

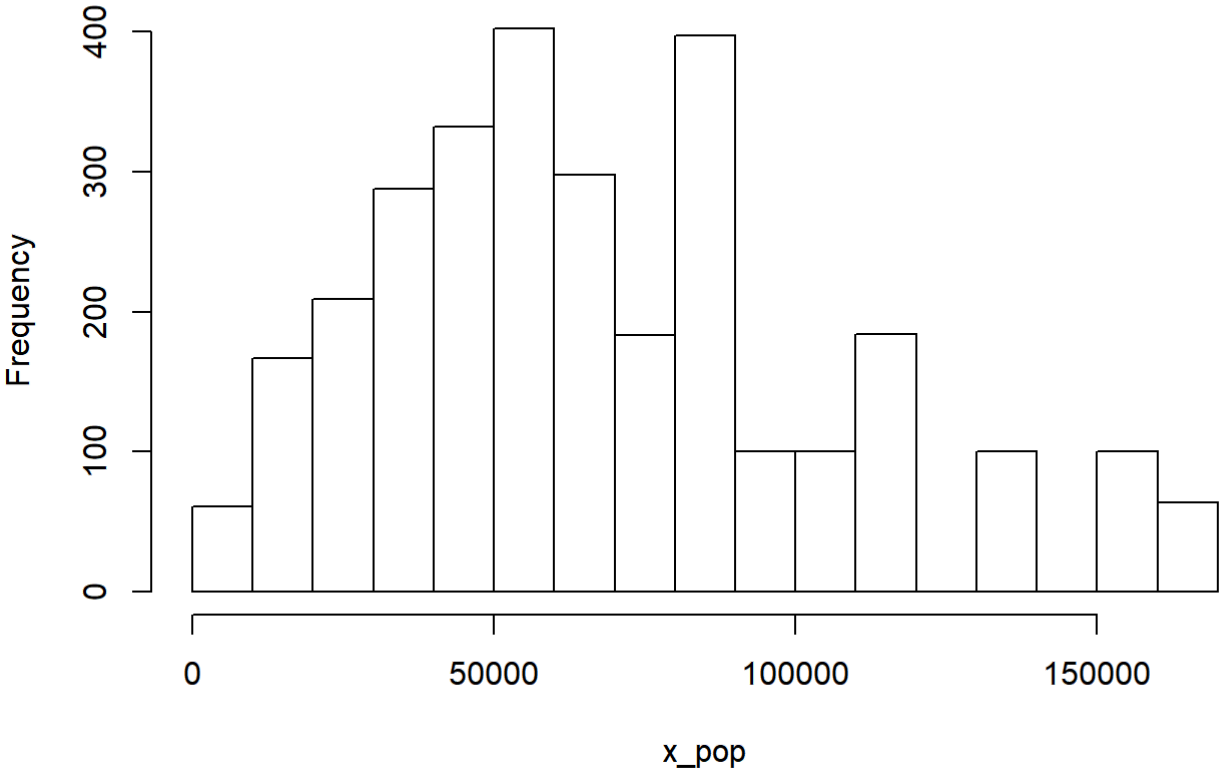
```
### 인천아파트 매매평균 vs. 3년 평균 인구수 EDA ###
## set working directory
setwd('C:/Users/wndy4/Desktop/Project_DEMA')
## csv file load
dat = read.csv('정동호/단지별변수(수정).csv',header=T,stringsAsFactors=F)
View(dat)
str(dat)
```

```
## 'data.frame':   2985 obs. of  19 variables:
## $ 구시군      : chr  "강화군" "강화군" "강화군" "강화군" ...
## $ 읍면동      : chr  "강화읍" "강화읍" "강화읍" "선원면" ...
## $ 아파트명    : chr  "BAYHILL" "BAYHILL" "BAYHILL" "강화1차세광엔리치빌" ...
## $ 분양면적    : num  105.9 95.1 79.8 109.1 86.7 ...
## $ 총세대수    : int   179 179 179 576 497 497 497 497 98 98 ...
## $ 세대수      : int    24 60 95 576 262 120 89 26 45 33 ...
## $ 매매평균..당가.만원.: int   182 173 169 138 173 168 167 164 158 142 ...
## $ X3.3.당가.만원. : num   601 573 559 455 572 ...
## $ 노후도      : num    14 14 14 17.1 14.1 ...
## $ 역세권점수   : int    0 0 0 0 0 0 0 0 0 0 ...
## $ 전세평균..당가.만원.: int    90 88 84 87 114 112 110 110 85 80 ...
## $ 토지면적.km2. : num    25 25 25 21.4 21.4 ...
## $ 문화시설수   : int    9 9 9 2 2 2 2 2 9 9 ...
## $ 쇼핑시설수   : int   27 27 27 4 4 4 4 4 27 27 ...
## $ 스타벅스      : int    0 0 0 0 0 0 0 0 0 0 ...
## $ X3년.평균.인구수 : int  22882 22882 22882 7667 7667 7667 7667 7667 22882 22882 ...
## $ X3년.평균.세대수 : int   9874 9874 9874 3219 3219 3219 3219 3219 9874 9874 ...
## $ 초중고현황    : int    9 9 9 1 1 1 1 1 9 9 ...
## $ 개발호재유무  : logi   NA NA NA NA NA NA ...
```

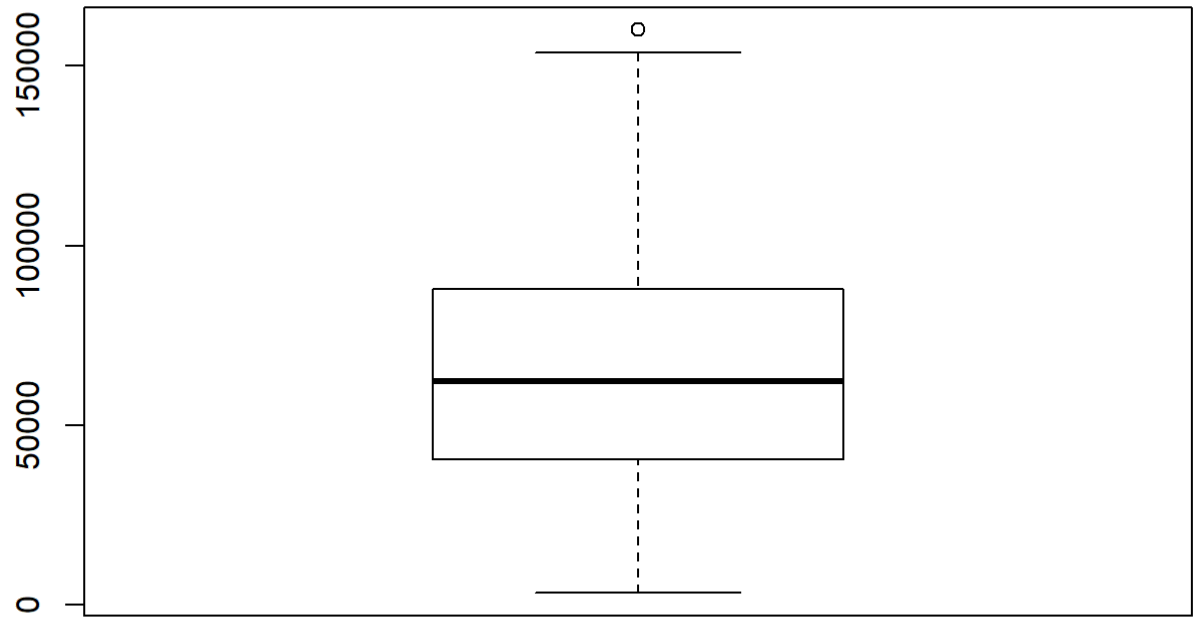
```
y = dat$매매평균..당가.만원.
x_pop = dat$X3년.평균.인구수
```

```
## 3년 평균 인구수 히스토그램
windows()
hist(x_pop,main = "인천 아파트에 대한 3년 평균 인구수 히스토그램")
```

인천 아파트에 대한 3년 평균 인구수 히스토그램

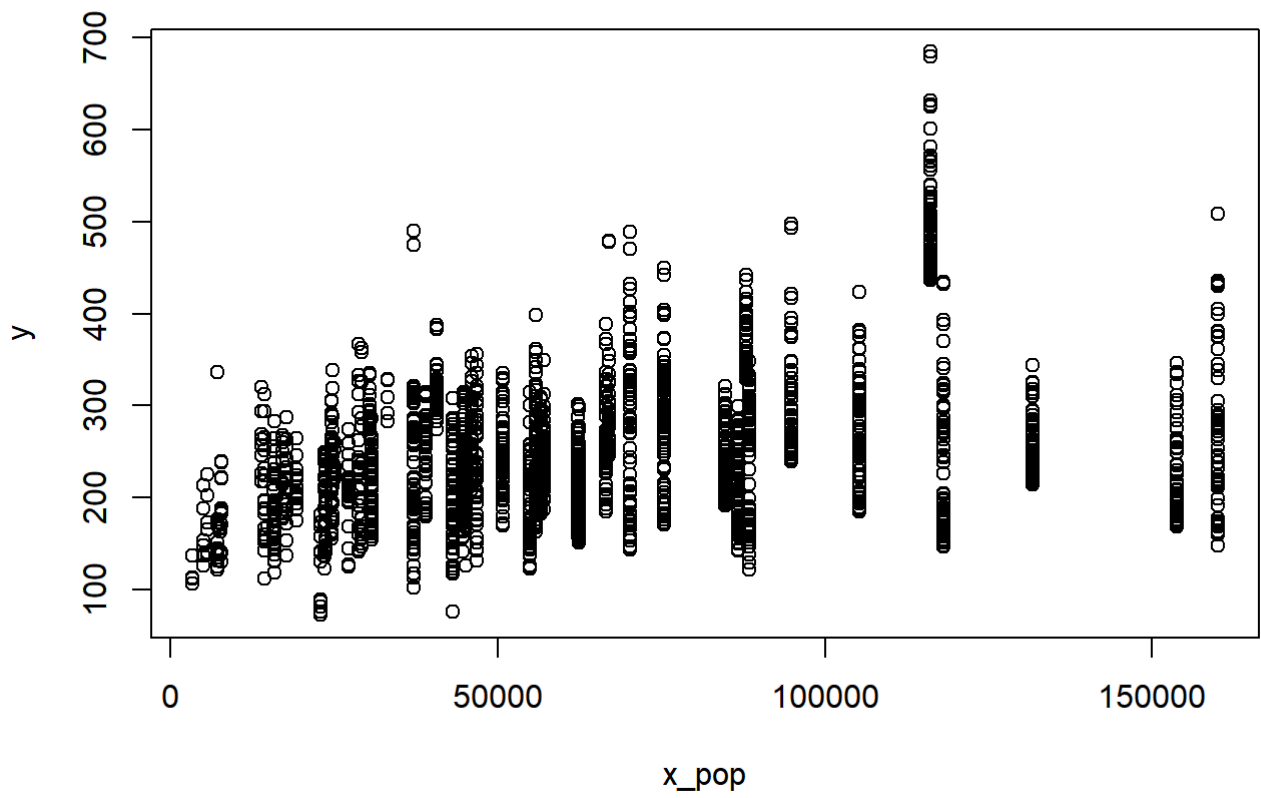


```
# 단봉보다는 쌍봉의 형태를 보이는 것 같다
# 왼쪽으로 쏠려있는 모양이다.
windows()
boxplot(x_pop)
```



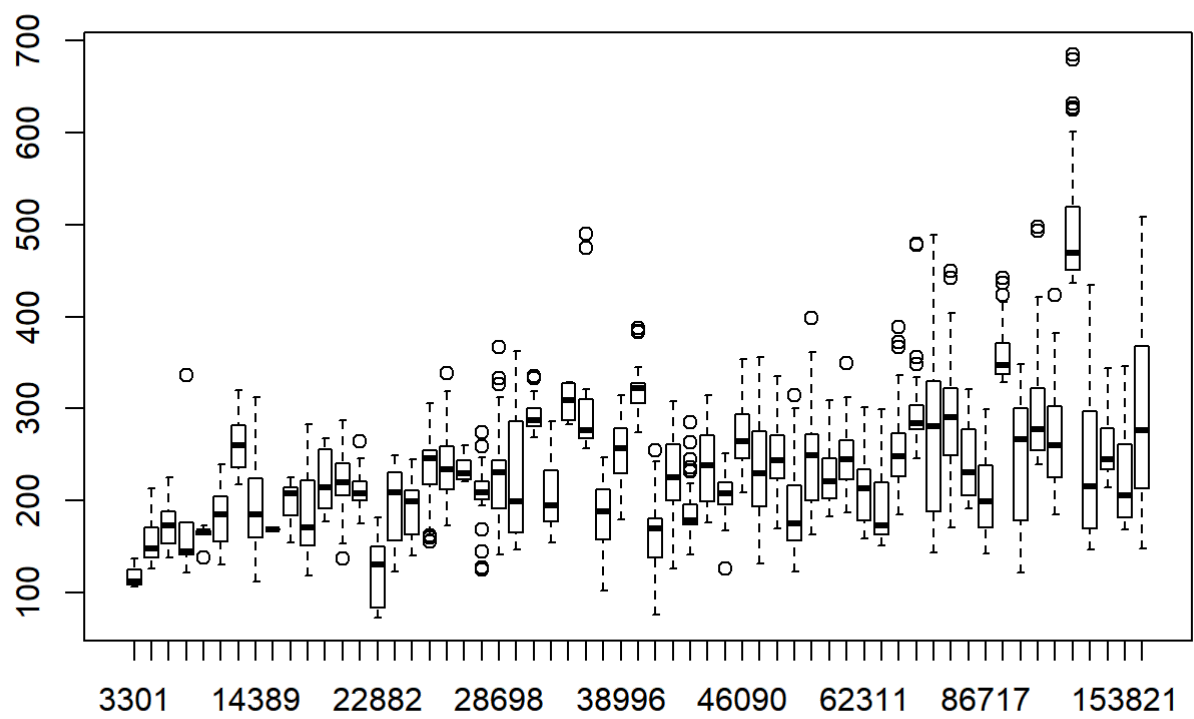
```
## 산점도  
windows()  
plot(x_pop,y,main='인천아파트 매매평균 vs. 3년 평균 인구수')
```

인천아파트 매매평균 vs. 3년 평균 인구수



```
## boxplot  
windows()  
boxplot(y ~ x_pop,main = '인천아파트 매매평균 vs. 3년 평균 인구수')
```

인천아파트 매매평균 vs. 3년 평균 인구수



```
## spearman correlation
cor.test(x = x_pop, y = y, method="spearman")
```

```
## Warning in cor.test.default(x = x_pop, y = y, method = "spearman"): Cannot
## compute exact p-value with ties
```

```
##
## Spearman's rank correlation rho
##
## data: x_pop and y
## S = 3012800000, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## 0.3203403
```

```
# rho가 0.32이기에 어느정도 양의 상관관계를 보인다고 할 수 있다.
## 상관관계
cor(x_pop,y)
```

```
## [1] 0.3320709
```

NA