Report

Of

Project Stage 2

Group: 1002_L32_G3

Group members:

500366554 yxia0083
530649584 ruli0058
520595273 hliu4477
530333591 zche9596

Part A

Introduction

This project aims to analyse the relationship between medical occupation density and mortality using a dataset that comes from a data file sent via email in response to a data request
data-request.
Understanding the density of medical occupations and their impact on mortality is particularly important for reducing mortality. We will use the data set to explore the distribution of medical occupation density under different categories and compare it with the corresponding mortality data.

Target audience:

Medical profession and government

Benefits:

Through data analysis, we hope to find out how the density of medical professions affects mortality, especially in chronic disease mortality, child mortality, infant mortality, suicide mortality, traffic accident mortality, etc. The study's results will provide a reference for improving the allocation of medical resources, optimizing medical services, and reducing mortality.

Dataset Description:

We decided to use the CSV dataset "SDG_goal3_clean" in the second stage, which has 164 rows and 29 columns. It covers mortality and health worker density data for various countries from 2000 to 2015. This dataset comes from a data file sent via email in response to a data request.

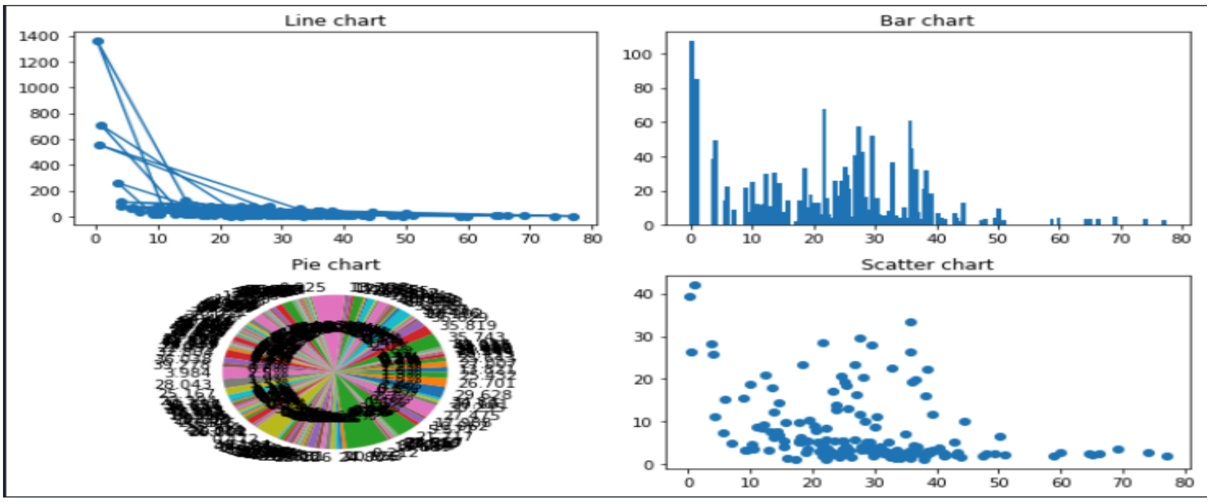
SID:500366554

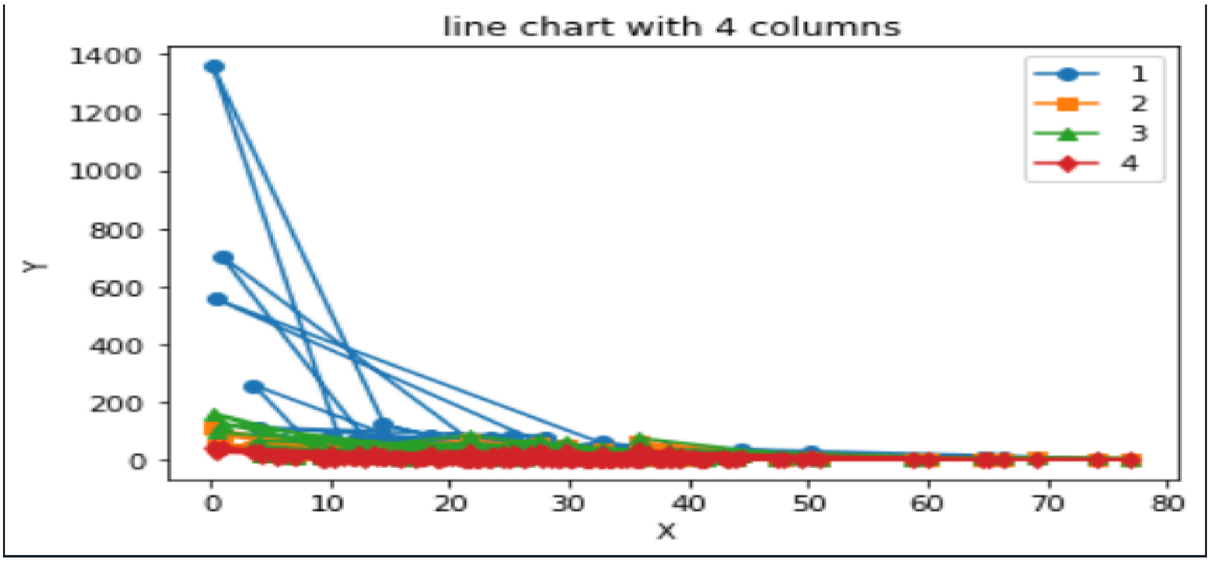
In this section I used the 4th column “Maternal mortality ratio”, the 6th column “Infant mortality rate (deaths per 1,000 live births) :::BOTHSEX”, the 9th column “Under-five mortality rate, by sex (deaths per 1,000 live births) :::BOTHSEX”, the 12th column “Neonanatal mortality rate (deaths per 1,000 live births)” and the 26th column “Health worker density, by type of occupation (per 10,000 population) ::PHYSICIAN” from the dataset SDG_goal3_clean.

Code:

```python
# -*- coding: utf-8 -*-
"""
Created on Fri Sep 27 15:22:34 2024

@author: Misooo
"""
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('SDG_goal3_clean.csv')
x = data['Health worker density, by type of occupation (per 10,000 population)::PHYSICIAN']
y1 = data['Maternal mortality ratio']
y2 = data['Infant mortality rate (deaths per 1,000 live births):::BOTHSEX']
y3 = data['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX']
y4 = data['Neonatal mortality rate (deaths per 1,000 live births)']

plt.figure(figsize=(10, 6))

# 1.
plt.subplot(2, 2, 1)
plt.plot(x, y1, marker='o')
plt.title('Line chart')

# 2.
plt.subplot(2, 2, 2)
plt.bar(x, y2)
plt.title('Bar chart')

# 3.
plt.subplot(2, 2, 3)
plt.pie(y3, labels=x, autopct='%1.1f%%')
plt.title('Pie chart')

# 4.
plt.subplot(2, 2, 4)
plt.scatter(x, y4)
plt.title('Scatter chart')

plt.tight_layout()
plt.show()


# 5.
plt.plot(x, y1, label=' 1', marker='o')
plt.plot(x, y2, label=' 2', marker='s')
plt.plot(x, y3, label=' 3', marker='^')
plt.plot(x, y4, label='4', marker='d')

plt.title('Line chart with 4 columns')
plt.xlabel('X ')
plt.ylabel('Y ')
plt.legend()

plt.show()
```

chart:





In this section we analyzed maternal mortality rates, infant mortality rates, under-five mortality rates, and neonatal mortality rates in relation to the density of doctors among healthcare workers using line charts, bar charts, pie charts, and scatter plots. I not only compared these different mortality rates with physician density separately, but also compared all four of them together with physician density as a line graph.
The findings indicate that as the density of doctors increases, mortality rates tend to decrease significantly. This underscores the importance of the distribution of medical resources, suggesting that increasing the density of doctors may effectively improve public health outcomes and reduce associated mortality rates.

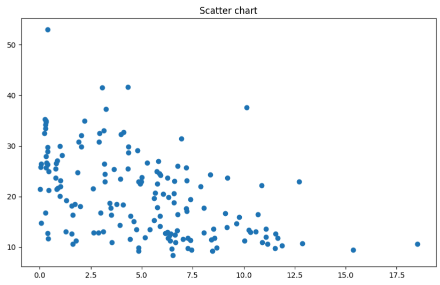
SID:530333591
zche9596

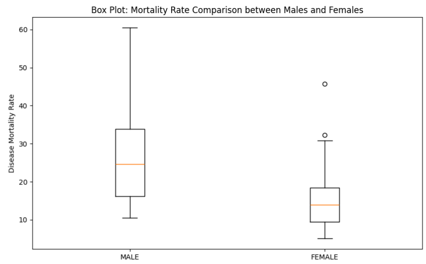
In this section, I focus on Pharmacist and use 4 different data to analysis the relationship between medical occupation density and mortality, which are **(28th column)** Health worker density, by type of occupation (per 10,000 population) ::PHARMACIST, **(13th column)**Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability) :::BOTHSEX, **(14th column)**Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability) :::MALE, **(15th column)**Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability) :::FEMALE. The points I focus on are what the situation is like about Mortality rate on cardiovascular disease, cancer and so on and is health worker density has a relation with the Mortality rate caused by this illness. Here is the code part:

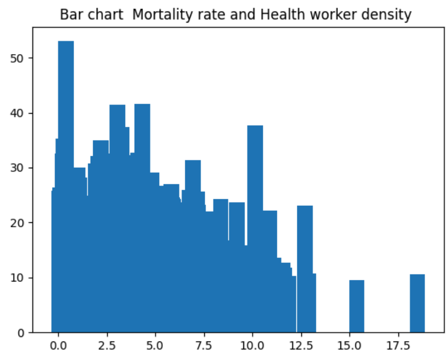
```python
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv("/Users/kyrie/Documents/悉尼及准备/学校资料/vscode-python/SDG_goal3_clean.csv")
x = data['Health worker density, by type of occupation (per 10,000 population)::PHARMACIST']
y1 = data['Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability):::BOTHSEX']
y2 = data['Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability):::MALE']
y3 = data['Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease (probability):::FEMALE']
y4 = data['Fraction of deaths due to diabetes, among deaths due to cardiovascular disease, cancer, diabetes or chronic respiratory disease']
y5 = ['Fraction of deaths due to cancer, among deaths due to cardiovascular disease, cancer, diabetes or chronic respiratory disease']
y6 = ['Universal health coverage (UHC) service coverage index']

plt.figure(figsize=(10, 6))
#1
plt.scatter(x, y1)
plt.title('Scatter chart')
plt.show()

#2
data_to_plot = [y2, y3]
plt.figure(figsize=(10, 6))
plt.boxplot(data_to_plot, labels=["MALE", "FEMALE"])
plt.title("Box Plot: Mortality Rate Comparison between Males and Females")
plt.ylabel("Disease Mortality Rate")
plt.show()
```

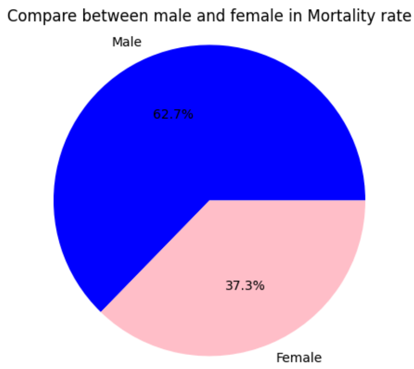
```python
#3
plt.bar(x,y1)
plt.title('Bar chart  Mortality rate and Health worker density ')
plt.show()

#4
total_y2 = y2.sum()
total_y3 = y3.sum()

sizes = [total_y2,total_y3]
labels = ['Male','Female']
colors = ['Blue','Pink']
plt.pie(sizes, colors=colors,labels=labels, autopct='%1.1f%%')
plt.axis('equal')
plt.title('Compare between male and female in Mortality rate')
plt.show()
```

And here are charts:









In this section, we use four charts to help me analyze the relation and the trend. Scatter chart shows that there exists a negative relation between pharmacist density and mortality rate. The data points are not closely clustered around a single line, indicating a certain level of dispersion. Look at the pir chart shows that Male has a majority in Mortality than female, which is obviously shown. But talking about the details, we can see the box plot, which not only shows the median(25 for male and 15 for female) Mortality rate compared between male and female, but shows Quartiles and IQR(between Q1 and Q3) which means most range in the distribution. Last, the bar chart shows the relation between density and mortality rate distribution, so we can easily compare.

SID: 530649584 ruli0058

This section examines Nursemidwife, utilising four distinct datasets to analyse the correlation between medical occupation density and mortality, including the 27th column. Health worker density categorised by occupation (per 10,000 population): NURSEMIDWIFE; 4th column, Maternal mortality ratio; 9th column Under-five mortality rate by sex (deaths per 1,000 live births): BOTHSEX; 19th column Suicide mortality rate by sex (deaths per 100,000 population): MALE; 20th column Suicide mortality rate by sex (deaths per 100,000 population): FEMALE.

Coding and figures part(including explanation–comments inside the codes):

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

```
df = pd.read_csv("SDG_goal3_clean.csv")
df.head()
```

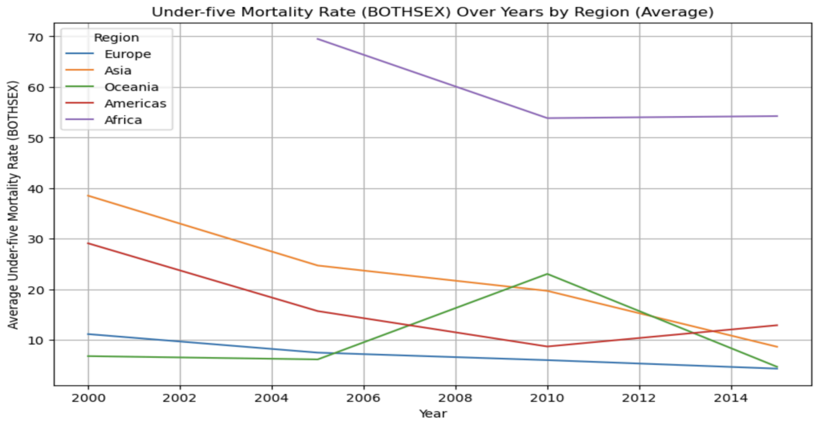
	Country	Year	Region	Maternal mortality ratio	Proportion of births attended by skilled health personnel (%)	Infant mortality rate (deaths per 1,000 live births):::BOTHSEX	Infant mortality rate (deaths per 1,000 live births):::MALE	Infant mortality rate (deaths per 1,000 live births):::FEMALE	Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX	Unde mortality (deat 1,0(births):::
0	Albania	2000	Europe	23	99.1	24.1	27.4	20.6	27.2	
1	Armenia	2000	Asia	43	96.8	27.0	29.8	24.1	30.7	
2	Armenia	2005	Asia	35	97.8	21.3	23.5	18.8	23.9	
3	Armenia	2010	Asia	32	99.5	16.5	18.3	14.6	18.5	
4	Australia	2000	Oceania	7	99.3	5.1	5.6	4.6	6.2	

5 rows × 28 columns

```
grouped_region = df.groupby('Region')['Maternal mortality ratio'].mean().reset_index()
grouped_region
```

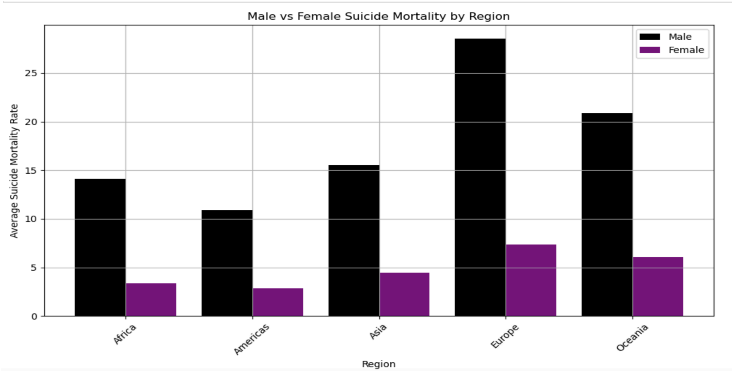
	Region	Maternal mortality ratio
0	Africa	367.250000
1	Americas	57.916667
2	Asia	37.395833
3	Europe	11.235294
4	Oceania	23.600000

```
# Group the data by 'Region' and 'Year', and calculate the mean of under-five mortality rate.
df_grouped = df.groupby(['Region', 'Year'])['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'].mean().reset_index()

# Create a new figure for the plot
plt.figure(figsize=(10, 6))

# Get a list of all unique regions from the 'Region' column of the original DataFrame.
regions = df['Region'].unique()

# Loop through each region to plot the under-five mortality rate over the years.
for region in regions:
    # Filter the grouped data to only include rows corresponding to the current region.
    region_data = df_grouped[df_grouped['Region'] == region]

    # Plot the 'Year' on the x-axis and the 'Under-five mortality rate (BOTHSEX)' on the y-axis and labeling the line.
    plt.plot(region_data['Year'], region_data['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'], label=region)

# Set the title of the plot to describe what is being shown.
plt.title('Under-five Mortality Rate (BOTHSEX) Over Years by Region (Average)')

# Set labels for the x-axis and y-axis.
plt.xlabel('Year')
plt.ylabel('Average Under-five Mortality Rate (BOTHSEX)')

# Add a legend to the plot to show which line corresponds to which region.
plt.legend(title='Region')

# Enable a grid for easier interpretation of the plot.
plt.grid(True)

# Display the final plot.
plt.show()
```



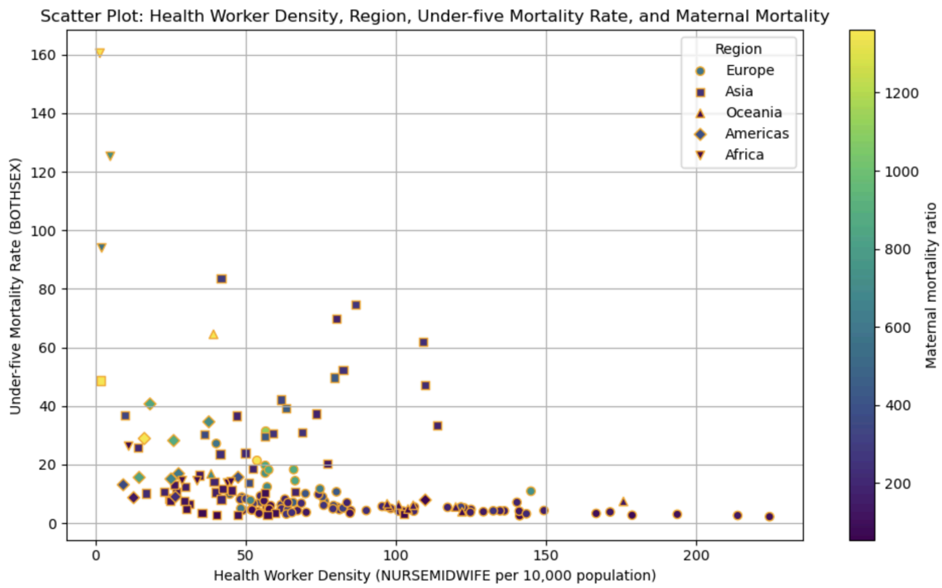
The diagram shows the under-five mortality rate trends from 2000 to 2015 across six regions: Africa, Asia, Europe, Oceania, and the Americas. All regions show a decline in child mortality, with Africa starting at the highest rate but gradually decreasing. Europe and the Americas maintain the lowest rates. Asia shows a significant drop, while Oceania has fluctuations. Overall, global child mortality is declining, but regional disparities persist.

```
# Group the data by 'Region' and calculate the average male and female suicide rates.
df_grouped = df.groupby('Region')[['Suicide mortality rate, by sex (deaths per 100,000 population):::MALE',\
                                   'Suicide mortality rate, by sex (deaths per 100,000 population):::FEMALE']].mean().reset_index()

# Set bar width and positions for each region.
bar_width = 0.4
index = np.arange(len(df_grouped))

# Create the figure.
plt.figure(figsize=(10, 6))

# Plot male suicide rates.
plt.bar(index, df_grouped['Suicide mortality rate, by sex (deaths per 100,000 population):::MALE'],\
width=bar_width, label='Male', color='black')

# Plot female suicide rates.
plt.bar(index + bar_width, df_grouped['Suicide mortality rate, by sex (deaths per 100,000 population):::FEMALE'],\
width=bar_width, label='Female', color='purple')

# Set x-axis and y-axis labels, and plot title.
plt.xlabel('Region')
plt.ylabel('Average Suicide Mortality Rate')
plt.title('Male vs Female Suicide Mortality by Region')

# Set x-ticks for regions and rotate for better visibility.
plt.xticks(index + bar_width / 2, df_grouped['Region'], rotation=45)

# Add legend and grid, then display the plot.
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```



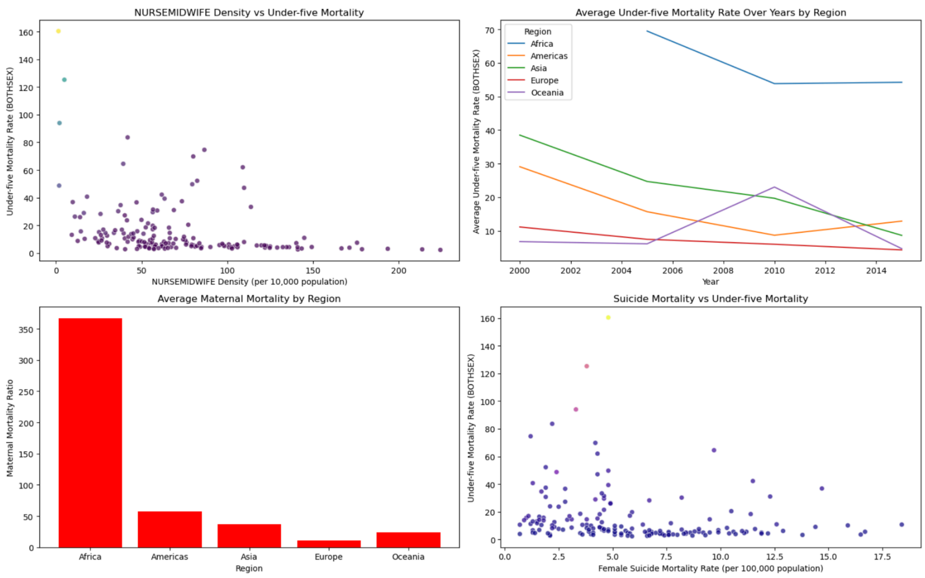
The bar chart compares suicide rates between males and girls across regions. Everywhere, male rates routinely exceed female rates. While Africa and the Americas have lower rates, Europe boasts the highest male suicide rate among all the continents; Asia follows. Overall, female suicide rates are relatively low in every area. This demonstrates quite evident gender differences in suicide death rates.

```python
# Extract data from the DataFrame for plotting.
x = df['Health worker density, by type of occupation (per 10,000 population)::NURSEMIDWIFE']  # Health worker density on x-axis
y = df['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX']  # Mortality rate on y-axis
colors = df['Maternal mortality ratio']  # Color points based on maternal mortality ratio
regions = df['Region'].unique()  # Get unique regions

# Assign different marker shapes for different regions.
shapes = ['o', 's', '^', 'D', 'v', '*']  # Define marker shapes

# Create the figure.
plt.figure(figsize=(10, 6))

# Plot a scatter plot for each region using different shapes.
for i, region in enumerate(regions):
    subset = df[df['Region'] == region]  # Filter data by region
    plt.scatter(subset['Health worker density, by type of occupation (per 10,000 population)::NURSEMIDWIFE'],
                subset['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'],
                c=subset['Maternal mortality ratio'],  # Color points by maternal mortality ratio
                cmap='viridis',  # Apply color map
                marker=shapes[i % len(shapes)],  # Assign unique marker shape to each region
                label=region,  # Label the region
                edgecolor='orange')  # Add an orange border to each marker

# Add a color bar to show variation in maternal mortality ratio.
plt.colorbar(label='Maternal mortality ratio')

# Set the plot title and labels for the x-axis (Health Worker Density) and y-axis (Under-five Mortality Rate).
plt.title('Scatter Plot: Health Worker Density, Region, Under-five Mortality Rate, and Maternal Mortality')
plt.xlabel('Health Worker Density (NURSEMIDWIFE per 10,000 population)')
plt.ylabel('Under-five Mortality Rate (BOTHSEX)')

# Add a legend to distinguish regions.
plt.legend(title='Region')

# Display grid, adjust layout, and show the plot.
plt.grid(True)
plt.tight_layout()
plt.show()
```



The scatter plot shows that as health worker density (NURSEMIDWIFE per 10,000 population) increases, under-five mortality rates tend to decrease across regions: Europe, Asia, Oceania, Americas, and Africa. Like Europe, areas with higher health worker density have lower mortality rates, while Africa has higher mortality with lower health worker density. The colour gradient, indicating the maternal mortality ratio, also suggests that regions with more health workers generally have lower maternal mortality. This highlights a negative correlation between healthcare access and both child and maternal mortality.

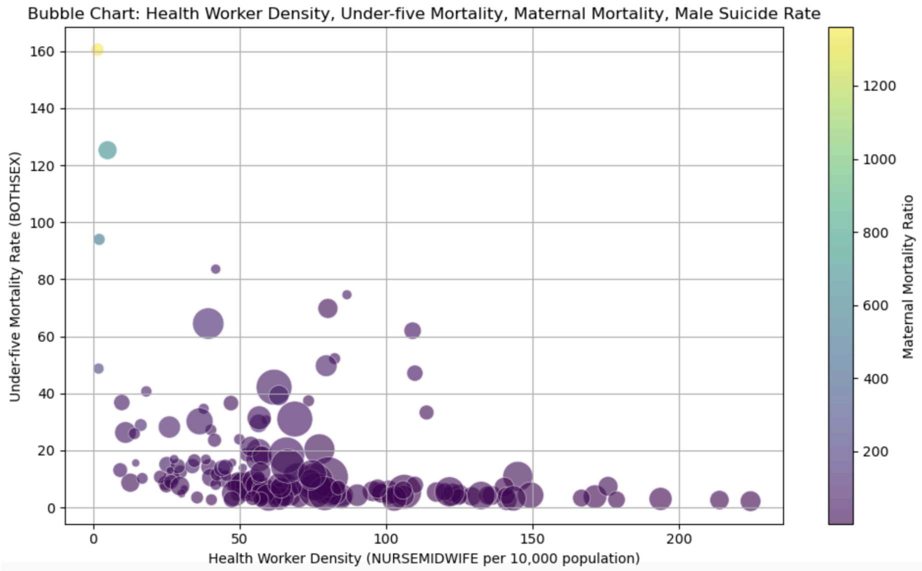
```python
# Create a 2x2 grid of subplots with a figure size of 16x10 inches.
fig, axs = plt.subplots(2, 2, figsize=(16, 10))

# Subplot 1: Scatter plot of NURSEMIDWIFE density vs under-five mortality, with colors representing maternal mortality ratio.
axs[0, 0].scatter(df['Health worker density, by type of occupation (per 10,000 population)::NURSEMIDWIFE'],
                  df['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'],
                  c=df['Maternal mortality ratio'], cmap='viridis', alpha=0.7, edgecolors="w", linewidth=0.6)
axs[0, 0].set_title('NURSEMIDWIFE Density vs Under-five Mortality')
axs[0, 0].set_xlabel('NURSEMIDWIFE Density (per 10,000 population)')
axs[0, 0].set_ylabel('Under-five Mortality Rate (BOTHSEX)')

# Subplot 2: Line plot of average under-five mortality rates by region over the years.
df_grouped = df.groupby(['Year', 'Region'])['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'].mean().unstack()
df_grouped.plot(ax=axs[0, 1], title='Average Under-five Mortality Rate Over Years by Region')
axs[0, 1].set_xlabel('Year')
axs[0, 1].set_ylabel('Average Under-five Mortality Rate (BOTHSEX)')

# Subplot 3: Bar plot showing average maternal mortality ratio by region.
df_grouped_maternal = df.groupby('Region')['Maternal mortality ratio'].mean().reset_index()
axs[1, 0].bar(df_grouped_maternal['Region'], df_grouped_maternal['Maternal mortality ratio'], color='red')
axs[1, 0].set_title('Average Maternal Mortality by Region')
axs[1, 0].set_xlabel('Region')
axs[1, 0].set_ylabel('Maternal Mortality Ratio')

# Subplot 4: Scatter plot of female suicide mortality vs under-five mortality, with colors representing maternal mortality ratio.
axs[1, 1].scatter(df['Suicide mortality rate, by sex (deaths per 100,000 population):::FEMALE'],
                  df['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX'],
                  c=df['Maternal mortality ratio'], cmap='plasma', alpha=0.7, edgecolors="w", linewidth=0.6)
axs[1, 1].set_title('Suicide Mortality vs Under-five Mortality')
axs[1, 1].set_xlabel('Female Suicide Mortality Rate (per 100,000 population)')
axs[1, 1].set_ylabel('Under-five Mortality Rate (BOTHSEX)')

# Adjust layout to avoid overlap and improve presentation.
plt.tight_layout()

# Display the plots.
plt.show()
```



The four subplots highlight the links among regions, death rates, and health resources. With Africa having high child death and low healthcare density, the top-left scatter plot demonstrates that greater NURSEMIDWIFE density correlates with lower under-five mortality. Though Africa still leads in high rates, the top-right line plot consistently drops under-five mortality over time. The bottom-left bar plot shows Africa's far greater maternal mortality than other areas. Finally, although Africa continues to be an outlier, the bottom-right scatter plot links female suicide rates with under-five mortality, so indicating reduced child mortality where female suicide rates are more significant. This underlines generally how important healthcare access is for lowering mortality.

Extra:

```
# X-axis: Nurse and midwife density; Y-axis: Under-five mortality rate;
# Bubble size: Male suicide mortality rate; Color: Maternal mortality ratio
x = df['Health worker density, by type of occupation (per 10,000 population)::NURSEMIDWIFE']
y = df['Under-five mortality rate, by sex (deaths per 1,000 live births):::BOTHSEX']
sizes = df['Suicide mortality rate, by sex (deaths per 100,000 population):::MALE']
colors = df['Maternal mortality ratio']

# Create a scatter plot (bubble chart) with bubble sizes and colors based on data
plt.figure(figsize=(10, 6))
bubble = plt.scatter(x, y, s=sizes*10, c=colors, cmap='viridis', alpha=0.6, edgecolors="w", linewidth=0.5)

# Add color bar for maternal mortality ratio
plt.colorbar(bubble, label='Maternal Mortality Ratio')

# Add title and axis labels
plt.title('Bubble Chart: Health Worker Density, Under-five Mortality, Maternal Mortality, Male Suicide Rate')
plt.xlabel('Health Worker Density (NURSEMIDWIFE per 10,000 population)')
plt.ylabel('Under-five Mortality Rate (BOTHSEX)')

# Show grid and optimize layout
plt.grid(True)
plt.tight_layout()
plt.show()
```



With bubble size representing the male suicide rate and colour showing the maternal mortality ratio, the bubble chart depicts the link between Health Worker Density (x-axis) and Under-five Mortality Rate (y-axis). While larger bubbles—more male suicide rates—cluster in areas with lower healthcare density, regions with greater healthworker density typically have lower under-five death rates. Darker bubbles, which indicate high mother mortality, are concentrated in areas with low health worker density and high under-five mortality, hence stressing the influence of healthcare availability on death rates.

SID : 520595273

```
In [118]: import pandas as pd
          import matplotlib.pyplot as plt
          import seaborn as sns
```

```
In [119]: data=pd.read_csv("SDG_goal3_clean.csv")
```

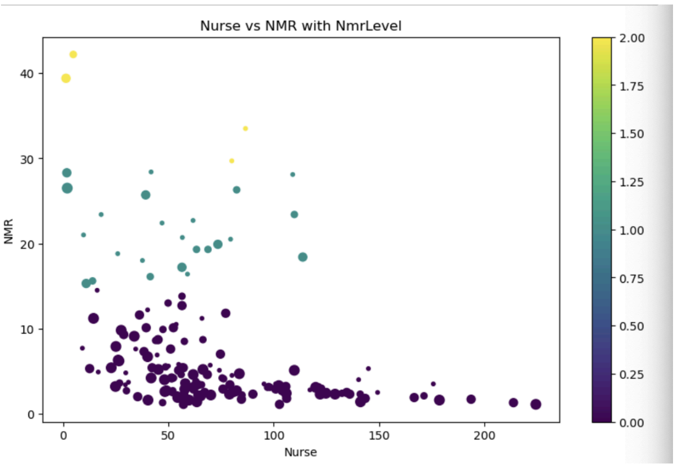
```
In [120]: data_new= data.rename({"Health worker density, by type of occupation (per 10,000 population)::NURSEMIDWIFE": "Nurse"
                              "Neonatal mortality rate (deaths per 1,000 live births)":"NMR",
                              "Infant mortality rate (deaths per 1,000 live births):::MALE":"boys",
                              "Infant mortality rate (deaths per 1,000 live births):::FEMALE":"girls"}, axis= 'columns')
```

```
In [121]: data_new["NmrLevel"]= pd.cut(data_new["NMR"], bins = 3, labels = ["low", "medium", "high"], ordered = True)
```

First, pandas, matplotlib and seaborn libraries were imported and csv files were read. In the data cleaning part, we renamed long names with the rename function to facilitate subsequent analysis. The NMR columns are grouped using the cut() function so that the low, medium, high intervals for analysis.

```
In [123]: year_size = {2000: 10, 2005: 30, 2010: 50, 2015: 70}

          data_new['YearSize'] = data_new['Year'].map(year_size)

          data_new.plot.scatter(
              x='Nurse',
              y='NMR',
              c=data_new['NmrLevel'].cat.codes,
              s=data_new['YearSize'],
              colormap='viridis',
              figsize=(10, 6),
              title='Nurse vs NMR with NmrLevel'
          )
          plt.show()
```

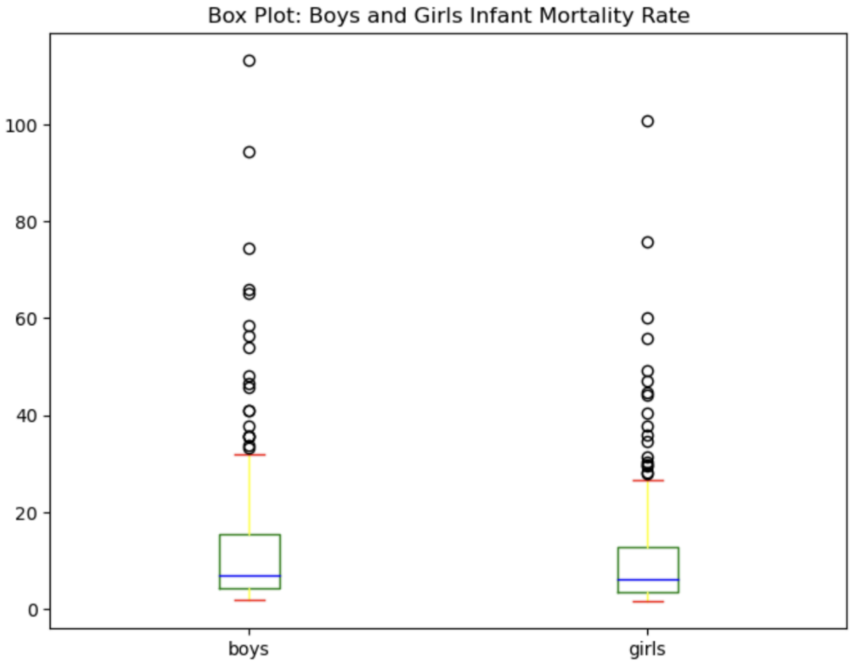
The plot.scatter function was used, with the X-axis set to nurse density and the Y-axis to neonatal mortality, and the size of the points controlled by the values in the YearSize column, indicating that the data point sizes varied across years. Also use the different colors of the NMRLevel control points and set the size of the chart to 10*6.



Purple points in the figure indicate lower mortality values, yellow points indicate higher mortality values, and a large number of data are concentrated on the number of 25 to 100 nurses. The size of the dots also represents the year, and we can find a significant reduction in infant mortality as the era progresses. In addition, there was an inverse relationship between nurse density and neonatal mortality, with lower mortality when nurse density was higher. In 2005, the infant mortality rate was higher than 40% in some areas because the number of nurses was almost zero, but in any year when the density of nurses was more than 150, the infant mortality rate was very close to zero.

```
In [124]: data_new[['boys', 'girls']].plot(kind='box', figsize=(8, 6),
                                          color=dict(boxes='green', whiskers='yellow', medians='blue', caps='red'))

          plt.title('Box Plot: Boys and Girls Infant Mortality Rate')

          plt.show()
```

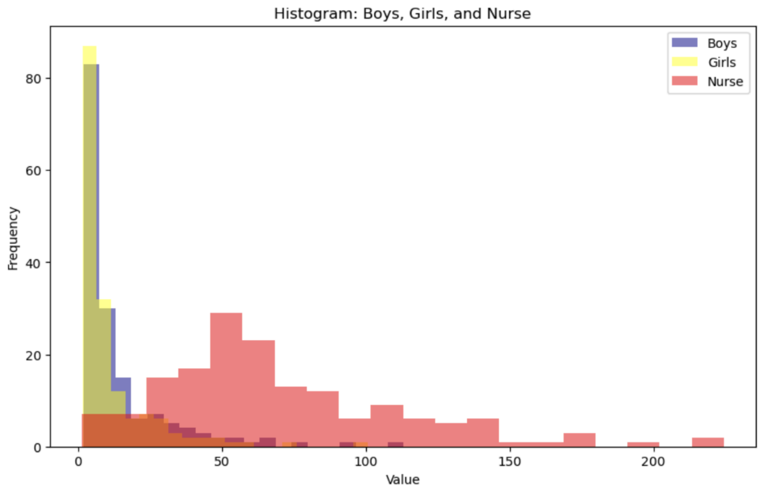
A boxplot of the data for boys and girls was created using the plot () function to show the distribution of infant male and female mortality rates. color=dict() defines the different colors, and figsize=() sets the size of the graph.



The boxplot shows that the median mortality rates for boys and girls are very close, about 5%. And their interquartile ranges were generally similar, suggesting that sex had a modest effect on infant mortality. However, there are a large number of discrete values in the boxplot, both for male and female infants. This may mean that some specific regions or time periods have abnormally high infant mortality rates.

```
In [125]: plt.figure(figsize=(10, 6))

plt.hist(data_new['boys'], bins=20, alpha=0.5, label='Boys', color='darkblue')

plt.hist(data_new['girls'], bins=20, alpha=0.5, label='Girls', color='yellow')

plt.hist(data_new['Nurse'], bins=20, alpha=0.5, label='Nurse', color='red')

plt.title('Histogram: Boys, Girls, and Nurse')
plt.xlabel('Value')
plt.ylabel('Frequency')


plt.legend()

plt.show()
```

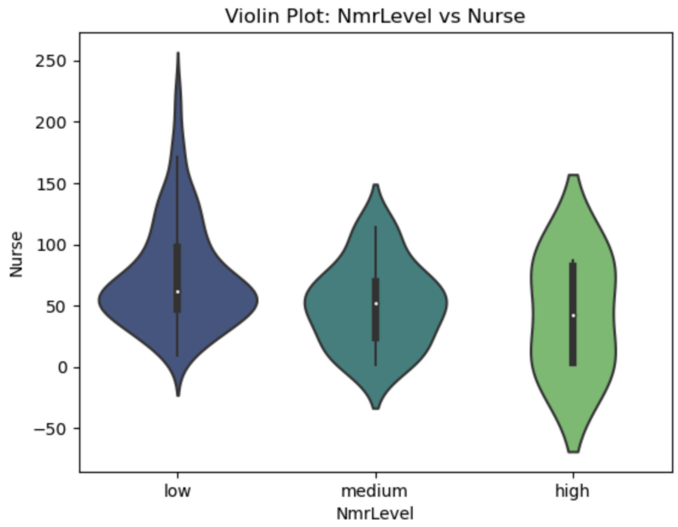
Use plt.hist(): to draw the frequency histogram, alpha=0.5 to make the image transparent, so that multiple histograms are easier to distinguish when superimposed, and different colors are also better to distinguish the data. Use the plt.legend() function to distinguish between disallowed data sets. The interval bins of each histogram were set to 20 to facilitate viewing the distribution of individual variables.



The infant mortality rates for both boys and girls are low in the 0-10 range, and the two decals are close, which means that the infant mortality rate is generally low between 2000 and 2015. Nurse density was distributed over a wide range, but mainly between 0 and 50. Nurse density is an important indicator to measure the resources of the health care system. A low density of nurses may indicate that there are insufficient medical resources in these areas, and the ability to provide medical services is relatively weak, especially when dealing with large numbers of patients or when long-term care is required.

```
sns.violinplot(x='NmrLevel', y='Nurse', data=data_new, palette="viridis")

plt.title('Violin Plot: NmrLevel vs Nurse')

plt.show()
```

sns.violinplot() was used to draw the violinplot with NmrLevel set to the X-axis and Nurse set to the Y-axis. palette= "viridis" gives the chart a more recognizable color. The plt.title() function is used to set the title for the chart.



The violin plot shows that the low NmrLevel is concentrated around 50-150, with high density data. The medium NmrLevel nurse density distribution was relatively concentrated, with most of the data in the 50 to 100 range and a smoother distribution without particularly high density. However, high is widely distributed and the overall density is relatively flat.

```
In [140]: data_selected = data_new[['Nurse', 'NMR', 'boys', 'girls']]

          median_values = data_selected.median()
          mean_values = data_selected.mean()
          std_values = data_selected.std()

          table_list = [
              ['Median'] + median_values.tolist(),['Mean'] + mean_values.tolist(),['Standard Deviation'] + std_values.tolist()
          ]

          headers = ['Statistic', 'Nurse', 'NMR', 'boys', 'girls']

          table = tabulate(table_list, headers=headers, tablefmt='grid')

          print(table)
```

The tolist () function was used to convert the data into a python list, while calculating its mean, median, and standard deviation. Use the tabulate function and select the grid style to create a readable table.

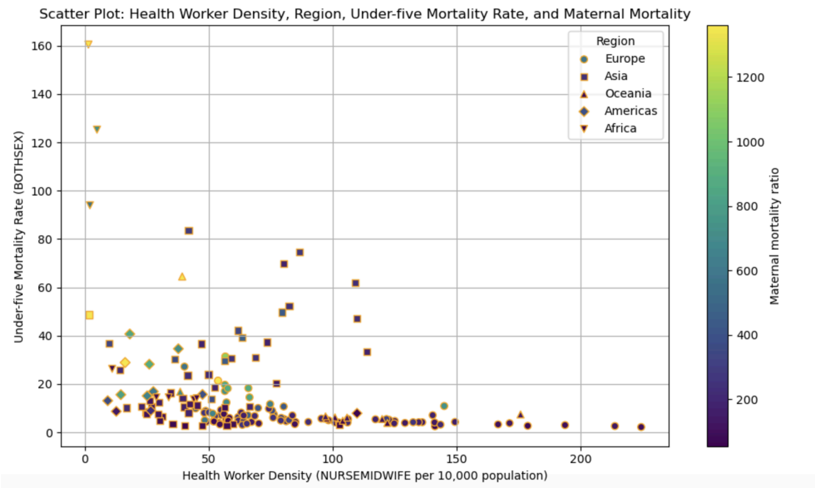
Statistic	Nurse	NMR	boys	girls
Median	59.089	4.2	7.1	6.1
Mean	70.0631	7.84908	14.2399	11.6577
Standard Deviation	42.4006	8.14126	17.3975	14.3163

The high standard deviation of the Nurse column shows that the distribution of nurse density is very scattered, indicating large differences in health resources across regions. The high standard deviation of NMR shows that neonatal mortality varies greatly across regions and may be influenced by medical resources, sanitation, and other factors.

PartB：

Limitations and uncertainties：

In this project, we analyzed the relationship between healthcare density and mortality. When analyzing CSV files, limitations may include insufficient representation of data, small sample size, potential confounding variables, and biases in the data collection process. For example, the dataset only used 2000, 2005, 2010, and 2015 data. The reliability of data sources cannot be guaranteed, and data may be inconsistent in some remote areas due to different collection times and methods. The range of data we used does not reflect changes in recent years. In addition, the data were unevenly distributed across the five continents. In addition, the units of each mortality rate were different. This may affect the reliability of the conclusions.



More specifically, according to the scatter plot, which shows the health worker density, region, under-five mortality rate, and maternal mortality, three outliers come from Africa. Calculating the mean values of each region will influence the degree of accuracy.

Conclusion：

Statistic	Nurse	NMR	boys	girls
Median	59.089	4.2	7.1	6.1
Mean	70.0631	7.84908	14.2399	11.6577
Standard Deviation	42.4006	8.14126	17.3975	14.3163

For example, this chart shows:

In recent years, infant mortality has decreased significantly, and medical staff has increased yearly. Infant mortality rates fell from more than 40 per cent in some areas in 2000 to less than 30 per cent in 2015 and less than 10 per cent in most areas. In the gender comparison, the distribution of infant mortality rates was similar for boys and girls, with close interquartile ranges. Still, both contained some discrete values, which may be related to the level of medical care and social differences across countries. We also found that higher nurse density was associated with lower infant mortality. In areas with low mortality rates, nurse density was concentrated around 50, while in areas with medium and high mortality rates, the number of nursing staff was more evenly distributed. Thus, the key factor in infant mortality may lie in the uniform distribution of care density rather than too high or too low.

Most of the variation in mortality rates is related to health care. Our results show that many mortality rates tend to decrease with changes in healthcare density.

Despite these conclusions, we only investigated a small part of mortality rates related to health care. Further research is needed to discover other aspects.