COVID-19 Trends in Rwanda (Apr 2020 - Jul 2021)

Sarah Agatoni

0: Summary

This project uses data from the Rwandan Ministry of Health Twitter account to:

- explore the spread of COVID-19 in Rwanda between 17 April 2020 to 23 July 2021
- explore how COVID-19 impacted various demographics during this period

1: Data Loading & Prep

loading packages

```
if (!require(pacman)) install.packages(pacman); library(pacman)
p_load(broom, choroplethrAdmin1, data.table, dplyr, ggplot2, ggpubr,
       ggthemr, janitor, lemon, lubridate, maps, rgdal, sp)
```

reading data in csv format

```
#data manually obtained from Rwanda Ministry of Health Twitter
data <- data.table::fread("covid_rw.csv")
demo <- data.table::fread("covid_demo.csv")
```

clean-up

```
data$Date <- dmy(data$Date)
demo$Date <- dmy(demo$Date)
data <- janitor::clean_names(data)
demo <- janitor::clean_names(demo)
```

2: Data Viz: Positivity Trend

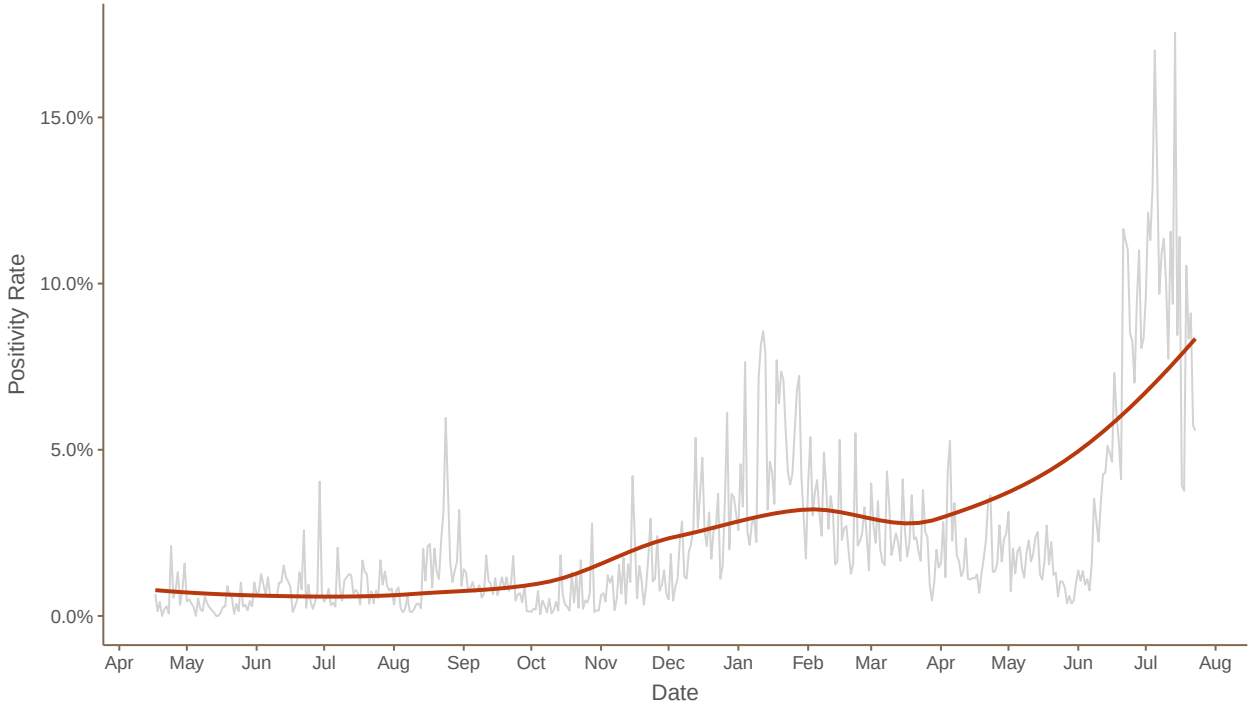
Data Prep

```
#calculating positivity rate as new daily cases per tests administered
data <- data %>%
  mutate(pos_rate = new_cases/tests_today)
```

Line Graph: Positivity Trend

```
ggthemr("fresh")
plot_1 <- ggplot(data = data, mapping = aes(x = date, y = pos_rate)) +
  geom_line(col = 'light grey') +
  geom_smooth(col = '#B8390E', se = FALSE) +
  scale_y_continuous("Positivity Rate", labels = scales::percent) +
  scale_x_date("Date", date_breaks = "1 month", date_labels = '%b') +
  no_gridlines() +
  theme(plot.margin = unit(c(1,1,1,1), "cm"))
plot_1
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



3: Data Viz: Deaths by Age & Gender

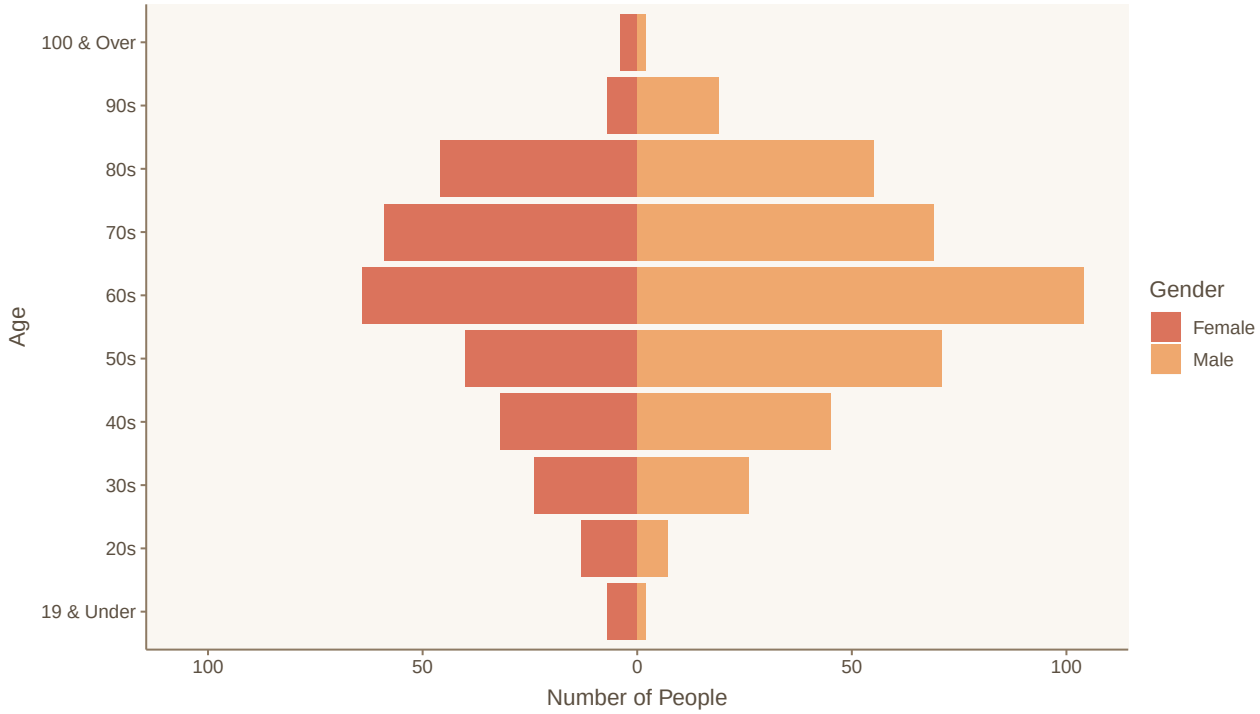
Data Prep

```
# grouping death counts by age and gender
by_age <- na.omit(demo) %>%

  mutate(gender = case_when(gender == "F" ~ "Female", gender == "M" ~"Male"),
         agegroup = case_when(age <= 19 ~ '19 & Under',
                              age >= 20  & age <= 29 ~ '20s',
                              age >= 30  & age <= 39 ~ '30s',
                              age >= 40  & age <= 49 ~ '40s',
                              age >= 50  & age <= 59 ~ '50s',
                              age >= 60  & age <= 69 ~ '60s',
                              age >= 70  & age <= 79 ~ '70s',
                              age >= 80  & age <= 89 ~ '80s',
                              age >= 90  & age <= 99 ~ '90s',
                              age >= 100 ~ '100 & Over',
                              is.na(age) == TRUE ~ "Unclear"),
         agegroup = factor(agegroup,
                           levels = c("Unclear", "19 & Under", "20s", "30s", "40s",
                                      "50s", "60s", "70s", "80s", "90s", "100 & Over"),
                           ordered = TRUE)) %>%

 count(agegroup, gender, name = "fatalities")
```

Population Pyramid: Deaths by Age & Gender

```
ggthemr('dust')
plot_2 <- ggplot(
  data = by_age,
```

```
  mapping = aes(x = ifelse(test = gender == "Female", yes = -fatalities, no = fatalities), y = agegroup
  geom_col() +
  lemon::scale_x_symmetric(labels = abs) +
  labs(x = "Number of People", y = "Age", fill = "Gender") +
  theme(plot.margin = unit(c(1,1,1,1), "cm")) +
  no_gridlines()

plot_2
```



4: Data Viz: Deaths by District

Data Prep

```
dist <- na.omit(demo) %>%
  group_by(district) %>%
  count(name = "deaths")

#adding Ngoma district; missing in original data file due to zero deaths
dist <- as.data.frame(dist)
by_dist <- rbind(dist, c("Ngoma",0))
by_dist$deaths <- as.integer(by_dist$deaths)


#normalizing death counts by district i.e. calculating deaths per population
#population counts obtained from PEPFAR
#https://www2.census.gov/programs-surveys/international-programs/tables/time-series/pepfar/rwanda.xlsx
by_dist <- by_dist %>%
 mutate(population = case_when(
    district == "Kigali"~ 1577326,
    district == "Nyarugenge" ~ 338862,
    district == "Gasabo" ~ 795658,
```

```
    district == "Kicukiro" ~ 442806,
    district == "Nyanza" ~ 410717,
    district == "Gisagara" ~ 361573,
    district == "Nyaruguru" ~ 341196,
    district == "Huye" ~ 369480,
    district == "Nyamagabe" ~ 381677,
    district == "Ruhango" ~ 374582,
    district == "Muhanga" ~ 331126,
    district == "Kamonyi" ~ 401146,
    district == "Karongi" ~ 361647,
    district == "Rutsiro" ~ 361121,
    district == "Rubavu" ~ 526725,
    district == "Nyabihu" ~ 301035,
    district == "Ngororero" ~ 359828,
    district == "Rusizi" ~ 443967,
    district == "Nyamasheke" ~ 409888,
    district == "Rulindo" ~ 303379,
    district == "Gakenke" ~ 331763,
    district == "Musanze" ~ 423647,
    district == "Burera" ~ 338043,
    district == "Gicumbi" ~ 432700,
    district == "Rwamagana" ~ 393112,
    district == "Nyagatare" ~ 767710,
    district == "Gatsibo" ~ 607320,
    district == "Kayonza" ~ 489568,
    district == "Kirehe" ~ 440339,
    district == "Ngoma" ~ 424047,
    district == "Bugesera" ~ 444823),

    norm_fat = (deaths/population) * 100000
    )

#reading map data as shape files
#obtained from Stanford's Libraries: https://earthworks.stanford.edu/catalog/stanford-qy869sx9298)

my_spdf <- readOGR(dsn = ("~/Documents/Data Projects/rwshapefiles/"))
```

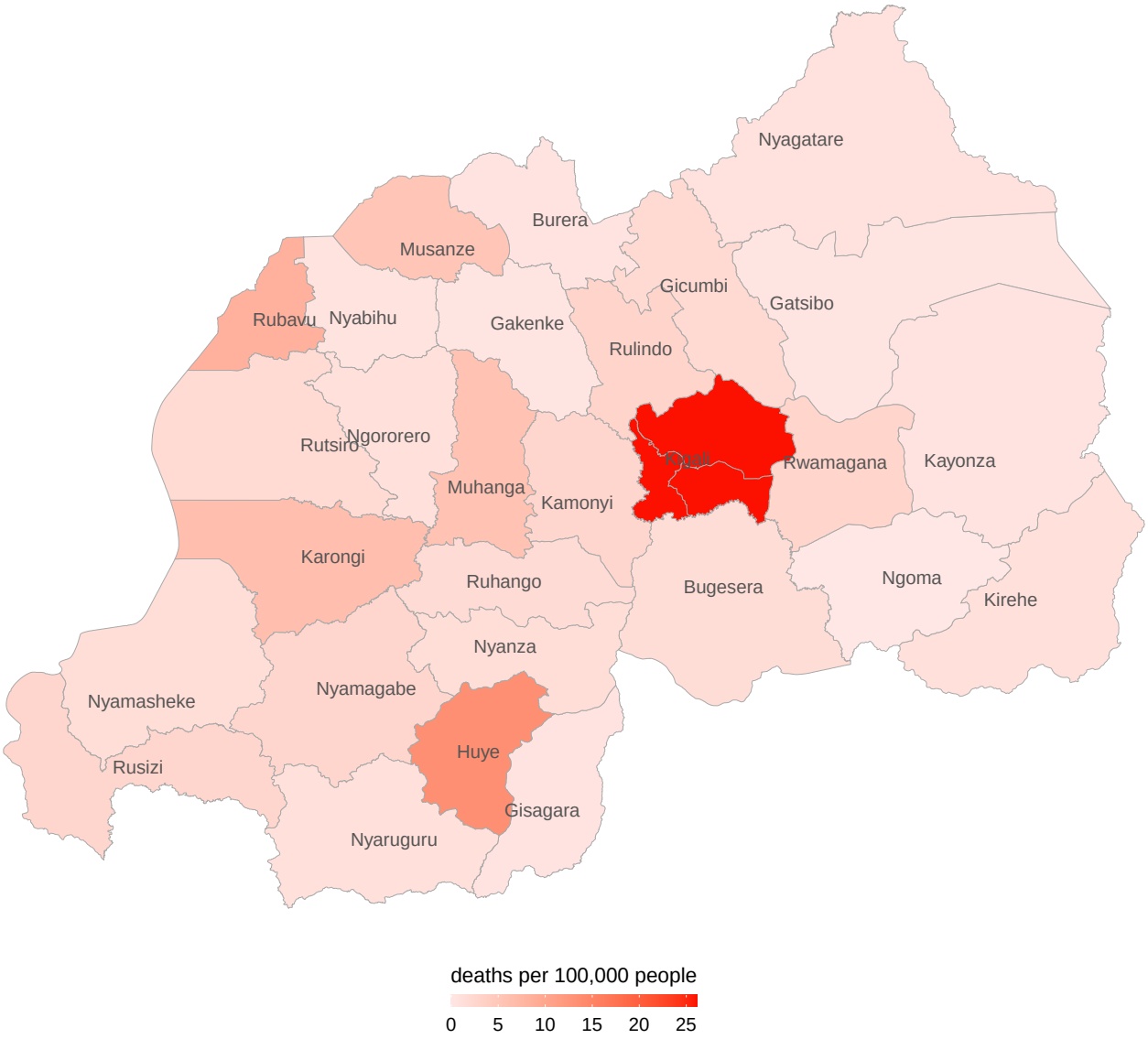
```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/sarahagatoni/Documents/Data Projects/rwshapefiles", layer: "RWA_adm2"
## with 30 features
## It has 14 fields
## Integer64 fields read as strings:  ID_0 ID_1 ID_2 CCN_2
```

```
#converting shape files to tidy data
#recoding "gasabo", "kicukiro" and "nyarugenge" as "kigali"; covid-19 data for Kigali isn't stratified
spdf_tidy <- tidy(my_spdf, region = "NAME_2") %>%
  mutate(group = as.character(group),
         region = case_when(
           grepl("Gasabo|Kicukiro|Nyarugenge", group, ignore.case = TRUE) ~ "Kigali",
           TRUE ~ id
    ))

#relating map data with covid-19 data
choro_df <- left_join(spdf_tidy, by_dist, by = c("region" = "district"))
```

```
#adding district labels
dist_names <- spdf_tidy %>%
  group_by(region) %>%
  summarise(long = mean(long), lat = mean(lat))
```

Choropleth Map: Deaths by District

```
plot_4 <- ggplot() +

  #gradient map
  geom_polygon(data = choro_df,
               mapping = aes(x = long, y = lat, group = group, fill = norm_fat), color = "dark grey", si
  theme_void() +

  #color scheme with horizontal legend at the bottom
  scale_fill_gradient(
    low ="#fee7e5",
    high = "#fc1100",
    guide = guide_colorbar(title = "deaths per 100,000 people",
                           direction = "horizontal",
                           barheight = unit(2, units = "mm"),
                           barwidth = unit(42, units = "mm"),
                           title.position = 'top',
                           title.theme = element_text(size = 10))) +

  #legend position & plot margins
  theme(legend.position = "bottom",
        plot.margin = unit(c(1,1,1,1), "cm")) +

  #adding district names
  geom_text(data = dist_names,
            mapping =  aes(x = long, y = lat, label = region),
            size = 3, hjust = "center") +

  #fixing aspect ratio
  coord_fixed(0.9)

plot_4
```


Nyagatare
Burera
Musanze
Gicumbi
Gatsibo
Rubavu
Nyabihu
Gakenke
Rulindo
Rutsiro
Ngororero
Muhanga
Kamonyi
Rwamagana
Kayonza
Karongi
Ruhango
Bugesera
Ngoma
Kirehe
Nyanza
Nyamagabe
Nyamasheke
Huye
Rusizi
Gisagara
Nyaruguru
deaths per 100,000 people
0
5
10
15
20
25