

What is Data Science?

Data Science Immersive

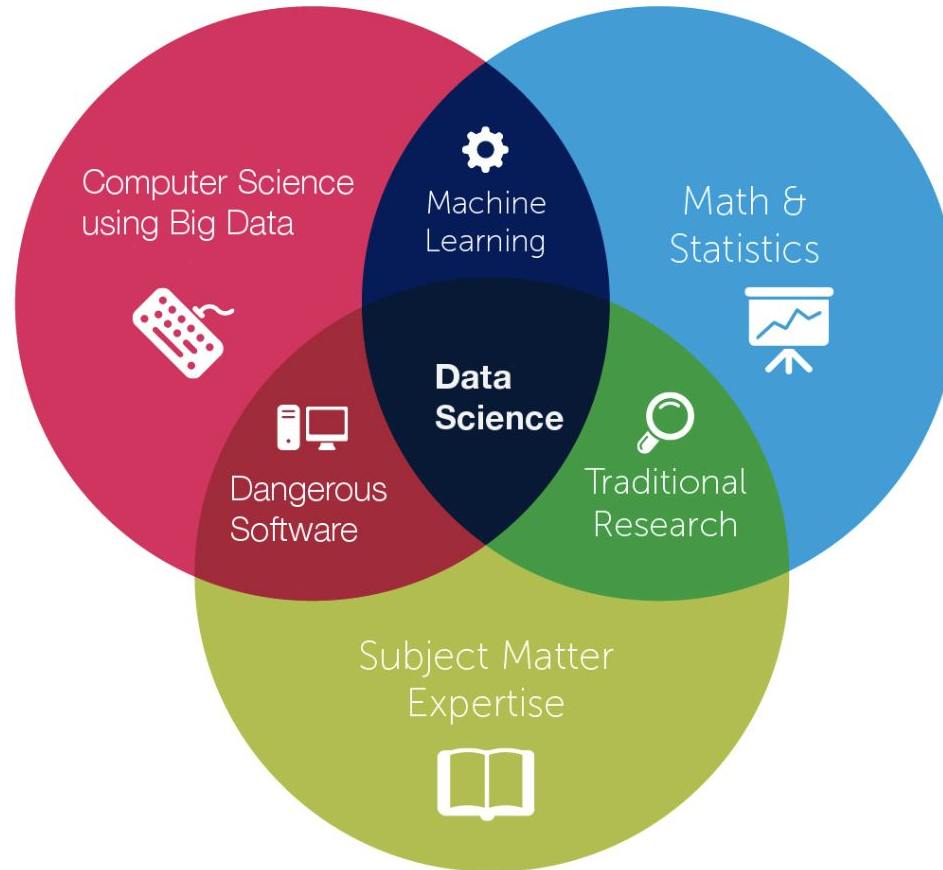
Overview of Day 1

- Introduction of Data Science
- Overview of Curriculum
- Setting up a data science environment
- Introduction to Git & Github

Introduction

- According to wikipedia, the definition of data science is “an **interdisciplinary** field that uses **scientific** methods, processes, algorithms and systems to extract **knowledge and insights from various forms**, both structured and unstructured”

What is data science?

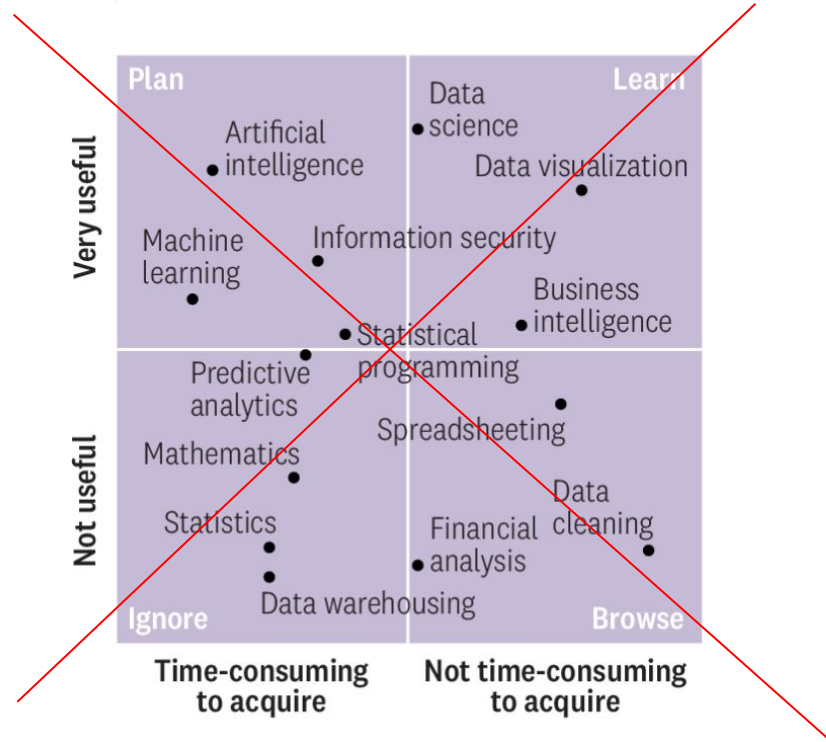


Introduction

“Data science is a "concept to unify **statistics**, **data analysis**, **machine learning** and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of **mathematics**, **statistics**, **information science**, and **computer science**.”

What is Data Science?

- What should you learn in data science?



What is Data Science?



Ok, but what kind of questions can DS answer?

- How many minutes, on average, is an American person exposed to advertisement?
- Given the items you have purchased in the past, what items are you more likely to purchase in the future?
- What makes a sound track successful?
- Is a given image mountain or glacier?
- How should the government allocate educational resources?
- Can you use NLP to predict whether two quora questions are duplicates?
- Based on the movies that you like, what are some of the similar movies you would also like?
- And many many more!

Activity

- Talk to your neighbor, and either pick one of the questions above, or come up with your own question that can be solved with data science tools, and lay out your plans or thought processes on solving that

Data Science Process

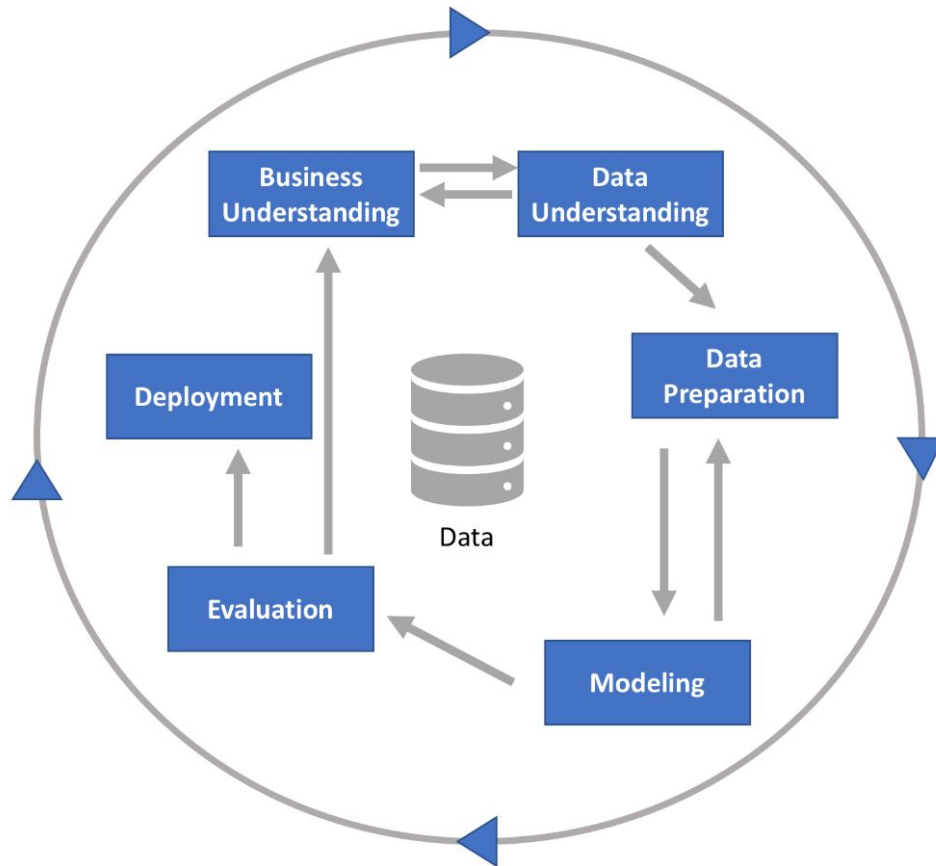
- What is the actual process of doing data science?
- How much time do data scientists spend on each of these components in the process?

Activity

- Break into small groups, talk about the problem or question that you discussed earlier, and tell us the steps it would take to execute your solutions

Data Science Process

- Crisp-DM



Overview of Your Flatiron Journey

- Learn.co as textbook
 - Readme's
 - Labs
- Lectures & presentations from instructor and coaches
- Code exercises
- Blogging
- Pair & Independent Projects
- Feelings Fridays
- Other community events

Module Overview

- Module 1 - Fundamentals of Python Programming

In this module, we will review and practice general programming fundamentals as well as data science specific programming, such as using packages like Numpy and Pandas. You will learn and get used to manipulate different data structures in Python, which lay the foundation for modeling data later in this course.

Module Overview

- **Module 2 - Data Engineering for Data Science**

In mod 2, we will learn (almost) everything you need to know about data acquisition and database manipulation. Topics such as webscraping, API, Object Oriented Programming will be covered.

Module Overview

- **Module 3 - Probabilities and Statistics for DS**

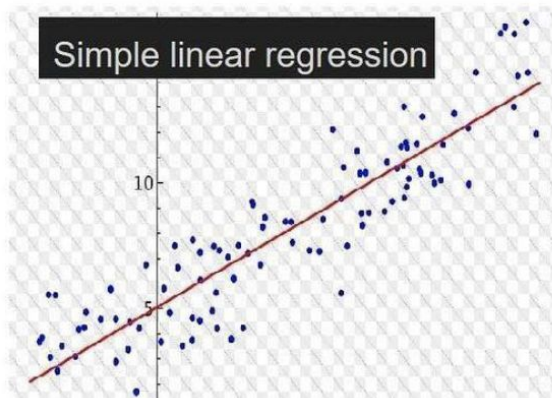
In this unit we will learn all about the the foundations for machine learning algorithms: probabilities and statistics. Topics include combinatorics, conditional probabilities, statistical distributions, and sampling.

Module Overview

- Module 4 - Statistical Modeling

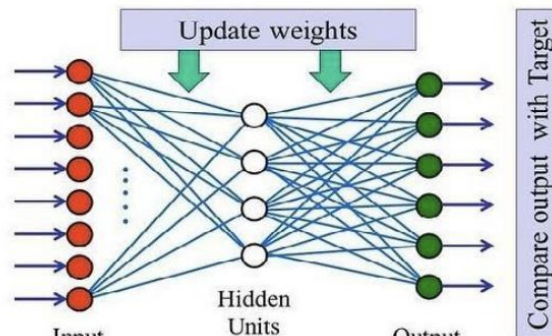
Linear models lay the foundation of statistical inference and prediction. We will dig into modeling and specifically focus on linear models such as linear regression, regularization, logistic regression, time series modeling, etc.

if you
don't love
me at my



then you
don't deserve
me at my

Artificial Neural Network



Module Overview

- **Module 5 - Machine Learning**

As great as linear models are, it is often not sufficient to account for the complicated behavior and relationship of variables. In this module, we will learn about nonlinear and nonparametric models such as decision trees, random forests, and boosting methods.

Module Overview

- Module 6 - Big Data & Advanced Topics in Machine Learning

You will learn advanced topics in the data science world such as recommendation engines, neural networks, Markov Chain Monte Carlo, and Natural Language Processing. You will be able to build a big data project using the techniques you learned from this chapter.

Setting Up a Data Science Environment

- Download Python 3.7 & Anaconda
- Slack
- Terminal 101
- Navigate Github & Git

Download Python 3.7

Anaconda 2018.12 For macOS Installer

Python 3.7 version *

↓ Download

[64-Bit Graphical Installer \(652.7 MB\)](#) ?

[64-Bit Command-Line Installer \(557 MB\)](#) ?

Python 2.7 version *

↓ Download

[64-Bit Graphical Installer \(640.7 MB\)](#) ?

[64-Bit Command-Line Installer \(547 MB\)](#) ?

Navigate Anaconda

- Opening Jupyter Notebook from Anaconda

Terminal 101

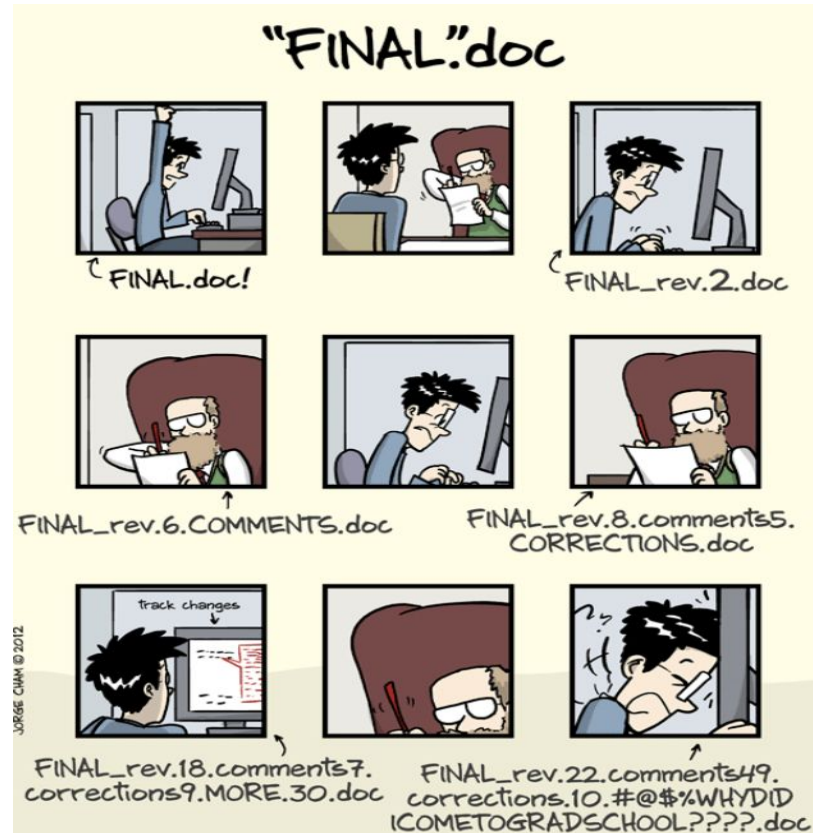
- Basic terminal command:

- pwd → print working directory
- mkdir → make a directory
- ls → list of documents in a directory
- cd → change directory
- cd .. → goes back to previous directory
- cd ~ → goes back to root directory
- mv → move stuff around; change name
- rm → remove stuff
 - rm -r directory
- Opening jupyter notebook from terminal by typing command “Jupyter Notebook”

- Python commands from terminal

- Pip install packages
- lpython vs python

Navigating Git & Github



Navigating Git & Github

- What is Git?
 - Git is an open source distributed version control tool for tracking changes in code.
- What is Github?
 - Github is a hosting services for git repositories (“repos”)



Navigating Git & Github

- Navigating Github

Example - <https://github.com/datasciencemasters/go>

Fork

- “A *fork* is a copy of a repository. Forking a repository allows you to freely experiment with changes without affecting the original project.”

Navigating Git & Github

- Setting up Github & Git
 - Github: to set up github, simply register on the website with your personal email
 - Git: to set up git, go to <https://help.github.com/articles/set-up-git/>
 - To configure Git
 - `git config --global user.name "firstname lastname"`
 - `git config --global user.email "your email here"`


Navigating Git & Github

- Create a new repo
- Cloning the repo to your local machine
- Making changes to the repo & push the changes
- Push the changes from

Git Commands


- create a file: **touch** [filename]
- Cloning a repo from github: **git clone** [url]
- check the status: **git status**
- track and stage a single file: **git add** [filename]
- track and stage all files: **git add .**
- commit with a message: **git commit -m "description of commit"**
- Go back to a previous commit **git revert**
- view the log: **git log --oneline**
- push changes: **git push** [remotename] [branchname]
- adding an upstream **git remote add** upstream [URL]
- Pull changes from remote: **git pull** [remotename] [branchname]

Git Commands



Apr 16 master

- line break after IPB training to render header on next li
- added todo for and
- added may 18 and 19 intro days
- fixed date for wed may 27. was 26
- Header changes for may 20, fix date and topic
- pain in the ass
- I'm comitted
- curriculum
- blah magitude
- blah 10
- more blah
- Blah 8
- blah 7
- blah 7
- blah 6
- blah 5
- blah 4
- blah blah blah
- bash topics
- blah
- basic outline for 2 days
- blah blah

 **D@niel Chen**
@chendaniely

how #not to write #git #commit messages -.-"