

GWSB ISTM 6212-10

Data Management for Analytics - week 1

Daniel Chudnov, instructor

2016-08-30

Agenda

- ❖ Introductions
- ❖ Purpose and Syllabus
- ❖ Schedule
- ❖ Jupyter notebooks
- ❖ Git and GitHub
- ❖ Exercise 01

About me

- ❖ BA '94, Economics, U of Michigan
 - ❖ MS '97, Information, U of Michigan
 - ❖ MS '15, Business Analytics, GWSB
 - ❖ 20 years as software developer & librarian:
Michigan, Yale, MIT, Library of Congress, GWU
 - ❖ Independent consultant
 - ❖ Faculty, District Data Labs
 - ❖ Reading, music, sports, baking
- at SciFoo '07 (ptufts)



About you

Purpose and Syllabus

Learning Objectives

- ❖ Develop theoretical foundation and practical experience working with a variety of traditional and contemporary data management tools, enabling students to work productively with any product or toolkit they might encounter
- ❖ Gain skill in wrangling and exploring data with a variety of tools inside and outside of databases
- ❖ Understand and be able to develop, deliver, and review reproducible data analyses

Elevator pitch

- ❖ Be capable and confident in taking data from any source, wrangling it into any tool, preparing to analyze it using descriptive, predictive, and prescriptive methods.

Developing your craft

- ✦ Do this work reliably, repeatably, and reproducibly

Syllabus

Schedule

Overview

- ❖ Four three-week segments: shell, SQL, warehouse, schemaless
- ❖ Readings every week
- ❖ Short exercises every week: start in class, finish by Friday
- ❖ Projects every third week: 10 day deadlines
- ❖ Reviews for every project

Assignments

- ❖ Exercises and projects in Jupyter notebooks
- ❖ Submit to both Blackboard (for record keeping) and GitHub (for sharing, grading)
- ❖ Automated notebook grading
- ❖ Reviews using Github comments

Recommendations

- ❖ Stay on top of readings
- ❖ Use class time wisely - focus on lectures, start exercises
- ❖ Start projects early

Schedule - will change over semester

(Break)

Jupyter notebooks

Git and GitHub

Git

- ❖ Widely used distributed version control tool
- ❖ "Distributed" == multiple people, multiple machines, multiple copies
- ❖ "Version control" == keeps track of changes over time
- ❖ Key concepts: **add** files to git, **commit** changes, **push** / **pull** / **merge** changes from one copy of the repository to another
- ❖ Multiple copies of a git repository are called **clones**

Working with Git

- ❖ Typically done with the command line
- ❖ Windows: Git for Windows (<https://git-for-windows.github.io/>)
- ❖ macOS: built-in on command line, or use homebrew
- ❖ GUI (both win/mac): GitHub Desktop (<https://desktop.github.com/>)
- ❖ **GitHub Desktop recommended**

GitHub

- ❖ Social network for Git repositories and the people who work with them
- ❖ Follow other users, clone or **fork** their repositories
- ❖ Report **issues** with code, assign, comment, resolve issues
- ❖ Send **pull requests** with your changes to other users; review, comment, merge

Exercise 01

Exercise 01 - your tasks, now

- ❖ Create a GitHub account
- ❖ Email your account name to **dchud@gwu.edu**
- ❖ Install GitHub Desktop on your machine
- ❖ Configure GitHub Desktop with your GitHub account
- ❖ Clone the course repository:

<https://github.com/gwsb-istm-6212-fall-2016/syllabus-and-schedule>



This repository

Search

Pull requests

Issues

Gist



gwsb-istm-6212-fall-2016 / syllabus-and-schedule

Unwatch 1

Star 0

Fork 0

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

Settings

Syllabus, schedule, and ancillary course materials as appropriate. — Edit

5 commits

1 branch

0 releases

1 contributor

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download



dchud Merge branch 'master' of github.com:gwsb-istm-6212-fall-2016/syllabus...



.gitignore

Initial commit



LICENSE.txt

added CC0 license



README.md

adding README



Schedule.pdf

first version of syllabus and schedule

4 hours ago



Syllabus.pdf

first version of syllabus and schedule

4 hours ago

Clone with SSH

Use HTTPS

Use an SSH key and passphrase from account.

git@github.com:gwsb-istm-6212-fall-2016



Open in Desktop

Download ZIP



README.md

GWSB ISTM 6212-10 Syllabus and Schedule

Syllabus, schedule, and ancillary course materials as appropriate.

Exercise 01 - get started

- ❖ Open up Jupyter at <http://datanotebook.org/>
- ❖ Read the Unix Shell lessons at <http://swcarpentry.github.io/shell-novice>
- ❖ Use a Jupyter terminal to follow the lessons
- ❖ Upload exercise-01.ipynb to Jupyter
- ❖ Complete the exercise
- ❖ Download your exercise-01.ipynb

Exercise 01 - finishing up

- ❖ Create your own new GitHub repository "istm-6212"
- ❖ Upload your downloaded exercise-01.ipynb to it using the GitHub web UI
- ❖ Go to Blackboard and complete the assignment there