

U6614 Assignment 5:

Water shutoffs, race, and income in Detroit:

Community impact by race and income

SAMPLE SOLUTION

2024-03-01

Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by the posted due date.

Remember to think carefully about what code you include in your knitted document. Only include code chunks that you need to generate the plots and statistics to answer the questions below. Don't include code from your working R script (that we started in class) that was only used to inspect and validate your results, and isn't necessary to answer the questions.

Load libraries

```
library(readstata13)
library(tidyverse)
library(lubridate)
library(weights)
```

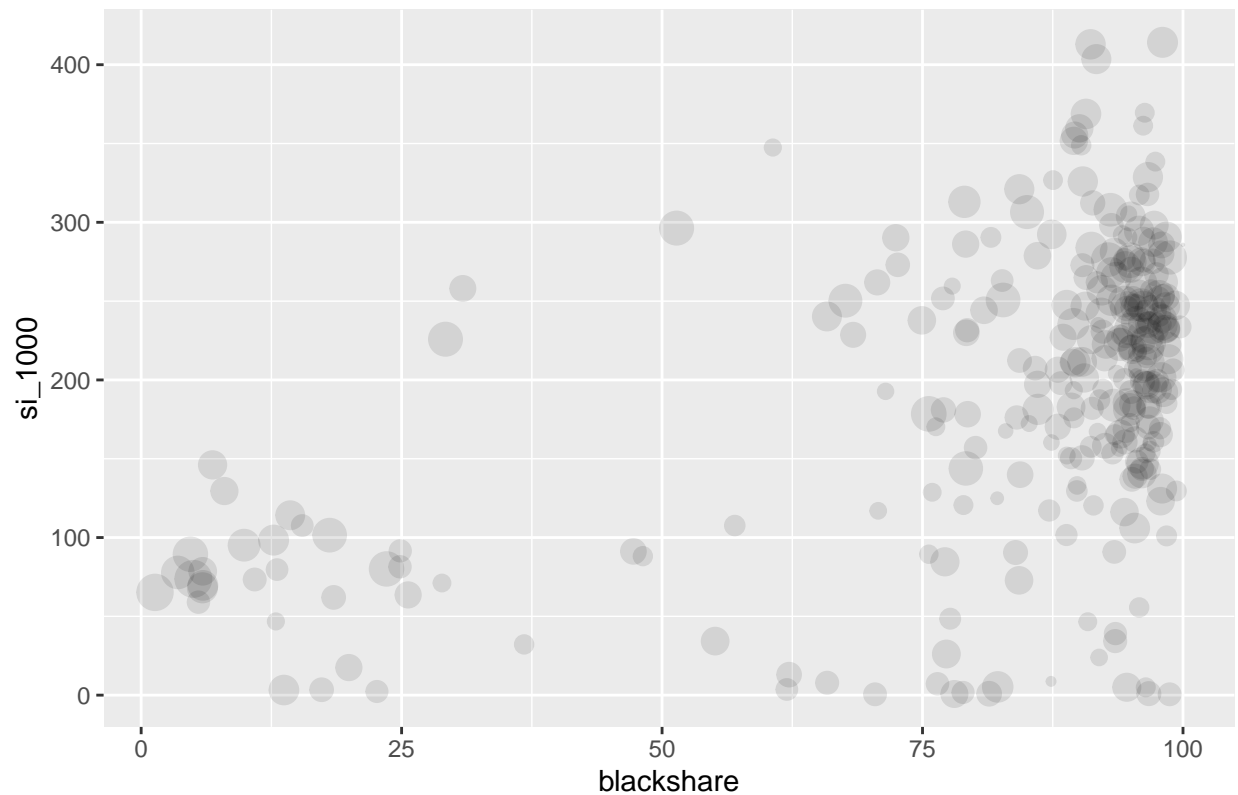
1 “Cross-sectional” analysis

In this section we'll explore variation in shutoffs *across* Census tracts (one observation per Census tract, summing shutoffs over the whole time period).

1.1 Visualize and interpret the relationship between share Black and shutoffs per capita across census tracts in Detroit.

```
#scatterplot: % black vs shutoffs
ggplot(data = tract, aes(x = blackshare,
                        y = si_1000, weight = pop, size = pop)) +
  geom_point(alpha = 0.1) +
  scale_size(range = c(0.1, 6), guide = "none") +
  ggtitle("Shutoffs per capita on Black share of population")
```

Shutoffs per capita on Black share of population



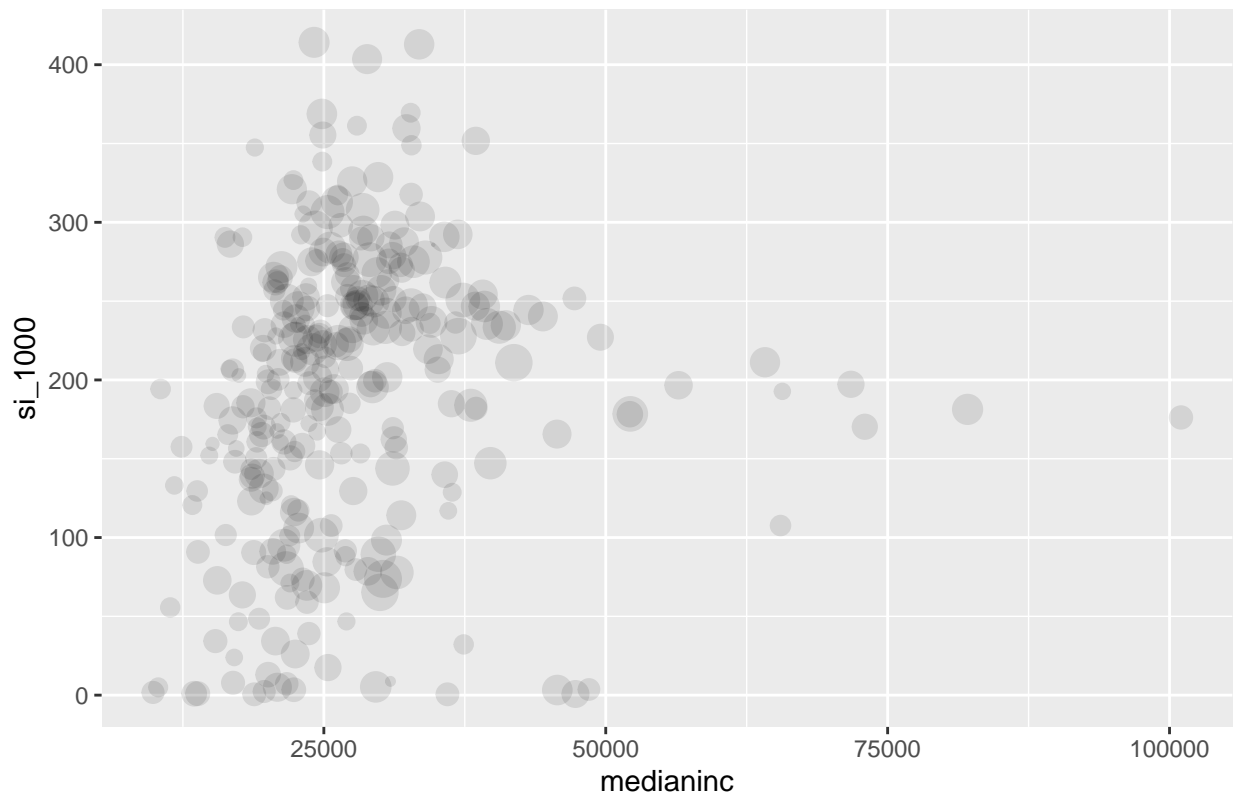
The above scatterplot plots the relationship between shutoffs per capita and the share of residents that are Black across Census tracts (Census tracts are weighted by population, with larger markers for more populous tracts). While the majority of tracts are predominantly Black, the concentration of tracts that are more than 75% Black have markedly higher shutoffs per capita (the population-weighted correlation between share Black and shutoffs per capita is 0.51).

Also note that all Census tract statistics based on the American Community Survey (for share Black, median household income, and population) are calculated as averages from 2010 through 2017.

1.2 Visualize and interpret the relationship between median income and shutoffs per capita across census tracts in Detroit.

```
#scatterplot: median income vs shutoffs
ggplot(data = tract, aes(x = medianinc,
                        y = si_1000, weight = pop, size = pop)) +
  geom_point(alpha = 0.1) +
  scale_size(range = c(0.1, 6), guide = "none") +
  ggtitle("Shutoffs per capita on median income")
```

Shutoffs per capita on median income



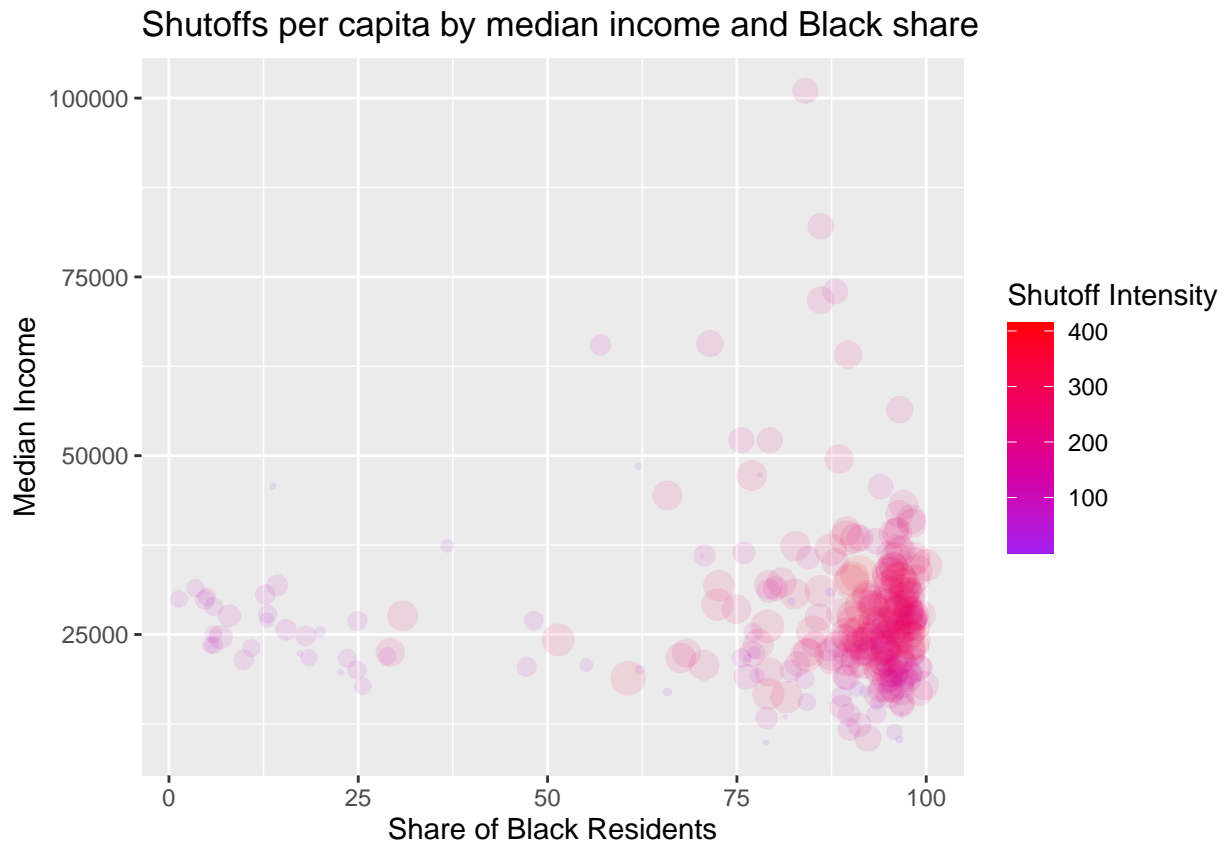
The above scatterplot plots the relationship between shutoffs per capita and median household income across Census tracts (Census tracts are again weighted by population). The overwhelming majority of tracts have a median income below 50,000, and the relationship between shutoffs per capita and median income is positive (albeit much weaker than the positive correlation between shutoffs per capita and share Black, 0.13 compared to 0.51).

1.3 Visualize and interpret how shutoffs per capita relate to both Black share and median income on the *same* plot. Does race or income appear to be more salient?

```
#scatterplot: median income & race vs shutoffs
#HINT: try plotting Black share on the X-axis and median income on the Y-axis
#      then experiment w/marker size and color to indicate greater shutoff rates

ggplot(data = tract,
       aes(x = blackshare,
           y = medianinc,
           size = si_1000,
           color = si_1000)) +
  geom_point(alpha = 0.1) +
  scale_size(range = c(0.1, 6), guide = "none") +
  scale_color_gradient(low="purple", high="red") +
  ggtitle("Shutoffs per capita by median income and Black share") +
  xlab("Share of Black Residents") +
  ylab("Median Income") +
```

```
labs(color = "Shutoff Intensity")
```



The above scatterplot shows how shutoffs per capita vary along with both share Black and median household income: tracts with more shutoffs per capita appear as larger, redder circles. While most tracts in Detroit are relatively low income and over 75% Black, it's clear that the tracts with the highest shutoffs per capita tend to be those with the highest share Black, regardless of median income.

2 Time-series analysis.

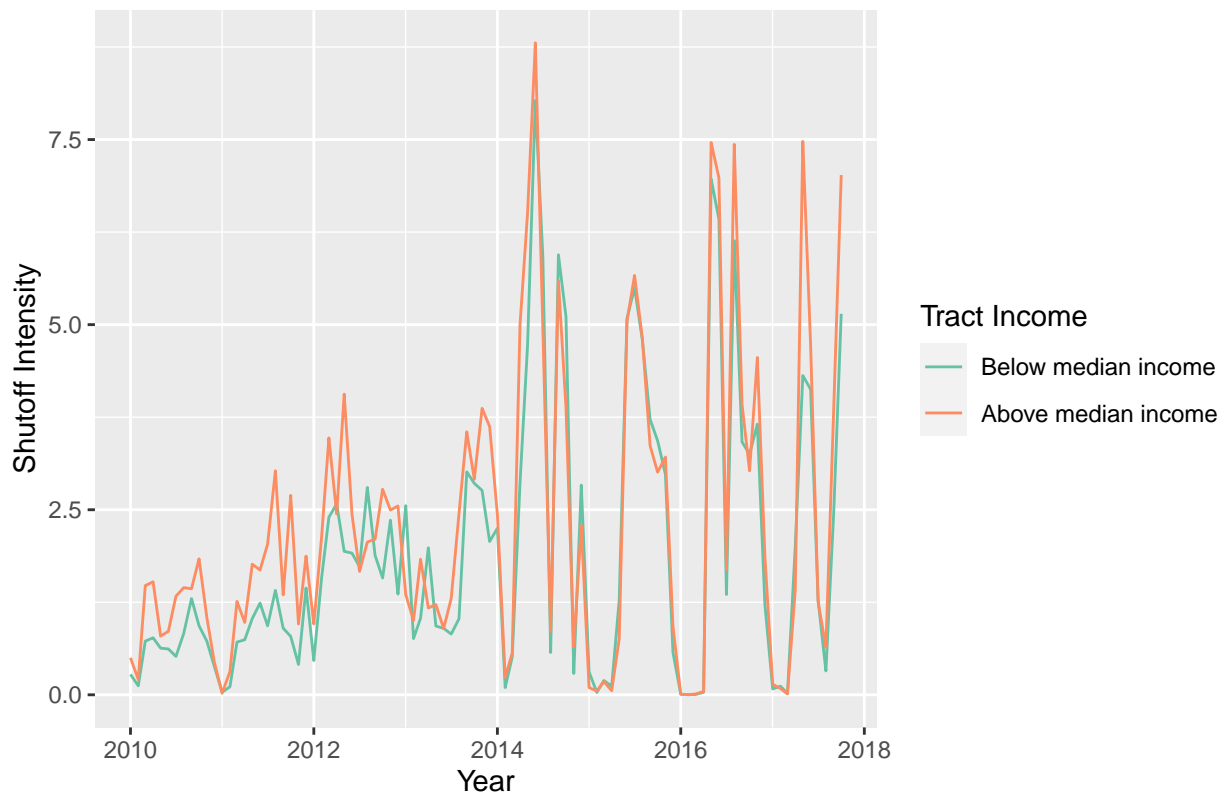
In this section, we'll explore variation *between* different groups of Census tracts and over time *within* groups (with groups defined based on tract-level income and racial composition).

2.1 Plot and interpret shutoffs per capita over time for tracts below/above citywide median housheold income (show two time series on a single plot).

```
#time series plot of SI per capita for tracts above/below citywide median income
ggplot(ym_inc,
  aes(x = date, y = si_1000)) +
  geom_line(aes(color = inc_above_median)) +
  ggtitle("Time series plot of shutoffs by tracts above/below median income") +
  xlab("Year") +
  ylab("Shutoff Intensity") +
  labs(color = "Tract Income") +
```

```
scale_color_brewer(palette = "Set2")
```

Time series plot of shutoffs by tracts above/below median income

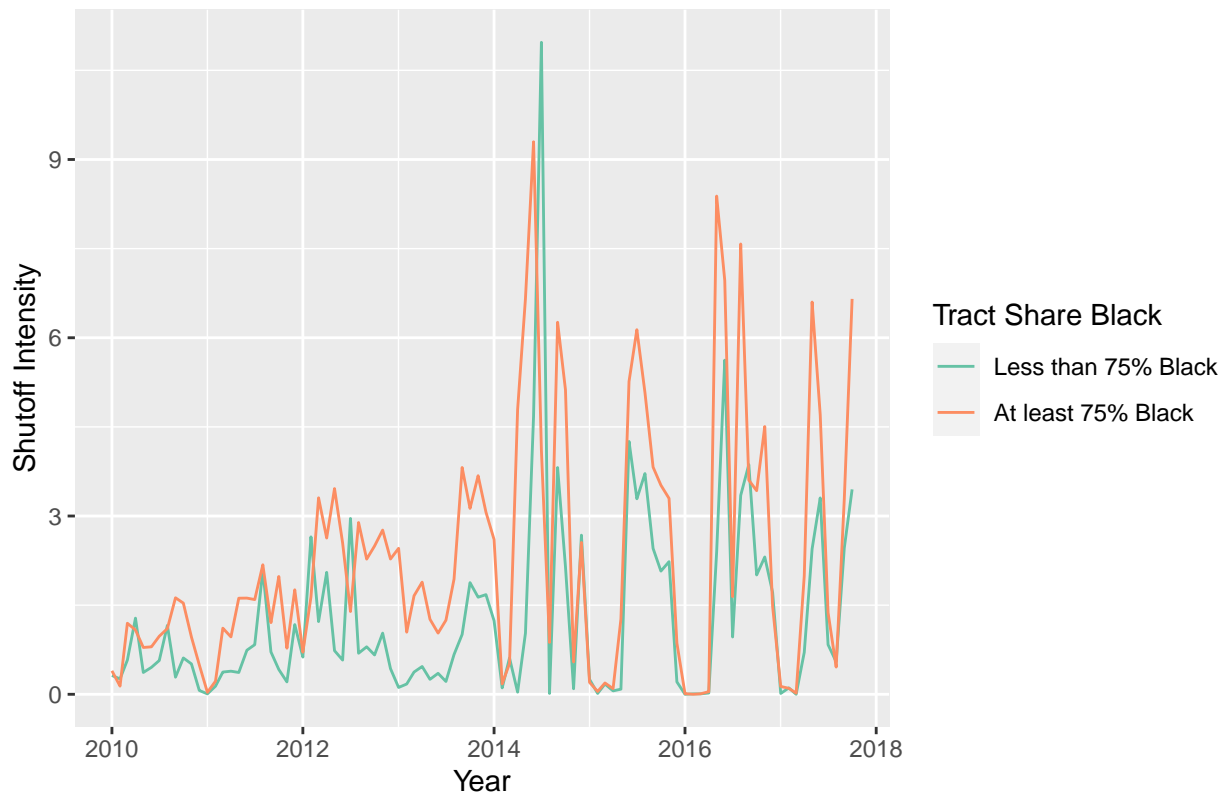


Shutoffs per capita exhibit pronounced seasonal variation, but beginning in 2014 monthly shutoffs per capita are nearly identical in Census tracts above and below the citywide median.

2.2 Plot and interpret shutoffs per capita over time for tracts that are at least 75% Black and those that aren't (show two time series on a single plot).

```
#time series plot of SI per capita by tracts above/below 75% black
ggplot(ym_race,
  aes(x = date, y = si_1000)) +
  geom_line(aes(color = black75)) +
  ggtitle("Time series plot of shutoffs by tracts above/below 75% Black") +
  xlab("Year") +
  ylab("Shutoff Intensity") +
  labs(color = "Tract Share Black") +
  scale_color_brewer(palette = "Set2")
```

Time series plot of shutoffs by tracts above/below 75% Black



When comparing shutoffs per capita over time for Census tracts that are at least 75% Black to other tracts, pronounced differences are visible: monthly shutoffs per capita are noticeably higher in predominantly Black Census tracts compared to other tracts in nearly every month.

3 Conclusion

3.1 Based on the “cross-sectional” and time series analysis conducted above, does race or income appear to be a more important factor for explaining what type of households are most affected by the public water shutoffs? Explain.

In both the cross-sectional and time series analysis, share Black appears to be a stronger predictor of shutoffs per capita than median income. This is consistent with disparate impact by race: income aligns more closely with the criteria used to suspend water service for nonpayment, but communities with the highest share of Black residents appear to be hit hardest regardless of income (Figure 1.3 is the most compelling visualization of this, with lower shutoffs per capita in lower-income tracts that are not predominantly Black).