

NYC Data Science Bootcamp

Manipulating data with “dplyr”

* Save all your code to *yourname.R* and push it to the homework Github repository.

Note: you may need to **pull** from origin before you **push** it to Github.

Question #1:

For the first two questions, we study the data mpg in the ggplot2 package.

- (a) Load ggplot2 package first and then type `data(mpg)`. Quickly go through the dataset and the help file.
- (b) Obtain a subset of data including: year, cyl, cty, hwy, and renames these variables as V1, V2, V3, V4.
- (c) In mpg data, obtain the average of *city miles per gallon* and *highway miles per gallon* for different numbers of cylinders. (Hint: the mean function calculates the average of a vector.)
- (d) For manufacturer, identify the car(s) that have the largest *city miles per gallon*.

Question #2:

We want to know the relationship between three variables: *engine displacement*, *city miles per gallon*, and *highway miles per gallon*.

1. Create a new variable ratioHVE showing the ratio between *highway miles per gallon* and *engine displacement*.
2. Create new variables rationCVE showing the ratio between *city miles per gallon* and *engine displacement*.
3. Obtain the average ratioHVE and ratioCVE by different years and manufacturers.
4. Find the biggest ratioHVE by different years and drv.

Question #3:

For this question, you are going to explore an online dataset and try to answer the questions. You can find the dataset here:

<https://data.cityofnewyork.us/City-Government/NYC-Jobs/kpav-sd4t>

1. What are the mean and median beginning and ending salaries for each agency?
Note that salaries can be annual, hourly, or daily. You need to convert all of them to annual.
2. Which agency has the highest average starting salary?
3. Does the type of posting (internal or external) have a big impact on the average salary range?
4. Rank the levels according to the average salary range for each level.
5. Find out the range of money currently being spent by each agency for new hires.
Note that the X.Of.Positions column shows how many positions are available.
6. What civil service title has the largest salary range?