Analysis

Amos Apreku

```
library(tidyverse)
library(dplyr)
library(ggrepel)
```

```
# Reading in our dat data

dat <- readRDS("../data/dat.rds")
```

```
colnames(dat)
```

```
 [1] "date"            "mmwr_year"       "mmwr_week"       "state"
 [5] "state_name"      "population"      "region"          "region_name"
 [9] "cases"           "hosp"            "deaths"          "series_complete"
[13] "booster"
```

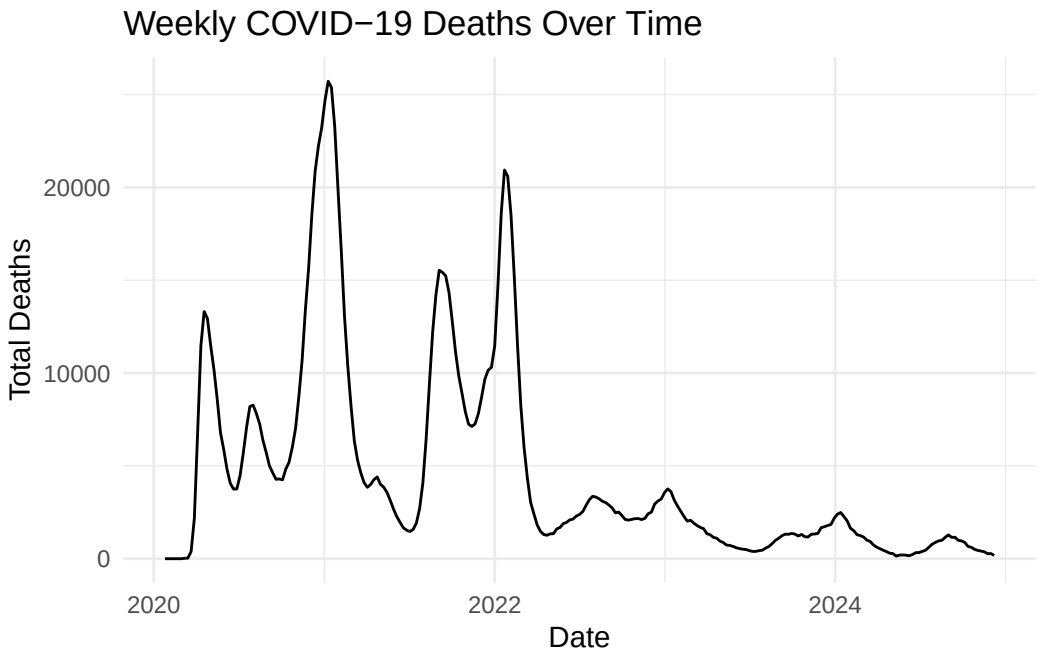
```
head(dat)
```

```
        date mmwr_year mmwr_week state state_name population region
1 2020-01-25      2020         4    AK     Alaska     732441     10
2 2020-02-01      2020         5    AK     Alaska     732441     10
3 2020-02-08      2020         6    AK     Alaska     732441     10
4 2020-02-15      2020         7    AK     Alaska     732441     10
5 2020-02-22      2020         8    AK     Alaska     732441     10
6 2020-02-29      2020         9    AK     Alaska     732441     10
        region_name cases hosp deaths series_complete booster
1 Pacific Northwest     0   NA      0              NA      NA
2 Pacific Northwest     0   NA      0              NA      NA
3 Pacific Northwest     0   NA      0              NA      NA
4 Pacific Northwest     0   NA      0              NA      NA
5 Pacific Northwest     0   NA      0              NA      NA
6 Pacific Northwest     0   NA      0              NA      NA
```

```r
str(dat)
```

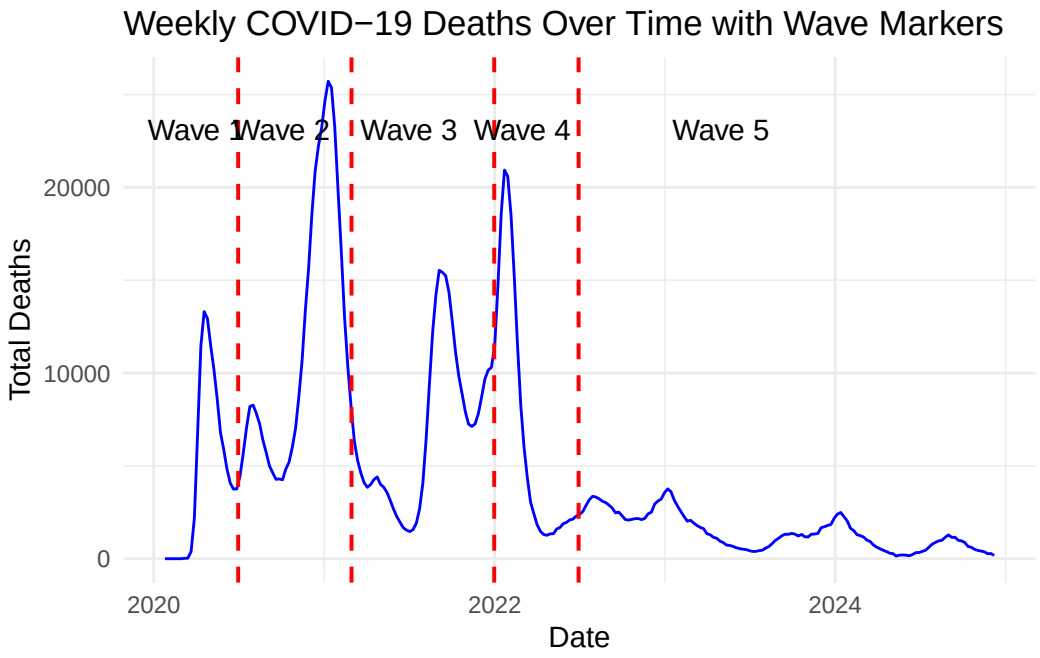
```
'data.frame':   13260 obs. of  13 variables:
 $ date           : Date, format: "2020-01-25" "2020-02-01" ...
 $ mmwr_year      : num  2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...
 $ mmwr_week      : num  4 5 6 7 8 9 10 11 12 13 ...
 $ state          : chr  "AK" "AK" "AK" "AK" ...
 $ state_name     : chr  "Alaska" "Alaska" "Alaska" "Alaska" ...
 $ population     : num  732441 732441 732441 732441 732441 ...
 $ region         : int  10 10 10 10 10 10 10 10 10 10 ...
 $ region_name    : chr  "Pacific Northwest" "Pacific Northwest" "Pacific Northwest" "Pacifi
 $ cases          : num  0 0 0 0 0 0 0 0 11 52 ...
 $ hosp           : num  NA NA NA NA NA NA NA NA NA NA ...
 $ deaths         : num  0 0 0 0 0 0 0 0 0 NA ...
 $ series_complete: num  NA NA NA NA NA NA NA NA NA NA ...
 $ booster        : num  NA NA NA NA NA NA NA NA NA NA ...
```

```r
library(ggplot2)
library(dplyr)
library(lubridate)
dat_weekly <- dat %>%
  group_by(date) %>%
  summarize(total_deaths = sum(deaths, na.rm = TRUE))
ggplot(dat_weekly, aes(x = date, y = total_deaths)) +
  geom_line() +
  labs(title = "Weekly COVID-19 Deaths Over Time",
       x = "Date",
       y = "Total Deaths") +
  theme_minimal()
```




```
library(ggplot2)
wave_periods <- data.frame(
  wave = c("Wave 1", "Wave 2", "Wave 3", "Wave 4", "Wave 5"),
  start = as.Date(c("2020-01-01", "2020-07-01", "2021-03-01", "2022-01-01", "2022-07-01")),
  end = as.Date(c("2020-06-30", "2021-02-28", "2021-12-31", "2022-06-30", "2024-12-31"))
)
ggplot(dat_weekly, aes(x = date, y = total_deaths)) +
  geom_line(color = "blue") +
  geom_vline(xintercept = as.Date(c("2020-06-30", "2021-02-28", "2021-12-31", "2022-06-30")),
             linetype = "dashed", color = "red", size = 0.7) +
  annotate("text", x = as.Date("2020-04-01"), y = max(dat_weekly$total_deaths, na.rm = TRUE)
  annotate("text", x = as.Date("2020-10-01"), y = max(dat_weekly$total_deaths, na.rm = TRUE)
  annotate("text", x = as.Date("2021-07-01"), y = max(dat_weekly$total_deaths, na.rm = TRUE)
  annotate("text", x = as.Date("2022-03-01"), y = max(dat_weekly$total_deaths, na.rm = TRUE)
  annotate("text", x = as.Date("2023-05-01"), y = max(dat_weekly$total_deaths, na.rm = TRUE)
  labs(title = "Weekly COVID-19 Deaths Over Time with Wave Markers",
       x = "Date",
       y = "Total Deaths") +
  theme_minimal()
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```




```r
# Define the wave periods
dat <- dat %>%
  mutate(wave = case_when(
    date >= as.Date("2020-01-01") & date <= as.Date("2020-06-30") ~ "Wave 1",
    date >= as.Date("2020-07-01") & date <= as.Date("2021-02-28") ~ "Wave 2",
    date >= as.Date("2021-03-01") & date <= as.Date("2021-12-31") ~ "Wave 3",
    date >= as.Date("2022-01-01") & date <= as.Date("2022-06-30") ~ "Wave 4",
    date >= as.Date("2022-07-01") & date <= as.Date("2024-12-31") ~ "Wave 5",
    TRUE ~ NA_character_
  ))

# Calculate death rates by state and wave
death_rates <- dat %>%
  group_by(state_name, wave) %>%
  summarize(
    total_deaths = sum(deaths, na.rm = TRUE),
    population = max(population, na.rm = TRUE),
    .groups = "drop" # Explicitly drop grouping after summarization
  ) %>%
  mutate(death_rate_per_100k = (total_deaths / population) * 100000)
```

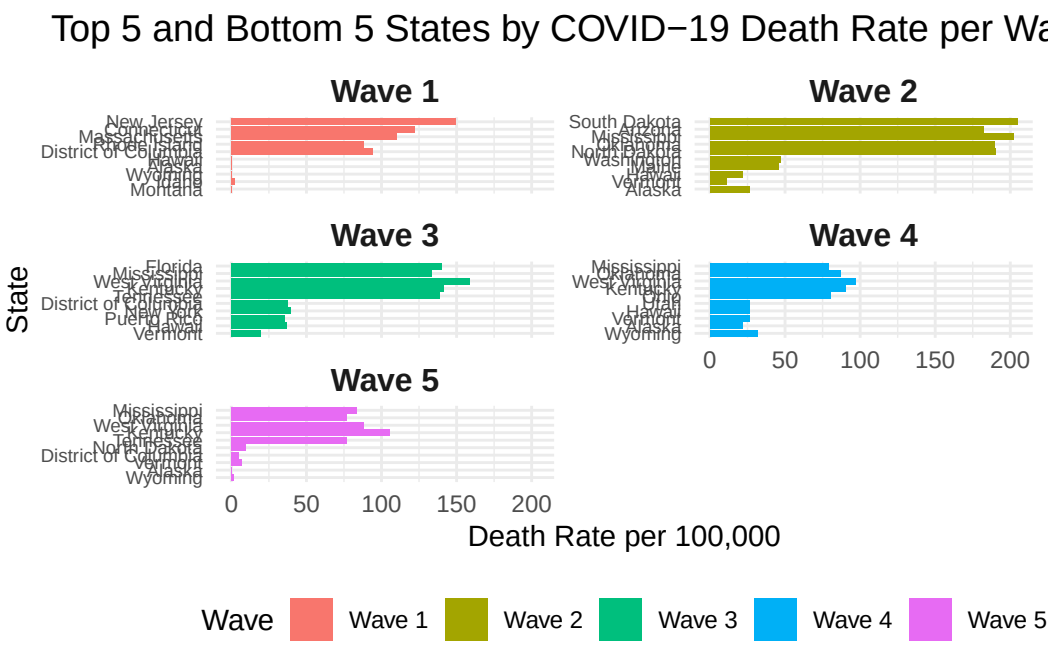


Figure 1: Figure 2: Top 5 and Bottom 5 States by COVID-19 Death Rate per 100,000 Population Across Waves

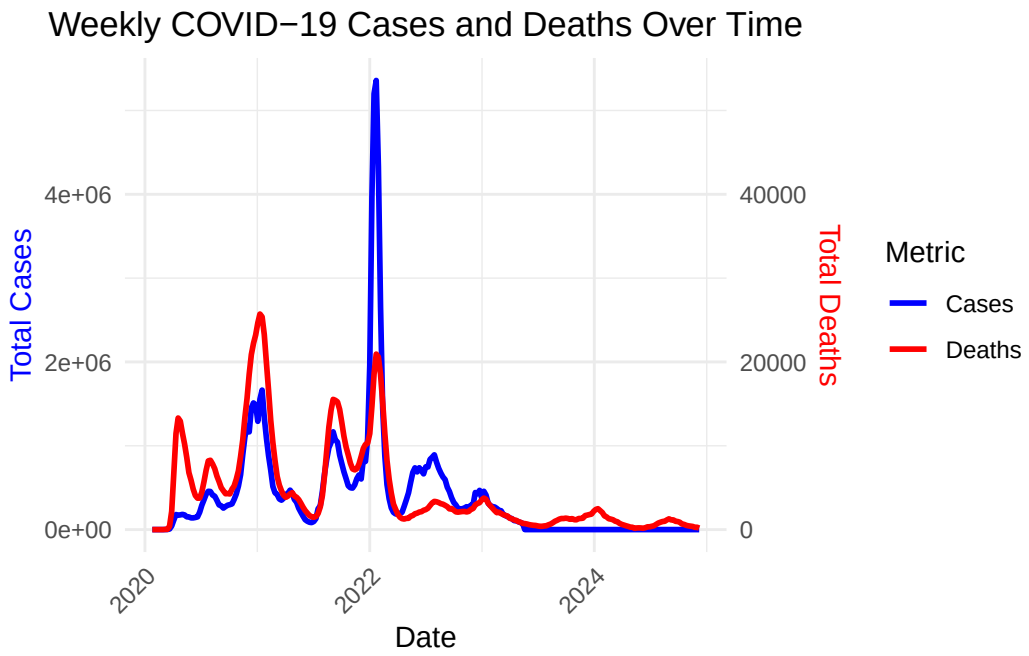


Figure 2: Figure 3: Weekly COVID-19 Cases and Deaths Over Time

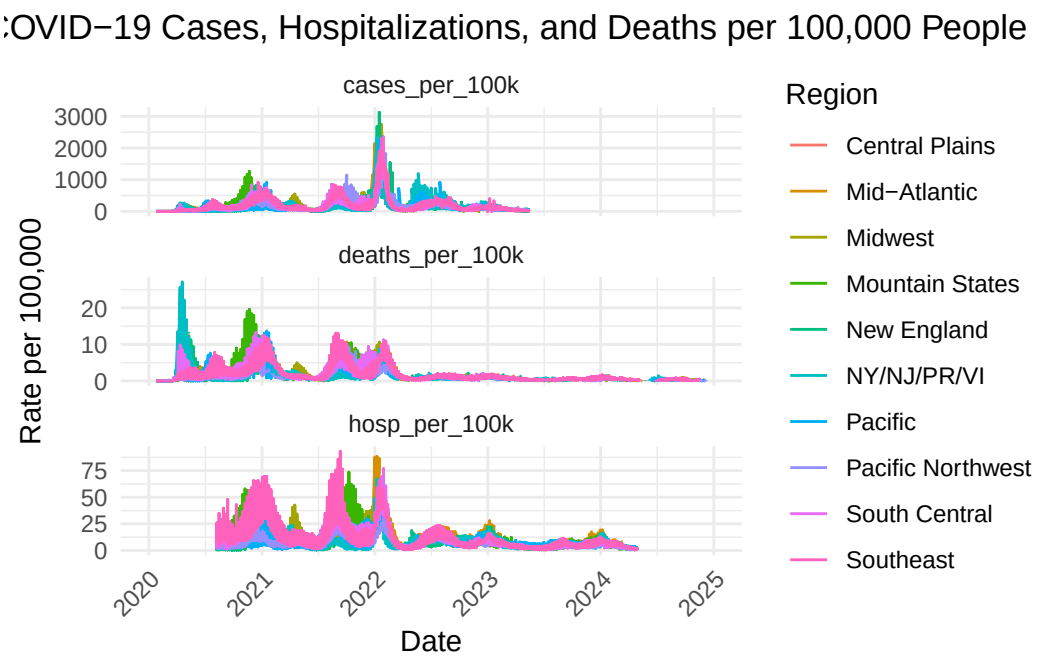


Figure 3: Figure 4: Trends in COVID-19 Cases, Hospitalizations, and Deaths per 100,000

```r
# Define the wave periods
dat <- dat %>%
  mutate(wave = case_when(
    date >= as.Date("2020-01-01") & date <= as.Date("2020-06-30") ~ "Wave 1",
    date >= as.Date("2020-07-01") & date <= as.Date("2021-02-28") ~ "Wave 2",
    date >= as.Date("2021-03-01") & date <= as.Date("2021-12-31") ~ "Wave 3",
    date >= as.Date("2022-01-01") & date <= as.Date("2022-06-30") ~ "Wave 4",
    date >= as.Date("2022-07-01") & date <= as.Date("2024-12-31") ~ "Wave 5",
    TRUE ~ NA_character_
  ))

# Calculate death rates by state and wave
death_rates <- dat %>%
  group_by(state_name, wave) %>%
  summarize(
    total_deaths = sum(deaths, na.rm = TRUE),
    population = max(population, na.rm = TRUE)
  ) %>%
  mutate(death_rate_per_100k = (total_deaths / population) * 100000)
```

```
`summarise()` has grouped output by 'state_name'. You can override using the
```

`.groups` argument.

```
# View the death rates
print(death_rates)
```

```
# A tibble: 260 x 5
# Groups:   state_name [52]
   state_name wave   total_deaths population death_rate_per_100k
   <chr>      <chr>         <dbl>      <dbl>                <dbl>
 1 Alabama    Wave 1         1168    5024803                 23.2
 2 Alabama    Wave 2         8905    5039877                177.
 3 Alabama    Wave 3         6273    5039877                124.
 4 Alabama    Wave 4         3184    5073903                 62.8
 5 Alabama    Wave 5         2911    5108468                 57.0
 6 Alaska     Wave 1            0     732441                  0
 7 Alaska     Wave 2          194     732673                 26.5
 8 Alaska     Wave 3          713     732673                 97.3
 9 Alaska     Wave 4          159     733276                 21.7
10 Alaska     Wave 5            0     733406                  0
# i 250 more rows
```

Show 10 entries Search:

Table 1: COVID-19 Total Cases, Deaths, Hospitalizations, and Rates by State

	State	Total Cases	Total Deaths	Total Hospitalizations	Case Rate (per 100k)
1	Alaska	297588	1066	10553	159
2	Rhode Island	441466	3641	14945	158
3	Kentucky	1743117	21936	165575	151
4	North Dakota	291093	2605	16657	146
5	West Virginia	650556	8621	47718	144
6	Tennessee	2542163	32182	142017	142
7	South Carolina	1852019	21979	104351	138
8	Puerto Rico	1122076	6885	32013	136
9	Wisconsin	2030717	18018	137940	135
10	Florida	7572282	86809	610281	134

Showing 1 to 10 of 52 entries Previous 1 2 3 4 5 6 Next

Show 10 entries

Search:

Table 2: COVID-19 Total Cases, Deaths, Hospitalizations, and Rates by Region

	Region	Total Cases	Total Deaths	Total Hospitalizations	Case Rate (per 100k)
1	Southeast	22952265	276458	1652525	131
2	Mountain States	3976119	32650	240836	124
3	Midwest	16587293	195735	1128868	123
4	Pacific	15975252	160922	930936	123
5	New England	4525215	45697	251914	117
6	South Central	13057010	168526	977866	117
7	Central Plains	4220756	52357	301693	116
8	Mid-Atlantic	8412794	114334	602312	105
9	Pacific Northwest	3757321	32907	179679	100
10	NY/NJ/PR/VI	7769390	89370	646881	94

Showing 1 to 10 of 10 entries

Previous 1 Next