

CSCI 127: Joy and Beauty of Data

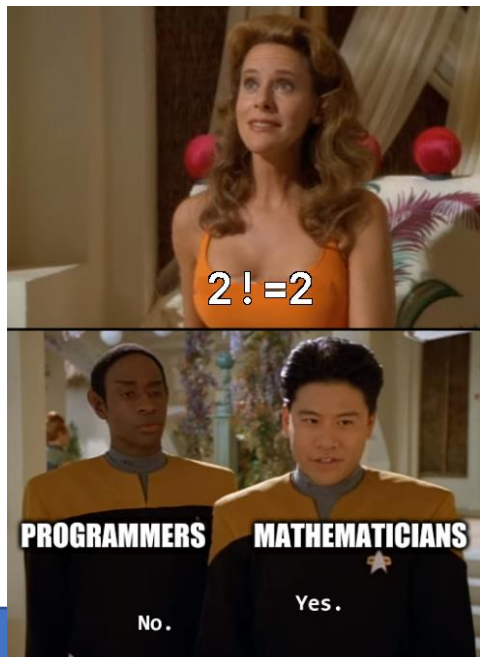
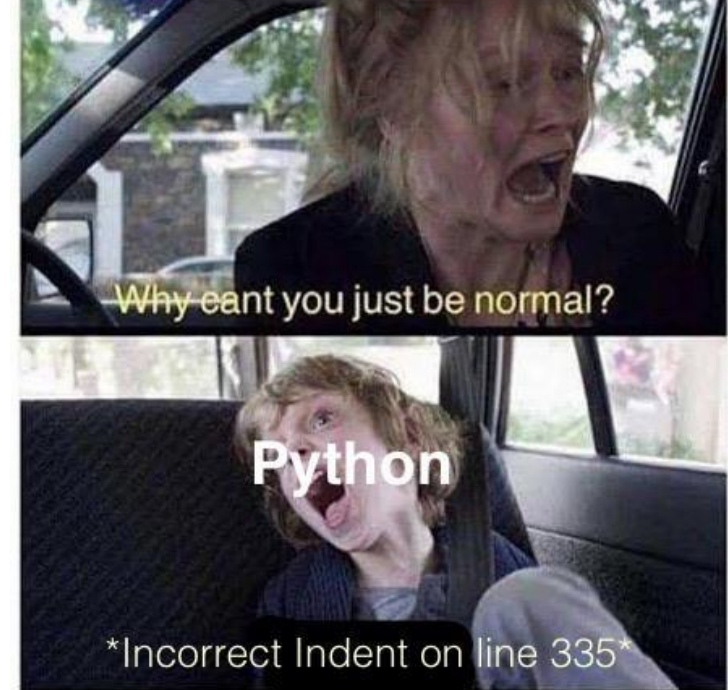
Lecture 16: Pandas

Reese Pearsall
Summer 2021

<https://reese.github.io/classes/summer2021/127/main.html>

Announcements

- No lecture tomorrow. Review for the final exam, work on program 5, lab 10, etc
- Review session in my Zoom room **tomorrow** at 2:00 PM
- Final exam on **Thursday**
- One last (short) required lecture will be posted on **Thursday** (wrap-up and conclusion)



Announcements

Lab 9 due **tonight** at 11:59 PM

- After today, you will be able to finish it (it shouldn't be too bad)

Lab 10 due **Wednesday** at 11:59 PM

- Two options
 - There will be a live, virtual review session on **Wednesday at 2:00 PM** in my Zoom room. If you attend, you will get a 10/10 for lab 10
 - OR-
 - You may do any two example exam questions from the second half of the class (or from the programming problems on the practice exam) and submit your solutions to D2L. If you attempted them, you will get a 10/10

Final exam on Thursday

Announcements

- Program 5 due **Sunday** June 20th at 11:59 PM
 - You may use Pandas on it (it will probably make things a lot easier!)
- Very open ended- You will create two graphs of your choice from a .csv file of your choice

CSV Datasets

CORGIS: The Collection of Really Great, Interesting, Situated Datasets

By Austin Cory Bart, Ryan Whitcomb, Jason Riddle, Omar Saleem, Dr. Eli Tilevich, Dr. Clifford A. Shaffer, Dr. Dennis Kafura



Search:



Aids

Records of AIDS related statistics from several countries. [View](#)

aids, death, disease, hiv, orphans, health, countries...



Broadway

This library holds data about Broadway shows, such as tickets sold. [View](#)

broadway, musical, theatre, tickets



Cars

This is a dataset about cars and how much fuel they use. [View](#)

cars, vehicles, fuel



Construction Spending

Estimates of the total dollar value of construction work done in the U.S. [View](#)

construction, spending, us, united states, america...



Airlines

Information about flight delays in major airports since 2003. [View](#)

airplane, airports, travel, plane, air, flights, delays...



Business Dynamics

The Business Dynamics Statistics (BDS) includes measures of establishment openings and closings, firm startups, job creation and destruction by firm size, age, and industrial sector, and several other statistics on business dynamics for the US. [View](#)

government, united states, us, usa, business, businesses...



Classics

Records and computed statistics about the top 1000 books on Project Gutenberg. [View](#)

classics, books, texts, text, book, classic, english...



County Demographics

Demographic information for counties in the United States. [View](#)

demographics, population, age, ethnicity, race, language...



Billionaires

Information about over 2000 billionaires from around the world. [View](#)

money, rich, wealthy, people, person, billionaire



Cancer

Cumulative cancer deaths for the period 2007-2013 are reported for each U.S. state. [View](#)

cancer, cancer deaths, medical, health



Construction Permits

This dataset provides data on the number and valuation of new housing units authorized by building permits. [View](#)

construction, permits, us, united states, america...



Covid

Since the beginning of the coronavirus pandemic, the Epidemic Intelligence team of the European Center for Disease Control and Prevention (ECDC) has been collecting on daily basis the number of COVID-19 cases and deaths, based on reports from health authorities worldwide. This helps to monitor and interpret the dynamics of the COVID-19 pandemic not only in the European Union (EU), the European Economic Area (EEA).

Important Announcement

Course evaluations are **OPEN**

EVALUATIONS CLOSE AT 11:59 PM ON THURSDAY

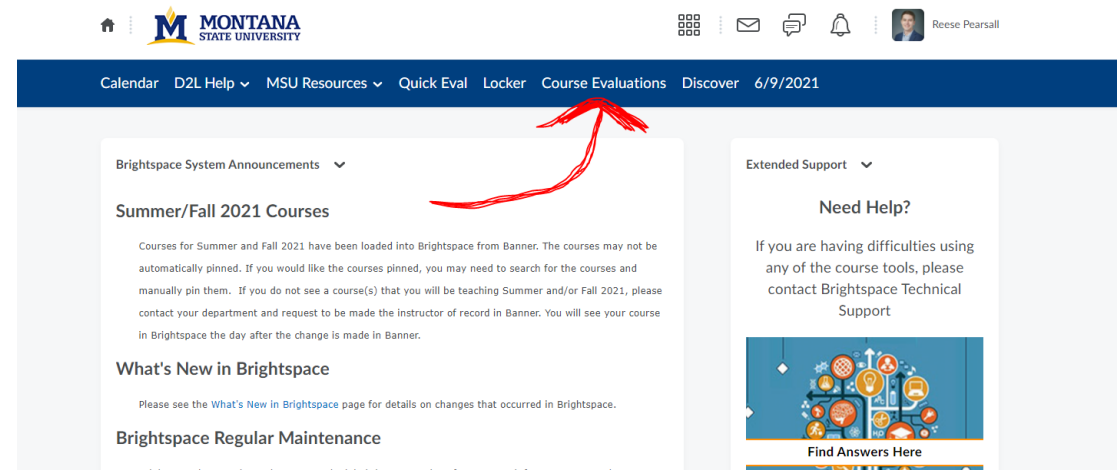
Because I am early in my teaching career,
your feedback is *very* important to me

Your feedback will help make this
class better for future students

Your feedback will also be read by the
CS department head, who is in charge
of hiring instructors 😊

If you submit a screenshot showing you completed the
course evaluation, I will add 2% to your final exam grade

If 90% of the class or more fills out the evaluation, I
will add on another bonus of some kind



Pandas

A really awesome data science module that gives lots of tools for data analysis

- We navigate through data using a **Pandas Dataframe**

We can easily read in an external file (.csv, .txt, .xls) and convert it to a Pandas Dataframe

With a Pandas Dataframe, it is very easy to

- Retrieve a specific row(s)
- Pull all data in a certain column
- Sort and aggregate data
- Plot data (using matplotlib)
- Rearrange and update data



Pandas

Examples:

Pulling and sorting data in pokemon.csv
Bar chart of top 5 strongest Water type pokemon

Failed Banks CSV file
Plot bar chart of # of failed banks by year

