

An Introduction and Comparison of R and Python Data Science Applications



Spencer McConville, Spring 2022 Computational Research Assistant
Math Faculty Computing Facility, University of Waterloo



1 - Introduction

This document is meant to be used as a reference and/or guide for students learning data science applications in R or Python for the first time. Individuals will most likely find this information more useful if they have prior knowledge in either R or Python and are trying to learn one or the other. There is a strong emphasis throughout this document on creating similar results in both languages and how to reproduce the results of one language in the other.

R and Python are two different languages created by different people for different reasons. R was designed for statistical computing, created by statisticians. On the other hand Python is a general purpose, object oriented language designed to highlight programmer productivity and flexibility. It wasn't until data science applications became more popular that people started to use the two languages for the same tasks. Modern day packages for both languages now allow one to perform such tasks in a way that there may be no clear advantage in using one over the other.

When it comes to data science applications, the two languages have code bases which are structured much differently. One can think of R as a collection of many smaller packages built on top of R's built-in functions. These smaller packages each have their own functions to perform specific tasks. Python can be thought of as a collection of larger packages. These larger packages can usually perform a wide range of tasks and can be used for multiple purposes. For example Python's PyTorch module can be used in many different areas of machine learning. Essentially, this can be broken down into a decision of either learning a large number of smaller R packages or learning a small number of larger Python modules.

If you have not learned either R or Python yet and do not know which to choose, I recommend doing some additional research on both of these languages. I would start off by watching all or a subset of these Youtube videos:

- [\[1\] R Vs Python | Which is Better for Data Analysis?](#)
- [\[2\] Python Vs R | Which is the best Programming Language](#)
- [\[3\] R vs Python - What should I learn in 2020?](#)

These videos will provide a great start in helping you choose a language to begin with. After watching these videos I would then look at more technical information. Try reading the short infographic created by DataCamp:

- [\[4\] Choosing Python or R for Data Analysis?](#)

Finally, a great document created by Lengersdorff, L. provides a short comparison of the two languages. This may be more suited for individuals with prior knowledge of R and Python:

- [\[5\] From R to Python-a short tutorial for \(data\) scientists.](#)

Both R and Python have their own web pages which are full of download/installation instructions, documentation information and other news/events. The official web pages can be found at:

- [\[6\] R Core Team, 2020. R: A Language and Environment for Statistical Computing](#)
- [\[7\] Van Rossum, G. & Drake, F.L., 2009. Python 3 Reference Manual](#)

The remainder of this document will cover many areas of data science applications in the form of examples (with code included). These sections include mathematical objects, mathematical operations, least square solutions, computational differences between the two languages, statistical analysis, data visualization and predictive modelling. It should be noted that throughout this document Python version 3.7.10 and R version 4.0.2 will be used.

```
!python --version
```

```
Python 3.7.10
```

```
version$version.string
```

```
'R version 4.0.2 (2020-06-22)'
```

[1] Alex The Analyst, 2021. R vs Python | Which is Better for Data Analysis?, [\[online video\]](#)

[2] Great Learning, 2021. Python Vs R | Which is the best Programming Language, [\[online video\]](#)

[3] Intellipaat, 2019. R vs Python - What should I learn in 2020? | R and Python Comparison, [\[online video\]](#)

[4] DataCamp Team, 2020. Choosing Python or R for Data Analysis? An Infographic, [\[link\]](#)

[5] Lengersdorff, L., From R to Python-a short tutorial for (data) scientists, [\[link\]](#)

[6] R Core Team, 2020. R: A Language and Environment for Statistical Computing, [\[link\]](#)

[7] Van Rossum, G. & Drake, F.L., 2009. Python 3 Reference Manual, [\[link\]](#)