

Education and the Wage Gap

An Analysis by States of the United States

Zachary Palmore

11/16/2020

Abstract

Although efforts have been made to reduce the gender wage gap and studies suggest that people support efforts to promote pay equity, evidence of this gap, measured through the average earnings of males and females, show that there has been little change since 2004. However, some studies point towards an increasing average annual salary for women during this time and attribute the cause to factors such as education. For this analysis we observe for those aged twenty-five and older, if the level of education attained has an effect on the average annual earnings of males and females across the United States in 2018. We extend this analysis further to determine if there is a difference at the state level in wages and discover where, if any, state has significantly reduced their gender wage gap in 2018. Preliminary results suggest that higher education does assist women in increasing their average annual earnings but it does not reduce the gender wage gap.

Introduction

Women who are 16 years of age or older and employed in the United States have lower average annual salaries than men. According to the U.S. Bureau of Labor Statistics, full-time wage and salaried workers had the lowest difference in their median weekly earnings in 2014 with women earning 82.5% of the median male earnings. This difference was greater in 1979 with women earning 62% of men's earnings. Since 1979 when comparable salary data became available, there has not been an occasion in which the average earnings of women in any state within the United States has met or exceeded the average earnings of men. However, some studies have shown that education helped increase the average earnings of women during this timespan.

Among the population of individuals who are 25 years of age or older and employed in the United States, those with higher levels of education have higher earnings on average. Using 1979 as a benchmark, workers with only high school diplomas have experienced an overall decline in wages but the decline for women was smaller with a 10% reduction compared to 29% for men. Additionally, workers with a bachelor's degree over this timespan have increased in wages overall with women experiencing a greater increase of 35% than men's 20%. This suggests education may have a significant impact on the reduction of the gender wage gap.

When considering this gender wage gap, or the difference in average earnings of the traditional sexes of male and female, it is important to acknowledge several factors outside of education that influence the average earnings of individuals. First, that certain industries are majority male while others are majority female. Some argue that this is the reason for discrepancies in pay, suggesting women choose lower-paying roles. However, these studies tend to omit the other variables influencing an individual's average earnings.

Two other factors often cited when considering why women earn less than men, is that women have fewer years of experience and work fewer hours thus earning less in wage on average. Both factors could be due to a restructuring and prioritization of time for children and family-related responsibilities. Since cultural norms traditionally rely on women to bear the brunt of family responsibilities, they are unable to dedicate as much time to their occupations as men. This results in lost opportunities for women to potentially increase

their wage. Given the prevalence of this phenomenon in society, these factors should be acknowledged as influences that, when controlled for, cause the analysis to miss the social impact these cultural norms have on the gender pay gap.

Lastly, although the Equal Pay Act of 1963 was enacted by Congress to “prohibit discrimination on account of sex in the payment of wages by employers” it must be acknowledged that certain types of discrimination remain and they can have a direct influence on an individual’s pay, especially for women and minorities. Discrimination is varied and widespread. To give just one example taken from the Center for American Progress, an employer may not actively choose to pay a woman less for the same position as a man, but may instead base her salary on her previous compensation and offer a certain percentage increase equal to the offer given to a man. In circumstances where any gendered wage discrimination has occurred in the past, this may only further the act of discrimination. Actions, such as this, are largely uncontrollable. For this reason, the best course of action when analyzing the gender wage gap is to take a comprehensive approach.

This analysis includes all of these major influencing factors as well as those which may not be named while reviewing the gender wage gap in education. It is unadjusted and describes the full-picture of the gender wage gap in the United States broken down by state.

A report published by the Economics Policy Institute in 2020, describes how “adjusted” measures (those that try to control for these factors) will always produce less realistic results. The tradeoff of these adjustments is informative and notably useful in understanding the gender wage gap from one perspective in which women and men are gauged on whether they are receiving equal pay for equal work. However, these ignore the differences in opportunities presented to each sex that change the nature of their connection with their employer prior to and during wage negotiations.

Additionally, for this analysis, reviewing the gender wage gap across education levels by state adds another host of variables that are difficult to control for when already considering the multitude of other factors that influence the gender wage gap. Although multivariate regression could assist in singling out those factors, it is beyond the scope of this analysis and would degrade the applicability of the results to the real-world.

To understand the gender wage gap and how education impacts it, we review the most recent data from the U.S. Census Bureau’s American Community Survey. Through this analysis we expect a continuation of trends in the gender wage gap- that the gap remains close to the same level as it has since 2004- but intend to look into this further for statistical significance. To apply a perhaps more useful measure, the data is calculated by state to reveal localized changes and discover recent trends in how education impacts the gender wage gap.

Data

This data comes from the American Community Survey (ACS) of the United States Census Bureau. They collect a random sample from the American population every year that is not decennial census year with no household ever receiving the survey more than once every five years. The ACS has a large sample size (approximately 2.143 million individuals) for this study and the data is widely considered a representative estimate of the population within their margins of error. Those margins are provided in the *income_2018_bystate* data frame. The ACS follows strict protocol on form-filled survey data collection and adheres to Approximate normality is reinforced in the small difference between the mean and median of the sample observations. See the reference to Census Bureau methodology listed in the references section for further details.

A copy of the data used in this survey was made available in the public Github repository listed below with the reference by the author. It was made available as a csv file and can be read remotely from any connected device to improve the reproducibility of this work.

In this analysis, the survey data is called directly from the remotely accessible repository and used throughout. Subsets of the spreadsheet are broken out into smaller tables and data frames for use in the analysis. We begin with the essentials of state identifiers, median earnings for the population of working age (16), and median earnings for educational attainment. Each is broken down by the traditional male or female gender for comparison.

```

# Load Data
income_2018_bystate <- read_csv("https://raw.githubusercontent.com/palmorezm/msdsdata606/master/Proje

## Parsed with column specification:
## cols(
##   .default = col_character()
## )

## See spec(...) for full column specifications.

# Select relevant income-specific variables by state
earnings <- income_2018_bystate[2:53,1:242] %>%
  select(c(
    # General Identifier
    NAME, # State Name

    # Population and overall earnings
    S2001_C03_002E, # Male median earnings 16+
    S2001_C05_002E, # Female median earnings 16+
    S2001_C03_013E, # Male full-time median earnings
    S2001_C05_013E, # Female full-time median earnings
    S2001_C03_015E, # Male median earnings 25+
    S2001_C05_015E, # Female median earnings 25+

    # Male earnings by education
    S2001_C03_016E, # Less than high school diploma
    S2001_C03_017E, # High school diploma
    S2001_C03_018E, # Some college or associates degree
    S2001_C03_019E, # Bachelors degree
    S2001_C03_020E, # Masters Degree

    # Female earnings by education
    S2001_C05_016E, # Less than high school diploma
    S2001_C05_017E, # High school diploma
    S2001_C05_018E, # Some college or associates degree
    S2001_C05_019E, # Bachelors degree
    S2001_C05_020E # Masters Degree

  )) %>%
  rename(
    State = NAME,
    m16 = S2001_C03_002E,
    f16 = S2001_C05_002E,
    mft = S2001_C03_013E,
    fft = S2001_C05_013E,
    m25 = S2001_C03_015E,
    f25 = S2001_C05_015E,
    flt = S2001_C03_016E,
    fhs = S2001_C05_017E,
    fas = S2001_C03_018E,
    fbs = S2001_C05_019E,
    fms = S2001_C03_020E,

```

```

mlt = S2001_C03_016E,
mhs = S2001_C03_017E,
mas = S2001_C03_018E,
mbs = S2001_C03_019E,
mms = S2001_C03_020E
)
# Convert numeric data to useful types
earnings[,2:17] <- lapply(earnings[,2:17], as.numeric)
head(earnings, 3)

```

```

## # A tibble: 3 x 17
##   State    m16    f16    mft    fft    m25    f25    mlt    mhs    mas    mbs    mms    flt
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Alaba~ 36576 24640 48414 35751 42319 28713 25734 34851 41996 63785 81597 15463
## 2 Alaska 45195 31929 60481 48900 53670 37676 30014 41992 53070 69624 89173 21379
## 3 Arizo~ 35820 27484 47584 39967 41672 32074 25199 32373 41541 62953 82617 17192
## # ... with 4 more variables: fhs <dbl>, fas <dbl>, fbs <dbl>, fms <dbl>

```

From this sample of the data we notice the column variables are not descriptive, they are coded. Other than the state name, there is *m16* which represents the median values for males aged 16 and older. Similarly, *f16* represents the same for age females. The variable *mft* represents the males that work full time while *mlt*, *mhs*, *mas*, *mbs*, and *mms* each represent median earnings of the education levels of individuals with less than a high school, high school diplomas, some college or associates degrees, bachelors degrees, masters or graduate degrees respectively. The same categories apply but for females where the variable name starts with an 'f'.

There are 23 variables and 52 observations in this data set, one for each state, the District of Columbia and Puerto Rico. Each case is the median earnings of individuals listed in their annual form for the year 2018.

Due to longstanding practice, data representation for individuals who do not fit within traditional gender rules, is lacking. This is not to say that this analysis intends to exclude anyone from the analysis, only that the data to perform it is absent. Furthermore, including these factors would only serve to improve the results and their applicability to the real-world as there is a proportion of the population that such data would serve better than that of typical data sets such as this one.

In addition to the earnings data extracted from the ACS survey, we will also extract some state-specific data for later use. To perform certain tests, the data will need enough cases to make inferences from. With the calculations of medians for each state already cleaned and organized we will need to reverse some of the process while maintaining the integrity of those same median calculations. Our goal here being, to supply a data set with observations of differences their gender and education, categories as well as their associated proportions. Notes are made for guidance in the process of extracting these additional variables.

```

# Further data extraction
income_2018_bystate <- income_2018_bystate[2:52,] %>%
  select(c(
    # General Reference Variables
    GEO_ID, # Unique Geographic Identifier of state
    NAME, # State Name
    S2001_C01_001E, # Total population 16+ MOE
    S2001_C01_001M, # Total population 16+

    # Mean Earnings for FULL-TIME Workers (age 16+)
    S2001_C03_002E,
    S2001_C03_002M, # Median Female Earnings
  ))

```

```

S2001_C05_002E,
S2001_C05_002M, # Median

# Male Variables
S2001_C03_016E,
S2001_C03_016M,
S2001_C03_017E,
S2001_C03_017M,
S2001_C03_018E,
S2001_C03_018M,
S2001_C03_019E,
S2001_C03_019M,
S2001_C03_020E,
S2001_C03_020M,

# Female Variables
S2001_C05_016E,
S2001_C05_016M,
S2001_C05_017E,
S2001_C05_017M,
S2001_C05_018E,
S2001_C05_018M,
S2001_C05_019E,
S2001_C05_019M,
S2001_C05_020E,
S2001_C05_020M
)) %>%

rename(
# Location Information
  GEO_ID = GEO_ID,
  State = NAME,

# General Statistics on Male and Female Earnings aged 16+
  TotalPop = S2001_C01_001E,
  TotalPop_moe = S2001_C01_001M,
  M_Earnings = S2001_C03_002E, # Female
  M_Earnings_moe = S2001_C03_002M,
  F_Earnings = S2001_C05_002E,
  F_Earnings_moe = S2001_C05_002M,

# Male Earnings by Level of Education (aged 25+)
  M_LTHS = S2001_C03_016E,
  M_HS = S2001_C03_017E,
  M_AS = S2001_C03_018E,
  M_BS = S2001_C03_019E,
  M_MS = S2001_C03_020E,

# Male Margins of Error by Education
  M_LTHS_moe = S2001_C03_016M,
  M_HS_moe = S2001_C03_017M,
  M_AS_moe = S2001_C03_018M,
  M_BS_moe = S2001_C03_019M,
  M_MS_moe = S2001_C03_020M,

```

```

# Female Earnings by Level of Education (aged 25+)
F_LTHS = S2001_C05_016E,
F_HS = S2001_C05_017E,
F_AS = S2001_C05_018E,
F_BS = S2001_C05_019E,
F_MS = S2001_C05_020E,

# Female Margins of Error by Education
F_LTHS_moe = S2001_C05_016M,
F_HS_moe = S2001_C05_017M,
F_AS_moe = S2001_C05_018M,
F_BS_moe = S2001_C05_019M,
F_MS_moe = S2001_C05_020M,
)

# Converting data types
income_2018_bystate[,3:28] <- lapply(income_2018_bystate[,3:28], as.numeric)
income_2018_bystate$GEO_ID <- as.factor(income_2018_bystate$GEO_ID)
income_2018_bystate <- as.data.frame(income_2018_bystate)

# Checking for missing values - there should be none
sum(is.na(income_2018_bystate))

```

```
## [1] 0
```

```

# Creating subsets of the data to isolate variables of interest

# excluding geo_id and moe for summary purposes
pop_income_2018 <- income_2018_bystate[,c(2,3,5,7)]
male_income_2018 <- income_2018_bystate[, c(2,seq(9, 18, 2))]
female_income_2018 <- income_2018_bystate[, c(2,seq(19, 28, 2))]

# These male and female stats can also be recombined
mf_income_2018 <- cbind(female_income_2018, male_income_2018[2:6])

# Calculate observed differences for the entire study
pop_obs <- pop_income_2018 %>%
  mutate(Obs_diff = M_Earnings - F_Earnings) %>%
  mutate(Pmf = F_Earnings/M_Earnings)

# Create a table with the highs and lows of states
# Alternatively tail could be used:
# min(tail(sort(pop_obs$Obs_diff),5))
top5 <- pop_obs %>%
  filter(Obs_diff >= min(head(sort(pop_obs$Obs_diff, decreasing=TRUE), 5)))
top5$Gap <- as.factor("Wide")
low5 <- pop_obs %>%
  filter(Obs_diff <= max(head(sort(pop_obs$Obs_diff, decreasing=FALSE), 5)))
low5$Gap <- as.factor("Narrow")
hilo_obs <- rbind(top5,low5)

# Calculating the observed differences of sex
mf_obs <- mf_income_2018 %>%
  mutate(obs_diff_lths = M_LTHS - F_LTHS) %>%

```

```
mutate(obs_diff_hs = M_HS - F_HS) %>%
mutate(obs_diff_as = M_AS - F_AS) %>%
mutate(obs_diff_bs = M_BS - F_BS) %>%
mutate(obs_diff_ms = M_MS - F_MS)

# Give each variable its own row in education
mf_stateobs <- melt(mf_obs)
```

Using State as id variables

```
mf_stateobs[4] <- as.numeric(seq(1:51))
mf_stateobs <- mf_stateobs %>%
  rename(Category = variable,
         Observation = value,
         ID = V4)
mf_state_obsdiffs <- melt(mf_obs[,12:16])
```

No id variables; using all as measure variables

```
mf_state_obsdiffs <- mf_state_obsdiffs %>%
  rename(Observation = variable,
         Difference = value)
mf_earnings_byedu <- melt(mf_obs[,2:11])
```

No id variables; using all as measure variables

```
# summarizing the education earning using means
mf_earnings_byedu <- mf_earnings_byedu %>%
  group_by(variable) %>%
  summarise(AvgEarning = mean(value)) %>%
  rename(Education = variable)
```

'summarise()' ungrouping output (override with '.groups' argument)

```
# Add in the variable of sex for later comparisons
mf_earnings_byedu$Sex <- c("Female", "Female", "Female", "Female", "Female", "Male", "Male", "Male", "Male", "Male")
# Rename the education observations to be descriptive of
# the entire data set and remove the male/female bounds of edu
mf_earnings_byedu$Education <-
  c("LTHS", "HS", "AS", "BS", "MS", "LTHS", "HS", "AS", "BS", "MS")
# Variables chosen to describe the data are not mutually exclusive
# For example 'BS' is not just for Bachelors of Science.
# That category includes all those individuals that
# attained a bachelors degree on the ACS in 2018
# and as another example 'AS' contains those with 'some college'
# from the ACS in 2018. These variables were only used for ease
# in describing the variables visually
```

This shortened table displays the first 10 statistics used in this study and the associated margins of error provided by the ACS for their calculations of the median and other relevant statistics in the year 2018.

These provide a general reference to those interested in determining the exact values of the medians and other statistics used throughout this analysis. They are used to consider which estimates are the safest in this work and help to ensure quality data is used in the figures but they are for reference only and are not included in the calculations made.

```
head(income_2018_bystate,10)
```

```
##      GEO_ID      State TotalPop TotalPop_moe M_Earnings
## 1  0400000US01      Alabama  2319177      7576      36576
## 2  0400000US02      Alaska   421213      1781      45195
## 3  0400000US04      Arizona  3396671      9044      35820
## 4  0400000US05      Arkansas 1434189      6303      33629
## 5  0400000US06      California 20152025  23699      40306
## 6  0400000US08      Colorado  3165808      6448      42051
## 7  0400000US09      Connecticut 2002828      5389      49879
## 8  0400000US10      Delaware   501672      2547      41018
## 9  0400000US11 District of Columbia  414178      2345      58864
## 10 0400000US12      Florida 10163248  19868      33644
##      M_Earnings_moe F_Earnings F_Earnings_moe M_LTHS M_LTHS_moe M_HS M_HS_moe
## 1      283      24640      230  25734      367 34851      434
## 2      881      31929      367  30014      2036 41992      840
## 3      184      27484      210  25199      323 32373      229
## 4      386      24899      263  26517      446 32858      490
## 5      102      30038      69  26283      104 35495      157
## 6      188      30792      143  31001      352 38125      491
## 7      405      34276      437  29215      1149 42667      594
## 8      358      31339      345  30155      1165 39066      1241
## 9     1340      49455     1011  27462     2263 31556      712
## 10     208      26728      82  24303      369 30959      129
##      M_AS M_AS_moe M_BS M_BS_moe M_MS M_MS_moe F_LTHS F_LTHS_moe F_HS
## 1  41996      367 63785      685  81597      1630  15463      387 21685
## 2  53070     1095 69624     1820  89173      3004  21379     1036 27865
## 3  41541      280 62953      569  82617      1050  17192      290 24957
## 4  39368      828 57730     1579  75169      1863  18002      769 22203
## 5  45717      316 73266      298 104547      429  17025      108 25295
## 6  44678      559 64896      643  86547      1214  18724      630 26728
## 7  51809      555 78938      925 103717     1216  19955      675 28918
## 8  44162     1382 64840     1901  84001     2563  20863      846 26725
## 9  41135     1181 75454     2016 104310     2011  21787     1120 29382
## 10 37693      381 54868      355  77608      864  16983      209 22627
##      F_HS_moe F_AS F_AS_moe F_BS F_BS_moe F_MS F_MS_moe
## 1      220 26815      285 40734      381 52630      351
## 2     1110 35356      932 47447     1434 67889     1560
## 3      445 31071      193 42530      356 53640      480
## 4      271 26761      312 40675      510 52641      745
## 5      131 32015      117 52101      202 72811      335
## 6      289 31448      258 43372      530 55820      730
## 7      627 36044      472 52209      617 69570      830
## 8      679 32435      923 49886     1063 64058     1150
## 9     1461 36795     2168 60319     1124 84124     1397
## 10     170 29475      196 40789      160 52321      356
```

Our final data frames with which the majority of the analysis will be used are the *earnings* and *income_2018_bystate*. From each of these we can evaluate the median earnings of education by state but

they differ in a few details.

In the *income_2018_bystate* the earnings variables are medians but include all the wage earners of the population. Therefore, it also lists the total population of each state, median earnings by gender and education by those same states, as well as the margins of error for each relevant statistic provided by the ACS. The geographic identifier of each state and the District of Columbia was also included as an option for visualizing the data further into the analysis. We extract from this basic data set the remaining subsets of the data frame that as necessary.

```
# Here we review the data as a whole for ACS. Since all states have
# larger incomes for males we can find the proportion of a females
# income to that of a males across the states. The mean dollar amounts
# are listed in the "Obs_diff" column.
kable(head(pop_obs[,c(1,5,6)], 5))
```

State	Obs_diff	Pmf
Alabama	11936	0.6736658
Alaska	13266	0.7064720
Arizona	8336	0.7672808
Arkansas	8730	0.7404026
California	10268	0.7452488

The *earnings* data frame lists the minimum necessary to perform the gender wage gap analysis. A sample is provided in the table below where the first 3 states are listed with their data. The unique features of this data frame are the inclusion of Puerto Rico as a state (although it is not formally) and the different median annual earnings variables such as that of male and female full-time and earnings of those aged 25+ instead of only 16 and older.

```
head(earnings, 3)
```

```
## # A tibble: 3 x 17
##   State    m16    f16    mft    fft    m25    f25    mlt    mhs    mas    mbs    mms    flt
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Alaba~ 36576 24640 48414 35751 42319 28713 25734 34851 41996 63785 81597 15463
## 2 Alaska 45195 31929 60481 48900 53670 37676 30014 41992 53070 69624 89173 21379
## 3 Arizo~ 35820 27484 47584 39967 41672 32074 25199 32373 41541 62953 82617 17192
## # ... with 4 more variables: fhs <dbl>, fas <dbl>, fbs <dbl>, fms <dbl>
```

From the data frame we calculate multiple new variables such as the observed difference in wages by gender, the proportion of median female earnings to median male earnings in categories consistent with the literature.

Analysis

First, we determine the presence or absence of a difference in the entire sample from the ACS. This provides a baseline set of information to build upon. We can then compute the observed differences between males and females average earnings in each state to examine the patterns among the states. With that, we can show the presence of a difference in earnings of male and female Americans in 2018 that has been observed in plethora of peer-reviewed studies.

As mentioned, there are several ways to approach this. We have the median annual earnings from those who are 16 years of age and older, 25 years and older, and those who are considered full-time workers with any income within the year. Since the literature often cites full-time workers earnings weekly or annually, we will follow their example.

```
pfm.ft <- earnings %>%
  summarise(Pfm_ft = mean(fft)/mean(mft))
pfm.ft <- round(pfm.ft, digits = 4)
pfm.ft
```

```
## # A tibble: 1 x 1
##   Pfm_ft
##   <dbl>
## 1  0.797
```

The proportion of the median wages of full-time female earners to full-time male earners in 2018 is 0.7966. Importantly, this statistic is considered *raw* by the standards of the literature due to its unadjusted features in the calculation. All variables that may influence the calculation of this proportion are included, regardless of consideration to the applicability of the statistic to the terms “equal pay for equal work”. In this analysis, unequal worker pay and majority-gender occupations and industries are considered because it is necessary to understand the scope of differences between men and women in this study.

Differences in the median earnings of each gender category are were calculated alongside the proportions of females and males. The proportion is found by observing how the female median earnings compare to the male median earnings. This is because the median male earnings were always larger than the median female earnings for the categories. The only exception to this rule is in Puerto Rico where the median earnings of women who work full time is slightly higher than median male earnings by 2.5540776 percent.

```
earnings <- earnings %>%
  mutate(dif.ft = mft - fft) %>%
  mutate(dif.16 = m16 - f16) %>%
  mutate(dif.25 = m25 - f25) %>%
  mutate(Pft = fft/mft) %>%
  mutate(P16 = f16/m16) %>%
  mutate(P25 = f25/m25)
max.pft <- summary(earnings$Pft)[6]
```

Given the other factors that are part of this exception, we will omit Puerto Rico from this analysis of the states. If the statehood status were to change and these values were to be included into the results, the data would shift towards a less realistic expectation of the experiences of individuals across the 50 states and District of Columbia. Summary statistics for each of the categories common to the states are displayed alongside with Puerto Rico separately.

```
earnings[which.max(earnings$Pft),]
```

```
## # A tibble: 1 x 23
##   State    m16    f16    mft    fft    m25    f25    mlt    mhs    mas    mbs    mms    flt
##   <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Puert~ 17136 17431 23531 24132 18528 18971 11534 15728 19033 30079 44335 9416
## # ... with 10 more variables: fhs <dbl>, fas <dbl>, fbs <dbl>, fms <dbl>,
## #   dif.ft <dbl>, dif.16 <dbl>, dif.25 <dbl>, Pft <dbl>, P16 <dbl>, P25 <dbl>
```

```
earnings <- earnings[1:51,]
earnings.summary <- summary(earnings[2:7])
education.summary <- summary(earnings[8:17])
kable(earnings.summary)
```

	m16	f16	mft	fft	m25	f25
	Min. :32127	Min. :21540	Min. :42245	Min. :32928	Min. :38295	Min. :26111
	1st Qu.:36568	1st Qu.:25635	1st Qu.:47436	1st Qu.:37067	1st Qu.:41821	1st Qu.:30202
	Median :39570	Median :26936	Median :51477	Median :39967	Median :45873	Median :31575
	Mean :40216	Mean :28430	Mean :52437	Mean :41668	Mean :46761	Mean :33226
	3rd Qu.:42424	3rd Qu.:30777	3rd Qu.:55145	3rd Qu.:45570	3rd Qu.:50472	3rd Qu.:36033
	Max. :58864	Max. :49455	Max. :77498	Max. :67692	Max. :67715	Max. :57420

Full time earnings of males and females in the 50 states and District of Columbia have a median of \$39,570 and \$26,936 for those 16 and older, \$51,477 and \$39,967 for those employed full-time with earnings in the past 12 months, along with \$45,873 and \$39,967 for those age 25 and older. Taking the difference of these medians and proportioning them create a new variable called the observed difference and proportion of median female earnings to male earnings. In the table these are abbreviated to *obs_dif* and *P'x'* where 'x' can take the place of the variable full-time ('ft'), the number 16 for those with earnings aged 16 and older, or 25 for those with earnings that are age 25 and older in these calculations.

```
kable(earnings[,c(1,18:23)])
```

State	dif.ft	dif.16	dif.25	Pft	P16	P25
Alabama	12663	11936	13606	0.7384434	0.6736658	0.6784896
Alaska	11581	13266	15994	0.8085184	0.7064720	0.7019937
Arizona	7617	8336	9598	0.8399252	0.7672808	0.7696775
Arkansas	7894	8730	10798	0.8131376	0.7404026	0.7236455
California	6554	10268	11711	0.8787599	0.7452488	0.7501707
Colorado	9419	11259	14177	0.8274561	0.7322537	0.7174095
Connecticut	11848	15603	16583	0.8192690	0.6871830	0.7118506
Delaware	8081	9679	10704	0.8490943	0.7640304	0.7735322
District of Columbia	9806	9409	10295	0.8734677	0.8401570	0.8479657
Florida	6032	6916	7952	0.8607089	0.7944359	0.7923489
Georgia	8929	9968	11097	0.8157982	0.7322086	0.7399466
Hawaii	9460	9283	11030	0.8166311	0.7751211	0.7658423
Idaho	11888	13340	15117	0.7461511	0.6175459	0.6333317
Illinois	12096	12776	16069	0.7842812	0.6986579	0.6815056
Indiana	13195	13680	14646	0.7390436	0.6532056	0.6751181
Iowa	11543	13650	14834	0.7733244	0.6631459	0.6810921
Kansas	11602	12938	14680	0.7700388	0.6700921	0.6806682
Kentucky	9917	10147	11435	0.7876173	0.7142093	0.7235185
Louisiana	16212	15320	17372	0.6850632	0.6198983	0.6213023
Maine	9618	9827	10835	0.8069334	0.7329257	0.7418394
Maryland	9576	12468	13120	0.8485745	0.7496134	0.7676354
Massachusetts	11888	14931	16313	0.8207749	0.7012725	0.7195200
Michigan	12283	13150	15399	0.7679388	0.6601452	0.6659001
Minnesota	10125	11918	14012	0.8186589	0.7248147	0.7252010
Mississippi	10125	9931	12201	0.7648248	0.7081178	0.6920028
Missouri	10921	10496	11662	0.7749686	0.7175913	0.7274024
Montana	11622	9615	12639	0.7542242	0.7158940	0.6891387
Nebraska	11077	12703	13858	0.7766869	0.6795328	0.6956760
Nevada	7935	7341	8967	0.8301657	0.7992068	0.7813194
New Hampshire	13931	14431	16480	0.7662662	0.6850846	0.6909749
New Jersey	12389	16172	16955	0.8095494	0.6817287	0.7059946
New Mexico	8243	8444	10609	0.8132237	0.7371681	0.7247276
New York	6989	9995	11820	0.8763053	0.7645465	0.7646122
North Carolina	7618	8909	10155	0.8346896	0.7506019	0.7538421
North Dakota	13862	14587	17604	0.7427962	0.6678961	0.6579822
Ohio	11540	12889	14189	0.7773705	0.6742734	0.6893555
Oklahoma	11900	11645	12919	0.7465713	0.6831379	0.6935648
Oregon	9949	10674	11781	0.8074809	0.7141250	0.7261188
Pennsylvania	11521	12815	15032	0.7855002	0.6913015	0.6887011
Rhode Island	9060	11165	13595	0.8359795	0.7336975	0.7307441
South Carolina	9724	9416	11350	0.7872629	0.7302006	0.7228733
South Dakota	10661	10857	11923	0.7700735	0.7098998	0.7189099
Tennessee	8780	9532	10805	0.8076291	0.7325177	0.7368677
Texas	10136	12094	13608	0.7974056	0.6889483	0.6977746
Utah	16170	18369	22433	0.6982252	0.5476396	0.5566250
Vermont	8290	8763	10189	0.8357115	0.7643468	0.7664145
Virginia	12059	13365	15821	0.7932873	0.7006652	0.7001327
Washington	13843	15617	17588	0.7728421	0.6625540	0.6699383
West Virginia	14019	12147	14235	0.7073461	0.6630233	0.6559184
Wisconsin	11021	13062	14274	0.7876248	0.6805107	0.6970327
Wyoming	16042	17298	20228	0.7086081	0.5979453	0.5984436

Observations made here display a trend of decreasing proportions of median female to male earnings as we change the variables from those with full-time employment to those who may not be employed full time. Between the ages of 16 and 25, the likelihood of being employed full-time is lower than for those who are 25 and older. However, there appears to be enough individuals not employed full-time in the 25 and up age group to cause the proportion of 25 and older to drop lower than for the category of full-time wage earners. We could speculate on a number of factors that could cause this but for now, we should only recognize this trend as it pertains the analysis of education across the gender wage gap.

The same process is used when computing the observed differences, proportions, and mean of proportions in the median earnings of educational attainment of individuals in this study. We organize the categories using abbreviations of science degrees for simplicity in visualizing. They are in no way exclusive and do not fully capture the depth of the data collected by the ACS. The ACS includes all educational attainment in the categories of *less than high school diploma*, *high school diploma*, *some college or an associates degree*, *bachelors degree*, and *graduate or professional degree*. We abbreviate these as LTHS, HS, AS, BS, and MS respectively.

```
male_earnings_byedu <- mf_earnings_byedu %>%
  filter(Sex == "Male")
female_earnings_byedu <- mf_earnings_byedu %>%
  filter(Sex == "Female")
mf_earn_ed <- cbind(male_earnings_byedu, female_earnings_byedu)
mf_earn_ed <- mf_earn_ed[-4] %>%
  rename(Median_M = AvgEarning,
         Median_F = AvgEarning.1,
         Sex.M = Sex,
         Sex.F = Sex.1) %>%
  mutate(Pfm = Median_F / Median_M) %>%
  mutate(obs_dif = Median_M - Median_F) %>%
  mutate(Mean_Pfm = mean(Pfm))
mf_earn_ed
```

##	Education	Median_M	Sex.M	Median_F	Sex.F	Pfm	obs_dif	Mean_Pfm
## 1	LTHS	28299.22	Male	17934.33	Female	0.6337396	10364.88	0.6723679
## 2	HS	37024.82	Male	24472.02	Female	0.6609625	12552.80	0.6723679
## 3	AS	44565.33	Male	30131.22	Female	0.6761133	14434.12	0.6723679
## 4	BS	63340.41	Male	43538.06	Female	0.6873662	19802.35	0.6723679
## 5	MS	83168.78	Male	58522.35	Female	0.7036577	24646.43	0.6723679

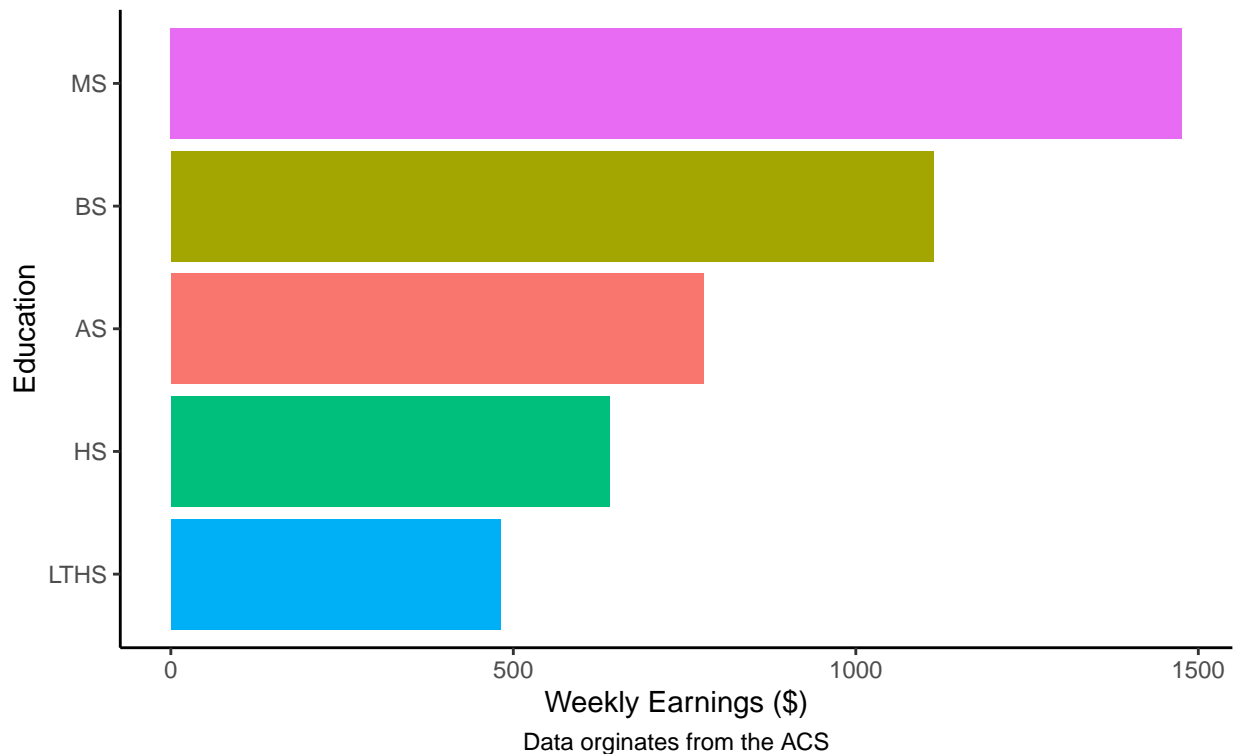
Through this table, it is clear that education increases the mean earnings of each gender and that the proportion of mean female to male earnings also increases with it. This can be read as the values where the proportion of mean female to male earnings changes from \$10,364.88 with less than a high school diploma to \$24,646.43 with a graduate degree. This is a decrease in the wage gap by \$14,281.55 percentage points.

Another pattern in this data show that as education attainment level increases, so too does the observed difference in dollars. That is, as the earnings of both genders increase with higher levels of education, the dollar amount of earnings that a male earns starts and increases at a greater rate than females. We can also visualize this in few ways.

This chart displays how a higher levels of education attainment results in higher median earnings. In this case, the earned wages are displayed as weekly earnings starting with zero and extending to nearly \$1,500.00 for the higher educational attainment category listed in this analysis, the master's degree. Importantly, this data takes the mean of the median average earnings to calculate its weekly earnings, making it slightly different than the commonly discrete median or mean selected in literature.

```
mf_earnings_byedu %>%
  group_by(Education) %>%
  mutate(Mean_Earnings = mean(AvgEarning)) %>%
  filter(Sex == "Male") %>%
  mutate(AvgEarn_wkdollars = Mean_Earnings / 48) %>%
  ggplot( aes(x = reorder(Education, AvgEarn_wkdollars), y = AvgEarn_wkdollars)) +
  geom_col(aes(fill = Education)) +
  labs(x = "Education",
       y = "Weekly Earnings ($)",
       title = "Average Income by Education in the U.S.",
       subtitle = "Using the Mean of Individual Earnings in 2018",
       caption = "Data originates from the ACS") +
  theme_classic() +
  coord_flip() +
  theme(legend.position = "none",
        plot.caption = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

Average Income by Education in the U.S.
Using the Mean of Individual Earnings in 2018



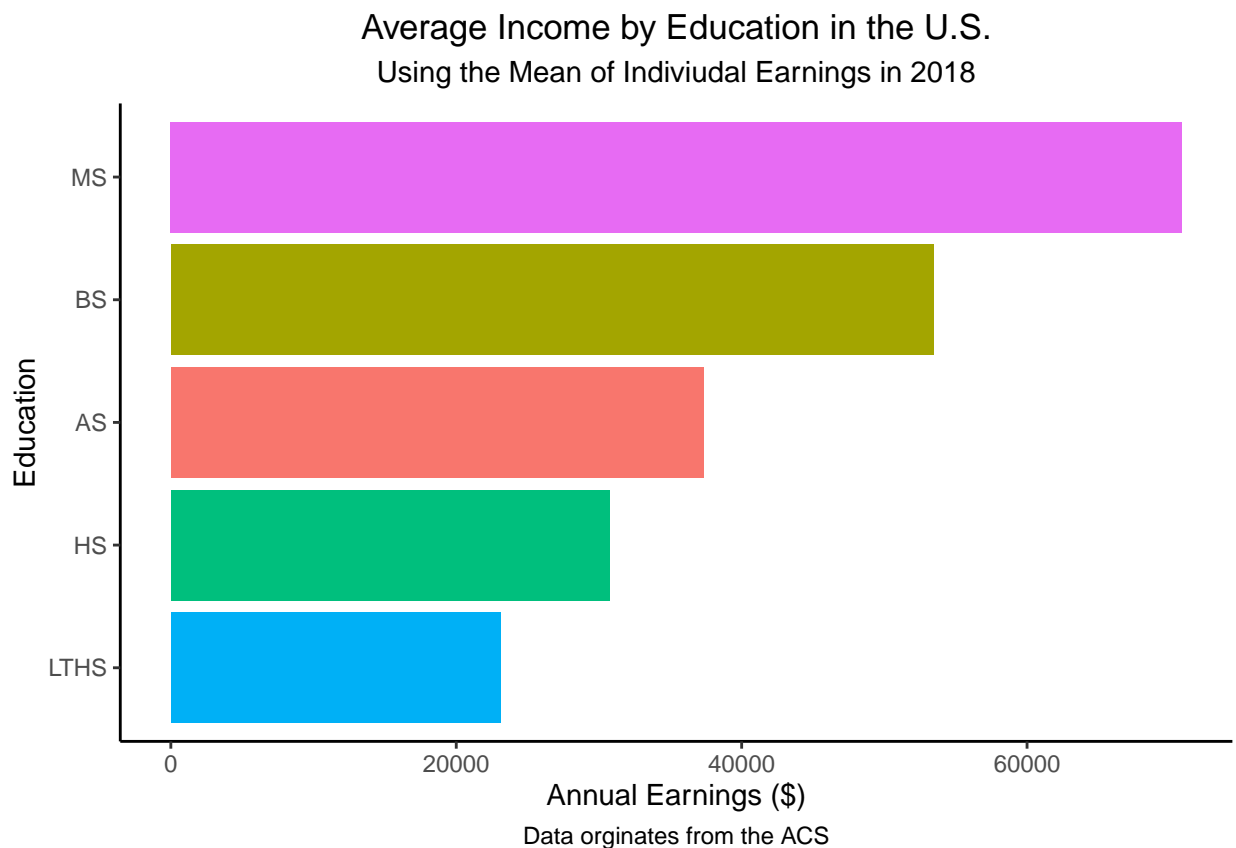
Information from the chart can also be displayed as annual earnings. To make things easier, that conversion has already been completed in the following chart. Note that these annual earnings also use the mean of the median. Results will vary slightly from commonly accepted values.

```
mf_earnings_byedu %>%
  group_by(Education) %>%
  mutate(Mean_Earnings = mean(AvgEarning)) %>%
```

```

filter(Sex == "Male") %>%
ggplot( aes(x = reorder(Education, Mean_Earnings), y = Mean_Earnings)) +
geom_col(aes(fill = Education)) +
labs(x = "Education",
     y = "Annual Earnings ($)",
     title = "Average Income by Education in the U.S.",
     subtitle = "Using the Mean of Individual Earnings in 2018",
     caption = "Data originates from the ACS") +
theme_classic() +
coord_flip() +
theme(legend.position = "none",
      plot.caption = element_text(hjust = 0.5),
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))

```



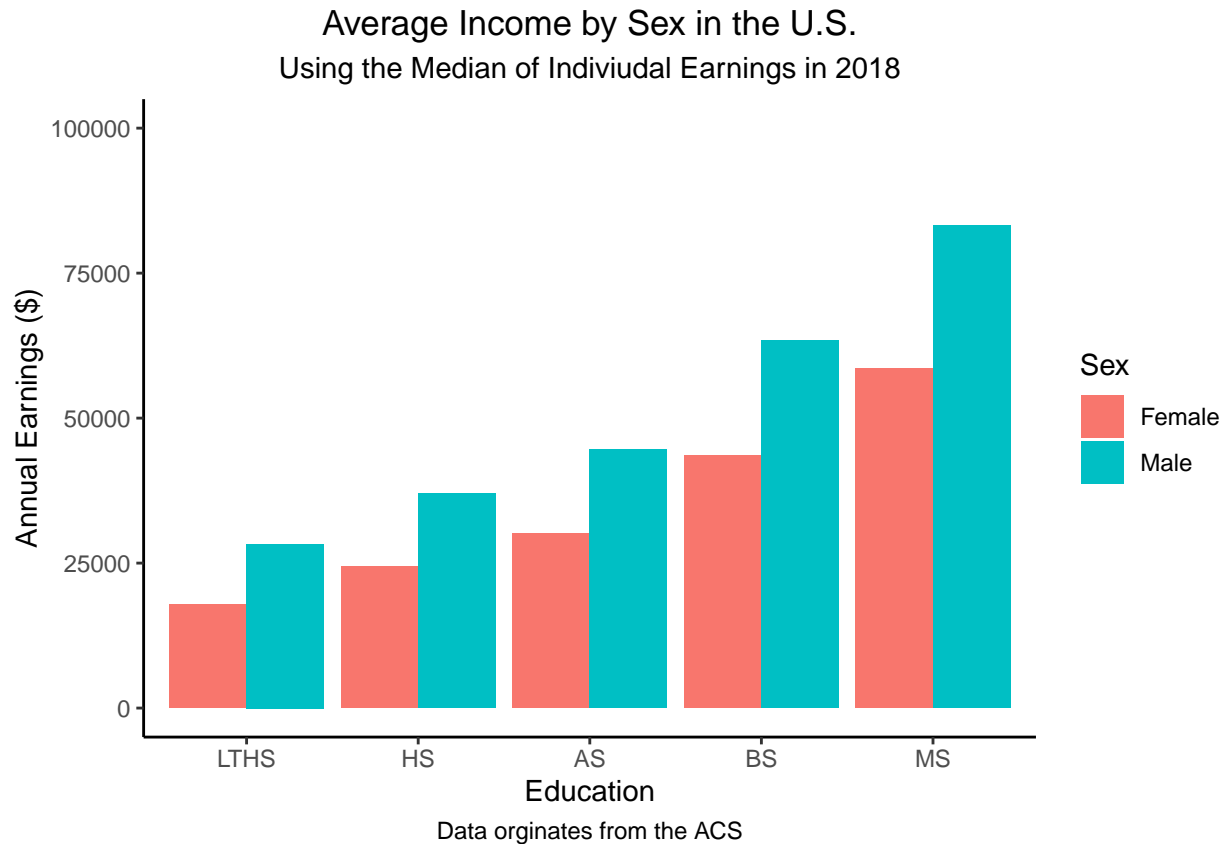
Another useful comparison of the median earnings of gender by each level of educational attainment is through this column chart. The same categories are used for the levels of education and median male and female annual earnings are reported for all individuals with any earnings in 2018.

```

ggplot(mf_earnings_byedu, aes(x = reorder(Education, AvgEarning), y = AvgEarning, fill = Sex)) +
geom_col(position = "dodge") +
ylim(0,100000) +
labs(x = "Education",
     y = "Annual Earnings ($)",
     title = "Average Income by Sex in the U.S.",
     subtitle = "Using the Median of Individual Earnings in 2018",

```

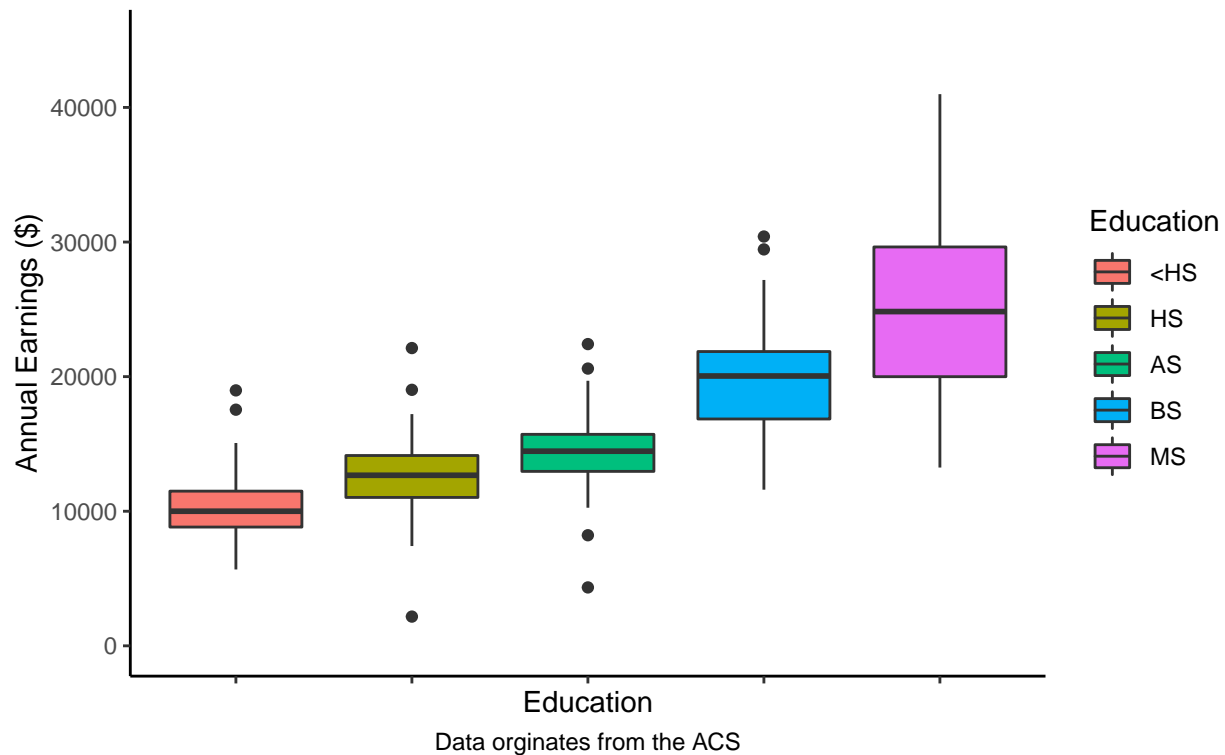
```
caption = "Data originates from the ACS") +
theme_classic() +
theme(plot.caption = element_text(hjust = 0.5),
      plot.title = element_text(hjust = 0.5),
      plot.subtitle = element_text(hjust = 0.5))
```



From this column chart we observe that in every mean of every education type, the male earnings bar in blue is taller than the bars of female earnings in red. Since each bar represents the median annual earnings of the individuals in each category of education, this visually displays how the annual earnings of females does not meet the median male earnings at any education level. Although, the median earning of females do increase with higher levels of education. A boxplot adds some detail to this.

```
ggplot(mf_state_obsdiffs, aes(x = reorder(Observation, Difference), y = Difference)) + geom_boxplot(aes(
  ylim(0, 45000) +
  scale_fill_discrete(name = "Education", labels = c("<HS", "HS", "AS", "BS", "MS"))) +
  labs(x = "Education",
       y = "Annual Earnings ($)",
       title = "Observed Differences in Earnings",
       subtitle = "Using the Median of Individual Earnings in 2018",
       caption = "Data originates from the ACS") +
  theme_classic() +
  theme(axis.text.x = element_blank(),
        plot.caption = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```

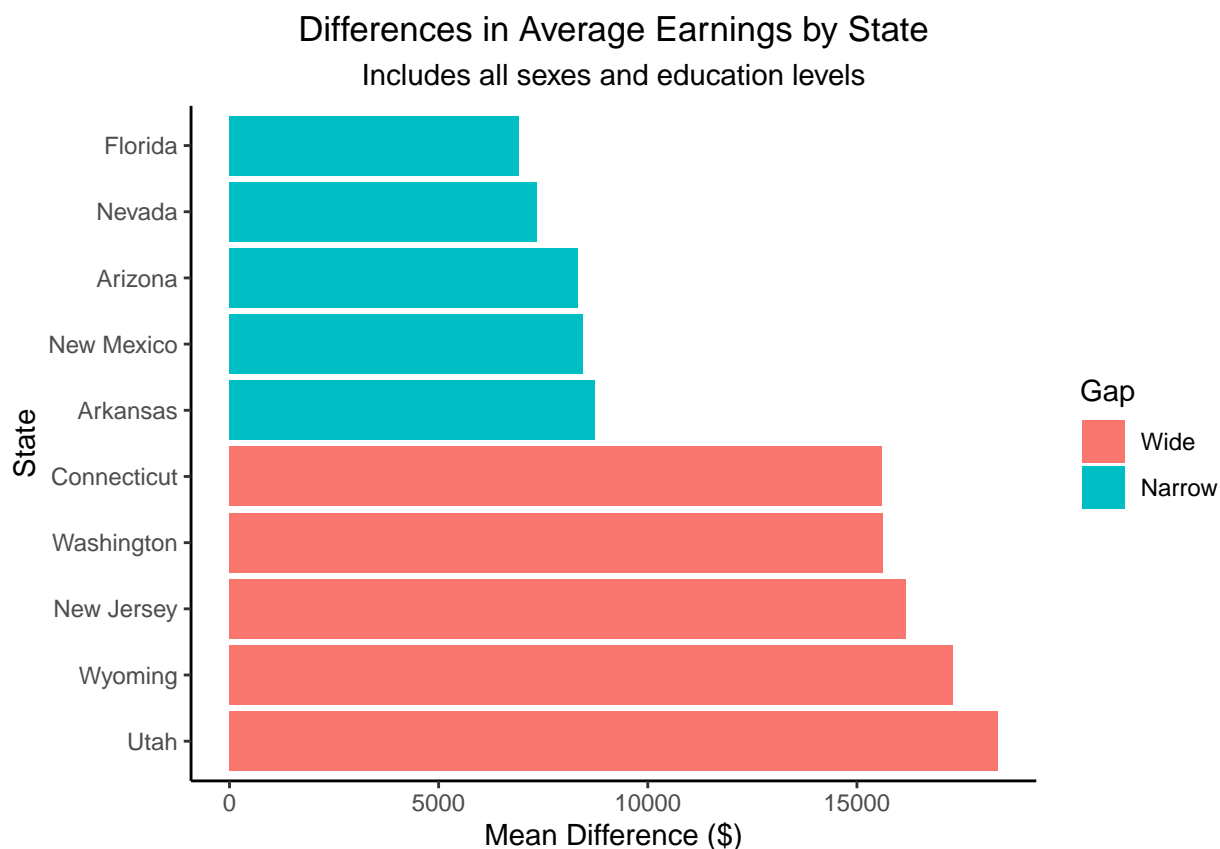

Observed Differences in Earnings Using the Median of Individual Earnings in 2018



As the level of education increases, so too do the average annual earnings of individuals. This boxplot also shows that the variation of individuals' earnings may increase with higher levels of education attainment too. Further research is needed to test these hypotheses and compare but strong trends are present.

To break the information into localized segments, the differences in individual earnings by state can be shown in the following chart. Only the states with the highest 5 and lowest 5 differences in earnings were selected for clarity. The states with the highest differences can be thought of as having a wider earnings gap. The reverse applies for the lowest where their earnings gap can be thought of as narrower relative to the rest of the states.

```
ggplot(hilo_obs, aes(x = reorder(State, -Obs_diff), y = Obs_diff)) +
  labs(x = "State",
       y = "Mean Difference ($)",
       title = "Differences in Average Earnings by State",
       subtitle = "Includes all sexes and education levels") +
  geom_col(aes(fill=Gap)) +
  coord_flip() +
  theme_classic() +
  theme(plot.caption = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



With the observational differences computed and the types of gap labeled based on the average median (or the median of the medians) we can see that there is not as much of change between those states with a wide gap and other states with wide gaps. The range of these wide gap states is \$2,766 with a maximum observed difference of \$18,369 and a minimum of \$15,603. They have a mean proportion of female to male annual earnings of 0.6354.

The same pattern applies to the states with narrowest gap having an observed maximum difference of \$8,730, minimum of \$6,916 and range of \$2,766 when compared against other states with relatively narrow differences in wage compared to other states. The mean proportion of female to male annual earnings of 0.7677 which is lower because the states themselves have a narrower gap in their earnings by gender overall.

In contrast, the difference between the gender wage gaps of narrow and wide states is greater than the range of the wide or narrow gaped states when compared within groups. The range of all 10 selected states is \$11,453 with a maximum observed difference of \$18,369 and minimum of \$6,916. This represents the full range of all states since the data frame includes the highest and lowest observational values of each state's gender gap type. A full table of these select states is shown below.

hilo_obs

##	State	TotalPop	M_Earnings	F_Earnings	Obs_diff	Pmf	Gap
## 1	Connecticut	2002828	49879	34276	15603	0.6871830	Wide
## 2	New Jersey	4789487	50812	34640	16172	0.6817287	Wide
## 3	Utah	1599204	40607	22238	18369	0.5476396	Wide
## 4	Washington	3949923	46280	30663	15617	0.6625540	Wide
## 5	Wyoming	328719	43024	25726	17298	0.5979453	Wide
## 6	Arizona	3396671	35820	27484	8336	0.7672808	Narrow
## 7	Arkansas	1434189	33629	24899	8730	0.7404026	Narrow

## 8	Florida	10163248	33644	26728	6916	0.7944359	Narrow
## 9	Nevada	1513020	36560	29219	7341	0.7992068	Narrow
## 10	New Mexico	994984	32127	23683	8444	0.7371681	Narrow

Notice that the proportion of earnings is not always higher when the observed difference in earnings is high. The same applies to states where the observed difference in earnings is low, the resultant proportion is not always lower.

For example in the state of Connecticut where the observed difference in wage is \$15,603 there is a proportional gap of 0.687. Now we observe the state of Washington with an observational difference of \$30,663. If trends were to hold steady, then the the proportional gap of Washington should show a proportion approximately equal to Connecticut or around 0.687. However, when reviewing the data we see that the trend is not as steady as originally thought. The proportional wage gap of Washington with an observational difference of \$30,663 has a proportion of 0.663 that is not approximately equal to the state of Connecticut. In fact, the state of New Jersey is the only other proportion that is nearly equal to what this observational difference should be with a proportion of 0.682 yet its observational difference is \$16,172.

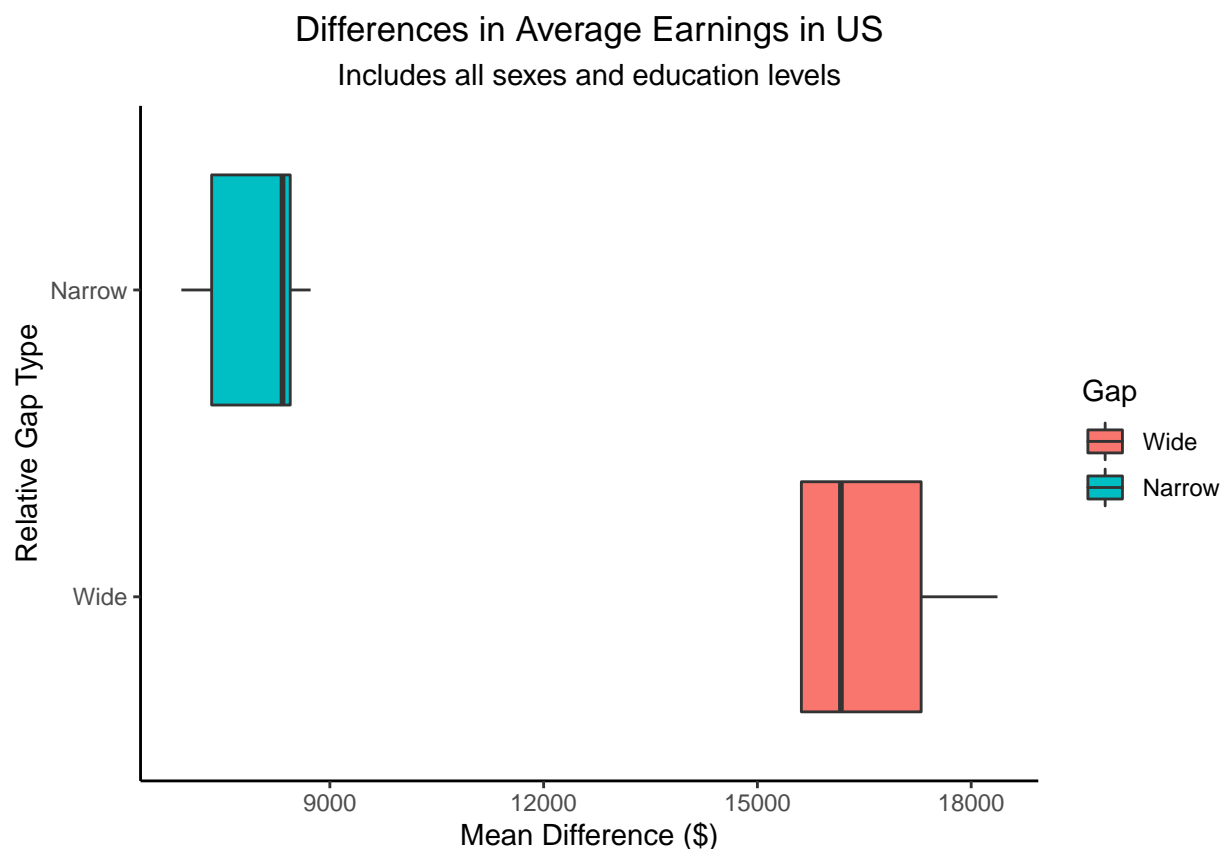
This could be due to random variation or the unique span of ranges between male and female populations of each state. Regardless, it seems to suggest that the observed difference is dollars between median male and female earning is not always an exact indicator of the gender wage gap. Given that the gender wage gap is measured by the difference in the median earnings between males and females, such a phenomenon presents a challenge in determining the significance of gender wage gaps by state.

Take another look at the gender wage gap by type. Keep in mind that both boxes represent a difference in the gender wages with their only differences stemming from their distribution and magnitude.

```
ggplot(hilo_obs, aes(x = Gap, y = Obs_diff, fill = Gap)) +

  labs(x = "Relative Gap Type",
       y = "Mean Difference ($)",
       title = "Differences in Average Earnings in US",
       subtitle = "Includes all sexes and education levels") +

  geom_boxplot() +
  theme_classic() +
  theme(plot.caption = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5)) +
  coord_flip()
```



Aside from the large gap between these two boxes, observe their distributions. States that contain relatively narrow gaps also seem to be more tightly distributed with a median located at high observational differences within in its interquartile range. On the bottom side in red, we see how wider gender wage gaps at the state level result in a looser distribution with greater upper range, little to no range below it interquartile and its median has situated itself much lower within its interquartile range. These larger differences in distribution may be stronger influencing factor on wages than previously considered.

As it stands in the boxplots, the difference in states gender wage gap exhibits a stretching of features that might be worthy of consideration for inference. The lower tail of the narrow states' distribution extends farther than it upper tail. Simultaneously, the lower tail of the wide states' distribution is shorter than its elongated upper tail. Each gap type appears as though their medians have been dragged in opposite directions. The narrow gap states whole distribution is getting pushed down while it median tried to maintain itself in one position. Again at the same time, the wider gap states whole distribution is being pulled up but the factors controlling this shift remain unknown.

Inference

The benchmark for this unadjusted full-picture gender wage gap analysis is 0.797 which suggests that women earn about 79.66 percent of the median annual earnings of men on average. With this we can infer that the gender gap in 2018 is approximately 0.203 using the median annual earnings of full-time men and women for its calculation. We can evaluate the significance of education at increasing the individuals median earnings but not the causality.

Our null hypothesis is that there is no effect on the median annual earnings of males and females in 2018 with higher education. Alternatively, would reject this null if there is an effect on median annual earnings of males and females in 2018 with higher education. We select the median observed differences of male and female variables in which the individuals education levels are categorized as having less than a high school

diploma, having a high school diploma, an associates degree, bachelors degree, or masters degree for this inference.

```
mfdif.mean <- mf_state_obsdiffs %>%  
  summarise(mean = mean(Difference))
```

The mean of the observed differences in earnings between genders in all educational categories is \$16,360.12.

Before proceeding, refer to the observed differences boxplot with median individual earnings categorized by education level from each state. These differences measured the specific amount of dollars that separated male and female earnings in 2018. It is clear from that figure that higher education does increase the average earnings of individuals as the individual attains higher levels of education. We evaluate the likelihood that this trend is random chance with an analysis of variance (anova) which will compare the differences in means across each education level.

```
mfobs.mean <- mf_stateobs[1:510,] %>%  
  summarise(mean = mean(Observation))  
mfobs.anova <- anova_test(mf_state_obsdiffs, Difference ~ Observation)
```

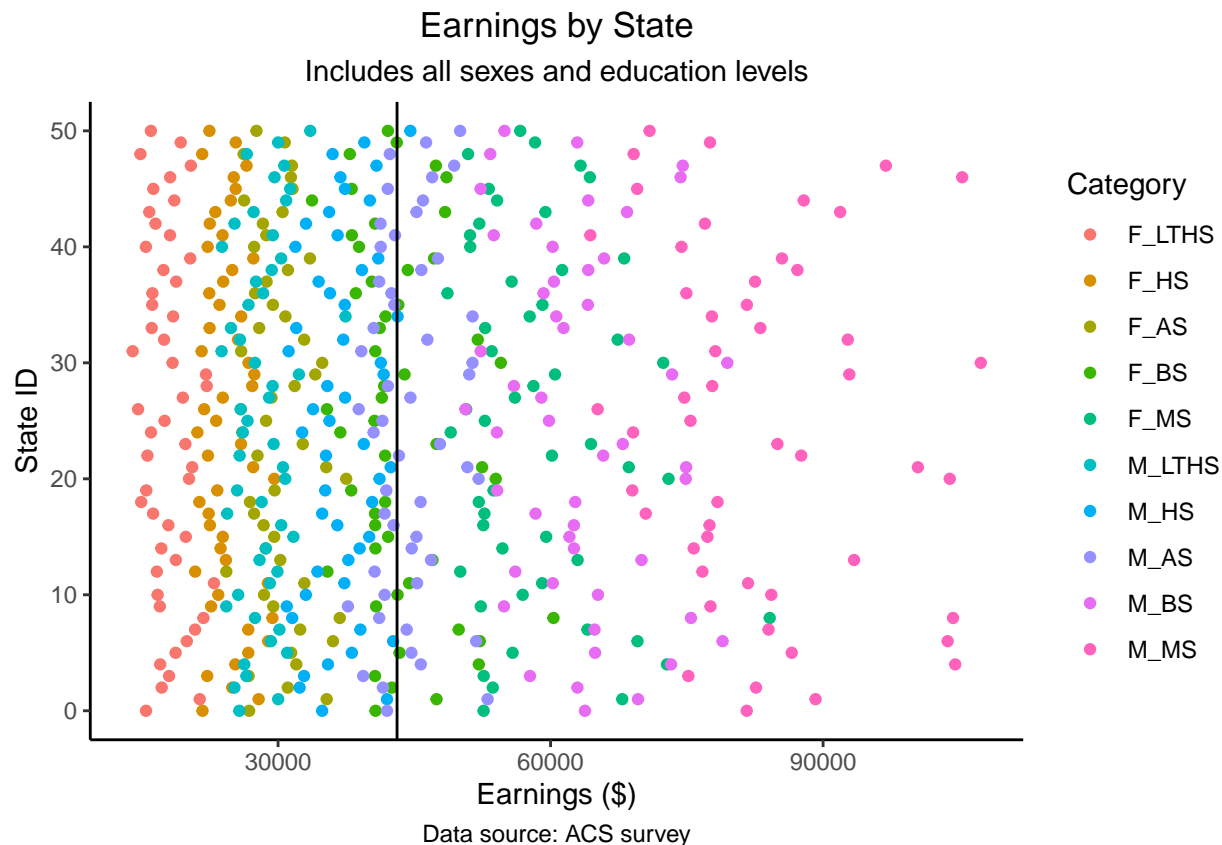
```
## Coefficient covariances computed by hccm()
```

```
mfobs.anova
```

```
## ANOVA Table (type II tests)  
##  
##           Effect DFn DFd           F          p p<.05    ges  
## 1 Observation    4 250 106.948 5.73e-53      * 0.631
```

We may infer from the results of the anova that this trend is significant at a level of 0.0001 with an p-value of 0 and f-value of 106.948. These indicate that the trend of increasing earnings with higher education is extremely unlikely to have occurred by chance. Another method of visualizing this is with the earnings in each category of male and female education levels plotted by each state.

```
ggplot(mf_stateobs[1:510,], aes(ID - 1, Observation)) + geom_point(aes(color = Category)) +  
  geom_hline(yintercept = mfobs.mean$mean) +  
  labs(x = "State ID",  
       y = "Earnings ($)",  
       title = "Earnings by State",  
       subtitle = "Includes all sexes and education levels",  
       caption = "Data source: ACS survey") +  
  coord_flip() +  
  theme_classic() +  
  theme(plot.caption = element_text(hjust = 0.5),  
        plot.title = element_text(hjust = 0.5),  
        plot.subtitle = element_text(hjust = 0.5))
```



The solid black line displays the mean earnings of all individuals in this study while the ID is a numerical representation of the state with the District of Columbia included. The median earnings of each individuals income calculated for each state is used to produce one point on the scatterplot.

Those that are above the line have above average income and tend to have higher levels of education. Perhaps most notably in this figure, we see males and females that have less than a high school diplomas sit more closely and tightly packed together with lower earnings toward the bottom of the figure. Higher earners are more educated and spread out with greater variation in their medians for each state. Here again, the trend shows that with higher education levels, higher earnings are likely.

Another analysis of variance can determine the significance of the difference in male and female earnings of each education category by state. Using the the same scatterplot, we would look to see how well the colors of the points line up in a singular bar (with upper and lower bounds) with one another. If each state had exactly the same observed differences in earnings by education then a perfectly ordered block of a single color would show for each education level. This does not appear to occur.

We can infer from the figure that there are differences in the observations of earnings for each state. We have already noted the top 5 states with the highest wage gap and the lowest 5 states with the narrowest gap. We can use the information from that analysis to determine the difference in every state. Using the observations of both genders presented in each category of this data set, we calculate the observed differences modeled by each state.

```
mfstobs.anova <- anova_test(mf_stateobs[1:510,], Observation ~ ID)
```

```
## Coefficient covariances computed by hccm()
```

```
mfstobs.anova
```

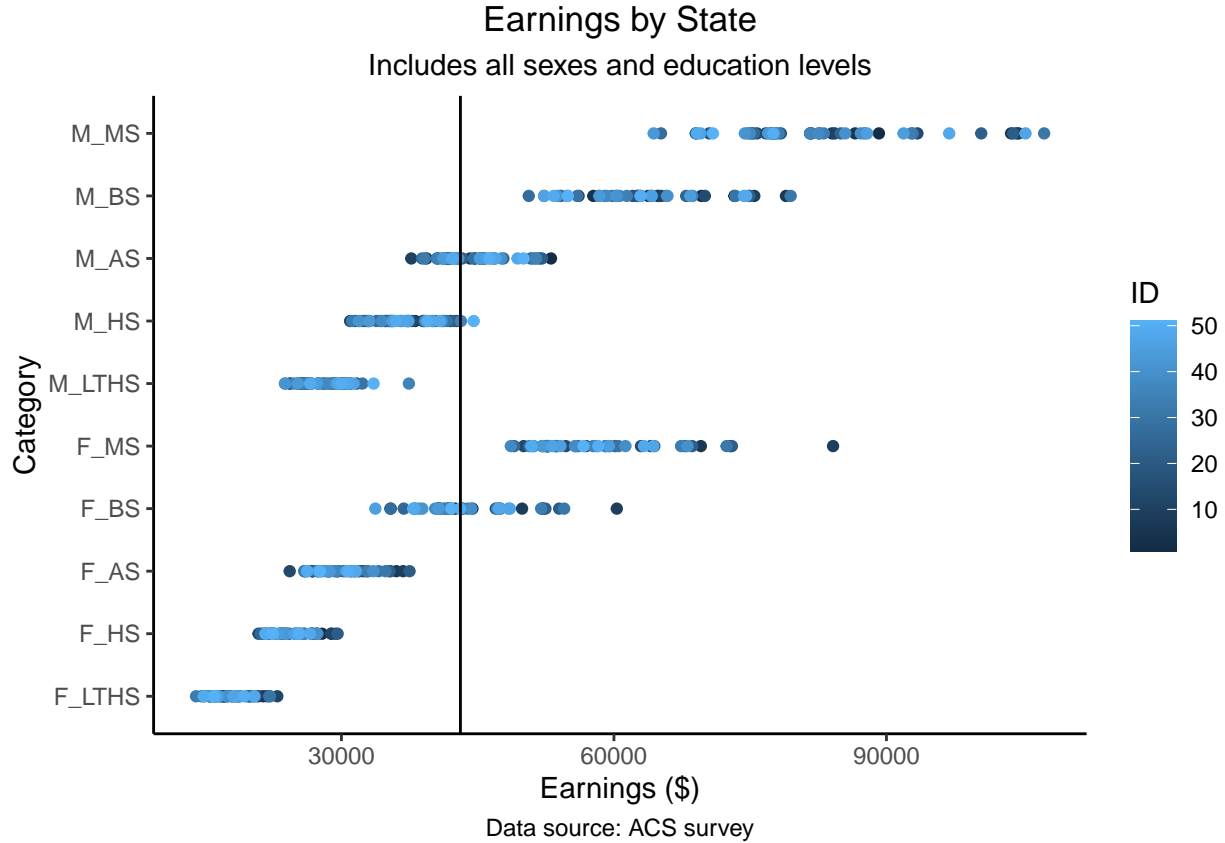
```
## ANOVA Table (type II tests)
##
##   Effect DFn DFd      F      p p<.05      ges
## 1      ID   1 508 0.402 0.526      0.00079
```

Our null hypothesis of this section would be that there is no significant difference by state in the observed differences of earnings by education. Alternatively, if this hypothesis were proven incorrect, we could say that this a significant difference in the observed differences of earnings by education in the states of the United States. However, this is not the case and we fail to reject the null hypothesis.

With an F-value of 0.402 and p-value of 0.526 the results of this hypothesis test are not significant and the variation in the median earnings of each education level by state is likely due to chance.

Now we can review all states by their ID values. These values were assigned to the states in alphabetical order starting with Alabama and include the District of Columbia. The darker the shade of blue the earlier the state occurs in the alphabetical list.

```
ggplot(mf_stateobs[1:510,], aes(Observation, Category, color = ID)) +
  geom_point() +
  geom_vline(xintercept = mfobs.mean$mean) +
  labs(x = "Earnings ($)",
       y = "Category",
       title = "Earnings by State",
       subtitle = "Includes all sexes and education levels",
       caption = "Data source: ACS survey") +
  theme_classic() +
  theme(plot.caption = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



Variables in the category of female start with an F, followed by the abbreviation for their education level. The same rule applies to male categories and both are concentrated on the upper and lower halves of the chart.

It appears as though there might be a cluster of slightly darker points on the right side of the data for the female education categories of *less than high school*, *high school*, and *some college or an associates degree*. However, the same does not occur for the rest of female education categories nor any male category. Additionally, there are a few darker points located on the far left sides of the data in these categories and they all seem to be randomly spread throughout as depicted in the results of the analysis of variance.

Conclusion

Through an analysis of variance on the observed differences in wages at the educational levels of less than high school, high school, some college or associates degree, bachelors degree, masters or graduate degree, we can conclude that higher levels of education did increase the average annual earnings of males and females age 25 and older in 2018. However, we were unable to prove that results at the state level were significantly different from one another, suggesting no state has reduced its wage gap beyond that of random variation in the same year.

Results for the increase in earnings by higher education attainment were significant at the 0.0001 level with a p-value of 0 and f-value of 106.948. Higher education has a strong correlation with higher average earnings. Variations in this data is extremely unlikely to have occurred by chance.

Results for the difference in earnings by state suggest that any variation in the observed differences in wages or proportions of female to male earnings is likely to have occurred by chance. It had a p-value of 0.526 and thus is not significant at the 0.05 alpha level. The associated f-value 0.402 being very near 1 demonstrates the increased probability that there is no state with a significant difference in the wage gap.

On average across this entire study, females working full-time make \$11,510 less than males per year. For females in their prime working ages of 25 to 54, this could result in a difference of \$333,790 over this 29 year span if trends in observed differences held constant. For reference, the average male earns \$51,477 per year and would accumulate \$1,492,833 over the same working duration if variables remained constant.

For both males and females education is likely to improve their average earnings but it does not mitigate the observed differences in earnings between men and women. Instead, the observed differences (the gap) increase with higher education.

For example, those in the category with the lowest level of education attainment earned \$28,353 for males and \$17,509 for females on average. This is a difference of \$11,510 for those with less than a high school education. Apply this to the the highest category of education, those with a master or graduate degree, and we see a difference of \$24,951 with males having earned \$81,613 and females having earned \$56,662 dollars on average.

The range of the highest and lowest earnings for education is \$53,260 dollars for the average male and \$39,153 dollars for the average female. This difference implies female earnings are distributed less widely than males. Additionally, there is a smaller gap in earnings at lower levels of education and lower earnings.

These results are unadjusted and include individuals choices of occupation, differences in experience levels, hours worked, salary negotiations, and all other variables associated with the earnings of individuals in United States. This analysis depicts a more realistic view of the gender wage gap in the United States for the year 2018 and elucidates some common misconceptions about education's effect on the gender wage gap in the United States.

References

- Blau, D. F., Kahn, M. L. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*. 14(4) 75-99 doi:10.1257/jep.14.4.75
- Bleiweis, Robin. (2020, March 24). Quick Facts about the Gender Pay Gap. Center for American Progress. Retrieved from https://cdn.americanprogress.org/content/uploads/2020/03/23133916/Gender-Wage-Gap-.pdf?_ga=2.138595639.58056853.1606596914-958559537.1606596914
- Bureau of Labor Statistics. (2018, April). Measuring the value of education : Career Outlook. Retrieved from <https://www.bls.gov/careeroutlook/2018/data-on-display/education-pays.htm>
- Bureau of Labor Statistics (2017, August). Rep. No. 1069. Highlights of women's earnings in 2016. Retrieved <https://www.bls.gov/opub/reports/womens-earnings/2016/home.htm>
- Card, D. (2000). The causal effect of education on earnings. Dept. of Economics, University of California, Berkeley. Retrieved from https://eml.berkeley.edu/~card/papers/causal_educ_earnings.pdf
- Graf, N., Brown, A., & Patten, E. (2020, August 14). The narrowing, but persistent, gender gap in pay. Pew Research Center. Retrieved from <https://www.pewresearch.org/fact-tank/2019/03/22/gender-pay-gap-facts/>
- Gould, R. (2016, October 20). What is the gender pay gap and is it real?: The complete guide to how women are paid less than men and why it can't be explained away. Economics Policy Institute. Retrieved from <https://www.epi.org/publication/what-is-the-gender-pay-gap-and-is-it-real/>
- Hamidullah, M. F., Riccucci, N. M., & Lee, I. P. (2020). Citizens' perceptions of closing the gender pay gap: An experimental study. *Public Management Review*, 1-24. doi:10.1080/14719037.2020.1722207
- Hegewisch, A., Heidi Hartmann, P. (2020, August 25). The Gender Wage Gap: 2018 Earnings Differences by Race and Ethnicity. Retrieved from <https://iwpr.org/iwpr-general/the-gender-wage-gap-2018-earnings-differences-by-race-and-ethnicity/>
- Lewis, G. B., Boyd, J., Pathak, R. (2017). Progress toward Pay Equity in State Government. *Public Administration Review*, 78(3), 386-397. doi:10.1111/puar.12897

The Equal Pay Act of 1963. (n.d.). Retrieved from <https://www.eeoc.gov/statutes/equal-pay-act-1963>

U.S. Census Bureau. (2020, August 27). Methodology. Retrieved December 01, 2020, from <https://www.census.gov/programs-surveys/acs/methodology.html>

Gould, Elise. (2019, February 20). State of Working America Wages 2018: Wage inequality marches on- and is even threatening data reliability. Economic Policy Institute. Retrieved from <https://www.epi.org/publication/state-of-american-wages-2018/>