

Using Data Visualizations to Find Patterns in New York Police Department Stop and Arrest Data

Introduction

Every year, the New York City Police Department (NYPD) stops individuals for suspected criminal involvement. From 2002 to the first quarter of 2015, there has been more than five million stop-and-frisk interrogations (New York Civil Liberties Union, 2015). Research on approximately 175,000 stops from January 1998 through March 1999 showed that Blacks and Hispanics represented 51% and 33% of the stops while they only represented 26% and 24% of the New York City population (Gelman, Fagan, & Kiss, 2012). While these statistics indicate large racial disparities, it is important to recognize that multiple factors can influence issues related to racial or gender discrimination.

In this activity, we will look at multiple graphics evaluating the number of stops and arrests based upon race, gender, type of crime, and the amount of force used by police. Information for this activity was collected from individual police reports from 2006 to 2014. This data is available at the [New York Police Department's website \(https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page\)](https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page).

Part A: Investigating Racial Disparities with Bar Charts

Go to the Grinnell College Rstudio site, <http://shiny.grinnell.edu/>, and select the [NYPD Bar Charts](#) (it may take a few seconds to open). Click the [Variable Description](#) link on the bottom left corner to understand the options within this app.

Before starting questions 1-5, make sure your app is set with the following options:

- **Y-axis Variable:** *Stopped*
 - **X-axis Variable:** *Race*
 - **Y-axis Measurement:** *Counts*
 - **Choose the Years:** 2006-2014
 - **Facet By:** *None*
 - **Color By:** *None*
- 1) As mentioned in the introduction, Gelman, Fagan, & Kiss found disparities in the percentage of Black and Hispanic stops in the 1990s. Does the trend in arrests based on race appear to have changed during the years 2006 through 2014? Using the available data from 2006 through 2014, approximate the percentage of stops for Black and Hispanic individuals.
 - 2) Choose *Year* in the **Facet By** menu. Review the data for each year from 2006 through 2014, looking at the race of individuals being stopped by NYPD. Does the same pattern

of stops based on race stay consistent for all 9 years? Are Black and Hispanic individuals consistently stopped more often than individuals of other races?

- 3) Due to a larger percentage of the population in New York City identifying as Black, White, and Hispanic, let's take a closer look at the record of individuals stopped among these races. Use the **Filters** tab to delete *Asian*, *Native*, and *Other* in the **Filter by Race** menu. Change the **Y-Axis Variable** to *Frisked* and **Facet By** to *None*. Now that you are able to see the differences among individuals of these three races, what patterns emerge regarding the number of individuals who were frisked by NYPD from 2006 through 2014?
- 4) Change the **Y-Axis Variable** to *Arrested* and look at the general trend throughout each year. What do you see? Is there a pattern of racial disparity, and if so, are there years during which this trend is not present?
- 5) In 2014, the New York Civil Liberties Union stated that New York City Mayor Bill de Blasio had "...made stop-and-frisk reform a central issue in his campaign, and shortly after his election he moved aggressively to honor his campaign promises." Change the **Y-Axis Variable** to *Stopped* and the **X-Axis Variable** to *Year*. Does the data support the conclusion reached by the NYCLU? How?

We've confirmed that in each year, NYPD police reports show that more Black and Hispanic people are stopped, frisked or arrested than people of other races. We also found that this pattern is fairly consistent throughout each year. While this data indicates a problem, there are numerous other factors that should be carefully considered, before we can say there is evidence of discrimination. The population, the location, the type of crime, and whether or not the suspect is guilty should all be taken into account. In addition the pattern based upon counts of frisked or arrested suspects is dramatically impacted by the number of stops. For example, there are more stops in a particular race, we would also expect more frisks. To address this issue, it can be helpful to consider percentages instead of counts.

Part B: Counts vs. Percentages

For questions 6 through 8, set your **Bar Chart App** options as follows:

- **Y-axis Variable:** *Frisked*
 - **X-axis Variable:** *Race*
 - **Y-axis Measurement:** *Percentage of Stops*
 - **Choose the Years:** 2006-2014
 - **Facet By:** *None*
 - **Color By:** *None*
 - **Filter by Race:** *Black Hispanic White*
- 6) Using the data from question 3, it is clear that the number of Black and Hispanic individuals being frisked by NYPD is larger than the number of White individuals being frisked. This data supports Ridgeway's conclusion that, "Nonwhites generally experienced slightly more intrusive stops, in terms of having more frequent frisks and

searches, than did similarly situated white suspects” (Ridgeway, 2007). As we look at frisks as a percentage of stops, does this pattern persist?

- 7) Change the **Y-Axis Measurement** to *Percentage of Arrests*. Is there a clear difference among Black, Hispanic, and White groups regarding NYPD arrest rates?
- 8) Using the count data from question 6, racial disparities among Black, Hispanic, and White groups are clear. Change the **Y-Axis Variable** to *Searched* and the **Facet By** to *Year*. Which race has the largest percentage of searches?

Part C: Incorporating Gender and Crime Type into the Analysis

We have seen that after changing the Counts to Percentages, the difference among Black, Hispanic, and White groups is much smaller. In fact, White individuals have a slightly higher percentage of searches during arrests throughout the years 2006 to 2014. Do these conclusions change when we add Gender and Crime Type? Are there differences among the racial groups when we consider those factors, and do they change when looking at the data in Counts and in Percentages?

To answer these questions, we are going to compare two graphs. Set the options in the **Bar Chart App** as follows:

- **Y-axis Variable:** *Frisked*
 - **X-axis Variable:** *Race*
 - **Y-axis Measurement:** *Counts*
 - **Choose the Years:** 2006-2014
 - **Facet By:** *Crime Type*
 - **Color By:** *Gender*
 - **Filter by Race:** *Black Hispanic White*
 - **Filter by Gender:** *Female Male*
- 9) In all three racial groups being analyzed (Black, Hispanic, White), there appear to be a large number of male individuals that are frisked. Is there a particular type of crime in which male individuals in all three groups are frisked more often than the other crime types?
 - 10) Change the **Y-Axis Measurement** to *Percentage of Arrests*. Notice that the percentage of individuals frisked while being arrested is around 76% for female individuals, and 89% for male individuals. Do most crime types share this pattern?

By comparing percentages rather than merely counts, we are able to gain additional perspective on the data. We see that even though there is a large difference in counts for the number of stops, frisks, and searches among Black, Hispanic, and White groups, the percentages of those groups are much closer to each other.

Part D: Evaluating the level of Force Used

It is generally claimed that the NYPD uses force against Black and Latinos more than Whites (Ridgeway, 2007). Fagan's research on data from 2006 to 2009 suggests that "force is used in nearly one stop in four, with force far more likely to be used against Black suspects (24.12 percent) and Hispanic suspects (24.75 percent) than White suspects (17.85 percent)" (Fagan, 2010, p.63).

To understand how much force was used on each racial group in stops that result in arrest, set the options in your app as:

- **Y-axis Variable:** *Force*
- **X-axis Variable:** *Race*
- **Y-axis Measurement:** *Percentage of Arrests*
- **Choose the Years:** 2006-2014
- **Facet By:** *None*
- **Color By:** *None*
- **Filter by Race:** *Black Hispanic White*

From this dataset, we see that NYPD used force on around 82% of Black and Hispanic individuals and 77% of White individuals. To get a better understanding of what this data means, we should break down the available information and answer the following questions.

- 11) Since a large amount of the force used by NYPD involved officers putting their hands on an individual, and this is not as likely to cause injury to an individual, delete *Hands* in **Select the Type of Force**. Does the data shown correspond to Fagan's claim that when a more restrictive definition of force is used (excluding police putting their hands on an individual), disparities are narrow, yet still present (Fagan, 2010)? Are there any other conclusions that can be reached based on the data in this question?
- 12) Choose *Year* in **Facet By**. Are there any historical trends based on the use of force from year to year among the three racial groups?
- 13) Choose *Percentage of Stops*, and then change the **Y-Axis Variable** to *Arrested*. Does this data support or negate Ridgeway's claim that "white suspects were slightly more likely to be issued a summons than were similarly situated nonwhite suspects" (Ridgeway, 2007)?

Part D: Drawing Conclusions

The first set of graphs clearly verified Gelman, Fagan, and Kiss's claim that there are more Black and Hispanic people stopped than White people. However, some of our graphs support Ridgeway's idea that "overall, after adjustment for stop circumstances, we generally found small racial differences in the rates of frisk, search, use of force, and arrest" (Ridgeway, 2007).

For example, when we remove the use of hands in the definition of force, the difference among the three groups was relatively small. This corresponds with Fagan's claim that when a more restrictive definition of force is used (excluding police putting hands on suspect), these disparities are narrow, yet still present (Fagan, 2010). Finally, in question 12) we see that the rates of Hispanic and White people being arrested are very close and are both higher than that of Black people, which supports Ridgeway's claim that "white suspects were slightly more likely to be issued summons than were similarly situated nonwhite suspects" (Ridgeway, 2007).

While all the graphs in the lab are based upon the same dataset, each graph can appear to tell a very different story. This demonstrates the importance of quantitative reasoning when viewing any graphical summary. Even when the data is accurate and the graphs are made correctly, data displays can be easily misunderstood.

In this dataset there could be reason to question the accuracy of the police forms. We cannot be certain that every form was carefully filled out and checked for accuracy. There may be errors in the type of force used or the way in which police identified the race of the suspect. While this data may be imperfect we can still glean some useful information based upon the patterns we see in the graph.

Endnotes:

This activity was created by Ying Long, Zachary Segall, Krit Petrachaianan and Shonda Kuiper. All rights reserved. Date: 7/25/2015

References:

Fagan. (2010). *Report of Jeffrey Fagan, Ph.D.* Retrieved from http://ccrjustice.org/sites/default/files/assets/files/Expert_Report_JeffreyFagan.pdf

Gelman, Fagan, & Kiss. (2012). *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias.* Retrieved from <http://www.stat.columbia.edu/~gelman/research/published/frisk9.pdf>

New York Civil Liberties Union. (2014). *STOP-AND-FRISK 2013* Retrieved from http://www.nyclu.org/files/publications/8.26.14_Stop-and-Frisk_2013_final.pdf

New York Civil Liberties Union. (2015). Retrieved from <http://www.nyclu.org/content/stop-and-frisk-data>

Ridgeway. (2007). *Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices.* Retrieved from http://www.rand.org/pubs/technical_reports/TR534.html