Local Government Training Institute (LGTI) Short Course Data Curriculum

Data Fundamentals Course

Module 5: Presenting Data

Disclaimer: This work is developed by School of Data with funding from The World Bank Tanzania Data programme. For more information visit https://www.schoolofdata.org.

Module 5: Presenting Data

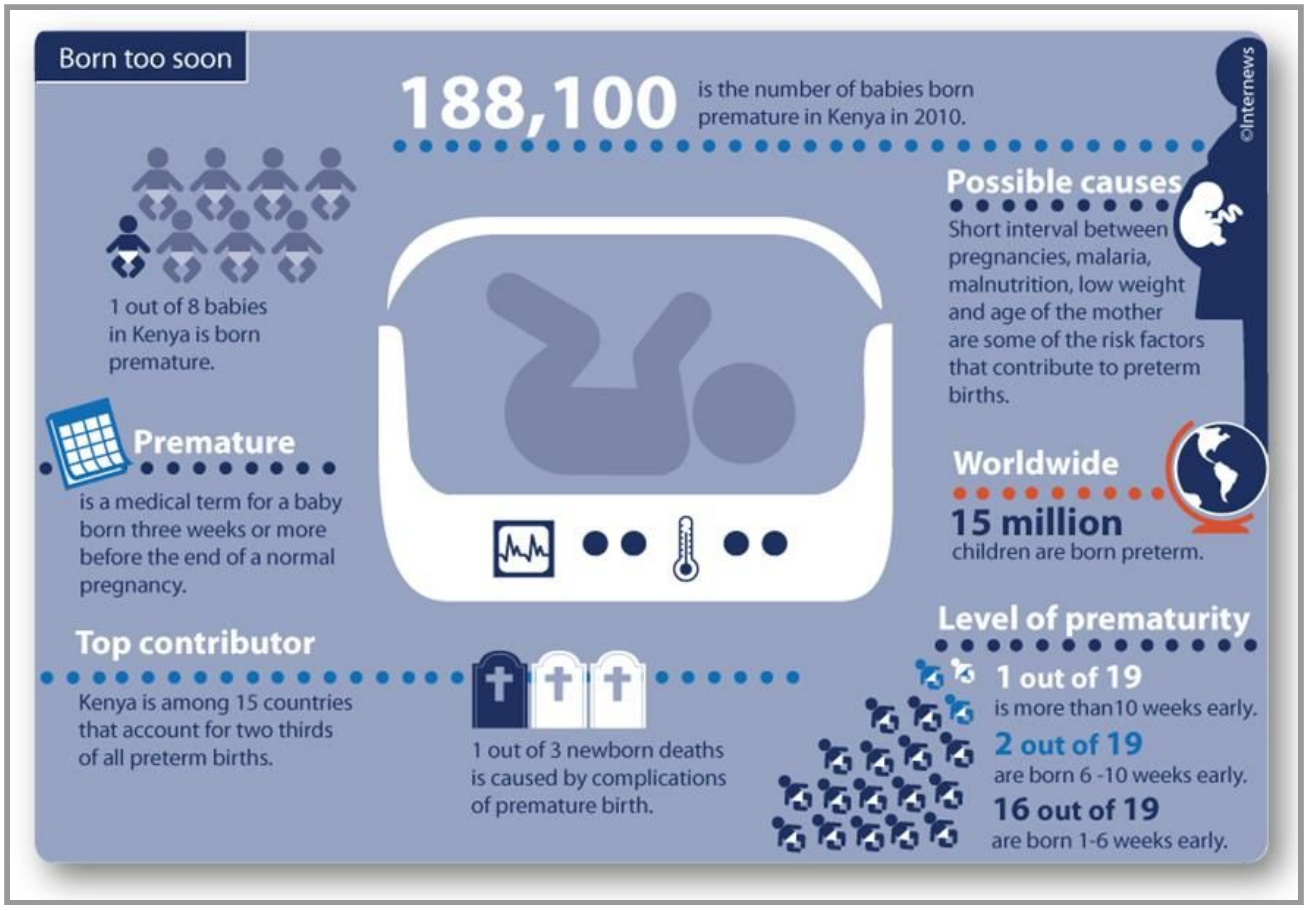
STUDENT WORKBOOK

In this module, you will learn the basics of effective communication with data visualisation. You will learn how the human brain is adapted for understanding data through visualisation and what effective sensory engagement looks like. This module will also introduce principles to matching data types to graphical forms for effective communication, basic design principles and data visualisation ethics. Students will learn to redesign data visualisations with this new knowledge. Most importantly this section is about conveying knowledge and information through visual stories, not just dry statistics.

At the end of this module, you will be able to:

- Appreciate the purpose of data visualisation
- Develop an effective data visualisation design strategy
- Match data and graph types
- Address ethical issues that arise with data visualisation
- Convey stories with data visualisation

Getting Started



Simple Fixes To Newborn Deaths

Kenya's high child death rate of 73 per 1,000 children could be greatly reduced if the country were to employ simple and practical solutions to reduce preterm births and complications, which are the leading
cause of death among newborns.

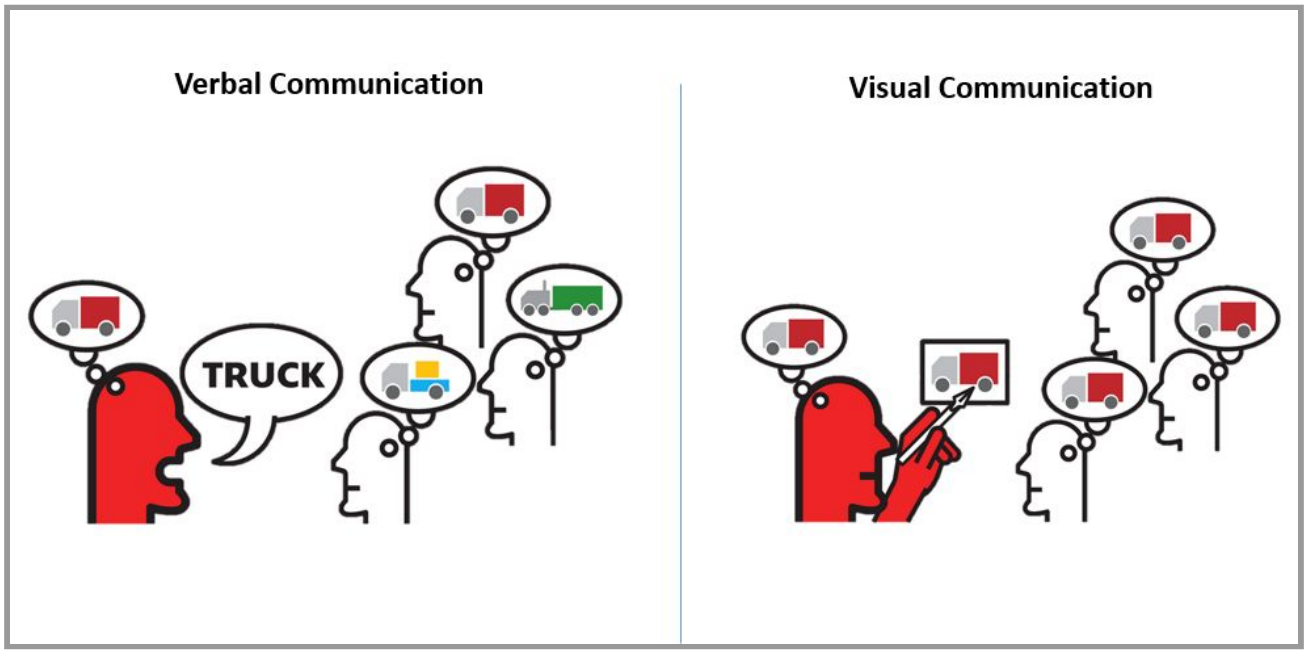
The fourth Millennium Development Goal (MDG 4) aims to reduce the 1990 mortality rate among under-five children by two thirds. Child mortality is also closely linked to MDG 5 - to improve maternal health. Since more than one third of all child deaths occur within the first month of life, providing skilled care to mothers during pregnancy, as well as during and after birth, greatly contributes to child survival.

In this module, you will use data to quantify and explain the problem of major development issues, like the rate, causes and solutions for newborn deaths.

Lesson 1: Purpose of Data Visualisation

In this lesson, we will review some best practices to follow when considering data visualisations to tell a
data story.

Why do we use data visualisation?



Visual communication has some distinct advantages over verbal or text communication. This section explores how visualisation can be more precise and concise than other forms of communication.

Why do we use data visualisation?

- The brain likes images.
- Makes the data easier to understand and compare

- Grabs the audience's attention
- Allows the audience to explore the data
- Tells a story about the data
- Exercise: Simplifying Numbers

Lesson 2: Matching Data and Graph Types

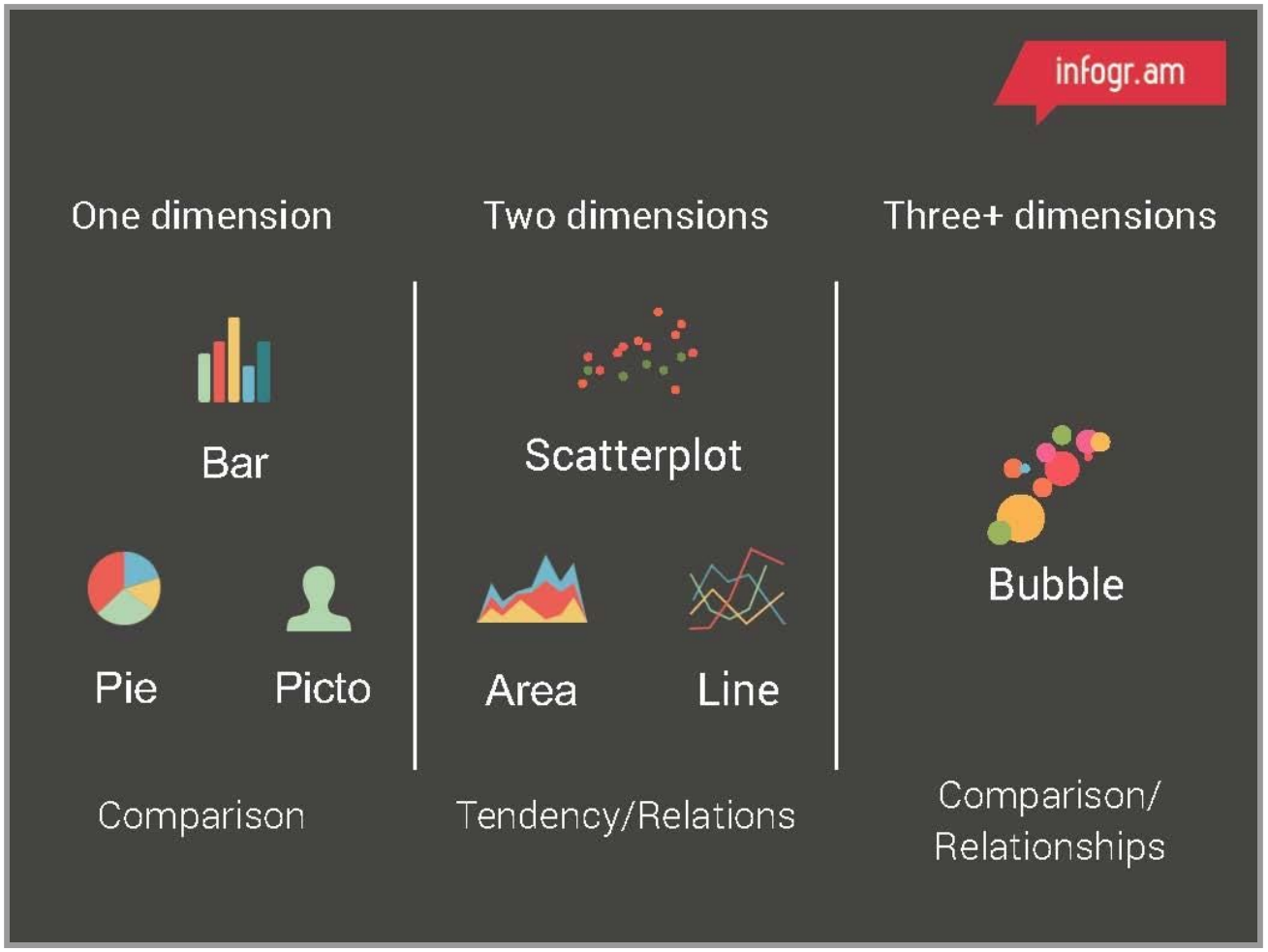
In this lesson, we will examine different data types, determine the message behind the data and match the data to an appropriate graphical form.

- **Bar chart:** Compare data across categories

- **Line graph:** Compare data across time
- **Scatter plot:** Compare interaction between two variables
- **Maps:** Compare data across geographical units
- **Pictogram:** Compare aspects of data through human depictions

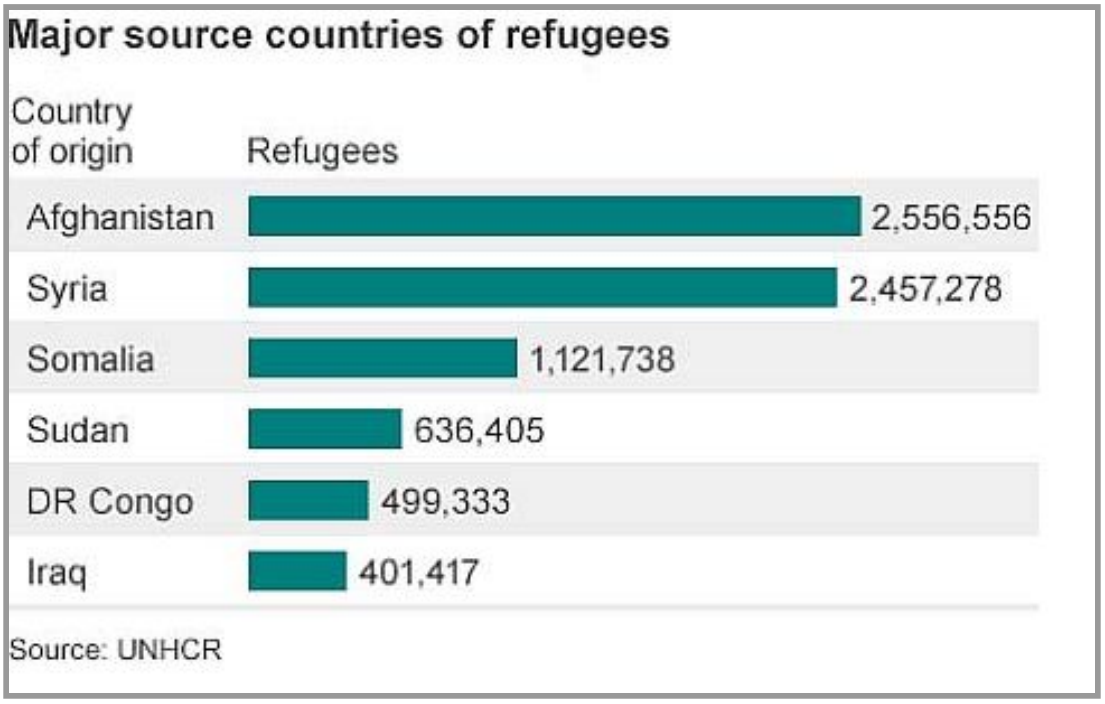
Before we begin working with datasets, let us begin our data literacy journey with learning how to interpret data visualisations.

Importantly, data visualisations help you to compare or connect two or more things, which are useful cues to find angles for data-driven stories. For instance, maps can help compare you compare administrative units (say the population density of two countries); while line graph can help compare time periods (say the population growth over two decades).



Let's look at the following types of commonly-used data visualisations: bar chart, line graph, pie charts, maps, and pictograms.

Interpreting Data Visualisations: Bar Chart

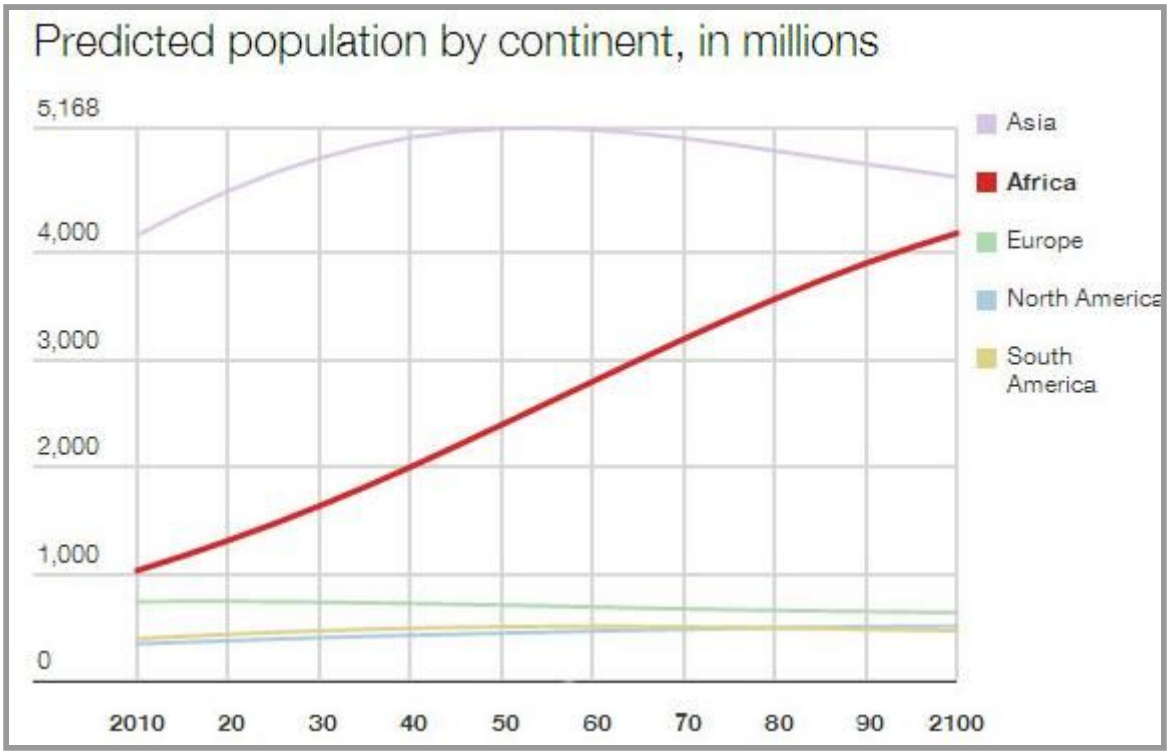


Bar charts are used to compare values across categories. In this example[1], total number of refugees is compared across countries of their origin.

Questions:

- Which countries produce the most refugees?
- How many times more refugees come from Afghanistan than from Iraq?
- What tells you that you can verify the data in this chart?

[1] http://stopwar.org.uk/news/war-drives-global-refugee-figure-past-50m-for-first-time-since-second-world-war



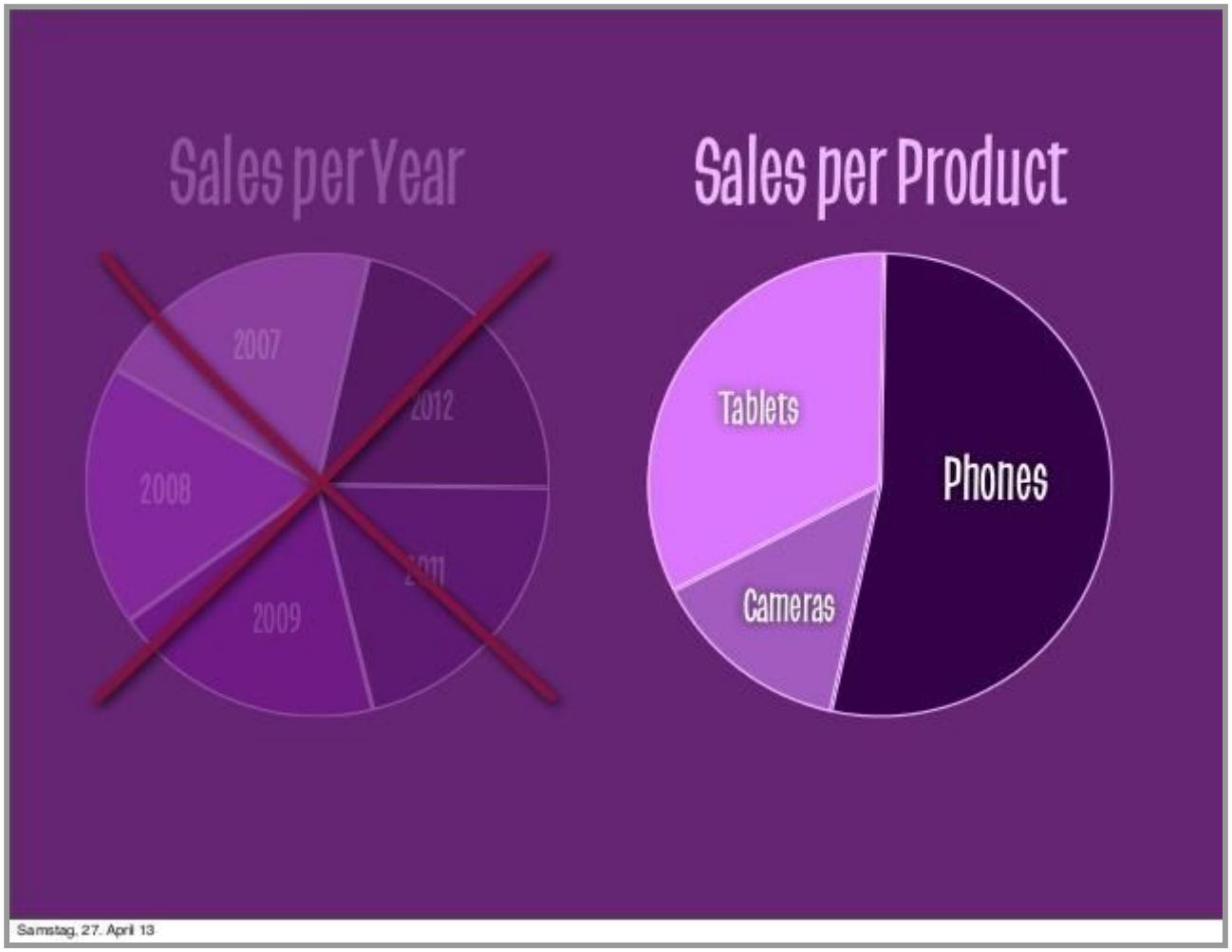
Line graphs can be used to compare data over time. In this example , predicted population of continents is compared across ten decades.

Questions:

- Which continent has the biggest population now?
- Which continent will see the biggest population growth?
- What makes you notice Africa right away?

Interpreting Data Visualisations: Pie Charts

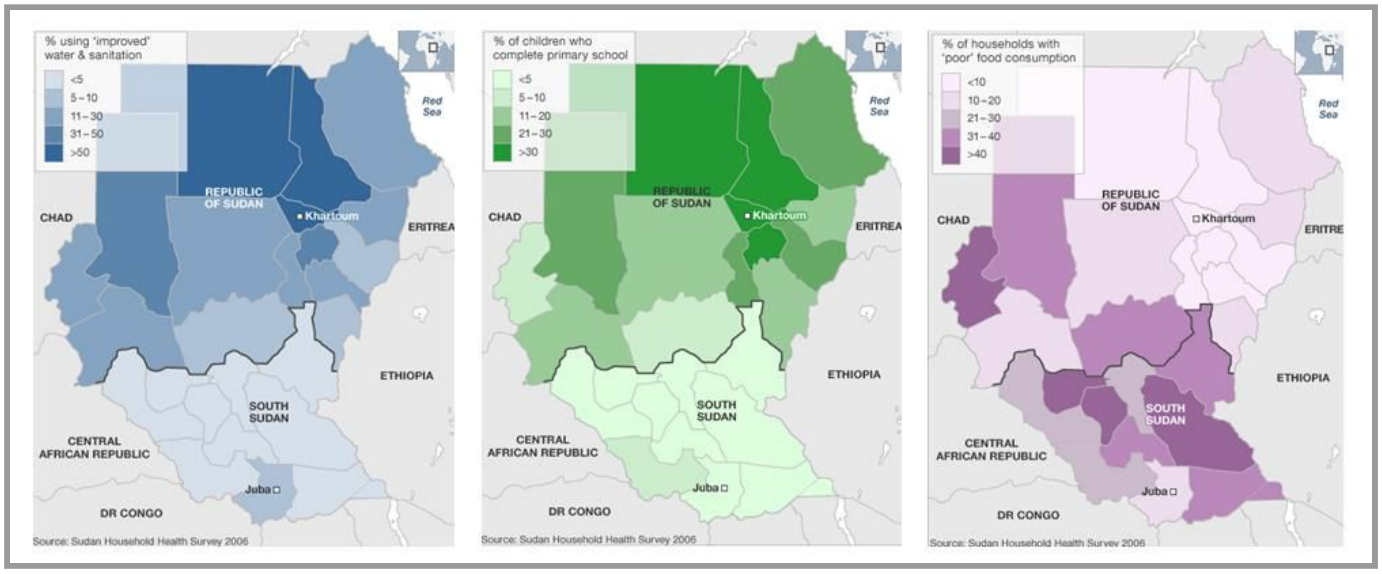
Pie charts are a popular visual form to show parts of a whole, for example, causes of death among young adults. Pie charts always have to add up to 100% (not the top two causes, all the causes). Pie charts should not have more than five slices because if there are more, the sizes are difficult to compare. They should also be different sizes.



Questions:

- What does the chart on the right measure?
- Which product has had the highest sales?
- What is incorrect about the first chart?

Interpreting Data Visualisations: Map



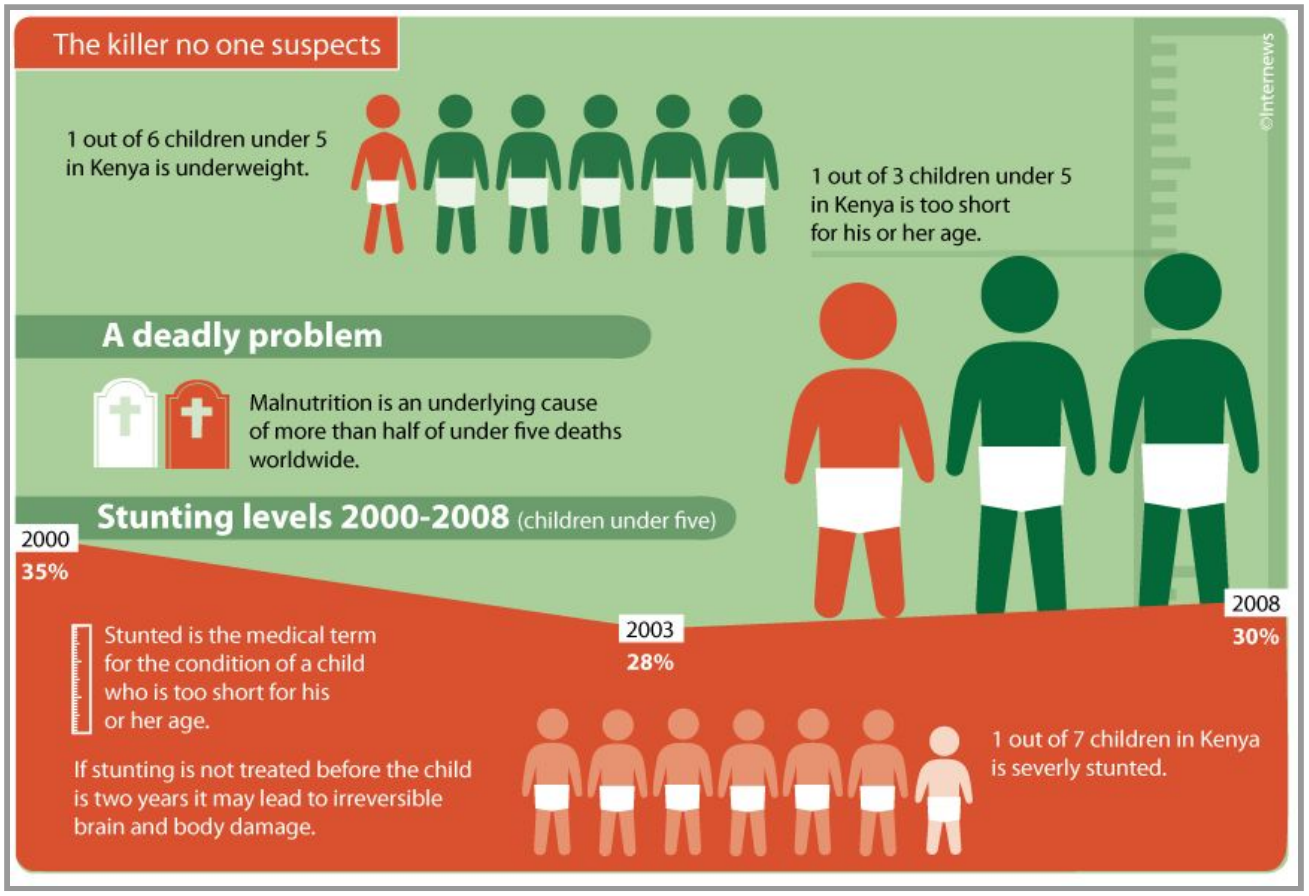
Maps are a familiar type of data visualisation – which are important for public interest stories as they let you compare geographies. There are various types of map visualisations; this example shows the visualisation type called "Choropleth". These maps use differences in shades or patterns inside defined areas in a map to indicate the average value of a variable – for instance population density or per-capita income. This example [2] shows three maps that display indicators for Sudan related to water and sanitation, schooling, and food consumption.

Questions:

- What do these different maps tell you?
- Do you see any trends?

[2] http://www.bbc.co.uk/news/world-africa-17126340

Interpreting Data Visualisations: Pictogram

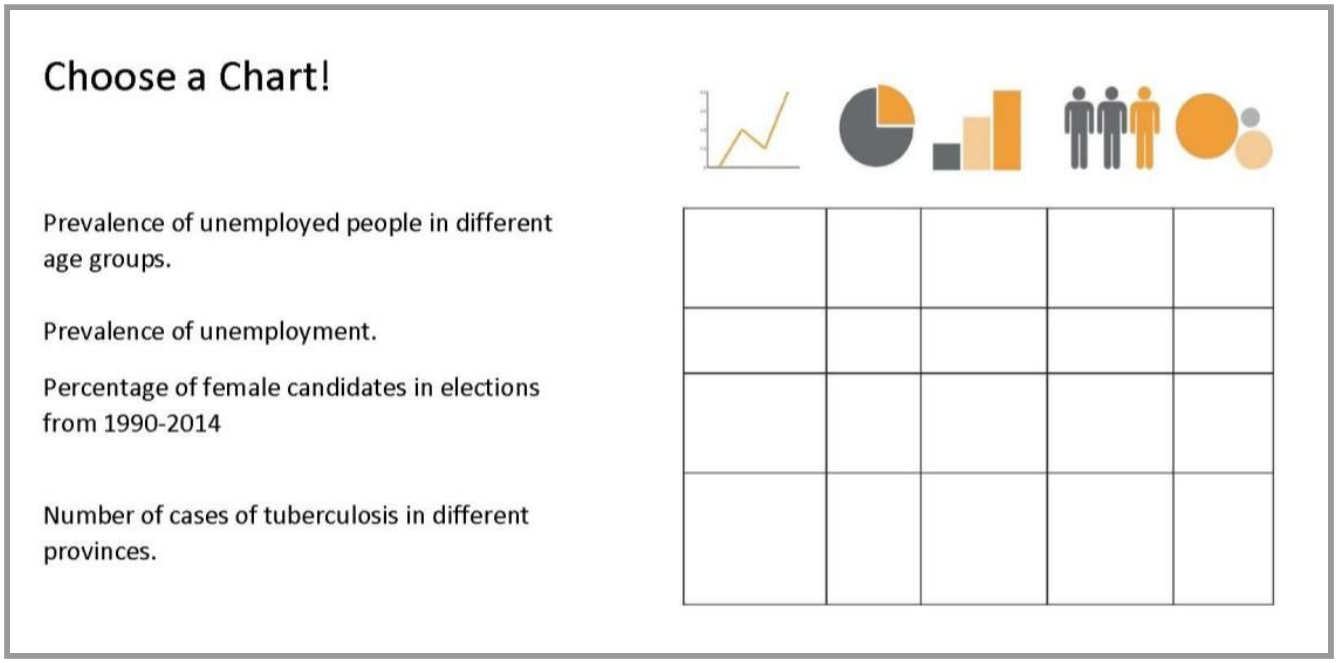


Pictograms are data visualisations that present information through pictorial resemblances to humans or other objects. This example highlights the impact of malnutrition on children's health in Kenya.

Questions:

- How many children are too short for their age in Kenya?
- How many children die of causes related to malnutrition?

Exercise: Matching Chart and Data Types



Indicate with an X in the box which chart type you would use for each of the listed data types.

Lesson 3: Design and Colour Basics

Basic Design Concepts

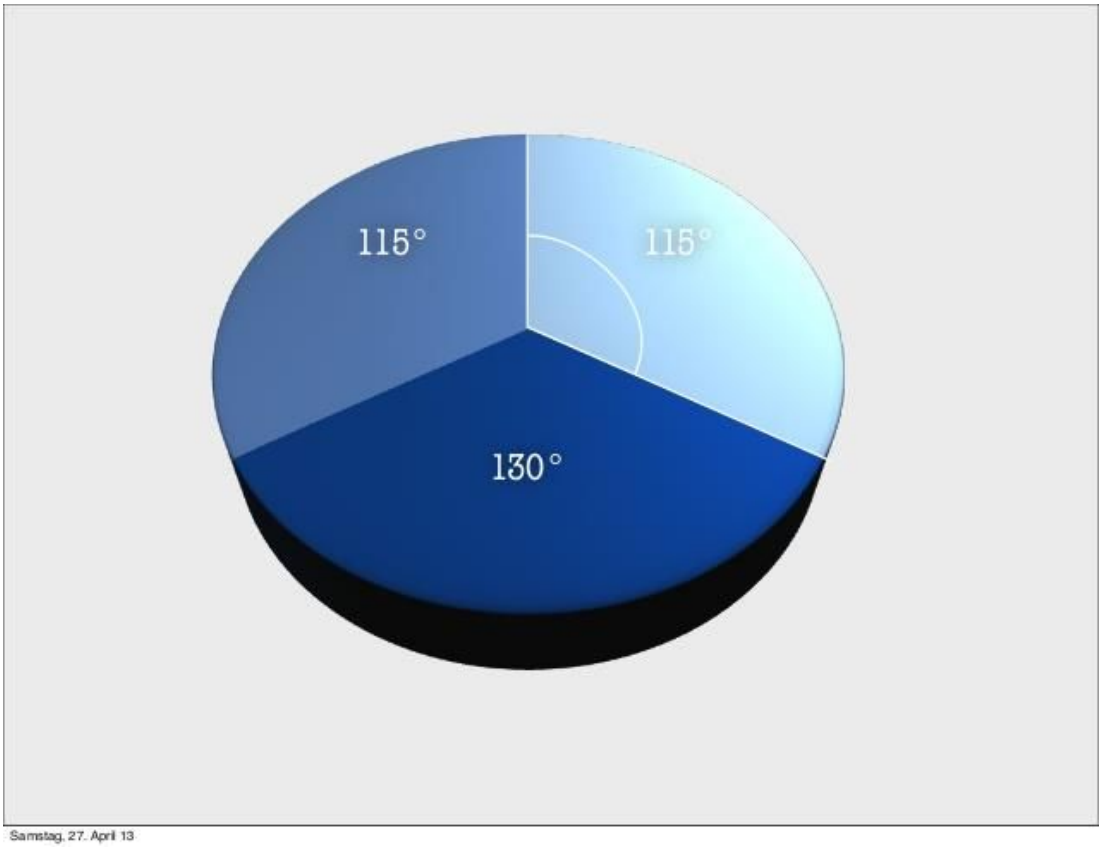
- **Simplicity:** Choose a maximum of three colors and fonts and stick with them consistently
- **Hierarchy:** All visualisations need a focal point to guide them through story

- **Brevity:** Keep text short and to the point
- **Creativity:** Incorporate playful design that related to the topic
- **Clarity:** Label clearly, specify units, use a legend when necessary

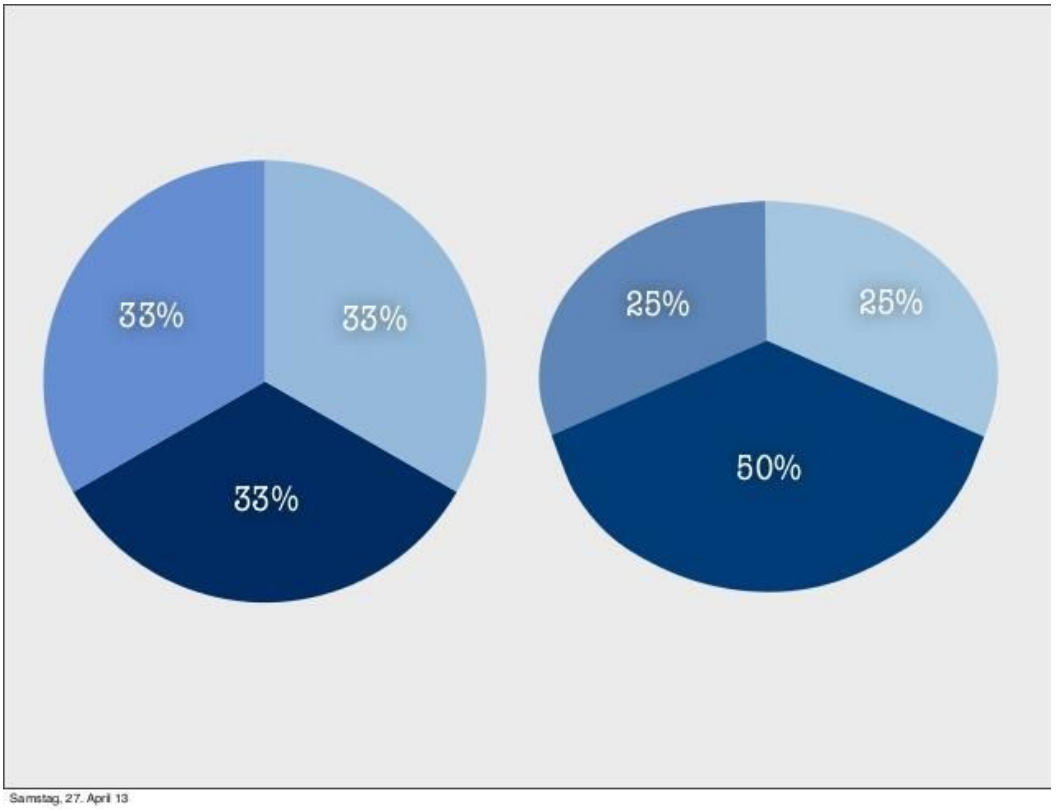
Five Golden Rules

1. **No 3D Effects**

   Three dimensional graphics distort the data. Whichever part of the data is closest to the reader is distorted to look larger. In the example below, see how a pie chart that is evenly divided into thirds, when 3D effect is added, distorts the data to look like the slice closest to the reader is larger.



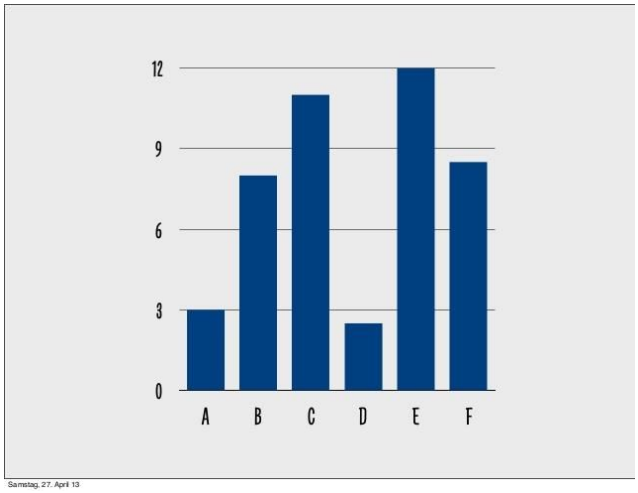
Samstag, 27. April 13

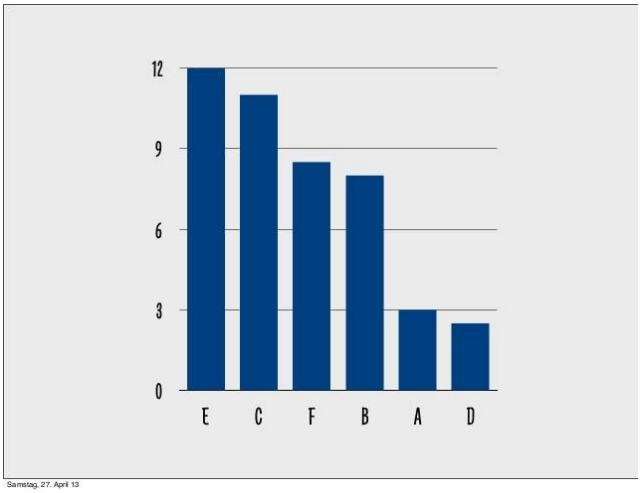


Samstag, 27. April 13

2. **Sort Data From Largest to Smallest**

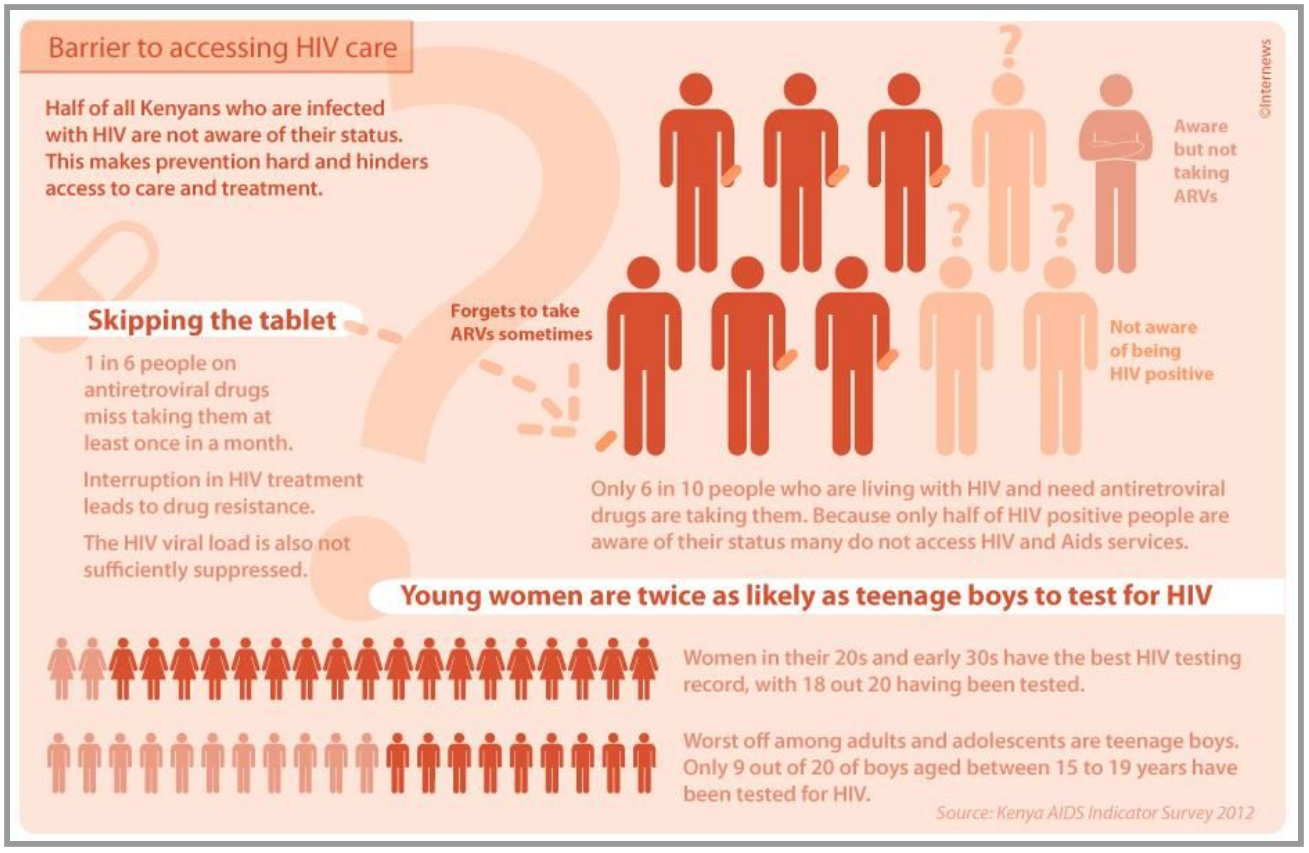
   You want your audience to easily be able to make a visual comparison across categories. To make this easier, sorting your data will order the bars in a bar graph or slices in a pie chart from greatest to least, making the visual comparison much easier.





3. **Choose maximum two colors or shades of the same color for your graphic. Stay away from rainbow colors.**

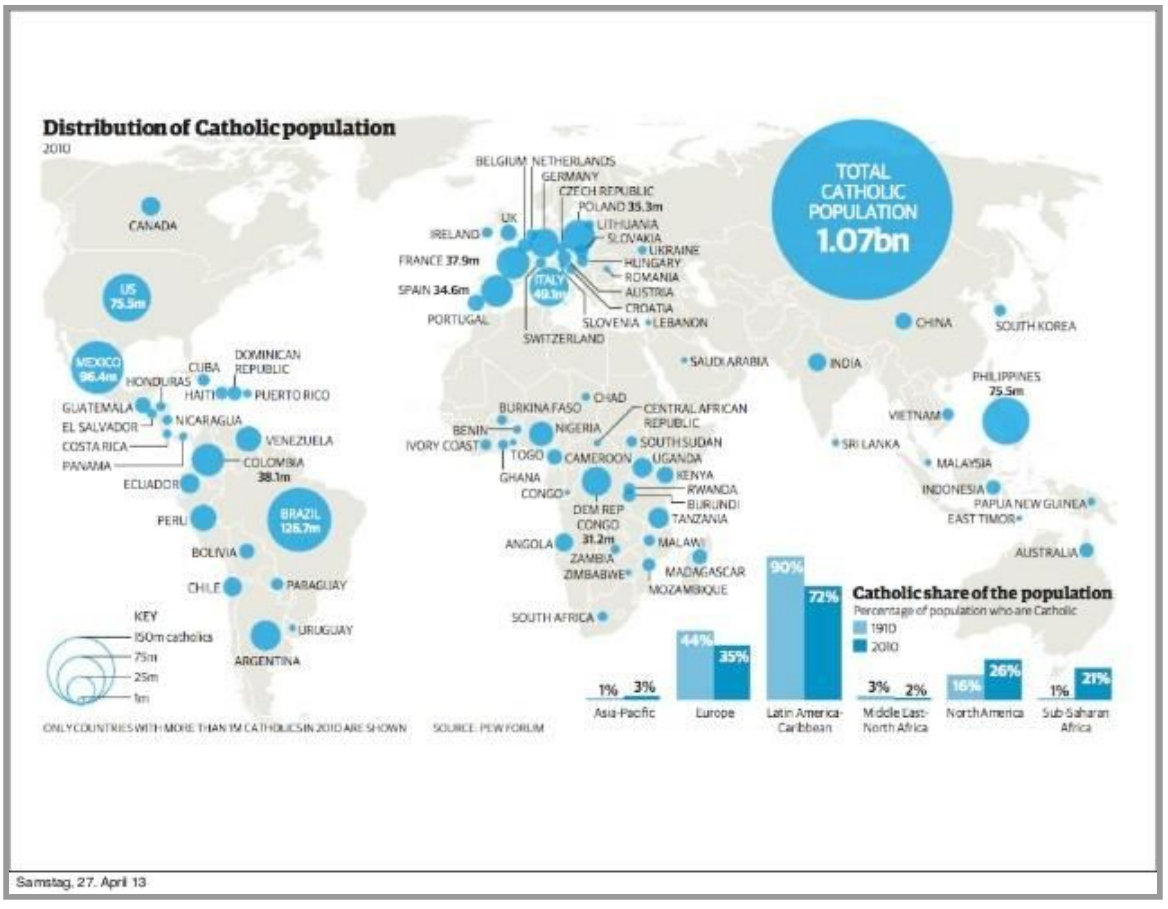
   Simple colors in the same color range and a consistent font make your visualizations look more professional and credible. Your headlines, labels and text should be clear and explain the visualization to the audience.



4. **Your headline should tell the story!**

   Before you create a data visualization, you should know what the message is. We create drafts in Excel to ensure that we have a clear headline and message before we start. It is important to be able to tell the audience where the data comes from but it's even more important to tell them through a strong headline, what the message of the story is.


?
What am I looking at?
Samstag, 27. April 13


Distribution of Catholic population
2010
TOTAL CATHOLIC POPULATION 1.07bn
CANADA
US 75.5m
MEXICO 96.4m
CUBA
DOMINICAN REPUBLIC
HONDURAS
HAITI
PUERTO RICO
GUATEMALA
EL SALVADOR
NICARAGUA
COSTA RICA
VENEZUELA
PANAMA
COLOMBIA 38.1m
ECUADOR
PERU
BRAZIL 126.7m
BOLIVIA
CHILE
PARAGUAY
URUGUAY
ARGENTINA
KEY
150m catholics
75m
25m
1m
BELGIUM
NETHERLANDS
GERMANY
CZECH REPUBLIC
POLAND 35.3m
LITHUANIA
SLOVAKIA
UKRAINE
HUNGARY
ROMANIA
AUSTRIA
CROATIA
SLOVENIA
LEBANON
IRELAND
UK
FRANCE 37.9m
SPAIN 34.6m
ITALY 49.3m
PORTUGAL
SWITZERLAND
CHINA
SOUTH KOREA
SAUDI ARABIA
INDIA
PHILIPPINES 75.5m
VIETNAM
SRI LANKA
MALAYSIA
INDONESIA
PAPUA NEW GUINEA
EAST TIMOR
AUSTRALIA
CHAD
BURKINA FASO
CENTRAL AFRICAN REPUBLIC
BENIN
NIGERIA
IVORY COAST
SOUTH SUDAN
TOGO
CAMEROON
UGANDA
GHANA
KENYA
CONGO
RWANDA
BURUNDI
DEM REP CONGO 31.2m
TANZANIA
ANGOLA
MALAWI
ZAMBIA
ZIMBABWE
MADAGASCAR
MOZAMBIQUE
SOUTH AFRICA
Catholic share of the population
Percentage of population who are Catholic
1910
2010
1% 3% Asia-Pacific
44% 35% Europe
90% 72% Latin America-Caribbean
3% 2% Middle East-North Africa
16% 26% North America
1% 21% Sub-Saharan Africa
ONLY COUNTRIES WITH MORE THAN 1M CATHOLICS IN 2010 ARE SHOWN
SOURCE: PEW FORUM
Samstag, 27. April 13

Samstag, 27. April 13

World Map of Catholic Population

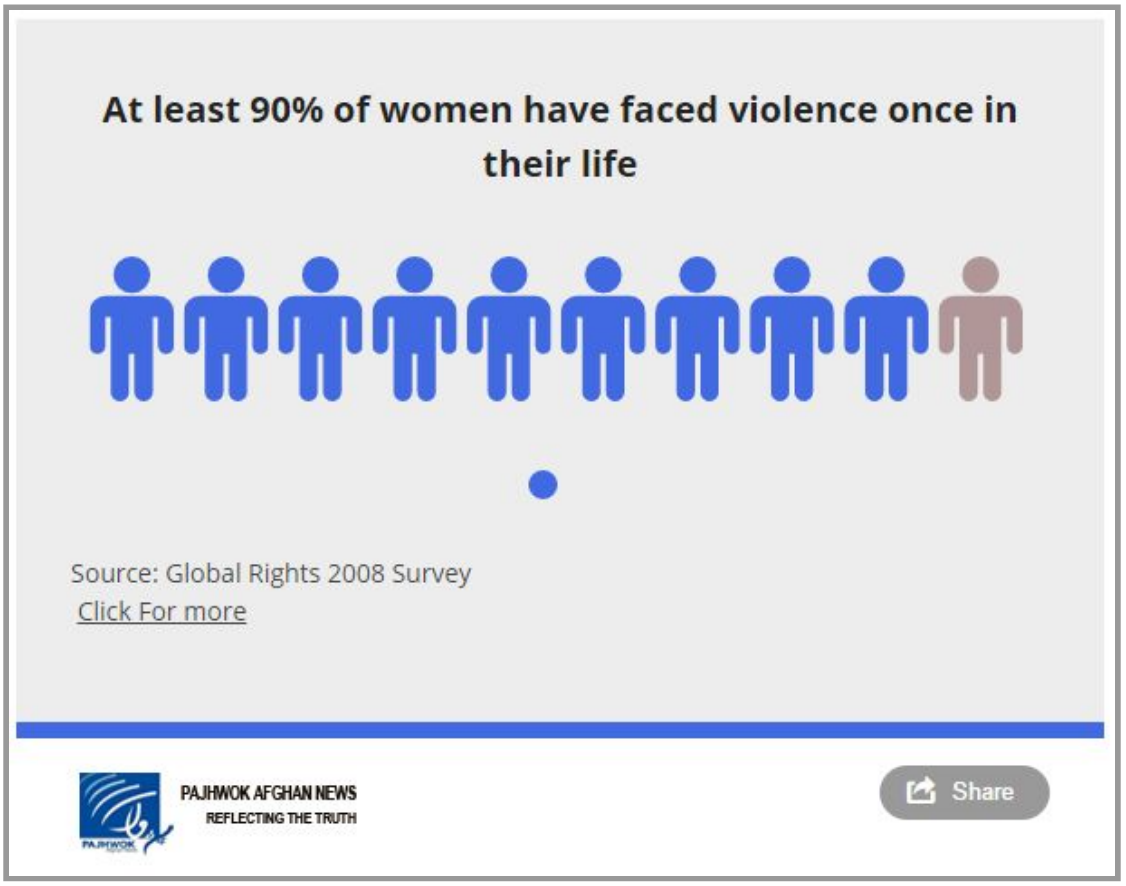
The Catholic Church Has Shifted Southwards over the Past Century

Samstag, 27. April 13

Examples:

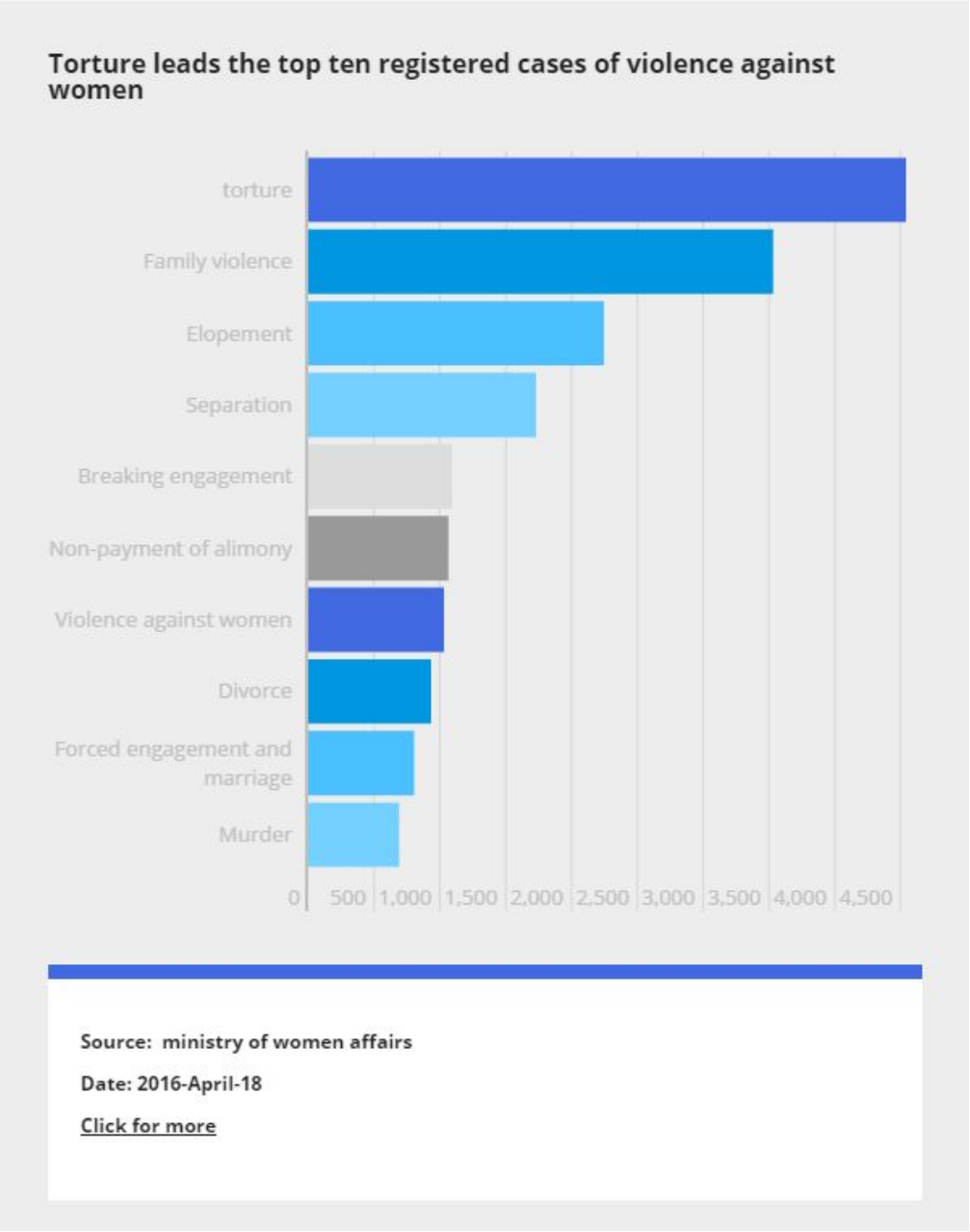
Bad headline: Rates of Domestic Violence in Afghanistan

Good Headline:



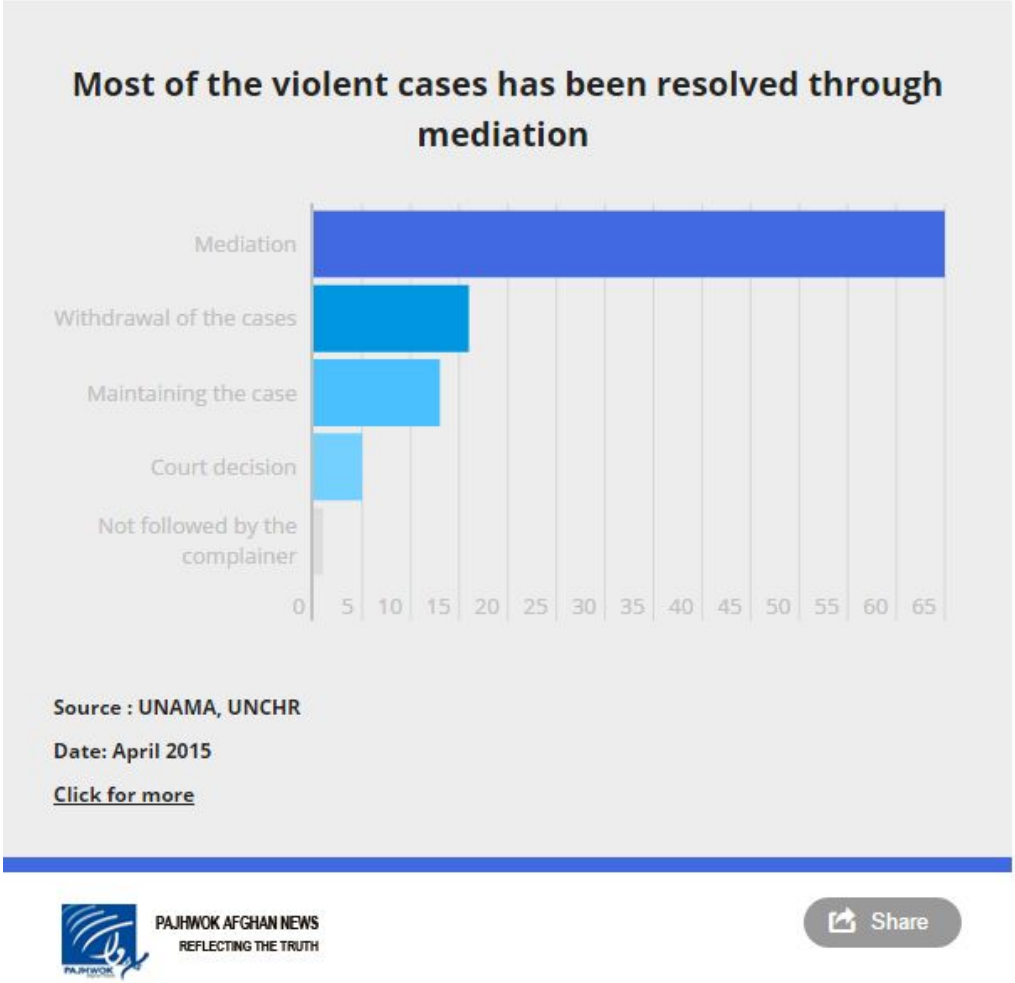
Bad Headline: Top 10 Types of Registered Domestic Violence Cases

Good Headline:



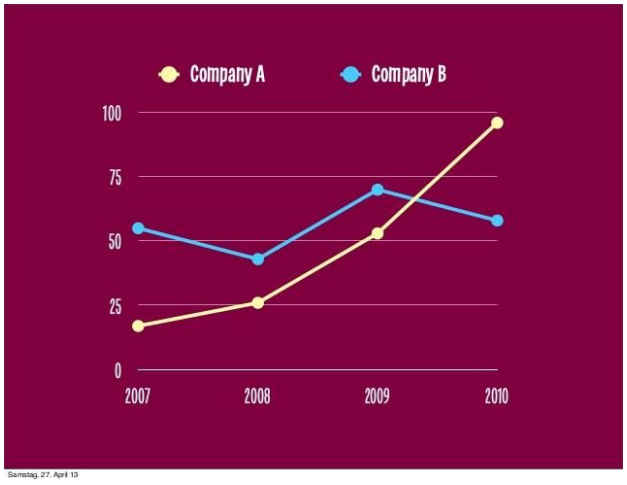
Bad Headline: Resolution Methods for Domestic Violence in Afghanistan

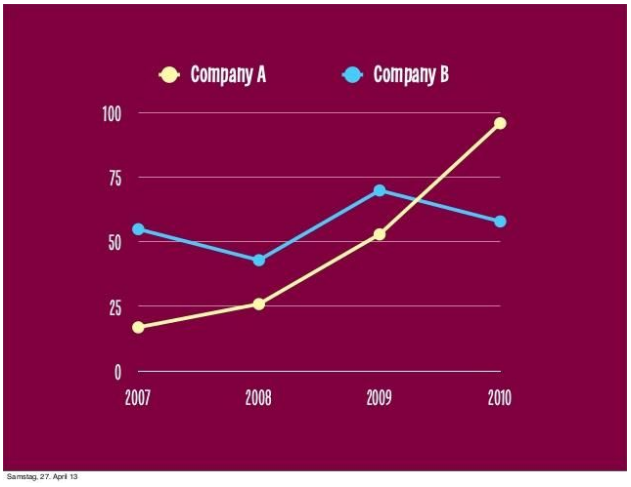
Good Headline:



5. Use Direct Labeling

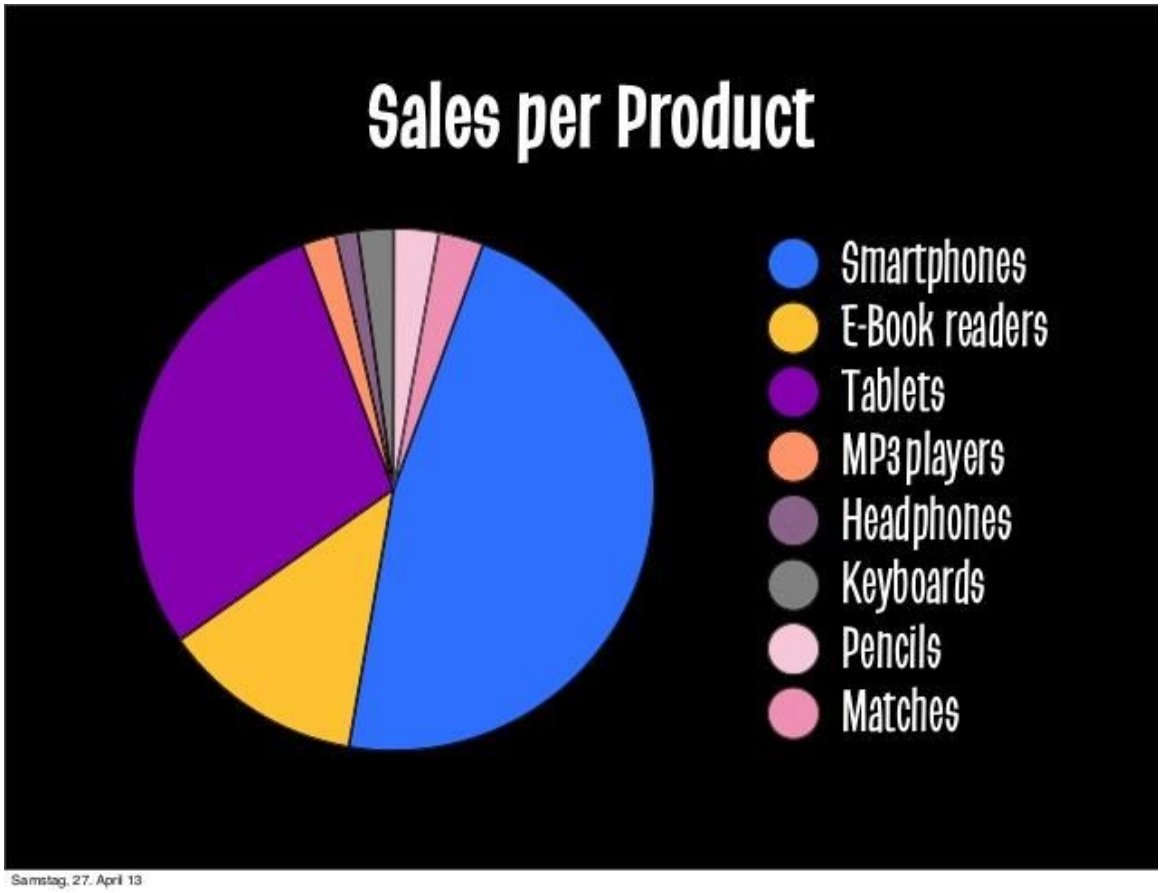
Having to jump back and forth between a key and a graph is difficult for readers. Whenever possible, use direct labeling on visualizations.





Exercise: Redesign a Data Visualisation

Examine the visualisations below and make a list of at least five changes you would make.



1. _____

   _____

2. _____

   _____

3. _____

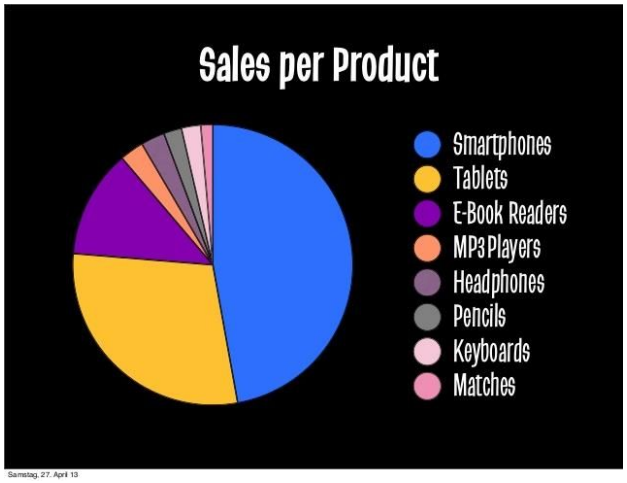
   _____

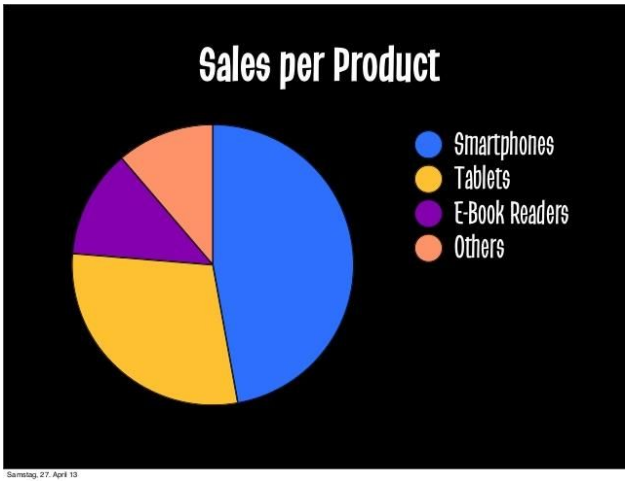
4. _____

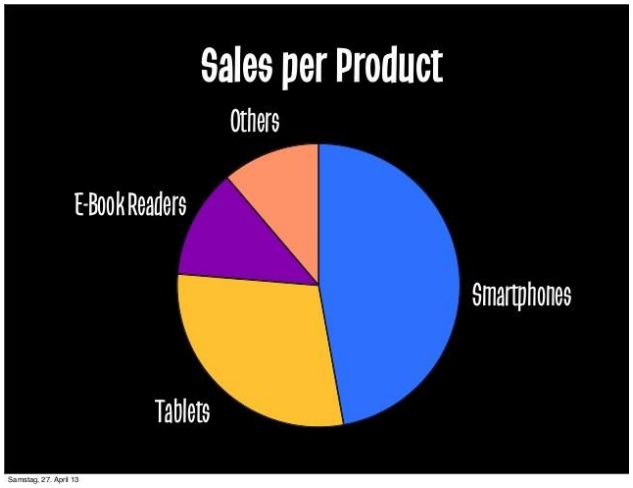
   _____

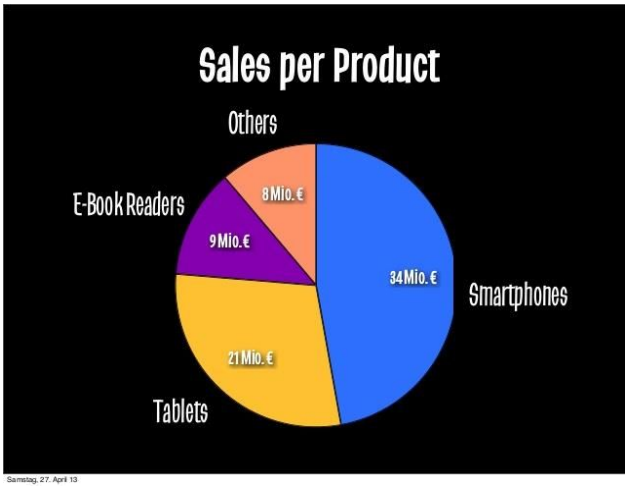
5. _____

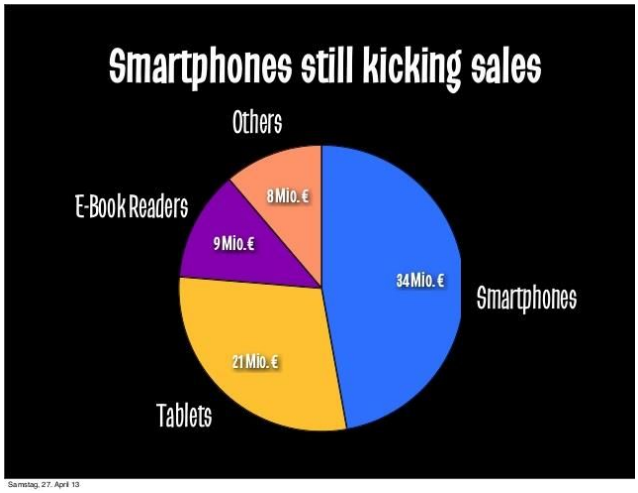
_____

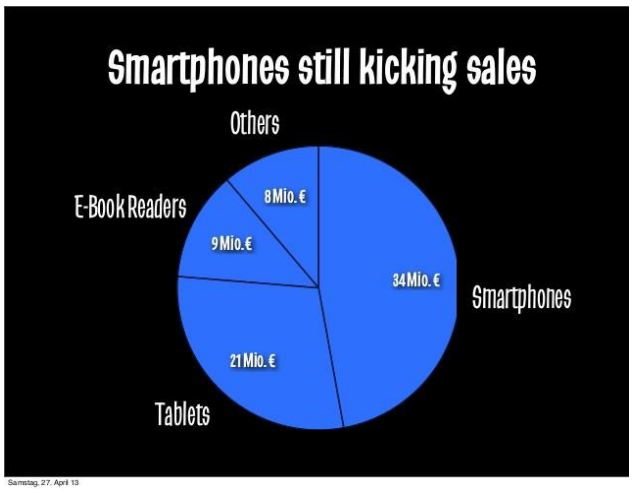

Sales per Product
Smartphones
Tablets
E-Book Readers
MP3 Players
Headphones
Pencils
Keyboards
Matches
Samstag, 27. April 13

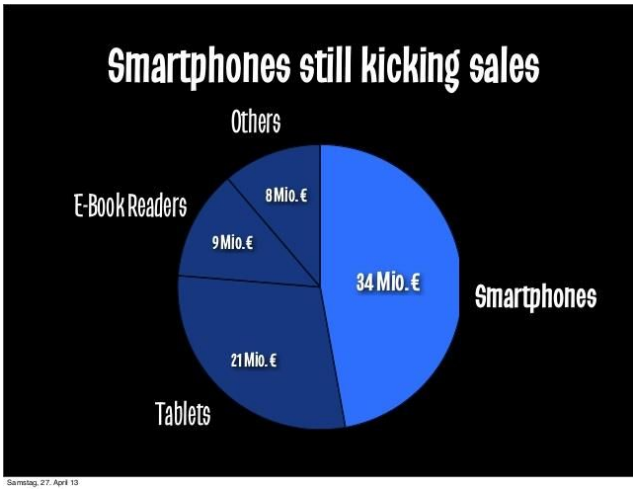

Sales per Product
Smartphones
Tablets
E-Book Readers
Others
Samstag, 27. April 13


Sales per Product
Others
E-Book Readers
Smartphones
Tablets
Samstag, 27. April 13


Sales per Product
Others
E-Book Readers
8 Mio. €
9 Mio. €
34 Mio. €
Smartphones
21 Mio. €
Tablets
Samstag, 27. April 13


Smartphones still kicking sales
Others
E-Book Readers
8 Mio. €
9 Mio. €
34 Mio. €
Smartphones
21 Mio. €
Tablets
Samstag, 27. April 13





At the end of this module, you will be able to:

- understand the design process of the data presentation stage
- choose a visualisation adapted to the needs of the project with the help of existing resources

The presentation stage of your data project is often the most exciting but also one of the most treacherous. Because of the visual aspect of data visualisation, it is easy to develop a feeling of familiarity with it, leading people to forgo best practices regarding their choices of visualisation. But data visualisation is an active field of research, meaning that the answer to « how should I visualise this data » is most of the time not obvious. This module will introduce key data visualisation concepts and best practices, allowing you to use the correct strategy and resources when deciding for a way to present your data.

Lesson 4 : Data presentation vs visualisation

So you've been through all the stages of the data pipeline, and you are ready to show your results to the world. You could, of course, simply click on the « graph » button of your spreadsheet software and be done with it. Or you could find a designer and request a nice infographic. But, more often than you would expect, the best option may be to simply publish an article/press release/small report that includes some figures but without any graph. Or talk about it during a radio

broadcast. This is the first reason why we're talking about presentation and not visualisation: **the end result of your data project does not have to be visual!** What you should focus on is making sure that your results are read and understood by your audience.

The second reason to avoid the word visualisation for this step is that data visualisation can happen all throughout the data pipeline. Consider this example dataset about the regions of Tanzania:

Region	Capital	Districts	Area (km2)	Population (2012)	Density	Postcode	
Dar es Salaam Region	Dar es Salaam	5	1393	4364541	3133.195262	11xxx	C
Mwanza Region	Mwanza	7	9467	2772509	292.860357	33xxx	L
Mbeya Region	Mbeya	7	35954	2,707,410	75.30205262	53xxx	S H
Kagera Region	Bukoba	8	25265	2458023	97.28964971	35xxx	L
Tabora Region	Tabora	7	76150	2291623	30.09353907	45xxx	C
Morogoro Region	Morogoro	7	70624	2218492	31.41272089	67xxx	C
Kigoma Region	Kigoma	8	37040	2127930	57.44951404	47xxx	W
Dodoma Region	Dodoma	7	41311	2083588	50.43663915	41xxx	C
Tanga Region	Tanga	10	26667	2045205	76.69422882	21xxx	N
Mara	Musoma	7	21760	1743830	80.13924632	31xxx	L

FIG1: data table of the Tanzania Regions
source: Wikipedia

To quickly verify the data, I can visualise it, in order to highlight potential anomalies

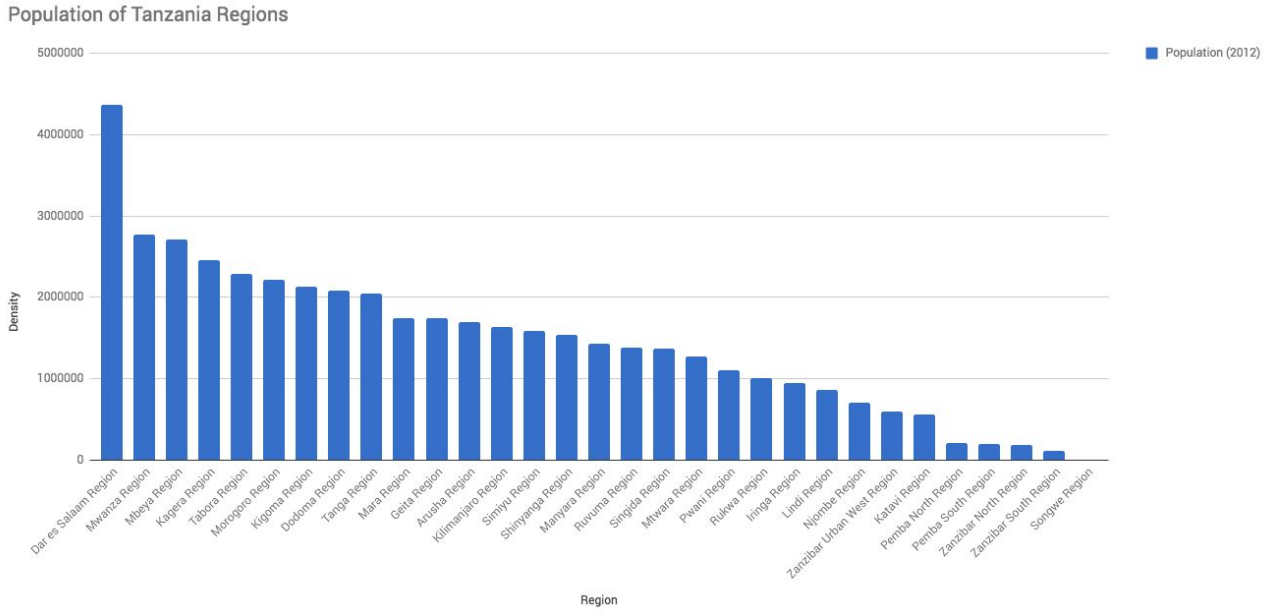


FIG2: Bar chart of the population of Tanzania regions

The Songwe region probably has more than 0 inhabitants. Missing data is a common problem, and visualisation can help spot that; this is especially useful for bigger datasets.

But say that I initially wanted to use this data to analyse where funds for urban development should be allocated. My idea is that the funds should go in priority to regions with high population density, which is where having well developed cities becomes critical. As part of my analysis, I will then calculate and visualise the density of each region (population / region area):

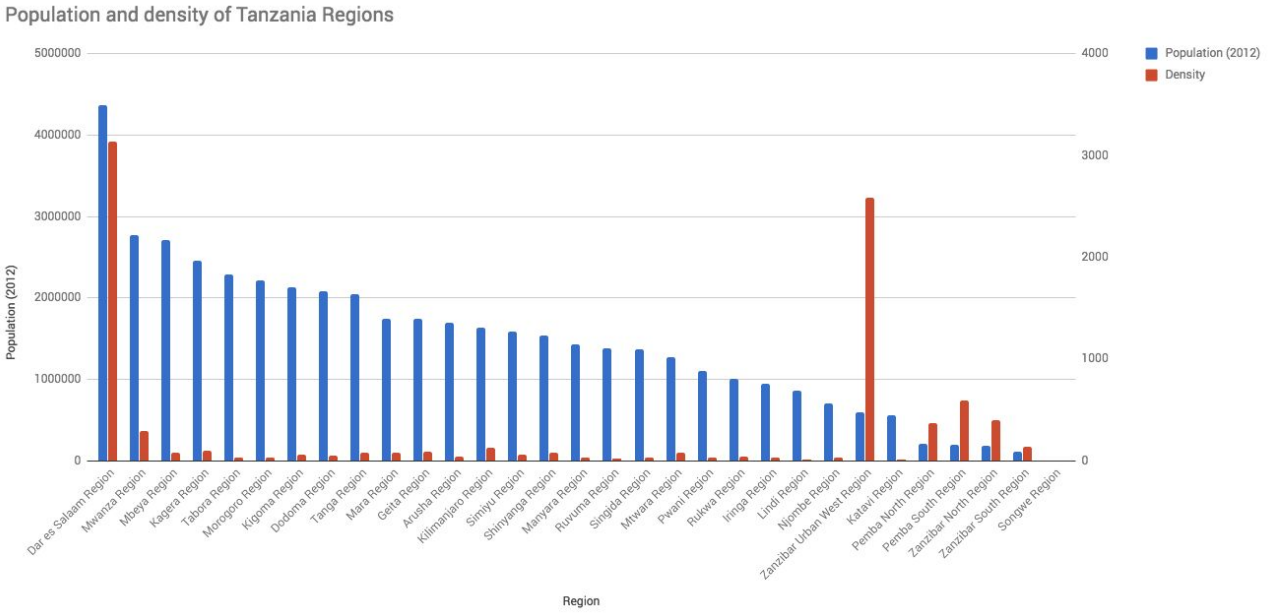


FIG3: bar chart of the population density of Tanzania regions

The chart confirms what I expected: the Dar es Salaam and Zanzibar regions are where people are the most concentrated. But even if they trail quite far behind, I should also look at the Pemba and Mwanza regions.

As part of this project, we have already visualised our data two times without even reaching the presentation stage! So our takeaway is that data visualisation doesn't come only at the end, and data projects do not always need a visual final deliverable.

Why does everyone insist in using data visualisations to show their results then? Because, as mentioned before, there is a feeling of familiarity and accessibility associated with visual elements. Additionally there is a certain authority to numbers, and inserting data visualisations can be a powerful way to strengthen your argument in the mind of your readers. But it can quickly become an exercise in vanity.

What's important here is to communicate accurately the insights you gained from working with the data. Accurately doesn't necessarily mean that you have to talk about everything: your audience doesn't necessarily care about everything. But you have the responsibility to reflect what the data shows, and not what you want

the data to show. This is a fine line, and you have to be careful about it throughout the data pipeline process.

Lesson 5 : Choosing the right presentation

Choosing the format of your presentation gets into the question of why we visualise data.

There are 2 main goals that you may want to pursue:

- Allow yourself and/or colleagues to explore further the data. Especially useful when there are not clear priority insights to extract from the data, or when you don't want to draw conclusions by yourself on the meaning of the data. In this case you will create an **exploratory visualisation**.
- Share with an audience the key insights from your data project. In this case your presentation will be more **explanatory**. Explanatory presentations do not need to be visual: you can explain very well the results of the data project via a radio broadcast or full-text newspaper article.

Exploratory visualisations are often more complex, because they tend to present a variety of parameters, allowing the reader to go into details to find the part of the data or the specific trends that are of interest for them.

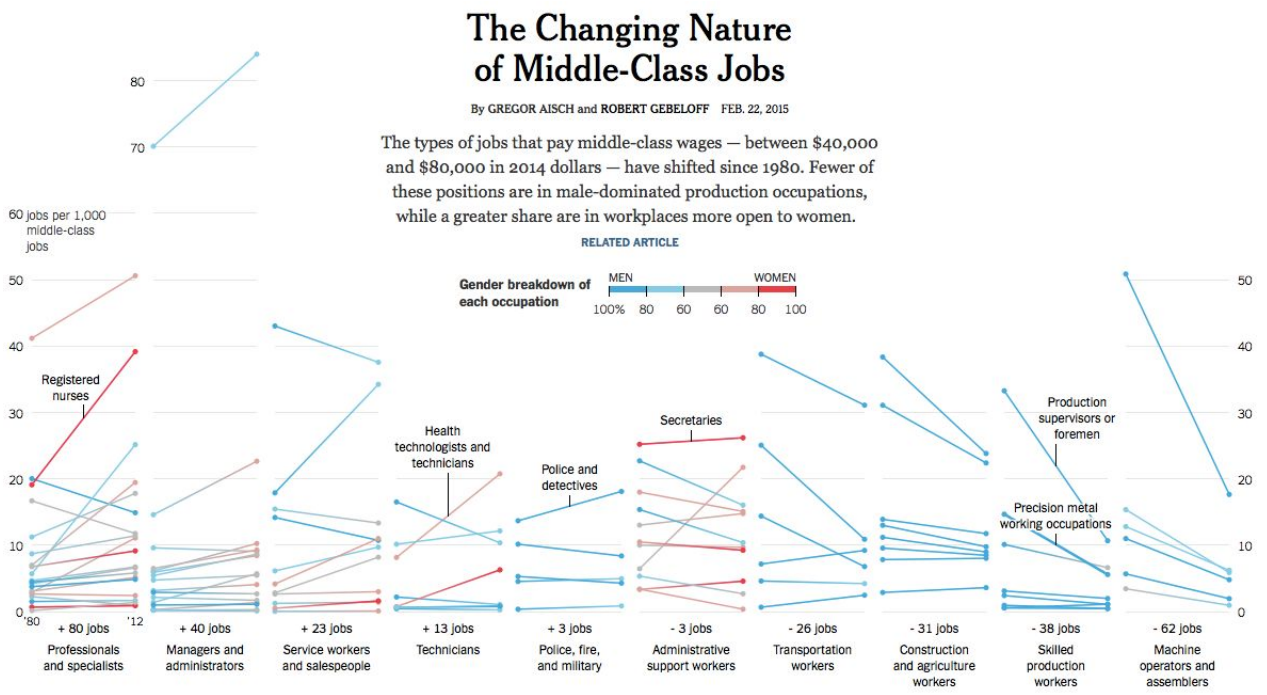


FIG4: exploratory visualisation example
source: The New York Times

Exploratory visualisations are less common, but become crucial when you want to enrich a discussion by pulling the relevant data. Instead of having to sift through huge tables of figures, the people involved in the discussion (of a new public policy, for example) can draw their own conclusions and exchange views quickly by looking at the visualisation.

While a couple of popular charts are specifically useful for explanatory analysis, such as the parallel coordinates chart, exploratory visualisations often incorporate multiple types of charts into one in order to allow the investigation of several parameters. Maps are well suited for that, as they allow to stack different types of visual representations.

FIG5: parallel coordinates chart about car engines and models and map with multiple variables

When the same type of chart is used multiple times but for several variables, we talk about a « small multiple » chart. Depending on its complexity, this chart can be used for both explanatory and exploratory purposes.

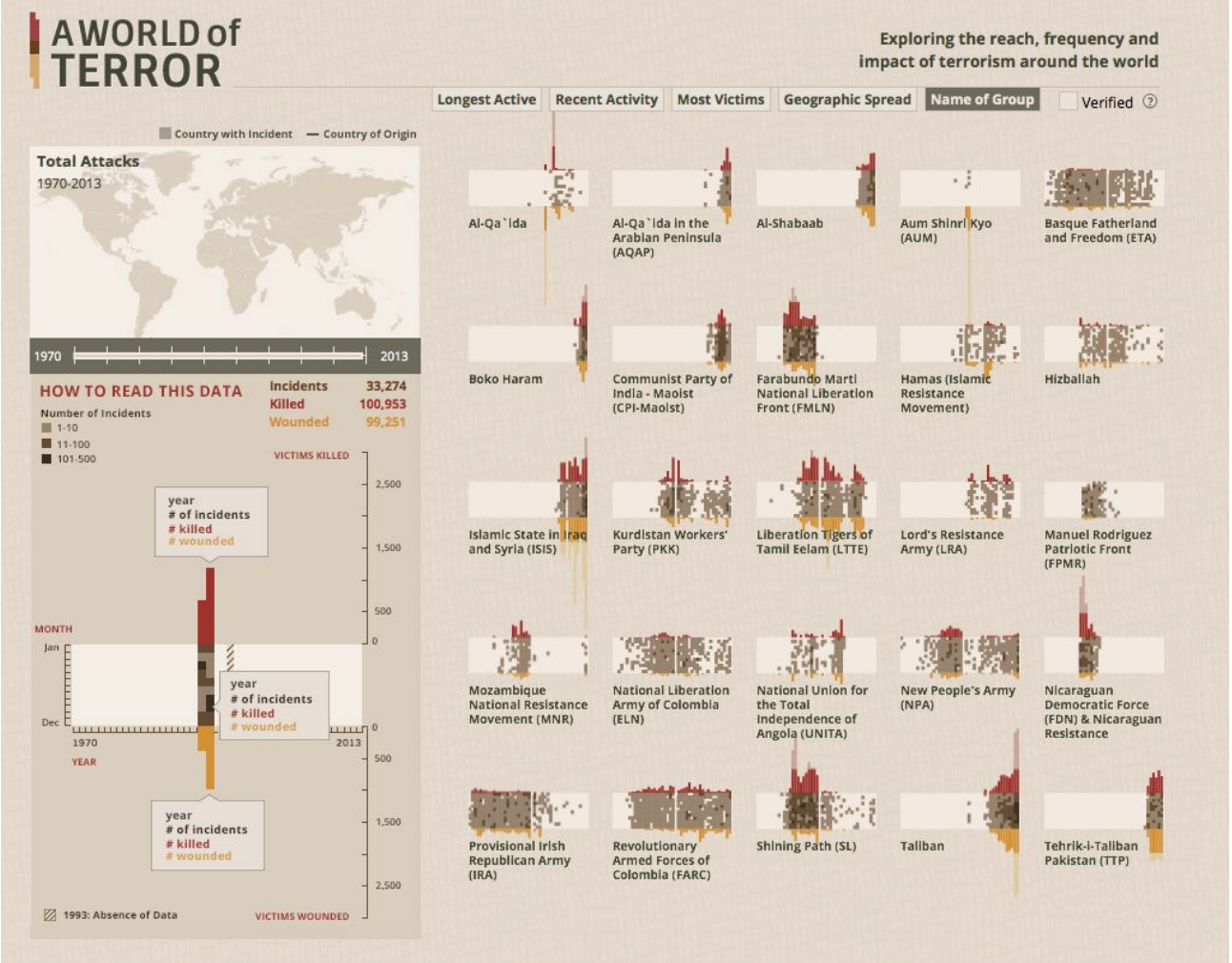


FIG6: small multiples chart
source: Global Terrorism Database

The development of internet and web browsers allow a new dimension to exploratory analysis: interactivity. From simple knobs that allow you to see how a graph evolves if you change a parameter to more complex interactions, it is now possible to give much more freedom to the reader to explore a dataset visually.

Explanatory presentations give you more latitude in your choice of presentation:

Visual presentation	Non visual presentation
Graphs/charts, physical visualisations,	text articles, radio broadcasts

interactive visualisations, static and animated infographics	

In order to pick the correct presentation, you have first to define:

- who your audience is
- their level of literacy (someone who struggles to read will understand infographics better than a graph, but using radio broadcasts may be even more relevant for this type of audience)
- their preferred mode of consuming information (maybe an animated infographic on TV is the best way to reach them)

For examples of text articles talking about data without using visualisations, check out websites such as Fivethirtyeight or the New York Time's blog The Upshot.

Lesson 6 : Choosing and designing your data visualisation

Data visualisations, while not the be-all-end-all of data presentation, still play a major role and are a key format to have knowledge of when working with data. You are probably familiar with what some researchers call the « Three Amigos »: the line chart, bar chart and pie chart. Those are the three most basic (and consequently most readable) visualisation formats, and they can be useful for the majority of datasets. They are also widely available as presets in common data processing tools such as spreadsheet software (Excel, Google Spreadsheets, LibreOffice…), which means that people default to them. But your choice is far from limited to those three options:

FIG7+: Screenshot of the Dataviz Catalogue

source: datavizcatalogue.com

Two key questions will allow you to find the visualisation adapted to your need:

- Of what type is the data that I want to visualise ?
- What do I want to highlight in my data?

Data comes in a number of different types, which determine what kinds of mapping can be used for them. The most basic distinction is that between continuous (or quantitative) and categorical data, which has a profound impact on the types of visualisations that can be used.

Quantitative data is data where the values can change continuously, and you cannot count the number of different values. Examples include weight, price, profits, counts, etc. Basically, anything you can measure or count is quantitative. **Categorical data**, in contrast, is for those aspects of your data where you make a distinction between different groups, and where you typically can list a small number of categories. This includes product type, gender, age group, etc.

If we go back to our earlier example about Tanzania regions, the name of the regions are categorical data while the population of each region is quantitative data.

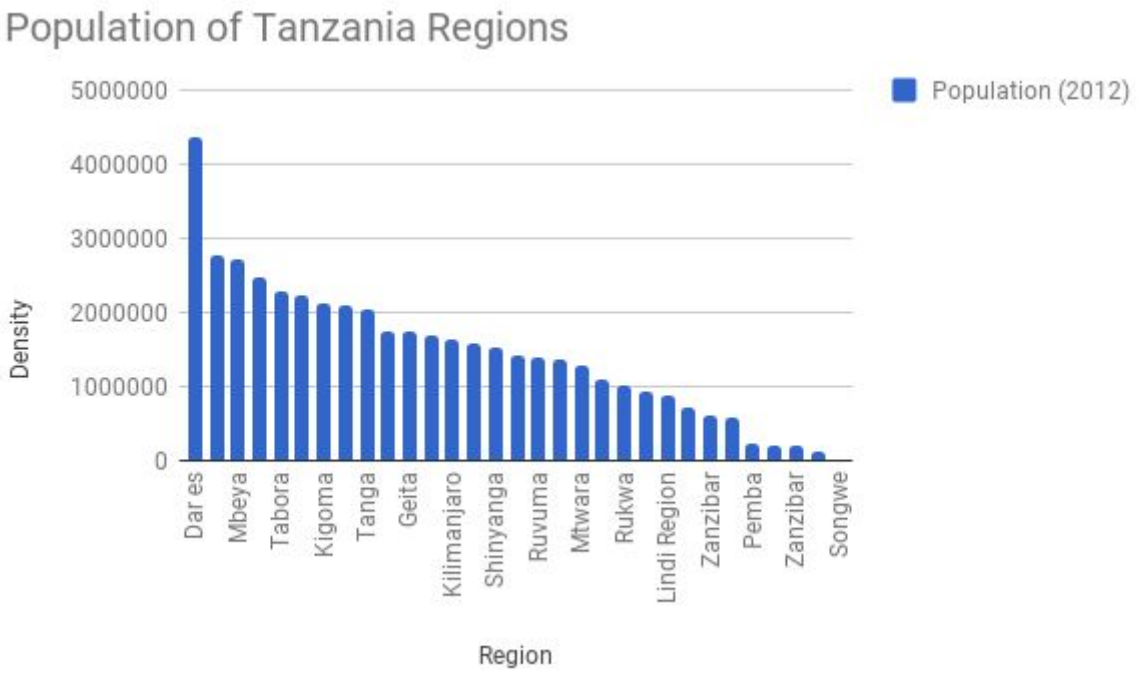


FIG2: Bar chart of the population of Tanzania regions

Because the regions are categorical data, it would be confusing to use a line chart, for example, because the line implies a continuity between the values. A line chart would be more suited to visualise the evolution of the population of Dar es Salaam over time, as time can be viewed as a continuous value.

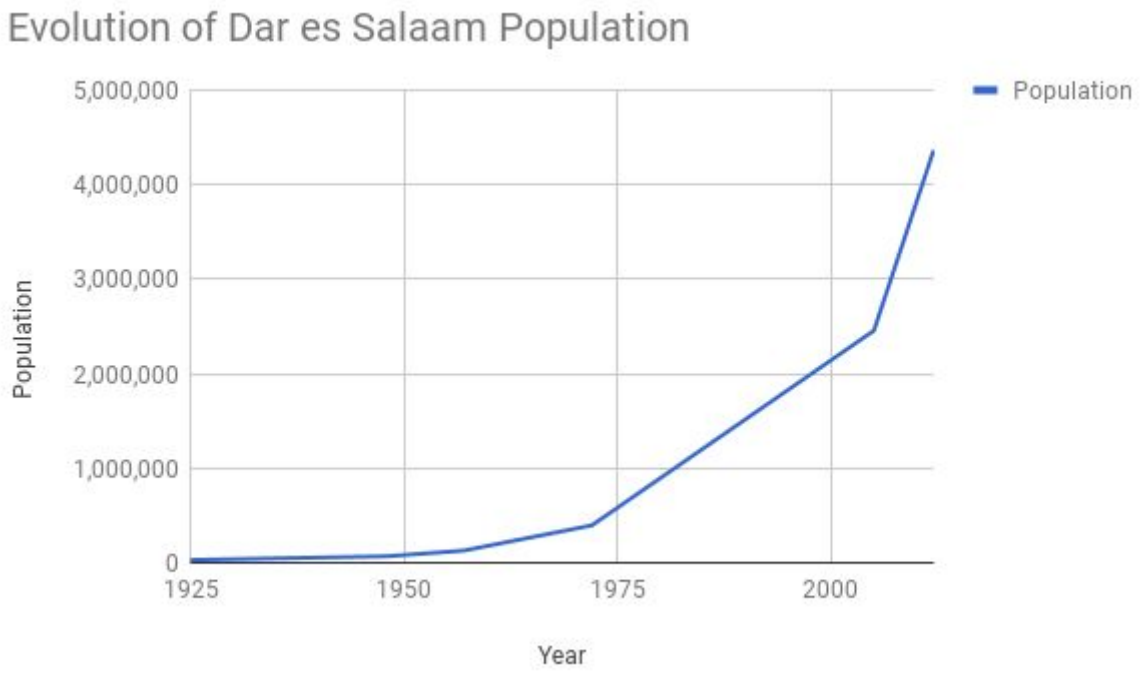


FIG8+: line chart of the evolution of Dar es Salaam's population
source: Wikipedia

Data types play an important role in visualisation because they determine what visualisation types can or should be used. That doesn't mean that there is only one chart for any combination of data types, but it does narrow down the possibilities.

But the type of data is not the end the story: you now have to decide what you want to highlight in the visualisation. The most common categories are the following:

- Comparison
- Distribution
- Composition
- Trend
- Relationship

Because no one expects you to know all possible visualisation by heart, here are two useful resources to help you pick and discover visualisations:

The Datavisualisation Catalogue lists a lot of common and uncommon visualisations with a short but precise description of their purpose and characteristics.

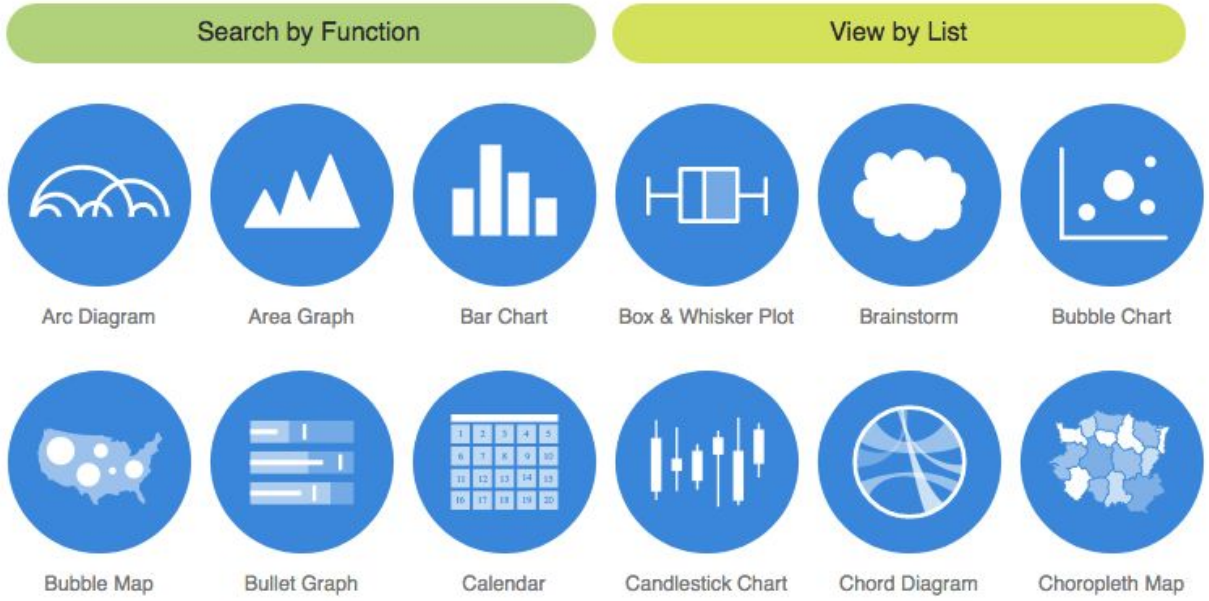


FIG9+: Screenshot of the Dataviz Catalogue

The Chart Chooser is an interactive website which allows you to filter a list of charts based on the category of visualisation you are interested in:

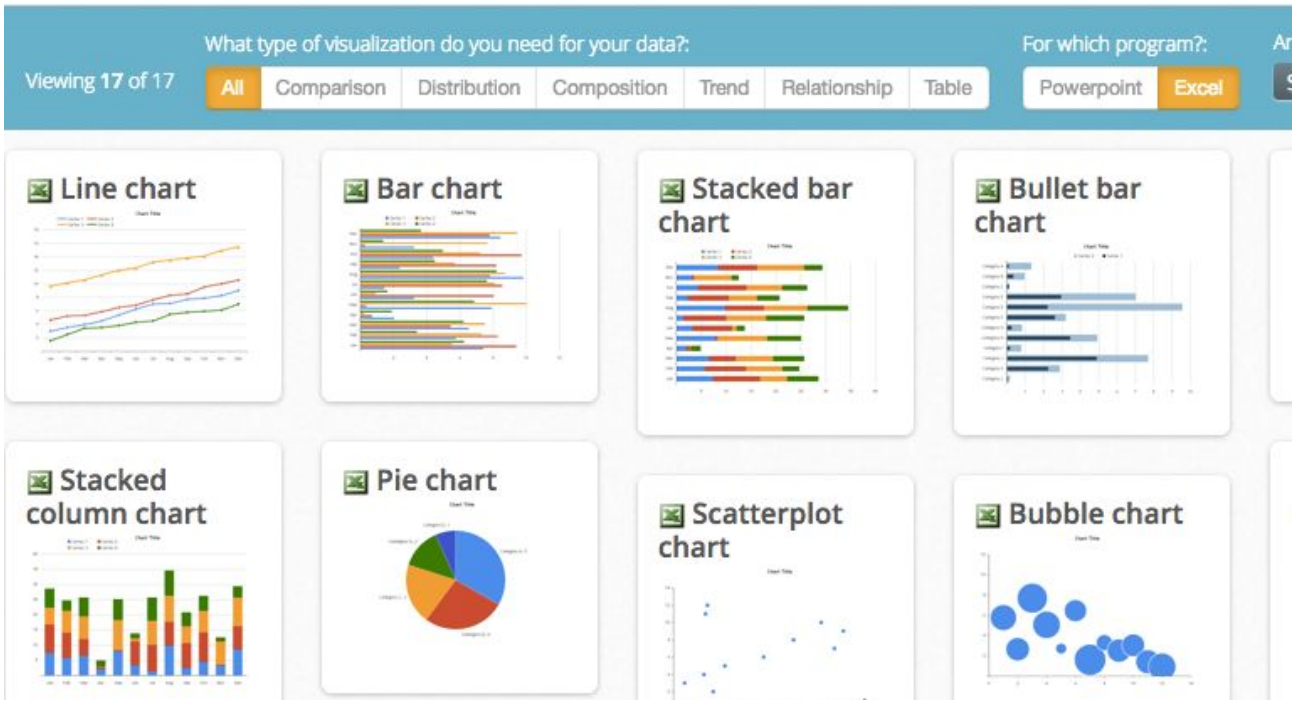


FIG10+: Screenshot of the chart chooser

If all of this sounds daunting, it's normal: Data visualisation is an active field of research, and a lot of parameters are taken into account when designing a visualisation. But there is no perfect answer to the question « what is the best visualisation for my data? », and you will often find that choosing an option over another boils down to personal preference or instinct. Which is why it's important to have at least a basic understanding of what works and what doesn't, in order to avoid always defaulting to the « Three Amigos » (or worse, using a fancy but nonsensical visualisation) regardless of their fit to the data.

The last important aspect of the design process of your visualisation is the choice of colour.

Colour is important because it plays a major role into the legibility of your visualisation and also helps guide the viewer's eye toward the key parts of your visualisation. People also naturally associate some colours with specific concepts (red = danger/hot, blue = cold/sea, green = nature/money), which can be confusing if it is not taken into account when designing the visualisation. Lastly, colour is also a question of accessibility: a substantial portion of the population is

colour blind, which means that they may not differentiate some colours from each other.

In the context of an organisation, it is advised to work with a designer to set up a colour scheme to be used across data visualisations. Fortunately, there are also tools for non-colour specialists which are handy when designing a visualisation. The main one is ColorBrewer2, which, while focused on cartography, is used by all data visualisation designers to choose colours because of its wealth of options: you can choose sets which are adapted for colour blindness; gradients for showing gradual change; or sets of contrasting colours for categorical data.

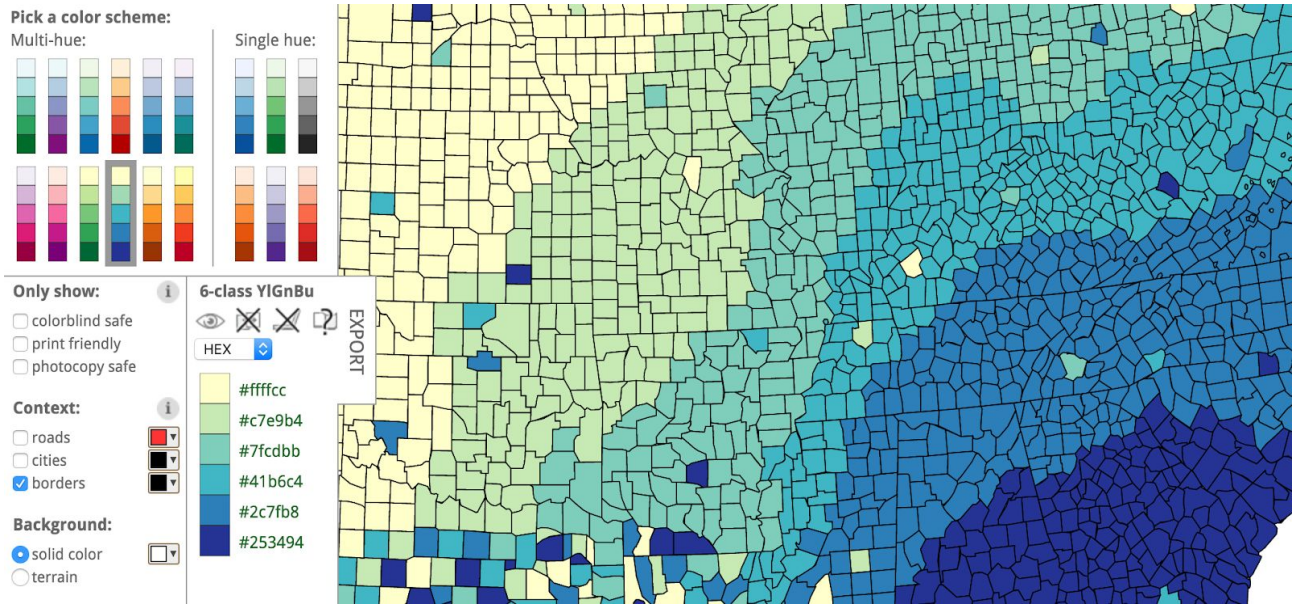


FIG11+ Screenshot of ColorBrewer2