title: “Peer-graded Assignment: Reproducible Report on COVID19 Data” author: “Mel Delgado” date: “2024-03-01” output: pdf_document: default html_document: default course: “dtsa-5301”

I will start by reading in the data from the four main csv files.

```
{r get_jhu_data} ## Get current Data in the four files # they all begin the same way
```

Get the needed library needed for analysis.

```
{r} library(tidyverse) library(lubridate)
```

Read the data ```{r} url_in <- “https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/ (https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/)”

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])

global_cases <- global_cases %>% pivot_longer(cols = -c(‘Province/State’, ‘Country/Region’, Lat, Long), names_to = “date”, values_to = “cases”) %>% select(-c(Lat,Long))

global_deaths <- global_deaths %>% pivot_longer(cols = -c(‘Province/State’, ‘Country/Region’, Lat, Long), names_to = “date”, values_to = “deaths”) %>% select(-c(Lat, Long))

Analyze te data

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

summary(global)

global <- global %>% filter(cases > 0)

summary(global)

global %>% filter(cases > 28000000)

US_cases

US_cases %>%
  pivot_longer(cols = -(UID:Combined_Key),
  names_to = "date",
  values_to = "cases")

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_cases

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)

US

global <- global %>%
  unite("Combined_Key",
  c(Province_State, Country_Region),
  sep = ", ",
```

```
  na.rm = TRUE,
  remove = FALSE)
```

Get additional data

```{r}
uid_lookup_url <- “https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/UID_ISO_FIPS_LookUp_Table.csv (https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/UID_ISO_FIPS_LookUp_Table.csv)” uid <- read_csv(uid_lookup_url) %>% select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

uid

uid <- read_csv(uid_lookup_url) %>% select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

uid

uid <- read_csv(uid_lookup_url) %>%select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

uid <- read_csv(uid_lookup_url) %>% select(-c(Lat, Long_, Combined_Key, code3, iso2, iso3, Admin2))

global <- global %>% left_join(uid, by = c(“Province_State”, “Country_Region”)) %>% select(-c(UID, FIPS)) %>% select(Province_State, Country_Region, date, cases, deaths, Population, Combined_Key)

US_by_state <- US %>% group_by(Province_State, Country_Region, date) %>% summarize(cases = sum(cases), deaths = sum(deaths), Population = sum(Population)) %>% mutate(deaths_per_mill = deaths *1000000 / Population) %>% select(Province_State, Country_Region, date, cases, deaths, deaths_per_mill, Population) %>% ungroup()

US_Totals <- US_by_state %>% group_by(Country_Region, date) %>% summarize(cases = sum(cases), deaths = sum(deaths), Population = sum(Population)) %>% mutate(deaths_per_mill = deaths *1000000 / Population) %>% select(Country_Region, date, cases, deaths, deaths_per_mill, Population) %>% ungroup()

state <- “New York”
```

```
## Visualize data

```{r}
US_Totals %>%
    filter(cases > 0) %>%
    ggplot(aes(x = date, y = cases)) +
    geom_line(aes(color = "cases")) +
    geom_point(aes(color = "cases")) +
    geom_line(aes(y = deaths, color = "deaths")) +
    geom_point(aes(y = deaths, color = "deaths")) +
    scale_y_log10() +
    theme(legend.position = "bottom",
        axis.text.x = element_text(angle=90)) +
    labs(title = "COVID19 in US", y = NULL)

max(US_Totals$date)

max(US_Totals$deaths)

US_by_state <- US_by_state %>%
    mutate(new_cases = cases - lag(cases),
        new_deaths = deaths - lag(deaths))

US_Totals <- US_Totals %>%
    mutate(new_cases = cases - lag(cases),
        new_deaths = deaths - lag(deaths))

tail(US_Totals)

tail(US_Totals %>% select(new_cases, new_deaths, everything()))

US_Totals %>%
    ggplot(aes(x = date, y = new_cases)) +
    geom_line(aes(color = "new_cases")) +
    geom_point(aes(color = "new_cases")) +
    geom_line(aes(y = new_deaths, color = "new_deaths")) +
    geom_point(aes(y = new_deaths, color = "new_deaths")) +
    scale_y_log10() +
    theme(legend.position = "bottom",
    axis.text.x = element_text(angle = 90)) +
    labs(title = "COVID19 in US", y = NULL)
```

Futher Analyze

```{r} US_state_totals <- US_by_state %>% + group_by(Province_State) %>% + summarize(deaths = max(deaths), cases = max(cases), + population = max(Population), + cases_per_thou = 1000* cases / population, + deaths_per_thou = 1000 * deaths / population) %>% + filter(cases > 0, population > 0)

US_state_totals %>% + slice_min(deaths_per_thou, n = 10) US_state_totals %>% + slice_min(deaths_per_thou, n = 10)

US_state_totals %>% + slice_min(deaths_per_thou, n = 10) %>% + select(deaths_per_thou, cases_per_thou, everything())

US_state_totals %>% + slice_min(deaths_per_thou, n = 10) %>% + select(deaths_per_thou, cases_per_thou, everything())

US_state_totals %>% + slice_max(deaths_per_thou, n = 10) %>% + select(deaths_per_thou, cases_per_thou, everything())

mod <- lm(deaths_per_thou ~ cases_per_thou, data = US_state_totals) summary(mod)

US_state_totals %>% slice_min(cases_per_thou)

US_state_totals %>% slice_max(cases_per_thou)

x_grid <- seq(1, 151)

new_df <- tibble(cases_per_thou = x_grid)

US_state_totals %>% mutate(pred = predict(mod))

us_tot_w_pred <- US_state_totals %>% mutate(pred = predict(mod))

us_tot_w_pred

us_tot_w_pred %>% ggplot() + + geom_point(aes(x = cases_per_thou, y = deaths_per_thou), color = “blue”) + + geom_point(aes(x = cases_per_thou, y = pred), color = “red”) ```