

# 2014 년 주거실태를 이용한 분석보고서

20196384 유나현

## 1. 어느 유형의 주택이 가장 많고, 만족도는 어떨까?

### R 소스코드

```
install.packages("dplyr")
```

```
install.packages("ggplot2")
```

```
install.packages("readxl")
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(readxl)
```

```
df<-read_excel("2014 년도 주거실태조사_공표자료(시군구 명칭 수정).xlsx")
```

```
df2 <- df %>% select (Stype, sido, dq1, q4, q6, q7, q9_1, q18, q23_13, q24_1, q24_2, q26_1,
q26_2, dq2_1, q41, q41_1, dq2_3, q40_2, q40_3, q49_5, q48, q8_1, q47_a2_1, q23_9, q42_2,
q47_a3_1)
```

```
df2 <- rename(df2, house_type = q4, house_location = q6, house_money = q7,
no_house_time = q9_1, house_when = q18, friendly = q23_13, satisfy = q24_1, satisfy2 =
q24_2, camera = q26_1, ring = q26_2, job = dq2_1, thinking = q41, thinking2 = q41_1, vehicle
= dq2_3, school = dq1, money1 = q40_2, money2 = q40_3, earn = q49_5, poor = q48, when
= q8_1, age = q47_a2_1, danger = q23_9, want = q42_2, master = q47_a3_1)
```

```
df2$house_type <- ifelse(df2$house_type == 1, '일반단독주택', ifelse(df2$house_type == 2,
"다가구단독주택", ifelse(df2$house_type == 3, '영업겸용단독주택', ifelse(df2$house_type == 4,
'아파트', ifelse(df2$house_type == 5, '연립주택', ifelse(df2$house_type == 6, '다세대주택',
ifelse(df2$house_type == 7, "비거주용건물", ifelse(df2$house_type == 8, "오피스텔",
ifelse(df2$house_type == 9, "판잣집", '기타')))))))))))
```

```
table(df2$house_type)
```

```
qplot(df2$house_type)+coord_flip()
```

```
count_sido <- df2 %>% filter(df2$house_type == '아파트') %>%
```

```
group_by(sido) %>%
```

```
summarise(n = n()) %>%
```

```
arrange(-n)
```

```
count_sido
```

```
sido_satisfy <- df2 %>% filter(df2$house_type == '아파트') %>%
```

```
group_by(sido) %>%
```

```
summarise(satisfy_mean = mean(satisfy),
```

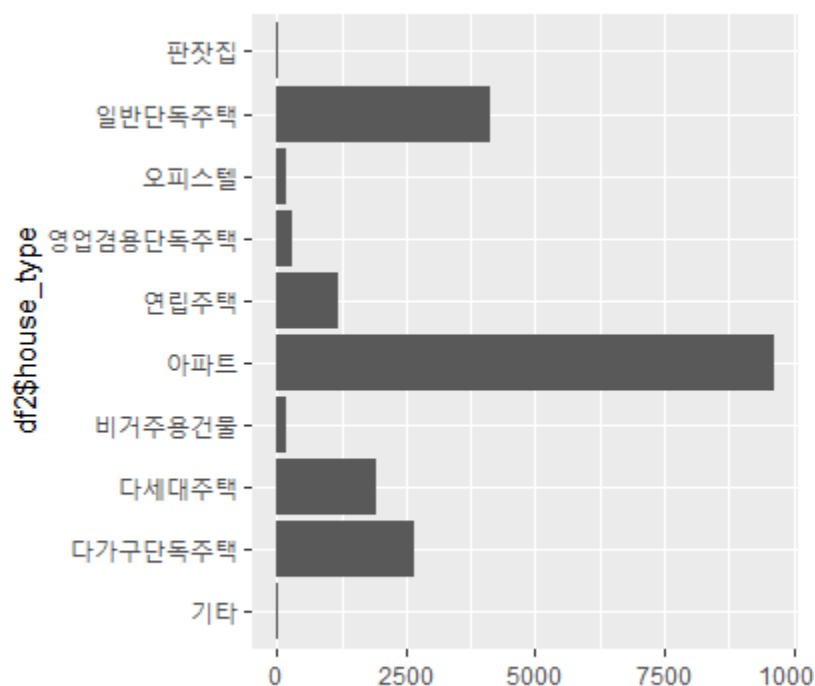
```
satisfy_mean2 = mean(satisfy2)) %>%
```

```
mutate(total = (satisfy_mean + satisfy_mean2) / 2) %>%
```

```
arrange(-total)
```

```
sido_satisfy
```

## 자료분석 결과물과 해석



거주 유형 빈도 분석을 한 그래프를 보면 아파트가 가장 많은 것을 볼 수 있다 이에 따라 1. 매우 불만족 2. 약간 불만족 3. 대체로 만족 4. 매우 만족으로 지역별 아파트의 주거환경 만족도와 전반적인 만족도의 조사 결과의 평균을 낸 것을 살펴보면 다음과 같다.

	sido	satisfy_mean	satisfy_mean2	total
1	전라북도	3.360656	3.000000	3.180328
2	대전광역시	3.140230	3.098851	3.119540
3	광주광역시	3.124434	3.052036	3.088235
4	대구광역시	3.099684	3.068038	3.083861
5	강원도	3.132143	3.007143	3.069643
6	제주도	3.052632	3.070175	3.061404
7	부산광역시	3.073583	2.993969	3.033776
8	서울특별시	3.015161	3.036639	3.025900
9	전라남도	2.955017	3.069204	3.012111
10	경상북도	3.029703	2.973597	3.001650
11	충청남도	3.123167	2.859238	2.991202
12	경상남도	2.986486	2.961712	2.974099
13	울산광역시	3.002994	2.934132	2.968563
14	경기도	2.999117	2.932039	2.965578
15	인천광역시	3.011551	2.914191	2.962871
16	충청북도	3.043011	2.845878	2.944444
17	세종특별자치시	2.914286	2.971429	2.942857

전라북도가 주거환경 만족도와 주택의 전반적인 만족도의 평균이 3.18 로 가장 높고 대전광역시, 광주광역시가 3.12, 3.08 로 상위권을 차지 하는 것을 확인 할 수 있다.

## 2. 최종 학력이 주거 형태에 영향을 미칠까?

### R 소스코드

```
df2$school<-ifelse(df2$school==1,"초등학교 졸업 이하",
                  ifelse(df2$school==2,"중학교 졸업",
                        ifelse(df2$school==3,"고등학교 졸업","대학 졸업 이상")))

school_df <- df2 %>%

group_by(school, house_type) %>%

summarise(n = n()) %>%
```

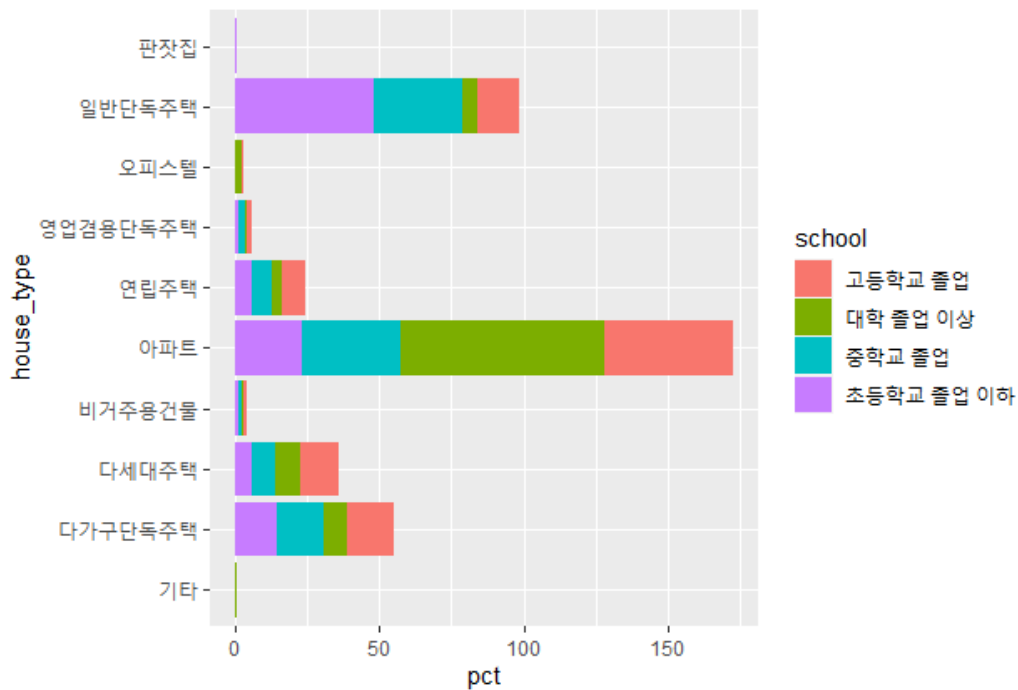
```
mutate(total = sum(n),
```

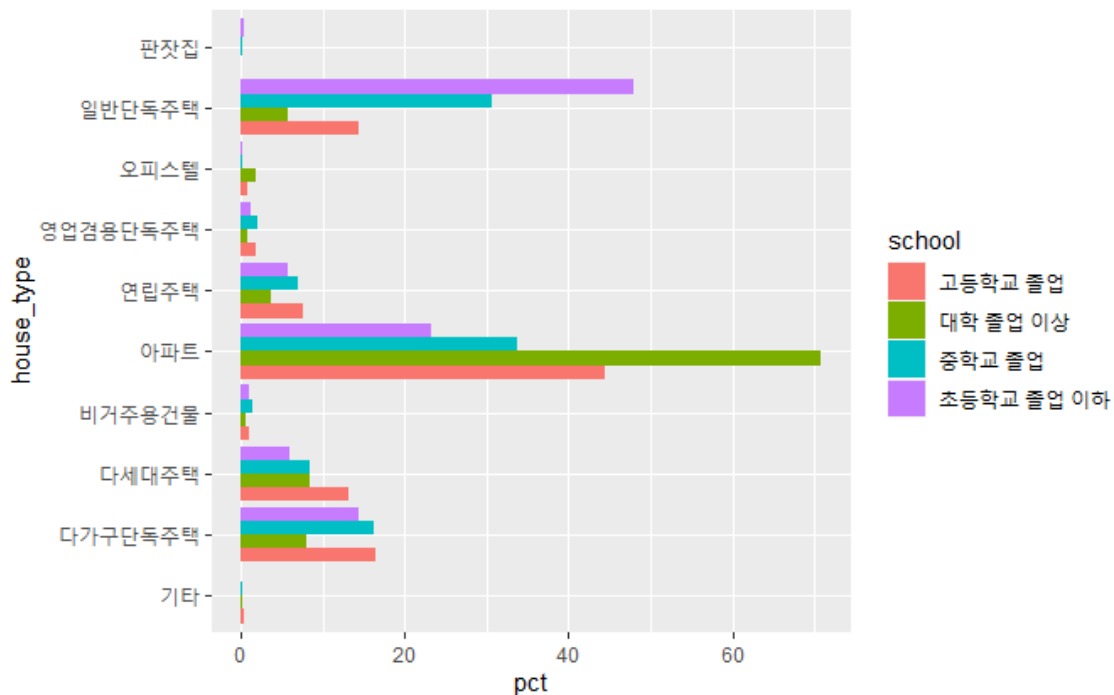
```
  pct = round(n / total * 100, 1))
```

```
ggplot(data = school_df, aes(x = house_type, y = pct, fill = school)) + geom_col() +  
coord_flip()
```

```
ggplot(data = school_df, aes(x = house_type, y = pct, fill = school)) +  
  
  geom_col(position='dodge') + coord_flip()
```

## 자료분석 결과물과 해석





학력 별 사는 곳에 대한 분석 그래프는 다음과 같다. 학력이 대학 졸업 이상과 고등학교를 졸업한 사람들은 아파트에 가장 많이 살고 있고, 초등학교 졸업 이하인 학력을 가진 사람들은 일반 단독 주택에 많이 사는 것을 확인할 수 있다.

### 3. 치안 및 범죄 등 방범상태가 가장 취약한 지역은 어디일까?

#### R 소스코드

```
df2$sido<-
ifelse(df2$sido==11,"서울특별시",ifelse(df2$sido==21,"부산광역시",ifelse(df2$sido==22,"대구
광역시",ifelse(df2$sido==23,"인천광역시",ifelse(df2$sido==24,"광주광역시",ifelse(df2$sido==2
5,"대전광역시",ifelse(df2$sido==26,"울산광역시",ifelse(df2$sido==29,"세종특별자치시",ifelse(df
2$sido==31,"경기도",ifelse(df2$sido==32,"강원도",ifelse(df2$sido==33,"충청북도",ifelse(df2$si
do==34,"충청남도",ifelse(df2$sido==35,"전라북도",ifelse(df2$sido==36,"전라남도",ifelse(df2$s
ido==37,"경상북도",ifelse(df2$sido==38,"경상남도",ifelse(df2$sido==39,"제주도","기타"))))))))
))))))

danger_table = df2 %>%

group_by(sido, danger) %>%
```

```
summarise(n = n()) %>%
```

```
mutate(total = sum(n),
```

```
pct = n / total * 100) %>%
```

```
filter(danger == 1) %>%
```

```
arrange(-pct)
```

danger\_table

#### 자료분석 결과물과 해석

	sido	danger	n	total	pct
1	경상북도	1	38	843	4.5077106
2	울산광역시	1	25	614	4.0716612
3	인천광역시	1	43	1129	3.8086802
4	전라북도	1	26	726	3.5812672
5	세종특별자치시	1	10	320	3.1250000
6	서울특별시	1	118	3840	3.0729167
7	경기도	1	127	4182	3.0368245
8	광주광역시	1	19	704	2.6988636
9	충청남도	1	18	776	2.3195876
10	대구광역시	1	26	1218	2.1346470
11	제주도	1	8	450	1.7777778
12	충청북도	1	11	660	1.6666667
13	부산광역시	1	22	1589	1.3845186
14	대전광역시	1	10	783	1.2771392
15	강원도	1	7	682	1.0263930
16	전라남도	1	7	694	1.0086455
17	경상남도	1	8	995	0.8040201

현재 주택 주거환경 만족도 중 치안 및 범죄 등 방법 상태에 대한 문항에 1. 매우 불만족으로 표기한 사람의 비율이 경상북도(4.50%), 울산광역시(4.07%), 인천광역시(3.80%) 순으로 많은 것을 보아 방법 상태가 취약한 곳이 많은 것을 알 수 있다. 반면에 경상남도(0.80%), 전라남도(1.00%), 강원도(1.02%)는 상대적으로 방법 상태가 좋은 것을 알 수 있다.

#### 4. 범죄 예방용 CCTV 가 부족한 곳은 어디일까?

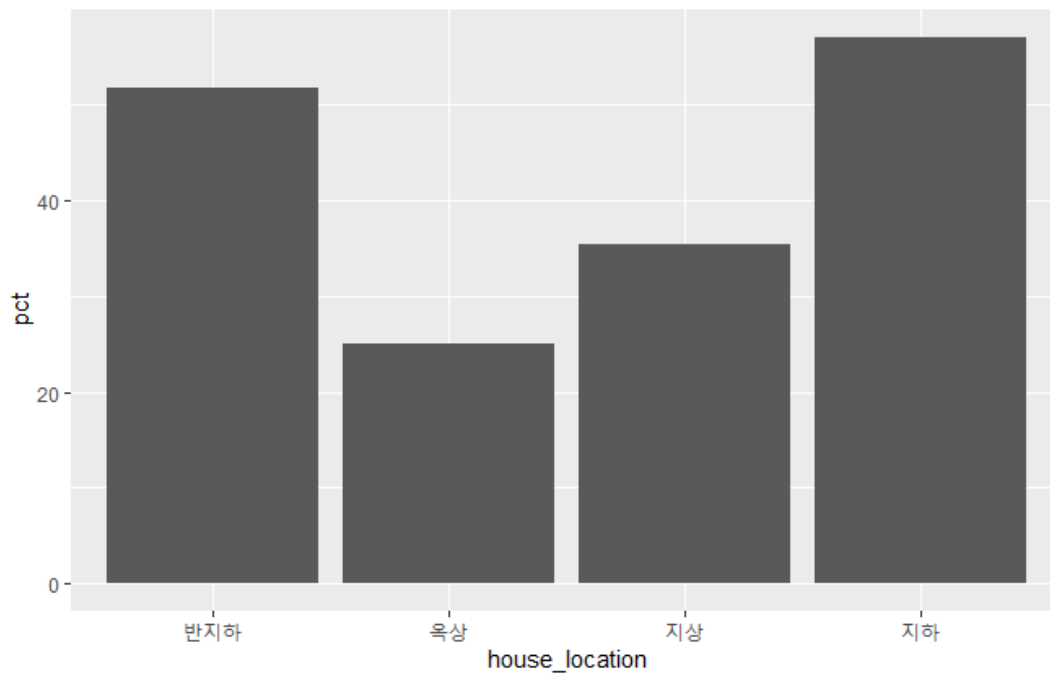
##### R 소스코드

```
df2$house_location <- ifelse(df2$house_location == 1, "지상",  
                             ifelse(df2$house_location == 2, "반지하",  
                                     ifelse(df2$house_location == 3, "지하", "옥상"))  
  
df2$camera <- ifelse(df2$camera == 1, "있음",  
                    ifelse(df2$camera == 2, "없음", "모르겠음"))  
  
camera_location <- df2 %>% group_by(house_location, camera) %>%  
  
  summarise(n = n()) %>%  
  
  mutate(total = sum(n),  
  
         pct = round(n / total * 100, 1)) %>%  
  
  filter(camera == '없음') %>%  
  
  select(house_location, pct)  
  
ggplot(camera_location, aes(x = house_location, y = pct)) + geom_col()
```

##### 자료분석 결과물과 해석

범죄 예방시설(CCTV) 설치 여부를 표기한 문항에서 2. 없음에 표기한 사람들이 많은 곳을 분석해 보면 다음과 같다.





반지하, 옥상, 지상, 지하 중 CCTV 가 없는 비율을 분석한 막대그래프를 보면, 반지하와 지하에 범죄 예방용 CCTV 가 적은 것을 확인할 수 있다. 사각지대가 많아 상대적으로 위험한 곳인 반지하와 지하에 CCTV 가 부족하여 추가적으로 설치되어야 할 것으로 보인다.

## 5. 연령대에 따라 향후 살고 싶은 생활양식이 다를까?

### R 소스코드

```
df2 <- df2 %>% mutate(ageg = ifelse(real_age < 30, 'young',
                                   ifelse(real_age <= 59, 'middle', 'old')))

df2$want<-ifelse(df2$want==1,"도시적 생활",
                 ifelse(df2$want==2,"전원생활","잘모르겠음"))

want_table = df2 %>%

  group_by(ageg, want) %>%

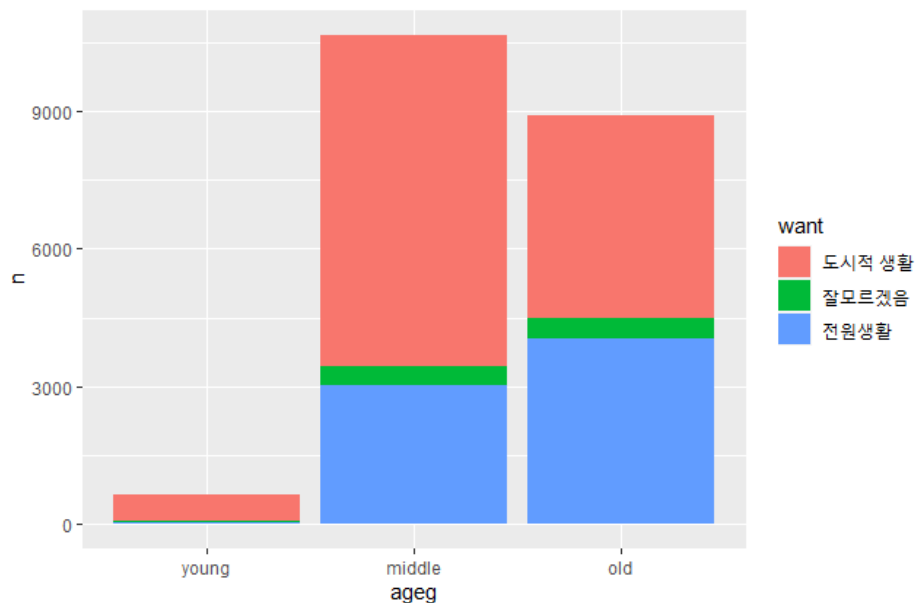
  summarise(n = n())

want_table
```

```
ggplot(data = want_table, aes(x = ageg, y = n, fill = want)) + geom_col() +  
  
  scale_x_discrete(limits = c('young', 'middle', 'old'))
```

## 자료분석 결과물과 해석

30 세 미만은 young, 30 세 이상 59 세 이하은 middle, 59 세 초과는 old 로 분류하여 향후 살고 싶은 생활양식(도시적 생활과 전원생활)의 선호도를 분석한 결과는 다음과 같다.



세 연령대 모두 전원생활을 선호하는 사람보다 도시적 생활을 선호하는 사람이 많지만 나이가 들수록 전원생활을 선호하는 사람의 비율이 커지는 것을 볼 수 있다.

## 6. 서울 아파트 자가 비율은 얼마일까?

### R 소스코드

```
df2$house_money<-  
ifelse(df2$house_money==1,"자가",ifelse(df2$house_money==2,"전세",ifelse(df2$house_mone  
y==3,"보증금 있는 월세",ifelse(df2$house_money==4,"보증금 없는  
월세",ifelse(df2$house_money==5,"사글세 또는  
연세",ifelse(df2$house_money==6,"일세","무상"))))))  
  
current_mine <- df2 %>%  
  
  filter(sido == "서울특별시" & house_type == "아파트") %>%
```

```
group_by(house_money) %>%
```

```
summarise(n = n()) %>%
```

```
mutate(total = sum(n), pct = n / total * 100)
```

current\_mine

## 자료분석 결과물과 해석

	house_money	n	total	pct
1	무상	30	1583	1.89513582
2	보증금 없는 월세	2	1583	0.12634239
3	보증금 있는 월세	150	1583	9.47567909
4	일세	1	1583	0.06317119
5	자가	932	1583	58.87555275
6	전세	468	1583	29.56411876

서울 아파트의 점유 형태를 조사한 결과 자가 비율이 58.8%으로 서울 아파트의 과반수 이상이 자가이다. 전세는 29.5% 인 것을 확인할 수 있다.

## 7. 직장까지 어떤 교통수단을 많이 이용 할까?

### R 소스코드

```
table(is.na(df2$vehicle))
```

```
df2$vehicle<-
```

```
ifelse(df2$vehicle==1,"승용차",ifelse(df2$vehicle==2,"대중교통",ifelse(df2$vehicle==3,"보도",ifelse(df2$vehicle==4,"자전거",ifelse(df2$vehicle==5,"오토바이","기타")))))
```

```
vehicle_nomiss<-df2 %>%
```

```
filter(!is.na(vehicle)) %>%
```

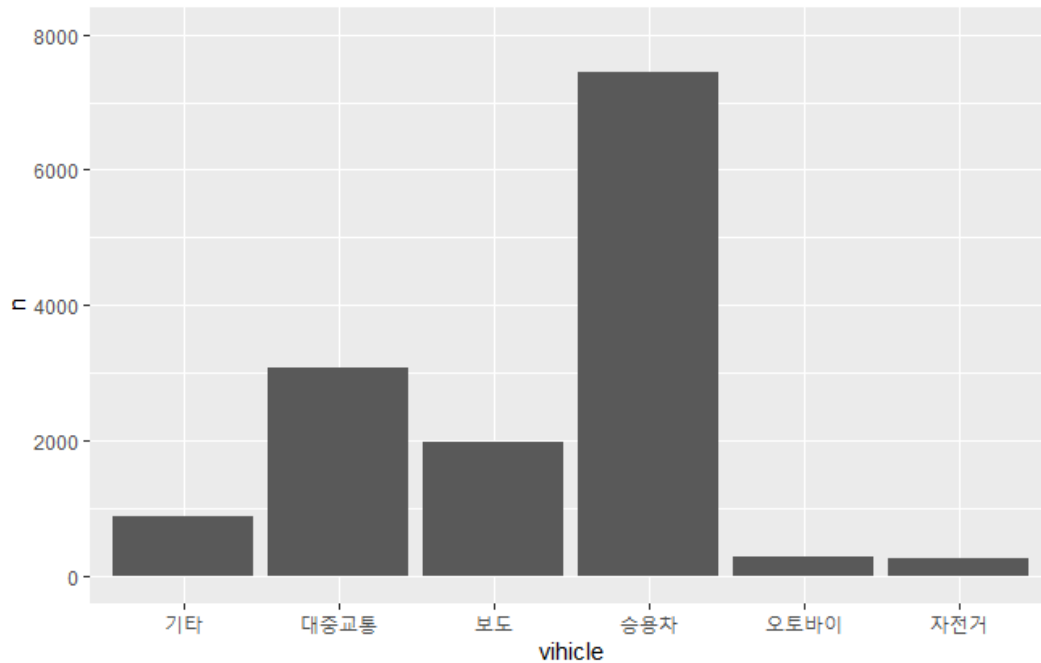
```
group_by(vehicle) %>%
```

```
summarise(n=n())
```

vehicle\_nomiss

```
ggplot(data = vehicle_nomiss, aes(x = vehicle, y = n)) + geom_col() + ylim(0, 8000)
```

#### 자료분석 결과물과 해석



결측치를 제외한 막대 그래프는 다음과 같다. 많은 사람이 출퇴근시 승용차를 이용하고 대중교통, 보도 순으로 이용하는 것을 확인할 수 있다.

8. 내 집을 꼭 마련 해야 한다고 생각하는 사람이 많을까? 그 이유는 무엇일까?

#### R 소스코드

```
df2$thinking <- ifelse(df2$thinking == 1, "그렇다", "아니다")
```

```
qplot(df2$thinking)
```

```
df2$thinking2 <- ifelse(df2$thinking2 == 1, "주거안정",
```

```
ifelse(df2$thinking2 == 2, "자산증식", "기타"))
```

```
think <- df2 %>%
```

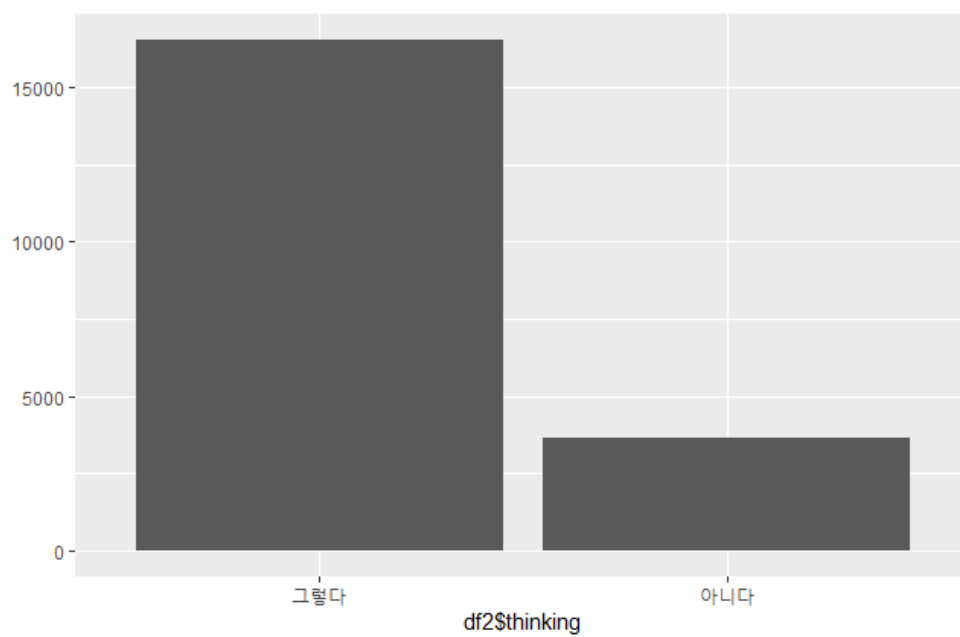
```
  group_by(thinking, thinking2) %>%
```

```
  summarise(n = n())%>%
```

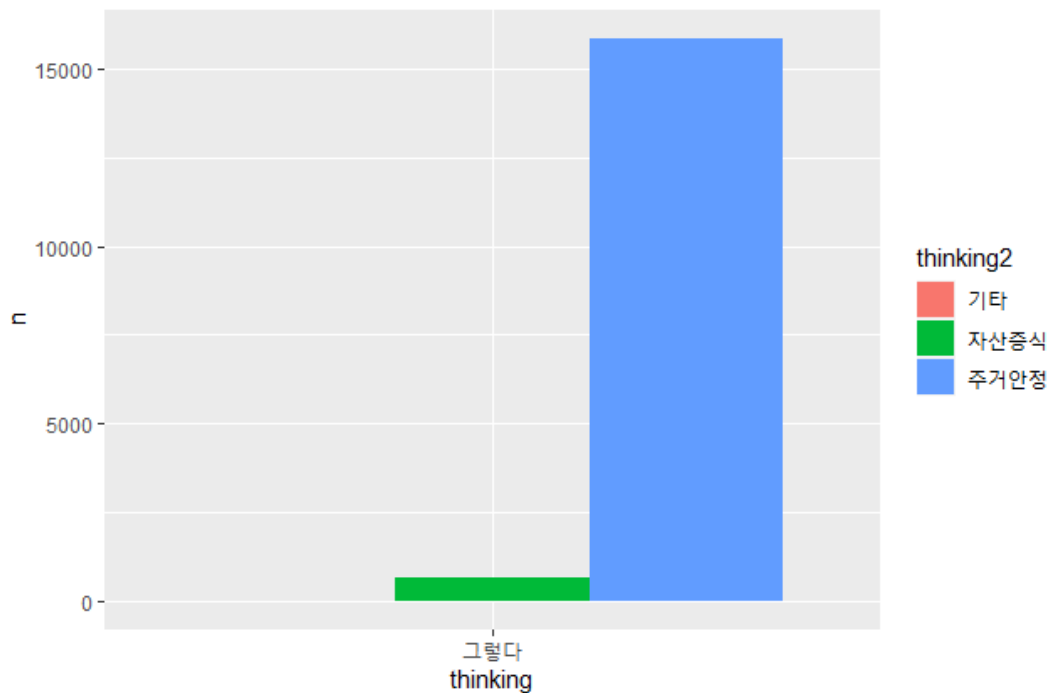
```
  na.omit()
```

```
ggplot(data = think, aes(thinking, y = n, fill = thinking2)) + geom_col(position = 'dodge')
```

#### 자료분석 결과물과 해석



“귀 닥은 내 집을 꼭 마련해야 한다고 생각하십니까?”의 항목에서 과반수가 그렇다고 표기했다.



그 이유로는 주거안정이 가장 많고 자산증식을 목적으로 내 집을 마련해야 한다고 생각하는 사람들도 있는 것으로 보인다.

## 9. 소득 별 주거 위치는 어떨까?

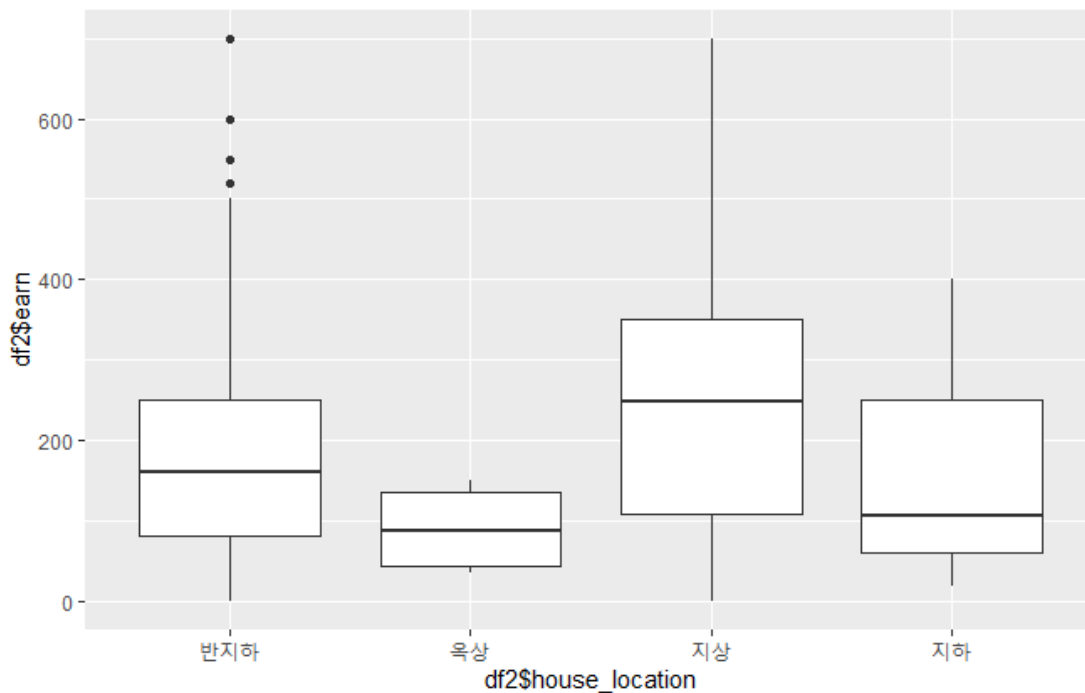
### R 소스코드

```
df2$house_location <- ifelse(df2$house_location == 1, "지상", ifelse(df2$house_location == 2, "반지하", ifelse(df2$house_location == 3, "지하", "옥상")))
```

```
df2$earn <- ifelse(df2$earn == 9999999, NA, df2$earn)
```

```
ggplot(data = df2, aes(x = df2$house_location, y = df2$earn)) + geom_boxplot() + ylim(0, 700)
```

### 자료분석 결과물과 해석



주거형태별 소득구간을 박스 플롯으로 표현해 보면 다음과 같다. 주거 위치가 지상에 있는 사람들의 소득이 가장 높게 분포하고, 옥상에 거주하고 있는 사람들이 제일 낮은 것으로 보인다.

## 10. 성별, 나이 구간별 평균 소득은 어떨까?

### R 소스코드

```
real_age <- 2014 - df2$age + 1
```

```
sex_age = df2 %>%
```

```
  filter(!is.na(earn)) %>%
```

```
  group_by(real_age, master) %>%
```

```
  summarise(mean_income = mean(earn))
```

```
ggplot(data = sex_age, aes(x = real_age, y = mean_income, col = master)) + geom_line()
```

```
sex_income <- df2 %>%
```

```
  group_by(ageg, master) %>%
```

```
summarise(mean_income = mean(earn, na.rm = T))
```

```
ggplot(data = sex_income, aes(x = age,
```

```
    y = mean_income,
```

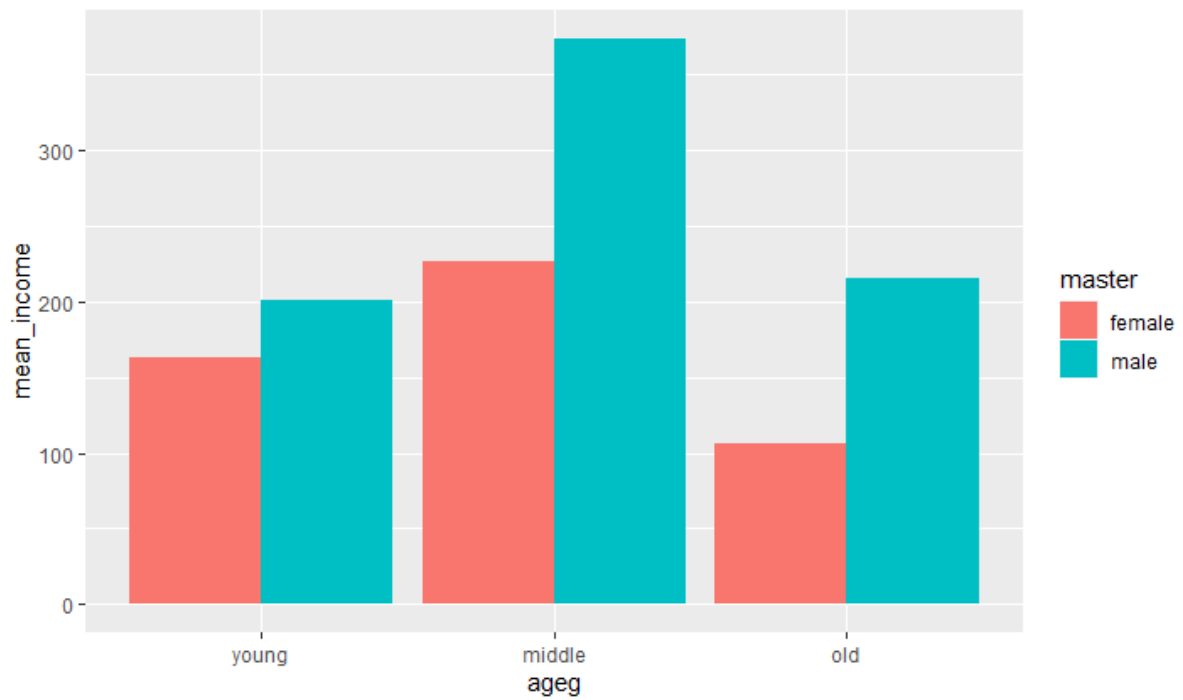
```
    fill = master)) +
```

```
geom_col(position='dodge')+scale_x_discrete(limits = c('young', 'middle', 'old'))
```

## 자료분석 결과물과 해석







남녀 평균 소득에 대해 출력된 그래프를 보면 남녀 모두 20 대에는 월급이 지속적으로 증가하다가, 여성은 30 대부터, 남성은 50 대 후반부터 하강 그래프를 그린다. 두번째 막대 그래프에서는 모든 연령층에서 남성이 여성보다 평균 소득이 높은 것으로 보이고, 여성은 중년, 초년, 노년순으로 평균소득이 감소하고 남성은 중년, 노년, 초년 순으로 평균 소득이 감소하는 것을 알 수 있다.