

Análises sobre o Avanço do Coronavírus

Daniel N. Rocha

01/07/2020

1. Metodologia

Este trabalho visa analisar a evolução dos dados de Covid-19. Tendo em vista que há atraso na notificação dos casos, consideramos a média móvel de uma semana, de modo a ter uma estimativa melhor da evolução dos óbitos e dos novos casos. Todos os dados deste trabalho foram coletado do site “brasil.io”, que divulga dados abertos da pandemia a partir dos boletins das Secretarias Estaduais de Saúde (SES).

Importando as bibliotecas

```
library(ggpubr)
library(purrr)
library(tidyr)
library(dplyr)
library(ggplot2)
library(tidyquant)
```

2. Óbitos de Coronavírus

Importando os dados

```
covid_nit <- read.csv('Downloads/niteroicovid19.csv')
```

Covid-19 em Niterói

Explorando o cabeçalho

```
covid_nit <- read.csv("/Users/pedrocerbino/Downloads/niteroicovid19.csv")
knitr::kable(head(covid_nit[,c(2,5,12,13,14)]))
```

date	city	new_deaths	last_available_death_rate	estimated_population_2019
2020-06-30	Niterói	45	0.0361	513584
2020-06-29	Niterói	0	0.0291	513584
2020-06-28	Niterói	1	0.0292	513584
2020-06-27	Niterói	1	0.0301	513584
2020-06-26	Niterói	-1	0.0300	513584
2020-06-25	Niterói	0	0.0311	513584

Explorando a estrutura

```
str(covid_nit)
```

```
## 'data.frame': 111 obs. of 16 variables:
## $ epidemiological_week : int 27 27 27 26 26 26 26 26 26 26 ...
## $ date : chr "2020-06-30" "2020-06-29" "2020-06-28" "2020-06-27" ...
## $ order_for_place : int 111 110 109 108 107 106 105 104 103 102 ...
## $ state : chr "RJ" "RJ" "RJ" "RJ" ...
## $ city : chr "Niterói" "Niterói" "Niterói" "Niterói" ...
## $ city_ibge_code : int 3303302 3303302 3303302 3303302 3303302 3303302 3303302 ...
## $ place_type : chr "city" "city" "city" "city" ...
## $ last_available_confirmed : int 6377 6357 6346 6121 6105 5922 5694 5540 5426 ...
## $ last_available_confirmed_per_100k_inhabitants : num 1242 1238 1236 1192 1189 ...
## $ new_confirmed : int 20 11 225 16 183 228 154 114 56 23 ...
## $ last_available_deaths : int 230 185 185 184 183 184 184 179 176 175 ...
## $ new_deaths : int 45 0 1 1 -1 0 5 3 1 0 ...
## $ last_available_death_rate : num 0.0361 0.0291 0.0292 0.0301 0.03 0.0311 0.032 ...
## $ estimated_population_2019 : int 513584 513584 513584 513584 513584 513584 513584 ...
## $ is_last : chr "True" "False" "False" "False" ...
## $ is_repeated : chr "False" "False" "False" "False" ...
```

Explorando o resumo

```
knitr::kable(summary(covid_nit[,c(2,5,12,13,14)]))
```

date	city	new_deaths	last_available_death_rate	estimated_population_2019
Length:111	Length:111	Min. :-1.000	Min. :0.00000	Min. :513584
Class	Class	1st Qu.: 0.000	1st Qu.:0.03140	1st Qu.:513584
:character	:character	Median :	Median :0.03630	Median :513584
Mode	Mode	Mean : 2.072	Mean :0.04555	Mean :513584
:character	:character	3rd Qu.: 2.000	3rd Qu.:0.06400	3rd Qu.:513584
NA	NA	Max. :45.000	Max. :0.14290	Max. :513584
NA	NA			

Transformando em variável “data” e colocando em ordem cronológica

```
covid_nit$date <- as.Date(covid_nit$date)
covid_nit <- map_df(covid_nit, rev)
```

Covid-19 no Rio de Janeiro

```
covid_riocity <- read.csv('/Users/pedrocerbino/Downloads/riocitycovid19.csv')
covid_riocity$date <- as.Date(covid_riocity$date)
covid_riocity <- map_df(covid_riocity, rev)
```

Covid-19 em São Gonçalo

```
sgcovid19 <- read.csv('/Users/pedrocerbino/Downloads/sgcovid19.csv')
sgcovid19$date <- as.Date(sgcovid19$date)
sgcovid19 <- map_df(sgcovid19, rev)
```

Plotando a evolução de óbitos em São Gonçalo

```
rjnitsg <- ggplot(sgcovid19,aes(x=date,y=new_deaths,linetype='-')) +  
  #Usando a média móvel de 7 dias para suavizar as curvas  
  geom_ma(size=1,n=7,aes(color="São Gonçalo")) +  
  #Repetindo os passos para Niterói e para o Rio  
  geom_ma(size=1,data=covid_nit,n=7,aes(color="Niterói")) +  
  geom_ma(size=1,data=covid_riocity,n=7,aes(color="Rio de Janeiro")) +  
  #Adicionando título, removendo nome dos eixos e renomeando legenda  
  ggtitle("Óbitos de Coronavírus (média móvel de 7 dias)") + labs(x = '', y = '') +  
  #Definindo a palheta de cores manualmente  
  scale_color_manual(values=c("São Gonçalo"="blue",  
    "Niterói"="#D10046","Rio de Janeiro"="#E1AF00")) +  
  #Removendo legenda da espessura da linha  
  scale_linetype(guide = "none") +  
  #Alterando tamanho do título, eixos e legenda,  
  #além de apagar título da legenda e mudar a posição da mesma  
  theme(plot.title = element_text(size=14),  
    axis.text = element_text(size=10),  
    legend.title = element_blank(),  
    legend.text = element_text(size=10),  
    legend.position=c(.12,.6))
```

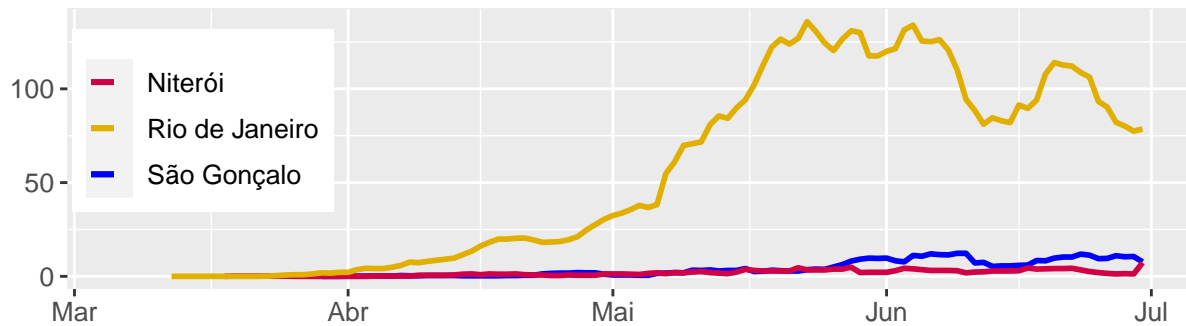
Plotando somente para Niterói e São Gonçalo

```
nitsg <- ggplot(sgcovid19,aes(x=date,y=new_deaths,linetype='-')) +  
  geom_ma(size=1,n=7,color='blue',show.legend=F) +  
  geom_ma(size=1,data=covid_nit,n=7,color='#D10046',show.legend=F) +  
  ggtitle('Óbitos de Coronavírus (média móvel de 7 dias) - Zoom') +  
  labs(x = '', y = '') +  
  theme(plot.title = element_text(size=14),  
    axis.text = element_text(size=10)) +  
  #Criando anotações no gráfico (ao invés de legenda)  
  annotate("text", x=c(as.Date('2020-05-24'),as.Date('2020-06-24')),  
    y=c(11,0.4),label=c("São Gonçalo","Niterói"),colour=c("blue","#D10046"), size=4)
```

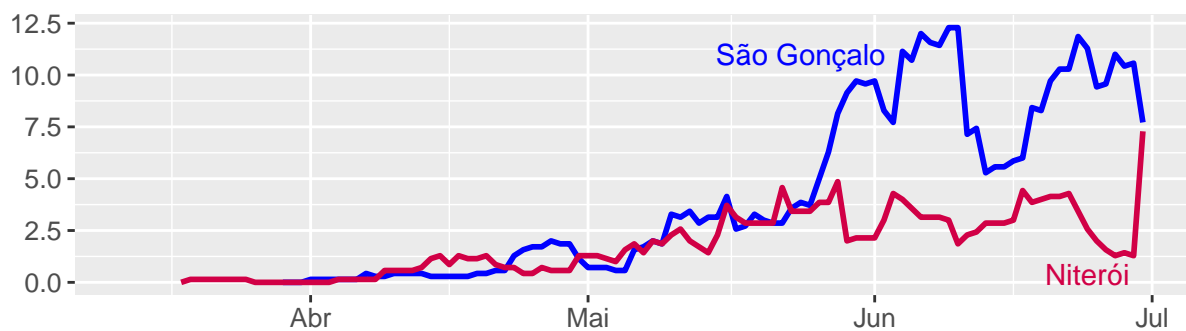
Plotando vários os 2 gráficos na mesma figura

```
ggarrange(rjnitsg, nitsg, ncol=1, nrow=2)
```

Óbitos de Coronavírus (média móvel de 7 dias)



Óbitos de Coronavírus (média móvel de 7 dias) – Zoom



3. Novos Casos de Coronavírus

Importando dados de Covid-19 de todos os estados

```
covid_estados <- read.csv('/Users/pedrocerbino/Downloads/covid19estados.csv')
```

Ajustando o nome da coluna para facilitar visualização da tabela

```
covid_estados <- covid_estados %>% rename(confper100khab=last_available_confirmed_per_100k_inhabitants)
```

Explorando o cabeçalho

```
knitr::kable(head(covid_estados[,c(2,5,8,9,10)]))
```

date	city	last_available_confirmed	confper100khab	new_confirmed
2020-06-30	NA	13253	1502.7185	107
2020-06-30	NA	35962	1077.5593	846
2020-06-30	NA	70823	1708.8030	930
2020-06-30	NA	28492	3368.9199	239
2020-06-30	NA	73307	492.8843	3840
2020-06-30	NA	110483	1209.8342	2258

Explorando a estrutura

```
str(covid_estados)
```

```
## 'data.frame': 2974 obs. of 16 variables:
## $ epidemiological_week : int 27 27 27 27 27 27 27 27 27 27 ...
## $ date : chr "2020-06-30" "2020-06-30" "2020-06-30" "2020-06-30" ...
## $ order_for_place : int 106 115 110 103 117 107 116 118 111 103 ...
## $ state : chr "AC" "AL" "AM" "AP" ...
## $ city : logi NA NA NA NA NA NA ...
## $ city_ibge_code : int 12 27 13 16 29 23 53 32 52 21 ...
## $ place_type : chr "state" "state" "state" "state" ...
## $ last_available_confirmed : int 13253 35962 70823 28492 73307 110483 49217 46893 24910 80451 ...
## $ confper100khab : num 1503 1078 1709 3369 493 ...
## $ new_confirmed : int 107 846 930 239 3840 2258 2146 1796 1531 1482 ...
## $ last_available_deaths : int 365 1052 2823 417 1853 6146 587 1647 491 2048 ...
## $ new_deaths : int 4 20 31 4 53 70 28 27 24 36 ...
## $ last_available_death_rate: num 0.0275 0.0293 0.0399 0.0146 0.0253 0.0556 0.0119 0.0351 0.0197 0.0197 ...
## $ estimated_population_2019: int 881935 3337357 4144597 845731 14873064 9132078 3015268 4018650 70 ...
## $ is_last : chr "True" "True" "True" "True" ...
## $ is_repeated : chr "False" "False" "False" "False" ...
```

Explorando o resumo

```
knitr::kable(summary(covid_estados[,c(2,5,8,9,10)]))
```

date	city	last_available_confirmed	confper100khab	new_confirmed
Length:2974	Mode:logical	Min. : 0.0	Min. : 0.002	Min. : -290.0
Class :character	NA's:2974	1st Qu.: 210.2	1st Qu.: 4.513	1st Qu.: 15.0
Mode :character	NA	Median : 2271.0	Median : 45.656	Median : 126.0
NA	NA	Mean : 12582.9	Mean : 224.164	Mean : 473.1
NA	NA	3rd Qu.: 12587.2	3rd Qu.: 266.075	3rd Qu.: 529.0
NA	NA	Max. :281380.0	Max. :3368.920	Max. :19030.0
NA	NA	NA	NA's :6	NA

Transformando em variável “data” e filtrando pelas últimas duas semanas

```
covid_estados$date <- as.Date(covid_estados$date)
covid_estados <- filter(covid_estados,date>'2020-06-16')
```

Agrupando a última semana por estado e tirando a média de casos novos

```
covid_estados_1 <- covid_estados %>% filter(date>'2020-06-23') %>%
  group_by(state) %>% summarise(avg=mean(new_confirmed))
```

Agrupando a penúltima semana por estado e tirando a média de casos novos

```
covid_estados_2 <- covid_estados %>% filter(date<'2020-06-24') %>%
  group_by(state) %>% summarise(avg=mean(new_confirmed))
```

Calculando, entre as duas semanas, a variação média percentual de novos casos

```
covid_estados_1$difnewcases <- covid_estados_1$avg - covid_estados_2$avg  
covid_estados_1$difnewcases <- covid_estados_1$difnewcases/covid_estados_2$avg
```

Fazendo uma cópia da coluna de estados. Adiante, a coluna original será sobrescrita por com as regiões

```
covid_estados_1$state_1 <- covid_estados_1$state  
covid_estados_2$state_1 <- covid_estados_2$state
```

Criando vetores de cada região

```
N <- c('N', 'AM', 'RR', 'AP', 'PA', 'TO', 'RO', 'AC')  
NE <- c('NE', 'MA', 'BA', 'AL', 'SE', 'PB', 'PE', 'RN', 'CE', 'PI')  
CO <- c('CO', 'MT', 'MS', 'GO', 'DF')  
SE <- c('SE', 'SP', 'RJ', 'ES', 'MG')  
S <- c('S', 'PR', 'RS', 'SC')  
regioes <- list(N, NE, S, SE, CO)
```

Substituindo os dados originais de estado pelos das regiões

```
for (k in regioes){  
  for (i in covid_estados_1$state){if(i %in% k){  
    covid_estados_1$state <- gsub(i,k[1],covid_estados_1$state)}}}
```

Renomeando as colunas para facilitar a identificação

```
covid_estados_1 <- covid_estados_1 %>% rename(estados=state_1, regioes=state)
```

Agrupando as linhas por região (para que, na hora de plotar, as colunas de estados de mesma região fiquem agrupadas)

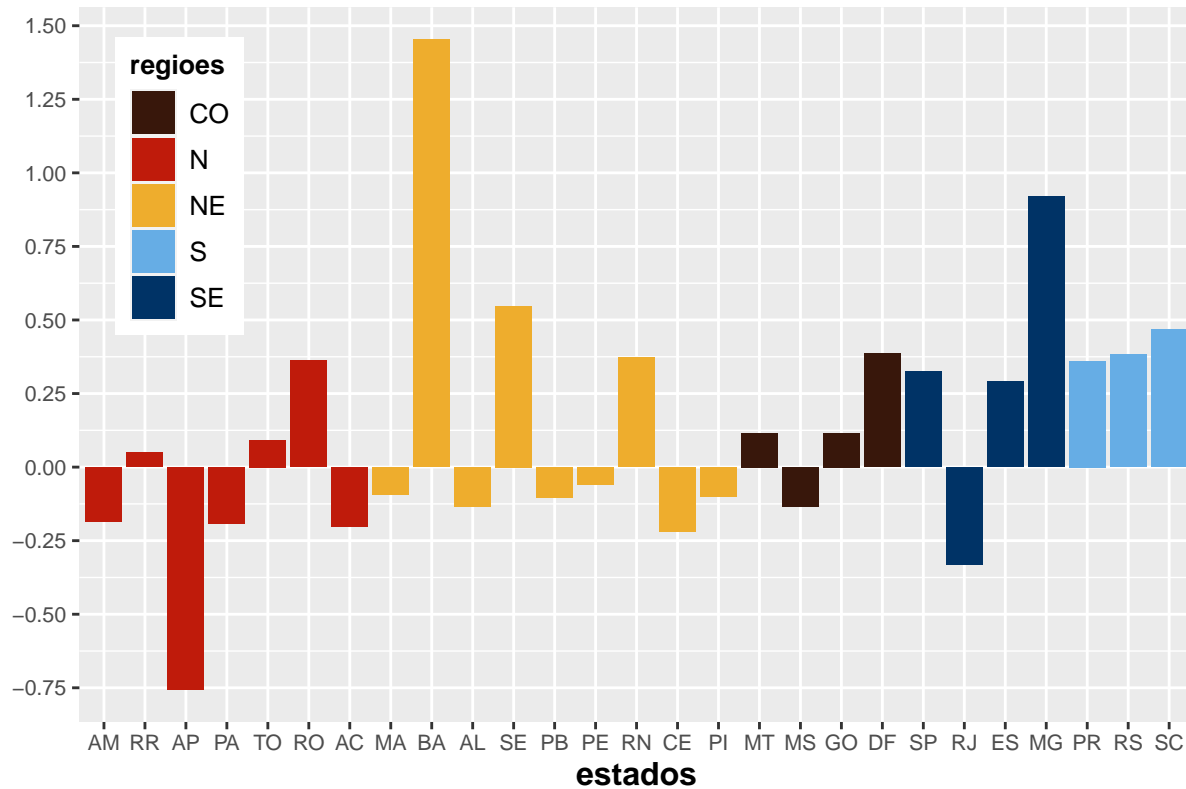
```
covid_estados_1$estados <- factor(covid_estados_1$estados, levels=  
  c(N[-1], NE[-1], CO[-1], SE[-1], S[-1]))
```

Plotando a variação percentual de óbitos para cada estado. Diferentemente do outro gráfico, plotamos barras ao invés de linhas

```
ggplot(covid_estados_1, aes(x=estados, y=difnewcases)) +  
  geom_bar(stat="identity", aes(fill=regioes)) + labs(y = '') +  
  scale_fill_manual(values=c("#38170B", "#BF1B0B", "#EEAD2D", "#66ADE5", "#003366")) +  
  ggtitle('Variação Percentual de Novos Casos (média móvel de 7 dias)') +  
  #Definindo a faixa de valores do eixo y  
  scale_y_continuous(breaks = seq(-0.75, 1.5, 0.25)) +  
  theme(plot.title = element_text(size=14),  
        axis.text = element_text(size=8),  
        axis.title = element_text(size=12, face="bold"),  
        legend.title = element_text(size=10, face="bold"),
```

```
legend.text = element_text(size=10),
legend.position=c(.09,.75))
```

Variação Percentual de Novos Casos (média móvel de 7 dias)



4. Referências:

Sthda - Tutorial Barplot
 Information is Beautiful - Estética de Gráficos
 UESC - R para Cientistas Sociais
 Brasil.io - Dados Coronavírus
 Rstudio - Cheatsheets
 R Documentation - Fonte Oficial
 R-Bloggers - Tutoriais
 R-Graph Gallery - Visualizações
 Stackoverflow - Dúvidas