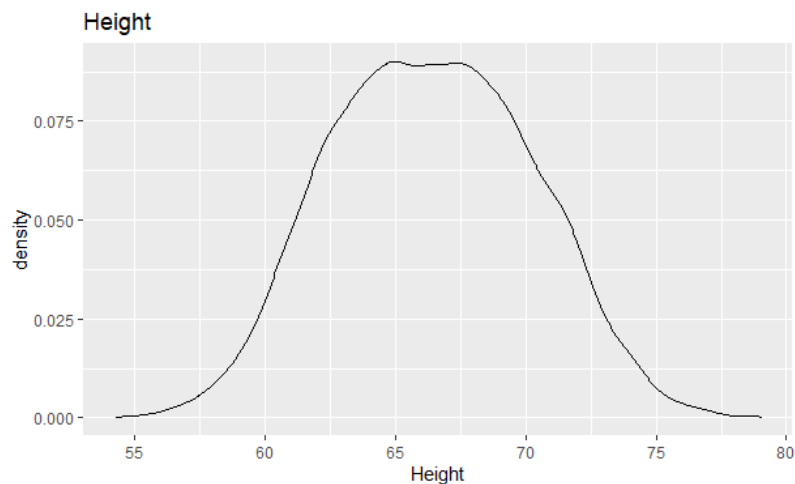


Khan Inan  
Assignment #3  
Applied Biostatistics

1. I used this code

```
library(ggplot2)
library(dplyr)
weight.height %>%
  ggplot() +
  geom_density(aes(x = Height)) +
  labs(title = "Height")
```

To get this



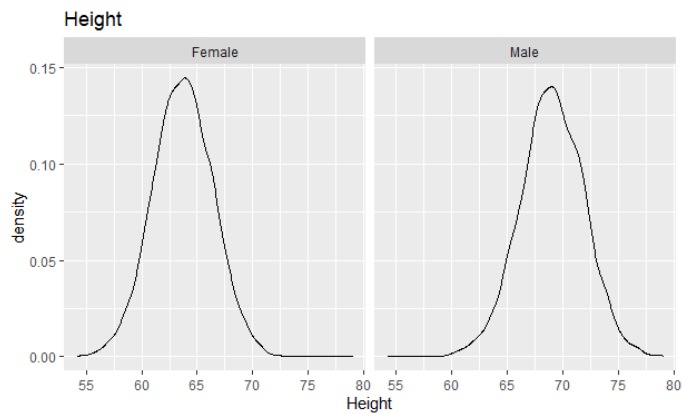
As we can see, this is a gaussian spread of data.

3. Based on the shape of the data, there doesn't seem to be any subgroups because no visible splits or divisions in the data. If something like several groups of outliers were present, we would see subgroups

4. This was my code to get the male and female height separately

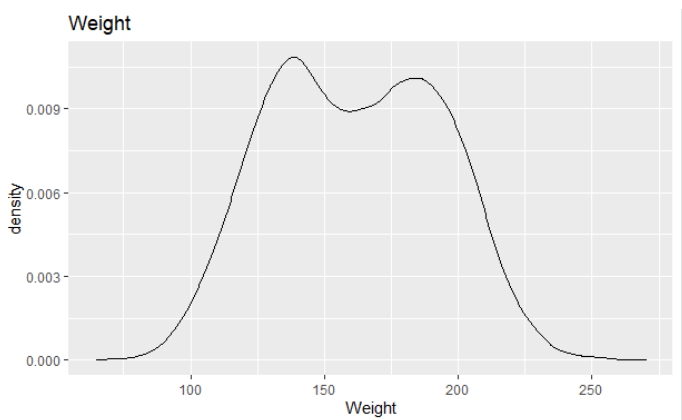
```
library(ggplot2)
library(dplyr)
weight.height %>%
  ggplot() +
  geom_density(aes(x = Height)) +
  facet_wrap(~Gender) +
  labs(title = "Height")
```

|

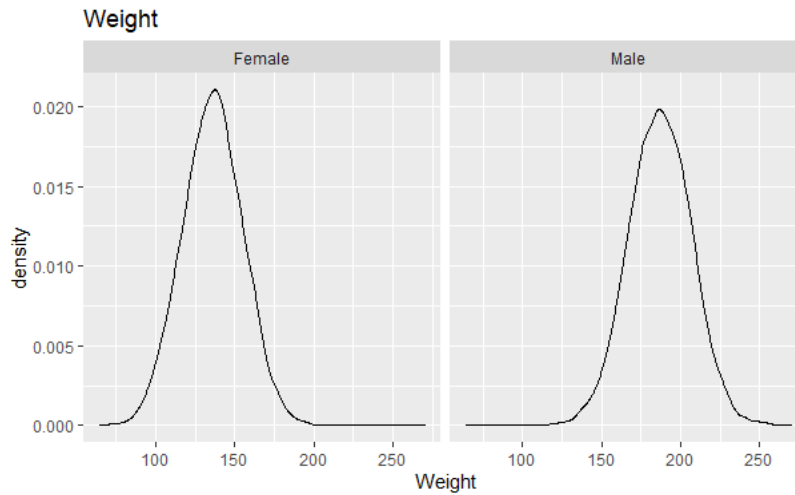


Now the code for the weight together and separately for gender

```
weight.height %>%  
  ggplot() +  
  geom_density(aes(x = weight)) +  
  labs(title = "weight")
```

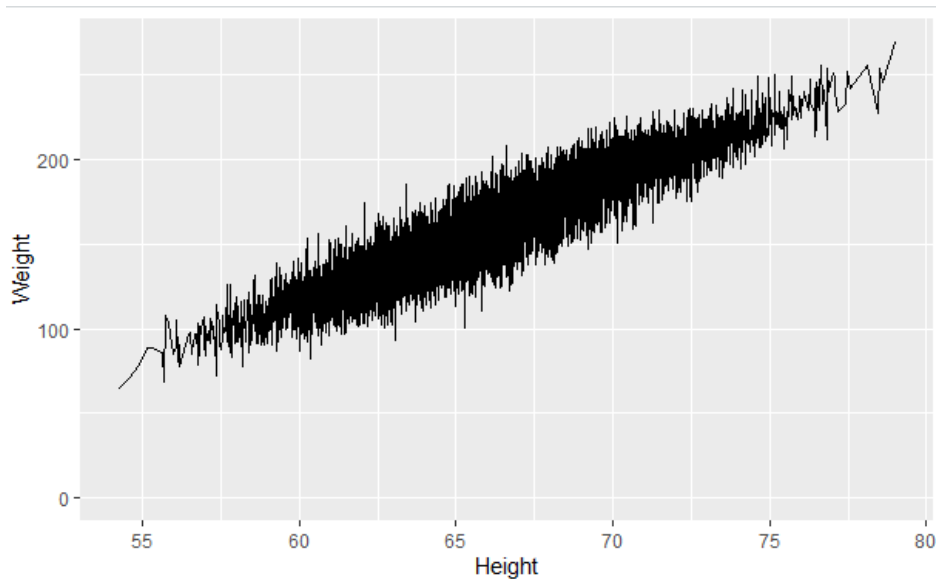


```
weight.height %>%
  ggplot() +
  geom_density(aes(x = weight)) +
  facet_wrap(~Gender) +
  labs(title = "weight")
```



Questions 6 , for plotting x axis Height and y axis Weight on the same graph

```
ggplot(weight.height) +
  geom_density(aes(x = Height, y = weight),
    stat = "identity")
```



Question 7 for plotting weight and height on the same graph for Men and Women separately

```
ggplot(weight.height) +  
  geom_density(aes(x = Height, y = weight),  
                stat = "identity") +  
  facet_wrap(~Gender) +  
  labs(title = "Height vs. weight")
```



8. I would say height is not a good proxy for Weight because you can see the the graphs are very chaotic going up and down seemingly randomly

That being said the rate of increase for men is higher than for women

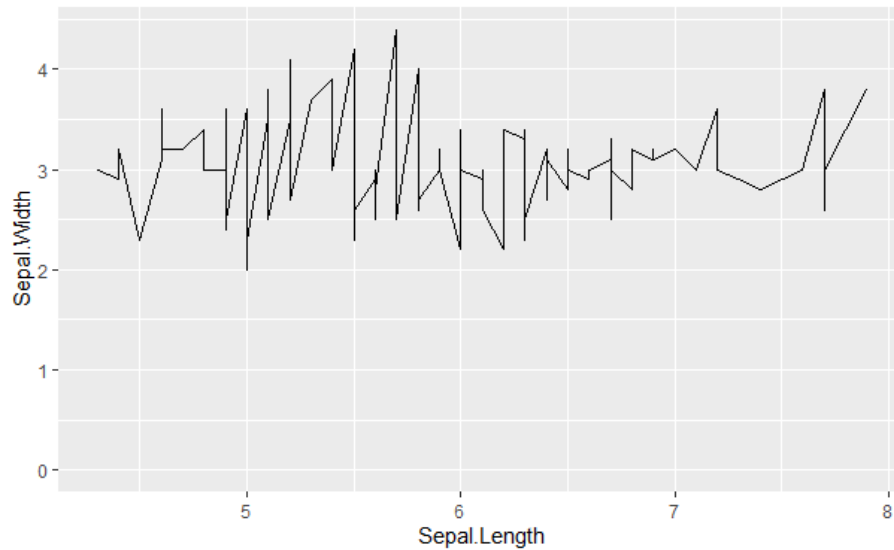
For women the Height would be a better proxy for Weight do to lower standard deviation and data being more congested in a smaller area.

Using a 3-D plot would be a be representation of this data because is it hard to visualize how each point in the graph represents a different person.

9. For the iris data

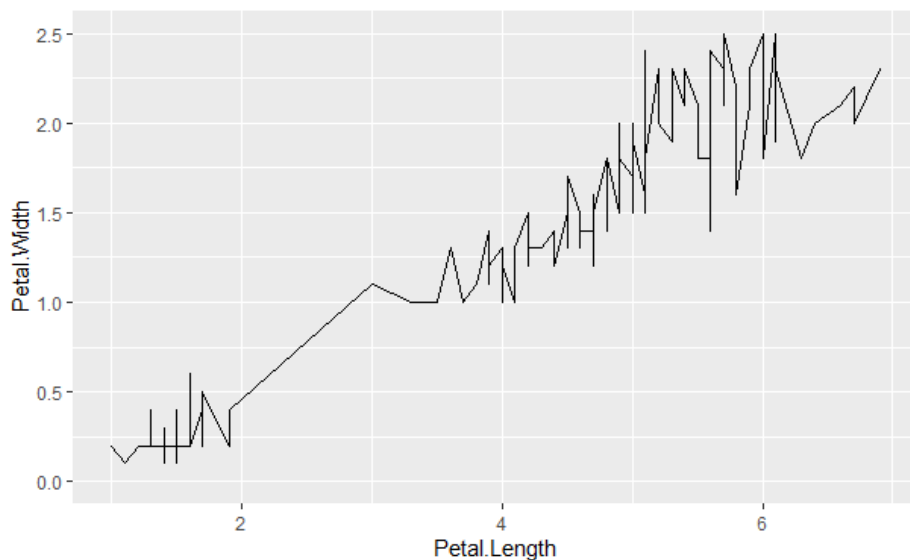
### Comparing Sepal Length and Width

```
ggplot(iris) +  
  geom_density(aes(x = Sepal.Length, y = Sepal.width),  
    stat = "identity")
```



### For Petal length and Width

```
ggplot(iris) +  
  geom_density(aes(x = Petal.Length, y = Petal.width),  
    stat = "identity")
```



Lastly this is the data split by species and put onto the same graph

```
ggplot(iris) +  
  geom_density(aes(x = Petal.Length, y = Petal.width),  
                stat = "identity") +  
  facet_wrap(~Species)  
|
```

