

CURSO R - 2023



—
NAIARA ALCANTARA



TÓPICOS

1-INTRODUÇÃO AO R

2- ANÁLISE EXPLORATÓRIA E MANIPULAÇÃO DOS DADOS

3- SALVAMENTO E ABERTURA

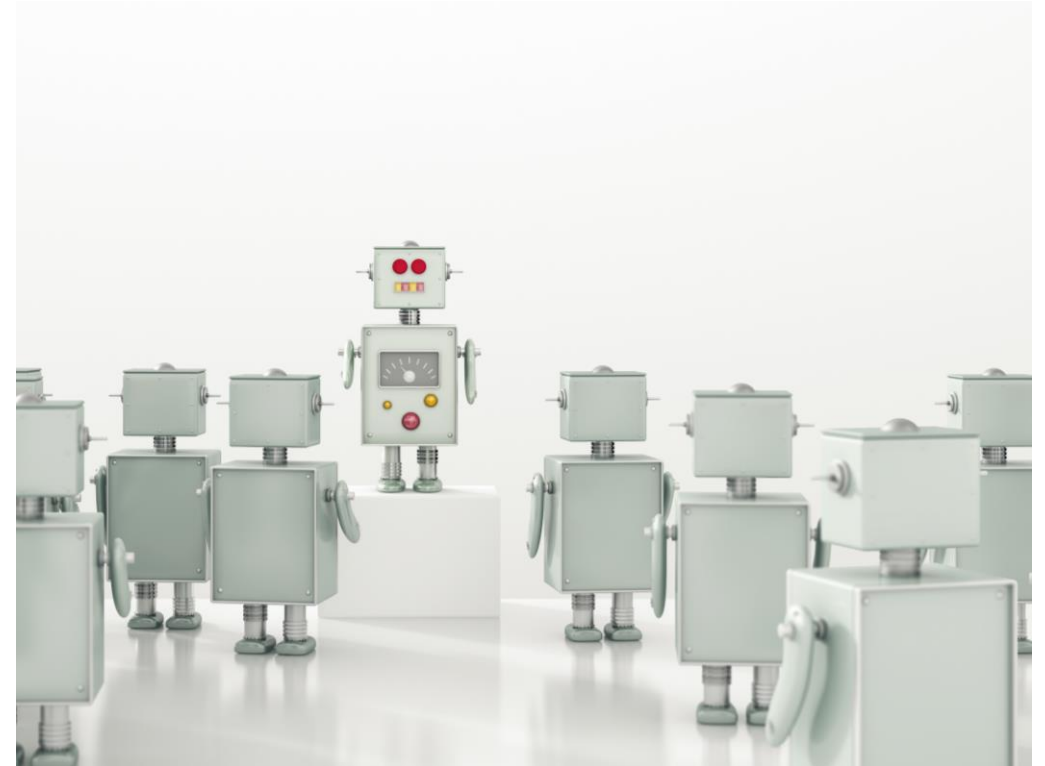
4- ANÁLISES DESCRITIVAS

5- APRESENTAÇÃO GRÁFICA

6- PROCESSAMENTO DE DADOS

7- ANÁLISES INFERENCIAIS

- **Introdução a análise descritiva**
- **Organização de uma base de dados real: seleção dos dados; recategorização; análise das medidas centrais (média, mediana e moda);**
- **Análises cruzadas**
- **Análise de frequências absolutas e relativas**
- **Introdução a análise gráfica**





ANÁLISE DESCRITIVA

ANÁLISE DESCRITIVA

Base de dados

- Para realizar a análise descritiva vamos utilizar a base de dados do LAPOP 2014 que baixamos anteriormente do site do LAPOP
- Quem não baixou, basta entrar nesse link: <http://datasets.americasbarometer.org/database/index.php?freeUser=true>
- Escrever “Brazil” e baixar a base em stata ou sav do ano de 2014

O que é uma análise descritiva.

Como o nome já diz, é um processo de descrição dos dados ou então é a fase inicial de um “processo de estudo dos dados coletados. Utilizamos métodos de Estatística Descritiva para organizar, resumir e descrever os aspectos importantes de um conjunto de características observadas ou comparar tais características entre dois ou mais conjuntos.”*

Após baixar a base de dados e o questionário:

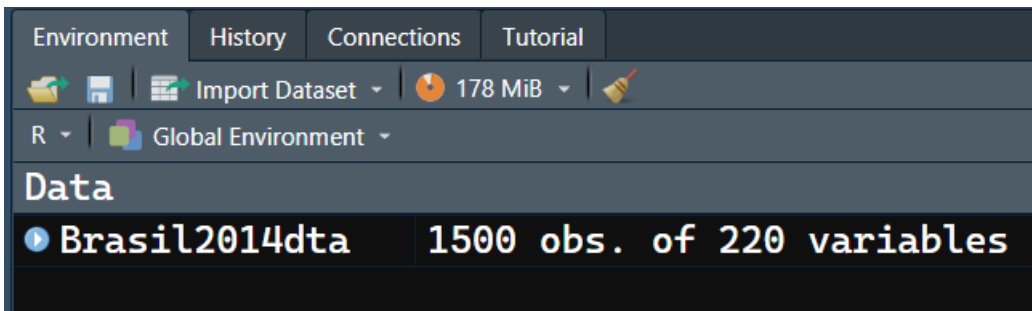
- Vamos olhar a base de dados no ambiente de programação do R, abrindo através do files ou usando a linha de comando que aprendemos na seção anterior.

```
#Análise descritiva####  
library(haven)  
  
Brasil2014dta ←  
  read_dta("636339374Brazil LAPOP AmericasBarometer 2014 v3.0_W.dta")
```

Para mais informações acessar Reis, E.A., Reis I.A. (2002) Análise Descritiva de Dados. Relatório Técnico do Departamento de Estatística da UFMG. Disponível em: www.est.ufmg.br ou <http://www.est.ufmg.br/portal/arquivos/rts/rte0202.pdf>

ANÁLISE DESCRITIVA

- Após carregar a base de dados no ambiente do R suas informações devem aparecer em que lugar?
- Isso mesmo, se você respondeu no Environment você acertou !



Temos uma base de dados com 1500 observações e 220 variáveis.

O número de observações está ótimo. Mas será que precisaremos analisar todas essas 220 variáveis em nossa pesquisa acadêmica? Teremos que fazer gráficos e testes usando todo esse material? Acho que não, então que tal selecionar o que será útil e trabalhar com uma base menor?

- Pode analisar dentro do R a base de dados, através da função View (ou clicando na base), mas usando o questionário ou um dicionário de códigos fica mais fácil, assim saberemos o que cada código significa e quais são suas categorias.

```
View(Brasil2014dta)
```

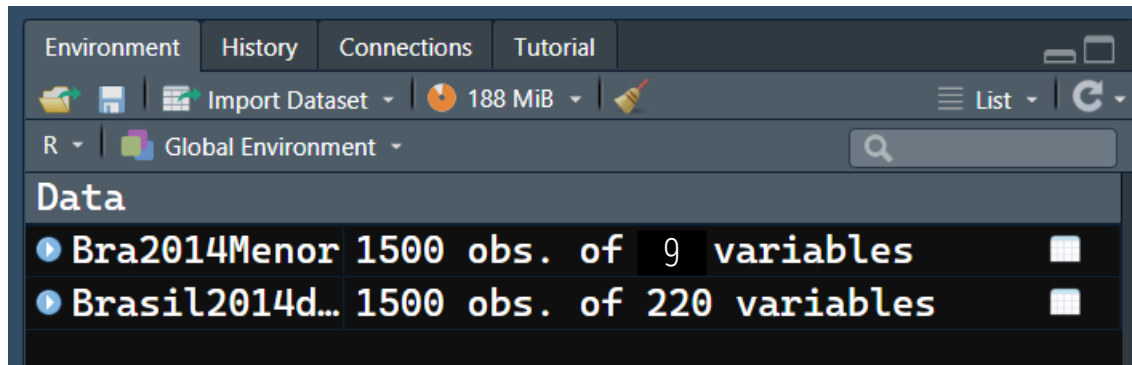
- Selecionei do questionário todas as questões que acredito serem suficientes para as análises nesse momento. Agora vamos fazer o mesmo procedimento no R.

ANÁLISE DESCRITIVA

```
#Seleção das variáveis####  
colnames(Brasil2014dta)  
Brasil2014dta$q11n  
  
# ls3 = Satisfação com a vida (SatVida) - inverter a escala  
# q2 = Idade  
# q1 = Sexo  
# q10new = Renda familiar (RendFam)  
# q10d = Percepção sobre a renda (PercRend) - Inverter a escala  
# q11n = Estado civil (EstCiv)  
# q12c = Quantidade de moradores no mesmo domicílio (QuantMor)  
# ocup4a = Ocupação (Ocupacao)  
# q12= Quantos filhos tem (Filhos)  
  
#Após definir todas as variáveis que iremos querer basta  
#utilizar a função subset que aprendemos na seção anterior  
  
Bra2014Menor ← subset(Brasil2014dta, select =  
                        c("ls3", "q2", "q1",  
                          "q10new", "q10d", "q10d",  
                          "q11n", "q12c", "ocup4a",  
                          "q12"))
```

ANÁLISE DESCRITIVA

- Agora veja como está nossa base de dados:



- Mas porque a base de dados com 220 variáveis ainda aparece?

Porque criamos uma outra base de dados, através de uma cópia de algumas variáveis existentes na base original.

Fique atento a esse procedimento.

Cuidado para NÃO substituir a base original e, perder dados que você poderá precisar posteriormente.



ANÁLISE DESCRITIVA

- Antes de iniciar as análises, temos que entender como estão organizadas as variáveis, fazer as recodificações e reorganizações, isso representa uma grande parte do trabalho de análise.
- Para essa etapa usaremos as funções **table** e **summary** que irão nos auxiliar mostrando o descritivo de cada variável
- Depois usaremos a função **recode** do pacote **memisc** para recategorizar as variáveis.

Agora a base está organizada, mas notem que ao longo de processo parte da análise descritiva já estava sendo feita, através, por exemplo:

- Distribuição da frequência, é comum querer entender e comparar a quantidade total de casos de cada categoria, por exemplo: i) total de homens em relação ao total de mulheres entrevistados i) total de pessoas casadas em relação a solteira, iii) total de pessoas que se sente satisfeita com salários em relação aos insatisfeitos.

Além das frequências absolutas e percentuais, as medidas centrais são também muito utilizadas para auxiliar nas descrições. As mais comuns são:

- Moda: representa o valor/categoria mais comum em uma distribuição
- Mediana: representa a medida central ou então a divisão das medidas centrais (quando são duas)
- Média: representa o valor médio dos casos

$$\bar{x} = \frac{\sum x}{N}$$

\bar{x} = soma de todas as observações/pelo N

$$M = \frac{N+1}{2}$$

N = número de observações

ANÁLISE DESCRITIVA

 Tanto a média, quanto a mediana nós já vimos que é possível obter através da função summary do rbase

```
Min. 1st Qu. Median Mean 3rd Qu. Max. NA's  
0.000 0.000 2.000 1.888 3.000 16.000 2
```

Em média as famílias possuem 2 filhos
A mediana (medida central) é 2

 Já a moda é possível obter através da função table para variáveis categóricas e também para as numéricas, o problema é que você tem olhar os dados para encontrar a moda. *É um problema se a quantidade de dados foi muito grande.

Número de filhos
Totais

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
453	290	319	214	83	50	35	23	11	9	1	4	2	1	1	1	1

Ocupação

Emprego_Rem	Emprego_N_Rem	Estudante	Aposentado	Nao_Empregado
812	200	125	180	183

 Vamos entender outros meios de obter essas medidas

```
#Funções específicas para média e mediana  
mean(Bra2014Menor$Filhos, na.rm = T)  
median(Bra2014Menor$Filhos, na.rm = T)
```

```
> mean(Bra2014Menor$Filhos, na.rm = T)  
[1] 1.88785  
> median(Bra2014Menor$Filhos, na.rm = T)  
[1] 2
```

ANÁLISE DESCRITIVA

```
506 #Para o cálculo da moda:http://www.dma.uem.br/kit/outros-arquivos/moda.pdf
507 #Explicações do help
508 moda ← function(x) {
509     modal ← unique(x)
510     modal[which.max(tabulate(match(x, modal)))]
511 }
512
513 #unique returns a vector or data frame like x
514 #but with duplicate elements/rows removed.
515
516 #which.max: Determines the location, i.e.,
517 #index of the (first) minimum or maximum of
518 #a numeric (or logical) vector.
519
520 #tabulate: counts the number of times each integer occurs in it.
521
522 #match returns a vector
523 #of the positions of its first argument in its second.
524
525 # Agora podemos rodar a função moda:
526 moda(Bra2014Menor$Filhos)
527 #[1] 0
```

ANÁLISE DESCRITIVA

Até esse momento aprendemos:

- 1- Organizar uma base de dados
- 2- Observar medidas de frequência absolutas
- 3- Observar medidas de tendência central

Isso não significa que esgotamos todo o assunto desses tópicos acima, pelo contrário, existem uma diversidade de funções que podem ser utilizadas para organizar e observar uma base de dados.

Muitas dessas questões vocês irão se deparar em situações práticas, por isso é importante entender as lógicas do que estamos fazendo, com esse conhecimento você será capaz de buscar soluções para outras questões.

Agora veremos mais algumas formas de analisar as variáveis, especialmente envolvendo mais de uma questão. Isso nos auxiliará no futuro, quando chegarmos as análise bivariadas

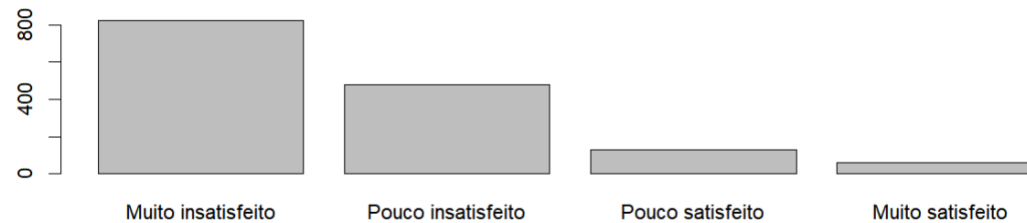


ANÁLISE DESCRITIVA

Frequência absoluta e percentual

```
#Frequência
#Pacote
library(descr)
freq(Bra2014Menor$SatVida)
```

```
Bra2014Menor$SatVida
Frequência Percentual
Muito insatisfeito      825      55.0000
Pouco insatisfeito      480      32.0000
Pouco satisfeito        131       8.7333
Muito satisfeito         62       4.1333
NA's                     2       0.1333
Total                   1500     100.0000
>
```



Com a função freq. além do percentual temos na saída dos plots um gráfico que é de ? _____

Só não teremos na saída um gráfico, se inserirmos uma função plot=F

!

Esse é o primeiro contato com os gráficos, ainda veremos alguns nesse tópico, porém não apresentaremos nenhuma elaboração gráfica.

O tópico seguinte será somente sobre GGLOT2

ANÁLISE DESCRITIVA

Análise cruzada - table

```
541 table(Bra2014Menor$SatVida, Bra2014Menor$Sexo)
542 #           Homem Mulher
543 # Muito insatisfeito  433   392
544 # Pouco insatisfeito  236   244
545 # Pouco satisfeito    51    80
546 # Muito satisfeito    28    34
547
548
549 #Vamos salvar o table que criamos dentro de um
550 #objeto e depois vamos chamar ele usando
551 #Uma outra função.
552 Obj ← table(Bra2014Menor$SatVida, Bra2014Menor$Sexo)
553
554 #Vamos utilizar a função: prop.table()
555 #Para apresentar os valores percentuais, contudo
556 #Teremos que multiplicar por 100, por isso salvamos
557 #tudo dentro de um novo objeto
558 Obj2 ← prop.table(Obj, margin = 1)
559 #Apresenta o percentual na linha
560 Obj2*100
561
562 Obj3 ← prop.table(Obj, margin = 2 )
563 #Apresenta o percentual na coluna
564 Obj3 * 100
```

	Homem	Mulher
Muito insatisfeito	52	48
Pouco insatisfeito	49	51
Pouco satisfeito	39	61
Muito satisfeito	45	55

	Homem	Mulher
Muito insatisfeito	58	52
Pouco insatisfeito	32	33
Pouco satisfeito	7	11
Muito satisfeito	4	5

ANÁLISE DESCRITIVA

```
570 ###  
571 plot(Bra2014Menor$Idade)  
572 plot(Bra2014Menor$Ocupacao)  
573 #  
574 hist(Bra2014Menor$Idade)
```

Gráfico de dispersão

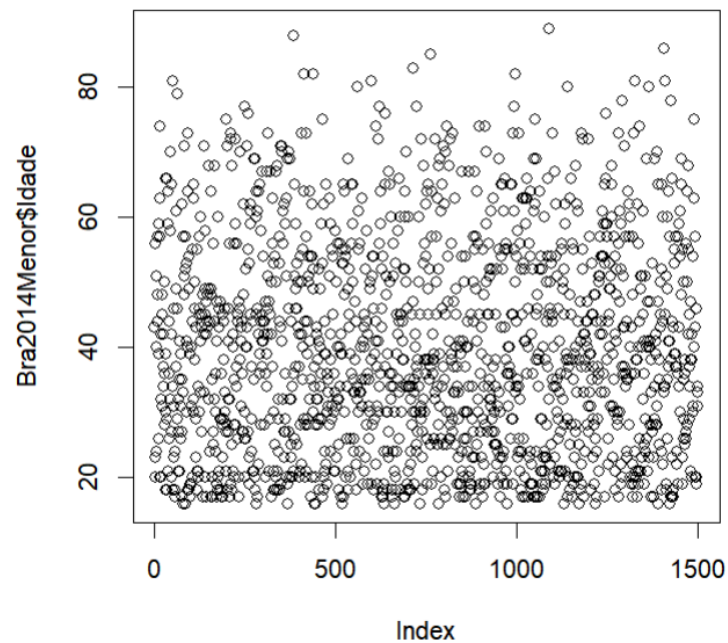
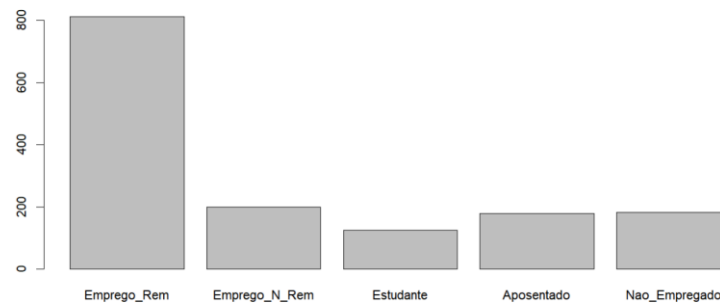
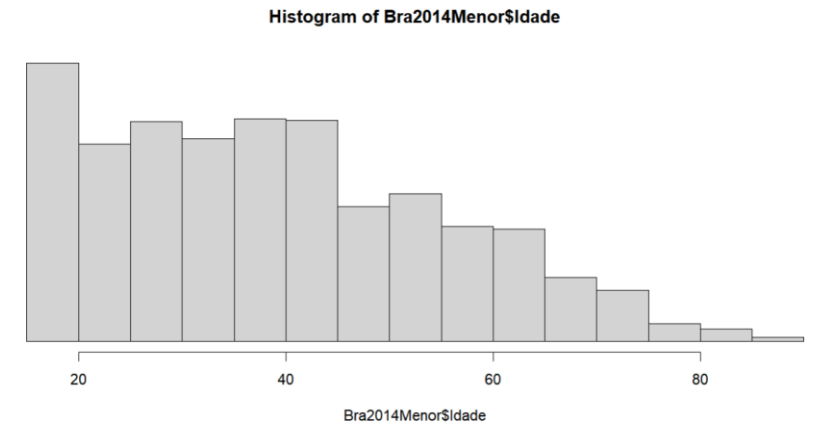


Gráfico de barras

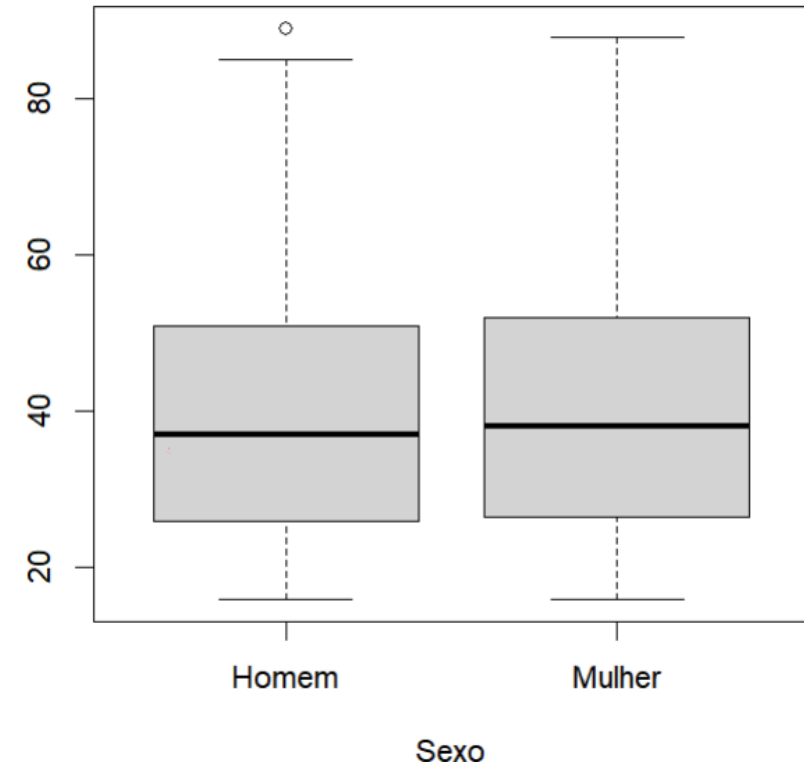
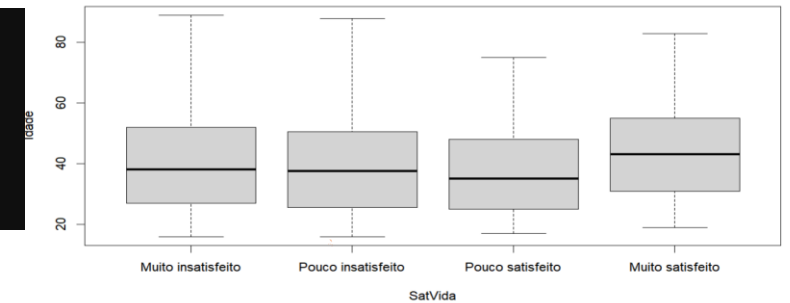
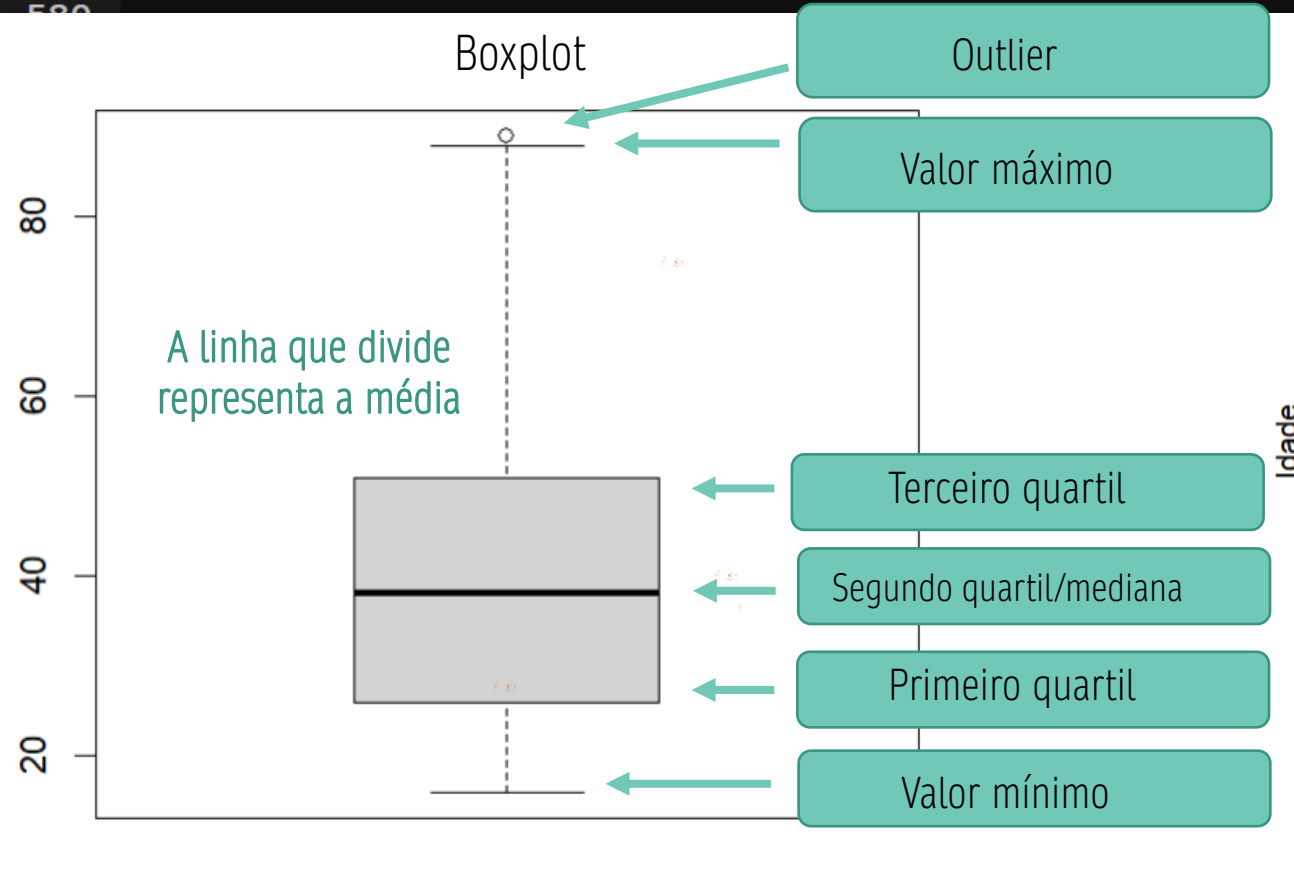


Histograma



ANÁLISE DESCRITIVA

```
575 boxplot(Bra2014Menor$Idade)
576 #
577 boxplot(Idade ~ Sexo, data= Bra2014Menor)
578 #
579 boxplot(Idade ~ SatVida, data= Bra2014Menor)
580
```



Nós vimos alguns tipos de gráficos, agora veremos esses mesmos gráficos de maneiras mais elaboradas, bem como, outros tipos



OBRIGADO

NAIARA ALCANTARA
NAYARA_SANDY@HOTMAIL.COM