Demographics and Financial Crisis

Questions

How were different demographics (sex, and race) affected by the 2008 recession? Also, how are different demographics compare in economic prosperity over the span of 2000 - 2014?

The project: an introduction

In the following pages we will describe our findings based on gathered statistical data from IPUMS USA (https://usa.ipums.org/usa/), regarding *household incomes* from 2000 to 2014. We analyzed data considering both *gender and race*. In particular, we divided the race in Caucasian, African American, Hispanic and Asian (Chinese and Japanese) and studied males and females respectively. Our thesis is based on proving the existence of some fundamental differences between both gender and race caused by the *financial crisis of 2008*. In conclusion, we manipulated the purchasing power and put it in terms of the year 2015 because we wanted to have an even perspective on all the values present, since the value of the dollar has changed in the last 15 years. Furthermore, we had to adjust our data to remove the **outliers** that went skewed the data to far up, since it would not show the true story of the average American household. For this reason, we eliminated every income which exceeded the the reasonable maximum of **2 million dollars**. In fact, some data may be not available or missing and was assigned a value of 9999999. eliminated the **outliers** in household income

Summary Statistics

Here we present all the main information that we used in order to proceed with our studies. Since we decided to assess the financial crisis and the impact in terms of household income, our summary statistics provides an overview of both the general situation over 15 years and the breakdown by gender and race.

Summary Statistics Overall Sample:

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -25090   33090   62070   82350  103000 1570000
```

It can be seen that the *average* income from 2000 to 2014 was **82'350 dollars**. The least amount of money earned was **-25'090 dollars**, experienced in *2003*. Whereas, the highest peak of **1'570'000 dollars** was recorded in *2006*. Curiously enough, during the very preceding years of the financial crisis. We can also estimate how volitile the income has been, with an overall **standard deviation** of:

```
## [1] 82187.61
```

Summary Statistics Race & Gender:

White

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -25090   35640   65180   85940  107000 1570000
```

Black

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -12550   19540   41100   54900   73290  915000
```

Hispanics

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -15410   27520   50420   65770   85020 1040000
```

Japanese

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0   39400   75290  101600  123000 1040000
```

Chinese

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -9531   36930   82600  106300  152000  940000
```

Males

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -25090   35790   64660   84830  106000 1570000
```

Females

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -13380   30520   59450   79860  101000 1570000
```

A lot of information is provided in the previous summaries. It can be seen that the highest income was earned by both a man and a woman of caucasian origin. On the other hand, the minimum, negative income, was associated to a white man. Unhappily, on average, the lowest income was experienced by **African Americans**, with **54'900 dollars**. Wheras, Asians recorded the highest

earnings, with **106'300** and **101'600 dollars** for *China and Japan* respectively. It can also be observed that the **Japanese** returns were the most volitile over the period considered, with a sample **standard deviation** of:

```
## [1] 106957.2
```

compared to the others of White, Black, Hispanics and Chinese respectively:

```
## [1] 84855.13
```

```
## [1] 53091.33
```

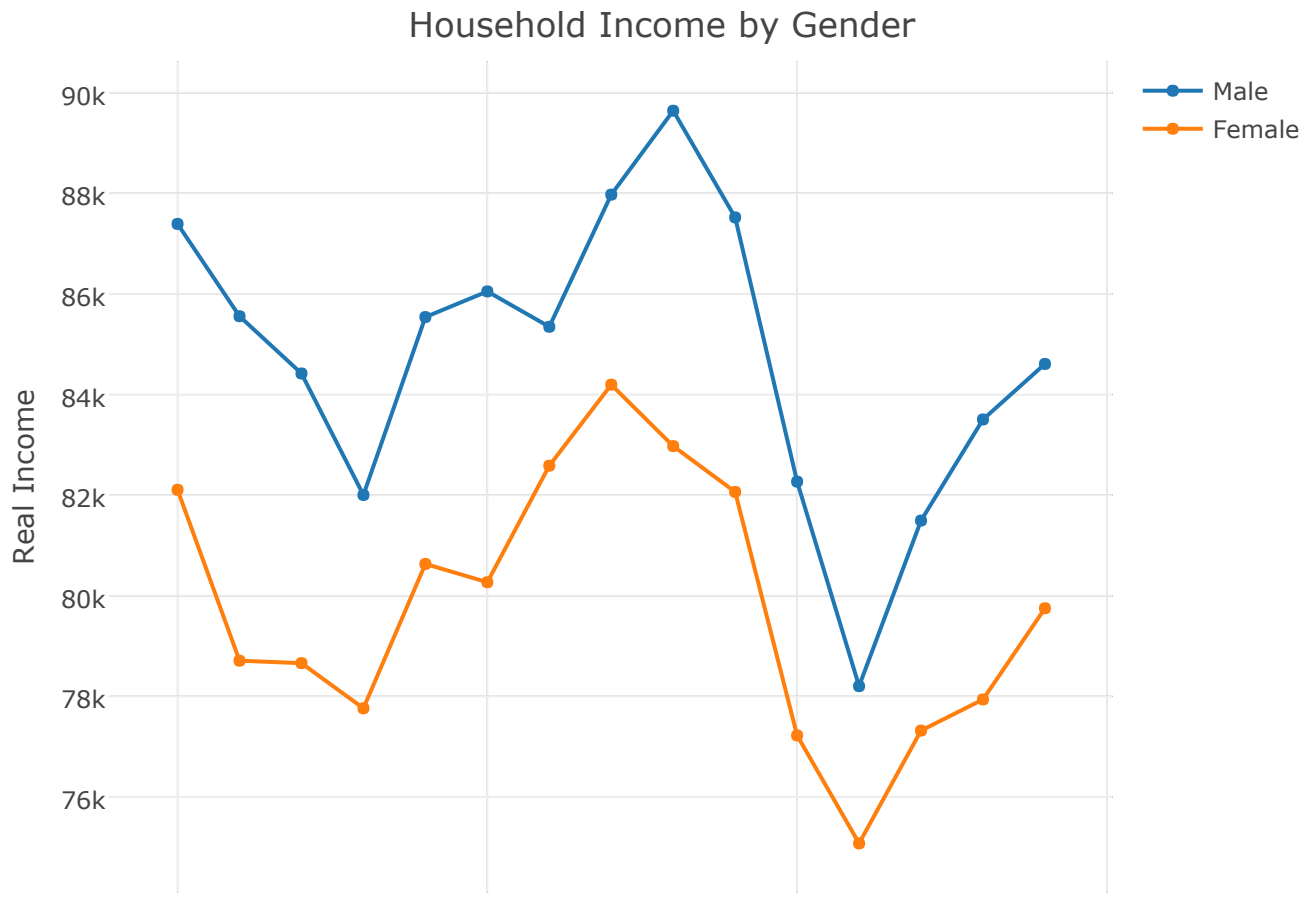
```
## [1] 61150.5
```

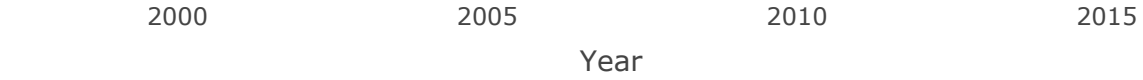
```
## [1] 99079.75
```

Graphs

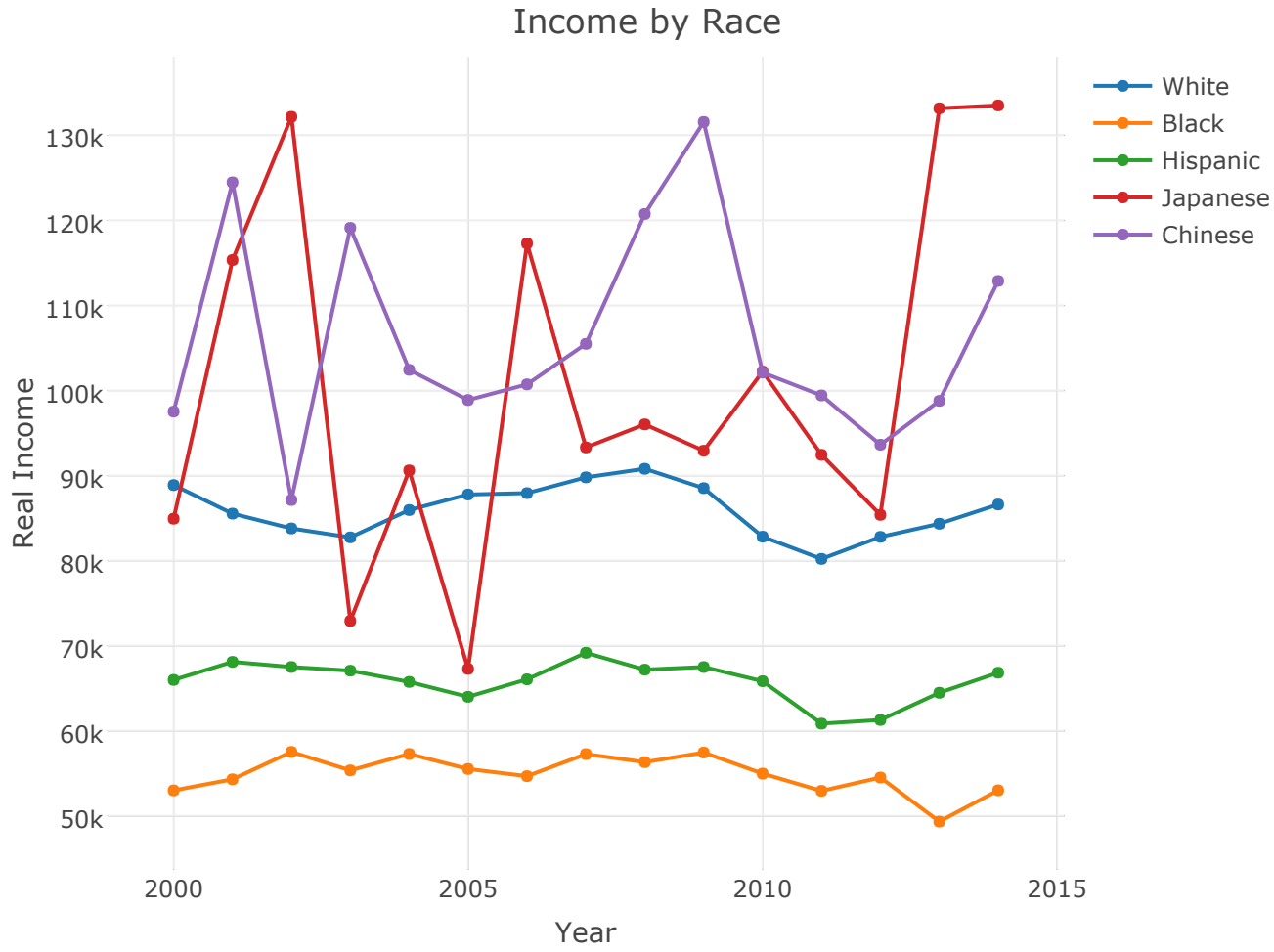
Here is a visual representation of inflation-adjusted household incomes between 2000 and 2014. Each number represents the average of income per sigle year.

Gender:





Race:



Hypothesis Tests

As far as our computations are concerned, it is important to outline two facts:

1. The tests statistics will be represented by the difference between two variables (income mean men and women).
2. We took advantage of the **Central Limit Theorem**, since our sample size, for both variables, is greater than *100'000 units*. For this reason, we decided to consider the distribution as a standard normal.

Overall population: Gender

The *Gender Chart* provides a good visualization of the difference in income between males and females. In order to provided a simplier representation, the chart outlines only the averages over a period of fifhteen years. However, in verifying our conclusions, we took into consideredtion all the sample size. This allowed us to present more reliable results.

Hypothesis tests:

Women and Men's income is the same. In other words, the two groups have the same compensation.

In mathematics terms, we presented our **NULL Hypothesis (Ho)** as follows:

$$\mu(Male) = \mu(Female)$$

On the other hand, the **Alternative Hypothesis (HA)** was defined under these terms:

There is an existential difference in income between the genders

$$\mu(Male) \neq \mu(Female)$$

Therefore, to confirm or disprove this hypothesis, we adopted a *two-sided test*. In particular, we chose to allow only a minimum degree of uncertainty, implying a **99% confidence interval**. Finally, before presenting the results, it is important to say that we assumed *independence* in the sample data. Indeed, this seems to be a reasonable deduction, since in almost every case, one and only one gender is attributed to one individual, and every person is different from each other.

The following steps indicate how we reached our conclusion:

- First we calculated our Test Statistics under the NULL Hypothesis, assuming therefore a difference of ZERO between the population means.

```
## [1] 14.5743
```

- We then derived the appropriate critical value, consistent with our confidence interval.

```
## [1] 2.575829
```

- Finally, we compared the two, in order to our hypothesis.

```
## [1] "14.5743 > 2.575829"
```

It can be seen, that the Test Statistics clearly exceeds our Critical Value, implying a *rejection* of the null with a 99% confidence. This confirms our empirical results, which outline significant differences between men and women's income.

Overall population: Race

How it can be seen in the previous findings, we proved, by rejecting the NULL, that there is a difference between men and women's compensation. We decided to investigate further a similar proposition focusing on income differences among races. Our data are presented in the *Race Chart*. The graph depicts five main demographics, however, we chose to focus our attention only on black and white people. For the same reasons as above, we decided to show only income averages from 2000 to 2014, but the hypothesis was stll tested considering the whole sample size.

Hypothesis tests:

Caucasians and African Americans earn the same. Or, there is not difference in income between the two demographics.

In mathematics terms, the **NULL Hypothesis (Ho)** is similar to the previous one:

NULL Hypothesis:

$$\mu(White) = \mu(Black)$$

The same applies to the **Alternative Hypothesis (HA)**: White people earns more than black people.

Alternative Hypothesis:

$$\mu(White) > \mu(Black)$$

As it can be seen from the Alternative Hypothesis, on this occasion, we considered only the upper tail of the distribution. In other words, we decided to adopt a *one-sided test*. In particular, we chose to set a 99% confidence interval distributed entirely on the right side. Finally, we assumed *independece* in our sample data.

Results:

- Instead of comparing the Test Statistics to the Critical Value, we decided to take advantage of the p-values. Always considering to operate under the NULL, we found the smallest alpha at which we would have rejected the null hypothesis.

```
## [1] 0
```

This means that, according to our findings, we would have never accepted the null hypothesis, since the smallasted level that would cause us to reject it is 0%.

In conclusion, even in this case, there is a strong evidence that leads us to reject the NULL Hypothesis with a 99% confidence interval, and necessarily to prove the existence of some differences in earnings between Caucasians and African Americans.

Results Discussion

Gender wise both men and women were increasing their household income up until the 2007-2008 years, then a drop occurred in household income. This drop, caused by the recession, ended up affecting, on average, men more than women. Men suffered from a steeper drop in income than women, with a plunge that continued until about 2011. Women seemed to suffer less under the recession, by recovering a year earlier than men.

As far as race is concerned, all races save for Asians seemed to have a steady increase in household income up until 2008. Caucasian Americans seemed to have a steeper drop in income more than any other race, however, in comparison to an Asian Nationality (Chinese), Caucasian American suffered relatively less. A drop of nearly **25k** in household income occurred to *Chinese Americans*. Hispanics

and Black Americans seemed to plateau up until about 2011. Hispanics then began to see an increase in household income while Black Americans seemed to suffer a dip form 2012-2013 however there seems to be a recovery occurring in the range of 2013-2014. The Asian Nationalities seem not to suffer from the same trend as Hispanics, Blacks and Caucasian Americans. They seem to undergo sporadic increases and decreases. The biggest jump however can be seen by **Japanese Americans** who say an increase in household income from about **82k** to about **134k**.

In the end, across most of our data, we did find that the recession did have an effect on about every demographic. In terms of economic prosperity, men seemed to be consistently more well off than women. Even though men suffered more than women during the recession, the formers still remained to keep their lead in income. In terms of race, Hispanics and black seemed to be relatively low income compared to Caucasians and Asians. However, we did see that Asian Americans did not seem to follow the same general trend that many other demographics presented. Also Japanese Americans seemed to make a relatively high jump in their overall earnings allowing them to be the most economically prosperous group by 2014.

Criticism and Extensions

The data seemed to have many issues that made us skeptical about their validity or trustworthiness. In particular, they suspiciously round perfect numbers that are not usually evident in income. Furthermore, the data recorded children in the dataset so those had to be removed from the sample since they represented outliers that would have skewed our results. One way to fix this would be to find out as to why the data was rounded to perfect numbers and work to take into account that manipulation or get new information from a different source. Moreover, we would have liked to have observed why women were less impacted by the recession than men. Was it the jobs that they took or some other unknown factor. Also we were unable to understand or glean where the sporadic jumps that Asian Americans experienced came from, and why they did not seem to follow the generally trend that Black, Caucasian and Hispanic Americans did. We would also have liked to see what industries were hit most and what demographics worked in those to better understand the trends that we were seeing. The data we were given trends but we have no means of identifying why these trends actually occurred. It would be in our best interest to actually expand the project to find what these factors were that influenced the trends we observed.

On top of the above critiques and suggestions, these other three criticisms are presented:

Criticism 1:

In considering our hypothesis we always assumed independence. However, this could be misleading, since we did not distinguish the sample between married couples and single people. Indeed, when combined income is considered, if one element of the couple earns a significant amount of money his or her partner could be disincentivized to work and consequently reduce the income. This could certainly imply a level of dependence within the sample.

Criticism 2:

We confirmed our hypothesis, by reporting the presence of a difference in revenues within the different demographics. However, we still do not know what is the most probable cause for these discrepancies. It could likely be a discrimination issue, but in order to be sure, we should have kept constant certain factors. For instance, level of education, working hours and conditions, personal merit and effort.

Criticism 3:

Our conclusions took into consideration only one limited variable to address possible inequalities. Indeed, we based our recommendation looking exclusively at household income. However, we did not consider possible elements which could have affected our results. Tax deductions is one of this, since some individuals could potentially employ different strategies in order to pay less taxes and report a higher income.