Research about the 2007 USA Tax and Income dataset

--By Zhanxi Shi

Summary of the research questions:

1. Which State and County has the most Adjusted Gross Income? Besides, is the County with the most Adjusted Gross Income belongs to the state with the most Adjusted Gross Income?

California and Los Angeles County. Yes

2. How much does the salaries and wages counts in the Adjusted Gross Income in USA? Does the state and county with the most Adjusted Gross Income falls into the same percentage?

About 72%. Yes.

3.Obviously the dataset doesn't give out all of the incomes, while how much does other income counts in the Adjusted Gross Income? Does that vary for the mean and the state/county with the most Adjusted Gross Income?

About 22%. No.

Motivation and background:

The economical background of the USA is kind of complex with sorts of investigation, organizations, and other international corporations. Since the time changes, the economic center of a country should also be changing. Several decades ago, Detroit and its surrounding cities were quite developed compared to others, and New York had it financial character that leads the world. While now, these cities' charm decreases and others' increases. To evaluate what is the modern leading region in economics in USA and how they contribute to that goal, we can analysis the dataset of Tax and Income.

Dataset:

The dataset used on this research is the Tax Year 2007 County Income Data released on the Data.Gov. Here is the link:

https://explore.data.gov/Population/Tax-Year-2007-County-Income-Data/wvps-imhx

I am operating this data with my code as a csv format since the Digit Reader for csv files is really convenient for cleaning up the data and obtaining information from dataset like this. Since I didn't make an assertion on the existence of the file, download it from that page in csv form in the following steps:

Export   -> CSV   -> Done

Then put the downloaded csv file within the same directory of the source code. Everything is set then.

Methodology:

First, if we want to find out the state and the county of the most Adjusted Gross Income, we could check over the dataset and find out the results related to the maximum Adjusted Gross Income. This is not that hard because we only need to do some simple comparison. Second, if we want to find out the percentage of a salaries and wages to the gross income of all states or counties, we shall find out each of them and compute the mean. Third, the other income that is not in the dataset should be the gross income minus the wages and salaries income minus the interest income minus the dividend income. This result is not the exact number of the other income since the tax is not counted in this formula. Yet, the tax of each county and state is not given so we could only calculate it like that to get a minimum possible value for other income. Besides, the percentage difference of A and B should be $(A-B)/((A+B)/2)$.

The reference of how to calculate the Adjusted Gross Income and other values could be found here: http://www.adjustedgrossincomecalculator.net

Results:

1.The state with the most Adjusted Gross Income is California and the county with the most Adjusted Gross Income is the Los Angeles County. These could be verified if you check the "More Views" button on that page. The result of this shows that California is kind of leading part on business environment.

2.The Adjusted Gross Income could show that it is the richest part in USA, but what actually contributes most to these? Check the dataset, and we could figure out that the wages and salaries is always the most of the Adjusted Gross Income. Yet, is it still the most part for the average and for the California or Los Angeles? The actual

computer results shows that:

The mean percentage of wages & salaries to gross income of all the States is: 0.722211091966

The mean percentage of wages & salaries to gross income of all the county is: 0.708222306204

The percentage of wages & salaries to gross income of the State with the most gross income is: 0.732793755065

The percentage of wages & salaries to gross income of the county with the most gross income is: 0.742530960704

The percentage difference between the mean and that of the most gross income state is: 0.014546567484

The percentage difference between the mean and that of the most gross income county is: 0.0472977111721

The percentage of wages and salaries to the overall income is always about 72% even though it shows a bit higher in California and Los Angeles, while the percentage difference is below 5% so it quite acceptable. This results shows that the wages and salaries is still the main income in California and Los Angeles. In other works, the structure of the income system doesn't change much.

3. Only few types of income is given so there must be some other ways of income. Since the salaries and wages only counts about 72%, the other shall be counted to see whether they should be neglect. After the computation, the results show like this:

The average percentage of the other income to the adjusted gross income of all states is: 0.22415195161

The average percentage of the other income to the adjusted gross income of all county is: 0.23788939329

The percentage of the other income to the adjusted gross income of the state with the most gross income is: 0.208237396377

The percentage of the other income to the adjusted gross income of the county with the most gross income is: 0.19424768419

The statistics shows that about 22% of the overall income is not shown in the dataset, and it could be more than that since the tax is counted as zero when calculated. This is not a small amount of the overall part. The reason that the government didn't post it could due to its complexity or it irrelevancy to the government.

Reproducing the results:

Just operating the source code on idle and the result would be reproduced since I setup three united print methods to print out the results of the research.

Collaboration & Reflection:

No one actually help me or work with me on this final project.

The environment of the USA is not that complex since the salaries and wages take over 70% of the overall income of the people. Yet there are some income not clearly shown by the government and the exact tax could not be calculated based on the government tax and income dataset. Is the government hiding it? Maybe yes, maybe not.