

Análise Exploratória SAEB 2019

Tailine J. S. Nonato

2024-06-24

Análise Exploratória SAEB 2019

Carregamento de pacotes e dados

```
if (!require(pacman)) install.packages("pacman")
```

Loading required package: pacman

```
pacman::p_load(vroom, tidyverse, knitr, openxlsx, kableExtra, lme4)  
setwd("C:/Users/User/Documents/GitHub/gradest-1/TCC/rel parcial")
```

```
df_aluno <- read.csv("TS_ALUNO_9EF_2019_co.csv")  
df_escola <- read.csv("TS_ESCOLA_2019_co.csv")
```

```
df <- left_join(df_aluno, df_escola, by = c("ID_ESCOLA", "ID_REGIAO", "ID_SAEB", "ID_UF", "ID_MUNICIPIO"))  
df <- df %>% mutate(ID_ESCOLA = as.factor(ID_ESCOLA))  
var_df <- names(df)  
dim(df)
```

```
[1] 203694    260
```

Filtragem dos dados para região Centro-Oeste

```
df_aluno <- read.csv("TS_ALUNO_9EF_2019.csv", sep = ";")
df_escola <- read.csv("TS_ESCOLA_2019.csv", sep = ";")

var_al <- names(df_aluno)
var_es <- names(df_escola)

df_aluno <- df_aluno %>% filter(ID_REGIAO == 5)
df_escola <- df_escola %>% filter(ID_REGIAO == 5)

write.csv(df_aluno, file = "TS_ALUNO_9EF_2019_co.csv")
write.csv(df_escola, file = "TS_ESCOLA_2019_co.csv")
```

Análise descritiva dos dados

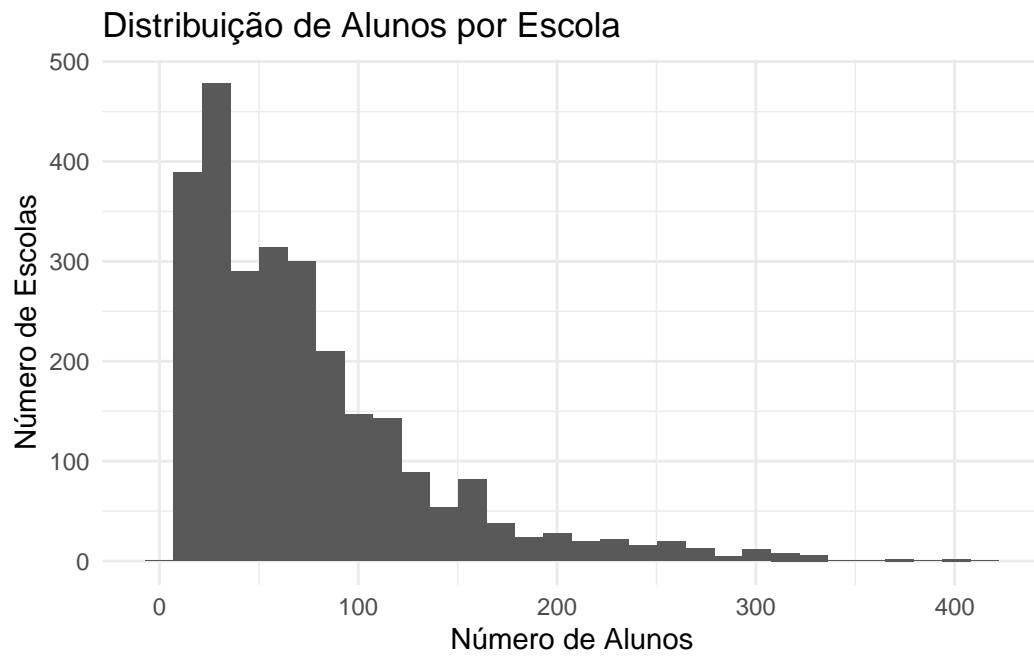
Quantidade de Escolas

```
df_aluno %>%
  pull(ID_ESCOLA) %>%
  n_distinct
```

```
[1] 2717
```

Quantidade de Alunos por Escola

```
df_aluno %>%
  group_by(ID_ESCOLA) %>%
  summarise(n = n()) %>%
  ggplot(aes(x = n)) +
  geom_histogram(bins = 30) +
  labs(title = "Distribuição de Alunos por Escola",
       x = "Número de Alunos",
       y = "Número de Escolas") +
  theme_minimal()
```



Nível Socioecononômico