race_income

Advait Rajagopal

April 29, 2017

Getting and preparing data

```
rm(list = ls())
setwd("/Users/Advait/Desktop/New_School/Spring17/GSS_experiments/race_income")
getwd()
```

```
## [1] "/Users/Advait/Desktop/New_School/Spring17/GSS_experiments/race_income"
```

```
#Load packages
library(foreign)
library(dplyr)
# get the data
test <- read.csv("race_income.csv", header = T,sep = ",", row.names = 1)
str(test)
```

```
## 'data.frame':    54505 obs. of  2 variables:
##  $ data.race  : Factor w/ 3 levels "black","other",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ data.income: Factor w/ 12 levels "$1000 to 2999",..: 2 10 2 2 2 2 7 10 1 3 ...
```

```
#######
#Reorder factor levels to match original data
#######
print(levels(test$data.race))
```

```
## [1] "black" "other" "white"
```

```
print(levels(test$data.income))
```

```
##  [1] "$1000 to 2999"  "$10000 - 14999" "$15000 - 19999" "$20000 - 24999"
##  [5] "$25000 or more" "$3000 to 3999"  "$4000 to 4999"  "$5000 to 5999"
##  [9] "$6000 to 6999"  "$7000 to 7999"  "$8000 to 9999"  "lt $1000"
```

```
#reordering here
test$data.income <- factor(test$data.income,
                           levels(test$data.income)[c(12,1,6:11,2:5)])
print(levels(test$data.income))
```

```
##  [1] "lt $1000"       "$1000 to 2999"  "$3000 to 3999"  "$4000 to 4999"
##  [5] "$5000 to 5999"  "$6000 to 6999"  "$7000 to 7999"  "$8000 to 9999"
##  [9] "$10000 - 14999" "$15000 - 19999" "$20000 - 24999" "$25000 or more"
```

```
df_race_income <- test

#total whites = 44109
dim(df_race_income[df_race_income$data.race == "white",])
```

```
## [1] 44109     2
```

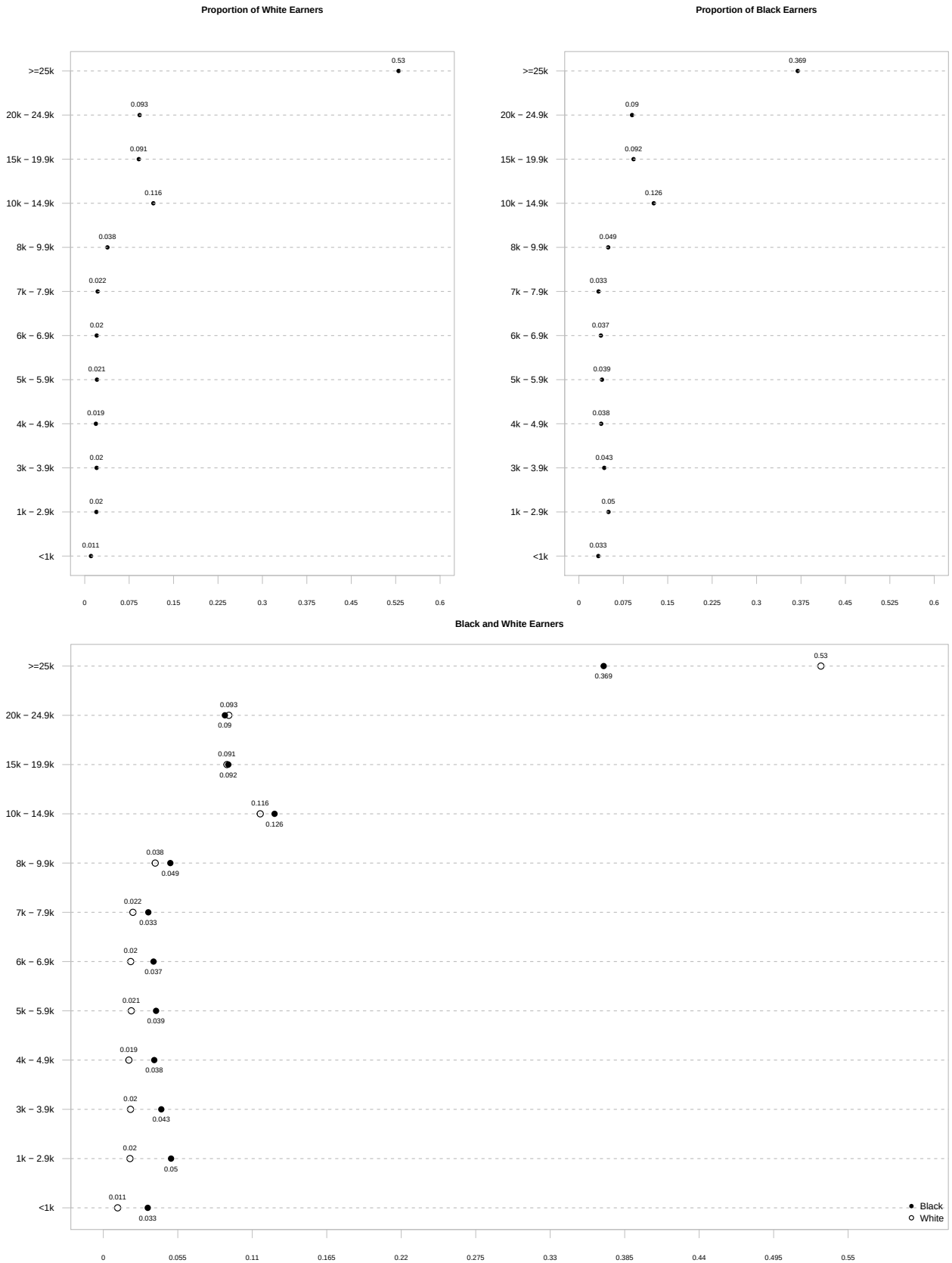
```
#total blacks = 7497
dim(df_race_income[df_race_income$data.race == "black",])
```

```
## [1] 7497    2
```

```r
#Part 1
#filter white people
#calculate proportion of white earners of each bracket
df_white <- df_race_income %>%
  filter(data.race == "white") %>%
   group_by(data.income) %>%
  summarize(prop_white = (n()/44109))

#Part 2
#filter black people
#calculate proportion of black earners of each bracket
df_black <- df_race_income %>%
  filter(data.race == "black") %>%
  group_by(data.income) %>%
  summarize(prop_black = (n()/7497))

#incomebracket vector to aid plots
income <- c("<1k","1k - 2.9k","3k - 3.9k",
            "4k - 4.9k", "5k - 5.9k", "6k - 6.9k",
            "7k - 7.9k", "8k - 9.9k", "10k - 14.9k",
            "15k - 19.9k", "20k - 24.9k",">=25k")

#creating dataframe for plots
df_plot_raceincome <- data.frame(inc_brack = df_black$data.income,
                                 prop_white = df_white$prop_white,
                                 prop_black = df_black$prop_black)
```

Plots

Plot White and Black Earners according to income bracket on separate plots and then plot on the same plot to facilitate comparison



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.