



MCA
CHIC
AGO

Museum of Contemporary Art Chicago

DS3500: Advanced Programming with Data

Prof. Rachlin

Introduction

In this assignment we survey the artists represented at the Museum of Contemporary Art (MCA) Chicago. The attached dataset is a JSON file containing over 10,000 artists whose works are displayed at the museum. Your task is to generate Sankey visualizations using the `plotly python` library that show connections between artist nationality, the decade when they were born, and their gender. Which countries, decades and genders are most represented? You may reuse any code presented in class for this assignment.

Specific Tasks

1. Convert the data into a Panda's dataframe containing three columns: nationality, gender, and the *decade* the artist was born. For example, if they were born between 1940 and 1949, you need only store the decade as 1940.
2. Aggregate the data, counting the number of artists grouped by both nationality and decade.
3. Filter any rows where the decade is 0 (presumably unknown).
4. Filter out rows whose artist count is below some threshold. You'll want to experiment with this value to produce a visually appealing visualization. I suggest trying a starting threshold around 20.
5. Generate a Sankey diagram with nationality on the left (sources) and decade of birth on the right (targets)
6. Repeat steps 2-5, but this time count the number of artists grouped by nationality and gender. Your Sankey diagram will show nationalities on the left and gender on the right.

7. Repeat steps 2-5, grouping by gender and decade. Your Sankey diagram will display gender on the left and decade of birth on the right.
8. Write a ½ to 1-page interpretation of your results. What insights do you glean from your visualizations? What does this data science exercise possibly tell us about diversity, inclusion, and bias in the art world?

Considerations

- A. When implementing steps 5-7, consider ways to make your code more reusable so that you aren't scripting the same tasks repeatedly.
- B. In general, as we transition from intermediate to advanced-level programming, we want to be thinking not only about functions and modularity, but the construction of packages and libraries. You should place code (functions, classes, etc.) that has the potential for future reuse in separate python files and then *import* these libraries into your main assignment code.
- C. Your code will be graded on correctness, efficiency, readability, documentation, and modularity, reusability, and your analysis.
- D. Extra credit (10 points): Extend the functionality of your code (or the code I developed in class) to generate multi-layered Sankey diagrams. Demonstrate your code by including a Sankey diagram with all three layers: nationality, gender, and decade of birth (in any order).

What to submit

Submit your python code as .py files (not Jupyter Notebooks!), your three Sankey diagrams, your optional extra-credit multi-layered Sankey diagrams, and your written analysis.