

Goal: Identify the “best” 5 and “worst” 5 precincts in the NYC Stop and Frisk data.

“Best” precincts are precincts whose race distribution in the NYC Stop and Frisk data most closely matches the race distribution of residents who live within this precinct. “Worst” precincts are those where the race distribution in the NYC Stop and Frisk data is most extreme against the race distribution of residents who live within this precinct.

Answer the following questions.

- Use the Chi-Square statistic shown below to measure the discrepancy between the race distribution from the NYC Stop and Frisk data to the race distribution of residents who live within this precinct.

$$Statistic = \sum_{all\ races} \frac{(Observed - Expected)^2}{Expected}$$

Obtain a total value for each precinct. Sort the list by total. Use this sorted list to identify the “best” 5 and “worst” 5 precincts. (10 pts)

“Best” Precincts	
Rank	Precinct #
1. (Best Overall)	113
2.	67
3.	69
4.	81
5.	73

Precinct	Statistic
113	4.1599070607
67	14.200847115
69	18.441682003
81	20.059776156
73	23.866151553
68	26.601400051

“Bad” Precincts	
Rank	Precinct #
1. (Worst overall)	19
2.	17
3.	13
4.	14
5.	62

Precinct	Statistic
62	1318.2853231
14	1505.4566906
13	1730.6265377
17	1773.5560466
19	2861.8149561

Comment: My list of “best” and “worst” precincts may differ than your list. There is no best way to combine races; thus, your method and my method may not agree.

- The race designations in the NYC Stop and Frisk data do not (exactly) match the race designations in the Census data. Explain what race designations were ignored or combined in order to compute the above measure of discrepancy between these two distributions. (4 pts)

From Stop & Frisk data, I decided to use only A, B, I, W, and Z as these race designations had obvious race designations in the Census data.

Stop & Frisk Reference		Census Data Reference	
A	ASIAN/PACIFIC ISLANDER	P0010006	
B	BLACK	P0010004	
I	AMERICAN INDIAN/ALASKAN NATIVE	P0010005	
P	BLACK-HISPANIC		P0020002
Q	WHITE-HISPANIC		P0020002
W	WHITE	P0010003	
X	UNKNOWN	P0010008	
Z	OTHER	P0010008	

Methodology:

- Compute a new total in the Stop & Frisk data using only Races: A, B, I, W, and Z.
- Compute the corresponding new total from the Census data, i.e. P0010006 + P0010004 + P0010005 + P0010003 + P0010008.
- Determine the proportion for each race, e.g. Percent_Asian = P0010006 / New Census Total, etc.
- Compute the expected count for each race, e.g. Expected_Asian = Stop & Frisk New Total * Percent_Asian, etc.
- Compute the Chi-Square Test Statistic for each Precinct as follows.

$$\frac{(N(\text{race}, A) - \text{Expected_Asian})^2}{\text{Expected_Asian}} + \frac{(N(\text{race}, B) - \text{Expected_Black})^2}{\text{Expected_Black}} + \frac{(N(\text{race}, I) - \text{Expected_AI})^2}{\text{Expected_AI}} + \frac{(N(\text{race}, W) - \text{Expected_White})^2}{\text{Expected_White}} + \frac{(N(\text{race}, Z) - \text{Expected_Other})^2}{\text{Expected_Other}}$$

- Rank the Precinct by this statistic - those with a high value are far from expected, those with low values are close to expected.

3. Consider your list of 5 “best” precincts and 5 “worst” precincts. Use Google Maps to identify the location of these precincts.

a. Do the 5 “best” tend to be in the same geographic area? Briefly discuss. (3 pts)

The “best” precincts, the precincts that most closely related to their expected counts for Stop & Frisk, tend to be concentrated in northeast Brooklyn. I believe the best precinct (113) is located in Queens.



4. Do the 5 “worst” tend to be in the same geographic area? Briefly discuss. (3 pts)

The “worst” precincts, those that tend to deviate the most from their expected counts, are most concentrated in Manhattan – mostly on the south end of Central Park. This might be a touristy area and maybe this is why the actual Stop & Frisk data does not closely relate to the Census data in this area.



Link: <https://data.cityofnewyork.us/Public-Safety/Police-Precincts/78dh-3ptz/data>