Final Project: The Gap in Earnings

Emerald Bismaputra[1]

[1] University of Arizona

Author Note

This paper was created for APCV 302 - Statistics in the Information Age, a class in the Applied Computing program at the College of Applied Science and Technology.

Final Project: The Gap in Earnings

The dataset that I chose is from a 2021 dataset called Employment and Earnings found at Mock (2021) and based on a 2018 article "blacks in the labor force" about the increasing number of blacks in the labor force that includes their participation rates, educational attainment, and employment in occupations. In this report, we will explore further the following: how race and gender will affect people's median weekly earn to see if there is any correlation between them.

Review of the literature

The topic about income gap between races and income gap between genders have been a controversial subject and a talking point for quite a while. In fact, a nationwide survey organized by CNN and Kaiser Family Foundation found that 49% of United States citizens thinks that racism is a serious complication in this era.

According to "Magnitude and evolution of gender and race contributions to earnings inequality across US regions" by Frederic et al (2020), income inequality among races and gender are highly significant. They did a study about how having a certain ethnic background or gender may have lower their chances of getting a good income. However, at the end they have also stated that the rates of racial contribution affecting earning inequality is only about 1 to 4 percent, which is really low. Later on in the analysis, it was concluded that while racism affecting income do exist, the gap between the blacks and whites are mostly associated and caused by other things such as low education, poverty, etc.

An investigation in a journal called "why do minority men earn less? a study of wage Differentials among the highly educated" written by Dan Black et al (2006), the researchers be of the opinion that income inequality is related to the poor education these people get. Dan black and the other authors said that "in particular, among college-educated men who speak English at home, our estimated wage gaps are very close

to 0 for Hispanic and Asian men. Similarly, the unexplained wage gap is approximately 0 for black men with college-educated parents not born in the South." The estimation claimed in this article, and the statement stated in the article before this proves that many researchers are having the same conclusion as to why a lot of minorities earns less.

In regards to gender, the income inequality between men and women is much larger than the inequality between races in the past years as discussed by Frederic et al (2020). Their research have shown that earning inequality associated with gender have decreased in the recent years, while on the contrary, earning inequality associated with race keeps on increasing going forward to the future. This was proven by frederic and his associates as they revealed that "Gender indeed explains a much higher share of earnings inequality (about 10%) whatever the geographical division. Unlike race, however, inequalities associated with gender were noticeably lower between 2005 and 2017." Even so, another study by Mandel and Semyonov (2014) stated that their findings indicated that working hours and occupational segregation are the ruling factors of the pay gap between men and women. Mandel and Smyonov believes that "Because a requirement for demanding working hours cannot meet the domestic division of labor at home, until fundamental changes in the latter take place, women will not be able to equal men in terms of their labor supply."

Despite of the cause of income inequality between races and genders, studies have shown that the pay gap do exist and keeps on going for at least the next few years. This interest me because I would like to find out the gap difference between all the variables and reveal the correlation they have between each other. I believe that there are a lot of unanswered questions regarding this issue such as: What makes women and minority race earn less inside of United States? how much really is the gap in income between them? and lastly, how much longer will this income imbalance continue?

We have far to go before equivalent income is a reality for women of color. Race and gender wage holes hurt women monetary security and their families, and amount to

significant misfortunes after some time. A women of color who works every day and all year, it is estimated that they can lose about 1 million dollars in pay for over a 40 year profession in view of the income gap. The wage hole is large, particularly for women of color, that is the reason we need to continue to push the government to make solutions that can solve the issue.

In investigating the research problem that will be discussed further in the next section of this report, a few things came to mind. While the dataset came from a survey conducted in the United States, being one who lives in Indonesia, I would definitely say that the gap in earnings are as significant. When I was in high school, it came to my attention that there were white teachers that earned more than south east asian teachers (even teachers of Indonesian descent) which was surprising to me. Aside from that, it is no question that the concept of 'women empowerment' is not normalized here and thus, most women are unemployed. However, women who are employed do earn way less than men which is really unfair considering they are putting the same effort/work. The two stories made me realize how widespread these issues are and that there should be a way to lower the gap between earnings.

Research problem

According to the article, the Blacks in the labor force have had an increase of employment rate and earnings within the United States. While the Blacks have had an increase in employment rates and earnings, it is possible that the two variables have increased to a certain extent. With that being said, I would like to investigate the extent to the said increase and figure out the differences of earnings in comparison to other races. Additionally, I will also be taking into account both genders (Men and Women) to see if it also contributes to the gap disparities.

Design

The data used in this project is part of the `tidytuesday` project (Mock, 2021). R packages used in the analysis include: R [Version 4.1.0; R Core Team (2021)] and the R-packages *ggplot2* [Version 3.3.5; Wickham (2016)], *magick* [Version 2.7.3; Ooms (2021)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], and *RColorBrewer* [Version 1.1.2; Neuwirth (2014)].

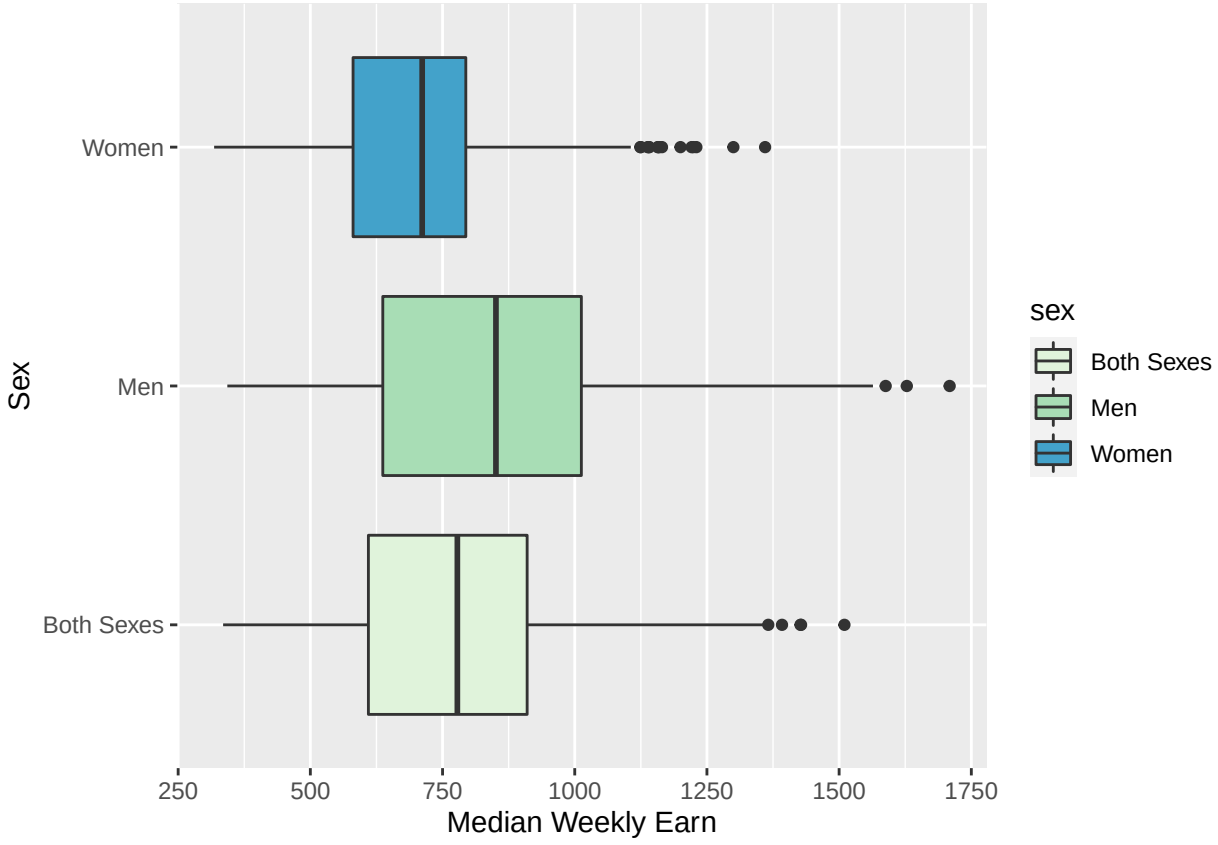
The chosen dataset was derived from the Current Population Survey (CPS) which is a monthly survey conducted by the Bureau of Census and the Bureau of Labor Statistics. For all people surveyed in the case of this dataset, it is assumed that they were all employed and in the labor force. In this investigation, I will be making use of the regression analysis (t-test) to find out if the median weekly income of men and women are significant. Moreover, I will also be making use of the correlation coefficient to find out if 'number of persons employed by group' and 'the median weekly earning in current dollars' are correlated. As for the plots, I will be illustrating a box plot for median weekly earning against race and another box plot for median weekly earning against gender.

Results

```
##      sex                race           ethnic_origin          age
##  Length:4224        Length:4224        Length:4224        Length:4224
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       year         quarter       n_persons         median_weekly_earn
##  Min.   :2010   Min.   :1.00   Min.   :   103000   Min.   : 318.0
```

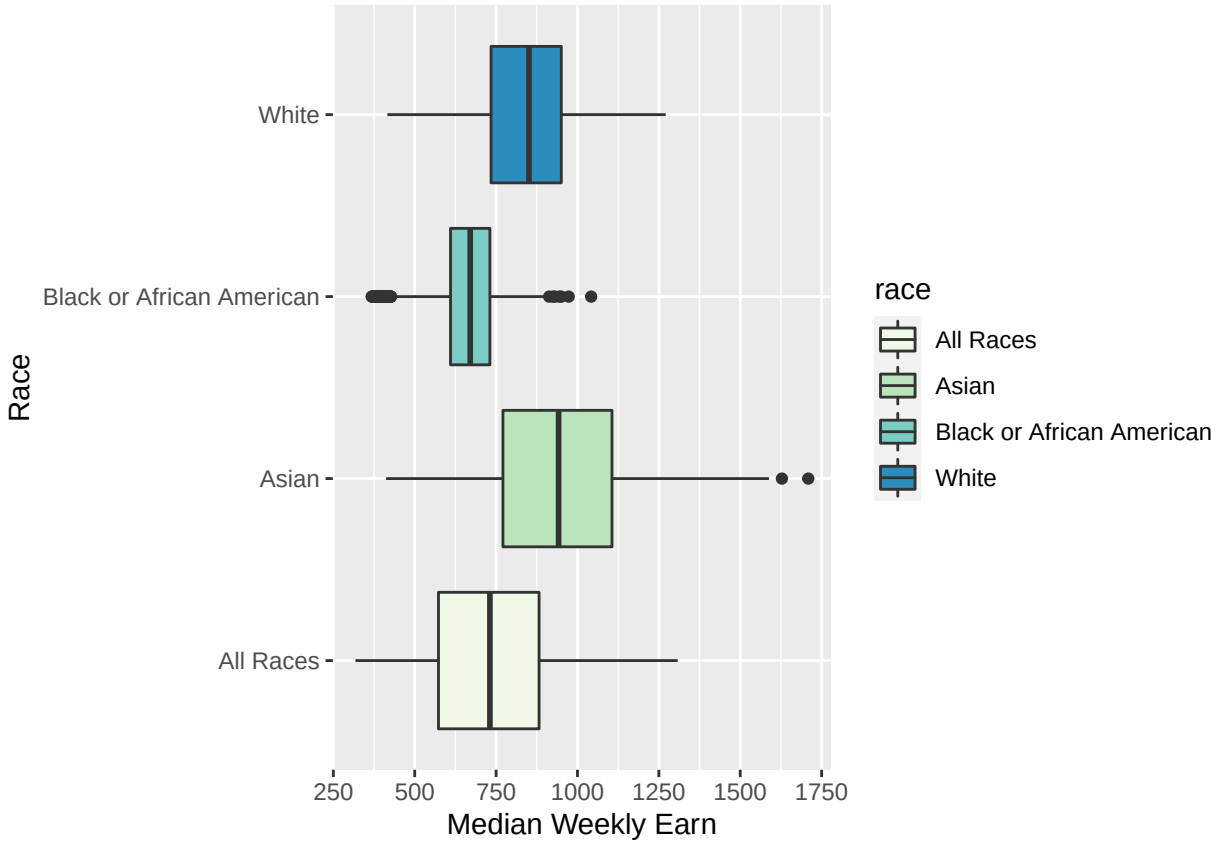
```
##  1st Qu.:2012   1st Qu.:1.75   1st Qu.:  2614000   1st Qu.: 605.0
##  Median :2015   Median :2.50   Median :  7441000   Median : 755.0
##  Mean   :2015   Mean   :2.50   Mean   : 16268338   Mean   : 762.2
##  3rd Qu.:2018   3rd Qu.:3.25   3rd Qu.: 17555250   3rd Qu.: 911.0
##  Max.   :2020   Max.   :4.00   Max.   :118358000   Max.   :1709.0
```

Seen above is the summary of the dataset which shows the maximum, minimum, median and quantiles of the chosen dataset. This summary also classifies the type of data that is available as the qualitative variables do not have the extra information mentioned previously while the quantitative variables do.



To visualize my findings, I made use of boxplots. The above plot is a boxplot that shows the number of people employed in regards to gender. As expected, the median weekly earn of the males are greater than of women. The male whiskers shows that the maximum median weekly earn of a male is just a little over 1500, with outliers beyond that

figure. On the other hand, the maximum median weekly earn of women is around 1125 (around 15% less than males), while outliers that exceed the maximum are still lower than of the males. However, the median of the weekly earn is not that far apart. Disregarding the 'both sexes' category, this visualization indicates that males are still earning more than women regardless of their race.



In terms of race, the boxplot above shows the differences of median weekly earnings in association to a person's race. Surprisingly, the median weekly earn of Asians are higher than the Whites, however, unsurprisingly, the Blacks or African Americans earn the least among the races. The median of the Asian's median weekly earn is close to a 1000 with the outlier reaching 1750, while the Whites median is over 750 and the Black's median is approximately 700.

```
##
##  Welch Two Sample t-test
```

```
##
## data:  median_weekly_earn[sex == "Men"] and median_weekly_earn[sex == "Women"]
## t = 17.915, df = 2457.9, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  127.1169 158.3660
## sample estimates:
## mean of x mean of y
##  833.1648  690.4233
```

A t-test was done to figure out if the 2 variables chosen are significant or insignificant. In a t-test, if the p value is less than alpha, it is considered statistically significant. According to the information above, the p value is 2.2e-16 which is the scientific notation of 0.00000000000000022, meaning it is very close to 0. Thus, indicating a significant result.

```
## [1] 0.2843811
```

Lastly the correlation of the median weekly earnings and the number of people employed by grouped was found using the cor function to identify if the two variables are correlated at all. Two variables have a perfect linear correlation when its correlation coefficient is an absolute value of 1. However, in the case of these two variables, the correlation coefficient appears to be 0.2843811 which shows that these two variables are very weakly correlated.

‘‘‘

Discussion

After investigating the dataset, I can conclude that the findings of the earning gaps between races supports the article that was used to base this investigation as the article

states that the Blacks earn less than the Whites due to various reasons such as education and poverty. The data about the income gap between men and women also supports the articles as it shows that men mostly are making more than women. It is unsurprising that both of the findings are in-line with the articles.

163 **References**

164 Chantreuil, Fourrey, Lebon, and Rebière (2021)

165 Toossi and Joyner (2018)

166 Labor Statistics (2017)

167 Mandel and Semyonov (2014)

168 Black, Haviland, Sanders, and Taylor (2006)

169 Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*.
170 Retrieved from https://github.com/crsh/papaja

171 Black, D., Haviland, A., Sanders, S., & Taylor, L. (2006). Why do minority men
172 earn less? A study of wage differentials among the highly educated. *The Review*
173 *of Economics and Statistics*, *88*(2), 300–313.

174 Chantreuil, F., Fourrey, K., Lebon, I., & Rebière, T. (2021). Magnitude and
175 evolution of gender and race contributions to earnings inequality across US
176 regions. *Research in Economics*, *75*(1), 45–59.

177 Labor Statistics, B. of. (2017). Labor force statistics from the current population
178 survey. US Department of Labor Washington, DC.

179 Mandel, H., & Semyonov, M. (2014). Gender pay gap and employment sector:
180 Sources of earnings disparities in the united states, 1970–2010. *Demography*,
181 *51*(5), 1597–1618.

182 Mock, T. (2021). Tidy tuesday: A weekly data project aimed at the r ecosystem.
183 Retrieved from https://github.com/rfordatascience/tidytuesday

184 Neuwirth, E. (2014). *RColorBrewer: ColorBrewer palettes*.

185 Ooms, J. (2021). *Magick: Advanced graphics and image-processing in r*.

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Toossi, M., & Joyner, L. (2018). Blacks in the labor force. *United States Bureau of Labor Statistics. Spotlight on Statistics*, 1–21.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org