COVID 19 Analysis

Tyler Kephart

2024-08-19

Required Packages

```
library(tidyverse)
library(lubridate)
library(usmap)
```

Part 1 - Basic Exploration of US Data

The New York Times (the Times) has aggregated reported COVID-19 data from state and local governments and health departments since 2020 and provides public access through a repository on GitHub. One of the data sets provided by the Times is county-level data for cumulative cases and deaths each day. This will be your primary data set for the first two parts of your analysis.

County-level COVID data from 2020, 2021, and 2022 has been imported below. Each row of data reports the cumulative number of cases and deaths for a specific county each day. A FIPS code, a standard geographic identifier, is also provided which you will use in Part 2 to construct a map visualization at the county level for a state.

Additionally, county-level population estimates reported by the US Census Bureau has been imported as well. You will use these estimates to calculate statistics per 100,000 people.

```
# Import New York Times COVID-19 data
# Import Population Estimates from US Census Bureau

us_counties_2020 <- read_csv(
    "https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2020.csv")
us_counties_2021 <- read_csv(
    "https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2021.csv")
us_counties_2022 <- read_csv(
    "https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2022.csv")

us_population_estimates <- read_csv("fips_population_estimates.csv")
```

Question 1

Your first task is to combine and tidy the 2020, 2021, and 2022 COVID data sets and find the total deaths and cases for each day since March 15, 2020 (2020-03-15). The data sets provided from the NY Times also includes statistics from Puerto Rico, a US territory. You may remove these observations from the data as they will not be needed for your analysis. Once you have tidied the data, find the total COVID-19 cases and deaths since March 15, 2020. Write a sentence or two after the code block communicating your results. Use inline code to include the `max_date`, `us_total_cases`, and `us_total_deaths` variables. To write inline code use `r`.

```
# combine the 2020, 2021, and 2022 COVID data sets
us_counties <- rbind(
                    us_counties_2020,
```

```
                    us_counties_2021,
                    us_counties_2022,
                    deparse.level = 1) %>%
                # exclude Puerto Rico, idk why other territories get to stay
                filter(state != "Puerto Rico",
                       # filter from March 15, 2020 and after
                       date >= as.Date("2020-03-15"))
# summarize total cases and deaths by date
daily_totals <- us_counties %>%
                    group_by(date) %>%
                    summarise(total_deaths = sum(deaths),
                              total_cases = sum(cases))
daily_totals
```

```
## # A tibble: 1,022 x 3
##    date       total_deaths total_cases
##    <date>            <dbl>       <dbl>
##  1 2020-03-15           68        3595
##  2 2020-03-16           91        4502
##  3 2020-03-17          117        5901
##  4 2020-03-18          162        8345
##  5 2020-03-19          212       12387
##  6 2020-03-20          277       17998
##  7 2020-03-21          359       24507
##  8 2020-03-22          457       33050
##  9 2020-03-23          577       43474
## 10 2020-03-24          783       53899
## # i 1,012 more rows
```

```
# set variables for inline code usage
max_date <- max(daily_totals$date)
us_total_cases <- max(daily_totals$total_cases)
us_total_deaths <- max(daily_totals$total_deaths)
```

```
# Your output should look similar to the following tibble:
#
#   A tibble: 657 x 3
#       date          total_deaths   total_cases
#      <date>            <dbl>         <dbl>
#   1 2020-03-15           68          3595
#   2 2020-03-16           91          4502
#   3 2020-03-17          117          5901
#   4 2020-03-18          162          8345
#   5 2020-03-19          212         12387
#   6 2020-03-20          277         17998
#   7 2020-03-21          359         24507
#   8 2020-03-22          457         33050
#   9 2020-03-23          577         43474
#  10 2020-03-24          783         53899
# ... with 647 more rows
```

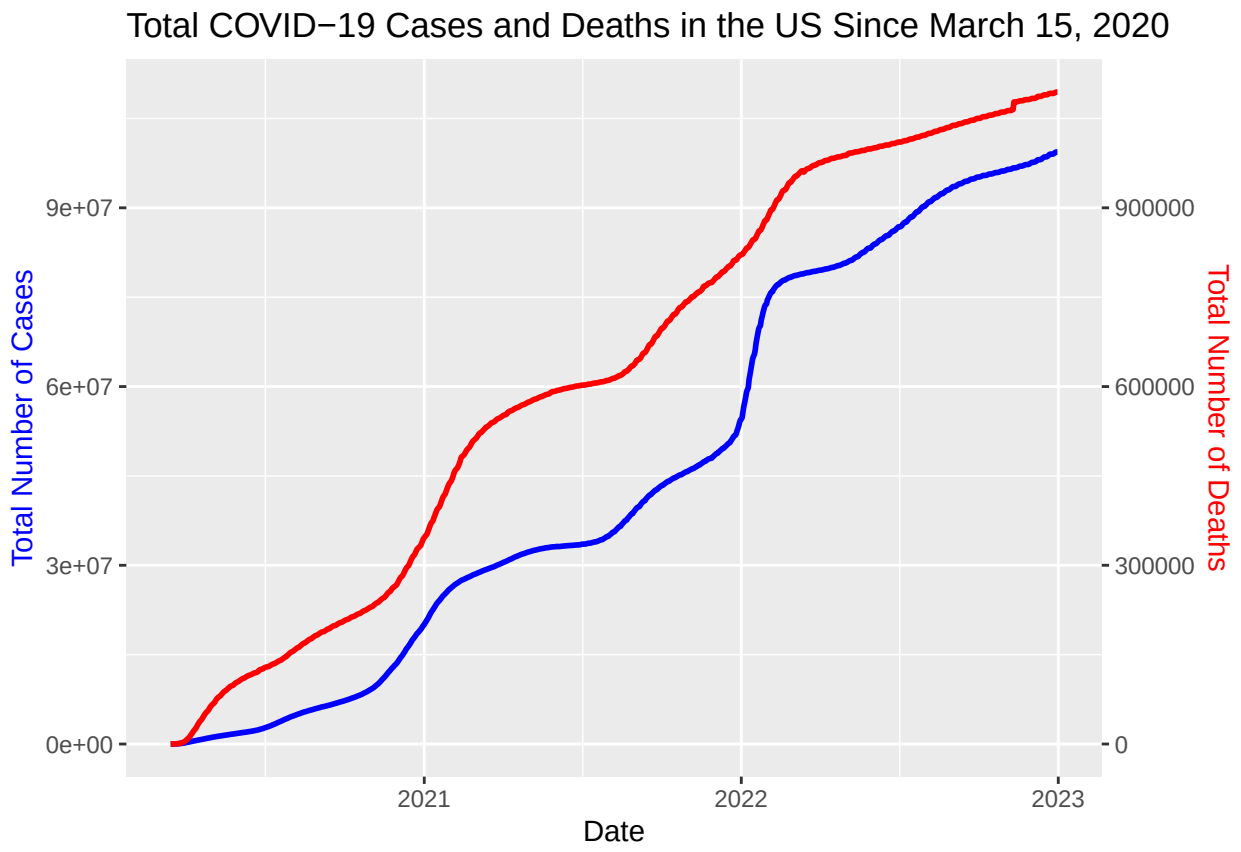
As of December 31, 2022, there was more than 99.37 million cases and 1.09 million deaths from COVID-19 in the United States (excluding Puerto Rico).

Question 2

Create a visualization for the total number of deaths and cases in the US since March 15, 2020. Before you create your visualization, review the types of plots you can create using the ggplot2 library and think about which plots would be

effective in communicating your results. After you have created your visualization, write a few sentences describing your visualization. How could the plot be interpreted? Could it be misleading?

```
# plot for the total number of US cases and deaths since March 15, 2020
ggplot(daily_totals, aes(x = date)) +
    geom_line(aes(y = total_cases, color = "Cases"), linewidth = 1) +
    # scaling deaths for visibility
    geom_line(aes(y = total_deaths * 100, color = "Deaths"), linewidth = 1) +
    # set up y axis for 2 variables
    scale_y_continuous(name = "Total Number of Cases",
                       # add second axis for deaths
                       sec.axis = sec_axis(~./100, name = "Total Number of Deaths")) +
    # color the y axis titles to match the plotted lines
    theme(axis.title.y.left = element_text(color = "blue"),
          axis.title.y.right = element_text(color = "red")) +
    labs(title = "Total COVID-19 Cases and Deaths in the US Since March 15, 2020",
        x = "Date") +
    # remove legend name
    theme(legend.position = "none") +
    scale_color_manual(values = c("Cases" = "blue", "Deaths" = "red"))
```



There seems to have been a spike in cases early 2022 that was possibly due to increased travel during the holiday season in the United States.

Question 3

While it is important to know the total deaths and cases throughout the COVID-19 pandemic, it is also important for local and state health officials to know the the number of new cases and deaths each day to understand how rapidly the virus is spreading. Using the table you created in Question 1, calculate the number of new deaths and cases each day and a seven-day average of new deaths and cases. Once you have organized your data, find the days that saw the largest number of new cases and deaths. Write a sentence or two after the code block communicating your results.

```r
# calculate number of new deaths and cases each day
daily_7day_avg <- daily_totals %>%
    mutate(
        delta_deaths_1 = total_deaths - lag(total_deaths),
        delta_cases_1 = total_cases - lag(total_cases))
# calculate rolling averages
calculate_rolling_average <- function(x, window_size) {
    sapply(seq_along(x), function(i) {
        if (i < window_size) return(NA)
        mean(x[(i - window_size + 1):i])
    })
}
# add 7-day rolling averages
daily_7day_avg <- daily_7day_avg %>%
    mutate(
        delta_deaths_7 = calculate_rolling_average(delta_deaths_1, 7),
        delta_cases_7 = calculate_rolling_average(delta_cases_1, 7)
    )
daily_7day_avg
```

```
## # A tibble: 1,022 x 7
##    date       total_deaths total_cases delta_deaths_1 delta_cases_1
##    <date>            <dbl>       <dbl>          <dbl>         <dbl>
##  1 2020-03-15           68        3595             NA            NA
##  2 2020-03-16           91        4502             23           907
##  3 2020-03-17          117        5901             26          1399
##  4 2020-03-18          162        8345             45          2444
##  5 2020-03-19          212       12387             50          4042
##  6 2020-03-20          277       17998             65          5611
##  7 2020-03-21          359       24507             82          6509
##  8 2020-03-22          457       33050             98          8543
##  9 2020-03-23          577       43474            120         10424
## 10 2020-03-24          783       53899            206         10425
## # i 1,012 more rows
## # i 2 more variables: delta_deaths_7 <dbl>, delta_cases_7 <dbl>
```

```r
# find the days with the largest number of new cases and deaths
max_new_cases_date <- daily_7day_avg %>%
    slice_max(order_by = delta_cases_1, n = 1)
max_new_deaths_date <- daily_7day_avg %>%
    slice_max(order_by = delta_deaths_1, n = 1)
```

```r
#  Your output should look similar to the following tibble:
#
#  date
#  total_deaths     >  the cumulative number of deaths up to and including the associated date
#  total_cases      >  the cumulative number of cases up to and including the associated date
#  delta_deaths_1   >  the number of new deaths since the previous day
#  delta_cases_1    >  the number of new cases since the previous day
#  delta_deaths_7   >  the average number of deaths in a seven-day period
#  delta_cases_7    >  the average number of cases in a seven-day period
#==
#  A tibble: 813 x 7
#    date          total_deaths   total_cases   delta_deaths_1    delta_cases_1  delta_deaths_7  delta_cases
#    <date>           <dbl>          <dbl>         <dbl>              <dbl>        <dbl>            <dbl>
#  1 2020-03-15          68          3600           0                    0         NA               NA
#  2 2020-03-16          91          4507          23                  907         NA               NA
#  3 2020-03-17         117          5906          26                 1399         NA               NA
```

```
#  4 2020-03-18          162          8350           45                2444        NA                NA
#  5 2020-03-19          212         12393           50                4043        NA                NA
#  6 2020-03-20          277         18012           65                5619        NA                NA
#  7 2020-03-21          360         24528           83                6516        NA                NA
#  8 2020-03-22          458         33073           98                8545      55.7              4210.
#  9 2020-03-23          579         43505          121               10432      69.7              5571.
# 10 2020-03-24          785         53938          206               10433      95.4              6862.
# ... with 803 more rows
```

As of December 31, 2022, the most amount of new cases in a single day happened on January 10, 2022 with 1.427097×10^6 cases. The most amount of new deaths in a single day happened on November 11, 2022 with 1.0037×10^4 deaths.

Question 4

Create a new table, based on the table from Question 3, and calculate the number of new deaths and cases per 100,000 people each day and a seven day average of new deaths and cases per 100,000 people.

```
# aggregate US population by year
us_population_totals <- us_population_estimates %>%
    group_by(Year) %>%
    summarize(total_population = sum(Estimate))

# add year to daily_totals
daily_7day_avg_yr <- daily_7day_avg %>%
    mutate(year = year(date))

# join with population totals
daily_7day_avg_yr_pop <- daily_7day_avg_yr %>%
    left_join(us_population_totals, by = c("year" = "Year"))

# calculate metrics per 100,000 people
daily_7day_avg_per_100k <- daily_7day_avg_yr_pop %>%
    mutate(
        total_deaths = (total_deaths / total_population) * 100000,
        total_cases = (total_cases / total_population) * 100000,
        delta_deaths_1 = (delta_deaths_1 / total_population) * 100000,
        delta_cases_1 = (delta_cases_1 / total_population) * 100000,
        delta_deaths_7 = calculate_rolling_average(delta_deaths_1, 7),
        delta_cases_7 = calculate_rolling_average(delta_cases_1, 7)
    ) %>%
    select(date,
            total_deaths,
            total_cases,
            delta_deaths_1,
            delta_cases_1,
            delta_deaths_7,
            delta_cases_7)
daily_7day_avg_per_100k
```

```
## # A tibble: 1,022 x 7
##    date       total_deaths total_cases delta_deaths_1 delta_cases_1
##    <date>            <dbl>       <dbl>          <dbl>         <dbl>
##  1 2020-03-15       0.0205        1.08       NA              NA
##  2 2020-03-16       0.0275        1.36        0.00694         0.274
##  3 2020-03-17       0.0353        1.78        0.00784         0.422
##  4 2020-03-18       0.0489        2.52        0.0136          0.737
##  5 2020-03-19       0.0640        3.74        0.0151          1.22
```

```
##  6 2020-03-20       0.0836        5.43        0.0196          1.69
##  7 2020-03-21       0.108         7.39        0.0247          1.96
##  8 2020-03-22       0.138         9.97        0.0296          2.58
##  9 2020-03-23       0.174        13.1         0.0362          3.14
## 10 2020-03-24       0.236        16.3         0.0621          3.14
## # i 1,012 more rows
## # i 2 more variables: delta_deaths_7 <dbl>, delta_cases_7 <dbl>
```

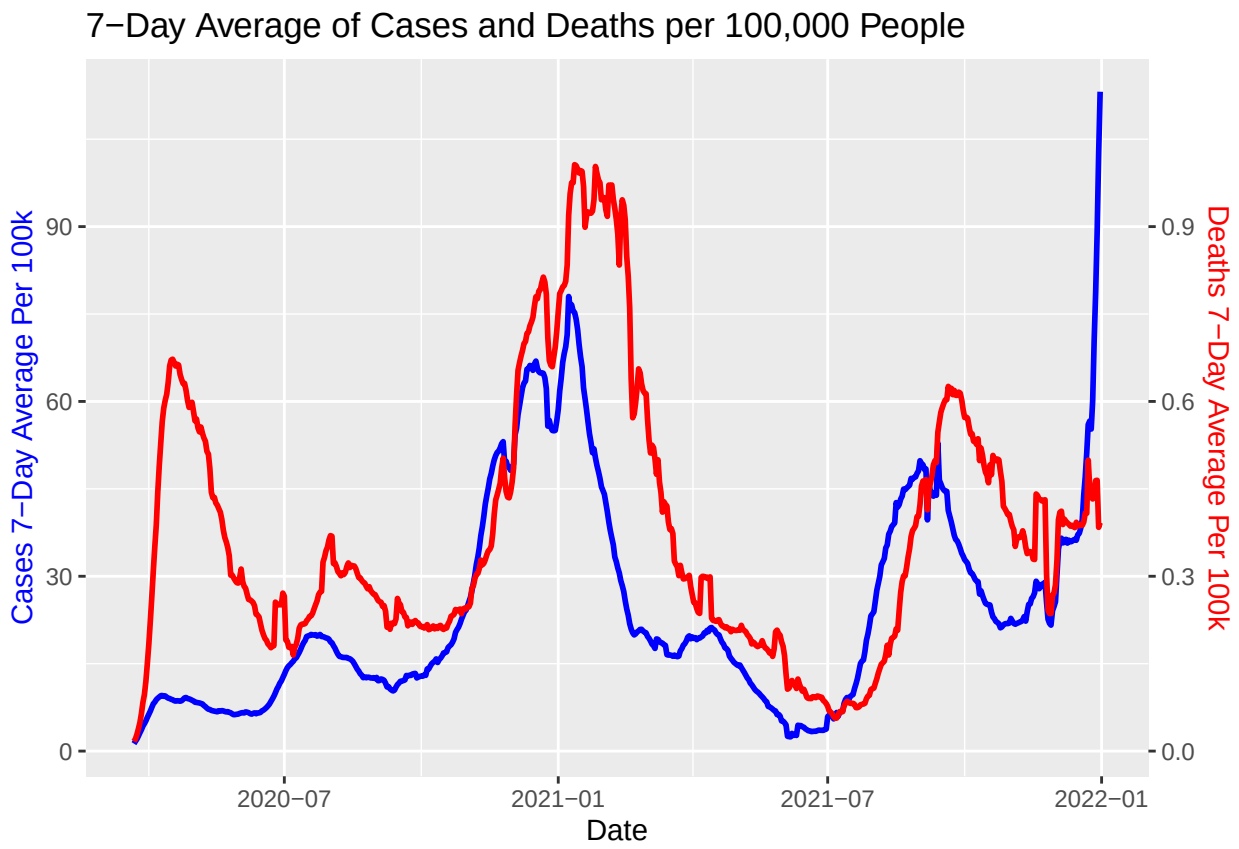
```
#  Your output should look similar to the following tibble:
#
#  date
#  total_deaths     >  the cumulative number of deaths up to and including the associated date
#  total_cases      >  the cumulative number of cases up to and including the associated date
#  delta_deaths_1   >  the number of new deaths since the previous day
#  delta_cases_1    >  the number of new cases since the previous day
#  delta_deaths_7   >  the average number of deaths in a seven-day period
#  delta_cases_7    >  the average number of cases in a seven-day period
#==
#  A tibble: 657 x 7
#       date         total_deaths   total_cases   delta_deaths_1   delta_cases_1 delta_deaths_7 delta_cases_7
#      <date>            <dbl>         <dbl>          <dbl>           <dbl>           <dbl>          <dbl>
#   1 2020-03-15         0.0205        1.08                0               0              NA             NA
#   2 2020-03-16         0.0275        1.36          0.00694           0.274              NA             NA
#   3 2020-03-17         0.0353        1.78          0.00784           0.422              NA             NA
#   4 2020-03-18         0.0489        2.52           0.0136           0.737              NA             NA
#   5 2020-03-19         0.0640        3.74           0.0151            1.22              NA             NA
#   6 2020-03-20         0.0836        5.43           0.0196            1.69              NA             NA
#   7 2020-03-21         0.108         7.39           0.0247            1.96              NA             NA
#   8 2020-03-22         0.138         9.97           0.0296            2.58          0.0168           1.27
#   9 2020-03-23         0.174        13.1            0.0362            3.14          0.0209           1.68
#  10 2020-03-24         0.236        16.3            0.0621            3.14          0.0287           2.07
```

The increasing values for delta_deaths_1 and delta_cases_1 suggest a rising trend in both metrics, which is further confirmed by the increasing delta_deaths_7 and delta_cases_7. This type of data can be useful for understanding the progression of the pandemic and the effectiveness of interventions over time.

Question 5

Create a visualization to compare the seven-day average cases and deaths per 100,000 people.

```
# compare 7-day average cases and deaths per 100k people
daily_7day_avg_per_100k %>% filter(!is.na(delta_cases_7)) %>%
    ggplot(aes(x = date)) +
    geom_line(aes(y = delta_cases_7, color = "Cases"), linewidth = 1) +
    # scale for visibility
    geom_line(aes(y = delta_deaths_7 * 100, color = "Deaths"), linewidth = 1) +
    # set up y axis for 2 variables
    scale_y_continuous(name = "Cases 7-Day Average Per 100k",
                       # add second axis for deaths
                       sec.axis = sec_axis(~./100, name = "Deaths 7-Day Average Per 100k")) +
    # color the y axis titles to match the plotted lines
    theme(axis.title.y.left = element_text(color = "blue"),
          axis.title.y.right = element_text(color = "red")) +
    labs(title = "7-Day Average of Cases and Deaths per 100,000 People",
         x = "Date") +
    # remove legend name
    theme(legend.position = "none") +
    scale_color_manual(values = c("Cases" = "blue", "Deaths" = "red"))
```



The plot really highlights the increased rates of cases and deaths around January 2021.

Part 2 - US State Comparison

While understanding the trends on a national level can be helpful in understanding how COVID-19 impacted the United States, it is important to remember that the virus arrived in the United States at different times. For the next part of your analysis, you will begin to look at COVID related deaths and cases at the state and county-levels.

Question 1

Your first task in Part 2 is to determine the top 10 states in terms of total deaths and cases between March 15, 2020, and December 31, 2021.

Once you have both lists, briefly describe your methodology and your results.

```
# Determine the top 10 states in terms of total deaths and cases between March 15, 2020, and December 31, 202
```

```
# Your transformed data should look similar to the following tibble:
#
# A tibble: 51 x 4
#    state                date       total_deaths  total_cases
#    <chr>                <date>            <dbl>        <dbl>
#  1 California           2021-12-31        76709      5515613
#  2 Texas                2021-12-31        76062      4574881
#  3 Florida              2021-12-31        62504      4166392
#  4 New York             2021-12-31        58993      3473970
#  5 Illinois             2021-12-31        31017      2154058
#  6 Pennsylvania         2021-12-31        36705      2036424
#  7 Ohio                 2021-12-31        29447      2016095
#  8 Georgia              2021-12-31        30283      1798497
```

```
#  9 Michigan             2021-12-31        28984      1706355
# 10 North Carolina       2021-12-31        19436      1685504
# ... with 41 more rows
```

– Communicate your methodology, results, and interpretation here –

Question 2

Determine the top 10 states in terms of deaths per 100,000 people and cases per 100,000 people between March 15, 2020, and December 31, 2021.

Once you have both lists, briefly describe your methodology and your results. Do you expect the lists to be different than the one produced in Question 1? Which method, total or per 100,000 people, is a better method for reporting the statistics?

```
# Determine the top 10 states in terms of deaths and cases per 100,000 people between March 15, 2020, and Dec
```

```
# Your transformed data should look similar to the following tibble:
#
# A tibble: 51 x 4
#    state                date           deaths_per_100k  cases_per_100k
#    <chr>                <date>               <dbl>          <dbl>
#  1 North Dakota         2021-12-31             265.           22482.
#  2 Alaska               2021-12-31             130.           21310.
#  3 Rhode Island         2021-12-31             280.           21093.
#  4 South Dakota         2021-12-31             278.           20014.
#  5 Wyoming              2021-12-31             264.           19979.
#  6 Tennessee            2021-12-31             296.           19783.
#  7 Kentucky             2021-12-31             269.           19173.
#  8 Florida              2021-12-31             287.           19128.
#  9 Utah                 2021-12-31             113.           19088.
# 10 Wisconsin            2021-12-31             190.           19008.
# ... with 41 more rows
```

– Communicate your methodology, results, and interpretation here –

Question 3

Now, select a state and calculate the seven-day averages for new cases and deaths per 100,000 people. Once you have calculated the averages, create a visualization using ggplot2 to represent the data.

```
# Select a state and then filter by state and date range your data from Question 1. Calculate the seven-day a
```

```
# Your transformed data should look similar to the following tibble:
#
# A tibble: 656 × 9
#    state    date       total_deaths total_cases population deaths_per_100k cases_per_100k deaths_7_day  ca
#    <chr>    <date>            <dbl>       <dbl>      <dbl>          <dbl>          <dbl>        <dbl>
#  1 Colorado 2020-03-15            2         136    5784308         0.0346           2.35           NA
#  2 Colorado 2020-03-16            2         161    5784308         0.0346           2.78           NA
#  3 Colorado 2020-03-17            3         183    5784308         0.0519           3.16           NA
#  4 Colorado 2020-03-18            3         216    5784308         0.0519           3.73           NA
#  5 Colorado 2020-03-19            5         278    5784308         0.0864           4.81           NA
#  6 Colorado 2020-03-20            5         364    5784308         0.0864           6.29           NA
#  7 Colorado 2020-03-21            6         475    5784308         0.104            8.21           NA
#  8 Colorado 2020-03-22            7         591    5784308         0.121            10.2         0.0123
```

```
#  9 Colorado 2020-03-23           10      721      5784308        0.173           12.5       0.0198
# 10 Colorado 2020-03-24           11      912      5784308        0.190           15.8       0.0198
# ... with 646 more rows
```

– Communicate your methodology, results, and interpretation here –

Question 4

Using the same state, identify the top 5 counties in terms of deaths and cases per 100,000 people.

```
# Using the same state as Question 2, filter your state and date range from the combined data set from Part 1
```

```
# Your transformed data should be similar to the following tibbles:
#
# Arranged by deaths:
# A tibble: 64 × 4
#    county       date       fips    total_deaths   total_cases
#    <chr>       <date>      <chr>       <dbl>         <dbl>
#  1 El Paso    2021-12-20    08041       1355         119772
#  2 Denver     2021-12-20    08031       1065         106747
#  3 Jefferson  2021-12-20    08059       1061          76732
#  4 Adams      2021-12-20    08001       1057          90476
#  5 Arapahoe   2021-12-20    08005       1046          95769
#  6 Pueblo     2021-12-20    08101        643          30739
#  7 Weld       2021-12-20    08123        569          55599
#  8 Mesa       2021-12-20    08077        445          29542
#  9 Larimer    2021-12-20    08069        393          47444
# 10 Douglas    2021-12-20    08035        361          48740
# ... with 54 more rows
#
#
# Arranged by cases:
# A tibble: 64 × 4
#    county       date       fips   total_deaths   total_cases
#    <chr>       <date>      <chr>     <dbl>         <dbl>
#  1 El Paso    2021-12-20    08041     1355         119772
#  2 Denver     2021-12-20    08031     1065         106747
#  3 Arapahoe   2021-12-20    08005     1046          95769
#  4 Adams      2021-12-20    08001     1057          90476
#  5 Jefferson  2021-12-20    08059     1061          76732
#  6 Weld       2021-12-20    08123      569          55599
#  7 Douglas    2021-12-20    08035      361          48740
#  8 Larimer    2021-12-20    08069      393          47444
#  9 Boulder    2021-12-20    08013      323          36754
# 10 Pueblo     2021-12-20    08101      643          30739
# ... with 54 more rows
```

– Communicate your methodology, results, and interpretation here –

Question 5

Modify the code below for the map projection to plot county-level deaths and cases per 100,000 people for your state.

```
# Copy and modify the code below for your state.
#
# plot_usmap arguments:
```

```
#   regions: can be one of ("states", "state", "counties", "county"). The default is "states"
#   include: The regions to include in the resulting map. If regions is "states"/"state", the value can be ei
#   data: values to plot on the map
#   values: the name of the column that contains the values to be associated with a given region.
#   color: the map outline color.
#
# Reference the plot_usmap documentation for further information using ?plot_usmap

#plot_usmap(regions = "county", include="CO", data = colorado_county, values = "total_deaths", color = "blue"
#  scale_fill_continuous(low = "white", high = "blue", name = "Deaths per 100,000")
```

– Communicate your methodology, results, and interpretation here –

Question 6

Finally, select three other states and calculate the seven-day averages for new deaths and cases per 100,000 people for between March 15, 2020, and December 31, 2021.

– Communicate your methodology, results, and interpretation here –

Question 7

Create a visualization comparing the seven-day averages for new deaths and cases per 100,000 people for the four states you selected.

– Communicate your methodology, results, and interpretation here –

Part 3 - Global Comparison

```
# Import global COVID-19 statistics aggregated by the Center for Systems Science and Engineering (CSSE) at Jo
# Import global population estimates from the World Bank.

#csse_global_deaths <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_
#csse_global_cases <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_1
#csse_us_deaths <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_d
#csse_us_cases <- read_csv("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_da

#global_population_estimates <- read_csv("global_population_estimates.csv")
```

Question 1

Using the state you selected in Part 2 Question 2 compare the daily number of cases and deaths reported from the CSSE and NY Times.

```
# To compare your state data between the two data sets, you will first need to tidy the US CSSE death and cas
# Hint: Review the documentation for pivot_longer().


# Once you have tidied your data, join the two CSSE US data sets to include cases and deaths in one table.


# Finally, create two visualizations with one plotting the CSSE and NY Times cases and the other plotting the
```

```
# Your tidied CSSE data for your selected state should look similar to the following tibble:
#
# A tibble: 43,362 × 6
#     fips  county  state        date      cases  deaths
#     <dbl> <chr>   <chr>       <date>     <dbl>  <dbl>
#  1  8001  Adams  Colorado   2020-03-15     6      0
#  2  8001  Adams  Colorado   2020-03-16     8      0
#  3  8001  Adams  Colorado   2020-03-17    10      0
#  4  8001  Adams  Colorado   2020-03-18    10      0
#  5  8001  Adams  Colorado   2020-03-19    10      0
#  6  8001  Adams  Colorado   2020-03-20    12      0
#  7  8001  Adams  Colorado   2020-03-21    14      0
#  8  8001  Adams  Colorado   2020-03-22    18      0
#  9  8001  Adams  Colorado   2020-03-23    25      0
# 10  8001  Adams  Colorado   2020-03-24    27      0
# ... with 43,352 more rows
```

– Communicate your methodology, results, and interpretation here –

Question 2

Now that you have verified the data reported from the CSSE and NY Times are similar, combine the global and US CSSE data sets and identify the top 10 countries in terms of deaths and cases per 100,000 people between March 15, 2020, and December 31, 2021.

```
# First, combine and tidy the CSSE death and cases data sets. You may wish to keep the two sets separate.
# Then, tidy the global population estimates. While tidying your data, remember to include columns that you u
# You will notice that the population estimates data does not include every country reported in the CSSE data
```

– Communicate your methodology, results, and interpretation here –

Question 3

Construct a visualization plotting the 10 countries in terms of deaths and cases per 100,000 people between March 15, 2020, and December 31, 2021. In designing your visualization keep the number of data you will be plotting in mind. You may wish to create two separate visualizations, one for deaths and another for cases.

– Communicate your methodology, results, and interpretation here –

Question 4

Finally, select four countries from one continent and create visualizations for the daily number of confirmed cases per 100,000 and the daily number of deaths per 100,000 people between March 15, 2020, and December 31, 2021.

– Communicate your methodology, results, and interpretation here –