Analysis of Covid-19 in South Africa

Mpilo Khumalo

2022-06-05

Loading Packages

```
library(tidyverse)
library(ggplot2)
library(cowplot)
library(dplyr)
library(lubridate)
library(plyr)
library(rvest)
```

Loading Data

```
raw_confirmedcases_data <- read.csv("time_series_covid19_confirmed_global.csv", sep = ",", header = T)
raw_recoveredcases_data <- read.csv("time_series_covid19_recovered_global.csv", sep = ",", header = T)
```

Web Scraping Lockdown Alerts Data

```
link = "https://www.gov.za/covid-19/about/about-alert-system"
page = read_html(link)

lockdowns_alerts = page %>% html_nodes("p:nth-child(3) , p:nth-child(4) , p:nth-child(5) , p:nth-child(

#Alert_level <- c("alert level 5", "alert level 3", "alert level 2", "alert level 1", "alert level 3",

#Dates <- c("26 March 2020", "01 June 2020", "18 August 2020", "21 September 2020", "29 December 2020",
```

Cleaning Confirmed Cases Data

```
southafrica_confirmedcases_df <- filter(raw_confirmedcases_data, Country.Region == "South Africa")
southafrica_confirmedcases_df <- southafrica_confirmedcases_df[, -c(1:4)]

southafrica_confirmedcases_df <- southafrica_confirmedcases_df %>%
  gather(key = "Date",
         value = "Confirmed_cases")
```

```
southafrica_confirmedcases_df$Date <- substr(southafrica_confirmedcases_df$Date, 2, 9)
southafrica_confirmedcases_df$Date <- gsub("\\.", "-", southafrica_confirmedcases_df$Date)
southafrica_confirmedcases_df$Date <- as.Date(southafrica_confirmedcases_df$Date, format="%m-%d-%y")
southafrica_confirmedcases_df$Date2 <- as.character(southafrica_confirmedcases_df$Date)
southafrica_confirmedcases_df$Year <- substr(southafrica_confirmedcases_df$Date2, start = 1, stop = 4)
southafrica_confirmedcases_df$Month <- substr(southafrica_confirmedcases_df$Date2, start = 6, stop = 7)
southafrica_confirmedcases_df$Day <- substr(southafrica_confirmedcases_df$Date2, start = 9, stop = 10)

southafrica_confirmedcases_df <- southafrica_confirmedcases_df %>%
  mutate(Month_name = case_when(startsWith(Month, "01") ~ "January",
                                startsWith(Month, "02") ~ "February",
                                startsWith(Month, "03") ~ "March",
                                startsWith(Month, "04") ~ "April",
                                startsWith(Month, "05") ~ "May",
                                startsWith(Month, "06") ~ "June",
                                startsWith(Month, "07") ~ "July",
                                startsWith(Month, "08") ~ "August",
                                startsWith(Month, "09") ~ "Sepember",
                                startsWith(Month, "10") ~ "October",
                                startsWith(Month, "11") ~ "November",
                                startsWith(Month, "12") ~ "December"))

southafrica_confirmedcases_df<- southafrica_confirmedcases_df%>%
  mutate(Alert_level = case_when(Date2 < "2020-03-26" ~ "No Lockdown",
                            Date2 >= "2020-03-26" & Date2 <= "2020-04-30" ~ "Level 5",
                            Date2 >= "2020-06-01" & Date2 <= "2020-08-17" ~ "Level 3",
                            Date2 >= "2020-08-18" & Date2 < "2020-09-21"~ "Level 2" ,
                            Date2 >= "2020-09-21" & Date2 <= "2020-12-28" ~ "Level 1",
                            Date2 >= "2020-12-29" & Date2 <= "2021-02-28" ~ "Adjusted Level 3",
                            Date2 >= "2021-03-01" & Date2 <= "2021-05-30" ~ "Adjusted Level 1",
                            Date2 >= "2021-05-31" & Date2 <= "2021-06-15" ~ "Adjusted Level 2",
                            Date2 >= "2021-06-16" & Date2 <= "2021-06-27" ~ "Adjusted Level 3",
                            Date2 >= "2021-06-28" & Date2 <= "2021-07-25" ~ "Adjusted Level 4",
                            Date2 >= "2021-07-26" & Date2 <= "2020-09-12" ~ "Adjusted Level 3",
                            Date2 >= "2021-09-13" & Date2 <= "2021-10-01" ~ "Adjusted Level 2",
                            Date2 > "2021-10-01" ~ "Adjusted Level 1",
                            TRUE ~ "Unspecified lockdown"))

southafrica_confirmedcases_df <- southafrica_confirmedcases_df[, c(1, 4,5,7, 6, 2, 8)]

attach(southafrica_confirmedcases_df)
```

Cleaning Recovered Cases Data

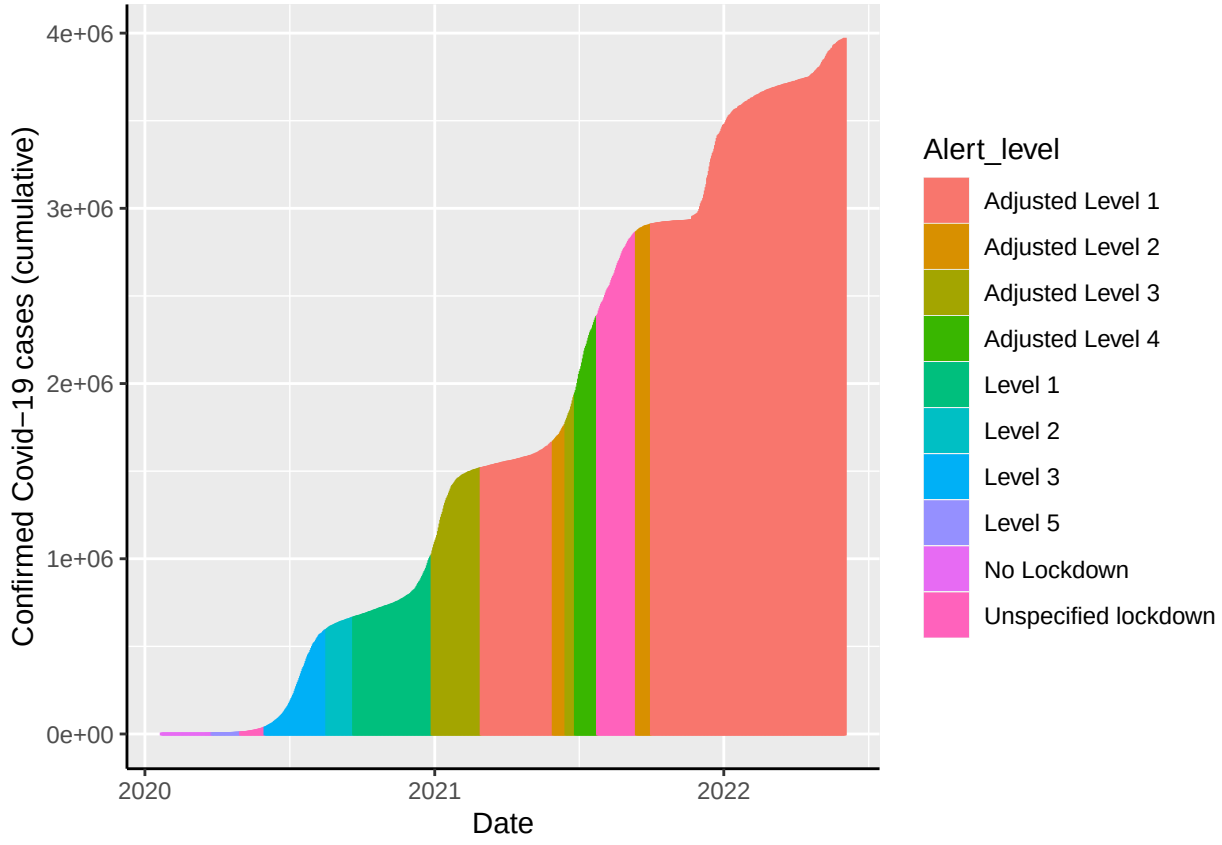
```
southafrica_recoveredcases_df <- filter(raw_recoveredcases_data, Country.Region == "South Africa")
southafrica_recoveredcases_df <- southafrica_recoveredcases_df[, -c(1:4)]

southafrica_recoveredcases_df <- southafrica_recoveredcases_df %>%
  gather(key = "Date",
         value = "Recovered_cases")
```

```
southafrica_recoveredcases_df <- filter(southafrica_recoveredcases_df, Recovered_cases > 0)
southafrica_confirmedcases_df2 <- filter(southafrica_confirmedcases_df, Date >= "2020-03-24" & Date <=
")
southafrica_confirmedcases_df2 <- southafrica_confirmedcases_df2[, 6]
southafrica_recoveredcases_df = cbind(southafrica_recoveredcases_df,southafrica_confirmedcases_df2)

southafrica_recoveredcases_df$Date <- substr(southafrica_recoveredcases_df$Date, 2, 9)
southafrica_recoveredcases_df$Date <- gsub("\\.", "-", southafrica_recoveredcases_df$Date)
southafrica_recoveredcases_df$Date <- as.Date(southafrica_recoveredcases_df$Date, format="%m-%d-%y")
southafrica_recoveredcases_df$Date2 <- as.character(southafrica_recoveredcases_df$Date)
southafrica_recoveredcases_df$Year <- substr(southafrica_recoveredcases_df$Date2, start = 1, stop = 4)
southafrica_recoveredcases_df$Month <- substr(southafrica_recoveredcases_df$Date2, start = 6, stop = 7)
southafrica_recoveredcases_df$Day <- substr(southafrica_recoveredcases_df$Date2, start = 9, stop = 10)

southafrica_recoveredcases_df <- southafrica_recoveredcases_df %>%
  mutate(Month_name = case_when(startsWith(Month, "01") ~ "January",
                                startsWith(Month, "02") ~ "February",
                                startsWith(Month, "03") ~ "March",
                                startsWith(Month, "04") ~ "April",
                                startsWith(Month, "05") ~ "May",
                                startsWith(Month, "06") ~ "June",
                                startsWith(Month, "07") ~ "July",
                                startsWith(Month, "08") ~ "August",
                                startsWith(Month, "09") ~ "Sepember",
                                startsWith(Month, "10") ~ "October",
                                startsWith(Month, "11") ~ "November",
                                startsWith(Month, "12") ~ "December"))

southafrica_recoveredcases_df<- southafrica_recoveredcases_df%>%
  mutate(Alert_level = case_when(Date2 < "2020-03-26" ~ "No Lockdown",
                            Date2 >= "2020-03-26" & Date2 <= "2020-04-30" ~ "Level 5",
                            Date2 >= "2020-06-01" & Date2 <= "2020-08-17" ~ "Level 3",
                            Date2 >= "2020-08-18" & Date2 < "2020-09-21"~ "Level 2" ,
                            Date2 >= "2020-09-21" & Date2 <= "2020-12-28" ~ "Level 1",
                            Date2 >= "2020-12-29" & Date2 <= "2021-02-28" ~ "Adjusted Level 3",
                            Date2 >= "2021-03-01" & Date2 <= "2021-05-30" ~ "Adjusted Level 1",
                            Date2 >= "2021-05-31" & Date2 <= "2021-06-15" ~ "Adjusted Level 2",
                            Date2 >= "2021-06-16" & Date2 <= "2021-06-27" ~ "Adjusted Level 3",
                            Date2 >= "2021-06-28" & Date2 <= "2021-07-25" ~ "Adjusted Level 4",
                            Date2 >= "2021-07-26" & Date2 <= "2020-09-12" ~ "Adjusted Level 3",
                            Date2 >= "2021-09-13" & Date2 <= "2021-10-01" ~ "Adjusted Level 2",
                            Date2 > "2021-10-01" ~ "Adjusted Level 1",
                            TRUE ~ "Unspecified lockdown"))

southafrica_recoveredcases_df <- southafrica_recoveredcases_df[, c(1, 5:8, 2, 3)]
colnames(southafrica_recoveredcases_df)[7] <- "Confirmed_cases"

southafrica_recoveredcases_df <- southafrica_recoveredcases_df %>%
  gather(key = "Cases.",
         value = "Cases",
         -c("Date", "Year", "Month", "Day", "Month_name"))

attach(southafrica_recoveredcases_df)
```

Data Visualisations

Confirmed Cases with Lockdowns

```
(graph1 <- ggplot(southafrica_confirmedcases_df, aes(Date, Confirmed_cases, fill = Alert_level, col= Al
  geom_bar(stat="identity", position = "dodge") +
  labs(x = "Date", y = "Confirmed Covid-19 cases (cumulative)") +
  theme(panel.background = element_rect(),
                    axis.line = element_line(colour ="black")))
```



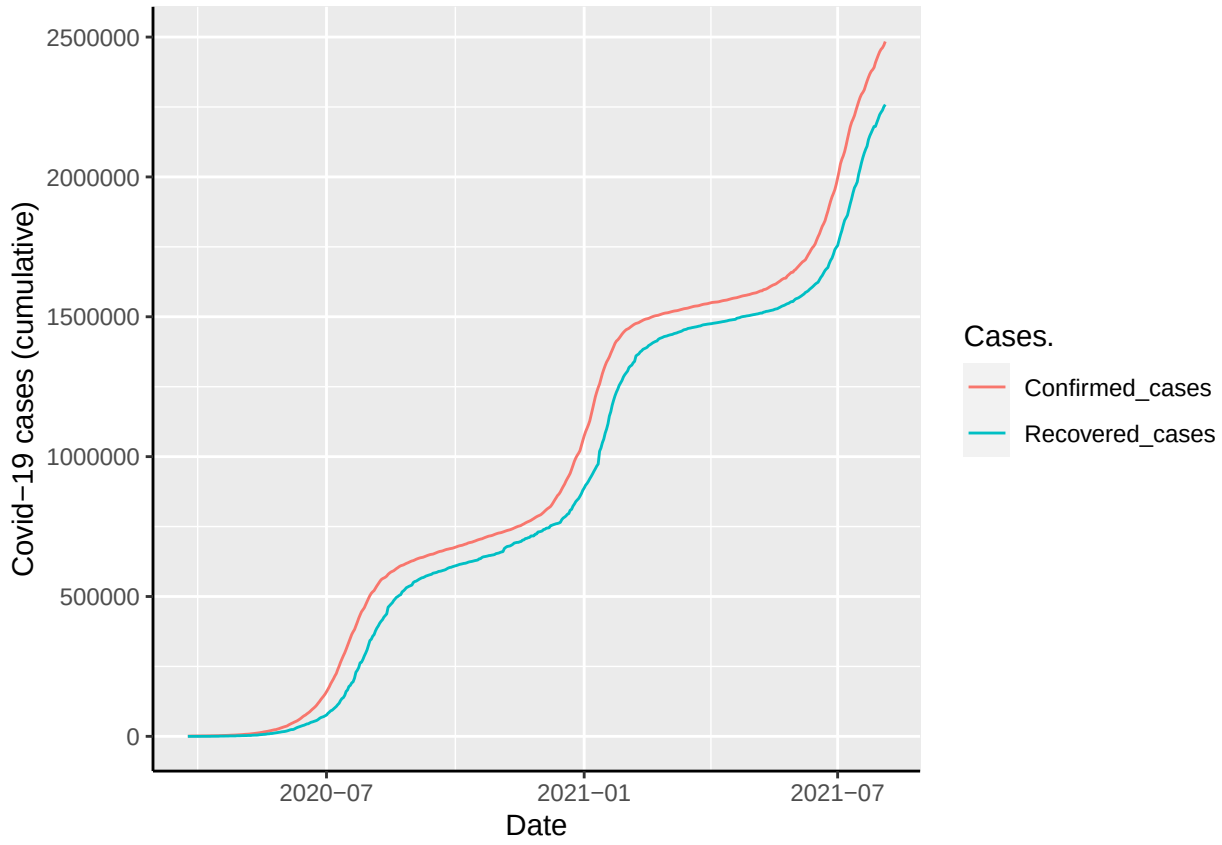

```
ggsave(filename = "confirmed_covid-19_cases_South_Africa.png", graph1)
```

```
## Saving 6.5 x 4.5 in image
```

Confirmed Cases vs Recorvered Cases

```
#ggplot(southafrica_confirmedcases_df, aes(Date, Confirmed_cases)) +
#  geom_line()

(graph2 <- ggplot(southafrica_recoveredcases_df, aes(x = Date, y = Cases, colour = Cases.)) +
   geom_line() +
   labs(x = "Date", y = "Covid-19 cases (cumulative)") +
```

```
  theme(panel.background = element_rect(),
                  axis.line = element_line(colour ="black")))
```




```
ggsave(filename = "covid-19_cases_South_Africa.png", graph2)
```

```
## Saving 6.5 x 4.5 in image
```