DA5020.A1.Hsiao-Yu.Peng

Hsiao-Yu Peng

2023-09-09

Q1-1.

Please see the attached certificate.

Q1-2.

Data frame stores data in a two-dimensional table of columns (variables) and rows (observation). We use it to solve a specific theme. Here is an example:

```
# Data frame example
df <- data.frame(
  patientID = c("U101", "U102", "U103", "U104"),
  gender = c("M", "F", "F", "M"),
  age = c(50, 66, 35, 56),
  grade = c("high", "low", "low", "medium")
  )

df
```

```
##   patientID gender age  grade
## 1      U101      M  50   high
## 2      U102      F  66    low
## 3      U103      F  35    low
## 4      U104      M  56 medium
```

Q2.

```
cars <- c("Truck", "Car", "SUV")
mpg <- c(11, 30, 24)
cost <- c(45000, 25000, 35000)
DF <- data.frame(cars, mpg, cost)
print(DF)
```

```
##    cars mpg  cost
## 1 Truck  11 45000
## 2   Car  30 25000
## 3   SUV  24 35000
```

Q2a.

```
# Select row 1 in column 3, what was selected?
DF[1, 3]
```

```
## [1] 45000
```

```
# Select rows 1 through 3, what was selected?
DF[1:3, ]
```

```
##    cars mpg  cost
## 1 Truck  11 45000
## 2   Car  30 25000
## 3   SUV  24 35000
```

```
# Select the last column, what was selected?
DF[, 3]
```

```
## [1] 45000 25000 35000
```

Q3.

```
# load dataset mtcars
data(mtcars)

# first 3 rows of the dataset
head(mtcars, 3)
```

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

```
# last 5 rows of the dataset
tail(mtcars, 5)
```

```
##                 mpg cyl  disp  hp drat    wt qsec vs am gear carb
## Lotus Europa   30.4   4  95.1 113 3.77 1.513 16.9  1  1    5    2
## Ford Pantera L 15.8   8 351.0 264 4.22 3.170 14.5  0  1    5    4
## Ferrari Dino   19.7   6 145.0 175 3.62 2.770 15.5  0  1    5    6
## Maserati Bora  15.0   8 301.0 335 3.54 3.570 14.6  0  1    5    8
## Volvo 142E     21.4   4 121.0 109 4.11 2.780 18.6  1  1    4    2
```

Q4.

```r
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```r
# Categorical variables
cat_var <- c("vs", "am")
cat("Categorical variables: ", cat_var)
```

```
## Categorical variables:  vs am
```

```r
# Continuous variables
con_var <- c("mpg", "wt")
cat("Continuous variables: ", con_var)
```

```
## Continuous variables:  mpg wt
```

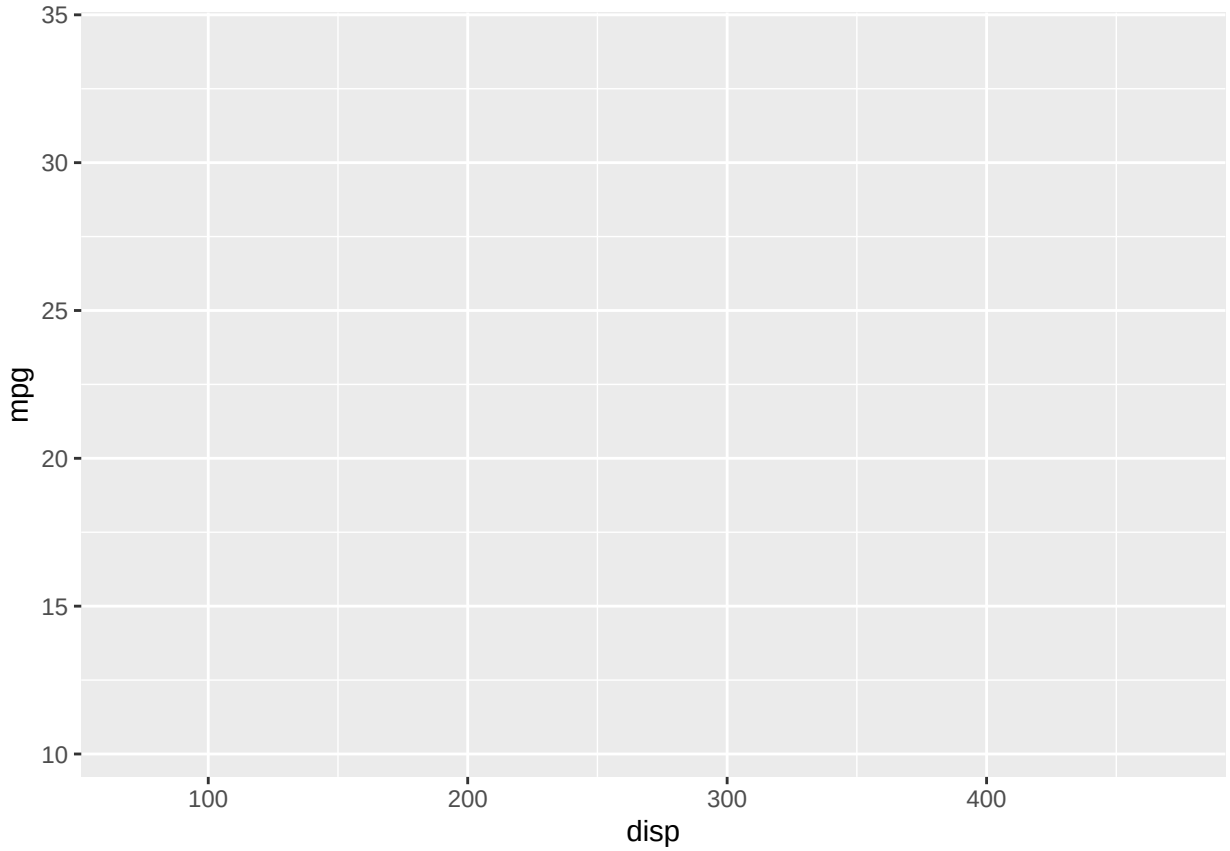
Q5.

```r
library(ggplot2)
```

```
##
## Attaching package: 'ggplot2'
```

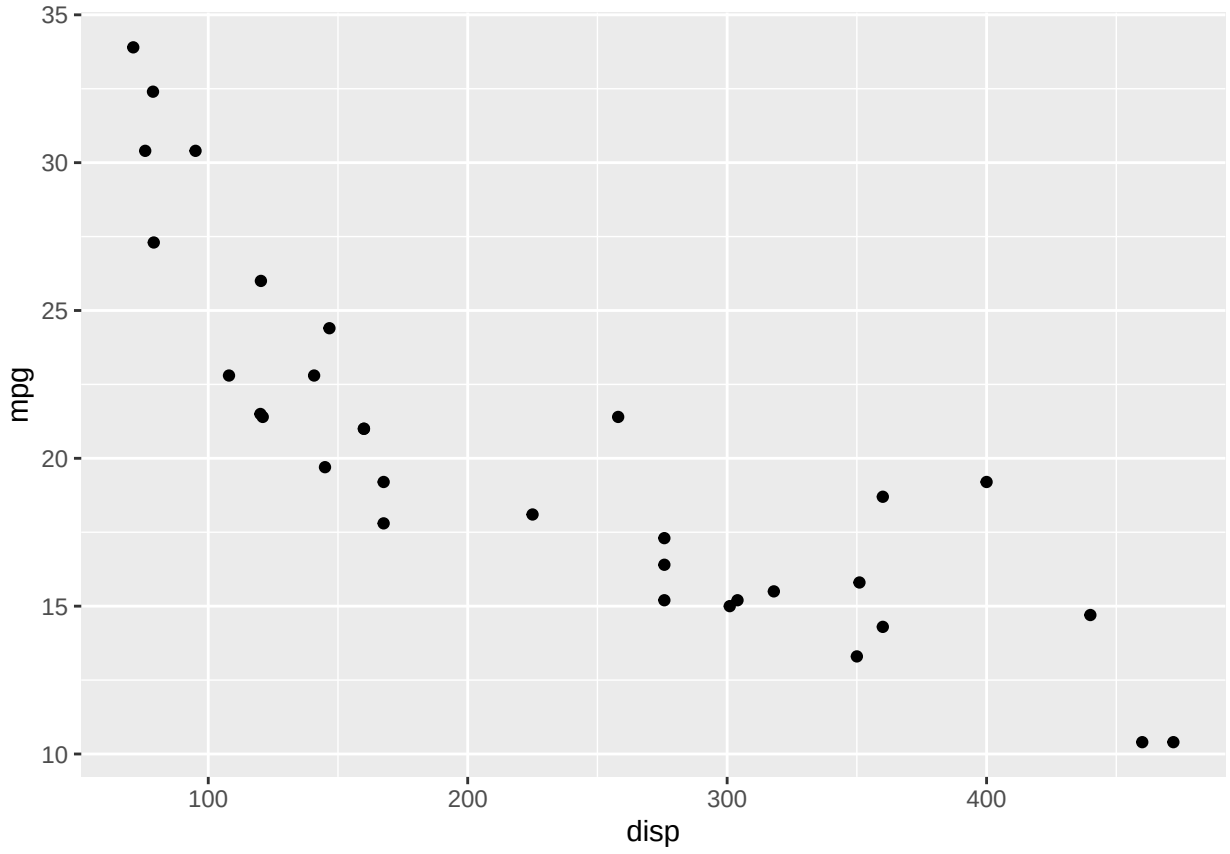
```
## The following object is masked _by_ '.GlobalEnv':
##
##     mpg
```

```r
data("mtcars")
ggplot(mtcars, aes(x = disp, y=mpg))
```



The code above did not actually displayed a chart, but it creates the base for a scatter plot. To show a scatter plot represents the relationship between "disp" and "mpg", we need to add a geom layer. The updated code is as follows:

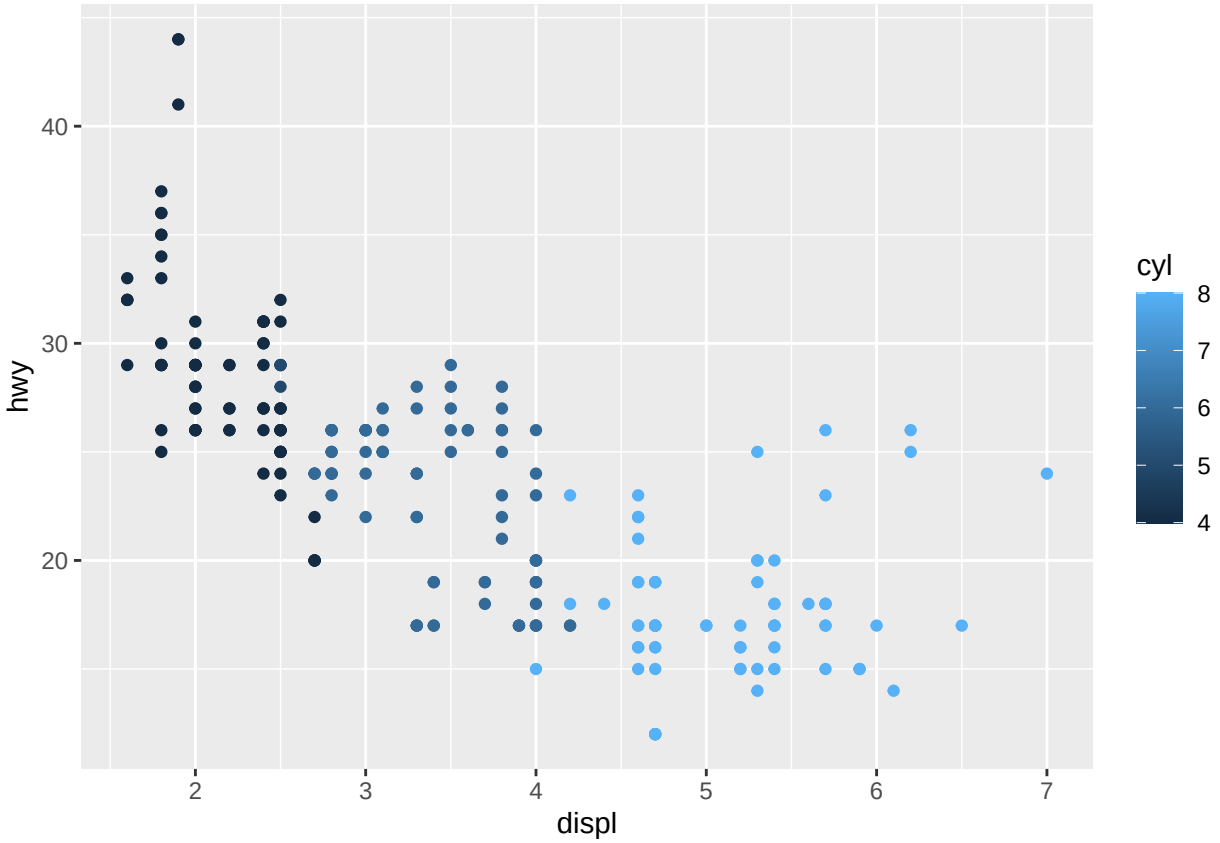
```
ggplot(mtcars, aes(x = disp, y=mpg)) + geom_point()
```



Q6.

```
# load the data frame
data("mpg")

# create a scatter plot: displ vs. hwy
ggplot(data=mpg, aes(displ, hwy, color = cyl)) + geom_point()
```

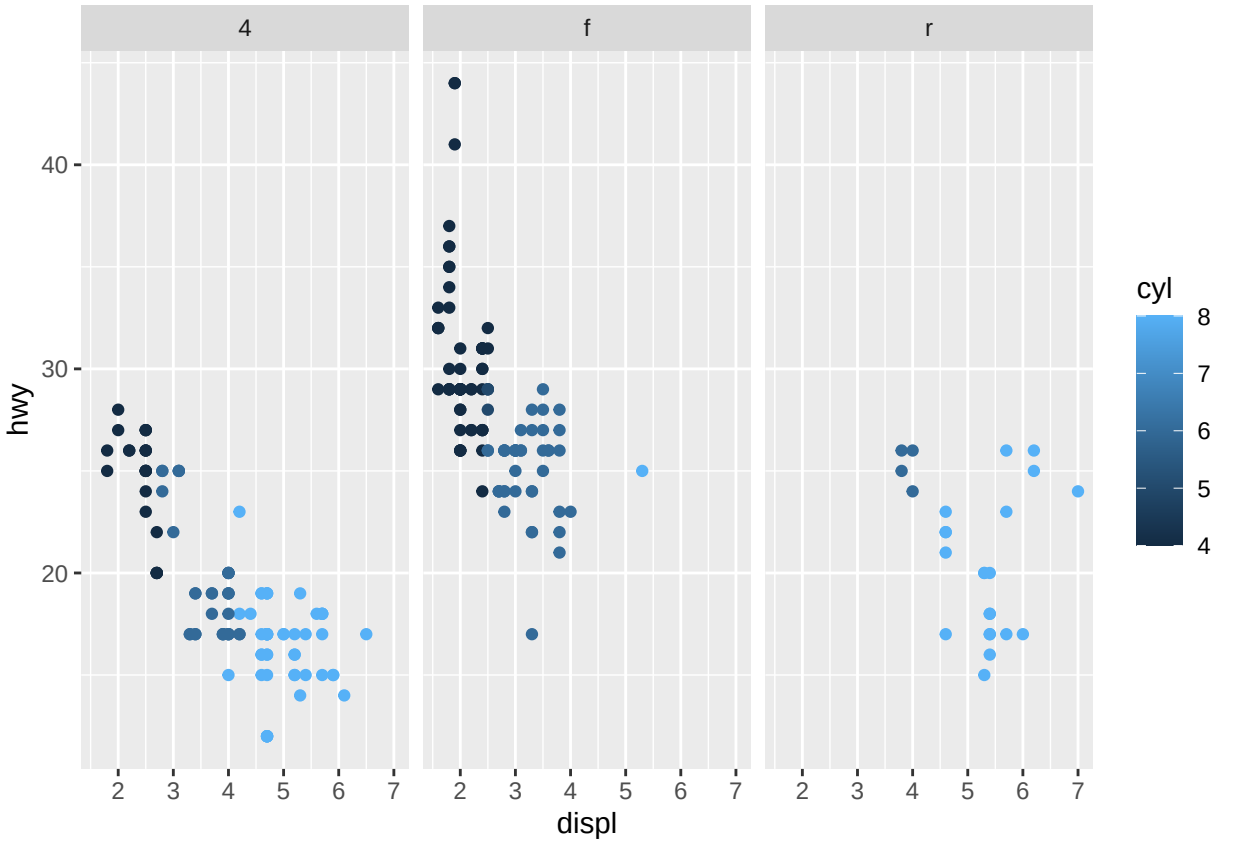


Q6 explanation

The scatter plot illustrate the relationship between engine displacement and highway miles per gallon. Each point's color indicates the number of cylinders in the car. It is evident that fewer engine displacement, for example, an engine displacement of 2, results in higher highway miles per gallon. However, as engine displacement increases, highway miles per gallon decreases. For instance, with an engine displacement of 5 or 6, the car achieves only about 20 miles per gallon or less on the highway.

Q7.

```
ggplot(data = mpg, aes(displ, hwy, color = cyl)) + geom_point() + facet_wrap(vars(drv))
```



Q7 explanation

Compared to 4WD and front-wheel drive (FWD), the rear-wheel drive (RWD) generally features engine displacements from 4 to 7. In contrast, FWD typically has engine displacements ranging from 1 to 4, while 4WD varies across several types, with engine displacements ranging from 1 to 7.

In terms of highway miles per gallon (MPG), RWD vehicles generally achieve a range of 15 to 25 MPG. Conversely, FWD vehicles typically achieve 25 to 30 MPG or even more on the highway. As for 4WD vehicles, they exhibit a wide range of MPG values due to various engine displacements. Generally, the higher the engine displacement, the lower the MPG 4WD achieves.