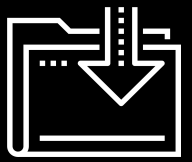
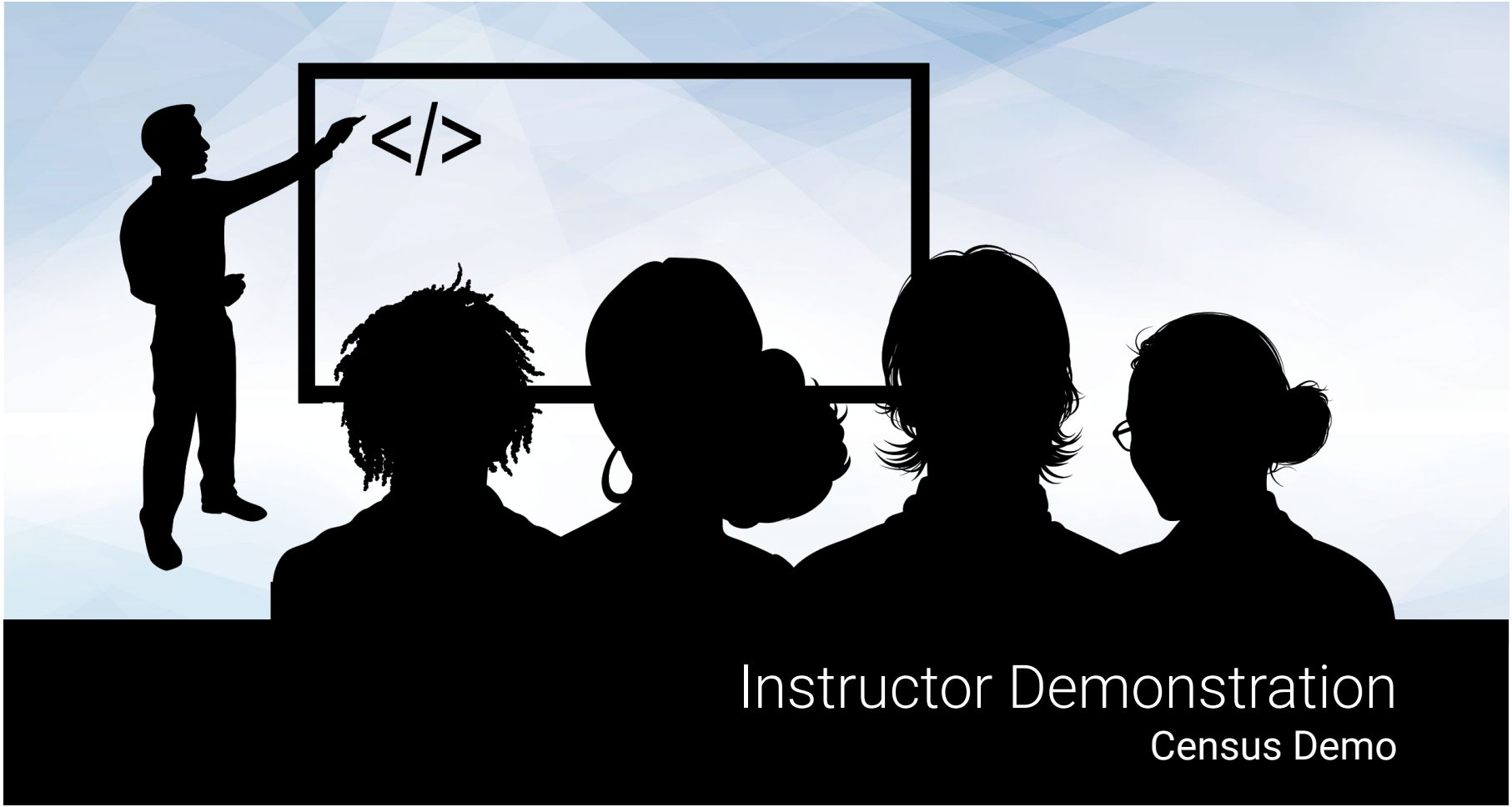
Projects & Collaboration with Git

Data Boot Camp

Lesson 7.1

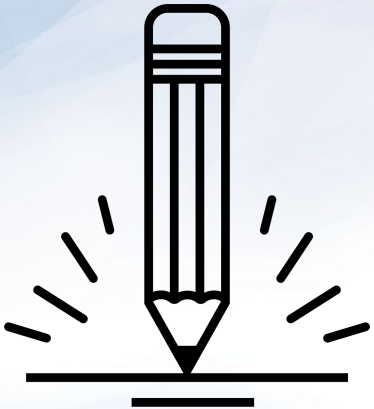




Instructor Demonstration

Census Demo

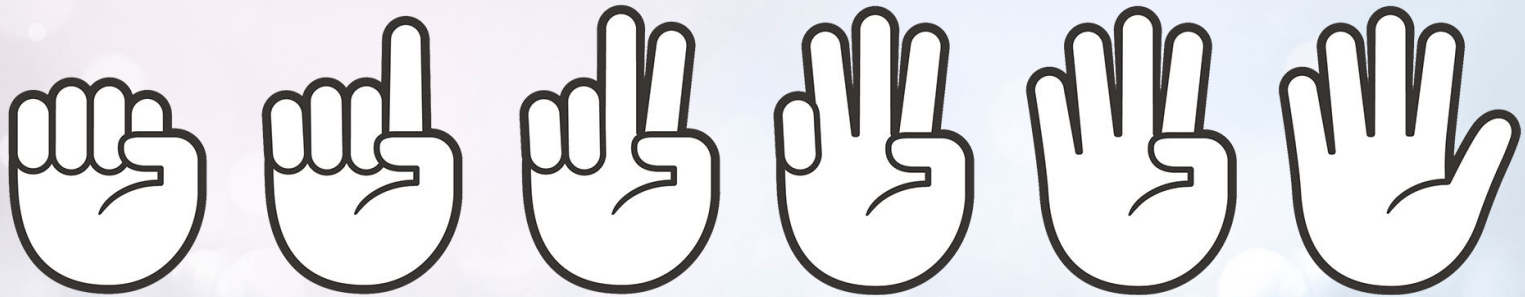
Census Activity



Instructions:

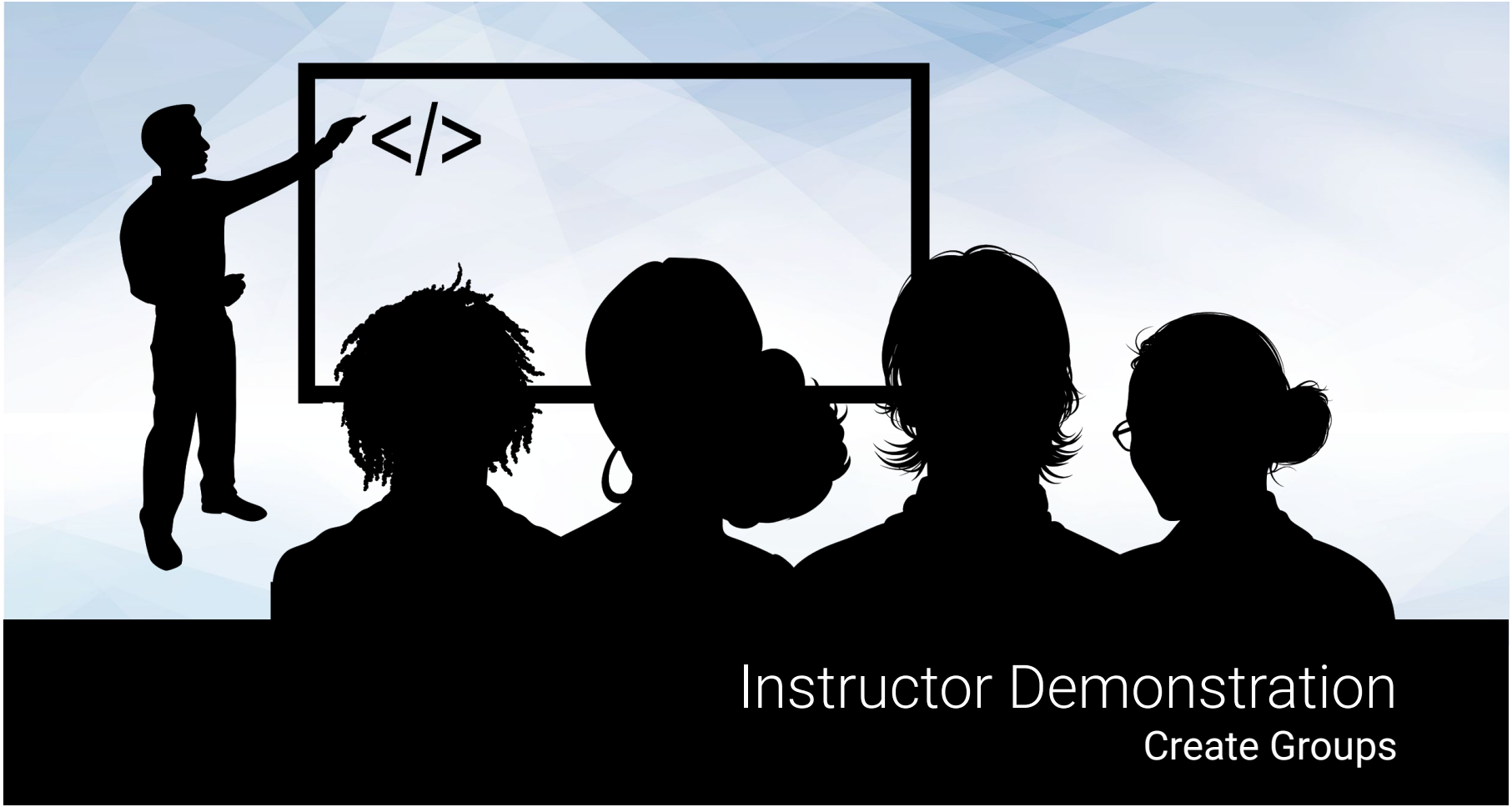
- Using 09-Stu_Census/Census_States.ipynb as a reference, create a completely new script that calculates each of the following fields at the state level:
  - Population
  - Median Age
  - Household Income
  - Per Capita Income
  - Poverty Count
  - Poverty Rate
  - Unemployment Rate
- Save the resulting data as a csv.
- Next, read in the provided csv containing state centroid coordinates and merge this data with your original census data.
- With the coordinates now appended to the dataframe, you have the ability to add markers to a base map, using the 'Poverty Rate' data to populate the info box.

Suggested Time:
15 minutes



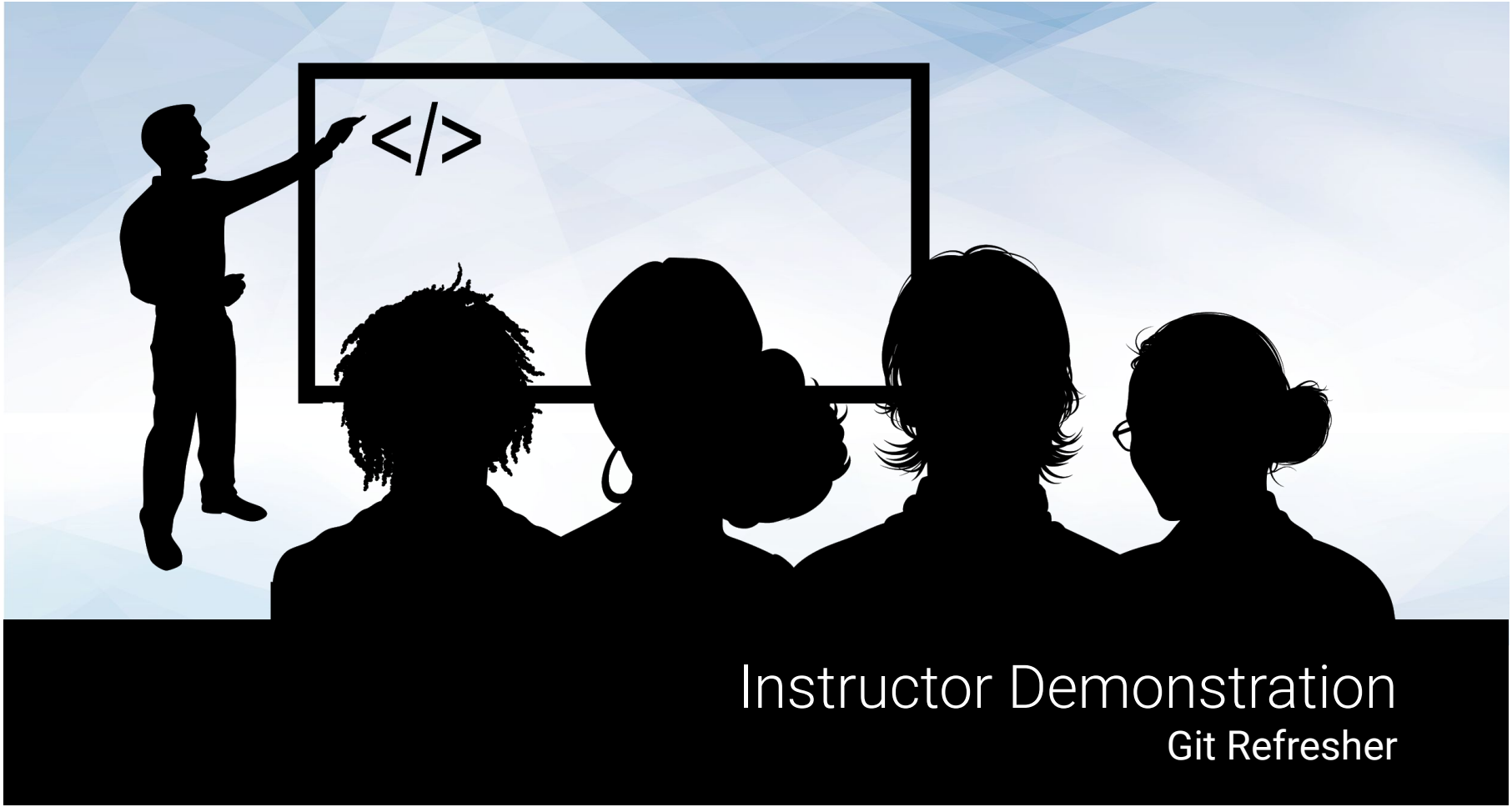
FIST TO FIVE:

APIs



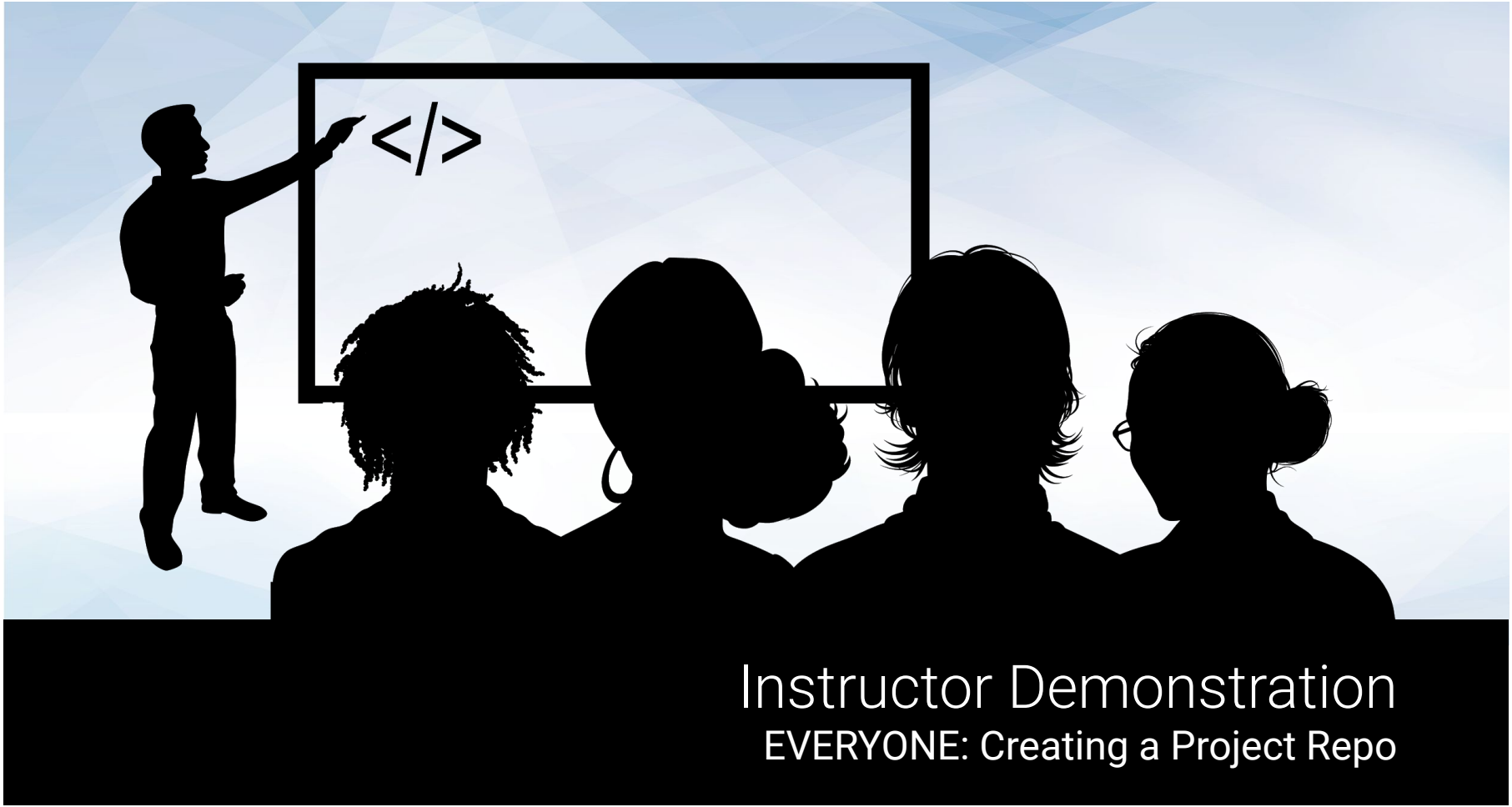
Instructor Demonstration

Create Groups



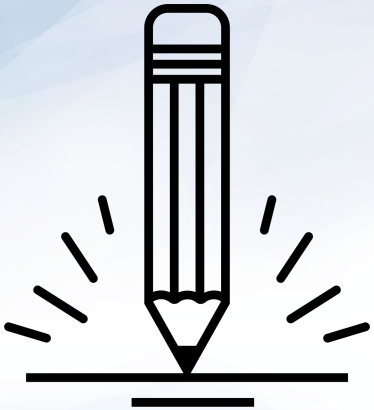
Instructor Demonstration

Git Refresher



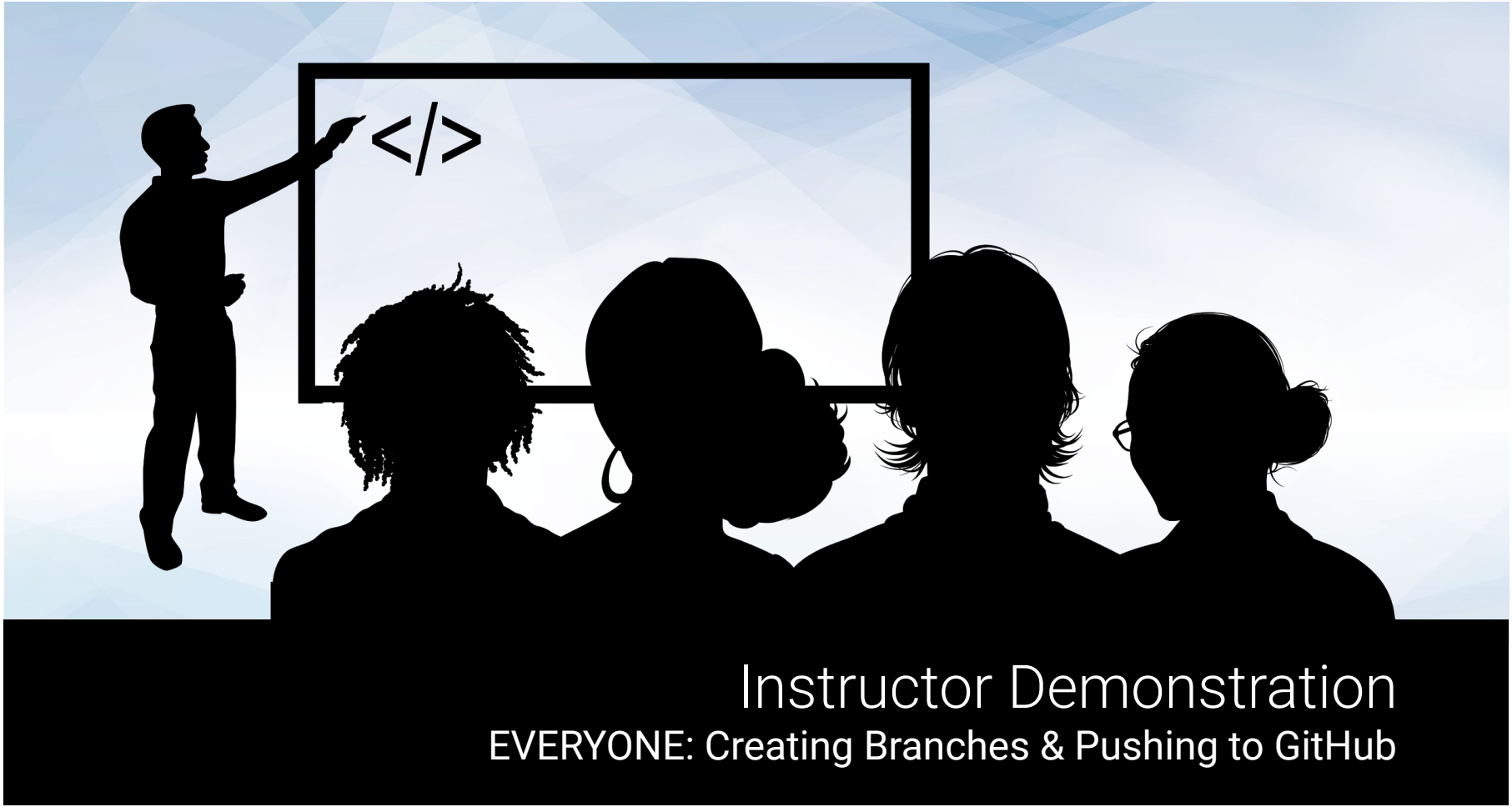
Instructor Demonstration

EVERYONE: Creating a Project Repo



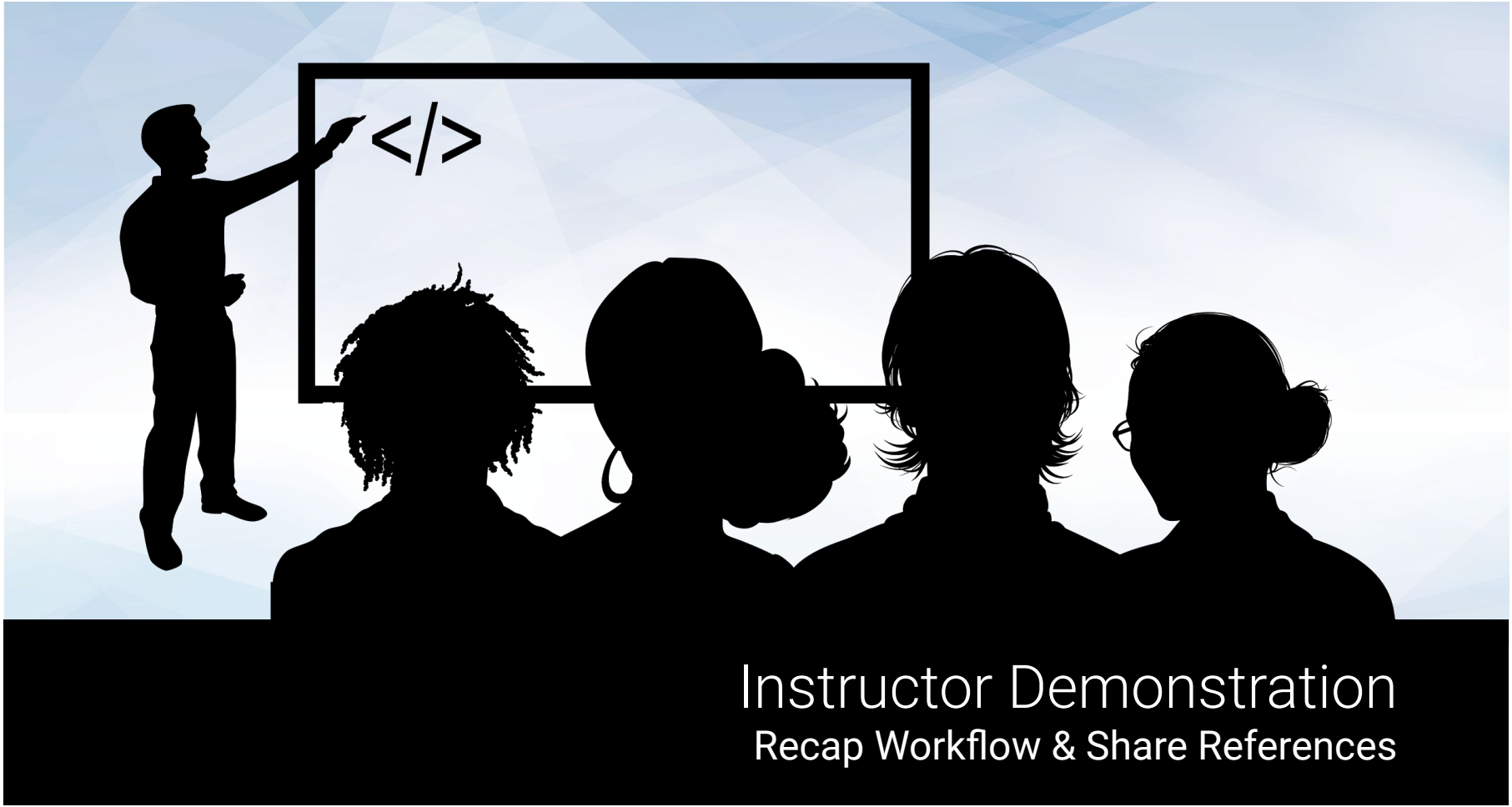
Workflows Exercise

Suggested Time:
5 minutes



Instructor Demonstration

EVERYONE: Creating Branches & Pushing to GitHub

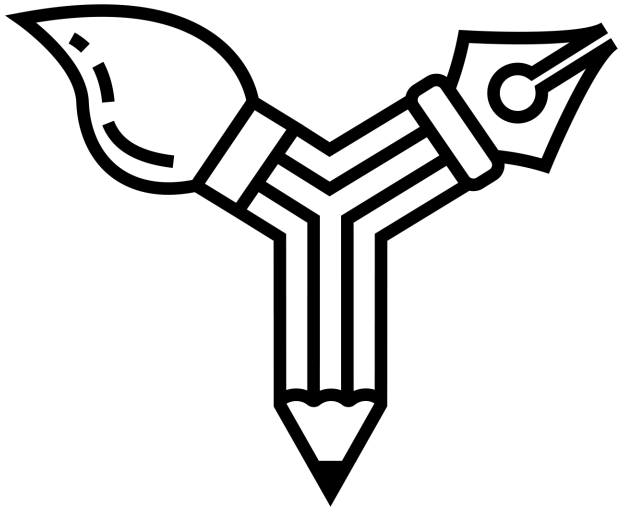

Instructor Demonstration
Recap Workflow & Share References

Time to divide into teams!

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:

Hardcore Development

Presentation Prep

Day 6:

Presentations

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, which should include a numerical summary and visualizations of that 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

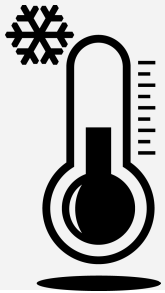
What do your results suggest about how police should plan their patrols?

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

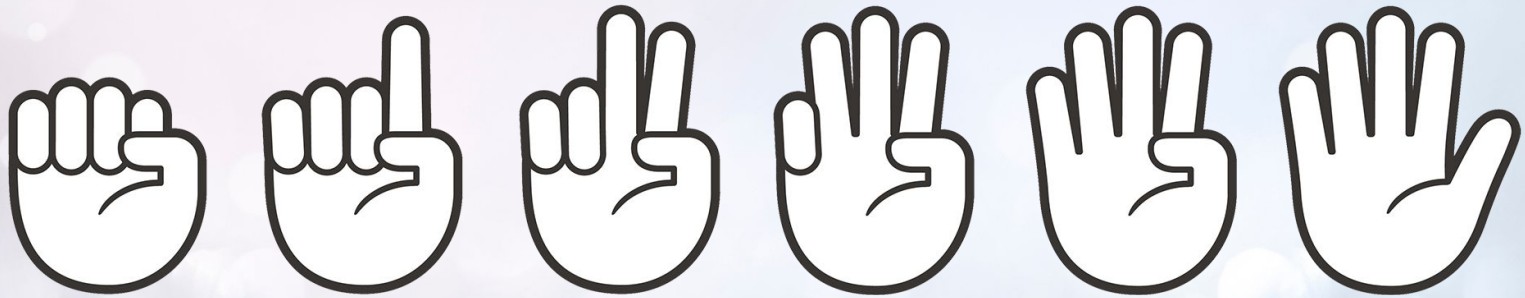
By the End of Today's Class:

- 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?



FIST TO FIVE:

Project 1

Take a Break!