

# Introdução à Linguagem R

Encontro 3/4

*Prof. Davi Moreira*

*29 de Abril, 2019*

## Sumário

<b>1</b>	<b>Encontro 3</b>	<b>2</b>
1.1	Dúvidas e revisão do conteúdo do encontro prévio . . . . .	2
1.2	Estrutura da encontro 3 . . . . .	2
<b>2</b>	<b>Dados para o encontro</b>	<b>2</b>
<b>3</b>	<b>Estatística Descritiva</b>	<b>5</b>
3.1	Medidas de posição . . . . .	5
3.2	Medidas de dispersão . . . . .	5
3.3	Atividade prática . . . . .	5
<b>4</b>	<b>Variáveis aleatórias</b>	<b>5</b>
4.1	Variáveis aleatórias discretas . . . . .	6
4.2	Variáveis aleatórias contínuas . . . . .	7
<b>5</b>	<b>Inferência Estatística</b>	<b>7</b>
5.1	Coefficiente de correlação . . . . .	7
<b>6</b>	<b>Atividade Prática</b>	<b>11</b>
<b>7</b>	<b>Comunicando nossas análises</b>	<b>11</b>
<b>8</b>	<b>Gráficos com o ggplot2</b>	<b>12</b>
8.1	Gráficos de dispersão . . . . .	12
8.2	Gráficos de coluna . . . . .	13
8.3	Histograma . . . . .	14
8.4	Gráficos de linha . . . . .	14
8.5	Faceting . . . . .	15
8.6	Box-plot . . . . .	16
<b>9</b>	<b>Relatórios</b>	<b>17</b>
9.1	Shiny . . . . .	17
9.2	R Markdown . . . . .	17

# 1 Encontro 3

## 1.1 Dúvidas e revisão do conteúdo do encontro prévio

- 15 minutos serão reservados para dúvidas e revisão do conteúdo do encontro prévio.

## 1.2 Estrutura da encontro 3

### 4. ESTATÍSTICA BÁSICA, VISUALIZAÇÃO DE DADOS E REPORTING

- Medidas de posição e dispersão;
- Medidas de associação: o coeficiente de correlação de Pearson; - Testes de médias e proporções;
- Tidyverse tools: Pacote ggplot2
- gráficos de colunas;
- boxplot;
- gráficos de linha;
- faceting;
- formatação;
- reporting (R Markdown)

Até o final do encontro o aluno deverá ser capaz de:

- Carregar bases de dados e realizar análises exploratórias
- Obter estatísticas básicas
- Produzir gráficos que permitam análise dos dados;
- Produzir relatórios usando o RMarkdown;

# 2 Dados para o encontro

```
# definindo diretório
setwd("./dados/")

# carregando arquivos CENSO ESCOLAR 2016
load("matricula_pe_censo_escolar_2016.RData")
load("docentes_pe_censo_escolar_2016.RData")
load("turmas_pe_censo_escolar_2016.RData")
load("escolas_pe_censo_escolar_2016.RData")

# carregando dados PNUD
if(require(tidyverse) == F) install.packages('tidyverse'); require(tidyverse)
if(require(readxl) == F) install.packages('readxl'); require(readxl)

setwd("./dados/")
pnud <- read_excel("atlas2013_dadosbrutos_pt.xlsx", sheet = 2)
head(pnud)
unique(pnud$ANO)

# selecionando dados de 2010 e do Estado de Pernambuco
pnud_pe_2010 <- pnud %>% filter(ANO == 2010 & UF == 26)

rm(pnud) # removendo base pnud

# Processando bases de dados do CENSO ESCOLAR conforme enunciado e adicionando
```

```

# outras variáveis

# Turmas
turmas_pe_sel <- turmas_pe %>% group_by(CO_MUNICIPIO) %>%
  summarise(n_turmas = n(),
            turmas_disc_prof = sum(IN_DISC_PROFISSIONALIZANTE, na.rm = T),
            turmas_disc_inf = sum(IN_DISC_INFORMATICA_COMPUTACAO, na.rm = T),
            turmas_disc_mat = sum(IN_DISC_MATEMATICA, na.rm = T),
            turmas_disc_pt = sum(IN_DISC_LINGUA_PORTUGUESA, na.rm = T),
            turmas_disc_en = sum(IN_DISC_LINGUA_INGLES, na.rm = T))

# verificacao
dim(turmas_pe_sel)[1] == length(unique(turmas_pe$CO_MUNICIPIO))
summary(turmas_pe_sel)

# Escolas
escolas_pe_sel <- escolas_pe %>% group_by(CO_MUNICIPIO) %>%
  summarise(n_escolas = n(),
            n_escolas_priv = sum(TP_DEPENDENCIA == 4, na.rm = T),
            escolas_func = sum(TP_SITUACAO_FUNCIONAMENTO == 1, na.rm = T),
            escolas_agua_inex = sum(IN_AGUA_INEXISTENTE, na.rm = T),
            escolas_energia_inex = sum(IN_ENERGIA_INEXISTENTE, na.rm = T),
            escolas_esgoto_inex = sum(IN_ESGOTO_INEXISTENTE, na.rm = T),
            escolas_internet = sum(IN_INTERNET, na.rm = T),
            escolas_alimentacao = sum(IN_ALIMENTACAO, na.rm = T))

# verificacao
dim(escolas_pe_sel)[1] == length(unique(escolas_pe$CO_MUNICIPIO))
summary(escolas_pe_sel)

# Docentes
docentes_pe_sel <- docentes_pe %>% group_by(CO_MUNICIPIO) %>%
  summarise(n_docentes = n(),
            docentes_media_idade = mean(NU_IDADE),
            docentes_fem_sx = sum(TP_SEXO == 2, na.rm = T),
            docentes_superior = sum(TP_ESCOLARIDADE == 4, na.rm = T),
            docentes_contrato = sum(TP_TIPO_CONTRATACAO %in% c(1, 4), na.rm = T)
            )

# verificacao
dim(docentes_pe_sel)[1] == length(unique(docentes_pe$CO_MUNICIPIO))
summary(docentes_pe_sel)

# Matrículas
matriculas_pe_sel <- matricula_pe %>% group_by(CO_MUNICIPIO) %>%
  summarise(n_matriculas = n(),
            alunos_media_idade = mean(NU_IDADE),
            alunos_fem_sx = sum(TP_SEXO == 2, na.rm = T),
            alunos_negros = sum(TP_COR_RACA %in% c(2, 3), na.rm = T),
            alunos_indigenas = sum(TP_COR_RACA == 5, na.rm = T),
            alunos_cor_nd = sum(TP_COR_RACA == 0, na.rm = T),
            matriculas_educ_inf = sum(TP_ETAPA_ENSINO %in% c(1, 2), na.rm = T),
            matriculas_educ_fund = sum(TP_ETAPA_ENSINO %in% c(4:21, 41), na.rm = T),
            matriculas_educ_medio = sum(TP_ETAPA_ENSINO %in% c(25:38), na.rm = T))

```

```

)

# verificacao
dim(matriculas_pe_sel)[1] == length(unique(matricula_pe$CO_MUNICIPIO))
summary(matriculas_pe_sel)

# UNINDO BASES CENSO E PNUD -----

# matriculas
censo_pnud_pe_sel <- pnud_pe_2010 %>% full_join(matriculas_pe_sel,
                                                by = c("Codmun7" = "CO_MUNICIPIO")
                                                )

dim(pnud_pe_2010)
dim(matriculas_pe_sel)
dim(censo_pnud_pe_sel)
names(censo_pnud_pe_sel)

# escolas
censo_pnud_pe_sel <- censo_pnud_pe_sel %>% full_join(escolas_pe_sel,
                                                    by = c("Codmun7" = "CO_MUNICIPIO")
                                                    )

dim(escolas_pe_sel)
dim(censo_pnud_pe_sel)
names(censo_pnud_pe_sel)

# turmas
censo_pnud_pe_sel <- censo_pnud_pe_sel %>% full_join(turmas_pe_sel,
                                                    by = c("Codmun7" = "CO_MUNICIPIO")
                                                    )

dim(turmas_pe_sel)
dim(censo_pnud_pe_sel)
names(censo_pnud_pe_sel)

# docentes
censo_pnud_pe_sel <- censo_pnud_pe_sel %>% full_join(docentes_pe_sel,
                                                    by = c("Codmun7" = "CO_MUNICIPIO")
                                                    )

dim(docentes_pe_sel)
dim(censo_pnud_pe_sel)
names(censo_pnud_pe_sel)

# salvando nova base -----
setwd("./dados")
save(censo_pnud_pe_sel, file = "2016_censo_pnud_pe_sel.RData")
write.csv2(censo_pnud_pe_sel, file = "2016_censo_pnud_pe_sel.csv",
           row.names = F)

rm(list = ls()) # limpando area de trabalho

# carregando nova base -----
setwd("./dados")
load("2016_censo_pnud_pe_sel.RData")

```

## 3 Estatística Descritiva

### 3.1 Medidas de posição

```
# Média Aritmética
mean(censo_pnud_pe_sel$n_matriculas)

# Mediana
median(censo_pnud_pe_sel$n_matriculas)

# Moda
y <- c(sample(1:10, 100, replace = T))
table(y)
table(y)[which.max(table(y))]

# Quantis
summary(censo_pnud_pe_sel$n_matriculas)

# Percentis / Decis...
?quantile
quantile(censo_pnud_pe_sel$n_matriculas, probs = seq(0,1, .01))
```

### 3.2 Medidas de dispersão

```
# Amplitude
max(censo_pnud_pe_sel$n_matriculas) - min(censo_pnud_pe_sel$n_matriculas)
# ou
range(censo_pnud_pe_sel$n_matriculas)[2] - range(censo_pnud_pe_sel$n_matriculas)[1]

# Variância
var(y)

# Desvio padrão
sd(y)
# ou
y %>% var %>% sqrt

# Coeficiente de variação
100*sd(y)/mean(y)
```

### 3.3 Atividade prática

Utilizando a variável IDHM, calcule: - a média - os decis - o segundo quartil ou mediana - a amplitude da amostra - a variância e o desvio padrão da amostra.

## 4 Variáveis aleatórias

Esta seção está baseada na apostila do [Minicurso de Estatística Básica: Minicurso de Estatística Básica: Introdução ao software R](#). O link [Univariate Distribution Relationships](#) apresenta o modo como as principais

distribuições se relacionam e pode ser útil para esta seção.

## 4.1 Variáveis aleatórias discretas

### 4.1.1 Distribuição Binomial

Um experimento binomial é experimento aleatório que consiste em repetidas tentativas que apresentam apenas dois resultados possíveis (sucesso ou fracasso) e possui as seguintes características: - As tentativas são independentes, ou seja, o resultado de uma não altera o resultado da outra; - Cada repetição do experimento admite apenas dois resultados: sucesso ou fracasso; - A probabilidade de sucesso ( $p$ ), em cada tentativa, é constante.

A variável aleatória  $X$  denota o número de tentativas que resultaram em sucesso e possui uma distribuição binomial com parâmetros  $p$  e  $n = 1, 2, 3, \dots$

```
# Distribuição Binomial

p <- 0.25 # probabilidade
n <- 100  # número de tentativas
x <- 20   # número de sucessos em n tentativas

dbinom(x, n, p)

x <- c(0:50)
bin <- dbinom(x, n, p)
plot(x, bin, type = "h", xlab = "Sucessos", ylab = "Probabilidade",
     main = "Distribuição binomial")
```

### 4.1.2 Distribuição De Poisson

A distribuição de Poisson expressa experimentos em que o número de amostras pode aumentar no tempo e a probabilidade de sucesso diminuir, mantendo a esperança  $E(X)$  constante. A variável aleatória  $X$  denota o número de contagens no intervalo.

```
x <- 2
lambda <- 2.3

# distribuição de Poisson com parâmetros x e lambda:
dpois(x, lambda)

x <- 0:10
poisson <- dpois(x, lambda)

plot(x, poisson, xlab = "Número de eventos no tempo",
     ylab = "Probabilidade de Poisson", main = "Distribuição de Poisson")
```

### 4.1.3 Atividade prática

Num determinado posto de gasolina, dados coletados indicam que um número médio de 6 clientes por hora param para colocar gasolina numa bomba.

- Qual é a probabilidade de 3 clientes pararem qualquer hora?
- Qual é a probabilidade de 3 clientes ou menos pararem em qualquer hora?
- Qual é o valor esperado, a média, e o desvio padrão para esta distribuição?

## 4.2 Variáveis aleatórias contínuas

### 4.2.1 Distribuição Normal

Um pesquisador coletou os dados da estatura de jovens em idade de alistamento militar. Sabe-se que a estatura de uma população segue a distribuição normal, com média 170 cm e variância 36 cm (desvio padrão de 6 cm).

- Qual a probabilidade de se encontrar um jovem com mais de 1,79 m de altura?

```
1-pnorm(179, 170, 6) # pnorm(x, média, desvio padrão)
```

- Qual a altura em que a probabilidade de encontrarmos valores menores que ela seja de 80%?

```
qnorm(0.8, 170, 6)
```

## 5 Inferência Estatística

Não sendo objetivo do curso, testes de médias e proporções não serão tópicos do encontro. Contudo, sendo de interesse, é fortemente recomendável que o aluno verifique o capítulo 7 da apostila do [Minicurso de Estatística Básica: Minicurso de Estatística Básica: Introdução ao software R](#).

### 5.1 Coeficiente de correlação

Para o conteúdo desta seção, foi feito o uso do conteúdo do R Correlation Tutorial.

```
cor(x, y, method = c("pearson", "kendall", "spearman"))
cor.test(x, y, method=c("pearson", "kendall", "spearman"))

movies <- read.csv(url("http://s3.amazonaws.com/dcwoods2717/movies.csv"))
head(movies)
str(movies)

# criando variável profit (lucro)
movies <- movies %>% mutate(profit = gross - budget)

# grafico de dispersao
plot(movies$rating, movies$profit)

# correlacao
cor(movies$rating, movies$profit)

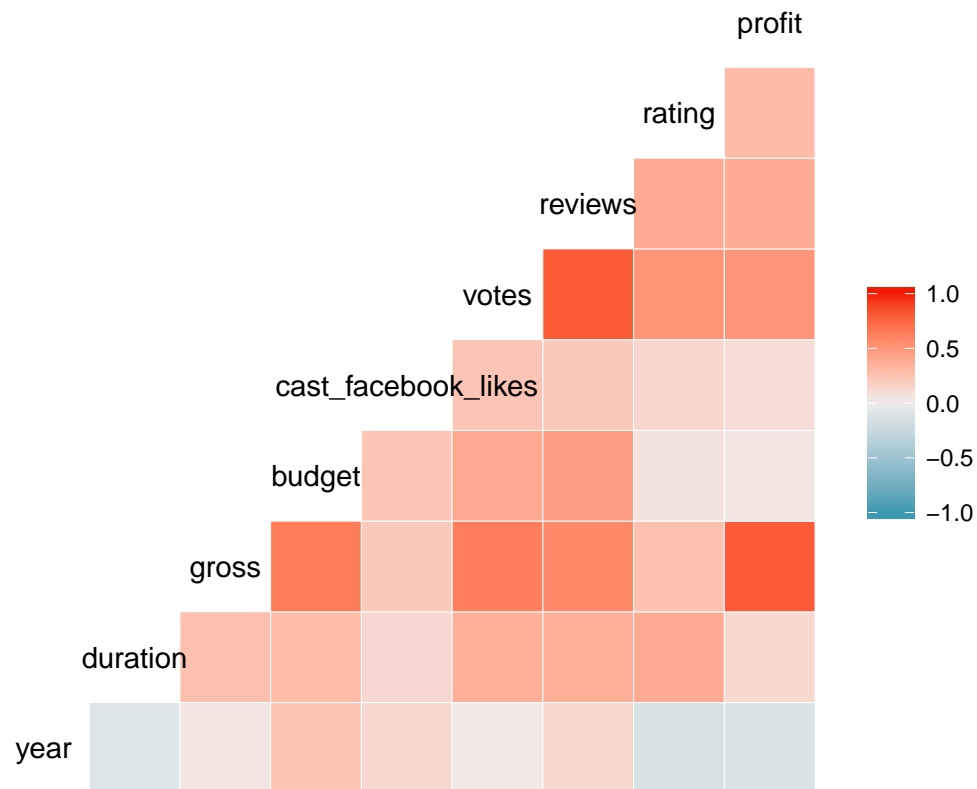
# teste de correlacao
cor.test(movies$rating, movies$profit) # p-valor < .00000000000000022

if(require(GGally) == F) install.packages('GGally'); require(GGally)

if(require(tidyverse) == F) install.packages('tidyverse'); require(tidyverse)
if(require(GGally) == F) install.packages('GGally'); require(GGally)

movies <- read.csv(url("http://s3.amazonaws.com/dcwoods2717/movies.csv"))
movies <- movies %>% mutate(profit = gross - budget)

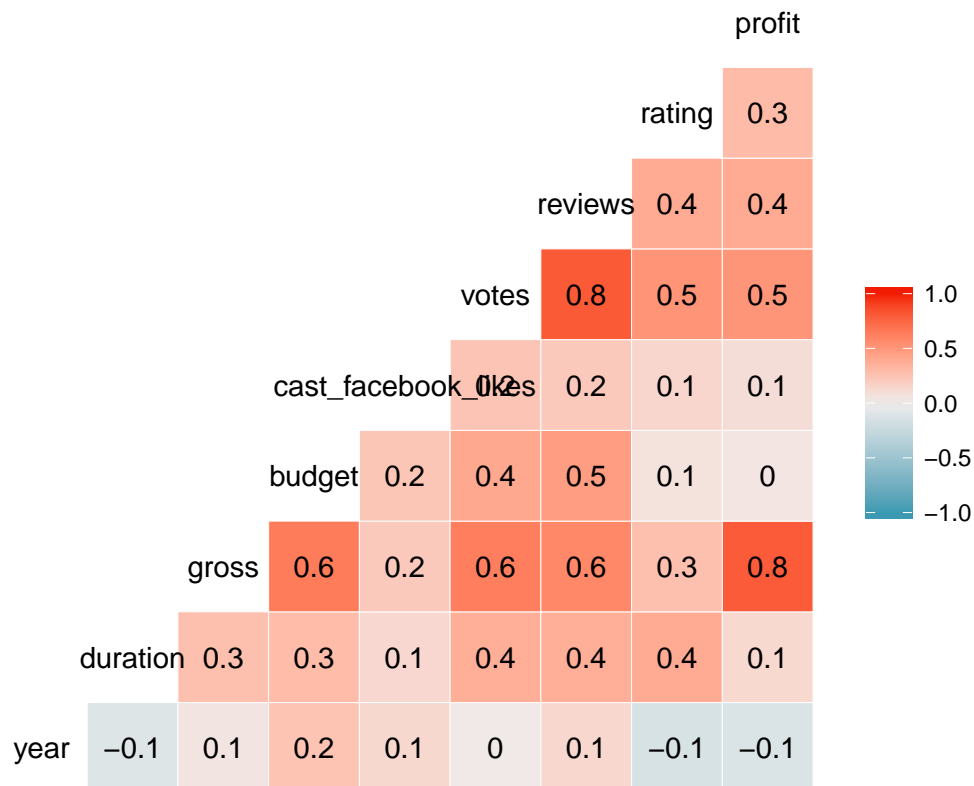
# correlacao
ggcorr(movies[, c(4:12)])
```



```
if(require(GGally) == F) install.packages('GGally'); require(GGally)

# correlacao
ggcorr(movies[, c(4:12)], label = T)
```



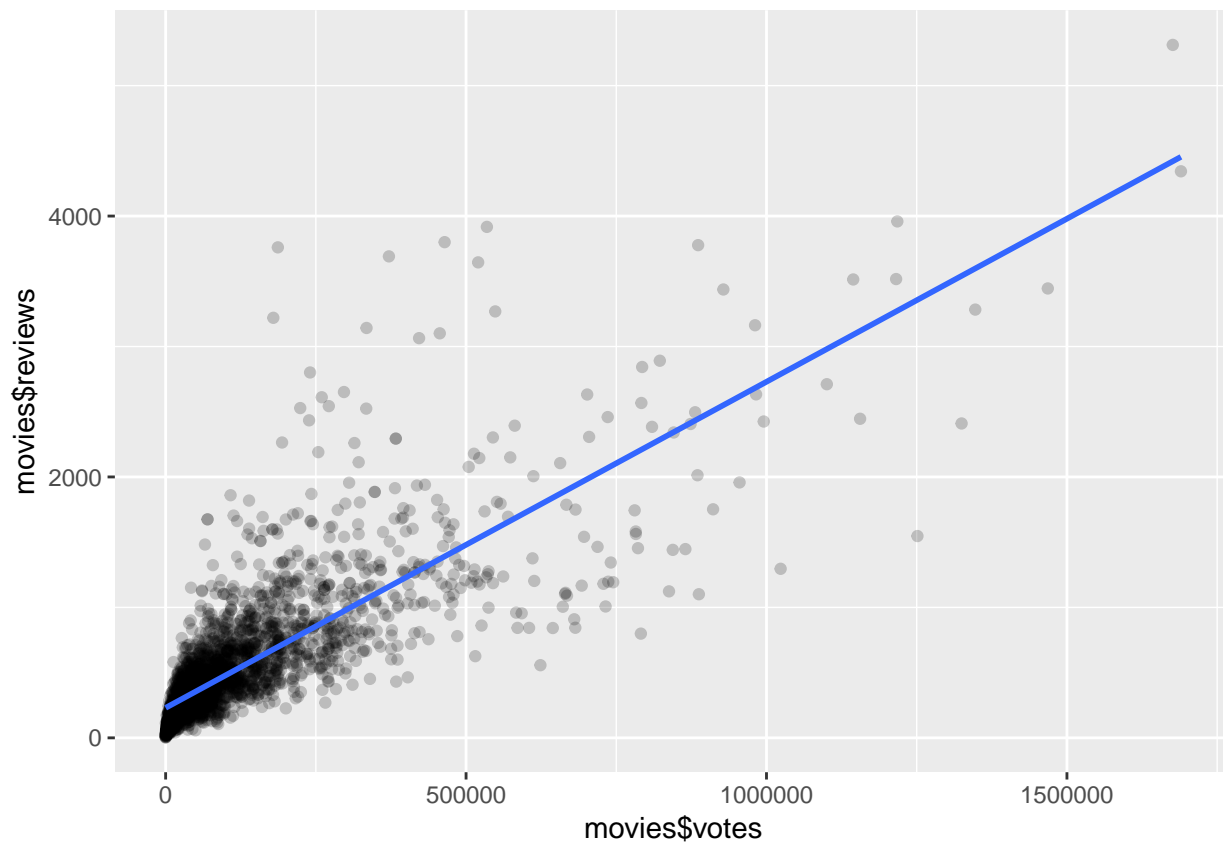


```
if(require(tidyverse) == F) install.packages('tidyverse'); require(tidyverse)

# Forte correlação positiva:

# Plot votes vs reviews
qplot(movies$votes,
      movies$reviews,
      data = movies,
      geom = c("point", "smooth"),
      method = "lm",
      alpha = I(1 / 5),
      se = F)
```

```
## Warning: Ignoring unknown parameters: method, se
```

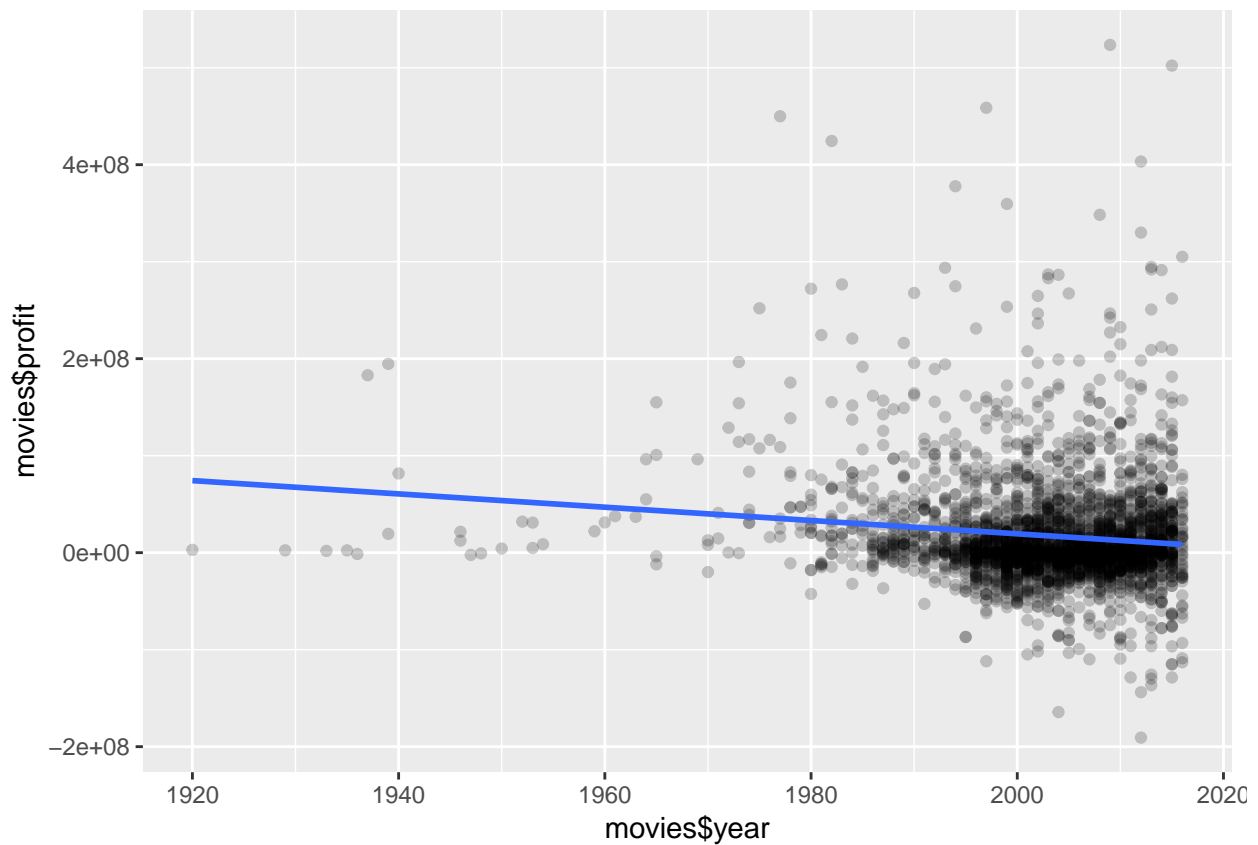


```
if(require(tidyverse) == F) install.packages('tidyverse'); require(tidyverse)

# Fraca correlação positiva:

# Plot profit over years
qplot(movies$year,
      movies$profit,
      data = movies,
      geom = c("point", "smooth"),
      method = "lm",
      alpha = I(1 / 5),
      se = F)
```

```
## Warning: Ignoring unknown parameters: method, se
```



## 6 Atividade Prática

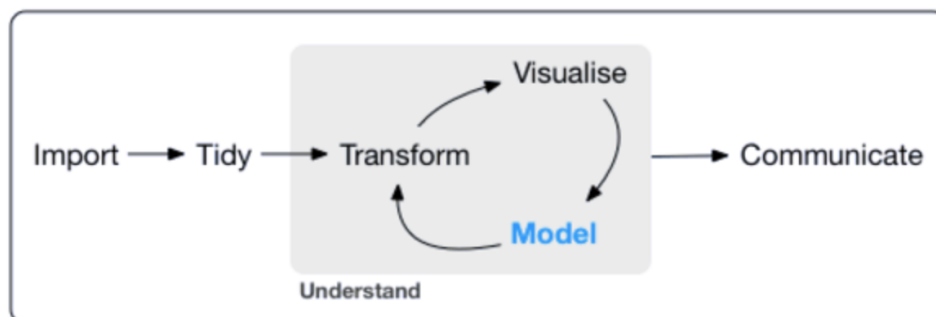
- Utilize a base `censo_pnud_pe_sel` para selecionar mais de duas variáveis contínuas de interesse e produzir graficamente a matriz de correlação, apresentando o coeficiente de correlação calculado.

## 7 Comunicando nossas análises

A comunicação da análise de dados é uma etapa tão importante que pode destruir todo um trabalho desenvolvido. É a partir dela que o público em geral, tomadores de decisão, especialistas e toda uma comunidade tem a possibilidade de compartilhar uma visão comum de aspectos complexos envolvidos na análise de dados.

```
## Loading required package: png
```

```
## Loading required package: grid
```



## 8 Gráficos com o ggplot2

Criador do pacote `ggplot2`, [Hadley Wickham](#) estabelece uma nova definição sobre o que é um gráfico. Em [A Layered Grammar of Graphics](#), sugere que os principais aspectos de um gráfico (dados, sistema de coordenadas, rótulos e anotações) podem ser divididos em camadas. É assim que o pacote `ggplot2` funciona e é com essa teoria em mente que seguiremos nessa aula.

Entre outras muitas referências disponíveis, estamos utilizando principalmente as seguintes: \* Pacote `ggplot2`  
\* Data Visualisation \* CursoR: `ggplot2`

```
if(require(ggplot2) == F) install.packages("ggplot2"); require(ggplot2)
if(require(tidyverse) == F) install.packages("tidyverse"); require(tidyverse)
if(require(lubridate) == F) install.packages("lubridate"); require(lubridate)
if(require(scales) == F) install.packages("scales"); require(scales)
```

### 8.1 Gráficos de dispersão

```
ggplot(data = censo_pnud_pe_sel, aes(x = n_matriculas, y = n_docentes) ) +
  geom_point(color = "red", size = 2) +
  labs(x = "Número de Matrículas", y = "Número de Docentes")
```

Observe que:

- a primeira camada é dada pela função `ggplot()` e recebe um `data frame`;
- a segunda camada é dada pela função `geom_point()`, especificando a forma geométrica utilizada no mapeamento das observações;
- as camadas são somadas com um `+`;
- o mapeamento na função `geom_point()` recebe a função `aes()`, responsável por descrever como as variáveis serão mapeadas nos aspectos visuais da forma geométrica escolhida, no caso, pontos.

Também usamos os argumentos mais comuns:

- `color` =: altera a cor de formas que não têm área (pontos e retas).
- `fill` =: altera a cor de formas com área (barras, caixas, densidades, áreas).
- `size` =: altera o tamanho de formas.
- `type` =: altera o tipo da forma, geralmente usada para pontos.
- `linetype` =: altera o tipo da linha no caso de gráficos de linha

A combinação da função `ggplot()` e de uma ou mais funções `geom_()` definirá o tipo de gráfico gerado.

```
# outro exemplo e matriz de correlacao
# pacotes
if(require(tidyverse) == F) install.packages('tidyverse'); require(tidyverse)
if(require(GGally) == F) install.packages('GGally'); require(GGally)

# dados
movies <- read.csv(url("http://s3.amazonaws.com/dcwoods2717/movies.csv"))
head(movies)
str(movies)

# criando variável profit (lucro)
movies <- movies %>% mutate(profit = gross - budget)

# Forte correlação positiva:

# Plot votes vs reviews
```

```

qplot(movies$votes,
      movies$reviews,
      data = movies,
      geom = c("point", "smooth"),
      method = "lm",
      alpha = I(1 / 5),
      se = F)

# Fraca correlação positiva:

# Plot profit over years
qplot(movies$year,
      movies$profit,
      data = movies,
      geom = c("point", "smooth"),
      method = "lm",
      alpha = I(1 / 5),
      se = F)

# correlacao
ggcorr(movies[, c(4:12)])

# correlacao
ggcorr(movies[, c(4:12)], label = T)

```

## 8.2 Gráficos de coluna

```

# cor / raça docentes
ggplot(docentes_pe, aes(as.factor(TP_COR_RACA))) +
  geom_bar()

# trabalhando com fatores

tamanho <- factor(c("pequeno", "grande", "médio", "pequeno", "médio"))
tamanho

tamanho <- factor(c("pequeno", "grande", "médio", "pequeno", "médio"),
                  levels = c("pequeno", "médio", "grande"))
tamanho

# verificando base censo_pnud

class(censo_pnud_pe_sel$Município)
censo_pnud_pe_sel$Município <- as.factor(censo_pnud_pe_sel$Município)
censo_pnud_pe_sel$Município

# ordenando factors por uma variável
censo_pnud_pe_sel$Município <- factor(censo_pnud_pe_sel$Município,
  censo_pnud_pe_sel$IDHM,
  decreasing = T))

```

- Atividade em aula:

Faça novamente o gráfico de barras com a variável TP\_COR\_RACA do banco `docentes`, mas de modo que a categoria 0 seja a última do lado direito do gráfico.

### 8.3 Histograma

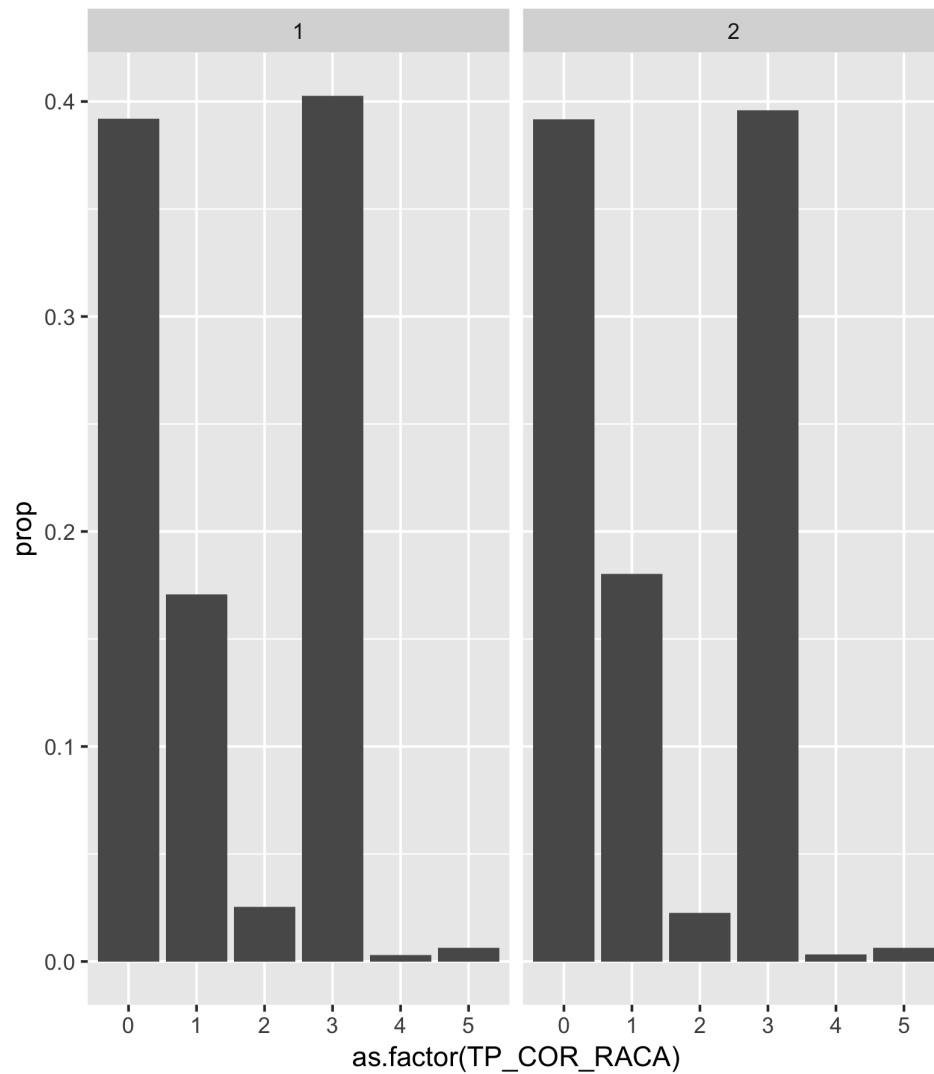
```
ggplot(censo_pnud_pe_sel, aes(n_escolas)) +  
  geom_histogram()  
  
ggplot(censo_pnud_pe_sel, aes(n_escolas)) +  
  geom_histogram(binwidth = 10)
```

### 8.4 Gráficos de linha

```
# ajustando dados  
matricula_pe_nasc <- matricula_pe %>%  
  select(NU_MES, NU_ANO) %>%  
  mutate(nascimento = make_date(NU_ANO, NU_MES)) %>%  
  filter(nascimento >= "2000-01-01" & nascimento <= "2009-01-01") %>%  
  group_by(nascimento) %>%  
  summarise(n_matriculas = n())  
  
summary(matricula_pe_nasc)  
  
ggplot(matricula_pe_nasc, aes(nascimento, n_matriculas)) +  
  geom_line() +  
  scale_x_date(labels = date_format("%m-%Y"))  
  
ggplot(matricula_pe_nasc, aes(nascimento, n_matriculas)) +  
  geom_line() +  
  scale_x_date(labels = date_format("%b-%Y"),  
              breaks = date_breaks("6 months"))  
  
ggplot(matricula_pe_nasc, aes(nascimento, n_matriculas)) +  
  geom_line() +  
  scale_x_date(labels = date_format("%b-%Y"),  
              breaks = date_breaks("year"))  
  
ggplot(matricula_pe_nasc, aes(nascimento, n_matriculas)) +  
  geom_line() +  
  scale_x_date(labels = date_format("%b-%Y"),  
              breaks = date_breaks("2 years"))  
  
ggplot(matricula_pe_nasc, aes(nascimento, n_matriculas)) +  
  geom_line() +  
  scale_x_date(labels = date_format("%b-%Y"),  
              breaks = date_breaks("year")) +  
  theme(axis.text.x = element_text(colour='black', angle = 45, hjust = 1, vjust = 1,  
                                   face = "bold"))
```

## 8.5 Faceting

```
ggplot(docentes_pe, aes(as.factor(TP_COR_RACA))) +  
  geom_bar()  
  
docentes_pe_sel <- docentes_pe %>%  
  select(TP_SEX0, TP_COR_RACA) %>%  
  group_by(TP_SEX0, TP_COR_RACA) %>%  
  summarise(n_docentes = n()) %>%  
  mutate(prop = n_docentes/sum(n_docentes))  
  
docentes_pe_sel  
  
ggplot(docentes_pe_sel, aes(as.factor(TP_COR_RACA), y = prop)) +  
  geom_bar(stat = "identity") + facet_wrap(~TP_SEX0)  
  
matriculas_pe_sel <- matricula_pe %>%  
  select(TP_SEX0, TP_COR_RACA) %>%  
  group_by(TP_SEX0, TP_COR_RACA) %>%  
  summarise(n_matriculas = n()) %>%  
  mutate(prop = n_matriculas/sum(n_matriculas))  
  
matriculas_pe_sel  
  
graph <- ggplot(matriculas_pe_sel, aes(as.factor(TP_COR_RACA), y = prop)) +  
  geom_bar(stat = "identity") + facet_wrap(~TP_SEX0)  
  
setwd("./imagens")  
ggsave(filename = "matriculas_cor_raca_sx.png", plot = graph)
```



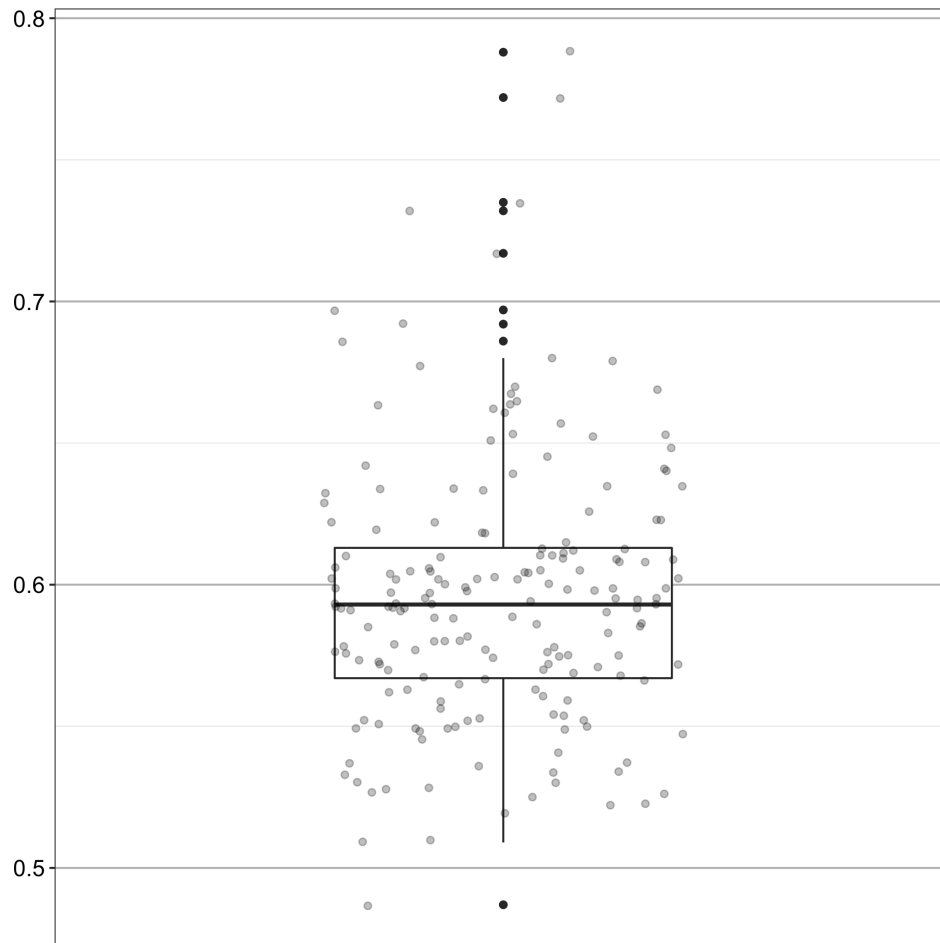
## 8.6 Box-plot

```
# mtcars
ggplot(mtcars) +
  geom_boxplot(aes(x = as.factor(cyl), y = mpg))

# censo_pnud
graph <- ggplot(censo_pnud_pe_sel, aes(0, IDHM)) +
  geom_boxplot() +
  scale_x_discrete() +
  geom_jitter(alpha = .25, size = 1.5) +
  xlab(NULL) +
  ylab(NULL) +
  theme_bw() +
  theme(panel.grid.major = element_line(colour = "grey")) +
  theme(axis.text.y = element_text(colour='black', angle = 0, size = 12, hjust = 0, vjust = 0.5))
```



```
graph
setwd("./imagens")
ggsave(filename = "box_plot.png", plot = graph)
```



## 9 Relatórios

### 9.1 Shiny

O [Shiny](#) é um pacote R que facilita a criação de aplicativos Web interativos diretamente do R. Você pode hospedar aplicativos em uma página da Web, incorporá-los em documentos R Markdown ou criar painéis. Para conhecer o potencial dessa ferramenta, acesse [esta galeria disponível](#).

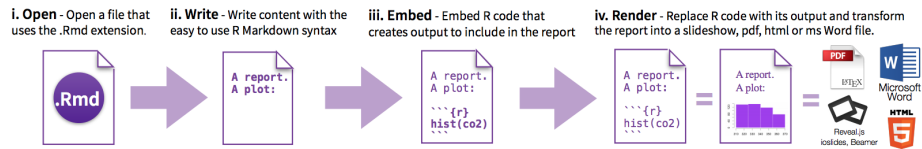
### 9.2 R Markdown

É com o R Markdown que foram produzidos todos os arquivos .pdf com o conteúdo do curso e é com ele que se faz possível desenvolver relatórios estáticos diretamente do R.

- Para começar, vamos olhar para os principais passos indicados no [cartão de dicas](#) do R Markdown.

## 9.2.1 Workflow

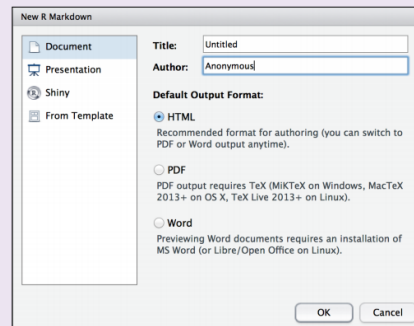
**1. Workflow** R Markdown is a format for writing reproducible, dynamic reports with R. Use it to embed R code and results into slideshows, pdfs, html documents, Word files and more. To make a report:



## 9.2.2 Começando um arquivo R Markdown

**2. Open File** Start by saving a text file with the extension .Rmd, or open an RStudio Rmd template

- In the menu bar, click **File ► New File ► R Markdown...**
- A window will open. Select the class of output you would like to make with your .Rmd file
- Select the specific type of output to make with the radio buttons (you can change this later)
- Click OK



## 9.2.3 Comandos básicos

Também é possível obter ajuda nos links abaixo:

- [Comandos básicos](#)
- [Tutorial](#)
- [R Markdown: The Definitive Guide](#)

**3. Markdown** Next, write your report in plain text. Use markdown syntax to describe how to format text in the final report.

#### syntax

Plain text  
 End a line with two spaces to start a new paragraph.  
`*italics*` and `_italics_`  
`**bold**` and `__bold__`  
`superscript^2^`  
`~~strikethrough~~`  
`[link](www.rstudio.com)`

`# Header 1`

`## Header 2`

`### Header 3`

`#### Header 4`

`##### Header 5`

`##### Header 6`

`endash: --`

`emdash: ---`

`ellipsis: ...`

`inline equation: $A = \pi * r^{2}$`

`image: `

`horizontal rule (or slide break):`

`***`

`> block quote`

`* unordered list`

`* item 2`  
`+ sub-item 1`  
`+ sub-item 2`

`1. ordered list`

`2. item 2`  
`+ sub-item 1`  
`+ sub-item 2`

`Table Header | Second Header`

`-----|-----`  
`Table Cell | Cell 2`  
`Cell 3 | Cell 4`

#### becomes

Plain text  
 End a line with two spaces to start a new paragraph.  
*italics* and *italics*  
**bold** and **bold**  
 superscript<sup>2</sup>  
 strikethrough  
[link](http://www.rstudio.com)

## Header 1

## Header 2

### Header 3

#### Header 4

#### Header 5

#### Header 6

endash: --

emdash: ---

ellipsis: ...

inline equation:  $A = \pi * r^2$

image: 

horizontal rule (or slide break):

block quote

- unordered list
- item 2
  - sub-item 1
  - sub-item 2

1. ordered list
2. item 2
  - sub-item 1
  - sub-item 2

Table Header      Second Header

Table Cell	Cell 2
Cell 3	Cell 4

### 9.2.4 Atividade em aula:

- Desenvolver relatório com os gráficos produzidos hoje.