한국인의 삶을 파악하라

- 2006~2015년까지 전국에서 7000여 가구를 선정하여 매년 추적 조사한 자료
- 데이터 셋 : Koweps_hpc10_2015_beta1.sav
    - 2016년도 발간한 복지패널 데이터 6,914가구, 16,664명에 대한 정보

In [1]:

```
install.packages("foreign")
```

```
Warning message:
"unable to access index for repository http://www.stats.ox.ac.uk/pub/RWin/bin/window
s/contrib/3.5: (http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/3.5:)
  URL 'http://www.stats.ox.ac.uk/pub/RWin/bin/windows/contrib/3.5/PACKAGES'를 열 수
 없습니다"

package 'foreign' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:₩Users₩WITHJS₩AppData₩Local₩Temp₩RtmpWlu2Cd₩downloaded_packages
```

In [2]:

```
library(foreign)
library(dplyr)
library(ggplot2)
library(readxl)
```

```

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

```

In [4]:

```
dat_welfare <- read.spss(file="D:₩₩dataset₩₩R_Doit₩₩Koweps_hpc10_2015_beta1.sav", to.data.frame=T)
welfare <- dat_welfare
```

```
Warning message in read.spss(file = "D:₩₩dataset₩₩R_Doit₩₩Koweps_hpc10_2015_beta1.sa
v", :
"D:₩dataset₩R_Doit₩Koweps_hpc10_2015_beta1.sav: Compression bias (0) is not the usua
l value of 100"
```

데이터 탐색해 보기

- head(welfare)

- tail(welfare)
- View(welfare)
- dim(welfare)
- str(welfare)
- summary(welfare)

In [6]:

```
welfare <- rename(welfare,
                  sex=h10_g3,
                  birth=h10_g4,
                  marriage=h10_g10,
                  religion=h10_g11,
                  income=p1002_8aq1,
                  code_job=h10_eco9,
                  code_region=h10_reg7)
names(welfare)
```

```
'h10_id'  'h10_ind'  'h10_sn'  'h10_merkey'  'h_new'  'h10_cobf'  'h10_reg5'
'code_region'  'h10_din'  'h10_cin'  'h10_flag'  'p10_wgl'  'p10_wsl'  'p10_wgc'
'p10_wsc'  'h10_hc'  'nh1001_1'  'nh1001_2'  'h1001_1'  'h10_pind'  'h10_pid'
'h10_g1'  'h10_g2'  'sex'  'birth'  'h10_g6'  'h10_g7'  'h10_g8'  'h10_g9'  'marriage'
'religion'  'h10_g12'  'h1001_110'  'h1001_5aq1'  'h1001_5aq2'  'h1001_5aq3'
'h1001_5aq4'  'h10_med1'  'h10_med2'  'h10_med3'  'h10_med4'  'h10_med5'
'h10_med6'  'h10_med7'  'h10_med8'  'h10_g9_1'  'h10_med9'  'h10_med10'
'h10_eco1'  'h10_eco2'  'h10_eco3'  'h10_eco4'  'h10_eco4_1'  'h10_eco5_1'
'h10_eco6'  'h10_eco_7_1'  'h10_eco_7_2'  'h10_eco_7_3'  'h10_eco8'  'code_job'
'h10_eco10'  'h10_eco11'  'h10_soc1'  'h10_soc_2'  'h10_soc_3'  'h10_soc_4'
'h10_soc_5'  'h10_soc_6'  'h10_soc_7'  'h10_soc_8'  'h10_soc_9'  'h10_soc_10'
'h10_soc_11'  'h10_soc8'  'h10_soc9'  'h10_soc11'  'h10_soc10'  'h10_soc_12'
'h10_soc_13'  'h1005_1'  'h1005_3aq1'  'h1005_2'  'h1005_3'  'h1005_4'  'h1005_5'
'h1005_6'  'h1005_7'  'nh1005_8'  'nh1005_9'  'h1005_3aq2'  'h1006_aq1'  'h1006_1'
'h1006_2'  'h1006_4'  'h1006_5'  'h1006_3'  'h1006_6'  'h1006_8'  'h1006_9'
'h1006_aq2'  'h1006_aq3'  'h1006_10'  'h1006_11'  'h1006_12'  'h1006_13'
'h1006_14'  'h1006_21'  'h1006_22'  'h1006_23'  'h1006_24'  'h1006_25'  'h1006_27'
```

미션 - 성별에 따른 월급 차이는 있을까?

- 변수 : 성별, 월급
- 성별, 월급 평균표 만들기
- 그래프 확인

01 성별 검토

In [7]:

```
class(welfare$sex)
```

```
'numeric'
```

In [8]:

```
table(welfare$sex)
```

```

   1    2 
7578 9086 
```

- 1: 남자
- 2: 여자
- 9: 응답 없음.

만약 존재할 수 있으므로, 결측치 처리해야함.

In [10]:

```
welfare$sex <- ifelse(welfare$sex == 9, NA, welfare$sex)
table(is.na(welfare$sex))  # 결측치 확인
```

```

FALSE 
16664 
```

변수 1(남자), 2(여자)로 전처리

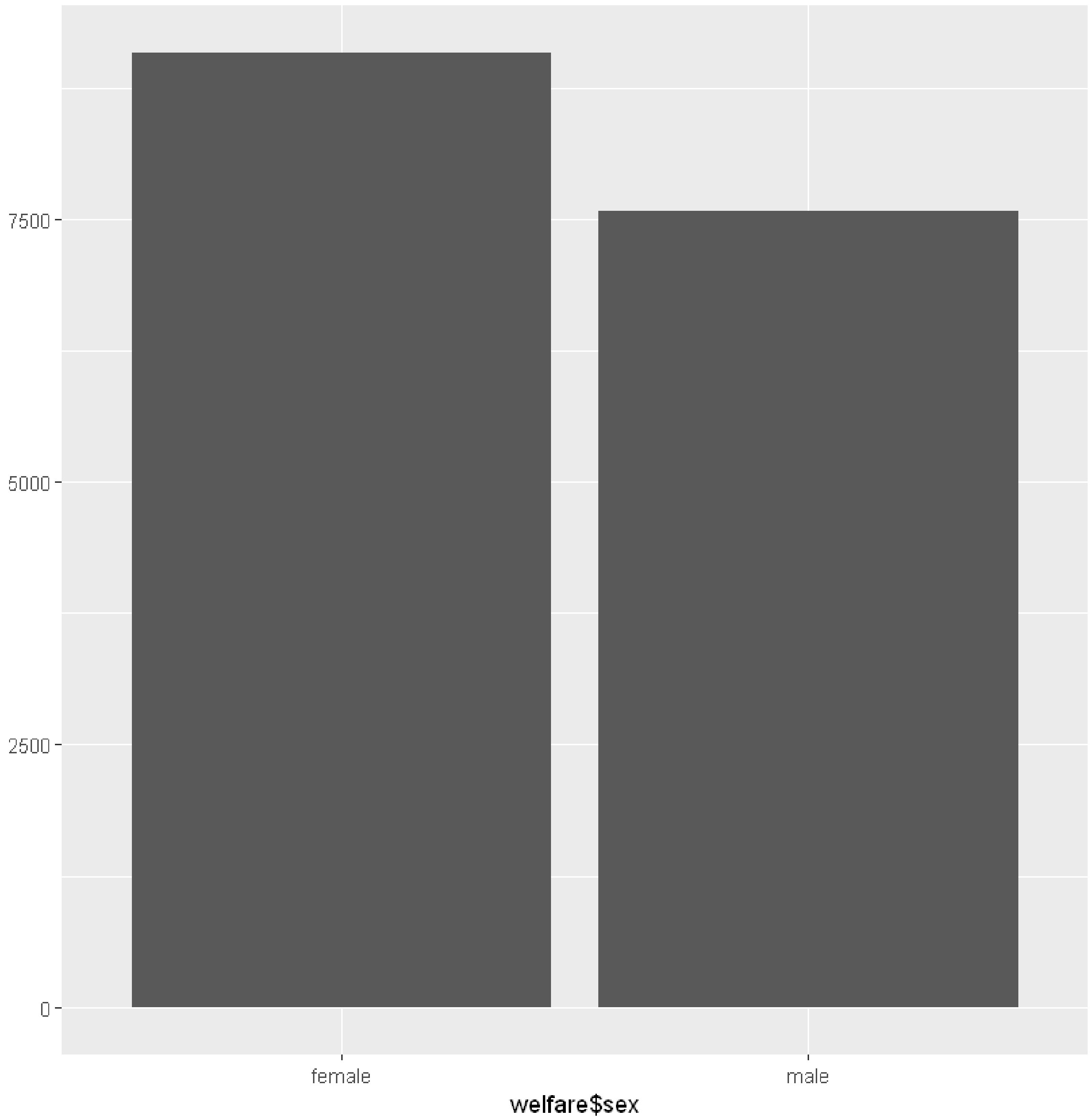
In [11]:

```
welfare$sex <- ifelse(welfare$sex == 1, "male", "female")
table(welfare$sex)
```

```

female   male 
  9086   7578 
```

In [12]:

```
qplot(welfare$sex)
```



02 월급 검토

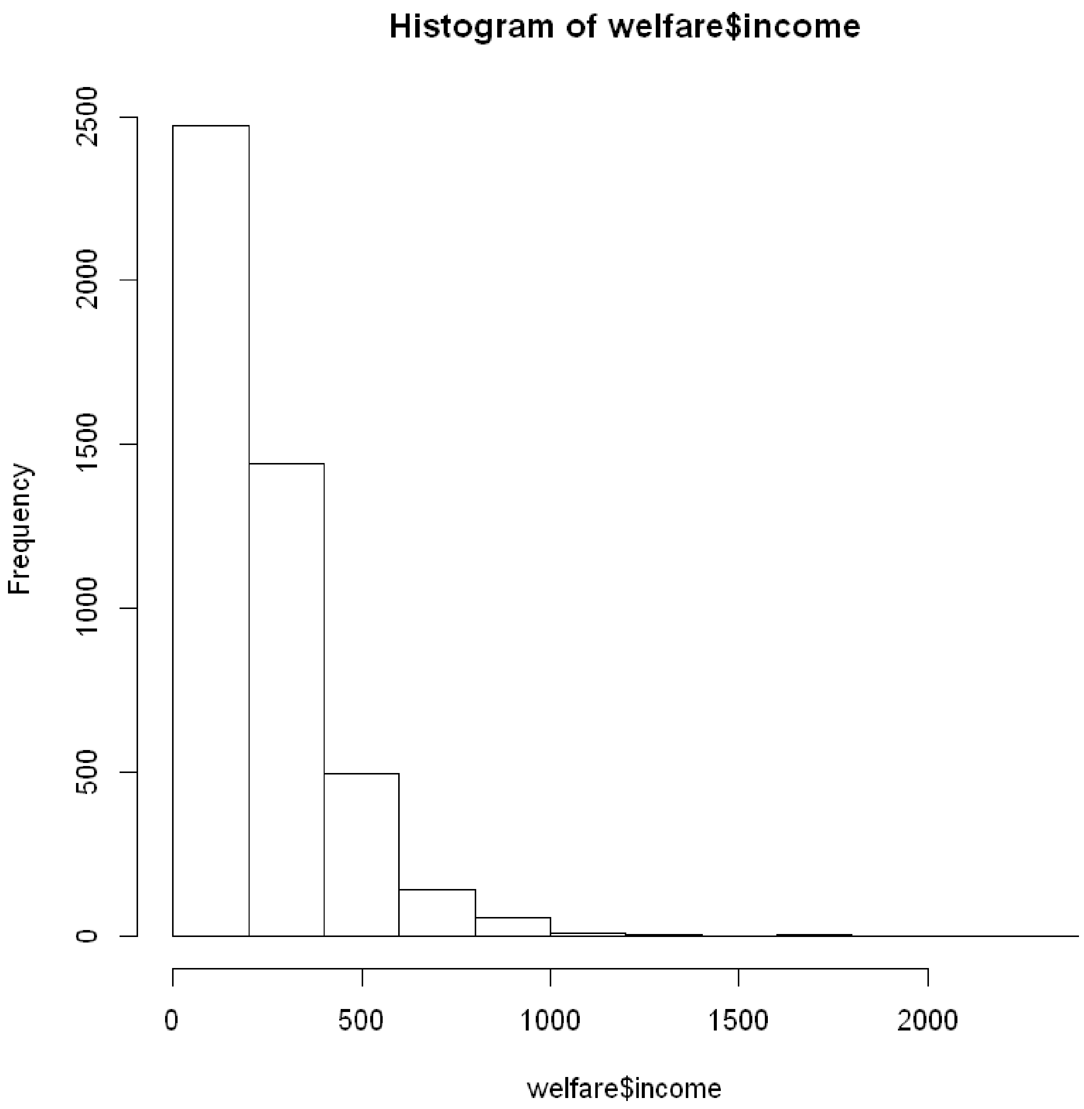
In [13]:

```
names(welfare)
```

```
'h1012_14'   'h1012_15'   'h1012_3aq1'   'h1012_16'   'h1012_17'   'h1012_18'
'h1012_19'   'h1012_1_4aq1'   'h1012_1_5aq1'   'h1012_1_5aq2'   'h1012_1_5aq3'
'h1012_1_5aq4'   'h1012_1_5aq5'   'h1012_1_4aq2'   'h1012_1_4aq3'   'h1013_2'
'h1013_6'   'h1013_10'   'h1013_14'   'h1013_18'   'h1013_22'   'h1013_26'
'h1013_8aq1'   'h1013_5aq1'   'h1013_8aq2'   'h1013_4aq1'   'h1013_4aq2'
'h1013_4aq4'   'h1013_4aq6'   'h1013_4aq8'   'h1013_4aq10'   'h1013_5aq4'
'h1013_5aq6'   'h1013_5aq8'   'h1013_6aq1'   'h1013_4aq14'   'h1013_4aq15'
'h1013_4aq16'   'h1013_4aq17'   'h1013_4aq18'   'h1013_4aq20'   'h1013_4aq22'
'h1013_4aq24'   'h1013_4aq26'   'h1013_4aq28'   'h1013_4aq30'   'h1013_4aq32'
'h1014_4'   'h1014_8'   'h1014_12'   'h1014_16'   'h1014_20'   'h1014_24'   'h1014_28'
'h1014_32'   'h1014_36'   'h1014_3aq1'   'h1014_4aq1'   'h1015_4'   'h1015_8'
'h1015_12'   'h1015_20'   'h1015_25'   'h1015_29'   'h1015_33'   'h1015_37'
'h1015_4aq1'   'h1015_7aq1'   'h1015_aq1'   'h1015_40'   'h1015_41'   'h1015_42'
'h1015_43'   'h1015_44'   'h1015_45'   'h1015_46'   'h1015_47'   'h1015_48'   'h1015_49'
'h1015_50'   'h1015_51'   'h1015_52'   'h1015_53'   'h1015_54'   'h1015_55'   'h1015_56'
'h1015_57'   'h1015_60'   'h1015_aq2'   'h1015_61'   'h1015_62'   'h1015_63'
'h1015_66'   'h1015_67'   'h1015_68'   'h1015_aq3'   'h1015_69'   'h1015_70'
```

In [15]:

```
hist(welfare$income)
```



In [16]:

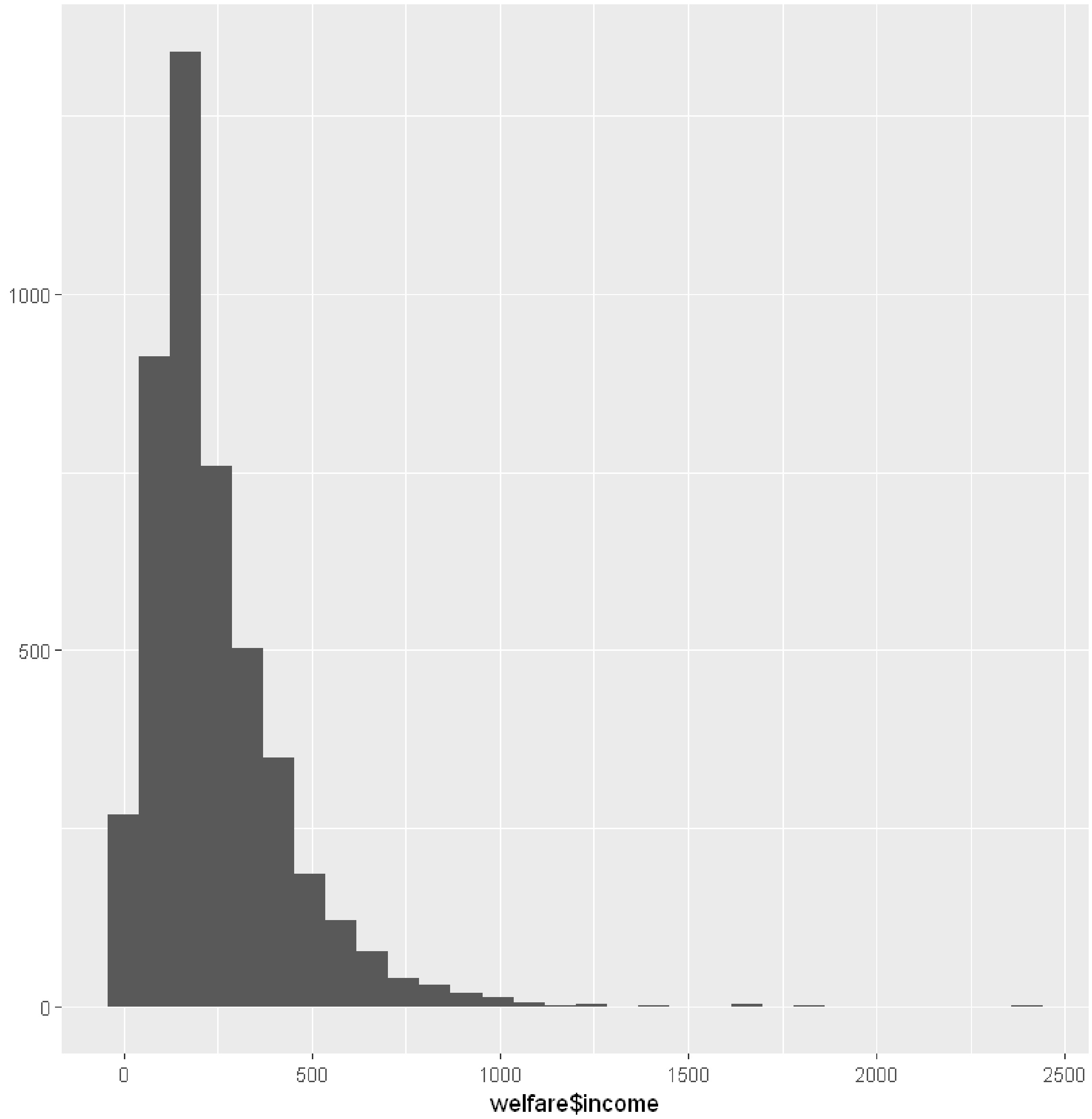
```
summary(welfare$income)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
    0.0   122.0   192.5   241.6   316.6  2400.0   12030 
```

In [17]:

```
### 자세히 보자.
qplot(welfare$income)
```

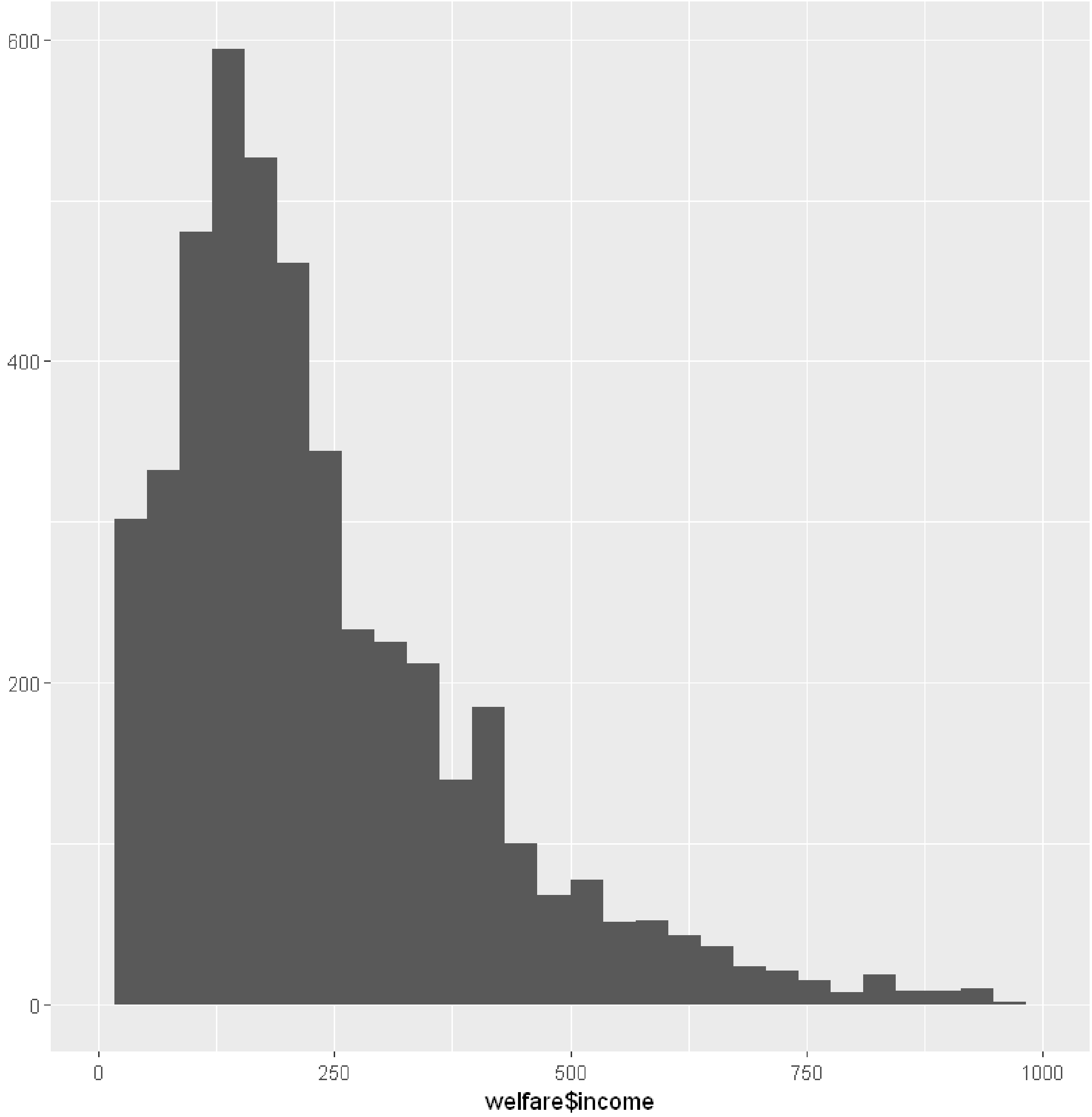
```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
Warning message:
"Removed 12030 rows containing non-finite values (stat_bin)."
```



In [18]:

```
### 자세히 보자.
qplot(welfare$income) + xlim(0,1000)
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
Warning message:
"Removed 12051 rows containing non-finite values (stat_bin)."
```



NA를 전처리

In [19]:

```
summary(welfare$income)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
    0.0   122.0   192.5   241.6   316.6  2400.0   12030 
```

In [20]:

```
### 모름/무응답 = 9999
### 범위 1~9998 이므로 0도 결측치 처리
welfare$income <- ifelse(welfare$income %in% c(0,9999), NA, welfare$income)
table(is.na(welfare$income))
```

```
FALSE  TRUE 
 4620 12044 
```

성별에 따른 월급 차이 분석

In [21]:

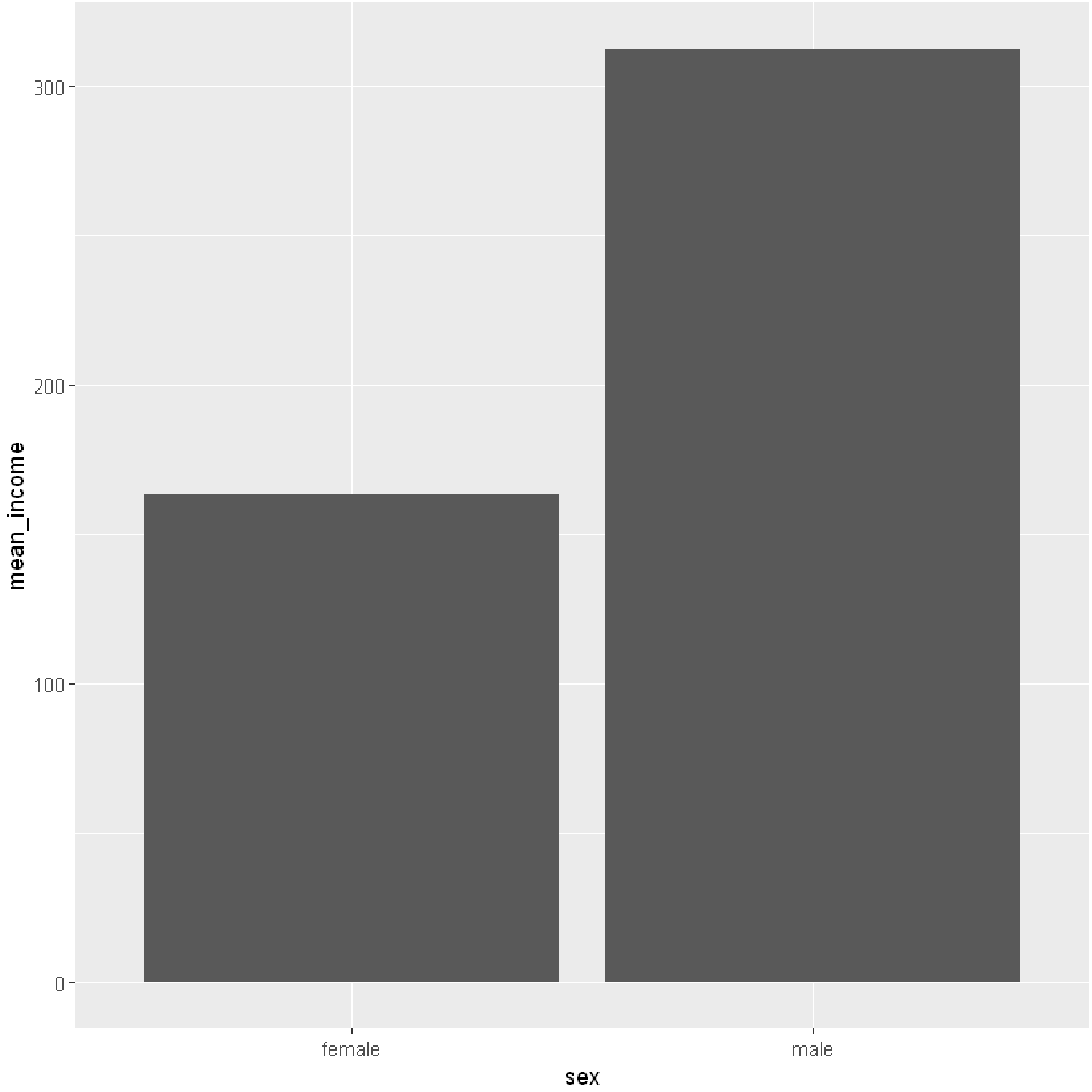
```
sex_income <- welfare %>% filter(!is.na(income)) %>%
                          group_by(sex) %>%
                          summarise(mean_income = mean(income))

sex_income
```

sex	mean_income
female	163.2471
male	312.2932

- 월급 평균은 남자가 312만원, 여자는 163만원으로 평균적으로 여성보다 남성의 월급이 약 150만원 많다.

In [24]:

```
ggplot(data = sex_income, aes(x=sex, y=mean_income)) + geom_col()
```



함수	설명
geom_point()	산점도
geom_col()	막대 그래프, X축, Y축을 모두 설정
geom_bar()	막대 그래프, X축만 설정, Y축은 해당 데이터의 수량
geom_line()	선 그래프
geom_boxplot()	박스 그래프

In [ ]: