

**Miranda Deane
Jack Tregurtha
Gabriel Gonzalez**

INFO/CS 3300

**Project 1: Static Visualization
TA: Brandon Giraldo**

**A Look into the post-grad Life for
Cornell Engineering Computer Science
Students**

Part A.

The data on the cities that recent graduates first started working was acquired from http://www.engineering.cornell.edu/resources/career_services/students/statistics/postgrad_reports.cfm. The variables that were included in the report from each year were: employer, job title, state, and city. We put the tables from each year into an excel sheet, added the year they graduated as a column, and then converted that to a csv file. To get the coordinates of each city we put a list of all the cities (with duplicates) into <http://www.gpsvisualizer.com/geocoding.html>, which gives a list of the coordinates. The coordinates were then appended to a csv file. We then used a csv-to-json converter to make that data into a json format we could use. From that json, we only considered the year and coordinates of each recent graduate to map the data to the map. The data for the map of the us was from the us.json file which was given in class. The same method of writing data to excel and converting to a JSON was used for the city data acquired from the Bureau of Labor Statistics.

The data on the median salaries was also acquired from http://www.engineering.cornell.edu/resources/career_services/students/statistics/postgrad_reports.cfm. We manually input into an excel sheet the year and median salary of the recent graduates, and then converted that into a csv file. The same csv-to-json converter was used to make that into a usable json file. Because we manually input the data, we used every attribute of each json object (the year and median salary). The data from the median CS related jobs in NYC, Seattle, and San Francisco came from <http://www.bls.gov/oes/2011/may/oes151131.htm#st> and was put into an excel sheet. As with the other data sets, it was converted to a csv file and then to a json file. The data set included the year and the low, high, average, and median salary. We only cared about the year and the median salary so those were the only parts used in the visualization. There were other cities included in this data set but because NYC, Seattle, and San Francisco was where recent grads from 2011-2014 most frequently ended up those were the only cities included in the visualization.

The data for the last chart was acquired from the Bureau of Labor Statistics website. Occupational Employment Statistics for Computer Programmers was found from <http://www.bls.gov/oes/2011/may/oes151131.htm#st>. There was a lot of data on the page, so we had to select which parts of it we wanted to include. The variables for this chart were median salary of computer programmers in the city, number of jobs available in the city, and percent of graduating students going to the city. Median salary data was recorded in order to compare these values with Cornell student values. The cities San Francisco, New York City, Chicago, Boston, Seattle, and Atlanta were chosen because they seemed to have the highest salaries in the nation according to the website. We wanted to compare if pay rate affected where students went, so choosing these cities seemed like a good idea. Additionally, they were chosen because they are known to be popular cities for students to go to so it made sense to sample from them. Data for the percentage of students going to each city was acquired from the Cornell CS website (<https://www.cs.cornell.edu/undergrad/cscareers/placementreport>) and also had to be manually recorded.

Part B.

In the map visualization of where recent graduates' first jobs were, a `d3.AlberUsa()` projection was used for each coordinate pair, and the map of USA was generated from the `us.json` file from class. The size of the circles corresponds to how many recent graduates ended up in a particular city. The circles were colored based on the year of graduation.

The visualization of the median salary has the x-axis correspond to the year of graduation and the y-axis correspond to the median salary. A d3 linear scale was used for both the x and y axes. The lines were colored based on whether the median salary came from the graduating class of cornell, NYC jobs, Seattle jobs, or San Francisco jobs. The lines were created with d3 functionality that was demonstrated in class.

For the final chart there were three variables to present. These were percentage of students going to each city, number of jobs available in each city, and median salary of computer programmers in each city. Percentage of students going to each city was represented with a line connecting cornell to the city. The thickness of this line was the percent value. If there is no line, no students reported going to that city. A linear scale was used. The domain varied from 0 to 0.35. Number of jobs available in each city was represented by size of the circle. This made sense because we also used a size to show percentage of students. This also makes it easy to see how big the job market is in each city as time goes on. A square root scale was used because the element representing the number was a circle. Thus the radius needs to be on a square root scale. Median computer programmer salary was showed with a color scale. Green was chosen because it represents money. A linear scale between two colors was used because we had minimum and maximum values for median salary. Text was used to label each city.

Part C.

We looked into Computer Science graduates from Cornell. We started with the individual, showing where people went after graduation and how much they made. Then we tied it into the country as a whole, showing the amount of job openings per city and average salary per city. Then we tied them together, comparing the two. Showing how many Cornellians go to each city, how much they make, and compared that to the average. It is surprising that so many Cornell graduates go to California for jobs when they do not have that many job offerings while almost no one goes to Atlanta or Chicago. It was interesting that despite how high-paying Seattle seemed to be, Cornell students don't flock there as much as San Francisco, which has fewer jobs but is growing. Also, New York had many job opportunities but still attracted fewer Cornell students than San Francisco. The data we have is not enough to determine why students might choose to go to one place over another. More research would have to be done into determining why a students are going to certain places.