# Data Selection Worksheet

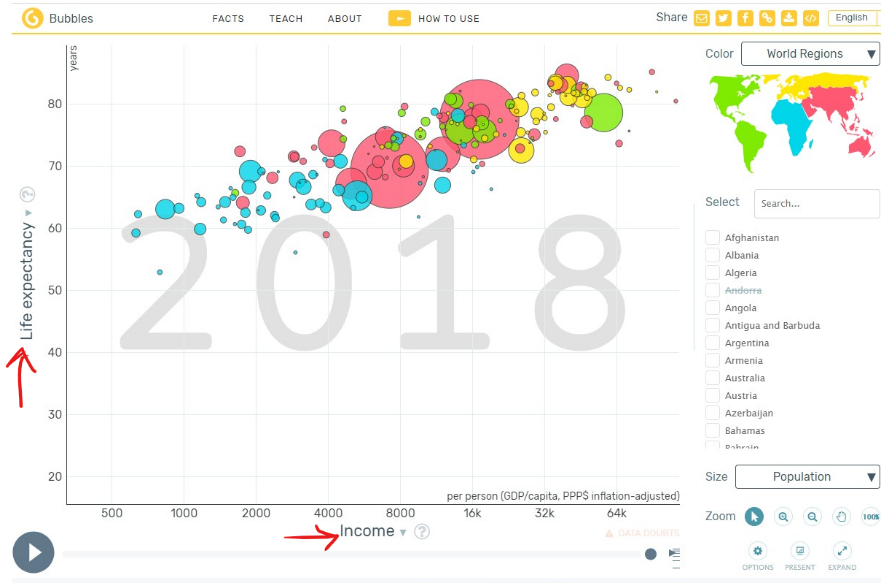
Statistical tests are used to interpret and present large numbers of numerical data. When observing a collection of data, you can identify features that allow you to classify the variables appropriately (see Data selection mini-lecture for details).

In order to select the appropriate datasets, let’s try to look for some interesting data sets and learn how to classify them. We elected a very interesting tool called Gapminder as our data source. This user-friendly resource will improve your understanding of datasets, graphs interpretation and introduce you to the wide amount of information that can be collected from their 4-parameter bubble graphs.

Gapminder is a foundation that identifies systematic misconceptions about important global trends and proportions and uses reliable data to develop easy to understand teaching materials to rid people of their misconceptions. Their mission is to fight devastating ignorance with a fact-based worldview everyone can understand.

## PART1- PLAYING WITH DATA IN GAPMINDER BUBBLE GRAPHS

1. Go to [GAPMINDER](https://www.gapminder.org/tools/) and Click on the X axis title (where it says “Income”).



Use the SEARCH BOX to select the variable “Diarrheal deaths in children” for the X axis

Now click on the Y axis title (where it says “Life expectancy”) and use the SEARCH BOX to select the variable “Malnutrition, weight for age”.



1. Classify these data sets according to the description above (you may need to click on the question mark by the axis titles for a detailed explanation of each variable):

* **Diarrhoeal deaths in children: (select one)**

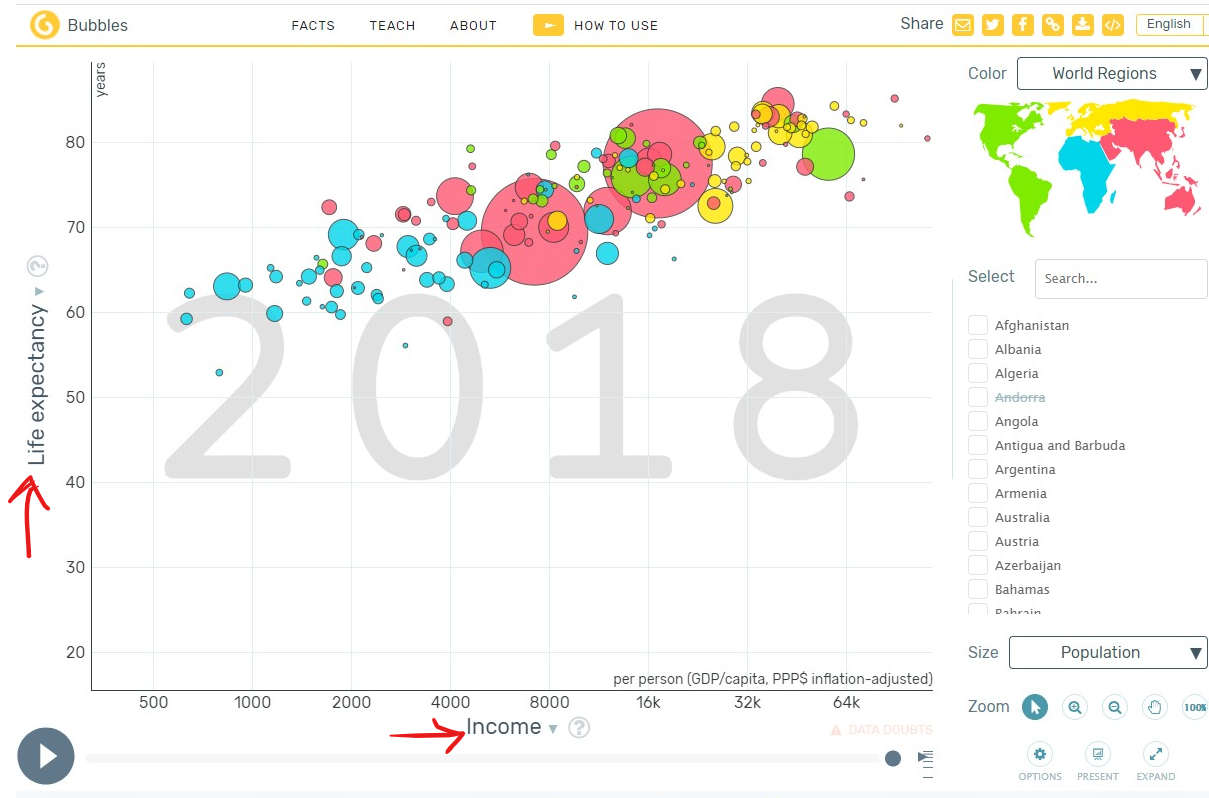
BINOMIAL   NOMINAL ORDINAL  INTERVAL orRATIO

* **Malnutrition (Prevalence of underweight children):** **(select one)**

BINOMIAL   NOMINAL ORDINAL  INTERVAL orRATIO 

1. Now in the right-side menu (see #1 in the picture) click on **South Korea** and **Pakistan** in the country list and identify the rate of malnutrition and diarrheal deaths for these 2 countries (You can now move the mouse on top of each bubble to find the individual data- see #2 in the picture).

Country list:





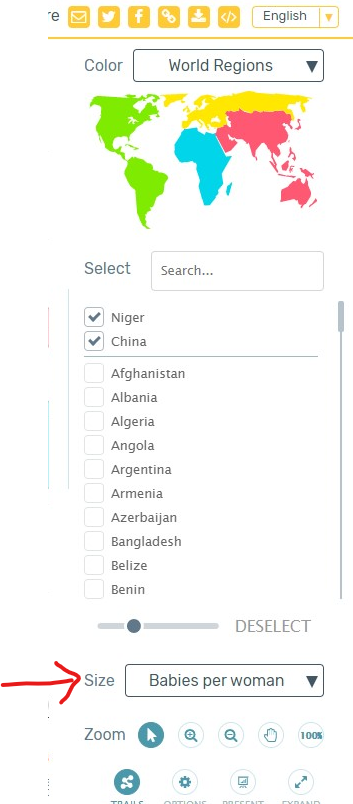
**2**

**1**



|  |  |  |
| --- | --- | --- |
|  | Diarrheal deaths in children | Malnutrition |
| South Korea |  |  |
| Pakistan |  |  |

1. On the right-side menu, click on size (see picture below) and select “Babies per woman”



1. Now your bubble size does not correspond to population size, but to babies per woman. In the country menu, select **China** and **Niger**. Complete the table below:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Babies per woman** | **Diarrheal deaths in children** | **Malnutrition** |
| China |  |  |  |
| Niger |  |  |  |

## PART 2 - FINDING DATASETS

You learned how to maneuver Gapminder and select variables, thus you can start your actual DATA SELECTION. Your aim here is to find variables classified as **interval or ratio**.

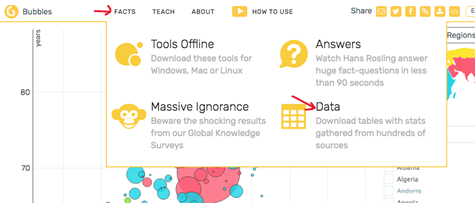
**Note that we are NOT interested in averages.**

Since we will calculate the median of each dataset in the next section (it is mathematically incorrect to average averages or calculate median of the averages). Here are examples of data that are **averages or means in the HEALTH folder and we are NOT interested in**:

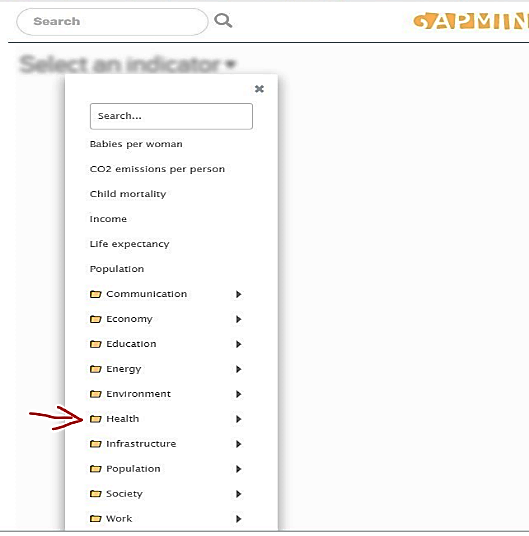
* Bad teeth per child
* Body mass index (BMI)
* Fat in blood
* Food supply
* Babies per woman
* Health spending
* Life expectancy

WHERE TO FIND GAPMINDER DATASETS:

1. Go to [GAPMINDER](https://www.gapminder.org/tools/) and click on “FACTS” then on “DATA” (see picture)



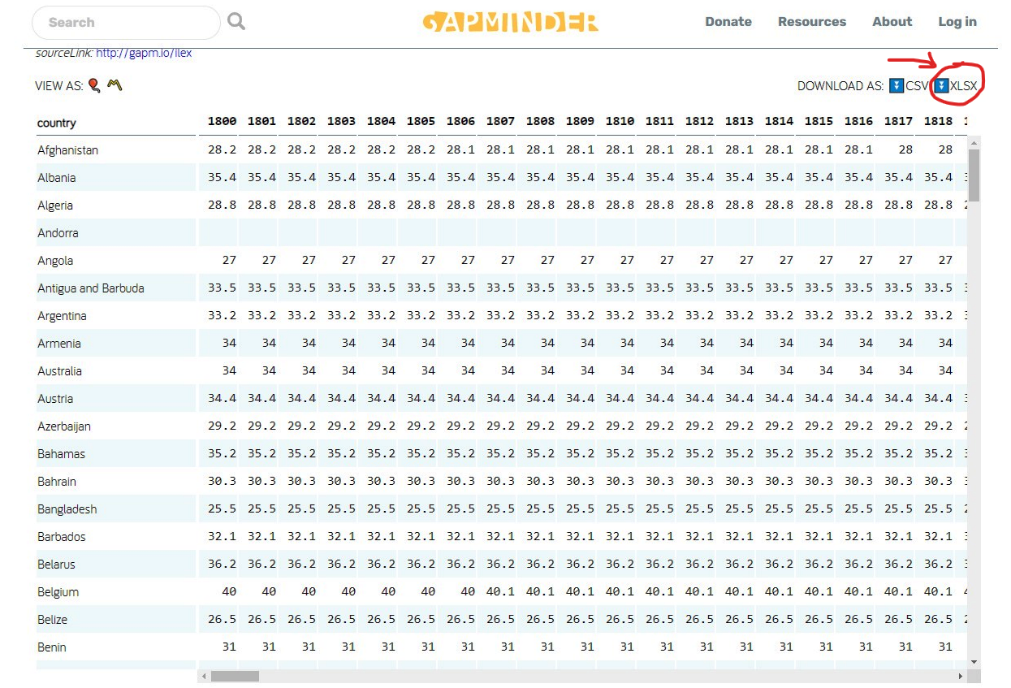
1. Scroll down to “Choose individual indicators”, click on the drop down menu for “Select an indicator” and select the folder “HEALTH”



1. Select 3 different Health variables that can be classified as INTERVAL or RATIO and list them below. (Check the list [Data we are NOT interested in](https://docs.google.com/document/d/1aOtcht8LKVyB8MrULzjzlagl-H93tpzWccjQRCG4KEk/edit#heading=h.2f6poiinj9jv))

|  |  |
| --- | --- |
| Variables | Interval OR Ratio |
|  |  |
|  |  |

4. Download and save your selected dataset (see red arrow in the picture below)



5. Proceed to the 5-point summary worksheet.



This material is based upon work supported by the National Science Foundation under Grant No.1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.