Analysis of Age, Race, and Income as Factors of Work Time

Audrey Kim, Lisa Li, Joy Mahoney, Grace Rade

12/16/2021

Abstract

The purpose of our analysis is to examine the relationship that age and race, split by income group, have on the number of hours per day spent on work and related activities. Using Annual American Time Use Survey data from the IPUMS database, spanning the years 2015 to 2020, we fit two multiple regression models with work hours per day as the response. Our models found that for all income levels, work hours do not decrease as the age of the working increases. Additionally, the race of the worker does not change the relationship between age and work hours for all income levels

Introduction

Neoliberal societies expect workers to commodify themselves and essentially sell their time and bodies to support themselves. The structure of the 40 hour work week demands that work becomes the utmost priority. As a result, one's work becomes an identifier, proving the old adage "you are what you do" correct. Studies show that Americans, on average, work 34.4 hours a week, longer than their counterparts in other economically developed nations. But why do Americans work so much? Examining how much time Americans spend working, on average, each day can inform policy decisions and proposals for modifications of the work week, wage and salary adjustments, and help workers advocate for their needs. Rather than broadly studying how much time Americans spent working per day, we investigated how that number changes based on age, family income, and race. Do Americans work fewer hours each day as they age? Does race influence how much people work over time? Does a higher total family income allow Americans to work less? Using such explanatory variables allowed us to see if different racial groups or ages face an undue work burden in order to make a certain amount of money. This project tests two hypotheses. The primary hypothesis claims that work hours will decrease with age, regardless of income level. The secondary hypothesis modifies the primary hypothesis, claiming that as age increases, the extent of the decrease in work hours will depend on the race of the respondent/worker, also regardless of income group.

Methods, Variables, and Model Choice

The dataset used in this project was synthesized from the IPUMS TimeUse database, which aggregates data about American Time Use, obtained from the Current Population Survey. The Current Population Survey is a survey sent to all households in the US that have at least 1 member who is at least 16 years old, resides in the 50 US states, and is not a part of any inmate institution. The data in the Time Use section was gathered from the federally administered Annual American Time Use Survey [https://www.atusdata.org/atus/]. The respondents are selected using a three-step process. From eligible households filling out the Current Population Survey, some in small states are removed from the pool since the CPS over-samples households in small states. Then, the remaining eligible households–which should now be more representative of the US–are stratified based on the race/ethnicity of the householder, the presence and age of children, and the number of adults in adult-only households. Lastly, a randomly selected individual (16+ years old) from each household is asked to track one day's activities.

For our models, we included age measured in years, income measured as salary in United States Dollars, and race. We chose the age variable because we had reason to believe that as one gets older, the amount of time dedicated to work may change over time. We included the income variable because we wanted to see if there was a negative association between the income of a respondent and the amount of time spent working. We also chose to include race because there have been studies done that show a significant pay gap between different races. We removed 48,000 missing observations and simplified the Race category to condense from over thirty categories to six. This was a difficult decision for us because it was important that our data was representative of the population while also keeping the number of coefficients in the model manageable. Additionally, we removed any values for ACT_WORK (and, by extension, ACT_WORK_hours) that equaled zero, to limit analysis to respondents currently working.

We believe that the parallel slopes model best addresses our first hypothesis; it describes the relationship between work, income and age well because it has the ability to show that age has a similar impact across all income levels. We selected an additive interaction model to address our secondary hypothesis. Having the interaction between the age and race of the respondent provides a way to quantify the extent of the effect that the race has on the effect of age on estimating time spent working. Both of the models highly violated the normality of variance, for all categories in each hypothesis, and the additive parallel slopes model satisfies the equality of variance and the linearity conditions more than the additive interaction model does.

Results

Please see data appenxid for model plots

Running a parallel slopes multiple linear regression with age and income explaining time working produces a model with an intercept of 7.245 and a slope on AGE of 0.001. These coefficients indicate that based on this data, we would expect a zero year old person with a family income of less than $10,000 to work 7.245 hours per day. Additionally, we would expect that, regardless of family income, the number of working hours per day would increase by 0.001 for each additional year of a person's age, on average. The intercept coefficient is statistically significant with p-value of nearly zero, indicating a very small probability of observing a coefficient of this magnitude or more extreme if the actual value was zero. The slope, however, has a p-value of 0.811, suggesting that there may not be a relationship between age and working hours. This is contrary to our hypothesis that there would be a negative relationship.

Our secondary hypothesis is that the extent of the decrease in work hours will depend on the race of the respondent, in addition to age and income. The explanatory variables will be AGE (numeric), FAMINCOME (categorical), and RACE (categorical). Even though we found through our first hypothesis' investigation that FAMINCOME does not explain variability in work hours beyond what AGE explains, we will keep FAMINCOME to stay consistent with our hypothesis. In order to test if the slope changes based on race, we ran a nested F-test to see if slope depends on race, using a reduced model with age, family income, and race as explanatory variables (all additive) and a full model with age, family income, and race as explanatory variables, with age and race interacting. The nested F-test returns a p-value of 0.2184, suggesting the slope of work hours on age dependent on race does not change the effectiveness of the model and that the slope on work hours are not dependent on the race of the respondent. This result provides evidence against our secondary hypothesis, which stated that race would affect the change in work hours.

Discussion

This project sought to investigate how/if time spent working varies by demographic group. We looked at age, family income, and race to predict the number of hours worked per person. In our primary hypothesis, we predicted that as people aged, they would work less and that this trend would hold up regardless of income level. Our results, however, provide evidence that there is no relationship between age and time worked. More specifically, our analysis determined that average daily work hours does not significantly change as the age of the worker increases.The slope of our model was very slightly positive, with a corresponding p-value that indicates that the slope is not statistically different from zero. These results may indeed be the case, for a number of reasons. For instance, employer-employee contracts may not be flexible to the whim of the employee, so the typical person may only have the option of working certain hours or none at all, especially as it aligns with the general nationwide standard for what constitutes a full or part time job in terms of expected work hours. The results may also provide evidence against a relationship due to some model limitations. In this case, we restricted the model to a linear shape when reality may demonstrate different working trends based on age group (for instance, increasing working hours until a certain age and then decreasing). A piecewise model that allows for different slopes based on age or an inverted quadratic may be more appropriate.

In our secondary hypothesis, we predicted that the relationship between age and amount of work time is impacted by race, when controlled for family income. Our result showed that race does not affect such relationships. However, we further demonstrated that when family income is not taken into account, race actually affects the relationship between age and work time. This result indicates that family income could be related to race, and by including family income to predict work time, it was not necessary to consider the effect of race. Thus, our model could be improved by accounting for the relationships between additional predictors and determining whether family income or race is more significant to predict the work time, in addition to age. Additionally, as suggested above, the relationship between age and work time might not be linear, which also accounts for the rather low significance of race, as the model with age and family income has less statistical power to begin with. Thus, a more accurate, non-linear model accounting for the relationship between age and work time is needed.

There are also a number of limitations to this study due to the data we used. For instance, the study was not comprehensive in the number of categories in each variable. Due to the large number of IPUMS classifications for categorical variables, we collapsed categories within family income and race and thus could not study intra-group differences. Also, using hours per day as a measurement of work time may not have been the most appropriate measure since it doesn't account for differences in the number of days worked per week. The data also does not distinguish between salaried and non-salaried workers, which could be an underlying factor determining how many hours per day a person works.

Overall, these investigations provide no evidence that age, income, and race are associated with time spent working when forming the multiple linear regressions as outlined in this report. However, this is not to say that there is no relationship between any or all of these variables: for instance, a different model may be needed. Next steps would include testing non-linear models and determining what (if any) data may be excluded as outliers. This study investigated broad demographic mappings, so it may be interesting to develop a more focused question inspired by the models here: for example, "Are pay gaps between genders and race reflected in time spent working?" While there were no significant findings, this study provides a jumping off point for future research and serves as an important beginning into the question of "Where are people spending their time, and why?"

Refrences

Brown, W. (2017). Undoing the Demos: Neoliberalism's Stealth Revolution. Zone Books.

Isidore, C., & Luhby, T. (2015, July 9). Turns out Americans work really hard... but some want to work harder. CNN Business. Retrieved December 13, 2021, from https://money.cnn.com/2015/07/09/news/economy/americans-work-bush/

Kenton, W. (2020, October 26). Neoliberalism Definition - Economy. Investopedia. Retrieved December 13, 2021, from https://www.investopedia.com/terms/n/neoliberalism.asp

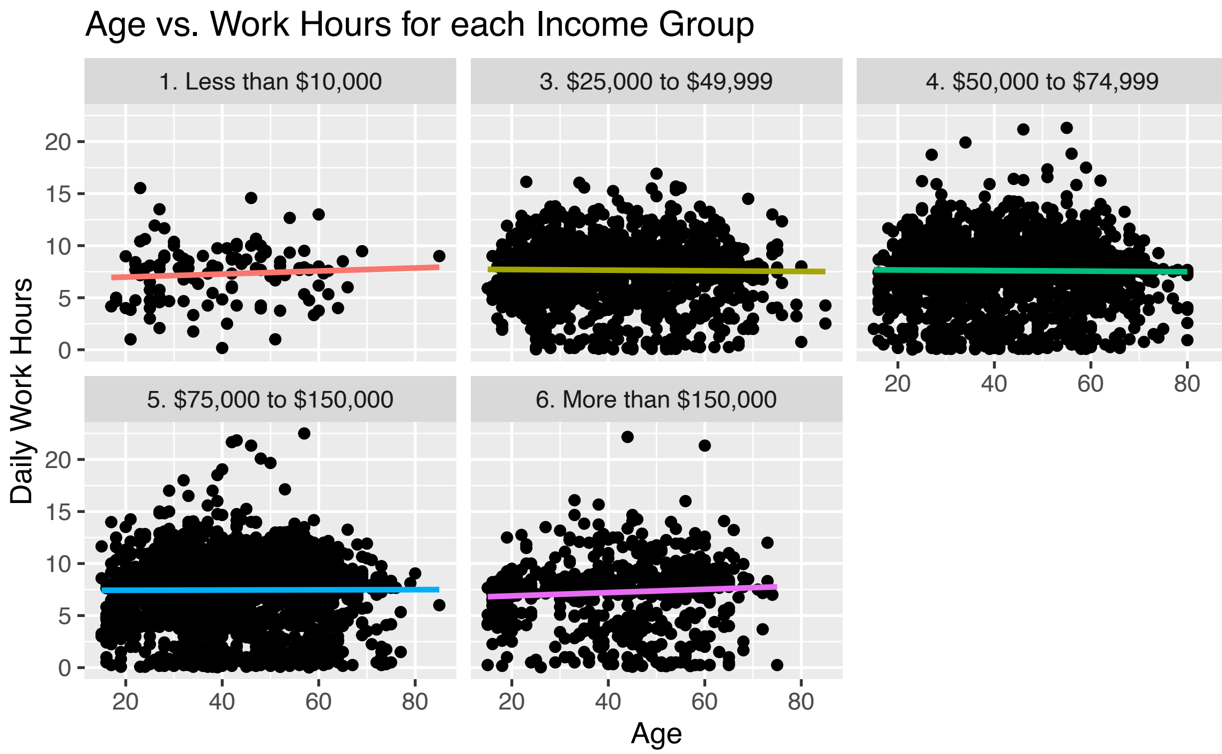
Miller, S. (2020, June 11). Black Workers Still Earn Less than Their White Counterparts. SHRM. Retrieved December 13, 2021, from https://www.shrm.org/resourcesandtools/hr-topics/compensation/pages/racial-wage-gaps-persistence-poses-challenge.aspx

Shriver, E. K. (n.d.). ATUS-X : ATUS Data Extract Builder. Retrieved December 13, 2021, from https://www.atusdata.org/atus/

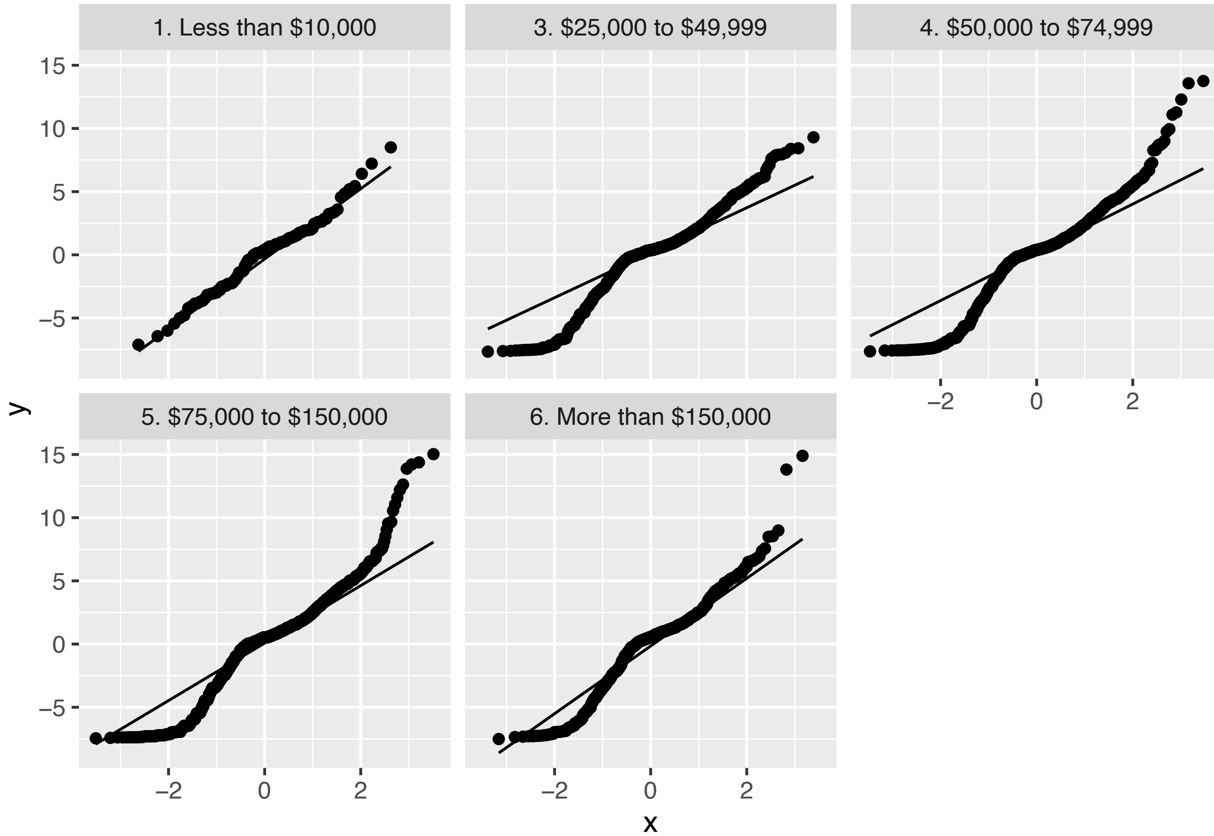
Data Appendix

Primary Model

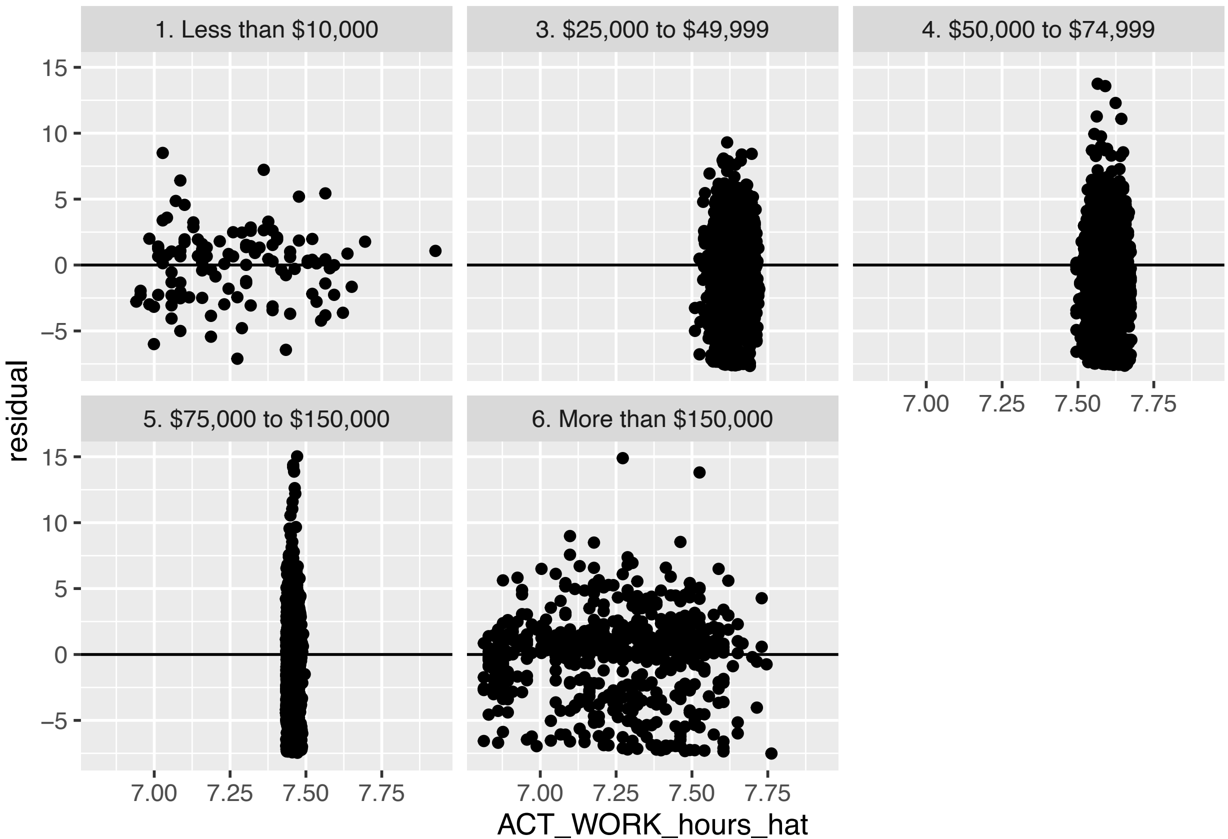
Plots



Normality Check



Linearity and Equal Variance Check



Linearity: In the residuals vs fitted values plots for income levels of less than 10,000 and more than 150,000, there does appear to be a random scatter, i.e., little discernible pattern, so the linearity assumption is met. However, for all the other income categories, the dots seem to be clustered around one region of x-values, so the linearity assumption is not met for the 25,000 to 49,999, 50,000 to 74,999, and 75,000 to 150,000 categories.

Independence of observations: The data were collected to be representative (of US inhabitants), part of which included random selection, so we have reason to believe that the independence assumption is met.

Normality of residuals: The QQ-plots compares the true pattern of the data's residuals against how a perfectly normally distributed set of residuals would fall. If our residuals are normally distributed, they fall along the reference line. For income less than 10,000, it seems that most of the points fall along the line. This is not the case with other categories of income. The left and right sides are somewhat symmetric but fall off the reference line. The normality plot (density vs residuals) clearly demonstrates that part of the issue is an overabundance of low residuals. Thus, the normality of residuals condition is not met. We will, however, proceed with caution.

Equal variance of residuals: All of the residuals vs fitted values plots appear to display equal variance: vertical spread is fairly consistent along different x-values. Importantly, there is no cone/fan shape. Thus, the equal variance assumption is met.

Model Coefficients and Summary

```
## # A tibble: 10 x 7
##    term                   estimate std_error statistic p_value lower_ci upper_ci
```

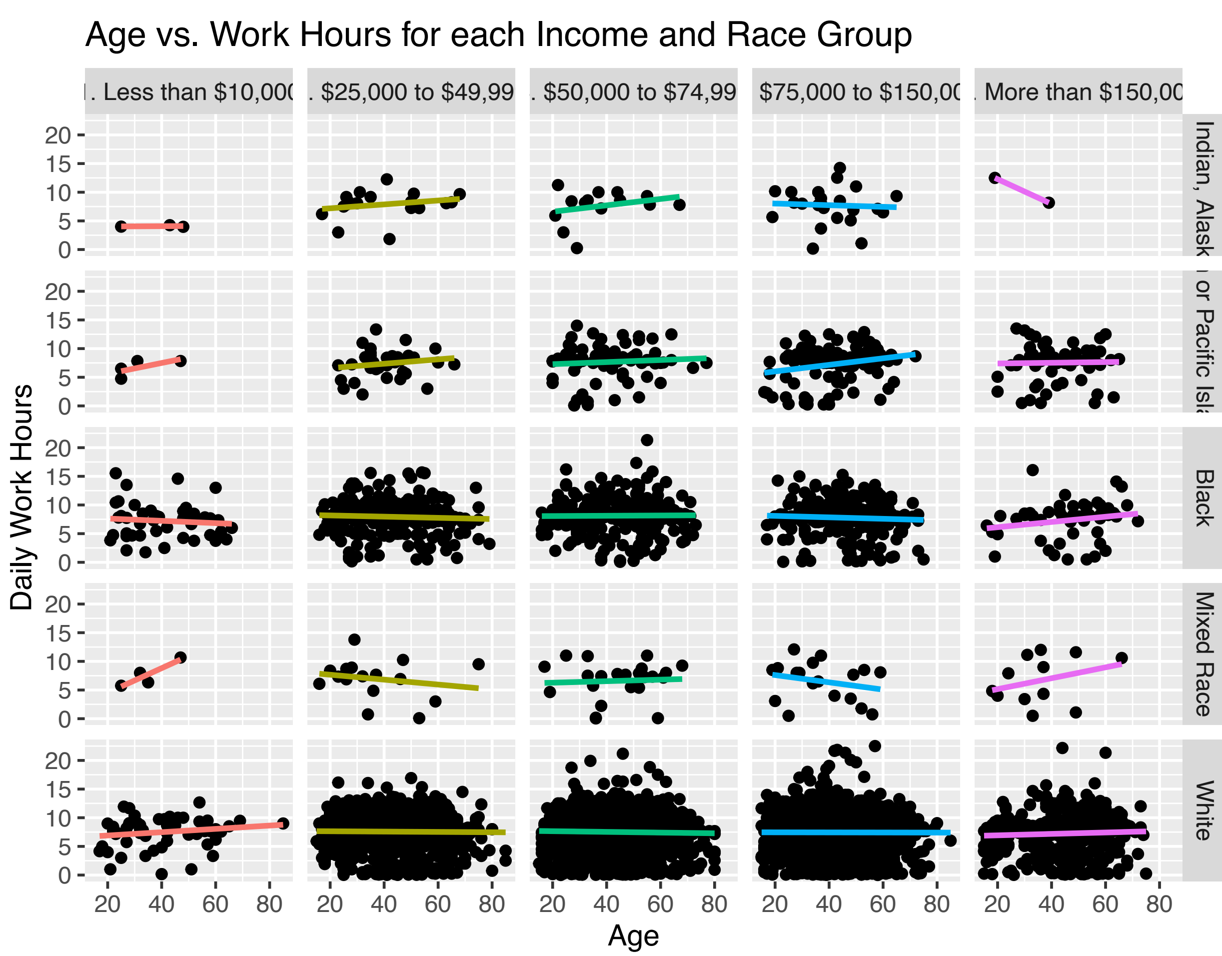
```
##    <chr>                         <dbl>     <dbl>     <dbl>   <dbl>     <dbl>     <dbl>
##  1 intercept                     6.70      0.845     7.92    0         5.04      8.35
##  2 AGE                           0.014     0.02      0.724   0.469    -0.025     0.054
##  3 FAMINCOME3. $25,000 t~        1.07      0.883     1.21    0.225    -0.659     2.80
##  4 FAMINCOME4. $50,000 t~        1.02      0.874     1.17    0.242    -0.69      2.74
##  5 FAMINCOME5. $75,000 t~        0.726     0.871     0.834   0.404    -0.981     2.43
##  6 FAMINCOME6. More than~       -0.118     0.928    -0.127   0.899    -1.94      1.70
##  7 AGE:FAMINCOME3. $25,0~       -0.018     0.021    -0.841   0.4      -0.058     0.023
##  8 AGE:FAMINCOME4. $50,0~       -0.017     0.021    -0.838   0.402    -0.058     0.023
##  9 AGE:FAMINCOME5. $75,0~       -0.014     0.021    -0.662   0.508    -0.054     0.027
## 10 AGE:FAMINCOME6. More ~        0.001     0.022     0.06    0.952    -0.041     0.044
```

```
## # A tibble: 1 x 9
##   r_squared adj_r_squared   mse  rmse sigma statistic p_value    df  nobs
##       <dbl>         <dbl> <dbl> <dbl> <dbl>     <dbl>   <dbl> <dbl> <dbl>
## 1     0.002         0.001  9.18  3.03  3.03      1.62   0.102     9  6347
```

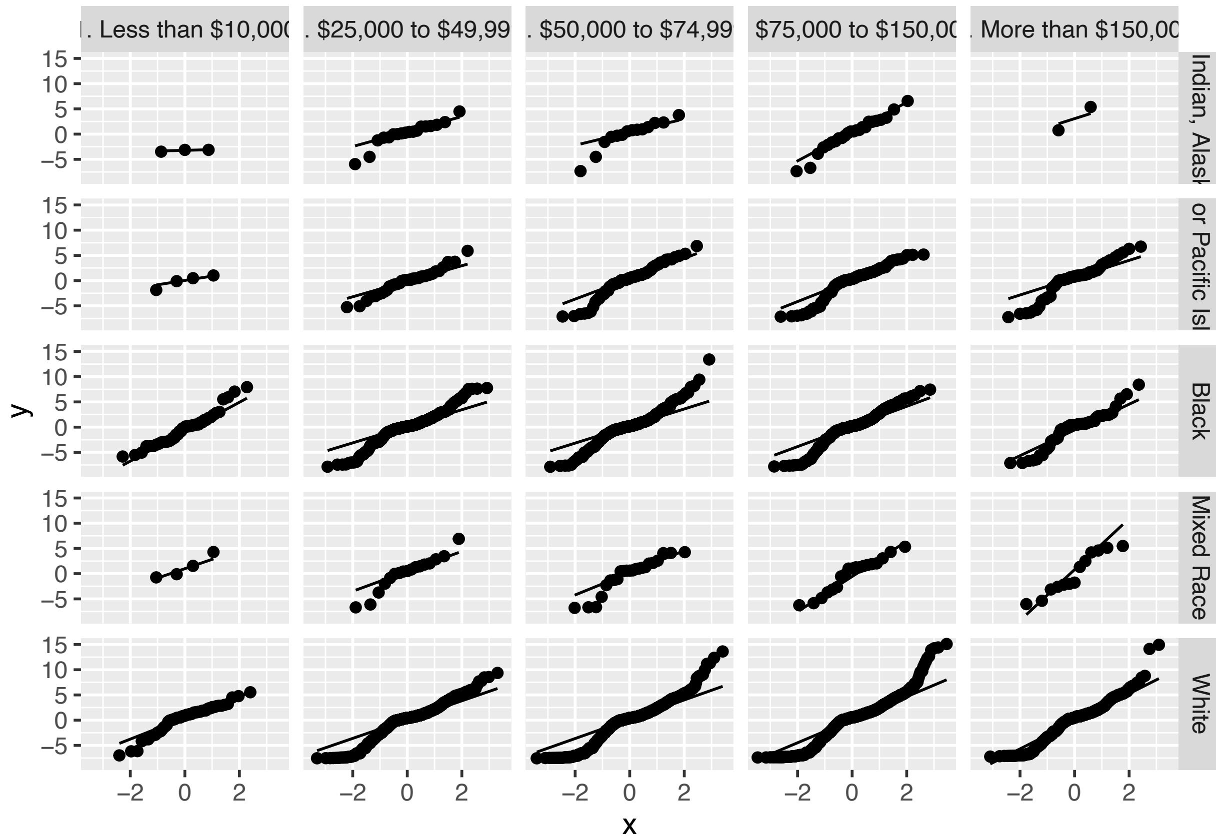
Model 1 Choice

Secondary Model

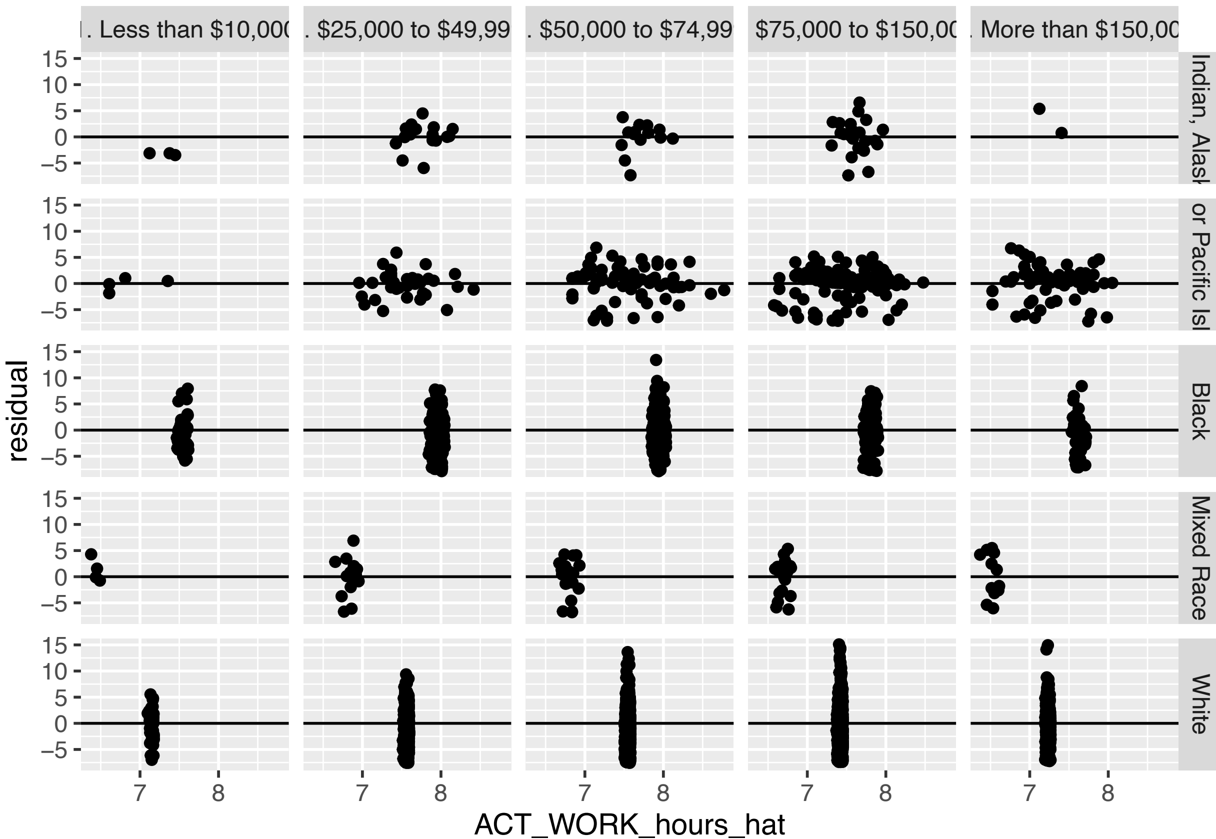
Plots



Normality Check



Linearty and Equal Variance Check



Linearity: In the residuals vs fitted values plots, most of the race categories display constant horizontal spread, so the linearity condition is met. For the black and white race categories, there appears to be a narrow cluster of dots, even though they seem equally distributed within the cluster, so the linearity condition is partially met for the income levels for these two racial groups.

Independence of observations: The data were collected to be representative (of US inhabitants), part of which included random selection, so we have reason to believe that the independence assumption is met.

Normality of residuals: All of the fitted residual plots have most of their points lining up along the reference line, so overall the normality condition is met among different race and income categories.

Equal variance of residuals: For all of the race categories across different family incomes, the condition is met because the points have consistent vertical spread.

Model Coefficients and Summary

```
## # A tibble: 14 x 7
##    term                      estimate std_error statistic p_value lower_ci upper_ci
##    <chr>                        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
##  1 intercept                    6.77      1.24      5.46    0         4.34     9.20
##  2 FAMINCOME3. $25,000 t~       0.417     0.294     1.42    0.155   -0.158    0.993
##  3 FAMINCOME4. $50,000 t~       0.402     0.291     1.38    0.167   -0.169    0.974
##  4 FAMINCOME5. $75,000 t~       0.273     0.291     0.941   0.347   -0.296    0.843
##  5 FAMINCOME6. More than~       0.086     0.308     0.279   0.78    -0.518    0.69
##  6 AGE                          0.014     0.029     0.486   0.627   -0.043    0.071
```

```
##  7 RACEAsian or Pacific ~   -1.01      1.37     -0.734   0.463   -3.70     1.68
##  8 RACEBlack                 0.915     1.26      0.726   0.468   -1.56     3.38
##  9 RACEMixed Race           -0.153     1.58     -0.097   0.923   -3.25     2.94
## 10 RACEWhite                 0.428     1.22      0.352   0.725   -1.96     2.82
## 11 AGE:RACEAsian or Paci~    0.02      0.033     0.602   0.547   -0.045    0.084
## 12 AGE:RACEBlack            -0.017     0.03     -0.578   0.563   -0.076    0.042
## 13 AGE:RACEMixed Race       -0.019     0.038    -0.501   0.616   -0.094    0.056
## 14 AGE:RACEWhite            -0.015     0.029    -0.523   0.601   -0.073    0.042
```

```
## # A tibble: 1 x 9
##   r_squared adj_r_squared   mse  rmse sigma statistic p_value    df  nobs
##       <dbl>         <dbl> <dbl> <dbl> <dbl>     <dbl>   <dbl> <dbl> <dbl>
## 1     0.005         0.003  9.15  3.03  3.03      2.56   0.002    13  6347
```

Model 2 Choice

```
## Analysis of Variance Table
##
## Model 1: ACT_WORK_hours ~ FAMINCOME + AGE + RACE
## Model 2: ACT_WORK_hours ~ FAMINCOME + AGE * RACE
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1   6337 58136
## 2   6333 58084  4    52.777 1.4386 0.2184
```

Data Summary

```
## # A tibble: 5 x 3
##   RACE                             Count MeanAge
##   <chr>                            <int>   <dbl>
## 1 American Indian, Alaskan Native     61    39.3
## 2 Asian or Pacific Islander          292    41.5
## 3 Black                              898    45.2
## 4 Mixed Race                          76    38.4
## 5 White                             5020    42.7
```

```
## # A tibble: 5 x 3
##   FAMINCOME               Count MeanAge
##   <chr>                   <int>   <dbl>
## 1 1. Less than $10,000      116    39.8
## 2 3. $25,000 to $49,999    1421    42.9
## 3 4. $50,000 to $74,999    1912    43.2
## 4 5. $75,000 to $150,000   2262    42.8
## 5 6. More than $150,000     636    42.9
```

Age Distribution by Race, Income Group