Using Curated Datasets

World Bank Open Data

Royal Geographical Society with IBG

Advancing geography and geographical learning

https://data.worldbank.org/

World Bank Open Data

The World Bank Open Data site houses an extremely comprehensive database relating to numerous development indicators and how the world's countries perform against them. It was started by the World Bank in an effort to make sure that decisions made by the organisation were based on sound and transparent data as well as monitor the impact any changes had on particular countries.

How do I access the data?

The above link is for the home page of World Bank Open Data. There are two routes through to data depending on what you wish to investigate. In the search box you have the ability to search by any keyword you choose, such as UK or GDP. Selecting **Country** will bring up a separate table listing all the countries and regions for which the World Bank holds data. Typing more than one country into the search box will allow you to compare two or more countries in the graph that subsequently gets produced.

Selecting **Indicator** brings up a separate table listing initially the most commonly used indicators for development. This can be navigated by scrolling through the list, or by using the quick find menu on the right hand side of the page. Alternatively one can select **All Indicators** to see the full list of data available.

Once a graph has been populated there are further options available to the reader. Firstly, scrolling to the bottom of the graph shows alternative countries that could be selected and miniature thumb nails of the general data trend those countries produce. Equally on the right hand side of the graph are alternative, but related, indicators for that country, again with thumbnails to highlight particular trends.

Hovering the cursor over the graph gives you precise data for any particular year and data can also be downloaded into an Excel spreadsheet to ease further use.

It is worth noted that often the same data is presented in different ways. For example, GDP might be shown as a total, as a measure per capita or altered to consider PPP.

Some example of data available in different categories:

GDP

GDP (current US$)

GDP (current local currency)

GDP growth (annual %)

GDP per capita

GDP adjusted to PPP

Population

Population total / total male / total female

Population % male / % female

Population in the largest city

Urban population

Population aged 0-14 (% of total)

Education

Primary school completion rate (% of relevant age group)

Literacy rate (% people)

Government expenditure per student (% GDP)

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Nuffield Foundation

How can I use this in my teaching?

The use of development indicators stretches far beyond the discrete topic of development and inequality. In this area, they can be used to define a country or a region and they enable students to place countries into categories and on spectrums according to how they compare against what are considered to be highly developed nations. When studied for different temporal periods they allow students to witness change and how countries have developed in different ways and at different rates over time, leading students to evaluate strategies designed to make countries more developed.

In other topic areas, development indicators can be a good starting point for students when they are exploring a new country and want to gain a quick understanding of what a country is like. For some topics, like the study of economic activity, this type of data can provide very detailed analysis of trade relationships and financial security while students looking at ecosystems and habitats can find data to show changes in land use and the harvesting of natural resources.

An important part of using and studying development indicators is to acknowledge their fallibility: no indicator gives a full picture of what a country is really like and reducing complex issues such as literacy or health into numbers can distort and simplify the issues. Students can use development indicator data to get a broad picture of where a country sits on a spectrum but should also critique the data itself and the methods used to collect it.

Curriculum Links

This curated dataset links to a number of parts of the National Curriculum and is relevant to GCSE and A Level Specifications.

Key Stage Three:	An understanding of human geography relating to international development.
	An understanding of geographical similarities, differences and links between places.
GCSE:	An understanding of the causes and consequences of uneven development at global level as the background for considering the changing context of population, economy and society and of technological and political development.
A Level:	An understanding of patterns of human development and life expectancy (for example, as understood through the human development index).
	An understanding of the nature of economic, political, social and environmental interdependence in the contemporary world.
	An understanding of how the demographic, socio-economic and cultural characteristics of places are shaped by shifting flows of people, resources, money and investment, and ideas at all scales from local to global.

The following specifications make particular reference to development indicators:

GCSE:			**A Level:**		
AQA	Cambridge IGCSE	Edexcel A	AQA	CIE	Edexcel
Edexcel B	Eduqas A	Eduqas B	Eduqas	OCR	
OCR A	OCR B				

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Nuffield Foundation

An example data walk-through

A student wishes to study find out how the values of two key indicators used to measure a country's progression in the Sustainable Development Goals have changed over time for a variety of different countries. The student believes the greatest change would be seen in Goal 4 (Quality Education) and far lower rates of change would be seen in Goal 3 (Good Health and Well Being). They base these ideas on the degree to which they believe change is possible, and that policy changes, in their eyes, have favoured education since 2015.

The student begins by going to the World Bank Data site and selecting **Browse by Indicator**. They then selected **Education** from the main list of indicators and looked at the many different indicators available to them. They then chose **Adjusted net enrolment rate, primary (% of primary school age children)**. This is the percentage of all primary-school aged children enrolled in primary education. The student then selected to **Download** the Excel file of all the data and began sorting it appropriately. They started by tidying the data: removing any data prior to 2015 (before the Sustainable Development Goals came into effect) and further removing any countries from the data set for whom no data had been collected.

The student then went back to the home page and repeated the process for a health indicator, choosing **Mortality rate under 5 (per 1,000 live births)** to study. This is the probability per 1,000 births that a new born baby will die before reaching the age of five. Again, the file was downloaded and tidied to produce a set of countries to study.

From the remaining comparable countries, the student chose four countries that interested them as they were all from the same continent: **Burkina Faso**, **The Gambia**, **Ivory Coast** and **Mozambique**. The data was separated for each (table below).

	Primary enrolment rate (%)		Mortality rate under 5 (per 1000 live births)	
	2015	2018	2015	2018
Burkina Faso	70.0	79.3	86.8	76.4
The Gambia	75.9	81.8	64.7	58.4
Ivory Coast	76.7	93.7	90.0	80.9
Mozambique	91.5	93.9	81.8	73.2

The student then calculated the percentage change for these countries in each of the two indicators between 2015 and 2018 (table below left).

	Percentage change primary enrolment rate (2015-2018)	Percentage change mortality rate under 5 (2015-2018)
Burkina Faso	+13.3%	-12.0%
The Gambia	+7.8%	-9.7%
Ivory Coast	+22.2%	-10.1%
Mozambique	+2.6%	-10.5%
MEAN	+11.5%	-10.6%

The results showed that there was actually very little difference in the amount of change witnessed between 2015 and 2018 in the two indicators, though with slightly higher levels of change seen in education than in health, as the student predicted. However, the student also recognised that the sample of data they used was extremely small and in fact had all countries been considered, they may have seen a very different result. The student also acknowledged that the two indicators chosen are just some of those used by the UN in monitoring the Sustainable Development Goals of Education and Health - a better study should have also looked at additional indicators to give a broader picture of change.

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Nuffield Foundation

Suggested delivery activities

The Importance of Units

While exploring the range of development indicators, students can look at how the units used to display the data can greatly influence how one views it and how easy it is to manipulate the reader by using units in a particular way. For example, students may wish to look at how the wealth of a country can be measured in vastly different ways and the key differences between GNI, GDP and GNP as well as the difference it makes to view wealth through per capita figures or by percentage change year on year.

This can create a further exercise which is for students to think about which is the fairest and most objective way of measuring that indicator of development and why.

Correlation or Causation?

In pairs, students download data for 15 different countries against two different development indicators. Student A's dataset should represent a social development indicator while Student B's should represent an economic development indicator. A scatter graph of the data should then be plotted.

The students should then work out if the graph shows a correlation between the data sets and the relative strength of that correlation. Where there is no correlation students might like to speculate whether there is any level of researcher bias in how they chose their countries. Then students can use their geographical knowledge and understanding to find reasons for any correlation (or for any lack of one).

Fact checking

Students should be encouraged to seek out 'facts' and statistical statements that are written about countries in the media and in reports about development. Good sources might include annual reviews for charities working in developing countries and newspaper reports about the plight of an underdeveloped nation.

Students are then tasked with having to fact check what the reports are saying by finding data evidence in support of the statement. They should evaluate whether, given the wider spread of data available, the 'fact' given is a true representation of what is actually happening in that country of region and if not, what a better statement would be to present information as accurately as possible.

Mann Whitney U test

The Mann Whitney U statistical test can be used to show how significant the degree of overlap is between the distribution of two sets of observed values. A large overlap would mean that there is little difference between the two sets of data in terms of their distribution while a small overlap would mean there is a significant difference between the distribution of one set of data compared to another.

A student can use the test to see if there is a significant difference between how a set of countries performs against a development indicator in one particular year to how the same set performs in a later time period. The test works best on samples of 20 or fewer.

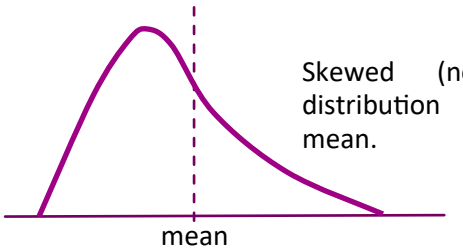
This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Nuffield Foundation

A Guide to Calculating the Mann Whitney U Test using Development Indicator Data

Year	GDP per capita (US$)	Rank
1990	7885	1
2018	9273	2
1990	9600	3
2018	12301	4
1990	13805	5
1990	14048	6
2018	14910	7
2018	15421	8
2018	16162	9
2018	17861	10
1990	19095	11
2018	19153	12
2018	19443	13
2018	20324	14
1990	20600	15
1990	20826	16
1990	21291	17
1990	21794	18
1990	22304	19
2018	23079	20
2018	23266	21
2018	23408	22
2018	26124	23
1990	26891	24
2018	28159	25
2018	30098	26
2018	30371	27
2018	34483	28
1990	34645	29
2018	41464	30
2018	42944	31
2018	47519	32
2018	47603	33
2018	50152	34
2018	51462	35
2018	53024	36
2018	54608	37
2018	61350	38
2018	78806	39
2018	116640	40

The Mann Whitney U test is a statistical measure of how significant any difference between two sets of data is. These sets of data should be values of the same attribute, though from different contexts, such as different time periods or from different locations. The test does not show where one set of data is significantly more or less evenly distributed than another, just whether they are significantly different from each other. It is designed to be used for data sets that may not exhibit a normal pattern of distribution around the mean.



Skewed (not normal) distribution around the mean.

A student wishes to see if there is a significant difference between the distribution of wealth in the countries making up the European Community and European Union in 1990 and in 2018 respectively. They believe there is likely to be a significant difference as more countries, who are considered to be economically weaker, have joined the union in that time period.

Worked example:

The student decides that their indicator for measuring the wealth of the different countries will be GDP per capita (current US$) and downloads the complete data set from the World Bank Data website. They research the names of the countries who were members of the European Community in 1990 and those that were members of the European Union in 2018. They then isolate this data and rank it, indicating at the same time which year it came from (table left).

The student then totals the rank for each of the years (r_1 and r_2)and counts the number of records in each year (n_1 and n_2):

1990: $\Sigma r_1 = 164$ $n_1 = 12$ 2018: $\Sigma r_2 = 656$ $n_2 = 28$

These values were then used in the Mann Whitney U equation for each year:

$$U_1 = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - \Sigma r_1 \quad \text{and} \quad U_2 = n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - \Sigma r_2$$

This gave the following values:

$U_1 = 250$ and $U_2 = 86$

The smaller of the two U values is taken to be the calculated value. One should then compare this with a critical value derived from a significance table. In this case the calculated value (86) is less than the critical value (101) so this tells the student that there is a significant difference between the distribution of wealth in the European Community of 1990 compared to the European Union of 2018.

This project was funded by the Nuffield Foundation, but the views expressed are those of the authors and not necessarily those of the Foundation.

Nuffield Foundation