Python Project Report

Fall 2019

Census 2000 PUMS Data

Submitted by

Snehal Ilawe 824655189

1. Introduction

The Census 2000 PUMS files have state-level data containing individual records for a 1

percent sample of people and housing units. The 1-percent files give users the maximum

amount of social, economic, and housing information available. The 2000 PUMS file

structure is hierarchical and contains two basic record types of 314 characters each: the

housing unit record and the person record. These data files are downloaded from

<https://www.census.gov/prod/cen2000/doc/pums.pdf> website. For the project the 1

percent data file is downloaded for following 10 states:

1. Alabama
2. California
3. Idaho
4. Louisiana
5. Hawaii
6. Mississippi
7. Montana
8. New Jersey
9. Alaska
10. Texas
11. Analysis and Results:

All of the above 10 state files are in the text format with a hierarchical structure with housing unit record and the person record. For each of the state file only required subset of columns are selected such as Race, Education, Income, Gender, Age, Relationship. The records are stored in a CSV file. The records are read from CSV file which is now formatted with headers. All of the state file records are appended into a single data frame with new column State representing the name of the state to which these records belong.

For analysis of income levels for a variety of levels of education state by state the records where Income is 0 are filtered out. To represent different levels of education the Education field is mapped into a new field Degrees which consist of following bins:

1. Less than High School where educational attainment is less than 9th grade
2. High School not Completed where educational attainment is between 9th -12th grade
3. High School Diploma
4. College Student with No Degree
5. Associate Degree
6. Bachelor`s Degree
7. Master’s Degree
8. Professional Degree
9. Doctorate Degree

Median Income is calculated by grouping the records by different Degrees for each State. Median incomes are used as measures because they tend to more accurately represent what people make in a given area. A Line Graph can help to visualize the change in median Income for different education levels for each of the 10 states.

Using Median Income, a Line Graph is plotted which shows median income(y-axis) in all of the states (x-axis) where colored lines indicate various education level.

A close up of text on a black background

Description automatically generated

Fig 1. Median Income in all 10 state for various Education levels

From Fig1 we can observe that highest level of education such as a Doctorate and Professional Degree have Median income approximately 5-10 times more than the lower education levels such as High School not completed. We can plot a horizontal bar chart as below to visualize it.

A screenshot of a cell phone

Description automatically generated

Fig 2. Mean Income for Doctorate and lower education levels by state

From Fig 2 we can observe that for states like Alaska and Montana the Median Income for Doctorate Degree is 10 times more than that for High School not completed. For states like California, Louisiana the Median Income for Doctorate Degree is 5 times more than that for High School not completed

From Fig1 we can also observe that the State of New Jersey and California have higher median Income than most of the states. We can plot a bar graph to study Income difference for states of California and New Jersey.

A screenshot of a cell phone

Description automatically generated

Fig 3. Median Income for different education levels for California and New Jersey

From Fig 3 we can observe that New Jersey state has higher median Income for most of the Education levels except for the education level High School not completed where the Median Income is same for California and New Jersey. For Professional Degree there is a marginal difference in median Income in both the states. Also, the Professional Degree Income in the state of California is higher than the Income for Master`s degree. For the New Jersey state the Median Income difference for Master`s Degree and Professional degree is not significant.

From the Fig1 we can also observe that Doctorate Degree has higher Median Income for most of the states except for states like Louisiana, Texas, Alaska and Mississippi. Therefore, we compare the incomes of the doctorate and professional degrees by plotting a boxplot these states.

A screenshot of a cell phone

Description automatically generated

Fig 4. Box Plot for states of Louisiana, Mississippi, Texas, Alaska

From the Boxplot we can observe that the median incomes for Professional degrees has higher distribution of Income for all of the 4 states. But the median Income in almost equal to the doctorate degree incomes for Louisiana and Texas. Whereas for Mississippi and Alaska the median Income for Professional Degree is higher than that of Doctorate Degree. This might be due to the higher salary commanded by professional degree holders working on offshore oil rigs and other similar high-risk jobs.

Also, from Fig1 we can observe that the Median Income for people with education level of “Less than High School” have more median Income than people with Education level of “High School not Completed” for all of the 10 states. Bar chart can be used to visualize this data.

A screenshot of a cell phone

Description automatically generated

Fig 5. Median Income for Lower Education levels

From Fig 5 we can observe that there is a significant difference in Income between “Less than High School” and “High School not Completed” education levels in states like Idaho Alaska and Montana.

**Part 2:**

For this part the income levels for a variety of levels of education for whites and non-whites are compared. Whites and Non-Whites are identified using a “get\_race” function. The function “get\_Race” returns Race as ‘White’ if the value in the argument being passed is ‘47’. For any other value it returns as ‘Non-White’.

Bar Chart is plotted for state of Alabama to compare the income levels for a variety of

levels of education for white and non-white population.

A screenshot of a cell phone

Description automatically generated

Fig 6. Whites vs Non-Whites Median Income for Alabama

From Fig 6 indicates that the median Income for White population is higher for most of the education levels except for Doctorate Degree where the Non-White Population have higher median Income and population with Less than High School have an approximately similar median income range. There is significant difference in the median Income for the White population for Professional Degree than the Non-White Population.

For state of Idaho a bar chart can be plotted to compare the income levels for a variety

of levels of education for white and non-white population.

A screenshot of a cell phone

Description automatically generated

Fig 7. Whites vs Non-Whites Median Income for Idaho

Fig 7 gives insights that in the state of Idaho the Lower Education levels have higher

median Income for the Non-White population. For Masters and Professional Degree, the median Income is same. For Doctorate Degree too there is not much of a significant difference. Whereas for bachelor’s and High School not Completed the White population has more median Income.

A screenshot of a cell phone

Description automatically generated

Fig 8. Whites vs Non-Whites Median Income for Alaska

From Fig 8, a surprising observation can be seen for the state of Alaska. There is no Income data for Non-White population for Doctorate Degree. Master’s degree and Professional Degree Median Income is same for White and Non-White population but for Professional Degree there is considerable difference in the median Income between White and non-White population.

A Line Graph can be plotted for different education levels to view a clear discrepancy in the Income between the White and Non-White population.

First Line Graph is for Doctorate Degree across all 10 states for the White and Non-White

Population.

A close up of a map

Description automatically generated

Fig 9. Doctorate Degree Median Income for White and Non-White

Fig 9, gives understanding of the Income levels for the Doctorate Degree across the 10

states. Most of the states have a higher median income for the White population except

for states like Alabama, Mississippi and Montana where the non-white population have a

considerably higher median income.

For Professional Degree except for Hawaii and Idaho all the other states have a higher

median income for the White Population.

A close up of text on a white background

Description automatically generated

Associates Degree have higher median income for White population throughout all of

the states.

A close up of text on a white background

Description automatically generated