

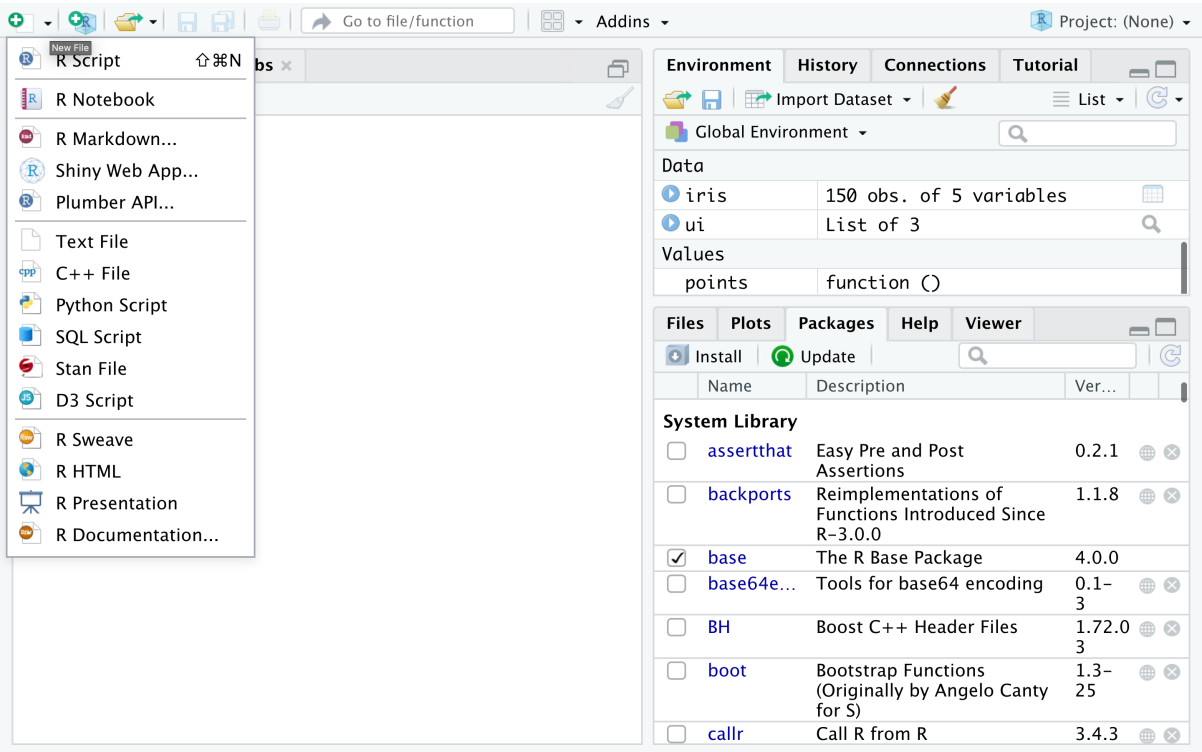


Objective for Exercise:

This lab introduces you to plotting in R with **ggplot** and **GGally**. **GGally** is an extension of **ggplot2**

Exercise:

1. Click in the tiny **plus** symbol top left and select **R Script** to create a new R script if you dont have one open already.



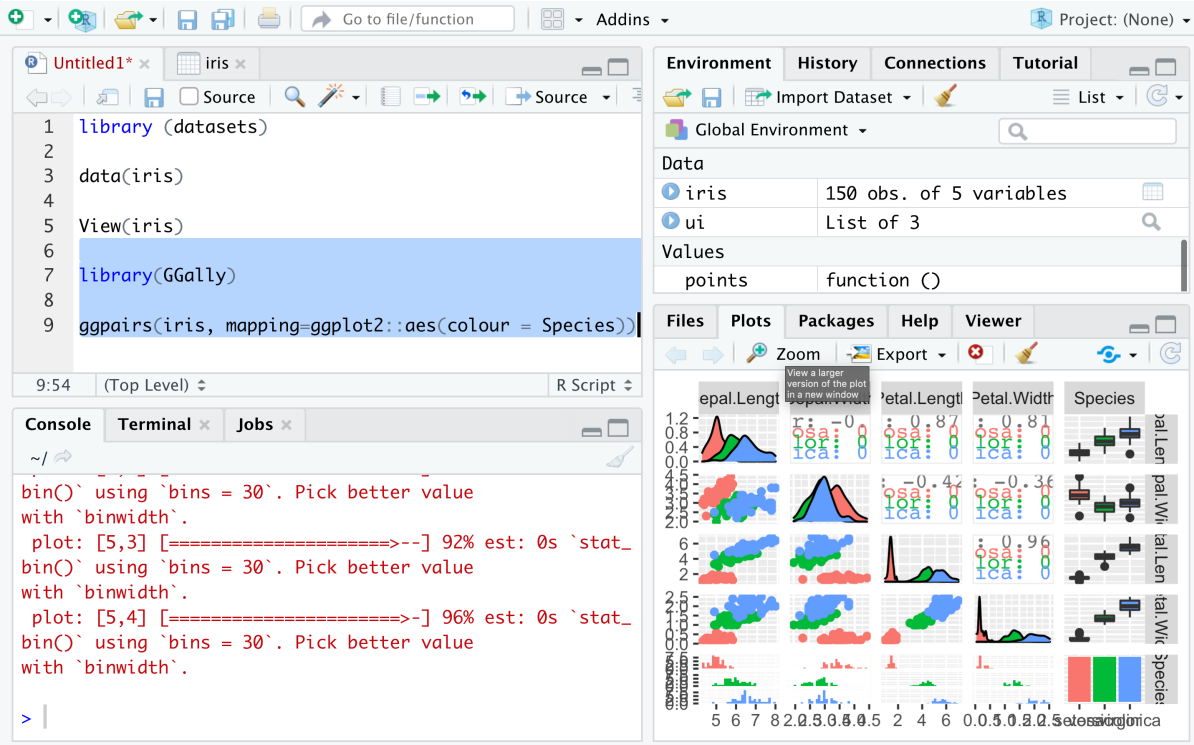
2. We will use the iris dataset. If you don't have it loaded, please copy and paste the following into your R script file.

```
library(datasets)
data(iris)
```

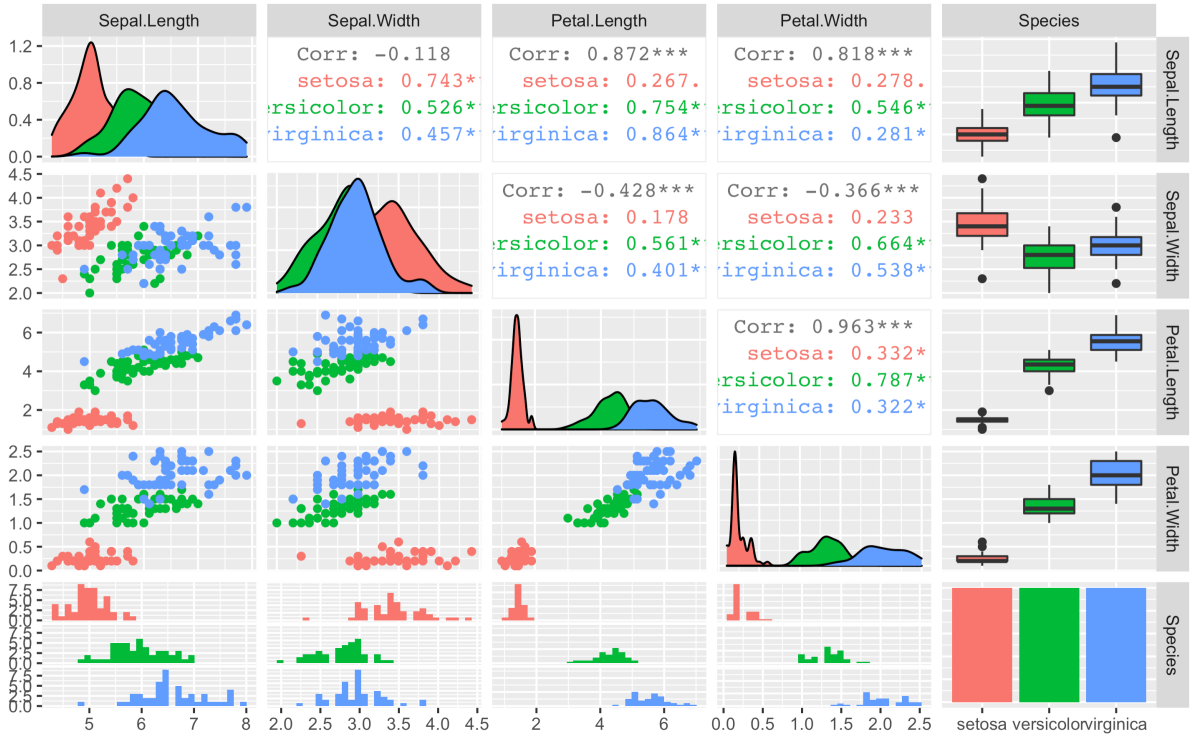
3. In the previous lab, you installed the libraries necessary to create some nice plots let’s execute the following commands:

```
library(GGally)
ggpairs(iris, mapping=ggplot2::aes(colour = Species))
```

4. Select the commands and click on run on the top. You’ll now see the following plot in the **plots** window:



5. Click on **Zoom** icon on the plot window to zoom and see the plot.



6. This gives us a lot of information for a single line of code. First, we see the data distributions per column and species on the diagonal. Then we see all pair-wise scatter plots on the tiles left to the diagonal, again broken down by color. It is, for example, obvious to see that a line can be drawn to separate **setosa** against **versicolor** and **virginica**. In later courses, we'll of course teach how the overlapping species can be separated as well. This is called supervised machine learning using non-linear classifiers by the way. Then you see the correlation between individual columns in the tiles right to the diagonal which confirms our thoughts that **setose** is more different, hence more easy to distinguish, than **versicolor** and **virginica** since a correlation value close to one signifies high similarity whereas a value closer to zero signifies less similarity. The remaining plots on the right are called **box-plots** and the ones at the bottom are called **histograms** bit we won't go into detail here and save this for a more advanced course in this series.

Author(s)

Romeo

Other Contributor(s)

Lavanya

Change log

Date	Version	Changed by	Change Description
2020-12-10	1.1	Aije	Created simplified version of the lab
2020-12-10	1.0	Malika Singla	Migrated lab to Markdown

© IBM Corporation 2020. All rights reserved.