

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI

FACULDADE DE CIÊNCIAS EXATAS

CURSO DE SISTEMAS DE INFORMAÇÃO

**DESCOBERTA DE CONHECIMENTO EM ÓBITOS PEDIÁTRICOS NA
REGIÃO DE DIAMANTINA**

Luiz Felipe Cordeiro Lopes

Diamantina

2014

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
FACULDADE DE CIÊNCIAS EXATAS

**DESCOBERTA DE CONHECIMENTO EM ÓBITOS PEDIÁTRICOS NA
REGIÃO DE DIAMANTINA**

Luiz Felipe Cordeiro Lopes

Orientador:

Marcus Vinícius Carvalho Guelpeli

Coorientadora:

Alison Cristine Pinto Guelpeli

Trabalho de Conclusão de Curso apresentado ao Programa de Graduação em Sistemas de Informação da Universidade Federal dos Vales do Jequitinhonha e Mucuri, requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Descoberta de Conhecimento.

Diamantina

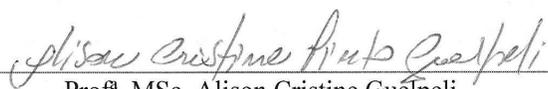
2014

Monografia de projeto final de graduação sob o título “Descoberta de Conhecimento em Óbitos Pediátricos na Região de Diamantina”, defendida por Luiz Felipe Cordeiro Lopes e aprovada em 25 de julho de 2014, em Diamantina, Minas Gerais.

Banca Examinadora:



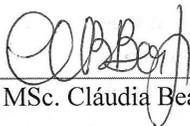
Prof. Dr. Marcus Vinícius Guelpeli
Orientador



Prof.ª MSc. Alison Cristine Guélpeli



Prof.ª Dr.ª Geruza Tomé Sabino



Prof.ª MSc. Cláudia Beatriz Berti

“Esforça-te e tem bom ânimo; não temas nem te espantes; porque o Senhor teu Deus é contigo, por onde quer que andares”.
Josué 1:9

AGRADECIMENTOS

Agradeço a Deus e Senhor Jesus Cristo pela Sua Fidelidade e pelo Seu Amor Incondicional; aos meus pais João e Fernanda pela dedicação, pelo apoio, pelo exemplo e todo o carinho transmitido; à minha irmã Renata pelo companheirismo e momentos de descontração; ao meu irmão Eduardo, que foi um dos grandes responsáveis por este momento. Te agradeço por todo incentivo aos meus estudos (essa conquista também é sua!). Ao HP, à família Revolução e toda a turma da GABADI (Vocês são demais!). Aos orientadores Prof.º Marcus Vinícius Guelpeli e Dra. Alison Guelpeli, que foram fundamentais para a realização desse trabalho. Aos meus colegas de curso, a todos os meus professores, funcionários, familiares, amigos e a todos aqueles que acreditaram nesta conquista. Divido este momento com todos vocês. Muito Obrigado!

RESUMO

Este trabalho aborda a descoberta de conhecimento em óbitos pediátricos na cidade de Diamantina-MG, através da mineração de textos, feita por meio da técnica de *Clusterização*. Os dados de óbitos pediátricos do ano de 2002 a 2012 foram coletados no Cartório de Registro Civil da cidade de Diamantina e para a realização do processamento foi utilizado o modelo Cassiopeia. A análise dos resultados obtidos após o processamento foi dividida em duas fases. A primeira exibiu os resultados adquiridos pelo modelo Cassiopeia, através de métricas textuais internas. A segunda fase dos experimentos, abordou a análise humana feita por uma especialista em pediatria que avaliou os *clusters* formados através da observação do relatório que contém os agrupamentos e as visualizações gráficas dos dados do *cluster*. Esses resultados são significativos para a descoberta de conhecimento, podendo servir como parâmetros na ajuda à tomada de decisões sobre fatores sociais e políticas públicas relacionadas à área da saúde da cidade de Diamantina.

Palavras-chave: descoberta de conhecimento, óbitos pediátricos, modelo Cassiopeia, cidade Diamantina, *clusterização*.

ABSTRACT

This paper discusses the knowledge discovery in pediatric deaths in the city of Diamantina, through text mining, taken through *Clustering* technique. Data from pediatric deaths of the year 2002-2012 were collected in the Civil Registry of the city of Diamantina, and for performing the processing, the Cassiopeia model was used. The results obtained after processing was divided into two phases, wherein the first displayed the results acquired by the Cassiopeia model through internal textual metrics. The second phase of the experiments, discussed the human analysis done by a pediatric specialist, which assessed the *Clusters* formed through observation of the report containing *Clusters* and graphical views of the data *Cluster*. These results are significant for knowledge discovery, which may serve as parameters to aid in making decisions about social factors and public policies related to health in the city of Diamantina.

Keywords: knowledge discovery, pediatric deaths, Cassiopeia, Diamantina city, *clustering* model.

LISTA DE FIGURAS

| | |
|---|----|
| Figura 1. Seleção dos atributos no modelo Cassiopeia (GUELPELI, 2012)..... | 33 |
| Figura 2. Planilhas com as informações de óbito. | 40 |
| Figura 3. Arquivos de texto com as informações de óbito que ocorreram no ano de 2002. | 41 |
| Figura 4. Arquivos de texto dos óbitos organizados por ano. | 42 |
| Figura 5. Tabela de um <i>cluster</i> com informações adicionais das crianças associadas ao óbito. | 44 |
| Figura 6. Quantidade de elementos agrupados em cada <i>cluster</i> | 47 |
| Figura 7. Índice de coesão em cada <i>cluster</i> | 47 |
| Figura 8. Valor do coeficiente Silhouette em cada <i>cluster</i> | 49 |
| Figura 9. Quantidade de óbitos por idade no <i>cluster</i> 01. | 52 |
| Figura 10. Quantidade de óbitos por naturalidade no <i>cluster</i> 01. | 53 |
| Figura 11. Quantidade de óbitos por mês no <i>cluster</i> 01. | 54 |
| Figura 12. Quantidade de óbitos por ano no <i>cluster</i> 01. | 55 |
| Figura 13. Quantidade de óbitos por mês/ano no <i>cluster</i> 02. | 56 |
| Figura 14. Quantidade de óbitos por ano no <i>cluster</i> 02. | 57 |
| Figura 15. Quantidade de óbitos por idade no <i>cluster</i> 03. | 59 |
| Figura 16. Quantidade de óbitos por ano no <i>cluster</i> 05. | 61 |
| Figura 17. Quantidade de óbitos por mês no <i>cluster</i> 09. | 64 |
| Figura 18. Quantidade de óbitos por sexo no <i>cluster</i> 01. | 82 |
| Figura 19. Quantidade de óbitos por idade no <i>cluster</i> 01. | 82 |
| Figura 20. Quantidade de óbitos por naturalidade no <i>cluster</i> 01. | 83 |
| Figura 21. Quantidade de óbitos por residente no <i>cluster</i> 01. | 83 |
| Figura 22. Quantidade de óbitos por mês no <i>cluster</i> 01. | 84 |
| Figura 23. Quantidade de óbitos por ano no <i>cluster</i> 01. | 84 |
| Figura 24. Quantidade de óbitos por sexo no <i>cluster</i> 02. | 86 |
| Figura 25. Quantidade de óbitos por idade no <i>cluster</i> 02. | 86 |

| | |
|--|-----|
| Figura 26. Quantidade de óbitos por naturalidade no <i>cluster</i> 02..... | 87 |
| Figura 27. Quantidade de óbitos por residente no <i>cluster</i> 02..... | 87 |
| Figura 28. Quantidade de óbitos por mês no <i>cluster</i> 02..... | 88 |
| Figura 29 Quantidade de óbitos por ano no <i>cluster</i> 02..... | 88 |
| Figura 30. Quantidade de óbitos por sexo no <i>cluster</i> 03..... | 90 |
| Figura 31. Quantidade de óbitos por idade no <i>cluster</i> 03..... | 91 |
| Figura 32. Quantidade de óbitos por naturalidade no <i>cluster</i> 03..... | 91 |
| Figura 33. Quantidade de óbitos por residente no <i>cluster</i> 03..... | 92 |
| Figura 34. Quantidade de óbitos por mês no <i>cluster</i> 03..... | 92 |
| Figura 35. Quantidade de óbitos por ano no <i>cluster</i> 03..... | 93 |
| Figura 36. Quantidade de óbitos por sexo no <i>cluster</i> 04..... | 94 |
| Figura 37. Quantidade de óbitos por idade no <i>cluster</i> 04..... | 94 |
| Figura 38. Quantidade de óbitos por naturalidade no <i>cluster</i> 04..... | 95 |
| Figura 39. Quantidade de óbitos por residente no <i>cluster</i> 04..... | 95 |
| Figura 40. Quantidade de óbitos por mês no <i>cluster</i> 04..... | 96 |
| Figura 41. Quantidade de óbitos por ano no <i>cluster</i> 04..... | 96 |
| Figura 42. Quantidade de óbitos por sexo no <i>cluster</i> 05..... | 98 |
| Figura 43. Quantidade de óbitos por idade no <i>cluster</i> 05..... | 98 |
| Figura 44. Quantidade de óbitos por naturalidade no <i>cluster</i> 05..... | 99 |
| Figura 45. Quantidade de óbitos por residente no <i>cluster</i> 05..... | 99 |
| Figura 46. Quantidade de óbitos por mês no <i>cluster</i> 05..... | 100 |
| Figura 47. Quantidade de óbitos por ano no <i>cluster</i> 05..... | 100 |
| Figura 48. Quantidade de óbitos por sexo no <i>cluster</i> 06..... | 102 |
| Figura 49. Quantidade de óbitos por idade no <i>cluster</i> 06..... | 102 |
| Figura 50. Quantidade de óbitos por naturalidade no <i>cluster</i> 06..... | 103 |
| Figura 51. Quantidade de óbitos por residente no <i>cluster</i> 06..... | 103 |
| Figura 52. Quantidade de óbitos por mês no <i>cluster</i> 06..... | 104 |

| | |
|--|-----|
| Figura 53. Quantidade de óbitos por ano no <i>cluster</i> 06..... | 104 |
| Figura 54. Quantidade de óbitos por sexo no <i>cluster</i> 07. | 106 |
| Figura 55. Quantidade de óbitos por idade no <i>cluster</i> 07..... | 106 |
| Figura 56. Quantidade de óbitos por naturalidade no <i>cluster</i> 07..... | 107 |
| Figura 57. Quantidade de óbitos por residente no <i>cluster</i> 07..... | 107 |
| Figura 58. Quantidade de óbitos por mês no <i>cluster</i> 07..... | 108 |
| Figura 59. Quantidade de óbitos por ano no <i>cluster</i> 07..... | 108 |
| Figura 60. Quantidade de óbitos por sexo no <i>cluster</i> 08. | 110 |
| Figura 61. Quantidade de óbitos por idade no <i>cluster</i> 08..... | 110 |
| Figura 62. Quantidade de óbitos por naturalidade no <i>cluster</i> 08..... | 111 |
| Figura 63. Quantidade de óbitos por residente no <i>cluster</i> 08..... | 111 |
| Figura 64. Quantidade de óbitos por mês no <i>cluster</i> 08..... | 112 |
| Figura 65. Quantidade de óbitos por ano no <i>cluster</i> 08..... | 112 |
| Figura 66. Quantidade de óbitos por sexo no <i>cluster</i> 09. | 114 |
| Figura 67. Quantidade de óbitos por idade no <i>cluster</i> 09..... | 114 |
| Figura 68. Quantidade de óbitos por naturalidade no <i>cluster</i> 09..... | 115 |
| Figura 69. Quantidade de óbitos por residente no <i>cluster</i> 09..... | 115 |
| Figura 70. Quantidade de óbitos por mês no <i>cluster</i> 09..... | 116 |
| Figura 71. Quantidade de óbitos por ano no <i>cluster</i> 09..... | 117 |
| Figura 72. Quantidade de óbitos por sexo no <i>cluster</i> 10. | 118 |
| Figura 73. Quantidade de óbitos por idade no <i>cluster</i> 10..... | 118 |
| Figura 74. Quantidade de óbitos por naturalidade no <i>cluster</i> 10..... | 119 |
| Figura 75. Quantidade de óbitos por residente no <i>cluster</i> 10..... | 119 |
| Figura 76. Quantidade de óbitos por mês no <i>cluster</i> 10..... | 120 |
| Figura 77. Quantidade de óbitos por ano no <i>cluster</i> 10..... | 120 |
| Figura 78. Quantidade de óbitos por sexo no <i>cluster</i> 11. | 122 |
| Figura 79. Quantidade de óbitos por idade no <i>cluster</i> 11..... | 122 |

| | |
|---|-----|
| Figura 80. Quantidade de óbitos por naturalidade no <i>cluster</i> 11..... | 123 |
| Figura 81. Quantidade de óbitos por residente no <i>cluster</i> 11..... | 123 |
| Figura 82. Quantidade de óbitos por mês no <i>cluster</i> 11..... | 124 |
| Figura 83. Quantidade de óbitos por ano no <i>cluster</i> 11..... | 124 |
| Figura 84. Quantidade de óbitos por sexo no <i>cluster</i> 12..... | 126 |
| Figura 85. Quantidade de óbitos por idade no <i>cluster</i> 12..... | 126 |
| Figura 86. Quantidade de óbitos por naturalidade no <i>cluster</i> 12..... | 127 |
| Figura 87. Quantidade de óbitos por residente no <i>cluster</i> 12..... | 127 |
| Figura 88. Quantidade de óbitos por mês no <i>cluster</i> 12..... | 128 |
| Figura 89. Quantidade de óbitos por ano no <i>cluster</i> 12..... | 128 |
| Figura 90. Quantidade de óbitos por sexo no <i>cluster</i> 13..... | 130 |
| Figura 91. Quantidade de óbitos por idade no <i>cluster</i> 13..... | 130 |
| Figura 92. Quantidade de óbitos por naturalidade no <i>cluster</i> 13..... | 131 |
| Figura 93. Quantidade de óbitos por residente no <i>cluster</i> 13..... | 131 |
| Figura 94. Quantidade de óbitos por mês no <i>cluster</i> 13..... | 132 |
| Figura 95. Quantidade de óbitos por ano no <i>cluster</i> 13..... | 132 |
| Figura 96. Quantidade de óbitos por sexo no <i>cluster</i> 14..... | 134 |
| Figura 97. Quantidade de óbitos por idade no <i>cluster</i> 14..... | 134 |
| Figura 98. Quantidade de óbitos por naturalidade no <i>cluster</i> 14..... | 135 |
| Figura 99. Quantidade de óbitos por residente no <i>cluster</i> 14..... | 135 |
| Figura 100. Quantidade de óbitos por mês no <i>cluster</i> 14..... | 136 |
| Figura 101. Quantidade de óbitos por ano no <i>cluster</i> 14..... | 136 |
| Figura 102. Quantidade de óbitos por sexo no <i>cluster</i> 15..... | 138 |
| Figura 103. Quantidade de óbitos por idade no <i>cluster</i> 15..... | 138 |
| Figura 104. Quantidade de óbitos por naturalidade no <i>cluster</i> 15..... | 139 |
| Figura 105. Quantidade de óbitos por residente no <i>cluster</i> 15..... | 139 |
| Figura 106. Quantidade de óbitos por mês no <i>cluster</i> 15..... | 140 |

| | |
|---|-----|
| Figura 107. Quantidade de óbitos por ano no <i>cluster</i> 15..... | 140 |
| Figura 108. Quantidade de óbitos por sexo no <i>cluster</i> 16. | 142 |
| Figura 109. Quantidade de óbitos por idade no <i>cluster</i> 16..... | 142 |
| Figura 110. Quantidade de óbitos por naturalidade no <i>cluster</i> 16..... | 143 |
| Figura 111. Quantidade de óbitos por residente no <i>cluster</i> 16..... | 143 |
| Figura 112. Quantidade de óbitos por mês no <i>cluster</i> 16..... | 144 |
| Figura 113. Quantidade de óbitos por ano no <i>cluster</i> 16..... | 144 |
| Figura 114. Quantidade de óbitos por sexo no <i>cluster</i> 17. | 146 |
| Figura 115. Quantidade de óbitos por idade no <i>cluster</i> 17..... | 146 |
| Figura 116. Quantidade de óbitos por naturalidade no <i>cluster</i> 17..... | 147 |
| Figura 117. Quantidade de óbitos por residente no <i>cluster</i> 17..... | 147 |
| Figura 118. Quantidade de óbitos por mês no <i>cluster</i> 17..... | 148 |
| Figura 119. Quantidade de óbitos por ano no <i>cluster</i> 17..... | 148 |
| Figura 120. Quantidade de óbitos por sexo no <i>cluster</i> 18. | 150 |
| Figura 121. Quantidade de óbitos por idade no <i>cluster</i> 18..... | 150 |
| Figura 122. Quantidade de óbitos por naturalidade no <i>cluster</i> 18..... | 151 |
| Figura 123. Quantidade de óbitos por residente no <i>cluster</i> 18..... | 151 |
| Figura 124. Quantidade de óbitos por mês no <i>cluster</i> 18..... | 152 |
| Figura 125. Quantidade de óbitos por ano no <i>cluster</i> 18..... | 152 |
| Figura 126. Quantidade de óbitos por sexo no <i>cluster</i> 19. | 154 |
| Figura 127. Quantidade de óbitos por idade no <i>cluster</i> 19..... | 154 |
| Figura 128. Quantidade de óbitos por naturalidade no <i>cluster</i> 19..... | 155 |
| Figura 129. Quantidade de óbitos por residente no <i>cluster</i> 19..... | 155 |
| Figura 130. Quantidade de óbitos por mês no <i>cluster</i> 19..... | 156 |
| Figura 131. Quantidade de óbitos por ano no <i>cluster</i> 19..... | 156 |
| Figura 132. Quantidade de óbitos por sexo no <i>cluster</i> 20. | 158 |
| Figura 133. Quantidade de óbitos por idade no <i>cluster</i> 20..... | 158 |

| | |
|---|-----|
| Figura 134. Quantidade de óbitos por naturalidade no <i>cluster</i> 20..... | 159 |
| Figura 135. Quantidade de óbitos por residente no <i>cluster</i> 20..... | 159 |
| Figura 136. Quantidade de óbitos por mês no <i>cluster</i> 20..... | 160 |
| Figura 137. Quantidade de óbitos por ano no <i>cluster</i> 20..... | 160 |
| Figura 138. Quantidade de óbitos por sexo no <i>cluster</i> 21..... | 162 |
| Figura 139. Quantidade de óbitos por idade no <i>cluster</i> 21..... | 162 |
| Figura 140. Quantidade de óbitos por naturalidade no <i>cluster</i> 21..... | 163 |
| Figura 141. Quantidade de óbitos por residente no <i>cluster</i> 21..... | 163 |
| Figura 142. Quantidade de óbitos por mês no <i>cluster</i> 21..... | 164 |
| Figura 143. Quantidade de óbitos por ano no <i>cluster</i> 21..... | 164 |
| Figura 144. Quantidade de óbitos por sexo no <i>cluster</i> 22..... | 166 |
| Figura 145. Quantidade de óbitos por idade no <i>cluster</i> 22..... | 166 |
| Figura 146. Quantidade de óbitos por naturalidade no <i>cluster</i> 22..... | 167 |
| Figura 147. Quantidade de óbitos por residente no <i>cluster</i> 22..... | 167 |
| Figura 148. Quantidade de óbitos por mês no <i>cluster</i> 22..... | 168 |
| Figura 149. Quantidade de óbitos por ano no <i>cluster</i> 22..... | 168 |
| Figura 150. Quantidade de óbitos por sexo no <i>cluster</i> 23..... | 170 |
| Figura 151. Quantidade de óbitos por idade no <i>cluster</i> 23..... | 170 |
| Figura 152. Quantidade de óbitos por naturalidade no <i>cluster</i> 23..... | 171 |
| Figura 153. Quantidade de óbitos por residente no <i>cluster</i> 23..... | 171 |
| Figura 154. Quantidade de óbitos por mês no <i>cluster</i> 23..... | 172 |
| Figura 155. Quantidade de óbitos por ano no <i>cluster</i> 23..... | 172 |
| Figura 156. Quantidade de óbitos por sexo no <i>cluster</i> 24..... | 174 |
| Figura 157. Quantidade de óbitos por idade no <i>cluster</i> 24..... | 174 |
| Figura 158. Quantidade de óbitos por naturalidade no <i>cluster</i> 24..... | 175 |
| Figura 159. Quantidade de óbitos por residente no <i>cluster</i> 24..... | 175 |
| Figura 160. Quantidade de óbitos por mês no <i>cluster</i> 24..... | 176 |

| | |
|--|-----|
| Figura 161. Quantidade de óbitos por ano no <i>cluster 24</i> | 176 |
| Figura 162. Quantidade de óbitos por sexo no <i>cluster 25</i> | 178 |
| Figura 163. Quantidade de óbitos por idade no <i>cluster 25</i> | 178 |
| Figura 164. Quantidade de óbitos por naturalidade no <i>cluster 25</i> | 179 |
| Figura 165. Quantidade de óbitos por residente no <i>cluster 25</i> | 179 |
| Figura 166. Quantidade de óbitos por mês no <i>cluster 25</i> | 180 |
| Figura 167. Quantidade de óbitos por ano no <i>cluster 25</i> | 180 |
| Figura 168. Quantidade de óbitos por sexo no <i>cluster 26</i> | 182 |
| Figura 169. Quantidade de óbitos por idade no <i>cluster 26</i> | 182 |
| Figura 170. Quantidade de óbitos por naturalidade no <i>cluster 26</i> | 183 |
| Figura 171. Quantidade de óbitos por residente no <i>cluster 26</i> | 183 |
| Figura 172. Quantidade de óbitos por mês no <i>cluster 26</i> | 184 |
| Figura 173. Quantidade de óbitos por ano no <i>cluster 26</i> | 184 |
| Figura 174. Quantidade de óbitos por sexo no <i>cluster 27</i> | 186 |
| Figura 175. Quantidade de óbitos por idade no <i>cluster 27</i> | 186 |
| Figura 176. Quantidade de óbitos por naturalidade no <i>cluster 27</i> | 187 |
| Figura 177. Quantidade de óbitos por residente no <i>cluster 27</i> | 187 |
| Figura 178. Quantidade de óbitos por mês no <i>cluster 27</i> | 188 |
| Figura 179. Quantidade de óbitos por ano no <i>cluster 27</i> | 188 |
| Figura 180. Quantidade de óbitos por sexo no <i>cluster 28</i> | 190 |
| Figura 181. Quantidade de óbitos por idade no <i>cluster 28</i> | 190 |
| Figura 182. Quantidade de óbitos por naturalidade no <i>cluster 28</i> | 191 |
| Figura 183. Quantidade de óbitos por residente no <i>cluster 28</i> | 191 |
| Figura 184. Quantidade de óbitos por mês no <i>cluster 28</i> | 192 |
| Figura 185. Quantidade de óbitos por ano no <i>cluster 28</i> | 192 |
| Figura 186. Quantidade de óbitos por sexo no <i>cluster 29</i> | 194 |
| Figura 187. Quantidade de óbitos por idade no <i>cluster 29</i> | 194 |

| | |
|---|-----|
| Figura 188. Quantidade de óbitos por naturalidade no <i>cluster</i> 29..... | 195 |
| Figura 189. Quantidade de óbitos por residente no <i>cluster</i> 29..... | 195 |
| Figura 190. Quantidade de óbitos por mês no <i>cluster</i> 29..... | 196 |
| Figura 191. Quantidade de óbitos por ano no <i>cluster</i> 29..... | 196 |
| Figura 192. Quantidade de óbitos por sexo no <i>cluster</i> 30..... | 198 |
| Figura 193. Quantidade de óbitos por idade no <i>cluster</i> 30..... | 198 |
| Figura 194. Quantidade de óbitos por naturalidade no <i>cluster</i> 30..... | 199 |
| Figura 195. Quantidade de óbitos por residente no <i>cluster</i> 30..... | 199 |
| Figura 196. Quantidade de óbitos por mês no <i>cluster</i> 30..... | 200 |
| Figura 197. Quantidade de óbitos por ano no <i>cluster</i> 30..... | 200 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 1. Estatísticas dos arquivos de texto do campo Óbito e Causa de Óbito. | 43 |
| Tabela 2. <i>Cluster</i> 12 gerado na clusterização realizada pelo modelo Cassiopeia. | 48 |
| Tabela 3. <i>Cluster</i> 08 gerado na clusterização realizada pelo modelo Cassiopeia. | 48 |
| Tabela 4. <i>Cluster</i> 01 gerado na clusterização realizada pelo modelo Cassiopeia. | 51 |
| Tabela 5. <i>Cluster</i> 02 gerado na clusterização realizada pelo modelo Cassiopeia. | 56 |
| Tabela 6. <i>Cluster</i> 03 gerado na clusterização realizada pelo modelo Cassiopeia. | 58 |
| Tabela 7. <i>Cluster</i> 05 gerado na clusterização realizada pelo modelo Cassiopeia. | 60 |
| Tabela 8. <i>Cluster</i> 07 gerado na clusterização realizada pelo modelo Cassiopeia. | 62 |
| Tabela 9. <i>Cluster</i> 09 gerado na clusterização realizada pelo modelo Cassiopeia. | 63 |
| Tabela 10. <i>Cluster</i> 15 gerado na clusterização realizada pelo modelo Cassiopeia. | 64 |
| Tabela 11. <i>Cluster</i> 21 gerado na clusterização realizada pelo modelo Cassiopeia. | 65 |
| Tabela 12. <i>Cluster</i> 24 gerado na clusterização realizada pelo modelo Cassiopeia. | 66 |
| Tabela 13. <i>Cluster</i> 25 gerado na clusterização realizada pelo modelo Cassiopeia. | 66 |
| Tabela 14. Tabela comparativa dos trabalhos correlatos..... | 69 |
| Tabela 15. <i>Cluster</i> 01 gerado na clusterização realizada pelo modelo Cassiopeia. | 81 |
| Tabela 16. <i>Cluster</i> 02 gerado na clusterização realizada pelo modelo Cassiopeia. | 85 |
| Tabela 17. <i>Cluster</i> 03 gerado na clusterização realizada pelo modelo Cassiopeia. | 89 |
| Tabela 18. <i>Cluster</i> 04 gerado na clusterização realizada pelo modelo Cassiopeia. | 93 |
| Tabela 19. <i>Cluster</i> 05 gerado na clusterização realizada pelo modelo Cassiopeia. | 97 |
| Tabela 20. <i>Cluster</i> 06 gerado na clusterização realizada pelo modelo Cassiopeia. | 101 |
| Tabela 21. <i>Cluster</i> 07 gerado na clusterização realizada pelo modelo Cassiopeia. | 105 |
| Tabela 22. <i>Cluster</i> 08 gerado na clusterização realizada pelo modelo Cassiopeia. | 109 |
| Tabela 23. <i>Cluster</i> 09 gerado na clusterização realizada pelo modelo Cassiopeia. | 113 |
| Tabela 24. <i>Cluster</i> 10 gerado na clusterização realizada pelo modelo Cassiopeia. | 117 |
| Tabela 25. <i>Cluster</i> 11 gerado na clusterização realizada pelo modelo Cassiopeia. | 121 |
| Tabela 26. <i>Cluster</i> 12 gerado na clusterização realizada pelo modelo Cassiopeia. | 125 |

| | |
|---|-----|
| Tabela 27. <i>Cluster</i> 13 gerado na clusterização realizada pelo modelo Cassiopeia. | 129 |
| Tabela 28. <i>Cluster</i> 14 gerado na clusterização realizada pelo modelo Cassiopeia. | 133 |
| Tabela 29. <i>Cluster</i> 15 gerado na clusterização realizada pelo modelo Cassiopeia. | 137 |
| Tabela 30. <i>Cluster</i> 16 gerado na clusterização realizada pelo modelo Cassiopeia. | 141 |
| Tabela 31. <i>Cluster</i> 17 gerado na clusterização realizada pelo modelo Cassiopeia. | 145 |
| Tabela 32. <i>Cluster</i> 18 gerado na clusterização realizada pelo modelo Cassiopeia. | 149 |
| Tabela 33. <i>Cluster</i> 19 gerado na clusterização realizada pelo modelo Cassiopeia. | 153 |
| Tabela 34. <i>Cluster</i> 20 gerado na clusterização realizada pelo modelo Cassiopeia. | 157 |
| Tabela 35. <i>Cluster</i> 21 gerado na clusterização realizada pelo modelo Cassiopeia. | 161 |
| Tabela 36. <i>Cluster</i> 22 gerado na clusterização realizada pelo modelo Cassiopeia. | 165 |
| Tabela 37. <i>Cluster</i> 23 gerado na clusterização realizada pelo modelo Cassiopeia. | 169 |
| Tabela 38. <i>Cluster</i> 24 gerado na clusterização realizada pelo modelo Cassiopeia. | 173 |
| Tabela 39. <i>Cluster</i> 25 gerado na clusterização realizada pelo modelo Cassiopeia. | 177 |
| Tabela 40. <i>Cluster</i> 26 gerado na clusterização realizada pelo modelo Cassiopeia. | 181 |
| Tabela 41. <i>Cluster</i> 27 gerado na clusterização realizada pelo modelo Cassiopeia. | 185 |
| Tabela 42. <i>Cluster</i> 28 gerado na clusterização realizada pelo modelo Cassiopeia. | 189 |
| Tabela 43. <i>Cluster</i> 29 gerado na clusterização realizada pelo modelo Cassiopeia. | 193 |
| Tabela 44. <i>Cluster</i> 30 gerado na clusterização realizada pelo modelo Cassiopeia. | 197 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|--------|--|
| AIDS | <i><u>A</u>quired <u>I</u>mmune de <u>D</u>eficiency <u>S</u>ndrome</i> |
| CKB | <i><u>C</u>onceptual <u>K</u>nowledge <u>B</u>ase</i> |
| CRM | <i><u>C</u>onselho <u>R</u>egional de <u>M</u>edicina</i> |
| DCD | <i><u>D</u>escoberta de <u>C</u>onhecimento em <u>D</u>ados</i> |
| DCT | <i><u>D</u>escoberta de <u>C</u>onhecimento em <u>T</u>extos</i> |
| HTML | <i><u>H</u>iper<u>T</u>ext <u>M</u>arkup <u>L</u>anguage</i> |
| MD | <i><u>M</u>ineração de <u>D</u>ados</i> |
| MLP | <i><u>P</u>erceptron de <u>M</u>últiplas <u>C</u>amadas</i> |
| MG | <i><u>M</u>inas <u>G</u>erais</i> |
| MT | <i><u>M</u>ineração de <u>T</u>extos</i> |
| PCA | <i><u>A</u>nálise de <u>C</u>omponentes <u>P</u>incipais</i> |
| RI | <i><u>R</u>ecuperação de <u>I</u>nformação</i> |
| SOM | <i><u>S</u>elf - <u>O</u>rganization <u>M</u>ap</i> |
| SP | <i><u>S</u>ão <u>P</u>aulo</i> |
| UNICEF | <i><u>F</u>undo das <u>N</u>ações <u>U</u>nidas para a <u>I</u>nfância</i> |

SUMÁRIO

| | |
|---|----|
| CAPÍTULO 1 – INTRODUÇÃO | 17 |
| 1.1 MOTIVAÇÃO | 20 |
| 1.2 PROBLEMA..... | 20 |
| 1.3 HIPÓTESE..... | 20 |
| 1.4 CONTRIBUIÇÕES | 20 |
| 1.5 METODOLOGIA DE PESQUISA | 21 |
| 1.6 ESTRUTURA DA PROPOSTA..... | 21 |
| CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA..... | 23 |
| 2.1 DESCOBERTA DE CONHECIMENTO | 23 |
| 2.1.1 DESCOBERTA DE CONHECIMENTO NA ÁREA DA SAÚDE..... | 27 |
| 2.2 ETAPAS DA DESCOBERTA DE CONHECIMENTO EM TEXTO | 29 |
| 2.2.1 PRÉ-PROCESSAMENTO | 29 |
| 2.2.2 PROCESSAMENTO | 30 |
| 2.2.3 PÓS-PROCESSAMENTO | 36 |
| CAPÍTULO 3 – METODOLOGIA | 39 |
| 3.1 REGISTRO DE ÓBITOS: DESCRIÇÃO DA BASE DE DADOS (PRÉ-PROCESSAMENTO)..... | 39 |
| 3.2 INSERÇÃO DOS ARQUIVOS DE TEXTO NO MODELO CASSIOPEIA (PROCESSAMENTO) | 42 |
| 3.3 VISUALIZAÇÃO DOS RESULTADOS (PÓS-PROCESSAMENTO)..... | 43 |
| CAPÍTULO 4 – RESULTADOS | 46 |
| 4.1 ANÁLISE DAS MÉTRICAS TEXTUAIS..... | 46 |
| 4.2 ANÁLISE HUMANA DOS <i>CLUSTERS</i> FORMADOS | 49 |
| 4.3 COMPROVAÇÃO DA HIPÓTESE..... | 67 |
| 4.4 TRABALHOS CORRELATOS | 67 |
| CAPÍTULO 5 – CONCLUSÕES | 70 |
| 5.1 LIMITAÇÕES | 72 |
| 5.2 TRABALHOS FUTUROS | 73 |
| REFERÊNCIAS | 74 |
| APÊNDICES | 77 |

CAPÍTULO 1 – INTRODUÇÃO

Cada vez mais a informação vem sendo um fator determinante na atualidade, tornando-se um bem precioso para obtenção de conhecimento. A informação pode ser usufruída pelas instituições e empresas para descobrir conhecimento, e conseqüentemente identificar fatores que influenciam na organização, auxiliando na criação de melhorias ou soluções para um problema encontrado.

Segundo Fialho *et al.* (2006), a informação é o conjunto de dados que são devidamente processados e tornam-se compreensíveis, ou seja, é a disposição dos dados de uma forma que apresentem um significado.

O conhecimento é a combinação da informação e das relações que levam os indivíduos à tomada de decisão e ao desenvolvimento de novas informações ou conhecimentos (FIALHO *et al.*, 2006).

Com o constante crescimento das informações, é necessário que se tenham técnicas que possam auxiliar na análise das mesmas, para que sejam descobertos conhecimentos através das informações existentes. A partir desta necessidade surgiu a área da descoberta de conhecimento.

Esta área consiste no processo de identificação de informações relevantes que podem ser computadas e agregadas ao conhecimento existente, mudando o estado do conhecimento atual para um novo estado, a fim de criar soluções para determinada situação ou problema (WIVES, 2004), (MATHERN *et al.*, 2012).

A descoberta de conhecimento pode ser classificada a partir do conteúdo a ser analisado. Se as informações tratadas são adquiridas a partir de dados previamente organizados e estruturados, ele é chamado de Descoberta de Conhecimento em Dados – DCD, caso as informações tratadas sejam encontradas em dados semiestruturados ou desestruturados, como documentos textuais, ele é classificado como Descoberta de Conhecimento em Texto – DCT (RAMOS e BRASCHER, 2009).

Os dados passam por três etapas, sendo que a primeira é a fase de pré-processamento, em que ocorre a preparação e organização dos dados. Esta é uma fase importante para a descoberta de conhecimento, pois a partir dela, os dados serão consistentes e poderão gerar um bom desempenho nas próximas fases.

A próxima etapa é a fase de processamento, que, como o próprio nome já diz, consiste em processar os textos que foram preparados e organizados na fase de pré-processamento. Esta etapa possui algoritmos que executam técnicas para realizar o processamento dos textos, em que métricas computacionais são geradas para fins de avaliação do processamento que foi realizado.

E finalmente, na fase de pós-processamento, os resultados que foram obtidos são analisados para identificar tendências e gerar um novo estado de conhecimento, com o auxílio de ilustrações ou análises gráficas que facilitam a avaliação dos documentos processados.

Essas informações, identificadas e trabalhadas na descoberta de conhecimento, podem ser de várias áreas, e neste trabalho abordará informações de óbito. Diversas causas de óbito podem ocorrer por vários fatores em comum como o clima, a ocorrência de doenças predominantes em uma determinada região, a época do ano e o índice de criminalidade.

A relação entre descoberta de conhecimento e causas de óbito se baseia na ideia de que óbitos são indicadores de saúde sociais e econômicos de uma população. A partir dos dados de óbitos, é possível diagnosticar diversos tipos de problemas que afetam uma cidade. Nesse contexto, a descoberta de conhecimento se torna uma ferramenta útil, podendo auxiliar na identificação e na criação de soluções para a resolução destes problemas (VIANNA e BARRA, 2005).

Como descoberta de conhecimento em casos de óbitos, destacam-se os óbitos pediátricos, pois óbitos nessa faixa etária demonstram um sinal dos problemas existentes em uma região, como a violência doméstica, a prematuridade extrema, a falta de saneamento básico, a exploração infantil, entre outros fatores sociais. (CIAMPO *et al.* 1997), (GUIMARÃES *et al.*, 2003), (VICTORA, 2009), (VIANNA *et al.*, 2010).

As fontes com estes tipos de informações, geralmente são encontradas em hospitais e nos próprios cartórios de registro civil das cidades do país, e não é diferente na cidade de Diamantina-MG. Porém, é impossível adquirir conhecimento dessas informações na forma em que elas são organizadas, uma que vez que não há como obter padrões que possam identificar novos conhecimentos, devido ao fato das informações

de óbitos não serem organizadas por temas, mas apenas por ano, sem nenhuma outra classificação.

Baseando-se nesse problema, este trabalho visa descobrir conhecimento através da padronização das informações dos registros de óbitos que ocorreram em crianças de 0 a 16 anos, nos anos de 2002 a 2012 na cidade de Diamantina - MG.

O motivo da seleção de uma faixa de idade é a existência de diversas doenças predominantes em uma determinada etapa da vida, pois crianças possuem causas de óbitos características da sua própria faixa etária, assim como um adulto ou um idoso também possuem causas de óbitos características para os seus períodos e condições de vida. Logo, a seleção de uma faixa etária proporciona um maior foco à pesquisa. Já o período de tempo escolhido, é estabelecido para que se possam ter informações suficientes para um estudo e comparação dos resultados obtidos dentro da faixa determinada.

O trabalho visa utilizar procedimentos de DCT que identificam informações relevantes de uma determinada quantidade de textos, para a descoberta de padrões, uma vez que os dados se encontram em formas textuais. A padronização das informações de óbitos será realizada na etapa de processamento por meio da técnica de agrupamento de textos, ou clusterização que consiste em gerar agrupamentos (*clusters*), que possuem informações similares entre si.

O agrupamento de textos é feito pelo modelo Cassiopeia que é um modelo desenvolvido por Guelpeli (2012), com o intuito de realizar agrupamentos textuais em domínios distintos, independentes do idioma, e com maior qualidade nas avaliações, que são feitas pelas métricas de análise textuais externas e internas.

Os resultados do processamento do modelo Cassiopeia são apresentados em duas fases. Primeiramente foi realizada a análise das métricas textuais que consiste na análise e avaliação das métricas internas (coesão, acoplamento e coeficiente Silhouette), geradas na fase de processamento. Já a segunda fase consiste na análise humana dos agrupamentos que foram formados, em que a especialista em pediatria, a Dra. Alison Guelpeli (CRM: 58624-MG), é responsável pela análise do relatório, contendo a descrição e ilustração das informações de cada *cluster* criado, tornando possível a geração de novos conhecimentos.

1.1 MOTIVAÇÃO

A descoberta de conhecimento de óbitos pediátricos auxilia na identificação de problemas existentes na sociedade e contribui para a reflexão e proposição de políticas públicas.

Desta forma, a motivação para este trabalho concentra-se na descoberta de conhecimento nos óbitos pediátricos ocorrentes na cidade de Diamantina e região, para identificar o que pode ter ocasionado esses óbitos, quais os problemas sociais que podem ser identificados na cidade e quais sugestões de encaminhamentos são passíveis de serem providenciados para prevenir futuros problemas dessa natureza.

Tais atitudes podem trazer resultados consideráveis na redução do índice de óbitos entre crianças da região e podem melhorar os recursos de tratamento e prevenções disponibilizados pela equipe médica a serviço da população diamantinense.

1.2 PROBLEMA

O problema deste trabalho consiste na inexistência de padrões em textos, dificultando a descoberta de conhecimento em informações de óbitos de crianças, com faixa etária entre zero a 16 anos, em um espaço de tempo de 2002 a 2012.

1.3 HIPÓTESE

A hipótese deste trabalho afirma que é possível descobrir conhecimento através dos óbitos pediátricos na cidade de Diamantina e região, por meio da técnica de agrupamento de textos, analisando os dados a partir do ano de 2002 até 2012.

1.4 CONTRIBUIÇÕES

As possíveis contribuições deste trabalho são:

- A capacidade de análise dos óbitos pediátricos como um todo e a descoberta de conhecimento por meio do agrupamento dos textos;
- A criação de dados estatísticos baseados nas tendências e padrões descobertos, para que sejam utilizados por instituições de pesquisa;
- Apoio a criação de políticas públicas de saúde na sociedade diamantinense, podendo ser aplicadas em hospitais, em postos de saúde, nas escolas e em outras instituições;

- Contribuição para o processo de prevenção de doenças que ocorrem em épocas específicas do ano;
- O desenvolvimento de campanhas direcionadas a determinadas regiões de Diamantina;
- A diminuição e controle dos índices de óbitos na cidade;
- O auxílio para a solicitação e requisição de aparelhos médicos, que poderão auxiliar em hospitais no tratamento e prevenção das doenças descobertas pela análise dos óbitos.

1.5 METODOLOGIA DE PESQUISA

A metodologia foi dividida em três passos: primeiramente foi abordada a fase de pré-processamento, em que foi descrito como foi feita a coleta dos dados de óbitos pediátricos no Cartório de registro Civil de Diamantina, seguido da preparação e organização dos textos e a criação de uma base de dados com as informações dos óbitos.

O segundo passo, consiste na fase de processamento, em que é descrito como foi feito a inserção dos arquivos de texto no modelo Cassiopeia e como estes dados foram processados. No terceiro passo, foi abordada a fase de pós-processamento, em que consiste em apresentar como foram feitas as visualizações dos resultados.

Na fase de pós-processamento, os *clusters* formados são inseridos em novas bases de dados e associados com as informações de óbitos, referentes a cada registro. Foram criados gráficos que demonstram de forma mais explícita os resultados obtidos com as clusterizações dos textos, além da criação de um relatório com informações detalhadas de cada *cluster*, e suas visualizações gráficas que auxiliaram a especialista em pediatria a descobrir novos conhecimentos.

1.6 ESTRUTURA DA PROPOSTA

Capítulo 2 – Fundamentação Teórica

Neste capítulo serão apresentados conceitos de Descoberta de Conhecimento, o seu processo evolutivo e a sua aplicação na área da saúde. Também serão abordados os conceitos de Descoberta de Conhecimento em Dados – DCD, Descoberta de Conhecimento em Textos – DCT, suas etapas (pré-processamento, processamento e pós-processamento), e suas técnicas (agrupamento de textos, sumarização, entre outros), além das métricas internas e externas de análise textual.

Capítulo 3 – Metodologia

Neste capítulo será abordada a metodologia do trabalho, descrevendo todo o processo de coleta e preparação dos dados, o processamento das informações de órbitas, e a criação de ilustrações para a visualização dos resultados do processamento e da análise humana.

Capítulo 4 – Resultados

O capítulo exibirá e discutirá os resultados obtidos através do modelo Cassiopeia e da análise humana. Também será abordado a comprovação da hipótese e os trabalhos correlatos.

Capítulo 5 – Conclusões

Neste capítulo serão feitas as conclusões, abordando as limitações do trabalho e os trabalhos futuros a serem realizados.

CAPÍTULO 2 – FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados os conceitos do estado da arte para o desenvolvimento deste trabalho. Inicialmente será abordado o conceito de descoberta de conhecimento, mostrando o seu processo evolutivo, enfatizando os tipos de descoberta de conhecimento. Também será abordada a descoberta de conhecimento na área da saúde, mais especificamente em óbitos pediátricos que é a delimitação deste trabalho, e por fim, serão apresentadas as etapas de pré-processamento, processamento com suas métricas de análise e pós-processamento.

2.1 DESCOBERTA DE CONHECIMENTO

Segundo Wives (2004) e Mathern *et al.* (2012), descoberta de conhecimento significa a identificação e o recebimento de informações relevantes que podem ser computadas e agregadas ao conhecimento prévio, mudando o estado do conhecimento atual, a fim de que determinada situação ou problema possa ser resolvido.

Descobrir conhecimento é encontrar a combinação de informações, dados e relações que levam os indivíduos à tomada de decisão, ao desenvolvimento de novas informações ou conhecimento, e à realização de tarefas como resposta ao conhecimento adquirido (FIALHO *et al.*, 2006 *apud* CECI, 2010).

Uma pessoa adquire novos conhecimentos a partir de experiências, valores e informações contextuais, a qual se apoia como base para a avaliação e a incorporação de novas experiências e informações (DAVENPORT e PRUSAK, 1998 *apud* LOPES, 2002).

O processo de descoberta de conhecimento é extremamente importante no contexto mundial atual e globalizado, pois a informação tornou-se cada vez mais importante para as organizações, sendo considerado um bem valioso. Logo, a informação faz com que o conhecimento seja fundamental para o sucesso das empresas e instituições (PINHEIRO, 2008 *apud* CECI, 2010).

O maior problema é que tomar decisões estratégicas, em um ambiente globalizado, formado por instituições e empresas, exige uma grande quantidade de informações. E quanto maior a quantidade de informações, mais difícil se torna a recuperação ou a identificação de informações que possuam relevância e a análise para descobrir novos conhecimentos (WIVES, 2004).

Para que essas grandes quantidades de informações possam ser recuperadas e processadas, é necessário aplicar uma metodologia que permita classificar, reconhecer e reutilizar os recursos de conhecimento existente nas organizações (VASCONCELOS, ROCHA e KIMBLE, 2003 *apud* CECI, 2010).

As primeiras aplicações de descoberta de conhecimento foram em base de dados de empresas, devido a grande quantidade de armazenamento de informações existentes nestes ambientes. A análise humana destes dados era inviável sem o auxílio de ferramentas computacionais apropriadas, tornando necessário o desenvolvimento de ferramentas que auxiliam o ser humano, de forma automática e inteligente, na tarefa de análise, interpretação e relação dos dados para a tomada de decisões estratégicas em cada contexto de aplicação (GOLDSCHIMIDT e PASSOS, 2005 *apud* BOENTE, ESTRELA, GOLDSCHIMIDT, 2006).

A partir dessa necessidade surgiu a área de *Descoberta de Conhecimento a partir de Dados – DCD* que é a área que busca descobrir co-relacionamentos e dados implícitos nos registros de um banco de dados, estudando e desenvolvendo um processo de extração de conhecimento novo, útil e interessante, para apresentá-lo de alguma forma acessível para o usuário (FAYYAD, 1996). Essa área permitiu que houvesse uma melhor utilização das informações nos bancos de dados, utilizando os mesmos com maior eficácia.

Segundo Barion e Lago (2008), as fases da DCD são:

- a) **Dados:** visualiza o tipo de conhecimento a ser extraído procurando primeiramente uma compreensão do domínio para posterior aplicação;
- b) **Seleção:** cria um conjunto ou um subconjunto de dados que será o foco da descoberta de novos conhecimentos;
- c) **Pré-Processamento:** realiza a limpeza dos dados;
- d) **Transformação:** os dados são armazenados de forma correta para facilitar o processamento dos dados, visando agregar valor semântico às informações;
- e) **Mineração de Dados:** seleção dos métodos a serem utilizados para localizar padrões através da descoberta de conhecimento sobre os dados;

- f) **Interpretação e Avaliação:** fase na qual o resultado da descoberta de conhecimento é exibido ao usuário, podendo ser analisado e interpretado.

Apesar da DCD, com o crescente uso de computadores, cada vez mais documentos eletrônicos estão sendo armazenados e colocados à disposição das pessoas. Em sua grande maioria, estes documentos, contêm informações codificadas na forma textual, tais como dicionários, manuais, enciclopédias, guias, mensagens e correio eletrônico. Estudos afirmam que mais de 80% das informações disponíveis estão no formato textual, podendo ser semiestruturado ou desestruturado (TAN, 1999).

A área que consiste em recuperar essas informações em formato de texto para serem acessadas por computador, é a Recuperação de Informação – RI.

A RI é a área de pesquisa baseada na ideia de obter resposta em meio às perguntas, utilizando sistemas que recuperam documentos com base em palavras-chave, por meio de métodos estatísticos usados para o processamento automático de dados de texto e comparação com a pergunta dada (HOTHO, NURNBERGER e PAAB, 2005).

Entretanto, encontrar tal conhecimento em informações com formatos textuais é uma tarefa árdua. Existem técnicas e ferramentas para RI, as quais auxiliam as pessoas a encontrarem documentos que contenham informações relevantes (SPARCK-JONES, 1997).

Esse tipo de informação textual não é tratada somente pelas ferramentas tradicionais de descoberta de conhecimento, pois possuem características que tornam sua análise complexa (ZANASI, 1998). Logo, a evolução da área de RI teve como consequência o surgimento da Descoberta de Conhecimento em Textos - DCT. O termo foi utilizado pela primeira vez por Feldman e Dagan (1995) para designar o processo de encontrar algo interessante em coleções de textos (artigos, histórias de revistas e jornais, mensagens de e-mail, páginas Web.) (LOH, 2001).

Segundo Tan (1999), DCT refere-se ao processo de extração de padrões interessantes e não triviais, ou conhecimento a partir de documentos de texto. É um método que abrange as áreas de RI, aprendizagem de máquina, estatística, linguística computacional e mineração de dados (HOTHO, NURNBERGER e PAAB, 2005).

A DCT visa descobrir conhecimento em textos que geram alguma informação relevante através dos dados recuperados por alguma forma de processamento. Ela vem solucionar grande parte dos problemas relacionados à busca, recuperação e análise de informações textuais.

A sobrecarga de informação, um dos maiores conflitos enfrentados pelos usuários da Internet, é um desses problemas. A grande quantidade de documentos torna impossível de serem totalmente absorvidos pelo homem, sendo necessário o auxílio da máquina, que desempenha um papel fundamental no gerenciamento da informação, fazendo o processamento prévio necessário do texto, com a finalidade de decodificá-lo e ajustá-lo às estruturas reconhecidas pelos computadores (SCHIESSL, 2007).

Assim, ao invés de encontrar textos que contenham informações e deixar que o próprio usuário procure o que lhe interessa, a DCT se preocupa em encontrar informações dentro dos textos e tratá-las de forma a apresentar ao usuário algum tipo de conhecimento útil e novo. Mesmo que tal conhecimento novo não seja a resposta direta às indagações do usuário, ele deve contribuir para satisfazer as necessidades de informação do mesmo (LOH, 2001).

Cabe salientar que é possível aplicar as técnicas tradicionais em informações textuais, porém as bases de texto apresentam-se desestruturadas, impossibilitando a aplicação somente das técnicas utilizadas em banco de dados estruturados. Sendo assim, técnicas específicas para tratamento de textos devem ser utilizadas a fim de se obter conhecimentos implícitos deste tipo de conteúdo (BARION e LAGO, 2008).

Silva Filho (2009) afirma que as fases mais importantes da aplicação em DCT são divididas da seguinte forma:

- a) **Coleta:** processo de busca e recuperação de textos que possui a finalidade de formar a base textual da qual se pretende extrair algum tipo de conhecimento;
- b) **Pré-processamento:** etapa posterior à coleta que visa à preparação do texto, eliminando termos irrelevantes, executando limpeza dos dados (retira dados que são irrelevantes) e correções ortográficas;

- c) **Indexação:** processo que visa organizar todos os termos que resultaram das fases anteriores a partir de fontes de dados que facilitam o acesso e recuperação;
- d) **Mineração:** aplicação de uma técnica ou método de mineração que identifique padrões e relacionamentos entre os dados;
- e) **Análise dos resultados:** fase que verifica a eficiência dos algoritmos da etapa de mineração e é responsável pela interpretação do conhecimento que foi identificado.

Sintetizando a ideia de Wives (2004), Hotho, Nurnenberger e Paab (2005) e Jesus (2010), dentre as técnicas conhecidas de DCT, as mais utilizadas são:

- a) **A Associação:** É uma técnica responsável por identificar padrões ou relações entre itens de um conjunto de dados.
- b) **A classificação:** consiste na alocação de elementos em classes que já estão pré-definidas, ou seja, já existem grupos em que os dados devem ser alocados;
- c) **Clusterização:** é a técnica utilizada para explorar semelhanças entre padrões e agrupar os elementos similares em grupos sem classe pré-definida;
- d) **A sumarização:** é a redução da dimensionalidade dos dados, eliminando dados que não possuem relevância para o documento;
- e) **A análise qualitativa e quantitativa:** consiste na análise dos resultados obtidos e na geração de relatórios pertinentes a tomada de decisão.

As aplicações em sistemas de DCT são inúmeras. Qualquer domínio que utilize intensivamente textos poderá beneficiar-se destes sistemas, tais como as áreas jurídicas e policiais, os cartórios e órgãos de registros, as empresas em geral, inclusive a área da saúde (LOH, 2001).

2.1.1 DESCOBERTA DE CONHECIMENTO NA ÁREA DA SAÚDE

A descoberta de conhecimento, por ser um processo em que gera conhecimentos que podem estar implícitos em dados semiestruturados ou não estruturados, é de extrema importância para a área da saúde, pois poderá auxiliar em resultados, em

diagnósticos e na descoberta de possíveis causas de doença que até podem levar ao falecimento, como por exemplo, os óbitos pediátricos, provocados por doenças predominantes em crianças.

Os óbitos pediátricos demonstram um sinal da situação da saúde e dos fatores sociais e econômicos de uma população. A partir de dados de óbitos de uma sociedade, é possível diagnosticar diversas tendências que estão ocorrendo em determinado local. Vianna *et al.* (2010), afirma que devido à vulnerabilidade da faixa etária de menores de um ano, a mortalidade infantil tem sido considerada um importante indicador para medir a situação da saúde de uma população.

Há várias causas de óbitos frequentes entre crianças e adolescentes que são consequências de determinadas características que predominam em uma comunidade ou região, como a violência, a falta de saneamento básico, podendo gerar doenças como epidemias, diarreias, entre outras enfermidades, por falta de cuidado com a saúde em locais mais pobres, como periferias e favelas (VICTORA, 2009), (CIAMPO *et al.* 1997) e (GUIMARÃES *et al.*, 2003).

Por exemplo, nas comunidades, os casos de falta de saneamento básico causam doenças como a diarreia. Segundo a UNICEF (2008), todos os anos, aproximadamente dois milhões de crianças ao redor do mundo morrem de diarreia. Nos países mais pobres, esta doença é a terceira causa mais comum em crianças menores de cinco anos, ficando logo atrás das causas neonatais e da pneumonia. Victora (2009) mostra um estudo baseado na diminuição dos casos de diarreia no Brasil, em que possui um aspecto muito interessante das análises realizadas sobre os óbitos pela doença que é a redução de disparidades geográficas, refletindo em uma queda nas desigualdades socioeconômicas.

Também existem casos de morte de crianças por motivos de acidentes, trauma ou lesão. Por exemplo, casos como os acidentes na infância são um destaque mundial em causas de morbimortalidade¹. Embora a evolução do conhecimento científico-tecnológico e os avanços da medicina ocorram ininterruptamente, na tentativa de

¹ Refere-se ao impacto das doenças e das mortes que incorrem em uma sociedade. Dicionário Online de Português. Disponível em: www.dicio.com.br/mortalidade. Acesso em: 28.07.14.

melhorar as condições de vida do ser humano, os acidentes representam cada vez mais uma importante causa de morbimortalidade em todo o mundo (CIAMPO *et al.* 1997).

Fatores sociais, como o desenvolvimento de grandes aglomerados humanos em precárias condições de subsistência na periferia das capitais e grandes cidades, o empobrecimento progressivo e a desinformação da grande maioria da população brasileira, têm contribuído de maneira significativa para o aumento do número de casos de trauma infantil (GUIMARÃES *et al.*, 2003).

Outros casos de morte comum em crianças é a morte prematura. McCallum *et al.* (2000) afirmam que apesar da existência da perícia médica moderna e tecnológica, existe um grupo de crianças para as quais a morte prematura é um resultado inevitável devido às suas doenças. Estes grupos incluem crianças com AIDS, distúrbios metabólicos, falência de órgãos em fase final, fibrose cística, anomalias congênitas graves, complicações de prematuridade extrema, e câncer.

Diante de tamanha importância, os dados de óbitos precisam ser tratados como ferramentas-chave para inúmeras soluções para os problemas tendenciosos que prejudicam a sociedade. Através das informações dos óbitos e das análises referente aos mesmos, é possível chegar a conclusões que irão auxiliar na tomada de decisões a respeito de fatores sociais, econômicos e políticos. Por isso a necessidade da aplicação dos processos de descoberta de conhecimento na área da saúde.

2.2 ETAPAS DA DESCOBERTA DE CONHECIMENTO EM TEXTO

A aplicação da DCT é dividida em fases de: coleta, limpeza e tratamento dos dados, indexação, mineração e interpretação. Estas fases formam as etapas de pré-processamento, processamento e pós-processamento que serão explicadas nas seções a seguir.

2.2.1 PRÉ-PROCESSAMENTO

A etapa de pré-processamento é a fase que consome maior parte do esforço de todo o processo da descoberta de conhecimento, pois envolve operações de entendimento, seleção, limpeza e transformação de dados. Muitas vezes, essas operações precisam ser repetidas de modo a aprimorar a qualidade dos dados e, conseqüentemente, melhorar também a acurácia e eficiência do processo de mineração (NEVES, 2003).

Após a organização, limpeza e preparação desses dados, eles estarão prontos para a etapa de processamento.

2.2.2 PROCESSAMENTO

A fase de processamento, como o próprio nome já diz, consiste em processar as informações preparadas na fase de pré-processamento, por meio de técnicas de mineração que são aplicadas através de métodos e algoritmos computacionais.

Na DCT, a fase de processamento ocorre na Mineração de Textos - MT, etapa em que padrões e conhecimento são extraídos a partir de documentos de texto, divulgando a informação oculta por meio de métodos que são capazes de lidar com um grande número de palavras e estruturas em linguagem natural e que permitem trabalhar com imprecisão e incerteza (MACHADO *et al.*, 2010).

Segundo Goebel e Gruenwald (1999 *apud* Wives, 2004, p. 23), existem vários métodos de mineração que são listados a seguir:

- a) **Métodos de processamento de dados:** são aplicados com o objetivo de selecionar, filtrar, agregar, exemplificar, limpar e transformar dados;
- b) **Método de predição:** buscam prever o valor de um atributo específico;
- c) **Método de Regressão:** analisam a dependência de valores de alguns atributos em relação a outros no mesmo item, gerando um modelo capaz de prever os valores de novos registros;
- d) **Métodos de classificação:** determinam a classe à qual determinado dado pertence;
- e) **Métodos de agrupamento textual (clusterização):** um tipo de análise exploratória de dados na qual um conjunto desconhecido de dados é particionado iterativamente em grupos de itens com características similares, a fim de se identificar padrões;
- f) **Métodos de análise de associações:** identificam relacionamentos entre atributos e itens (a fim de detectar se a presença de um padrão implica a presença de outro);

- g) **Métodos de visualização:** tornam o conhecimento entendível (gráficos e animações 2D e 3D).

Diversas ferramentas são criadas para auxiliar a fase de processamento dos textos, como por exemplo, o modelo Cassiopeia. Esta ferramenta é um modelo de Mineração de Textos desenvolvido por Guelpeli (2012), com o intuito de realizar agrupamentos textuais.

O modelo Cassiopeia realiza um novo método que é uma técnica de redução da alta dimensionalidade e dos dados esparsos baseado na curva da Lei de Zipf², em que se pode definir um limite inferior (corte que retira palavras muito específicas, encontradas uma única vez nos documentos) e superior (corte que retira as stopwords³) como limiares de corte, favorecendo encontrar termos relevantes, assumindo que as palavras mais significativas para discriminar o conteúdo estão em um pico imaginário, posicionado no meio dos dois pontos de corte (VAN RIJSBERGEN, 1979 *apud* GUELPELI, 2012).

Um novo corte médio na distribuição das frequências das palavras é proposto no modelo Cassiopeia, para que seja viabilizada por meio de centroides, como forma de representação do espaço amostral, e para organização dos textos nos agrupamentos (GUELPELI, 2012).

Os textos ao serem minerados passam pela técnica de clusterização através do uso de um algoritmo para a formação de agrupamento de textos hierárquicos, onde os mesmos são agrupados por grau de similaridade. Cada um dos agrupamentos criados possui um vetor de palavras, ordenado de forma decrescente denominado centroide dos agrupamentos, ao qual possui palavras pertinentes aos textos do agrupamento (GUELPELI, 2009). Caso haja entrada de novos textos, é possível surgir novos agrupamentos, sub-agrupamentos, ou a fusão dos dois, sendo, portanto um processo iterativo (LOH, 2001).

² Lei de Zipf é uma distribuição estatística específica, que se encontra em raros fenômenos estocásticos como a distribuição da frequência da ocorrência de palavras em um texto, em que nas ordenadas, se tem o valor dessa frequência, e nas abscissas, o valor da posição de ordenação relativa dessa palavra, em termos da sua frequência em relação ao das outras palavras do texto (QUONIAM, 2001, CUMMINS e O'RIORDAN 2005 e NOGUEIRA, 2009 *apud* GUELPELI, 2012).

³ Stopwords são classes de palavras que não carregam significados, tais como artigos, pronomes, interjeições e preposições (PARDO, 2002 *apud* GUELPELI, 2012).

O modelo Cassiopeia identifica características das palavras no texto, utilizando a frequência relativa, que define a importância de um termo, de acordo com a quantidade de vezes com que a palavra é encontrada no documento. Quanto mais um termo aparece em um texto, mas ele é importante para aquele documento. A frequência relativa é calculada para que o resultado da frequência absoluta das palavras seja normalizado, e evite que documentos sejam representados por vetores de tamanhos diferentes e sejam proporcionais ao tamanho do documento (GUELPELI, 2012).

Segundo Guelpeli (2012), o processo de seleção de atributos ocorre da seguinte forma: primeiramente a frequência relativa é calculada conforme é demonstrado na equação 1, ou seja, calcula-se quantas vezes cada palavra aparece no documento, dividido pelo número total de palavras do mesmo.

$$F_r X = \frac{F_{abs} X}{N} \quad (1)$$

Onde $F_r X$ é igual à frequência relativa de X , $F_{abs} X$ é igual à frequência absoluta de X , ou seja, a quantidade de vezes que a palavra aparece no documento e N é igual ao número total de palavras no documento (GUELPELI, 2012).

Em seguida, as palavras são ordenadas em ordem decrescente de frequência, e após a ordenação, a frequência média das palavras é calculada para definir o novo corte na distribuição dos termos que mais aparecem, conforme é exibida na equação 2. A soma das frequências relativas de cada documento é dividida pelo número total das palavras no texto.

$$f(P, N) = \frac{\sum_{n=1}^N F_r P_n}{N} \quad (2)$$

Onde N é igual ao número total de palavras no documento; $F_r P_n$ é igual frequência relativa de P_n ; onde P é o conjunto de palavras no documento e P_n refere-se à quantidade de vezes que uma palavra aparece no documento, e $f(P, N)$ é a frequência média das palavras na distribuição (GUELPELI, 2012).

O próximo passo é o modelo encontrar a primeira palavra no documento cuja frequência está mais próxima da média. Ao encontrar esta palavra, ela é marcada e

incluída com as 24 palavras anteriores (à esquerda do corte do vetor), e também são escolhidas as outras 25 palavras posteriores (à direita do corte do vetor), conforme a Figura 1.

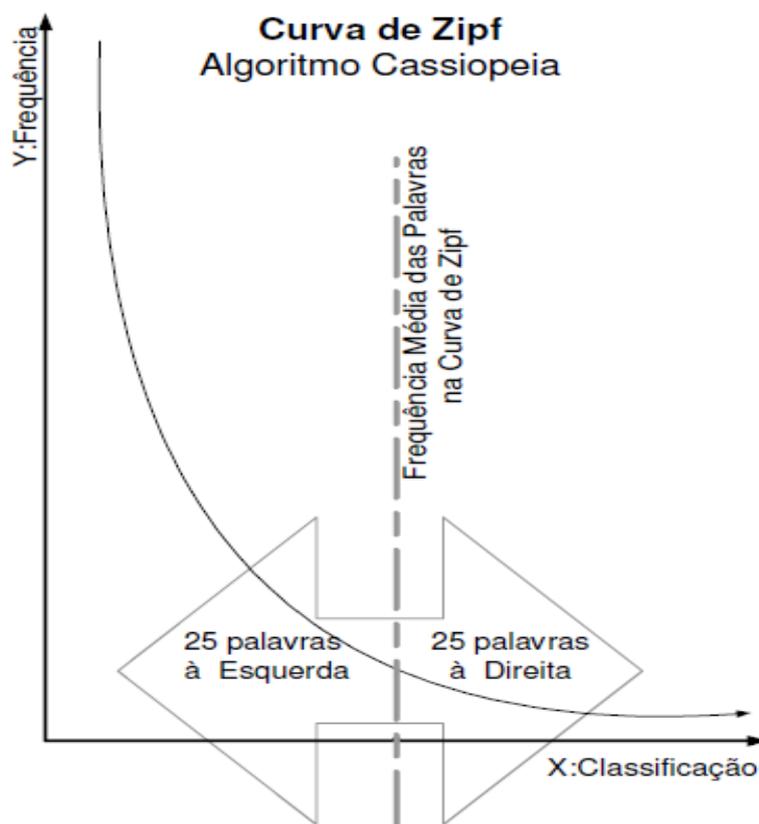


Figura 1. Seleção dos atributos no modelo Cassiopeia (GUELPELI, 2012).

E finalmente, após a escolha de todas as palavras, o vetor é montado com as 50 palavras escolhidas em ordem decrescente.

A clusterização, utilizada na fase de processamento do modelo Cassiopeia, é um método atribuído de métricas que auxiliam na análise do conglomerado dos textos, ao serem processados. Essas métricas são abordadas nas seções seguintes.

2.2.2.1 MÉTRICAS EXTERNAS

Métricas externas são métricas em que os resultados dos agrupamentos são avaliados por uma estrutura de classes pré-definidas que refletem a opinião de um especialista humano. As medidas usadas neste tipo de métricas são: Recall, Precision e F-Measure, que é a média harmônica entre as duas primeiras medidas (GUELPELI, 2012).

O Recall, representado pela equação 3, mede a proporção de elementos que foram alocados de forma correta no agrupamento em relação ao total dos elementos da classe associada a este agrupamento (RIJSBERGEN, 1979) e (MANNING *et al.*, 2008).

$$R = \frac{n(A)}{n(A \cup D)} \quad (3)$$

Onde $n(A)$ é o número de elementos do subconjunto A de acertos e $n(D)$ é o número de elementos do subconjunto D de falsos negativos⁴ e $n(A \cup D)$ é o número total de elementos da classe correspondente (GUELPELI, 2012).

A Precision, representada pela equação 4, mede a proporção de elementos que foram alocados de forma correta no agrupamento em relação ao total de elementos deste agrupamento (RIJSBERGEN, 1979) e (MANNING *et al.*, 2008).

$$P = \frac{n(A)}{n(A \cup B)} \quad (4)$$

Onde $n(A)$ é o número de elementos do subconjunto de A de acertos e $n(B)$ é o número de elementos do subconjunto B , e $n(A \cup B)$ é o número total de elementos do grupo (GUELPELI, 2012).

O F-Measure, representado na equação 5, é a média harmônica entre Recall e Precision, assumindo valores que estão no intervalo $[0,1]$. O valor zero indica que nenhum elemento foi agrupado corretamente, e o valor um, indica que todos os elementos foram corretamente agrupados. Portanto, o agrupamento ideal deve retornar valor igual a um (RIJSBERGEN, 1979) e (MANNING *et al.*, 2008).

$$2 * \frac{Precision(P) * Recall(R)}{Precision(P) + Recall(R)} \quad (5)$$

Cada uma das métricas descritas acima é calculada em cada um dos grupos obtidos, fornecendo assim a qualidade de cada grupo. A medida de avaliação, para todo o agrupamento, é obtida através do cálculo da média entre cada uma das medidas de todos os grupos (GUELPELI, 2012).

⁴ Falsos negativos são elementos que deveriam ter sido alocados a um grupo, mas que foram alocados a outros (GUELPELI, 2012).

2.2.2.2 MÉTRICAS INTERNAS

As métricas internas utilizam informações contidas nos grupos que foram formados para realizar a análise dos resultados, portanto não utilizam informações externas. As medidas usadas nesse tipo de métricas são: Coesão, Acoplamento e Coeficiente Silhouette que é uma espécie de média harmônica entre as duas primeiras medidas (GUELPELI, 2012).

A Coesão representada pela equação 6, mede a similaridade entre os elementos do mesmo grupo. Quanto maior a Coesão, mais os elementos do mesmo grupo possuem características semelhantes:

$$\frac{\sum_{i>j} Sim(P_i, P_j)}{\frac{n(n-1)}{2}} \quad (6)$$

$Sim(P_i, P_j)$ é o cálculo da similaridade entre os textos i e j que pertencem ao agrupamento P , n é o número de textos contidos no agrupamento P , e P_i e P_j são membros do agrupamento P (GUELPELI, 2012).

Já o Acoplamento representado pela equação 7, mede a similaridade entre elementos de grupos diferentes. Dois elementos que possuem grupos diferentes um do outro são comparados, para verificar se possuem grau de similaridade.

$$\frac{\sum_{i>j} Sim(C_i, C_j)}{\frac{n_a(n_a-1)}{2}} \quad (7)$$

C é o centroide de um agrupamento, ou seja, organiza os elementos deste agrupamento que está presente em P , $Sim(C_i, C_j)$, é o cálculo da similaridade do texto i pertencente ao agrupamento P e o texto j que não pertence a P , C_i é centroide do agrupamento P e C_j é centroide do agrupamento P_i e n_a é o número de agrupamentos presentes em P (GUELPELI, 2012).

E o Coeficiente Silhouette, representado na equação 8, é a média harmônica entre as duas medidas de métrica interna, baseando-se em quanto um elemento é similar em relação ao outro elemento do mesmo grupo, e o quanto ele é similar em relação ao elemento de um grupo diferente.

$$S = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

Onde $a(i)$ é a distância média entre o i -ésimo elemento do grupo e os outros do mesmo grupo. O $b(i)$ é o valor mínimo de distância entre o i -ésimo elemento do grupo e qualquer outro grupo, que não contenha o elemento referenciado, e max é a distância máxima entre $a(i)$ e $b(i)$ (GUELPELI, 2012).

2.2.3 PÓS-PROCESSAMENTO

A DCT, por possuir uma grande quantidade de informações, necessita utilizar de formas de visualização que sejam possíveis observar os dados que foram trabalhados e processados. Essas formas de visualização compõem a fase de pós-processamento.

O pós-processamento consiste na avaliação das descobertas, analisando os resultados obtidos e sua relevância. Nesta etapa ocorrem a interpretação, a visualização e a validade das tendências descobertas, dos padrões extraídos e possivelmente retorna aos passos anteriores, caso ocorra a necessidade de melhorias do resultado (SILVA e ROVER, 2011).

Os dados podem ser visualizados por gráficos, tabelas, relatórios, histogramas, entre outras formas, para facilitar a interpretação dos resultados encontrados através das etapas anteriores.

Em DCT, métodos de visualização podem melhorar e simplificar a descoberta ou a extração de padrões, ou então, as informações relevantes. Informações que permitem uma representação visual compreendem aspectos da coleção de documentos ou conjunto de resultados (HOTH0, NURNBERGER e PAAB, 2005).

Visualizações gráficas das informações, frequentemente fornecem informações mais abrangentes e melhores, sendo mais rápido do que é possível compreender por meio do texto puro. Muitas das abordagens desenvolvidas para fins de mineração de texto são motivadas por métodos que haviam sido propostos na exploração das áreas de análise de dados, visualização de informação e mineração de dados visuais (HOTH0, NURNBERGER e PAAB, 2005).

Como exemplo de métodos de visualização em pós-processamento, pode ser citado o *Self - Organization Map – SOM*, uma auto-organização de mapas⁵ que

⁵ Mapas são formas esquemáticas utilizadas na análise das estatísticas dos contextos de palavras para fornecer informações sobre as suas relações (VISA et al., 2000).

consistem em organizar documentos com base em sua similaridade (VISA *et al.*, 2000) e (MEDEIROS e COSTA, 2008).

O SOM define um mapeamento de um espaço multidimensional para um conjunto finito de vetores-referência ou neurônios, dispostos na forma de um arranjo espacial regular, normalmente bidimensional. O objetivo principal é reduzir a dimensionalidade do espaço de entrada ao mesmo tempo em que se busca preservar ao máximo a topologia (VISA *et al.*, 2000) e (MEDEIROS e COSTA, 2008).

Visa *et al.* (2000) utilizam o SOM para descobrir conhecimento em documentos de texto baseado em mapas de parágrafos, sendo estes mapas comparados e interpretados por meio de histogramas utilizados para esta análise. Ele utiliza o SOM e codificação inteligente de palavras para hierarquizar os dados de um texto em vários níveis como palavra, frase e parágrafo. O trabalho é focado no nível mais alto, que é o nível de parágrafo.

Rajaraman e Tan (2002) desenvolveram uma técnica interativa de visualização de mapas de conceito⁶ que é apresentada como um guia do usuário para a descoberta de conhecimento, tendo como apoio a construção da base de conhecimentos de conceitos chamada de *Conceptual Knowledge Base – CKB*. Na *CKB*, um banco de dados estruturado de importantes conceitos e relações é extraído a partir dos textos.

O mapa de conceito fornece uma interface acessível para visualizar o conteúdo do documento. O usuário pode escolher um conceito e ver sua relação com outros conceitos, tornando possível uma visão baseada pelo tema dos documentos. Isto implica que a interface pode permitir diferentes perspectivas do conteúdo (RAJARAMAN e TAN, 2002).

Existem outros métodos de visualização como a *Análise de Componentes Principais – PCA*, um dos métodos de visualização mais antigos que consiste na extração de atributos, baseada em uma transformação linear sobre os dados, obtendo projeções sobre um novo sistema de coordenadas ortogonais que permitem representar a

⁶ Conceito é uma forma utilizada para descoberta de conhecimento em textos por meio do significado em relação aos fenômenos da realidade (pessoas, objetos, entidades, eventos e situações do mundo real), e não analisando somente palavras ou sentenças (LOH, 2001).

maior variância possível dos dados, numa sequência decrescente do percentual dessa variância, correspondendo à ordem desses componentes (MEDEIROS e COSTA, 2008).

O método aplica uma transformação linear sobre um conjunto de dados de entrada e encontra um novo sistema de coordenadas para representar esse conjunto numa forma mais adequada à análise dos dados (MEDEIROS e COSTA, 2008).

Um modelo de rede neural chamado *Perceptron de Múltiplas Camadas – MLP* também tem sido muito utilizado em mapeamentos e reconhecimento de padrões. Redes MLP são compostas de múltiplas camadas de neurônios, completamente conectadas (MEDEIROS e COSTA, 2008)..

Tipicamente apresentam fluxo de dados para frente (*feed-forward*) e utilizam algoritmo de treinamento de retro propagação (*back-propagation*) da diferença entre a resposta da rede e a resposta desejada para uma determinada entrada. O treinamento da rede é baseado na apresentação sucessiva de um conjunto de dados de treinamento, num processo iterativo (MEDEIROS e COSTA, 2008).

No modelo Cassiopeia, a etapa de pós-processamento gera formas de visualização através da criação de um arquivo em HTML e uma planilha, exibindo informações dos resultados dos agrupamentos que foram formados. Porém, essas formas de visualização não são ilustrativas o suficiente para analisar os resultados de um modo geral.

Sendo assim, outros métodos de visualização foram utilizados neste trabalho para exibir os resultados dos agrupamentos e de suas métricas textuais de uma forma mais ilustrativa. Esses métodos serão abordados na metodologia deste trabalho.

Com as técnicas de visualização de textos na etapa de pós-processamento descritas, dentre outras, o conhecimento pode ser descoberto e adquirido para ser aplicado ao mundo real, tornando-o um complemento do conhecimento já existente.

CAPÍTULO 3 – METODOLOGIA

Neste capítulo, será apresentada a metodologia utilizada para a descoberta de conhecimento em óbitos pediátricos. As ferramentas utilizadas foram: o modelo Cassiopeia, desenvolvido por Guelpeli (2012) que tem o intuito de realizar o agrupamento dos arquivos textuais e avaliar os mesmos por métricas internas; o Navicat Lite, versão 10.0.5 (não comercial), responsável por gerenciar os bancos de dados que foram formados, fazendo com que apresentem uma melhor visualização das informações; o FineCount, versão 2.6, usado para gerar informações estatísticas sobre os arquivos de texto; e o Microsoft Office Excel 2010 que é usado para a criação dos gráficos que auxiliam na visualização das informações geradas após a clusterização dos documentos.

3.1 REGISTRO DE ÓBITOS: DESCRIÇÃO DA BASE DE DADOS (PRÉ-PROCESSAMENTO)

Primeiramente foi realizado um levantamento dos dados de óbito de crianças de zero a 16 anos, no período de 2002 a 2012, utilizando como fonte de base de dados o Cartório de Registro Civil em Diamantina, no estado de Minas Gerais. A outra fonte de obtenção dos dados era no Hospital Nossa Senhora da Saúde em Diamantina, mas devido a trâmites burocráticos, não foi possível ter acesso a essas informações. No entanto, os dados de óbitos contidos no cartório são enviados pelo próprio hospital, ou seja, ambas as fontes possuem as mesmas informações.

Neste Cartório foi colhido um total de 147 registros de óbitos. Aparentemente parecem poucas as informações para a pesquisa em questão, mas é importante considerar que a cidade de Diamantina é uma cidade do interior de Minas Gerais, e possui uma população equivalente a 45.880 habitantes, sendo um número consideravelmente reduzido ao ser comparado com as grandes cidades do país como Belo Horizonte - MG com 2.375.151 habitantes e São Paulo - SP com 11.253.503 habitantes (IBGE, 2010).

E além do número reduzido da população, os dados de óbitos colhidos foram apenas óbitos pediátricos, filtrando ainda mais as informações para esta pesquisa. Logo, 147 registros de óbitos pediátricos é um número significativo, baseado nas estimativas de morte pediátrica por crianças nascidas vivas.

Os dados de óbitos, ao serem colhidos, foram inseridos em uma planilha separando as informações em campos como: Nome, Sexo, Idade, Cor, Naturalidade, Residente, Profissão, Óbito, Data do Óbito e Causa do Óbito, conforme a Figura 2.

| Sexo: | Idade: | Cor: | Naturalidade: | Residente: | Profissão: | Óbito | Data de Óbito | Causa do Óbito: |
|----------|---------|----------|---------------------|------------------|------------|--|---------------|--|
| Feminino | 16 anos | ignorada | Corinto - MG | Diamantina - MG | nda consta | choque misto, hemorragia digestiva alta, diabetes mellitus descompensado | 04/12/2003 | Dando como causa da morte choque misto, hemorragia digestiva alta, diabetes mellitus descompensado declarou que no dia 4 de dezembro de 2003, às 20 hs, em Hospital de Nossa Senhora da Saúde, nesta cidade Diamantina |
| Feminino | 6 anos | ignorada | Rio de Janeiro - RJ | Mauá - SP | nda consta | traumatismo crânio-encefálico | 16/12/2003 | Dando como causa da morte traumatismo crânio-encefálico, declarou que no dia 16 de dezembro de 2003, às ignorada, em Hospital de Nossa Senhora da Saúde, nesta cidade Diamantina |
| Feminino | 8 anos | ignorada | Santo Amaro - SP | Santo Amaro - SP | nda consta | traumatismo crânio-encefálico contuso | 16/12/2003 | Dando como causa da morte traumatismo crânio-encefálico contuso, declarou que no dia 16 de dezembro de 2003, às "ignorada", em Hospital de Nossa Senhora da Saúde, nesta cidade Diamantina |
| Feminino | 10 anos | ignorada | Independência - CE | São Paulo - SP | nda consta | insuficiência respiratória, trauma cervical contuso | 16/12/2003 | Dando como causa da morte insuficiência respiratória, trauma cervical contuso, declarou que no dia 16 de dezembro de 2003, às "ignorada", em Hospital de Nossa Senhora da Saúde, nesta cidade Diamantina |
| Feminino | 8 anos | ignorada | Cratêus - CE | São Paulo - SP | nda consta | politrauma contuso | 16/12/2003 | Dando como causa da morte politrauma contuso, declarou que no dia 16 de dezembro de 2003, às "ignorada", em Hospital de Nossa Senhora da Saúde, nesta cidade Diamantina |

Figura 2. Planilhas com as informações de óbito.

Após as informações terem sido coletadas, os campos Nome, Cor e Profissão foram descartados na preparação das informações. O campo Nome foi retirado das informações para preservar a identidade da criança que veio ao óbito e de seus familiares.

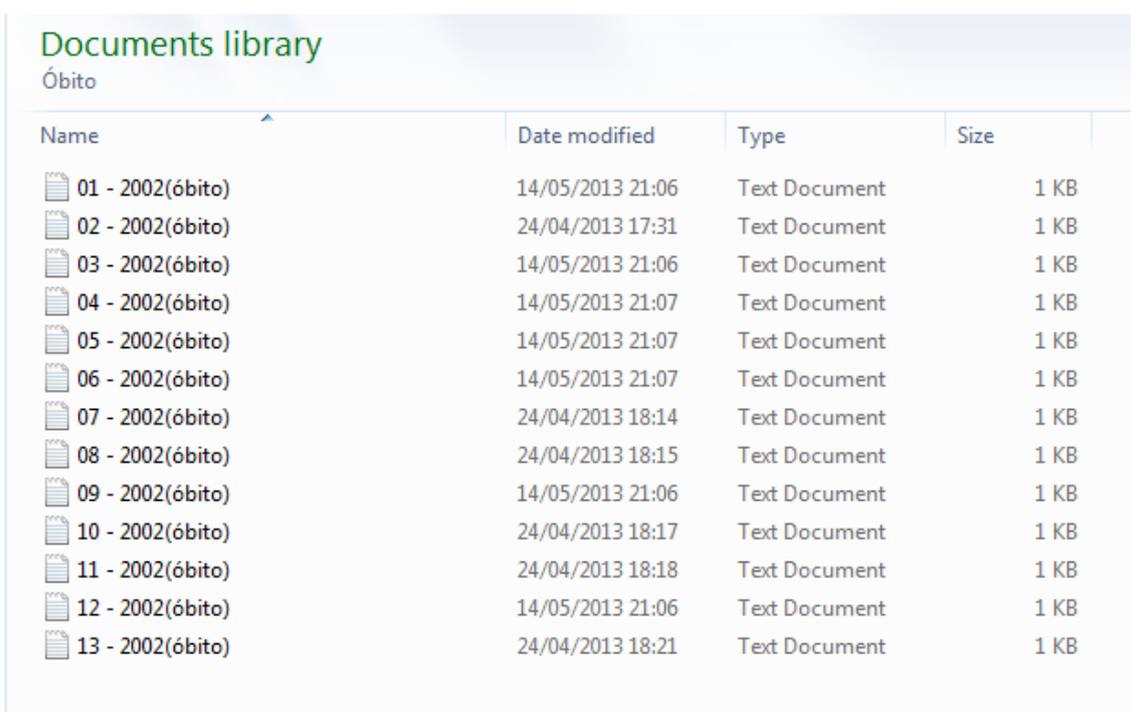
O campo Cor também foi retirado, pois grande parte dos registros não possui este tipo de dado, estando apenas declarada a expressão “ignorada”, conforme mostra a Figura 2. Logo, tal campo se tornou inviável para a pesquisa devido à ausência de dados.

A profissão foi descartada pelo fato de que, como os dados dos óbitos coletados são de menores de idade, não haveria informações consistentes que pudessem auxiliar nessa pesquisa, pois crianças não possuem uma profissão definida.

O campo Data do Óbito foi transformado em dois novos campos: Mês do Óbito e Ano do Óbito. Esta mudança auxilia na análise dos meses e dos anos em que ocorreram os óbitos. Os dados da planilha foram conferidos pela especialista em pediatria, a Dra. Alison Guelpeli (CRM: 58624-MG), para evitar possíveis erros de escrita das informações de causas de óbito.

O fato de inserir os dados de óbitos em uma planilha não traz uma descoberta de conhecimento dos registros, pois nesse estado, eles ainda não estão organizados por um padrão que proporcione tal conhecimento. É necessário que haja o agrupamento desses registros para que ocorra similaridade entre os dados que foram coletados.

Sendo assim, após a inserção dos dados nas planilhas e a verificação de possíveis erros ortográficos, as informações dos campos óbito e causa de óbito foram retiradas das outras informações e divididas em arquivos, em formato de texto, para a fase de processamento. A Figura 3 ilustra um exemplo de como os arquivos de texto foram formados, baseados nas informações de óbito ocorridas no ano de 2002.



| Name | Date modified | Type | Size |
|------------------|------------------|---------------|------|
| 01 - 2002(óbito) | 14/05/2013 21:06 | Text Document | 1 KB |
| 02 - 2002(óbito) | 24/04/2013 17:31 | Text Document | 1 KB |
| 03 - 2002(óbito) | 14/05/2013 21:06 | Text Document | 1 KB |
| 04 - 2002(óbito) | 14/05/2013 21:07 | Text Document | 1 KB |
| 05 - 2002(óbito) | 14/05/2013 21:07 | Text Document | 1 KB |
| 06 - 2002(óbito) | 14/05/2013 21:07 | Text Document | 1 KB |
| 07 - 2002(óbito) | 24/04/2013 18:14 | Text Document | 1 KB |
| 08 - 2002(óbito) | 24/04/2013 18:15 | Text Document | 1 KB |
| 09 - 2002(óbito) | 14/05/2013 21:06 | Text Document | 1 KB |
| 10 - 2002(óbito) | 24/04/2013 18:17 | Text Document | 1 KB |
| 11 - 2002(óbito) | 24/04/2013 18:18 | Text Document | 1 KB |
| 12 - 2002(óbito) | 14/05/2013 21:06 | Text Document | 1 KB |
| 13 - 2002(óbito) | 24/04/2013 18:21 | Text Document | 1 KB |

Figura 3. Arquivos de texto com as informações de óbito que ocorreram no ano de 2002.

Esses arquivos de texto foram organizados em pastas de acordo com o ano em que ocorreu cada órbita, conforme é ilustrado na Figura 4.

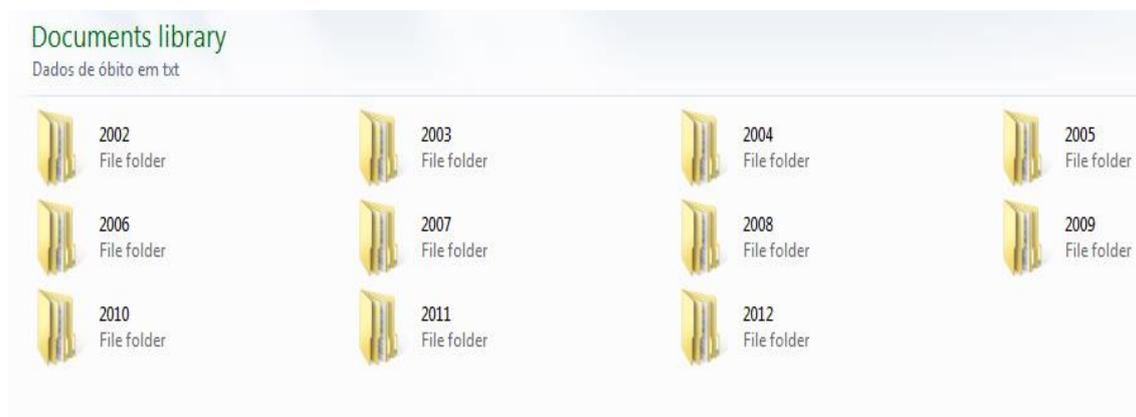


Figura 4. Arquivos de texto dos órbitas organizados por ano.

Como os arquivos possuem pouco conteúdo textual, não chegando a completar um parágrafo em cada arquivo, não houve a necessidade de aplicar a técnica de sumarização para a diminuição da dimensionalidade dos dados. A interação humana nessa fase foi o suficiente para a organização dos documentos, fazendo com que eles possuíssem um grau de informação considerável para a fase de processamento. As demais informações que estavam na planilha foram inseridas em uma base de dados para o uso futuro, após a fase de agrupamento dos textos.

Após os dados serem limpos, organizados e preparados, eles ficaram prontos para a etapa de processamento com o uso do modelo Cassiopeia.

3.2 INSERÇÃO DOS ARQUIVOS DE TEXTO NO MODELO CASSIOPEIA (PROCESSAMENTO)

Os dados foram processados pelo modelo Cassiopeia através da técnica de agrupamento de textos (clusterização). Todos os arquivos de texto formados na etapa de pré-processamento foram inseridos no modelo para serem agrupados de acordo com similaridade entre esses dados.

Ao processar os arquivos, métricas internas de análise textual foram geradas. Essas métricas auxiliam na análise do processamento dos textos, pois utilizam informações contidas nos grupos que foram formados para realizar a análise dos resultados.

Ao terminar o processamento foi possível perceber que o agrupamento dos textos dos campos Óbito e Causa de Óbito era ineficaz quando inseridos juntos para a clusterização dos textos, pois ao processar os dois campos, os clusters gerados acabavam não sendo agrupados pela causa de óbito, mas por outras palavras contidas nos textos do campo Causa de Óbito.

Este campo, além de conter as mesmas informações do campo Óbito, também possui algumas informações irrelevantes que prejudicam a eficiência do agrupamento dos documentos. Já o campo óbito possui as mesmas informações, porém mais relevantes e objetivas.

Por exemplo, em um arquivo de texto, enquanto o campo Óbito é descrito apenas a informação “prematuridade extrema”, no campo Causa de Óbito vem descrito “Dando como causa da morte prematuridade extrema, declarou que no dia 28 de abril de 2002, às 12,09 hrs, no Hospital Nossa Senhora da Saúde, nesta cidade Diamantina”.

Como ambos os campos possuem a mesma informação sobre o motivo do óbito, porém uma de forma mais objetiva e a outra de forma menos relevante e mais prolixa para o estudo, seria redundante inserir os dois campos com o mesmo tipo de informação. Baseado nesses resultados iniciais, o campo Causa do Óbito foi retirado para a realização de um segundo processamento.

Na Tabela 1, são exibidas as estatísticas de texto do campo Óbito e Causa de Óbito, comparando suas quantidades de palavras, de números, de caracteres, além da quantidade de sentenças e a quantidade de linhas de cada um.

Tabela 1. Estatísticas dos arquivos de texto do campo Óbito e Causa de Óbito.

| Arquivo | Palavras | Palavras e Números | Números (%) | Caracteres | Caracteres e espaços | Sentenças | Linhas |
|----------------|----------|--------------------|-------------|------------|----------------------|-----------|---------|
| Óbito | 885 | 885 | 0,00% | 5675 | 6184 | 150 | 112,44 |
| Causa de Óbito | 4611 | 5154 | 10,54% | 23300 | 27704 | 150 | 503, 71 |

3.3 VISUALIZAÇÃO DOS RESULTADOS (PÓS-PROCESSAMENTO)

Com o processamento dos arquivos de texto, foram gerados pelo modelo Cassiopeia um total de 30 *clusters*, separados por grau de similaridade entre as suas informações. Esses *clusters* foram inseridos em uma nova base de dados, e as

informações adicionais que estavam armazenadas na base de dados, formada anteriormente na fase de pré-processamento, foram associadas novamente com o campo Óbito na nova base de dados.

Porém, essas informações não estão mais organizadas por ano, mas sim por *clusters*, ou seja, as informações de óbitos estão organizadas em grupos gerados pelo modelo Cassiopeia, sendo formados por causas de óbito em comum, conforme mostra a Figura 5.

The image shows a screenshot of a database table with the following columns: txt, obito, sexo, idade, naturalidade, residencia, mes_obito, and ano_obito. The table contains 20 rows of data, all representing deaths in Diamantina - MG. The causes of death are grouped into a cluster, with 'prematividade extrema' being the most frequent cause.

| txt | obito | sexo | idade | naturalidade | residencia | mes_obito | ano_obito |
|---------|---|-----------|---------------------|-----------------|-----------------|-----------|-----------|
| 03-2002 | prematividade extrema | Feminino | 1 dia | Diamantina - MG | Nada Consta | Abril | 2002 |
| 02-2005 | prematividade extrema | Feminino | 2 dias | Diamantina - MG | Nada Consta | Maio | 2005 |
| 06-2004 | prematividade extrema | Feminino | 5 dias | Diamantina - MG | Nada Consta | Novembro | 2004 |
| 01-2002 | prematividade | Masculino | 4 dias | Diamantina - MG | Nada Consta | Março | 2002 |
| 04-2002 | prematividade | Masculino | 23 dias | Diamantina - MG | Nada Consta | Maio | 2002 |
| 05-2002 | prematividade | Masculino | 2 dias | Diamantina - MG | Nada Consta | Junho | 2002 |
| 06-2011 | ilegível, prematividade extrema | Feminino | 6 dias (0 anos) | Diamantina - MG | Diamantina - MG | Setembro | 2011 |
| 05-2008 | distúrbio hidroeletrólítico, prematividade | Feminino | 1 mês | Diamantina - MG | Nada Consta | Maio | 2008 |
| 08-2012 | prematividade extrema | Masculino | 15 dias (0 anos) | Diamantina - MG | Diamantina - MG | Maio | 2012 |
| 02-2002 | ilegível | Masculino | 6 dias | Diamantina - MG | Nada Consta | Março | 2002 |
| 08-2002 | ilegível | Feminino | 9 meses | Diamantina - MG | Diamantina - MG | Agosto | 2002 |
| 10-2002 | ilegível | Masculino | 3 dias | Diamantina - MG | Nada Consta | Novembro | 2002 |
| 02-2012 | parada cardio respiratória, prematividade extrema | Feminino | 3 horas (0 anos) | Diamantina - MG | Diamantina - MG | Dezembro | 2012 |
| 01-2012 | parada cardio respiratória, sepse | Masculino | 1 ano | Curvelo - MG | Corinto - MG | Dezembro | 2012 |
| 09-2012 | choque cardiogênico, pneumotorax hipertensivo, p | Feminino | 0 anos | Diamantina - MG | Diamantina - MG | Maio | 2012 |
| 14-2012 | prematividade extrema | Feminino | 1,30 horas(0 anos) | Diamantina - MG | Diamantina - MG | Julho | 2012 |
| 33-2012 | prematividade extrema | Feminino | 30 minutos (0 anos) | Diamantina - MG | Diamantina - MG | Novembro | 2012 |
| 27-2012 | prematividade extrema | Feminino | 30 minutos (0 anos) | Diamantina - MG | Diamantina - MG | Outubro | 2012 |

Figura 5. Tabela de um *cluster* com informações adicionais das crianças associadas ao óbito.

Com o resultado das métricas internas de análise textual, foi possível a criação de gráficos que demonstram de forma ilustrativa e mais clara os resultados obtidos com a clusterização dos arquivos de texto. Esses gráficos exibem a quantidade de elementos por *cluster*, o índice de coesão, o acoplamento e o índice de Coeficiente Silhouette de cada *cluster* formado.

Também foi feita a criação de um relatório com informações importantes, como, o número total de elementos colhidos para a pesquisa, o número de *clusters* formados na fase de processamento, tabelas com as informações de óbito de cada *cluster* e gráficos de comparação das informações da base de dados. Tal documento pode ser visualizado no Apêndice A.

Este relatório foi criado para uma análise humana que implica em possíveis descobertas de conhecimento através dos agrupamentos e das ilustrações criadas, baseadas nas informações dos *clusters*.

Assim como a correção das informações de óbito na fase de pré-processamento, a análise do relatório com as informações de cada *cluster* formado, também foi feita pela especialista em pediatria, a Dra. Alison (CRM: 58624-MG). Cada *cluster* foi analisado, e baseando nos resultados adquiridos, foi possível a descoberta de conhecimento fundamentada nessas informações, cujas mesmas, foram separadas em grupos, podendo ser visualizadas no Apêndice B.

CAPÍTULO 4 – RESULTADOS

Este capítulo abordará os resultados que foram obtidos por meio das métricas de avaliação dos textos processados pelo modelo Cassiopeia. Também serão apresentados os resultados da análise humana, em que foi possível a descoberta de conhecimento sobre os *clusters* que foram gerados. Alguns desses *clusters* são mencionados neste capítulo para ilustrar de forma objetiva a coerência dos resultados.

4.1 ANÁLISE DAS MÉTRICAS TEXTUAIS

Após a clusterização dos textos que continham os dados de óbitos pediátricos, foram obtidos os seguintes resultados:

Como o número de palavras dos registros é pequeno, ou seja, um total de 885, de acordo com a Tabela 1, o modelo Cassiopeia mostrou-se sensível a este número, apresentando um índice médio do coeficiente Silhouette de -0,016%.

Este resultado leva a entender que as clusterizações realizadas pelo modelo Cassiopeia não foram válidas, pois valores do coeficiente Silhouette que estiverem próximos de -1% representam *clusters* que não foram bem agrupados. Valores do Coeficiente Silhouette que forem próximos de 0% representam agrupamentos neutros, já os coeficientes próximos a 1%, representam elementos que estão devidamente agrupados em seus *clusters*.

Apesar do baixo índice dos coeficientes Silhouette, a análise humana, apresentada na segunda fase dos resultados, comprova que os *clusters* formados foram agrupados corretamente.

A métrica de acoplamento dos agrupamentos obteve um índice médio de 0,12%, ou seja, o acoplamento possui um índice baixo. Já a métrica de coesão dos agrupamentos obteve um índice médio de 1,9%. Logo, a similaridade entre elementos de um grupo e elementos de grupos diferentes é baixa, o que conseqüentemente leva a entender que a coesão, ou seja, a similaridade entre elementos do mesmo grupo é alta, sendo, portanto, um fator positivo para a pesquisa.

Os gráficos abaixo demonstram um resumo dos resultados do processamento feito através do modelo Cassiopeia. No primeiro gráfico, Figura 6, é exibida a quantidade de elementos que foram agrupados em cada *cluster*.

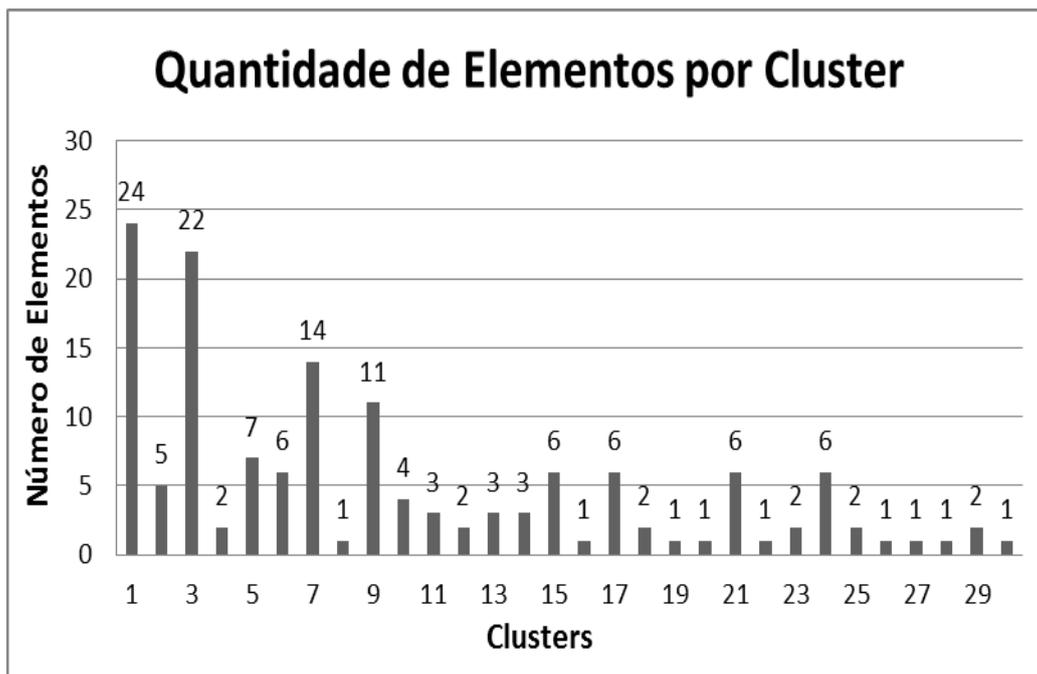


Figura 6. Quantidade de elementos agrupados em cada *cluster*.

Na Figura 7, o gráfico exibe os índices de coesão de cada *cluster* formado pelo modelo Cassiopeia.

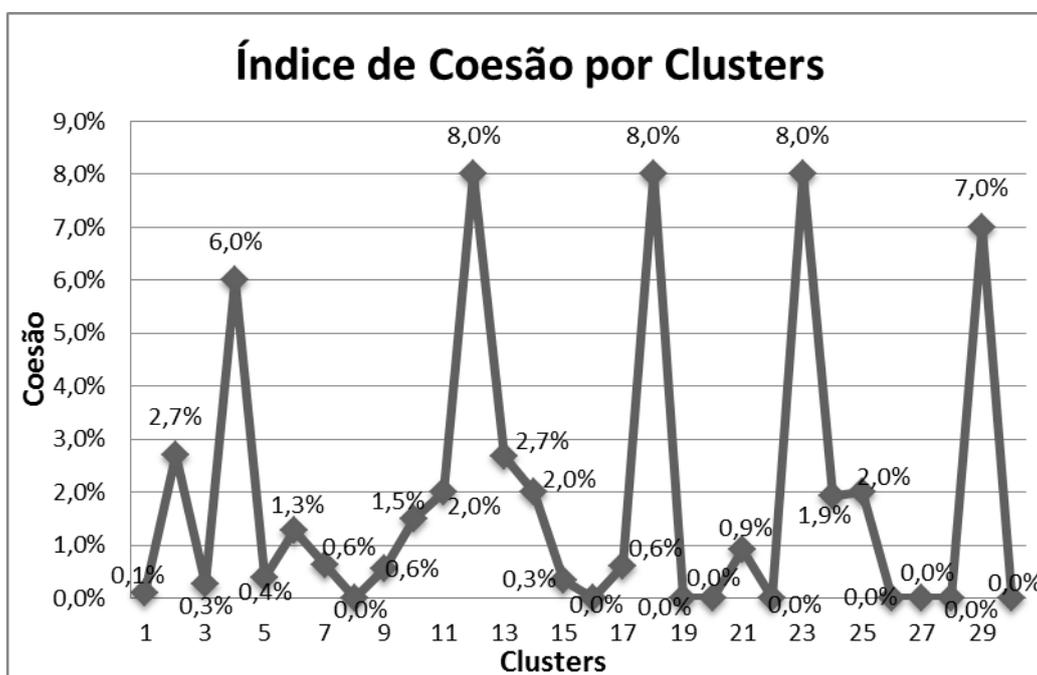


Figura 7. Índice de coesão em cada *cluster*.

Através dos gráficos da Figura 6 e da Figura 7, foi possível observar os seguintes resultados: *Clusters* com maiores índices de coesão são *clusters* que possuem menor quantidade de registros. Ao analisar *Clusters* com índices de coesão de 6% a 8%,

é possível perceber que esses grupos possuem apenas dois registros. Como exemplo, pode-se citar o *cluster* 12, na Tabela 2, que possui um índice de coesão de 8% e contém apenas dois registros de óbito.

Tabela 2. Cluster 12 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|---------|-----------------|----------------|-----------|-----------|
| 1 | 06 2012 | Politraumatismo, vítima de acidente de trânsito | Masculino | 14 anos | Minas Novas MG | Minas Novas MG | Abril | 2012 |
| 2 | 05 2012 | Traumatismo crânio encefálico, vítima de acidente de trânsito | Masculino | 16 anos | Itamarandiba MG | Carbonita MG | Março | 2012 |

Os *clusters* com menores índices de coesão possuem uma maior quantidade de registros em seus grupos. Por exemplo, o *cluster* 01, Figura 6, possui um índice de coesão de 0,1% e contém 24 elementos em seu grupo.

Clusters que possuem um índice de 0% de coesão são grupos que contém apenas um registro de óbito, ou seja, neste grupo não há como medir o nível de similaridade entre os elementos, por haver somente um elemento pertencente ao grupo. No *cluster* 08, na Tabela 3 abaixo, é possível perceber essa situação.

Tabela 3. Cluster 08 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|-------------------------------|----------|--------|-------------------|-----------|-----------|-----------|
| 1 | 02 2003 | Traumatismo crânio-encefálico | Feminino | 6 anos | Rio de Janeiro RJ | Mauá SP | Dezembro | 2003 |

E finalmente, o gráfico abaixo, Figura 8, exibe a média harmônica das métricas internas em cada *cluster* formado pelo modelo Cassiopeia.

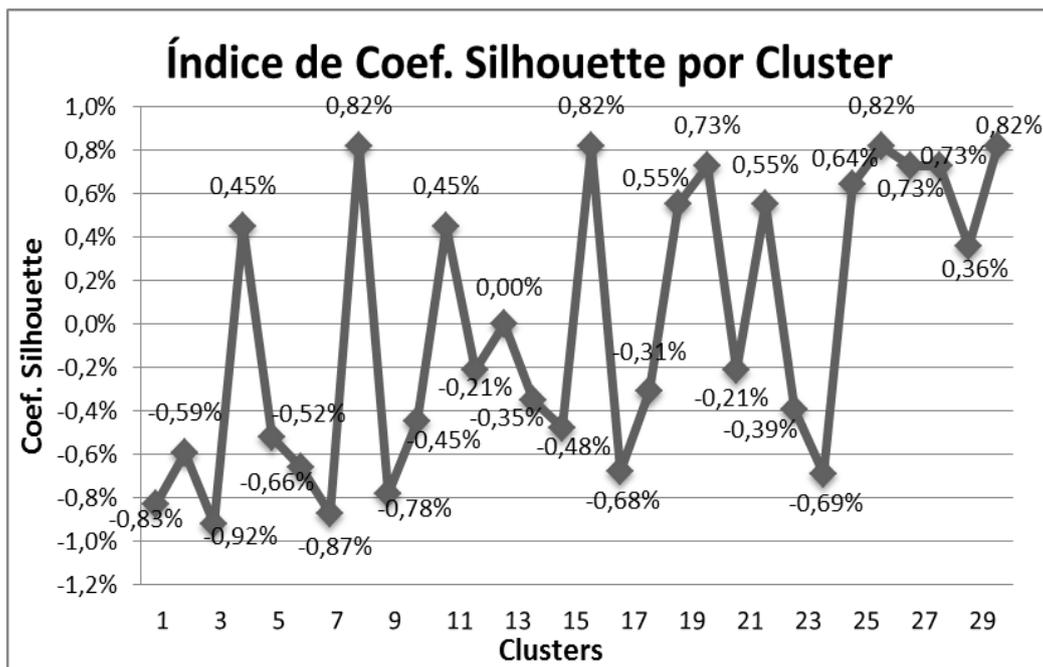


Figura 8. Valor do coeficiente Silhouette em cada cluster.

Devido à sensibilidade do modelo Cassiopeia em relação ao baixo número de palavras dos registros de óbitos, a Figura 8 passa um entendimento de que grande parte dos *clusters* que foram formados não possuem coeficientes Silhouette válidos, pois a maioria dos agrupamentos possui este índice abaixo de 0%. Porém, como mencionado no início desta seção, foi possível comprovar a validade desses agrupamentos através da análise humana dos *clusters* formados.

4.2 ANÁLISE HUMANA DOS CLUSTERS FORMADOS

Além da comparação dos resultados das métricas, foi possível obter alguns resultados práticos com os registros dos óbitos, através da análise dos agrupamentos e da criação de visualizações gráficas que foram baseadas no relatório criado para a realização da análise humana, conforme Apêndice A.

Através desse relatório é possível identificar alguns padrões nos agrupamentos formados pelo modelo Cassiopeia, em que os mesmos podem ser estudados, gerando novos conhecimentos. Os *clusters* a seguir demonstram alguns padrões que foram descobertos na pesquisa através da análise da especialista em pediatria, a Dra. Alison (CRM: 58624-MG). A análise completa pode ser encontrada no Apêndice B.

No *cluster* 01, Tabela 4, 62% das causas dos óbitos em recém-nascidos são por prematuridade ou prematuridade extrema. Geralmente esta patologia necessita de aporte

de uma UTI neonatal para melhorar a sobrevida destes pacientes, porém estes óbitos ocorreram entre 2002 e 2011, e neste período a cidade ainda não contava com este serviço, que foi inaugurado somente em junho de 2012, contribuindo significativamente para ocorrência dos óbitos em questão.

Quanto ao óbito das três crianças (uma de três anos, natural e residente do Serro, uma de um ano de Presidente Kubitscheck, e uma de um ano, residente em Corinto) com quadros de sepse e broncopneumonia, a distância das cidades em relação à Diamantina e o tempo de evolução da doença podem ter colaborado com os casos de óbito. Estas cidades tem como referência a cidade de Diamantina, com exceção de Corinto que deve ter transferido a criança por falta de leito local.

A criança de Corinto já chegou à Diamantina num período que já possuía UTI neonatal, mas ainda não possuía UTI pediátrica que são geralmente para as crianças acima de 28 dias até 12 anos de idade.

Tabela 4. Cluster 01 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|------------------------|-------------------|---------------------------------|-----------|-----------|
| 1 | 03 2002 | prematuridade extrema | Feminino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Abril | 2002 |
| 2 | 02 2005 | prematuridade extrema | Feminino | 2 dias (0 ano) | Diamantina MG | Nada Consta | Maio | 2005 |
| 3 | 06 2004 | prematuridade extrema | Feminino | 5 dias (0 ano) | Diamantina MG | Nada Consta | Novembro | 2004 |
| 4 | 01 2002 | prematuridade | Masculino | 4 dias (0 ano) | Diamantina MG | Nada Consta | Março | 2002 |
| 5 | 04 2002 | prematuridade | Masculino | 23 dias (0 ano) | Diamantina MG | Nada Consta | Maio | 2002 |
| 6 | 05 2002 | prematuridade | Masculino | 2 dias (0 ano) | Diamantina MG | Nada Consta | Junho | 2002 |
| 7 | 06 2011 | ilegível, prematuridade extrema | Feminino | 6 dias (0 ano) | Diamantina MG | Diamantina MG | Setembro | 2011 |
| 8 | 05 2008 | distúrbio hidroeletrólítico, prematuridade | Feminino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Maio | 2008 |
| 9 | 08 2012 | prematuridade extrema | Masculino | 15 dias (0 ano) | Diamantina MG | Diamantina MG | Maio | 2012 |
| 10 | 02 2002 | ilegível | Masculino | 6 dias (0 ano) | Diamantina MG | Nada Consta | Março | 2002 |
| 11 | 08 2002 | ilegível | Feminino | 9 meses (0 ano) | Diamantina MG | Diamantina MG | Agosto | 2002 |
| 12 | 10 2002 | ilegível | Masculino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Novembro | 2002 |
| 13 | 02 2012 | parada cardio respiratória, prematuridade extrema | Feminino | 3 horas (0 ano) | Diamantina MG | Diamantina MG | Dezembro | 2012 |
| 14 | 01 2012 | parada cardio respiratória, sepse | Masculino | 1 ano | Curvelo MG | Corinto MG | Dezembro | 2012 |
| 15 | 09 2012 | choque cardiogênico, pneumotórax hipertensivo, prematuridade extrema | Feminino | 0 ano | Diamantina MG | Diamantina MG | Maio | 2012 |
| 16 | 14 2012 | prematuridade extrema | Feminino | 1, 30 horas (0 ano) | Diamantina MG | Diamantina MG | Julho | 2012 |
| 17 | 33 2012 | prematuridade extrema | Feminino | 30 minutos (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |
| 18 | 27 2012 | prematuridade extrema | Feminino | 30 minutos (0 ano) | Diamantina MG | Diamantina MG | Outubro | 2012 |
| 19 | 08 2003 | septicemia, distúrbio eletrolítico | Masculino | 1 ano | Diamantina MG | Nada Consta | Fevereiro | 2003 |
| 20 | 05 2011 | choque séptico, broncopneumonia | Masculino | 1 ano | Diamantina MG | Presidente Kubitscheck MG | Julho | 2011 |
| 21 | 10 2007 | choque séptico, septicemia, diarreia aguda | Masculino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Novembro | 2007 |
| 22 | 07 2012 | choque séptico, septicemia, prematuridade | Feminino | 0 ano | Sabinópolis MG | Sabinópolis MG | Abril | 2012 |
| 23 | 22 2012 | choque séptico, broncopneumonia extensa | Feminino | 3 anos | Serro MG | Serro MG | Setembro | 2012 |
| 24 | 10 2009 | choque séptico, septicemia, hemorragia pulmonar | Feminino | 12 dias (0 ano) | Diamantina MG | Nada Consta | Setembro | 2009 |

Quanto à faixa etária, Figura 9, geralmente os óbitos ocorrem até o primeiro ano de vida pela fragilidade e vulnerabilidade destas crianças neste período e este é um dos grandes desafios do Ministério da Saúde: tentar a cada ano diminuir estes índices.

Por isso, a necessidade de manter programas de atenção primária com controle de baixo peso, campanhas de vacinação e acompanhamento de gestantes, dentre outros.

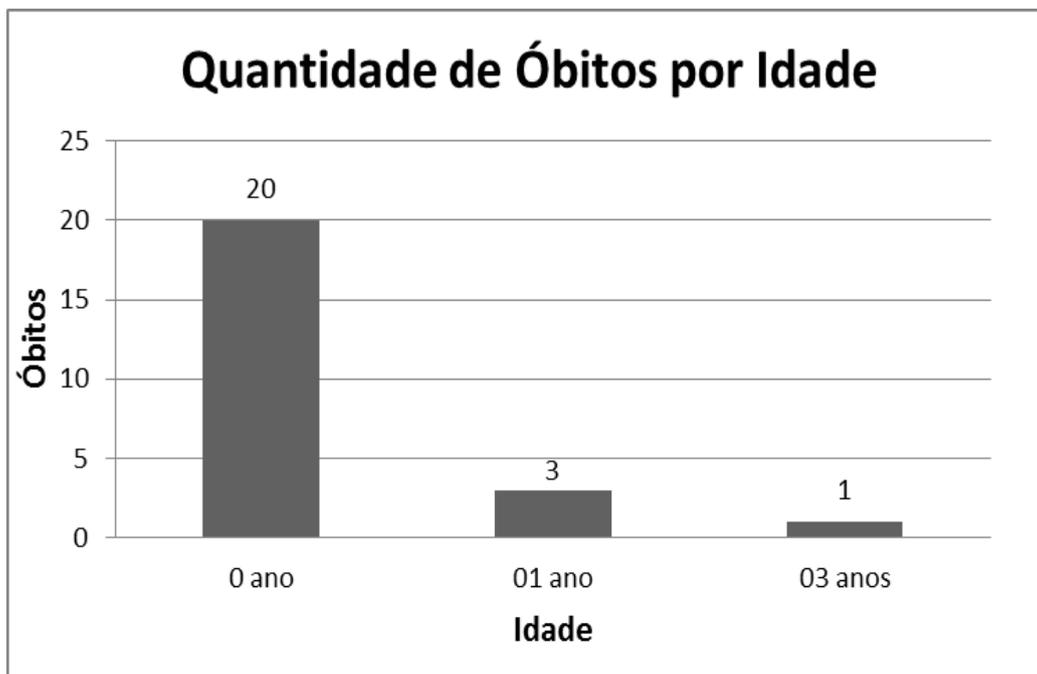


Figura 9. Quantidade de óbitos por idade no *cluster* 01.

Quanto à naturalidade, Figura 10, o predomínio de nascimentos é de Diamantina, porque esta cidade possui maternidade e está como referência para os pacientes das cidades vizinhas. Logo, o predomínio de óbitos no que tange ao critério naturalidade, reflete este fato.

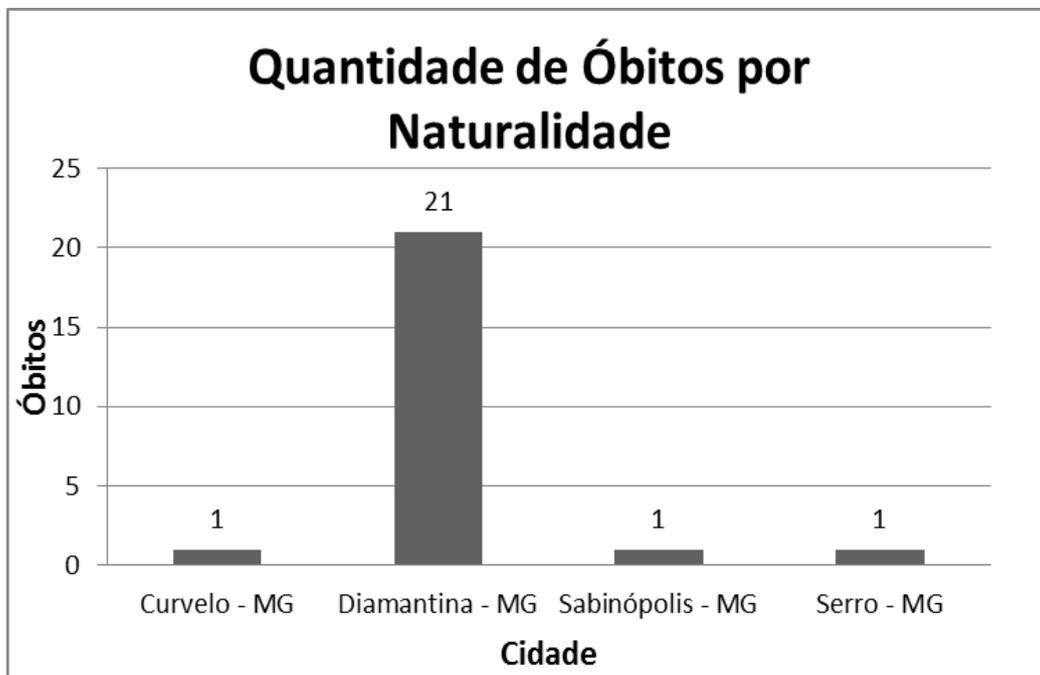


Figura 10. Quantidade de óbitos por naturalidade no *cluster* 01.

Nos óbitos por mês, Figura 11, os casos de hipertensão arterial aumentam no período que antecede diretamente o inverno, o que pode ser justificável o grande índice de óbitos em maio. Já os casos de doenças infecciosas e sepses, aconteceram no período próximo ao verão, onde ocorrem as predominâncias dessas enfermidades.

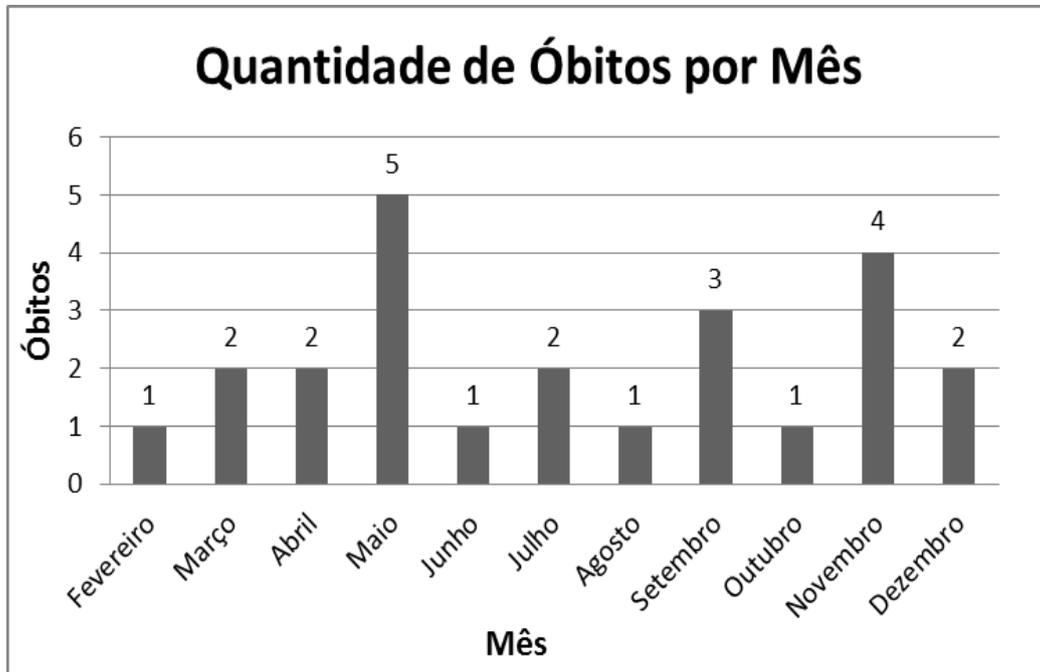


Figura 11. Quantidade de óbitos por mês no *cluster* 01.

Quanto à quantidade de óbitos por ano, Figura 12, talvez a dificuldade em encontrar um profissional especializado, como, no caso, um pediatra, um neonatologista, ou um intensivista pediátrico, possa justificar o predomínio de óbitos no ano de 2012.

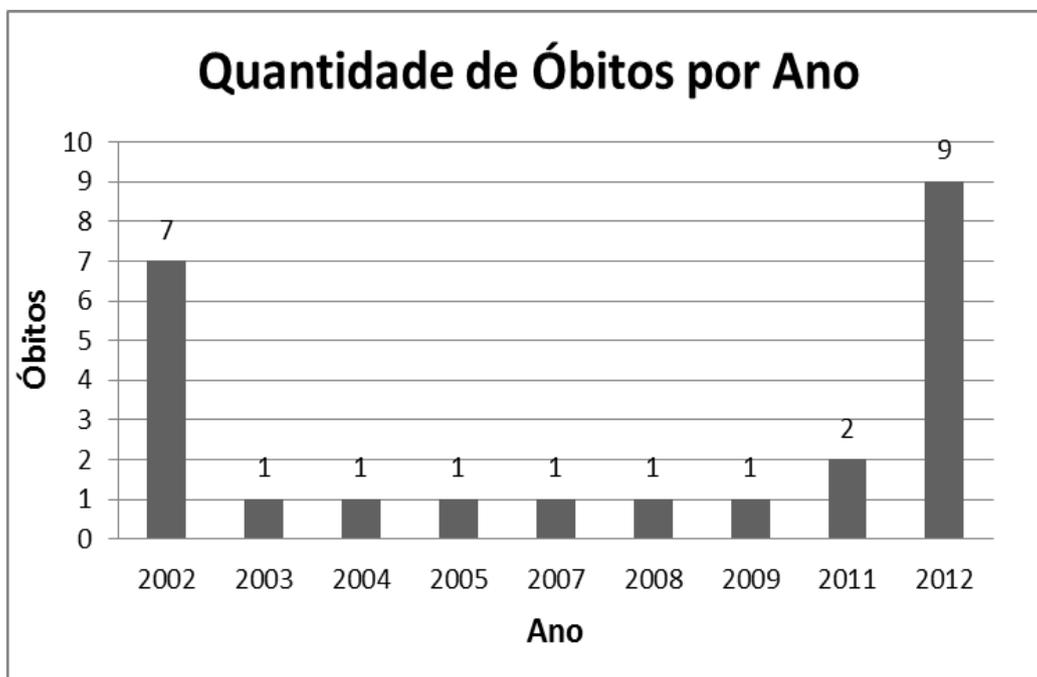


Figura 12. Quantidade de óbitos por ano no *cluster* 01.

Desde 2010 a saúde pública vem enfrentando sérios problemas com a falta de profissionais da área de pediatria. Nos programas de saúde da família, não há a presença de pediatra, contando somente com a presença de um clínico geral. E quando há este profissional especializado, isto só vem a piorar a qualidade da saúde infantil, devido ao excesso de trabalho a um único médico. Além da falta de pediatras, há a evasão destes profissionais da rede pública, principalmente em zonas rurais e regiões mais distantes, como, o Vale do Jequitinhonha.

No *cluster* 02, exibido na Tabela 5, é possível perceber que todos os óbitos foram agrupados por insuficiência respiratória e prematuridade extrema, e como consequência, todas as crianças morreram recém-nascidas. Mesmo após a implantação da UTI neonatal em Diamantina, não foi possível a sobrevivência destas crianças em decorrência da imaturidade física das mesmas.

Tabela 5. Cluster 02 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|---------------------|---------------|---------------|-----------|-----------|
| 1 | 04 2010 | insuficiência respiratória aguda, prematuridade extrema | Feminino | 0 ano | Diamantina MG | Diamantina MG | Dezembro | 2010 |
| 2 | 16 2012 | insuficiência respiratória, hipertensão pulmonar, prematuridade extrema | Masculino | 15 Dias (0 ano) | Diamantina MG | Diamantina MG | Julho | 2012 |
| 3 | 29 2012 | insuficiência respiratória, doença da membrana hialina, prematuridade extrema | Masculino | 20 minutos (0 ano) | Diamantina MG | Diamantina MG | Outubro | 2012 |
| 4 | 31 2012 | insuficiência respiratória, prematuridade extrema | Masculino | 2, 49 horas (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |
| 5 | 30 2012 | insuficiência respiratória, prematuridade extrema | Feminino | 3 horas (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |

Assim como no *cluster* 01, todas as crianças nasceram e viveram em Diamantina-MG.

Na Figura 13, é possível visualizar a quantidade de pessoas que morreram em cada mês/ano.

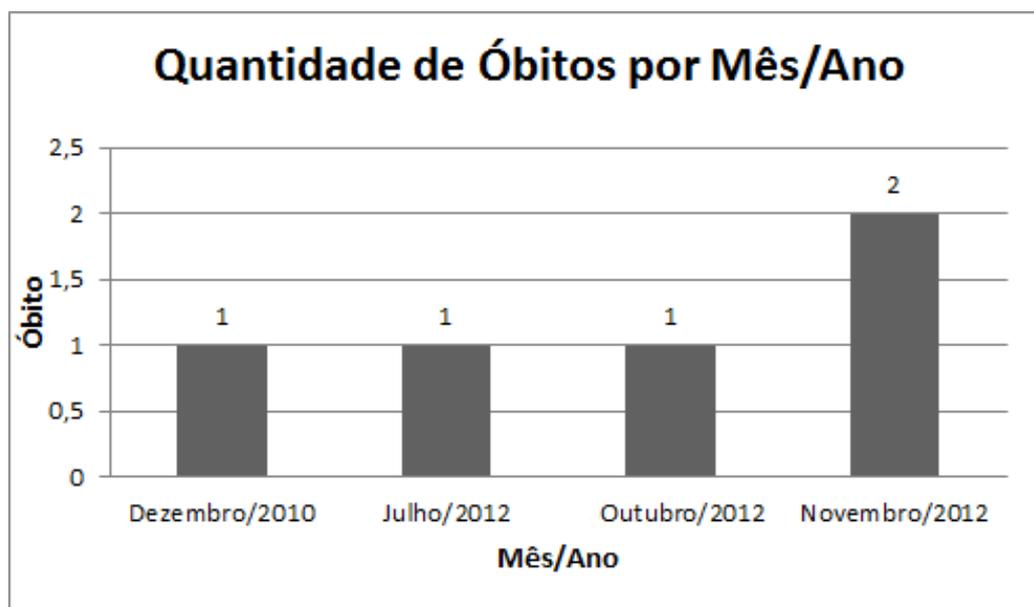


Figura 13. Quantidade de óbitos por mês/ano no cluster 02.

Ao analisar a Figura 13, é possível perceber que dos cinco óbitos, quatro ocorreram no final do ano, nos meses de outubro, novembro e dezembro.

Já na Figura 14, foi ilustrada a quantidade de óbitos que ocorreram por ano no *cluster* 02.

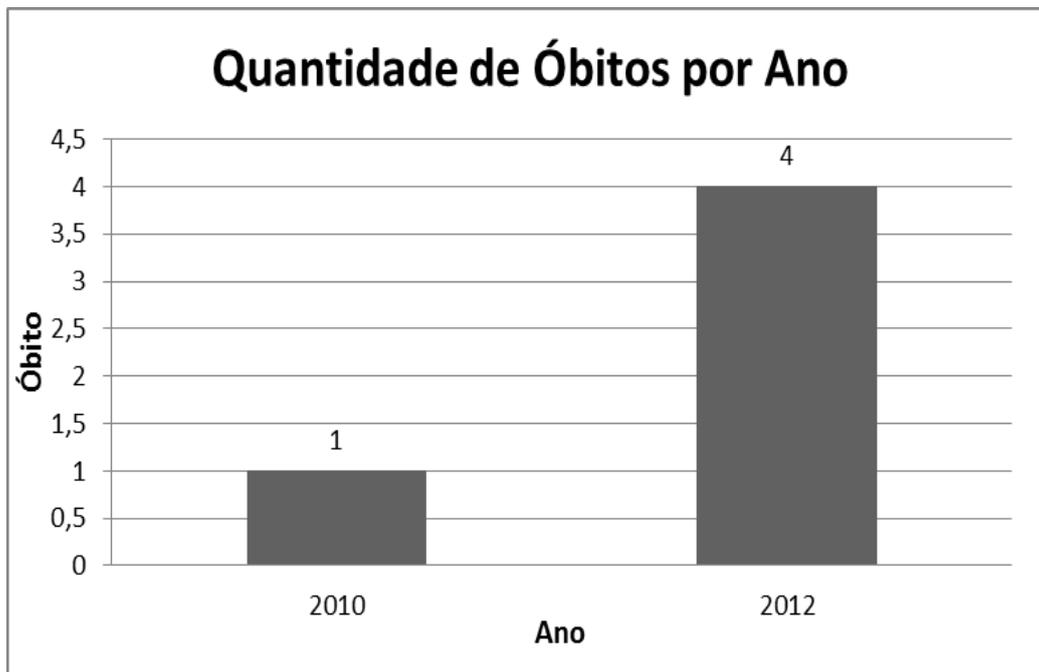


Figura 14. Quantidade de óbitos por ano no *cluster* 02.

Na Figura 14, 80% dos óbitos ocorreram no ano de 2012. Baseando-se no fato de que a maioria desses óbitos aconteceu no final do ano, conforme a Figura 13, e por dados publicados em sites especializados em tempo, como o G1: (<http://goo.gl/xfOf7K>), neste ano de 2012 houve um verão bastante rigoroso, com temperaturas acima da média, e na cidade de Diamantina não foi diferente. No ano de 2010, também houve temperaturas elevadas, se igualando ao recorde de um dos anos mais quentes da história, segundo o site BBC Brasil: (<http://goo.gl/k3ULHW>).

Devido a essas condições climáticas, houve possivelmente um aumento da pressão arterial das gestantes, obrigando-as a entrar em trabalho de parto antes do tempo correto, levando a criança recém-nascida à prematuridade extrema, e como consequência, à insuficiência respiratória.

No *cluster* 03, Tabela 6, os óbitos foram agrupados por diagnóstico de doenças relacionadas à insuficiência respiratória. As idades são variadas, conforme apresentadas na Figura 15.

Tabela 6. Cluster 03 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|-----------------------|------------------|------------------------|-----------|-----------|
| 1 | 07 2005 | insuficiência respiratória aguda | Masculino | 1 ano | Diamantina MG | Diamantina MG | Agosto | 2005 |
| 2 | 04 2003 | insuficiência respiratória, trauma cervical contuso | Feminino | 10 anos | Independência CE | São Paulo SP | Dezembro | 2003 |
| 3 | 01 2006 | insuficiência respiratória aguda grave, aspiração de vômito, crise convulsiva | Feminino | 1 ano | Diamantina MG | Couto de Magalhães MG | Março | 2006 |
| 4 | 04 2008 | Insuficiência respiratória aguda | Feminino | 4 meses (0 ano) | Diamantina MG | Nada Consta | Abril | 2008 |
| 5 | 07 2006 | insuficiência respiratória aguda, sepse | Feminino | 2 anos | Diamantina MG | Diamantina MG | Agosto | 2006 |
| 6 | 08 2009 | insuficiência respiratória aguda, broncopneumonia bilateral extensa | Feminino | 8 anos | Diamantina MG | Diamantina MG | Agosto | 2009 |
| 7 | 10 2010 | insuficiência respiratória aguda | Masculino | 6 anos | Diamantina MG | Cardoso, Diamantina MG | Maio | 2010 |
| 8 | 09 2010 | insuficiência respiratória aguda | Masculino | 1, 10 minutos (0 ano) | Diamantina MG | Nada Consta | Maio | 2010 |
| 9 | 21 2012 | falência múltipla de Órgãos e tecidos, leucemia | Feminino | 15 anos | Serro MG | Milho Verde, Serro MG | Setembro | 2012 |
| 10 | 06 2008 | insuficiência respiratória aguda, anemia crônica descompensada | Masculino | 2 anos | Diamantina MG | Couto de Magalhães MG | Junho | 2008 |
| 11 | 03 2008 | insuficiência respiratória, sepse neonatal, prematuridade | Feminino | 10 dias (0 ano) | Diamantina MG | Nada Consta | Abril | 2008 |
| 12 | 15 2011 | leucemia mieloide aguda, seqüela neurológica por PCR | Feminino | 15 anos | Diamantina MG | Diamantina MG | Abril | 2011 |
| 13 | 02 2010 | insuficiência hepática, anemia falciforme, insuficiência renal aguda, CIVD | Feminino | 13 anos | Coluna MG | Coluna MG | Agosto | 2010 |
| 14 | 12 2009 | insuficiência respiratória sofrimento fetal agudo grave | Masculino | 4 dias (0 ano) | Diamantina MG | Nada Consta | Outubro | 2009 |
| 15 | 07 2011 | insuficiência respiratória, desnutrição grave | Feminino | 16 anos | Diamantina MG | Diamantina MG | Setembro | 2011 |
| 16 | 23 2012 | insuficiência respiratória, desnutrição grave, síndrome Coccayne | Feminino | 5 anos | Itamarandiba MG | Itamarandiba MG | Setembro | 2012 |
| 17 | 05 2010 | sofrimento fetal agudo, prolapso de cordão | Feminino | 0 ano | Diamantina MG | Nada Consta | Janeiro | 2010 |
| 18 | 08 2007 | falência orgânica múltipla, choque séptico, pneumonia de aspiração | Feminino | 14 anos | Minas Novas MG | Minas Novas MG | Setembro | 2007 |
| 19 | 06 2007 | sepse, falência múltiplos órgãos | Masculino | 12 anos | Diamantina MG | Diamantina MG | Maio | 2007 |
| 20 | 09 2002 | caquexia neoplásica, câncer renal | Masculino | 7 anos | Diamantina MG | Inháí MG | Outubro | 2002 |
| 21 | 08 2005 | choque séptico, gangrena, injeção de anabolizante | Masculino | 14 anos | Diamantina MG | Diamantina MG | Agosto | 2005 |
| 22 | 05 2007 | broncopneumonia | Feminino | 1 ano | Diamantina MG | Diamantina MG | Abril | 2007 |

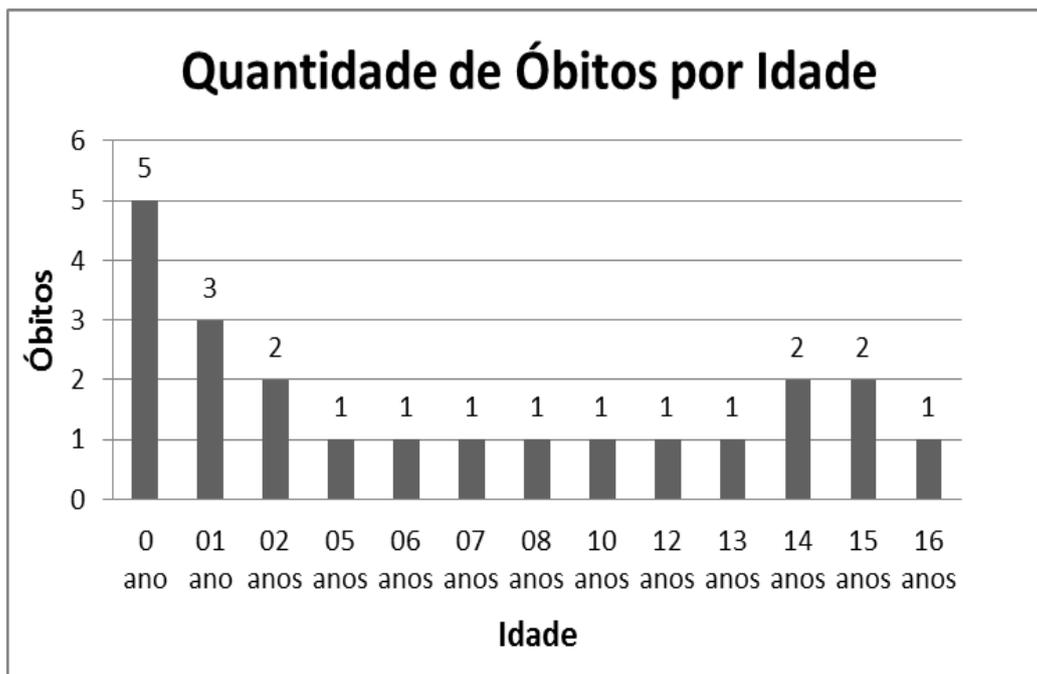


Figura 15. Quantidade de óbitos por idade no *cluster* 03.

Quanto ao ano, dois casos ocorreram após a implantação da UTI neonatal e pediátrica, ou seja, em 2012. Porém um caso tratou-se de falência múltipla por leucemia mieloide. Nesta situação, o paciente já estava em estágio terminal, e este óbito seria inevitável.

A leucemia na faixa etária descrita (15 anos), é mais comum no sexo feminino, o que foi de fato observado. Como mostrado na Figura 15, houve dois óbitos nesta idade, e ambos os casos possuíam o diagnóstico de leucemia e eram do sexo feminino. Quanto aos critérios de tratamento e sobrevida dos pacientes com este tipo de doença, cabe uma discussão à parte.

A leucemia mieloide aguda, encontrada em um dos dois casos, é menos comum e muito mais agressiva, com um prognóstico muito ruim. Já nas crianças, predominam a leucemia linfóide aguda, que tem muitas chances de cura e menos casos de óbito.

No *cluster* 05, Tabela 7, os registros relatam sepse e choque séptico como causas de óbito predominantes.

Tabela 7. Cluster 05 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|--------------------|---|----------------------------|-----------|-----------|
| 1 | 05 2009 | insuficiência respiratória, sepse, anemia | Feminino | 1 dia (0 ano) | Gouveia MG | Nada Consta | Fevereiro | 2009 |
| 2 | 03 2011 | choque séptico, sepse neonatal precoce | Feminino | 5 dias (0 ano) | Diamantina MG | Datas MG | Julho | 2011 |
| 3 | 03 2006 | sepse neonatal tardia | Masculino | 20 dias (0 ano) | Couto de Magalhães de Minas MG | Nada Consta | Março | 2006 |
| 4 | 06 2005 | choque séptico | Masculino | 14 anos | Diamantina MG | Diamantina MG | Julho | 2005 |
| 5 | 32 2012 | choque não especificado | Masculino | 9 dias (0 ano) | Serro MG | Serro MG | Novembro | 2012 |
| 6 | 28 2012 | choque séptico | Feminino | 0 ano | Curvelo MG | Curvelo MG | Outubro | 2012 |
| 7 | 20 2012 | choque séptico, broncopneumonia infectada | Masculino | 8 anos | Serro MG | Alvorada de Minas MG | Agosto | 2012 |

Pode-se perceber que a maioria estava na faixa etária menor que 30 dias, e dentre eles, dois eram de outra cidade distante de Diamantina, o que pode ter agravado as complicações, além do processo de transporte das crianças.

Quanto ao ano de 2012, este ano faz parte do período de carência dos profissionais médicos, o que pode justificar, mesmo com a presença da UTI neonatal e pediátrica, a ocorrência dos óbitos, Figura 16. Já o caso de óbito do adolescente de 14 anos, no item 4, é exceção, uma vez que é raro o óbito por sepse nesta idade, a menos que, já sofresse com uma doença de base, o que não foi mencionado no atestado de óbito.

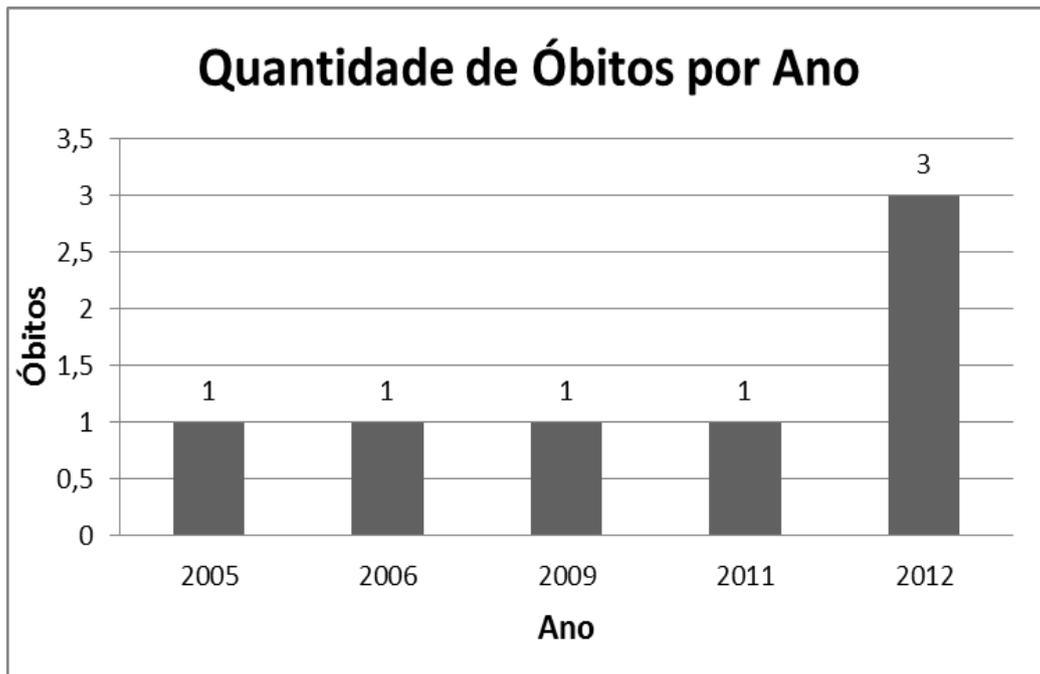


Figura 16. Quantidade de óbitos por ano no *cluster* 05.

No *cluster* 07, Tabela 8, os registros relacionam patologias respiratórias (insuficiência respiratória, trombose pulmonar, pneumonias) e más formações congênitas, chamando a atenção para a evolução do tempo de sobrevivência aumentado de acordo com o ano. Em 2002 o óbito foi bem mais precoce em termos de faixa etária do que em 2012, em função de uma melhor assistência com a evolução dos anos.

Tabela 8. Cluster 07 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|------------------|-------------------|-------------------------|-----------|-----------|
| 1 | 01 2007 | pneumonia, má formação SNC | Feminino | 4 meses (0 ano) | Belo Horizonte MG | Carbonita MG | Fevereiro | 2007 |
| 2 | 02 2004 | insuficiência respiratória, pneumonia aguda | Masculino | 15 anos | Diamantina MG | Diamantina MG | Maio | 2004 |
| 3 | 04 2012 | bronco pneumonia aspirativa, insuficiência cardíaca, cardiopatia congênita | Feminino | 3 anos | Curvelo MG | Gouveia MG | Março | 2012 |
| 4 | 07 2009 | AVC hemorrágico, má formação arterio venosa | Feminino | 15 anos | Curvelo MG | Gouveia MG | Julho | 2009 |
| 5 | 08 2010 | insuficiência cardiorrespiratória, cardiopatia congênita | Masculino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Março | 2010 |
| 6 | 05 2004 | pneumonia aspirativa | Masculino | 3 meses (0 ano) | Diamantina MG | Nada Consta | Agosto | 2004 |
| 7 | 09 2007 | parada cardiorrespiratória, insuficiência cardíaca congestiva, cardiopatia grave | Masculino | 11 anos | Diamantina MG | Diamantina MG | Setembro | 2007 |
| 8 | 07 2002 | cardiopatia congênita completa | Feminino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Julho | 2002 |
| 9 | 13 2009 | tromboembolismo pulmonar, trombose venosa profunda | Masculino | 14 anos | Itamarandiba MG | Carbonita MG | Outubro | 2009 |
| 10 | 01 2010 | insuficiência respiratória aguda, pneumonia bacteriana aguda, paralisia cerebral não especificada | Feminino | 5 anos | Diamantina MG | Pinheiro, Diamantina MG | Junho | 2010 |
| 11 | 17 2012 | insuficiência renal aguda, choque hipotérmico, pneumonia, paralisia cerebral | Masculino | 3 anos | Serro MG | Alvorada de Minas MG | Julho | 2012 |
| 12 | 19 2012 | insuficiência respiratória aguda, choque séptico, pneumonia, crises convulsivas | Masculino | 2 meses (0 ano) | Capelinha MG | Capelinha MG | Agosto | 2012 |
| 13 | 08 2011 | insuficiência respiratória, pneumonia aspirativa, paralisia cerebral, síndrome de West | Feminino | 6 anos | Rio Vermelho MG | Serro MG | Setembro | 2011 |
| 14 | 14 2009 | choque séptico, meningite bacteriana, hidrocefalia aguda | Feminino | 10 meses (0 ano) | Sete Lagoas MG | Corinto MG | Outubro | 2009 |

Faz-se exceção o item 6, pois pneumonia aspirativa é um acidente após uma bronco aspiração (engasgo mais severo) de crianças abaixo de um ano, não dependendo de mês ou ano, mas sim da vigilância dos responsáveis pela criança. Já com relação ao item 13, a criança possuía seis anos, mas cursava com patologia neurológica, o que a torna semelhante a uma criança com menos de um ano. Com relação ao item 14, a meningite é uma patologia grave, que pode levar ao óbito independente da idade.

No *cluster* 09, Tabela 9, grande parte dos registros está relacionada a traumatismos crânio encefálicos, e também está associado às épocas do ano com festividades, Figura 17, o que possivelmente pode ter contribuído para a causa destes óbitos em decorrência de imprudência no trânsito e no convívio com a coletividade.

Tabela 9. Cluster 09 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|---------------------|------------------------|-----------------|-----------|-----------|
| 1 | 25 2012 | traumatismo crânio encefálico | Masculino | 16 anos | Guanhães MG | Guanhães MG | Outubro | 2012 |
| 2 | 11 2002 | traumatismo crânio encefálico | Feminino | 9 anos | Turmalina MG | Pitangueiras SP | Dezembro | 2002 |
| 3 | 06 2002 | parada cardíaca respiratória, traumatismo crânio encefálico | Masculino | 16 anos | Diamantina MG | Diamantina MG | Junho | 2002 |
| 4 | 03 2003 | traumatismo crânio-encefálico contuso | Feminino | 8 anos | Santo Amaro SP | Santo Amaro SP | Dezembro | 2003 |
| 5 | 04 2005 | Insuficiência cardíaca, respiratória, septicemia, hidrocefalia | Feminino | 2, 20 horas (0 ano) | Diamantina MG | Nada Consta | Maio | 2005 |
| 6 | 02 2008 | insuficiência respiratória, hidrocefalia infectada | Masculino | 3 meses (0 ano) | Belo Horizonte-MG | Itamarandiba MG | Abril | 2008 |
| 7 | 03 2005 | traumatismo cranio encefálico, acidente automobilístico | Feminino | 5 meses (0 ano) | Diamantina MG | Nada Consta | Maio | 2005 |
| 8 | 09 2009 | traumatismo craniano encefálico, projétil de arma de fogo | Feminino | 16 anos | Serra Azul de Minas MG | Diamantina MG | Agosto | 2009 |
| 9 | 11 2012 | traumatismo craniano, acidente automobilístico | Feminino | 14 anos | Guanhães MG | Guanhães MG | Junho | 2012 |
| 10 | 10 2011 | acidente vascular encefálico hemorrágico | Feminino | 14 anos | Turmalina MG | Turmalina MG | Novembro | 2011 |
| 11 | 04 2009 | estado de mal epiléptico, tumor cerebral | Feminino | 4 anos | Capelinha MG | Capelinha MG | Janeiro | 2009 |

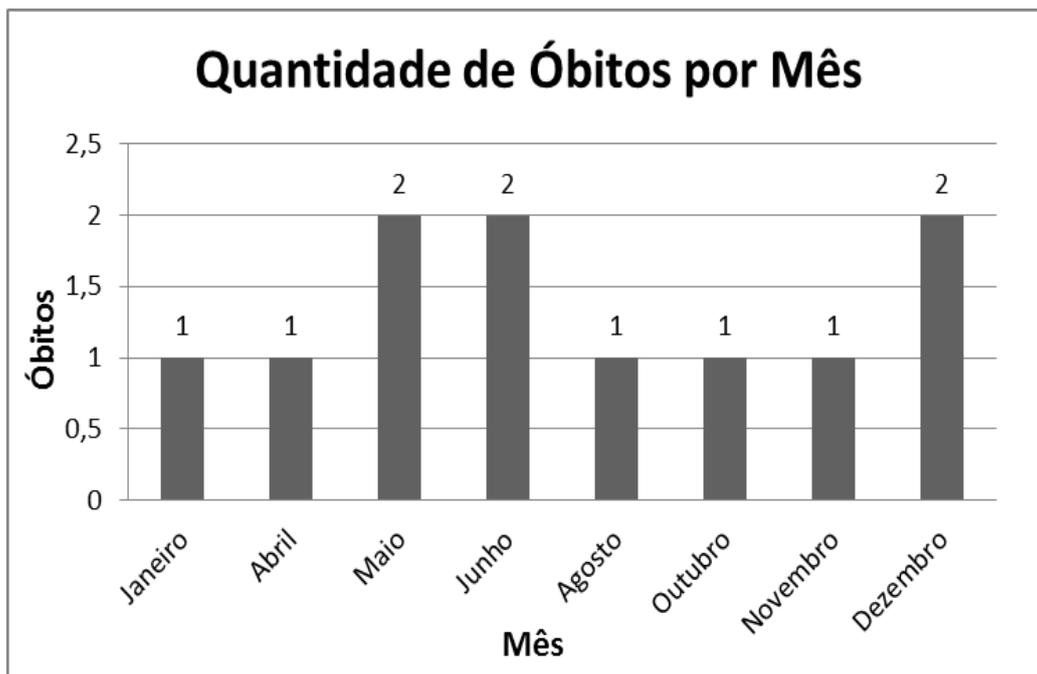


Figura 17. Quantidade de óbitos por mês no *cluster* 09.

No *cluster* 15, Tabela 10, quase todos os registros relacionam-se a sepse e parada cardio respiratória, que ocorreram em sua maioria em Diamantina e distritos, levando a questionar casos de infecção hospitalar na cidade de Diamantina.

Tabela 10. *Cluster* 15 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|------------------|---------------|-------------------------------|-----------|-----------|
| 1 | 04 2007 | parada cardiorrespiratória, septicemia | Feminino | 4 anos | Diamantina MG | Felício dos Santos MG | Abril | 2007 |
| 2 | 06 2005 | parada cardiorrespiratória, morte súbita neonatal | Masculino | 14 dias (0 ano) | Diamantina MG | Nada Consta | Novembro | 2005 |
| 3 | 07 2003 | septicemia, desnutrição | Masculino | 9 meses (0 anos) | Diamantina MG | Diamantina MG | Dezembro | 2003 |
| 4 | 09 2008 | desnutrição III | Feminino | 7 anos | Diamantina MG | Senador Mourão, Diamantina MG | Novembro | 2008 |
| 5 | 07 2010 | pneumonia adquirida na comunidade, amiotrofia Espanhol grau III | Masculino | 16 anos | Diamantina MG | Sopa, Diamantina MG | Fevereiro | 2010 |
| 6 | 05 2006 | parada cardiorrespiratória | Feminino | 2 anos | Diamantina MG | Diamantina MG | Junho | 2006 |

No *cluster* 21, Tabela 11, os registros se destacam, devido aos óbitos de causa indeterminada, pois são casos de óbitos que chegam ao hospital, e o médico pode emitir

o atestado de óbito como causa indeterminada, ou enviar para o IML para necropsia, que seria o mais correto.

Tabela 11. Cluster 21 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|--------------------|------------------|---|-----------|-----------|
| 1 | 04 2004 | causa indeterminada | Masculino | 3 anos | Diamantina MG | Diamantina MG | Julho | 2004 |
| 2 | 17 2009 | óbito de causa indeterminada | Masculino | 15 anos | Diamantina MG | Ribeirão de Areia, Diamantina MG | Dezembro | 2009 |
| 3 | 18 2012 | causa indeterminada, hidrocefalia, paralisia cerebral | Masculino | 12 anos | Diamantina MG | Diamantina MG | Julho | 2012 |
| 4 | 06 2006 | indeterminada | Masculino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Junho | 2006 |
| 5 | 01 2004 | indeterminada | Masculino | 3 meses (0 ano) | Diamantina MG | Nada Consta | Janeiro | 2004 |
| 6 | 09 2011 | indeterminada, epilepsia | Masculino | 16 anos | Diamantina MG | Diamantina MG | Novembro | 2011 |

Chama a atenção este excesso de causas indeterminadas neste *cluster*, e a ação do médico na emissão de atestado de óbito. Este comportamento profissional muda totalmente as estatísticas de causa de óbito por ano no país e deve ser evitado. Todos os óbitos foram do sexo masculino, mas como não se sabe as causas, fica difícil uma avaliação baseada nessa informação.

No *cluster* 24, Tabela 12, os registros 1, 2 e 3 trata-se de asfixia por afogamento e os mesmos estão relacionados a uma faixa etária maior (adolescentes), e a uma época do ano ainda de temperaturas mais altas, o que possibilita estes tipos de acidentes, principalmente em piscinas e cachoeiras, que são os tipos de entretenimento mais comuns na região. Geralmente os adolescentes tendem a se aventurar e se acham onipotentes, podendo levar a estes desfechos.

Tabela 12. Cluster 24 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|---------------|---------------|---------------|-----------|-----------|
| 1 | 03 2007 | asfixia mecânica por obstrução das vias aéreas por meio líquido (afogamento) | Feminino | 12 anos | Diamantina MG | Diamantina MG | Abril | 2007 |
| 2 | 02 2007 | asfixia mecânica por obstrução das vias aéreas por meio líquido (afogamento) | Feminino | 10 anos | Diamantina MG | Diamantina MG | Abril | 2007 |
| 3 | 07 2004 | afogamento | Masculino | 14 anos | São Paulo SP | Serro MG | Dezembro | 2004 |
| 4 | 13 2002 | Síndrome de aspiração de mecônio, sofrimento fetal crônico | Masculino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Dezembro | 2002 |
| 5 | 06 2010 | asfixia por sufocação direta | Masculino | 4 anos | Diamantina MG | Diamantina MG | Fevereiro | 2010 |
| 6 | 08 2008 | asfixia por soterramento | Masculino | 6 anos | Diamantina MG | Diamantina MG | Agosto | 2008 |

No item 4, a síndrome de aspiração de mecônio ocorre quando a criança passa do tempo de nascimento ou o líquido amniótico torna-se infectado por bolsa rota. Trata-se de um líquido atípico que pode levar a morte por asfixia durante o nascimento ou horas após. Os itens 5 e 6 registram asfixia por soterramento e sufocação direta, que nesse caso são situações particulares e ocasionais sem relação às anteriores.

No *cluster* 25, Tabela 13, o item 1 do registro trata de um choque séptico por leptospirose em uma criança mais velha. Chama a atenção para a investigação epidemiológica da doença, que poderia ter sido feita a fim de rastrear esta enfermidade em familiares, vizinhos e afins.

Tabela 13. Cluster 25 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|------------------------------|-----------|---------------|---------------|---------------|-----------|-----------|
| 1 | 07 2007 | choque séptico, leptospirose | Feminino | 10 anos | Diamantina MG | Diamantina MG | Agosto | 2007 |
| 2 | 03 2012 | choque séptico (septicemia) | Masculino | 1 mês (0 ano) | Curvelo MG | Nada Consta | Dezembro | 2012 |

No item 2 é mais provável de acontecer este tipo de doença nesta faixa etária, porém chama a atenção quanto ao problema da vigilância da infecção hospitalar. As outras informações detalhadas do relatório podem ser encontradas no Apêndice A.

Com a clusterização, houve a possibilidade da análise dessas informações e, conseqüentemente, a geração de descoberta de conhecimento. Essas análises e outras

dos agrupamentos gerados pelo modelo Cassiopeia, estão no relatório criado, Apêndice B, pela especialista em pediatria, a Dra. Alison (CRM: 58624-MG), cujo mesmo relatório será encaminhado à Secretária de Saúde de Diamantina para análise e conhecimento público.

4.3 COMPROVAÇÃO DA HIPÓTESE

Com os resultados obtidos e apresentados na seção, foi comprovada a hipótese de que é possível descobrir conhecimento em informações de óbito de crianças da cidade de Diamantina e região, com faixa etária entre zero e 16 anos, em um espaço de tempo de 2002 a 2012.

4.4 TRABALHOS CORRELATOS

Nesta seção são exibidos os trabalhos correlatos, onde seus resultados são comparados aos resultados obtidos através dos agrupamentos formados pelo modelo Cassiopeia, e da análise do relatório dos *clusters*.

Loh (2001) aborda os conceitos para realizar a descoberta de conhecimento em textos. Ao invés de aplicar as técnicas de mineração sobre palavras ou dados estruturados extraídos de textos, a abordagem explora conceitos identificados no texto. A ideia é analisar o conhecimento codificado em textos num nível acima das palavras, ou seja, não analisando somente os termos e expressões presentes nos textos, mas seu significado em relação aos fenômenos da realidade (pessoas, objetos, entidades, eventos e situações do mundo real) (LOH, 2001).

Nas conclusões das avaliações, Loh (2001) mostra que a diferença de desempenho entre as abordagens baseadas em conceitos e baseada em palavras não foi grande na sua pesquisa. Pode-se concluir que não há vantagem de um método de representação sobre outro em processos de classificação de textos, podendo-se usar palavras ou conceitos como características representantes do conteúdo dos textos. Entretanto, as regras utilizadas pelo método baseado em conceitos são mais fáceis de serem entendidas. Isto facilita a explicação ou mesmo a validação do raciocínio usado no processo de classificação (LOH, 2001).

Wives (2004), assim como Loh (2001), avalia a utilização dos conceitos como descritores de texto e forma de descoberta de conhecimento. O autor define conceito como uma estrutura capaz de representar uma ideia ou um assunto contido em um

documento. Os conceitos foram utilizados para representar o conteúdo dos documentos, sendo usados como descritores no processo de identificação de similaridade entre os textos e, para que, conseqüentemente, estes sejam agrupados em conglomerados de assuntos similares.

Os resultados obtidos no trabalho de Wives (2004) indicam que a metodologia proposta apresenta resultados interessantes, relativamente melhores, mais fáceis de serem interpretados do que, quando, palavras são utilizadas. E também possuem menor complexidade, conseqüentemente levando menos tempo para processar uma grande quantidade de documentos.

Ramos e Brascher (2009) relatam sobre resultados de pesquisa aplicando a DCT em conteúdos textuais, considerados como importantes fontes de informação para tomada de decisão. O objetivo central da pesquisa foi verificar a eficácia da DCT na descoberta de informações para apoio à construção de indicadores de políticas públicas, utilizando como estudo de caso o Serviço Brasileiro de Respostas Técnicas (SBRT), por meio do agrupamento de documentos a partir dos termos minerados na base de dados.

Ao classificar os documentos no trabalho de Ramos e Brascher (2009), pôde-se obter resultados que demonstraram a coerência entre os documentos, que de maneira geral, sugeriam tratar de temas diversos, contudo, com uma visão central de tratamento e reaproveitamento de materiais de diferentes naturezas e para vários fins. Já na análise dos termos minerados, os resultados mostraram que a partir da extração, foi possível aprofundar as análises, tornando evidente o diferencial de se utilizar a DCT para a exploração de informações ocultas em documentos textuais (RAMOS E BRASCHER, 2009).

A aplicação da DCT proporcionou a visualização das fortes relações existentes entre documentos que aparentemente não teriam interação entre si, caso fossem consideradas as classificações e palavras-chave que lhes foram atribuídas (RAMOS E BRASCHER, 2009).

Vianna *et al.* (2010) realizaram o estudo que busca identificar padrões de características materno-fetais na predição da mortalidade infantil, por meio da incorporação de técnicas de mineração de dados, que se mostram relevantes em saúde pública. A ferramenta utilizada para a mineração foi o WEKA (Waiko Environment for

Knowledge Analysis) (<http://www.cs.waikato.ac.nz/ml/weka>), que permite a descoberta de padrões na forma de um classificador representado como árvore de decisão. Os resultados mostraram que a integração e o relacionamento das bases de dados, diferentes dos sistemas de informação de saúde do governo do Paraná, auxiliam a complementar, recuperar e rever as informações, permitindo uma aproximação epidemiológica real, existente na população estudada (VIANNA *et al.*, 2010).

A Tabela 14 exibe um comparativo entre trabalhos correlatos que foram citados e utilizados como base para este trabalho e este próprio trabalho em questão, mostrando o nome dos autores dos trabalhos e suas formas de Pré-processamento, Processamento, e Pós-processamento, demonstrando se houve utilização de interação humana em alguma das fases da pesquisa.

Tabela 14. Tabela comparativa dos trabalhos correlatos.

| Autor (Ano) | Pré Processamento | Processamento | Pós Processamento | Interação Humana |
|-----------------------------|---|---|--|------------------|
| LOH (2001) | Stopword + Descritores de vocabulário controlado ou padronizado. | Agrupamento hierárquico aglomerativo por conceitos. | Descoberta de Conhecimento em Textos – DCT | Sim |
| WIVES (2004) | Stopword + Descritores de vocabulário controlado ou padronizado. | Agrupamento hierárquico aglomerativo por conceitos. | Descoberta de Conhecimento em Textos – DCT | Sim |
| RAMOS E BREASCHER (2009) | Remoção de palavras não significativas + Lematização | Agrupamento dos documentos | Descoberta de Conhecimento em Textos - DCT + Interação Humana | Sim |
| VIANNA <i>et al.</i> (2010) | Remoção de palavras + Limpeza dos dados não significativos + Correção | Classificação representada como árvore de decisão por conceitos | Descoberta de Conhecimento em Dados - DCD + Interação Humana | Sim |

CAPÍTULO 5 – CONCLUSÕES

Com a análise dos resultados das métricas de análise textual e do relatório dos *clusters* que foram formados, pode-se concluir que o modelo Cassiopeia apresenta dados interessantes e concisos, comprovando que a clusterização dos óbitos pediátricos possui um alto grau de similaridade e trazem descobertas de conhecimento importantes para o contexto pediátrico da cidade de Diamantina e região.

A obtenção dos dados de óbitos na fase de pré-processamento, proporcionou uma fonte de informações textuais, para que fosse clusterizada através do modelo Cassiopeia, na fase de processamento. Logo após, os resultados foram analisados para que padrões pudessem ser descobertos.

A etapa de pré-processamento foi feita por meio da análise dos textos, pela especialista em pediatria, a Dra. Alison Guelpeli (CRM: 58624-MG) que averiguou e verificou a necessidade de manter ou retirar alguns termos técnicos e possíveis erros de ortografia que pudessem atrapalhar esta fase, que é a parte mais custosa da DCT.

Não houve a necessidade de executar a técnica de sumarização, pois devido ao pouco volume de palavras e termos, era possível que os dados fossem tratados sem a utilização do modelo Cassiopeia na fase de pré-processamento. Bastou apenas, uma síntese eficiente e justificada por meio da análise humana.

Na fase de processamento, foi realizado o método de agrupamento de textos, através do modelo Cassiopeia, onde os mesmos foram agrupados em *clusters* por níveis de similaridades de seus termos. Com este método, foi possível gerar 30 *clusters* que foram utilizados para serem avaliados através das métricas textuais do modelo.

A análise dos resultados obtidos após o processamento foi dividida em duas fases, em que a primeira exibiu os resultados obtidos pelo modelo Cassiopeia, através de métricas textuais internas. Essas métricas forneceram os índices de coesão, de coeficiente Silhouette de cada *cluster* e os seus índices médios, além da média dos índices de acoplamento dos agrupamentos que foram gerados.

As métricas comprovaram um bom rendimento do modelo Cassiopeia, apesar dos registros terem um número inferior a 885 palavras, demonstrando que os resultados

foram válidos, por representarem o bom nível de similaridade existente entre os elementos dos *clusters*.

A segunda fase dos experimentos, abordou a análise humana nos *clusters* formados, dos quais foram obtidos resultados práticos com os registros dos óbitos, através da observação dos agrupamentos e da criação de visualizações gráficas que auxiliaram na análise do relatório, feito pela especialista em pediatria, a Dra. Alison Guelpeli (CRM: 58624-MG).

Os relatórios possuíam informações como: o número total de registros, o número de registros de cada agrupamento, as informações de óbitos de cada *cluster*, e suas respectivas ilustrações gráficas.

Com a análise da primeira fase dos resultados, pode-se concluir que o agrupamento de textos, feito pelo modelo Cassiopeia, apresentaram resultados animadores, pois os textos clusterizados possuem um baixo índice de acoplamento e um alto índice de coesão. Conseqüentemente, os dados possuem um alto índice de similaridade entre os elementos do mesmo grupo, enquanto que a similaridade entre os elementos de grupos diferentes é baixa.

Na análise da segunda fase dos resultados, através do relatório criado, com informações e ilustrações gráficas dos *clusters*, foi possível identificar padrões nos agrupamentos formados pelo modelo Cassiopeia e a análise desses padrões possibilitaram a descoberta de conhecimento.

Casos, como, possíveis infecções hospitalares, a falta de profissionais da área pediátrica e casos de acidentes em épocas específicas do ano, foram alguns dos principais conhecimentos descobertos que chamaram a atenção nesta etapa.

Os resultados encorajam a dizer que o modelo Cassiopeia pode trazer benefícios, quando usado como uma ferramenta que auxilia na descoberta de conhecimento em óbitos pediátricos.

É importante ressaltar a importância da descoberta de conhecimento na área de óbitos pediátricos, pois através dela é possível identificar diversos problemas relacionados com fatores que podem levar ao falecimento de crianças. A relação entre descoberta de conhecimento e óbitos pediátricos, possivelmente poderá gerar diversas

contribuições em questões de pesquisas e prevenções para a sociedade diamantinense e região, podendo trazer um impacto altamente positivo.

Através destas evidências, este trabalho possivelmente poderá auxiliar no desenvolvimento de propostas para políticas públicas, que poderão prevenir causas de óbito por consequência de doenças, acidentes ou agressões em crianças.

Com este trabalho, possivelmente poderá ocorrer o desenvolvimento de diversos serviços que trarão benefícios para a cidade de Diamantina e região, gerando alertas e prevenções sobre erros do setor médico ou de ambientes que possam estar contribuindo para o óbito na faixa etária estudada. Novas formas de tratamento de doenças poderão ser desenvolvidas, além da criação de campanhas que fornecerão modos de prevenção, podendo auxiliar na diminuição dos óbitos, através da mudança do modo de vida dos moradores da região analisada.

Técnicas para descoberta de conhecimento são muito úteis para serem usadas como aplicações na vida real, através de métodos computacionais que auxiliam na tomada de decisões importantes para a sociedade.

Pode-se concluir que a descoberta de conhecimento é importante para a identificação de informações relevantes que podem ser computadas, sendo possível gerar descobertas que mudam o estado atual do conhecimento que se tinha em questão. Através desta descoberta, o que era um conhecimento definido, se torna uma forma inicial de conhecimento, em que o mesmo é complementado pelo novo conhecimento adquirido.

5.1 LIMITAÇÕES

O processo de obtenção dos dados de óbitos foi demorado e cansativo, pois os dados não são digitalizados. Todas as informações estavam em cadernos registrados por máquina de datilografar e alguns dados estavam ilegíveis devido às rasuras feitas ao digitar os dados incorretamente.

Como não era permitida a retirada desses dados do cartório, foi preciso obter a cópia dessas informações manualmente, o que demandou um alto custo de tempo.

5.2 TRABALHOS FUTUROS

Como trabalhos futuros, poderão ser apresentadas mais análises dos gráficos e dos relatórios gerados para a descoberta de conhecimento por meio da análise de mais de um especialista humano. Também poderão ser feitas comparações dos resultados do modelo Cassiopeia com outra ferramenta computacional de clusterização, capaz de realizar o processamento dos textos obtidos neste trabalho.

Ampliar a faixa etária para a aplicação de descoberta de conhecimento em pessoas idosas que são propensas a causas de óbitos diferentes da faixa etária estudada.

Ampliar a região de abrangência do trabalho aumentando assim o espectro, saindo da questão regional.

REFERÊNCIAS

- Beppler, M., Fernandes, A. (2005) “Aplicação de text mining para a extração de conhecimento Jurisprudencial”. Primeiro Congresso Sul Catarinense de Educação.
- Boente, A. N. P., Estrela, V.V. e Goldschmidt, R.R. (2006) “Uma metodologia para apoio à realização do processo de descoberta de conhecimento em base de dados”. II WCC.
- Ceci, F. (2010) “Um Modelo Semiautomático para a Construção e Manutenção de Ontologias a partir de Bases de Documentos Não Estruturados”. 131 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Florianópolis.
- Clark, P., Harrison, P. (2009) “Large-Scale Extraction and Use of Knowledge From Text”. Fifth International Conference on Knowledge Capture.
- Feldman, R., Dagan, I. (1995) “Knowledge discovery in textual databases (KDT)”. Knowledge Discovery and Data Mining, p. 112–117.
- Fialho, F.A.P. *et al.* (2006) “Gestão do conhecimento e aprendizagem: as estratégias competitivas da sociedade pós-industrial”. Florianópolis: Visualbooks.
- Filho, L. A. da S. (2009) “Mineração de Regras de Associação Utilizando KDD e KDT: Uma Aplicação em Segurança Pública”. 85 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará. Belém.
- Guelpeli M. V. C. (2012) "Cassiopeia: Um modelo de agrupamento de textos baseado em sumarização". 220 f. Tese (Doutorado em Inteligência Artificial) - Programa de Pós-Graduação em Computação da UFF. Universidade Federal Fluminense, Niterói.
- Hotho, A., Nurnberger, A., Paass, G. (2005) "A brief survey of text mining". LDV Forum - GLDV Journal for Computational Linguistics and Language Technology, v. 20, n. 1, p. 19-62.
- IBGE. Minas Gerais – Belo Horizonte. Censo Demográfico 2010. Disponível em: <<http://cidades.ibge.gov.br/painel/populacao.php?lang=&codmun=310620&search=|belo-horizonte>> Acesso em: 05 mai. 2014.
- IBGE. Minas Gerais – Diamantina. Censo Demográfico 2010. Disponível em: <<http://cidades.ibge.gov.br/painel/populacao.php?lang=&codmun=312160&search=|diamantina>> Acesso em: 05 mai. 2014.
- IBGE. São Paulo – São Paulo. Censo Demográfico 2010. Disponível em: <<http://cidades.ibge.gov.br/xtras/perfil.php?lang=&codmun=355030&search=sao-paulo|sao-paulo>> Acesso em: 05 mai. 2014.
- Jain, A. K. *et al.* (1999) "Data Clustering: a review". ACM Computing Surveys, New York, v. 31, n. 3, p. 264-323.

- Jesus, D. N. (2010) “Estudo e Utilização da Ferramenta Text Mining Suite”. 42f. Relatório apresentado como requisito parcial para Estágio Supervisionado em Sistemas de Informação. Centro Universitário Luterano de Palmas, Palmas.
- Loh, S., Wives, L. K., Oliveira, J. P. M. de, (2000) “Concept based knowledge discovery from texts extracted from the web”. ACM SIGKDD Explorations, vol. 2, pp. 29–40.
- Machado, A. P. *et al.* (2010) "Mineração de Texto em Redes Sociais Aplicada à Educação a Distância". Colabor@ - Revista Digital da CVA - Ricesu, ISSN 1519-8529 , nº 23.
- Mathern B. *et al.* (2012) “Towards a Knowledge-Intensive and Interactive Knowledge Discovery Cycle”. Dans 20th ICCBR Workshop Proceedings, Luc Lamontagne, Juan A. Recio-Garcia ed. Lyon, France. pp. 151-162.
- Milani, C. S., Carvalho, D. R. (2013) "Pós - Processamento em KDD". Revista de Engenharia e Tecnologia ISSN 2176-72 , Paraná, nº 1.
- Morais, E. A., Ambrósio, A. P. (2008) "Automatic Domain Classification of Jurisprudence Documents". EATIS, Aracaju, nº 08.
- Pinheiro, C. A. R. (2008) “Inteligência analítica: mineração de dados e descoberta de conhecimento”. Rio de Janeiro: Ciência Moderna apud Ceci, F. (2010) “Um Modelo Semiautomático para a Construção e Manutenção de Ontologias a partir de Bases de Documentos Não Estruturados”. 131 f. Dissertação (Mestrado) - Universidade Federal de Santa Catarina, Florianópolis.
- Rajaraman, K., Tan, A. (2002) "Knowledge Discovery from Texts: A Concept Frame Graph Approach”. Eleventh International Conference on Information and Knowledge Management (CIKM'02). McLean, Virginia, USA.
- Silva, E. R. G., Rover, A. J. (2011) “O Processo de descoberta do conhecimento como suporte à análise criminal: minerando dados da Segurança Pública de Santa Catarina”. International Conference on Information Systems and Technology Management, São Paulo. Anais da International Conference on Information Systems and Technology Management. São Paulo: FEA, 2011. v. 8.
- Tan, A. H. (1999). "Text Mining: the state of the art and the challenges". Workshop on Knowledge Discovery from Advanced Databases. Proceedings... Heidelberg, p. 491-502 (Lecture Notes in Computer Science, 1574).
- Vianna, R. C. X. F, Barra, C. M. C. M. (2005). “Aplicação da descoberta de conhecimento em base de dados para a identificação de características relacionadas à mortalidade infantil”. Proceedings of International Association for Development of the Information Society. Lisboa: IADIS Press. p. 678-80

- Vianna, R. C. X. F et al (2010). "Mineração de dados e características da mortalidade infantil". Cad. Saúde Pública, Rio de Janeiro, 26(3): 535-542.
- Visa, A. *et al.* (2000) "Knowledge Discovery from Text Documents Based on Paragraph Map". 33rd Hawaii International Conference on System Sciences, Hawaií.
- Xu, K. *et al.* (2000) "Self-Teaching Semantic Annotation Method for Knowledge Discovery from Text". 42nd Hawaii International Conference on System Sciences, Hawaií.
- Wives, L. K. (2004) "Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (*Clustering*) de documentos". 136f. Tese (Doutorado em Ciência da Computação) - Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Zanasi, A. (1998) "Competitive intelligence through datamining public sources". Competitive Intelligence Review, Alexandria, Virginia, v. 9(1), p. 44-54.

APÊNDICES

APÊNDICE A

RELATÓRIOS DOS *CLUSTERS*

O Apêndice A apresenta em detalhes o relatório com as informações dos *Clusters* e dos gráficos de comparação, analisados na seção 4.2, que abordou a análise humana dos *Clusters* formados. Neste relatório são exibidas informações como o número total de registros, o número total de *Clusters*, o número de registros de cada *Cluster* e comparações de informações de cada *Cluster* por meio de ilustrações gráficas, como: a quantidade de óbitos por sexo, a quantidade de óbitos por idade, a quantidade de óbitos por naturalidade e a quantidade de óbitos por residente.

Mineração de Óbitos Pediátricos em Diamantina

Relatório dos *Clusters*

Total de Registros: 147

Total de *Clusters*: 30

Cluster 01:

Número de registros: 24

Tabela:

Tabela 15. Cluster 01 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|------------------------|-------------------|---------------------------------|-----------|-----------|
| 1 | 03 2002 | prematuridade extrema | Feminino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Abril | 2002 |
| 2 | 02 2005 | prematuridade extrema | Feminino | 2 dias | Diamantina MG | Nada Consta | Maio | 2005 |
| 3 | 06 2004 | prematuridade extrema | Feminino | 5 dias | Diamantina MG | Nada Consta | Novembro | 2004 |
| 4 | 01 2002 | prematuridade | Masculino | 4 dias | Diamantina MG | Nada Consta | Março | 2002 |
| 5 | 04 2002 | prematuridade | Masculino | 23 dias (0 ano) | Diamantina MG | Nada Consta | Maio | 2002 |
| 6 | 05 2002 | prematuridade | Masculino | 2 dias (0 ano) | Diamantina MG | Nada Consta | Junho | 2002 |
| 7 | 06 2011 | ilegível, prematuridade extrema | Feminino | 6 dias (0 ano) | Diamantina MG | Diamantina MG | Setembro | 2011 |
| 8 | 05 2008 | distúrbio hidroeletrólítico, prematuridade | Feminino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Maio | 2008 |
| 9 | 08 2012 | prematuridade extrema | Masculino | 15 dias (0 ano) | Diamantina MG | Diamantina MG | Maio | 2012 |
| 10 | 02 2002 | ilegível | Masculino | 6 dias (0 ano) | Diamantina MG | Nada Consta | Março | 2002 |
| 11 | 08 2002 | ilegível | Feminino | 9 meses (0 ano) | Diamantina MG | Diamantina MG | Agosto | 2002 |
| 12 | 10 2002 | ilegível | Masculino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Novembro | 2002 |
| 13 | 02 2012 | parada cardio respiratória, prematuridade extrema | Feminino | 3 horas (0 ano) | Diamantina MG | Diamantina MG | Dezembro | 2012 |
| 14 | 01 2012 | parada cardio respiratória, sepse | Masculino | 1 ano | Curvelo MG | Corinto MG | Dezembro | 2012 |
| 15 | 09 2012 | choque cardiogênico, pneumotórax hipertensivo, prematuridade extrema | Feminino | 0 ano | Diamantina MG | Diamantina MG | Maio | 2012 |
| 16 | 14 2012 | prematuridade extrema | Feminino | 1, 30 horas (0 ano) | Diamantina MG | Diamantina MG | Julho | 2012 |
| 17 | 33 2012 | prematuridade extrema | Feminino | 30 minutos (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |
| 18 | 27 2012 | prematuridade extrema | Feminino | 30 minutos (0 ano) | Diamantina MG | Diamantina MG | Outubro | 2012 |
| 19 | 08 2003 | septicemia, distúrbio eletrólítico | Masculino | 1 ano | Diamantina MG | Nada Consta | Fevereiro | 2003 |
| 20 | 05 2011 | choque séptico, broncopneumonia | Masculino | 1 ano | Diamantina MG | Presidente Kubitscheck MG | Julho | 2011 |
| 21 | 10 2007 | choque séptico, septicemia, diarreia aguda | Masculino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Novembro | 2007 |
| 22 | 07 2012 | choque séptico, septicemia, prematuridade | Feminino | 0 ano | Sabinópolis MG | Sabinópolis MG | Abril | 2012 |
| 23 | 22 2012 | choque séptico, broncopneumonia extensa | Feminino | 3 anos | Serro MG | Serro MG | Setembro | 2012 |
| 24 | 10 2009 | choque séptico, septicemia, hemorragia pulmonar | Feminino | 12 dias (0 ano) | Diamantina MG | Nada Consta | Setembro | 2009 |

Na Tabela 15 é apresentada o Cluster 1 com 24 registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 16% dos 147 registros processados.

Gráficos:

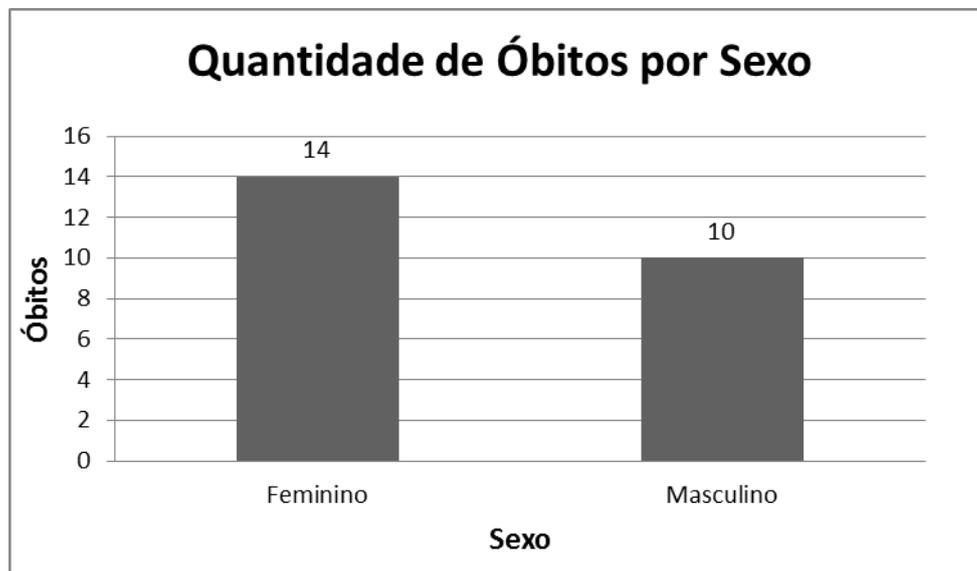


Figura 18. Quantidade de óbitos por sexo no *cluster* 01.

A Figura 18 mostra a quantidade de óbitos por sexo no *cluster* 01 obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com 14 registros, sendo equivalente a 58% dos 24 registros agrupados no *cluster* 01.

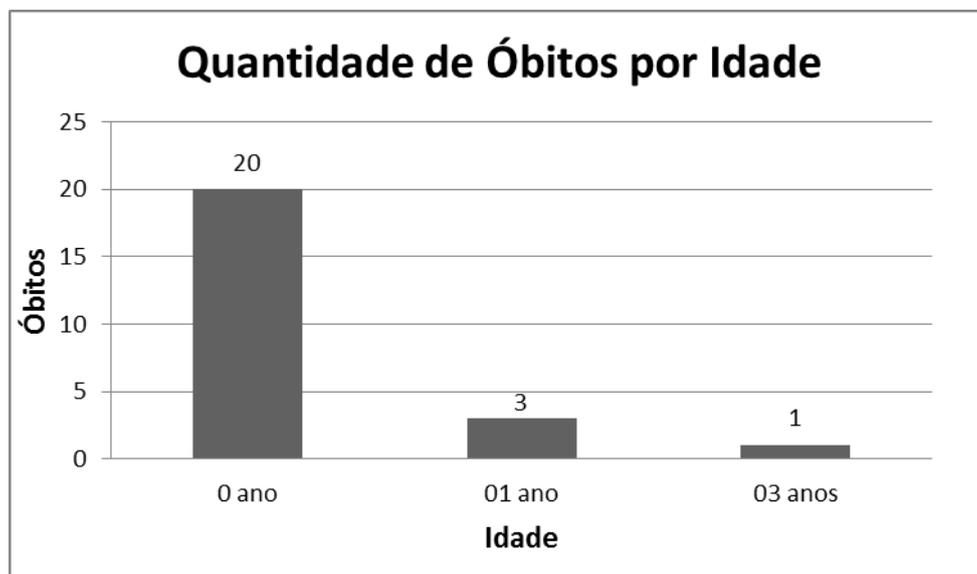


Figura 19. Quantidade de óbitos por idade no *cluster* 01.

A Figura 19 mostra a quantidade de óbitos por idade no *Cluster* 01 obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano com 20 registros, sendo equivalente a 83% dos 24 registros.

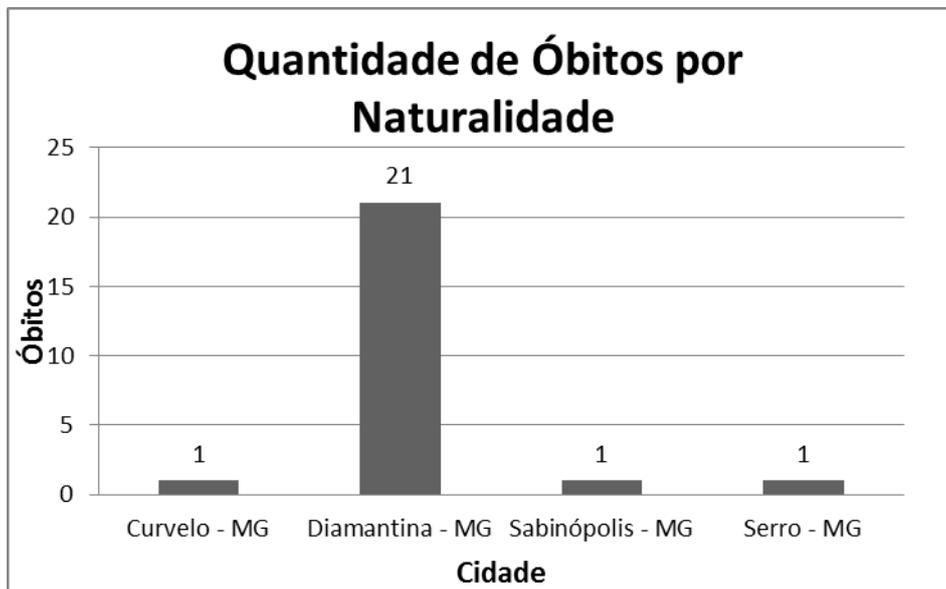


Figura 20. Quantidade de óbitos por naturalidade no *cluster* 01.

A Figura 20 mostra a quantidade de óbitos por naturalidade no *Cluster* 01, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em crianças naturais de Diamantina-MG com 21 registros, sendo equivalente a 87% dos 24 registros.

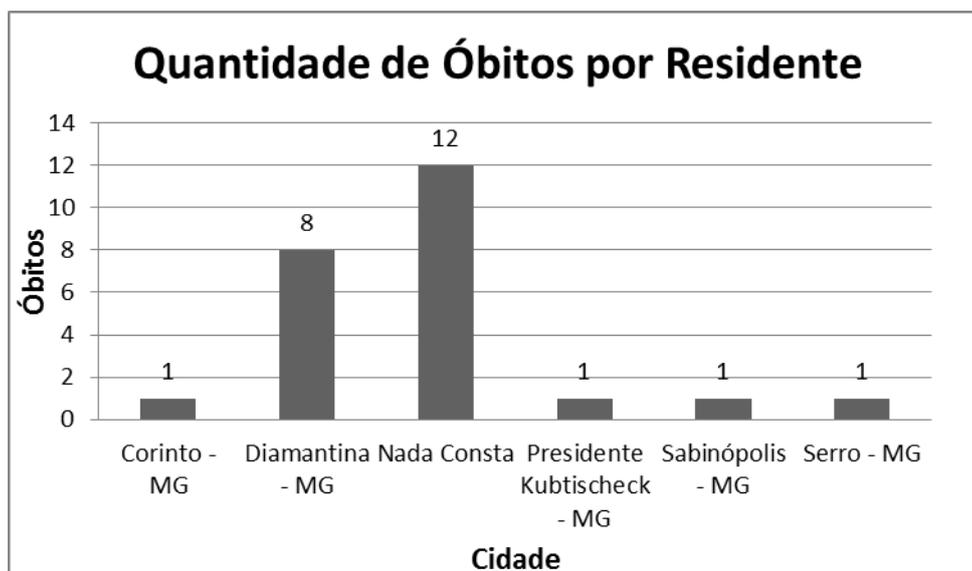


Figura 21. Quantidade de óbitos por residente no *cluster* 01.

A Figura 21 mostra a quantidade de óbitos por residente no *Cluster 01*, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de registros não declarados com um total de 12 óbitos, sendo equivalente a 50% dos 24 registros.

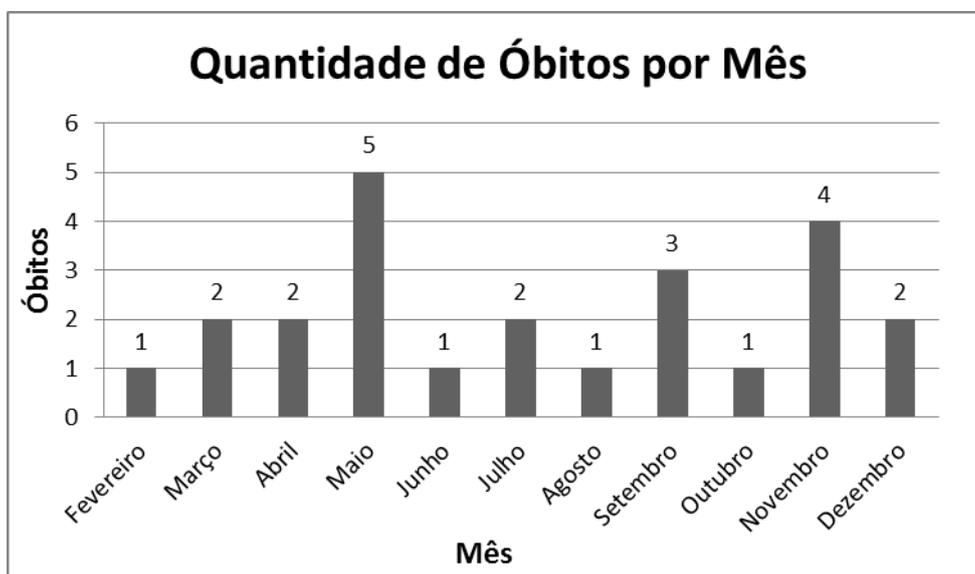


Figura 22. Quantidade de óbitos por mês no *cluster 01*.

A Figura 22 mostra a quantidade de óbitos por mês no *Cluster 01*, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de maio com 5 registros e no mês de novembro com 4 registros, sendo equivalente a 20% e a 16% dos 24 registros respectivamente.

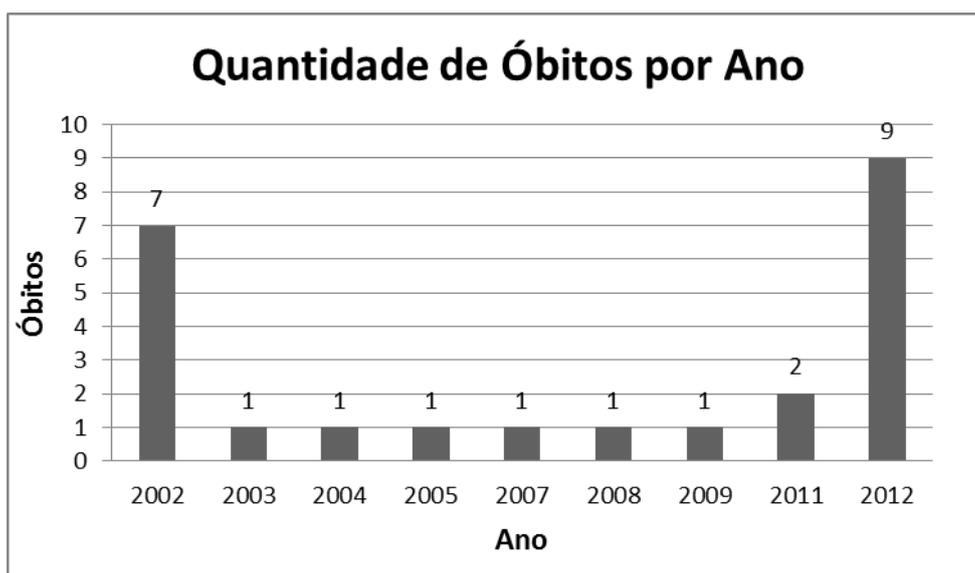


Figura 23. Quantidade de óbitos por ano no *cluster 01*.

A Figura 23 mostra a quantidade de óbitos por ano no *Cluster* 01, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram nos anos de 2002 com 7 registros e 2012 com 9 registros, sendo equivalente a 29% e 37% dos 24 registros respectivamente.

Cluster 02:

Número de Registros: 5

Tabela:

Tabela 16. *Cluster* 02 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|---------------------|---------------|---------------|-----------|-----------|
| 1 | 04 2010 | insuficiência respiratória aguda, prematuridade extrema | Feminino | 0 ano | Diamantina MG | Diamantina MG | Dezembro | 2010 |
| 2 | 16 2012 | insuficiência respiratória, hipertensão pulmonar, prematuridade extrema | Masculino | 15 dias (0 ano) | Diamantina MG | Diamantina MG | Julho | 2012 |
| 3 | 29 2012 | insuficiência respiratória, doença da membrana hialina, prematuridade extrema | Masculino | 20 minutos (0 ano) | Diamantina MG | Diamantina MG | Outubro | 2012 |
| 4 | 31 2012 | insuficiência respiratória, prematuridade extrema | Masculino | 2, 49 horas (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |
| 5 | 30 2012 | insuficiência respiratória, prematuridade extrema | Feminino | 3 horas (0 ano) | Diamantina MG | Diamantina MG | Novembro | 2012 |

Na Tabela 16 é apresentada o *cluster* 2 com cinco registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 3% dos 147 registros processados.

Gráficos:

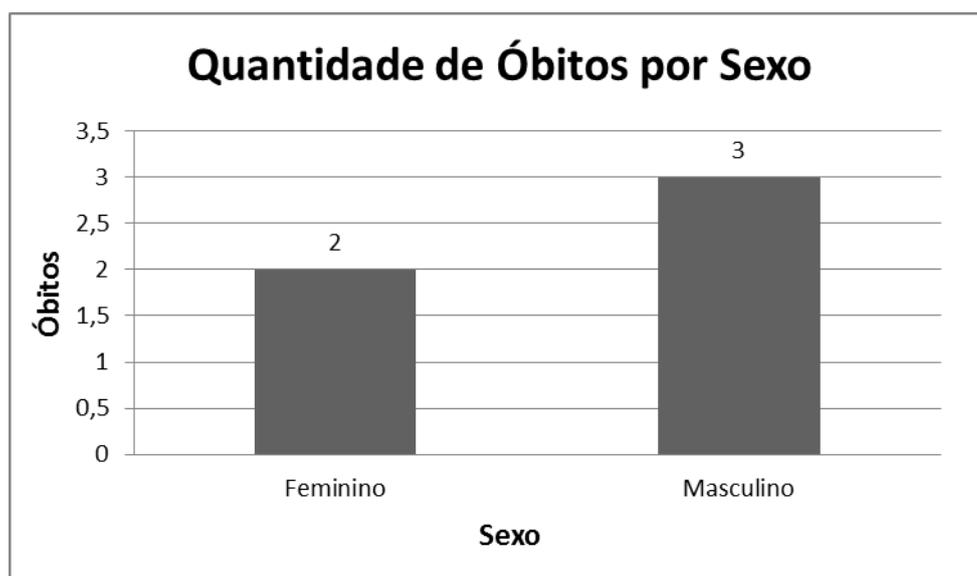


Figura 24. Quantidade de óbitos por sexo no *cluster* 02.

A Figura 24 mostra a quantidade de óbitos por sexo no *Cluster* 02 obtidos pelo modelo Cassioéia. Nesta Figura é possível perceber a predominância de óbitos masculinos com três registros, sendo equivalente a 60% dos cinco registros agrupados no *Cluster* 2.

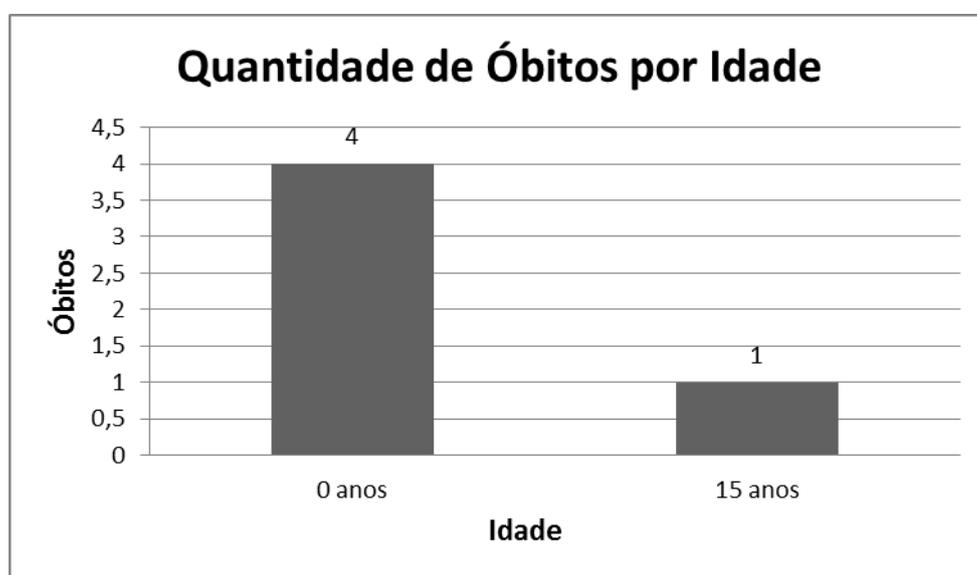


Figura 25. Quantidade de óbitos por idade no *cluster* 02.

A Figura 25 mostra a quantidade de óbitos por idade no *Cluster* 02 obtidos pelo modelo Cassioeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano com quatro registros, sendo equivalente a 80% dos cinco registros agrupados no *Cluster*.

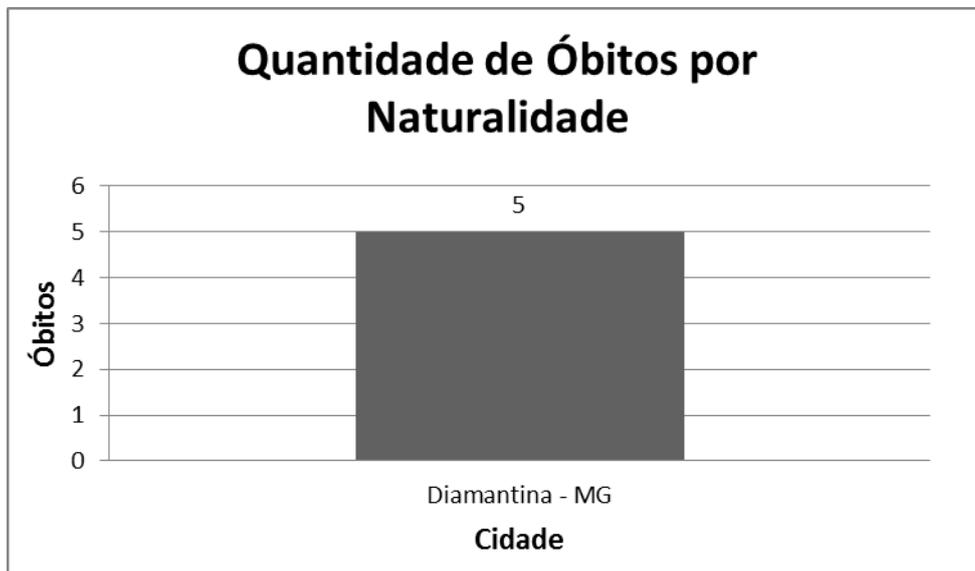


Figura 26. Quantidade de óbitos por naturalidade no *cluster* 02.

A Figura 26 mostra a quantidade de óbitos por naturalidade no *Cluster* 02, obtidos pelo modelo Cassiopeia. Nesta Figura 100% dos óbitos são de crianças naturais de Diamantina-MG com cinco registros agrupados no *Cluster*.

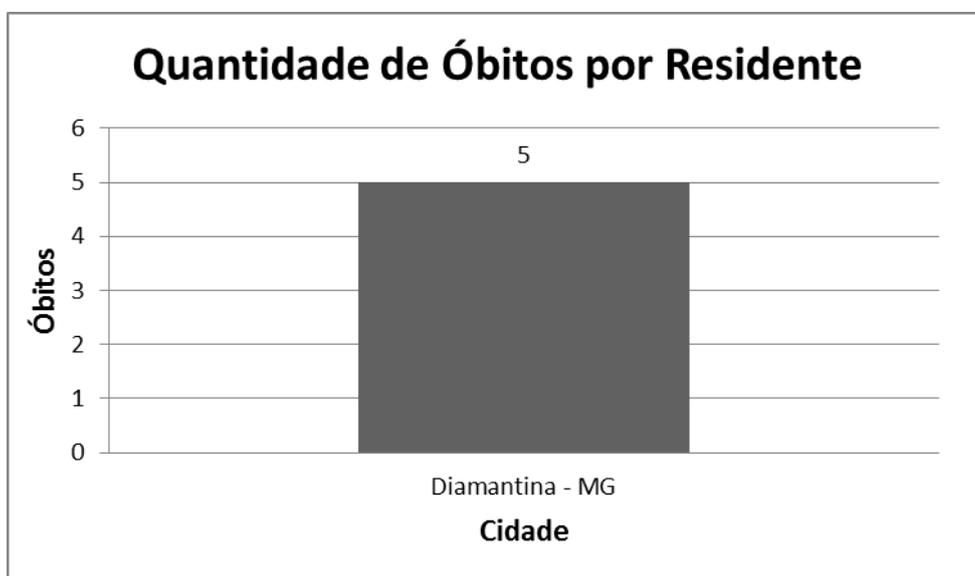


Figura 27. Quantidade de óbitos por residente no *cluster* 02.

A Figura 27 mostra a quantidade de óbitos por residente no *Cluster* 02, obtidos pelo modelo Cassiopeia. Nesta Figura 100% dos óbitos são de crianças que eram residentes em Diamantina-MG, sendo um total de 5 registros agrupados no *Cluster*.

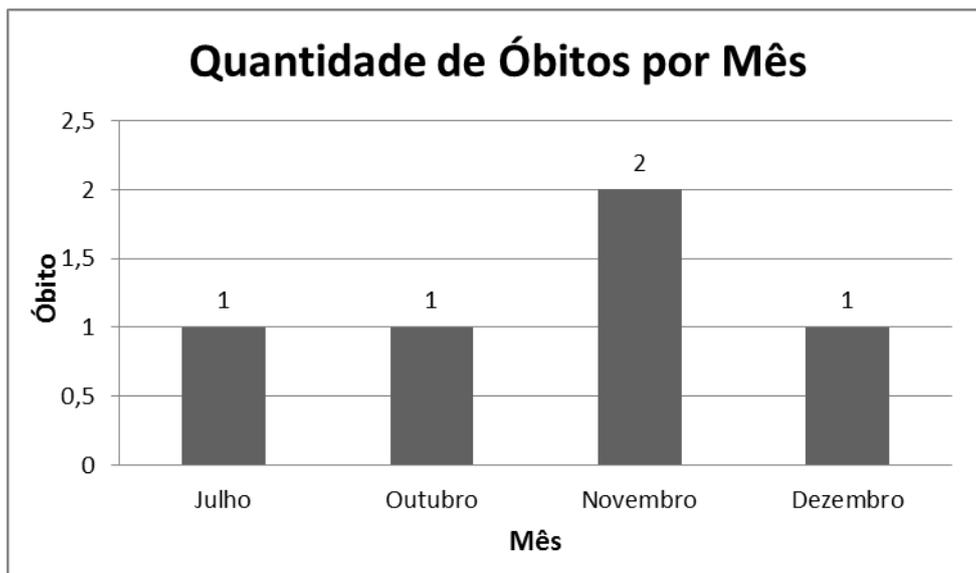


Figura 28. Quantidade de óbitos por mês no *cluster* 02.

A Figura 28 mostra a quantidade de óbitos por mês no *Cluster* 02, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de novembro com dois registros, sendo equivalente a 40% dos cinco registros.

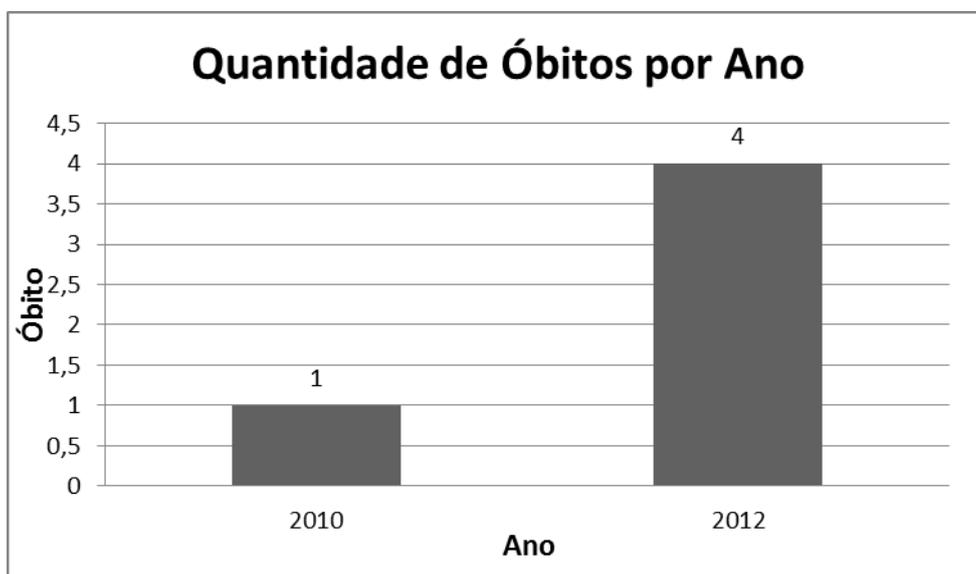


Figura 29 Quantidade de óbitos por ano no *cluster* 02.

A Figura 29 mostra a quantidade de óbitos por ano no *Cluster* 02, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no ano de 2012 com quatro registros, sendo equivalente a 80% dos cinco registros do *Cluster*.

Cluster 03:

Número de registros: 22

Tabela 17. Cluster 03 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|-----------------------------|----------------------|------------------------------|-----------|-----------|
| 1 | 07 2005 | insuficiência respiratória Aguda | Masculino | 1 ano | Diamantina MG | Diamantina MG | Agosto | 2005 |
| 2 | 04 2003 | insuficiência respiratória, trauma cervical contuso | Feminino | 10 anos | Independência CE | São Paulo SP | Dezembro | 2003 |
| 3 | 01 2006 | insuficiência respiratória aguda grave, aspiração de vômito, crise convulsiva | Feminino | 1 ano | Diamantina MG | Couto de Magalhães MG | Março | 2006 |
| 4 | 04 2008 | Insuficiência respiratória aguda | Feminino | 4 meses (0 ano) | Diamantina MG | Nada Consta | Abril | 2008 |
| 5 | 07 2006 | insuficiência respiratória aguda, sepse | Feminino | 2 anos | Diamantina MG | Diamantina MG | Agosto | 2006 |
| 6 | 08 2009 | insuficiência respiratória aguda, broncopneumonia bilateral extensa | Feminino | 8 anos | Diamantina MG | Diamantina MG | Agosto | 2009 |
| 7 | 10 2010 | insuficiência respiratória aguda | Masculino | 6 anos | Diamantina MG | Cardoso, Diamantina MG | Maior | 2010 |
| 8 | 09 2010 | insuficiência respiratória aguda | Masculino | 1, 10 minutos (0 ano) | Diamantina MG | Nada Consta | Maior | 2010 |
| 9 | 21 2012 | falência múltipla de Órgãos e tecidos, leucemia | Feminino | 15 anos | Serro MG | Milho Verde, Serro MG | Setembro | 2012 |
| 10 | 06 2008 | insuficiência respiratória aguda, anemia crônica descompensada | Masculino | 2 anos | Diamantina MG | Couto de Magalhães MG | Junho | 2008 |
| 11 | 03 2008 | insuficiência respiratória, sepse neonatal, prematuridade | Feminino | 10 dias (0 ano) | Diamantina MG | Nada Consta | Abril | 2008 |
| 12 | 15 2011 | leucemia mieloide aguda, sequela neurológica por PCR | Feminino | 15 anos | Diamantina MG | Diamantina MG | Abril | 2011 |
| 13 | 02 2010 | insuficiência hepática, anemia falciforme, insuficiência renal aguda, CIVD | Feminino | 13 anos | Coluna MG | Coluna MG | Agosto | 2010 |
| 14 | 12 2009 | insuficiência respiratória sofrimento fetal agudo grave | Masculino | 4 dias (0 ano) | Diamantina MG | Nada Consta | Outubro | 2009 |
| 15 | 07 2011 | insuficiência respiratória, desnutrição grave | Feminino | 16 anos | Diamantina MG | Diamantina MG | Setembro | 2011 |
| 16 | 23 2012 | insuficiência respiratória, desnutrição grave, síndrome Cocayne | Feminino | 5 anos | Itamarandiba MG | Itamarandiba MG | Setembro | 2012 |
| 17 | 05 2010 | sofrimento fetal agudo, prolapso de cordão | Feminino | 0 ano | Diamantina MG | Nada Consta | Janeiro | 2010 |
| 18 | 08 2007 | falência orgânica múltipla, choque séptico, pneumonia de aspiração | Feminino | 14 anos | Minas Novas MG | Minas Novas MG | Setembro | 2007 |
| 19 | 06 2007 | sepse, falência múltiplos Órgãos | Masculino | 12 anos | Diamantina MG | Diamantina MG | Maior | 2007 |
| 20 | 09 2002 | caquexia neoplásica, câncer renal | Masculino | 7 anos | Diamantina MG | Inhá MG | Outubro | 2002 |
| 21 | 08 2005 | choque séptico, gangrena, injeção de anabolizante | Masculino | 14 anos | Diamantina MG | Diamantina MG | Agosto | 2005 |
| 22 | 05 2007 | broncopneumonia | Feminino | 1 ano | Diamantina MG | Diamantina MG | Abril | 2007 |

Na Tabela 17 é apresentada o *Cluster* 3 com 22 registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 15% dos 147 registros processados.

Gráficos:

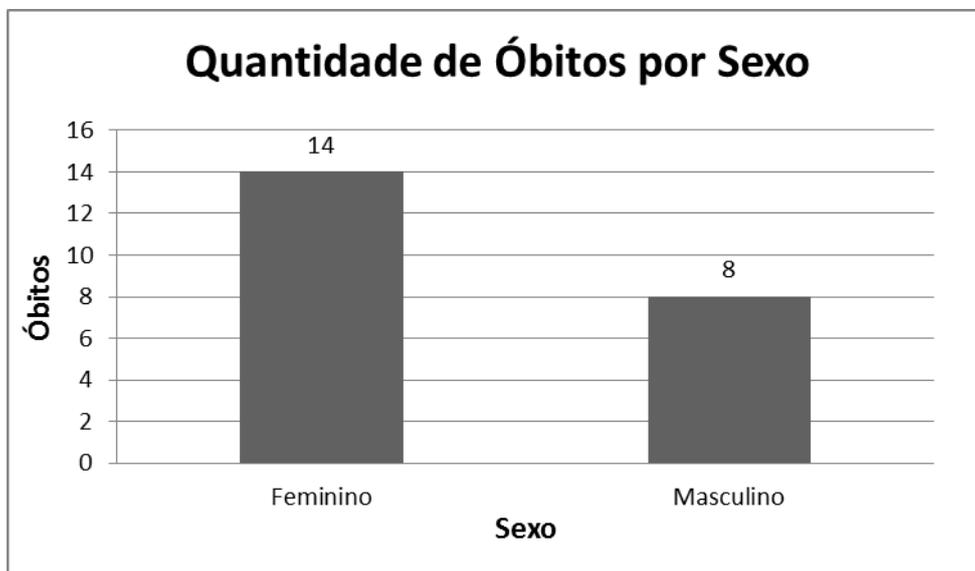


Figura 30. Quantidade de óbitos por sexo no *cluster* 03.

A Figura 30 mostra a quantidade de óbitos por sexo no *Cluster* 03 obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com 14 registros, sendo equivalente a 63% dos 22 registros agrupados no *Cluster* 02.

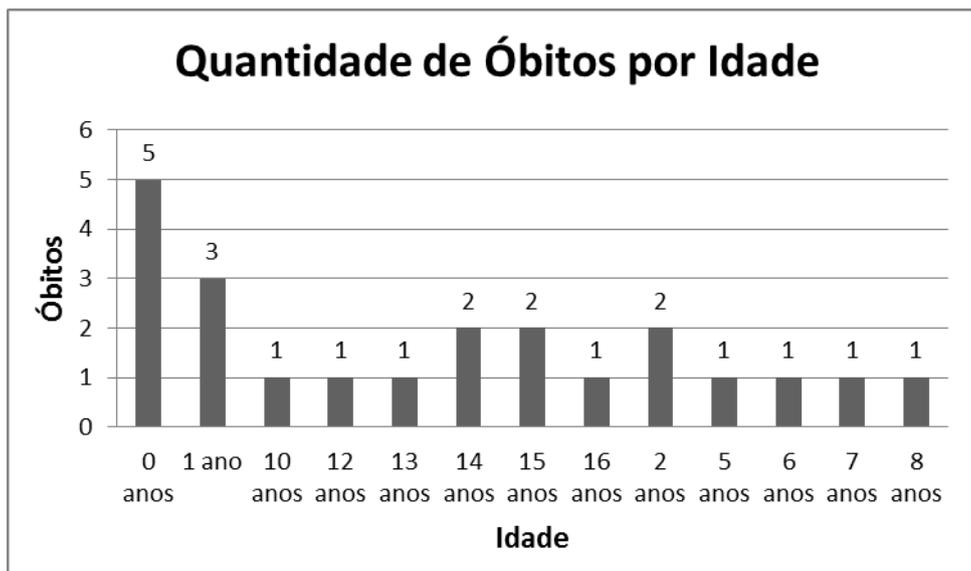


Figura 31. Quantidade de óbitos por idade no *cluster* 03.

A Figura 31 mostra a quantidade de óbitos por idade no *Cluster* 03 obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que ocorreram na faixa de idade abaixo de um ano com cinco registros, sendo equivalente a 22% dos 22 registros agrupados no *Cluster*.

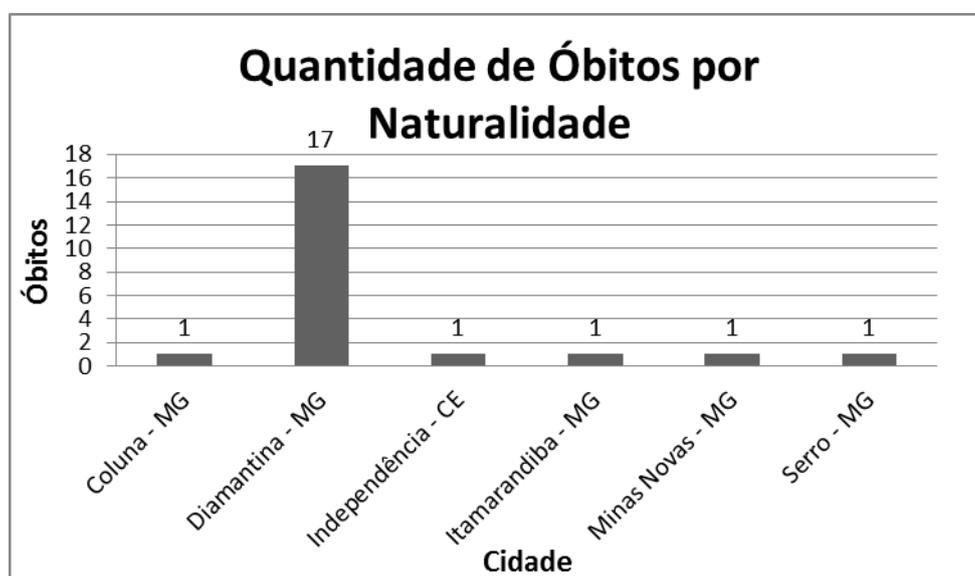


Figura 32. Quantidade de óbitos por naturalidade no *cluster* 03.

A Figura 32 mostra a quantidade de óbitos por naturalidade no *Cluster* 03, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância óbitos em crianças naturais de Diamantina-MG com 17 registros, sendo equivalente a 77% dos 22 registros.

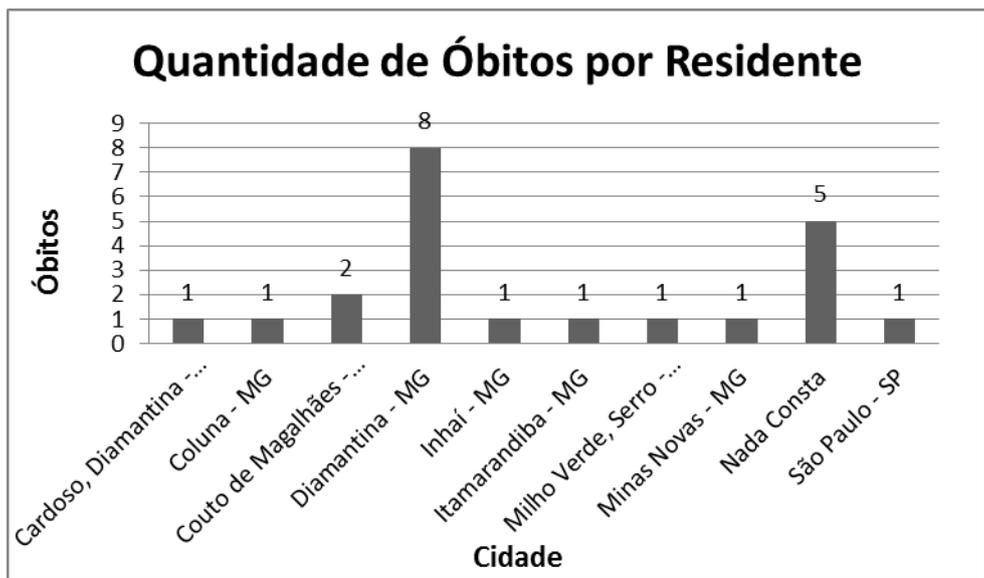


Figura 33. Quantidade de óbitos por residente no *cluster* 03.

A Figura 33 mostra a quantidade de óbitos por residente no *Cluster* 03, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em crianças que eram residentes em Diamantina-MG com oito registros, sendo equivalente a 36% dos 22 registros.

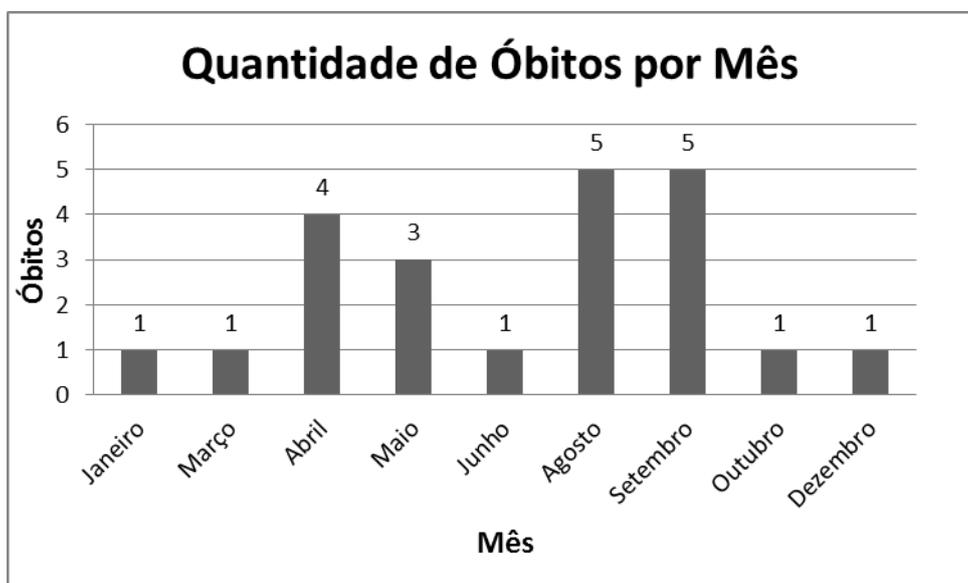


Figura 34. Quantidade de óbitos por mês no *cluster* 03.

A Figura 34 mostra a quantidade de óbitos por mês no *Cluster* 03, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram nos meses de agosto e novembro, ambos com cinco registros, sendo cada um, equivalentes a 22% dos 22 registros do *Cluster*.



Figura 35. Quantidade de óbitos por ano no *cluster* 03.

A Figura 35 mostra a quantidade de óbitos por ano no *Cluster* 03, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no ano de 2010 com quatro registros, sendo equivalente a 18% dos 22 registros do *Cluster*.

Cluster 04:

Número de Registros: 2

Tabela:

Tabela 18. *Cluster* 04 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|-----------------|---------------|---------------|-----------|-----------|
| 1 | 04 2006 | insuficiência respiratória, prematuridade | Masculino | 18 dias (0 ano) | Diamantina MG | Nada Consta | Maio | 2006 |
| 2 | 07 2008 | insuficiência respiratória, prematuridade | Feminino | 3 dias (0 ano) | Diamantina MG | Diamantina MG | Agosto | 2008 |

Na Tabela 18 é apresentada o *Cluster* 4 com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

Gráficos:

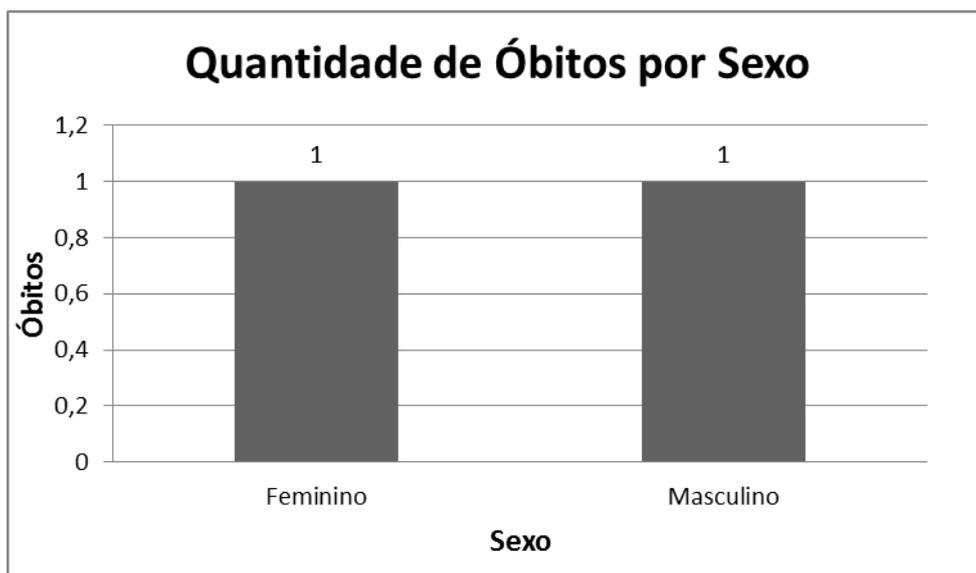


Figura 36. Quantidade de óbitos por sexo no *cluster* 04.

A Figura 36 mostra a quantidade de óbitos por sexo no *Cluster* 04, obtidos pelo modelo Cassiopeia. Nesta Figura há um óbito do sexo masculino e um óbito do sexo feminino, sendo os dois registros pertencentes ao *Cluster*.

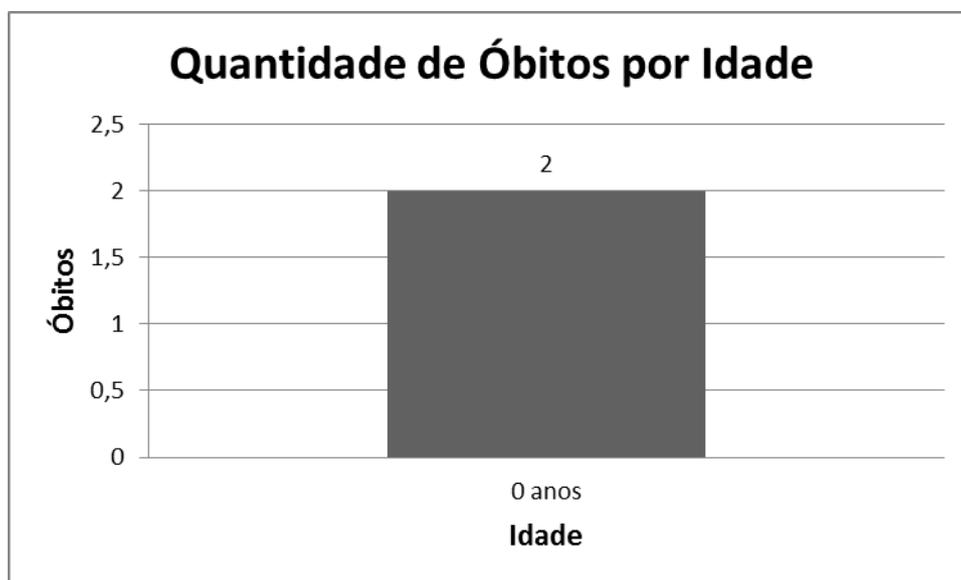


Figura 37. Quantidade de óbitos por idade no *cluster* 04.

A Figura 37 mostra a quantidade de óbitos por idade no *Cluster* 04 obtidos pelo modelo Cassiopeia. Nesta Figura, todos os óbitos ocorreram na faixa de idade abaixo de um ano, com dois registros agrupados no *Cluster*.

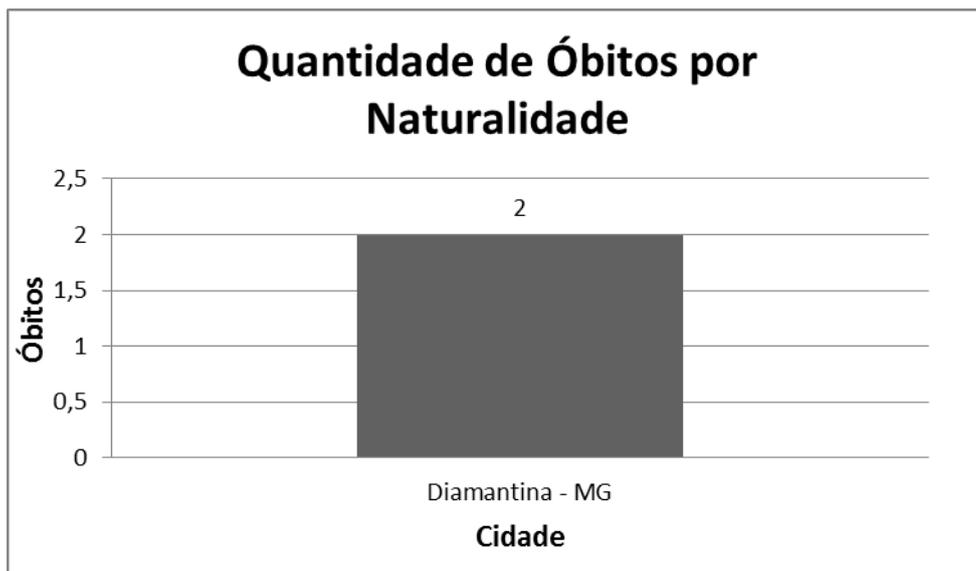


Figura 38. Quantidade de óbitos por naturalidade no *cluster* 04.

A Figura 38 mostra a quantidade de óbitos por naturalidade no *Cluster* 04, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças que faleceram eram naturais de Diamantina-MG, com dois registros agrupados no *Cluster*.

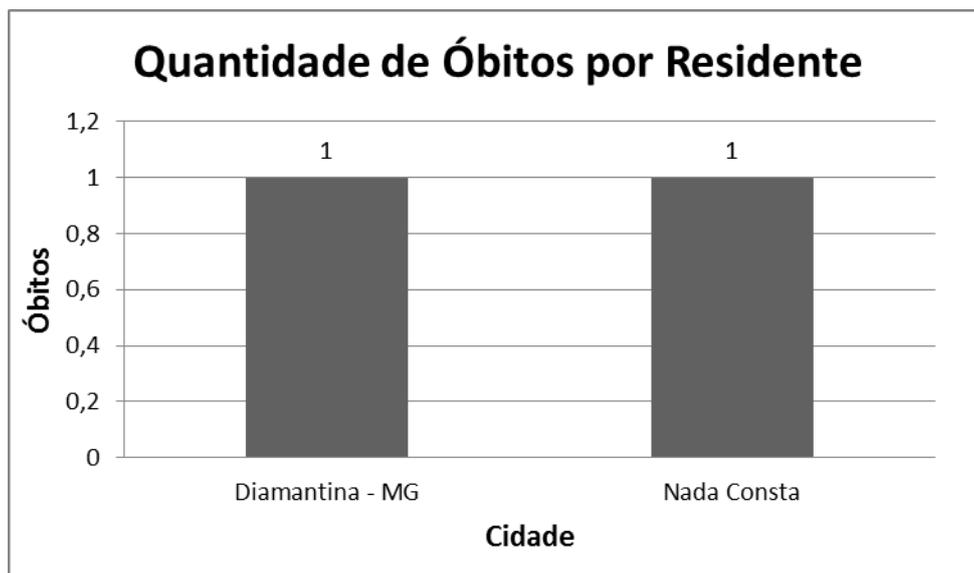


Figura 39. Quantidade de óbitos por residente no *cluster* 04.

A Figura 39 mostra a quantidade de óbitos por residente no *Cluster* 04, obtidos pelo modelo Cassiopeia. Uma das crianças era residente de Diamantina-MG, e a outra não constava o local em que a criança residia.

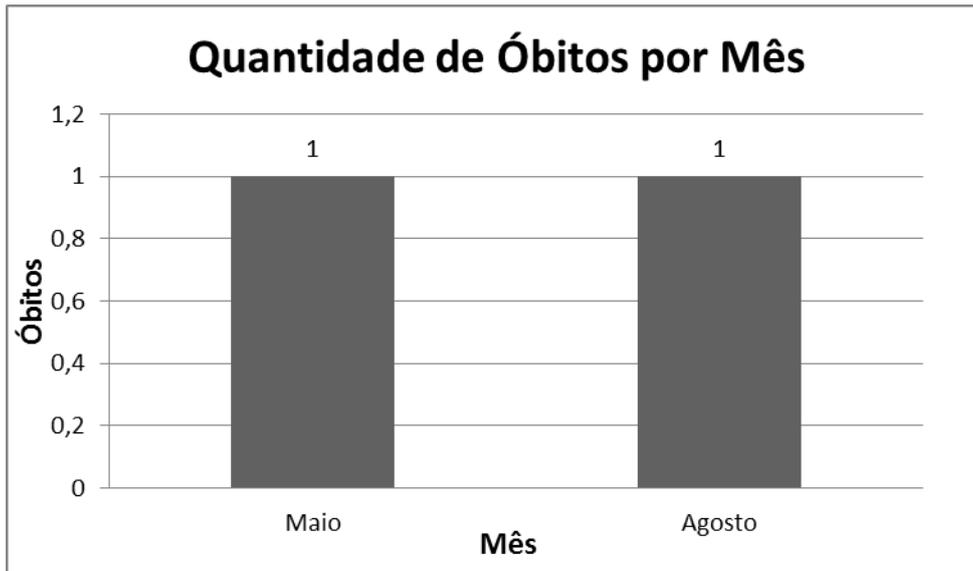


Figura 40. Quantidade de óbitos por mês no *cluster* 04.

A Figura 40 mostra a quantidade de óbitos por mês no *Cluster* 04, obtidos pelo modelo Cassiopeia. Nesta Figura um óbito ocorreu no mês de maio e o outro ocorreu no mês de agosto.

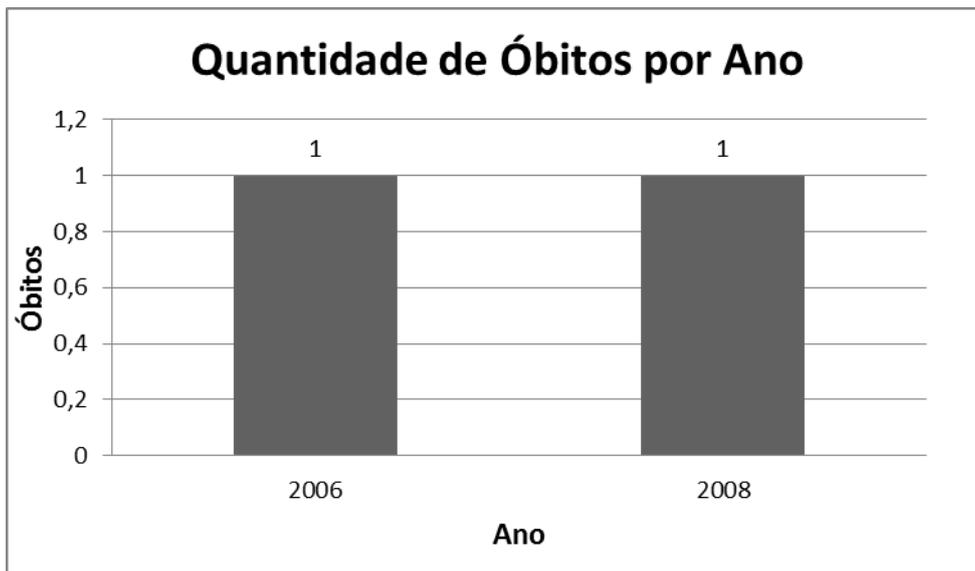


Figura 41. Quantidade de óbitos por ano no *cluster* 04.

A Figura 41 mostra a quantidade de óbitos por ano no *Cluster* 04, obtidos pelo modelo Cassiopeia. Nesta Figura um óbito ocorreu no ano de 2006 e o outro ocorreu no ano de 2008.

Cluster 05:

Número de Registros: 7

Tabela:

Tabela 19. Cluster 05 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|--------------------|---|----------------------------|-----------|-----------|
| 1 | 05 2009 | insuficiência respiratória, sepse, anemia | Feminino | 1 dia (0 ano) | Gouveia MG | Nada Consta | Fevereiro | 2009 |
| 2 | 03 2011 | choque séptico, sepse neonatal precoce | Feminino | 5 dias (0 ano) | Diamantina MG | Datas MG | Julho | 2011 |
| 3 | 03 2006 | sepse neonatal tardia | Masculino | 20 dias (0 ano) | Couto de Magalhães de Minas MG | Nada Consta | Março | 2006 |
| 4 | 06 2005 | choque séptico | Masculino | 14 anos | Diamantina MG | Diamantina MG | Julho | 2005 |
| 5 | 32 2012 | choque não especificado | Masculino | 9 dias (0 ano) | Serro MG | Serro MG | Novembro | 2012 |
| 6 | 28 2012 | choque séptico | Feminino | 0 ano | Curvelo MG | Curvelo MG | Outubro | 2012 |
| 7 | 20 2012 | choque séptico, broncoespasmo infectado | Masculino | 8 anos | Serro MG | Alvorada de Minas MG | Agosto | 2012 |

Na Tabela 19 é apresentado o *Cluster 5*, com sete registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

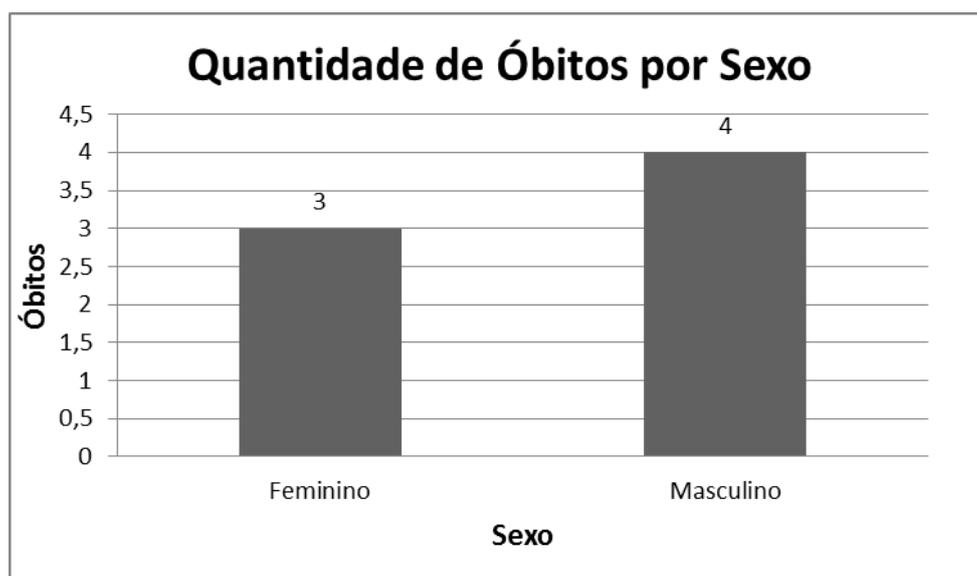


Figura 42. Quantidade de óbitos por sexo no *cluster* 05.

A Figura 42 mostra a quantidade de óbitos por sexo no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos masculinos, com quatro registros, sendo equivalente a 57% dos sete registros agrupados no *Cluster*.

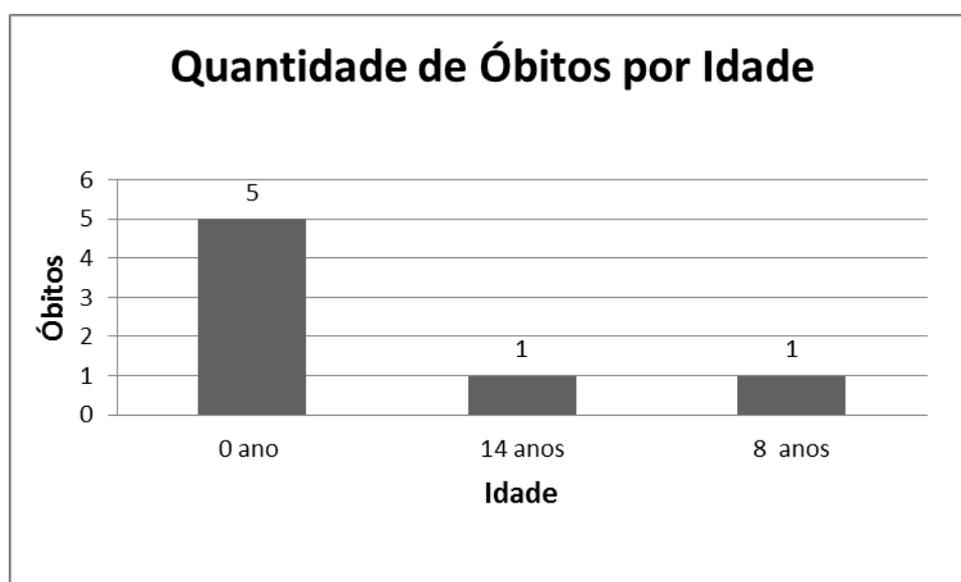


Figura 43. Quantidade de óbitos por idade no *cluster* 05.

A Figura 43 mostra a quantidade de óbitos por idade no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano, com cinco registros, sendo equivalente a 71% dos sete registros que estão agrupados no *Cluster*.

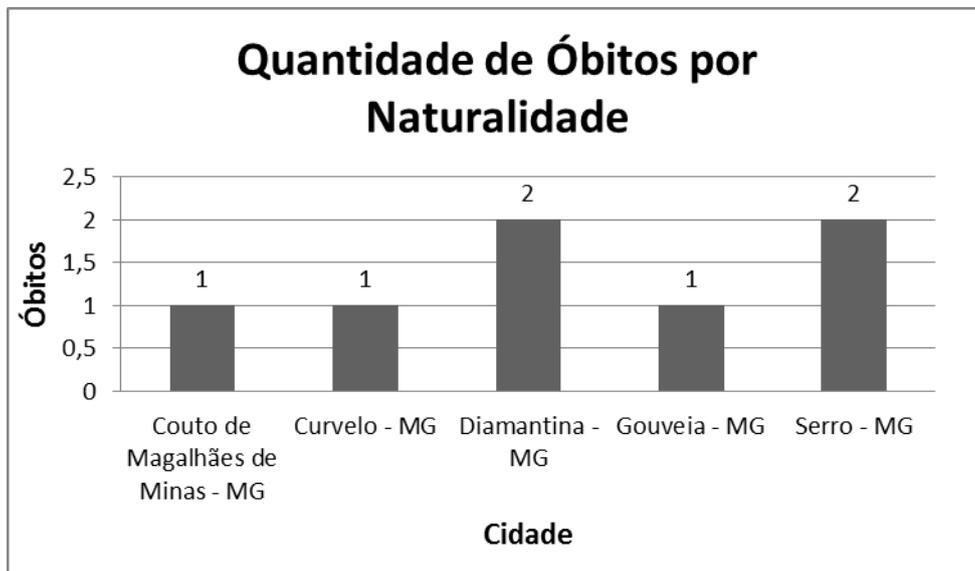


Figura 44. Quantidade de óbitos por naturalidade no *cluster* 05.

A Figura 44 mostra a quantidade de óbitos por naturalidade no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância óbitos em crianças naturais de Diamantina-MG e Serro-MG, ambos com dois registros, sendo cada um equivalente a 28% dos sete registros do *Cluster*.

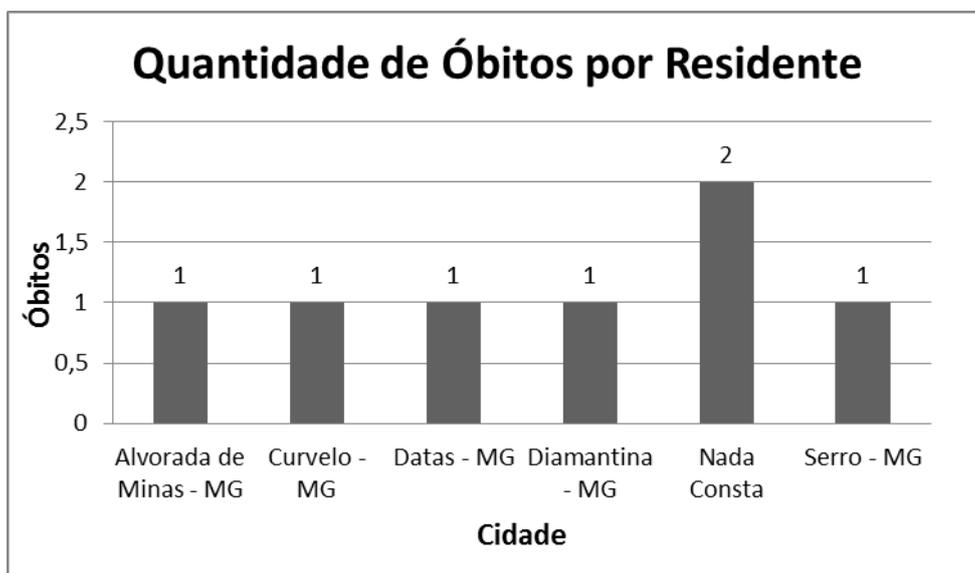


Figura 45. Quantidade de óbitos por residente no *cluster* 05.

A Figura 45 mostra a quantidade de óbitos por residente no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos

que não possuem a informação de onde a criança morava, com dois registros, sendo equivalente a 28% dos sete registros do *Cluster*.

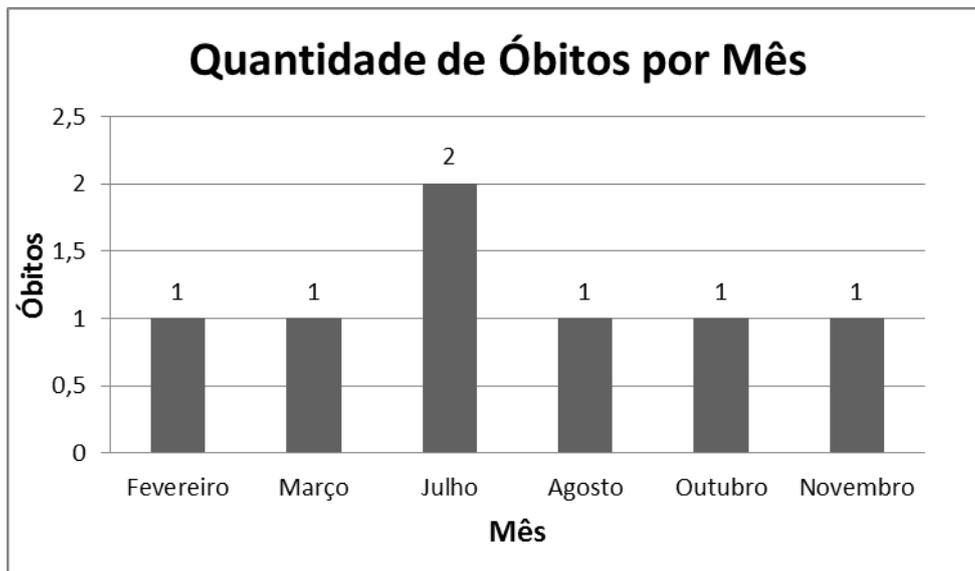


Figura 46. Quantidade de óbitos por mês no *cluster* 05.

A Figura 46 mostra a quantidade de óbitos por mês no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de julho, com dois registros, sendo equivalente a 28% dos sete registros do *Cluster*.

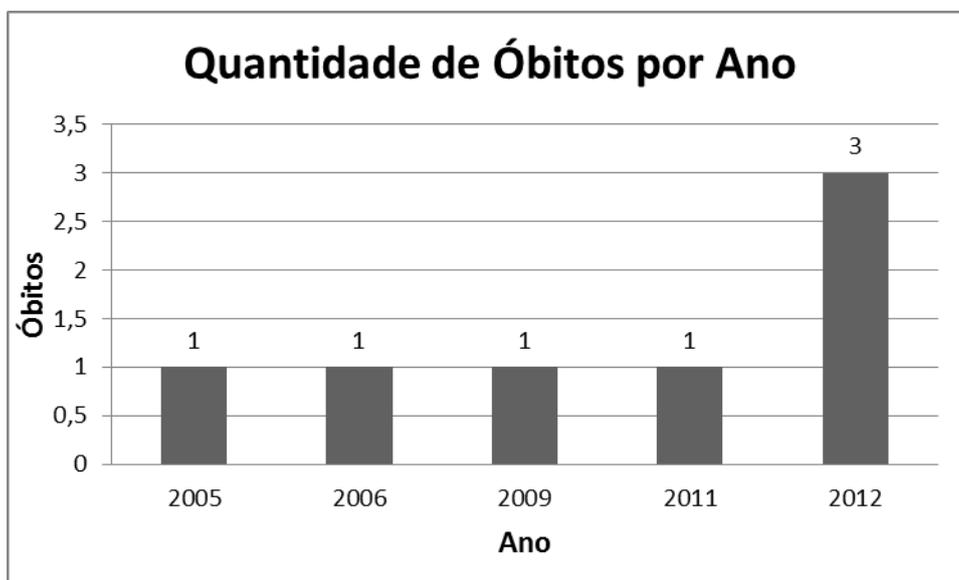


Figura 47. Quantidade de óbitos por ano no *cluster* 05.

A Figura 47 mostra a quantidade de óbitos por ano no *Cluster* 05, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que

ocorreram no ano de 2012 com três registros, sendo equivalente a 42% dos sete registros do *Cluster*.

Cluster 06:

Número de registros: 6

Tabelas:

Tabela 20. *Cluster* 06 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|--------------------|------------------|------------------|-----------|-----------|
| 1 | 02 2011 | insuficiência respiratória aguda, cardiopatia congênita | Masculino | 1 dia (0 ano) | Diamantina MG | Diamantina MG | Julho | 2011 |
| 2 | 10 2012 | insuficiência respiratória aguda, cardiopatia congênita, pneumonia bacteriana, má formação SNC | Masculino | 7 meses (0 ano) | Serro MG | Serro MG | Junho | 2012 |
| 3 | 13 2012 | insuficiência | Feminino | 9 meses (0 ano) | Capelinha MG | Capelinha MG | Julho | 2012 |
| 4 | 24 2012 | insuficiência respiratória aguda, pneumonia | Feminino | 1 ano | Capelinha MG | Capelinha MG | Setembro | 2012 |
| 5 | 12 2007 | pneumonia, má formação SNC | Feminino | 0 ano | Gouveia MG | Gouveia MG | Fevereiro | 2007 |
| 6 | 01 2009 | peritonite, má formação trato digestiva | Feminino | 2 meses (0 ano) | Turmalina MG | Nada Consta | Janeiro | 2009 |

Na Tabela 20 é apresentada o *Cluster* 6 com seis registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

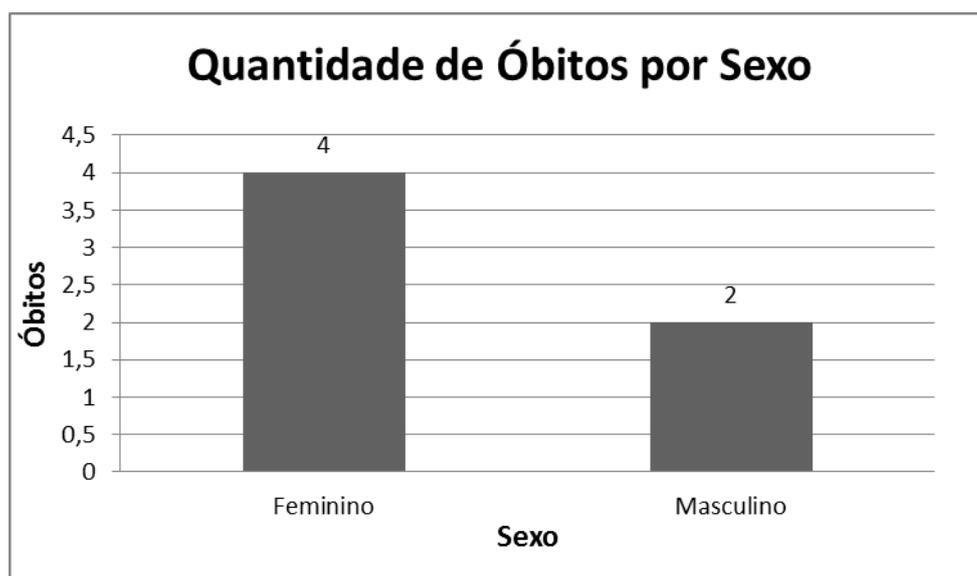


Figura 48. Quantidade de óbitos por sexo no *cluster* 06.

A Figura 48 mostra a quantidade de óbitos por sexo no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com quatro registros, sendo equivalente a 66% dos seis registros agrupados no *Cluster*.

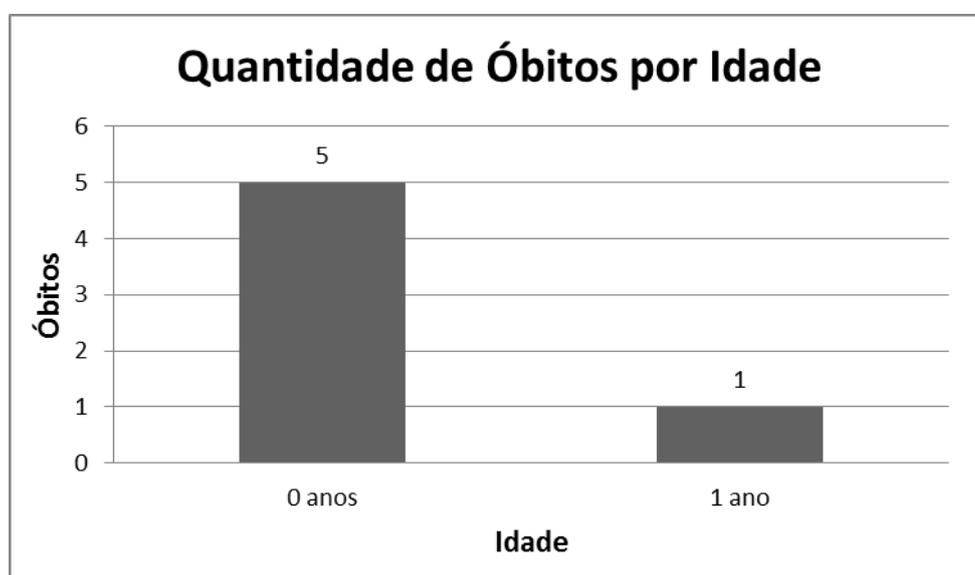


Figura 49. Quantidade de óbitos por idade no *cluster* 06.

A Figura 49 mostra a quantidade de óbitos por idade no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano com cinco registros, sendo equivalente a 83% dos seis registros agrupados no *Cluster*.

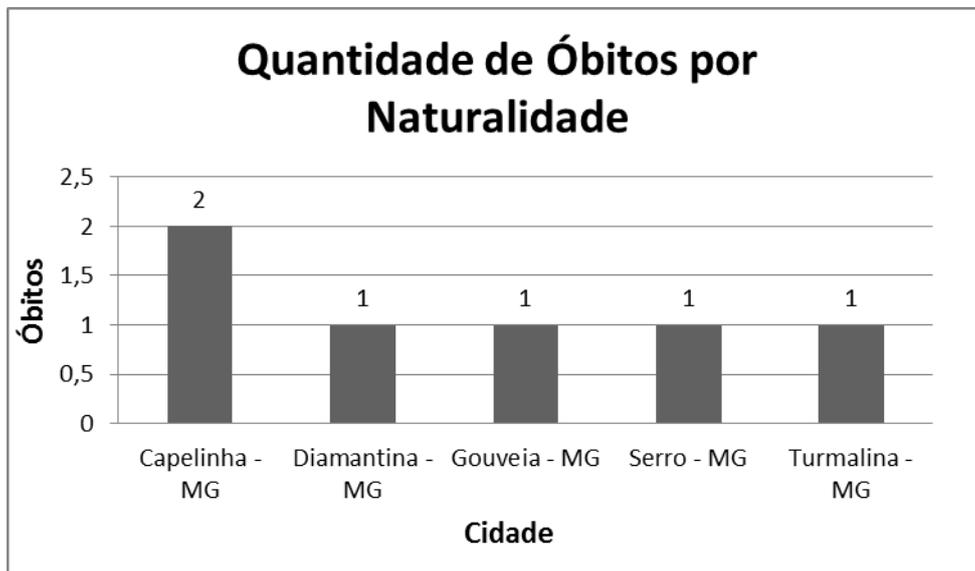


Figura 50. Quantidade de óbitos por naturalidade no *cluster* 06.

A Figura 50 mostra a quantidade de óbitos por naturalidade no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em crianças que são naturais de Capelinha-MG, com dois registros, sendo equivalente a 33% dos seis registros do *Cluster*.

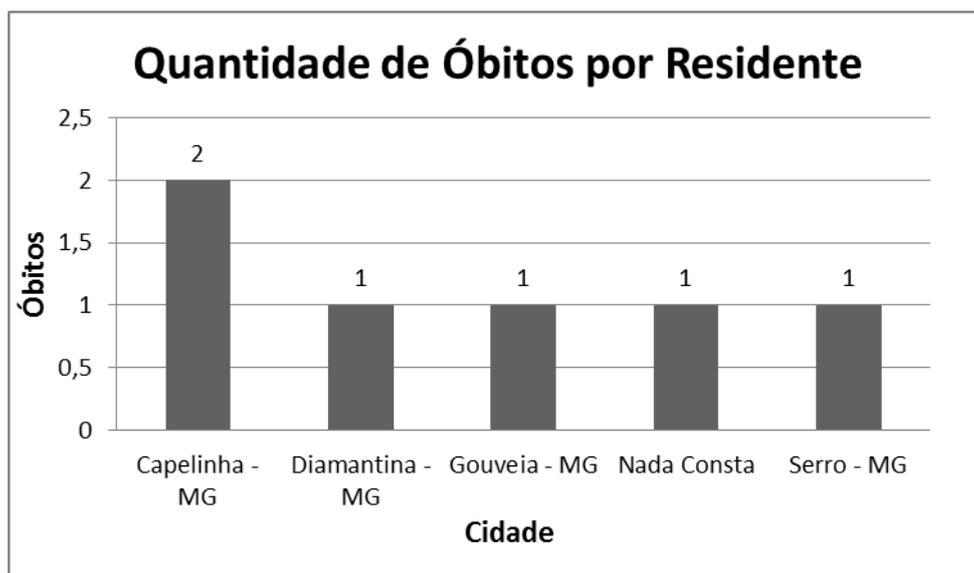


Figura 51. Quantidade de óbitos por residente no *cluster* 06.

A Figura 51 mostra a quantidade de óbitos por residente no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças

falecidas que eram residentes em Capelinha-MG, com dois registros, sendo equivalente a 33% dos seis registros do *Cluster*.

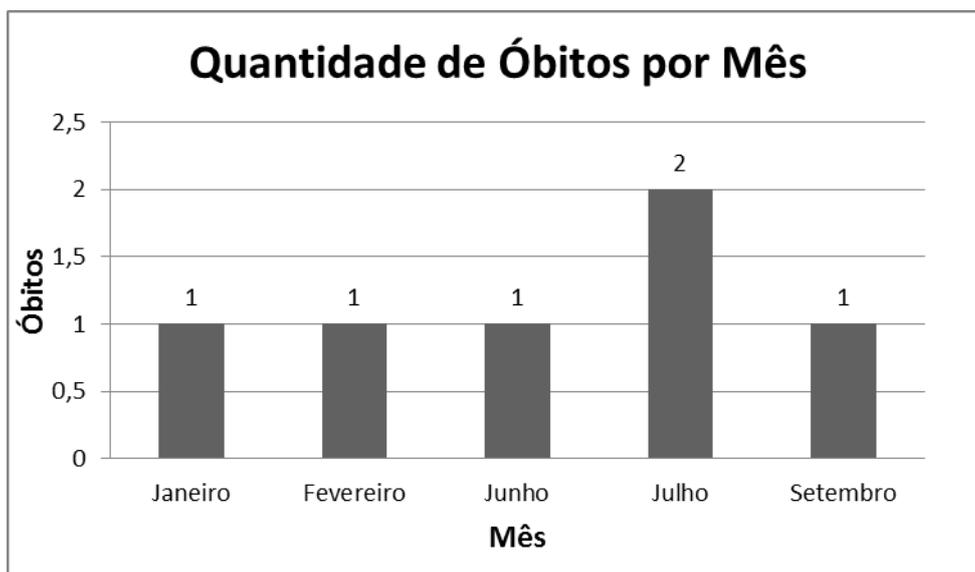


Figura 52. Quantidade de óbitos por mês no *cluster* 06.

A Figura 52 mostra a quantidade de óbitos por mês no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de julho, com dois registros, sendo equivalente a 33% dos sete registros do *Cluster*.

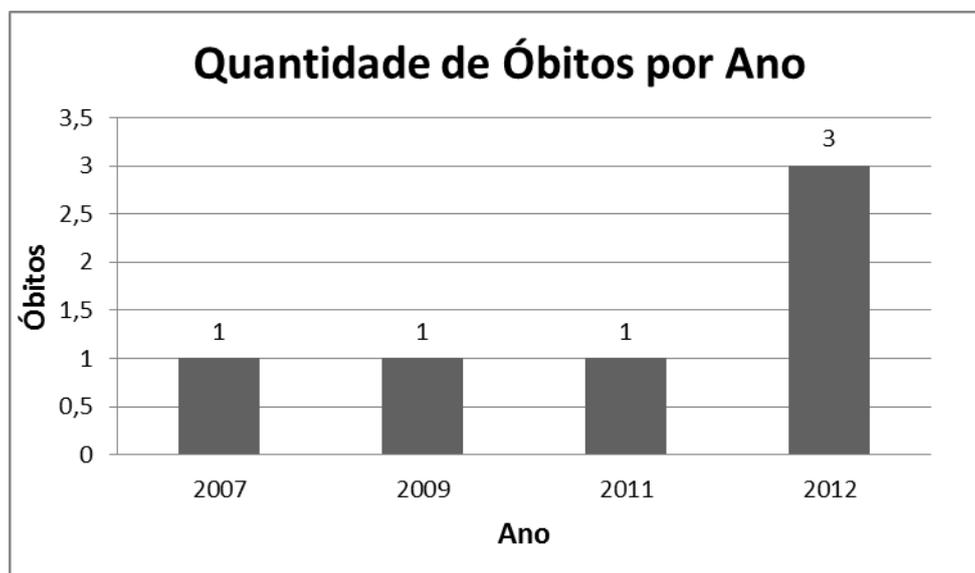


Figura 53. Quantidade de óbitos por ano no *cluster* 06.

A Figura 53 mostra a quantidade de óbitos por ano no *Cluster* 06, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que

ocorreram no ano de 2012 com três registros, sendo equivalente a 50% dos seis registros do *Cluster*.

Cluster 07:

Número de Registros: 14

Tabela:

Tabela 21. Cluster 07 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|------------------|-------------------|-------------------------|-----------|-----------|
| 1 | 01 2007 | pneumonia, má formação SNC | Feminino | 4 meses (0 ano) | Belo Horizonte MG | Carbonita MG | Fevereiro | 2007 |
| 2 | 02 2004 | insuficiência respiratória, pneumonia aguda | Masculino | 15 anos | Diamantina MG | Diamantina MG | Maio | 2004 |
| 3 | 04 2012 | bronco pneumonia aspirativa, insuficiência cardíaca, cardiopatia congênita | Feminino | 3 anos | Curvelo MG | Gouveia MG | Março | 2012 |
| 4 | 07 2009 | AVC hemorrágico, má formação arterio venosa | Feminino | 15 anos | Curvelo MG | Gouveia MG | Julho | 2009 |
| 5 | 08 2010 | insuficiência cardiorrespiratória, cardiopatia congênita | Masculino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Março | 2010 |
| 6 | 05 2004 | pneumonia aspirativa | Masculino | 3 meses (0 ano) | Diamantina MG | Nada Consta | Agosto | 2004 |
| 7 | 09 2007 | parada cardiorrespiratória, insuficiência cardíaca congestiva, cardiopatia grave | Masculino | 11 anos | Diamantina MG | Diamantina MG | Setembro | 2007 |
| 8 | 07 2002 | cardiopatia congênita completa | Feminino | 3 dias (0 ano) | Diamantina MG | Nada Consta | Julho | 2002 |
| 9 | 13 2009 | tromboembolismo pulmonar, trombose venosa profunda | Masculino | 14 anos | Itamarandiba MG | Carbonita MG | Outubro | 2009 |
| 10 | 01 2010 | insuficiência respiratória aguda, pneumonia bacteriana aguda, paralisia cerebral não especificada | Feminino | 5 anos | Diamantina MG | Pinheiro, Diamantina MG | Junho | 2010 |
| 11 | 17 2012 | insuficiência renal aguda, choque hipotérmico, pneumonia, paralisia cerebral | Masculino | 3 anos | Serro MG | Alvorada de Minas MG | Julho | 2012 |
| 12 | 19 2012 | insuficiência respiratória aguda, choque séptico, pneumonia, crises convulsivas | Masculino | 2 meses (0 ano) | Capelinha MG | Capelinha MG | Agosto | 2012 |
| 13 | 08 2011 | insuficiência respiratória, pneumonia aspirativa, paralisia cerebral, síndrome de West | Feminino | 6 anos | Rio Vermelho MG | Serro MG | Setembro | 2011 |
| 14 | 14 2009 | choque séptico, meningite bacteriana, hidrocefalia aguda | Feminino | 10 meses (0 ano) | Sete Lagoas MG | Corinto MG | Outubro | 2009 |

Na Tabela 21 é apresentada o *Cluster 7* com 14 registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 9% dos 147 registros processados.

Gráficos:

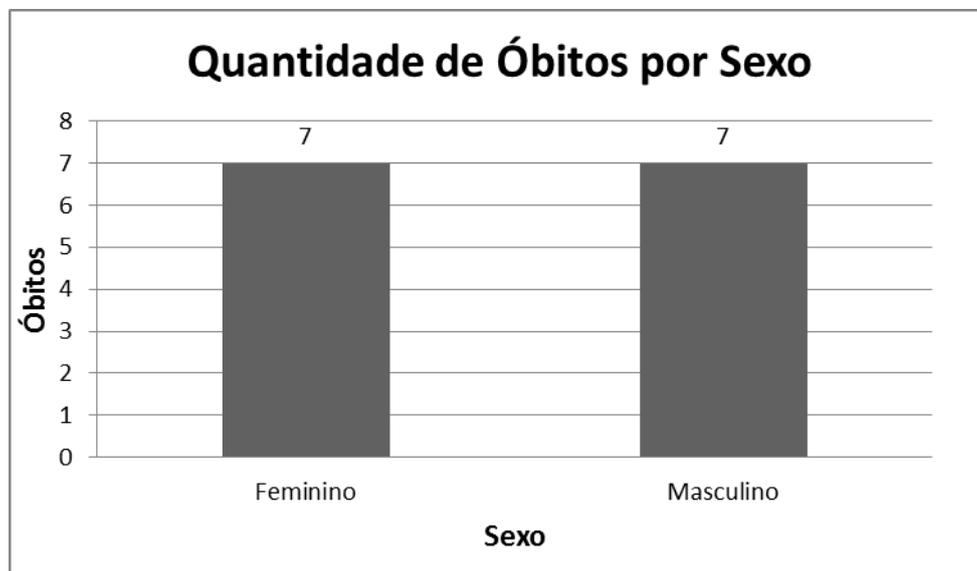


Figura 54. Quantidade de óbitos por sexo no *cluster 07*.

A Figura 54 mostra a quantidade de óbitos por sexo no *Cluster 07*, obtidos pelo modelo Cassiopeia. Nesta Figura ambos os sexos possuem sete registros de óbito, sendo cada um, equivalente a 50% dos 14 registros agrupados no *Cluster*.

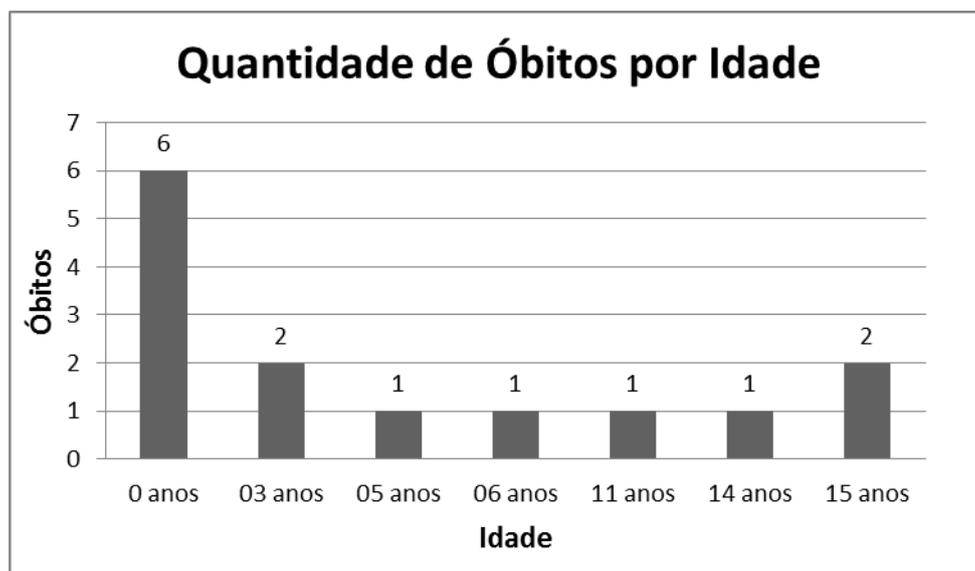


Figura 55. Quantidade de óbitos por idade no *cluster 07*.

A Figura 55 mostra a quantidade de óbitos por idade no *Cluster 07*, obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano com seis registros, sendo equivalente a 42% dos 14 registros agrupados no *Cluster*.

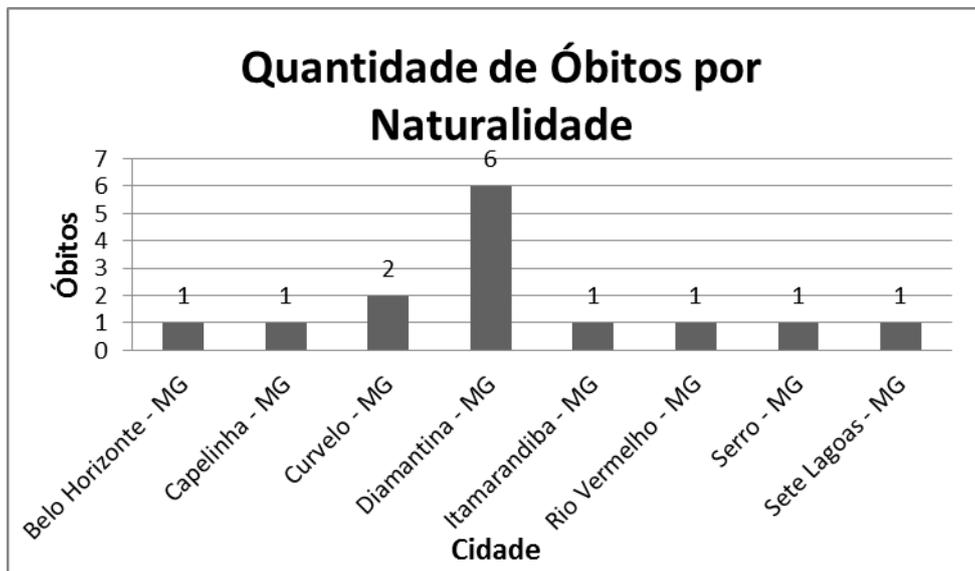


Figura 56. Quantidade de óbitos por naturalidade no *cluster* 07.

A Figura 56 mostra a quantidade de óbitos por naturalidade no *Cluster* 07, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em crianças naturais de Diamantina-MG, com seis registros, sendo equivalente a 42% dos 14 registros.

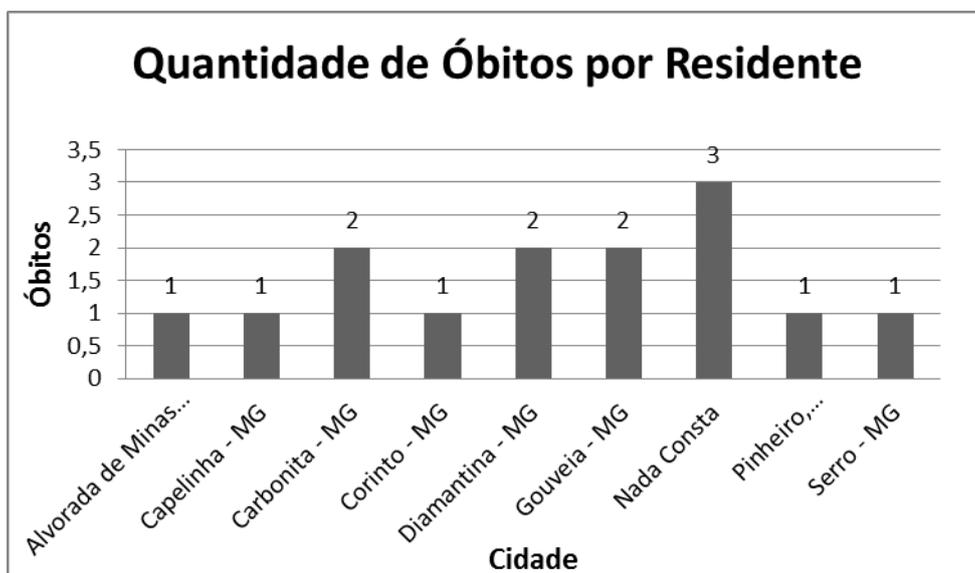


Figura 57. Quantidade de óbitos por residente no *cluster* 07.

A Figura 57 mostra a quantidade de óbitos por residente no *Cluster* 07, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos

que não possuem a informação de onde a criança residia, com três registros, sendo equivalente a 21% dos 14 registros do *Cluster*.

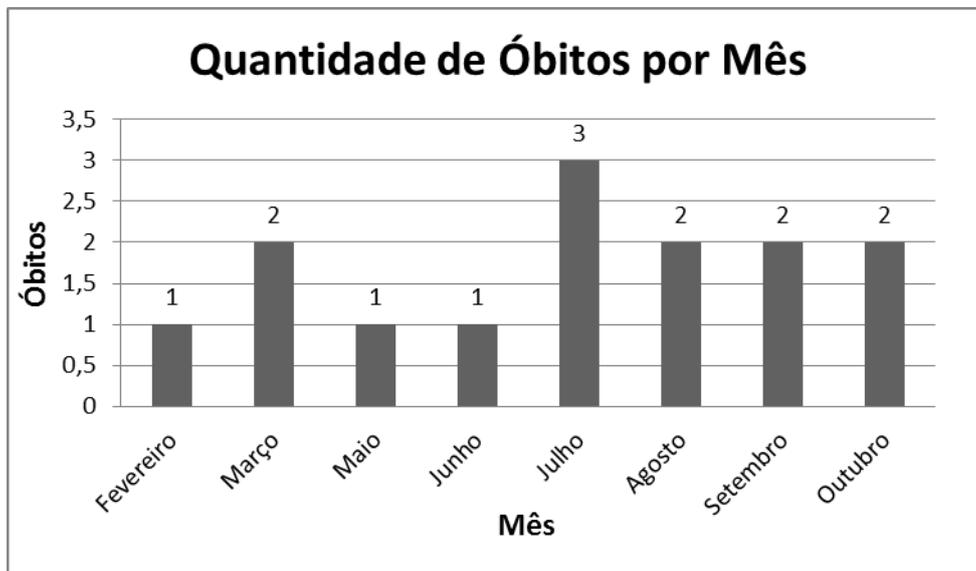


Figura 58. Quantidade de óbitos por mês no *cluster* 07.

A Figura 58 mostra a quantidade de óbitos por mês no *Cluster* 07, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de julho, com três registros, sendo equivalente a 21% dos 14 registros do *Cluster*.

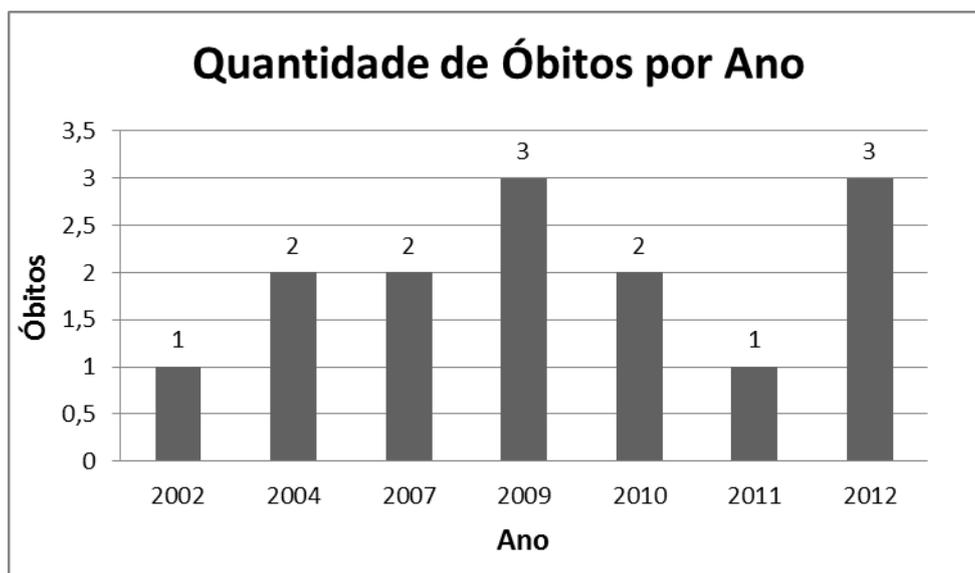


Figura 59. Quantidade de óbitos por ano no *cluster* 07.

A Figura 59 mostra a quantidade de óbitos por ano no *Cluster* 07, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que

ocorreram nos anos de 2009 e 2012, ambos com três registros, sendo cada um, equivalente a 21% dos 14 registros.

Cluster 08:

Número de Registros: 01

Tabela:

Tabela 22. Cluster 08 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|-----------------------------------|----------|-----------|----------------------|------------|-----------|-----------|
| 1 | 02 2003 | traumatismo crânio- encefálico | Feminino | 6 anos | Rio de Janeiro RJ | Mauá SP | Dezembro | 2003 |

Na Tabela 22 é apresentada o *Cluster 8*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíam relevância para pesquisa, pelo fato de só haver um registro no agrupamento.

Gráficos:

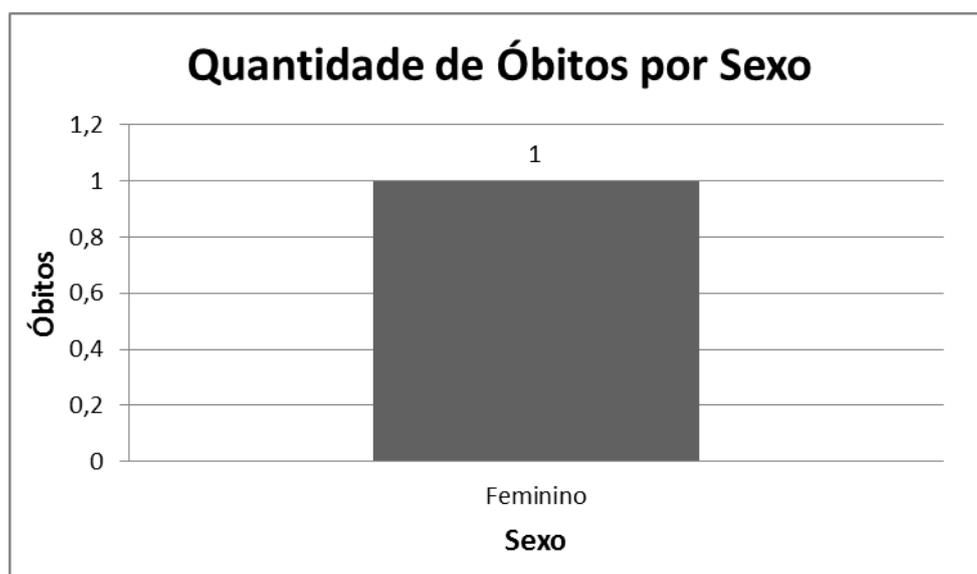


Figura 60. Quantidade de óbitos por sexo no *cluster* 08.

A Figura 60 mostra a quantidade de óbitos por sexo no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, sendo do sexo feminino.

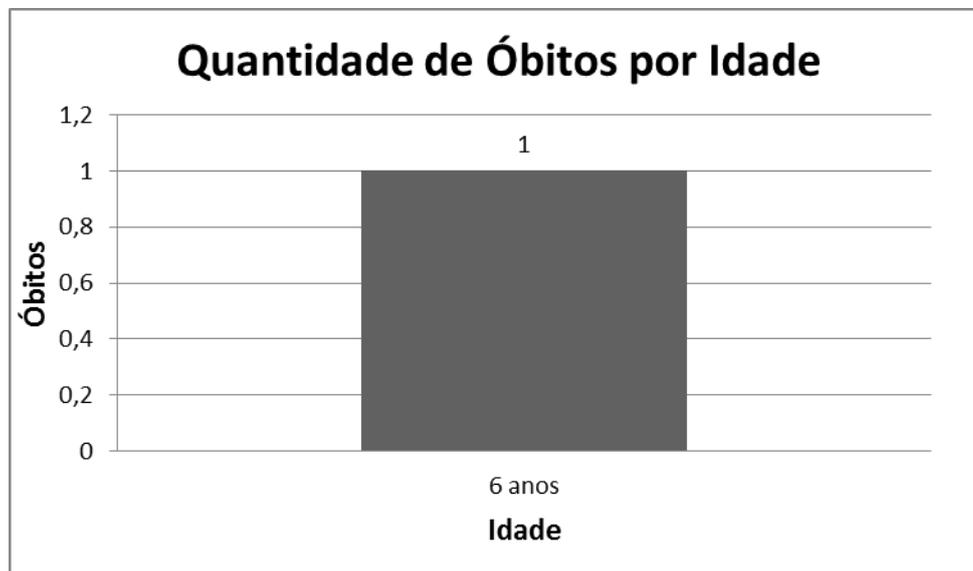


Figura 61. Quantidade de óbitos por idade no *cluster* 08.

A Figura 61 mostra a quantidade de óbitos por idade no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, que possui seis anos de idade.



Figura 62. Quantidade de óbitos por naturalidade no *cluster* 08.

A Figura 62 mostra a quantidade de óbitos por naturalidade no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, sendo natural do Rio de Janeiro-RJ.

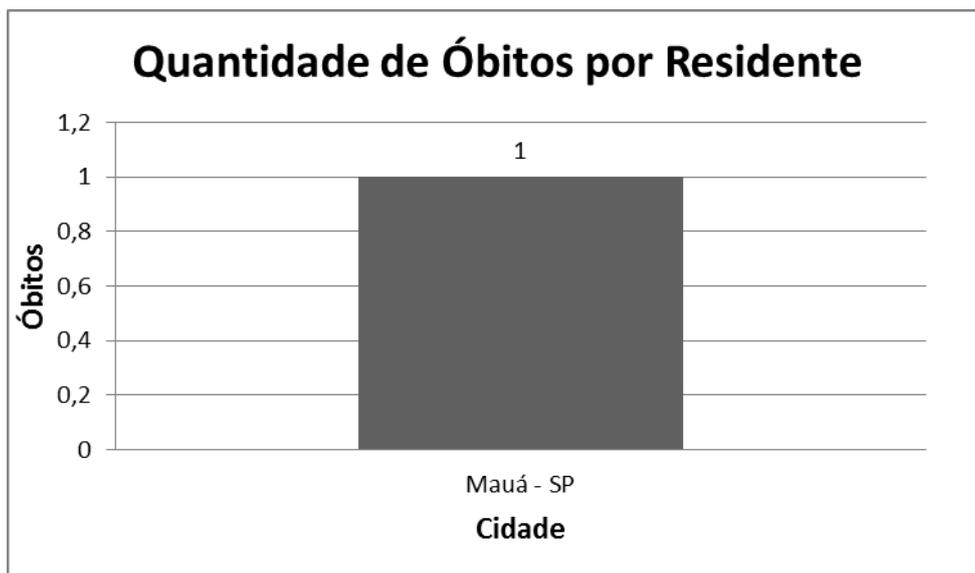


Figura 63. Quantidade de óbitos por residente no *cluster* 08.

A Figura 63 mostra a quantidade de óbitos por residente no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, sendo residente de Mauá-SP.

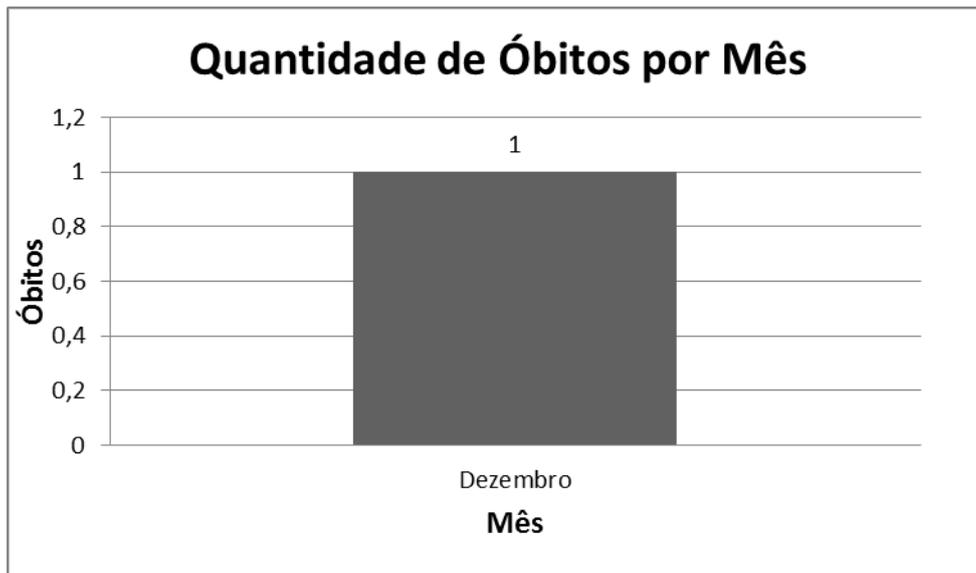


Figura 64. Quantidade de óbitos por mês no *cluster* 08.

A Figura 64 mostra a quantidade de óbitos por mês no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, que faleceu no mês de dezembro.

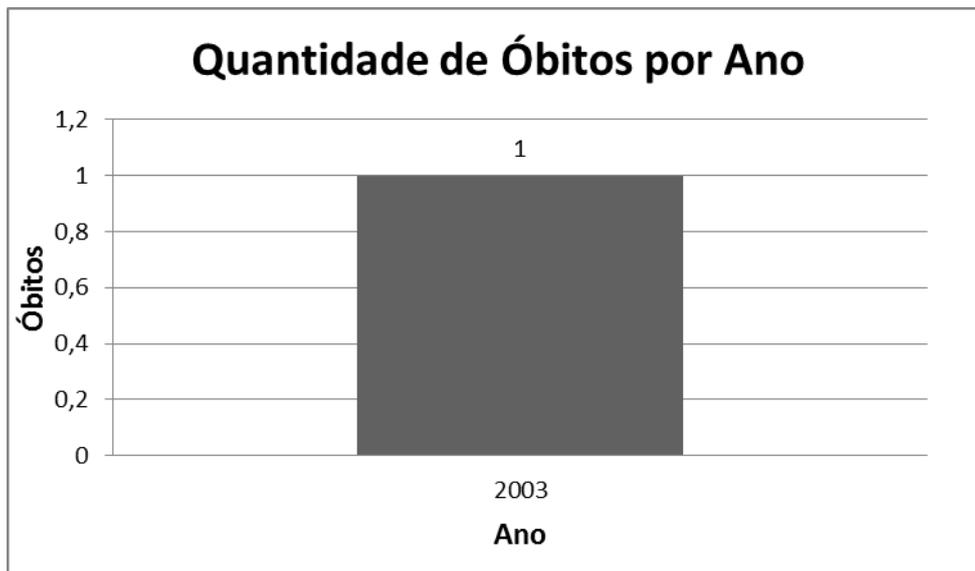


Figura 65. Quantidade de óbitos por ano no *cluster* 08.

A Figura 65 mostra a quantidade de óbitos por ano no *Cluster* 08, obtidos pelo modelo Cassiopeia. Nesta Figura é apresentado o único óbito do *Cluster*, que faleceu no ano de 2003.

Cluster 09:

Número de Registros: 11

Tabela:

Tabela 23. Cluster 09 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|---------------------|------------------------|-----------------|-----------|-----------|
| 1 | 25 2012 | traumatismo crânio encefálico | Masculino | 16 anos | Guanhães MG | Guanhães MG | Outubro | 2012 |
| 2 | 11 2002 | traumatismo crânio encefálico | Feminino | 9 anos | Turmalina MG | Pitangueiras SP | Dezembro | 2002 |
| 3 | 06 2002 | parada cardio respiratória, traumatismo crânio encefálico | Masculino | 16 anos | Diamantina MG | Diamantina MG | Junho | 2002 |
| 4 | 03 2003 | traumatismo crânio-encefálico contuso | Feminino | 8 anos | Santo Amaro SP | Santo Amaro SP | Dezembro | 2003 |
| 5 | 04 2005 | insuficiência cardíaca, respiratória, septicemia, hidrocefalia | Feminino | 2, 20 horas (0 ano) | Diamantina MG | Nada Consta | Maio | 2005 |
| 6 | 02 2008 | insuficiência respiratória, hidrocefalia infectada | Masculino | 3 meses (0 ano) | Belo Horizonte MG | Itamarandiba MG | Abril | 2008 |
| 7 | 03 2005 | traumatismo cranio encefálico, acidente automobilístico | Feminino | 5 meses (0 ano) | Diamantina MG | Nada Consta | Maio | 2005 |
| 8 | 09 2009 | traumatismo craniano encefálico, projétil de arma de fogo | Feminino | 16 anos | Serra Azul de Minas MG | Diamantina MG | Agosto | 2009 |
| 9 | 11 2012 | traumatismo craniano, acidente automobilístico | Feminino | 14 anos | Guanhães MG | Guanhães MG | Junho | 2012 |
| 10 | 10 2011 | acidente vascular encefálico hemorrágico | Feminino | 14 anos | Turmalina MG | Turmalina MG | Novembro | 2011 |
| 11 | 04 2009 | estado de mal epiléptico, tumor cerebral | Feminino | 4 anos | Capelinha MG | Capelinha MG | Janeiro | 2009 |

Na Tabela 23 é apresentada o Cluster 9 com 11 registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 7% dos 147 registros processados.

Gráficos:

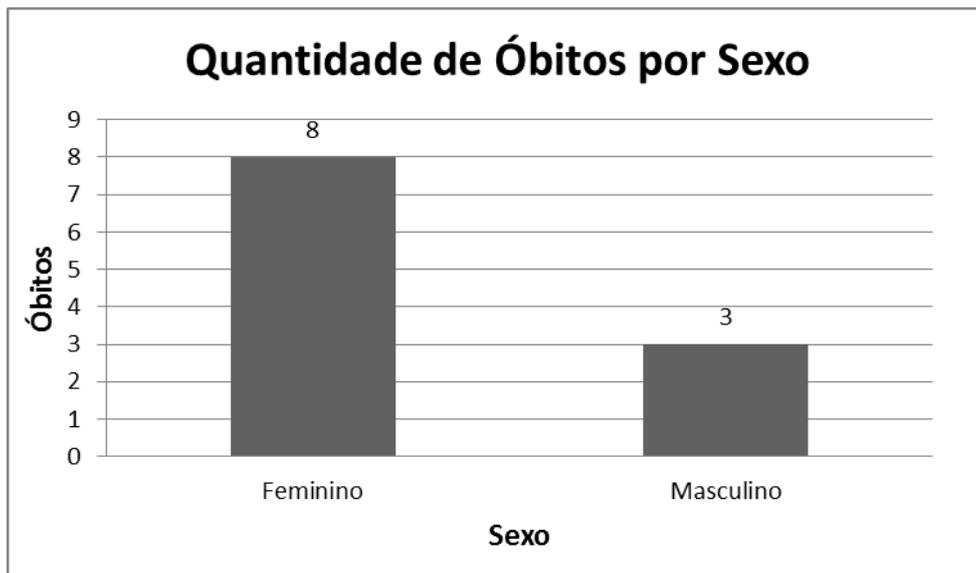


Figura 66. Quantidade de óbitos por sexo no *cluster* 09.

A Figura 66 mostra a quantidade de óbitos por sexo no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com oito registros, sendo equivalente a 72% dos 11 registros agrupados no *Cluster*.

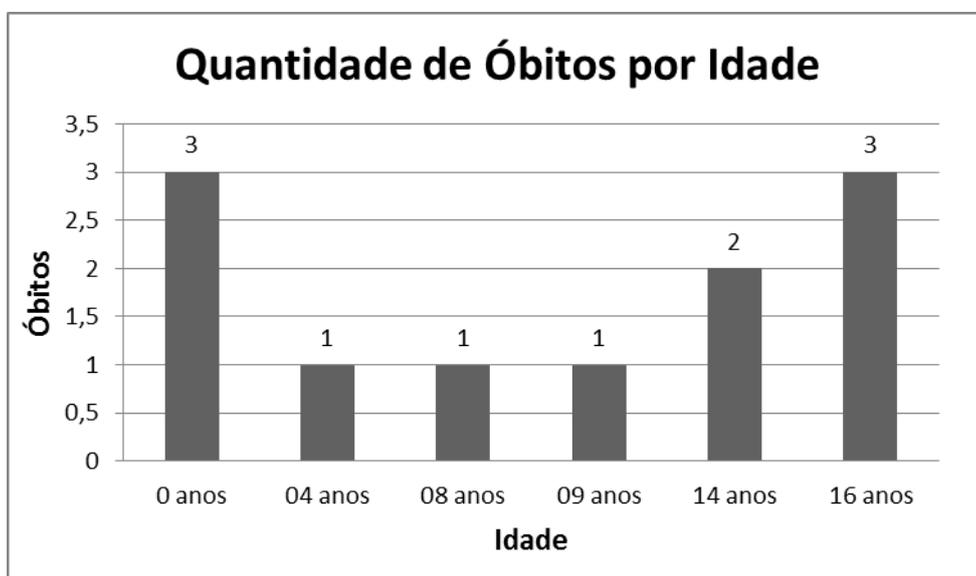


Figura 67. Quantidade de óbitos por idade no *cluster* 09.

A Figura 67 mostra a quantidade de óbitos por idade no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura, é possível perceber a predominância de óbitos que

ocorreram na faixa de idade abaixo de um ano e com 16 anos, ambos com três registros, sendo cada um, equivalente a 27% dos 11 registros agrupados no *Cluster*.

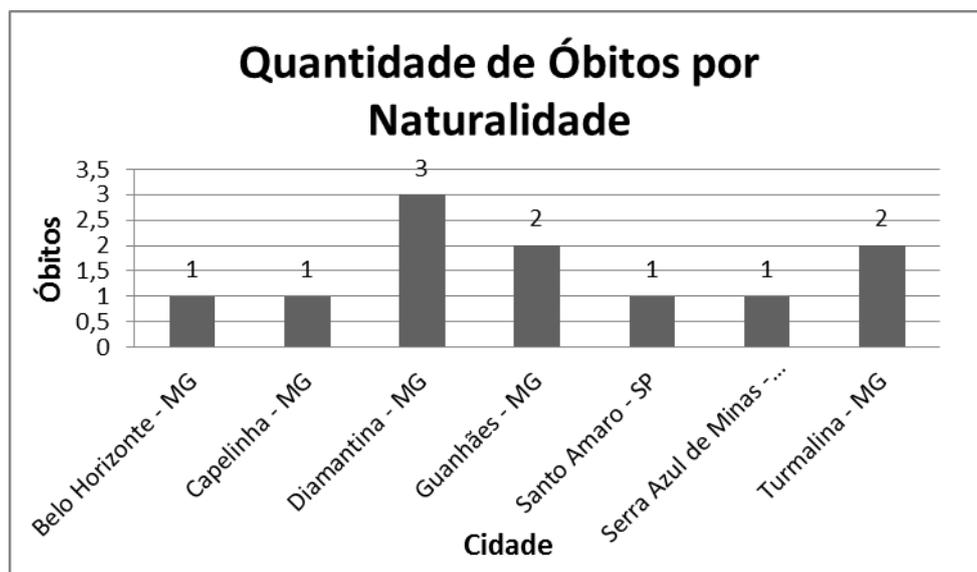


Figura 68. Quantidade de óbitos por naturalidade no *cluster* 09.

A Figura 68 mostra a quantidade de óbitos por naturalidade no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em crianças naturais de Diamantina-MG, com três registros, sendo equivalente a 27% dos 11 registros.

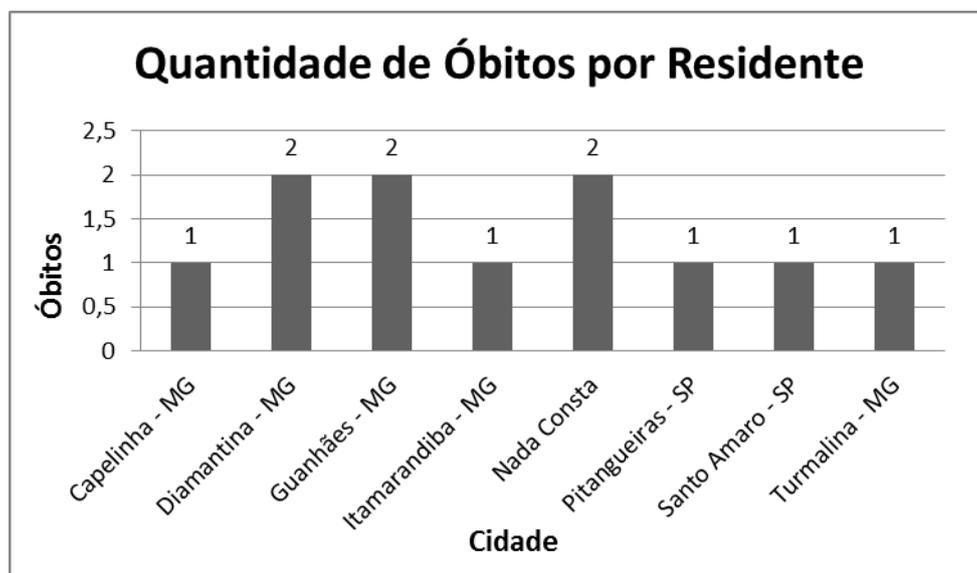


Figura 69. Quantidade de óbitos por residente no *cluster* 09.

A Figura 69 mostra a quantidade de óbitos por residente no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de

crianças que residiam em Diamantina-MG, Guanhães-MG e também dos óbitos que possuíam informação de residência desconhecida. As três predominâncias possuíam dois registros, sendo cada um, equivalente a 18% dos 11 registros do *Cluster*.

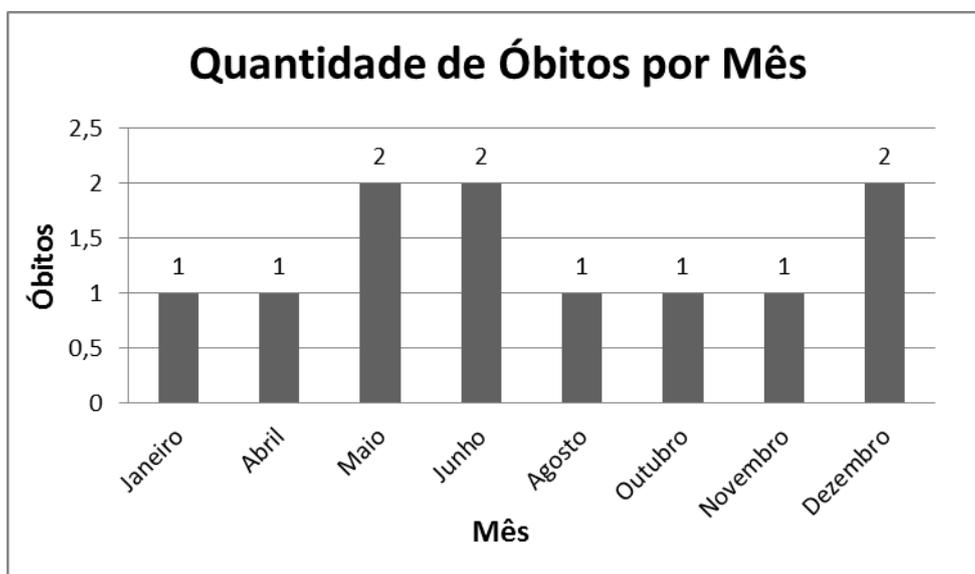


Figura 70. Quantidade de óbitos por mês no *cluster* 09.

A Figura 70 mostra a quantidade de óbitos por mês no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de maio, junho e dezembro, todos com dois registros, sendo equivalente a 18% dos 11 registros do *Cluster*.

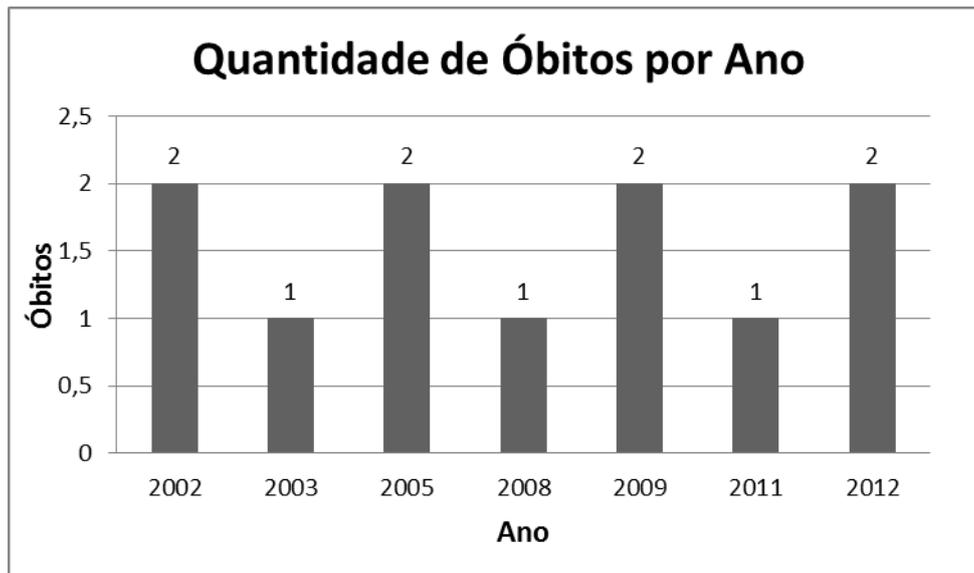


Figura 71. Quantidade de óbitos por ano no *cluster* 09.

A Figura 71 mostra a quantidade de óbitos por ano no *Cluster* 09, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram nos anos de 2002, 2005, 2009 e 2012, todos com dois registros, sendo cada um, equivalente a 18% dos 11 registros.

Cluster 10:

Número de Registros: 04

Tabela:

Tabela 24. *Cluster* 10 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|---------------|---------------|---------------------------------|-----------|-----------|
| 1 | 11 2007 | parada cardio respiratória, cardiopatia congênita | Masculino | 1 ano | Diamantina MG | Diamantina MG | Novembro | 2007 |
| 2 | 10 2005 | insuficiência respiratória, cardiopatia congênita | Masculino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Outubro | 2005 |
| 3 | 03 2010 | insuficiência cardio respiratória, edema agudo do pulmão, acidente escorpiónico | Masculino | 9 anos | Diamantina MG | Conselheiro Mata, Diamantina MG | Agosto | 2010 |
| 4 | 09 2006 | não consta do mandado | Feminino | 16 anos | Diamantina MG | Nada Consta | Mai | 2006 |

Na Tabela 24 é apresentado o *Cluster* 10 com quatro registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 2% dos 147 registros processados.

Gráficos:

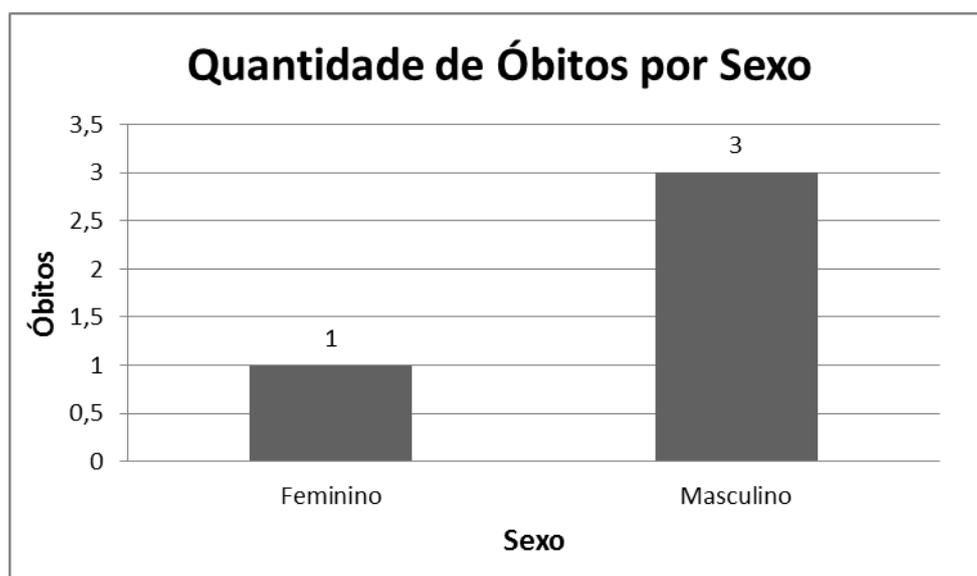


Figura 72. Quantidade de óbitos por sexo no *cluster* 10.

A Figura 72 mostra a quantidade de óbitos por sexo no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos masculinos com três registros, sendo equivalente a 75% dos quatro registros agrupados no *Cluster*.

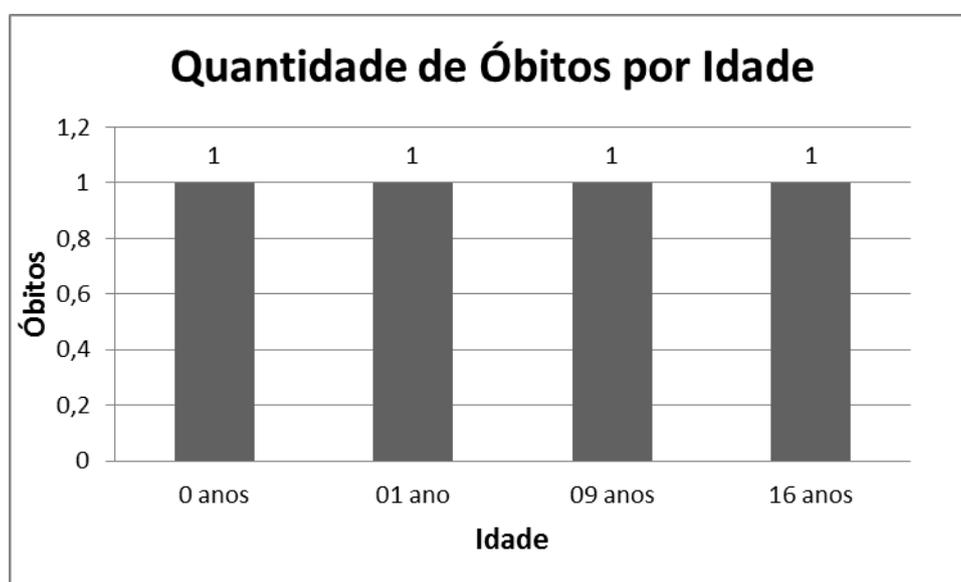


Figura 73. Quantidade de óbitos por idade no *cluster* 10.

A Figura 73 mostra a quantidade de óbitos por idade no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura, cada óbito possuía uma faixa etária diferente, sendo um total de quatro registros que estão agrupados no *Cluster*.

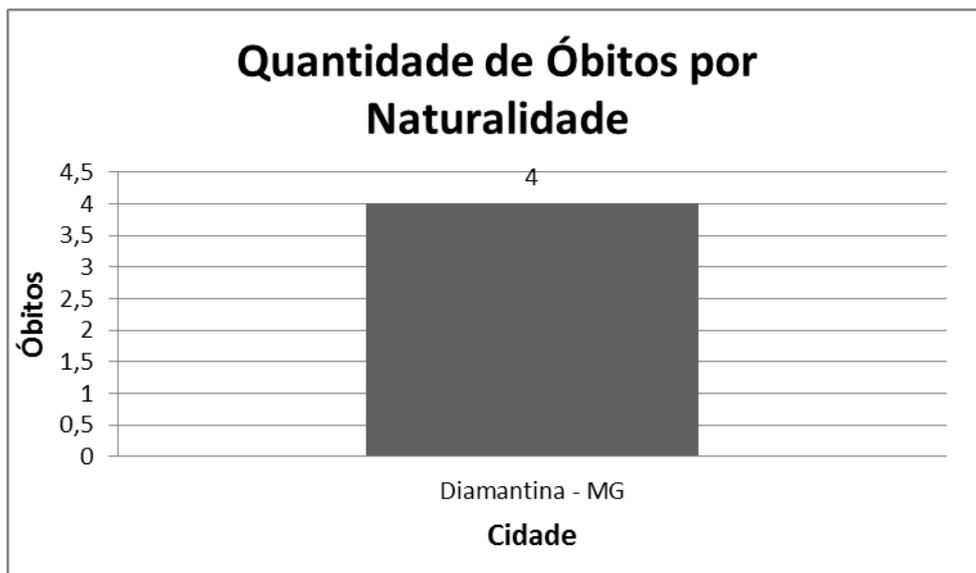


Figura 74. Quantidade de óbitos por naturalidade no *cluster* 10.

A Figura 74 mostra a quantidade de óbitos por naturalidade no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura todas as crianças falecidas são naturais de Diamantina-MG, sendo um total de quatro registros no *Cluster*.

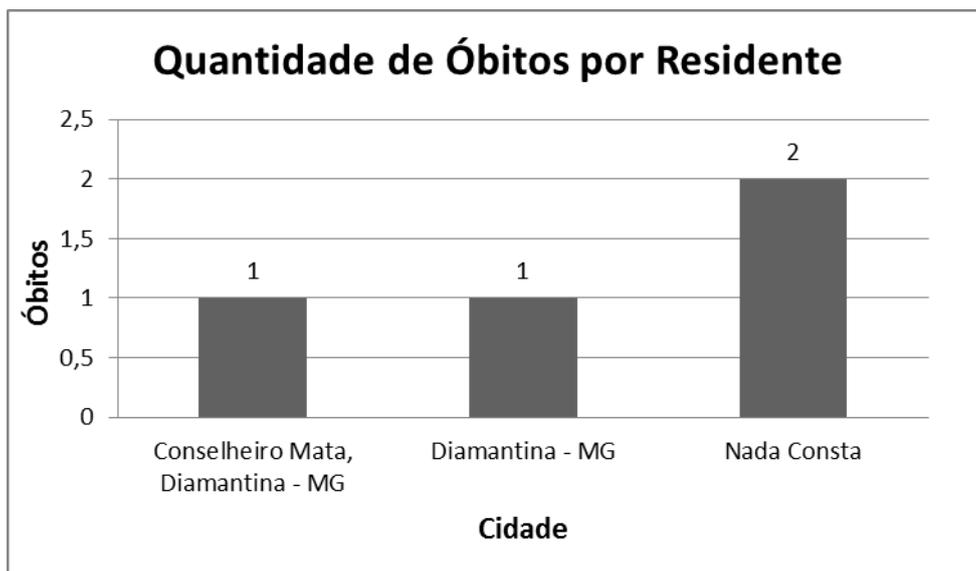


Figura 75. Quantidade de óbitos por residente no *cluster* 10.

A Figura 75 mostra a quantidade de óbitos por residente no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças que não continham a informação de onde elas residiam, sem um total de dois registros. Esta quantidade equivale a 50% dos quatro registros.

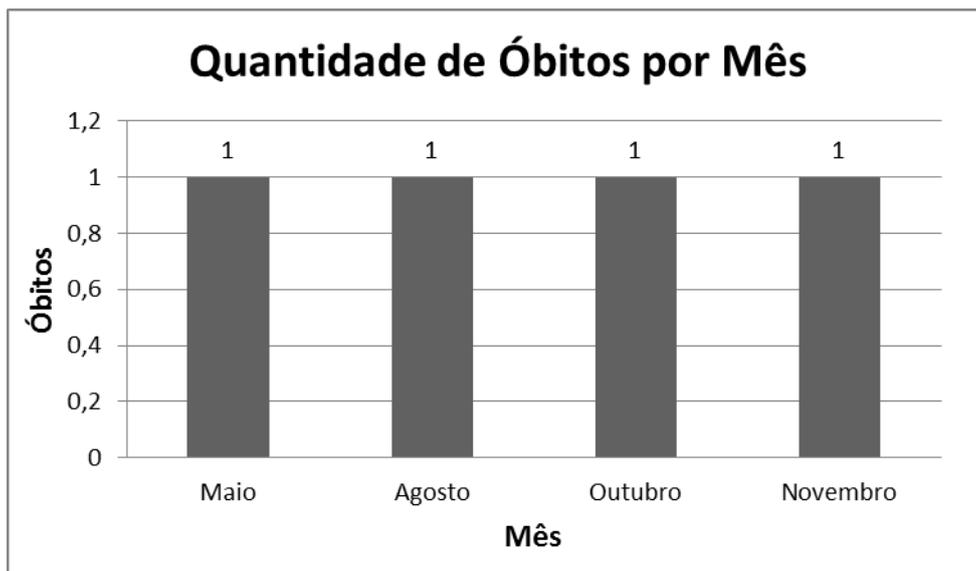


Figura 76. Quantidade de óbitos por mês no *cluster* 10.

A Figura 76 mostra a quantidade de óbitos por mês no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um mês diferente, sendo um total de quatro registros.

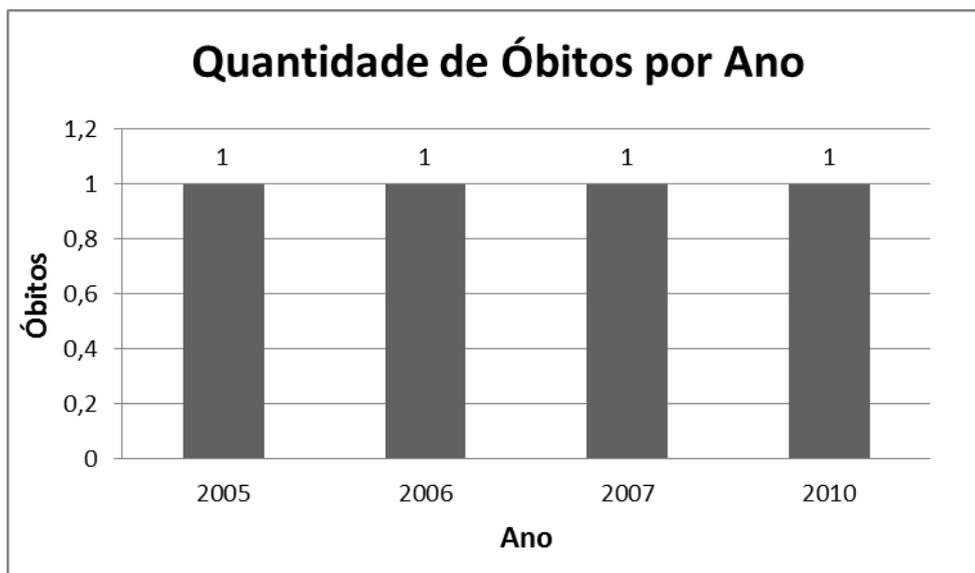


Figura 77. Quantidade de óbitos por ano no *cluster* 10.

A Figura 77 mostra a quantidade de óbitos por ano no *Cluster* 10, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um ano diferente, sendo um total de quatro registros.

Cluster 11:

Número de Registros: 3

Tabela:

Tabela 25. *Cluster 11* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|-----------------------|-----------|-----------|------------------|------------------|-----------|-----------|
| 1 | 06 2003 | politrauma contuso | Feminino | 4 anos | Boa Viagem CE | São Paulo SP | Dezembro | 2003 |
| 2 | 15 2012 | politrauma contuso | Masculino | 4 anos | Diamantina MG | Diamantina MG | Julho | 2012 |
| 3 | 05 2003 | politrauma contuso | Feminino | 8 anos | Crateús CE | São Paulo SP | Dezembro | 2003 |

Na Tabela 25 é apresentado o *Cluster 11* com três registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 2% dos 147 registros processados.

Gráficos:

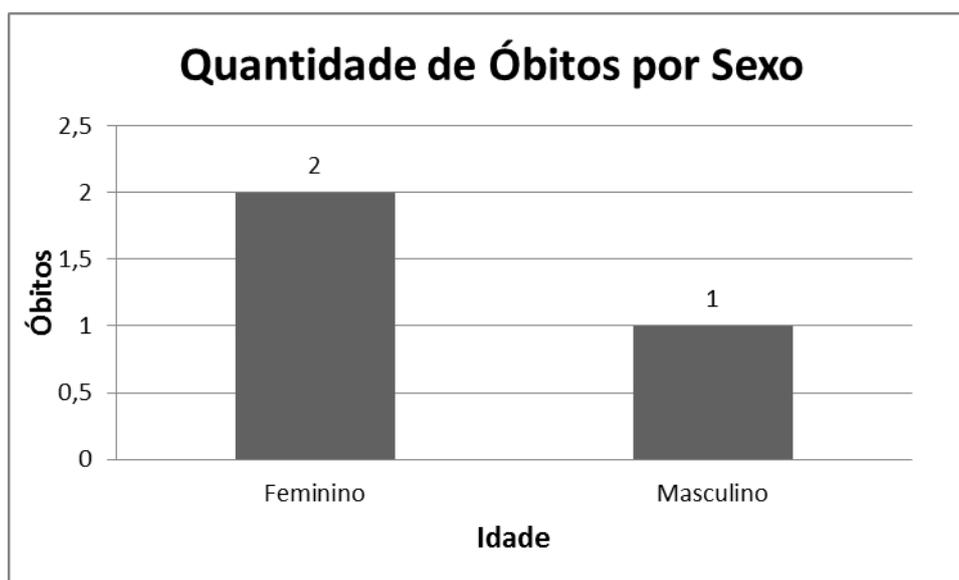


Figura 78. Quantidade de óbitos por sexo no *cluster* 11.

A Figura 78 mostra a quantidade de óbitos por sexo no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

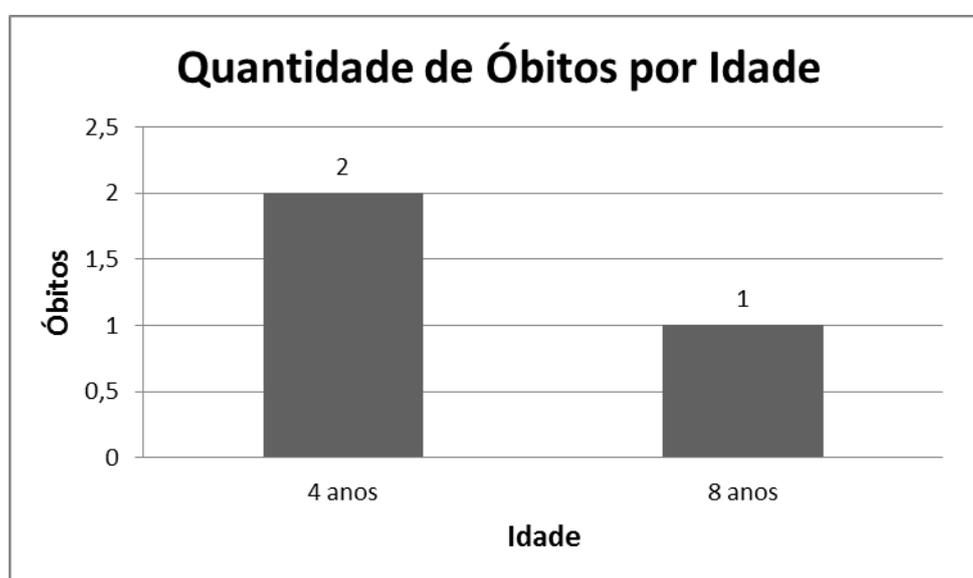


Figura 79. Quantidade de óbitos por idade no *cluster* 11.

A Figura 79 mostra a quantidade de óbitos por idade no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças que

faleceram aos quatro anos de idade, com dois registros, sendo equivalente a 66% dos três registros que foram agrupados no *Cluster*.

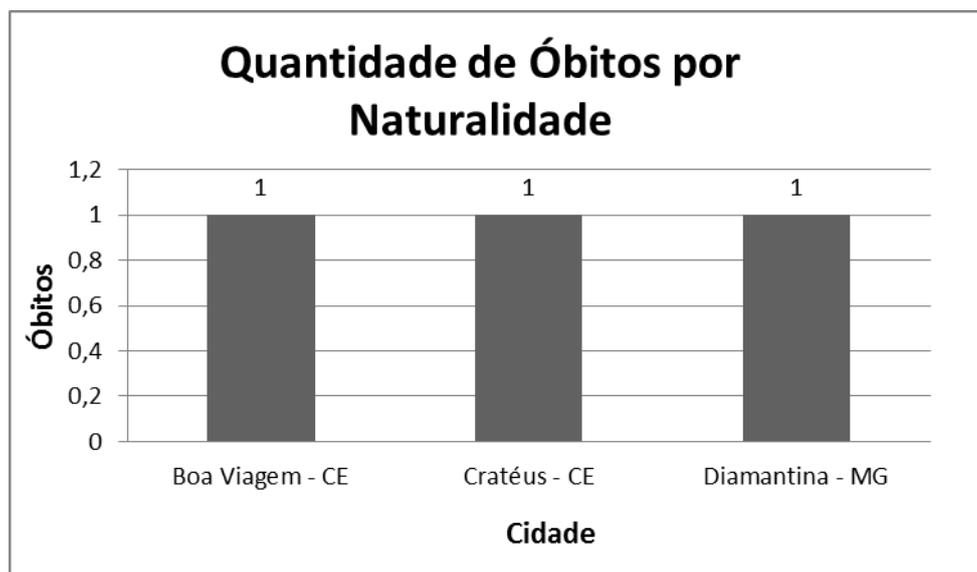


Figura 80. Quantidade de óbitos por naturalidade no *cluster* 11.

A Figura 80 mostra a quantidade de óbitos por mês no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um mês diferente, sendo um total de três registros.

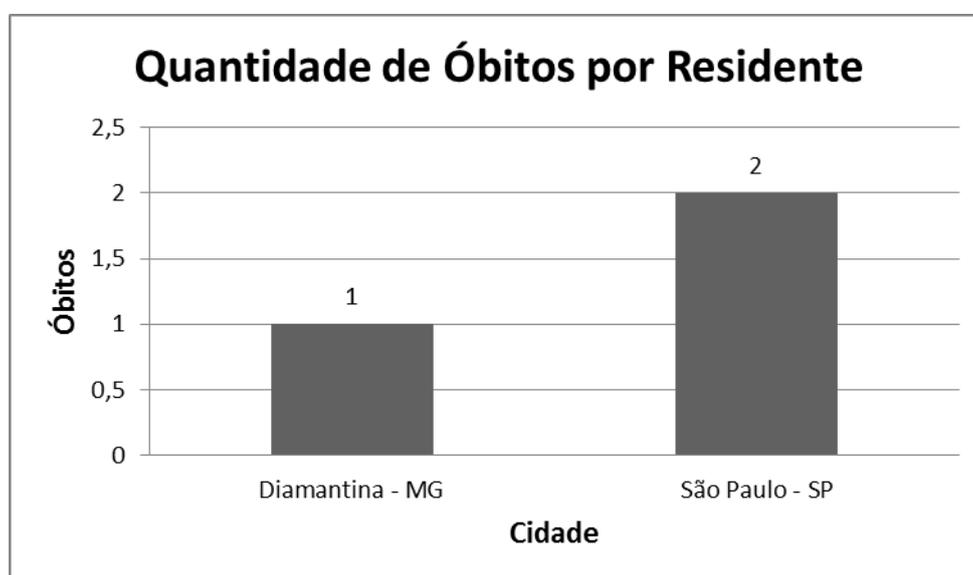


Figura 81. Quantidade de óbitos por residente no *cluster* 11.

A Figura 81 mostra a quantidade de óbitos por residente no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças

falecidas que eram residentes de São Paulo-SP, com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

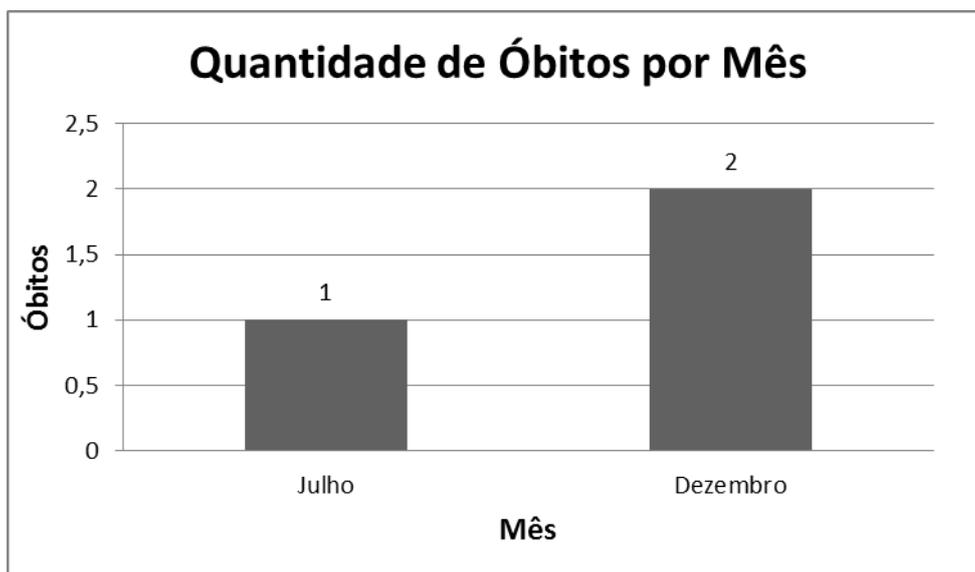


Figura 82. Quantidade de óbitos por mês no *cluster* 11.

A Figura 82 mostra a quantidade de óbitos por mês no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças que faleceram no mês de dezembro, com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

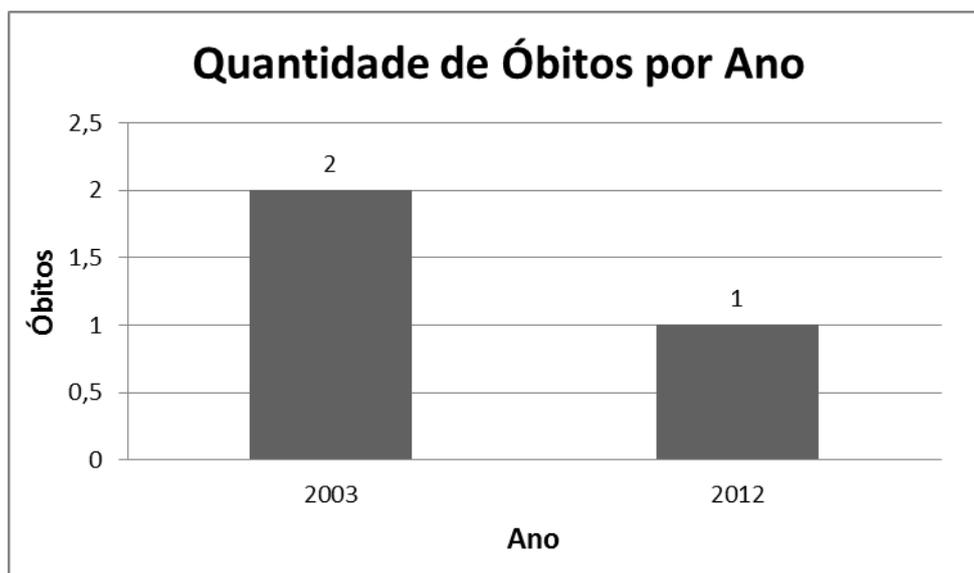


Figura 83. Quantidade de óbitos por ano no *cluster* 11.

A Figura 83 mostra a quantidade de óbitos por ano no *Cluster* 11, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças que

faleceram no ano de 2003, com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

Cluster 12:

Número de Registros: 02

Tabela:

Tabela 26. *Cluster* 12 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|---------|-----------------|----------------|-----------|-----------|
| 1 | 06 2012 | politraumatismo, vítima de acidente de trânsito | Masculino | 14 anos | Minas Novas MG | Minas Novas MG | Abril | 2012 |
| 2 | 05 2012 | traumatismo crânio encefálico, vítima de acidente de trânsito | Masculino | 16 anos | Itamarandiba MG | Carbonita MG | Março | 2012 |

Na Tabela 26 é apresentado o *Cluster* 12 com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

Gráficos:

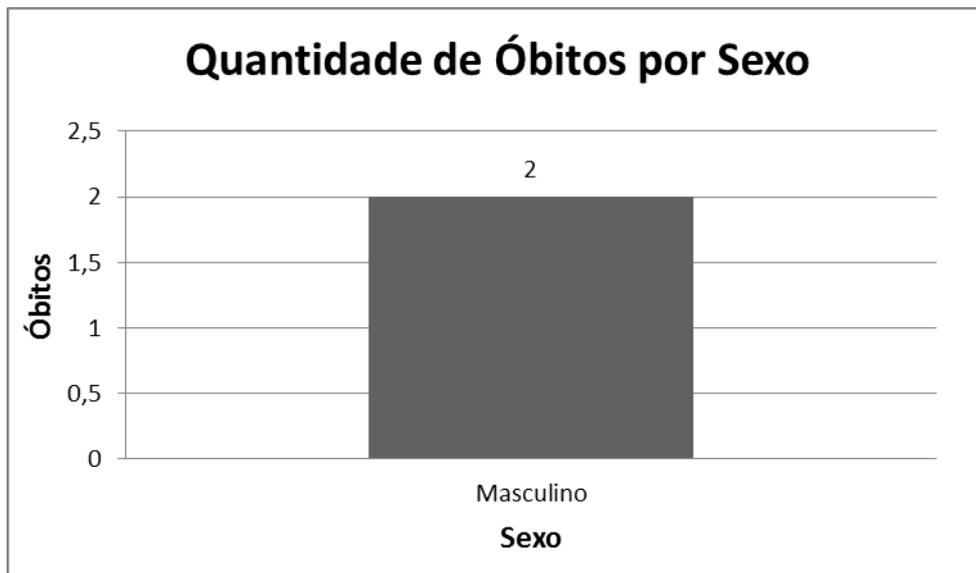


Figura 84. Quantidade de óbitos por sexo no *cluster* 12.

A Figura 84 mostra a quantidade de óbitos por sexo no *Cluster* 12, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças que faleceram eram do sexo masculino, com dois registros agrupados no *Cluster*.

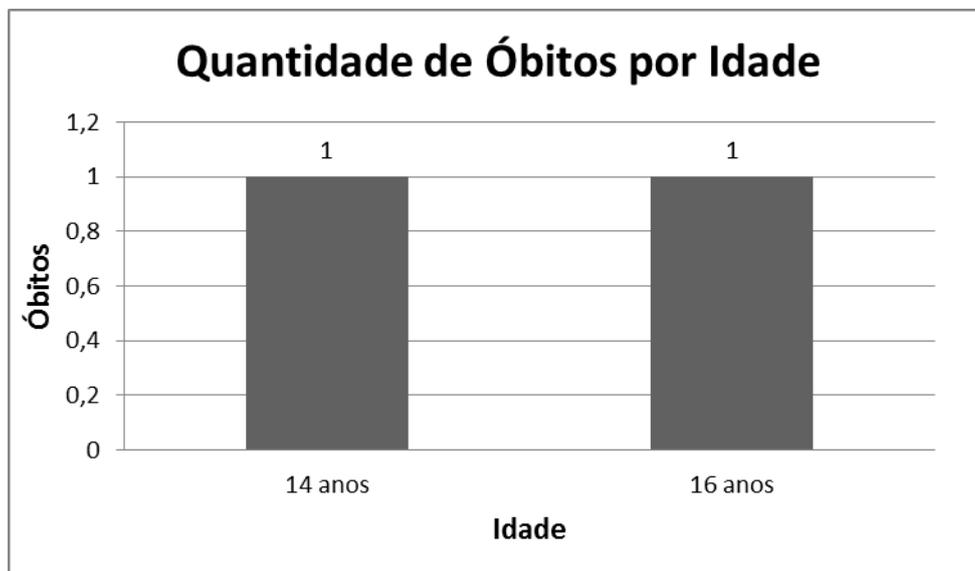


Figura 85. Quantidade de óbitos por idade no *cluster* 12.

A Figura 85 mostra a quantidade de óbitos por idade no *Cluster 12*, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu com uma idade diferente, sendo um total de dois registros.

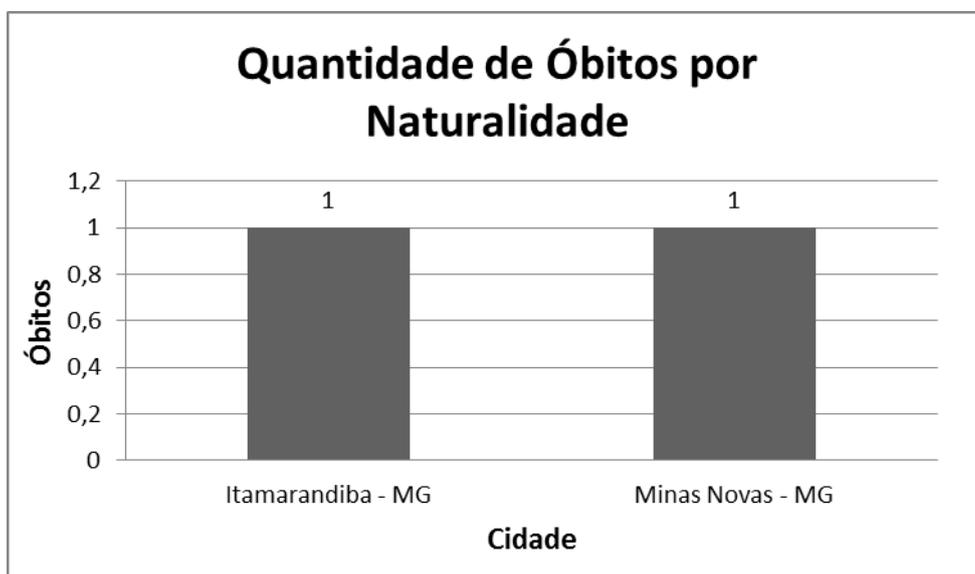


Figura 86. Quantidade de óbitos por naturalidade no *cluster 12*.

A Figura 86 mostra a quantidade de óbitos por naturalidade no *Cluster 12*, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança era natural de uma cidade diferente, sendo um total de dois registros.

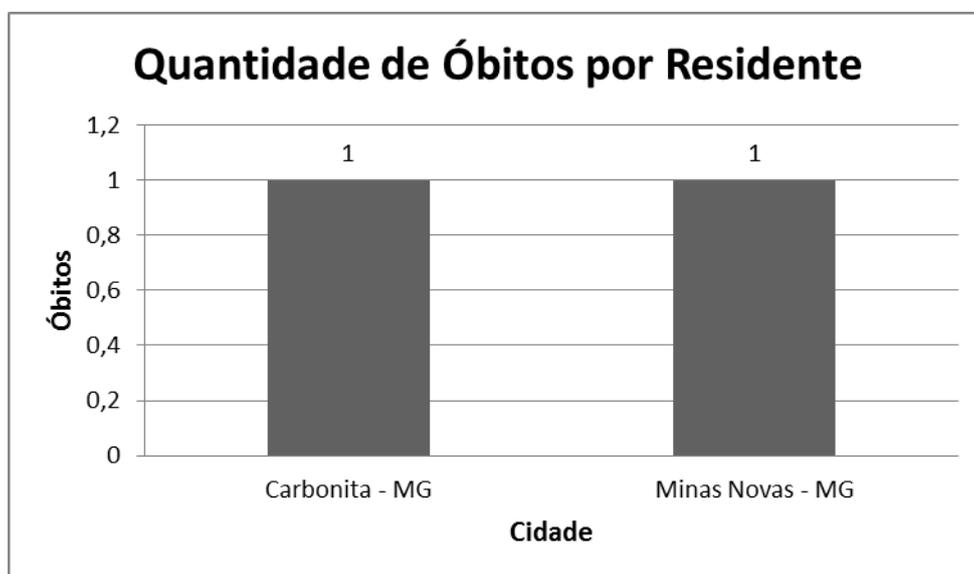


Figura 87. Quantidade de óbitos por residente no *cluster 12*.

A Figura 87 mostra a quantidade de óbitos por residente no *Cluster 12*, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança era residente de uma cidade diferente, sendo um total de dois registros.

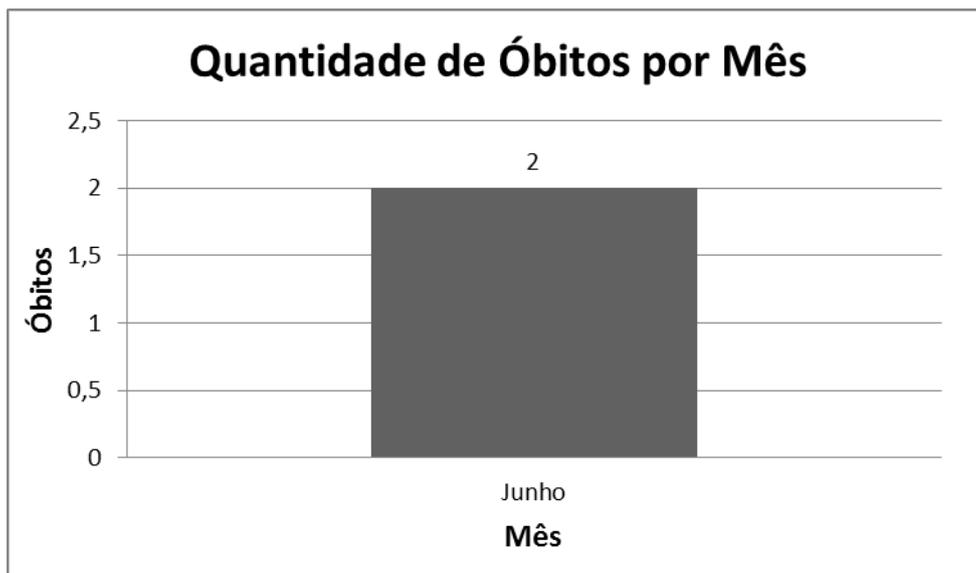


Figura 88. Quantidade de óbitos por mês no *cluster* 12.

A Figura 88 mostra a quantidade de óbitos por mês no *Cluster* 12, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças faleceram no mês de junho, com dois registros agrupados no *Cluster*.

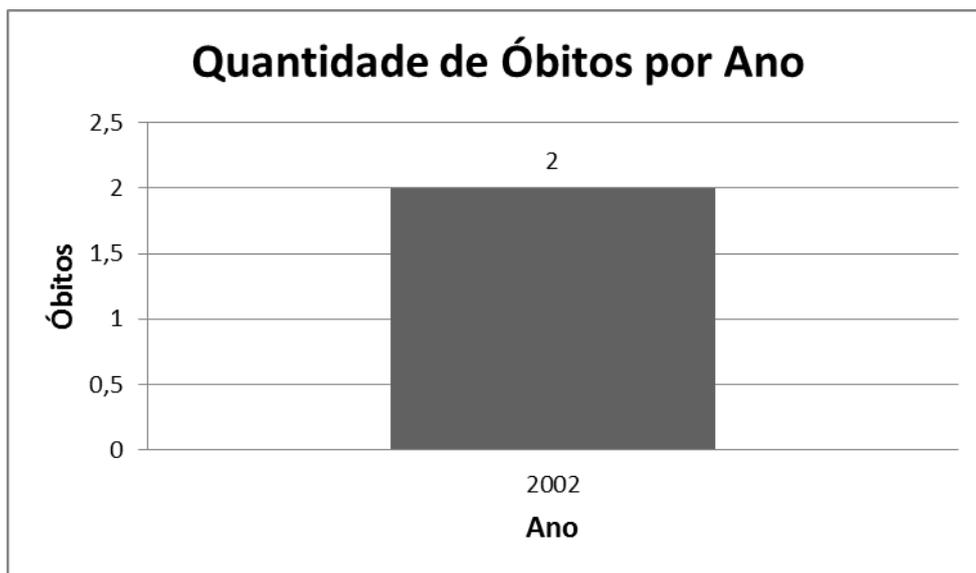


Figura 89. Quantidade de óbitos por ano no *cluster* 12.

A Figura 89 mostra a quantidade de óbitos por ano no *Cluster* 12, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças faleceram no ano de 2002, com dois registros agrupados no *Cluster*.

Cluster 13:

Número de Registros: 03

Tabela:

Tabela 27. Cluster 13 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------------|-------------------|--------------------|--------------------|---------------|-----------|
| 1 | 02 2009 | choque cardiogênico, cardiopatía congénita | Femi- nino | 2 dias (0 ano) | Diamantina MG | Nada Consta | Janeiro | 2009 |
| 2 | 34 2012 | choque, neuropatia congénita grave | Femi- nino | 6 anos | Itamarandiba MG | Itamarandiba MG | Novem- bro | 2012 |
| 3 | 09 2005 | choque cardiogênico, miocardiopatía dilatada | Mascu- -lino | 1 ano | Diamantina MG | Diamantina MG | Setem- bro | 2005 |

Na Tabela 27 é apresentado o *Cluster* 13 com três registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 2% dos 147 registros processados.

Gráficos:

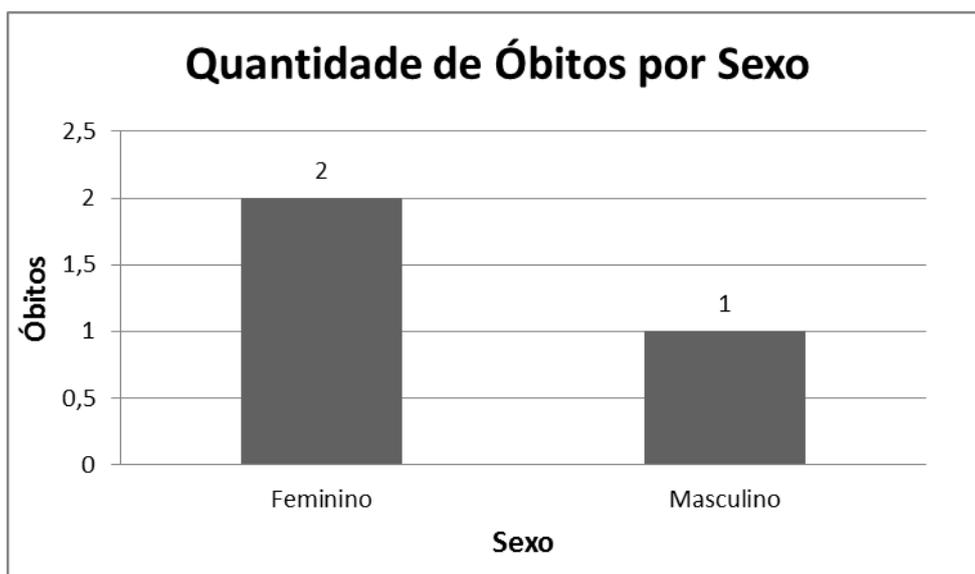


Figura 90. Quantidade de óbitos por sexo no *cluster* 13.

A Figura 90 mostra a quantidade de óbitos por sexo no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

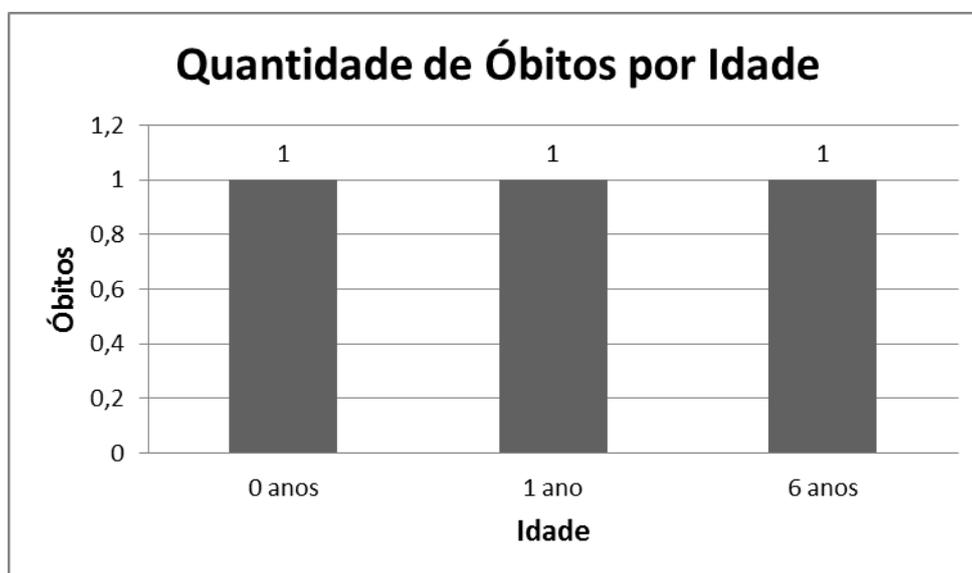


Figura 91. Quantidade de óbitos por idade no *cluster* 13.

A Figura 91 mostra a quantidade de óbitos por idade no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu com uma idade diferente, sendo um total de três registros.

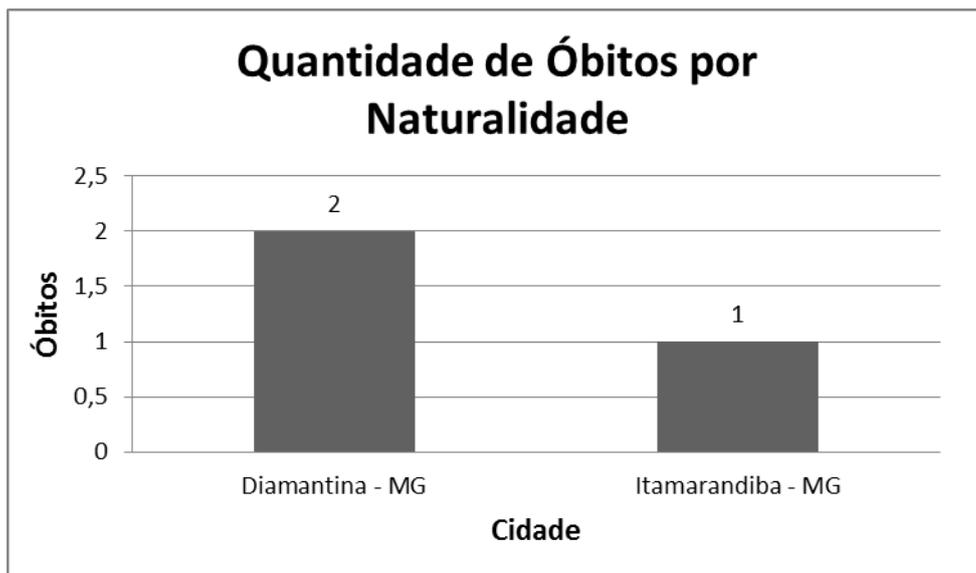


Figura 92. Quantidade de óbitos por naturalidade no *cluster* 13.

A Figura 92 mostra a quantidade de óbitos por naturalidade no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças naturais de Diamantina-MG, com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

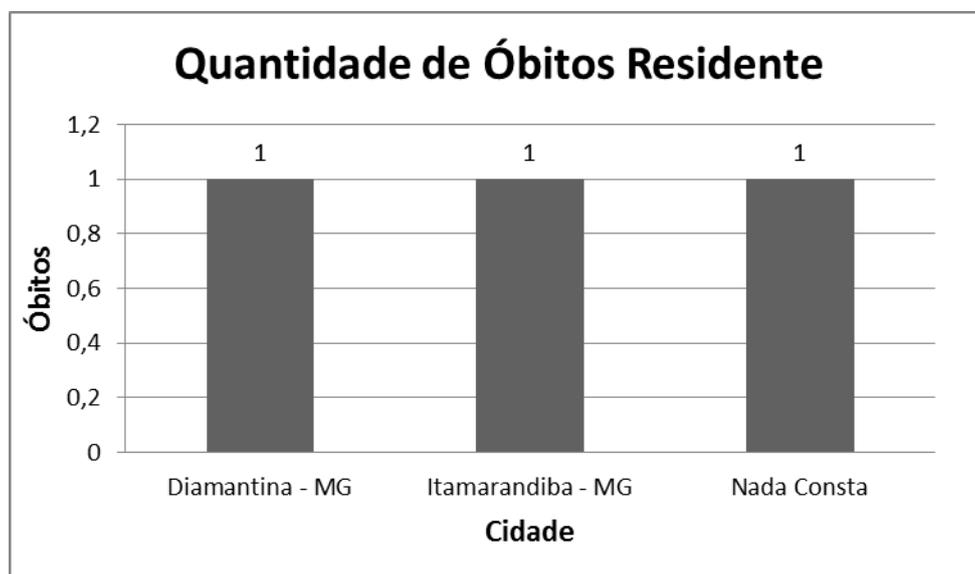


Figura 93. Quantidade de óbitos por residente no *cluster* 13.

A Figura 93 mostra a quantidade de óbitos por residente no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança falecida era residente de uma cidade diferente ou não constava em suas informações, sendo um total de três registros.

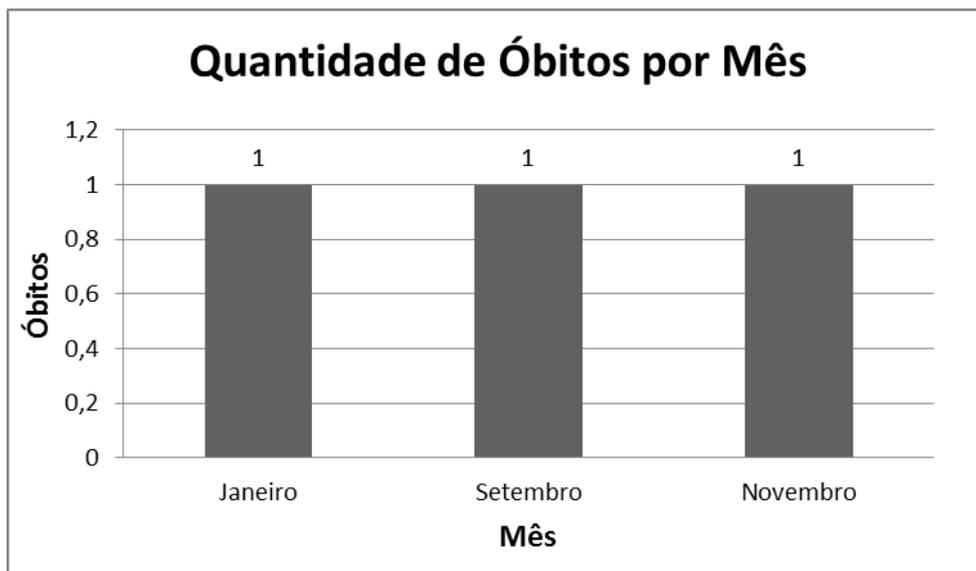


Figura 94. Quantidade de óbitos por mês no *cluster* 13.

A Figura 94 mostra a quantidade de óbitos por mês no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um mês diferente, sendo um total de três registros.

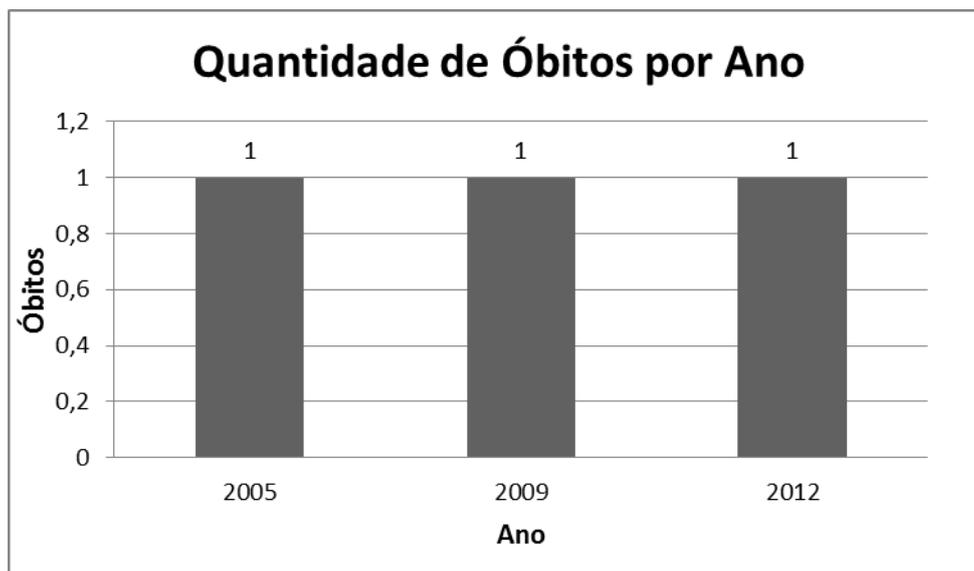


Figura 95. Quantidade de óbitos por ano no *cluster* 13.

A Figura 95 mostra a quantidade de óbitos por ano no *Cluster* 13, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um ano diferente, sendo um total de três registros.

Cluster 14:

Número de Registros: 03

Tabela:

Tabela 28. Cluster 14 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|------------|------------------|--|-----------|-----------|
| 1 | 13 2011 | choque neurogênico, edema cerebral, encefalite | Feminino | 5 anos | Diamantina MG | Senador Mourão, Diamantina MG | Fevereiro | 2011 |
| 2 | 01 2011 | choque neurogênico, encefalopatia hipóxica, hipoxia pós-parada, pneumonia, diabetes tipo I | Feminino | 16 anos | Diamantina MG | Diamantina MG | Junho | 2011 |
| 3 | 08 2004 | encefalite | Masculino | 8 anos | Diamantina MG | Inhaí SP | Dezembro | 2004 |

Na Tabela 28 é apresentado o *Cluster 14* com três registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 2% dos 147 registros processados.

Gráficos:

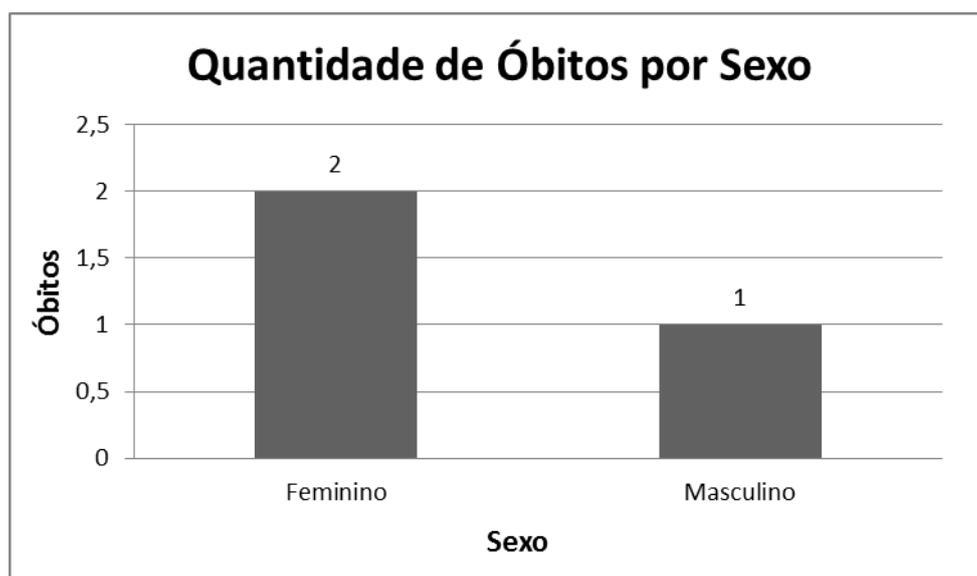


Figura 96. Quantidade de óbitos por sexo no *cluster* 14.

A Figura 96 mostra a quantidade de óbitos por sexo no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com dois registros, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

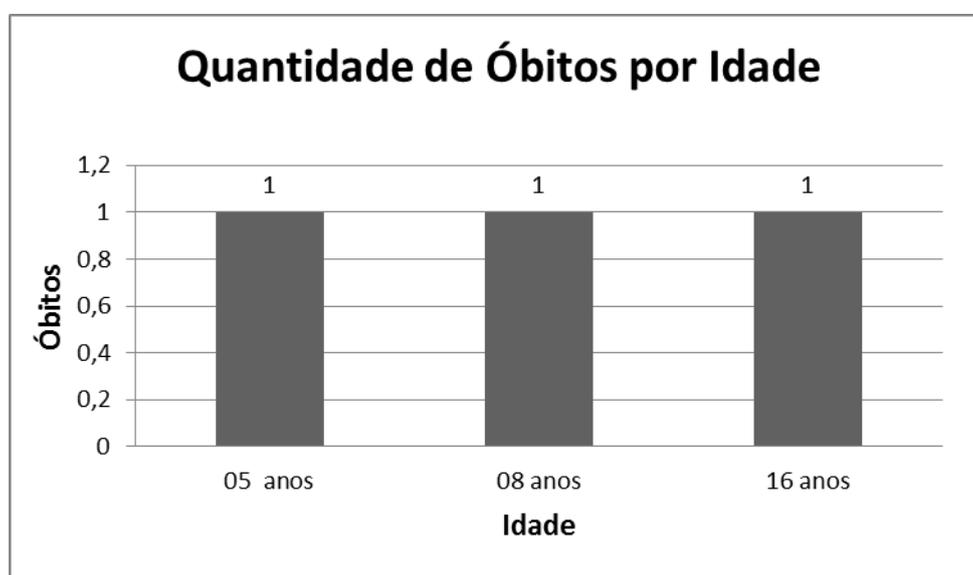


Figura 97. Quantidade de óbitos por idade no *cluster* 14.

A Figura 97 mostra a quantidade de óbitos por idade no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu com uma idade diferente, sendo um total de três registros.

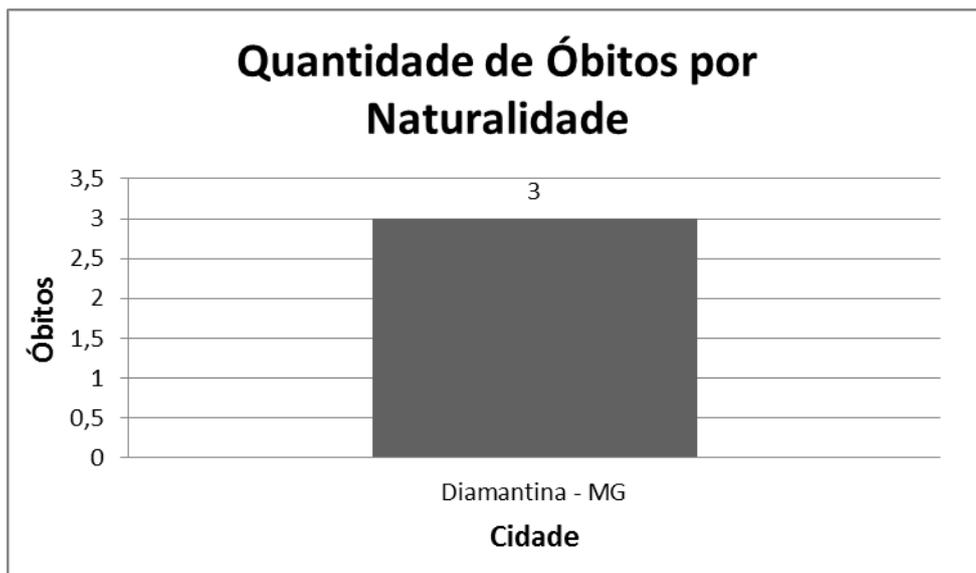


Figura 98. Quantidade de óbitos por naturalidade no *cluster* 14.

A Figura 98 mostra a quantidade de óbitos por naturalidade no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram naturais de Diamantina-MG, com três registros agrupados no *Cluster*.

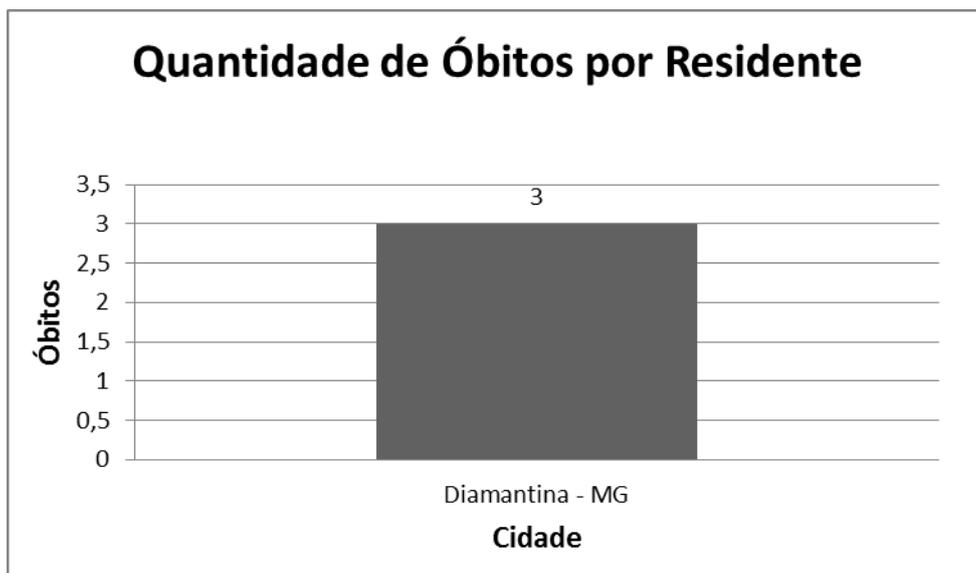


Figura 99. Quantidade de óbitos por residente no *cluster* 14.

A Figura 99 mostra a quantidade de óbitos por residente no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram residentes em Diamantina-MG, com três registros agrupados no *Cluster*.

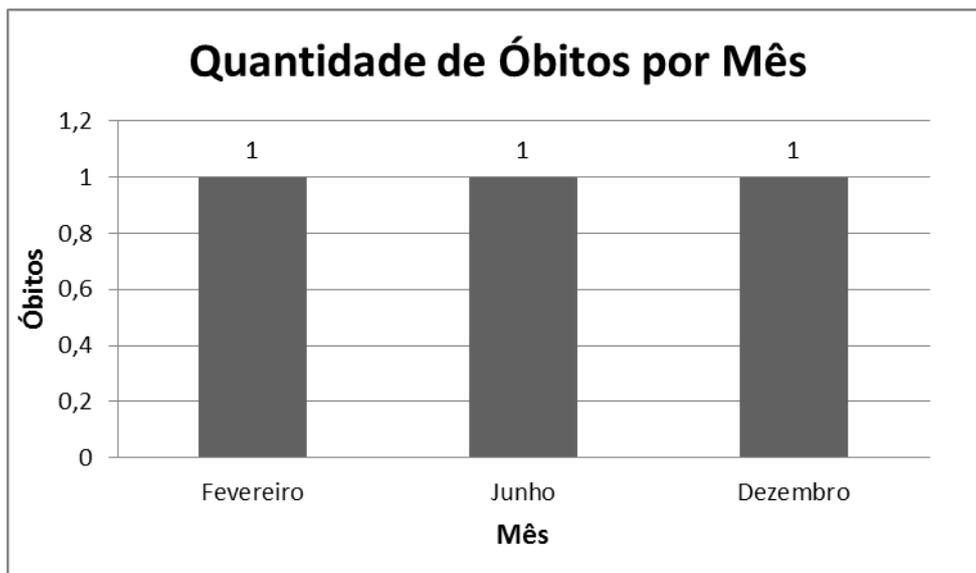


Figura 100. Quantidade de óbitos por mês no *cluster* 14.

A Figura 100 mostra a quantidade de óbitos por mês no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um mês diferente, sendo um total de três registros.

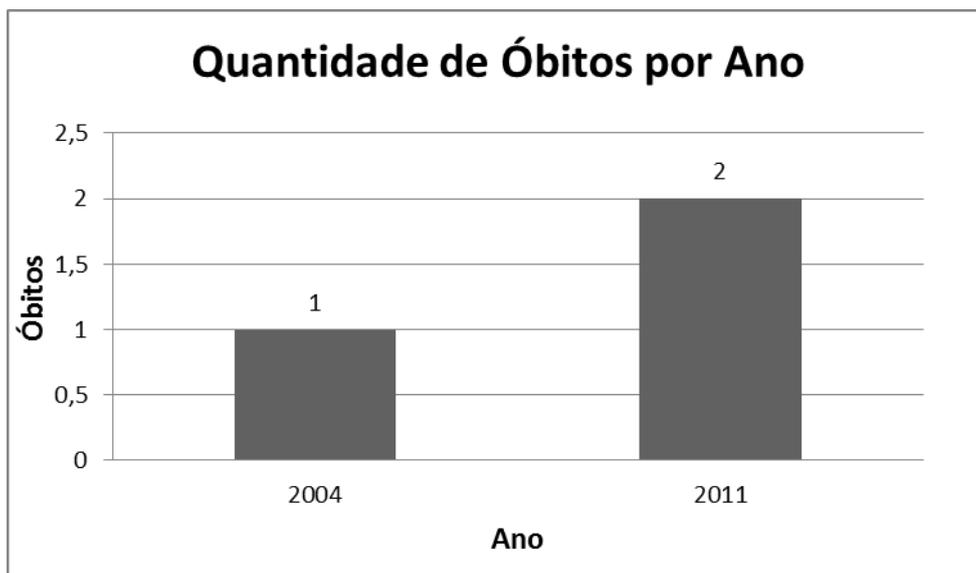


Figura 101. Quantidade de óbitos por ano no *cluster* 14.

A Figura 101 mostra a quantidade de óbitos por ano no *Cluster* 14, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram em 2011, sendo equivalente a 66% dos três registros agrupados no *Cluster*.

Cluster 15:

Número de Registros: 06

Tabela:

Tabela 29. Cluster 15 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|----------------|--------------------|------------------|--|----------------|-----------|
| 1 | 04 2007 | parada cardiorrespiratória, septicemia | Femi- nino | 4 anos | Diamantina MG | Felício dos Santos MG | Abril | 2007 |
| 2 | 06 2005 | parada cardiorrespiratória, morte súbita neonatal | Mascu- lino | 14 dias (0 ano) | Diamantina MG | Nada Consta | Novem- bro | 2005 |
| 3 | 07 2003 | septicemia, desnutrição | Mascu- lino | 9 meses (0 ano) | Diamantina MG | Diamantina MG | Dezem- bro | 2003 |
| 4 | 09 2008 | desnutrição III | Femi- nino | 7 anos | Diamantina MG | Senador Mourão, Diamantina MG | Novem- bro | 2008 |
| 5 | 07 2010 | pneumonia adquirida na comunidade, amiotrofia Espanhol grau III | Mascu- lino | 16 anos | Diamantina MG | Sopa, Diamantina MG | Feve- reiro | 2010 |
| 6 | 05 2006 | parada cardiorrespiratória | Femi- nino | 2 anos | Diamantina MG | Diamantina MG | Junho | 2006 |

Na Tabela 29 é apresentado o *Cluster 15* com seis registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

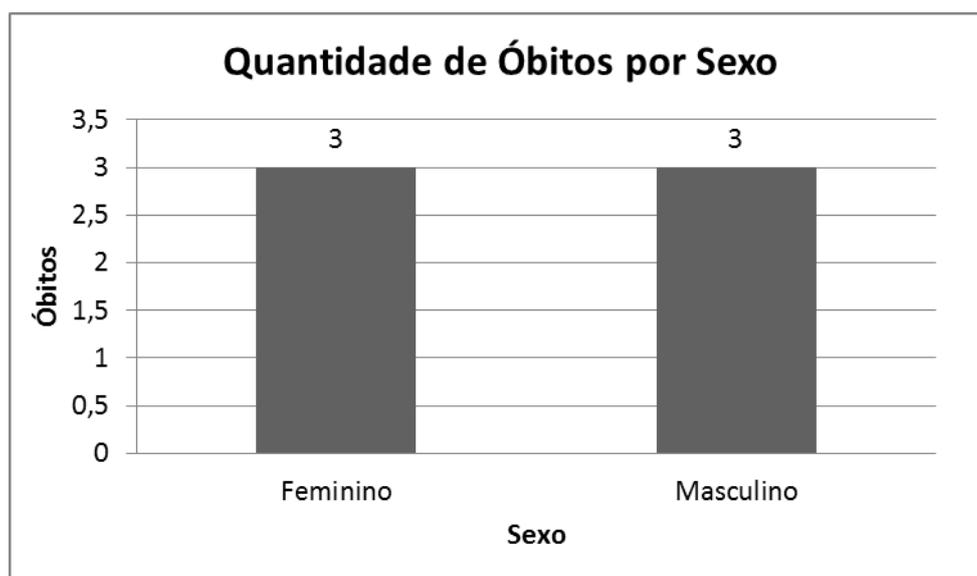


Figura 102. Quantidade de óbitos por sexo no *cluster* 15.

A Figura 102 mostra a quantidade de óbitos por sexo no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber que houve a mesma quantidade de óbitos tanto para o sexo feminino, quanto para o masculino, sendo ambos equivalente a 50% dos seis registros agrupados no *Cluster*.

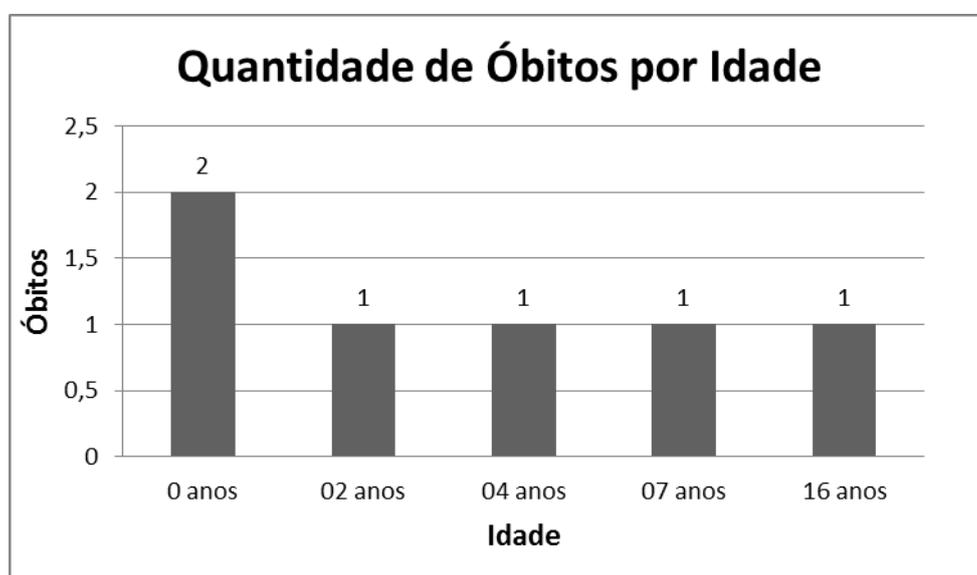


Figura 103. Quantidade de óbitos por idade no *cluster* 15.

A Figura 103 mostra a quantidade de óbitos por idade no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos

que ocorreram na faixa etária abaixo de um ano, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

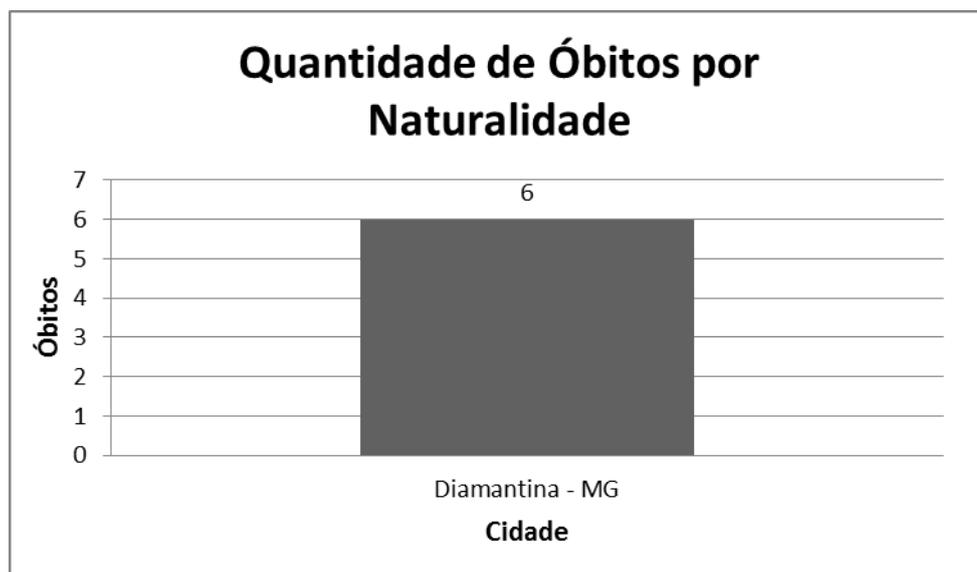


Figura 104. Quantidade de óbitos por naturalidade no *cluster* 15.

A Figura 104 mostra a quantidade de óbitos por naturalidade no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram naturais de Diamantina-MG, com seis registros agrupados no *Cluster*.

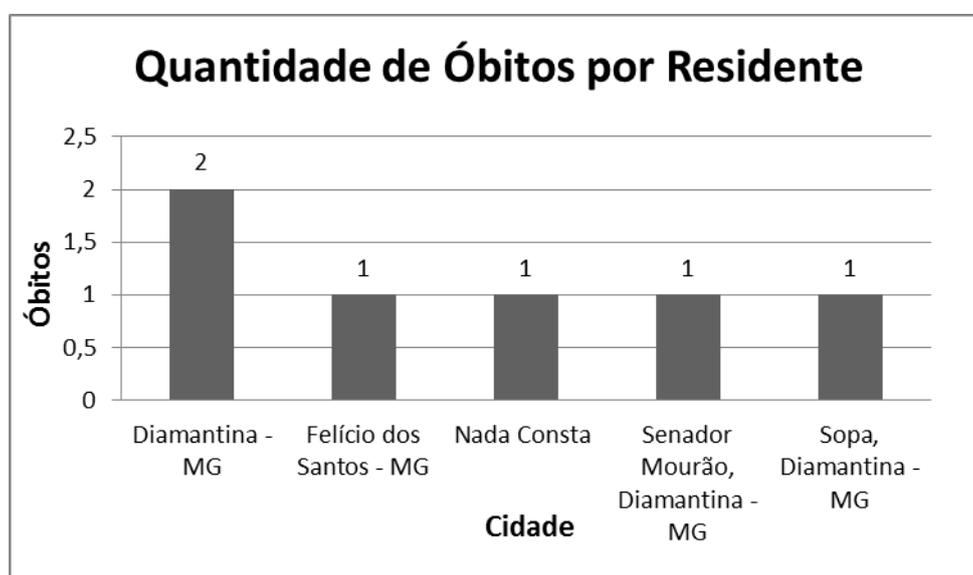


Figura 105. Quantidade de óbitos por residente no *cluster* 15.

A Figura 105 mostra a quantidade de óbitos por residente no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos em

crianças que residiam em Diamantina-MG, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

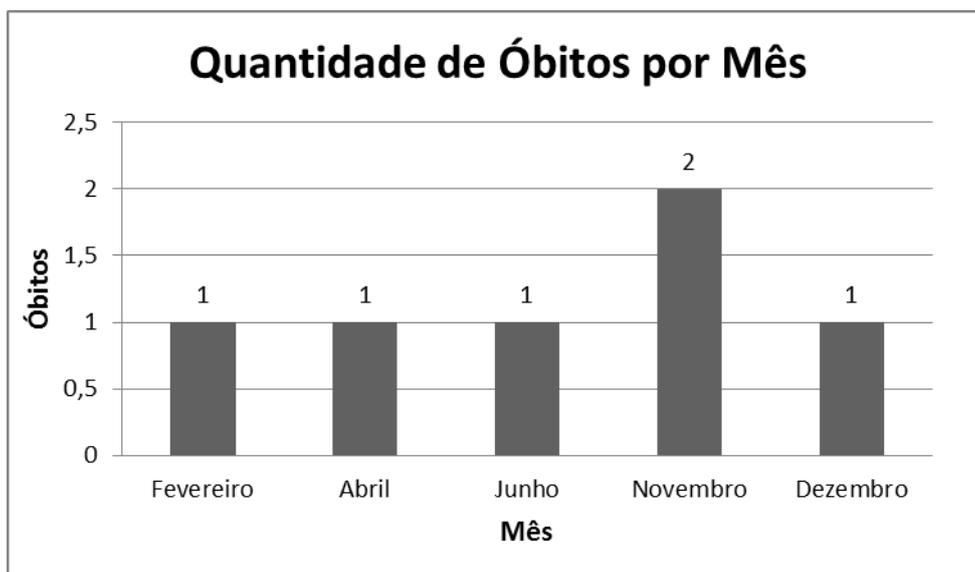


Figura 106. Quantidade de óbitos por mês no *cluster* 15.

A Figura 106 mostra a quantidade de óbitos por mês no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos no mês de novembro, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

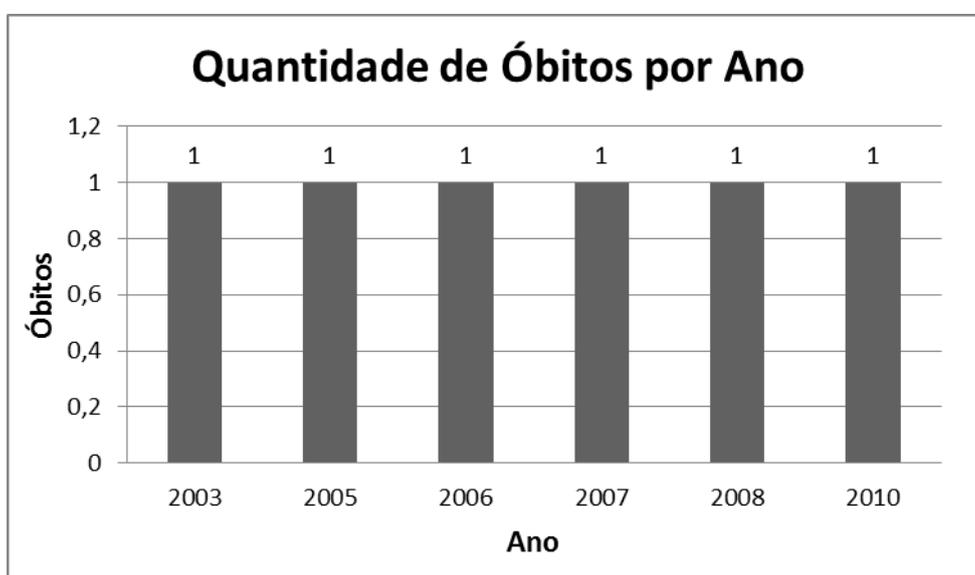


Figura 107. Quantidade de óbitos por ano no *cluster* 15.

A Figura 107 mostra a quantidade de óbitos por ano no *Cluster* 15, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um ano diferente, sendo um total de seis registros.

Cluster 16

Número de Registros: 01

Tabela:

Tabela 30. Cluster 16 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--------------------|-----------|-----------|------------------|------------------|-----------|-----------|
| 1 | 10 2008 | causa desconhecida | Masculino | 2 anos | Diamantina MG | Diamantina MG | Março | 2008 |

Na Tabela 30 é apresentado o *Cluster* 16 com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de só haver um registro no agrupamento.

Gráficos:

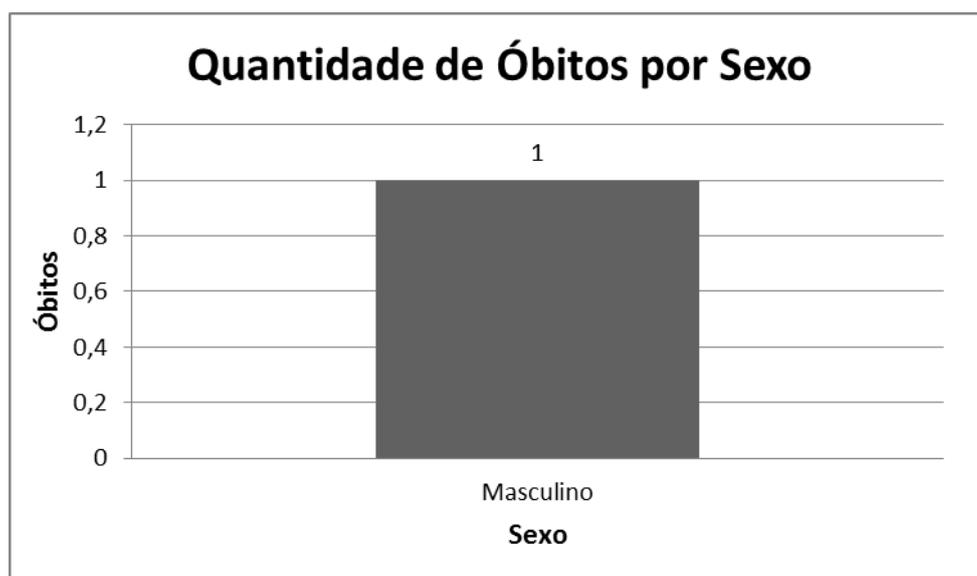


Figura 108. Quantidade de óbitos por sexo no *cluster* 16.

A Figura 108 mostra a quantidade de óbitos por sexo no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, o único óbito do *Cluster* pertencia ao sexo feminino.

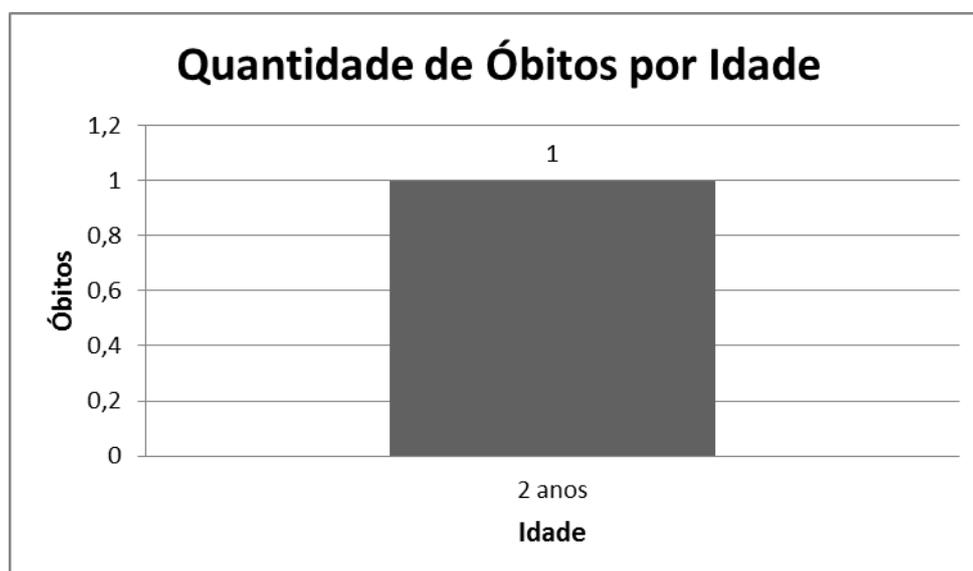


Figura 109. Quantidade de óbitos por idade no *cluster* 16.

A Figura 109 mostra a quantidade de óbitos por idade no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, o único óbito do *Cluster* possuía a idade de dois anos.

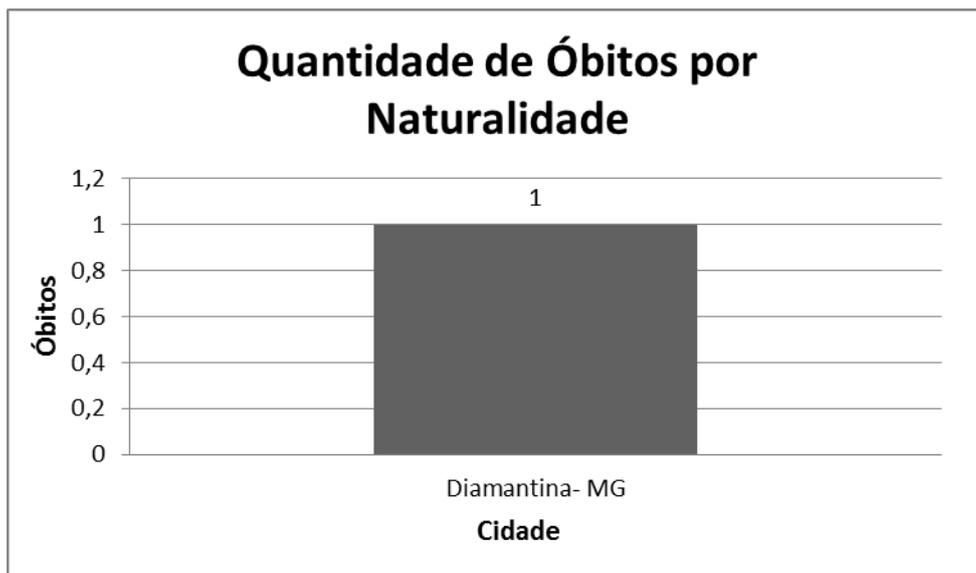


Figura 110. Quantidade de óbitos por naturalidade no *cluster* 16.

A Figura 110 mostra a quantidade de óbitos por naturalidade no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural de Diamantina-MG.

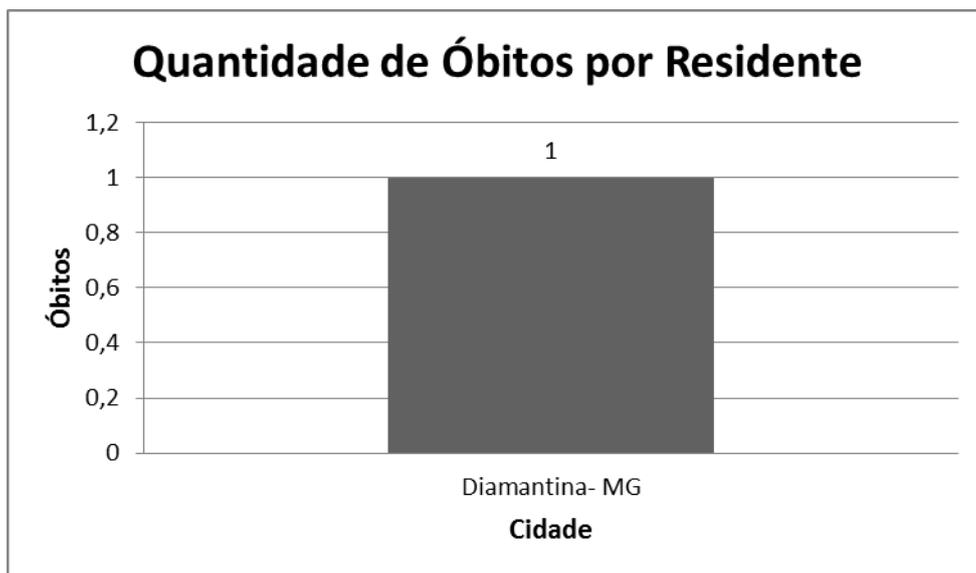


Figura 111. Quantidade de óbitos por residente no *cluster* 16.

A Figura 111 mostra a quantidade de óbitos por residente no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* residia em Diamantina-MG.

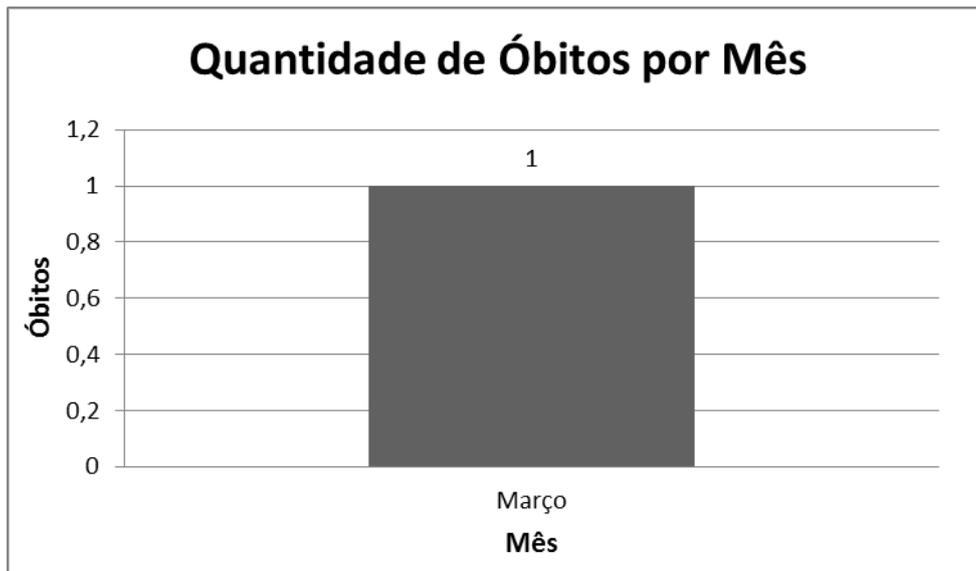


Figura 112. Quantidade de óbitos por mês no *cluster* 16.

A Figura 112 mostra a quantidade de óbitos por mês no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, o único óbito ocorreu no mês de março.

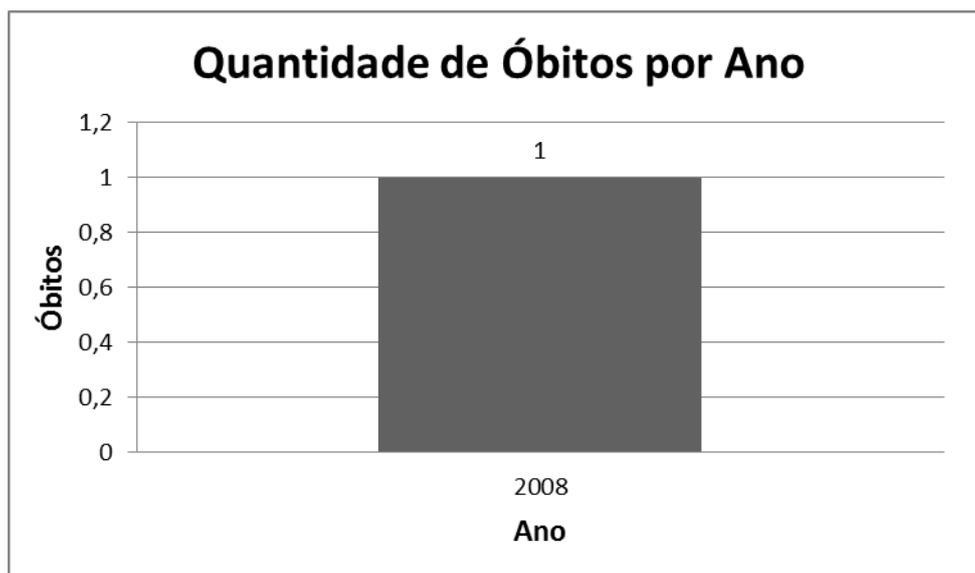


Figura 113. Quantidade de óbitos por ano no *cluster* 16.

A Figura 113 mostra a quantidade de óbitos por ano no *Cluster* 16, obtidos pelo modelo Cassiopeia. Nesta Figura, o único óbito ocorreu no ano de 2008.

Cluster 17:

Número de Registros: 06

Tabela:

Tabela 31. Cluster 17 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|------------|------------------|------------------|-----------|-----------|
| 1 | 01 2008 | hipoxemia grave, parada cardiorrespiratória | Feminino | 0 ano | Diamantina MG | Diamantina MG | Fevereiro | 2008 |
| 2 | 12 2002 | parada cardiorrespiratória, causa desconhecida | Masculino | 2 anos | Diamantina MG | Diamantina MG | Dezembro | 2002 |
| 3 | 01 2005 | S.A.R.A-hipoxemia, insuficiência hepática | Feminino | 0 ano | Diamantina MG | Nada Consta | Abril | 2005 |
| 4 | 26 2012 | choque refratário, acidose metabólica grave | Feminino | 3 anos | Curvelo MG | Gouveia MG | Outubro | 2012 |
| 5 | 03 2009 | morte sem assistência médica de causa indeterminada | Feminino | 0 ano | Diamantina MG | Nada Consta | Janeiro | 2009 |
| 6 | 03 2004 | morte de causa violenta e etiologia a esclarecer | Masculino | 16 anos | Diamantina MG | Diamantina MG | Agosto | 2004 |

Na Tabela 31 é apresentado o *Cluster 17*, com seis registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

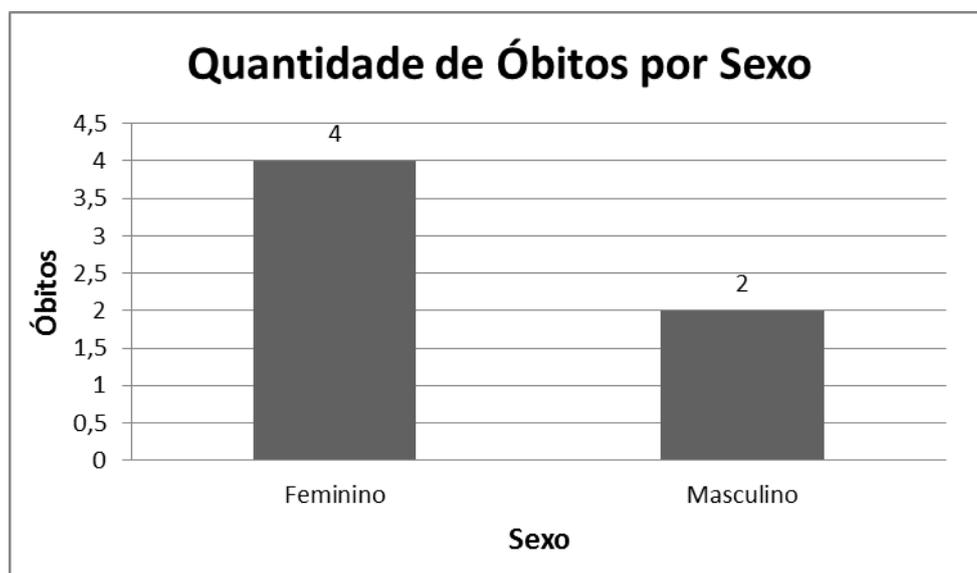


Figura 114. Quantidade de óbitos por sexo no *cluster* 17.

A Figura 114 mostra a quantidade de óbitos por sexo no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos femininos com quatro registros, sendo equivalente a 66% dos seis registros agrupados no *Cluster*.

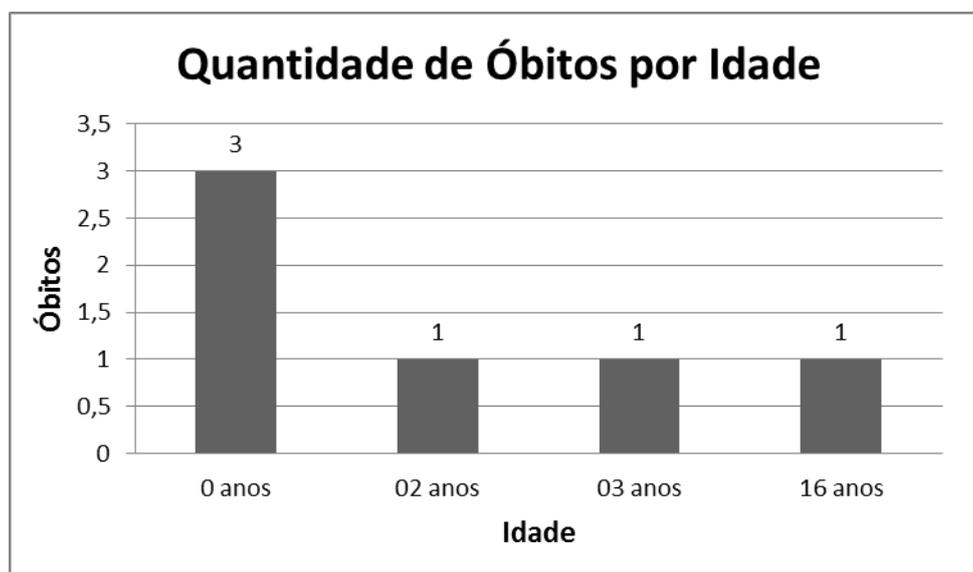


Figura 115. Quantidade de óbitos por idade no *cluster* 17.

A Figura 115 mostra a quantidade de óbitos por idade no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos na

faixa de idade abaixo de um ano, com três registros, sendo equivalente a 50% dos seis registros agrupados no *Cluster*.

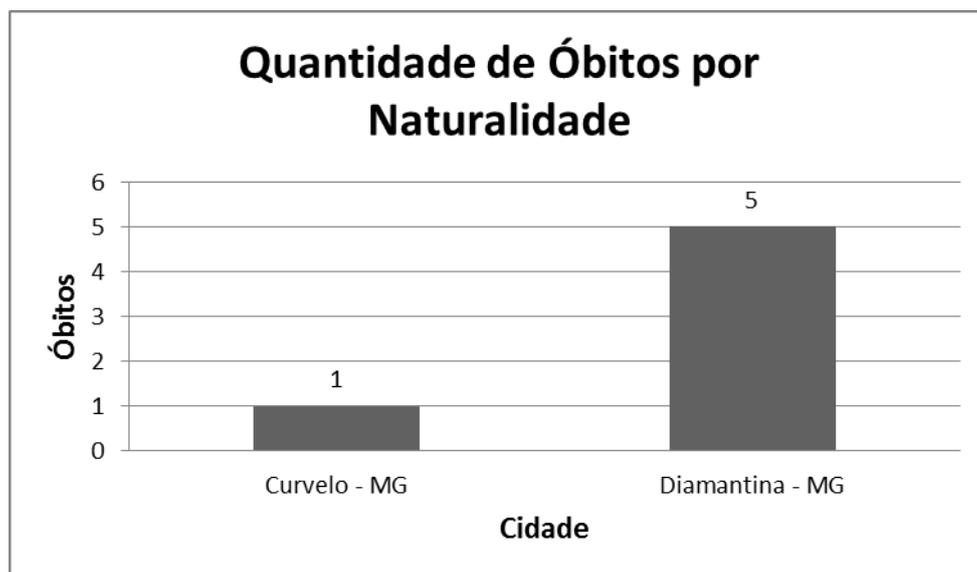


Figura 116. Quantidade de óbitos por naturalidade no *cluster* 17.

A Figura 116 mostra a quantidade de óbitos por naturalidade no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças falecidas naturais de Diamantina-MG, com cinco registros, sendo equivalente a 83% dos seis registros agrupados no *Cluster*.

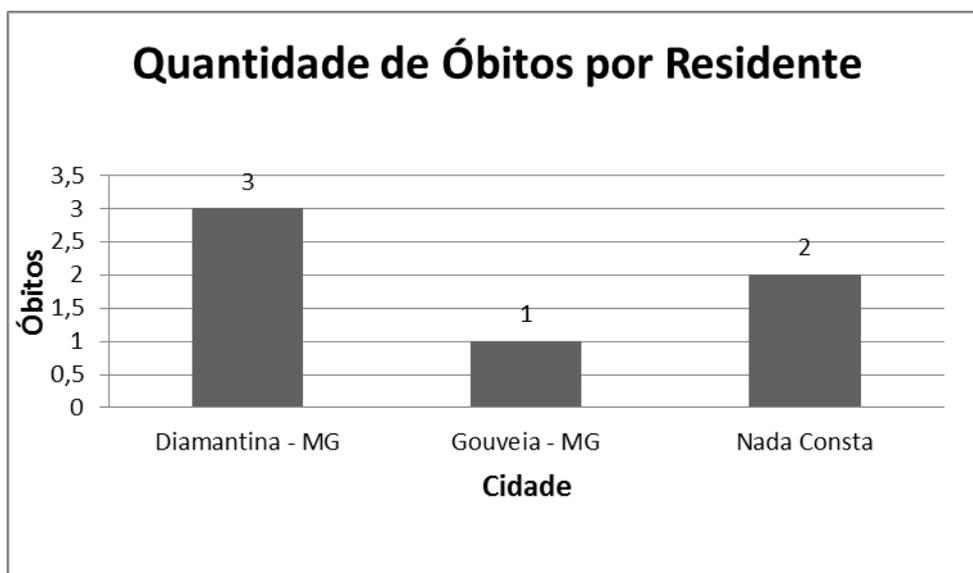


Figura 117. Quantidade de óbitos por residente no *cluster* 17.

A Figura 117 mostra a quantidade de óbitos por residente no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças

falecidas que eram residentes em Diamantina-MG, com três registros, sendo equivalente a 50% dos seis registros agrupados no *Cluster*.

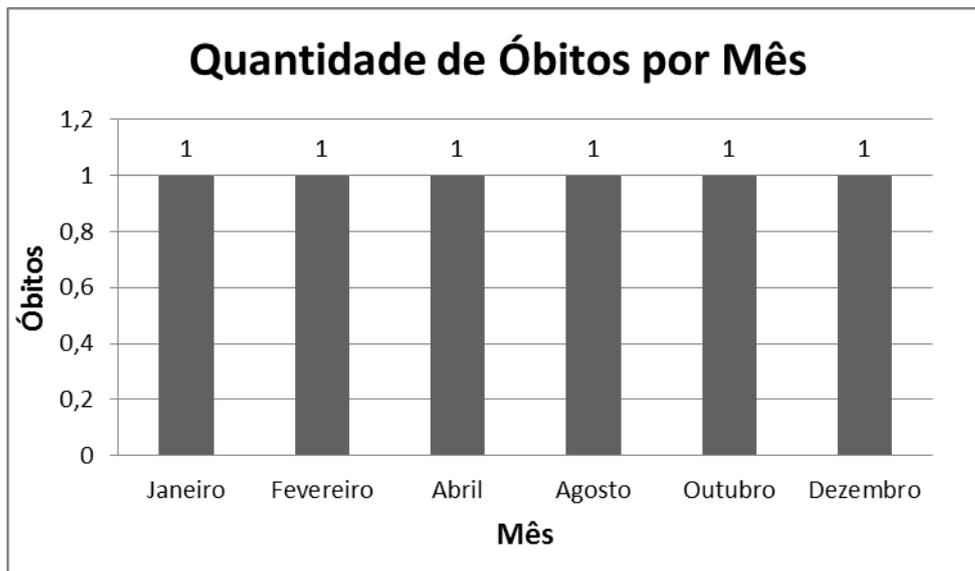


Figura 118. Quantidade de óbitos por mês no *cluster* 17.

A Figura 118 mostra a quantidade de óbitos por mês no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um mês diferente, sendo um total de seis registros.

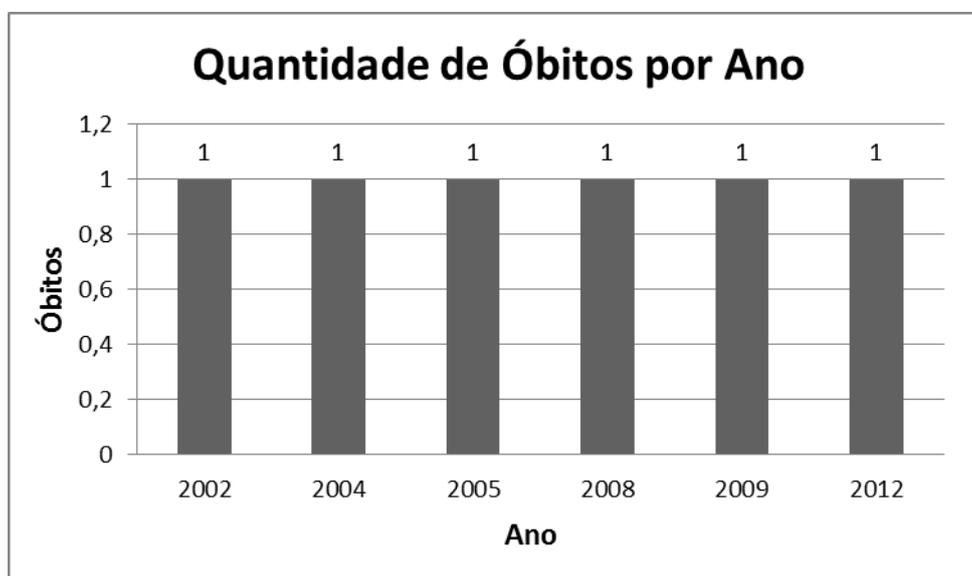


Figura 119. Quantidade de óbitos por ano no *cluster* 17.

A Figura 119 mostra a quantidade de óbitos por ano no *Cluster* 17, obtidos pelo modelo Cassiopeia. Nesta Figura cada criança faleceu em um ano diferente, sendo um total de seis registros.

Cluster 18:

Número de Registros: 02

Tabela:

Tabela 32. *Cluster 18* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|-----------------|--------------|---------------|-----------|-----------|
| 1 | 01 2003 | choque misto, hemorragia digestiva alta, diabetes mellitus descompensado | Feminino | 16 anos | Corinto MG | Diamantina MG | Dezembro | 2003 |
| 2 | 08 2006 | choque hipovolêmico, hemorragia digestiva alta, sepse neonatal lardia | Masculino | 15 dias (0 ano) | Carbonita MG | Nada Consta | Setembro | 2006 |

Na Tabela 32 é apresentado o *Cluster 18*, com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

Gráficos:

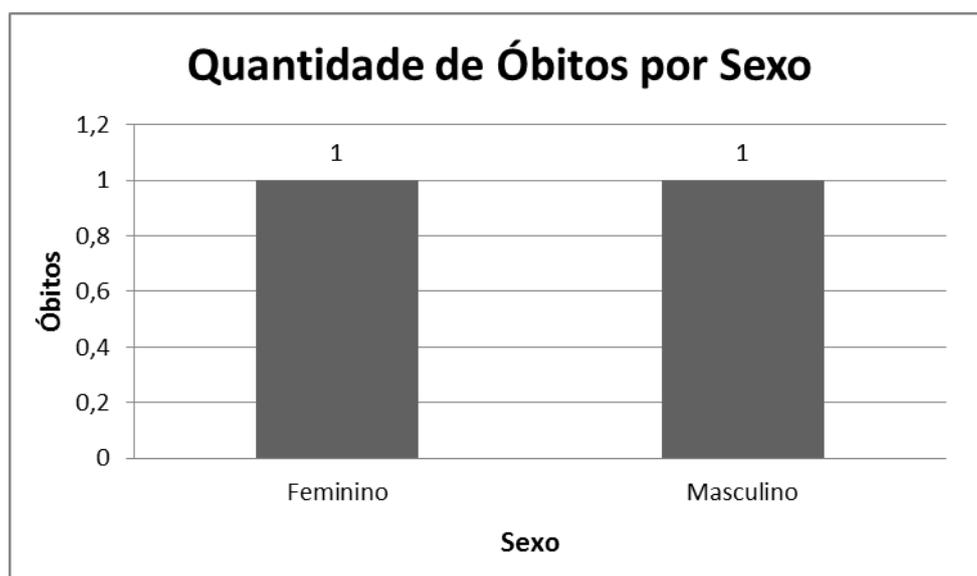


Figura 120. Quantidade de óbitos por sexo no *cluster* 18.

A Figura 120 mostra a quantidade de óbitos por sexo no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* eram de sexos opostos.

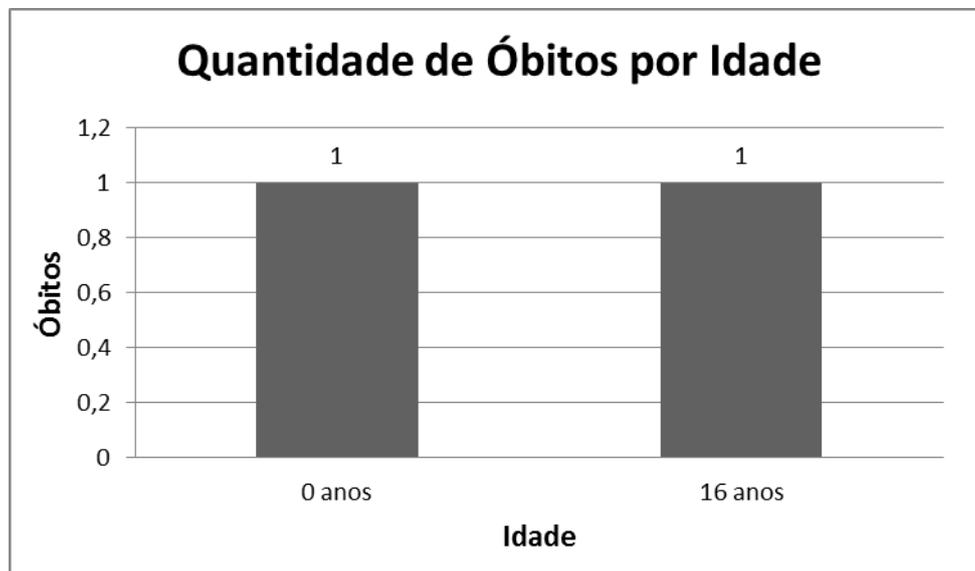


Figura 121. Quantidade de óbitos por idade no *cluster* 18.

A Figura 121 mostra a quantidade de óbitos por idade no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* possuíam idades diferentes.

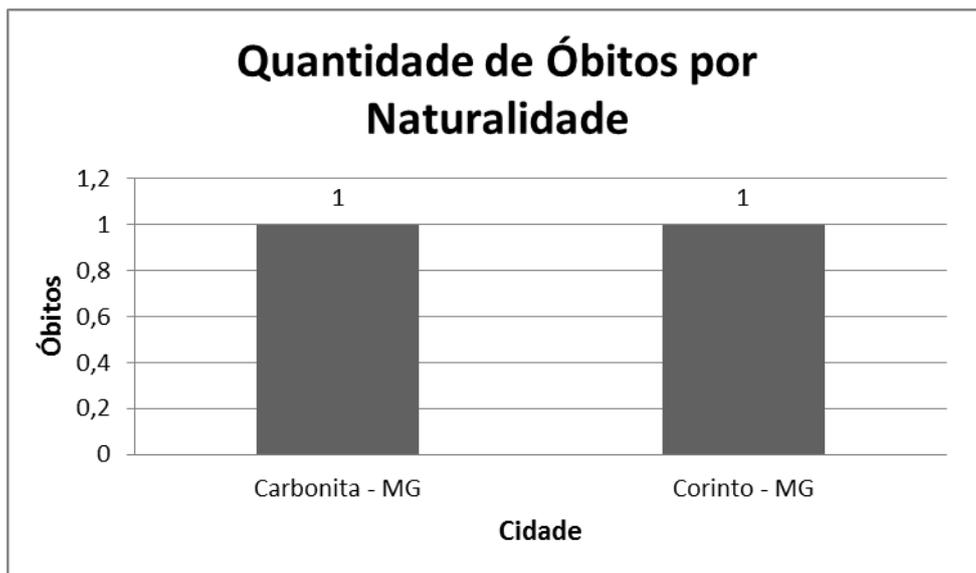


Figura 122. Quantidade de óbitos por naturalidade no *cluster* 18.

A Figura 122 mostra a quantidade de óbitos por naturalidade no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* eram naturais de cidades diferentes.

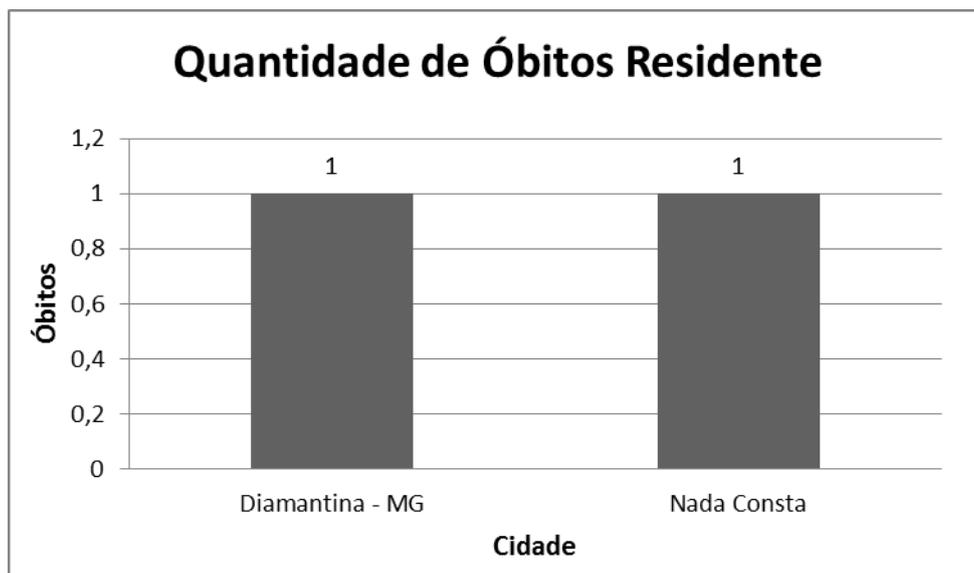


Figura 123. Quantidade de óbitos por residente no *cluster* 18.

A Figura 123 mostra a quantidade de óbitos por residente no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* residiam em cidades diferentes.

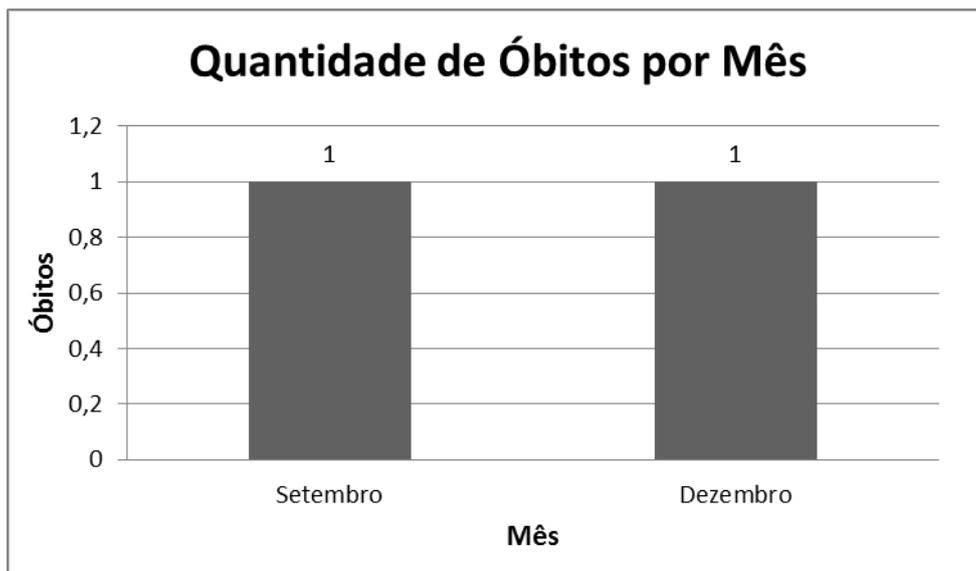


Figura 124. Quantidade de óbitos por mês no *cluster* 18.

A Figura 124 mostra a quantidade de óbitos por mês no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em meses diferentes.

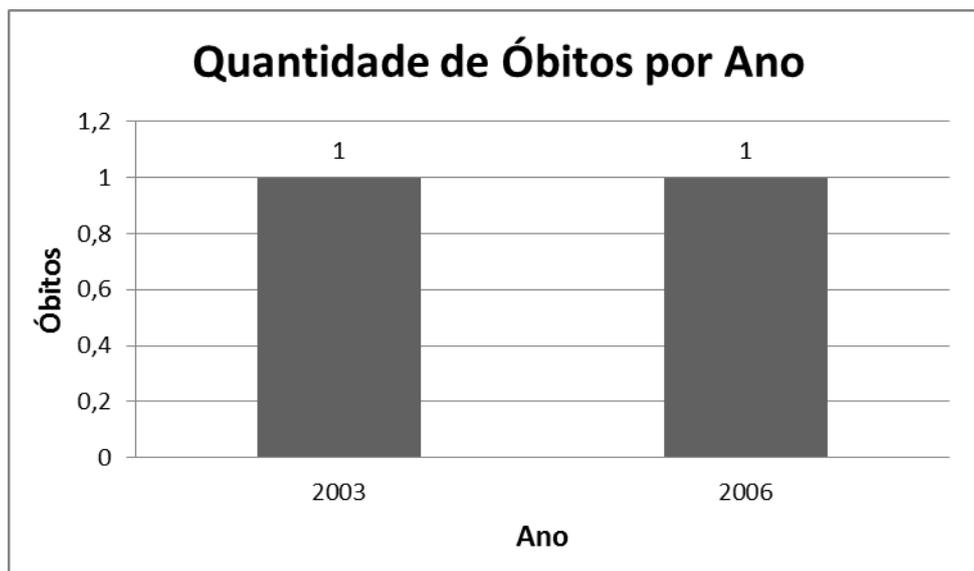


Figura 125. Quantidade de óbitos por ano no *cluster* 18.

A Figura 125 mostra a quantidade de óbitos por ano no *Cluster* 18, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em anos diferentes.

Cluster 19:

Número de Registros: 01

Tabela:

Tabela 33. *Cluster* 19 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|----------|-------------------|------------------|----------------|-----------|-----------|
| 1 | 06 2009 | choque séptico, sepse precoce, prematuridade | Feminino | 4 dias (0 ano) | Diamantina MG | Nada Consta | Junho | 2009 |

Na Tabela 33 é apresentado o *Cluster* 19, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de só haver um registro no agrupamento.

Gráficos:

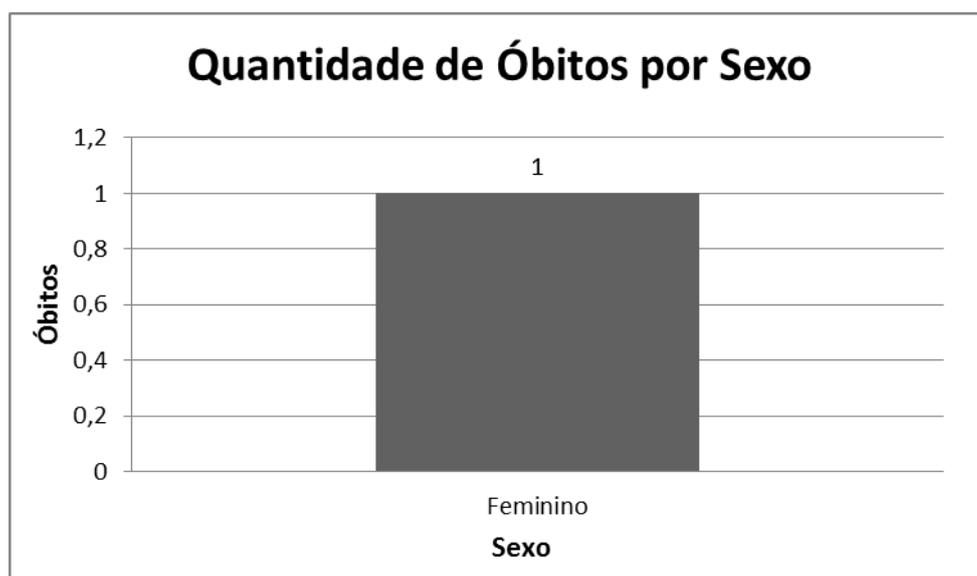


Figura 126. Quantidade de óbitos por sexo no *cluster* 19.

A Figura 126 mostra a quantidade de óbitos por sexo no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo feminino.

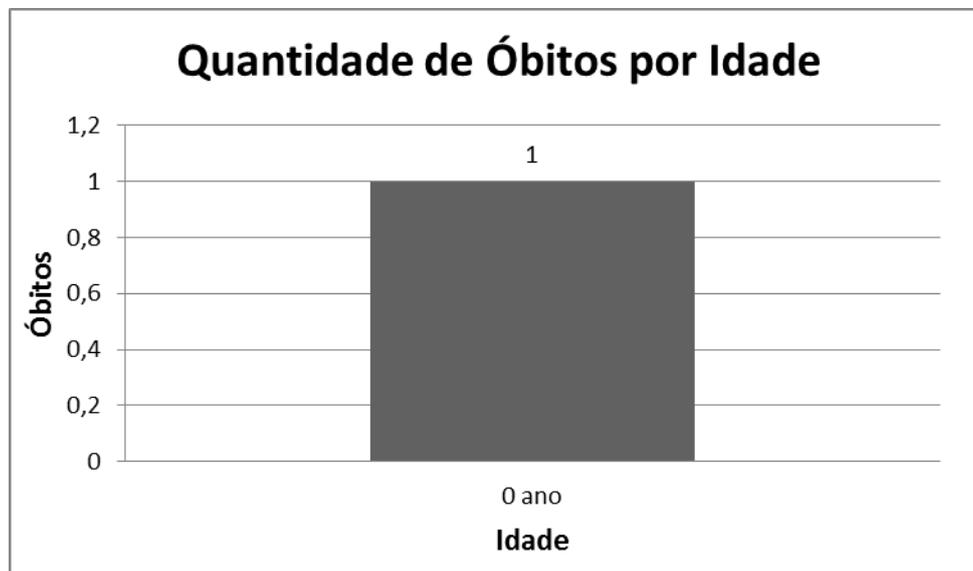


Figura 127. Quantidade de óbitos por idade no *cluster* 19.

A Figura 127 mostra a quantidade de óbitos por idade no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía idade abaixo de um ano.



Figura 128. Quantidade de óbitos por naturalidade no *cluster* 19.

A Figura 128 mostra a quantidade de óbitos por naturalidade no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural de Diamantina-MG.

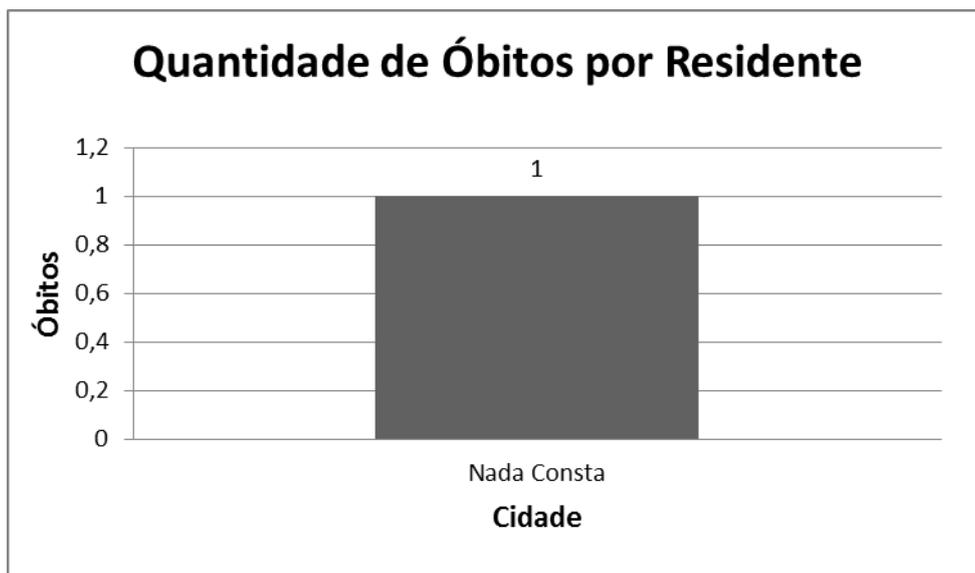


Figura 129. Quantidade de óbitos por residente no *cluster* 19.

A Figura 129 mostra a quantidade de óbitos por residente no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* não possuía informação de onde residia.

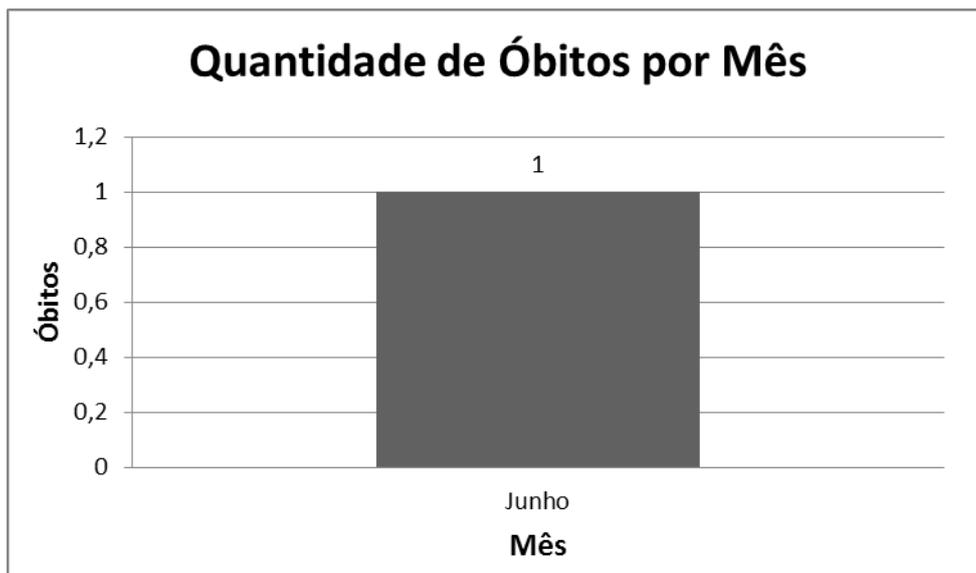


Figura 130. Quantidade de óbitos por mês no *cluster* 19.

A Figura 130 mostra a quantidade de óbitos por mês no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de junho.

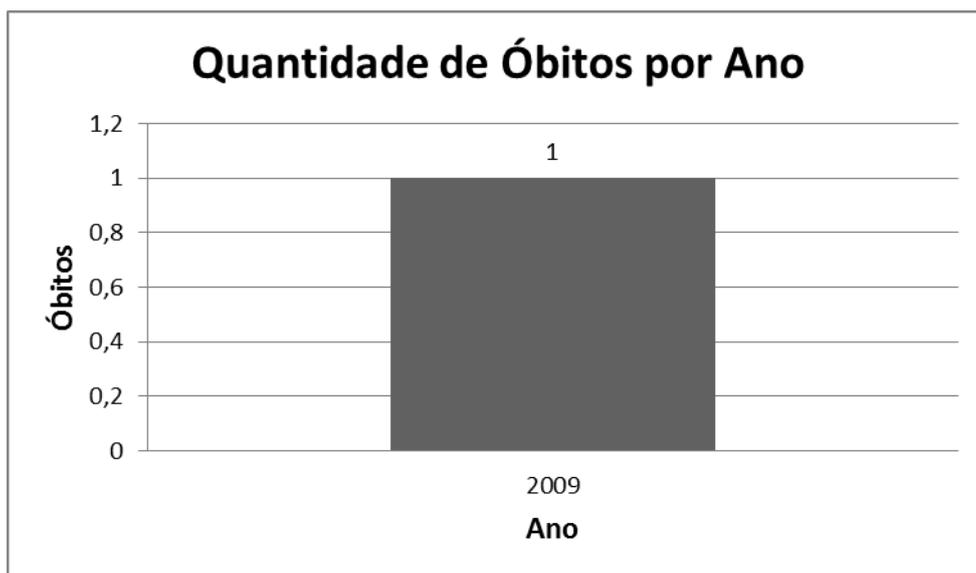


Figura 131. Quantidade de óbitos por ano no *cluster* 19.

A Figura 131 mostra a quantidade de óbitos por ano no *Cluster* 19, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2009.

Cluster 20:

Número de Registros: 01

Tabela:

Tabela 34. *Cluster 20* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---------------------------------|-----------|-----------|-----------------|-----------------|---------------|-----------|
| 1 | 15 2009 | sepsis, meningite bacteriana | Masculino | 5 anos | São Paulo SP | Carbonita MG | Novem- bro | 2009 |

Na Tabela 34 é apresentado o *Cluster 20*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de só haver um registro no agrupamento.

Gráficos:

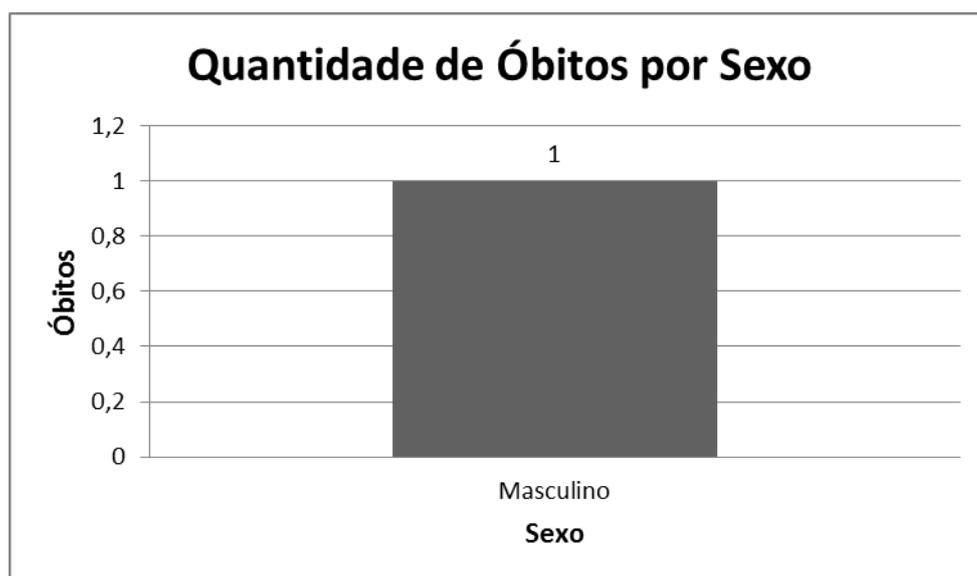


Figura 132. Quantidade de óbitos por sexo no *cluster* 20.

A Figura 132 mostra a quantidade de óbitos por sexo no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo masculino.

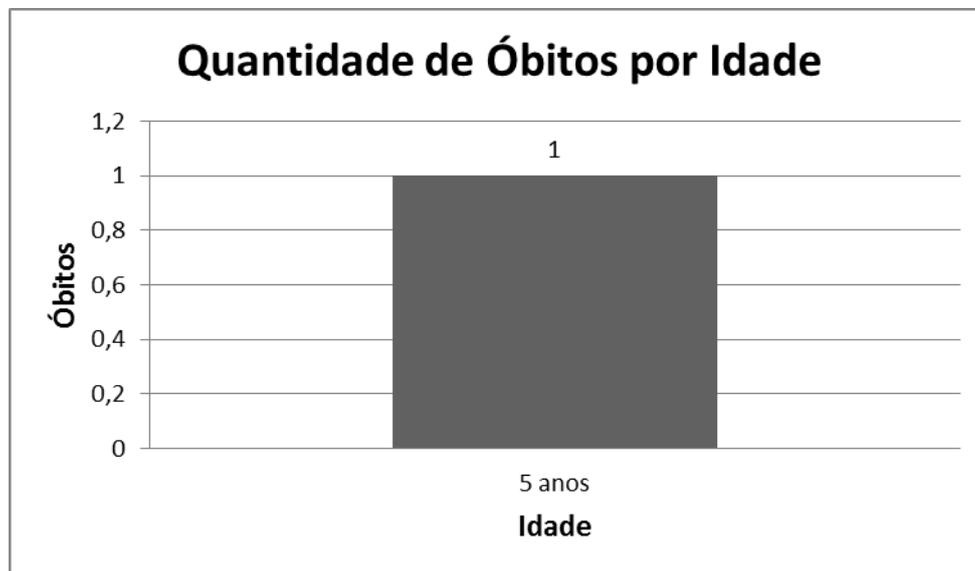


Figura 133. Quantidade de óbitos por idade no *cluster* 20.

A Figura 133 mostra a quantidade de óbitos por idade no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía cinco anos de idade.

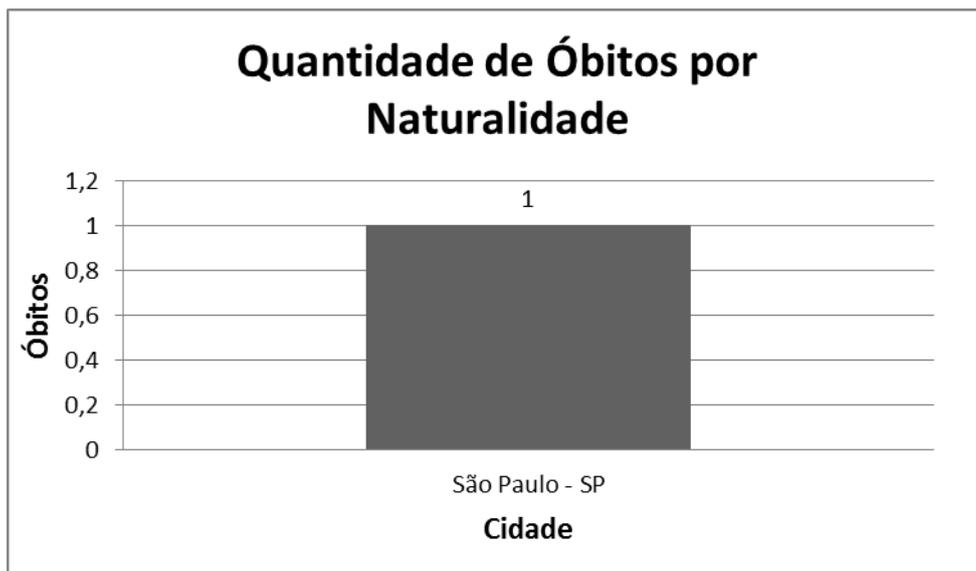


Figura 134. Quantidade de óbitos por naturalidade no *cluster* 20.

A Figura 134 mostra a quantidade de óbitos por naturalidade no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* era natural de São Paulo-SP.

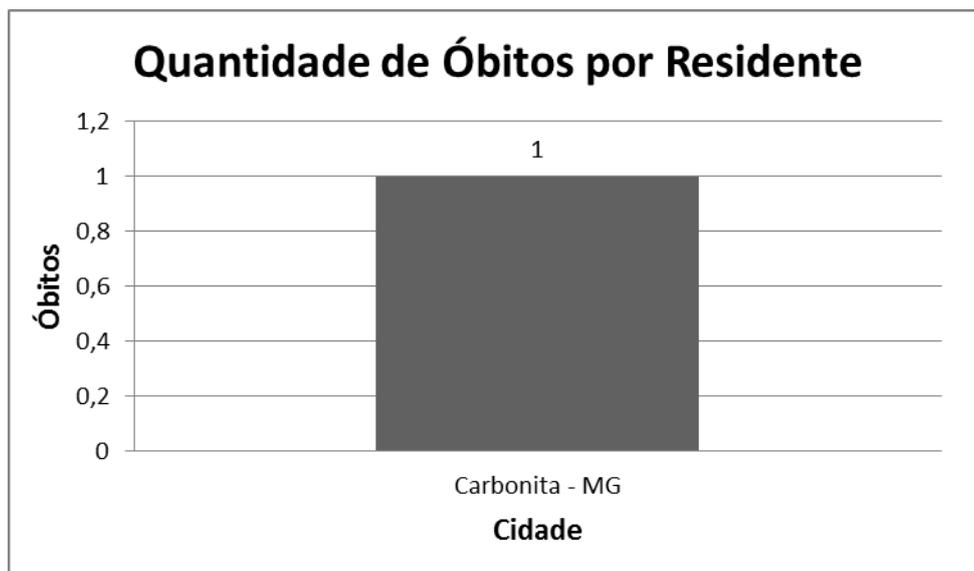


Figura 135. Quantidade de óbitos por residente no *cluster* 20.

A Figura 135 mostra a quantidade de óbitos por residente no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* residia em Carbonita-MG.

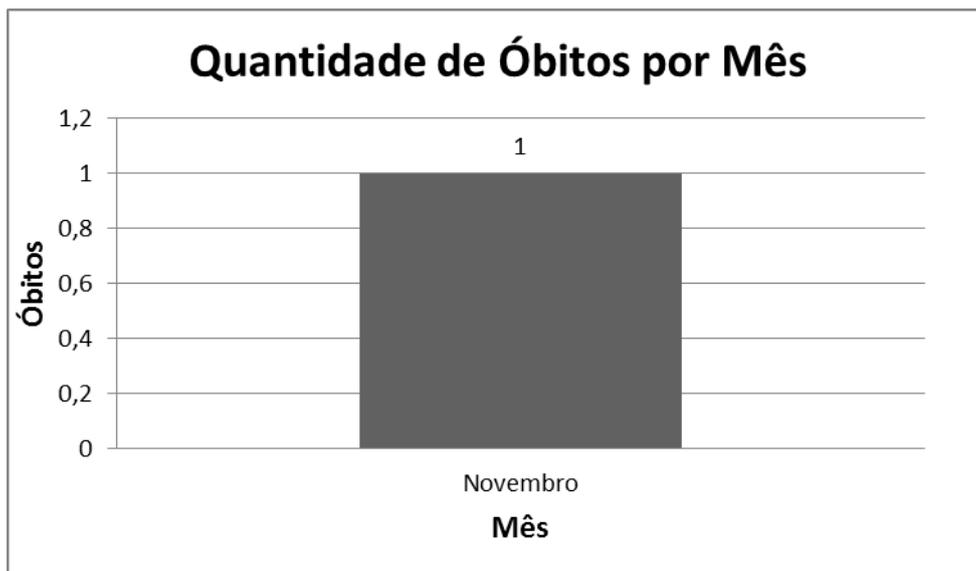


Figura 136. Quantidade de óbitos por mês no *cluster* 20.

A Figura 136 mostra a quantidade de óbitos por mês no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de novembro.

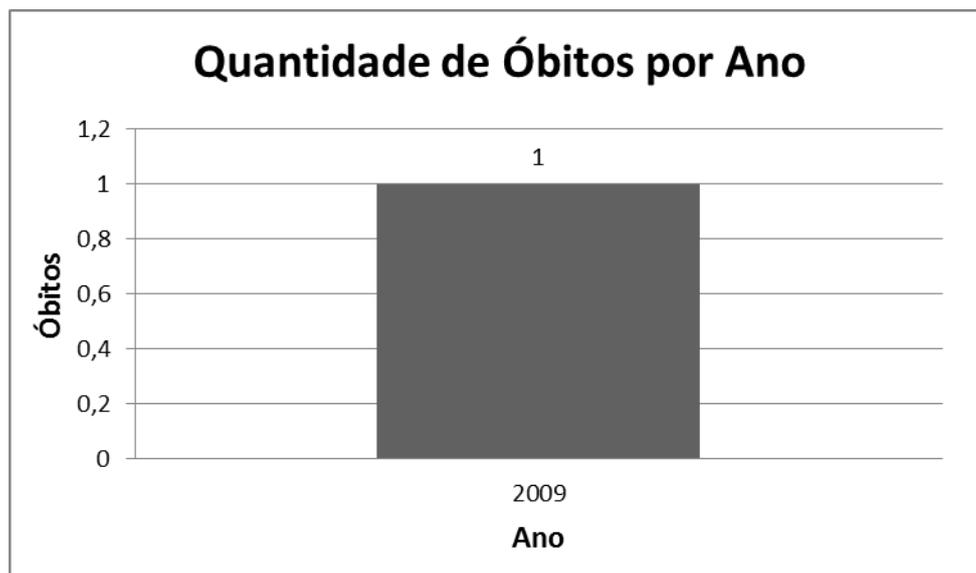


Figura 137. Quantidade de óbitos por ano no *cluster* 20.

A Figura 137 mostra a quantidade de óbitos por ano no *Cluster* 20, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2009.

Cluster 21:

Número de Registros: 06

Tabela:

Tabela 35. Cluster 21 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|--------------------|------------------|---|-----------|-----------|
| 1 | 04 2004 | causa indeterminada | Masculino | 3 anos | Diamantina MG | Diamantina MG | Julho | 2004 |
| 2 | 17 2009 | óbito de causa indeterminada | Masculino | 15 anos | Diamantina MG | Ribeirão de Areia, Diamantina MG | Dezembro | 2009 |
| 3 | 18 2012 | causa indeterminada, hidrocefalia, paralisia cerebral | Masculino | 12 anos | Diamantina MG | Diamantina MG | Julho | 2012 |
| 4 | 06 2006 | indeterminada | Masculino | 1 mês (0 ano) | Diamantina MG | Nada Consta | Junho | 2006 |
| 5 | 01 2004 | indeterminada | Masculino | 3 meses (0 ano) | Diamantina MG | Nada Consta | Janeiro | 2004 |
| 6 | 09 2011 | indeterminada, epilepsia | Masculino | 16 anos | Diamantina MG | Diamantina MG | Novembro | 2011 |

Na Tabela 35 é apresentado o *Cluster 21*, com seis registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

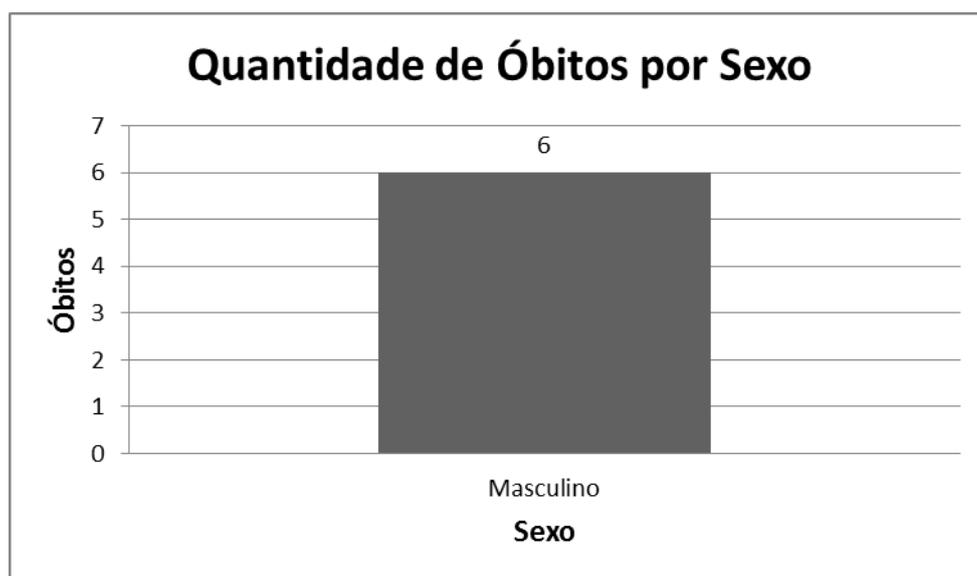


Figura 138. Quantidade de óbitos por sexo no *cluster* 21.

A Figura 138 mostra a quantidade de óbitos por sexo no *Cluster* 21, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram do sexo masculino, com seis registros agrupados no *Cluster*.

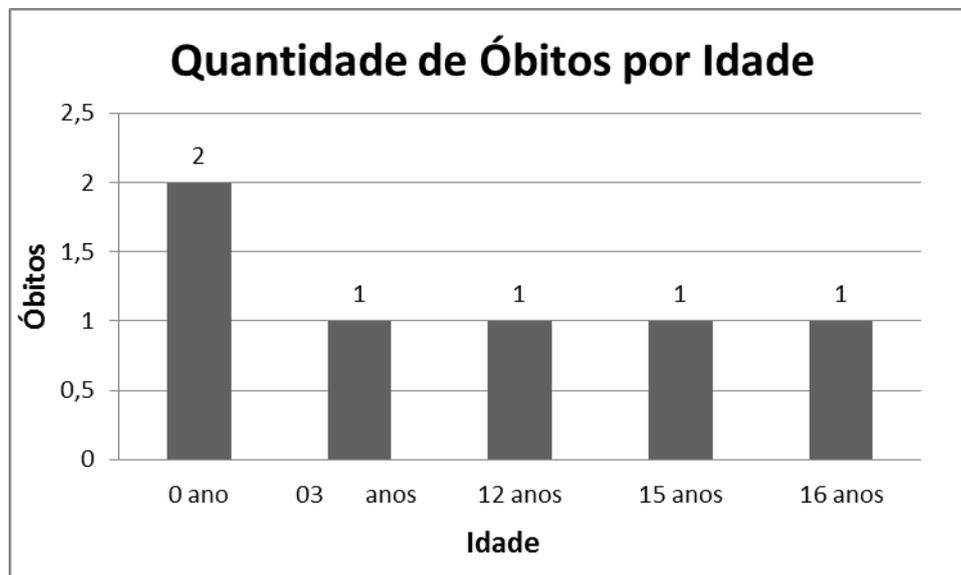


Figura 139. Quantidade de óbitos por idade no *cluster* 21.

A Figura 139 mostra a quantidade de óbitos por idade no *Cluster* 21, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos na faixa de idade abaixo de um ano, com dois registros, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

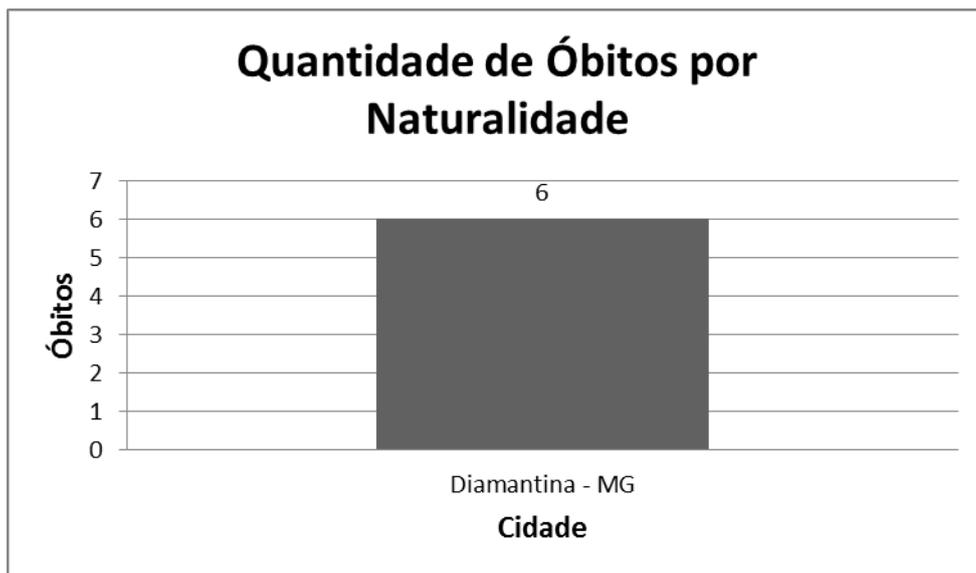


Figura 140. Quantidade de óbitos por naturalidade no *cluster* 21.

A Figura 140 mostra a quantidade de óbitos por naturalidade no *Cluster* 21, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram naturais de Diamantina-MG, com seis registros agrupados no *Cluster*.

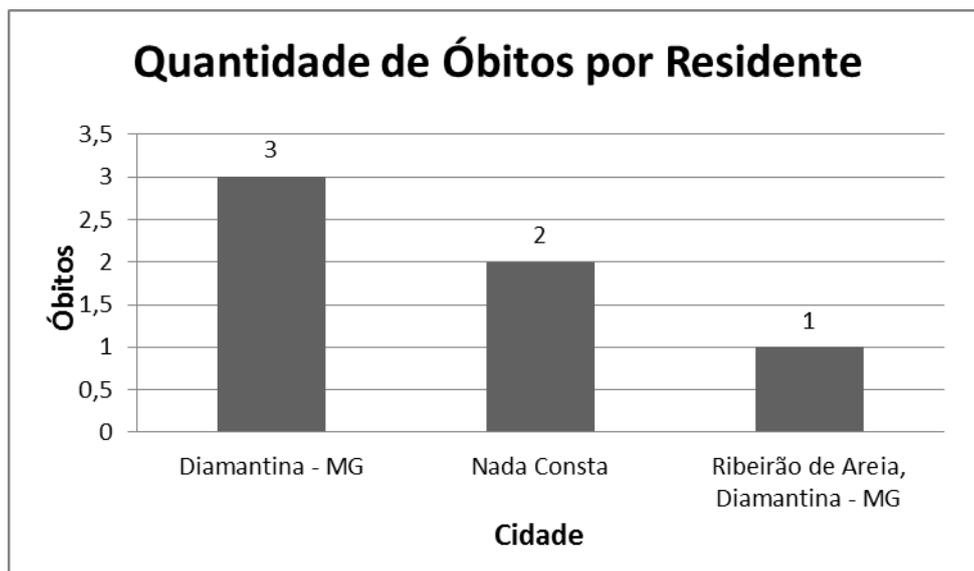


Figura 141. Quantidade de óbitos por residente no *cluster* 21.

A Figura 141 mostra a quantidade de óbitos por residente no *Cluster* 21, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de crianças falecidas que residiam em Diamantina-MG, com três registros, sendo equivalente a 50% dos seis registros agrupados no *Cluster*.

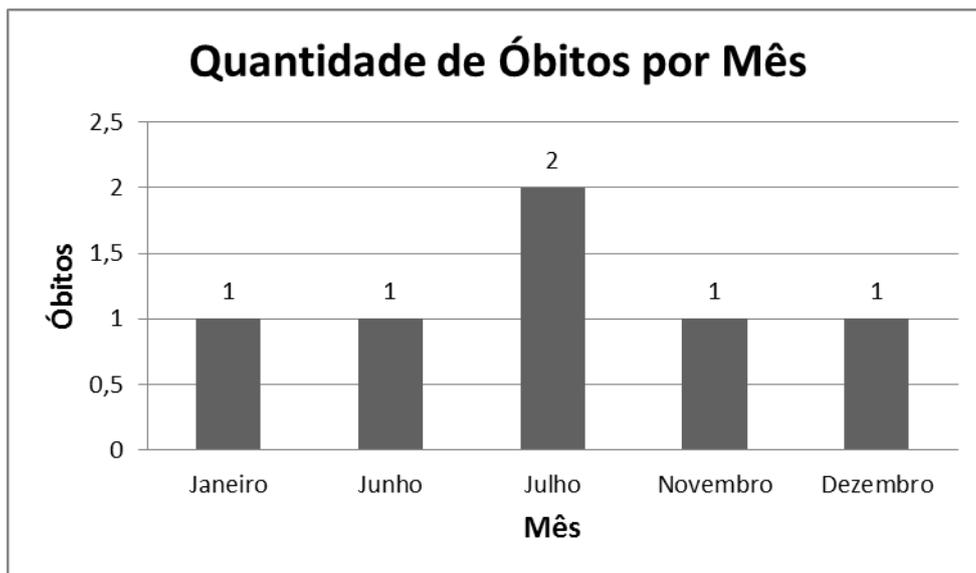


Figura 142. Quantidade de óbitos por mês no *cluster 21*.

A Figura 142 mostra a quantidade de óbitos por mês no *Cluster 21*, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no mês de julho, com dois registros, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

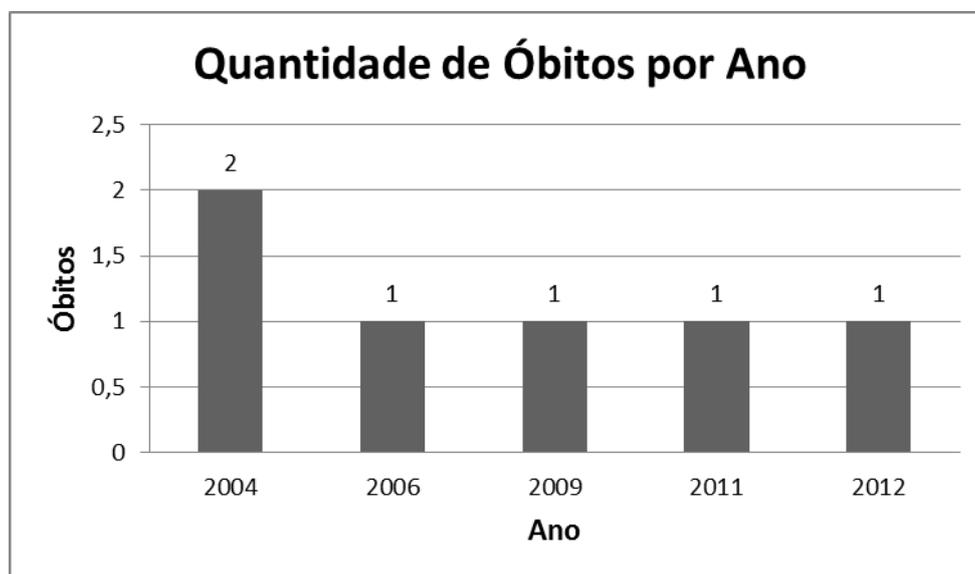


Figura 143. Quantidade de óbitos por ano no *cluster 21*.

A Figura 143 mostra a quantidade de óbitos por ano no *Cluster 21*, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos que ocorreram no ano de 2004, com dois registros, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

Cluster 22:

Número de Registros: 01

Tabela:

Tabela 36. *Cluster 22* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|---------------|---------------------------|------------------|----------------|-----------|-----------|
| 1 | 05 2005 | sofrimento fetal, más formações múltiplas | Femi- nino | 2, 20 horas (0 ano) | Diamantina MG | Nada Consta | Julho | 2005 |

Na Tabela 36 é apresentado o *Cluster 22*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de só haver um registro no agrupamento.

Gráficos:

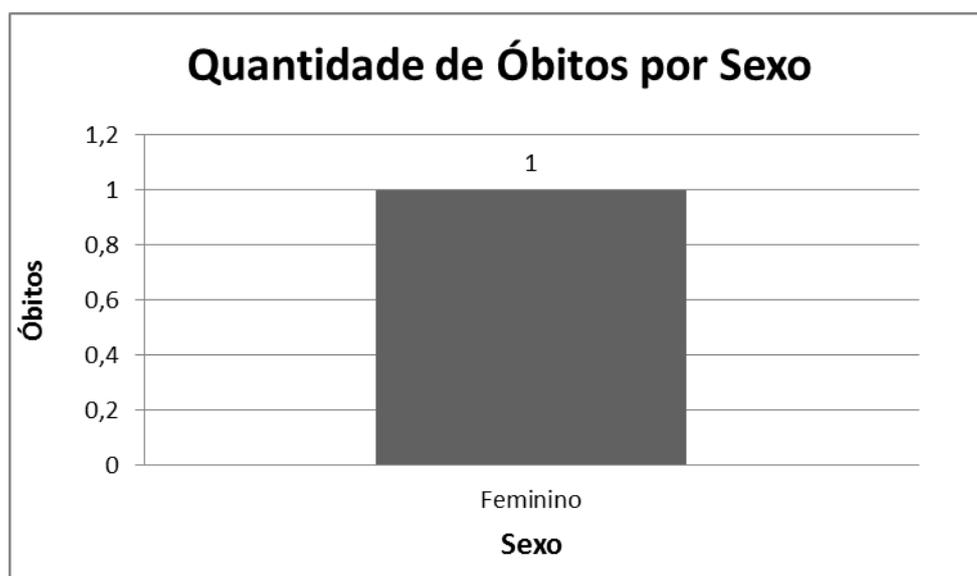


Figura 144. Quantidade de óbitos por sexo no *cluster 22*.

A Figura 144 mostra a quantidade de óbitos por sexo no *Cluster 22*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo feminino.

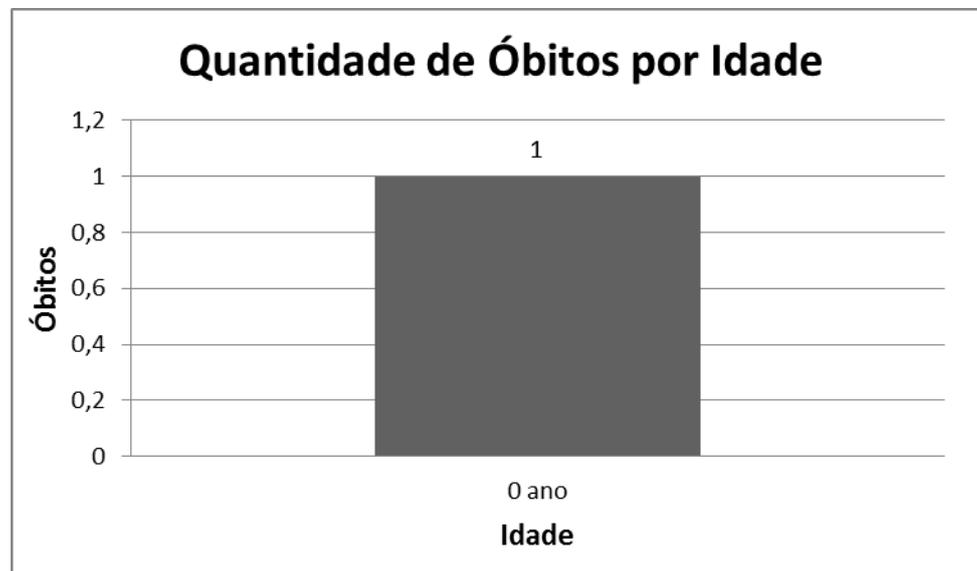


Figura 145. Quantidade de óbitos por idade no *cluster 22*.

A Figura 145 mostra a quantidade de óbitos por idade no *Cluster 22*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía faixa de idade abaixo de um ano.

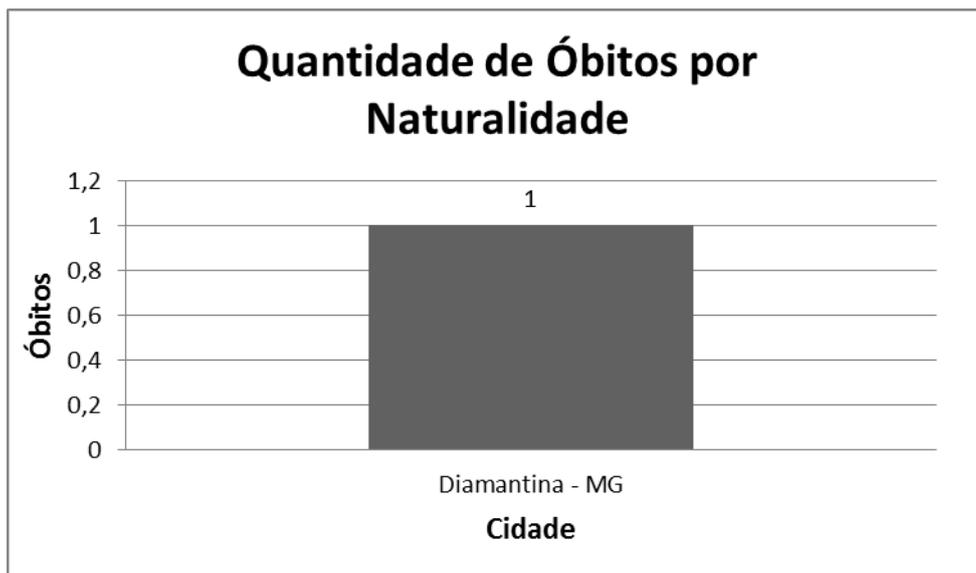


Figura 146. Quantidade de óbitos por naturalidade no *cluster* 22.

A Figura 146 mostra a quantidade de óbitos por naturalidade no *Cluster* 22, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural de Diamantina-MG.

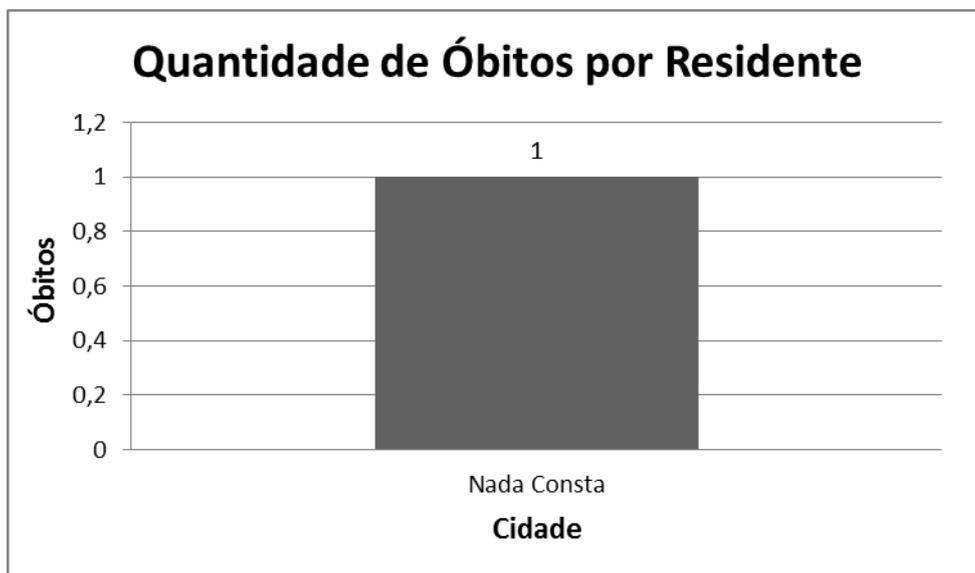


Figura 147. Quantidade de óbitos por residente no *cluster* 22.

A Figura 147 mostra a quantidade de óbitos por residente no *Cluster* 22, obtidos pelo modelo Cassiopeia. Nesta Figura, o único registro de óbito do *Cluster* não continha a informação de onde a criança residia.

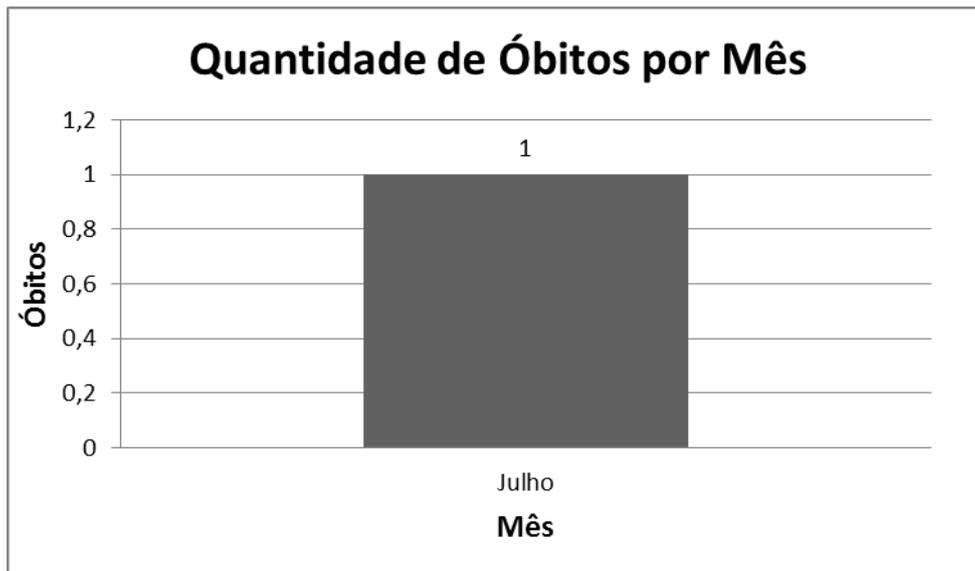


Figura 148. Quantidade de óbitos por mês no *cluster 22*.

A Figura 148 mostra a quantidade de óbitos por mês no *Cluster 22*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de julho.

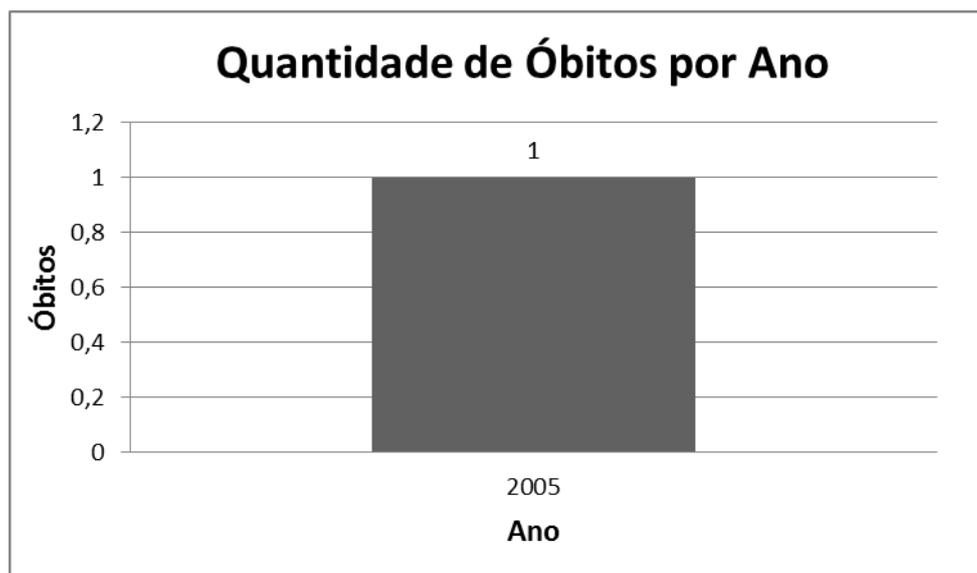


Figura 149. Quantidade de óbitos por ano no *cluster 22*.

A Figura 149 mostra a quantidade de óbitos por ano no *Cluster 22*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2005.

Cluster 23:

Número de Registros: 02

Tabela:

Tabela 37. *Cluster 23* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|----------|--------------------|--------------------|-----------------------------|-----------|-----------|
| 1 | 12 2011 | encefalopatia hipóxico-isquêmica, sofrimento fetal agudo, sepse precoce | Feminino | 1 dia (0 ano) | Diamantina MG | Felício dos Santos MG | Fevereiro | 2011 |
| 2 | 11 2011 | encefalopatia hipóxico isquêmica, síndrome de West, distúrbio metabólico, BCP aspiração | Feminino | 9 meses (0 ano) | Itamarandiba MG | Carbonita MG | Dezembro | 2011 |

Na Tabela 37 é apresentado o *Cluster 23*, com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

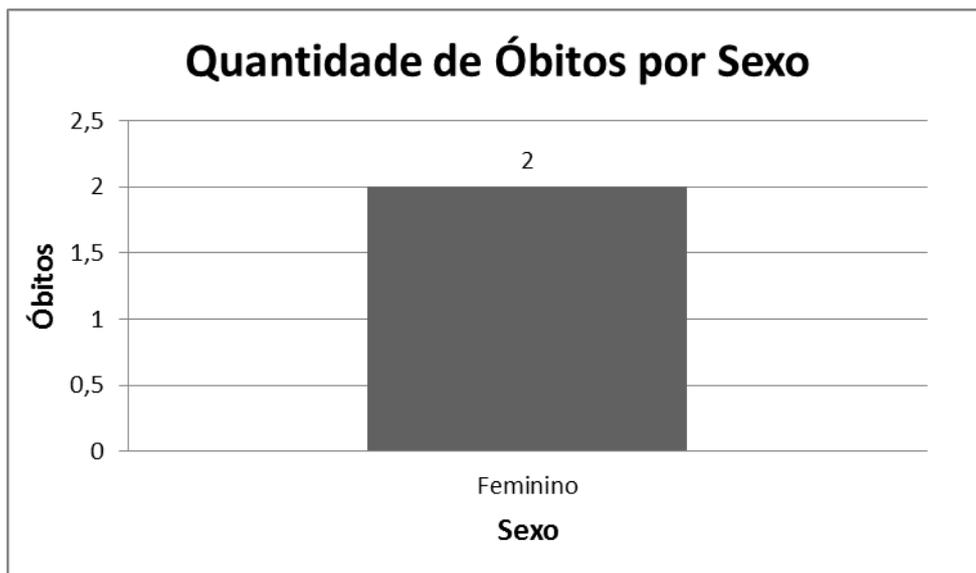


Figura 150. Quantidade de óbitos por sexo no *cluster* 23.

A Figura 150 mostra a quantidade de óbitos por sexo no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram do sexo feminino, com dois registros agrupados no *Cluster*.

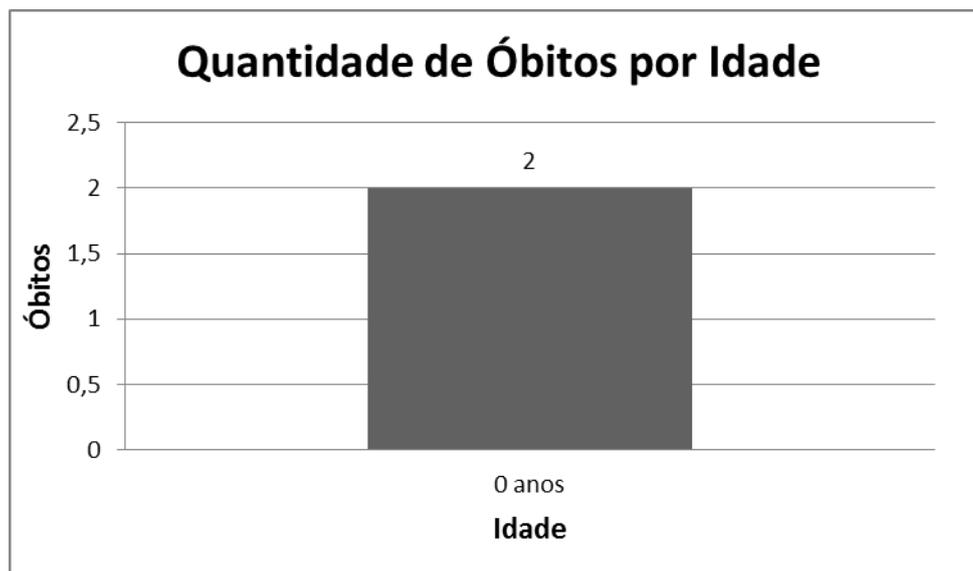


Figura 151. Quantidade de óbitos por idade no *cluster* 23.

A Figura 151 mostra a quantidade de óbitos por idade no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças possuíam idade abaixo de 1 ano, com dois registros agrupados no *Cluster*.

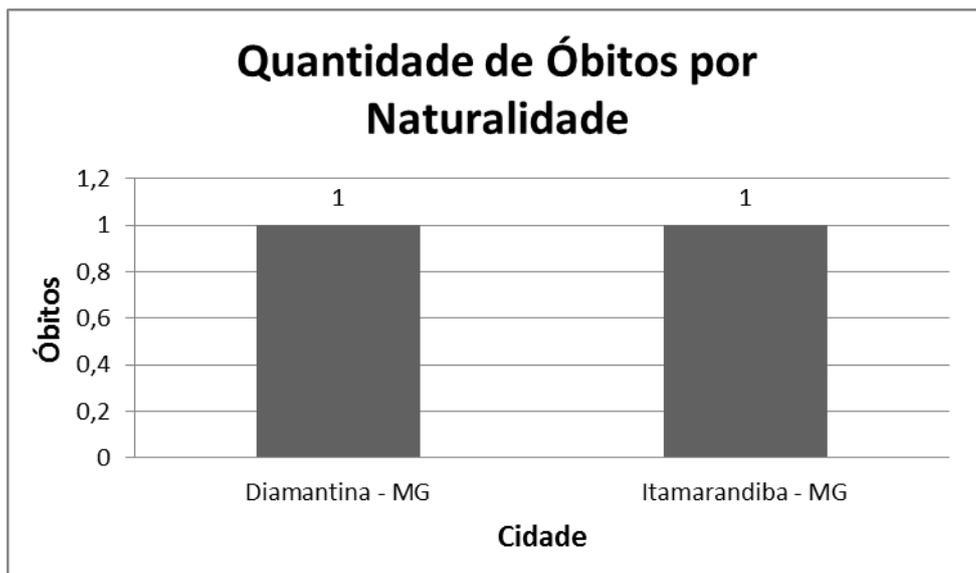


Figura 152. Quantidade de óbitos por naturalidade no *cluster* 23.

A Figura 152 mostra a quantidade de óbitos por naturalidade no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* nasceram em cidades diferentes.

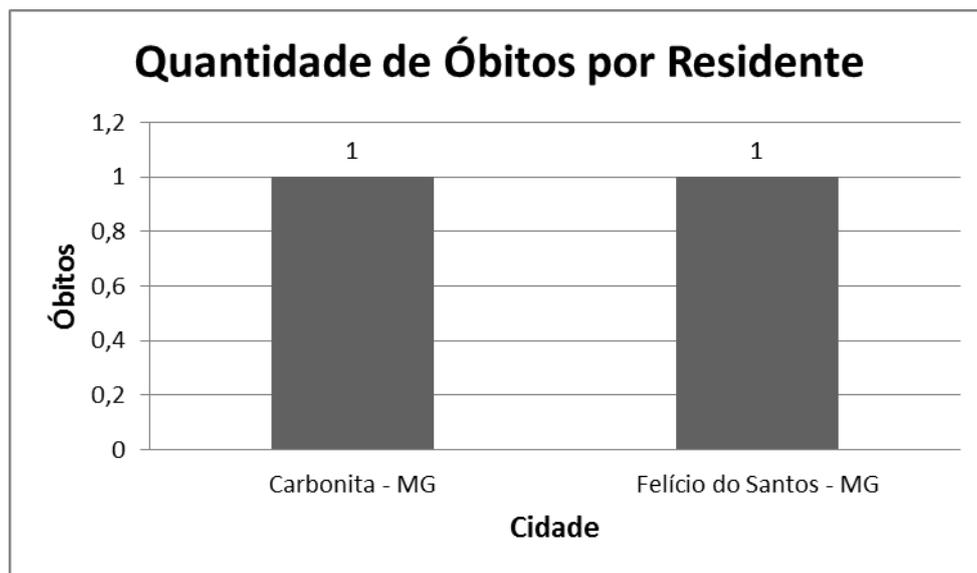


Figura 153. Quantidade de óbitos por residente no *cluster* 23.

A Figura 153 mostra a quantidade de óbitos por residente no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* residiam em cidades diferentes.

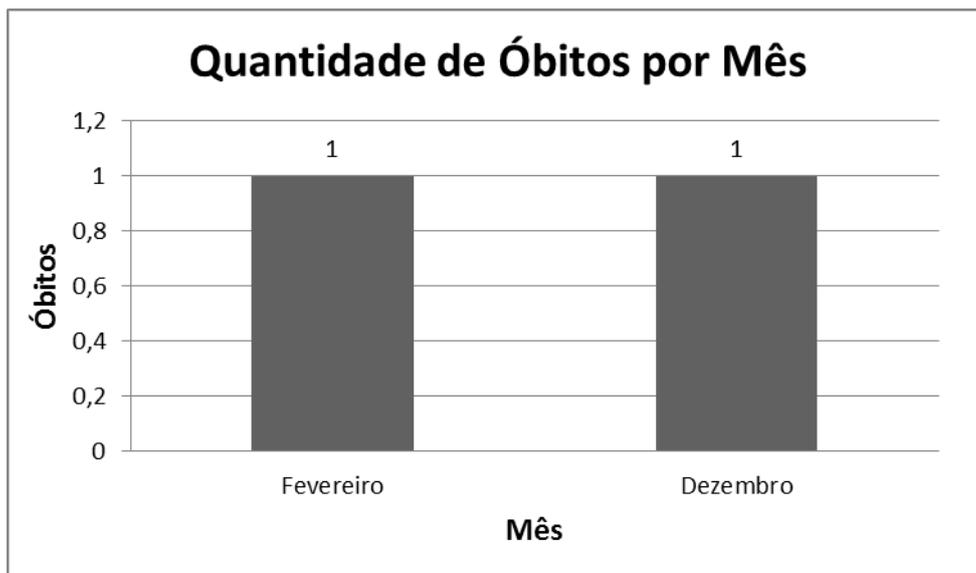


Figura 154. Quantidade de óbitos por mês no *cluster* 23.

A Figura 154 mostra a quantidade de óbitos por mês no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em meses diferentes.

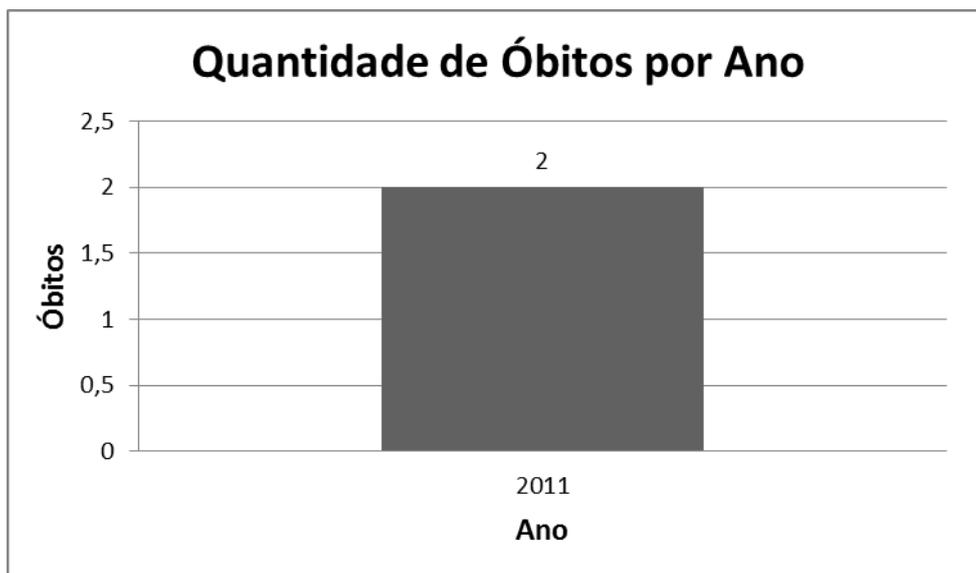


Figura 155. Quantidade de óbitos por ano no *cluster* 23.

A Figura 155 mostra a quantidade de óbitos por ano no *Cluster* 23, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças faleceram no ano de 2011, com dois registros agrupados no *Cluster*.

Cluster 24:

Número de Registros: 06

Tabela:

Tabela 38. Cluster 24 gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|-----------|---------------|---------------|---------------|-----------|-----------|
| 1 | 03 2007 | asfixia mecânica por obstrução das vias aéreas por meio líquido (afogamento) | Feminino | 12 anos | Diamantina MG | Diamantina MG | Abril | 2007 |
| 2 | 02 2007 | asfixia mecânica por obstrução das vias aéreas por meio líquido (afogamento) | Feminino | 10 anos | Diamantina MG | Diamantina MG | Abril | 2007 |
| 3 | 07 2004 | afogamento | Masculino | 14 anos | São Paulo SP | Serro MG | Dezembro | 2004 |
| 4 | 13 2002 | síndrome de aspiração de mecônio, sofrimento fetal crônico | Masculino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Dezembro | 2002 |
| 5 | 06 2010 | asfixia por sufocação direta | Masculino | 4 anos | Diamantina MG | Diamantina MG | Fevereiro | 2010 |
| 6 | 08 2008 | asfixia por soterramento | Masculino | 6 anos | Diamantina MG | Diamantina MG | Agosto | 2008 |

Na Tabela 38 é apresentado o *Cluster 24*, com seis registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 4% dos 147 registros processados.

Gráficos:

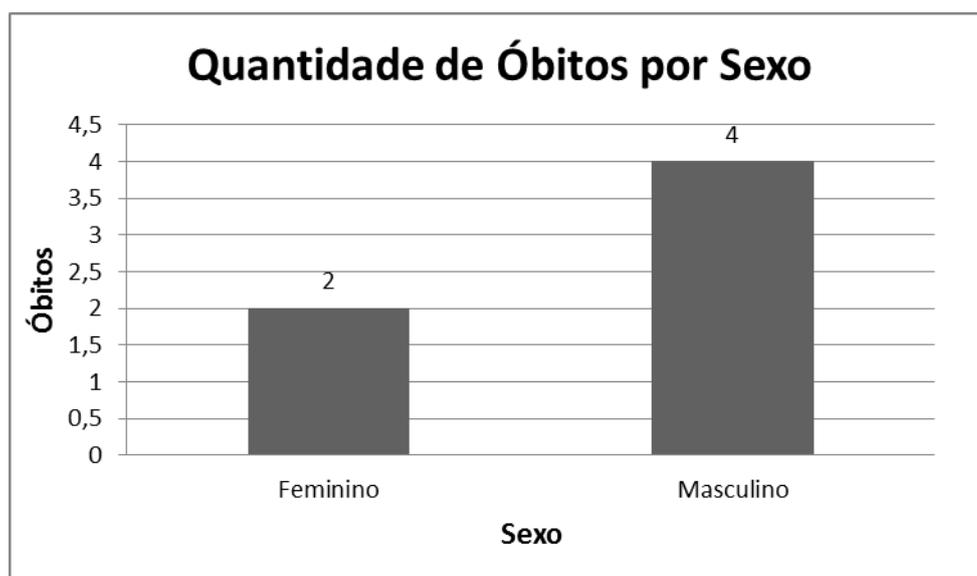


Figura 156. Quantidade de óbitos por sexo no *cluster* 24.

A Figura 156 mostra a quantidade de óbitos por sexo no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos do sexo masculino, com quatro registros, sendo equivalente a 66% dos seis registros agrupados no *Cluster*.

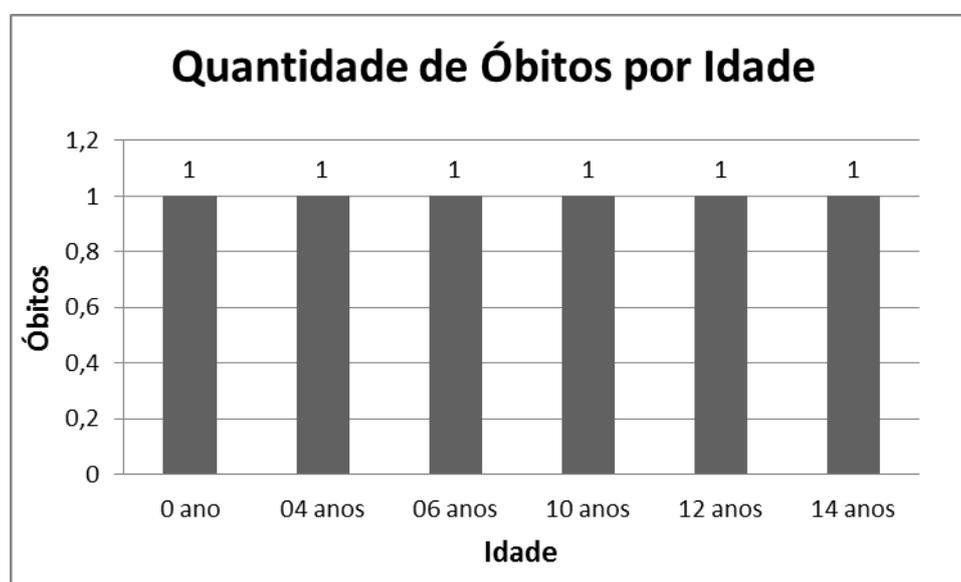


Figura 157. Quantidade de óbitos por idade no *cluster* 24.

A Figura 157 mostra a quantidade de óbitos por idade no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura as seis crianças do *Cluster* faleceram com idades diferentes.

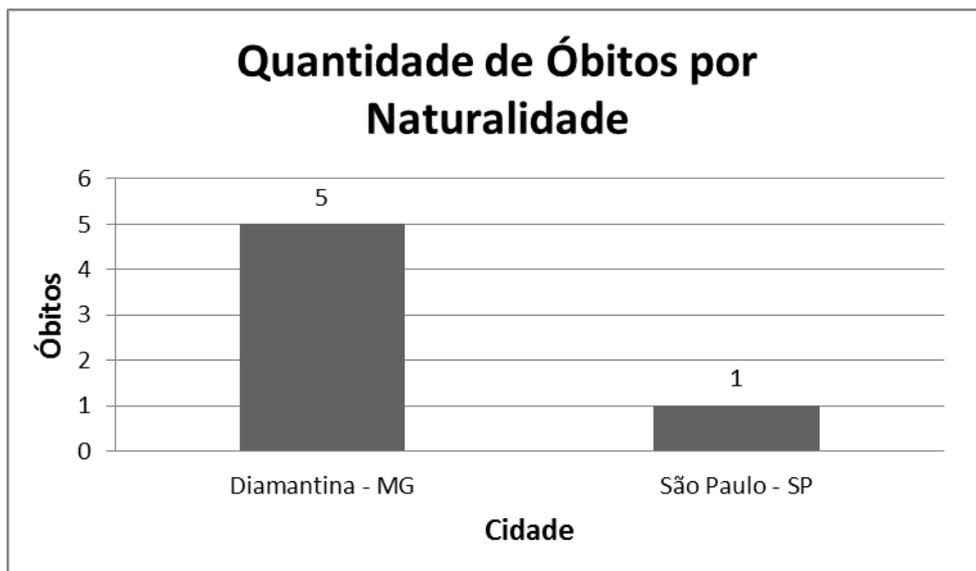


Figura 158. Quantidade de óbitos por naturalidade no *cluster* 24.

A Figura 158 mostra a quantidade de óbitos por naturalidade no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças naturais de Diamantina-MG, com cinco registros, sendo equivalente a 83% dos seis registros agrupados no *Cluster*.

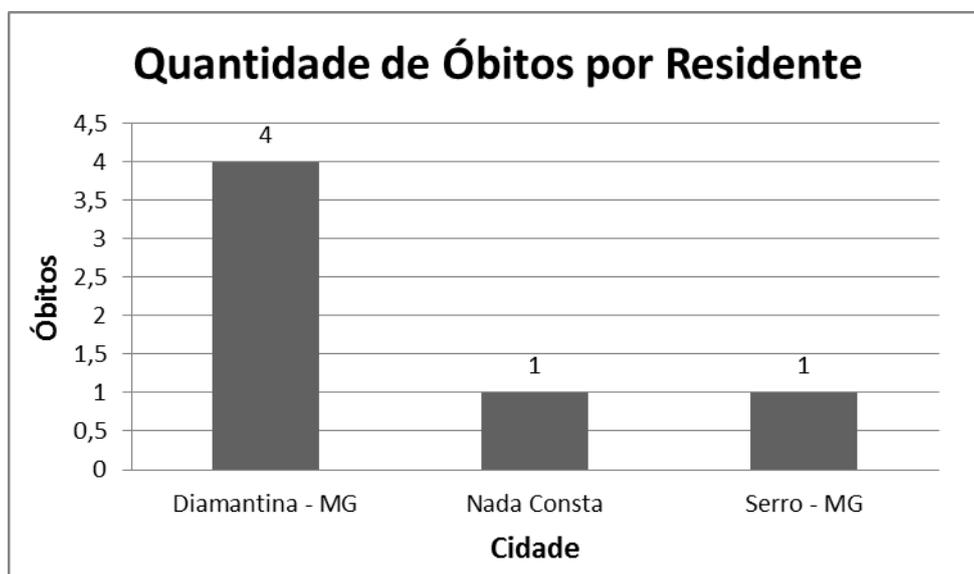


Figura 159. Quantidade de óbitos por residente no *cluster* 24.

A Figura 159 mostra a quantidade de óbitos por residente no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças que residiam em Diamantina-MG, com quatro registros, sendo equivalente a 66% dos seis registros agrupados no *Cluster*.

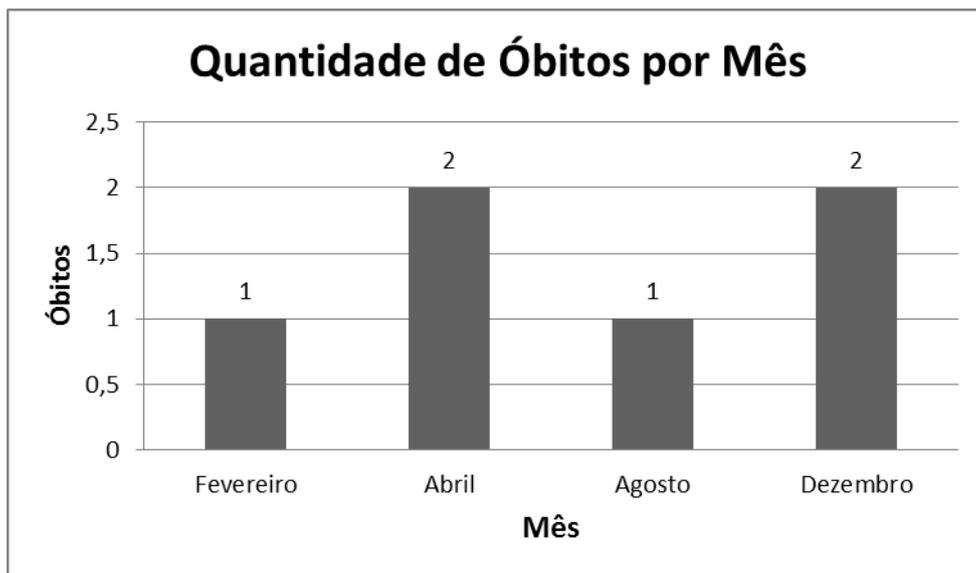


Figura 160. Quantidade de óbitos por mês no *cluster* 24.

A Figura 160 mostra a quantidade de óbitos por mês no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças que ocorreram nos meses de abril e dezembro, com dois registros cada, sendo ambos, equivalente a 33% dos seis registros agrupados no *Cluster*.

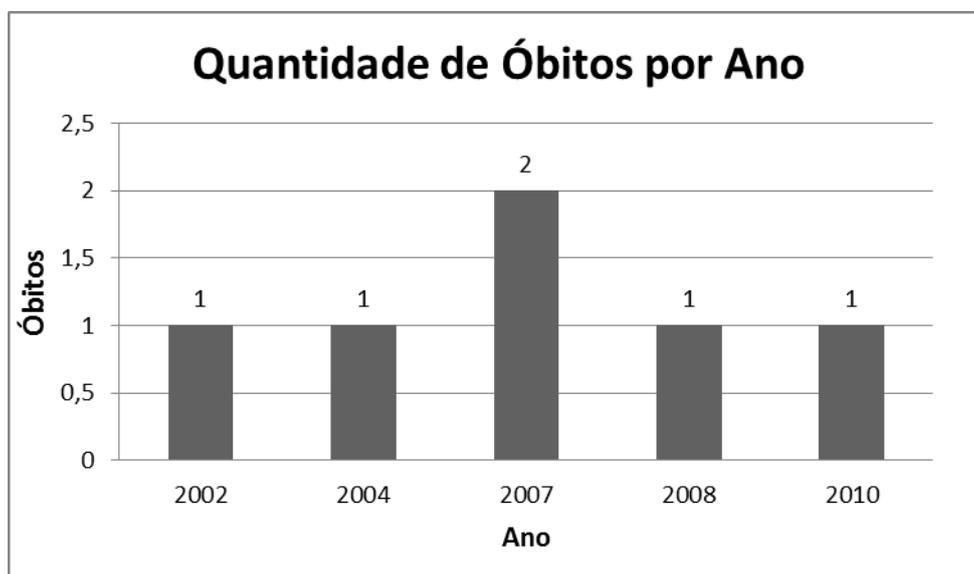


Figura 161. Quantidade de óbitos por ano no *cluster* 24.

A Figura 161 mostra a quantidade de óbitos por ano no *Cluster* 24, obtidos pelo modelo Cassiopeia. Nesta Figura é possível perceber a predominância de óbitos de crianças que faleceram no ano de 2007, com dois registros, sendo equivalente a 33% dos seis registros agrupados no *Cluster*.

Cluster 25:

Número de Registros: 02

Tabela:

Tabela 39. *Cluster 25* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---------------------------------|-----------|-----------------|------------------|------------------|-----------|-----------|
| 1 | 07 2007 | choque séptico, leptospirose | Feminino | 10 anos | Diamantina MG | Diamantina MG | Agosto | 2007 |
| 2 | 03 2012 | choque séptico (septicemia) | Masculino | 1 mês 0 ano) | Curvelo MG | Nada Consta | Dezembro | 2012 |

Na Tabela 39 é apresentado o *Cluster 25*, com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

Gráficos:

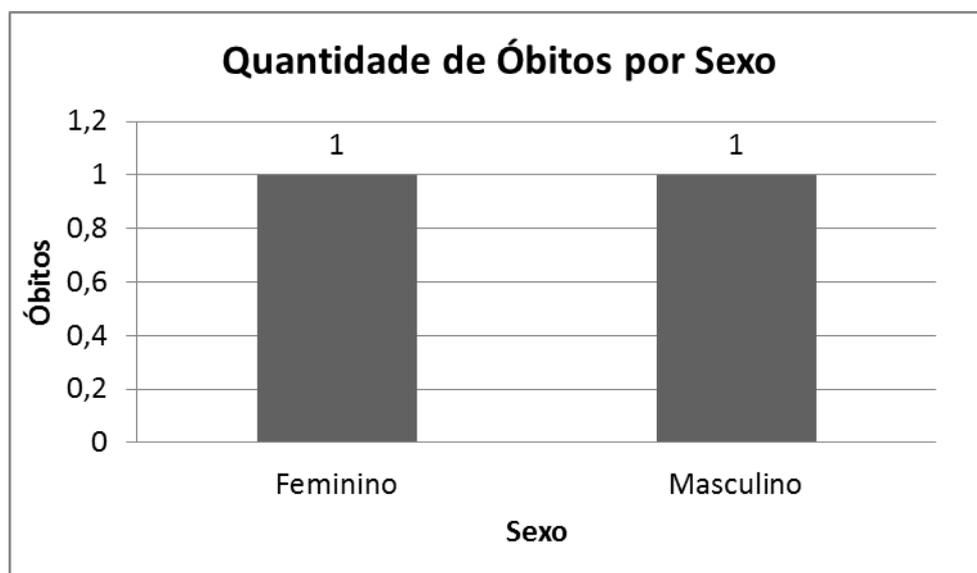


Figura 162. Quantidade de óbitos por sexo no *cluster* 25.

A Figura 162 mostra a quantidade de óbitos por sexo no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* eram de sexos diferentes.

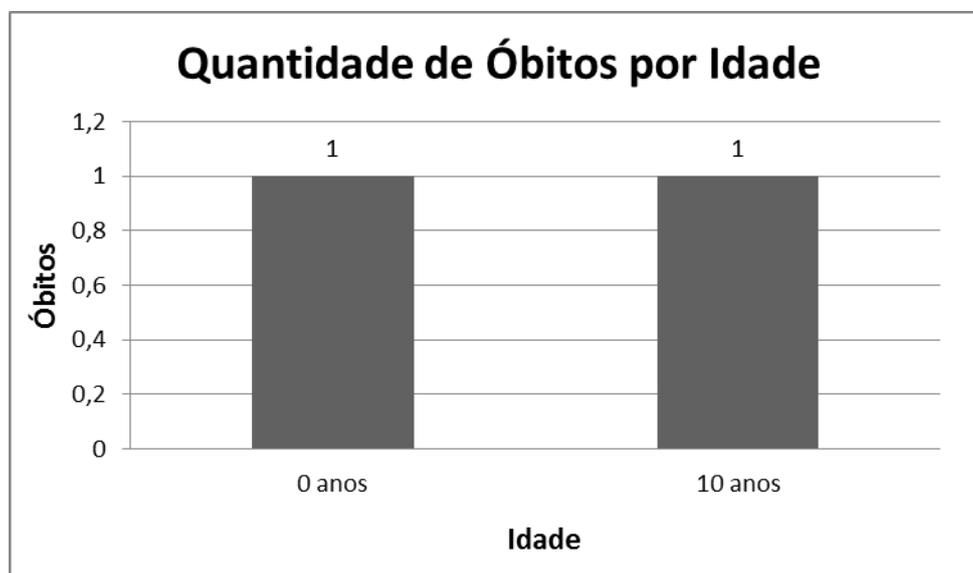


Figura 163. Quantidade de óbitos por idade no *cluster* 25.

A Figura 163 mostra a quantidade de óbitos por idade no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* possuíam idades diferentes.

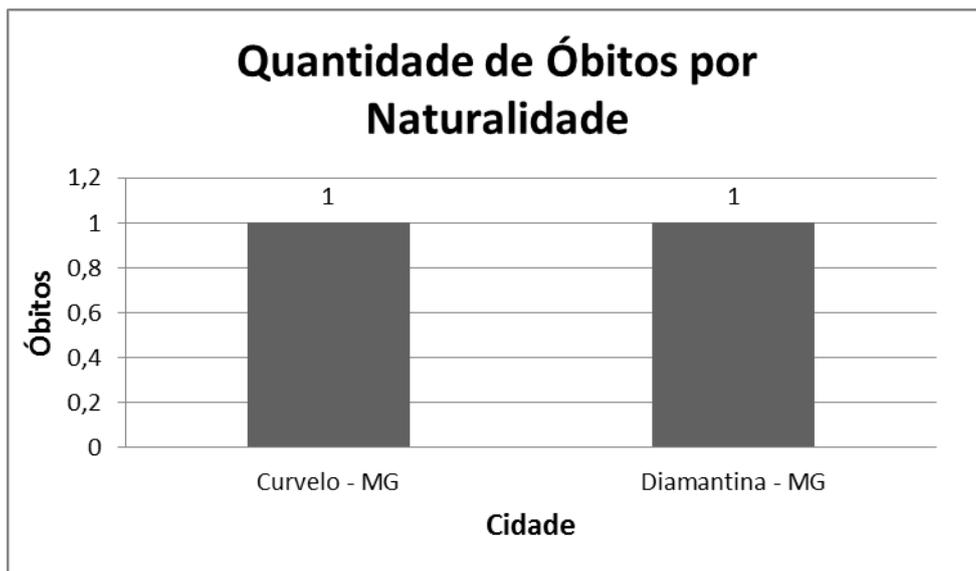


Figura 164. Quantidade de óbitos por naturalidade no *cluster* 25.

A Figura 164 mostra a quantidade de óbitos por naturalidade no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* são naturais de cidades diferentes.

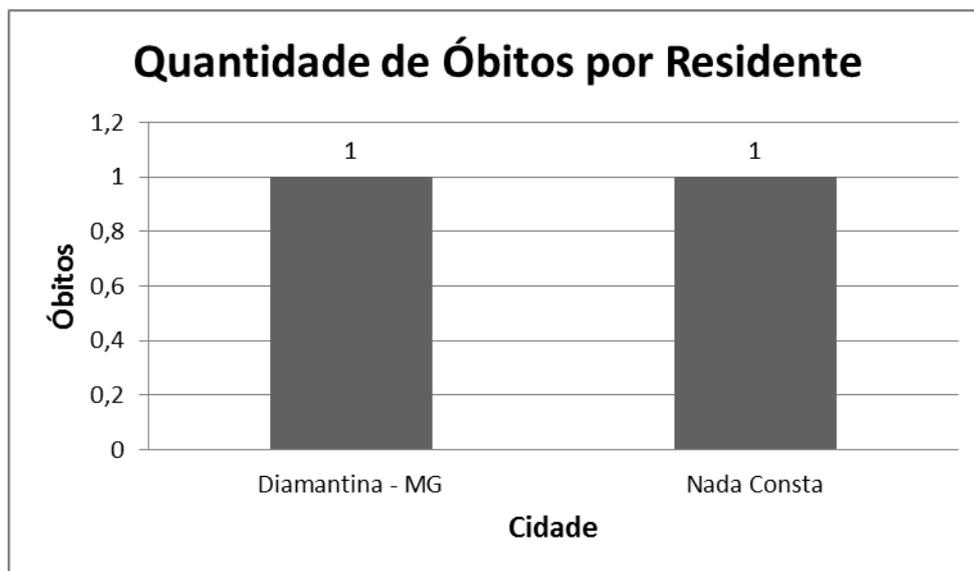


Figura 165. Quantidade de óbitos por residente no *cluster* 25.

A Figura 165 mostra a quantidade de óbitos por residente no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura uma das crianças residia em Diamantina-MG, e a outra não consta nos seus registros onde ela morava.

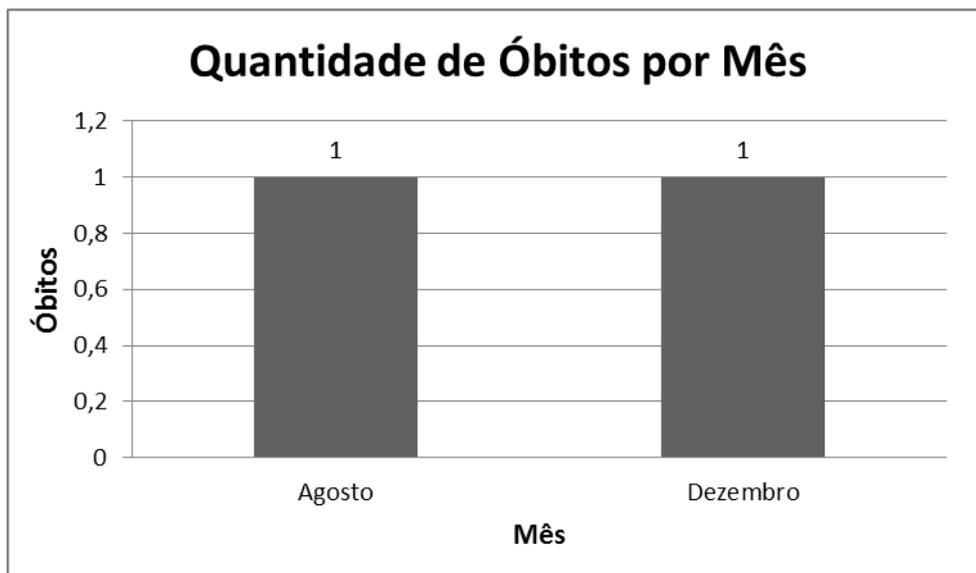


Figura 166. Quantidade de óbitos por mês no *cluster* 25.

A Figura 166 mostra a quantidade de óbitos por mês no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em meses diferentes.

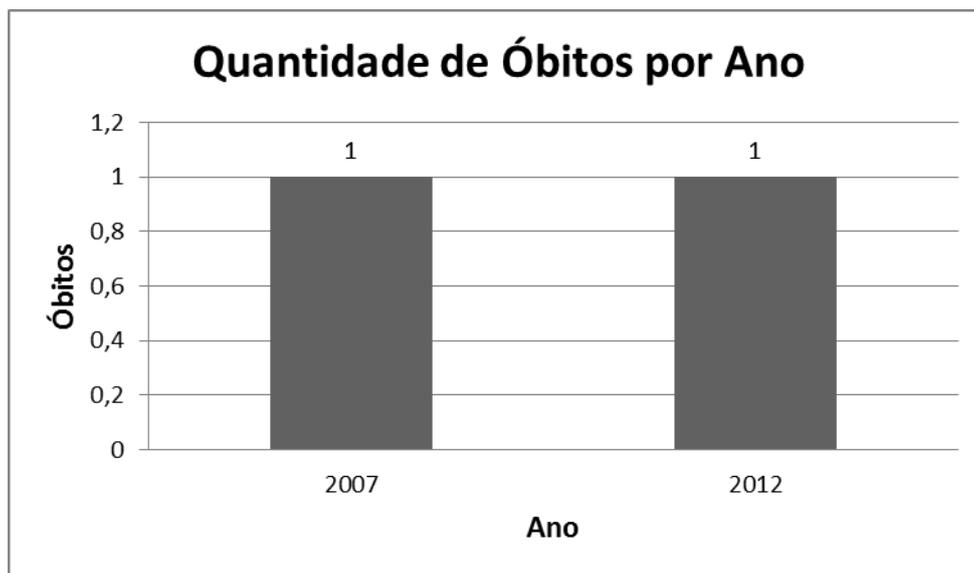


Figura 167. Quantidade de óbitos por ano no *cluster* 25.

A Figura 167 mostra a quantidade de óbitos por ano no *Cluster* 25, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em anos diferentes.

Cluster 26:

Número de Registros: 01

Tabela:

Tabela 40. *Cluster 26* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--------------------------|---------------|------------|--|------------------|-----------|-----------|
| 1 | 02 2006 | cetoacidose diabética | Femi- nino | 12 anos | São Gonçalo do Rio das Pedras MG | Diamantina MG | Fevereiro | 2006 |

Na Tabela 40 é apresentado o *Cluster 26*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de haver somente um registro no agrupamento.

Gráficos:

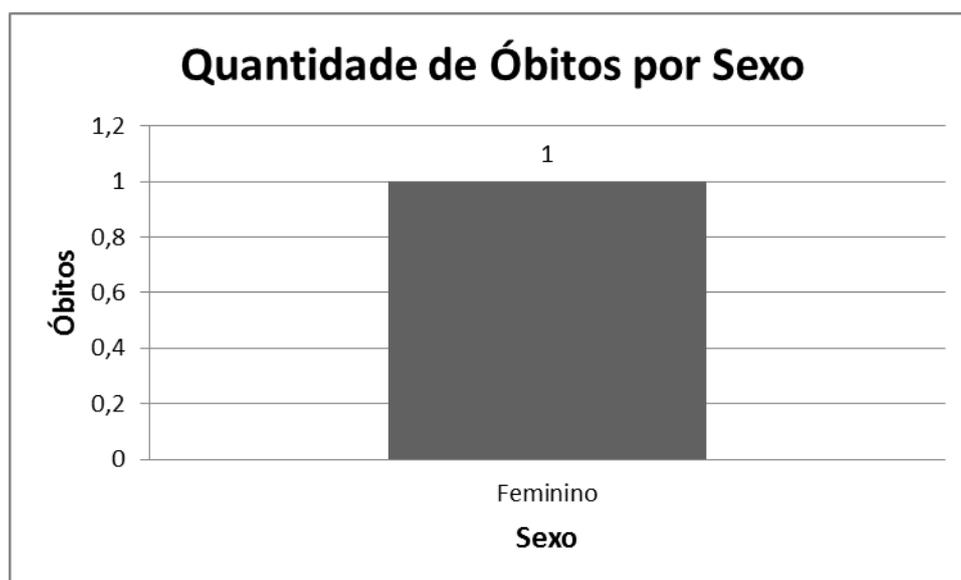


Figura 168. Quantidade de óbitos por sexo no *cluster* 26.

A Figura 168 mostra a quantidade de óbitos por sexo no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo feminino.

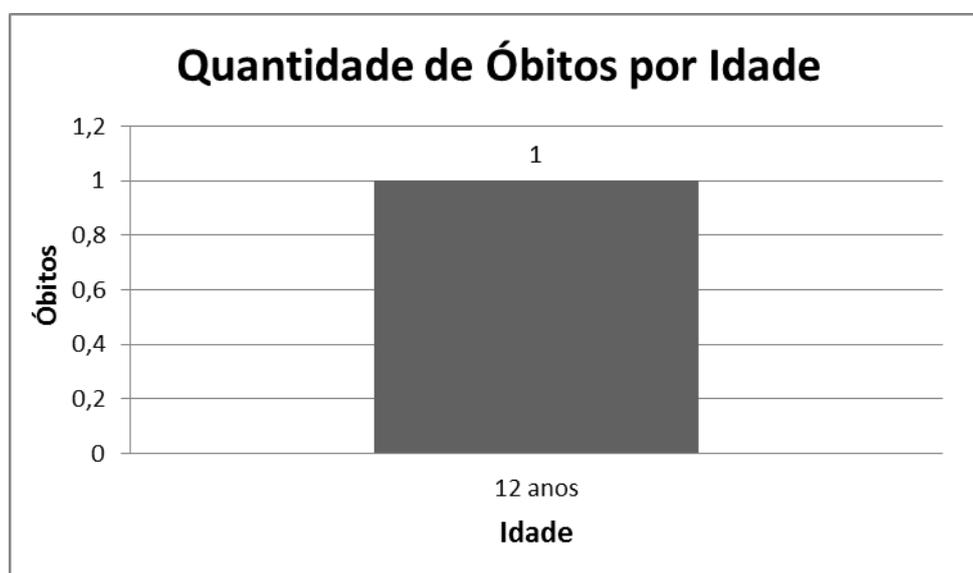


Figura 169. Quantidade de óbitos por idade no *cluster* 26.

A Figura 169 mostra a quantidade de óbitos por idade no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía 12 anos.

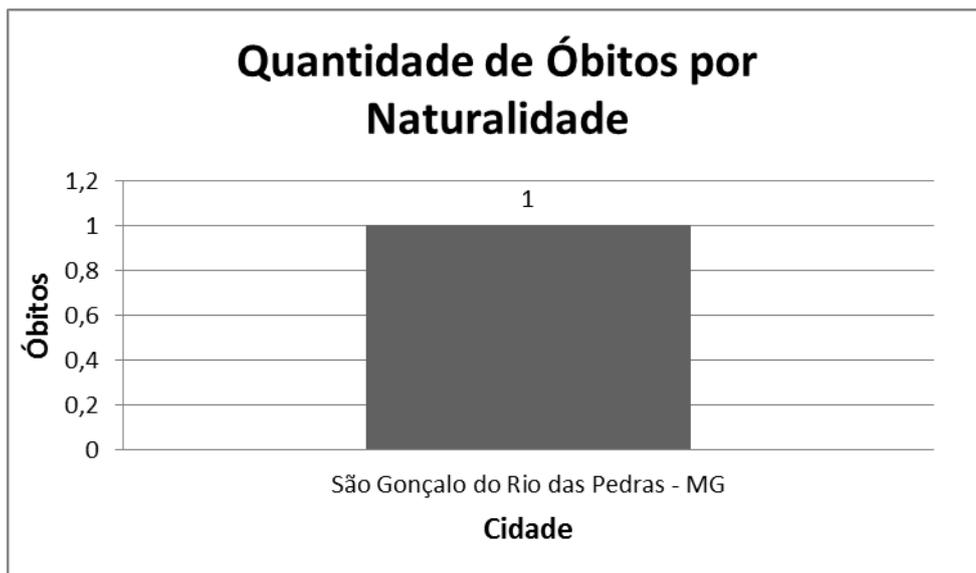


Figura 170. Quantidade de óbitos por naturalidade no *cluster* 26.

A Figura 170 mostra a quantidade de óbitos por naturalidade no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural de São Gonçalo do Rio das Pedras-MG.

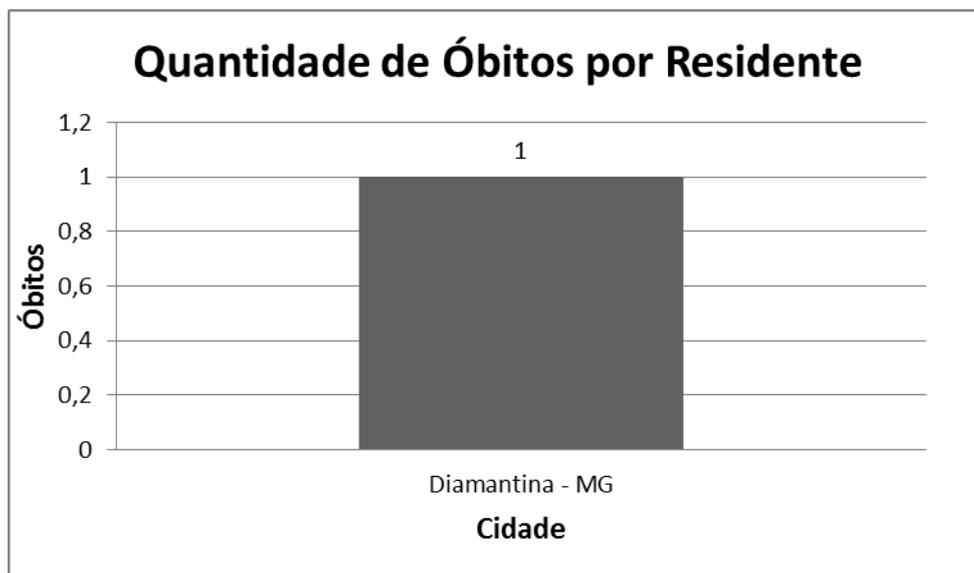


Figura 171. Quantidade de óbitos por residente no *cluster* 26.

A Figura 171 mostra a quantidade de óbitos por residente no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* residia em Diamantina-MG.

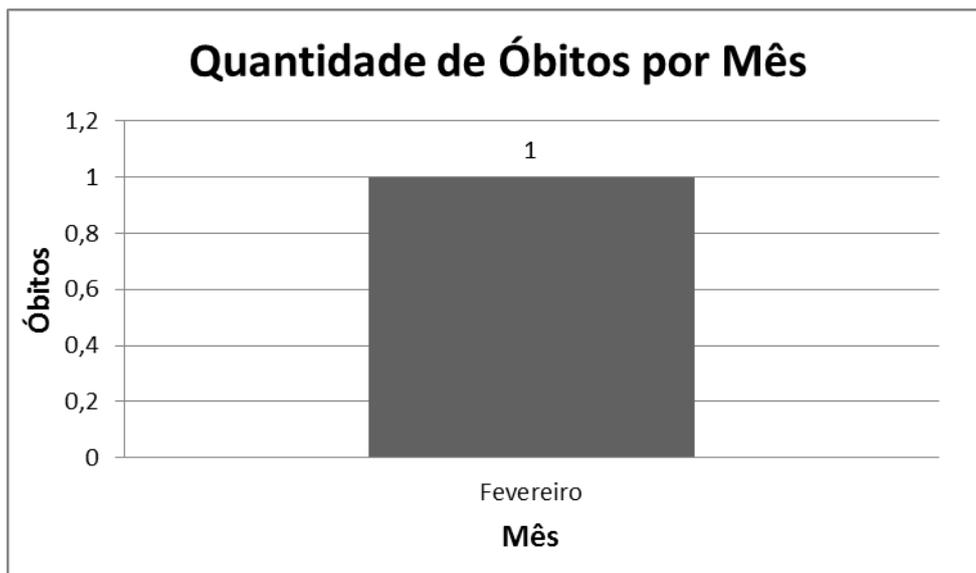


Figura 172. Quantidade de óbitos por mês no *cluster* 26.

A Figura 172 mostra a quantidade de óbitos por mês no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de fevereiro.

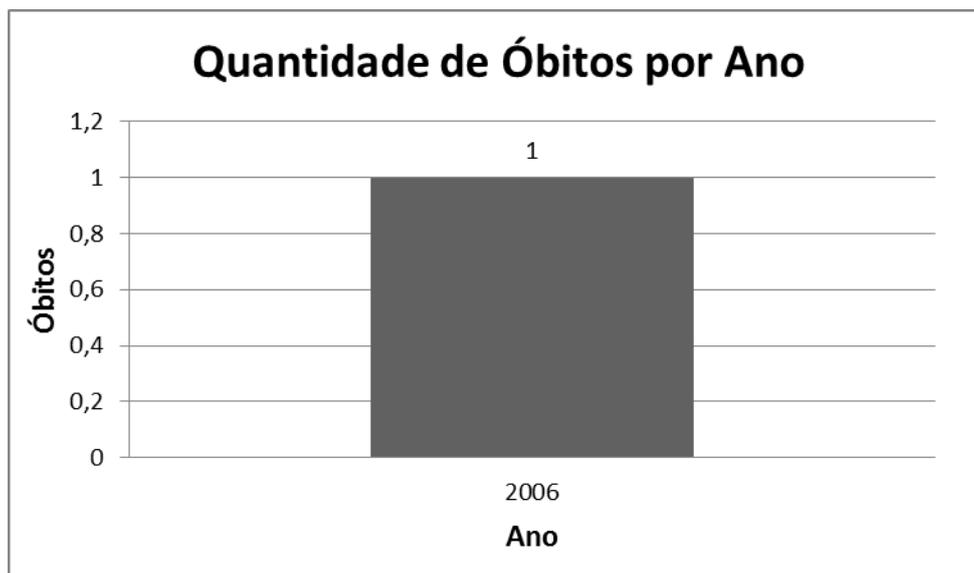


Figura 173. Quantidade de óbitos por ano no *cluster* 26.

A Figura 173 mostra a quantidade de óbitos por ano no *Cluster* 26, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2006.

Cluster 27:

Número de Registros: 01

Tabela:

Tabela 41. *Cluster 27* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--|----------|------------|--------------|-------------|-----------|-----------|
| 1 | 11 2009 | envenenamento, medicamentos psicotrópicos | Feminino | 13 anos | Datas MG | Datas MG | Setembro | 2009 |

Na Tabela 41 é apresentado o *Cluster 27*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de haver somente um registro no agrupamento.

Gráficos:

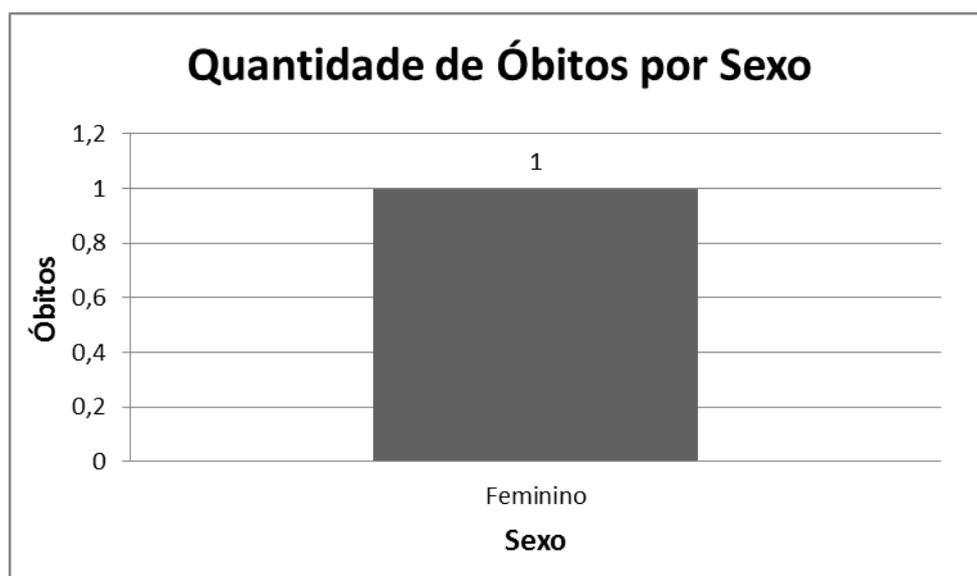


Figura 174. Quantidade de óbitos por sexo no *cluster 27*.

A Figura 174 mostra a quantidade de óbitos por sexo no *Cluster 27*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo feminino.

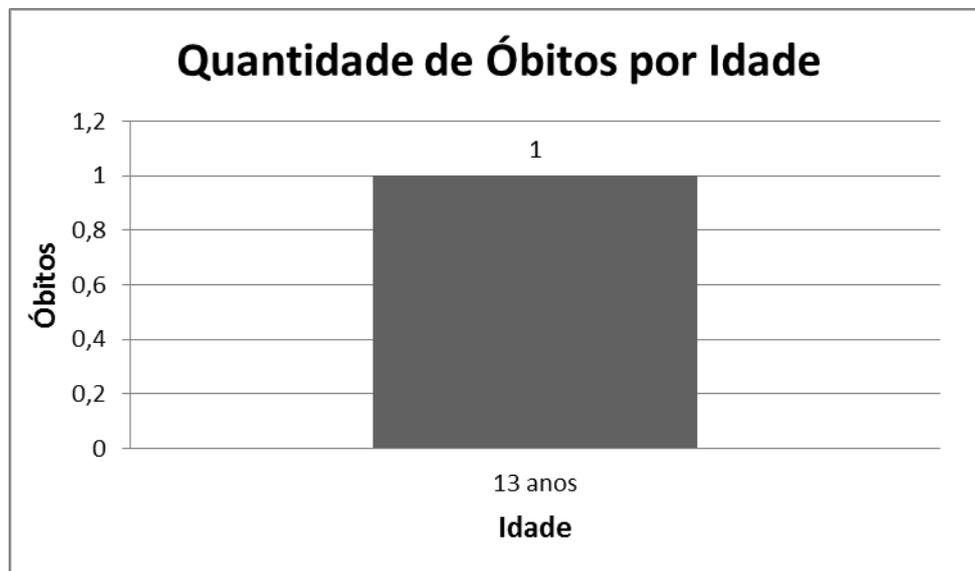


Figura 175. Quantidade de óbitos por idade no *cluster 27*.

A Figura 175 mostra a quantidade de óbitos por idade no *Cluster 27*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía 13 anos de idade.

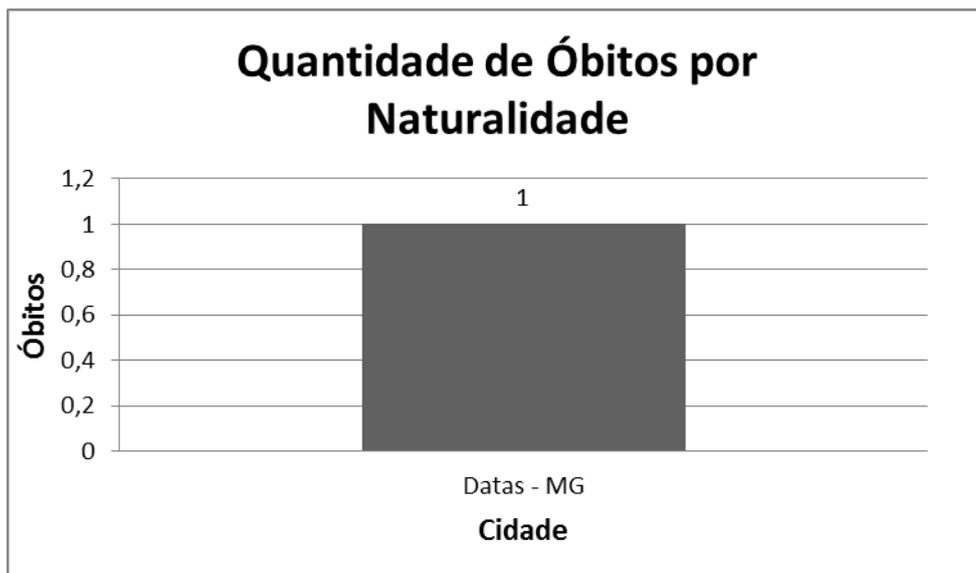


Figura 176. Quantidade de óbitos por naturalidade no *cluster 27*.

A Figura 176 mostra a quantidade de óbitos por naturalidade no *Cluster 27*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* era natural de Dadas-MG.

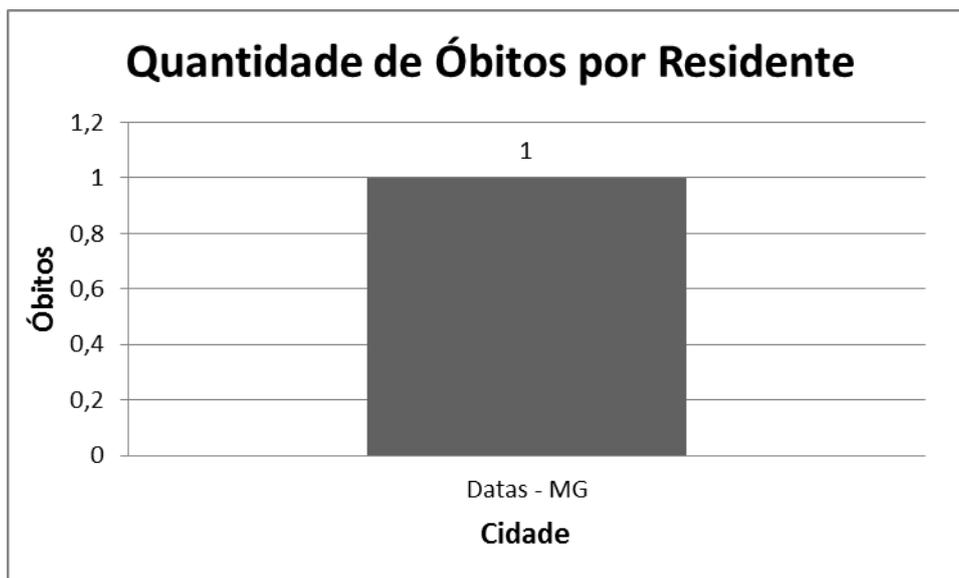


Figura 177. Quantidade de óbitos por residente no *cluster 27*.

A Figura 177 mostra a quantidade de óbitos por residente no *Cluster 27*, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* residia em Dadas-MG.

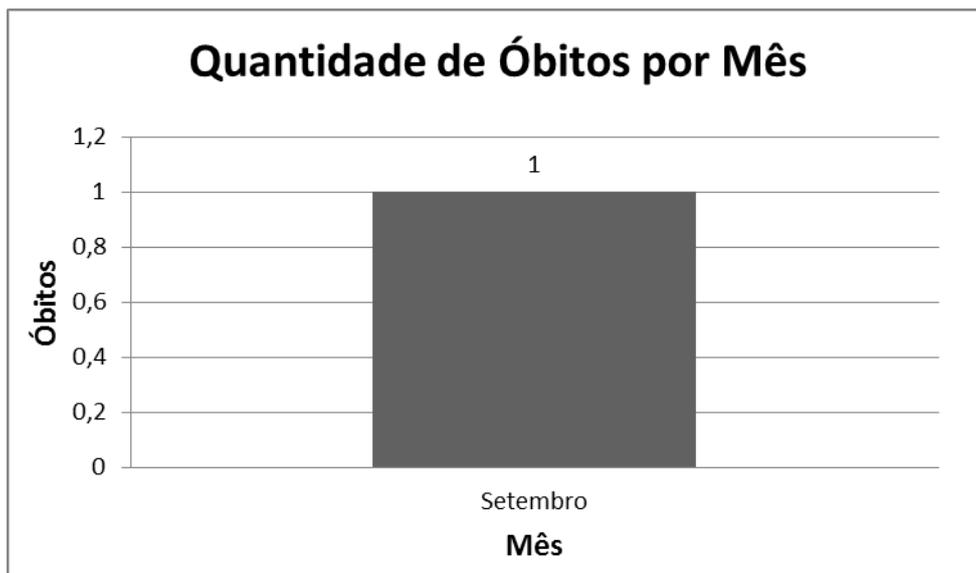


Figura 178. Quantidade de óbitos por mês no *cluster* 27.

A Figura 178 mostra a quantidade de óbitos por mês no *Cluster* 27, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de setembro.

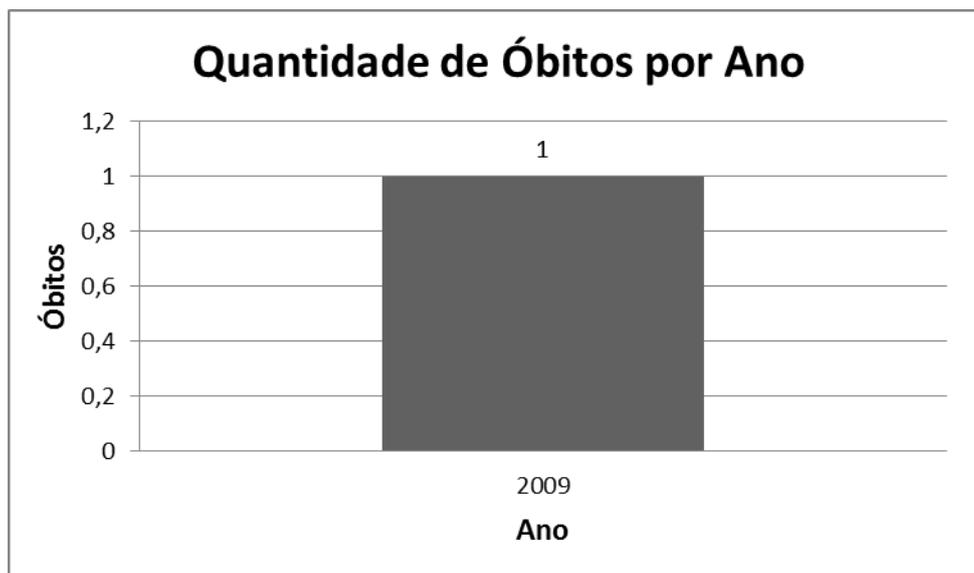


Figura 179. Quantidade de óbitos por ano no *cluster* 27.

A Figura 179 mostra a quantidade de óbitos por ano no *Cluster* 27, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2009.

Cluster 28:

Número de Registros: 01

Tabela:

Tabela 42. *Cluster 28* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|----------|------------------|------------------|----------------|-----------|-----------|
| 1 | 16 2009 | coagulação intravascular disseminada | Feminino | 1 dia (0 ano) | Diamantina MG | Nada Consta | Dezembro | 2009 |

Na Tabela 42 é apresentado o *Cluster 28*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de haver somente um registro no agrupamento.

Gráficos:

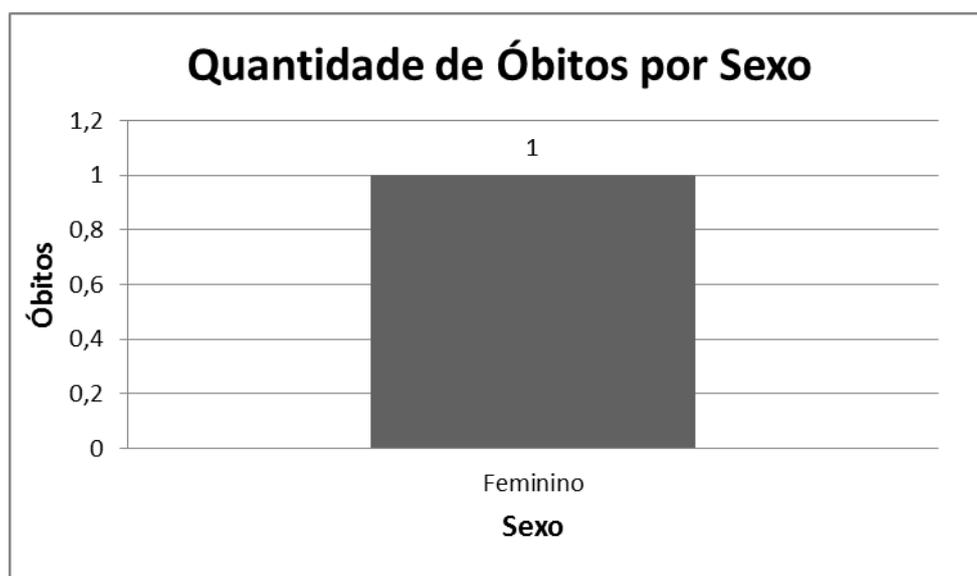


Figura 180. Quantidade de óbitos por sexo no *cluster* 28.

A Figura 180 mostra a quantidade de óbitos por sexo no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era do sexo feminino.

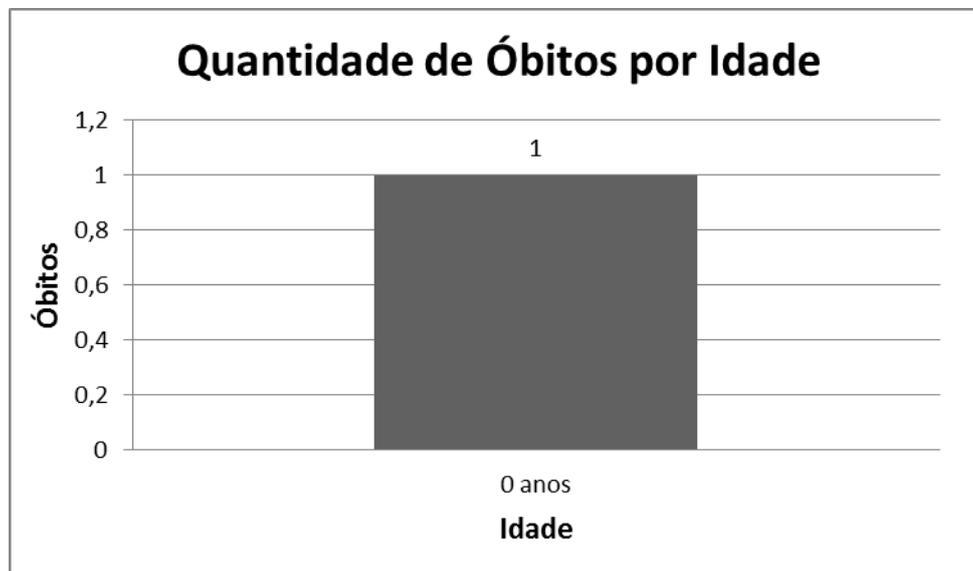


Figura 181. Quantidade de óbitos por idade no *cluster* 28.

A Figura 181 mostra a quantidade de óbitos por idade no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* possuía idade abaixo de um ano.

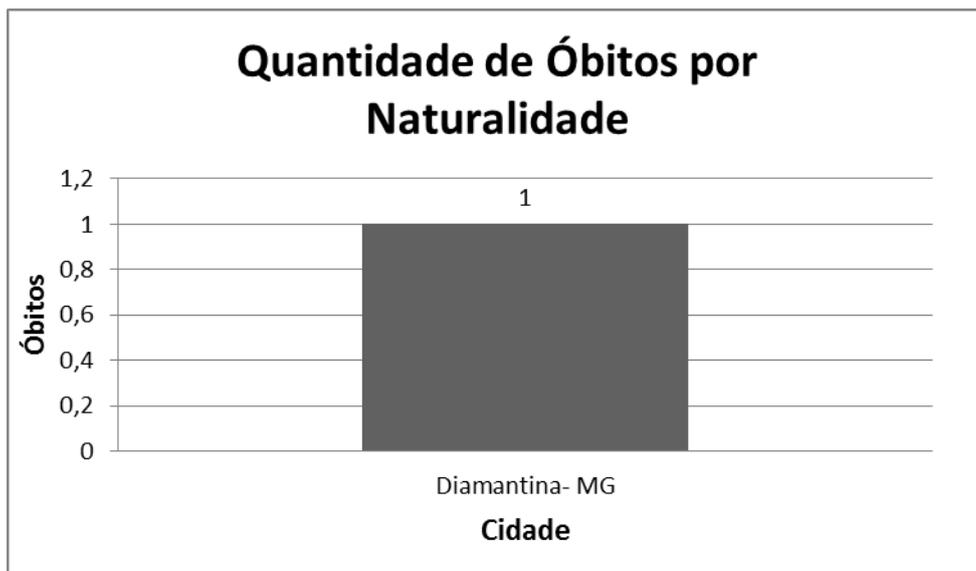


Figura 182. Quantidade de óbitos por naturalidade no *cluster* 28.

A Figura 182 mostra a quantidade de óbitos por naturalidade no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural de Diamantina-MG.

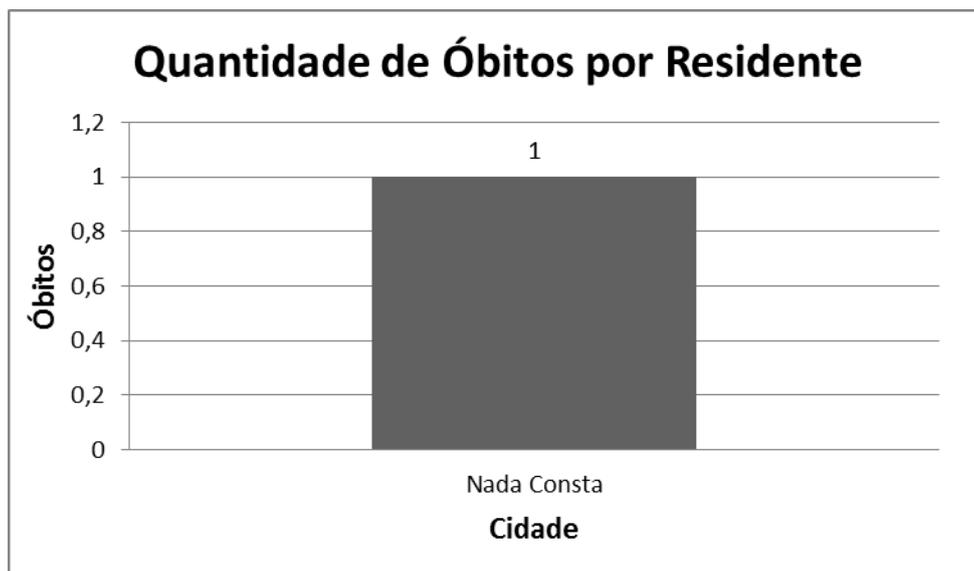


Figura 183. Quantidade de óbitos por residente no *cluster* 28.

A Figura 183 mostra a quantidade de óbitos por residente no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* não possuía a informação de onde residia.

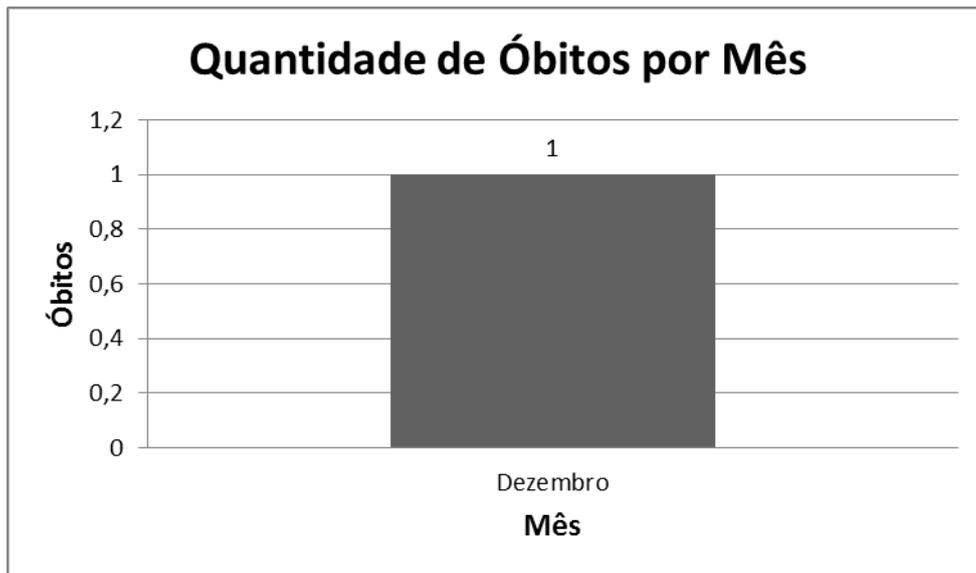


Figura 184. Quantidade de óbitos por mês no *cluster* 28.

A Figura 184 mostra a quantidade de óbitos por mês no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de dezembro.

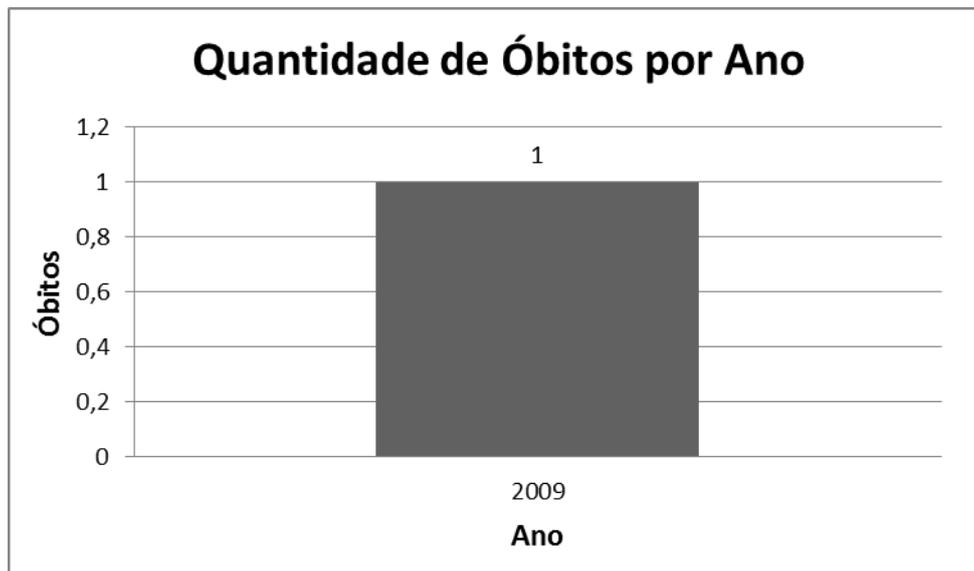


Figura 185. Quantidade de óbitos por ano no *cluster* 28.

A Figura 185 mostra a quantidade de óbitos por ano no *Cluster* 28, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2009.

Cluster 29:

Número de Registros: 02

Tabela:

Tabela 43. *Cluster 29* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|---|-----------|--------------------|------------------|------------------|-----------|-----------|
| 1 | 14 2011 | síndrome de desconforto respiratório, prematuridade extrema | Feminino | 3 horas (0 ano) | Diamantina MG | Diamantina MG | Abril | 2011 |
| 2 | 04 2011 | prematuridade extrema, desconforto respiratório | Masculino | 2 anos | Diamantina MG | Diamantina MG | Julho | 2011 |

Na Tabela 43 é apresentado o *Cluster 29*, com dois registros gerados pelo modelo Cassiopeia, o que equivale ao agrupamento de 1% dos 147 registros processados.

Gráficos:

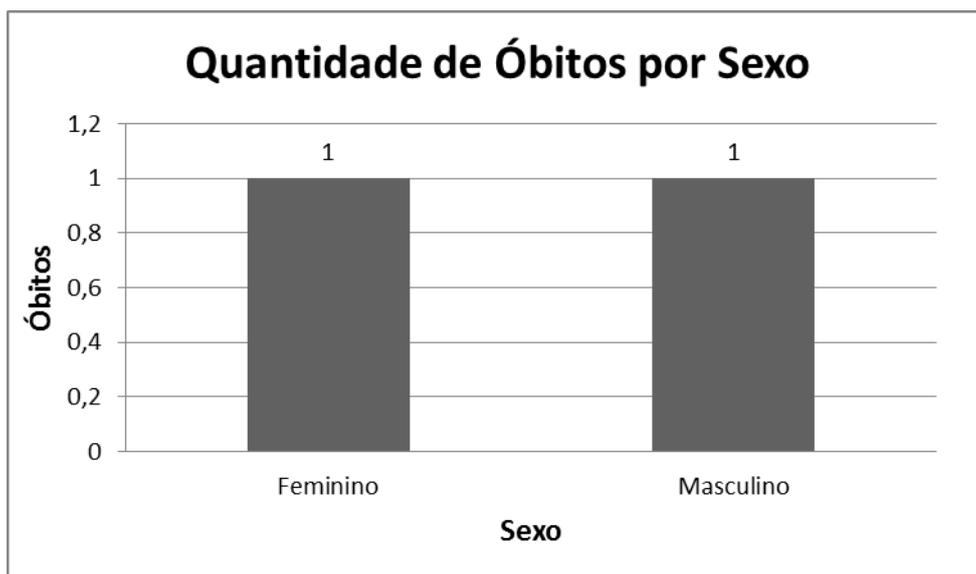


Figura 186. Quantidade de óbitos por sexo no *cluster 29*.

A Figura 186 mostra a quantidade de óbitos por sexo no *Cluster 29*, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* eram de sexos diferentes.

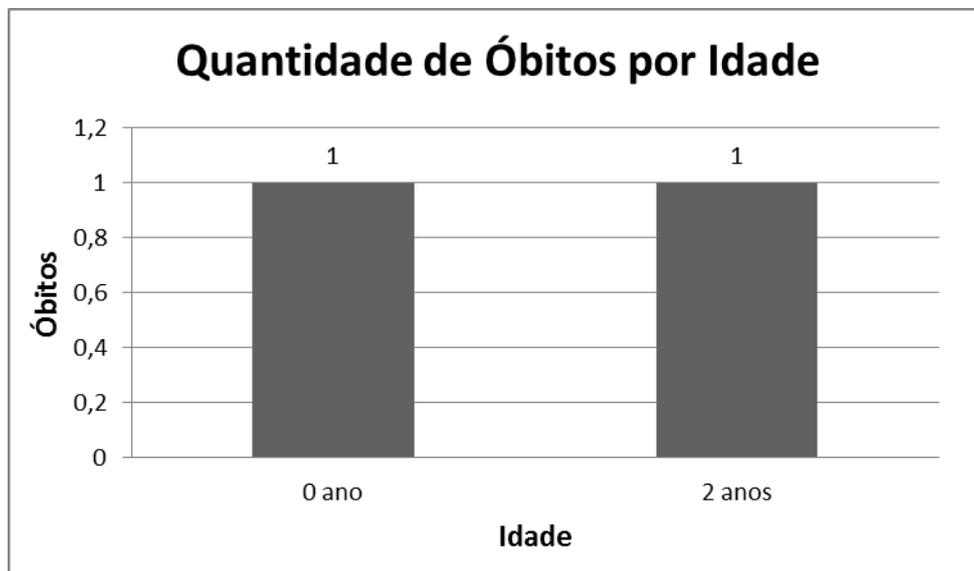


Figura 187. Quantidade de óbitos por idade no *cluster 29*.

A Figura 187 mostra a quantidade de óbitos por idade no *Cluster 29*, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças falecidas do *Cluster* eram de idades diferentes.

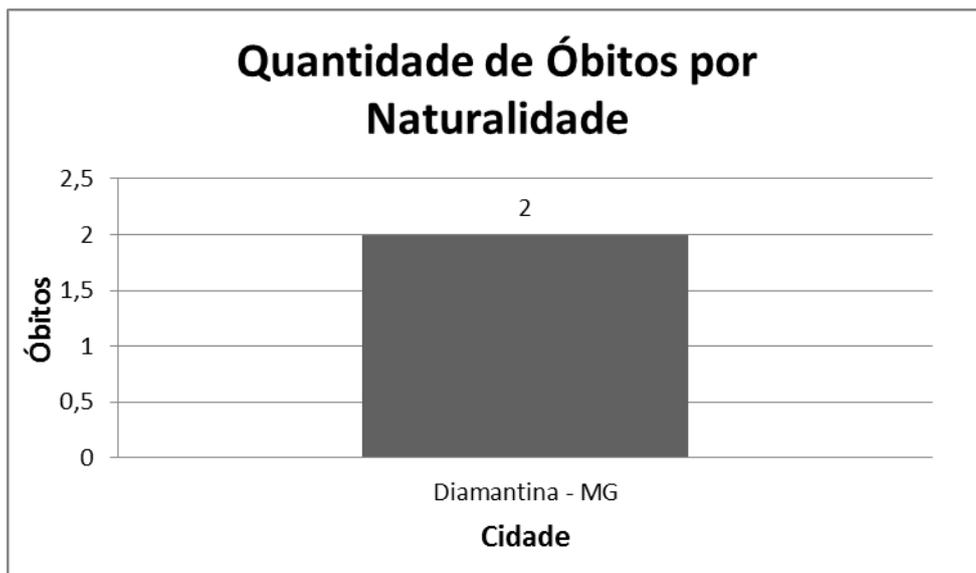


Figura 188. Quantidade de óbitos por naturalidade no *cluster* 29.

A Figura 188 mostra a quantidade de óbitos por naturalidade no *Cluster* 29, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças eram naturais de Diamantina-MG.

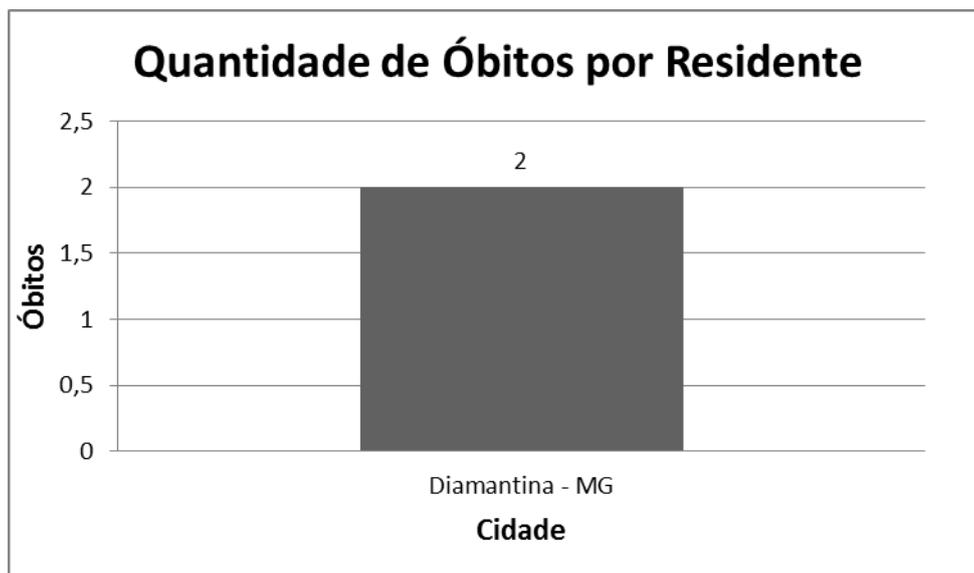


Figura 189. Quantidade de óbitos por residente no *cluster* 29.

A Figura 189 mostra a quantidade de óbitos por residente no *Cluster* 29, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças residiam em Diamantina-MG.

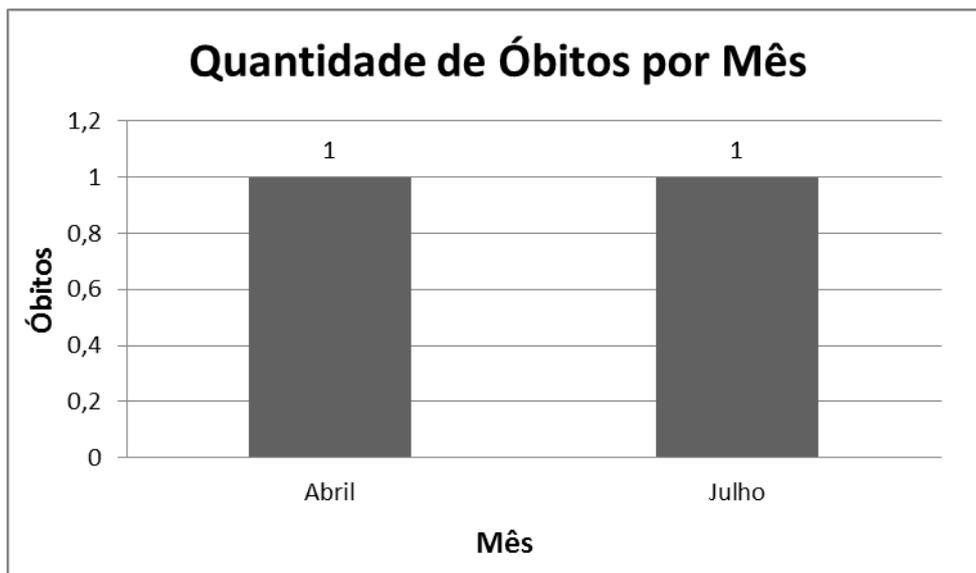


Figura 190. Quantidade de óbitos por mês no *cluster* 29.

A Figura 190 mostra a quantidade de óbitos por mês no *Cluster* 29, obtidos pelo modelo Cassiopeia. Nesta Figura as duas crianças do *Cluster* faleceram em meses diferentes.



Figura 191. Quantidade de óbitos por ano no *cluster* 29.

A Figura 191 mostra a quantidade de óbitos por ano no *Cluster* 29, obtidos pelo modelo Cassiopeia. Nesta Figura, todas as crianças faleceram no ano de 2011.

Cluster 30:

Número de Registros: 01

Tabela:

Tabela 44. *Cluster 30* gerado na clusterização realizada pelo modelo Cassiopeia.

| Nº | Txt | Óbito | Sexo | Idade | Naturalidade | Residente | Mês Óbito | Ano Óbito |
|----|------------|--------------------------------|----------|----------|---------------|---------------|-----------|-----------|
| 1 | 12 2012 | meningite, hibranencefância | Feminino | 0 ano | Gouveia MG | Gouveia MG | Junho | 2012 |

Na Tabela 44 é apresentado o *Cluster 30*, com um registro gerado pelo modelo Cassiopeia, o que equivale ao agrupamento de 0% dos 147 registros processados. *Clusters* com apenas um registro não possuíram relevância para pesquisa, pelo fato de haver somente um registro no agrupamento.

Gráficos:

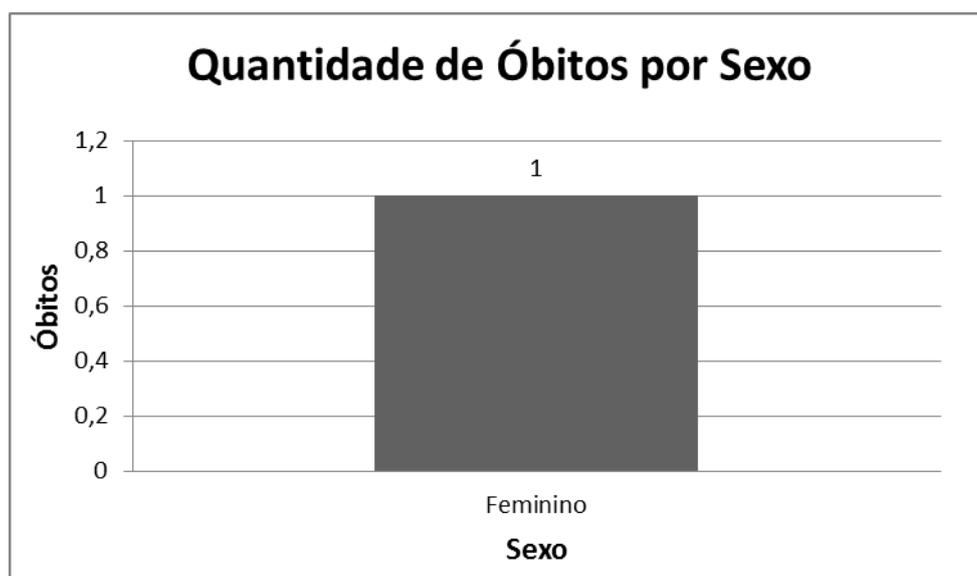


Figura 192. Quantidade de óbitos por sexo no *cluster* 30.

A Figura 192 mostra a quantidade de óbitos por sexo no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* era do sexo feminino.



Figura 193. Quantidade de óbitos por idade no *cluster* 30.

A Figura 193 mostra a quantidade de óbitos por idade no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* possuía faixa de idade abaixo de um ano.



Figura 194. Quantidade de óbitos por naturalidade no *cluster* 30.

A Figura 194 mostra a quantidade de óbitos por naturalidade no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* era natural da cidade de Gouveia-MG.

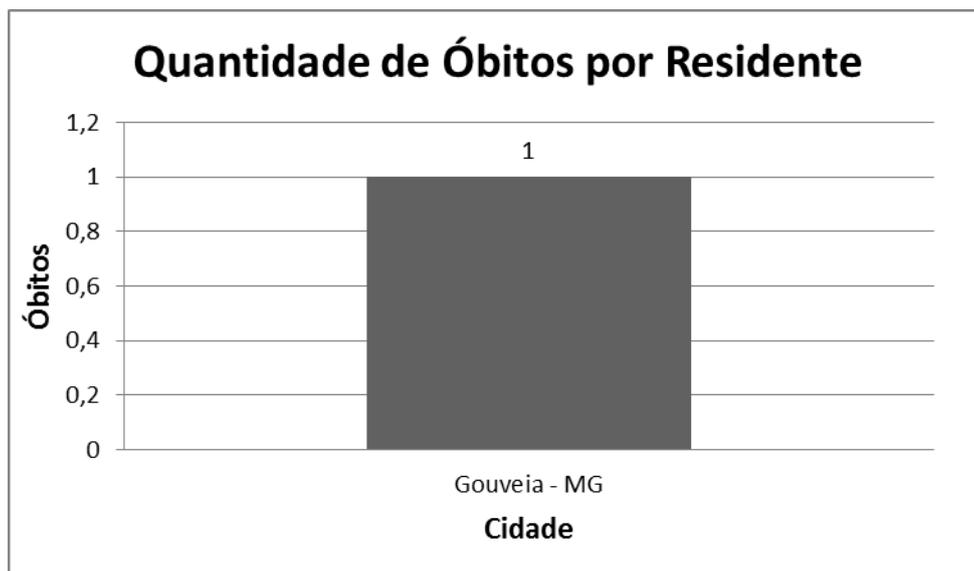


Figura 195. Quantidade de óbitos por residente no *cluster* 30.

A Figura 195 mostra a quantidade de óbitos por residente no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança falecida do *Cluster* residia em Gouveia-MG.

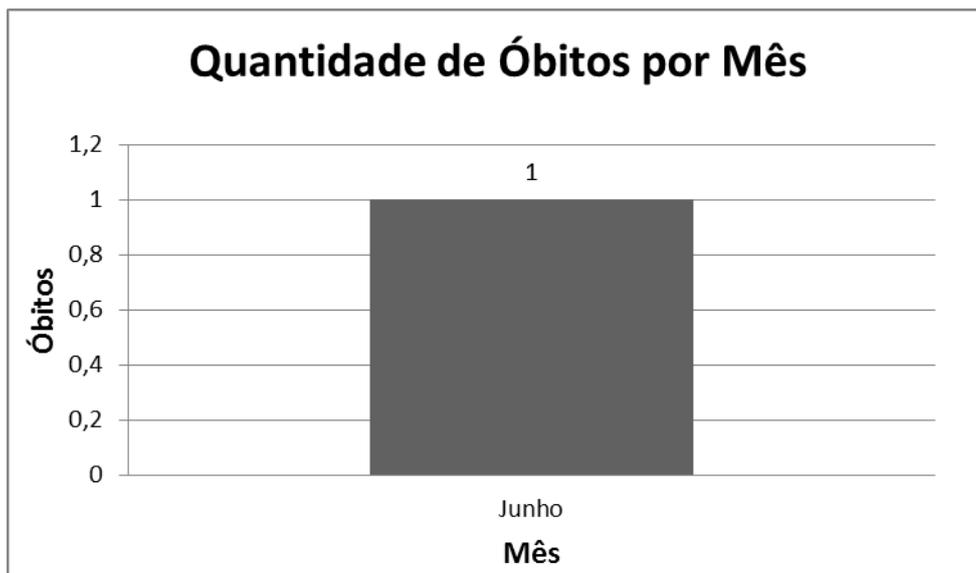


Figura 196. Quantidade de óbitos por mês no *cluster* 30.

A Figura 196 mostra a quantidade de óbitos por mês no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no mês de junho.

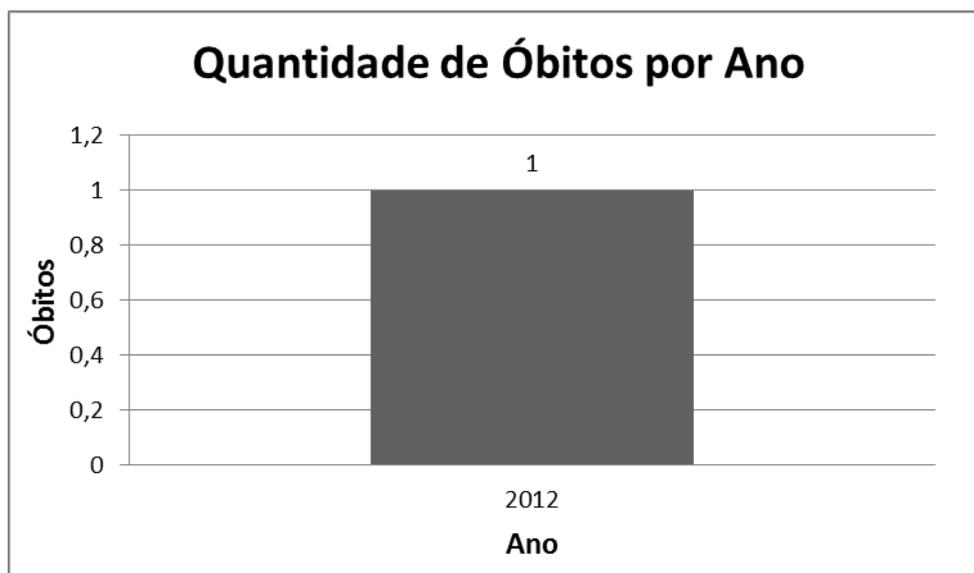


Figura 197. Quantidade de óbitos por ano no *cluster* 30.

A Figura 197 mostra a quantidade de óbitos por ano no *Cluster* 30, obtidos pelo modelo Cassiopeia. Nesta Figura, a única criança do *Cluster* faleceu no ano de 2012.

APÊNDICE B

ANÁLISE DO RELATÓRIO

O Apêndice B apresenta a continuidade das análises feita pela especialista em pediatria, a Dra. Alison Guelpeli (CRM: 58624-MG). Neste apêndice são apresentadas as análises de cada *Cluster* que foi formado durante o processamento do modelo Cassiopeia.

Análise do Relatório

CLUSTER 1:

Nesta pesquisa pode-se observar a predominância da prematuridade extrema como causa dos óbitos em recém - nascidos. Geralmente esta patologia necessita de aporte de uma UTI Neonatal para melhorar a sobrevivência destes pacientes, porém estes óbitos ocorreram no período de 2002 a 2011 e neste período a cidade ainda não contava com este serviço que só foi inaugurado em junho de 2012, justificando assim os óbitos em questão.

Quanto ao óbito das três crianças (uma de três anos, natural e residente do Serro, uma de um ano de Presidente Kubitscheck, e uma de um ano, residente em Corinto) com quadros de sepse e broncopneumonia, a distância das cidades em relação à cidade de Diamantina e o tempo de evolução da doença, podem ter colaborado com os casos de óbito. Estas cidades tem como referência a cidade de Diamantina, com exceção de Corinto que deve ter transferido a criança por falta de leito local.

A criança de Corinto já chegou à Diamantina num período que já possuía UTI neonatal, mas ainda não possuía UTI pediátrica, que são para as crianças acima de 28 dias até 12 anos de idade geralmente.

QUANTO AOS GRÁFICOS:

1- Sexo

A predominância do sexo feminino nos óbitos se deve proporcionalmente ao maior número de nascimento do mesmo sexo (feminino).

2- Faixa Etária

Geralmente os óbitos ocorrem mais até o 1º ano de vida pela fragilidade e vulnerabilidade destas crianças neste período e este é um dos grandes desafios do Ministério da Saúde, tentar a cada ano diminuir estes índices. Por isto a necessidade de manter programas de atenção primária, como, controle de baixo peso, campanhas de vacinação e acompanhamento de gestantes, dentre outros.

3- Naturalidade

O predomínio de nascimentos é de Diamantina porque esta cidade possui maternidade e referencia os pacientes de outras cidades. Logo, o predomínio de óbitos no que tange ao critério naturalidade reflete este fato.

4- Local de Residência

Fica difícil a análise deste gráfico uma vez que o percentual de nada consta quase se iguala ao percentual da cidade de Diamantina.

5- Mês

Os casos de Hipertensão arterial aumentam neste período que antecede diretamente o inverno o que pode ser justificável, o grande índice de óbitos em maio. Os casos de doenças infecciosas e sepses acontecem mais no período próximo ao verão, onde se vê as predominâncias de setembro e novembro.

6- Ano

A dificuldade em encontrar profissional especializado, no caso, um pediatra e um neonatologista, além de um intensivista pediátrico, talvez possa justificar o predomínio do ano de 2012. Desde 2010 a saúde pública vem enfrentando sérios problemas com a falta de profissionais da área de pediatria. Nos programas de saúde da família, não há a presença de pediatra, contando somente com a presença de um clínico geral. E quando há este profissional especializado, isto só vem a piorar a qualidade da saúde infantil, devido ao excesso de trabalho a um único médico. Além da falta de pediatras, há a evasão destes profissionais da rede pública, principalmente em zonas rurais e regiões mais distantes, como, o Vale do Jequitinhonha.

CLUSTER 2:

Os óbitos aconteceram em função do diagnóstico de prematuridade extrema. Mesmo após a implantação da UTI neonatal não foi possível a sobrevivência destas crianças em decorrência da imaturidade física das mesmas.

CLUSTER 3:

Os óbitos aqui agrupados estão relacionados com o diagnóstico de Insuficiência respiratória. As idades são variadas.

Quanto ao ano só um caso ocorreu após a implantação da UTI neonatal e pediátrica, porém tratou-se de um caso de falência múltipla por Leucemia. O paciente já estava em estágio terminal e este óbito seria inevitável.

A Leucemia na faixa etária descrita (15 anos) é mais comum no sexo feminino, o que foi de fato observado. Quanto aos critérios de tratamento e sobrevida dos pacientes cabe uma discussão a parte. A leucemia mieloide aguda é menos comum e muito mais agressiva com prognóstico muito ruim. Nas crianças predomina a leucemia linfóide aguda que tem muitas chances de cura e menos casos de óbitos.

CLUSTER 4:

Estes registros agrupam insuficiência respiratória e prematuridade, confirmando o aumento dos óbitos na idade inferior a 30 dias, associados a esta patologia no período que antecedeu a implantação da UTI neonatal na região.

CLUSTER 5:

Estes registros relatam sepse e choque séptico como causas de óbito predominantes. Pode-se perceber que a maioria estava na faixa etária menor que 30 dias, e dentre eles, dois eram de outra cidade, distante de Diamantina, o que pode ter agravado as complicações, além do processo de transporte das mesmas.

Quanto ao ano de 2012, este ano faz parte do período de carência dos profissionais médicos, o que pode justificar, mesmo com a presença da UTI neonatal e pediátrica, a ocorrência dos óbitos, Figura 16. Já o caso de óbito do adolescente de 14 anos, no item 4, é exceção, uma vez que é raro o óbito por sepse nesta idade, a menos que, já sofresse com uma doença de base, o que não foi mencionado no atestado de óbito.

CLUSTER 6:

Os registros agruparam insuficiência respiratória e má formação congênita com exceção do item 3 que está incompleto e não pode ser avaliado. Estas patologias

necessitam de atendimento especializado e muito específico dos quais a região não dispõe, somente nas capitais, e na sua maioria são pouco compatíveis com a vida, o que justificam os óbitos numa faixa etária menor de um ano.

CLUSTER 7:

Os registros relacionam patologias respiratórias (insuficiência respiratória, trombose pulmonar, pneumonias) e más formações congênitas, chamando a atenção para a evolução do tempo de sobrevivência aumentado de acordo com o ano. Em 2002 o óbito foi bem mais precoce em termos de faixa etária do que em 2012, em função de uma melhor assistência com a evolução dos anos. Faz-se exceção o item 6, pois pneumonia aspirativa é um acidente após uma bronco aspiração (engasgo mais severo) de crianças abaixo de um ano, não dependendo de mês ou ano, mas sim da vigilância dos responsáveis pela criança. Já com relação ao item 13, a criança possuía seis anos, mas cursava com patologia neurológica, o que a torna semelhante a uma criança com menos de um ano. Com relação ao item 14, a meningite é uma patologia grave que pode levar ao óbito independente da idade.

CLUSTER 8:

Registro único de traumatismo crânio encefálico em decorrência talvez de acidente no sentido de fatalidade. A criança de outra região estava de passagem em trânsito pela região (passeando de férias ou passando por uma estrada próxima), quando sofreu o acidente que a levou ao óbito e este não necessariamente foi um acidente automobilístico.

CLUSTER 9:

Os registros 1, 2, 3, 4, 7, 8 e 9 estão relacionados a traumatismos crânio encefálicos e estão associados a épocas do ano com festividades o que pode justificar todos estes óbitos (em decorrência de imprudência no trânsito e no convívio com a coletividade). Os itens 5, 6, 10 e 11, nenhuma descoberta se aplica.

CLUSTER 10:

Os registros 1 e 2 estão relacionados com insuficiência respiratória e cardiopatia congênita em crianças até 1 ano, antes do período da implantação da UTI neonatal e pediátrica e em função de serem patologias severas que necessitam de suporte bem mais

especializado. O item 3 foi causa acidental por animal peçonhento (o que é muito comum na região de diamantina), e que talvez pela distância e o tempo da picada, tenham sido imperiosos para este óbito que independe da época do ano. O item 4 está incompleto e deve ser desconsiderado.

CLUSTER 11:

Estes registros estão relacionados com traumas contusos, onde se percebe a predominância de outros locais de naturalidade e residência, o que pode justificar o caráter acidental e a presença das pessoas de forma passageira pela cidade, com exceção do item 2.

CLUSTER 12:

Estes registros que indicam óbitos por acidente de trânsito ocorreram em épocas festivas do ano, conseqüentemente aumentando os casos de imprudência no trânsito. Todos estavam de passagem pela região.

CLUSTER 13:

Trata-se de registros de más formações congênitas pouco compatíveis com a vida, e necessitam de assistência bem mais especializada, disponíveis somente em grandes centros, para aumentar a sobrevivência.

CLUSTER 14:

Neste *Cluster*, os registros foram agrupados devido às patologias neurológicas (choque neurogênico, encefalites), que na sua maioria aconteceram nos meses em que ocorre o verão, onde predominam as doenças infecciosas, no caso, encefalites. Faz-se exceção o item 2, cuja causa básica foi o diabetes.

CLUSTER 15:

Quase todos os registros relacionam-se a sepse e parada cardio respiratória, que ocorreram em sua maioria em Diamantina e distritos, levando a questionar casos de infecção hospitalar na cidade de Diamantina.

CLUSTER 16:

Pouco relevante, porém quando a pessoa chega ao óbito, o médico pode colocar causa desconhecida e fornecer o atestado de imediato ou enviar para o IML para investigação da causa do óbito. Este procedimento seria o mais correto, principalmente, se tratando de criança com baixa faixa etária, a fim de descartar envenenamento, intoxicação, traumas, violência entre outros.

CLUSTER 17:

Sem relevância para a pesquisa.

CLUSTER 18:

Sem relevância para a pesquisa.

CLUSTER 19:

Sem relevância para a pesquisa.

CLUSTER 20:

Apresenta registro de meningite e choque. Esta patologia é comum próximo aos meses de verão, e dependendo do agente etiológico, possui caráter fatal.

CLUSTER 21:

Nestes registros chama a atenção os óbitos de causa indeterminada, que conforme já mencionado, são casos de óbitos que chegam ao hospital e o médico pode emitir o atestado de óbito como causa indeterminada ou enviar para o IML para necropsia, que seria o mais correto. Chama a atenção este excesso de causas indeterminadas emitido e o papel do médico na emissão de atestado de óbito. Com este comportamento, o profissional altera totalmente as estatísticas de causa de óbito por ano no país e deve ser evitado. Todos foram do sexo masculino, porém como não sabemos as causas, fica difícil uma avaliação.

CLUSTER 22:

Registro único de sofrimento fetal e malformações múltiplas. Estas condições geralmente são incompatíveis com a vida ou tem uma pequena sobrevivência de algumas horas, mesmo recebendo atenção mais especializada, o que não foi o caso.

CLUSTER 23:

Neste registro há a encefalopatia hipóxico- isquêmica, que seria a má oxigenação da criança, levando a lesões cerebrais irreversíveis ou não, dependendo do grau e tempo de hipóxia do cérebro. No registro 1 houve o óbito mais precoce em decorrência do sofrimento fetal agudo e da sepse, ou seja, houve complicações severas que levaram ao óbito. No registro 2 a hipoxia não foi tão severa, o que permitiu um tempo de vida maior da criança, porém esta era sindrômica e provavelmente foi a óbito por complicações de sua síndrome.

CLUSTER 24:

Os registros 1, 2 e 3 tratam-se de asfixia por afogamento e estão relacionados a uma faixa etária maior (adolescentes), e a uma época do ano ainda de temperaturas mais altas, o que possibilita estes tipos de acidentes, principalmente em piscinas e cachoeiras, que são tipos de entretenimento mais comuns na região. Geralmente os adolescentes tendem a se aventurar e se acham onipotentes, podendo levar a estes desfechos. No item 4, a síndrome de aspiração de mecônio ocorre quando a criança passa do tempo de nascimento ou o líquido amniótico torna-se infectado por bolsa rota. Trata-se de um líquido atípico que pode levar a morte por asfixia durante o nascimento ou horas após. O item 5 e 6 registram asfixia por soterramento e sufocação direta, que nesse caso são situações particulares e ocasionais sem relação às anteriores.

CLUSTER 25:

O item 1 do registro trata de um choque séptico por leptospirose em uma criança mais velha. Chama a atenção para a investigação epidemiológica da doença, que poderia ter sido feita a fim de rastrear esta enfermidade em familiares, vizinhos e afins. No item 2 é mais provável de acontecer este tipo de doença nesta faixa etária, porém chama a atenção quanto ao problema da vigilância da infecção hospitalar

CLUSTER 26:

Este registro é de um caso de óbito mais esporádico de um adolescente de 12 anos. Geralmente não evolui para óbito por esta causa, mas pela época do ano. Foi época festiva de carnaval e a mesma pode ter se descuidado de seu tratamento, além de ser uma época mais difícil para assistência médica. Esta gama de fatores pode ter colaborado para o óbito.

CLUSTER 27:

Registra um óbito por envenenamento por uso de psicotrópico em adolescente de 13 anos sexo feminino. As adolescentes vivem nesta fase um mundo de conflitos e dúvidas, e quando não são bem conduzidas, tendem a ideias de autoextermínio. A cidade e as regiões ao redor apresentam um índice grande destes casos.

CLUSTER 28:

Sem relevância para a pesquisa.

CLUSTER 29:

São registros relacionados a desconforto respiratório e prematuridade extrema, ocorridos no mesmo ano, sendo condições que podem precipitar o óbito.

CLUSTER 30:

Sem relevância para a pesquisa.