TU Dortmund

Introductory Case Studies

Project 1: Descriptive analysis of demographic data

Lecturers:
Prof. Dr. Katja Ickstadt
M. Sc. Zeyu Ding
M. Sc. Yassine Talleb

Author: Mohammad Sakhawat Hossain

Matriculation No: 231838

Group number: 5

Group members: Md Shahabub Alam, Shahed Iqbal Chowdhury, Hasan Zamil Ahmed, Sazeda Sultana

May 12, 2023

Contents

1 Introduction

Demographic data refers to the statistical information collected about some populations attributes including factors like age, sex, income, education, race, and ethnicity. To examine social, economic, and political matters, this data plays a very important role. Evolving various characteristics over the time period can be evaluated by demographic data. Comparison between different countries, regions or sexes can also be determined by demographic data. This information can be helpful in several fields, like economics, social systems, healthcare, and education (Lundquist et al., 2014, p. 02).

The goal of this study is to examine the demographic data of 227 countries consisting of five regions and 21 subregions, including information on life expectancy of male, female, both sexes and under age 5 mortality of male, female, both sexes for the years 2002 and 2022. Initially, frequency distributions and measures of central tendency are used to assess each variable. Subsequently, measures of variability are utilized to evaluate the differences among regions and subregions. Here, one region is considered only. Correlation coefficients are then applied to ascertain the relationship between two different variables. Finally, a comparative analysis is conducted between the datasets of 2002 and 2022.

Section 2 gives an overview of the given data set and its structure is presented precisely. Here, explanations of all the variables and their characteristics are given in detail. Section 3 explains the statistical methodology used to analyze the dataset. Section 4 presents and explains the result of the analysis. Section 5 contains a summary of the findings and an outlook for possible further analysis.

2 Problem statement

2.1 Dataset description and data quality

The dataset used in this study is provided by the instructors of the "Introductory Case Studies" course at TU Dortmund during the summer of 2023. The data is collected from the International Database (IDB) of the United States Census Bureau which comprises demographic information from over 200 countries with populations of 5,000 or above. Censuses, surveys, administrative records, and estimated and projected figures are the main resources of this dataset (U.S Census Bureau, 2022).

The dataset includes information on U5 mortality under the age of five and life expectancy at birth for males, females and both sexes in the years 2002 and 2022. It has a total of 227 countries grouped into 5 regions and then again divided into 21 subregions. Overall, the dataset contains 454 observations and 11 variables, including index (refer to indexing), country, region, subregion, year, Under age 5 mortality for male, female, and both sexes, as well as life expectancy for male, female, and both sexes. The mortality of under age 5 and life expectancy variables are numerical, while country, region, and subregion are nominal variables. All quantitative variables consist of positive integers or decimal numbers represented with 1 or 2 decimal places. The three variables named under 5 mortality rate of male, female and both sexes are renamed and used in the analysis as 'U5 morality of male', 'U5 mortality of females' and 'U5 mortality of both sexes'. The data set has two distinct years: 2002 and 2022. Here, Country refers to the name of a country that falls under the subregion and a region is consist of some subregion. The mortality under 5 age for both sexes can be referred to as the number of babies that die within 5 years of their birth from a group of 1000 live births. Life expectancy can be referred to as the average lifespan of a group of people who were all born in the same year, assuming that mortality rates at each age do not change in the future. The dataset used in this study has 44 missing values, where region and subregion have four missing values each. These eight observations are replaced manually with their original values, leaving 36 missing values for six numerical variables. These 36 missing values are again replaced with the mean of their respective subregion, using a technique called mean imputation. The study's analysis, therefore, utilized all 454 observations and 11 variables.

2.2 Project objective

The objective of this project is to conduct a descriptive analysis on the provided demographic dataset. Initially, the frequency distribution of variables for the year 2022 is examined using histograms, box plots, and measures of central tendency, such as mean and median. Subsequently, the variability between regions and subregions is analyzed using box-whisker plots. Pearson correlation coefficients and scatter plots are then used to determine any linear correlations. Lastly, a comparison of all observations for the years 2002 and 2022 is performed using scatter plots.

3 Statistical Methods

Several statistical methods are presented in this section. The software R (version 4.0.5)(R Core Team, 2022) with library patchwork (Pedersen, 2022), tidyverse (Wickham and et al., 2019) and ggplot2 (Wickham, 2016) are used for the visualization as well as calculation. To merge multiple grid-based plot gridExtra (Auguie, 2017) and cowplot (Wilke, 2020) has been used. xtable (Dahl et al., 2019) is used to export the table in the latex format.

3.1 Arithmetic mean

The arithmetic mean can be applied when the underlying random variable for a sample is continuous. If $x_1, x_2, ..., x_n$ are numeric observations then arithmetic mean is defined by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

where n is the total number of observations. The mean differs for each sample (or population) and is not always present in the dataset used to compute it. This measure is best suited for fairly homogeneous data because it is sensitive to extreme values in one of the data tails (Hay-Jahans, 2019, p. 74).

3.2 Median

The median is determined if the sample $y_1, y_2, ..., y_n$ of a particular variable Y is sorted in ascending order. It is defined as

$$\tilde{y} = \begin{cases} y_{\frac{n+1}{2}} & \text{, if } n \text{ is odd.} \\ \frac{1}{2}\left(y_{\frac{n}{2}} + y_{(\frac{n}{2}+1)}\right) & \text{, if } n \text{ is even.} \end{cases}$$

In cases where a dataset contains extreme values on one end, this central tendency measure is often preferred as it is not affected by extreme values in the sample (Hay-Jahans, 2019, p. 75).

3.3 Standard deviation (SD) and variance

To determine how data are spread out from the mean value, we normally use empirical variance or standard deviation. If $x_1, x_2, ..., x_n$ are sample data then the empirical variance is defined as

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad .$$

In the formula, squared difference helps to get non-negative values where $\bar{x}$ represents mean value of the sample data. The standard deviation is simply $\sqrt{\hat{\sigma}^2}$. While a high standard deviation suggests that the values are dispersed throughout a larger range, a low standard deviation suggests that the values tend to be near to the mean of the set (Lee et al., 2015, p. 220).

3.4 Pearson correlation coefficient

The Pearson correlation coefficient is a statistical measure that is used to determine the linear relationship between two quantitative variables. The linear relationship can be defined when two variables are directly related, meaning that if the value of x changes, y must likewise change in the same manner. We can use a scatter plot to visualize the correlation between the variables. The strength of the relationship can be presented by the correlation coefficient. The range of the coefficient is -1 to 1 (Benesty et al., 2009).

If there are two variables X and Y with observations $x_1, x_2, ..., x_n$ and $y_1, y_2, ..., y_n$ then Pearson correlation can be defined by

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

If the value of r is between 0 and 1 then the variables are positively correlated and if less than 0 then negatively correlated.

$r = 1$: Exact positive linear correlation i.e when one variable increases another variable also increases in the same direction.
$r = 0$: No linear correlation exists i.e there is no linear relationship between the variables.
r = -1 : Negative linear correlation i.e when one variable increases another variable decrease.

3.5 Boxplot

A boxplot is a graphical representation with a five-number summary, i.e., maximum, minimum, 1st quartile ($Q1$), median, and third quartile ($Q3$). This summary can be used as an early indicator of symmetry violations or the potential existence of extreme values (outliers) in the data. Here, the minimum is the lowest, and the maximum is the highest value in the dataset. Quantiles are a statistical concept that divides a set of observations into subgroups of equal size. Quartiles are a commonly used example of quantiles, which involve dividing a dataset into four equal parts. The first quartile ($Q1$), also referred to as the lower quartile (0.25), is the median value between the minimum value and the median. Similarly, the third quartile ($Q3$), also referred to as the upper quartile (0.75), is the median value between the maximum value and the median. Assuming that there are $x_0, x_1, x_2.......x_n$ data points where n is the number of data points, then the quartiles can be written as follows:

$$Q_1 = x_{\left(\frac{n+1}{4}\right)}\text{th term}, Q_2 = x_{\left(\frac{n+1}{2}\right)}\text{th term}, Q_3 = x_{\left(\frac{3(n+1)}{4}\right)}\text{th term}.$$

In a box plot, a longer right whisker (a T-shaped line) denotes right-skewed data, while a longer left whisker denotes left-skewed data. The sample is deemed symmetrical if the whiskers are roughly equal in length and the median is roughly in the center of the box. Since the median is more robust to extreme values than the mean, box plots are often used in statistical measures. The whisker represents how data is skewed. The box always represents the range of the middle 50% of data values, which is called the interquartile range $IQR = Q_3 - Q_1$

When observations fall 1.5 IQR or more units below the first quartile or rise above the third quartile, they are classified as extreme values (or outliers) in the sample (Hay-Jahans, 2019, p. 137-142).

3.6 Scatterplot

When one continuous variable (X) depends on another variable (Y) or when the two continuous variables are independent, a scatter plot can be used to analyze their relationships. In the scatter plot, points are on the 2-dimensional coordinate axes for ordered pairs (x_i, y_i), where $i = 1, 2, ..., n$ is used to define different values along the x and y-axis. A linear relationship between the variables can be determined by drawing a best-fit line. A positive correlation exists between the variables if the pattern of the dots slopes from lower left to upper right. If the dot pattern slopes downward from upper left to lower right, there is a negative correlation (Hay-Jahans, 2019, p. 159-168).

3.7 Histogram

A histogram is a graph that is used to visualize the grouped frequency distribution of a continuous variable. The horizontal axis of a histogram shows the ranges of values (grouped), and the density of the distribution is represented by vertically scaled bars. We can count how many values fall into each interval after dividing the entire range of values into a series of intervals, which are typically referred to as bins. The area of the constructed rectangle is proportional to the number of cases in the bin even if the bins are not of equal width and in the density histogram the area of the entire histogram equals to 1 (Hay-Jahans, 2019, p.131-136).

4 Statistical Analysis

In this section, all the 6 numerical variables such as U5 mortality of males, U5 mortality of females, U5 mortality of both sexes, life expectancy of male, female, both sexes are used for the year 2022. Finally, a comparison between the year 2022 and 2002 is outlined in the last subsection. All the decimal numbers which are used in the analysis are corrected to 2 decimal places.

4.1 Frequency distributions of the variables

In this subsection, we will analyze our data using a histogram and box plot. First, we observe the frequency distributions of U5 mortality for males, females, and both sexes.

From Figure 1 and Table 1 on page 18 in the appendix section, we see the mean value of males is 29.23, which is 3.93 higher than the U5 mortality of females, and the median value of males (17.55) is also higher than the of females (13.62). The range of male U5 mortality is calculated as (159.75) which is also higher than the female U5 mortality (144.45). This indicates that there are more countries on the right side of U5 male mortality than on the left side of U5 female mortality. The standard deviation (30.09) is also higher for infant mortality in males than in females. However, when we see the U5 mortality of both sexes, the mean value (26.67) is less than the mean value of males (29.23) but higher than the mean value of females (24.01).

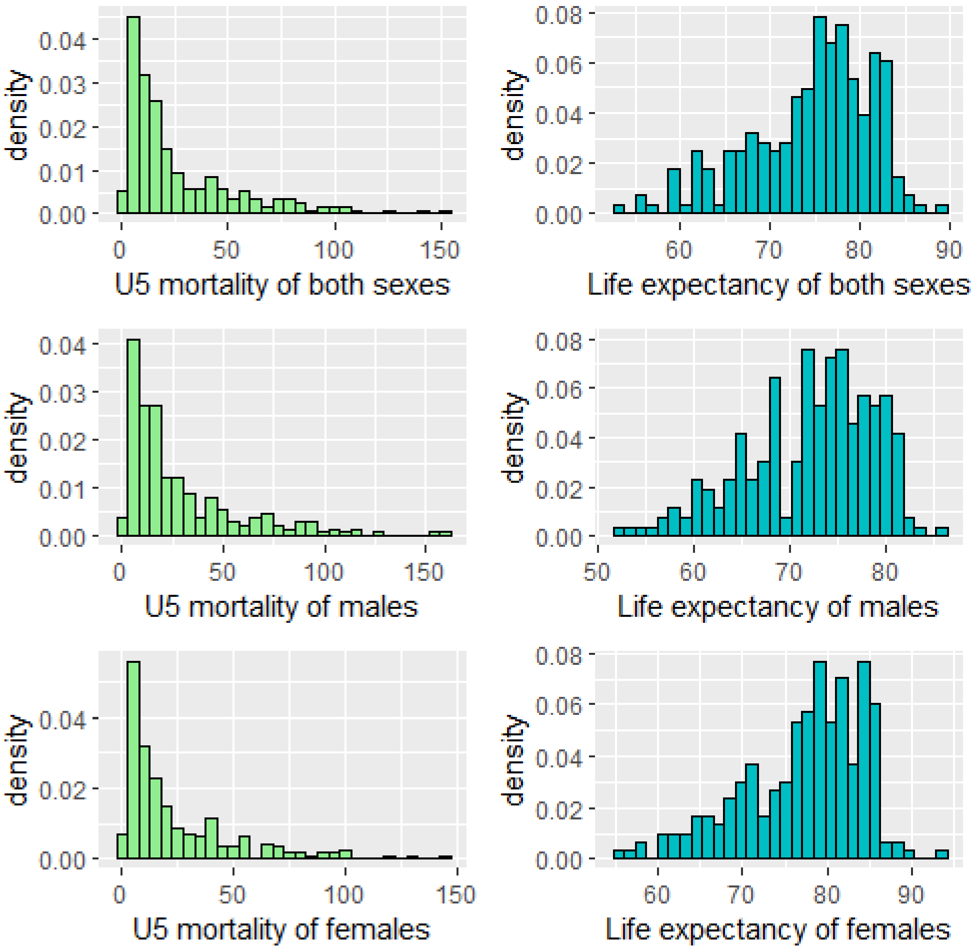

Figure 1: Histogram showing the frequency distributions of life expectancy and U5 mortality of male, female, and both sexes.

While considering the life expectancy we observe, female life expectancy (77.18) is 5.08 years higher than that of males (72.10). Similarly, the median value of females is also higher (5.08) than the median value of males. This indicates there are more countries on the right side for females than males, which we can also see in Figure 1. Standard deviations and ranges are also higher for females than males life expectancy.

From Figures 2 and 3, we observe the differences between regions and sexes. All the numerical values used in this part are obtained using the R code.

If we compare the life expectancy of males in 5 regions, we see that the African region has the lowest average life expectancy (64.40), Europe has the highest average life expectancy (76.99), and America has the second highest average life expectancy of males, which is 74.55. Since the whisker is longer on the right side, the Asia region is right-skewed compared to America, where the mean and median values of America (74.55 and 75.15) are a bit higher than Asia (72.90 and 73.55). Oceania has the lowest IQR (6.13) among the regions whereas Africa has the highest variance.

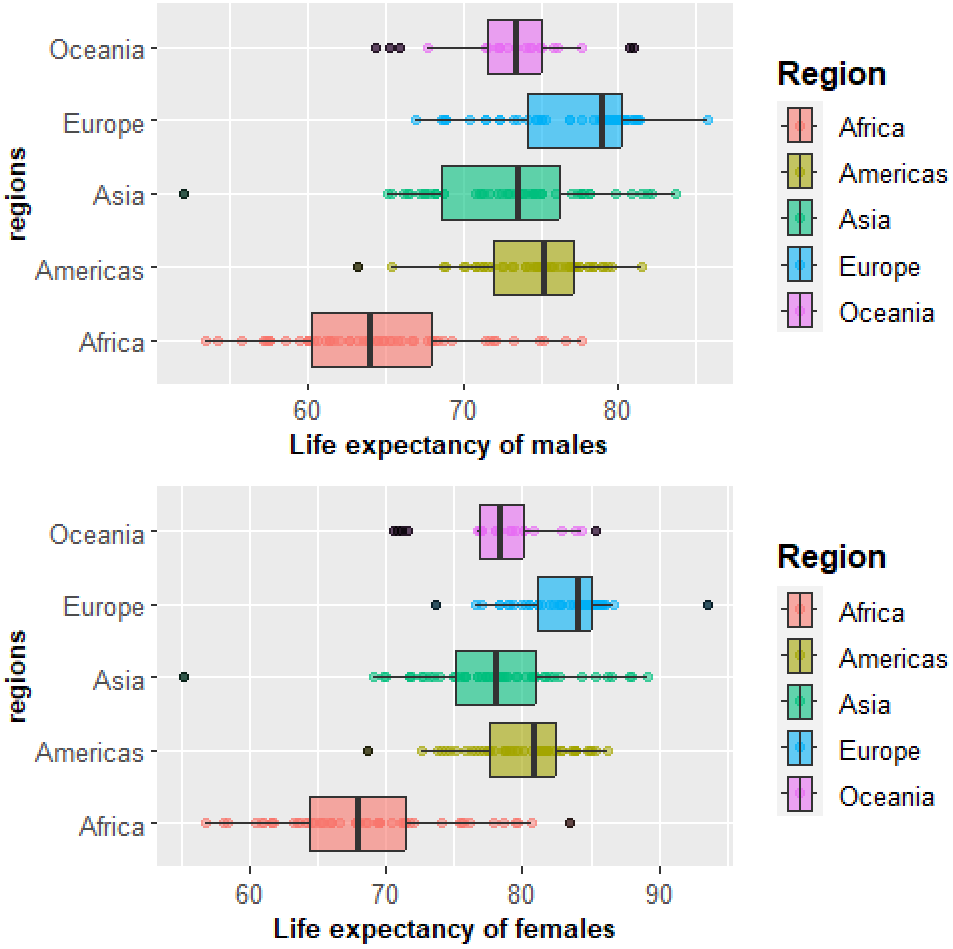

Figure 2: Box plot showing life expectancy of males and females of all regions.

Similar to males, female life expectancy is highest in the Europe region (83.05), and average life expectancy is lowest in the African region (68.49). The life expectancy of females in Africa is almost symmetrical, with the highest IQR of 7.08 and Oceania having the lowest IQR of 3.2. Since the whisker is longer on the left side, Europe is left-skewed compared to Oceania.

On the contrary, we see the opposite scenario in Figure 3. Here, the U5 mortality of males and females can be observed. In the African region, average infant mortality (65.95) is higher than in any other region. Since the median (65.19) and mean are almost the same, it suggests that the data are roughly distributed symmetrically around the center with almost no skewness. Europe has the lowest variance (27.45) as well as the lowest mean and median among the regions. Asia is right-skewed as the whisker is longer on the right side, with the second-largest IQR (22.89).

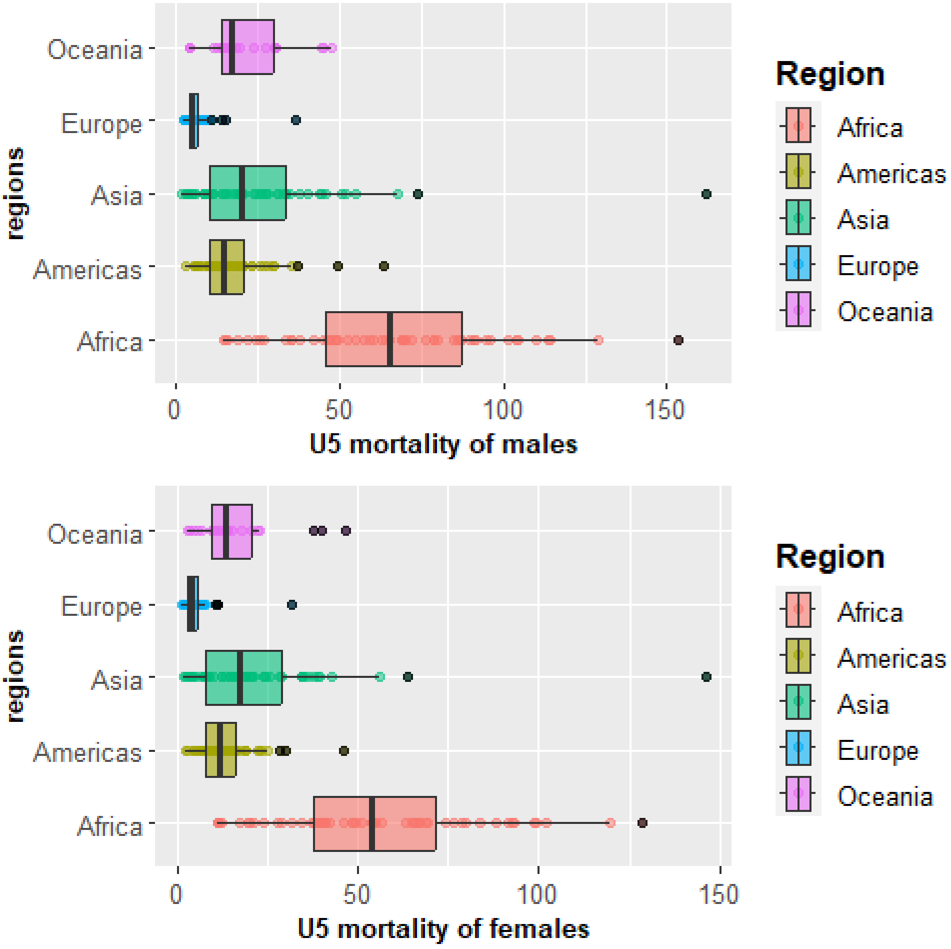

Figure 3: Box plot showing U5 mortality rate of males and females for all regions.

In addition, the average female U5 mortality is highest in Africa, at 55.37, followed by average values from Asia and Oceania. Europe has the lowest mean and median (5.19 and 4.03) in terms of female U5 mortality. This region also has the lowest variance (19.93) and lowest IQR (2.36). The Americas region has the lowest difference (1.21) in mean and median, suggesting data are roughly symmetrically distributed around the center with positive skewness.

4.2 Variablility analysis

The analysis of the data by region and subregion will be the main focus of this section. Since box plots allow us to compare data from various subregions, we will use it to analyze the data. Our data set consists of 5 regions and 21 subregions. However, we will consider only the African region in the analysis. All the values for this subsection are calculated and summarized in Table 2, on page 19, in the appendix section.

At first, if we see the life expectancy of males in Eastern Africa and Middle Africa, then the difference between mean and median is very low for these two subregions, which are 0.32 and 0.30, respectively. This suggests that almost all of the countries in these regions are symmetrically distributed. High variability is seen in northern Africa, which is 48.35, as there is a country that has an extremely low life expectancy for males. As shown in Figure 4, the IQR in Western Africa is high for both male and female life expectancy. Middle and eastern Africa has some countries that have exceptionally lower or higher life expectancies for males and females. For both males and females, western Africa is right-skewed, meaning the whisker is longer on the right side. Female life expectancy in East Africa has an almost symmetrical distribution where the mean and median difference is very low .05 and four countries have extremely high and low life expectancies in this subregion. Both sexes average life expectancy (84.76) for Middle Africa is higher than male and female, whereas for all other subregions, both sexes have the lowest average life expectancy than male and female individually.

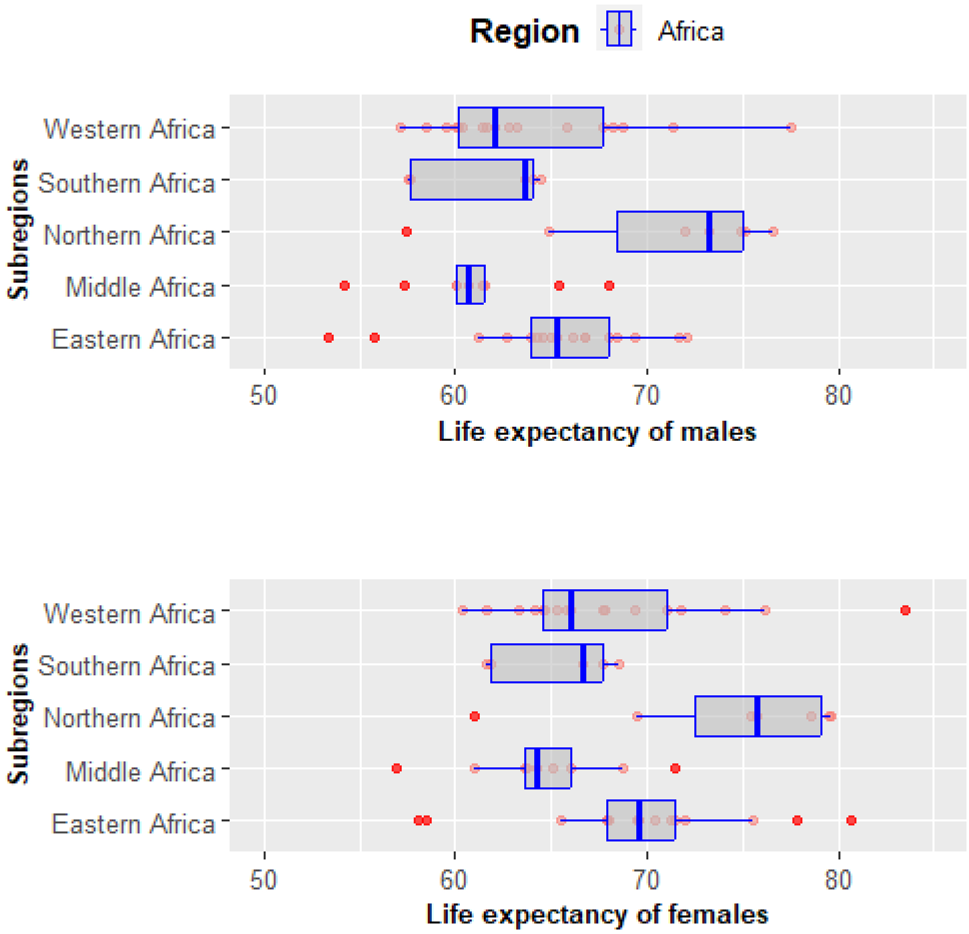

Figure 4: Box plot of all subregions with respect to male life expectancy and female life expectancy of Africa.

Now from Figure 5, we will analyze the U5 mortality for males. Here, middle Africa has the highest IQR (37.63), as well as the highest mean and median (90.11 and 87.58). Eastern Africa is roughly symmetrically distributed, where the median is in the middle position and both sides whisker lengths are almost the same, except for one country with an extremely high mortality rate. The Northern African region is right-skewed with the lowest mean value (39.30) and lowest median value, where the mean is affected by a country that has the highest U5 mortality in this subregion. Similar to the U5 mortality of males, the middle Africa region has the highest mean and median values, which are 90.11 and 87.58 respectively. IQR is also high in this subregion. Northern Africa is positively skewed with the lowest mean, which is affected by a country with higher infant mortality, where median values (24.40) is also lowest. The average infant mortality of both sexes (84.76) is lower than males but 5.56 higher than females in the middle African subregion. If we compare the life expectancy and U5 mortality of males, females, and both sexes, then the variance and IQR are higher in U5 mortality than the life expectancy for males, females, or both sexes in general. In short, life expectancy category has a lower variance and the mortality rate has a high variance for males, females, and both sexes.

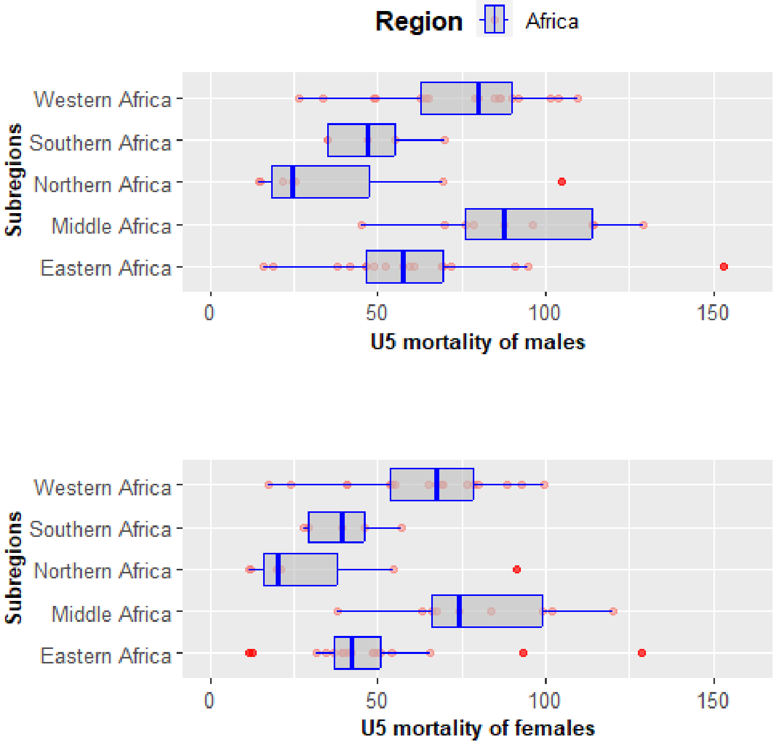

Figure 5: Box plot showing the U5 mortality of males and females for all African subregions.

4.3 Relationship between the variables

In this part, we will try to figure out if any two variables are related to each other or not. We find that as the U5 mortality of both sexes increases along x-axis, the life expectancy of male, female and both sexes decreases along y-axis. Most precisely, U5 mortality in both sexes, males and females are nearly identical and negatively correlated (Figure 6) where the values are -0.90,-0.88 and -0.91 respectively. This indicates there is a strong negative correlation between all the sexes and the U5 mortality of both sexes, where females have the largest negative linear correlation.

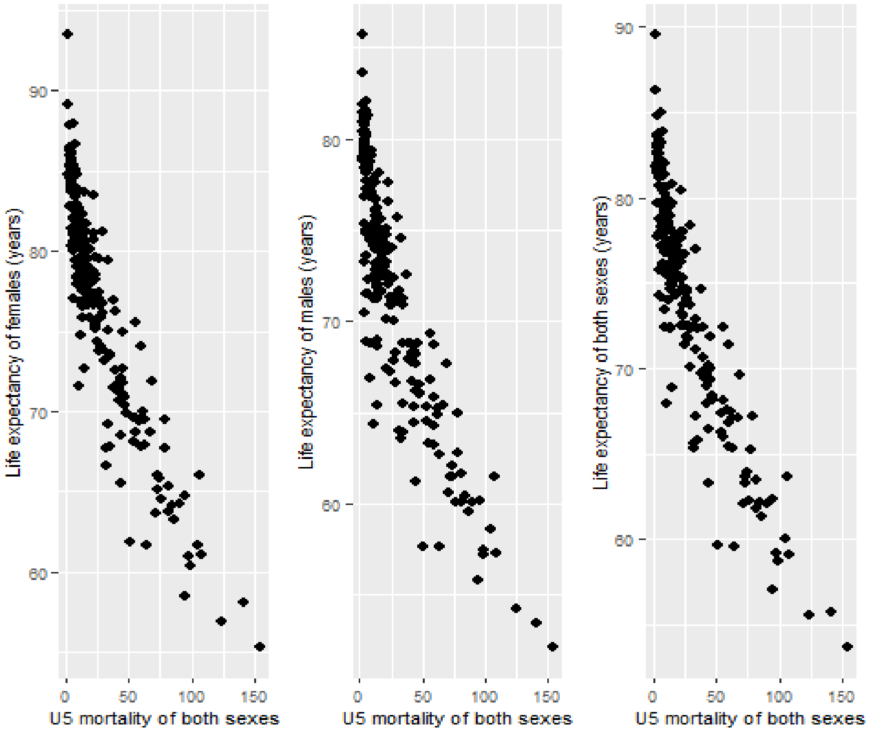

Figure 6: Scatter plot showing negative correlations between two different variables.

Contrarily, since the observations run along an increasing line (Figure 9 on page 18 in the appendix) and is nearly linear, there is a positive linear correlation between life expectancy for both sexes when compared to female life expectancy (0.99) and male life expectancy (0.99). Furthermore, there is a strong correlation between male and female life expectancy (0.97). As a result, we can draw the conclusion that there are positive correlations between the sexes, meaning that as female life expectancy rises, so does male life expectancy.

4.4 Comparison of the variables between 2002 and 2022

A scatter plot (Figure 7) is drawn to demonstrate the changes between 2022 and 2002. All the regions are marked with different colors. The horizontal axis shows data for 2022, and the vertical axis shows data for 2002. A line in the scatter plot helps to distinguish changes between 2022 and 2002. Points falling above this line are considered to be decreasing in 2022, and falling below it indicates an increase in 2022 compared to 2002. We observe that the life expectancy of both sexes increases in all regions except two American and two African countries. Two Asian and one American country slightly touches the line, meaning countries in these three regions life expectancy of both sexes does not differ that much between 2022 and 2002. Europe, America, and some Asian countries will have a higher life expectancy in 2022. In contrast, countries in the African region tend to have the lowest life expectancy in 2022 compared to 2002, and one Asian country will have an exceptionally low life expectancy in 2022.

We also observe that the overall U5 mortality of both sexes has sharply decreased in 2022 compared to 2002 since most of the countries are lying above the line except for two African countries whose mortality rate increases in 2022. However, the entire European region tends to differ very little in terms of U5 mortality for both sexes between 2022 and 2002. Some Asian and American regions also fall into the same category. In contrast, the African region, including one exceptional Asian country, has a higher mortality rate in 2022 than other countries in other regions.

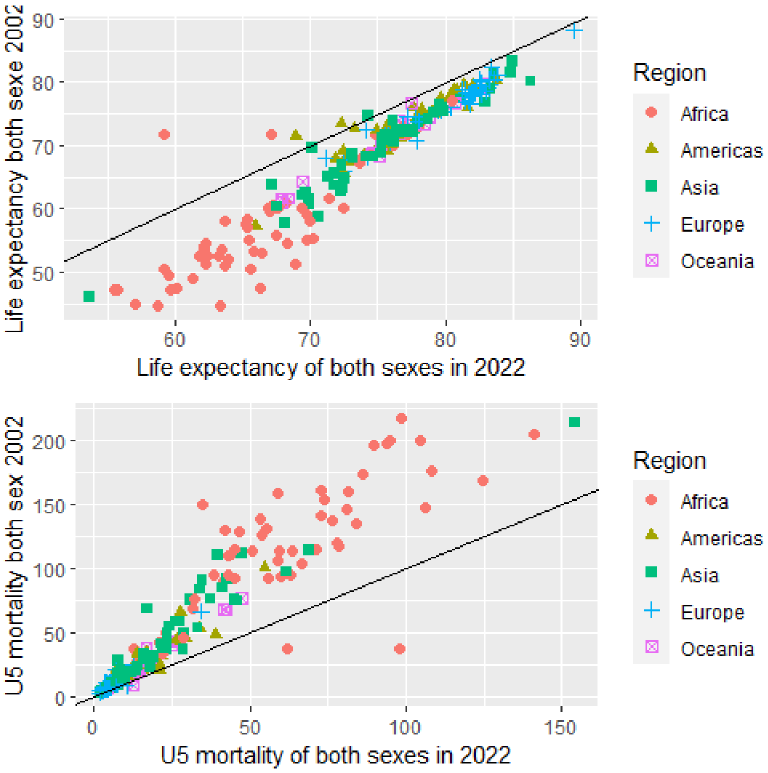

Figure 7: Scatter plot of life expectancy and U5 mortality of both sexes in 2022 and 2002.

5 Summary

In this project, a descriptive analysis of demographic data was conducted which includes variables of under 5 mortality rate of male, female and both sexes as well as life expectancy of male, female and both sexes in 227 countries in the year 2022 and 2002. This data was collected from the International Data Base (IDB) of the U.S. Census Bureau and compiled by the instructors of the course Introductory Case studies. The frequency distributions of the variables, the difference between sexes and all 5 regions, the dependency and variability between the variables, the relationship between the variables of U5 mortality and life expectancy categories and how the variables changed between 2022 and 2002 was analyzed with the help of statistical measures and graphical presentations. We mainly used the data of 2022 to answer all of the above-mentioned questions except the comparison of the variables between 2022 and 2002. Presenting with a histogram, we saw that mortality rate and life expectancy are changing in reverse. This suggests that life expectancy is higher in countries with lower U5 mortality. From the box plot, we saw European countries have more life expectancy than any other region where U5 mortality is higher in Africa. Variability existed between the subregions of an individual variable. High variability was observed in U5 mortality whereas low variability was observed in the life expectancy. We also noticed that there is negative correlation between U5 mortality and life expectancy category. Life expectancy increased in 2022 compared to 2002. On the contrary, U5 mortality decreased in 2022 compared to 2002. The main reasons for the longer life expectancy may be better lifestyles, better education, and improvements in health care and medicine. To reduce the mortality rate better sanitation, access to clean drinking water, immunization against infectious diseases, and other public health initiatives can be taken.

We analyzed only six numerical variables, and other external factors like weather, lifestyle, and economy might have a meaningful effect in any particular region, so the data sets might not be completely perfect. In the future, finding a regression model between the under-5 age mortality rate and life expectancy related to each subregion might be interesting.

Bibliography

Baptiste Auguie. *gridExtra: Miscellaneous Functions for "Grid" Graphics*, 2017. URL `https://CRAN.R-project.org/package=gridExtra`. R package version 2.3.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

David B. Dahl, David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton. *xtable: Export Tables to LaTeX or HTML*, 2019. URL `https://CRAN.R-project.org/package=xtable`. R package version 1.8-4.

Christopher Hay-Jahans. *R Companion to Elementary Applied Statistics*. CRC Press, 2019.

Dong Lee, Junyong In, and Sangseok Lee. Standard deviation and standard error of the mean. *Korean journal of anesthesiology*, 68:220, 06 2015. doi: 10.4097/kjae.2015.68.3.220.

Jennifer Hickes Lundquist, Douglas L Anderton, and David Yaukey. *Demography: the study of human population*. Waveland Press, 2014. [Visited on 29-04-2023].

Thomas Lin Pedersen. *patchwork: The Composer of Plots*, 2022. URL `https://CRAN.R-project.org/package=patchwork`. R package version 1.1.2.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022. URL `https://www.R-project.org/`.

U.S Census Bureau. International data base, 2022. URL `https://www.census.gov/glossary/`. [Visited on 30-04-2023].

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

Hadley Wickham and et al. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686. URL `https://doi.org/10.21105/joss.01686`.

Claus O. Wilke. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*, 2020. URL `https://CRAN.R-project.org/package=cowplot`. R package version 1.1.1.

Appendix

A Additional figures

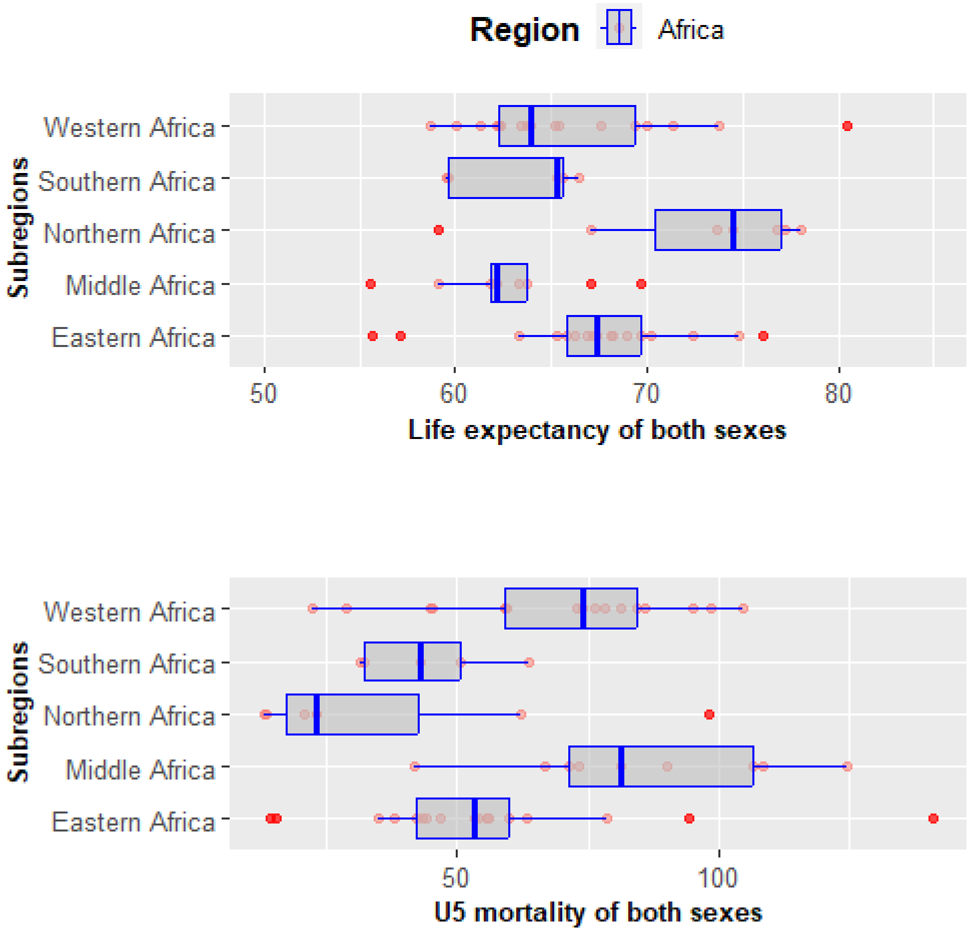

Figure 8: Box plot of all African subregions for life expectancy of both sexes and infant mortality of both sexes.

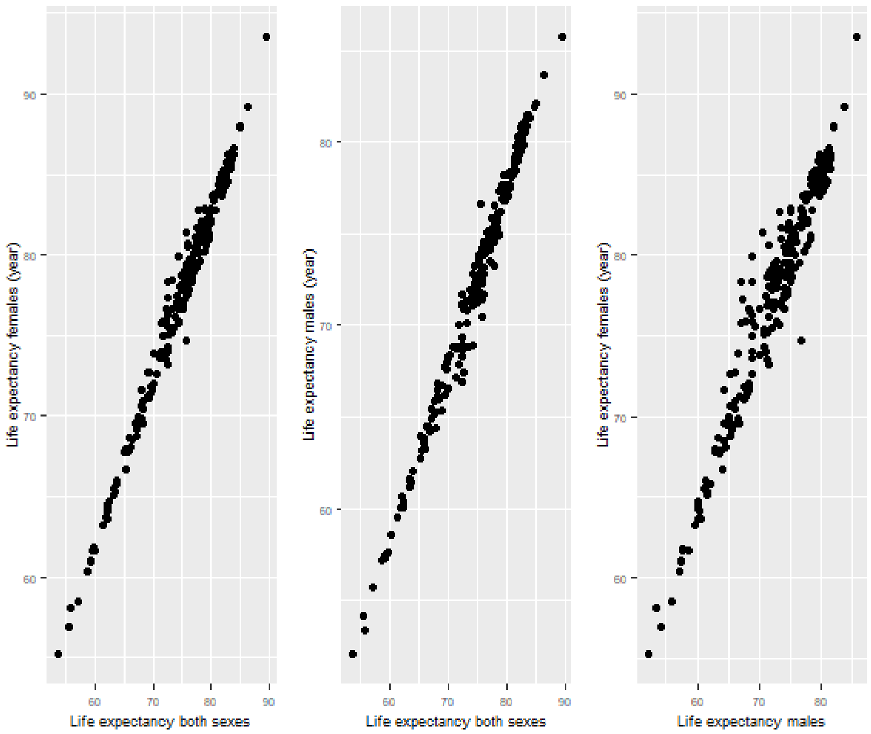

Figure 9: Scatter plot showing the positive correlations between the variables.

B Additional tables

Table 1: Descriptive statistics of 6 numerical variables for 2022

Variables	minimum	maximum	Mean	Median	Standard deviation
Life expectancy of both sexes	53.65	89.52	74.58	75.82	6.84
Life expectancy of males	52.10	85.70	72.10	73.26	6.67
Life expectancy of females	55.28	93.49	77.18	78.69	7.13
U5 mortality of both sexes	1.94	154.13	26.68	15.08	28.09
U5 mortality of males	2.03	161.78	29.23	17.55	30.09
U5 mortality of females	1.64	146.09	24.01	13.62	26.14

Table 2: Summary statistics of all subregions for Africa region in 2022

Variables	Northern	Middle	Western	Southern	Eastern
U5 mortality both sexes					
Mean	36.21	84.76	68.33	44.25	54.95
Median	22.85	81.09	73.29	43.11	53.35
Variance	1028.07	644.27	590.99	179.98	876.53
IQR	25.33	35.31	29.27	18.45	17.74
Min	12.95	41.73	22.14	31.61	14.30
Max	98.26	124.58	104.72	63.57	141.20
U5 mortality males					
Mean	39.30	90.11	73.65	48.48	60.92
Median	24.40	87.58	79.44	46.80	57.61
Variance	1182.78	682.80	618.94	218.78	1009.72
IQR	28.93	37.63	30.93	20.03	23.20
Min	14.43	45.29	26.61	35.07	15.97
Max	104.67	129.08	109.77	70.16	153.23
U5 mortality females					
Mean	32.95	79.25	62.82	39.91	48.80
Median	20.23	74.39	66.97	39.32	42.02
Variance	880.19	608.09	565.58	144.17	770.16
IQR	22.06	32.91	28.38	16.83	13.76
Min	11.38	38.07	17.43	28.09	11.51
Max	91.52	119.95	99.52	56.79	128.81
Life expectancy of both sexes					
Mean	72.35	62.72	66.19	63.34	67.28
Median	74.45	62.11	64.56	65.32	67.42
Variance	47.18	16.87	32.08	11.64	27.55
IQR	6.60	1.87	7.16	5.95	3.84
Min	59.16	55.52	58.76	59.57	55.72
Max	78.03	69.70	80.48	66.47	76.10
Life expectancy of males					
Mean	70.60	60.95	64.15	61.45	65.00
Median	73.26	60.65	62.41	63.60	65.32
Variance	48.35	16.34	29.26	12.46	23.93
IQR	6.60	1.46	7.49	6.37	4.09
Min	57.43	54.19	57.16	57.57	53.39
Max	76.57	67.98	77.58	64.46	72.04
Life expectancy of females					
Mean	74.19	64.53	68.30	65.28	69.62
Median	75.72	64.24	66.87	66.68	69.57
Variance	46.33	17.58	35.30	10.97	32.46
IQR	6.61	2.42	6.70	5.93	3.47
Min	60.97	56.88	60.41	61.64	58.12
Max	79.57	71.48	83.51	68.53	80.66