



### Welcome to CS 205: “Data Driven Discovery”

Course Website: <https://courses.engr.illinois.edu/cs199205/>

...or just Google Search: **cs205 uiuc**

**Description:** This course will expand on the basic programming you have learned in an introductory programming course, specifically focusing on data discovery, analysis, and visualization.

**Pilot Course:** CS 205 is a pilot course this semester, everything is subject to change. You will get to shape the experience and development of this course!

#### Required Materials:

- Laptop, brought to class every day
- Colored pencils (for visualization drafts)

#### Semester at a Glance:

Meta	Major Topics
#1: Structured Data	CSV/TSV Datasets Classifiers Web-based APIs
#2: Images & Natural Language	Color representation, histograms Clustering NLP (grammar trees, similarity)
#3: Geographic & Social	Heat Maps Graphs (Dijkstra’s, TSP) Social Network Trees
#4: Integrated	Solving big problems in small pieces

#### Coursework and Grade

- Assignments nearly every day (560 points)
- Projects (440 points):
  - 3 Meta Projects (3 projects x 100 pts = 300 points)
  - Final Project (140 points)
- No exam, no final exam, no scantrons!
- Standard breakpoints; available on the course website.

#### CSV File Format

The first few weeks of data will come to us in the form of CSV files, which stands for Comma Separated Values. Any file that can be opened in a spreadsheet application (Excel, Google Sheets, etc) can be saved as a CSV file.

As a file, CSVs store each row of the spreadsheet as one line of the file. Each column in the spreadsheet is separated in each row by a comma. For example:

	A	B	C	D	E	F
1	CRN	Course Subject	Course Number	Course Title	Course Section	Sched Type
2	41758	AAS	100	Intro Asian American Studies	AD1	DIS
3	47100	AAS	100	Intro Asian American Studies	AD2	DIS
4	47102	AAS	100	Intro Asian American Studies	AD3	DIS
5	51248	AAS	100	Intro Asian American Studies	AD4	DIS
6	51249	AAS	100	Intro Asian American Studies	AD5	DIS
7	51932	AAS	100	Intro Asian American Studies	AD6	DIS
8	56496	AAS	100	Intro Asian American Studies	B	LCD
9	53947	AAS	211	Asian Americans and the Arts	A	LCD

File in Microsoft Excel

...is saved as the following CSV...

1	CRN, Course Subject, Course Number, Course Title, Course Section, Sched Type
2	41758, AAS, 100, Intro Asian American Studies, AD1, DIS
3	47100, AAS, 100, Intro Asian American Studies, AD2, DIS
4	47102, AAS, 100, Intro Asian American Studies, AD3, DIS
5	51248, AAS, 100, Intro Asian American Studies, AD4, DIS
6	51249, AAS, 100, Intro Asian American Studies, AD5, DIS

#### CSV Files as a Dictionary/Object

In CS 205, we will always pre-clean our CSV files so that the first row of data will always be column headers. By ensuring this condition, we will set up our data so we can access each row by its header. For example:

```
1 subject = row["Course Subject"]
```

However, before we can actually program this, we need to get Python! ...we’ll pick up here on Thursday!