TU Dortmund

Introductory Case Studies

Project 1 - Descriptive analysis of demographic data

Lecturers:

Prof. Dr. Jörg Rahnenführer

Dr. Philipp Adämmer

Dr. Andrea Bommert

M. Sc. Hendrik Dohme

Author: Jorge Eduardo Durán Vásquez

Group 6

Group Members: Daniel Casas, Sadat Mahmud

May 14, 2021

Contents

1. Introduction

The analysis of demographic data is vital for evidence-based decision making at all stages of a country's development (United States Census Bureau, 2013). Tracking the development of population indicators enables a better understanding of needs, not only of a country but also of subregions and entire world regions. The purpose of this report is to understand the variables Life expectancy at Birth and Total Fertility Rate in a specific year by analyzing the frequency distribution of these variables. In addition, the difference of Life Expectancy at Birth between Males and Females is analyzed using a relative frequency distribution of the absolute differences. Complementing that the Pearson correlation coefficient is used to obtain the bivariate correlations of the demographic variables. Using information regarding the geographical location of each country, the subregions' Life Expectancy at Birth of Both Sexes and Total Fertility Rate for the year 2020 is compared and the variability of these variables is assessed. The previous analysis is conducted using multi-box plot diagrams and comparing measures of dispersion like total range, interquartile range, and standard deviation. Finally, the change between 2000 and 2020 of Life Expectancy at Birth of Both Sexes and Total Fertility Rate is evaluated with a direct graphical comparison.

In Section 2, information of the data set, as well as the process of data collection and data quality is described. Section 3 explains the software tools that are used and the statistical methods, which include the relative frequency histogram, the box plot, and the Pearson correlation. In section 4, the results of the analysis are presented. Finally, section 5 concludes the report, discusses the current constraints of this analysis, and provides recommendations for future research.

2. Problem statement

The data set used in this report was provided by the teachers of the subject Introductory Case Studies of the Master of Science in Data Science at TU Dortmund University. It is a subset of the data set International Data Base (IDB) published by the U.S. Census Bureau on their web page.

The U.S. Census Bureau has prepared population estimates and projections for other countries since the 1960s. The IDB provides accurate annual demographic measures of over 200 countries and respective regions of the world (United States Census Bureau, 2021). According to the documentation provided by the U.S. Census Bureau, the data collection process begins

by collecting demographic data from censuses, surveys, vital registries, and administrative records from a variety of sources (United States Census Bureau, 2021). The U.S. Census Bureau publishes the data after carefully evaluating the internal and temporal consistency between their existing data and the new data.

The sample size amounts to 442 observations that include information of 221 countries in the years 2000 and 2020. These countries are classified into one of the five world regions and 21 subregions according to their location. The number of countries that can be found in each subregion is presented in Table B1. The data set has eleven variables, which contain information either about the geographic location or demographic data by year. The numeric and ordinal variable year takes the values 2000 or 2020.

The geographic location information for each observation is given by three variables. These are the name of the country, the corresponding region and subregion, all of them as a string parameter, and the latter two being nominal measurements. According to (United States Census Bureau, 2020), the countries and areas covered in the data set are recognized by the U.S. Department of State and have a population of 5,000 or more. Furthermore, the geographical information is complemented by two country codes, the Federal Information Processing Standard (FIPS) 10-4 and the Geopolitical Entities, Names, and Codes (GENC) standard (United States Census Bureau, 2020). The codes FIOS and GENC are string value compounds with 2 capital letters.

The demographic data is given in a ratio scale and covers the following variables. First, the Total Fertility Rate is an integer value that shows the average number of children that would be born per woman if all women lived to the end of their childbearing years and bore children according to a given set of age-specific fertility rates (United States Census Bureau, 2017). Second, the Life Expectancy at Birth of Both Sexes is also an integer value and represents the average number of years a group of people born in the same year can be expected to live if mortality at each age remains constant in the future (United States Census Bureau, 2017). The other two features that are given in the data set result from the stratification of the Life Expectancy at Birth of Both Sexes into Life Expectancy at Birth of Males, respectively Females. They are also integer values and refer to the same information as the Life expectancy at Birth of Both Sexes, but in this case by gender.

It can be assumed that the quality of the IDB data set is assured by the U.S. Census Bureau's evaluation process regarding internal and temporal consistency of the data (United States Census Bureau, 2020). It is important to remark that the data set contains missing values for 7 countries in the year 2000. This is not a problem for the analysis of this report, though, because the values of the year 2000 are only used for analyzing a general global behavior and not for assessing individual countries.

To understand the distribution of the Life expectancy at Birth and the Total Fertility Rate in 2020 for the group of 221 countries, the relative frequency histograms of the variables are plotted. The aggrupation of the countries into subregions and regions as given by the U.S. Department of State is used in this report as well. In addition, the stratification of the variable Life Expectancy at Birth in Life Expectancy at Birth for Males and Life Expectancy at Birth for Females is analyzed with the histogram of the absolute difference between these two variables. Next, the bivariate correlations of the variables are presented using the Pearson correlation coefficient and the scatter plot analysis. Other factors that are compared for the year 2020 are the variables Life Expectancy at Birth of Both Sexes and Total Fertility Rate within the subregions and between them. This comparison is presented using multi-box plot diagrams. Finally, the change between 2000 and 2020 of Life Expectancy at Birth of Both Sexes and Total Fertility Rate is evaluated with a direct graphical comparison of the variables.

3. Statistical methods

3.1 Relative frequency histogram

The relative frequency histograms are focused on the overall pattern of how the values are distributed rather than on the counts themselves (Sharpe et al., 2012). On the vertical axis the share of cases falling in each bin in the total number of cases is displayed. This is called the relative frequency and is calculated as follows:

$$Relative\ frequency = \frac{Cases\ in\ a\ specific\ bin}{Total\ number\ of\ cases}$$

To plot the diagram it is also necessary to choose the width of the bin. There is not a specific rule to select the number of bins. It rather needs to be evaluated which bin size is the most adequate in each case depending on the unit and overall distribution of the data. For this study a bin width of 0.5 for Total Fertility Rate and 2 for Life Expectancy at Birth is used.

To analyze the relative frequency histogram, the evaluation of the shape, center and spread of each distribution is important. First, the shape is described in terms of its mode, its symmetry and whether it has any gaps or outlying values (Sharpe et al., 2012). A distribution can be classified into uniform, unimodal, bimodal, and multimodal, depending on the number of humps. In addition, a distribution can also be called symmetric or skewed. A histogram is symmetric if the right and left tail have the same, but mirrored, shape. In turn, it is skewed if one tail stretches out farther than the other (Sharpe et al., 2012). Second, the center is the number that describes the typical value or class of the variable that is being analyzed (Sharpe et al., 2012). The mean is an indicator of the center of unimodal and symmetric distributions, but it can be misleading for skewed data or for distributions with gaps or outliers (Sharpe et al., 2012). To avoid this, in skewed distributions the median is used as a value that splits the histogram into two equal areas (Sharpe et al., 2012).

To calculate the mean, it is necessary to add up all the values of a variable in the data set and divide the sum by the total number of observations (n) (Sharpe et al., 2012). In the following formula for the mean the variable is represented by "x_i", where "i" represents the position of the observation in the data set:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

To obtain the value of the median, the data is sorted from the lowest to the highest value and then the value located in the middle of the distribution is selected (Sharpe et al., 2012). The position of the median depends on the number of observations (n):

$$Possition = \begin{cases} \frac{n+1}{2} & if\ n\ is\ odd \\ \frac{n}{2}\ and\ \frac{n}{2}+1 & if\ n\ is\ even \end{cases}$$

Finally, the spread of the distribution is measured, among others, with the interquartile range (IQR) and the standard deviation (SD). The IQR summarizes the spread by focusing on the middle half of the data. It is defined by:

$$IQR = Q3 - Q1$$

where Q1 represents the first quartile, i.e. the 25% of the observations with the lowest values, and Q3 stands for the third quartile, i.e. the 75% of the observations with the lowest values. Additionally, the SD provides an idea of how far each value is from the mean. The SD is mainly used for symmetric data and can be influenced by outliers. It is calculated as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where x_i represents the observation i and $\bar{x}$ the mean value.

3.2 Box plot

The box plot of a distribution is considered a suitable means to summarize the main characteristics of a variable. The box in the plot is delimited by Q1 and Q3, which enclose the 50% of the data that are in the middle of the distribution (Sharpe et al., 2012). The median is represented by a horizontal bold line inside the box (Sharpe et al., 2012).

Furthermore, the box plot has two whiskers that represent the minimum and maximum range of the data and are located on the left and on the right side of the box respectively. To obtain the minimum and maximum value the following procedure is used. If the observation with the lowest value is larger than $(Q1 - 1.5 * IQR)$, the minimum whisker takes on the value of that observation, otherwise the minimum is $(Q1 - 1.5 * IQR)$. If the observation with the highest value is smaller than $(Q3 + 1.5 * IQR)$, the maximum whisker takes on the value of that observation, otherwise the maximum is $(Q3 + 1.5 * IQR)$. All values that fall out of these whiskers are classified as outliers or extreme values.

When comparing several groups, boxplots usually do a better job than many histograms. Box plots can be plotted side by side which makes it is easy to compare multiple groups (Sharpe et al., 2012).

3.3 Correlation

The correlation, also called bivariate correlation, helps to show the linear dependency between two variables (Sharpe et al., 2012). It takes values between -1 and +1 and the sign of the coefficient gives the direction of the association (Sharpe et al., 2012). The closer the correlation coefficient is to the extremes (-1, +1), the stronger is the correlation between those variables (Sharpe et al., 2012). One of the measures used to calculate the correlation is the Pearson correlation coefficient (Sharpe et al., 2012). It is computed in terms of two variables y and x as follows:

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

where x_i represents the observation i of the first set and y_i represents the observation i of the other set. Furthermore, $\bar{x}$ and $\bar{y}$ represent the mean values of each group of observations.

3.4 Software tools

The Software used for this report is R in the version R 4.0.5 GUI 1.74 and RStudio Version 1.4.1106, both for MAC iOS.

4. Statistical analysis

The statistical methods explained above are used to evaluate the data set. First, to have a better understanding of the variables Total Fertility Rate and Life Expectancy at Birth of Both Sexes the frequency distribution of each variable for the year 2020 is presented. Using the same statistical method, the differences of sexes is analyzed. Then, the comparison between the regions and subregions is carried out using multi-boxplot diagrams and, finally, the change of the variables Total Fertility Rate and Life Expectancy at Birth of Both Sexes between 2000 and 2020 is studied.

4.1 Frequency distributions

4.1.1 Total Fertility Rate 2020

As can be seen in Figure 1, the relative frequency distribution of the Total Fertility Rate for 2020 is positively skewed since in the majority of countries the average number of children per woman is low. The distribution presents many extreme values, for that reason the center of the distribution is better described by the median than by the mean. In this case the former is 2 children. It means that at least 50% of the countries has a total fertility rate that is smaller or equal to 2 children per woman.

In addition, to describe the distribution of the Figure 1, it is important to analyze the spread of the data. The interquartile range is 1.16 children, which means that the middle 50% of the data has a maximum difference of 1.16 children per woman. The standard deviation is 1.15 children which confirms that the variation of the data is low and that most of the observations are close to the mean.

Min	Q1	Median	Mean	Q3	Max
1.06	1.71	2.00	2.46	2.87	7.00

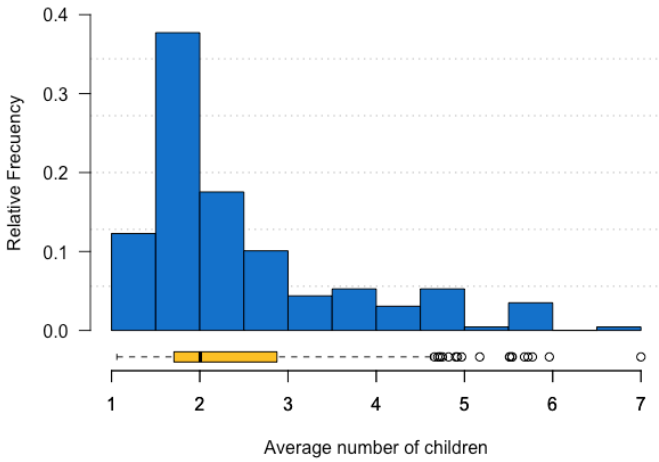

Figure 1. *Relative Frequency Histogram Total Fertility Rate 2020.*

4.1.2 Life Expectancy at Birth of Both Sexes 2020

Life Expectancy at Birth of Both Sexes in 2020 is slightly negatively skewed (see Figure 2) and the median is 75.5 years for both sexes. It can be observed that in at least 50% of the countries the population lives between 69.6 and 79.2 years. The spread of the Life Expectancy at Birth of Both Sexes is again defined by the interquartile range and the standard deviation. The interquartile range is 9.6 years, and the standard deviation is 7.03 years. With these two values and the total range (36.49 years) it can be concluded that the variation of the data is relatively high.

Min	Q1	Median	Mean	Q3	Max
52.84	69.59	75.53	74.04	79.22	89.27

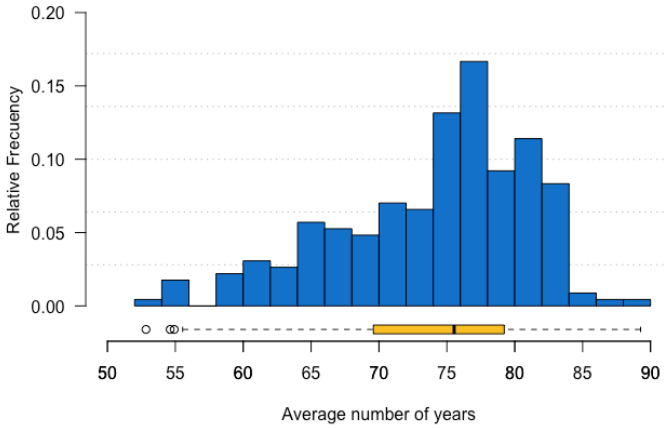

Figure 2. *Relative Frequency Histogram Life Expectancy at Birth of Both Sexes 2020.*

4.1.3 Life Expectancy at Birth of Males and Females 2020

The classification of the variable Life Expectancy at Birth in two classes, males, and females, creates two new distributions (see Figure A1 and A2). These new distributions are similar to each other and to the distribution of the previous variable Life Expectancy of Both Sexes. However, in comparison to Life Expectancy for Males, Life Expectancy for Females is shifted to the right by 5 years since, on average, women live 5 years more than men. The median of the Life Expectancy is 78.35 years for women and 72.95 years for men.

Analyzing the difference between both sexes leads to a better understanding of the life expectancy variable. For that reason, in Figure 3, the result of the difference between the Life Expectancy at Birth of Females and the Life Expectancy at Birth of Males is plotted. More than 95% of the countries obtain a positive value. Based on this, it can be concluded that the probability that a woman lives longer than a man is higher than the other way around. In 50% of the countries, women have at least 4.9 years more life expectancy than men. The variation of the differences is also high, it goes from -3 to 11. However, the interquartile range is located between 4 and 6 years, which means that in the middle 50% of the countries the difference between Females and Males is located in that range.

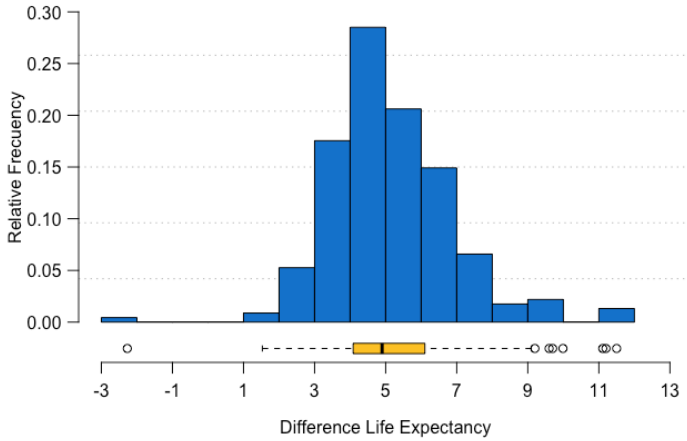

Figure 3. *Relative frequency histogram of the difference between Life Expectancy at Birth of Females 2020 and Life Expectancy at Birth of Males 2020.*

There is just one country where the men live tendentially longer than the women, namely Monserrat in the Caribbean subregion. It is represented by the extreme value that is on the left side of the graph. On the other hand, countries from North-Eastern Europe like Russia and Belarus are concentrated on the right side of the graph with a difference of 11 years.

4.2 Correlation

To compare the linear relation between the variables, the Pearson correlation coefficient is used.

Table 1. Bivariate Pearson correlation variables 2020.

	Total Fertility Rate	**Life Expectancy at Birth Both Sexes**	**Life Expectancy at Birth Males**	**Life Expectancy at Birth Females**
Total Fertility Rate	1	-0.8	-0.77	-0.82
Life Expectancy at Birth Both Sexes	-0.8	1	0.99	0.99
Life Expectancy at Birth Males	-0.77	0.99	1	0.97
Life Expectancy at Birth Females	-0.82	0.99	0.97	1

As can be seen in Table 1, the correlation between Total Fertility Rate and Life Expectancy of Both Sexes for the year 2020 is strongly negative. This means that for countries with higher Fertility Rates, Life Expectancy at Birth is lower. This behavior can also be analyzed using a scatterplot. In Figure 4, the axes depict the Life Expectancy at Birth of Both Sexes, respectively the Total Fertility Rate, and each point represents a country.

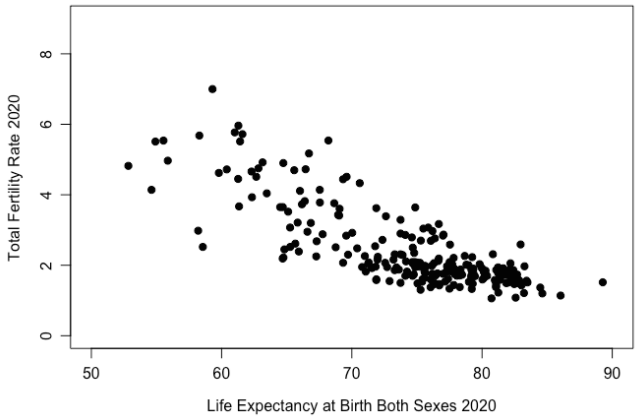

Figure 4. *Scatterplot Life Expectancy at Birth of Both Sexes and Total Fertility Rate 2020.*

Figure 4 shows the linear dependency of Life Expectancy of Both Sexes and the Total Fertility Rate. It becomes clear that the correlation between these two variables is strongly negative. Finally, Table 1 shows that the Pearson correlation between all variables related with life expectancy is high. This is because the values of Life Expectancy of Males and Females are stratified and have a high linear dependency with the value of Life expectancy of Both Sexes.

4.3 Comparison between regions and subregions

The multi-boxplot graph depicted in Figure 5 allows to compare the behavior of the variables within subregions and between them. The representation of the values of Total Fertility Rate

and Life Expectancy of Both Sexes during the year 2020 is compared. First, the Fertility Rate for the year 2020 is analyzed within subregions, then compared between subregions of one region, finally, compared between overall regions. Each box plot represents a subregion, which are grouped according to their region. All the subregions that belong to one region have similar colors that are explained in the legend.

Figure 5 and Table B2 shows that the variability of the Total Fertility Rate, as measured by the total range and interquartile range, is larger in Middle, Western, Eastern, and Northern Africa than in the other subregions. That means that the values of Total Fertility Rate for 2020 in the countries within the African subregions have a high variation. In addition, the difference between the median values of the African subregions is larger than the difference between the median values of the subregions of other regions.

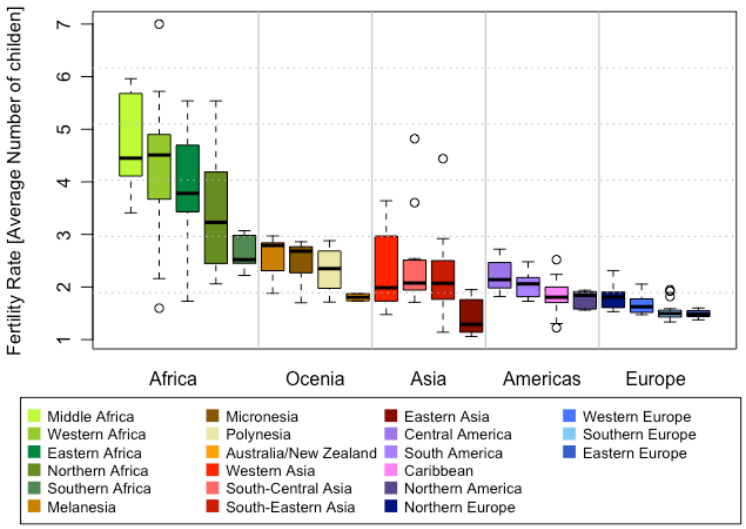

Figure 5. *Comparison of Total Fertility Rate 2020 between regions and subregions.*

Furthermore, the variability of the Total Fertility Rate is the second highest in Asia, third highest in Oceania, and is followed by Americas, and lastly, Europe. Also, the median values of the subregions in these regions are closer to each other than the median values of the African subregions. Figure 5 illustrates a clear tendency that the average number of children that could be born per woman in the countries of Africa is higher than in the rest of the world. There are subregions with countries that are represented as extreme values, but they do not change this analysis at all.

In Figure 6 the Life Expectancy of Both Sexes for 2020 complements the analysis. The subregions are sorted in the same order as before to facilitate the comparison and analysis.

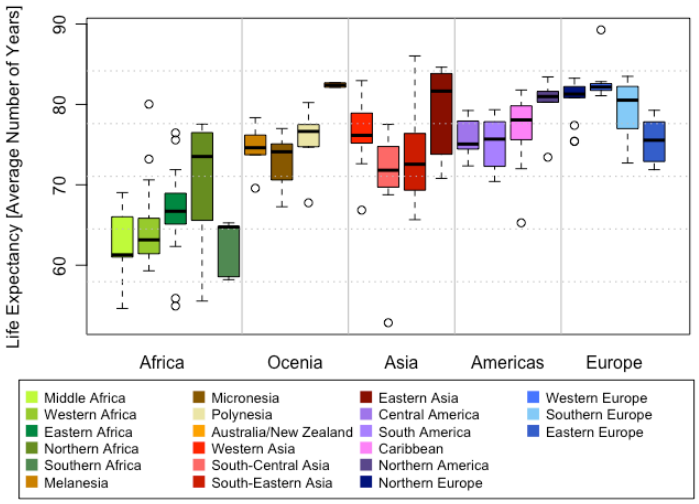

***Figure 6**. Comparison of Life Expectancy at Birth of Both Sexes 2020 between regions and subregions.*

In the case of life expectancy, the new shape of the multi-box plot needs to be noted. The effect of the negative correlation between the Total Fertility Rate and Life Expectancy at Birth of Both Sexes can be observed. The subregions with a high Total Fertility Rate have a low Life Expectancy.

In terms of variability within the subregions, Life Expectancy at Birth of Both Sexes shows a different behavior than the Total Fertility Rate. The African subregions are not the only ones with a high variability of Life Expectancy (Table B2) since the variability for Eastern and Southern Europe, Asia and Americas are similar like Africa in some cases.

To conclude this section, it can be said that the variability of the Total Fertility Rate depends on the subregion. For example, for Middle Africa the variability shows a heterogenous behavior but for Eastern Europe it is homogeneous in the ten countries of this subregion. In a similar way, for Life Expectancy at Birth of Both Sexes the homogeneity or heterogeneity of the variability depends on the region and subregion.

To increase the understanding of the variables in the countries, the extreme values as presented in both Figures 5 and 6 need to be analyzed individually since they are affected by specific conditions, among others, culture or governmental policies.

4.4 Change of the variables between 2000 and 2020

For the following analysis, it is important to note that the data set contains seven countries with missing values for the year 2000. For that reason, these countries are withdrawn from the

analysis. The focus of the following analysis is on the overall change, but some cases that can be pointed out from the graphs are presented. Moreover, for a better understanding of the information all the countries are plotted as a point belonging to a subregion (same colors). Also, some reference lines, that visualize the degree of variation from 2000 to 2020, are depicted. If a point is above the continuous line in the middle, the value of the variable that is analyzed has increased, while a point below the continuous line indicates that the value of that country has decreased.

Figure 7 displays the change of the Total Fertility Rate between 2000 and 2020. It is possible to see that the Total Fertility Rate of many countries in African subregions has decreased between 40% and 20%, 1 or 2,6 children per woman. Furthermore, it can be observed that even though many European countries have experienced a relative change of their Total Fertility Rate of around 20% between 2000 and 2020, the absolute change is small. The Total Fertility Rate for most of the countries in Americas, Asia and Europe has varied between 2.5 and 1.5, a range in which the medians of both years are located.

In Figure 7, the extreme countries that have either the highest or the lowest Total Fertility Rate in 2000 or 2020 are labelled. For example, Hong Kong had the lowest value for 2000 but it has increased over the last 20 years. Niger and Afghanistan are the countries with the highest Total Fertility Rate in 2000 but they have decreased since then by almost 15% in the former and by around 40% in the latter case.

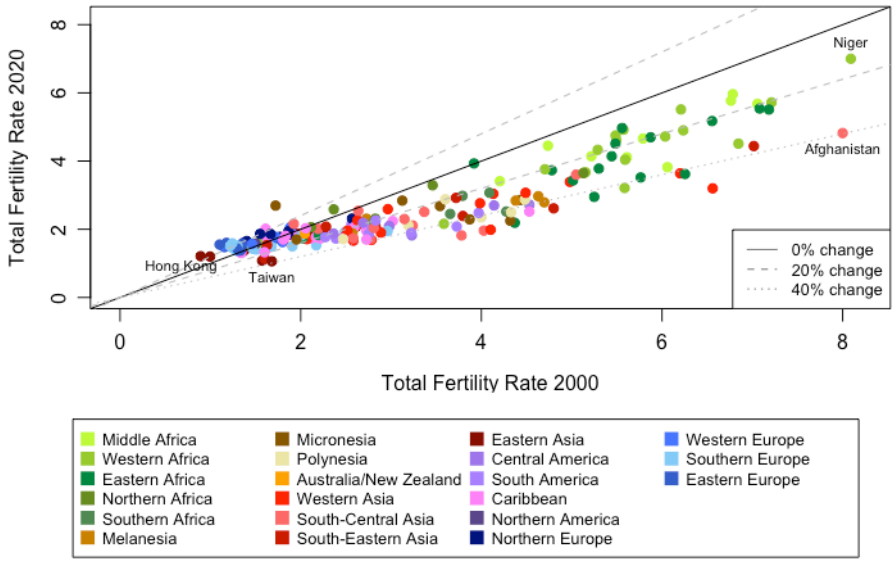

Figure 7. *Change in Total Fertility Rate between 2000 to 2020 by country.*

Figure 8 represents the change of the Life Expectancy at Birth of Both Sexes between 2000 and 2020. It shows that the Life Expectancy of at least 98% of the countries has increased over the last 20 years. The most pronounced increases belong to African countries with percentages over 20%, which represents between 10 and 15 years of Life Expectancy for both sexes. It can also be observed that, in general, all regions experience an increase of Life Expectancy at Birth of Both Sexes. For a vast majority of the countries Life Expectancy of Both Sexes increase between 2 and 15 years, reaching a total between 70 and 85 years in 2020. The median increase among the countries is 4.8 years. The change of the Life Expectancy at Birth for Males and Life Expectancy at Birth for Females between 2000 and 2020 is plotted in Figures A3 and A4.

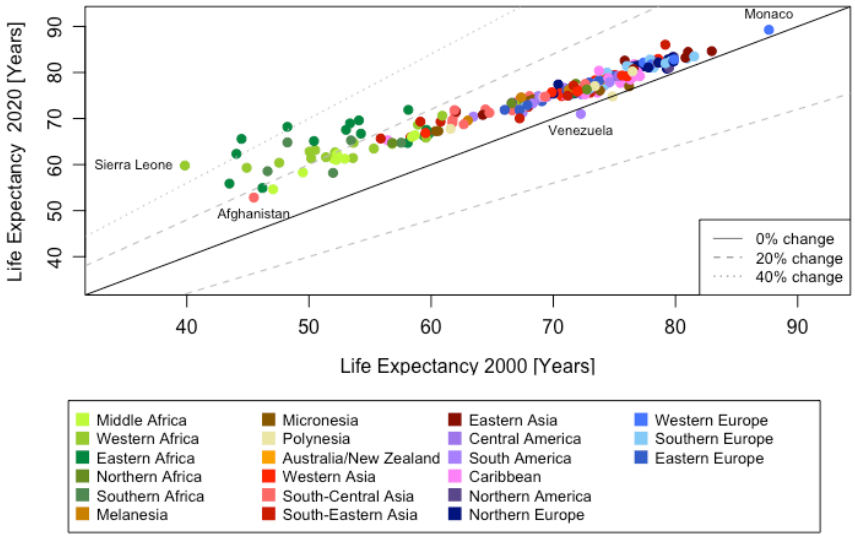

***Figure 8**. Change in Life Expectancy of Both Sexes between 2000 and 2020 by country.*

In Figure 8 Monaco can be identified as the country with the highest life expectancy for both years, while Zambia, Uganda and Sierra Leone are the countries with the largest absolute and relative increase of around 20 years from 2000 to 2020. The country in which the Life Expectancy of Both Sexes has decreased is Venezuela. Furthermore, Afghanistan is an extreme value that is characterized by a far lower life expectancy than the other countries in the South-Central Asian subregion.

5. Summary

With the International Data Base (IDB) the U.S. Census Bureau provides population estimates and projections for over 200 countries and areas of the world on their web page. The data set used in this report is a subset of the IDB. The subset includes demographic and geographical information about 221 countries in the years 2000 and 2020. Regarding the geographic

information it provides the region, subregion, and name of each country. Furthermore, two types of country codes are given. The variables of the demographic information are Total Fertility Rate, given in the average number of children per woman, and Life Expectancy at Birth of Both Sexes, Life Expectancy at Birth of Males, respectively Females given in years.

To understand the distribution of the Life expectancy at Birth and the Total Fertility Rate in 2020 for the group of 221 countries, the relative frequency histogram of the variables is described. In addition, the difference of Life Expectancy at Birth between Males and Females is analyzed. The Pearson correlation coefficient is used to obtain the bivariate correlations of the demographic variables. Other factors that are compared for the year 2020 are the variables Life Expectancy at Birth of Both Sexes and Total Fertility Rate within the subregions and between them. This comparison is presented using multi-box plot diagrams. Finally, the change of Life Expectancy at Birth of Both Sexes and Total Fertility Rate between 2000 and 2020 is evaluated with a direct graphical comparison.

The relative frequency distribution for Total Fertility Rate for 2020 is positively skewed with low variability and presents many extremely high values. At least 50% of the countries has a total fertility rate that is smaller or equal to 2 children per woman. In contrast, Life Expectancy at Birth of Both Sexes in 2020 is slightly negatively skewed, has a high variation and a median of 75.5 years. Total Fertility Rate and Life Expectancy at Birth of Both Sexes are negatively strongly correlated. For the difference of Life Expectancy between Males and Females is found that in more than 95% of the countries women live longer than men and in 50% of the countries women have at least 4.9 years more life expectancy.

The variability of the Total Fertility Rate and Life Expectancy for both sexes depends on the subregion. For example, for Middle Africa the variability shows a heterogenous behavior but for Eastern Europe it shows a homogeneous behavior. The Total Fertility Rate shows homogeneity within regions, while Life Expectancy at Birth is heterogeneous within regions.

The Total Fertility Rate of most of the countries in African subregions has decreased between 40% and 20% between 2000 and 2020, which represents an absolute change between 2 and 2.5 children. Even though European countries have also experienced a relative change of their Total Fertility Rate of around 20% between 2000 and 2020, their absolute change is comparatively small, namely between 0.3 and 0.5 children. Furthermore, the Life Expectancy of at least 98%

of the countries has increased over the last 20 years. The absolute increase of Life Expectancy between 2000 and 2020 ranges between 2 and 15 years, with a median value of 4.8 years.

For further analysis, it is recommended to evaluate the reasons for the high linear dependency between Total Fertility Rate and Life Expectancy. Furthermore, finding relations to other variables like child mortality rates, poverty rates, social conditions or economic development could lead a better understanding of the demographic data. In addition, a metric that could complement this analysis is the population of each country because with this it is possible to assign weights to the variables and get more representative findings for regions and subregions.

Bibliography

Sharpe, N. R., De Veaux, R. D. & Velleman P. F. (2012). *Business Statistics, 2nd Edition*. United States of America Pearson Education, inc.

United States Census Bureau (2013). Demographic and Economic Analysis. URL: https://www2.census.gov/programs-surveys/international-programs/about/factsheets/factsheet6-analysis.pdf (Visited on 12/05/2021)

United States Census Bureau (2017). Glossary. URL: https://www.census.gov/programs-surveys/international-programs/about/glossary.html (Visited on 12/05/2021)

United States Census Bureau (2020). *International Data Base: Population Estimates and Projections Methodology.* URL: https://www2.census.gov/programs-surveys/international-programs/technical-documentation/methodology/idb-methodology.pdf (Visited on 12/05/2021)

United States Census Bureau (2021). *International Data Base*. URL: https://www.census.gov/programs-surveys/international-programs/about/idb.html (Visited on 12/05/2021)

Appendix

Appendix A Additional figures

Min	Q1	Median	Mean	Q3	Max
51.350	67.255	72.950	71.529	76.650	85.400

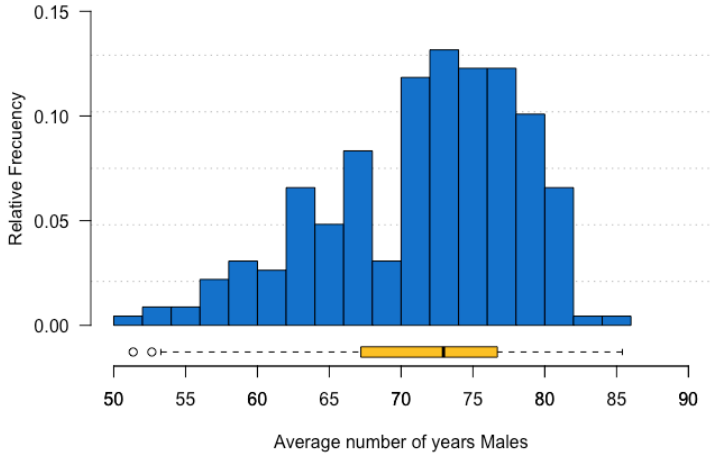

Figure A1. *Relative Frequency Histogram Life Expectancy at Birth of Males 2020.*

Min	Q1	Median	Mean	Q3	Max
54.410	71.915	78.350	76.666	82.205	93.300

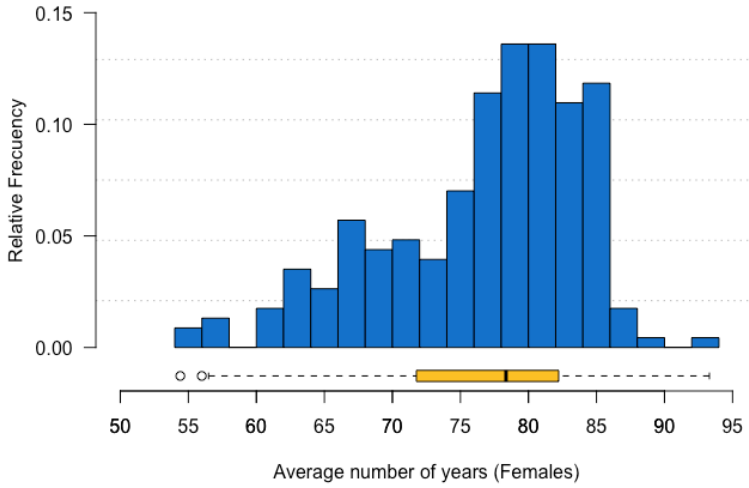

Figure A2. *Relative Frequency Histogram Life Expectancy at Birth of Females 2020.*

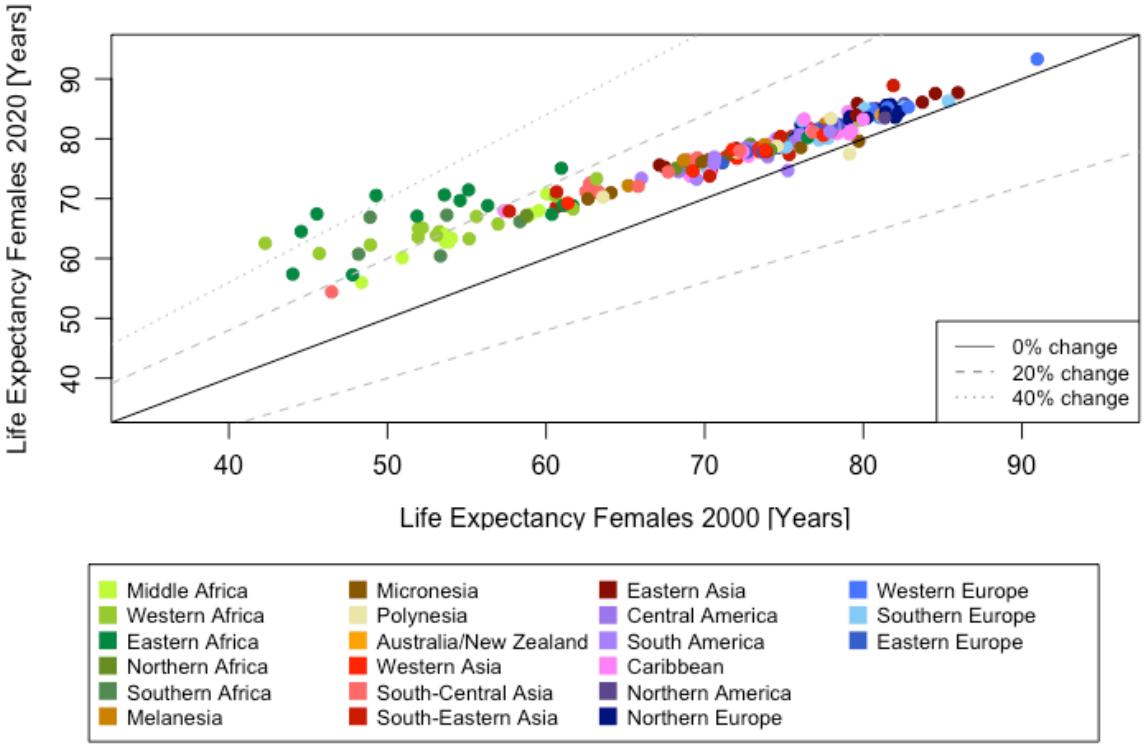

Figure A3. *Change in Life Expectancy at Birth Females between 2000 and 2020 by country.*

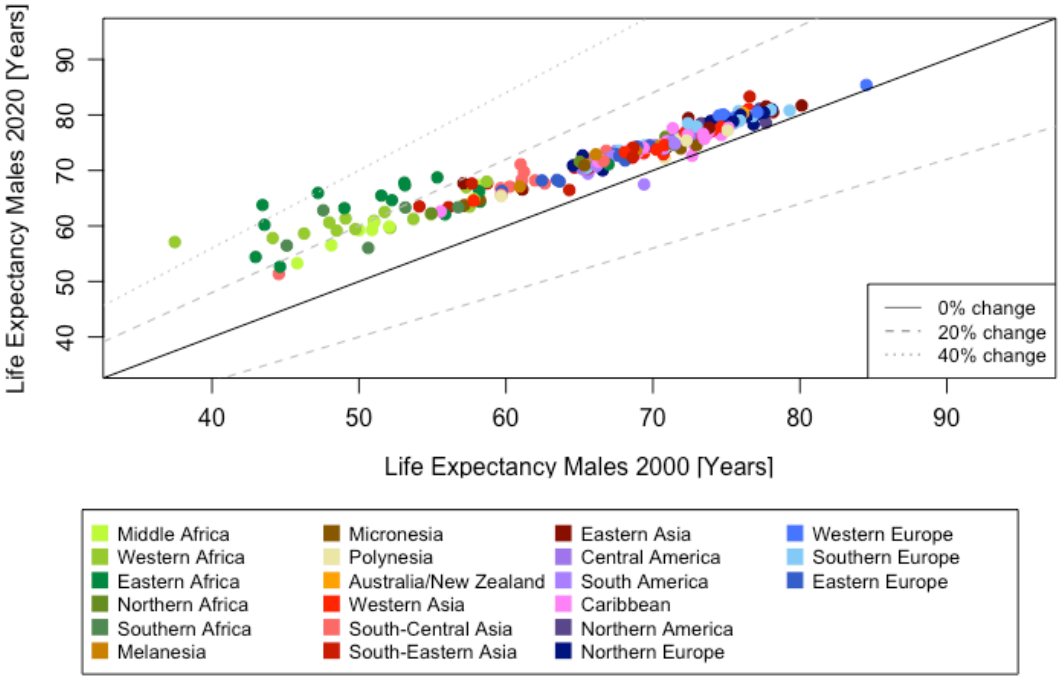

Figure A4. *Change in Life Expectancy at Birth Males between 2000 and 2020 by country.*

Appendix B Additional tables

Table B1. *Number of countries by subregion.*

Subregion	# of countries	Subregion	# of countries
Central America	7	Eastern Europe	10
Northern America	4	Northern Europe	14
South America	12	Southern Europe	16
South-Eastern Asia	11	Eastern Asia	8
Micronesia	7	South-Central Asia	14
Caribbean	24	Eastern Africa	17
Western Africa	17	Western Asia	18
Australia/New Zealand	2	Polynesia	7
Northern Africa	5	Melanesia	5
Southern Africa	5	Western Europe	9
Middle Africa	9		

Table B2. *SD and IQR by region for Life Expectancy at Birth of Both Sexes 2020 and for Total Fertility Rate 2020.*

Region	Subregions	SD Life Expectancy 2020	IQR Life Expectancy 2020	SD Total Fertility Rate 2020	IQR Total Fertility Rate 2020
Africa	Northern Africa	5.025	3.020	0.670	0.980
Africa	Middle Africa	4.397	5.010	0.925	1.570
Africa	Western Africa	5.439	4.420	1.282	1.230
Africa	Southern Africa	3.602	6.260	0.363	0.530
Africa	Eastern Africa	5.597	3.840	1.182	1.267
America	Caribbean	3.504	3.835	0.260	0.296
America	South America	3.106	5.043	0.230	0.328
America	Central America	2.560	3.390	0.360	0.483
America	Northern America	4.411	2.978	0.205	0.343
Asia	South-Central Asia	5.905	4.560	0.851	0.511
Asia	Western Asia	3.461	3.530	0.699	1.249
Asia	South-Eastern Asia	6.012	7.060	0.878	0.736
Asia	Eastern Asia	5.688	8.610	0.359	0.510
Europe	Southern Europe	3.125	4.993	0.186	0.115
Europe	Western Europe	2.488	0.840	0.186	0.257
Europe	Eastern Europe	2.692	4.438	0.076	0.094
Europe	Northern Europe	2.590	1.305	0.207	0.274
Oceania	Polynesia	3938	2.755	0.465	0.710
Oceania	Australia/New Zealand	0.396	0.280	0.097	0.068
Oceania	Melanesia	3.263	2.460	0.454	0.530
Oceania	Micronesia	3.954	4.490	0.420	0.497