Installing R

Now that we've got a handle on what a data scientist is, how to find answers, and then spent some time going over a data science example, it's time to get you set up to start exploring on your own. And the first step of that is installing R.

What is R? What is CRAN?

First, let's remind ourselves exactly what R is and why we might want to use it.

R is both a programming language and an environment, focused mainly on statistical analysis and graphics. It will be one of the main tools you use in this and following courses.

R is downloaded from the Comprehensive R Archive Network, or CRAN, and while this might be your first brush with it, we will be returning to CRAN time and time again, when we install packages - so keep an eye out!

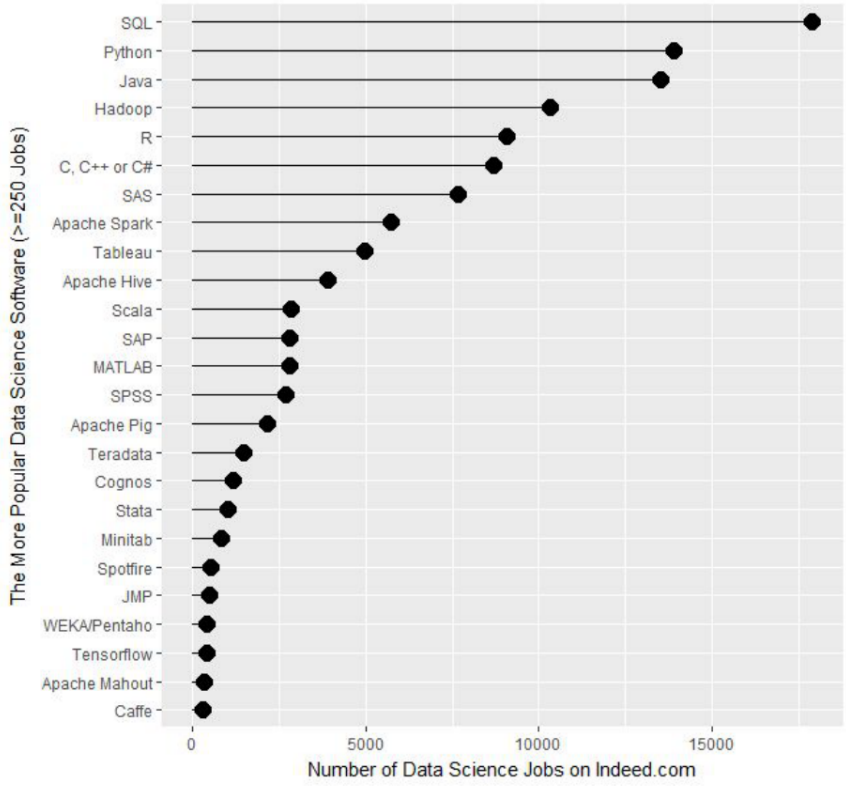
Why should you use R?

Outside of this course, you may be asking yourself - why should I use R?

The reasons for using R are myriad, but some big ones are:

1) Its popularity

R is quickly becoming the standard language for statistical analysis. This makes R a great language to learn as the more popular a software is, the quicker new functionality is developed, the more powerful it becomes, and the better the support there is! Additionally, as you can see in the graph below, knowing R is one of the top five languages asked for in data scientist job postings!



http://r4stats.com/articles/popularity/

R's popularity among data scientists from r4stats.com

2) Its cost

FREE!

This one is pretty self-explanatory - every aspect of R is free to use, unlike some other stats packages you may have heard of (eg: SAS, SPSS), so there is no cost barrier to using R!

3) Its extensive functionality

R is a very versatile language - we've talked about its use in stats and in graphing, but its use can be expanded to many different functions - from making websites, making maps using GIS data, analysing language… and even making these lectures and videos! For whatever task you have in mind, there is often a package available for download that does exactly that!

4) Its community

And the reason that the functionality of R is so extensive is the community that has been built around R. Individuals have come together to make "packages" that add to the functionality of R - and more are being developed every day!

Particularly for people just getting started out with R, its community is a huge benefit - due to its popularity, there are multiple forums that have pages and pages dedicated to solving R problems. We talked about this in the Getting Help lesson; these forums are great both for finding other people who have had the same problem as you, and posting your own new problems.