Challenge-11

Ho Zhi Yi

2023-10-30

Load libraries

```
library(httr)
library(jsonlite)
library(tidyverse)
```

```
historic_state_data_url <- "https://api.covidactnow.org/v2/states.timeseries.json?apiKey=c85924a8e71d478
raw_data <- GET(historic_state_data_url)
```

Extract data

```
data <- fromJSON(rawToChar(raw_data$content))
```

Explore data

```
glimpse(data)
```

```
## Rows: 53
## Columns: 25
## $ fips                          <chr> "02", "01", "05", "04", "06", "08", "09~
## $ country                       <chr> "US", "US", "US", "US", "US", "US", "US~
## $ state                         <chr> "AK", "AL", "AR", "AZ", "CA", "CO", "CT~
## $ county                        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ hsa                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ hsaName                       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ level                         <chr> "state", "state", "state", "state", "st~
## $ lat                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ locationId                    <chr> "iso1:us#iso2:us-ak", "iso1:us#iso2:us-~
## $ long                          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ population                    <int> 731545, 4903185, 3017804, 7278717, 3951~
## $ hsaPopulation                 <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ metrics                       <df[,14]> <data.frame[26 x 14]>
## $ riskLevels                    <df[,6]> <data.frame[26 x 6]>
```

```
## $ cdcTransmissionLevel           <int> 2, 4, 3, 3, 1, 4, 4, 1, 4, 4, 2, 3,~
## $ communityLevels                <df[,2]> <data.frame[26 x 2]>
## $ actuals                        <df[,19]> <data.frame[26 x 19]>
## $ annotations                    <df[,30]> <data.frame[26 x 30]>
## $ lastUpdatedDate                <chr> "2023-10-30", "2023-10-30", "2023-10~
## $ url                            <chr> "https://covidactnow.org/us/alaska-ak",~
## $ metricsTimeseries              <list> [<data.frame[1334 x 14]>], [<data.fr~
## $ actualsTimeseries              <list> [<data.frame[1334 x 20]>], [<data.f~
## $ riskLevelsTimeseries           <list> [<data.frame[1334 x 3]>], [<data.fr~
## $ cdcTransmissionLevelTimeseries <list> [<data.frame[1334 x 2]>], [<data.frame[~
## $ communityLevelsTimeseries      <list> [<data.frame[1334 x 3]>], [<data.frame[~
```

Questions

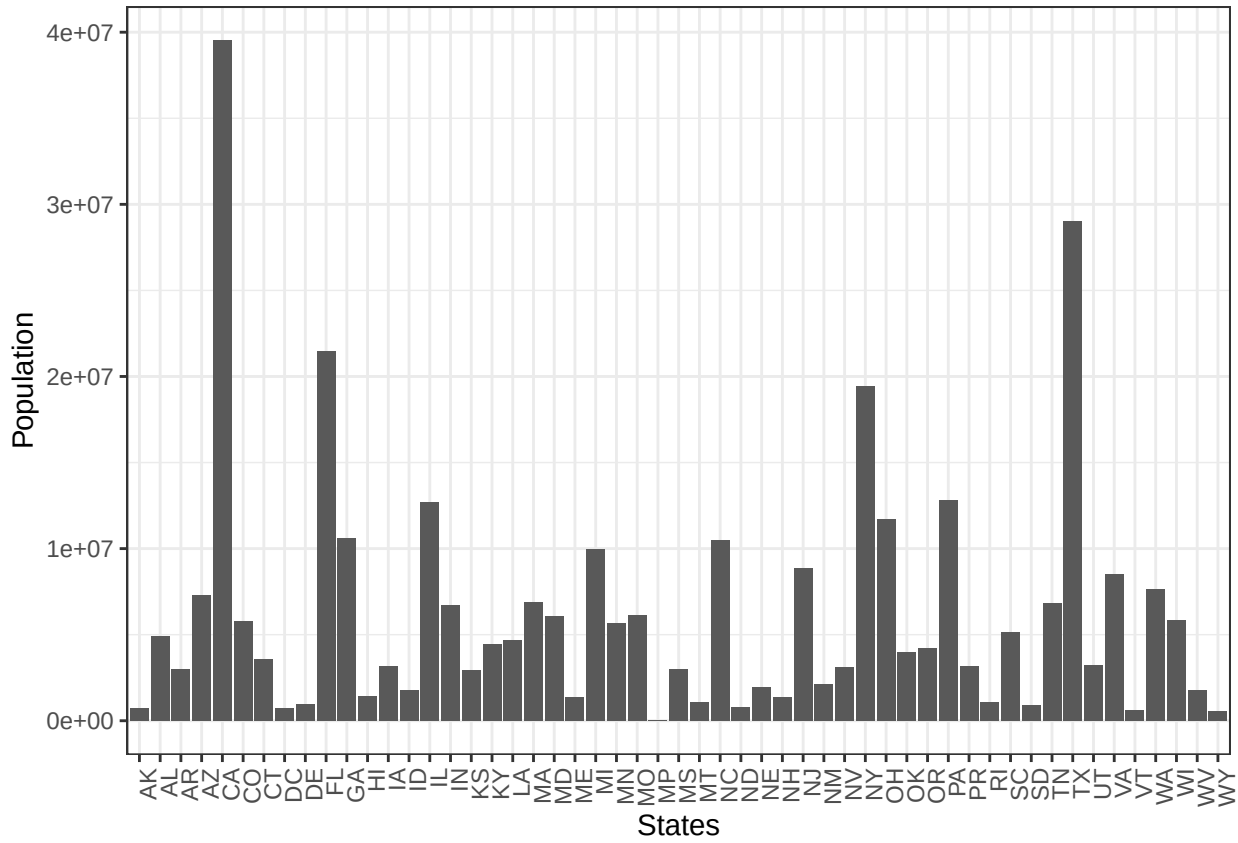
i. What is the population in various states of U.S.A?

```
data$population
```

```
##  [1]   731545  4903185  3017804  7278717 39512223  5758736  3565287   705749
##  [9]   973764 21477737 10617423  1415872  3155070  1787065 12671821  6732219
## [17]  2913314  4467673  4648794  6892503  6045680  1344212  9986857  5639632
## [25]  6137428    53605  2976149  1068778 10488084   762062  1934408  1359711
## [33]  8882190  2096829  3080156 19453561 11689100  3956971  4217737 12801989
## [41]  3193694  1059361  5148714   884659  6829174 28995881  3205958  8535519
## [49]   623989  7614893  5822434  1792147   578759
```

plot

```
ggplot(data, aes(x=state,y=population)) +
  geom_bar(stat="identity") +
  labs(x="States",y="Population") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.justification = c(0, 1))
```



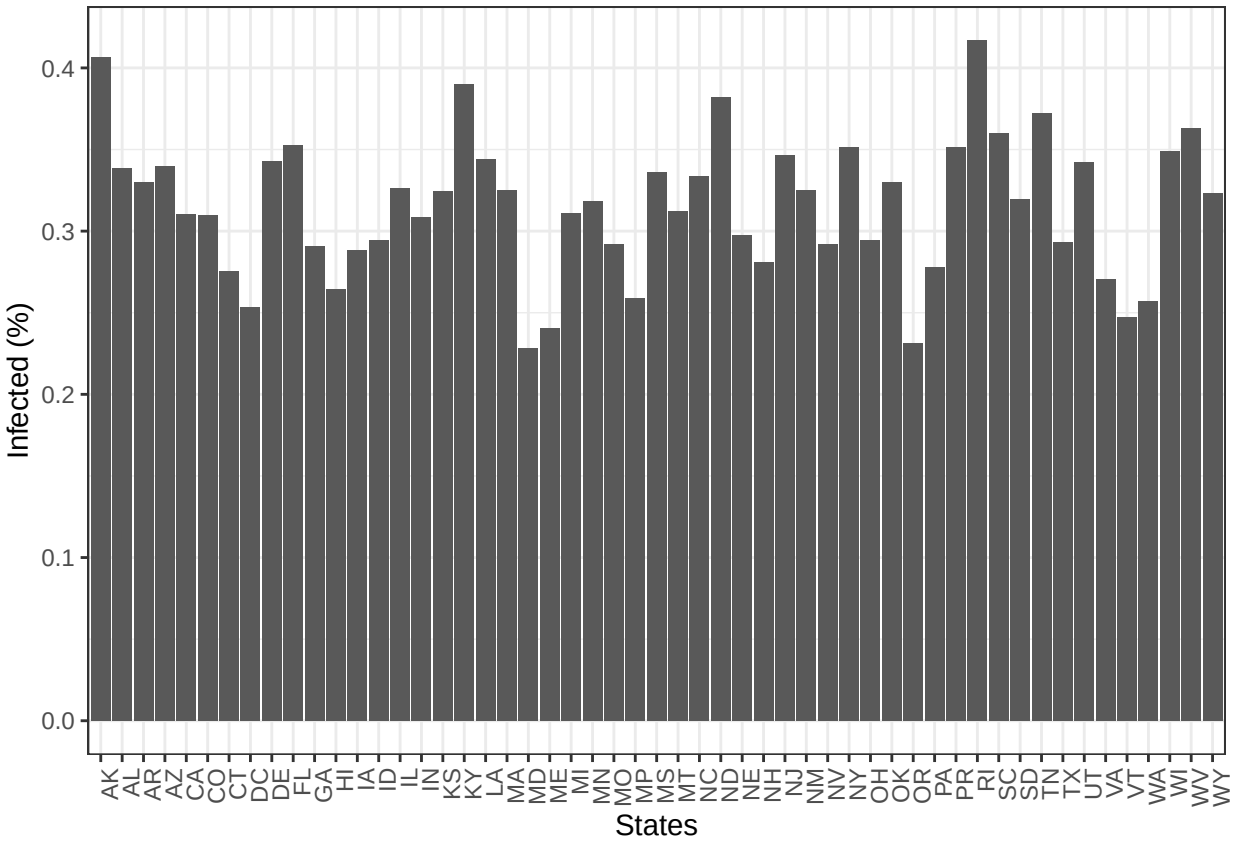
ii. What fraction of the population was infected ?

```
sum(data$actuals$cases)/sum(data$population)
```

```
## [1] 0.3154839
```

plot

```
ggplot(data, aes(x=state,y=(data$actuals$cases/population))) +
  geom_bar(stat="identity") +
  labs(x="States",y="Infected (%)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.justification = c(0, 1))
```



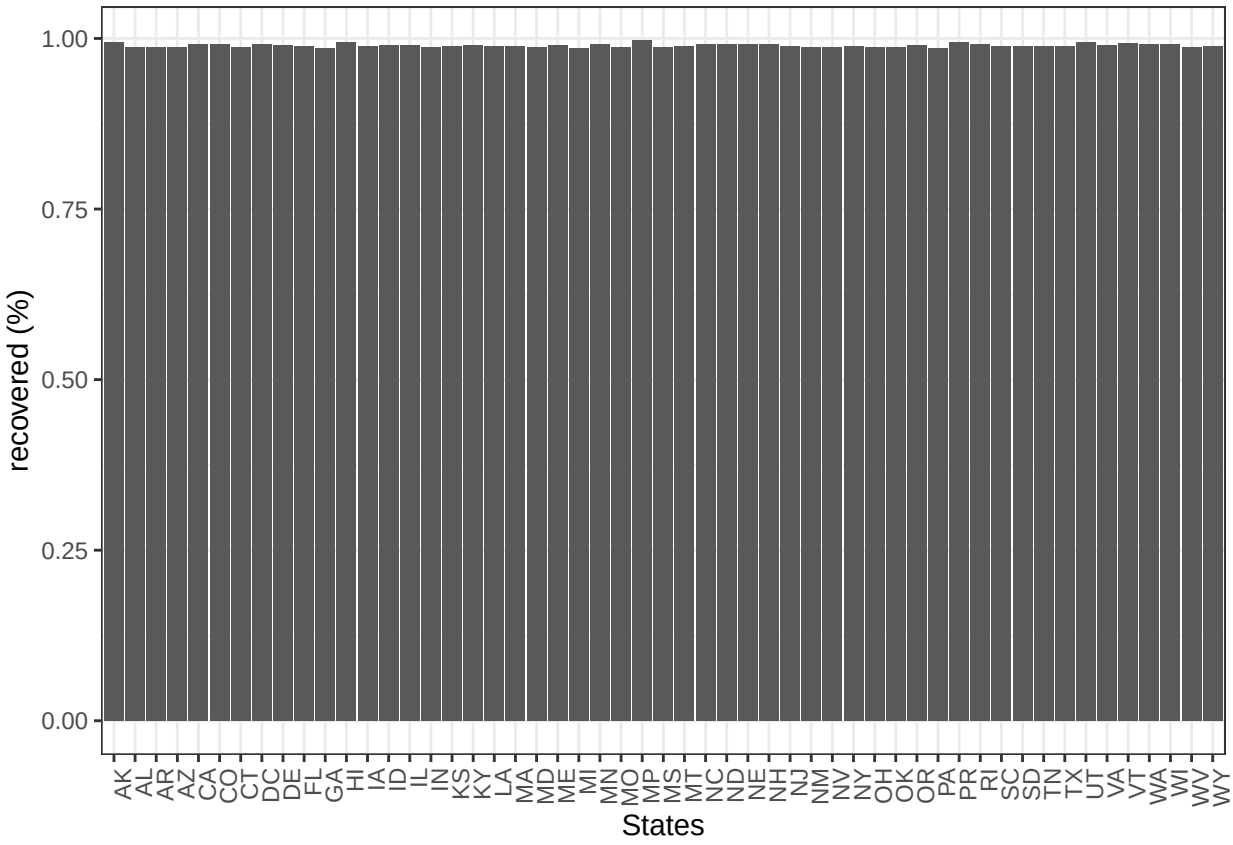
iii. What fraction of infected persons recovered ?

```
sum(data$actuals$cases - data$actuals$deaths)/sum(data$actuals$cases)
```

```
## [1] 0.9891648
```

plot

```
ggplot(data, aes(x=state,y=((data$actuals$cases-data$actuals$deaths)/data$actuals$cases))) +
  geom_bar(stat="identity") +
  labs(x="States",y="recovered (%)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.justification = c(0, 1))
```



iv. What fraction of the population is currently vaccinated ?

```
sum(data$actuals$vaccinationsCompleted)/sum(data$population)
```

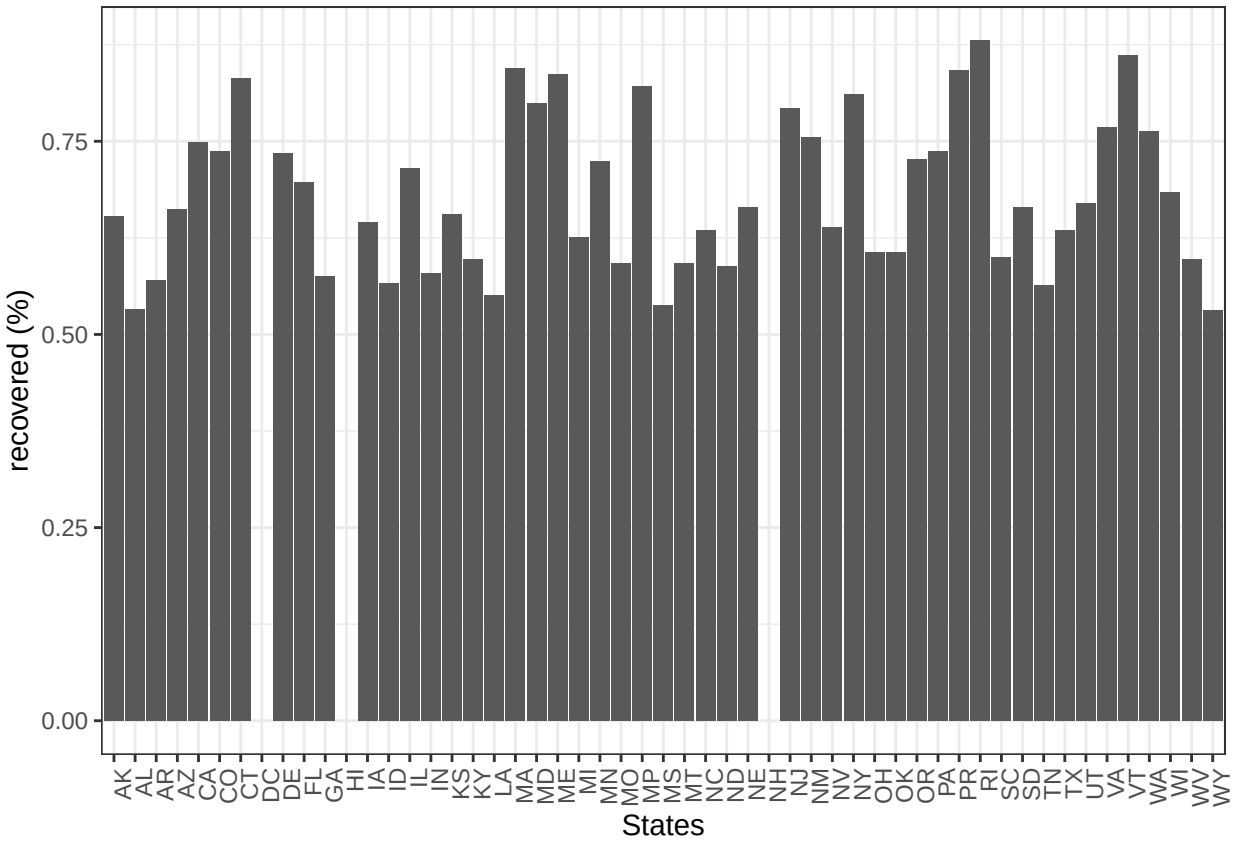
```
## [1] NA
```

plot

```
ggplot(data, aes(x=state,y=((data$actuals$vaccinationsCompleted)/data$population))) +
  geom_bar(stat="identity") +
  labs(x="States",y="recovered (%)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.justification = c(0, 1))
```

```
## Warning: Use of `data$population` is discouraged.
## i Use `population` instead.
```

```
## Warning: Removed 3 rows containing missing values (`position_stack()`).
```



v. What was the transmission-like in the various states ?

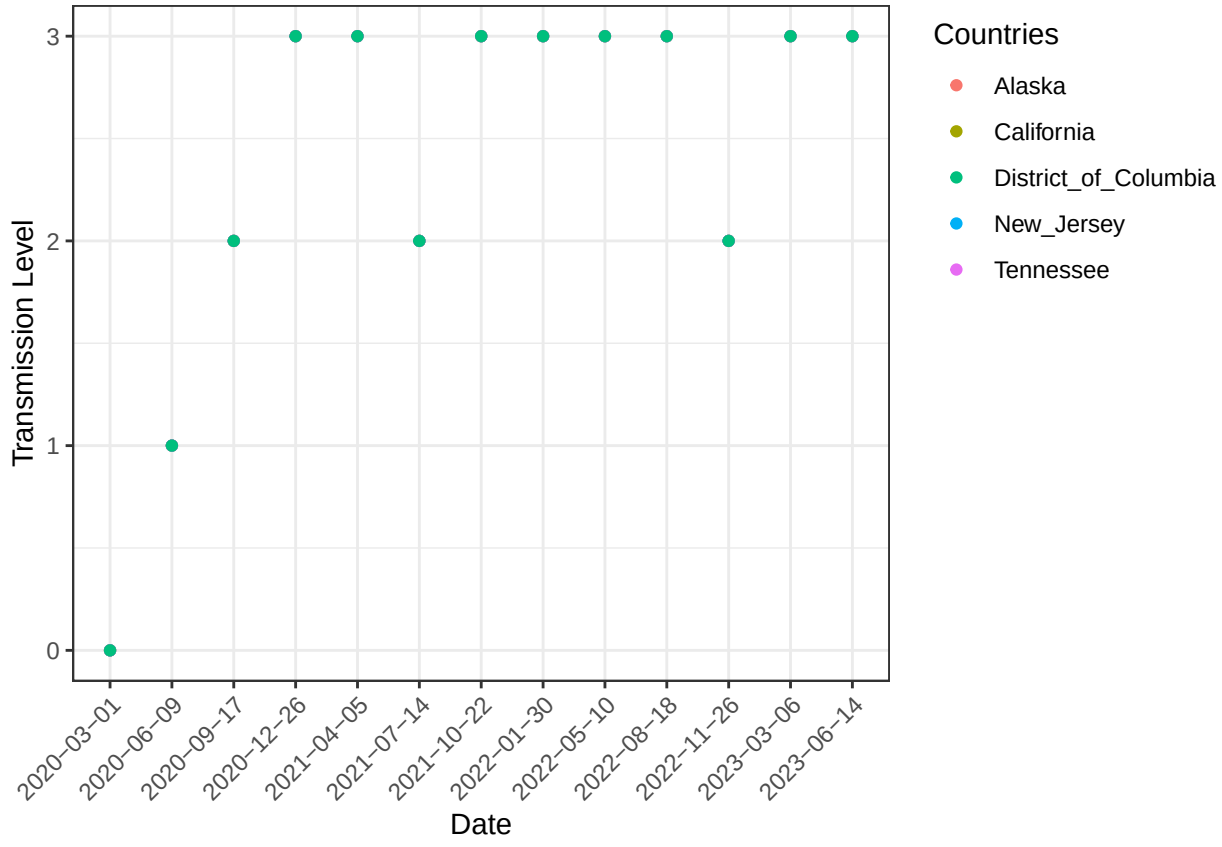
```
time_series <- data %>%
  unnest(actualsTimeseries)

# Creating a new dataframe with needed data
# Save date
time_series_transmission <- tibble(Date=time_series$cdcTransmissionLevelTimeseries[[which(data$state=="
# Transmission levels in each state
time_series_transmission$Alaska <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="AK")]]

time_series_transmission$California <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="CA

time_series_transmission$New_Jersey <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="NJ

time_series_transmission$Tennessee <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="TN

time_series_transmission$District_of_Columbia <- time_series$cdcTransmissionLevelTimeseries[[which(data$

print(head(time_series_transmission))
```

```
## # A tibble: 6 x 6
##   Date       Alaska California New_Jersey Tennessee District_of_Columbia
```

```
##   <chr>       <int>      <int>      <int>     <int>                 <int>
## 1 2020-03-01      0          0          0         0                     0
## 2 2020-03-02      0          0          0         0                     0
## 3 2020-03-03      0          0          0         0                     0
## 4 2020-03-04      0          0          0         0                     0
## 5 2020-03-05      0          0          0         0                     0
## 6 2020-03-06      0          0          0         0                     0
```

plot

```
time_series_transmission[seq(1,1300,by=100),]%>%
  pivot_longer(cols=Alaska:District_of_Columbia,names_to="Countries",values_to="Transmission") %>%
  ggplot(aes(x=Date,y=Transmission,colour=Countries,group=Countries)) +
  geom_point(show.legend=TRUE) +
  labs(x="Date",y="Transmission Level") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.justification = c(0, 1))
```



vi. How did the disease progress since it started ?

it increased at an increasing rate

```
# New data-frame with dates
time_series_cases <- list(Alaska = time_series %>%
```

```
                            filter(state=="AK") %>%
                            select(date,cases))

# Cases of each state
(time_series_cases$California <- time_series %>%
  filter(state=="CA") %>%
  select(date,cases))
```

```
## # A tibble: 1,368 x 2
##    date       cases
##    <chr>      <int>
##  1 2020-01-25     1
##  2 2020-01-26     2
##  3 2020-01-27     2
##  4 2020-01-28     2
##  5 2020-01-29     2
##  6 2020-01-30     2
##  7 2020-01-31     3
##  8 2020-02-01     3
##  9 2020-02-02     6
## 10 2020-02-03     6
## # i 1,358 more rows
```

```
(time_series_cases$New_Jersey <- time_series %>%
  filter(state=="NJ") %>%
  select(date,cases))
```

```
## # A tibble: 1,330 x 2
##    date       cases
##    <chr>      <int>
##  1 2020-03-01    NA
##  2 2020-03-02    NA
##  3 2020-03-03    NA
##  4 2020-03-04     1
##  5 2020-03-05     2
##  6 2020-03-06     4
##  7 2020-03-07     4
##  8 2020-03-08     6
##  9 2020-03-09    11
## 10 2020-03-10    15
## # i 1,320 more rows
```

```
(time_series_cases$Tennessee <- time_series %>%
  filter(state=="TN") %>%
  select(date,cases))
```

```
## # A tibble: 1,333 x 2
##    date       cases
##    <chr>      <int>
##  1 2020-03-01    NA
##  2 2020-03-02    NA
##  3 2020-03-03    NA
```

```
##  4 2020-03-04    NA
##  5 2020-03-05     1
##  6 2020-03-06     1
##  7 2020-03-07     1
##  8 2020-03-08     3
##  9 2020-03-09     4
## 10 2020-03-10     7
## # i 1,323 more rows
```

```
(time_series_cases$District_of_Columbia <- time_series %>%
  filter(state=="DC") %>%
  select(date,cases))
```

```
## # A tibble: 1,331 x 2
##    date       cases
##    <chr>      <int>
##  1 2020-03-06    NA
##  2 2020-03-07     1
##  3 2020-03-08     1
##  4 2020-03-09     4
##  5 2020-03-10     4
##  6 2020-03-11    10
##  7 2020-03-12    10
##  8 2020-03-13    10
##  9 2020-03-14    16
## 10 2020-03-15    17
## # i 1,321 more rows
```

data to plot

```
data_to_plot <- tibble(Date_Alaska = time_series_cases$Alaska$date[seq(1,1300,by=100)],
 Cases_Alaska = time_series_cases$Alaska$cases[seq(1,1300,by=100)],
 Date_California = time_series_cases$California$date[seq(1,1300,by=100)],
 Cases_California = time_series_cases$California$cases[seq(1,1300,by=100)],
 Date_New_Jersey = time_series_cases$New_Jersey$date[seq(1,1300,by=100)],
 Cases_New_Jersey = time_series_cases$New_Jersey$cases[seq(1,1300,by=100)],
 Date_Tennessee = time_series_cases$Tennessee$date[seq(1,1300,by=100)],
 Cases_Tennessee = time_series_cases$Tennessee$cases[seq(1,1300,by=100)],
 Date_District_of_Columbia = time_series_cases$District_of_Columbia$date[seq(1,1300,by=100)],
 Cases_District_of_Columbia = time_series_cases$District_of_Columbia$cases[seq(1,1300,by=100)])
```

plots

```
library(cowplot)

fig1 <- ggplot(data_to_plot, aes(x = Date_Alaska, y = Cases_Alaska)) +
  geom_point() +
  labs(x = "Date", y = "Cases", title = "Alaska") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1), legend.justification = c(0, 1))

fig2 <- ggplot(data_to_plot, aes(x = Date_California, y = Cases_California)) +
```

```
  geom_point() +
  labs(x = "Date", y = "Cases", title = "California") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1), legend.justification = c(0, 1))

fig3 <- ggplot(data_to_plot, aes(x = Date_New_Jersey, y = Cases_New_Jersey)) +
  geom_point() +
  labs(x = "Date", y = "Cases", title = "New Jersey") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1), legend.justification = c(0, 1))

fig4 <- ggplot(data_to_plot, aes(x = Date_Tennessee, y = Cases_Tennessee)) +
  geom_point() +
  labs(x = "Date", y = "Cases", title = "Tennessee") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1), legend.justification = c(0, 1))

fig5 <- ggplot(data_to_plot, aes(x = Date_District_of_Columbia, y = Cases_District_of_Columbia)) +
  geom_point() +
  labs(x = "Date", y = "Cases", title = "District of Columbia") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 35, hjust = 1), legend.justification = c(0, 1))

plot_grid(
  fig1, fig2, fig3, fig4, fig5,
  ncol = 2, labels = LETTERS[1:5],
  align = "v", axis = "lr"
)
```

screenshot

```
knitr::include_graphics("screenshot.jpg")
```