**Data 8 Spring 2020**

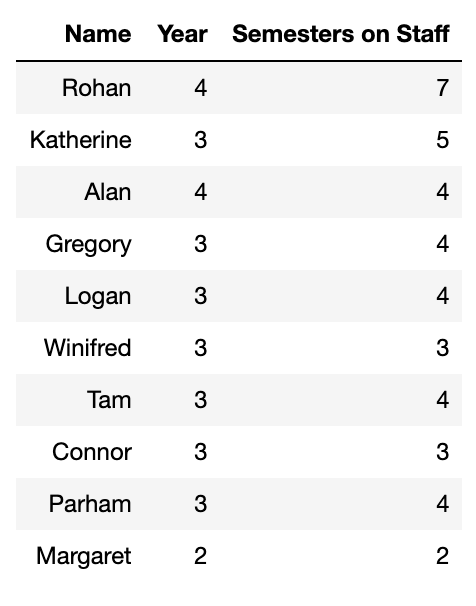
**Discussion: Introduction to Tables**

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

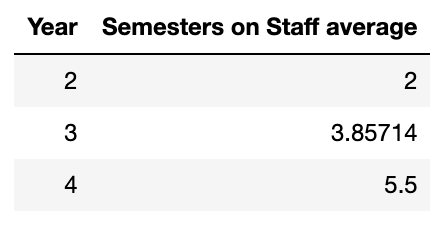
* a sequence of named columns that each describe a single attribute of all entries in a data set, or
* a sequence of rows that each contain all information about a single individual in a data set.

**1. Ready, Willing and Table**

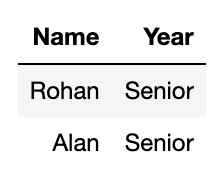
Let’s look at an example table called staff



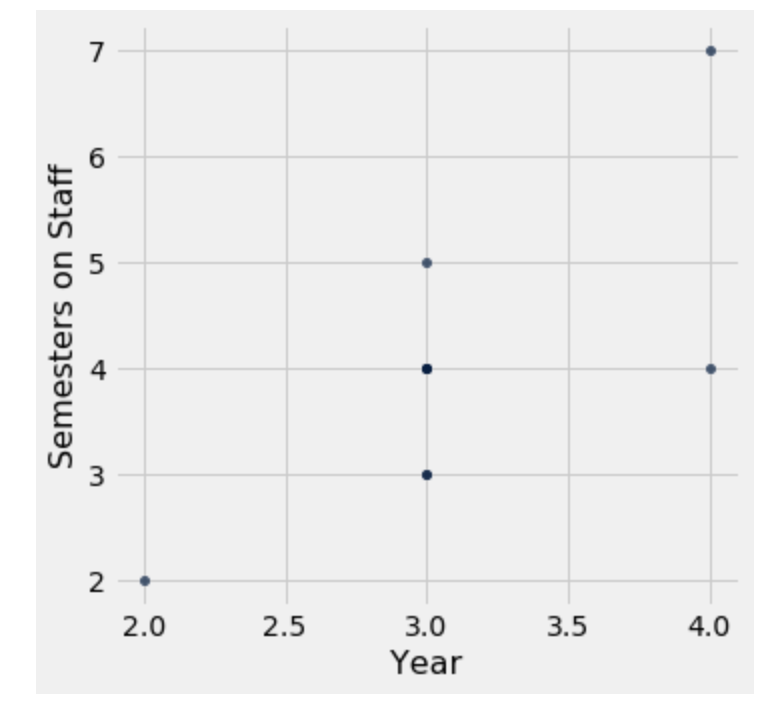
The table has 10 rows, each corresponding to one member of Data 8 Staff. Each row has three attributes, the staff member’s name, year and how many semesters they have been on staff.Using just the information from the staff table can we can generate the following?



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



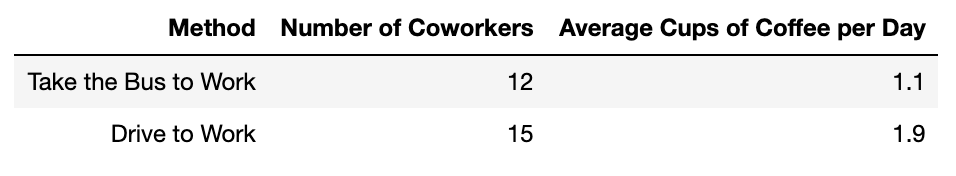
**B.** True / False We don’t have the data in this table that tells us that a fourth year is called a senior! However, we could filter the table down to only have fourth years, and remove the column with semesters on staff.



**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.

**2. Casuality, Coworkers and Coffee**

Divyesh collected the following information about his coworkers’ methods of getting to work and their coffee consumption.



**A.** Divyesh is trying to compute the total yearly difference between the cups of coffee that his driving coworkers drink and the cups of coffee his coworkers who bus drink. He 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\_day\_difference = (12(1.1 - 15(1.9)))

number\_cups\_week\_difference = number\_cups\_difference \* 7

yearly cups = number\_cups\_week\_difference \* 52

number\_cups\_day\_difference = 15\*1.9 - 12\*1.1explanation: can’t use () for multiplication, the nesting is incorrect.

number\_cups\_week\_difference = number\_cups\_day\_difference \* 7

explanation: the variable name was wrong, you can use tab to autocomplete!

yearly\_cups = number\_cups\_week\_difference \* 52

explanation: variable names cannot have spaces

yearly\_cups   
explanation: a cell will not output anything unless a variable names 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.