

Estatísticas e Gráficos

Ricardo Theodoro

OBSCOOP/USP

- Estatísticas descritivas
 - média, mediana, desvio padrão e quartis
- Gráficos
 - Histograma, boxplot, barras e pizza

Por que utilizar programação para estatística?

Importar



Arrumar

(Armazenar os dados
consistentemente)



Transformar

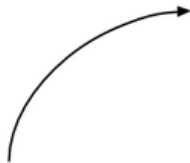
(Criar novas variáveis e
agregações)

Visualizar

(Surpreende, mas não é
escalável)

Modelar

(É escalável, mas não
surpreende)



```
survAluno_alterado <-  
  read.csv("data/survAluno_alterado.csv")
```

Crie um script no seu projeto e rode esses códigos para ver os resultados.

Pacotes utilizados

```
install.packages(c(  
  "dplyr",  
  "ggplot2",  
  "ggthemes",  
  "psych",  
  "summarytools",  
  "tidyselect"  
))
```

Sobre estatística básica com R

Sobre estatísticas descritivas

Exemplos de ggplot2

Galeria de Gráficos

Resumo do ggplot2

Estatísticas descritivas com o R

Variáveis numéricas

Média

```
mean(survAluno_alterado$n_livros_ano)
```

Desvio Padrão

```
sd(survAluno_alterado$n_livros_ano)
```

Mediana

```
median(survAluno_alterado$n_livros_ano)
```

Quartil

```
quantile(survAluno_alterado$n_livros_ano)
```

Estatísticas descritivas

```
survAluno_alterado |>
```

```
  dplyr::select(where(is.numeric)) |> # rodar library(tidyselect)
```

```
  psych::describe(quant = c(0, .25, .5, .75, 1))
```

Estatísticas descritivas com o R

Variáveis categóricas

■ Resumo dos dados

```
library(tidyselect)

psych::describeBy(survAluno_alterado |>
  dplyr::select(where(is.numeric)),
  survAluno_alterado$religiao)
```

```
# Frequência
survAluno_alterado$animais_domesticos |>
  as.factor() |>
  summary()
```

```
# Proporções
survAluno_alterado$animais_domesticos |>
```


Estatísticas descritivas com o R

Variáveis categóricas - Tabelas de frequência e proporções

```
# Por linha
table(survAluno_alterado$animais_domesticos,
      survAluno_alterado$casa_pais) |>
  proportions() |>
  round(2)
```

```
# Por coluna
table(survAluno_alterado$animais_domesticos,
      survAluno_alterado$casa_pais) |>
  proportions(2) |>
  round(2)
```

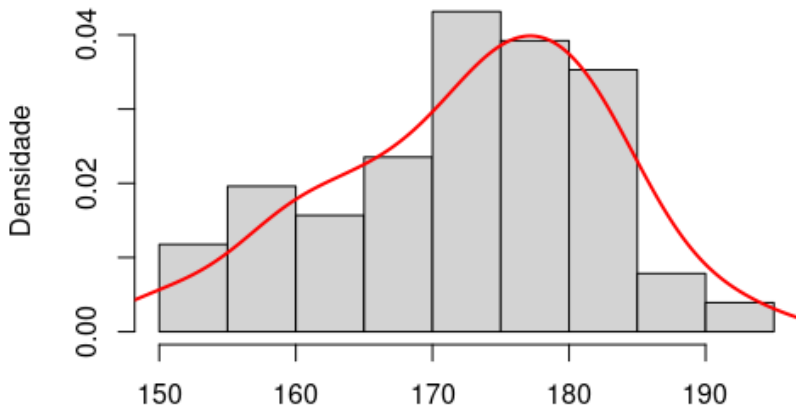
```
summarytools::ctable(
  x = survAluno_alterado$animais_domesticos,
  y = survAluno_alterado$casa_pais,
```

Histograma

- Utilizado para verificar a distribuição dos dados

```
hist(  
  survAluno_alterado$altura,  
  freq = FALSE,  
  main = "Distribuição da Altura dos Alunos",  
  ylab = "Densidade",  
  xlab = "Valores da altura"  
)  
  
# Curva de densidade (deve ser plotada sobre histograma de densidade)  
lines(density(survAluno_alterado$altura),  
      lwd = 2,  
      col = "red")
```

Distribuição da Altura dos Alunos

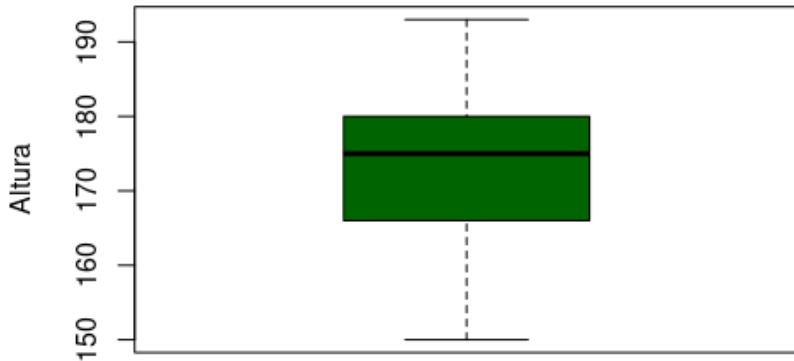


Boxplot

- Utilizado para verificar a distribuição dos dados

```
boxplot(  
  survAluno_alterado$altura,  
  main = "Boxplot da Altura dos Alunos",  
  ylab = "Altura",  
  xlab = "Distribuição",  
  col = "darkgreen"  
)
```

Boxplot da Altura dos Alunos



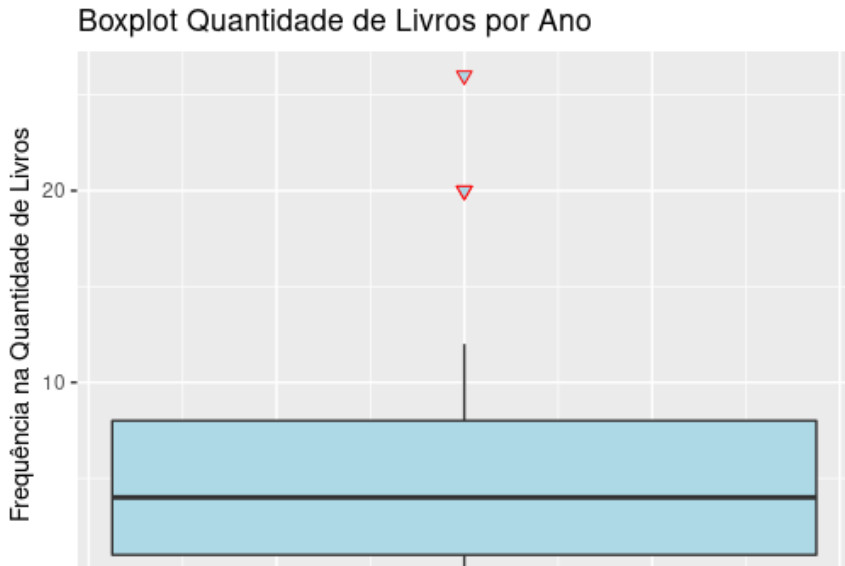
Exportando gráficos

Como salvar o gráfico como imagem .png

```
png("graficos/boxplot_n_livros.png")
```

```
survAluno_alterado |>  
  ggplot2::ggplot() +  
  ggplot2::aes(x = n_livros_ano) +  
  ggplot2::geom_boxplot(  
    fill = "lightblue",  
    outlier.colour = "red",  
    outlier.shape = 25,  
    outlier.size = 2  
  ) +  
  ggplot2::coord_flip() +  
  ggplot2::labs(title = 'Boxplot Quantidade de Livros por Ano') +  
  ggplot2::xlab("Frequência na Quantidade de Livros")
```

Exportando gráficos



Gráficos com ggplot2

Boxplot com ggplot2

- Ao invés de `|>`, usamos o `+`

```
survAluno_alterado |>
  ggplot2::ggplot() + # primeira camada
  ggplot2::aes(x = n_livros_ano) + # definição estética e eixos
  ggplot2::geom_boxplot( # definimos que será usado um boxplot e suas camadas
    fill = "lightblue",
    outlier.colour = "red",
    outlier.shape = 25,
    outlier.size = 2
  ) +
  ggplot2::coord_flip() + # alteramos as coordenadas
  ggplot2::labs(title = 'Boxplot Quantidade de Livros por Ano') + # adicionamos título
  ggplot2::xlab("Frequência na Quantidade de Livros") # adicionamos legenda
```


Gráficos com ggplot2

Histograma

```
survAluno_alterado |>
  ggplot2::ggplot() +
  ggplot2::aes(n_livros_ano) +
  ggplot2::geom_histogram(      # tipo de gráfico
    ggplot2::aes(y = ..density..),
    bins = 13,
    colour = "#011e5a",
    fill = "lightblue"
  ) +
  # Curva de densidade
  ggplot2::stat_function(fun = dnorm, # cálculo da curva (dnorm = distrib
    args = list( # argumentos utilizados para o cálculo
      mean = mean(survAluno_alterado$n_livros_ano, na.rm=T),
      sd = sd(survAluno_alterado$n_livros_ano, na.rm=T)
```

Histograma da Quantidade de Livros Lidos por Ano

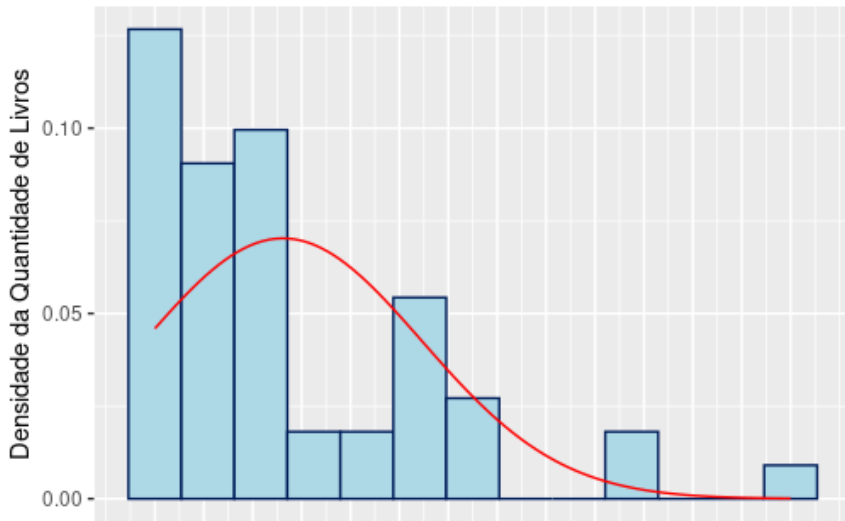


Gráfico de Pizza

```
survAluno_alterado |>
  dplyr::select(oculos) |>
  dplyr::group_by(oculos) |>
  dplyr::count() |> # Reparem que aqui editamos nossa base antes de gerar
  ggplot2::ggplot() +
  ggplot2::aes(x = "", y = n, fill = oculos) + # Gráfico de pizza não tem
  ggplot2::geom_bar(stat = "identity") +
  ggplot2::coord_polar("y") +
  ggplot2::geom_text(
    ggplot2::aes(label = n),
    col = "white",
    size = 8,
    position = ggplot2::position_stack(vjust = 0.5)
  ) +
  ggplot2::labs(title = "Gráfico de pizza sobre usuários de óculos") +
  ggplot2::xlab("") +
```

Gráfico de pizza sobre usuários de óculos

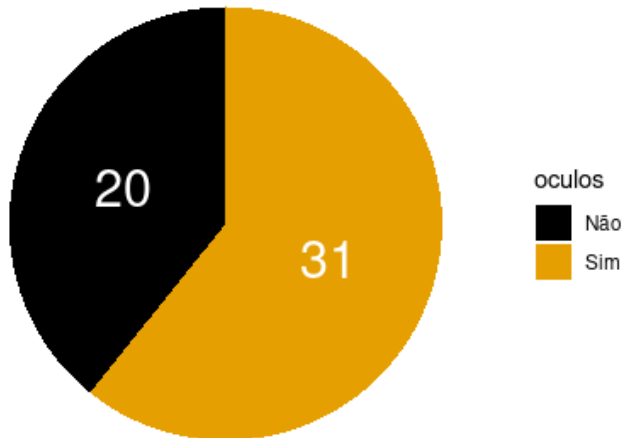


Gráfico de Pontos

```
survAluno_alterado |>  
  ggplot2::ggplot() +  
  ggplot2::aes(x = altura, y = peso) +  
  ggplot2::geom_point(colour = "#011e5a") +  
  ggplot2::xlab("Altura") +  
  ggplot2::ylab("Peso") +  
  ggplot2::labs(title = "Relação Peso X Altura")
```



Gráfico de Pontos

```
survAluno_alterado |>  
  ggplot2::ggplot() +  
  ggplot2::aes(x = altura, y = n_calçado) +  
  ggplot2::geom_point(colour = "#011e5a",  
                      fill = "lightblue") +  
  
  ggplot2::xlab("Altura") +  
  ggplot2::ylab("Nº Calçado") +  
  ggplot2::labs(title = "Relação Nº de Calçado X Altura") +  
  ggplot2::geom_smooth(method = "lm", se = FALSE)
```

