

Exercícios – R para Ciência dos Dados

I. Exploração dos Dados

3.1. Missing values:

- 3.1.1. Quantas observações com missing values existem em `flights`?
- 3.1.2. Usa `flights` para criar uma nova tabela sem *missing values*.
- 3.1.2. Calcula o atraso médio das chegadas por dia sem usar `na.rm=TRUE`.

3.2. Outliers:

- 3.2.1. Usando `flights`, calcula a média de atraso na chegada por avião (usa `tailnum`).
- 3.2.2. Observa a distribuição do atraso médio nas chegadas por avião.
- 3.2.3. Cria uma nova tabela com o número de voos e a média de atraso na chegada por avião.
- 3.2.4. Observa a relação entre o número de voos e o atraso médio graficamente.
- 3.2.5. Retira as médias com menos de 25 observações e reconstrói o gráfico.

3.3. Distribuição 1D:

- 3.3.1. Em `diamonds`, compara a distribuição de `price` com a distribuição Gaussiana
- 3.3.2. Quais os valores dos momentos de `price` sem transformação?
- 3.3.3. Compara graficamente a distribuição de `carat` com a distribuição Gaussiana.
- 3.3.4. Será que `carat` também beneficiaria de transformação logarítmica?

3.4. Distribuição 2D:

- 3.4.1. Em `diamonds`, reproduz o estudo gráfico entre `clarity` e `price` para o corte "Fair".
- 3.4.2. Qual é o valor da correlação no estudo anterior?
- 3.4.3. Em `diamonds`, entre `cut`, `clarity` e `color`, quais têm uma correlação maior?