

R_Activity_Assignment_3

Dahee Ahn

2024-09-13

```
rm(list=ls())
```

```
options(repos=c(CRAN="https://cran.r-project.org"))  
install.packages("ggplot2")
```

```
## Installing package into 'C:/Users/chemk/AppData/Local/R/win-library/4.4'  
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked  
##  
## The downloaded binary packages are in  
## C:\Users\chemk\AppData\Local\Temp\Rtmpg1Cfpx\downloaded_packages
```

1. Download the cherrytree data set from Canvas and load it into R using `read.table()`. Note this file is a .csv file (comma separated values).

```
cherrytrees_df_csv <- read.table(file="C:/Users/chemk/OneDrive/Desktop/Classes/ENT6707_DataAnalysis/week4/cherrytrees.csv",header=TRUE, sep=",")  
summary(cherrytrees_df_csv)
```

```
##      Girth      Height      Volume      Variety  
## Min.   : 8.30   Min.   :63   Min.   :10.20   Length:31  
## 1st Qu.:11.05   1st Qu.:72   1st Qu.:19.40   Class :character  
## Median :12.90   Median :76   Median :24.20   Mode  :character  
## 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
```

```
head(cherrytrees_df_csv)
```

```
##   Girth Height Volume Variety
## 1   8.3     70   10.3        B
## 2   8.6     65   10.3        B
## 3   8.8     63   10.2        A
## 4  10.5     72   16.4        B
## 5  10.7     81   18.8        B
## 6  10.8     83   19.7        A
```

```
tail(cherrytrees_df_csv)
```

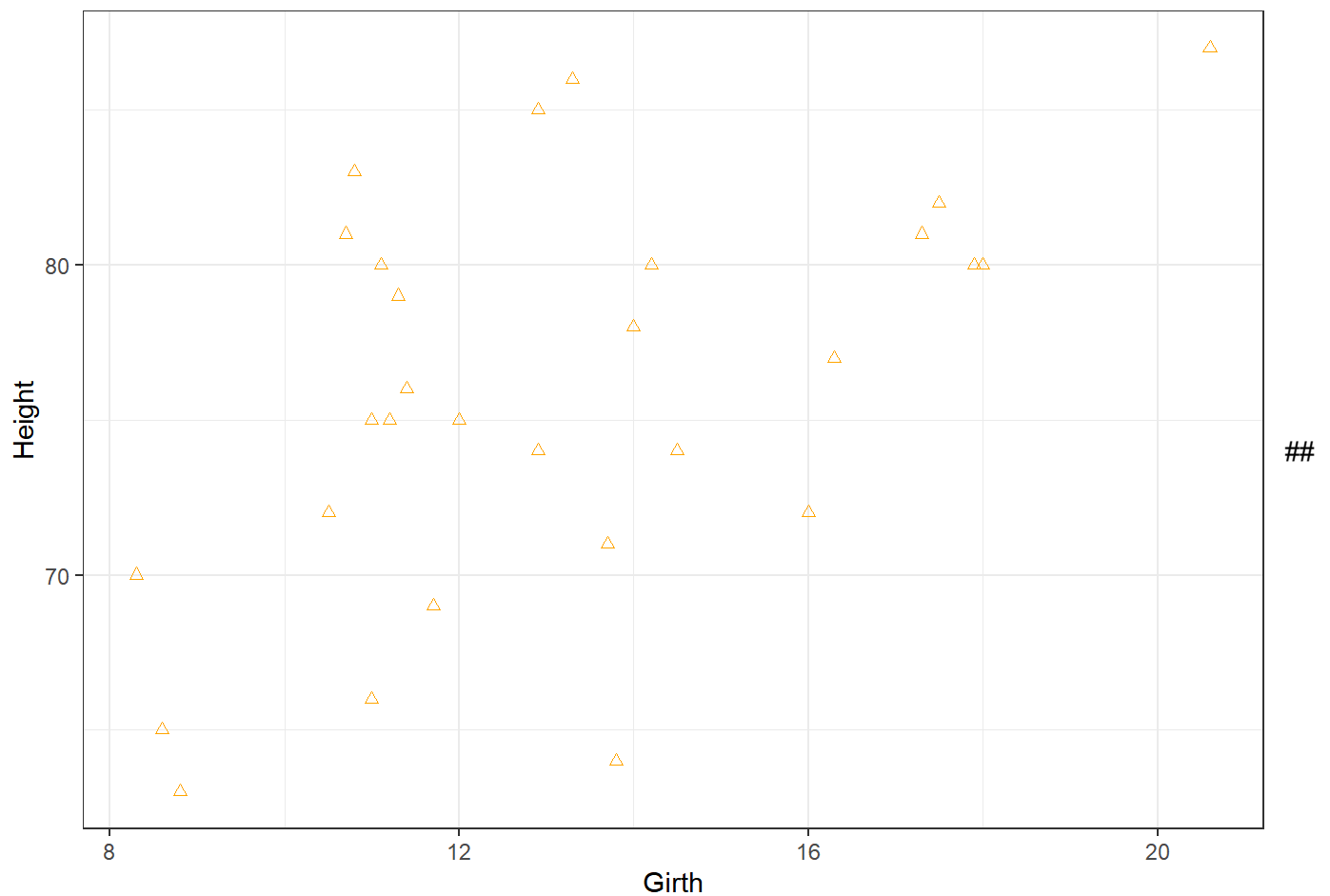
```
##   Girth Height Volume Variety
## 26  17.3     81   55.4        A
## 27  17.5     82   55.7        B
## 28  17.9     80   58.3        A
## 29  18.0     80   51.5        A
## 30  18.0     80   51.0        A
## 31  20.6     87   77.0        A
```

```
str(cherrytrees_df_csv)
```

```
## 'data.frame':   31 obs. of  4 variables:
##  $ Girth  : num  8.3 8.6 8.8 10.5 10.7 10.8 11 11 11.1 11.2 ...
##  $ Height : int  70 65 63 72 81 83 66 75 80 75 ...
##  $ Volume : num  10.3 10.3 10.2 16.4 18.8 19.7 15.6 18.2 22.6 19.9 ...
##  $ Variety: chr  "B" "B" "A" "B" ...
```

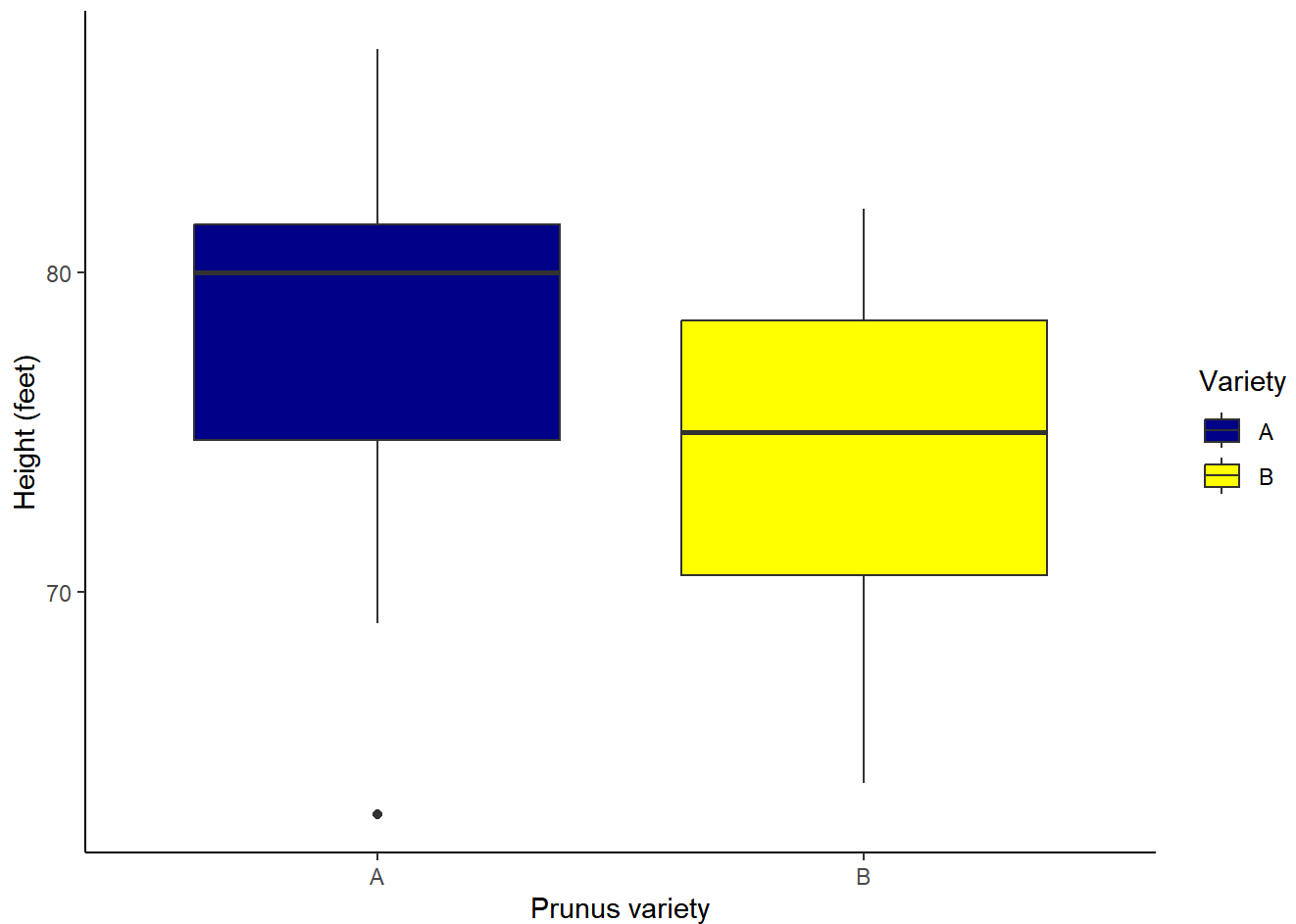
2. Make a scatterplot of Height as a function of Girth using ggplot2. Use the theme theme_bw() and change the black circles to orange, hollow triangles.

```
library("ggplot2")
ggplot(data=cherrytrees_df_csv, mapping=aes(x=Girth, y=Height))+geom_point(shape=2, col="orange")+theme_bw()
```



3. Make a boxplot of Height and as a function of Variety using ggplot2. Change the x-axis label to “Prunus variety” and the y-axis label to “Height (feet)”. Use the theme `theme_classic()` and fill the variety A box as dark blue and the variety B box as yellow.

```
ggplot(cherrytrees_df_csv, mapping=aes(x=Variety, y=Height, fill=Variety))+geom_boxplot()+scale_
fill_manual(values=c("A"="darkblue", "B"="yellow"))+theme_classic()+xlab("Prunus variety")+ylab
("Height (feet)")
```



4. Make the same plot as the previous question, but make it a violin plot instead of a box plot and remove the legend.

```
ggplot(cherrytrees_df_csv, mapping=aes(x=Variety, y=Height, fill=Variety))+geom_violin()+scale_fill_manual(values=c("A"="darkblue", "B"="yellow"))+theme_classic()+xlab("Prunus variety")+ylab("Height (feet)")+theme(legend.position="none")
```

