Suicidal World

Norbert Hartkamp, M.D., M.Sc.

21.11.2022

Contents

1 Executive Summary

An analysis is done of worldwide data on suicide rates as part of the HarvardX PH125.9x data science-course, capstone-module.

Suicide a relevant health issue that is repeatedly adressed e.g. by the WHO.

A publicly available data set from *Kaggle* being too fragmentary, a analysis data set from various sources (Our World in Data, WHO, PEW research and others) was compiled.

The time trends of suicidality in different entities (i.e. nations) were first analyzed using cluster analysis. In a second step determinants of suicidality were investigated using bi-variate and multiple correlation analysis.

The results of cluster analysis only show partial accordance with results reported in the literature.

The bi-variate and multivariate correlations show some accordance with findings from the literature, although it remains unclear, how the results are affected by a sometimes limited quality of the primary data, from which the data set for the present analysis was compiled.

2 Introduction and Overview

There is little doubt among scientists (Preti, 2011) that the ability to intentionally end one's own life is uniquely human. Although it apparently defies the logic of evolutionary biology (Chatterjee & Rai, 2021) there is each year an estimated number of more than 700,000 people dying from suicide, with suicide being a leading cause of death particularly in the younger population (Ritchie et al., 2015).

Although data on suicide are published widely on the net (see *Suicide Worldwide in 2019*, n.d.), in the present analysis I attempt at describing patterns and trends in suicide data, hoping to be able to show some of the relevant patterns in a concise way. It has to noted however, that the standards of data reporting in this field are not optimal. As Bertolote & Fleischmann (2002) note, developed countries report these data mostly on a regular basis, while most developing countries from Latin America, Asia, and the East Mediterranean report only irregularly, and only very few countries from Africa do report on mortality to the WHO at all.

What I obviously cannot do in my data based analysis, is to focus on the existentially important questions of the psychopathology of suicidality (Kulawik, 1977; Little, 2009; Maltsberger, 2004; Modestin, 1990), the individual developmental preconditions of the development of suicidal tendencies (Hardt et al., 2007), the question how suicidal tendencies can be prevented (De Leo, 2002; Hegerl et al., 2019; Taylor et al., 1997; Zalsman et al., 2017), or how suicidality can or should be treated medically (Althaus & Hegerl, 2004; Hegerl, 2005; Wolfersdorf, 2008).

2.1 R libraries being used

The following libraries will be used in the present project:

```
library(knitr)
library(data.table)
library(tidyverse)
library(RSelenium)
library(stringi)
library(ggthemes)
library(gridExtra)
library(viridis)
library(owidR)
library(lubridate)
library(kableExtra)
library(psych)
library(latex2exp)
library(rvest)
library(pdftools)
library(openxlsx)
library(readxl)
library(fpc)
library(TSclust)
library(factoextra)
library(hopkins)
library(clValid)
library(ggpubr)
library(compiler)
```

2.2 Data ingestion

The data on which the present analysis is based are provided by Kaggle, which since 2017 is a subsidiary of Google. As a corresponding Wikipedia article states, it is the aim of Kaggle to provide users, mainly data scientists and machine learning practitioners, a platform from which publicly available data can be downloaded to do data analyses, which then even can be subject to a competition of multiple users over who does data analysis best. However, the present analysis does not attempt to be part of such an endeavor, but is part of the fulfillment of the requirements of the HarvardX PH125.9x course on Data Science.

The data used in the present analysis can be found under the heading of Suicide Rates Overview 1985 to 2016 on the Kaggle website as a zipped comma separated file.

3 Methods

The methods I'll be using are first and foremost *descriptive*. Furthermore, I will try to illustrate the data with geographical and statistical figures. I'll also use classification methods, and methods to reduce complexity. Where possible and appropriate, I will attempt to make connections between the data on suicide incidence and other relevant data that are publicly available.

4 Describing the *Kaggle*-data

In the *Kaggle*-data read in from the CSV-file there are a total of 27,820 entries, each comprising of 12 variables. The variables are:

(1) **country** with *101* alphabetically sorted entries ranging from *Albania* to *Uzbekistan*,

(2) **year** ranging from *1985* to *2016*,

(3) **sex** with the two entries *male* and *female*,

(4) **age**, where ages are grouped in the different classes *5-14*, *15-24*, *25-34*, *35-54*, *55-74*, *75+* years,

(5) **suicides_no** indicating the number of suicides in the respective subgroup (country, year, sex, age-group),

(6) **population** indicating the population size in the respective subgroup,

(7) **suicides.100k.pop** suicide rate standardized per population size of 100,000

(8) **country.year** indicating country and year, an aggregated variable following the pattern *Nnnnnnnnnnyyyy*, which might be useful for grouping the data

(9) **HDI.for.year** with the Human Development Index being "a summary measure of average achievement in key dimensions of human development" (Nations, n.d.)

(10) **gdp_for_year** indicating the Gross Domestic Product for the respective country and year

(11) **gdp_per_capita** indicating the GDP per person for the respective country and year

(12) **generation** a characterization of "generations" according to an age-group (which may or may not be useful in discussing worldwide data from different cultural backgrounds). The terms used here are *Generation X*, *Silent*, *G.I. Generation*, *Boomers*, *Millenials*, and *Generation Z*.

4.1 General Overview

Over the whole time span covered by the data, there were a number of 6,748,420 suicides reported. The mean annual number of which ($\overline{m}$ = 210,888.1) is showing a sharp decline (as shown in fig. 1) in the last reported year 2016, which most likely is due to an incompleteness of the data from that year. As a consequence the data from year 2016 are being excluded from the further analysis. Also the lower numbers from before 1990 might indicate an incompleteness of data related to the countries of the former *Eastern Europe* and the *Soviet Union* having not provided their suicide data.

The following fig. 2 shows that for the *Russian Federation* there are no data from before 1989 (when the *Soviet Union* still existed), and there is also no country *Soviet Union* in the *countries* list.

The same applies to *Germany*, where no data from before 1990 (year of German unification) are found in the Kaggle data set, or e.g. for *Romania* where no data exist from before 1989, which was the year of the Romanian Revolution.

Macau is another special case: Macau was formerly a Portuguese colony, and was transferred to China in 1999 (see: "Macau," 2022). In the data set there are no data in suicide in China. In the case of *Macau* there are only data available for one single year (*1994*), and because of this the *Macau*-data were excluded from the further analyses.

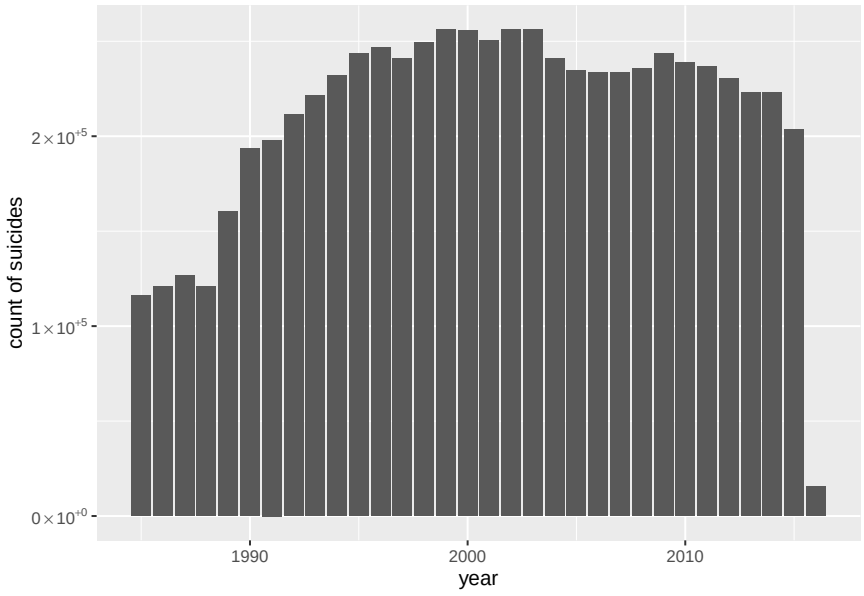

Figure 1: average number of suicides per year

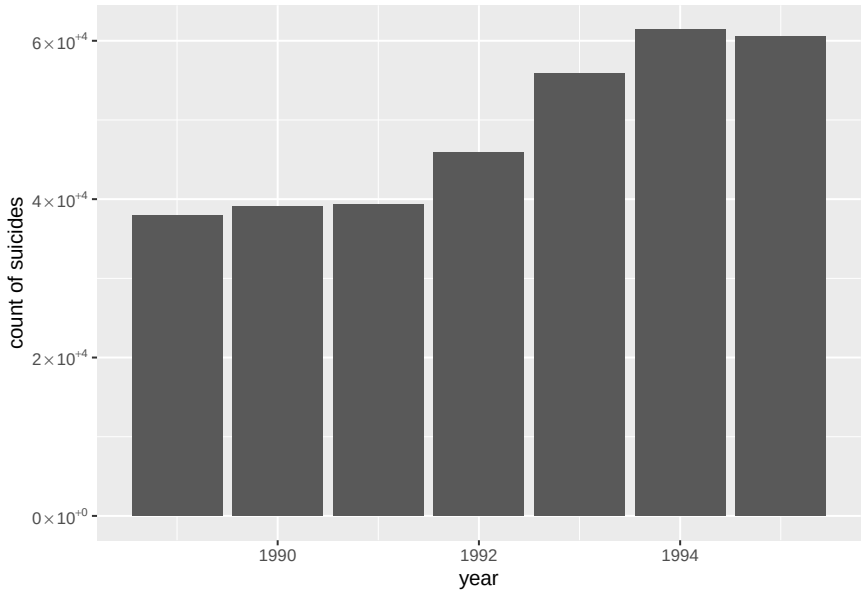

Figure 2: mean suicides/year in Russ. Fed. (partial view)

In summary, due to the incompleteness of the data, it at first seemed safe to **restrict** the analyses to the time span from **1990** to **2015**, so that the analysis would be based on 24,636 entries in the data set.

As the following map (fig. 3) shows, there are still many countries from which data are missing, which of course is a grave limitation to the analysis to be made.

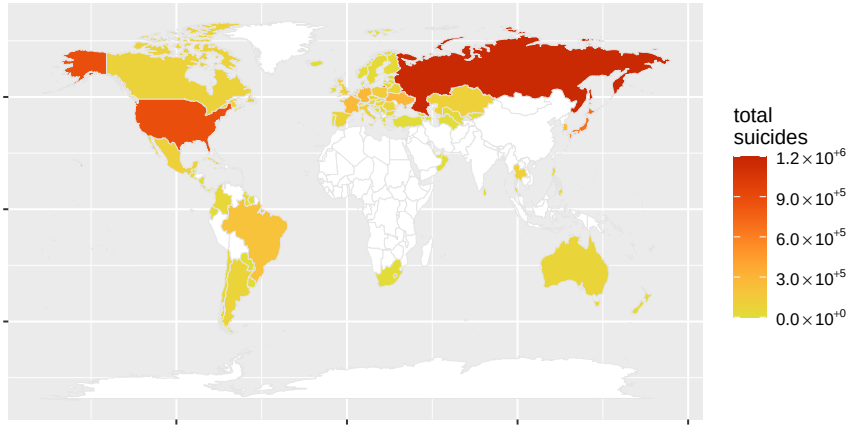

Figure 3: map of total suicide numbersin different countries according to kaggle

4.2 Turning to data from a different source

The incompleteness of the data being unsatisfactory, there was the need to look for data from a different source. The online publication *Our World in Data* (n.d.) is one such source of data. As an respective Wikipedia article states, Our World in Data (OWID) focuses on large global problems such as poverty, disease, hunger, climate change, war, existential risks, and inequality.

OWID is a charity, founded by Max Roser, a social historian and development economist. Its research team is based at the University of Oxford. OWID-data have been used in academic scientific journals and major newspapers of the world.

The recently published published owidR-package (York, 2022) provides a handy R-interface to the OWID-data. Using the functions of the owidR-package it is possible to retrieve a number of data sets on the topics listed in table 1.

Table 1: topics covered in the OWID tables on suicide

Topics
Suicide rate vs. death rate from violence
Suicide rate vs. homicide rate
Suicide death rate vs. prevalence of mental & substance use disorders
Suicide rates vs. prevalence of depression
Share of suicide deaths from pesticide poisoning
Death rate from suicides
Deaths from suicide, by age
Male vs. female suicide rate
Male-to-female ratio of suicide rates
Number of suicide deaths from firearms
Share of deaths from suicide
Suicide death rate by age
Suicide death rate by sex
Suicide death rates
Suicide rate
Suicide rate by firearm
Suicide rate in 1990 vs. 2019
Suicide rate vs. income inequality
Suicide rates by age

In addition to the *suicide*-related data other from OWID were used as well. These Topics of these data (mainly epidemiological and societal measures) are presented in table 10 in the Appendix.

All OWID-data have three fields in common: *"entity"*, *"code"*, and *"year"*, with *"entity"* designating the country or region from which the data were collected, *"code"* designating a country-code (where applicable), and *"year"* designating the year in which the respective data were collected. Among the *"entities"* there are a number of items that are already aggregated according to criteria e.g. of WHO or WorldBank. When doing analyses based on countries, these records have to be excluded from the OWID-tables.

Overall the *OWID*-data seem to cover more regions of the world than the *Kaggle*-data, as shown in figure 4. However a small number of countries that are in the OWID-data, but not in the world-map provided with ggplot (*Eswatini, Tokelau, Tuvalu*) had to be excluded from the plot[1].

4.3 Sources and date range of the OWID-data

The **OWID**-website adds to each chart or table it provides information on the **source of the data**. This information is accessible using a **"SOURCES"**-tab on each of the respective pages where OWID presents its data[2]. The tabulated info on the *sources* of the OWID-data extracted from the OWID-website are to be found in the Appendix (see Ch. 9 of this paper).

From this it can be seen that the time scale of the data in particular is very different. E.g. data for *Male-to-female ratio of suicide rates* range from **1990** to **2017**, while data on *historical estimation of population*

[1]It might be noteworthy, that for a number of countries no suicide-data are given in the OWID-data set. These are: *Anguilla, Antarctica, Aruba, Ascension Island, Azores, Bonaire, Canary Islands, Cayman Islands, Chagos Archipelago, Christmas Island, Cocos Islands, Curacao, Falkland Islands, Faroe Islands, French Guiana, French Polynesia, French Southern and Antarctic Lands, Guadeloupe, Guernsey, Heard Island, Isle of Man, Jersey, Kosovo, Liechtenstein, Madeira Islands, Martinique, Mayotte, Montserrat, New Caledonia, Norfolk Island, Pitcairn Islands, Reunion, Saba, Saint Barthelemy, Saint Helena, Saint Martin, Saint Pierre and Miquelon, Siachen Glacier, Sint Eustatius, Sint Maarten, South Georgia, South Sandwich Islands, Swaziland, Turks and Caicos Islands, Vatican, Wallis and Futuna, Western Sahara.*

[2]As the OWID-website relies heavily on the use of Javascript driven technologies that dynamically change the structure of the web pages, it was not possible to just use e.g. `rvest` to scrape this information from the OWID-website. Instead it was necessary to use a *browser* (in the present case *chrome*) remotely administered from *R* using the *Selenium WebDriver* and the `RSelenium`-package. The data were then imported into *R* using the functions `minimal_html()`, `html_elements()`, and `html_text()` from the `rvest`-package.

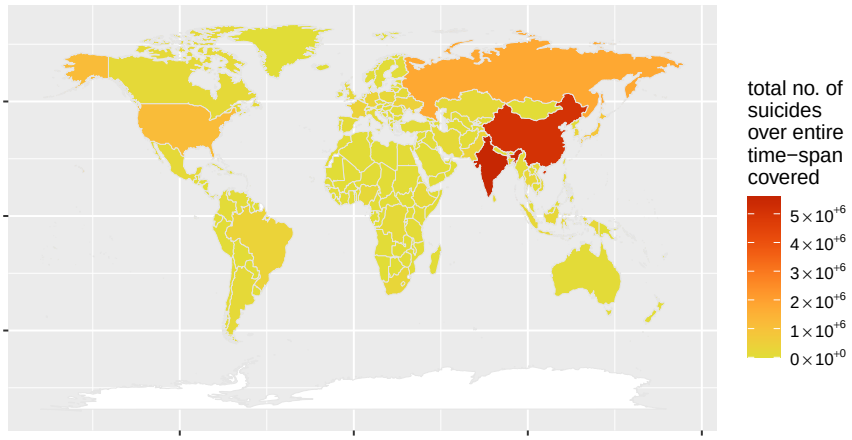

Figure 4: map of OWID data on total suicide numbers worldwide

(relevant within the context of suicide rates vs. prevalence of depression) range from **10,000 BCE** to **2021**. Data for *share of suicide deaths from pesticide poisoning* are available for only **one** year (*2014*).

So, for further analysis, I will restrict the time interval to consider to the range of **1990** to **2017**.

4.4 Additional data sources

Besides the OWID-data other data sources have been used, as the data from these sources were more complete than those found in the OWID-web space.

Population data were collected from the web site of the United Nations Population Division. This site holds both historical data and future projections of multiple population related variables. In the present project *total* population, *male*, and *female* population, each as of July, 1st of each year were included.

Data on the *Human Development Index* were acquired from the web site of the United Nations Human Development Reports. "The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living." (Nations, n.d.) The HDI is a composite measure of a *health*-dimension, a *standard of living*-dimension, and an *education*-dimension.

Also included is the *Gender Inequality Index* (GII) from the same source. This index encompasses measures of *reproductive health* (including maternal mortality ratio and adolescent birth rate), *empowerment* (e.g. share of women with at least secondary education) and the *labor market* (e.g. female labor force participation rate).

The Gross Domestic Product per capita is acquires from the Maddison Project web site of the Faculty of Economics and Business from Groningen/NL University (Bolt & van Zanden, n.d.).

Though a measure of *People Living in Poverty* is available from the OWID-website, a more complete data set is available via the OWID github. From this source the data on the percentage of people living on less than 1$/day, and the percentage of people living on less than 10$/day were extracted (both based on a 2017 purchasing power parity). From the same source the *Gini* coefficient was extracted.

A measure of religious commitment (which might be of relevance to the rates of suicide) is available from Pew Research. The relevant data can be found in the "Complete Report", from where the data had to be extracted using the *R* `pdftools`-library.

Multinational data on people's *happiness* can be found in the World Happiness Report. Levels of *happiness* might be of relevance to the analysis of suicidality. The relevant data are available from the web site of the World Happiness Report as a spreadsheet-file[3]. The *happiness* is measured by means of a *life ladder*, where a rung *0* represents *the worst possible life*, and the rung *10* represents the *best possible life* for the informants on whose interviews the *happiness index* is based (see: (Cantril, 1965)). Typically around 1000 informants had been interviewed from each country in the database to determine the happiness index for the respective country.

Data on shares of expenditures for primary care physicians (who in many cases of suicidality are likely to be among the first to whom patients or their relatives turn to) are part of the the WHO Global Health Expenditure Database. Unfortunately these data are too incomplete to be useful in the present analysis.

Some of the data are not covering the whole time span of the present analysis. E.g. for *Albania* the happiness-(*life-ladder*-) data are not available before *2007* and not for the year *2008*. To handle this, *imputations* of the missing values were calculated using *linear modeling*. This imputation was done with missing values from *HDI*, *GII*, *life ladder*, *positive affect*, *negative affect*, *less_1_ratio*, *less_10_ratio*, and *gini* variables. However it is not known, if *linear modeling* in all these cases is an adequate method for imputation, this methodological question going beyond the scope of this paper.

For the purpose of analyzing the data on the level of *entities* (e.g. *nations*), while leaving the *years* out of consideration, a condensed data set was calculated holding the means of the relevant variables after grouping for entity.

4.5 The data set for the present analysis

In figure 4 it could be seen that the data set compiled for the present analysis covers more regions of the world than the Kaggle data set did. However, the total number of suicides in a country over a certain period of time does not give a meaningful picture.

A better overall picture is obtained by measuring the total number of suicides against the population of the respective country. In the present analysis this is done by averaging the numbers of suicides per year and also the population of the countries over the observed time interval to calculate suicide mortality rates per 100,000 population.

At a first glance some countries and world regions apparently have quite high, while others have quite low rates of suicides. The question however is, how these differences might be explained. In the following it'll be tried to identify characteristics of countries that are related to either high or low suicide rates.

Before that can be done, it is useful to take a look at the data as they have been compiled for the present analysis. The following table 2 lists variable names and their meaning.

[3]There is a difference between the happiness-data from OWID and the data from the WHR 22. OWID apparently uses the wrong *years*, as can be seen from the first rows of the resp. tables:

OWID					WHR 22		
Entity	Code	Year	Life satisfaction in Cantril Ladder		Country name	year	Life Ladder
Afghanistan	AFG	2006	3.72358989715576		Afghanistan	2008	3.72358989715576
Afghanistan	AFG	2007	4.40177822113037		Afghanistan	2009	4.40177822113037
Afghanistan	AFG	2008	4.75838088989258		Afghanistan	2010	4.75838088989258
Afghanistan	AFG	2009	3.83171916007996		Afghanistan	2011	3.83171916007996

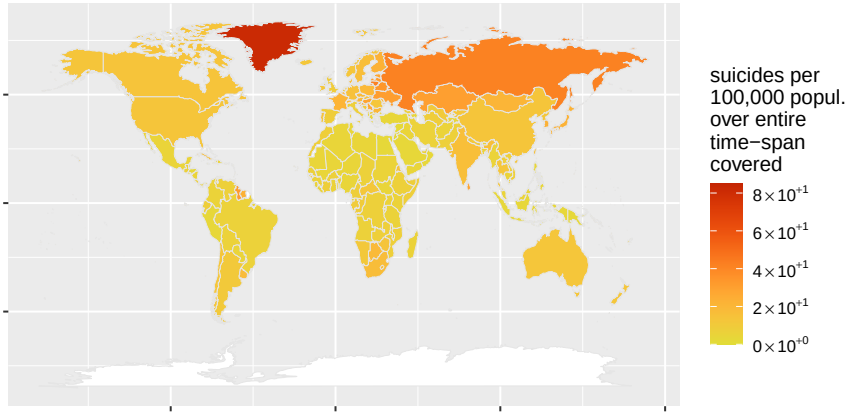

Figure 5: map of OWID data on suicide rates / 100,000 population worldwide

Table 2: variable names in the analysis data set and their descriptions

var label	description
entity	country, or group of countries
age70plus	suicides in the age group 70+ years
age50__69	suicides in the age group 50 to 69 years
age15__49	suicides in the age group 15 to 49 years
age5__14	suicides in the age group 5 to 14 years
yearly__sum	sum of suicides in the resp. year
population	population on July, 1st, from UN Population Division
m__population	male population on July, 1st, from UN Population Division
f__population	female population on July, 1st, from UN Population Division
popage65plus	population of age 65+ on July, 1st, from UN Population Division
popage25to64	population of age between 24 and 64 on July, 1st, from UN Population Division
popage15to24	population of age between 15 and 24 on July, 1st, from UN Population Division
popage5to14	population of age between 5 and 14 on July, 1st, from UN Population Division
popage0to4	population of age between 0 and 4 on July, 1st, from UN Population Division
median__age	population's median ag on July, 1st, from UN Population Division
m__suic__rate	male suicides per 100,000 population
f__suic__rate	female suicides per 100,000 population
suic__rate	overall suicides per 100,000 population
mf__slfhrmdeath__age__stand	rate of deaths by self-harm, age-standradized
mf__ipvioldeath__age__stand	rate of detah from interpersonal violence, age-standradized
mf__prev__md__s__ab__age__stand	prevalence of mental diseases and substance use disorders, both sexes, age-standradized
m__prev__adhd	prevalence of ADHD (males)
f__prev__adhd	prevalence of ADHD (females)
m__prev__alco	prevalence of alcohol related disorders (males)
f__prev__alco	prevalence of alcohol related disorders (females)
m__prev__depr	prevalence of depressive disorders (males)
f__prev__depr	prevalence of depressive disorders (females)
m__prev__anxiety	prevalence of anxiety disorders (males)
f__prev__anxiety	prevalence of anxiety disorders (females)

(Continued on Next Page...)

Table 2: variable names in the analysis data set and their descriptions *(continued)*

var label	description
m__prev__bipolar	prevalence of bipolar disorders (males)
f__prev__bipolar	prevalence of bipolar disorders (females)
m__prev__schizoph	prevalence of schizophrenia (males)
f__prev__schizoph	prevalence of schizophrenia (females)
hdi	Human Development Index, from United Nations Human Development Reports
gii	Gender Inequality Index, United Nations Human Development Reports
gdppc	Gross Domestic Product per capita from the Maddison Project web site
dollars__per__day	calulated by dividing gdppc by 365
less__1__ratio	precentage of population living in households with an income or expenditure per person below $1 a day
less__10__ratio	percentage of population living in households with an income or expenditure per person below $10 a day
gini	measure of inequality between 0 and 1, with higher values indicating greater inequality
self__rep__trust	share of people agreeing with the statement “most people can be trusted”
all__veryimp	percentage of population who see religion as “very important” (all age groups)
life__ladder	Cantril ladder value, from World Happiness Report 2022 (WHR22)
positive__affect	laughter, enjoyment, and learning or doing something interesting on the previous day from WHR22
negative__affect	worry, sadness, and anger on the previous day from WHR22

5 Results of clustering of suicide time-series

The following figure 6 gives a first impression of the development of suicide rates over time. For this diagram the very small entities (e.g. *Antigua and Barbuda* with a population less than 100,000 people) were excluded. Among those small entities *Greenland* is a special case, as it started in 1990 with an excepionally high suicide rate of max. *95.23* in the year 1991, and a minimal rate of *62.58* in 2017. *Greenland's* suicide average rate from 1990 to 2017 of *82.21* is about 8 times higher than that of all countries averaged *except Greenland*, which is 10.88.

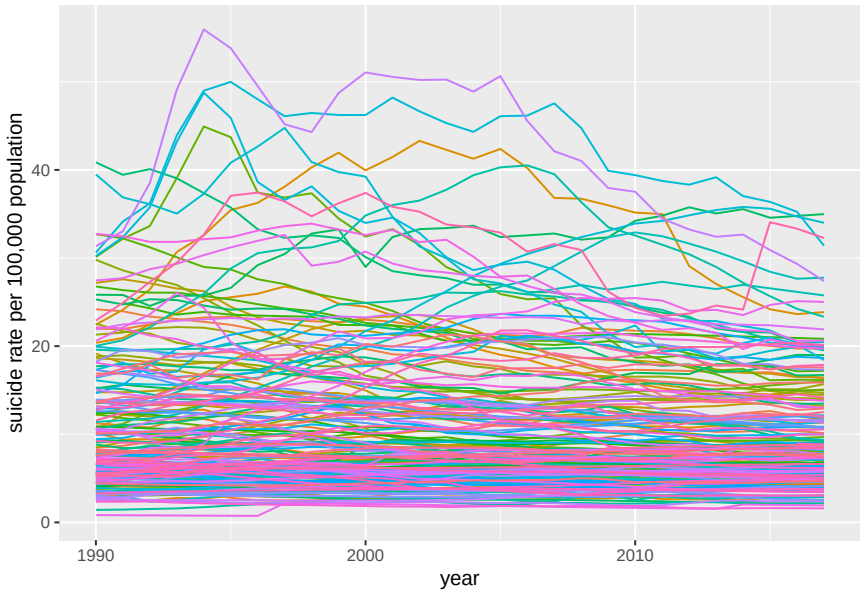

Figure 6: temporal development of rates of suicide per 100,000 population

From inspecting the graph a couple of things can be seen: (1) there seem to be a large proportion of entities in the lower range of *suicide rates* for which these rates seem to not change much, (2) there seem to be a handful of entities that show a *decrease in suicide rates*, partly with in intermediate *high*, (3) there also seem to be some entities in which an increase of *suicide rates* over time may be present. In the very low range there (4) also seems to exist entities for which ever other year a *zero-* or *close to zero*-value seems to have been reported, which might indicate so kind of reporting error.

Looking at a *violin plot* (fig. 7) of the suicide rates the marked asymmetry of the distribution of suicide-rates clearly can be seen. 63.2% of reported suicide rates are below the average suicide rate of 11.225338 (with the median = 8.11).

The question now arises whether it is possible to group (cluster) the trajectories of suicide rates objectively in a meaningful way. Such grouping can be done according to the general trajectory of the curves, their shape or their level. Methods for clustering time series data have been discussed e.g. by (Warren Liao, 2005) or (Rani & Sikka, 2012). (Warren Liao, 2005) point out that an application of time-series clustering methods typically has to be preceded by some data preparation, e.g. to extract features, which then can be attempted to cluster.

5.1 Is it possible to cluster the data?

This question can be answered by using the Hopkins statistic. It measures the cluster tendency of some data, and tests whether the data are uniform or randomly distributed or if they are clustered. The Hopkins

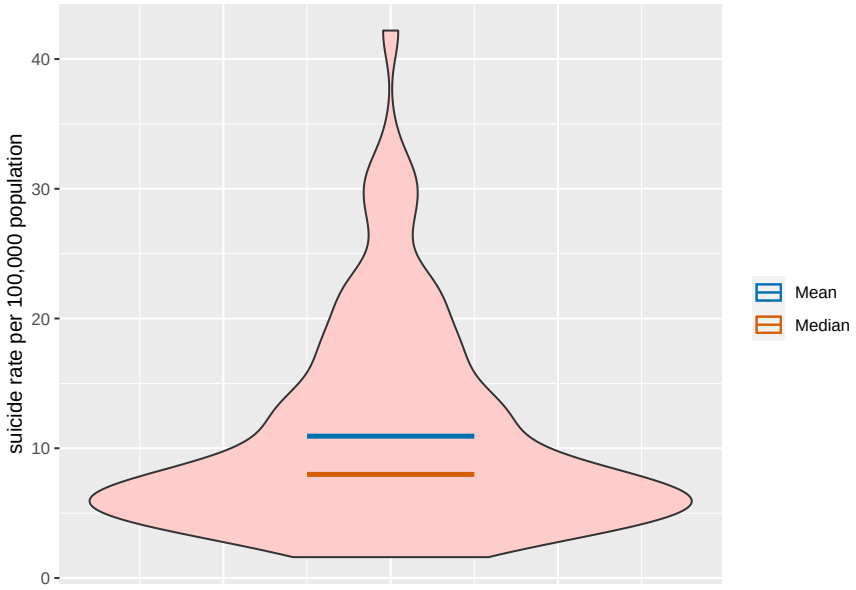

Figure 7: violin plot of suicide rates in different entities (> 100,000 population)

statistic can be calculated using the `hopkins` *R*-package. As the package's help-file states, a Hopkins statistic of **0.7 - 1.0** indicates, that data are clustered, and thus a cluster analysis is feasible.

For the *scaled* and *centered* set of complete suicide trend-lines a Hopkins statistic of **** is computed which indicates that a clustering of the data could work.

A visual inspection of the dissimilarity matrix confirms this: in figure 8 there seems to be quite some non-randomness in the ordering of the data which is a precondition for them to be clustered.

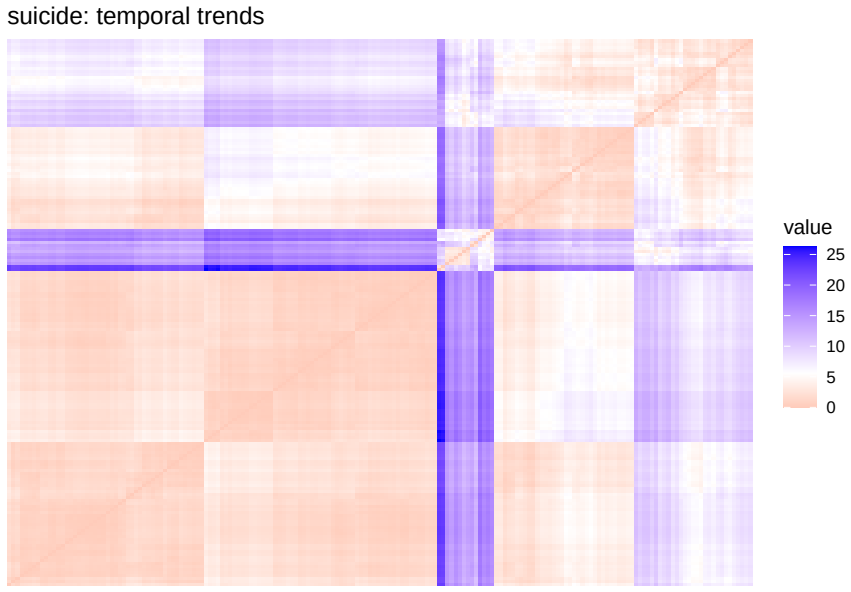

Figure 8: distance matrix of complete records showing non-randomness

5.2 What clustering method is best?

R provides a number of different functions for clustering. Frequently the *hierarchical* clustering, the *pam* clustering, and the *k-means* clustering methods are being used. Other clustering algorithms are **diana** (*divisive hierarchical clustering*), **clara** (*time-efficient clustering of large datasets*), or **fanny** (*fuzzy clustering*). The *clValid*-package (*clValid*, 2008) allows for validating the results of a clustering analysis. The following paragraph shows the output of the *clValid()* function:

```
##
## Clustering Methods:
##  hierarchical pam kmeans fanny
##
## Cluster sizes:
##  2 3 4 5 6 7
##
## Validation Measures:
##                                  2       3       4       5       6       7
##
## hierarchical Connectivity   7.9972 12.1468 13.4325 19.3401 23.9032 35.3254
##              Dunn           0.2140  0.2140  0.2140  0.0960  0.0984  0.1132
##              Silhouette     0.6621  0.6241  0.5906  0.5846  0.5820  0.5393
## pam          Connectivity   6.8913 25.8425 38.0984 40.8183 49.6587 55.6234
##              Dunn           0.0364  0.0199  0.0181  0.0220  0.0148  0.0177
##              Silhouette     0.5822  0.5178  0.3762  0.4013  0.3704  0.3717
## kmeans       Connectivity   7.2536 15.1262 24.7464 35.4944 36.1738 38.7579
##              Dunn           0.0584  0.0561  0.0382  0.0910  0.0603  0.0655
##              Silhouette     0.6448  0.5947  0.5214  0.5645  0.5297  0.5320
## fanny        Connectivity   6.8913 28.9933 38.4254 49.8198 57.5230 60.8389
##              Dunn           0.0364  0.0115  0.0120  0.0132  0.0154  0.0160
##              Silhouette     0.5822  0.4655  0.3755  0.3842  0.3495  0.3016
##
## Optimal Scores:
##
##              Score  Method       Clusters
## Connectivity 6.8913 pam          2
## Dunn         0.2140 hierarchical 2
## Silhouette   0.6621 hierarchical 2
```

From this analysis it seems that *hierarchical* and *pam* clustering methods are preferable. If in addition to this a ranking of the validation measures (using the `getRanksWeights()` and the `RankAggreg()` functions) is taken into account, then *k-means* also seems feasible.

```
## The optimal list is:
##         hierarchical-2 kmeans-2 hierarchical-3 pam-2 hierarchical-4
##
##   Algorithm:   CE
##   Distance:    Spearman
##   Score:       4.645767
```

So in the following all three of the above mentioned methods will be used.

In addition to this the *R*-package *TSclust* (Montero & Vilar, 2015) will be used, which provides a wide set of dissimilarity measures between time series. These dissimilarity measures can then be used for clustering

using model-free approaches, model-based approaches, complexity based approaches, and prediction-based approaches.

In the context of the current project the model-free approaches appear to be the most appropriate.

5.3 Preparing the data for clustering

Feature scaling will be used using the min-max normalization and the Z-standardization method.

Min-max normalization performs a linear transformation on the original data with the aim of getting all the scaled data into some range, typically between *zero* and *one*. This transformation is done by applying the formula

$$x' = \frac{x - min(x)}{max(x) - min(x)}$$

to each trend line. The transformation keeps both the trajectory and the shape of the original trend lines, while equalizing the different mean levels and ranges.

The standardization (Z-score Normalization) is done by applying the formula

$$x' = \frac{x - \bar{x}}{\sigma}$$

to each trend line, thus standardizing on a mean value of *zero* and a variance of *one*.

For the calculation of dissimilarity matrices and for the cluster analyses the data (using only *complete* cases) are transformed into a *wide* format such that each row represents a *entity*, and each column represents one *year*. To have a *"numeric-only"* matrix the first column (containing the country-codes) is being stripped. This information can be added later on for graphical presentation of the results (see following table 3).

Table 3: partial view of suicide rates-data transformed into a wide format

1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
0.79	0.96	1.00	0.87	0.76	0.71	0.7	0.68	0.61	0.54	0.54	0.61	0.48	0.40
0.02	0.21	0.15	0.13	0.00	0.18	0.3	0.61	0.83	0.73	0.55	0.50	0.67	0.86

The results of a first attempt at clustering the suicide trends using *hierarchical clustering* is shown in the next fig. 9. This procedure[4] yields rather unconclusive results, as to how many clusters should be extracted. From the graph a choice of both **two** or **three** clusters seems reasonable.

5.4 Determining the optimal number of clusters to choose

5.4.1 Elbow method

A frequently recommended procedure to determine the number of clusters to choose is the "elbow"-method, where an increasing number of clusters are extracted, and from a plot of the sum of squares against the number of clusters (fig. 10) it can be seen where the slope of the plot turns from "steep" to "shallow" (the so-called "elbow")[5]. Beyond this point the additional clusters do not add much to a meaningful ordering of the underlying variables. In the present case however a *two-cluster solution* seems favorable, as two clusters already extract more than 50 percent (51.6 %) of the total variance, although a *three-cluster solution* also seems applicable.

[4]The clustering based on *min-max normalization* and on *Z-standardization* lead to identical results, hence only the results of using min-max normalized data are presented here.

[5]Again, the results of *min-max normalization* and of *Z-standardization* lead to (in his case: almost) identical results, so only the results of using min-max normalized data are presented here.

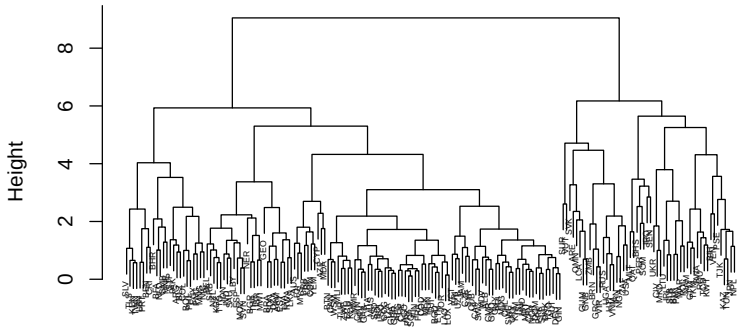

Figure 9: hierarchical clustering of suicide trends

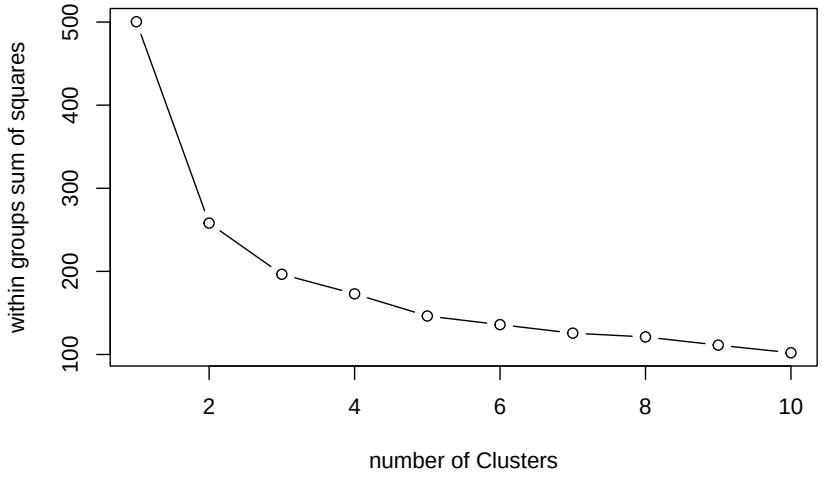

Figure 10: "elbow"-plot of number of clusters to extract

5.4.2 Gap-Statistic

The *gap*-statistic estimates the number of clusters by taking the output of any clustering algorithm and compare the change in within-cluster dispersion with that expected under an appropriate reference null distribution (see: Tibshirani et al., 2001). A maximized *gap*-statistic indicates that the clustering structure is as far as possible from a random uniform distribution of measurement points.

In the present case the *gap*-statistic for up to 5 clusters using *k-means* (*pam* is almost identical) results in:

```
##          logW   E.logW       gap     SE.sim
## [1,] 4.592996 4.877247 0.2842518 0.01361133
## [2,] 4.259743 4.737942 0.4781992 0.01534709
## [3,] 4.125604 4.659096 0.5334914 0.01271031
## [4,] 4.055335 4.599640 0.5443058 0.01366301
## [5,] 3.987806 4.560553 0.5727470 0.01188767
```

A plot of the results (fig. 11) reveals that using this technique a two-cluster solution appears to be the best, while a three-cluster solution also seems feasible, as it extracts even more variance without becoming too confusing.

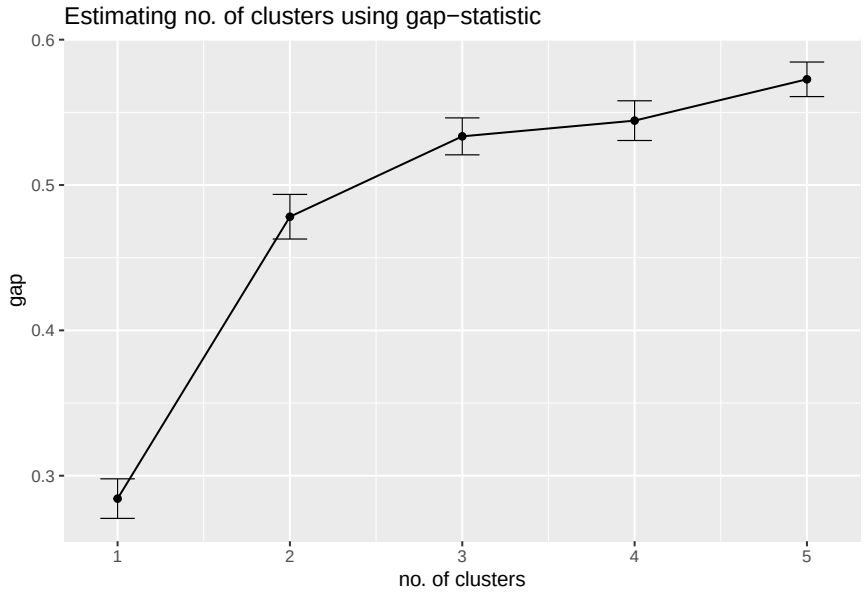

Figure 11: Plot of gap-statistic showing that two clusters are best

5.4.3 Partitioning around medoids (PAM)

The "elbow"-method being inexact and somewhat intuitive, there are other ways to determine the number of clusters to choose. One such method is *partitioning around medoids with estimation of number of clusters* (function `pamk()`) from the *fpc*-package. The method (Kaufman & Rousseeuw, 2009) aims at being a more robust (meaning: less sensitive to outliers) form of the k-means algorithm. *Medoids* can be understood as *centers of gravity*, where the data points are more grouped together and are more similar to each other (see: Manvailer, n.d.). Applying this procedure results in an estimated number of clusters of **two**.

However, an inspection of the clusters shows a considerable overlap of the two clusters (fig. 12 and fig. 13). A silhouette plot (fig. 14) of the clustering reveals that the two clusters are not vey well structured. According

to the criteria proposed by Kaufman & Rousseeuw (2009) (to be found in Manvailer (n.d.)), a mean S_i-value between **0.26** and **0.50** indicates a weak cluster structure, which could be artificial. In the present case the mean silhouette-width S_i is only **0.43**, which indicates a rather weak structure.

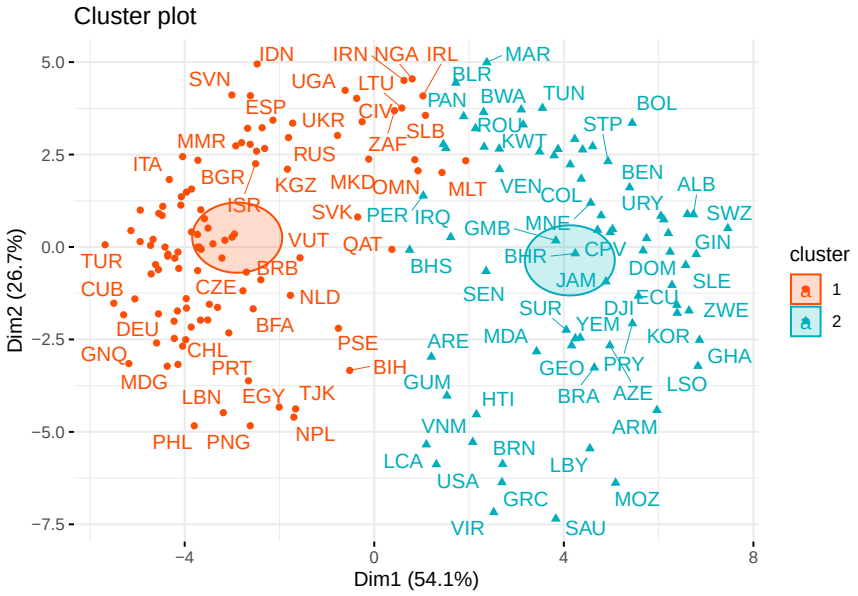

Figure 12: plot of clusters based on PAM

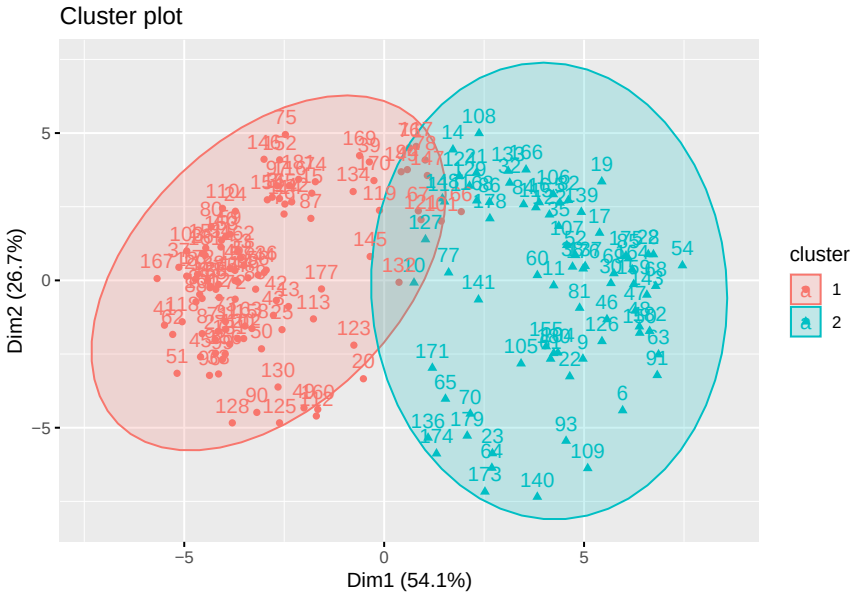

Figure 13: plot of clusters based on PAM

5.4.4 Using the NbClust-package

An even more comprehensive routine for estimating the number of clusters is provided by the *R* package *NbClust*. As ITS authors state, it "provides 30 indices which determine the number of clusters in a data set

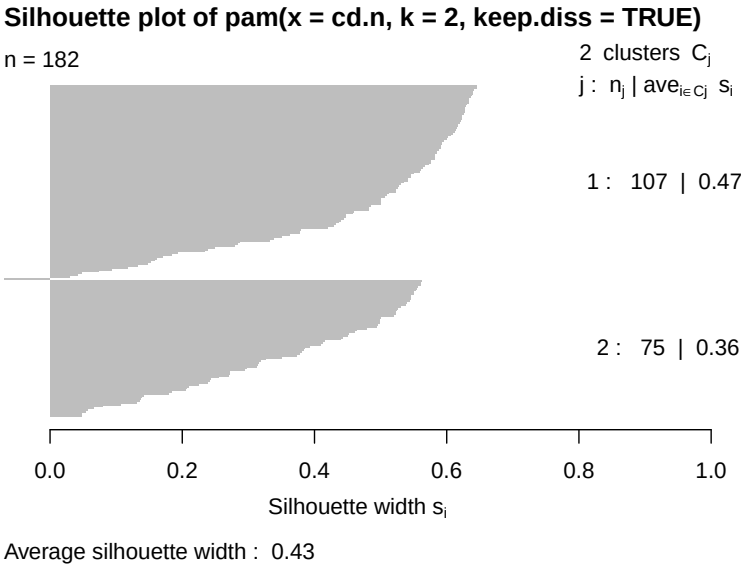

Figure 14: Silhouette-plot of clusters based on PAM

and it offers also the best clustering scheme from different results to the user" (Charrad et al., 2014). The results of applying the `NbClust()`-function are presented below.

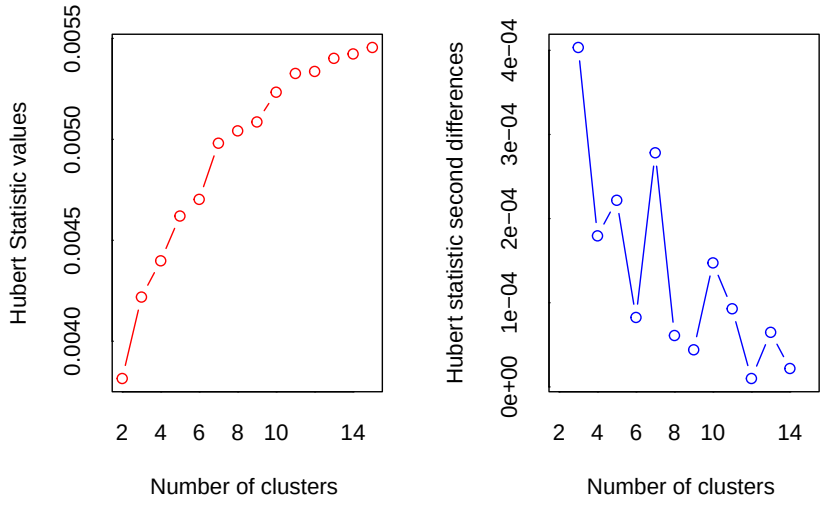

Figure 15: estimation of number of clusters using NbClust()

```
## *** : The Hubert index is a graphical method of determining the number of clusters.
##                In the plot of Hubert index, we seek a significant knee that corresponds to a
##                significant increase of the value of the measure i.e the significant peak in Hubert
##                index second differences plot.
##
```

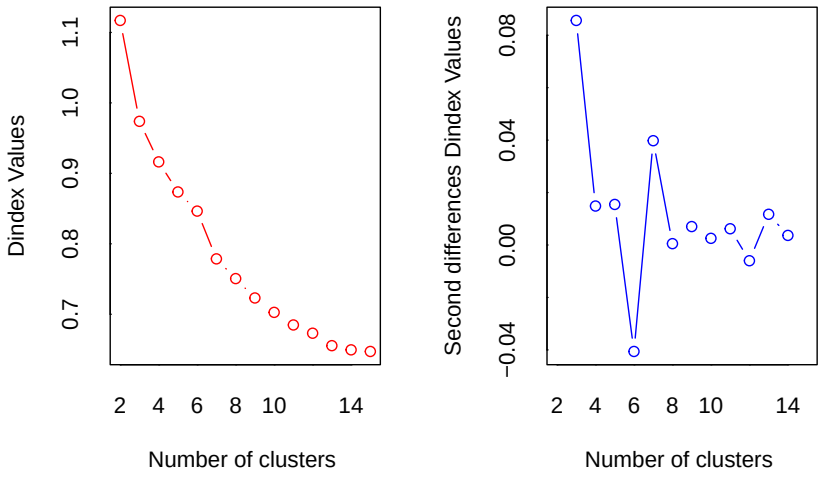

Figure 16: estimation of number of clusters using NbClust()

```
## *** : The D index is a graphical method of determining the number of clusters.
##                 In the plot of D index, we seek a significant knee (the significant peak in Dindex
##                 second differences plot) that corresponds to a significant increase of the value of
##                 the measure.
##
## *******************************************************************
## * Among all indices:
## * 8 proposed 2 as the best number of clusters
## * 15 proposed 3 as the best number of clusters
## * 1 proposed 7 as the best number of clusters
## * 1 proposed 9 as the best number of clusters
## * 1 proposed 14 as the best number of clusters
## * 1 proposed 15 as the best number of clusters
##
##                    ***** Conclusion *****
##
## * According to the majority rule, the best number of clusters is  3
##
##
## *******************************************************************
```

According to this the preferable number of clusters is **three**.

In sum: There is no clear preference as to whether to extract **two** or **three** clusters. So, in the following it is proceeded with **three** clusters, as this choice will likely extract a bit more of the variance in the time-series data.

5.5 Time-series clustering using *TSclust*

Besides the usual *pam* and *hierarchical* methods for time-series custering approaches given in the *TSclust* *R*-package were used. As Montero & Vilar (2015, p. 2) state, in time series "conventional distances between raw data (Euclidean or Manhattan, among others) evaluating a one-to-one mapping of each pair of sequences can produce satisfactory result". However there is "a wide range of dissimilarity measures to perform time series clustering" (Montero & Vilar, 2015, p. 3) assembled in the *TSclust*-package from which the **COR**relation, the **EUC**lidean Distance, and

the **DTWARP** (Dynamic Time Warping) were used in the present project. The following plot (fig. 17) depicts the results of applying these procedures.

In essence, the differences between *pam-* and *k-means*-clustering are not impressive. A more considerable difference exists between clustering with the *DTWARP*-method on the one hand and *COR*-, and *EUC*-methods on the other.

The question is, however, which is the preferable cluster solution?

Obviously the question of how to determine the validity of a clustering is not an easy one to answer. Different methods are discussed by Saitta et al. (2008), Zhao et al(2009) or Chouikhi et al. (2015). For the purpose of the present paper **only one measure** will be considered, the *within-cluster-sum-of-squares* (wcss). When clusters are "neat" and well separated it can be assumed that the *wcss* is minimized, as it indicates that the points within the clusters are *close* to each other.

The following table 4 summarizes the findings concerning this matter. It shows that clustering using the *k-means* method, with the *euclidic* time-series distances computed using the *TSclust*-package is the best.

Table 4: within cluster sum of squares

clustering method	wcss
hierarchical 3 clusters	**473.11**
pam 3 clusters	**197.9**
time-series (using TSclust) pam (cor, 3 clu.)	**199.46**
time-series (using TSclust) pam (euc, 3 clu.)	**200.73**
time-series (using TSclust) pam (dtwarp, 3 clu.)	**232.49**
time-series (using TSclust) k-means (cor, 3 clu.)	**197.09**
time-series (using TSclust) k-means (euc, 3 clu.)	**196.75**
time-series (using TSclust) k-means (dtwarp, 3 clu.)	**211.55**

Using the results from this clustering procedure (clustering using the *k-means* method, with the *euclidic* time-series distances computed using the *TSclust*-package) the world map looks like in figure 18.

Undoubtedly these results are not very convincing. Intuitively there seems to be little that countries have in common like **USA**, **Libya** and **Moldavia** (which are all in Cluster 1), or **North Korea**, **Canada** and **Malawi** (which are all in Cluster 2). On the other hand, the clustering developed here shows agreement with a grouping of nations reported by Bertolote & Fleischmann (2002): These authors observed a similarity of suicide rates in "Eastern Europe, in a group of countries that share similar historical and sociocultural characteristics, such as Estonia, Latvia, Lithuania and, to a lesser extent, Finland, Hungary and the Russian Federation", and "in the Eastern Mediterranean Region, which comprises mostly countries that follow Islamic traditions" (BERTOLOTE & FLEISCHMANN, 2002, p. 181).

Reasons for this – in spite of some agreement – somewhat disappointing result might be (a) the fact that the time-series data on suicidality do not contain as much information as it could be hoped for, (b) the clustering methodology is not adequate for extracting information, that in fact *is* in the data, or (c) the primary collection of the data in the different countries suffers from methodological flaws, and is not sufficiently valid and reliable.

It might, however, be useful to try to relate data on suicidality to the other variables which had been gathered in the analysis data set.

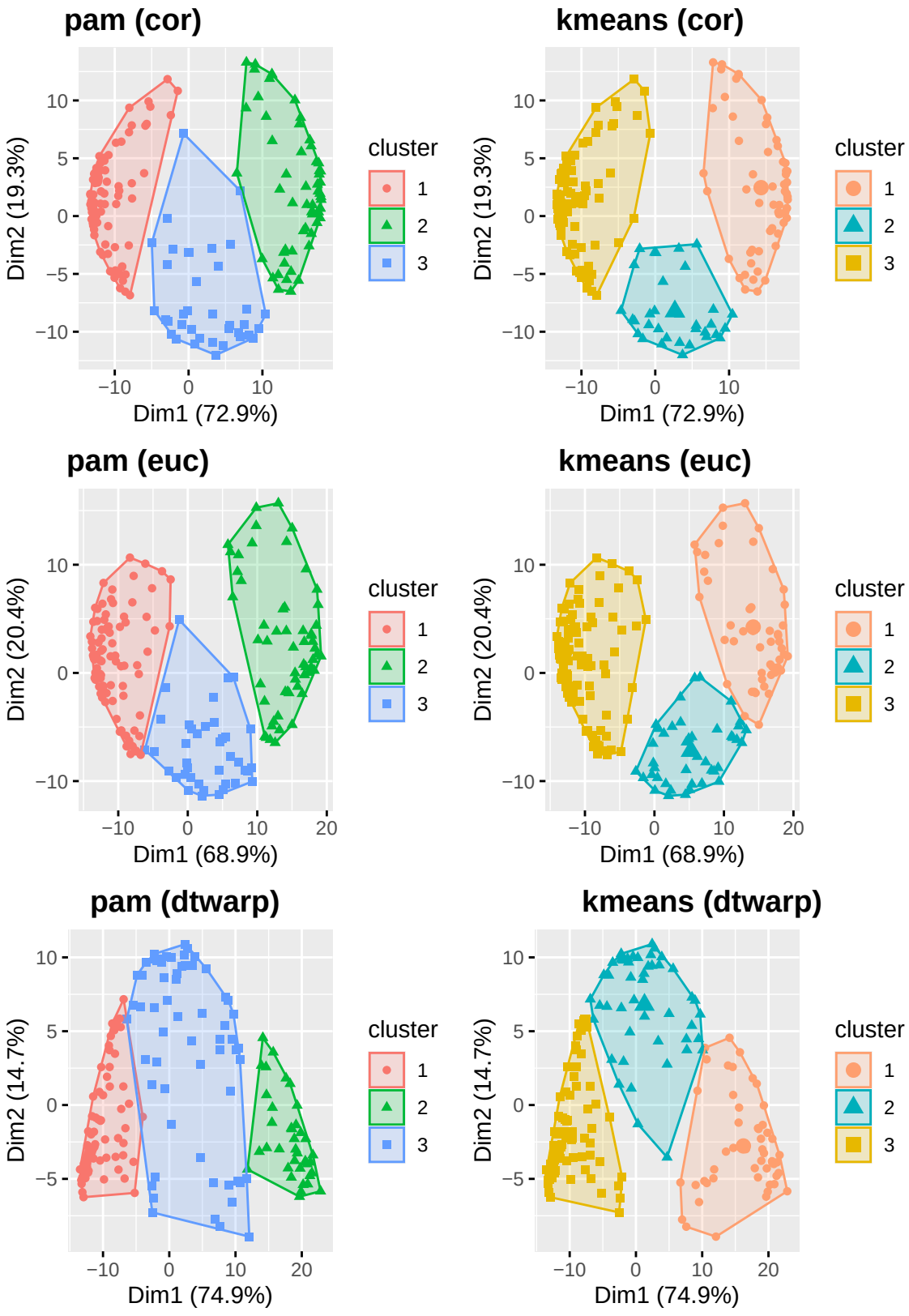

Figure 17: plots of time-series clustering using pam- and kmeans-methods

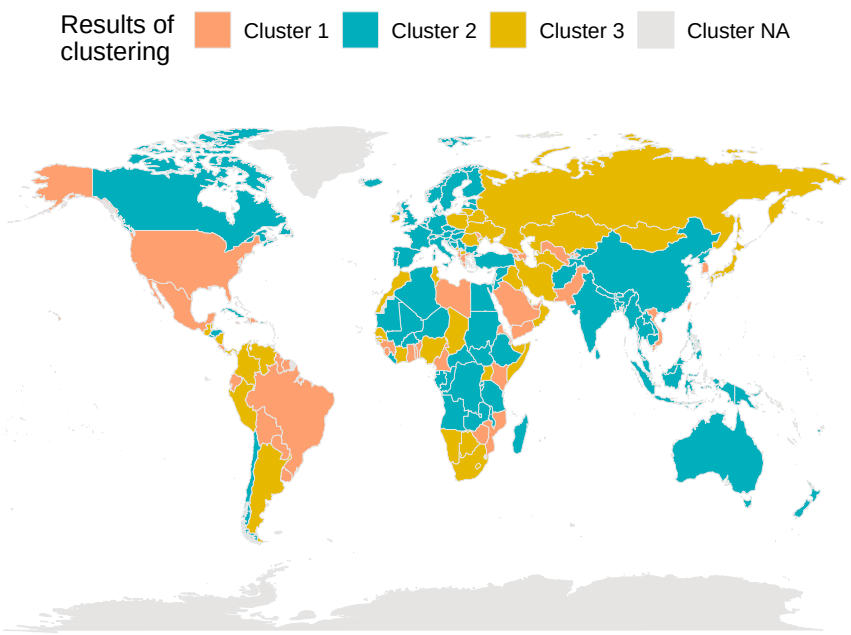

Figure 18: map of countries assigned to results of cluster analysis

6 Searching for determinants of suicidality

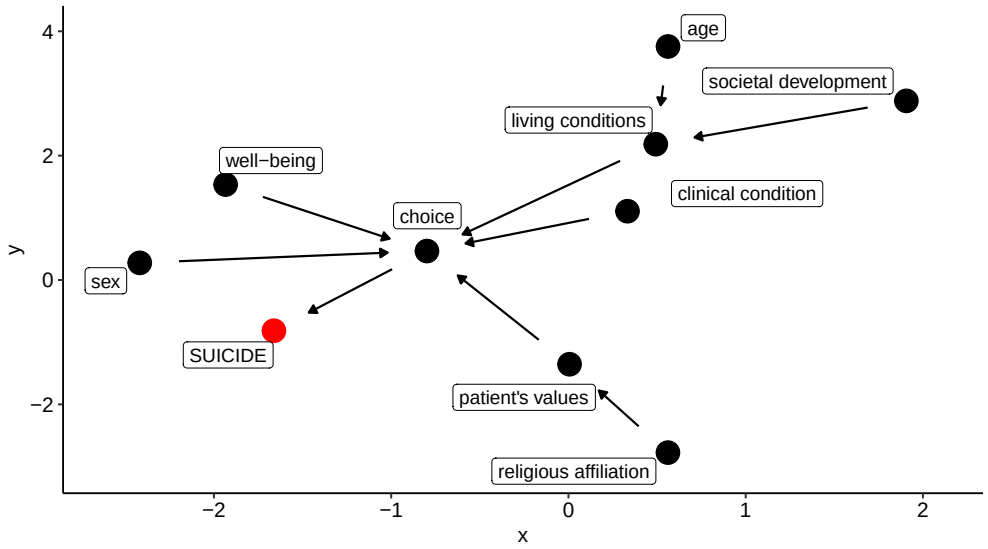

Figure 19: possible causal factors contributing to a resulting suicidality

Suicide results from a complex interplay of psychological, biogenetic, sociocultural, and existential variables (Pompili & Tatarelli, 2010). Within the scope of the present project, it will not be possible to do sufficient justice to this complexity. A preliminary attempt to summarize factors contributing to suicidality in a graph can be found in figure 19. In the following at first bivariate relations between different variables and suicidality are examined, and then an attempt is made to develop a multivariate model of factors influencing suicidality.

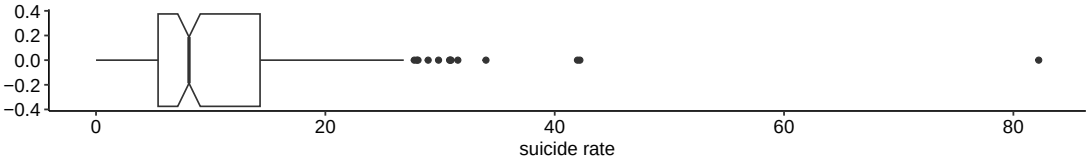

Figure 20: boxplot of rates of suicides per 100.000

6.1 Suicidality and the sexes

In almost all (precisely, in *98.96%*) of the entities under consideration in the present project there is a higher suicide rate among *men* than among *women*. Over all world regions that are considered in this project on average the *male* suicide rate is **3.26** times higher, than that of women. It can thus be assumed, that the overall rate of suicide in a society correlates with the *proportion of men* in the respective society's population.

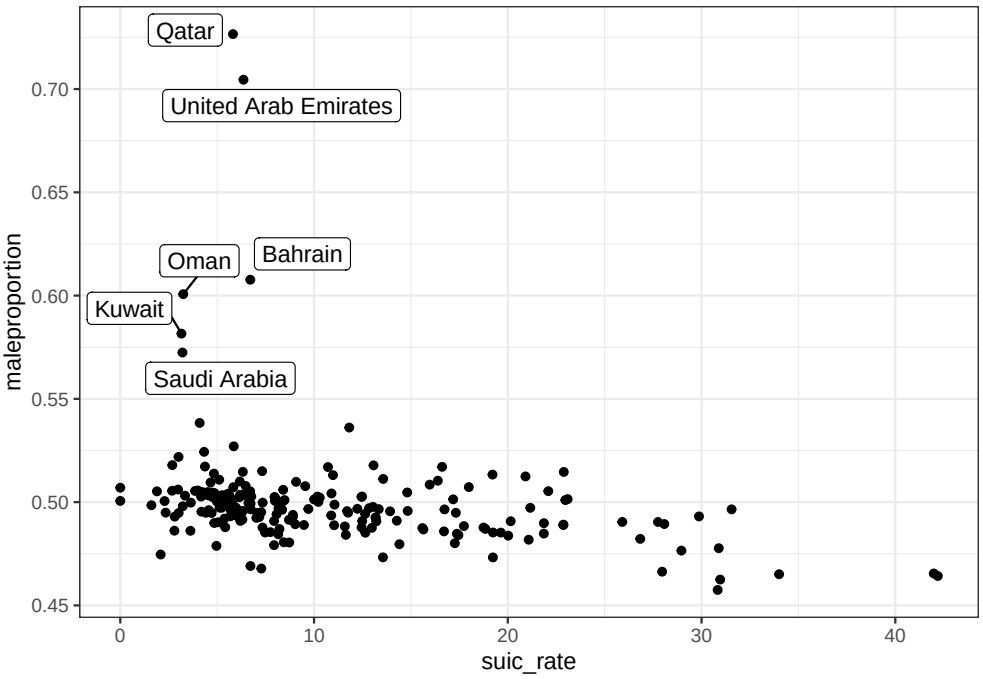

Figure 21: proportion of males in the population vs. the suicide rates

However, as a plot of the proportion of males in the population *vs.* the suicide rates shows, there are some entities with a apparently abnormal high proportion of males (fig. 21). These entities, all of them from the *Arabian Peninsula* would likely introduce a bias when correlating *proportion of males* and *suicide rates*. Inasmuch as it is unlikely that the sex ratio in the Arabian Peninsula differs significantly from that in other parts of the world, it can be assumed that there are errors in data collection here. Therefore, the entities in question should be excluded from the data evaluation.

Even after the exclusion of this data from the evaluation there does not show up a *positive* relation between a population sex ratio that favors men and a higher suicide rate. Instead a *negative* correlation of $r =$ **-0.45** , (*p-value* =) can be observed, meaning that suicide rate appears to be *lower* in societies where the ratio of the sexes is tilted towards the *male* side.

6.2 Suicidality and age factors

It is well known that the rates of suicide increase with age (O'Connell et al., 2004). This can also (tentatively) be found in the data set for the present analysis. The difficulty here is that the reported *suicides* × *age* groups are different from the *population* × *age* groups in the UN population data. In the former case the age groups are: *5 to14*, *15* to *49*, *50* to *69*, and *70+* years, in the latter case the age greoups are: *0* to *4*, *5* to *14*, *15* to *24*, *25* to *64*, and *65+*. Consequently the data from suicide rates by age and population-groups by age cannot be perfectly matched. The following table 5 thus gives only a rough estimate of the relation of suicide rates and age groups averaged over all entities and years under consideration, and of the total number of suicides over all years in all entities. The table also shows that the raw numbers (not the *rate*) of suicide is highest in the *younger* to *middle age* group, as this group constitutes the largest part of the world's population.

Table 5: suicide rates and suicides in different age groups

	5 to 14	15 to 49	50 to 69	70+
suicide rate per 100.000	0.62	14.96	15.43	17.52
total no. of suicides	11226	499488	183844	98251

The question of whether there is a relation between the suicide rates and the overall age structure can be answered using the *median age* of an entity's population. Calculating the *pearson*-correlation between these two measures results in r=**0.467** (p=*0.00*) , which can be interpreted as indicating that approx. **22%** of the *variance of the suicide rates* might be explained by the *median population age*.

6.3 Suicidality and mental disorder

Suicide risk is related to depression and bipolar disorder (Sakinofsky, 2010), and schizophrenia (Meltzer, 2010), as well as to other mental illnesses. However, as can be seen from a table showing correlations between suicide rates and rates of mental illnesses on the entity (national) level, this doesn't shows up as clearly as expected in the data set for the present analysis (table 6). However as the box-plot in figure 20 shows there is one outlier with an exceptionally high suicide rate (*Greenland*), that will likely introduce some distortion into any correlation or regression. So *Greenland* had to be excluded from the further analyses (see chapter 5).

Table 6: correlations between suicide rates and rates of mental illnesses

	Both Sexes		Males		Females	
	r	p	r	p	r	p
prevalence ADHD	-0.01	*(0.86)*	-0.04	*(0.60)*	0.19	*(0.01)*
prevalence alcohol rel. dis.	0.46	*(0.00)*	0.03	*(0.65)*	0.22	*(0.00)*
prevalence anxiety dis.	-0.05	*(0.47)*	-0.05	*(0.46)*	0.22	*(0.00)*
prevalence depressive dis.	-0.06	*(0.41)*	-0.03	*(0.68)*	0.28	*(0.00)*
prevalence bipolar dis.	-0.12	*(0.09)*	-0.05	*(0.47)*	0.18	*(0.01)*
prevalence schizophrenia	0.23	*(0.00)*	-0.03	*(0.66)*	0.25	*(0.00)*

Note: r: Pearson correlation, p: two-sided p-value

An inspection of the scatterplots of the correlations (see fig. 22) shows that in the majority of cases the correlations are low, maybe because of the overall low reported suicide rates. Also in the case of ADHD prevalence, bipolar prevalence, and schizophrenia prevalence, the charts give the impression that prevalences were recorded discontinuously in some cases, suggesting possible inadequacies in the collection of primary data.

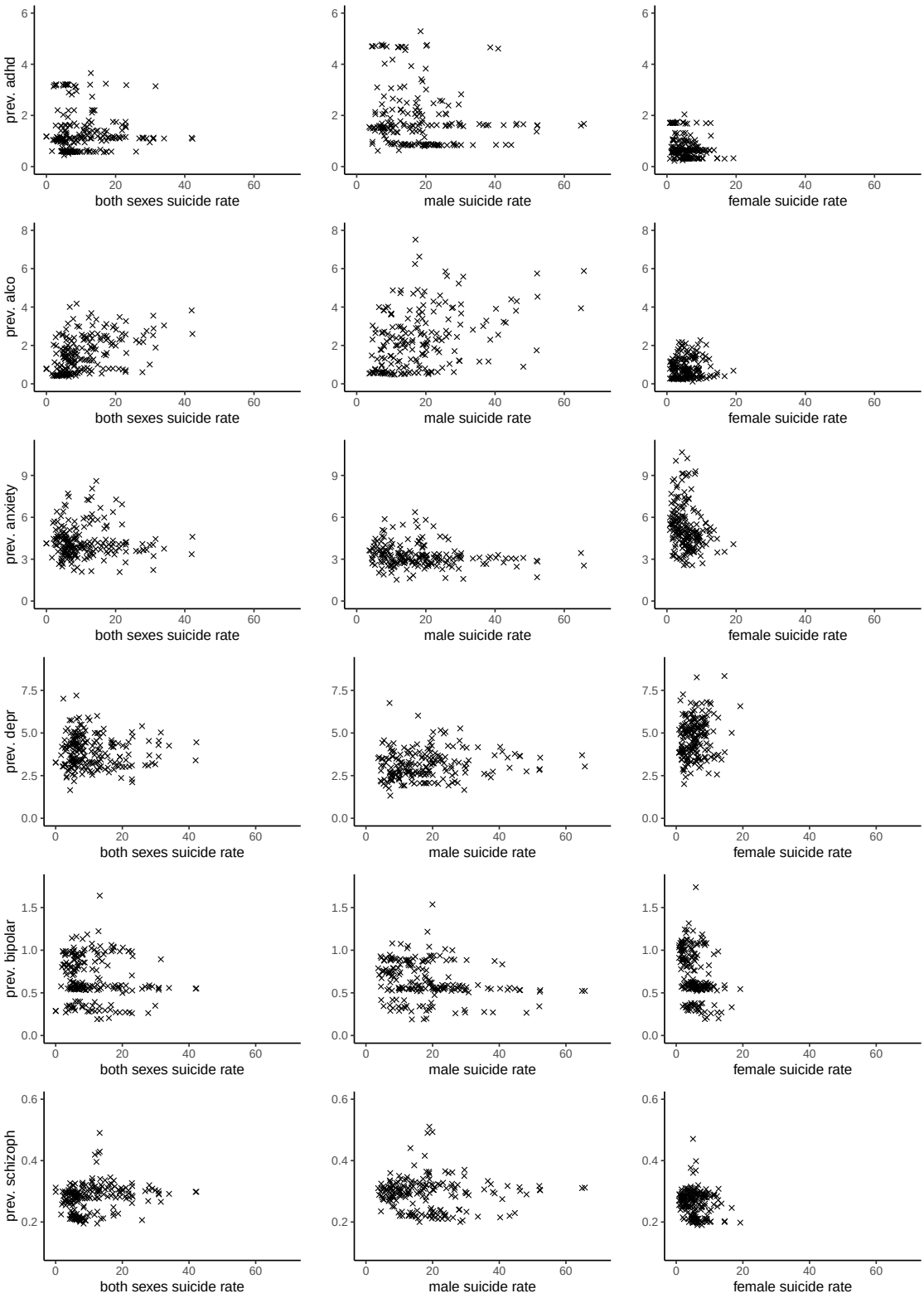

Figure 22: scatter plots of suicide rates vs. prevalence of mental illnesses

6.4 Suicidality and religious commitment

There seems to be a negative relation of *suicidality* and *importance of adherence to religious practice*, and this relation seems to be more prominent in men as can be seen from table 7, meaning the higher on a societal level is the religious commitment, the lower are the rates of suicidality in this entity.

Table 7: correlations between suicide rates and importance of religious affiliation

	Both Sexes		Males		Females	
	r	p	r	p	r	p
importance of religious practice	-0.67	*(0.00)*	-0.45	*(0.00)*	-0.22	*(0.03)*

Note: r: Pearson correlation, p: two-sided p-value

6.5 Suicidality and well-being

It can be assumed that suicidality is lower whenever and wherever the overall *psychological well-being* is higher. In the context of the present project *psychological well-being* can be seen as a function of an ability to have *trust* in other people, the degree to which one's own life is being seen as a good one on a "life ladder" ranging from *the worst* to *the best* possible life. Further contributions to psychological well-being might be the degree to which *positive* and *negative* affects are experienced, with *positive* affects presumably being related to lower suicidality and *negative* affects being related to higher suicidality.

Table 8: correlations between suicide rates and psychological well-being

	Both Sexes		Males		Females	
	r	p	r	p	r	p
self-reported trust	0.25	*(0.01)*	0.09	*(0.37)*	0.21	*(0.04)*
happines acc. to cantril ladder	0.22	*(0.00)*	-0.02	*(0.76)*	-0.17	*(0.04)*
positive affect today or prev. day	-0.02	*(0.77)*	-0.01	*(0.92)*	-0.03	*(0.68)*
negative affect today or prev. day	-0.33	*(0.00)*	-0.30	*(0.00)*	-0.25	*(0.00)*

Note: r: Pearson correlation, p: two-sided p-value

Table 8 shows bi-variate correlations of variables related to *psychological well-being* and suicidality. The presumed associations can only be found in part: the self-reported trust does not show a correlation with suicidality in men, and positive affect also does not seem to have a relation to suicidality in men and women. The relation of *happiness* to suicidality seems ambiguous, which may be the result of overall differences in suicidality rates between men and women, resulting in different trends in the association between happiness and suicidality in the two sexes (see: figure 23).

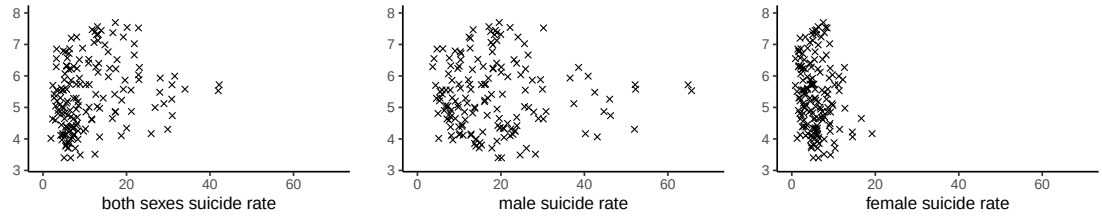

Figure 23: psychological well-being in women, men, and both sexes combined

6.6 Suicidality and societal conditions

One might assume that under social conditions that offer little prospect of a healthy and satisfying life, the tendency to end one's own life is greater than in societies that promise better in this respect. In the context of the present project societal conditions are represented by the *Human Development Index* (HDI), the *Gender Inequality Index* (GII), the *gross domestic product per capita*, the *GINI*-index, and the percentage of people living on *less than $10 per day* (measured in 2011 Purchasing Power Parities).

The respective bi-variate associations can be seen in table 9 below.

Table 9: correlations between suicide rates and societal factors

	Both Sexes		Males		Females	
	r	p	r	p	r	p
Human Development Index	0.32	*(0.00)*	0.01	*(0.91)*	-0.21	*(0.00)*
Gender Inequality Index	-0.46	*(0.00)*	-0.16	*(0.04)*	0.03	*(0.66)*
Gross Domestic Prod. per capita	0.23	*(0.00)*	-0.01	*(0.93)*	-0.10	*(0.22)*
Percentage of people living on < $10 / day	-0.34	*(0.00)*	-0.05	*(0.55)*	0.11	*(0.16)*
GINI-Index	-0.27	*(0.00)*	-0.02	*(0.82)*	-0.06	*(0.49)*

Note: r: Pearson correlation, p: two-sided p-value

Again, not all presumed associations can be found in the data. The *HDI* seems to be influential only in women (with a *higher* HDI being associated with *lower* suicidality), while – surprisingly – a lower *GII* does not seem to be beneficial to women in lowering their rate of suicides. Interestingly the *GDP per capita* appears to have little influence on suicide rates when looking at both sexes separately, while there being a positive and statistically significant relation, if both sexes are taken together. This again might be a result of a confounding effect of the gender variable with regard to suicide rates. The poverty index of *living in a household that has less the 10$ available per day* is correlated inversely to the suicide rates if both sexes are taken together, which would mean that in a society, where the rate of poor people is higher, there would be a lower rate of suicides. The *GINI*-index apparently shows no association with suicide rates.

6.7 Multivariate modeling of suicidality

In summary the results of the bivariate associations between various measures like sex, age, mental disorder, well-being, social measures, or religious commitment on the one hand, and suicide rates on the other seem somewhat equivocal and inconclusive.

It might be asked then, if a multivariate modeling (in this case: a *multiple regression*) of possible influences on suicidality could lead to further insights.

To perform a multiple regression first a data set performing this analysis had to be prepared, where prevalence for different mental illnesses which in the original data set had been reported separately for men and women were computed to reflect the prevalence in the total population (including both sexes). Then the gender specific prevalence rates were removed from the data. The sex ratio was computed as the rate of males in the entire population (with entities removed which had a male to female-ratio > 0.55 as discussed in section 6.1). Other variables that would not be useful in the coomputation of a multiple regression were excluded as well. The remaining variables are:

```
##  [1] "median_age"              "mf_ipvioldeath_age_stand"
##  [3] "suic_rate"               "hdi"
##  [5] "gii"                     "gdppc"
##  [7] "dollars_per_day"         "less_10_ratio"
##  [9] "gini"                    "self_rep_trust"
## [11] "all_veryimp"             "life_ladder"
## [13] "positive_affect"         "negative_affect"
## [15] "maleproportion"          "prev_adhd"
```

```
## [17] "prev_alco"                "prev_anxiety"
## [19] "prev_depr"                "prev_bipolar"
## [21] "prev_schizoph"
```

For doing the multiple regression first a full model was taken with all possible predictors. Only complete rows were included, which might however limit the explanatory power of model (in effect only 77 entities were included in the present calculation). As a next step all possible were computed using the function `ols_step_all_possible()` from the `olsrr`-package. From the result of this procedure the models with the **highest** *adjusted* R^2, and with the **lowest** *AIC* and *Mallow's CP* were selected. As this procedure is, however, quite highly time consuming, in the present report there are only included the final results of this calculation (in the accompanying *R*-source the relevant code is commented out, but can, of course, be run after deleting the comment-markings).

```
model <- lm(suic_rate ~ . , data = na.omit(lmdataset))
```

In the present analysis the model with the *maximum adjusted* R^2, as well as with the *minimum AIC* had **eight** predictors: `lm(suic_rate ~ median_age + gii + all_veryimp + life_ladder + maleproportion + prev_alco + prev_bipolar + prev_schizoph, data = na.omit(lmdataset))`. The model with *minimum Mallow's CP* had **six** predictors: `lm(suic_rate ~ median_age + gii + all_veryimp + maleproportion + prev_alco + prev_bipolar, data = na.omit(lmdataset))`.

When finally doing a *stepwise* regression with the `ols_step_both_p()`-function, a model results with **ten** predictors: `m10 <- lm(suic_rate ~ all_veryimp + maleproportion + prev_bipolar + less_10_ratio + prev_alco + gii + prev_depr + median_age + life_ladder + prev_schizoph, data = na.omit(lmdataset))`.

The summary of the **six**-predictor model is:

```
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.828       RMSE                 5.464
## R-Squared               0.685       Coef. Var           42.563
## Adj. R-Squared          0.658       MSE                 29.854
## Pred R-Squared          0.616       MAE                  3.870
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                 Sum of
##                Squares        DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression    4539.356         6        756.559    25.342    0.0000
## Residual      2089.795        70         29.854
## Total         6629.152        76
## --------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##          model        Beta    Std. Error    Std. Beta      t        Sig        lower      upper
## ----------------------------------------------------------------------------------------------------
##    (Intercept)      99.092        37.107                 2.670    0.009      25.084    173.100
##     median_age       0.505         0.197        0.446     2.563    0.013       0.112      0.897
##            gii      15.922         9.617        0.310     1.656    0.102      -3.260     35.103
##    all_veryimp      -0.103         0.049       -0.345    -2.124    0.037      -0.200     -0.006
## maleproportion    -200.035        67.331       -0.281    -2.971    0.004    -334.322    -65.747
##      prev_alco       1.820         0.781        0.199     2.330    0.023       0.262      3.378
##   prev_bipolar      -8.779         2.913       -0.216    -3.014    0.004     -14.588     -2.969
## ----------------------------------------------------------------------------------------------------
```

The summary of the **eight**-predictor model is:

```
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.838       RMSE                 5.382
## R-Squared               0.703       Coef. Var           41.922
## Adj. R-Squared          0.668       MSE                 28.962
## Pred R-Squared          0.620       MAE                  3.869
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -------------------------------------------------------------------
##                 Sum of
##                Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    4659.713         8        582.464    20.111    0.0000
## Residual      1969.439        68         28.962
## Total         6629.152        76
## -------------------------------------------------------------------
##
##                                    Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##          model        Beta    Std. Error    Std. Beta      t        Sig        lower      upper
## ----------------------------------------------------------------------------------------------------
##    (Intercept)     107.524        38.054                2.826    0.006      31.588    183.460
##     median_age       0.584         0.203        0.516    2.881    0.005       0.179      0.988
##            gii      20.842        10.076        0.406    2.068    0.042       0.735     40.949
##    all_veryimp      -0.100         0.048       -0.334   -2.082    0.041      -0.195     -0.004
##    life_ladder       2.436         1.329        0.257    1.833    0.071      -0.216      5.088
## maleproportion    -226.898        74.452       -0.318   -3.048    0.003    -375.465    -78.331
##      prev_alco       1.503         0.806        0.164    1.864    0.067      -0.106      3.112
##   prev_bipolar     -13.423         3.814       -0.330   -3.519    0.001     -21.034     -5.812
##  prev_schizoph     -31.581        20.275       -0.155   -1.558    0.124     -72.040      8.878
## ----------------------------------------------------------------------------------------------------
```

The summary of the **ten**-predictor model is:

```
##                         Model Summary
## ----------------------------------------------------------------
## R                       0.855       RMSE                 5.116
## R-Squared               0.732       Coef. Var           39.850
## Adj. R-Squared          0.700       MSE                 26.169
## Pred R-Squared          0.651       MAE                  3.702
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                               ANOVA
## -------------------------------------------------------------------
##                 Sum of
##                Squares        DF    Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    4849.641         8        606.205    23.165    0.0000
## Residual      1779.510        68         26.169
```

```
## Total         6629.152        76
## ------------------------------------------------------------------
##
##                                    Parameter Estimates
## ---------------------------------------------------------------------------------------------
##          model        Beta    Std. Error    Std. Beta      t        Sig        lower      upper
## ---------------------------------------------------------------------------------------------
##    (Intercept)      96.276        38.321                2.512    0.014      19.808    172.745
## maleproportion    -189.450        66.343       -0.266  -2.856    0.006    -321.835    -57.066
##    all_veryimp      -0.124         0.046       -0.414  -2.692    0.009      -0.215     -0.032
##   prev_bipolar     -15.560         3.456       -0.382  -4.502    0.000     -22.457     -8.663
##  less_10_ratio      -0.085         0.041       -0.320  -2.058    0.043      -0.167     -0.003
##      prev_alco       2.274         0.743        0.248   3.058    0.003       0.790      3.757
##      prev_depr       1.864         0.745        0.180   2.503    0.015       0.378      3.349
##            gii      24.322         9.666        0.473   2.516    0.014       5.034     43.610
##     median_age       0.388         0.211        0.343   1.838    0.070      -0.033      0.808
## ---------------------------------------------------------------------------------------------
```

It is the understanding of the author of this report[6] that all three models have some limitations, not only because of the large number of rows that were not included in the calculation because of missing values. All three are *statistically significant*, as can be seen from the respective ANOVAs. However in each of the models one or more individual predictors are *not* statistically significant (e.g. the *gender inequality index* in the six-predictors model or the *prevalence of schizophrenia* in the eight-predictors and ten-prdictors model).

As shown in the plots of *fit* and *residuals* (see: fig. 24) in all three models the spread of the residuals is larger than the spread of the centered fit, thus indicating a less that satisfactory modeling of the data. This is possibly the result of outliers, i.e. data rows that have a particularly strong influence on the model (see: fig. 25).

In summary, notwithstanding the previously mentioned limitations, the results of the *multivariate modeling of suicidality* do show some interesting findings. As it could be expected from the previous research, the *median age* of an entity's population plays a role in predicting suicide rates. Also the *Gender Inequality Index* seems to be of quite some importance, as it was already reported in section 6.6 on the bi-variate influence of societal conditions.

[6] Obviously it takes quite some statistical knowledge and experience (which the author of this paper does only have to some limited extent) to skilfully conduct a complicated analysis like the one presented here. So please, bear with me ...

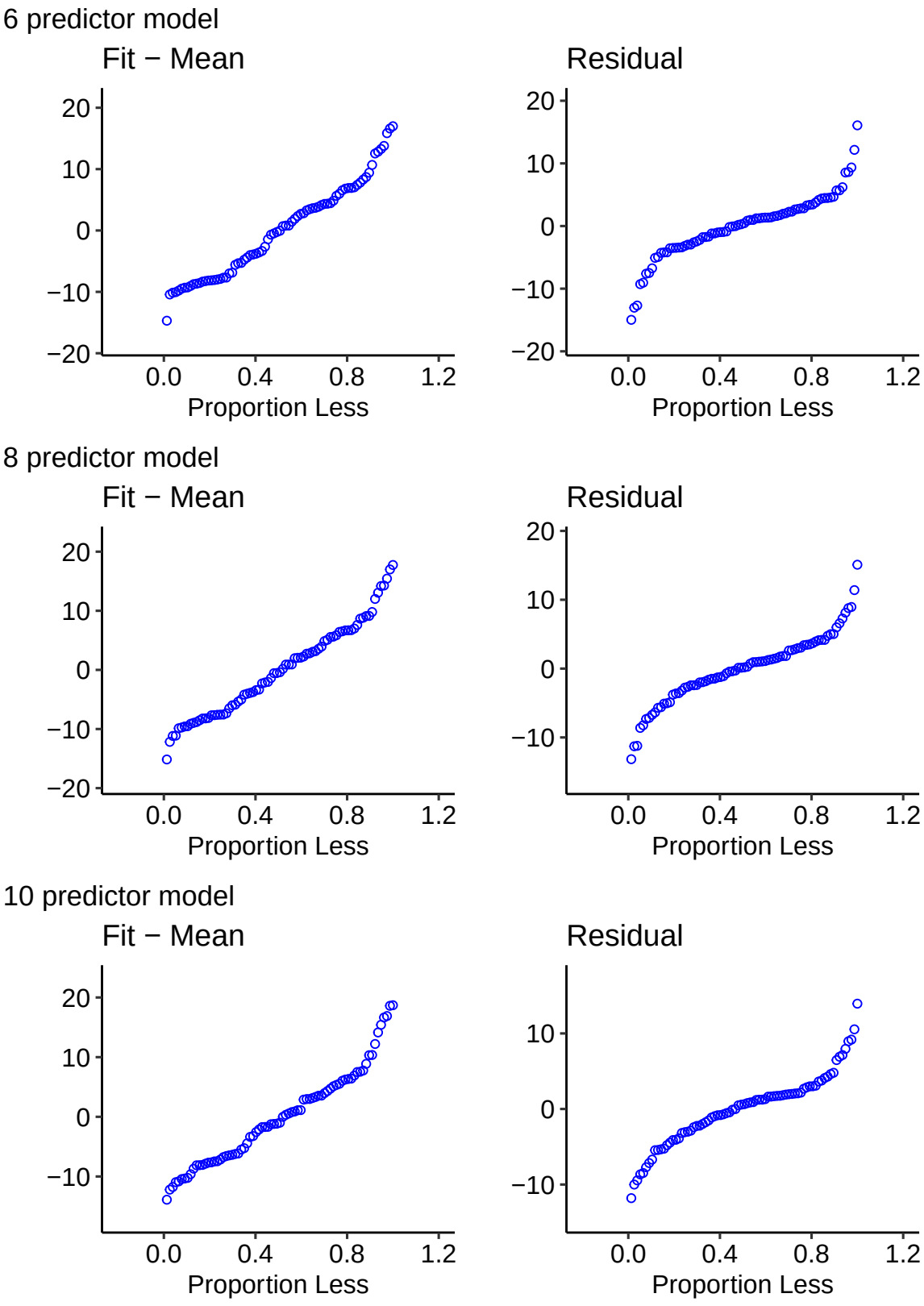

Figure 24: plots of residuals and fits

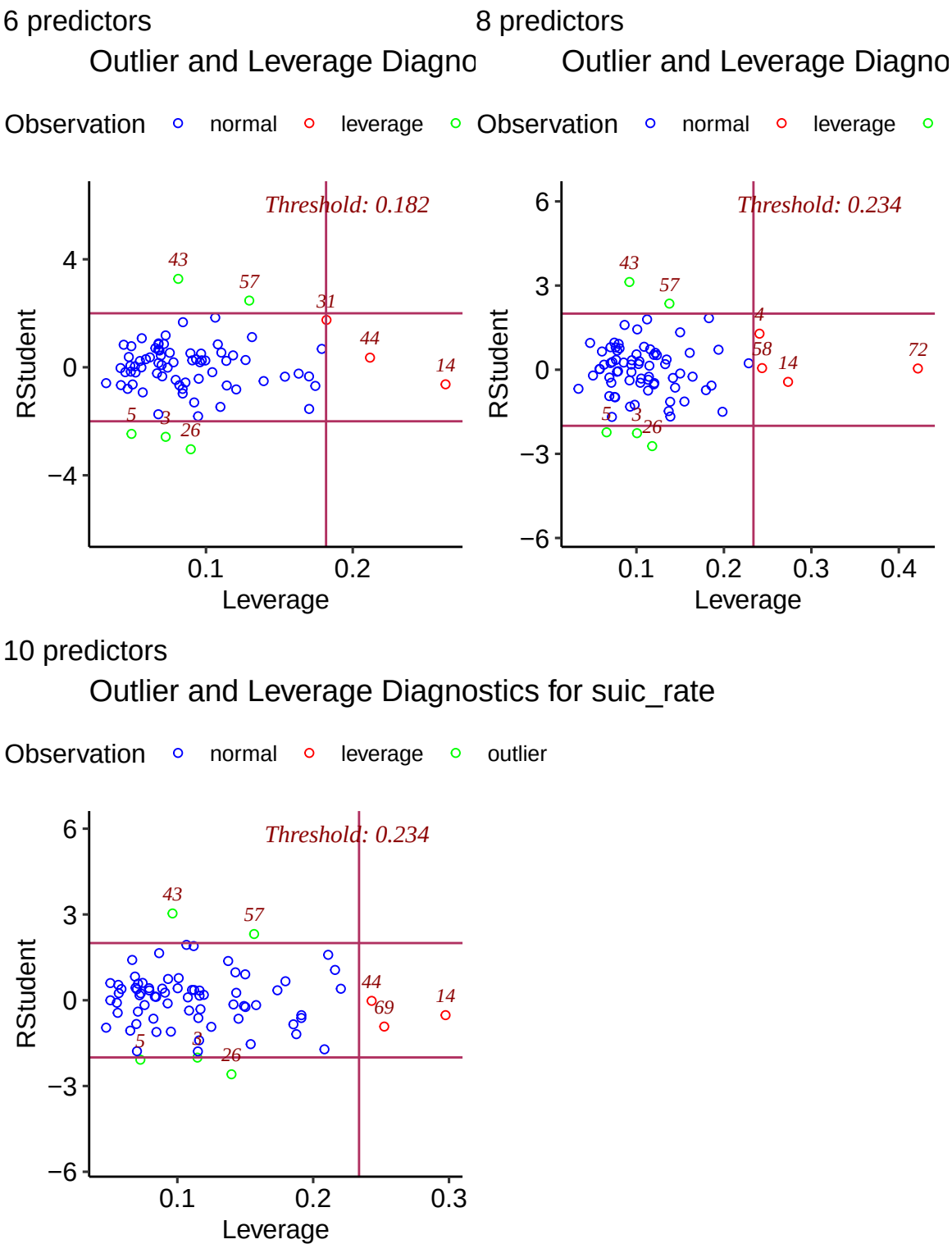

Figure 25: plots of outliers and lerverage

7 Concluding Remarks

The present project gave me the opportunity to delve into the topic of suicidality from an epidemiological perspective, a topic with which I every now and then am confronted with from a more clinical point of view.

I could see that quite a plethora of data on a wide range of topics related to suicidality can be found on the net in different formats, and I had the chance to learn how to extract these for the present analysis.

At the same time I had to realize the limitations that are also there: for a number of regions of the world data are only fragmentary (maybe due to lack of regular reporting) or maybe even wrong (this might be the case in the reported sex ratio for nations from the Arabian peninsula).

I also had to realize that – although the present Data Science course was exceptionally valuable – for myself a whole lot is missing, particularly in the field of statistics as it comes to applying the more sophisticated methods of *cluster analysis* or *multiple regression* (and other methods as well).

In the end I am particularly grateful to all the people who contributed such extraordinary tools like *R* or *RStudio* and all the tools and packages that I could use in this project. I remember well my first steps in medical statistics (cf: Davies-Osterkamp et al., 1992), when it still was necessary to use eighty-column punch card, a pile of which was carried to the university computing center, only to get back a batch of fan-fold paper a couple of days later. How much easier is it today, and how much more fun!

8 References

Althaus, D., & Hegerl, U. (2004). [Causes, diagnosis and treatment of suicidality]. *Der Nervenarzt*, *75*(11), 1123–1134; quiz 1135. https://doi.org/10.1007/s00115-004-1824-2

BERTOLOTE, J. M., & FLEISCHMANN, A. (2002). Suicide and psychiatric diagnosis: a worldwide perspective. *World Psychiatry*, *1*(3), 181–185. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489848/

Bolt, J., & van Zanden, J. L. (n.d.). *Maddison style estimates of the evolution of the world economy. A new 2020 update*. 43.

Cantril, H. (1965). *Pattern of human concerns*. Rutgers University Press.

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61*, 1–36. https://doi.org/10.18637/jss.v061.i06

Chatterjee, D., & Rai, R. (2021). Choosing Death Over Survival: A Need to Identify Evolutionary Mechanisms Underlying Human Suicide. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.689022

Chouikhi, H., Charrad, M., & Ghazzali, N. (2015). A comparison study of clustering validity indices. *2015 Global Summit on Computer & Information Technology (GSCIT)*, 1–4. https://doi.org/10.1109/GSCIT.2015.7353330

clValid: An R Package for Cluster Validation by Guy Brock, Vasyl Pihur, Susmita Datta, Somnath Datta. (2008). https://www.jstatsoft.org/article/view/v025i04

Davies-Osterkamp, S., Hartkamp, N., Heigl-Evers, A., & Standke, G. (1992). Zur Diagnostik von Ich-Funktionen und Objektbeziehungen anhand von Ratingskalen. *Zeitschrift für Psychosomatische Medizin und Psychoanalyse*, *38*(1), 17–30. https://www.jstor.org/stable/23997122

De Leo, D. (2002). Why are we not getting any closer to preventing suicide? *The British Journal of Psychiatry: The Journal of Mental Science*, *181*, 372–374. https://doi.org/10.1192/bjp.181.5.372

Hardt, J., Egle, U. T., & Johnson, J. G. (2007). Suicide attempts and retrospective reports about parent-child relationships: evidence for the affectionless control hypothesis. *Psycho-Social Medicine*, *4*, Doc12.

Hegerl, U. (2005). Depression und Suizidalität. *Verhaltenstherapie*, *15*(1), 6–11. https://doi.org/10.1159/000083697

Hegerl, U., Maxwell, M., Harris, F., Koburger, N., Mergl, R., Székely, A., Arensman, E., Van Audenhove, C., Larkin, C., Toth, M. D., Quintão, S., Värnik, A., Genz, A., Sarchiapone, M., McDaid, D., Schmidtke, A., Purebl, G., Coyne, J. C., Gusmão, R., & OSPI-Europe Consortium. (2019). Prevention of suicidal behaviour: Results of a controlled community-based intervention study in four European countries. *PloS One*, *14*(11), e0224602. https://doi.org/10.1371/journal.pone.0224602

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Kulawik, H. (1977). Zur Psychopathologie der Suizidalität. *Psychiatrie, Neurologie Und Medizinische Psychologie*, *29*(5), 257–265. https://www.jstor.org/stable/45251961

Little, R. (2009). Understanding the Psychodynamics of Suicidal Clients: Exploring Suicidal and Presuicidal States. *Transactional Analysis Journal*, *39*(3), 219–228. https://doi.org/10.1177/036215370903900305

Macau. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Macau&oldid=1114194176

Maltsberger, J. T. (2004). The descent into suicide. *The International Journal of Psycho-Analysis*, *85*(Pt 3), 653–667. https://doi.org/10.1516/002075704774200799

Manvailer, V. (n.d.). *Partition Around Medoids (PAM) and Silhouette plots*. https://www.google.com/search?client=firefox-b-d&q=Partition+Around+Medoids+%28PAM%29+and+Silhouette+plots

Meltzer, H. Y. (2010). Evidence-based treatment for reducing suicide risk in schizophrenia. *M. Pompili, & R. Tatarelli (Red.), Evidence-Based Practice in Suicidology*, 317–328.

Modestin, J. (1990). Über die präsuizidale Verfassung. *Über Die Präsuizidale Verfassung*, *61*(3), 159–163.

Montero, P., & Vilar, J. A. (2015). TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software*, *62*, 1–43. https://doi.org/10.18637/jss.v062.i01

Nations, U. (n.d.). *Human Development Index*. United Nations. https://hdr.undp.org/data-center/human-development-index

O'Connell, H., Chin, A.-V., Cunningham, C., & Lawlor, B. A. (2004). Recent developments: Suicide in older people. *BMJ*, *329*(7471), 895–899. https://doi.org/10.1136/bmj.329.7471.895

Our World in Data. (n.d.). Our World in Data. https://ourworldindata.org

Pompili, M., & Tatarelli, R. (2010). *Evidence-Based Practice in Suicidology: A Source Book*. Hogrefe Publishing.

Preti, A. (2011). Do Animals Commit Suicide? Does It Matter? *Crisis*, *32*(1), 1–4. https://doi.org/10.1027/0227-5910/a000077

Rani, S., & Sikka, G. (2012). Recent techniques of clustering of time series data: a survey. *International Journal of Computer Applications*, *52*(15).

Ritchie, H., Roser, M., & Ortiz-Ospina, E. (2015). Suicide. *Our World in Data*. https://ourworldindata.org/suicide

Saitta, S., Raphael, B., & Smith, I. F. C. (2008). A comprehensive validity index for clustering. *Intelligent Data*

Analysis, *12*(6), 529–548. https://doi.org/10.3233/IDA-2008-12602

Sakinofsky, I. (2010). *Evidence-Based Approaches for Reducing Suicide Risk in Major Affective Disorders*. https://doi.org/10.13140/2.1.2360.9287

Suicide worldwide in 2019. (n.d.). https://www.who.int/publications-detail-redirect/9789240026643

Taylor, S. J., Kingdom, D., & Jenkins, R. (1997). How are nations trying to prevent suicide? An analysis of national suicide prevention strategies. *Acta Psychiatrica Scandinavica*, *95*(6), 457–463. https://doi.org/10.1111/j.1600-0447.1997.tb10132.x

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Warren Liao, T. (2005). Clustering of time series data—a survey. *Pattern Recognition*, *38*(11), 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025

Wolfersdorf, M. (2008). Suizidalität. *Der Nervenarzt*, *79*(11), 1319–1336. https://doi.org/10.1007/s00115-008-2478-2

York, P. (2022). *owidR: Import Data from Our World in Data*. https://CRAN.R-project.org/package=owidR

Zalsman, G., Hawton, K., Wasserman, D., van Heeringen, K., Arensman, E., Sarchiapone, M., Carli, V., Höschl, C., Winkler, P., Balazs, J., Purebl, G., Kahn, J. P., Sáiz, P. A., Bobes, J., Cozman, D., Hegerl, U., Rancāns, E., Hadlaczky, G., Van Audenhove, C., … European Evidence-Based Suicide Prevention Program [EESPP] Group by the Expert Platform on Mental Health, Focus on Depression. (2017). Evidence-based national suicide prevention taskforce in Europe: A consensus position paper. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *27*(4), 418–421. https://doi.org/10.1016/j.euroneuro.2017.01.012

Zhao, Q., Xu, M., & Fränti, P. (2009). Sum-of-Squares Based Cluster Validity Index and Significance Analysis. In M. Kolehmainen, P. Toivanen, & B. Beliczynski (Eds.), *Adaptive and Natural Computing Algorithms* (pp. 313–322). Springer. https://doi.org/10.1007/978-3-642-04921-7_32

9 Appendix

9.1 Sources of the OWID-data on suicide

9.1.1 Suicide rate vs. death rate from violence

Deaths - Self-harm - Sex: Both - Age: All Ages (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - All forms of violence - OWID - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable description	Variable calculated by OWID: the sum of ['Deaths - Interpersonal violence - Sex: Both - Age: Age-standardized (Rate)', 'Deaths - Conflict and terrorism - Sex: Both - Age: Age-standardized (Rate)', 'Deaths - Executions and police conflict - Sex: Both - Age: Age-standardized (Rate)']
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.1.2 Suicide rate vs. homicide rate

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Interpersonal violence - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.1.3 Suicide death rate vs. prevalence of mental & substance use disorders

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Prevalence - Mental and substance use disorders - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	11-February-18

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.1.4 Suicide rates vs. prevalence of depression

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.1.5 Share of suicide deaths from pesticide poisoning

Share of suicide deaths from pesticides (Mew et al. 2017)

Var	Source
Variable description	Share of deaths from suicide which result from pesticide poisoning
Variable time span	2014 – 2014
Data published by	Mew, E. J., Padmanathan, P., Konradsen, F., Eddleston, M., Chang, S. S., Phillips, M. R., & Gunnell, D. (2017). The global burden of fatal self-poisoning with pesticides 2006-15: systematic review. Journal of affective disorders, 219, 93-104.
Link	https://www.sciencedirect.com/science/article/pii/S016503271730280X
Retrieved	18th November 2019

9.1.6 Death rate from suicides

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.7 Deaths from suicide, by age

Deaths - Self-harm - Sex: Both - Age: 70+ years (Number)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 50-69 years (Number)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 15-49 years (Number)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 5-14 years (Number)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.8 Male vs. female suicide rate

Male suicide rate (age-standardized)

Var	Source
Variable description	Age-standardized suicide rate in males, measured as the number of deaths per 100,000.
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	18th November 2019

Female suicide rate (age-standardized)

Var	Source
Variable description	Age-standardized suicide rate in females, measured as the number of deaths per 100,000.
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	18th November 2019

9.1.9 Male-to-female ratio of suicide rates

Male:female suicide ratio

Var	Source
Variable description	Male:female suicide ratio is measured as the number of suicides in men versus the number in women.
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	18th November 2019

9.1.10 Number of suicide deaths from firearms

Deaths - Self-harm by firearm - Sex: Both - Age: All Ages (Number)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.11 Share of deaths from suicide

Deaths - Self-harm - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.12 Suicide death rate by age

Deaths - Self-harm - Sex: Both - Age: All Ages (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 70+ years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 50-69 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 15-49 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 5-14 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.13 Suicide death rate by sex

Female suicide rate (age-standardized)

Var	Source
Variable description	Age-standardized suicide rate in females, measured as the number of deaths per 100,000.
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	18th November 2019

Male suicide rate (age-standardized)

Var	Source
Variable description	Age-standardized suicide rate in males, measured as the number of deaths per 100,000.
Variable time span	1990 – 2017
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	18th November 2019

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.14 Suicide death rates

Suicide mortality rate (per 100,000 population)

Var	Source
Variable description	Suicide mortality rate is the number of suicide deaths in a year per 100,000 population. Crude suicide rate (not age-adjusted).
Variable time span	2000 – 2019
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Global Health Observatory Data Repository - World Health Organization
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

9.1.15 Suicide rate

Suicide mortality rate (per 100,000 population)

Var	Source
Variable description	Suicide mortality rate is the number of suicide deaths in a year per 100,000 population. Crude suicide rate (not age-adjusted).
Variable time span	2000 – 2019
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Global Health Observatory Data Repository - World Health Organization
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

9.1.16 Suicide rate by firearm

Deaths - Self-harm by firearm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.17 Suicide rate in 1990 vs. 2019

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.1.18 Suicide rate vs. income inequality

Suicide mortality rate (per 100,000 population)

Var	Source
Variable description	Suicide mortality rate is the number of suicide deaths in a year per 100,000 population. Crude suicide rate (not age-adjusted).
Variable time span	2000 – 2019
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Global Health Observatory Data Repository - World Health Organization
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Gini coefficient

Var	Source
Variable description	The Gini coefficient measures inequality on a scale between 0 and 1, where higher values indicate greater inequality.
Variable time span	1967 – 2021
Data published by	World Bank Poverty and Inequality Platform
Data publisher's source	World Bank Data are based on primary household survey data obtained from government statistical agencies and World Bank country departments.
Link	https://pip.worldbank.org/
Retrieved	2022-10-03

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.1.19 Suicide rates by age

Deaths - Self-harm - Sex: Both - Age: 70+ years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 50-69 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: All Ages (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 5-14 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

Deaths - Self-harm - Sex: Both - Age: 15-49 years (Rate)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-09-22

9.2 Topics of additional OWID-data related to the project

Table 10: topics covered in additional OWID tables

Topics
Human Development Index vs. Corruption Perception Index
Human Development Index vs. GDP per capita
Gender Inequality Index from the Human Development Report
Human Development Index
Life expectancy vs. expected years lived with disability or disease
Life expectancy vs. GDP per capita
Inequality in life expectancy vs. health expenditure per capita
Life expectancy vs. health expenditure
Life expectancy vs. health expenditure per capita
Life expectancy at birth, including the UN projections
Share in poverty vs. Life expectancy
Life expectancy
Population by broad age group, World
Life satisfaction vs. life expectancy
Healthy life expectancy and years lived with disability
Prevalence of alcohol use disorders by age
Prevalence of alcohol use disorders in males vs. females
Child mortality vs. Prevalence of child wasting
Child mortality vs. Prevalence of stunting
Prevalence of ADHD in males vs females
Prevalence of anxiety disorders by age
Prevalence of anxiety disorders, males vs. females
Prevalence of bipolar disorder by age
Prevalence of bipolar disorder in males vs. females
Prevalence of depression by age
Prevalence of depression, males vs. females
Prevalence of eating disorders by age
Prevalence of schizophrenia by age
Prevalence of schizophrenia in males vs. females
Average years of schooling vs. GDP per capita
How important is God in your life? vs GDP per capita, 2014
Self-reported life satisfaction, 2003 to 2020
Happiness vs. life satisfaction, 2015
Civil liberties, 2021
Share of people agreeing with the statement "most people can be trusted"

9.3 Sources of the additional OWID-data related to the project

9.3.1 Human Development Index vs. Corruption Perception Index

Human Development Index (UNDP)

Var	Source
Variable description	The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. Measured by:– Life expectancy at birth– Mean years of schooling & Expected years of schooling– GNI per capita (in PPP adjusted international-$)The HDI is the geometric mean of normalized indices for each of the three dimensions.
Variable geographic coverage	Global by country
Variable time span	1980 – 2017
Data published by	United Nations Development Programme, Human Development Report 2020
Link	http://hdr.undp.org/en/indicators/137506
Retrieved	08/05/2019

Corruption Perception Index - Transparency International (2018)

Var	Source
Variable description	Annual ranking of countries by their perceived levels of corruption, as determined by expert assessments and opinion surveys. Scale is from 100 (very clean) to 0 (highly corrupt).
Variable time span	2012 – 2018
Data published by	Transparency International
Data publisher's source	Various expert surveys including (as available for each country): African Development Bank Governance Ratings, Bertelsmann Foundation Sustainable Governance Indicators, Bertelsmann Foundation Transformation Index, Economist Intelligence Unit Country Risk Ratings, Freedom House Nations in Transit, Global Insight Country Risk Ratings, IMD World Competitiveness Yearbook, Political and Economic Risk Consultancy Asian Intelligence, Political Risk Services International Country Risk Guide, World Bank - Country Policy and Institutional Assessment, World Economic Forum Executive Opinion Survey, World Justice Project Rule of Law Index
Link	https://www.transparency.org/cpi2018
Retrieved	07/05/2018

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.2 Human Development Index vs. GDP per capita

Human Development Index (UNDP)

Var	Source
Variable description	The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. Measured by:– Life expectancy at birth– Mean years of schooling & Expected years of schooling– GNI per capita (in PPP adjusted international-$)The HDI is the geometric mean of normalized indices for each of the three dimensions.
Variable geographic coverage	Global by country
Variable time span	1980 – 2017
Data published by	United Nations Development Programme, Human Development Report 2020
Link	http://hdr.undp.org/en/indicators/137506
Retrieved	08/05/2019

GDP per capita, PPP (constant 2017 international $)

Var	Source
Variable description	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the country plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2017 international dollars.Statistical concept and methodology: For the concept and methodology of 2017 PPP, please refer to the International Comparison Program (ICP)'s website (https://www.worldbank.org/en/programs/icp).
Variable time span	1990 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Household surveys, UN Population Division
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.3 Gender Inequality Index from the Human Development Report

Gender Inequality Index (Human Development Report (2015))

Var	Source
Variable description	The GII is an inequality index. It shows the loss in potential human development due to disparity between female and male achievements in two dimensions, empowerment and economic status. The GII ranges between 0 and 1. Higher GII values indicate higher inequalities and thus higher loss to human development.
Variable time span	1995 – 2015
Data published by	UN Development Programme
Data publisher's source	Relies on data from the UN's MMEIG which includes WHO, UNICEF, UNFPA, and the World Bank as well as the UN Department of Economic and Social Affairs, UNESCO, IPU, ILO
Link	http://hdr.undp.org/en/data#
Retrieved	23/03/2018

9.3.4 Human Development Index

Human Development Index (UNDP)

Var	Source
Variable description	The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living. Measured by:– Life expectancy at birth– Mean years of schooling & Expected years of schooling– GNI per capita (in PPP adjusted international-$)The HDI is the geometric mean of normalized indices for each of the three dimensions.
Variable geographic coverage	Global by country
Variable time span	1980 – 2017
Data published by	United Nations Development Programme, Human Development Report 2020
Link	http://hdr.undp.org/en/indicators/137506
Retrieved	08/05/2019

9.3.5 Life expectancy vs. expected years lived with disability or disease

Life Expectancy (IHME)

Var	Source
Variable geographic coverage	global by country
Variable time span	1990 – 2016
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	15th February 2018

Years lived with disability

Var	Source
Variable geographic coverage	global by country
Variable time span	1990 – 2016
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	15th February 2018

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.6 Life expectancy vs. GDP per capita

Life expectancy

Var	Source
Variable description	Life expectancy at birth is defined as the average number of years that a newborn could expect to live if he or she were to pass through life subject to the age-specific mortality rates of a given period.
Variable time span	1543 – 2019
Data published by	James C. Riley (2005) – Estimates of Regional and Global Life Expectancy, 1800–2001. Issue Population and Development Review. Population and Development Review. Volume 31, Issue 3, pages 537–543, September 2005., Zijdeman, Richard; Ribeira da Silva, Filipa, 2015, "Life Expectancy at Birth (Total)", http://hdl.handle.net/10622/LKYT53, IISH Dataverse, V1, and UN Population Division (2019)
Data publisher's source	https://www.lifetable.de/RileyBib.pdf
Link	https://datasets.socialhistory.org/dataset.xhtml?persistentId=hdl:10622/LKYT53, http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4457.2005.00083.x/epdf, https://population.un.org/wpp/Download/Standard/Population/
Retrieved	14/09/2019

GDP per capita

Var	Source
Variable time span	1 – 2018
Data published by	Bolt, Jutta and Jan Luiten van Zanden (2020), "Maddison style estimates of the evolution of the world economy. A new 2020 update".
Data publisher's source	The Maddison Project Database is based on the work of many researchers that have produced estimates ofeconomic growth for individual countries. The full list of sources for this historical data is given for each country below.
Link	https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020
Retrieved	2022-04-12

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.7 Inequality in life expectancy vs. health expenditure per capita

Inequality in life expectancy (2015-2020)

Var	Source
Variable time span	2015 – 2015
Data published by	United Nations Development Programme
Data publisher's source	Within-country surveys
Link	http://hdr.undp.org/en/composite/IHDI
Retrieved	13 February 2019

Current health expenditure per capita, PPP (current international $)

Var	Source
Variable description	Current expenditures on health per capita expressed in international dollars at purchasing power parity.Statistical concept and methodology: The health expenditure estimates have been prepared by the World Health Organization (WHO) under the framework of the System of Health Accounts 2011 (SHA 2011). The Health SHA 2011 tracks all health spending in a given country over a defined period of time regardless of the entity or institution that financed and managed that spending. It generates consistent and comprehensive data on health spending in a country, which in turn can contribute to evidence-based policy-making. WHO converted the expenditure data using PPP time series extracted from WDI (based on ICP 2017) and OECD data. Where WDI/OECD data were not available, IMF or WHO estimates were utilized. Detailed metadata are available at .Notes from original source: The World Health Organization (WHO) has revised health expenditure data using the new international classification for health expenditures in the revised System of Health Accounts (SHA 2011). WHO's Global Health Expenditure Database in this new version is the reference source for health expenditure for international comparison imbedded in a standardized framework. The SHA 2011 clarifies the financing mechanisms and introduces new dimensions which improve the comparability of health expenditures in the perspective of universal health coverage.
Variable time span	2000 – 2019
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Global Health Observatory Data Repository - World Health Organization
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.8 Life expectancy vs. health expenditure

Life expectancy at birth, total (years)

Var	Source
Variable description	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.Limitations and exceptions: Annual data series from United Nations Population Division's World Population Prospects are interpolated data from 5-year period data. Therefore they may not reflect real events as much as observed data.Statistical concept and methodology: Life expectancy at birth used here is the average number of years a newborn is expected to live if mortality patterns at the time of its birth remain constant in the future. It reflects the overall mortality level of a population, and summarizes the mortality pattern that prevails across all age groups in a given year. It is calculated in a period life table which provides a snapshot of a population's mortality pattern at a given time. It therefore does not reflect the mortality pattern that a person actually experiences during his/her life, which can be calculated in a cohort life table.High mortality in young age groups significantly lowers the life expectancy at birth. But if a person survives his/her childhood of high mortality, he/she may live much longer. For example, in a population with a life expectancy at birth of 50, there may be few people dying at age 50. The life expectancy at birth may be low due to the high childhood mortality so that once a person survives his/her childhood, he/she may live much longer than 50 years.
Variable time span	1960 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Household surveys, UN Population Division
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Health Expenditure and Financing (per capita) (OECDstat (2017))

Var	Source
Variable description	Per capita health expenditure and financing in OECD countries, measured in 2010 international dollars
Variable geographic coverage	OECD countries
Variable time span	1970 – 2015
Data published by	Organisation for Economic Co-operation and Development (OECD) Statistics
Data publisher's source	Joint OECD, EUROSTAT and WHO Health Accounts SHA Questionnaires (JHAQ)
Link	https://stats.oecd.org/
Retrieved	24/05/2017

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.9 Life expectancy vs. health expenditure per capita

Life expectancy - Sex: all - Age: at birth - Variant: estimates

Var	Source
Variable description	The average number of remaining years of life expected by a hypothetical cohort of individuals who already reached age x and would be subject during the remainder of their lives to the mortality rates of a given period.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Current health expenditure per capita, PPP (current international $)

Var	Source
Variable description	Current expenditures on health per capita expressed in international dollars at purchasing power parity.Statistical concept and methodology: The health expenditure estimates have been prepared by the World Health Organization (WHO) under the framework of the System of Health Accounts 2011 (SHA 2011). The Health SHA 2011 tracks all health spending in a given country over a defined period of time regardless of the entity or institution that financed and managed that spending. It generates consistent and comprehensive data on health spending in a country, which in turn can contribute to evidence-based policy-making. WHO converted the expenditure data using PPP time series extracted from WDI (based on ICP 2017) and OECD data. Where WDI/OECD data were not available, IMF or WHO estimates were utilized. Detailed metadata are available at .Notes from original source: The World Health Organization (WHO) has revised health expenditure data using the new international classification for health expenditures in the revised System of Health Accounts (SHA 2011). WHO's Global Health Expenditure Database in this new version is the reference source for health expenditure for international comparison imbedded in a standardized framework. The SHA 2011 clarifies the financing mechanisms and introduces new dimensions which improve the comparability of health expenditures in the perspective of universal health coverage.
Variable time span	2000 – 2019
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Global Health Observatory Data Repository - World Health Organization
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.10 Life expectancy at birth, including the UN projections

Life expectancy - Sex: all - Age: at birth - Variant: estimates

Var	Source
Variable description	The average number of remaining years of life expected by a hypothetical cohort of individuals who already reached age x and would be subject during the remainder of their lives to the mortality rates of a given period.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Life expectancy - Sex: all - Age: at birth - Variant: medium

Var	Source
Variable description	The average number of remaining years of life expected by a hypothetical cohort of individuals who already reached age x and would be subject during the remainder of their lives to the mortality rates of a given period.
Variable time span	2022 – 2100
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

9.3.11 Population by broad age group, World

Population by broad age group - Sex: all - Age: 65+ - Variant: estimates

Var	Source
Variable description	De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands. Alternative metric to 'population', used for different (and groader) age groups.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Population by broad age group - Sex: all - Age: 25-64 - Variant: estimates

Var	Source
Variable description	De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands. Alternative metric to 'population', used for different (and groader) age groups.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Population by broad age group - Sex: all - Age: 15-24 - Variant: estimates

Var	Source
Variable description	De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands. Alternative metric to 'population', used for different (and groader) age groups.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Population by broad age group - Sex: all - Age: 5-14 - Variant: estimates

Var	Source
Variable description	De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands. Alternative metric to 'population', used for different (and groader) age groups.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Population by broad age group - Sex: all - Age: 0-4 - Variant: estimates

Var	Source
Variable description	De facto population in a country, area or region as of 1 July of the year indicated. Figures are presented in thousands. Alternative metric to 'population', used for different (and groader) age groups.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

9.3.12 Share in poverty vs. Life expectancy

Life expectancy at birth, total (years)

Var	Source
Variable description	Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.Limitations and exceptions: Annual data series from United Nations Population Division's World Population Prospects are interpolated data from 5-year period data. Therefore they may not reflect real events as much as observed data.Statistical concept and methodology: Life expectancy at birth used here is the average number of years a newborn is expected to live if mortality patterns at the time of its birth remain constant in the future. It reflects the overall mortality level of a population, and summarizes the mortality pattern that prevails across all age groups in a given year. It is calculated in a period life table which provides a snapshot of a population's mortality pattern at a given time. It therefore does not reflect the mortality pattern that a person actually experiences during his/her life, which can be calculated in a cohort life table.High mortality in young age groups significantly lowers the life expectancy at birth. But if a person survives his/her childhood of high mortality, he/she may live much longer. For example, in a population with a life expectancy at birth of 50, there may be few people dying at age 50. The life expectancy at birth may be low due to the high childhood mortality so that once a person survives his/her childhood, he/she may live much longer than 50 years.
Variable time span	1960 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Household surveys, UN Population Division
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

$3.65 a day - share of population below poverty line

Var	Source
Variable description	% of population living in households with an income or expenditure per person below $3.65 a day
Variable time span	1967 – 2021
Data published by	World Bank Poverty and Inequality Platform
Data publisher's source	World Bank Data are based on primary household survey data obtained from government statistical agencies and World Bank country departments.
Link	https://pip.worldbank.org/
Retrieved	2022-10-03

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.13 Life expectancy

Life expectancy

Var	Source
Variable description	Life expectancy at birth is defined as the average number of years that a newborn could expect to live if he or she were to pass through life subject to the age-specific mortality rates of a given period.
Variable time span	1543 – 2019
Data published by	James C. Riley (2005) – Estimates of Regional and Global Life Expectancy, 1800–2001. Issue Population and Development Review. Population and Development Review. Volume 31, Issue 3, pages 537–543, September 2005., Zijdeman, Richard; Ribeira da Silva, Filipa, 2015, "Life Expectancy at Birth (Total)", http://hdl.handle.net/10622/LKYT53, IISH Dataverse, V1, and UN Population Division (2019)
Data publisher's source	https://www.lifetable.de/RileyBib.pdf
Link	https://datasets.socialhistory.org/dataset.xhtml?persistentId=hdl:10622/LKYT53, http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4457.2005.00083.x/epdf, https://population.un.org/wpp/Download/Standard/Population/
Retrieved	14/09/2019

9.3.14 Life satisfaction vs. life expectancy

Life expectancy - Sex: all - Age: at birth - Variant: estimates

Var	Source
Variable description	The average number of remaining years of life expected by a hypothetical cohort of individuals who already reached age x and would be subject during the remainder of their lives to the mortality rates of a given period.
Variable time span	1950 – 2021
Data published by	United Nations - Population Division (2022)
Link	https://population.un.org/wpp/Download/
Retrieved	2022-09-09

Life satisfaction in Cantril Ladder (World Happiness Report 2022)

Var	Source
Variable description	Average of survey responses to the 'Cantril Ladder' question in the Gallup World Poll. The survey question asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.
Variable geographic coverage	Global by country
Variable time span	2003 – 2020
Data published by	World Happiness Report 2022
Data publisher's source	Gallup World Poll surveys (life evaluation question)
Link	https://worldhappiness.report/ed/2022/#appendices-and-data
Retrieved	2022-05-12

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.15 Healthy life expectancy and years lived with disability

Years lived with disability

Var	Source
Variable geographic coverage	global by country
Variable time span	1990 – 2016
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	15th February 2018

Healthy Life Expectancy (IHME)

Var	Source
Variable geographic coverage	global by country
Variable time span	1990 – 2016
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2016 (GBD 2016) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2017.
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	15th February 2018

9.3.16 Prevalence of alcohol use disorders by age

Prevalence - Alcohol use disorders - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 15 to 19 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 15-49 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 5-14 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.17 Prevalence of alcohol use disorders in males vs. females

Prevalence - Alcohol use disorders - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Alcohol use disorders - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.18 Child mortality vs. Prevalence of child wasting

Mortality rate, under-5 (per 1,000 live births)

Var	Source
Variable description	Under-five mortality rate is the probability per 1,000 that a newborn baby will die before reaching age five, if subject to age-specific mortality rates of the specified year.Limitations and exceptions: Complete vital registration systems are fairly uncommon in developing countries. Thus estimates must be obtained from sample surveys or derived by applying indirect estimation techniques to registration, census, or survey data. Survey data are subject to recall error, and surveys estimating infant/child deaths require large samples because households in which a birth has occurred during a given year cannot ordinarily be preselected for sampling. Indirect estimates rely on model life tables that may be inappropriate for the population concerned. Extrapolations based on outdated surveys may not be reliable for monitoring changes in health status or for comparative analytical work.Statistical concept and methodology: The main sources of mortality data are vital registration systems and direct or indirect estimates based on sample surveys or censuses. A "complete" vital registration system - covering at least 90 percent of vital events in the population - is the best source of age-specific mortality data.Estimates of neonatal, infant, and child mortality tend to vary by source and method for a given time and place. Years for available estimates also vary by country, making comparisons across countries and over time difficult. To make neonatal, infant, and child mortality estimates comparable and to ensure consistency across estimates by different agencies, the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), which comprises the United Nations Children's Fund (UNICEF), the World Health Organization (WHO), the World Bank, the United Nations Population Division, and other universities and research institutes, developed and adopted a statistical method that uses all available information to reconcile differences. The method uses statistical models to obtain a best estimate trend line by fitting a country-specific regression model of mortality rates against their reference dates.
Variable time span	1960 – 2020
Unit conversion factor for chart	0.1
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	UN Inter-agency Group for Child Mortality Estimation
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Prevalence of wasting, weight for height (% of children under 5)

Var	Source
Variable description	Prevalence of wasting is the proportion of children under age 5 whose weight for height is more than two standard deviations below the median for the international reference population ages 0-59.
Variable time span	1983 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Joint child malnutrition estimates - UNICEF / World Health Organization / World Bank
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

9.3.19 Child mortality vs. Prevalence of stunting

Mortality rate, under-5 (per 1,000 live births)

Var	Source
Variable description	Under-five mortality rate is the probability per 1,000 that a newborn baby will die before reaching age five, if subject to age-specific mortality rates of the specified year.Limitations and exceptions: Complete vital registration systems are fairly uncommon in developing countries. Thus estimates must be obtained from sample surveys or derived by applying indirect estimation techniques to registration, census, or survey data. Survey data are subject to recall error, and surveys estimating infant/child deaths require large samples because households in which a birth has occurred during a given year cannot ordinarily be preselected for sampling. Indirect estimates rely on model life tables that may be inappropriate for the population concerned. Extrapolations based on outdated surveys may not be reliable for monitoring changes in health status or for comparative analytical work.Statistical concept and methodology: The main sources of mortality data are vital registration systems and direct or indirect estimates based on sample surveys or censuses. A "complete" vital registration system - covering at least 90 percent of vital events in the population - is the best source of age-specific mortality data.Estimates of neonatal, infant, and child mortality tend to vary by source and method for a given time and place. Years for available estimates also vary by country, making comparisons across countries and over time difficult. To make neonatal, infant, and child mortality estimates comparable and to ensure consistency across estimates by different agencies, the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), which comprises the United Nations Children's Fund (UNICEF), the World Health Organization (WHO), the World Bank, the United Nations Population Division, and other universities and research institutes, developed and adopted a statistical method that uses all available information to reconcile differences. The method uses statistical models to obtain a best estimate trend line by fitting a country-specific regression model of mortality rates against their reference dates.
Variable time span	1960 – 2020
Unit conversion factor for chart	0.1
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	UN Inter-agency Group for Child Mortality Estimation
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Prevalence of stunting, height for age (% of children under 5)

Var	Source
Variable description	Prevalence of stunting is the percentage of children under age 5 whose height for age is more than two standard deviations below the median for the international reference population ages 0-59 months. For children up to two years old height is measured by recumbent length. For older children height is measured by stature while standing. The data are based on the WHO's new child growth standards released in 2006.
Variable time span	1983 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Joint child malnutrition estimates - UNICEF / World Health Organization / World Bank
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

9.3.20 Prevalence of ADHD in males vs females

Prevalence - Attention-deficit/hyperactivity disorder - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Attention-deficit/hyperactivity disorder - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.21 Prevalence of anxiety disorders by age

Prevalence - Anxiety disorders - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: Under 5 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 15 to 19 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 5-14 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 15-49 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.22 Prevalence of anxiety disorders, males vs. females

Prevalence - Anxiety disorders - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Anxiety disorders - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.23 Prevalence of bipolar disorder by age

Prevalence - Bipolar disorder - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 5-14 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 15-49 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.24 Prevalence of bipolar disorder in males vs. females

Prevalence - Bipolar disorder - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Bipolar disorder - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.25 Prevalence of depression by age

Prevalence - Depressive disorders - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 15 to 19 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Both - Age: 15-49 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.26 Prevalence of depression, males vs. females

Prevalence - Depressive disorders - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Depressive disorders - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.27 Prevalence of eating disorders by age

Prevalence - Eating disorders - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 5-14 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Eating disorders - Sex: Both - Age: 15 to 19 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.28 Prevalence of schizophrenia by age

Prevalence - Schizophrenia - Sex: Both - Age: 15 to 19 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 10 to 14 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 25 to 29 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 5-14 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 50-69 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 15-49 years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 70+ years (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 20 to 24 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: 30 to 34 (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Both - Age: All Ages (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

9.3.29 Prevalence of schizophrenia in males vs. females

Prevalence - Schizophrenia - Sex: Male - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Prevalence - Schizophrenia - Sex: Female - Age: Age-standardized (Percent)

Var	Source
Variable time span	1990 – 2019
Data published by	Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021.
Data publisher's source	Institute for Health Metrics and Evaluation, Global Burden of Disease (2019)
Link	http://ghdx.healthdata.org/gbd-results-tool
Retrieved	2021-10-05

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.30 Average years of schooling vs. GDP per capita

Average Total Years of Schooling for Adult Population (Lee-Lee (2016), Barro-Lee (2018) and UNDP (2018))

Var	Source
Variable description	Mean years of total schooling across all education levels.
Variable time span	1870 – 2017
Data published by	Lee-Lee (2016); Barro-Lee (2018) and UNDP HDR (2018). See dataset description for full references.
Link	http://www.barrolee.com/Lee_Lee_LRdata_dn.htm; http://www.barrolee.com/data/yrsch2.htm; and http://hdr.undp.org/en/indicators/103006
Retrieved	24/05/19

GDP per capita, PPP (constant 2017 international $)

Var	Source
Variable description	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the country plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2017 international dollars.Statistical concept and methodology: For the concept and methodology of 2017 PPP, please refer to the International Comparison Program (ICP)'s website (https://www.worldbank.org/en/programs/icp).
Variable time span	1990 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Household surveys, UN Population Division
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.31 How important is God in your life? vs GDP per capita, 2014

How important is God in your life - wvs_godimp

Var	Source
Variable description	How important is God in your life? \ \1. Not at all important \2. \3. \4. \5. \6. \7. \8. \9. \10. Very important
Variable time span	1981 – 2014
Data published by	The Quality of Government Institute
Data publisher's source	World Values Survey / European Values Survey
Link	http://qog.pol.gu.se/data, http://www.worldvaluessurvey.org/
Retrieved	25-July-17

GDP per capita, PPP (constant 2017 international $)

Var	Source
Variable description	GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser's prices is the sum of gross value added by all resident producers in the country plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2017 international dollars.Statistical concept and methodology: For the concept and methodology of 2017 PPP, please refer to the International Comparison Program (ICP)'s website (https://www.worldbank.org/en/programs/icp).
Variable time span	1990 – 2020
Data published by	World Development Indicators - World Bank (2022.05.26)
Data publisher's source	Demographic and Health Surveys, Multiple Indicator Cluster Surveys, Household surveys, UN Population Division
Link	https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators
Retrieved	2022-06-02

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.32 Self-reported life satisfaction, 2003 to 2020

Life satisfaction in Cantril Ladder (World Happiness Report 2022)

Var	Source
Variable description	Average of survey responses to the 'Cantril Ladder' question in the Gallup World Poll. The survey question asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.
Variable geographic coverage	Global by country
Variable time span	2003 – 2020
Data published by	World Happiness Report 2022
Data publisher's source	Gallup World Poll surveys (life evaluation question)
Link	https://worldhappiness.report/ed/2022/#appendices-and-data
Retrieved	2022-05-12

9.3.33 Happiness vs. life satisfaction, 2015

Life satisfaction in Cantril Ladder (World Happiness Report 2022)

Var	Source
Variable description	Average of survey responses to the 'Cantril Ladder' question in the Gallup World Poll. The survey question asks respondents to think of a ladder, with the best possible life for them being a 10, and the worst possible life being a 0.
Variable geographic coverage	Global by country
Variable time span	2003 – 2020
Data published by	World Happiness Report 2022
Data publisher's source	Gallup World Poll surveys (life evaluation question)
Link	https://worldhappiness.report/ed/2022/#appendices-and-data
Retrieved	2022-05-12

IR HEADINGS[2]

Var	Source
Variable description	Share of people who say they are 'very happy' or 'rather happy'. Full question asks: "Taking all things together, would you say you are (i) Very happy, (ii) Rather happy, (iii) Not very happy, (iv) Not at all happy, (v) Don't know".
Variable geographic coverage	Global by country
Variable time span	1984 – 2014
Data published by	World Value Survey (Longitudinal Multiple-Wave file, version 2015)
Data publisher's source	Household surveys (representative at the national level)
Link	http://www.worldvaluessurvey.org/
Retrieved	27/07/2016

Population (historical estimates)

Var	Source
Variable description	Population by country, available from 10,000 BCE to 202110,000 BCE - 1799: Historical estimates by HYDE (v3.2).1800-1949: Historical estimates by Gapminder.1950-2022: Population records by the United Nations - Population Division (2022).
Variable time span	10,000 BCE – 2021
Data published by	Gapminder (v6); United Nations - Population Division (2022); HYDE (v3.2)
Link	https://www.gapminder.org/data/documentation/gd003/ ; https://population.un.org/wpp/Download/Standard/Population/ ; https://dataportaal.pbl.nl/downloads/HYDE/
Retrieved	October 8, 2021

9.3.34 Civil liberties, 2021

INDIV__LIBS__VDEM__OWID

Var	Source
Variable description	The variable denotes the best estimate of the extent to which extent to which citizens enjoy physical integrity rights, freedoms of religion, of movement, and from forced labor, as well as access to justice, laws are transparent and public administration is impartial.It is based on the V-Dem variable v2xcl_rol, and we expand it to cover more years for some countries.The variables ranges from 0 to 1 (most rights).
Variable time span	1789 – 2021
Data published by	Our World in Data, Bastian Herre
Data publisher's source	Data comes from the Varieties of Democracy project (v12), sometimes using Lührmann et al's (2018) Regimes of the World classification, and own expansions and refinements.
Link	http://v-dem.net/vdemds.html
Retrieved	March 2, 2022

INDIV__LIBS__VDEM__HIGH__OWID

Var	Source
Variable description	The variable denotes the upper-bound estimate of the extent to which extent to which citizens enjoy physical integrity rights, freedoms of religion, of movement, and from forced labor, as well as access to justice, laws are transparent and public administration is impartial.It is based on the V-Dem variable v2xcl_rol_codehigh, and we expand it to cover more years for some countries.The variables ranges from 0 to 1 (most rights).
Variable time span	1789 – 2021
Data published by	Our World in Data, Bastian Herre
Data publisher's source	Data comes from the Varieties of Democracy project (v12), sometimes using Lührmann et al's (2018) Regimes of the World classification, and own expansions and refinements.
Link	http://v-dem.net/vdemds.html
Retrieved	March 2, 2022

INDIV__LIBS__VDEM__LOW__OWID

Var	Source
Variable description	The variable denotes the lower-bound estimate of the extent to which extent to which citizens enjoy physical integrity rights, freedoms of religion, of movement, and from forced labor, as well as access to justice, laws are transparent and public administration is impartial.It is based on the V-Dem variable v2xcl_rol_codelow, and we expand it to cover more years for some countries.The variables ranges from 0 to 1 (most rights).
Variable time span	1789 – 2021
Data published by	Our World in Data, Bastian Herre
Data publisher's source	Data comes from the Varieties of Democracy project (v12), sometimes using Lührmann et al's (2018) Regimes of the World classification, and own expansions and refinements.
Link	http://v-dem.net/vdemds.html
Retrieved	March 2, 2022

9.3.35 Share of people agreeing with the statement "most people can be trusted"

Trust in others (World Values Survey (2014))

Var	Source
Variable description	Self-reported trust in others is constructed as the number of people responding to the question "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Possible answers include "Most people can be trusted", "Don't know" and "Can't be too careful".
Variable geographic coverage	Global by country
Variable time span	1984 – 2014
Data published by	World Values Survey (2014)
Data publisher's source	World Values Survey Data, Longitudinal
Link	http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp
Retrieved	27/07/2016