PYTHON AND SQL REVIEW

Justin Gould (gould29@purdue.edu)

HONR 39900: Foundations of Geospatial Analytics

Fall 2021

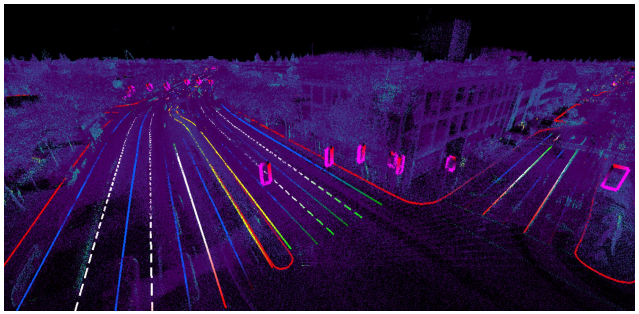
PURDUE UNIVERSITY

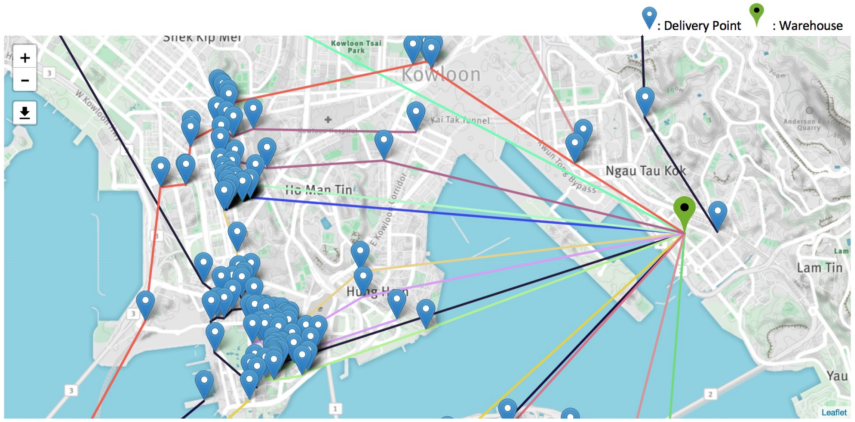
Topics

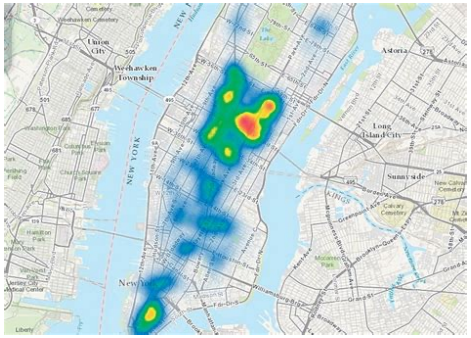
- Welcome!
- Housekeeping and course overview
- Homework Submission and Formatting
- Review Python and SQL
- Homework #1

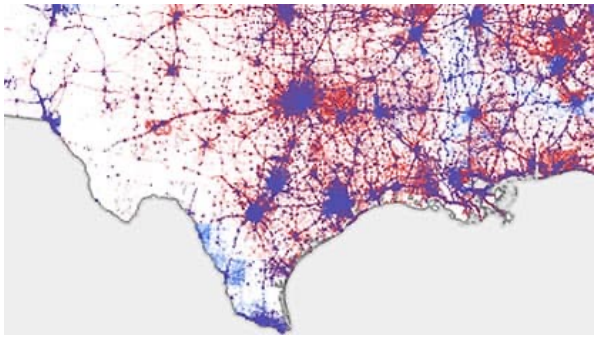
Welcome!

- Welcome to HONR 39900: Foundations of Geospatial Analytics









Who am I?

- Current Role: Senior Data Scientist in the Data Mine
  - Technical advisor to students working on corporate projects
  - Data Science consultant with faculty/staff for Fortune 100 companies
  - Research in energy-efficient distributed computing infrastructure

- Hobbies:
  - Rock and ice climbing
  - Classical music
  - Reading (https://gouldju1.github.io/gouldju1/pages/reading-list.html)

PURDUE UNIVERSITY

Who am I?

- DTE Energy
  - Route optimization and coordinate system conversion API
  - Web-based SQL query generation app (asset management)

- Stryker Orthopaedics
  - Correlate procedures which use Stryker products to demographics, and predict demographic growth in US regions for sales/lead generation

Who am I?

- Product Sustainability Analytics
  - On-road vehicle data to quantify fuel and emissions savings of engine start/stop feature
- Connected Vehicles/Driver Assist Technology (DAT)
  - Hands-free feature in Mach-E!
  - Route prediction (from key on/key off events)
  - DAT feature map generation and analysis tool
    - ETL pipeline for on-road vehicle data, build feature availability map, virtually “drive” map via predicted routes, and quantify where feature performance can be improved
  - Product Owner of CV data discovery tool

PURDUE UNIVERSITY

Who am I?

- Natural Language Processing
  - Question & Answering
  - NL2SQL (Natural Language to SQL)
    - Input table and natural language question; translate into executable SQL query
  - Product Owner of NLP products and model API catalog
    - Machine translation, Q&A, document classification
    - All NLP work reusable via microservice architecture and published in API catalog for widespread adoption
  - Business engagement
    - How can we convince Ford colleagues to “purchase” NLP models and applications

Housekeeping and Course Overview

- Required materials:
  - Required Textbook:
    - Obe, Regina O., and Leo S. Hsu. PostGIS in Action. 3rd ed., Manning, 2021.
  - Software:
    - Laptop (either MacOS or Windows is fine)
    - Python >= 3.6.x & < 3.9.x: https://www.python.org/downloads/
    - QGIS 3.16: https://www.qgis.org/en/site/forusers/index.html
    - PostGIS >= 3.1.0: https://postgis.net/install/

Housekeeping and Course Overview

- Where to find course content:
  - Brightspace: https://purdue.brightspace.com/d2l/login
  - GitHub: https://github.com/gouldju1/honr39900-foundations-of-geospatial-analytics
- Everything you need to know about the course is clearly stated in the syllabus (check Brightspace and GitHub if you lose your copy)

Housekeeping and Course Overview

- Office Location: Convergence Center
- Purdue Email Address: gould29@purdue.edu
  - Emails are read 7:00 to 17:00 EST during school days, and will be responded to within 24 hours.
  - I receive many emails during the day. It will really help me find your email if you put our class number in the subject line of the email. Thank you in advance!
- Scheduling Time with Instructor: https://calendly.com/gould29
  - If you are unable to make my office hours, please schedule a time with me via Calendly (link above). We can meet in my office, virtually via WebEx, or even take a walk around campus.
- Office Hours:
  - My dedicated office hours will be:
    - Wednesdays: 9:00-9:50 EDT/EST
    - Fridays: 8:30-9:20 EDT/EST

Housekeeping and Course Overview

- Grading scale

Grade	Points Earned
A+	965-1000
A	940-964
A-	900-939
B+	865-899
B	840-864
B-	800-839
C+	765-799
C	735-764
C-	700-734
D+	665-699
D	635-664
D-	600-634
F	<=599

Housekeeping and Course Overview

- Assignment point weights

Assignments	Due	Points
Homework (Interactive Labs/“Notebook”-Style Assignments)	23:59 EST/EDT the day before our class; ongoing	**250**
Midterm Project	10/11 at 23:59 EDT	**300**
Final Project	12/13 at 23:59 EST	**450**
TOTAL		**1000**

Housekeeping and Course Overview

- Assignment descriptions:
  - **Homework** (10 assignments at 25 points each, for a total of 250 points; ongoing, but due at 23:59 EST/EDT the day before our class): These will be interactive assignments, such as Jupyter notebooks, or .py files which must pass a series of unit tests. Grades will be provided based on passing unit tests, accuracy, and completing the assignment, per the provided instructions.
  - **Midterm project** (300 points; due October 11 at 23:59 EDT): Details about this project are in Brightspace under Assignments, including the grading rubric. A presentation will be given in class on October 18.
  - **Final project** (450 points; due December 13 at 23:59 EST): Details about this project are in Brightspace under Assignments, including the grading rubric. A final presentation will be given in class during finals week.

Housekeeping and Course Overview

- Missed work policy:
  - **Homework**: If you miss a due date, and I do not hear from you before the prescribed due date, you will receive a 0 for that assignment, and there will be no makeup work available.
  - **Projects**: If you miss a due date, and I do not hear from you before the prescribed due date, your project portion of the grade (**not** the presentation portion) will reduce 100 points every day the project is not submitted.

Housekeeping and Course Overview

Week	Date	Topic(s)	Assignment(s) Due
Week 1	August 24	• Welcome! • Review Python and SQL	• N/A

Week 2	August 31	• What are maps?
		  ○ Map design principles
		• Basic map creation and visualization
		  ○ Preprocessing spatial data
		  ○ Python and QGIS
			• HW #1 due 08/30 23:59 EDT

Week 3	September 07	• Basic map creation and visualization
		  ○ Preprocessing spatial data
		  ○ Python and QGIS
			• HW #2 due 09/06 23:59 EDT

Week 4	September 14	• Introduction to PostGIS ○ Basic map manipulation	• HW #3 due 09/13 23:59 EDT
Week 5	September 21	• Advanced PostGIS	• HW #4 due 09/20 23:59 EDT
Week 6	September 28	• Advanced PostGIS	• HW #5 due 09/27 23:59 EDT
Week 7	October 05	• Class time to work on midterm project	• HW #6 due start of class 10/05
Week 8	October 12	• Midterm project: public health	• Midterm project due 10/11 23:59 EDT • Midterm presentation during class

*Subject to change; see syllabus

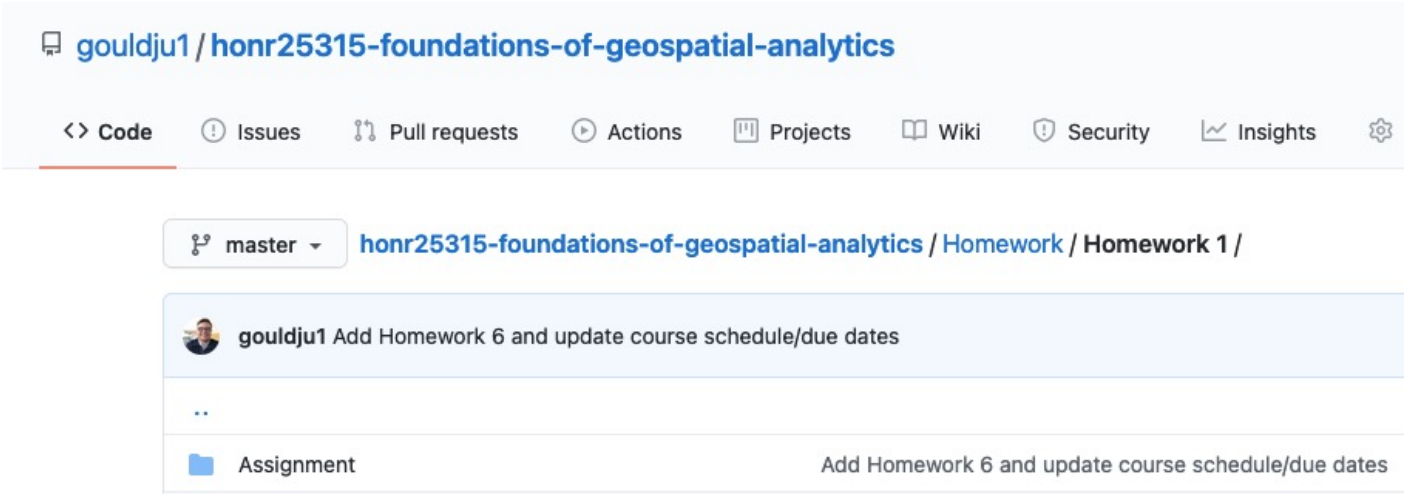
Housekeeping and Course Overview

Week	Date	Topic(s)	Assignment(s) Due
Week 9	October 19	• Midterm project: public health	• Midterm presentations during class
Week 10	October 26	• Using "dirty" spatial data ○ Map matching algorithms	• N/A
Week 11	November 02	• Using "dirty" spatial data ○ Map matching algorithms	• HW #7 due 11/01 23:59 EDT
Week 12	November 09	• Routing and optimization	• HW #8 due 11/08 23:59 EST
Week 13	November 16	• Routing and optimization	• HW #9 due 11/15 23:59 EST
Week 14	November 23	**NO CLASS – THANKSGIVING BREAK**	• HW #10 due 11/22 23:59 EST
Week 15	November 30	• Class time to work on final project	• N/A
Week 16	December 07	• Semester wrap-up • Final project: autonomous vehicles	• Final presentation during class
Finals Week	December 14	• Final project: autonomous vehicles	• Final project due 12/17 23:59 EST • Final presentation during class

*Subject to change; see syllabus

Homework Submission and Formatting

- Where to find homework: https://github.com/gouldju1/honr39900-foundations-of-geospatial-analytics/tree/master/Homework
  - Includes PDF of assignment requirements
  - Boilerplate submission .py file



Homework Submission and Formatting

HONR 25315 – Homework SAMPLE

Justin A. Gould
gould29@purdue.edu

March 1, 2021

Homework Instructions
To receive credit for the assignment, do the following:

1. Create a .py file, and name it: purduealias_honr25315_homework_number.py
   (e.g., gould29_honr25315_homework_1.py)

2. Create a function for each problem, accepting the input and providing the desired output (both of which will be defined in the homework assignment).
   (e.g., def problem_1() for Problem #1)

3. Submit the .py file to Brightspace by the due date.

For grading, I will leverage unit tests, to ensure you aren't hard-coding your work. These unit tests are hidden. To test your code, I suggest using a Jupyter Notebook to ensure you're following directions. An example .py file is on our Brightspace and GitHub.

Problem 1

Python Basics: Addition – 5 points
Given a non-negative integer, add one to it.

Input: Any non-negative integer, n.

Desired Output: An integer, $n + 1$.

Problem 2

Python Basics: Sorting a List – 5 points
Given a list of items, either strings or integers, sort the list (ascneding).

Input: A list of items (strings or integers).

Desired Output: A sorted list (ascending).

Problem 3

Python Basics: Sorting a List – 10 points
Given a float, round it to 3 digits.

Input: A float.

Desired Output: A float, rounded to 3 digits.

Homework Submission and Formatting

- The image on the right is my submission file for the sample homework assignment:
  - Gould29_honr39900_homework_sample.py

- Each problem has its own function, following specified input/output requirements (this is ***critical*** to receive credit)

```
"""
Name: Justin A. Gould
Date: 2021-02-10
"""

#Required Packages
import numpy as np

#Problem 1
def problem_1(x):
    return x + 1

#Problem 2
def problem_2(ls):
    ls.sort()
    return ls

#Problem 3
def problem_3(x):
    return np.round(x, decimals=3)
```

Homework Submission and Formatting

- The image on the top shows the test cases for the sample homework assignment:
  - This file will be hidden for all other assignments

```python
#Problem 1
def test_problem_1():
    assert problem_1(1) == 2
    assert problem_1(2) == 3
    assert problem_1(3) == 4

#Problem 2
def test_problem_2():
    assert problem_2(["Ford", "BMW", "Volvo", "GM"]) == ["BMW", "Ford", "GM", "Volvo"]
    assert problem_2([0,3,1,5,6,2,5]) == [0, 1, 2, 3, 5, 5, 6]
    assert problem_2(["Tesla", "Apple", "Microsoft", "Dell", "Google"]) == ["Apple", "Dell", "Google", "Microsoft", "Tesla"]

#Problem 3
def test_problem_3():
    assert problem_3(3.14159265358979) == 3.142
    assert problem_3(7.17937876452) == 7.179
    assert problem_3(10.742309745) == 10.742
```

- The image on the bottom shows how your homework is evaluated, via a unit testing framework

```
============================= test session starts ==============================
platform darwin -- Python 3.7.9, pytest-4.0.0, py-1.10.0, pluggy-0.13.1
rootdir: /Users/gould29/OneDrive - purdue.edu/GitHub/honr-foundations-of-geospatial-analytics/Homework/0_SAMPLE, inifile:
plugins: anyio-2.1.0, notebook-0.6.1
collected 3 items

test_sample.py ...                                                       [100%]

============================== 3 passed in 0.73 seconds ==============================
```

Homework Submission and Formatting

- To receive credit, outside of the strict adherence to formatting in the aforementioned slides, you must:
  - Submit the .py file to Brightspace **before** the given due date (23:59 EST/EDT the day before our class)

- Tips and comments:
  - Start early! I am willing to answer questions (to an extent), so be proactive.
  - Leverage Jupyter Notebooks, PyCharm, etc. to work on your assignments to ensure your code behaves as expected
  - Please, please, please follow the input/output instructions in the assignment **exactly**
    - Example: if an integer is expected as output, but you return str(10) instead of int(10), you will **not** receive credit for that question!

Review Python

- To be successful in this class, Python skill are required. Every homework assignment is to be written in Python.

- To review in our class today, we will create the conda environment for our course and run through a couple exercises!

- Course virtual environment

1. Open your terminal (Mac) or the Anaconda command prompt (Windows)
2. Run: `conda create env --name honr39900 pip`
    1. This will create a virtual environment called honr490 with the `pip` package
3. Download `requirements.txt` from our GitHub under `./env`
4. Run: `conda activate honr39900`
5. `cd` to where `requirements.txt` is located and run: `pip install –r requirements.txt`

Review Python and SQL

- Once installed, let's go through some basic techniques…run: `jupyter lab` to open a jupter lab session
  - We will leverage Jupyter Lab to run our lecture/in-class experience notebooks
  - Today's notebook is located here: https://github.com/gouldju1/honr39900-foundations-of-geospatial-analytics/tree/master/Lectures/Week%201
- Supplementary review materials:
  - Python review link: https://developers.google.com/edu/python/introduction
  - SQL review link: https://www.w3schools.com/sql/

Homework #1

- Available on GitHub: https://github.com/gouldju1/honr39900-foundations-of-geospatial-analytics/tree/master/Homework/Homework%201

- Due 2021/08/29 at 23:59 EDT via Brightspace