For our income data we used datasets from the IRS website:

<https://www.irs.gov/uac/soi-tax-stats-individual-income-tax-statistics-zip-code-data-soi>

We focused on four years – from 2006 to 2010 since only the datasets from these years had similar format.

Using csvcut to extract the columns that interest us the most.

The columns that we extracted from the original datasets were the following:

1. STATE: The State associated with the ZIP code
2. ZIPCODE: 5-digit zip code
3. A00100: Adjusted Gross Income
4. A00200: Salaries and Wages

Similarly to the education part, we loaded our newly acquired csv tables with income data to the database.

As a result we come to over 700,000 records.

Voila! In view, we have 5 relatively comprehendible tables containing information about income by zip code across the nation.

We did not run quires for income alone, as I recall

Overall, the major difficulty we faced during our work was getting and cleaning data from IRS. The lack of consistency in datasets and unclear structure limited the timeline we could analyze to 5 years. However, information collected by IRS proved to be the only free source of information on income by zip code in the US. Given this fact, our project in its current form already has a lot of value, as well as potential for further development.

The quality of our analysis could have been significantly better had we included population data, and that is something that we will take into consideration in our future projects.