

GUIDE TO DATA REPORTING for authors and editors

AUGUST 2019

Contents

Introduction	3
Dataset Submission Guidelines.....	3
Data organisation	3
Data analysis.....	4
Data assurance	4
Graphs/Figures/Charts Submission Guidelines	5
Metadata Submission Guidelines	7

Introduction

The SA Cities Network (SACN) is an established network of South African cities and partners that encourages the exchange of information, experience and best practices on urban development and city management. Our objectives include collecting, collating, analysing, assessing, disseminating and applying knowledge and experience for large city government in a South African context.

SACN produces reports and publications that have a strong evidence base informing their findings. This evidence base needs to be kept and archived appropriately. As far as possible, all core data and indicators are shared through SACN's open data portal SCODA (South African Cities Open Data Almanac) and reflected in the State of the Cities Report published every five years.

Authors are required to provide data and underlying methodologies and calculations with the final reports submitted to SACN. All data files should be given to SACN in an easy-to-read format, and which can be made openly available for city and public interrogation. Therefore, data must be provided as a Microsoft Excel file (.xls) and a .csv file.

This guide to data reporting is divided into three sections for authors to submit data: dataset submission, graphs/figures submission, and meta data submission.

Dataset Submission Guidelines

The goal of the data submission guidelines is for authors to produce descriptive datasets based on data used within research papers/reports that can be used, shared, integrated and checked for errors.

Data organisation

Setting up your spreadsheet

- The first sheet of the Excel spreadsheet must be the standard metadata sheet (see metadata submission guidelines). All fields must be completed, and additional fields added if needed.
- Consider creating a template for use during data collection, especially if there are different data sources. This will allow for consistency across data.
- Define each parameter (row/column). For example, each row/column should include: the data format, the units used (e.g. Rands, litres, kilometres, etc), and codes (values) for missing values (i.e. does a zero mean no data or a 0 numerical value?).
- The first row can be used for a heading.
- Open rows should be kept to a minimum.
- Data must be kept together in a single table.
- Other columns may be used for date, units of measure, data source etc., to provide context to the dataset.
- Numerical data must reflect appropriate precision (e.g. significant decimal places). For example (Rands use two decimal places for cents).

Formatting and naming

- Be consistent: one dataset or table should be on one sheet. Multiple data tables should be split across multiple sheets.
- Use the same format throughout the file – do not rearrange columns or rows within the file.
- Use plain text characters for parameter names, file names, and data, so that the data file can be read by the maximum number of software applications and uses.
- Assign descriptive file names: File names ideally describe the project, file contents, location, and date, and should be unique enough to stand alone as file descriptions. File names do not replace complete metadata records.

Note: Keep your raw data raw so that it can be checked later against any analysis performed. Supply the data at the level or resolution it was collected. The only exception is where proprietary conditions prevent sharing and such instances should be shared and discussed with the project manager.

Data analysis

- All calculations performed should be kept within the spreadsheet, i.e. do not paste data as values
- Data calculations and analysis should be done within accepted standards and methods. The method discussion should form part of project documentation.
- All calculations should be kept with the original dataset in a single table if possible. (for example, if calculating percentages add a row to the dataset and do not start a new table).

Data assurance

Basic quality control must be performed on your data during data collection, entry and analysis.

- Identify values that are estimated, double-check data that is entered by hand or copied from pdf documents.
- Check and identify missing values.
- Check the format of the data to be sure consistency across the dataset.
- In larger datasets, perform statistical and graphical summaries (e.g. max/min, average, range) to check for questionable values and to identify outliers. This will also allow checking the quality of data for possible data capture errors.
- Clearly indicate any data quality concerns within the dataset in the metadata or data documentation. The project manager must be made aware of any concerns around data quality as this can influence the project findings and analysis
- Check data using similar datasets to identify potential problems.

Graphs/Figures/Charts Submission Guidelines

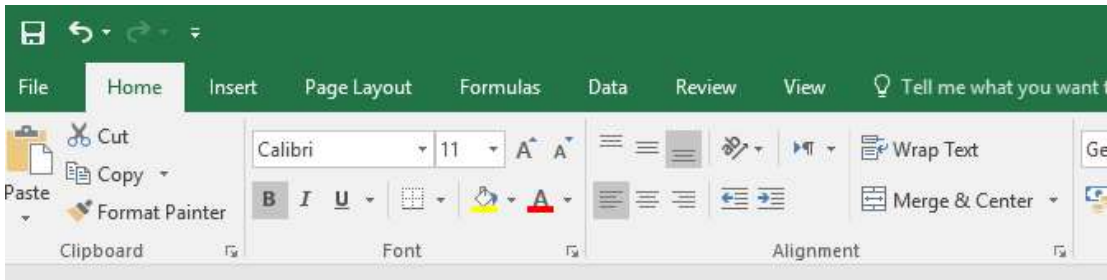
The goal for the graphs, figures, and charts submission guidelines is to ensure compatibility with supporting diagrams and SACN's data portal. Authors must provide data of graphs/figures/charts in an Excel spreadsheet (see next page for an example), and ensure the following:

- Use the figure number in the index of the report/document as the name of the worksheet within the Excel workbook.
- Include only one figure in each worksheet.
- Place graph near to the dataset on the worksheet.
- Ensure all figures have titles.
- Ensure the figure title is the same as in the report.
- Check consistency and logic of axis units and labels.
- If applicable, ensure that the key is legible, relates to the correct data point, and is located in the same place across graphs.

As far as possible, graphs should be drawn from the original datasets to eliminate room for error. If the calculations are done within a spreadsheet, the graph should be drawn from the data with the formulas within it, to enable data checking and eliminate room for errors. For example, do not copy-paste data as values and then draw a graph – draw the graph from the original data.

Authors are also required to provide any graphs/tables and photos (as applicable) in a resolution and format that is suitable for printing, i.e. .jpeg. The data used to develop the graphs and tables must also be provided in .csv file format, so that the information can be loaded on SACN's open data platform – SCODA.

SACN supports innovation when displaying data and city trends and encourages the use of a range of graph types, maps, graphical representations and photos and other means of displaying data.

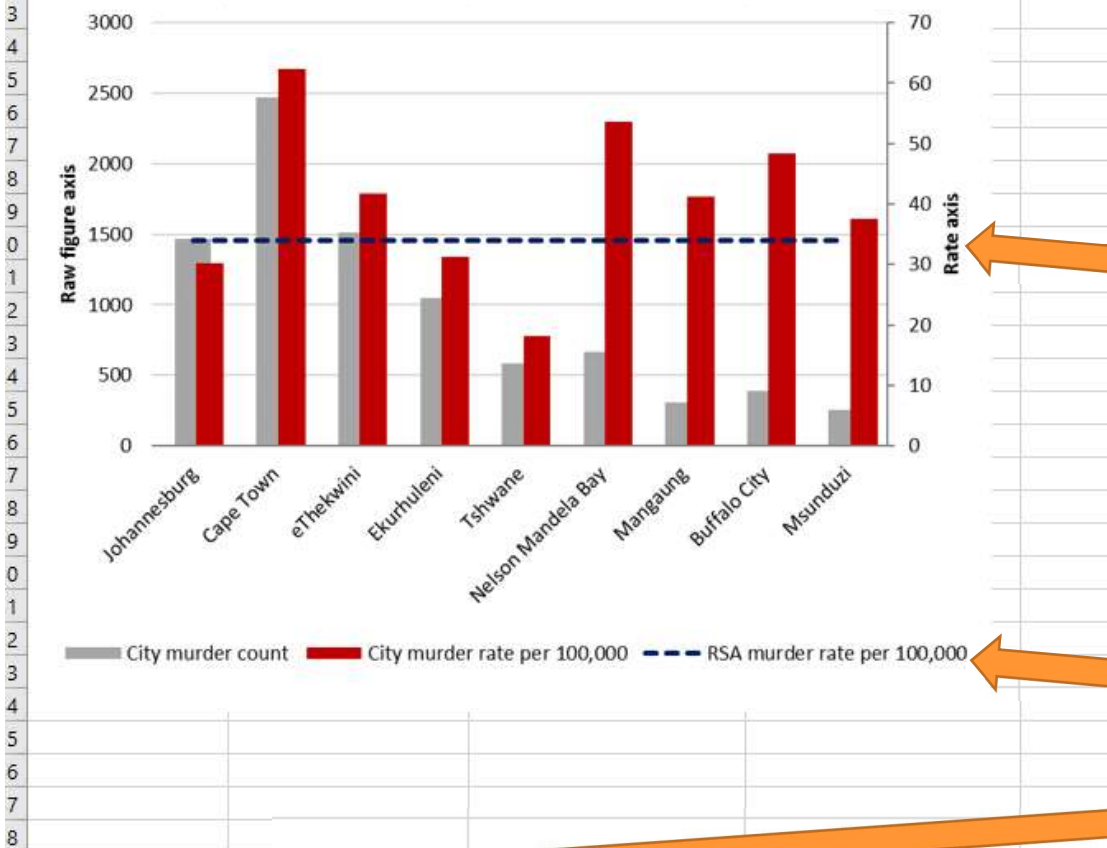


A1 : fx Figure 2: Raw murder figures compared to murder rates per 100 000

City	City murder count 15/16	City murder rate per 100,000	RSA murder rate per 100,000
JHB	1467	30	34
CPT	2469	62	34
ETH	1514	42	34
EKU	1046	31	34
TSH	584	18	34
NMB	668	54	34
MAN	311	41	34
BCM	388	48	34
MSU	252	38	34

Figure number and name

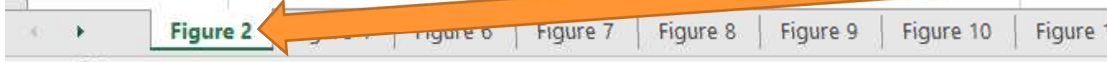
Data used to generate figure and figure located on same worksheet



Clearly labelled axes

Key / legend

Sheet name = figure number



Metadata Submission Guidelines

The goal for the metadata submission guideline is to ensure the provision of adequate metadata to be included into SACN's open data portal. Metadata is data that provides descriptive information (content, context, quality, structure, source and so on) about a dataset and allows others to search use the data for different applications.

The table below provides a standard metadata template to be used. The table is available in an excel format and should be the first tab of a spreadsheet. Additional information can be added if needed.

This spreadsheet contains the datasets used within:

[Report title]
[Report sub title]

Metadata	
Title:	
Project:	
Month and Year:	
Project manager:	
Project team:	
Project reference:	
Data sources:	
Data collection frequency:	
Contact details:	
Data access (e.g. website):	
Revisions/updates:	
Additional information	
Notes	