Challenge-11

Hariz Emran

2023-10-30

Step 1: Procuring Data

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

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()  masks stats::filter()
## x purrr::flatten() masks jsonlite::flatten()
## x 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=9b75214f4f3041
raw_data <- GET(historic_state_data_url)
```

Extracting data

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

Exploring 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> 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[~
```

Mapping Variables to Questions

```
data$state
```

```
##  [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MP" "MS" "MT" "NC" "ND"
## [31] "NE" "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "PR" "RI" "SC" "SD" "TN"
## [46] "TX" "UT" "VA" "VT" "WA" "WI" "WV" "WY"
```

```
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
```

```
data$actuals$cases
```

```
##  [1]   297588  1659936   995043  2474154 12251820  1783663   982335   178904
##  [9]   333772  7572282  3087729   374264   908936   526118  4136659  2076326
## [17]   945923  1743117  1600610  2242176  1379385   322982  3106362  1795771
## [25]  1790763    13886  1000415   333959  3501404   291093   575134   382013
## [33]  3075271   681242   898511  6838769  3441458  1306350   975856  3559331
## [41]  1122076   441466  1852019   282895  2542163  8508204  1097475  2310846
## [49]   154243  1957759  2030717   650556   187034
```

```
data$state
```

```
##  [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MP" "MS" "MT" "NC" "ND"
## [31] "NE" "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "PR" "RI" "SC" "SD" "TN"
## [46] "TX" "UT" "VA" "VT" "WA" "WI" "WV" "WY"
```

```
data$actuals$cases - data$actuals$deaths
```

```
##  [1]   296120  1638798   981928  2440703 12149934  1769278   969998   177469
##  [9]   330378  7484034  3044947   372408   898139   520639  4094654  2049880
## [17]   935709  1724494  1581639  2217532  1362645   319944  3063489  1781001
## [25]  1767832    13842   986941   330244  3472345   288580   570045   378961
## [33]  3039153   672006   886503  6761247  3399251  1290193   966306  3508284
## [41]  1116168   437551  1832096   279664  2512614  8414424  1092113  2287103
## [49]   153276  1941573  2014117   642431   185003
```

```
data$state
```

```
##  [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MP" "MS" "MT" "NC" "ND"
## [31] "NE" "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "PR" "RI" "SC" "SD" "TN"
## [46] "TX" "UT" "VA" "VT" "WA" "WI" "WV" "WY"
```

```
data$actuals$vaccinationsCompleted
```

```
##  [1]   477592  2611593  1720209  4821350 29588939  4248431  2967081       NA
##  [9]   715787 14971549  6103647       NA  2036465  1012257  9054864  3896474
## [17]  1909898  2669260  2561641  5825021  4835390  1124112  6254485  4082263
## [25]  3634453    44044  1600364   633531  6659428   448757  1286860       NA
## [33]  7043583  1582545  1966594 15763340  7089473  2399774  3065513  9444480
## [41]  2688410   932541  3091956   588257  3852924 18406327  2146291  6556391
## [49]   537193  5814062  3981606  1071540   307619
```

```
data$state
```

```
##  [1] "AK" "AL" "AR" "AZ" "CA" "CO" "CT" "DC" "DE" "FL" "GA" "HI" "IA" "ID" "IL"
## [16] "IN" "KS" "KY" "LA" "MA" "MD" "ME" "MI" "MN" "MO" "MP" "MS" "MT" "NC" "ND"
## [31] "NE" "NH" "NJ" "NM" "NV" "NY" "OH" "OK" "OR" "PA" "PR" "RI" "SC" "SD" "TN"
## [46] "TX" "UT" "VA" "VT" "WA" "WI" "WV" "WY"
```

```
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
```

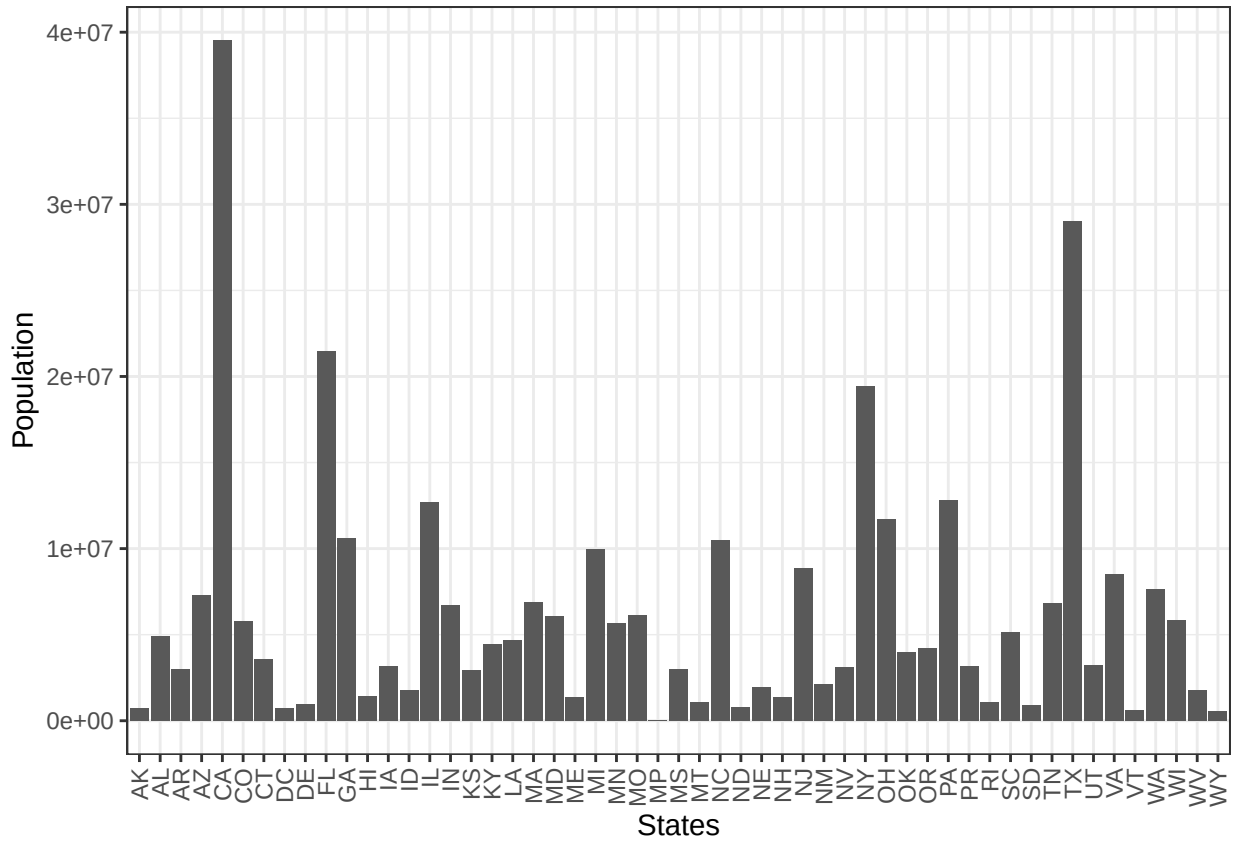
```
time_series_transmission$Alaska <-
  time_series$cdcTransmissionLevelTimeseries[[which(data$state=="AK")]]$ cdcTransmissionLevel

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))
```

```
## # 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
```

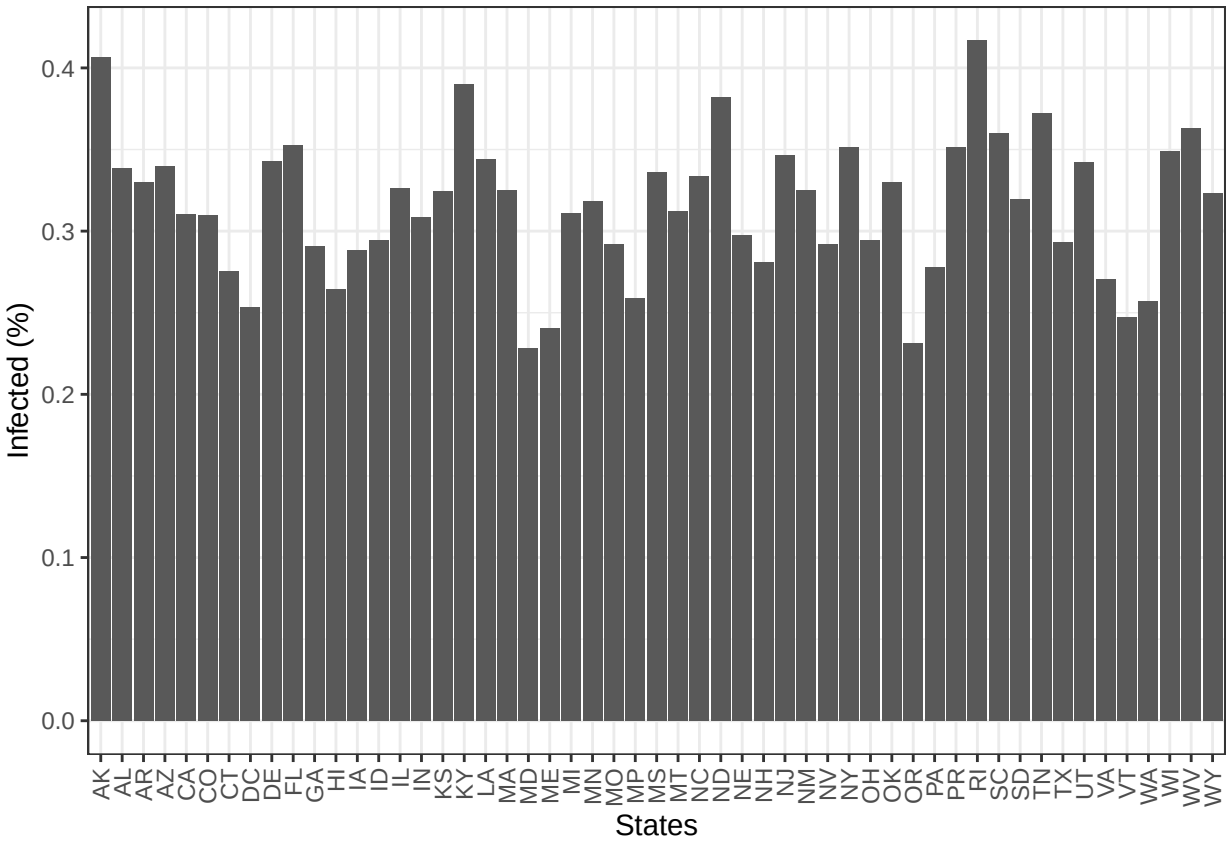
```
# New data-frame with dates
time_series_cases <- list(Alaska = time_series %>% filter(state=="AK") %>% select(date,cases))

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

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)
```

Analysing Data

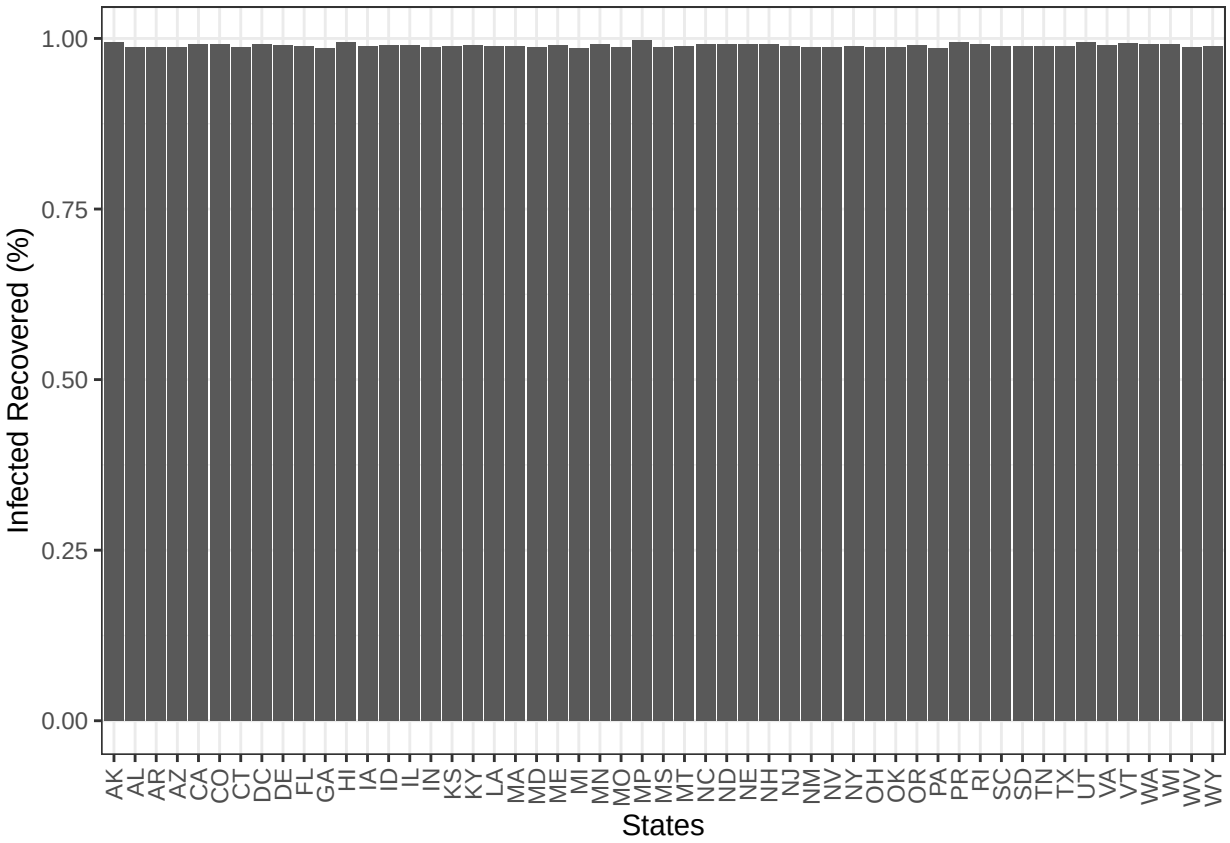
```
ggplot(data, aes(x=state,y=population)) +
  geom_bar(stat="identity") +
  labs(x="States",y="Population") +
  theme_bw() + scale_x_discrete(guide = guide_axis(angle=90))
```



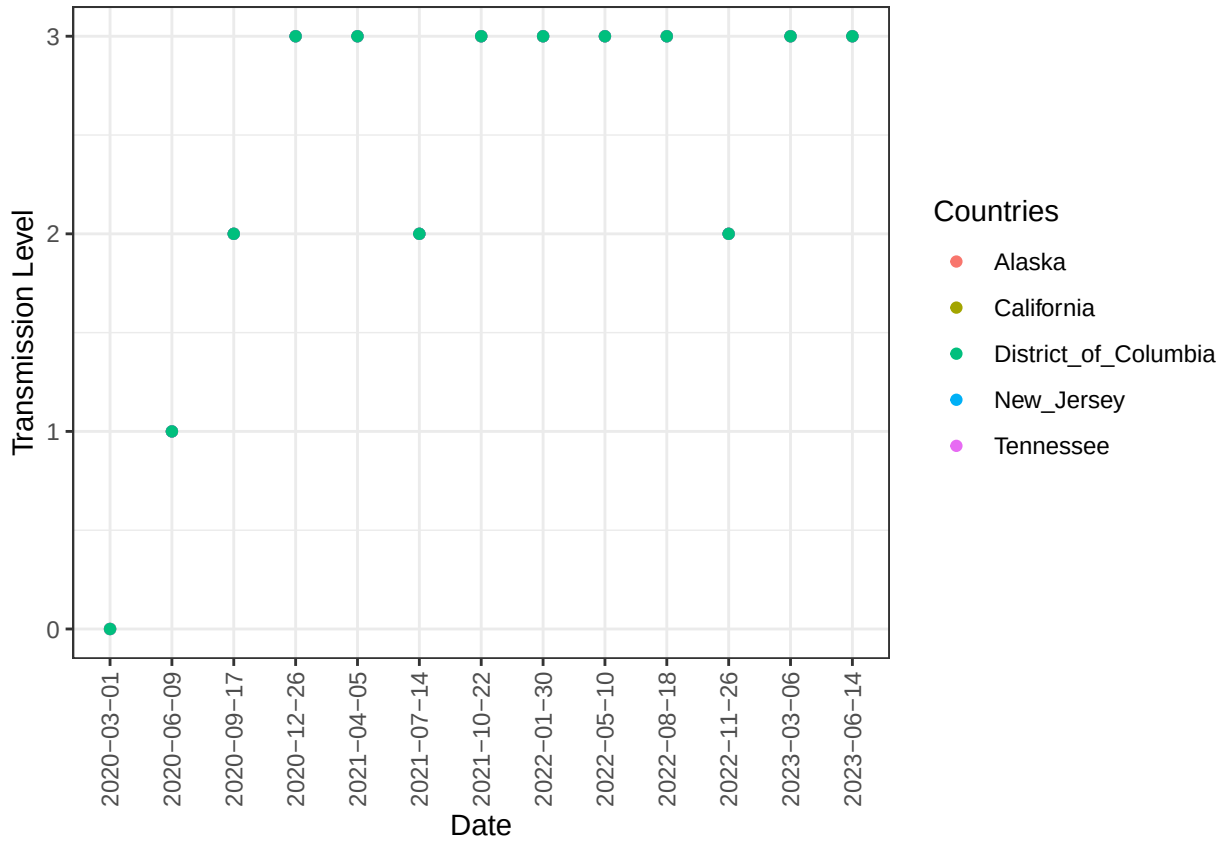

```
ggplot(data, aes(x=state,y=(data$actuals$cases/population))) +
  geom_bar(stat="identity") +
  labs(x="States",y="Infected (%)") +
  theme_bw() + scale_x_discrete(guide = guide_axis(angle=90))
```




```
ggplot(data, aes(x=state,y=((data$actuals$cases - data$actuals$deaths)/data$actuals$cases))) +
  geom_bar(stat="identity") +
  labs(x="States",y="Infected Recovered (%)") +
  theme_bw() + scale_x_discrete(guide = guide_axis(angle=90))
```




```
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() + scale_x_discrete(guide = guide_axis(angle=90))
```




```
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(1,130
                       Cases_District_of_Columbia = time_series_cases$District_of_Columbia$cases[seq(1,13

data_to_plot
```

```
## # A tibble: 13 x 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>
```

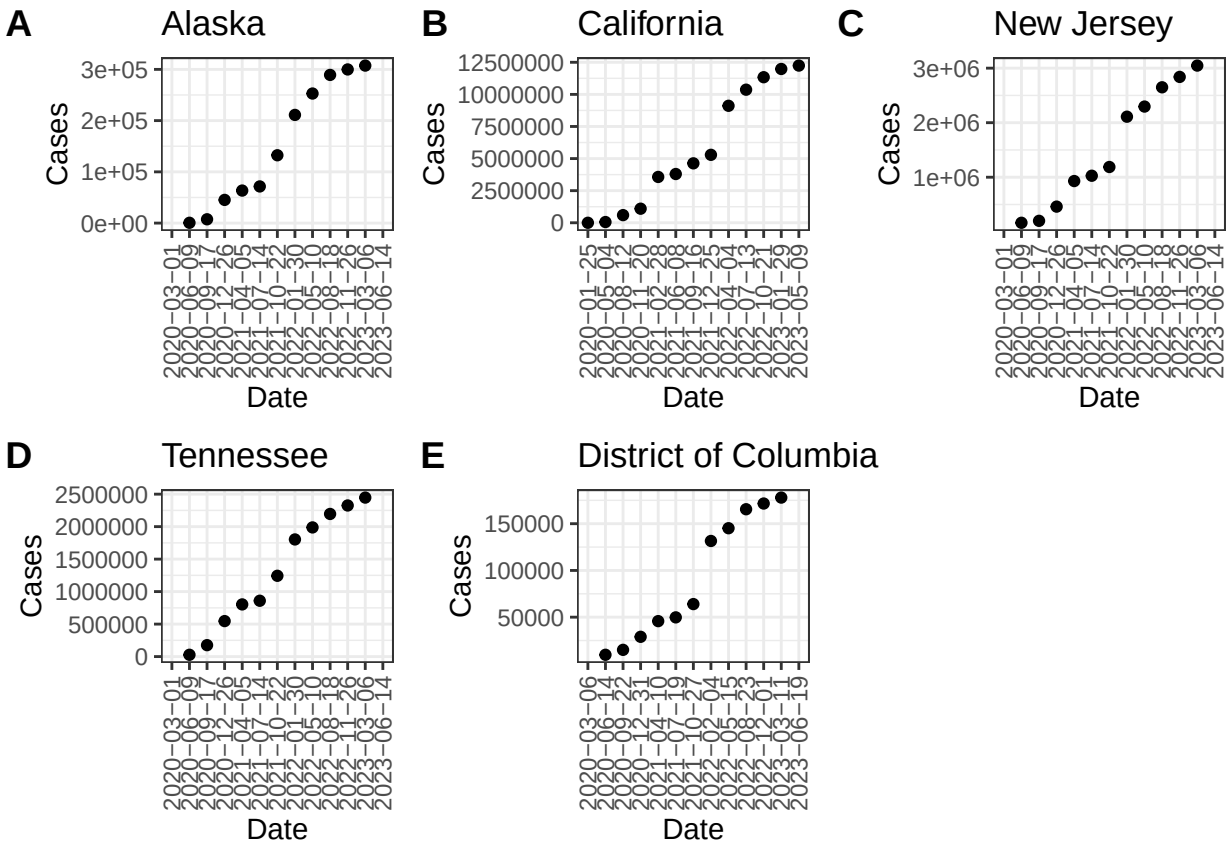
```
library(cowplot)
```

```
##
## 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() + scale_x_discrete(guide = guide_a

fig2<- ggplot(data_to_plot, aes(x=Date_California,y=Cases_California)) +
 geom_point() + labs(x="Date",y="Cases", title="California") + theme_bw() + scale_x_discrete(guide = gui

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() + scale_x_discrete(guide = gui

fig4<- ggplot(data_to_plot, aes(x=Date_Tennessee,y=Cases_Tennessee)) +
 geom_point() + labs(x="Date",y="Cases", title="Tennessee") + theme_bw() + scale_x_discrete(guide = gui

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() + scale_x_discrete(guide = guide_axis(angle=90))

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= 2,
          ncol = 3,labels = LETTERS[1:5],
          rel_heights = c(1,1))
```

```
## 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
2020−03−01
2020−06−09
2020−09−17
2020−12−26
2021−04−05
2021−07−14
2021−10−22
2022−01−30
2022−05−10
2022−08−18
2022−11−26
2023−03−06
2023−06−14
Date
B
California
Cases
12500000
10000000
7500000
5000000
2500000
0
2020−01−25
2020−05−04
2020−08−12
2020−11−20
2021−02−28
2021−06−08
2021−09−16
2021−12−25
2022−04−04
2022−07−13
2022−10−21
2023−01−29
2023−05−09
Date
C
New Jersey
Cases
3e+06
2e+06
1e+06
2020−03−01
2020−06−09
2020−09−17
2020−12−26
2021−04−05
2021−07−14
2021−10−22
2022−01−30
2022−05−08
2022−08−18
2022−11−26
2023−03−06
2023−06−14
Date
D
Tennessee
Cases
2500000
2000000
1500000
1000000
500000
0
2020−03−01
2020−06−09
2020−09−17
2020−12−26
2021−04−05
2021−07−14
2021−10−22
2022−01−30
2022−05−10
2022−08−18
2022−11−26
2023−03−06
2023−06−14
Date
E
District of Columbia
Cases
150000
100000
50000
2020−03−06
2020−06−14
2020−09−22
2020−12−31
2021−04−10
2021−07−19
2021−10−27
2022−02−04
2022−05−15
2022−08−23
2022−12−01
2023−03−11
2023−06−19
Date