

Final Project

2022-10-11

Final Project 2: Reproducible Report on COVID19 Data

##Project Step 1: Start an Rmd Document, Install Packages and Import the data

```
library(tidyverse)
library(lubridate)
library(ggplot2)
library(dplyr)

url_in <- ("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series.csv")

file_names <-
  c("time_series_covid19_confirmed_global.csv",
  "time_series_covid19_deaths_global.csv",
  "time_series_covid19_confirmed_US.csv",
  "time_series_covid19_deaths_US.csv")

urls <- str_c(url_in, file_names)

global_cases <- read_csv(urls[1])
global_deaths <- read_csv(urls[2])
US_cases <- read_csv(urls[3])
US_deaths <- read_csv(urls[4])
```

Project Step 2: Tidy and Transform Data

```
global_cases <- global_cases %>%
  pivot_longer(cols =
    -c('Province/State',
      'Country/Region', Lat, Long),
    names_to = "date",
    values_to = "cases")

global_deaths <- global_deaths %>%
  pivot_longer(cols =
    -c('Province/State',
      'Country/Region', Lat, Long),
    names_to = "date",
    values_to = "deaths")
```

```

global <- global_cases %>%
  full_join(global_deaths) %>%
  rename(Country_Region = 'Country/Region',
         Province_State = 'Province/State') %>%
  mutate(date = mdy(date))

## Joining, by = c("Province/State", "Country/Region", "Lat", "Long", "date")

summary(global)

##   Province_State      Country_Region          Lat           Long
##   Length:286977      Length:286977      Min.   :-71.950   Min.   :-178.12
##   Class :character    Class :character    1st Qu.: 3.934    1st Qu.: -42.60
##   Mode  :character    Mode  :character    Median : 21.513   Median : 20.94
##                                         Mean   : 19.719   Mean   : 22.18
##                                         3rd Qu.: 40.464   3rd Qu.: 90.36
##                                         Max.   : 71.707   Max.   : 178.06
##                                         NA's   :1986     NA's   :1986
##   date              cases        deaths
##   Min.   :2020-01-22   Min.   :      0   Min.   :      0
##   1st Qu.:2020-09-26   1st Qu.:    412   1st Qu.:       2
##   Median :2021-06-01   Median : 9980   Median :    113
##   Mean   :2021-06-01   Mean   :762339   Mean   : 11892
##   3rd Qu.:2022-02-04   3rd Qu.:169396   3rd Qu.:   2451
##   Max.   :2022-10-10   Max.   :96721113   Max.   :1062681
## 

US_cases <- US_cases %>%
  pivot_longer(cols = -(UID:Combined_Key),
               names_to = "date",
               values_to = "cases") %>%
  select(Admin2:cases) %>%
  mutate(date = mdy(date)) %>%
  select (-c(Lat, Long_))

US_deaths <- US_deaths %>%
  pivot_longer(cols = -(UID:Population),
               names_to = "date",
               values_to = "deaths") %>%
  select(Admin2:deaths) %>%
  mutate(date = mdy(date)) %>%
  select (-c(Lat, Long_))

US <- US_cases %>%
  full_join(US_deaths)

## Joining, by = c("Admin2", "Province_State", "Country_Region", "Combined_Key",
## "date")

```

```
summary(US)
```

```
##      Admin2      Province_State      Country_Region      Combined_Key
##  Length:3318606  Length:3318606  Length:3318606  Length:3318606
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      date      cases      Population      deaths
##  Min.   :2020-01-22  Min.   :-3073  Min.   :    0  Min.   :-82.0
##  1st Qu.:2020-09-26  1st Qu.:   215  1st Qu.: 9917  1st Qu.:   2.0
##  Median :2021-06-01  Median :  1789  Median : 24892  Median :  30.0
##  Mean   :2021-06-01  Mean   : 11684  Mean   : 99604  Mean   : 165.8
##  3rd Qu.:2022-02-04  3rd Qu.:  6631  3rd Qu.: 64979  3rd Qu.: 105.0
##  Max.   :2022-10-10  Max.   :3464157  Max.   :10039107 Max.   :33740.0
```

```
##Project Step 3: Add Visualizations and Analysis for State of New Jersey
```

```
nj <- US %>%
  filter(Province_State == "New Jersey", cases > 0) %>%
  group_by(date, Admin2)

nj_counties <- nj %>%
  group_by/Admin2, date) %>%
  mutate(mortality_rate = deaths / cases) %>%
  select/Admin2, date, cases, deaths, Population, mortality_rate)

nj_totals <- nj %>%
  group_by(date) %>%
  summarize(cases = sum(cases), deaths = sum(deaths), Population = sum(Population)) %>%
  select(date, cases, deaths, Population) %>%
  ungroup()

current_counties <- nj_counties %>%
  group_by/Admin2) %>%
  mutate(county_mortality_rate = deaths/cases) %>%
  select(date, Admin2, cases, deaths, Population, county_mortality_rate) %>%
  ungroup()

# Total NJ date.
max(nj_totals$cases)
```

```
## [1] 2759767
```

```
# Total NJ deaths to date.
max(nj_totals$deaths)
```

```
## [1] 34792
```

```

# NJ mortality rate:
max(nj_totals$deaths) / max(nj_totals$cases)

## [1] 0.01260686

# NJ county with the highest mortality rate:
current_counties %>% slice_max(county_mortality_rate)

## # A tibble: 1 x 6
##   date      Admin2    cases deaths Population county_mortality_rate
##   <date>     <chr>    <dbl>  <dbl>      <dbl>                  <dbl>
## 1 2020-05-28 Unassigned    32    1809        0                 56.5

# NJ county with the lowest mortality rate:
current_counties %>% slice_min(county_mortality_rate)

## # A tibble: 1,096 x 6
##   date      Admin2    cases deaths Population county_mortality_rate
##   <date>     <chr>    <dbl>  <dbl>      <dbl>                  <dbl>
## 1 2020-03-18 Atlantic     3      0    263670                  0
## 2 2020-03-19 Atlantic     3      0    263670                  0
## 3 2020-03-20 Atlantic     3      0    263670                  0
## 4 2020-03-21 Atlantic     4      0    263670                  0
## 5 2020-03-22 Atlantic     5      0    263670                  0
## 6 2020-03-23 Atlantic     6      0    263670                  0
## 7 2020-03-24 Atlantic     6      0    263670                  0
## 8 2020-03-25 Atlantic    12      0    263670                  0
## 9 2020-03-26 Atlantic    10      0    263670                  0
## 10 2020-03-27 Atlantic   14      0    263670                  0
## # ... with 1,086 more rows
## # i Use 'print(n = ...)' to see more rows

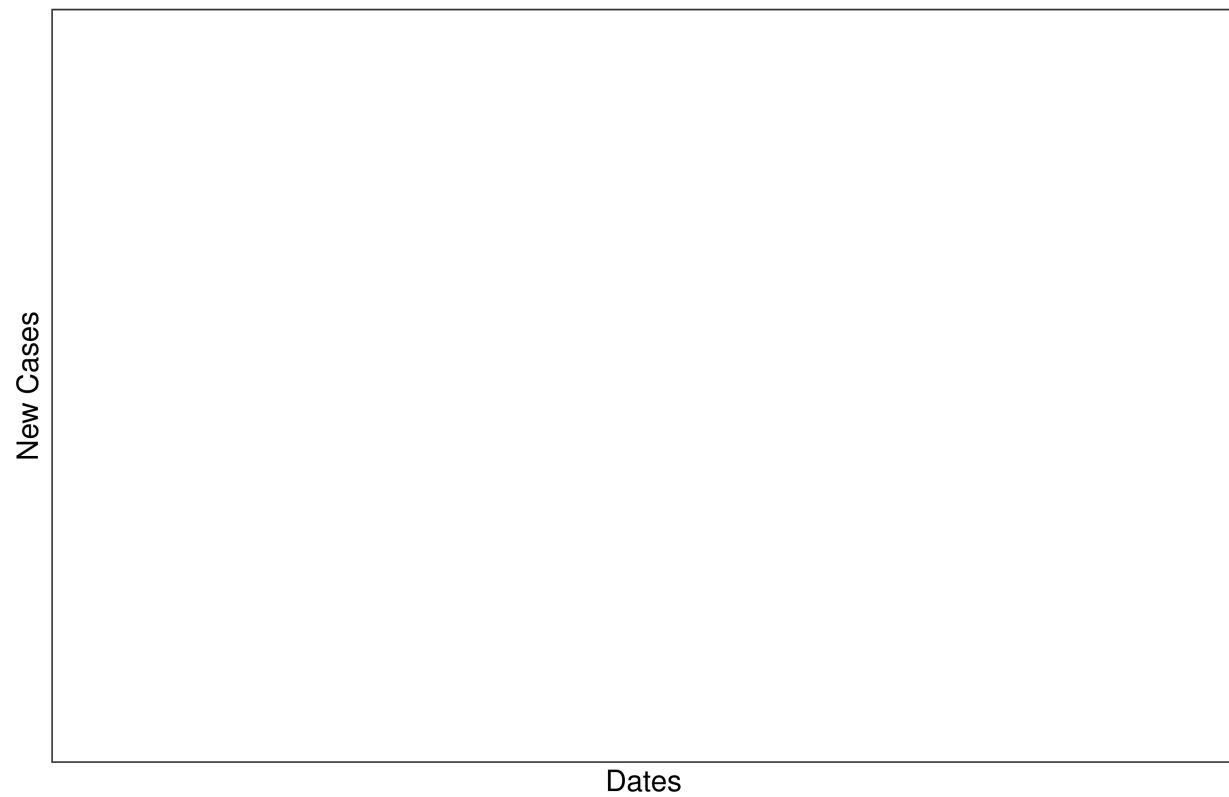
nj_counties <- nj_counties %>%
  mutate(new_cases = cases - lag(cases), new_deaths = deaths - lag(deaths)) %>%
  select(date, Admin2, cases, deaths, Population, new_cases, new_deaths)

nj_counties <- nj_counties %>%
  filter(new_cases >= 0, new_deaths >=0)

ggplot(nj_counties, aes(x=date)) +
  geom_line(aes(y = new_cases), color="blue4") +
  scale_x_date(date_labels = "%Y %b %d", date_breaks = "4 month") +
  theme_bw() +
  labs(x = "Dates",
       y = "New Cases",
       title = "NJ COVID19 Cases")

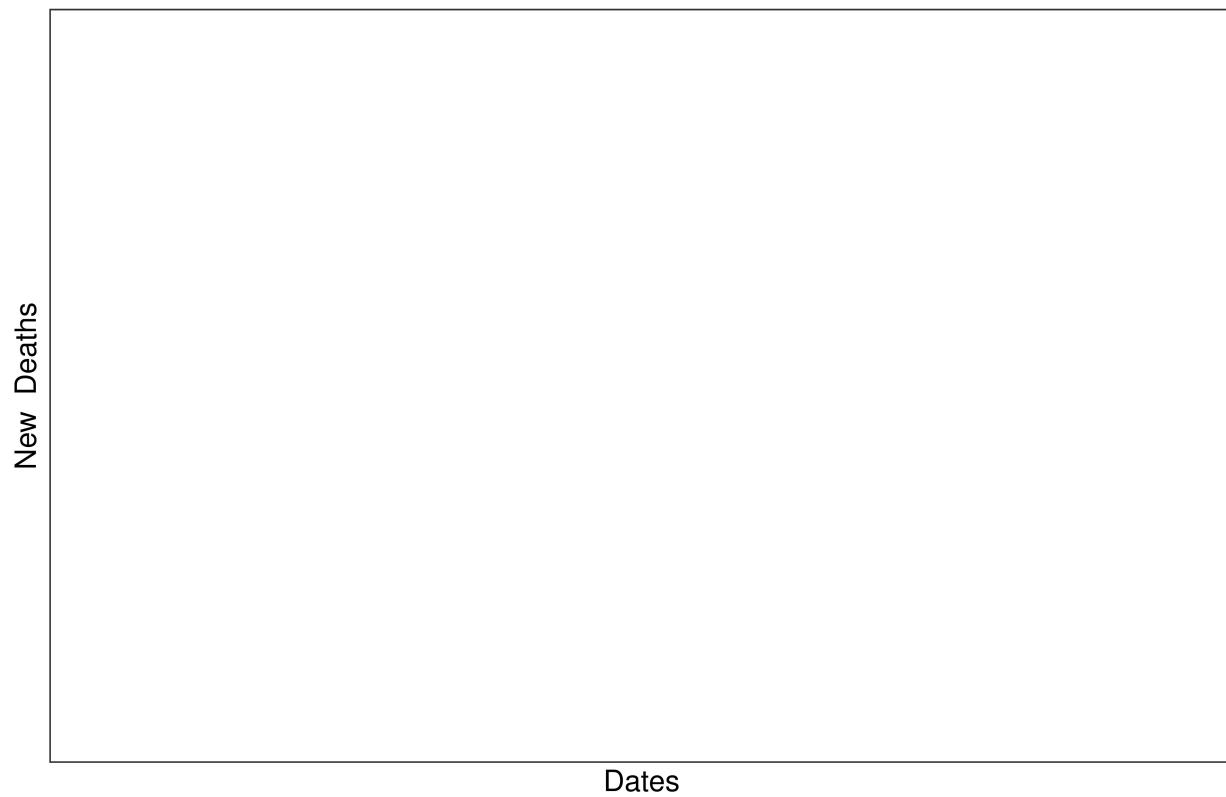
```

NJ COVID19 Cases



```
ggplot(nj_counties, aes(x=date)) +  
  geom_line(aes(y = new_deaths), color = "darkred") +  
  scale_x_date(date_labels = "%Y %b %d", date_breaks = "4 month") +  
  theme_bw() +  
  labs(x = "Dates",  
       y = "New Deaths",  
       title = "NJ COVID19 Deaths")
```

NJ COVID19 Deaths



##Project Step 4: Add Bias Identification The models or opinions during Covid gave several false conclusions due to inherent bias such as people of certain race/ethnicity/national origin being more susceptible to the disease, as time passed more analysis these were proven to be false.