

CS 106 Winter 2019

Lab 10: Tables

Due: Wednesday, March 27th, 11:59pm

SAVE your sketches as "L10Q01" and "L10Q02".

Question 1 Marvel Comic Characters

The statistical analysis website FiveThirtyEight has a great habit of publishing the datasets they build in the process of their reporting, resulting in a veritable treasure trove of fun and unusual CSV files. We'll use one of those datasets to get you comfortable working with tables in Processing. Your sketch won't draw anything, it'll just crunch some data and report summaries to the console.

1. In 2014, FiveThirtyEight ran a story titled [Comic Books Are Still Made By Men, For Men And About Men](#). To create that story, they assembled a large dataset of Marvel and DC comic book characters. A copy of that dataset in the file `marvel.csv` is provided to you.
2. Study the dataset to get a general sense of what data is in it and how it's laid out. You can load the CSV file into Excel to see the data.
3. In the sketch, create a `setup()` function. In `setup()`, load the CSV file into the sketch and store it in a local variable of type `Table`. Note that this table has a header row. Run the sketch to verify that the table was initialized without any errors.
4. Add code to `setup()` to count the number of bisexual characters in Marvel comics. Gather this information by walking over the table and checking whether each character is bisexual, adding one to a local integer variable each time they are. At the end of `setup()`, use `println()` to print a line of text to the console of the form

```
Bisexual characters: 123
```

where 123 is replaced by the number of bisexual characters you found in the table (123 isn't the answer). Don't draw this or any other text in the sketch window; print it to the console.

5. Add code to `setup()` to measure the average number of appearances by all characters. The top characters have thousands of appearances, but there's a "long tail" of bit players with one or two appearances each. Where does the average fall? Print a line of the form

```
Average number of appearances: 12.3456
```

where `12.3456` is replaced by the average you calculate.

6. What percentages of characters are male or female? Presumably there must also be characters who are neither. Gather information about the proportions of characters who are listed as male, female, or anything else, and print a line of the form

```
Sex: 40.1% male, 39.9% female, 20.0% other
```

Those are *definitely* not the right answers! Note that to calculate a percentage, use an expression like `float(count)/float(total)*100.0`, where `count` is the number of characters you found with the trait and `total` is the total number of characters in the dataset.

When you're done, the sketch's console window should look like this:

```
Bisexual characters: 123
Average number of appearances: 12.3456
Sex: 40.1% male, 39.9% female, 20.0% other
```

Nothing should be drawn in the sketch window. You don't need any code in your sketch apart from a `setup()` function. And remember: it's incorrect to figure out the answers to these questions in, say, Excel, and then to write a sketch that just prints out the answers. A "solution" of that form will receive no marks.

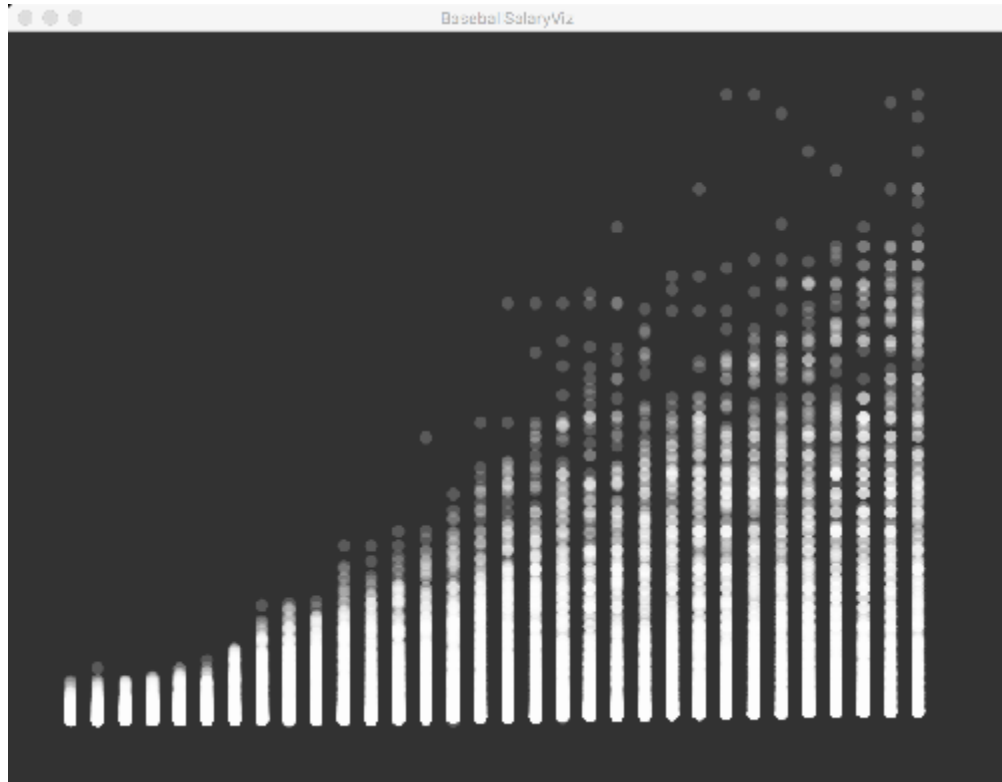
Question 2 Baseball Salaries

The highest single-year salary ever paid to a Major League Baseball player was \$33,000,000 (thirty-three million dollars), which Alex Rodriguez received in 2009 and 2010, and Clayton Kershaw received in 2016. How have player salaries been evolving over time? Is everybody getting paid more, or only top players? You will write a sketch to visualize player salaries from 1985 to the present day.

1. From Sean Lahman's Baseball Archive we have downloaded `Salaries.csv` for you.
2. Write a `setup()` function that sets the window size to at least 500x500. Add the file `Salaries.csv` to the sketch. For this sketch, that's the only table you'll need.
3. Add code to `setup()` to load the table from `Salaries.csv`. Note that this table includes a header row.

In this exercise, we care about two fields in the table: the first field, named `yearID`, and the last field, named `salary`.

4. Loop over the rows of the table. For every row, draw a small semi-transparent circle whose `x` position is determined from the year and whose `y` position is determined from the given player's salary in that year. Use `map()` (or something like it) to ensure that the range of years 1985–2016 fills the width of the sketch (minus a margin on the left and right) and the range of salaries 0–33000000 fills the height of the sketch (minus a margin on the top and bottom). You'll end up with an image like this:



For this question, you're allowed to "know" going in that the highest salary in the table is 33000000. That is, you can put that into your sketch as a constant, you don't need to walk through the table to find the maximum salary (this will save a bit of coding).

You can complete this sketch in under 20 lines of code, not counting comments.

Submission

Submit all sketch directories from this lab as one ZIP file called L10.zip to the lab dropbox on Learn.

It is your responsibility to submit to the correct dropbox with the correct files before the deadline. Otherwise you will receive a mark of 0.