

Exercise Problem Set 03

Basic R Skills

"**Car04.csv**" file contains information of various car models. The dataset records characteristics on all of the new models of cars for sale in the US in a certain year.

(Korean Translation) "**Car04.csv**"파일은 미국에서 특정 연도에 판매된 여러 자동차 모델에 대한 정보를 담고 있습니다.

1. Load the data in the file into R as a data.frame named "**car_df**"

(Korean Translation) 파일로부터 데이터를 R로 읽어오시오. 데이터 프레임 이름은 **car_df**

2. How many variables are in the dataset? And how many observations are there?

(Korean Translation) 변수는 몇 개며 observation은 몇 개 입니까?

3. The "**name**" column represents the name of car models. What would be the proper data type of "**name**", character or factor? Convert the type into proper one if necessary.

(Korean Translation) name columns은 자동차 모델의 이름을 담고 있습니다. character와 factor type 중 어떤 것이 적절할까요? 필요하다면 type 변환을 하시오.

4. "**msrp**" means Manufacturer Suggested Retail Price. Calculate average difference between **msrp** and **dealer_cost**.

(Korean Translation) msrp는 소비자 권장 가격입니다. dealer_cost와 비교해서 평균적으로 차이가 얼마나 나나요?

5. Find the car models with highest **city_mpg**. What is the difference of **city_mpg** and **hwy_mpg** for the car? Is it the same car with the car of highest **hwy_mpg**?

(Korean Translation) city_mpg값이 가장 큰 자동차 모델은 무엇인가요? 그 차의 city_mpg와 hwy_mpg의 차이는 얼마인가요? 이 차는 hwy_mpg가 가장 높은 차와 같은 차종인가요?

6. How many cars models are in each car type of (**sport car, suv, wagon, minivan, pickup**)? How many cars are in none of these car types?

(Korean Translation) 각 각의 자동차 종류(**sport car, suv, wagon, minivan, pickup**)마다 자동차 모델이 몇 개씩 있나요? 어떤 종류에도 속하지 않는 자동차는 몇 개나 있나요?

7. Compare the weight of **suv** and **minivan**. Which car type is heavier on average?

(Korean Translation) SUV와 minivan 차종의 무게를 비교하시오. 평균적으로 어떤 차종이 더 무겁나요?

8. Add new column "**avg_mpg**" which represent average mpg value of **city_mpg** and **hwy_mpg**.

(Korean Translation) 새로운 column "avg_mpg"를 추가하시오. avg_mpg는 city_mpg와 hwy_mpg의 평균입니다.

9. Add new column of "**eco_grade**" that has value of "**good**" for cars with high 20% avg_mpg, "**bad**" for ones with low 20% mpg, and "**normal**" for the rest.

(Korean Translation) "eco_grade"라는 columns을 추가하시오. avg_mpg 상위 20%에는 "good" 하위 20%에는 "bad", 나머지에는 "normal" 값을 부여하시오.

10. Compare horse power of all wheel drive cars and rear wheel drive cars.

(Korean Translation) 4륜 구동 자동차와 후륜구동 자동차의 마력을 비교하시오.

Data Transformation

```
> data(anscombe)
> anscombe
  x1 x2 x3 x4 y1 y2 y3 y4
1 10 10 10 8 8.04 9.14 7.46 6.58
2  8  8  8 8 6.95 8.14 6.77 5.76
3 13 13 13 8 7.58 8.74 12.74 7.71
4  9  9  9 8 8.81 8.77 7.11 8.84
5 11 11 11 8 8.33 9.26 7.81 8.47
6 14 14 14 8 9.96 8.10 8.84 7.04
7  6  6  6 8 7.24 6.13 6.08 5.25
8  4  4  4 19 4.26 3.10 5.39 12.50
9 12 12 12 8 10.84 9.13 8.15 5.56
10 7  7  7 8 4.82 7.26 6.42 7.91
11 5  5  5 8 5.68 4.74 5.73 6.89
```

"**anscombe**" is a data.frame combined with 4 different datasets of "x" and "y" variable. Answer the following questions.

"**anscombe**" data.frame은 4개의 서로 다른 데이터 셋을 하나로 묶어 놓은 데이터 프레임이다. 각 데이터 셋은 X와 Y를 변수로 가진다. 아래에 답하시오. (x1, y1) (x2, y2) (x3, y3) (x4, y4)

1. Calculate average and standard deviation of 4 different sets of "x" and "y". What can you

infer from the result?

4개의 data set에서 각각의 x 변수와 y변수의 평균과 분산을 계산하여 비교해 보시오. 이를 통해서 알 수 있는 것은 무엇입니까?

2. Visualize 4 different sets of 'x' and 'y' on a scatter plot with **plot** function. (`plot(x = variableX, y = variableY)`). What can you infer from the visualization?

4개의 data set을 각각 시각화 해보시오. x축에는 x변수 y축에는 y변수를 표현한 산점도 그래프로 나타내시오. 총 4개의 그래프가 그려지면 4개의 그래프를 그려보고 알 수 있는 점을 설명해보시오.

3. Transform **anscombe** data frame into following format. You need to have 44 rows in the resulting data frame.

anscombe data frame을 아래와 같은 형태로 변환해보시오. (참고) 결과 data frame의 행의 수는 44개 입니다.

```
> anscombe.tidy
  key.x value.x key.y value.y
1     x1      10    y1    8.04
2     x1       8    y1    6.95
3     x1      13    y1    7.58
4     x1       9    y1    8.81
5     x1      11    y1    8.33
6     x1      14    y1    9.96
7     x1       6    y1    7.24
8     x1       4    y1    4.26
9     x1      12    y1   10.84
10    x1       7    y1    4.82
11    x1       5    y1    5.68
12    x2      10    y2    9.14
13    x2       8    y2    8.14
14    x2      13    y2    8.74
15    x2       9    y2    8.77
16    x2      11    y2    9.26
17    x2      14    y2    8.10
18    x2       6    y2    6.13
19    x2       4    y2    3.10
20    x2      12    y2    9.13
```

...