K L E F

KONERU LAKSHMAIAH EDUCATION FOUNDATION

(Deemed to be university estd, u/s, 3 of the UGC Act, 1956)

(NAAC Accredited "A" Grade University)

PROJECT REPORT

On

THE DEATH OF AMERICA'S RURAL HOSPITALS

Submitted in partial fulfilment of the

Requirements for the award of the Degree of

Bachelor of Technology

In

Computer Science and Engineering

Under the esteemed guidance of

B. Manjula

By

Maganti Sreelekha Chowdary

(DST-FIST Sponsored Department)

K L EDUCATION FOUNDATION

Green Fields, Vaddeswaram, Guntur District-522 502

2019-2020

K L EDUCATION FOUNDATION

K L EDUCATION FOUNDATION

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(DST-FIST Sponsored Department)

CERTIFICATE

This is to certify that this project based lab report entitled "**THE DEATH OF AMERICA'S RURAL HOSPITALS** " is a bonafide work done by **Maganti Sreelekha (180030523)** in the course 18CS3262S DATA MODELING & VISUALIZATION in partial fulfilment of the requirements for the award of Degree in Bachelor of Technology in COMPUTER SCIENCE AND ENGNEERING during the Even Semester of Academic year 2020-2021

Faculty in Charge
B.Manjula

Head of the Department
Mr Hari Kiran Vege

K L EDUCATION FOUNDATION

DEPT OF COMPUTER SCINCE AND ENGINEERING

(DST-FIST Sponsored Department)

DECLARATION

We hereby declare that this project based lab report entitled "**THE DEATH OF AMERICA'S RURAL HOSPITALS**" has been prepared by us in the course 18CS3262S DATA MODELING & VISUALIZATION in partial fulfilment of the requirement for the award of degree bachelor of technology in COMPUTER SCIENCE AND ENGINEERING during the Even Semester of the academic year 2020-2021. We also declare that this project-based lab report is of our own effort and it has not been submitted to any other university for the award of any degree.

Date: 3/11/2020

Place:VIJAYAWADA

ACKNOWLEDGEMENT

Our sincere thanks to B.Manjula mam in the Lab for outstanding support throughout the project for the successful completion of the work.

We express our gratitude to Dr P.Vidyullatha Course Co-ordinator for 18CS3262S DATA MODELING & VISUALIZATION course in the Computer Science and Engineering Department for providing us with adequate planning and support and means by which we can complete this project-based Lab.

We express our gratitude to Mr. V. HARIKIRAN, Head of the Department for computer science and Engineering for providing us with adequate facilities, ways and means by which we can complete this project-based Lab.

We would like to place on record the deep sense of gratitude to the honourable Vice Chancellor, K L University for providing the necessary facilities to carry the project-based Lab.

Last but not the least, we thank all Teaching and Non-Teaching Staff of our department and especially our classmates and our friends for their support in the completion of our project-based Lab.

Name of the student

Maganti Sreelekha

180030523

TABLE OF CONTENTS

ABSTRACT:

The project is totally based on nonmetropolitan and metropolitan areas have significant numbers of deaths from the five leading causes of death. These figures accompany this report by presenting information about deaths in nonmetropolitan and metropolitan areas at the state level i.e rural states of America. They also add additional years of data and options for selecting different age ranges and benchmarks.

The dataset presents number of deaths and age-adjusted death rates for the 5 leading causes of death i.e cancer,unintentional injury,stroke,respiratory and heart dieases in the United States and by rural states for 2005-2013. This real time dataset i.e National Centre for Health Statistics by simply showing a dataset that contains 165125 more or less people may confuse while seeing a patient records In these project we have applied data preprocessing,dataanalysis and data visualization.techniques

INTRODUCTION:

This Project is mainly about visualizing the number of deaths in Rural States of America.

In the United States, there is a rural-urban gap in age-adjusted death rates and potentially excess deaths from the five leading causes of death. Rural communities experience higher age-adjusted death rates and a higher number of potentially excess deaths from the five leading causes compared with urban areas. Higher death rates and potentially excess deaths are often associated with various interconnected societal, geographic, behavioral, and structural factors. Although rural communities are at higher risk for death from the five leading causes of death, funding to address risk factors is allocated on a population basis often resulting in underfunded rural programs. An increased emphasis on need and epidemiologic burden of disease as major factors in targeting future allocation of public health and prevention funding might contribute, among other factors, to bridging the mortality gap from the five leading causes of death between rural and urban areas in the United States

PROBLEM AND DEFINITION:

This Project is about visualizing the cause of deaths in Rural states of America. It is used to demonstrate and visualize the no of deaths in rural states of America using tableau and python, here we represented the no of deaths from various aspects and the ways to cure them. major no of deaths are causing to the people

For many life-threatening diseases, the common risk factors include overweight,smoking,drinking,stressful lifestyle etc like symptoms'Cancer,stroke,respiratory and heart dieases are some very serious diseases across the states that have resulted in a lot of deaths.Causes of deaths can be life-threatening and requires immediate emergency

1.3-SCOPE:

Compared with metropolitan areas, nonmetropolitan areas have higher age-adjusted death rates and greater percentages of potentially excess deaths from the five leading causes of death, nationally and across public health regions.

Not all **potentially excess deaths** can be prevented; some areas might have characteristics that predispose them to higher rates of death. However, many potentially excess deaths might represent deaths that could be prevented through improved public health programs that support healthier behaviors and neighborhoods or better access to health care services.

PURPOSE:

To demonstrate and visualize the cause of deaths in rural states of America using tableau and python, here we represented the no of deaths from various rural states of America and the ways to cure them. Major no of deaths are causing due to cancer,unintentional injury,heart and repository dieases

PROBLEM AND EXISTING TECHNOLOGY:

To better understand digital visualization challenges, first let us understand whatexactly data visualization is. As defined by Oracle, "Data visualization is the presentation of abstract information in graphical form. Data visualization allows us to spot patterns, trends, and correlations that otherwise might go unnoticed in traditional reports, tables, or spreadsheets."In the early days of data sciences, businesses were suffering to calculate margins around the globe. They used to maintain data by Excel, CSV, MS Access Application,Flat files etc. It was very tough to maintain & visualize data. Especially when connecting to legacy systems, fetching data takes so much time & estimating the time to extract is not constant every time. Even after development, we must check whether the data visualization is similar on all the browser platforms or not.After completing development, we need to host manually on either the individual server or cloud servers. This takes huge time to publish the report. After publishing the report, we need to provide authentication & authorization for security purpose.Authorization is to validate the user and ensure that user is having permission to view or not.

PROPOSED SYSTEM:

Mortality data for U.S. residents from the National Vital Statistics System were used to calculate age-adjusted death rates and potentially excess deaths for nonmetropolitan and metropolitan areas for the five leading causes of death. Age-adjusted death rates included all ages and were adjusted to the 2005 U.S. standard population by the direct method. Potentially excess deaths are defined as deaths among persons aged <80 years that exceed the numbers that would be expected if the death rates of states with the lowest rates (i.e., benchmark states) occurred across all states. (Benchmark states were the three states with the lowest rates for each cause during 2008–2010.) Potentially excess deaths were calculated separately for nonmetropolitan and metropolitan areas. Data are presented for the United States and the 10 U.S. Department of Health and Human Services public health regions.

REQUIREMENTS AND ANALYSIS:

PLATFORM REQUIREMENTS:

OS WINDOWS 7 OR LATER/MAC

TOOL:PYTHON JUPITER NOTEBOOK,TABLEAU

LANGUAGE:PYTHON

RAM:1GB/MORE

HARDDISK:6GB/MORE

PROCESSOR:ANY INTEL PROCESSOR

MODULE DESCRIPTION:

This module makes us understand the statistics of the american states death rate. We have analysed the data and visualized to clearly understand the death rate and how it got affected with each attribute provided in the dataset

ALGORITHM:

IN JUPYTER:

Step 1 : Open jupyter notebook.

Step 2 : import the packages what ever we need for this project.

Step 3 : read input data in csv

Step 4 : Use pandas datafram for reading csv file

Step 5 : Display the data present in the input file

Step 6 : Now visualize the data that represents the opioid deaths in Connecticut.

IN TABLEAU:

Step 1 : Open Tableau Desktop

Step 2 : Click on the text file or Microsoft excel based on the data

Step 3 : Import dataset

Step 4 : Perform the visualization based on problem statement and drag the values into rows and columns whatever we need.

Step 5 : Apply different colours, shapes to the visualization for understanding easily

Step 6 : Rename the sheets Step

7: Save the book.

PSEUDO CODE:

#180030523

```
import pandas as pd import numpy as np from

sklearn.model_selection import train_test_split

import seaborn as sns import matplotlib.pyplot as

plt import missingno as msno


%matplotlib inline import warnings warnings.filterwarnings("ignore")

import chart_studio.plotly as pt import plotly.graph_objs as go from

plotly.offline import download_plotlyjs,init_notebook_mode,plot,iplot

init_notebook_mode(connected = True)

df.describe()
df.info() df.isnull().sum() dfcod=df['Cause of Death'].value_counts() dfcod
names=['Heart Disease','Stroke','Cancer','Respiratory disease','injury'] fig =
plt.figure(figsize =(10, 7))
plt.pie(dfcod,colors=['yellow','lightblue','lightgreen','pink','hotpink'],labels=names)
plt.show() dfl=df['Locality'].value_counts()
dfl
dfmet=df[df['Locality']=='Metropolitan']
```

```
dfem=dfmet['Expected                Deaths'].sum()

dfpedm=dfmet['Potentially Excess Deaths'].sum()

dfnon=df[df['Locality']=='Nonmetropolitan']

dfen=dfnon['Expected Deaths'].sum()

dfpedn=dfnon['Potentially Excess Deaths'].sum()
locality=['Metropolitan','Nonmetropolitan']

Expected=[dfem,dfen] Ped=[dfpedm,dfpedn]

dfplt=pd.DataFrame({'Locality':locality,'Expected

Deaths':Expected,'PED':Ped}) width=0.35

fig = plt.figure(figsize=(5,5)) p1=plt.bar(dfplt['Locality'],dfplt['Expected

Deaths'],width,color='crimson')

p2=plt.bar(dfplt['Locality'],dfplt['PED'],width)

plt.yticks(np.arange(10000000,50000000,10000000))

plt.legend((p1,p2),('Excess Deaths','Expected Deaths'))

dfstate=df['State'].value_counts() dfstate

msno.heatmap(df) msno.bar(df,color='green')

plt.figure(figsize=(8,6))

sns.heatmap(df.corr(),annot=True)

df.isnull().sum()

df.hist(bins=10,figsize=(20,20)) plt.show()
```

SCREENSHOTS:

Output:1

	Year	Cause of Death	State	State FIPS Code	HHS Region	Age Range	Benchmark	Locality	Observed Deaths	Population	Expected Deaths	Potentially Excess Deaths	Percent Potentially Excess Deaths
0	2005.0	Cancer	Alabama	AL	4.0	0-49	2005 Fixed	All	756.0	3148377.0	451.0	305.0	40.3
1	2005.0	Cancer	Alabama	AL	4.0	0-49	2005 Fixed	Metropolitan	556.0	2379871.0	341.0	217.0	39.0
2	2005.0	Cancer	Alabama	AL	4.0	0-49	2005 Fixed	Nonmetropolitan	200.0	768506.0	111.0	89.0	44.5
3	2005.0	Cancer	Alabama	AL	4.0	0-49	2010 Fixed	All	756.0	3148377.0	421.0	335.0	44.3
4	2005.0	Cancer	Alabama	AL	4.0	0-49	2010 Fixed	Metropolitan	556.0	2379871.0	318.0	238.0	42.8

Output 2

Out[4]:

	Year	HHS Region	Observed Deaths	Population	Expected Deaths	Potentially Excess Deaths	Percent Potentially Excess Deaths
count	165124.00000	165124.000000	156824.000000	1.608820e+05	156824.000000	156824.000000	156824.000000
mean	2008.93405	5.237216	2964.486303	6.965942e+06	2095.181898	879.752015	35.618019
std	2.55710	2.822844	17159.402910	2.839582e+07	13138.396045	4472.576572	18.537710
min	2005.00000	0.000000	10.000000	5.666500e+04	2.000000	0.000000	0.000000
25%	2007.00000	3.000000	153.000000	6.819730e+05	91.000000	41.000000	21.100000
50%	2009.00000	5.000000	503.000000	1.608654e+06	296.000000	158.000000	35.600000
75%	2011.00000	8.000000	1571.000000	4.642670e+06	1011.000000	539.000000	49.800000
max	2041.00000	45.000000	486080.000000	3.100880e+08	442680.000000	175703.000000	84.300000

Output 3

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 165125 entries, 0 to 165124
Data columns (total 13 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   Year                            165124 non-null  float64
 1   Cause of Death                  165124 non-null  object
 2   State                           165124 non-null  object
 3   State FIPS Code                 165124 non-null  object
 4   HHS Region                      165124 non-null  float64
 5   Age Range                       165124 non-null  object
 6   Benchmark                       165124 non-null  object
 7   Locality                        165124 non-null  object
 8   Observed Deaths                 156824 non-null  float64
 9   Population                      160882 non-null  float64
 10  Expected Deaths                 156824 non-null  float64
 11  Potentially Excess Deaths       156824 non-null  float64
 12  Percent Potentially Excess Deaths  156824 non-null  float64
dtypes: float64(7), object(6)
memory usage: 16.4+ MB
```

Output 4

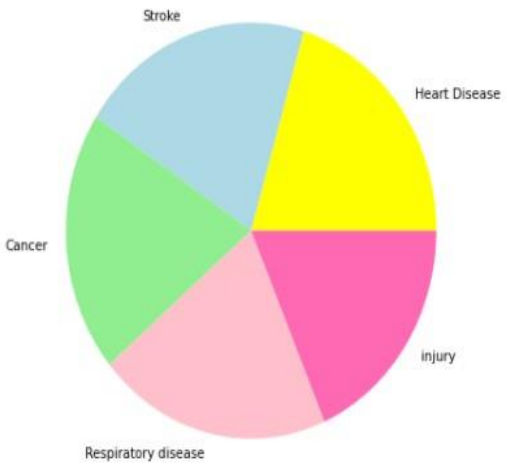
```
In [6]: df.isnull().sum()

Out[6]: Year                                1
        Cause of Death                      1
        State                               1
        State FIPS Code                     1
        HHS Region                          1
        Age Range                           1
        Benchmark                           1
        Locality                            1
        Observed Deaths                  8301
        Population                       4243
        Expected Deaths                  8301
        Potentially Excess Deaths        8301
        Percent Potentially Excess Deaths    8301
        dtype: int64
```

Output:5

```
']: Cancer                                33700
    Stroke                                33649
    Chronic Lower Respiratory Disease     33625
    Heart Disease                         33581
    Unintentional Injury                  30569
    Name: Cause of Death, dtype: int64
```

Output 6

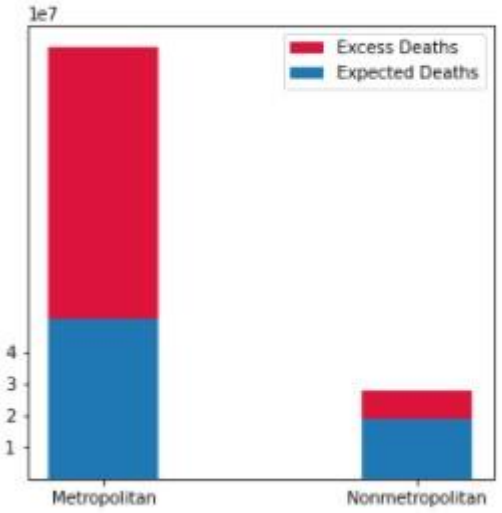


Output:7

```
Nonmetropolitan    55061
All                55034
Metropolitan       55029
Name: Locality, dtype: int64
```

Output 8

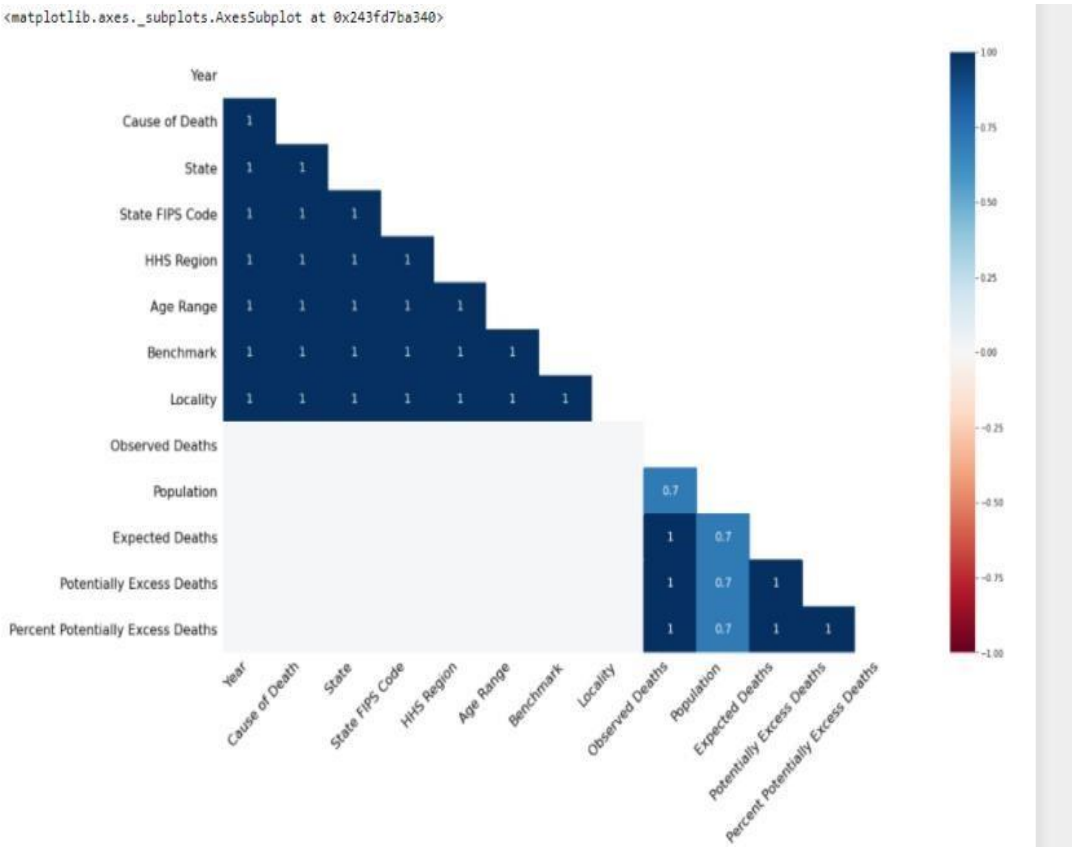
```
<matplotlib.legend.Legend at 0x243fcbd5dc0>
```



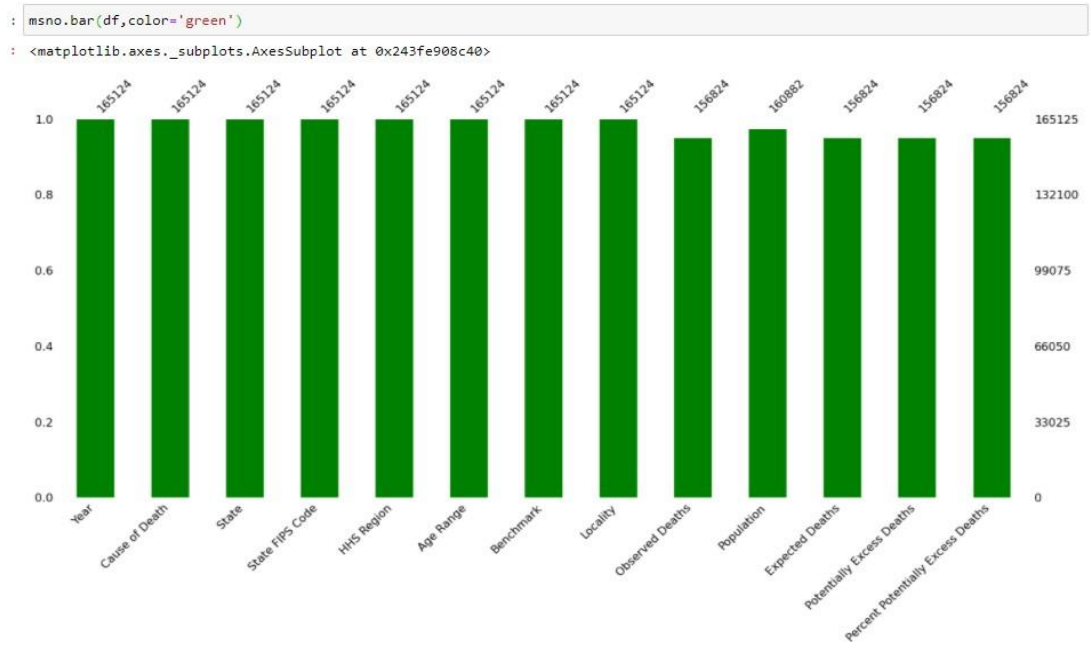
Output 9

```
Alabama                     3241
Alaska                      3241
Arizona                     3241
Arkansas                    3241
Colorado                    3240
California                  3240
Connecticut                 3238
Delaware                    3202
District of\nColumbia       3184
Rhode Island                3169
New Jersey                  3169
Nebraska                    3168
South Carolina              3168
Indiana                     3168
Oklahoma                    3168
Texas                       3168
Michigan                    3168
Tennessee                   3168
Ohio                        3168
Utah                        3168
Louisiana                   3168
Pennsylvania                3168
Mississippi                 3168
Virginia                    3168
Illinois                    3168
Kentucky                    3168
North Dakota                3168
Oregon                      3168
Minnesota                   3168
North Carolina              3168
New Hampshire               3168
South Dakota                3168
West Virginia               3168
Washington                  3168
New Mexico                  3168
New York                    3168
United States               3168
Vermont                     3168
Wisconsin                   3167
Maine                       3167
Missouri                    3164
Kansas                      3161
Massachusetts               3161
Georgia                     3160
Nevada                      3156
Florida                     3156
Hawaii                      3154
Wyoming                     3152
Iowa                        3152
Idaho                       3150
Maryland                    3144
Montana                     3138
Name: State, dtype: int64
```

Output 10

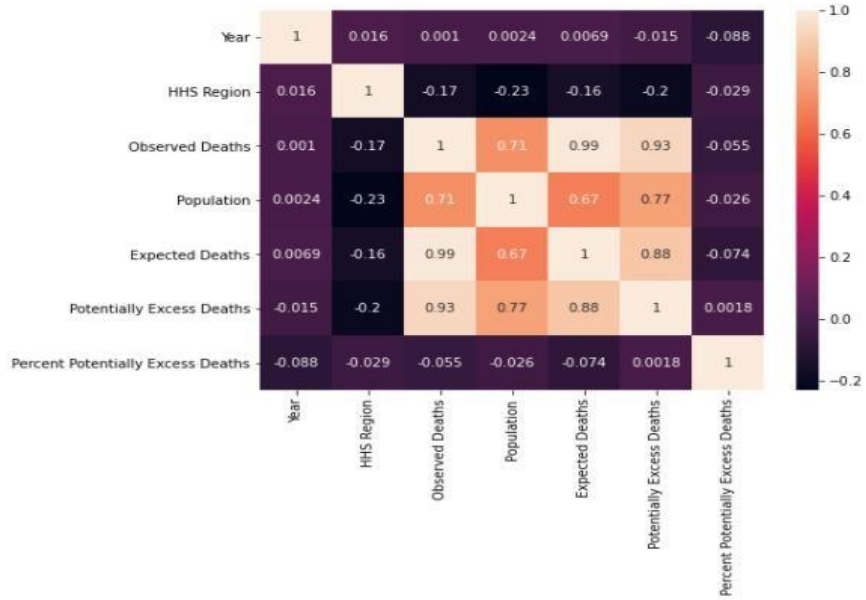


Output 11

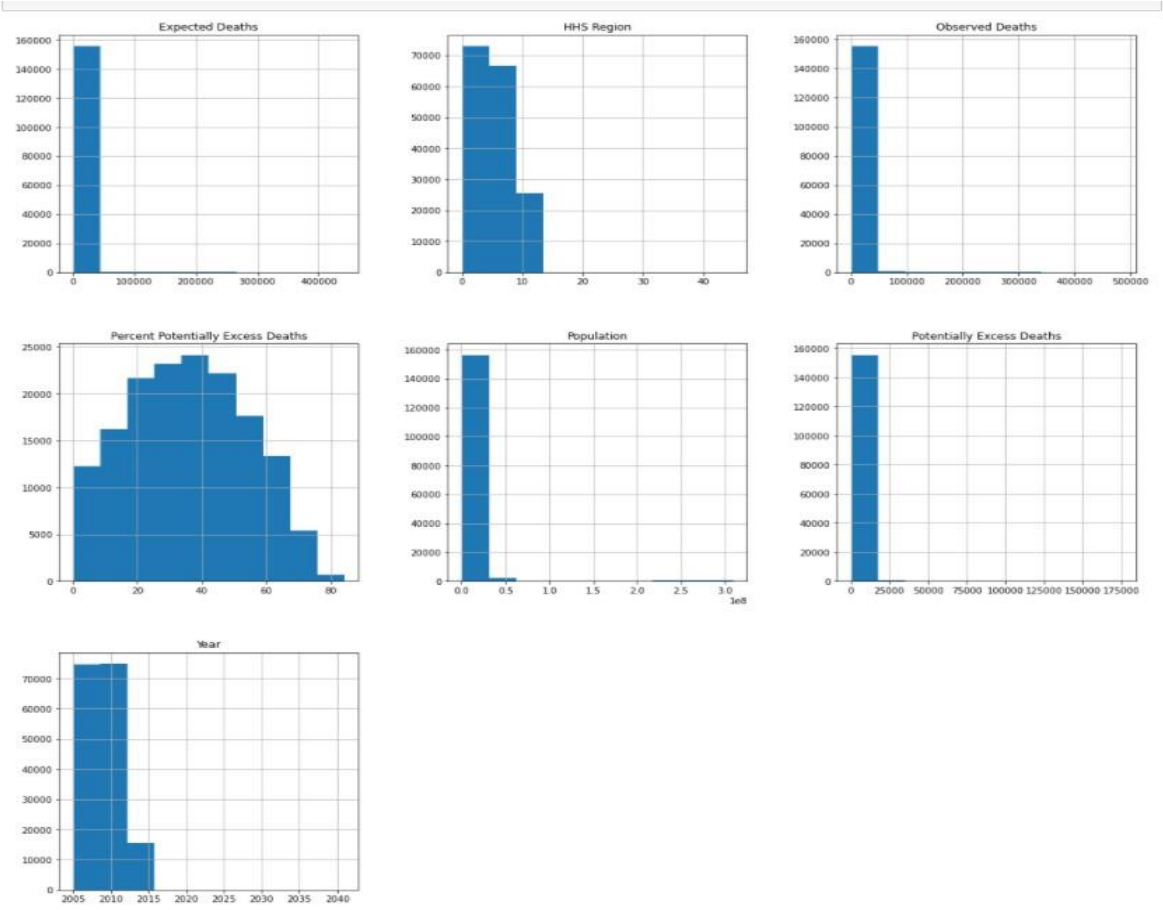


Output 12

```
<matplotlib.axes._subplots.AxesSubplot at 0x243fdd0ffa0>
```



Output 13



Tableau

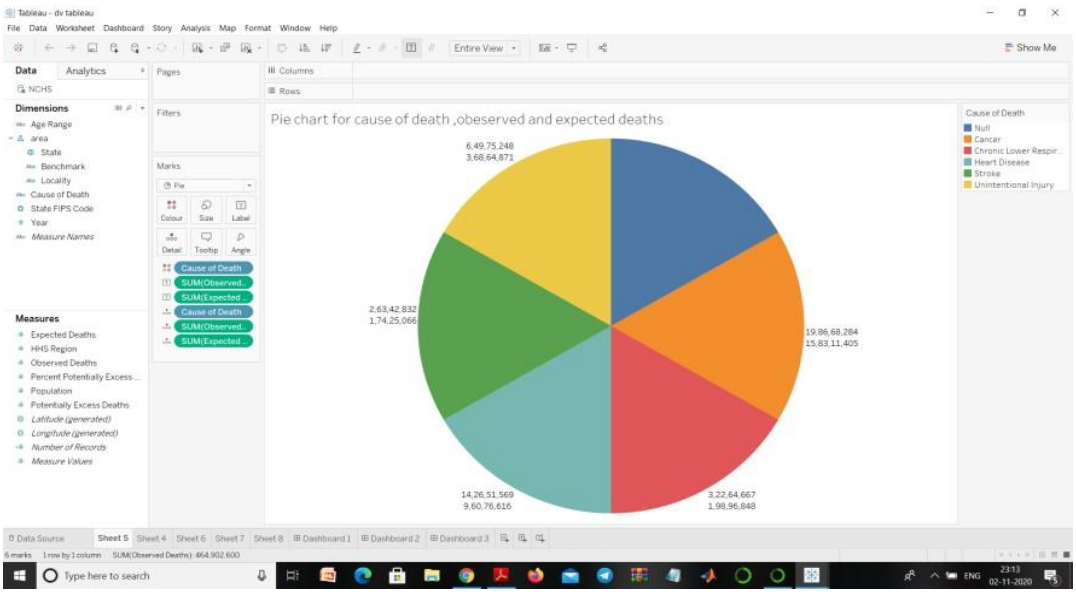


Tableau - dv tableau
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Pie chart for cause of death ,obeserved and expected deaths
6,49,75,248
3,68,64,871
19,86,68,284
15,83,11,405
3,22,64,667
1,98,96,848
14,26,51,569
9,60,76,616
2,63,42,832
1,74,25,066
Cause of Death
Null
Cancer
Chronic Lower Respir...
Heart Disease
Stroke
Unintentional Injury

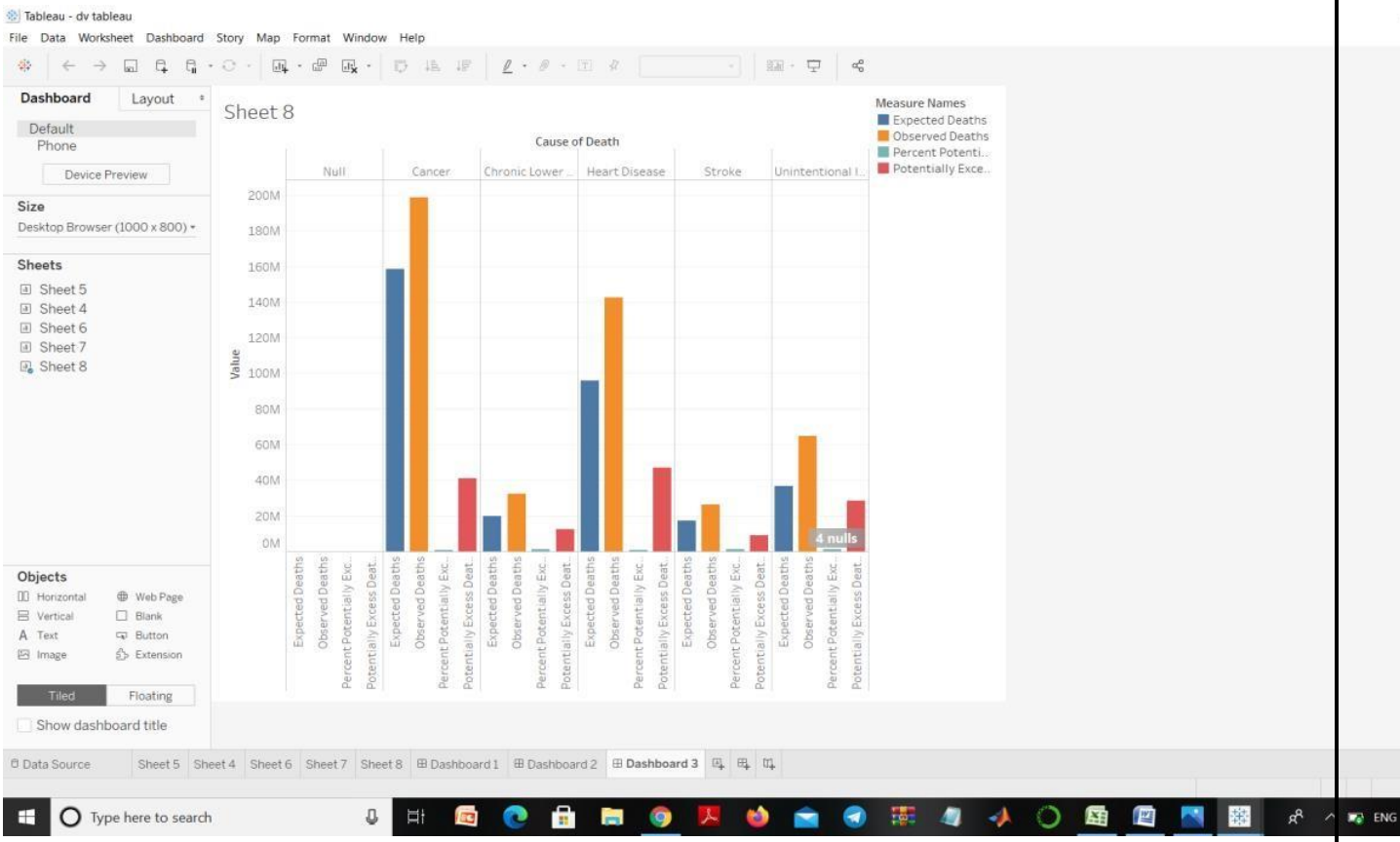


Tableau - dv tableau
File Data Worksheet Dashboard Story Map Format Window Help
Dashboard
Layout
Default
Phone
Device Preview
Size
Desktop Browser (1000 x 800)
Sheets
Sheet 5
Sheet 4
Sheet 6
Sheet 7
Sheet 8
Objects
Horizontal
Vertical
Text
Image
Web Page
Blank
Button
Extension
Tiled
Floating
Show dashboard title
Sheet 8
Cause of Death
Null
Cancer
Chronic Lower ...
Heart Disease
Stroke
Unintentional I...
Measure Names
Expected Deaths
Observed Deaths
Percent Potenti...
Potentially Exce...
Value
200M
180M
160M
140M
120M
100M
80M
60M
40M
20M
0M
4 nulls
Data Source
Sheet 5
Sheet 4
Sheet 6
Sheet 7
Sheet 8
Dashboard 1
Dashboard 2
Dashboard 3
Type here to search

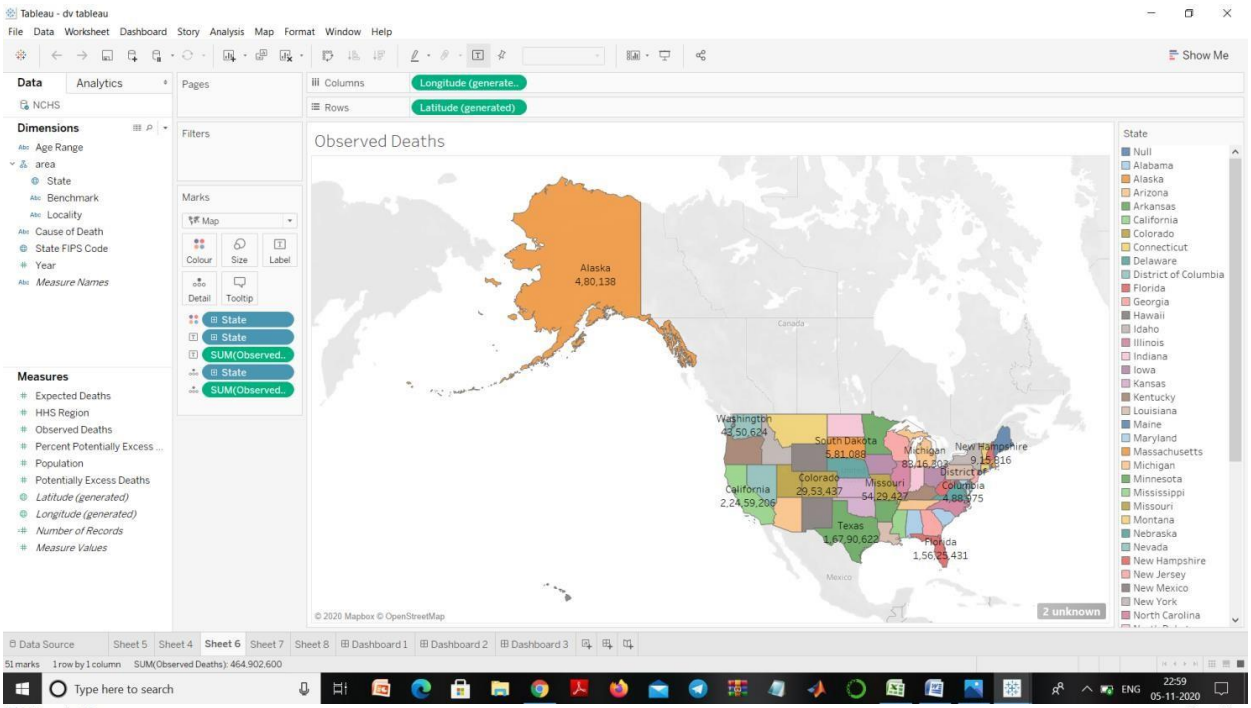


Tableau - dv tableau
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Show Me
Data Analytics
NCHS
Dimensions
Age Range
area
State
Benchmark
Locality
Cause of Death
State FIPS Code
Year
Measure Names
Measures
Expected Deaths
HHS Region
Observed Deaths
Percent Potentially Excess ...
Population
Potentially Excess Deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Map
Colour
Size
Label
Detail
Tooltip
State
State
SUM(Observed..
State
SUM(Observed..
Columns
Longitude (generate..
Rows
Latitude (generated)
Observed Deaths
Alaska
4,80,138
Canada
Washington
43,50,624
South Dakota
5,81,088
Michigan
80,16,384
New Hampshire
9,15,816
District of
Columbia
4,88,575
Colorado
29,53,437
Missouri
54,29,427
California
2,24,59,206
Texas
1,67,90,623
Florida
1,56,25,431
Mexico
© 2020 Mapbox © OpenStreetMap
2 unknown
State
Null
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Data Source
Sheet 5
Sheet 4
Sheet 6
Sheet 7
Sheet 8
Dashboard 1
Dashboard 2
Dashboard 3
51 marks
1 row by 1 column
SUM(Observed Deaths): 464,902,600
Type here to search
ENG
22:59
05-11-2020

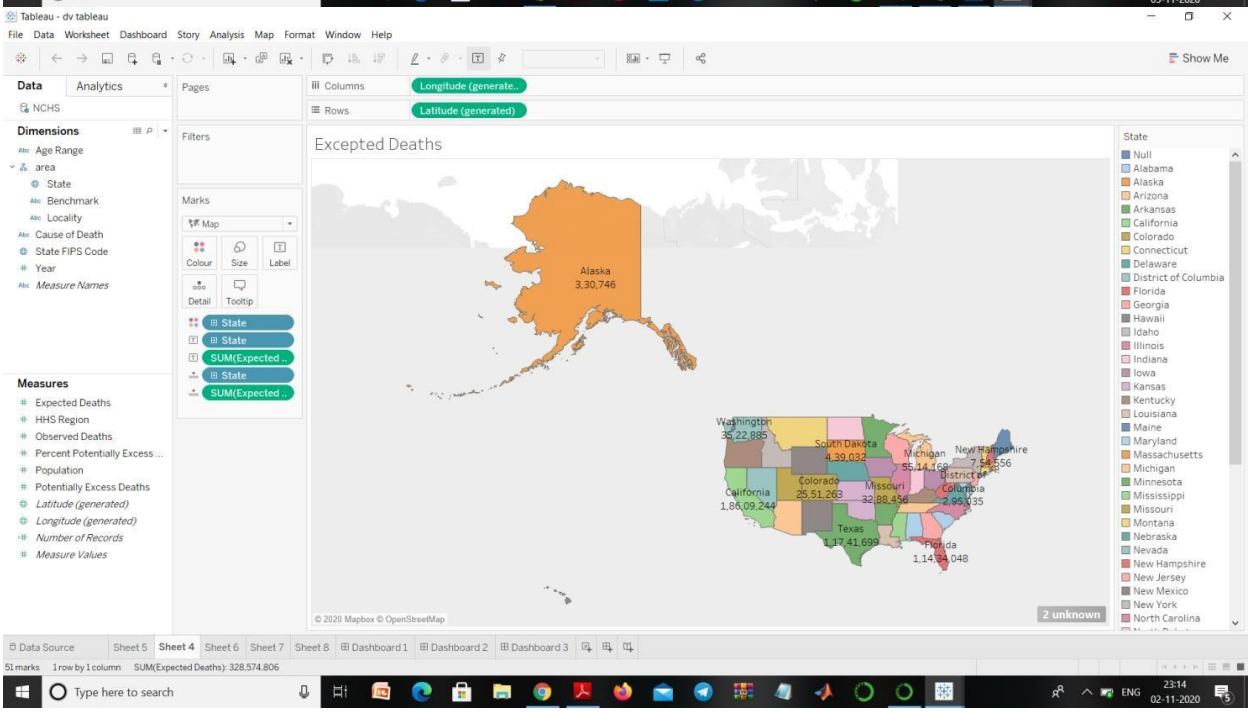


Tableau - dv tableau
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Show Me
Data Analytics
NCHS
Dimensions
Age Range
area
State
Benchmark
Locality
Cause of Death
State FIPS Code
Year
Measure Names
Measures
Expected Deaths
HHS Region
Observed Deaths
Percent Potentially Excess ...
Population
Potentially Excess Deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Map
Colour
Size
Label
Detail
Tooltip
State
State
SUM(Expected ..
State
SUM(Expected ..
Columns
Longitude (generate..
Rows
Latitude (generated)
Excepted Deaths
Alaska
3,30,746
Washington
35,22,885
South Dakota
4,39,032
Michigan
55,14,784
New Hampshire
7,54,556
District of
Columbia
2,95,035
Colorado
25,51,263
Missouri
32,88,456
California
1,86,09,244
Texas
1,17,41,693
Florida
1,14,34,048
© 2020 Mapbox © OpenStreetMap
2 unknown
State
Null
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
Data Source
Sheet 5
Sheet 4
Sheet 6
Sheet 7
Sheet 8
Dashboard 1
Dashboard 2
Dashboard 3
51 marks
1 row by 1 column
SUM(Expected Deaths): 328,574,806
Type here to search
ENG
23:14
02-11-2020

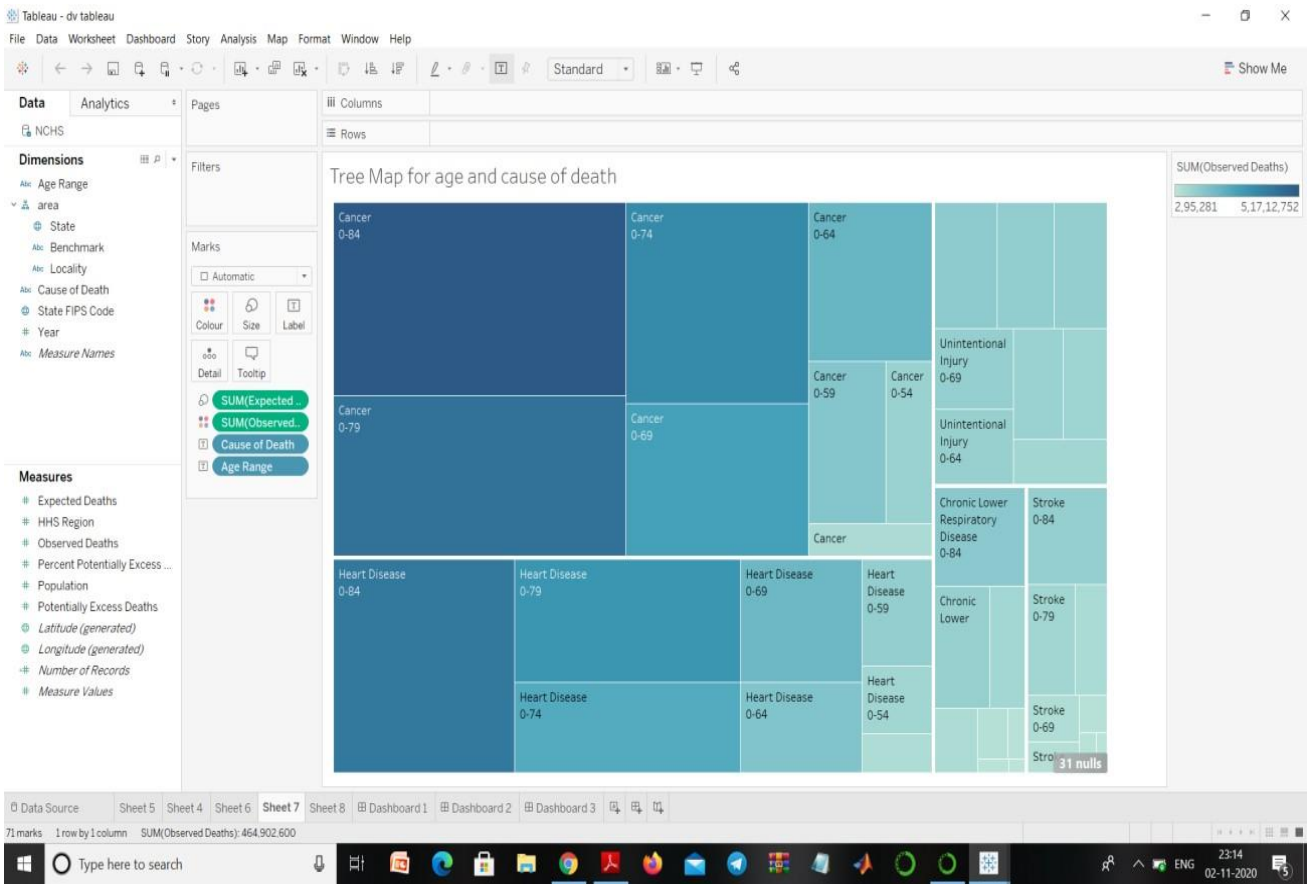


Tableau - dv tableau
File Data Worksheet Dashboard Story Analysis Map Format Window Help
Standard
Show Me
Data
Analytics
NCHS
Dimensions
Age Range
area
State
Benchmark
Locality
Cause of Death
State FIPS Code
Year
Measure Names
Measures
Expected Deaths
HHS Region
Observed Deaths
Percent Potentially Excess ...
Population
Potentially Excess Deaths
Latitude (generated)
Longitude (generated)
Number of Records
Measure Values
Pages
Filters
Marks
Automatic
Colour
Size
Label
Detail
Tooltip
SUM(Expected ..
SUM(Observed..
Cause of Death
Age Range
Columns
Rows
Tree Map for age and cause of death
Cancer 0-84
Cancer 0-74
Cancer 0-64
Cancer 0-79
Cancer 0-69
Cancer 0-59
Cancer 0-54
Cancer
Unintentional Injury 0-69
Unintentional Injury 0-64
Chronic Lower Respiratory Disease 0-84
Stroke 0-84
Chronic Lower
Stroke 0-79
Stroke 0-69
Heart Disease 0-84
Heart Disease 0-79
Heart Disease 0-69
Heart Disease 0-59
Heart Disease 0-74
Heart Disease 0-64
Heart Disease 0-54
31 nulls
SUM(Observed Deaths)
2,95,281
5,17,12,752
Data Source
Sheet 5
Sheet 4
Sheet 6
Sheet 7
Sheet 8
Dashboard 1
Dashboard 2
Dashboard 3
71 marks
1 row by 1 column
SUM(Observed Deaths): 464,902,600
Type here to search
ENG
23:14
02-11-2020

CONCLUSION:

Nonmetropolitan areas have higher age-adjusted death rates and greater percentages of potentially excess deaths from the five leading causes of death. Routine tracking of potentially excess deaths from the five leading causes of death in nonmetropolitan and metropolitan areas might help public health officials monitor important rural health disparities and select effective programs and policies to improve the health of residents of rural areas. Additional information on potentially excess deaths might be used to evaluate the success of public health interventions and to help determine where to allocate resources in areas with the greatest need. State and local public health officials in rural areas might seek advice from officials in rural areas with fewer potentially excess deaths for ways to reduce mortality in their jurisdictions or increase coordination with urban areas to ensure rural residents have timely access to specialized services REFERENCES:

https://www.cdc.gov/nchs/data-visualization/potentially-excess-deaths/index.htm

https://www.cdc.gov/mmwr/volumes/66/ss/ss6602a1.htm[xxi]