The Working Lives of Married Families

Yicheng Wang, Ao Liu, Arnold Lau, Yanyu Zheng

February 3, 2016

Load libraries

```
library("dplyr") # for data wrangling
library("data.table") # for data handling
library("ggplot2") # for plotting
library("choroplethr") # for mapping
library("choroplethrMaps") # contains state map
library("mapproj") # contains projections
library("gridExtra") # allows multiple ggplot2 plots on the same window
library("RColorBrewer") # colors!
library("splitstackshape") # for assistance with weighting
library("cowplot") # for saving plots
source("../lib/state_choropleth_wl.R") # modified state_choropleth function that removes labels
```

Getting to know the basics of the data

There are a total of 1,476,313 households and 3,249,588 people in the sample. In this analysis we focus on a subset of the households via the variable **FES** (Family Type and Employment Status), looking only at married-couple families.

```
table(hp$FES, useNA = "ifany")
```

```
##
##       1       2       3       4       5       6       7       8    <NA>
## 1007309  457181  144836  287154  121762   36907  294869  139788  759782
```

```
sum(table(hp$FES)[1:4])
```

```
## [1] 1896480
```

What are we focusing on?

There are 759,782 missing values in **FES** in the sample, and our subset will have 1,896,480 people (617,090 families). In our subset, **FES** has four categories:

1. Married-couple family, both husband and wife in the labor force
2. Married-couple family, only husband in the labor force
3. Married-couple family, only wife in the labor force
4. Married-couple family, neither husband nor wife in the labor force

To be "in the labor force" means that one is a civilian above 18 years of age who is either employed, or unemployed and seeking work.

We merge in a .csv file matching full state names and state abbreviations by PUMS state ID. We then create two separate datasets, one to be weighted using the household weight variable **WGTP** and the other to be weighted using the person weight variable **PWGTP**.

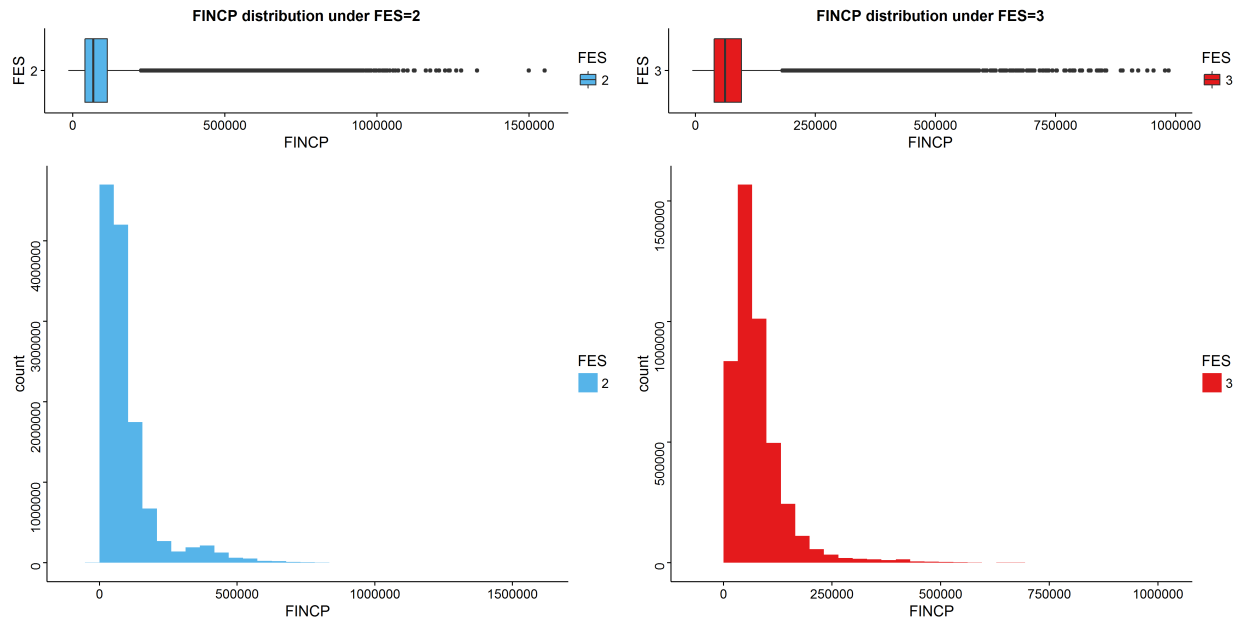
```
statenames <- read.csv("../data/statenames.csv", strip.white = TRUE)
statenames <- rename(statenames, ST = code, ST_name = name, ST_abbr = abbr)
hp <- tbl_df(hp)
hp <- left_join(hp, statenames)
```

```
## Joining by: "ST"
```

```
hp_fes1to4_wgtp <- select(hp, SERIALNO, WGTP, FES, ST_name, FINCP, MAR, SEX, INDP2, SCHL) %>%
              filter(FES >= 1 & FES <= 4) %>%
              distinct(SERIALNO)
hp_fes1to4_wgtp$FINCP <- hp_fes1to4_wgtp$FINCP * 1.007549 # adjust to 2013 dollars
hp_fes1to4_pwgtp <- select(hp, PWGTP, FES, FINCP, PINCP, MAR, SEX, INDP2, SCHL) %>%
              filter(FES >= 1 & FES <= 4)
hp_fes1to4_pwgtp$FINCP <- hp_fes1to4_pwgtp$FINCP * 1.007549
hp_fes1to4_pwgtp$PINCP <- hp_fes1to4_pwgtp$PINCP * 1.007549
rm(hp)
```

Families with Single-Earning Husbands vs. Single-Earning Wives

We are mainly interested in comparing families in which only the husband is in the labor force versus families in which only the wife is in the labor force. Are there any obvious differences between these subgroups of families along the following:
total family income, which might include both the money made by whoever is in the labor force plus any passive income earned by that person's spouse, as well as other family members?
education level of the person in the labor force?
industry type of the person in the labor force?




```
summary(FINCP_FES2$FINCP)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -14110   40300   68510   99950  114800 1563000
```

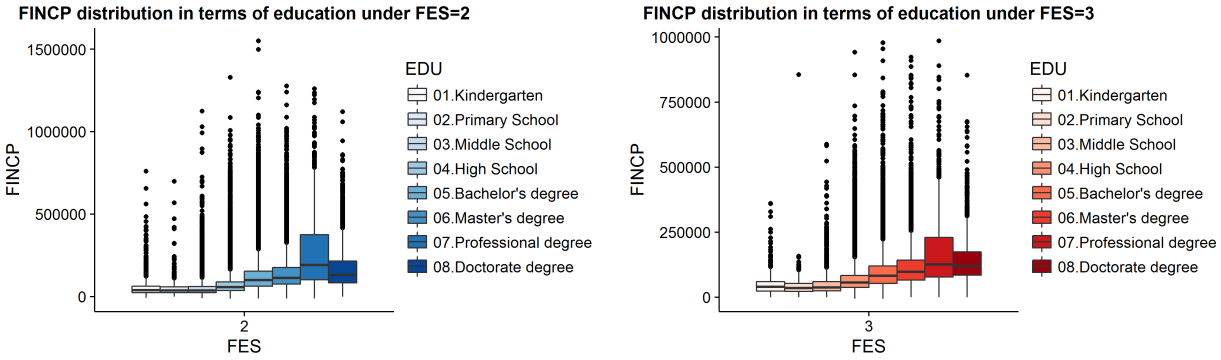
```
summary(FINCP_FES3$FINCP)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -6045   39290   62470   79190   96720  992900
```

The combined boxplot and histogram show fairly unsurprising income distributions, with large right skews. Median income is similar for both groups, but outliers are stronger among families with only the husband in the labor force compared to familes with only the wife in the labor force. Based on weighted estimates, there are around 12.5 million families where only the husband is in the labor force, compared to only 4.5 million families vice-versa. Notice also that the bin on the right peaks at a slightly higher amount than on the left.

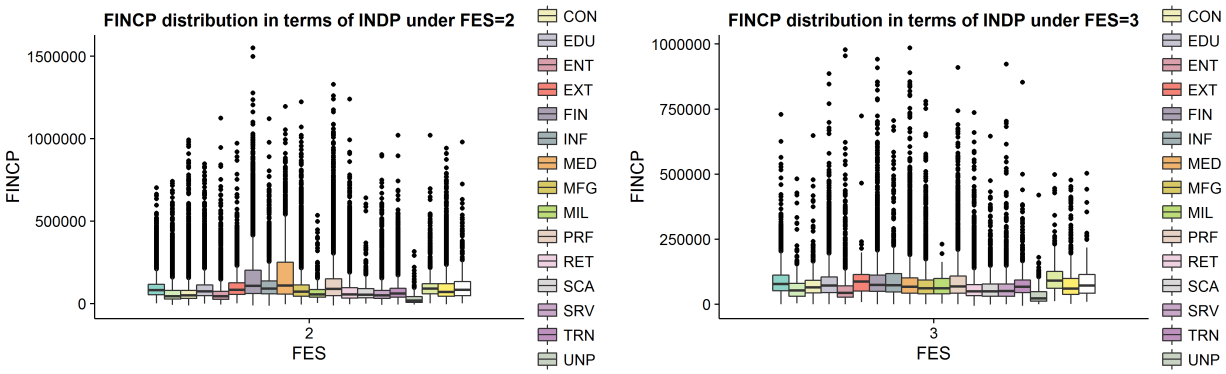
```
## Joining by: "SCHL"
```

Next, we look at how these two groups differ across the levels of education of the person in the labor force, using a recoded version of the **SCHL** variable with less categories.



While the two plots look fairly similar, a closer look indicates that the median income for the families of married women where they are the sole person in the labor force is slightly higher if the wife has master's and doctorate degrees, while married men's families make much more use out of the husband having professional degrees. There also appear to be more outliers among the family incomes of married men who finished only kindergarten, primary school, or middle school. The general unsurprising trend is that family income tends to rise as the education level of the "laborer" in the family rises. In this sense, however, post-baccalaureate professional degrees are associated with higher median incomes and higher outliers compared to doctorate degrees.

Finally, we look at the distribution of family incomes by occupational industry between the two subgroups. If we assume that, given that they are within a particular industry, there are no differences in the outlook within that industry for single-earning husbands and single-earning wives, then we would expect the distributions per industry to be very similar between the two subgroups.



The leftmost two categories got cut out of the legend; they are ADMINISTRATION and AGRICULTURE.

With the exception of the finance and medical industries for single-earning husbands, they enjoy smaller boxes (i.e. smaller differences betwen 25th and 75th percentiles) across all industries. In other words, within each industry, the incomes of married-couple families with only the wife in the labor force exhibit higher variance. For both subgroups, people in the military and in social services had the lowest income spreads.

```
sort(table(FINCP_FES2$INDP2))
```

```
##
##     UNP     SCA     MIL     UTL     EXT     INF     AGR     WHL     EDU
##   38509   54356  197348  199757  222700  282485  337195  497894  561137
##     SRV     ENT     MED     ADM     TRN     FIN     RET     CON     PRF
##  595737  615830  636390  654653  809165  835687 1054911 1537994 1735888
##     MFG
## 2045346
```

```
sort(table(FINCP_FES3$INDP2))
```

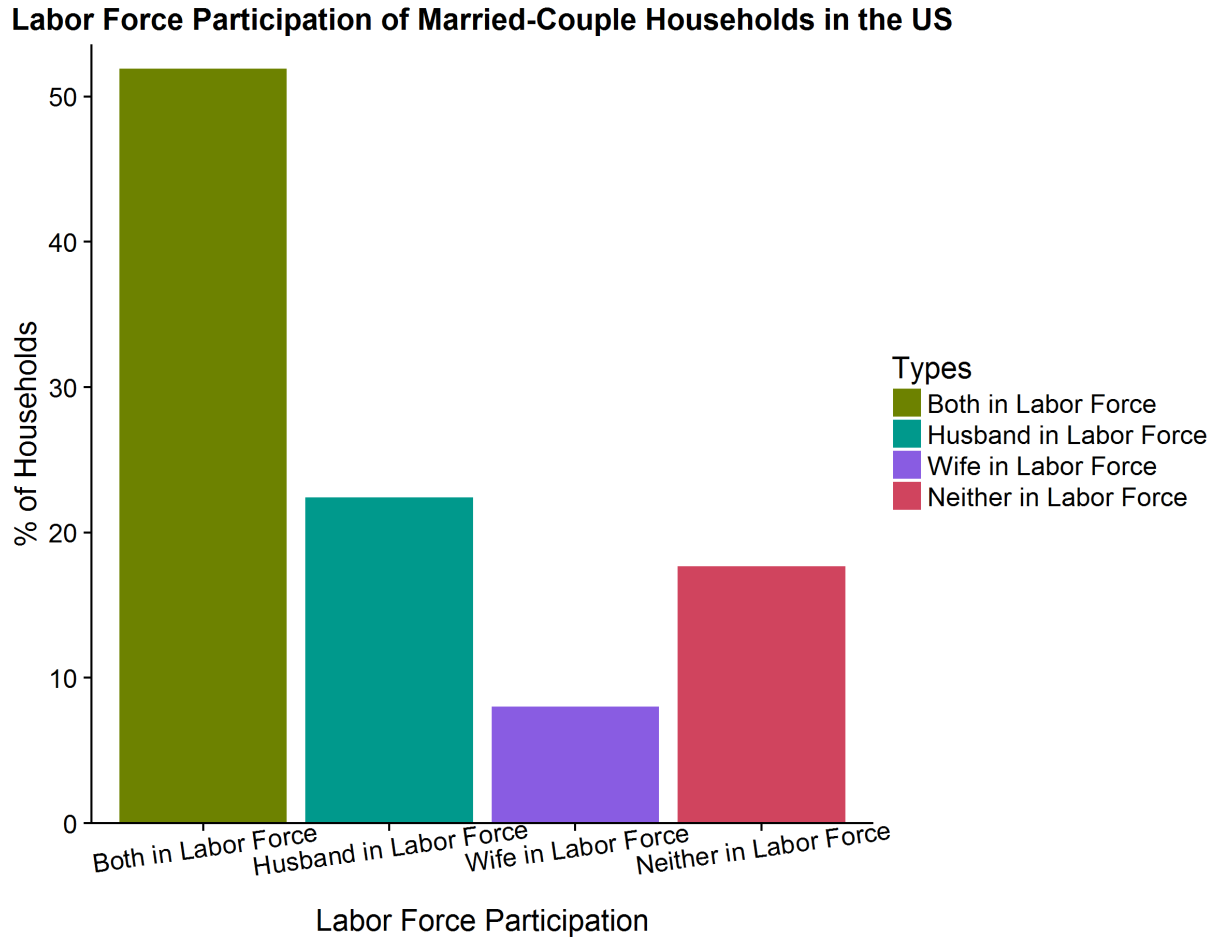
```
##
##    MIL    EXT    UTL    AGR    UNP    CON    WHL    INF    TRN    SCA
##   5267   9630  25738  27330  27408  53327  76444  78097 103978 179250
##    ADM    SRV    ENT    MFG    FIN    PRF    RET    EDU    MED
## 248190 255711 276935 343555 353399 397630 505689 623078 937736
```

The top 5 common industries among husbands alone in the labor force: manufacturing, professional services, construction, retail, and finance.

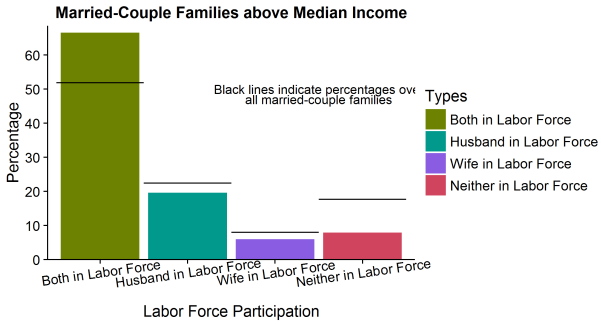
The top 5 common industries among wives alone in the labor force: medicine, education, professional services, finance and manufacturing.

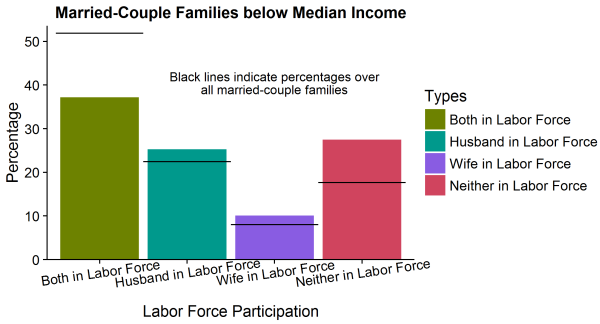
Distribution of types of married-couple families

Then, we look at the proportion of different types of married-couple families over all married-couple families:

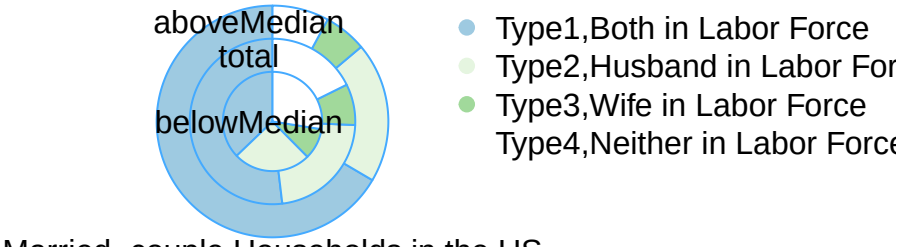


Does this proportion differ for married-couple families at or above median income ($77,000), versus below median income?



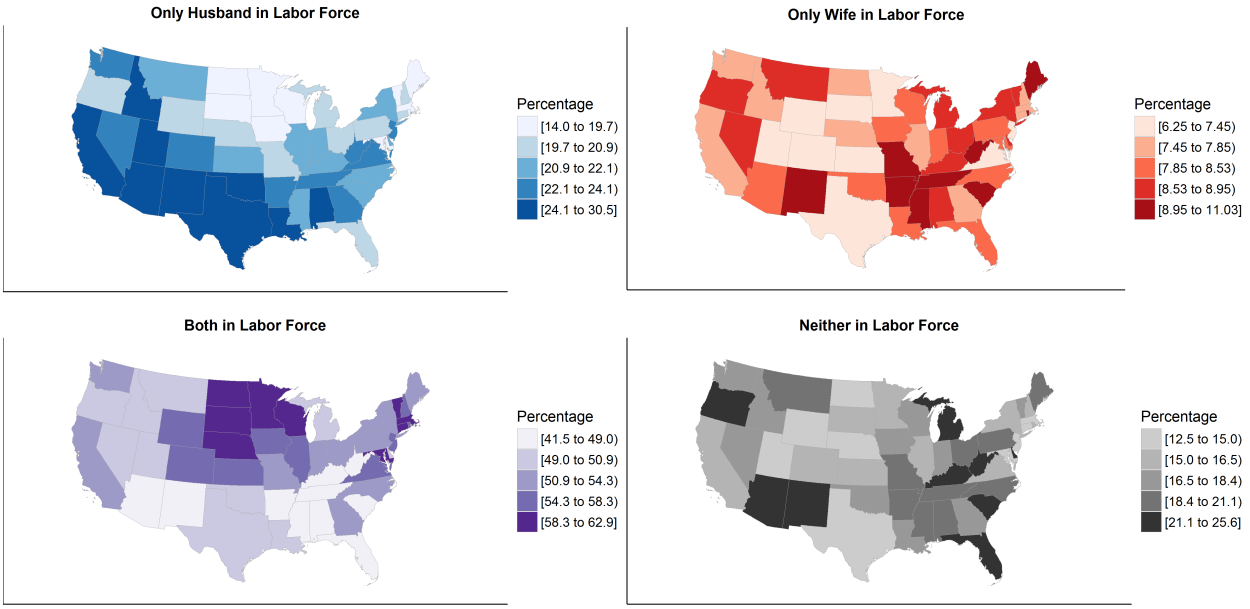


As we can see, above median income, there are many more married-couple families with both in the labor force, while below median income, there are many more married-couple families with only one person or none in the labor force. This isn't surprising when you think about it, but it also doesn't necessarily mean that being in the labor force is a family's only source of income, as we'll see later.



Married−couple Households in the US

Next, we look at the distribution of different types of married-couple families across the continental United States. Do certain states or groups of states have greater proportions of families with both husband and wife in the labor force, as opposed to only one of them or neither of them?



Across the 48 continental states:

- The proportion of married-couple families with only the husband in the labor force varies from 14 to 30%, with evident clusters of high proportions towards the southwest and south of the country and clusters of low proportions towards the north-central part.

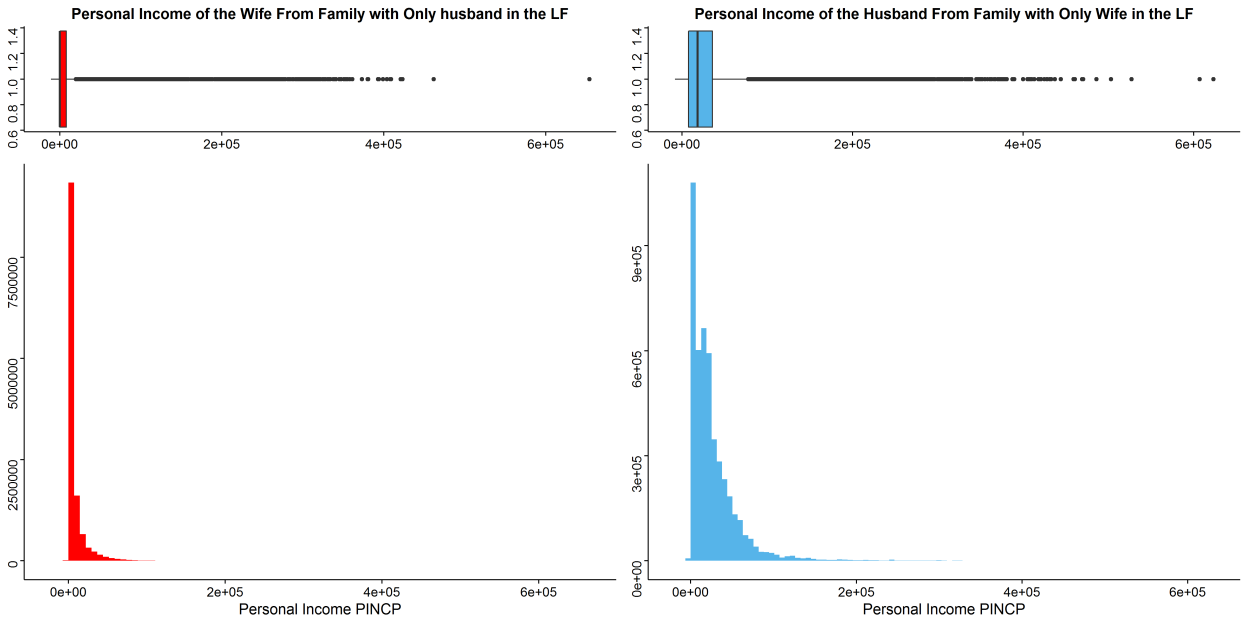
- The proportion of married-couple families with only the wife in the labor force varies from 6 to 11% (in other words, it's far less common in every state for a family to have only the wife in the labor force). There appears to be clusters of high proportions around the southern states.

- The proportion of married-couple families with both the husband and the wife in the labor force varies from 41 to 63%, with the lowest proportions clustering around the southern states and the highest proportions in the north-central part of the country.

- The proportion of married-couple families with neither the husband or the wife in the labor force varies from 12 to 26%. We are not sure if there's any pattern here, but places like Florida with lots of retirement communities exhibit high proportions here (Florida is ranked #1 by this metric).

Do Spouses Not in the Labor Force Still Have Income?

When we looked into the data carefully, we found that the family income of the family from which has only husband or wife in the labor force (LF) did not always coincide with the personal income of the husband or wife who is the LF. This suggests that for several families, the spouse who was classified as not in the labor force may also have personal income from different sources. For example, he or she may have stock market dividends. We produce a plot similar to the ones earlier of the personal incomes of spouses who aren't in the labor force alongside their significant others.



These two plots shows the boxplot and distribution of the Personal Income from the family which has only one person in the labor force (LF). We can see that the median of the two plots are generally close to the zero, which is reasonable. Therefore, both distributions are extremely right skewed. This is the usual shape of an income distribution, except nobody here is classified as being in the labor force.

The median personal income for husbands not in the labor force is higher than for wives, and as you can see from the plot, the percentage of males who are making money is MUCH greater than the percentage of females. Suprisingly, there are quite a few non-labor-force participants with a great amount of personal income. One example of an extreme data point is a wife not in the labor force who has a personal income of $635,000 dollars in a year, compared to her husband who IS in the labor force but earns only $109,000 dollars in a year. We find this intriguing and worth looking deeper into.