

Curso de Iniciação com R com exemplos de Política Internacional(USP)

Trabalho Final

Humberto Bezerra de Meneses Júnior

29/04/2022

Análise de dados sobre Homicídios no RJ

1. CABEÇALHO

NOME: HUMBERTO BEZERRA DE MENESES JÚNIOR

TÍTULO: Análise de Dados sobre Homicídios no RJ

2. BASE DE DADOS, MOTIVAÇÃO E OBJETIVOS

- A base de dados é um recorte do número de Homicídios especificamente no Estado do RJ com o IDH presente por município, ressaltando a não exposição de dado pessoal individual. Base disponível em:
 - https://github.com/jrhumberto/R_politicas_internacionais/blob/main/municipios_rj.csv
 - Destaca-se que o **Índice de Desenvolvimento Humano é importante métrica para verificação de um dado populacional acerca de Políticas Internacionais**. Portanto, a motivação do presente estudo considera atributo que pode ser de interesse quando há reflexão acerca da Violência e de motivações implícitas.
 - Objetivos:
 - **Verificar alguma existência de outlier presente na coluna Taxa de Homicídios.**
 - **Analisar um diagnóstico da distribuição assimétrica dos dados.**
-

3. RESULTADOS

Uma prévia sanitização dos dados é feita através do recurso de **select combinado com filter, em prol de separar linhas que contenham taxas maiores que zero e sem o fenômeno de dados faltantes(NA) no dataset**.

Para verificar a base de dados de Homicídios no RJ, também utiliza-se o recurso Kable em prol de determinar as primeiras linhas da tabela. A função Head contribui para que o recurso disponibilize as primeiras linhas da tabela.

Não é necessário trazer todas as colunas do dataset para esta análise, as que importam são a taxa de Homicídios, o Município e a Região do RJ que contém esse numeral de taxa, por isso no parâmetro do dataset passado há um operação de slice(recorte) de colunas que interessam.

Esse parâmetro “recortado” está sendo passado para função head() combinada com recurso Kable:

===

```
url_base <- paste0("https://raw.githubusercontent.com/jrhumberto/",
                  "R_politicas_internacionais/", "master/",
                  "municipios_rj.csv", sep="")
dados0 <- read.csv(url_base)

dados1 <- dados0 %>%
  select(REGIAO, MUNICIPIO, TX_HOMICIDIO_100MIL) %>%
  filter( TX_HOMICIDIO_100MIL > 0 , !is.na(TX_HOMICIDIO_100MIL) )
knitr::kable(head(dados1[, 1:3]), "simple")
```

REGIAO	MUNICIPIO	TX_HOMICIDIO_100MIL
COSTA VERDE	Angra dos Reis	45.95
NOROESTE	Aperibé	8.96
BAIXADA LITORANEA	Araruama	53.63
BAIXADA LITORANEA	Armação dos Búzios	41.04
BAIXADA LITORANEA	Arraial do Cabo	34.39
MEDIO PARAIBA	Barra do Pirai	23.67

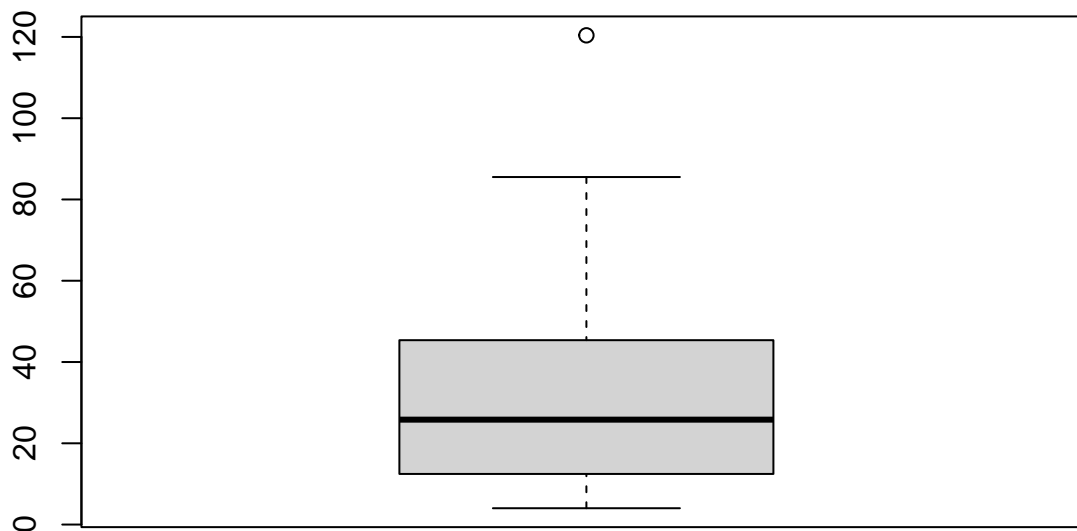
4. GRÁFICOS

Diagrama de Caixa - Boxplot

Para efetuar a primeira verificação sobre a hipótese de existência de outliers no dataset, o gráfico abaixo (também conhecido como “gráfico do bigode”) permite uma visualização com o ponto assimétrico. Nessa representação visual é possível perceber quartis, mediana e outliers no mesmo gráfico, considerando valores da coluna **Taxa de Homicídios**.

===

```
boxplot(dados1$TX_HOMICIDIO_100MIL)
```



Histograma ou Gráfico de Barras

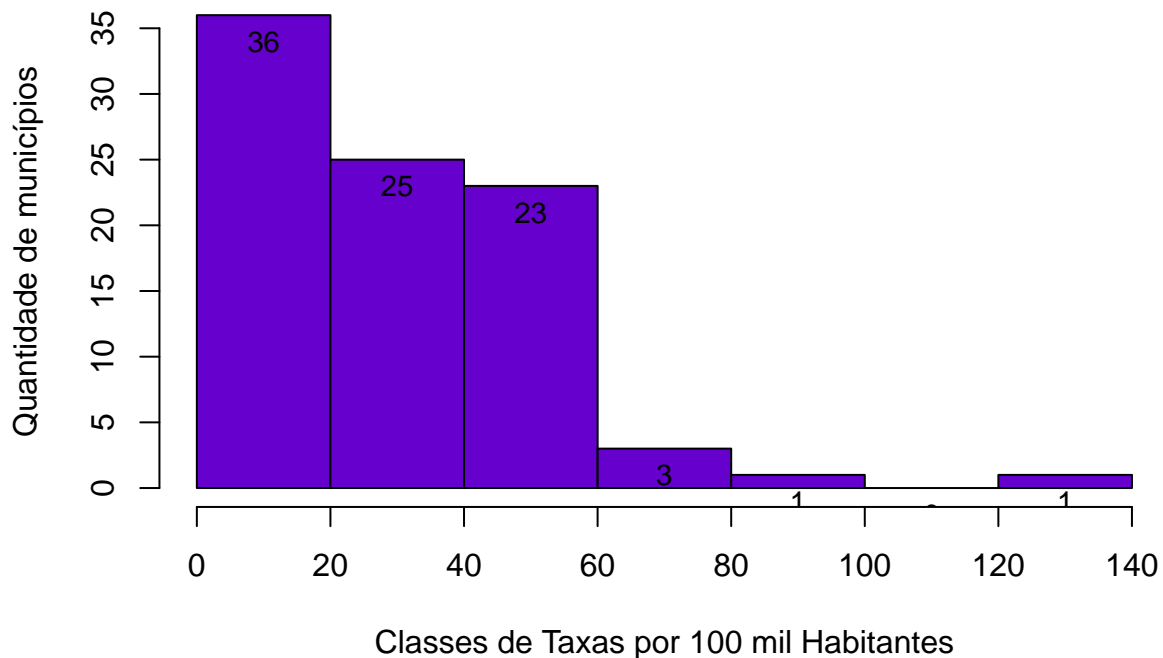
Em vista que foi percebido o ponto distante do último quartil, é possível verificar a quantidade de municípios com intervalos de classes de taxas.

Em propósito, 5 classes serão consideradas e o outlier na sexta classe. Percebe-se que realmente apenas uma cidade é o outlier evidente nos 2 gráficos. As classes facilitam o diagnóstico onde se encontra a assimetria.

====

```
bp= hist(dados1$TX_HOMICIDIO_100MIL,
  main = "Histograma Municípios com suas Taxa de Homicídios",
  xlab = "Classes de Taxas por 100 mil Habitantes",
  ylab = "Quantidade de municípios",
  col = "#6600cc",
  labels = FALSE , breaks=5 )
text(x=bp$mids, y=bp$counts, labels=bp$counts ,cex=0.9,pos=1.9)
```

Histograma Municípios com suas Taxa de Homicídios



Outlier

De posse do valor aproximado da taxa de homicídio do Outlier, é possível ativar um select com filter para um valor de taxa maior que 100(valor limite da quinta classe) e determina-se a identidade do Município que está provocando essa discrepância em relação aos demais.

====

```
dados2 <- dados1 %>%
  select(REGIAO, MUNICIPIO, TX_HOMICIDIO_100MIL) %>%
  filter( TX_HOMICIDIO_100MIL >= 100.0 )
head(dados2)
```

```
##          REGIAO MUNICIPIO TX_HOMICIDIO_100MIL
## 1 METROPOLITANA Queimados          120.39
```

5. CONCLUSÃO

Gráficos demonstrados nesta análise como o histograma e o diagrama de caixa permitiram a identificação do problema dos valores discrepantes e localizar o exato elemento que propicia ser um outlier em taxas de homicídios. Esses esquemas visuais podem ser interessantes recursos em futuras análises exploratórias. Inclusive ressalte-se em toda análise de dados a avaliação prévia de dados faltantes no dataset do problema e também uma investigação preliminar sobre outliers, que possam estar prejudicando a uniformidade amostral.