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.

a. True / False

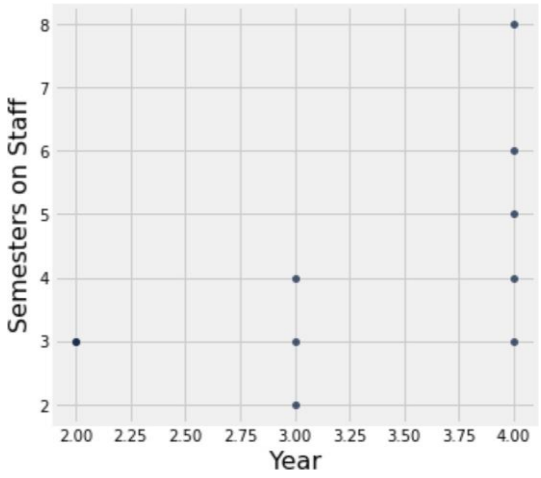
Year	Semesters on Staff average
2	3
3	3
4	5.2

b. True / False

technically would need year names instead of number for this table. Although one could assume that a third year is called a junior, this could technically be fabricating unsent data.

Name	Year
James	Junior
Oswaldo	Junior
Peter	Junior

c. True / False



d. True / False

year

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

given all necessary data for the "true" tables

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_bus = 12*1.1
number_cups_drive = 15*1.9
number_cups_day_difference = abs(number_cups_bus - number_cups_drive)
number_cups_week_difference = number_cups_day_difference*7
yearly_cups = number_cups_week_difference*52
yearly_cups
```

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 higher coffee consumption and driving oneself to work. It is impossible to say if there is a causal relationship because this seems to be an epidemiological study--wherein Ciara did not force people to have changes in their lifestyles, but simply studied people in their typical habits. In order to establish causation, Ciara might want to consider blind, randomized control trials. Some possible explanations for this data could be that driving oneself to work is more tiring than having the bus driver drive themselves, and the exercise it takes to get to the bus station could help wake people up. Conversely, one might argue unfoundedly that a bus taker might need to wake up earlier to get to the bus. But again, it is impossible to say if this is a causal relationship; therefore, it is merely an associative relationship.