

Programação em R para Geoprocessamento



Pedro José Farias Fernandes
pj_fernandes@id.uff.br
Laboratório de Geografia Física

Data Frames

The diagram illustrates a data frame structure. At the top, the word "Columns" is written in blue, with three blue arrows pointing down to the column headers: "Name", "Team", and "Number". To the left of the table, the word "Rows" is written in orange, with three orange arrows pointing to the row indices 0, 1, and 2. A purple bracket labeled "Data" is positioned at the bottom right, spanning the data cells of the "Team" and "Number" columns for rows 2 through 6. The table itself has a light green background and a thin black border. The columns are labeled "Name", "Team", "Number", "Position", and "Age". The rows are indexed from 0 to 6.

	<i>Name</i>	<i>Team</i>	<i>Number</i>	<i>Position</i>	<i>Age</i>
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

æG

<https://www.geeksforgeeks.org/r-data-frames/>

Data Frames

alunos	notas
Jemisson	7
Pedro	8
Raphael	9
Junior	10
Jairson	10
Marcelo	0

Data Frames

```
notas<-c(7,8,9,10,10,0)
alunos<-c("Jemisson","Pedro","Raphael","Junior","Jairson","Marcelo")
```

```
df<-data.frame(alunos=alunos, notas=notas)
df
```

```
df[, "notas"]
df[, "alunos"]
```

```
df[,1]
df[,2]
```

```
df[1,]
df[c(2,3),]
```

```
df$notas
df$alunos
```

Data Frames

```
##Pegando alunos nota 10
```

```
alunos_nota_10 <- df[df$notas==10,]
```

```
#Pior aluno
```

```
pior_aluno <- df[df$notas==0,]
```

Data Frames

```
#Leitura de dados
```

```
tabela<-read.csv("VGs__melhores.csv", h=T, sep=",")
```

```
names(tabela)
```

```
tabela$class_name
```

```
head(tabela)
```

Estatística descritiva

Média: soma dos valores /total de observações

Notas: 7 8 9 10 10 0 5

$$(7+8+9+10+10+0)/7 = 7$$

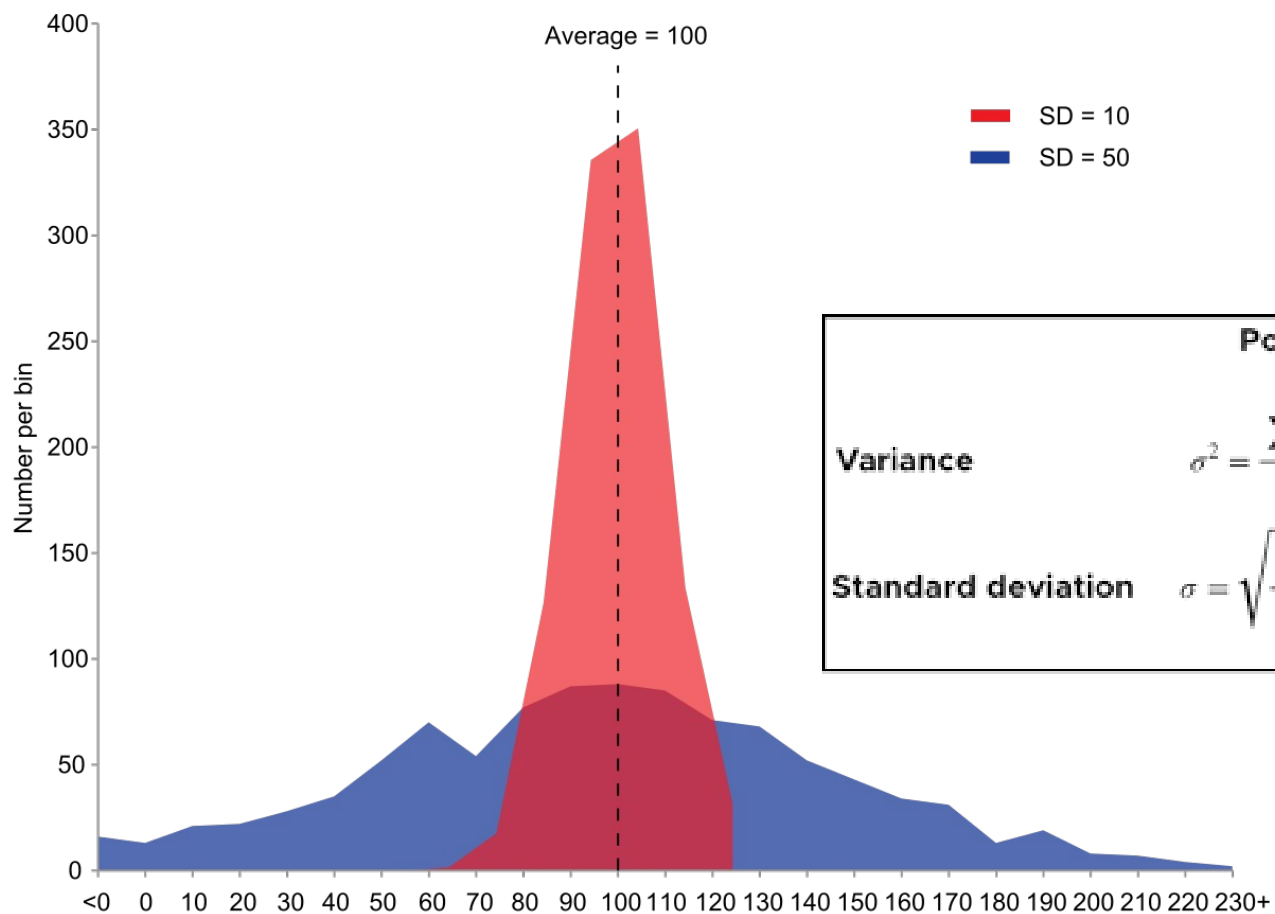
Mediana: valor central dos dados ordenados:

$$0 \ 5 \ 7 \ 8 \ 9 \ 10 \ 10 = 8$$

Moda: valor mais frequente = 10

Variância: medida de dispersão = 12.66667

Desvio padrão: $\sqrt{\text{variância}}$ = 3.559026

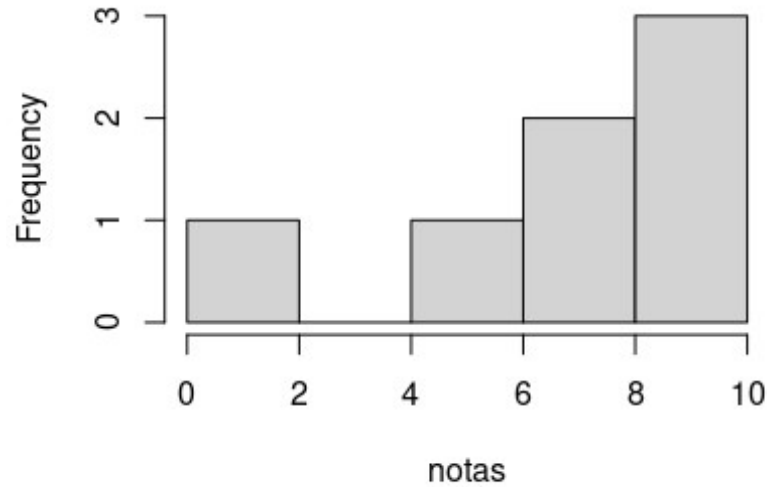


	Population	Sample
Variance	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

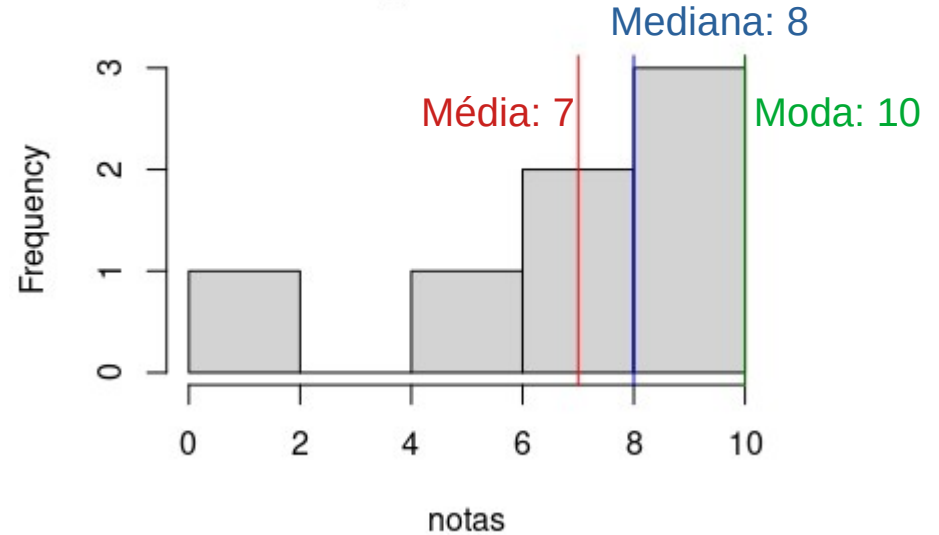
<https://en.wikipedia.org/wiki/Variance>

Histograma das notas

Histogram of notas



Histogram of notas



#Estatística descritiva e medidas de das notas

summary(notas)

mean(notas)

median(notas)

var(notas)

sd(notas)

hist(notas)

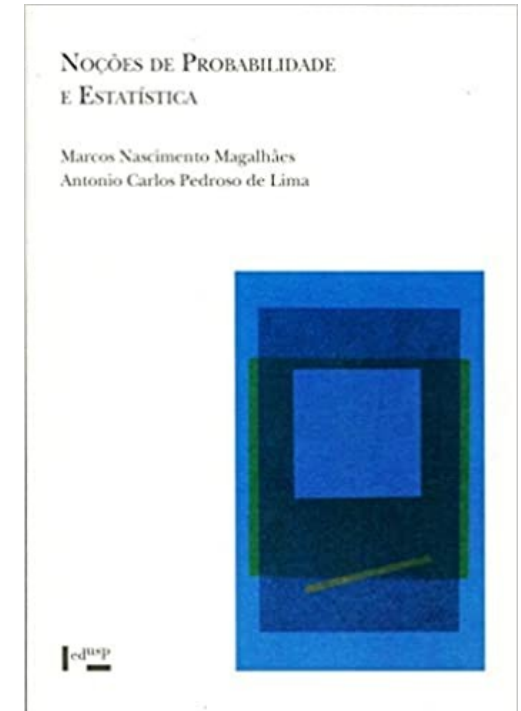
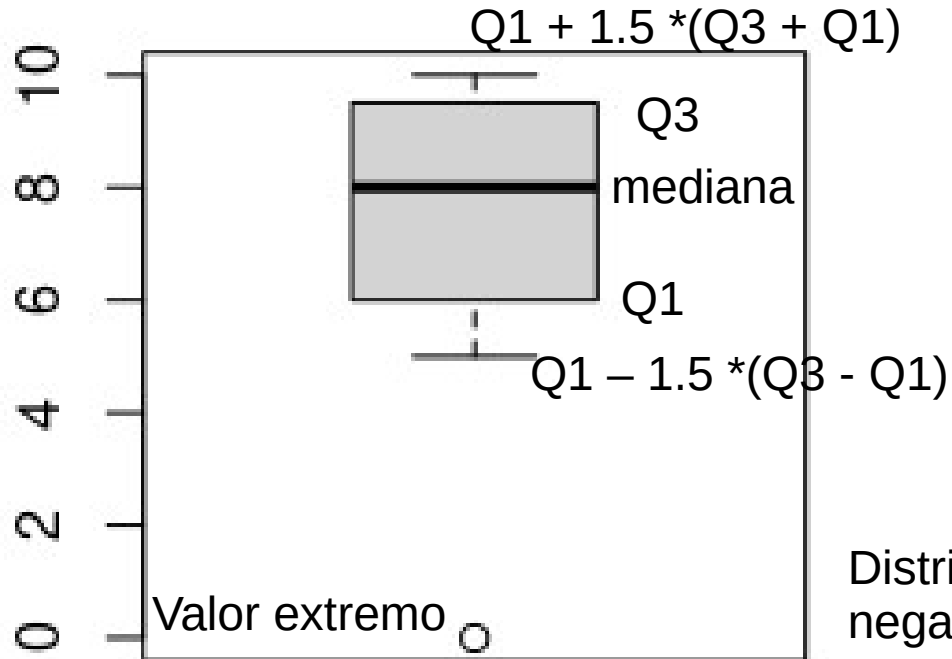
abline(v=c(mean(notas, 0)), col="red") #adicionando média

abline(v=c(median(notas, 0)), col="blue") #adicionando mediana

abline(v=10, col="darkgreen") #adicionando moda

Boxplot

“... permite visualizar diversos aspectos da distribuição dos dados...” (MAGALHÃES; LIMA, 2015)



Boxplot e assimetria

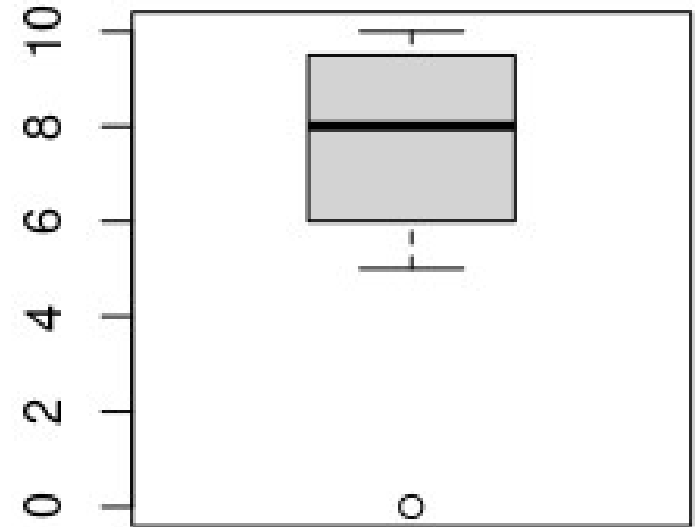
#Boxplot e Assimetria

```
boxplot(notas)
```

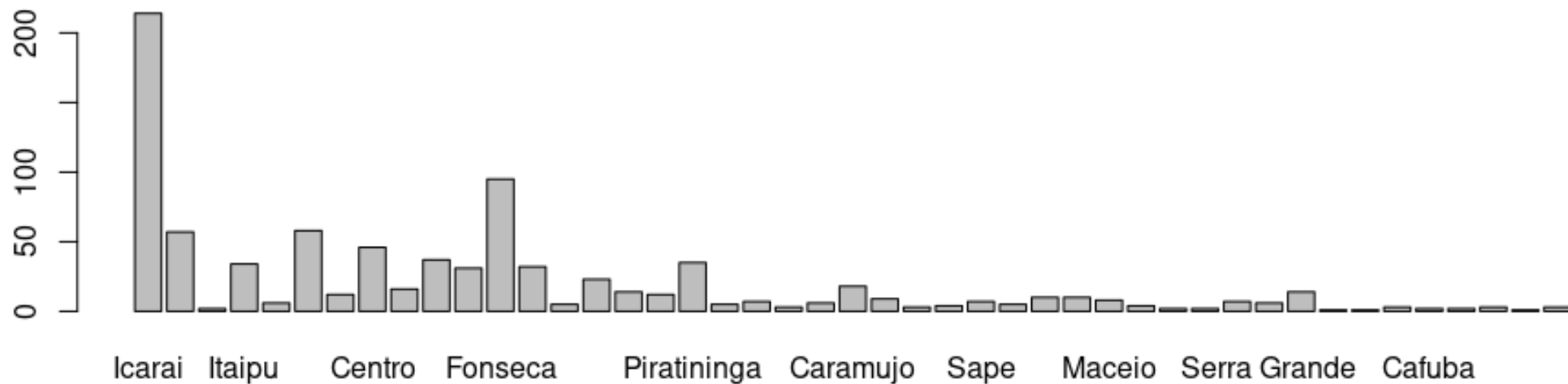
```
install.packages("moments")
```

```
library(moments)
```

```
skewness(notas)
```



Barplot – Casos COVID-19 em Niterói RJ (13/05/2020)



Barplot – Casos COVID-19 em Niterói RJ (13/05/2020)

#barplot

```
casos_covid_niteroi <-  
read.csv("https://raw.githubusercontent.com/pjfernandes/pjfernandes.github.io/  
master/casos.csv", sep=";", h=T)
```

```
head(casos_covid_niteroi)
```

```
barplot(casos_covid_niteroi$Casos, names.arg = casos_covid_niteroi$NOME)
```

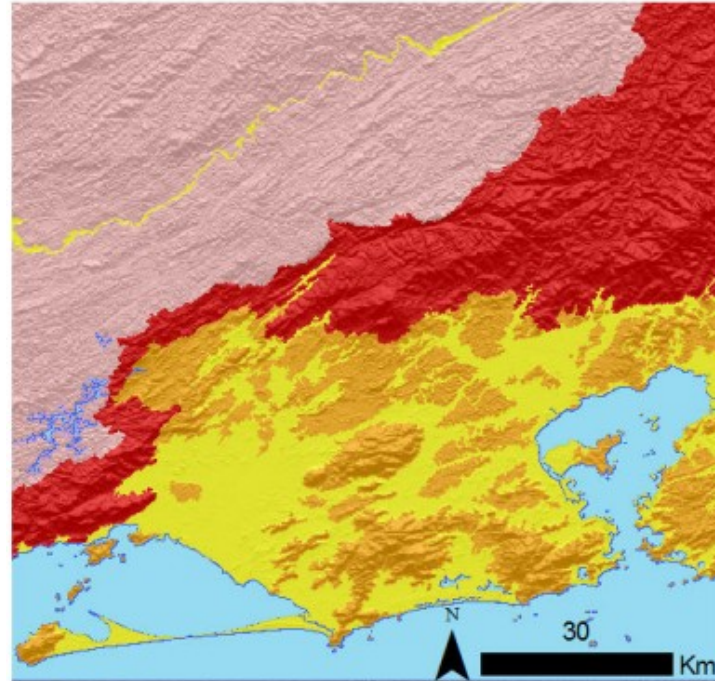
Biblioteca plotly

```
#plotly
install.packages("plotly")
library(plotly)

fig <- plot_ly(
  x = casos_covid_niteroi$NOME,
  y = casos_covid_niteroi$casos,
  name = "Casos COVID (13/05/2020)",
  type = "bar"
)

fig
```

Boxplot das classes de relevo



- Colinas e Maciços Costeiros
- Escarpas e Reversos da Serra do Mar
- Planícies Litorâneas
- Vale do Paraíba do Sul

Boxplot das classes de relevo

```
tabela<-read.csv("VGs__melhores.csv", h=T, sep=",")
```

```
ersm_declividade<-tabela[tabela$class_name=="Escarpas e Reversos da Serra do Mar",  
$Mean.Declividade
```

```
cmc_declividade<-tabela[tabela$class_name=="Colinas e Maciços Costeiros",  
$Mean.Declividade
```

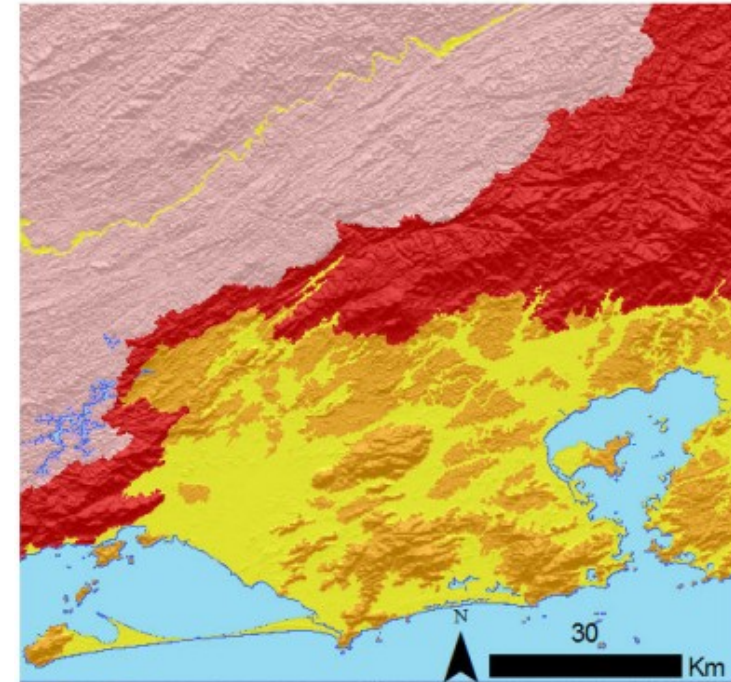
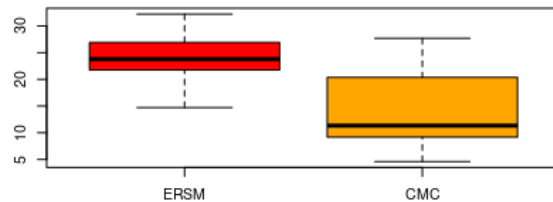
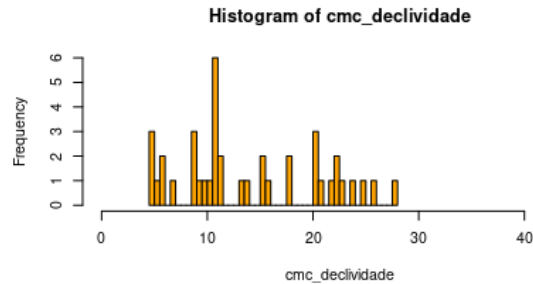
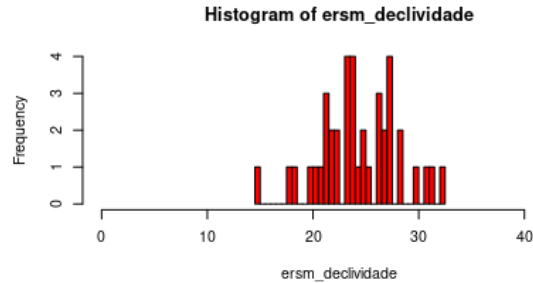
```
par(mfrow=c(3,1))
```

```
hist(ersm_declividade,breaks=50,xlim=c(0,45), col="red")
```

```
hist(cmc_declividade,breaks=50,xlim=c(0,45), col="orange")
```

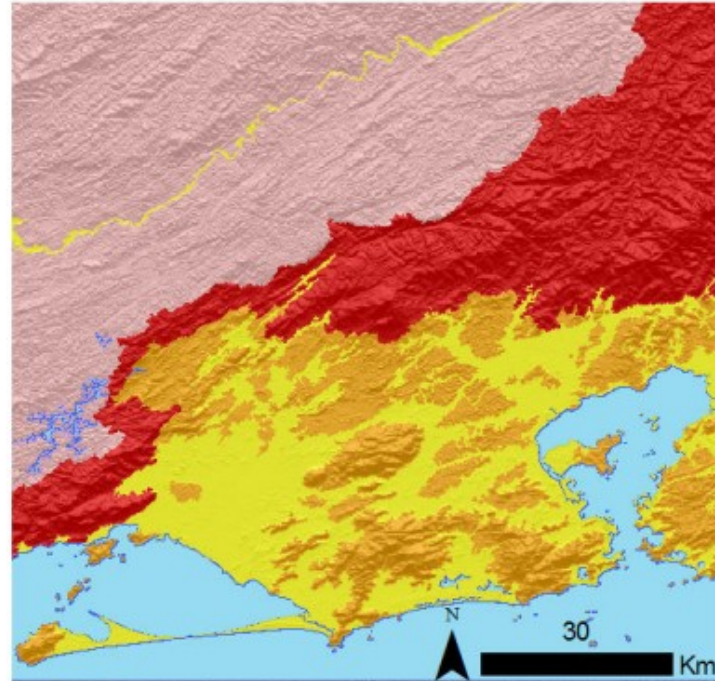
```
boxplot(ersm_declividade, cmc, col=c("red", "orange"), names=c("ERSM", "CMC"))
```

Boxplot das classes de relevo



- Colinas e Maciços Costeiros
- Escarpas e Reversos da Serra do Mar
- Planícies Litorâneas
- Vale do Paraíba do Sul

Boxplot das classes de relevo



- Colinas e Maciços Costeiros
- Escarpas e Reversos da Serra do Mar
- Planícies Litorâneas
- Vale do Paraíba do Sul

Teste t de Student

Para verificar diferença significativa entre a média de dois grupos

No caso, CMC e ERSM

- 1) Verificar se as variâncias dos grupos são significativamente iguais ou diferentes
- 2) Executar o teste t

Teste F

```
var.test(cmc_declividade, ersm_declividade)
```

- F test to compare two variances
-
- data: cmc_declividade and ersm_declividade
- $F = 3.1918$, num df = 39, denom df = 39, p-value = 0.0004527
- alternative hypothesis: true ratio of variances is not equal to 1
- 95 percent confidence interval:
- 1.688116 6.034708
- sample estimates:
- ratio of variances
- 3.191753

Teste t heterocedástico

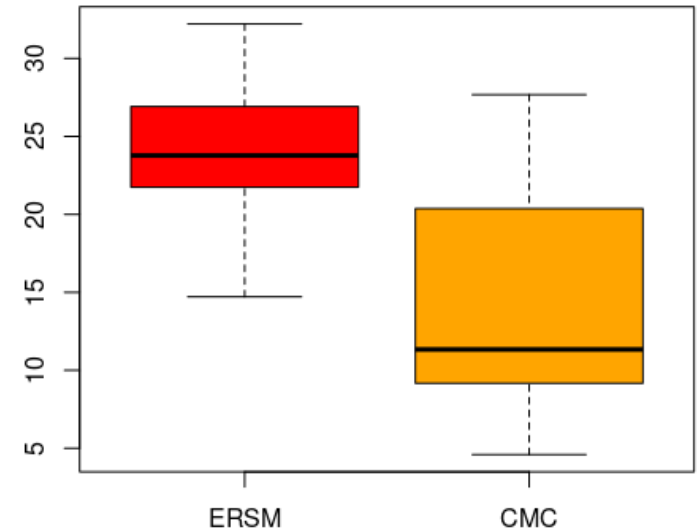
```
t.test(cmc_declividade, ersm_declividade, var.equal = F)
```

- Welch Two Sample t-test
-
- data: cmc_declividade and ersm_declividade
- $t = -8.4485$, $df = 61.254$, $p\text{-value} = 7.239e-12$
- alternative hypothesis: true difference in means is not equal to 0
- 95 percent confidence interval:
 - -12.630132 -7.796018
- sample estimates:
 - mean of x mean of y
 - 14.09792 24.31100

```
boxplot(ersm_declividade, cmc, col=c("red", "orange"),  
names=c("ERSM", "CMC"))
```

```
mean(ersm_declividade)
```

```
mean(cmc_declividade)
```



NOÇÕES DE PROBABILIDADE E ESTATÍSTICA

Marcos Nascimento Magalhães
Antônio Carlos Pedrosa de Lima

edusp



Correlação e Regressão linear simples

Coeficiente de correlação fornece a dependência linear entre variáveis

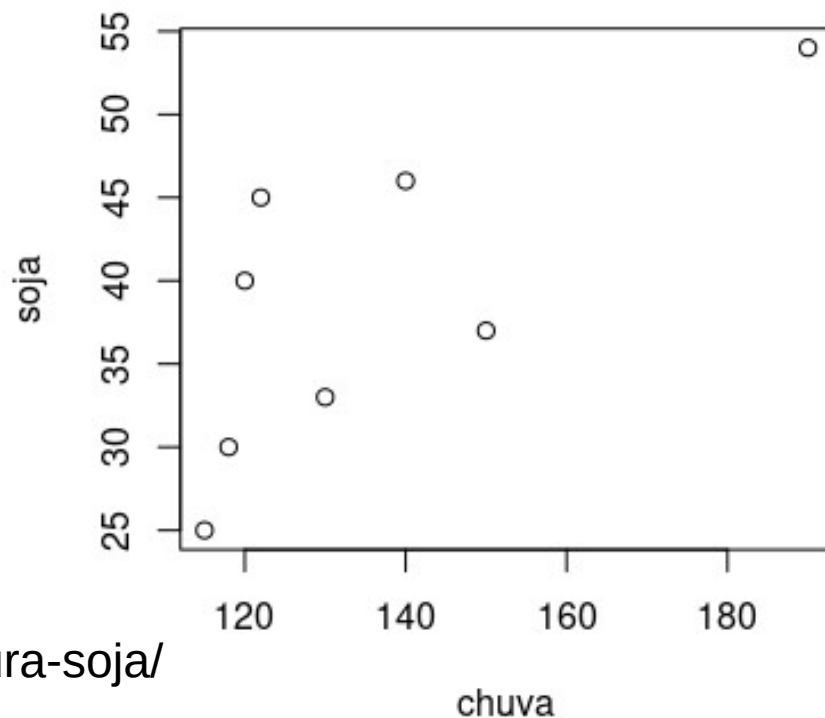
- **Exemplo:**

```
chuva<-c(120, 140, 122, 150, 115, 190, 130, 118) #em mm  
soja<-c(40, 46, 45, 37, 25, 54, 33, 30) #em toneladas  
cor(chuva, soja)
```

Quanto mais chuva, mais soja!

plot(chuva, soja)

“Mas a falta ou excesso de chuvas é o principal fator que afeta a produção brasileira de soja. O regime de precipitação exerce papel fundamental para a cultura, uma vez que esta espécie apresenta uma demanda hídrica importante, chegando a alcançar um pico de consumo de 8mm/dia/planta no período reprodutivo.”



<https://blog.syngentadigital.ag/influencia-chuvas-cultura-soja/>

Como estimar tonelada de soja a partir dos valores de chuva?

- Regressão linear simples!
- A regressão linear simples fornece uma equação para estimar soja a partir da chuva

$$\text{SOJA (t)} = 0.2743 * \text{CHUVA (mm)} + 1.5527$$

Uma equação $y = A * x + B$

```
summary(lm(soja~chuva))
```

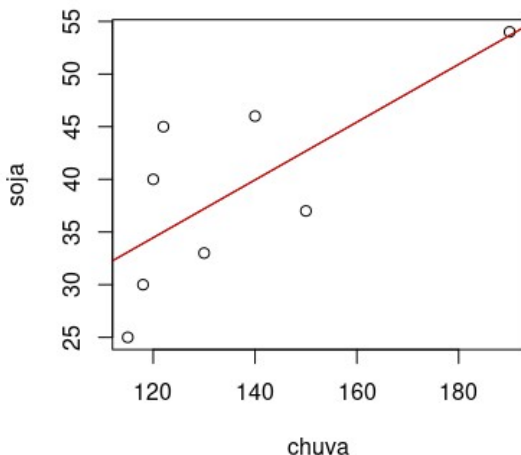
- Residuals:
- Min 1Q Median 3Q Max
- -8.093 -4.579 -1.790 5.664 9.987
-
- Coefficients:
- Estimate Std. Error t value Pr(>|t|)
- (Intercept) 1.5527 14.6568 0.106 0.919
- chuva 0.2743 0.1065 2.575 0.042 *
- ---
- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-
- Residual standard error: 7.048 on 6 degrees of freedom
- Multiple R-squared: 0.525, Adjusted R-squared: 0.4459
- F-statistic: 6.633 on 1 and 6 DF, p-value: 0.04203

Regressão linear simples

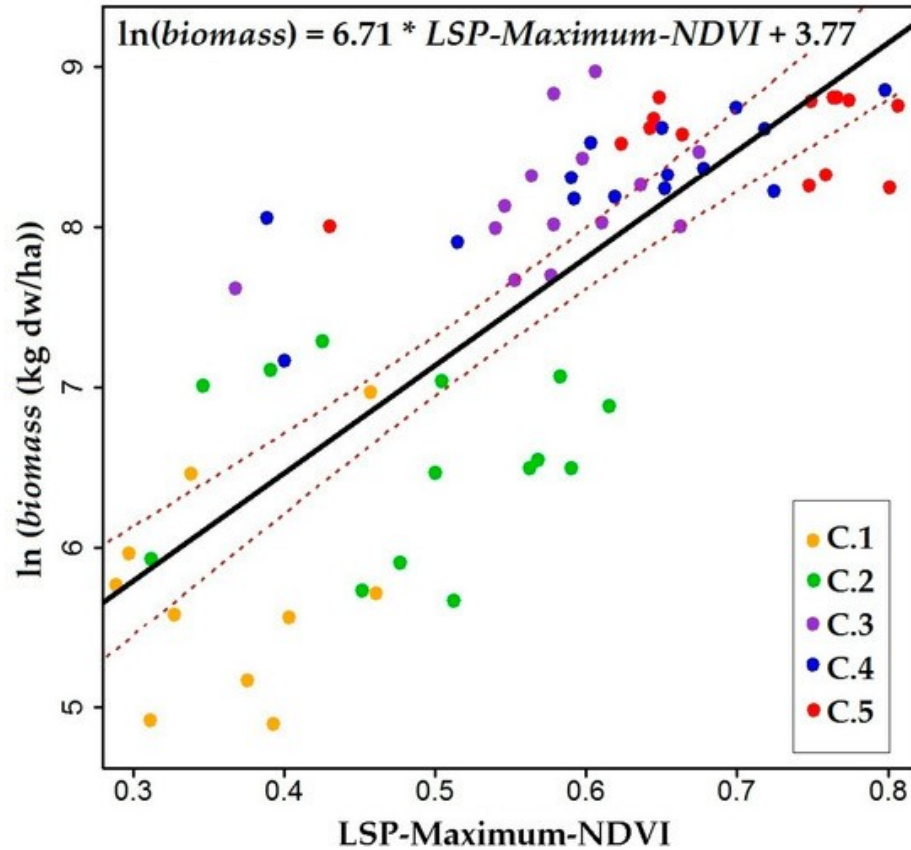
- A reta é ajustada pelo método dos mínimos quadrados.
- É a reta que minimiza o erro (minimiza a soma dos resíduos quadrados)
- Soma dos resíduos igual a zero

$$R^2=0,525$$

$P<0.05$ (o coeficiente angular é significativamente diferente de 0)



```
plot(chuva, soja)  
abline(lm(soja~chuva), col="red")
```



LUMBIERRES et al. (2017). Remote Sensing
<https://www.mdpi.com/2072-4292/9/4/392>

- Correlação (r) dá o grau de relacionamento entre duas variáveis (mede dependência linear).
- Coeficiente de determinação (r^2) fornece o quanto a variação de uma variável é explicada pela variação da outra.

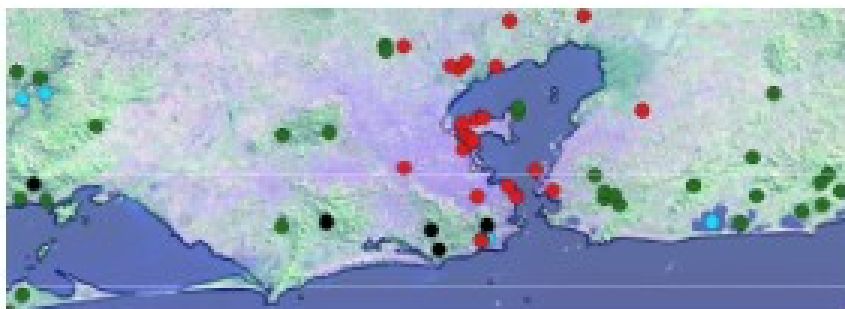
- ##correlação e regressão linear simples
chuva<-c(120, 140, 122, 150, 115, 190, 130, 118)
soja<-c(40, 46, 45, 37, 25, 54, 33, 30)

cor(chuva, soja)
summary(lm(soja~chuva))

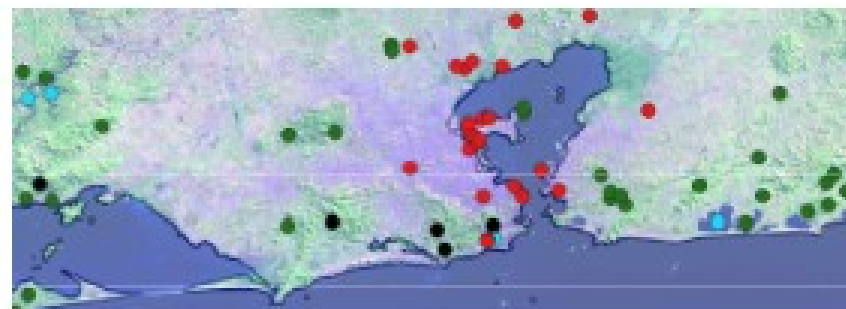
Normalização radiométrica (LANDSAT-5)

A normalização radiométrica ajusta as propriedades radiométricas de uma imagem para corresponder com as de uma imagem de referência, por um modelo de regressão por mínimos quadrados, a partir de pontos pseudoinvariantes.

Coloca as duas imagens em uma escala comum sem o uso de parâmetros extras (CHEN et al., 2005)

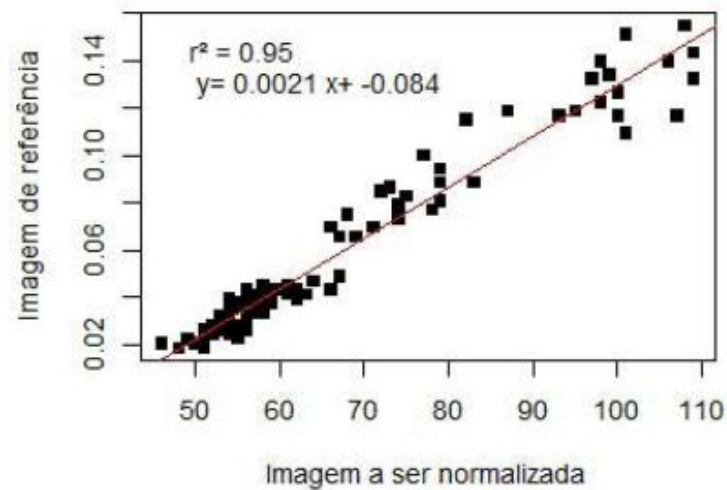


1984



2011

Banda 1 (antes)



Gerando modelo de normalização radiométrica para a banda 1 (TM-5)

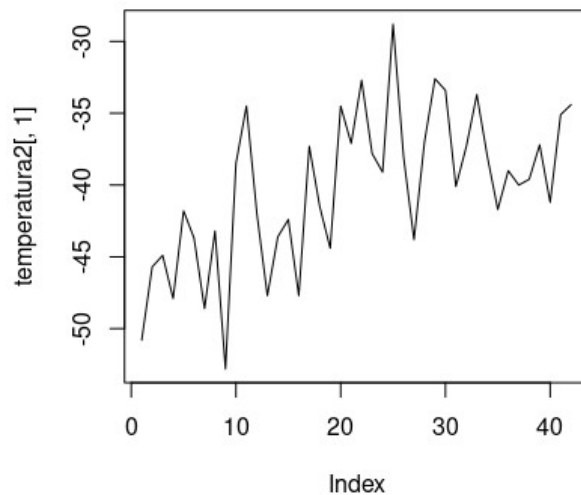
- Objetivo: Colocar a imagem de 1984 (em níveis de cinza) na mesma escala de reflectância da imagem de 2011 (referência)
- Qual a variável dependente?
- Estimar valores de reflectância a partir do nível de cinza de 1984

Normalização radiométrica

- 1) Ler arquivo com dados extraídos
- 2) Criar modelo
- 3) Plotar gráfico
- 4) Interpretar

Séries temporais

- De acordo com Morettin e Toloi (2006), uma série temporal é um conjunto qualquer de observações com ordenação temporal



Séries temporais

```
#####Séries temporais
```

```
install.packages("Kendall")
```

```
library(Kendall)
```

```
temperatura<-read.csv("https://raw.githubusercontent.com/pjfernandes/tendencias_dados_antartica/master/dados.txt", sep="\t", h=T)
```

```
head(temperatura)
```

```
plot(temperatura[,1], t="l")
```

```
MannKendall(temperatura[,1])
```

```
#####
```

```
temperatura2<-read.csv("https://raw.githubusercontent.com/pjfernandes/dados_clima_media/master/lon_75.txt", sep="\t", h=T)
```

```
plot(temperatura2[,1], t="l")
```

```
MannKendall(temperatura2[,1])
```