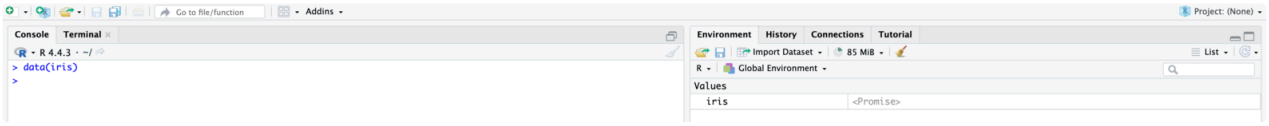
Shao Yun Gao–Week 1: R Output

step 0

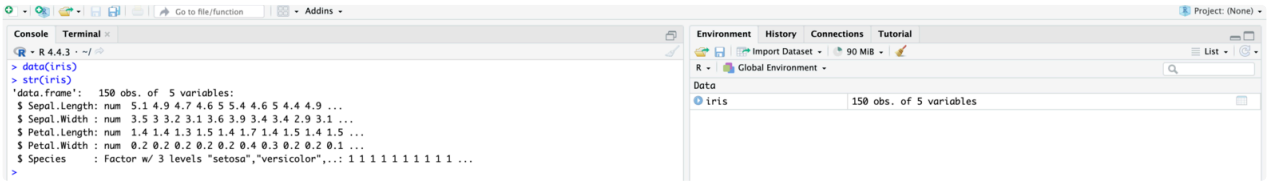
load data set

```R
data(iris)
```

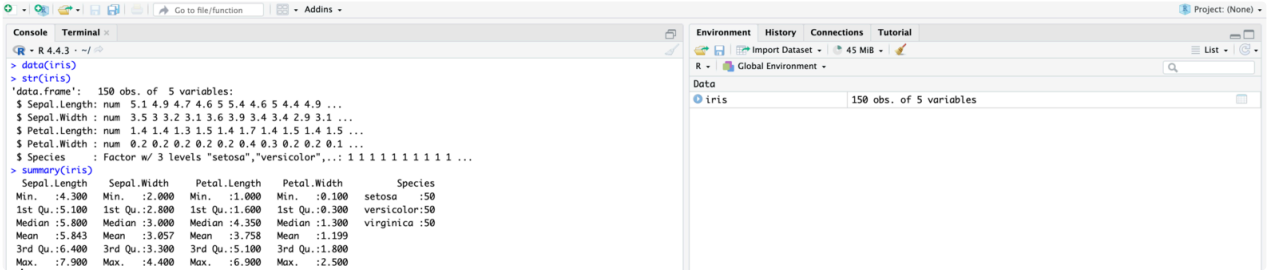


Step 1: Describe the Data

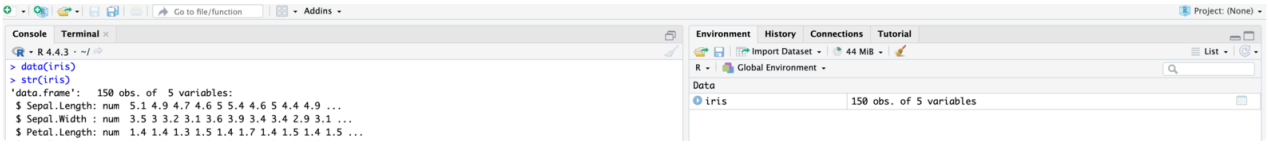
```R
str(iris)
```




```R
summary(iris)
```



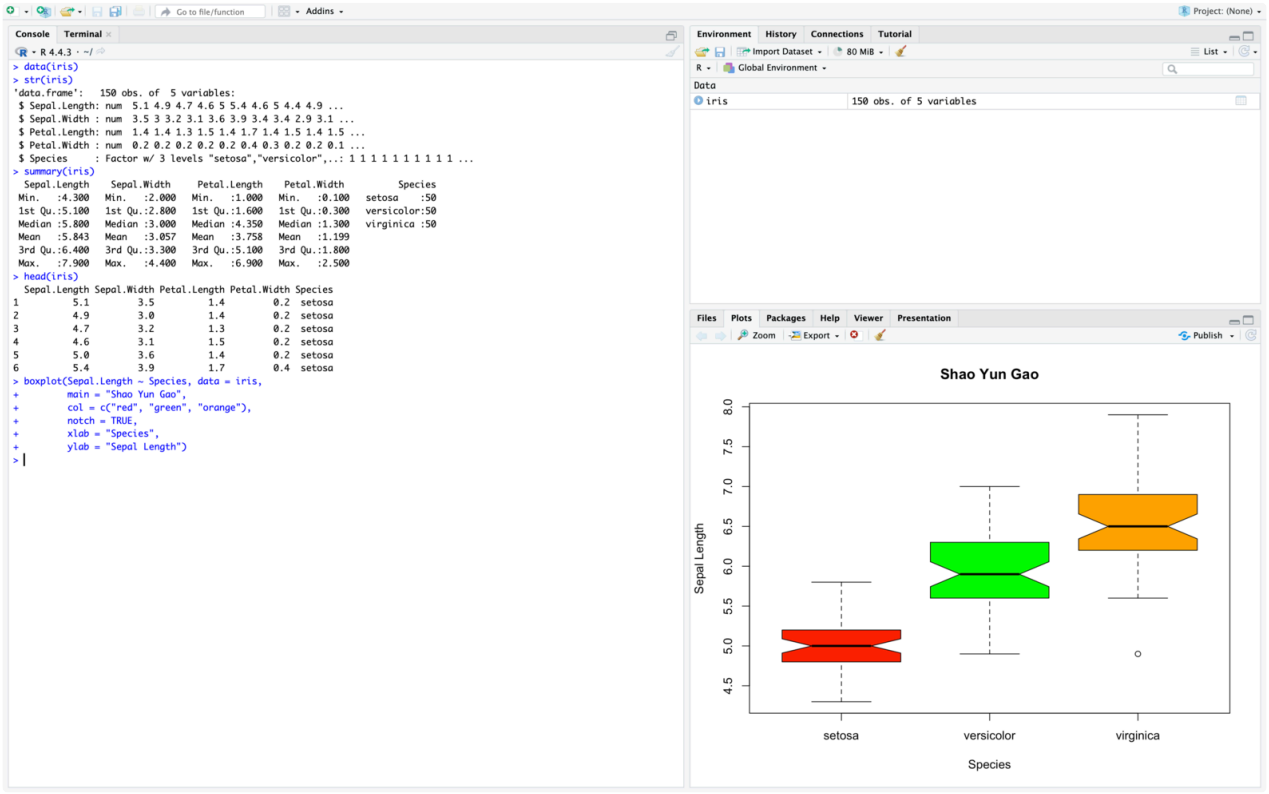

```R
head(iris)
```



Step 2: Box–Whisker Plots

R

```r
boxplot(Sepal.Length ~ Species, data = iris,
        main = "Shao Yun Gao",
        col = c("red", "green", "orange"),
        notch = TRUE,
        xlab = "Species",
        ylab = "Sepal Length")
```

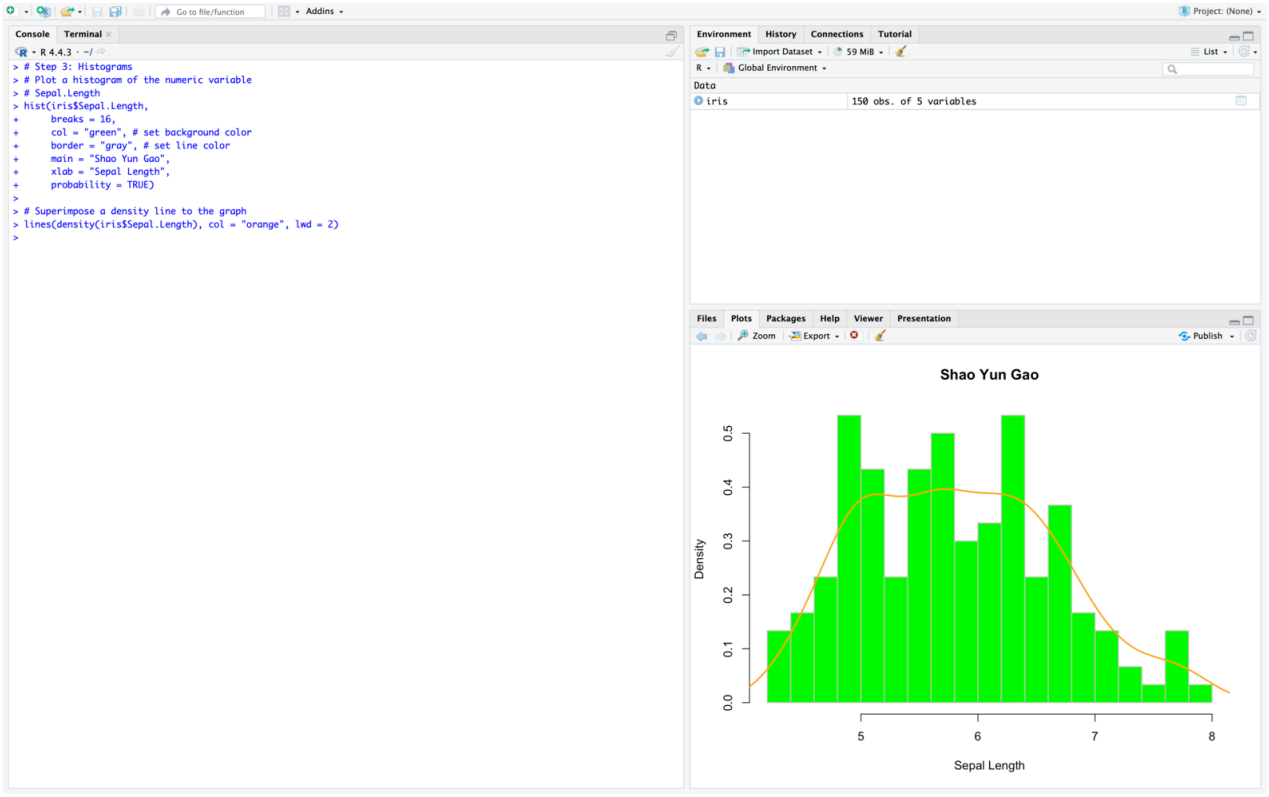


Step 3: Histograms

R

```r
# Step 3: Histograms
# Plot a histogram of the numeric variable
# Sepal.Length
hist(iris$Sepal.Length,
     breaks = 16,
     col = "green", # set background color
     border = "gray", # set line color
     main = "Shao Yun Gao",
     xlab = "Sepal Length",
     probability = TRUE)
```

```r
# Superimpose a density line to the graph
lines(density(iris$Sepal.Length), col = "orange", lwd = 2)
```

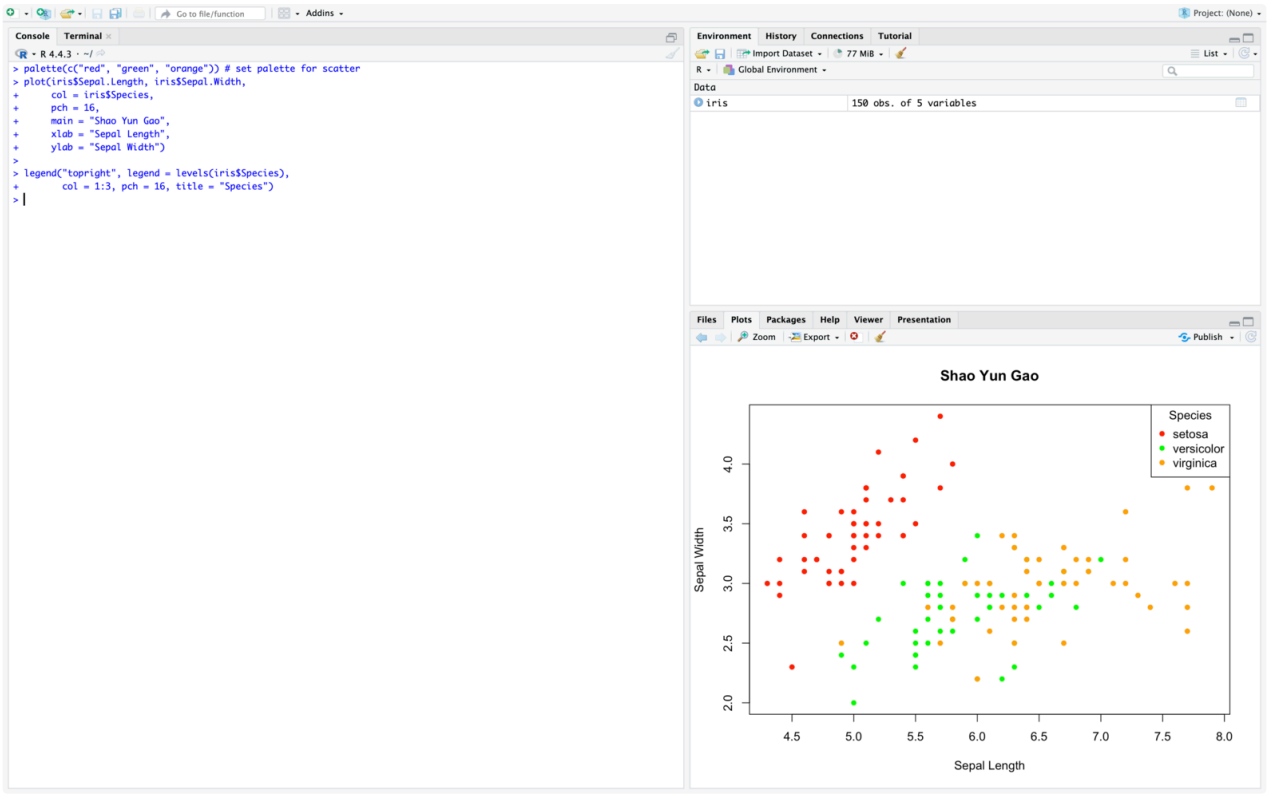


Step 4: Scatter Plots

R

```r
# Step 4: Scatter Plots
# Scatter plot of Sepal.Length vs Sepal.Width colored by Species
palette(c("red", "green", "orange")) # set palette for scatter
plot(iris$Sepal.Length, iris$Sepal.Width,
     col = iris$Species,
     pch = 16,
     main = "Shao Yun Gao",
     xlab = "Sepal Length",
     ylab = "Sepal Width")

legend("topright", legend = levels(iris$Species),
       col = 1:3, pch = 16, title = "Species")
```



Step 5: Simple Math

R

```r
# Step 5: Simple Math Calculations
# For the numeric variable, compute the following statistics
# Sepal.Length
cat("Mean:", mean(iris$Sepal.Length), "\n")
cat("Median:", median(iris$Sepal.Length), "\n")
cat("Min:", min(iris$Sepal.Length), "\n")
cat("Max:", max(iris$Sepal.Length), "\n")
cat("Standard Deviation:", sd(iris$Sepal.Length), "\n")

# Calculate the Median for the numeric for each group member.
medianSpecies <- aggregate(Sepal.Length ~ Species, data = iris, FUN = median)
#Sort the result in Descending order.
sortedMedian <- medianSpecies[order(-medianSpecies$Sepal.Length), ]
# print the sorted result
print(sortedMedian)
```

```
> # Step 5: Simple Math Calculations
> # For the numeric variable, compute the following statistics
> # Sepal.Length
> cat("Mean:", mean(iris$Sepal.Length), "\n")
Mean: 5.843333
> cat("Median:", median(iris$Sepal.Length), "\n")
Median: 5.8
> cat("Min:", min(iris$Sepal.Length), "\n")
Min: 4.3
> cat("Max:", max(iris$Sepal.Length), "\n")
Max: 7.9
> cat("Standard Deviation:", sd(iris$Sepal.Length), "\n")
Standard Deviation: 0.8280661
>
> # Calculate the Median for the numeric for each group member.
> medianSpecies <- aggregate(Sepal.Length ~ Species, data = iris, FUN = median)
> #Sort the result in Descending order.
> sortedMedian <- medianSpecies[order(-medianSpecies$Sepal.Length), ]
> # print the sorted result
> print(sortedMedian)
     Species Sepal.Length
3  virginica          6.5
2 versicolor          5.9
1     setosa          5.0
>
```