HIV/AIDS ANNUAL REPORT DATA ANALYSIS

Name: Rahul Yadav Suresh

Student ID: 23072902

1. Brief Description of the data

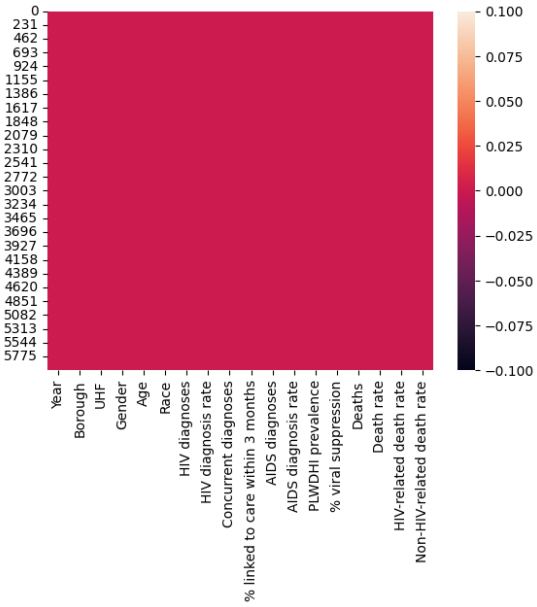
The dataset for this research is collected from Kaggle and this dataset contains information on the global AIDS/HIV epidemic for many years. This dataset contains 18 features such as Year, Race, HIV diagnoses, HIV diagnosis rate, Borough, UHF, Gender, Age, etc. This dataset provides information related to HIV/AIDS data from year 2011-2015. It also covers statistics for facilities, deaths, cases, and treatment of HIV/AIDS in countries worldwide annually. This information can be highly useful for the creation of the model that predicts HIV/AIDS rate in people based on historical data (Williams et al., 2023

2. Exploratory Data Analysis

```
# Data information
data.info()
```

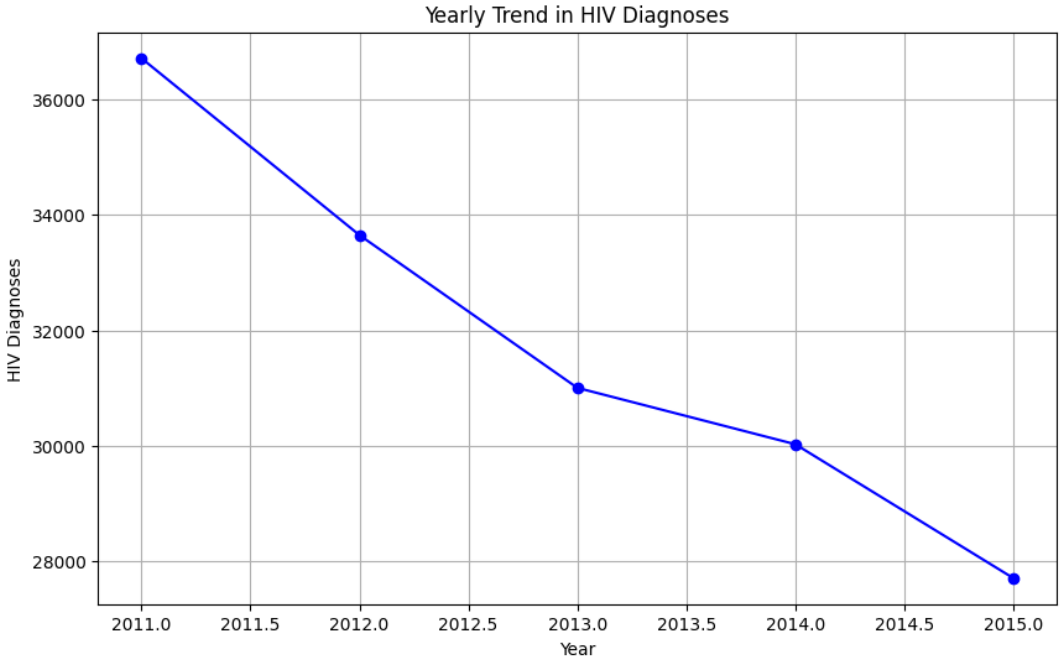
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6005 entries, 0 to 6004
Data columns (total 18 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Year                            6005 non-null   int64
 1   Borough                         6005 non-null   object
 2   UHF                             6005 non-null   object
 3   Gender                          6005 non-null   object
 4   Age                             6005 non-null   object
 5   Race                            6005 non-null   object
 6   HIV diagnoses                   6005 non-null   int64
 7   HIV diagnosis rate              6005 non-null   float64
 8   Concurrent diagnoses            6005 non-null   int64
 9   % linked to care within 3 months  6005 non-null   int64
 10  AIDS diagnoses                  6005 non-null   int64
 11  AIDS diagnosis rate             6005 non-null   float64
 12  PLWDHI prevalence               6005 non-null   float64
 13  % viral suppression             6005 non-null   int64
 14  Deaths                          6005 non-null   int64
 15  Death rate                      6005 non-null   float64
 16  HIV-related death rate          6005 non-null   float64
 17  Non-HIV-related death rate      6005 non-null   float64
dtypes: float64(6), int64(7), object(5)
memory usage: 844.6+ KB
```

This image provides function used for data information printing. **data.info()** function is used to display information of data.



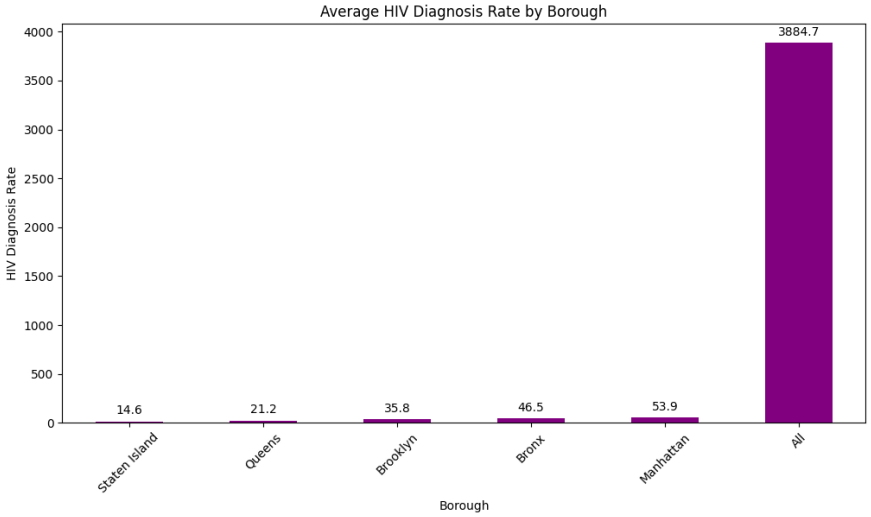
This image shows a heatmap from the seaborn library in Python, which says there are no null values across columns such as Year, Gender, borough, HIV Diagnoses, Death Rate, and Deaths, ensuring that all data are complete for health-related metrics.

1. Yearly Trend in HIV Diagnoses



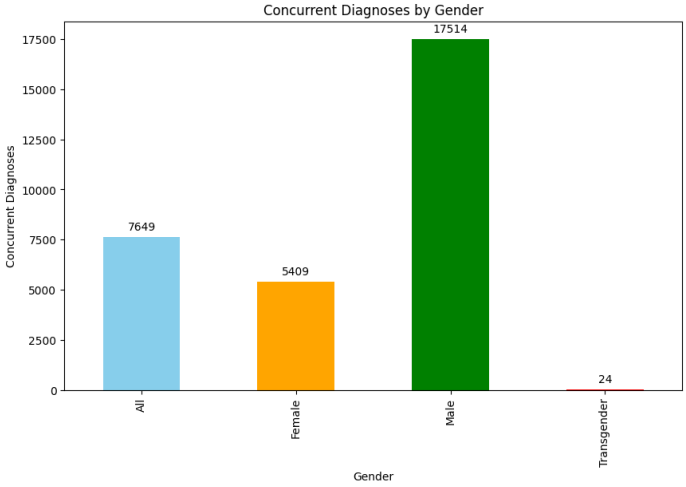
This graph shows yearly trend in HIV Diagnoses between year 2011 to 2015 (Giguère et al., 2021). Annual decreases underscore not only the positive impact that prevention strategies bring but also improved health outcomes.

2. HIV Diagnosis Rate by Borough



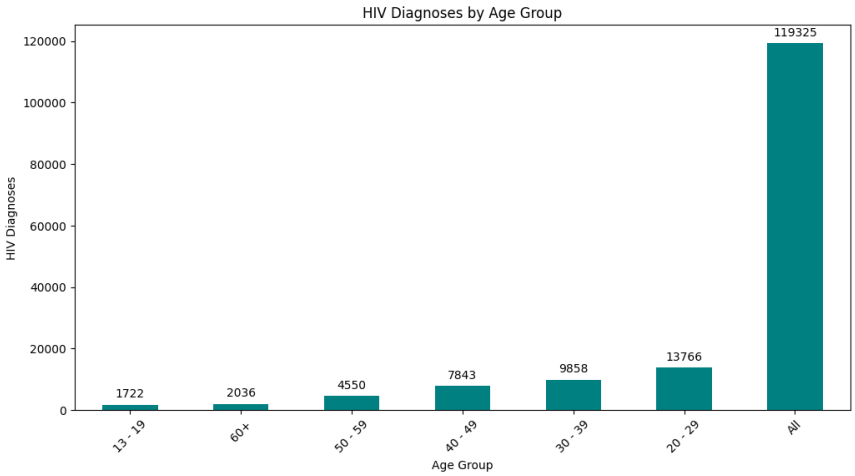
The above image shows the HIV diagnosis rate by borough from 2011 to 2015, HIV diagnoses remained steadily declining. The numbers declined from slightly above 36,000 to just below 28,000. This decrease indicates improved prevention strategies and better health outcomes over the five years.

3. Gender Disparity in Concurrent Diagnoses



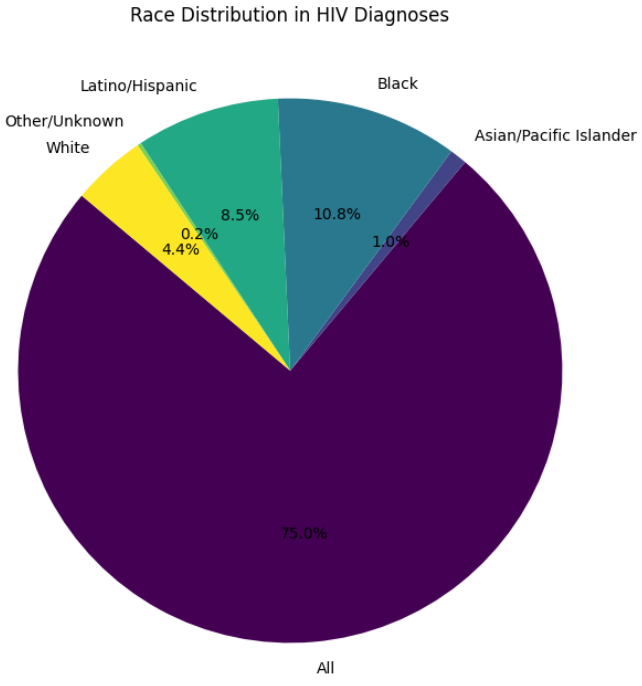
The bar plot illustrates a huge difference in the number of concurrent diagnoses between males and females and transgender, which is highest among males at approximately 17,500. Females and transgender have relatively fewer numbers, so there is a need for interventions that are specific to this group.

4. Age Group Distribution in HIV Diagnoses



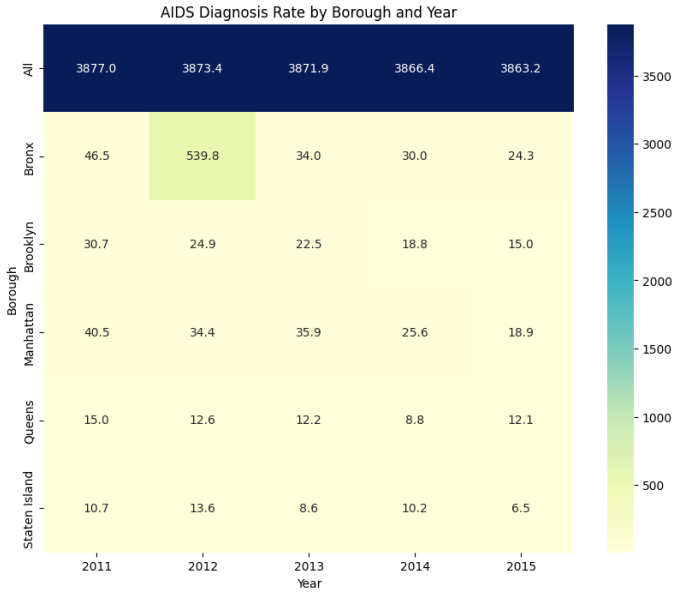
This bar graph shows HIV diagnoses by age group, showing the highest numbers in the "All" category and ages 30-39 (Carvalho, 2019). This demonstrates the widespread influence of HIV and the need for broad awareness and prevention efforts.

5. Race Distribution in HIV Diagnoses



The pie chart reveals that 75% of HIV diagnoses are unspecified, which highlights a critical data gap in addressing racial disparities.

6. AIDS Diagnosis Rate by Borough



The above heatmap shows that the HIV diagnoses are decreased from 2011 through 2015, indicating positive public health intervention and prevention methods.

7. Viral Suppression by Gender and Age

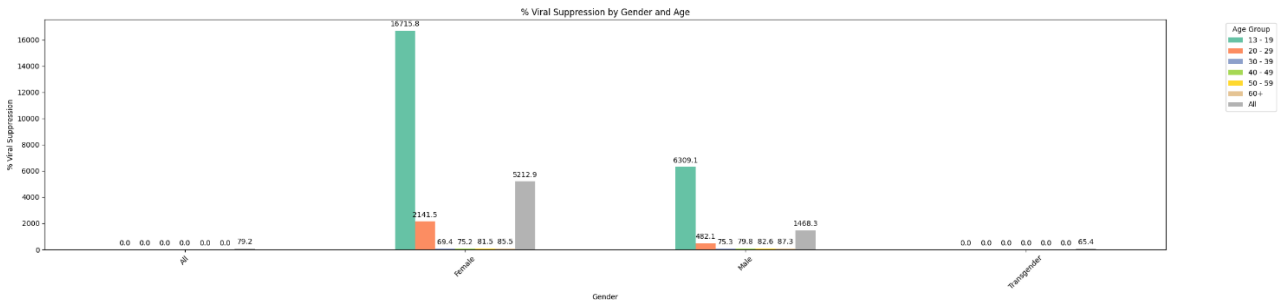
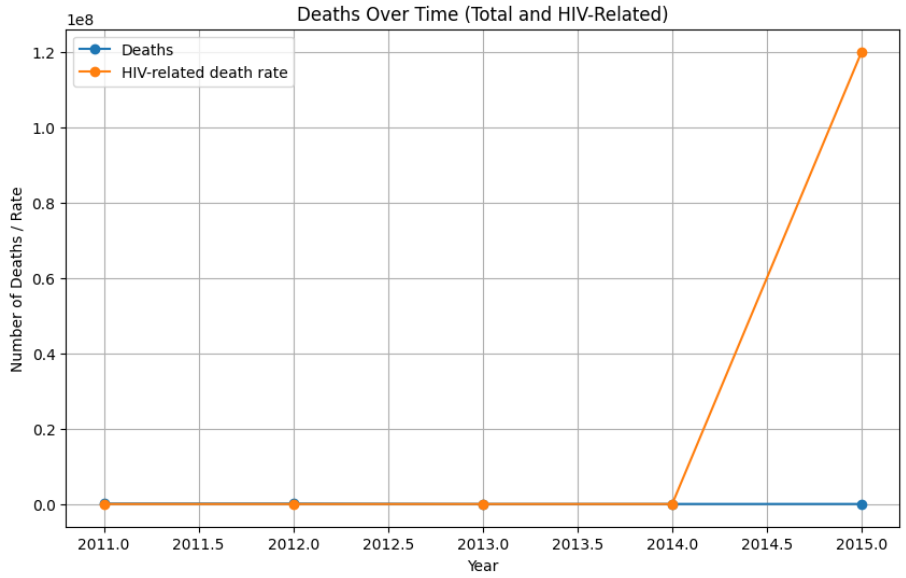


Figure 1: Viral Suppression by Gender and Age

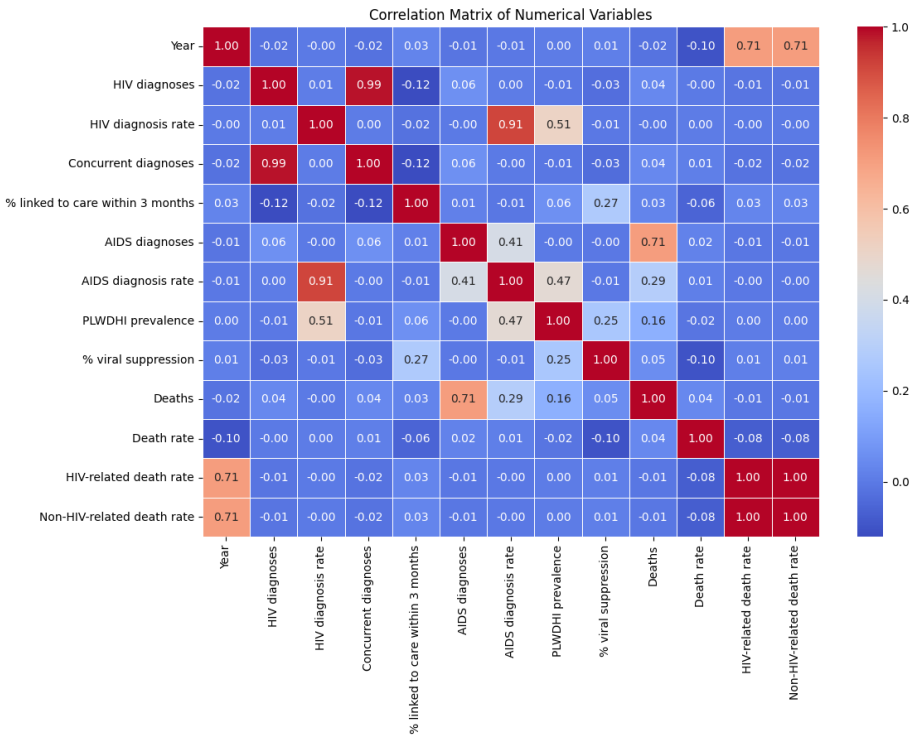
The above plot shows that females aged 60+ have the highest viral suppression rates, with the need to implement targeted interventions in other populations.

8. Deaths Over Time (Total and HIV-related)



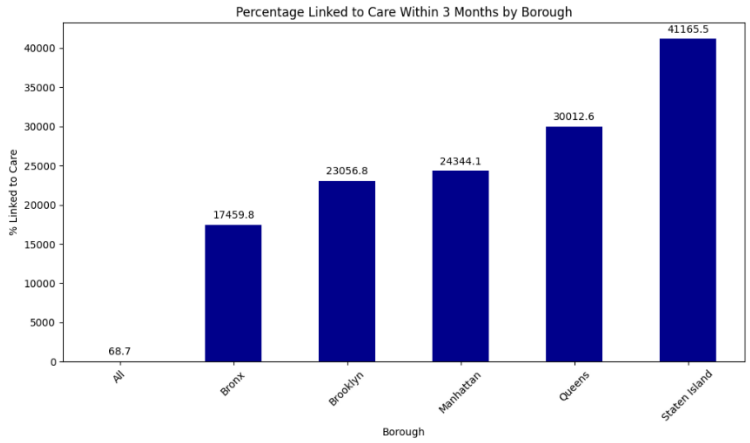
The above plot images show that in 2015, there was an increase in deaths due to HIV while total deaths remained steady; this may reflect a sudden change in health scenarios or reporting practices

9. Correlation Matrix of Numerical Columns



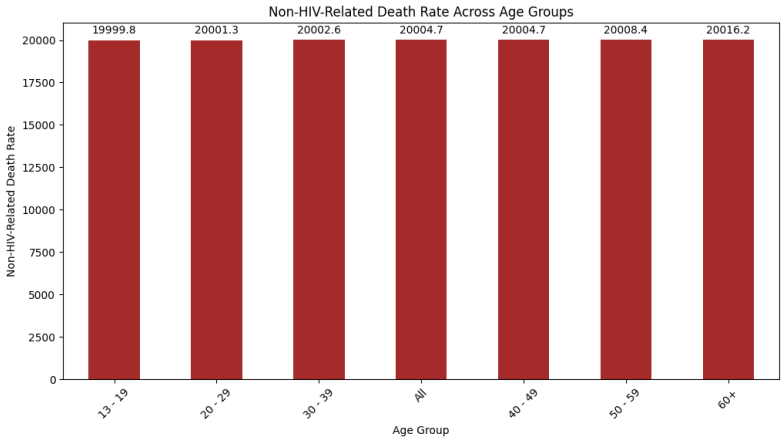
The correlation matrix shows strong correlations between HIV diagnosis and concurrent diagnoses as well as death rates, which reflect critical health insights.

10. Linked Care Analysis by Borough



The bar graph indicates that Staten Island (41165.5) leads in timely linkage to healthcare, while the Bronx lags (17459.8), which is a reflection of regional disparities in access.

12. Non-HIV-Related Death Rate Across Age Groups



From the bar graph, one realizes that non-HIV-related deaths have consistently occurred in large quantities throughout the entire age ranges studied.

Summary Statistics Table

	Borough	HIV diagnoses	AIDS diagnosis rate	% viral suppression
0	All	448.615385	3870.371538	75.392308
1	Bronx	21.868000	134.919700	1273.328000
2	Brooklyn	20.314667	22.380933	2074.791333
3	Manhattan	20.453818	31.063200	2330.571636
4	Queens	13.515636	12.118182	3712.434182
5	Staten Island	2.771200	9.891520	5190.884800

The table above summarizes statistics for HIV/AIDS in the New York City boroughs. It tabulates average HIV diagnoses, AIDS diagnosis rates, and viral suppression percentages. Manhattan and Brooklyn have shown higher rates of suppression, whereas the Bronx presents the highest rate of AIDS diagnosis (Cosenza-Nashat, 2016).

Recommendations

This could include a focus on trends and key summary statistics to derive actionable insights. More advanced visualization techniques, such as interactive plots or geospatial mapping, will enhance the interpretation of data and storytelling. All these approaches will provide a better understanding of the patterns, disparities, and progress in combating HIV/AIDS, hence supporting more targeted public health strategies.