

# 과제 3

32180331\_김가영

2020 12 6

```
## install.packages("githubinstall")  
library(ggmap)
```

```
## Loading required package: ggplot2
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(ggplot2)  
library(jsonlite)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
## filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

```
library(httr)  
library(leaflet)  
register_google(key='AIzaSyBoDPocehhHWmx-aXmvTbdylY7Soy_sJ_Y')
```

## 1. 지역별 코로나 확진자 현황 (지도를 이용해 서 표현)

```

API_KEY <- '72b2bb63260c981620256ec2b5677d6d'
data <- read.csv(file = "data.csv", header = TRUE)
names <- c("서울", "부산", "대구", "인천", "광주", "대전", "울산", "세종", "경기", "강원", "충북", "충
남", "전북", "전남", "경북", "경남"
          , "제주")
addr <- c("서울", "부산", "대구", "인천", "광주", "대전", "울산", "세종", "경기", "강원", "충북", "충
남", "전북", "전남", "경북", "경남"
          , "제주")
gc.list <- c()
for(i in 1:17){
  res <- GET(url = 'https://dapi.kakao.com/v2/local/search/address.json',
             query = list(query = addr[i]),
             add_headers(Authorization = paste0("KakaoAK ", API_KEY)))

  df <- res %>%
    content(as = 'text') %>%
    fromJSON()

  gc <- df$documents[1,] %>%
    select(x,y) %>%
    as.numeric()

  gc.list <- rbind(gc.list, gc)
}

gc.list

```

```

##      [,1]      [,2]
## gc 126.9787 37.56683
## gc 129.0751 35.17982
## gc 128.6018 35.87138
## gc 126.7053 37.45600
## gc 127.2551 37.42939
## gc 127.3848 36.35054
## gc 129.3115 35.53948
## gc 127.2892 36.48006
## gc 127.0089 37.27498
## gc 127.7298 37.88533
## gc 127.4915 36.63536
## gc 126.6728 36.65883
## gc 127.1090 35.82020
## gc 126.4628 34.81608
## gc 128.5058 36.57600
## gc 128.6919 35.23777
## gc 126.4982 33.48891

```

```

df <- data.frame(name = names,
                 lon = gc.list[,1],
                 lat = gc.list[,2])
str(df)

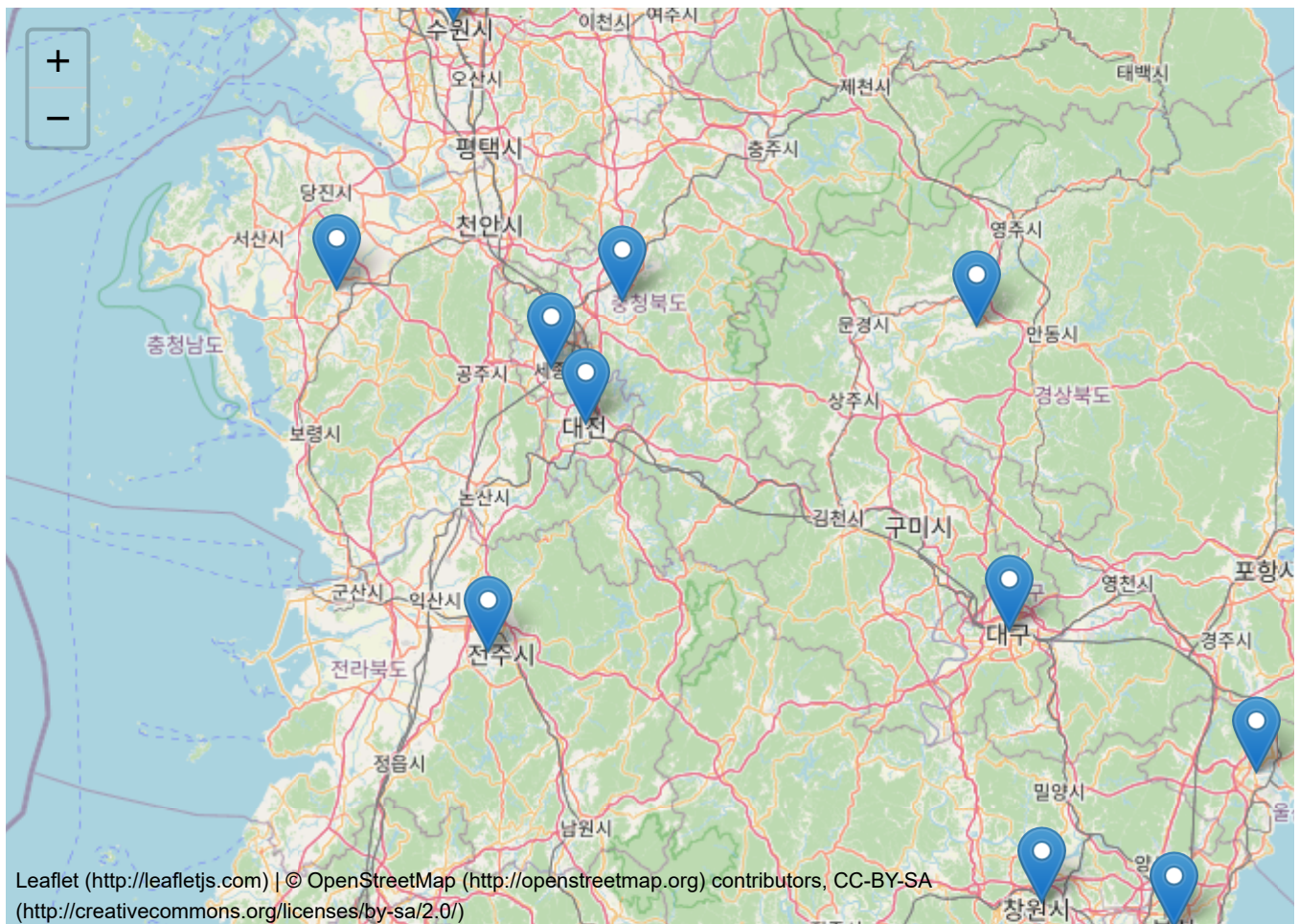
```

```
## 'data.frame':   17 obs. of  3 variables:
## $ name: chr  "서울" "부산" "대구" "인천" ...
## $ lon : num  127 129 129 127 127 ...
## $ lat : num  37.6 35.2 35.9 37.5 37.4 ...
```

```
cen <- c(mean(df$lon), mean(df$lat))

map <- leaflet(data = df) %>%
  addTiles() %>%
  addMarkers(~lon, ~lat, label = ~name) %>%
  setView(lng = cen[1], lat = cen[2], zoom = 3)
```

map

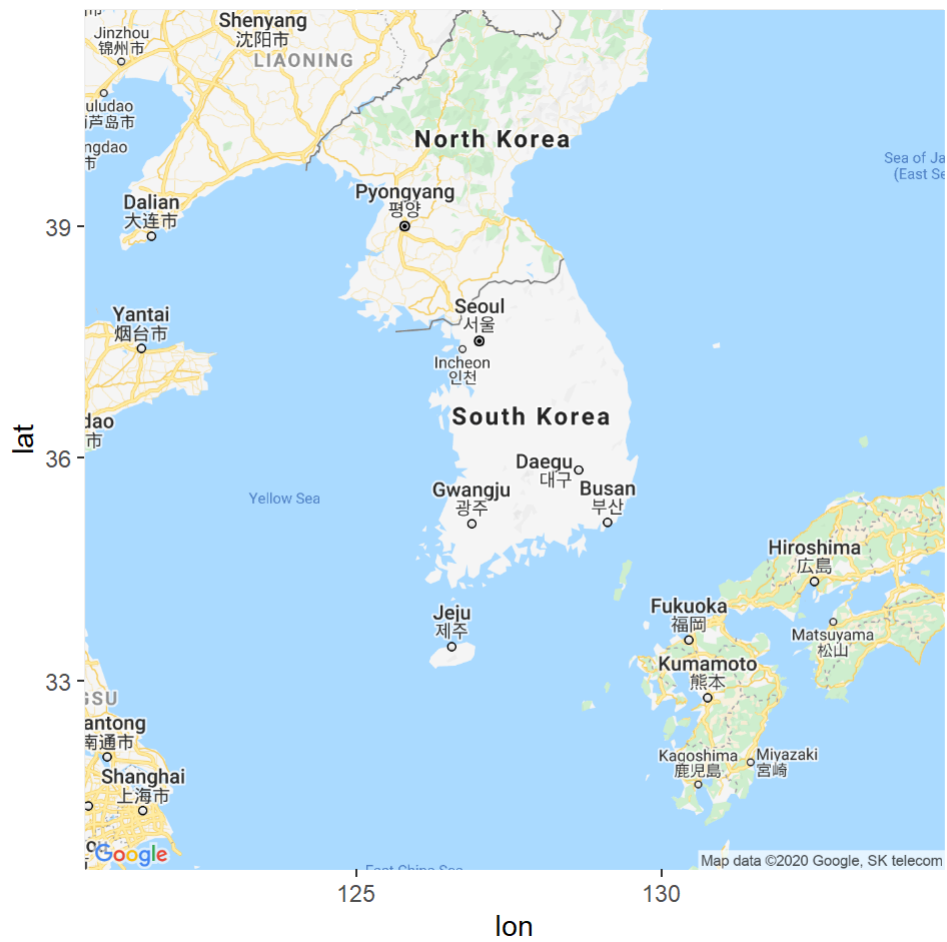


```
library(ggmap)
data <- read.csv(file = "data.csv", header = TRUE)

map <- get_googlemap(center=cen, maptype = "roadmap", zoom=6)
```

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=36.250998,127.574831&zoom=6&size=640x640&scale=2&maptype=roadmap&key=xxx-aXmvTbdyIY7Soy_sJ_Y
```

```
gmap <- ggmap(map)
gmap
```



```
nrow(data)
```

```
## [1] 234
```

```
ncol(data)
```

```
## [1] 304
```

```
s <- c(data[58,304],data[68,304],data[78,304],data[88,304],data[98,304],data[108,304],data[118,304],data[128,304],data[138,304],
        ,data[148,304],data[158,304],data[168,304],data[178,304],data[188,304],data[198,304],
        data[208,304],data[218,304])
name <- c(addr)
data.frame(name,s)
```

```
##      name      s
## 1   서울  7900
## 2   부산   661
## 3   대구  7223
## 4   인천  1286
## 5   광주   632
## 6   대전   464
## 7   울산   178
## 8   세종    89
## 9   경기  6681
## 10  강원   575
## 11  충북   231
## 12  충남   817
## 13  전북   252
## 14  전남   373
## 15  경북  1679
## 16  경남   492
## 17  제주    70
```

```
## df = data.frame(lon,lat,s)
## df
## gmap+geom_point(data=df,aes(x=lon,y=lat,size=s),alpha=0.5,col="blue") +
##   scale_size_continuous(range=c(1,18))
```

서울이 확진자가 제일 많고, 대구, 경기 순으로 많다. 아무래도 수도권에 사람들이 제일 밀집해있기 때문에 확진자가 많다.

## 2. 지역별 코로나 사망자 현황(지도를 이용해 표현)

```
s <- c(data[64,304],data[74,304],data[84,304],data[94,304],data[104,304],data[114,304],data[124,304],data[134,304],data[144,304],
        ,data[154,304],data[164,304],data[174,304],data[184,304],data[194,304],data[204,304],
        data[214,304],data[224,304])
name <- c(addr)
data.frame(name,s)
```

```
##      name      s
## 1  서울      87
## 2  부산      15
## 3  대구     199
## 4  인천      11
## 5  광주       3
## 6  대전       6
## 7  울산       2
## 8  세종       0
## 9  경기     111
## 10 강원       6
## 11 충북       3
## 12 충남       8
## 13 전북       2
## 14 전남       2
## 15 경북      57
## 16 경남       1
## 17 제주       0
```

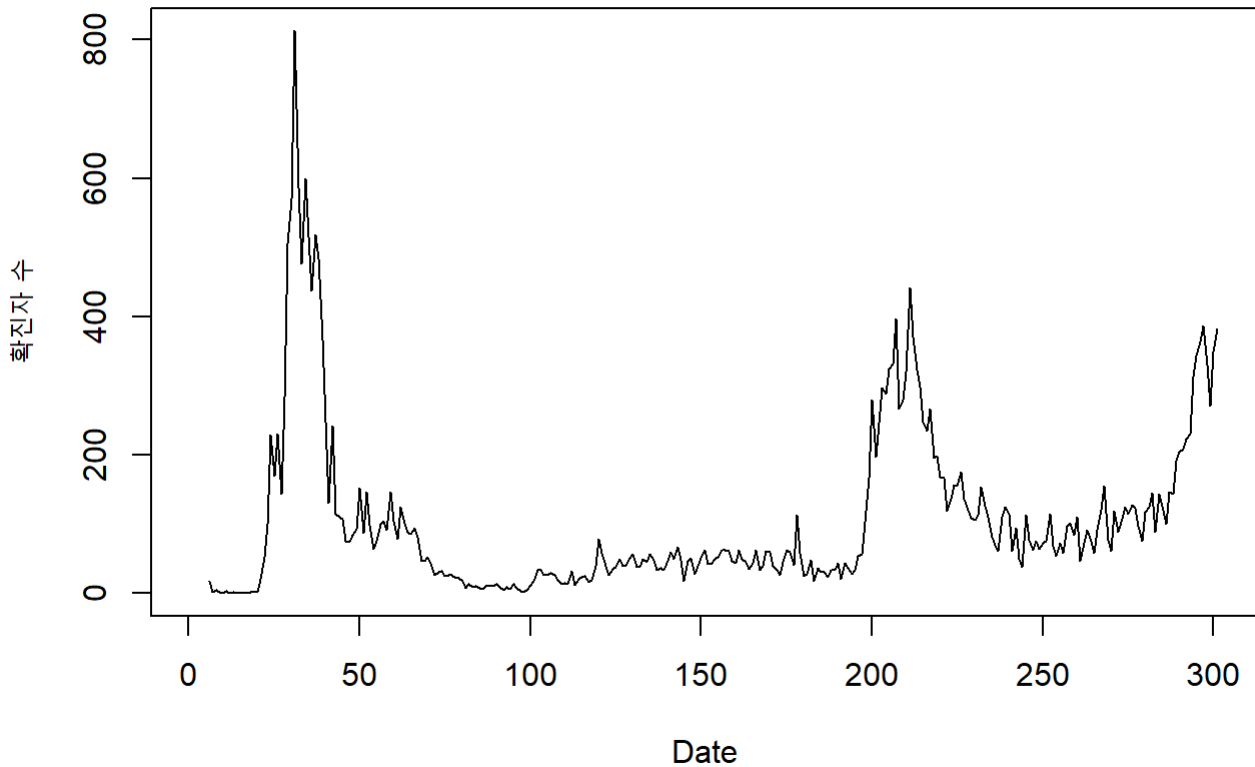
1번에서 확진자 수는 서울이 제일 많았지만, 사망자 수는 대구, 경기순으로 많고, 그 다음이 서울이다. 확진자 수와 사망자 수는 비례하지 않는다.

### 3. 일별 전국 확진자 현황 (선 그래프 이용)

```
date = 1:301
number = c(data[3,4:304])
plot(date,
      number,
      main = "일별 전국 확진자 현황",
      type = 'l',
      lty=1,
      lwd=1,
      xlab="Date",
      ylab="확진자 수")
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 강제형변환에 의해 생성된 NA 입
## 니다
```

## 일별 전국 확진자 현황

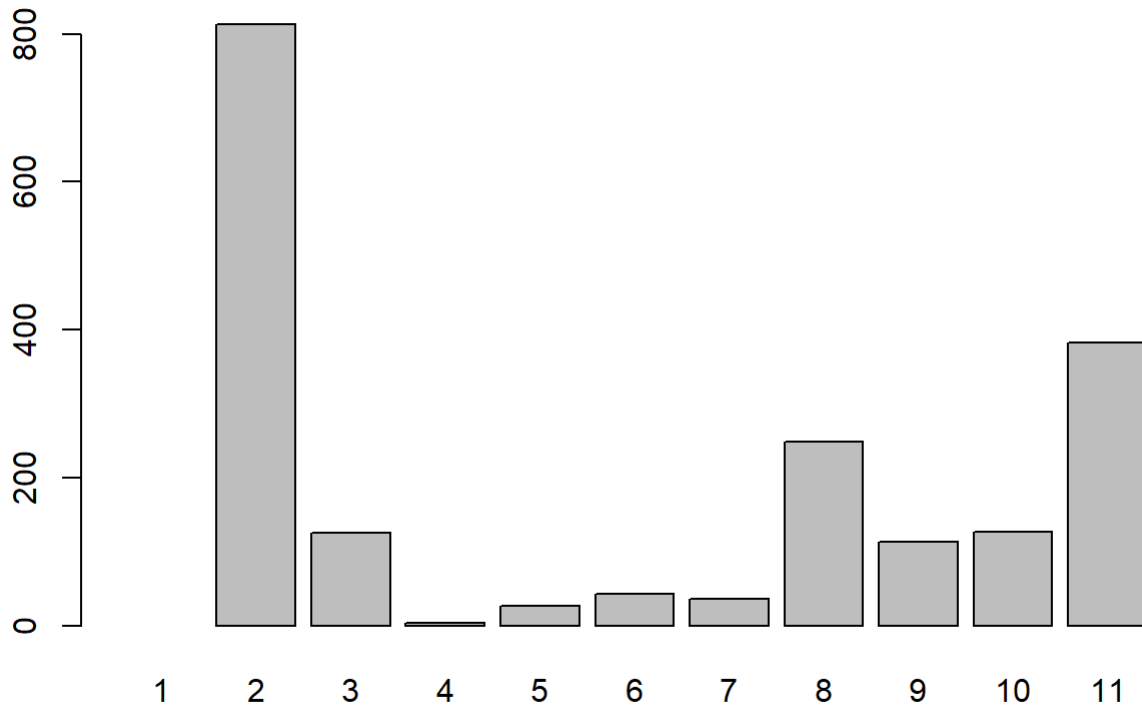


1월,2월에는 확진자가 별로 없다가 2월 후반부에 약 800명으로 급격히 늘었고, 확 줄었다가 9월쯤 다시 늘어나고 지금까지도 확진자가 발생하는 추세이다.

## 4. 월별 전국 확진자 현황

```
month = 1:11
number <- c(data[3,5], data[3,34], data[3,65], data[3,95], data[3,126], data[3,156], data[3,187], data[3,218], data[3,248],
            data[3,279], data[3,304])
barplot(number, names=c(month), main='월별 전국 확진자 현황')
```

월별 전국 확진자 현황



확진자는 2월을 기준으로 확 늘어났고, 3월부터는 줄어들었다가 8월에 휴가철의 영향으로 급격히 증가하였고, 11월에도 많이 증가하였다.

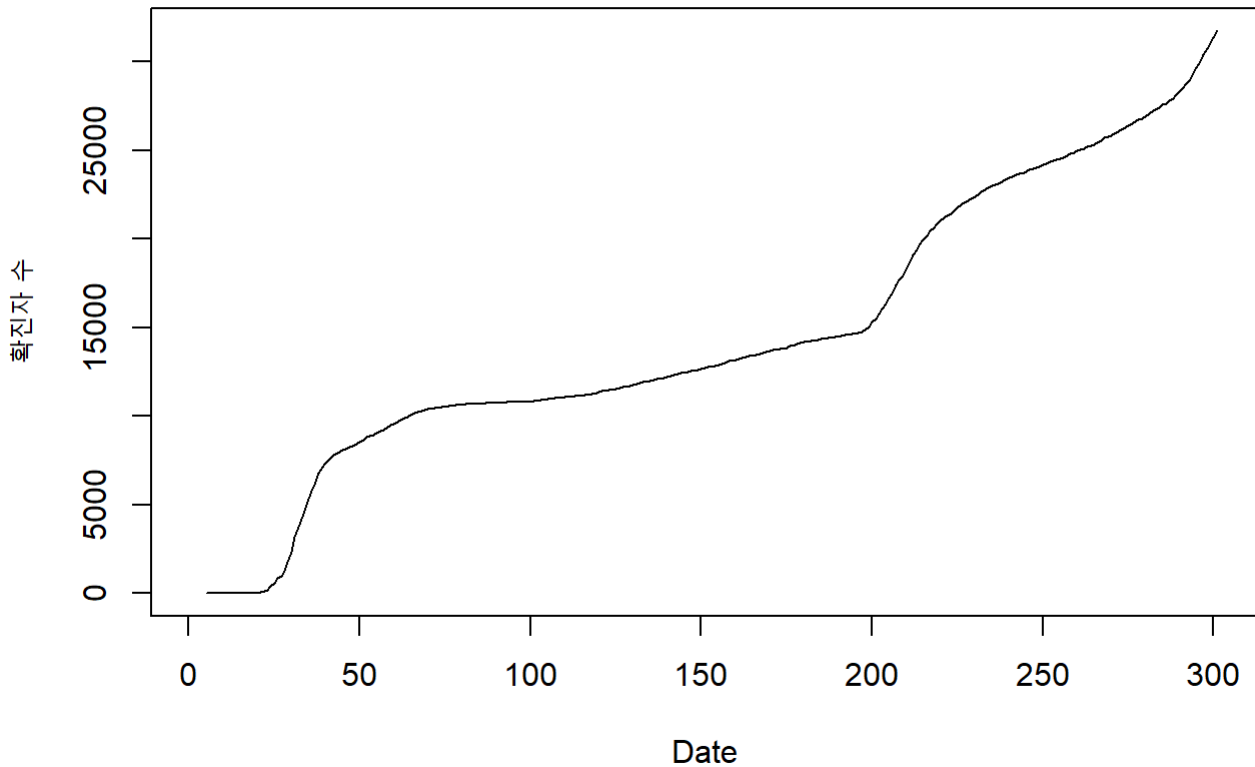
## 5. 누적 확진자 현황 (선 그래프 이용)

```
date = 1:301
number = c(data[4,4:304])
plot(date,
      number,
      main = "누적 확진자 현황 (2020.01.31~2020.11.25)",
      type = 'l',
      lty=1,
      lwd=1,
      xlab="Date",
      ylab="확진자 수")
```

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 강제형변환에 의해 생성된 NA 입
## 니다
```



## 누적 확진자 현황(2020.01.31~2020.11.25)



2월부터 확진자가 발생하였고, 3월부터 7월정도까지는 조금씩 누적되다가 8월부터 휴가철의 영향으로 다시 급격히 늘어나 누적 확진자가 25000명에 돌파하게 되었고, 11월 25일까지 계속 증가하는 추세이다.

## 6. 남녀 코로나 확진자 수 비교(막대 그래프)

```
sex <- c('male','female')
people_number <- c(data[12,304], data[16,304])
barplot(people_number, names=c(sex), main='남녀 전국 확진자 현황')
```

## 남녀 전국 확진자 현황



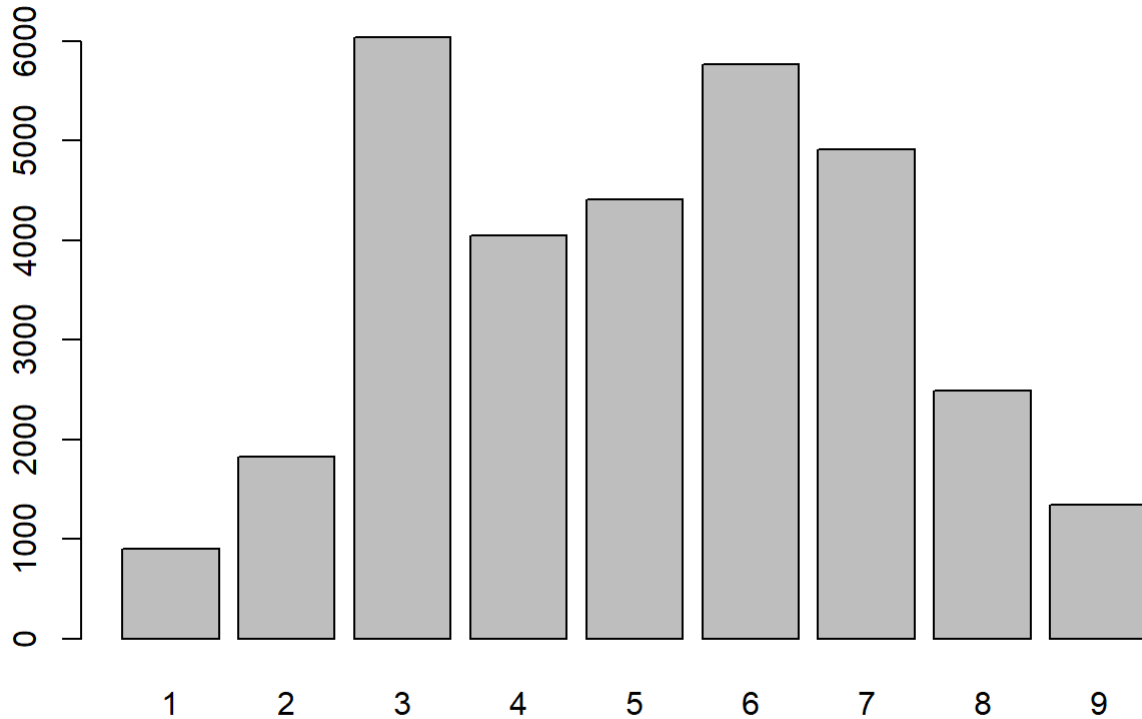
11월 25일 기준으로 남녀 전국 누적 확진자 수를 비교해보았을 때, 큰 차이가 없는 것으로 보아 코로나는 남,녀 구별없이 걸리는 병이라는 것을 알 수 있다.

## 7. 나이별 코로나 확진자 수 비교

```
group1 <- data[20,304]
group2 <- data[24,304]
group3 <- data[28,304]
group4 <- data[32,304]
group5 <- data[36,304]
group6 <- data[40,304]
group7 <- data[44,304]
group8 <- data[48,304]
group9 <- data[52,304]

age<-c('1','2','3','4','5','6','7','8','9')
age_number <- c(group1,group2,group3,group4,group5,group6,group7,group8,group9)
barplot(age_number,names =c(age), main ='나이별 코로나 확진자 수')
```

나이별 코로나 확진자 수



1 : 0~9세, 2 : 10~19세, 3 : 20~29세, 4 : 30~39세, 5 : 40~49세, 6 : 50~59세, 7 : 60~69세, 8 : 70~79세, 9 : 80세 이상이라고 했을 때 3구간과 6구간, 그 다음으로는 7구간이 제일 확진자 수가 많은 것을 볼 수 있다. 노인들이 제일 취약하다고 알려져있지만 실제로는 20대가 가장 많이 나오고 있다는 것을 알 수 있다.