مـــركـــز الإحـــصـــاء
STATISTICS CENTRE

Statistical Report Writing and Data Presentation Guide

Methodology and Quality Guides - No. (17)

17

www.scad.ae

Table of Contents

Purpose of the Guide

This guide's intent is a practical tool to assist authors (statisticians) to produce high quality and consistent statistical publications, reports, and data presentations. The guide provides advice on the use of text, tables, charts, infographics, and maps to help bring statistics to life. It contains suggestions, guidelines, examples, and templates - to be applied during the production of statistical publications.

Introduction

Statisticians spend a lot of time collecting and preparing statistics. Therefore, it is crucial that the statistics are presented in an optimal way. To achieve this, the author (statistician) is encouraged to apply the international best practice dissemination techniques presented in this guide.

Good presentation of statistics is important to ensure the message is easily and quickly understood, and to reduce likelihood of statistics being misused.

An effective data release uses a combination of text, tables, and graphics to maximize its strength in conveying various types of information.

References

This guide was prepared by referencing international best practice dissemination techniques. In particular:

- The United Nations Economic Commission for Europe's (UNECE) Making Data Meaningful Part 2
- The UK Office of National Statistics' (ONS) Sty.
- Generic Statistical Business Process Model (GSBPM)

The production activities related to this guide fall under the GSBPM main process *7 Disseminate*, and the sub-process *7.2 Produce Dissemination Products*.

Part I: Statistical Report Writing

1. Audience and Types

Audience

The first and one of the most important steps, when writing a statistical report, is to consider the intended audience.

Ask, who is going to read the report – journalists, policy makers, statisticians, or the public?

Different audiences have different levels of understanding and different interests. Potential audiences can be categorised into three bands:

- **Basic**. The individual will be aware of, understand and appreciate the type of study or data source used (e.g. survey, census, or administrative data set), be able to find definitions of statistical terms and understand basic statistical measures and charts.
- **Intermediate**. The individual will understand and appreciate the limitations of the methods used, understand the more commonly used statistical concepts and terms (e.g. labour-force participation rate), understand variability and uncertainty.
- **Advanced**. The individual will be able to understand the more sophisticated statistical terminology, understand sampling errors and non-sampling errors, understand what constitutes a valid statistical study, make valid statistical inferences and critically assess claims made about the data.

Types

Report Type	Content	Frequency	Audience
Standard statistical publication, compendium	Structured layout, key points, tables and charts, explanatory notes. Repeatable and consistent.	Regular (e.g. monthly, quarterly, annual)	General public
Media release	Short summary of key points.	Upon important publication releases	Journalists

Special Report	Specific topic of interest. Includes observations and trends.	Ad hoc, upon request	Policy makers
Scientific paper	Detailed report which includes wider research and references.	Ad hoc, conferences or projects.	Researchers and academics

2. Style and Layout

Corporate Style

SCAD has specific guidelines on styles, fonts, and colours for its statistical publications. The style requirements can be found in the *SCAD Style Guidelines.*

Benefits of applying the SCAD style:

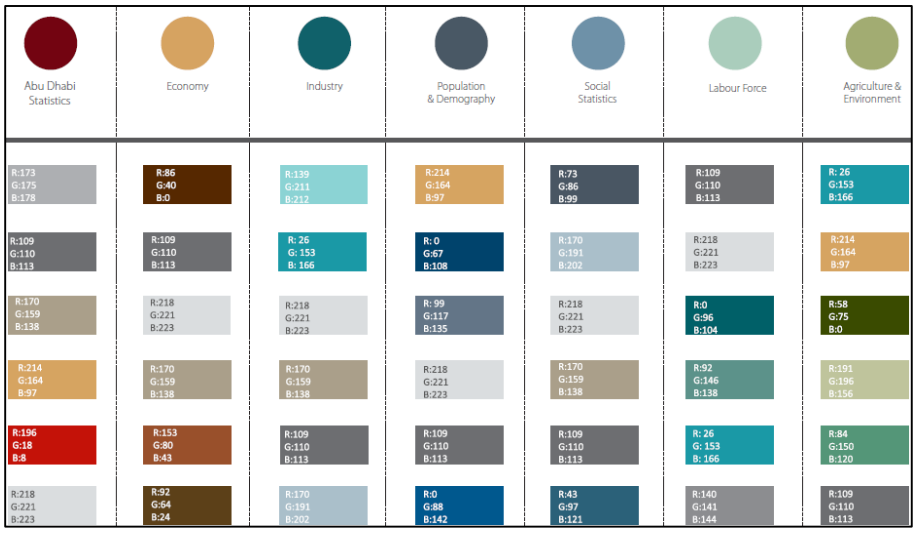
1. Consistency for readers
2. Efficiency for authors/approvers
3. Recognition / branding for organisation

Font Style

Heading 1: **Font: Tahoma (Arb), Arial (Eng)** **Size: 14pt (Arb, Eng)** **Location: Follows title.**	Body: Font: Tahoma (Arb), Arial (Eng) Size: 10pt (Arb, Eng) Location: Follows Heading1 or 2.	Source: Font: Tahoma (Arb), Arial (Eng) Size: 8pt (Arb, Eng) Location: Under Table / Chart.

Colours

The use of standardized colour scheme can be useful to maintain corporate image and consistency.



Layout / Sections

Statistical publications should follow a common structure and include specific components. At a minimum, a standard report should include:

- Introduction / Foreword (not mandatory – use if explanation of data required)
- Key Points (like an executive summary of the major points from release)
- Statistical Commentary (general text, charts and small tables about the statistics)
- Statistical Tables (listed separately – includes detailed and time-series tables)
- Explanatory Notes:
    - Glossary / Technical Notes
    - Release information
    - Disclaimer / Terms of Use

Typically, in SCAD, there are three types of statistical publications produced:

1. Regular publications (e.g. monthly CPI)
2. Annual subject-specific publications (e.g. Population Estimates)
3. Major releases (e.g. Statistical Year Book)

Components to include in publication types			
	1. Regualr (Monthly)	2. Annual	3. Major releases
Table of contents			
Foreword			
List of statistical tables			
List of figures			
Introduction			
Key points			
Statistical commentary			
Statistical tables			
Explanatory notes			
- Glossary			
- Technical notes			
- Data sources			
Release information			
Disclaimer / Terms of use			

Publication Titles

All publication titles should be succinct and include only the subject matter title and the reference period for the data relate. For example, *Foreign Trade Statistics January 2010.*

If a publication is ready for release in 2016, but it contains tables with data related to 2015, then it should be titled for 2015. For example, Education Statistics 2015

Authors should not use the term 'Report' unless it contains substantial analytical content. The Poverty Report is an example of where this might be used.

The following words should not be used in report titles: indicators, publication, series, and bulletin.

Exceptions to this for example the Abu Dhabi Statistical Yearbook and Abu Dhabi in Figures because they are considered special or one-off type compendia.

Examples:

- Transport and Communication Industry Statistics, 2009
- Consumer Price Index, 2010
- Tourism Statistics, 2010
- National Accounts, 2009
- Foreign Trade Statistics January, 2010
- Abu Dhabi Statistical Yearbook, 2011
- Building Materials Price Index, 2010
- Environmental Statistics, 2009

Templates

One method for controlling the structure, fonts, and colours of a publication is to use templates. Templates provide all the required corporate styles and structures in a blank document – saving the author time and effort and allowing the author to concentrate just on the content. SCAD has created English and Arabic MS Word templates for each of the six statistical themes and for the following types of reports:

- Key Points
- Monthly
- Quarterly and Annual

Additionally, MS Excel templates are available for tables and charts. All templates can be found in Smart SCAD at:
http://smartscad/taskmgmt/default.aspx?redirect=/smartpublisher/SPMS_Template_%20Library.aspx

3. Writing Rules

The following writing rules are editing text standards and conventions that apply for all SCAD publications. Consistent use of these will improve quality of publications.

SCAD English reports should adopt the English UK dictionary in MS Word, and not use English US.

Terminology

- Use “Citizens”, not “Nationals”
- Use “Al Dhafra Region”, not “Western Region” or “Al Gharbia”, and Al Ain Region, not “Eastern Region”
- Use the term “gender”, not “sex”
- Always use an initial capital letter when referring to an emirate. E.g. Abu Dhabi Emirate
- Gender and citizenship should be plural. E.g. males / citizens / male citizens / female citizens

Punctuation

- Use “compared with”, not “compared to” for contrasts between figures
- Write out “and” at all times. Ampersands (&) should never be used. E.g. hotels and restaurants, and transport, storage and communication sectors
- Use ‘per’ or the slash ‘/’ to indicate rates. E.g. Males per 100 Females; km/day

- Use “a” and “an” as they would be said. E.g. an 18% increase; a NATO paper; a UAE organisation; an IT solution
- Use “-ise”, not “-ize” as a word ending. E.g. organise; prioritise
- Bullets points following Main Points should start with a capital letter and finish with a full stop.
- Keep bullet points short (no more than 1 sentence). E.g. Main points:
    - There were 240,854 marriages in 2013, a decrease of 8.6% compared with 2012 and the first decline since 2009.
    - Civil ceremonies accounted for 72% of all marriages in 2013.

Numbers

- Write numbers from 1 – 9 as text. E.g. the region consists of six sectors
- Write numbers 10 and above as numerals. E.g. more than 67 persons were recorded
- Write out rankings first to ninth, then use numerals. E.g. Canada ranked first and USA 10th
- Don’t use abbreviations of “numbers”, such as “no” or “nos”
- Use commas above 999, in text, tables and charts. E.g. 6,530,722
- Use a 0 where there’s no digit before the decimal point in a number. E.g. 0.61
- Don’t start a sentence with a numeral. Rearrange the sentence accordingly. E.g. The number of persons who drive a car is 52.4 million. *Not* 52.4 million persons drive a car.
- Use the percent symbol (%) with no space between it and the number. E.g. 76.8%
- Use the same number of decimal places for a sequence of decimal numbers. E.g. 43.5, 18.0
- Try to use one decimal point as standard. However, there may be times when further decimals are necessary.
- Use percentage points for differences between percentages. E.g. a value of 28.4% falling by one percentage point becomes 27.4%. However, a fall of 1% would result in a value of 28.1%.
- Use the format “aged [age] years”. E.g. persons aged 9 years; employed persons aged 15 to 64 years. An exception is when written in parenthesis, then the word aged is not needed. E.g. the minority was older persons (65+ years), while the majority is working age population (15 to 64 years).
- Use the major currency unit before the amount. E.g. AED 15.5 million, USD 23,992

- Don't use "0.X million" for amounts less than 1 million, unless in a sequence. E.g. 3.5 million, 6.5 million, and 0.9 million
- Use negative sign for values less than 0. Do not use brackets. E.g. -15.3%, -2.84
- Conventions: 25.4 mm, 1.83 m, 100 km, 490 ha, 110 kg, 25 °C, −15%
- In tables the precise figures (e.g. 123,493) are to be used. While in the text, you may wish to apply rounding for ease of reading (e.g. In 2016, almost 123,500 persons were…)

Date Formats

- Range: 2009 to 2010, July to September 2014, July 2013 to September 2014
- Financial/school year: 2009–10 (not 2009–2010 or 2009/10)
- Specific years: 2008 and 2009 (not 2008–2009)
- Decades: 1980s, 2000s (no apostrophe)

4. Writing Well

Find the Story

On their own, statistics are just numbers. To mean anything to the person in the street, their value must be brought to life. A statistical story doesn't just recite data in words - it tells a story about the data. Readers tend to recall ideas more easily than they do data.

A statistical story conveys a message that tells readers what happened, who did it, when and where it happened, and hopefully, why and how it happened. A statistical story can:

- Provide general awareness/perspective/context;
- Inform debate on specific issues.

A statistical agency should want to tell a story because:

- It is the mandate of most agencies is to inform the general public about the population, society, economy and culture of the nation. This information will guide citizens in doing their jobs, raising their families, making purchases and in making many other decisions.
- An agency should want to demonstrate the relevance of its data to government and the public.

Authors of statistical reports need to balance the aspiration to be interesting and informative, with the need to be impartial and objective.

Tips

Understand what your users want:

- Keep in mind who the users are when writing your statistical report.
- What do they need to know and what is the best way to tell them?

Don't bury the story in statistics:

- Decide what the most important messages are and only include these.
- Include everything else in other, more detailed, tables and methodology reports.

Every section should be self-contained:

- Statistical publications are not news stories with a beginning, middle and end.
- Structure your report into sections which discuss different sub-topics.

Use lead headlines:

- A lead almost reads like a newspaper headline. Tells a complete story in itself.
- States something new and summarises the most important finding.
- Contains few numbers, if any at all.
- E.g. Abu Dhabi sees significant growth in consumer prices while Al Ain remains static.

Plain Language

Plain language is clear language, with no jargon, that is understood by all readers. It is language that avoids obscurity, inflated vocabulary and convoluted sentence construction. Writers of plain language let their audience concentrate on the message instead of being distracted by complicated language.

The main goal in writing is to put your message across clearly and concisely. Readers want an effortless, readable and clear writing style. Don't use formal or long words when easy or short ones will do.

Example:

> *Before*
>
> High-quality learning environments are a necessary precondition for facilitation and enhancement of the ongoing learning process.
>
> *After*
>
> Children need good schools if they are to learn properly.
>
> ---

Before

Your enquiry about the use of the entrance area at the library for the purpose of displaying posters and leaflets about Welfare and Supplementary Benefit rights, gives rise to the question of the provenance and authoritativeness of the material to be displayed. Posters and leaflets issued by the Central Office of Information, the Department of Health and Social Security and other authoritative bodies are usually displayed in libraries, but items of a disputatious or polemic kind, whilst not necessarily excluded, are considered individually.

After

Thank you for your letter asking for permission to put up posters in the library. Before we can give you an answer we will need to see a copy of the posters to make sure they won't offend anyone.

Paragraphs

- Tightly written with compact sentences that follow a logical order.
- No more than six sentences.
- Start with a lead sentence that introduces the information in paragraph.
- Be able to make complete sense on its own.
- Cover one subject.

Sentences

- Should be not be too long. Ideally, no more than 25 words.
- A sentence should not start with a number. E.g. 4.1% was the average annual growth rate for Abu Dhabi between 1995 and 2005. *Better to re-structure* E.g. Between 1995 and 2005, Abu Dhabi had an average annual growth rate of 4.1%.
- Be precise in the numbers used in text. Percentages generally to first decimal. E.g. The population estimates showed that the population of the Emirate of Abu Dhabi increased by 7.1% from 2,161,619 persons at the mid-year 2011 to 2,314,819 persons at the mid-year 2012.

5. Editing and Proofreading

Editing

The editing process ensures your written message matches what you were trying to say. It helps to condense and improve the efficiency of your writing. Questions the flow of thoughts, ensuring there's

good logic. Editing tells you if your content is too technical, or if it does not make sense, at least to the general public. It asks questions or presents an alternative perspective that you might not have considered.

Editing is re-reading a draft to see, for example, whether the paper is well-organized, the transitions between paragraphs are smooth, and that evidence for statements exist. You can edit on several levels:

- Content (e.g. relevant, accurate)
- Overall structure (e.g. meets the layout guidelines)
- Clarity (e.g. spell out acronyms)
- Style (e.g. use of plain English)
- Citations (e.g. identified sources)

In statistical publications, it is important that each number is validated, whether in a table, chart, infographic, or text. Validation usually involves someone other than the author (e.g. a team member) reviewing the data and checking against confirmed official statistics. Formulas for rates, totals, percentages, etc. also need to be checked for accuracy.

Proofreading

In SCAD's publication approval process, the publication should be reviewed or proofread by both the subject matter expert and the team manager. Proofreading is the final stage of the editing process, focusing on errors such as misspellings and mistakes in grammar and punctuation. Proofreading occurs only after editing revisions.

6. Approving

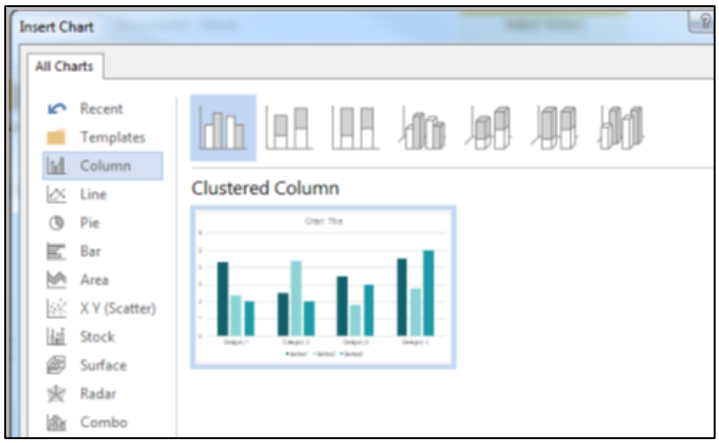
Most major statistical offices have a process for managing the approvals of their publications. In 2014, SCAD produced a set of detailed business process documentation called Statistical Standard Operating Procedures (SSOP). The specific process related to approving SCAD publications is '*Produce Standard Publication SOP'*.

The workflow of this approval process was automated into a system called Smart Publisher. A *New Publication Job* should be created in this Smart Publisher by an author (statistician) when a publication has reached final draft stage and all internal editing and proofreading checks have been completed.

Part II: Data Presentation

1. Templates

Similar to the publication templates, SCAD has developed MS Excel templates to use when constructing tables and charts. It is recommended to use these templates as they contain all the necessary approved SCAD styles, colours and fonts. Use of the templates also promotes consistency in SCAD's releases. SCAD's MS Excel templates are found in SmartSCAD at: http://smartscad/taskmgmt/default.aspx?redirect=/smartpublisher/SPMS_Template_%20Library.aspx



2. Tables

Why tables are important

- Good tables are an integral part of dissemination.
- Whether it is a news release, a statistical report, or a research paper - using tables effectively helps minimize the number of data values in your text.
- It also eliminates the need to discuss less significant variables that are not essential to the story line.

Types

Reference tables contain large amounts of cells and are for reference. Often included in an annex of a statistical report. **Summary tables** communicate a small number of messages clearly and simply. These are found in the body of the report and are often referenced in the commentary.

Guidelines

A table, with its components (title, source, etc.), should be self-contained. That is, if someone were to copy and paste elsewhere, all information required to understand the data would be available. When referring to tables in text, introduce each table/figure in text before table/figure appears. Refer to its number in text. E.g. 'Table 3.1 shows' or 'as shown in Table 3.1'.

Layout

- Consider the main message you wish to convey and design table appropriately.
- Lines and shading can be used to encourage users to read horizontally or vertically.
- Columns should be evenly spaced and not too far apart. The table should only be as wide as the data content requires.
- Generally, display 'totals' on last row of table. Only exception is when using a hierarchy within the table. Totals can be bolded.
- Use comma ',' separator for thousands. E.g. 9,876,543
- Align all numbers and column headings to the right (for both English and Arabic)
- Use consistent decimal points (generally, a single decimal)
- Do not leave any data cell empty. If nil, use '- ', or 'n.a.' for not available.
- In general, try to make all data columns the same width.
- Always include a source. Generally, use a smaller font size for source (e.g. 9 point). E.g. Source: Statistics Centre – Abu Dhabi

Columns and Rows

- Generally, users find it easier to scan down columns, rather than across rows. For example, if the key message is change over time periods, it might be best to place years in the rows (to scan down the column).

Good

Year	Population
1975	211,812
1985	566,036
1995	942,463
2005	1,399,484

Poor

	1975	1985	1995	2005
Population	211,812	566,036	942,463	1,399,484

- When designing two-way tables, consider what is the key message.

  Where the time series is important…

Year	Males	Females	Total
1975	155,058	56,754	211,812
1985	380,253	185,783	566,036
1995	650,744	291,719	942,463
2005	926,814	472,670	1,399,484

  Where the data items are important. Also look for natural or size order in data series (e.g. ascending order).

	1975	1985	1995	2005
Females	56,754	185,783	291,719	472,670
Males	155,058	380,253	650,744	926,814
Total	211,812	566,036	942,463	1,399,484

Titles

- Use the word “Table” at the beginning of the title and number the table.
- Title describes what the table is showing. Structured as follows:
- Subject matter that the statistics relate to, e.g. Employed persons
- How the statistics are classified, e.g. Age by Gender
- Then by region/place, e.g. Abu Dhabi Emirate
- Then by period, e.g. 30 June 2010.
- E.g. Table 3.1: Employed Persons by Age by Gender, Abu Dhabi Emirate, 30 June 2010

Rounding

Tables should contain the precise figure (e.g. 183,493). However, can change the order of magnitude if it assists with interpretation.

Year	Population
1975	211,812
1985	566,036
1995	942,463
2005	1,399,484

Year	Population (‘000)
1975	212
1985	566
1995	942
2005	1,399

Year	Population (million)
1975	0.2
1985	0.6
1995	0.9
2005	1.4

Columns - Headings and headnotes

- Align first column heading left (i.e. the one above the row headings and non-data columns). Align data columns headings middle and right (in both English and Arabic).
- For headnotes spanning two or more columns, use em dash 2 spaces headnote 2 spaces em dash (e.g. — number —). Using numeric keypad type ALT 0151. Centre the spanned headnote across the two or more columns.
- Use % rather than per cent; number (small n) rather than no.
- Use ’000, million, billion

Poor

3.1.3 Average Annual Population Growth Rates by Citizenship and Gender, Abu Dhabi Emirate, Mid-years 2005-2015

Citizenship and Gender	Mid Year 2005	Mid Year 2015	Average annual growth rates
Grand Total	**1,374,169**	**2,784,490**	**7.3**
Males	911,864	1,831,741	7.2
Females	462,305	952,749	7.5
Citizens	**344,350**	**536,741**	**4.5**
Males	173,861	275,950	4.7
Females	170,489	260,791	4.3
Non-Citizens	**1,029,819**	**2,247,749**	**8.1**
Males	738,003	1,555,791	7.7
Females	291,816	691,958	9.0

Good

3.1.3 Average Annual Population Growth Rate by Citizenship and Gender, Abu Dhabi Emirate, Mid Years 2005 and 2015

Citizenship and Gender	Mid Year 2005	Mid Year 2015	Average annual growth rate
	— persons —		*— % —*
Grand Total	**1,374,169**	**2,784,490**	**7.3**
Males	911,864	1,831,741	7.2
Females	462,305	952,749	7.5
Citizens	**344,350**	**536,741**	**4.5**
Males	173,861	275,950	4.7
Females	170,489	260,791	4.3
Non-Citizens	**1,029,819**	**2,247,749**	**8.1**
Males	738,003	1,555,791	7.7
Females	291,816	691,958	9.0

Rows – Headings

- Total in bold, subtotals in italics.
- If several levels of row headings (e.g. hierarchy), bold highest level, and indent subsequent levels.
- Align row headings left (in English) and right (in Arabic).

Footnotes

- Label footnotes (a), (b), etc. Make sure (a) appears in the table before (b)…
- Usually start each footnote on a new line at bottom of table before source.
- Any table showing abbreviations and acronyms should be included in Explanatory Notes.

- Some common statistical footnote annotations:

r = revised	**n.e.c.** = not elsewhere classified
p = preliminary	**n.a.** = not available
– nil or rounded to zero	**n.y.a.** =not yet available
. . not applicable	**n.p.** =not published

3. Charts

Introduction

- Statistics can often be better understood when presented in a chart than in a table.
- A chart is a visual representation of statistical data, in which the data are represented by symbols such as bars or lines.
- It is a very effective visual tool, as it displays data quickly and easily, facilitates comparison and can reveal trends and relationships within the data.

You should consider using charts if you want to show:

- Comparison: How much? Which item is bigger or smaller?
- Changes over time: How does a variable evolve?
- Frequency distribution: How are the items distributed? What are the differences?
- Correlation: Are two variables linked?
- Relative share of a whole: How does one item compare to the total?

Cognitive Load

- Cognitive load is the effort a reader requires to understand what you are trying to communicate in your chart.
- A chart with a high cognitive load will be hard to understand and to remember.
- A chart with a low cognitive load will be understood at a glance. Its message will be obvious.
- When you design a chart, you can control its cognitive load.

A good chart:

- Grabs the reader's attention
- Presents the information simply, clearly, and accurately
- Does not mislead
- Displays the data in a concentrated way (e.g. one line chart instead of many pie charts)
- Facilitates data comparison and highlights trends and differences
- Illustrates messages, themes, or storylines in the accompanying text

Strength of a chart:

- Communicating broad trends and comparisons
- Information can be taken in quickly
- Convey strong and lasting images
- Attractive

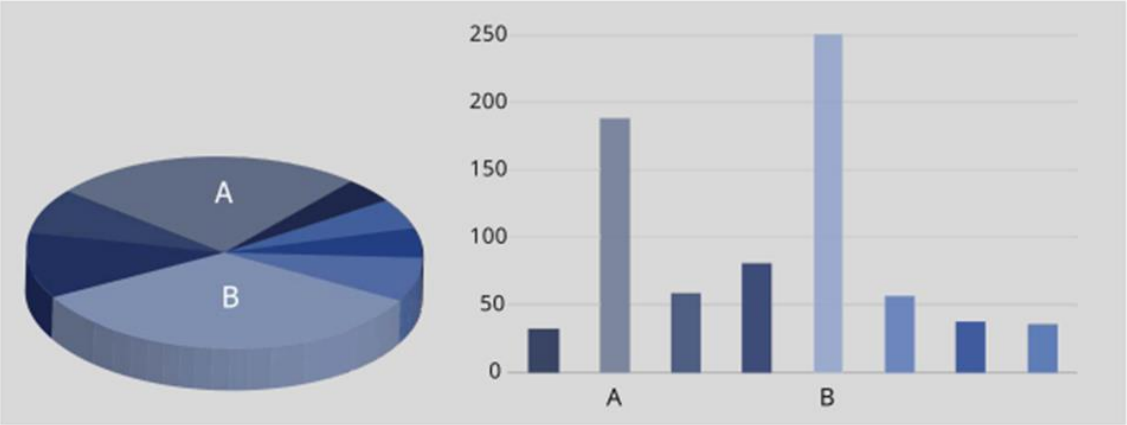
Weakness of a chart:

- Time consuming to create effective chart
- Do not convey specific amounts well
- A poorly constructed chart can mislead

Guidelines

- The same guidelines and rules for tables apply to charts (e.g. charts should have a title, a source, clearly labelled, same fonts, etc.).
- As with text and tables, for charts, you need to think about:
    - who is the audience?
    - what is the message?
- Recommend using SCAD's MS Excel templates for consistency in style, layout, colours, and fonts.
- There are numerous types of charts, with the most common ones being: line charts, bar or column charts, and pie or donut charts.

- Use these simple charts unless you have a good reason to use another type, such as a scatter chart.
- The detailed data should be available in an accompanying table (either shown near the chart, or referenced elsewhere in the publication).
- Avoid 3D charts. They can false perspective will distort the data.
- For example below, categories A and B seem equal when plotted in 3D. However, category B is noticeably larger, as shown when plotted in 2D.

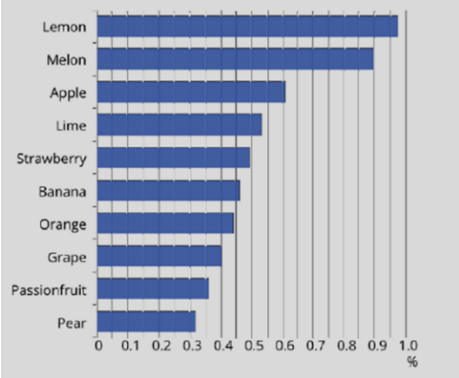


- Do not add chart borders.
- Apply style “Chart title” to allow automatic generation of a List of Charts.
- For multiple time periods where space is limited, then each periods name does not need to be included in the chart, every third month for example could be named.
- Horizontal bar charts are often used when data sets have long names.
- Double bar charts or pyramid charts are usually confined to age-gender profiles.
- Pie charts or donut charts can be used as an alternative to a bar or column chart, e.g. to show a dissection of government expenditure by type.
- Pie charts should be used with caution as they can misrepresent the data especially if there are several segments of relatively similar size.
- Also, it is preferred that the segments are not more than 8.

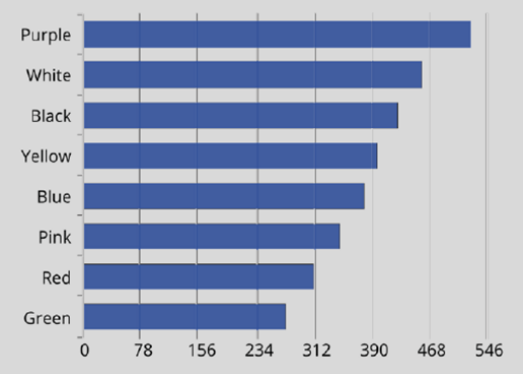
Gridlines and Gaps

- Use gridlines sparingly. There should usually be between four and eight gridlines per chart.
- Generally show grid lines in a light grey colour.
- Don't use too many gridlines.
- Choose sensible intervals

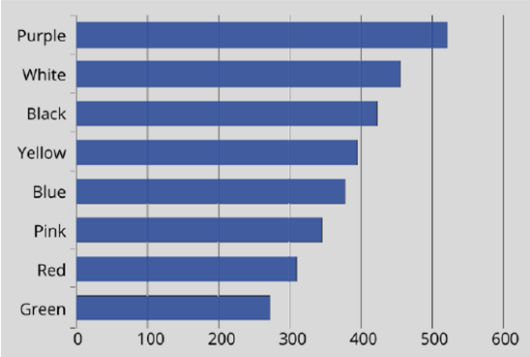
Poor



Poor



Good

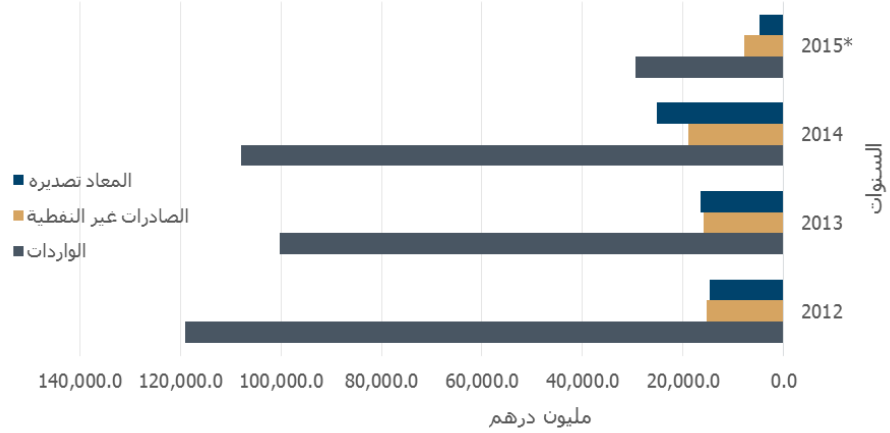


- A standard bar chart should have gaps between bars that are slightly narrower than the bars. The exception to this is a histogram.

Y Axis

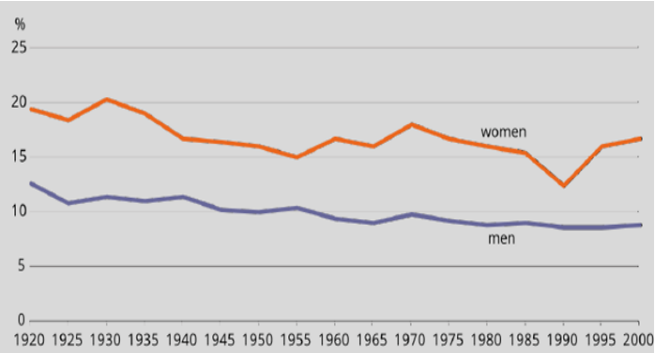
- The Y axis is the vertical axis.
- Align the Y-axis to the left side (in English) and to the right side (in Arabic).
- Place axis title at left for English, and right for Arabic (see example below).

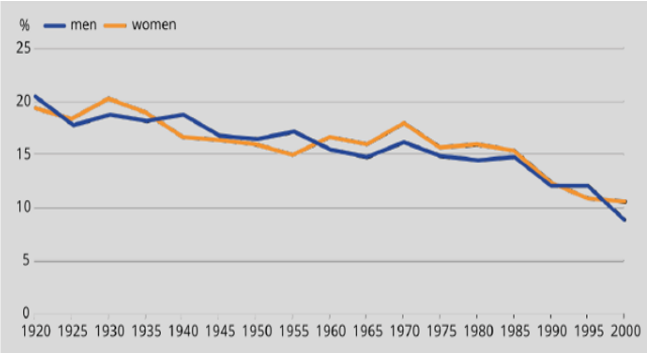
- Show zero if appropriate; note that sometimes relative differences in values can look misleading if no zero.
- Y-axis values should be ranked in ascending order



Legend

- Try not to use legends. Instead label the data directly.
- If a legend or key is necessary, place it on the chart as close as possible to the data.

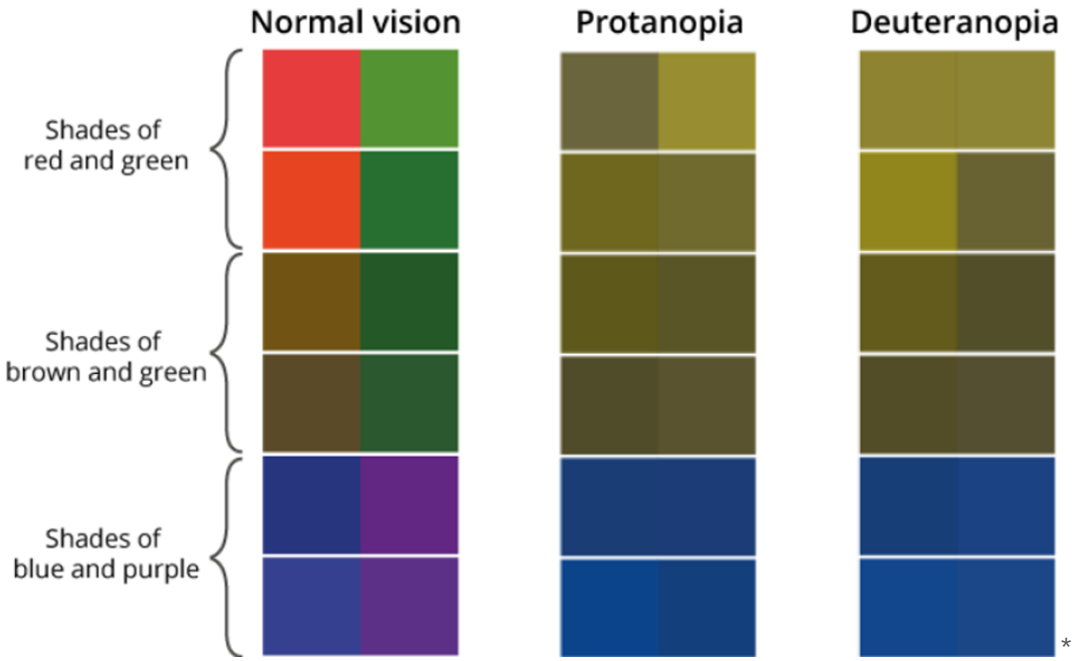




Colours

- Get it right in black and white
  - Make sure your design works in greyscale (black and white) before adding in colour.
- The safest hue is blue
  - The richest colour across all types of colour blindness is blue, therefore blue is often a good colour choice.

- Red and green should never be seen
    - Avoid using red and green together as they are difficult to distinguish from one another in the more common types of colour blindness.
- Lines have a smaller area than bars, which means fewer colours can be distinguished.
- Try to avoid using more than four shades of one colour on a line chart without using another mechanism to differentiate between them.



Types

Choosing the correct chart

Prioritise what you want to highlight in the data and choose the chart type accordingly. There are eight common relationships that charts display.

1. Comparisons of magnitude (size)
2. Time series
3. Ranking
4. Part-to-whole
5. Deviation
6. Distribution
7. Correlation
8. Spatial (maps)

Some resources to assist:

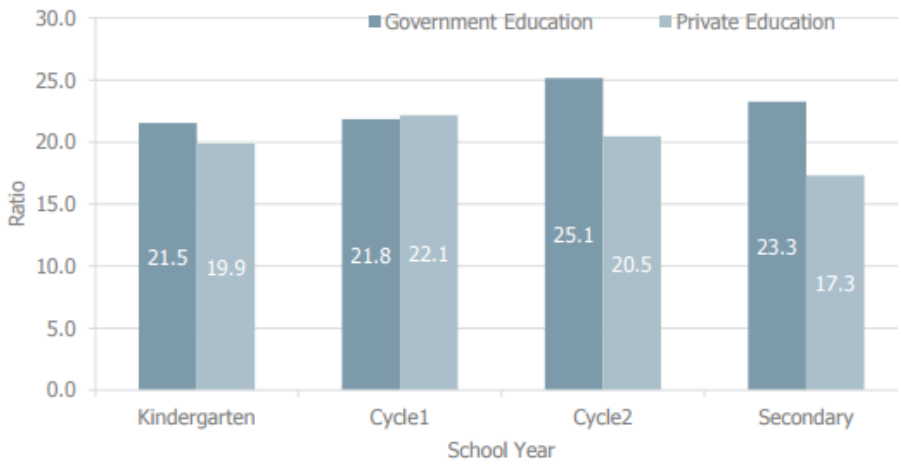
- Visual Vocabulary → https://github.com/ft-interactive/chart-doctor/raw/master/visual-vocabulary/Visual-vocabulary.pdf
- Visual Catalogue → http://www.datavizcatalogue.com/index.html

Bar Charts

- Best used for:
    - Comparisons of magnitude (size)
    - Ranking
    - Deviation
    - Time series
- Shown as horizontal or vertical bars
- There should always be a gap between the bars, but the gap should not be wider than the bars themselves
- Unless there is a natural order, bars should be arranged in order of size, from the smallest to the largest

- Generally, show start of y-axis as zero. If there are small differences between values sometimes it is necessary to start at a non-zero value (rare).

Figure 2: Pupils per Classrooms by Educational Stage and Sector 2015-16

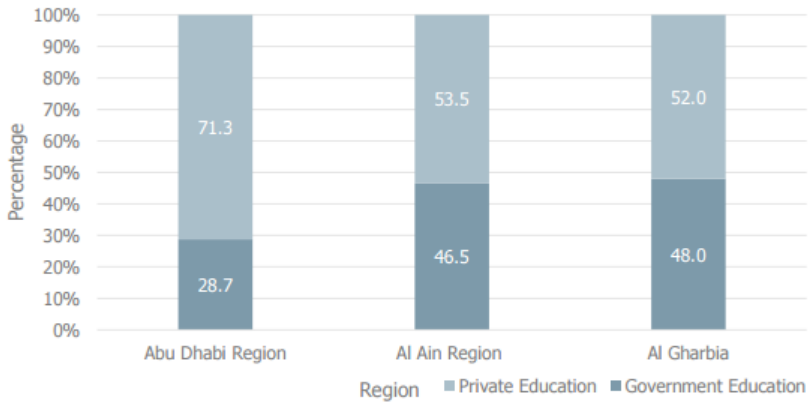


Source: Statistics Centre – Abu Dhabi and Abu Dhabi Education Council

Stacked/Grouped Bar Chart

- Best for :
    - Part to whole.
- Sometimes it is useful to show segments of a total for comparison.
- Stacked bar charts can be shown either in actual numbers or percentages.
- However, percentages tend to be more useful for comparisons between bars as well as within.

Figure 3: Percentage Distribution of Enrolled Pupils by Region and Sector 2015-16

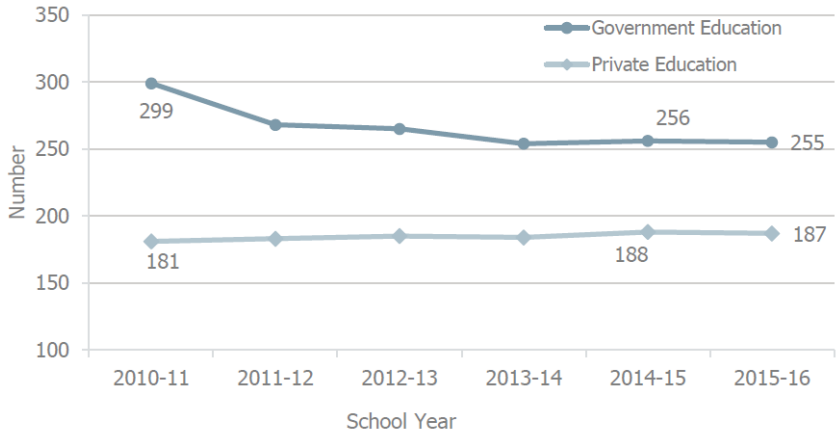


Source: Statistics Centre – Abu Dhabi and Abu Dhabi Education Council

Line charts

- Best used for:
    - Change over time (time series)
- A time series line can show:
    - Change
    - Trend
    - Growth
    - Decline
- Time should always run from left to right along the horizontal axis (English).
- Can be presented with bar charts for comparing rate versus number.

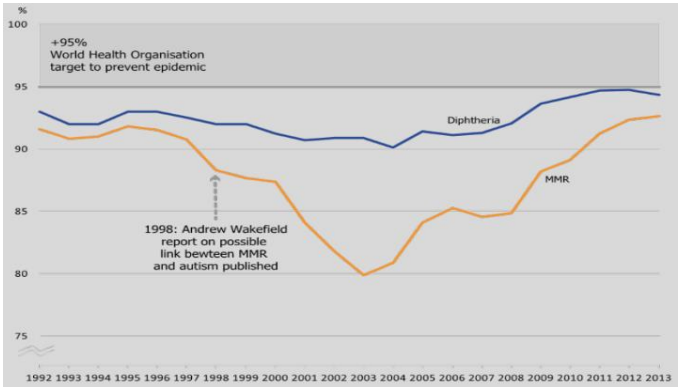
Figure 1: Schools by Sector, 2010-11 to 2015-16



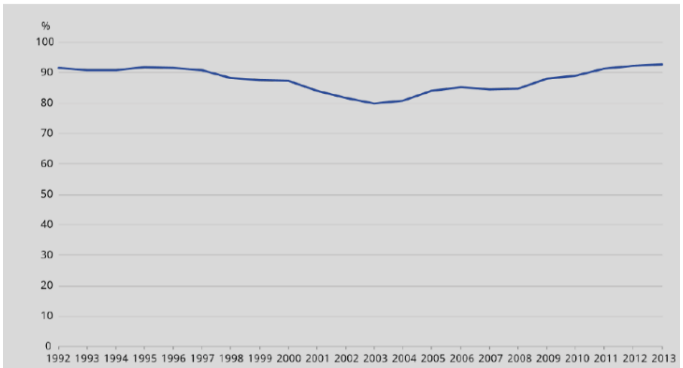
Source: Statistics Centre – Abu Dhabi and Abu Dhabi Education Council

Scale

- Time series charts don't have to begin at 0, if a chart doesn't start at 0 this must be indicated by breaking the y-axis in an obvious way.
- A chart can tell a very different story depending on the scale of the axes. See below MMR example where the adjustment of scale provides a more accurate picture of MMR vacancy uptake (compare blue line in top chart with orange line in bottom chart).

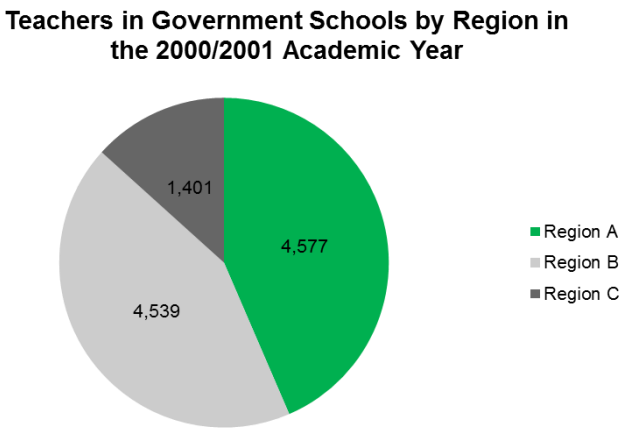


MMR vaccination uptake at age 1
UK, 1992 to 2012



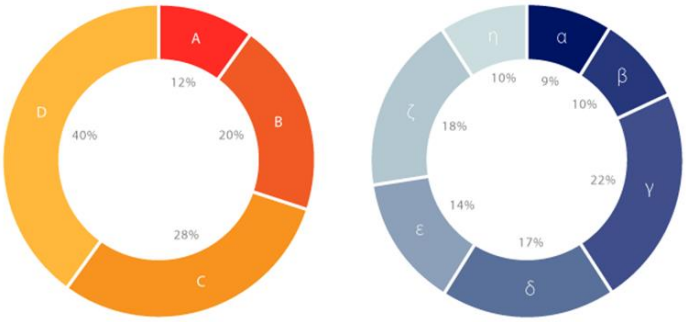
Pie Chart

- Best used for:
    - Part to whole
- Part to whole shows:
    - ratios
    - percentages
    - proportions
    - shares
    - hierarchies
- Used to demonstrate the percentage distribution of one variable.
- Many analysts prefer bar charts.
- Pie charts can be confusing for the reader – the size of the different segments can be difficult to compare. Often the segments are labelled with the appropriate values to prevent this.
- Don't use more than six segments.



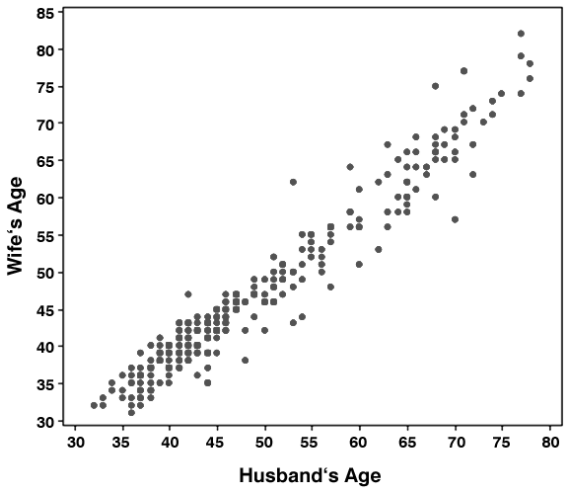
Donut Pie Chart

- A donut chart is essentially a Pie Chart with an area of the center cut out.
- Visually, Donut Charts seem easier to determine relative sizes of the pieces to one another.



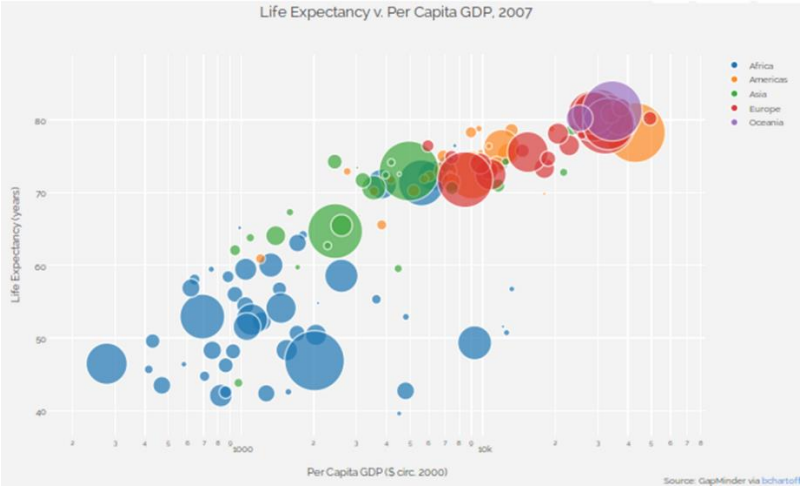
Scatter Plots

- Best used for:
    - Correlations
- Scatter plots show the relationship between two variables.
- They are often used as a first step to determine whether further analysis such as regression is worthwhile.
- Consider your audience before using a scatter plot – it may be better to use bar charts
- Trend lines are sometimes added to scatter plots if there is evidence of correlation between the two variables.



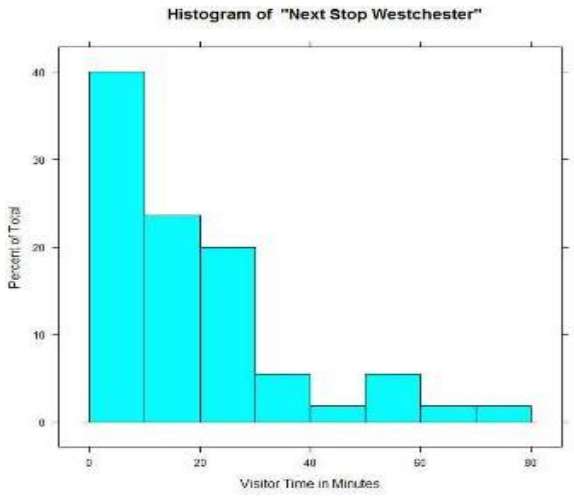
Bubble Charts

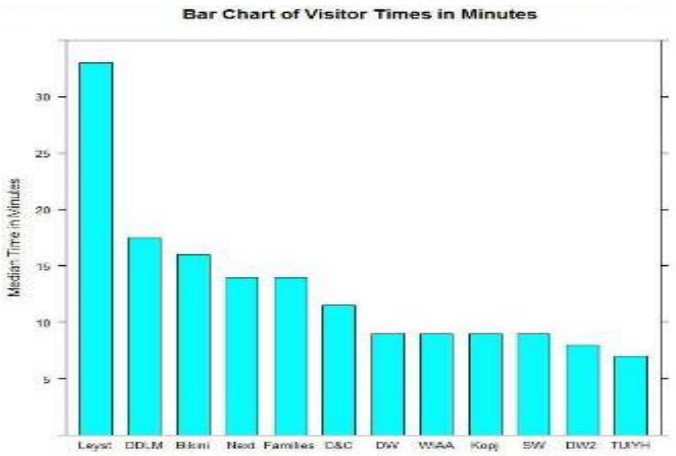
- Best used for:
    - Correlations
- Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.



Histograms

- Histograms are used to show distributions of a continuous variable (e.g. age), while bar charts are used to compare variables.
- Histograms plot binned quantitative data, while bar charts plot categorical data.
- Bars cannot be reordered in a histogram, while they can in a bar chart.
- There are no gaps between bars in a histogram (because continuous), there should be gaps in bar charts.



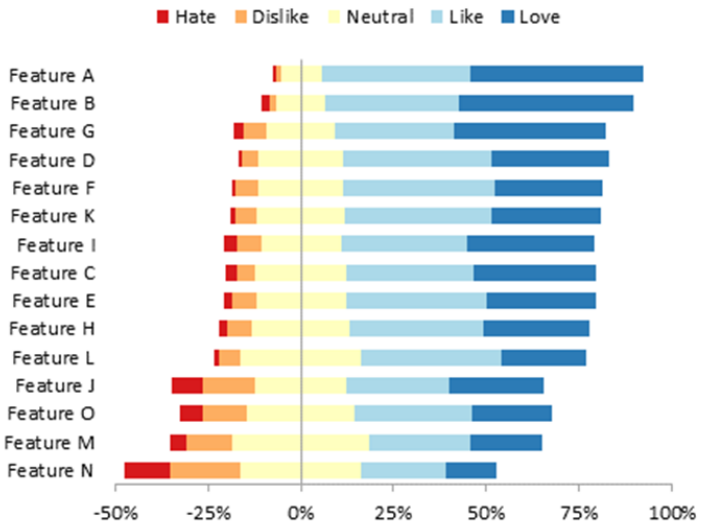


Diverging Bar Chart

- Best used for:

- Deviation (from a known point)

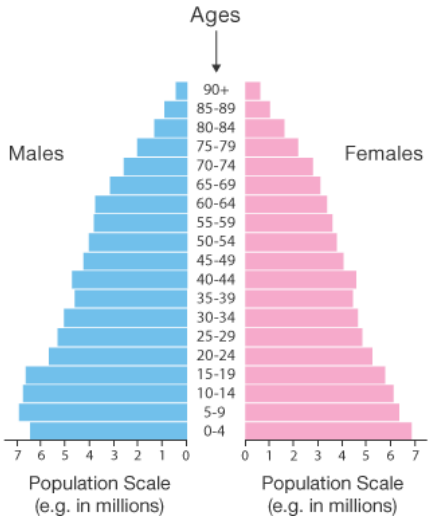
- Simple standard bar chart that can handle both negative and positive magnitude values.



Population Pyramid

- Best used for:
  - Deviation

- A Population Pyramid is a back-to-back pair of Histograms (for each gender) that displays the distribution of population in all age groups.



4. Infographics

Introduction

- A picture paints a thousand words.
- Great way to understanding data is to visualize the numbers as a picture.
- Makes it easier to see a pattern and can reveal patterns that might otherwise have been missed.
- An infographic is a self-contained visual story, presenting information, data or knowledge clearly, with meaning and context, and without bias.
- There must be a balance between design and function. Complicated visualisations often fail to communicate.

Process

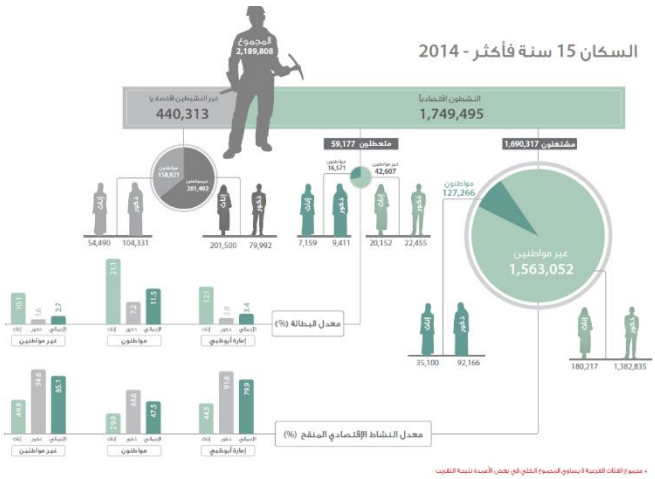
- Before drafting infographic make sure team (graphic designer, statistician, project leader) all know the story the infographic is going to tell.
- Turn the story ideas into rough concepts. Include titles, data presentation and annotations. Refine concept.
- Apply design rules to the drawing. Refine hierarchy and flow of story.
- Review final design and discuss if story is effective.
- Send for approvals.
- Release via channels (e.g. website, media, social media, etc).

Types

- Individual infographic - presents a single picture where the graphical elements are focused on a specific subject (e.g.Agricultural holdings, or employed persons).
- Storyboard Infographic - is a collection of individual infographics that are related to the same topic. Starts with a general topic and drill down to specifics. Can incorporate data from other official sources to add extra context to the story.

Individual

Storyboard



Design Principles

Never deceive the viewer

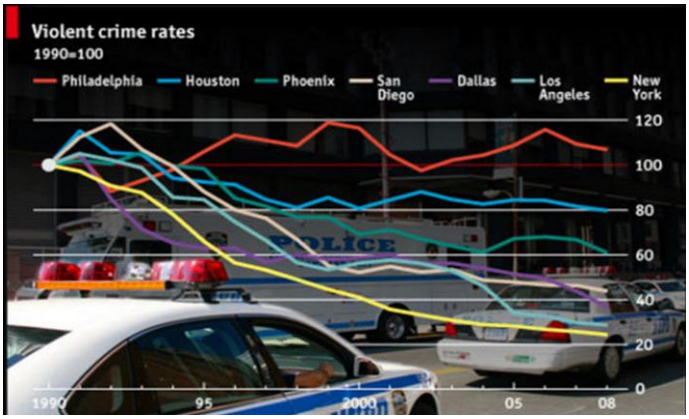
- Visualisation ethics relate to the potential deception that can be created, intentionally, or otherwise, from ineffective and inappropriate representation of data.
- Sometimes it can be through a simple lack of understanding of visual perception.

Have an objective / story

- Before you start making an infographic make sure you have a clear objective.
- Avoid making infographics simply because it is the current trend. It is advisable to think about what your end result will be.
- Try and find the interesting story in the data. Not all data will result in an interesting infographic, so be selective.

Functional and clean

- This infographic is a **poor** example.



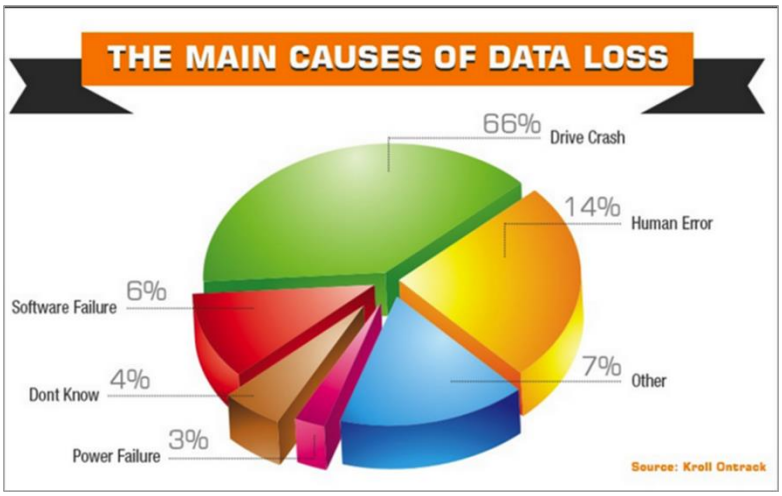
Simple colour palette

- Some of the best and most effective infographics use basic colour schemes.
- Avoid overwhelming your audience with different colours that do not match each other.

Source and accuracy

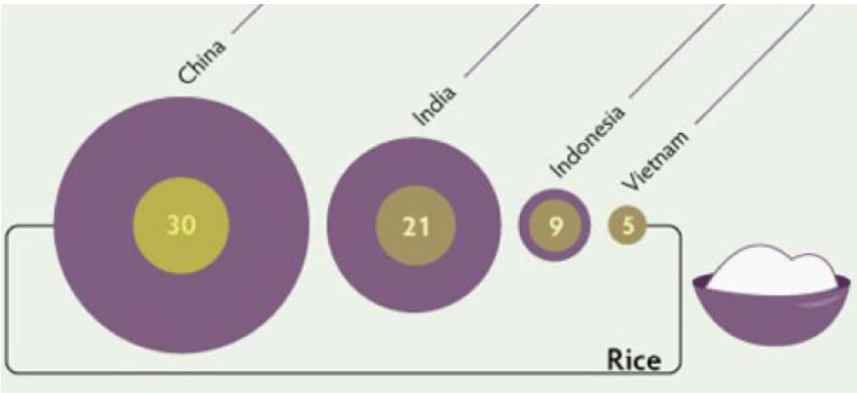
- The audience trust statistics from SCAD and they must also trust that the infographics presented are true and factual.
- Make sure that you have well-researched and factual data.
- To make your infographic more trustworthy, cite your resources at the bottom of the infographics so that your audience can check whether the pieces of information provided are unbiased and true. Ensure citations for all sources used are included.

Poor



Area of infograph shapes

- The human visual system naturally experiences a circle size in terms of its area and the area of a disk—unlike its diameter or circumference—is not proportional to its radius.
- So, if one chooses to scale the bubbles using the radius, then one must scale each new disk's radius to the corresponding data value.
- For example, if the value and radius of a bubble (r1) is 3cm, then the area of the bubble will be 28.26cm2 (pi x $r1^2$).
- To calculate the correct area for other data values we only need to know the scale of change between the data values. Say the next bubble has a value of 15, then the change in area will be 5 times.
- So the area of the second bubble will need to be 141.3cm^2 (5x28.26). To determine this area use (change x pi x $r1^2$) or (5 x 3.14 x 9) = 141.3cm^2.
- In the example below the production of rice is incorrect in the purple colour. The real proportions have been shown in the yellow (smaller bubbles).



Pictograms

- Repeating icons can be used to represent units of data.
- Provide a key to denote the icon meaning and value that it represents.
- Never represent slices of icons, instead show the full icon and use shading for any portions that the value represents.

5. Maps

Introduction

- Useful tool for putting statistical information into geographical context.
- Used to demonstrate patterns within data.
- Attractive – improve the look of statistical report.
- Should be able to consider the map in isolation.
- Should have suitable title, labels and source.
- Maps should be used to show geographical locations and spatial distribution of the data and compare regions.
- As with charts, maps should not be used when there is no variation to demonstrate.
- It will not be possible to produce a map if there is no geographical breakdown of the data.

Types

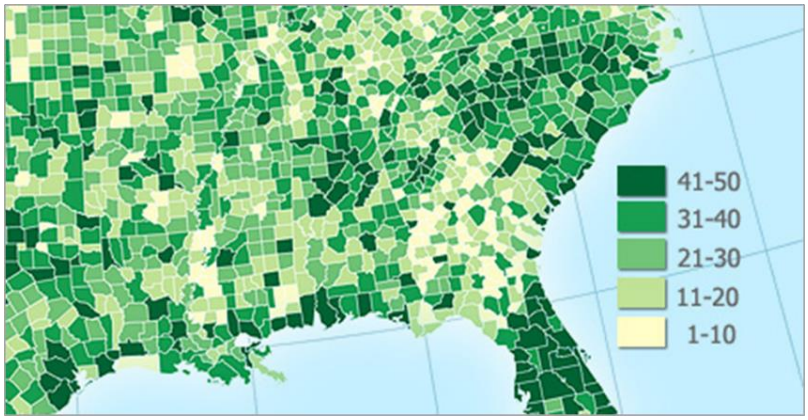
There are two main categories of maps:

- Reference maps: These are used to show borders, location and interesting features (e.g. lakes). These maps are also known as topographic maps.
- Thematic maps: These are used to show the spatial distribution of statistics and are also known as statistical maps. Three main types of thematic maps:
    - Choropleth maps
    - Graduated symbol maps
    - Cluster / Dot maps

Choropleth

- Most common type of statistical map. Used to show the spatial distribution of ratios.

- Only ratios should be displayed – proportions, rates, and densities.
- Choropleth maps should not be used to display absolute values.

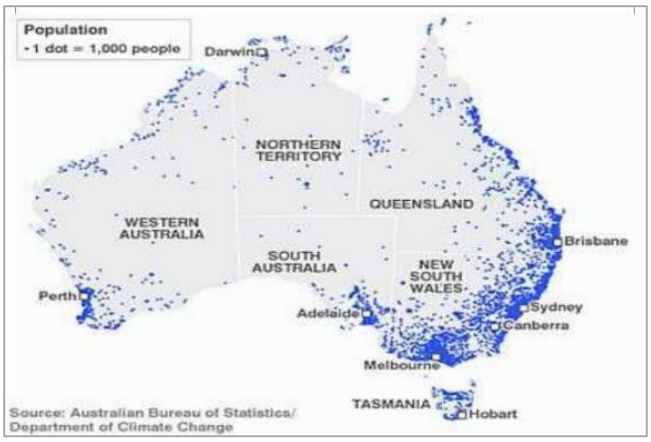


Graduated symbol

- Graduated symbol maps, or proportional symbol maps, are used to display absolute values, where the size of the symbol (often a dot) is proportional to value it represents.
- Circles are often used, as they are easy to scale.

Cluster / Dot

- These are similar to graduated symbol maps, in that dots are attached to locations.
- However, all dots are the same size and each dot represents a value.
- Therefore, locations with greater values attached have a higher number of dots.

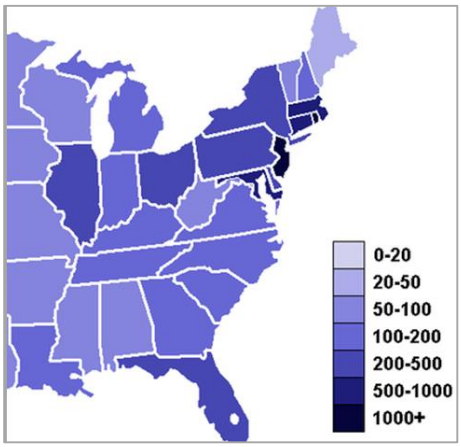


Interactive maps

- Geographical Information Systems (GIS) and statistical mapping software have become readily available in recent years
- Maps can be static or interactive. Static maps are produced using images and cannot be edited by the reader.
- Interactive maps allow the reader (user) to alter the design of the map, changing the data required.
- The data and metadata sits behind the map therefore allowing the user to changes the parameters without recreating the map.

Guidelines

- Ranges set to sensible classes (no more than 8). Clear limits and no gaps. Shades of a colour better than many different colours.



- Make sure that all aspects of the map are explained, including north arrow, distances, symbols, patterns and colours.

- Where there are no data available for an area, make sure this is identified. Usually with white background.
- Location map insert can be useful if the audience is not familiar with the geographical area in the map, as it puts the location in context.