

## Exercícios – R para Ciência dos Dados

### I. Exploração dos dados

#### 1. Visualização

- 1.1. Que geom se utiliza para fazer uma curva de densidade? E um boxplot? E um histograma?
- 1.2. Corre o comando `ggplot(data=mpg)`. O que faz?
- 1.3. Quantas linhas e colunas existem em `mpg`?
- 1.4. Quais as variáveis contínuas e as categóricas em `mpg`?
- 1.5. A que corresponde a variável `drv`?
- 1.6. Cria um scatterplot de `hwy` e `cyl`. Como podes melhorar a visualização?
- 1.7. Quantas observações e variáveis existem em `diamonds`?
- 1.8. Como se relacionam as variáveis `color` e `cut`?
- 1.9. Como se relacionam as variáveis `color` e `price`?

#### 2. Manipulação de Dados

- 2.1. `filter()` :
  - 2.1.1. Seleciona os voos cuja chegada atrasou mais de 2 horas.
  - 2.1.2. Seleciona os voos para Houston (IAH ou HOU).
  - 2.1.3. Seleciona os voos que chegaram mais de 2 horas atrasados mas não partiram atrasados.
  - 2.1.4. Seleciona os voos que partiram entre meia-noite e as 6 da manhã.
- 2.2. `arrange()` :
  - 2.2.1. Ordena `flights` para encontrar os voos com a partida mais atrasada.
  - 2.2.2. Ordena `flights` para encontrar o voo mais rápido.
  - 2.2.3. Que voos viajaram distâncias maiores? E as menores?
  - 2.2.4. Coloca no principio os valores com missing values (usa a função `is.na()`).
- 2.3. `select()` :
  - 2.3.1. Tenta usar várias vezes a mesma variável no `select()`.
  - 2.3.2. Para que serve a função `one_of()`? (usa `vars <- c("year", "years")`).
- 2.4. `mutate()` :
  - 2.4.1. Converte `dep_time` de HHMM (ou HMM) para horas e minutos.
  - 2.4.2. Recria a variável `dep_delay`.
- 2.5. `summarize()` e `group_by()` :
  - 2.5.1. Quando parte o primeiro e o último voo de cada dia?
  - 2.5.2. Que destino tem mais aviões diferentes a voar para lá?
  - 2.5.3. Quantos voos partem antes das 5:00 am?
  - 2.5.4. Qual a proporção de voos atrasados mais de 1 hora por dia do ano?