
Network ID: GA-Guest

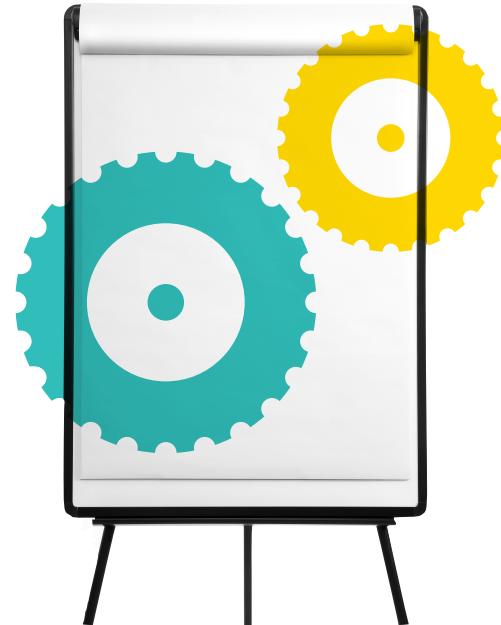


Password: yellowpencil

WELCOME TO OPENING DAY OF DATA SCIENCE

ORIENTATION AGENDA

- 1. Meet Your Course Producer**
- 2. Meet The Instructional Team**
- 3. General Course Info**
- 4. GA Mission**
- 5. Course Expectations**
- 6. Student Experience**
- 7. Meet Your Classmates**



OPENING DAY

HELLO

**WHAT IS A PROGRAM
PRODUCER?**

FOCUS: STUDENT SUCCESS + INSTRUCTOR SUPPORT

timothy.payne@galaxyedu.com

OPENING DAY

MEET YOUR INSTRUCTIONAL TEAM

INSTRUCTOR & TA

**IVAN CORNEILLET
GEORGE MCINTIRE**

COURSE INFO

DS25

**July 13th - September 21st (No
class on September 5th)**

Mondays and Wednesday 6:30-9:30

Classroom 3

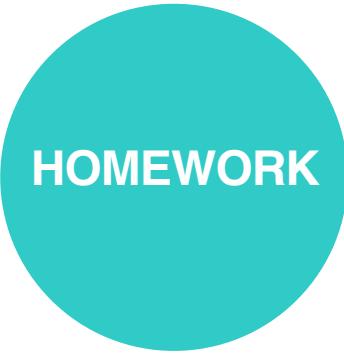
**GENERAL ASSEMBLY IS
A GLOBAL COMMUNITY
OF INDIVIDUALS
EMPOWERED TO
PURSUE THE WORK WE
LOVE.**

**GENERAL ASSEMBLY'S
MISSION IS TO BUILD
OUR COMMUNITY BY
TRANSFORMING
MILLIONS OF THINKERS
INTO CREATORS.**

COURSE EXPECTATIONS

A photograph of two individuals working on laptops at a wooden desk. The person in the foreground, wearing a black beanie and glasses, is looking down at their screen which displays code. The person in the background, wearing a light-colored shirt and glasses, is also looking at their laptop. Behind them is a chalkboard with various handwritten words and symbols, including "TOPPERS", "Julian", "Kappa", "IDEAS", "Ceremony", and "Laptop".

SUCCESSFUL GRADUATE PROFILE



HOMEWORK

COMPLETE 80% OF
HOMEWORK/LABS



ATTENDANCE

BE PRESENT FOR AT
LEAST 18 CLASSES



FINAL
PROJECT

ROCK IT!

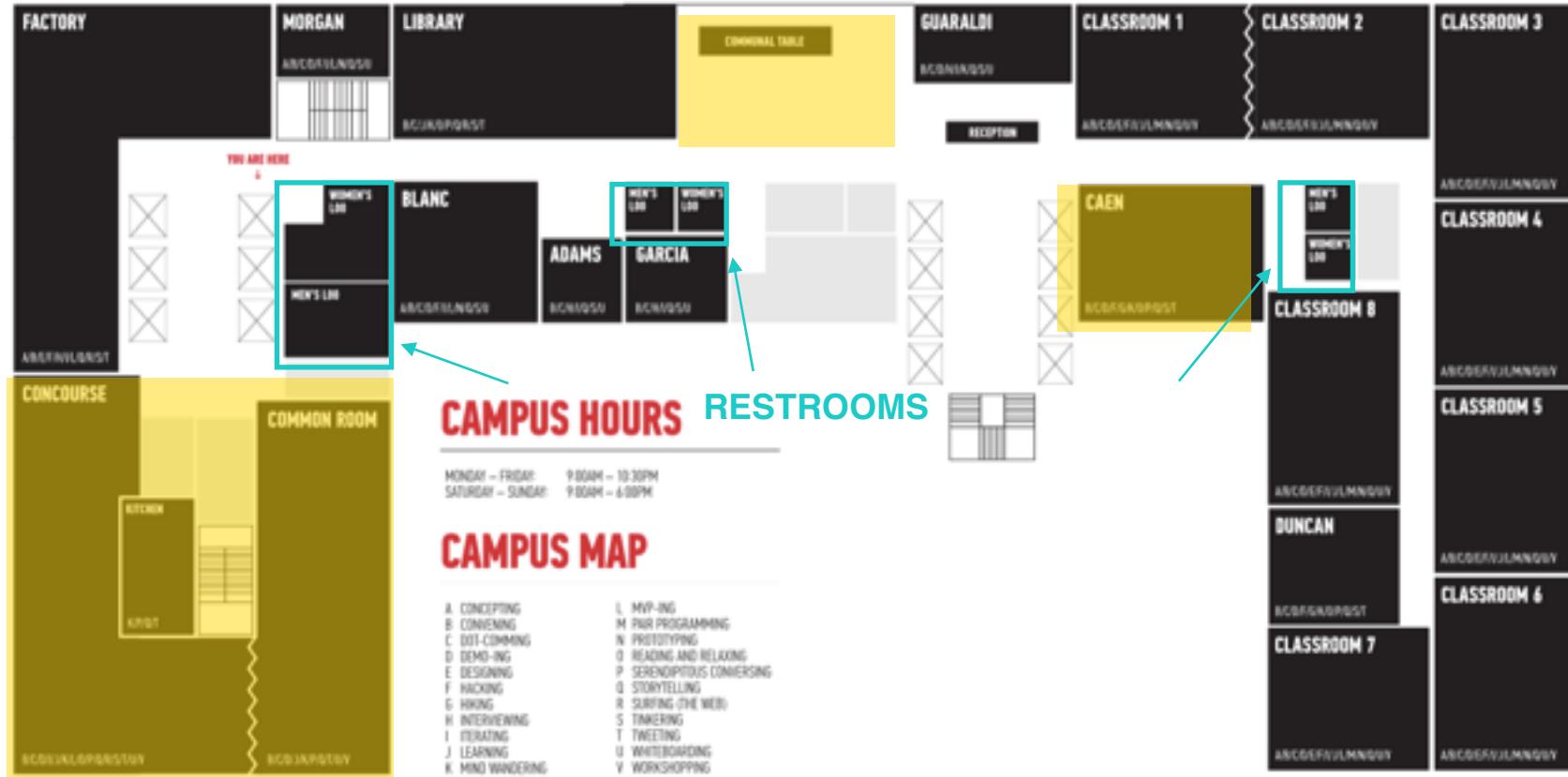


COMMUNITY

GET INVOLVED!

COURSE EXPECTATIONS

- 1. Arrive on time.**
- 2. Turn in your assignments.**
- 3. Ask questions.**
- 4. Share with peers.**
- 5. Complete exit tickets.**
- 6. Complete mid & end of course surveys.**
- 7. Make friends.**



STUDENT SPACE

DATA SCIENCE ORIENTATION





STUDENT EXPERIENCE

DATA SCIENCE ORIENTATION



DISCOUNTS!

15% off all classes
and workshops

code:

currentstudentdiscount15



GA Directory

The GA Directory is a place for students, alumni and instructors to connect.

- Find your classmates
- Reach out to alumni and instructors
- Hire talent based on skills and experience

directory.generalassemb.ly

The screenshot shows a web-based directory interface. At the top, there's a search bar and filters for 'Name', 'Course', 'Campus', 'Location', and 'Skill'. Below these filters, a section titled 'My Cohort' lists 'Front End Web Development' and 'Data Science Immersive'. The main area displays a list of profiles for users like Phillip Arnold, Marie Bass, Willie Brewer, Ryan Darker, Helen Johnson, Benjamin Matthews, and Ralph Silva, each with a small profile picture, their current location (e.g., New York City), a brief bio, and a 'View Profile' button.

| Name | Location | Bio | Action |
|-------------------|---------------|---|------------------------------|
| Phillip Arnold | New York City | Former analyst, Gif enthusiast, coffee addict, excited to collaborate on new projects with fellow alumni and talented people. Currently I am a UI Designer at Balsamiq | View Profile |
| Marie Bass | New York City | I like to solve problems. Currently working at an e-commerce startup in the art space. Looking for designers and engineers to help! Currently I am a Product Manager/General Manager/Studio Executive at Cool Army Stories | View Profile |
| Willie Brewer | New York City | Previously I was a Software Engineer at Bitly My skills include JavaScript, HTML/CSS, Node.js | View Profile |
| Ryan Darker | New York City | Previously I was a Full Stack Developer at Magnetic | View Profile |
| Helen Johnson | New York City | Users come first, I research and create useful products and platforms for our users. Currently I am a UI Designer at Medium | View Profile |
| Benjamin Matthews | San Francisco | Previously I was a Web Developer at Balsamiq My skills include JavaScript/Words, Bootstrap 3, HTML, SASS | View Profile |
| Ralph Silva | New York City | Previously a PhD candidate in Political Science, moving towards Data Science in the start-up scene. Currently I am a Data Scientist at Full Stack Everywhere | View Profile |

INTRODUCTIONS

MEET YOUR CLASSMATES

**WE'RE ALL
IN THIS
TOGETHER.**

—

QUESTIONS?

Welcome to Data Science

*Ivan Corneillet
Data Scientist*

Learning Objectives

After this lesson, you should be able to:

- Describe the roles and components of a successful learning environment
- Define data science and data scientists
- Setup your development environment and practice the different workflows used in the course



Today

Here's what's happening today:

- Welcome to GA and DS!
- Pre-work
- Setting you up for success
- What is data science and who are data scientists?
- Installfest
- Lab – Python Review
- Review
- Exit Tickets



Welcome to GA and DS!



Pre-Work

Pre-Work

Before this lesson, you should already be able to:

- Define basic data types used in object-oriented programming
- Recall the Python syntax for lists, dictionaries, and functions
- Create files and navigate directories using the command line interface



Setting You Up for Success

Meet Your Team

- Ivan Corneillet, Lead Instructor



- George McIntire, Associate Instructor

- Tim Payne, Course Producer



Course Logistics

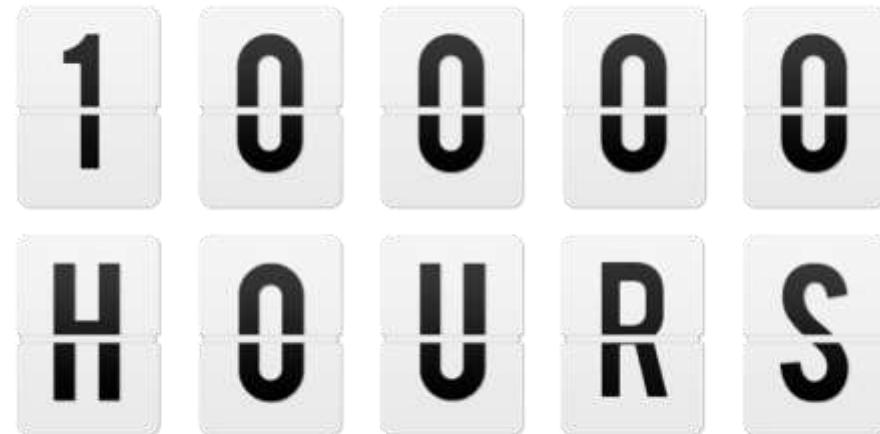
- Lead Instructor
 - Ivan Corneillet (ivan+GA@paspeur.com)
- Associate Instructor
 - George McIntire (geo.mcintire@gmail.com)
- Course Producer
 - Tim Payne (timothy.payne@generalassemb.ly)
- Class
 - July 13 – September 21, Mondays and Wednesdays,
6:30PM – 9:30PM (no class on September 5)
 - Classroom 3
- Slack
 - <https://ds-sf-25.slack.com>
- GitHub
 - <https://github.com/ga-students/DS-SF-25>
- Exit Tickets
 - <http://tiny.cc/ds-sf-25>

What skills will I learn in this class?

| | | | | |
|--|--------------------------------------|--|--------------------------------|---|
| Research Design and Data Analysis | Research Design | Data Visualization in <i>pandas</i> | Statistics | Exploratory Data Analysis in <i>pandas</i> |
| Foundations of Modeling | Linear Regression | Classification Models | Evaluating Model Fit | Presenting Insights from Data Models |
| Data Science in the Real World | Decision Trees and Random Forests | Time Series Data | Natural Language Processing | Databases |

Gladwell's 10,000 Hour Rule

- “Greatness requires enormous time”
- It takes roughly ten thousand hours of practice to achieve mastery in a field



The image shows a digital timer or scoreboard consisting of two rows of five square segments each. The top row displays the digits '1', '0', '0', '0', and '0'. The bottom row displays the letters 'H', 'O', 'U', 'R', and 'S', which together read '10000 HOURS'.

How will I apply and reinforce these new skills?

| | | | | |
|---|------------------------|---------------------------|----------------------|--|
| Unit Project You will design a research project, perform exploratory data analysis and build a logistic model to determine what factors affect admission the most | Research Design | Exploratory Data Analysis | Logistic Modeling | Executive Summary with Findings |
| Final Project Using a dataset of your choosing, you will design a project, build a data science model and present their finding to the course | Lightning Presentation | Experimental Write-up | Exploratory Analysis | Notebook Draft Final Presentation |

Past Student Projects

(<https://gallery.generalassemb.ly/DS>)

The screenshot displays the homepage of "THE GALLERY" at <https://gallery.generalassemb.ly/>. The page features a header with the GA logo, a search bar for "Data Science" and "All Cities", and a "LOOK" button. The main title "THE GALLERY" is prominently displayed in large, bold letters, with the subtitle "A COLLECTION OF STUDENT-UPLOADED PROJECTS" below it. The page is organized into a grid of project cards, each with a thumbnail, title, description, and author information.

- Read Like You Tweet** by Karissa Errea: A New York Times Article Recommendation System based on Your Twitter Timeline.
- RETHINKING THE COMMUTE** by Gregory Bushik: An analysis of urban sprawl and its impact on transportation.
- PREDICTING REDDIT POPULARITY** by Noah Fine: A project comparing General Assembly and reddit logos.
- PREDICT DISCOUNTS: 2POINTB** by Daron Hargan: An app that predicts discounts at 2PointB locations.
- YELP REVIEW DATA ANALYSIS** by Zeph Dean: An analysis of Yelp review data.
- TALK TO ME** by Tess Shashere: A project involving media advertising and family.
- TWITTER & SENTIMENT ANALYSIS** by Monisha Chauhan: An analysis of tweets related to gender equality.
- P2P LOAN ANALYSIS** by Clark Dubois: An analysis of P2P loan data across the United States.
- BILLBOARD TOP 40 ANALYSIS** by Matthew Lavelle: An analysis of Billboard Top 40 charts from 1960 to 2010.
- HATER NEWS** by Karen McNamee: An analysis of Hater News user engagement.

Typical Class

- Today's objectives
- Announcements and exit tickets
- Review of the previous class
- Series of:
 - Short lectures
 - (deck, whiteboard, codealongs, and demos)
 - Practices
 - (individual and group exercises, and codealongs)
- Lab/Independent study
- Review of today's class
- Office hours for final projects (for the last 3 weeks of the course)
- Exit Tickets



Q & A



Setting You Up for Success

Slack (<https://ds-sf-25.slack.com>)

GitHub (<https://github.com/ga-students/DS-SF-25>)

Exit Tickets (<http://tiny.cc/ds-sf-25>)



Q & A



What is Data Science and Who are Data Scientists?

Activity | What is Data Science and Who are Data Scientists?

EXERCISE

DIRECTIONS (10 minutes)

1. What is data science? What are its applications? Why now? What's next?
2. Who are data scientists? How do they add value? What makes a good data scientist?
3. When finished, share your answers with your table

DELIVERABLE

Answers to the above questions

Harvard Business Review | “Data Scientists: The Sexiest Job of the 21st Century” (2012)

(<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>)



SPOTLIGHT ON BIG DATA

Spotlight

ARTWORK: Yasser Iskander, Andrew J. Dabatz
SIZE: 10K screens (or a page three high across
your desk, 16" x 11")

Data Scientist: *The Sexiest Job of the 21st Century*

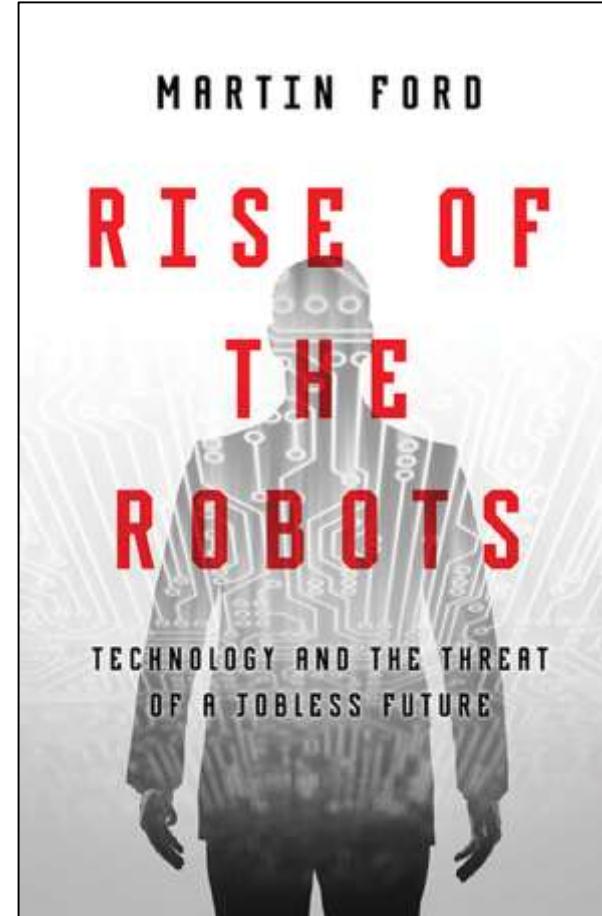
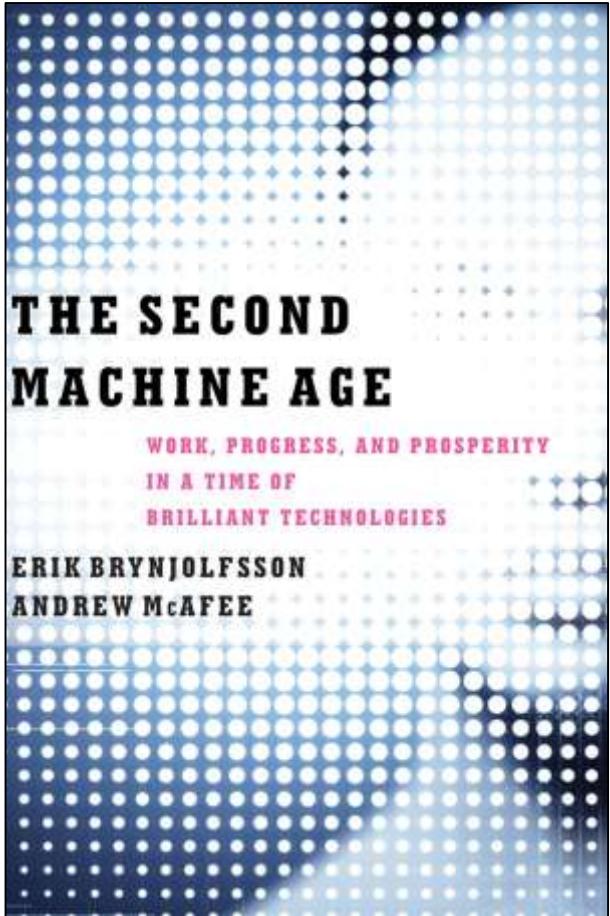
Meet the people who can coax treasure out of messy, unstructured data.
by Thomas H. Davenport and D.J. Patil

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a startup. The company had just under 8 million members, and the member base was growing quickly as existing members invited their friends and colleagues to join. But users weren't making connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—and you probably leave early."

© Harvard Business Review. October 2012

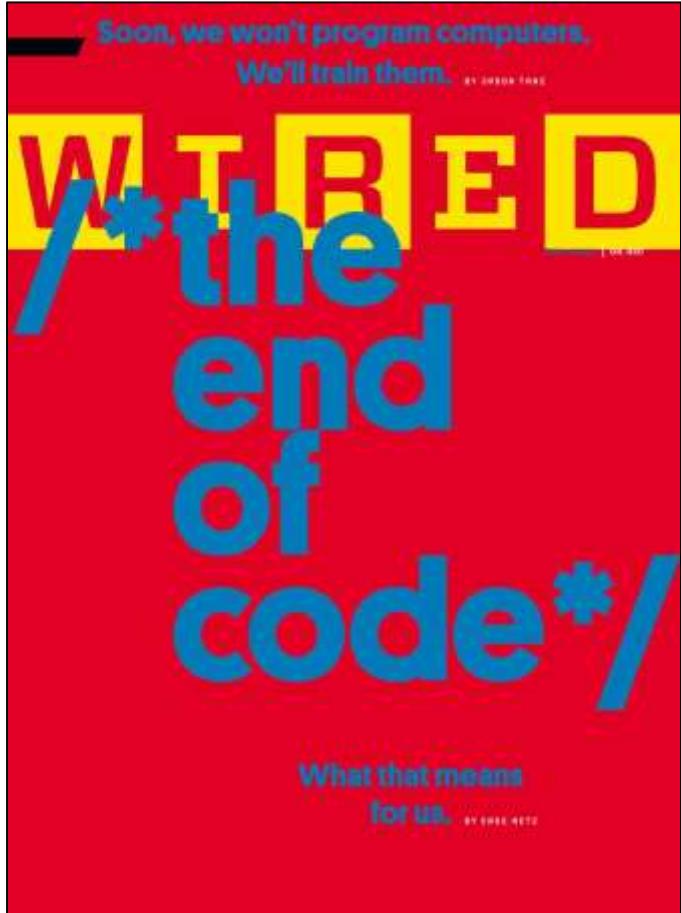
Source: Harvard Business Review

“The Second Machine Age” and “Rise of the Robots” (2016)



Wired | “The Rise of Artificial Intelligence and the End of Code” (2016)

(<http://www.wired.com/2016/05/the-end-of-code/>)



Source: Wired

Data Science Businesses

FiveThirtyEight

Walmart

Google



U B E R



NETFLIX

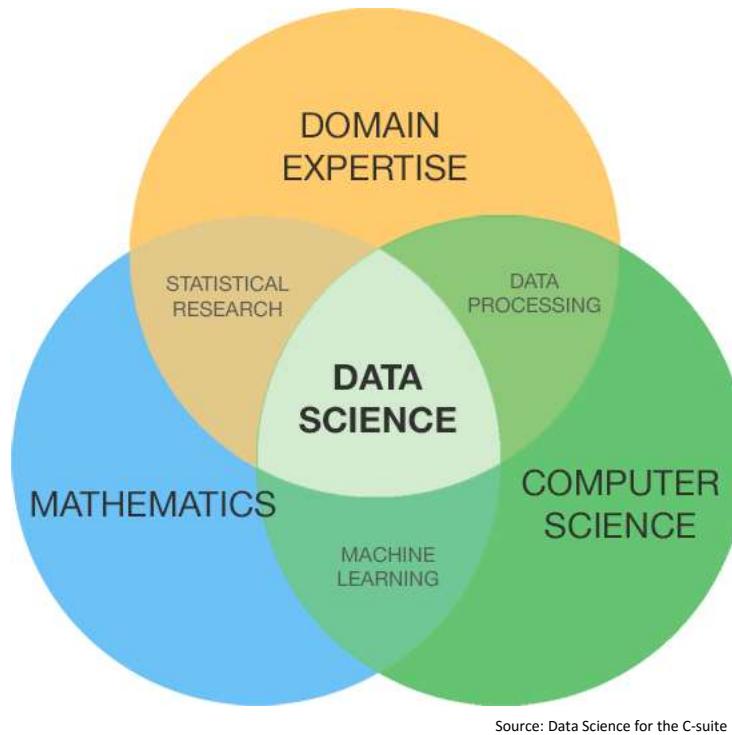
amazon



facebook

LinkedIn

Data Science involves a variety of skillsets



Source: Data Science for the C-suite

Data Scientists in ≤ 140 characters

Zvi (@nivertech)
"Data Scientist" is a Data Analyst who lives in California.
RETWEETS 162 LIKES 82
5:55 PM - 14 Mar 2012

Josh Wills (@josh_wills)
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.
RETWEETS 1,339 LIKES 799
9:55 AM - 3 May 2012

A
B
C
D

Harvard IACS (@Harvard_IACS)
Data Scientist: Someone better at statistics than a software engineer, and better at software engineering than a statistician?
#datastorm14
RETWEETS 43 LIKES 15
6:42 AM - 24 Jan 2014

Javier Nogales (@fjnogales)
Data Scientist (2/2): person who is worse at statistics than any statistician and worse at software engineering than any software engineer
RETWEET 1 LIKES 5
8:08 AM - 27 Jan 2014

Source: Twitter



Q & A



Installfest



Installfest

“GA” user



Installfest

GitHub Desktop

GitHub Desktop

- As part of its installation, GitHub Desktop will seamlessly setup encryption (SSH) keys (so you don't have to) to authenticate you against your repositories
- These keys will work within GitHub Desktop as well as on the *Terminal* when using “git” directly

GitHub Desktop

- › Checking that you are correctly authenticated after installing GitHub Desktop and setting up your user profile:
 - › Launch *Terminal* and check the output of the following command:

```
ssh -T git@github.com (for GitHub)
```

- › If you get the following message, you are good to go:

Hi ███! You've successfully authenticated, but GitHub does not provide shell access.

- › Set some Git configuration:

```
git config --global push.default simple
```



Installfest

Continuum's Anaconda (Python 2.7)



Q & A



Git and Anaconda's Jupyter Notebook Primer

Git and Anaconda's Jupyter Notebook Primer

Practice #1 | Fork the course repository and clone your fork; update your clone and fork as needed

Practice #1 | How to *fork* the course repository and *clone* your *fork* (one-time only)

- ① Fork the GitHub course's repository into a new GitHub repository
 - Login into GitHub; open <https://github.com/ga-students/DS-SF-25>; click on the Fork button on the top right; your fork is at <https://github.com/paspeur/DS-SF-25> (replace *paspeur* with your username)
- ② *Clone* your *fork* (a.k.a., the “*origin*” repository) into your computer
 - Open a new *Terminal*; type “`git clone https://github.com/paspeur/DS-SF-25`”; your clone is under the DS-SF-25 folder (type “`cd DS-SF-25`” to change the current directory to your clone's root directory)
- ③ Set the GitHub course's repository as the “*upstream*” repository
 - Within *Terminal*, type “`cd DS-SF-25`” to change the current directory to your *clone*'s root directory then “`git remote add upstream https://github.com/ga-students/DS-SF-25`”

Practice #1 | How to update your *clone* (*local* repository) and *fork* (*origin/remote* repository on GitHub) with new commits from the *course* repository (*upstream/remote* repository on GitHub)

- ① Fetch the newer commits from the *course* repository
 - `git fetch upstream`
- ② Merge and commit changes from the *course* repository to your *clone*
 - `git merge --no-commit upstream/master`
 - `git commit -m "Merged commits from ga-students/SF-DAT-25 up to 2016-07-13"`
 - (if the merge was “Fast-forward”, i.e., trivial, there is no need to commit these changes)
- ③ Push from your *clone* to your *fork*
 - `git push`
 - (Git will ask you your GitHub credentials the first time around but will remember them afterwards)

Git and Anaconda's Jupyter Notebook Primer

Practice #2 | Run Anaconda's Jupyter Notebook

Practice #2 | How to run Anaconda's Jupyter Notebook on your *clone*

- ① Open a new *Terminal* and change the current directory to your *clone*'s root directory
 - `cd DS-SF-25`
- ② Start Jupyter Notebook
 - `jupyter notebook`
 - (Jupyter Notebook will be running on the background and won't return the prompt back to you; open a new *Terminal* as needed)

Git and Anaconda's Jupyter Notebook Primer

Practice #3 | How to add or modify existing and commit them in your clone and fork (e.g., your Unit Project #1)

Practice #3 | How to add or modify existing and commit them in your *clone* and *fork* (e.g., your Unit Project #1)

- ① Open a new *Terminal* and “cd” (change directory) to your *clone*’s root repository
 - cd DS-SF-25
- ② Make a copy of the Unit Project #1 Jupyter notebook
 - cp unit-projects/1/code/unit-project-1-starter-code.ipynb unit-projects/1/code/unit-project-1-ivan.ipynb
- ③ Update and save the notebook
 - git add unit-projects/1/code/unit-project-1-ivan.ipynb
- Repeat ② and ③ as often as you need it
 - Commit early and often (<http://blog.codinghorror.com/check-in-early-check-in-often>)

Practice #3 | How to add or modify existing and commit them in your *clone* and *fork* (e.g., your Unit Project #1) (cont.)

- ▶ ④ Commit your Jupyter notebook to your *clone*
 - ▶ `git commit -m "a descriptive message so you know what this change is about in 3 months or next week..."`
- ▶ ⑤ Push from your *clone* to your *fork*
 - ▶ `git push`
- ▶ But if you can't *push*, you need to *merge*
 - ▶ Git may complain that changes in your fork aren't in your clone and that you first need to synchronize from your fork back to your clone first; check the `git fetch/merge` slide from Practice #1

Practice #3 | How to add or modify existing and commit them in your *clone* and *fork* (e.g., your Unit Project #1) (cont.)

- And if you can't *merge*, you have *conflict(s)* you need to *resolve*
 - E.g., you are trying to push changes that modify the same cell that a previous commit on the fork but not on the clone also changed. Because both commits aren't compatible with each other, Git errors out to let you resolve the merge manually)
 - You won't be able to resolve the conflict with Jupyter Notebook: Git's annotation of the conflict broke the structure of the notebook
- ① Take notes of conflicts
 - `git diff unit-projects/1/code/unit-project-1-ivan.ipynb`
 - The output will tell you in plain text where the conflicts occur. Make notes of them
- ② Undo the merge for the time being
 - `git reset --hard`

Practice #3 | How to add or modify existing and commit them in your *clone* and *fork* (e.g., your Unit Project #1) (cont.)

- ③ Make a copy of your pre-merge changes because you will apply them manually
 - `cp unit-projects/1/code/unit-project-1-ivan.ipynb unit-projects/1/code/unit-project-1-ivan-pre-merge.ipynb`
- ④ Merge again; of course, you'll get the same conflicts
 - `git merge`
- ⑤ Checkout the copy from the fork
 - `git checkout --theirs unit-projects/1/code/unit-project-1-ivan.ipynb`

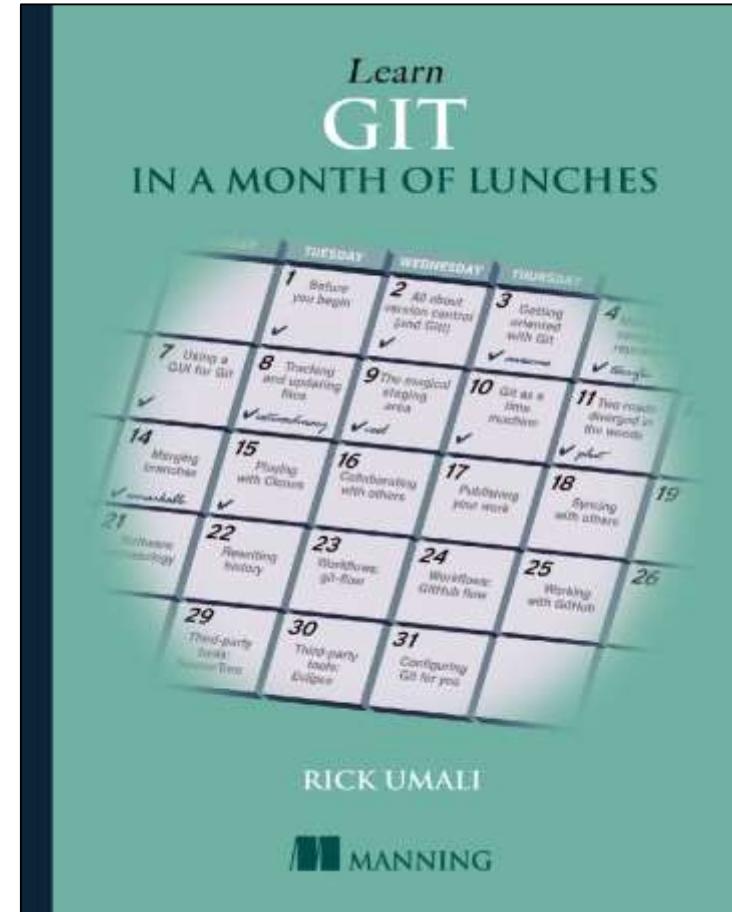
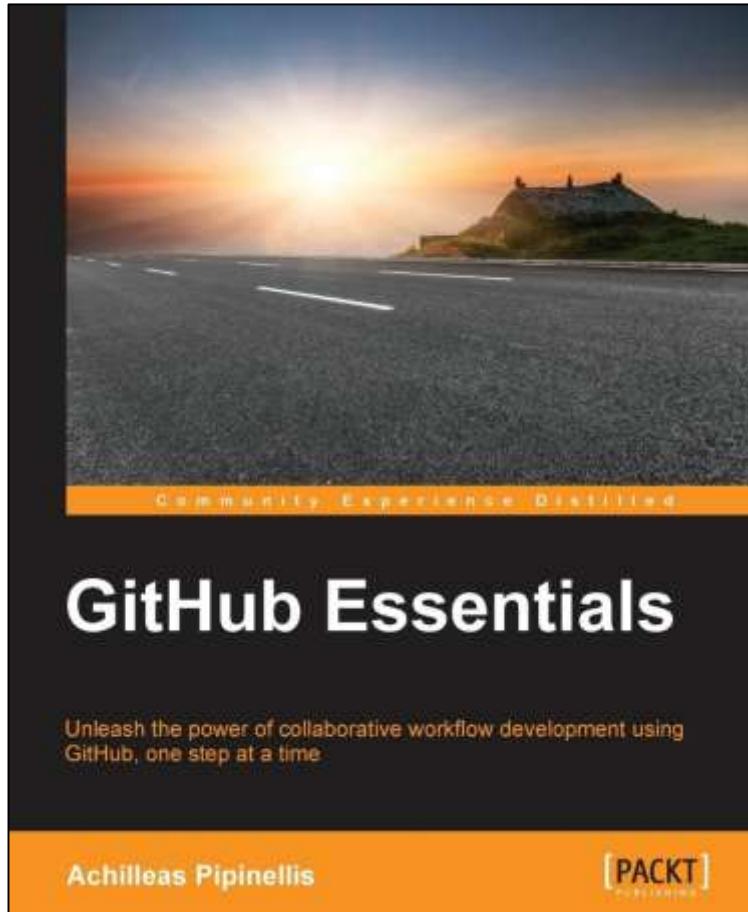
Practice #3 | How to add or modify existing and commit them in your *clone* and *fork* (e.g., your Unit Project #1) (cont.)

- ⑥ Now you can use Jupyter Notebook to open both *unit-project-1-ivan.ipynb* (the last version in your *fork*) and *unit-project-1-ivan-pre-merge.ipynb* (the last version in your *clone* that you wanted to push)
 - Copy and paste the relevant sections from *unit-project-1-ivan-pre-merge.ipynb* back to *unit-project-1-ivan.ipynb*
- ⑦ Tell Git you resolve the conflicts
 - `git add unit-projects/1/code/unit-project-1-ivan.ipynb`
- ⑧ Finally, commit and push... (like we've done a few times already)
 - `git commit -m "Updated Unit Project #1"`
 - `git push`



Q & A

A couple of resources to get started with Git (*optional; not required for the course*) (the styles are different but the content overlaps so only pick one if any)

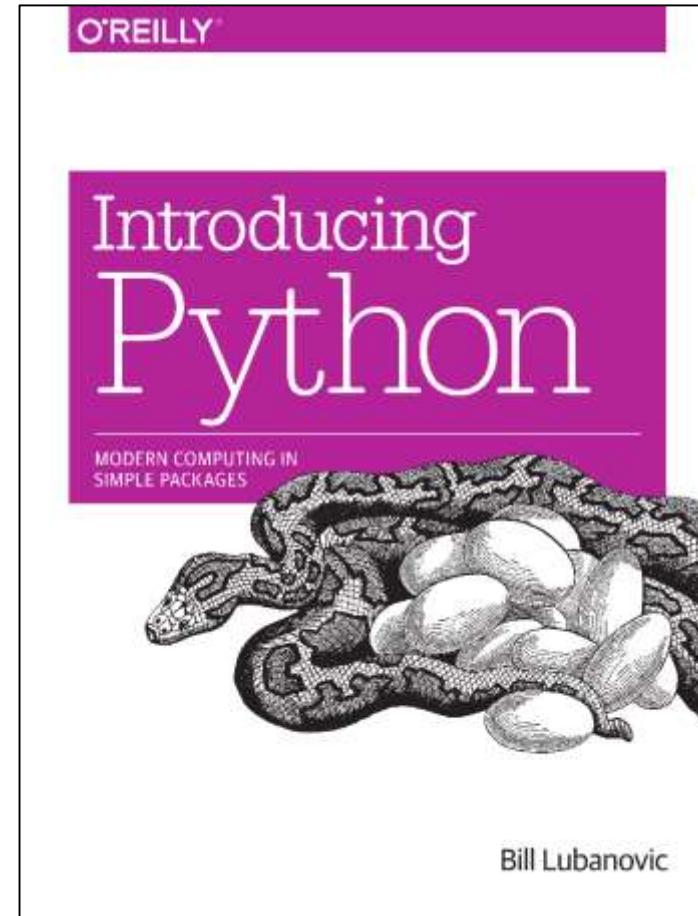
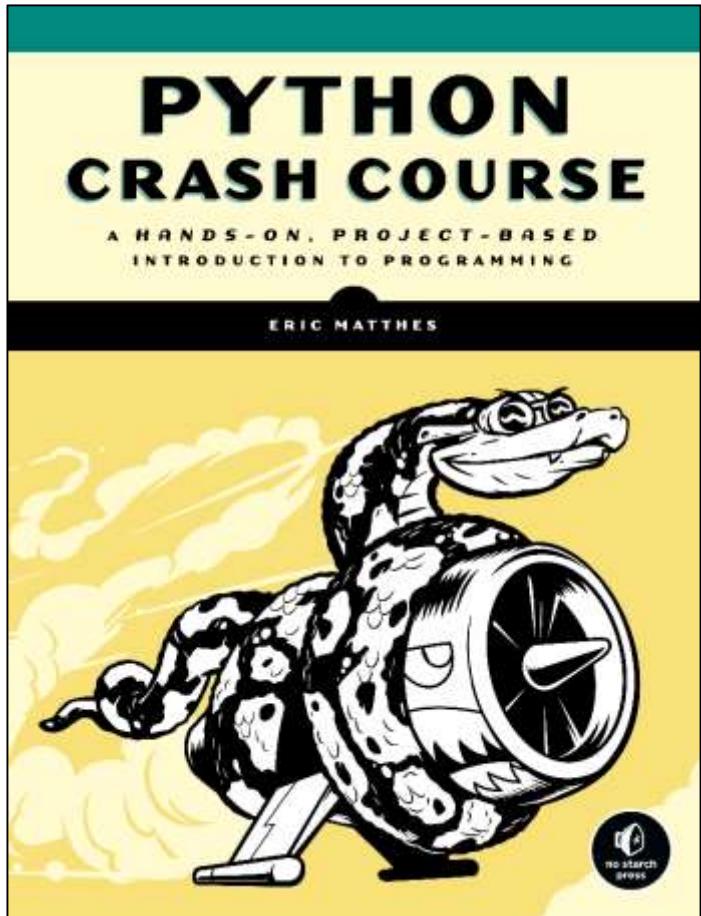




Lab

Python Review

A couple of resources to get started with Python (*optional; not required for the course*) (the styles are different but the content overlaps so only pick one if any)





Review

Review

You should now be able to answer the following questions:

- Describe the roles and components of a successful learning environment
- Define data science and data scientists
- Setup your development environment and practice the different workflows used in the course



Q & A



Before Next Class

Before Next Class

- Complete your development environment setup and practice the different workflows used in the course
- Complete the onboarding pre-work
- Look into the first unit project
- Start ideating about your final project's topic

Next Class

Research Design and pandas

Learning Objectives

After the next lesson, you should be able to:

- Define the data science workflow
- Define a problem and types of data
- Identify dataset types
- Apply the data science workflow in the *pandas* context
- Write an Jupyter notebook to import, format, and clean data using the *pandas* library



Exit Ticket

Don't forget to fill out your exit ticket [here](#)

Slides © 2016 Ivan Corneillet Where Applicable
Do Not Reproduce Without Permission