Challenge-11

Marzuki Nooranas

2023-11-06

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

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.3     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter()  masks stats::filter()
## ✖ purrr::flatten() masks jsonlite::flatten()
## ✖ dplyr::lag()     masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
 historic_state_data_url <-
"https://api.covidactnow.org/v2/states.timeseries.json?apiKey=38b7e9895fd84dd9a8f9035
4f9eb3b75"
 raw_data <- GET(historic_state_data_url)
```

STEP-2 : Extracting data

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

STEP-3: Explore the 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> 3, 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> "2023-11-05", "2023-11-05", "2023-11…
## $ url                              <chr> "https://covidactnow.org/us/alaska-ak",…
## $ metricsTimeseries                <list> [<data.frame[1340 x 14]>], [<data.fr…
## $ actualsTimeseries                <list> [<data.frame[1340 x 20]>], [<data.f…
## $ riskLevelsTimeseries             <list> [<data.frame[1340 x 3]>], [<data.fr…
## $ cdcTransmissionLevelTimeseries <list> [<data.frame[1340 x 2]>], [<data.frame[…
## $ communityLevelsTimeseries        <list> [<data.frame[1340 x 3]>], [<data.frame[…
```

STEP-4 : Explore the data

i. What is the population in various states of U.S.A?
ii. What fraction of the population was infected ?
iii. What fraction of infected persons recovered ?
iv. What fraction of the population is currently vaccinated ?
v. What was the transmission-like in the various states ? vi. How did the disease progress since it started ?

STEP-4 : Mapping variables to questions

```
population_data <- data[, c('state', 'population')]
print(population_data)
```

```
##    state population
## 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
## 11    GA   10617423
## 12    HI    1415872
## 13    IA    3155070
## 14    ID    1787065
## 15    IL   12671821
## 16    IN    6732219
## 17    KS    2913314
## 18    KY    4467673
## 19    LA    4648794
## 20    MA    6892503
## 21    MD    6045680
## 22    ME    1344212
## 23    MI    9986857
## 24    MN    5639632
## 25    MO    6137428
## 26    MP      53605
## 27    MS    2976149
## 28    MT    1068778
## 29    NC   10488084
## 30    ND     762062
## 31    NE    1934408
```

```
## 32    NH    1359711
## 33    NJ    8882190
## 34    NM    2096829
## 35    NV    3080156
## 36    NY   19453561
## 37    OH   11689100
## 38    OK    3956971
## 39    OR    4217737
## 40    PA   12801989
## 41    PR    3193694
## 42    RI    1059361
## 43    SC    5148714
## 44    SD     884659
## 45    TN    6829174
## 46    TX   28995881
## 47    UT    3205958
## 48    VA    8535519
## 49    VT     623989
## 50    WA    7614893
## 51    WI    5822434
## 52    WV    1792147
## 53    WY     578759
```

```
data$case_fraction <- data$actuals$cases / data$population
print(data[, c('state', 'case_fraction')])
```

```
##    state case_fraction
## 1     AK     0.4067938
## 2     AL     0.3385424
## 3     AR     0.3297242
## 4     AZ     0.3399162
## 5     CA     0.3100767
## 6     CO     0.3097317
## 7     CT     0.2755276
## 8     DC     0.2534952
## 9     DE     0.3427648
## 10    FL     0.3525642
## 11    GA     0.2908172
## 12    HI     0.2643346
## 13    IA     0.2880874
## 14    ID     0.2944034
## 15    IL     0.3264455
## 16    IN     0.3084163
## 17    KS     0.3246897
## 18    KY     0.3901622
## 19    LA     0.3443065
## 20    MA     0.3253065
## 21    MD     0.2281604
## 22    ME     0.2402761
## 23    MI     0.3110450
## 24    MN     0.3184199
## 25    MO     0.2917774
## 26    MP     0.2590430
## 27    MS     0.3361441
## 28    MT     0.3124681
```

```
## 29    NC     0.3338459
## 30    ND     0.3819807
## 31    NE     0.2973178
## 32    NH     0.2809516
## 33    NJ     0.3462289
## 34    NM     0.3248915
## 35    NV     0.2917096
## 36    NY     0.3515433
## 37    OH     0.2944160
## 38    OK     0.3301389
## 39    OR     0.2313696
## 40    PA     0.2780295
## 41    PR     0.3513411
## 42    RI     0.4167286
## 43    SC     0.3597052
## 44    SD     0.3197786
## 45    TN     0.3722504
## 46    TX     0.2934280
## 47    UT     0.3423236
## 48    VA     0.2707329
## 49    VT     0.2471887
## 50    WA     0.2570961
## 51    WI     0.3487746
## 52    WV     0.3630037
## 53    WY     0.3231639
```

```
data$estimated_recoveries <- data$actuals$cases - data$actuals$deaths
data$recovery_fraction <- data$estimated_recoveries / data$actuals$cases
print(data[, c('state', 'recovery_fraction')])
```

```
##    state recovery_fraction
## 1     AK         0.9950670
## 2     AL         0.9872658
## 3     AR         0.9868197
## 4     AZ         0.9864798
## 5     CA         0.9916840
## 6     CO         0.9919351
## 7     CT         0.9874411
## 8     DC         0.9919789
## 9     DE         0.9898314
## 10    FL         0.9883459
## 11    GA         0.9861445
## 12    HI         0.9950409
## 13    IA         0.9881213
## 14    ID         0.9895860
## 15    IL         0.9898457
## 16    IN         0.9872631
## 17    KS         0.9892021
## 18    KY         0.9893163
## 19    LA         0.9881476
## 20    MA         0.9890089
## 21    MD         0.9878642
## 22    ME         0.9905939
## 23    MI         0.9861983
## 24    MN         0.9917751
```

```
## 25    MO         0.9871948
## 26    MP         0.9968313
## 27    MS         0.9865316
## 28    MT         0.9888759
## 29    NC         0.9917008
## 30    ND         0.9913670
## 31    NE         0.9911516
## 32    NH         0.9920107
## 33    NJ         0.9882553
## 34    NM         0.9864424
## 35    NV         0.9866357
## 36    NY         0.9886643
## 37    OH         0.9877357
## 38    OK         0.9876320
## 39    OR         0.9902137
## 40    PA         0.9856583
## 41    PR         0.9947348
## 42    RI         0.9911318
## 43    SC         0.9892426
## 44    SD         0.9885788
## 45    TN         0.9883764
## 46    TX         0.9889777
## 47    UT         0.9951142
## 48    VA         0.9897254
## 49    VT         0.9937307
## 50    WA         0.9917324
## 51    WI         0.9918255
## 52    WV         0.9875107
## 53    WY         0.9891410
```

```
data$vaccination_fraction <- data$actuals$vaccinationsCompleted / data$population
print(data[, c('state', 'vaccination_fraction')])
```

```
##    state vaccination_fraction
## 1     AK            0.6528539
## 2     AL            0.5326320
## 3     AR            0.5700201
## 4     AZ            0.6623901
## 5     CA            0.7488553
## 6     CO            0.7377367
## 7     CT            0.8322138
## 8     DC                   NA
## 9     DE            0.7350724
## 10    FL            0.6970729
## 11    GA            0.5748709
## 12    HI                   NA
## 13    IA            0.6454579
## 14    ID            0.5664355
## 15    IL            0.7145669
## 16    IN            0.5787800
## 17    KS            0.6555757
## 18    KY            0.5974609
## 19    LA            0.5510335
## 20    MA            0.8451242
## 21    MD            0.7998091
```

```
## 22    ME            0.8362609
## 23    MI            0.6262716
## 24    MN            0.7238527
## 25    MO            0.5921785
## 26    MP            0.8216398
## 27    MS            0.5377298
## 28    MT            0.5927620
## 29    NC            0.6349518
## 30    ND            0.5888720
## 31    NE            0.6652475
## 32    NH                   NA
## 33    NJ            0.7930007
## 34    NM            0.7547325
## 35    NV            0.6384722
## 36    NY            0.8103061
## 37    OH            0.6065029
## 38    OK            0.6064674
## 39    OR            0.7268146
## 40    PA            0.7377354
## 41    PR            0.8417870
## 42    RI            0.8802863
## 43    SC            0.6005298
## 44    SD            0.6649534
## 45    TN            0.5641859
## 46    TX            0.6347911
## 47    UT            0.6694695
## 48    VA            0.7681303
## 49    VT            0.8609014
## 50    WA            0.7635120
## 51    WI            0.6838388
## 52    WV            0.5979085
## 53    WY            0.5315148
```

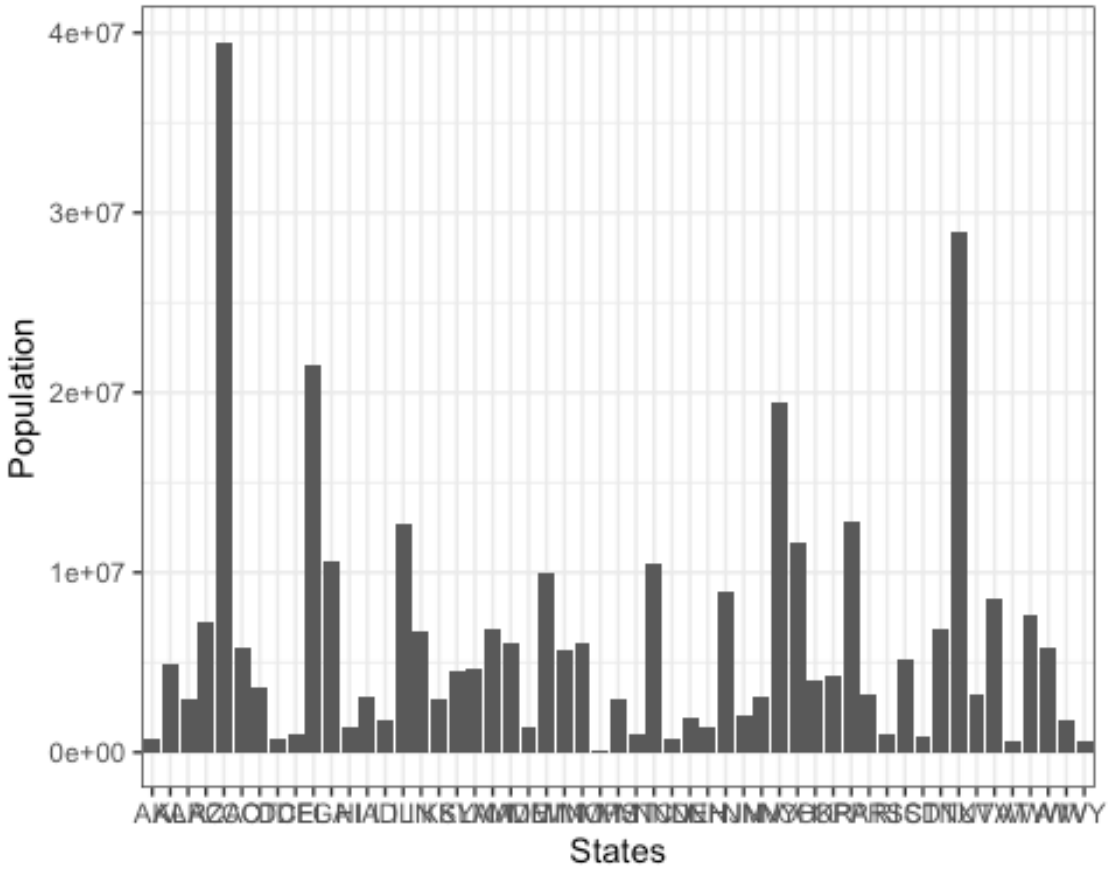
```
time_series <- data %>%
unnest(actualsTimeseries)

time_series_transmission <-
tibble(Date=time_series$cdcTransmissionLevelTimeseries[[which(data$state=="CA")]]$dat
e)
time_series_transmission$Alaska <-
time_series$cdcTransmissionLevelTimeseries[[which(data$state=="AK")]]$
cdcTransmissionLevel
time_series_transmission$California <-
time_series$cdcTransmissionLevelTimeseries[[which(data$state=="CA")]]$cdcTransmission
Level
time_series_transmission$New_Jersey <-
time_series$cdcTransmissionLevelTimeseries[[which(data$state=="NJ")]]$cdcTransmission
Level
time_series_transmission$Tennessee <-
time_series$cdcTransmissionLevelTimeseries[[which(data$state=="TN")]]$cdcTransmission
Level
time_series_transmission$District_of_Columbia <-
time_series$cdcTransmissionLevelTimeseries[[which(data$state=="DC")]]$cdcTransmission
Level
 print(head(time_series_transmission))
```

```
## # A tibble: 6 × 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
```

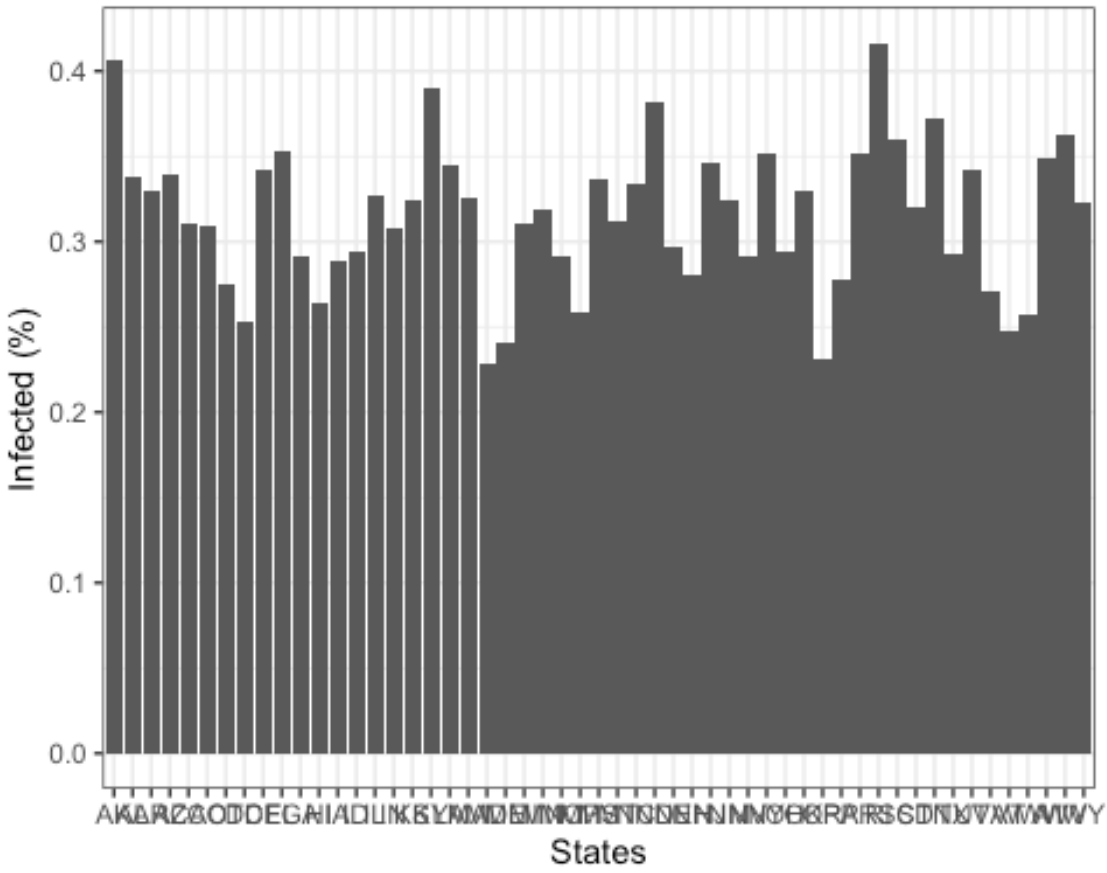
```
time_series_cases <- list(Alaska = time_series %>% filter(state=="AK") %>%
select(date,cases))
# Cases of each state
time_series_cases$California <- time_series %>% filter(state=="CA") %>%
select(date,cases)
time_series_cases$New_Jersey <- time_series %>% filter(state=="NJ") %>%
select(date,cases)
time_series_cases$Tennessee <- time_series %>% filter(state=="TN") %>%
select(date,cases)
time_series_cases$District_of_Columbia <- time_series %>% filter(state=="DC") %>%
select(date,cases)
```

STEP-6: Analysing Data

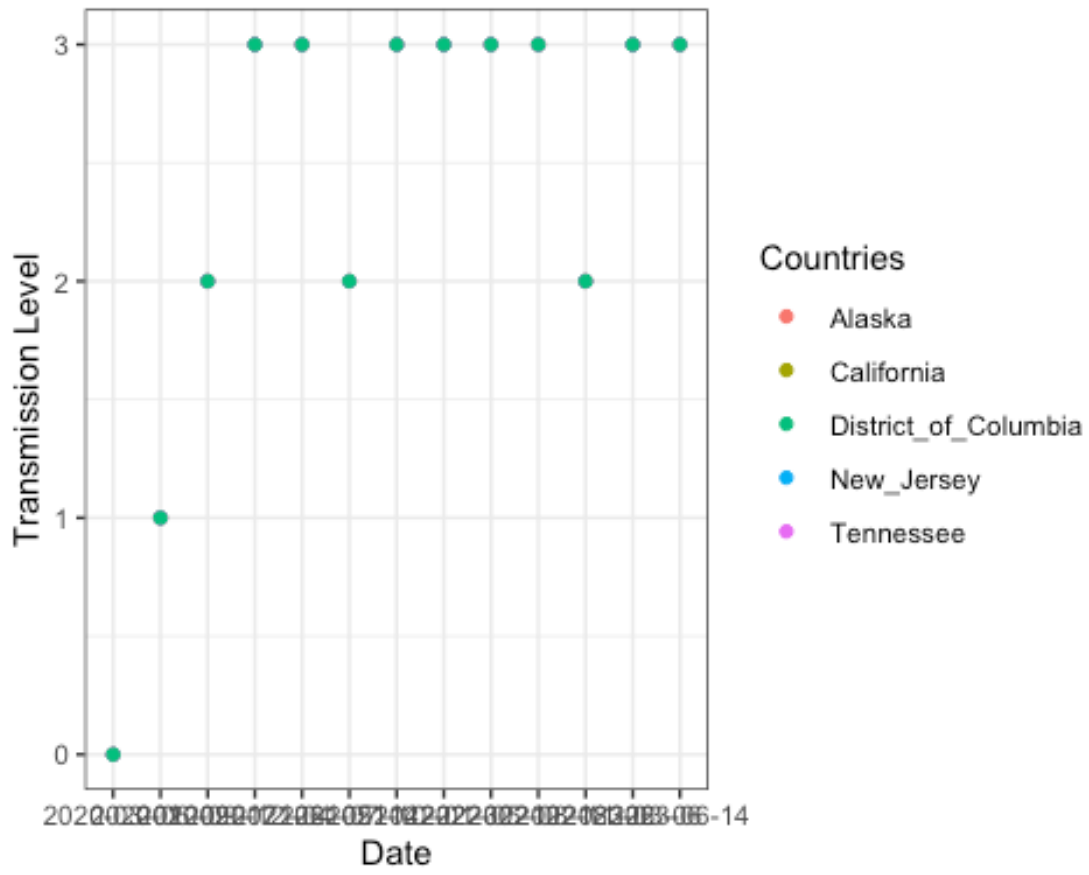
```
ggplot(data, aes(x=state,y=population)) + geom_bar(stat="identity") +
labs(x="States",y="Population") + theme_bw()
```



```
ggplot(data, aes(x=state,y=(data$actuals$cases/population))) +
geom_bar(stat="identity") + labs(x="States",y="Infected (%)")+theme_bw()
```




```
time_series_transmission[seq(1,1300,by=100),]%>%

pivot_longer(cols=Alaska:District_of_Columbia,names_to="Countries",values_to="Transmi
ssion") %>%
 ggplot(aes(x=Date,y=Transmission,colour=Countries,group=Countries)) +
 geom_point(show.legend=TRUE) + labs(x="Date",y="Transmission Level")+theme_bw()
```




```
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)])

data_to_plot

## # A tibble: 13 × 10
##    Date_Alaska Cases_Alaska Date_California Cases_California Date_New_Jersey
##    <chr>              <int> <chr>                      <int> <chr>
##  1 2020-03-01            NA 2020-01-25                     1 2020-03-01
##  2 2020-06-09           620 2020-05-04                 56333 2020-06-09
##  3 2020-09-17          7413 2020-08-12                595097 2020-09-17
##  4 2020-12-26         45247 2020-11-20               1096427 2020-12-26
```

```
##  5 2021-04-05         63486 2021-02-28               3569578 2021-04-05
##  6 2021-07-14         71539 2021-06-08               3798225 2021-07-14
##  7 2021-10-22        132393 2021-09-16               4629146 2021-10-22
##  8 2022-01-30        211117 2021-12-25               5291605 2022-01-30
##  9 2022-05-10        252847 2022-04-04               9110544 2022-05-10
## 10 2022-08-18        289203 2022-07-13              10365785 2022-08-18
## 11 2022-11-26        299841 2022-10-21              11338846 2022-11-26
## 12 2023-03-06        307377 2023-01-29              11980312 2023-03-06
## 13 2023-06-14            NA 2023-05-09              12242634 2023-06-14
## # ℹ 5 more variables: Cases_New_Jersey <int>, Date_Tennessee <chr>,
## #   Cases_Tennessee <int>, Date_District_of_Columbia <chr>,
## #   Cases_District_of_Columbia <int>
```

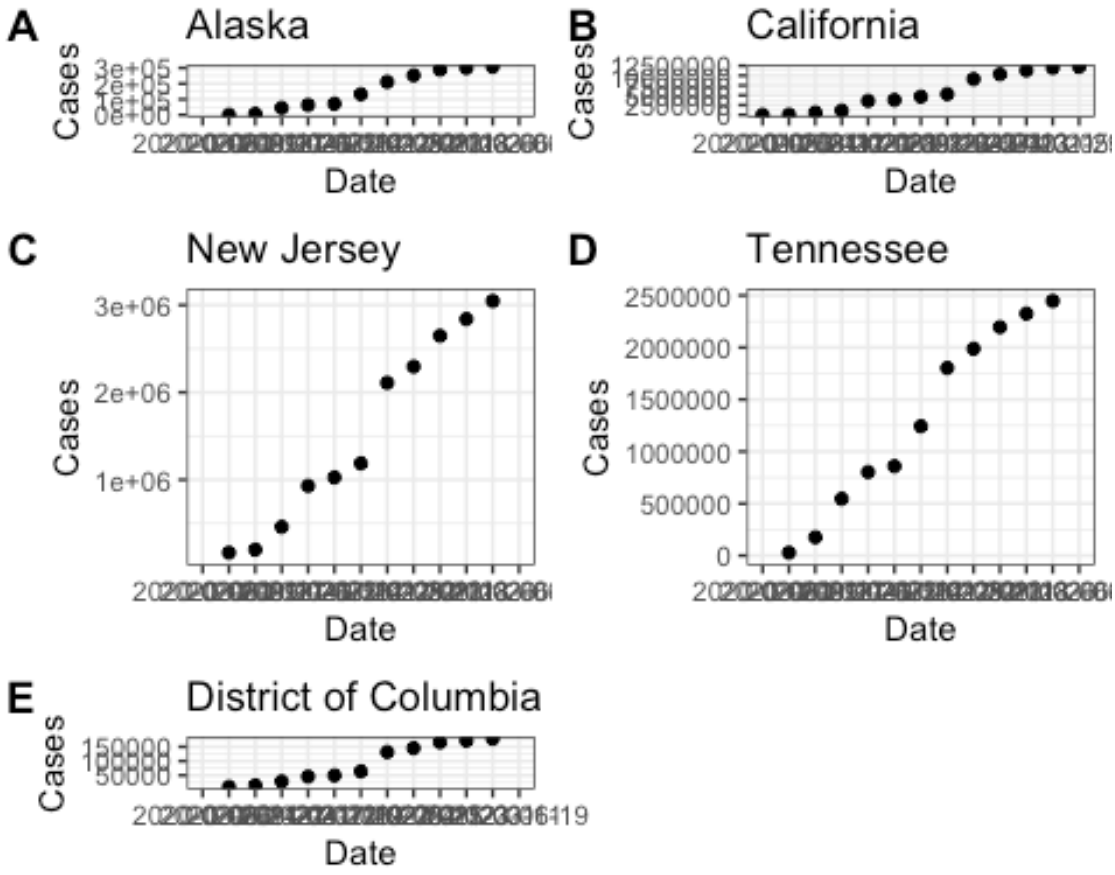
```
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

```
## The following object is masked from 'package:lubridate':
##
##     stamp
```

```
fig1<- ggplot(data_to_plot, aes(x=Date_Alaska,y=Cases_Alaska)) +
 geom_point() + labs(x="Date",y="Cases", title="Alaska") + theme_bw()
 fig2<- ggplot(data_to_plot, aes(x=Date_California,y=Cases_California)) +
 geom_point() + labs(x="Date",y="Cases", title="California") + theme_bw()
 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()
 fig4<- ggplot(data_to_plot, aes(x=Date_Tennessee,y=Cases_Tennessee)) +
 geom_point() + labs(x="Date",y="Cases", title="Tennessee") + theme_bw()
 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()

 plot_grid(fig1 + theme(legend.justification = c(0,1)),
           fig2 + theme(legend.justification = c(1,0)),
           fig3 + theme(legend.justification = c(0,1)),
           fig4 + theme(legend.justification = c(1,0)),
           fig5 + theme(legend.justification = c(0,1)),
           align = "v", axis = "lr", nrow=3,
           ncol = 2,labels = LETTERS[1:5],
           rel_heights = c(1,2))
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_point()`).
## Removed 2 rows containing missing values (`geom_point()`).
```


A
Alaska
Cases
3e+05
2e+05
1e+05
0e+00
Date
B
California
Cases
12500000
0
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
50000
Date