The Effect on Wages

Justin Papagelis

May 24, 2021

Abstract

This report aims to determine the significance that predictors such as gender, health insurance coverage, hours worked, and age have in predicting a person's individual salary over the past 12 months. In particular, we would like to examine the gender gap that is often talked about in society. The data we used was collected from the American Community Survey (ACS). We will use two different types of statistical tests, a Two-Way ANOVA and a Multiple Linear Regression to create models that will be useful in determining this relationship between the predictors and wage. From these tests, we were able to determine that the significant predictors of wage are health insurance coverage, sex, region, work hours, and age.

Background & Significance

We intend to examine a person's wages from the past 12 months based on different predictors that relate to the characteristics and position that a person is in. We will see if it is possible to use these variables to predict a person's salary income, and if so, which of these variables are the best predictors. We hope to create a model that can use to see the trends in individual wages through characteristics such as gender or what language is spoken. In today's society, the difference in wages between certain groups of people is highly talked about and in the process of changing. Hopefully, this model can be used to show the disparities and patterns that shape how a person's salary is affected by these predictors. Some of the research questions that we hope to investigate are as follows. How do individual wages vary across gender and to what extent is there a gender wage gap? How does the number of hours worked correlate to the overall salary that a person earns? Does the region in which a person lives influence their wages such that different paying jobs are located in certain regions? These questions are important to examine because, in the modern world, a person's living conditions are often determined by their salary and to see what factors affect this value is to investigate how the underlying differences in position can inhibit or benefit a person's overall life. Ultimately, the research hypothesis that we will be focusing on how these factors affect individual wages and to what extent.

Methods

The data that we will be using is collected from the American Community Survey (ACS) which is a survey that is done every year of people living in the United States. The ACS is used to determine how funds are distributed across the country which means that the survey includes a multitude of questions that pertain to a person's position and traits. In particular, we will be looking at the person/individual data subset to examine a person's individual wages over the past 12 months. Some biases that could found in this dataset include a coverage bias because there could be some places in the United States that could face under coverage or over coverage bias. Hopefully, this bias is minimized through the use of random sampling of the population. The larger population that we will be trying to generalize to is the United States population of adults (aged greater than 15) who earned a salary in the year 2019 because this survey was conducted with a random sampling of the United States population.

The response variable that we will be examining is a person's wages for the past 12 months. In this dataset, the person's individual salary is determined by adults older than 15 years old. The range of values is lower bounded at 0 dollars in which the person does not earn a salary and upper bounded at 999,999 dollars. This is because any values greater than 999,999 dollars are top-coded to conserve the anonymity of respondents. This means there could some concerns with people that have salaries greater than 1 billion dollars because they are not included in this data set.

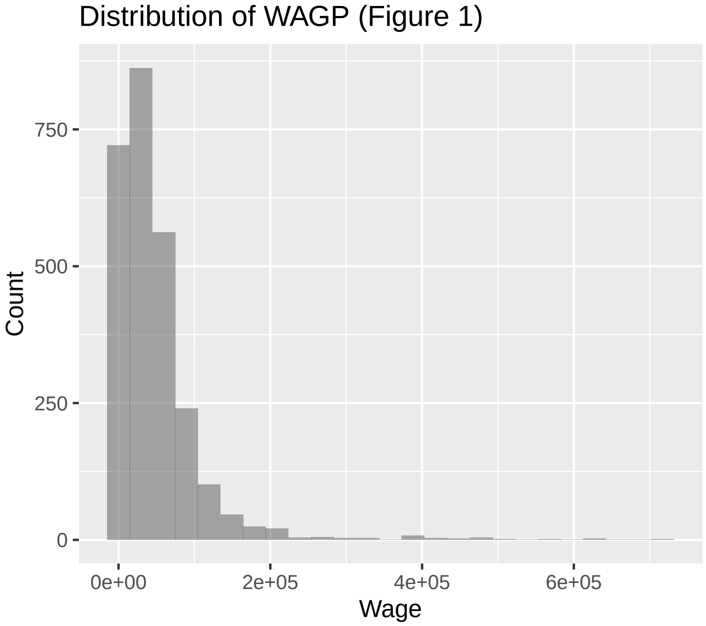
We will be attempting to use two different models to analyze the data: a two-way ANOVA model and a multiple linear regression model. These models will be used to determine the relationship between the response variable and the predictors. For these tests, we will be using a 5% significance level.

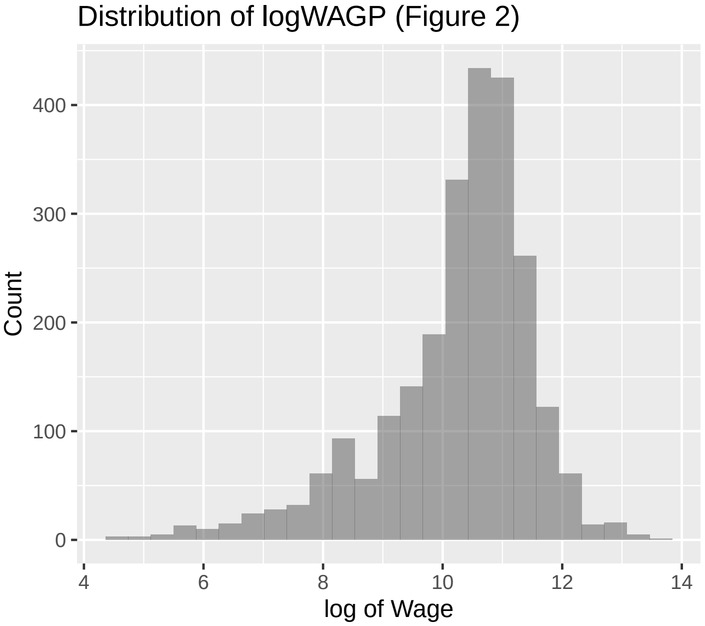
Results

Univariate EDA on Response Variable

The distribution of WAGP (Figure 1) is unimodal and heavily right-skewed because the salary is always positive and that there are potential outliers in the high-income bracket that drag the distribution to the right. Because of this, we decided to perform a log transformation on WAGP.

After the transformation, the distribution of logWAGP (Figure 2) ranges from 4.382 log-dollars to 13.483 log-dollars and is unimodal and symmetric, with a slight left skew. logWAGP has a median of 10.521 log-dollars. There does not appear to be any outliers in this distribution.





Univariate EDA on Predictors

WKHP (Figure 3) The variable WKHP represents the usual hours worked per week in the past 12 months. This was measured from adults (greater than 16 years of age) who worked in the past 12 months. The range of 1-98 represents 1 to 98 usual hours while 99 represents 99 or more usual hours. The distribution of Number of Hours Worked per week is unimodal and symmetric with a mean of 38.179 and a range of 1 to 99 hours. The could be some outliers at extremely high values of work such as 99 hours.

AGEP (Figure 4) The variable AGEP represents the age of the person in a range of 1 to 99 with the ages above 99 top-coded to preserve the anonymity of the participants. The distribution of age is appeared to be evenly distributed with a slight right skew. The ages ranged from 16 to 90 years old with a median of 43 years old. There does not appear to be any outliers.

HICOV (Figure 5) The variable HICOV represents the health insurance coverage records of the person. This factor has 2 levels: The person has health insurance coverage or the person does not have health insurance coverage. The bar chart shows that of the participants, there is a significantly larger portion of people that have health insurance coverage.

SEX (Figure 6) The variable SEX represents the gender of the participant. For this factor, there are 2 levels: Male or Female. The bar chart displays that there is approximately the same amount of each gender participating in the survey with slightly more males than females.

LANX (Figure 7) The variable LANX represents if a language other than English is spoken at home. There are 2 levels for this factor: speaks another language at home or speaks English at home. Of the participants, there is a greater number who speak English at home than the number of people who speak another language at home.

REGION (Figure 8) The variable REGION represents the region code based on the 2010 Census definitions. For this factor, there are 4 levels: Northeast, Midwest, South, and West. The category of Puerto Rico was removed from this variable because there was no data from this region. Each level appears to have roughly the same number of participants with a greater number in the South and the least amount in the Northeast.

Bivariate EDA between Response Variable and Predictors

Relationship Between `logWAGP` and `WKHP` (Figure 9) There appears to be a moderately strong, positive, and linear relationship between `logWAGP` and `WKHP`. There is a correlation of 0.61464 between these two variables.

Relationship Between `logWAGP` and `AGEP` (Figure 10) There appears to be a moderate, positive relationship between `logWAGP` and `AGEP`. There is a correlation of 0.27615. The relationship is curved which means that in our multiple linear regression, we should add another order term to use a quadratic model.

Relationship Between `logWAGP` and `HICOV` (Figure 11) It appears that the median `logWAGP` for having health insurance coverage is 10.597 log-dollars which is greater than the median for not having health insurance coverage which is 9.971 log-dollars. With health insurance coverage appears to have more variability as its IQR is 1.289 log-dollars (from 9.7981 log-dollars to 11.082 log-dollars) which is slightly greater than the IQR of no health insurance coverage which is 1.1283 log-dollars (from 9.3927 log-dollars to 10.521 log-dollars). There are a couple of outliers on the lower side for both with and without health insurance coverage.

Relationship Between `logWAGP` and `SEX` (Figure 12) It appears that the median `logWAGP` for males is greater than that of females at 10.714 log-dollars and 10.342 log-dollars, respectively. The IQR of females (1.4053 log-dollars) is greater than the IQR of males (1.2131 log-dollars). There are some outliers on the lower side for both genders.

Relationship Between `logWAGP` and `LANX` (Figure 13) It appears that the median `logWAGP` for speaking another language at home is less than the the median `logWAGP` for speaking English at home at 10.390 log-dollars compared to 11.051 log-dollars, respectively. The IQR for speaking English at home is 1.346 log-dollars which is slightly higher than the IQR for speaking another language at home which is 1.1701 log-dollars. As before, there are some outliers on the lower side for both levels.

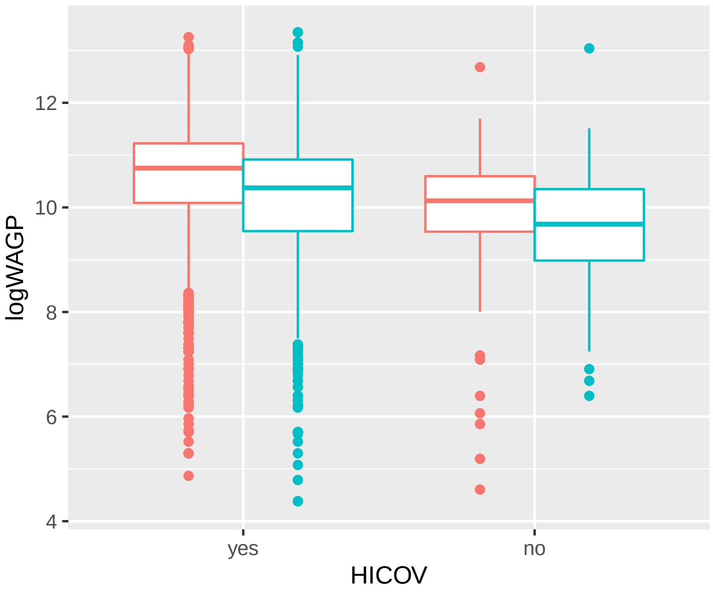
Relationship Between `logWAGP` and `REGION` (Figure 14) It appears that the median `logWAGP` for each region is approximately the same at 10.621 log-dollars for the northeast and 10.491 log-dollars for the midwest, south, and west. The IQRs of each region are also approximately the same for the northeast, midwest, south, and west at 1.3597, 1.3217, 1.2848, and 1.3239 log-dollars, respectively. Each region has a couple outliers on the lower side.

Modeling

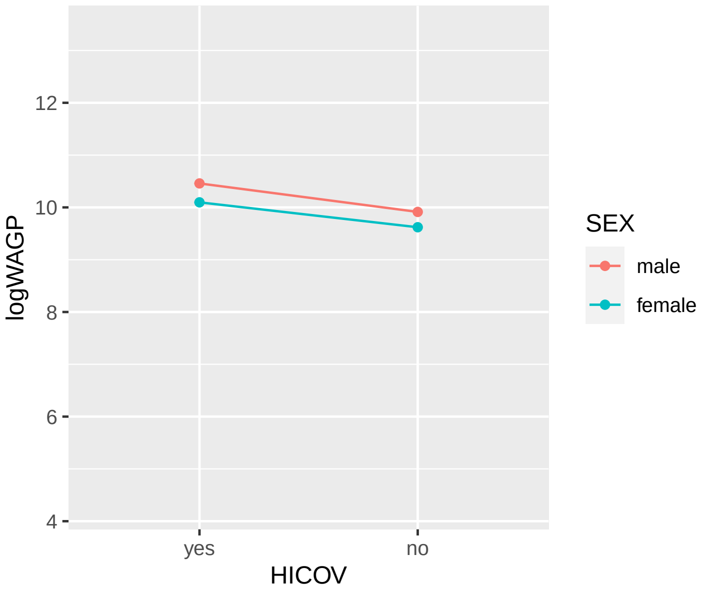
Two-Way ANOVA

First, we will use a Two-Way ANOVA test to determine the relationship between `logWAGP` and the two levels of the categorical variables, `HICOV` and `SEX`. This test can be used to determine if there is an interaction effect between these two variables and to what extent we need to factor that in.

side−by−side boxplots (Figure 15)



interaction plot (Figure 16)



Using the side-by-side boxplots above, the distributions of male and female participants with health insurance coverage as well as the distributions of male and female participants without health insurance coverage appear to be roughly symmetric which satisfies the Normality condition. The median `logWAGP` of participants with health insurance is greater than that of participants without health insurance. The median `logWAGP` of male participants with health insurance coverage is greater than the median `logWAGP` of female participants with health insurance coverage. The IQRs for each of the distributions seem to be relatively similar which satisfies the Equal Variance Condition. The ratio of the highest SD (no.male) to the lowest SD (no.female) is equal to 1.2123 (= 1.3080/1.0789) which is less than 2 and further suggests that our Equal Variance Condition is satisfied. There are some outliers for the lower values of log-dollars for each of the plots, so we will proceed with caution when using the two-way ANOVA model. Using the interaction plot, there does not appear to be an interaction effect between `SEX` and `HICOV` because the difference in median `logWAGP` between males and females with health insurance coverage is 0.3791 log-dollars which is similar to the difference in median `logWAGP` between males and females without health insurance coverage which is 0.4463 log-dollars. On the interaction plot, the two line segments are nearly parallel which shows that there might not be a significant interaction effect.

Using the Analysis of Variance Table (Table 2), there appears to be a significant `SEX` effect and a significant `HICOV` effect. This is because the F-test gives an F-statistic that is high for these predictors which in turn yields a low p-value. This means that there is significant evidence of a difference in means due to the `SEX` effect and the `HICOV` effect. On average, male participants have a greater `logWAGP` than female participants. On average, participants with health insurance coverage have a greater `logWAGP` than participants without health insurance coverage. Since the interaction term is not significant, this suggests that perhaps a model without an interaction term would be a better fit.

Using the two-way ANOVA test without interaction provides a model that has a significant SEX and significant HICOV effect (Table 3). It appears that the residuals in the Residual vs Fitted (see Appendix) plot are evenly spread around zero and there are no trends or patterns which means the Equal Variance of Residuals condition is satisfied. The QQ-plot is roughly linear with some deviation from the reference line at the left end of the plot which suggests we should proceed with caution with the Normality condition. Because the participants were randomly selected, the Randomness condition is verified and we can assume Independence from person to person. Comparing the two-way ANOVA model with and without interaction, it appears that the model without interaction is a better fit because the adjusted R^2 of the model with interaction is 0.0302 which is slightly greater than the adjusted R^2 of the model without interaction which is 0.0298. The R^2 of both models is extremely weak which suggests that we will need to add some more predictors to get a stronger model.

ANOVA Model Summary:

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.4585     0.0368  284.19  < 2e-16 ***
## SEXfemale    -0.3587     0.0513   -6.99  3.5e-12 ***
## HICOVno      -0.5137     0.0934   -5.50  4.2e-08 ***
##
## Residual standard error: 1.27 on 2454 degrees of freedom
## Multiple R-squared:  0.031,  Adjusted R-squared:  0.0302
## F-statistic: 39.2 on 2 and 2454 DF,  p-value: <2e-16
```

Multiple Linear Regression

Next, we will use Multiple Linear Regression to model the relationship the response variable, `WAGP` has with both the quantitative and categorical predictors. To start off, we created a full model that uses all of the predictors. This model uses the predictors, `HICOV`, `SEX`, `LANX`, `REGION`, `WKHP`, and `AGEP`. While checking the conditions for this model, it appears that the Linearity condition is satisfied because the Residuals vs Fitted Plot (see Appendix) shows a linear relationship. In this same plot, the residuals appear to be evenly spread with little pattern although there is some tapering of the residuals for the higher fitted values which suggests we have a mild concern with the Equal Variance condition. The Independence condition is satisfied because the participants of this survey were randomly sampled. The QQ-plot shows some deviation from the reference line near both tails which suggests a mild concern with the Normality conditions. There are a couple of outliers, but they do not seem to affect the model in any adverse ways. Therefore, we will proceed with caution with this model. We also checked the VIFs of this model, but since all of the values were below 5, we do not detect any signs of multicollinearity (Table 5).

The model is overall significant because the F-statistic of the F-test (232) gives a p-value that is less than 2×10^{-16} which is under the accepted significance level of 0.05. The model also has three significant predictors because the p-values of their individual t-tests fall below 0.05 as well: `REGION`, `WKHP`, and `AGEP`. The test is moderately strong with a Multiple R^2 of 0.431 and a Residual Standard Error (RSE) of 0.976 (Table 4)

Next, we performed a Stepwise Regression which is a variable selection procedure that starts with no predictors and uses forward and backward selection to create a model with the most significant predictors. This resulted in a model with the predictors, WKHP, AGEP, HICOV, and REGION. We see the same concerns that we saw in the previous model with the conditions, with the rest holding (see Appendix). Therefore, we will proceed with caution when using this model as well.

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## logWAGP ~ 1
##
## Final Model:
## logWAGP ~ WKHP + AGEP + HICOV + REGION
##
##
##       Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                          2456     4095.6 1257.434
## 2   + WKHP  1 1547.235      2455     2548.3   93.667
## 3   + AGEP  1  186.441      2454     2361.9  -91.007
## 4  + HICOV  1   21.381      2453     2340.5 -111.351
## 5 + REGION  3    8.649      2450     2331.9 -114.447
```

The stepwise model is overall significant with an F-statistic of 309 which gives a p-value less than 2×10^{-16}. The test appears to have a similar strength as the previous full model at a Multiple R^2 of 0.431. This model also has a RSE of 0.976 (Table 6). This is expected, however, because the stepwise regression model gives the full model with the non-significant predictors removed.

The next model we will use is a quadratic model to account for the curved relationship that we noticed AGEP had with logWAGP. We decided to add a higher-order term by squaring AGEP to create AGEPsq. We incorporated this into the full model using all of the predictors as well as AGEPsq. This resulted in a model where all of the predictors except for LANX were significant and if LANX is removed from the model, there is no difference in the strength of the model or the RSE. The conditions of this model hold about the same as the previous two MLR models and we will proceed with caution for this model as well (see Appendix) . The strength of the model is moderately strong with a Multiple R^2 of 0.484 and a RSE of 0.929. All of the predictors of this model are significant.

When choosing which model to use as our final model, the full model and the quadratic model are quite similar although the quadratic model has a greater adjusted R^2 than the full model and a lower RSE. The Stepwise Regression produces a model that is simpler as it has fewer predictors than the quadratic model, but as with the full model, the strength of the stepwise model is less than that of the quadratic model. Additionally, the quadratic model accounts for the curved relationship that appeared in our EDA which suggests that it is also more appropriate to use the quadratic model rather than try and fit a linear relationship. The last thing that we did is look at the outliers of the model and if removing them would improve our model. We found there to be 35 large outliers to the data because they had standardized residual values greater than 3. There were no influential points. We decided to remove the large outliers from the data set to see if the model could better fit the data. The refitted model has a Multiple R^2 of 0.488 and a RSE of 0.924 which is an improvement in both of these calculations. Therefore, our final model will be the quadratic model with the large outliers removed.

Quadratic Model Summary:

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.65e+00   1.44e-01   39.20  < 2e-16 ***
## HICOVno       -3.90e-01   6.85e-02   -5.69  1.4e-08 ***
## SEXfemale     -1.18e-01   3.83e-02   -3.09  0.00202 **
## REGIONmidwest -2.04e-01   5.70e-02   -3.58  0.00035 ***
## REGIONsouth   -1.85e-01   5.44e-02   -3.40  0.00068 ***
## REGIONwest    -1.35e-01   5.60e-02   -2.41  0.01592 *
## WKHP           4.89e-02   1.55e-03   31.54  < 2e-16 ***
## AGEP           1.32e-01   7.29e-03   18.12  < 2e-16 ***
## AGEPsq        -1.30e-03   8.12e-05  -16.04  < 2e-16 ***
##
## Residual standard error: 0.924 on 2445 degrees of freedom
## Multiple R-squared:  0.488,  Adjusted R-squared:  0.486
## F-statistic:  291 on 8 and 2445 DF,  p-value: <2e-16
```

Interpretations

Two-Way ANOVA

For the Two-Way ANOVA test, we used `HICOV` and `SEX` to predict the wage of a participant. The final model is $log\widehat{WAGP} = 10.4585 - 0.3587(SEXfemale) - 0.5137(HICOVno)$. This means that the predicted mean wage of a female participant with no health insurance would be 14,560.97 dollars ($e^{9.5861}$ dollars). If the participant was instead a male with no health insurance, the predicted wage would be 20,843.55 dollars ($e^{9.9448}$ dollars). Having health insurance for a male would increase his predicted wage by 67.15% to get 34,839.25 dollars ($e^{10.4585}$ dollars). Because we found there to be no interaction effect between `SEX` and `HICOV`, if a female participant had health insurance, her predicted wage would also increase by 67.15% to get 24,338.14 dollars ($e^{10.0998}$ dollars).

Multiple Linear Regression

For the Multiple Linear Regression test, we ended up using a quadratic model to predict the wage of a participant. The final model is: $log\widehat{WAGP} = 5.65 - 0.39(HICOVno) - 0.118(SEXfemale) - 0.204(REGIONmidwest) - 0.185(REGIONsouth) - 0.135(REGIONwest) + 0.0489(WKHP) + 0.132(AGEP) - 0.0013(AGEPsq)$. The predicted wage of a 0-year-old male participant with health insurance living in the northeast that works 0 hours per week is 284.29 dollars. Since this scenario is highly impossible, it would not make sense to interpret the intercept. This could also be because of some extrapolation because no data was collected from participants under the age of 16. While we intercept the coefficients of the model, we will allow for simultaneous changes in the other variables. If the participant has does not have health insurance, their predicted wage would decrease by 47.70% on average. If the participant was female, their predicted wage would decrease by 12.52% on average. Depending on the region in which the participant lives, their predicted wage on average would decrease 22.63% if they lived in the Midwest, decrease 20.32% if they lived in the South, and decrease 14.45% if they lived in the West. For every one work hour increase, the predicted wage of the participant increases on average by 5.01%. We used a higher-order term to represent `AGE` (to account for the curved relationship) which means that it is more difficult to interpret the coefficient than the linear relationships/indicator variables. As the participant increases their age by one year, their predicted wage has a greater positive percent change when they are younger, a small percent change when they are near the age of 50, and a larger negative percent change as they get older. Another way to say this is that as a person ages, their predicted average wage percent change for a year is positive but the amount by which it is increasing is decreasing each year until the age of 50 when the percent change is negative and the amount at which it is decreasing every year is increasing.

Conclusions

The objective of this study was to determine the connections that a person's individual wage from the past 12 months has with several other factors that include sex, health insurance coverage, hours worked, age, etc. We were able to create two final models that were able to predict the average individual wage of a person. Our multiple linear regression model had an R^2 value of 0.488 which means that the model created explains 48.8% of the variation in wages. This means that the model we created is moderately strong and can be useful in predicting wages.

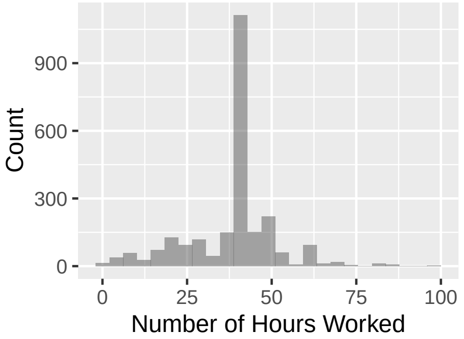
Overall, we answered many of the research questions we had initially posed and many of the answers we found were not surprising. For example, the Two-Way ANOVA test clearly showed a sex effect on wages which clarifies the gender wage gap. This wage gap is present but significantly smaller than we expected it to be. There is a significant correlation between the number of hours worked and a person's income. This is most likely explained through the fact that many individual's salaries are based on an hourly wage rather than a fixed contract in which the individual is expected to work many hours. Therefore it follows that a higher number of hours worked results in a higher salary. One surprising finding was that the region in which the individual lived was a significant predictor of wage. We did not expect this to occur because there are a variety of different paying occupations around the country, but it seems that, on average, there is a higher concentration of higher salary jobs in places such as the northeast. It was interesting to analyze the relationship that age had with wages because the relationship was not linear.

There are limitations to the findings we came to because of the bias mentioned earlier in the study. There is no way of verifying that the information an individual submits is truthful. There is under coverage and over coverage that might misrepresent our country as a whole. It is difficult to account for all of these biases, but the models we created can still be useful in predicting wages, on average. Ultimately, many of the factors that make up a person (including gender) will continue to determine the wages of that individual although there is hope that as there is more equality introduced into the world, the characteristics that make up a person will become less important in determining their wages and defining how that person can live their life.

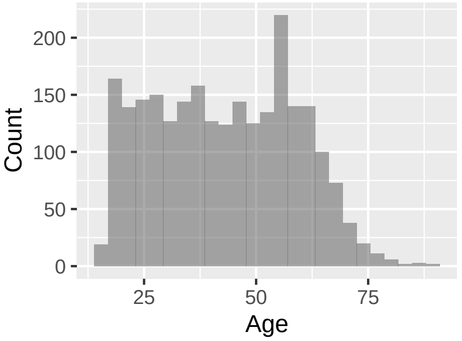
Appendix

Univariate EDA on Predictors

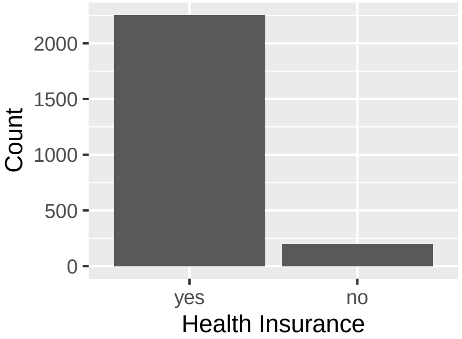
Distribution of WKHP (Figure 3



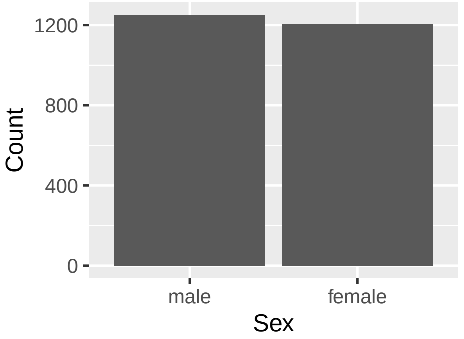
Distribution of AGEP (Figure 4)



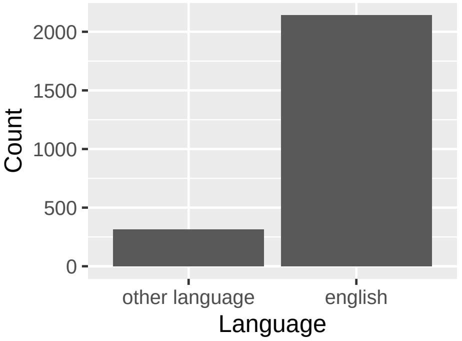
Distribution of HICOV (Figure



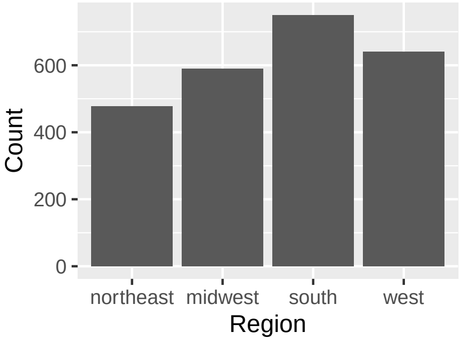
Distribution of SEX (Figure 6)



Distribution of LANX (Figure 7



Distribution of REGION (Figure



Bivariate EDA between Response Variable and Predictors

Figure 9

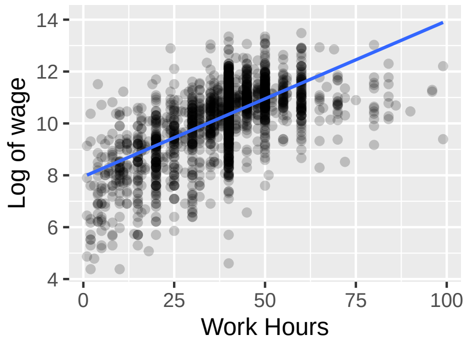


Figure 10

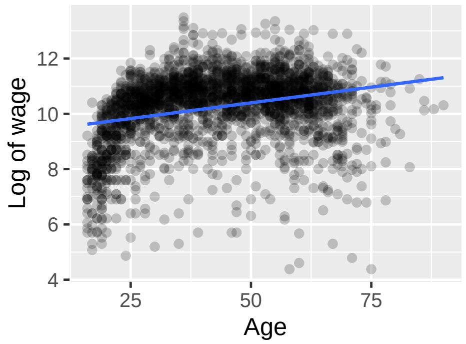


Figure 11

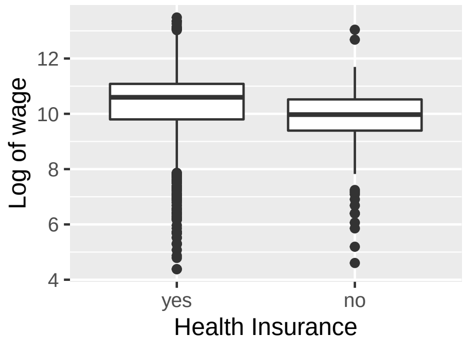


Figure 12

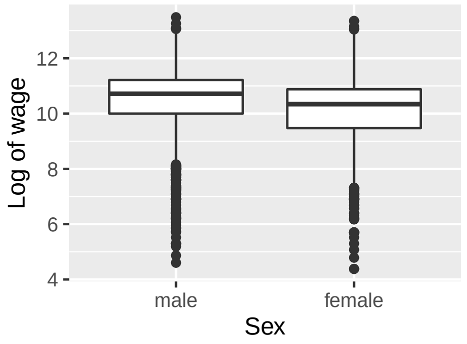


Figure 13

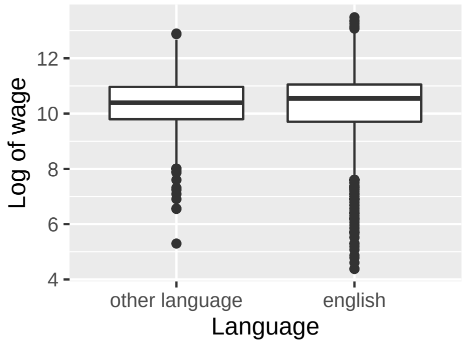
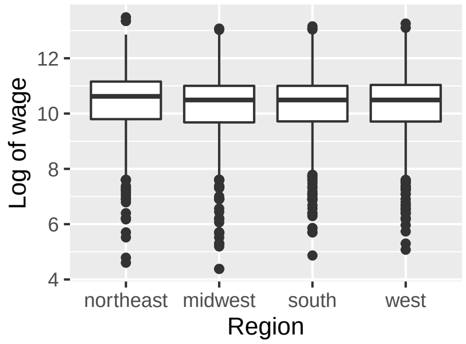


Figure 14



Two-Way ANOVA

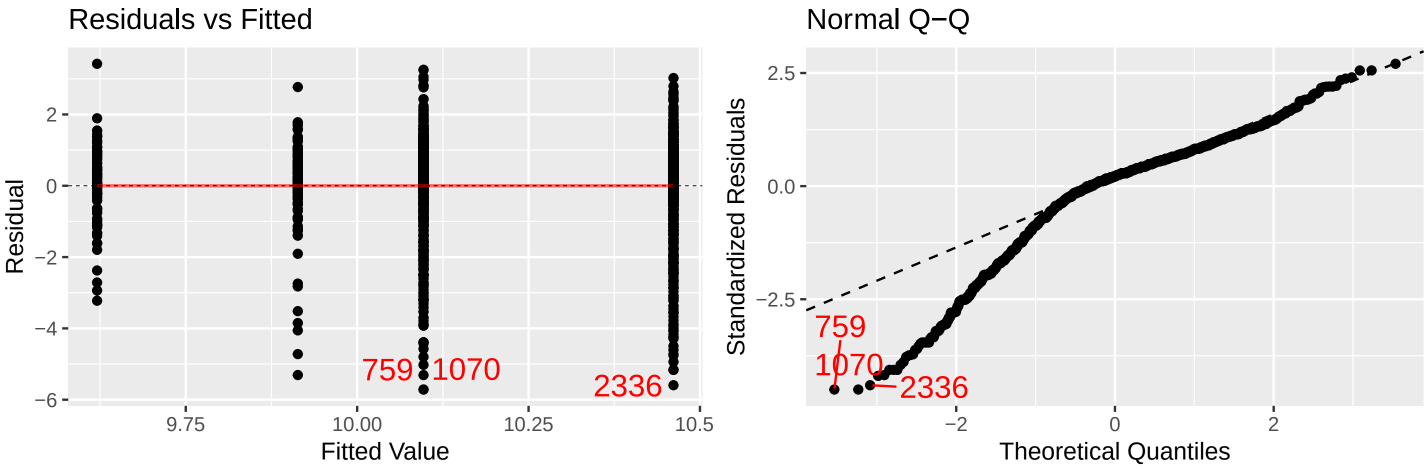
Table 1

```
##    HICOV.SEX    min      Q1  median     Q3    max    mean     sd    n missing
## 1   yes.male 4.8675 10.0858 10.7515 11.225 13.483 10.4592 1.2702 1145       0
## 2    no.male 4.6052  9.5360 10.1266 10.597 12.682  9.9134 1.3080  106       0
## 3 yes.female 4.3820  9.5468 10.3735 10.915 13.349 10.0942 1.2817 1108       0
## 4  no.female 6.3969  8.9809  9.6803 10.301 11.513  9.5847 1.0250   95       0
```

ANOVA model with interaction

Table 2

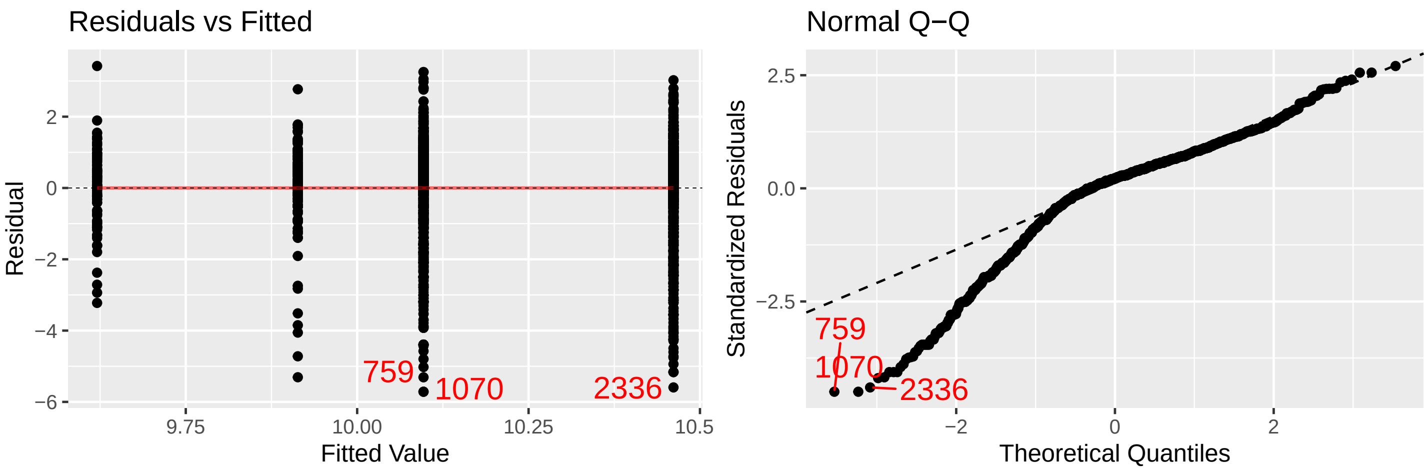
```
## Analysis of Variance Table
##
## Response: logWAGP
##             Df Sum Sq Mean Sq F value  Pr(>F)
## SEX          1     78    77.9   48.14 5.1e-12 ***
## HICOV        1     49    48.9   30.24 4.2e-08 ***
## SEX:HICOV    1      0     0.2    0.15     0.7
## Residuals 2453   3969     1.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



ANOVA model without interaction

Table 3

```
## Analysis of Variance Table
##
## Response: logWAGP
##             Df Sum Sq Mean Sq F value  Pr(>F)
## SEX          1     78    77.9    48.2 5.0e-12 ***
## HICOV        1     49    48.9    30.2 4.2e-08 ***
## Residuals 2454   3969     1.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Multiple Linear Regression

Full Model

Table 4

```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.50544    0.10597   70.82  < 2e-16 ***
## HICOVno         -0.33740    0.07245   -4.66  3.4e-06 ***
## SEXfemale       -0.05524    0.04028   -1.37   0.1704
## LANXenglish     -0.03519    0.05975   -0.59   0.5559
## REGIONmidwest -0.17131    0.06023   -2.84   0.0045 **
## REGIONsouth     -0.13142    0.05736   -2.29   0.0220 *
## REGIONwest      -0.11096    0.05921   -1.87   0.0611 .
## WKHP             0.05744    0.00153   37.53  < 2e-16 ***
## AGEP             0.01712    0.00127   13.53  < 2e-16 ***
##
## Residual standard error: 0.976 on 2448 degrees of freedom
## Multiple R-squared:  0.431,  Adjusted R-squared:  0.429
## F-statistic:  232 on 8 and 2448 DF,  p-value: <2e-16
```

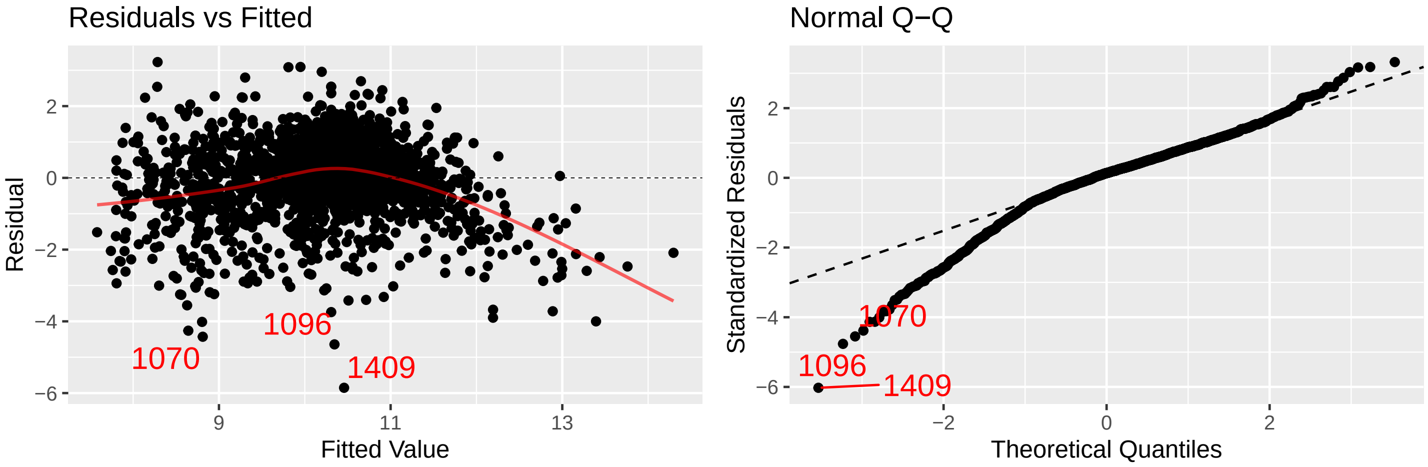


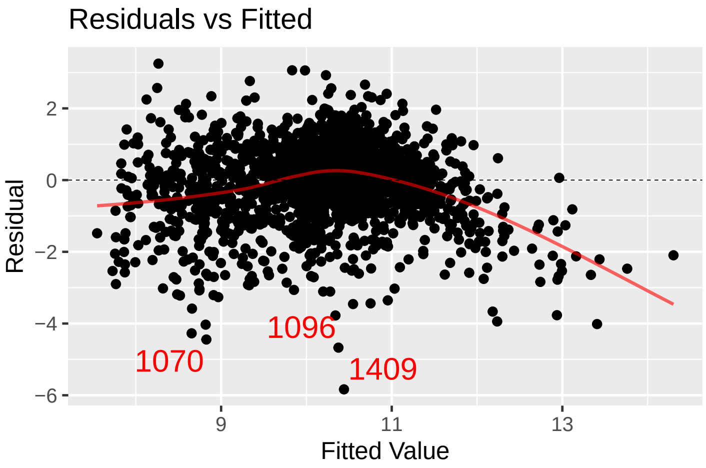
Table 5: VIFs of full model

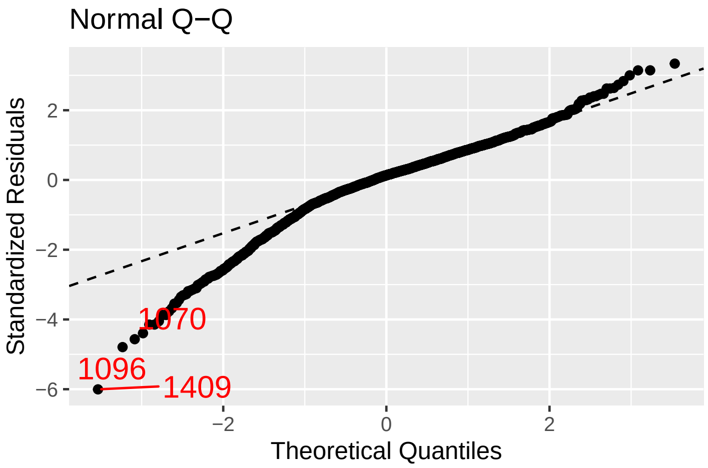
```
##          GVIF Df GVIF^(1/(2*Df))
## HICOV  1.0224  1          1.0112
## SEX    1.0467  1          1.0231
## LANX   1.0274  1          1.0136
## REGION 1.0296  3          1.0049
## WKHP   1.0558  1          1.0275
## AGEP   1.0189  1          1.0094
```

Stepwise Regression

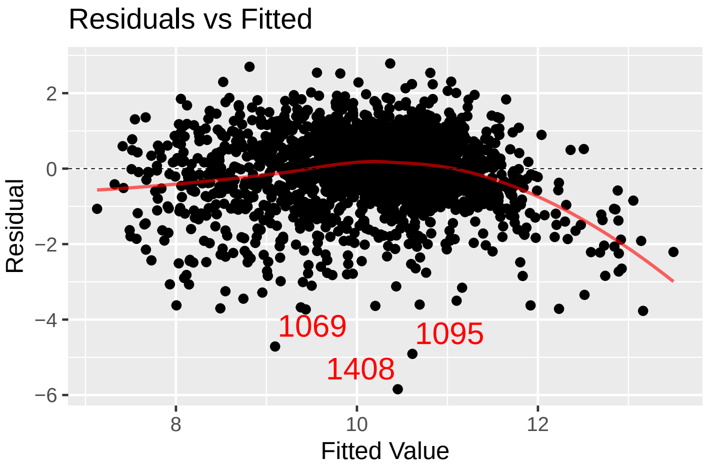
Table 6

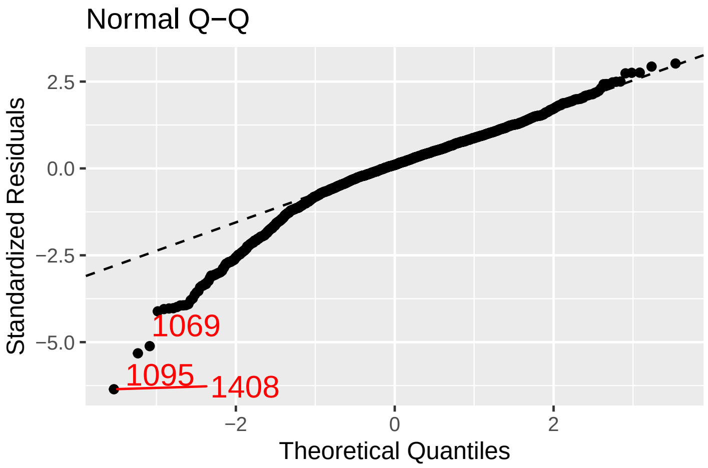
```
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.43280    0.08728   85.16  < 2e-16 ***
## WKHP             0.05787    0.00150   38.59  < 2e-16 ***
## AGEP             0.01708    0.00127   13.50  < 2e-16 ***
## HICOVno         -0.33192    0.07212   -4.60  4.4e-06 ***
## REGIONmidwest   -0.17493    0.06014   -2.91   0.0037 **
## REGIONsouth     -0.13339    0.05734   -2.33   0.0201 *
## REGIONwest      -0.10594    0.05906   -1.79   0.0730 .
##
## Residual standard error: 0.976 on 2450 degrees of freedom
## Multiple R-squared:  0.431,  Adjusted R-squared:  0.429
## F-statistic:  309 on 6 and 2450 DF,  p-value: <2e-16
```





Quadratic Model


Residuals vs Fitted
Residual
Fitted Value


Normal Q−Q
Standardized Residuals
Theoretical Quantiles