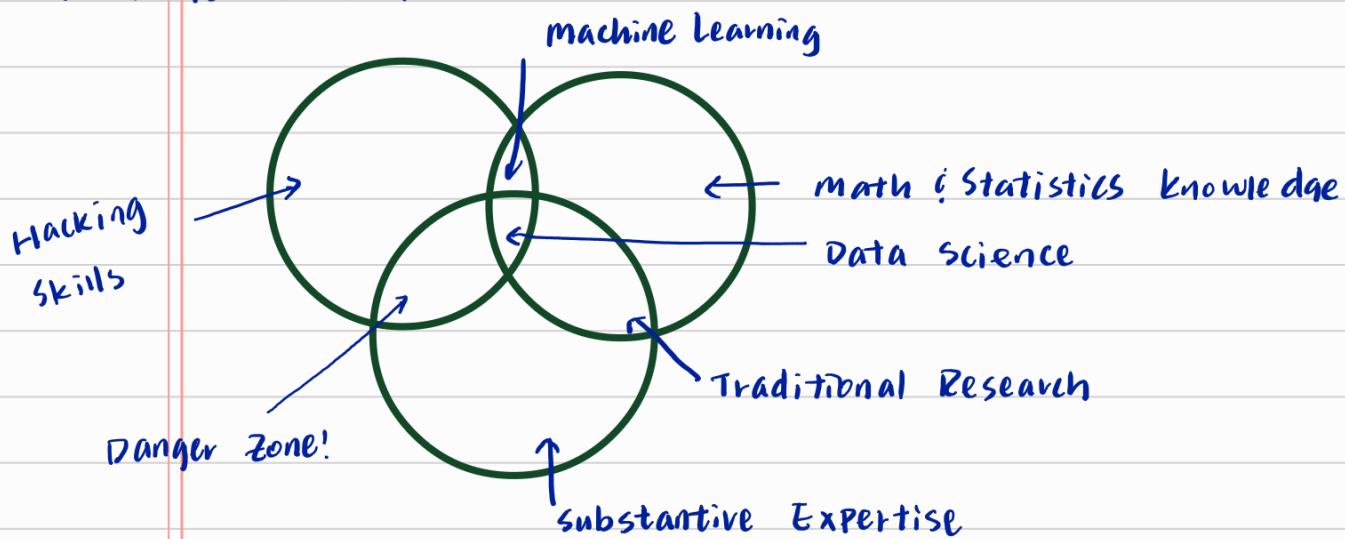


Coursera: Data Science: Foundations using R Specialization

Course 1: The Data Scientist's Toolbox

what is Big Data? volume, Velocity, Variety

What is Data Science?



Data: A set of values of qualitative or quantitative variables

In statistics, the population

You are trying to discover sth about

measurements or characteristics of an item

measurements or information

about qualities

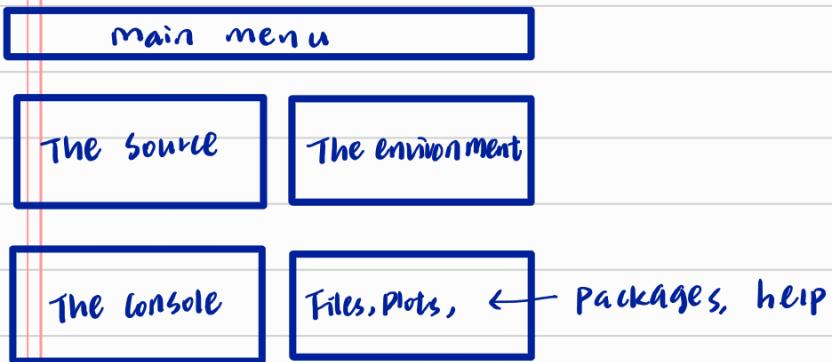
measurements or info.

about quantities or numerical items.

Data is of secondary importance: data is important, but it is secondary to your question. A good data scientist asks questions first and seek out relevant data second.

CRAN: Comprehensive R Archive Network

R studio



install Pkg: `install.packages(" ")`

Load Pkg: `library()`

What Pkgs are installed? `installed.packages()` or `library()`

Updating Pkgs: `old.packages() → update.packages()`



`install.packages(" ")`

Unloading Pkgs: `detach()`

`detach(" ", unload = TRUE)`

Uninstall Pkgs: `remove.packages(" ")`

Help. ? function

OR vignettes → extended help files. most of time include detailed examples.

Example: `browseVignettes("ggplot2")`

Project : Create new Project . Note: You can run multiple Project at the same time.

Structured vs unstructured Data



long tables, spreadsheet, etc.



Text files, audio, video etc.

GitHub:

① Repository: Project's folder / directory

(repo)

② Commit: Save your edits / changes made.

③ Push: Updating repo with your edits.

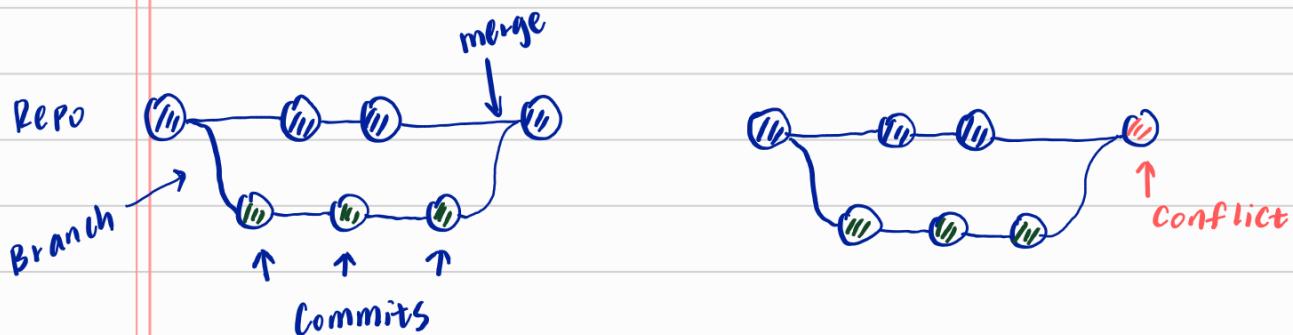
④ Pull: Updating your local version of the repo to current version

/// Staging: act of preparing a file for a commit. Allows you to separate out file changes into separate commits.

/// Branch: same file has two simultaneous copies.

/// Merge: independent edits of the same file are incorporated into a single, unified file. If both people made an edit to the same sentence that precludes one of the edits from being possible, we have a problem! Git recognizes this conflict and asks for user assistance in picking which edit to keep.

/// Conflict: multiple people make changes to the same file and Git is unable to merge the edits.



/// Clone: Making a copy of an existing Git Repo

/// Fork: A personal copy of a repo that you have taken from another person. The edits are logged on your repo, not theirs

Descriptive analysis: describe or summarize a set of data

Exploratory analysis: examine or explore the data and find relationships that weren't previously known.

Inferential Analysis: use a relatively small sample of data to infer to say sth. about the population at large.

Predictive Analysis: use current data to make predictions about future data.

Causal analysis: see what happens to one variable when we manipulate another variable - looking at the cause and effect of a relationship.

Mechanistic analysis: understand the exact changes in variables that lead to exact changes in other variables.

Experimental Design

④ Independent variable: x

④ Dependent Variable: y

〃 Sample size, confounder, control



An extraneous variable that may affect the relationship b/w the dependent and independent variables

〃 control group. treatment group, blind, placebo effect
group not manipulated

Rep vs. no Rep.