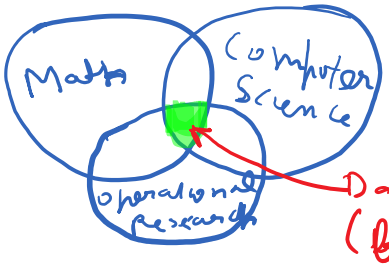
Tuesday, December 26, 2017 11:19 PM

We talked about the following things in the class today.

Data Science Venn Diagram



Topics for data sets of interest

- financial data (like stock market)
- ??? (Sanjit please remind us)
- Astronomy data (lets say maps of the sky)
- FDA data (food borne diseases)

I will send out some useful links related to these data sets (See Slack Channel).

API (Application Programming Interface)

Think of this as someone creating a set of "functions" (reusable code) which you can use in your code without having to deal with the implementation details. All you need to know is how to call the function/API i.e. what arguments/parameters to pass as inputs and what return values to expect.

R/Python and all other Programming languages

are accompanied with a whole host of libraries that provide APIs to do a lot of useful work (like downloading data from the internet).

As part of this course we would be using a lot of APIs. The first example of a library that we would be using would be "http", we will talk about it in the next class.

CSV (Comma Seperated Value) format

CSV format is one of the most commonly used formats for storing data (especially if we are not dealing with the Big Data).

It simply refers to data stored in a tabular format with column values being separated by "," (comma).

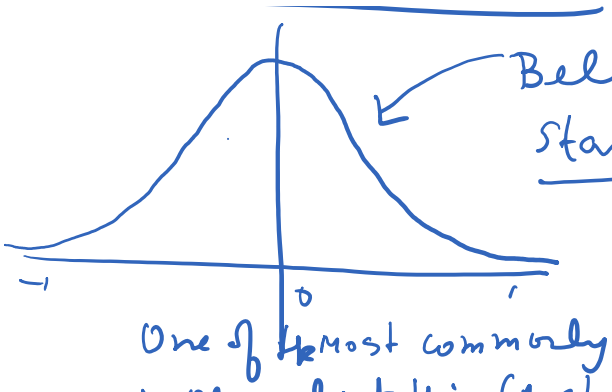
For example:

Name, Age, Place
John, 10, Clarksburg

[Note that each field is seperated by a comma each row is on a different line.]

We might also talk about another very commonly used format called JSON (Javascript Object Notation). More convenient for Programming rather than for human beings. (If you are interested we may talk about this some more.)

Normal distribution



Bell curve

Standard Normal

↳ Mean (average) = 0

↳ Standard deviation = 1

One of the most commonly found distributions. We will talk more about this (just a little bit more) in next class.