

WP7 Submission

Meghan Edgerton

2023-02-27

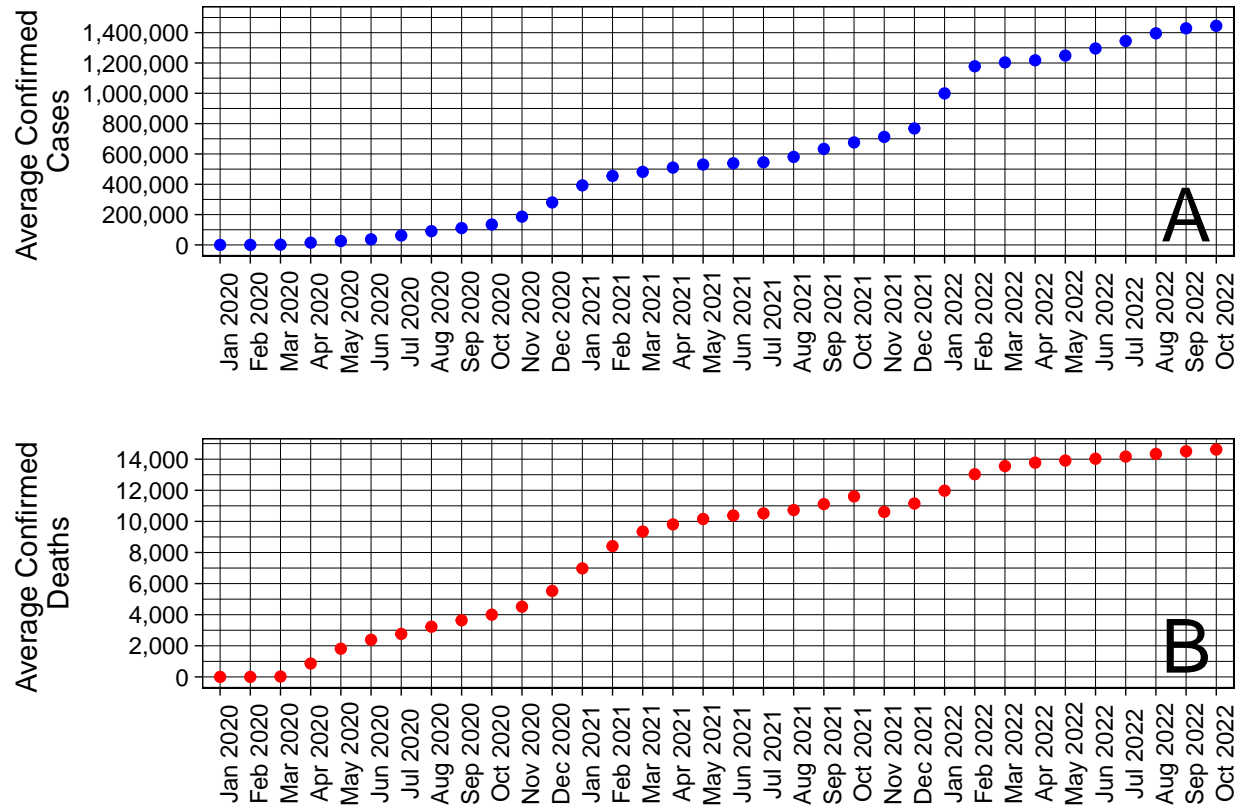


Figure 1. Two scatter plots displaying the quick increase in average number of confirmed cases (Plot A) of COVID-19 and the average number of confirmed deaths (Plot B) caused by COVID-19 in the United States from January 2020 to October of 2022. The number of cases and deaths are reported on a daily basis for each state in the United States by the Centers of Disease Control and Prevention (CDC). The data was pulled from the CDC's publicly available data repository and the average was calculated of the counts of confirmed cases and deaths for each month in the dataset's timespan.

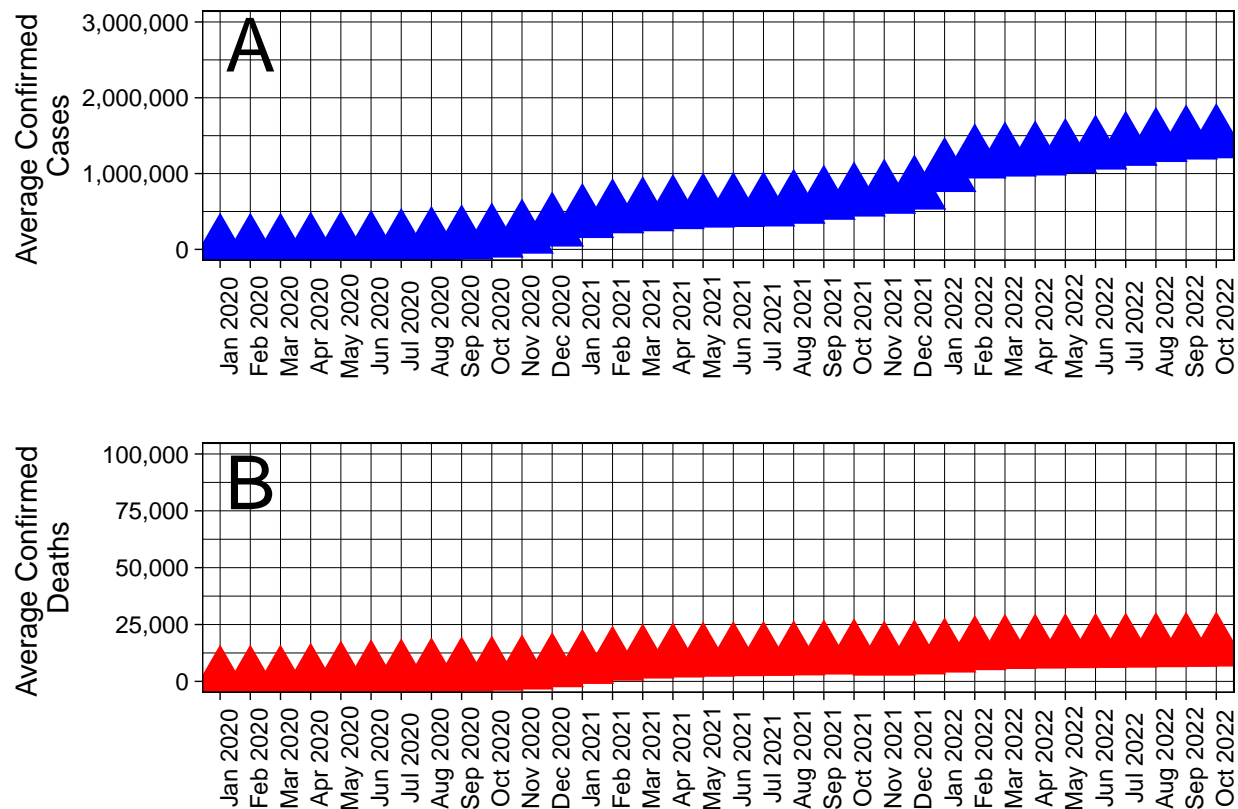


Figure 2. Two scatter plots displaying the average number of confirmed cases (Plot A) of COVID-19 and the average number of confirmed deaths (Plot B) caused by COVID-19 in the United States from January 2020 to October of 2022. Plot A shows a slight, steady increase in confirmed cases, while Plot B hardly shows an increase in confirmed deaths. The number of cases and deaths are reported on a daily basis for each state in the United States by the Centers of Disease Control and Prevention (CDC). The data was pulled from the CDC’s publicly available data repository and the average was calculated of the counts of confirmed cases and deaths for each month in the dataset’s timespan.

Data Source and Questions Being Answered

Weekly Plot 3: Advanced Plots

How was the United States affected by the COVID-19 Pandemic? Using the “United States COVID-19 Cases and Deaths by State Over Time” dataset provided by the Centers for Disease Control and Prevention (CDC), make a plot that answers the following questions:

- 1) How did the number of cases in the United States change throughout the months and years of the COVID-19 Pandemic?
- 2) How did the number of deaths in the United States change throughout the months and years of the COVID-19 Pandemic?

Data Source Link: <https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36>

Data Description: This is an archived dataset made publicly available by the Centers for Disease Control and Prevention (CDC) that consists of daily counts by state of confirmed COVID-19 cases and deaths in the United States as recorded from January 2020 to October 2022. This dataset is archived (meaning that it is no longer subject to updates and changes) due to the CDC beginning to aggregate this data on a weekly basis rather than a daily basis as of October 20th, 2022. The data is verified and reported to the CDC by state, local, and territorial public health departments. The purpose of recording this data is to track trends of the COVID-19 Pandemic, detect outbreaks, and monitor whether public health measures that were put into place were working.

R Code Used To Make These Plots

```
library(dplyr)
library(ggplot2)
library(rmarkdown)
library(gridExtra)
library(scales)
library(readxl)

options(scipen = 9999)

path <- "/Users/meghanedgerton/Documents/DataVizINF0526/COVID19Data.xlsx"

data <- read_xlsx(path)
```

```
avgs <- data %>% group_by(YYYYMM) %>%
  summarize(avgcases = mean(conf_cases, na.rm = TRUE),
            avgdeaths = mean(conf_death, na.rm = TRUE),
            avgnewcases = mean(new_case, na.rm = TRUE),
            avgnewdeaths = mean(new_death, na.rm = TRUE))
```

Rounding down to the nearest integer for simplicity

```
avgs$avgcases <- floor(avgs$avgcases)
avgs$avgdeaths <- floor(avgs$avgdeaths)
```

```
xlabels <- c("Jan 2020", "Feb 2020", "Mar 2020", "Apr 2020", "May 2020", "Jun 2020",
             "Jul 2020", "Aug 2020", "Sep 2020", "Oct 2020", "Nov 2020", "Dec 2020",
             "Jan 2021", "Feb 2021", "Mar 2021", "Apr 2021", "May 2021", "Jun 2021",
             "Jul 2021", "Aug 2021", "Sep 2021", "Oct 2021", "Nov 2021", "Dec 2021",
             "Jan 2022", "Feb 2022", "Mar 2022", "Apr 2022", "May 2022", "Jun 2022",
             "Jul 2022", "Aug 2022", "Sep 2022", "Oct 2022")
```

Average Count of Confirmed Cases in the US

```
p1 <- ggplot(avgs, aes(x = YYYYMM, y = avgcases)) +
  geom_point(color = "blue") +
  theme_linedraw() +
  scale_x_discrete(labels = xlabels) +
  scale_y_continuous(limits = c(0, 1500000), n.breaks = 10, labels = comma) +
  theme(axis.text.x = element_text(angle = 90),
        panel.grid.major = element_line(colour = "black", linewidth = 0.1),
        panel.grid.minor = element_line(colour = "black", linewidth = 0.1)) +
  labs(x = "", y = "Average Confirmed \nCases") +
  annotate("text", label = "A", x = 33, y = 200000, size = 10)
```

```

# Average Count of Confirmed Deaths in the US
p2 <- ggplot(avgs, aes(x = YYYYMM, y = avgdeaths)) +
  geom_point(color = "red") +
  theme_linedraw() +
  scale_x_discrete(labels = xlabels) +
  scale_y_continuous(n.breaks = 10, labels = comma) +
  theme(axis.text.x = element_text(angle=90),
        panel.grid.major = element_line(colour = "black", linewidth = 0.1),
        panel.grid.minor = element_line(colour = "black", linewidth = 0.1)) +
  labs(x = "", y = "Average Confirmed \nDeaths") +
  annotate("text", label = "B", x = 33, y = 2000, size = 10)

```

```

# Stacking Plots
a <- ggplotGrob(p1)
b <- ggplotGrob(p2)
grid::grid.newpage()
grid::grid.draw(rbind(a, b))

```

```

# Illusion Plots-----

p3 <- ggplot(avgs, aes(x = YYYYMM, y = avgcases)) +
  geom_point(color = "blue", size = 8, shape = 17) +
  theme_linedraw() +
  scale_x_discrete(labels = xlabels) +
  scale_y_continuous(limits= c(0,3000000), labels = comma) +
  theme(axis.text.x = element_text(angle=90),
        panel.grid.major = element_line(colour = "black", linewidth = 0.1),
        panel.grid.minor = element_line(colour = "black", linewidth = 0.1)) +
  labs(x = "", y = "Average Confirmed \nCases") +
  annotate("text", label = "A", x = 2, y = 2700000, size = 10)

```

```

p4 <- ggplot(avgs, aes(x = YYYYMM, y = avgdeaths)) +
  geom_point(color = "red", size = 8, shape = 17) +
  theme_linedraw() +
  scale_x_discrete(labels = xlabels) +
  scale_y_continuous(limits= c(0,100000), labels = comma) +
  theme(axis.text.x = element_text(angle=90),
        panel.grid.major = element_line(colour = "black", linewidth = 0.1),
        panel.grid.minor = element_line(colour = "black", linewidth = 0.1)) +
  labs(x = "", y = "Average Confirmed \nDeaths") +
  annotate("text", label = "B", x = 2, y = 87500, size = 10)

```

```

# Stacking Plots
c <- ggplotGrob(p3)
d <- ggplotGrob(p4)
grid::grid.newpage()
grid::grid.draw(rbind(c, d))

```