

Workshop: Assessing Data Quality and Disclosure Risk in Numeric Data

Exercise 2: Using the QAMyData tool

In this exercise you will use the **QAMyData** tool to assess the data quality of a dataset by configuring a number of tests.

Follow these steps to open, configure and run the **QAMYData** tool.

Steps:

- 1) Open QAMyData's help
 - a) On a *Windows machine* i) open Command Prompt (type in **command** in the search taskbar and click on **Command Prompt**) and ii) type in **qamd help**
 - b) On a *Mac* i) open the Terminal (in the search bar in the right corner type in Terminal and click on it) and ii) type in **qamd help**

- 2) Load in the teaching dataset and run QAMyData

Type in:

```
qamd run "location where you saved the data\name of the file.extension" --include-locators --output-format html --output your_results_file_name.html --config "location where you saved the configuration file\name of the file.extension"
```

```
Example: qamd run "I:\dcmagd\Open-source-tools\QAMyData\data\teaching data%set.sav" --include-locators --output-format html --output results_teaching_data.html --config "I:\dcmagd\Open-source-tools\QAMyData\config\config.toml"
```

--include-locators will include the locators for each "fail", so for which variable and if applicable, in which row, the issues can be found

--output-format html specifies the format of the output QAMyData will produce, in this case, html

`--output results_teaching_data.html` specifies the name of the output file, in this case `results_teaching_data`

`--config "I:\dcmagd\Open-source-tools\QAMyData\config\config.toml"` specifies the location of the configuration file you want to use and its name

3) Compare the results QAMyData found with the issues you detected in Exercise 1

- a) Did you find the same issues?
- b) Did QAMyData find any issues you already didn't?
- c) Did you find any issues QAMyData did not find?

4) Create a new configuration file to include a **RegEx** search for UK mobile numbers

- a) Open the configuration file with a text editor (notepad ++, notepad, etc.)
- b) Copy the contents of the default configuration file, create a new file titled "configregex.toml" and copy the contents

Try to find the RegEx pattern for mobile phone number and include the check in your new configuration file

OR

Copy the following in your new file and save it

```
[value_config.regex_patterns]
```

```
setting = ['^(+44\s?7\d{3})|(07\d{3})?\s?\d{3}\s?\d{3}$']
```

```
desc = "Values matching the UK mobile phone numbers regex pattern fail the check."
```

- c) Run QAMyData using the new "configregex.toml" configuration file

Trainer: Cristina Magder, UK Data Service

2019