

Maria Eduarda Rodrigues Garcia - 11796621  
Mirela Mei - 11208392

Roteiro - Atividade 2 de ACH2036

## 1. Identificação do dataset escolhido: Violência contra mulheres e meninas (Violence Against Women & Girls)

- Quem fez o levantamento dos dados (empresa, instituição de pesquisa, universidade, endereço)?
  - Demographic and Health Surveys (DHS) Program
  - Fonte de acesso: DataWorld (<https://data.world/>)
  - Instituição de pesquisa: Demographic and Health Surveys (DHS) Program, promovida pela empresa estadunidense United States Agency for International Development
  - Fonte:  
<https://www.kaggle.com/datasets/whenamancodes/violence-against-women-girls>
- Quando o levantamento foi feito?
  - Última atualização em 12 de Setembro de 2022
  - Indicadores desde 2000
- Quantas linhas e quantas colunas tem o dataset?
  - 12601 linhas e 8 colunas

## 2. Procedimentos de amostragem

Os dados foram retirados de uma pesquisa com homens e mulheres em países africanos, asiáticos e sul-americanos, explorando as atitudes e justificativas percebidas para cometer atos de violência contra as mulheres. Os dados também exploram diferentes grupos sociodemográficos aos quais os entrevistados pertencem, incluindo: Nível de escolaridade, Estado civil, Emprego e Faixa etária.

## 3. Variáveis

- Record ID - quantitativa discreta
- País - qualitativa nominal
- Gênero - qualitativa nominal
- Pergunta demográfica - qualitativa nominal
- Resposta demográfica - qualitativa nominal e ordinal
- Pergunta - qualitativa nominal
- Ano da pesquisa - quantitativa contínua

- Valor - quantitativa contínua
- Há missing data? Em quais variáveis? Indique como vocês trataram os missing data.
  - Há missing data para o campo Value
  - Tratamento: No caso em que as colunas inteiras estão vazias, essas serão excluídas. Em caso de missing data pontual, usaremos o listwise deletion, desconsiderando todo o registro em caso de variável sem valor.

#### 4. Observações, casos ou instâncias

- População observada entre países africanos, asiáticos e sul-americanos.

#### 5. Estatística descritiva

- tabelas de distribuição de frequências adequadas ao tipo de variável
- gráficos adequados ao tipo de variável
- medidas de tendência central adequadas ao tipo de variável
- medidas de dispersão adequadas ao tipo de variável
- box-plot (caso se aplique).

Importando o dataset:

```
# Importando DataSet e omitindo valores nulos
dados <- read.csv('database.csv', sep = ',', stringsAsFactors = F)
dados <- na.omit(dados)
```

Frequências absolutas e relativas das variáveis qualitativas:

```
# Variáveis qualitativas
# Frequências absolutas:
table(dados$Gender)
table(dados$Country)
table(dados$DemographicsQuestion)
table(dados$DemographicsResponse)
table(dados$Question)

# Frequências relativas:
prop.table(table(dados$Gender))
prop.table(table(dados$Country))
prop.table(table(dados$DemographicsQuestion))
prop.table(table(dados$DemographicsResponse))
prop.table(table(dados$Question))
```

Frequências das variáveis quantitativas:

```

# Variáveis quantitativas contínuas
# Analisando as amplitudes
range(dados$SurveyYear)
range(dados$Value)

# Avaliando a quantidade de categorias necessárias
nclass.Sturges(dados$SurveyYear)
nclass.Sturges(dados$Value)

# Criando as tabelas com as faixas
table(cut(dados$SurveyYear, seq(01/01/2000, 01/01/2018, l = 15)))
table(cut(dados$Value, seq(0.0, 87, l = 15)))

```

Medidas de tendência central e dispersão:

```

# Função summary (média, mediana, quartis e valores min e máx)
summary(dados$Value)

# Mediana para data
medianYear <- median(dados$SurveyYear)
print(medianYear)

# Moda
getmode <- function(dados) {
  uniqv <- unique(dados)
  uniqv[which.max(tabulate(match(dados, uniqv)))]
}
resultCountry <- getmode(dados$Gender)
resultDemoQuestion <- getmode(dados$DemographicsQuestion)
resultDemoResponse <- getmode(dados$DemographicsResponse)
resultQuestion <- getmode(dados$Question)
resultGender <- getmode(dados$Gender)
resultSurveyYear <- getmode(dados$SurveyYear)
print(resultCountry)
print(resultDemoQuestion)
print(resultDemoResponse)
print(resultQuestion)
print(resultSurveyYear)

# Função describe e describeBy (média, desvio padrão[sd], erro, mediana)
describe(dados$Value)
describeBy(dados$Value, group = dados$Gender)

# Erro
std_mean <- function(x) sd(x)/sqrt(length(x))
std_mean(dados$Value)

# Variância
var(dados$Value)

```

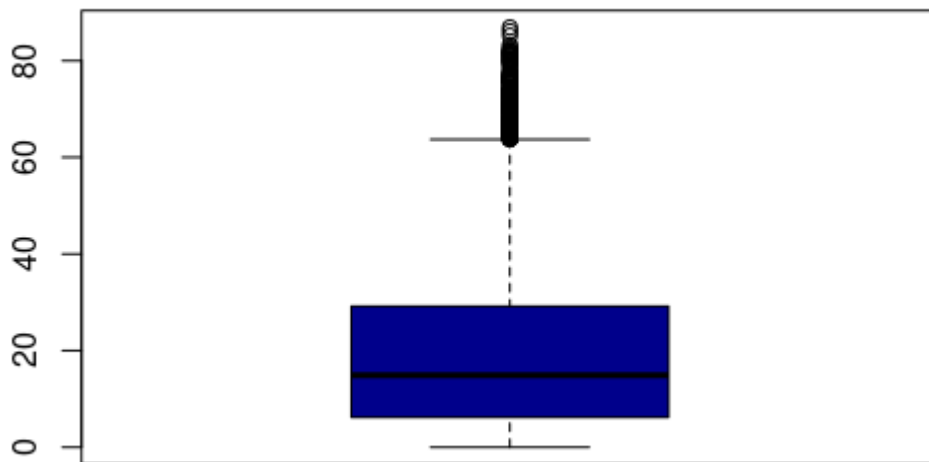
Gráficos:

```

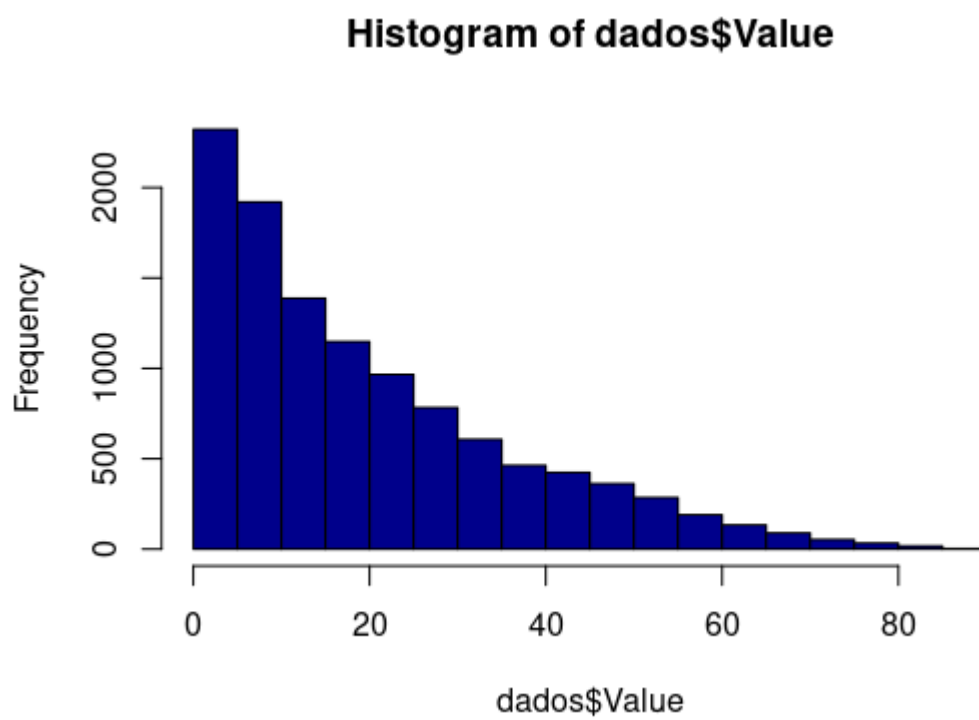
## GRÁFICOS QUANTITATIVAS
boxplotValue <- boxplot(dados$Value, col="darkblue", border="black")
hist(dados$Value, col="darkblue", border="black")

```

Boxplot:



Histograma:



```
## GRAFICOS QUALITATIVAS
x <- xtabs(~Gender, data = dados)
barplot(x,
        xlab = "Gêneros",
        ylab = "Frequência absoluta",
        col = c("seagreen", "yellowgreen"))

pie(x, col = c("#5398ed", rgb(12, 58, 114, max = 255)), main = "Gêneros")

x <- xtabs(~Country, data = dados)
barplot(x,
        xlab = "Países",
        ylab = "Frequência absoluta",
        col = c("seagreen", "yellowgreen"))

pie(x, col = c("#5398ed", rgb(12, 58, 114, max = 255)), main = "Países")

x <- xtabs(~DemographicsQuestion, data = dados)
barplot(x,
        xlab = "Perguntas demográficas",
        ylab = "Frequência absoluta",
        col = c("seagreen", "yellowgreen"))
```

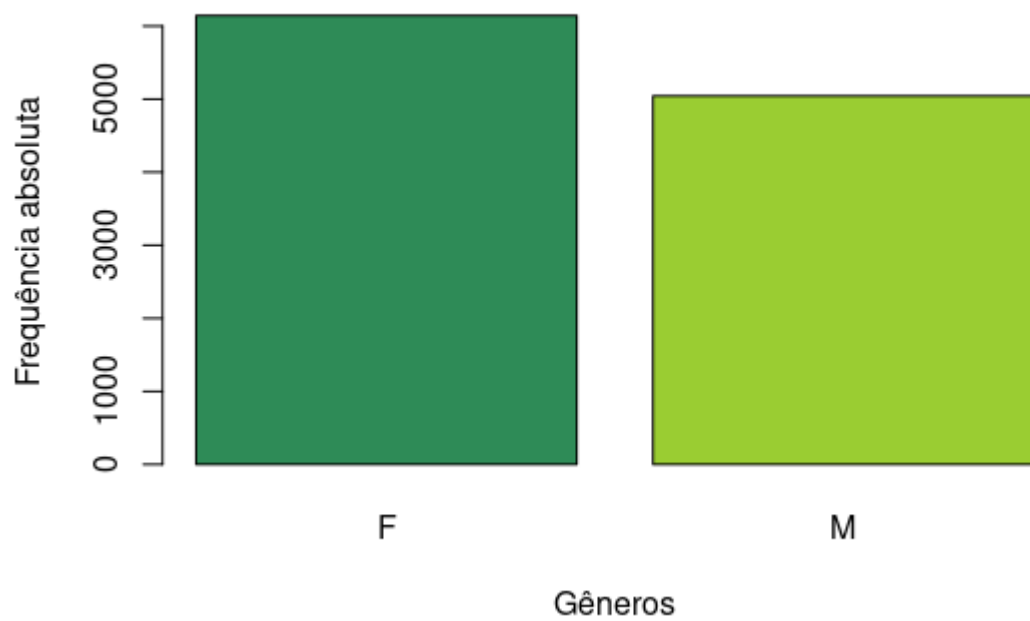
```
pie(x, col = c("#5398ed", rgb(12, 58, 114, max = 255)), main = "Perguntas demográficas")

x <- xtabs(~DemographicsResponse, data = dados)
barplot(x,
        xlab = "Respostas demográficas",
        ylab = "Frequência absoluta",
        col = c("seagreen", "yellowgreen"))

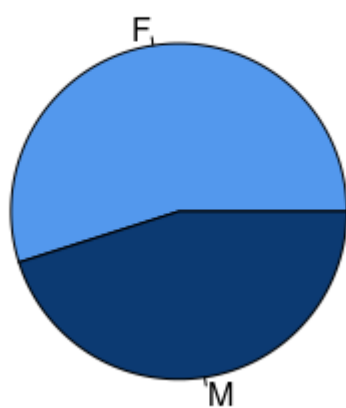
pie(x, col = c("#5398ed", rgb(12, 58, 114, max = 255)), main = "Respostas demográficas")

x <- xtabs(~Question, data = dados)
barplot(x,
        xlab = "Perguntas",
        ylab = "Frequência absoluta",
        col = c("seagreen", "yellowgreen"))
pie(x, col = c("#5398ed", rgb(12, 58, 114, max = 255)), main = "Perguntas")
```

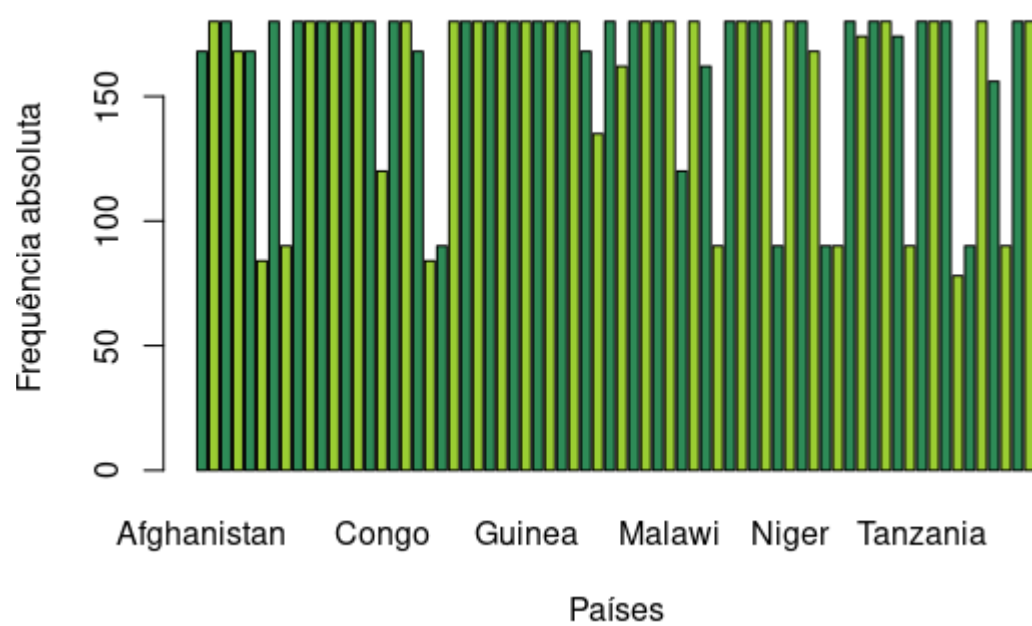
Frequência dos gêneros:



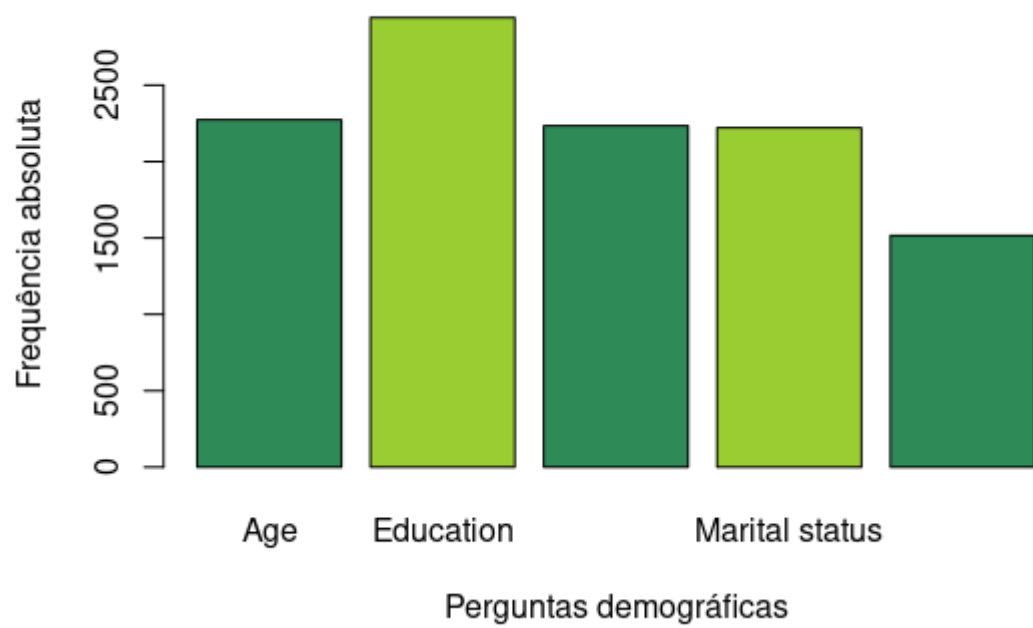
### Gêneros



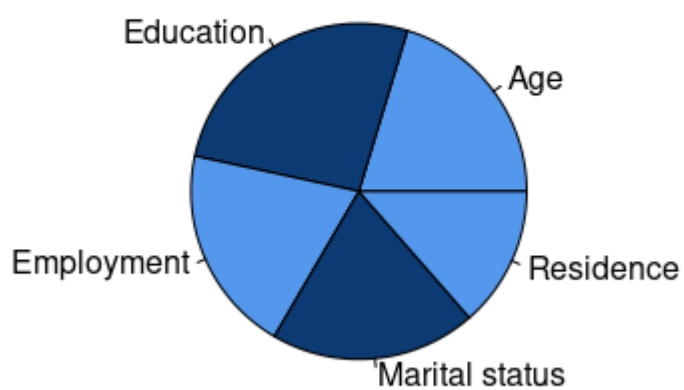
Frequência absoluta dos países:



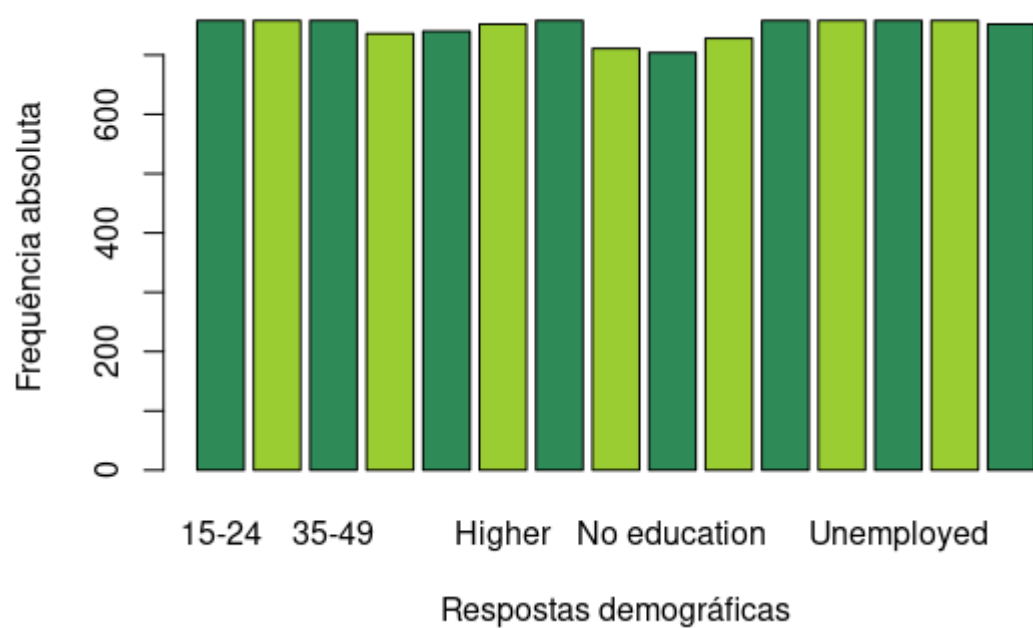
Frequência das perguntas:



## Perguntas demográficas

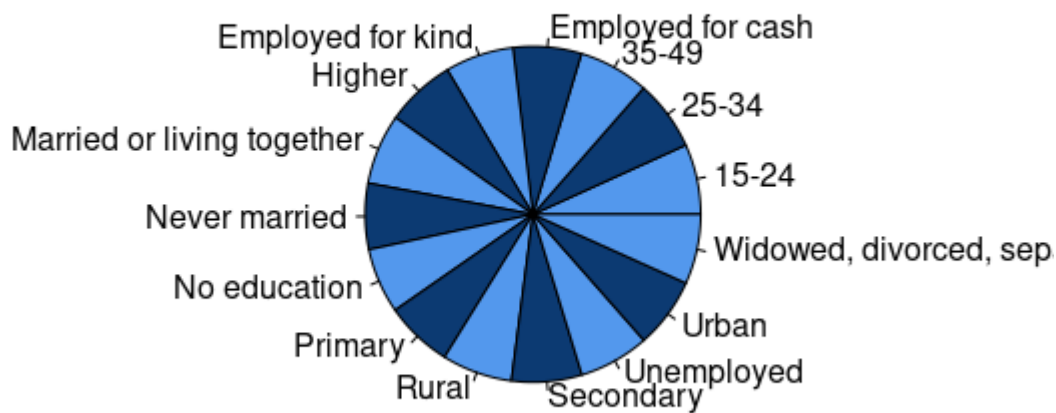


Frequência das respostas:



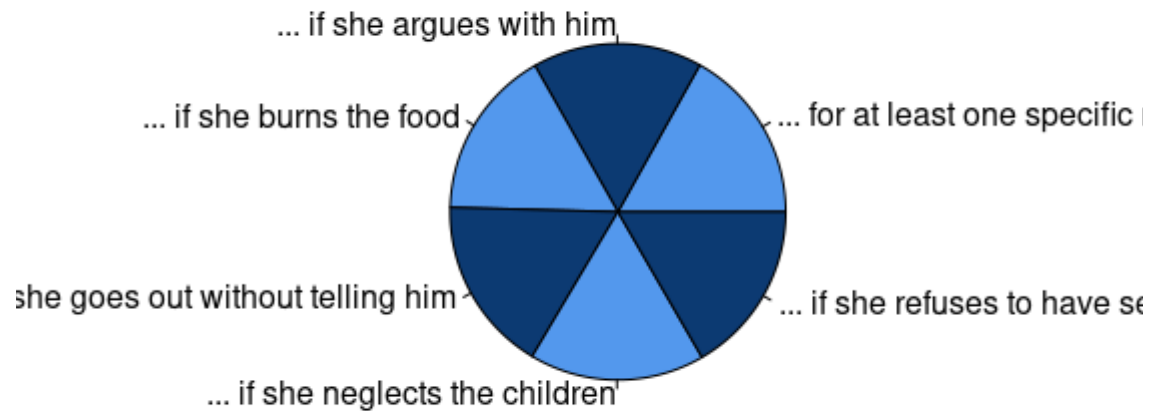


## Respostas demográficas



Perguntas:

## Perguntas



### 6. Que tipo de pesquisa/pergunta você pretende fazer com este dataset?

- Quais países têm maiores taxas de violência?
- Qual o IDH desses países?

- Qual a renda média dos habitantes dos países com menores taxas?
  - Qual a diferença ao analisar as respostas de homens e mulheres?
  - Qual idade se encontra índices menores de violência?
  - Homens que nunca foram casados possuem taxas maiores ou menores de violência?
  - Há diferença nas taxas ao analisar pessoas da área rural e urbana?
- Qual seria um título adequado para o seu trabalho?
  - Violência contra mulher: quais os maiores indicativos?