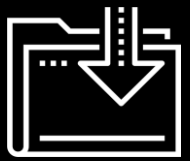
Projects & Collaboration with Git

Data Boot Camp

Lesson 7.1



© 2019 Trilogy Education Services, Inc.

Class Objectives

By the end of today's class you will be able to:

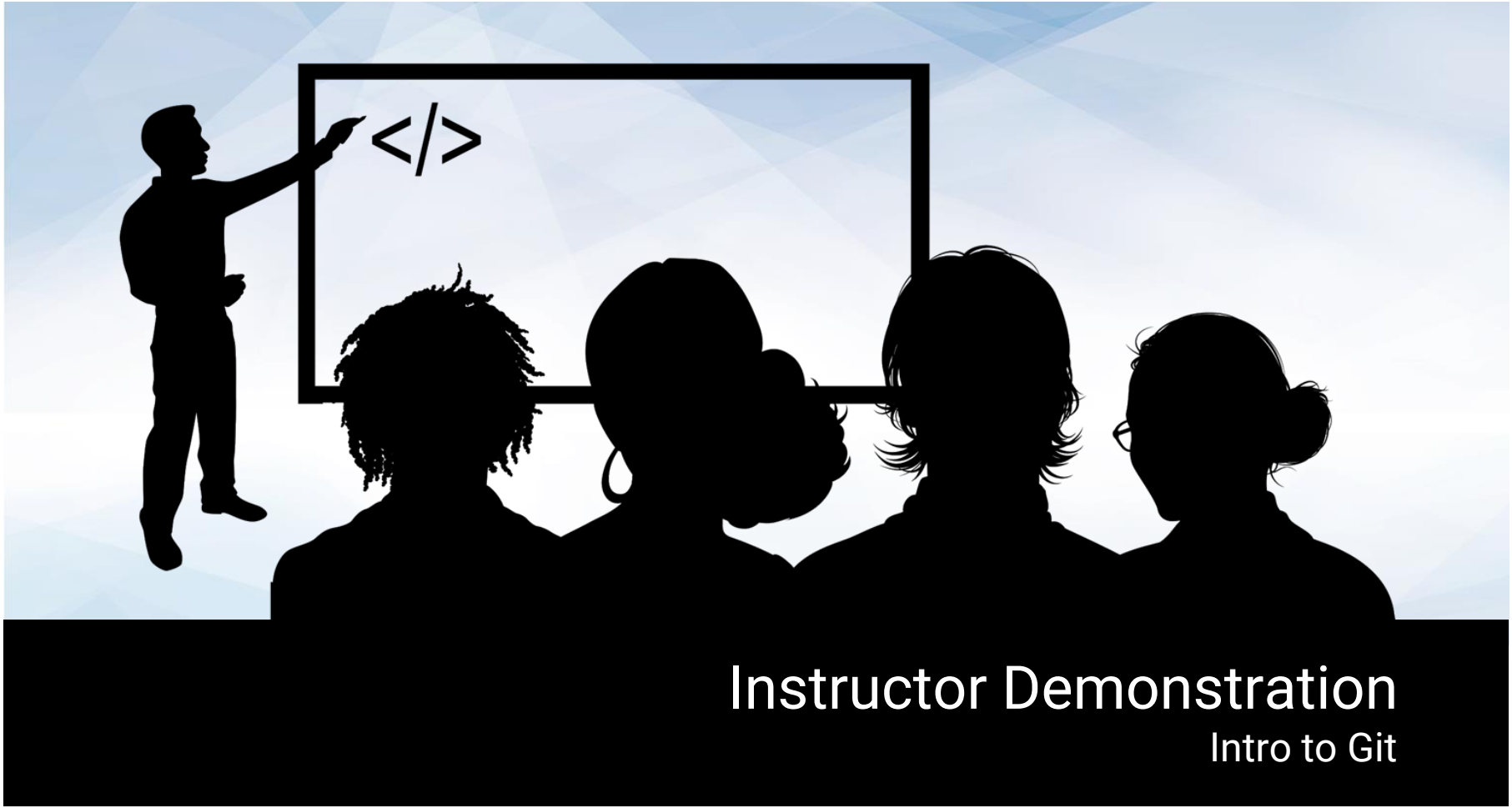
Articulate the requirements for Project 1.

Draw and interpret diagrams of Git branching workflows.

Create new branches with Git.

Push local branches to GitHub.

START



Instructor Demonstration

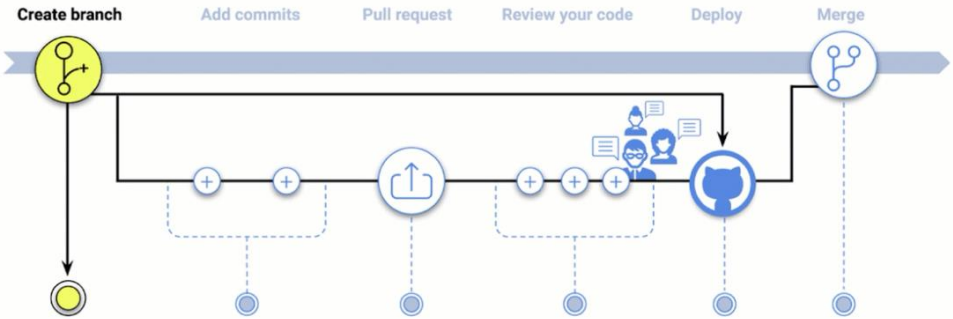
Intro to Git

Intro to Git

What is Git?

git --everything-is-local

- Git is a distributed version-control system for **tracking changes** in source code during software development.



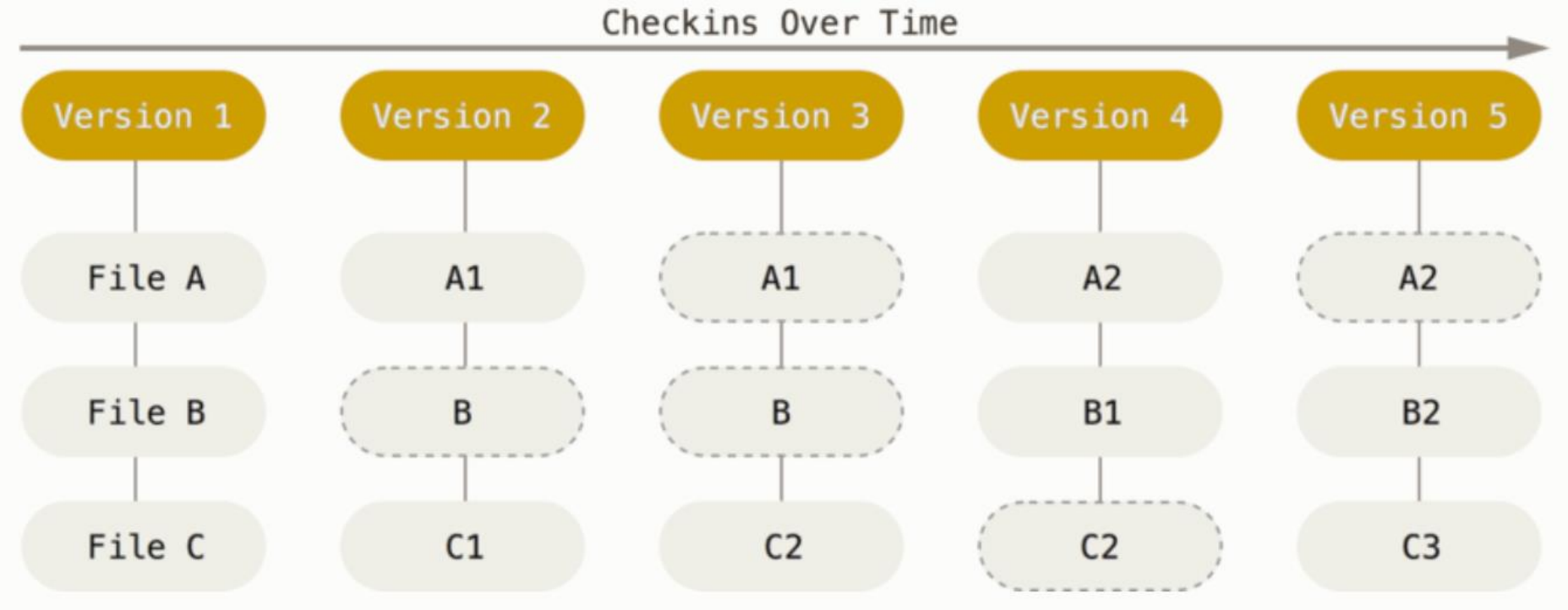
Intro to Git

Git and your Project!

- **Scenario**: Your group has been working with Uber's rider data, and you decided to analyze the average age of the riders:
    - Git essentially allows us to write this code, and save it with the name: `age_analysis`.
    - The code in `age_analysis` differs from the root code.
    - The root code for the project is called `master`.
    - `age_analysis` is a branch originated from the `master` branch. It contains updates that will be added to the main branch when it's ready to merge.

Intro to Git

Git's "Snapshot model"



Activity: Creating a Project Repo

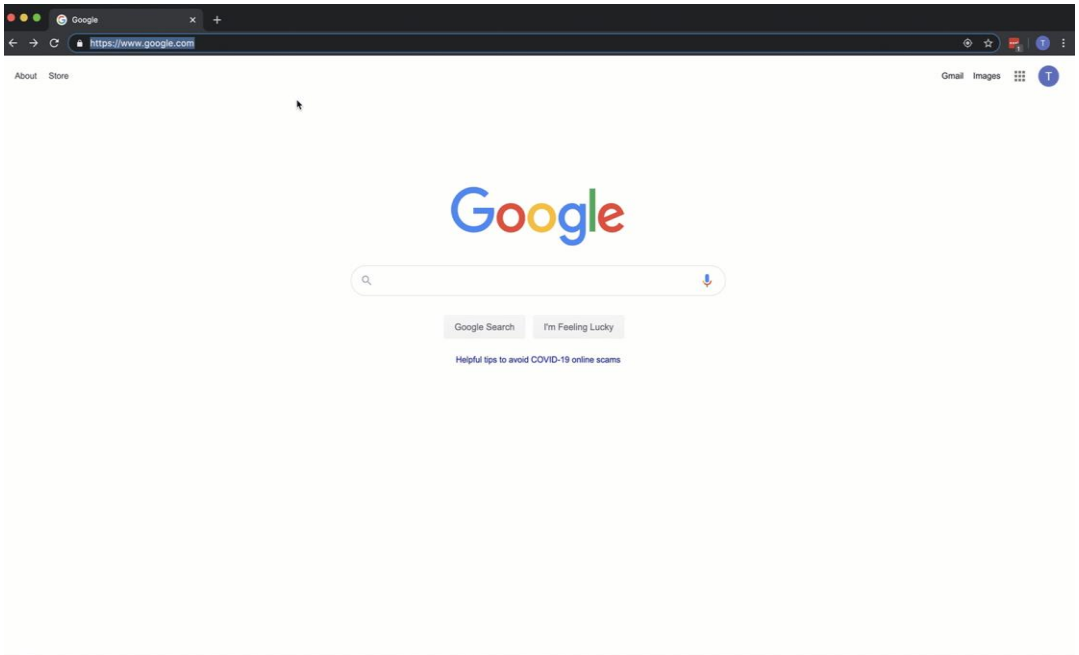
In this activity, everyone will set up a GitHub repository that we can use for our projects.

Suggested Time:
10 Minutes

Activity: Creating a Project Repo

Instructions Part 1 - Creating a Repo:

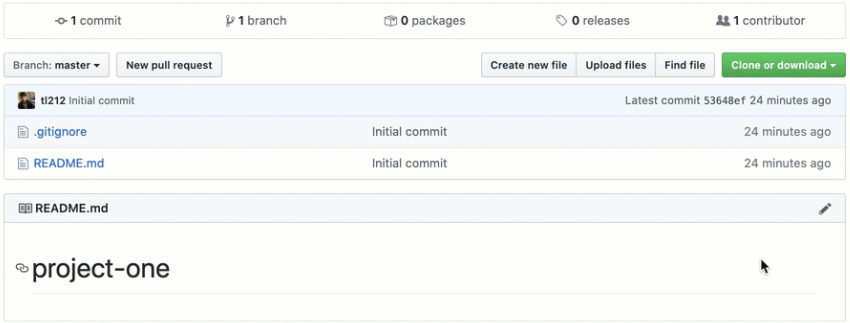
- Nominate one of your group member to create the Project Repository.
- Go to GitHub and do the following:
    - Click on the plus sign next to your profile picture.
    - Click on ‘New Repository’.
    - Initialize with `.gitignore`.
    - Choose `Python` in the gitignore dropdown.



Activity: Creating a Project Repo

Instructions Part 2 - Cloning the Repo:

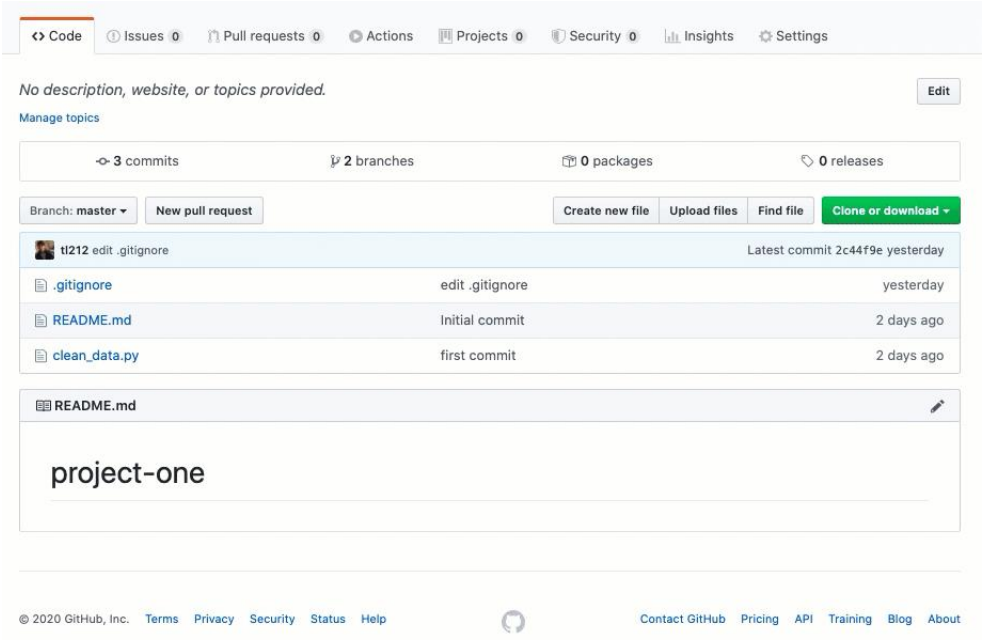
- Slack the remote URL to your team members.
- `git clone` the repository.
    - Once in the repo page click ‘Clone or download’ and copy.
    - Open Terminal and `git clone`



Activity: Creating a Project Repo

Instructions Part 3 - Adding Collaborators:

- In your Project Repo page click on the “Settings” tab.
    - Click on “Manage access” on the left panel.
    - Click on ‘Invite collaborator”.
    - Type in your colleague’s GitHub username and click add.



Activity: Workflows

In this activity, you will take a few minutes to review the concepts we have learned.

Suggested Time:
5 Minutes

Activity: Workflows

Instructions:

- This is a diagramming exercise. You can either draw your solutions on paper, or use the interface provided at Git Viz.

- Check your slack for: Activities/02-Stu_Workflows/README.md or open directly from your students repo.

Time’s Up! Let’s Review.

Activity: Creating Branches

In this activity, everyone will review how to create a branch in GitHub.

Suggested Time:
10 Minutes

Activity: Creating Branches

Creating a Branch

- To create a new, isolated development history, we must create **branches**.

- Alternatively, we can create a branch and then switch to it as two separate steps, though this is uncommon.

Activity: Pushing to GitHub

In this activity, everyone will review how to push a commit in GitHub.

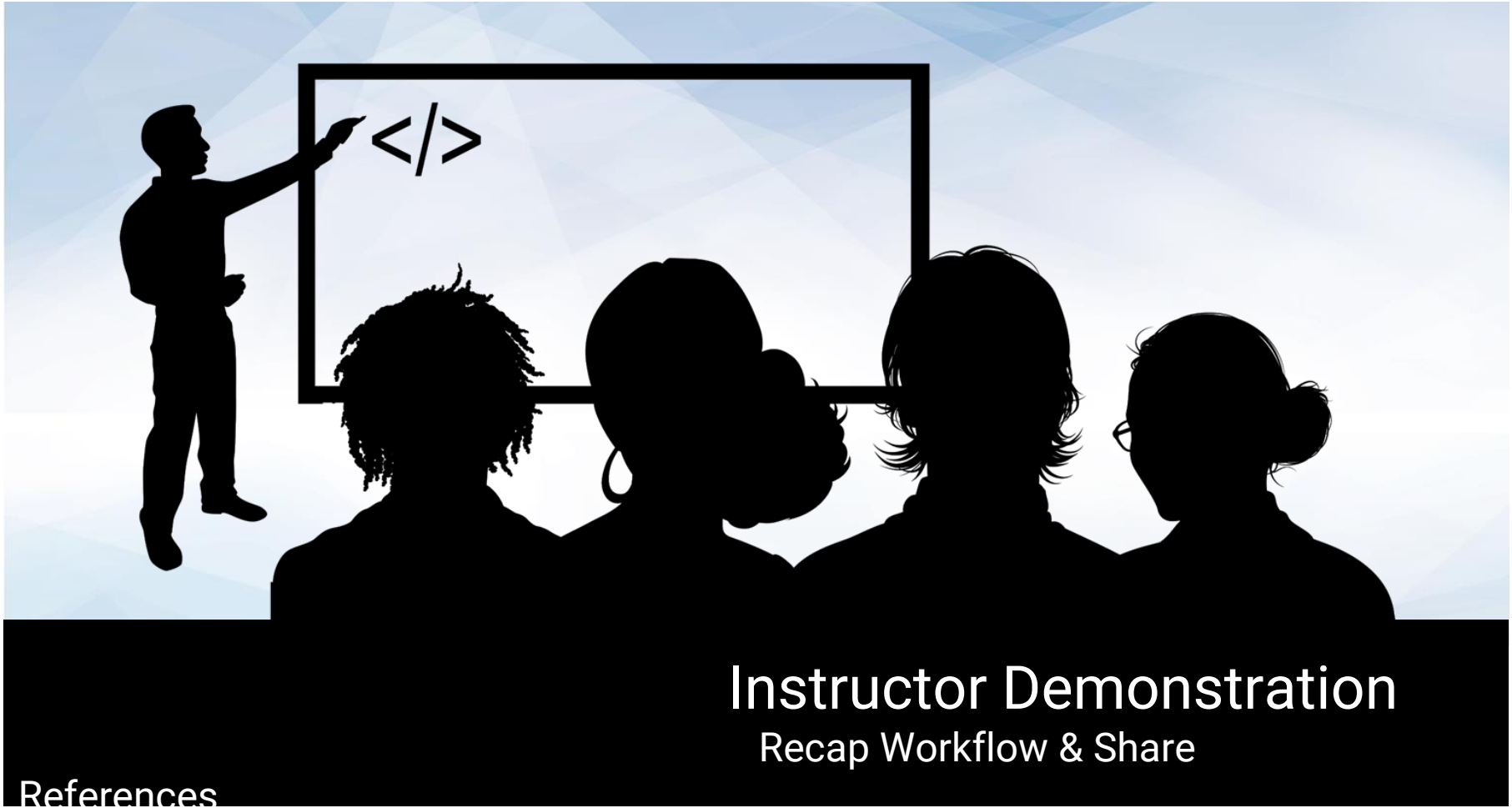
Suggested Time:
10 Minutes

Activity: Pushing to GitHub

git push origin <branch_name>

Activity: Pushing to GitHub

git pull

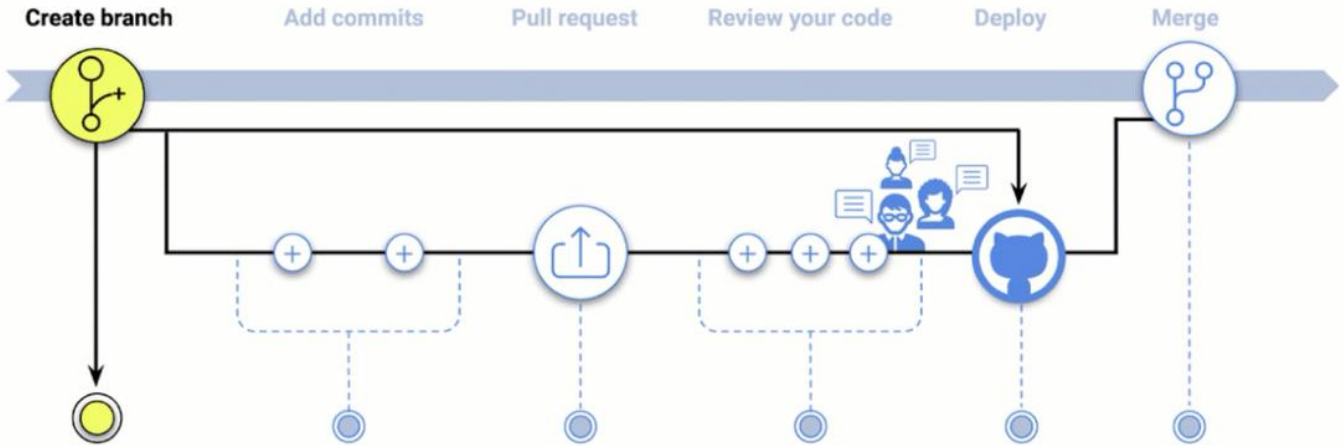


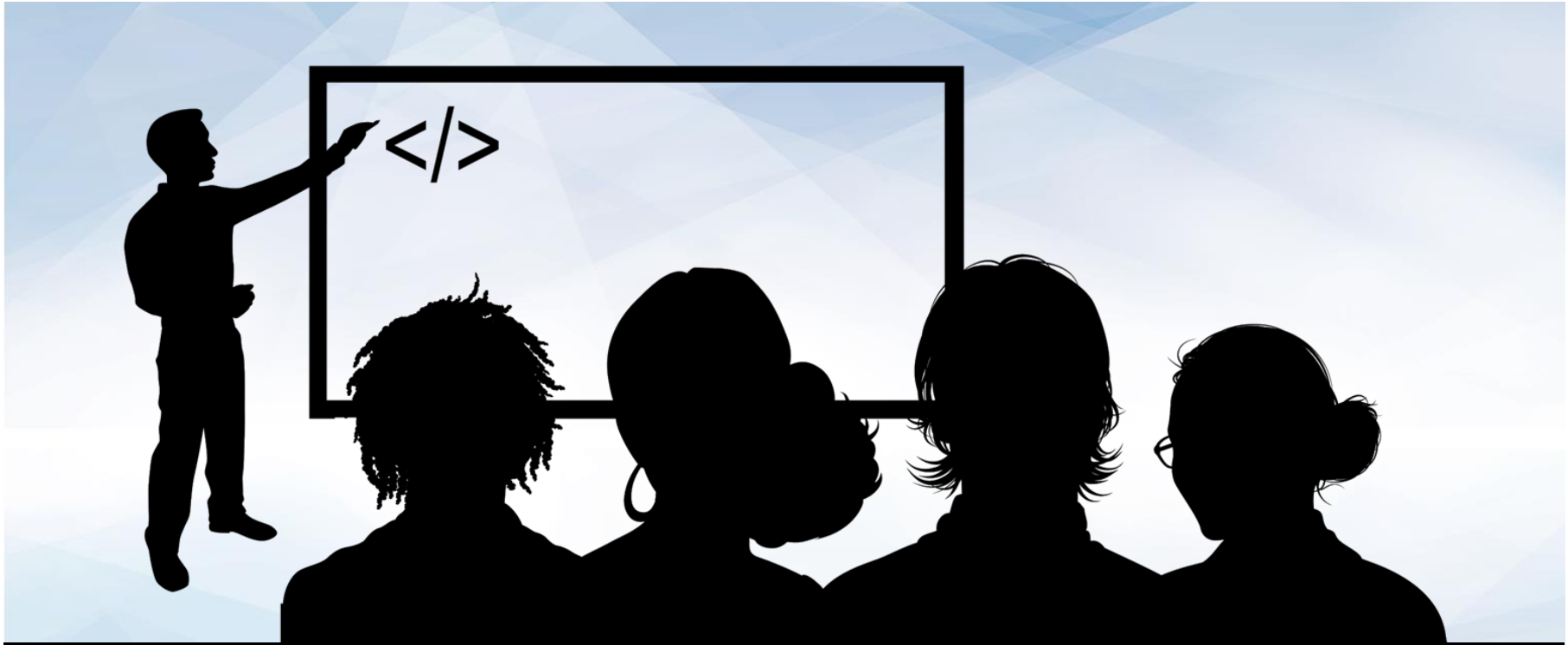
Instructor Demonstration

Recap Workflow & Share

References

Recap Workflow & Share Preferences





Instructor Demonstration

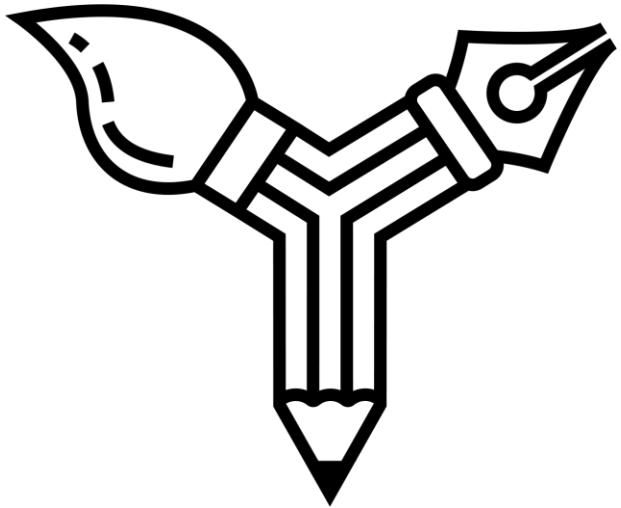
Introduce

Projects

Project Week! (This Week)

Day 1:

- Form Groups
- Outline Project Ideas
- Initial Data Exploration
- Begin Research of Datasets
- Submit Project Proposal for Approval



Day 2:

- Hardcore Development

Day 3:

- Hardcore Development

Project Week! (Next Week)

Day 4:

Hardcore Development

Day 5:

Presentation Prep

Polish Project

Day 6:

Presentations

Time to divide into teams!

Project Requirements

Development Requirements

- Use Pandas to clean and format your dataset(s).
- Create a Jupyter Notebook **describing the data exploration and cleanup** process.
- Create a Jupyter Notebook **illustrating the final data analysis**.
- Use Matplotlib to create a total of 6–8 visualizations of your data (ideally, at least 2 per ”question” you ask of your data).
- Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation.
- (Optional) Use at least one API, if you can find an API with data pertinent to your primary research questions.
- Create a write-up summarizing your major findings. This should include a heading for each “question” you asked of your data and a short description of your findings and any relevant plots.

Presentation Requirements

You will also be responsible for preparing a formal, 10-minute presentation that covers:

- Questions you found interesting and what motivated you to answer them
- Where and how you found the data you used to answer these questions
- The data exploration and cleanup process (accompanied by your Jupyter Notebook)
- The analysis process (accompanied by your Jupyter Notebook)
- Your conclusions including a numerical summary and visualizations of the summary
- The implications of your findings: what do your findings mean?

Suggested Data Sources

Suggestions for Data Sources

Feel free to ask us (the instructional staff) for input, but our general advice is to stick to data sources that:

Are sufficiently large.

Have a consistent format.

Ideally, contain more data than needed.

Are well-documented.

Example Project Ideas

Private Investigator

01

Use aggregate crime data from different police precincts in a city to uncover patterns in criminal activity.

02

Most crime in NYC takes place in the summer.

Can you uncover similar patterns in your city?

03

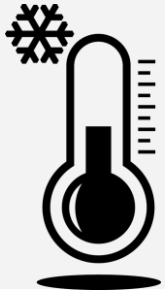
What do your results suggest about how police should plan their patrols?

What do your results suggest about how best to distribute law enforcement resources over the calendar year?

Uber Rides and Weather

01

No one likes to walk in subzero temperatures *or* scorching heat. Do people use Uber more when the weather is uncomfortable?



02

Using Uber ride data from Kaggle and data from a weather API, find out if people take Uber more during summer and winter, and if there are relationships between daily temperature and ride frequency.

03

What do the results tell you about surge pricing strategies and commuter habits?

Bullying and Crime Rates

01

Bullying and violent crime seem like they should be related. Can we find a correlation between frequency of bullying and rates of violent crime?

02

Using Data.gov's data on bullying and data from police districts of your choice, investigate relationships between bullying and violent crime frequency and location (zip code, city, etc.).

03

Are these two activities correlated?

What do the results suggest about society and public policy?

Today’s Focus

By the End of Today's Class:

- Brainstorm possible project ideas.
- Begin data research.
- Write a description of the scope of your research.

Create a short, 1-page project proposal that covers the following:

- Project Title
- Team Members
- Project Description/Outline
- Research Questions to Answer
- Datasets to Be Used
- Rough Breakdown of Tasks

Questions?

TRADING PLATFORM
ETH USD LTC
190.34 +3.44 +1.88%
199.31 +0.31 +0.05% 19:55 Jun 01.06
SELL
BUY
Sales
UP 23%
DOWN 14%
Month
190.34

The End