Reports in Rmarkdown

Justin and Arturo

Main Title

Report for Species versicolor

In this Rmarkdown we will show how to build a workflow around RMarkdown that allows for the automated generation of reports.

Step 1: Repository Organisation

There are several key files that are often useful when writing remports in `RMarkdown` (especially when the project is *big*):

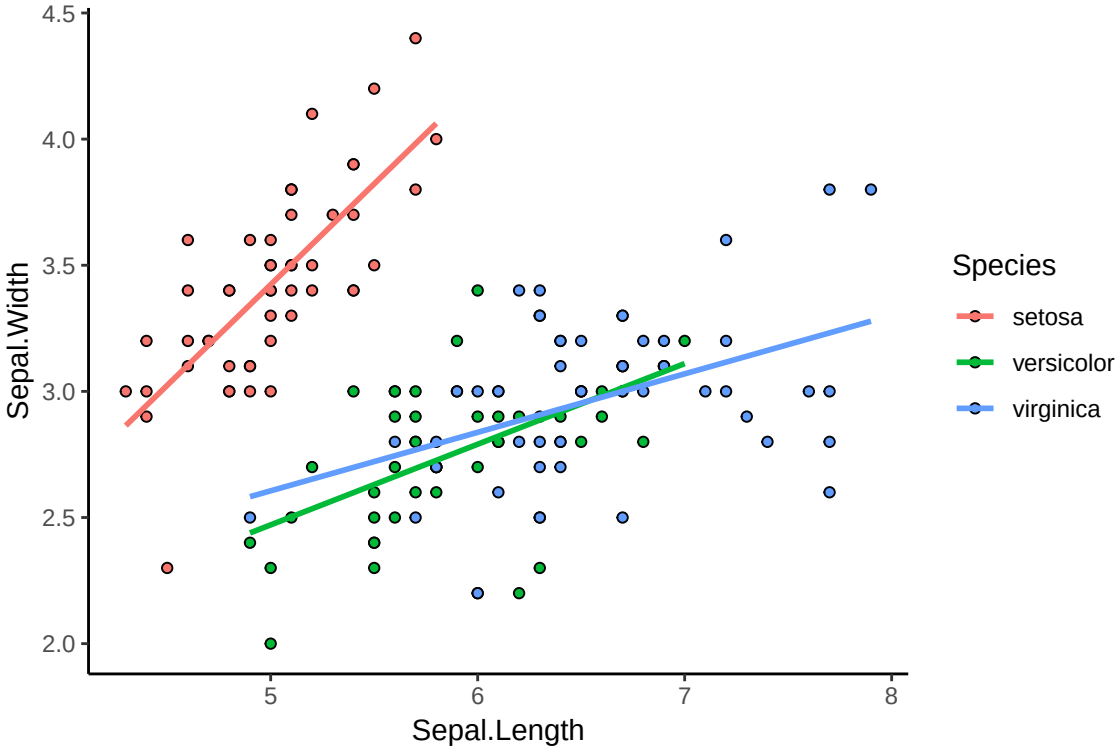
- RMarkdown document (this one)
- Data (usually excel/csv files)
- Control file: used to automatically generate reports
- R Scripts and additional data wrangling (helps keep long code out of this document)
- Git/version control (we can leave this for another time)

Rmarkdown Basics

RMarkdown is a document written in the markdown language that has embedded `R code`. For instance, I can add a calculation into the the text, such as the sum of two integers (three and five): 8. We can also add *code chunks*:

setosa	versicolor	virginica
50	50	50

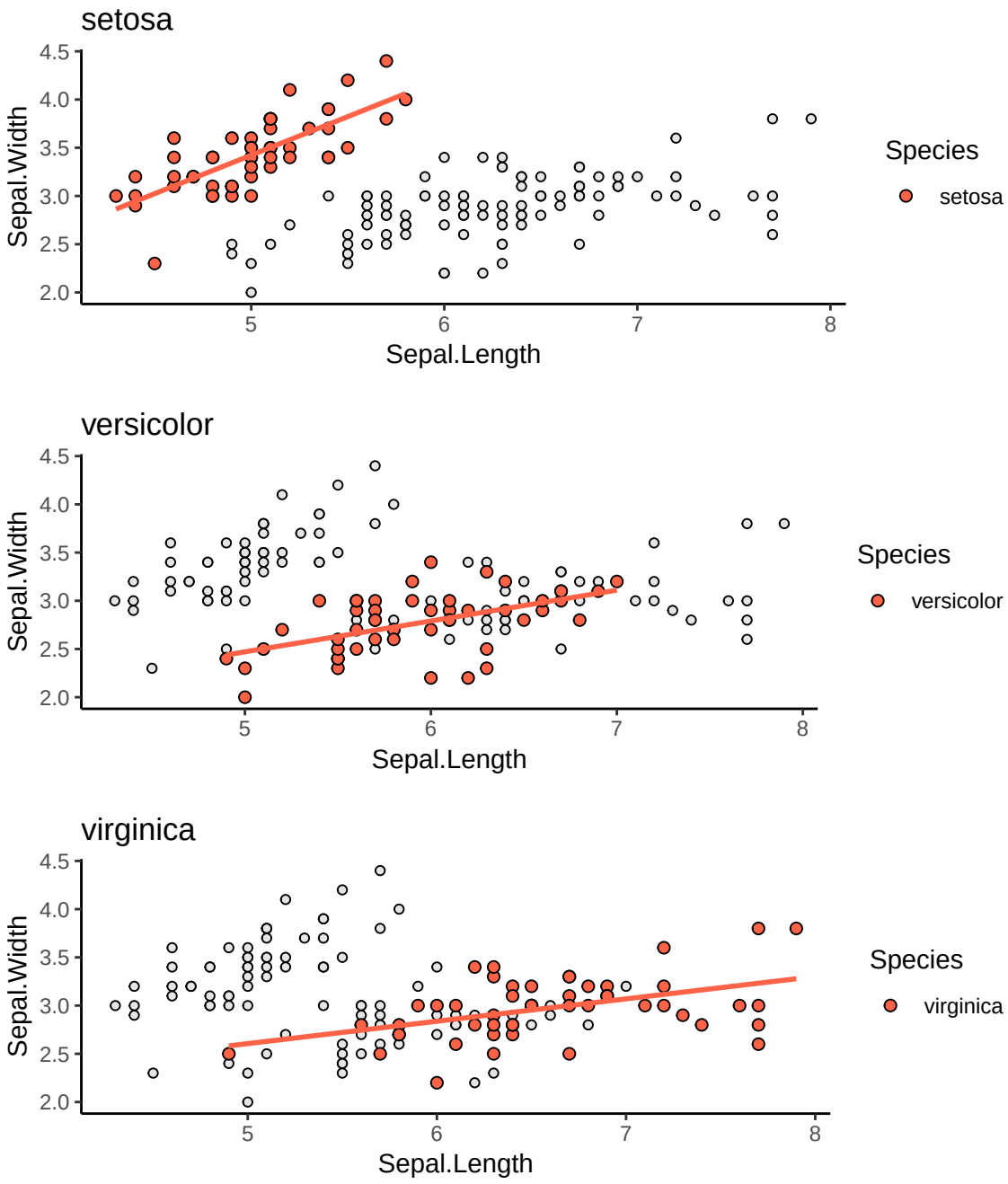
Ultimately, your report will begin to blend text, code, tables and figures:



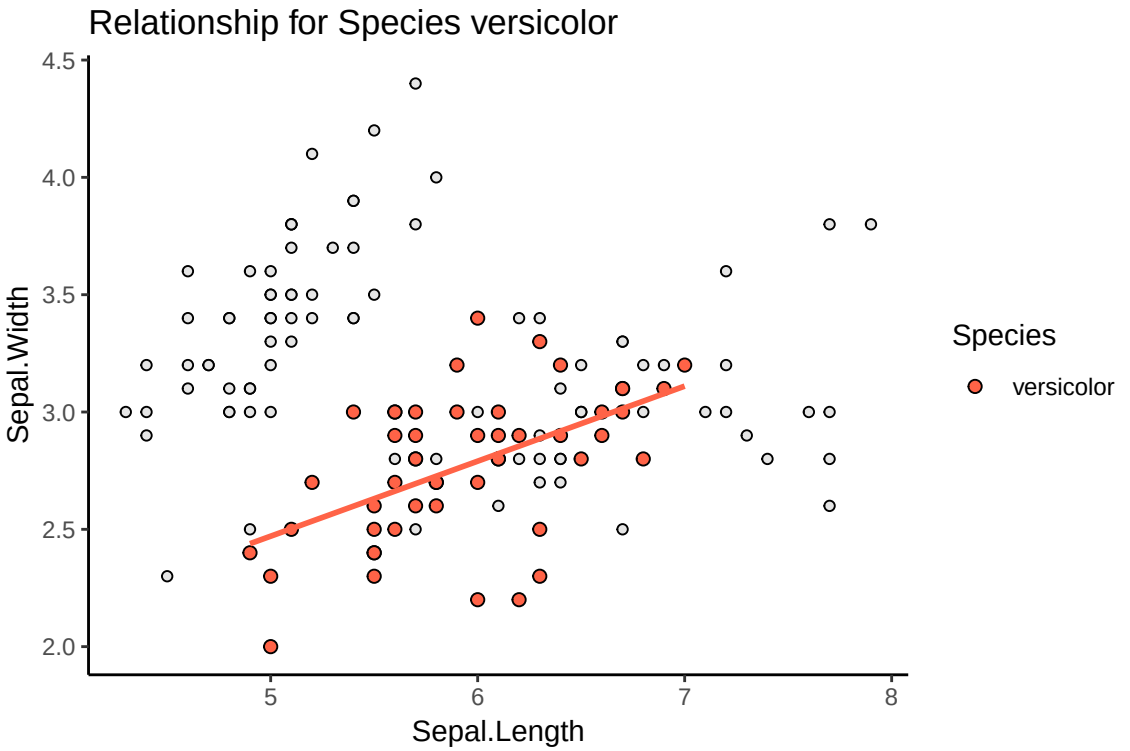
Generating multiple plots/reports

Now imagine that we want to look at the correlation between sepal lengths and sepal widths for each of the iris plant species, **BUT** we want to keep the other two species unidentifiable... For instance, imagine we want to make a report for *setosa* but we don't want the person reading it to know what the correlation is for the other two species.

If you wanted to make three figures for each species, you could make a loop within this RMarkdown:



But, we are interested in generating multiple reports, so the loop should be placed outside of this RMarkdown, for this purpose we make a control file. The following plot will than render for each unique species



Finally, we can customiser large sections of text that someone has written externally and placed in a csv. To do that we map a *csv* file with the custom text to the specific report parameter (Species). The following come from the wikipedia entries for each species:

Iris versicolor is also commonly known as the blue flag, harlequin blueflag, larger blue flag, northern blue flag, and poison flag, plus other variations of these names, and in Britain and Ireland as purple iris. It is a species of Iris native to North America, in the Eastern United States and Eastern Canada. It is common in sedge meadows, marshes, and along streambanks and shores. The specific epithet versicolor means "variously coloured". It is one of the three Iris species in the Iris flower data set outlined by Ronald Fisher in his 1936 paper "The use of multiple measurements in taxonomic problems" as an example of linear discriminant analysis.