

O objetivo desta prática é exercitar a análise exploratória de um conjunto de dados. Para isso, espera-se que o aluno consiga, utilizando observações estatísticas e apresentações gráficas, responder questionamentos sobre as relações em um conjunto de dados.

Para esta prática você vai precisar:

- RapidMiner
- Editor de texto com suporte para figuras (ex: Microsoft Office Word, LibreOffice Writer, LaTeX)
- Base Adult (<http://archive.ics.uci.edu/ml/datasets/Adult>)

A seguir são apresentados 5 questionamentos. Para cada um deles, você deve apresentar um ou mais gráficos juntamente com um texto descrevendo suas observações. Deve constar no relatório como você gerou os gráficos apresentados.

O relatório deve estar no formato pdf para submissão.

1. Observe a relação existente entre gênero e estado civil. Descreva suas observações. Lembre-se que a base de dados Adult é uma amostra com apenas maiores de idade ($age > 16$) e que trabalham ($hours\text{-}per\text{-}week > 0$).
2. Investigue a relação entre ganho de capital e ganho anual: qual é o ganho médio de capital de acordo com o ganho anual? Observe também o histograma do ganho de capital. O que podemos inferir sobre essas observações?
3. Quais são as profissões em que mais pessoas tem ganho anual $\geq 50K$? Quais profissões tem mais $\leq 50K$?
4. Quem trabalha mais horas em média: o marido, a esposa, pessoas com filhos, etc? Este comportamento varia dependendo da idade?
5. O que mais você consegue explorar nesta base? Apresente pelo menos mais uma relação que conseguir encontrar utilizando uma visualização.