In [1]:
```
import pandas  as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:
```
url="https://www.worldlifeexpectancy.com/country-health-profile/india/"
tables = pd.read_html(url)
for i, table in enumerate(tables):
    print(f"Table {i+1}:")
    print(table)
    print("\n")
```

```
34           06/30/21  30.411m  2.237m  412688   57207
35           05/31/21  28.174m  9.017m  355481  137080
36           04/30/21  19.157m  6.936m  218401   50931
37           03/31/21  12.221m  1.109m  167470    5911
38           02/28/21  11.112m  353437  161559    2839
39           01/31/21  10.759m  472290  158720    5589
40           12/31/20  10.286m  823075  153131   11701
41           11/30/20   9.463m   1.28m  141430   15954
42           10/31/20   8.183m  1.873m  125476   24074
43           09/30/20    6.31m  2.622m  101402   34122
44           08/31/20   3.688m  1.991m   67280   29783
45           07/31/20   1.697m  1.111m   37497   19704
46           06/30/20   585792  395183   17793   12293
47           05/31/20   190609  155746    5500    4341
48           04/30/20    34863   33466    1159    1123
49           03/31/20     1397    1394      36      36
50           02/29/20        3       3       0       0
51           01/31/20        0       0       0       0
```

In [8]:
```
if tables:
    # Access and print the second table
    second_table = tables[1]  # Correct variable name
    print("Second Table:")
    print(second_table)  # Use the correct variable name
else:
    print("No tables found on the webpage.")
```

```
Second Table:
    TOP 50 CAUSES OF DEATH TOP 50 CAUSES OF DEATH.1    Rate  World Rank
0                      1.0   Coronary Heart Disease  140.72          55
1                      2.0              Lung Disease   87.90           3
2                      3.0                    Stroke   64.97         113
3                      4.0       Diarrhoeal diseases   64.45          20
4                      5.0  Influenza and Pneumonia   35.30          77
5                      6.0              Tuberculosis   32.84          32
6                      7.0         Diabetes Mellitus   25.49          97
7                      8.0             Liver Disease   22.24          83
8                      9.0                     Falls   21.87           1
9                     10.0            Kidney Disease   19.90         107
10                    11.0                    Asthma   18.49          24
11                    12.0    Road Traffic Accidents   16.33          90
12                    13.0     Alzheimers & Dementia   14.60         137
13                    14.0             Breast Cancer   13.62         125
14                    15.0                   Suicide   12.91          41
15                    16.0            Other Injuries   12.19          44
16                    17.0          Low Birth Weight   11.98          19
17                    18.0               Oral Cancer   11.45           6
18                    19.0  Rheumatic Heart Disease   10.94           8
19                    20.0              Hypertension   10.68         128
20                    21.0              Lung Cancers    7.82         117
21                    22.0           Cervical Cancer    7.46         100
22                    23.0            Stomach Cancer    7.07          76
23                    24.0      Colon-Rectum Cancers    7.03         118
24                    25.0            Prostate Cancer    6.69         147
```

In [6]:
```
if tables:
    # Access and print the first table
    first_table = tables[0]
    print("First Table:")
    print(first_table)
else:
    print("No tables found on the webpage.")
```

```
First Table:
   Unnamed: 0_level_0 Unnamed: 1_level_0 Unnamed: 2_level_0 World Rank
   Unnamed: 0_level_1               Male             Female          M    F
0           At Birth               69.5               72.2        107  122
1              Age 5               71.9               74.7         98  119
2             Age 10               72.1               74.9         98  120
3             Age 15               72.2               75.1         96  120
4             Age 20               72.4               75.4         98  118
5             Age 25               72.8               75.7         97  116
6             Age 30               73.2               76.0         97  116
7             Age 35               73.6               76.3         94  115
8             Age 40               74.2               76.7         93  115
9             Age 45               74.9               77.1         91  113
10            Age 50               75.7               77.7         91  115
11            Age 55               76.7               78.5         87  113
12            Age 60               78.1               79.5         85  110
13            Age 65               79.7               80.9         84  106
14            Age 70               81.7               82.6         88  106
15            Age 75               84.0               84.7         84  104
16            Age 80               86.7               87.1         86  109
17            Age 85               90.1               90.4         77   95
```

In [9]:
```
if tables:
    # Access and print the second table
    third_table = tables[2]  # Correct variable name
    print("Third Table:")
    print(third_table)  # Use the correct variable name
else:
    print("No tables found on the webpage.")
```

```
Third Table:
    TOP 50 CAUSES OF DEATH TOP 50 CAUSES OF DEATH.1  Rate  World Rank
0                     26.0     Peptic Ulcer Disease  6.24          42
1                     27.0             Birth Trauma  5.11          64
2                     28.0     Congenital Anomalies  4.70         122
3                     29.0      Parkinson's Disease  4.57          65
4                     30.0             Encephalitis  4.12           3
5                     31.0                 Violence  3.84         104
6                     32.0                Drownings  3.76          51
7                     33.0             Ovary Cancer  3.73         120
8                     34.0                Lymphomas  3.44         136
9                     35.0        Oesophagus Cancer  3.40          57
10                    36.0                  HIV/AIDS  3.35          87
11                    37.0          Pancreas Cancer  2.97         113
12                    38.0                 Leukemia  2.79         129
13                    39.0             Liver Cancer  2.67         175
14                    40.0               Meningitis  2.52          55
15                    41.0                 Epilepsy  2.44          47
16                    42.0                    Fires  1.84          79
17                    43.0                  Alcohol  1.81          68
18                    44.0      Maternal Conditions  1.79          74
19                    45.0             Skin Disease  1.62          92
20                    46.0                   Dengue  1.51           5
21                    47.0           Bladder Cancer  1.34         140
22                    48.0      Endocrine Disorders  1.29         169
23                    49.0     Rheumatoid Arthritis  1.19           4
24                    50.0             Uterin Cancer  1.19         155
```

In [10]:
```
if tables:
    # Access and print the second table
    fourth_table = tables[1]  # Correct variable name
    print("Fourth Table:")
    print(fourth_table)  # Use the correct variable name
else:
    print("No tables found on the webpage.")
```

```
Fourth Table:
    TOP 50 CAUSES OF DEATH TOP 50 CAUSES OF DEATH.1    Rate  World Rank
0                      1.0   Coronary Heart Disease  140.72          55
1                      2.0             Lung Disease   87.90           3
2                      3.0                   Stroke   64.97         113
3                      4.0      Diarrhoeal diseases   64.45          20
4                      5.0  Influenza and Pneumonia   35.30          77
5                      6.0             Tuberculosis   32.84          32
6                      7.0        Diabetes Mellitus   25.49          97
7                      8.0            Liver Disease   22.24          83
8                      9.0                    Falls   21.87           1
9                     10.0           Kidney Disease   19.90         107
10                    11.0                   Asthma   18.49          24
11                    12.0   Road Traffic Accidents   16.33          90
12                    13.0    Alzheimers & Dementia   14.60         137
13                    14.0            Breast Cancer   13.62         125
14                    15.0                  Suicide   12.91          41
15                    16.0           Other Injuries   12.19          44
16                    17.0         Low Birth Weight   11.98          19
17                    18.0              Oral Cancer   11.45           6
18                    19.0  Rheumatic Heart Disease   10.94           8
19                    20.0             Hypertension   10.68         128
20                    21.0             Lung Cancers    7.82         117
21                    22.0          Cervical Cancer    7.46         100
22                    23.0           Stomach Cancer    7.07          76
23                    24.0     Colon-Rectum Cancers    7.03         118
24                    25.0           Prostate Cancer    6.69         147
```

In [15]:
```
first_table.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   TOP 50 CAUSES OF DEATH    25 non-null     float64
 1   TOP 50 CAUSES OF DEATH.1  25 non-null     object
 2   Rate                      25 non-null     float64
 3   World Rank                25 non-null     int64
dtypes: float64(2), int64(1), object(1)
memory usage: 932.0+ bytes
```

In [16]:
```
first_table.head()
```

Out[16]:

	TOP 50 CAUSES OF DEATH	TOP 50 CAUSES OF DEATH.1	Rate	World Rank
0	1.0	Coronary Heart Disease	140.72	55
1	2.0	Lung Disease	87.90	3
2	3.0	Stroke	64.97	113
3	4.0	Diarrhoeal diseases	64.45	20
4	5.0	Influenza and Pneumonia	35.30	77

In [17]:
```
first_table.tail()
```

Out[17]:

	TOP 50 CAUSES OF DEATH	TOP 50 CAUSES OF DEATH.1	Rate	World Rank
20	21.0	Lung Cancers	7.82	117
21	22.0	Cervical Cancer	7.46	100
22	23.0	Stomach Cancer	7.07	76
23	24.0	Colon-Rectum Cancers	7.03	118
24	25.0	Prostate Cancer	6.69	147

In [18]:
```
first_table.describe()
```

Out[18]:

	TOP 50 CAUSES OF DEATH	Rate	World Rank
count	25.000000	25.000000	25.000000
mean	13.000000	27.397600	70.720000
std	7.359801	31.340038	47.609978
min	1.000000	6.690000	1.000000
25%	7.000000	10.940000	24.000000
50%	13.000000	14.600000	77.000000
75%	19.000000	25.490000	113.000000
max	25.000000	140.720000	147.000000

In [19]:
```
first_table.dropna()
```

Out[19]:

	TOP 50 CAUSES OF DEATH	TOP 50 CAUSES OF DEATH.1	Rate	World Rank
0	1.0	Coronary Heart Disease	140.72	55
1	2.0	Lung Disease	87.90	3
2	3.0	Stroke	64.97	113
3	4.0	Diarrhoeal diseases	64.45	20
4	5.0	Influenza and Pneumonia	35.30	77
5	6.0	Tuberculosis	32.84	32
6	7.0	Diabetes Mellitus	25.49	97
7	8.0	Liver Disease	22.24	83
8	9.0	Falls	21.87	1
9	10.0	Kidney Disease	19.90	107
10	11.0	Asthma	18.49	24
11	12.0	Road Traffic Accidents	16.33	90
12	13.0	Alzheimers & Dementia	14.60	137
13	14.0	Breast Cancer	13.62	125
14	15.0	Suicide	12.91	41
15	16.0	Other Injuries	12.19	44
16	17.0	Low Birth Weight	11.98	19
17	18.0	Oral Cancer	11.45	6
18	19.0	Rheumatic Heart Disease	10.94	8
19	20.0	Hypertension	10.68	128
20	21.0	Lung Cancers	7.82	117
21	22.0	Cervical Cancer	7.46	100
22	23.0	Stomach Cancer	7.07	76
23	24.0	Colon-Rectum Cancers	7.03	118
24	25.0	Prostate Cancer	6.69	147

In [21]:
```
x=first_table.mean(numeric_only=True)
```

In [22]:
```
x
```

Out[22]:
```
TOP 50 CAUSES OF DEATH    13.0000
Rate                      27.3976
World Rank                70.7200
dtype: float64
```

In [23]:
```
first_table_filled=first_table.fillna(x)
```

In [24]:
```
first_table_filled
```

Out[24]:

	TOP 50 CAUSES OF DEATH	TOP 50 CAUSES OF DEATH.1	Rate	World Rank
0	1.0	Coronary Heart Disease	140.72	55
1	2.0	Lung Disease	87.90	3
2	3.0	Stroke	64.97	113
3	4.0	Diarrhoeal diseases	64.45	20
4	5.0	Influenza and Pneumonia	35.30	77
5	6.0	Tuberculosis	32.84	32
6	7.0	Diabetes Mellitus	25.49	97
7	8.0	Liver Disease	22.24	83
8	9.0	Falls	21.87	1
9	10.0	Kidney Disease	19.90	107
10	11.0	Asthma	18.49	24
11	12.0	Road Traffic Accidents	16.33	90
12	13.0	Alzheimers & Dementia	14.60	137
13	14.0	Breast Cancer	13.62	125
14	15.0	Suicide	12.91	41
15	16.0	Other Injuries	12.19	44
16	17.0	Low Birth Weight	11.98	19
17	18.0	Oral Cancer	11.45	6
18	19.0	Rheumatic Heart Disease	10.94	8
19	20.0	Hypertension	10.68	128
20	21.0	Lung Cancers	7.82	117
21	22.0	Cervical Cancer	7.46	100
22	23.0	Stomach Cancer	7.07	76
23	24.0	Colon-Rectum Cancers	7.03	118
24	25.0	Prostate Cancer	6.69	147

In [25]:
```
first_table_filled.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 4 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   TOP 50 CAUSES OF DEATH    25 non-null     float64
 1   TOP 50 CAUSES OF DEATH.1  25 non-null     object 
 2   Rate                      25 non-null     float64
 3   World Rank                25 non-null     int64  
dtypes: float64(2), int64(1), object(1)
memory usage: 932.0+ bytes
```

In [26]:
```
first_table_filled.to_csv("highest chance of death",index=False)
```

In [28]:
```
url2="https://www.worldatlas.com/articles/the-leading-causes-of-death-in-india.html"
tables = pd.read_html(url)
for i, table in enumerate(tables):
    print(f"Table {i+1}:")
    print(table)
    print("\n")
```

```
Table 1:
   Unnamed: 0_level_0 Unnamed: 1_level_0 Unnamed: 2_level_0 World Rank
   Unnamed: 0_level_1               Male             Female          M    F
0           At Birth               69.5               72.2        107  122
1              Age 5               71.9               74.7         98  119
2             Age 10               72.1               74.9         98  120
3             Age 15               72.2               75.1         96  120
4             Age 20               72.4               75.4         98  118
5             Age 25               72.8               75.7         97  116
6             Age 30               73.2               76.0         97  116
7             Age 35               73.6               76.3         94  115
8             Age 40               74.2               76.7         93  115
9             Age 45               74.9               77.1         91  113
10            Age 50               75.7               77.7         91  115
11            Age 55               76.7               78.5         87  113
12            Age 60               78.1               79.5         85  110
13            Age 65               79.7               80.9         84  106
14            Age 70               81.7               82.6         88  106
15            Age 75               84.0               84.7         84  104
```

In [29]:
```
if tables:
    # Access and print the second table
    first_table = tables[0]  # Correct variable name
    print("First Table:")
    print(first_table)  # Use the correct variable name
else:
    print("No tables found on the webpage.")
```

```
First Table:
   Unnamed: 0_level_0 Unnamed: 1_level_0 Unnamed: 2_level_0 World Rank
   Unnamed: 0_level_1               Male             Female          M    F
0           At Birth               69.5               72.2        107  122
1              Age 5               71.9               74.7         98  119
2             Age 10               72.1               74.9         98  120
3             Age 15               72.2               75.1         96  120
4             Age 20               72.4               75.4         98  118
5             Age 25               72.8               75.7         97  116
6             Age 30               73.2               76.0         97  116
7             Age 35               73.6               76.3         94  115
8             Age 40               74.2               76.7         93  115
9             Age 45               74.9               77.1         91  113
10            Age 50               75.7               77.7         91  115
11            Age 55               76.7               78.5         87  113
12            Age 60               78.1               79.5         85  110
13            Age 65               79.7               80.9         84  106
14            Age 70               81.7               82.6         88  106
15            Age 75               84.0               84.7         84  104
16            Age 80               86.7               87.1         86  109
17            Age 85               90.1               90.4         77   95
```

In [30]:
```
first_table.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18 entries, 0 to 17
Data columns (total 5 columns):
 #   Column                                      Non-Null Count  Dtype
---  ------                                      --------------  -----
 0   (Unnamed: 0_level_0, Unnamed: 0_level_1)  18 non-null     object
 1   (Unnamed: 1_level_0, Male)                 18 non-null     float64
 2   (Unnamed: 2_level_0, Female)               18 non-null     float64
 3   (World Rank, M)                            18 non-null     int64
 4   (World Rank, F)                            18 non-null     int64
dtypes: float64(2), int64(2), object(1)
memory usage: 852.0+ bytes
```

In [31]:
```
first_table.describe()
```

Out[31]:

	Unnamed: 1_level_0	Unnamed: 2_level_0	World Rank	
	Male	Female	M	F
count	18.000000	18.000000	18.000000	18.000000
mean	76.638889	78.638889	91.722222	112.888889
std	5.693561	4.791308	7.209516	6.850409
min	69.500000	72.200000	77.000000	95.000000
25%	72.500000	75.475000	86.250000	109.250000
50%	74.550000	76.900000	92.000000	115.000000
75%	79.300000	80.550000	97.000000	117.500000
max	90.100000	90.400000	107.000000	122.000000

In [32]:
```
x=first_table.mean(numeric_only=True)
```

In [33]:
```
x
```

Out[33]:
```
Unnamed: 1_level_0  Male       76.638889
Unnamed: 2_level_0  Female     78.638889
World Rank          M          91.722222
                    F         112.888889
dtype: float64
```

In [35]:
```
First_Table=first_table.fillna(x)
```

In [36]:
```
First_Table
```

Out[36]:

	Unnamed: 0_level_0	Unnamed: 1_level_0	Unnamed: 2_level_0	World Rank	
	Unnamed: 0_level_1	Male	Female	M	F
0	At Birth	69.5	72.2	107	122
1	Age 5	71.9	74.7	98	119
2	Age 10	72.1	74.9	98	120
3	Age 15	72.2	75.1	96	120
4	Age 20	72.4	75.4	98	118
5	Age 25	72.8	75.7	97	116
6	Age 30	73.2	76.0	97	116
7	Age 35	73.6	76.3	94	115
8	Age 40	74.2	76.7	93	115
9	Age 45	74.9	77.1	91	113
10	Age 50	75.7	77.7	91	115
11	Age 55	76.7	78.5	87	113
12	Age 60	78.1	79.5	85	110
13	Age 65	79.7	80.9	84	106
14	Age 70	81.7	82.6	88	106
15	Age 75	84.0	84.7	84	104
16	Age 80	86.7	87.1	86	109
17	Age 85	90.1	90.4	77	95

In [38]:
```
First_Table.describe().T
```

Out[38]:

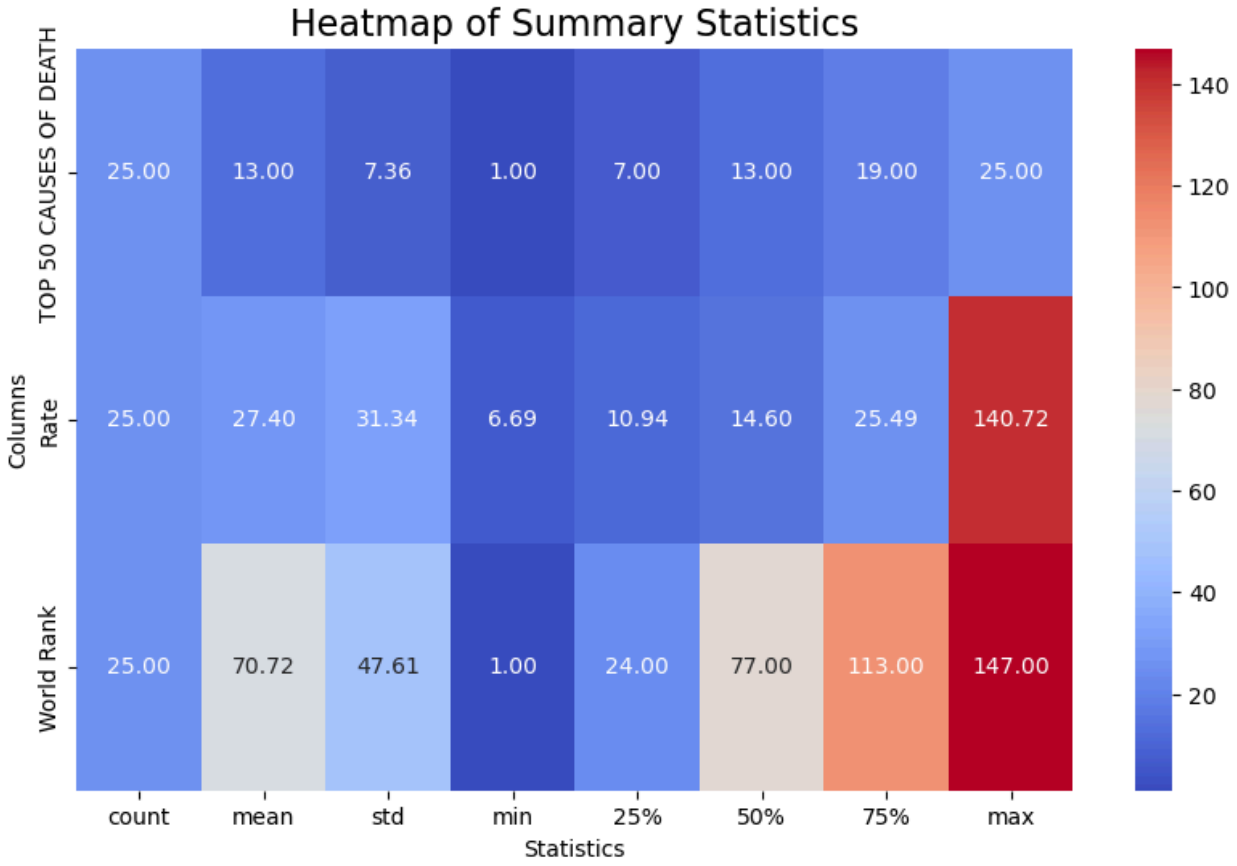
		count	mean	std	min	25%	50%	75%	max
Unnamed: 1_level_0	Male	18.0	76.638889	5.693561	69.5	72.500	74.55	79.30	90.1
Unnamed: 2_level_0	Female	18.0	78.638889	4.791308	72.2	75.475	76.90	80.55	90.4
World Rank	M	18.0	91.722222	7.209516	77.0	86.250	92.00	97.00	107.0
	F	18.0	112.888889	6.850409	95.0	109.250	115.00	117.50	122.0

In [40]:
```
first_table_filled.describe().T
```

Out[40]:

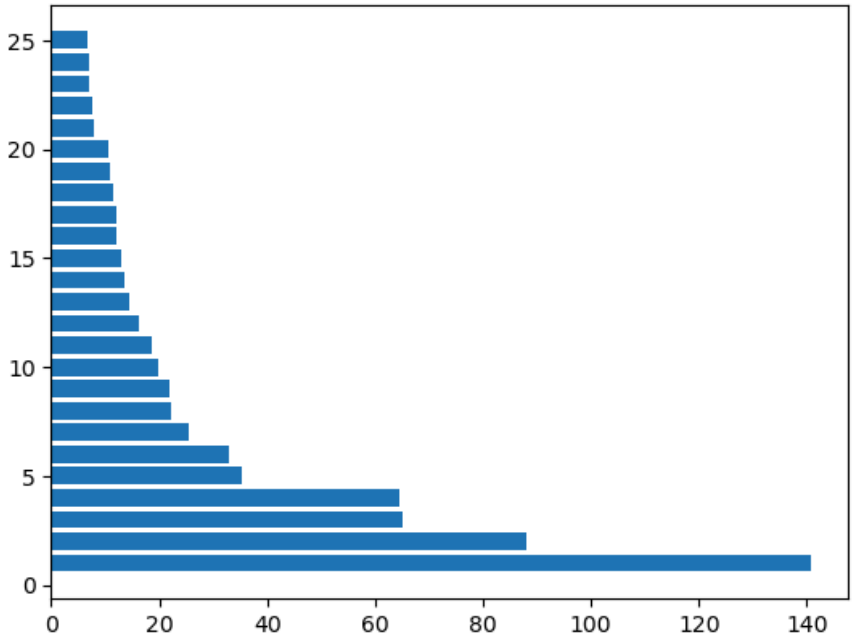
	count	mean	std	min	25%	50%	75%	max
TOP 50 CAUSES OF DEATH	25.0	13.0000	7.359801	1.00	7.00	13.0	19.00	25.00
Rate	25.0	27.3976	31.340038	6.69	10.94	14.6	25.49	140.72
World Rank	25.0	70.7200	47.609978	1.00	24.00	77.0	113.00	147.00

In [41]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Compute the summary statistics and transpose the result
summary_stats = first_table_filled.describe().T

# Create a heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(summary_stats, annot=True, cmap="coolwarm", fmt=".2f")

# Add titles and labels
plt.title("Heatmap of Summary Statistics", fontsize=16)
plt.xlabel("Statistics")
plt.ylabel("Columns")

# Show the heatmap
plt.show()
```



In [49]:
```
plt.barh(first_table_filled['TOP 50 CAUSES OF DEATH'],first_table_filled["Rate"])
```

Out[49]:
```
<BarContainer object of 25 artists>
```



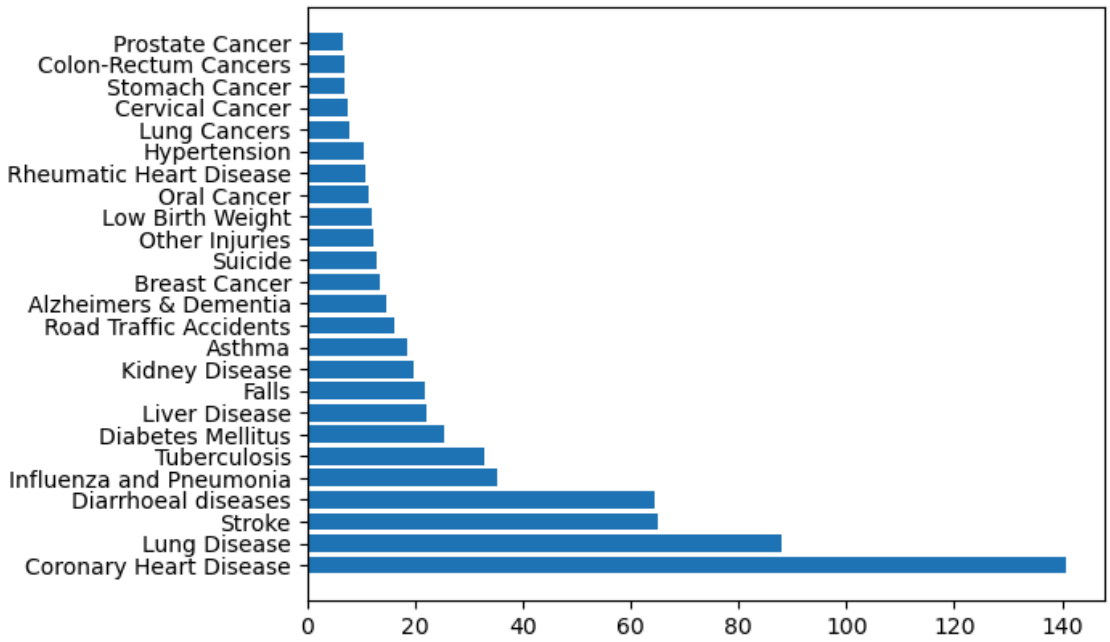
In [47]:
```
print(first_table_filled.columns)
```

```
Index(['TOP 50 CAUSES OF DEATH', 'TOP 50 CAUSES OF DEATH.1', 'Rate',
       'World Rank'],
      dtype='object')
```

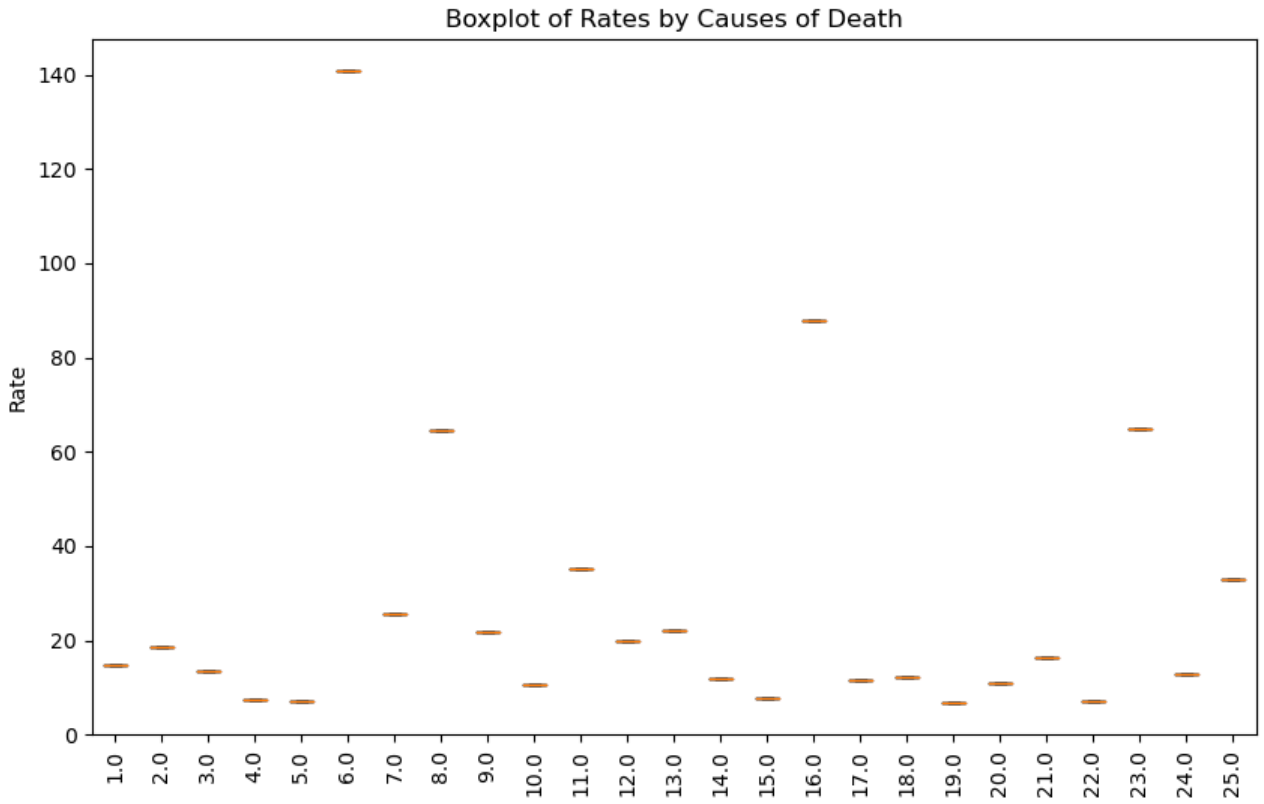
In [ ]:
```

```

In [58]:
```
plt.barh(first_table_filled["TOP 50 CAUSES OF DEATH.1"],first_table_filled["Rate"])
```

Out[58]:
```
<BarContainer object of 25 artists>
```



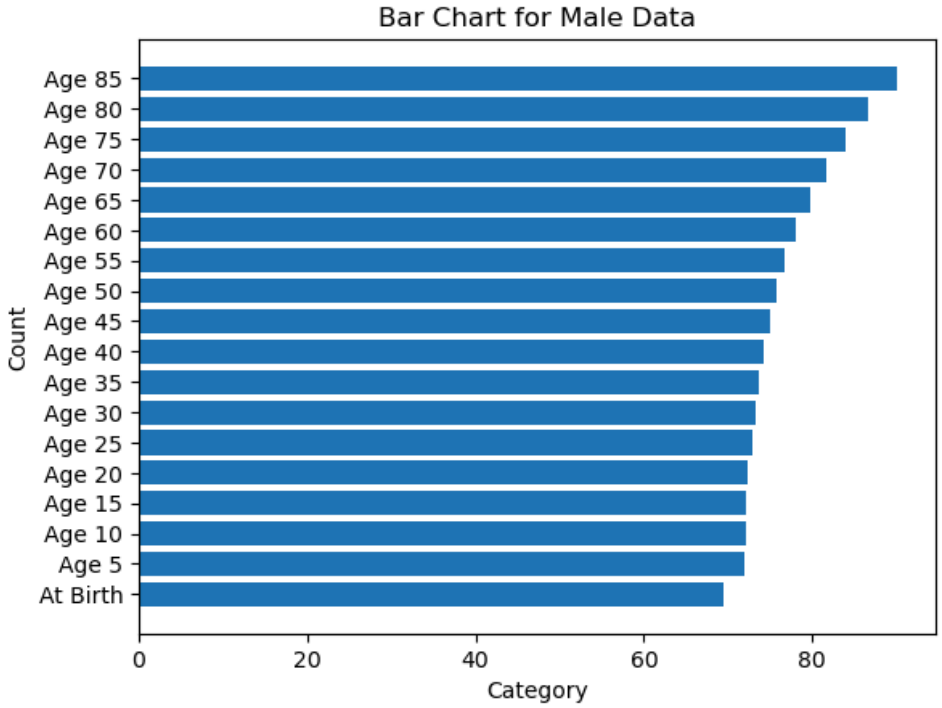
In [64]:
```
import matplotlib.pyplot as plt

# Group the data
grouped_data = [group["Rate"].values for _, group in first_table_filled.groupby("TOP 50 CAUSES OF DEATH.

# Create the boxplot
plt.figure(figsize=(10, 6))
plt.boxplot(grouped_data, labels=first_table_filled["TOP 50 CAUSES OF DEATH"].unique(), vert=True)
plt.xticks(rotation=90)  # Rotate labels for better readability
plt.ylabel("Rate")
plt.title("Boxplot of Rates by Causes of Death")
plt.show()
```



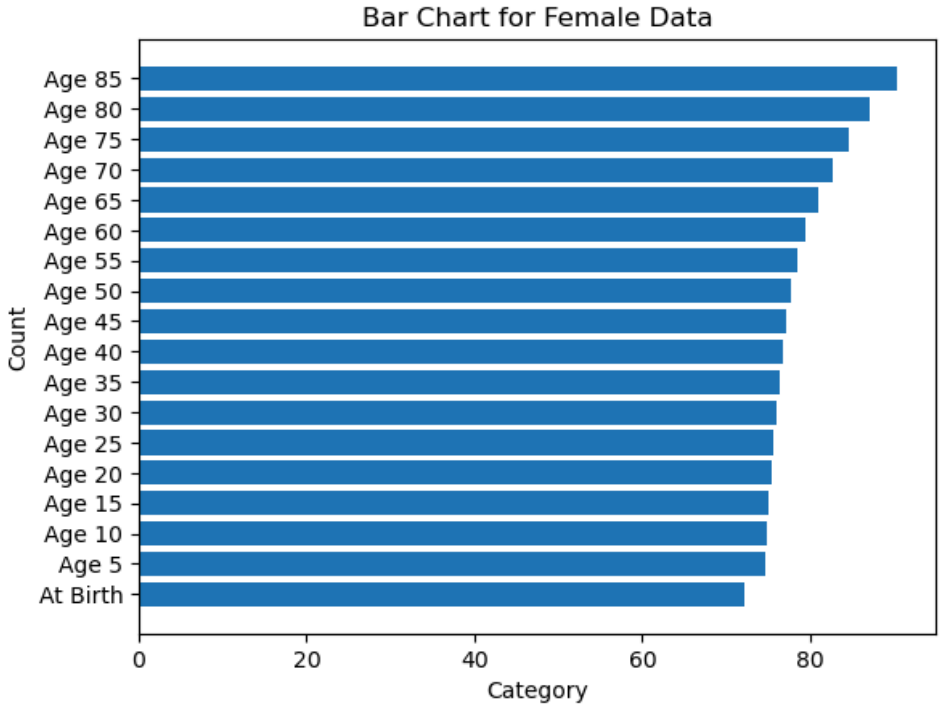
In [68]:
```
print(First_Table.columns)
```

```
MultiIndex([('Unnamed: 0_level_0', 'Unnamed: 0_level_1'),
            ('Unnamed: 1_level_0',              'Male'),
            ('Unnamed: 2_level_0',            'Female'),
            (        'World Rank',                 'M'),
            (        'World Rank',                 'F')],
           )
```

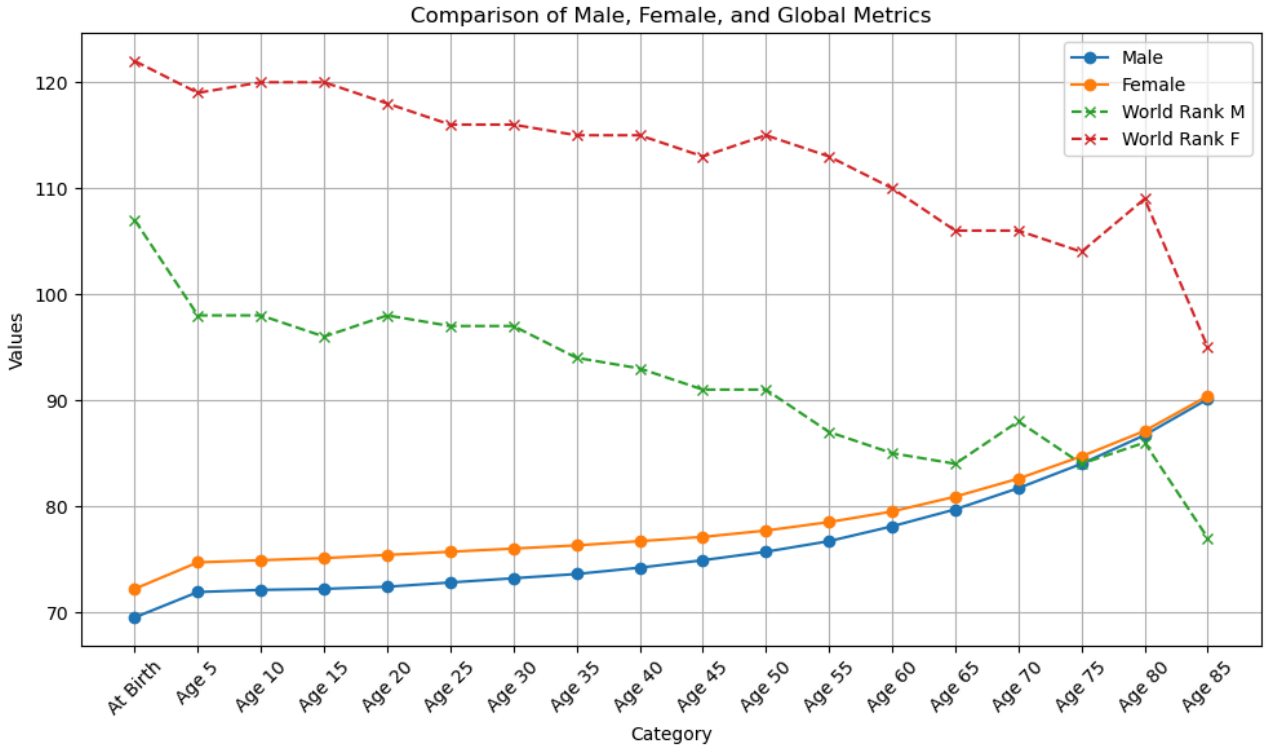
In [73]:
```
plt.barh(First_Table[('Unnamed: 0_level_0', 'Unnamed: 0_level_1')],
        First_Table[('Unnamed: 1_level_0', 'Male')])
plt.xlabel('Category')  # Replace with appropriate label
plt.ylabel('Count')     # Replace with appropriate label
plt.title('Bar Chart for Male Data')
plt.show()
```



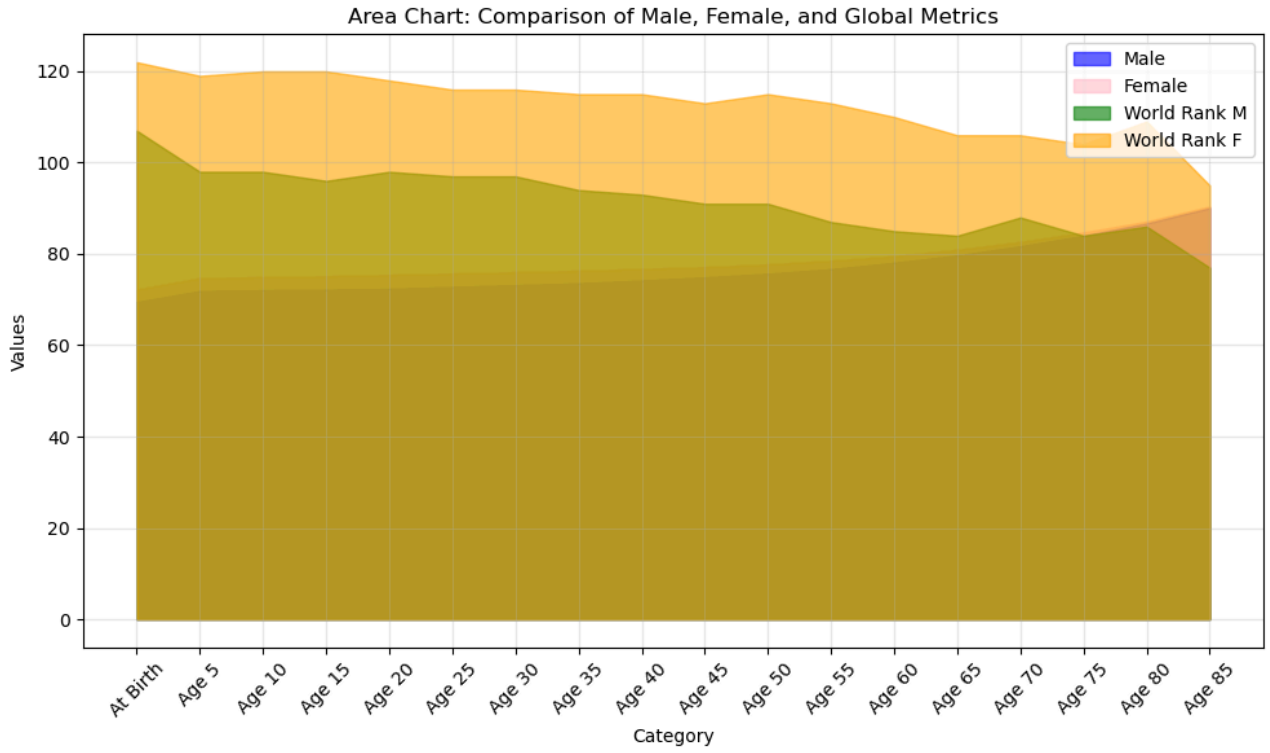
In [75]:
```
plt.barh(First_Table[('Unnamed: 0_level_0', 'Unnamed: 0_level_1')],
        First_Table[('Unnamed: 2_level_0', 'Female')])
plt.xlabel('Category')  # Replace with appropriate label
plt.ylabel('Count')     # Replace with appropriate label
plt.title('Bar Chart for Female Data')
plt.show()
```



In [76]:
```
import matplotlib.pyplot as plt

# Extracting columns
categories = First_Table[('Unnamed: 0_level_0', 'Unnamed: 0_level_1')]  # X-axis labels
male_values = First_Table[('Unnamed: 1_level_0', 'Male')]
female_values = First_Table[('Unnamed: 2_level_0', 'Female')]
world_rank_m = First_Table[('World Rank', 'M')]
world_rank_f = First_Table[('World Rank', 'F')]

# Plotting
plt.figure(figsize=(10, 6))
plt.plot(categories, male_values, label='Male', marker='o')
plt.plot(categories, female_values, label='Female', marker='o')
plt.plot(categories, world_rank_m, label='World Rank M', linestyle='--', marker='x')
plt.plot(categories, world_rank_f, label='World Rank F', linestyle='--', marker='x')

# Adding labels and title
plt.xlabel('Category')
plt.ylabel('Values')
plt.title('Comparison of Male, Female, and Global Metrics')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True)

# Display the chart
plt.tight_layout()
plt.show()
```



In [77]:
```
import matplotlib.pyplot as plt

# Extracting columns
categories = First_Table[('Unnamed: 0_level_0', 'Unnamed: 0_level_1')]  # X-axis labels
male_values = First_Table[('Unnamed: 1_level_0', 'Male')]
female_values = First_Table[('Unnamed: 2_level_0', 'Female')]
world_rank_m = First_Table[('World Rank', 'M')]
world_rank_f = First_Table[('World Rank', 'F')]

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

# Area for Male
plt.fill_between(categories, male_values, label='Male', alpha=0.6, color='blue')
# Area for Female
plt.fill_between(categories, female_values, label='Female', alpha=0.6, color='pink')
# Area for World Rank M
plt.fill_between(categories, world_rank_m, label='World Rank M', alpha=0.6, color='green')
# Area for World Rank F
plt.fill_between(categories, world_rank_f, label='World Rank F', alpha=0.6, color='orange')

# Adding labels and title
plt.xlabel('Category')
plt.ylabel('Values')
plt.title('Area Chart: Comparison of Male, Female, and Global Metrics')
plt.legend()
plt.xticks(rotation=45)
plt.grid(True, alpha=0.3)

# Display the chart
plt.tight_layout()
plt.show()
```



In [ ]: