A step-change in quantitative social science skills

Funded by the Nuffield Foundation, ESRC and HEFCE

Q-Step

Manchester Metropolitan University

RISE: Introduction to Statistics

Independent Task: Introduction to Google Sheets

Google Sheets

In this guide, we will begin to look at real data using a free online software, Google Sheets. The purpose of this guide is to help you become familiar with variable types and raw data.

This guide will help you in completing other RISE intensives, such as Univariate and Bivariate Analysis.

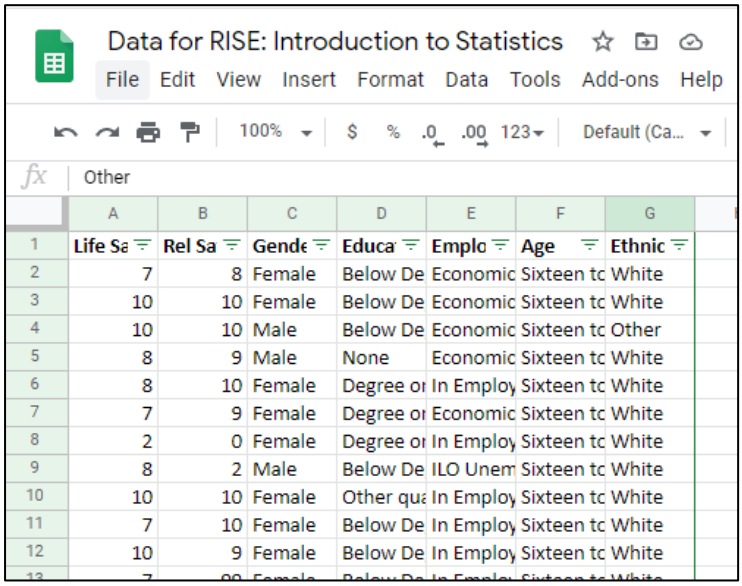
Step 1 – Accessing the data

The data we are going to analyse is a cutdown dataset from the 'Opinions and Lifestyle Survey: Well-Being Module, 2015.

Let's get the data open. Copy and paste the link to access the Google Sheets cutdown version of this data. Details about the study can be found at the end of this guide.

Link:
https://docs.google.com/spreadsheets/d/1wT5KeAV7Oobo89nliwpa_hPzRLLqwbvtcksdHnHbTz8/edit#gid=602299064

Your data should look like the screenshot below. Be sure to save your own copy of this data in your Google Drive so you can always come back to this original version in the future.



	A	B	C	D	E	F	G
1	Life Sa	Rel Sa	Gende	Educa	Emplo	Age	Ethnic
2	7	8	Female	Below De	Economic	Sixteen to	White
3	10	10	Female	Below De	Economic	Sixteen to	White
4	10	10	Male	Below De	Economic	Sixteen to	Other
5	8	9	Male	None	Economic	Sixteen to	White
6	8	10	Female	Degree or	In Employ	Sixteen to	White
7	7	9	Female	Degree or	Economic	Sixteen to	White
8	2	0	Female	Degree or	In Employ	Sixteen to	White
9	8	2	Male	Below De	ILO Unem	Sixteen to	White
10	10	10	Female	Other qua	In Employ	Sixteen to	White
11	7	10	Female	Below De	In Employ	Sixteen to	White
12	10	9	Female	Below De	In Employ	Sixteen to	White

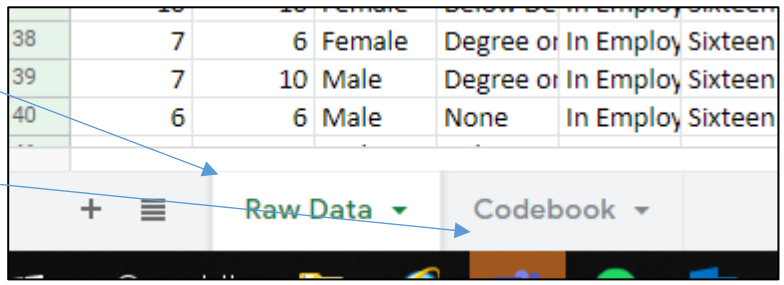
Step 2 – Understanding the Codebook

There are two sheets in this dataset:

The Raw Data

&

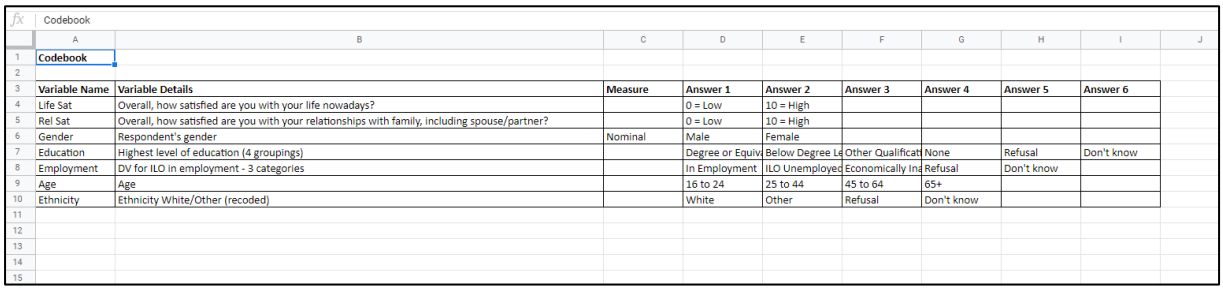
The Codebook



The Raw Data contains the actual responses from real individuals when answering a survey for the Opinions and Lifestyle Survey study. The codebook provides you an overview of the variables and values.

Let's take a look at the Codebook

The screenshot of the codebook below provides you with details of each variable.



Codebook

	A	B	C	D	E	F	G	H	I	J
1	**Codebook**									
2										
3	**Variable Name**	**Variable Details**	**Measure**	**Answer 1**	**Answer 2**	**Answer 3**	**Answer 4**	**Answer 5**	**Answer 6**	
4	Life Sat	Overall, how satisfied are you with your life nowadays?		0 = Low	10 = High					
5	Rel Sat	Overall, how satisfied are you with your relationships with family, including spouse/partner?		0 = Low	10 = High					
6	Gender	Respondent's gender	Nominal	Male	Female					
7	Education	Highest level of education (4 groupings)		Degree or Equiv	Below Degree Le	Other Qualificati	None	Refusal	Don't know	
8	Employment	DV for ILO in employment - 3 categories		In Employment	ILO Unemployed	Economically Ina	Refusal	Don't know		
9	Age	Age		16 to 24	25 to 44	45 to 64	65+			
10	Ethnicity	Ethnicity White/Other (recoded)		White	Other	Refusal	Don't know			
11										
12										
13										
14										
15										

Looking at the codebook, see if you can complete the 'Measure' column by filling in the cells with each variables measure. Gender has been done for you.

Gender is a nominal variable because…. It is a categorical variable (that contains words, not numbers) that can be put into a particular order.

Here are some tips to help you

Ordinal Data is categorical (contains words, not numbers) and can be put into a particular order.

Scale Data is numerical (contains numbers, not words) and therefore does not contain categories.

Once you have attempted to identify the types of variables, move on to the next page. We will come back to this later to see if you are right…