Assignment 4 Report

1951567 梁潇怡

1. Introduction

This project is a project to visualize the relationship between salary data, regions and university types of graduates at different stages of their careers, further clarify their internal relationships in the form of graphs, and clearly show data relationships to users.

2. Dataset

2.1 dataset introduction

The data set selected for the project includes three files, which mainly describe the relationship between school type and salary, the relationship between region and salary, and the relationship between major and salary.

In each file, the salary includes data for multiple stages, including Starting Median Salary, Mid-Career Median Salary, Mid-Career 10th Percentile Salary, Mid-Career 25th Percentile Salary, Mid-Career 75th Percentile Salary, Mid-Career 90th Percentile Salary data.

2.2 characteristics of dataset

By observing the data set, we can find that the data set has clear correspondence characteristics, and in each file, the data is classified according to the requirements in the file name. This means that we don't need to spend too much time processing the dataset.

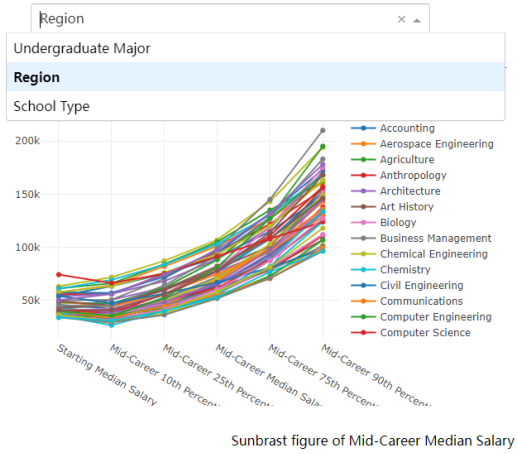
2.3 tasks

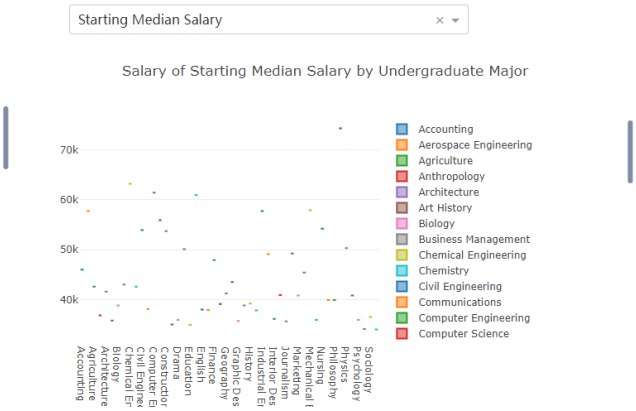
According to the characteristics of the dataset, we specify the following three task scenarios, aiming to provide users with a clearer description of the data by utilizing the visual charts.

1. By choosing majors, regions, and school types, get the corresponding average salary for different career stages.
2. The user selects different stages, and the system displays the salary of the specified stage under different aggregation methods according to the selected major, region, and school type.
3. The system displays Mid-Career Median Salary according to the selected major, region and school type

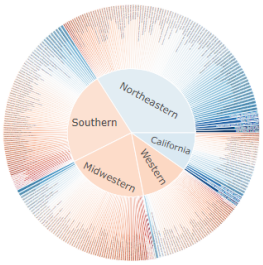
3. Display

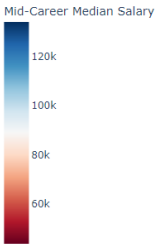
The project is based on dash, flask and plotly. Flask provides the server interface, plotly provides images, and dash is responsible for carrying it all.



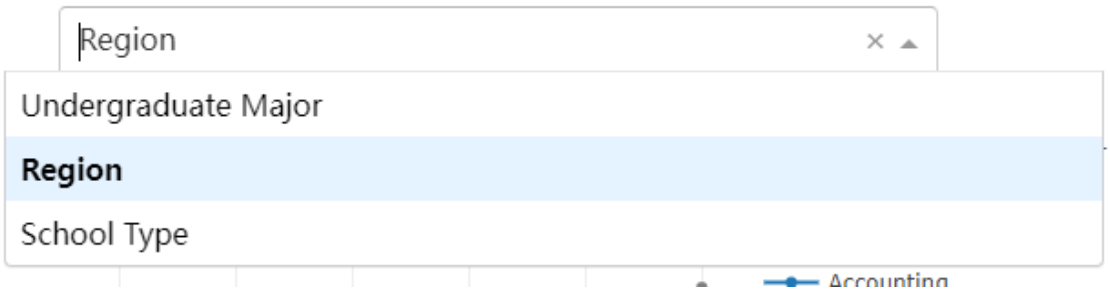


Sunbrast figure of Mid-Career Median Salary

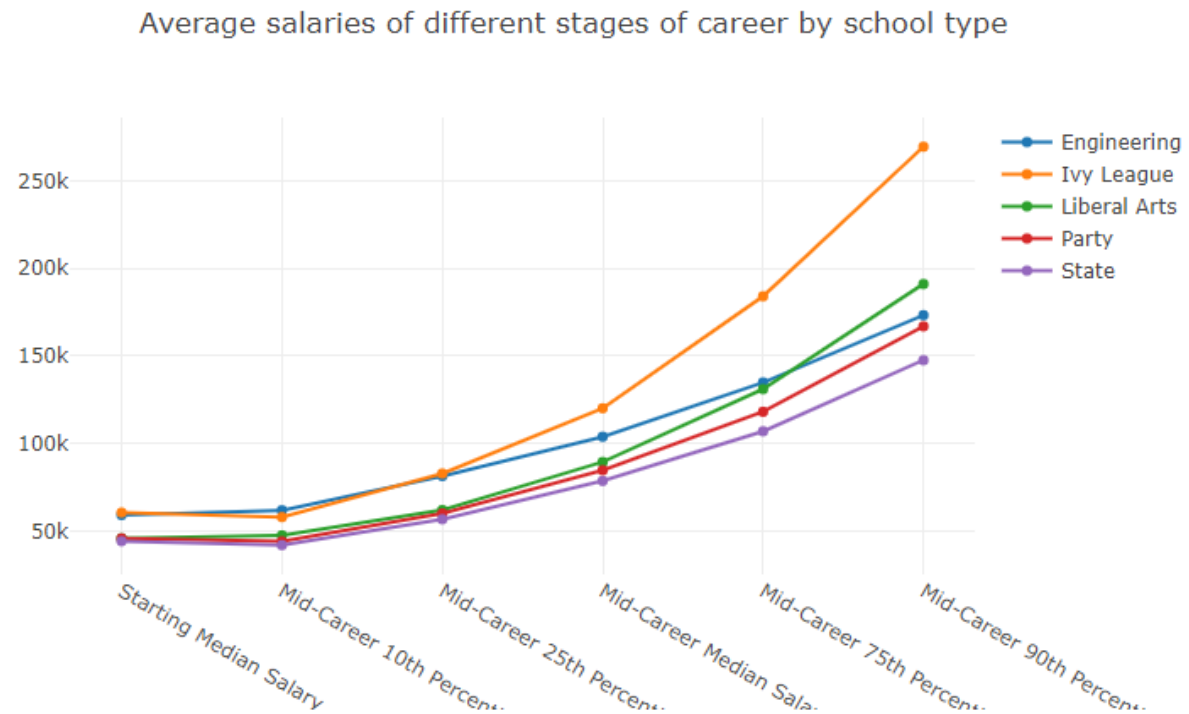




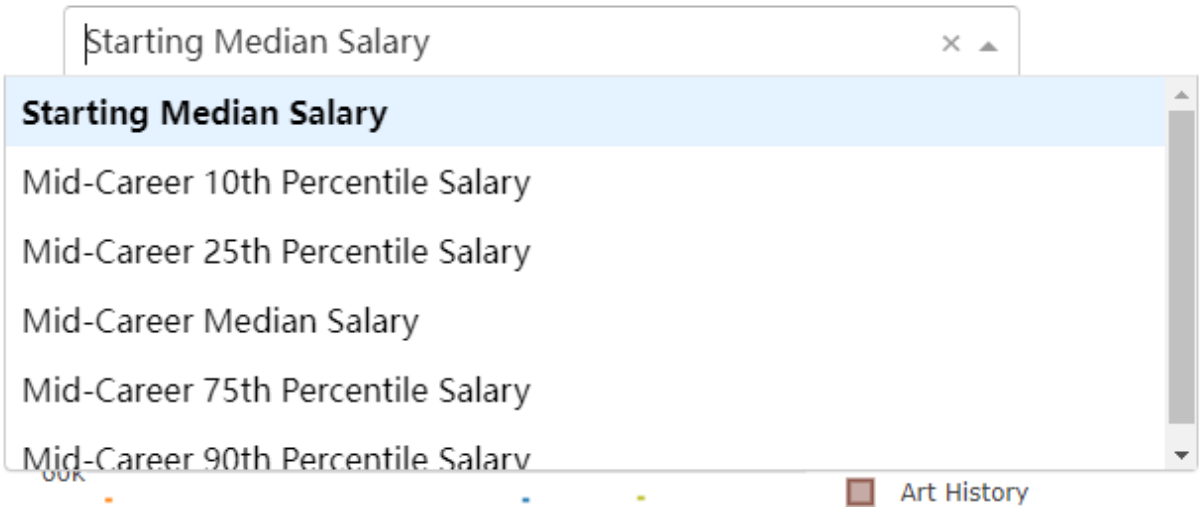
The project layout is shown below. The first selector is a global selector that controls the next two graphs. To show the trend clearly, I used a line chart.



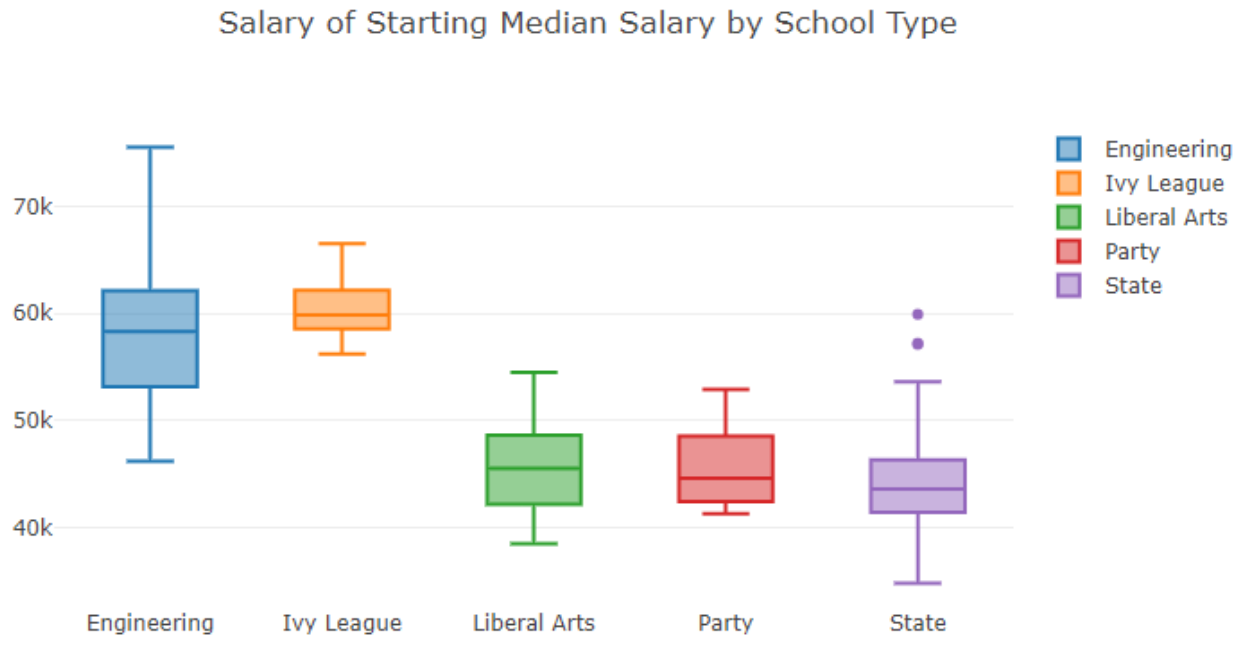
The first graph shows salary growth for different filters.



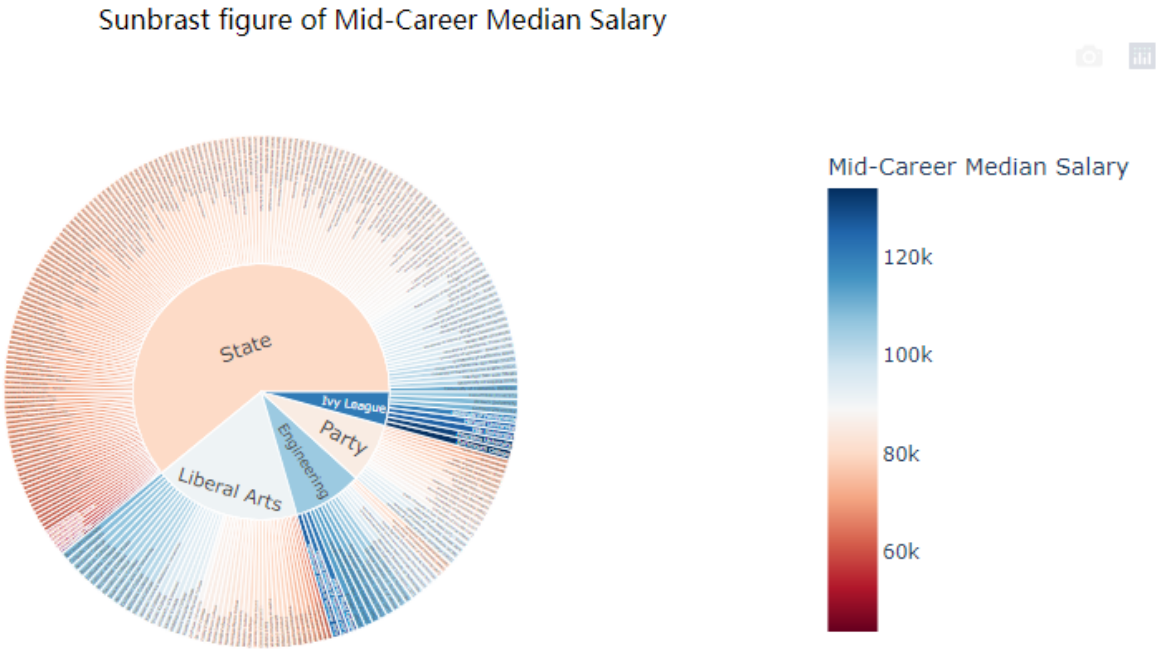
The second selector is used to select the stage of the occupation, thus controlling the display of the second graph.



The second picture shows the distribution of salary in the specified stage under the same aggregation method. In order to display the highest and lowest data, I use a box plot to display it.



The third picture shows the corresponding Mid-Career Median Salary. Sunbrast chart is an upgraded version of pie chart, which can clearly show proportion and distribution.



4. How to Run Code

Install the requirements.

```
pip install -r requirements.txt
```

Run the server.

```
python lab3-dashboard.py
```

```
Dash is running on http://127.0.0.1:8050/

 * Serving Flask app 'lab3-dashboard' (lazy loading)
 * Environment: production
   WARNING: This is a development server. Do not use it in a production deployment.
   Use a production WSGI server instead.
 * Debug mode: off
 * Running on http://127.0.0.1:8050 (Press CTRL+C to quit)
```

visit http://127.0.0.1:8050/.