Data Visualization Of gender inequality and HIV- AIDS Dataset using Tableau Software

Table of Contents

1 Introduction

- Original Title: **2021/W2: Gender Inequality in HIV Infections in Adolescents summary**
- The dataset contains the information regarding gender inequality in new HIV infections among adolescents
- Tableau can quickly connect the data model, visualize, and securely share insights even embed it all into app or website. It can connect with hundreds of data sources.
- The selection of the dataset was based on publicly available data. The data was collected from one of the open sources named as http://data.world/.
- Source: makeovermonday/2021w2 | Workspace | data.world
- Last updated at https://health.data.ny.gov : 2021-03-02
  Original data Source: UNICEF

2 Attributes of Dataset

Country
UNICEF Region
Year
Sex
Age
Estimated number of Annual new HIV infections
Estimated number of People living with HIV
Estimated number of Annual AIDS deaths
Rate of annual AIDS related deaths per 100000 population
Rate of new HIV infection per 1000 uninfected population

3 Data Visualization Tool

Tableau Software

4 Python Code for Data Cleaning

4.1 Data Preparation

In order to attain a successful data visualization, we a proper data preparation should be done. The data should be closely watched and remove all the unnecessary noises from the data. Higher quality data that can be processed and analyzed more quickly and efficiently leads to more timely, efficient, and high-quality decisions.

```
In [5]: data.isnull().sum()

Out[5]: Country                                                                  0
        UNICEF Region                                                            0
        Year                                                                     0
        Sex                                                                      0
        Age                                                                      0
        Estimated incidence rate of new HIV infection per 1 000 uninfected population     120
        Estimated number of annual AIDS related deaths                           60
        Estimated number of annual new HIV infections                           120
        Estimated number of people living with HIV                               0
        Estimated rate of annual AIDS related deaths  per 100 000 population     60
        dtype: int64

In [6]: df= data.dropna()
        df.isnull().sum()

Out[6]: Country                                                                  0
        UNICEF Region                                                            0
        Year                                                                     0
        Sex                                                                      0
        Age                                                                      0
        Estimated incidence rate of new HIV infection per 1 000 uninfected population     0
        Estimated number of annual AIDS related deaths                           0
        Estimated number of annual new HIV infections                            0
        Estimated number of people living with HIV                               0
        Estimated rate of annual AIDS related deaths  per 100 000 population     0
        dtype: int64

In [7]: data.duplicated().sum()

Out[7]: 0

In [22]: df.to_excel(r'C:\Users\reshm\Desktop\neha\1205\Assignments\assignment 4\updated.xlsx', index = False)
```

5 Data Cleaning

Importing to python was able to check for duplicates and null values. There were no duplicates present but null values were present.
120 null values were present in Rate of new HIV infection per 1000 uninfected population, Estimated number of Annual new HIV infections.
60 null values were present in Estimated number of Annual AIDS deaths, Rate of annual AIDS related deaths per 100000 population.
Hence all the Null Values had to be removed.

6 Visualization techniques

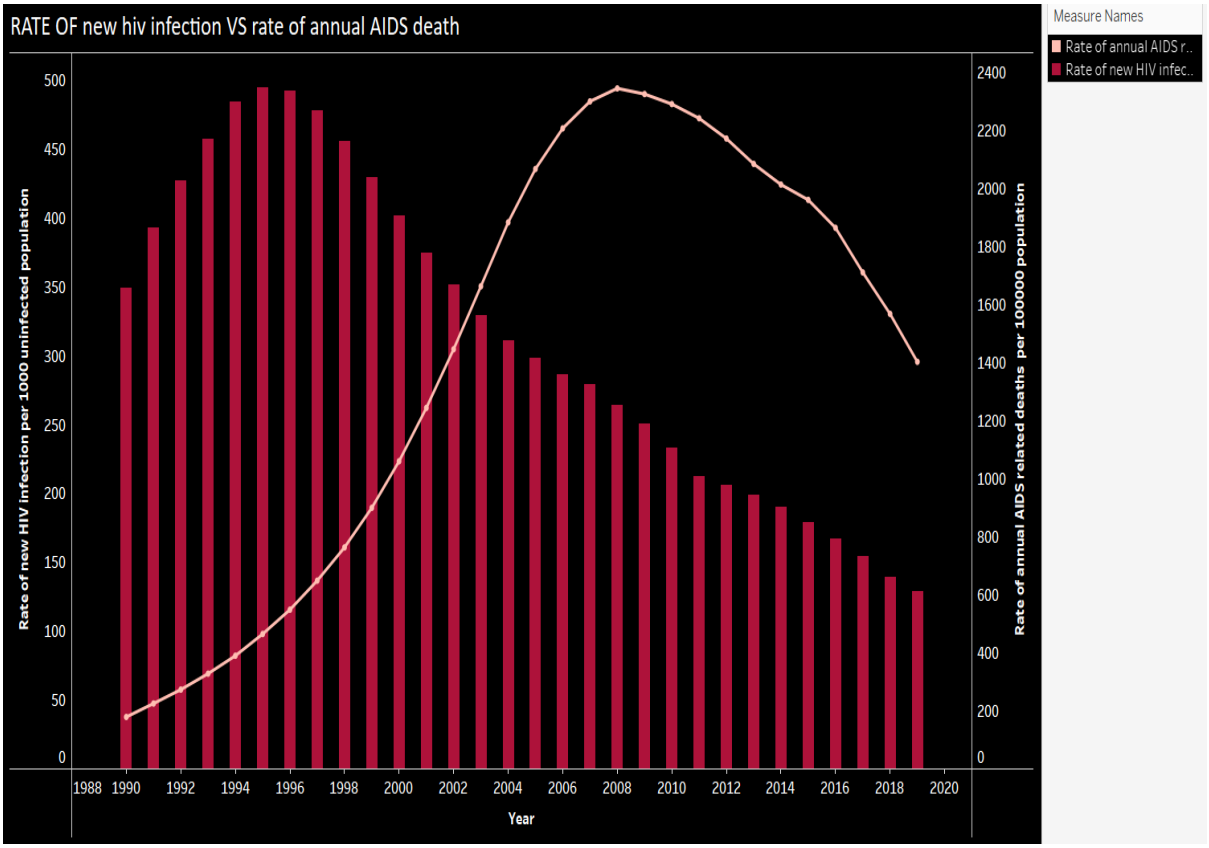
6.1 Simple Map

- Symbol maps are great for showing quantitative data for individual locations.
- Size of the green circle defines the count of new infection in each country

- Out of 41 countries are mentioned, South Africa has the highest with 21,01,6000 new infection rate. Which is followed by Mozambique with 5,8030.
- Gambia & Somalia has the lowest count with 6000 rates, followed by Djibouti and Madagascar with 6500 & 6700 rate respectively.
- It's clear that Only 8 countries have the rate greater than 3 lakhs.

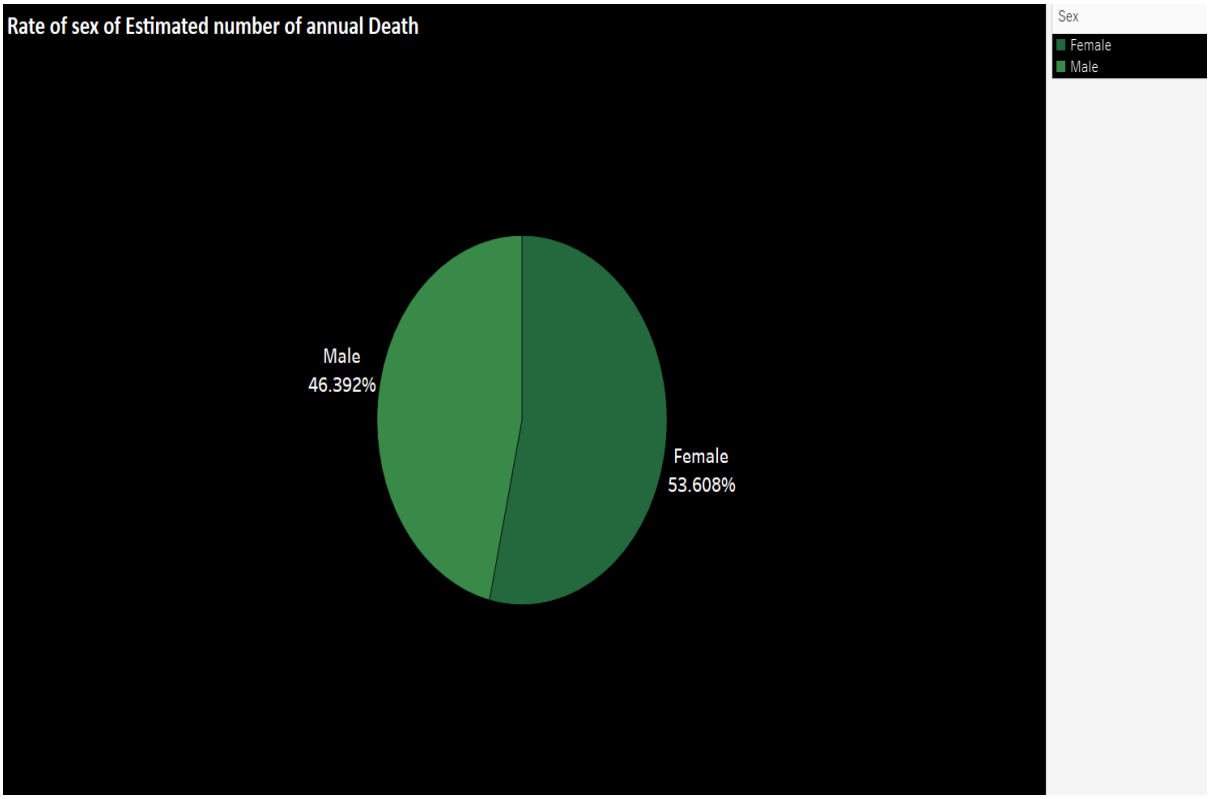
6.2 Dual combination Graph

- The Tableau Dual Combination Chart is handy to compare one Measure against Other measures visually.
- This graph shows the comparison between rate of newly HIV infected to rate of annual death 1990 to 2019.
- In 1990 the rate of new HIV infection was 350, while the rate of annual death was 179.
- In 1995 both the rate of newly infected and the death rate were almost same i.e. 495 & 465 respectively.
- By 2019 the rate of newly infected was lower compared to 1990. while that of annual death rate was comparatively higher.



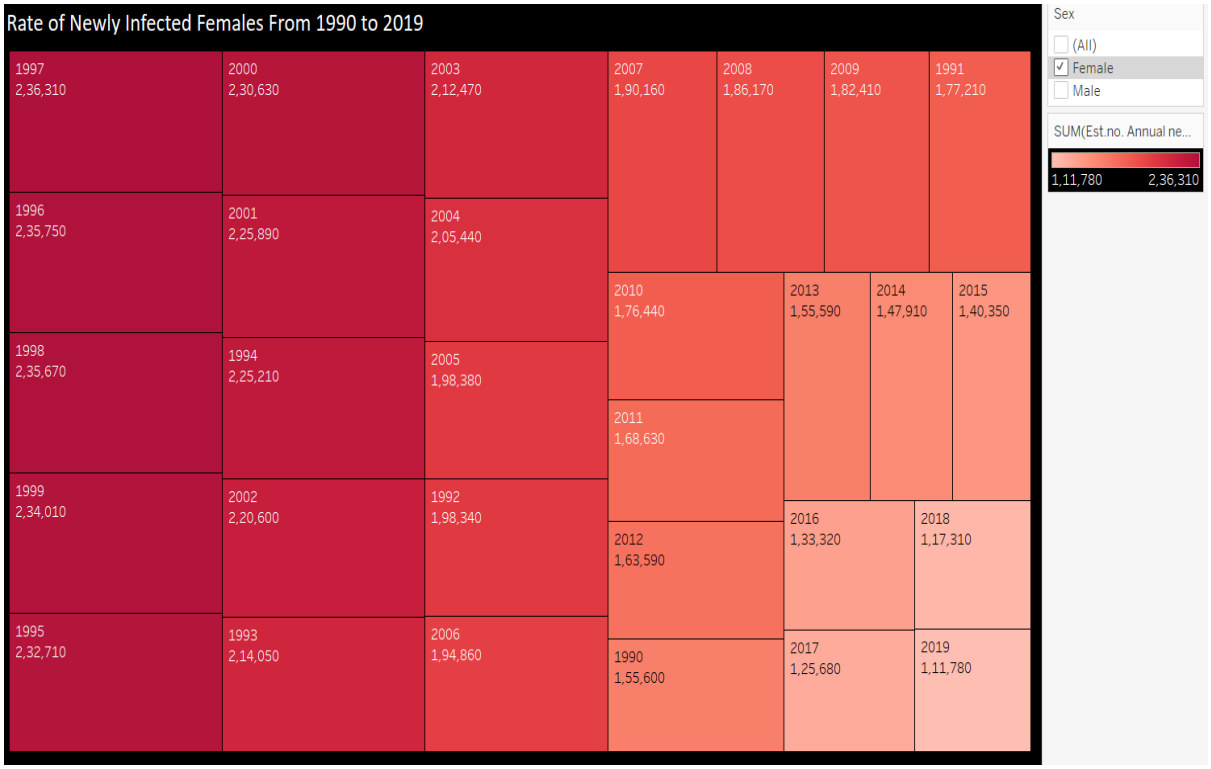
6.3 Pie Chart

- Pie charts helps to find out the composition of male and females died due to mentioned causes.
- The pie chart describes the ratio of males and females over the annual aids related death.
- In total 4,60730 males died annually due to AIDS. That is 46.392%.
- 53.608% of females died due to AIDS that is 532,400 of total population.



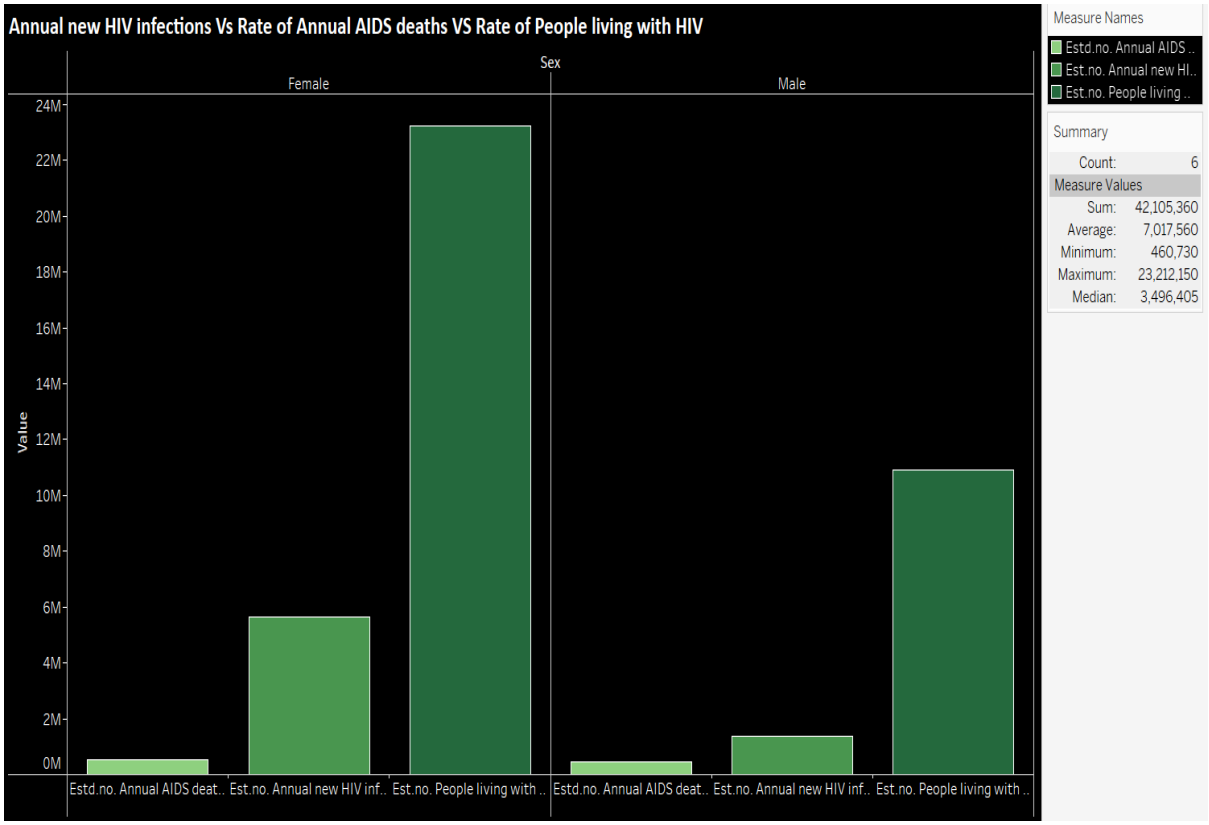
6.4 Tree Map

- Tree map in Tableau is a basic chart type that is represented by nested rectangular boxes. This chart can be useful for large datasets for visualization. Tree map is the graph that can mark the hierarchical data for comparative analysis.
- Size and color of the rectangles in tree map depends on the count of newly infected females over the given years.
- From 1993 to 2002 the count was highest it varied from 2,14050 to 2,36,310.
- While the least count was measured from the year 2016 to 2019 which varied between 1,11780 to 1,33,320



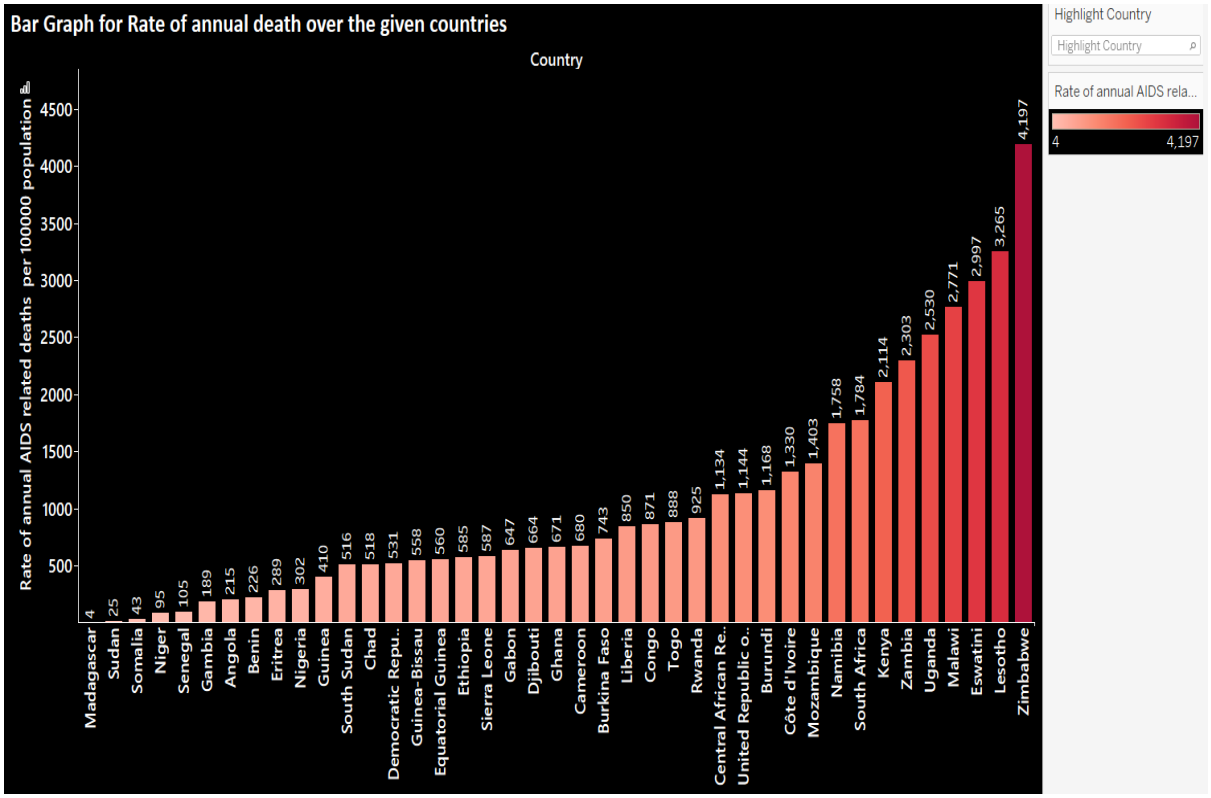
6.5 Side-by-Side Bar for Comparison

- In a side-by side bar chart, the bars are split into colored bar segment. This helps to compare between different variables of the given dataset.
- In this side-by-side bar graph three variables are compared which are Comparison of Estimated number of Annual new HIV infections, Annual AIDS deaths & People living with HIV.
- The Estimated annual death of Males is 460,730, while that of Females is 532,400.
- The Estimated annual new HIV infection of Males is 136,0340 while that of Females is 563,2470.
- Finally, the estimated number of females living with HIV is 232,12150 and that of males is 109,07270.
- Comparing all three variables estimated number of people living with HIV is very much high than other to for both females and males.

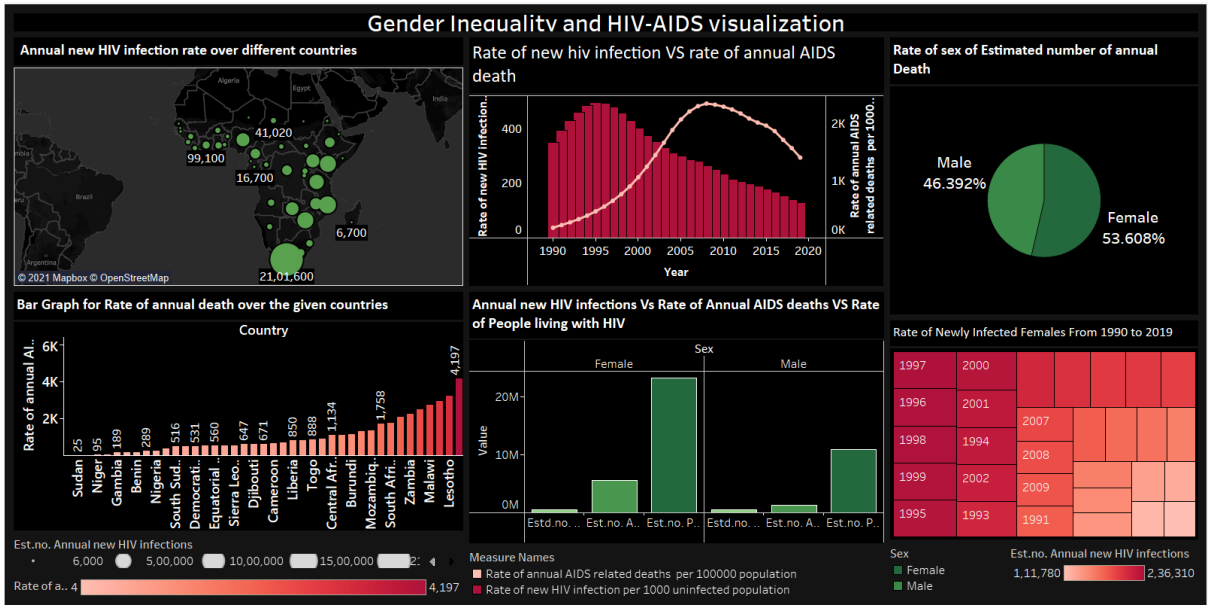


6.6 Bar Graph

- The Bar Graph is used to identify the frequency of the rate of annual death of given countries.
- The color shade gets darker as rate increase.
- Madagascar got the least value of 4 followed by Sudan and Somalia with 25 and 43 respectively.
- Zimbabwe has the annual death rate which is 4,197. followed by Lesotho Eswatini with 3,265 and 2,997 respectively.
- Out 41 countries 27 countries have rate less than 1000.



6.7 Statistical Dashboard



7 Conclusion

- Out of 41 countries are mentioned, South Africa has the highest with 21,01,6000 new infection rate.
- By 2019 the rate of newly infected was lower compared to 1990. while that of annual death rate was comparatively higher.
- Among the people died annually due to AIDS 46.392% were males & 53.608% were females.
- From 1993 to 2002 the count of newly infected females was the highest it varied from 2,14050 to 2,36,310.
- Comparing all three variables, estimated number of Annual new HIV infections, Annual AIDS deaths & People living with HIV, the estimated number of people living with HIV is very much high than other to for both females and males.
- Zimbabwe has the annual death rate which is 4,197. followed by Lesotho Eswatini with 3,265 and 2,997 respectively.

8 References

- Data sources you need right now | Tableau Public
- Tableau Course: Download Practice Datasets - SuperDataScience Pages - Machine Learning | AI | Data Science Career | Analytics | Success
- Tutorial: Get Started with Tableau Desktop – Tableau
- Tableau Tutorial | Step by Step Guide to Learn Tableau | Edureka
- About HIV/AIDS | HIV Basics | HIV/AIDS | CDC