

Paper on NLSY97 Data of Incarceration Status

Xiaohan Wu

Spring 2022

1 Introduction

This report describes patterns in incarceration status by race and gender in the year 2002, using NLSY97 publicly available data. Before analyzing the dataset, the raw dataset was mined into a clean dataset with readable variables wanted.

The three variables to be analyzed are **race**, **gender**, and **total_incarcerated**. The variable **race** includes **Black**, **Hispanic**, and **Non-Black or Non-Hispanic**. The variable **total_incarcerated** represents the count of months of incarcerations in 2002 for the respondent who has the crime record.

2 Analysis

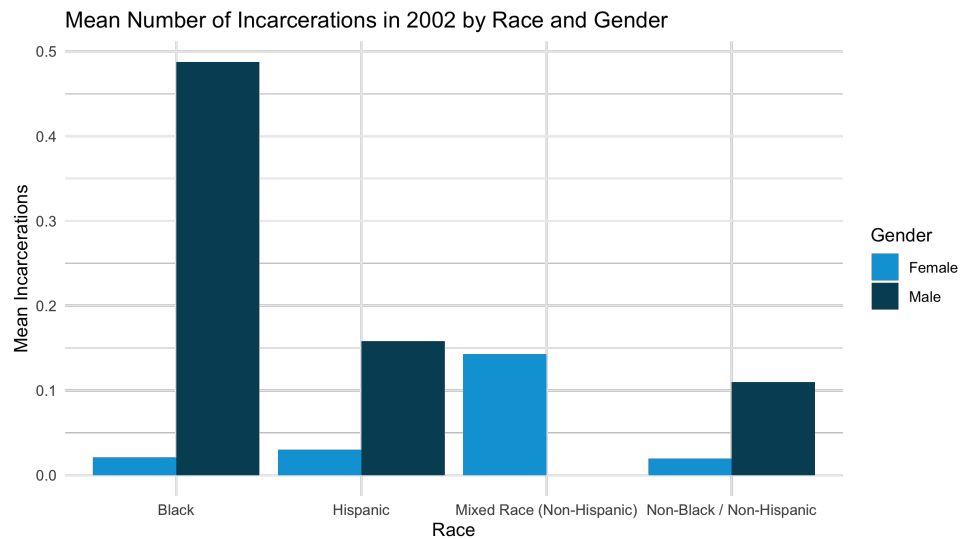


Figure 1: Mean Number of Incarcerations in 2002 by Race and Gender (this is the LaTeX caption, not the ggplot title)

Below is the Figure 1, which is a barplot showing the mean number of incarcerations in 2002 by race and gender.

Table 1: Mean incarcerations in 2002 by Race and Gender

Gender	Black	Hispanic	Mixed Race Non Hispanic	Non Black Non Hispanic
Female	0.0211268	0.0298013	0.1428571	0.0193192
Male	0.4876712	0.1579509	0.0000000	0.1099476

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

<i>Dependent variable:</i>	
Incarcerations 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 ²	0.015
Adjusted R ²	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	

3 The Next Steps