STAT6020: Predictive Analytics
Project (20%)

Project Overview

This assessment involves writing a report that summarizes a statistical learning-related investigation that you have conducted on the data provided.

The data set is provided, and an objective is very broadly defined, but it is up to you to clarify in your report exactly what you will investigate your methods and results. It is quite possible that different students might come up with very different yet still valid analyses based on the same data set. You don't necessarily have to use all of the data, and you don't necessarily need to provide an exhaustive analysis that extracts every possible shred of information from this data, but you do need to clearly document your targeted investigation and how your results relate to the broad objective provided. You don't need to use every possible method, but there should be some justification for the methods that you do use.

We don't expect a Nobel prize-winning analysis, but your report should demonstrate that:

- You have grasped important concepts associated with this course.
- You can communicate your investigation in a formal written manner.

Your investigation should use R or Python to analyse data using methods from at least two of the following course content areas:

- Classification (using LDA, QDA, KNN, or logistic regression)
- Linear regression, possibly including regularisation
- Decision trees and/or ensembles for classification or regression
- Principal Component Analysis (PCA)
- Cluster Analysis
- Outlier Detection
- Support Vector Machines
- Formal Quantitative Assessment of Results (e.g., cross-validation for model assessment and selection in regression or classification)
- Qualitative Assessment through Visualisation of Results (e.g. visualisation and interpretation of clustering hierarchies, visual interpretation of PCA, etc.)

If you need to clarify the project requirements or if you have technical issues you may post in the appropriate Discussion Forum but your post should not given any indication of your planned investigation, methods or results.

Your submission must consist of your own work in accordance with the Academic Integrity Policy.

Report Format and Marking Guide

For students who prepare their work using R: You need to submit a source code file (.rmd) and a PDF report. The markers may NOT necessarily run your code, but they may refer to this file while marking your report, if some clarification is needed.

For students who prepare their work using Python: You need to submit a source code file (.ipynb) and a PDF report. The markers may NOT necessarily run your code, but they may refer to this file while marking your report, if some clarification is needed.

The report should have the following sections marked clearly:

Title: (2 marks) In today's busy world, it is very important to make the most of your title. Make the title concise (< 20 words) and 'eye-catching' yet an informative and accurate representation of the contents of the report. Author should be listed below the title.

Abstract: (6 marks) The abstract provides a brief overview of the report contents in around 200 – 300 words. An abstract typically consists of:

- Introductory statement: background to the study, important issue(s) the report addresses. (approximately 1 to 2 sentences)
- Purpose of the report: state the objectives (1-2 sentences)
- Methodological approach: overview the data and methods (2-3 sentences)
- Findings or Achievements: list one or two of the main findings or achievements from your investigation (1-2 sentences)
- Conclusions and Implications: what conclusions can be drawn from your investigation? How can the findings/achievements in your report deliver a benefit to people, things, systems or processes? (1-2 sentences)

Introduction: (12 marks) The introduction sets the scene for the investigative efforts. It provides motivation for the work and relevant background information and references that will enable the reader to put in context the key objectives and achievements in your report. Address the important issues that have motivated your investigation. At the end of the introduction clearly state the objectives of the report. Do not put any results from your investigation in the introduction. Do not discuss details about the data and methods in this section. Do not discuss your conclusions or key findings in the introduction.

Data: (10 marks) This section should enable the reader to understand how the data was obtained, pre-processed (if applicable) and what the data represent.

Methods: (30 marks) This section should summarise the statistical learning methods that were used to process and to analyse the data. The methods should be appropriate to ensure that the objectives of the report are met. You are strongly encouraged to interleave your text with key calls to R functions that generate relevant results that you may want to highlight, just like the weekly course notes and labs. This can be achieved straightforwardly using R Markdown. You can use the R Markdown chunk option echo=FALSE to hide chunks of code that you judge less relevant, but these must still be present in the source code for verification, if necessary. Your code will be assessed for correctness, organisation and clarity as part of the Methods section. However, there should be enough detail in your textual description for your methodology to be repeated by an independent person, without having to refer to and decipher your code.

Results and Discussion: (20 marks) The results are explained correctly, clearly, and in sufficient detail. The results and discussion clearly follow from the data collection and the methods. (In fact, although it is more usual to have separate sections, you may find it more appropriate to merge the Results and Discussion with your explanation of the methodology.) The discussion centres on the outputs from the statistical learning procedures that you have performed. For example, what are the main outcomes? Why are they useful and what for? How are they interesting and why? What are the main achievements and their implications?

Conclusions: (10 marks) Final remarks about the key achievements of the investigations and what makes them "interesting" or "useful". Achievements or findings should be linked with the original objectives or hypotheses of the project. Are there any recommended actions from your analysis? Make sure that you mention any limitations of your work here. Limit the conclusion to no more than two or three paragraphs.

References: List any sources your investigation has drawn from. Note that all references should be referred to in the text.

Appendices: Appendices can be useful when the incorporation of material in the body of the work would make it poorly structured or too long and detailed. Appendices may be used for helpful, supporting or essential material that would otherwise clutter, break up or be distracting to the text.

In addition to the marks allocated to each section, there are an extra **10 marks** associated with the **general quality of the writing and presentation**. For example, in a high-quality report:

The material is coherently organized and the logic is easy to follow. There are no spelling or grammatical errors and terminology is clearly defined. Writing is clear and concise and persuasive. Each Figure/Table is numbered, followed by a caption, and referred to in the body of the text, most noticeably in the results and/or discussion section. The Figures/Tables provided reinforce the most relevant achievements of the work. Any references are listed at the end of the report, with citations in the body of the report. Appendices are appropriately used.

The report proper (i. e., not including references or appendices) should be no more than about 10 single-column pages when printed in A4 format using R Markdown default settings for font, font size, line spacing, margins, etc. If you exceed this recommendation (including through inappropriate use of appendices to include extra material) then the

marker may stop reading (and marking) after 10 pages.

World Bank Climate Change Data

Climate change is the critically important issue of our time and affects everyone on the planet. It is also a subject where there is a huge amount of data collected, analysed, written about and argued about. The wbcc bc.csv file is available on Canvas. The variables (columns) correspond to the country characteristics or indicators collected by the World Bank that it has categorised as relevant to climate change. Each record (row) corresponds to a particular country. Although much climate change analysis focusses on how things are changing over time, this data set is cross-sectional rather than longitudinal. It is a snapshot of a single recent value of these characteristics for each country. Table 3 lists the indicators and their descriptions.

Your task is to explore this data set and report on any interesting and relevant relationships you may discover between these indicators. It is also possible that the lack of an expected relationship might also be interesting. Do these climate change related indicators suggest particular groupings of countries? How does Australia compare to other countries in the world based on these indicators? Are there any findings that you think would be relevant to highlight to Australian leaders or to world leaders at the UN Climate Change Conference later this month in Glasgow?

Hints:

- Yes, the objective here is very ill defined. Part of your task is to clarify a more specific objective. This should form part of the introduction for your report.
- The questions in the previous paragraph are just starting points. They should not limit your project, or even necessarily be considered at all if there are other aspects of the data you would prefer to explore.
- Some initial exploratory analysis is likely to prove useful. (May or may not be relevant to include in your final report.)
- Probably of even more importance is some careful thought about the variables before rushing into analysis.
- It may not be appropriate to use all of the variables or all of the countries in your analysis (or all parts of your analysis).
- Some variables may be appropriate to use as response or outcome variables but others not.
- Some of the variables are clearly derived from others so be careful in choosing which to include in an analysis.
- Note that this is real data extracted using the code shown in the appendix. There may be mistakes and certainly plenty of missing data that you will need to deal with.
- Some of the data is more recent than others. If you look at the Appendix you will see that each value is the most recent value of that indicator available for that country between 2001 and 2020. So the lack of a common time point for our snapshot of data is certainly a flaw that should be acknowledged.

Appendix: World Bank Data

The World Bank Group is a global partnership comprising five related institutions "working for sustainable solutions that reduce poverty and build shared prosperity in developing countries." https://www.worldbank.org/en/who-we-are

One aspect of their work is the provision of free and open access to global development data. There is a huge amount of data available at https://data.worldbank.org/ in many different forms. The World Bank uses the term "indicators" for the variables it collects to characterise countries and/or regions over time. The data is compiled from officially recognized international sources and presents what is considered to be some of the most current and accurate global development data available.

The https://databank.worldbank.org/source/world-development-indicators web page provides an interface that can be used to extract and download data from the World Development Indicators (WDI) collection. This is the primary World Bank collection of development indicators and is the default collection (or database) that is pre-selected for this web interface but other World Bank databases can be selected instead. Alternatively, as shown below, there is an R package that can be used to access data directly. There is no particular need for you to use this code (or the web interface) as the data has already been extracted for you. However, the code is shown here for your interest. Some of this code may not have been explicitly covered in this course but you should be able to follow it with the help of the comments.

```
library(tidyverse) # for more data manipulation functions
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1
--
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts()
--
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(wbstats) # for easy access to World Bank databases from R
library(kableExtra) # for nicer printing of tables
```

```
##
## Attaching package:  'kableExtra'
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
wbind <- wb_indicators() # list of the indicators (variables) available
```

Note that there are 19220 different indicators available but as is common with real world data there are a lot of missing values. Not all indicators are available for all countries for all time periods.

The indicators are organised into a number of different ‘collections’ or databases which are listed in Table 1.

```
# list World Bank collections
wbsrc <- unique(wbind[c("source_id", "source")]) %>%
  arrange(source_id)

kbl(wbsrc, longtable=TRUE, booktabs=TRUE,
    caption="List of World Bank data `collections'.") %>%
  kable_styling(latex_options = c("repeat_header"))
```

Table 1: List of World Bank data 'collections'.

source_id	source
1	Doing Business
2	World Development Indicators
3	Worldwide Governance Indicators
5	Subnational Malnutrition Database
11	Africa Development Indicators
12	Education Statistics
13	Enterprise Surveys
14	Gender Statistics
15	Global Economic Monitor
16	Health Nutrition and Population Statistics
18	IDA Results Measurement System
19	Millennium Development Goals
20	Quarterly Public Sector Debt
22	Quarterly External Debt Statistics SDDS
23	Quarterly External Debt Statistics GDDS
24	Poverty and Equity
25	Jobs
27	Global Economic Prospects
28	Global Financial Inclusion
29	The Atlas of Social Protection: Indicators of Resilience and Equity
30	Exporter Dynamics Database – Indicators at Country-Year Level
32	Global Financial Development
33	G20 Financial Inclusion Indicators
34	Global Partnership for Education
35	Sustainable Energy for All
36	Statistical Capacity Indicators
37	LAC Equity Lab
39	Health Nutrition and Population Statistics by Wealth Quintile
41	Country Partnership Strategy for India (FY2013 - 17)
43	Adjusted Net Savings
45	Indonesia Database for Policy and Economic Research
46	Sustainable Development Goals
50	Subnational Population

Table 1: List of World Bank data 'collections'. *(continued)*

source_id	source
54	Joint External Debt Hub
59	Wealth Accounts
60	Economic Fitness
61	PPPs Regulatory Quality
63	Human Capital Index
64	Worldwide Bureaucracy Indicators
65	Health Equity and Financial Protection Indicators
66	Logistics Performance Index
67	PEFA 2011
69	Global Financial Inclusion and Consumer Protection Survey
70	Economic Fitness 2
71	International Comparison Program (ICP) 2005
73	Global Financial Inclusion and Consumer Protection Survey (Internal)
75	Environment, Social and Governance (ESG) Data
78	ICP 2017
80	Gender Disaggregated Labor Database (GDLD)
81	International Debt Statistics: DSSI
82	Global Public Procurement
83	Statistical Performance Indicators (SPI)
84	Education Policy
86	Global Jobs Indicators Database (JOIN)
87	Country Climate and Development Report (CCDR)

```
# extract topic info to a more easily managed form
# there are more efficient ways of doing this but this code
# is probably more easily understood
wbit <- NULL # wb indicators with topics
for (i in 1:nrow(wbind)) { # for each available indicator
  nrc <- dim(wbind$topics[[i]]) # rows and columns for the topics dataframe
                                # for that indicator
  if (nrc[1]!=0 & nrc[2]!=0) {  # if there is valid topic info
    topics <- cbind(indicator_id=rep(wbind$indicator_id[i],nrc[1]),
                    wbind$topics[[i]])
    wbit <- rbind(wbit, topics)
    }
}
wbit$id <- as.numeric(wbit$id)
wbtop <- unique(wbit[c("id", "value")]) %>%
  arrange(id)

kbl(wbtop, booktabs=TRUE, position="t",
    caption="Topics used to categorize World Bank Indicators")
```

5508 of the indicators are categorized into the 'topics' shown in Table 2. Some indicators are associated with more than one topic.

Table 2: Topics used to categorize World Bank Indicators

id	value
1	Agriculture & Rural Development
2	Aid Effectiveness
3	Economy & Growth
4	Education
5	Energy & Mining
6	Environment
7	Financial Sector
8	Health
9	Infrastructure
10	Social Protection & Labor
11	Poverty
12	Private Sector
13	Public Sector
14	Science & Technology
15	Social Development
16	Urban Development
17	Gender
18	Millenium development goals
19	Climate Change
20	External Debt
21	Trade

```
# create list of indicator ids within the "Climate Change" topic
ccvars <- wbit %>%
  filter(value=="Climate Change") %>%
  select(indicator_id)

ccvars <- ccvars$indicator_id # change from data frame to vector
# extract details for indicators categorized as "Climate Change" related
wbindcc <- wbind %>%
  filter(indicator_id %in% as.vector(ccvars)) %>%
  select(indicator_id,
         indicator,
         indicator_desc)

kbl(wbindcc, booktabs=TRUE, longtable=TRUE,
    caption="World Bank Indicators categorised under Climate Change topic") %>%
  column_spec(2, width="60mm") %>%
  column_spec(3, width="120mm") %>%
  kable_styling(font_size=8) %>%
  landscape()
```

Table 3: World Bank Indicators categorised under Climate Change topic

indicator_id	indicator	indicator_desc
AG.LND.AGRI.K2	Agricultural land (sq. km)	Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures. Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Land under permanent crops is land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber. This category includes land under flowering shrubs, fruit trees, nut trees, and vines, but excludes land under trees grown for wood or timber. Permanent pasture is land used for five or more years for forage, including natural and cultivated crops.
AG.LND.AGRI.ZS	Agricultural land (% of land area)	Agricultural land refers to the share of land area that is arable, under permanent crops, and under permanent pastures. Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded. Land under permanent crops is land cultivated with crops that occupy the land for long periods and need not be replanted after each harvest, such as cocoa, coffee, and rubber. This category includes land under flowering shrubs, fruit trees, nut trees, and vines, but excludes land under trees grown for wood or timber. Permanent pasture is land used for five or more years for forage, including natural and cultivated crops.
AG.LND.ARBL.ZS	Arable land (% of land area)	Arable land includes land defined by the FAO as land under temporary crops (double-cropped areas are counted once), temporary meadows for mowing or for pasture, land under market or kitchen gardens, and land temporarily fallow. Land abandoned as a result of shifting cultivation is excluded.
AG.LND.EL5M.RU.K2	Rural land area where elevation is below 5 meters (sq. km)	Rural land area below 5m is the total rural land area in square kilometers where the elevation is 5 meters or less.
AG.LND.EL5M.RU.ZS	Rural land area where elevation is below 5 meters (% of total land area)	Rural land area below 5m is the percentage of total land where the rural land elevation is 5 meters or less.
AG.LND.EL5M.UR.K2	Urban land area where elevation is below 5 meters (sq. km)	Urban land area below 5m is the total urban land area in square kilometers where the elevation is 5 meters or less.
AG.LND.EL5M.UR.ZS	Urban land area where elevation is below 5 meters (% of total land area)	Urban land area below 5m is the percentage of total land where the urban land elevation is 5 meters or less.
AG.LND.EL5M.ZS	Land area where elevation is below 5 meters (% of total land area)	Land area below 5m is the percentage of total land where the elevation is 5 meters or less.
AG.LND.FRST.K2	Forest area (sq. km)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.
AG.LND.FRST.ZS	Forest area (% of land area)	Forest area is land under natural or planted stands of trees of at least 5 meters in situ, whether productive or not, and excludes tree stands in agricultural production systems (for example, in fruit plantations and agroforestry systems) and trees in urban parks and gardens.

AG.LND.IRIG.AG.ZS	Agricultural irrigated land (% of total agricultural land)	Agricultural irrigated land refers to agricultural areas purposely provided with water, including land irrigated by controlled flooding.
AG.LND.PRCP.MM	Average precipitation in depth (mm per year)	Average precipitation is the long-term average in depth (over space and time) of annual precipitation in the country. Precipitation is defined as any kind of water that falls from clouds as a liquid or a solid.
AG.YLD.CREL.KG	Cereal yield (kg per hectare)	Cereal yield, measured as kilograms per hectare of harvested land, includes wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains. Production data on cereals relate to crops harvested for dry grain only. Cereal crops harvested for hay or harvested green for food, feed, or silage and those used for grazing are excluded. The FAO allocates production data to the calendar year in which the bulk of the harvest took place. Most of a crop harvested near the end of a year will be used in the following year.
BX.KLT.DINV.WD.GD.ZS	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows net inflows (new investment inflows less disinvestment) in the reporting economy from foreign investors, and is divided by GDP.
EG.ELC.ACCS.ZS	Access to electricity (% of population)	Access to electricity is the percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources.
EG.ELC.COAL.ZS	Electricity production from coal sources (% of total)	Sources of electricity refer to the inputs used to generate electricity. Coal refers to all coal and brown coal, both primary (including hard coal and lignite-brown coal) and derived fuels (including patent fuel, coke oven coke, gas coke, coke oven gas, and blast furnace gas). Peat is also included in this category.
EG.ELC.HYRO.ZS	Electricity production from hydroelectric sources (% of total)	Sources of electricity refer to the inputs used to generate electricity. Hydropower refers to electricity produced by hydroelectric power plants.
EG.ELC.NGAS.ZS	Electricity production from natural gas sources (% of total)	Sources of electricity refer to the inputs used to generate electricity. Gas refers to natural gas but excludes natural gas liquids.
EG.ELC.NUCL.ZS	Electricity production from nuclear sources (% of total)	Sources of electricity refer to the inputs used to generate electricity. Nuclear power refers to electricity produced by nuclear power plants.
EG.ELC.PETR.ZS	Electricity production from oil sources (% of total)	Sources of electricity refer to the inputs used to generate electricity. Oil refers to crude oil and petroleum products.
EG.ELC.RNEW.ZS	Renewable electricity output (% of total electricity output)	Renewable electricity is the share of electrity generated by renewable power plants in total electricity generated by all types of plants.
EG.ELC.RNWX.KH	Electricity production from renewable sources, excluding hydroelectric (kWh)	Electricity production from renewable sources, excluding hydroelectric, includes geothermal, solar, tides, wind, biomass, and biofuels.
EG.ELC.RNWX.ZS	Electricity production from renewable sources, excluding hydroelectric (% of total)	Electricity production from renewable sources, excluding hydroelectric, includes geothermal, solar, tides, wind, biomass, and biofuels.
EG.FEC.RNEW.ZS	Renewable energy consumption (% of total final energy consumption)	Renewable energy consumption is the share of renewables energy in total final energy consumption.
EG.USE.COMM.GD.PP.KD	Energy use (kg of oil equivalent) per $1,000 GDP (constant 2017 PPP)	Energy use per PPP GDP is the kilogram of oil equivalent of energy use per constant PPP GDP. Energy use refers to use of primary energy before transformation to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport. PPP GDP is gross domestic product converted to 2017 constant international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as a U.S. dollar has in the United States.

EG.USE.ELEC.KH.PC	Electric power consumption (kWh per capita)	Electric power consumption measures the production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants.
EG.USE.PCAP.KG.OE	Energy use (kg of oil equivalent per capita)	Energy use refers to use of primary energy before transformation to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport.
EN.ATM.CO2E.EG.ZS	CO2 intensity (kg per kg of oil equivalent energy use)	Carbon dioxide emissions from solid fuel consumption refer mainly to emissions from use of coal as an energy source.
EN.ATM.CO2E.GF.KT	CO2 emissions from gaseous fuel consumption (kt)	Carbon dioxide emissions from liquid fuel consumption refer mainly to emissions from use of natural gas as an energy source.
EN.ATM.CO2E.GF.ZS	CO2 emissions from gaseous fuel consumption (% of total)	Carbon dioxide emissions from liquid fuel consumption refer mainly to emissions from use of natural gas as an energy source.
EN.ATM.CO2E.KD.GD	CO2 emissions (kg per 2010 US$ of GDP)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
EN.ATM.CO2E.KT	CO2 emissions (kt)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
EN.ATM.CO2E.LF.KT	CO2 emissions from liquid fuel consumption (kt)	Carbon dioxide emissions from liquid fuel consumption refer mainly to emissions from use of petroleum-derived fuels as an energy source.
EN.ATM.CO2E.LF.ZS	CO2 emissions from liquid fuel consumption (% of total)	Carbon dioxide emissions from liquid fuel consumption refer mainly to emissions from use of petroleum-derived fuels as an energy source.
EN.ATM.CO2E.PC	CO2 emissions (metric tons per capita)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
EN.ATM.CO2E.PP.GD	CO2 emissions (kg per PPP $ of GDP)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
EN.ATM.CO2E.PP.GD.KD	CO2 emissions (kg per 2017 PPP $ of GDP)	Carbon dioxide emissions are those stemming from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring.
EN.ATM.CO2E.SF.KT	CO2 emissions from solid fuel consumption (kt)	Carbon dioxide emissions from solid fuel consumption refer mainly to emissions from use of coal as an energy source.
EN.ATM.CO2E.SF.ZS	CO2 emissions from solid fuel consumption (% of total)	Carbon dioxide emissions from solid fuel consumption refer mainly to emissions from use of coal as an energy source.
EN.ATM.GHGO.KT.CE	Other greenhouse gas emissions, HFC, PFC and SF6 (thousand metric tons of CO2 equivalent)	Other greenhouse gas emissions are by-product emissions of hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride.
EN.ATM.GHGO.ZG	Other greenhouse gas emissions (% change from 1990)	Other greenhouse gas emissions are by-product emissions of hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride. Each year of data shows the percentage change to that year from 1990.
EN.ATM.GHGT.KT.CE	Total greenhouse gas emissions (kt of CO2 equivalent)	Total greenhouse gas emissions in kt of CO2 equivalent are composed of CO2 totals excluding short-cycle biomass burning (such as agricultural waste burning and savanna burning) but including other biomass burning (such as forest fires, post-burn decay, peat fires and decay of drained peatlands), all anthropogenic CH4 sources, N2O sources and F-gases (HFCs, PFCs and SF6).

EN.ATM.GHGT.ZG	Total greenhouse gas emissions (% change from 1990)	Total greenhouse gas emissions are composed of CO2 totals excluding short-cycle biomass burning (such as agricultural waste burning and savanna burning) but including other biomass burning (such as forest fires, post-burn decay, peat fires and decay of drained peatlands), all anthropogenic CH4 sources, N2O sources and F-gases (HFCs, PFCs and SF6). Each year of data shows the percentage change to that year from 1990.
EN.ATM.HFCG.KT.CE	HFC gas emissions (thousand metric tons of CO2 equivalent)	Hydrofluorocarbons, used as a replacement for chlorofluorocarbons, are used mainly in refrigeration and semiconductor manufacturing.
EN.ATM.METH.KT.CE	Methane emissions (kt of CO2 equivalent)	Methane emissions are those stemming from human activities such as agriculture and from industrial methane production.
EN.ATM.METH.ZG	Methane emissions (% change from 1990)	Methane emissions are those stemming from human activities such as agriculture and from industrial methane production. Each year of data shows the percentage change to that year from 1990.
EN.ATM.NOXE.KT.CE	Nitrous oxide emissions (thousand metric tons of CO2 equivalent)	Nitrous oxide emissions are emissions from agricultural biomass burning, industrial activities, and livestock management.
EN.ATM.NOXE.ZG	Nitrous oxide emissions (% change from 1990)	Nitrous oxide emissions are emissions from agricultural biomass burning, industrial activities, and livestock management. Each year of data shows the percentage change to that year from 1990.
EN.ATM.PFCG.KT.CE	PFC gas emissions (thousand metric tons of CO2 equivalent)	Perfluorocarbons, used as a replacement for chlorofluorocarbons in manufacturing semiconductors, are a byproduct of aluminum smelting and uranium enrichment.
EN.ATM.SF6G.KT.CE	SF6 gas emissions (thousand metric tons of CO2 equivalent)	Sulfur hexafluoride is used largely to insulate high-voltage electric power equipment.
EN.CLC.DRSK.XQ	Disaster risk reduction progress score (1-5 scale; 5=best)	Disaster risk reduction progress score is an average of self-assessment scores, ranging from 1 to 5, submitted by countries under Priority 1 of the Hyogo Framework National Progress Reports. The Hyogo Framework is a global blueprint for disaster risk reduction efforts that was adopted by 168 countries in 2005. Assessments of "Priority 1" include four indicators that reflect the degree to which countries have prioritized disaster risk reduction and the strengthening of relevant institutions.
EN.CLC.GHGR.MT.CE	GHG net emissions/removals by LUCF (Mt of CO2 equivalent)	GHG net emissions/removals by LUCF refers to changes in atmospheric levels of all greenhouse gases attributable to forest and land-use change activities, including but not limited to (1) emissions and removals of CO2 from decreases or increases in biomass stocks due to forest management, logging, fuelwood collection, etc.; (2) conversion of existing forests and natural grasslands to other land uses; (3) removal of CO2 from the abandonment of formerly managed lands (e.g. croplands and pastures); and (4) emissions and removals of CO2 in soil associated with land-use change and management. For Annex-I countries under the UNFCCC, these data are drawn from the annual GHG inventories submitted to the UNFCCC by each country; for non-Annex-I countries, data are drawn from the most recently submitted National Communication where available. Because of differences in reporting years and methodologies, these data are not generally considered comparable across countries. Data are in million metric tons.

EN.CLC.MDAT.ZS	Droughts, floods, extreme temperatures (% of population, average 1990-2009)	Droughts, floods and extreme temperatures is the annual average percentage of the population that is affected by natural disasters classified as either droughts, floods, or extreme temperature events. A drought is an extended period of time characterized by a deficiency in a region's water supply that is the result of constantly below average precipitation. A drought can lead to losses to agriculture, affect inland navigation and hydropower plants, and cause a lack of drinking water and famine. A flood is a significant rise of water level in a stream, lake, reservoir or coastal region. Extreme temperature events are either cold waves or heat waves. A cold wave can be both a prolonged period of excessively cold weather and the sudden invasion of very cold air over a large area. Along with frost it can cause damage to agriculture, infrastructure, and property. A heat wave is a prolonged period of excessively hot and sometimes also humid weather relative to normal climate patterns of a certain region. Population affected is the number of people injured, left homeless or requiring immediate assistance during a period of emergency resulting from a natural disaster; it can also include displaced or evacuated people. Average percentage of population affected is calculated by dividing the sum of total affected for the period stated by the sum of the annual population figures for the period stated.
EN.POP.EL5M.RU.ZS	Rural population living in areas where elevation is below 5 meters (% of total population)	Rural population below 5m is the percentage of the total population, living in areas where the elevation is 5 meters or less.
EN.POP.EL5M.UR.ZS	Urban population living in areas where elevation is below 5 meters (% of total population)	Urban population below 5m is the percentage of the total population, living in areas where the elevation is 5 meters or less.
EN.POP.EL5M.ZS	Population living in areas where elevation is below 5 meters (% of total population)	Population below 5m is the percentage of the total population living in areas where the elevation is 5 meters or less.
EN.URB.MCTY.TL.ZS	Population in urban agglomerations of more than 1 million (% of total population)	Population in urban agglomerations of more than one million is the percentage of a country's population living in metropolitan areas that in 2018 had a population of more than one million people.
ER.H2O.FWTL.K3	Annual freshwater withdrawals, total (billion cubic meters)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use (including withdrawals for cooling thermoelectric plants). Withdrawals for domestic uses include drinking water, municipal use or supply, and use for public services, commercial establishments, and homes. Data are for the most recent year available for 1987-2002.
ER.H2O.FWTL.ZS	Annual freshwater withdrawals, total (% of internal resources)	Annual freshwater withdrawals refer to total water withdrawals, not counting evaporation losses from storage basins. Withdrawals also include water from desalination plants in countries where they are a significant source. Withdrawals can exceed 100 percent of total renewable resources where extraction from nonrenewable aquifers or desalination plants is considerable or where there is significant water reuse. Withdrawals for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use (including withdrawals for cooling thermoelectric plants). Withdrawals for domestic uses include drinking water, municipal use or supply, and use for public services, commercial establishments, and homes. Data are for the most recent year available for 1987-2002.

ER.LND.PTLD.ZS	Terrestrial protected areas (% of total land area)	Terrestrial protected areas are totally or partially protected areas of at least 1,000 hectares that are designated by national authorities as scientific reserves with limited public access, national parks, natural monuments, nature reserves or wildlife sanctuaries, protected landscapes, and areas managed mainly for sustainable use. Marine areas, unclassified areas, littoral (intertidal) areas, and sites protected under local or provincial law are excluded.
ER.MRN.PTMR.ZS	Marine protected areas (% of territorial waters)	Marine protected areas are areas of intertidal or subtidal terrain–and overlying water and associated flora and fauna and historical and cultural features–that have been reserved by law or other effective means to protect part or all of the enclosed environment.
ER.PTD.TOTL.ZS	Terrestrial and marine protected areas (% of total territorial area)	Terrestrial protected areas are totally or partially protected areas of at least 1,000 hectares that are designated by national authorities as scientific reserves with limited public access, national parks, natural monuments, nature reserves or wildlife sanctuaries, protected landscapes, and areas managed mainly for sustainable use. Marine protected areas are areas of intertidal or subtidal terrain–and overlying water and associated flora and fauna and historical and cultural features–that have been reserved by law or other effective means to protect part or all of the enclosed environment. Sites protected under local or provincial law are excluded.
IC.BUS.EASE.XQ	Ease of doing business index (1=most business-friendly regulations)	Ease of doing business ranks economies from 1 to 190, with first place being the best. A high ranking (a low numerical rank) means that the regulatory environment is conducive to business operation. The index averages the country's percentile rankings on 10 topics covered in the World Bank's Doing Business. The ranking on each topic is the simple average of the percentile rankings on its component indicators.
IQ.CPA.PUBS.XQ	CPIA public sector management and institutions cluster average (1=low to 6=high)	The public sector management and institutions cluster includes property rights and rule-based governance, quality of budgetary and financial management, efficiency of revenue mobilization, quality of public administration, and transparency, accountability, and corruption in the public sector.
IS.ROD.PAVE.ZS	Roads, paved (% of total roads)	Paved roads are those surfaced with crushed stone (macadam) and hydrocarbon binder or bituminized agents, with concrete, or with cobblestones, as a percentage of all the country's roads, measured in length.
NV.AGR.TOTL.ZS	Agriculture, forestry, and fishing, value added (% of GDP)	Agriculture corresponds to ISIC divisions 1-5 and includes forestry, hunting, and fishing, as well as cultivation of crops and livestock production. Value added is the net output of a sector after adding up all outputs and subtracting intermediate inputs. It is calculated without making deductions for depreciation of fabricated assets or depletion and degradation of natural resources. The origin of value added is determined by the International Standard Industrial Classification (ISIC), revision 3. Note: For VAB countries, gross value added at factor cost is used as the denominator.
SE.ENR.PRSC.FM.ZS	School enrollment, primary and secondary (gross), gender parity index (GPI)	Gender parity index for gross enrollment ratio in primary and secondary education is the ratio of girls to boys enrolled at primary and secondary levels in public and private schools.
SE.PRM.CMPT.ZS	Primary completion rate, total (% of relevant age group)	Primary completion rate, or gross intake ratio to the last grade of primary education, is the number of new entrants (enrollments minus repeaters) in the last grade of primary education, regardless of age, divided by the population at the entrance age for the last grade of primary education. Data limitations preclude adjusting for students who drop out during the final year of primary education.
SH.DYN.MORT	Mortality rate, under-5 (per 1,000 live births)	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.

SH.MED.CMHW.P3	Community health workers (per 1,000 people)	Community health workers include various types of community health aides, many with country-specific occupational titles such as community health officers, community health-education workers, family health workers, lady health visitors and health extension package workers.
SH.STA.MALN.ZS	Prevalence of underweight, weight for age (% of children under 5)	Prevalence of underweight children is the percentage of children under age 5 whose weight for age is more than two standard deviations below the median for the international reference population ages 0-59 months. The data are based on the WHO's child growth standards released in 2006.
SI.POV.DDAY	Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)	Poverty headcount ratio at $1.90 a day is the percentage of the population living on less than $1.90 a day at 2011 international prices. As a result of revisions in PPP exchange rates, poverty rates for individual countries cannot be compared with poverty rates reported in earlier editions.
SP.POP.GROW	Population growth (annual %)	Annual population growth rate for year t is the exponential rate of growth of midyear population from year t-1 to t, expressed as a percentage . Population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship.
SP.POP.TOTL	Population, total	Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates.
SP.URB.GROW	Urban population growth (annual %)	Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects.
SP.URB.TOTL	Urban population	Urban population refers to people living in urban areas as defined by national statistical offices. It is calculated using World Bank population estimates and urban ratios from the United Nations World Urbanization Prospects. Aggregation of urban and rural population may not add up to total population because of different country coverages.
SP.URB.TOTL.IN.ZS	Urban population (% of total population)	Urban population refers to people living in urban areas as defined by national statistical offices. The data are collected and smoothed by United Nations Population Division.

Table 3 lists and describes the 77 indicators which are categorized under the 'Climate Change' topic. Note that all of these are from the World Development Indicators collection. The source for each indicator is listed in Table 4.

```
wbindccsource <- wbind %>%
  filter(indicator_id %in% as.vector(ccvars)) %>%
  select(indicator_id,
         source_org)

kbl(wbindccsource, booktabs=TRUE, longtable=TRUE,
    caption="Sources for World Bank Climate Change Indicators") %>%
  column_spec(2, width="110mm") %>%
  kable_styling(font_size=8, latex_options = c("repeat_header"))
```

Table 4: Sources for World Bank Climate Change Indicators

indicator_id	source_org
AG.LND.AGRI.K2	Food and Agriculture Organization, electronic files and web site.
AG.LND.AGRI.ZS	Food and Agriculture Organization, electronic files and web site.
AG.LND.ARBL.ZS	Food and Agriculture Organization, electronic files and web site.
AG.LND.EL5M.RU.K2	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
AG.LND.EL5M.RU.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
AG.LND.EL5M.UR.K2	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
AG.LND.EL5M.UR.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
AG.LND.EL5M.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
AG.LND.FRST.K2	Food and Agriculture Organization, electronic files and web site.
AG.LND.FRST.ZS	Food and Agriculture Organization, electronic files and web site.
AG.LND.IRIG.AG.ZS	Food and Agriculture Organization, electronic files and web site.
AG.LND.PRCP.MM	Food and Agriculture Organization, electronic files and web site.
AG.YLD.CREL.KG	Food and Agriculture Organization, electronic files and web site.
BX.KLT.DINV.WD.GD.ZS	International Monetary Fund, International Financial Statistics and Balance of Payments databases, World Bank, International Debt Statistics, and World Bank and OECD GDP estimates.
EG.ELC.ACCS.ZS	World Bank Global Electrification Database from "Tracking SDG 7: The Energy Progress Report" led jointly by the custodian agencies: the International Energy Agency (IEA), the International Renewable Energy Agency (IRENA), the United Nations Statistics Division (UNSD), the World Bank and the World Health Organization (WHO).
EG.ELC.COAL.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.HYRO.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/

Table 4: Sources for World Bank Climate Change Indicators *(continued)*

indicator_id	source_org
EG.ELC.NGAS.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.NUCL.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.PETR.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.RNEW.ZS	IEA Statistics © OECD/IEA 2018 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.RNWX.KH	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.ELC.RNWX.ZS	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.FEC.RNEW.ZS	World Bank, Sustainable Energy for All (SE4ALL) database from the SE4ALL Global Tracking Framework led jointly by the World Bank, International Energy Agency, and the Energy Sector Management Assistance Program.
EG.USE.COMM.GD.PP.KD	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.USE.ELEC.KH.PC	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EG.USE.PCAP.KG.OE	IEA Statistics © OECD/IEA 2014 (http://www.iea.org/stats/index.asp), subject to https://www.iea.org/t&c/termsandconditions/
EN.ATM.CO2E.EG.ZS	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.GF.KT	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.GF.ZS	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.KD.GD	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. See NY.GDP.MKTP.KD for the denominator's source.
EN.ATM.CO2E.KT	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions.
EN.ATM.CO2E.LF.KT	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.LF.ZS	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.PC	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. See SP.POP.TOTL for the denominator's source.
EN.ATM.CO2E.PP.GD	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. See NY.GDP.MKTP.PP.CD for the denominator's source.
EN.ATM.CO2E.PP.GD.KD	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions. See NY.GDP.MKTP.PP.KD for the denominator's source.
EN.ATM.CO2E.SF.KT	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.
EN.ATM.CO2E.SF.ZS	Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.

Table 4: Sources for World Bank Climate Change Indicators *(continued)*

indicator_id	source_org
EN.ATM.GHGO.KT.CE	World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/.
EN.ATM.GHGO.ZG	World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/.
EN.ATM.GHGT.KT.CE	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions.
EN.ATM.GHGT.ZG	World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/.
EN.ATM.HFCG.KT.CE	European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/
EN.ATM.METH.KT.CE	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions.
EN.ATM.METH.ZG	World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/.
EN.ATM.NOXE.KT.CE	Data for up to 1990 are sourced from Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States. Data from 1990 are CAIT data: Climate Watch. 2020. GHG Emissions. Washington, DC: World Resources Institute. Available at: https://www.climatewatchdata.org/ghg-emissions.
EN.ATM.NOXE.ZG	World Bank staff estimates from original source: European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/.
EN.ATM.PFCG.KT.CE	European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/
EN.ATM.SF6G.KT.CE	European Commission, Joint Research Centre (JRC)/Netherlands Environmental Assessment Agency (PBL). Emission Database for Global Atmospheric Research (EDGAR): http://edgar.jrc.ec.europa.eu/
EN.CLC.DRSK.XQ	(UNISDR, 2009-2011 Progress Reports, http://www.preventionweb.net/english/hyogo).
EN.CLC.GHGR.MT.CE	United Nations Framework Convention on Climate Change.
EN.CLC.MDAT.ZS	EM-DAT: The OFDA/CRED International Disaster Database: www.emdat.be, Université Catholique de Louvain, Brussels (Belgium), World Bank.
EN.POP.EL5M.RU.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
EN.POP.EL5M.UR.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
EN.POP.EL5M.ZS	Center for International Earth Science Information Network (CIESIN)/Columbia University. 2013. Urban-Rural Population and Land Area Estimates Version 2. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://sedac.ciesin.columbia.edu/data/set/lecz-urban-rural-population-land-area-estimates-v2.
EN.URB.MCTY.TL.ZS	United Nations, World Urbanization Prospects.
ER.H2O.FWTL.K3	Food and Agriculture Organization, AQUASTAT data.

Table 4: Sources for World Bank Climate Change Indicators *(continued)*

indicator_id	source_org
ER.H2O.FWTL.ZS	Food and Agriculture Organization, AQUASTAT data.
ER.LND.PTLD.ZS	World Database on Protected Areas (WDPA) where the compilation and management is carried out by United Nations Environment World Conservation Monitoring Centre (UNEP-WCMC) in collaboration with governments, non-governmental organizations, academia and industry. The data is available online through the Protected Planet website (https://www.protectedplanet.net/).
ER.MRN.PTMR.ZS	World Database on Protected Areas (WDPA) where the compilation and management is carried out by United Nations Environment World Conservation Monitoring Centre (UNEP-WCMC) in collaboration with governments, non-governmental organizations, academia and industry. The data is available online through the Protected Planet website (https://www.protectedplanet.net/).
ER.PTD.TOTL.ZS	World Database on Protected Areas (WDPA) where the compilation and management is carried out by United Nations Environment World Conservation Monitoring Centre (UNEP-WCMC) in collaboration with governments, non-governmental organizations, academia and industry. The data is available online through the Protected Planet website (https://www.protectedplanet.net/).
IC.BUS.EASE.XQ	World Bank, Doing Business project (http://www.doingbusiness.org/). NOTE: Doing Business has been discontinued as of 9/16/2021. For more information: https://bit.ly/3CLCbme
IQ.CPA.PUBS.XQ	World Bank Group, CPIA database (http://www.worldbank.org/ida).
IS.ROD.PAVE.ZS	International Road Federation, World Road Statistics and electronic files, except where noted.
NV.AGR.TOTL.ZS	World Bank national accounts data, and OECD National Accounts data files.
SE.ENR.PRSC.FM.ZS	UNESCO Institute for Statistics (http://uis.unesco.org/). Data as of September 2020.
SE.PRM.CMPT.ZS	UNESCO Institute for Statistics (http://uis.unesco.org/). Data as of September 2020.
SH.DYN.MORT	Estimates developed by the UN Inter-agency Group for Child Mortality Estimation (UNICEF, WHO, World Bank, UN DESA Population Division) at www.childmortality.org.
SH.MED.CMHW.P3	World Health Organization's Global Health Workforce Statistics, OECD, supplemented by country data.
SH.STA.MALN.ZS	UNICEF, WHO, World Bank: Joint child malnutrition estimates (JME). Aggregation is based on UNICEF, WHO, and the World Bank harmonized dataset (adjusted, comparable data) and methodology.
SI.POV.DDAY	World Bank, Development Research Group. Data are based on primary household survey data obtained from government statistical agencies and World Bank country departments. Data for high-income economies are from the Luxembourg Income Study database. For more information and methodology, please see PovcalNet (http://iresearch.worldbank.org/PovcalNet/index.htm).
SP.POP.GROW	Derived from total population. Population source: (1) United Nations Population Division. World Population Prospects: 2019 Revision, (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme.
SP.POP.TOTL	(1) United Nations Population Division. World Population Prospects: 2019 Revision. (2) Census reports and other statistical publications from national statistical offices, (3) Eurostat: Demographic Statistics, (4) United Nations Statistical Division. Population and Vital Statistics Reprot (various years), (5) U.S. Census Bureau: International Database, and (6) Secretariat of the Pacific Community: Statistics and Demography Programme.
SP.URB.GROW	World Bank staff estimates based on the United Nations Population Division's World Urbanization Prospects: 2018 Revision.
SP.URB.TOTL	World Bank staff estimates based on the United Nations Population Division's World Urbanization Prospects: 2018 Revision.
SP.URB.TOTL.IN.ZS	United Nations Population Division. World Urbanization Prospects: 2018 Revision.

```
# this is the function that actually extracts the data for all countries
wbcc <- wb_data(indicator=ccvars, country="countries_only", start_date=2001,
                end_date=2020)
```

```r
lastne <- function (x) {
  # Extracts the last value of the vector x that is not NA
  indx <- which(!is.na(x))
  return(ifelse(length(indx)==0, NA, x[max(indx)]))
}

# the wbcc dataframe has data for each variable and each country for the twenty
# year period from 2001 to 2020.
# This data is collapsed to a single row for each country that contains the
# latest data within this period for each variable.
wbcc_bc <- wbcc %>%
  group_by(iso3c) %>%
  summarise(country=last(country), across(all_of(ccvars), lastne))
```

```r
# write data frames to files
write.csv(wbcc_bc, "./WorldBank/wbcc_bc.csv", row.names=FALSE)
write.csv(wbindcc, "./WorldBank/wbindcc.csv", row.names=FALSE)
write.csv(wbindccsource, "./WorldBank/wbindccsource.csv", row.names=FALSE)
```

**** This is the end of Project ****