

project

June 5, 2024

1 Wage Indifferences in the USA

Cassie Hoang | CSE 163 Intermediate Programming | Final Project

1.1 Summary

This project aims to analyze and explore racial and gender inequalities in the United States. Despite the USA being racially and ethnically diverse, minority groups seem to struggle financially, economically, and socially as opposed to dominant and privileged individuals. By researching these inequalities through average wages and incomes, we can determine if racial groups are disadvantaged or equally compensated, based on their levels of education and gender. Understanding this problem will help address racial inequalities and discriminations in the working force, which can be proven to hire or underpay individuals based on their race or gender, despite their levels of education. ## Research Questions *1. Does a higher level of educational attainment contribute to median and average wages? Is this growth exponential or is there a preferred minimum level of education?*

2. What is the significant difference, if any, in wages between Black and White men and women? Does gender affect wages and income between individuals?

3. Is there a significant difference in wages between Black and White individuals in the USA? How do racial difference play a part in income differences in addition to levels of education and gender?

2 Challenge Goals

1. Multiple Datasets: I will be using 3 different datasets for this project. Every dataset was found and extracted from Kaggle. All 3 datasets were chosen in consideration to the project topic to analyze similar variables. Two datasets will be merged to focus on analyzing similar variables. Each of my research questions will target information from all 3 datasets, with some questions being an accumulation of datasets.

2. New Library: With my project, I will display my results and analysis through data visualizations, specifically with graphs and charts to display numerical and categorical data. I will be using a new library, *plotly*, to visualize my data and hypothesis. I have imported *plotly express* and *plotly subplots* from the *Plotly* library.

2.1 Collaboration and Conduct

Students are expected to follow Washington state law on the [Student Conduct Code for the University of Washington](#). In this course, students must:

- Indicate on your submission any assistance received, including materials distributed in this course.
- Not receive, generate, or otherwise acquire any substantial portion or walkthrough to an assessment.
- Not aid, assist, attempt, or tolerate prohibited academic conduct in others.

Update the following code cell to include your name and list your sources. If you used any kind of computer technology to help prepare your assessment submission, include the queries and/or prompts. Submitted work that is not consistent with sources may be subject to the student conduct process.

```
[1]: your_name = "Cassie Hoang"
sources = [
    "education.ipynb",
    "mapping.ipynb",
    "https://plotly.com/python/plotly-express/",
    "https://stackoverflow.com/questions/73488602/line-graph-using-plotly",
    "https://www.geeksforgeeks.org/how-to-rename-columns-in-pandas-dataframe/",
    "https://plotly.com/python/bar-charts/",
    "https://www.geeksforgeeks.org/how-to-group-bar-charts-in-python-plotly/",
    "https://stackoverflow.com/questions/56727843/
    ↪how-can-i-create-subplots-with-plotly-express"
]

assert your_name != "", "your_name cannot be empty"
assert ... not in sources, "sources should not include the placeholder ellipsis"
assert len(sources) >= 6, "must include at least 6 sources, inclusive of
    ↪lectures and sections"
```

2.2 Data Setting

1. Median & Average Hourly Wages in the USA (1973-2022)

This dataset contains information on median and average hourly wages in the USA between the years 1973-2022. Wages are calculated to the 2022 dollar value in the USA. This dataset examines wage differences between men, women, and several racial groups. Wages for recent highschool or college graduate statuses are also included. This data is sourced from the Economic Policy Institute's (EPI) State of Working America Data Library. The context of this dataset specifically examines indifferences between Black, White, and Hispanic racial groups. I will be using data on Black and White individuals to correlate with my additional datasets. This dataset will give me insight on how race, gender, and graduate status can affect median and average wages in the USA, as well as how this information can change over time in years.

2. Black/White Wage Gaps in the USA

This dataset contains information on the median and average hourly wages in the USA, specifically between Black and White racial groups, between the years 1973-2022. This dataset examines wage gaps between racial groups, as well as men and women. Data is sourced from the Economic Policy Institute's (EPI) State of Working America Data Library. The context of

this dataset similarly examines median and average hourly wages in the USA, for only Black and White racial groups. This dataset is simpler than the previous one, and only examines racial groups and gender, but can give additional mean and medians to solidify correlations. This data also shows how data for these variables can change over time in years.

3. Wages by Education in the USA (1973-2022)

This dataset contains information on the average hourly wages of individuals in the USA based on their highest level of education, between the years of 1973-2022. Education attainment levels include less than high school, high school, some college, and bachelor's degree. This data also compares these education attainment levels between men and women. Data is sourced from the Economic Policy Institute's (EPI) State of Working America Data Library. The context of this dataset examines how higher levels of education can correlate with average hourly wages in the USA. Instead of focusing on racial differences, this dataset compares levels of education between men and women in the USA as a whole. This dataset can additionally translate how other factors, such as level of educational attainment can affect wages, in addition to race. It will be used to examine different factors and variables that play a role.

2.3 Methods

My datasets were all retrieved and extracted from Kaggle. To prepare my data, I loaded in my datasets and cleaned them to rid of column information that I would not be using in my analysis. By cleaning my datasets of unnecessary data, I can better relate all of my datasets using their similar variables and goals. Through out my code, I will be using multiple datasets and the new *plotly* library as my challenge goals. I merged two datasets based on their most similarity to enhance the emphasize of certain column variables.

For my first research question, I analyzed the impact of differing levels of educational attainment on average wage values. To do so, I plotted a line plot overtime displaying multiple lines based on educational attainment levels in the general USA population. Plotting a line plot over time will help me identify and trends or patterns overtime, describing any significant changes in growth.

My second research question uses the merged dataset to analyze if gender plays a role in influencing average wages. I use a line plot to display wage growth between men and women overtime, showing if data grows parallel and at what gap. This research question also further examines specifically if gender influences are present in racial differences. Comparing variables between White and Black individuals based on their gender will help me identify any significant differences in wage. This is displayed in bar charts, to easily identify significant gaps in value and the proportion of gender differences.

The last research question aims to address how racial differences in wage are amplified or influenced by education and gender, using the merged and education datasets. This research question targets an accumulation of the two prior research questions. Using *plotly* to graph a bar chart showing racial differences across all differing levels of educational attainment, it is easy to pinpoint significant differences in wage values.

2.4 Results

```
[ ]: !pip install -q folium mapclassify
!pip install --upgrade plotly
!pip install "jupyterlab>=3" "ipywidgets>=7.6"
!pip install jupyter-dash
```

```
[3]: import geopandas as gpd
import matplotlib.pyplot as plt
import pandas as pd
import folium
#Challenge Goal - New Library: plotly
import plotly.express as px
import plotly.subplots as sp
```

Load Data The following functions are used to create a merged dataset and load csv file datasets into data frames and assigned to variables.

```
[ ]: def merge_data(first, second):
    """This function takes two parameters, which are two different csv files,
    and merges them into
    one dataframe."""
    median = pd.read_csv(first)
    by_race = pd.read_csv(second)
    merge = median.merge(by_race, on=None, how='left')
    merge.set_index('year', inplace=True)
    return merge

merged = merge_data("median_average_wages.csv", "black_white_wage_gap.csv")
display(merged)
```

```
[5]: def load_data(csv):
    """This function takes a csv file as a parameter and loads it into a data
    frame variable."""
    data = pd.read_csv(csv)
    data.set_index('year', inplace=True)
    if (csv == "wages_by_education.csv"):
        drop_columns = ['hispanic_women_less_than_hs',
        'hispanic_women_high_school',
        'hispanic_women_some_college',
        'hispanic_women_bachelors_degree',
        'hispanic_women_advanced_degree']
        data.drop(drop_columns, inplace=True, axis=1)
    return data
#Challenge Goal - Multiple Datasets
education = load_data("wages_by_education.csv")
median_averages = load_data("median_average_wages.csv")
```

```
wage_gap = load_data("black_white_wage_gap.csv")
```

2.4.1 R2: Does a higher level of educational attainment contribute to average wages? Has average wages changed for levels of education overtime, and if so, is there a trend/pattern occurring?

For this research question, I will be focusing on using the dataset titled *Wages by Education in the USA (1973-2022)*. This research question aims to explore how differing levels of educational attainment plays a role in contributing to average wages. I hypothesize that higher levels of educational attainment correlates to higher averaging wages. Analyzing this dataset to answer this research question, I plotted a line plot using plotly to display average wages over time based on the minimum level of educational attainment.

```
[ ]: def plot_education(data):  
    """This function takes a dataframe as a parameter and returns a line plot  
    of corresponding average wages  
    based on minimum educational attainment levels over time."""  
    data = data.filter(items=['less_than_hs', 'high_school', 'some_college',  
    'bachelors_degree', 'advanced_degree'])  
    data.rename(columns = {'less_than_hs':'Less than highschool', 'high_school':  
    'Highschool',  
    'some_college':'Some college', 'bachelors_degree':  
    'Bachelors degree', 'advanced_degree':'Advanced degree'},  
    inplace=True)  
    fig = px.line(data, x=data.index, y=data.columns, title="Average Wage by  
    Educational Attainment")  
    fig.update_layout(  
        xaxis_title="Year",  
        yaxis_title="Average Wage",  
        legend_title="Minimum Educational Attainment")  
    fig.show()  
  
plot_education(education)
```

```
[ ]:
```

Between the years of 1973 to approximately 1982 there is a slight decline in average wage values before beginning a trend for representative educational attainment values. For attainment levels including less than highschool, highschool, and some college, the values stay relatively steady without any major increase or decrease in average values, with values in corresponding least to highest wage value between approximately 12-25 in USD. This numeric range is consistent between the years of 1973 up to most recently, 2022. Comparitively, there is a higher value trend for individuals with a bachelor's degree or advance degree. Although this has always been the case since the beginning of the observation time frame, the growth trend of average wages for a bachelor's and advanced degree follows a steeper growth. After 1985, wage values continuously increase, furthering the gap of wage values between a minimum bachelor's degree as compared to some college and below. The plot predicts that the variables that correlate with lower wage values,

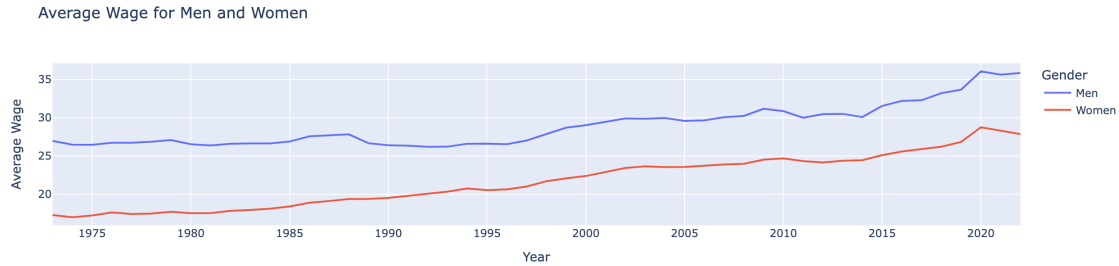
less than highschool, highschool, and some college, will continue at a steady trend, without any significant growth predicted. Whereas the trend line for a bachelor's and advanced degree suggests a further incline.

With this visualization, it can be supported that levels of educational attainment does have an influence on average wages. These results support my hypothesis that higher levels of educational attainment will result in higher average wage values. Overtime, average wages for lower educational attainment levels, such as less than highschool, highschool, and some college have not had any significant difference or shift in value. Average wages for bachelor's degree and advanced degree have consistently increased. Therefore, by receiving a higher level of educational attainment, an individual's wage will be higher than otherwise. As educational attainment levels rise, average wage values will also increase. Especially a bachelor's or advanced degree will grant a significantly higher average wage than the next level after, which is some college.

2.4.2 R2: Does gender affect wages and income between individuals? What is the significant difference, if any, in wages between Black and White men and women?

For this research question, I will be working with a merged dataset that I merged from previous code above. The two datasets merged were the median_average_wages.csv file and the black_white_wage_gap.csv file. This research question aims to examine whether gender plays a role in influencing wages between individuals, specifically racially different individuals. I hypothesize that gender plays a significant role in determining wage indifferences, specifically that male wages will be especially higher than that of women wages. To add to this hypothesis, I predict that average wages for black individuals will be notably less than white individuals. Analyzing this research question will help identify inequalities in the working force due to social stigma and stereotype, as well as injustices.

```
[ ]: def plot_line_gender(data):  
    """This function takes a dataframe as a parameter and returns a line plot of  
    corresponding average wages based on gender over time."""  
    data = data.filter(items=['men_average', 'women_average'])  
    data.rename(columns = {'men_average': 'Men', 'women_average': 'Women'},  
    inplace=True)  
    fig = px.line(data, x=data.index, y=data.columns, title="Average Wage for  
    Men and Women")  
    fig.update_layout(  
        xaxis_title="Year",  
        yaxis_title="Average Wage",  
        legend_title="Gender")  
    fig.show()  
  
plot_line_gender(merged)
```

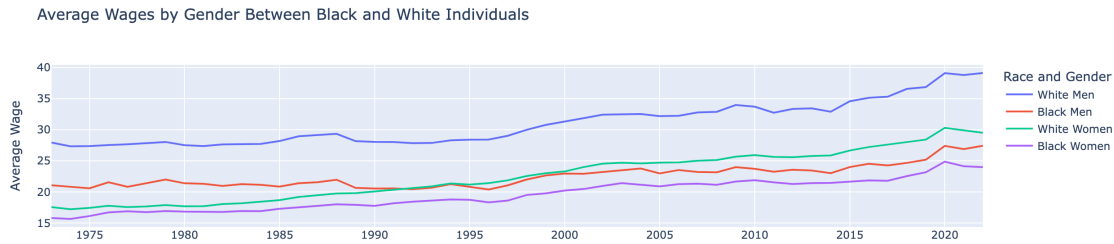


This plot takes the merged dataset to analyze how average wages have changed overtime based on gender. The earliest observations in 1973 show that there was a significantly large gap in average wage between men and women at an early year. The difference in 1973 is approximately 9-10 dollars between men and women, with the former gender having the higher average. As time progresses the gap continues to be visually present as the growth lines increase for both genders overtime. However, the gap does slowly close overtime, most notably at approximately 1994, where the average wage for women increases near the men's growth line before slightly dipping back lower, increasing the gap. Although the gap slightly decreases overtime, the average wage for women never reaches the men's, and the gap stays present despite an overall increase in average wage over time for both genders.

This visualization suggests that men consistently have higher average wages than women. It supports my hypothesis that gender does play a role in influencing average wages, specifically that men have higher average wages than women.

```
[ ]: #Create a line plot of for wages between genders of both races over time.
def plot_line_race_gender(data):
    """This function takes a dataframe as a parameter and returns a line plot
    of corresponding average
    wages for black and white individuals based on their gender over time."""
    data = data.filter(items=['white_men_average', 'black_men_average',
                              'white_women_average', 'black_women_average'])
    data.rename(columns= {'white_men_average': 'White Men', 'black_men_average':
    'Black Men',
                              'white_women_average': 'White Women',
    'black_women_average': 'Black Women'},
                inplace=True)
    fig = px.line(data, x=data.index, y=data.columns, title="Average Wages by
    Gender Between Black and White Individuals")
    fig.update_layout(
        xaxis_title="Year",
        yaxis_title="Average Wage",
        legend_title="Race and Gender")
    fig.show()

plot_line_race_gender(merged)
```



This visualization plots a line plot to show average wages by gender between black and white individuals overtime. This visualization is a more complex version of the one above, as it includes additional variables to observe racial influence. At a glance, it can be seen that at the earliest observations in 1973, both White and Black men have higher average wages than that of White and Black women. This supports the results from the visualization prior that men have higher average wages than women. Between the years of 1973 to approximately 1987, both White and Black women have a similar wage, with White women having a slightly higher wage value. This changes afterwards as the gap increases, and White women begin to consistently broaden the gap between them and Black women, increasing their wages at a higher difference from Black women. For men, overtime, White men have always had a significantly higher average wage value than that of Black men, with an almost approximately 10 dollar difference. This continues overtime, but the trend line of White men grows at a slightly faster pace per year, and by the end of the dataset timeframe of 2022, there is a larger gap between the two variable observations.

At the year 1992, the gender trends change as the average wage value of White women passes the line for Black men. As they grow similarly between 1991-2000, the White women trend line starts increasing at a faster pace, and a gap between the two occurs. Overtime up to 2022, White women have consistently had a higher average wage value than that of Black women. Although this does not entirely support the results from the first plot, it actually supports my earlier hypothesis that race will also play a role in influencing wage differences. White men and women earn higher average wages than that of both Black men and women. It also continues to support the hypothesis that gender has an impact because although all White individuals earn higher wages, White women do not earn as much as White men, who leads as the most dominant group to earn the highest average wage.

This visualization supports my hypothesis that both gender and race have a strong influence on average wages between White and Black populations. It suggest that men earn higher wages than women, but is also triumphed by how White individuals ultimately earn higher average wages than Black individuals.

```
[ ]: def plot_bars(data):
    """This function takes a dataframe as a parameter and returns a bar plot_
    ↪displaying the all time
    average of avergae wages for individuals based on their race and gender."""
    white_men_avg = data['white_men_average'].mean()
    black_men_avg = data['black_men_average'].mean()

    white_women_avg = data['white_women_average'].mean()
    black_women_avg = data['black_women_average'].mean()
```



```

datanew = {
    'Race/Gender': ['White Men', 'Black Men', 'White Women', 'Black Women'],
    'Average Wage': [white_men_avg, black_men_avg, white_women_avg,
↪black_women_avg]}
avg_data = pd.DataFrame(datanew)

fig = sp.make_subplots(rows=1, cols=2, subplot_titles=['Men', 'Women'])

fig.add_trace(
    px.bar(avg_data.loc[avg_data['Race/Gender'] == 'White Men'], y='Average_
↪Wage', x='Race/Gender',
            color_discrete_sequence=['orange'], barmode='group').data[0],
    row=1, col=1)
fig.add_trace(
    px.bar(avg_data.loc[avg_data['Race/Gender'] == 'Black Men'], y='Average_
↪Wage', x='Race/Gender',
            color_discrete_sequence=['blue'], barmode='group').data[0],
    row=1, col=1)

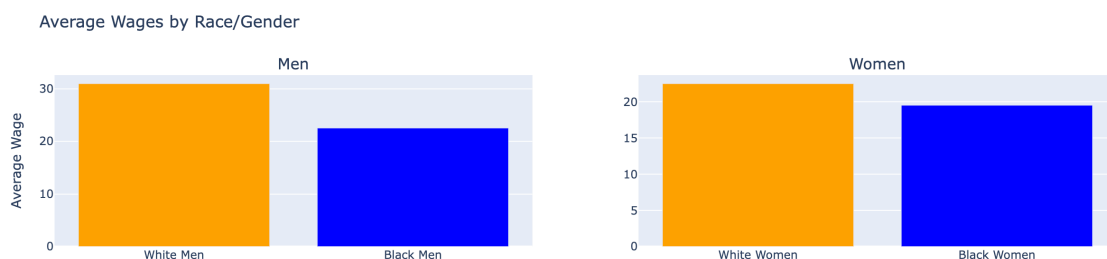
fig.add_trace(
    px.bar(avg_data.loc[avg_data['Race/Gender'] == 'White Women'],
↪y='Average Wage', x='Race/Gender',
            color_discrete_sequence=['orange'], barmode='group').data[0],
    row=1, col=2)
fig.add_trace(
    px.bar(avg_data.loc[avg_data['Race/Gender'] == 'Black Women'],
↪y='Average Wage', x='Race/Gender',
            color_discrete_sequence=['blue'], barmode='group').data[0],
    row=1, col=2)

fig.update_layout(
    title_text='Average Wages by Race/Gender',
    xaxis_title='',
    yaxis_title='Average Wage',)

fig.show()

plotBars(merged)

```



This plot aims to simplify the difference of average wage between Black and White men and women as identified in the prior plots. The two plots are divided based on gender, and further analyzed by race. These values have been calculated as an overall average for alltime, across the time period of 1973 to 2022. Wage values are rounded to the nearest one hundreth. The bar plot on the left shows that the average wage for White men is 30.96, while Black men have an average wage of 22.52. The difference between the two values is approximately 8.44. On the right, the bar plot shows that the average wage for White women is 22.53, while Black women have an average wage of 19.54.

This visualization further supports the hypothesis, as it specifies the values and differences, showing how wages differ between genders, specifically highlighting racial genders, despite gender. Comparing the two bar plots together shows how the average value of wage for White women even surpasses that of Black women, supporting the hypothesis that race plays an influential role in determining race.

2.4.3 R3: Is there a significant difference in wages between Black and White individuals in the USA? How do racial difference play a part in income differences in addition to levels of education and gender?

This research question uses the *Wages by Education in the USA (1973-2022)* dataset. This research question focuses on average wages based on the level of educational attainment in individuals. This research question differs from the previous research questions by aiming to find racial disparities. It takes analysis from the prior research questions to determine factors that are define racial inequalities in wage and income within the USA. I hypothesize that across all levels of educational attainment, the average wage for White individuals will consistently be higher than the average wage for Black individuals.

```
[ ]: def plot_all(data):
    data_filtered = data.filter(items=[
        'white_less_than_hs', 'black_less_than_hs', 'white_high_school',
        ↪ 'black_high_school',
        'white_some_college', 'black_some_college', 'white_bachelors_degree',
        ↪ 'black_bachelors_degree',
        'white_advanced_degree', 'black_advanced_degree'])

    education_levels = [
        'Less than High School', 'High School', 'Some College',
        "Bachelor's Degree", 'Advanced Degree']
    white_avg_wages = []
    black_avg_wages = []

    white_columns = data_filtered.filter(like='white_').columns
    black_columns = data_filtered.filter(like='black_').columns

    for col in white_columns:
        white_avg_wages.append(data_filtered[col].mean())

    for col in black_columns:
        black_avg_wages.append(data_filtered[col].mean())
```

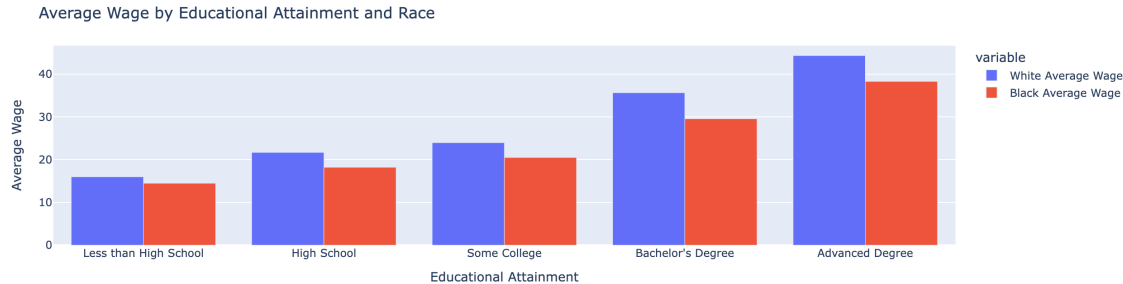
```

data = {
    'Education Level': education_levels,
    'White Average Wage': white_avg_wages,
    'Black Average Wage': black_avg_wages}

fig = px.bar(data, x='Education Level', y=['White Average Wage', 'Black_
Average Wage'], barmode='group')
fig.update_layout(
    title='Average Wage by Educational Attainment and Race',
    xaxis_title='Educational Attainment',
    yaxis_title='Average Wage',)
fig.show()

plot_all(education)

```



This bar plot displays multiple comparisons between the average wage of White and Black individuals based on the increasing level of educational attainment. There is a clear correlation between average wage and the level of educational attainment. As supported in the first research question, but also this visual plot, higher levels of educational attainment increase the value of average wage. As this relationship was established and supported by our analysis in the first research question, this plot addresses how the inclusion of a determining factor like race influences these averages. For White individuals, the lowest value of average wage at an education level of less than high school is approximately 16, ranging up to 44.35 for the highest level of education. On the other hand, Black individuals have the lowest average wage value of 14.48 and the highest value of 38.30 for respective lowest and highest levels of education.

At every stage of educational attainment level, White individuals have higher average wages as compared to Black individuals. The growth of average wages for White individuals at progressive stages of education level also progresses at a faster growth rate, with a 177.197 percent increase. The growth rate of average wages for Black individuals is lower, at a 164.563 percent growth rate. In addition to our earlier research analysis that identified a correlation between educational attainment level and average wage, this research question has developed a further relationship.

The results from this visualization support my hypothesis that the average wage for White individuals will consistently be higher than the average wage for Black individuals. There is a relationship between race and educational attainment in influencing average wages.

2.4.4 Result Summary

The results from analyzing data visualizations based on the information from datasets recording average wages in the USA have helped determine influential factors that impact different populations of individuals. It was observed that education has a role in increasing average wages. That is, higher levels of educational attainment directly correlate with higher wages. This can be that higher levels of education prepare individuals with stronger skillsets and experiences that allow them to be better equipped for a job, and makes a stronger career candidate. Additionally, we determined that there are clear differences in wages based on gender. Men are found to earn significantly more than women. It was also observed that in addition to the gender differences, White men and women earn more than Black men and women. An accumulation of these research questions and results also found that white individuals consistently earned higher average wages than Black individuals in every level of educational attainment.

White individuals have an advantage in increased average wages, regardless of their educational attainment or gender as compared to Black individuals. Black individuals have a larger struggle in receiving higher average wages as compared to White individuals. This difference establishes a clear instability between racial groups, and should be addressed as a social injustice and inequality.

2.5 Implications and Limitations

The purpose of this project and the research questions performed is to determine and address racial and gender inequalities in the work force in the USA. By determining if minority groups are disadvantaged, despite their gender or educational level, social solutions can be made to bridge the gap to better equip individuals for more economically and financially stable lives and careers. The results of this analysis can help advocate for gender and racial inequalities, emphasizing the importance of these imbalances to help improve labor and education inequalities. This data and further in regards to this topic can be used to inform and formulate policies and actions made towards closing wage gaps. It will also help individuals pursue higher levels of education to achieve higher outputs in their career aspirations.

There are several limitations to this research analysis. Although this project aims to address racial inequalities, it only focuses on data for Black and White individuals. Excluding data for other racial or ethnic groups makes it difficult to examine all racial inequalities and different severities. Many different racial or ethnic groups face different social issues and have different dynamics. Another limitation is that by grouping all advanced degrees together as a level of educational attainment, it obscures the potential differences in income for higher degree types. This includes a master's, doctorate's, or even multiple types of degrees, as these could suggest an increase in or slow down the rate of increase for average wage values. The third limitation is that the information provided in the datasets used is very general, and does not account for different occupations or the amount of hours worked. These are factors that could impact wages for individuals, despite or including the variables that we analyzed; gender, race, and educational attainment. This project research analysis gives a broad overview to establish a significance between wages in different populations or individuals. Including more factors would allow for a more in depth and inclusive analysis to examine and predict further solutions.