M2-03: Grouping Data in Python

Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: *https://discovery.cs.illinois.edu/m2-03/*

New Dataset: Illinois GPA Dataset

A new dataset we'll explore is the GPAs of nearly every course at Illinois. This dataset is similar to the "Course Explorer" dataset that lists every course offered, but this data lists less about the course and details the number of "A+", "A", "A-", "B+", and so on given by every instructor:

Year	Term	Subject	Number	Course Title	A+	A	A-	...	D-	F	W	Primary Instructor
2019	Spring	AAS	100	Intro Asian American Studies	2	13	4	...	0	2	0	Espiritu, Augusto F
...												
2018	Fall	CS	225	Data Structures	12	205	24	...	5	9	2	Fagen-Ulmschnei, Wade A
2018	Fall	CS	225	Data Structures	13	203	17	...	6	12	7	Fagen-Ulmschnei, Wade A
...												
2019	Spring	STAT	100	Statistics	84	72	38	...	5	7	12	Flanagan, Karle A
2019	Spring	STAT	100	Statistics	114	99	53	...	7	4	6	Flanagan, Karle A
...												

Dataset URL: https://waf.cs.illinois.edu/discovery/gpa.csv

There's a few interesting bits in this dataset:

- The grade of "W" is a "Withdraw" and we will ignore it for all of our purposes.
- Grades are split at The University of Illinois **by section**, **not by course**. This means that there may be multiple entries for a single course in one semester. You can see this where both CS 225 (Fall 2018) and STAT 100 (Spring 2019) are listed twice.
- At The University of Illinois, letter grades contribute to a weighted average known as a Grade Point Average or GPA. An "A" is worth 4.00, "B" is worth 3.00, "C" is worth 2.00, "D" is worth 1.00, and can be modified with a "+" (giving an extra 0.33, except in the case of an A+) or a "-" (giving a reduction of 0.33). Here's the full table:

Letter Grade	GPA	Letter Grade	GPA	Letter Grade	GPA
A+	4.00	A	4.00	A-	3.67
B+	3.33	B	3.00	B-	2.67
C+	2.33	C	2.00	C-	1.67
D+	1.33	D	1.00	D-	0.67
		F	0.00		

M2-03: Grouping Data in Python

Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: *https://discovery.cs.illinois.edu/m2-03/*

Operations on Groups

We explored only a few operations on groups, let's discovery the results on the GPA dataset by first creating a variable called subject_group with the groups:

Initial Code:	`subject_group = df.groupby("Subject")`
Description:	

Working with the two operations we will use on groups:

Python:	`subject_group.size()`
Result:	

Working with the two operations we will use on groups:

Python:	`subject_group.describe()`
Result:	

Groups on Multiple Variables

The groupby function also accepts multiple variables as a list, similar to below:

Initial Code:	`subject_group = df.groupby(["Subject", "Number"])`
Description:	