R입문 기말고사

김민준(202121836)

12/20/2021

1번

```
### 1-1
epsilon_i<-rnorm(1000)
head(epsilon_i)
```

```
## [1] -0.3961434 -0.9573362 -0.6673329  0.4419802  0.7406195  0.6378518
```

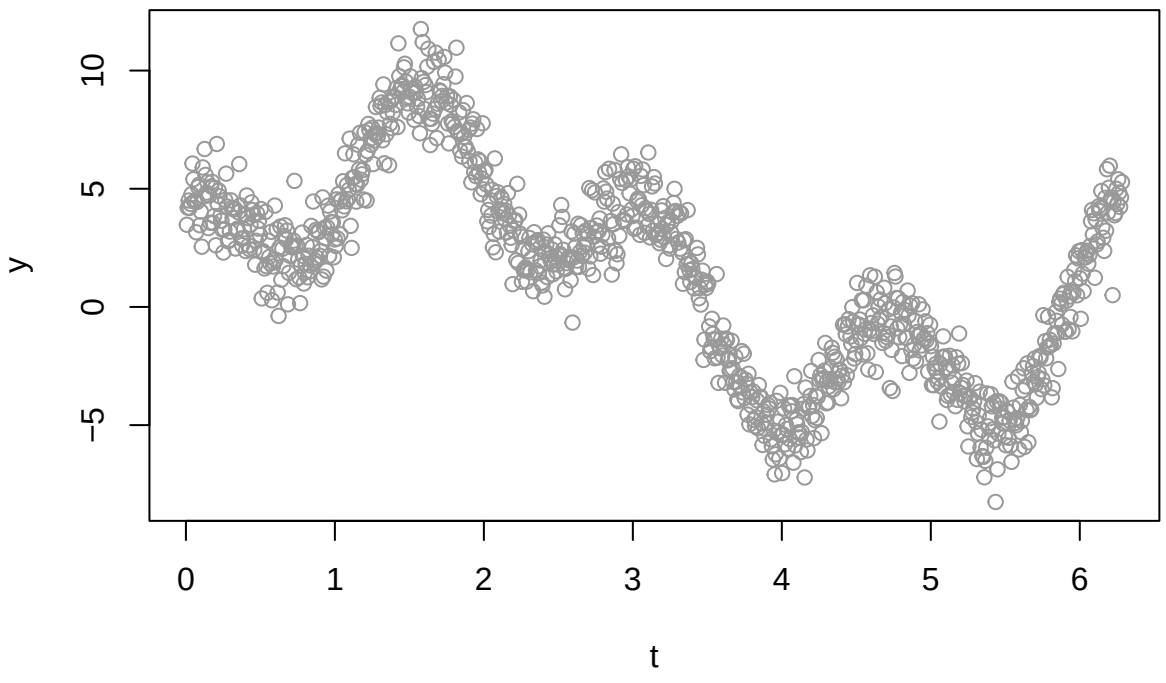
```
### 1-2
i=1:1000
X1<- sin(2*pi*i/1000)
j=1:1000
X2<- cos(8*pi*j/1000)
head(X1)
```

```
## [1] 0.006283144 0.012566040 0.018848440 0.025130095 0.031410759 0.037690183
```

```
head(X2)
```

```
## [1] 0.9996842 0.9987370 0.9971589 0.9949510 0.9921147 0.9886517
```

```
###1-3
epsilon_i<-rnorm(1000)
k=1:1000
t<- 2*pi*k/1000
l=1:1000
y<- 1.5 + 5*sin(2*pi*l/1000) + 3*cos(8*pi*l/1000) + epsilon_i
plot(t,y,col='gray60')
```



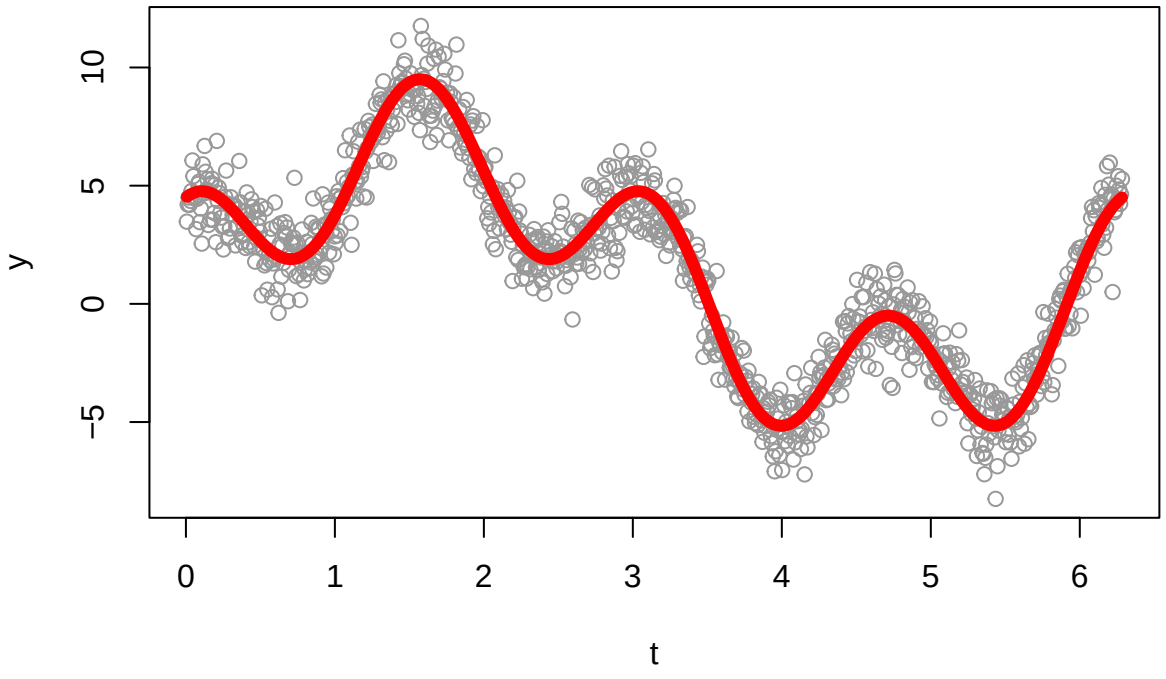
```
###1-4
X<- cbind(1,X1,X2)
head(X)
```

```
##                  X1        X2
## [1,] 1 0.006283144 0.9996842
## [2,] 1 0.012566040 0.9987370
## [3,] 1 0.018848440 0.9971589
## [4,] 1 0.025130095 0.9949510
## [5,] 1 0.031410759 0.9921147
## [6,] 1 0.037690183 0.9886517
```

```
###1-5
β <-rbind(1.5,5,3)
β
```

```
##      [,1]
## [1,]  1.5
## [2,]  5.0
## [3,]  3.0
```

```
m <- X %*% β
plot(t,y,col='gray60')
lines(t, m, col='red', lwd='6')
```




```
###1-6
X<- cbind(1,X1,X2)
Y<- cbind(y)
```

```
a<- t(X)
b<- solve(a %*% X)
hatβ<- b %*% a %*% Y
hatβ
```

```
##           y
##    1.464048
## X1 4.907243
## X2 2.968219
```

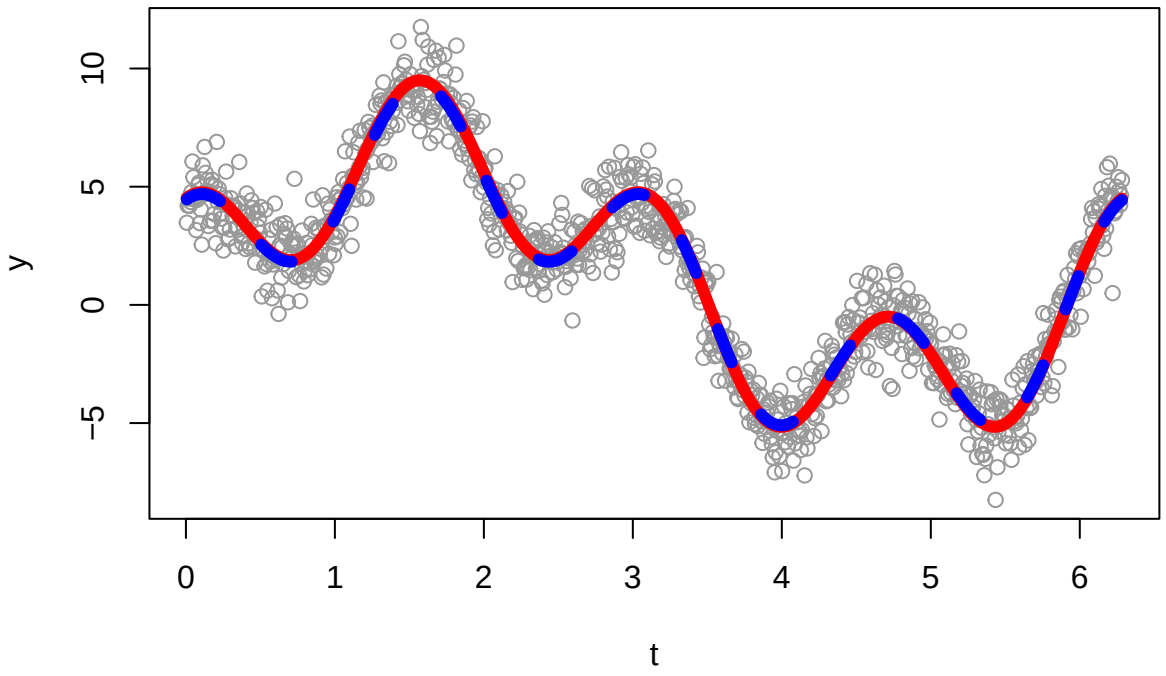
```
c(hatβ)
```

```
## [1] 1.464048 4.907243 2.968219
```

```
c(β)
```

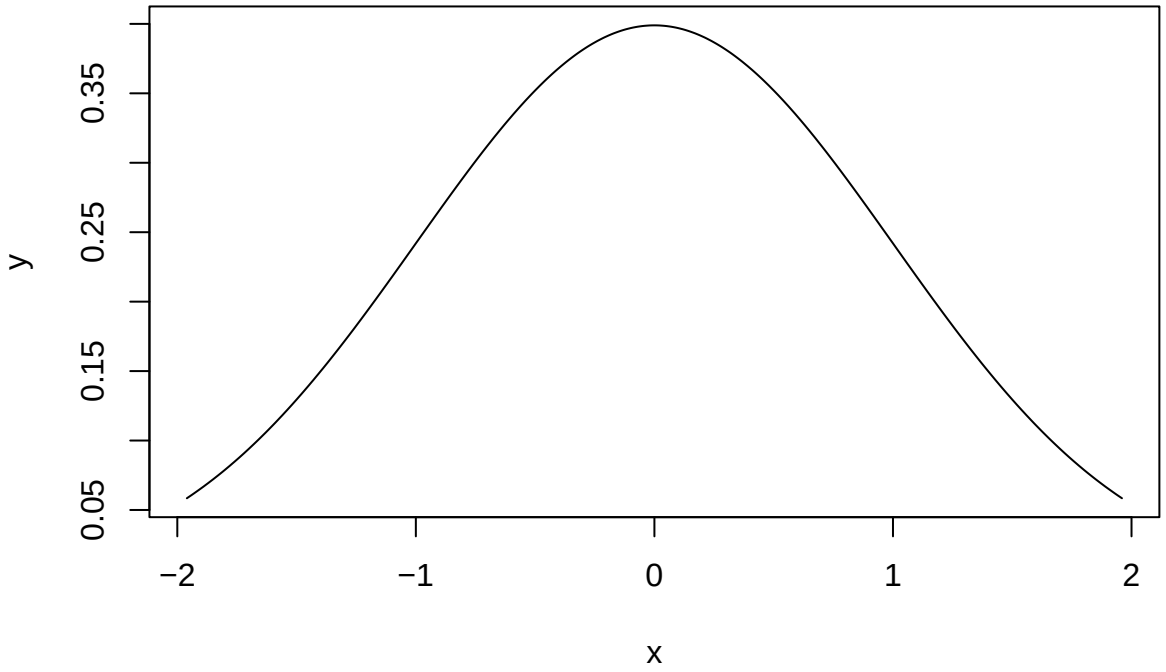
```
## [1] 1.5 5.0 3.0
```

```
###1-7
Xhatβ <- X %*% hatβ
plot(t,y,col='gray60')
lines(t,m,col='red',lwd='6')
lines(t,Xhatβ, lty="dashed", col='blue', lwd='6')
```

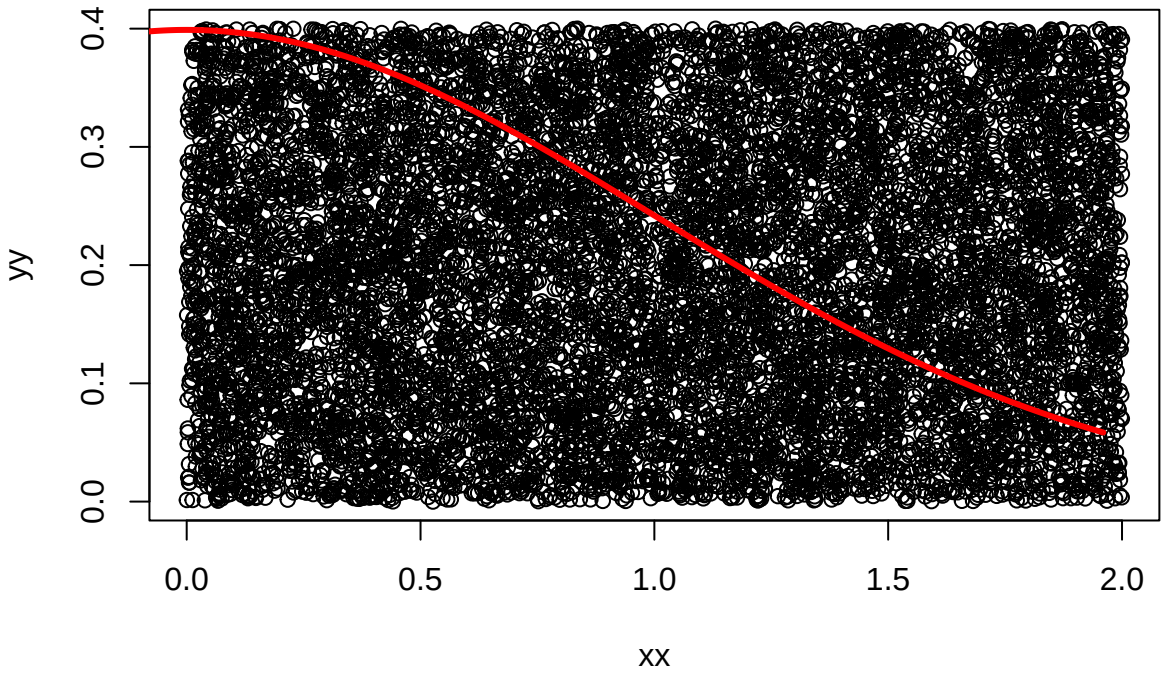


2번

```
### 2-1
x=seq(from=-1.96, to=1.96, by=0.01)
y=(1/sqrt(2*pi))*exp(-x^2/2)
plot(x,y,type='l')
```




```
xx=runif(10000)
xx=xx*2
yy=runif(10000)*0.4
plot(xx,yy)
lines(x,y,col='red',lwd=3)
```




```
test=function(xx,yy){
  yy< (1/sqrt(2*pi))*exp(-xx^2/2)
}
print(c(xx[1],yy[1]))
```

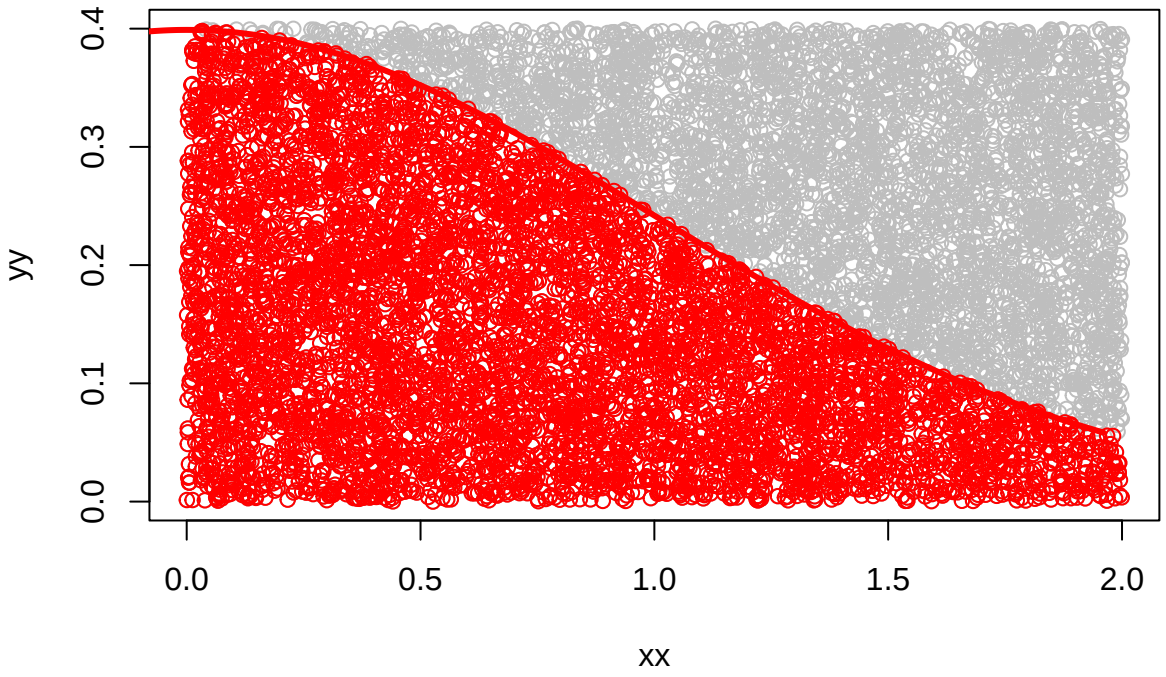
```
## [1] 0.9327435 0.1800020
```

```
print(1/sqrt(2*pi))*exp((-1/2)*xx[1]^2)
```

```
## [1] 0.3989423
```

```
## [1] 0.2582199
```

```
tst=c()
for(i in 1:10000) tst[i]= test(xx[i],yy[i])
plot(xx,yy,col='gray')
lines(x,y,col='red',lwd=3)
points(xx[tst],yy[tst],col='red')
```




```
sum(tst)
```

```
## [1] 5997
```

```
sum(tst)/10000
```

```
## [1] 0.5997
```

```
sum(tst)/10000*0.8*2
```

```
## [1] 0.95952
```

```
###2-2
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
df<-c(rnorm(1000))
head(df)
```

```
## [1] -0.1351214 -0.3360343  0.4503999 -1.3411654  0.6760814 -0.4031555
```

```
tb=as_tibble(df)
head(tb)
```

```
## # A tibble: 6 x 1
##    value
##    <dbl>
## 1 -0.135
## 2 -0.336
## 3  0.450
## 4 -1.34
## 5  0.676
## 6 -0.403
```

```
tb2= tb %>% filter(df>-1.96,df<1.96)
head(tb2)
```

```
## # A tibble: 6 x 1
##    value
##    <dbl>
## 1 -0.135
## 2 -0.336
## 3  0.450
## 4 -1.34
## 5  0.676
## 6 -0.403
```

```
nrow(tb2)
```

```
## [1] 948
```

3번

```
### TypeA
APR=c('N1','N2','N3','N4','N5','N6','N7','N8','A','N10')
SURV=10
PLAYER=APR[SURV]
STAGE=0
PROB=0.5
TOSSRSLT=NA
library(tidyverse)
toss=function(p) rbinom(n=1,size=1,prob=p) %>% as.logical

reset=function(){
  TOSSRSLT<<-NA
  SURV<<- 10
  STAGE<<- 0
  PLAYER<<- APR[SURV]
}

record=function(){
  list(PRE_TOSSRSLT=TOSSRSLT, SURV=SURV, STAGE=STAGE, PLAYER=PLAYER)
}

go= function(){
  PROB<<- 0.5+(PLAYER=='A')*0.45
  TOSSRSLT<<-toss(PROB)
  if (TOSSRSLT==FALSE) SURV<<-SURV-1
  STAGE<<-STAGE+1
  PLAYER<<-APR[SURV]
}

gogo=function(){
      for(i in 1:20){
        go()
       if (SURV==0) break
  }
}

gogo_history=function() {
  rslt_=as_tibble(record())
  for(i in 1:20){
    go()
    rslt_=rbind(rslt_,as_tibble(record()))
  }
```

```
  print(rslt_)
}

simulate_once=function(){
  reset()
  gogo()
  return(record()$SURV)
}

simrslt=c()
for (i in 1:1000) simrslt[i]=simulate_once()
mean(simrslt)
```

```
## [1] 5.584
```

```
APR=c('N10','A','N8','N7','N6','N5','N4','N3','N2','N1')
SURV=10
PLAYER=APR[SURV]
STAGE=0
PROB=0.5
TOSSRSLT=NA
library(tidyverse)

toss=function(p) rbinom(n=1,size=1,prob=p) %>% as.logical

reset=function(){
  TOSSRSLT<<-NA
  SURV<<- 10
  STAGE<<- 0
  PLAYER<<- APR[SURV]
}

record=function(){
  list(PRE_TOSSRSLT=TOSSRSLT, SURV=SURV, STAGE=STAGE, PLAYER=PLAYER)
}

go= function(){
  PROB<<-0.5+(PLAYER=='A')*0.45
  TOSSRSLT<<-toss(PROB)
  if (TOSSRSLT==FALSE) SURV<<-SURV-1
  STAGE<<-STAGE+1
  PLAYER<<-APR[SURV]
}

gogo=function(){
```

```
  for(i in 1:20){
    go()
    if (SURV==0) break
  }
}

gogo_history=function() {
  rslt_=as_tibble(record())
  for(i in 1:20){
    go()
    rslt_=rbind(rslt_,as_tibble(record()))
  }
  print(rslt_)
}

simulate_once=function(){
  reset()
  gogo()
  return(record()$SURV)
}

simrslt=c()
for (i in 1:1000) simrslt[i]=simulate_once()
mean(simrslt)
```

```
## [1] 1.842
```

```
## 따라서 8번 참가자는 TypeA에서 살아남을 확률이 높다.
```

4번

```
df=read_csv('https://raw.githubusercontent.com/guebin/2021IR/master/_notebooks/covid19.c
```

```
## Rows: 12294 Columns: 5
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): prov
## dbl (4): year, month, day, cases
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
head(df)
```

```
## # A tibble: 6 x 5
##    year month   day prov  cases
##   <dbl> <dbl> <dbl> <chr> <dbl>
## 1  2020     1    20 서울      0
## 2  2020     1    20 부산      0
## 3  2020     1    20 대구      0
## 4  2020     1    20 인천      1
## 5  2020     1    20 광주      0
## 6  2020     1    20 대전      0
```

```
library(tidyverse)
```

```
###4-1
df %>% filter(year==2020) %>% summarise(sum(cases))
```

```
## # A tibble: 1 x 1
##   `sum(cases)`
##          <dbl>
## 1        60726
```

```
df %>% filter(year==2021) %>% summarise(sum(cases))
```

```
## # A tibble: 1 x 1
##   `sum(cases)`
##          <dbl>
## 1       396886
```

```
###4-2
df %>% filter(year==2020 & month==2 & day<=15) %>% group_by(prov) %>% summarise(confirm=
```

```
## # A tibble: 18 x 2
##    prov  confirm
```

```
##    <chr>   <dbl>
##  1 강원        0
##  2 검역        0
##  3 경기        9
##  4 경남        0
##  5 경북        0
##  6 광주        2
##  7 대구        0
##  8 대전        0
##  9 부산        0
## 10 서울        5
## 11 세종        0
## 12 울산        0
## 13 인천        0
## 14 전남        1
## 15 전북        0
## 16 제주        0
## 17 충남        0
## 18 충북        0
```

```
## 따라서 가장 많은 확진자가 발견된 지역은 경기도다.
```

```
###4-3
df %>% filter(year==2020 &month==2 & day>=16 & day<=29) %>% group_by(prov) %>% summarise
```

```
## # A tibble: 18 x 2
##    prov  confirm
##    <chr>   <dbl>
##  1 강원        7
##  2 검역        0
##  3 경기       65
##  4 경남       59
##  5 경북      472
##  6 광주        7
##  7 대구     2055
##  8 대전       13
##  9 부산       75
## 10 서울       62
## 11 세종        1
## 12 울산       17
## 13 인천        5
## 14 전남        1
## 15 전북        4
## 16 제주        2
## 17 충남       48
## 18 충북       10
```

```
## 따라서 가장 많은 확진자가 발견된 지역은 대구다.
```