GENERAL ASSEMBLY

WELCOME TO GA

Please write your name on your whiteboard tent and introduce yourself to your new classmates

Wi-fi: GA-Guest
pw: yellowpencil

Gus Ostow

TODAY'S SCHEDULE

- Logistics
  - Meet the team
  - Class resources
- What is Data Science?
- Python Review

LOGISTICS

INSTRUCTIONAL TEAM

Gus Ostow

Lead Instructor

augustustostow@gmail.com

INSTRUCTIONAL TEAM

Stewart Knox

Instructional Assistant

Stewart.knox@gmail.com

STUDENT SERVICES

Matt Jones

Slack: @studentservicessf

Things he can do for you:

- Access to tools
- Feedback about the course
- Enrollment and finances
- Graduation certificates
- GA facilities
- Extra-curricular events
- Discounts for other courses

COURSE DETAILS

- Lead Instructor
  - Me :)
- Instructional Assistant
  - Stewart
- Class Schedule
  - January 16 - March 22
  - Tuesdays and Thursdays
  - 6:30 PM - 9:30 PM

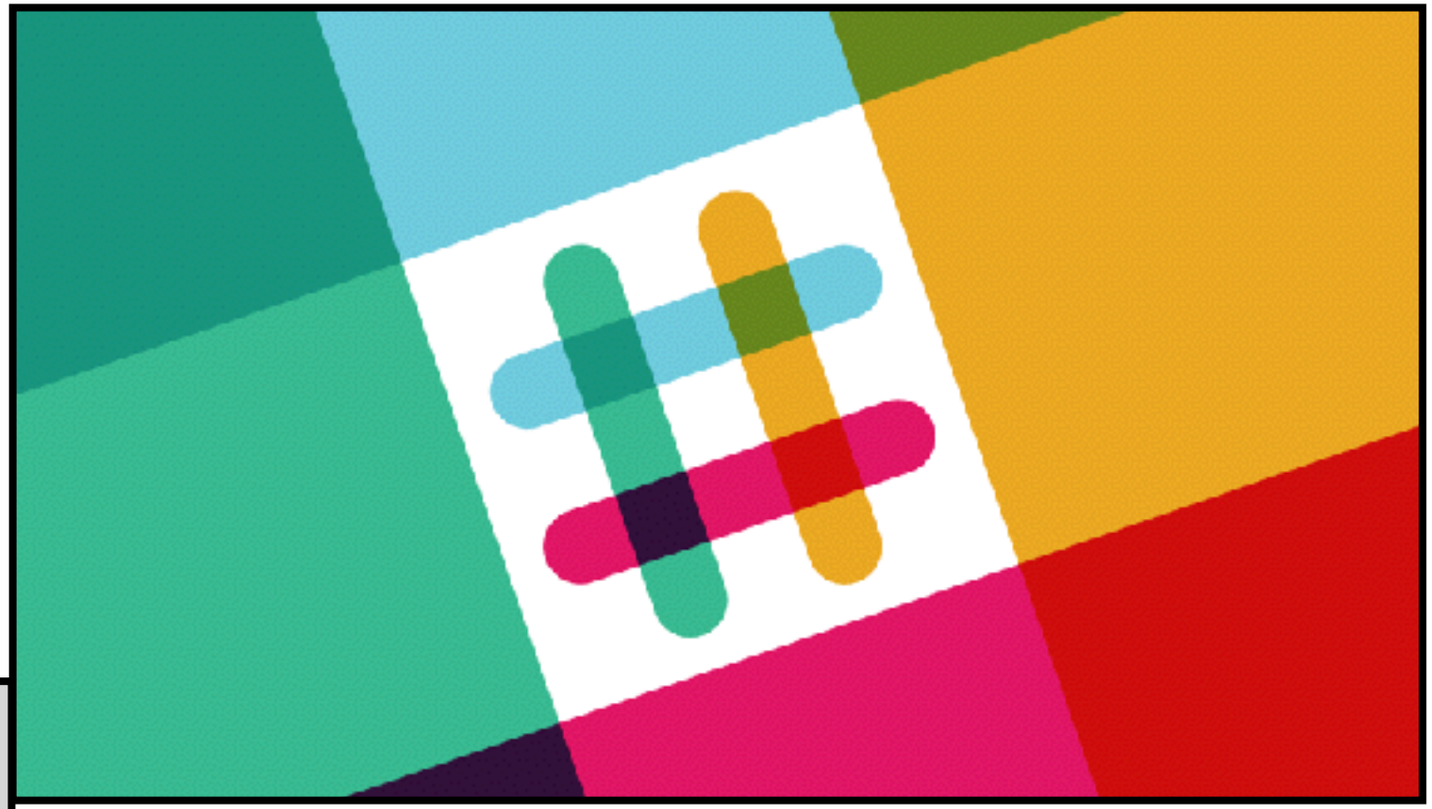
COURSE TOOLS

- Slack
  - https://dat-sf-42.slack.com/
- Github
  - https://github.com/ga-students/DS-SF-42/
- Exit Tickets
  - https://tinyurl.com/ds-sf-42

SLACK

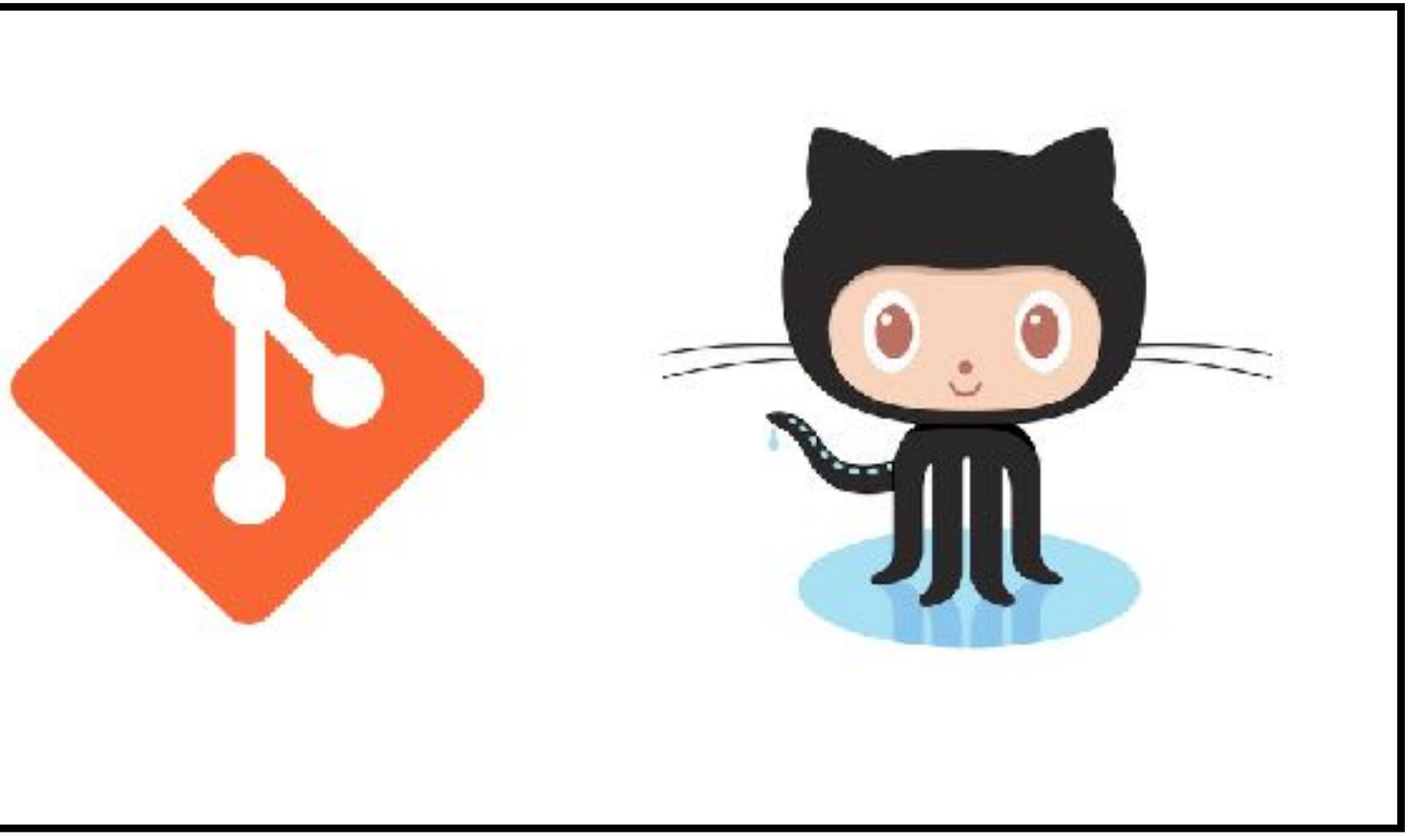
- All course communications with each other and instructors will happen here.
- Great for sharing code snippets during class
- Please don't make me write an email. Please.

- Take 3 minutes to install Slack



GITHUB

- “Dropbox for code” (loosely)
- How you will receive all class materials
- Backup (and show off) your work
- Learn how real teams collaborate in production



Git is important

EXIT TICKETS

At General Assembly we are data scientists; we need data on how we're doing

- **Give us valuable information:**
  - Your engagement
  - Your opinion on the success of the lesson
  - Remaining questions
  - Suggestions

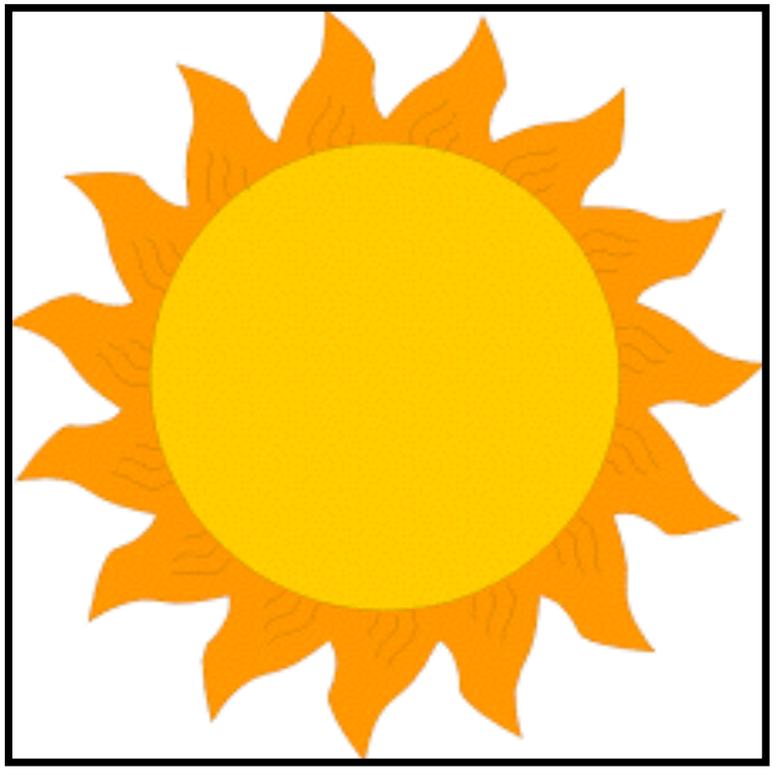
CLASSROOM CULTURE

- We should all agree to:
  - Set ambitious learning goals
  - Pursue a ~Growth Mindset~
  - Collaborate and connect
  - Be radically transparent

- **What would you like to add?**

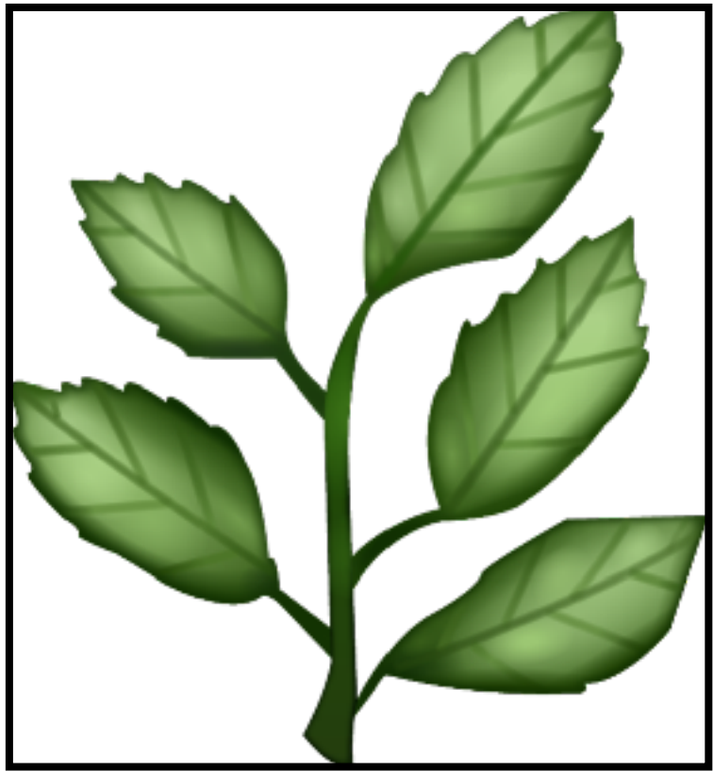
MY INSTRUCTIONAL GOALS

- Find a brisk, but manageable pace
- Facilitate an informal, open learning environment
- Make myself very available
- Feedback is a two-way street
  - To give it you early and often (and by request over Slack)
  - To take feedback well, myself. Everybody needs a ***~Growth Mindset~***!



CONTENT PHILOSOPHY

- Offer a variety of resources, to be accessible to diverse learning styles
- Balance theory with applications
- Maintain a high level of rigor, despite our time-constraints
- The deep learning happens during work on projects



YOUR KEYS TO SUCCESS

- Lots of effort
- Outside study and preparation
- Keep track of your goals, and progress toward them
- Work together on reviews and projects
- When you get frustrated take a walk outside
- Stewart

HOW TO USE STEWART

- During lecture:
  - If you have a question that nobody else will benefit from:
    - Slack him. Even though we are in the same room.
    - Move to where he is sitting if necessary
- During independent practice:
  - However you would normally communicate with a human
- Office Hours:
  - **Slack:** Monday 6-7PM **At GA:** Tuesday 5:30-6:30PM (directly before class)

A TYPICAL CLASS

- Pre-readings found on the Github wiki
- Class objectives (also on wiki)
- Previous class review

- Lecture and theory
- Guided-practice
- Misc activities
- Independent practice

- Class review
- Exit ticket
- Additional resources

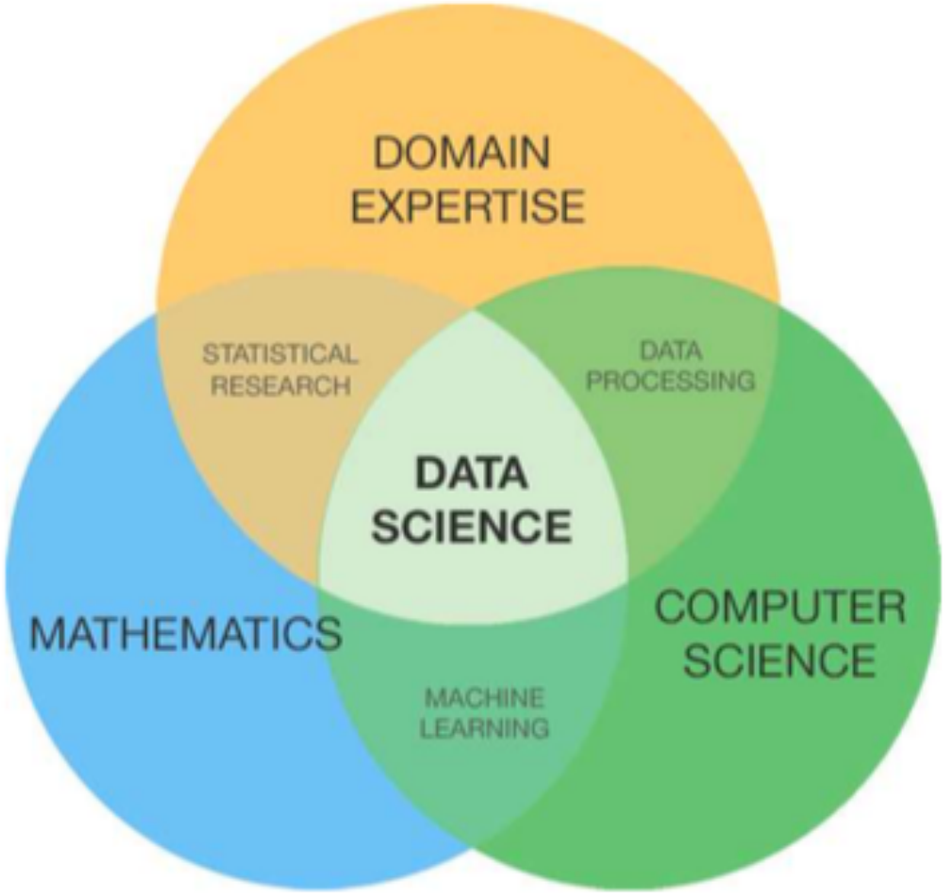
QUESTIONS?

WHAT IS DATA SCIENCE?

DATA SCIENCE IS EVERYWHERE

DATA SCIENTISTS MAKE DATA USEFUL

- Key skills:
  - Software engineering and scripting
    - Python, R, Java, C
  - Statistics and experimental design
  - Predictive modeling
  - (Big) data engineering
    - SQL, Spark, MapReduce
  - Business acumen and communication



Source: Data Science for the C-suite

EXAMPLE #1: CHURN PREDICTION

Problem:

- Churn is a ubiquitous problem for companies with reoccurring revenue business models
- Risk of cancelling a subscription severely limits customer lifetime value

Goal:

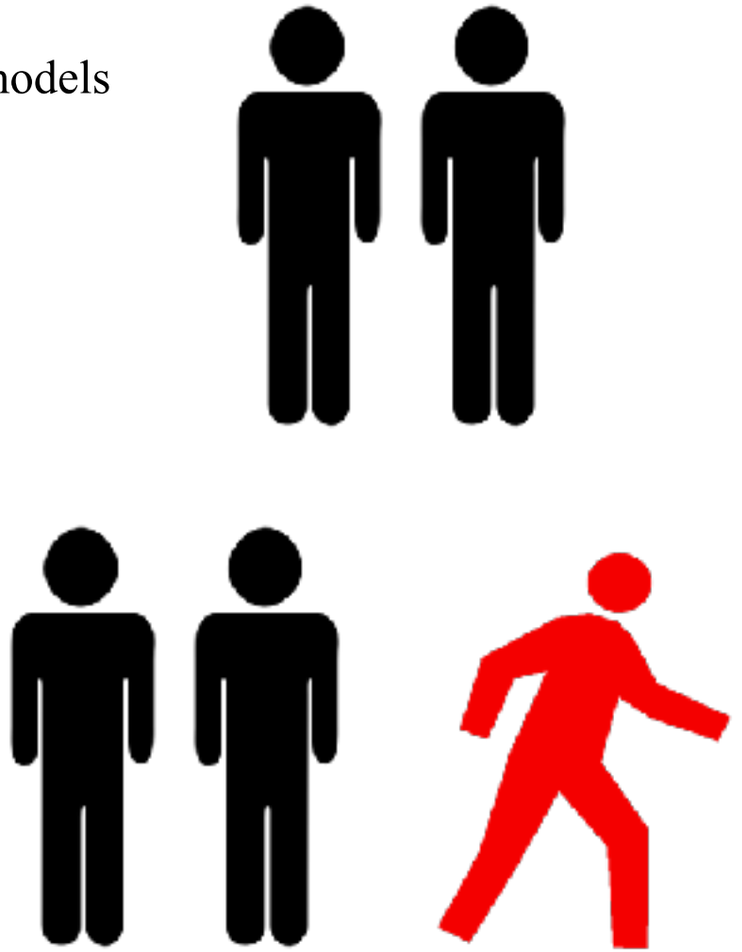
- Identify customers who are likely to cancel telecomm subscriptions

Data:

- Telco Customer Churn Dataset: customers who left, services subscribed to, account information, demographics

Impact:

- Interventions can be made to retain customers at a high-risk of churn, which are considerably cheaper than acquiring a new customer.



Source: https://blog.insightdatascience.com/deep-learning-for-disaster-recovery-45c8cd174d7a

EXAMPLE #2: DETECTING FLOODED ROADS

Problem:

- After natural disasters, flooded roads are deadly to motorists.
- Two-thirds of of U.S. deaths in flash floods occurred in vehicles.

Goal:

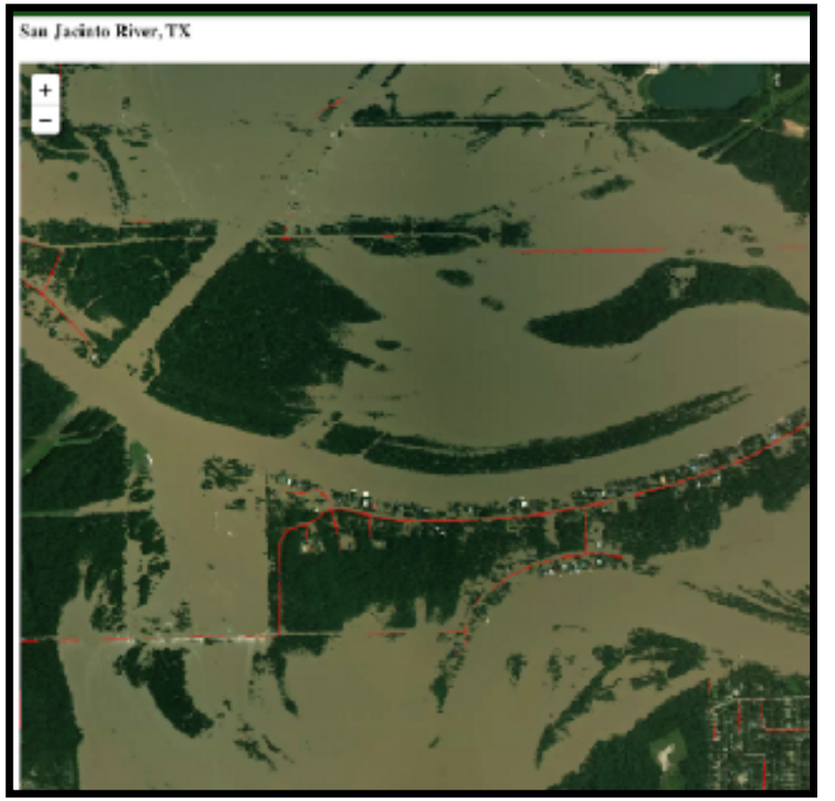
- Automatically mark flooded roads on an interactive map

Data:

- Pre-flood Mapbox satellite and streetmap tiles from Boston, NYC, Atlantic City, Miami, and New Orleans

Impact:

- Information on road anomalies could be instantly available after natural disasters



Source: https://blog.insightdatascience.com/deep-learning-for-disaster-recovery-45c8cd174d7a

MORE COMMON QUESTIONS ASKED IN DS

REGRESSION

- What will fourth quarter sales be?
- How many retweets will this post get?
- How much is this house worth?
- What should the steering angle be based on self-driving car video frames?

CLASSIFICATION

- Will this person default on their loan?
- What is in this image?
- Will this person click on the advertisement?
- Is the person in the fMRI viewing a happy, or scary video clip?

UNSUPERVISED

- Which Netflix subscribers like the same kind of movies?
- Which topics consistently appear in patent proposals
- What are the different types of coffee drinkers?
- How can you visualize the similarity between car models in a 2-d space?

OUR CURRICULUM

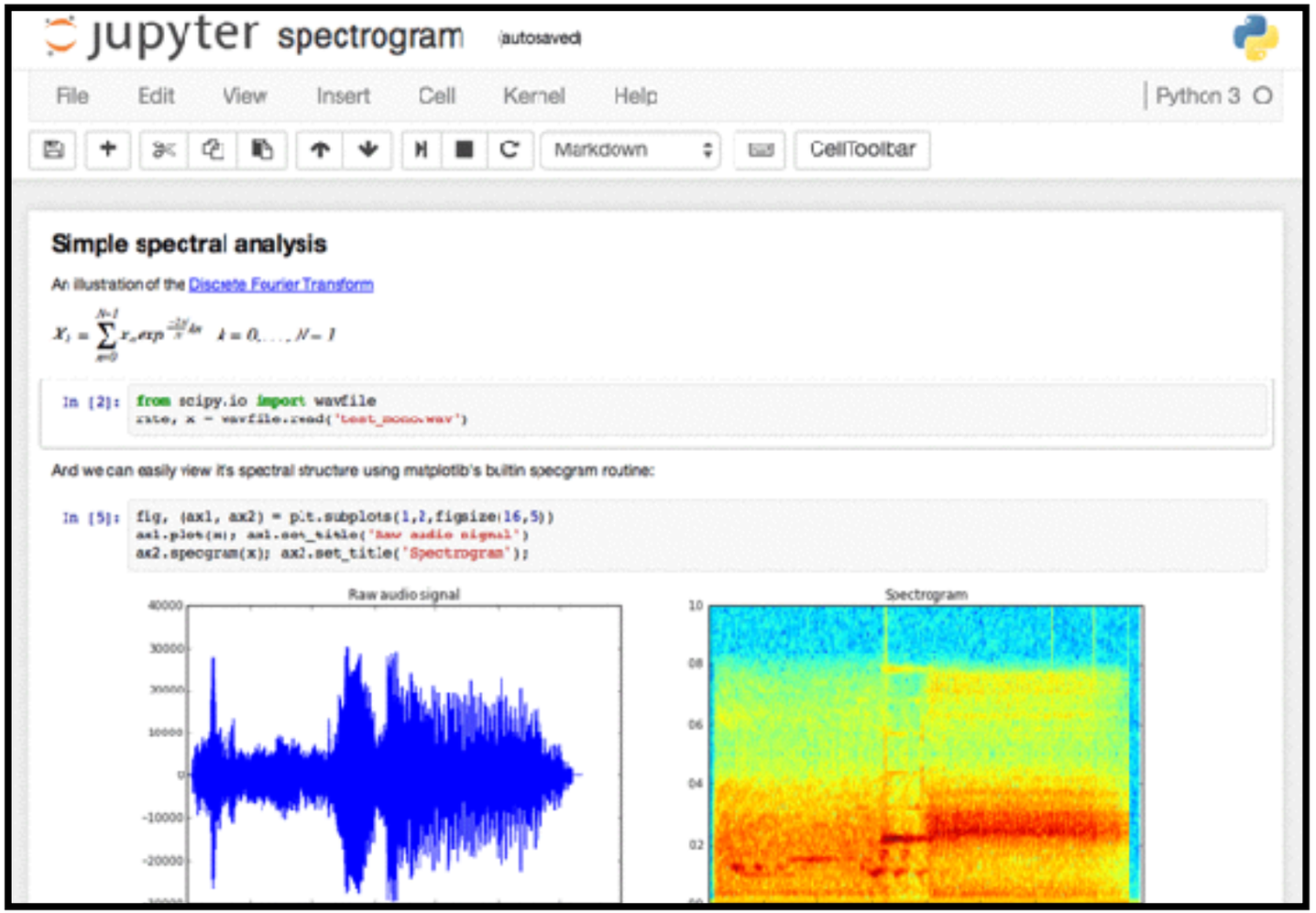
What is Data Science?	Git, command line	Python review	Pandas
Exploratory Data Analysis, Visualization	Databases	k-Nearest Neighbors	Linear Regression
Regularization	Logistic Regression	Advanced sklearn	Applied ML
Unsupervised Learning: Clustering, PCA	Decision Trees	Ensembles	Natural Language Processing

QUESTIONS?

LET'S GET OUR HANDS DIRTY

OUR STACK

- We write out Python code in Jupyter notebooks
- We access and manage our Jupyter notebooks with Git and Github
- We interact with Git using the command line, a text-based interface (e.g. terminal on Mac)



FOR THE REST OF CLASS

- Practice command line
- Learn how to access course materials on Github
- Practice Python in a Jupyter notebook

EXERCISE 29

KEY OBJECTIVE(S)

Insert learning/exercise objectives

AGENDA

Time 1. Insert key steps

DELIVERABLE

Insert deliverable/outcome

RESOURCES

List resources required / used

INSERT COMPANY NAME

SUMMARY

Insert context and background including relevant data/evidence

KEY CHALLENGE / QUESTION

Insert problem or challenge faced and key case questions

QUESTIONS?

DISCUSSION TIME