

# STEPS Survey

Data analysis using R software  
*User Guide*

Surveillance, Monitoring and Reporting

Department of Noncommunicable Diseases, Rehabilitation and Disabilities

World Health Organization, Geneva, Switzerland



World Health  
Organization



---

## Software used for STEPS data analysis

MS Excel



R software



RStudio



---

## R Software

- R is a powerful and free software environment for statistical computing and graphics.
- Download & Setup Instructions:
  - R Project website: <https://cran.r-project.org/>
  - Download the appropriate version for your system (Windows, macOS, or Linux).
  - Run the installer and follow the on-screen instructions.

---

## RStudio

- RStudio is an integrated development environment (IDE) for R software. It provides a user-friendly interface with additional features such as syntax highlighting, code completion, and tools for data visualization and reporting.
- Download & Setup Instructions:
  - RStudio website: <https://www.rstudio.com/>
  - Choose the free RStudio Desktop version for your operating system
  - Download and run the installer. Follow the installation steps.
  - On the first launch, RStudio will automatically detect your R installation. If prompted, point RStudio to your R installation path.

---

## Quick Tips:

- R is the core programming language, while RStudio makes it easier to work with R
- Make sure you install R first, as RStudio depends on it
- RStudio can be used to run R scripts, manage projects, and visualize data

---

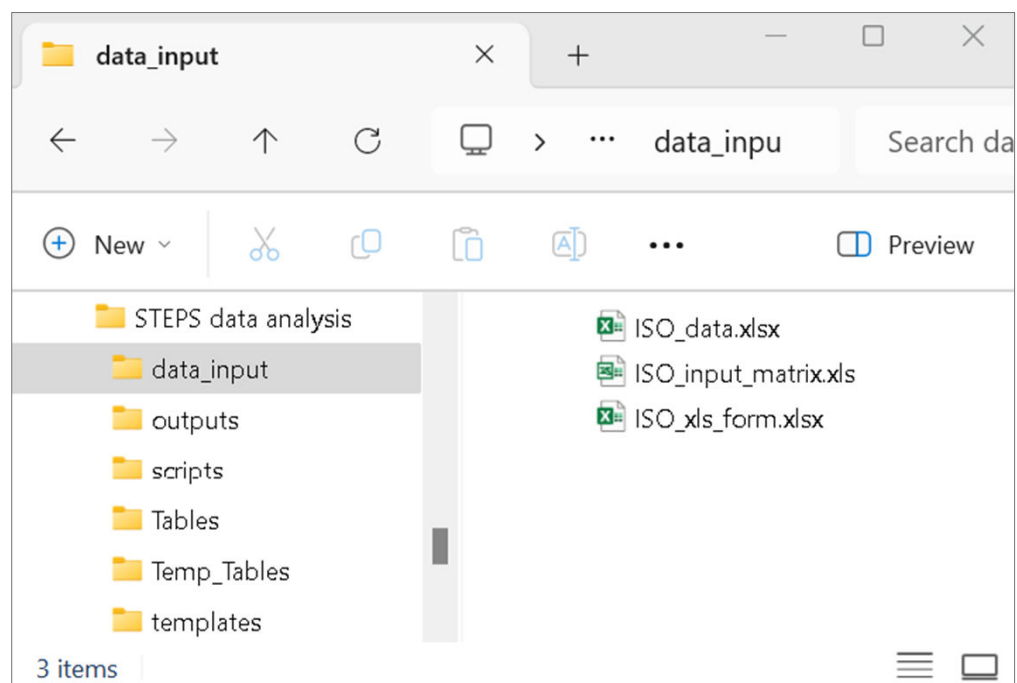
## List of Required R Packages:

openxlsx	stringr	dplyr
readxl	officer	haven
tidyverse	flextable	polyglotr
survey	readstata13	

All required packages shall be automatically installed when running the scripts for the first time.  
Alternatively the packages can be installed manually using the following command:

```
install.packages('package_name')
```

# File structure

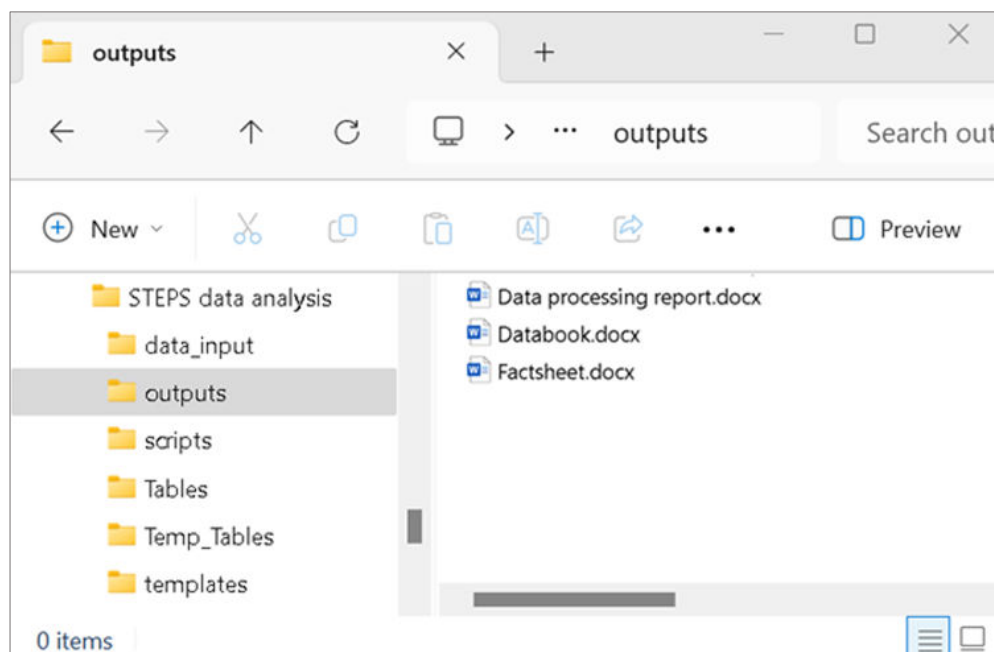


## “STEPS data analysis/data\_input” Folder

- ISO\_data.xls: STEPS dataset
- ISO\_input\_matrix.xls: STEPS indicators matrix
- ISO\_xls\_form.xls: ODK XLS form (steps1-2)

“ISO” will be replaced by three-letter country specific ISO code

# File structure



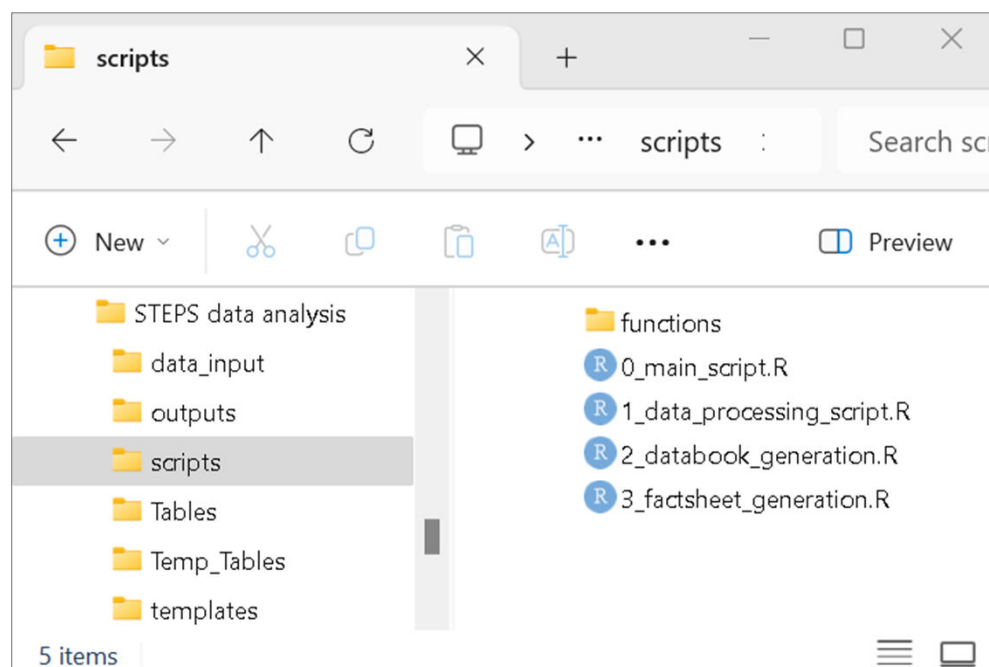
## **“STEPS data analysis/outputs” Folder**

- Data processing report.docx: Analysis report
- Databook.docx: STEPS Databook
- Factsheet.docx: STEPS Factsheet

Output files will be generated every time the analysis script is executed



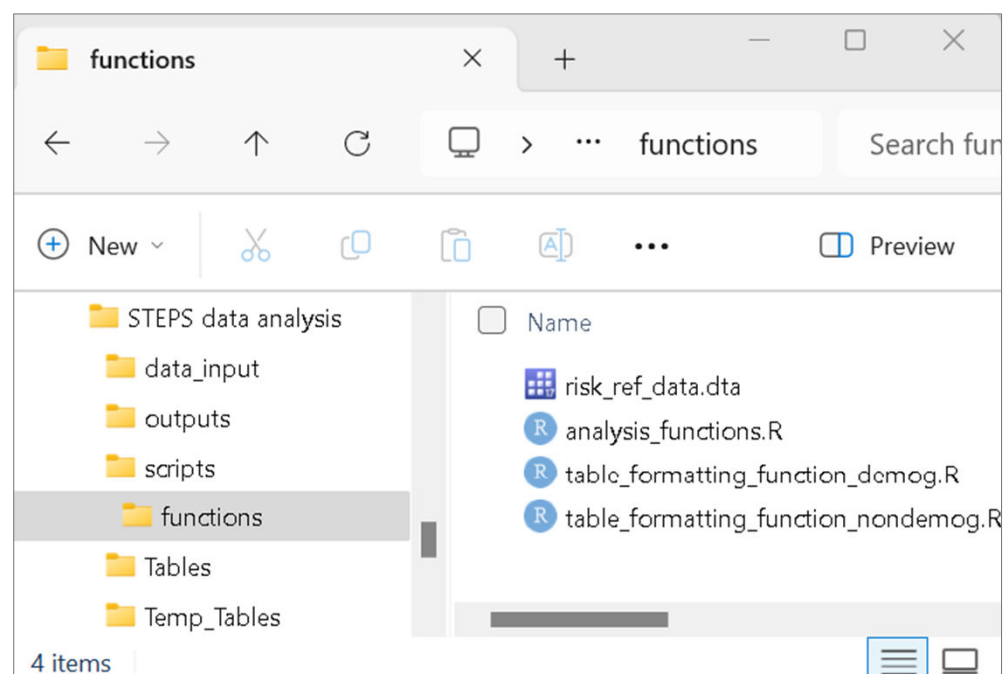
# File structure



## “STEPS data analysis/scripts” Folder

- Functions (folder): special functions
- 0\_main\_script.R: initiation R script
- 1\_data\_processing\_script.R: Data processing script
- 2\_databook\_generation.R: Databook script
- 3\_factsheet\_generation.R: Factsheet script

# File structure



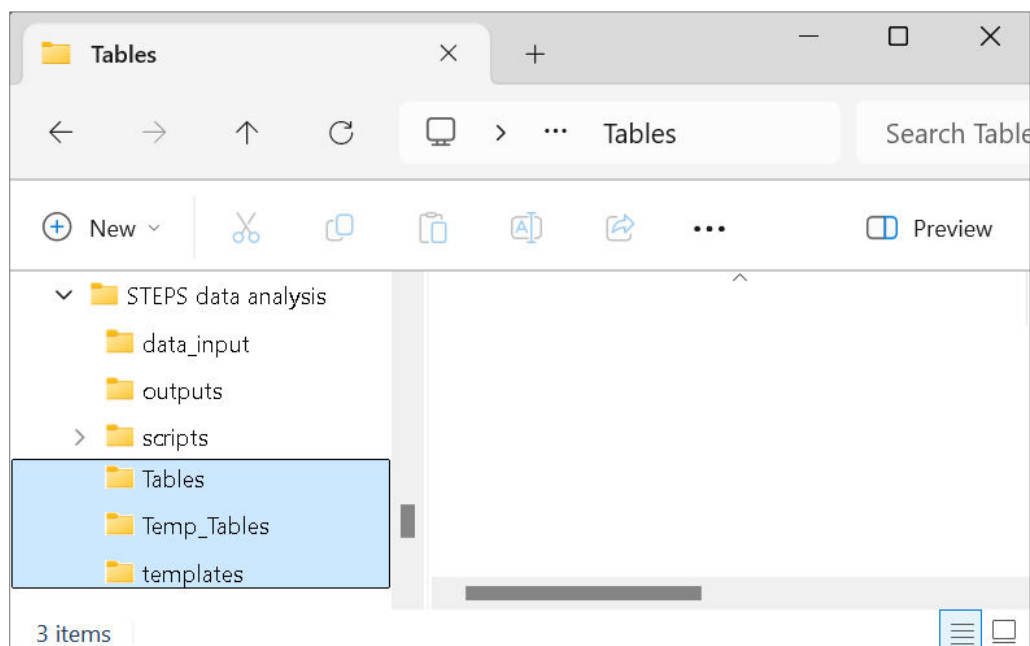
## “STEPS data analysis/scripts/unctions” Folder

- analysis\_functions.R
- table\_formatting\_function\_nondemog.R
- table\_formatting\_function\_demog.R
- risk\_ref\_data.dta CVD risk database

(Used for CVD risk calculation using WHO risk charts)

\*specific functions are used to format output tables for demographic and non-demographic indicators

# File structure



## **“Tables, Temp\_Tables, templates” Subfolders**

- include temporary and formatting files, and templates for Databook and Factsheet generation

# XLS form file - “xml\_file.xlsx”

File Home Insert Draw Page Layout Formulas Data Review View Developer Help Foxit PDF Acrobat										
Clipboard		Font		Alignment		Number		Styles		Cells
D103										
1	A	B	C	D	E	F	G	H	I	J
	type	name	choice_filter	constraint	required	relevant	calculation	appearance	label::English	constraint_message::Engl
84	<b>begin group</b>	tobacco_use							<b>Tobacco Use</b>	
85	note	note1							Now I am going to ask you some questions about tobacco	
86										
87	<b>begin group</b>	T1grp						field-list		
88	select_one yn	T1		yes					T1. Do you currently smoke any tobacco products, such	
89	note	smoked_tb_showcard							Smoked tobacco products	
90	<b>end group</b>									
91	select_one yn	T2		yes					T2. Do you currently smoke tobacco products daily?	
92	integer	T3		(.>=8 and .<=age);yes					T3. How old were yo Age must be at least 8 Enter '77' if	
93	<b>begin group</b>	T4grp								
94	integer	T4							T4. Do you remember how long ago it was? Leave blank	
95	select_one time2	T4type		(\${T4type} = 'weeks' yes					Please select whether answer is in years, months, or we	
96	<b>end group</b>									
97	integer	T5a		(.>=0 and .<=50) or .yes					T5a. On average, ho Number must be betwe Enter '77' if	
98	integer	T5aw		(.>=0 and .<=350) or yes					T5aw. On average, t Number must be betwe Enter '777' i	
99	integer	T5b		(.>=0 and .<=50) or .yes					T5b. On average, ho Number must be betwe Enter '77' if	
100	integer	T5bw		(.>=0 and .<=350) or yes					T5bw. On average, t Number must be betwe Enter '777' i	
101	integer	T5c		(.>=0 and .<=50) or .yes					T5c. On average, ho Number must be betwe Enter '77' if	

---

## Data Input - dataset

- STEPS raw dataset has to be cleaned and prepared for the analysis including:
  - All steps data files merged in a final dataset based on unique identifiers (ex: QR code)
  - All duplicates removed
  - Testing/training data removed if still there
  - Mandatory variables to be generated:
    - Id: unique identifier/QR code
    - age: age of survey participant
    - agerange: age groups
    - sex: sex ("Male"/"Female")
    - valid: valid=1 if age and sex are valid, else valid=2
    - stratum: stratification variable (e.g. regions, urban/rural). If not used, stratum=1
    - psu: sampling units before household selection
    - wstep1/ 2 /3: weights for Steps 1, 2 and 3
  - Save the final dataset: "data.xlsx"

# Dataset file example - “data.xlsx”

File

Home

Insert

Draw

Page Layout

Formulas

Data

Review

View

Developer

Help

Foxit PDF

Acrobat

Paste

Clipboard

Aptos Narrow

11

A<sup>^</sup>

A<sub>v</sub>

---

## Input Matrix file - “input\_matrix.xls”

- The file is used to assign indicators for data analysis and to compute derived variables
- STEPS **standard indicators** are predefined
- **Survey specific** or optional **indicators** can be added into the input matrix file for the analysis
- The xls file has 3 worksheets: “**indicators**”; “**derivedvars**”; and “**other**”
- Each worksheet has multiple columns

## Input Matrix file – indicators worksheet

include_in_analysis	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_var	pop_subset	primary_variable	weight_step	Table_title	subtitle1	subtitle2	column_strategy	Type	pop_level_num_denom	sheet_desc	ind_position
Yes	Demogr.	Demographic	dem1a:Ag	Summary info C1: Sex	C1	C1	All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr.	Demographic	dem1b:Ec	Mean number C4: In total, how many	C4	C4	All	C4	wstep1	Mean number mean			All	mean			
Yes	Demogr.	Demographic	dem1c:Hij	Highest level C5: What is the highest level of education	C5	C5	All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr.	Demographic	dem1d:Et	Summary response C6: What is your ethnic group?	C6	C6	All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr.	Demographic	dem1e:M	Marital status C7: What is your marital status?	C7	C7	All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr.	Demographic	dem1f:En	Proportion of C8: Which of the following best describes your employment status?	C8	C8	All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr.	Demographic	dem1g:Ur	Proportion of C8: Which of the following best describes your unpaid work status?	C8=4;C8=5;C8=6;C8=7	C8=4   C8=5	C8	C8	wstep1	Unpaid work status: Non-paid; Student; Homemaker	All			categorical			
Yes	Demogr.	Demographic	dem1h:Pe	Per capita annual consumption C9: How many cigarettes do you consume per year?	C9	C9	All	C9	wstep1	Mean annual consumption			total_overall	mean			
Yes	Demogr.	Demographic	dem1i:Est	Estimated household earnings C11: If you don't know, please estimate the household's monthly income from all sources.	C11	C11	All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker T1: Do you currently smoke?	T1	T1	All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w/ tobacco	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status T1: Do you currently smoke daily, occasionally, or not at all?	T1	T1	All	T1	wstep1	Smoking status: Current smoker; Former smoker; Never smoker			All	categorical		2. Percentage w/ tobacco	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Daily smoking T1: Do you currently smoke daily?	T1	T1	T1=1	T1;T2	wstep1	Current daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Initiation age T1: Do you currently smoke? If yes, at what age did you start smoking?	T1	T1	T1=1 & T2=1	T1;T2;t3age	wstep1	Mean age of initiation			All	mean		3. Average age of initiation	

## include\_in\_analysis

- Indicates whether to include the indicator in the final analysis
- Value = Yes if indicator shall be included. If No/missing – indicator will be excluded



## Input Matrix file – indicators worksheet

include_in_analysis	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_variable	pop_subset	primary_variable	weight_step	Table_title	subtitle1	subtitle2	column_strategy	Type	pop_level_num_denom	factsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info	C1: Sex	C1	All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number	C4: In total, how many	C4	All	C4	wstep1	Mean number of cigarettes smoked per day	All			mean			
Yes	Demogr	Demographic	dem1c:Hij	Highest level of education	C5: What is the highest level of education completed	C5	All	C5	wstep1	Highest level of education	All			categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary response	C6: What is your ethnic group	C6	All	C6	wstep1	Ethnic group of respondents	total_col			categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status	C7: What is your marital status	C7	All	C7	wstep1	Marital status	All			categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of respondents	C8: Which of the following best describes your employment status	C8	All	C8	wstep1	Employment status	All			categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of respondents	C8: Which of the following best describes your employment status	C8=4;C8=5;C8=6;C8=7	C8=4   C8=5	C8	wstep1	Unpaid worker/Non-paid;Student;Housewife	All			categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita annual expenditure	C9: How many cigarettes do you smoke per year	C10peryear	All	c10type;c9;C10	wstep1	Mean annual expenditure	total_overall			mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of estimated household earnings	C11: If you don't know, please estimate the household earnings	C11	All	C11	wstep1	Estimated household earnings	total_col			categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker	T1: Do you currently smoke	T1=1	All	T1	wstep1	Percentage Current smoker	All			categorical		1. Percentage who are current smokers	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status	T1: Do you currently smoke	T1=1 & T2=1;T1=1 & T2=2	All;All;All	T1;T2;T8	wstep1	Smoking status: Current smoker/Daily/Non-daily	All			categorical		2. Percentage who are daily smokers	
Yes	Tobacco	Tobacco Use	tob1c:Da	Daily smoker	T1: Do you currently smoke	T1=1 & T2=1	T1=1	T1;T2	wstep1	Current daily smoker	All			categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Initiation age	T1: Do you currently smoke	tob_initiation_age	T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age of initiation	All			mean		3. Average age of initiation	

### 'Section' and 'section\_translated'

- Include the STEPS Module/section name.
- The name will appear on top of each Databook and Factsheet section

## Input Matrix file – indicators worksheet

include_in_analysis	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_variable	pop_subset	primary_variable	weight_step	Table_title	subtitle1	subtitle2	column_strategy	Type	pop_level_num_denominator	worksheet_desc	ind_position
Yes	Demogr.	Demographic	dem1a:Ag	Summary info C1: Sex	C1	C1	All	C1	wstep1	Age group and sex of responder	All		categorical				
Yes	Demogr.	Demographic	dem1b:Ec	Mean number C4: In total, how many times did you smoke?	C4	C4	All	C4	wstep1	Mean number of cigarettes smoked per day		All	mean				
Yes	Demogr.	Demographic	dem1c:Hij	Highest level of education C5: What is the highest level of education completed?	C5	C5	All	C5	wstep1	Highest level of education		All	categorical				
Yes	Demogr.	Demographic	dem1d:Et	Summary response C6: What is your ethnicity?	C6	C6	All	C6	wstep1	Ethnic group of respondents		total_col	categorical				
Yes	Demogr.	Demographic	dem1e:M	Marital status C7: What is your marital status?	C7	C7	All	C7	wstep1	Marital status		All	categorical				
Yes	Demogr.	Demographic	dem1f:En	Proportion of employment C8: Which of the following best describes your current employment status?	C8	C8	All	C8	wstep1	Employment status		All	categorical				
Yes	Demogr.	Demographic	dem1g:Ur	Proportion of unpaid work C8: Which of the following best describes your current employment status?	C8=4;C8=5;C8=6;C8=7	C8=4   C8=5	All	C8	wstep1	Unpaid work status	Non-paid;Student;Housewife	All	categorical				
Yes	Demogr.	Demographic	dem1h:Pe	Per capita annual consumption C9: How many cigarettes do you consume per year?	C10peryear	C10peryear	All	C10type;c9;c10	wstep1	Mean annual consumption		total_overall	mean				
Yes	Demogr.	Demographic	dem1i:Est	Estimated household earnings C11: If you don't know, please estimate the household's monthly income from all sources.	C11	C11	All	C11	wstep1	Estimated household earnings		total_col	categorical				
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker T1: Do you currently smoke?	T1=T1=1	T1	All	T1	wstep1	Percentage Current smoker		All	categorical			1. Percentage w/ tobacco use	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status T1: Do you currently smoke?	T1=T1=1 & T2=T1=1 & T3=T1=1	T1=T1=1 & T2=T1=1 & T3=T1=1	All;All;All;All	T1;T2;T3	wstep1	Smoking status	Current smoker;Former smoker;Never smoker	All	categorical			2. Percentage w/ tobacco use	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Daily smoking T1: Do you currently smoke daily?	T1=T1=1 & T2=T1=1	T1=T1=1 & T2=T1=1	T1=1	T1;T2	wstep1	Current daily smokers		All	categorical				
Yes	Tobacco	Tobacco Use	tob1d:Init	Initiation age T1: Do you currently smoke? Initiation age	T1=T1=1 & T2=T1=1	T1=T1=1 & T2=T1=1	T1=1 & T2=T1=1	T1;T2;t3age	wstep1	Mean age of initiation	Mean age	All	mean			3. Average age of initiation	

### Indicator

- Contains the indicator code and name (e.g. “tob1a:Current smoking”)
- Codes must be unique and cannot be variable codes from the dataset. (e.g. “tob1a”)
- The indicator name will appear on the top left corner of each Databook indicator page (e.g. “Current smoking”)

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, how	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary response C6: What is your	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is your	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of the	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of the	C8	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	wstep1	Unpaid worker Non-paid;Student;H	All		All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita annual C9: How many	C10	per year	All	c10type;c9;c	wstep1	Mean annual mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of C11: If you don't	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker T1: Do you currently	T1	T1=1	All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w	1
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status T1: Do you currently	T1	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	wstep1	Smoking status Current smoker	Daily;Non		All	categorical		2. Percentage w	1
Yes	Tobacco	Tobacco Use	tob1c:Dai	Daily Percentage of T1: Do you currently	T1	T1=1 & T2=1	T1=1	T1;T2	wstep1	Current daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of initiation T1: Do you currently	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age of initiation	Mean age		All	mean		3. Average age	1

## Description

- Indicator description will appear on the top of each Databook indicator page

## Input Matrix file – indicators worksheet

include_in_analysis	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info	C1: Sex	C1	All	C1	wstep1	Age group and sex of responder			All	categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number	C4: In total, how many	C4	All	C4	wstep1	Mean number of cigarettes smoked per day			All	mean			
Yes	Demogr	Demographic	dem1c:Hij	Highest level of education	C5: What is the highest level of education completed	C5	All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary of ethnic group	C6: What is your ethnic group	C6	All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status	C7: What is your marital status	C7	All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of employment status	C8: Which of the following best describes your current employment status	C8	All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of unpaid work	C8: Which of the following best describes your current employment status	C8=4;C8=5;C8=6;C8=7	C8=4   C8=5	C8	wstep1	Unpaid work status	Non-paid;Student;Housework		All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita annual income	C9: How many cigarettes do you smoke per year	C10peryear	All	c10type;c9;c10	wstep1	Mean annual income			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of estimated household earnings	C11: If you don't know, please estimate the total household earnings	C11	All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker	T1: Do you currently smoke	T1=1	All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage who are current smokers	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status	T1: Do you currently smoke	T1=1 & T2=1;T1=1 & T2=2	All;All;All	T1;T2;T8	wstep1	Smoking status	Current smoker;Daily;Non-smoker		All	categorical		2. Percentage who are daily smokers	
Yes	Tobacco	Tobacco Use	tob1c:Da	Daily smoker	T1: Do you currently smoke	T1=1 & T2=1	T1=1	T1;T2	wstep1	Current daily smoker	Daily smokers		All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Initiation age	T1: Do you currently smoke	tob_initiation_age	T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age of initiation	Mean age of initiation		All	mean		3. Average age of initiation	

## Instrument\_questions

- All questions from the STEPS instrument used to compute a specific indicator
- Questions will be listed in separate lines in one cell

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	wstep1	Unpaid wor Non-paid;Student;H	All			categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita an C9: How many	C10	peryear	All	c10type;c9;c	wstep1	Mean annu; mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stat T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8		wstep1	Smoking sta Current s; Daily;Non	All			categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s; Mean age			All	mean		3. Average age ;1	

## logic\_condition\_var

- The column contains the logical conditions for the indicator nominator.
- For example: if the desired nominator is those who smoke tobacco, the cell will contain T1=1
- If the indicator has several reported categories, the **logic\_condition\_var** will have to be repeated the same number of times (e.g. "C1=1;C1=2;C1=3 etc.")

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder			All	categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H			All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita an C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s	Daily;Non		All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s Mean age			All	mean		3. Average age ;1	

## pop\_subset

- Logical conditions for the indicator denominator.
- If the indicator is percentage of current daily smokers among current smokers, then the cell will contain T1=1
- If the denominator is all population, the cell value will be “All”
- If the indicator has several reported categories, the pop\_subset will have to be repeated the same number of times (e.g. “All;All;All”)

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H	All		All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita ani C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s; Daily;Non			All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s; Mean age			All	mean		3. Average age ;1	

## Primary\_variables

- List of all variables used for each indicator
- If any of the listed variables are missing in the dataset the indicator will not be computed

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H	All		All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita ani C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s; Daily;Non			All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s; Mean age			All	mean		3. Average age ;1	

## weight\_step

- Indicates which weight variable to be used for the computation of specific indicators
- Example of weighting variables: wstep1, wstep2 or wstep3



# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder			All	categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H			All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita an C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu; mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stat T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s; Daily;Non			All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s; Mean age			All	mean		3. Average age ;1	

## Table\_title

- Indicates the title of the indicator table (tables) in the final Databook

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H	All			categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita ani C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s	Daily;Non		All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s Mean age			All	mean		3. Average age ;1	

## subtitle1 and subtitle2

- These two columns contain the final tables subtitles based on the indicators' values
- For instance, if we need the final table to contain separate columns for smoking status, the subtitle1 column will contain: Current smoker;Non-smokers
- If each of the above mentioned columns we want to be divided again, then the subtitle2 can be: Daily;Non-Daily;Former smoker;Never smoker

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All		All	categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H			All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita ani C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current sr Daily;Non			All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dai Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age sr Mean age			All	mean		3. Average age ;1	

## column\_strat

- This column indicates the type of stratification we want to report in the final table
- Accepted values: **All** – for all stratifications by agegroup and sex; **collevel1\_2** – for Men and Women only (no totals); **collevel1** - for Men only ; **collevel2** - for Women only; **total\_col** - only for both sexes; **total\_overall** – just total figure without stratification by age and sex.

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheets_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All		All	categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, how	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary response C6: What is your	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is your	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of the	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of the	C8	C8=4;C8=5;C8=6;C8=7	All	C8	wstep1	Unpaid worker Non-paid;Student;Household			All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita annual C9: How many	C10	per year	All	C10	wstep1	Mean annual	mean		total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of C11: If you don't	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker T1: Do you currently	T1	T1=1	All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status T1: Do you currently	T1	T1=1 & T2=1;T1=1 & T2=2;T1=1 & T2=3;T1=1 & T2=4;T1=1 & T2=5;T1=1 & T2=6;T1=1 & T2=7;T1=1 & T2=8	All	T1;T2;T8	wstep1	Smoking status Current smoker	Daily;Non-daily		All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Daily smokers T1: Do you currently	T1	T1=1 & T2=1	T1=1	T1;T2	wstep1	Current daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of initiation T1: Do you currently	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age of initiation	Mean age		All	mean		3. Average age ;1	

## Type

- Type of the indicator
- Indicators can be categorical or numerical
- Possible values:
  - categorical - if reported indicator has categories
  - mean - if mean is reported
  - median - if median is reported

# Input Matrix file – indicators worksheet

include_in_ana	Section	section_trans	Indicator	Description	Instrument_q	logic_condition_var	pop_subset	primary_var	weight_step	Table_title	subtitle1	subtitle2	column_strat	Type	pop_level_num_denom	fsctsheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Age	Summary info C1: Sex	C1		All	C1	wstep1	Age group and sex of responder	All			categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number C4: In total, ho	C4		All	C4	wstep1	Mean number	mean		All	mean			
Yes	Demogr	Demographic	dem1c:Hi	Highest level C5: What is the	C5		All	C5	wstep1	Highest level of education			All	categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary res C6: What is you	C6		All	C6	wstep1	Ethnic group of respondents			total_col	categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status C7: What is you	C7		All	C7	wstep1	Marital status			All	categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of C8: Which of th	C8		All	C8	wstep1	Employment status			All	categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of C8: Which of th	C8=4;C8=5;C8=6;C8=	C8=4   C8=5	C8	C8	wstep1	Unpaid wor Non-paid;Student;H			All	categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita ani C9: How many	C10peryear		All	c10type;c9;c	wstep1	Mean annu: mean			total_overall	mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of f C11: If you don	C11		All	C11	wstep1	Estimated household earnings			total_col	categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smok T1: Do you curi	T1=1		All	T1	wstep1	Percentage Current smoker			All	categorical		1. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking stati T1: Do you curi	T1=1 & T2=1;T1=1 &	All;All;All;All	T1;T2;T8	T1;T2;T8	wstep1	Smoking sta Current s	Daily;Non		All	categorical		2. Percentage w1	
Yes	Tobacco	Tobacco Use	tob1c:Dai	Percentage o T1: Do you curi	T1=1 & T2=1		T1=1	T1;T2	wstep1	Current dail Daily smokers			All	categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Mean age of i T1: Do you curi	tob_initiation_age		T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age s Mean age			All	mean		3. Average age ;1	

## pop\_level\_num\_denom

- Population level numeric denominator (target population size based on sample weight)
- Used in some specific indicators, e.g. unrecorded alcohol consumption

## Input Matrix file – indicators worksheet

include_in_analysis	Section	section_trans	Indicator	Description	Instrument_question	logic_condition_variable	pop_subset	primary_variable	weight_step	Table_title	subtitle1	subtitle2	column_strategy	Type	pop_level_num_denominator	sheet_desc	ind_position
Yes	Demogr	Demographic	dem1a:Ag	Summary info	C1: Sex	C1	All	C1	wstep1	Age group and sex of responder		All		categorical			
Yes	Demogr	Demographic	dem1b:Ec	Mean number	C4: In total, how many	C4	All	C4	wstep1	Mean number of cigarettes smoked per day		All		mean			
Yes	Demogr	Demographic	dem1c:Hij	Highest level of education	C5: What is the highest level of education completed	C5	All	C5	wstep1	Highest level of education		All		categorical			
Yes	Demogr	Demographic	dem1d:Et	Summary response	C6: What is your ethnic group	C6	All	C6	wstep1	Ethnic group of respondents		total_col		categorical			
Yes	Demogr	Demographic	dem1e:M	Marital status	C7: What is your marital status	C7	All	C7	wstep1	Marital status		All		categorical			
Yes	Demogr	Demographic	dem1f:En	Proportion of respondents	C8: Which of the following best describes your employment status	C8	All	C8	wstep1	Employment status		All		categorical			
Yes	Demogr	Demographic	dem1g:Ur	Proportion of respondents	C8: Which of the following best describes your unpaid work status	C8=4;C8=5;C8=6;C8=7	C8=4   C8=5	C8	wstep1	Unpaid work status	Non-paid;Student;Housework	All		categorical			
Yes	Demogr	Demographic	dem1h:Pe	Per capita annual expenditure	C9: How many cigarettes do you smoke per year	C10peryear	All	c10type;c9;C10	wstep1	Mean annual expenditure		total_overall		mean			
Yes	Demogr	Demographic	dem1i:Est	Summary of estimated household earnings	C11: If you don't know, please estimate	C11	All	C11	wstep1	Estimated household earnings		total_col		categorical			
Yes	Tobacco	Tobacco Use	tob1a:Cur	Current smoker	T1: Do you currently smoke	T1=1	All	T1	wstep1	Percentage Current smoker		All		categorical		1. Percentage who smoke	
Yes	Tobacco	Tobacco Use	tob1b:Sm	Smoking status	T1: Do you currently smoke	T1=1 & T2=1;T1=1 & T2=2	All;All;All	T1;T2;T8	wstep1	Smoking status	Current smoker;Daily;Non-smoker	All		categorical		2. Percentage who smoke daily	
Yes	Tobacco	Tobacco Use	tob1c:Da	Daily smoker	T1: Do you currently smoke	T1=1 & T2=1	T1=1	T1;T2	wstep1	Current daily smoker	Daily smokers	All		categorical			
Yes	Tobacco	Tobacco Use	tob1d:Init	Initiation age	T1: Do you currently smoke	tob_initiation_age	T1=1 & T2=1	T1;T2;t3;age	wstep1	Mean age of initiation	Mean age of initiation	All		mean		3. Average age of initiation	

### “fscsheet\_desc” and “ind\_position”

- The two columns specify the indicator name and indicator column position in the Factsheet.
- Ind\_position = 1 if output table has only one column. If multiple columns, the ind\_position will correspond to the column number to be reported.
- If the cells are left empty, the indicator will not be added to the Factsheet

# Input Matrix file – derived variables worksheet

	A	B	C	D
1	<b>section</b>	<b>derived_vars</b>	<b>primary_vars</b>	<b>logic_exp</b>
8	Tobacco Use	total_cig	t5a;t5b	data =data %>% rowwise %>% mutate(total_cig = ifelse(!is.null(t5a) & !is.null(t5b), sum(c(t5a,t5b), na.rm = T),NA))
	Tobacco Use	freq_T14a;freq_T14b;freq_T14c	t14a;t14aw;t14b;t14bw;t14c;t14cw	data = data %>% rowwise %>%mutate(freq_t14a = ifelse(!is.null(t14a), t14a, ifelse(is.null(t14aw), t14aw*7, NA)), freq_t14b = ifelse(!is.null(t14b), t14b, ifelse(is.null(t14bw), t14bw*7, NA)), freq_t14c = ifelse(!is.null(t14c), t14c, ifelse(is.null(t14cw), t14cw*7, NA)), freq_t14d = ifelse(!is.null(t14d), t14d,

**section:** STEPS Module/section name

**derived\_vars:** derived variable(s) name(s)

**primary\_vars:** variables from the STEPS dataset used to compute specific derived variables. If a variable is missing, the derived variable will not be computed.

**logic\_exp:** code snippet used to compute the derived variable (in R programming language)

# Input Matrix linked to the generated Databook table example

#	Input Matrix column	value
1	include_in_analysis	Yes
2	Section	Tobacco Use
3	section_translated	Tobacco Use
4	Indicator	tob1b:SmokingStatus
5	Descritpion	Smoking status among all population
6	Instrument_questions	T1: Do you currently smoke any tobacco products, such as cigarettes, cigars, or pipes? T2: Do you currently smoke tobacco products daily? T8: In the past, did you ever smoke any tobacco products?
7	logic_condition_var	T1=1 & T2=1;T1=1 & T2=2;T1=2 & T8=1;T8=2
8	pop_subset	All;All;All;All
9	primary_variables	T1;T2;T8
10	weight_step	wstep1
11	Table_title	Smoking status
12	subtitle1	Current smokers;Non-smokers
13	subtitle2	Daily;Non-Daily;Former smokers;Never smokers
14	column_strat	All
15	Type	categorical
16	pop_level_num_denom	
17	fsctsheet_desc	2. Percentage who currently smoke tobacco daily
18	ind_position	1

Tobacco Use									
<b>Smoking Status</b>									
Description: Smoking status of all respondents Instrument question: T1: Do you currently smoke any tobacco products, such as cigarettes, cigars, or pipes? T2: Do you currently smoke tobacco products daily? T8: In the past, did you ever smoke any tobacco products?									
<b>Smoking status</b>									
<b>Male</b>									
Current smoker									
Non-smokers									
Age Categories (Years)	n	% Daily	95% CI	% Non-Daily	95% CI	% Former smoker	95% CI	% Never smoker	95% CI
18-29	648	8.2	5.2 - 12.7	4.2	2.7 - 6.7	6.2	3.7 - 10.1	81.3	76.1 - 85.7
30-44	522	20.8	16.3 - 26.2	7.1	4.5 - 11.2	7.3	5.0 - 10.5	64.8	58.7 - 70.4
45-59	297	24.9	19.7 - 31.0	3.4	1.5 - 7.8	11.6	7.1 - 18.5	60.0	52.5 - 67.1
60-69	124	18.8	10.4 - 31.5	7.6	3.2 - 17.1	13.5	6.5 - 26.2	60.1	45.9 - 72.7
<b>Total</b>	<b>1591</b>	<b>15.3</b>	<b>12.7 - 18.5</b>	<b>5.1</b>	<b>3.8 - 6.7</b>	<b>7.9</b>	<b>6.1 - 10.1</b>	<b>71.7</b>	<b>67.8 - 75.3</b>
<b>Location</b>									
Rural	829	15.5	11.9 - 20.0	5.5	3.8 - 7.9	8.4	6.1 - 11.4	70.6	65.2 - 75.5
Urban	762	15.0	11.7 - 18.9	4.3	2.8 - 6.6	6.9	4.4 - 10.6	73.9	68.9 - 78.3



# STEPS data analysis - preparation

## R Studio:

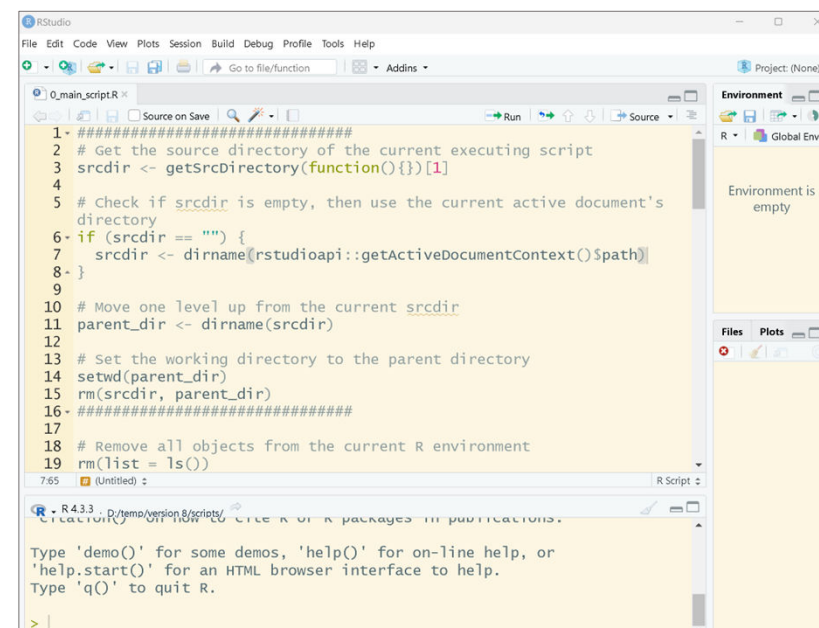
- Ensure you have R and R Studio installed on your computer.
- Open R Studio.

## Script:

- Open the “**scripts/0\_main\_script.R**” file in R Studio.
- Adapt the code with survey specific analysis inputs (specified on next page).

## Input files:

- Survey specific files shall be updated and saved under /data\_input/ folder
  - ISO\_input\_matrix.xls
  - ISO\_data.xlsx
  - ISO\_xls\_form.xlsx
- “ISO” in the file names to be replaced by three-letter country specific code



```
1 #####
2 # Get the source directory of the current executing script
3 srcdir <- getSrcDirectory(function(){})[1]
4
5 # Check if srcdir is empty, then use the current active document's
  directory
6 if (srcdir == "") {
7   srcdir <- dirname(rstudioapi::getActiveDocumentContext()$path)
8 }
9
10 # Move one level up from the current srcdir
11 parent_dir <- dirname(srcdir)
12
13 # Set the working directory to the parent directory
14 setwd(parent_dir)
15 rm(srcdir, parent_dir)
16 #####
17
18 # Remove all objects from the current R environment
19 rm(list = ls())
20
21 R 4.3.3 : D:/demo/version 8/scripts/
22 Creator: On how to cite R or R packages in publications.
23 Type 'demo()' for some demos, 'help()' for on-line help, or
24 'help.start()' for an HTML browser interface to help.
25 Type 'q()' to quit R.
26
27 >
```

# STEPS data analysis - survey parameters

## Set survey specific analysis inputs

country:	Country name.
survey_year:	Year when survey was conducted.
col_strat_variables:	Variables to be used as column stratifiers in the output tables.
row_strat_variables:	Variables to be used as row stratifiers in the output tables.
denom_limit:	The minimum number of respondents required within a stratification category to produce estimates.
language:	Report language to use. 1 = English. (For other languages translation shall be added to the indicators table).
country_ISO:	ISO code of the country (used for CVD risk calculation).

```
##### Specifying analysis inputs #####

### Define the country and year of the survey
country = 'Country Name' # Country for the survey
survey_year = 2024 # Year of the survey

### Define stratification variables for the analysis
col_strat_variable = c('sex') # Column stratifier (e.g., sex)
row_strat_variables = c('agerange', 'ruralurban') # Row stratifiers (e.g., age range, nationality)
row_strat_variable_titles = c("Age Categories (Years)", 'Location') # Titles for row stratifiers
vars_exempt_77_88 = c('') # Variables exempt from specific conditions (e.g., missing codes)

### Denominator limit: Minimum sample size required for estimating point estimates and confidence intervals
denom_limit = 10

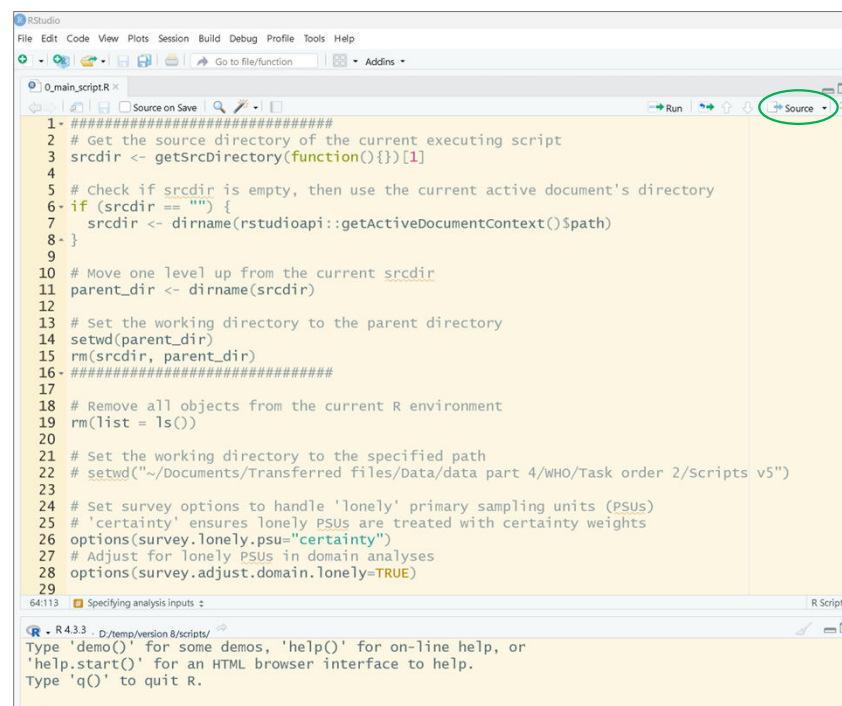
### Language setting for the analysis
language = c('ENGLISH', 'FRENCH', 'ARABIC', 'SPANISH', 'RUSSIAN', 'OTHER')[1] # Select language
if (language == 'OTHER'){language = 'SPECIFY'} else{} # Handle 'OTHER' language option
language = tolower(language) # Convert language to lowercase for consistency

### Inputs for cardiovascular disease (CVD) risk computation
country_ISO = 'ISO' # ISO code for the country

### Check if the specified ISO code exists in the reference dataset
ISO_existence = country_ISO %in% unique(read_dta('scripts/functions/risk_ref_data.dta')$ccode)
# If the ISO code does not exist, stop the script and display an error message
if(!ISO_existence)stop('The supplied ISO code does not exist in the reference dataset for cvd risk calculation')
```

# STEPS data analysis – running the analysis

- Ensure all the above preparation steps are completed.
- Execute the “**scripts/0\_main\_script.R**” file in R Studio by pressing on “**Source**” button.
- The script includes additional routines for generating Databook and Factsheet documents.
- Upon successful execution, survey-specific Databook and Factsheet files will be saved in the **/outputs/** directory.



The screenshot shows the RStudio interface with the file `0_main_script.R` open in the editor. The script contains R code for setting the source directory, checking for an empty `srcdir`, moving up one level to the parent directory, setting the working directory to the parent directory, removing the `srcdir` object, removing all objects from the current R environment, and setting the working directory to the specified path. The 'Source' button in the top right corner of the editor is circled in green.

```
1 #####
2 # Get the source directory of the current executing script
3 srcdir <- getSrcDirectory(function(){})[1]
4
5 # Check if srcdir is empty, then use the current active document's directory
6 if (srcdir == "") {
7   srcdir <- dirname(rstudioapi::getActiveDocumentContext()$path)
8 }
9
10 # Move one level up from the current srcdir
11 parent_dir <- dirname(srcdir)
12
13 # Set the working directory to the parent directory
14 setwd(parent_dir)
15 rm(srcdir, parent_dir)
16 #####
17
18 # Remove all objects from the current R environment
19 rm(list = ls())
20
21 # Set the working directory to the specified path
22 # setwd("~/Documents/Transferred files/Data/data part 4/WHO/Task order 2/Scripts v5")
23
24 # Set survey options to handle 'lonely' primary sampling units (PSUs)
25 # 'certainty' ensures lonely PSUs are treated with certainty weights
26 options(survey.lonely.psu="certainty")
27 # Adjust for lonely PSUs in domain analyses
28 options(survey.adjust.domain.lonely=TRUE)
29
30 #####
```