

R Basics Notes

Lesson 1:

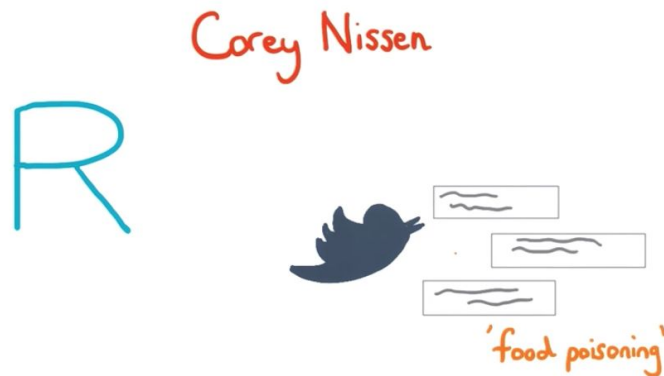
This lesson comes from the 'Data Analysis with R' course, which is part of the Data Analysis Nanodegree.

It is included in this orientation for students enrolling in the 'Data Visualization' course.

Data Visualization uses 'R' throughout the course.

This lesson tells students how to download and get started with the 'R' programming language.

Lesson 2:



- Corey Nissen helped citizens of Chicago by searching for quotes like 'food poisoning' to help the healthcare system.
 - <http://blog.revolutionanalytics.com/2013/08/foodborne-chicago.html>
 - <http://blog.yhat.com/posts/roc-curves.html>
-

Lesson 3:

Video: Why R

To install ggplot2 and RColorBrewer, you can run the following lines of code in RStudio.

```
install.packages('ggplot2', dependencies = TRUE)
library(ggplot2)
install.packages('RColorBrewer', dependencies = TRUE)
library(RColorBrewer)
```

R

■ the leading tool



■ many packages

■ active community



Lesson 5:

- Download from the following links:

- <https://cran.rstudio.com/>
- <https://www.rstudio.com/>

Video: Install RStudio On A Mac

Download and install the R programming language **FIRST** at <http://cran.rstudio.com>.

After you install R, you can download and install RStudio from <http://www.rstudio.com>.

There have been a few reports of the latest version of Rstudio failing to initialize. If you encounter these issue you can try using an older stable version:

<https://s3.amazonaws.com/rstudio-dailybuilds/RStudio-0.98.944.dmg>

If you have trouble installing RStudio, please google your problem first to get help. After that, we recommend posting a discussion. Be as specific as possible with your problem and include the current version of your operating system.

We do not recommend installing R and RStudio using Homebrew or MacPorts.

Lesson 8:

- Guide of how to use R.
 - <http://www.computerworld.com/article/2497143/business-intelligence/business-intelligence-beginner-s-guide-to-r-introduction.html>

Quiz: RStudio Layout

All four panels are configurable. The layout discussed is the default layout for RStudio.

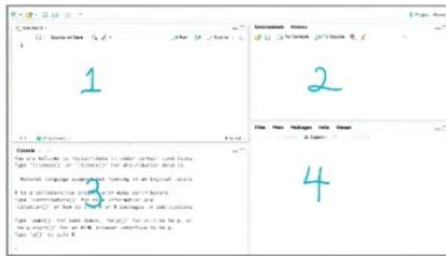
You can change the default layout by going to Tools -> Options...

This quiz assumes the default layout.

[A Beginner's Guide to R: Introduction](#) by Sharon Machlis

Quick Guide to R Layout by Oscar Torres-Reyna.

Quiz clarification: "Clear the workspace" means removing all variables from the current working memory.



- 1 R scripts/files
- 2 Environment/History
- 3 Console
- 4 Files/Plots/Packages/Help

- | | |
|--------------------------------------|---|
| 1 Save on R script | 1 Run a bunch of commands from a file |
| 2 Review the log of commands entered | 4 Look at a plot |
| 4 Read help documentation | 2 See a list of objects in memory |
| 2/3 Clear the workspace | 3 Read the results from functions or calculations |

Lesson 9:

- Swirl Courses to teach you R
 - https://github.com/swirldev/swirl_courses

Quiz: Demystifying R

We recommend anyone new to R and RStudio to try **Swirl** (statistics with interactive R learning). Swirl is a software package for the R statistical programming language. Its purpose is to teach statistics and R commands interactively.

Type the following commands in the **Console**, pressing Enter or Return after each line:

```
install.packages("swirl")
library(swirl)
swirl()
```

Note that the `>` symbol at the beginning of the line is R's prompt for you type something into the console. We include it here so you know that the above commands are to be typed into the console and not elsewhere. The part you type begins after `>`.

Lesson 10:

- Helpful Websites to problem solve with R:
 - <http://www.statmethods.net/>
 - <http://www.cookbook-r.com/>
 - <https://www.r-bloggers.com/>
 - <http://stackoverflow.com/tags/r/info>
 - <http://stackoverflow.com/questions/tagged/r-faq%20>
 - <https://google.github.io/styleguide/Rguide.xml>

Lesson 11:

- Loading Data in R:
 - <http://flowingdata.com/2015/02/18/loading-data-and-basic-formatting-in-r/>
 - Subsetting Data:
 - <http://www.statmethods.net/management/subset.html>
-

Lesson 12:

- An **R Markdown document (RMD)** is a simple formatting syntax for authoring web pages.

There is a difference between using the hash or pound symbol `#` inside and outside of a `{r}` block of code.

```
```{r}
The hash or pound symbol inside the r block creates
a comment. These three lines of are not code and cannot be # executed.
x <- [1:10]
mean(x)
```
```

If you use the `#` symbol outside of an `{r}` block of code, you can create headers of text.

Markdown Tutorial: You can also use Markdown to format your posts in the discussions.

You'll need to install and load the package `knitr` in order to use the **KNIT HTML** button . Run the following commands in the RStudio console to install and load knitr.

```
install.packages('knitr', dependencies = T)
library(knitr)
```

Download the file, [demystifyingR2.Rmd](#).

- Markdown Tutorial
 - <https://www.youtube.com/watch?v=6A5EpgqDOdk>
-

Lesson 13:

- R Data Types:
 - <http://www.statmethods.net/input/datatypes.html>
- Who is reddit?
 - <https://redditblog.com/2011/09/12/who-in-the-world-is-reddit-results-are-in/>

Lesson 14:

To run the `qplot` function you must have the `ggplot2` library installed and loaded. You can do that by running these two lines of code.

```
install.packages('ggplot2', dependencies = T)
library(ggplot2)
```

For those on Mac OS X 10.9 (or more recent OS X updates), the above line of code may display an error in the Console and fail to load `ggplot2`. This is due to a recent software update deleting the X11 symlink. You will need to reinstall XQuartz from <http://xquartz.macosforge.org/> and then run the following lines of code:

```
install.packages('devtools', dependencies = T)
library(devtools)
install_version("colorspace", "1.2-4")
```

You will then be able to load `ggplot2` in RStudio. See the following stackoverflow links for reference [here](#) and [here](#).

- Links from the above image:
 - <http://stackoverflow.com/questions/28984243/ggplot2-sourcing-error-x11-library-is-missing>
 - <http://stackoverflow.com/questions/28952128/loading-ggplot2-colorspace-actually-o-pens-up-x11>
 - <https://www.xquartz.org/>
- **Ordered Factors** are variables that can be naturally grouped within levels.
 - E.g. `age.range` (0-18, 18-28) or `income.range` (0-£20k, £21k - 31K)

Lesson 15:

- **Set and Order Factor Variables:**
 - http://statistics.ats.ucla.edu/stat/r/modules/factor_variables.htm

Lesson 16:

Video: Data Munging

Data Wrangling with MongoDB: Data Manipulation and Retrieval

Read this **fantastic tutorial** on loading data into R by Nathan Yau.

Tidy Data by Hadley Wickham

Tidy Data Presentation by Hadley Wickham (June 2012)