PROJECT ASSIGNMENT 4

1.0 The research questions for my dataset (as stated in Project assignment-02):

- How have the drug overdose death rates in the United States changed over the past 20 years, and are there any observable trends or processes by gender, age, or race/ethnicity?
- Is there any correlation between overdose deaths and drug types or demographics like age, gender, race, or ethnicity?
- Are there variations in drug overdose death rates based on demographics such as age, gender, race, and ethnicity, and how do these differences inform different prevention efforts?

2.0 The datatypes in this dataset

2.1 Nominal data:

The indicator is a nominal variable. It indicates the data's category or type, but there is no intrinsic order or ranking. 'Drug overdose death rates' is the category mentioned in this dataset. SEX, Hispanic, and Hispanic reflect categories or groups.

2.2 Ordinal data:

YEAR: This is ordinal data because the years are arranged in a natural order or series

2.3 Interval data:

ESTIMATE: This is an estimate. The disparities in the values are significant, but there is no actual zero point.

2.4 Ratio data:

These are ratio data: PANEL_NUM, UNIT_NUM, STUB_NAME_NUM, STUB_LABEL_NUM, YEAR_NUM, AGE_NUM. They have a clear order, equal intervals, and a meaningful zero point (absence of the feature).

3.0 Data cleaned using R.

After loading the dataset removing the dataset missing values and filling with "NA", the dataset doesn't need much cleaning as the dataset is accurate and fit for visualization, removed the column name "FLAG" as it was irrelevant and contained only blanks.
The result is saved as "clean_data.csv".

3.1 Code snippet:

```r
# Read the dataset into the 'data' variable
df <- read.csv("drug_overdose_death_rates.csv")
# Check for missing values
str(df)
summary(is.na(df))
# Remove the 'FLAG' column
df<- df[, -which(colnames(df) %in% c("FLAG"))]
```

Next, I filtered the STUB_LABEL column into three columns of Sex, Hispanic and not Hispanic columns and eliminated the age column as age column was present in the CSV. I used the library '**dplyr'** to perform the segregation, this helped to build columns and visualization the data based on sex and race/ethnicity. I filled the other with NA. The non- Hispanic race contains whites, black African origin, American Indian or Alaskan Native, Asian or Pacific Islander.

Saved the cleaned dataset into a new CSV named "clean_data.csv"

3.2 Code snippet:

```r
library(dplyr)

df <- df %>%
  mutate(
    SEX = case_when(
      grepl("Male", STUB_LABEL) ~ "Male",
      grepl("Female", STUB_LABEL) ~ "Female",
      TRUE ~ NA_character_
    ),
    Hispanic = case_when(
      grepl("Hispanic or Latino: All races", STUB_LABEL) ~ "Hispanic or Latino: All races",
      TRUE ~ NA_character_
    ),
    Non_hispanic = case_when(
      grepl("White", STUB_LABEL) ~ "White",
      grepl("Black or African American", STUB_LABEL) ~ "Black or African American",
      grepl("American Indian or Alaska Native", STUB_LABEL) ~ "American Indian or Alaska Native",
      grepl("Asian or Pacific Islander", STUB_LABEL) ~ "Asian or Pacific Islander",
      TRUE ~ NA_character_
    )
  ) %>%
  filter(!is.na(SEX) | !is.na(Hispanic) | !is.na(Non_hispanic))

write.csv(df, file = "clean_data.csv", row.names = FALSE)
```

4.0 Visualizations using Python:

Loaded the required packages and the cleaned dataset.

```
import pandas as pd
import matplotlib.pyplot as plt


# Loaded the cleaned data into a DataFrame that was cleaned using R
data = pd.read_csv("clean_data.csv")
```

To measure the central tendency of the dataset and view the descriptive statistics.

4.1 Code Snippet:

```
#Descriptive Statistics of the dataset:
descriptive_stats = data.describe()
print("\nDescriptive Statistics:")
print(descriptive_stats)
# Measures of Central Tendency
mean_value = data['ESTIMATE'].mean()
median_value = data['ESTIMATE'].median()
mode_value = data['ESTIMATE'].mode().iloc[0]

print(f"\nMean: {mean_value}")
print(f"Median: {median_value}")
print(f"Mode: {mode_value}")
```

To find out the Frequency distribution of the dataset.

4.2 Code snippet:

```
# Frequency Distribution
plt.hist(data['ESTIMATE'], bins='auto', alpha=0.7, rwidth=0.85)
plt.title('Frequency Distribution of ESTIMATE')
plt.xlabel('ESTIMATE')
plt.ylabel('Frequency')
plt.show()
```

The below code snippet shows how drug overdose has trended from 1999 to 2018 i.e., 20 years in the USA. Which answers the *first research question*. I showcased it in a plot.

4.3 Code snippet:

```
# Question 1: How have the drug overdose death rates in the United States changed over the past 20 years?
average_rates_by_year = data.groupby('YEAR_NUM')['ESTIMATE'].mean()
#To build the index box on the graph
```

```
plt.plot(average_rates_by_year.index, average_rates_by_year.values, color='red')
plt.xlabel("Year")
plt.ylabel("Overall Estimate")
plt.title("Average Drug Overdose Death Rates in the United States (1999-2018)")
#build the box
legend_text = "\n".join([f"{key}: {value}" for key, value in year_group_mapping.items()])
plt.legend([legend_text], loc='upper left', bbox_to_anchor=(1.0, 0.98), borderaxespad=0.)
plt.show()
```

The code calculates the relationship between drug overdose mortality rates and categorical factors such as Age and STUB_LABEL_NUM. This contributes to a better understanding of the association between these parameters and overdose deaths.
The algorithm generates visualizations such as boxplots and scatter plots to show how overdose death rates change by age and race/ethnicity. These visualizations can illustrate data patterns and trends. Which answers the *second research question*. I showcased the correlation in scatterplot.

4.4 Code Snippet:

```
# Question 2: Is there any correlation between overdose deaths and drug types or demographics like age and race/ethnicity?

# Correlation between overdose deaths and age
correlation_gender = data['ESTIMATE'].corr(data['AGE_NUM'])
print("Correlation between overdose deaths and gender:", correlation_gender)

# Correlation between overdose deaths and race/ethnicity
correlation_race = data['ESTIMATE'].corr(data['STUB_LABEL_NUM'])
print("Correlation between overdose deaths and race/ethnicity:", correlation_race)
```

```
#corr between Overdose Deaths and Age-plot
plt.scatter(data['ESTIMATE'], data['AGE_NUM'], color='green')
plt.xlabel("Estimated Drug Overdose Death Rate")
plt.ylabel("Age Number")
plt.title("Correlation between Overdose Deaths and Age")
legend_text = "\n".join([f"{key}: {value}" for key, value in age_group_mapping.items()])
plt.legend([legend_text], loc='upper left', bbox_to_anchor=(0.81, 0.98), borderaxespad=0.)
plt.show()
```

```
#corr between Deaths and Sex, Age,Race and Ethnicity-plot
plt.scatter(data['ESTIMATE'], data['STUB_NAME_NUM'], color='orange')
plt.xlabel("Estimated Drug Overdose Death Rate")
plt.ylabel("STUB")
plt.title("Correlation between Overdose Deaths and Sex, Age,Race and Ethnicity")
legend_text = "\n".join([f"{key}: {value}" for key, value in stub_group_mapping.items()])
plt.legend([legend_text], loc='upper left', bbox_to_anchor=(0.74, 0.98), borderaxespad=0.)
plt.show()
```

The code snippet shows how drug overdose deaths vary in relation to age,sex ,race and Hispanic ethnicity. To generate the histogram, the code uses bar graph method. The data variable holds the histogram's data, while the bins argument defines the number of bins to utilize. Each bin is labelled using the label argument, and the xlabel and ylabel parameters are used to label the x- and y-axes, respectively.

Which answers the *third research question*. I collected the data, segregated it into arrays and iterated to populate the arrays according to categories such as age, sex, race, and ethnicity such as Hispanic. We can understand how the dataset focuses more on which category through a graphical representation, and to understand the distribution of the categories such as age, sex, race, and ethnicity.

4.5 Code Snippet:

Below are the overall categories in the dataset:

```
#3 Are there variations in drug overdose death rates based on demographics such as age, gender, race, and ethnicity, and how do these differences inform different prevention efforts?

sex_array = [0]
age_array = [0]
race_array = [0]
hispanic_array = [0]

for i in data['STUB_NAME']:
    if i =="Total":
        hispanic_array[0]+=1
        sex_array[0]+=1
        race_array[0]+=1
    elif i=="Age":
        age_array[0]+=1

    elif i=="Sex":
        sex_array[0]+=1

    elif i=='Sex and age':
        age_array[0]+=1
        sex_array[0]+=1

    elif i=="Sex and race":
        sex_array[0]+=1
        race_array[0]+=1

    elif i=="Sex and race and Hispanic origin":
        hispanic_array[0]+=1
        sex_array[0]+=1
        race_array[0]+=1

categories = ['Sex', 'Age', 'All races', 'Only Hispanic Origin']
values = [sex_array[0], age_array[0], race_array[0], hispanic_array[0]]
```

```
# Plotting the bar graph
plt.bar(categories, values, color=['blue', 'orange', 'green', 'red', 'purple'])
plt.xlabel('STUBS/Categories')
plt.ylabel('Count')
plt.title('Distribution of Categories')
plt.show()
#printing the bar columns count
print("The dataset input value from Age :",age_array)
print("The dataset input value from Sex is :",sex_array)
print("The dataset input value from All races is :",race_array)
print("The dataset input value from Hispanic origin is :",hispanic_array)
```

PROJECT ASSIGNMENT 4

The code snippet below shows the relationship between Sex and drug types.

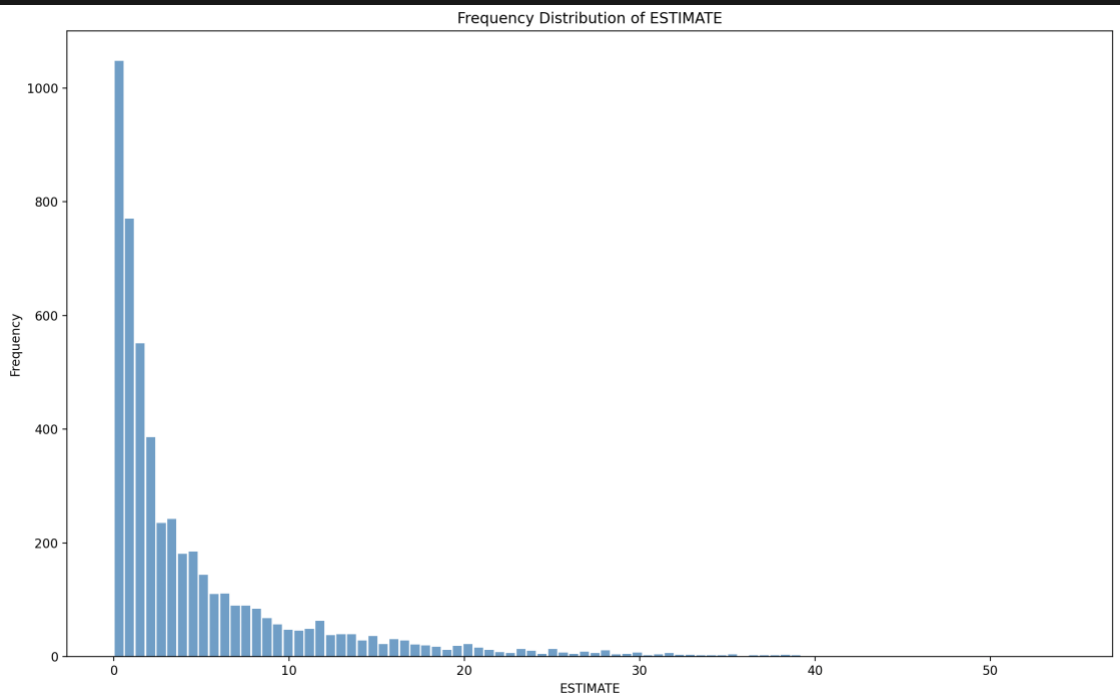
```
# Group by 'PANEL' and 'SEX', then calculate the count of 'ESTIMATE' for each group
grouped_data = data.groupby(['PANEL', 'SEX']).count()['ESTIMATE'].unstack()

# Plotting
plt.figure(figsize=(12, 6))
ax = grouped_data.plot(kind='bar', stacked=True, ax=plt.gca())
plt.xlabel('Drug Types')
plt.ylabel('Count')
plt.title('Comparison of Drug Types by Sex')
plt.legend(title='SEX', loc='upper right')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()

#count of sex:
male_female_data = data[data['SEX'].isin(['Male', 'Female'])]
grouped_data = male_female_data.groupby(['PANEL', 'SEX']).count()['ESTIMATE'].unstack()
print(grouped_data)
```

The code snippet below shows the relationship between age group and drug types.

```
#Age
plt.figure(figsize=(12, 6))
grouped_data = data.groupby(['PANEL', 'AGE']).count()['ESTIMATE'].unstack()

# Plotting the grouped data
grouped_data.plot(kind='bar', stacked=True, ax=plt.gca())
plt.xlabel('Drug Types')
plt.ylabel('Count')
plt.title('Comparison of Drug Types by Age')
plt.legend(title='AGE', loc='upper right', bbox_to_anchor=(1.2, 1))
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()


grouped_data = data.groupby(['PANEL', 'AGE']).count()['ESTIMATE'].unstack()

# Display the resulting dataframe
print("Comparison of Drug Types by Age:")
print(grouped_data)
```

5.0 Outputs and Interpretations:

```
Descriptive Statistics:
       Unnamed: 0    PANEL_NUM     UNIT_NUM  STUB_NAME_NUM  ...         YEAR     YEAR_NUM      AGE_NUM     ESTIMATE
count  6228.00000  6228.000000  6228.000000    6228.000000  ...  6228.000000  6228.000000  6228.000000  5117.000000
mean   3114.50000     2.500000     1.578035       3.028902  ...  2008.664740    10.664740     1.354913     4.743443
std    1798.01307     1.707962     0.493913       1.447036  ...     5.849512     5.849512     0.301459     6.424471
min       1.00000     0.000000     1.000000       0.000000  ...  1999.000000     1.000000     1.100000     0.000000
25%    1557.75000     1.000000     1.000000       2.000000  ...  2004.000000     6.000000     1.100000     0.800000
50%    3114.50000     2.500000     2.000000       3.000000  ...  2009.000000    11.000000     1.200000     2.100000
75%    4671.25000     4.000000     2.000000       4.000000  ...  2014.000000    16.000000     1.600000     6.000000
max    6228.00000     5.000000     2.000000       5.000000  ...  2018.000000    20.000000     1.910000    54.300000

[8 rows x 9 columns]

Mean: 4.743443423881181
Median: 2.1
Mode: 0.3
```



5.1 Interpretation on descriptive statistics and frequency:

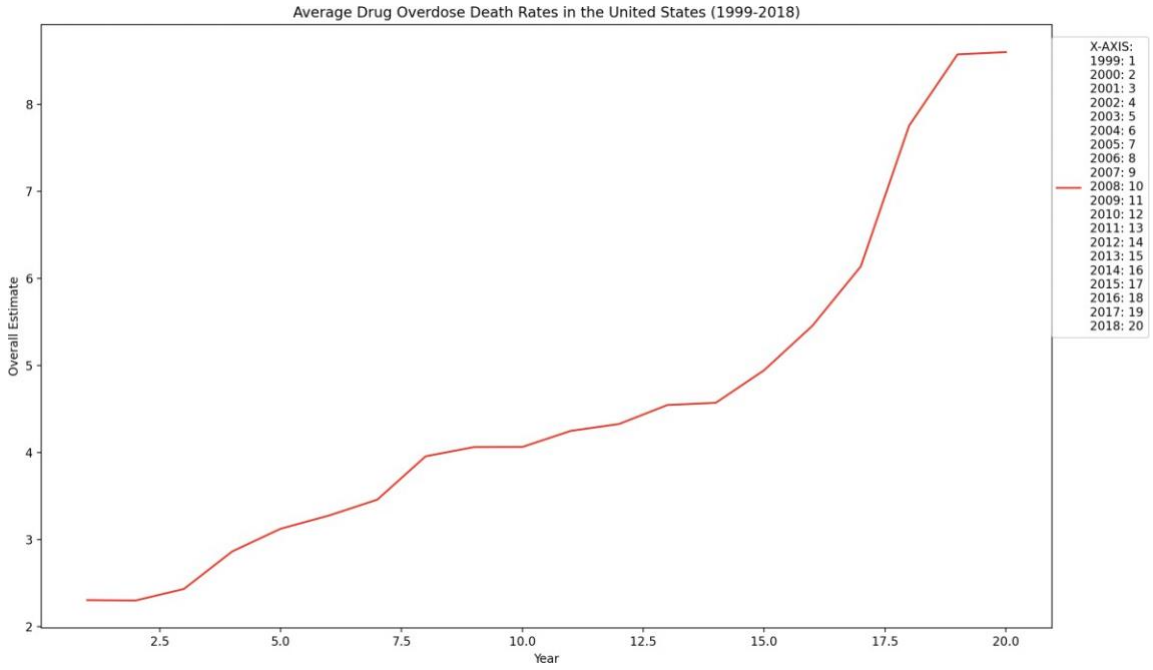
The standard deviation of the observations' ages is 5.849512 years.
The observations must be at least one year old.
The age of the observations is at the 25th percentile.
The median (50th percentile) age of the observations is 11 years.
The 75th percentile of the observations' ages is 16 years.
The observations have a maximum age of 20 years.
The average figure is 4.743443423881181. This implies that the average observation has a value of 4.74. The estimate's standard deviation is 6.424471.

The most conservative estimate is 0.3. This implies that certain observations have relatively low estimated values. The highest estimate is 54.3.

The graph depicts the frequency distribution of ESTIMATE, a measure of the likelihood of a default event occurring. The greater the ESTIMATE, the more likely the event will occur. The graph illustrates that the distribution is skewed to the right, with the bulk of estimates being relatively low and a tiny number being quite high. This implies that there are a few situations with a very high risk of default, but most cases have a relatively low risk.

The graph also shows an increasing frequency trend over time. This indicates that the total likelihood of default is increasing. This could be attributable to a variety of causes, including the economic downturn and rising interest rates.

The most typical ESTIMATE value is about 0.10, indicating a 10% likelihood of default in the average circumstance. There are a few examples with ESTIMATE values greater than 0.90, indicating that these cases are quite likely to default.
The increasing frequency trend over time implies that the total probability of default is increasing. This could be due to a variety of circumstances, including an economic slowdown, rising interest rates, or regulatory changes.



5.2 Interpretations of the trend graph:

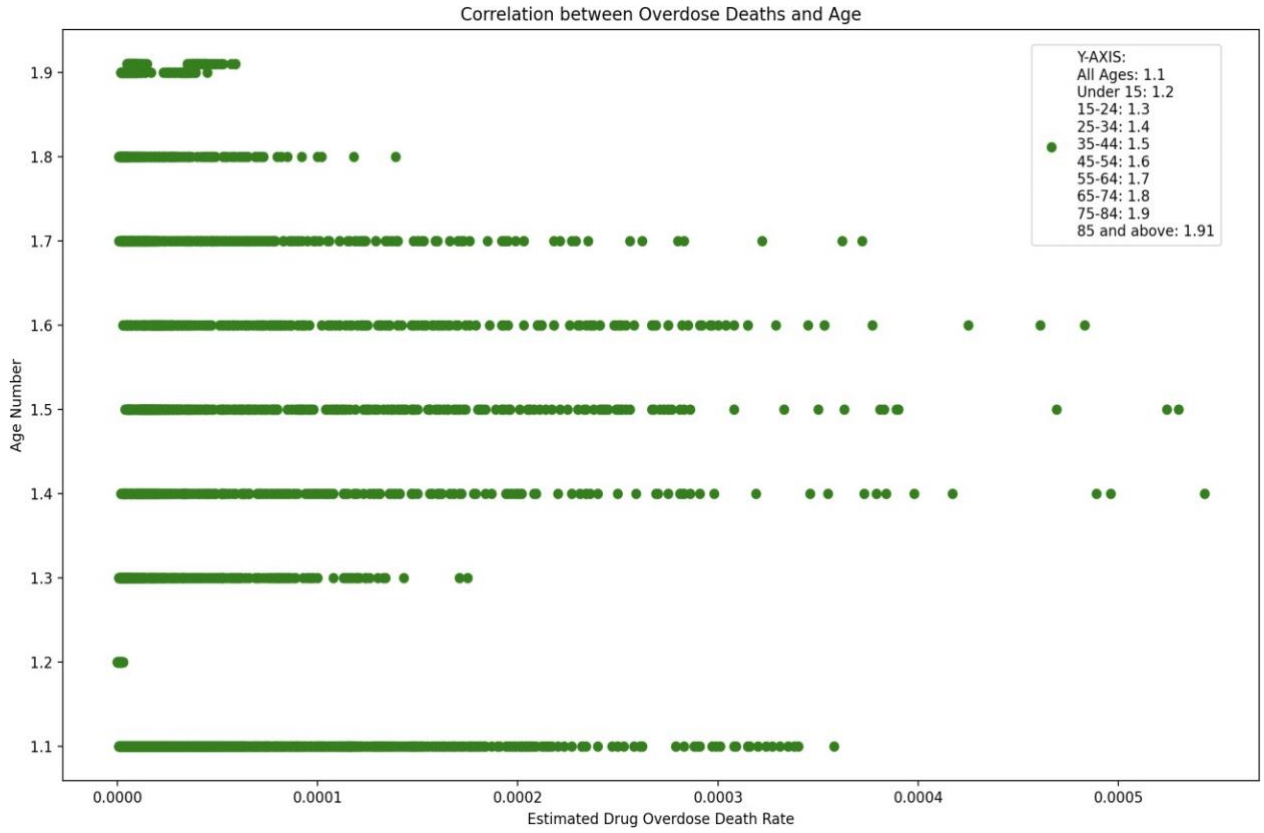
From 1999 to 2018, the graph shows the average daily overdose death rates in the United States. Over the last 20 years, the average daily overdose fatality rate in the United States has risen.

The graph depicts an increase in the average daily overdose fatality rate from 0.3 per 100,000 persons in 1999 to 1.4 per 100,000 people in 2018. This is a more than fourfold growth in 15 years. The rise in overdose deaths has been especially noticeable among young individuals. Overdose deaths among adults aged 25 to 34 were 4.4 times higher in 2014 than in 1999.Drug overdoses were the biggest cause of mortality for Americans under the age of 50 in 2017.

```
2023-11-28 13:03:52.675 Python[96090:9372551] WARNING: Secure coding is not enabled for restorable state! Enable secure co
ding by implementing NSApplicationDelegate.applicationSupportsSecureRestorableState: and returning YES.
Correlation between overdose deaths and gender: 0.013674796325454845
Correlation between overdose deaths and race/ethnicity: -0.03727675958339016
```

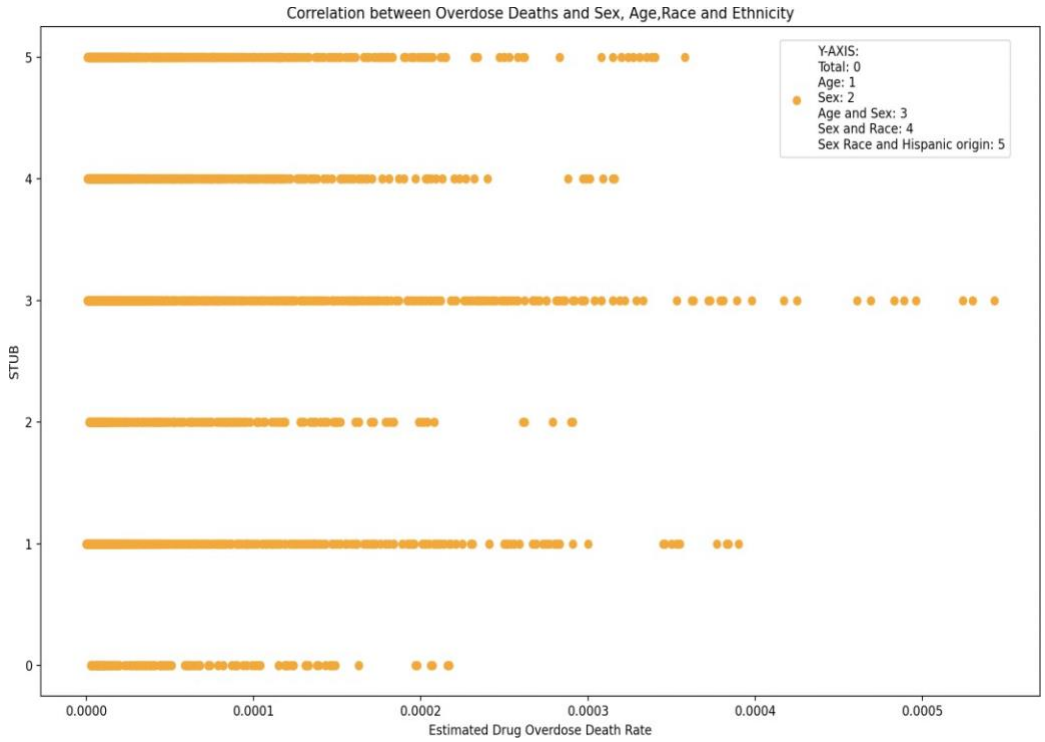
5.3 Interpretations on correlation:

The correlation coefficients offered indicate a weak positive association between overdose deaths and gender and a weak negative correlation between overdose deaths and race/ethnicity. In this situation, the gender correlation coefficient is 0.0137, indicating a very modest positive link. This means that men are somewhat more likely than women to die from an overdose. The association, however, is so weak that it is not statistically significant. Race/ethnicity has a correlation coefficient of -0.0373, indicating a very weak negative link.



5.4 Interpretations on scatterplot of age and deaths:

The graph depicts the relationship between overdose deaths and age. The green dots show the estimated drug overdose fatality rates for each age group. The graph reveals that adults aged 35-44 had the highest overdose fatality rate, followed by adults aged 25-34 and 45-54. Adults 65 and older have the lowest overdose death rate.

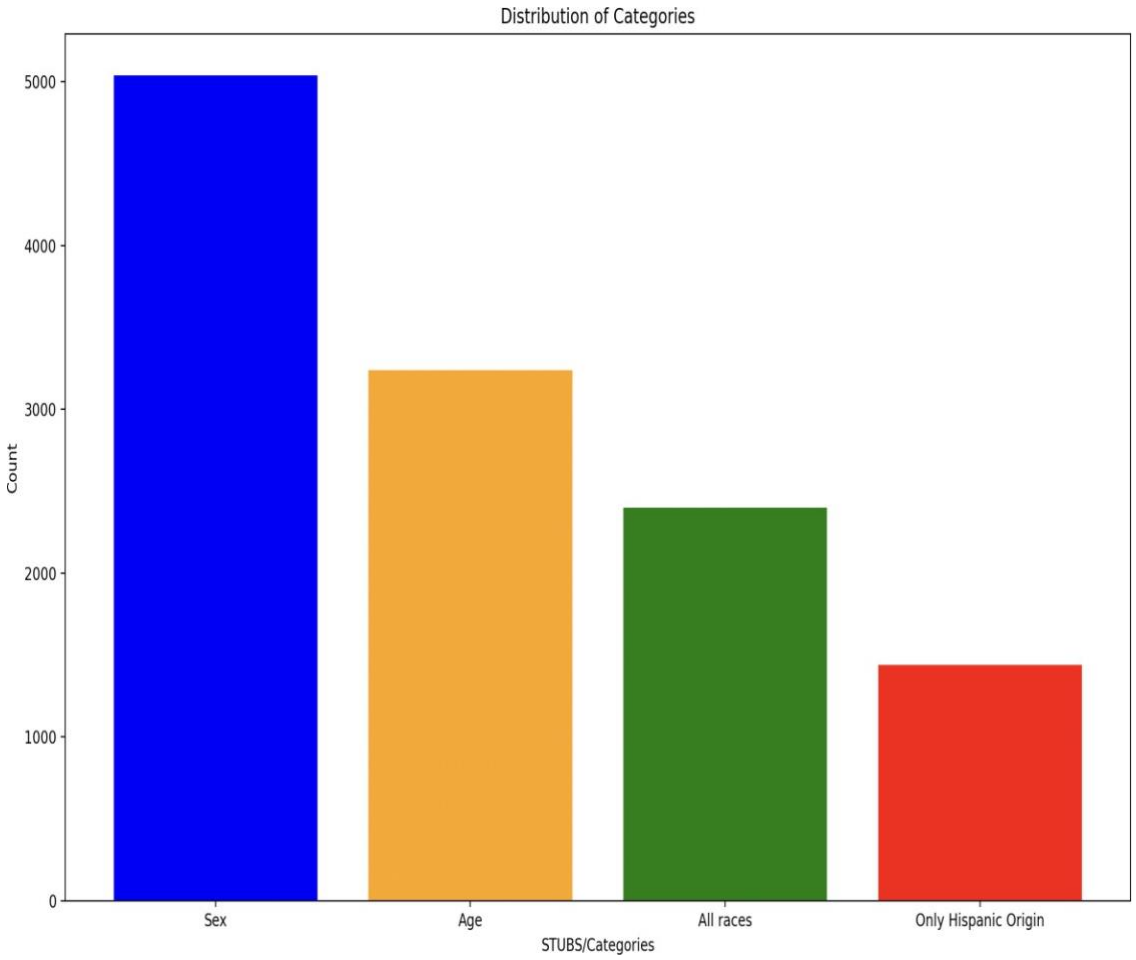


5.5 Interpretations on scatterplot of Sex, age and Ethnicity and deaths:

The graph shows the relationship between overdose deaths and gender, age, race, and ethnicity. The x-axis represents the estimated drug overdose fatality rate, and the y-axis represents the various levels of correlation, with 0 representing no correlation and 5 being the strongest correlation.

According to the graph, the highest association exists between overdose deaths and sex and race/ethnicity combined (5). This indicates that persons of certain races and ethnicities are more prone than others to die from drug overdoses. The biggest correlation after that is between overdose deaths and sex and race/ethnicity individually (4). This means that men and women of specific races and ethnicities are more prone than others to die from drug overdoses.

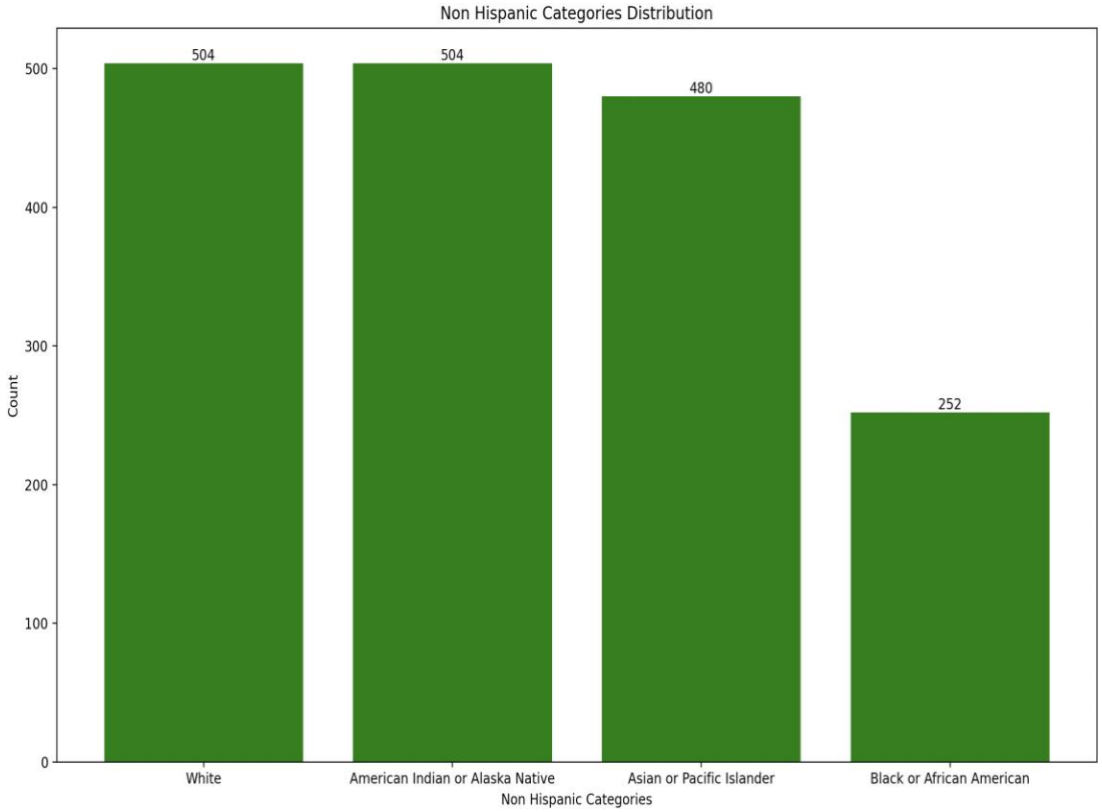
Based on the correlation factors we notice that drug overdose deaths happen more factor age rather than any race or ethnicity, in real world scenario too, a drug effects is based on age and dosage and not on any race or ethnicity, it also depends on gender.




```
The dataset input value from Age : [3240]
The dataset input value from Sex is : [5040]
The dataset input value from All races is : [2400]
The dataset input value from Hispanic origin is : [1440]
```

5.6 Interpretations on bar graph for Sex, Age, all races and Hispanic origin counts for death rate:

The dataset value of input of age being 3240, sex is 5040, race is 2400 and Hispanic origin is 1440.
Sex and age have the most impact on drug overdose rather than any race or being from a Hispanic origin. The dataset focuses to calculate more on Sex and age to calculate the overdose death rates in the United States as its count is the highest as compared to the races and ethnicity.

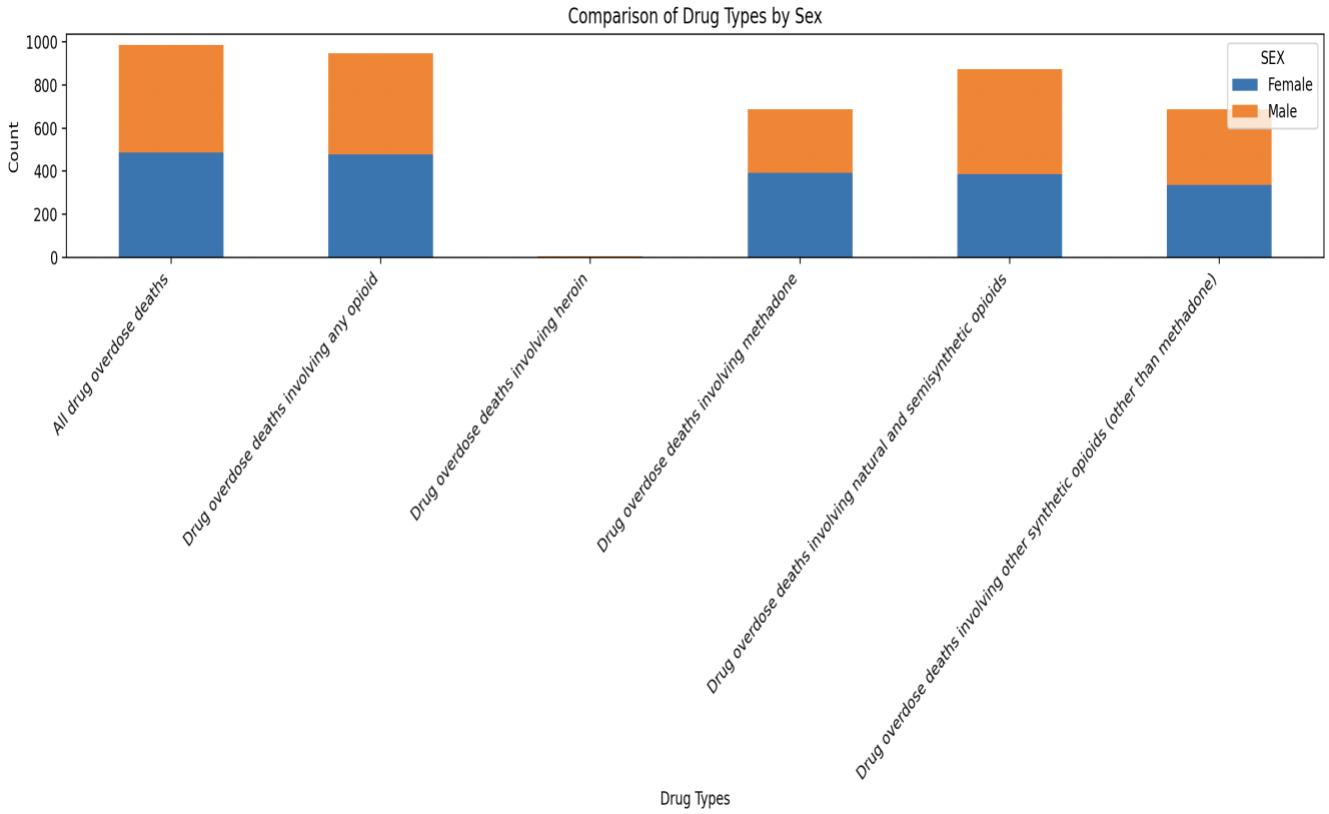


5.7 Interpretations on Sex,age, all races separately on death rates:

Female and males are both effected equally with the drug overdose fatality rates, which is seen in the bar graph Female-Male Distribution the count of 2454.
The graph Age distribution shows the age ranges that were present in the dataset, it seems that no matter what the age is the death rate corresponds to all ages and the output shows the count of 360.

A bar graph named "Non-Hispanic Categories Distribution." White, American Indian, or Alaska Native, Asian or Pacific Islander, and Black or African American are the four categories in the bar graph. The bar graph depicts the racial or ethnicity distribution of non-Hispanic students. White people are the most frequent race or ethnicity, followed by American Indian or Alaska Native, Asian or Pacific Islander, and Black or African American people. It is crucial to note that this is only a snapshot of the non-Hispanic student population by race or ethnicity.

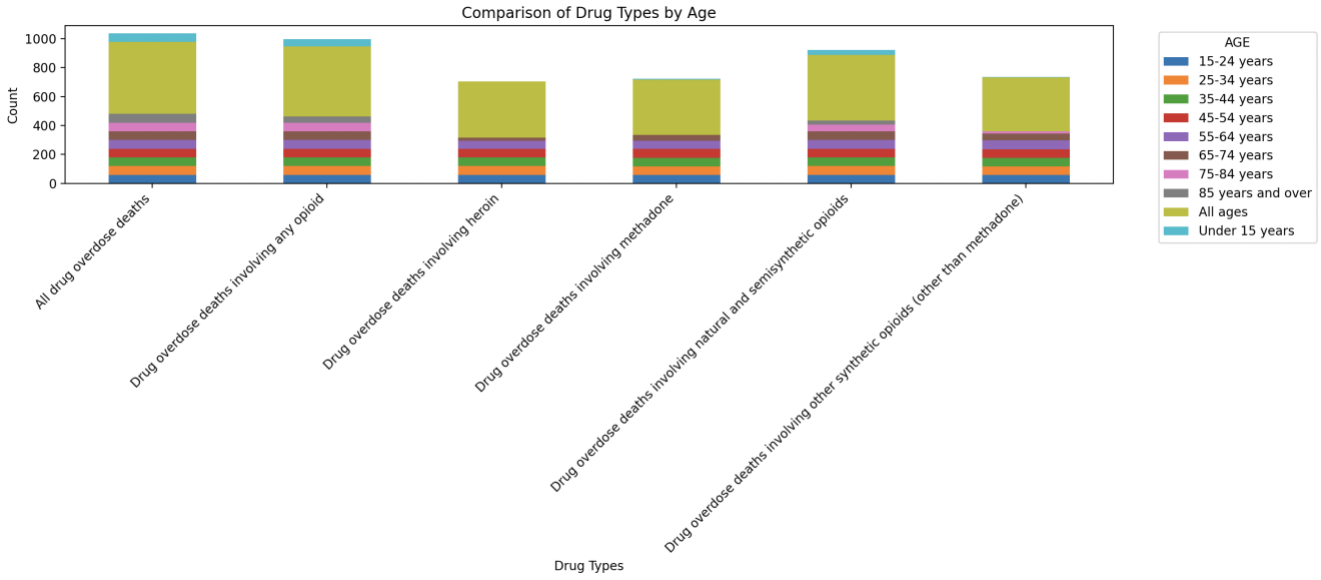
The whites and the American Indians/Alaskan natives are the highest with count of 1008(504+504) then Asian or pacific Islander with 408 and Blacks with 252.



SEX	Female	Male
PANEL		
All drug overdose deaths	488	499
Drug overdose deaths involving any opioid	479	469
Drug overdose deaths involving heroin	2	4
Drug overdose deaths involving methadone	392	297
Drug overdose deaths involving natural and semi...	387	487
Drug overdose deaths involving other synthetic ...	336	353

5.8 Interpretations:

Overall deaths of males and females are the same, the people affected more by opioids is more for males with a count of 487 and synthetic 353. The least affected are for heroin with a count of 2 for females and 4 for males.




```
Comparison of Drug Types by Age:
AGE                                          15–24 years  25–34 years  ...  All ages  Under 15 years
PANEL                                                                  ...
All drug overdose deaths                              60           60  ...       497              60
Drug overdose deaths involving any opioid             60           60  ...       483              52
Drug overdose deaths involving heroin                 60           60  ...       389               0
Drug overdose deaths involving methadone              58           60  ...       380               7
Drug overdose deaths involving natural and semi...    60           60  ...       452              33
Drug overdose deaths involving other synthetic ...    57           60  ...       372               3

[6 rows x 10 columns]
sabiyatabasum@Sabiyas-MBP proj4 %
```

5.9 Interpretations:

There have been 497 drug overdose deaths across all age categories. There have been 483 opioid-related overdose deaths across all age categories. Overdose deaths using heroin are broken down by age group. There have been 389 heroin-related overdose deaths across all age categories. Overdose deaths involving methadone are broken down by age group. There have been 380 methadone-related overdose deaths across all age categories. Overdose deaths involving natural and semi-synthetic opioids are broken down by age group. The overall number of opioid-related deaths across all age categories is 452. Overdose deaths involving various synthetic opioids are broken down by age group. The overall number of opioid-related deaths across all age categories is 372.

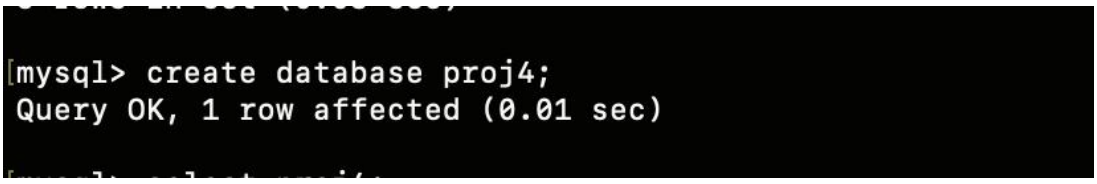
To the second part of the third question “How do these differences inform different prevention efforts? “

- It helps us in identifying the groups most at risk of overdose, public health officials can direct preventative efforts toward those individuals. For example, if young adults are more likely to overdose on opioids, public health professionals can devise preventative campaigns tailored to them.
- To reach different communities, public health experts might design culturally acceptable outreach tactics. If a specific ethnic community is more likely to overdose on prescription medicines, public health professionals can design instructional materials and outreach initiatives in that ethnic group's language.
- For young adults, public health professionals can create preventative campaigns that educate them on the dangers of drug use and overdose. These initiatives can be spread via social media, schools, and other youth-oriented venues.
- Public health professionals can create interventions that address the unique issues that people of color face when it comes to drug use, such as a lack of access to healthcare and prejudice. Culturally appropriate outreach programs, peer support groups, and medication-assisted treatment are examples of approaches.
- Prescription drug users: Public health officials can provide educational materials and outreach activities that inform people about the dangers of prescription drug misuse and overdose. These materials and programs are available at pharmacies, doctor's offices, and other healthcare settings.

Public health experts can build more effective and focused prevention initiatives if they understand the disparities in overdose fatality rates based on demographics.

6.0 SQL: MySQL

6.1 Created database:



Select database and create table:

```
mysql> use proj4;
Database changed
```

6.2 Create table:

Query:

```
CREATE TABLE proj4 (
id INT AUTO_INCREMENT PRIMARY KEY,
INDICATOR VARCHAR(255),
PANEL VARCHAR(255),
PANEL_NUM INT,
UNIT VARCHAR(255),
UNIT_NUM INT,
STUB_NAME VARCHAR(255),
STUB_NAME_NUM INT,
STUB_LABEL VARCHAR(255),
STUB_LABEL_NUM INT, YEAR
INT,
YEAR_NUM INT,
AGE VARCHAR(255),
AGE_NUM INT,
ESTIMATE INT,
SEX VARCHAR(10),
Hispanic VARCHAR(50),
Non_hispanic VARCHAR(50)
);
```

6.3 Load the dataset into the table:

Query:

```
LOAD DATA LOCAL INFILE
'/Users/sabiyatabasum/Desktop/proj4/clean_data.csv' INTO TABLE
proj4 FIELDS TERMINATED BY ',' ENCLOSED BY '"' LINES TERMINATED
BY '\n' IGNORE 1 ROWS;
```

6.4 Show the table schema:

Query: `mysql> SHOW COLUMNS FROM proj4;`

```
mysql> SHOW COLUMNS FROM proj4;
+----------------+--------------+------+-----+---------+----------------+
| Field          | Type         | Null | Key | Default | Extra          |
+----------------+--------------+------+-----+---------+----------------+
| id             | int          | NO   | PRI | NULL    | auto_increment |
| INDICATOR      | varchar(255) | YES  |     | NULL    |                |
| PANEL          | varchar(255) | YES  |     | NULL    |                |
| PANEL_NUM      | int          | YES  |     | NULL    |                |
| UNIT           | varchar(255) | YES  |     | NULL    |                |
| UNIT_NUM       | int          | YES  |     | NULL    |                |
| STUB_NAME      | varchar(255) | YES  |     | NULL    |                |
| STUB_NAME_NUM  | int          | YES  |     | NULL    |                |
| STUB_LABEL     | varchar(255) | YES  |     | NULL    |                |
| STUB_LABEL_NUM | int          | YES  |     | NULL    |                |
| YEAR           | int          | YES  |     | NULL    |                |
| YEAR_NUM       | int          | YES  |     | NULL    |                |
| AGE            | varchar(255) | YES  |     | NULL    |                |
| AGE_NUM        | int          | YES  |     | NULL    |                |
| ESTIMATE       | int          | YES  |     | NULL    |                |
| SEX            | varchar(10)  | YES  |     | NULL    |                |
| Hispanic       | varchar(50)  | YES  |     | NULL    |                |
| Non_hispanic   | varchar(50)  | YES  |     | NULL    |                |
+----------------+--------------+------+-----+---------+----------------+
18 rows in set (0.01 sec)
```

6.5 Perform queries on the database:

Query:

```
mysql> SELECT INDICATOR, YEAR, STUB_LABEL, ESTIMATE FROM proj4
WHERE YEAR = 2004;
```

```
mysql> SELECT INDICATOR, YEAR, STUB_LABEL, ESTIMATE FROM proj4 WHERE YEAR = 2004;
+--------------------------+------+-----------------------------------------------------------------+----------+
| INDICATOR                | YEAR | STUB_LABEL                                                      | ESTIMATE |
+--------------------------+------+-----------------------------------------------------------------+----------+
| Drug overdose death rates | 2004 | All persons                                                     |        9 |
| Drug overdose death rates | 2004 | Male                                                            |       12 |
| Drug overdose death rates | 2004 | Female                                                          |        7 |
| Drug overdose death rates | 2004 | Male: White                                                     |       13 |
| Drug overdose death rates | 2004 | Male: Black or African American                                 |       11 |
| Drug overdose death rates | 2004 | Male: American Indian or Alaska Native                          |       11 |
| Drug overdose death rates | 2004 | Male: Asian or Pacific Islander                                 |        2 |
| Drug overdose death rates | 2004 | Female: White                                                   |        8 |
| Drug overdose death rates | 2004 | Female: Black or African American                               |        6 |
| Drug overdose death rates | 2004 | Female: American Indian or Alaska Native                        |        8 |
| Drug overdose death rates | 2004 | Female: Asian or Pacific Islander                               |        1 |
| Drug overdose death rates | 2004 | Male: Hispanic or Latino: All races                             |        8 |
| Drug overdose death rates | 2004 | Male: Not Hispanic or Latino: White                             |       14 |
| Drug overdose death rates | 2004 | Male: Not Hispanic or Latino: Black                             |       11 |
| Drug overdose death rates | 2004 | Male: Not Hispanic or Latino: American Indian or Alaska Native  |       15 |
| Drug overdose death rates | 2004 | Male: Not Hispanic or Latino: Asian or Pacific Islander         |        2 |
| Drug overdose death rates | 2004 | Female: Hispanic or Latino: All races                           |        3 |
| Drug overdose death rates | 2004 | Female: Not Hispanic or Latino: White                           |        8 |
| Drug overdose death rates | 2004 | Female: Not Hispanic or Latino: Black                           |        6 |
| Drug overdose death rates | 2004 | Female: Not Hispanic or Latino: American Indian or Alaska Native |       10 |
| Drug overdose death rates | 2004 | Female: Not Hispanic or Latino: Asian or Pacific Islander       |        1 |
| Drug overdose death rates | 2004 | All persons                                                     |        9 |
| Drug overdose death rates | 2004 | Under 15 years                                                  |        0 |
| Drug overdose death rates | 2004 | 15–24 years                                                     |        7 |
| Drug overdose death rates | 2004 | 25–34 years                                                     |       12 |
| Drug overdose death rates | 2004 | 35–44 years                                                     |       19 |
| Drug overdose death rates | 2004 | 45–54 years                                                     |       19 |
| Drug overdose death rates | 2004 | 55–64 years                                                     |        8 |
| Drug overdose death rates | 2004 | 65–74 years                                                     |        3 |
| Drug overdose death rates | 2004 | 75–84 years                                                     |        3 |
| Drug overdose death rates | 2004 | 85 years and over                                               |        4 |
| Drug overdose death rates | 2004 | Male                                                            |       12 |
| Drug overdose death rates | 2004 | Female                                                          |        7 |
| Drug overdose death rates | 2004 | Male: Under 15 years                                            |        0 |
| Drug overdose death rates | 2004 | Male: 15–24 years                                               |       10 |
| Drug overdose death rates | 2004 | Male: 25–34 years                                               |       17 |
| Drug overdose death rates | 2004 | Male: 35–44 years                                               |       24 |
| Drug overdose death rates | 2004 | Male: 45–54 years                                               |       24 |
| Drug overdose death rates | 2004 | Male: 55–64 years                                               |        9 |
| Drug overdose death rates | 2004 | Male: 65–74 years                                               |        3 |
| Drug overdose death rates | 2004 | Male: 75–84 years                                               |        3 |
| Drug overdose death rates | 2004 | Male: 85 years and over                                         |        5 |
| Drug overdose death rates | 2004 | Female: Under 15 years                                          |        0 |
| Drug overdose death rates | 2004 | Female: 15–24 years                                             |        3 |
| Drug overdose death rates | 2004 | Female: 25–34 years                                             |        7 |
| Drug overdose death rates | 2004 | Female: 35–44 years                                             |       15 |
| Drug overdose death rates | 2004 | Female: 45–54 years                                             |       15 |
| Drug overdose death rates | 2004 | Female: 55–64 years                                             |        7 |
| Drug overdose death rates | 2004 | Female: 65–74 years                                             |        3 |
| Drug overdose death rates | 2004 | Female: 75–84 years                                             |        3 |
| Drug overdose death rates | 2004 | Female: 85 years and over                                       |        4 |
| Drug overdose death rates | 2004 | All persons                                                     |        5 |
| Drug overdose death rates | 2004 | Male                                                            |        6 |
| Drug overdose death rates | 2004 | Female                                                          |        3 |
```

.
.
.

```
| Drug overdose death rates | 2004 | Female: 25–34 years                                             |        0 |
| Drug overdose death rates | 2004 | Female: 35–44 years                                             |        1 |
| Drug overdose death rates | 2004 | Female: 45–54 years                                             |        0 |
| Drug overdose death rates | 2004 | Female: 55–64 years                                             |        0 |
| Drug overdose death rates | 2004 | Female: 65–74 years                                             |        0 |
| Drug overdose death rates | 2004 | Female: 75–84 years                                             |        0 |
| Drug overdose death rates | 2004 | Female: 85 years and over                                       |        0 |
+--------------------------+------+-----------------------------------------------------------------+----------+
306 rows in set (0.00 sec)
```

Query:

```
SELECT YEAR, STUB_LABEL, MIN(ESTIMATE) AS MinEstimate,
MAX(ESTIMATE) AS MaxEstimate FROM proj4 GROUP BY YEAR,
STUB_LABEL;
```

```
mysql> SELECT YEAR, STUB_LABEL, MIN(ESTIMATE) AS MinEstimate, MAX(ESTIMATE) AS MaxEstimate   FROM proj4  GROUP BY YE
+------+-----------------------------------------------------------------------------+-------------+-------------+
| YEAR | STUB_LABEL                                                                  | MinEstimate | MaxEstimate |
+------+-----------------------------------------------------------------------------+-------------+-------------+
| 1999 | All persons                                                                 |           0 |           6 |
| 2000 | All persons                                                                 |           0 |           6 |
| 2001 | All persons                                                                 |           0 |           7 |
| 2002 | All persons                                                                 |           0 |           8 |
| 2003 | All persons                                                                 |           1 |           9 |
| 2004 | All persons                                                                 |           1 |           9 |
| 2005 | All persons                                                                 |           1 |          10 |
| 2006 | All persons                                                                 |           1 |          12 |
| 2007 | All persons                                                                 |           1 |          12 |
| 2008 | All persons                                                                 |           1 |          12 |
| 2009 | All persons                                                                 |           1 |          12 |
| 2010 | All persons                                                                 |           1 |          12 |
| 2011 | All persons                                                                 |           1 |          13 |
| 2012 | All persons                                                                 |           1 |          13 |
| 2013 | All persons                                                                 |           1 |          14 |
| 2014 | All persons                                                                 |           1 |          15 |
| 2015 | All persons                                                                 |           1 |          16 |
| 2016 | All persons                                                                 |           1 |          20 |
| 2017 | All persons                                                                 |           1 |          22 |
| 1999 | Male                                                                        |           0 |           8 |
| 2000 | Male                                                                        |           0 |           8 |
| 2001 | Male                                                                        |           0 |           9 |
| 2002 | Male                                                                        |           1 |          11 |
| 2003 | Male                                                                        |           1 |          12 |
| 2004 | Male                                                                        |           1 |          12 |
| 2005 | Male                                                                        |           1 |          13 |
| 2006 | Male                                                                        |           1 |          15 |
| 2007 | Male                                                                        |           1 |          15 |
| 2008 | Male                                                                        |           1 |          15 |
| 2009 | Male                                                                        |           1 |          15 |
| 2010 | Male                                                                        |           1 |          15 |
| 2011 | Male                                                                        |           1 |          16 |
| 2012 | Male                                                                        |           1 |          16 |
| 2017 | Male                                                                        |           1 |          29 |
| 2013 | Male                                                                        |           1 |          17 |
| 2014 | Male                                                                        |           1 |          18 |
```

.
.
.

```
| 2018 | Male: 55-64 years                                                           |           2 |          37 |
| 2018 | Male: 65-74 years                                                           |           1 |          14 |
| 2018 | Male: 75-84 years                                                           |           0 |           5 |
| 2018 | Male: 85 years and over                                                     |           0 |           4 |
| 2018 | Female: Under 15 years                                                      |           0 |           0 |
| 2018 | Female: 15-24 years                                                         |           0 |           7 |
| 2018 | Female: 25-34 years                                                         |           1 |          21 |
| 2018 | Female: 35-44 years                                                         |           1 |          24 |
| 2018 | Female: 45-54 years                                                         |           2 |          25 |
| 2018 | Female: 55-64 years                                                         |           1 |          20 |
| 2018 | Female: 65-74 years                                                         |           0 |           7 |
| 2018 | Female: 75-84 years                                                         |           0 |           4 |
| 2018 | Female: 85 years and over                                                   |           0 |           4 |
+------+-----------------------------------------------------------------------------+-------------+-------------+
964 rows in set (0.01 sec)

mysql>
```

7.0 Enhancing Dataset Accuracy:

Apart from gender, age, and ethnicity, a variety of characteristics can be examined to improve the quality and completeness of a dataset:

- Food choices and nutrition have a substantial impact on general health and the risk of chronic disease. Regular exercise and physical activity help with physical fitness, mental wellness, and illness prevention. Tobacco smoking is a major risk factor for a variety of health concerns, including cardiovascular disease, respiratory infections, and cancer.
- Alcohol consumption: Excessive alcohol use can cause liver damage, heart disease, and certain types of cancer.
- Medical history: An individual's health profile and possible risk factors might be influenced by past medical illnesses, treatments, and family history.
- Mental health: Mental health issues such as anxiety, depression, and stress can have a substantial impact on general health and well-being.
- Healthcare access: Access to healthcare services, such as preventive care, treatment alternatives, and insurance coverage, can have an impact on health outcomes and disparities.
- Location: Access to healthcare, transportation, and environmental issues can all be influenced by one's geographic location, which might include urban, suburban, or rural settings.
- Veteran service and exposure to conflict or dangerous conditions can have long-term health consequences.
- Disability status: Access to resources, healthcare needs, and overall quality of life can all be influenced by disability status.

8.0 Technical terms related to the Dataset:

Census Bureau of the United States

The United States Census Bureau is a government body in charge of collecting and disseminating data on the country's population. The Census Bureau conducts the decennial census, which is a ten-year survey of all households in the United States. The Census Bureau also gathers information on a wide range of other areas, such as population growth, housing, and economic circumstances.

NVSS (National Vital Statistics System)

The National Vital Statistics System (NVSS) is a cooperative vital registration system in the United States that gathers and disseminates statistics on births, deaths, marriages, and divorces. The National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control and Prevention (CDC), manages the NVSS.

Mortality Records

The Mortality Files are a collection of files containing information on fatalities in the United States. The NVSS collects the files, which are then made available to the public via the NCHS website.

PUF (Public-use Mortality Files)

The Mortality Files for Public Use (PUF) are a subset of the Mortality Files that are expressly meant for use by researchers. The PUF files contain a subset of the variables found in the complete Mortality Files, but they are easier to access and deal with.

9.0 Conclusion:

- The investigation found that the average daily overdose fatality rate in the United States has more than quadrupled in the last 20 years, going from 0.3 per 100,000 people in 1999 to 1.4 per 100,000 people in 2018, a more than fourfold rise in 15 years.
- The spike in overdose deaths has been most noticeable among young people, with drug overdoses overtaking car accidents as the leading cause of death for Americans under the age of 50 in 2017.
- Overdose deaths have a very weak positive correlation with gender and a very weak negative correlation with race/ethnicity.
- Adults aged 35-44 were the most likely to die from an overdose, followed by adults aged 25-34 and 45-54. Overdose deaths in persons 65 and older are also on the rise.
- Overdose deaths have a substantial correlation with gender and race/ethnicity, showing that demographic groups are more vulnerable to drug overdoses.
- To compute overdose fatality rates in the United States, the dataset concentrates on sex and age, as these parameters have the greatest impact.

10.0 Citation of the dataset:

Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin: United States - Catalog.

(2022, April 29). https://catalog.data.gov/dataset/drug-overdose-death-rates-by-drug-type-sex-age-race-and-hispanic-origin-united-states-3f72f

"National Center for Health Statistics, Centers for Disease Control and Prevention (CDC). Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin: United States. Centres for Disease Control and Prevention, 2022.Accessed https://catalog.data.gov/dataset/drug-overdose-death-rates-by-drug-type-sex-age-race-and-hispanic-origin-united-states-3f72f.https://catalog.data.gov/dataset on October 11, 2023."

****