VIS Creative Assignment 1

Julia Meinhardt

9/12/2020

```
library(ggplot2)
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──
```

```
## ✓ tibble  3.0.3     ✓ dplyr   1.0.2
## ✓ tidyr   1.1.2     ✓ stringr 1.4.0
## ✓ readr   1.3.1     ✓ forcats 0.5.0
## ✓ purrr   0.3.4
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(tidycensus)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
library(extrafont)
```

```
## Registering fonts with R
```

```
library(xkcd)
```

```
vars_2010 <- load_variables(2010, "sf1")
```

Variables

For this assignment, I'm pulling my variable set directly from Carole's written tutorial and changing my county of observation to Monterey County, CA. As noted in the written tutorial, variables are pulled from the 2010 Census and are as follows:

1.Urban or rural (categorical) 2.Majority race/ethnicity (categorical) 3.Average household size (continuous) 4.Percent of households who live in rented homes (continuous) 5.Population of the tract (continuous) 6.Median age of residents (continuous)

```
vars <- c(tot_hh = "H001001",
         tot_rural_hh = "H002005",
         tot_rent_hh = "H004004",
         avg_hh_size = "H012001",
         tot_pop = "P003001",
         nh_white_pop = "P005003",
         nh_black_pop = "P005004",
         nh_asian_pop = "P005006",
         hispanic_pop = "P005010",
         med_age = "P013001")

Monterey_tracts <- get_decennial(geography = "tract", variables = vars,
                           state = "CA", county = "Monterey",
                           output = "wide") %>%
  mutate(rural = (tot_rural_hh == tot_hh) ,
         pct_rental = tot_rent_hh / tot_hh,
         majority_race_eth = case_when(
           nh_white_pop / tot_pop > 0.5 ~ "nh_white",
           nh_black_pop / tot_pop > 0.5 ~ "nh_black",
           nh_asian_pop / tot_pop > 0.5 ~ "nh_asian",
           hispanic_pop / tot_pop > 0.5 ~ "hispanic",
           (nh_white_pop + nh_black_pop + nh_asian_pop + hispanic_pop) /
             tot_pop < 0.5 ~ "other",
           TRUE ~ "None")) %>%
  filter(tot_pop > 0, tot_hh > 0) %>%
  select(GEOID, rural, majority_race_eth,
         pct_rental, avg_hh_size, tot_pop, med_age)
```

```
## Getting data from the 2010 decennial Census
```

```
## Using Census Summary File 1
```

Figure 1: Scatterplot

I started with a simple scatter plot. I learned from **Jennifer's** assignment on github how to add title and subtitle labels. She also told me about the viridis package, which I downloaded and used later on!

```
ggplot(Monterey_tracts,
       aes(x = med_age,
           y = pct_rental,
           color = rural)) +
  geom_point()+
  stat_smooth(method = "lm", se = FALSE)+
  scale_y_continuous(name = "Percent Households in Rented Homes",
                     breaks = seq(0, 1, by = 0.20),
                     labels = paste(seq(0, 100, by = 20), "%", sep = "")) +
  scale_x_continuous(name = "Median age within tract",
                     breaks = seq(10, 80, by = 10)) +
  scale_color_discrete(name = "",
                       labels = c("Urban Tract",
                                  "Rural Tract")) +
   labs(title = 'Scatter Plot of Median Age and Percent of Households in Rented Units',
       subtitle = 'Monterey County, CA')+

theme_minimal()
```

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

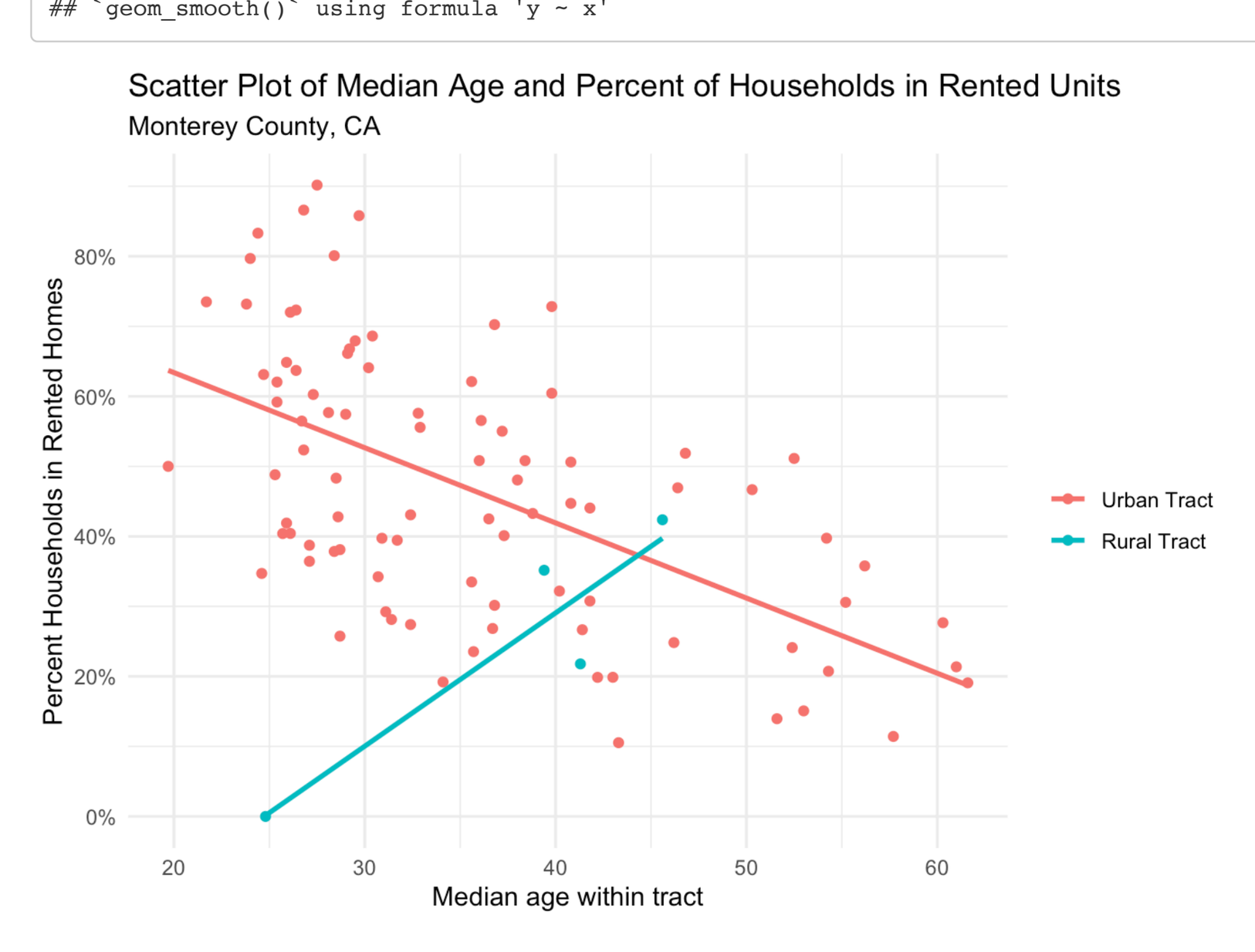


Figure 2: Total Population Histogram

```
ggplot(Monterey_tracts,
       aes(x = tot_pop)) +
  geom_histogram(binwidth = 500, color = "black", fill = "light blue") +
  scale_x_continuous(name = "Total Population",
                     breaks = seq(0, 12000, by = 1000)) +
  scale_y_continuous(name = "Number of Tracts",
                     breaks = seq(0, 20, by = 1)) +
  labs(title = 'Population Histogram',
       subtitle = 'Monterey County, CA')+

  theme_light()
```

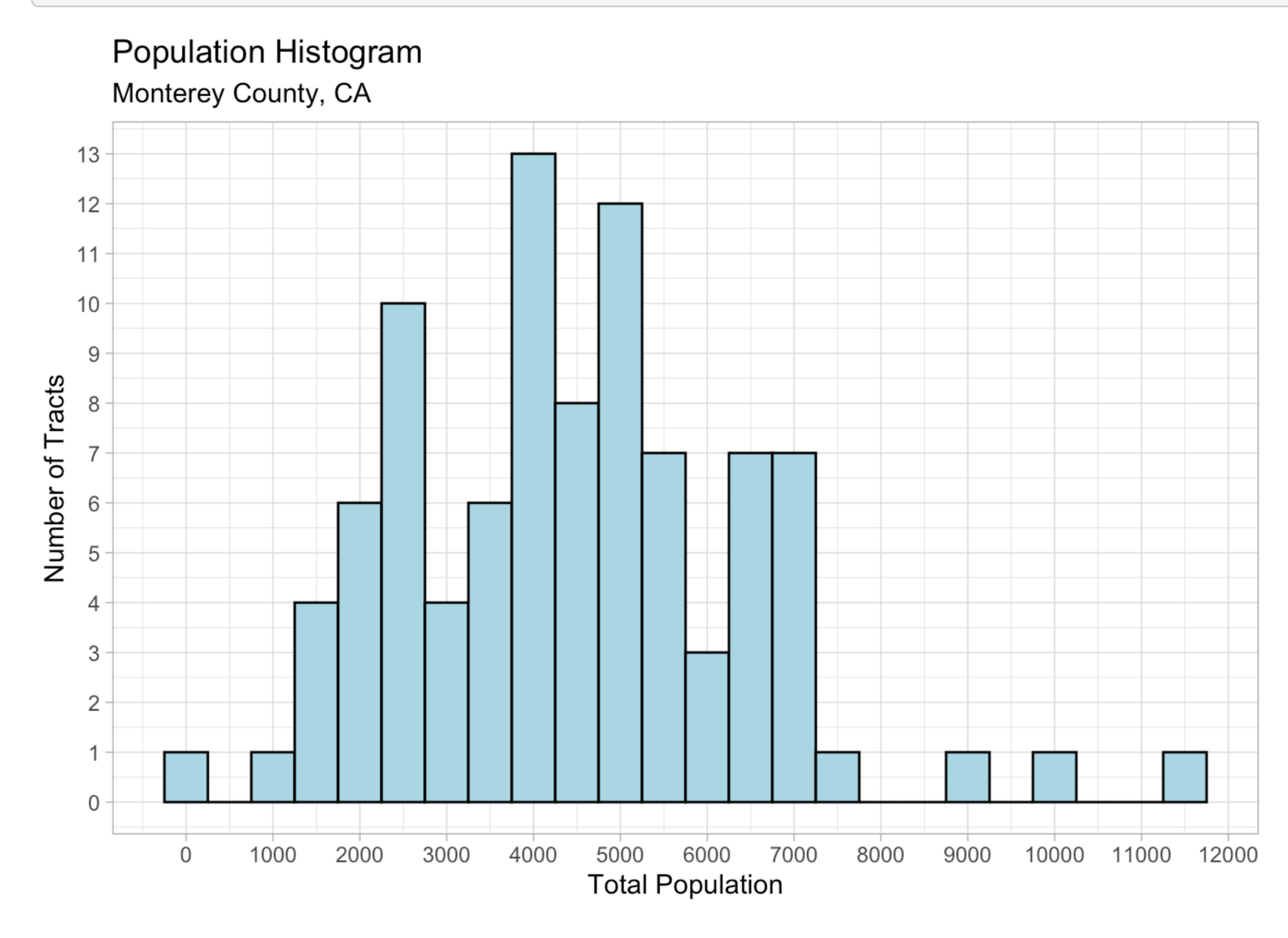


Figure 3:Average Household Histogram

```
ggplot(Monterey_tracts,
       aes(x = avg_hh_size)) +
geom_histogram(aes(fill = ..count..), binwidth = .5, color = "black", weight = 10) +

scale_x_continuous(name = "Average Household Size",
                   breaks = seq(0, 7, by = 1)) +
scale_y_continuous(name = "Count",
                   breaks = seq(0, 20, by = 2)) +
labs(title = 'Average Household Size',
     subtitle = 'Monterey County, CA')+

scale_fill_viridis()
```

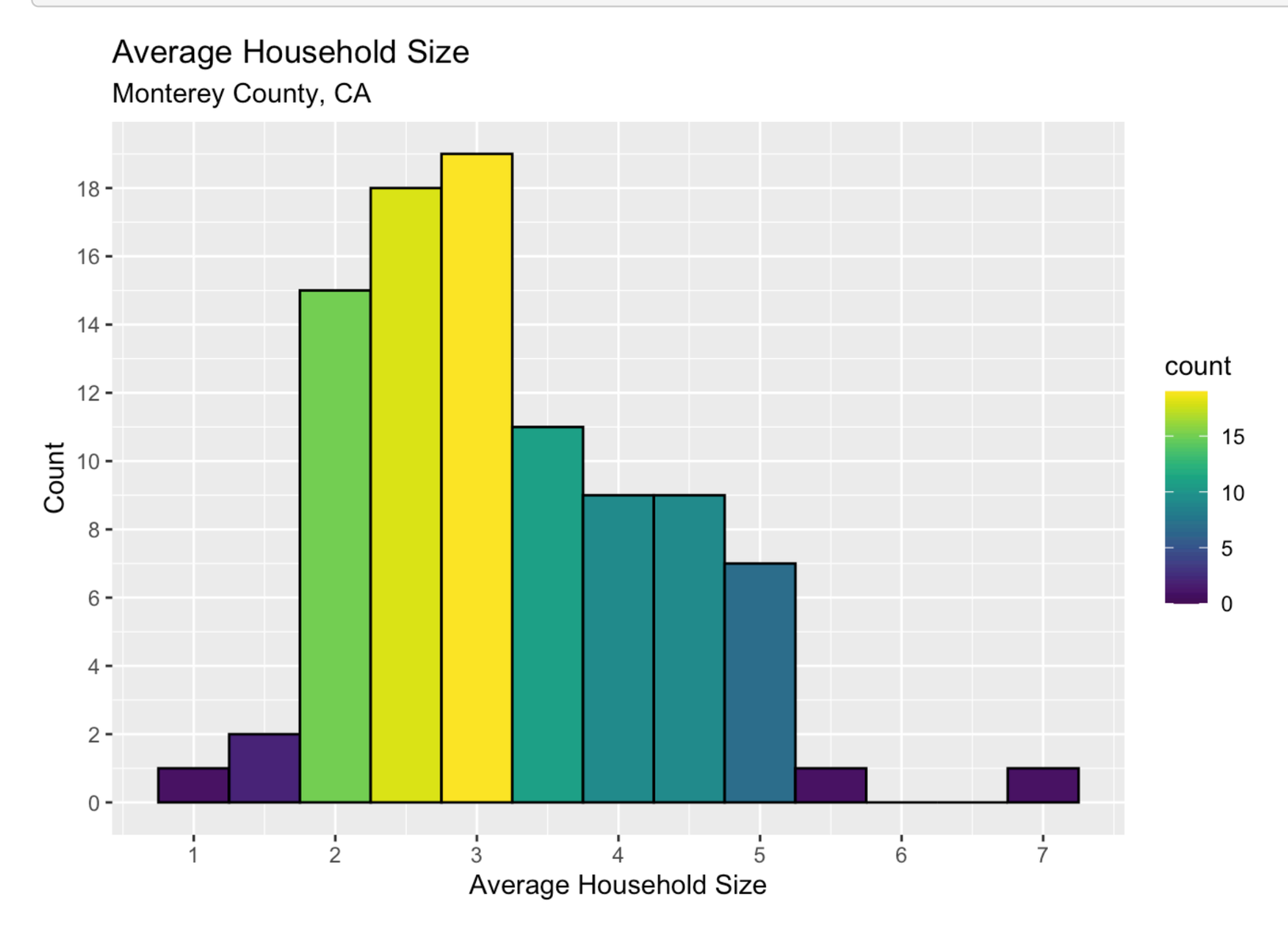


Figure 4: Boxplot

```
ggplot(Monterey_tracts,
       aes(x = majority_race_eth,
           y = med_age,
           fill = rural )) +
  geom_boxplot(linetype = 1, color = "black") +
  scale_x_discrete(name = "Majority Race/Ethnicity",
                   labels = c("Hispanic",
                              "Non-Hispanic white",
                              "No majority")) +
  scale_y_continuous(name = "Median Age",
                     breaks = seq(10, 80, by = 10)) +
  scale_fill_discrete (name = "",
                       labels = c("Urban Tract",
                                  "Rural Tract")) +
  labs(title = 'Majority Race/Ethnicity By Age and Tract Type',
    subtitle = 'Monterey County, CA')+


  theme_bw()
```

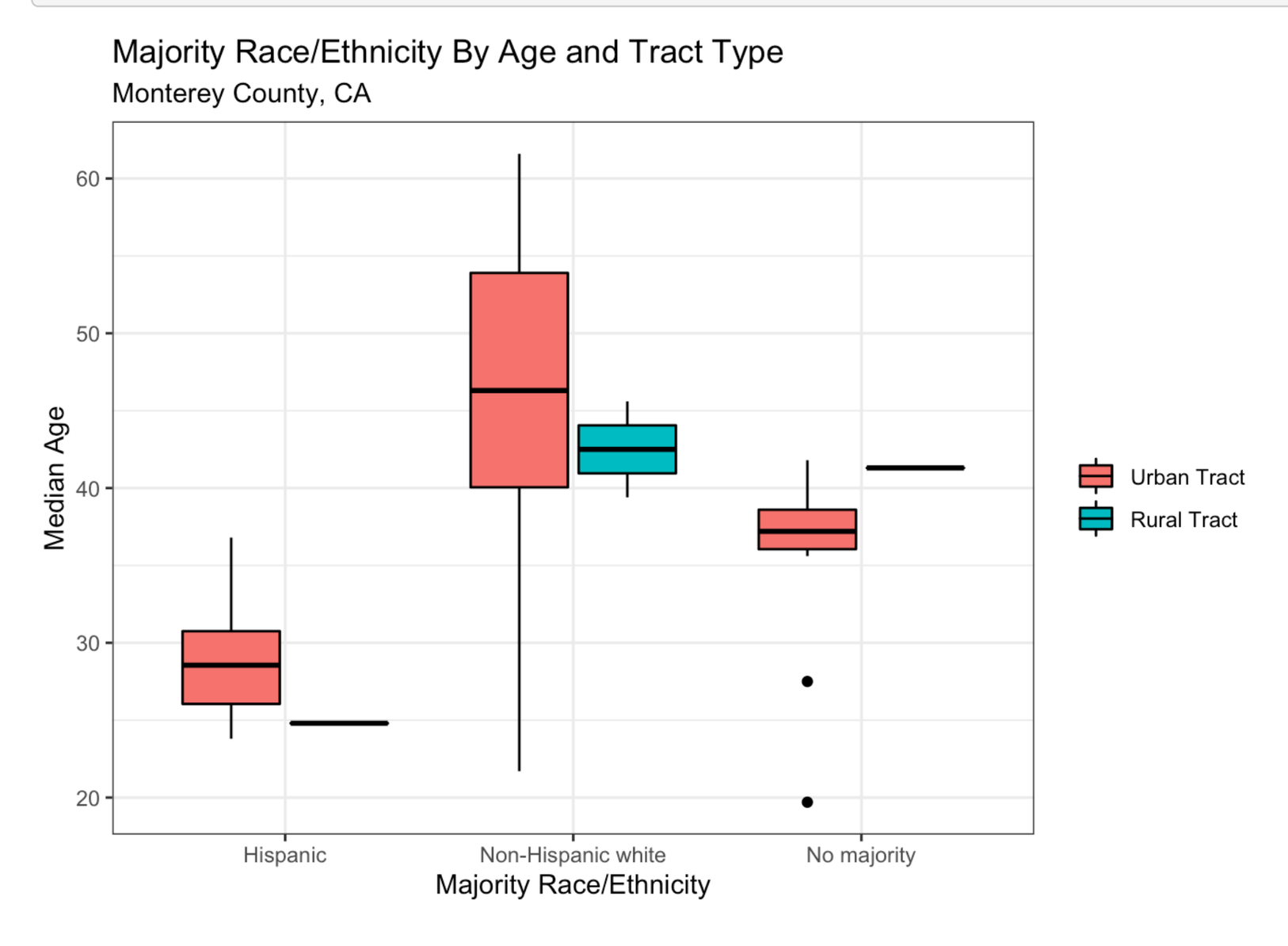


Figure 5: Density Map

While not a great or exceptionally clear graphic, this type of plot is intriguing to me and I would like to play with it further! I got the idea for this density plot from the cheat sheet Carole shared, with which I've been plugging in variously geoms and seeing what happens!

```
ggplot(Monterey_tracts)+
   geom_density2d(mapping = aes(x = med_age, y = pct_rental, color = rural))+
   scale_x_continuous(name = "Median Age") +
   scale_y_continuous(name = "Percent of Households in Rented Homes",
                      breaks = seq(0, 1, by = 0.20),
                      labels = paste(seq(0, 100, by = 20), "%", sep = "")) +
   scale_color_discrete(name = "",
                      labels = c("Urban Tract",
                                 "Rural Tract")) +
   labs(title = 'Majority Race/Ethnicity By Age and Tract Type',
        subtitle = 'Monterey County, CA')+

theme_bw()
```

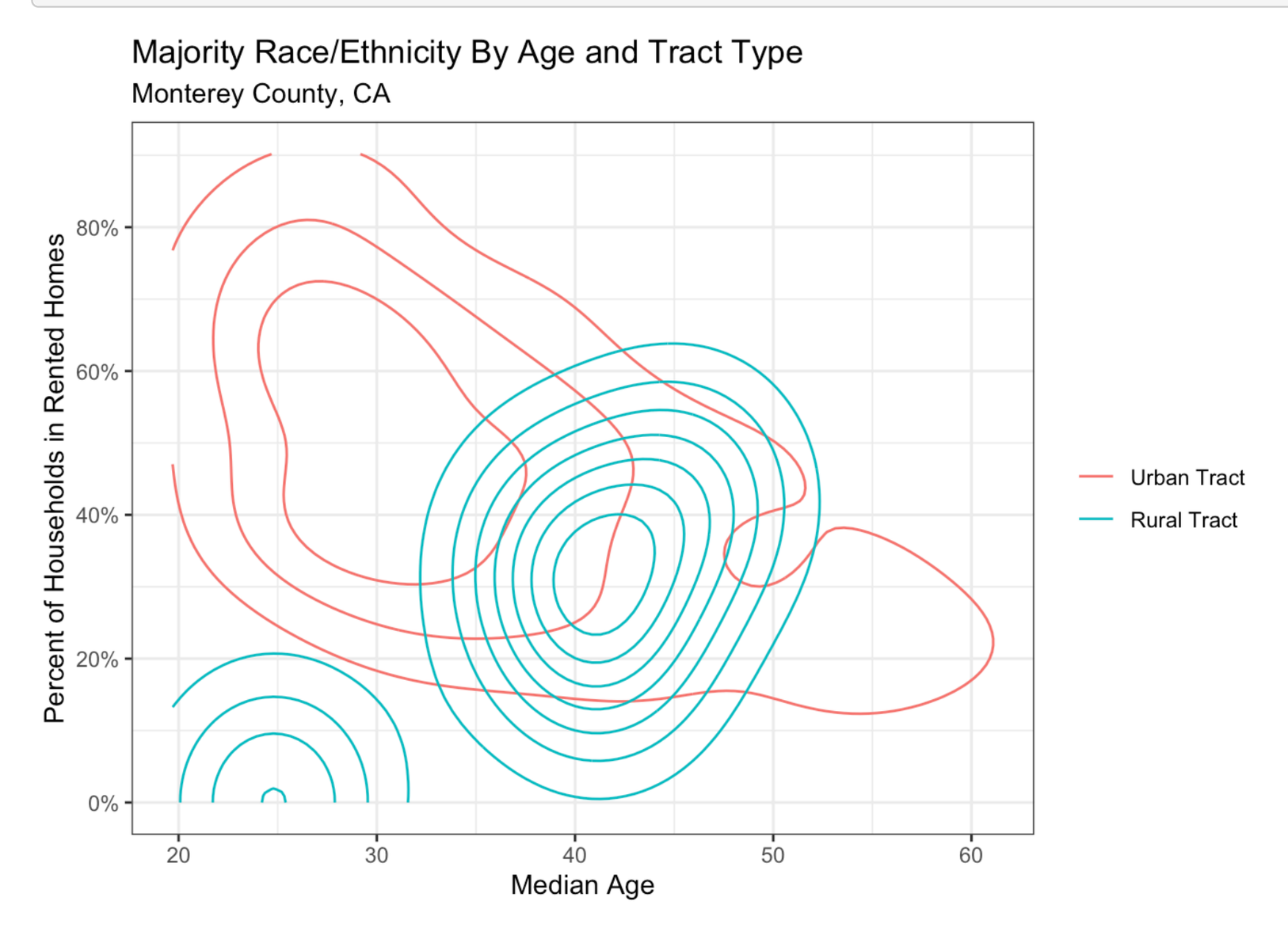
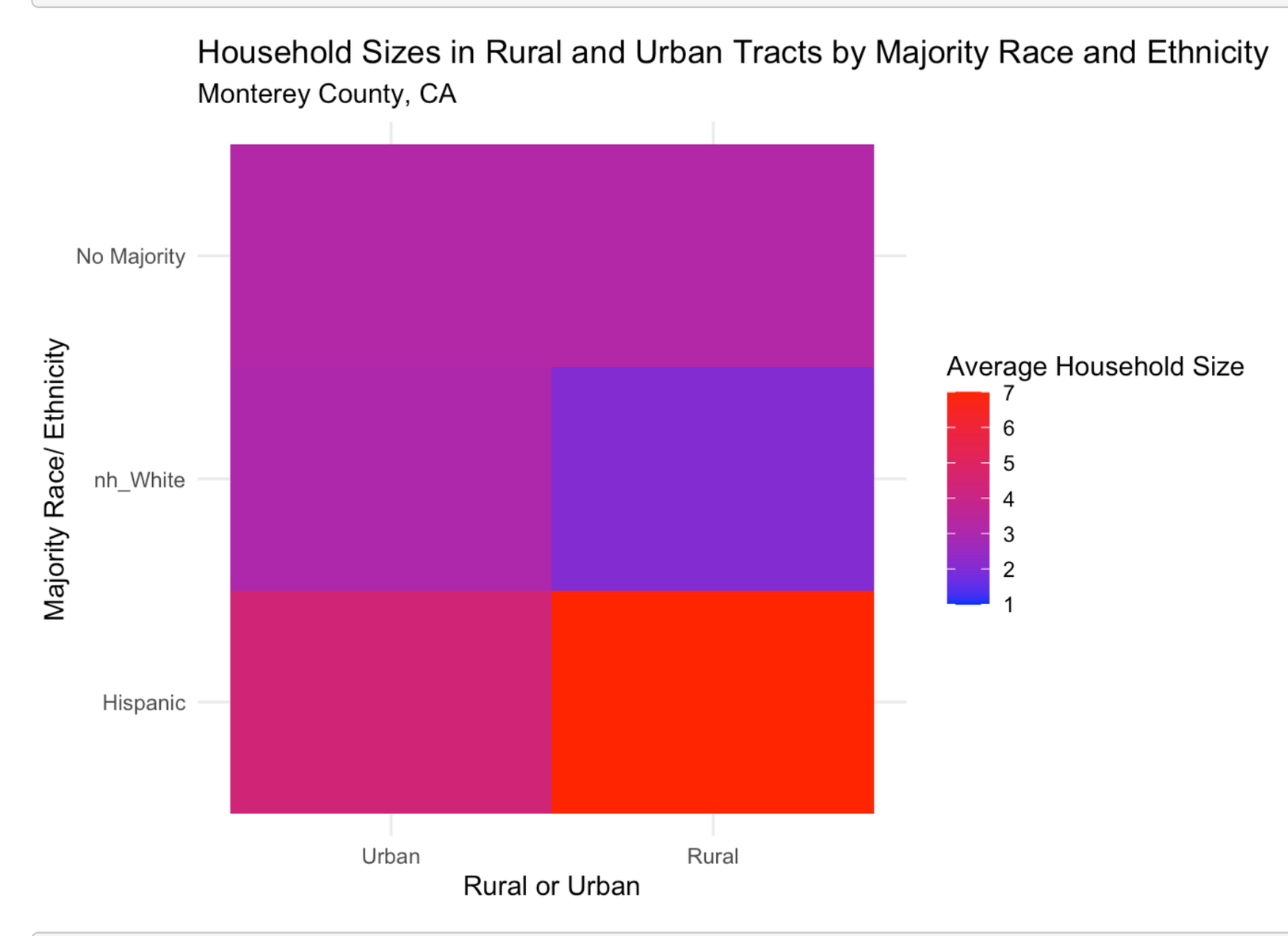


Figure 6: Tile

The got the idea for the tile graphic from **Taelor's** document on github! This is a graphic I would like to explore further with a larger dataset and more refined gradient.

```
ggplot(Monterey_tracts,
  aes(x = rural,
         y = majority_race_eth,
         fill = avg_hh_size))+
  geom_tile() +
  scale_x_discrete(name = "Rural or Urban",
                   labels = c("Urban",
                              "Rural",
                              "No data")) +
  scale_y_discrete(name = "Majority Race/ Ethnicity",
                   labels = c("Hispanic",
                              "nh_White",
                              "No Majority")) +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(fill = "Average Household Size",
       title = "Household Sizes in Rural and Urban Tracts by Majority Race and Ethnicity",
       subtitle = 'Monterey County, CA')+

  theme_minimal()
```




```
palette.pals()
```

```
##  [1] "R3"              "R4"               "ggplot2"         "Okabe-Ito"
##  [5] "Accent"          "Dark 2"           "Paired"          "Pastel 1"
##  [9] "Pastel 2"        "Set 1"            "Set 2"           "Set 3"
## [13] "Tableau 10"      "Classic Tableau"  "Polychrome 36"   "Alphabet"
```

Figure 7: Violin Plot

```
ggplot(Monterey_tracts,
       aes(x = rural,
           y = med_age,
           color = rural)) +
  geom_violin() +
  geom_point(position = "jitter",
             alpha = 0.5,
             size = 3) +
  scale_x_discrete(name = "Rural or Urban",
                   labels = c("Urban",
                              "Rural")) +
  scale_y_continuous(name = "Median Age",
                     breaks = seq(10, 80, by = 10)) +
   labs(title = "Median Ages in Rural and Urban Tracts",
       subtitle = 'Monterey County, CA')+
  theme_bw()
```

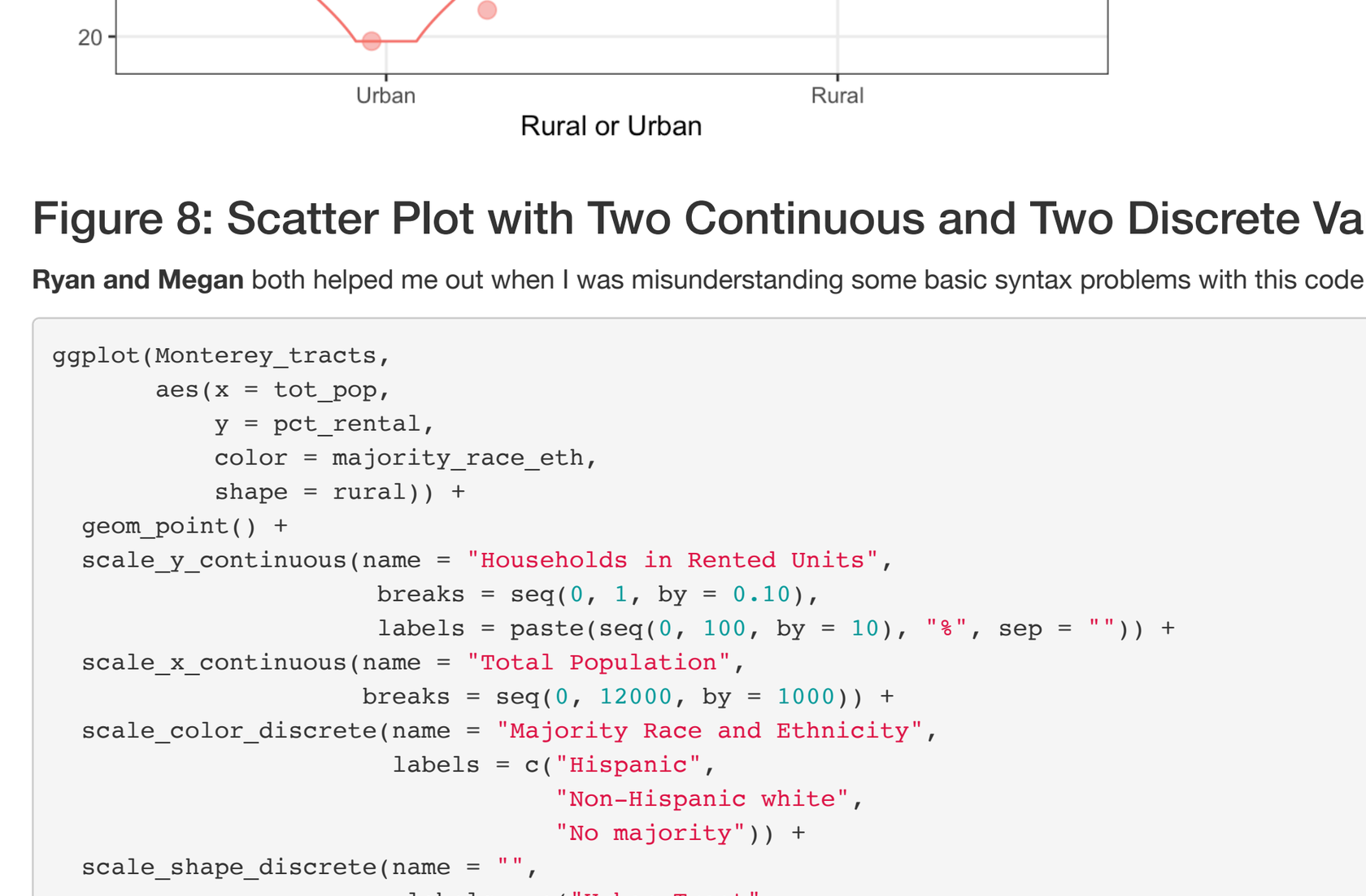


Figure 8: Scatter Plot with Two Continuous and Two Discrete Variables

Ryan and Megan both helped me out when I was misunderstanding some basic syntax problems with this code chunk.

```
ggplot(Monterey_tracts,
       aes(x = tot_pop,
           y = pct_rental,
           color = majority_race_eth,
           shape = rural)) +
  geom_point() +
  scale_y_continuous(name = "Households in Rented Units",
                     breaks = seq(0, 1, by = 0.10),
                     labels = paste(seq(0, 100, by = 10), "%", sep = "")) +
  scale_x_continuous(name = "Total Population",
                     breaks = seq(0, 12000, by = 1000)) +
  scale_color_discrete(name = "Majority Race and Ethnicity",
                       labels = c("Hispanic",
                                  "Non-Hispanic white",
                                  "No majority")) +
  scale_shape_discrete(name = "",
                       labels = c("Urban Tract",
                                  "Rural Tract")) +
  labs(title = "Percentage of Households in Rented Units",
       subtitle = 'Monterey County, CA')+

  theme_bw()
```

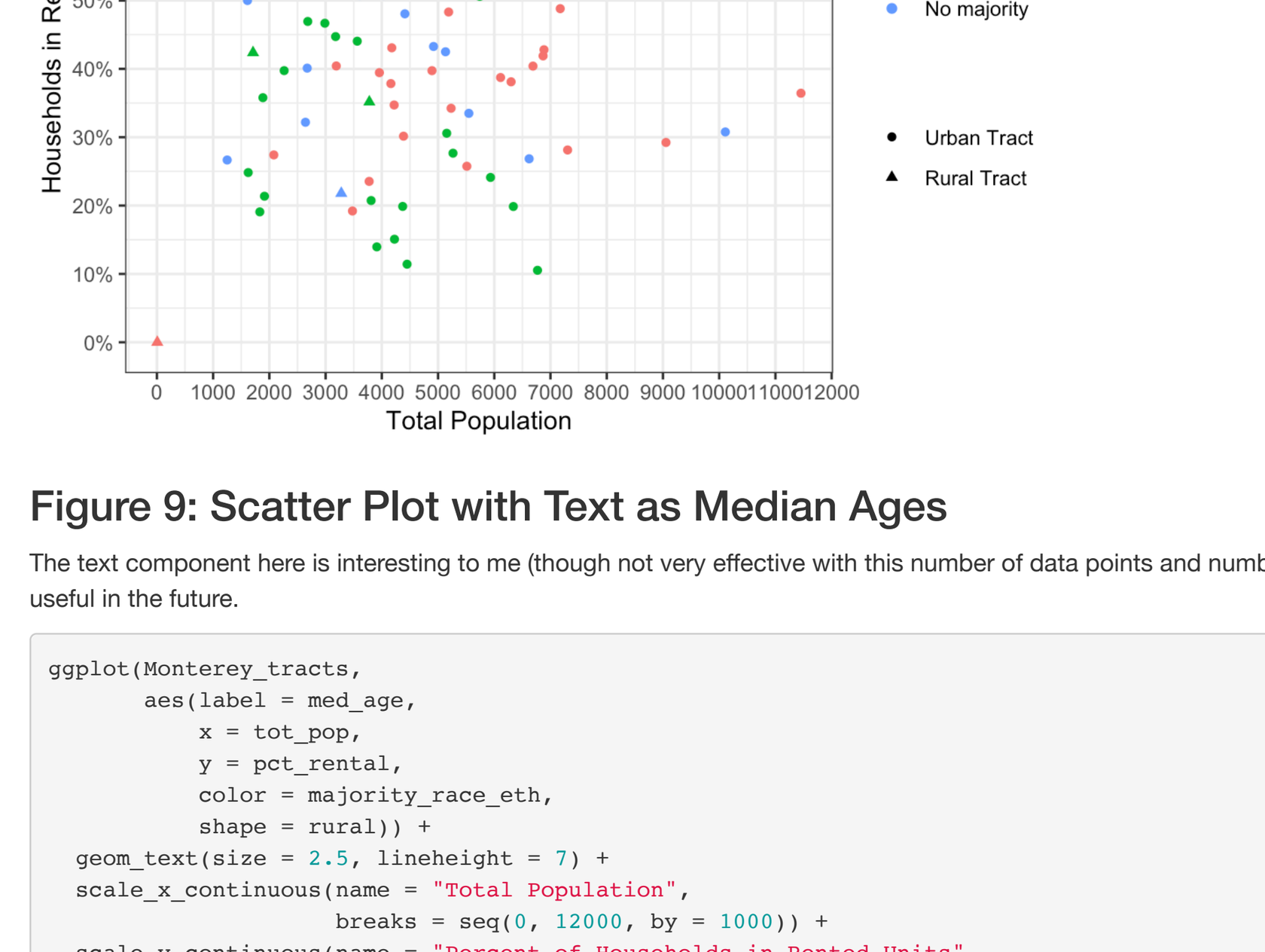


Figure 9: Scatter Plot with Text as Median Ages

The text component here is interesting to me (though not very effective with this number of data points and number of variables), but it could be useful in the future.

```
ggplot(Monterey_tracts,
       aes(label = med_age,
           x = tot_pop,
           y = pct_rental,
           color = majority_race_eth,
           size = rural)) +
  geom_text(size = 2.5, lineheight = 7) +
  scale_x_continuous(name = "Total Population",
                     breaks = seq(0, 12000, by = 1000)) +
  scale_y_continuous(name = "Percent of Households in Rented Units",
                     breaks = seq(0, 1, by = 0.20),
                     labels = paste(seq(0, 100, by = 20), "%", sep = "")) +
  scale_color_discrete(name = "Majority Race and Ethnicity",
                       labels = c("Hispanic",
                                  "Non-Hispanic white",
                                  "No majority")) +
  labs(subtitle = 'Monterey County, CA')+

  theme_classic()
```

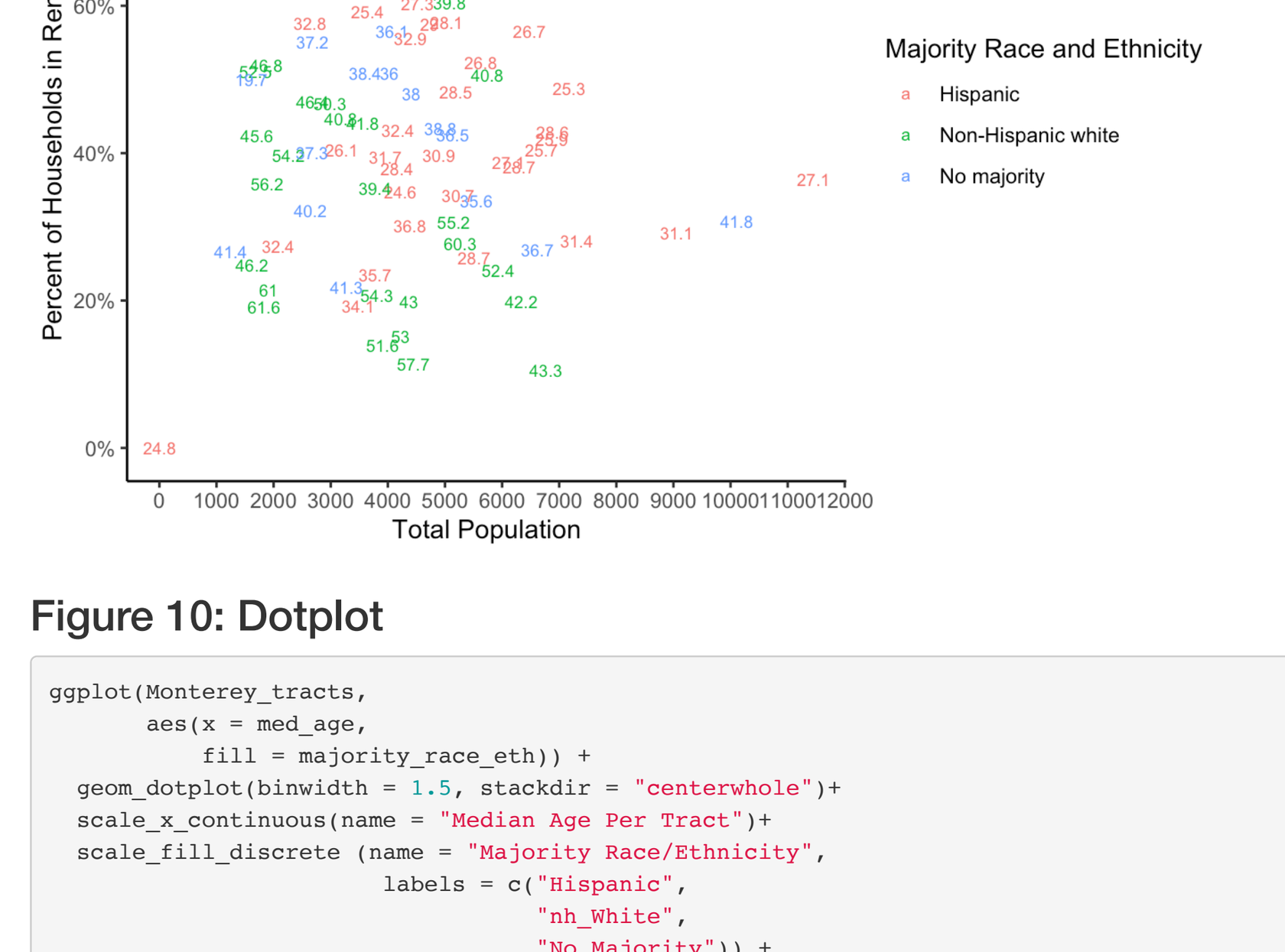


Figure 10: Dotplot

```
ggplot(Monterey_tracts,
       aes(x = med_age,
       fill = majority_race_eth)) +
  geom_dotplot(binwidth = 1.5, stackdir = "centerwhole")+
  scale_x_continuous(name = "Median Age Per Tract")+
  scale_fill_discrete (name = "Majority Race/Ethnicity",
                       labels = c("Hispanic",
                                  "nh_White",
                                  "No Majority")) +
  labs(title = 'Median Ages',
    subtitle = 'Monterey County, CA')+

  theme_bw()
```