Medidas de tendência central e dispersão

Jairo Nicolau

13/03/2022

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ ggplot2 3.3.5 ✓ purrr 0.3.4  
## ✓ tibble 3.1.6 ✓ dplyr 1.0.8  
## ✓ tidyr 1.2.0 ✓ stringr 1.4.0  
## ✓ readr 2.1.2 ✓ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(here)

## here() starts at /Users/jaironicolau/Documents/GitHub/estatistica\_2022

## Banco do Rio de Janeiro

library(readxl) # para ler arquivos excel  
  
#library(here) # para ler dados nas pastas  
  
#df <- read\_excel(here("dados\_originais","bairro\_rio\_2008.xlsx"))  
  
df <- read\_excel("bairro\_rio\_2008.xlsx")

glimpse(df)

## Rows: 158  
## Columns: 79  
## $ CdGPP <chr> "5", "4", "5", "2", "4", "2", "6", "3", "7", "1", …  
## $ GPP <chr> "Central", "Periferia Norte", "Central", "Centro-N…  
## $ CdRA <chr> "33045570517", "33045570530", "33045570517", "3304…  
## $ RegAdmin <chr> "Méier", "Pavuna", "Méier", "Tijuca", "Anchieta", …  
## $ CBai <chr> "330455705060", "330455705138", "330455705057", "3…  
## $ Bairro\_IPP <chr> "Abolicao", "Acari", "Agua Santa", "Alto da Boa Vi…  
## $ NumLoclVot <dbl> 2, 3, 2, 7, 5, 11, 4, 3, 55, 19, 3, 1, 6, 8, 9, 25…  
## $ NumSecoes <dbl> 9, 27, 10, 29, 46, 86, 25, 32, 492, 165, 11, 9, 72…  
## $ Aptos2008 <dbl> 3728, 12013, 4470, 11822, 21353, 34936, 11750, 152…  
## $ T1Abst <dbl> 585, 1797, 742, 2158, 3304, 7011, 1494, 2751, 3223…  
## $ T1Abst\_pc <dbl> 15.69206, 14.95879, 16.59955, 18.25410, 15.47324, …  
## $ T1Comp <dbl> 3143, 10216, 3728, 9664, 18049, 27925, 10256, 1249…  
## $ T1Comp\_pc <dbl> 84.30794, 85.04121, 83.40045, 81.74590, 84.52676, …  
## $ T1Vali <dbl> 2701, 8792, 3239, 8414, 15560, 24620, 9158, 10788,…  
## $ T1Vali\_pcC <dbl> 85.93700, 86.06108, 86.88305, 87.06540, 86.20976, …  
## $ T1Bran <dbl> 150, 736, 188, 548, 1093, 1264, 471, 650, 10844, 1…  
## $ T1Bran\_pcC <dbl> 4.772510, 7.204385, 5.042918, 5.670530, 6.055737, …  
## $ T1Nulo <dbl> 292, 688, 301, 702, 1396, 2041, 627, 1057, 13855, …  
## $ T1Nulo\_pcC <dbl> 9.290487, 6.734534, 8.074034, 7.264073, 7.734501, …  
## $ Gabeira <dbl> 724, 687, 717, 2593, 1898, 9723, 2378, 2353, 21345…  
## $ Gab\_pcC <dbl> 23.035317, 6.724745, 19.232833, 26.831540, 10.5158…  
## $ Eduardo <dbl> 822, 3073, 1166, 1988, 5618, 7000, 3185, 3718, 549…  
## $ Edu\_pcC <dbl> 26.15336, 30.08027, 31.27682, 20.57119, 31.12638, …  
## $ Crivella <dbl> 438, 3020, 591, 1383, 3956, 2732, 1836, 2336, 3897…  
## $ Criv\_pcC <dbl> 13.935730, 29.561472, 15.853004, 14.310844, 21.918…  
## $ Jandira <dbl> 307, 942, 326, 900, 1815, 1884, 601, 1178, 16555, …  
## $ Jan\_pcC <dbl> 9.767738, 9.220830, 8.744635, 9.312914, 10.055959,…  
## $ Molon <dbl> 185, 376, 143, 457, 845, 1170, 473, 442, 10302, 13…  
## $ Mol\_pcC <dbl> 5.886096, 3.680501, 3.835837, 4.728891, 4.681700, …  
## $ Chico <dbl> 62, 79, 75, 189, 209, 736, 99, 212, 2644, 509, 38,…  
## $ Chi\_pcC <dbl> 1.9726376, 0.7732968, 2.0118026, 1.9557119, 1.1579…  
## $ Solange <dbl> 93, 295, 154, 281, 700, 878, 421, 230, 9281, 884, …  
## $ RSol\_pcC <dbl> 2.958956, 2.887627, 4.130901, 2.907699, 3.878331, …  
## $ PauloR <dbl> 52, 113, 49, 495, 365, 347, 125, 138, 3646, 549, 1…  
## $ RPau\_pcC <dbl> 1.6544703, 1.1061081, 1.3143777, 5.1221026, 2.0222…  
## $ `1FilipePereira` <dbl> 10, 154, 10, 105, 64, 76, 18, 35, 1129, 99, 28, 14…  
## $ `2ViniciusCordeiro` <dbl> 3, 40, 6, 10, 77, 46, 14, 138, 570, 46, 8, 11, 132…  
## $ `3EduardoSerra` <dbl> 4, 11, 1, 12, 10, 22, 7, 8, 109, 26, 2, 1, 13, 25,…  
## $ `4AntonioCarlos` <dbl> 1, 2, 1, 1, 3, 6, 1, 0, 36, 9, 0, 0, 10, 10, 3, 13…  
## $ Outros\_1\_4 <dbl> 18, 207, 18, 128, 154, 150, 40, 181, 1844, 180, 38…  
## $ ROut\_pcC <dbl> 0.5727012, 2.0262334, 0.4828326, 1.3245033, 0.8532…  
## $ T2Abst <dbl> 651, 2356, 895, 2421, 3915, 7522, 1906, 3060, 3840…  
## $ T2Abst\_pc <dbl> 17.46245, 19.61209, 20.02237, 20.47877, 18.33466, …  
## $ T2Comp <dbl> 3077, 9657, 3575, 9401, 17438, 27414, 9844, 12186,…  
## $ T2Comp\_pc <dbl> 82.53755, 80.38791, 79.97763, 79.52123, 81.66534, …  
## $ T2Vali <dbl> 2781, 8721, 3291, 8595, 15778, 25330, 9165, 11075,…  
## $ T2Vali\_pcC <dbl> 90.38024, 90.30755, 92.05594, 91.42644, 90.48056, …  
## $ T2Bran <dbl> 82, 340, 68, 244, 522, 516, 188, 318, 5263, 672, 1…  
## $ T2Bran\_pcC <dbl> 2.664933, 3.520762, 1.902098, 2.595469, 2.993463, …  
## $ T2Nulo <dbl> 214, 596, 216, 562, 1138, 1568, 491, 793, 11698, 1…  
## $ T2Nulo\_pcC <dbl> 6.954826, 6.171689, 6.041958, 5.978087, 6.525978, …  
## $ T2Gabe <dbl> 1385, 3138, 1591, 5635, 5718, 15511, 4685, 5036, 6…  
## $ T2Gabe\_pcC <dbl> 45.01137, 32.49456, 44.50350, 59.94043, 32.79046, …  
## $ T2Gabe\_pcV <dbl> 49.80223, 35.98211, 48.34397, 65.56137, 36.24033, …  
## $ T2Edua <dbl> 1396, 5583, 1700, 2960, 10060, 9819, 4480, 6039, 9…  
## $ T2Edua\_pcC <dbl> 45.36887, 57.81299, 47.55245, 31.48601, 57.69010, …  
## $ T2Edua\_pcV <dbl> 50.19777, 64.01789, 51.65603, 34.43863, 63.75967, …  
## $ Pop2000 <dbl> 12346, 24650, 7243, 8254, 53808, 38540, 21551, 121…  
## $ PopM <dbl> 5621, 11969, 3940, 3901, 25676, 17067, 10272, 5815…  
## $ PopM\_pc <dbl> 45.52892, 48.55578, 54.39735, 47.26193, 47.71781, …  
## $ PopF <dbl> 6725, 12681, 3303, 4353, 28132, 21473, 11279, 6311…  
## $ PopF\_pc <dbl> 54.47108, 51.44422, 45.60265, 52.73807, 52.28219, …  
## $ Pop\_adulta <dbl> 9563, 15343, 5616, 6250, 36621, 30181, 16059, 8808…  
## $ Pop70a\_mais <dbl> 1010, 587, 372, 657, 2085, 3507, 1028, 565, 9960, …  
## $ Pop70pc <dbl> 10.561539, 3.825849, 6.623932, 10.512000, 5.693455…  
## $ Dom <dbl> 4024, 6808, 2903, 2596, 15299, 13122, 6345, 3539, …  
## $ DomFav <dbl> 0, 5711, 203, 917, 2586, 1349, 557, 1295, 7578, 27…  
## $ DomFav\_pc <dbl> 0.0000000, 83.8866040, 6.9927661, 35.3235747, 16.9…  
## $ RDom\_Media <dbl> 527.31, 171.63, 380.68, 719.70, 318.98, 1036.49, 8…  
## $ RDom\_SM\_Md <dbl> 3.492119, 1.136623, 2.521060, 4.766225, 2.112450, …  
## $ AEstu\_Media <dbl> 7, 4, 7, 7, 6, 9, 8, 7, 6, 11, 5, 5, 6, 7, 7, 10, …  
## $ Rlg\_CAR <dbl> 8277, 11983, 4753, 6227, 27700, 25581, 14283, 7078…  
## $ Rlg\_CAR\_pc <dbl> 67.04196, 48.61258, 65.62198, 75.44221, 51.47933, …  
## $ Rlg\_EMis <dbl> 373, 1408, 262, 334, 3543, 1624, 968, 866, 17454, …  
## $ Rlg\_EMis\_pc <dbl> 3.019844, 5.713712, 3.621290, 4.045675, 6.585340, …  
## $ Rlg\_EPtc <dbl> 976, 3945, 638, 666, 10756, 2394, 1589, 1424, 4553…  
## $ Rlg\_EPtc\_pc <dbl> 7.901426, 16.003043, 8.804363, 8.062757, 19.988682…  
## $ Rlg\_Sem <dbl> 1135, 6446, 1290, 698, 8571, 3395, 2744, 1954, 402…  
## $ Rlg\_Sem\_pc <dbl> 9.193261, 26.150101, 17.810300, 8.456506, 15.92885…

#### selecionando as variaveis importantes

bairro <- df %>%   
 select(Bairro\_IPP, GPP, RegAdmin, Dom, DomFav, Pop2000, PopF,PopM, RDom\_Media, Rlg\_CAR, Rlg\_EMis\_pc,Rlg\_EPtc\_pc, AEstu\_Media)  
  
glimpse(bairro)

## Rows: 158  
## Columns: 13  
## $ Bairro\_IPP <chr> "Abolicao", "Acari", "Agua Santa", "Alto da Boa Vista", "A…  
## $ GPP <chr> "Central", "Periferia Norte", "Central", "Centro-Norte", "…  
## $ RegAdmin <chr> "Méier", "Pavuna", "Méier", "Tijuca", "Anchieta", "Vila Is…  
## $ Dom <dbl> 4024, 6808, 2903, 2596, 15299, 13122, 6345, 3539, 75040, 3…  
## $ DomFav <dbl> 0, 5711, 203, 917, 2586, 1349, 557, 1295, 7578, 275, 0, 37…  
## $ Pop2000 <dbl> 12346, 24650, 7243, 8254, 53808, 38540, 21551, 12126, 2445…  
## $ PopF <dbl> 6725, 12681, 3303, 4353, 28132, 21473, 11279, 6311, 125134…  
## $ PopM <dbl> 5621, 11969, 3940, 3901, 25676, 17067, 10272, 5815, 119384…  
## $ RDom\_Media <dbl> 527.31, 171.63, 380.68, 719.70, 318.98, 1036.49, 856.93, 4…  
## $ Rlg\_CAR <dbl> 8277, 11983, 4753, 6227, 27700, 25581, 14283, 7078, 126886…  
## $ Rlg\_EMis\_pc <dbl> 3.019844, 5.713712, 3.621290, 4.045675, 6.585340, 4.214297…  
## $ Rlg\_EPtc\_pc <dbl> 7.901426, 16.003043, 8.804363, 8.062757, 19.988682, 6.2129…  
## $ AEstu\_Media <dbl> 7, 4, 7, 7, 6, 9, 8, 7, 6, 11, 5, 5, 6, 7, 7, 10, 7, 8, 7,…

#### criando variáveis

bairro %>%   
mutate(pct\_favela = DomFav/ Dom,  
 pct\_evangelico = Rlg\_EMis\_pc + Rlg\_EPtc\_pc) -> bairro

#### salvando o banco

#library(readr)  
write\_csv(bairro, "bairros\_rio.csv")

## Começão de um banco criado

df2 <- read\_csv(here("bairros\_rio.csv"))

## Rows: 158 Columns: 15  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (3): Bairro\_IPP, GPP, RegAdmin  
## dbl (12): Dom, DomFav, Pop2000, PopF, PopM, RDom\_Media, Rlg\_CAR, Rlg\_EMis\_pc...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

### Fazendo sinteses

bairro %>%   
 group\_by(GPP) %>%  
 summarize(total\_bairros = n(),  
 media\_favela = mean (pct\_favela)\*100,  
 mediana\_evangelico = median(pct\_evangelico),  
 desvio\_favela = sd(pct\_favela),  
 total\_pop = sum (Pop2000)) -> SUMARIO

#### Sumário de 5 Números

bairro %>%   
 summarize( n = n(),   
 Min = min(pct\_favela),  
 Q1 = quantile(pct\_favela, .25),  
 Media = mean(pct\_favela),   
 Q3 = quantile(pct\_favela, .75),  
 Max = max(pct\_favela)  
 )

## # A tibble: 1 × 6  
## n Min Q1 Media Q3 Max  
## <int> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 158 0 0.0217 0.175 0.232 1

bairro %>%   
 summary

## Bairro\_IPP GPP RegAdmin Dom   
## Length:158 Length:158 Length:158 Min. : 29   
## Class :character Class :character Class :character 1st Qu.: 3698   
## Mode :character Mode :character Mode :character Median : 7728   
## Mean :11633   
## 3rd Qu.:14161   
## Max. :85569   
## DomFav Pop2000 PopF PopM   
## Min. : 0.0 Min. : 136 Min. : 71 Min. : 65   
## 1st Qu.: 202.2 1st Qu.: 12139 1st Qu.: 6374 1st Qu.: 5630   
## Median : 887.0 Median : 25026 Median : 13327 Median : 11955   
## Mean : 1953.0 Mean : 37075 Mean : 19682 Mean : 17393   
## 3rd Qu.: 2282.5 3rd Qu.: 45862 3rd Qu.: 24642 3rd Qu.: 20319   
## Max. :20776.0 Max. :297494 Max. :154914 Max. :142580   
## RDom\_Media Rlg\_CAR Rlg\_EMis\_pc Rlg\_EPtc\_pc   
## Min. : 158.0 Min. : 50 Min. : 0.000 Min. : 0.00   
## 1st Qu.: 345.4 1st Qu.: 7595 1st Qu.: 3.094 1st Qu.: 7.90   
## Median : 487.3 Median : 15200 Median : 4.398 Median :11.20   
## Mean : 713.3 Mean : 22506 Mean : 5.018 Mean :11.63   
## 3rd Qu.: 779.6 3rd Qu.: 28133 3rd Qu.: 6.207 3rd Qu.:15.42   
## Max. :3375.4 Max. :153328 Max. :38.934 Max. :26.21   
## AEstu\_Media pct\_favela pct\_evangelico   
## Min. : 3.000 Min. :0.00000 Min. : 3.213   
## 1st Qu.: 6.000 1st Qu.:0.02171 1st Qu.:11.382   
## Median : 7.000 Median :0.09681 Median :16.158   
## Mean : 6.937 Mean :0.17496 Mean :16.646   
## 3rd Qu.: 8.000 3rd Qu.:0.23176 3rd Qu.:20.759   
## Max. :11.000 Max. :1.00000 Max. :38.934

bairro %>%   
select(pct\_favela) %>%  
 summary()

## pct\_favela   
## Min. :0.00000   
## 1st Qu.:0.02171   
## Median :0.09681   
## Mean :0.17496   
## 3rd Qu.:0.23176   
## Max. :1.00000

### Encontrando quantis específicos

bairro %>%   
 summarize(   
 bottom\_10 = quantile(pct\_evangelico, .10),  
 mediana = mean(pct\_favela),   
 top\_25 = quantile(pct\_evangelico, .75),  
 top\_1 = quantile(pct\_evangelico, .99)  
 )

## # A tibble: 1 × 4  
## bottom\_10 mediana top\_25 top\_1  
## <dbl> <dbl> <dbl> <dbl>  
## 1 8.05 0.175 20.8 32.7

#### Calculando o z-score

mean(bairro$pct\_favela)

## [1] 0.1749603

sd(bairro$pct\_favela)

## [1] 0.2171894

bairro %>%   
 mutate(zscore = (pct\_favela- mean(pct\_favela))/sd(pct\_favela)) ->bairro

ggplot(bairro, aes(zscore)) +  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

