

U6614: Subway Fare Evasion Arrests and Racial Bias

Your Name (your-uni)

2020-09-29

Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by 11:59pm on Monday, October 5th.

1 Load libraries

2 Subway station-level arrest totals

- 2.1 Load cleaned/appended arrest microdata (`arrests_all.csv`) w/all strings as factors.
- 2.2 Create new data frame (`st_arrests`) that aggregates microdata to station-level observations including the following information.
 - `st_id`, `loc2`, `total arrests`
- 2.3 Plot histogram of arrests and briefly describe the distribution of arrests across stations.

3 Joining subway ridership and neighborhood demographic data

- 3.1 Import and inspect secondary data sources (`station_povdataclean_2016.csv` and *Subway Ridership by Station - BK.csv*).
- 3.2 Join ridership and demographic data to `st_arrests` and inspect results (store new data frame as `st_joined`).
 - *Inspect results from joins, drop unnecessary columns from the ridership data, and group `st_joined` by `st_id` and `mta_name`.*
- 3.3 Print the top 10 stations by total arrest counts, along with the poverty rate and share black for each subway station area shown.

4 Explore relationship between arrest intensity and poverty rates across subway station (areas)

4.1 Compute arrest intensity and prep data.

- *Drop the observation for the Coney Island station and very briefly explain your logic*
- *Create new column of data representing fare evasion enforcement intensity (`arrperswipe_2016` = arrests per 100,000 swipes)*
- *Create dummy variable indicating whether a station is high poverty (`highpov` = 1 if pov rate is > median pov rate across all Brooklyn station areas)*
- *Create new dummy for majority Black station areas (`nblack` = 1 if `shareblack` > .5)*
- *coerce new dummy variables into factors w/category labels*
- *Assign results to new data frame called `stations`*

4.2 Examine the relationship between arrest intensity and poverty rates

- *Show a scatterplot of arrest intensity vs. poverty rates along with the regression line you think best fits this relationship*
- *Which regression specification do you prefer: linear or quadratic? Be clear about your logic and if applicable cite statistical evidence to support your decision.*
- *Report diff in mean arrest intensity between high/low pov areas and assess statistical significance*
- *For this assignment: weight observations by the number of MetroCard swipes when calculating difference in group means, but not for your regressions*

5 How does neighborhood racial composition mediate the relationship between poverty and arrest intensity? Examine the relationship between arrest intensity & poverty by Black vs non-Black station area (`nblack`).

5.1 Present a table showing the difference in mean arrests per swipe for each group in a 2x2 table of `highpov` vs `nblack`.

- *Remember to weight by the number of MetroCard swipes at each station*
- *Could the difference in arrest intensity be explained by differences in poverty rate?*

5.2 Show a scatterplot of arrest intensity vs. poverty rates along with the regression line you think best fits this relationship.

- *Which regression specification do you prefer: linear or quadratic? Be clear about your logic and if applicable cite statistical evidence to support your decision.*
- *Interpret your preferred regression specification (carefully!).*

6 Examine the relationship between arrest intensity and crime

6.1 Load the crime data and join to the existing stations data frame (`nypd_criminalcomplaints_2016.csv`)

NOTE: For the next two subsections, present your preferred plots to inform the relationships in question, along with any additional data manipulation and evidence to support your decisions/interpretation/conclusions. You'll want to explore the data before arriving at your preferred plots, but don't present everything you tried along the way such as intermediate versions of your preferred plot. Focus on the analysis you eventually settled on to best inform the question at hand, and any critical observations that led you down this path.

6.2 Examine the overall relationship between arrest intensity and crime (without taking neighborhood racial composition or poverty into account) (comparable to Section 4.2).

6.3 Examine how neighborhood racial composition mediates the relationship between arrest intensity and crime (comparable to Section 5.2).

7 Summarize and interpret your findings with respect to subway fare evasion enforcement bias based on race

- *Is there any additional analysis you'd like to explore with the data at hand?*
- *Are there any key limitations to the data and/or analysis affecting your ability to assess enforcement bias based on race?*
- *Is there any additional data you'd like to see that would help strengthen your analysis and interpretation?*