

01_TEST

- Guide
 - 1. Data Wrangling Problem
 - 2. Visualization Problem(Using ggplot2)
-

- 필요 패키지 : dplyr , ggplot2
 - 필요 데이터 : house_price.csv
-

Guide

- 시험시간 : 13:30 ~ 15:00 (총90분)
- R markdown을 이용하여 생성한 html파일 모두 제출
- Dropbox에 제출시 html 파일과 rmd 파일 모두 업로드 해야함
 - 공유된 폴더안에 시험날짜 폴더를 만들고 (예 : 202101234)
 - 관련 데이터는 Data폴더 를 만들어 제출
 - html,rmd파일은 Rmd폴더 에 학번_이름으로 작성 제출 (예 : 202101234_홍길동)
- 문제에서 주어진 결과와 같게 작성
 - ex) 작성방식 : 1.1번 문제라고 적고 R Chunk를 생성하여 그 안에 해당 code 기입
- 업로드 및 수정시간이 모두 기록 됨에 유의
- 코드공유 적발시 0점 처리
- 문제에서 출력한대로 출력(전체데이터 출력은 지양)
- 위의 유의사항을 어길 시 감점
- 시험시간을 넘겨서 제출하면 0점 처리



0. Packages

```
library(dplyr)
library(ggplot2)
```

1. Data Wrangling Problem

1.1 house_price.csv 를 불러오시오.[5점]

- head()로 출력할 것

```
price <- read.csv(file = "../Data/house_price.csv")
head(price)
```

##	자치구명	법정동명	신고년도	건물면적	건물주용도	물건금액	건물명
## 1	성북구	장위동	2020	59.92	아파트	7.49e+08	래미안 장위포레카운티
## 2	성북구	길음동	2020	84.77	아파트	1.20e+09	롯데캐슬 클라시아
## 3	구로구	구로동	2020	36.90	아파트	1.05e+08	궁전아트 빌라
## 4	용산구	산천동	2020	59.55	아파트	8.30e+08	리버힐
## 5	용산구	이촌동	2020	64.43	아파트	7.35e+08	삼성강변
## 6	용산구	이태원동	2020	105.75	아파트	1.43e+09	청화1

1.2 데이터의 변수명을 확인하시오[5점]

```
colnames(price)
```

```
## [1] "자치구명"    "법정동명"    "신고년도"    "건물면적"    "건물주용도"
## [6] "물건금액"    "건물명"
```

1.3 데이터의 속성을 확인하시오[5점]

```
str(price)
```

```
## 'data.frame':    176001 obs. of  7 variables:
## $ 자치구명      : chr   "성북구" "성북구" "구로구" "용산구" ...
## $ 법정동명      : chr   "장위동" "길음동" "구로동" "산천동" ...
## $ 신고년도      : int    2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 .
##
## $ 건물면적      : num    59.9 84.8 36.9 59.5 64.4 ...
## $ 건물주용도     : chr   "아파트" "아파트" "아파트" "아파트" ...
## $ 물건금액      : num    7.49e+08 1.20e+09 1.05e+08 8.30e+08 7.35e+08 1.43e+
## 09 4.00e+08 8.75e+08 3.13e+08 7.40e+08 ...
## $ 건물명        : chr   "래미안 장위포레카운티" "롯데캐슬 클라시아 " "궁전아트빌라" "리
## 버힐삼성" ...
```

1.4 다음과 같이 일부 column만 선택하여 price2를 만들고 출력하시오[5점]

- head()로 출력할 것

```
price2 <- price %>% select(c("자치구명", "법정동명", "건물면적", "건물주용도", "물건금액"))
head(price2)
```

```
##   자치구명 법정동명 건물면적 건물주용도 물건금액
## 1   성북구   장위동    59.92   아파트 7.49e+08
## 2   성북구   길음동    84.77   아파트 1.20e+09
## 3   구로구   구로동    36.90   아파트 1.05e+08
## 4   용산구   산천동    59.55   아파트 8.30e+08
## 5   용산구   이촌동    64.43   아파트 7.35e+08
## 6   용산구   이태원동  105.75   아파트 1.43e+09
```

1.5 건물용도 별 count를 구하시오[10점]

- 정렬하여 나타낼 것[5점]

```
price2 %>% group_by(건물주용도) %>% summarise(cnt=n()) %>% arrange(desc(cnt))
```

```
## # A tibble: 4 x 2
##   건물주용도    cnt
##   <chr>      <int>
## 1 아파트      88553
## 2 연립주택     60351
## 3 오피스텔     14934
## 4 단독주택     12163
```

1.6 지역별 아파트 가격 평균을 구하시오 [25점]

- 조건 필터링, 그룹화, 통계량 등 [5점]
- 정렬하여 나타낼 것 [5점]
- “평균집값”이라는 이름으로 열을 생성해서 00억으로 나타낼 것 [10점]
 - 힌트: round()와 paste()를 활용할 것
- 상위 5개 자치구만 뽑아낼 것 [5점]

```
price2 %>% filter(건물주용도=="아파트") %>% group_by(자치구명) %>% summarise(m
ean_price=mean(물건금액)) %>% arrange(desc(mean_price)) %>% mutate(평균집값
= paste(round(mean_price/100000000, 2), "억", sep="")) %>% head(5)
```

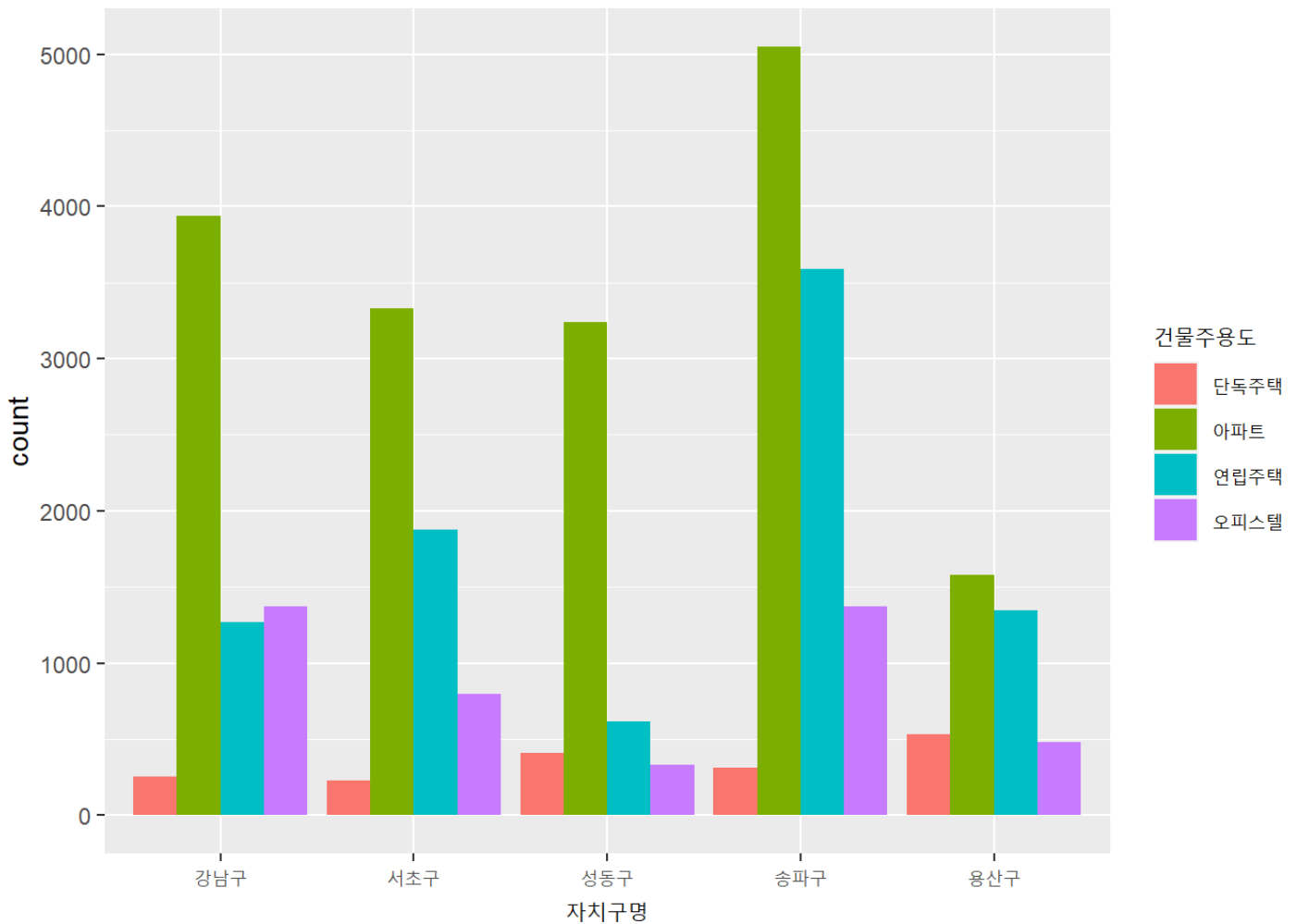
```
## # A tibble: 5 x 3
##   자치구명 mean_price 평균집값
##   <chr>      <dbl> <chr>
## 1 강남구    1802629827. 18.03억
## 2 서초구    1719564776. 17.2억
## 3 용산구    1495991173. 14.96억
## 4 송파구    1290320644. 12.9억
## 5 성동구    1112411834. 11.12억
```

2. Visualization Problem(Using ggplot2)

2.1 지역별 건물을 barplot을 활용해 그리시오 [20점]

- 위에서 구한 상위 5개 자치구만 활용할 것
 - 상위 5개 자치구(=강남구, 서초구, 용산구, 송파구, 성동구)
- 조건 필터링 [5점]
- 기본 그림 [15점]

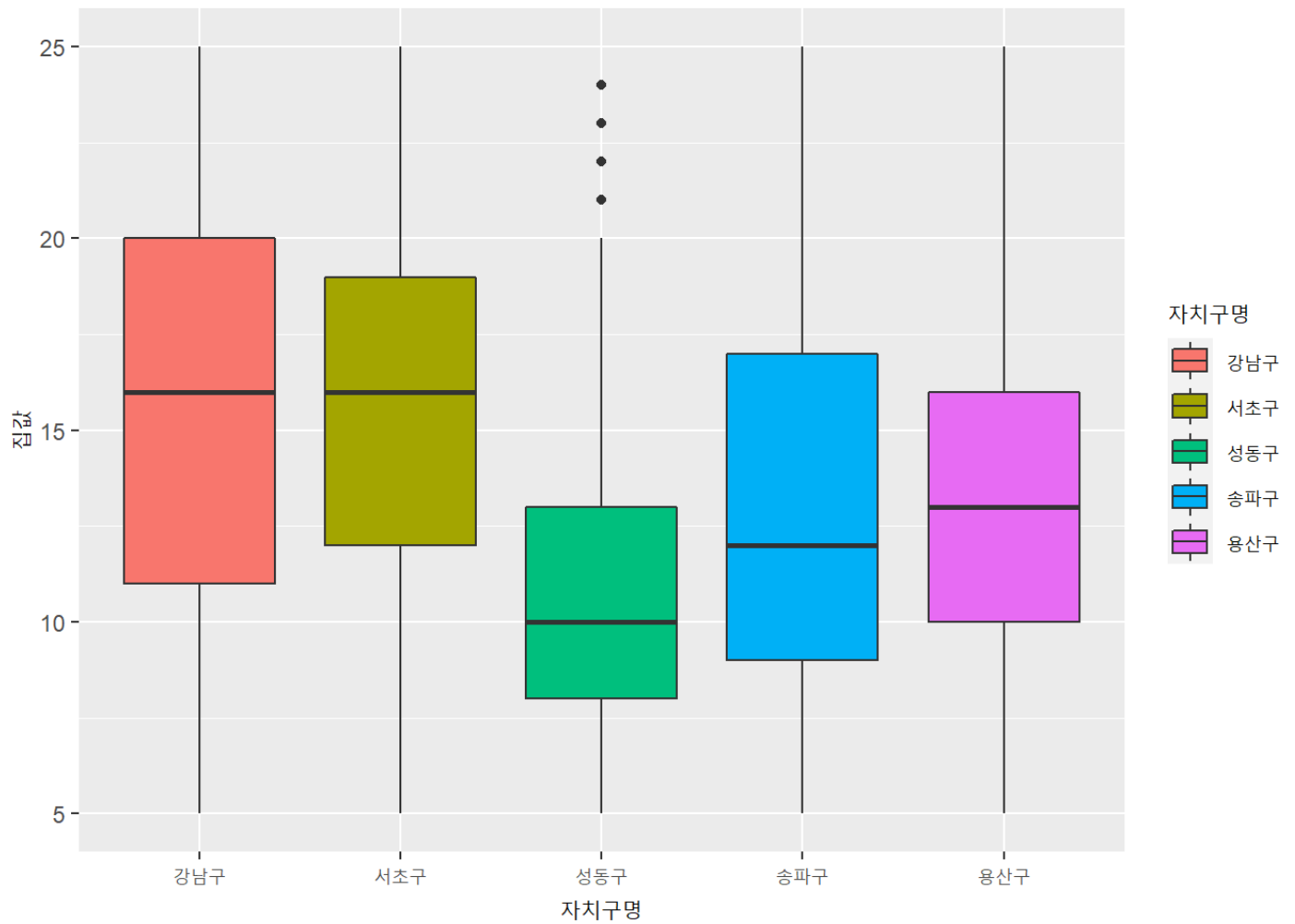
```
price2 %>% filter(자치구명 %in% c("강남구", "서초구", "용산구", "송파구", "성동구")) %>%
ggplot(aes(자치구명, fill=건물주용도)) + geom_bar(position="dodge")
```



2.2 지역별 아파트 가격을 boxplot을 활용해 그리시오 [25점]

- 위에서 구한 상위 5개 자치구만 활용할 것
 - 상위 5개 자치구(=강남구, 서초구, 용산구, 송파구, 성동구)
- 기본 그림 [15점]
- 가격은 억단위로 바뀌서 “집값”이라는 열을 생성할 것 [5점]
 - 힌트: round()
- y축 범위는 5에서 25로 할 것 [5점]

```
price2 %>% filter(자치구명 %in% c("강남구", "서초구", "용산구", "송파구", "성동구"))
%>% filter(건물주용도 %in% "아파트") %>% mutate(집값 = round(물건금액/1000000000
),2) %>% ggplot(aes(자치구명, 집값, fill=자치구명)) + geom_boxplot() + scale
_y_continuous(limits = c(5, 25))
```



© 2021 Advice, All Rights Reserved

No part of this contents may be reproduced, copied, modified or adapted, without the prior written consent of the author, unless otherwise indicated for stand-alone materials.