Creative 1: Relationships + Distributions

Meghan Casey

September 14, 2020

Variables of Interest

Milwaukee, WI will be the city I analyze for this assignment.

Unit of Analysis: Census Tract

Population: TBD

1. Total Population of the tract (continuous): tot_pop
2. Percent of households owned (continuous): per_own
3. Majority race of the tract (categorical): maj_race
4. Median age of tract (continuous): med_age
5. Majority household type of the tract (categorical): maj_hh
6. Average Household size of the tract (continuous): avg_hh

Load Libraries

```
library(ggplot2)
library(tidycensus)
library(tidyverse)
library(ggthemes)
```

Get Data

```
area_vars_2010 <- load_variables(2010, "sf1")
```

Loading Milwaukee County Tract Level Data from the 2010 Census

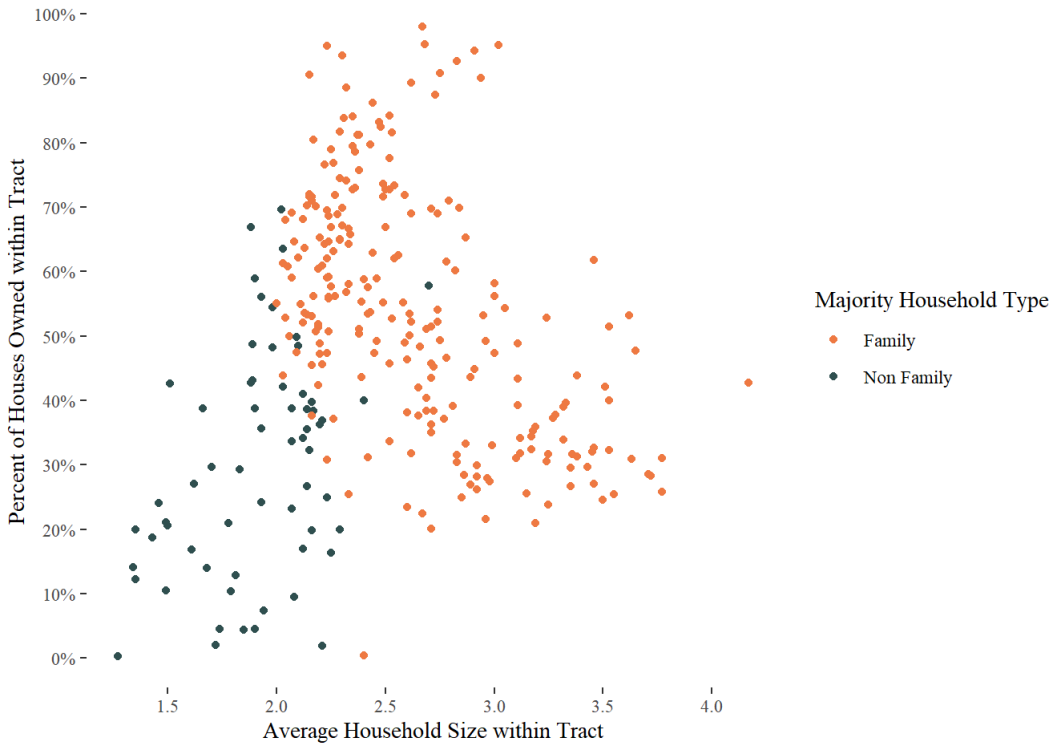
```
vars <- c(tot_pop = "P001001",
          white =   "P008003",
          black =   "P008004",
          native =  "P008005",
          asian =   "P008006",
          pac_isl = "P008007",
          other =   "P008008",
          multi =   "P008010",

          med_age = "P013001",

          avg_hh =  "H012001",

          tot_ten =  "H004001",
          own_mort = "H004002",
          own_free = "H004003",
          rent =     "H004004",

          tot_hh =    "P018001",
          family_hh = "P018002",
          nonfam_hh = "P018007")

mke_tracts <- get_decennial(geography = "tract", variables = vars,
                            state = "WI", county = "Milwaukee",
                            output = "wide") %>%

  mutate(per_own = (own_mort + own_free) / tot_ten) %>%

  mutate(maj_race = case_when(
    white/tot_pop > 0.5 ~ "white",
    black/tot_pop > 0.5 ~ "black",
    native/tot_pop > 0.5 ~ "native",
    asian/tot_pop > 0.5 ~ "asian",
    pac_isl/tot_pop > 0.5 ~ "pac_isl",
    other/tot_pop > 0.5 ~ "other",
    multi/tot_pop > 0.5 ~ "multi",
    (white + black + native + asian +
       asian + pac_isl + other + multi) / tot_pop < 0.5 ~ "other",
    TRUE ~ "None")) %>%

  mutate(maj_hh = case_when(
     family_hh / tot_hh > 0.5 ~ "family",
    nonfam_hh / tot_hh > 0.5 ~ "nonfamily")) %>%

  filter(avg_hh > 1.00) %>%

  select(GEOID, tot_pop, maj_race, maj_hh, med_age, per_own, avg_hh)
```

Plotting Data

Plot #1

Average Household Size and Percent of Houses Owned by Majority Household Type

```
ggplot(mke_tracts,
       aes(x = avg_hh,
           y = per_own,
           color = maj_hh)) +
  geom_point() +
  scale_x_continuous(name = "Average Household Size within Tract",
                     breaks = seq(0, 4.5, by = 0.5)) +
  scale_y_continuous(name = "Percent of Houses Owned within Tract",
                     breaks = seq(0, 1, by = 0.1),
                     labels = paste(seq(0, 100, by = 10), "%", sep = "")) +
  scale_color_manual(values = c("sienna2", "darkslategrey"),
                     name = "Majority Household Type",
                     labels = c("Family", "Non Family")) +
  theme_tufte()
```



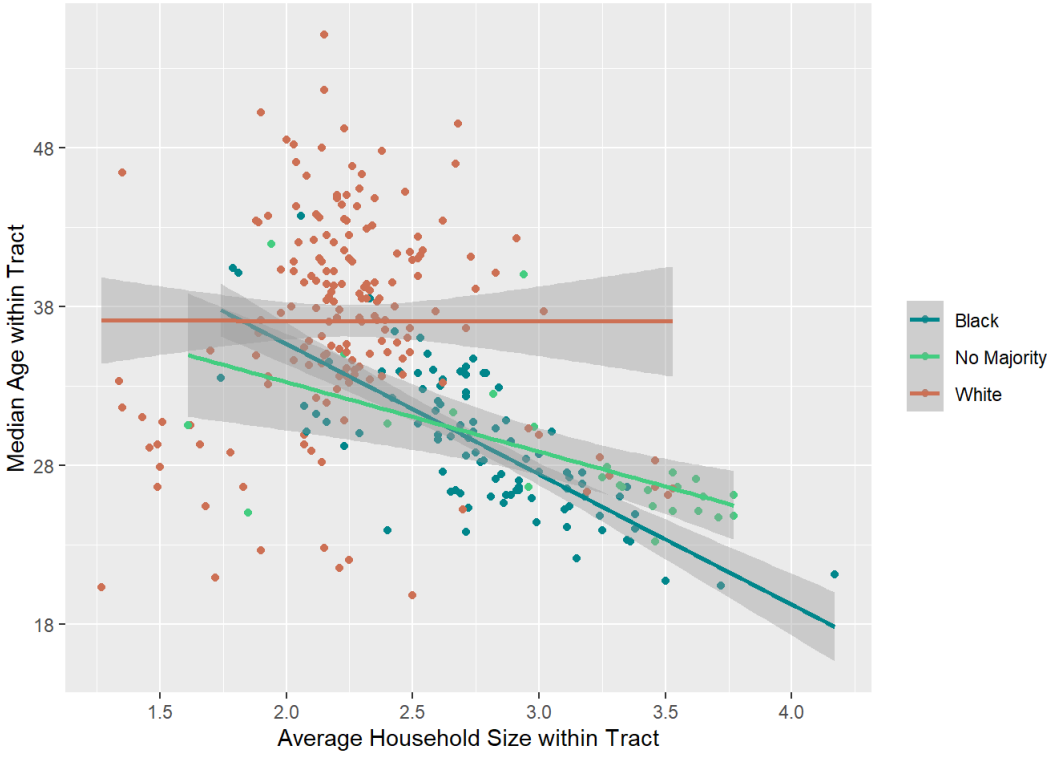
In general, majority family household tracts consist of more owned homes, with more people living within those homes.

Plot #2

Average Household Size and Median Age by Majority Race

```
ggplot(mke_tracts,
       aes(x = avg_hh,
           y = med_age,
           color = maj_race)) +
  geom_point() +
  stat_smooth(method = "lm") +
  scale_x_continuous(name = "Average Household Size within Tract",
                     breaks = seq(0, 4.5, by = 0.5)) +
  scale_y_continuous(name = "Median Age within Tract",
                     breaks = seq(18, 58, by = 10)) +
  scale_color_manual(values = c("turquoise4", "seagreen3", "salmon3"),
                     name = element_blank(),
                     labels = c("Black",
                                "No Majority",
                                "White")) +
  scale_fill_discrete(name = element_blank(),
                      labels = c("Black",
                                 "No Majority",
                                 "White")) +
  theme_gray()
```

```
## `geom_smooth()` using formula 'y ~ x'
```



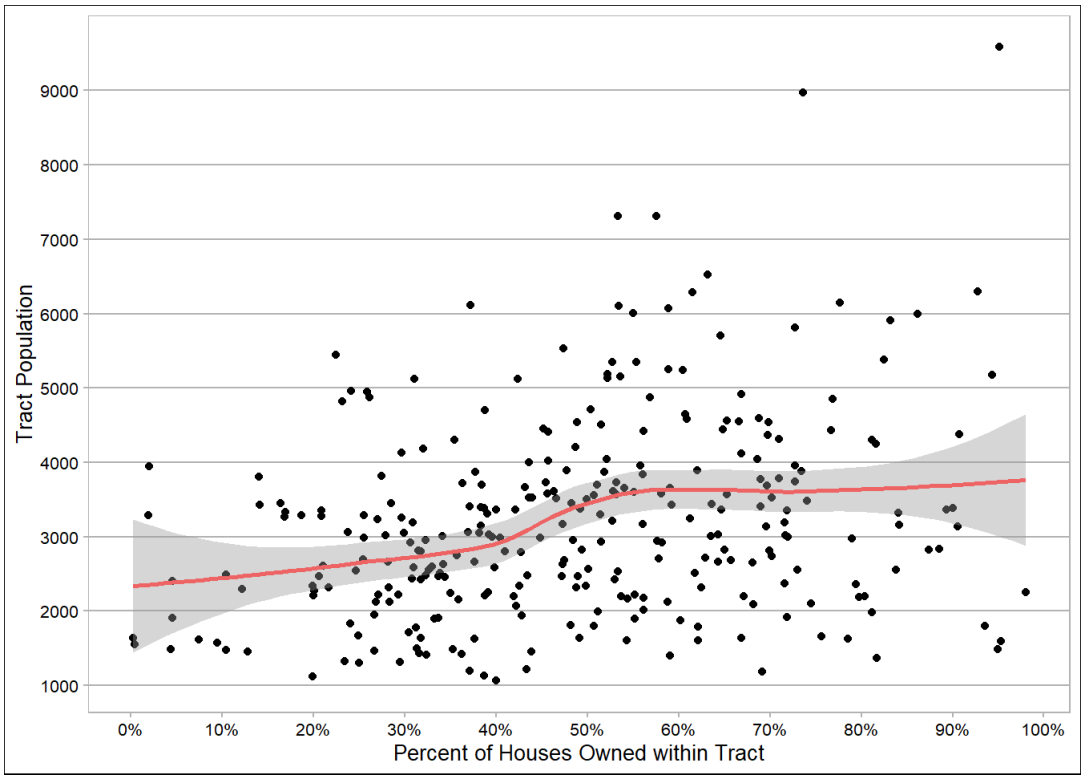
The grouping of several tracts with larger households size and younger age could relate to rental populations among students or young professionals.

Plot #3

Percent of Houses Owned and Tract Population

```
ggplot(mke_tracts,
       aes(x = per_own,
           y = tot_pop)) +
  geom_point() +
  stat_smooth(color = "indianred2") +
  scale_x_continuous(name = "Percent of Houses Owned within Tract",
                     breaks = seq(0, 1, by = 0.1),
                     labels = paste(seq(0, 100, by = 10), "%", sep = "")) +
  scale_y_continuous(name = "Tract Population",
                     breaks = seq(1000, 9000, by = 1000)) +
  theme_calc()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



The clustering of tract data points in the middle section might indicate that there is not a strong relationship between the percent of houses owned within a tract and that tract's population.

Plot #4

Median Age and Tract Population

```
ggplot(mke_tracts,
       aes(x = med_age,
           y = tot_pop)) +
  geom_point(color = "grey33", alpha = 0.5, size = 0.5) +
  stat_smooth(color = "darksalmon", size = 1) +
  scale_x_continuous(name = "Median Age of Tract",
                     breaks = seq(18, 58, by = 10),
                     labels = c("58 or 18", "28", "38", "48", "58")) +
  scale_y_continuous(name = "Tract population",
                     trans = "log",
                     breaks = c(1000, 3000, 5000, 7000, 9000),
                     labels = c("1,000", "3,000", "5,000", "7,000", "9,000")) +
  coord_polar(theta = "x") +
  theme_hc()
```

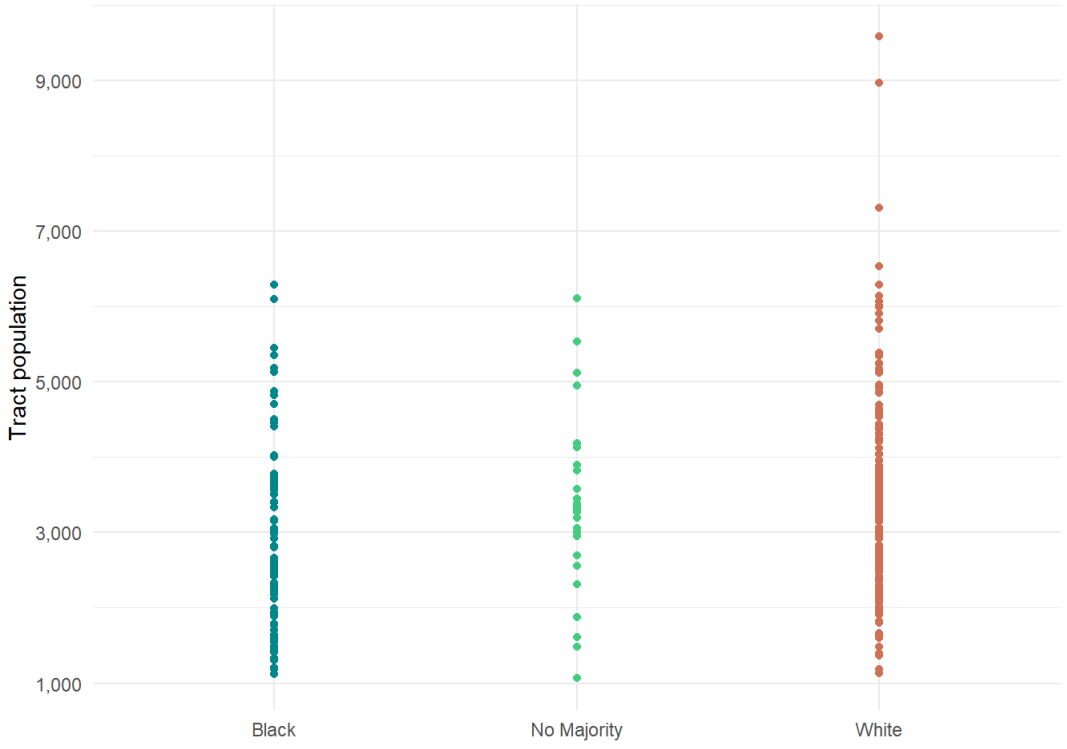
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Again, not seeing a strong correlation between median age of tract and the tract population.

Plot #5

Majority Race and Tract Population

```
ggplot(mke_tracts,
       aes(x = maj_race,
           y = tot_pop,
           color = maj_race)) +
  geom_point() +
  scale_x_discrete(name = element_blank(),
                   labels = c("Black",
                              "No Majority",
                              "White")) +
  scale_y_continuous(name = "Tract population",
                     breaks = c(1000, 3000, 5000, 7000, 9000),
                     labels = c("1,000", "3,000", "5,000", "7,000", "9,000")) +
  scale_color_manual(values = c("turquoise4", "seagreen3", "salmon3")) +
  theme_minimal() +
  theme(legend.position = "none")
```

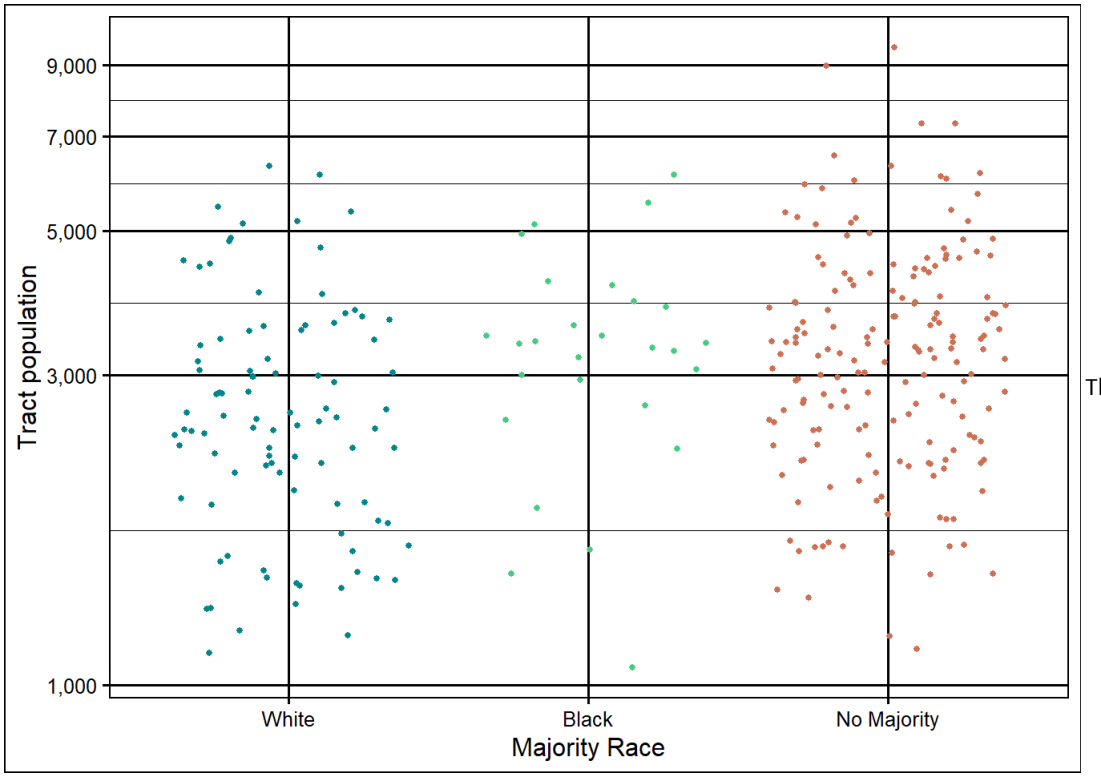


The distribution seems more or less even, with a few outliers in the White section.

Plot #6

Majority Race and Tract Population

```
ggplot(mke_tracts,
       aes(x = maj_race,
           y = tot_pop,
           color = maj_race)) +
  geom_point(position = "jitter",
             size = 1.0,
             alpha = 1.0) +
  scale_x_discrete(name = "Majority Race",
                       labels = c("White",
                                  "Black",
                                  "No Majority")) +
  scale_y_continuous(name = "Tract population",
                     trans = "log",
                     breaks = c(1000, 3000, 5000, 7000, 9000),
                     labels = c("1,000", "3,000", "5,000", "7,000", "9,000")) +
  scale_color_manual(values = c("turquoise4", "seagreen3", "salmon3")) +
  theme_foundation() +
  theme(legend.position = "none")
```

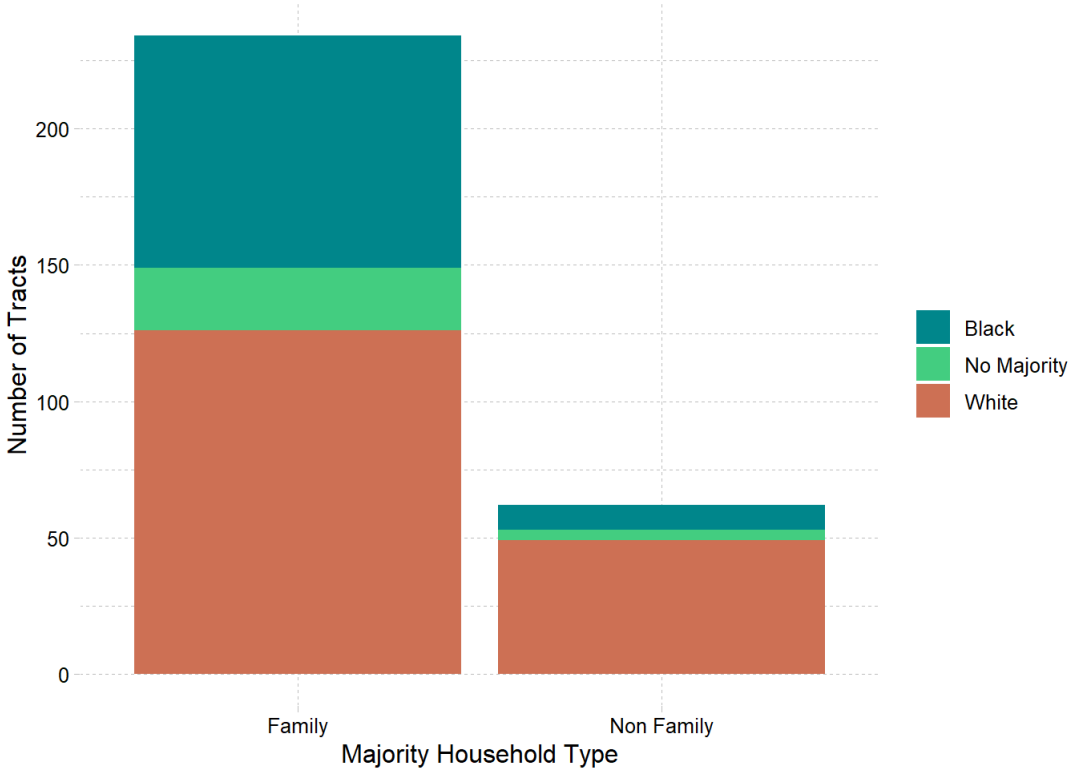


This is a similar plot to the previous one, but also shows the range of majority percentages within each racial category.

Plot #7

Majority Household Type and Number of Tracts by Majority Race

```
ggplot(mke_tracts,
       aes(x = maj_hh,
           fill = maj_race)) +
  geom_bar() +
  scale_y_continuous(name = "Number of Tracts") +
  scale_x_discrete(name = "Majority Household Type",
                   labels = c("Family",
                              "Non Family")) +
  scale_fill_manual(values = c("turquoise4", "seagreen3", "salmon3"),
                    name = element_blank(),
                    labels = c("Black",
                               "No Majority",
                               "White")) +
  theme_pander()
```

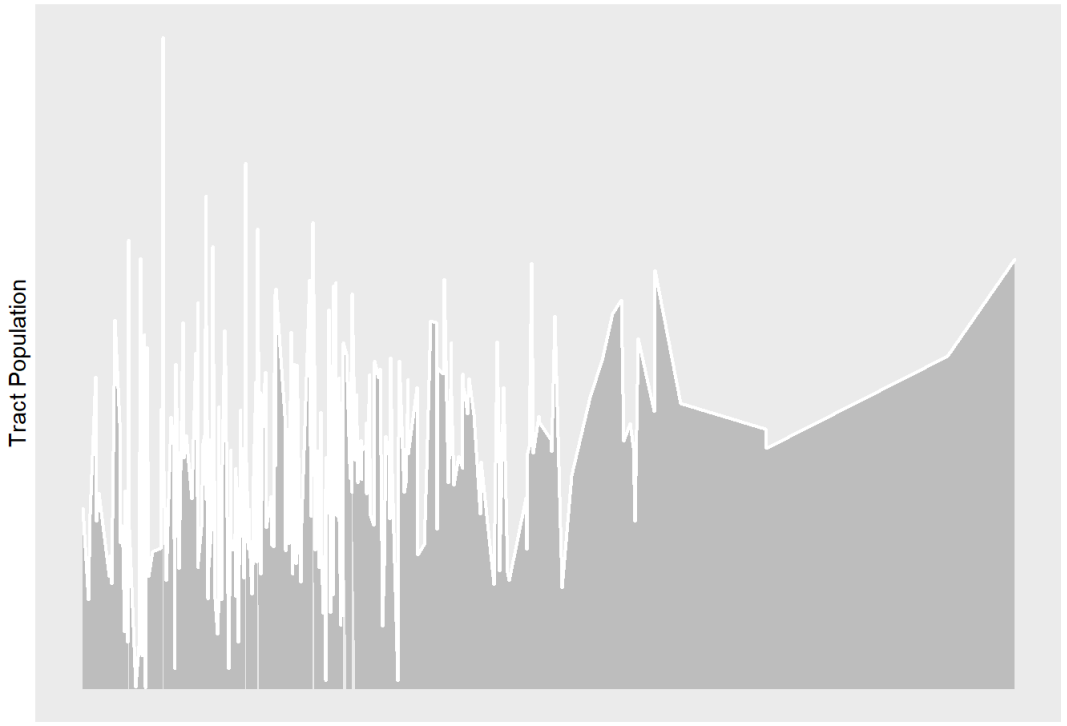


This is an effective way to show the racial majorities in each household type.

Plot #8

Percent of Houses Owned and Tract Population

```
ggplot(mke_tracts,
       aes(x = tot_pop,
           y = per_own)) +
  geom_area(alpha = 0.25, color = "white", size = 0.75) +
  scale_y_continuous(name = "Tract Population",
                     breaks = seq(1000, 9000, by = 1000)) +
  scale_x_continuous(name = "Percent of Houses Owned within Tract",
                     breaks = seq(0, 1, by = 0.1),
                     labels = paste(seq(0, 100, by = 10), "%", sep = "")) +
  theme_gray()
```

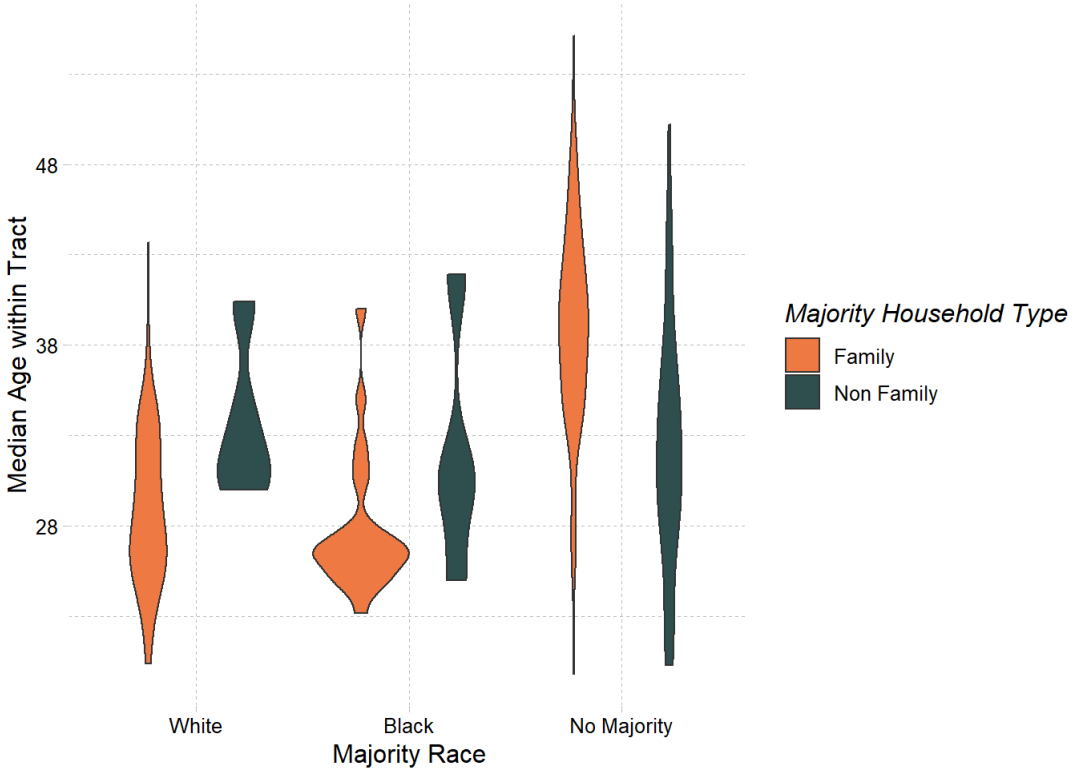


This plot doesn't do a great job of indicating a relationship between tract population and percent of houses owned, there is a lot of variety.

Plot #9

Majority Race and Median Age by Majority Household Type

```
ggplot(mke_tracts,
       aes(x = maj_race,
           y = med_age,
           fill = maj_hh)) +
  geom_violin(scale = "area") +
  scale_x_discrete(name = "Majority Race",
                       labels = c("White",
                                  "Black",
                                  "No Majority")) +
  scale_y_continuous(name = "Median Age within Tract",
                     breaks = seq(18, 58, by = 10)) +

  scale_fill_manual(values = c("sienna2", "darkslategrey"),
                     name = "Majority Household Type",
                     labels = c("Family", "Non Family")) +

  theme_pander()
```



There is a lot more variation in no majority tracts - could be interesting to see what other variables this is true for.

Plot #10

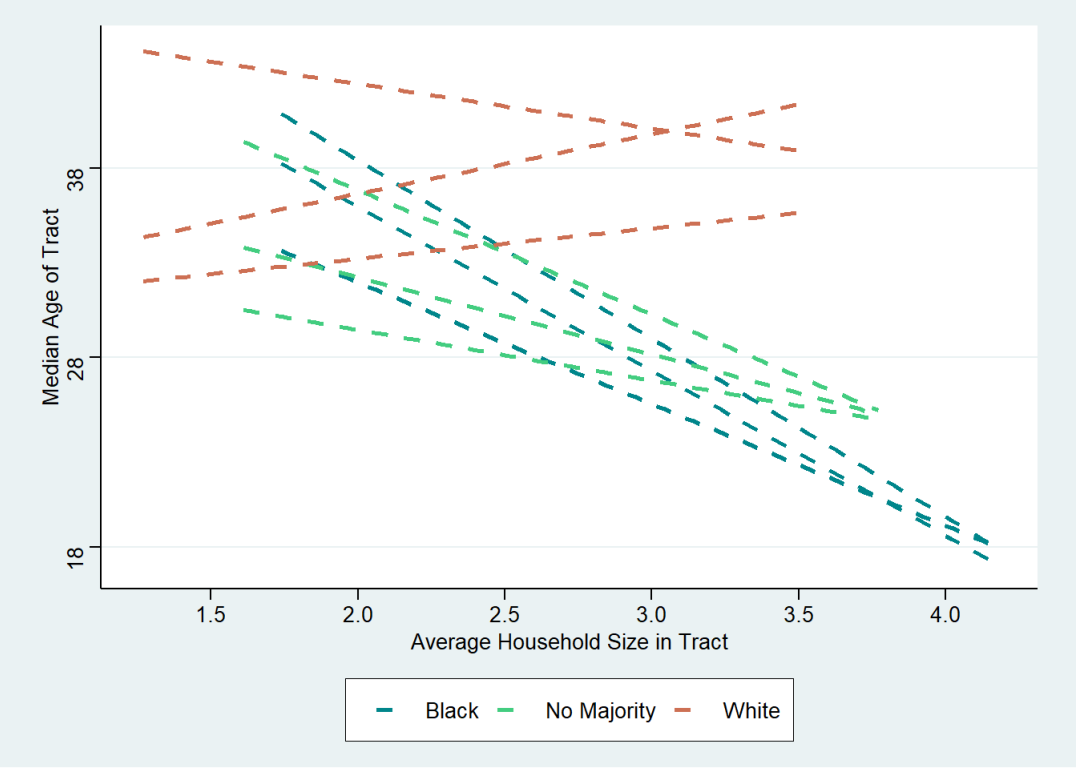
Average Household Size and Median Age by Majority Race

```
ggplot(mke_tracts,
       aes(x = avg_hh,
           y = med_age,
           color = maj_race)) +
  geom_quantile(linetype = 2, size = 1.0) +
  scale_x_continuous(name = "Average Household Size in Tract",
                     breaks = seq(0, 4.5, by = 0.5)) +
  scale_y_continuous(name = "Median Age of Tract",
                     breaks = seq(8, 68, by = 10),
                     labels = c("8", "18", "28", "38", "48", "58", "68")) +
 scale_color_manual(values = c("turquoise4", "seagreen3", "salmon3"),
                     name = element_blank(),
                     labels = c("Black",
                                "No Majority",
                                "White")) +
  theme_stata()
```

```
## Smoothing formula not specified. Using: y ~ x
## Smoothing formula not specified. Using: y ~ x
```

```
## Warning in rq.fit.br(wx, wy, tau = tau, ...): Solution may be nonunique
```

```
## Smoothing formula not specified. Using: y ~ x
```

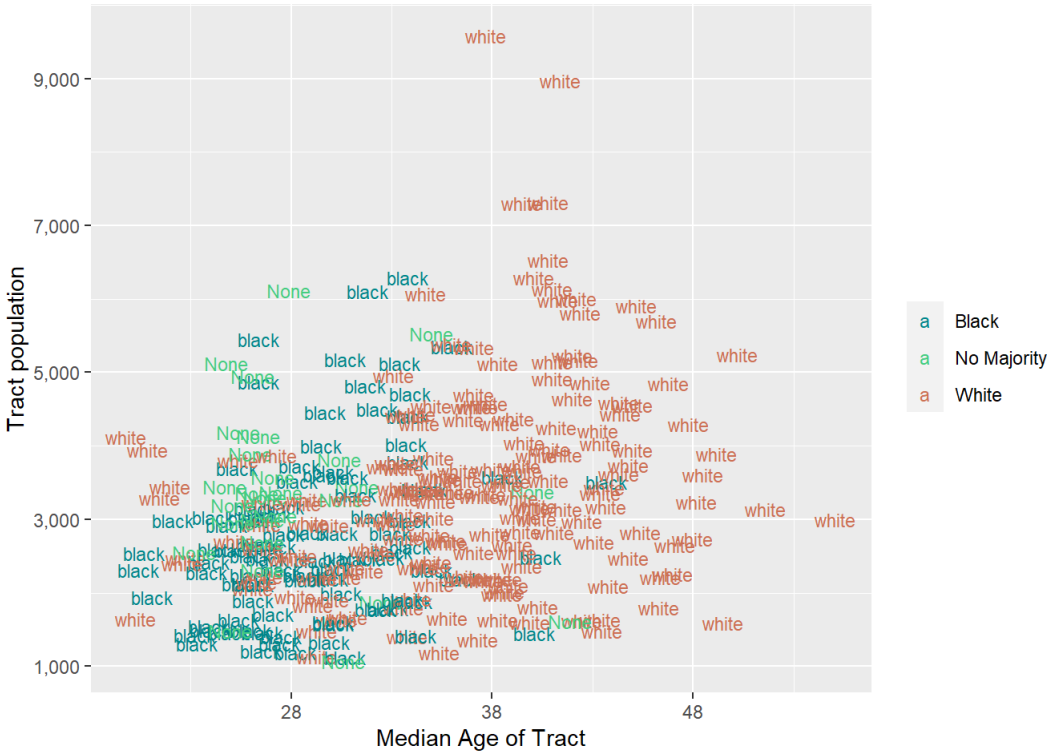


Like an earlier plot, the trend towards larger households for younger median ages could be indicative of student populations, curious that the white category does not follow this trend.

Plot #11

Median Age and Tract Population by Majority Race

```
ggplot(mke_tracts,
       aes(x = med_age,
           y = tot_pop,
           label = maj_race,
           color = maj_race)) +
  geom_text(size = 3) +
  scale_x_continuous(name = "Median Age of Tract",
                     breaks = seq(8, 68, by = 10),
                     labels = c("8", "18", "28", "38", "48", "58", "68")) +
  scale_y_continuous(name = "Tract population",
                     breaks = c(1000, 3000, 5000, 7000, 9000),
                     labels = c("1,000", "3,000", "5,000", "7,000", "9,000")) +
  scale_color_manual(values = c("turquoise4", "seagreen3", "salmon3"),
                     name = element_blank(),
                     labels = c("Black",
                                "No Majority",
                                "White")) +
  theme_grey()
```

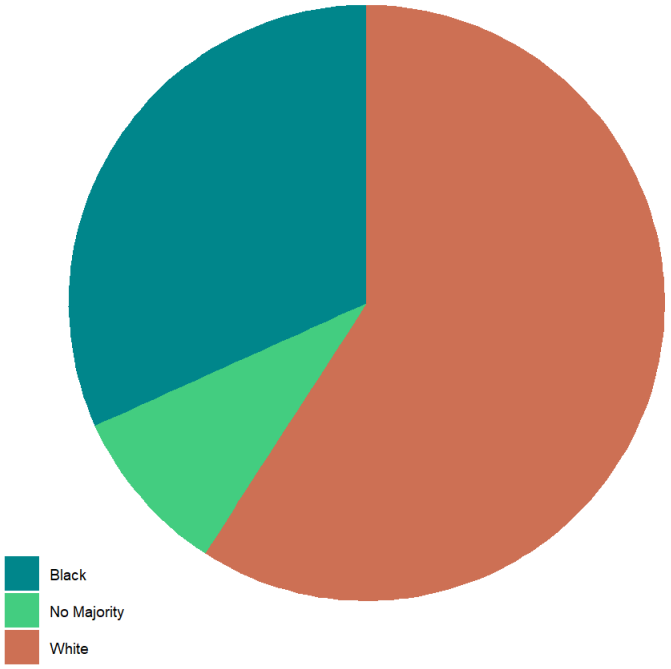


An interesting take on a scatterplot!

Plot #12

Majority Race

```
pie_chart <- ggplot(mke_tracts, aes(x = 1, fill = maj_race)) +
  geom_bar(position = "fill") +
  coord_polar(theta = "y") +
  scale_y_continuous(name = "Number of tracts") +
  scale_fill_manual(values = c("turquoise4", "seagreen3", "salmon3"),
                    name = element_blank(),
                    labels = c("Black",
                               "No Majority",
                               "White")) +
  theme_map()

pie_chart
```



Very clear way to see racial majorities across tracts.

Save Dataset

```
write_csv(mke_tracts, "mke_tracts.csv")
```