Case Study

Rifa

12/6/2020

Importing various libraries

```
library(data.table)
library(magrittr)
library(ggplot2)
```

Load all the tables

```
tables = list.files("data/", full.names = T)
tables
```

```
##  [1] "data//Case.csv"          "data//PatientInfo.csv"
##  [3] "data//Policy.csv"        "data//Region.csv"
##  [5] "data//SearchTrend.csv"   "data//SeoulFloating.csv"
##  [7] "data//Time.csv"          "data//TimeAge.csv"
##  [9] "data//TimeGender.csv"    "data//TimeProvince.csv"
## [11] "data//Weather.csv"
```

Read individual csv files

```
case_dt = fread(tables[1])
pinfo_dt = fread(tables[2])
policy_dt = fread(tables[3])
region_dt = fread(tables[4])
search_dt = fread(tables[5])
seoul_dt = fread(tables[6])
time_dt = fread(tables[7])
tage_dt = fread(tables[8])
tgender_dt = fread(tables[9])
tprovince_dt = fread(tables[10])
weather_dt = fread(tables[11])
```

WORKING WITH TIME_PROVINCE FILE

```
head(tprovince_dt,n=5)
```

```
##          date time province confirmed released deceased
## 1: 2020-01-20   16    Seoul         0        0        0
## 2: 2020-01-20   16    Busan         0        0        0
## 3: 2020-01-20   16    Daegu         0        0        0
## 4: 2020-01-20   16  Incheon         1        0        0
## 5: 2020-01-20   16  Gwangju         0        0        0
```

```
summary(tprovince_dt)
```

```
##       date                 time          province           confirmed
##  Min.   :2020-01-20   Min.   : 0.000   Length:2771        Min.   :   0.0
##  1st Qu.:2020-02-29   1st Qu.: 0.000   Class :character   1st Qu.:   9.0
##  Median :2020-04-10   Median : 0.000   Mode  :character   Median :  42.0
##  Mean   :2020-04-10   Mean   : 4.123                      Mean   : 444.3
##  3rd Qu.:2020-05-21   3rd Qu.:16.000                      3rd Qu.: 133.0
##  Max.   :2020-06-30   Max.   :16.000                      Max.   :6906.0
##     released         deceased
##  Min.   :   0.0   Min.   :  0.00
##  1st Qu.:   1.0   1st Qu.:  0.00
##  Median :  21.0   Median :  0.00
##  Mean   : 320.7   Mean   :  9.24
##  3rd Qu.:  92.0   3rd Qu.:  1.00
##  Max.   :6700.0   Max.   :189.00
```

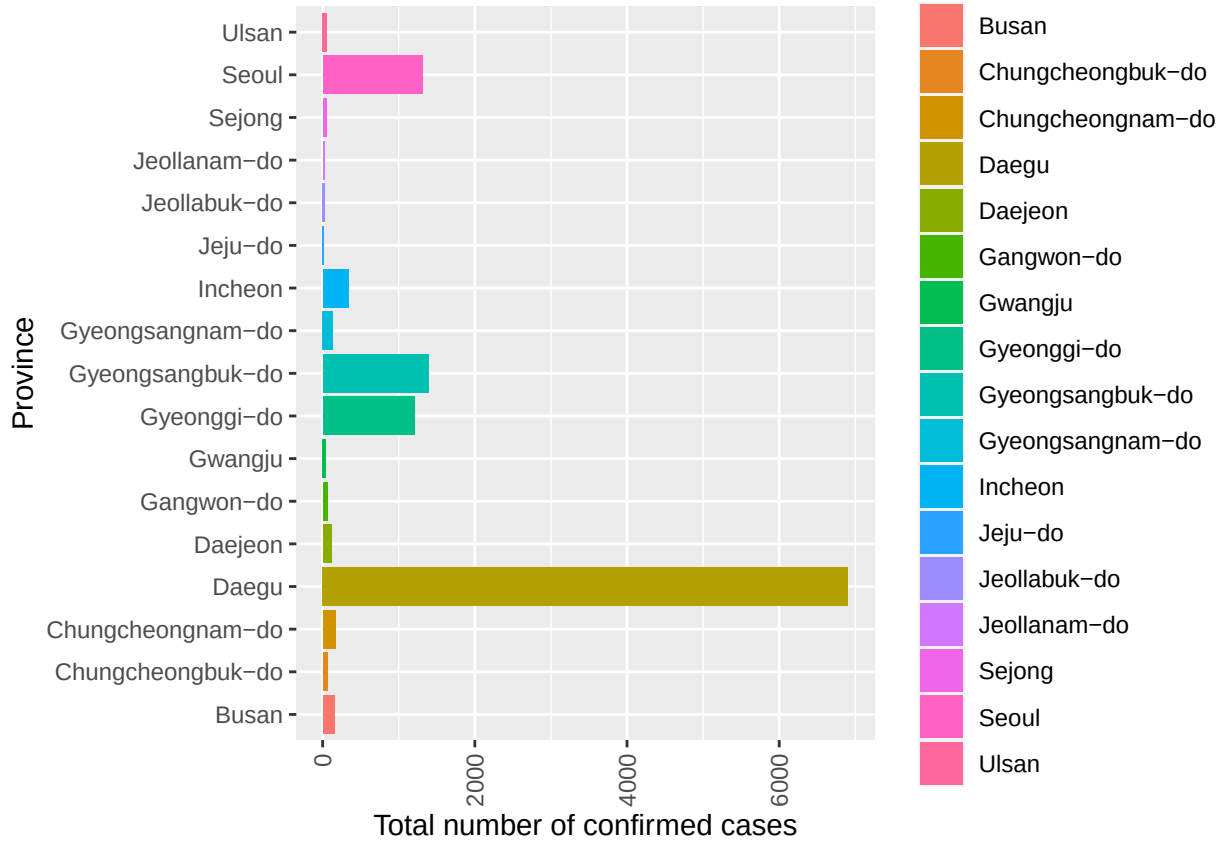
```
#Printing the name of the various provinces
tprovince_dt[, unique(province)]
```

```
##  [1] "Seoul"             "Busan"             "Daegu"
##  [4] "Incheon"           "Gwangju"           "Daejeon"
##  [7] "Ulsan"             "Sejong"            "Gyeonggi-do"
## [10] "Gangwon-do"        "Chungcheongbuk-do" "Chungcheongnam-do"
## [13] "Jeollabuk-do"      "Jeollanam-do"      "Gyeongsangbuk-do"
## [16] "Gyeongsangnam-do"  "Jeju-do"
```

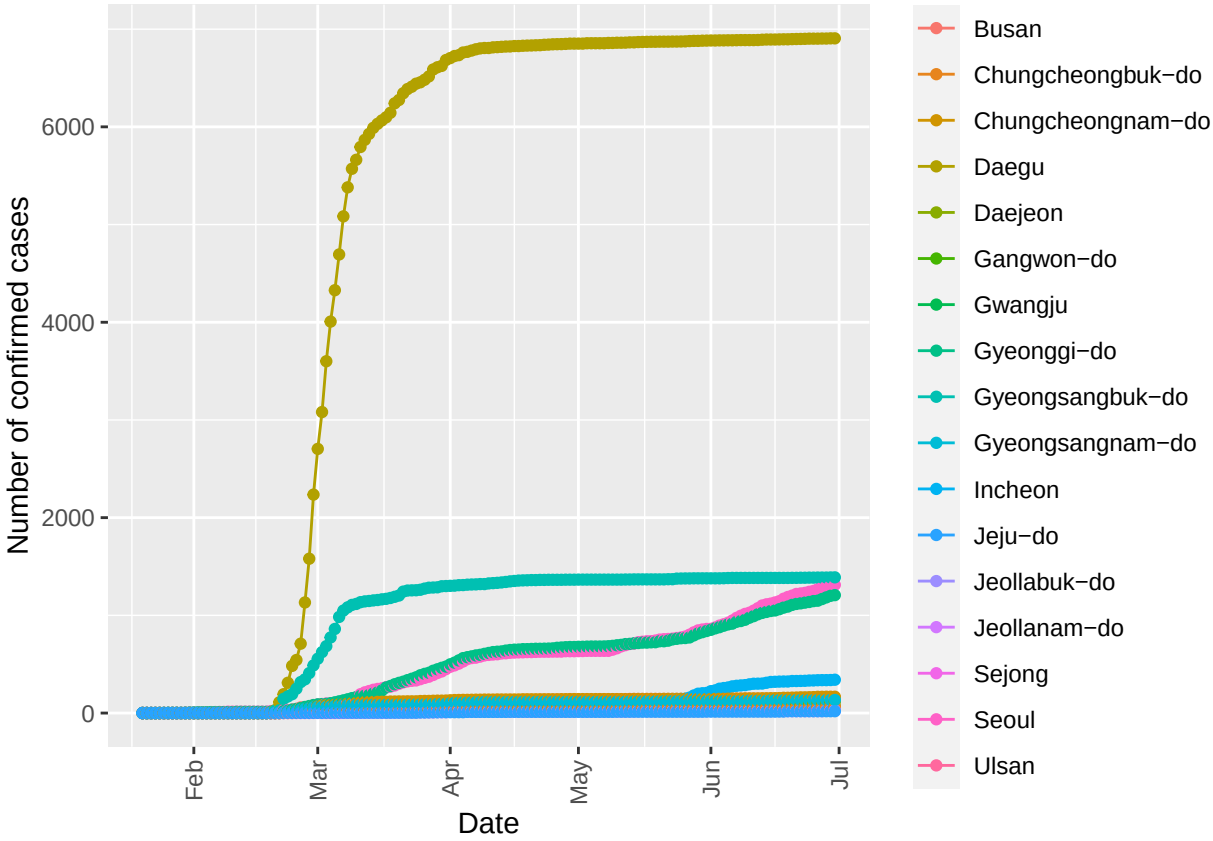
```
# Making a new table with province and the corresponding total number of cases
province_cases<-tprovince_dt[, .(number_of_cases=max(confirmed)), by='province']
province_cases
```

```
##              province number_of_cases
##  1:             Seoul            1312
##  2:             Busan             154
##  3:             Daegu            6906
##  4:           Incheon             341
##  5:           Gwangju              44
##  6:           Daejeon             117
##  7:             Ulsan              55
##  8:            Sejong              50
##  9:       Gyeonggi-do            1207
## 10:        Gangwon-do              65
## 11: Chungcheongbuk-do              65
## 12: Chungcheongnam-do             167
## 13:      Jeollabuk-do              27
## 14:      Jeollanam-do              24
## 15:  Gyeongsangbuk-do            1389
## 16:  Gyeongsangnam-do             134
## 17:           Jeju-do              19
```

```
# Plotting a bar graph for number of total cases for various provinces
ggplot(province_cases, aes(x=province, y=number_of_cases , fill=province)) + geom_bar(stat='identity') +
labs(x='Province', y='Total number of confirmed cases')
```




```
# Plotting growth in the total number of cases for various provinces
ggplot(tprovince_dt, aes(x=date, y=confirmed, color=province)) + geom_line() + geom_point() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x='Date', y='Number of confirmed cases')
```



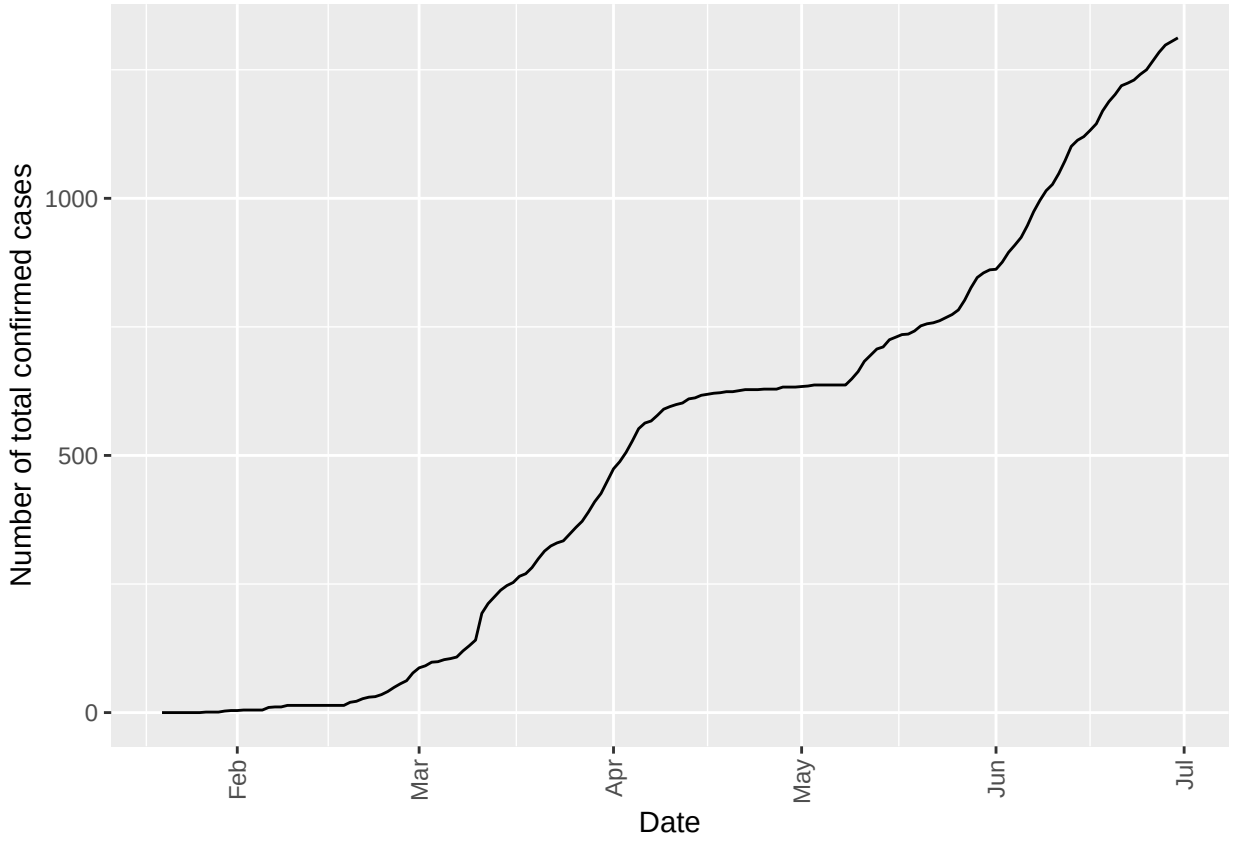
The number of cases is quite high (in thousands) for some provinces and comparatively smaller (in hundreds or less) for other provinces.

Growth in total number of cases for various provinces

```
tprovince_Seoul=tprovince_dt[province=="Seoul"]
#Busan
#Daegu
#Incheon
#Gwangju
#Daejeon
#Ulsan
#Sejong
#Gyeonggi-do
#Gangwon-do
#Chungcheongbuk-do
#Chungcheongnam-do
#Jeollabuk-do
#Jeollanam-do
#Gyeongsangbuk-do
#Gyeongsangnam-do
#Jeju-do
```

We look at various provinces individually

```
# Plotting total number of cases for Seoul
ggplot(tprovince_Seoul, aes(x=date, y=confirmed)) + geom_line() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Date", y="Number of total confirmed cases")
```



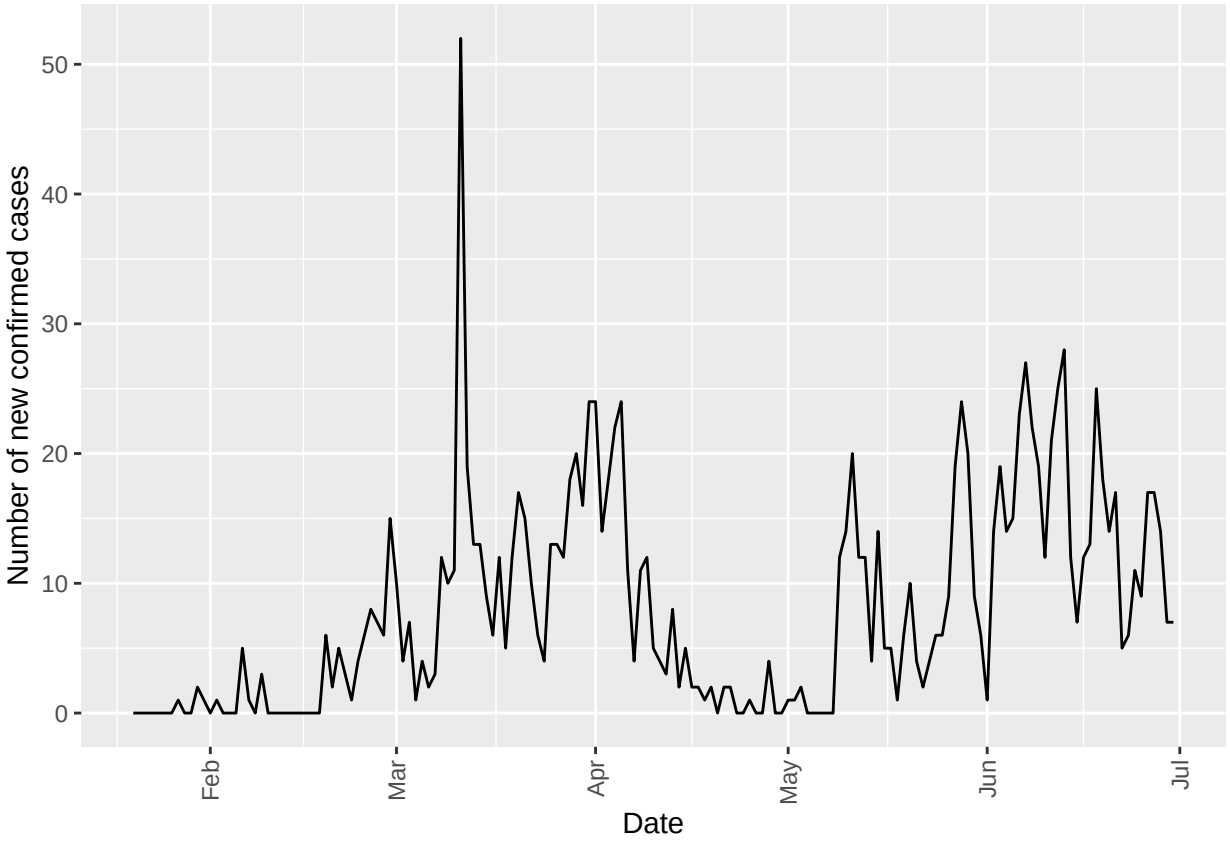
Seoul

The same could be done for other provinces.

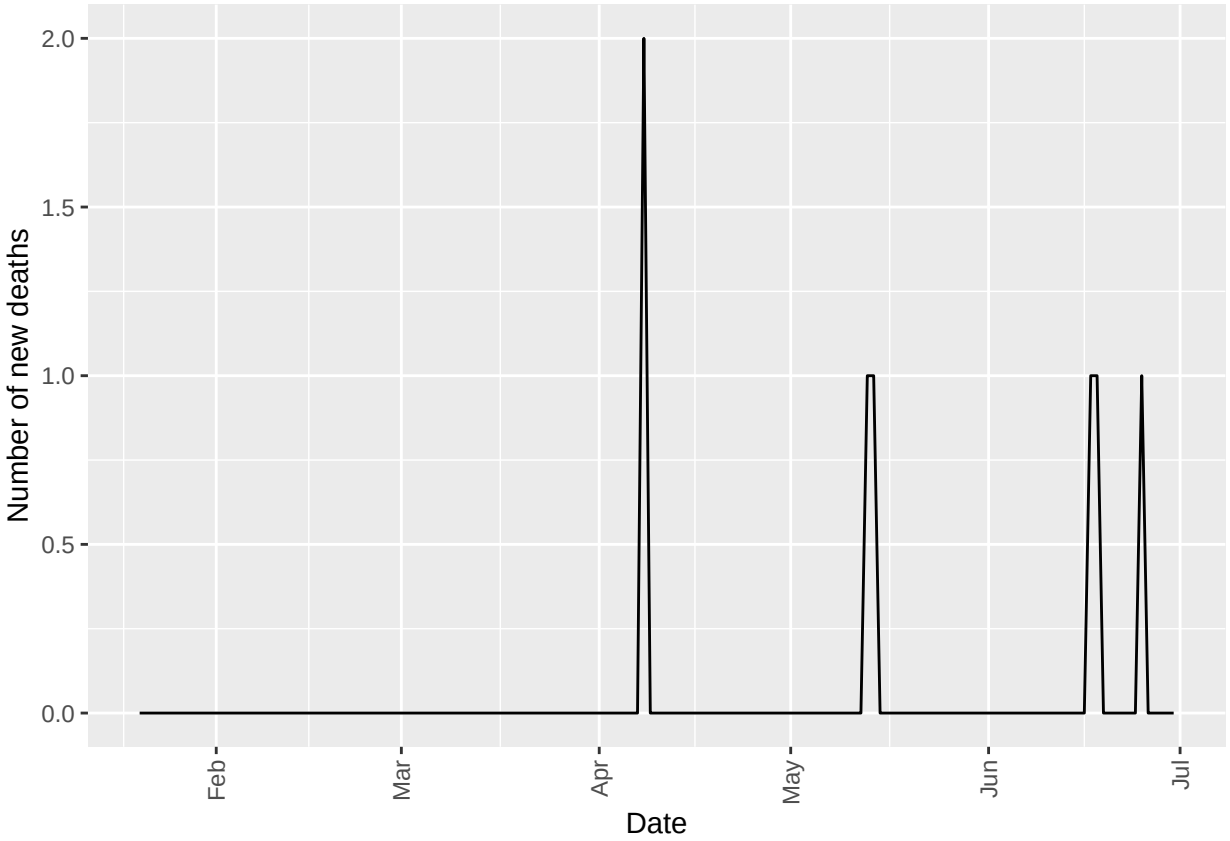
Visualizing the number of new cases and new deaths per day for various provinces Seoul

```
#Adding column of daily new cases and daily new deaths
tprovince_Seoul[,daily_new_case:=confirmed-shift(confirmed,fill=first(confirmed))]
tprovince_Seoul[, daily_new_deaths:=  deceased-shift(deceased,fill=first(deceased)) ]
```

```
#Plotting daily new cases for Seoul
ggplot(tprovince_Seoul, aes(x=date, y=daily_new_case)) + geom_line() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Date", y="Number of new confirmed cases")
```




```
#Plotting daily new deaths for Seoul
ggplot(tprovince_Seoul, aes(x=date, y=daily_new_deaths)) + geom_line() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Date", y="Number of new deaths")
```



Analyzing effect of temp on number of covid cases

Weather file

```
head(weather_dt, n=5)
```

```
##     code province       date avg_temp min_temp max_temp precipitation
## 1: 10000    Seoul 2016-01-01      1.2     -3.3      4.0             0
## 2: 11000    Busan 2016-01-01      5.3      1.1     10.9             0
## 3: 12000    Daegu 2016-01-01      1.7     -4.0      8.0             0
## 4: 13000  Gwangju 2016-01-01      3.2     -1.5      8.1             0
## 5: 14000  Incheon 2016-01-01      3.1     -0.4      5.7             0
##    max_wind_speed most_wind_direction avg_relative_humidity
## 1:            3.5                  90                  73.0
## 2:            7.4                 340                  52.1
## 3:            3.7                 270                  70.5
## 4:            2.7                 230                  73.1
## 5:            5.3                 180                  83.9
```

```
summary(weather_dt)
```

```
##       code         province             date              avg_temp
##  Min.   :10000   Length:26271       Min.   :2016-01-01   Min.   :-14.80
##  1st Qu.:13500   Class :character   1st Qu.:2017-02-14   1st Qu.:  6.00
##  Median :20000   Mode  :character   Median :2018-04-01   Median : 14.60
```

```
##  Mean   :32125                    Mean   :2018-03-31   Mean   : 13.86
##  3rd Qu.:50500                    3rd Qu.:2019-05-16   3rd Qu.: 21.90
##  Max.   :70000                    Max.   :2020-06-29   Max.   : 33.90
##                                                        NA's   :15
##     min_temp          max_temp      precipitation     max_wind_speed
##  Min.   :-19.200   Min.   :-11.90   Min.   :  0.000   Min.   : 1.000
##  1st Qu.:  1.400   1st Qu.: 10.90   1st Qu.:  0.000   1st Qu.: 3.800
##  Median :  9.900   Median : 19.80   Median :  0.000   Median : 4.700
##  Mean   :  9.665   Mean   : 18.78   Mean   :  1.487   Mean   : 5.109
##  3rd Qu.: 18.200   3rd Qu.: 26.70   3rd Qu.:  0.000   3rd Qu.: 6.000
##  Max.   : 30.300   Max.   : 40.00   Max.   :266.000   Max.   :29.400
##  NA's   :5         NA's   :3                          NA's   :9
##  most_wind_direction avg_relative_humidity
##  Min.   : 20.0       Min.   : 10.4
##  1st Qu.: 90.0       1st Qu.: 53.6
##  Median :200.0       Median : 66.9
##  Mean   :195.9       Mean   : 65.7
##  3rd Qu.:290.0       3rd Qu.: 78.6
##  Max.   :360.0       Max.   :100.0
##  NA's   :29          NA's   :20
```

```
weather_Seoul=weather_dt[province == "Seoul"]
weather_Seoul
```

Getting weather tables for various provinces (Currently only for Seoul province)

```
##        code province       date avg_temp min_temp max_temp precipitation
##    1: 10000    Seoul 2016-01-01      1.2     -3.3      4.0           0.0
##    2: 10000    Seoul 2016-01-02      5.7      1.0      9.5           0.0
##    3: 10000    Seoul 2016-01-03      6.5      5.1      9.4           0.0
##    4: 10000    Seoul 2016-01-04      2.0     -2.5      5.3           0.0
##    5: 10000    Seoul 2016-01-05     -2.7     -4.8      1.5           0.0
##   ---
## 1638: 10000    Seoul 2020-06-25     21.6     20.1     23.4          13.3
## 1639: 10000    Seoul 2020-06-26     21.8     19.2     25.7           2.1
## 1640: 10000    Seoul 2020-06-27     24.1     20.4     29.6           0.0
## 1641: 10000    Seoul 2020-06-28     25.2     21.5     30.1           0.0
## 1642: 10000    Seoul 2020-06-29     23.8     20.6     26.5          11.9
##       max_wind_speed most_wind_direction avg_relative_humidity
##    1:            3.5                  90                  73.0
##    2:            4.5                 320                  76.9
##    3:            4.0                 320                  80.6
##    4:            5.1                 320                  54.4
##    5:            4.6                  20                  39.4
##   ---
## 1638:            4.3                 180                  91.0
## 1639:            5.3                 230                  82.1
## 1640:            5.4                 250                  70.4
## 1641:            4.7                 270                  70.5
## 1642:            5.9                  50                  77.0
```

```
#Listing various columns of both tables
colnames(tprovince_Seoul)
```

Combining Weather and time info for Seoul province

```
## [1] "date"             "time"             "province"         "confirmed"
## [5] "released"         "deceased"         "daily_new_case"   "daily_new_deaths"
```

```
colnames(weather_Seoul)
```

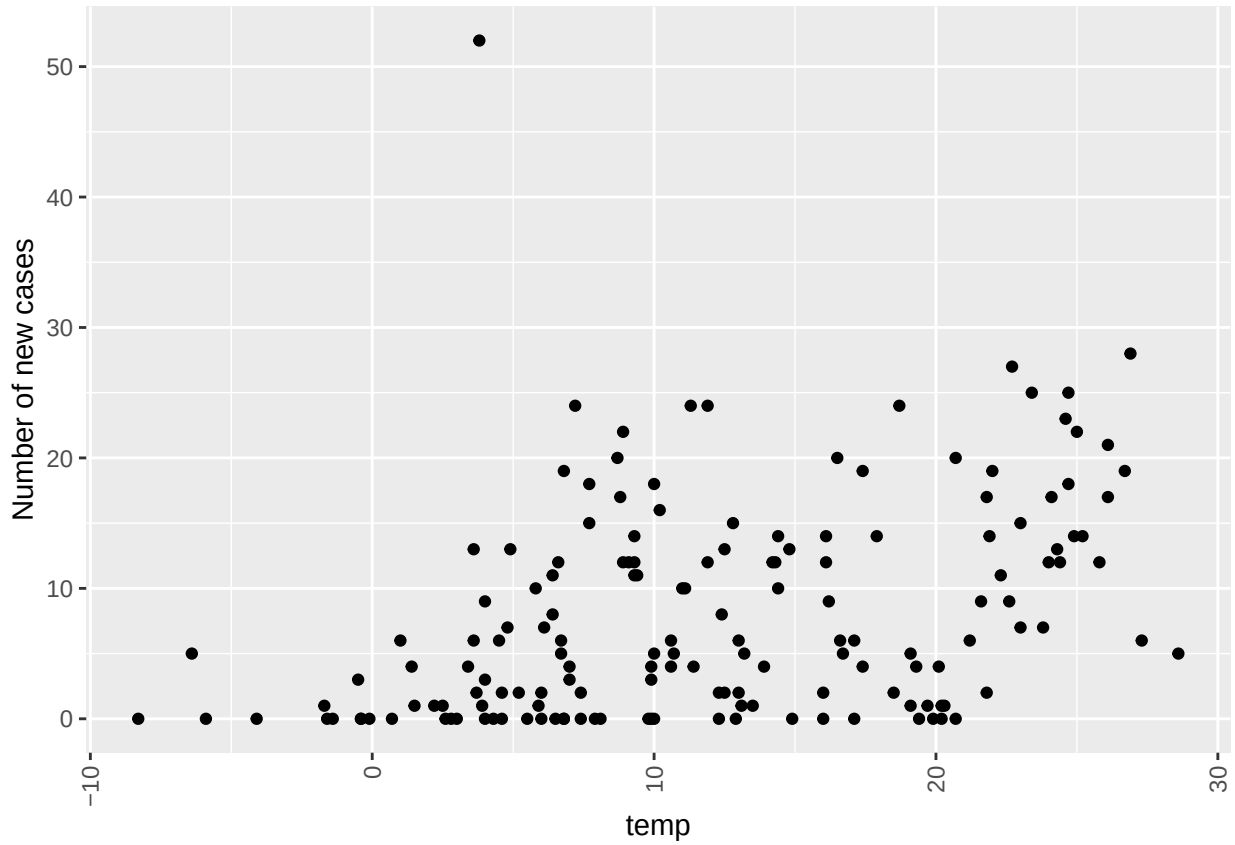
```
##  [1] "code"                 "province"             "date"
##  [4] "avg_temp"             "min_temp"             "max_temp"
##  [7] "precipitation"        "max_wind_speed"       "most_wind_direction"
## [10] "avg_relative_humidity"
```

```
# Deleting the columns not required for this analysis from both tables and creating new tables
Seoul_time<-tprovince_Seoul[,c("province","time"):=NULL]
Seoul_weather<-weather_Seoul[,c("province", "code"):=NULL]
```

```
# Merging the two tables (weather and time)
Seoul_weather_time <- merge(Seoul_weather, Seoul_time, by = "date", all = FALSE)
Seoul_weather_time[,.(avg_temp,daily_new_case)]
```

```
##      avg_temp daily_new_case
##   1:      0.7              0
##   2:     -0.4              0
##   3:      3.0              0
##   4:      4.6              0
##   5:      2.8              0
##  ---
## 158:     21.6              9
## 159:     21.8             17
## 160:     24.1             17
## 161:     25.2             14
## 162:     23.8              7
```

```
#Plotting daily new cases against temp
ggplot(Seoul_weather_time, aes(x=avg_temp, y=daily_new_case)) + geom_point() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="temp", y="Number of new cases")
```

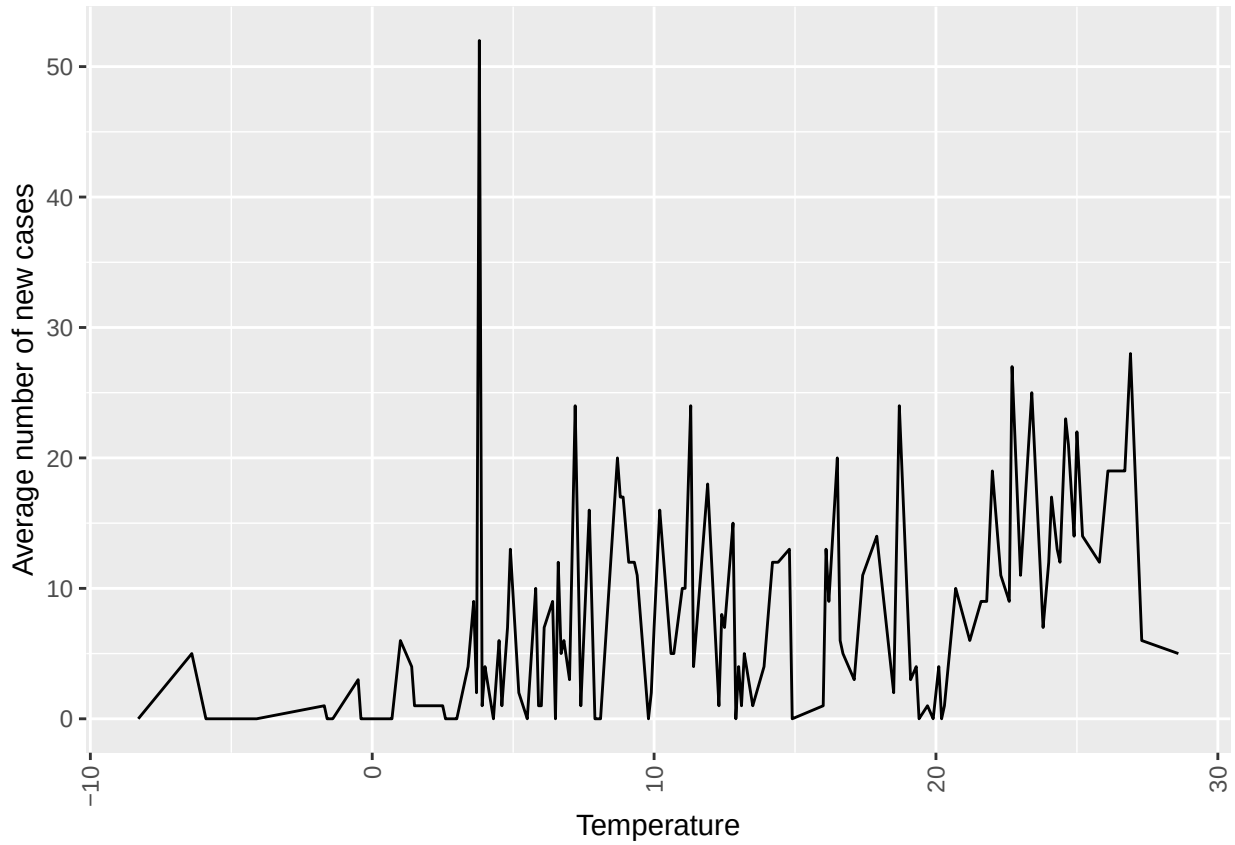


As temperature values repeat, we take average number of cases corresponding to a particular temperature.

```
Seoul_weather_time[, avg_no_of_new_cases:=as.integer(mean(daily_new_case)), by=avg_temp]
Seoul_weather_time[,.(avg_temp,avg_no_of_new_cases)][order(-avg_no_of_new_cases)]
```

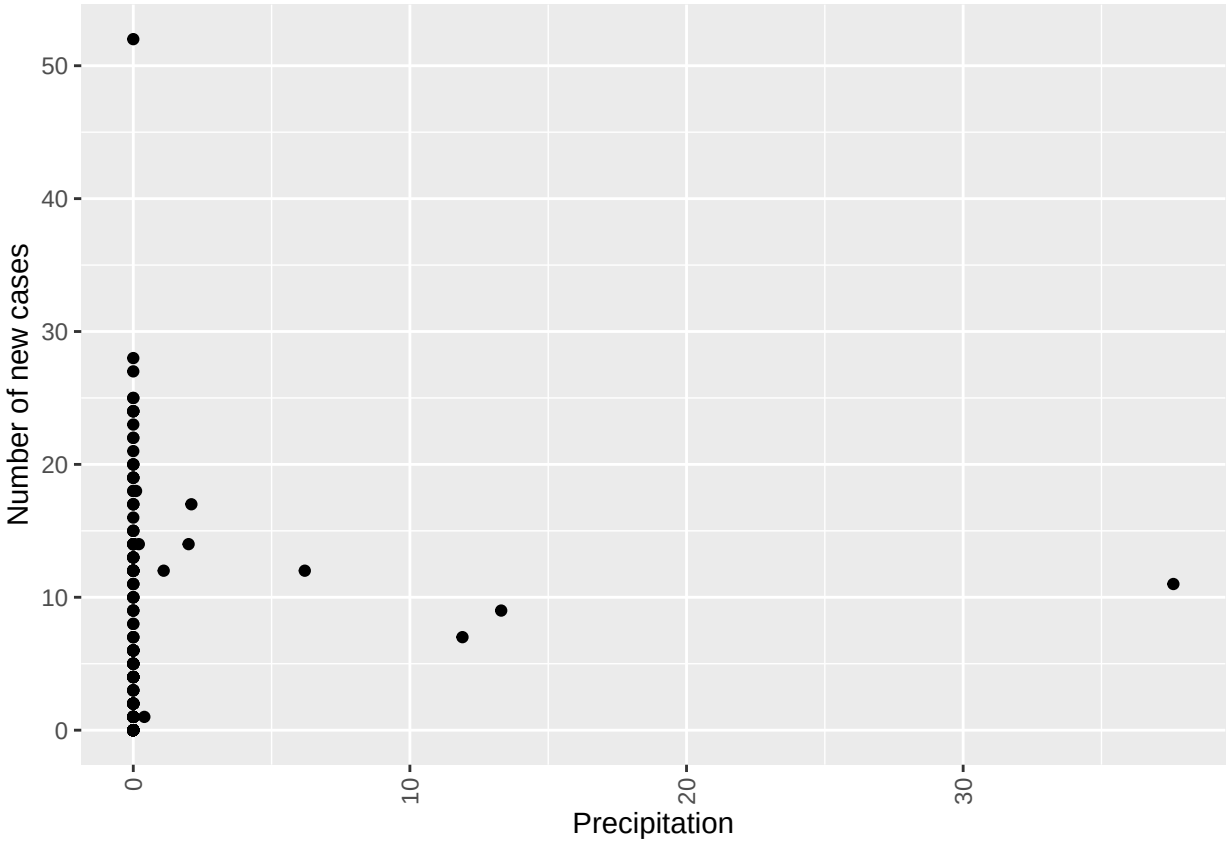
```
##      avg_temp avg_no_of_new_cases
##   1:      3.8                  52
##   2:     26.9                  28
##   3:     22.7                  27
##   4:     23.4                  25
##   5:     11.3                  24
##  ---
## 158:     14.9                   0
## 159:     20.2                   0
## 160:     20.2                   0
## 161:     19.4                   0
## 162:     19.9                   0
```

```
#Plotting average number of new cases against temp
ggplot(Seoul_weather_time, aes(x=avg_temp, y=avg_no_of_new_cases)) + geom_line() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Temperature", y="Average number of new cases")
```



From the above graph we infer that the average number of cases is generally higher for a temperature above 3.8 and comaparatively lower for lower temperature ranges.

```
#Plotting daily new cases against precipitation
ggplot(Seoul_weather_time, aes(x=precipitation, y=daily_new_case)) + geom_point() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Precipitation", y="Number of new cases")
```

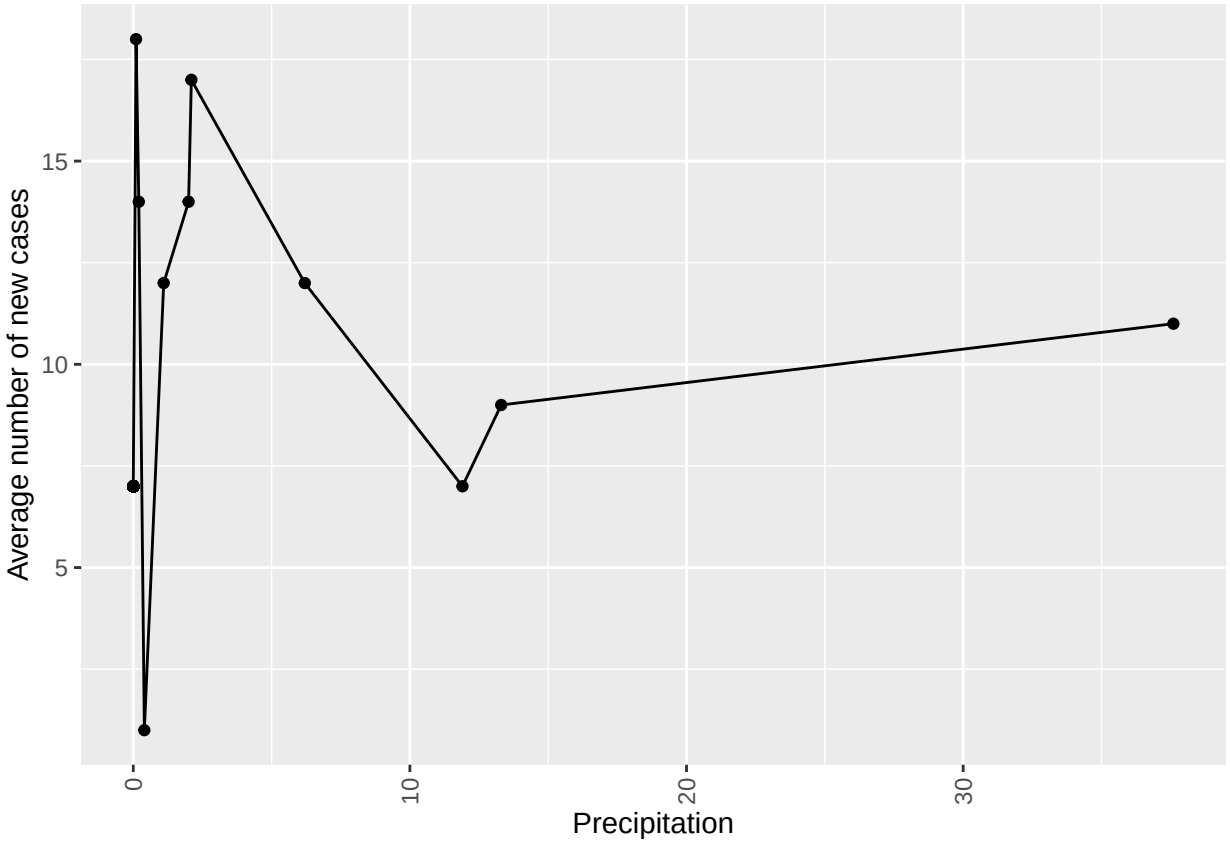


Here also, we need to look at average number of cases for a particular precipitation value, as precipitation value is same for multiple days.

```
Seoul_weather_time[, avg_no_of_new_cases_prep:=as.integer(mean(daily_new_case), na.rm=TRUE), by=precipi
Seoul_weather_time[, .(precipitation, avg_no_of_new_cases_prep)]
```

```
##      precipitation avg_no_of_new_cases_prep
##   1:           0.0                        7
##   2:           0.0                        7
##   3:           0.0                        7
##   4:           0.0                        7
##   5:           0.0                        7
##  ---
## 158:          13.3                        9
## 159:           2.1                       17
## 160:           0.0                        7
## 161:           0.0                        7
## 162:          11.9                        7
```

```
#Plotting average number of new cases against precipitation
ggplot(Seoul_weather_time, aes(x=precipitation, y=avg_no_of_new_cases_prep)) + geom_line() + geom_point
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) +
labs(x="Precipitation", y="Average number of new cases")
```



Precipitation changes seem to have no effect on the number of COVID cases.