Welcome to Data Structures Honors: *A Data-centric Dive into Data Science*

Course Website: https://courses.engr.illinois.edu/cs225/sp2019/pages/honors/

…or from the CS 225 Website -> "Course Info" -> "Honors Section"

Course Overview

Spring 2019 Theme: *"A Data-centric Dive into Data Science"*

The highlight of this course will be **data**; we will be looking to create meaningful and impactful data discovery.

- Data analysis will be done primarily in Python.
- Data visualization will be done using d3.js (JavaScript).
- We will often relate data structures back to CS 225.

Instructor: Wade Fagen-Ulmschneider (waf@illinois.edu)

Weekly Meeting: Thursdays, 5:00pm - 5:50pm (1404 Siebel Center)

Weekly Experiences:

- [Always]: Lecture w/ Data
- [Usually]: Homework or Project

Coursework and Grading

Your coursework for CS 296-25 is completely independent from your work in CS 225. You will submit it to a different repository, your grades in CS 225 do not impact CS 296 and your grades in CS 296 do not impact CS 225 in any way.

There will be only two assignments:

- **500 points**: Weekly Homeworks:
    - Smaller assignments to dive into data and learn the language/tools
    - Approx 5 homeworks, equally weighted among 500 points
- **500 points**: Projects:
    - "Meta" Projects (200 points)
    - Final Project (300 points)

Final CS 296-25 Grades

Your final course grade is determined by the number of points you earned during the semester:

Points	Grade	Points	Grade	Points	Grade
Exceptional	A+	[930, 1000)	A	[900, 930)	A-
[870, 900)	B+	[830, 870)	B	[800, 830)	B-
[770, 800)	C+	[730, 770)	C	[700, 730)	C-
[670, 700)	D+	[630, 670)	D	[600, 630)	D-
		(600, 0]	F		

Python Pandas: Reading a CSV

Read a CSV from the same directory, saving the contents as a **DataFrame** df:

Python:	```import pandas as pd
df = pd.read_csv('file.csv')```

Pandas Cheat Sheet

The pandas organization provides a tool that nearly every data scientist has on their desk -- the **pandas cheat sheet**! This document provides a basic overview of the pandas library in a clean, organized format.

Distance Metrics

Consider two points is k-dimensional space. How do we determine how far apart they are from one-another?

$$< \_\_\_\_\_ , \_\_\_\_\_ , \_\_\_\_\_ > \quad \Leftrightarrow \quad < \_\_\_\_\_ , \_\_\_\_\_ , \_\_\_\_\_ >$$
$$\text{dist?}$$

Distance Metric #1:	
Distance Metric #2:	
Distance Metric #3:	
Distance Metric #4:	
Distance Metric #5:	
Distance Metric #6:	