

Draft Analysis for Married Family & Family Income

Yicheng Wang

Saturday, January 30, 2016

Contents

Preparation	1
Reading data	1
Data Manipulation &Format	2
Visuliazation	2
Exploraring the basic statistics	2
Mapping Tool	4

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Preparation

```
library("dplyr")
library("data.table")
library("ggplot2")
library("choroplethr")
```

Reading data

```
readIn = function(reRead, colsToKeep, pathA, pathB){
  if(reRead == TRUE){
    popDataA = fread(pathA, select = colsToKeep)
    popDataB = fread(pathB, select = colsToKeep)
    popData1 = rbind(popDataA, popDataB)
    rm(popDataA, popDataB)
    save(popData1, file = "popData1.RData")
  }else{
    load("popData1.RData")
  }
  return(popData1)
}
```

```
data1 <- readIn(T,c("ST","FES","FINCP"),"/Users/cheeselopeicecream/Documents/columbia/W4249/ss13husa.cs
```

```
##
Read 15.9% of 756065 rows
Read 56.9% of 756065 rows
Read 97.9% of 756065 rows
Read 756065 rows and 3 (of 231) columns from 0.574 GB file in 00:00:05
##
Read 22.2% of 720248 rows
Read 65.3% of 720248 rows
Read 720248 rows and 3 (of 231) columns from 0.546 GB file in 00:00:04
```

Data Manipulation & Format

```
data1 <- tbl_df(data1)
ds <- data1 %>%
  na.omit() %>%
  filter(FES %in% c(1,2,3,4)) %>%
  group_by(FES)
rm(data1)
```

Visuliazation

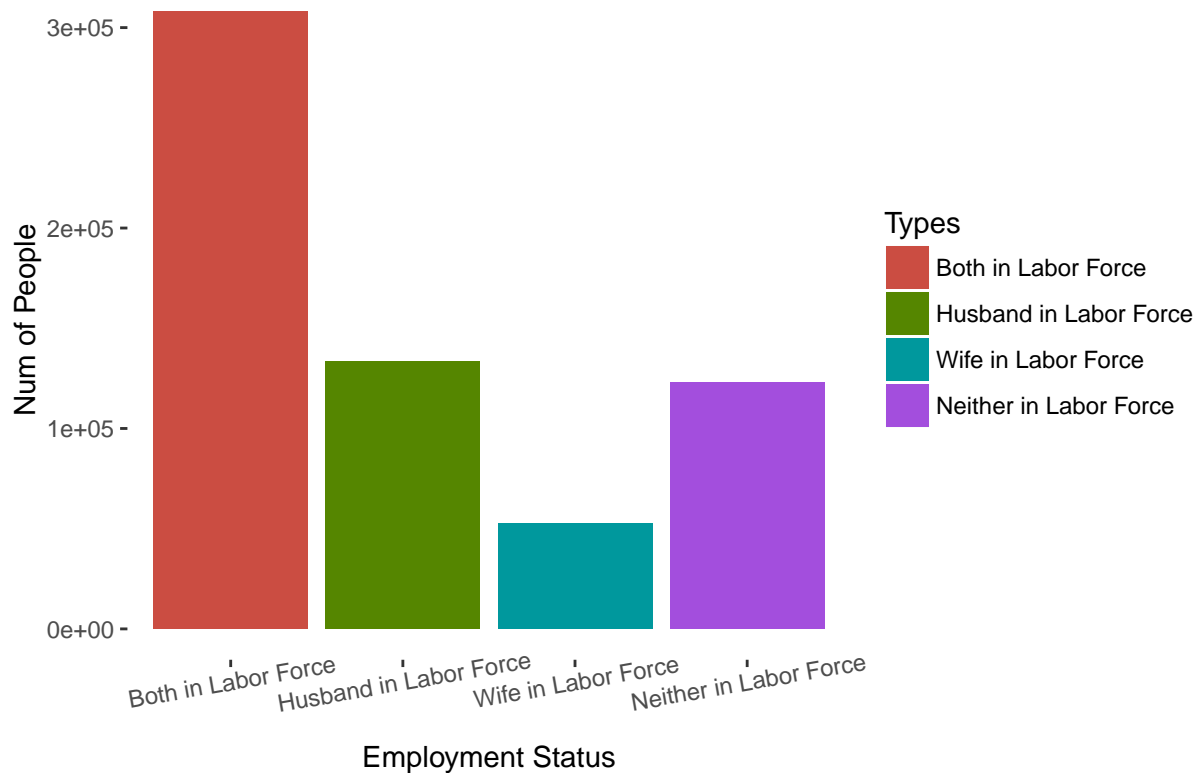
```
fescode <- "FES,MarriedFamilyEmployType
1,Both in Labor Force
2,Husband in Labor Force
3,Wife in Labor Force
4,Neither in Labor Force"
fescodes <- fread(fescode)

fesnumbers <- summarise(ds,count=n())
fesnumbers <- left_join(fesnumbers,fescodes,by.x=c("FES"))
Types <- factor(fesnumbers$MarriedFamilyEmployType,levels=unique(fesnumbers$MarriedFamilyEmployType))
```

Exploraring the basic statistics

```
ggplot(fesnumbers, aes(x= Types , y=fesnumbers$count, fill= Types)) +
  geom_bar(stat="identity") + scale_fill_hue(l=50) +
  ylab("Num of People") +
  xlab("Employment Status") + ggtitle("Comparing Employment status of Married Family in the US") +
  theme(axis.text.x = element_text(angle = 10, hjust = 0.5),
  panel.background = element_rect(fill = 'white' ))
```

Comparing Employment status of Married Family in the US



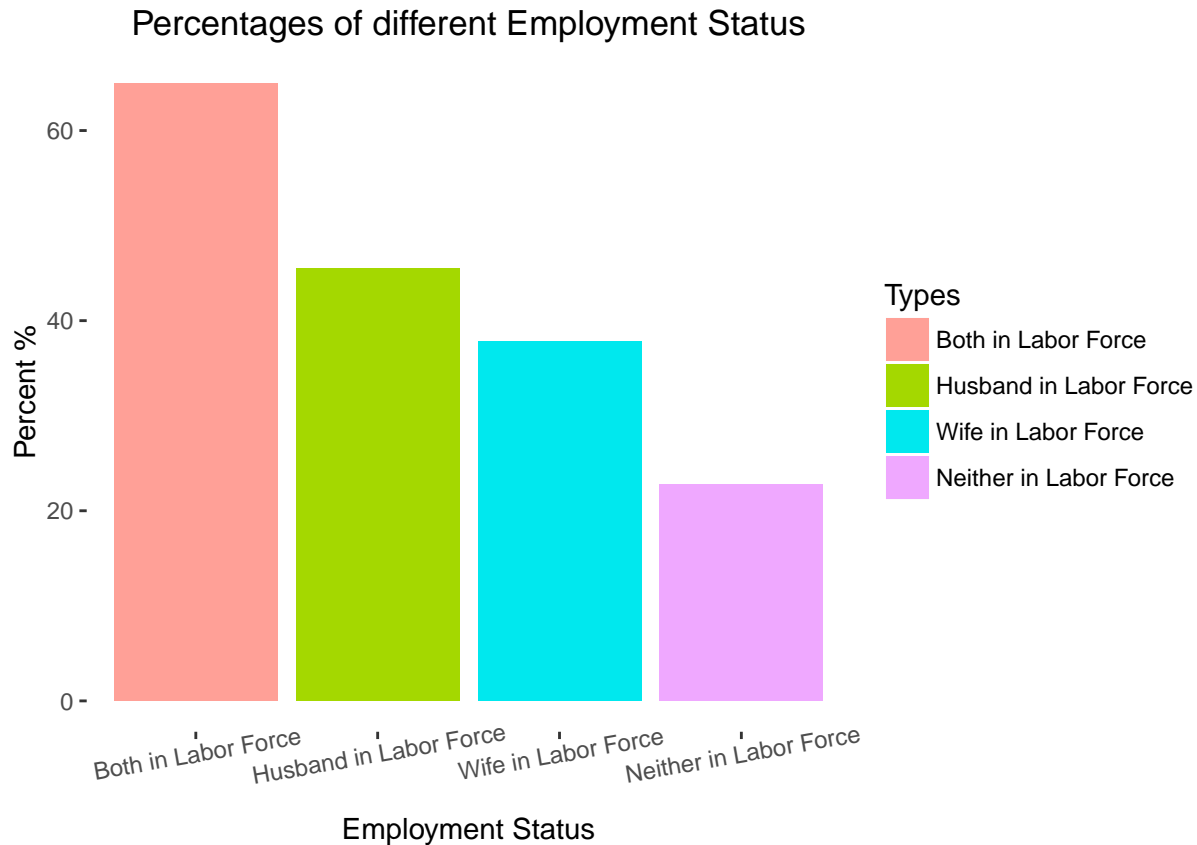
```
summary(ds$FINCP)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -19770   45800   77000  100600  121200 2060000
```

The median family income is 77000 dollars. We use this as bench line. We want to see the percentage under each type of family who has family income more than 77000 which is approxiamatly middle class.

```
medianper <- ds %>%
  filter(FINCP>=77000) %>%
  group_by(FES) %>%
  summarise(count = n()) %>%
  mutate(Percent = count/fesnumbers$count*100)
```

```
ggplot(medianper, aes(x= Types , y=medianper$Percent, fill= Types)) +
  geom_bar(stat="identity") + scale_fill_hue(l=80) +
  ylab("Percent %") +
  xlab("Employment Status") + ggtitle("Percentages of different Employment Status")+
  theme(axis.text.x = element_text(angle = 10, hjust = 0.5),
  panel.background = element_rect(fill = 'white' ))
```



We can see an interesting result comparing to the previous one that the married family with only wife in the labor force has absolute less numbers of people than the family who has only husband in the labor force, while showing almost the same percentage with more than 77000 family income. This could be investigated later.

Mapping Tool

This map shows percentage of family which neither of the member is in the labor force while having more than 121200 dollars (3rd quantile) family income.

```
stateTotalBoth <- ds %>%
  filter(FES==4) %>%
  group_by(ST) %>%
  summarise(count=n())

medianBoth <- ds %>%
  filter(FES==4, FINCP >= 121200) %>%
  group_by(ST) %>%
  summarise(count=n())

medianBoth <- right_join(medianBoth, stateCodes, by.x=c("ST"))

## Joining by: "ST"
```

```
medianBoth[is.na(medianBoth)] <- 0
medianBoth <- mutate(medianBoth,value=medianBoth$count/stateTotalBoth$count*100)

state_choropleth(medianBoth,title="Percentage of Neither People in labor earning more than 121200",num_c
```

Percentage of Neither People in labor earning more than 121200

