

# Data Analysis with Tableau

Andre Mendes (amd871)

February 7, 2016

## 1 Introduction

The goal of this project is to use a information visualization tool called Tableau to analyze a dataset called parole. Through the use of this visualization software, it is possible to try to explore and understand the relationship between decisions made by the Parole Board in New York State. The goal is to see if race and time spent in prison somehow affect the outcome of the parole interview.

## 2 Background Story

Every year in New York State over 10,000 prisoners appear in front of the Parole Board and are denied release. Many legal experts argue that there is little reason to the way that the New York State Parole Board makes these decisions. In response to this problem, Nikki Zeichner (a lawyer and ex criminal defense attorney) built a dataset of New York State Parole Board hearings based on publicly published parole hearing records scraped from the Parole Board's website. More information about this project can be found here: [parolehearingdata.org](http://parolehearingdata.org).

## 3 Objectives

Through analyzing this dataset, the goal is to find answers for the following questions:

- Who are the prisoners?
- When were they arrested and how long they wait for the interview?
- When do the interviews happen and how are the decisions so far?
- How do time and race relate to the decisions?

### 3.1 Who are the prisoners?

First, let's explore the data to try to understand basic information about the prisoners. Figure 1 shows how the prisoners are distributed with regards to sex and race. The reference material says that race was obtained through self declaration, but nothing was said about how the sex information was collected. Considering sex, it is possible to see that there are many more male (22,706) than female (1,456) prisoners. Since the difference in the number of male and female

## Sex and Race

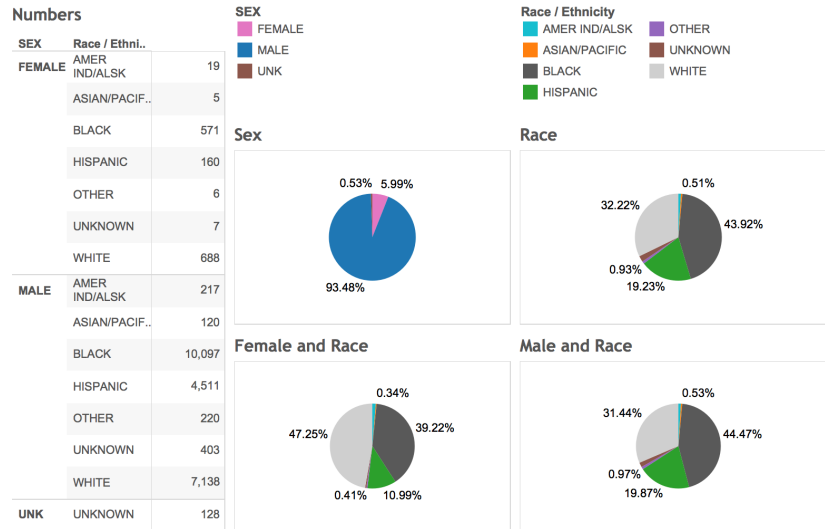


Figure 1: Distribution with regards to Race and Sex.

prisoners is huge, it is better to perform different analysis for each sex so that the data about men does not dominate the data about women. One example of this dominance is the distribution for race. When considering all people, the great majority is Black, followed by White and Hispanic. This is the same distribution observed when analyzing the race for men only. However, when taking only women in account, this distribution change because the number of white female prisoners is higher than black female prisoners. Lastly, the race and sex for 126 prisoners are classified as unknown in this dataset.

Another important basic information is the age for the prisoners. Data shows that most of men and women in our observations were born between 1960 and 1989 (around 84%). Since this data has been analyzed in 2016, it represents that they are between 27 and 56 years old.

### 3.1.1 When were they arrested and how long they wait to interview?

An analysis is performed comparing the age for when the prisoners entered the prison, the time they spend there before doing the last parole interview, and the age they were by the time of the interview.

When considering women, Figure 2 shows the entrance age follows an almost normal distribution with mean in 36 years and standard deviation of 11 years. As shown in the second column, more then 88% of the women do parole interview within the first two years in prison and for this reason the shape of the distribution in the last column, is similar to the first one (mean = 39 and std = 11). It is also important to notice that although the range says that most of them are younger than 56, there are much older prisoners in their 70s. With regards to men, in Figure 3, first column, it seems like the distribution shows that many of them are arrested in younger ages, such as 16 to 30. However, the

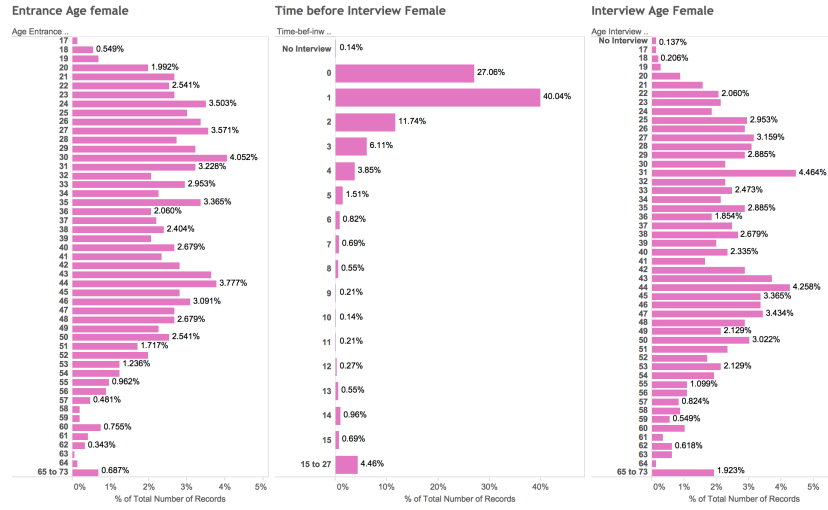


Figure 2: Age distribution for female prisoners.

mean for the distribution is 34 years and the standard deviation is 11 years, not much different from women. On the other hand, the second column shows that only 53% of male prisoner do parole interview before two years. This affects the last graph (mean = 40 and std = 12), showing that male tend to stay longer than female prisoners.

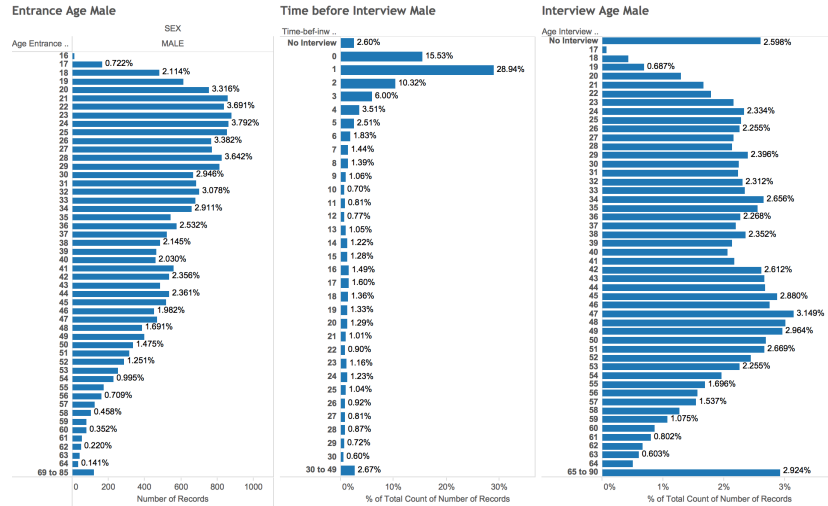


Figure 3: Age distribution for male prisoners.

There are many reasons to justify why this happens. One possible assumption is that the crimes committed by men are more serious than women, thus the sentence is longer. Therefore they stay longer in prison and take more time to do parole interview. Useful information to verify this, would be the type of crime committed. Unfortunately it was not possible find it in neither the data provided, nor in the website.

### 3.1.2 When do the interviews happen and how are the decisions so far?

In our dataset, the interviews available happened in 2011, 2012 and 2013. There are different types of interview, such as Initial, merit base, deportation, medical and reappear. Figure 4 shows how these different types are distributed along the 3 years considered and a logarithm scale is used to better compare the variation among the types.

The top graph shows the distribution in months. For this graph, it looks like there is pattern in the frequencies of the interviews but it is not exactly clear. However, it is possible to say that there are many variations with high peaks and low valleys in all almost types of interviews. When analyzing the medium graph, where the interval are quarters, it is noted that there is a clearer pattern that indicates that frequencies of all types of interviews are somehow cyclical. They increase in the first two quarters of the year and then drop towards the last two quarters. Finally, the bottom graph presents the distribution along the years and it shows that the number of interviews are increasing in the period.



Figure 4: Interview distributions over time.

Now let's analyze the decisions in each type of interview. In order to do that, two groups of decision are focused: Negative (Denied and not granted) and Positive (Granted and paroled). Also, the three most common types of interviews are used: Initial (51%), merit base (14%) and reappear (28%).

Data shows that is indeed hard to be granted paroled in the system for both male and female. For men, Figure 5, shows that the number of interviews are increasing and with that, the number of denied decisions. Meanwhile, the granted decision rate stays constant or even decreases as happened in Initial and Reappear types. For women, the scenario is even worst since in 2013 for example, all the Reappear interviews were denied, Figure 6.



Figure 5: Parole interview decisions for male prisoners.



Figure 6: Parole interview decisions for female prisoners.

Additionally, Figure 7 presents a curve that shows how the entrance year for the prisoners in our dataset. This shows that the number of new prisoners tend to increase along the years. All this information together confirms Zeichner concerns and if the parole board keeps denying requests and new prisoners arrive at this rate, the number of prisoners can increase significantly in the next years.

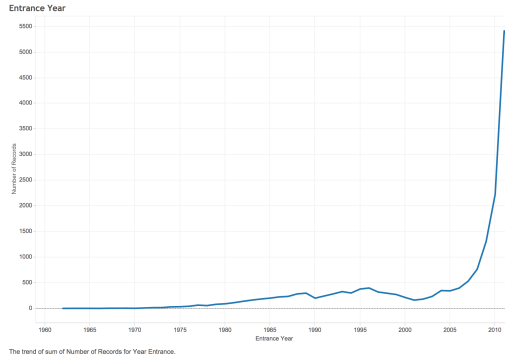


Figure 7: Entrance year for prisoners who did parole interview.

### 3.1.3 How do time and race relate to the decisions?

Finally, in this last section it is investigated the relations between the information previously explored. The goal is to understand if there is certain race that tends to be denied or accepted more often. Also, it is tried to determine if the time spent in prison before the interview have impact in the decisions.

With regards to time, since the majority of decisions are negative, the time distribution with regards to decisions is very similar to the original distribution presented in Figure 2 for women and Figure 3 for men. However, there are interesting points when taking race into account.

For women for example, Figure 8 shows that even though there are more White than Black prisoners, data shows that Black women received more positive responses than White women. For men, the surprise in Figure 9 is related to the Hispanic race. More than 50% of the positive responses are for Hispanic, although they represent only 19% of the male prisoners as shown in Figure 1.

## 4 Conclusion

This work presents basic analyzes of a dataset using Tableau, a powerful information visualization tool. From the results obtained, is possible to see that since 2011, more prisoners arrived and this rate tends to increase for next years. On the other hand, the parole board accept very few requests and it contributes for increasing the number of prisoners generating more costs for the state and possibly many problems for the inmates and their families. With regards to sex, race and time affecting the decisions, the analysis presented here is not sufficient to affirm that is any relation between these aspects and the decisions. Further analysis with more information such as type of crime and reason for parole decisions would improve the conclusions.



Figure 8: Relationship between decisions, time and race for female prisoners.

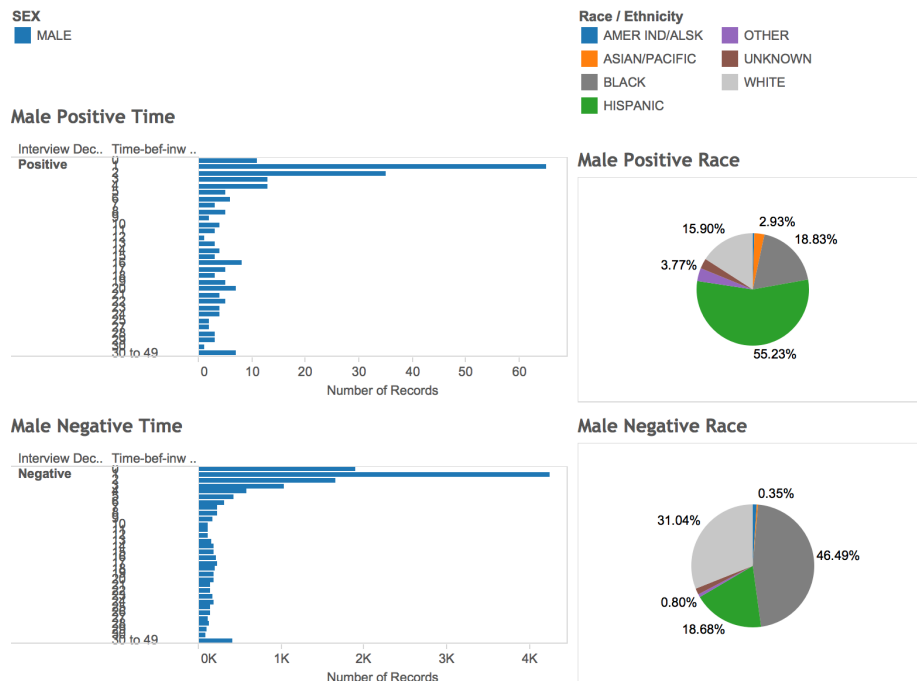


Figure 9: Relationship between decisions, time and race for male prisoners.