

FIAP

NBA



# MBA EM DATA SCIENCE & AI

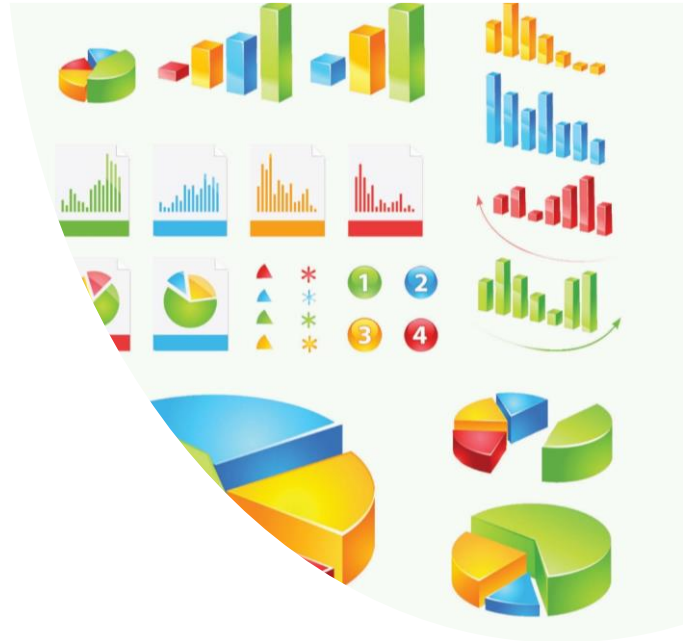
## STATISTICS WITH R

# AULA 3

## Plots no R

### Distribuição de probabilidade Normal





# Plots

Análises Gráficas  
com o R

# Por onde começar

- Você já conhece minimamente seus dados?
- Quais perguntas quer responder?

# Pacotes de plotagens mais conhecidos

- Base Plot (pacote graphics)
  - Core do R, possui os comandos básicos
- lattice
  - Permite criar o plot em uma única função, ótimo para subplots
- ggplot2
  - Mistura elementos do core com do lattice. Este é o pacote mais usado
- plotly
  - Pacote mais novo, com gráficos interativos em HTML.
  - Possui versão para python, matlab, node.js, etc.
- plot3D
  - advinha

# GGPLOT2



# GGPLOT2



- Construção de gráficos por camadas.
- Gráficos naturalmente mais bonitos;
- Fácil personalização (mais simples deixar o gráfico do jeito que você quer);
- A estrutura padronizada das funções deixa o aprendizado muito mais intuitivo;
- A diferença no código entre tipos diferentes de gráficos é muito pequena.



# Instalação do pacote

```
install.packages("ggplot2")  
library(ggplot2)
```

# Preparando o terreno...

```
#carregando a base
```

```
imdb <- readr::read_rds("imdb.rds")
```

```
#chamando o ggplot
```

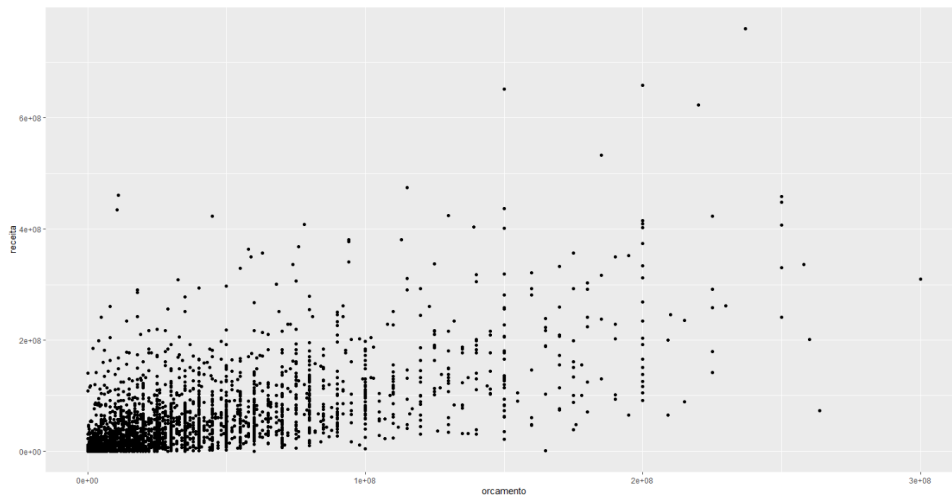
```
ggplot(data = imdb)
```

```
#Com o pipe
```

```
imdb %>% ggplot()
```

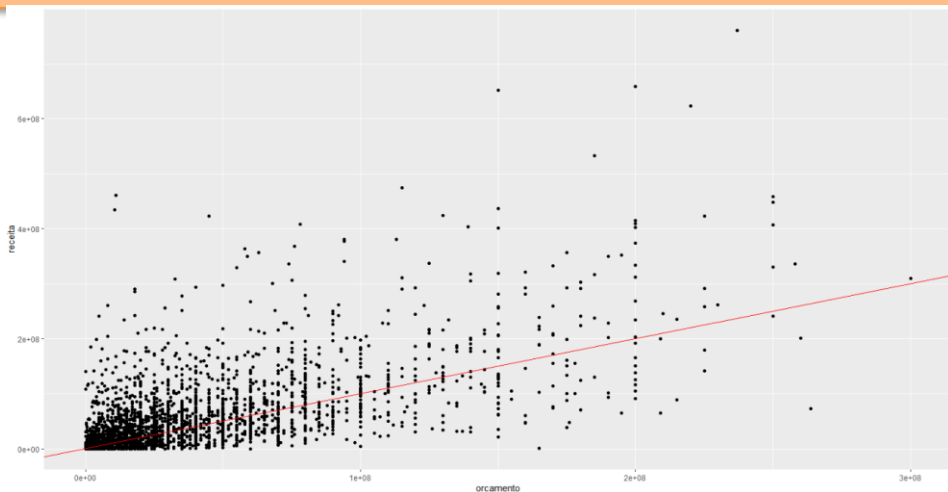
# Gráfico de Dispersão

```
ggplot(imdb) +  
  geom_point(mapping = aes(x =  
    orcamento, y = receita))
```



# Gráfico de Dispersão

```
ggplot(imdb) +  
  geom_point(mapping = aes(x = orcamento, y =  
    receita)) +  
  geom_abline(intercept = 0, slope = 1, color =  
    "red")
```



# Gráfico de Dispersão

```
ggplot(imdb) +  
  geom_abline(intercept = 0, slope = 1, color =  
    "red") + geom_point(mapping = aes(x = orcamento,  
    y = receita))
```

O que acontece se invertermos a ordem do código?

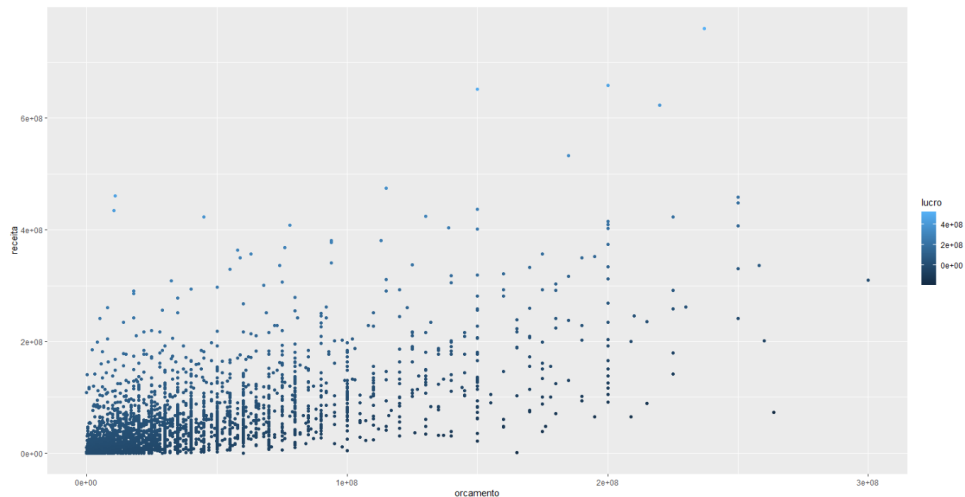
# Gráfico de Dispersão

```
imdb %>%
```

```
  mutate(lucro = receita - orcamento) %>%
```

```
  ggplot() +
```

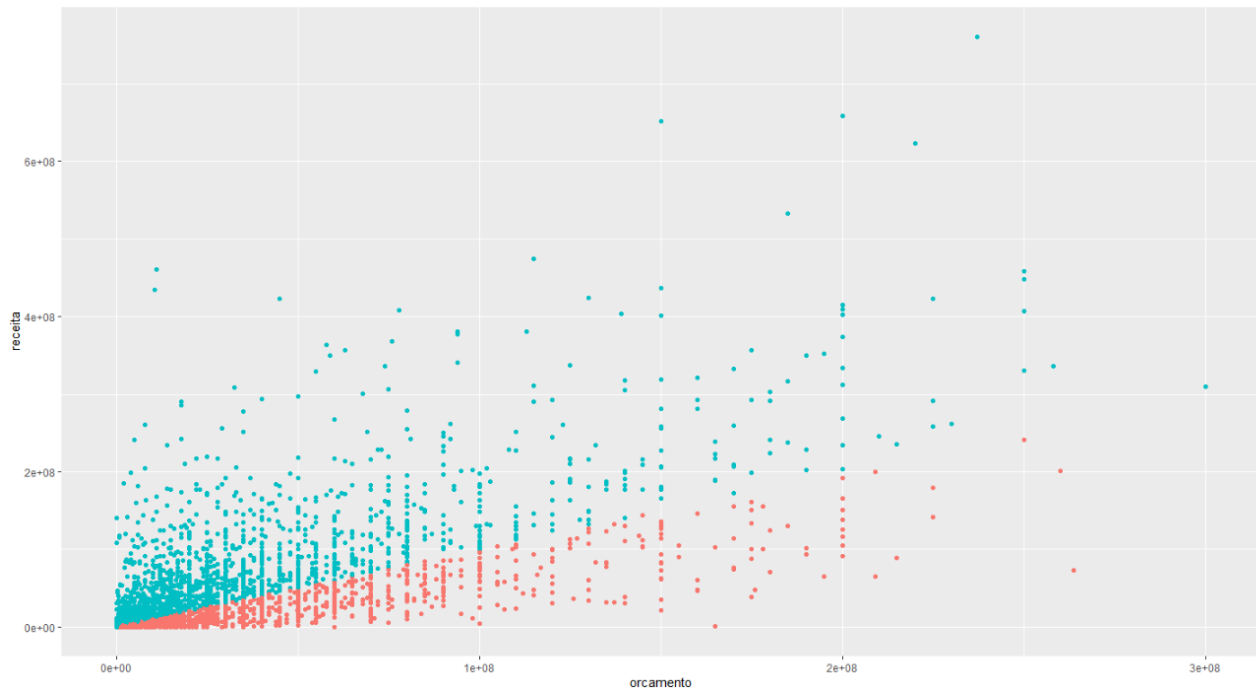
```
  geom_point(aes(x = orcamento, y = receita, color =  
lucro))
```



# Gráfico de Dispersão

```
imdb %>%  
  mutate(  
    lucro = receita - orcamento,  
    lucro = ifelse(lucro <= 0, "Não", "Sim")  
  ) %>%  
  filter(!is.na(lucro)) %>%  
  ggplot() +  
    geom_point(mapping = aes(x = orcamento, y =  
      receita, color = lucro))
```

# Gráfico de Dispersão





# Gráfico de Dispersão

Cuidado!!!

```
ggplot(imdb) +  
  geom_point(aes(x = orcamento, y = receita, color =  
    "blue"))
```

# Gráfico de Dispersão

Cuidado!!!

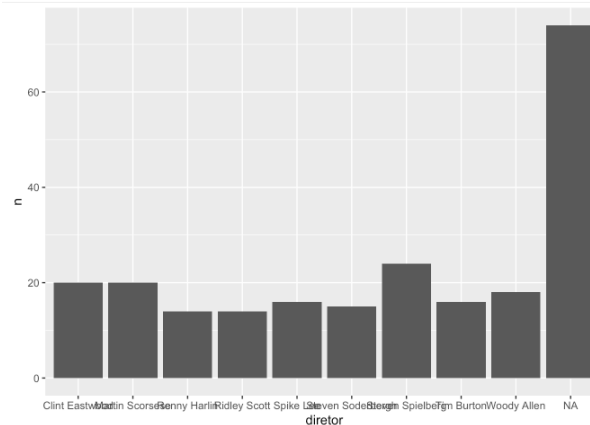
```
ggplot(imdb) +  
  geom_point(aes(x = orcamento, y = receita, color =  
    "blue"))
```

Color fica do lado de fora

```
ggplot(imdb) +  
  geom_point(aes(x = orcamento, y = receita), color  
    = "blue")
```

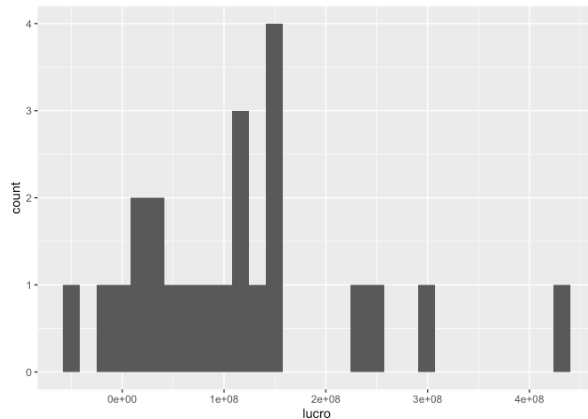
# Gráficos de barras

```
imdb %>%  
  count(diretor) %>%  
  top_n(10, n) %>%  
  ggplot() +  
  geom_col(aes(x = diretor, y = n))
```



# Histogramas

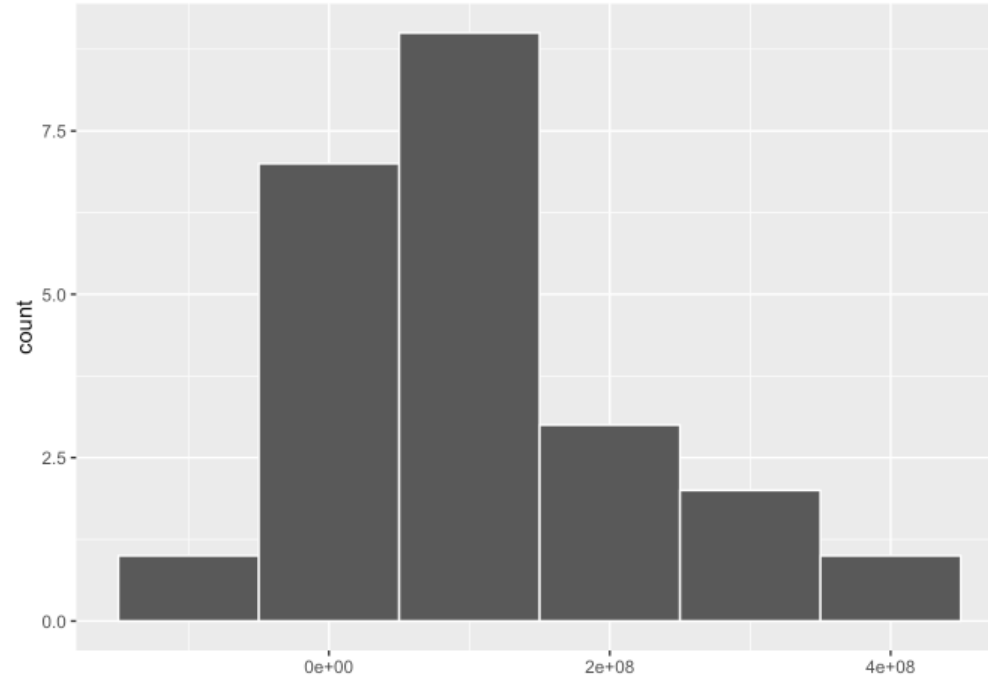
```
imdb %>%  
  filter(diretor == "Steven Spielberg") %>%  
  mutate(lucro = receita - orcamento) %>%  
  ggplot() +  
  geom_histogram(aes(x = lucro))
```



# Histogramas - binwidth

```
imdb %>%  
  filter(diretor == "Steven Spielberg") %>%  
  mutate(lucro = receita - orcamento) %>%  
  ggplot() +  
    geom_histogram(  
      aes(x = lucro),  
      binwidth = 1000000000,  
      color = "white"  
    )
```

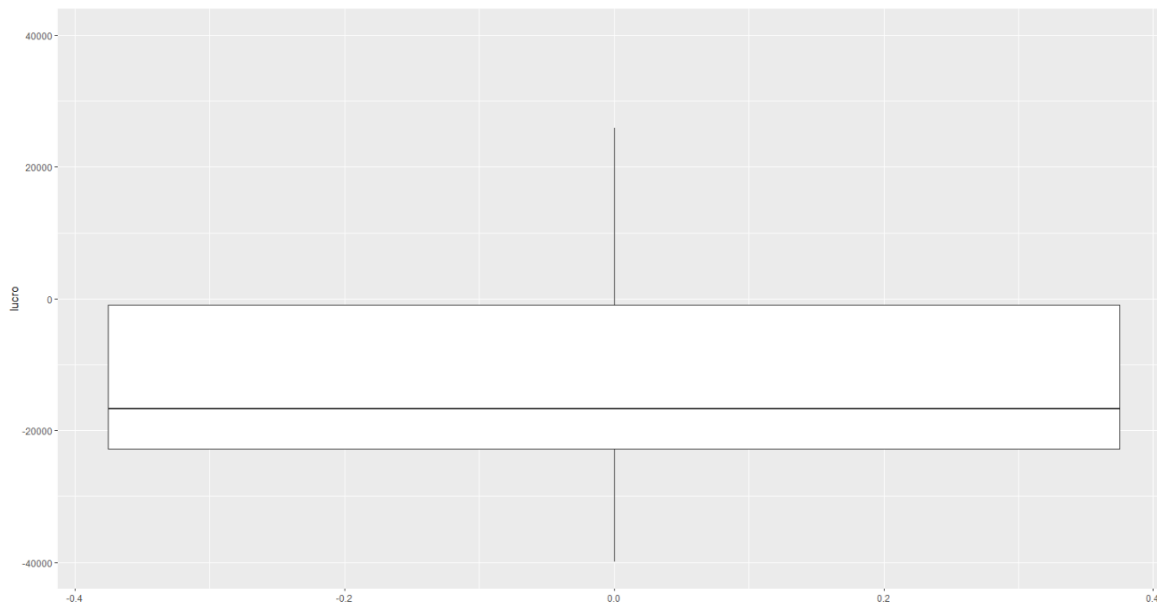
# Histogramas - binwidth



# Boxplots

```
imdb %>%  
  mutate(lucro = receita - orcamento) %>%  
  ggplot() +  
  geom_boxplot(aes(y = lucro)) +  
  ylim(c(-40000, 40000))
```

# Boxplots

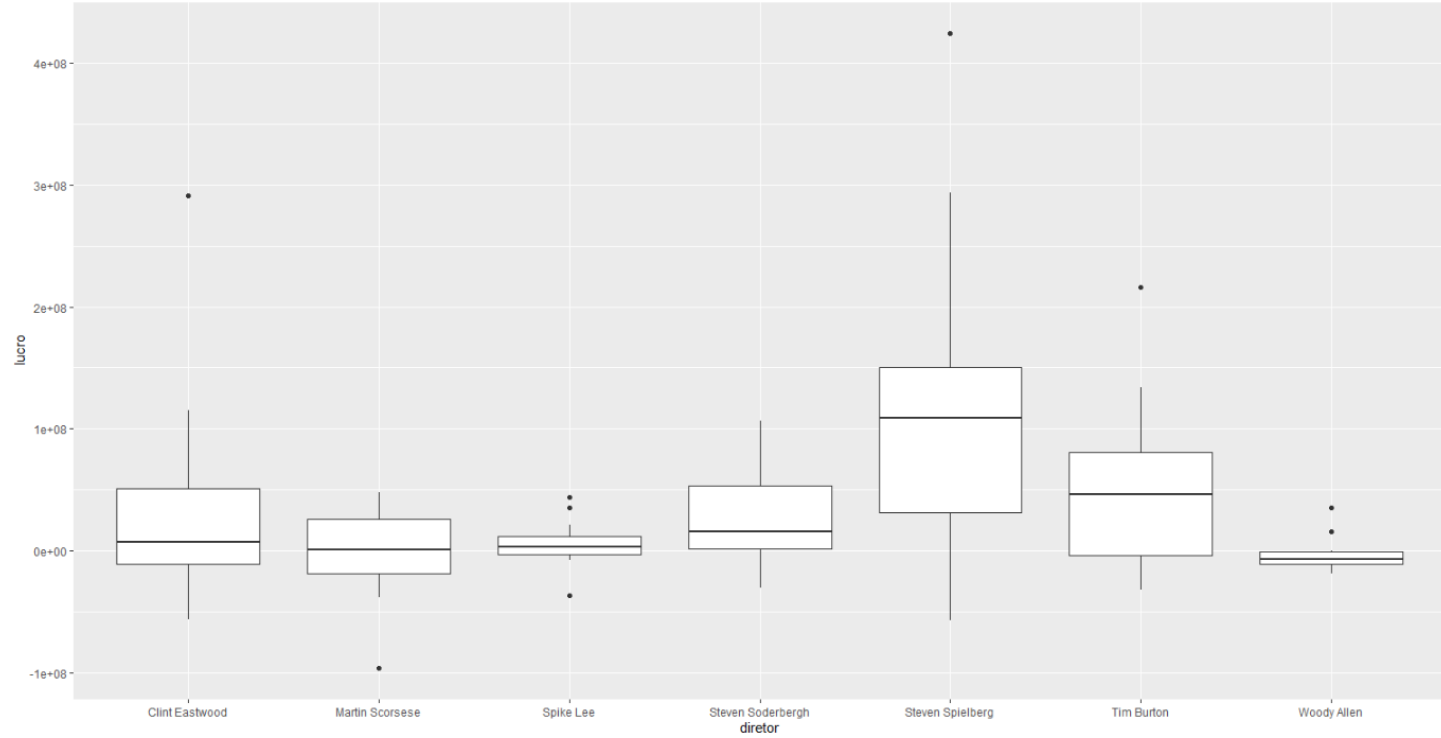




# Boxplots

```
imdb %>%  
  filter(!is.na(diretor)) %>%  
  group_by(diretor) %>%  
  filter(n() >= 15) %>%  
  mutate(lucro = receita - orcamento) %>%  
  ggplot() +  
  geom_boxplot(aes(x = diretor, y = lucro))
```

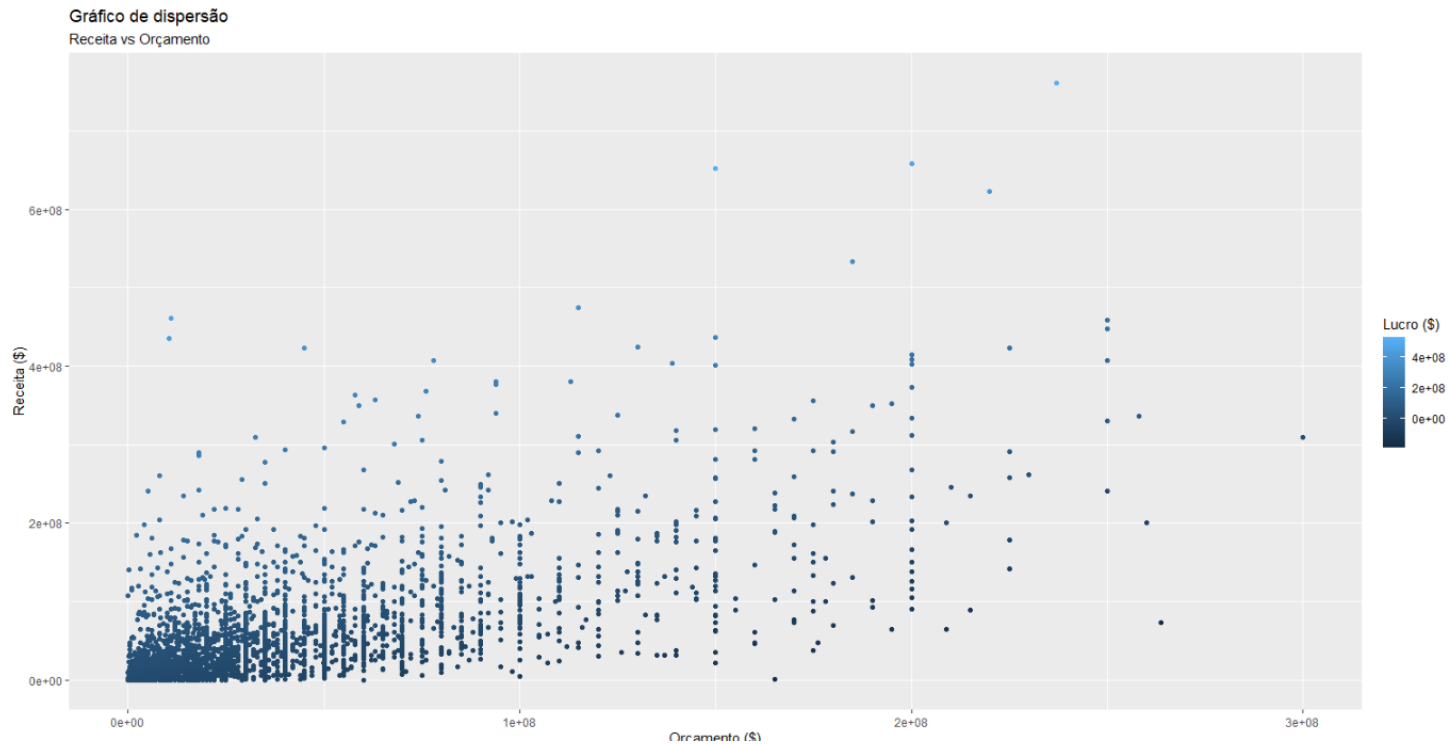
# Boxplots



# Títulos e labels

```
imdb %>%  
  mutate(lucro = receita - orcamento) %>%  
  ggplot() +  
  geom_point(mapping = aes(x = orcamento, y =  
    receita, color = lucro)) +  
  labs(  
    x = "Orçamento ($)",  
    y = "Receita ($)",  
    color = "Lucro ($)",  
    title = "Gráfico de dispersão",  
    subtitle = "Receita vs Orçamento"  
  )
```

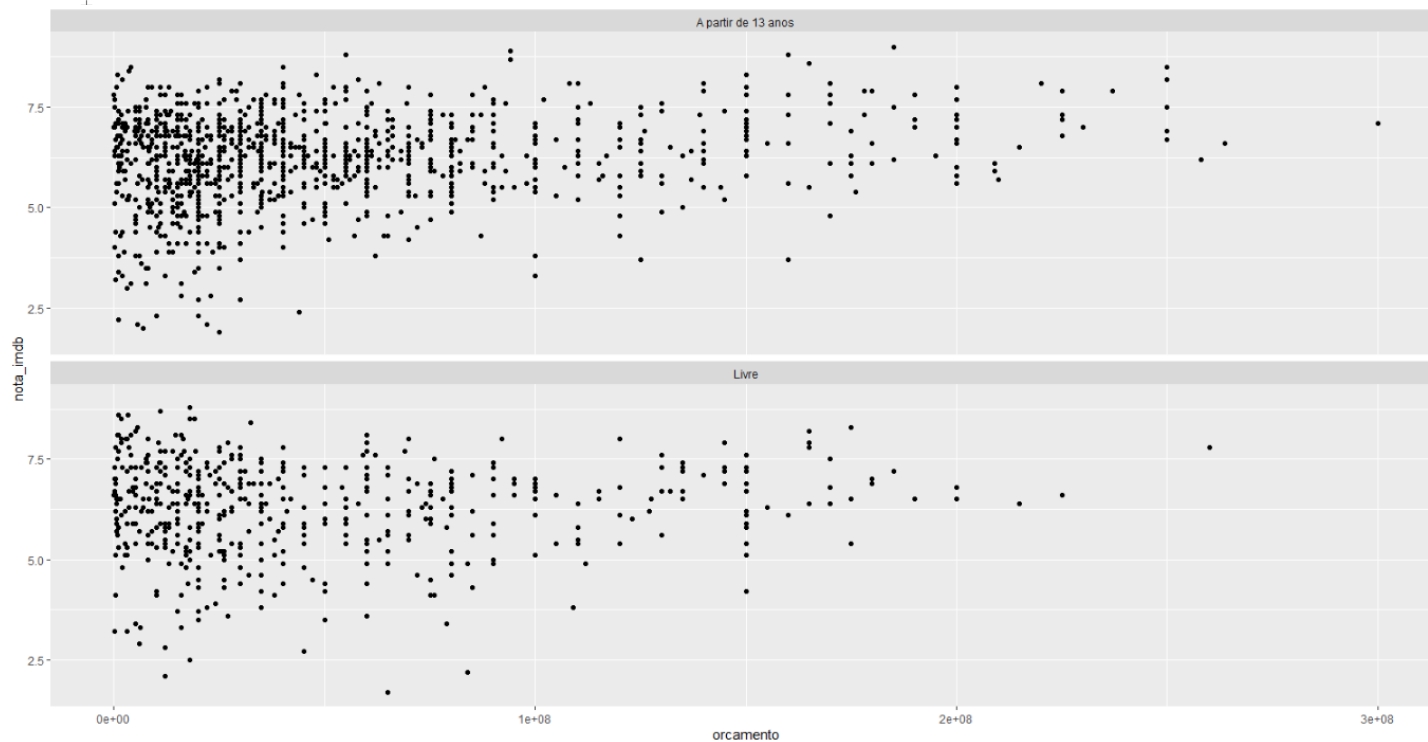
# Títulos e labels



# Facet

```
imdb %>%  
  filter(classificacao %in% c("Livre", "A partir de 13  
anos")) %>%  
  ggplot() +  
  geom_point(aes(x = orcamento, y = nota_imdb)) +  
  facet_wrap(~classificacao, nrow = 2)
```

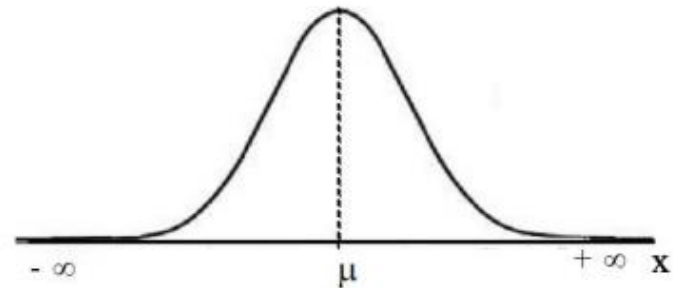
# Facet



# Exercícios

1. Rode cada um dos exemplos
2. O que acontece quando rodamos o código `ggplot(data = mtcars)`?
3. Utilizando a base `mtcars`, faça um gráfico de dispersão de `mpg` por `qsec`.
4. Usando a base `imdb`, crie um gráfico de dispersão da nota do `imdb` pelo orçamento.
5. Descubra quais são os 5 atores que mais aparecem na coluna `ator_1` e faça um boxplot do lucro dos filmes desses atores.

# Distribuição de probabilidade Normal



## CARACTERÍSTICAS:

- A) A variável pode assumir qualquer valor no conjunto real.
- B) O gráfico da distribuição é uma curva em forma de sino, simétrica em torno da média  $\mu$ , que é igual à mediana e à moda.
- C) A área sob a curva é igual a 1, e corresponde à probabilidade de a variável assumir valores entre  $[-\infty \dots +\infty]$ .
- D)  $\mu$ ;  $\sigma$  (Mi e Sigma) representam os parâmetros de posição e dispersão da distribuição.
- E) Os pontos de inflexão da curva ocorrem nos valores definidos por  $(\mu - \sigma$  e  $\mu + \sigma)$ .
- F) A expressão da função densidade de probabilidade é:

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2[(X-\mu)/\sigma]^2}$$



## ● Família normal

- `rnorm`: Para obter número aleatórios seguindo uma distribuição normal
- `dnorm` - Avalia a probabilidade da normal de um valor (dada a média  $\mu$  e o desvio padrão  $\sigma$ )
- `pnorm` - Avalia a probabilidade ACUMULADA da normal de um valor (dada  $\mu$  e  $\sigma$ )
  - Esta função deve formar uma curva sigmóide!

# Exemplo

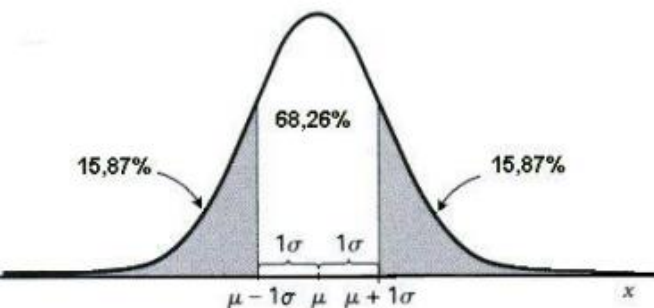
```
rnorm(100, 0, 10)
```

```
dnorm(5, 0, 10)
```

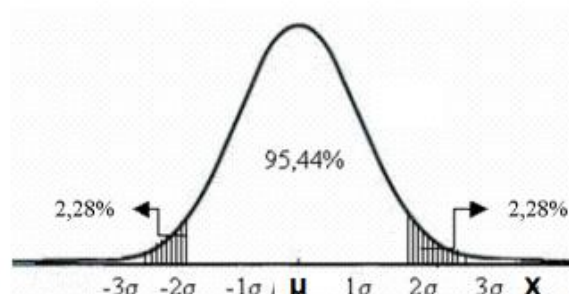
```
pnorm(5, 0, 10)
```

# Distribuição Normal

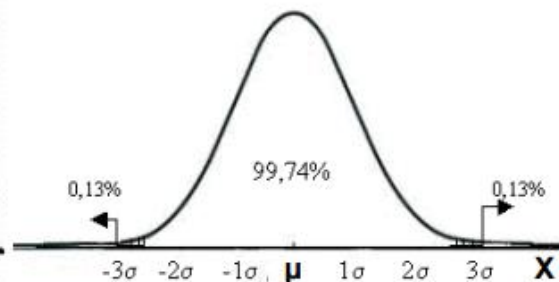
$$Z \sim N(0,1)$$



$$P[(\mu - \sigma) < X < (\mu + \sigma)] = 68.25\%$$



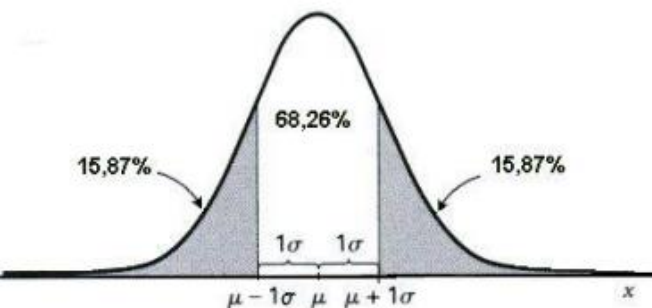
$$P[(\mu - 2\sigma) < X < (\mu + 2\sigma)] = 95.44\%$$



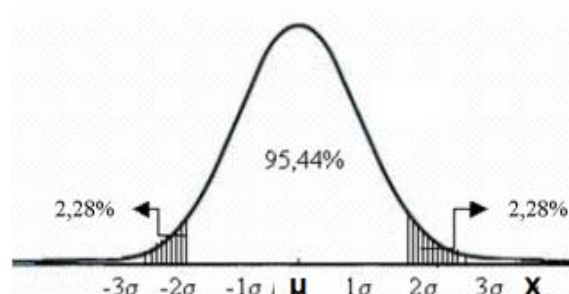
$$P[(\mu - 3\sigma) < X < (\mu + 3\sigma)] = 99.74\%$$

# Distribuição Normal

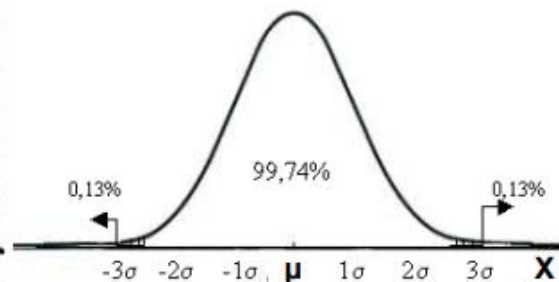
$$Z \sim N(0,1)$$



$$P[(\mu - \sigma) < X < (\mu + \sigma)] = 68.25\%$$



$$P[(\mu - 2\sigma) < X < (\mu + 2\sigma)] = 95.44\%$$



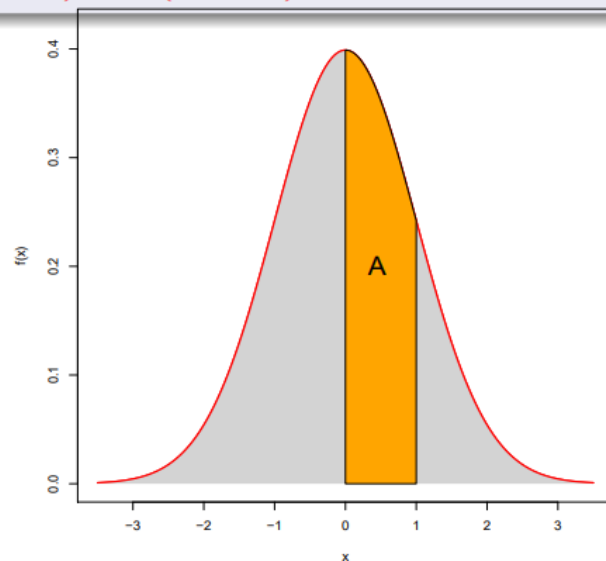
$$P[(\mu - 3\sigma) < X < (\mu + 3\sigma)] = 99.74\%$$

# Distribuição Normal

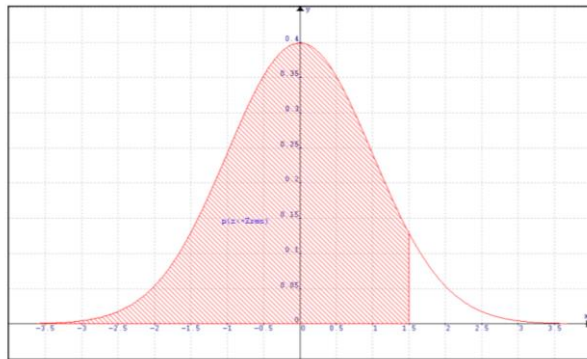
## Cálculo de probabilidades

Por exemplo, a probabilidade  $A = P(0 \leq X \leq 1)$  pode ser calculada pela diferença

$$P(X \leq 1) - P(X \leq 0) = 0,841 - 0,5 = 0,341.$$



## FIAP MBA+



$$P(Z < 1.38) = ?$$

[illegible]

# Exemplo 1

Foi realizado um estudo sobre o salário dos economistas. Sabe-se que o salário desses profissionais é uma variável aleatória que possui distribuição normal com média de R\$ 5.000,00 e um desvio-padrão de R\$ 800,00. Levando em consideração essa situação, qual é a probabilidade de o salário dos economistas ser maior do que R\$ 7.500,00?

# Exemplo 2

A concentração de um poluente em água liberada por uma fábrica tem distribuição  $N(8, 1.5)$ . Qual a chance, de que num dado dia, a concentração do poluente exceda o limite regulatório de 10 ppm?



# Exercício

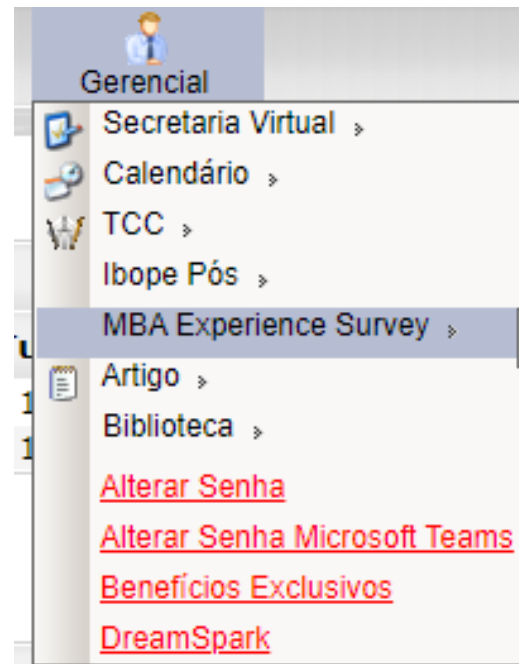
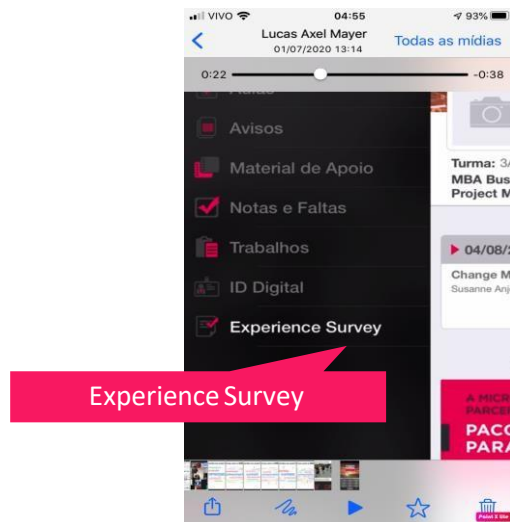
**01.** Uma empresa produz televisores de dois tipos, tipo *A* (comum) e tipo *B* (luxo), e garante a restituição da quantia paga se qualquer televisor apresentar defeito grave no prazo de seis meses. O tempo para ocorrência de algum defeito grave nos televisores tem distribuição normal sendo que, no tipo *A*, com média de 10 meses e desvio padrão de 2 meses e no tipo *B*, com média de 11 meses e desvio padrão de 3 meses. Os televisores de tipo *A* e *B* são produzidos com lucro de 1200 u.m. e 2100 u.m. respectivamente e, caso haja restituição, com prejuízo de 2500 u.m. e 7000 u.m., respectivamente.

(a) Calcule as probabilidades de haver restituição nos televisores do tipo *A* e do tipo *B*.

# O que você achou da aula de hoje?

Pelo aplicativo da FIAP

(Entrar no FIAPP, e no menu clicar em Experience Survey)



# OBRIGADO



**in** /lafphd

profleandro.ferreira@fiap.com.br

**FIAP** MBA<sup>+</sup>

Copyright © 2019 | Professor (a) Nome do Professor  
Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente  
proibido sem consentimento formal, por escrito, do professor/autor.

FIAP