Assignment 1 Question 2

Sheen Thusoo

h. Using R, generate a scatterplot of sepal length (y-axis) vs iris species (x-axis). Add to this plot three red dots indicating the median sepal length of each iris species. Connect the medians (red dots) with red lines.

```
iris_data <- read.csv("Iris.csv")

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
iris_data <- iris_data %>%
  group_by(Species) %>%
  mutate(medians = median(c(SepalLength)))

library("ggplot2")

setosa.SepalLength.median <- first(iris_data[iris_data$Species == "Iris-setosa",
                                             "medians"])[1]
versicolor.SepalLength.median <- first(iris_data[iris_data$Species == "Iris-versicolor",
                                                 "medians"])[1]
virginica.SepalLength.median <- first(iris_data[iris_data$Species == "Iris-virginica",
                                                "medians"])[1]

medians <- c(setosa.SepalLength.median, versicolor.SepalLength.median,
             virginica.SepalLength.median)
medians.data <- data.frame(SepalLength = medians,
                           Species = c("Iris-setosa", "Iris-versicolor", "Iris-virginica" ))

ggplot(data = iris_data) +
  aes(x = Species, y = SepalLength, group = 1) +
  geom_point(size = 1) +
  geom_point(data = medians.data, color = 'red') +
```

```
geom_path(data = medians.data, color = 'red') +
labs(
  title ='Relationship between Species and Sepal Length',
  x = 'Species',
  y = 'Sepal Length (cm)'
)
```