Week 11 Challenge

Code

API Link: https://api.covidactnow.org/v2/states.timeseries.json?apiKey=af425098c233446aa11a347a9dbd4d6a (https://api.covidactnow.org/v2/states.timeseries.json?apiKey=af425098c233446aa11a347a9dbd4d6a)

Hide

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

Hide

```
historic_state_data_url <- "https://api.covidactnow.org/v2/states.timeseries.json?apiKey=af425098c233446aa11a347a9dbd4d6a"

raw_data <- GET(historic_state_data_url)
```

##Step2: Extracting data

1. json format to data-frame

Hide

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

##Step3: Exploring data

Hide

```
glimpse(data)
```

```
Rows: 53
Columns: 25
$ fips                          <chr> "02", "01", "05", "04", "06", "08", "09", "1
1", "10", "12", "13", "15", "19", "16", "17", "18"…
$ country                       <chr> "US", "US", "US", "US", "US", "US", "US", "U
S", "US", "US", "US", "US", "US", "US", "US", "US"…
$ state                         <chr> "AK", "AL", "AR", "AZ", "CA", "CO", "CT", "D
C", "DE", "FL", "GA", "HI", "IA", "ID", "IL", "IN"…
$ county                        <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ hsa                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ hsaName                       <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ level                         <chr> "state", "state", "state", "state", "state",
"state", "state", "state", "state", "state", "sta…
$ lat                           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ locationId                    <chr> "iso1:us#iso2:us-ak", "iso1:us#iso2:us-al", "i
so1:us#iso2:us-ar", "iso1:us#iso2:us-az", "iso1:…
$ long                          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ population                    <int> 731545, 4903185, 3017804, 7278717, 39512223, 5
758736, 3565287, 705749, 973764, 21477737, 10617…
$ hsaPopulation                 <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N
A, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA…
$ metrics                       <df[,14]> <data.frame[45 x 14]>
$ riskLevels                    <df[,6]> <data.frame[45 x 6]>
$ cdcTransmissionLevel          <int> 2, 4, 3, 3, 1, 4, 4, 1, 4, 4, 2, 3, 4, 3, 2,
3, 3, 3, 2, 1, 2, 2, 3, 4, 4, 1, 2, 3, 4, 3, …
$ communityLevels               <df[,2]> <data.frame[45 x 2]>
$ actuals                       <df[,19]> <data.frame[45 x 19]>
$ annotations                   <df[,30]> <data.frame[45 x 30]>
$ lastUpdatedDate               <chr> "2023-10-30", "2023-10-30", "2023-10-30", "202
3-10-30", "2023-10-30", "2023-10-30", "2023-1…
$ url                           <chr> "https://covidactnow.org/us/alaska-ak", "http
s://covidactnow.org/us/alabama-al", "https://covi…
$ metricsTimeseries             <list> [<data.frame[1334 x 14]>], [<data.frame[1330
x 14]>], [<data.frame[1332 x 14]>], [<data.fra…
$ actualsTimeseries             <list> [<data.frame[1334 x 20]>], [<data.frame[1330
x 20]>], [<data.frame[1332 x 20]>], [<data.fr…
$ riskLevelsTimeseries          <list> [<data.frame[1334 x 3]>], [<data.frame[1330 x
3]>], [<data.frame[1332 x 3]>], [<data.frame…
$ cdcTransmissionLevelTimeseries <list> [<data.frame[1334 x 2]>], [<data.frame[1330 x
2]>], [<data.frame[1332 x 2]>], [<data.frame[137…
$ communityLevelsTimeseries     <list> [<data.frame[1334 x 3]>], [<data.frame[1330 x
3]>], [<data.frame[1332 x 3]>], [<data.frame[137…
```

##Step5: Mapping Variables to Questions

Hide

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

##Step5.5: Answers to Questions a. i. What is the population in various states of U.S.A

Hide

```
data %>% select(state, population)
```

	state <chr>	population <int>
1	AK	731545
2	AL	4903185
3	AR	3017804
4	AZ	7278717
5	CA	39512223
6	CO	5758736
7	CT	3565287
8	DC	705749
9	DE	973764
10	FL	21477737

1-10 of 53 rows Previous **1** 2 3 4 5 6 Next

ii. What fraction of the population was infected?

Hide

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

```
[1] 0.3154839
```

iii. What fraction of infected persons recovered ?

Hide

```
n_recovered <- sum(actuals_unnested$cases) - sum(actuals_unnested$deaths)
percent_recovered <- n_recovered / sum(actuals_unnested$cases)
percent_recovered
```

```
[1] 0.9891648
```

iv. What fraction of the population is currently vaccinated?

Hide

```
mean(metrics_unnested$vaccinationsCompletedRatio, na.rm = TRUE)
```

```
[1] 0.6811
```

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

Hide

```
data %>% select(state, cdcTransmissionLevel)
```

	state <chr>	cdcTransmissionLevel <int>
1	AK	2
2	AL	4
3	AR	3
4	AZ	3
5	CA	1
6	CO	4
7	CT	4
8	DC	1
9	DE	4
10	FL	4

1-10 of 53 rows Previous **1** 2 3 4 5 6 Next

Hide

```
time_series_transmission <- tibble(Date=time_series$cdcTransmissionLevelTimeseries[[which(data$state=="CA")]]$date)

time_series_transmission$Alaska <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="AK")]]$cdcTransmissionLevel

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

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

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

time_series_transmission$District_of_Columbia <- time_series$cdcTransmissionLevelTimeseries[[which(data$state=="DC")]]$cdcTransmissionLevel

print(head(time_series_transmission))
```

Date <chr>	Alaska <int>	California <int>	New_Jersey <int>	Tennessee <int>	District_of_Columbia <int>
2020-03-01	0	0	0	0	0
2020-03-02	0	0	0	0	0
2020-03-03	0	0	0	0	0
2020-03-04	0	0	0	0	0

Date <chr>	Alaska <int>	California <int>	New_Jersey <int>	Tennessee <int>	District_of_Columbia <int>
2020-03-05	0	0	0	0	0
2020-03-06	0	0	0	0	0

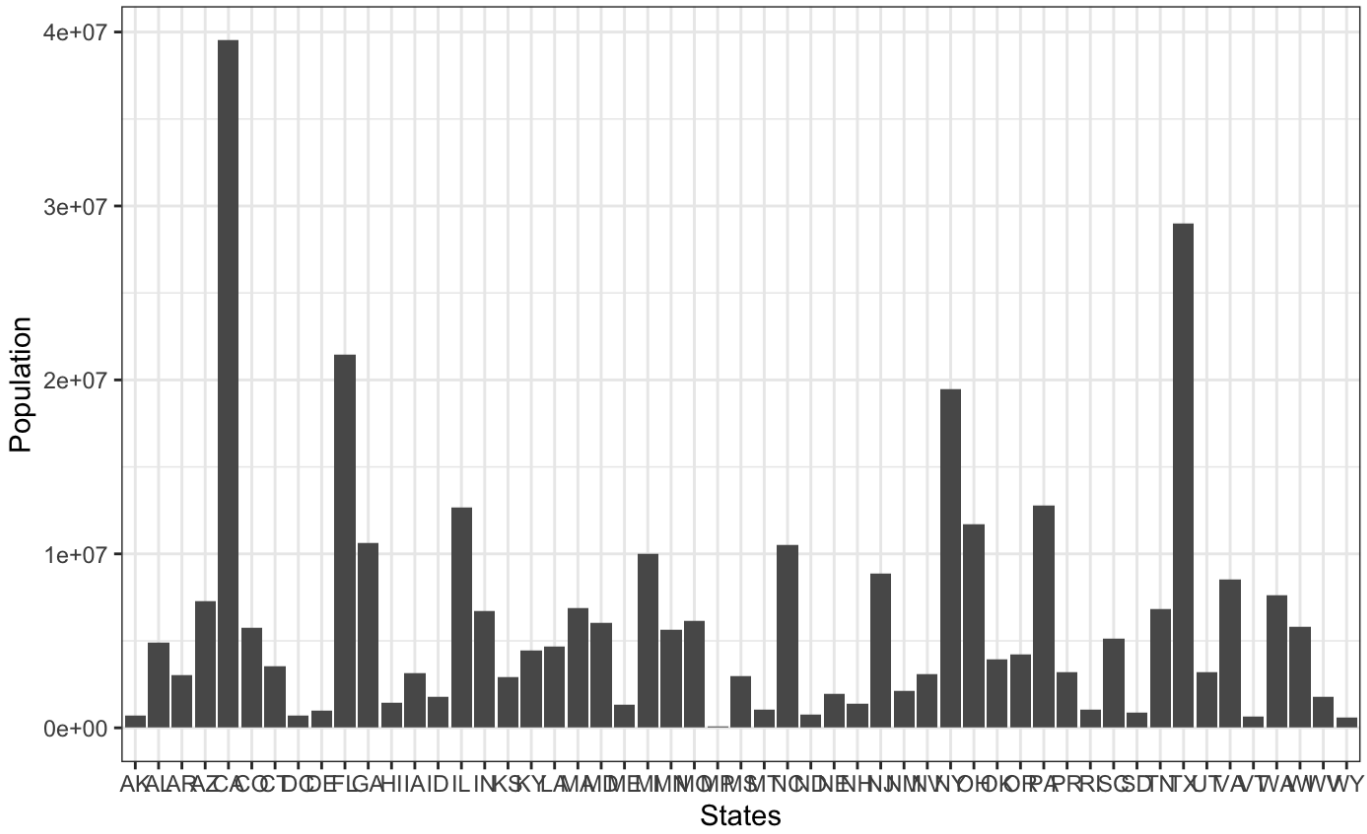
6 rows

vi. How did the disease progress since it started ?

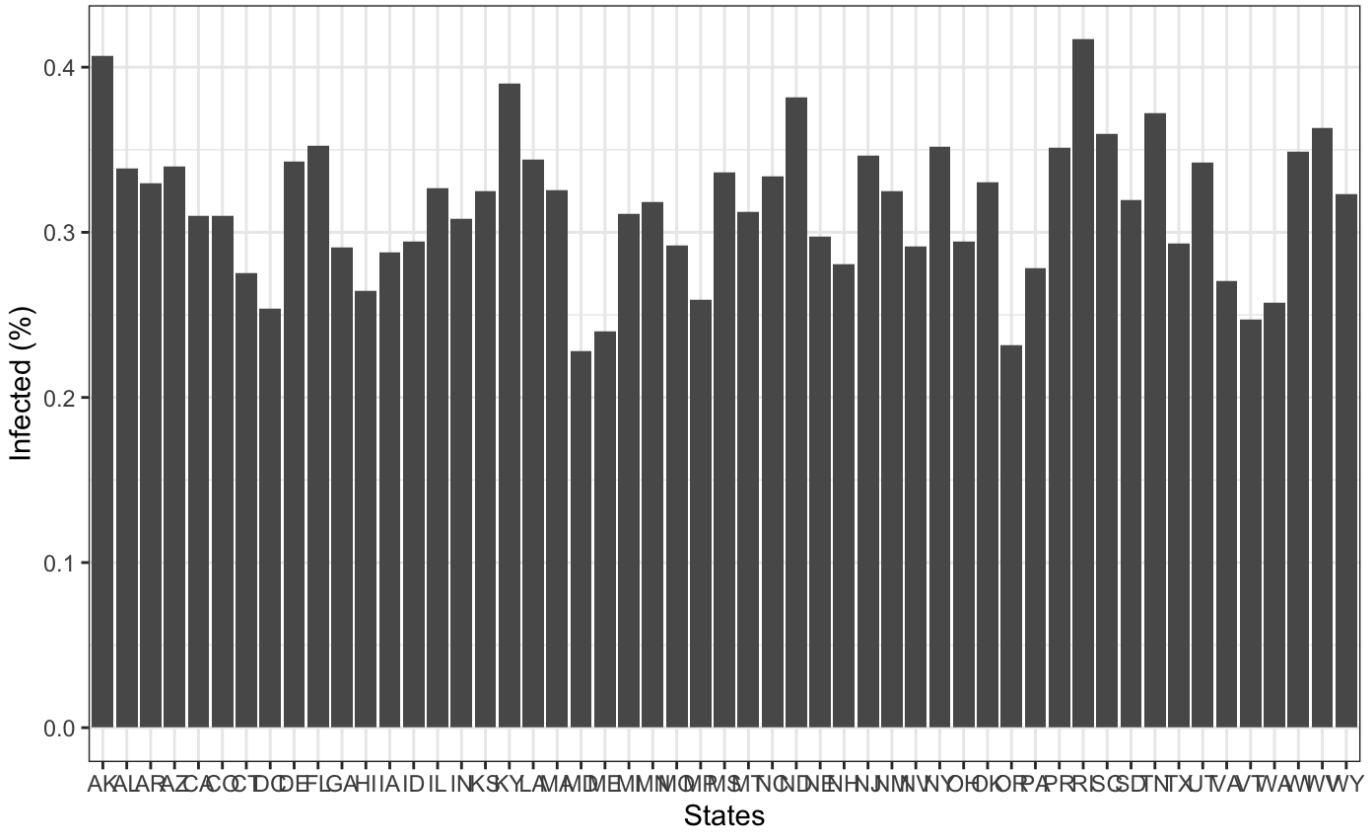
Hide

```
# New data-frame with dates
time_series_cases <- list(Alaska = time_series %>% filter(state=="AK") %>% select(dat
e,cases))
# Cases of each state
time_series_cases$California <- time_series %>% filter(state=="CA") %>% select(date,c
ases)
time_series_cases$New_Jersey <- time_series %>% filter(state=="NJ") %>% select(date,c
ases)
time_series_cases$Tennessee <- time_series %>% filter(state=="TN") %>% select(date,ca
ses)
time_series_cases$District_of_Columbia <- time_series %>% filter(state=="DC") %>% sel
ect(date,cases)
```

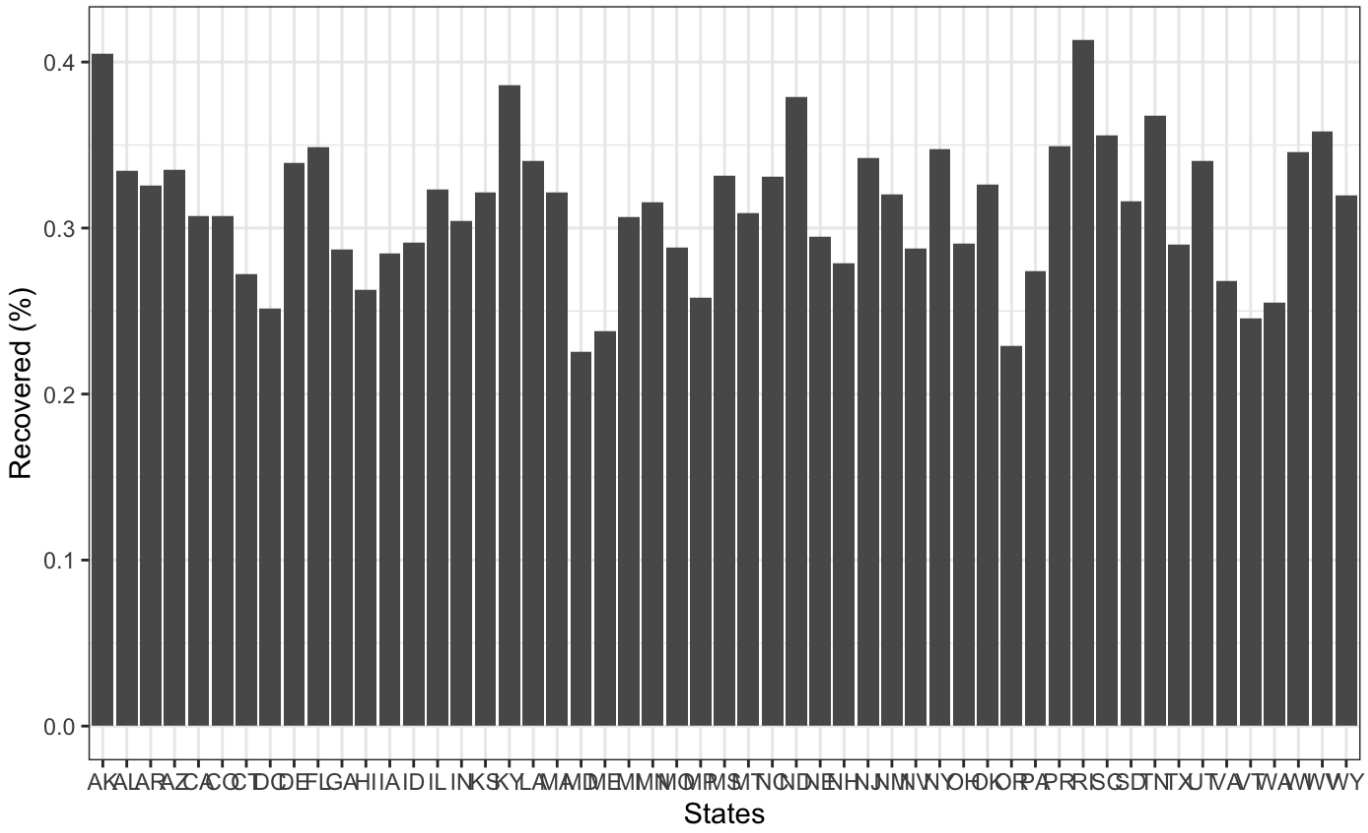
##Step6: Analysing Data i.



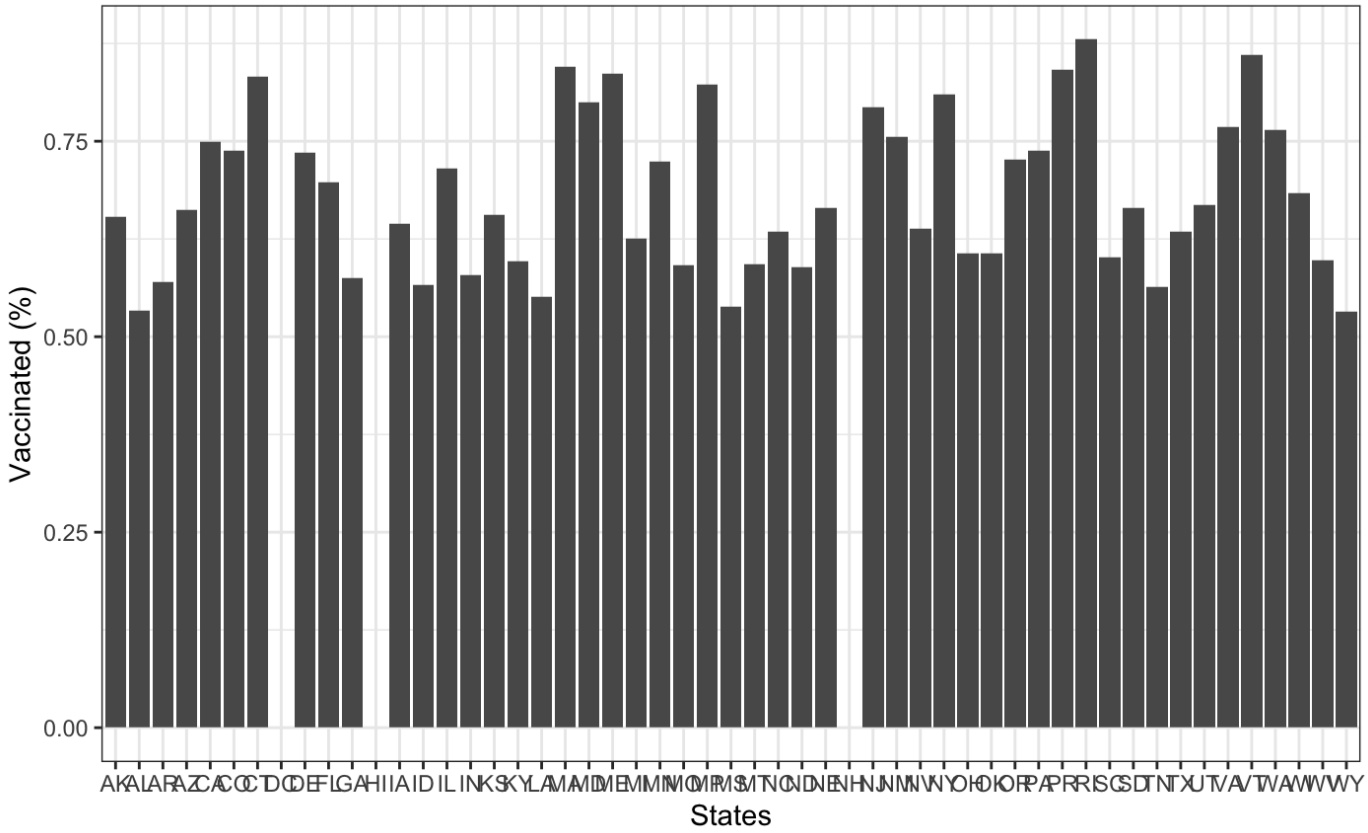
ii.



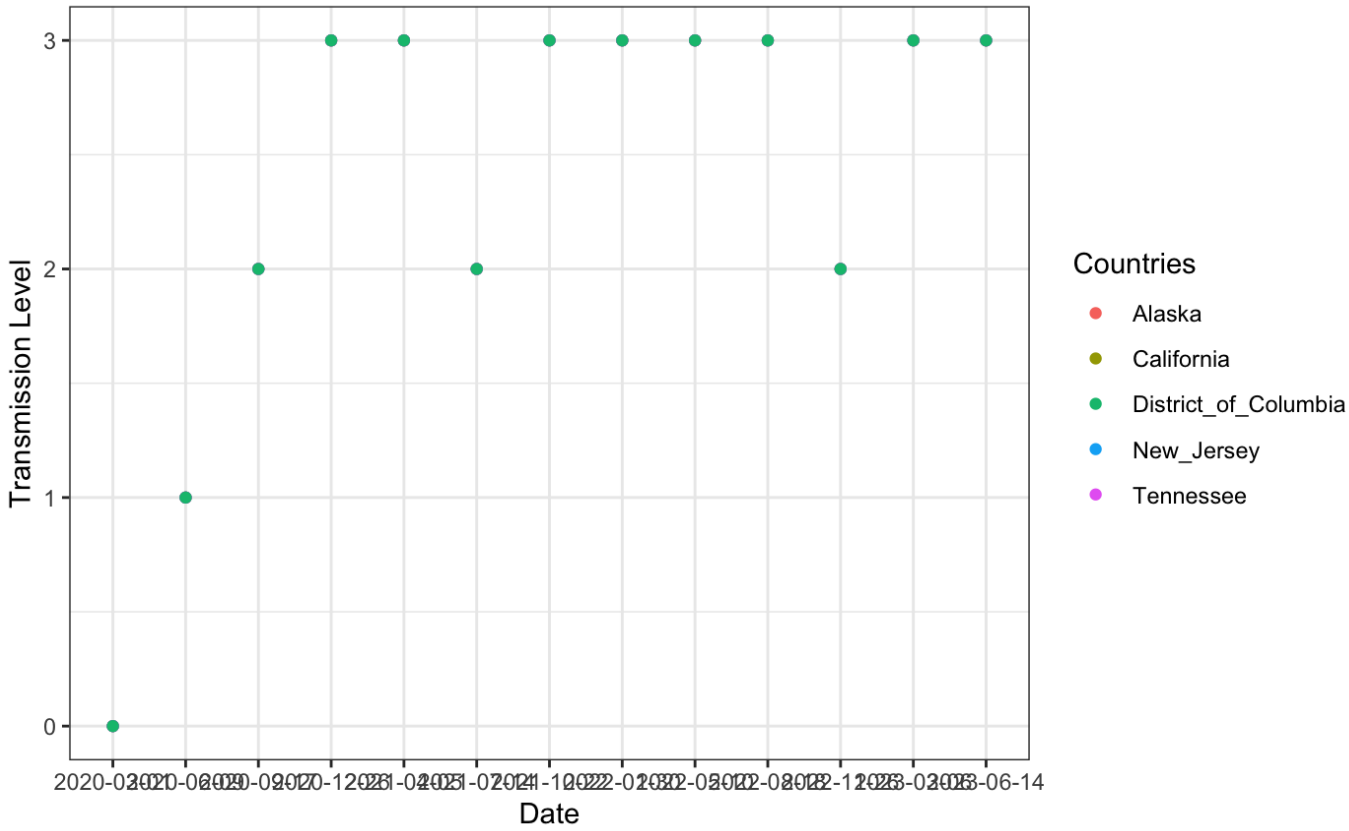
iii.



iv.


Vaccinated (%)
0.75
0.50
0.25
0.00
States

v.


Transmission Level
3
2
1
0
Date
Countries
Alaska
California
District_of_Columbia
New_Jersey
Tennessee

vi.

Hide

```
data_to_plot <- tibble(Date_Alaska = time_series_cases$Alaska$date[seq(1,1300,by=10
0)],
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)])
data_to_plot
```

Date_Alaska <chr>	Cases_Alaska <int>	Date_California <chr>	Cases_California <int>	Date_New_Jersey <chr>
2020-03-01	*NA*	2020-01-25	1	2020-03-01
2020-06-09	620	2020-05-04	56333	2020-06-09
2020-09-17	7413	2020-08-12	595097	2020-09-17
2020-12-26	45247	2020-11-20	1096427	2020-12-26
2021-04-05	63486	2021-02-28	3569578	2021-04-05
2021-07-14	71539	2021-06-08	3798225	2021-07-14
2021-10-22	132393	2021-09-16	4629146	2021-10-22
2022-01-30	211117	2021-12-25	5291605	2022-01-30
2022-05-10	252847	2022-04-04	9110544	2022-05-10
2022-08-18	289203	2022-07-13	10365785	2022-08-18

1-10 of 13 rows | 1-5 of 10 columns Previous **1** 2 Next

Hide

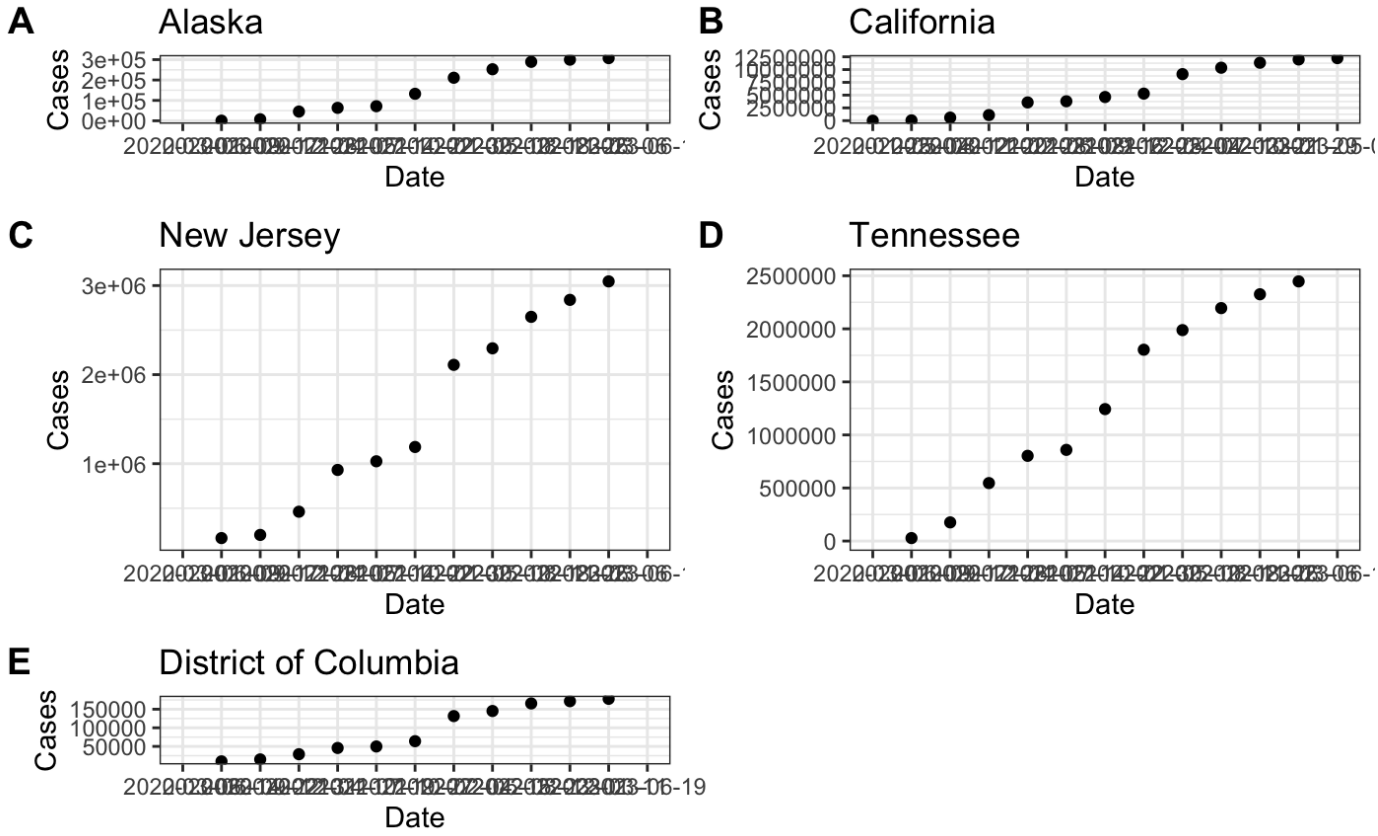
```
install.packages("cowplot")
```

```
Error in install.packages : Updating loaded packages
```

Hide

```
library(cowplot)
```

```
Warning: Removed 2 rows containing missing values (`geom_point()`).Warning: Removed 2
rows containing missing values (`geom_point()`).Warning: Removed 2 rows containing mi
ssing values (`geom_point()`).Warning: Removed 2 rows containing missing values (`geo
m_point()`).
```


A
Alaska
Cases
3e+05
2e+05
1e+05
0e+00
Date
B
California
Cases
Date
C
New Jersey
Cases
3e+06
2e+06
1e+06
Date
D
Tennessee
Cases
2500000
2000000
1500000
1000000
500000
0
Date
E
District of Columbia
Cases
150000
100000
50000
Date