Race and ownership of residential properties

Mehdi Khan

October 8, 2017

Introduction:

This data was extracted from 2000 Census with information about race, rental and ownership of residences in Anne Arundel County, Maryland. The data was used by Anne Arundel County Community Development Services, Inc - a nonprofit agency in their housing research.

The data was downloaded as an excel file which was converted to a CSV file, which then needed to be made tidy to do analysis in R. The analysis was done to see if the data could provide any relationship in race and ownership of residential properties.

Load necessary libraries:

```
suppressMessages(suppressWarnings(library(stringr)))
suppressMessages(suppressWarnings(library(tidyr)))
suppressMessages(suppressWarnings(library(dplyr)))
suppressMessages(suppressWarnings(library(data.table)))
suppressMessages(suppressWarnings(library(ggplot2)))
```

load data:

```
houseDS <- read.csv("C:\\Temp\\race versus tenure of owner & rental housing units.csv",
    sep = ",", stringsAsFactors = FALSE, fill = TRUE)
# str(houseDS)
```

Since data is distributed under two categories of ownership (rental and owned) two datsets were created with only relevant rows and columns, all other rows and columns were avoided:

```
owned_houses <- houseDS[2:9, 1:2]
rental_houses <- houseDS[12:19, 1:2]
```

A funtion was created to do untidy operations on these two dataset, the function does the following: a. update with meaningful column names b. update race column with simplified race names c. add and populate new column with ownership information

```
untidyData <- function(dat) {
    colnames(dat) <- c("race", "population")
    dat$population <- str_replace(dat$population, ",", "")

    dat$population <- as.numeric(dat$population)
    for (i in (1:length(dat$race))) {
        if (str_detect(dat$race[i], "White")) {
            dat$race[i] <- "White"
        } else if (str_detect(dat$race[i], "Black")) {
            dat$race[i] <- "African American"
        } else if (str_detect(dat$race[i], "Indian")) {
            dat$race[i] <- "American Indian and Alaska Native"
```

```
        } else if (str_detect(dat$race[i], "Asian")) {
            dat$race[i] <- "Asian"
        } else if (str_detect(dat$race[i], "Native")) {
            dat$race[i] <- "Native Hawaiian, Pacific Islander"
        } else if (str_detect(dat$race[i], "other race")) {
            dat$race[i] <- "Others"
        } else if (str_detect(dat$race[i], "Two or more")) {
            dat$race[i] <- " Two or more races"
        }
    }
    mutate(dat, ownership = "")
    if (str_detect(dat$race[1], "Owner") == TRUE) {
        dat$ownership <- "owner"
    } else if (str_detect(dat$race[1], "Renter")) {
        dat$ownership <- "renter"
    }

    dat <- dat[-1, ]

    ## dat$percentage=dat$population/sum(dat$population)

    return(dat)
}
owned_houses <- untidyData(owned_houses)
rental_houses <- untidyData(rental_houses)

HouseData <- rbind(owned_houses, rental_houses)

head(HouseData)
```

```
##                                race population ownership
## 3                               White     119609     owner
## 4                   African American      11615     owner
## 5 American Indian and Alaska Native        350     owner
## 6                               Asian       1807     owner
## 7 Native Hawaiian, Pacific Islander         57     owner
## 8                              Others        337     owner
```

Some statistics:

```
house_stat <- HouseData %>% group_by(ownership) %>% summarise(total_population = sum(population),
    max_population = max(population), race_max_occupancy = race[which.max(population)],
    min_population = min(population), race_min_occupancy = race[which.min(population)])

head(house_stat)
```

```
## # A tibble: 2 x 6
##   ownership total_population max_population race_max_occupancy
##       <chr>            <dbl>          <dbl>              <chr>
## 1     owner           134922         119609              White
## 2    renter            43748          31035              White
## # ... with 2 more variables: min_population <dbl>,
## #   race_min_occupancy <chr>
```

Figure 1:

```
ggplot(HouseData, aes(race, population)) + geom_bar(aes(fill = ownership),
    stat = "identity", position = "dodge") + labs(title = "Occupancy of residence by race and ownership
    y = "Population") + theme(axis.text.x = element_text(angle = 45,
    hjust = 1))
```

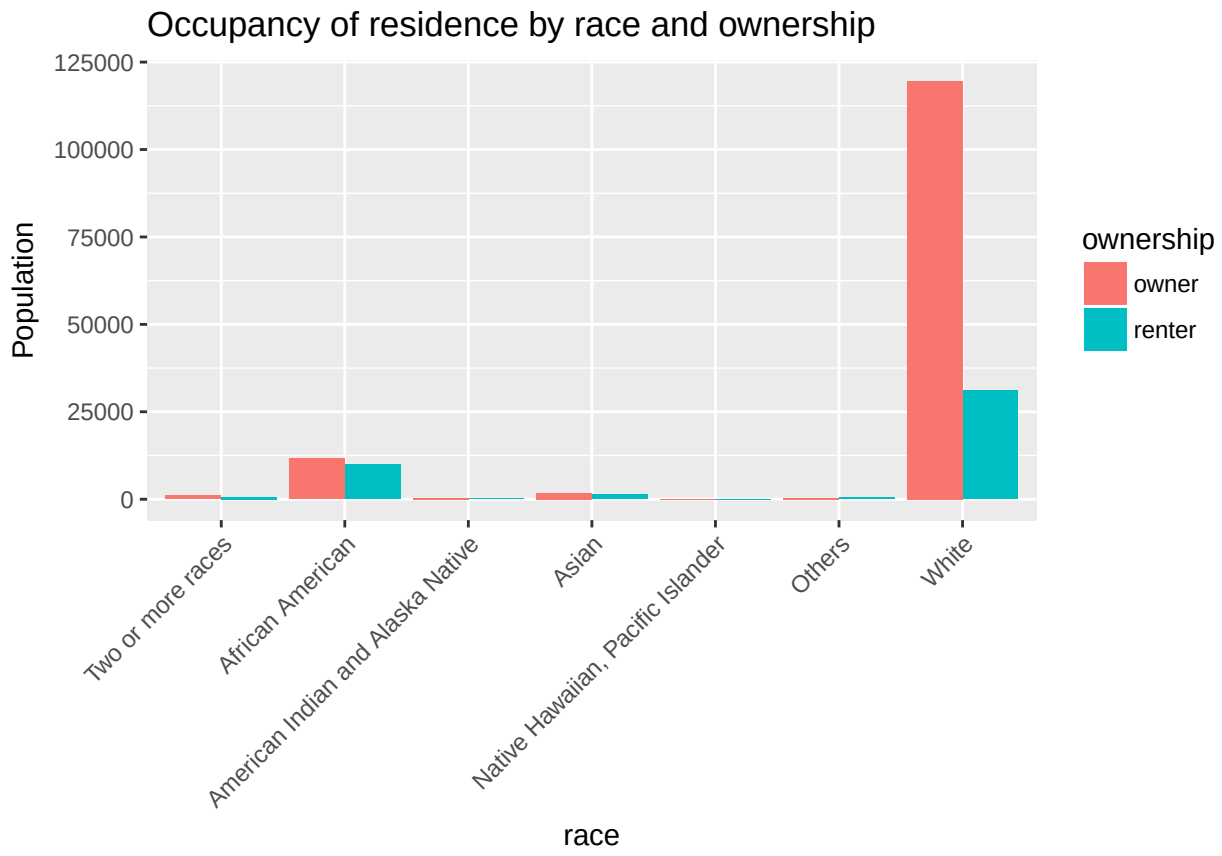

Figure 1 shows that the white population has bigger occupancy in both owned and rental properties.

spread function is used to untidy the data again:

```
HouseData2 <- spread(HouseData, 3, 2)
```

Figure 2:

```
ggplot(HouseData2, aes(race, owner/renter)) + geom_bar(aes(fill = race),
    stat = "identity", position = "dodge") + labs(title = "Occupancy of residence by race and ownership
    y = "owner renter ratio") + theme(axis.text.x = element_blank())
```

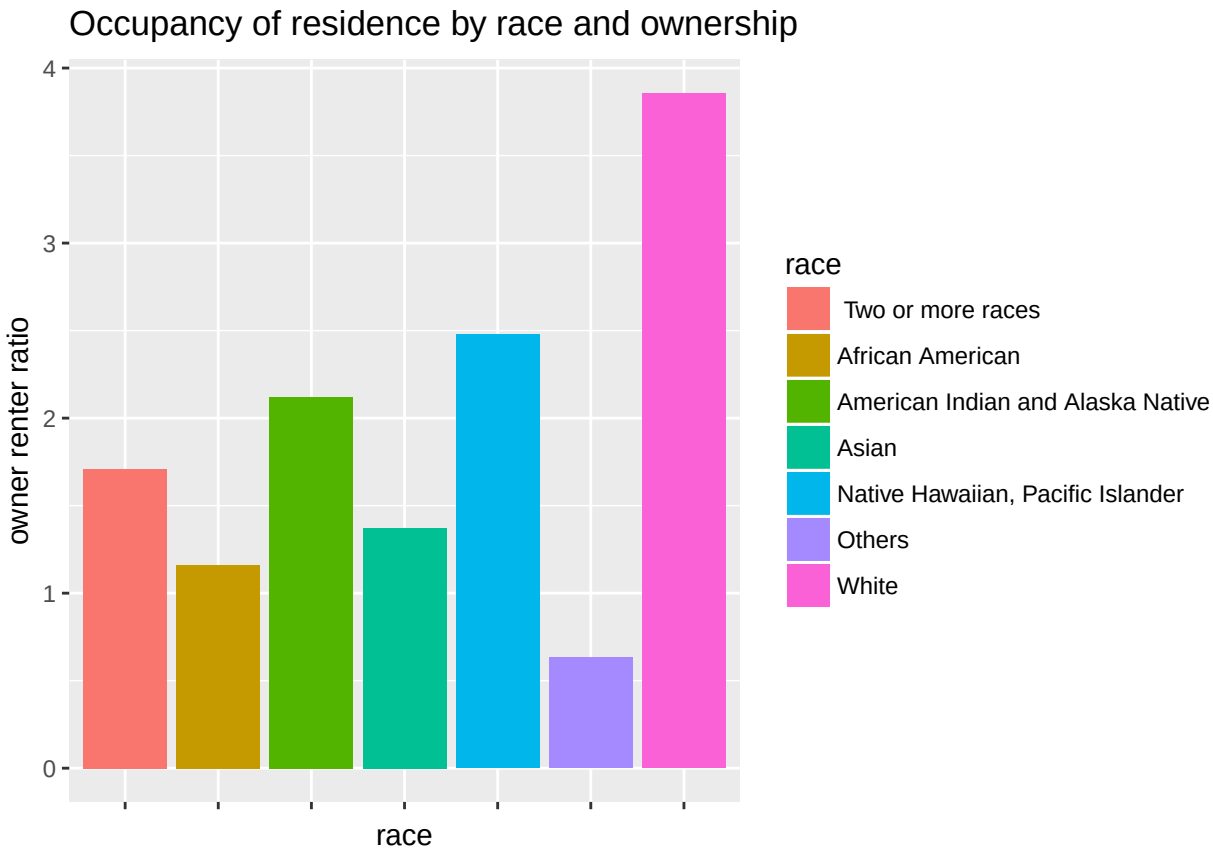

Figure 2 shows that the ratio of white population has more house ownership compared to other races.

Figure 3:

```
ggplot(HouseData2, aes(x = race, y = owner/sum(owner))) + geom_point(aes(color = race,
    size = owner)) + labs(title = "Ratio house ownership by race",
    y = "ownership ratio") + theme(axis.text.x = element_blank())
```

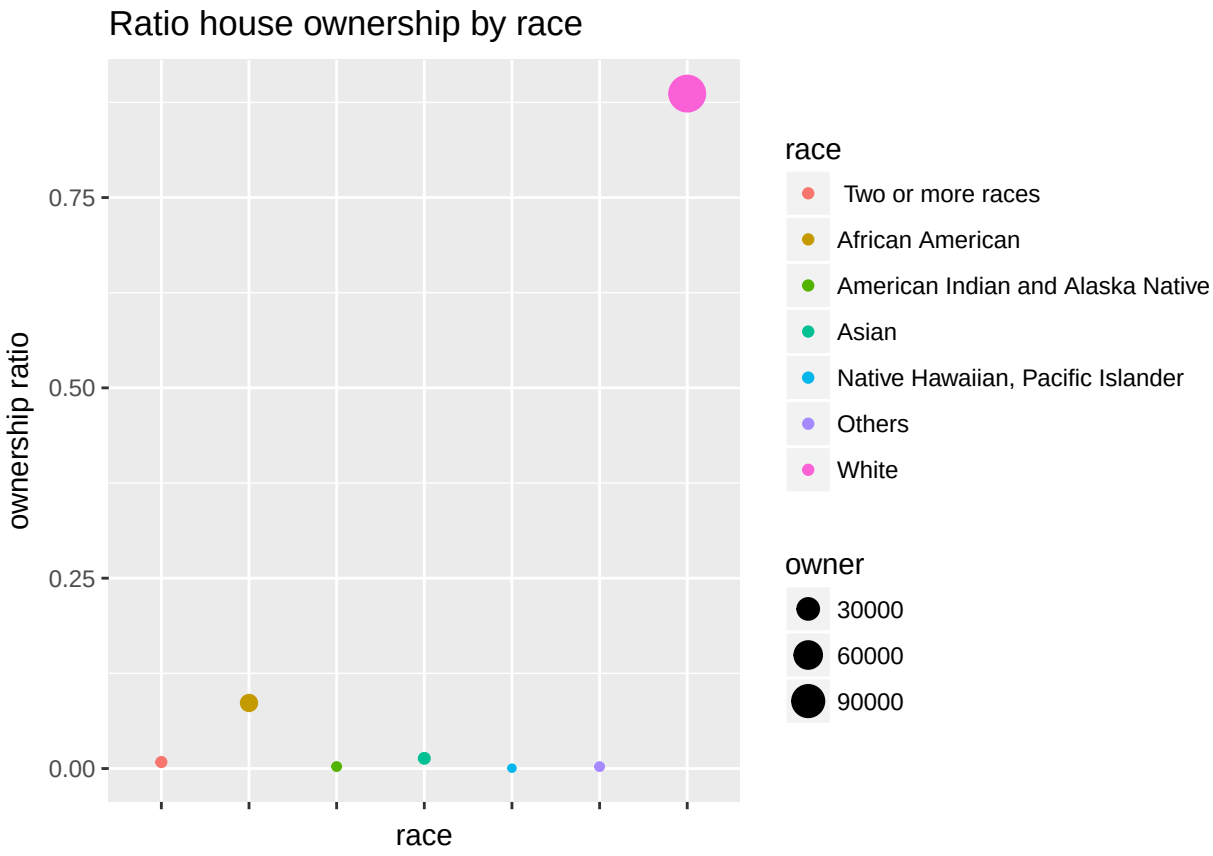

Figure 3 also suggest the white population has a bigger percentage of home ownership while african american population hold the distant second position.