

COGS 108 Notes

1 9/27

Class covers Statistics and Programming portions of data science, as well as ways to think about data science Project:

Formulate a data science project, explain and carry out analyses in Python, communicate clearly, and learn new tools

Python Data Science Handbook can be useful

2 9/30

Encouraged to choose a topic that matters to the world (family, neighborhood, etc.)

Avoid things where there are a lot of examples online

Version Control:

Enables multiple people to simultaneously work on a single project

Each person edits their own local copy and chooses whether to share those changes

A way to manage the evolution of a set of files

The set of files is called a repository

git is the version control system – software

GitHub is the home where git-based projects live on the Internet

git for: Collaboration: make local changes with a stable copy

Exposure: Other people can see the work you do

Tracking: Can see other people's work :)

Git: Clone the repository

Add files to keep track of (staging)

Options: -u stages modified and deleted files, -A stages new, modified, and deleted files, * stages everything

Commit files (takes a snapshot)

Push changes to remote so other people have access

Pull other people's changes

status to get status of local copy