# A Mini-Course on R

Arnab Sinha

Microsoft Research

Email: arnabsinha@gmail.com

# What is R?

R is an open source (free) programming language and software environment for statistical computing and graphics. Personally, I like R as a scripting language for quick inspection and visualization of (not too big) data. ([Learn more](https://en.wikipedia.org/wiki/R_(programming_language)))

# Why learn R?

Data is ubiquitous the new oil of the digital economy and Harvard Business Review [article](https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1) (2012) claims that data scientist ([what is data science](https://datajobs.com/what-is-data-science) and [who is a data scientist](http://www-01.ibm.com/software/data/infosphere/data-scientist/) anyway?) is "the sexiest job of the 21st century". In the recent past, large number of data scientists are being employed in both start-ups and large corporations for mining insights from data. Still, we need 140,000-190,000 data scientists with deep analytical talent according to a [McKinsey report](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation) (2011). Further, with growth of other technologies such as Internet of Things (IoT), there is a surge of data scientists even outside the tech industry such as manufacturing. Hence it is safe to claim that data scientists will be in high demand in the next decade.

Today, data scientists combine several sub-disciplines of computer science, mathematics and statistics together in order to analyze data such as data mining, natural language processing, fundamental statistics, predictive analytics, machine learning, and distributed computing. However, before applying any sophisticated technique data scientists often like to explore the data first. To that end, the language R has become a de facto standard among the data scientists for quick design and testing of hypothesis.

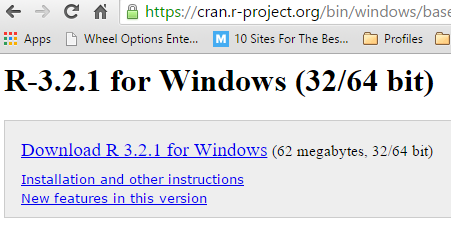
# Who is this mini-course for?

Anybody interested/curious about learning/practicing data science using open source tools. I would assume knowledge of at least one programming language and basic probability and statistics.

# How do I get started?

Download the appropriate R binaries from the links below and install in your machine. (Make sure you select the right binary depending on 32/64 bit machine. If unsure, try this in a command prompt window or the Start > Run box: “control /name microsoft.system”.)

* <http://cran.r-project.org/bin/windows/base/>

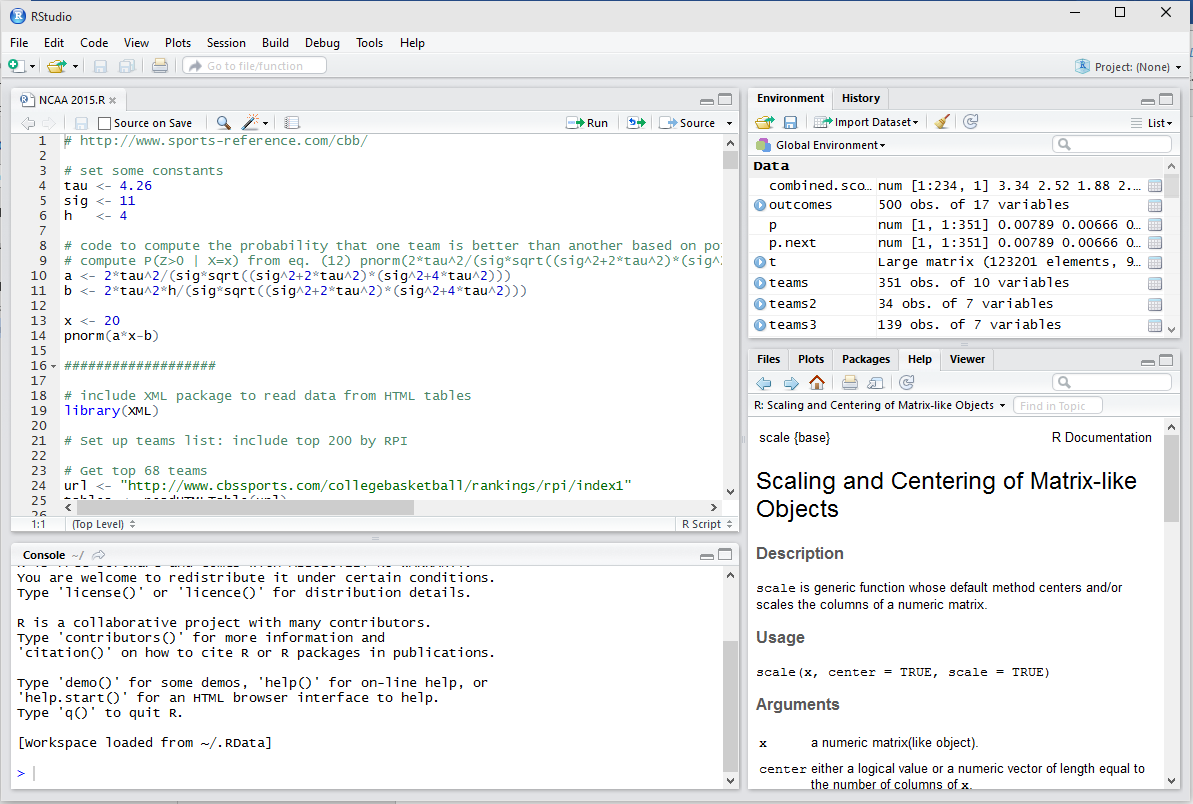


* <http://cran.r-project.org/bin/macosx/>
* <http://cran.r-project.org/bin/linux/>

Download and install the latest version of RStudio (This is the IDE, much like Eclipse to Java)

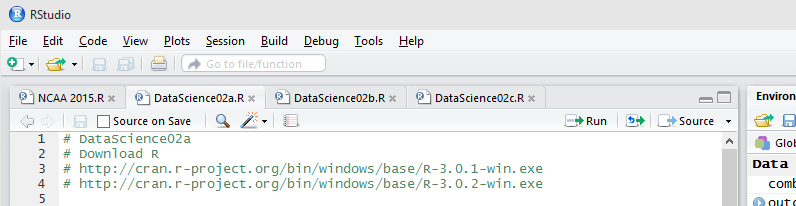
* <https://www.rstudio.com/products/rstudio/> (Follow the link saying “Download RStudio Desktop”)

A snapshot of RStudio.



# Install Packages in RStudio

Open and load the file “DataScience02a.R” and click on “Run” (as shown below) to execute the single lines of code and inspect the console output. Otherwise, “Ctrl + Enter” also has the same effect.

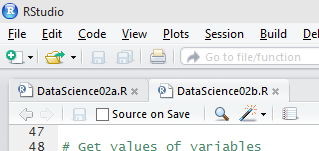


The following lines might install several packages. So, please be patient.

|  |
| --- |
| if(!require(installr)) {install.packages("installr", dep=TRUE, repos=reposURL)} else {"Latest R version is already Installed."} #load / install+load installr  library(installr) |

# Familiarize and Have Fun with R

Once you are done with “DataScience02a.R”, open “DataScience02b.R” in a separate tab (as shown below). Start executing that file and see if it makes sense. The file is self-explanatory with the comments.



If you are unable to figure out yourself, do not worry. We’ll go over them in person in the first lecture. But make sure that you install R and RStudio before the first lecture.

# Acknowledgements

Thanks to Ernst Henle and Richard Sharp III for sharing the contents for this mini course with me. Also, my sincere thanks to my friend and colleague Arjun Vijayakumar (Microsoft) who agreed to become my first student and testing out all the links and making sure the instructions make sense.

## Enjoy your vacation!