**Programming Training Resources**

Author: Xianbin (Eric) Cheng

# Introduction

This document is mix of official tutorials and personal suggestions and is only intended to cover some basic concepts in specific programming languages for people with limited or no programming background. Although there is no specific code in this document, it teaches you the basic terminologies in programming so that you will learn to google about coding more effectively. And yes, Google is your friend. You only take a tutorial when you don’t know what to google.

Table of Contents

[Introduction 1](#_Toc10152331)

[R 3](#_Toc10152332)

[Objective 3](#_Toc10152333)

[Before you start 3](#_Toc10152334)

[Tutorials 3](#_Toc10152335)

[Useful tools and packages 4](#_Toc10152336)

[Practice dataset 4](#_Toc10152337)

[Python 4](#_Toc10152338)

[Objective 4](#_Toc10152339)

[Before you start 5](#_Toc10152340)

[Tutorials 5](#_Toc10152341)

[Access 5](#_Toc10152342)

[Objective 5](#_Toc10152343)

[Before you start 5](#_Toc10152344)

[Tutorials 5](#_Toc10152345)

[Git and GitHub 6](#_Toc10152346)

[Objective 6](#_Toc10152347)

[Before you start 6](#_Toc10152348)

[Tutorials 6](#_Toc10152349)

[Bash 7](#_Toc10152350)

[Objective 7](#_Toc10152351)

[Before you start 7](#_Toc10152352)

[Tutorials 7](#_Toc10152353)

[Versions 7](#_Toc10152354)

# R

## Objective

1. To learn the basics of R
2. To learn the following packages: *dplyr*, *ggplot2*
3. To learn to read, clean, analyze, visualize, and output data

## Before you start

1. Print out the cheat sheets for base R, *dplyr*, *ggplot2*. They will make your life much easier.

* <https://www.rstudio.com/resources/cheatsheets/>

## Tutorials

The following tutorials are ordered by the level of difficulty. If you are a beginner of R, I would highly recommend you take the **tutorials 1, 2** and practice a lot before moving on to **tutorials 3, 4.** If you want to learn advanced techniques of R programming, **tutorial 5** would be an option. Eventually, **tutorial 6** is for those who want to use R as not only a data analysis tool, but also a programming tool.

1. **Dataquest: Introduction to Programming in R**

* This interactive tutorial teaches you the basics of R, including arithmetic calculation, vector, matrix, list, data frame.
  + <https://app.dataquest.io/course/intro-to-r>

1. **Lynda.com: Learning R**

* This complements the first tutorial, adding contents on basic statistics.
* <https://www.lynda.com/R-tutorials/Up-Running-R/120612-2.html?srchtrk=index%3a2%0alinktypeid%3a2%0aq%3aR%0apage%3a1%0as%3arelevance%0asa%3atrue%0aproducttypeid%3a2>

1. **Lynda.com: Learning the R *Tidyverse***

* This tutorial teaches you how to clean data using the package *Tidyverse*. It is important to go over **section 1 to 6**. Section 7 is optional.
* <https://www.lynda.com/R-tutorials/Learning-R-Tidyverse/586672-2.html?srchtrk=index%3a2%0alinktypeid%3a2%0aq%3aR%0apage%3a1%0as%3arelevance%0asa%3atrue%0aproducttypeid%3a2>

1. **R Studio Primers: Data Visualization Basics**

* This tutorial helps you learn the package for data visualization: *ggplot2*, which is essential for our work.
  + <https://rstudio.cloud/learn/primers/1.1>

1. **Coursera: Advanced R Programming by Johns Hopkins University**
   * This 4-week online class offers swirl() tutorials on functions and functional programming. It also offers content on object-oriented programming, debugging, and profiling. You may find it useful to refer to the “**Advanced R**” book below for more detailed explanation.
   * <https://www.coursera.org/learn/advanced-r/home/welcome>
2. **Advanced R by Hadley Wickham**

* This online book is for advanced R users. If you want to learn the working mechanisms behind R, this is the perfect book.
* I highly recommend the **chapter 1 and 2: *Foundation* and *Functional Programming*.** You don’t have to read everything, just skim through the material and learn some basic concepts. The chapter *Foundation* will help you understand the structure of R in a systematic way. The chapter *Functional Programming* will help you write code MUCH more efficiently by taking advantage of vectorization.
* <https://adv-r.hadley.nz/>

## Useful tools and packages

1. **R markdown**

* I HIGHLY recommend R markdown. This is very useful for presenting your code with instantaneous visual feedback. It is also very useful when you like to run things line by line and see what happens.
* Print the cheat sheet. It will help a lot.

1. **Package: *purrr***

* This package offers you a better way to do repetitive work. Basically, it gives you a cleaner version of the apply family functions. Again, for loops always work. It just takes time.

1. **Package: *caret***

* This is an important package if you need to do machine learning.
* Class recommendation: STAT 432 Basics of Statistical Learning. This class teaches you how to train machine learning models in R using caret. It also helps you learn R Markdown.

1. **R Shiny**

* This gives you a tool to make interactive apps using R.

## Practice dataset

1. R’s default datasets: “mtcars”, “iris”, etc. Type *library(help = "datasets")* in the R console for more information.
2. Real data: The Yao’s kernel dataset.

# Python

## Objective

1. To learn the basics of Python

## Before you start

1. If you just want to do common data cleaning, analysis and visualization, I would recommend R as it is more user friendly and easier to learn. However, if you need to interact with the operating system or you want to do advanced data science projects like machine learning and web scraping, Python may be better.

## Tutorials

1. **Dataquest: Python for data science: fundamentals**
   * This tutorial gives you the opportunity to learn Python interactively by actually typing code and working on a project, which I find more interesting to learn compared to watching videos.
   * <https://app.dataquest.io/course/python-for-data-science-fundamentals>

# Access

## Objective

1. To learn to use Access database as an end user.
2. To learn the concepts of tables, queries, forms, and reports.
3. To learn to extract data by queries.

## Before you start

1. Excel will be enough if your datasets are small and independent, because Excel is easier to use and more interactive. I would recommend you use Access only if your datasets are large (i.e. > 10 excel spread sheets) and somewhat inter-connected.

## Tutorials

1. **Lynda.com: Access essential training (Office 365)**
   * This tutorial gives a comprehensive introduction to Access. Skimming from **section 1 to section 5** should help you understand the basic concepts. The last few sections are optional.
   * <https://www.lynda.com/Access-tutorials/Access-Essential-Training-Office-365/737759-2.html?srchtrk=index%3a5%0alinktypeid%3a2%0aq%3aaccess%0apage%3a1%0as%3arelevance%0asa%3atrue%0aproducttypeid%3a2>

# Git and GitHub

## Objective

1. To learn the basics of Git.
2. To learn to manage local and remote repo.

## Before you start

1. Git is useful only when you want to keep track of different versions of text files (e.g. Word documents) or code.
2. Prior experience with Bash will be tremendously helpful.

## Tutorials

1. **Git-it**

* This tutorial is very visual and teaches you how to create a local repository, add things to the staging area, commit changes, and push the commits to the remote repository, which is GitHub.
  + <https://github.com/jlord/git-it-electron#what-to-install>

1. **GitHub Learning Lab**
   * This online tutorial offers an interactive program to teach you how to use GitHub
   * <https://lab.github.com/>

# Bash

## Objective

1. To learn the basic concepts of Bash
2. To be able to use Bash to run Git

## Before you start

1. Bash is very useful if you need to deal with bioinformatics.
2. Bash can be harder to learn as it is less interactive and more like hardcore programming.

## Tutorials

1. **Lynda.com: Learning Bash Scripting**
   * This tutorial offers a relatively comprehensive introduction to Bash. Reading through **section 1** should be enough if you are only using Bash for Git. The remaining sections are useful if you need to run repetitive scripts on the Linux system.
   * <https://www.lynda.com/Bash-tutorials/Learning-Bash-Scripting/142989-2.html>

# Versions

|  |  |
| --- | --- |
| Date | Description |
| 1/17/2019 | Created the first version |
| 5/30/2019 | Added Python section; replaced git tutorial with Git-it; replaced datacamp with Dataquest in R; removed Swirl() for introduction to R; rearranged sequence of tutorials for R; Updated ggplot2 tutorial link |