

Métodos Quantitativos II - Lista 1

Professor Manoel Galdino e Monitor Davi Veronese

September 4, 2023

Esta primeira lista destina-se a exercitar funções básicas no software R. Nos exercícios, utilizaremos dados demográficos disponibilizados pelo IBGE.

Os alunos deverão entregar um arquivo PDF gerado no próprio R e o script para replicação. Para gerar o PDF, podem utilizar o RMarkdown ou o Rsweave.

```
# Material de apoio para esta lista:  
# https://jonnyphillips.github.io/Analise\_de\_Dados\_2022/
```

1

O pacote `ribge` permite importar bases de dados diretamente para o environment do R. Primeiro, instale e ative o pacote. Depois, importe os dados de 2020 por meio do código abaixo.

```
# install.packages("devtools")  
# devtools::install_github("tbrugz/ribge")  
library(ribge)  
pop2020 <- populacao_municipios(2020)  
  
# Para mais informações: https://github.com/tbrugz/ribge#readme
```

Qual é a unidade de análise desse banco de dados?

2

Para este exercício e os posteriores, vamos analisar apenas dados relativos ao estado de São Paulo. Selecione apenas as observações referentes ao estado de São Paulo. Isso é equivalente a remover da base as observações relativas a outros estados brasileiros.

Antes da análise, limpe a base de dados conforme os seguintes passos: (i) remova as variáveis "codigo uf" e "populacao str", (ii) renomeie "nome munic" para "municipio" e (iii) para todos os nomes de municipios contidos na sua nova variável municipio, coloque todos os caracteres em letra minúscula.

Quantos municípios há no estado de São Paulo?

Qual é o menor município do estado? Quantos habitantes ele tem?

```
library(tidyverse)
library(tidylog)

pop2020 <- pop2020 %>%
  filter(uf == "SP") %>%
  select(-codigo_uf, -populacao_str) %>%
  rename(
    municipio = nome_munic
  ) %>%
  mutate(municipio = tolower(municipio))

# A base resultante tem 645 observações.
# O menor município é Borá, com 838 habitantes.
```

3

Agora vamos utilizar estatísticas para conhecer as características de nossa base de dados, especialmente da variável "populacao".

Para a variável "populacao", calcule: (i) a média, (ii) a mediana, (iii) o desvio padrão e (iv) a variância.

Apresente os resultados em uma tabela (dica: usar a função "kable" do pacote "knitr"). Coloque um título na sua tabela.

```
# install.packages("knitr")
library(knitr)
```

```
média <- pop2020 %>%
  summarize(média = mean(populacao))
mediana <- pop2020 %>%
```

Table 1: Estatísticas Descritivas da População do Estado de São Paulo em 2020

média	mediana	desvio_padrão	variância
71766.41	14141	498489.9	248492160624

```

summarize(mediana = median(populacao))
desvio_padrão <- pop2020 %>%
  summarize(desvio_padrão = sd(populacao))
variância <- pop2020 %>%
  summarize(variância = var(populacao))

tabela1 <- bind_cols(média, mediana, desvio_padrão, variância)

tabela1 %>% kable(caption = "Estatísticas Descritivas da População do Estado de São Paulo e

```

4

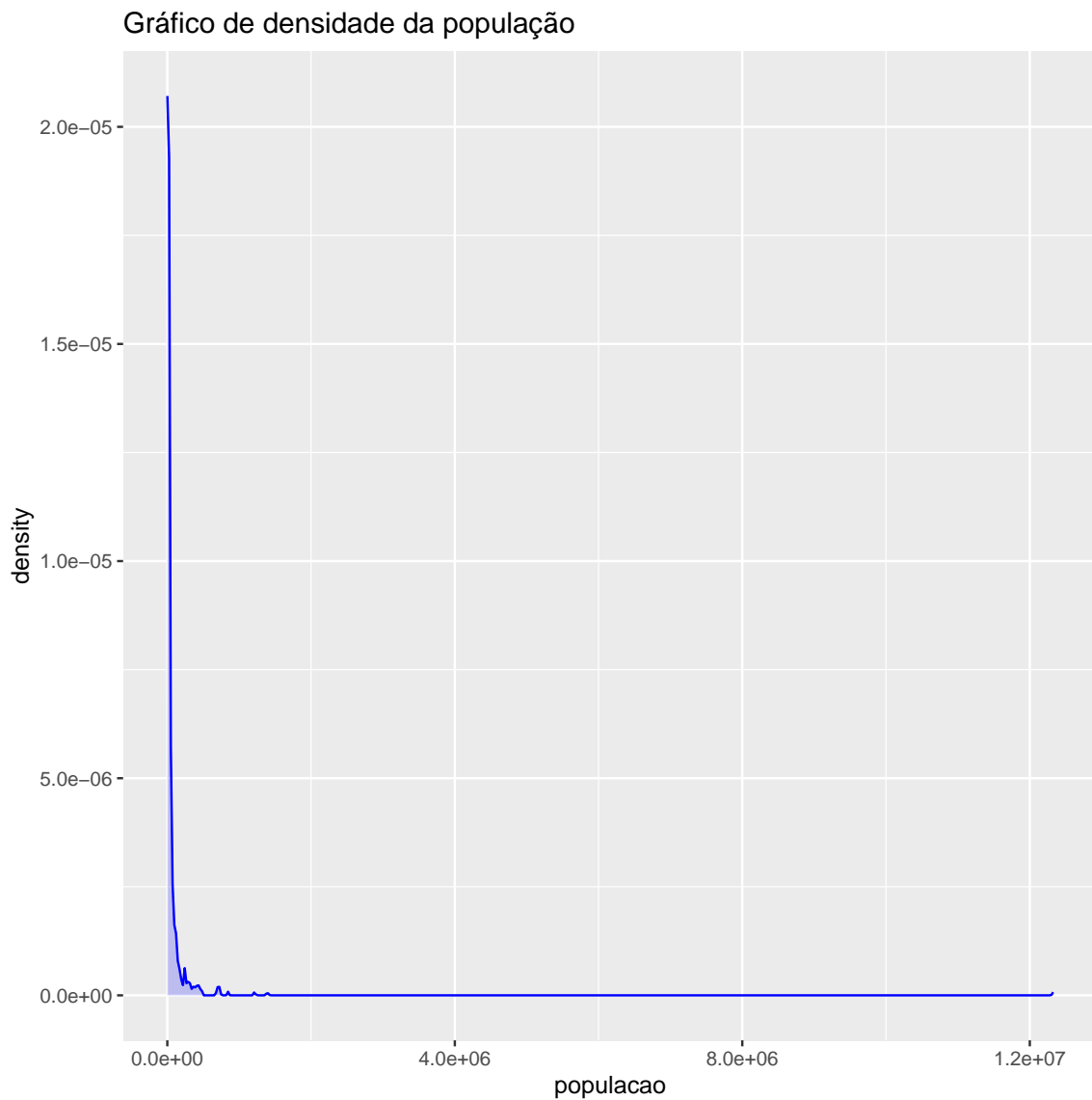
Note que essas estatísticas não fornecem informações suficientes sobre a distribuição da população. Crie um gráfico de densidade que permita visualizar essa distribuição (dica: use o pacote ggplot2).

```

library(ggplot2)

pop2020 %>%
  ggplot() +
  geom_density(aes(x=populacao), colour="blue", fill="blue", alpha=0.2) +
  ggtitle("Gráfico de densidade da população")

```



O que você observa?

Qual parece ser a medida mais adequada de tendência central: a média ou a mediana?

A distribuição é muito assimétrica, de modo que a mediana é uma medida descritiva melhor.

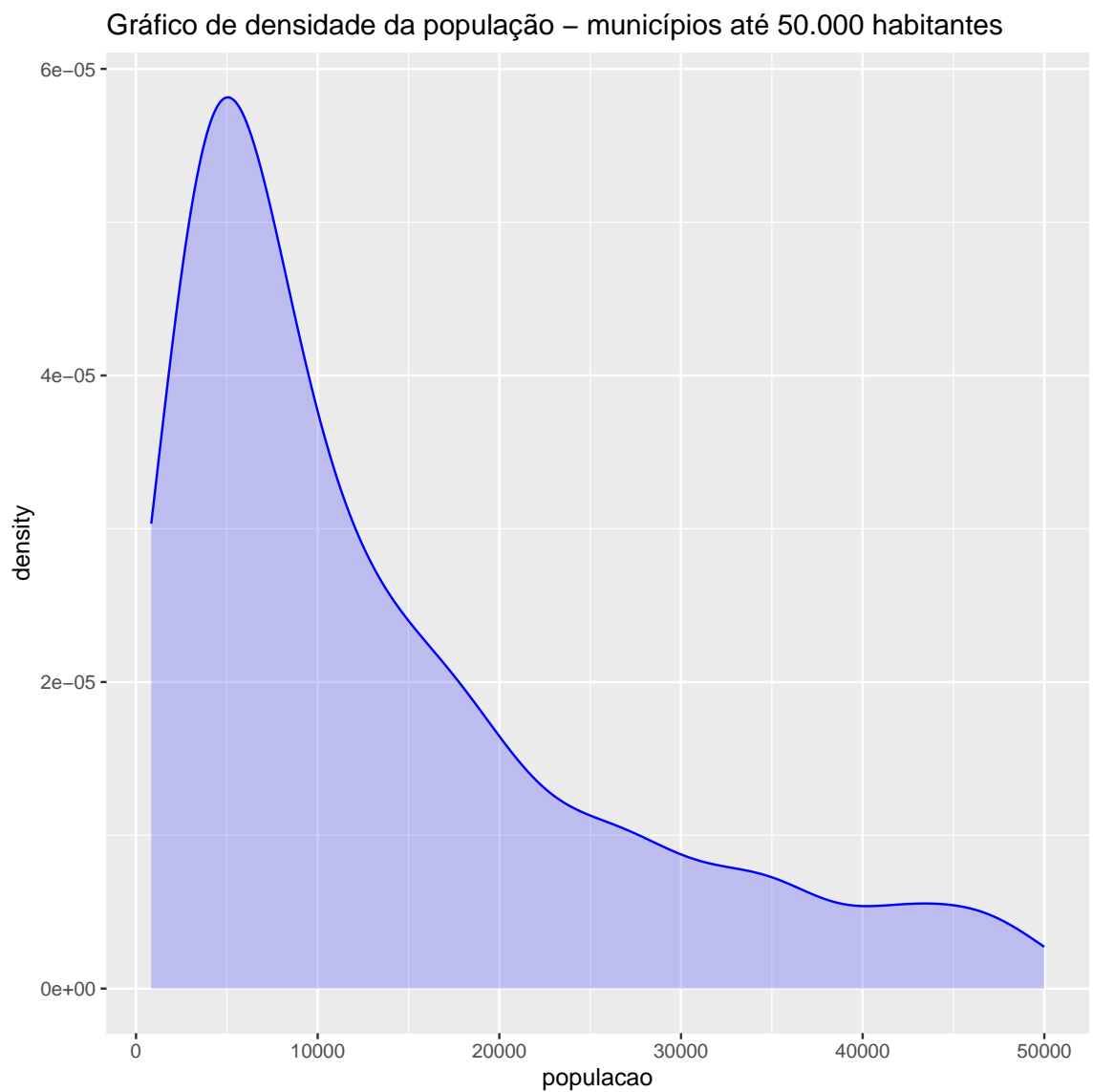
5

Agora crie novamente o gráfico de densidade, mas apenas para os municípios com menos de 50.000 habitantes.

Quantos municípios com menos de 50.000 há?

Em comparação ao gráfico anterior, o que você observa?

```
pop2020 %>%  
  filter(populacao < 50000) %>%  
  ggplot() +  
  geom_density(aes(x=populacao), colour="blue", fill="blue", alpha=0.2) +  
  ggtitle("Gráfico de densidade da população – municípios até 50.000 habitantes")
```



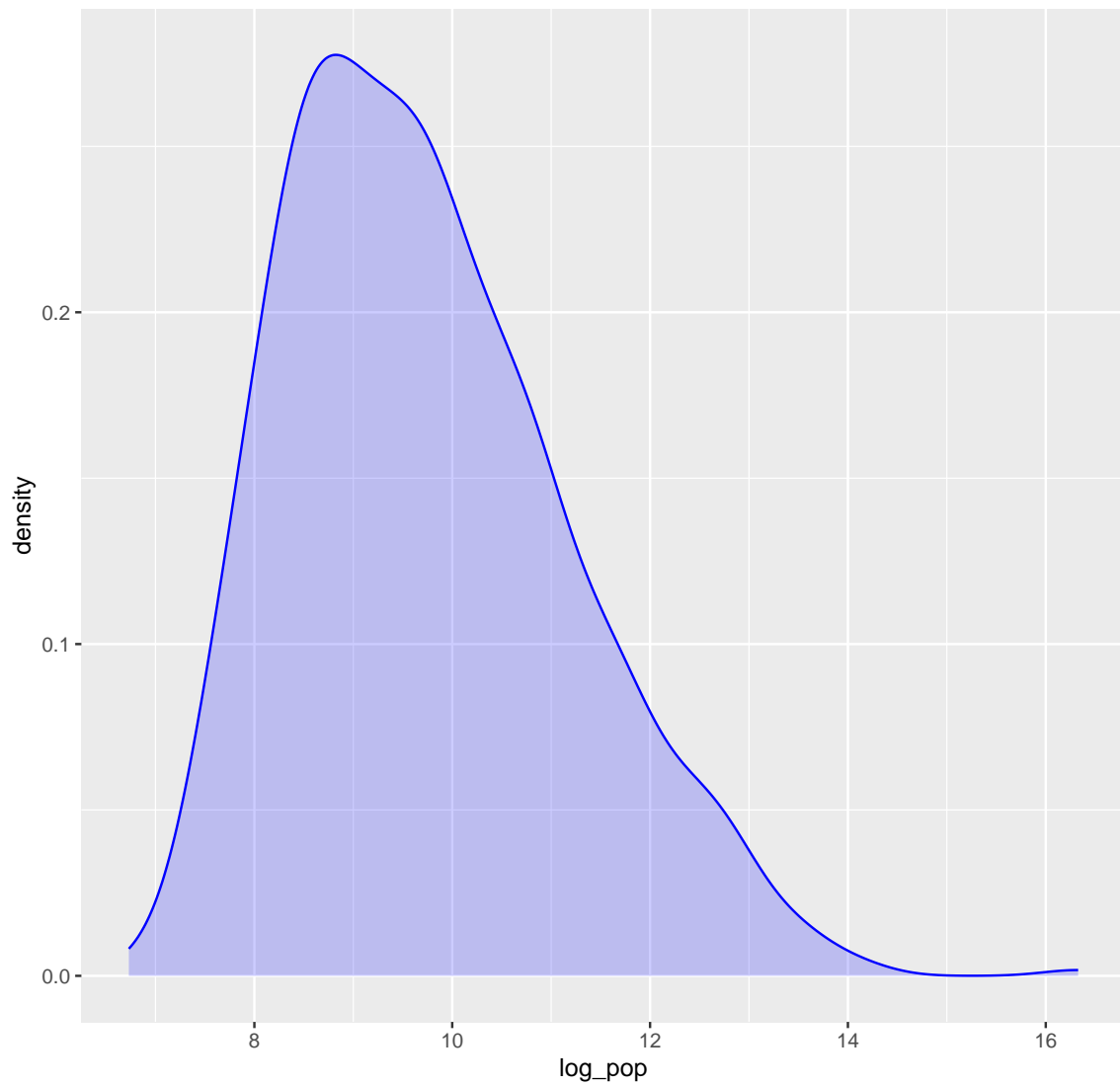
```
# Há 504 municípios com menos de 50.000 habitantes.  
# A distribuição ainda é assimétrica, porém consideravelmente menos.
```

6

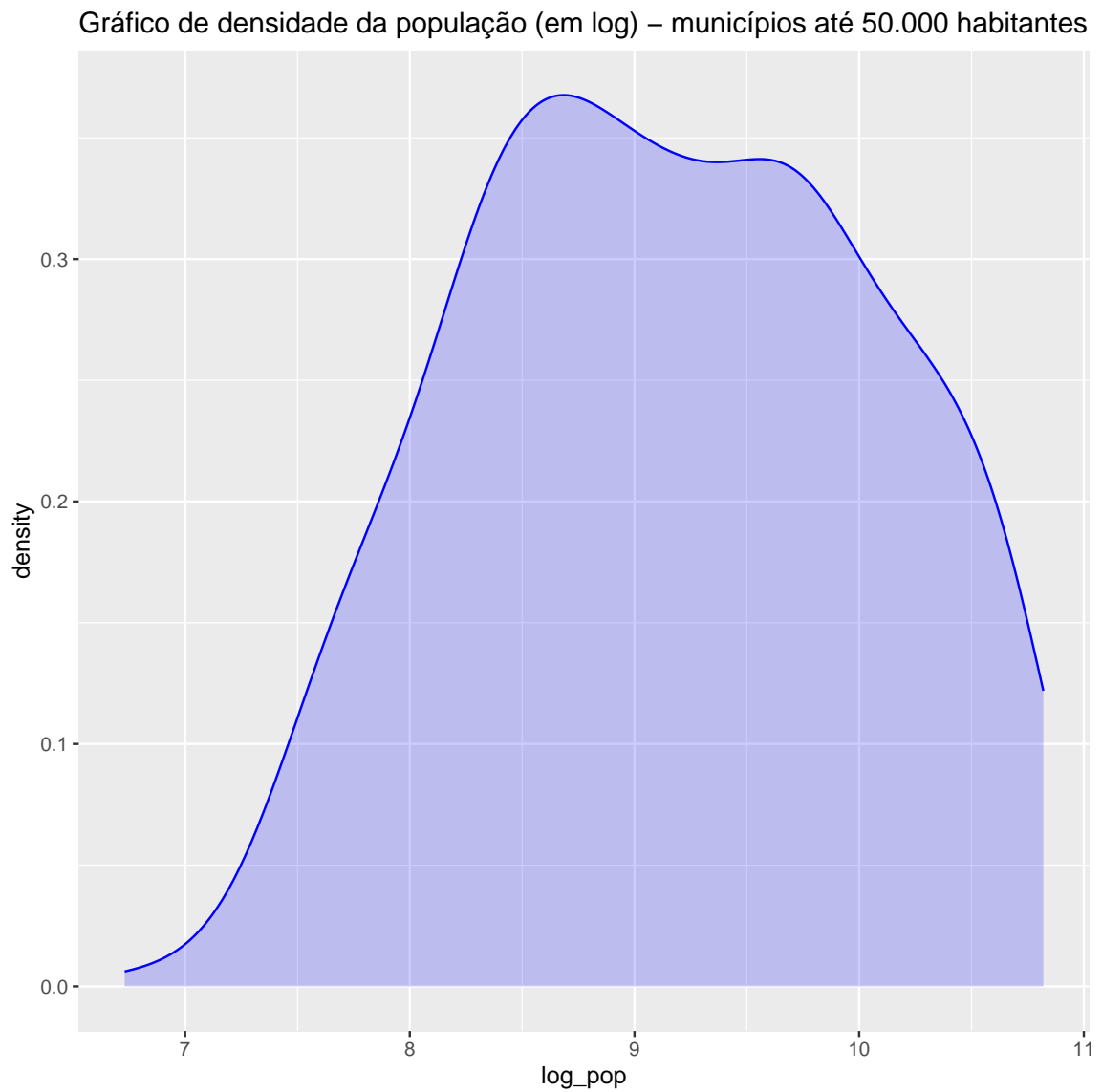
Repita o mesmo exercício, mas agora utilize a variável de população em logaritmo. O que você observa em seu novo gráfico de densidade?

```
pop2020 <- pop2020 %>%  
  mutate(log_pop = log(populacao))  
  
pop2020 %>%  
  ggplot() +  
  geom_density(aes(x=log_pop), colour="blue", fill="blue", alpha=0.2) +  
  ggtitle("Gráfico de densidade da população (em log)")
```

Gráfico de densidade da população (em log)



```
pop2020 %>%  
  filter(populacao < 50000) %>%  
  ggplot() +  
  geom_density(aes(x=log_pop), colour="blue", fill="blue", alpha=0.2) +  
  ggtitle("Gráfico de densidade da população (em log) - municípios até 50.000 habitantes")
```



7

Para esta questão, importe novamente a base de dados original (para 2020).

Calcule a média da população para cada um dos estados brasileiros e informe quais deles possuem maior e menor população média por município. Apresente seus resultados em uma tabela.

Table 2: Média da população nos municípios por estado

uf	pop_mean
TO	11440.63
PI	14649.46
PB	18113.35
RN	21162.66
RS	22983.85
SC	24584.75
MG	24962.09
MT	25008.65
PR	28864.26
GO	28916.83
SE	30917.63
MA	32786.17
AL	32858.26
RO	34547.31
MS	35561.95
BA	35804.88
AC	40657.73
RR	42078.73
CE	49929.91
PE	51981.74
ES	52103.23
AP	53860.81
PA	60352.40
AM	67866.35
SP	71766.41
RJ	188762.92
DF	3055149.00

```
pop2020 <- populacao_municipios(2020)

medias_pop_estados <- pop2020 %>%
  group_by(uf) %>%
  summarize(pop_mean = mean(populacao)) %>%
  arrange(pop_mean)

medias_pop_estados %>% kable(caption = "Média da população nos municípios por estado")
```

Tocantins tem menor média, enquanto Rio de Janeiro possui maior média.

8

Apresente seus resultados em um arquivo PDF compilado diretamente a partir do R. Garanta que seu arquivo esteja limpo, contendo as respostas, os gráficos e as tabelas, mas não eventuais mensagens e erros. Para esta primeira lista, inclua no arquivo PDF os códigos utilizados (siga o exemplo apresentado na questão 1). Tenha em mente, no entanto, que para a maioria dos trabalhos acadêmicos, devemos omitir os códigos no arquivo PDF e fornecê-los separadamente para replicação.