sdv-project-g4

April 25, 2024

##python

```
[1]: #importing the nessasary libraries
     import pandas as pd
     # Load the first dataset (Population Dynamics)
     ct_ds_url = "https://raw.githubusercontent.com/viteshc/SDV_5320_Project_Group-4/
      ⊖main/1"
     ct_ds = pd.read_csv(ct_ds_url)
     # Load the second dataset (Health Systems)
     hs ds url = "https://raw.githubusercontent.com/viteshc/SDV 5320 Project Group-4/
     ⇔main/2"
     hs_ds = pd.read_csv(hs_ds_url)
     # Load the third dataset (Gender Equality)
     gi_ds_url = "https://raw.githubusercontent.com/viteshc/SDV_5320_Project_Group-4/
      ⇒main/3"
     gi_ds = pd.read_csv(gi_ds_url)
[2]: #printing the columns of the dataset
     print("Columns in Population Dynamics Dataset:")
     print(ct_ds.columns)
     print("\nColumns in Health Systems Dataset:")
     print(hs_ds.columns)
     print("\nColumns in Gender Inequality Dataset:")
     print(gi_ds.columns)
    Columns in Population Dynamics Dataset:
    Index(['country', 'rank', 'area', 'landAreaKm', 'cca2', 'cca3', 'netChange',
           'growthRate', 'worldPercentage', 'density', 'densityMi', 'place',
           'pop1980', 'pop2000', 'pop2010', 'pop2022', 'pop2023', 'pop2030',
           'pop2050'],
          dtype='object')
    Columns in Health Systems Dataset:
```

```
Index(['Country_Region', 'Province_State', 'World_Bank_Name',
           'Health_exp_pct_GDP_2016', 'Health_exp_public_pct_2016',
           'Health_exp_out_of_pocket_pct_2016', 'Health_exp_per_capita_USD_2016',
           'per_capita_exp_PPP_2016', 'External_health_exp_pct_2016',
           'Physicians per 1000 2009-18', 'Nurse midwife per 1000 2009-18',
           'Specialist_surgical_per_1000_2008-18',
           'Completeness of birth reg 2009-18',
           'Completeness_of_death_reg_2008-16'],
          dtype='object')
    Columns in Gender Inequality Dataset:
    Index(['Country', 'Human development', 'GII', 'Rank', 'Maternal mortality',
           'Adolescent_birth_rate', 'Seats_parliament', 'F_secondary_educ',
           'M_secondary_educ', 'F_Labour_force', 'M_Labour_force'],
          dtype='object')
[3]: # Display null values for each dataset
     print("Null values in Population Dynamics Dataset:")
     print(ct_ds.isnull().sum())
     print("\nNull values in Health Systems Dataset:")
     print(hs_ds.isnull().sum())
     print("\nNull values in Gender Inequality Dataset:")
     print(gi_ds.isnull().sum())
    Null values in Population Dynamics Dataset:
    country
                       0
    rank
    area
                       0
    landAreaKm
                       0
    cca2
                        1
    cca3
                       0
                       8
    netChange
    growthRate
                       0
    worldPercentage
    density
                       0
    densityMi
                       0
    place
                       0
    pop1980
                       0
                       0
    pop2000
                       0
    pop2010
    pop2022
                       0
    pop2023
                       0
    pop2030
                       0
    pop2050
                       0
    dtype: int64
```

```
Country_Region
                                              23
    Province_State
                                              196
    World_Bank_Name
                                               0
    Health exp pct GDP 2016
                                              24
    Health_exp_public_pct_2016
                                              24
    Health_exp_out_of_pocket_pct_2016
                                              24
    Health_exp_per_capita_USD_2016
                                              24
    per_capita_exp_PPP_2016
                                              24
    External_health_exp_pct_2016
                                              43
    Physicians_per_1000_2009-18
                                              21
    Nurse_midwife_per_1000_2009-18
                                              21
    Specialist_surgical_per_1000_2008-18
                                              35
    Completeness_of_birth_reg_2009-18
                                              47
    Completeness_of_death_reg_2008-16
                                             103
    dtype: int64
    Null values in Gender Inequality Dataset:
    Country
                               0
    Human development
                               4
    GII
                              25
    Rank
                              25
    Maternal_mortality
                              11
    Adolescent_birth_rate
                               0
    Seats_parliament
                               2
    F_secondary_educ
                              18
    M_secondary_educ
                              18
    F_Labour_force
                              15
    M_Labour_force
                              15
    dtype: int64
[4]: # Remove null values from Population Dynamics Dataset
     ct_ds_clean = ct_ds.dropna()
     # Remove null values from Health Systems Dataset
     hs_ds_clean = hs_ds
     # Remove null values from Gender Inequality Dataset
     gi_ds_clean = gi_ds.dropna()
[5]: ct_ds_clean.isnull().sum()
[5]: country
                        0
                        0
     rank
                        0
     area
                        0
     landAreaKm
     cca2
                        0
```

Null values in Health Systems Dataset:

```
0
     netChange
     growthRate
                         0
     worldPercentage
     density
     densityMi
                         0
     place
                         0
                         0
     pop1980
                         0
    pop2000
     pop2010
                         0
     pop2022
                         0
     pop2023
                         0
     pop2030
                         0
                         0
     pop2050
     dtype: int64
[6]: hs_ds_clean.isnull().sum()
[6]: Country_Region
                                                23
     Province State
                                               196
     World_Bank_Name
                                                 0
     Health_exp_pct_GDP_2016
                                                24
     Health_exp_public_pct_2016
                                                24
     Health_exp_out_of_pocket_pct_2016
                                                24
     Health_exp_per_capita_USD_2016
                                                24
     per_capita_exp_PPP_2016
                                                24
     External_health_exp_pct_2016
                                                43
     Physicians_per_1000_2009-18
                                                21
     Nurse_midwife_per_1000_2009-18
                                                21
     Specialist_surgical_per_1000_2008-18
                                                35
     Completeness_of_birth_reg_2009-18
                                                47
     Completeness_of_death_reg_2008-16
                                               103
     dtype: int64
[7]: gi_ds_clean.isnull().sum()
[7]: Country
                               0
     Human_development
                               0
     GII
                               0
     Rank
                               0
     Maternal_mortality
                               0
     Adolescent_birth_rate
                               0
     Seats_parliament
                               0
     F_secondary_educ
                               0
     M_secondary_educ
                               0
     F_Labour_force
                               0
     M_Labour_force
                               0
```

cca3

0

dtype: int64

```
[8]: # Drop unused columns
     ct_ds_clean.drop(['rank', 'area', 'landAreaKm', 'cca2', 'cca3', 'pop1980', _

¬'pop2000', 'pop2010', 'pop2022', 'pop2030', 'pop2050'], axis=1, inplace=True)

     <ipython-input-8-506ae55fc7f7>:2: SettingWithCopyWarning:
     A value is trying to be set on a copy of a slice from a DataFrame
     See the caveats in the documentation: https://pandas.pydata.org/pandas-
     docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
       ct_ds_clean.drop(['rank', 'area', 'landAreaKm', 'cca2', 'cca3', 'pop1980',
     'pop2000', 'pop2010', 'pop2022', 'pop2030', 'pop2050'], axis=1, inplace=True)
 [9]: # Drop unused columns
     hs ds clean.drop(['Province State', 'World Bank Name', |
       → 'Health_exp_out_of_pocket_pct_2016', 'Health_exp_per_capita_USD_2016', '

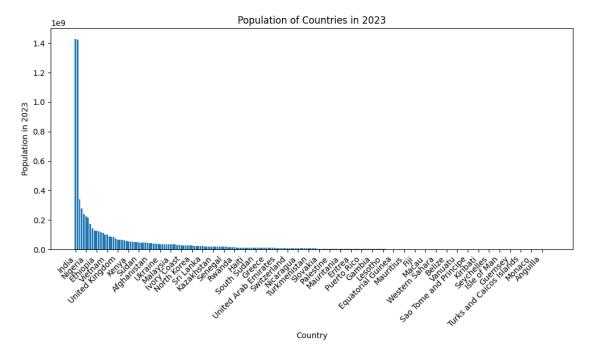
¬'Nurse_midwife_per_1000_2009-18', 'Specialist_surgical_per_1000_2008-18'],

       ⇔axis=1, inplace=True)
[10]: # Drop unused columns
     gi_ds_clean.drop(['Human_development', 'Rank'], axis=1, inplace=True)
     <ipython-input-10-1a275f5b5126>:2: SettingWithCopyWarning:
     A value is trying to be set on a copy of a slice from a DataFrame
     See the caveats in the documentation: https://pandas.pydata.org/pandas-
     docs/stable/user guide/indexing.html#returning-a-view-versus-a-copy
       gi_ds_clean.drop(['Human_development', 'Rank'], axis=1, inplace=True)
[11]: # Display all columns of cleaned datasets
     print("Columns in Population Dynamics Dataset:")
     print(ct_ds_clean.columns)
     print("\nColumns in Health Systems Dataset:")
     print(hs_ds_clean.columns)
     print("\nColumns in Gender Inequality Dataset:")
     print(gi_ds_clean.columns)
     Columns in Population Dynamics Dataset:
     Index(['country', 'netChange', 'growthRate', 'worldPercentage', 'density',
            'densityMi', 'place', 'pop2023'],
           dtype='object')
     Columns in Health Systems Dataset:
     Index(['Country_Region', 'Health_exp_pct_GDP_2016',
```

```
'Health_exp_public_pct_2016', 'Physicians_per_1000_2009-18',
            'Completeness_of_birth_reg_2009-18',
            'Completeness_of_death_reg_2008-16'],
           dtype='object')
     Columns in Gender Inequality Dataset:
     Index(['Country', 'GII', 'Maternal mortality', 'Adolescent birth rate',
            'Seats_parliament', 'F_secondary_educ', 'M_secondary_educ',
            'F_Labour_force', 'M_Labour_force'],
           dtype='object')
[12]: #Saving cleaned datasets
      ct_ds_clean.to_csv("population_dynamics_clean.csv", index=False)
      hs_ds_clean.to_csv("health_systems_clean.csv", index=False)
      gi_ds_clean.to_csv("gender_inequality_clean.csv", index=False)
[13]: import pandas as pd
      import matplotlib.pyplot as plt
      # Load cleaned datasets
      ct_ds_clean = pd.read_csv("population_dynamics_clean.csv")
      hs_ds_clean = pd.read_csv("health_systems_clean.csv")
      gi_ds_clean = pd.read_csv("gender_inequality_clean.csv")
[14]: # Print the first few rows of Population Dynamics dataset
      print("Population Dynamics Dataset:")
      print(ct_ds_clean.head())
      # Print the first few rows of Health Systems dataset
      print("\nHealth Systems Dataset:")
      print(hs_ds_clean.head())
      # Print the first few rows of Gender Inequality dataset
      print("\nGender Inequality Dataset:")
      print(gi_ds_clean.head())
     Population Dynamics Dataset:
              country netChange growthRate worldPercentage
                                                                density densityMi \
     0
                India
                          0.4184
                                      0.0081
                                                       0.1785 480.5033 1244.5036
                        -0.0113
                                                       0.1781 151.2696 391.7884
     1
                China
                                     -0.0002
     2 United States
                          0.0581
                                                       0.0425
                                                                         96.2666
                                      0.0050
                                                                37.1686
                                                       0.0347 147.8196
     3
            Indonesia
                         0.0727
                                      0.0074
                                                                         382.8528
             Pakistan
                        0.1495
                                      0.0198
                                                       0.0300 311.9625
                                                                         807.9829
        place
                  pop2023
     0
          356 1428627663
     1
          156 1425671352
     2
          840
                339996563
```

```
3
     360
           277534122
     586
           240485658
Health Systems Dataset:
  Country_Region Health_exp_pct_GDP_2016 Health_exp_public_pct_2016 \
0
     Afghanistan
                                      10.2
                                                                    5.1
                                       6.7
                                                                   41.4
         Albania
1
                                                                   67.7
         Algeria
                                       6.6
3
         Andorra
                                      10.4
                                                                   49.1
4
          Angola
                                       2.9
                                                                   44.1
   Physicians_per_1000_2009-18 Completeness_of_birth_reg_2009-18 \
0
                            0.3
                                                               42.3
1
                            1.2
                                                               98.4
2
                            1.8
                                                              100.0
3
                            3.3
                                                              100.0
4
                            0.2
                                                               25.0
   Completeness_of_death_reg_2008-16
0
                                  NaN
                                 53.0
1
2
                                  NaN
                                 80.0
3
4
                                  NaN
Gender Inequality Dataset:
                  GII Maternal_mortality Adolescent_birth_rate \
       Country
                                       5.0
                                                               2.2
0
   Switzerland 0.018
                                       2.0
                                                               2.3
1
               0.016
        Norway
2
       Iceland 0.043
                                       4.0
                                                               5.4
3
     Australia 0.073
                                       6.0
                                                               8.1
       Denmark 0.013
                                       4.0
                                                               1.9
   Seats_parliament F_secondary_educ M_secondary_educ F_Labour_force \
0
               39.8
                                  96.9
                                                     97.5
                                                                     61.7
               45.0
                                  99.1
                                                     99.3
1
                                                                     60.3
2
               47.6
                                  99.8
                                                     99.7
                                                                     61.7
                                  94.6
3
               37.9
                                                     94.4
                                                                     61.1
4
               39.7
                                  95.1
                                                     95.2
                                                                     57.7
   M_Labour_force
0
             72.7
1
             72.0
2
             70.5
3
             70.5
4
             66.7
```

```
[15]: # Visualize Population Dynamics with adjusted xticks
plt.figure(figsize=(10, 6))
plt.bar(ct_ds_clean['country'], ct_ds_clean['pop2023'])
plt.xlabel('Country')
plt.ylabel('Population in 2023')
plt.title('Population of Countries in 2023')
# Rotate labels for better readability
plt.xticks(rotation=45, ha='right')
# Display every 5th label
plt.gca().set_xticks(plt.gca().get_xticks()[::5])
# Adjust layout to prevent clipping
plt.tight_layout()
plt.show()
```



below: Different stakeholders, including policy makers, economists and social scientists, would find it useful to know which countries are projected to have the largest populations in the foreseeable future since global demographics are constantly evolving. This code furnishes a snapshot of demographic trends that shape our world.

When we look at the horizontal bar plot, it is clear that some nations dominate for their large population. On top of these populous nations include China, India and United States having populations running into billions or millions. These countries have always been recognized for their massive influence on global and regional demography.

However, besides regular giants, other states with large populations like Indonesia Pakistan and Brazil can also be seen on this graph that calls one's attention due to its huge demographic weight. These countries belong to different regions and carry implications for such aspects of global affairs

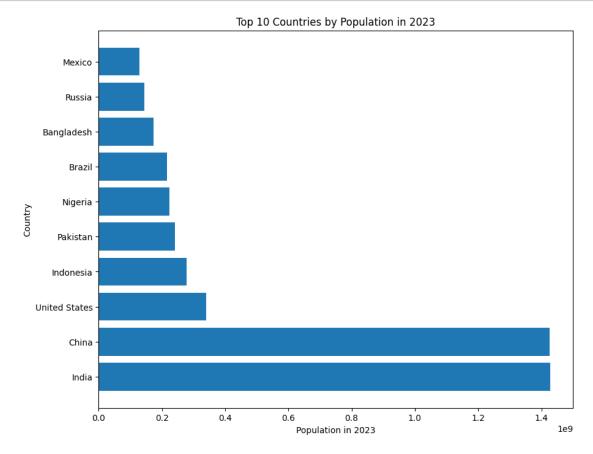
as economic growth, social development and environmental sustainability.

Therefore for policymakers as well as other stakeholders looking at this visualization becomes crucial since they will be able to get prepared for any upcoming challenges while benefiting from opportunities that may arise out of these demographic changes.

```
[16]: # Sort the dataset by population in descending order
    ct_ds_sorted = ct_ds_clean.sort_values(by='pop2023', ascending=False)

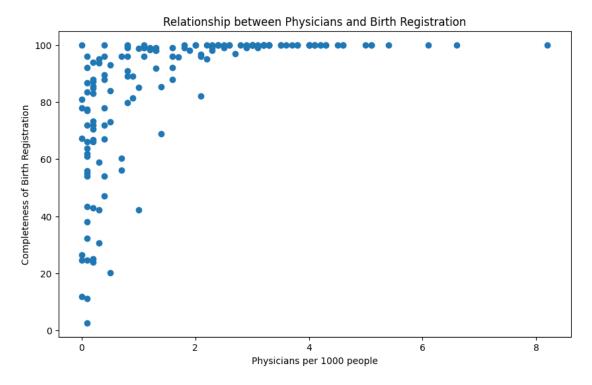
# Select the top 10 countries with the highest population
    top_10_countries = ct_ds_sorted.head(10)

# Visualize the top 10 countries with a horizontal bar plot
    plt.figure(figsize=(10, 8))
    plt.barh(top_10_countries['country'], top_10_countries['pop2023'])
    plt.xlabel('Population in 2023')
    plt.ylabel('Country')
    plt.title('Top 10 Countries by Population in 2023')
    plt.show()
```

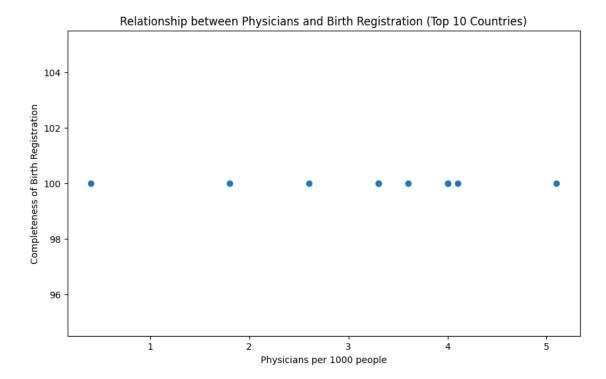


below: The diagram above is used to represent how the number of doctors per 1000 people relates with birth registration completeness. One point in the scatter plot shows one country and

its x-coordinate is the number of doctors per thousand people while y-coordinate represents the percentage birth registration complete.







below: The visualization is about Gender Inequality Index (GII) among the top ten most gender unequal countries. GII is a measure that combines measures for gender gaps in health, education, and economic participation.

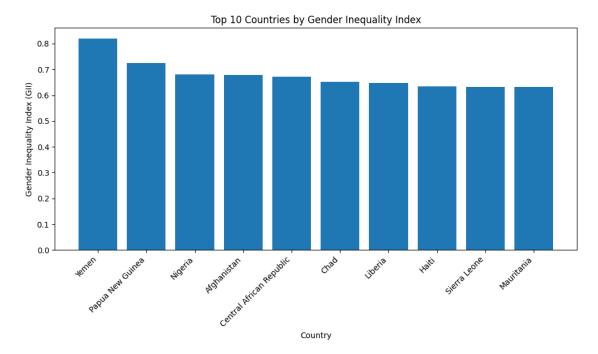
Looking at the chart we observe that higher GII values indicate more inequality between genders in the countries. Those are larger bars representing countries where inequality between women and men is greater across various elements of life such as educational access, medical care and employment.

For example, Yemen has the highest GII among the top 10 countries indicating considerable gender inequalities in areas like maternal mortality, adolescent birth rates, educational attainment and labour force morbidity. Others like Afghanistan, Niger and Chad also reveal high levels of gender inequality by their respective GII values.

This visualization underlines that addressing the question of gender inequality as one of most important aspects for social and economic development. It demonstrates specific interventions requiring focused approaches towards ensuring gender equality and empowering girls as well as women globally that will eventually create to an all-inclusive society-both aspect for human rights indeed.

```
[19]: # Sort the DataFrame by GII and select the top 10 countries top_10_gii = gi_ds_clean.sort_values(by='GII', ascending=False).head(10) # Visualize Gender Inequality with adjusted xticks for top 10 countries
```

```
plt.figure(figsize=(10, 6))
plt.bar(top_10_gii['Country'], top_10_gii['GII'])
plt.xlabel('Country')
plt.ylabel('Gender Inequality Index (GII)')
plt.title('Top 10 Countries by Gender Inequality Index')
plt.xticks(rotation=45, ha='right') # Rotate labels for better readability
plt.tight_layout() # Adjust layout to prevent clipping
plt.show()
```



below: Population dynamics: A visualization of the top ten countries by population in 2023 lays bare the demographic landscape of the most populous nations on earth. Leading this list are giants such as India and China, whose huge populations have global implications. These countries, like United States, Indonesia and Pakistan among others, show different demographics, that is; growth rates and population densities. It is clear from these population dynamics that meeting the needs of rapidly growing populations while ensuring sustainable development will be a critical issue for years to come.

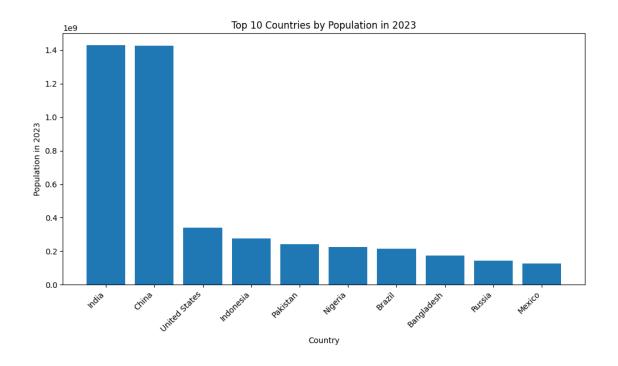
Health systems: An examination of physician distribution per 1000 persons reveals disparities in terms of access to healthcare and resources across countries. While some countries have a good physician-population ratio, others face challenges in their healthcare infrastructure's limitedness or workforce shortage. This difference shows how important it is to strengthen health systems so that every person can get quality healthcare. Also, looking at indicators like health expenditures as percentage of GDP points out the extent to which different countries invest in health care and therefore where interventions could improve health status through better delivery globally.

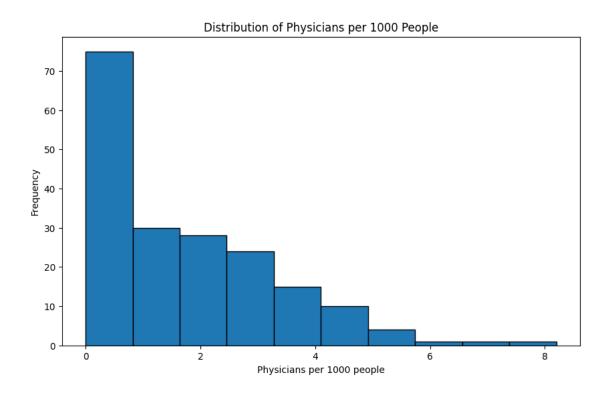
Gender Inequality: When it comes to gender inequality and maternal health, a scatter plot that shows the correlation between maternal mortality ratios and adolescent birth rates would be the

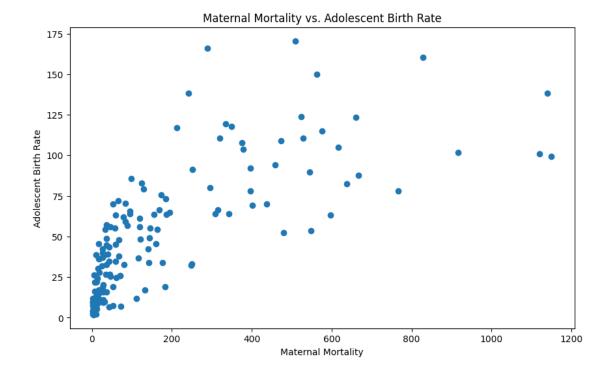
best idea. It is observed that countries with higher numbers of maternal deaths tend to have elevated levels of teenage pregnancy which implies a lack of reproductive health services and education. Similarly, the Gender Inequality Index (GII) draws attention to the fact that these disparities are systemic and not limited to one sector such as education, political representation or employment. Consequently, mitigating such disproportions is indispensable towards attaining global gender parity and promoting all-embracing societies.

```
[20]: import pandas as pd
      import matplotlib.pyplot as plt
      # Load cleaned datasets
      population_df = pd.read_csv("population_dynamics_clean.csv")
      health_systems_df = pd.read_csv("health_systems_clean.csv")
      gender_inequality_df = pd.read_csv("gender_inequality_clean.csv")
      # Example 1: Population Dynamics Visualization
      plt.figure(figsize=(10, 6))
      top_countries_population = population_df.nlargest(10, 'pop2023')
      plt.bar(top_countries_population['country'],__
       →top_countries_population['pop2023'])
      plt.xlabel('Country')
      plt.ylabel('Population in 2023')
      plt.title('Top 10 Countries by Population in 2023')
      plt.xticks(rotation=45, ha='right')
      plt.tight_layout()
      plt.show()
      # Example 2: Health Systems Visualization
      plt.figure(figsize=(10, 6))
      plt.hist(health_systems_df['Physicians_per_1000_2009-18'], bins=10,__
       ⇔edgecolor='black')
      plt.xlabel('Physicians per 1000 people')
      plt.ylabel('Frequency')
      plt.title('Distribution of Physicians per 1000 People')
      plt.show()
      # Example 3: Gender Inequality Visualization
      plt.figure(figsize=(10, 6))
      plt.scatter(gender_inequality_df['Maternal_mortality'],__

→gender_inequality_df['Adolescent_birth_rate'])
      plt.xlabel('Maternal Mortality')
      plt.ylabel('Adolescent Birth Rate')
      plt.title('Maternal Mortality vs. Adolescent Birth Rate')
      plt.show()
```







below: In a world of opposites, where ups and downs on progress meet with dissimilarities that are hard to bear, the Population Health Heatmap and Gender Inequality Index (GII) Map give windows into global dynamics portrayed in their intricate tapestry.

Part 1: Population Health Heatmap

As we start our journey through the Population Health Heatmap, the landscape unravels before us in vibrant hues, each shade reflecting the density of human life. From crowded Asian cities to African plains stretching out as far as the eye can see, it tells us about people over there beating by billions.

At times within bustling cities and other times in rural areas full of tranquility, gaps of inequality emerge. Areas with high population densities show strains in healthcare systems that expose the need for equal access to medical services and resources. Still from those difficulties there is light at the end of tunnel as communities rise up to address health issues that affect their people towards healthier and stronger future full of unity.

Part 2: Gender Inequality Index Map

When it comes to the Gender Inequality Index Map, things change a bit with colors giving way to another form of human life story. This is because different shades of inequity inform us about women's issues that have been influenced by cultural norms and societal expectations.

Where women voices are raised high and their rights are protected; on the other hand, the map shows softer colors denoting an advance toward gender parity. On the contrary, in regions where there is systemic discrimination and deeply rooted gender stereotyping, this landscape has darker shades of inequality that obstruct progress for women and girls.

Nevertheless, some embers of resilience are ignited from within as grass root movements together with advocacy groups strive to remove obstacles that hinder equal treatment based on gender. It is all about educating, empowering and advocating; hence forging paths towards a more inclusive and just society as regards sex inequality narrative in communities.

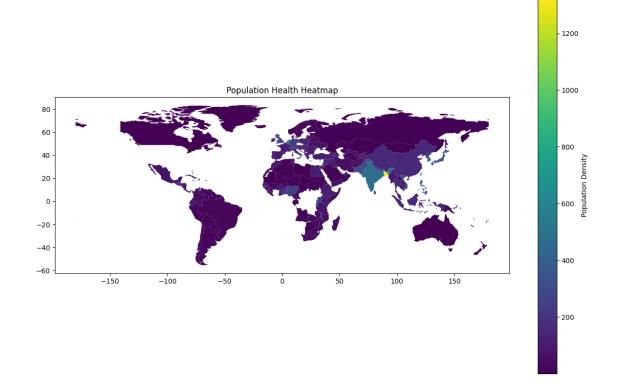
```
[21]: import geopandas as gpd
      import matplotlib.pyplot as plt
      # Load shapefile for country boundaries
      world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))
      # Load population density data
      population_density = pd.read_csv("/content/population_dynamics_clean.csv") #__
       →Replace with your data file
      # Merge shapefile with population density data
      world = world.merge(population_density, how='left', left_on='name',_

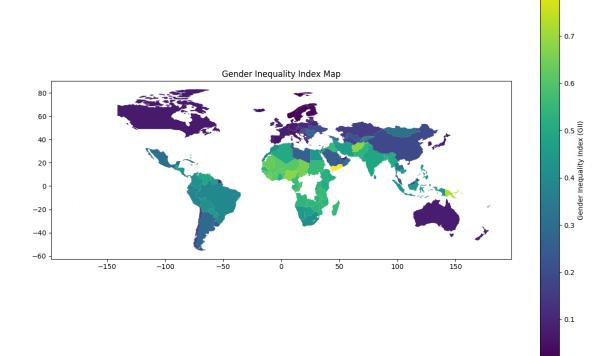
¬right_on='country')
      # Plot Population Health Heatmap
      fig, ax = plt.subplots(1, 1, figsize=(15, 10))
      world.plot(column='density', ax=ax, legend=True,
                 legend_kwds={'label': "Population Density", 'orientation':

¬"vertical"})
      ax.set title('Population Health Heatmap')
      plt.show()
      # Load Gender Inequality Index (GII) data
      gender_inequality = pd.read_csv("gender_inequality_clean.csv") # Replace with_
       →your data file
      # Merge shapefile with GII data
      world = world.merge(gender_inequality, how='left', left_on='name',_
       →right_on='Country')
      # Plot Gender Inequality Index Map
      fig, ax = plt.subplots(1, 1, figsize=(15, 10))
      world.plot(column='GII', ax=ax, legend=True,
                 legend_kwds={'label': "Gender Inequality Index (GII)", 'orientation':
       → "vertical"})
      ax.set_title('Gender Inequality Index Map')
      plt.show()
```

<ipython-input-21-c3699117767e>:5: FutureWarning: The geopandas.dataset module
is deprecated and will be removed in GeoPandas 1.0. You can get the original
'naturalearth_lowres' data from
https://www.naturalearthdata.com/downloads/110m-cultural-vectors/.

world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))





0.8

Basically in this world the economic growth progress is generally measured its progress by considering the gender inequality of the population and also based on the social justice and the challenges are generally based on the world gender equality of the population.

Next colors on the map are more vibrant in which it gives the whole information of the hue easily. Next the gender parity of the scadinavia tales are in the soft blues. This, in turn, has led to situations where both men and women grow and prosper.

And yet not all areas are benefiting from this progress. Across great lengths of land in Africa or Asia it is a tragic fact that the color red is used to represent inequality. Women here experience lack of access to education, health services and economic opportunities which has continued throughout generations owing to traditional customs and circumstances.

In such disparities however, there is still a ray of hope. In certain places, women have emerged as leaders, entrepreneurs or catalysts for change in a time when overall conditions were bleak. Rwanda and Bangladesh were previously characterized by violence and destitution but now they epitomize progress with their path towards achieving gender equality lighting the way for others.

Navigating the contours of GII map shows that gender inequality has no boundaries. It thrives in the silence of norms in society and indifference as well. However, it also fades away before commitment and unity.

We are one step closer to a world where people's fate is not determined by gender for every stroke painted on the canvas of progress. This GII map is both a mirror reflecting our journey together and a compass steering us into an age where equality will eventually prevail.

In this global tapestry of humanity, let us commit ourselves to creating a future that radiates with all colors, speaks through every voice, and enables any person to reach heights that were never dreamt possible due to gender identity.

```
legend_kwds={'label': "Gender Inequality Index (GII)", 'orientation':
    "vertical"})
ax.set_title('Gender Inequality Index Map')
plt.show()
```

<ipython-input-22-d1eb86f90135>:5: FutureWarning: The geopandas.dataset module
is deprecated and will be removed in GeoPandas 1.0. You can get the original
'naturalearth_lowres' data from

https://www.naturalearthdata.com/downloads/110m-cultural-vectors/.
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

