Julie Rosenbaum // jar892
Database Design and Web Implementation
Assignment 4: Working with Data in SQLite
Part 3: Write-up

This dataset was obtained for the Data Visualization Assignment (#3). I retrieved this dataset, Quarterly Census of Employment and Wages Historical Annual Data: 1975 - 2000, from data.ny.gov. It includes data about the average wages and average employment across several New York State over 25 years. It offers a very interesting view of employment in New York state, broken down by county. The Standard Industrial Classification System outlines the 10 industry categories as see in the dataset, such as Agriculture, Finance, and Retail.

Through my queries and analysis, some obvious conclusions were reached. New York, NY and its surrounding counties were the most highly-employed with wages to match. Additionally, industries like finance and manufacturing employed the most New Yorkers, depending on the decade. For instance, in 2000, finance ruled, but in 1975, categories like wholesale trade and manufacturing did. These reflect a very clear shift in society away from blue-collar / low-paying jobs during this time period. In 1975, the average weekly wage in a manufacturing job in Tioga County was comparable to a job in New York City in wholesale trade or construction, where as in 2000, a finance job in New York City was making over double the amount any most other industries and counties (see query #2). The county-by-county breakdowns were what made this analysis both interesting, yet difficult to understand and manage.

Because this dataset was retrieved from a government website, I trust the accuracy and precision of the data. In some of my queries I was able to calculate some of the numbers that are pre-calculated in the table and they matched so at the very least the math is correct. The limitations of this dataset come from the lumping of large subcategories with broad industries like "services," which I found vague, lacking inclusion, and somewhat outdated. Most of the categories are for working-class professions, which are decreasingly common today.

To make this dataset more interesting, other data about New York State would be useful for comparison, like population and race information, as well as more specificity within each category of employment. For this assignment, I was attempting to become better acquainted with SQLite, and therefore did not focus as much on making my results as interesting as possible. To remedy this, I might have included more calculations of the data and compared sums and averages to other counties and industries, as I did in my last two queries. Additionally, it would be interesting to look at the changes over time (perhaps within each county, compared to the whole), but I was limited by my knowledge of SQL at this time.