Money, Money, Money!

Ziyue Jin, Hexiu Ye, Yimin Zhang, Juan Campos

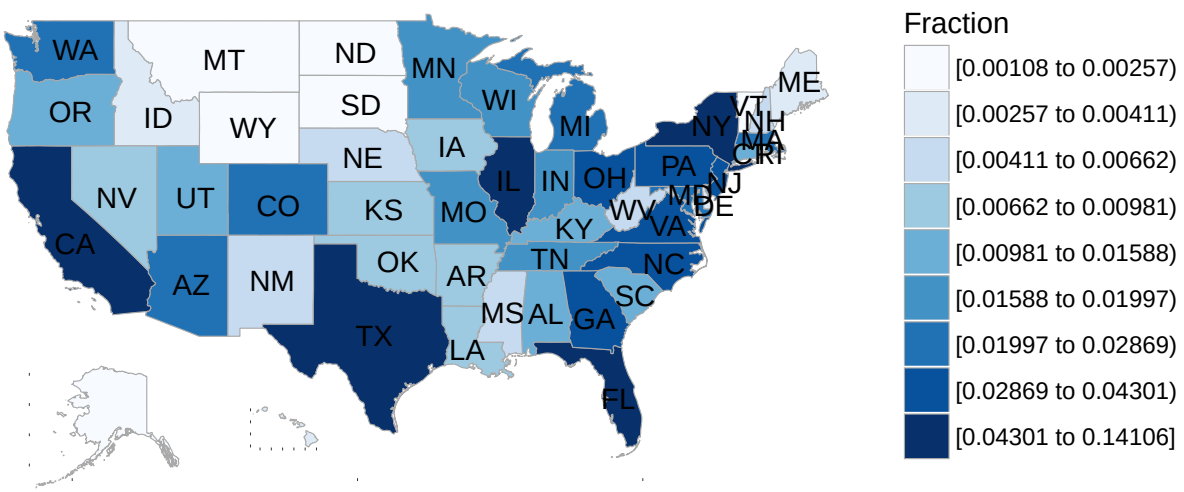
February 2, 2016

Ziyue Jin

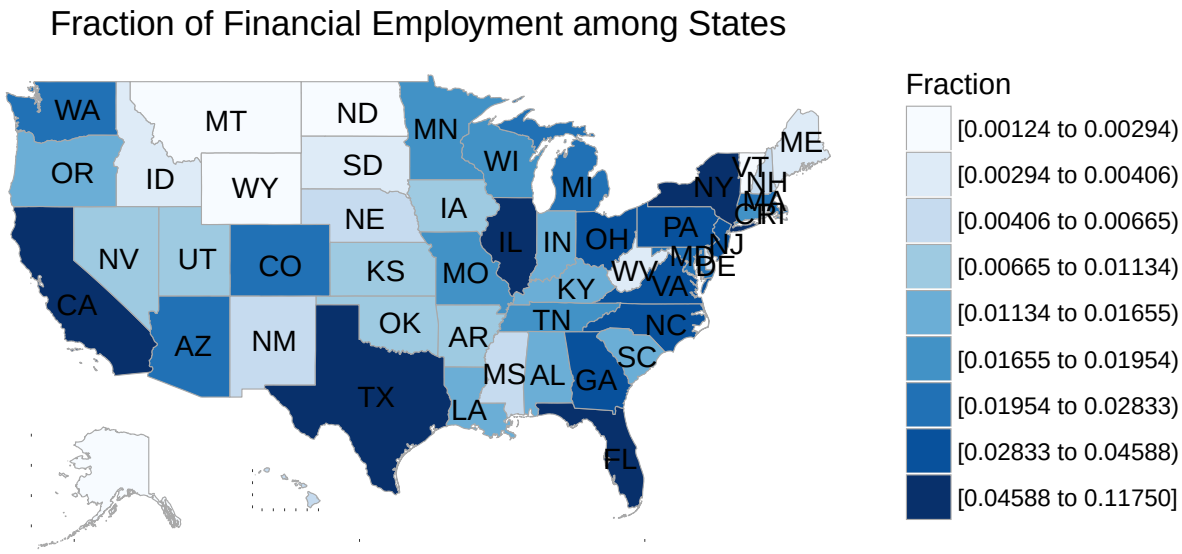
First Part

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

setwd("~/Documents/cycle1-2/data")
control <- 1

if(control == 0) {
  # Select variables: state, income, occupation, race, language spoken at home
  var <- c("ST", "WAGP","WKL", "OCCP", "INDP")
  dA <- fread("~/Documents/cycle1-2/data/csv_pus/ss13pusa.csv", select = var)
  dB <- fread("~/Documents/cycle1-2/data/csv_pus/ss13pusb.csv", select = var)
  data <- rbind(dA, dB)
  setwd("~/Documents/cycle1-2/data")
  save(data, file="4249Data.RData")
  rm(dA, dB)
} else {
  load(file="4249Data.RData")
}

# Add a new column to test whether a person is computer related job or not
computerCode <- c(0110, 1005, 1006, 1010, 1050, 1105, 1106, 1107, 1400, 5800,
                  5900, 7010, 7900)
infoCode <- c(6470 ,6480 ,6490 ,6570 ,6590 ,6670 ,6672 ,6680 ,6690 ,6695 ,6770 ,6780)

finanCode <- c(0120, 0800, 0810, 0820, 0830, 0840, 0850, 0860, 0900, 0910, 0930, 0940, 0950)
indFCode <- c(6870, 6880, 6890, 6970, 6990, 7070, 7080, 7170, 7180, 7190)

techOrNot <- function(x, y) {
  ifelse(x %in% computerCode | y %in% infoCode, 1, 0)
}
finOrNot <- function(x, y) {
  ifelse(x %in% finanCode | y %in% indFCode, 1, 0)
}

techLabel <- mutate(data, computerCheck = techOrNot(OCCP, INDP))
finLabel <- mutate(data, finanCheck = finOrNot(OCCP, INDP))

# Separate data by their state
gfinLabel <- group_by(finLabel, ST)
```

```
gtechLabel <- group_by(techLabel, ST)
# Calculate total number of computer related workers
techByState <- summarize(gtechLabel, value=sum(computerCheck, na.rm = T))
# Total number of financial related workers
finByState <- summarize(gfinLabel, value=sum(finanCheck, na.rm = T))
# Get the proportion of related workers in that state
total <- sum(techByState$value)
ftotal <- sum(finByState$value)
techByState <- mutate(techByState, value = value/total)
finByState <- mutate(finByState, value = value/ftotal)
# Transfer state code to state name
state_list =list("1"="alabama", "2"="alaska", "4"="arizona", "5" = "arkansas",
                 "6" = "california", "8" = "colorado", "9" = "connecticut",
                 "10" = "delaware", "11" = "district of columbia", "12" = "florida",
                 "13" = "georgia", "15" = "hawaii",  "16" = "idaho",
                 "17" = "illinois", "18" = "indiana", "19" = "iowa", "20" = "kansas",
                 "21" = "kentucky", "22" = "louisiana",  "23" = "maine",
                 "24" = "maryland", "25" = "massachusetts", "26" = "michigan",
                 "27" = "minnesota",  "28" = "mississippi", "29" = "missouri",
                 "30" = "montana", "31" = "nebraska", "32" = "nevada",
                 "33" = "new hampshire", "34" = "new jersey", "35" = "new mexico",
                 "36" = "new york", "37" = "north carolina", "38" = "north dakota",
                 "39" = "ohio",    "40" = "oklahoma", "41" = "oregon",  "42" = "pennsylvania",
                 "44" = "rhode island", "45" = "south carolina", "46" = "south dakota",
                 "47" = "tennessee",  "48" = "texas",   "49" = "utah",    "50" = "vermont",
                 "51" = "virginia", "53" = "washington", "54" = "west virginia", "55" = "wisconsin",
                 "56" = "wyoming", "72" = "puerto rico")
regionTransfer <- function(x){return ( state_list[[as.character(x)]]) }
plotData <- techByState %>% mutate(region = vapply(ST, regionTransfer, "")) %>%
  select(region, value)
plotfData <- finByState %>% mutate(region = vapply(ST, regionTransfer, "")) %>%
  select(region, value)
state_choropleth(plotData, title = "Fraction of Technology Employment among States",
                 legend = "Fraction", num_colors = 9)
```



```
state_choropleth(plotfData, title = "Fraction of Financial Employment among States",
                 legend = "Fraction", num_colors = 9)
```




```
plotData[which.max(plotData$value),]$region
```

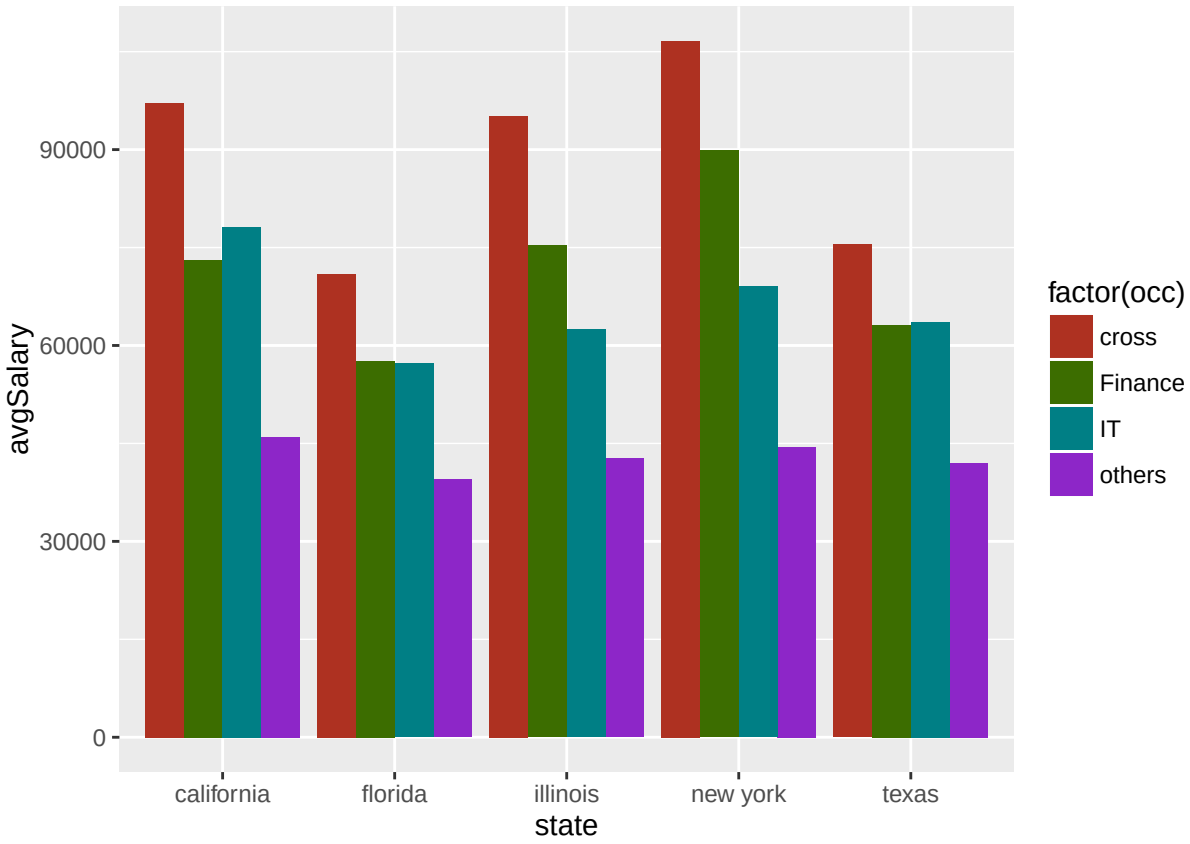
```
## [1] "california"
```

```
plotfData[which.max(plotfData$value),]$region
```

```
## [1] "california"
```

Second Part

```
finLabel$finanCheck <- finLabel$finanCheck*2
combine <- mutate(techLabel, check=computerCheck+finLabel$finanCheck)
groupData <- combine %>%
  filter(ST %in% c(6,12,17,36,48)) %>%
  na.omit() %>%
  filter(WAGP!=000000) %>% #exclude no salary person
  filter(WAGP!='bbbbbb') %>% # exclude N/A
  filter(WKL==1) %>% #only person who last worked in 12 month
  group_by(ST, check)
sumData <- summarize(groupData, avgSalary = mean(WAGP))
sumData$state <- vapply(sumData$ST, regionTransfer, "")
occupationList <- c("0"="others", "1"="IT", "2"="Finance", "3"="cross")
occupationTransfer <- function(x){return ( occupationList[[as.character(x)]]) }
sumData$occ <- vapply(sumData$check, occupationTransfer, "")
ggplot(sumData, aes(x=state , y=avgSalary, fill=factor(occ))) +
  geom_bar(stat="identity",position="dodge") + scale_fill_hue(l=40)
```


avgSalary
90000
60000
30000
0
california
florida
illinois
new york
texas
state
factor(occ)
cross
Finance
IT
others