

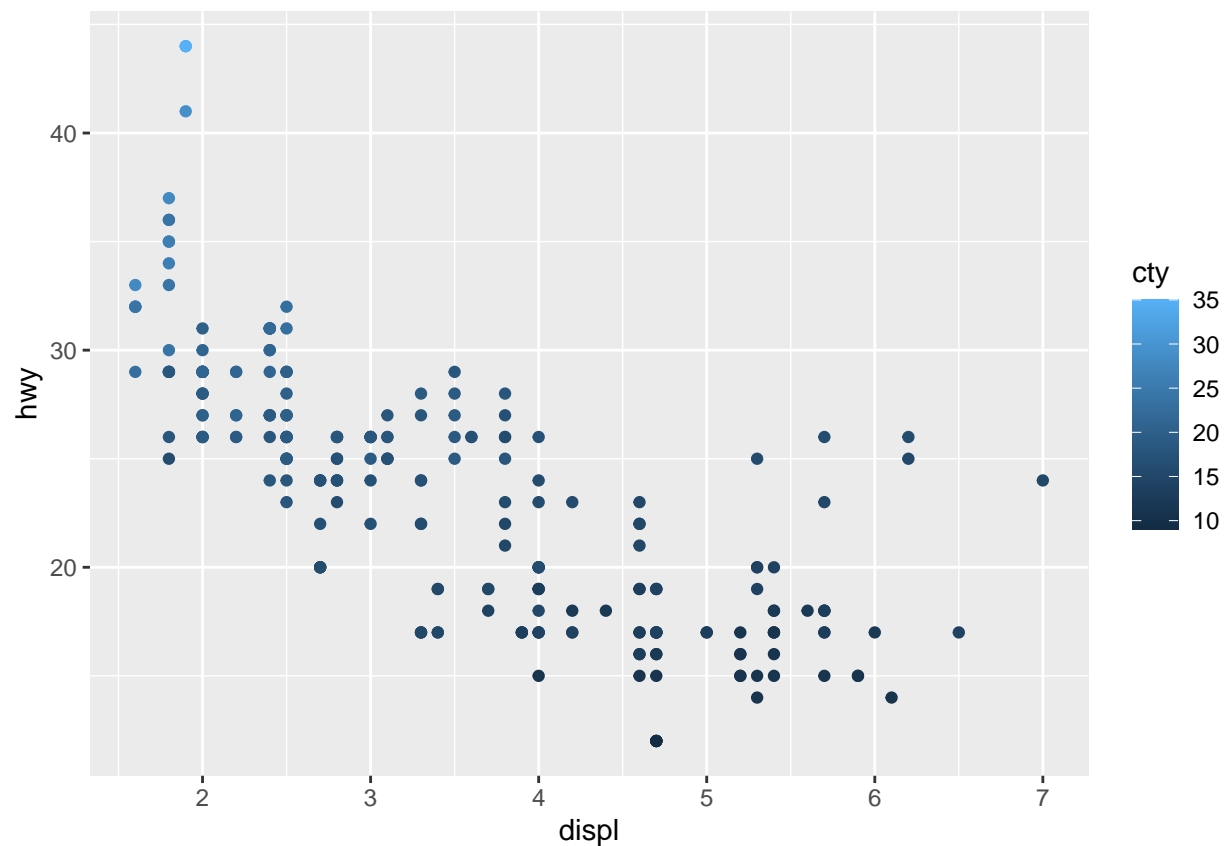
HW1_jaeyounglee

Jaeyoung Lee

August 28, 2020

Problem 1

```
### Problem 1 : Some examples from Primers ###  
### Visualization basics  
library(ggplot2)  
# Map cty to color from mpg data  
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy, color = cty))
```



```
### Programming Basics  
list_ex <- list(number = 100, logical = TRUE, string = "stories")  
list_ex[['number']]
```

```
## [1] 100
```

Problem 2

Part A

What I hope to get out of this class as follows,

1. How to be reproducible to collaborate with others and even myself
2. How to write effective codes
3. How to make nice reports using R Markdown (Be a nice statistician)
4. Getting used to R Markdown and GitHub

I feel lucky to take this class.

Part B

Binomial, Gamma, Beta pdfs from Casella & Berger

$$P(X = x|n, p) = \binom{n}{x} p^x (1-p)^{1-x} \quad (1)$$

$$P(X|\alpha, \beta) = \frac{1}{\gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (2)$$

$$P(X|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (3)$$

Problem 3

With Rule 1, it is useful to make comments for every steps. There is no need to use rule 2 for problem 4, but in the future, it is essential not to use Excel or other programs for data manipulation. From Rule 3, it is essential to write the version of R console whenever we make different version of R code to avoid errors. Using GitHub, we can follow the rule 4. According to rule 5, it is better to record intermediate results, so that we can fix problems easily. To control randomness and show the same result, setting a random seed is important by rule 6. For readability, following rule 7 is helpful. R Markdown is useful to follow rule 8. ReadMe file is a good example for rule 9. For rule 10, GitHub is again a great tool.

Problem 4

```
### Problem 4 : A Scatter Plot and A Histogram ###
```

```
# R version 4.0.2
```

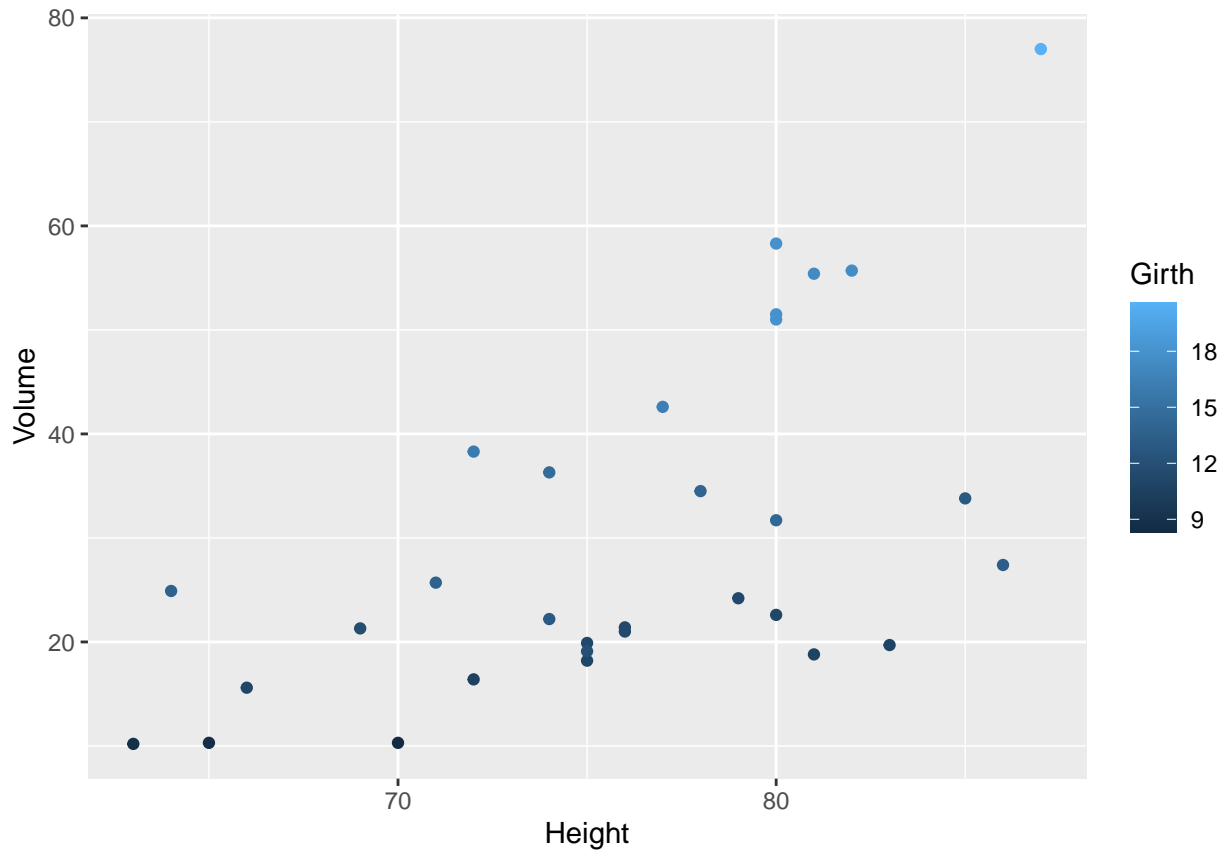
```
library(help = 'datasets') # To get a list of the datasets
```

```
summary(trees) # Summary of trees data
```

```
##      Girth      Height      Volume
##  Min.   : 8.30   Min.   :63   Min.   :10.20
## 1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40
##  Median :12.90   Median :76   Median :24.20
##  Mean   :13.25   Mean   :76   Mean   :30.17
## 3rd Qu.:15.25   3rd Qu.:80   3rd Qu.:37.30
##  Max.   :20.60   Max.   :87   Max.   :77.00
```

```
### A basic scatter plot from 'trees' data using ggplot2
```

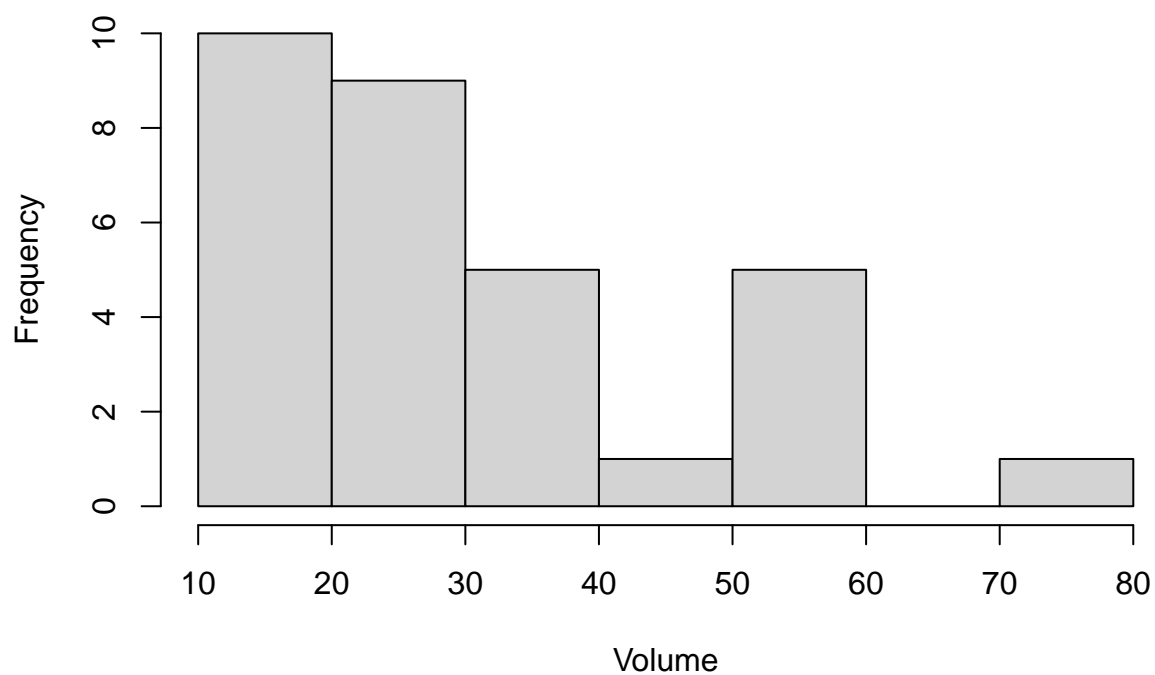
```
ggplot(data = trees) +  
  geom_point(mapping = aes(x = Height, y = Volume, color = Girth))
```



```
### Histogram for 'trees' data
```

```
hist(trees$Volume, main = 'Histogram of Volume of Trees', xlab = 'Volume')
```

Histogram of Volume of Trees



Problem 5

[Push to GitHub](#)

Appendix : R codes

```
knitr::opts_chunk$set(echo = TRUE)
### Problem 1 : Some examples from Primers ###
### Visualization basics
library(ggplot2)
# Map cty to color from mpg data
ggplot(data = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy, color = cty))

### Programming Basics
list_ex <- list(number = 100, logical = TRUE, string = "stories")
list_ex[['number']]

### Problem 4 : A Scatter Plot and A Histogram ###

# R version 4.0.2
library(help = 'datasets') # To get a list of the datasets
summary(trees)             # Summary of trees data

### A basic scatter plot from 'trees' data using ggplot2
ggplot(data = trees) +
  geom_point(mapping = aes(x = Height, y = Volume, color = Girth))

### Histogram for 'trees' data
hist(trees$Volume, main = 'Histogram of Volume of Trees', xlab = 'Volume')
```