First Project

20180968 박서현

2020 10 11

데이터 불러오기

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------- tidyverse 1.3.0 --
```

```
## √ ggplot2 3.3.2     √ purrr   0.3.4
## √ tibble  3.0.3     √ dplyr   1.0.2
## √ tidyr   1.1.2     √ stringr 1.4.0
## √ readr   1.4.0     √ forcats 0.5.0
```

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

```
DF <- read_csv("C:/Users/nabib/Documents/GitHub/Statics/1011/toefl.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   id = col_double(),
##   gender = col_character(),
##   listen = col_double(),
##   grammar = col_double(),
##   reading = col_double()
## )
```

```
head(DF)
```

```
## # A tibble: 6 x 5
##      id gender listen grammar reading
##   <dbl> <chr>   <dbl>   <dbl>   <dbl>
## 1     1 FEMALE     40      50      56
## 2     2 FEMALE     33      44      46
## 3     3 MALE       39      44      47
## 4     4 FEMALE     44      55      66
## 5     5 MALE       43      56      53
## 6     6 MALE       57      49      53
```

성(Female,Male}를 새로운 변수 성별{0=남,1=여}로 변환하고 값을 지정하시오.

변수유형 변경(factor)

```
DF <- DF %>% mutate(gender=factor(gender))
str(DF)
```

```
## tibble [100 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id     : num [1:100] 1 2 3 4 5 6 7 8 9 10 ...
##  $ gender : Factor w/ 2 levels "FEMALE","MALE": 1 1 2 1 2 2 1 1 1 1 ...
##  $ listen : num [1:100] 40 33 39 44 43 57 54 40 43 50 ...
##  $ grammar: num [1:100] 50 44 44 55 56 49 43 50 51 48 ...
##  $ reading: num [1:100] 56 46 47 66 53 53 56 53 57 52 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_double(),
##   ..   gender = col_character(),
##   ..   listen = col_double(),
##   ..   grammar = col_double(),
##   ..   reading = col_double()
##   .. )
```

수준재조정

```
DF$gender <- factor(DF$gender, levels=c('MALE', 'FEMALE'))

DF$gender
```

```
##   [1] FEMALE FEMALE MALE   FEMALE MALE   MALE   FEMALE FEMALE FEMALE FEMALE
##  [11] MALE   MALE   MALE   MALE   FEMALE FEMALE FEMALE MALE   FEMALE FEMALE
##  [21] FEMALE MALE   FEMALE FEMALE MALE   MALE   FEMALE FEMALE MALE   FEMALE
##  [31] MALE   MALE   FEMALE FEMALE MALE   FEMALE MALE   FEMALE FEMALE FEMALE
##  [41] MALE   FEMALE FEMALE MALE   FEMALE MALE   FEMALE MALE   FEMALE MALE
##  [51] FEMALE FEMALE FEMALE FEMALE MALE   MALE   FEMALE FEMALE MALE   MALE
##  [61] FEMALE MALE   MALE   FEMALE FEMALE FEMALE MALE   FEMALE MALE   MALE
##  [71] MALE   FEMALE FEMALE MALE   FEMALE FEMALE FEMALE FEMALE FEMALE FEMALE
##  [81] MALE   FEMALE MALE   FEMALE MALE   FEMALE FEMALE FEMALE FEMALE FEMALE
##  [91] MALE   FEMALE MALE   FEMALE MALE   FEMALE MALE   MALE   FEMALE MALE
## Levels: MALE FEMALE
```

factor를 수치 변수로 변환

```
DF$gendernum <-as.numeric(DF$gender)

head(DF)
```

```
## # A tibble: 6 x 6
##      id gender listen grammar reading gendernum
##   <dbl> <fct>   <dbl>   <dbl>   <dbl>     <dbl>
## 1     1 FEMALE     40      50      56         2
## 2     2 FEMALE     33      44      46         2
## 3     3 MALE       39      44      47         1
## 4     4 FEMALE     44      55      66         2
## 5     5 MALE       43      56      53         1
## 6     6 MALE       57      49      53         1
```

값 조정

```
DF <- mutate(DF, gendernum = ifelse(gendernum==2, 1, 0))

head(DF)
```

```
## # A tibble: 6 x 6
##      id gender listen grammar reading gendernum
##   <dbl> <fct>   <dbl>   <dbl>   <dbl>     <dbl>
## 1     1 FEMALE     40      50      56         1
## 2     2 FEMALE     33      44      46         1
## 3     3 MALE       39      44      47         0
## 4     4 FEMALE     44      55      66         1
## 5     5 MALE       43      56      53         0
## 6     6 MALE       57      49      53         0
```

토플성적은 (듣기+문법+독해)/3*10를 소수점 첫째자리에서 반올

림한 것이다. 새 변수 성적을 생성하시오.

```
DF <- mutate(DF, grade = round((listen + grammar + reading) / 3 * 10), 1)

head(DF)
```

```
## # A tibble: 6 x 8
##      id gender listen grammar reading gendernum grade   `1`
##   <dbl> <fct>   <dbl>   <dbl>   <dbl>     <dbl> <dbl> <dbl>
## 1     1 FEMALE     40      50      56         1   487     1
## 2     2 FEMALE     33      44      46         1   410     1
## 3     3 MALE       39      44      47         0   433     1
## 4     4 FEMALE     44      55      66         1   550     1
## 5     5 MALE       43      56      53         0   507     1
## 6     6 MALE       57      49      53         0   530     1
```

성적의 평균, 분산, 표준편차, 최소, 최대값을 계산하시오. 상자그림, 히스토그램, 정규확률그림을 작성하시오

성적의 평균, 분산, 표준편차, 최소, 최대값 계산

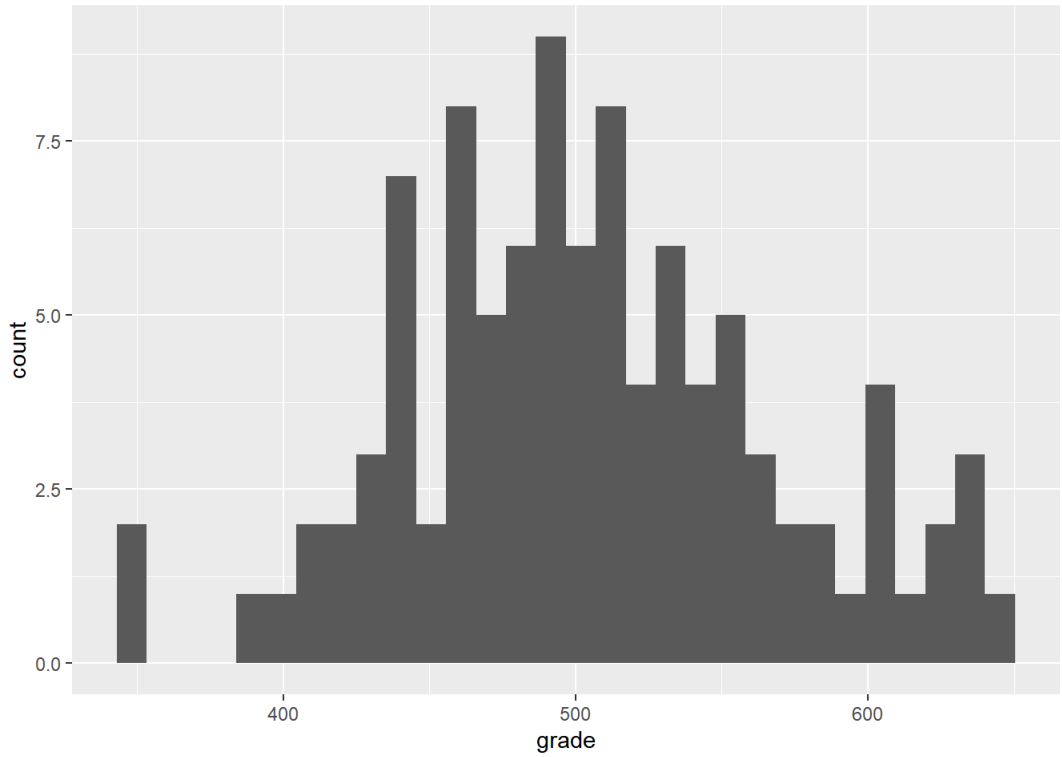
```
summarize(DF, n=n(), mean=mean(grade), var=var(grade), sd=sd(grade), min = min(grade), max(grade))
```

```
## # A tibble: 1 x 6
##       n  mean   var    sd   min `max(grade)`
##   <int> <dbl> <dbl> <dbl> <dbl>        <dbl>
## 1   100  505. 4003.  63.3   350          647
```

히스토그램

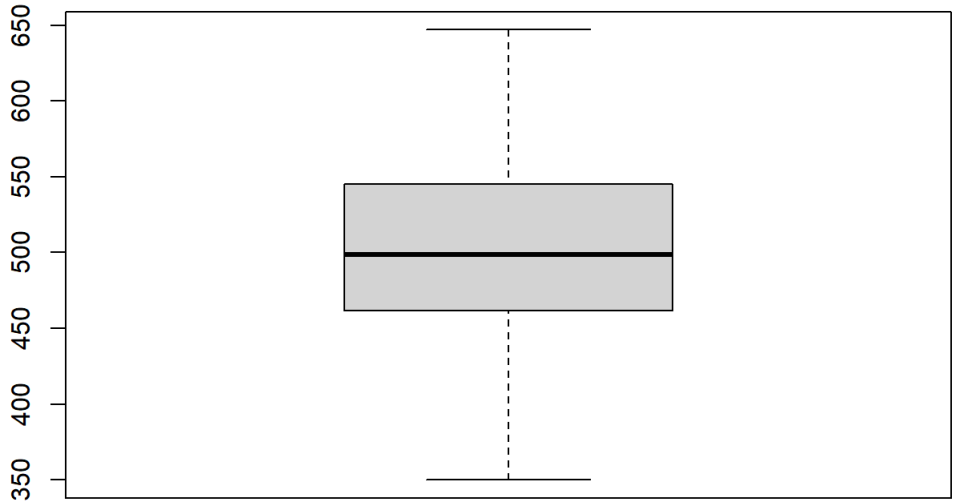
```
ggplot(DF, aes(x=grade)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



상자 그림

```
boxplot(DF$grade)
```



성별 성적의 평균, 분산, 표준편차, 최소, 최대값을 계산하시오. 성별 상자그림,히스토그램, 정규확률그림등을 작성하시오.

성별 성적의 평균, 분산, 표준편차, 최소, 최대값

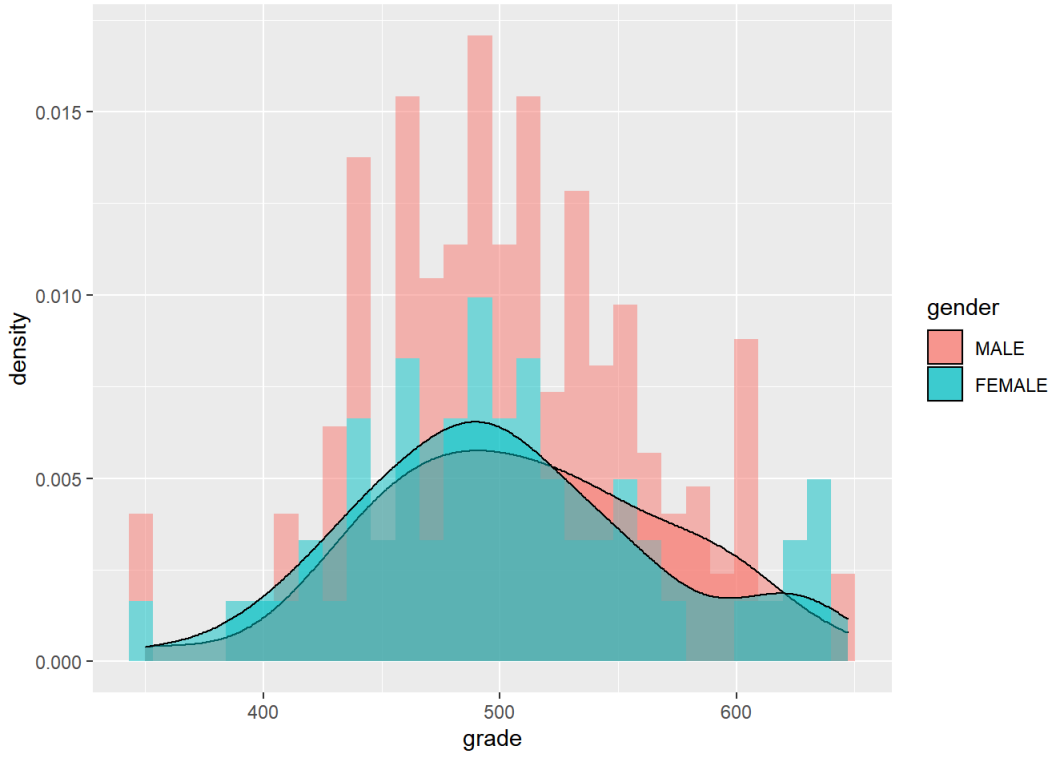
```
DF %>%
  group_by(gender) %>%
  summarize_at(vars(grade), list(mean=mean, var=var, sd=sd, min = min, max))
```

```
## # A tibble: 2 x 6
##   gender  mean   var    sd   min   fn1
##   <fct>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 MALE    510. 3921.  62.6   350   647
## 2 FEMALE  501. 4097.  64.0   350   640
```

성별 히스토그램

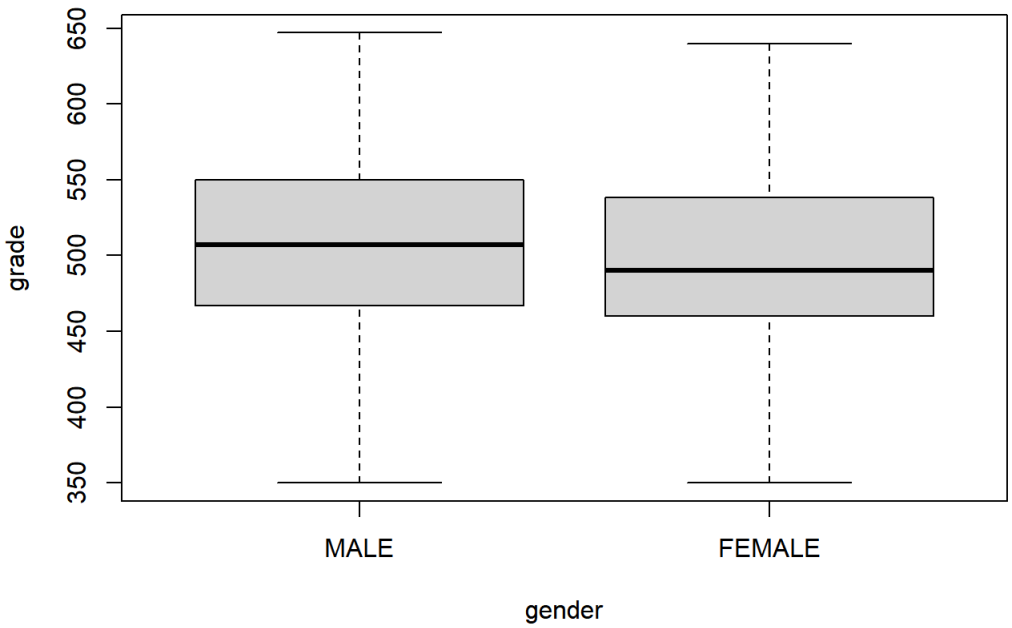
```
ggplot(DF) +
  geom_histogram(aes(x=grade, y=..density.., fill=gender), alpha=0.5) +
  geom_density(aes(x=grade, fill=gender), alpha=0.5)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



성별 상자그림

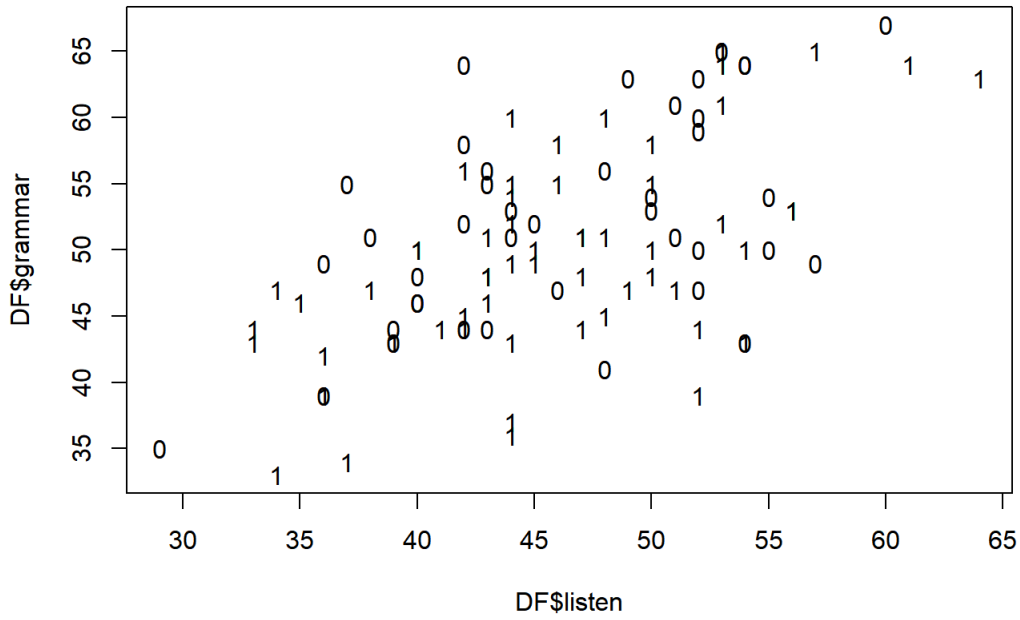
```
boxplot(grade~gender, data=DF)
```



듣기, 문법, 독해, 성적들의 산점도를 그리시오.

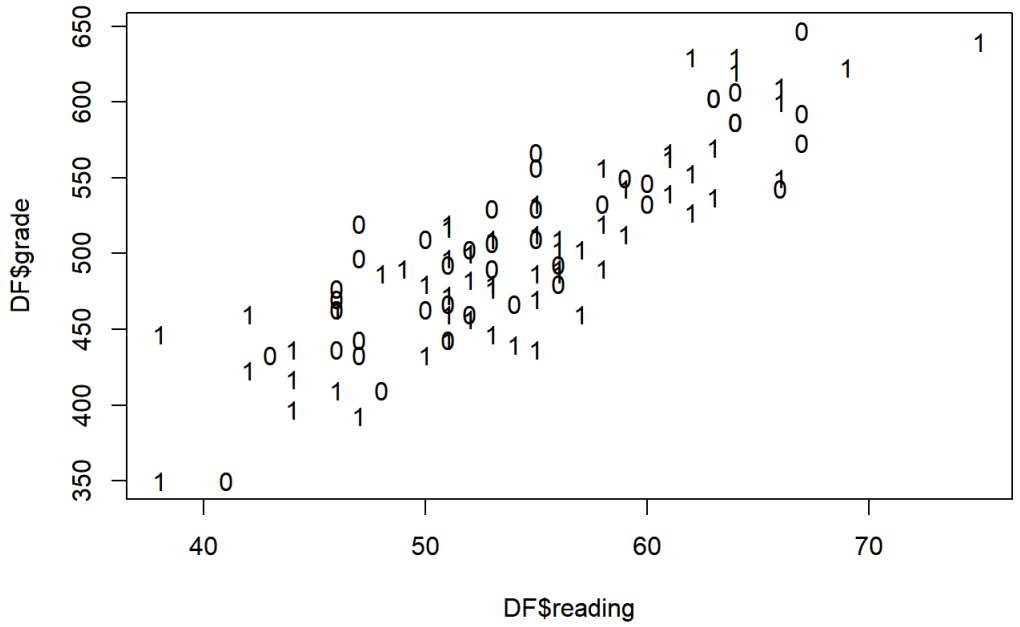
듣기 vs 문법

```
plot(DF$listen, DF$grammar, type='n')
text(DF$listen, DF$grammar, label=DF$gendernum)
```



읽기 vs 성적

```
plot(DF$reading, DF$grade, type='n')
text(DF$reading, DF$grade, label=DF$gendernum)
```



□성적이 500이상이면 1=응시료와 졸업시험면제, 450이상 500미만이면 2=졸업시험면제, 400이상 450미만이면 3=지원없음, 400미만이면 4=특강수강 하기로 하였다. 변수 지원{1,2,3,4}을 작성하시오

```
DF$support <- ifelse(DF$grade >= 500, 1,
                     ifelse(DF$grade >= 450, 2,
                            ifelse(DF$grade >= 400, 3, 4)))

head(DF)
```

```
## # A tibble: 6 x 9
##      id gender listen grammar reading gendernum grade   `1` support
##   <dbl> <fct>   <dbl>   <dbl>   <dbl>     <dbl> <dbl> <dbl>   <dbl>
## 1     1 FEMALE     40      50      56         1   487     1       2
## 2     2 FEMALE     33      44      46         1   410     1       3
## 3     3 MALE       39      44      47         0   433     1       3
## 4     4 FEMALE     44      55      66         1   550     1       1
## 5     5 MALE       43      56      53         0   507     1       1
## 6     6 MALE       57      49      53         0   530     1       1
```

성적이 450이상인 학생들은 졸업시험을 면제해주기로 하였다. 변수 면제 (1=성공,0=실패)를 만드시오.

```
DF$exemption <- ifelse(DF$support <= 2, 1, 0)

head(DF)
```

```
## # A tibble: 6 x 10
##      id gender listen grammar reading gendernum grade   `1` support exemption
##   <dbl> <fct>   <dbl>   <dbl>   <dbl>     <dbl> <dbl> <dbl>   <dbl>     <dbl>
## 1     1 FEMALE     40      50      56         1   487     1       2         1
## 2     2 FEMALE     33      44      46         1   410     1       3         0
## 3     3 MALE       39      44      47         0   433     1       3         0
## 4     4 FEMALE     44      55      66         1   550     1       1         1
## 5     5 MALE       43      56      53         0   507     1       1         1
## 6     6 MALE       57      49      53         0   530     1       1         1
```

지원의 빈도표와 면제의 빈도표를 작성하시오.

지원의 빈도표

```
sufreq <- table(DF$support)
margin.table(sufreq)
```

```
## [1] 100
```

비율로 변경

```
prop.table(sufreq)
```

```
##
##    1    2    3    4
## 0.50 0.30 0.16 0.04
```

면제의 빈도표

```
exfreq <- table(DF$exemption)
margin.table(exfreq)
```

```
## [1] 100
```

비율로 변경

```
prop.table(exfreq)
```

```
##
##   0   1
## 0.2 0.8
```

성별*지원, 성별*면제의 교차표를 작성하시오.

성별 vs 지원

```
sutbl <- table(DF$gender, DF$support)
sutbl <- xtabs(~gender + support, data=DF)
sutbl
```

```
##         support
## gender    1  2  3  4
##   MALE   22 12  6  1
##   FEMALE 28 18 10  3
```

비율로 변경

```
prop.table(sutbl)
```

```
##         support
## gender      1    2    3    4
##   MALE   0.22 0.12 0.06 0.01
##   FEMALE 0.28 0.18 0.10 0.03
```

성별 vs 면제

```
extbl <- table(DF$gender, DF$exemption)
extbl <- xtabs(~gender+exemption, data=DF)
extbl
```

```
##         exemption
## gender    0  1
##   MALE    7 34
##   FEMALE 13 46
```

비율로 변경

```
prop.table(extbl)
```

```
##         exemption
## gender      0    1
##   MALE   0.07 0.34
##   FEMALE 0.13 0.46
```