



# Pratica 3

## Roteiro 3 - Análise Exploratória II

<https://luisfelipebr.github.io/mti2020/roteiros/roteiro3.html>

Codigo e dados disponivel em →

[https://github.com/wesleyclzns/analise\\_dados\\_planejamento\\_territorial\\_UFABC](https://github.com/wesleyclzns/analise_dados_planejamento_territorial_UFABC)

## Importando e elendo os dados

```
getwd()
"D:/CM/ADPT/scripts"

agua1 <- read.csv2("https://raw.githubusercontent.com/luisfelipebr/mti/master/dados/agua1.csv", encoding="UTF-8")
head(agua1)

rede1 <- read.csv2("https://raw.githubusercontent.com/luisfelipebr/mti/master/dados/rede1.csv", encoding="UTF-8")
head(rede1)

names(agua1)
names(rede1)
```

## Instalações e carregamento de pacs

```
##INTALAÇÕES
#install.packages("tidyverse")

#Carregar Pacotes
library(tidyverse)
```

## Filtros e transformações

```
#Testando Filtros
agua1[which(agua1$UF == "SP" & agua1$IDH > 0.85),c("NOME_MUN", "IDH")]

agua1 %>%
  filter(UF == "SP" & IDH > 0.85) %>%
  select(NOME_MUN, IDH)

#Transformações
agua1$CONSUMO1 <- agua1$AG020 * 1000 / agua1$GE012
agua1$CONSUMO2 <- agua1$AG020 * 1000 / agua1$AG001
```

## Criando novos DataFrames

```
#Criamos o df agua2
agua2 <- agua1 %>%
  mutate(CONSUMO1 = AG020 * 1000 / GE012,
         CONSUMO2 = AG020 * 1000 / AG001)
head(agua2)

#Criando df Agua Rede 1 usando o Join
agua_rede1 <- rede1 %>%
  mutate(PROPREDE = REDE/DOMICIL) %>%
  select(-c(UF, REGIAO)) %>%
  full_join(agua2, by = "ID_IBGE")

head(agua_rede1)
```

## Agrupando e resumindo

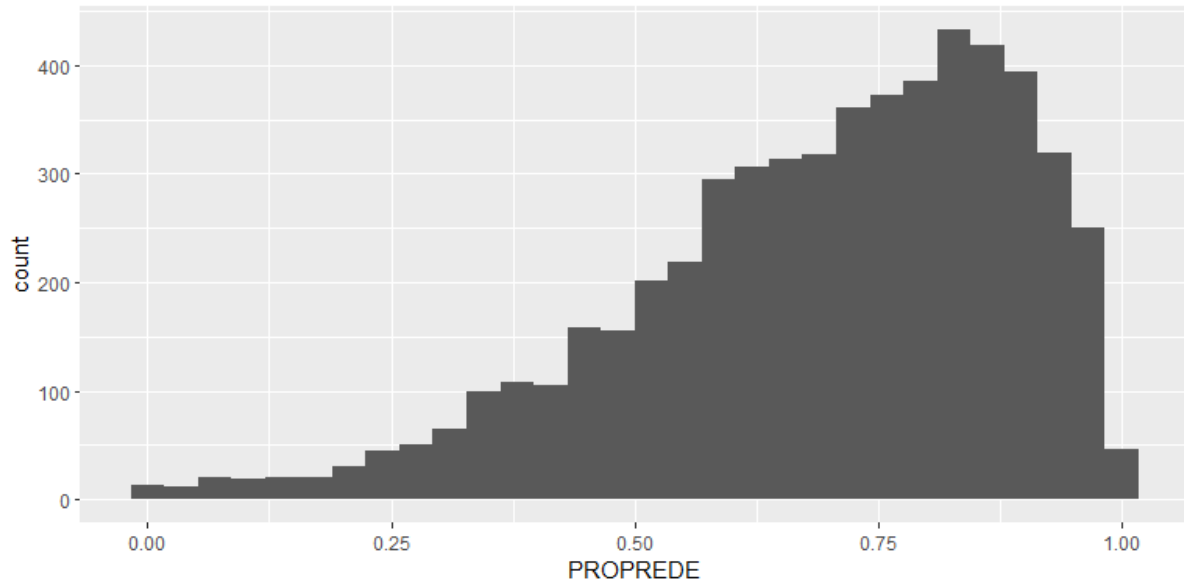
```
#Agrupando e resumindo
tabela_PROPREDE <- agua_rede1 %>%
  drop_na(PROPREDE) %>% # Limpa os NAN
  group_by(REGIAO) %>%
  summarize(n_obs = n(),
           media = mean(PROPREDE),
           desvio_padrao = sd(PROPREDE)) %>%
  mutate(erro = 1.96*desvio_padrao/sqrt(n_obs),
         limite_superior = media + erro,
         limite_inferior = media - erro)

tabela_PROPREDE
```

# Graficos

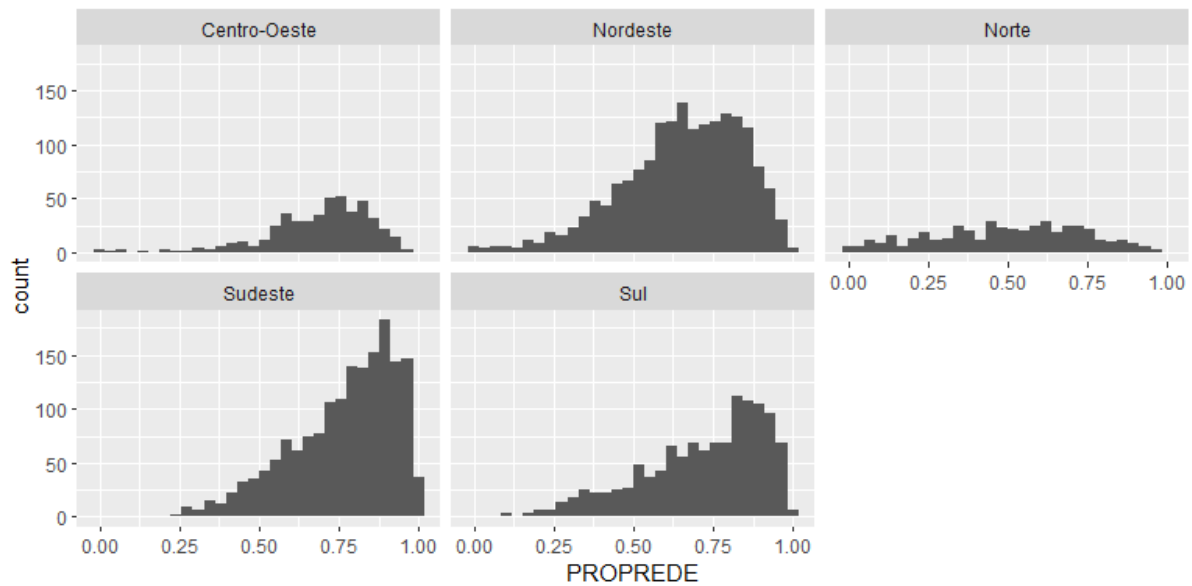
## Histograma geral

```
ggplot(data = agua_rede1, aes(x = PROPREDE)) +  
  geom_histogram()
```



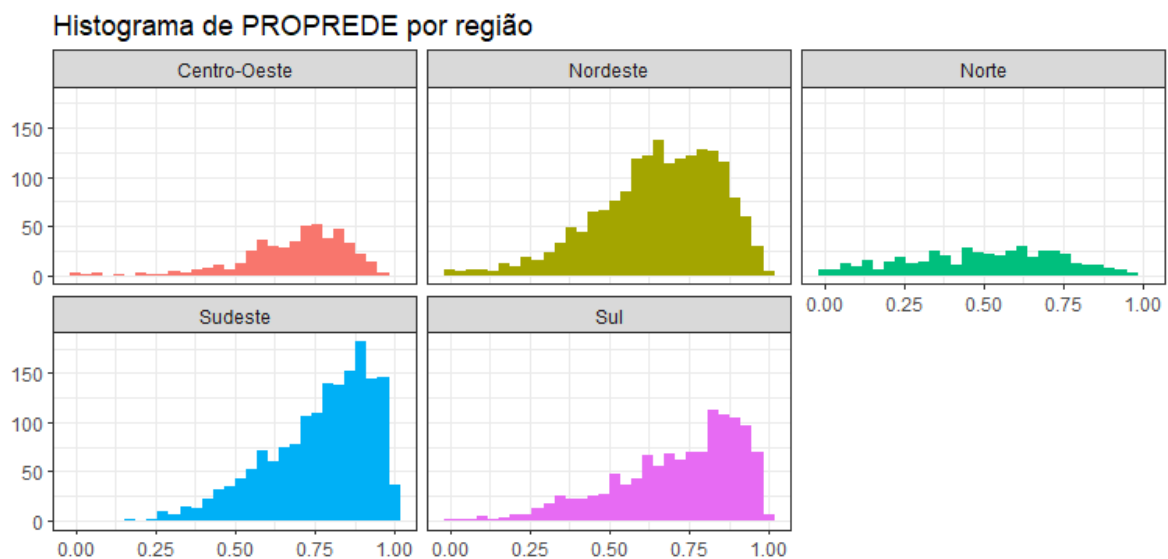
## Histograma regiões

```
ggplot(data = agua_rede1, aes(x = PROPREDE)) +  
  geom_histogram() +  
  facet_wrap(~REGIAO)
```



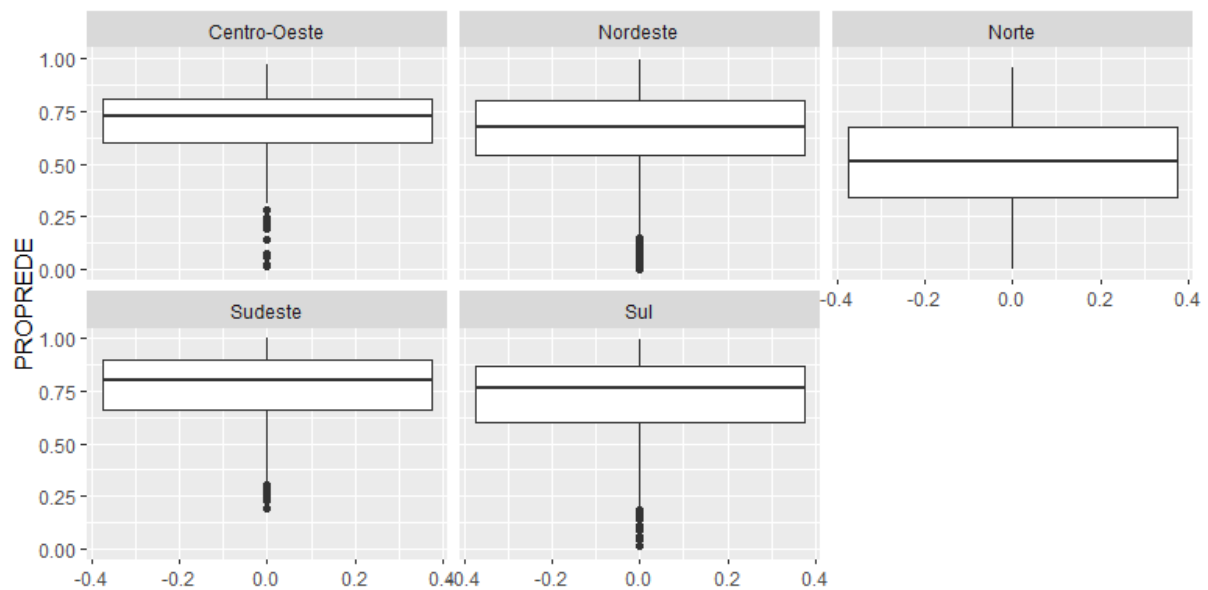
## Histograma regiões coloridos

```
#Deixando o graph anterior mais bonito
ggplot(data = agua_rede1, aes(x = PROPDE, fill = REGIAO)) +
  geom_histogram() +
  facet_wrap(~REGIAO) +
  ggtitle("Histograma de PROPDE por região") +
  xlab("") +
  ylab("") +
  theme_bw() +
  theme(legend.position = "none")
```



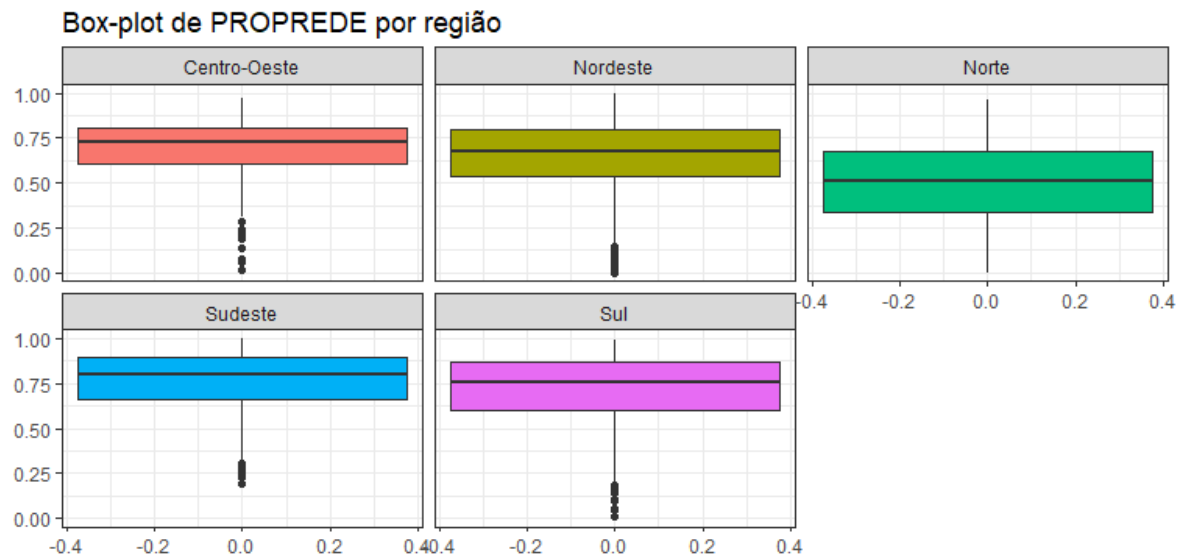
## Boxplot das regiões

```
#Boxplot do GG
ggplot(data = agua_rede1, aes(y = PROPREDE)) +
  geom_boxplot() +
  facet_wrap(~REGIAO)
```



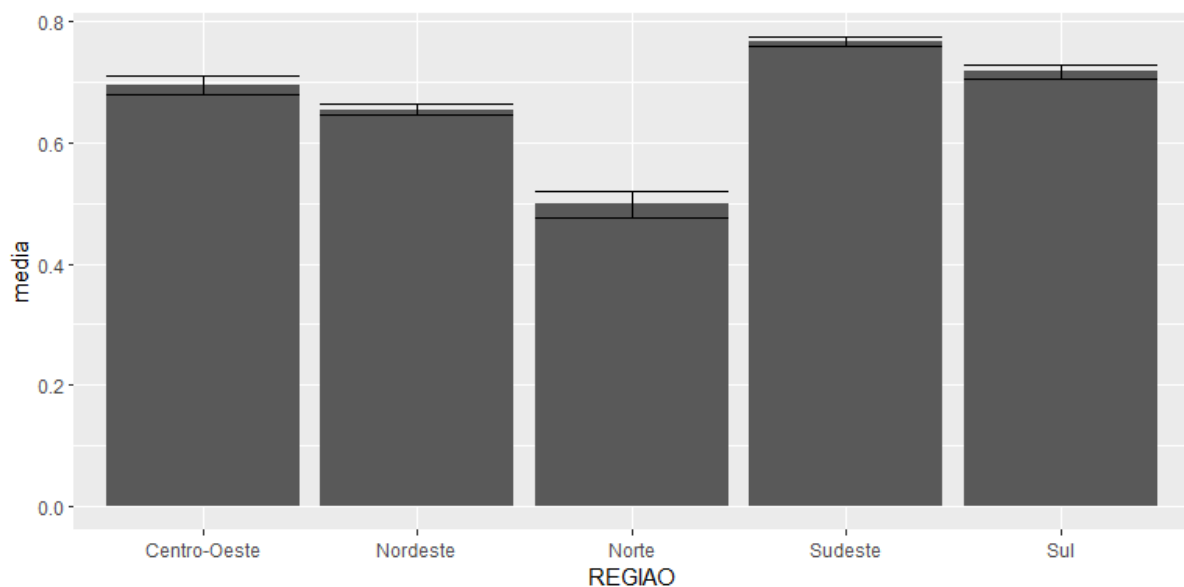
## Boxplot das regiões colorido

```
#boxplot bonitinho
ggplot(data = agua_rede1, aes(y = PROPREDE, fill = REGIAO)) +
  geom_boxplot() +
  facet_wrap(~REGIAO) +
  ggtitle("Box-plot de PROPREDE por região") +
  xlab("") +
  ylab("") +
  theme_bw() +
  theme(legend.position = "none")
```



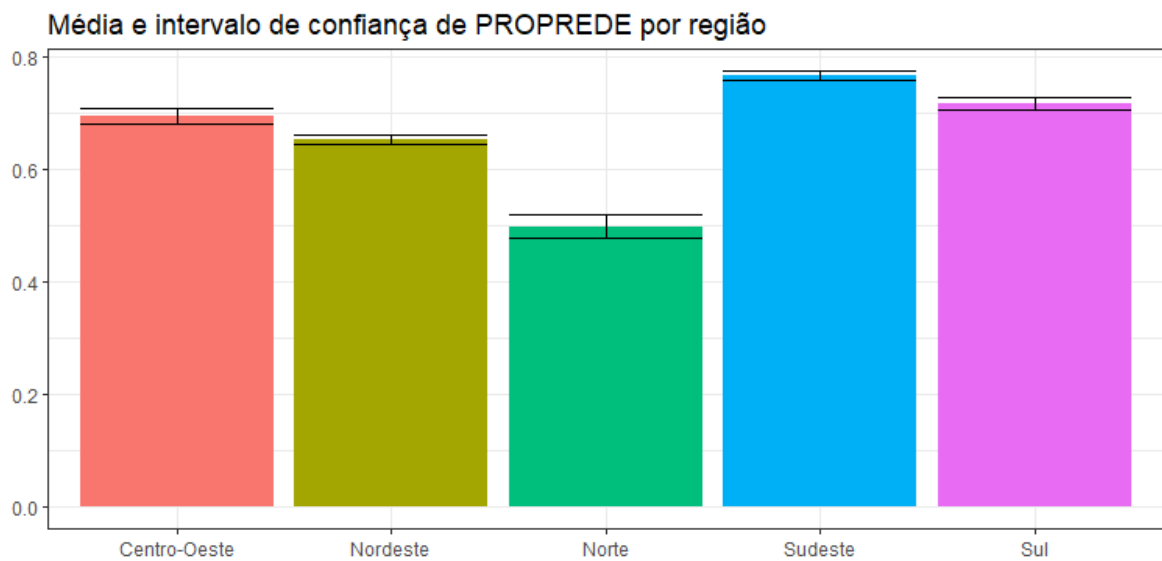
## Media e intervalo de confiança

```
#Media e intervalo de conf
ggplot(data = tabela_PROPREDE, aes(x = REGIAO, y = media)) +
  geom_col() +
  geom_errorbar(aes(ymin = limite_inferior, ymax = limite_superior))
```



## Media e intervalo de confiança colorido

```
#media e intervalo bonitinho
ggplot(data = tabela_PROPREDE, aes(x = REGIAO, y = media, fill=REGIAO)) +
  geom_col() +
  geom_errorbar(aes(ymin = limite_inferior, ymax = limite_superior)) +
  ggtitle("Média e intervalo de confiança de PROPREDER por região") +
  xlab("") +
  ylab("") +
  theme_bw() +
  theme(legend.position = "none")
```



## Exportação do df criado

```
write.csv2(agua_rede1, "D:/CM/ADPT/dados/praticas/agua_rede1.csv", row.names = FALSE)
```