

Visualização de grandes conjuntos de dados: exemplos dos Censos Agropecuários e de precipitações diárias

Alfredo José Barreto Luiz¹ Pedro Vernier² Rafaela Ferraz Molina³

Resumo

O uso de gráficos ou outras formas de representação visual na análise exploratória de dados é recomendado e utilizado há muito tempo. Entretanto, nem sempre está claro o que fazer e o que não fazer nas diversas opções de visualização de forma a auxiliar na interpretação de grandes conjuntos de dados. O presente trabalho fornece exemplos do uso do diagrama em caixa e bigodes (*box and whisker plot*) na visualização de dados relacionados à produção de cana-de-açúcar oriundos dos Censos Agropecuários 2006 e 2017, do IBGE. São comparados gráficos com e sem a transformação logarítmica dos dados. Também é oferecido um exemplo de visualização da série histórica de muitos anos, da precipitação pluviométrica diária registrada em centenas de estações meteorológicas, localizadas no estado de Goiás e no Distrito Federal, em uma única figura. Os exemplos são usados para demonstrar a utilidade da visualização nos passos iniciais de análises complexas, que abrangem grandes e heterogêneos conjuntos de dados. Também é ressaltada a importância do conhecimento profundo da forma de coleta e da natureza de cada variável na escolha do gráfico, da transformação ou da forma de visualização dos dados, de maneira que as representações pictóricas ajudem na interpretação. Por fim, é feito um alerta para se evitar o uso de diagramas ou gráficos em excesso, repetitivos ou que produzam falsas impressões.

Palavras-chave: diagrama em caixa e bigodes; transformação logarítmica; gráficos de barras; estratificação; análise descritiva.

¹ Embrapa Meio Ambiente, Jaguariúna, SP – alfredo.luiz@embrapa.br

² Graduando em Engenharia Agrícola, FEAGRI/UNICAMP, Campinas, SP - p204768@dac.unicamp.br

³ Mestranda em Produção Animal Sustentável, IZ/APTA, Nova Odessa, SP - rafamolina.rafamolina@gmail.com