

# Preliminary Analysis of Incarceration by Race and Gender

Sonali Mishra

Spring 2022

## 1 Overview

This paper does a prima facie analysis of incarceration status in United States in the year 2002 faceted at race and gender level. The way we calculate duration is by looking at [NLS investigator](#). We look at NLSY97 from 1997 to 2019 and download the incarceration history captured for the year 2002. This dataset contains month-wise data and observations are recorded at individual level. Upon some data manipulation, we only consider the months when the individual was incarcerated in that month partially or entirely. We also have indicators on race (black, hispanic, mixed race and non-black). We further look at incarceration rate which is calculated as number of people incarcerated as proportion of 100,000 individuals as that is the most commonly used metric to analysis incarceration trends in United States.

## 2 Graphical representation

The below graph compares incarceration rate between male and female across ethnic groups. Incarceration rate for black men is roughly 8 times more than black women on average. The other two groups have similar trend except for mixed race (which is to be ignored for this assignment).

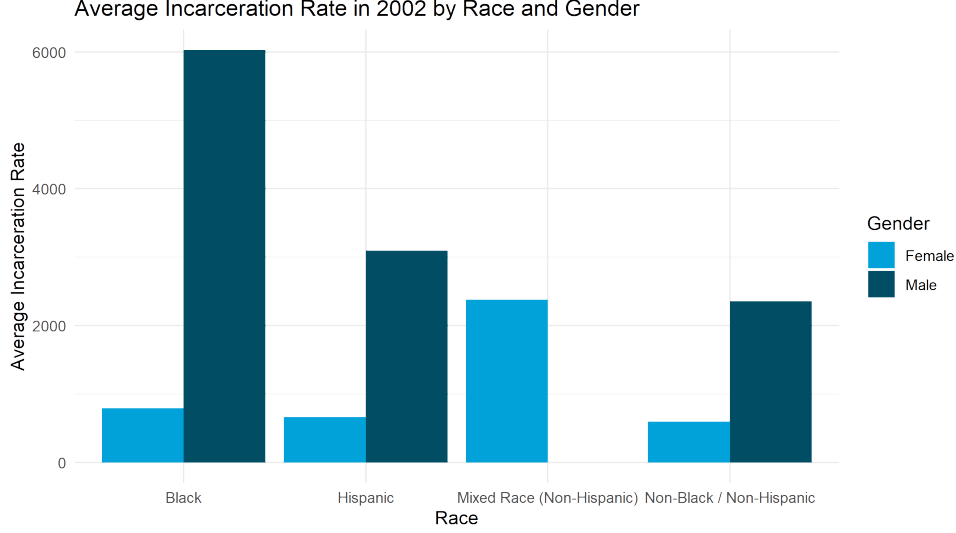


Figure 1: Incarceration Rate on an average (for a given 100,00 individuals) in 2002 by Race and Gender

### 3 OLS Regression results

Table summarises the graph

Table 1: Incarceration Rate (as a proportion of 100,000 individuals) in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	792.2535	662.2517	2380.952	597.9761
Male	6027.3973	3094.9840	0.000	2356.0209

Following is a result of OLS regression based on the given model:

$$y = \beta_0 + x_1\beta_{male} + x_2\beta_{mixedrace} + x_3\beta_{nonblack} + x_4\beta_{hispanic} + \varepsilon$$

All covariates here are dummy variables. Race black and Gender female is omitted so everything is measured in reference to these classes. As expected all other ethnic groups have lower chances of being incarcerated compared to blacks. Men have higher chances of being incarcerated compared to women.

The outcome variable here gives the duration of incarceration. So coefficient of male would be interpreted as, Black males are incarcerated for 0.194 months longer than black females. The estimates in this regression seem to be statistically significant.

I used this variable because it varied among individuals. I could have used incarceration status (Y/N) but it would have led to an LPM model which i wanted to avoid for this assignment

Table 2: Regression Output. Omitted category is Black Females.

	<i>Dependent variable:</i>
	Incarceration Length in 2002
Hispanic	-0.159*** (0.038)
Mixed Race (Non-Hispanic)	-0.174** (0.083)
Non-Black / Non-Hispanic	-0.189*** (0.035)
Male	0.194*** (0.022)
Constant	0.155*** (0.026)
Observations	8,621
R <sup>2</sup>	0.015
Adjusted R <sup>2</sup>	0.014
Residual Std. Error	1.019 (df = 8616)
F Statistic	32.033*** (df = 4; 8616)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01