

DSPC 7514 Assignment 4

Water Shutoffs in Detroit: Community Impact by Race and Income

Arnav Sahai

2026-03-02

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

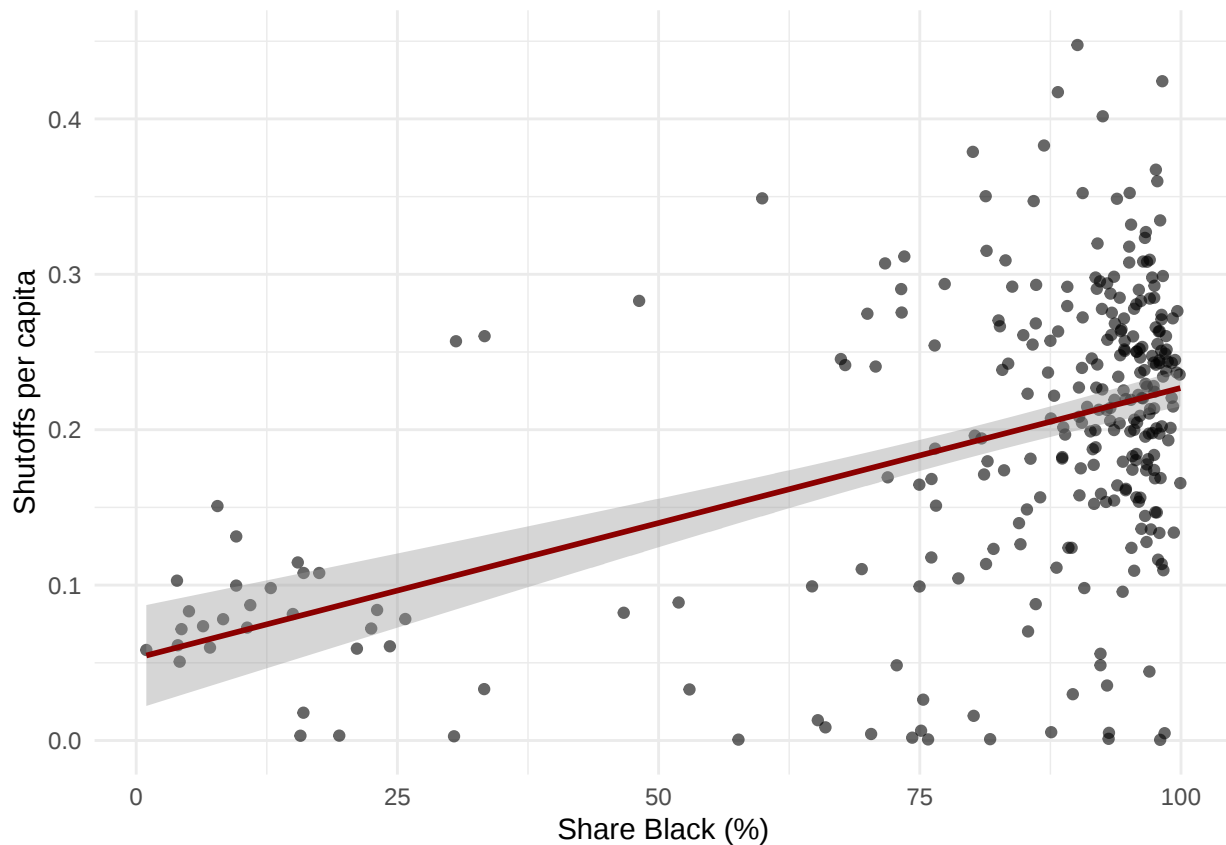
```
library(tidyverse)
library(haven)
```

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. Make sure to clearly label your plot axes.

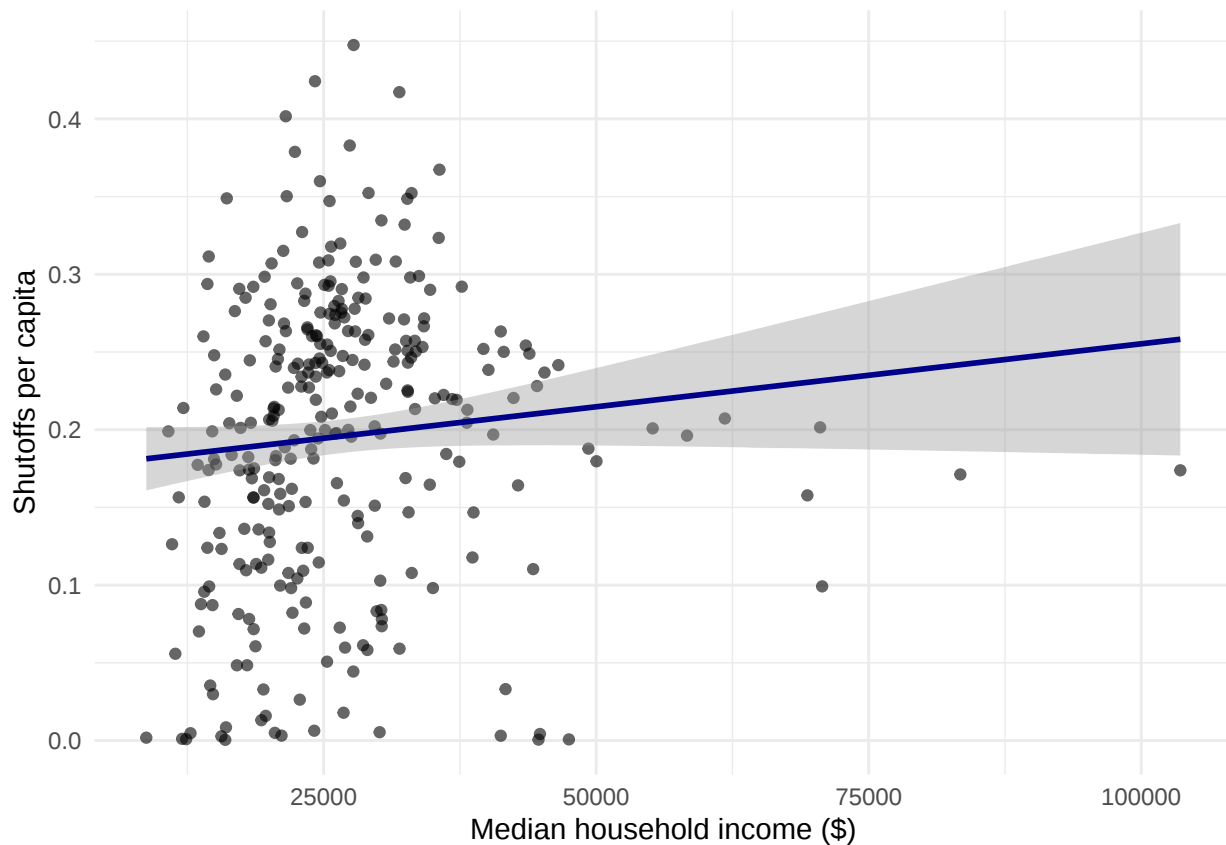
```
cross_section %>%
  ggplot(aes(x = blackshare, y = shutoffs_per_capita)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, colour = "darkred") +
  labs(x = "Share Black (%)", y = "Shutoffs per capita") +
  theme_minimal()
```



Interpretation: Tracts with a higher share of Black residents tend to have higher shutoffs per capita. The positive slope suggests that race is associated with greater exposure to water shutoffs across Detroit census tracts.

1.2 Visualize and interpret the relationship between median income and shutoffs per capita across census tracts in Detroit. Make sure to clearly label your plot axes.

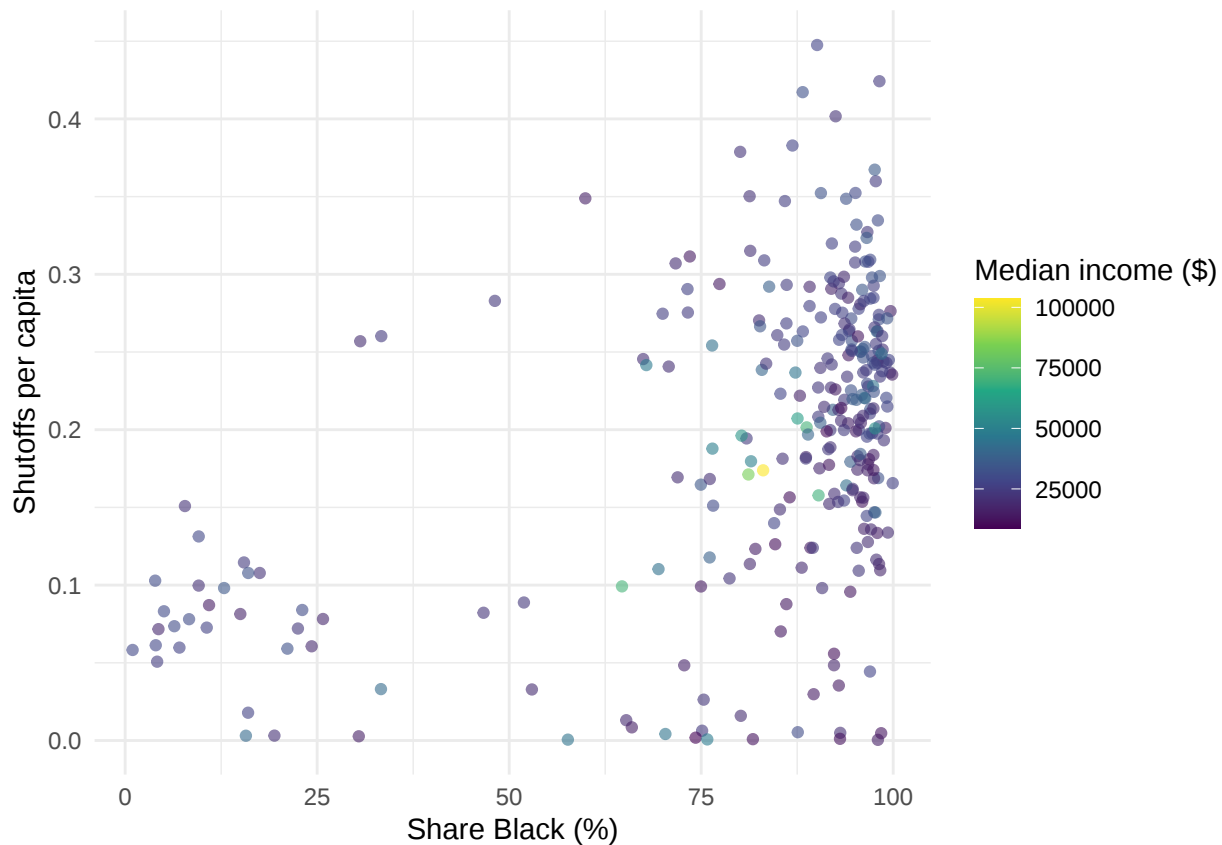
```
cross_section %>%
  ggplot(aes(x = medianinc, y = shutoffs_per_capita)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, colour = "darkblue") +
  labs(x = "Median household income ($)", y = "Shutoffs per capita") +
  theme_minimal()
```



Interpretation: Tracts with lower median household income tend to have higher shutoffs per capita. The negative relationship is consistent with income being a factor in shutoff exposure—poorer tracts experience more shutoffs per capita.

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

```
cross_section %>%
  ggplot(aes(x = blackshare, y = shutoffs_per_capita, colour = medianinc)) +
  geom_point(alpha = 0.6) +
  scale_colour_viridis_c(name = "Median income ($)" +
  labs(x = "Share Black (%)", y = "Shutoffs per capita") +
  theme_minimal() +
  theme(legend.position = "right")
```



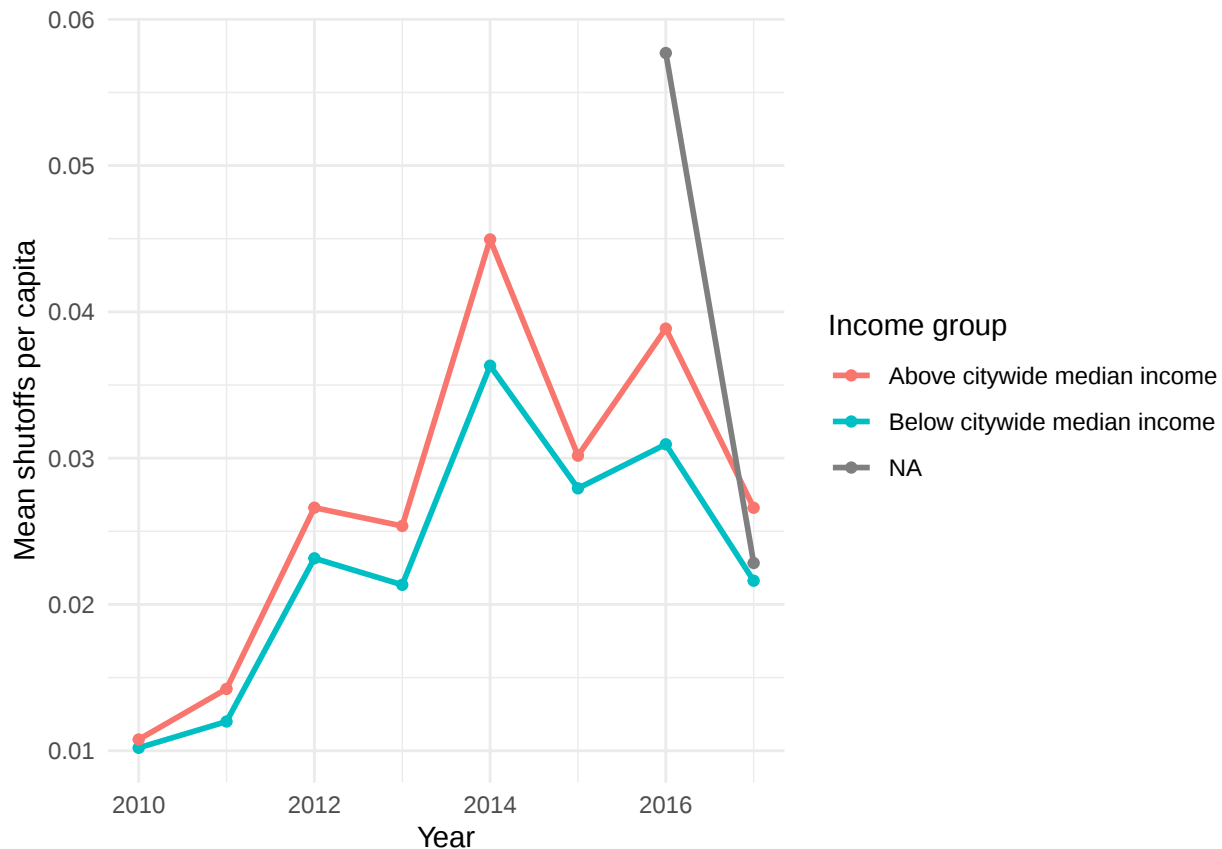
Interpretation: When both share Black and median income are shown on the same plot (income as colour), the gradient by income is visible within different levels of Black share: at a given share Black, lower-income tracts tend to have higher shutoffs per capita. The strong positive relationship with share Black remains. Both race and income appear salient; the cross-sectional patterns do not clearly show one as dominant—race and income are correlated across tracts, so further analysis (e.g. regression controlling for the other) would be needed to separate their roles.

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 the shutoffs per capita over time for tracts below/above citywide median household income (show two time series on a single plot). Make sure to clearly label your plot axes and legend.

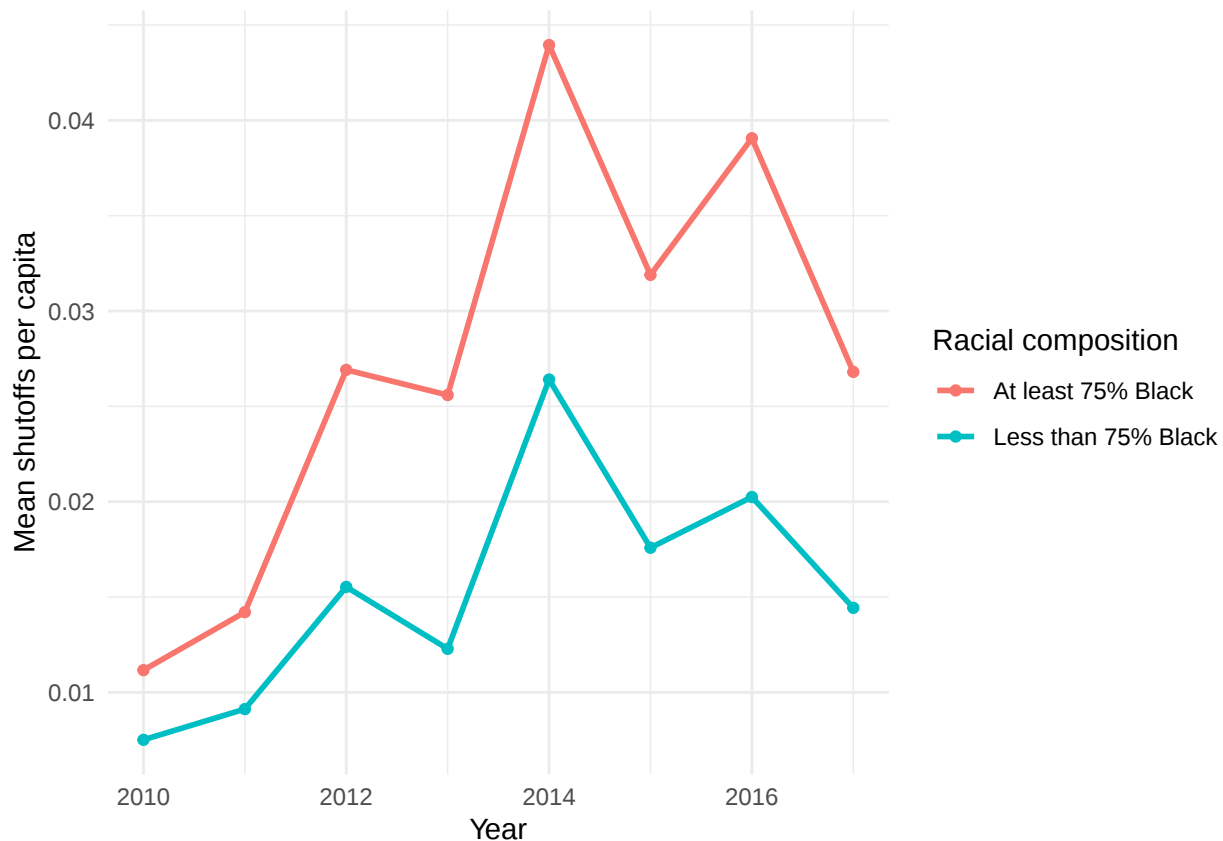
```
ts_income %>%
  ggplot(aes(x = year, y = mean_shutoffs_pc, colour = income_group, group = income_group)) +
  geom_line(linewidth = 1) +
  geom_point() +
  labs(x = "Year", y = "Mean shutoffs per capita", colour = "Income group") +
  theme_minimal()
```



Interpretation: In the data, tracts *above* the citywide median household income have consistently *higher* mean shutoffs per capita than tracts below the median in every year—so the line for “Above citywide median income” lies above the line for “Below citywide median income.” That is the opposite of a simple “poorer tracts have more shutoffs” story. It may reflect confounding (e.g. race, housing type, or the fact that “above median” in Detroit is still relatively low income), or different exposure to shutoffs (e.g. more single-family accounts in somewhat higher-income tracts). The time series shows a clear income-group gap that persists over time, but with higher-income tracts having higher mean shutoffs per capita in this descriptive plot.

2.2 Plot and interpret the 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). Make sure to clearly label your plot axes and legend.

```
ts_race %>%
  ggplot(aes(x = year, y = mean_shutoffs_pc, colour = race_group, group = race_group)) +
  geom_line(linewidth = 1) +
  geom_point() +
  labs(x = "Year", y = "Mean shutoffs per capita", colour = "Racial composition") +
  theme_minimal()
```



Interpretation: Tracts that are at least 75% Black have higher mean shutoffs per capita than tracts that are less than 75% Black in every year. The gap is substantial and persistent over time, consistent with race being an important factor in which communities are most affected by water shutoffs.

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.

Both race and income are associated with higher shutoff rates in Detroit. In the cross-section, tracts with a higher share Black and lower median income both show higher shutoffs per capita. Because Black share and income are correlated (predominantly Black tracts tend to have lower income), the descriptive plots alone do not cleanly separate which factor is “more” important. In the time series, the 75% Black split shows clear, persistent gaps in mean shutoffs per capita, with predominantly Black tracts having higher rates in every year. By contrast, the income time series shows the opposite of a simple “poorer = more shutoffs” pattern: tracts above the citywide median income have higher mean shutoffs per capita than tracts below the median, which may reflect confounding with race or other factors. On balance, the patterns are consistent with both race and income mattering; without regression controls we cannot say one is definitively more important than the other, but the evidence suggests that households in predominantly Black and lower-income tracts are most affected by public water shutoffs.