

O relatório deverá conter a descrição e explicação dos cálculos realizados e gráficos obtidos de acordo com o conteúdo da disciplina, inclusive testes de hipótese. Naturalmente, nem todos os itens devem ser abordados. O grupo deverá determinar as análises a serem realizadas, ou seja, a descrição do problema e a solução a partir do uso dos dados e do R. Faça “perguntas” à base de dados e mostre o resultado.

Nomes do grupo: Erico, Daniel, Guilherme Nara e Gustavo Falconi

Tema: Registro de Ocupação Hospitalar COVID-19

Primeiro foi carregado o dataframe, retiradas colunas incompletas e retirados dados negativos

```
#Arrumar df

df_dados2022=read.csv("/cloud/project/esus-vepi.LeitoOcupacao_2022.csv")

df_dados2022<- subset(df_dados2022, select = -c( ocupacaoSuspeitoUti,ocupacaoSuspeitoCli,
ocupacaoConfirmadoCli, ocupacaoConfirmadoUti))

df_dados2022$dataNotificacao<-as.POSIXct(df_dados2022$dataNotificacao, origin="1970-01-
01")

#df_dados2021=read.csv("/cloud/project/esus-vepi.LeitoOcupacao_2021.csv")

library(dplyr)

df_dados2022<-df_dados2022 %>% filter(ocupacaoHospitalarUti >-1)
df_dados2022<-df_dados2022 %>% filter(ocupacaoHospitalarCli >-1)
df_dados2022<-df_dados2022 %>% filter(saidaSuspeitaObitos >-1)
df_dados2022<-df_dados2022 %>% filter(saidaSuspeitaAltas >-1)
df_dados2022<-df_dados2022 %>% filter(saidaConfirmadaObitos >-1)
df_dados2022<-df_dados2022 %>% filter(saidaConfirmadaAltas >-1)
df_dados2022<-df_dados2022 %>% filter(ocupacaoCovidCli >-1)
```

Em primeira análise usando o summary(df_dados2022)

Os dados de ocupação na CLI de covid, uti, saídaSuspeitaObitos e SaindaSuspeita altas tendem a zero tendo entre 2 a 3 quartis zerados e a maior média sendo ocupacaoHospitalarCli. Analisando os maiores valores vemos que o maior valor de ocupacaoCovidCli é 21701, será que isso está correto?

Analisando o registro logo é constatado que isso foi notificado em Barra do Piraí, Rio de Janeiro. Com uma rápida pesquisa descobre-se que a cidade possui pouco mais de 100000 habitantes, o que significaria 1/5 da cidade internado por covid-19 em 2022. Obviamente isso não é factível, inclusive um report da própria cidade reportou cerca de 10000 casos confirmados no período. Portanto esses dados serão limpos de nossa base. Após filtramos os dados ocupacaoCovidCli, saidaConfirmadaAltas, saidaSuspeitaAltas.

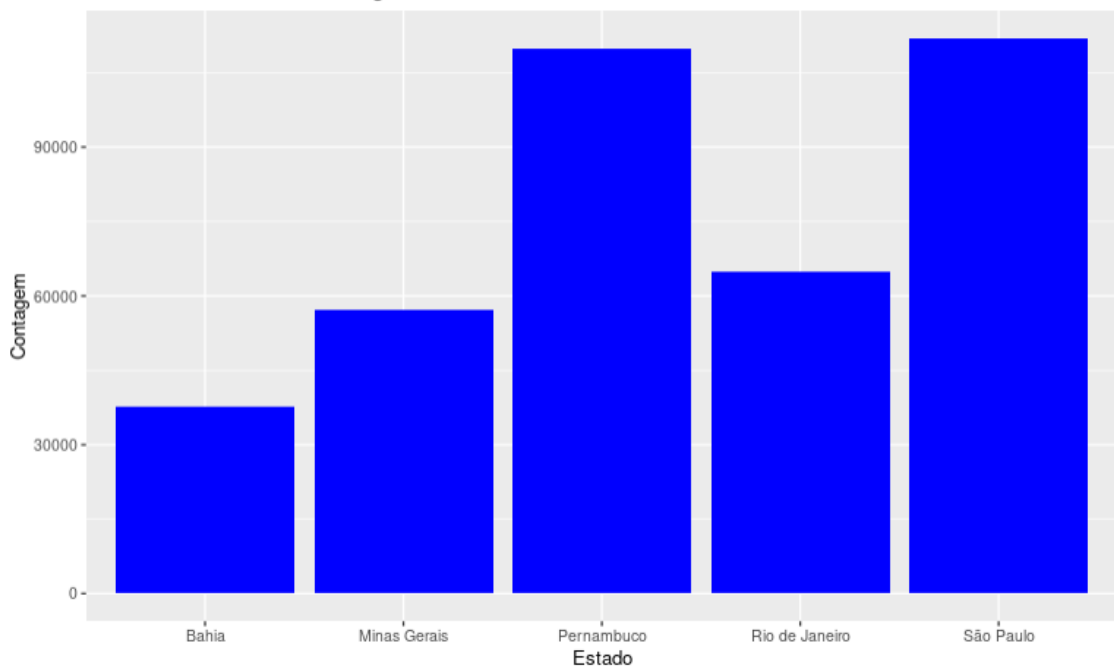
```
df_dados2022<-df_dados2022 %>% filter(ocupacaoCovidCli <20000)

df_dados2022<-df_dados2022 %>% filter(saidaConfirmadaAltas <1000)

df_dados2022<-df_dados2022 %>% filter(saidaSuspeitaAltas <2000)
```

Então quais os 5 estados com maior número de ocupação somada na uti por covid-19 são

Bar Plot of Estado vs. Contagem



São Paulo, Pernambuco, Rio de Janeiro, Minas Gerais, Bahia

```
contagem<-df_dados2022 %>% group_by(estado) %>%
```

```
summarise(somaOcupa=sum(ocupacaoCovidUti),
```

```
.groups = 'drop')
```

```
contagem_top5 <- contagem %>%
```

```
arrange(desc(somaOcupa)) %>%
```

```
head(5)
```

```
ggplot(contagem_top5, aes(x = estado, y = somaOcupa)) +
```

```
geom_bar(stat = "identity", fill = "blue") + # You can choose a different fill color if needed
```

```
labs(title = "Bar Plot of Estado vs. Contagem",
```

```
x = "Estado",
```

```
y = "Contagem")
```

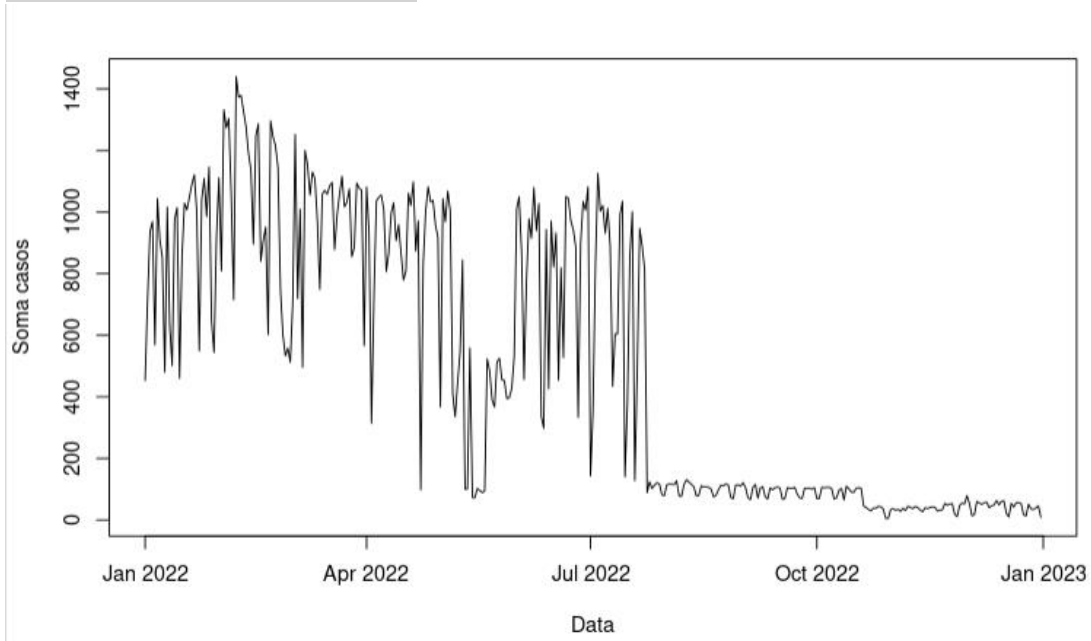
Como foi a evolução de óbitos?

```
contagem_Obitos<-df_dados2022 %>% group_by(dataNotificacao) %>%
```

```
summarise(somaOcupa=sum(saidaConfirmadaObitos),
```

```
.groups = 'drop')
```

```
plot(contagem_Obitos$dataNotificacao, y = contagem_Obitos$somaOcupa, type = "l", lty = 1,  
ylab = "Soma casos", xlab = "Data")
```



vemos que a contagem de casos foi reduzindo ao longo dos meses, em especial após julho.

Qual município foi mais atingido por óbitos de casos confirmados?

```
contagemMortes<-df_dados2022 %>% group_by(municipio) %>%
```

```
summarise(somaOcupa=sum(saidaConfirmadaObitos),
```

```
.groups = 'drop')
```

```
contagem_top6Mortes <- contagemMortes %>%
```

```
arrange(desc(somaOcupa)) %>%
```

```
head(6)
```

```
ggplot(contagem_top6Mortes, aes(x = municipio, y = somaOcupa)) +
```

```
geom_bar(stat = "identity", fill = "blue") +
```

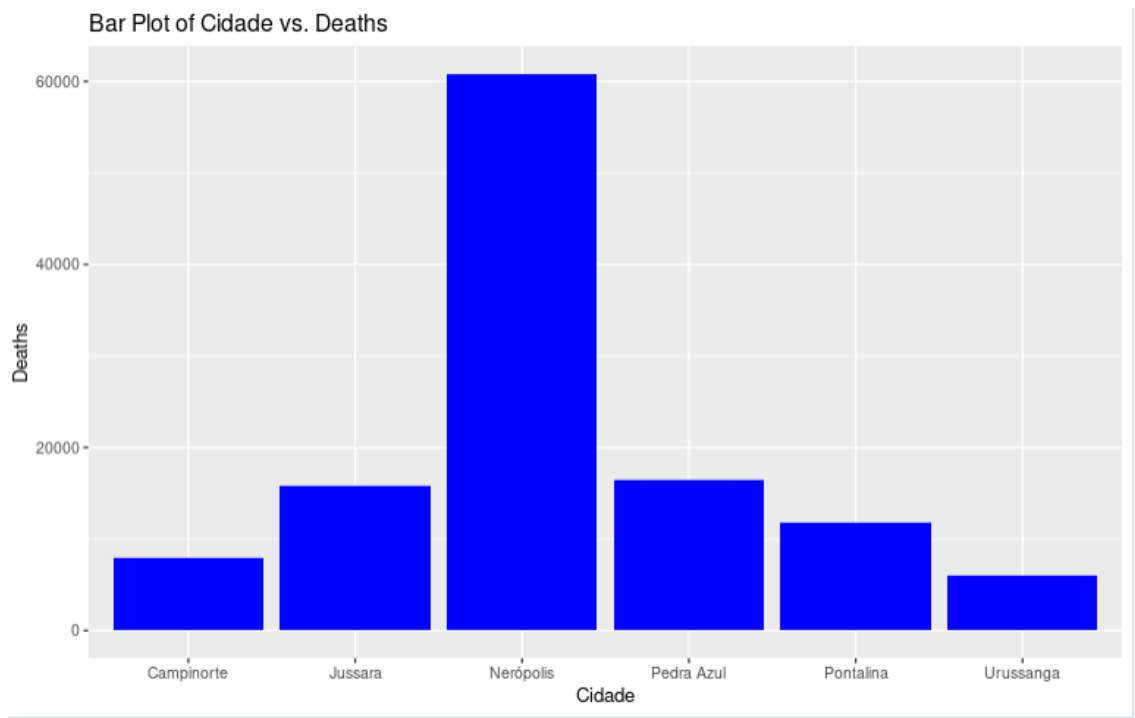
```
labs(title = "Bar Plot of Cidade vs. Deaths",
```

```
x = "Cidade",
```

```
y = "Deaths")
```

```
nome_da_variavel_x <- "municipio"
```

```
nome_da_variavel_y <- "saidaSuspeitaObitos "
```



Aqui nos deparamos com uma questão: Nerópolis aparece com o maior número, porém ao consultar outras fontes nos deparamos com valores máximos de 400, não 60000. Ao cruzar dados constata-se que isso é porque a cidade listou as mortes como diárias com o valor acumulado.

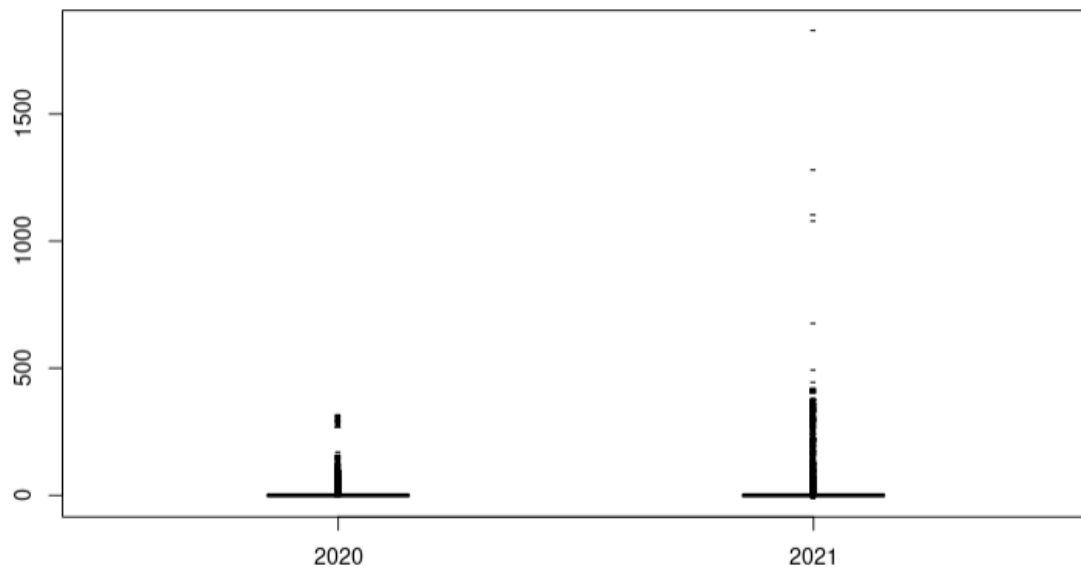
Casos no Município



<https://www.neropolis.go.gov.br/coronavirus/>, Acesso em 22/11/2023

Contudo continuaremos mostrando as respostas baseando-se nessa fonte de dados pois é baseado nelas que responderemos as perguntas.

É possível afirmar que aumentaram óbitos de 2020 para 2021?



```
df_dados2020=read.csv("/cloud/project/esus-vepi.LeitoOcupacao_2020.csv")
```

```
df_dados2021=read.csv("/cloud/project/esus-vepi.LeitoOcupacao_2021.csv")
```

```
summary(df_dados2020)
```

```
boxplot(df_dados2021$saidaConfirmadaObitos )
```

```
boxplot(df_dados2020$saidaConfirmadaObitos)
```

```
#Criando o boxplot
```

```
xp<-list(df_dados2020$saidaConfirmadaObitos,df_dados2021$saidaConfirmadaObitos)
```

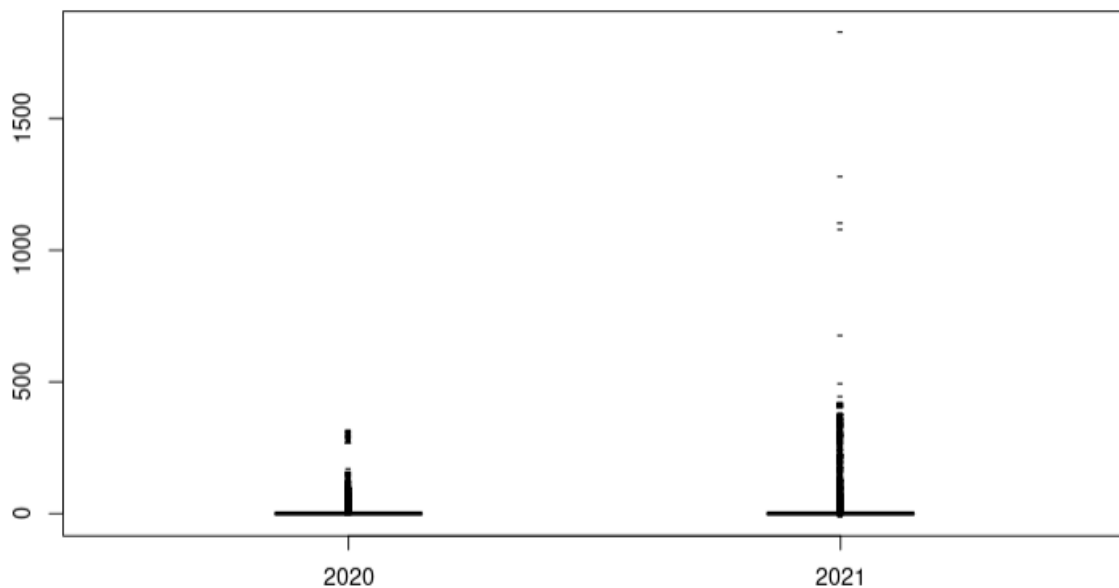
```
boxplot(xp,pch="-", col="lightblue", border="black", boxwex=0.3, names=c("2020","2021"))
```

Digno de nota que a distribuição de 2021 é mais abrangente, com dias de maior pico de acordo com o boxplot.

É possível afirmar que aumentaram óbitos de 2020 para 2021?

Primeiro comparamos as médias: 2020 com 0,373 e 2021 com 0,8594, o que é um indicativo de 2021 ser maior, mas após isso analisamos se estatisticamente isso faz sentido.

Gerando ambos os boxplots se notou uma diferença logo de início



```
summary(df_dados2020$saidaConfirmadaObitos)
```

```
summary(df_dados2021$saidaConfirmadaObitos)
```

```
xp<-list(df_dados2020$saidaConfirmadaObitos,df_dados2021$saidaConfirmadaObitos)
```

```
boxplot(xp,pch="-", col="lightblue", border="black", boxwex=0.3, names=c("2020","2021"))
```

Analisando se estatisticamente é possível afirmar a existência de uma variância estatística usamos o teste de variância, que mostrou que sim, pois $p < 0.05$.

Ao realizar o teste de hipótese, temos a hipótese nula que 2021 é menor ou igual a 2020 em mortes e a hipótese alternativa que 2021 é maior.

Sabendo disso fomos para o teste t em que o valor p é 0.000000000000000022, ou seja < 0.05 (alfa padrão), assim rejeitamos a hipótese nula e aceitamos a hipótese alternativa. Baseado nisso, um ano supera outro com alta evidência.

```
#Teste F para comparar as variâncias
```

```
r_0 = 1
```

```
var.test(x = df_dados2020$saidaConfirmadaObitos,
```

```
  y = df_dados2021$saidaConfirmadaObitos,
```

```
  ratio = 1,
```

```
  alternative = "two.sided",
```

```
  conf.level = .95)
```

```
#diferentes pois o p é 0.000000000000000022, que é menor que 0.05
```

```
#Teste t
```

```
t.test(x = df_dados2020$saidaConfirmadaObitos,
```

```
      y = df_dados2021$saidaConfirmadaObitos,
```

```
      alternative = "less",
```

```
      mu = 0,
```

```
      conf.level = 0.95,
```

```
      var.equal = FALSE)
```