

Lista 1

Rafaela Polachini

Instalando e carregando os pacotes

```
# options(repos = list(CRAN= "http://cran.rstudio.com/"))
# install.packages("devtools")
# devtools::install_github("tbrugz/ribge")
# install.packages("tidyverse")

library(devtools)
library(ribge)
library(dplyr)
library(tidyverse)
```

Importando a base de dados

```
pop_2020 <- populacao_municipios(2020)
```

Perguntas:

1. Qual é a unidade de análise desse banco de dados? Quantas observações estão listadas?
2. Quantas variáveis estão contidas nesse banco de dados?
3. Comente sobre outros aspectos da base de dados que considerar interessante, curioso ou que tenha gerado algum tipo de dúvida.

Respostas:

1. As unidades de análise são os municípios.
2. No banco estão contidas oito variáveis.
3. A base de dados é bem extensa, com 5570 observações e há mais variáveis numéricas do que caracteres. Fiquei em dúvida quanto às variáveis 'cod_munic6' e 'cod_municipio', as quais não soube identificar sobre o que se tratam.

Limpando a base de dados usando a função select()

```
pop_2020_novo <- pop_2020 %>%  
  select("uf", "nome_munic", "populacao")
```

Renomeando a variável "nome_munic"

```
pop_2020_renomeado <- pop_2020_novo %>% rename("municipio" = "nome_munic")
```

Alterando os nomes dos municípios para letra minúscula

```
pop_2020_alterado <- pop_2020_renomeado %>%  
  mutate(municipio = as.factor(municipio) %>%  
    tolower())
```

Analisando o Estado de São Paulo

Filtrando a base de dados

```
unique(pop_2020$uf)
```

```
[1] "RO" "AC" "AM" "RR" "PA" "AP" "TO" "MA" "PI" "CE" "RN" "PB" "PE" "AL" "SE"  
[16] "BA" "MG" "ES" "RJ" "SP" "PR" "SC" "RS" "MS" "MT" "GO" "DF"
```

```
pop_2020_sp <- pop_2020 %>%  
  filter(uf == "SP")
```

Perguntas:

1. Quantos municípios há no estado de São Paulo?

Resposta: Há 645 municípios no estado de São Paulo.

2. Qual o menor município do estado para o ano de 2020? Quantos habitantes ele tinha?

Resposta:

```
min(pop_2020_sp$populacao)
```

```
[1] 838
```

```
menor_munic <- pop_2020_sp %>% filter(populacao == 838)
```

O menor município do estado para o ano de 2020 foi Borá, com 838 habitantes.

3. Quantos municípios paulistas tem uma população maior que um milhão de habitantes? Liste os, mencionando também a população de cada um.

Resposta:

```
munic_pop_milhao <- pop_2020_sp %>%  
  filter(populacao > 1e6)
```

Há três municípios no estado de São Paulo com uma população maior que um milhão. São eles: Campinas, com 1.213.792 habitantes; Guarulhos, com 1.392.121; e São Paulo, com 12.325.232.

Estatística descritiva

Perguntas:

Para a variável “populacao”, calcule:

1. Média—use a função `mean()`.
2. Mediana—use a função `median()`.
3. Desvio-padrão— use a função `sd()`.
4. Variância— use a função `var()`.

Respostas:

```
est_descritiva <- pop_2020_sp %>% summarise(  
  media = mean(populacao),  
  mediana = median (populacao),  
  desvio_padrao = sd(populacao),  
  variancia = var(populacao))
```

A media da população é de 71.766; a mediana é de 14.141; o desvio padrão, 498.490; e, por fim, a variância é de 2.48e+11.

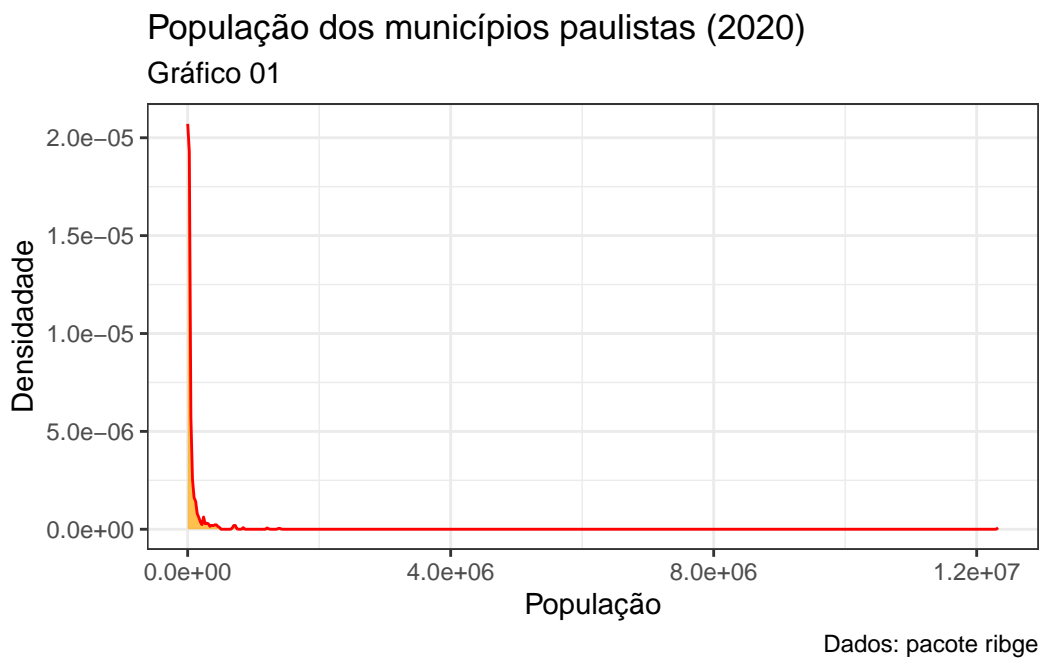
Trabalhando com gráficos

Perguntas:

1. Construa um gráfico de densidade (`geom_density()`) da variável população dos municípios paulistas. Extra: tente mudar a cor (argumento `colour =`), e preenchimento da curva (argumento `fill =`), e a transparência da curva (argumento `alpha =`) de seu gráfico.

Resposta:

```
pop_2020_sp %>% ggplot2::ggplot(aes(x = populacao)) +  
  geom_density(  
    alpha = 0.7,  
    colour = "red",  
    fill = "orange") +  
  ggplot2::labs(  
    x = "População",  
    y = "Densidade",  
    title = "População dos municípios paulistas (2020)",  
    subtitle = "Gráfico 01",  
    caption = "Dados: pacote ribge") +  
  ggplot2::theme_bw()
```



2. Comente o gráfico gerado. O que você observa?

Resposta: O gráfico apresenta um pico à esquerda, e uma mínima variação à direita. Isso indica uma grande quantidade de municípios com menos de 1 milhão de habitantes, e uma baixa quantidade daqueles com mais de um milhão.

3. A partir da observação do gráfico, qual parece ser a medida mais adequada de tendência central da população: a média ou a mediana? Justifique sua resposta

Resposta: Como há uma grande discrepância nas quantidades de municípios com menos 1 milhão de habitantes e aqueles com mais, a medida mais adequada de tendência central da população é a mediana, uma vez que a partir dela será possível ter uma melhor noção dessa assimetria entre os municípios paulistas. Como já feito visto antes, a mediana da população é de 14.141, enquanto a média apresenta um valor de 71.766, indicando um valor muito maior do que os municípios têm de fato.

Gráfico de densidade de pequenos municípios

Pergunta:

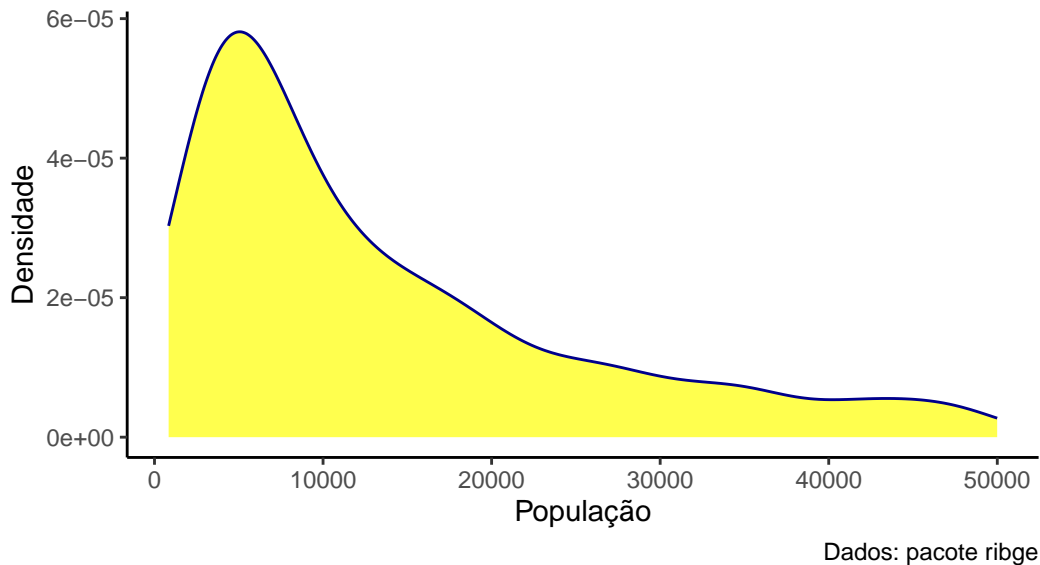
1. Crie um gráfico de densidade apenas para os municípios com menos de 50.000 habitantes. Utilize novamente a função `filter()`, e perceba que você não precisa criar um novo banco de dados, pois pode conectar todos os comandos através do “pipe” `%>%`.

Resposta:

```
pop_2020_sp %>% filter(
  populacao < 50000) %>%
  ggplot2::ggplot(aes(x = populacao)) +
  geom_density(
    alpha = 0.7,
    colour = "darkblue",
    fill = "yellow") +
  labs(
    x = "População",
    y = "Densidade",
    title = "Municípios paulistas com menos de 50 mil habitantes",
    subtitle = "Gráfico 02",
    caption = "Dados: pacote ribge") +
  theme_classic()
```

Municípios paulistas com menos de 50 mil habitantes

Gráfico 02



2. Quantos são os municípios paulistas com menos de 50.000 habitantes? Qual a porcentagem dessas cidades em relação ao conjunto de municípios do estado? Tente fazer a conta pelo R.

Quantidade de municípios com menos de 50.000 habitantes.

```
pop_menos_50 <- pop_2020_sp %>%  
  filter (populacao < 50000)
```

Cálculo da porcentagem

```
porcentagem_sp <- nrow(pop_menos_50)/nrow(pop_2020_sp)*100
```

Resposta: Há 504 municípios paulistas com menos de 50.000 habitantes. A porcentagem dessas cidades em relação ao conjunto de municípios do estado é de 78.13953.

3. Em comparação com o gráfico da questão anterior, o que você observa?

Resposta: O gráfico em questão apresenta uma variação maior do que o gráfico gerado anteriormente. Analisando sua distribuição, é possível afirmar que há mais municípios com

menos de 10.000 habitantes. Os municípios próximos ao número de 50.000 habitantes são poucos comparados aos com menos de 10.000.

Trabalhando com escala logarítmica

Perguntas:

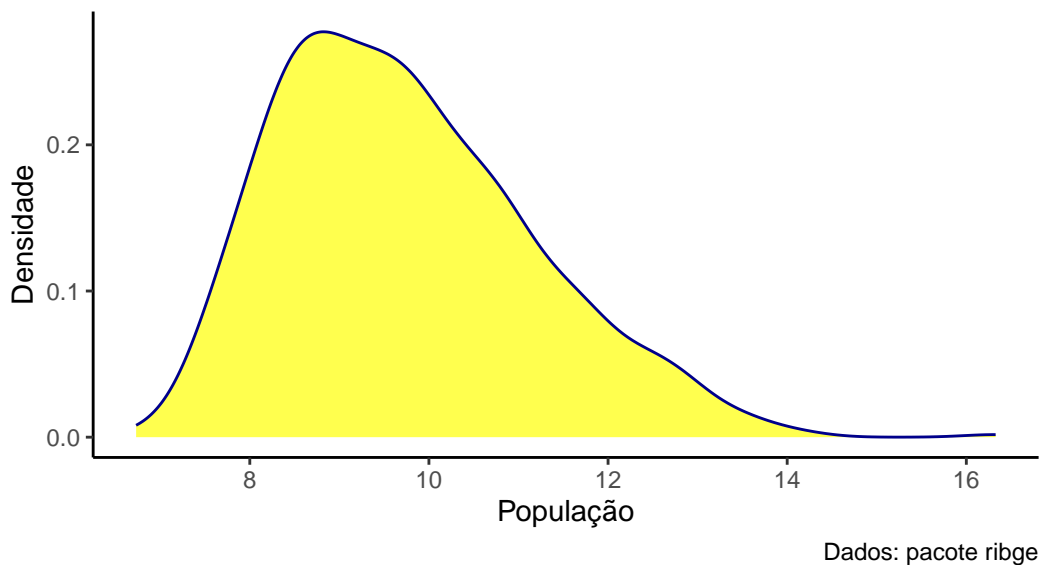
1. Crie novamente um gráfico de densidade para todos os municípios de São Paulo, mas agora utilize a variável de população em logaritmo. Um caminho possível é criar uma nova variável com a função `mutate()` e usar a função `log()` para transformar a variável em logaritmo natural.

Resposta:

```
pop_2020_sp %>% mutate(  
  pop_log = log(populacao)) %>%  
  ggplot2::ggplot(aes(x = pop_log)) +  
  ggplot2::geom_density(  
    alpha = 0.7,  
    colour = "darkblue",  
    fill = "yellow",) +  
  labs(  
    x = "População",  
    y = "Densidade",  
    title = "População nos municípios paulistas (população em logaritmo)",  
    subtitle = "Gráfico 03",  
    caption = "Dados: pacote ribge") +  
  theme_classic()
```


População nos municípios paulistas (população em logarítimo)

Gráfico 03



2. Comente o gráfico de densidade, comparando com o gráfico do exercício 4.1.

Resposta: O gráfico gerado, comparado com o gráfico 4.1, apresenta uma maior distribuição, não havendo uma assimetria tão marcante, como ocorreu no anterior. Além disso, os números do eixo x (população) não chegam na casa dos milhares, como ocorreu no gráfico 4.1.

Média das populações por estado

Pergunta:

1. Assim como foi feito para os municípios do estado de São Paulo, calcule a média da população para cada um dos estados brasileiros. Utilize as funções `group_by()` (pacote `dplyr`) para agrupar os estados (variável “uf”), além da função `summarise()` usada no exercício 3. Atribua esses comandos a um novo objeto.

Resposta:

```
media_pop_uf <- pop_2020 %>%  
  group_by(uf) %>%  
  summarise(  
    media = mean(populacao))
```

```
head(media_pop_uf)
```

```
# A tibble: 6 x 2
  uf      media
  <chr>  <dbl>
1 AC    40658.
2 AL    32858.
3 AM    67866.
4 AP    53861.
5 BA    35805.
6 CE    49930.
```

2. Perceba que a tabela gerada não segue uma ordem crescente ou decrescente. Acrescente a função `arrange()` (pacote `dplyr`). Quais dos estados possuem maior e menor população média por município?

Resposta:

Menor média

```
media_pop_uf %>% dplyr::arrange(media) %>% head(5)
```

```
# A tibble: 5 x 2
  uf      media
  <chr>  <dbl>
1 TO    11441.
2 PI    14649.
3 PB    18113.
4 RN    21163.
5 RS    22984.
```

Maior média

```
media_pop_uf %>% dplyr::arrange(desc(media)) %>% head(5)
```

```
# A tibble: 5 x 2
  uf      media
  <chr>  <dbl>
```

1	DF	3055149
2	RJ	188763.
3	SP	71766.
4	AM	67866.
5	PA	60352.

Conforme a primeira tabela gerada, o estado com menor população média é o Tocantins, com 11.441 de média; já conforme a segunda tabela, o estado com maior população média é o Distrito Federal, com média de 305.5149.