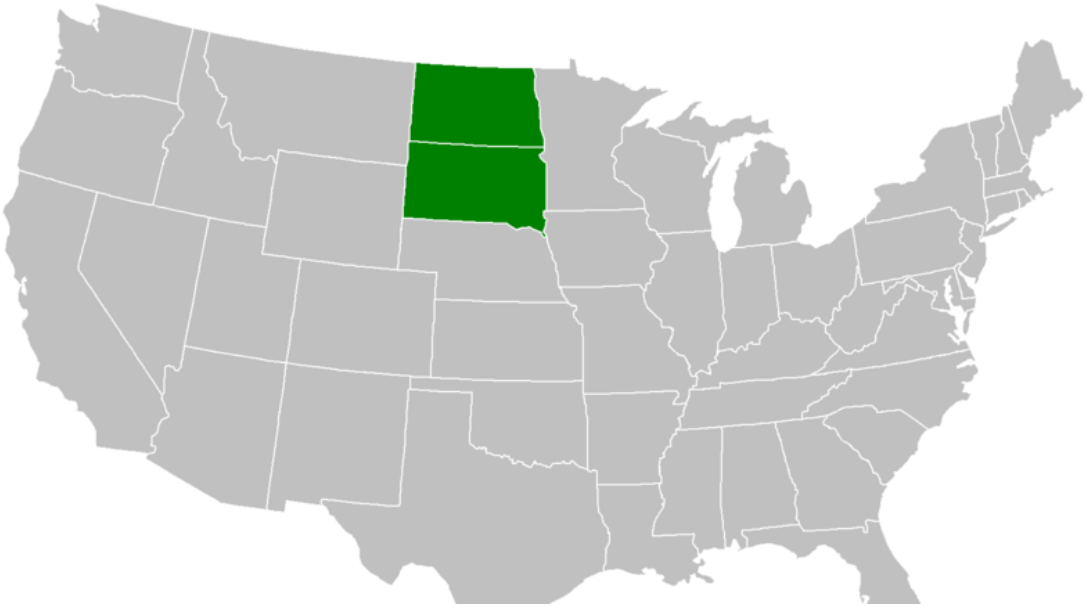


NORTH DAKOTA AND SOUTH DAKOTA'S EDUCATION AND EMPLOYMENT

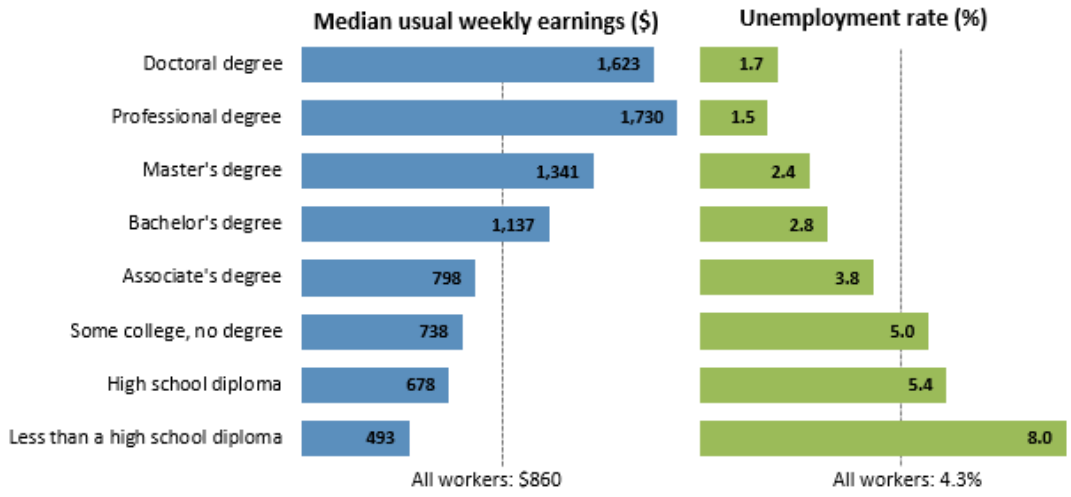
Is education having an impact on employment in the two states?

Group member: Zhekai Jiang, Jie Xu, Zijin Yu

INTRODUCTION

Problem statement and background

We are interested in the relationship between unemployment and education levels. According to data from the *U.S. Bureau of Labor Statistics* (BLS), earnings increase and unemployment decreases as educational attainment rises. Grouping workers by education level, the chart below shows that, those with more education have higher earnings and lower rates of unemployment than those with less education.[1]



BLS's data tells the total trend, however, as state governments are the exact policy executive parties, the situation may vary among different states. North Dakota and South Dakota are two neighbouring states in the middle of the country. The regional history can be traced back to 1861. President James Buchanan signed the bill creating the Dakota Territory, which originally included the area covered today by both Dakotas as well as Montana and Wyoming. Beginning about 1877, efforts were made to bring Dakota into the Union as both a single state and as two states. The latter was successful and on November 2, 1889, both North and South Dakota were admitted.[2] However, as time pass by, the two 'Dakota's went into

[1] U.S. Bureau of Labour Statistics, https://www.bls.gov/careeroutlook/2016/data-on-display/education-matters.htm.
[2] State of North Dakota, https://www.nd.gov/state-facts/origin-name.

different ways: North Dakota is built on oil and gas production while South Dakota mainly relies on farming. The economic development bears different societies, which can be an interesting background for social study.

In this report, we want to focus on the relationship between education and employment within the Dakota States. These are the data-driven questions we intend to answer:

1. Are there any variances in education level between males and females in different race categories?
2. What is the relationship between unemployment and education level?
3. Did employment rates and average levels of education vary among the two states?
4. Did the education level of citizens change from one time period to another?
5. Is there any correlation between weekly earnings and years of education?

We would try to dig into the relationship of education and unemployment with quantitative tools, and explore if there are any lurking variable, so that we can draw the conclusion on the relationship and provide our idea of policy improvement direction.

Dataset

The original idea is to collect CPS data across the United States; however, we tried many ways to reduce the size of the official file so that we could use it on *google colab* and all of them didn't work. Therefore, we have to focus on part of the data set and we choose South Dakota and North Dakota, as illustrated above, to study the relationship between unemployment, education level, race and gender. For better visualization, according to the codebook, we divided the specific education time from the raw data into four categories

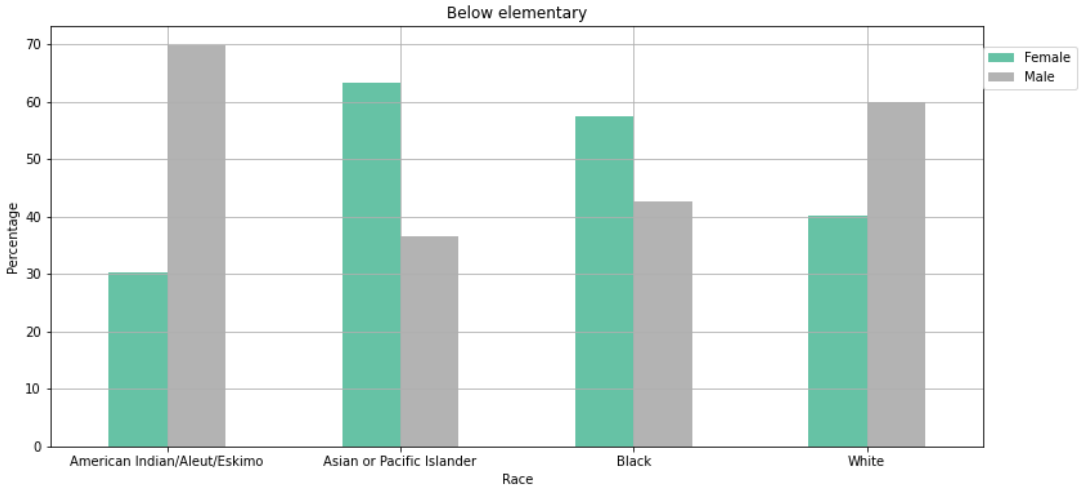
which are *Above high school below bachelor*, *Above bachelor*, *Below high school*, and *Below the primary school*, and divided the race into five categories which are *White*, *Black*, *Other*, *American Indian/Aleut/Eskimo* and *Asian or Pacific Islander*.

ANALYSIS

Task 1: Are there any variances in education level between males and females in different race categories?

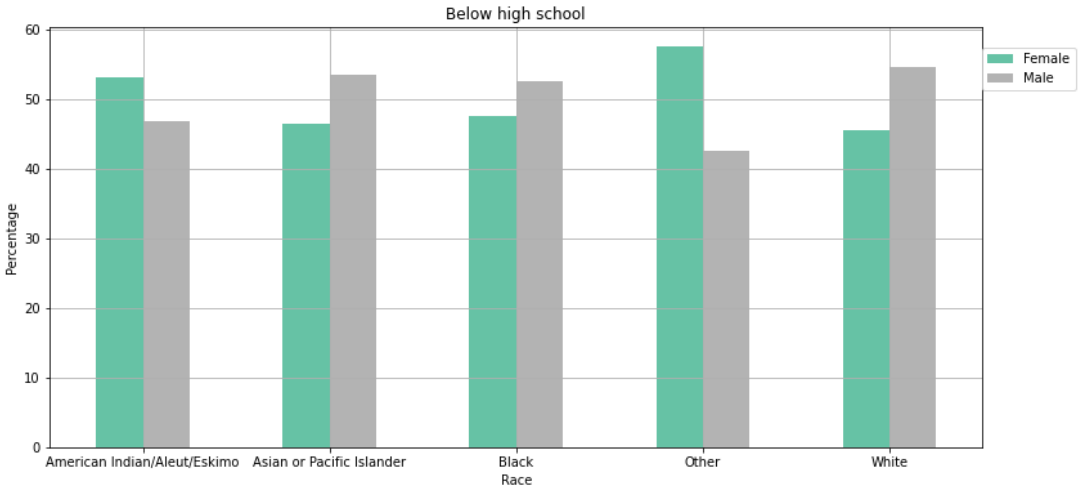
We first examined the differences between men and women of different races at the same educational level.

Figure 1.1: The difference of Sex among Races at Below Elementary



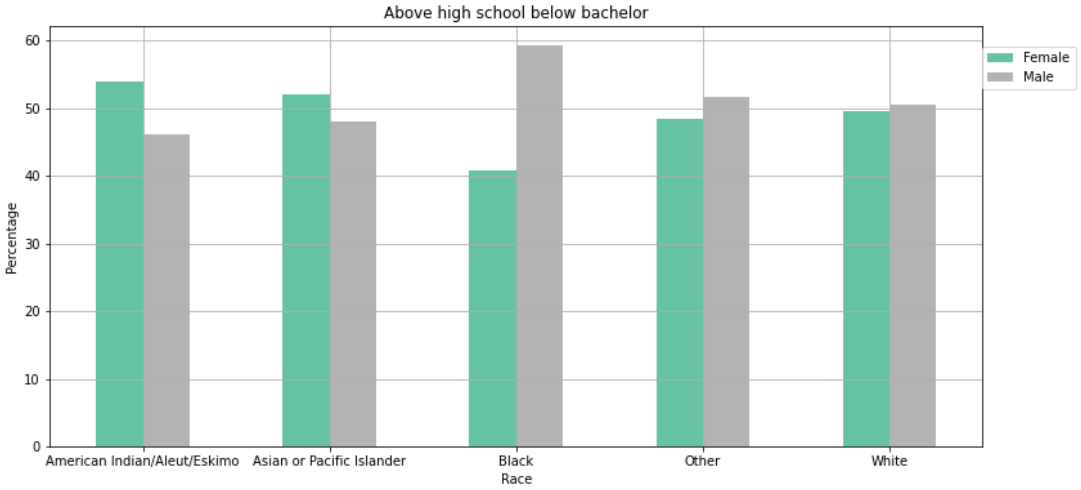
Due to the lack of valid data, we do not have the type of "Other" race in this table. We find that the number of males of American Indian/Aleut/Eskimo and White is much higher than that of females at a Below elementary level. The difference between American Indian men and women is even nearly 40 percent. On the contrary, the number of males of Black and Asian or Pacific Islander is much smaller than that of females at a Below elementary level.

Figure 1.2: The difference of Sex among Races at Below High School



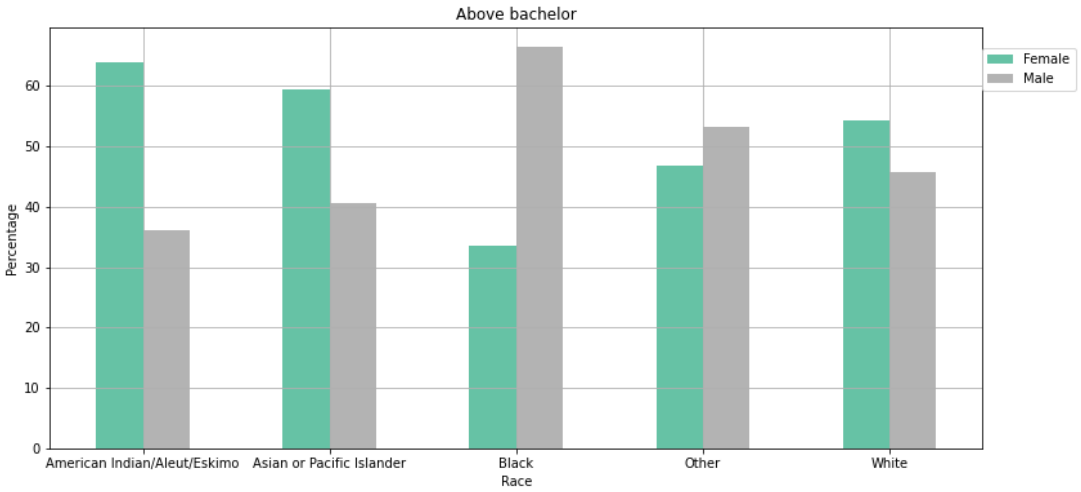
We find that the number of males of Black and Asian or Pacific Islander and White is much higher than that of women at a Below high school level. On the contrary, the number of males of American Indian/Aleut/Eskimo is much smaller than that of women at a Below high school level.

Figure 1.3: The difference of Sex among Races at Below Bachelor



We find that at the above high school below bachelor level, the proportion of men and women is similar across all races except Black people. More Black males have this educational level than females.

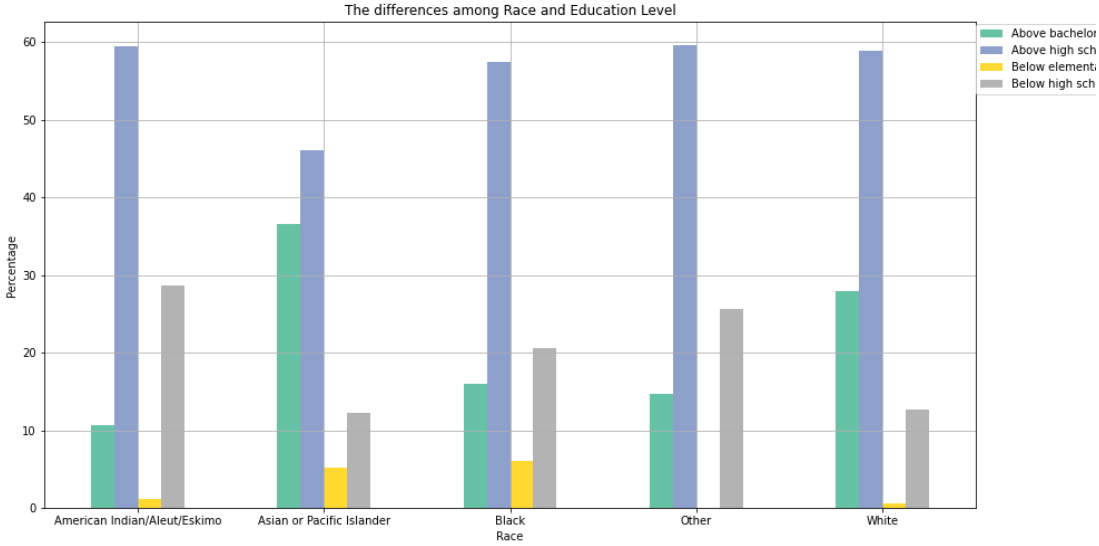
Figure 1.4: The difference of Sex among Races at Above Bachelor



We find that as with a bachelor level, the number of black males with high education is far greater than that of females. However, the number of males of American Indian/Aleut/Eskimo and Asian or Pacific Islander are much smaller than that of females while there no obvious difference between White and Other race.

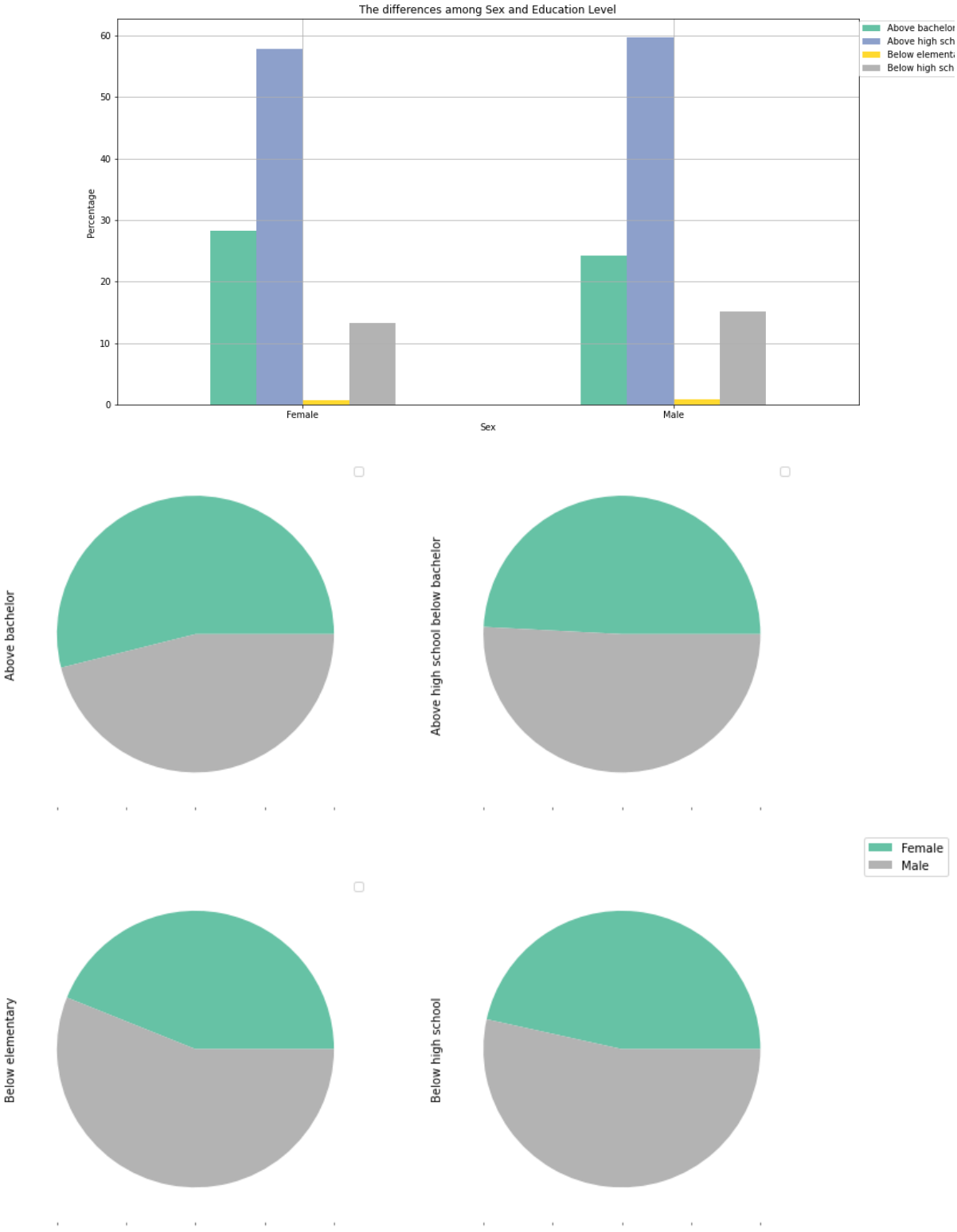
Then we focus on two specific variables.

Figure 2: The difference Among Race and Education Levels



Focusing on *Race* and *Education Levels*, we find that, regardless of race, the education level of the vast majority of people is at above high school below bachelor level and the average education level of White and Asian or Pacific Islander will be higher. It is worth noting that, the proportion of Asians at higher education levels is even higher than that of Whites. What's more, the proportion of Asians at bottom education level is also significant compared to other races and Black is even higher than Asians. And we run a Chi-Square Test between race and education level which shows there is a relationship between race and education level and p value is smaller than 0.0000.

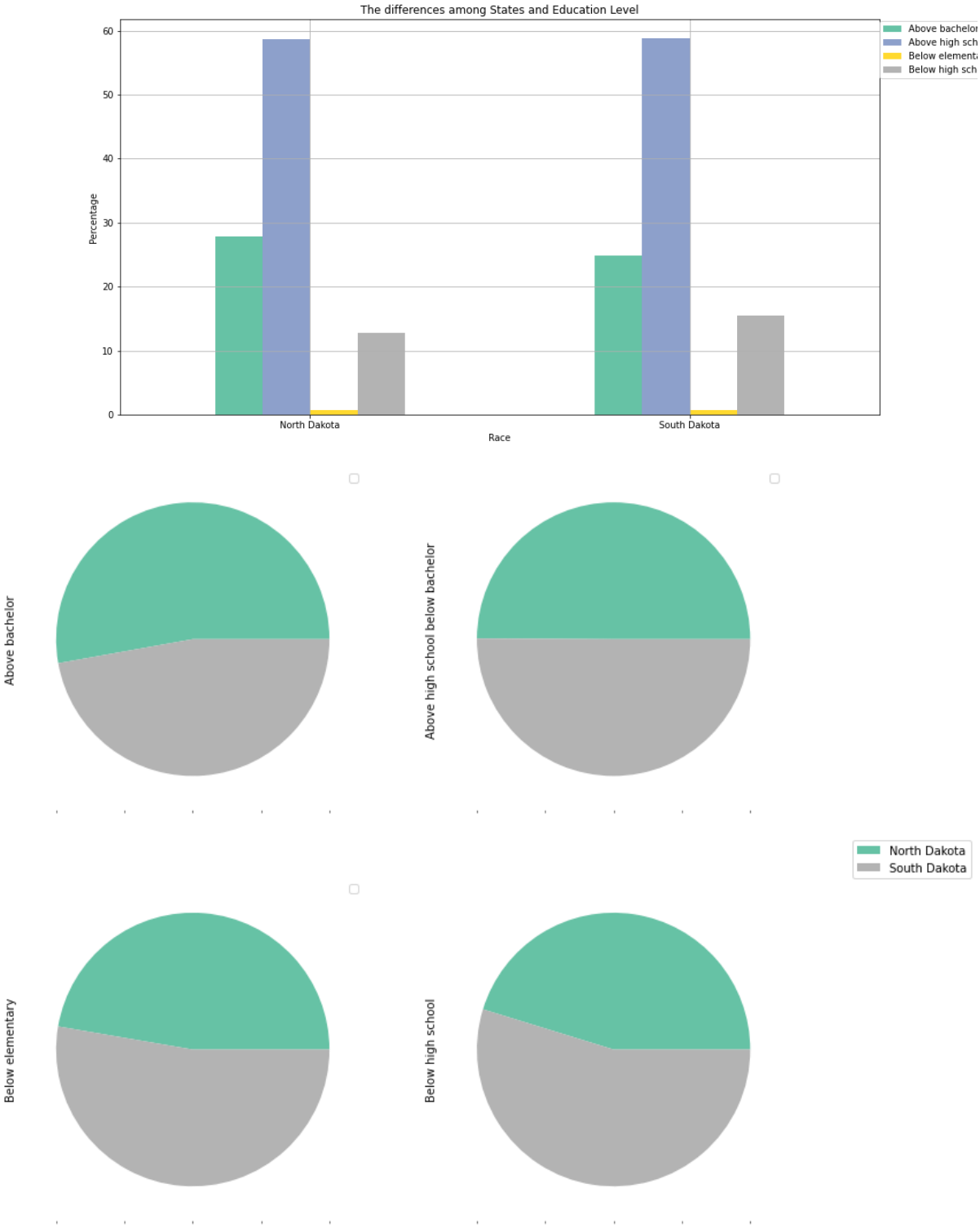
Figure 3: The difference Among Sex and Education Levels



Focusing on *Sex* and *Education Levels*, we find that, at the overall level, there is no obvious difference between male and female in terms of education level. The proportion of female at higher education level is slightly higher than that of male while the proportion of male at

lower education level is slightly higher than that of female. However, there still shows a relationship between sex and education level and p value is smaller than 0.0000.

Figure 4: The difference Among States and Education Levels

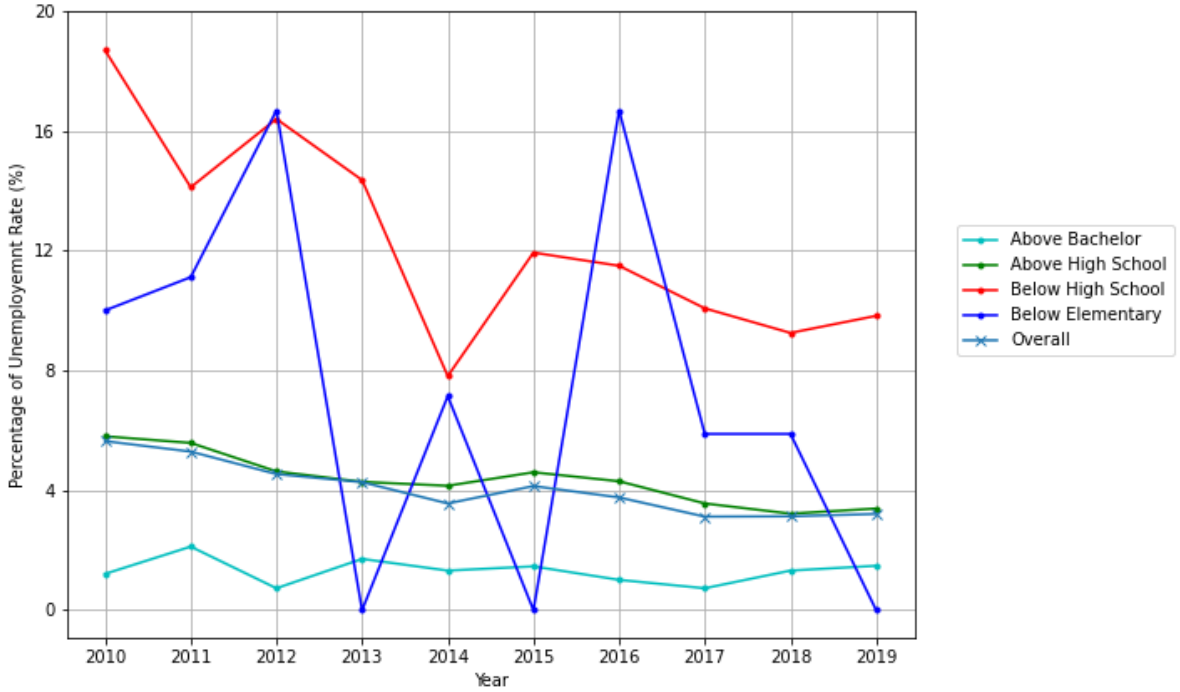


Focusing on *States* and *Education Levels*, we find that, at the overall level, there is no obvious difference between North Dakota and South Dakota in terms of education level. The proportion of North Dakota at higher education level is slightly higher than that of South Dakota while the proportion of South Dakota at lower education level is slightly higher than that of North Dakota. However, we also find a relationship between states and education level and p value is smaller than 0.0000.

Task 2: What is the relationship between unemployment and education level?

The variable "EMPSTAT" in the original CPS dataset indicates whether persons were part of the labour force and whether they were currently unemployed. The variable also provides information on the activity or status of persons in or not in the labour force. However, for research purpose, we simply categorize the variable into "Employed" and "Unemployed", two status of employment to be studied.

Figure 5: The Relationship between Unemployment Rate and Different Education Levels



The figure above shows both the overall and categorical percentages of unemployment rates. Four lines with dots shows the respective percentages of unemployment rates among different levels of education, while the line with crosses shows the overall percentage of unemployment rate. In spite of the particular values “0” within “Below Elementary”, people educated below elementary and below high school have a much higher unemployment rate than the overall level. The line representing “Above High School” education level shows a similar trend with the overall rate, and people educated “Above Bachelor” keeps a rather low unemployment rate. We can read through the result and see a positive impact of education on employment.

Therefore, we would like to further explore the correlation between unemployment and education level with the tool of *Chi-square Test*s.

Table 1: Chi-square Test Result of Unemployment Rate and Different Education Levels

Year	**χ^2**	**Reject H_0**
2010	113.071	✓
2011	55.380	✓
2012	108.761	✓
2013	NaN	✗
2014	24.649	✓
2015	NaN	✗
2016	42.827	✓
2017	37.966	✓
2018	29.227	✓
2019	NaN	✗

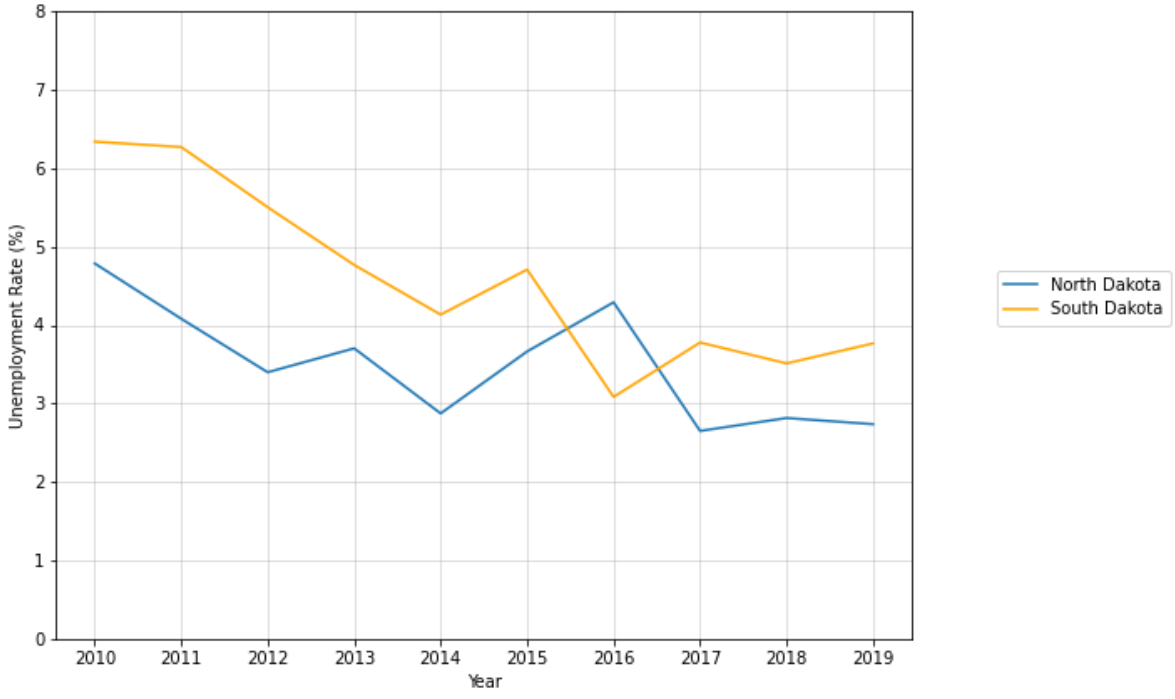
According to the result of the *Chi-square Test* above, 7 out of 10 years calculates rather small p-value (p-value<0.01) and therefore provide convincing evidence to reject the null hypothesis, while there is missing values in data of other three years (year 2013, 2015 and 2019) and fail to proceed the test. Considering the results and the probability of rejection, we can draw the conclusion that there is a difference in the distributions of unemployment under different levels of education, which can be the evidence that **education does have an impact on unemployment**.

Task 3: Did employment rates and average levels of education vary among the two states?

To identify if states could be an influencing factor (lurking variable) in the relationship of unemployment and education, we conduct out research on the levels of employment and education among the two states.

To start with, we try to visualize the unemployment rates in the two states.

Figure 6: The Unemployment Rate of North Dakota and South Dakota



According to the figure above, we can read that North Dakota has been kept a lower unemployment rate during the past decade, except for year 2016. It can be conjectured that there is difference in unemployment at the level of states. However, the figure only provides us with a general concept on the correlation, and therefore we then conduct a *Chi-square Test* to verify the speculation.

Table 2: Chi-square Test Result of Different States and Unemployment Rate

Year	χ^2	**Reject H_0**
2010	3.013	✗
2011	6.428	✓
2012	6.497	✓
2013	1.659	✗
2014	2.767	✗
2015	1.387	✗
2016	1.785	✗
2017	1.827	✗
2018	0.644	✗
2019	1.552	✗

According to the result of the *Chi-square Test* above, 8 out of 10 years calculates rather big p-value (p-value>0.05) and therefore fail to reject the null hypothesis, while only two years (year 2011 and 2012) achieved a positive result but still with of little significance. Considering the results and the probability of rejection, no evidence shows that there is a difference in the distributions of unemployment between the two states. The result only shows that there is no linear relationship between the two variables, but there may exist other relationship which cannot be detected by *Chi-square Test*.

Table 3: Chi-square Test Result of Different States and Education Levels

Year	χ^2	Reject H_0
2010	9.910	✓
2011	4.220	✗
2012	25.933	✓
2013	15.000	✓
2014	11.384	✓
2015	10.524	✓
2016	8.445	✓
2017	10.719	✓
2018	12.928	✓
2019	8.676	✓

To excluded the variable 'state' as a lurking variable, we check the relationship between states and education levels through *Chi-square Test*. According to the result above, 9 out of 10 years calculates small p-value (p-value<0.05), even 4 calculating rather small p-value (p-value<0.01) and therefore provide evidence to reject the null hypothesis. Considering the results and their significance, we can reach the conclusion that there is a difference in the distributions of education levels between the two states. Although it does not matter our research, the conclusion indicates that average education levels are different at the level of states, which can be another interesting topic in further studies.

Task 4: Did the education level of citizens change from time to time?

Figure 7: The Unemployment Rate with Different Education Levels in Two States

We plot changes of percentage point in education levels across North Dakota and South Dakota. According to the graph, the majority of people in both states have above high school below bachelor education, about 45% while people with below elementary education are about 1% in the two states.

By contrast, though people above bachelor and people below high school represent the second largest and third largest population in both states, the gap between the two levels in North Dakota is getting bigger as the amount of people above bachelor is gradually increasing and the amount of people below high school decreases. However, the amount of people with above bachelor degree in South Dakota shows a faster increase than North Dakota, indicating more and more people in South Dakota pursue a higher education degree.

In summary, percentages of education levels from 2010 to 2019 in North and South Dakota have not changed too much overtime. Thus, we can take education as an independent variable within this time period.

Task 5: Is there any correlation between weekly earnings and years of education?

To figure out this question, we regressed weekly earnings on years of education in the two states. The output is presented as below.

Table 4: OLS Regression Results of Weekly Earnings on Years of Education in Two States

OLS Regression Results

Dep. Variable:	EARNWEEK	R-squared:	0.105
Model:	OLS	Adj. R-squared:	0.105
Method:	Least Squares	F-statistic:	451.0
Date:	Mon, 21 Dec 2020	Prob (F-statistic):	1.06e-94
Time:	02:52:41	Log-Likelihood:	-29515.
No. Observations:	3827	AIC:	5.903e+04
Df Residuals:	3825	BIC:	5.905e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-356.2550	54.243	-6.568	0.000	-462.602	-249.908
Years_of_education	79.2754	3.733	21.236	0.000	71.957	86.594

Omnibus:	1103.089	Durbin-Watson:	1.886
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3020.176
Skew:	1.528	Prob(JB):	0.00
Kurtosis:	6.099	Cond. No.	90.5

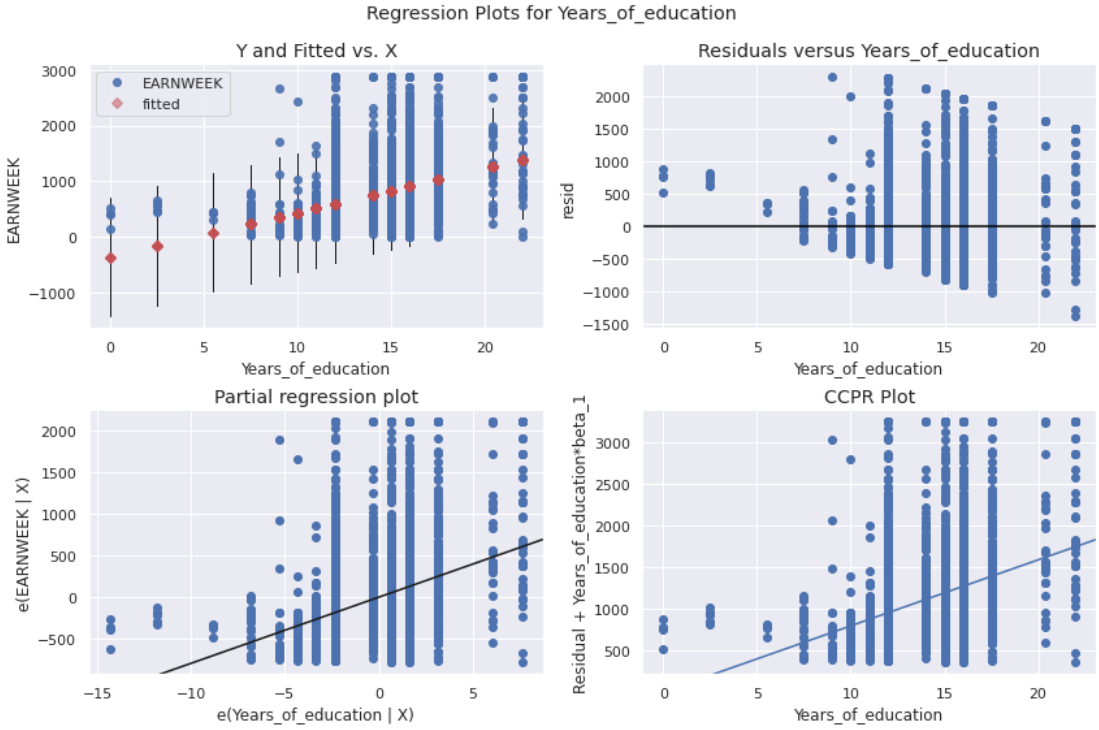
Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

According to the model summary, there is a non-robust positive relationship between years of education and weekly earnings with an R-squared of 0.105, which indicates about 10.5% of the variation in weekly earnings can be explained by the explanatory variable years of education. However, the relatively low R-squared value indicates that years of education is not completely predictive of weekly earnings. Further research should focus on confounding or lurking variables that might better explain weekly earnings than years of educations.

The estimate slope is 79.3 with a confidence interval from approximate 72.0 to 86.6, meaning that with one unit of increase in years of education, there is a 79.3 increase in weekly earnings. The p-value of 0.000 for the slope indicates that the effect of years of education is statistically significant under 0.05 significance level.

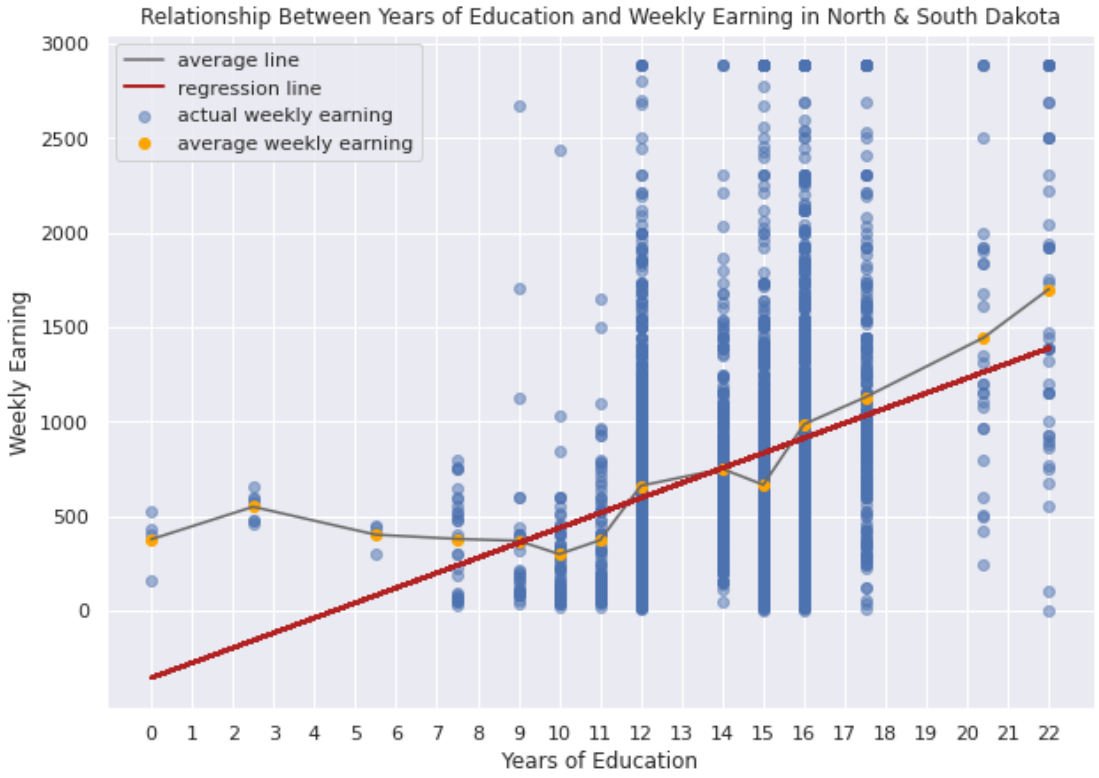
At the bottom of the summary, Durbin-Watson indicates tests for homoscedasticity, which means that the variance of errors is consistent across the entire data set. The value 1.886 is within (1, 2); thus, we do not have error rate grows in a particular direction. However, the condition number measuring the sensitivity of output compared to its input is 90.5, which is much higher than 30, our expected number. This means we have multicollinearity, and therefore much higher fluctuations to small changes in the data.

Figure 8: The Regression Plots for Years of Education



Then we produced four regression plots. The upper right one is residual vs. fitted plot, whose x-axis on shows the actual values of years of education and the y-axis shows residuals. We can thus argue that although there is some deviation of residuals with observed value greater than expected value, errors seem to randomly scattered across the data. The upper left plot visualizing fitted values versus years of education while also includes prediction intervals. The bottom left plot is partial regression plot, which we discuss later in a bigger picture. The bottom right plot helps us to judge the effect of one independent variable on the response variable by taking into account of the effects of the other independent variable; since we have only one independent variable, this plot looks the same as the partial regression plot.

Figure 9: The Relationship Between Years of Education and Weekly Earning in North & South Dakota



From this regression model, we can see a positive relationship between years of education and weekly earnings as consistent with the numeric summary. However, this effect of

education appears to be mitigated as people get at least 12 years of education, which means that they either graduated from high school or received 12 Grade education, have roughly similar opportunity to get higher earnings no matter how long they get educated in the future. On average, people received 0-7 years and 16-22 years of education have weekly earnings well above expected while people with 10, 11, and 15 years of education have below expected weekly earnings.

CONCLUSION

Results

We can conclude that education level has an impact on employment and higher educated people seem to find jobs easily. At the same time, there are also great differences in the educational level between different race groups, especially between white and black people. A large proportion of black people below the level of primary education. Moreover, Black men generally have higher education than women. Although the gap between men and women in each segment is not particularly large, there are still some differences. Education levels also vary between the two states and the proportion has not changed much over the past decade. Also, no evidence shows that there is a difference in the distributions of unemployment between two states. What's more, higher education level represents higher income to a certain extent

Discussion

Since the proportion of black people below the level of primary education is very large, we can infer that these people may not have access to education, which involves the issue of social equity. This study can be further extended and deepened as the focus of relevant policies. Moreover, the issue of gender equality seems to be particularly serious among black people in the two states we are concerned about. We are confident to put this conclusion on the whole country scale so that the policy of gender equality for black people is imminent. And the existing education policies across two states do not seem to improve the general education level of people. Base on the conclusion that higher education level people are more likely have higher income, the existing education policies are no doubt unsuccessful. It is worth mentioning that, although there is no evidence of differences in the distribution of unemployment rates between the two states, there are still two years that show a relationship between the two variables when we subdivide the data into years. We suspect there may be other factors that we have not observed.

Limitation

Since our data are only limited to two states and one decade from 2010-2019, whether our conclusion can be accurately extended to the whole country remains to be verified. Moreover, since many data are collected by family, some younger people may still be in the school or have not started receiving education. To a certain extent, it would lead to a statistical deviation (average education level will decline in a whole range). What's more, we only

roughly classify race and education level, which may not reflect the difference very precisely. Also, as for the relationship between education level and income, our conclusion may not be representative due to the lack of most data, therefore, further investigation and research are needed. And the specific situation of unemployment might change according to different jobs, here we only focus on the condition of unemployment and not studied this area in detail.