

Projeto

Leonardo Ewbank

2024-03-07

```
##carregar bibliotecas
```

```
library(renv) ##utilizado para criar pasta que irá armazenar info das bibliotecas usadas
library(tidyverse)
library(knitr) ##RMarkdown -> integrar código e exposição gerada por RMarkdown e YAML
library(summarytools)
library(ggplot2)
library(stringr) ##manipulação de strings
##library(stringi) ##converter char pra utf8 para manipulação
library(stats)
```

```
##abrir o arquivo com os dados
```

```
base_inic <- read.csv2("~/ProjetoEst/R/aux/BaseDPEvolucaoMensalCisp.csv")
##Visualizar Base
##View(base_inic)
```

```
##Função auxiliar para criar coluna no formato YYYY-MM-DD
```

```
formData <- function(YY,MM)
{
  formatted<-stringr::str_c(YY,MM,"01",sep="-")
  date_t<-as.Date(formatted,format="%Y-%m-%d")
  return(date_t)
}
```

```
##selecionando variaveis de interesse e criando uma nova base
```

| mes | ano | munic | Regiao | total_furtos | total_roubo | stentat_hon | hom_dolos | datrocínio | DataCompleta |
|-----|------|----------------|---------|--------------|-------------|-------------|-----------|------------|--------------|
| 1 | 2003 | Rio de Janeiro | Capital | 169 | 100 | 1 | 0 | 0 | 2003-01-01 |
| 1 | 2003 | Rio de Janeiro | Capital | 115 | 82 | 0 | 3 | 0 | 2003-01-01 |
| 1 | 2003 | Rio de Janeiro | Capital | 216 | 123 | 1 | 3 | 0 | 2003-01-01 |
| 1 | 2003 | Rio de Janeiro | Capital | 90 | 143 | 2 | 6 | 0 | 2003-01-01 |
| 1 | 2003 | Rio de Janeiro | Capital | 35 | 55 | 2 | 4 | 0 | 2003-01-01 |

| mes | ano | munic | Regiao | total_furtos | total_roubo | tentat_hon | hom_dolos | da | atrocinio | DataCompleta |
|-----|------|----------------|---------|--------------|-------------|------------|-----------|----|-----------|--------------|
| 1 | 2003 | Rio de Janeiro | Capital | 268 | 124 | 3 | 1 | | 0 | 2003-01-01 |

```
##removendo espaços em branco da variavel Regiao
```

```
##Media de roubos e furtos por ano para cada região
```

```
media_roubosefurtos<-base_select%>%dplyr::group_by(ano)%>%summarize(media_furto=mean(total_furtos),media_roubo=mean(total_roubo),na.rm=TRUE)
```

```
## # A tibble: 20 x 4
##   ano media_furto media_roubo na.rm
##   <int>      <dbl>      <dbl> <lgl>
## 1 2003        78.3        77.9 TRUE
## 2 2004        76.1        71.1 TRUE
## 3 2005        81.0        72.7 TRUE
## 4 2006        90.3        79.5 TRUE
## 5 2007       101.         88.1 TRUE
## 6 2008       108.         90.1 TRUE
## 7 2009       108.         88.0 TRUE
## 8 2010       110.         75.9 TRUE
## 9 2011       108.         65.4 TRUE
##10 2012       110.         63.6 TRUE
##11 2013       112.         77.2 TRUE
##12 2014       110.         95.4 TRUE
##13 2015       109.         89.3 TRUE
##14 2016       102.        126. TRUE
##15 2017        87.7        139. TRUE
##16 2018        86.9        140. TRUE
##17 2019        94.0        121. TRUE
##18 2020        67.5         75.4 TRUE
##19 2021        69.7         70.3 TRUE
##20 2022        98.6         66.0 TRUE
```

```
##Desvio padrão das medidas de total de roubos e furtos por ano
```

```
desvpad_roubosefurtos <- base_select%>% dplyr::group_by(ano)%>%summarize(desv_furto=sd(total_furtos),desv_roubo=sd(total_roubo),na.rm=TRUE)
```

```
## # A tibble: 20 x 4
##   ano desv_furto desv_roubo na.rm
##   <int>      <dbl>      <dbl> <lgl>
## 1 2003        73.4        99.1 TRUE
## 2 2004        71.2        94.7 TRUE
## 3 2005        79.4        93.8 TRUE
## 4 2006        90.8        98.7 TRUE
## 5 2007        99.0       109. TRUE
## 6 2008       110.       110. TRUE
## 7 2009       110.       109. TRUE
## 8 2010       115.        94.9 TRUE
## 9 2011       116.        86.6 TRUE
```

```
## 10 2012      121.      85.3 TRUE
## 11 2013      126.      99.4 TRUE
## 12 2014      127.     125.  TRUE
## 13 2015      123.     115.  TRUE
## 14 2016      113.     167.  TRUE
## 15 2017       98.3     188.  TRUE
## 16 2018       89.3     187.  TRUE
## 17 2019      102.     165.  TRUE
## 18 2020       84.7     113.  TRUE
## 19 2021       78.0     95.3 TRUE
## 20 2022      121.     86.9 TRUE
```

Os quantis: 25% e 75% para total de furtos

```
quant25_furtos <- quantile(base_select$total_furtos,probs = (0.25),na.rm = TRUE)
quant25_furtos
```

```
## 25%
## 18
```

```
quant75_furtos <- quantile(base_select$total_furtos,probs = (0.75),na.rm = TRUE)
quant75_furtos
```

```
## 75%
## 137
```

```
print(paste0("25%:", quant25_furtos, ", 75%:", quant75_furtos))
```

```
## [1] "25%:18, 75%:137"
```

Utilizando o pacote summarytools (função descr), descreva estatisticamente a sua base de dados.

```
summarytools::descr(base_select%>%select(tentat_hom,total_furtos,total_roubos,latrocinio,hom_doloso),na.rm = TRUE)
```

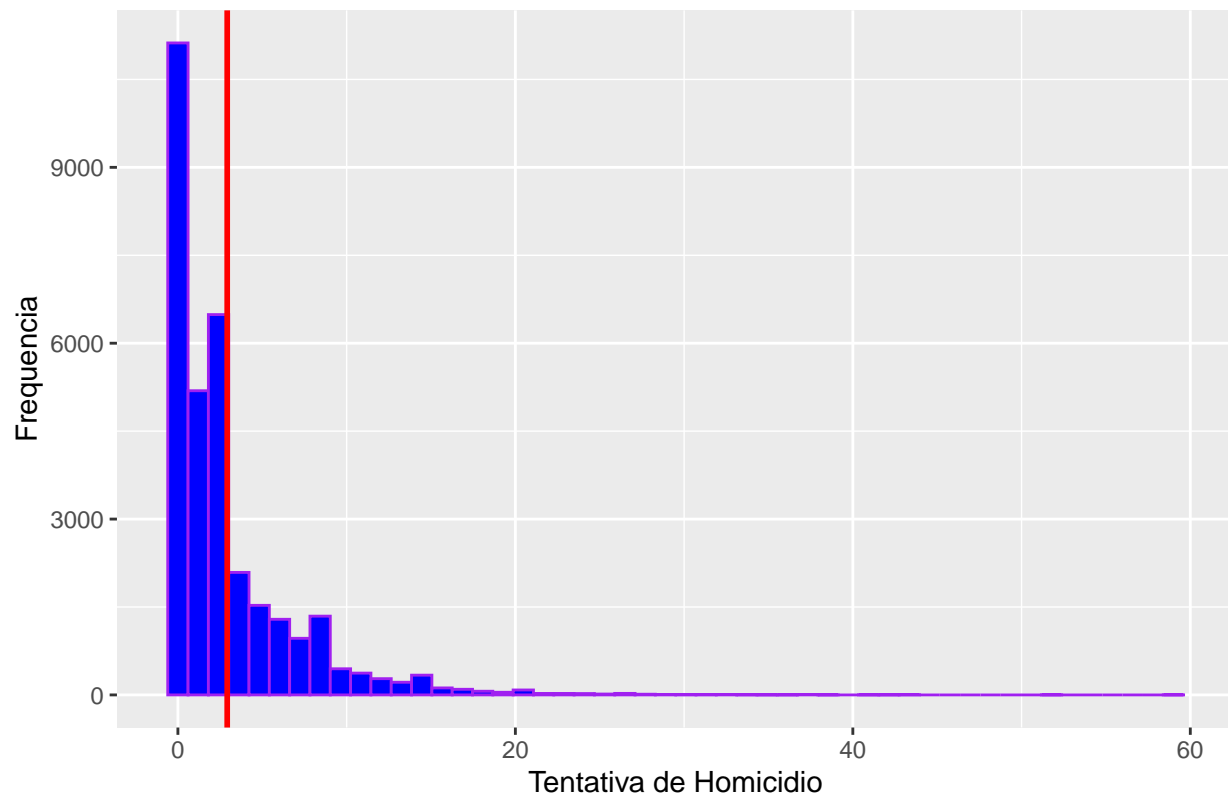
```
## Descriptive Statistics
```

```
##
##          hom_doloso  latrocinio  tentat_hom  total_furtos  total_roubos
## -----
##          Mean        3.10         0.10         2.92         95.42         88.86
##          Std.Dev      4.51         0.38         4.06        105.06        123.25
##          Min          0.00         0.00         0.00          0.00          0.00
##          Q1           0.00         0.00         0.00         18.00          3.00
##          Median       1.00         0.00         1.00         64.00         32.00
##          Q3           4.00         0.00         4.00        137.00        138.00
##          Max          43.00         9.00        59.00       1665.00       1160.00
##          MAD          1.48         0.00         1.48         75.61         47.44
##          IQR          4.00         0.00         4.00        119.00        135.00
##          CV           1.46         3.72         1.39          1.10          1.39
##          Skewness     2.49         5.21         2.66          2.62          2.27
##          SE.Skewness   0.01         0.01         0.01          0.01          0.01
##          Kurtosis      7.98        43.76        11.63         16.32          7.35
##          N.Valid     32245.00     32245.00     32245.00     32245.00     32245.00
##          Pct.Valid    100.00     100.00     100.00     100.00     100.00
```

```
## Escolha uma variável e crie um histograma. Justifique o número de bins usados.
##a distribuição dessa variável se aproxima de uma "normal"? Justifique.
##his<-hist(base_select$tentat_hom)
ggplot(base_select,aes(x=tentat_hom))+geom_histogram(color="purple",fill="blue",bins=50)+
  labs(x="Tentativa de Homicidio",y="Frequencia")+ggtitle("Histograma de Tentativas de Homicidio entre 20xx e 2022")
  geom_vline(aes(xintercept=mean(tentat_hom,na.rm=TRUE)),color="red",size=1)
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

Histograma de Tentativas de Homicidio entre 20xx e 2022



Não se aproxima de uma normal, visto que a curva não se distribui em torno do ponto médio

Calcule a correlação entre todas as variáveis dessa base. Quais são as 3 pares de variáveis mais correlacionadas?

```
correlacao <- cor(base_select%>%select(total_furtos,total_roubos,tentat_hom,hom_doloso,latrocinio))
kable(correlacao)
```

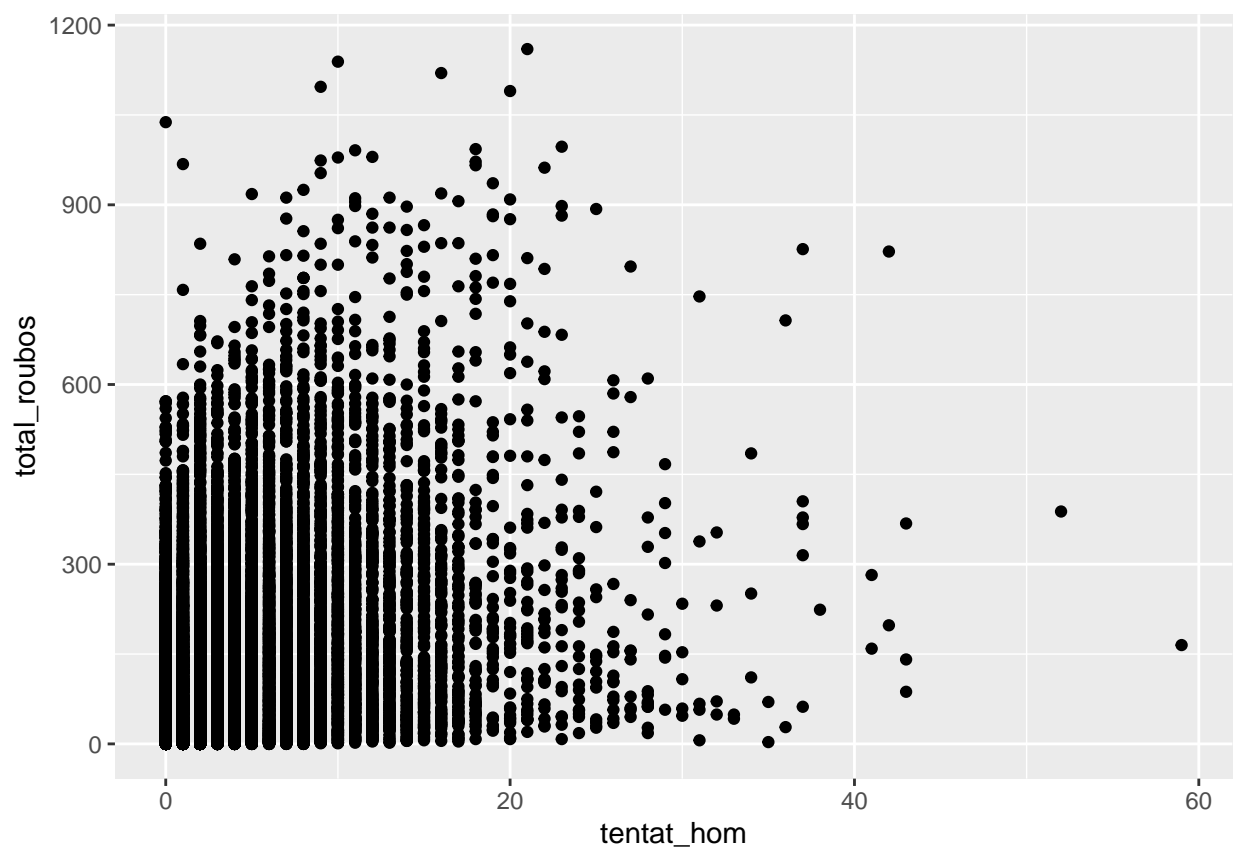
| | total_furtos | total_roubos | tentat_hom | hom_doloso | latrocinio |
|--------------|--------------|--------------|------------|------------|------------|
| total_furtos | 1.0000000 | 0.5434236 | 0.2737988 | 0.2437570 | 0.1217073 |
| total_roubos | 0.5434236 | 1.0000000 | 0.4252534 | 0.4900958 | 0.2570745 |

| | total_furtos | total_roubos | tentat_hom | hom_doloso | latrocinio |
|------------|--------------|--------------|------------|------------|------------|
| tentat_hom | 0.2737988 | 0.4252534 | 1.0000000 | 0.4986792 | 0.1560309 |
| hom_doloso | 0.2437570 | 0.4900958 | 0.4986792 | 1.0000000 | 0.2178095 |
| latrocinio | 0.1217073 | 0.2570745 | 0.1560309 | 0.2178095 | 1.0000000 |

Variáveis mais Correlacionadas: Homicídio Doloso com Total Roubo e Tentativa de Homicídio com Total Roubos

##Crie um scatterplot entre duas variáveis das resposta anterior. ##Qual a relação da imagem com a correlação entre as variáveis.

```
ggplot(base_select,aes(x=tentat_hom,y=total_roubos))+geom_point()
```



##crie um gráfico linha de duas das variáveis. Acrescente uma legenda e rótulos nos eixos.

```
s <-base_select%>%dplyr::group_by(ano)%>%dplyr::filter(munic=="Rio de Janeiro")%>%summarize(sumtr=sum(h
ggplot(s,aes(x=ano,y=sumtr,color=ano))+geom_line()+labs(x="Ano",y="Homicidio Doloso",title="Homicidios p
```

