**NYC Daily Inmates in Custody (1991-2020)**

**Introduction:**

The proliferation of coronavirus around the world is one of the most significant pandemics that the current generation has ever faced. Millions of people are infected with this coronavirus and are undergoing treatment. But have we thought for the inmates who are in custody? Are the inmates healthy? Do they have any mental illness? These are a few questions that we have never bothered to think about, and hence, I find this topic interesting to explore.

New York has the highest number of inmates who are in custody, and its crucial to know about their mental health, which can be explained well with the help of few visualizations. This project is based on NYC Daily Inmates in Custody from 1991-2000. The dataset provided by the Department of Correction that falls into the Public Safety category and taken from the NYC Open Data. This dataset has daily update frequency, so there may be some inconsistencies with the number of rows and the number of columns in the dataset. There are 4143 observations and 13 columns in the dataset.

**Data Dictionary:**

As mentioned previously, there are 13 columns (variables) which have significant description and datatypes already made available by the NYC Open Data. Here is the list which of variables that I will be using for my analysis:

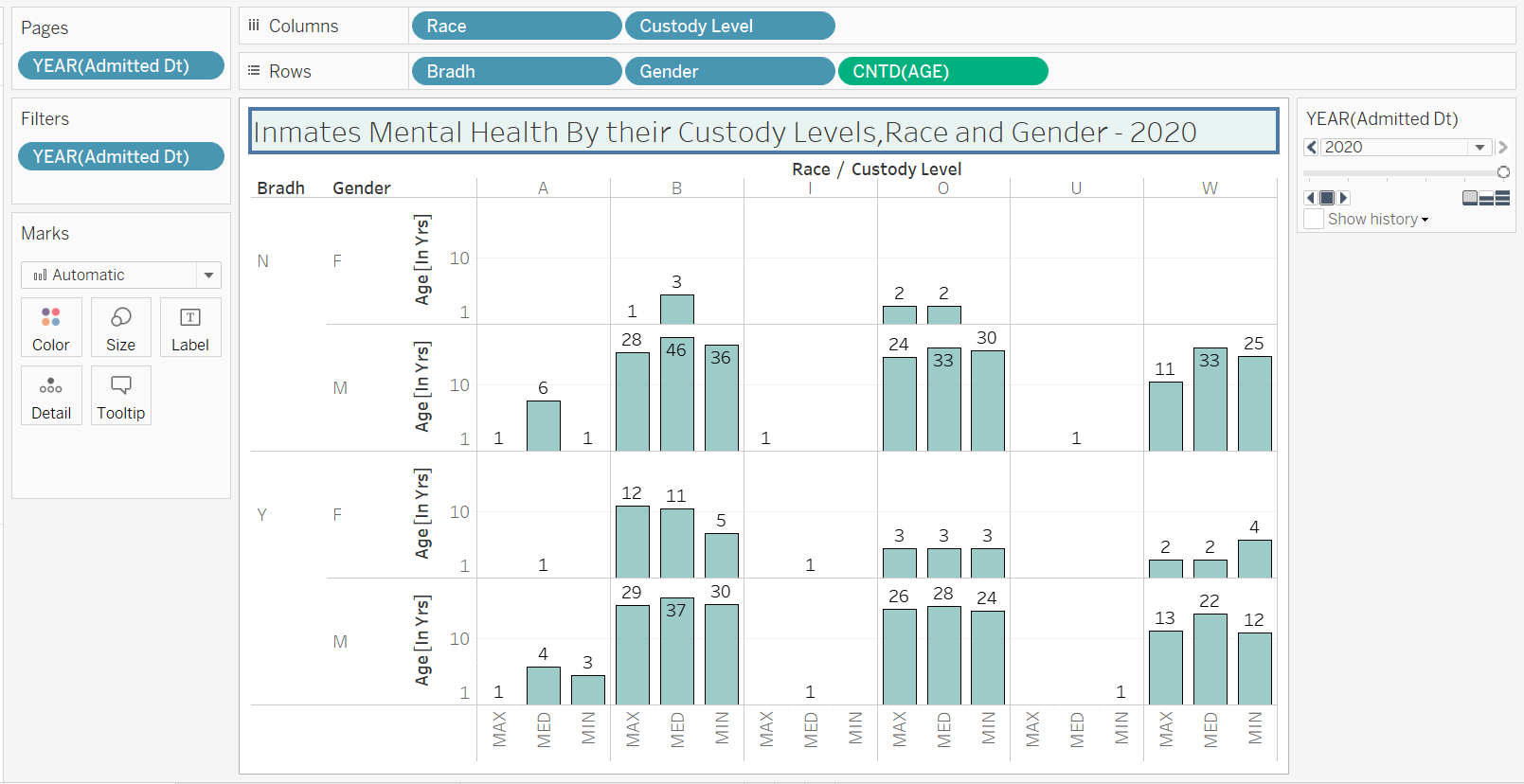
|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Description | Data Type | Expected Value |
| Inmateid | A primary key that uniquely identifies each inmate record. | Number |  |
| Admitted\_Dt | Admitted date and time of the incident | Date & Time |  |
| Custody\_level | Level of custody provided for the inmate. | String (Plain Text) | MIN, MED, and MAX |
| Bradh | Whether the inmate is under mental observation or not. | String (Plain Text) | YES or NO |
| Race | Tells us the Race of the inmate. | String (Plain Text) | A B, I, O, U and W |
| Gender | Gender of the inmate | String (Plain Text) | Male or Female |
| Age | Calculate the Age of the inmate | Number |  |
| Inmate\_Status\_Code | Provides the inmate status example if the inmate is a detainee. | String (Plain Text) | CS= City Sentenced CSP= City Sentenced - with VP Warrant DE= Detainee DEP= Detainee - with Open Case & VP Warrant DNS= Detainee- Newly Sentenced to State Time DPV= Detainee- Technical Parole Violator SCO= State Prisoner- Court Order SSR= State Ready |
| Infraction | Indicates whether the inmate has an infraction | String (Plain Text) | Yes or No |

**Data Cleaning:**

Data Cleaning is the process of detecting and correcting corrupt or inaccurate records from a table, database, a recordset, and refers to identifying incomplete, incorrect, inaccurate, or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data. I chose to do the data cleaning in python. The inmate’s dataset contained an entire column called “DISCHARGED\_DT,” which was null, i.e., all the values were missing. It was a better decision to omit this column from the dataset. To find the null values in the entire dataset, I have used the IsNull() function in python, which outputs that 661 values are invalid or missing in the dataset. These values are minimal in comparison to the existing observations, so I have dropped all the rows from the database. I will be using the dataset now, which has 3553 rows and 12 columns. There are three essential questions which I plan to answer with this dataset.

**Data Visualization:**

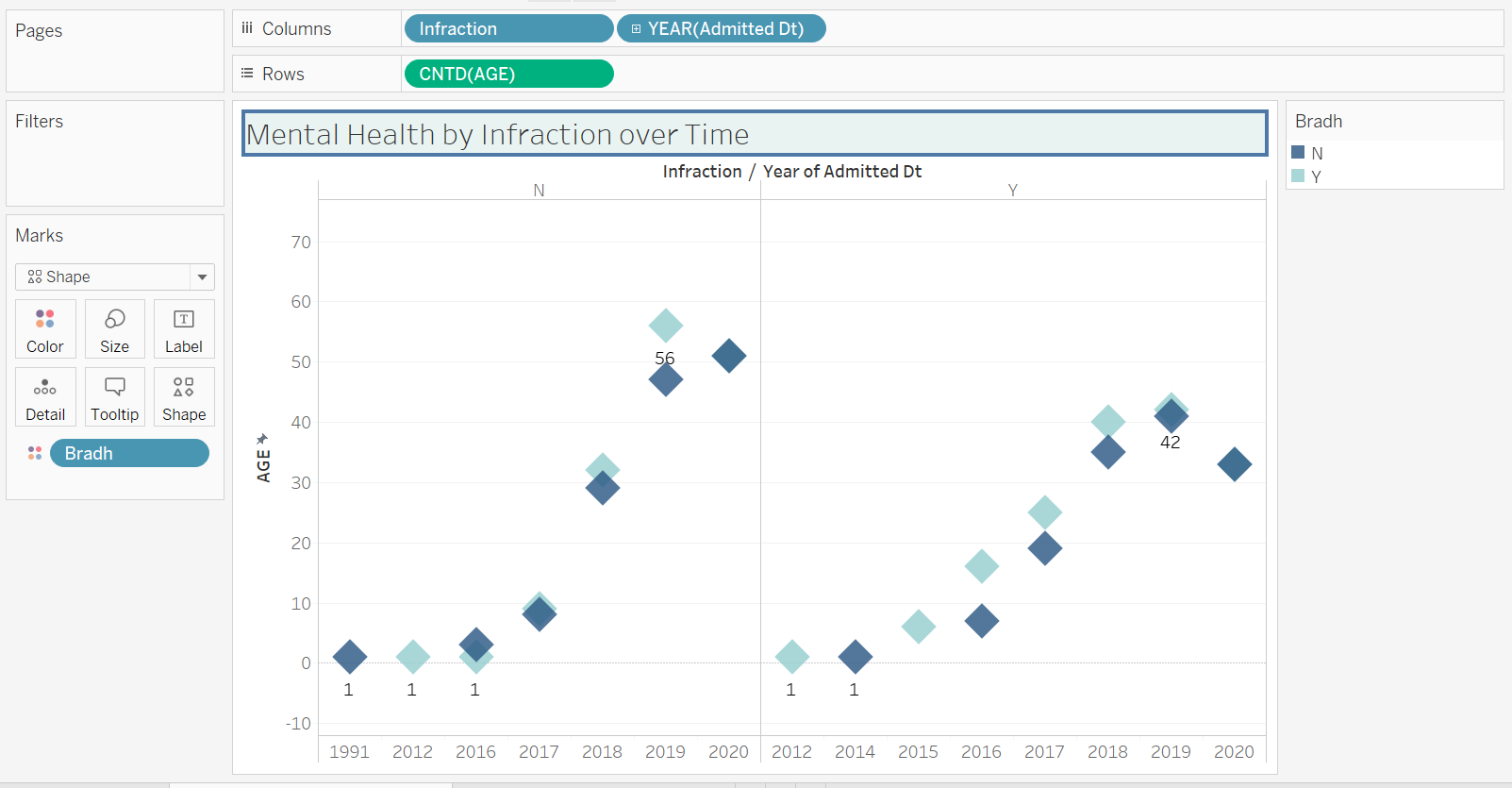
**1) What are the Race and the custody level that impacted the inmate’s mental health (2015-2020)?**



The above graph displays Race of the inmates which are A = American Native, B = Black or African American, I = American Indian, O = White (Non-Hispanic), U = Unknown, W = White against Custody levels that are Min, Med, and Max. Bradh has two values (Y= the inmate is under mental observation and N= not under mental observation). I have used Race, Custody level on the x-axis, while Bradh, Gender, and Age (Distinct of Age) has been placed on the y-axis. As there are no values for the years 1991, 2012, and 2014, I have removed them from Year (Admitted Dt) filter tile. I’ve place Year (Admitted Dt) on the tile Page. For the borders which are visible on the bar charts, I have selected Colors under the Marks tile, and Under Effects, I chose borders. For the numbers shown in the chart area, I have chosen tab Analysis 🡪Show Mark Labels.

**Conclusion:** The above graph tells us that there were very few inmates (5) in 2015, 12 in 2016, 25 in 2017 in NYC custody. These figures show that the number of inmates has continuously been increasing, and their mental conditions have been affected adversely. 2019 seems to have the maximum people who are under mental observation while 2015 have the least (5) people under subjective scrutiny. There is a slight deviation in the number of female criminals in 2019, as they are almost the same (13F and 16M). However, in other years the male criminals have dominated the females. A general trend has found that Black, White (Non-Hispanic) and White that they are the one who has been charged most for crimes in New York.

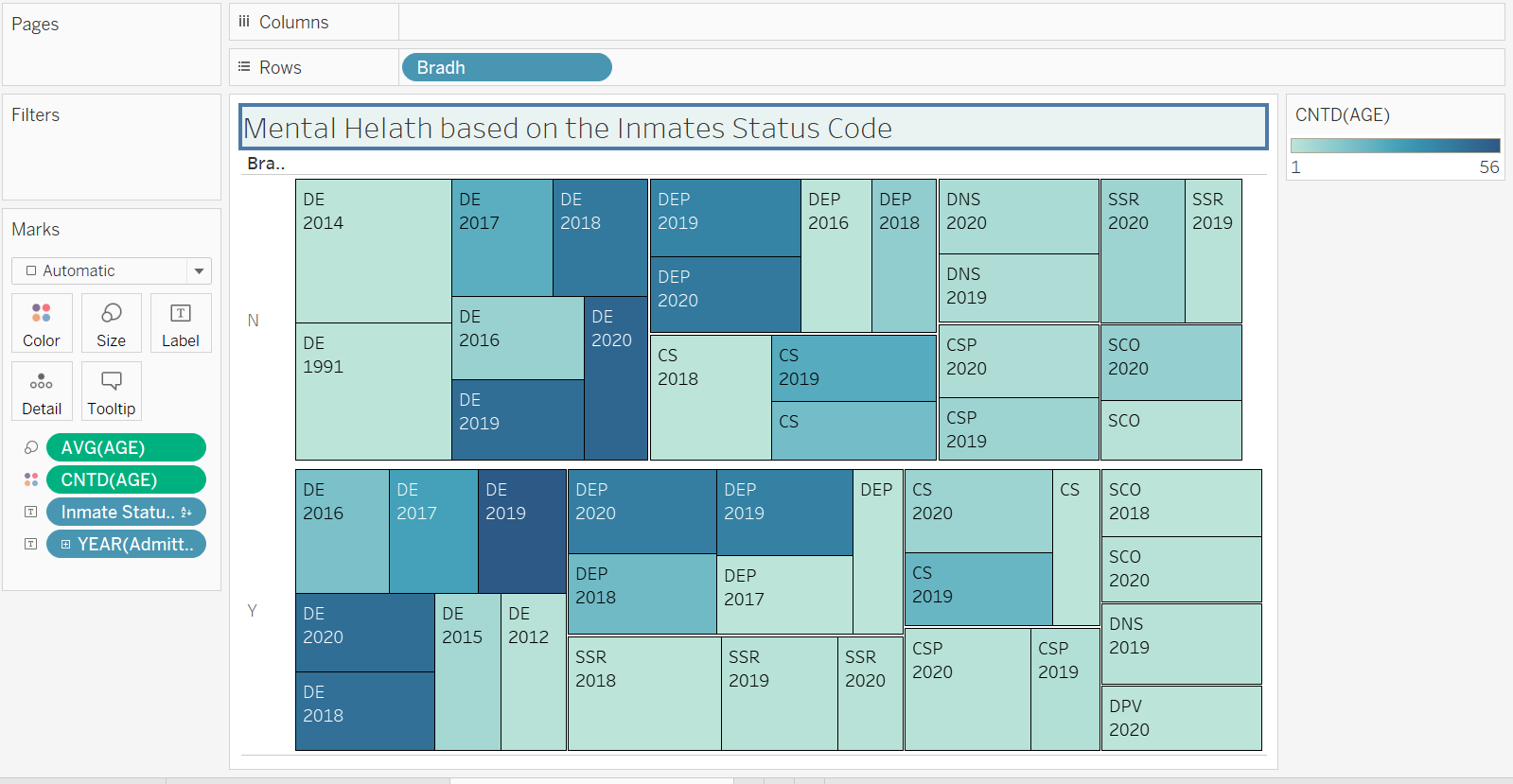
**2) Are the inmates arrested for Infraction or some severe crimes over the years?**



The second visualization here is a scatterplot between the Age of the inmates, the Year of admission, and Infraction. An infraction has two possible values (Y and N). For Age on the y-axis, I have used the logarithm scale. I have given colors for the binary values of Bradh. The scatterplot has the maximum and minimum age values displayed for both the Infraction conditions.

**Conclusion:** One conclusion we can draw from the visualization is that inmates with no mental illness and no infraction have been increasing every year. There is an increase in the infraction rate from 2012 to 2019, and it has shown a decline in 2020.

**3) What is the inmate’s status code concerning their mental health?**



Above is a treemap for Bradh, Inmate Status Code, and the Year of admission. I have used Age (Distinct) as my color filter, where light blue color depicts the lowest Age (1), and the darkest shade of blue has the maximum value for Age. I have labeled the treemap for different inmate’s status code and the Year (Admitted). The size of the boxes depicts the average Age of these inmates. There were plenty of inmates who belonged to the age group of 30-50 years so I thought taking an average of the age group would serve as a better option.

**Conclusion:** The average age of inmates is around 35-45 years and the inmates admitted in 2019 and 2020 suffer the most from mental illness.

The dashboard for all these three visualizations is shown below:

