

HW 01 Autograder

● Graded

Student

Sangwon Ji

Total Points

79 / 74 pts

Autograder Score

74.0 / 74.0

Passed Tests

Public Tests

Question 2

Early Submission Bonus

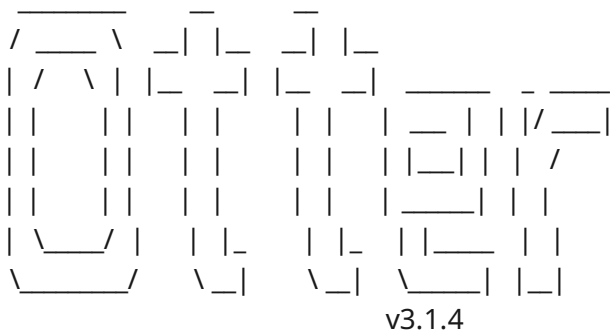
5 / 0 pts

✓ + 5 pts Early Submission Bonus

+ 0 pts No bonus

Autograder Results

Autograder Output



----- GRADING SUMMARY -----

Error encountered while trying to verify scores with log:
'TestCaseResult' object has no attribute 'hidden'

Successfully uploaded submissions for: sangwon@berkeley.edu

Total Score: 74.000 / 74.000 (100.000%)

	name	score	max_score
0	Public Tests	NaN	NaN
1	q2_1	4.0	4.0
2	q2_2	4.0	4.0
3	q3_1	4.0	4.0
4	q3_2	4.0	4.0
5	q3_3	4.0	4.0
6	q4_1	5.0	5.0
7	q4_2	4.0	4.0
8	q4_3	5.0	5.0
9	q4_4	4.0	4.0
10	q6_1	4.0	4.0
11	q7_1	4.0	4.0
12	q7_2	4.0	4.0
13	q7_3	4.0	4.0
14	q7_4	4.0	4.0
15	q7_5	4.0	4.0
16	q7_6	4.0	4.0
17	q7_7	3.0	3.0
18	q7_8	4.0	4.0
19	q8_1	1.0	1.0

Public Tests

q2_1 results: All test cases passed!

q2_2 results: All test cases passed!

q3_1 results: All test cases passed!

q3_2 results: All test cases passed!

q3_3 results: All test cases passed!

q4_1 results: All test cases passed!

q4_2 results: All test cases passed!

q4_3 results: All test cases passed!

q4_4 results: All test cases passed!

q6_1 results: All test cases passed!

q7_1 results: All test cases passed!

q7_2 results: All test cases passed!

q7_3 results: All test cases passed!

q7_4 results: All test cases passed!

q7_5 results: All test cases passed!

q7_6 results: All test cases passed!

q7_7 results: All test cases passed!

q7_8 results: All test cases passed!

q8_1 results: All test cases passed!

Submitted Files

In [7]:

```
# Initialize Otter
import otter
grader = otter.Notebook("hw01.ipynb")
```

Homework 1: Causality and Expressions

Please complete this notebook by filling in the cells provided. Before you begin, run the previous cell to load the provided tests.

Recommended Readings:

- [What is Data Science?](#)
- [Causality and Experiments](#)
- [Programming in Python](#)

For all problems that you must write explanations and sentences for, you **must** provide your answer in the designated space. Moreover, throughout this homework and all future ones, please be sure to not re-assign variables throughout the notebook! For example, if you use `max_temperature` in your answer to one question, do not reassign it later on. Otherwise, you will fail tests that you thought you were passing previously!

Deadline:

This assignment is due Friday, 6/24 at 11:59pm PT. Turn it in by Thursday, 6/23 at 11:59pm PT for 5 extra credit points. Late work will not be accepted as per the [policies](#) page.

Note: This homework has hidden tests on it. That means even though tests may say 100% passed, it doesn't mean your final grade will be 100%. We will be running more hidden tests for correctness once everyone turns in the homework.

Directly sharing answers is not okay, but discussing problems with the course staff or with other students is encouraged. Refer to the policies page to learn more about how to learn cooperatively.

You should start early so that you have time to get help if you're stuck. Office hours are held on Tuesdays, Thursdays, and Fridays. The schedule appears on <http://data8.org/su22/office-hours.html>.

1. Scary Arithmetic

An ad for ADT Security Systems says,

"When you go on vacation, burglars go to work ... According to FBI statistics, over 25% of home burglaries occur between Memorial Day to Labor Day."

Do the data in the ad support the claim that burglars are more likely to go to work during the time between Memorial Day to Labor Day? Please explain your answer. **(6 Points)**

Note: You can assume that "over 25%" means only slightly over. Had it been much over, say closer to 30%, then the marketers would have said so.

Note: Memorial Day is observed on the last Monday of May and Labor Day is observed on the first Monday of September.

From Memorial Day to Labor Day is indicated between the last Monday of May to first Monday of September. Thinking how much that period could be in a year, it's approximately three months and a week or so. That's almost the quarter of the year, which is also 25 percent. With the numbers given, it might be slightly higher than other periods in the year, but I think it's not concentrated in this period and would be the similar for other periods that are given at three months or so. Therefore, there isn't causation found between these two, I think the data might look convincing but thinking of the real time, it isn't that much of a difference and the data in the ad isn't supporting the claim strongly.

2. Characters in Little Women

In lecture, we counted the number of times that the literary characters were named in each chapter of the classic book, [Little Women](#). In computer science, the word "character" also refers to a letter, digit, space, or punctuation mark; any single element of a text. The following code generates a scatter plot in which each dot corresponds to a chapter of *Little Women*. The horizontal position of a dot measures the number of periods in the chapter. The vertical position measures the total number of characters.

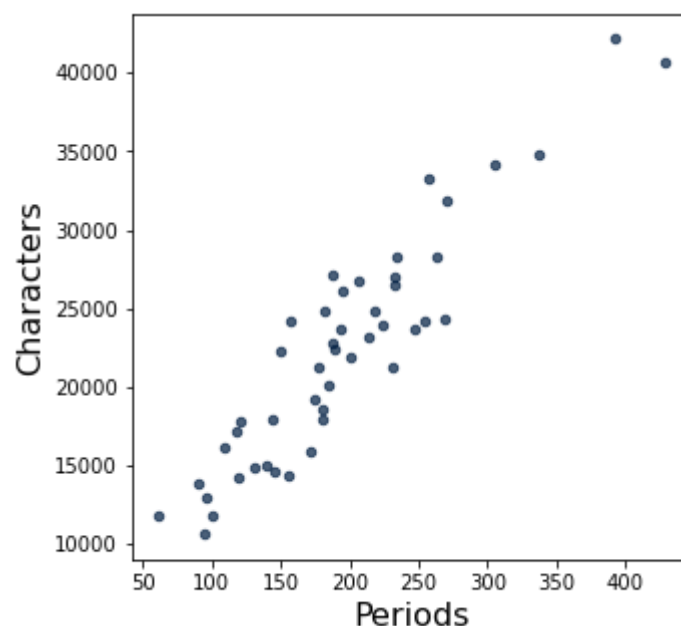
In [8]:

```
# Just run this cell.  
  
# This cell contains code that hasn't yet been covered in the course,
```

but you should be able to interpret the scatter plot it generates.

```
from datascience import *
from urllib.request import urlopen
import numpy as np
import d8error
%matplotlib inline

little_women_url = 'https://www.inferentialthinking.com/data/little_women.txt'
chapters = urlopen(little_women_url).read().decode().split('CHAPTER ')[1:]
text = Table().with_column('Chapters', chapters)
Table().with_columns(
    'Periods', np.char.count(chapters, '.'),
    'Characters', text.apply(len, 0)
).scatter(0)
```



Question 1. Around how many periods are there in the chapter with the most characters? Assign either 1, 2, 3, 4, or 5 to the name `characters_q1` below. (4 Points)

1. 250
2. 390
3. 440
4. 32,000
5. 40,000

Note: If you run into a `NameError: name 'grader' is not defined` error in the autograder cell below (and in any assignment), please re-run the first cell at the very top of this notebook!

In [9]:

```
characters_q1 = 2
```

In [10]: `grader.check("q2_1")`

Out [10]: q2_1 results: All test cases passed!

The test above checks that your answers are in the correct format. **This test does not check that you answered correctly**, only that you assigned a number successfully in each multiple-choice answer cell.

Question 2. Which of the following chapters has the most characters per period? Assign either 1, 2, or 3 to the name `characters_q2` below. **(4 Points)**

1. The chapter with about 60 periods
2. The chapter with about 350 periods
3. The chapter with about 440 periods

In [11]: `characters_q2 = 1`

In [12]: `grader.check("q2_2")`

Out [12]: q2_2 results: All test cases passed!

Again, the test above checks that your answers are in the correct format, but not that you have answered correctly.

To discover more interesting facts from this plot, check out [Section 1.3.2](#) in the textbook.

3. Names and Assignment Statements

Question 1. When you run the following cell, Python produces a cryptic error message.

In [13]: `4 = 2 + 2`

Input In [13]

`4 = 2 + 2`

^

SyntaxError: cannot assign to literal

Choose the best explanation of what's wrong with the code, and then assign 1, 2, 3, or 4 to `names_q1` below to indicate your answer. **(4 Points)**

1. Python is smart and already knows `4 = 2 + 2`.
2. In Python, it's a rule that the `=` sign must have a variable name to its left, and `4` isn't a variable name.
3. It should be `2 + 2 = 4`.
4. I don't get an error message. This is a trick question.

In [14]: `names_q1 = 2`

In [15]: `grader.check("q3_1")`

Out [15]: q3_1 results: All test cases passed!

Question 2. When you run the following cell, Python will produce another cryptic error message.

In [16]: `two = 3
six = two plus two`

Uh-o it seems we have an error!

It seems we have a `SyntaxError`. `SyntaxErrors` are usually because of:

1. Adding extra parenthesis
2. Not following python define or for loop structure, not adding a colon

The Data 8 Reference might be helpful to look over for examples and usage: [Data 8 Reference](#)

If you are having more trouble please feel free to consult a staff member at [Office Hours](#) or see the error message below

Accordion(children=(VBox(children=(VBox(children=(Label(value='Was the message you

Input In [16]

```
six = two plus two
```

^

SyntaxError: invalid syntax

Choose the best explanation of what's wrong with the code and assign 1, 2, 3, or 4 to `names_q2` below to indicate your answer. **(4 Points)**

1. The `plus` operation only applies to numbers, not the word "two".
2. The name "two" cannot be assigned to the number 3.
3. Two plus two is four, not six.
4. The name `plus` isn't a built-in operator; instead, addition uses `+`.

In [17]: `names_q2 = 4`

In [18]: `grader.check("q3_2")`

Out [18]: q3_2 results: All test cases passed!

Question 3. Run the following cell.

In [19]:

```
x = 2
y = 3 * x
x = 4
```

What is `y` after running this cell, and why? Choose the best explanation and assign 1, 2, 3, or 4 to `names_q3` below to indicate your answer. **(4 Points)**

1. `y` is equal to 6, because the second `x = 4` has no effect since `x` was already defined.
2. `y` is equal to 6, because `x` was 2 when `y` was assigned, and $3 * 2$ is 6.
3. `y` is equal to 12, because `x` is 4 and $3 * 4$ is 12.
4. `y` is equal to 12, because assigning `x` to 4 will update `y` to 12 since `y` was defined in terms of `x`.

In [20]: `names_q3 = 2`

In [21]: `grader.check("q3_3")`

Out [21]: q3_3 results: All test cases passed!

4. Differences Between Majors

Berkeley's Office of Planning and Analysis provides data on numerous aspects of the campus. Adapted from the OPA website, the table below displays the number of degree recipients in three majors in the 2008-2009 and 2017-2018 academic years.

Major	2008-2009	2017-2018
Gender and Women's Studies	17	28
Linguistics	49	67
Rhetoric	113	56

Question 1. Suppose you want to find the **biggest** absolute difference between the number of degree recipients in the two years, among the three majors.

In the cell below, compute this value and call it `biggest_change`. Use a single expression (a single line of code) to compute the answer. Let Python perform all the arithmetic (like subtracting 49 from 67) rather than simplifying the expression yourself. The built-in `abs` function takes a numerical input and returns the absolute value. The built-in `max` function can take in 3 arguments and returns the maximum of the three numbers. **(5 Points)**

In [22]: `biggest_change = max(abs(17-28),abs(49-67),abs(113-56))`
`biggest_change`

Out [22]: 57

In [23]: `grader.check("q4_1")`

Out [23]: q4_1 results: All test cases passed!

Question 2. Which of the three majors had the **smallest** absolute difference? Assign `smallest_change_major` to 1, 2, or 3 where each number corresponds to the following major:

1. Gender and Women's Studies
2. Linguistics
3. Rhetoric

Choose the number that corresponds to the major with the smallest absolute difference.

You should be able to answer by rough mental arithmetic, without having to calculate the exact value for each major. **(4 Points)**

```
In [24]: smallest_change_major = 1
        smallest_change_major
```

```
Out [24]: 1
```

```
In [25]: grader.check("q4_2")
```

```
Out [25]: q4_2 results: All test cases passed!
```

Question 3. For each major, define the “relative change” to be the following: $\frac{\text{absolute difference}}{\text{value in 2008-2009}} * 100$

Fill in the code below such that `gws_relative_change`, `linguistics_relative_change` and `rhetoric_relative_change` are assigned to the relative changes for their respective majors. **(5 Points)**

```
In [26]: gws_relative_change = (abs(17-28) / 17) * 100
        linguistics_relative_change = (abs(49-67) / 49) * 100
        rhetoric_relative_change = (abs(113-56) / 113) * 100
        gws_relative_change, linguistics_relative_change, rhetoric_relative_change
```

```
Out [26]: (64.70588235294117, 36.734693877551024, 50.442477876106196)
```

```
In [27]: grader.check("q4_3")
```

```
Out [27]: q4_3 results: All test cases passed!
```

Question 4. Assign `biggest_rel_change_major` to 1, 2, or 3 where each number corresponds to the following:

1. Gender and Women's Studies
2. Linguistics
3. Rhetoric

Choose the number that corresponds to the major with the biggest relative change. **(4 Points)**

In [28]: `biggest_rel_change_major = 1`
`biggest_rel_change_major`

Out [28]: 1

In [29]: `grader.check("q4_4")`

Out [29]: q4_4 results: All test cases passed!

5. Nearsightedness Study

[Myopia](#), or nearsightedness, results from a number of genetic and environmental factors. In 1999, Quinn et al studied the relation between myopia and ambient lighting at night (for example, from nightlights or room lights) during childhood.

Question 1. The data were gathered by the following procedure, reported in the study. "Between January and June 1998, parents of children aged 2-16 years ... that were seen as outpatients in a university pediatric ophthalmology clinic completed a questionnaire on the child's light exposure both at present and before the age of 2 years." Was this study observational, or was it a controlled experiment? Explain. **(5 Points)**

This study is an observational study. Since it wasn't done at a lab or wasn't controlled by the ones who conducted the study, and was just done with observation of the changes that happened throughout with the questionnaire solved by the parents and the children, the study is an observational study.

Question 2. The study found that of the children who slept with a room light on before the age of 2, 55% were myopic. Of the children who slept

with a night light on before the age of 2, 34% were myopic. Of the children who slept in the dark before the age of 2, 10% were myopic. The study concluded the following: "The prevalence of myopia ... during childhood was strongly associated with ambient light exposure during sleep at night in the first two years after birth."

Do the data support this statement? Why or why not? You may interpret "strongly" in any reasonable qualitative way. **(5 Points)**

I believe the data supports this statement, but partially. Without any interfering, they had the three groups with the same age, exposed to different light sets. Just comparing the results, the difference between the percentages is significant and can easily relate the relation between prevalence of myopia and exposure to ambient light. However, there might be a confounding factor that the researchers might not have caught in the research. Other factors might have been included but since it's an observational study, they had only the information that they could observe. Therefore, I do see the relation but believe that the data does not strongly support the statement.

Question 3. On May 13, 1999, CNN reported the results of this study under the headline, "Night light may lead to nearsightedness." Does the conclusion of the study claim that night light causes nearsightedness? **(5 Points)**

Like the headline, I think nightlight "may" lead to nearsightedness. Similar answer to the previous one, the results look relevant, but the result of the study does not strongly supports the conclusion. So no, it doesn't claim that night light causes nearsightedness. There might be other confounding factor that worked in the relation between the results. However, through the results, we could see that it might be one of the factor that cause myopia.

Question 4. The final paragraph of the CNN report said that "several eye specialists" had pointed out that the study should have accounted for heredity.

Myopia is passed down from parents to children. Myopic parents are more likely to have myopic children, and may also be more likely to leave lights on habitually (since the parents have poor vision). In what way does the knowledge of this possible genetic link affect how we interpret the data from the study? Explain. **(5 Points)**

Unlike how the researchers have seen the relevance between myopia and exposure to light, the conclusion may come out completely different. That being said, that the myopic parents may habitually leave the lights on, and myopia has to do with the genetics reasons, myopia and light exposure may have no relevance after all. This certain data, we have to be aware of all the time. Just like this case, confounding variable could be existing in the observational study, and the results of the data might look convincing enough that could easily lead to an error. It would change the whole of how we interpret the data from the study.

6. Studying the Survivors

The Reverend Henry Whitehead was skeptical of John Snow's conclusion about the Broad Street pump. After the Broad Street cholera epidemic ended, Whitehead set about trying to prove Snow wrong. (The history of the event is detailed [here](#).)

He realized that Snow had focused his analysis almost entirely on those who had died. Whitehead, therefore, investigated the drinking habits of people in the Broad Street area who had not died in the outbreak.

What is the main reason it was important to study this group? Assign either 1, 2, or 3 to the name `survivor_answer` below. **(4 Points)**

1. If Whitehead had found that many people had drunk water from the Broad Street pump and not caught cholera, that would have been evidence against Snow's hypothesis.
2. Survivors could provide additional information about what else could have caused the cholera, potentially unearthing another cause.
3. Through considering the survivors, Whitehead could have identified a cure for cholera.

In [30]:

```
survivor_answer = 1
```

In [31]:

```
grader.check("q6_1")
```

Out [31]:

q6_1 results: All test cases passed!

Note: Whitehead ended up finding further proof that the Broad Street pump played a central role in spreading the disease to the people who lived

near it. Eventually, he became one of Snow's greatest defenders.

7. Policies and Administrivia

This section of the homework is to ensure that you have read over the policies and frequently asked questions for the course.

It's important that you read through this section of the homework very carefully. If you can get through all of this section and are sure you have all of the correct resources set up, you will be able to focus on the actual material this semester!

Reading through the [policies](#) and the [FAQ](#) will help you get through this section very easily. It is recommended you do this before proceeding.

Question 1. You have a question regarding the grading of your assignments that has not been previously answered on Ed or the FAQ. Who do you contact? Assign to the number corresponding to the best choice below. **(4 Points)**

1. The Instructors
2. Post on Ed
3. Contact your Lab TA

In [32]:

```
contact = 3
```

In [33]:

```
grader.check("q7_1")
```

Out [33]:

```
q7_1 results: All test cases passed!
```

Question 2. Why are there typically 2 items listed on Gradescope for each homework assignment? Assign to the number corresponding to the best choice below. **(4 Points)**

1. There was a mistake in the grading. I should contact someone about this.
2. One assignment is for coding questions (which I will submit to), and the other is automatically submitted for me and contains my written work.
3. Trick question.

In [34]:

```
grades = 2
```

In [35]: `grader.check("q7_2")`

Out [35]: q7_2 results: All test cases passed!

Question 3. Regrade deadline dates will always be posted on the same Ed post that releases the assignment grades, common mistakes, and solutions. Can you ask for parts of your assignment reggraded after the regrade request window has passed? Assign `regrade` to the number corresponding to the best choice below. **(4 Points)**

1. Yes
2. No

In [36]: `regrade = 2`

In [37]: `grader.check("q7_3")`

Out [37]: q7_3 results: All test cases passed!

Question 4. Do you have an Gradescope account? Head to gradescope.com and check if you see Data 8. If you do not, please send your Lab TA an email with your email and student ID number.

Once you have been enrolled, go to the Data 8 Gradescope course website. At the end of the url link, you should see a six-digit number. Assign `gradescope` to that number. **(4 Points)**

In [38]: `gradescope = 397747`

In [39]: `grader.check("q7_4")`

Out [39]: q7_4 results: All test cases passed!

Question 5. Given the following scenarios, assign `acceptable` to the scenario that is permissible given the guidelines on the [policies](#) page. **(4 Points)**

1. Yuqi gets stuck on a homework assignment, so she googles a fix. She stumbles across a pdf of the solutions for the homework assignment

from a previous semester's offering of Data 8. After inspecting the solution, Yuqi writes her own solution and submits the assignment.

2. After getting confused by a project, Varun asks his friend for help. His friend Angela helps by walking Varun through her own logic, without showing her code, pointing out areas that are important given the context of the question. Upon hearing his friend's logic, Varun writes his own code and completes the project.
3. Ciara has an extremely busy schedule, so she really wants to leave lab early by finishing it and getting checked off. Her neighbor, Prasann, simply turns his computer so Ciara can see how he completed some questions. After looking at his code, Ciara finishes the lab and gets checked off.

In [40]: `acceptable = 2`

In [41]: `grader.check("q7_5")`

Out [41]: q7_5 results: All test cases passed!

Question 6. To make sure you have read through the [policies](#) and the [FAQ](#), carefully, how many HW and lab drops are there? Assign `drops` to the number corresponding to the best choice below. **(4 Points)**

1. Three homework drops and three lab drops
2. One homework drop and one lab drop
3. Only two homework drops
4. One homework drop and two lab drops

In [42]: `drops = 2`

In [43]: `grader.check("q7_6")`

Out [43]: q7_6 results: All test cases passed!

Question 7. Does Data 8 offer alternate final exam to those with class conflicts? Assign `exams` to the number corresponding to the best choice below. **(3 Points)**

1. Yes
2. No

In [44]: exams = 2

In [45]: grader.check("q7_7")

Out [45]: q7_7 results: All test cases passed!

Question 8: Are you actually checking Ed? Go to this semester's [Data 8 Ed](#) and find an instructor posted thread with a certain secret phrase. Assign `secret` to this secret phrase in quotes (i.e. as a string). **(4 Points)**

In [46]: secret = 'squid'

In [47]: grader.check("q7_8")

Out [47]: q7_8 results: All test cases passed!

8. Welcome Survey

Question 1. Please complete the welcome survey below in order to receive credit for homework 1. **(1 Point)**

- Summer 2022 Welcome Survey: <https://forms.gle/stjdzC6PmkSrNHBW9>

Assign `survey` to the secret phrase given at the end of the welcome survey. Make sure the phrase is in quotes (i.e. is a string)!

In [48]: survey = 'data8 is data gr8'

In [49]: grader.check("q8_1")

Out [49]: q8_1 results: All test cases passed!

You're done with Homework 1!

Important submission information: Be sure to run the tests and verify that they all pass, then choose **Save Notebook** from the **File** menu, then **run the final cell** and click the link to download the zip file. Then, go to

[Gradescope](#) and submit the zip file to the corresponding assignment. The name of this assignment is "HW 01 Autograder". **It is your responsibility to make sure your work is saved before running the last cell.**

Once you have submitted, your Gradescope assignment should look something like the following image if you have passed all tests.

Note: *This is a photo of a generic Gradescope submission result, and it does not included the same test numbers as this assignment. Please check that all test cases have passed for each question.*

In []:

To double-check your work, the cell below will rerun all of the autograder tests.

In [50]:

```
grader.check_all()
```

Out [50]:

q2_1 results: All test cases passed!

q2_2 results: All test cases passed!

q3_1 results: All test cases passed!

q3_2 results: All test cases passed!

q3_3 results: All test cases passed!

q4_1 results: All test cases passed!

q4_2 results: All test cases passed!

q4_3 results: All test cases passed!

q4_4 results: All test cases passed!

q6_1 results: All test cases passed!

q7_1 results: All test cases passed!

q7_2 results: All test cases passed!

q7_3 results: All test cases passed!

q7_4 results: All test cases passed!

q7_5 results: All test cases passed!

q7_6 results: All test cases passed!

q7_7 results: All test cases passed!

q7_8 results: All test cases passed!

q8_1 results: All test cases passed!

Submission

Make sure you have run all cells in your notebook in order before running the cell below, so that all images/graphs appear in the output. The cell below will generate a zip file for you to submit. **Please save before exporting!**

In [51]:

```
# Save your notebook first, then run this cell to export your submission.  
grader.export(pdf=False)
```

<IPython.core.display.HTML object>

▼ .OTTER_LOG

Download

1 Binary file hidden. You can download it using the button above.

▼ __zip_filename__

Download

1 hw01_2022_06_23T23_19_08_020241.zip