Lab 02: Introduction to Tables

Data 8 Discussion Worksheet

Tables are a fundamental way of representing data sets. A table can be viewed in two ways:

- A sequence of named columns that each describe a single attribute of all entries in a data set, or
- A sequence of rows where each row contains all the attribute information about that entry in the data set

1. Ready, Willing and Table

Let's look at an example table called `staff`

Name	Year	Semesters on Staff
Devarsh	4	3
Raymond	4	5
Miranda	2	3
Carlos	4	6
Peter	3	2
Olivia	4	4
James	3	4
Oswaldo	3	3
Meghan	4	8
Diana	2	3

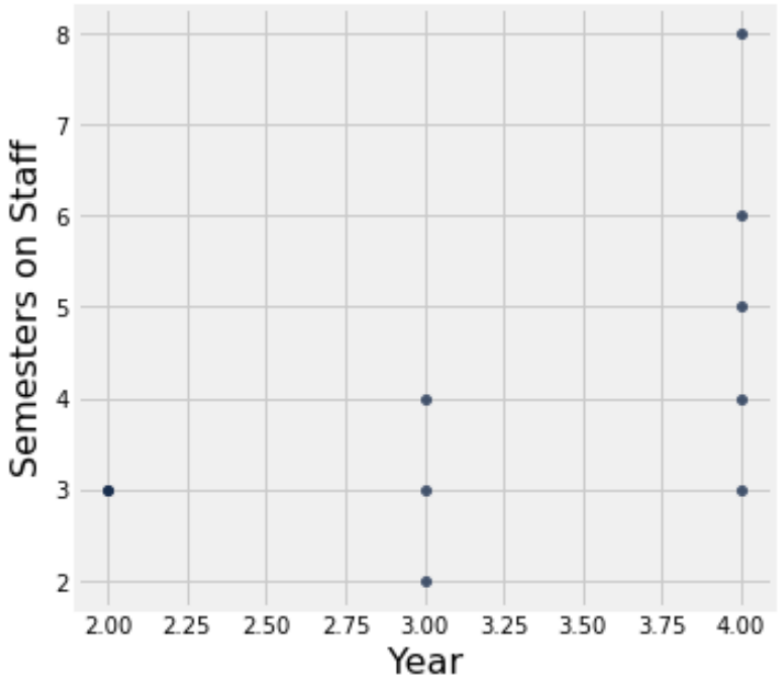
The table has 10 rows, each corresponding to one member of Data 8 Staff. Each row has three attributes, the staff member's name, year, and how many semesters they have been on staff. Using just the information from the `staff` table, do we have enough information to generate the following by hand? If not, what additional information do you need? *(You don't need to worry about how you'd do it in Python.)* You can assume all calculations are correct.

Year	Semesters on Staff average
2	3
3	3
4	5.2

A. True / False We can look at all the staff members of each unique year and average their semesters on staff.

Name	Year
James	Junior
Oswaldo	Junior
Peter	Junior

B. True / False We don't have the data in this table that tells us that a third year is called a Junior! However, we could filter the table down to only have third years, and remove the column with semesters on staff.



C. True / False Each point on this plot represents one staff member, the x value is their year and the y value is the number of semesters they have been on staff.

Semesters on Staff	2	3	4
2	0	1	0
3	2	1	1
4	0	1	1
5	0	0	1
6	0	0	1
8	0	0	1

D. True / False This is a really tricky example because the column labels are not super intuitive! The first column is semesters on staff, and the subsequent columns refer to the years. The value in the "Semesters on Staff" column combines with the column (either 2, 3, or 4) to tell us how many rows (staff members) from the original table had that many semesters on staff and are that year. That's a little bit hard to understand, so as an example look at the cell in the 2 column and in the row of 3 semesters on staff. That tells us that in the original table there were 2 (the value in that cell) members of staff that have 3 semesters on staff and are second years. As another example, we can see that there is 1 member of staff with 8 semesters of experience who is a fourth year.

2. Causality, Coworkers and Coffee

Ciara collected the following information about her coworkers' methods of getting to work and their coffee consumption.

Method	Number of Coworkers	Average Cups of Coffee per Day
Take the Bus to Work	12	1.1
Drive to Work	15	1.9

a. Ciara is trying to compute the absolute value of the difference between the total number of cups drunk by driving coworkers per year vs the total number of cups drunk by bussing co-workers per year. She will do all of this in a single cell. Identify the errors in the following cell and correct them. *Make sure that the code cell outputs a single positive number.*

```
number_cups_bus = 12(1.1)
number_cups_drive = 15(1.9)
number_cups_day_difference = ((number_cups_bus -
number_cups_drive)
```

```
   number_cups_week_difference = number_cups_difference * 7
   yearly cups = number_cups_week_difference * 52
```

```
number_cups_bus = 12(1.1)
number_cups_drive = 15(1.9)
```
#1 Error - Explanation: can't use () for multiplication

```
number_cups_day_difference = number_cups_bus - number_cups_drive
```
#2 Error - Parentheses were wrong! In jupyter you can use put the cursor on parentheses to see if they're matched! Syntax errors are often reported on the wrong line.

```
number_cups_week_difference = number_cups_difference * 7
```
#3 Error - Explanation: the variable name was wrong, you can use tab to autocomplete!

#4 Error - Also, we want to use absolute value (by calling abs) at some point (can do it in any of the last three lines) because the question asked us to! We could also use a different subtraction order to make sure the answer is positive.

```
yearly cups = number_cups_week_difference * 52
```
#5 Error - Explanation: variable names cannot have spaces. It's always good to have descriptive variable names, including ones that are multiple words, but we need to use underscores to separate them instead of spaces.

```
yearly_cups
```
#6 Error - Explanation: a cell will not output anything unless a variable name is the last line or a print statement was executed.

b. Is there a relationship between transportation method and coffee consumption—an association, a causal relationship or something else? Why?

There is an association between transportation method and coffee consumption: those who drive to work consume a higher amount of coffee than those that take the bus to work. We cannot say there is a causal link because this was an observational study, the groups were not randomly assigned a treatment so there can be confounding factors. Examples of confounding factors could be age.