

Análise Socioeconômica dos Municípios Brasileiros

Andres Kindel Barbosa

Luiz Adriano Augusto dos Santos

Gustavo Beretta Gonçalves

→ Introdução e objetivos:

A base de dados escolhida contém 81 informações diferentes sobre todas as cidades que compõem o Brasil. Os dados são de alta confiabilidade, já que foram obtidos de organizações como IBGE, Banco Central do Brasil, Nações Unidas, ANATEL, DENATRAN, entre outras. A data de referência dos dados varia para cada um, mas a maioria é posterior a 2016, sendo que os mais antigos são do censo de 2010. Todas as análises foram realizadas com auxílio dos softwares estatísticos R e RStudio.

O objetivo deste trabalho é analisar diferentes dados socioeconômicos dos municípios do Brasil e relacioná-los entre si para que seja possível tirar *insights* sobre os assuntos analisados e ter uma melhor noção da situação do nosso país.

→ Materiais e métodos:

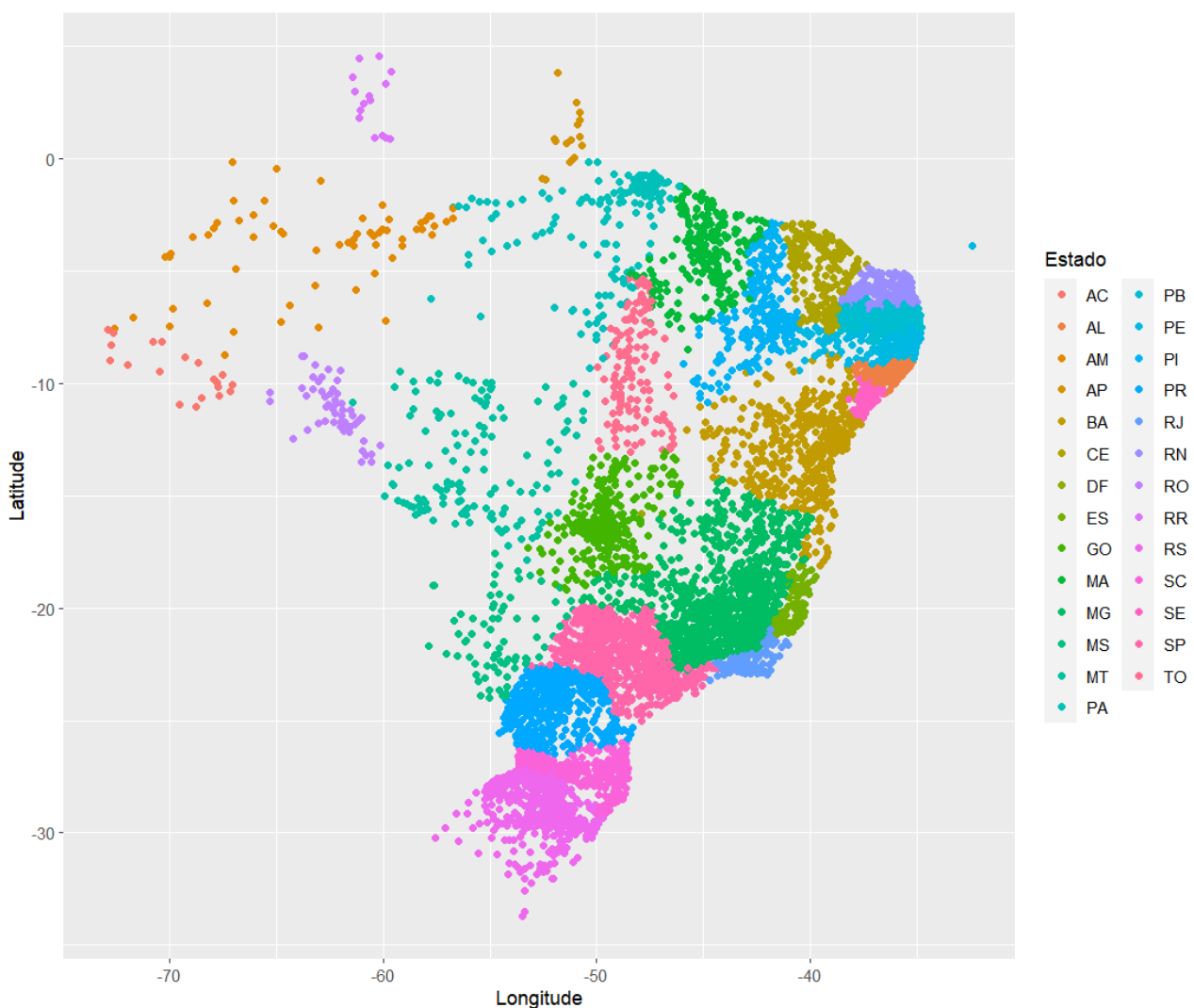
As variáveis utilizadas para a análise foram as seguintes:

Nome	Descrição	Data	Tipo
STATE	Estado a que pertence	Atual	Qualitativa nominal
ESTIMATED_POP	População estimada	07/2018	Quantitativa discreta
TAXES	Impostos arrecadados	2016	Quantitativa discreta
MUN_EXPENDIT	Gastos municipais	2016	Quantitativa discreta
IDHM	IDH	2010	Quantitativa discreta
LONG	Longitude	2010	Quantitativa discreta
LAT	Latitude	2010	Quantitativa discreta

- A variável STATE foi escolhida para possibilitar a estratificação dos dados por estados, permitindo a análise destes também, além dos municípios.
- A variável ESTIMATED_POP foi escolhida para possibilitar a análise de acordo com a população dos municípios.
- As variáveis TAXES e MUN_EXPENDIT foram escolhidas para possibilitar análises econômicas em relação aos municípios.
- A variável IDHM foi escolhida por ser o índice de maior importância na representação da qualidade de vida de uma população.
- As variáveis LONG e LAT foram escolhidas pois possibilitam o geoposicionamento dos municípios em um mapa.

→ Resultados e discussões:

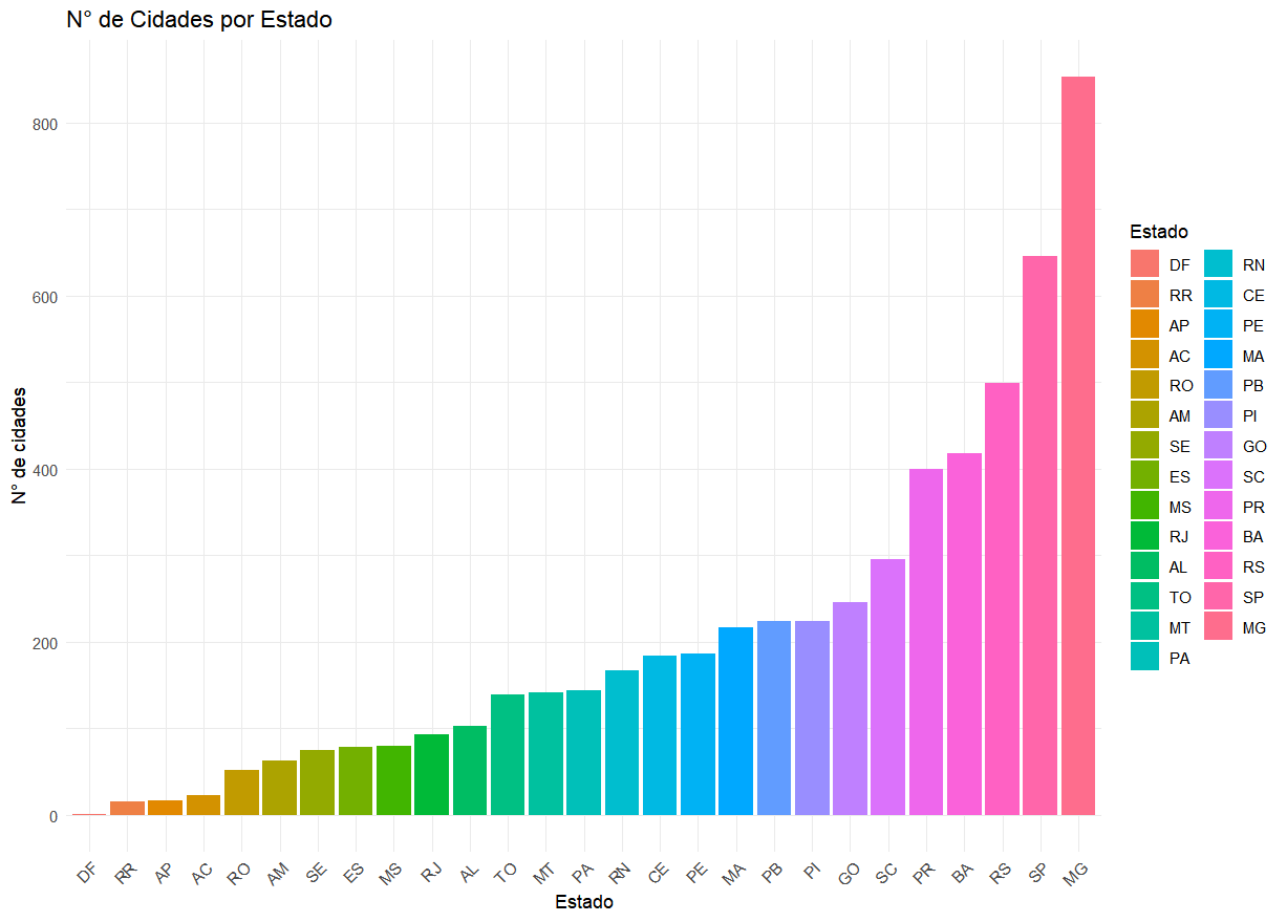
- Distribuição geográfica dos municípios, separados por seus respectivos estados:



Através deste gráfico de dispersão dos pontos, é possível verificar muitos espaços vazios, que representam municípios com grandes áreas. Estes espaços estão presentes, principalmente, nas regiões Centro-Oeste e Norte.

Enquanto isso, nas demais regiões, há uma grande concentração de municípios, causada pela maior quantidade dos mesmos, o que pode indicar menores populações para cada um.

- Número de municípios por estado:



Após analisar esse gráfico de barras, é possível confirmar a disparidade no número de municípios em cada estado. O estado com menos municípios (com exceção do Distrito Federal) é Roraima, com apenas 15, enquanto Minas Gerais, o estado com mais municípios, contém 853.

Apesar dessa confirmação, não há como afirmar se há ou onde há algum problema na divisão de municípios. Para isso, é preciso analisar a população destes, já que as prefeituras devem se manter através dos impostos de seus residentes e, se houver poucos contribuintes, a arrecadação será baixa.

- Medidas de posição da população dos municípios:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
786	5454	11591	37439	25305	12176866

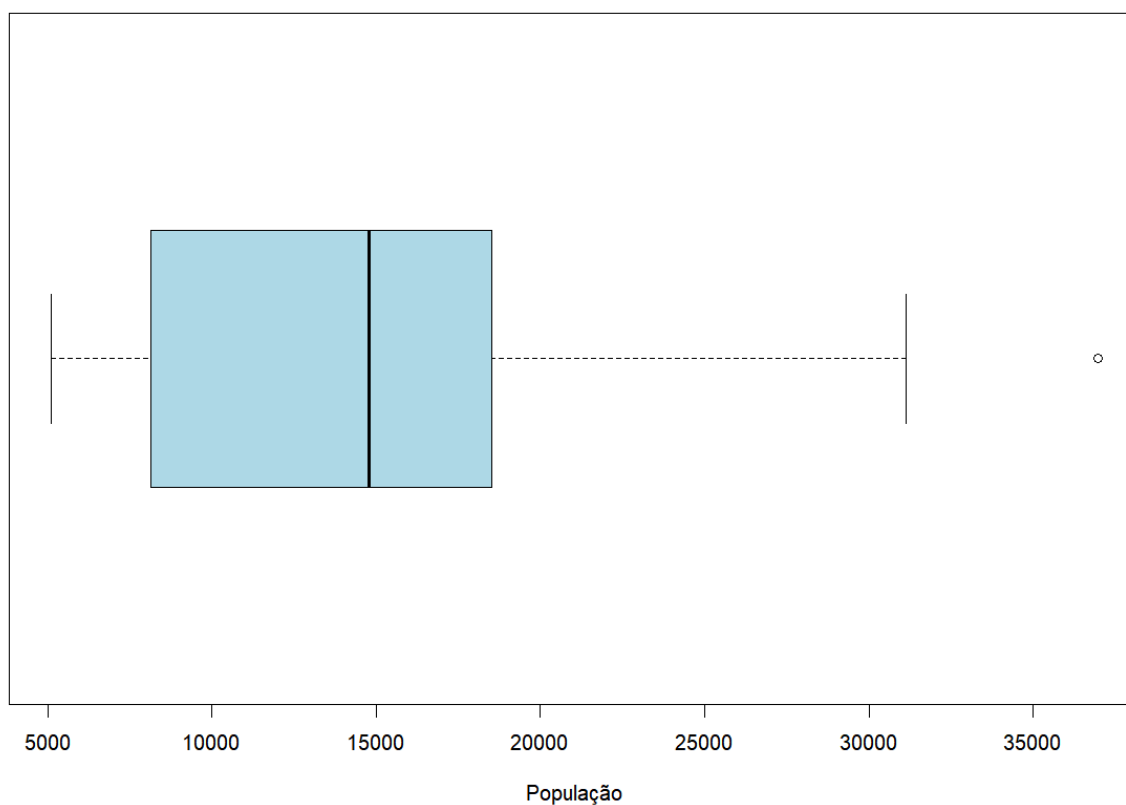
Com estas medidas de posição, pode-se confirmar que há municípios com baixíssima população.

Algumas informações obtidas:

- ❖ O município com menor população no Brasil tem apenas 786 habitantes.
- ❖ 25% dos municípios tem menos de 5454 habitantes.
- ❖ Metade dos municípios tem menos de 11591 habitantes.
- ❖ A média da população dos municípios brasileiros é mais de 3 vezes maior do que a mediana, o que indica uma grande concentração de habitantes em algumas cidades, provavelmente as capitais.

- Mediana da população dos municípios por estado:

Mediana da população das cidades por estado brasileiro



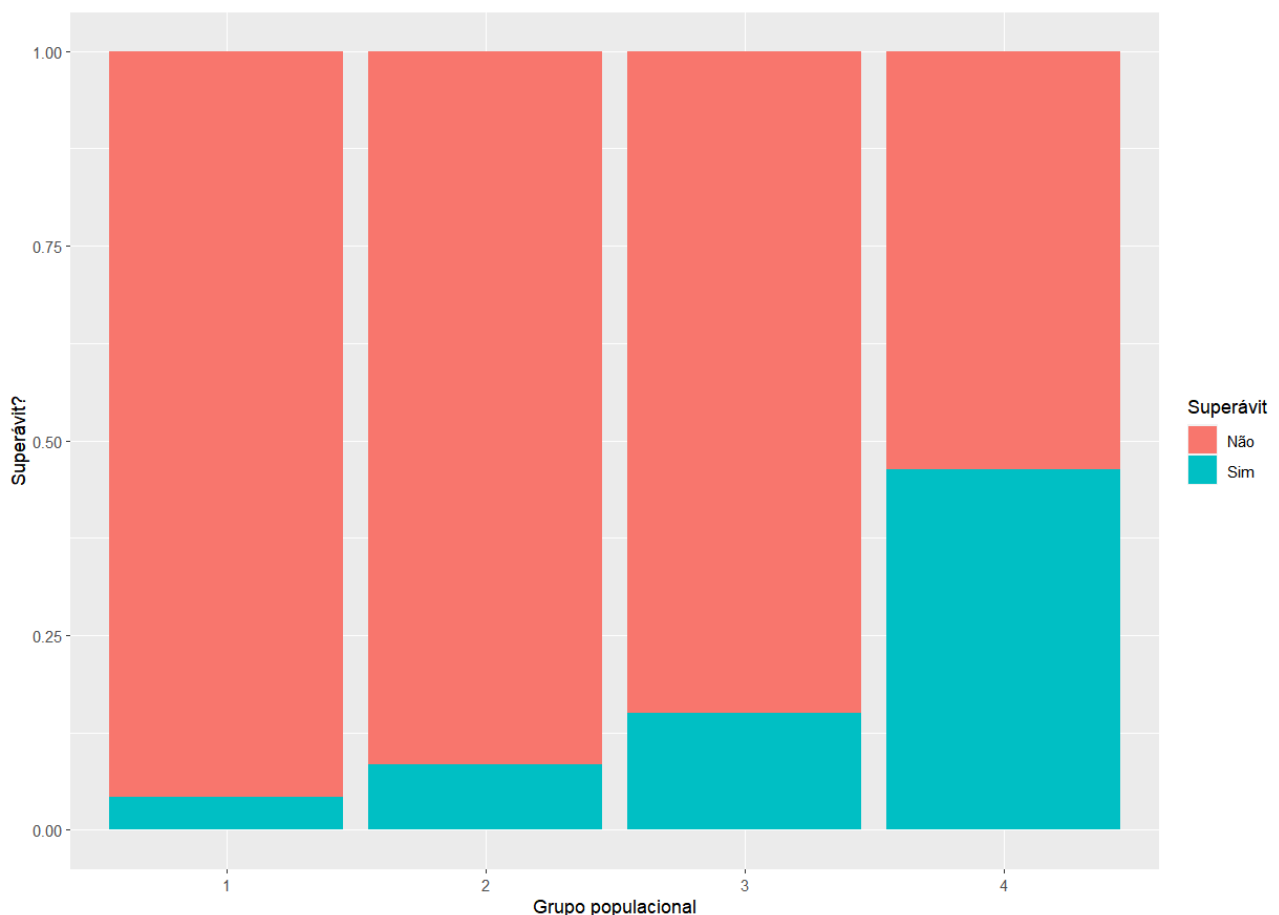
O gráfico de caixa acima mostra que há uma disparidade na mediana da população dos municípios por estado. Enquanto Tocantins tem uma mediana de 5094 habitantes por município, a mesma métrica para o Rio de Janeiro aponta um valor de 36985 habitantes. É importante ressaltar que o Distrito Federal não está contido nesta análise, já que possui apenas um município e atrapalhou a análise quando considerado.

Esses dados que confirmam que há municípios com populações baixíssimas trazem um questionamento à tona: esses municípios conseguem se manter por conta própria ou demandam ajuda externa?

- Análise do fluxo de caixa dos municípios, divididos por quartis de população:

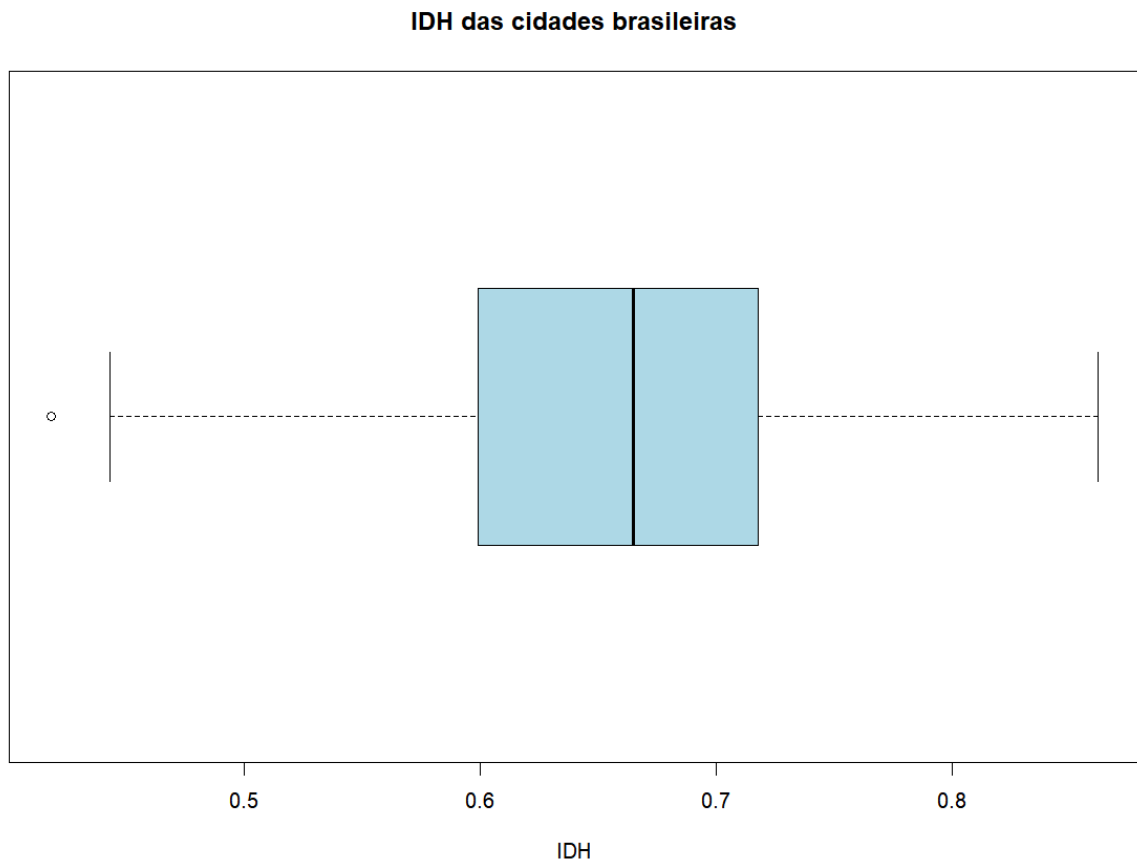
Para realizar essa análise, foi obtida a diferença entre os impostos arrecadados e o gasto de cada município. Caso essa diferença seja positiva, o município é superavitário e consegue se manter com recursos próprios. Caso negativo, o município demanda ajuda externa, normalmente da União.

O Grupo 1 contém municípios com população até 5.354, o Grupo 2 de 5.355 a 11.640, o Grupo 3 de 11.641 a 25.802 e o Grupo 4 acima de 25.803 habitantes.



Constata-se claramente uma correlação positiva entre a população de um município e a capacidade do mesmo ser superavitário. Ademais, a própria lógica econômica também corrobora com essa análise.

- População e IDH:



Analisando o gráfico de caixa exibido por último, é perceptível uma grande desigualdade na qualidade de vida entre os municípios brasileiros. Apesar disso, a maior parte deles possuem IDH médio (0,555 a 0,699) ou alto (0,700 a 0,799).

Entretanto, esse dado não basta para concluir que a população brasileira no geral vive em boas condições, já que vimos anteriormente que há uma grande concentração de população em determinados municípios. Para fazer essa análise, precisa-se classificar os municípios de acordo com seu IDH e, depois, soma-se a população dos municípios em cada classificação.

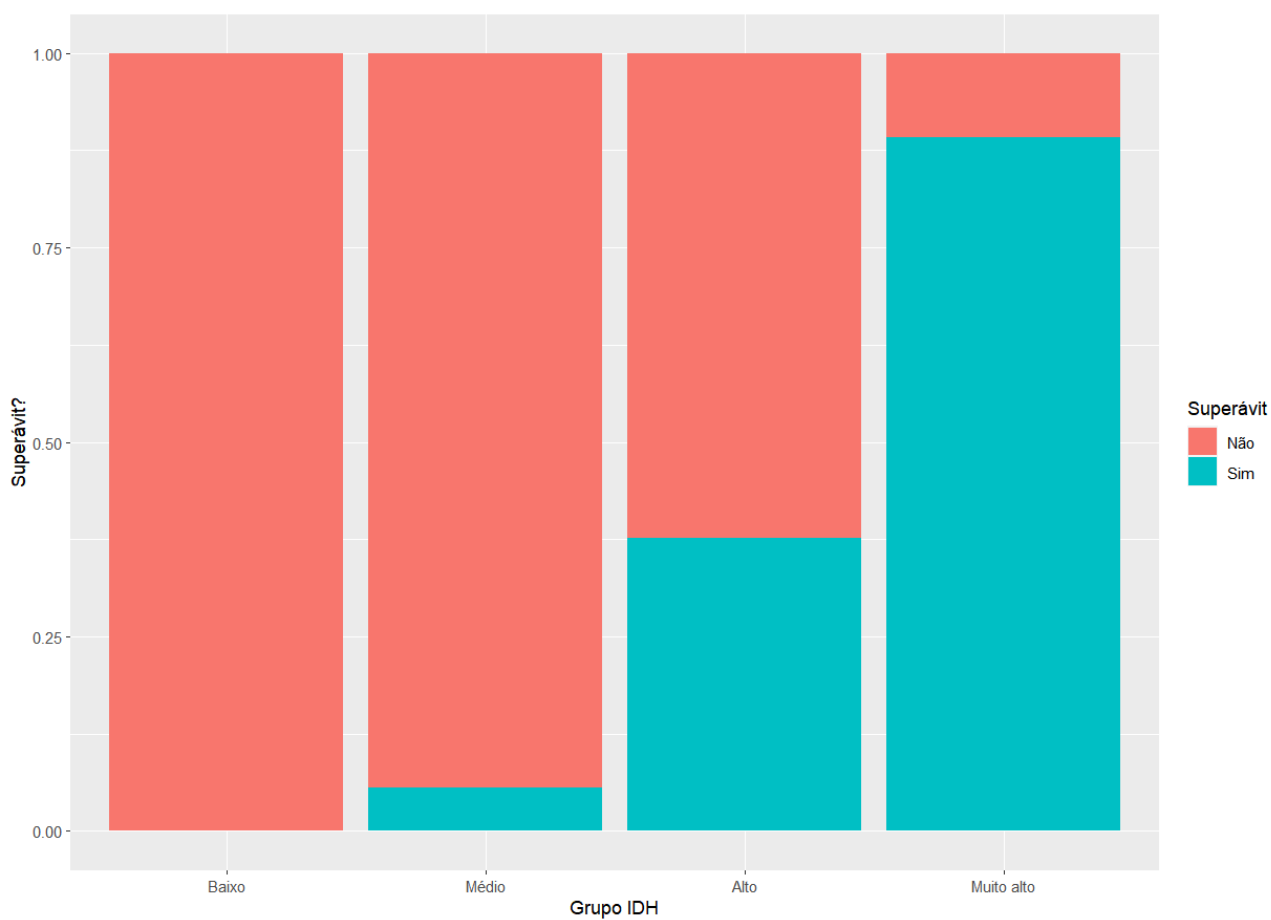
Esse método resulta nos seguintes dados tabelados:

Classificação do IDH	População
Baixo	7.151.491
Médio	60.148.575
Alto	108.374.954
Muito Alto	32.916.621

Constata-se, então, que a maior parte dos brasileiros vivem em cidades com alto IDH.

- Superávit e IDH:

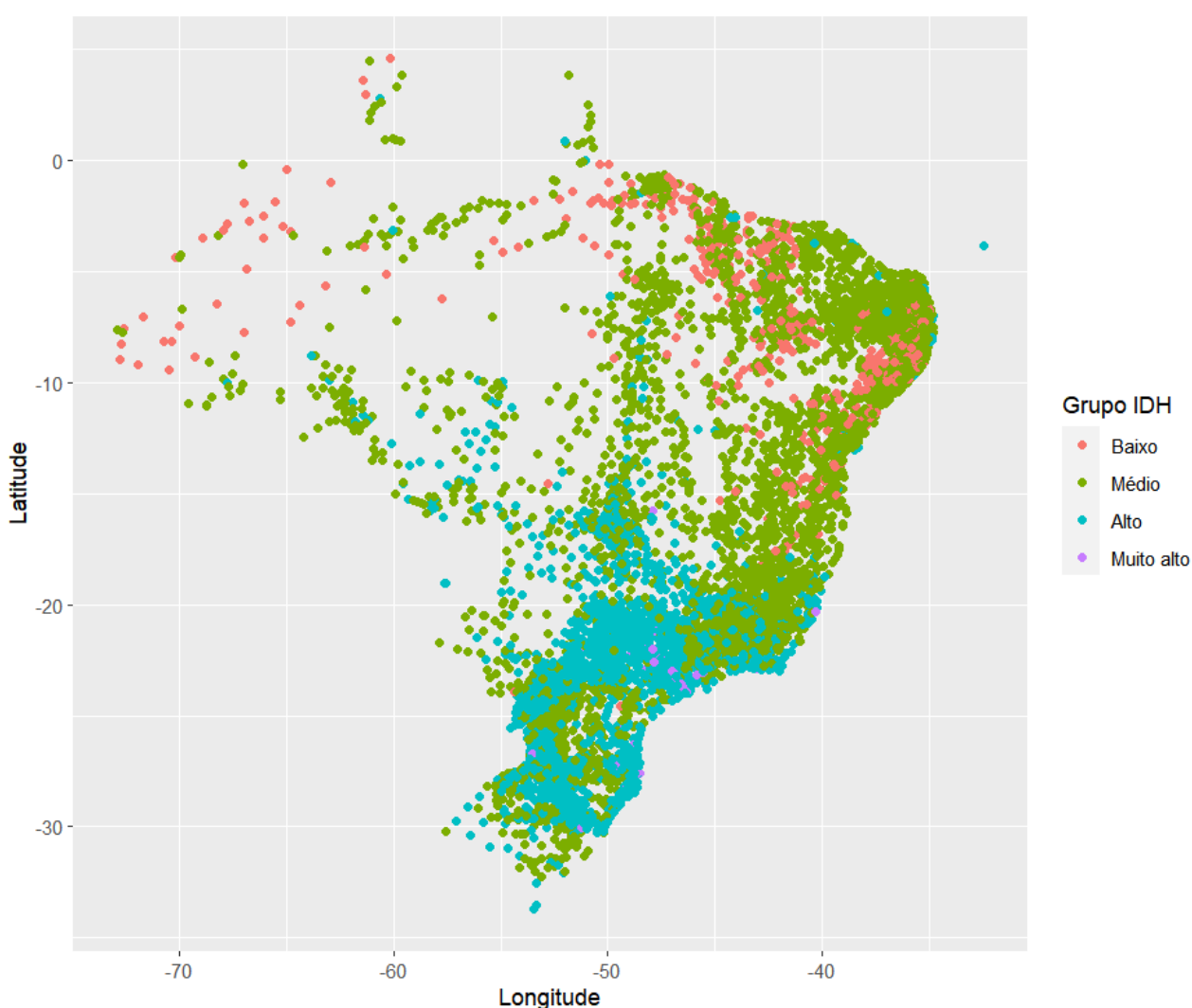
Voltando à análise do superávit dos municípios, podemos verificar se há alguma tendência no IDH dos municípios superavitários.



Pelo gráfico acima, podemos verificar uma forte correlação entre IDH e Superávit dos municípios. É possível argumentar que municípios superavitários têm tendência a terem maior IDH, mas é necessário lembrar que esses não são os únicos fatores em jogo, e análises socioeconômicas são muito mais profundas do que pode-se fazer neste trabalho.

- Distribuição geográfica dos municípios, identificando a classificação do IDH:

Com os dados da classificação do IDH de cada município, pode-se, novamente, distribuir os municípios no plano de acordo com suas coordenadas e analisar se há tendência de algum local ser mais desenvolvido do que outro.



Fica evidente que as regiões Sul e Sudeste concentram a grande maioria dos municípios de alto e muito alto IDH, enquanto as regiões Norte e Nordeste concentram os municípios de baixo IDH. Novamente, é muito complicado fazer qualquer inferência sobre os motivos disso, já que são diversos fatores que resultam nesses dados.

→ Considerações finais:

Com base nos dados confiáveis de diversas fontes, como IBGE, Banco Central do Brasil, Nações Unidas, ANATEL, DENATRAN e outros, presentes na base de dados e usando os softwares estatísticos R e RStudio, este estudo analisou informações socioeconômicas de cidades do Brasil. Foram examinados 7 indicadores diferentes da base para entender melhor a situação do país.

Os resultados destacaram as diferenças entre as cidades do Brasil, revelando tendências e padrões interessantes. Isso pode ser útil para tomar decisões em áreas como políticas públicas e planejamento regional, visando entender melhor a situação atual do Brasil e melhorar a qualidade de vida em todo o país.

Ao relacionar os dados, identificamos oportunidades e desafios em áreas como educação, saúde, economia e infraestrutura. Esses insights são cruciais para o desenvolvimento do país.

Em resumo, a análise feita forneceu informações valiosas sobre o Brasil e serve como base para futuras pesquisas e ações voltadas para o progresso e o bem-estar de todos os brasileiros. O uso de dados confiáveis e ferramentas analíticas é fundamental para tomar decisões informadas e construir um futuro melhor para o país.

→ Referências bibliográficas

PARADA, CRISTIANA. Brazilian Cities. 2022. Disponível em: <<https://www.kaggle.com/datasets/crisparada/brazilian-cities>>. Acesso em: 24, Setembro, 2023

trabalho.R

Luiz Adriano

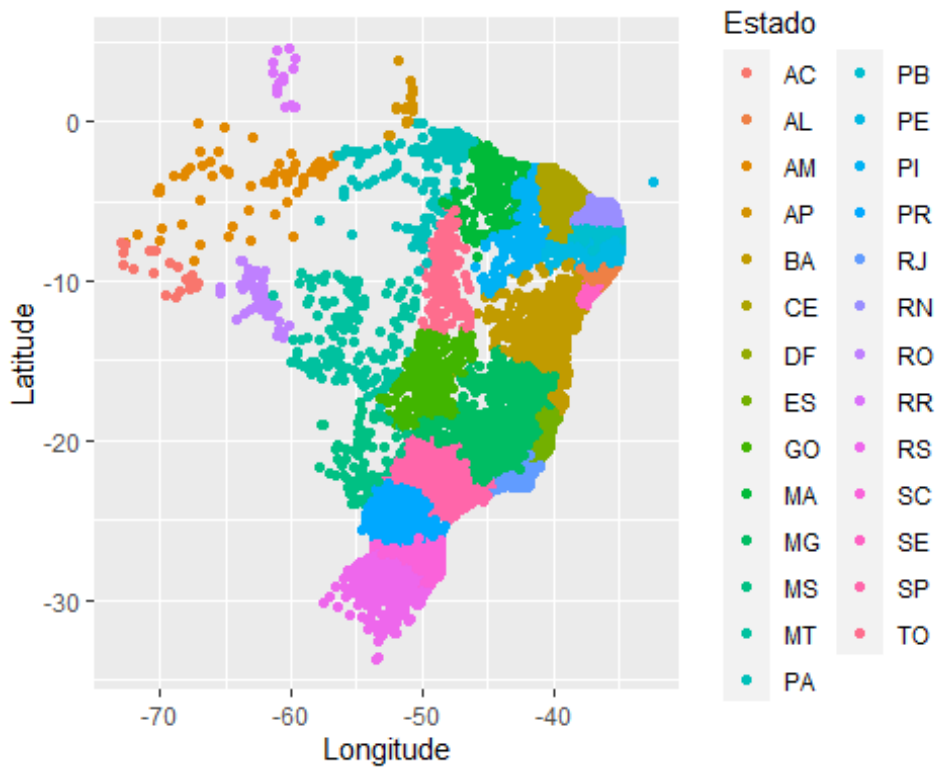
Andres Kindel

Gustavo Bereta

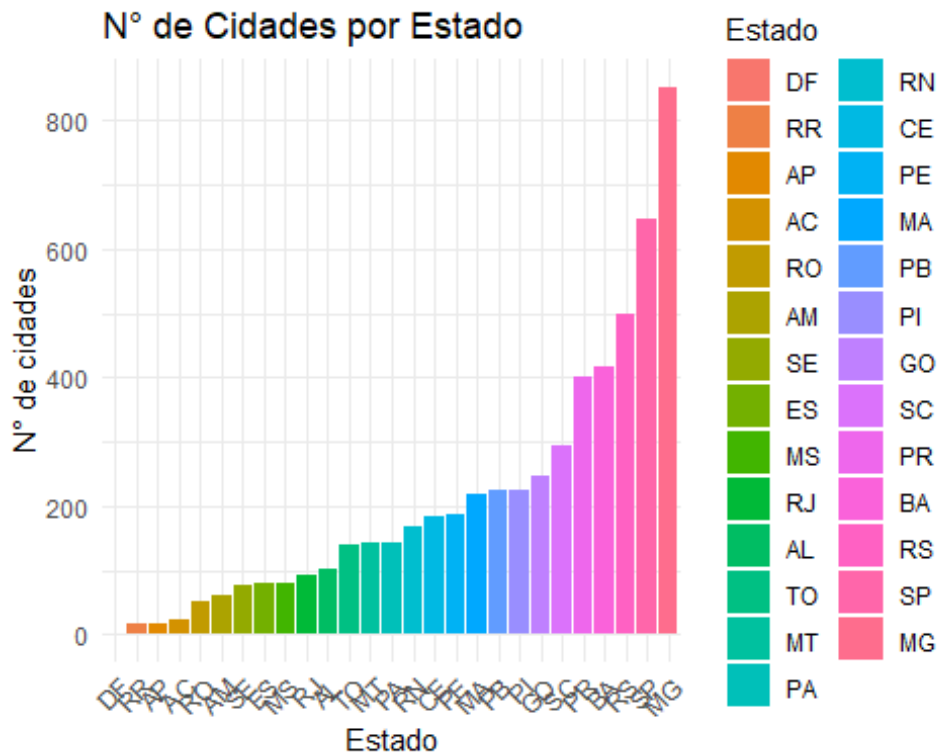
```
#carregando a base de dados
library(data.table)
base <- fread(input = paste0("brazil_cities_2022.csv"), header = T,
na.strings = "NA", data.table = FALSE, dec=",")

#organizando variaveis
base$IDHM <- as.numeric(base$IDHM)
base$LAT <- as.numeric(base$LAT)
base$LONG <- as.numeric(base$LONG)

#grafico posições geograficas das cidades
coordenadas <- data.frame(base$STATE, base$LONG, base$LAT, base$IDHM)
colnames(coordenadas) <- c("Estado", "Longitude", "Latitude", "IDH")
coordenadas <- subset(coordenadas, Longitude != 0)
coordenadas <- subset(coordenadas, Latitude != 0)
library(ggplot2)
ggplot(data = coordenadas, aes(Longitude, Latitude, color=Estado)) +
  geom_point()
```



```
#grafico de cidades por estado
valor_cidades_por_estado <- table(base$STATE)
valor_cidades_por_estado_ordenado <-
valor_cidades_por_estado[order(valor_cidades_por_estado)]
tabela_cidades_por_estado <-
data.frame(valor_cidades_por_estado_ordenado)
colnames(tabela_cidades_por_estado) <- c("Estado", "Nº de cidades")
ggplot(tabela_cidades_por_estado, aes(x = Estado, y = `Nº de cidades`,
fill = Estado)) +
  geom_bar(stat = "identity") +
  labs(title = "Nº de Cidades por Estado", x = "Estado", y = "Nº de
cidades") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



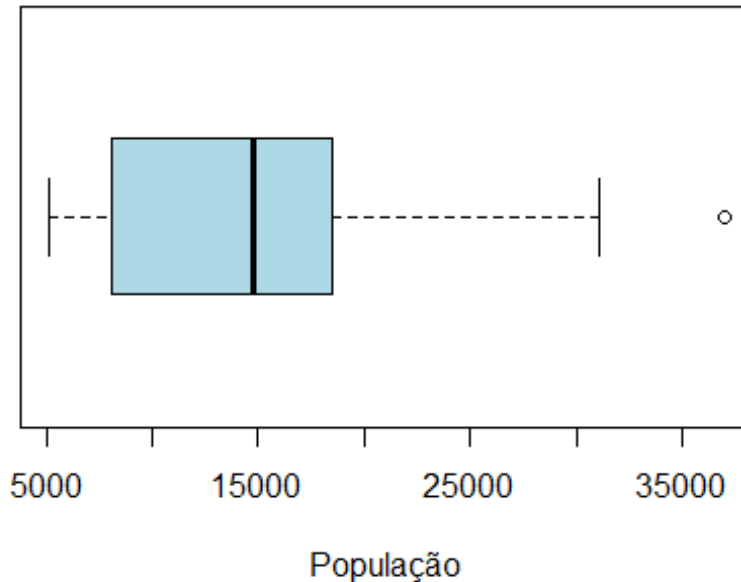
```
#analise da população das cidades
populacao_nao_zerada <- data.frame(base$ESTIMATED_POP, base$STATE)
colnames(populacao_nao_zerada) <- c("População", "Estado")
populacao_nao_zerada <- subset(populacao_nao_zerada, População != 0)
summary(populacao_nao_zerada$População)

##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
##      786     5454    11591    37439    25305   12176866

tabela_populacao_estados <- aggregate(População ~ Estado, data =
populacao_nao_zerada, FUN = sum)
colnames(tabela_populacao_estados) <- c("Estado", "População")
tabela_populacao_estados$`Nº de cidades` <- valor_cidades_por_estado
tabela_populacao_estados$`Média de população por cidade` <-
tabela_populacao_estados$`População` / tabela_populacao_estados$`Nº de
cidades`
mediana_populacao_estados <- aggregate(ESTIMATED_POP ~ STATE, data =
base, FUN = median)
mediana_populacao_estados_para_tabela <- mediana_populacao_estados[, -1]
tabela_populacao_estados$`Mediana de população por cidade` <-
mediana_populacao_estados_para_tabela
mediana_populacao_estados_semDF <-
mediana_populacao_estados_para_tabela[-7]
boxplot(mediana_populacao_estados_semDF,
        main = "Mediana da população das cidades por estado brasileiro",
        xlab = "População",
        ylab = "",
```

```
col = "lightblue",
horizontal = TRUE)
```

Mediana da população das cidades por estado bras



```
#analise do fluxo de caixa das cidades e comparação com nº de habitantes
fluxo_caixa_cidades <- data.frame(base$ESTIMATED_POP, base$TAXES*1000,
base$MUN_EXPENDIT, base$IDHM)
colnames(fluxo_caixa_cidades) <- c("População", "Impostos", "Gastos
municipais", "IDH")
fluxo_caixa_cidades <- subset(fluxo_caixa_cidades, `Gastos municipais` !=
0)
fluxo_caixa_cidades <- subset(fluxo_caixa_cidades, Impostos != 0)
fluxo_caixa_cidades <- subset(fluxo_caixa_cidades, População != 0)
summary(fluxo_caixa_cidades$População)

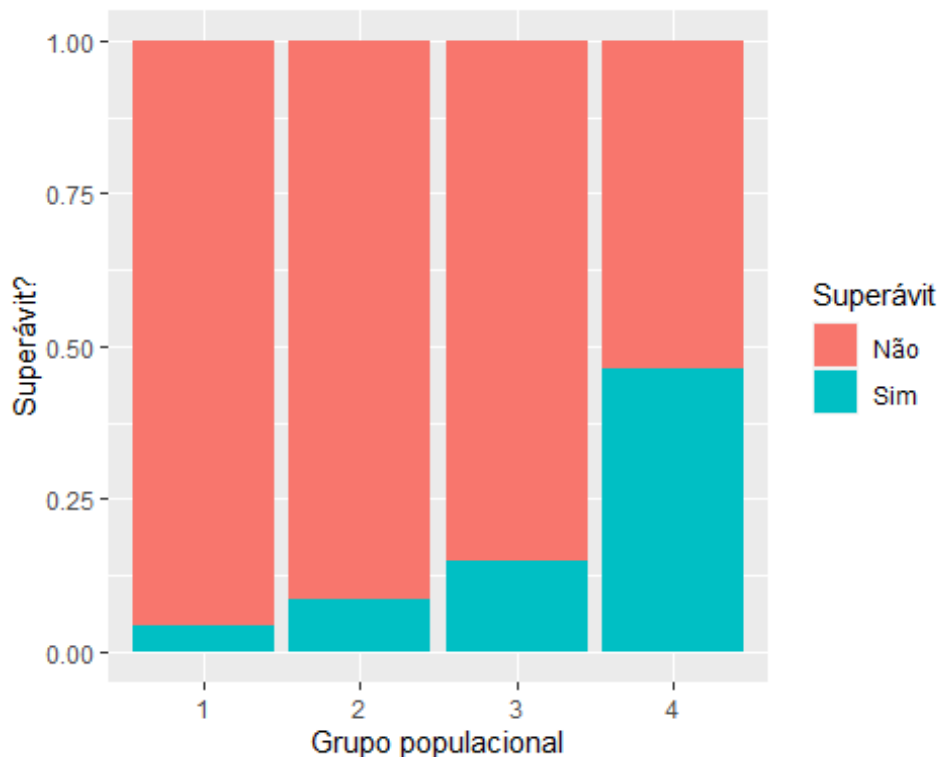
##      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
##      786     5354    11640    40951    25802 12176866

fluxo_caixa_cidades$`Grupo populacional` [fluxo_caixa_cidades$População
<= 5354] <- "1"
fluxo_caixa_cidades$`Grupo populacional` [fluxo_caixa_cidades$População >
5354 & fluxo_caixa_cidades$População <= 11640] <- "2"
fluxo_caixa_cidades$`Grupo populacional` [fluxo_caixa_cidades$População >
11640 & fluxo_caixa_cidades$População <= 25802] <- "3"
fluxo_caixa_cidades$`Grupo populacional` [fluxo_caixa_cidades$População >
25802] <- "4"
fluxo_caixa_cidades$Diferença <- fluxo_caixa_cidades$Impostos -
fluxo_caixa_cidades$`Gastos municipais`
```

```

fluxo_caixa_cidades$Superávit [fluxo_caixa_cidades$Diferença <= 0] <-
"Não"
fluxo_caixa_cidades$Superávit [fluxo_caixa_cidades$Diferença > 0] <-
"Sim"
grafico_lucro <- data.frame(table(fluxo_caixa_cidades$Superávit,
fluxo_caixa_cidades$`Grupo populacional`))
colnames(grafico_lucro) <- c("Superávit", "Grupo populacional", "Freq")
ggplot(grafico_lucro, aes(fill=Superávit, y=Freq, x=`Grupo
populacional`)) +
  geom_bar(position="fill", stat="identity") +
  ylab("Superávit?")

```



```

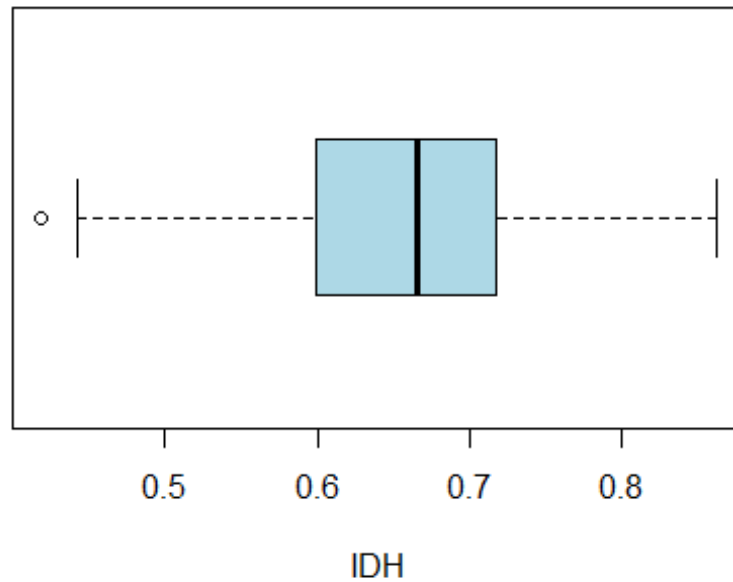
#IDH e população
idh_nao_zerado <- data.frame(base$ESTIMATED_POP, base$IDHM)
colnames(idh_nao_zerado) <- c("População", "IDH")
idh_nao_zerado <- subset(idh_nao_zerado, IDH != 0)
summary(idh_nao_zerado$IDH)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4180  0.5990  0.6650  0.6592  0.7180  0.8620

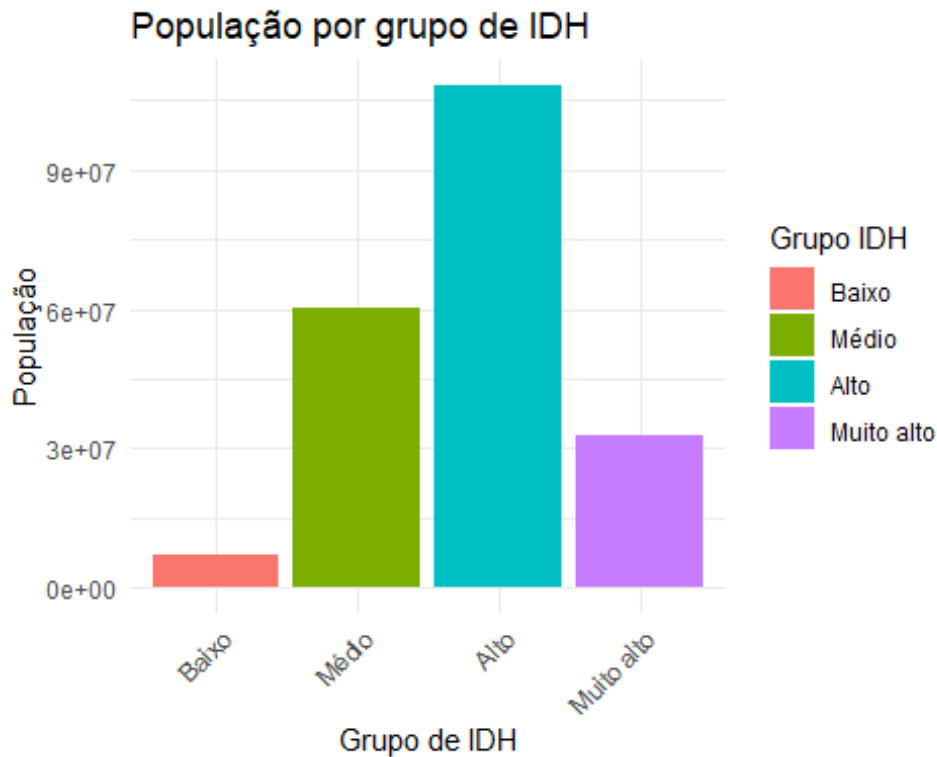
boxplot(idh_nao_zerado$IDH,
  main = "IDH das cidades brasileiras",
  xlab = "IDH",
  ylab = "",
  col = "lightblue",
  horizontal = TRUE)

```

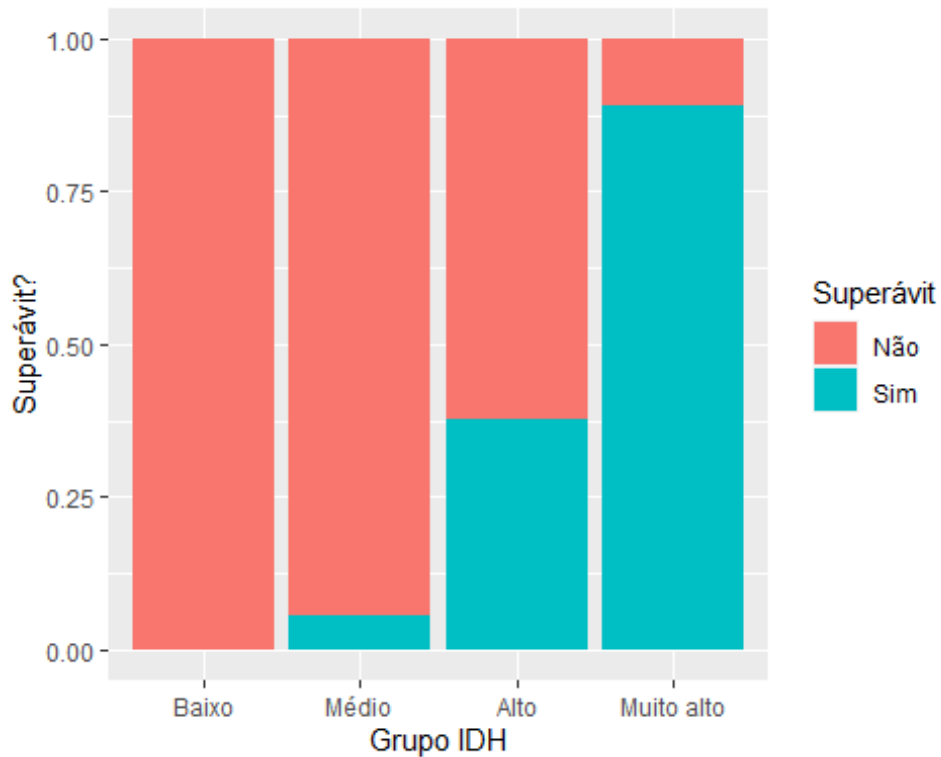
IDH das cidades brasileiras



```
idh_nao_zerado$`Grupo IDH` [idh_nao_zerado$IDH < 0.555] <- "Baixo"
idh_nao_zerado$`Grupo IDH` [idh_nao_zerado$IDH >= 0.555 &
idh_nao_zerado$IDH < 0.7] <- "Médio"
idh_nao_zerado$`Grupo IDH` [idh_nao_zerado$IDH >= 0.7 &
idh_nao_zerado$IDH < 0.8] <- "Alto"
idh_nao_zerado$`Grupo IDH` [idh_nao_zerado$IDH >= 0.8] <- "Muito alto"
tabela_para_grafico_grupoidh <- aggregate(População ~ `Grupo IDH`, data =
idh_nao_zerado, FUN = sum)
tabela_para_grafico_grupoidh$`Grupo IDH` <-
factor(tabela_para_grafico_grupoidh$`Grupo IDH`, levels = c("Baixo",
"Médio", "Alto", "Muito alto"))
ggplot(tabela_para_grafico_grupoidh, aes(x = `Grupo IDH`, y = População,
fill = `Grupo IDH`)) +
  geom_bar(stat = "identity") +
  labs(title = "População por grupo de IDH", x = "Grupo de IDH", y =
"População") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
#IDH e superavit
fluxo_caixa_cidades <- subset(fluxo_caixa_cidades, IDH != 0)
fluxo_caixa_cidades$`Grupo IDH` [fluxo_caixa_cidades$IDH < 0.555] <-
"Baixo"
fluxo_caixa_cidades$`Grupo IDH` [fluxo_caixa_cidades$IDH >= 0.555 &
fluxo_caixa_cidades$IDH < 0.7] <- "Médio"
fluxo_caixa_cidades$`Grupo IDH` [fluxo_caixa_cidades$IDH >= 0.7 &
fluxo_caixa_cidades$IDH < 0.8] <- "Alto"
fluxo_caixa_cidades$`Grupo IDH` [fluxo_caixa_cidades$IDH >= 0.8] <-
"Muito alto"
grafico_idh <- data.frame(table(fluxo_caixa_cidades$Superávit,
fluxo_caixa_cidades$`Grupo IDH`))
colnames(grafico_idh) <- c("Superávit", "Grupo IDH", "Freq")
grafico_idh$`Grupo IDH` <- factor(grafico_idh$`Grupo IDH`, levels =
c("Baixo", "Médio", "Alto", "Muito alto"))
ggplot(grafico_idh, aes(fill=Superávit, y=Freq, x=`Grupo IDH`)) +
  geom_bar(position="fill", stat="identity") +
  ylab("Superávit?")
```

#IDH e mapa:

```
coordenadas$`Grupo IDH` [coordenadas$IDH < 0.555] <- "Baixo"
coordenadas$`Grupo IDH` [coordenadas$IDH >= 0.555 & coordenadas$IDH <
0.7] <- "Médio"
coordenadas$`Grupo IDH` [coordenadas$IDH >= 0.7 & coordenadas$IDH < 0.8]
<- "Alto"
coordenadas$`Grupo IDH` [coordenadas$IDH >= 0.8] <- "Muito alto"
coordenadas$`Grupo IDH` <- factor(coordenadas$`Grupo IDH`, levels =
c("Baixo", "Médio", "Alto", "Muito alto"))
ggplot(data = coordenadas, aes(Longitude, Latitude, color=`Grupo IDH`)) +
  geom_point()
```

