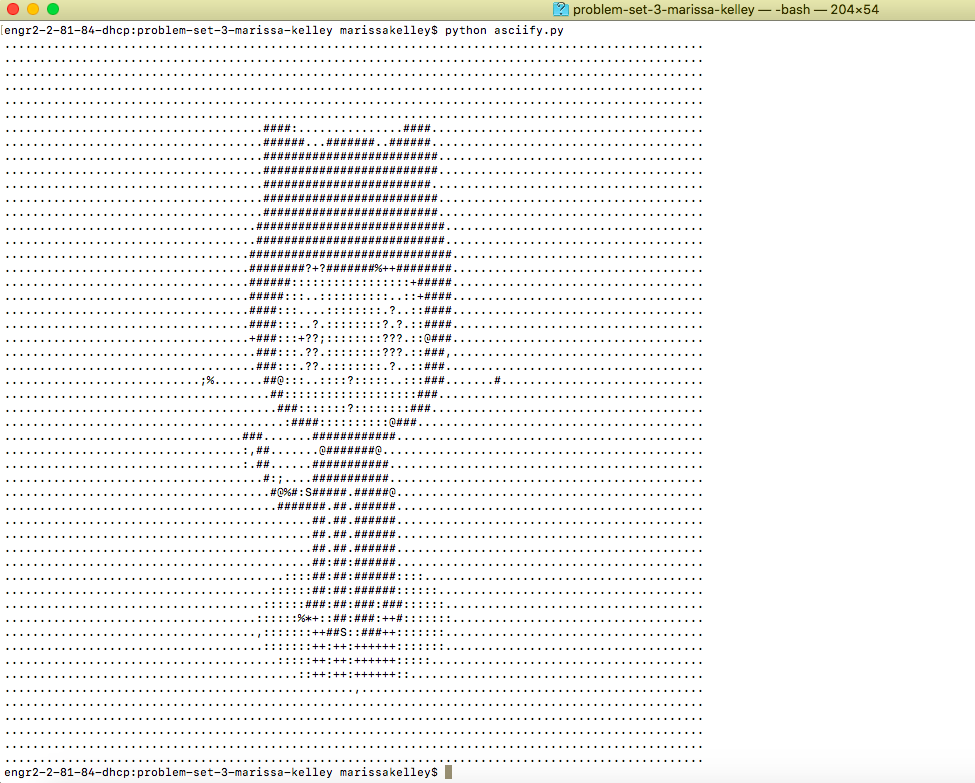
Problem Set #3

Marissa Kelley

Due: 9.16.18

1. Use the command line to run asciify.py (included in the assignment repo). What does this script do? Include a screenshot of your results.

The asciify.py script takes in an image as an input, and returns a text output that resembles the input image. In other words, it replicates an image with certain characters.



1. Install the Delorean Python package. What command did you use? Then, from the command line, launch Python. Use it to execute the following commands:

I used the command: pip install Delorean

What prints to the command line?

Delorean(datetime=datetime.datetime(2018, 9, 10, 16, 44, 33, 235224), timezone='US/Eastern')

1. Use the command line to download <http://www.colorado.edu/>. What command did you use? Make sure to include this file as part of your submission.

I used the command wget <http://www.colorado.edu>

1. Search the downloaded file from #3 for the word "Colorado." What command did you use? How many lines did this find?

I used the command grep Colorado index.html-c (grab Colorado index.html –c (counts).” It printed out a bunch of seemingly random lines. Each one is a line that contains the world “Colorado”. There were 17 lines.

Now, print the lines containing the word "Colorado" to a file to "search\_results.html". What command did you use? Include this file as part of your submission.

I used the command grep Colorado index.html > search\_results.html

**Wednesday:**

CU has recently come across a variety of findings stating that regular gym attendance can lead to a higher GPA. In an effort to understand this effect and how they might update their services and facilities to encourage students to be active, they want to turn to the data to both understand the current state of gym use, its potential effects on students, and how they might target specific facilities or programs to improve attendance.

Three data sets they could use:

\* Buff Card data

\* Gym card swiping

\* mycuhub data to determine GPA

CoNVO method:

**Context:**

**Stakeholders**: Our stakeholders include CU Boulder faculty and staff, marketing companies associated with CU, and health/fitness businesses

**What are they trying to achieve?:** They are trying to determine if there is a correlation between gym usage and a student’s GPA in order to understand if exercise can lead to a higher GPA.

**What work will the project advance?** The project would help advance the rec center or any outdoor center in order to encourage students to be more active.

**Target Audience**: CU students

**Needs:**

**What problems could be fixed using this data?** This data could demonstrate why students may not be utilizing their gym. This data would also show another way for students to improve their GPA.

**Specific utility** of the results to the organization: The specific utility would be to improve the usage of the CU rec center

**Vision:**

**What techniques could we use to address our needs?** We could gather said data, use polls/surveys in an email, ask students at the rec center or at another building on campus.

**How do our methods integrate our context?** These methods would allow us toobtain both qualitative and quantitative data. For CU to understand what the current rate of attendance and usage of their facility is it’d be smart to track how many students swipe into the gym during the week, along with how many people attend the classes, and correlate it to their GPAs. It would be personalized to CU-Boulder students.

**What will be have when we are done?** We would have information that CU resources could use to improve the rec center usage, resources for certain departments, and for marketing.

**Outcome:**

**How will our results be used?/How will new knowledge be integrated back into the organization?** The results will be used to then improve where they see issues such as attendance, figure out ways to make the gym more appealing or market classes or activities offered. Then once the changes are made it can then be looked at again to see the difference in students attending, and then try and see if there is an increase, has students GPA’s also increased.

**What will happen when we are done?** Improvements can happen, then once improvements are made, we can then try and track again to see if there is a difference.

**Who handles the data next?** It would then be given back to whoever handles CU’s rec design or marketing so then they can make changes based on the data that was revealed.

**Friday:**

In class today, you learned about a historical instance of data science in practice. Find three additional problems in history that were addressed using data. For each of these problems, describe the following:

1. What was the problem?   
2. What data was available?   
3. What methods were used?   
4. What did they find?   
5. If you were to tackle the same problem today, what would you do differently and why?

Example 1: The Hollerith Tabulating Machine (1881)

1. The problem was summarizing all of the information stored on a punch card. The tabulating machine was developed to help process data for the U.S census in 1890.
2. The data came from punch cards from passengers on a railway ticket. It recorded details like gender and age.
3. The methods used included the creation of the tabulating machine in 1881. Prior to that, it took years to process a census for one year. It combined a large staff and recording systems.
4. They found a quicker way to record census data
5. If I were to tackle the same problem today, I would use data from hospitals, such as birth records. I could also use people’s mobile devices and social medias to supplement my findings.
6. Resources: <http://www.columbia.edu/cu/computinghistory/hollerith.html>

<http://www.columbia.edu/cu/computinghistory/hollerith.html>

Example 2:

1. The problem was the rising number of deaths and sicknesses in both the US and Canada.
2. The data available was taken from the hospitals of those people, in which
3. The methods used included inspecting the plants after workers informed the FDA that the area was filthy. The FDA inspectors found mold on the plant’s ceiling and walls, insects and holes made from rodents. They also found dirty equipment.
4. They found a virulent strain of salmonella within Peanut Corp’s products.
5. If I were to tackle the same problem today, I would use the same methods and look further into the medical data of the patients.
6. Resources: <https://www.investopedia.com/financial-edge/0512/the-5-largest-food-recalls-in-history.aspx>

Example 3:

1. The problem for UPS was finding the quickest route to deliver mail
2. The data available came from maps, however the maps didn’t give the quickest routes
3. The methods used included creating a software system, called Orion, that differentiated more important left-hand turns from the normal turns. It was made to learn about parking lots, private driveways, variable speed limits and roads that were inaccessible for trucks.
4. They found that taking left hand turns took longer and used up more gas, due to more stopping at lights.
5. If I were to tackle the same problem today, I would use GPS systems in trucks to help the system learn more efficient, fuel saving routes
6. Resources: <https://www.cnn.com/2017/02/16/world/ups-trucks-no-left-turns/index.html>