

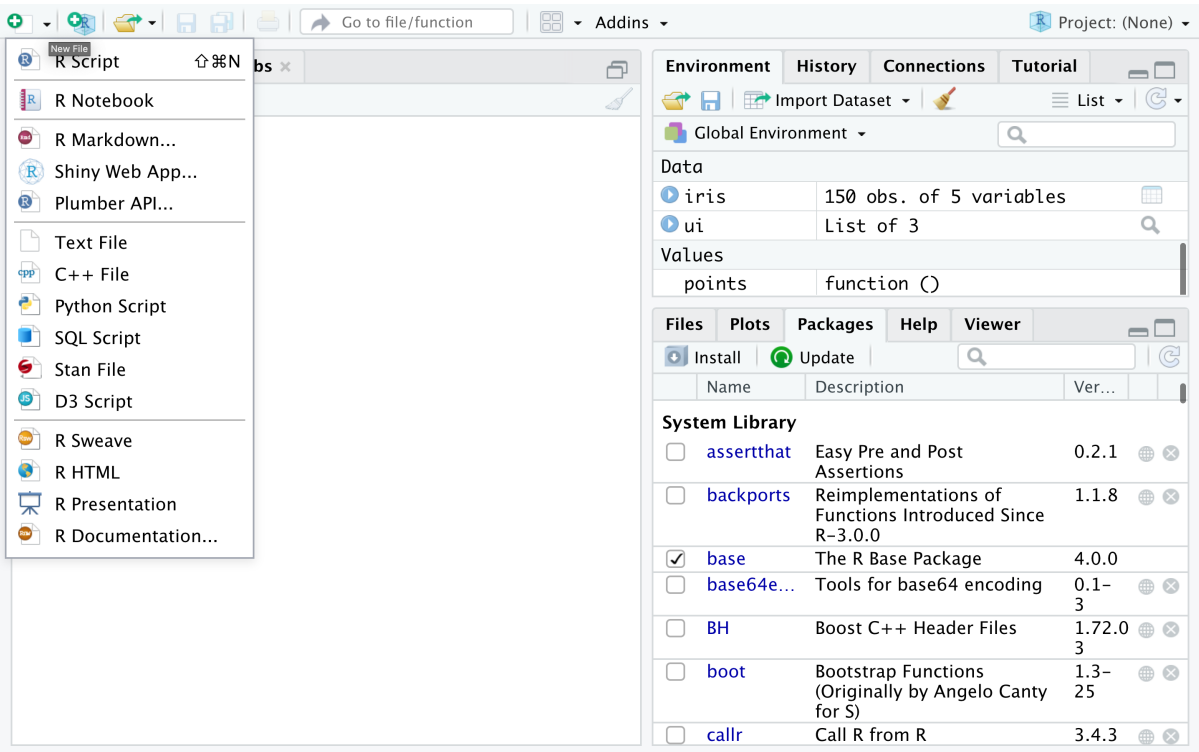


Objective for Exercise:

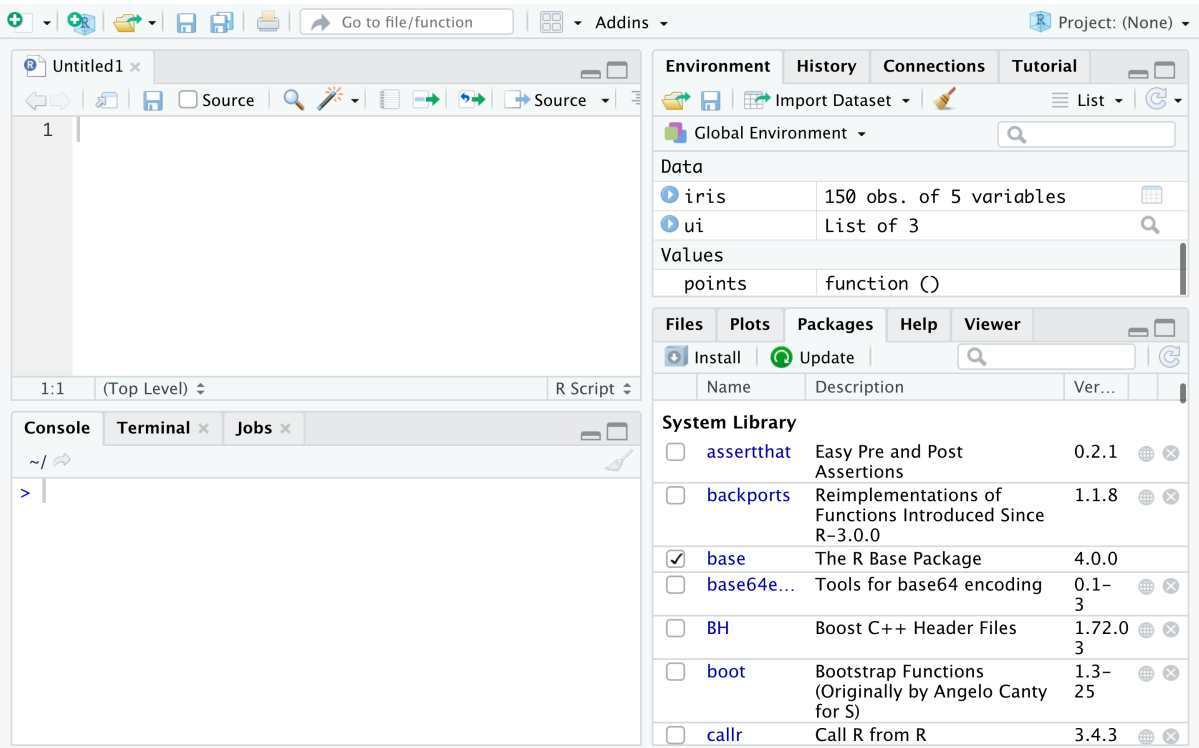
This lab introduces you to R and RStudio

Exercise:

1. Please install R appropriate for your operating system. You can obtain it [here](#)
2. Once you’ve installed R, please download and install RStudio appropriate for your operating system. You can find it [here](#)
3. Start RStudio. An R-Studio window will open up. Now click in the tiny **plus** symbol top left and select **R Script**

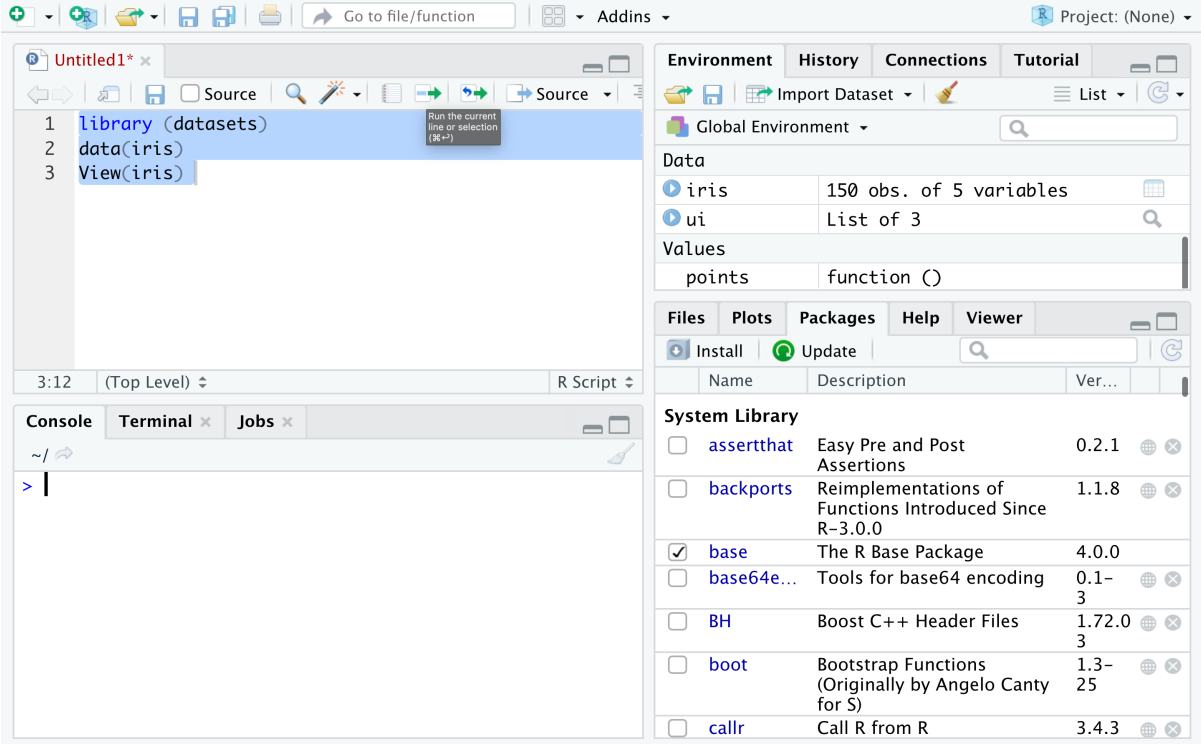


4. An untitled R Script panel opens. It would look like this.

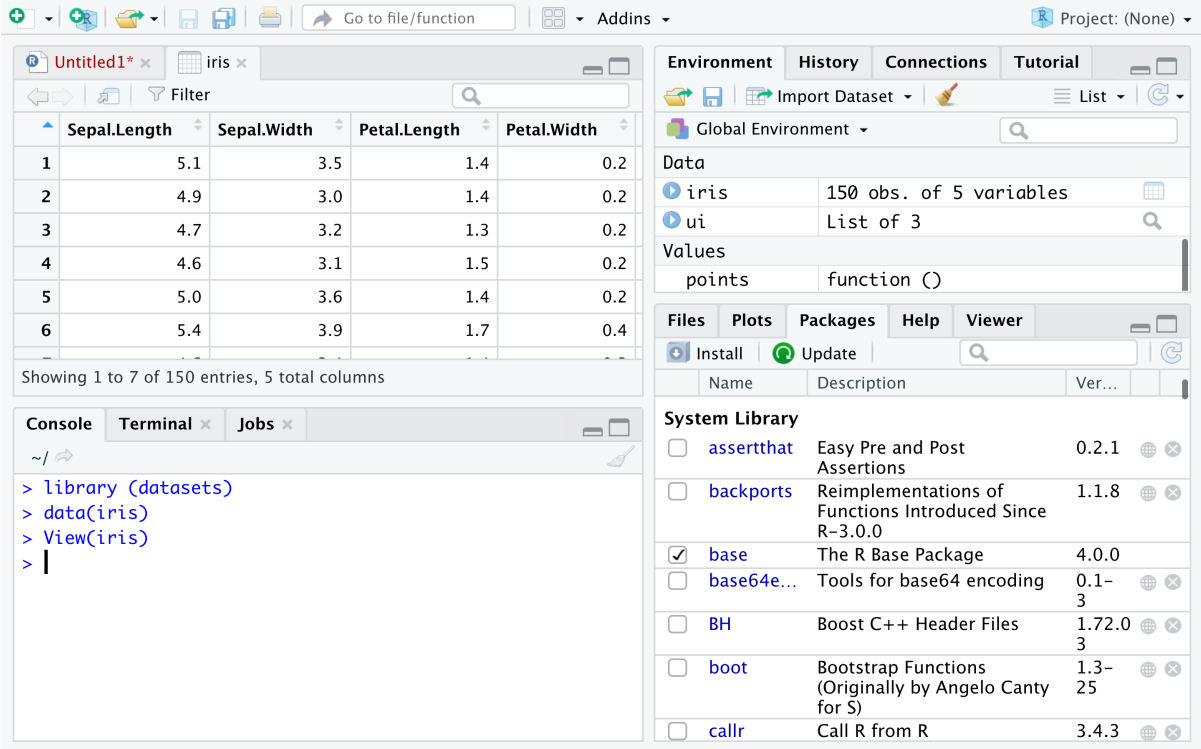


5. Now we load the iris dataset which you already should be familiar with from the previous labs. Please enter the following lines into the editor window which just appeared. Then select them all such that they turn blue. Then click on the tiny run icon just above the editor window.

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

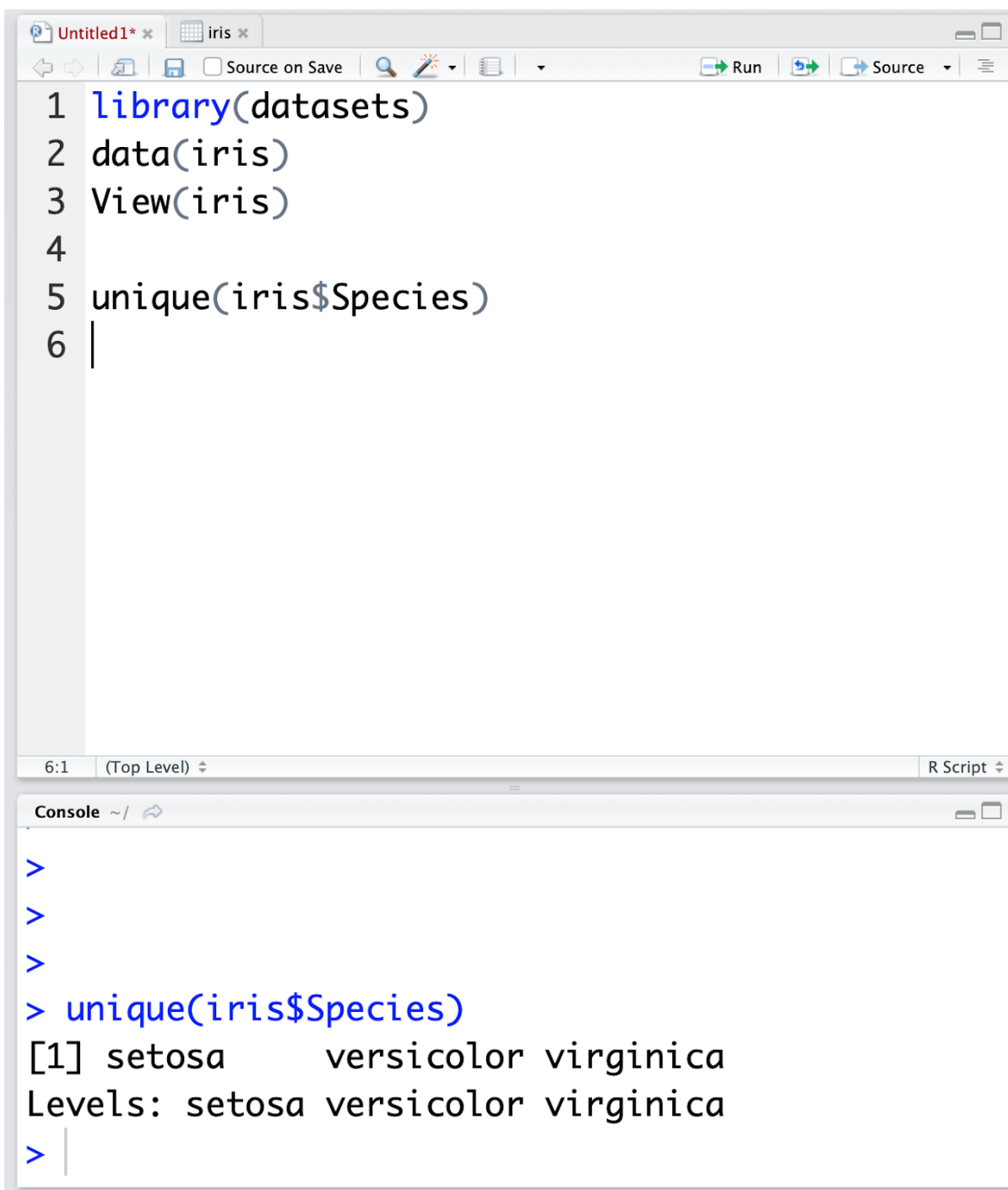


6. You are directly taken to the data view tab to inspect your dataset. We can see that there are five columns in this data set and that the first four are floating point and the last one is label of data type string which contains the category value of our data set. We also see that we have 150 entries in total of which we are seeing the first 19.



7. Now let's find how many different species there are present in the data set. Type the following command into the editor window and click the run.

```
unique(iris$Species)
```



The screenshot shows the RStudio interface. The top pane displays an R script with the following code:

```
1 library(datasets)
2 data(iris)
3 View(iris)
4
5 unique(iris$Species)
6 |
```

The bottom pane shows the console output for the command `unique(iris$Species)`:

```
>
>
>
> unique(iris$Species)
[1] setosa    versicolor virginica
Levels: setosa versicolor virginica
> |
```

In the Console window at the bottom you'll see the result of the executed command and will know that there are only three different species present in the data set.

8. Now it's time to look into the data set in more detail. We'll create a similar plot as we've already done in the Jupyter lab. To do so we first need to install the ggplot2 package. Please type and execute the following command in the console window.

```
install.packages(c("GGally", "ggplot2"))
```



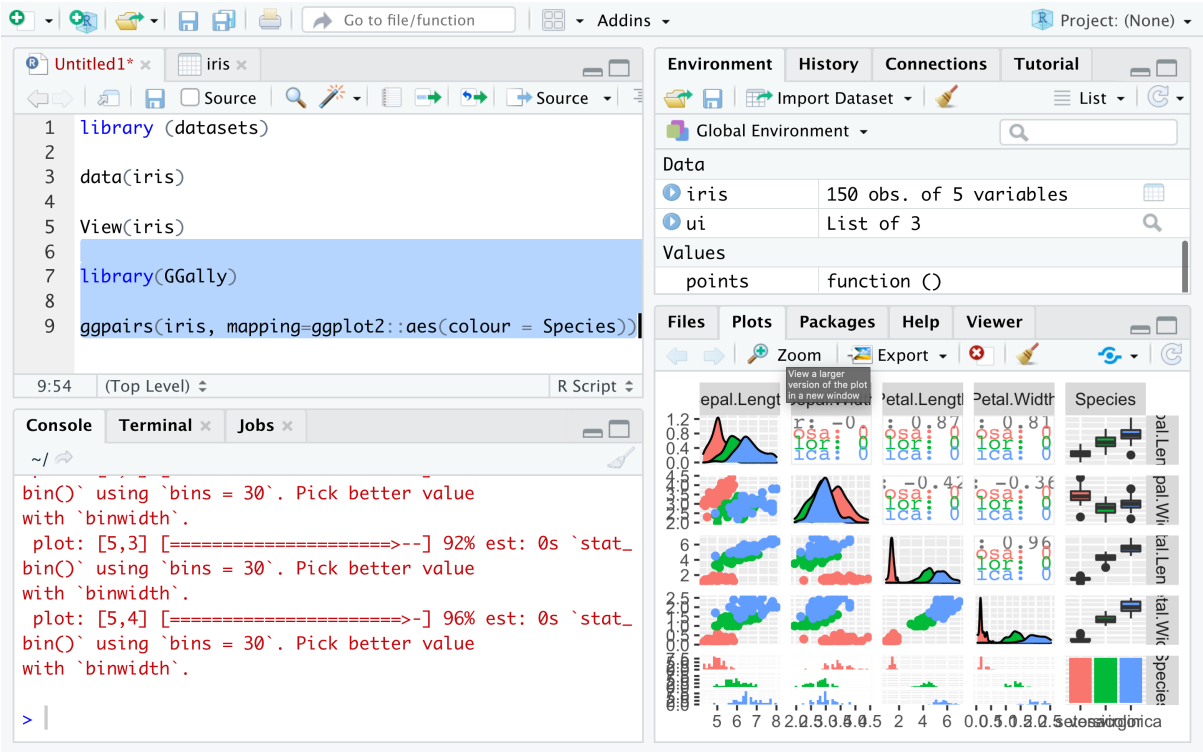
The screenshot shows the RStudio console window with the following command entered:

```
> install.packages(c("GGally", "ggplot2"))
```

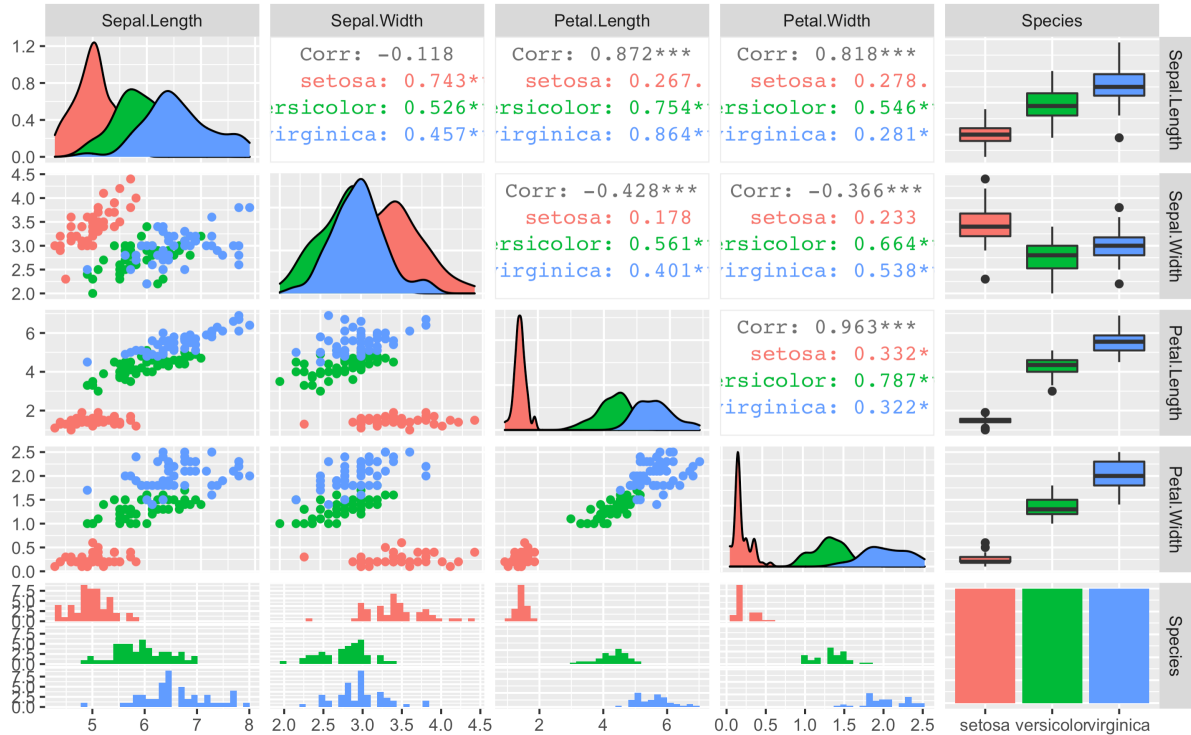
9. Now that you have installed the libraries necessary to create some nice plots let's execute the following commands:

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

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



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



12. 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.

This concludes this lab, I hope you've enjoyed it!

Author(s)

Romeo

Other Contributor(s)

Lavanya

Change log

Date	Version	Changed by	Change Description
------	---------	------------	--------------------

Date	Version	Changed by	Change Description
2020-08-25	2.0	Lavanya	Migrated Lab to Markdown and added to course repo in GitLab

© IBM Corporation 2020. All rights reserved.