CS196

Data Science

Ben Congdon

Course Staff CS 199ACC
Data / Python Enthusiast

–

@benrcongdon

tyler kim

Course Staff CS 199ACC
Aesthetic Hippie

–

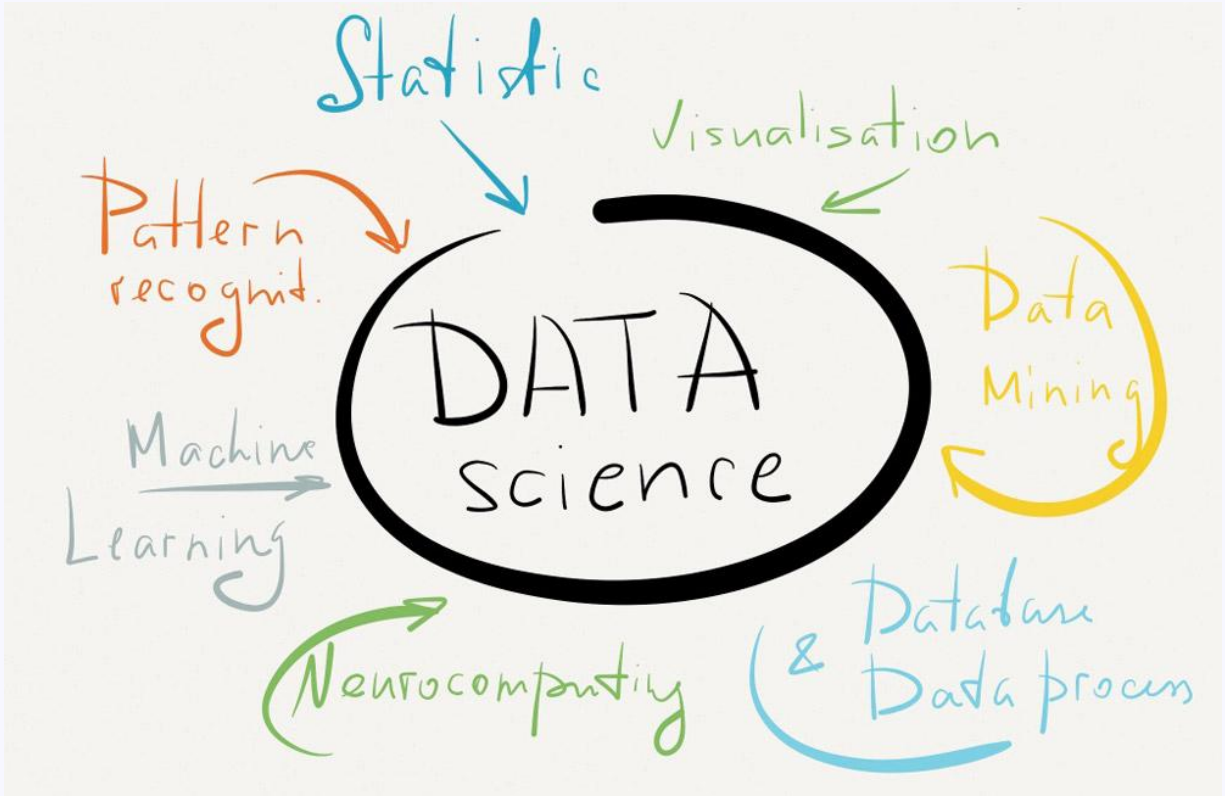
@tyler-thetyrant

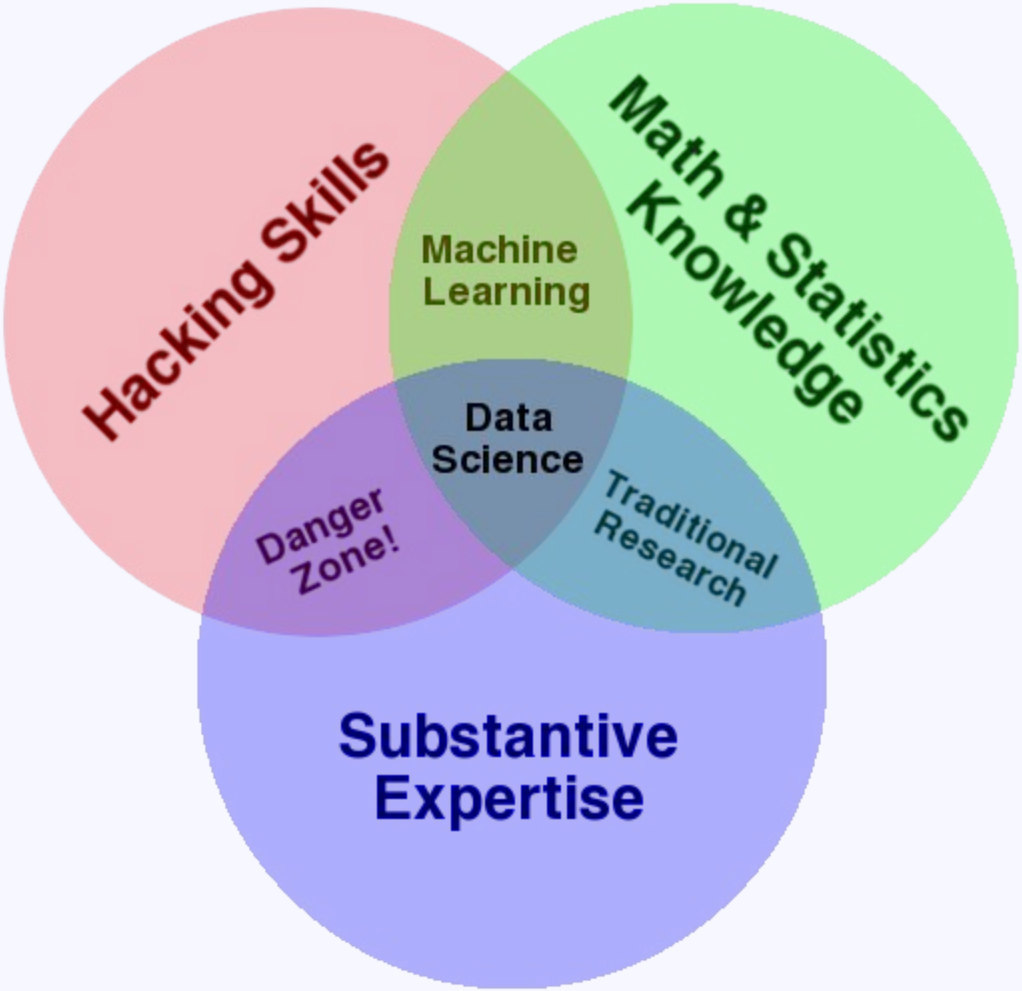
Josh Dunigan

Course Staff CS 199ACC
Not Aesthetic
Normal Human Being

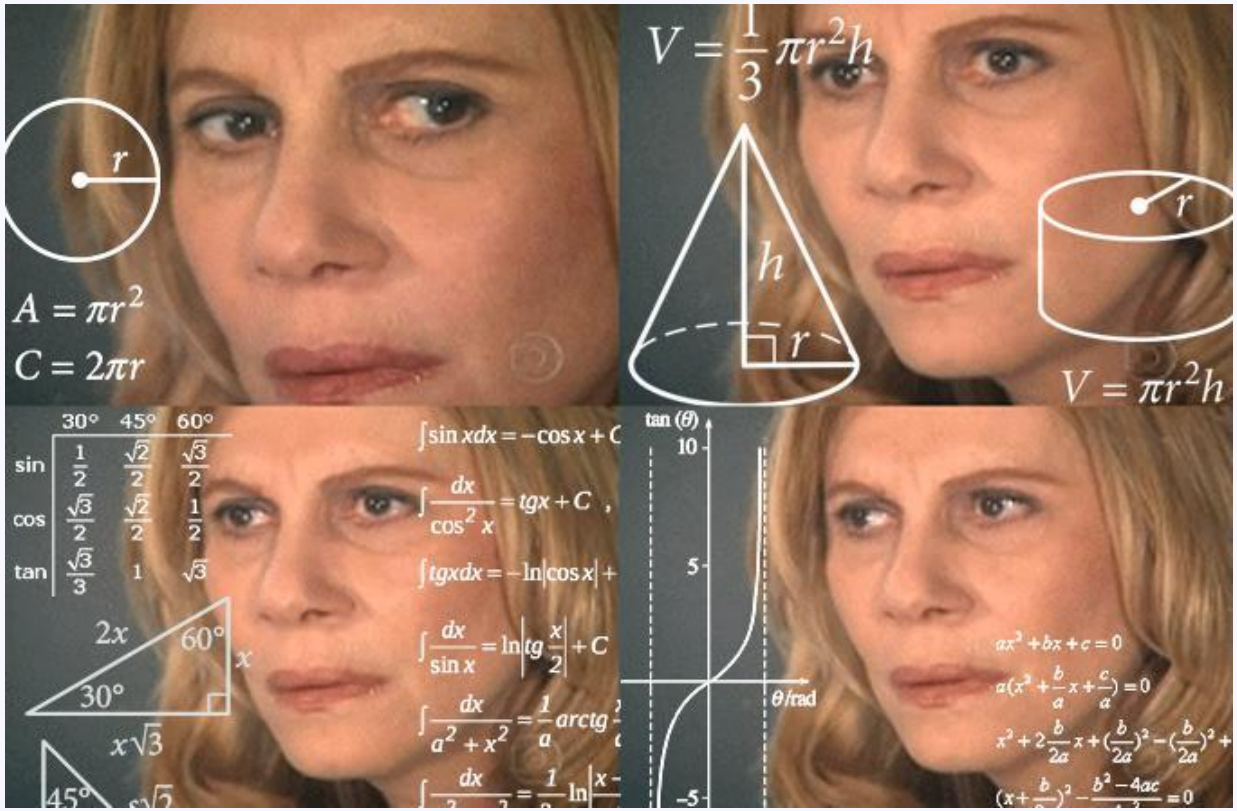
What is Data Science?

I mean, seriously, what is it?

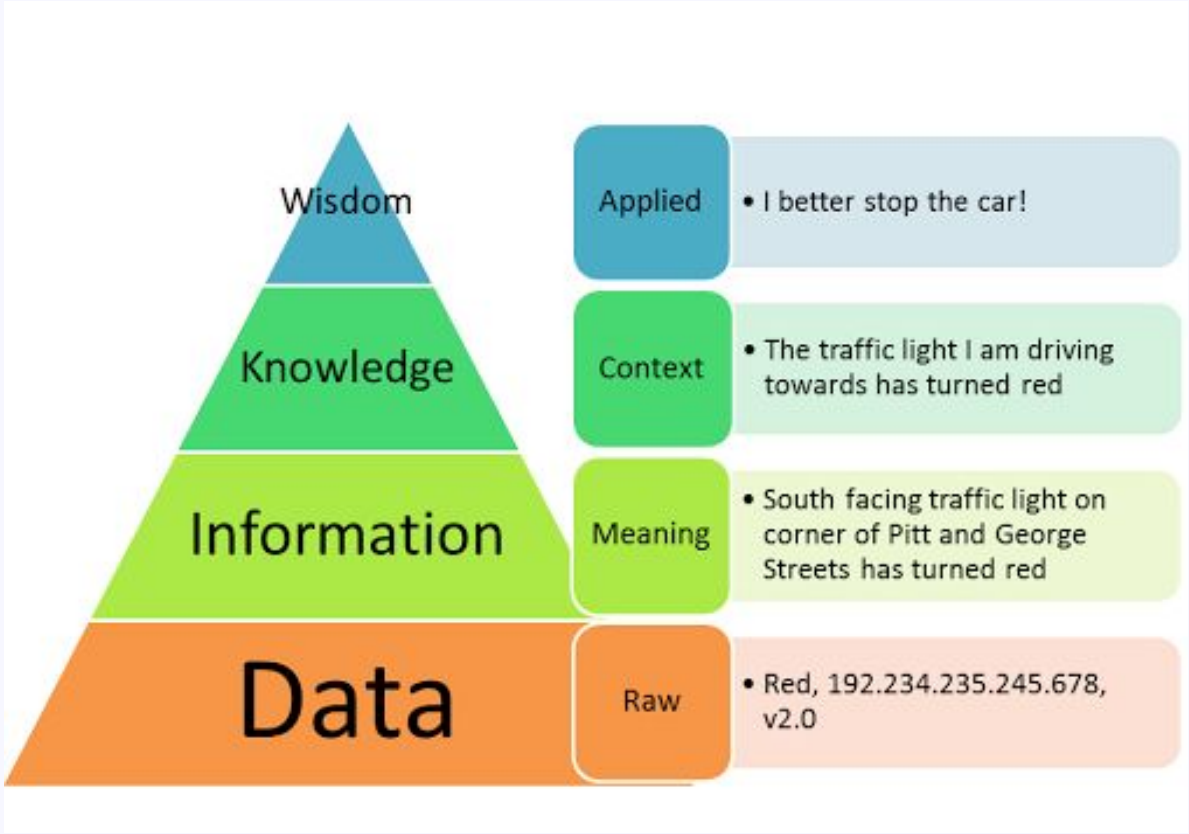

Statistic
Visualisation
Pattern recognit.
DATA
science
Data Mining
Machine Learning
Neurocomputing
& Database Data proces


Hacking Skills
Math & Statistics Knowledge
Substantive Expertise
Machine Learning
Data Science
Danger Zone!
Traditional Research

Why Data Science?



Data driven thinking is the future


Wisdom
Knowledge
Information
Data
Applied
• I better stop the car!
Context
• The traffic light I am driving towards has turned red
Meaning
• South facing traffic light on corner of Pitt and George Streets has turned red
Raw
• Red, 192.234.235.245.678, v2.0

Data Scientists?

What now?

(((Josh Wills)))
@josh_wills

Follow

Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

RETWEETS 1,486 LIKES 1,015

8:55 AM - 3 May 2012

1.5K 1K

Course Objectives

High-level:

- Know how to think in scale.
- Given a problem, know how to approach.

Low-level:

- Familiarity with the core tools of data science/analytics
- Enough theoretical background to get you started
- Several data science projects to get a feel for the breadth of the field

Mini Syllabus ~ 15% of 196 Grade

Grading: Attendance + Notebooks

Expectation: Attendance / Participation is Mandatory.

Also, be kind.

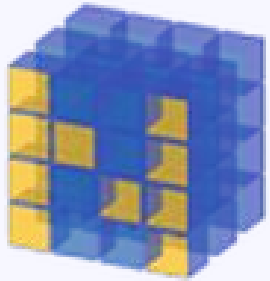
Course Plan

Week 2	Data Science Overview. Work environment setup.
Week 3	Basic Data Structures + Statistics for Data Science
Week 4	Numpy & Pandas
Week 5	Data Visualization
Week 6	Data Mining & Web Scraping
Week 7	Web APIs
Week 8	MapReduce

Course Plan

Week 9	Relational Databases and SQL for Data Science
Week 10	Intro to NoSQL Databases
Week 11	Intro to Machine Learning with Scikit Learn
Week 12	???
Week 13	???

This is an open-ended course. We'll fill the rest of topics based on your interests.



scikit learn
machine learning in Python

pandas
$y_i t = \beta' x_{it} + \mu_i + \epsilon_{it}$

python™

DASK

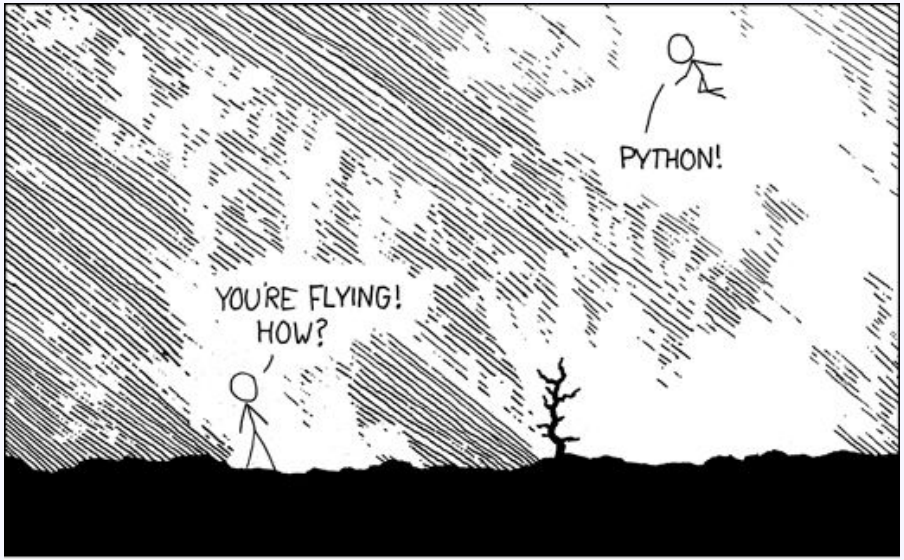
jupyter

CONDA

Hello, World

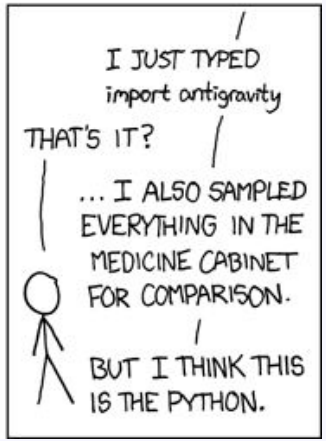
print(“Hello, World”)

Why Python™?

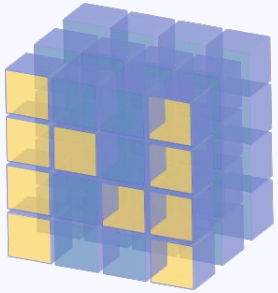

PYTHON!
YOU'RE FLYING! HOW?


I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
HELLO WORLD IS JUST print "Hello, world!"


I DUNNO... DYNAMIC TYPING? WHITESPACE?
COME JOIN US! PROGRAMMING IS FUN AGAIN! IT'S A WHOLE NEW WORLD UP HERE!
BUT HOW ARE YOU FLYING?


I JUST TYPED import antigravity
THAT'S IT?
... I ALSO SAMPLED EVERYTHING IN THE MEDICINE CABINET FOR COMPARISON.
BUT I THINK THIS IS THE PYTHON.

Base **Numpy** (2006)



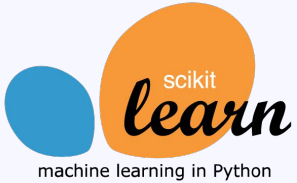
NumPy

"Prior to Python, I used Pearl (for a year) and then Matlab and shell scripts & Fortran & C/C++ libraries.
When I discovered Python, I really liked the language.... But, it was very nascent and lacked a lot of libraries. I felt like I could add value to the world by connecting low-level libraries to high-level usage in Python."

- Travis Oliphant, 2015,
  Creator of Numpy & Scipy

Era of Pydata

Motto: “Python as alternative to R”



2010: Machine Learning

2011: Labeled Data, Data Frame

2012: Package & Environment Manager

2012: Compute Environment

2015: Multi-language support


jupyter

Jupyter Notebooks

- Interactive, web-based data science platform
- Can be run on your laptop, or on a high-performance cluster
- Useful for exploring, transforming, and presenting data

Today's Goal

- Get everyone set up with Jupyter/Python
- Load into the example Jupyter Notebook
- Have fun exploring!


KEEP
CALM
AND
CONDA
INSTALL

OS-agnostic, system-level binary package manager

Install Miniconda3:

https://conda.io/miniconda.html

Add conda to your PATH

$ conda install jupyter notebook

$ git clone https://github.com/CS196Illinois/Data_Hacker space_FA17.git

Clone our Notebook for the week.

Questions? AUA