

Homework 1: Causality and Expressions

Please complete this notebook by filling in the cells provided. When you're done:

1. Select `Run All` from the `Cell` menu to ensure that you have executed all cells.
2. Select `Download as PDF via LaTeX (.pdf)` from the `File` menu
3. Read that file! If any of your lines are too long and get cut off, we won't be able to see them, so break them up into multiple lines and download again.
4. Submit that downloaded file to Gradescope.

If you cannot submit online, come to office hours for assistance. The office hours schedule appears on data8.org/fa16/weekly.html (<http://data8.org/fa16/weekly.html>). (It might not be posted by the time this assignment comes out.)

This assignment is due Thursday, September 1 at 5PM. You will receive an early submission bonus point if you turn it in by Wednesday, August 31 at 5PM. Directly sharing answers is not okay, but discussing problems with course staff or with other students is encouraged.

Reading:

- Textbook chapters [1](http://www.inferentialthinking.com/chapters/01/what-is-data-science.html) (<http://www.inferentialthinking.com/chapters/01/what-is-data-science.html>), [2](http://www.inferentialthinking.com/chapters/02/causality-and-experiments.html) (<http://www.inferentialthinking.com/chapters/02/causality-and-experiments.html>), and [3](http://www.inferentialthinking.com/chapters/03/programming-in-python.html) (<http://www.inferentialthinking.com/chapters/03/programming-in-python.html>)

Tests

Automated tests are provided for many questions. If a question has automated tests, there will be a runnable code cell after the question's answer cell. Run your answer cell, then run the test cell to check your answer.

Passing the automatic tests does not guarantee full credit on any question. The tests are provided to help catch some common errors, but it is *your* responsibility to answer the questions correctly.

Run the cell below to prepare the notebook and automated tests.

```
In [ ]: # Run this cell to set up the notebook, but please don't change it.
        from client.api.assignment import load_assignment
        tests = load_assignment('hw01.ok')
```

1. Arithmetic Didn't Take

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 and Labor Day."

Do the data in the ad support the claim that burglars are more likely to go to work during the summer vacation period than at other times?

Write your answer here, replacing this text.

2. Breaking Down Expressions

The most important idea in Computer Science is that complicated, useful things can be built by putting together simple parts according to simple rules. Python code is an important example of this principle. Once you understand the basic rules, you can code with confidence. These exercises are designed to give you some practice with those rules.

First, a brief review of subexpressions.

You can take any Python expression that has a value and combine it with other expressions. For example, you can combine two number-valued expressions by putting a `+` between them to add their values together. This forms a new, larger expression called a *compound expression*. The expressions that were combined together are called *subexpressions*.

You can tell if something is a subexpression by checking whether it would make sense to write it in a line by itself. For example, in the expression `2 * 3`, `2` is a subexpression, but `2 *` isn't, because `2 *` isn't a valid expression. (Try executing it!)

Question 1. List all the subexpressions of the following expression:

`2 + 3`

Put each subexpression on its own line in the next cell.

Hint: There are two of them.

In []: ...
...

Question 2. Consider the following expression:

$(1 + 2) * ((3 / 4) ** 5)$

Here is a list of almost all the subexpressions of that expression. One is missing.

1. 1
2. 2
3. 3
4. 4
5. 5
6. $(1 + 2)$
7. $((3 / 4) ** 5)$

In the next cell, write the missing expression.

In [12]: ...

Question 3. List all the subexpressions of the following expression:

$((2**3) / 4) / 5 - 6$

Put each subexpression on its own line in the next cell.

In []: ...
...
...
...
...
...
...
...

3. Errors in Naming

Question 1. When you run the following cell, Python will produce an slightly-cryptic error message. Explain in the text cell below, in your own words, what's wrong with the code. (Remember, double-click the cell to edit it, and then click the Run button when you're done.)

In []: 4 = 2 + 2

Write your answer here, replacing this text.

Question 2. When you run the following cell, Python will produce an slightly-cryptic error message. **Fix the error**, and then **explain below** in your own words what was wrong with the code.

```
In [2]: two = 2  
       four = two plus two
```

```
In [3]: _ = tests.grade('q3_2')
```

Write your answer here, replacing this text.

4. Job Opportunity Education in Rural India

A [study \(http://www.nber.org/papers/w16021.pdf\)](http://www.nber.org/papers/w16021.pdf) at UCLA investigated factors that might result in greater attention to the health and education of girls in rural India. One such factor is information about job opportunities for women. The idea is that if people know that educated women can get good jobs, they might take more care of the health and education of girls in their families, as an investment in the girls' future potential as earners.

The study focused on 160 villages outside the capital of India, all with little access to information about call centers and similar organizations that offer job opportunities to women. In 80 of the villages chosen at random, recruiters visited the village, described the opportunities, recruited women who had some English language proficiency and experience with computers, and provided ongoing support free of charge for three years. In the other 80 villages, no recruiters visited and no other intervention was made.

At the end of the study period, the researchers recorded data about the school attendance and health of the children in the villages.

Question 1. Did this analysis have a treatment group and a control group? If so, describe the two groups.

Write your answer here, replacing this text.

Question 2. Was this an observational study or a randomized controlled experiment?

Write your answer here, replacing this text.

Question 3. The study reported, "Girls aged 5-15 in villages that received the recruiting services were 3 to 5 percentage points more likely to be in school and experienced an increase in Body Mass Index, reflecting greater nutrition and/or medical care. However, there was no net gain in height. For boys, there was no change in any of these measures." Why do you think the author points out the lack of change in the boys?

Write your answer here, replacing this text.

5. Differences between Universities

Question 1. Suppose you're choosing a university to attend, and you'd like to *quantify* how *dissimilar* any two universities are. You rate each university you're considering on 3 traits, using a 0 to 10 scale for each trait:

1. Cost to attend (0 for the cheapest)
2. Graduation rate
3. How cool its mascot is

You decide that the dissimilarity between two universities is:

- the maximum of
- the absolute values of
- the 3 differences in their trait values.

Using this method, compute the dissimilarity between Stanford (whose traits are 8, 9, and 0, respectively) and Berkeley (whose traits are 7, 9, and 10, respectively). Call your answer `dissimilarity`. Use a single line of code to compute the answer. Use Python to do all the steps, including arithmetic (like subtracting 8 from 7).

```
In [4]: dissimilarity = ...
         dissimilarity
```

```
In [5]: _ = tests.grade('q5_1')
```

Question 2. Identify all the subexpressions in your answer to the previous question. Write each on its own line.

Hint: If your answer to the previous question was as straightforward as possible, there should be 12 subexpressions, not including the whole expression itself.

```
In [ ]: ...  
...  
...  
...  
...  
...  
...  
...  
...  
...  
...
```

6. More Children Living at Home?

A USA Today [article \(http://usatoday30.usatoday.com/news/nation/2006-03-16-failure_x.htm\)](http://usatoday30.usatoday.com/news/nation/2006-03-16-failure_x.htm) from 2006 includes this sentence: “Since 1970, the percentage of people ages 18 to 34 [in the United States] who live at home with their family increased 48%, from 12.5 million to 18.6 million, the Census Bureau says.”

Question 1. The word “percentage” isn’t used correctly in the context of the rest of the sentence. What word should replace it?

Write your answer here, replacing this text.

Question 2. Can you give a simple explanation for these data? Feel free to include other sources of data to support your explanation, but please keep your answer to 1-3 sentences.

Write your answer here, replacing this text.