Lecturer Series on Data Science (LSDS 2018)

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



اَلْسَلَامْ عَلَيْكُمْ

Table of contents

Lecture 1: Introduction

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Definition of Data

- Data is facts and statistics collected together for refence or nalysis.
- the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
- things known or assumed as facts, making the basis of reasoning or calculation.

Source: Oxford Dictionary English

Definition of Science

- The intellectual and practical activity encompassing the sytematic study of the structure and behavior of the physical and natural world through observation and experiment.

Source: Oxford Dictionary English

Definition of Scientific Method

- Method of procedure that has characterized natural science since the 17th century, consisting in systematic observation, measurement, and experiment, and the formulation, testing, and modification of hypotheses.

Source: Oxford Dictionary English

Definition of Data

- Information, usually numerical or categorical.

Source: Oxford Dictionary of Statistics

Definition of Data Science

- As a field, data science is hard to define, so it is easier to consider this through the ingredients of the role of data scientists. They posses a broad repertoire of capabilities covering the gathering, handling and analysing data. Typically this data is of a large size and complexity and originates from multiple sources.

  Data scientist will have strong mathematical, statistical and computer science skills, not to mention astute business experience and many notable 'softer' skills like problem solving, communication and presentation. If you find somebody with all these skills, tie them to a desk (legally) and never ever let them leave your organisation.

Source: Data Visualisation: A Handbook for Data Driven Design

Lecture 2: Learning about Data

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Discrete Variable

A random variable whose set of possible values is a finite or infinite sequence of numbers $x_1, x_2, \ldots$. The set of values is often a subset of the non-negative integers $0, 1, \ldots$. The probability distribution of a discrete random variable is referred to as a discrete distribution.

- Discrete distribution. Poisson, binomial, bernoulli etc.
- Discrete random variables.
- Issues with discrete data. Histogram or barplot?
- Examples.

Continuous Variable

A variable whose set of possible values is a continuous interval of real numbers x, such that $a < x < b$, in which a can be $-\infty$ and b can be ∞.

- Continuous distribution. Normal, t , F, χ^2 etc.
- Continuous random variables.
- Issues with continous data. Treated continous as discrete.
- Examples.

Circular data

A type of cyclic data in which measurements are directions. Axial data ocur when a data item has known orientation but its direction is not known - for example, an iron filing will align itself either way around in a magnetic field. In this case the effective orientations lie in 180^o range. To utilize the standard methods for cyclic data it is usual to double all the angles (substracting 360^o if required). Other example is wind & current data.

Spherical data - 3D data.

Dirty data

Data with unusual, sparse, missing, or incorrect values. Most real sets of data are dirty.

Extreme data

Data are either maximum or minimum values.

Categorical data

Classification of/from real data. Table form data.

Temporal data

Data that are related to time.

Longitudinal data

Data collected at sequence of time points for each of a sample of individuals. Because each individual contributes several observations, these observations are usually not independent and this has to be taken into account in the analysis.

Likert scale

A scale, usually of approval or agreement, used in questionnaires. The respondent is asked to say whether, for example, they ’ Strongly agree’, ’Agree’, ’Disagree’, or ’Strongly disagree’ with some statement.

Ordinal variable - A categorical variable in which the categories have an obvious order (e.g. strongly disagree, diagree, neutral, agree, strongly agree). Specialist models are required for ordinal variiables to take account of their ordered categories)

UPM UNIVERSITI PUTRA MALAYSIA BERILMU BERBAKTI

Grouped data

The frequencies of occurrence of values in specified intervals. Histograms and frequency polygons are used to illustrate grouped data.

Index number

A measure of the value of variable relative to its value at some base date or state (the base period). The index is often scaled so that its base value is 100. Such an index may be described as a base-weighted index. Eg. Price Index.

Transformation

A function of a random variable (or of a statistics). Examples are $\sqrt{X}$ or $\log(X)$, where X is the variable. Using the transformed variable may, for example, simplify as model or stabilize the variance.

- Pre-process
- Reformulate (or indexing)
- Score/rank
- Re-units or Scaling
- Others

Univariate, Bivariate and Multivariate

Bivariate data is the the data that consist of pairs of values $(x_1, y_{)}, (x_2, y_2), \ldots$ taken from a bivariate distribution. The term bivariate dates from an article published in 1920.

- Response and explanatory variables
- Marginal, Joint and conditional terms

Model

Model: A simple description of a probabilistic process that may have given rise to observed data. For example, if the data consist of the numbers shown by a fair die during a game of Snakes and Ladders, then a simple model would state that for each roll, and independent of the outcomes of other roll, the distribution of the number shown is a discrete uniform, on 1, 2 , 3, 4, 5, & 6.

Models form the bedrock of Statistics. Specific distributions are often invoked. Ther are many types of model can be found.

Lecture 3: Show Me the Numbers: Designing Tables and Graphs to Enlighten [Few, S. (2012) Analytics Press.]

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Your are given 6 Scenarios to response

- Table or graph?
- If a table, which kind?
- If a graph, what kind of relationship?
- If a graph, which graphical objects for quantitative encoding?
- Anything else?

Lecture 4: Exploratory Data Analysis - EDA

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018

Lecture 5: The Truthful Art: Data, Charts, and Maps for Communication, Cairo, A. (2016)

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Source: Cairo, A. (2016). The Truthful Art: Data, Charts, and Maps for Communication. New Riders, US.

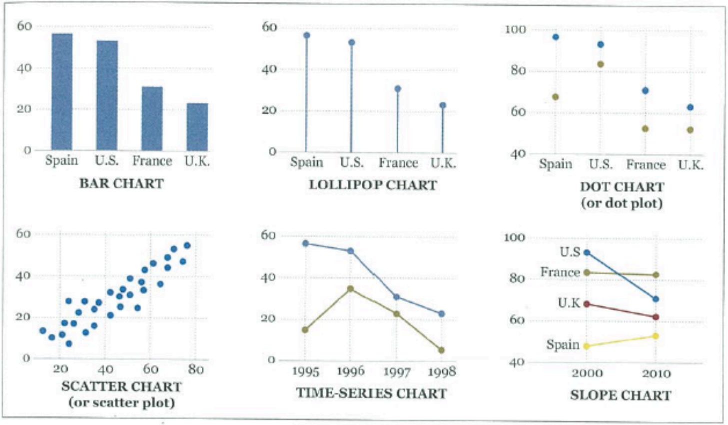
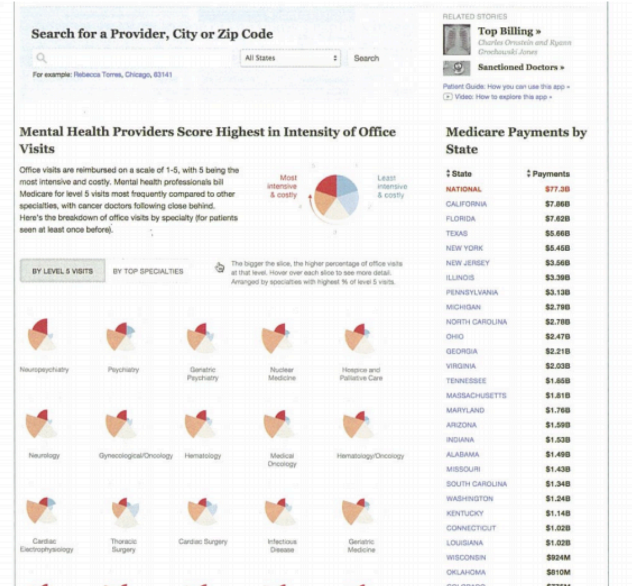


Figure 1.1 Examples of charts. Not all charts have an X-axis and a Y-axis. Pie charts, for instance, aren't based on a Cartesian coordinate system.

The Truthful Art: Data, Charts, and Maps for Communication.



THE TRUTHFUL ART: DATA, CHARTS, AND MAPS FOR COMMUNICATION.

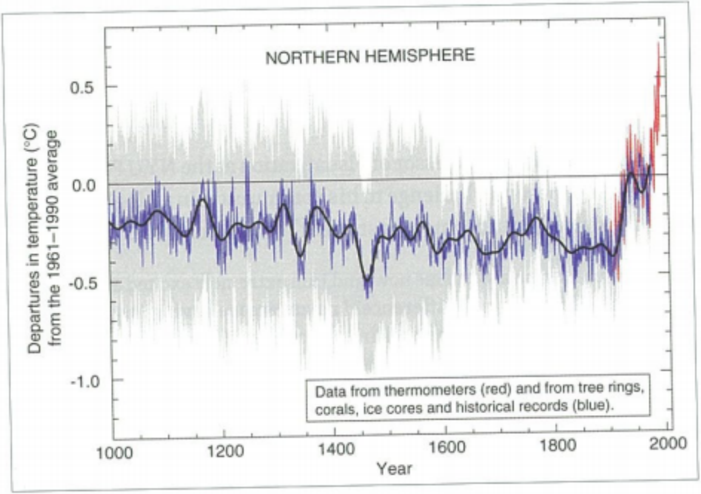


Figure 2.1 The hockey stick chart. Summary For Policymakers of the 2001 Third Assessment Report of the Intergovernmental Panel on Climate Change.

UPM UNIVERSITI PUTRA MALAYSIA BERILMU BERBAKTI

The Truthful Art: Data, Charts, and Maps for Communication.

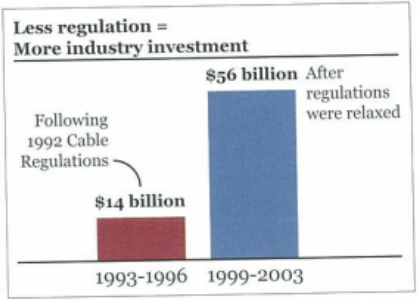


Figure 2.2 In 10 seconds, tell me everything that looks fishy in this chart.

The Truthful Art: Data, Charts, and Maps for Communication.

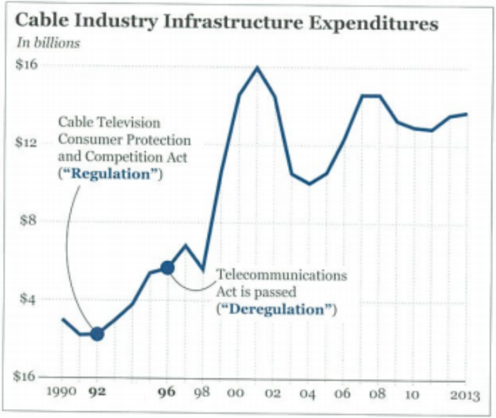


Figure 2.3 Do government regulations really hinder investment in cable infrastructure?

The Truthful Art: Data, Charts, and Maps for Communication.

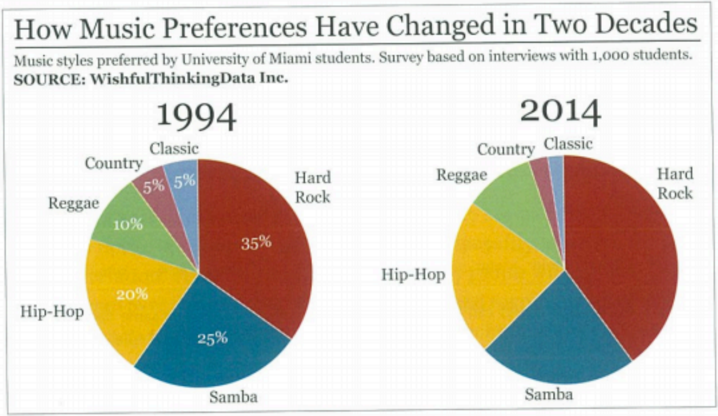


Figure 2.6 Notice the source of the graphic. Quite a bit of wishful thinking on my part was involved in making up the data, as I like hard rock.

The Truthful Art: Data, Charts, and Maps for Communication. .

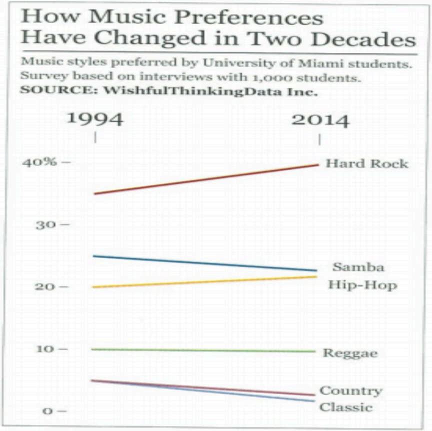


Figure 2.7 A slope chart is much better to represent change between two points in time.

The Truthful Art: Data, Charts, and Maps for Communication.

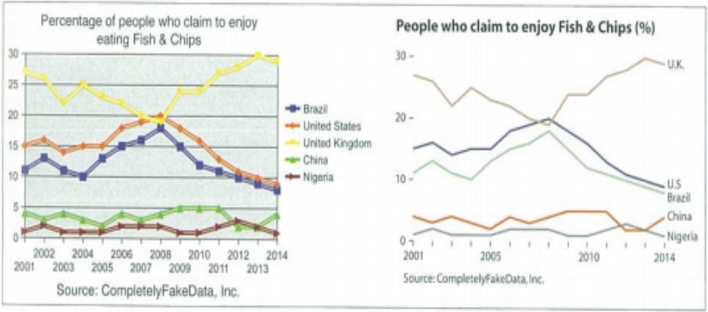


Figure 2.11 Which chart is more aesthetically pleasing?

The Truthful Art: Data, Charts, and Maps for Communication.

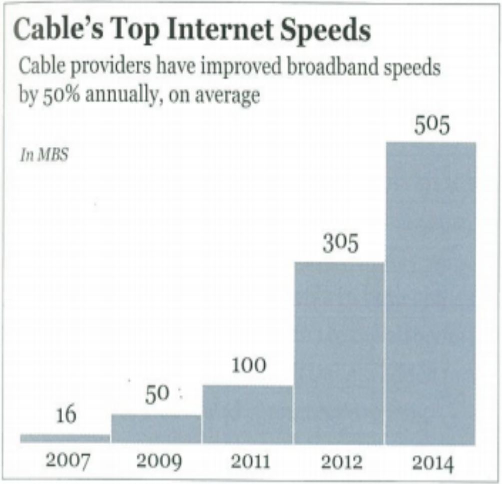


Figure 3.1 Misleading your audience may yield benefits in the short term. In the long term, however, it may destroy your credibility.

The Truthful Art: Data, Charts, and Maps for Communication.

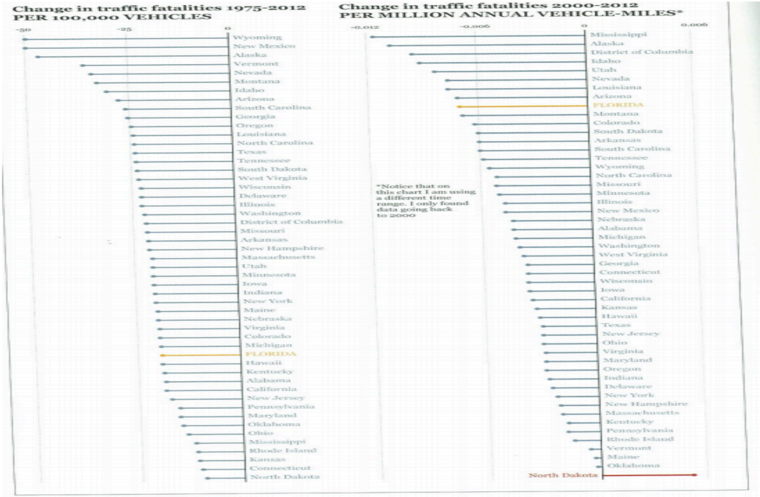


Figure 3.3 When we apply some controls, the results are quite different.

The Truthful Art: Data, Charts, and Maps for Communication.

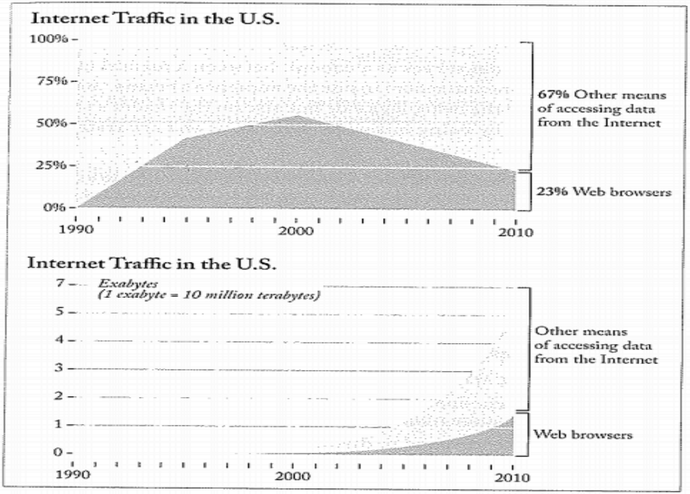


Figure 3.4 Two versions of the same story. (Sources: *Wired* magazine and BoingBoing.net.)

The Truthful Art: Data, Charts, and Maps for Communication.

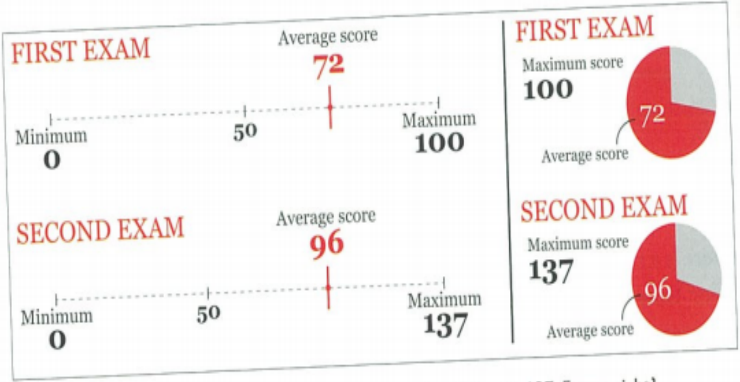


Figure 5.1 Seventy-two over 100 is a better score than 96 over 137. Funny, right?

The Truthful Art: Data, Charts, and Maps for Communication.

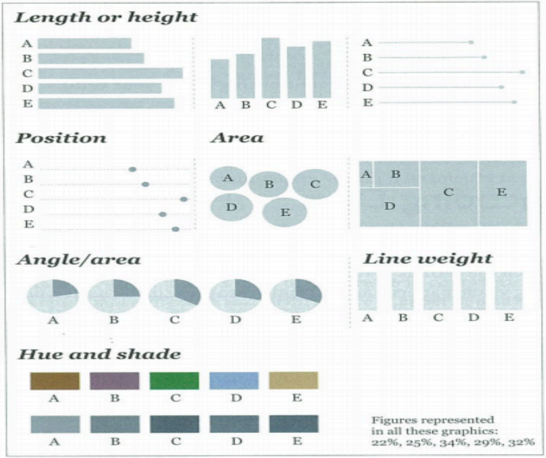


Figure 5.2 Different methods of encoding the same small data set. Remember that, perhaps because our client requested it, countries are organized alphabetically. Otherwise, it'd make more sense to arrange the figures from largest to smallest.

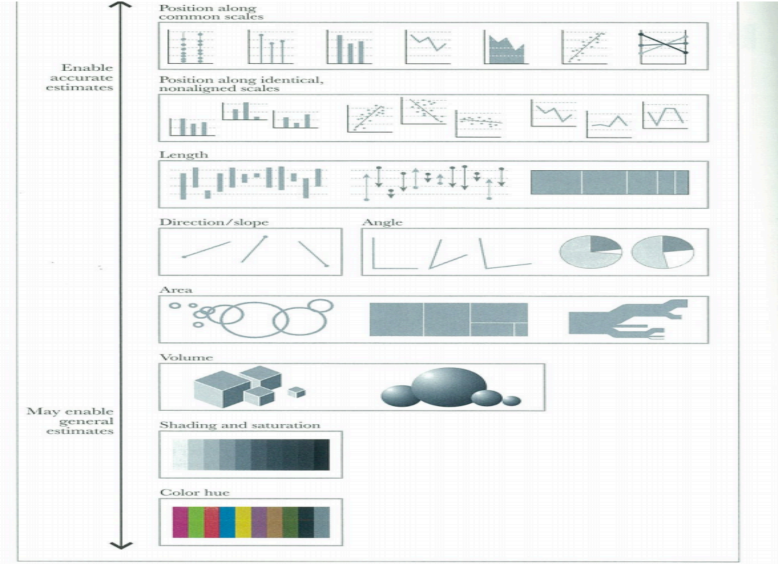


Figure 5.5 Scale of elementary perceptual tasks, inspired by William Cleveland and Robert McGill.

The Truthful Art: Data, Charts, and Maps for Communication.

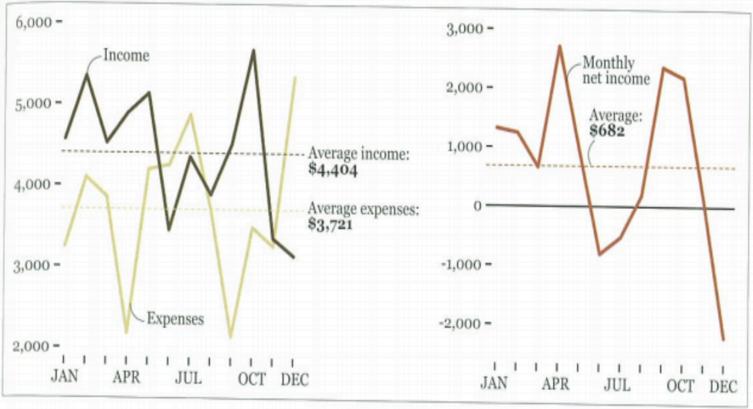


Figure 5.10 Which chart is better? It all depends on if you want to emphasize income versus expenses or if you wish to display the monthly net income.

The Truthful Art: Data, Charts, and Maps for Communication.

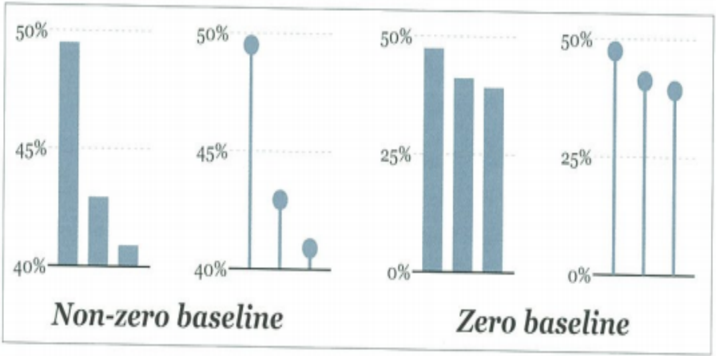


Figure 5.11 Don't truncate the Y-axis in bar charts and lollipop charts.

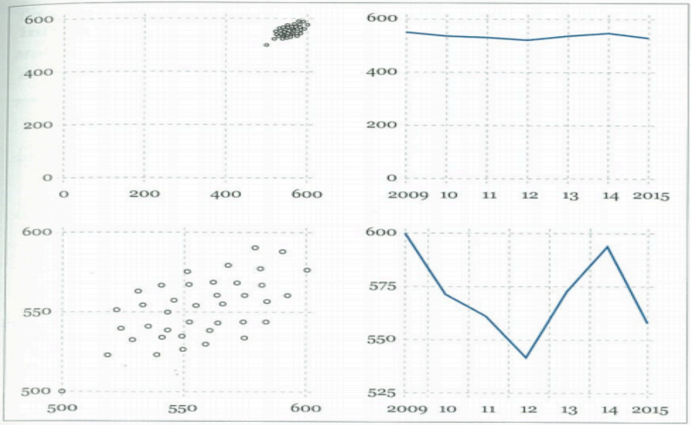


Figure 5.13 Are 0-baselines always necessary?

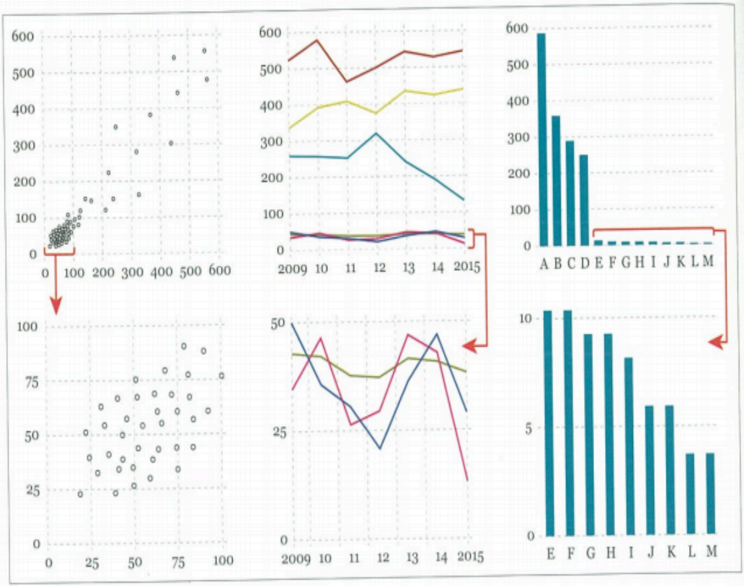


Figure 5.14 Two different scales for subsets of the same data.

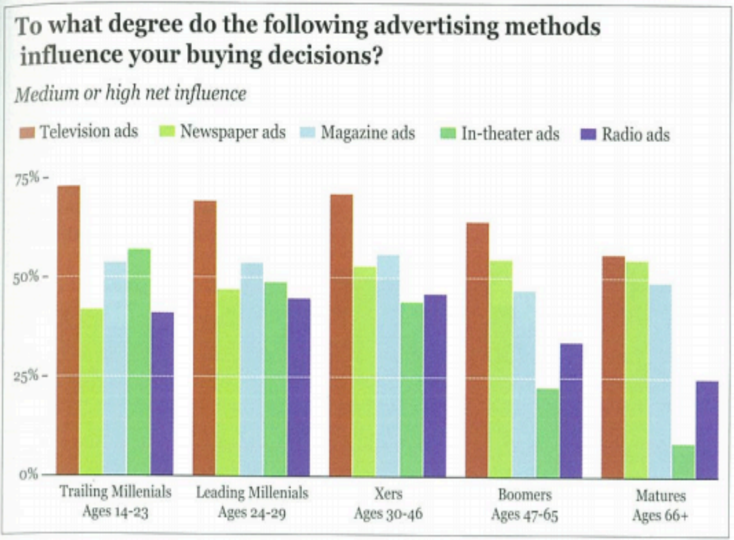


Figure 5.15 Data source: Deloitte's Digital Democracy Survey.

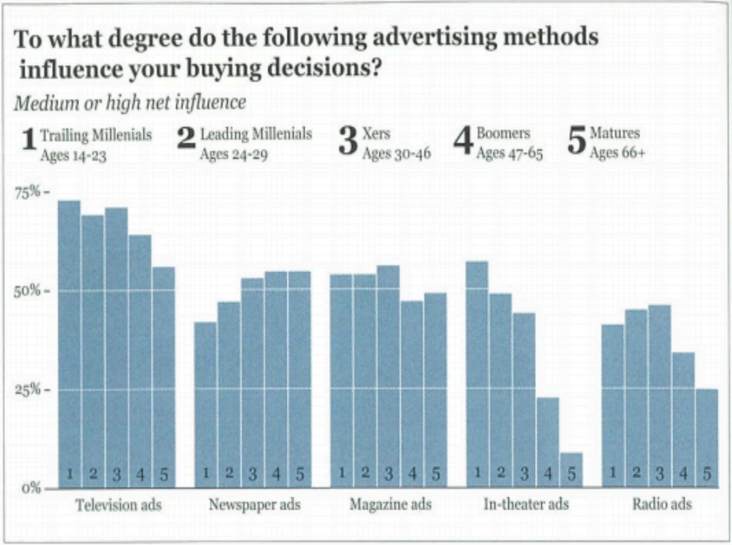


Figure 5.16 Reorganizing the data from Figure 5.15.

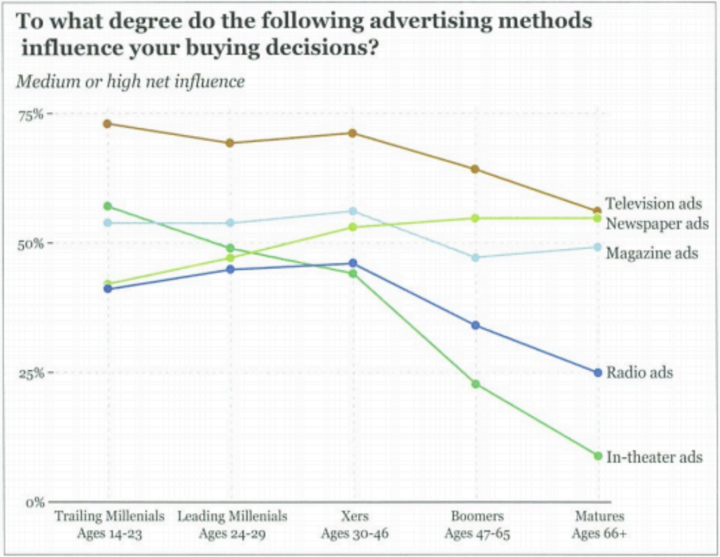


Figure 5.17 Line chart with the same data used in Figure 5.15.

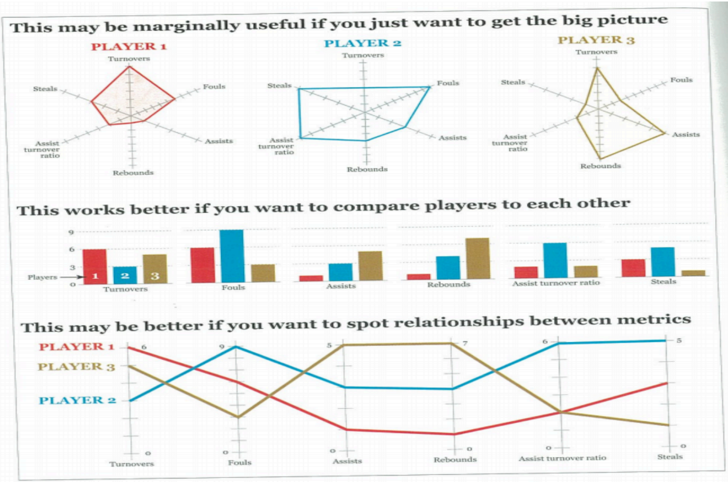


Figure 5.19 Radar charts aren't usually very effective. These are all fake charts, by the way.

The Truthful Art: Data, Charts, and Maps for Communication.

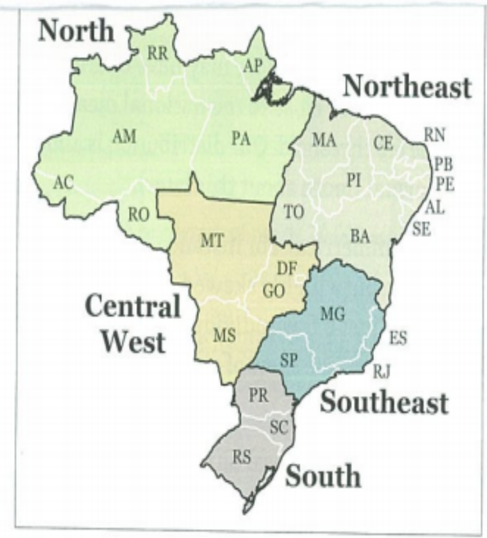


Figure 6.8 Map of Brazil. States in the north and the northeast are much poorer than those in the south and southeast.

The Truthful Art: Data, Charts, and Maps for Communication.

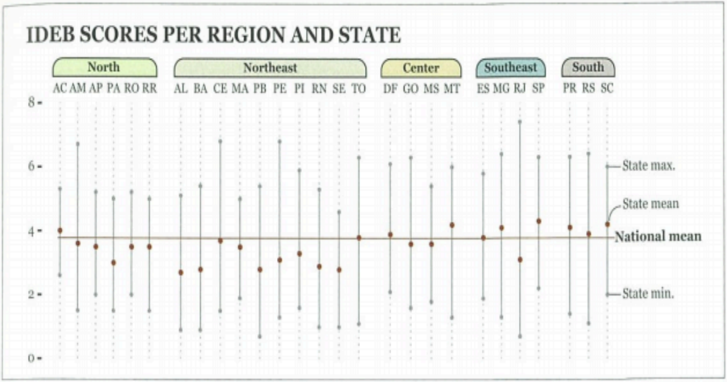


Figure 6.9 Summary data, state by state.

The Truthful Art: Data, Charts, and Maps for Communication.

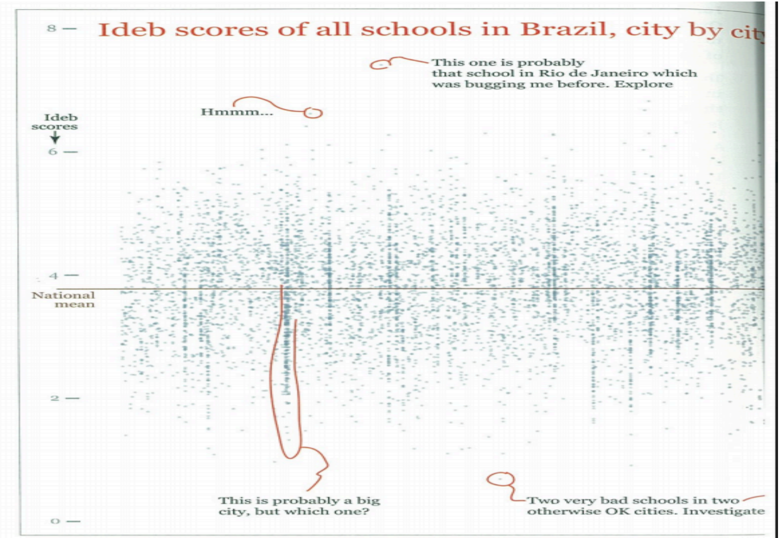
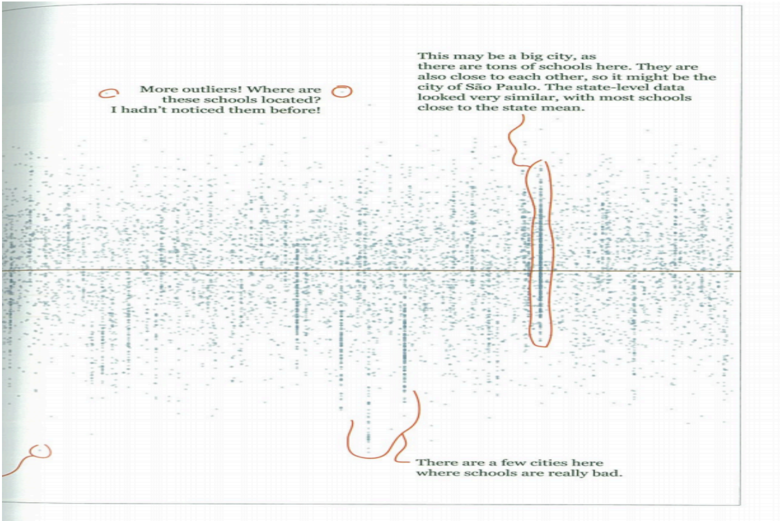


Figure 6.12 A strip plot of all schools in Brazil. Each dot represents one school, and each column of dots is a city. The annotations mimic what I do in a real project, which is to write down reminders of things that look promising in the data.


More outliers! Where are these schools located? I hadn't noticed them before!
This may be a big city, as there are tons of schools here. They are also close to each other, so it might be the city of São Paulo. The state-level data looked very similar, with most schools close to the state mean.
There are a few cities here where schools are really bad.

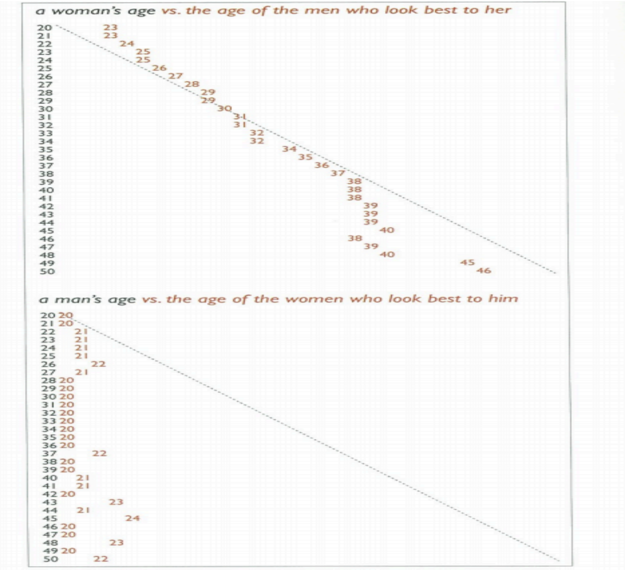


Figure 7.1 Chart by Christian Rudder.

Figure 7.2 Five possible (and fictitious) distributions for the data corresponding to the preferences of men of 40. All of them have the same mode: 21.

The Truthful Art: Data, Charts, and Maps for Communication.

Figure 8.21 Always use equal intervals on the X-axis and emphasize missing scores.

The Truthful Art: Data, Charts, and Maps for Communication.

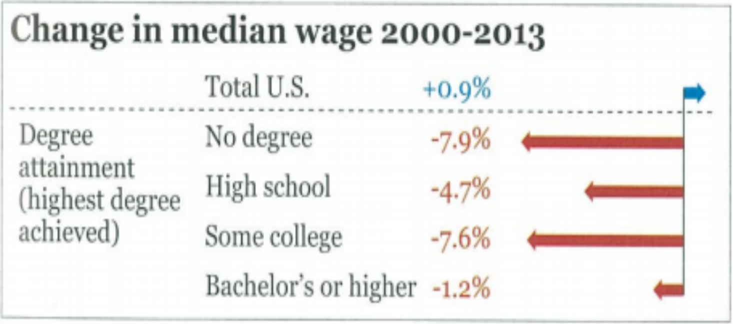


Figure 8.23 Isn't there a contradiction in this chart?

The Truthful Art: Data, Charts, and Maps for Communication.

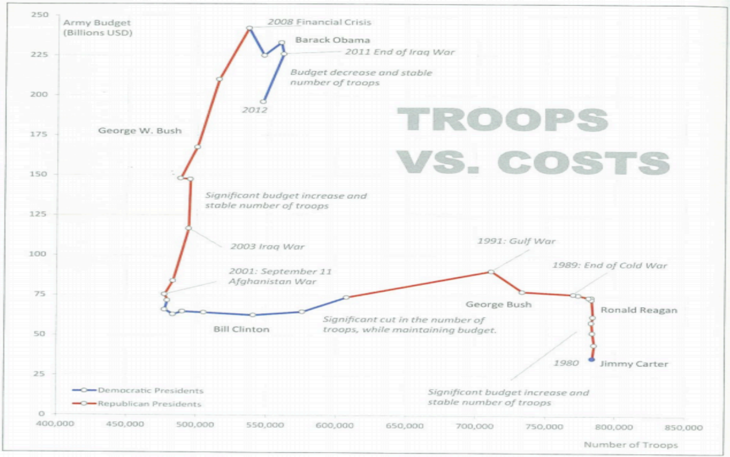


Figure 8.31 Connected scatter plot by Jorge Camões.

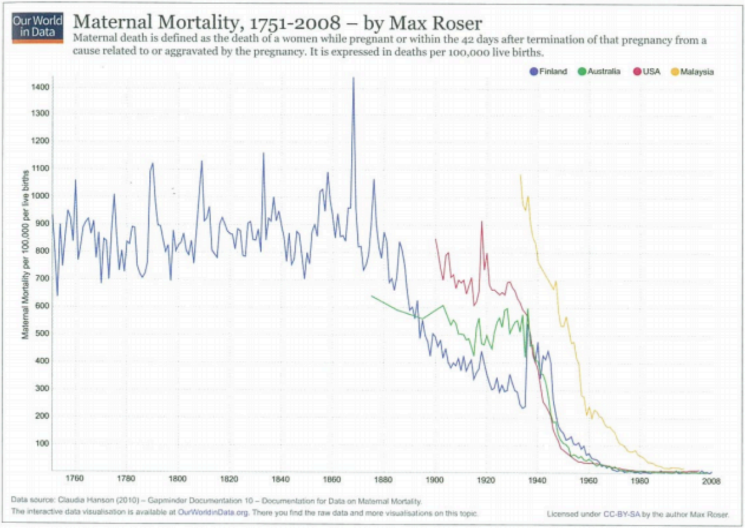


Figure 8.32 Time series line chart by Max Roser, http://ourworldindata.org/.

The Truthful Art: Data, Charts, and Maps for Communication.

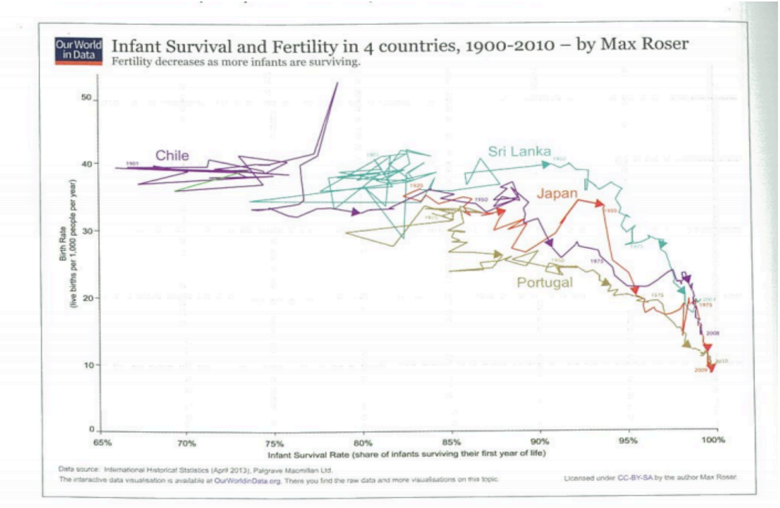


Figure 8.34 Connected scatter plot by Max Roser, http://ourworldindata.org/.

The Truthful Art: Data, Charts, and Maps for Communication.

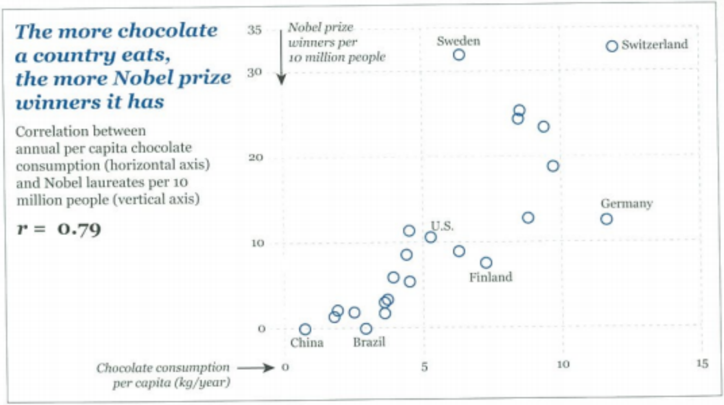


Figure 9.1 Chocolate and Nobel prizes.

The Truthful Art: Data, Charts, and Maps for Communication.

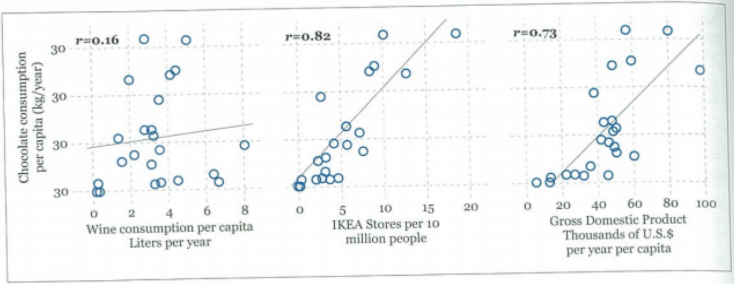


Figure 9.3 The relationship between chocolate consumption—and IKEA stores!

The Truthful Art: Data, Charts, and Maps for Communication.

Students	Math scores	English scores
Maria	10	16
Helen	11	17.5
Alice	14	22
Jorge	15	23.5
Lucius	16	25
Finn	18	28
Ricky	21	32.5
Alisha	22	34
Joe	23	35.5
Liz	24	37

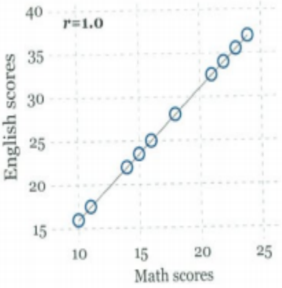


Figure 9.4 A perfect correlation.

The Truthful Art: Data, Charts, and Maps for Communication.

Figure 9.5 A non-linear relationship. (Source: Transportation Energy Data Book Edition 3: http://tinyurl.com/oaevnrx.)

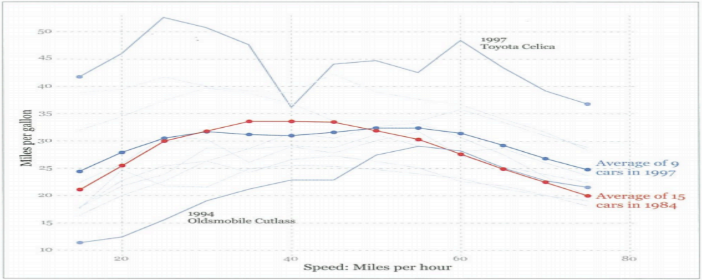


Figure 9.6 Comparing the average of 15 cars tested in 1984 to the average of nine cars in 1997 and to the models tested that year (light blue).

The Truthful Art: Data, Charts, and Maps for Communication.

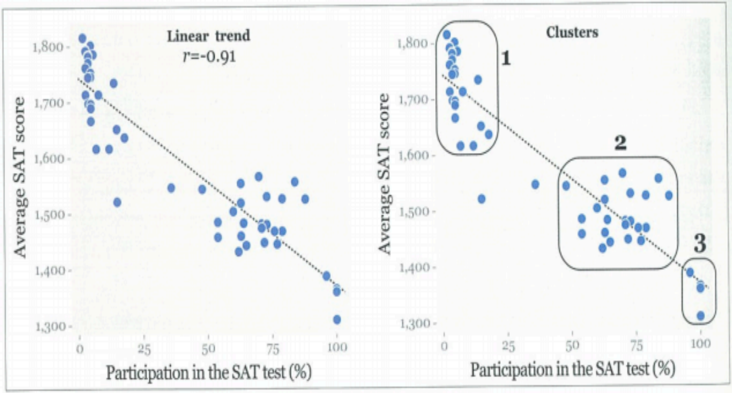


Figure 9.7 Straight trend line and data clusters.

The Truthful Art: Data, Charts, and Maps for

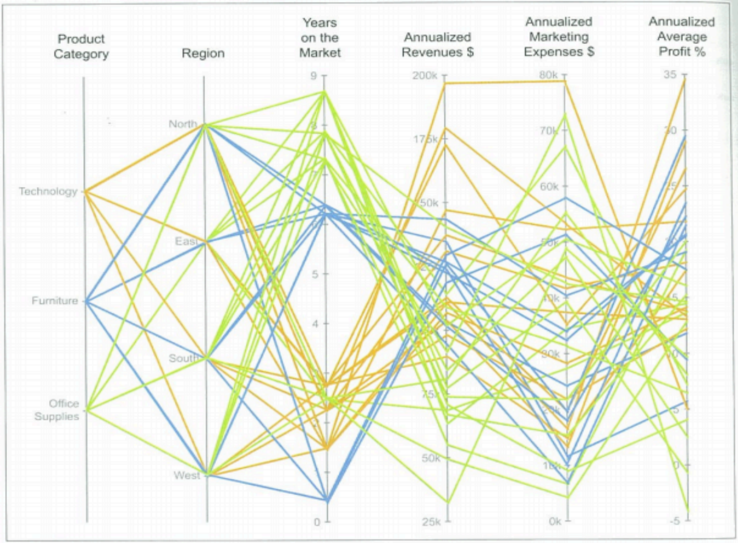


Figure 9.15 A parallel coordinates plot by Stephen Few.

The Truthful Art: Data, Charts, and Maps for Communication.

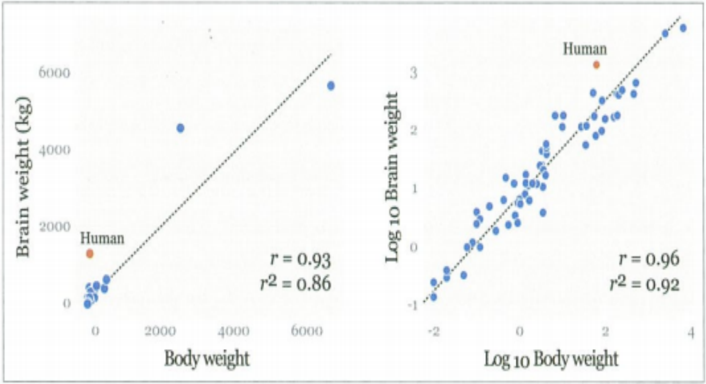


Figure 9.24 Correlation between body weight and brain weight for 62 mammal species. Raw data and logs. Data from Allison, T. and Cicchetti, D.V. (1976), "Sleep in mammals: ecological and constitutional correlates," *Science*, v. 194, pp. 732–34. Full data set at http://tinyurl.com/pbm9pzg.

The Truthful Art: Data, Charts, and Maps for Communication.

Figure 10.1 The same region of the world represented at two different scales.

The Truthful Art: Data, Charts, and Maps for Communication.

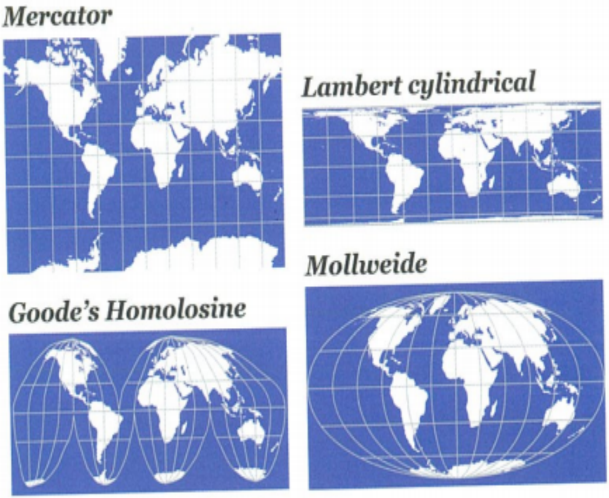


Figure 10.4
Four very popular map projections.

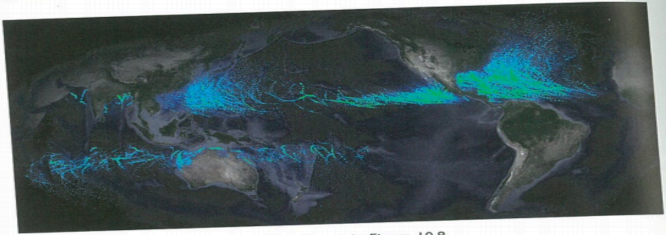


Figure 10.7 A draft for the map of hurricanes in Figure 10.8.

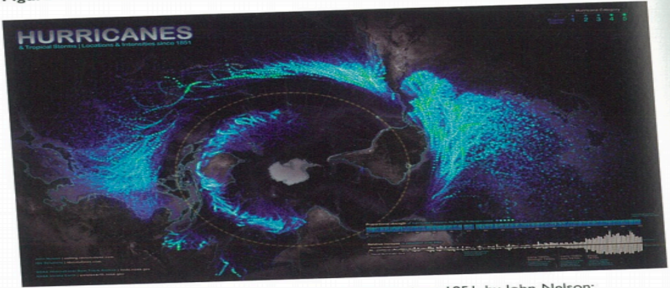


Figure 10.8 A map of hurricanes and tropical storms since 1851, by John Nelson: http://tinyurl.com/9y2axf4.

The Truthful Art: Data, Charts, and Maps for Communication.

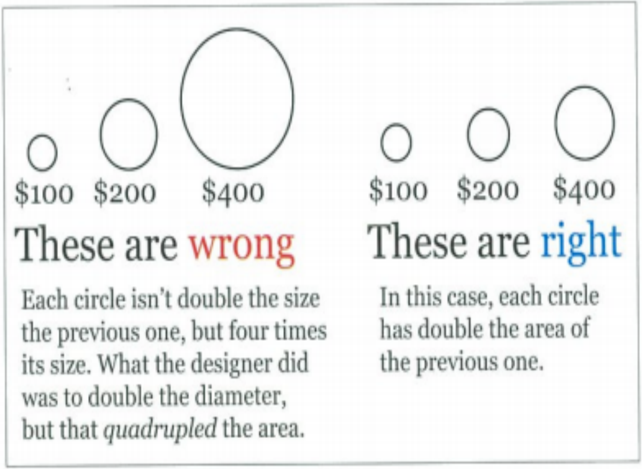


Figure 10.13 How to size circles incorrectly and correctly.

The Truthful Art: Data, Charts, and Maps for Communication.

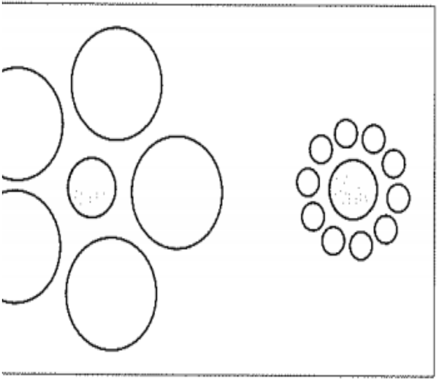


Figure 10.15 The Ebbinghaus illusion, named after German psychologist Hermann Ebbinghaus. The gray circle surrounded by large white circles looks smaller than the one surrounded by small white circles. In truth, they are the same size.

The Truthful Art: Data, Charts, and Maps for Communication.

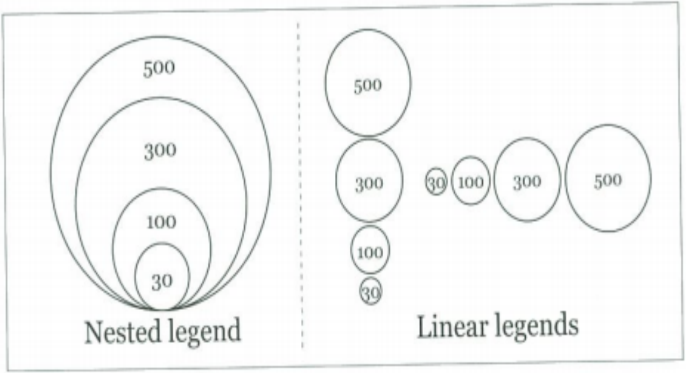


Figure 10.17 Legends for proportional symbol maps.

The Truthful Art: Data, Charts, and Maps for Communication.

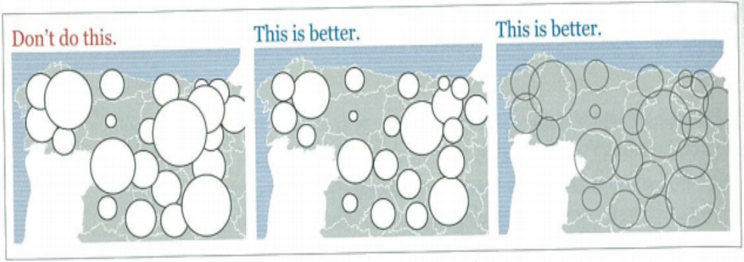


Figure 10.16 How to make proportional symbol maps clearer.

The Truthful Art: Data, Charts, and Maps for Communication.

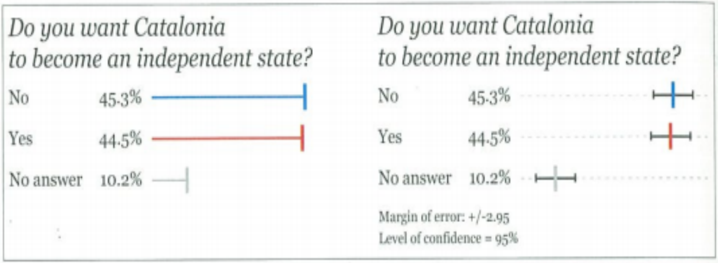


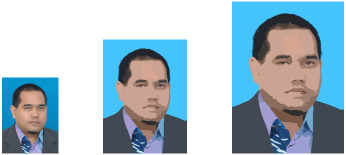
Figure 11.1 Displaying the margin of error can change your view of the data.

Lecture 6: Show Me the Numbers

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Practice in Table design

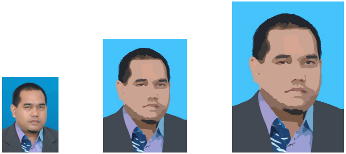
Nothing helps learning take root like practice. You will strengthen your developing expertise in table design by working through a few real-world scenarios. Six exercises are given.

Lecture 7: Data Visualisation: A Handbook for Data Driven Design. London, Sage Publication. Kirk, A. (2016)

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Defining Data Visualisation

The representation and presentation of data to facilitate understanding. The process of understanding a data visualisation involves three stages, namely:

- Perceiving: what can I see?
- Interpreting: what does it mean?
- Comprehending: what does it mean to me?

Data Visualisation

Three principles of good visualisation design:

- Good data visualisation is trustworthy.
- Good data visualisation is accessible.
- Good data visualisation is elegant.

Visualisation Workflow

Four key stages

- Formulating your brief: planning, defining and initiating your project.
- Working with data: going through the mechanics of gathering, handling and preparing your data.
- Establishing your editorial thinking: defining what you will show your audience.
- Developing your design solution: considering all the design options and beginning the production cycle.

UPM UNIVERSITI PUTRA MALAYSIA BERILMU BERBAKTI

Visualisation Workflow

Partition your mindset:

- Thinking: conceptual task, decision making.
- Doing: practical undertakings like sketching, visually examining data.
- Making: technical duties like analysisng data, constructing the solution.

Some notes in Page 60.

The Hidden Thinking: Establishing Your Project's Context

Defining your origin curiosity. Why are we doing it: what type of curiosity has motivated the decision/desire to undertake this visualisation project?

- Personal intrigue: "I wonder what ..."
- Stakeholder intrigue: "He/she needs to know ..."
- Audience intrigue: "They need to know ..."
- Anticipated intrigue: "They might be interested in knowing ..."
- Potential intrigue: "There should be something interesting..."

The Hidden Thinking: Establishing Your Project's Context

Circumstances. The key factors that will impact on your critical thinking and shape your ambitions:

- People: stakeholders, audience.
- Constraints: pressure, rules.
- Consumption: frequency, setting.
- Deliverables: quantity, format.
- Resources: skills, technology.

The Hidden Thinking: Establishing Your Project's Context

Defining your purpose. The ’so what?’: what are we trying to accomplish with this visualisation? What is a successful ’outcome’?

The Hidden Thinking: Establishing Your Project's Vision

Purpose Map. Plotting your expectation of what will be the best-fit type of solution to facilitate the desired purpose:

- What kind of experience? Explanatory, exhibitory or exploratory?
- What tone of voice will it offer? The efficiency and perceptibility of reading data vs the high-level, affective nature of feeling data?

The Hidden Thinking: Establishing Your Project's Vision

Harnessing Idea. What mental images, ideas and keywords instinctively come to mind when thinking about the subject matter of this challenge? What infuence and inspiration can you source from elsewhere that might start to shape your thinking?

Some notes in Page 95.

Working with Data

Key foundations for the requisite data literacy (the different types of data):

- Textual (qualitative): e.g. 'Any other comments?' data submitted in a survey.
- Nominal (qualitative): e.g. The 'gender' selected by survey participant.
- Ordinal (qualitative): e.g. The response to a survey question, based on a scale of 1 (unhappy) to 5 (very happy).
- Interval (quantitative): e.g. The shoe size of a survey participant.
- Ratio (quantitative): e.g. The age of a survey participant in years.

Others should be mentioned i.e. texts, colors, speeches, smell, touches etc. From all the senses.

Working with Data

Four steps involved in working with data:

- Acquisition - Different sources and methods for getting your data. Curated by you and by others.
- Examination - Developing an intimate appreciation of the characteristics of this critical raw material. Physical properties and meaning.
- Transformation - Getting your data into shape, ready for its role in your exploratory analysis and visualisation design. Clean, create and consilidate.
- Exploration - Using visual and statistical techniques to see the data's qualities: what insight does it reveal to you as deepen your familiarity with it?

More in Page 128.

ESTABLISHING YOUR EDITORIAL THINKING

Three perspectives that underpin your editorial thinking:

Angle

- Must be relevent in its potential interest for your audience.
- Must have sufficient quantities to cover all relevent views - but no more than required.

Establishing Your Editorial Thinking

Three perspectives that underpin your editorial thinking:

Framing

- Applying filters to your data to determine the inclusion and exclusion criteria.
- Framing decisions must provide access to the most salient content but also avoid any distorting of the view of the data.

Establishing Your Editorial Thinking

Three perspectives that underpin your editorial thinking:

Focus

- Which features of the display to draw particular attention to?
- How to organise the visibility and hierarchy of the content?

More in Page 142.

Visual Encoding

All charts are based on combinations of marks and attributes:

- Marks: represent records (or aggregation of records) and can be points, lines, areas or forms.
- Attributes: represent variable values held for each record and can include visual properties like position, size, colour, connection.

Chart Types

If visual encoding is the fundamental theoretical understanding of data representation, chart types are the practical application. There are five families of chart types:

CATEGORICAL	Comparing categories and distributions of quantitatives values
HIERARCHICAL	Charting part-to-whole relationships and hierarchies
RELATIONAL	Graphing relationships to explore correlations and connections
TEMPORAL	Showing trends and activities over time
SPATIAL	Mapping spatial patterns through overlays and distortions

Influencing Factors and Considerations

Formulating the brief: skills and resources - what charts can you make and how efficiently? From the definitions across the 'purpose map' what 'tone' did you detremine this project might demonstrates?

Influencing Factors and Considerations

Working with data: what is the shape of the data and how might that impact on your chart design? Have you already used a chart type to explore your data that might prove to be the best way to communicate it to others?

Influencing Factors and Considerations

Establishing your editorial thinking: what is the specific angle of the enquiry that you want to potray visually? Is it relevent and representative of the most interesting analysis of your data?

Influencing Factors and Considerations

Trustworthy design: avoid deception through mistaken geometric calculations, 3D decoration, truncated axis scales, corrupt charts.

Influencing Factors and Considerations

Accessible design: the use of encoded overlays, such as bandings, markers, reference lines, can aid readability and interpretation.

Influencing Factors and Considerations

Elegant design: consider the scope of certain design flourishes that might enhance the visual appeal through the form of your charts whilst also prevailing their function.

Refer to Page 221 for Tips and Tactics

Interactivity

Data adjustments affect what data is displayed and may include the following features:

- Framing: isolate, include or exclude data.
- Navigating: expand or explore greater levels of detail in the displayed data.
- Animating: portray temporal data via animated sequences.
- Sequencing: navigate through discrete sequences of different angles of analysis.
- Contributing: customising experiences through user-inputted data.

INTERACTIVITY

Presentation adjustment affect how the data is displayed and may include the following features:

- Focusing: control what data is visually emphasised.
- Annotating: interatc with marks to bring up more detail.
- Orientating: make better sense of your location within a display.

Influencing Factors and Considerations

Formulating the brief: skills and resources, timescales, setting, and format will all influence the scope of interactivity. What experience are your facilitating and how might interactive options help achieve this?

Influencing Factors and Considerations

Working with data: what range of data do you wish to include? Large datasets with diverse values may need interactive features to help users filter views and interrogate the contents.

Influencing Factors and Considerations

Establishing your editorial thinking: choices made about your chosen angle, as well as definitions for framing and focus will all influence interactive choices, especially if users must navigate to view multiple angles of analysis or representations potrayed through animated sequences.

Influencing Factors and Considerations

Data representation: certain chart choices may require interactivity to enable readability.

Influencing Factors and Considerations

Trustworthy design: functional performance and reliability will substantiate the perception of trust from your users.

INTERACTIVITY

Influencing Factors and Considerations

Accessible design: any interactive feature should prove to be useful and unobtrusive. Interactivity can also assist with challenges around visual accessibility.

Influencing Factors and Considerations

Elegent design: beware of feature creep, minimise the clicks, but embrace the pleasure of playability.

More in Page 246.

Annotation

Project annotation help viewers understand what the project is about and how to use it, and may include the following features:

- Headings: titles, sub-titles and section headings.
- Introductions: providing background and aims of the project.
- User guides: advice or instruction for how to use any interactive features.
- Multimedia: the potential to enhance your project using appropriate imagery, videos or illustrations.
- Footnotes: potentially includes data sources, credits, usage information, and time/date stamps.

UPM
UNIVERSITI PUTRA MALAYSIA
BERILMU BERBAKTI

Annotation

Chart annotation help viewers perceive the charts and optmise their potential interpretations and may include the following features:

- Charts apparatus: axis lines, gridlines, tick marks.
- Labels: axis titles, axis labels, value labels.
- Legend: providing detailed keys for colour or size associations.
- Reading guides: detailed instructions advising readers how to perceive and interpret the chart.
- Chaptions: drawing out key findings and commentaries.

Typography
Most of the annotation features you include are based on text and so you will need to consider carefully the legibility of the typeface you choose and the logic behind the-font-size hierarchy you display.

Annotation

Influencing Factors and Considerations

- Formulating the brief: consider the characteristics and needs of the audience. Certain chart choices and subjects may require more explanation. From the 'purpose map' what type of tone and experience are you trying to create and what role might annotation play?
- Establishing your editorial thinking: what things do you want to emphasise or direct the eye forwards (focus)?
- Trustworthy design: maximise the information viewers have to ensure all your data work is transparent and clearly explained.
- Accessible design: what is the right amount and type of annotation suitable to the setting and complexity of your subject?
- Elegant design: minimise the clutter.

Some notes in Page 261.

Colour

Data legibility involves using colours to represent different types of data. The most appropriate colour association or scale decisions will depend on the data type: nominal (qualitative), ordinal (qualitative), interval and ratio (quantitative).

Colour

Editorial salience is about using colour to direct the eye. For which features and to what degree of emphasis do you want to create contrast?

Functional harmony concerns deciding about every other colour property as applied to all interactive features, annotations and aspect of your composition thinking.

Influencing Factors and Considerations

- Formulating the brief: format, setting, colour rules and imposed guidelines all have a significant impact. Your definitions about both tone and experience, on the purpose map, will lead to specific choices being more suitable than others. What initial ideas did you form? Have any sources of inspiration already implanted ideas inside your head about which colours you could use?
- Working with data: what type of data and what range of values/number of classifications have you got?

Influencing Factors and Considerations

- Establishing your editorial thinking: what things do you want to emphasise or direct the eye towards (focus)?
- Data representation: certain chart type choices will already include colour as an encoded attribute.
- Trustworthy design: ensure that your colour choices are faithful to the shape of your data and the integrity of your insight. If something looks meaningful it should be, otherwise it will confuse or deceive.

Influencing Factors and Considerations

- Accessible design: once you've commited colour to mean something preserve the consistency of association for as long as possible. Be aware of the sensitivies around visual accessibility and positive/negative colour connotations.
- Elegant design: the perception of colours is relative so unity of your choices needs to be upheld. Ensure that you can justify every dot of colour used and, ultimately, rely on your own judgment to determine when your final palette feels right.

Some notes in Page 291.

Composition

Project composition defines the layout and hierarchy of the entire visualisation project and may include the following features:

- Visual hierarchy - layout: how to arrange the position of elements?
- Visual hierarchy - size: how to manage the hierarchy of element sizes?
- Absolute positioning: where specifically should certain elements be placed?

Composition

Chart composition defines the shape, size and layout choices for all components within your charts and may include the following features:

- Chart size: don't be afraid to shrink charts, so long as any labels are still readable, and especially embrace the power of small multiple.
- Chart scales: what are the most meaningful range of values given the nature of the data?
- Chart orientation: which way is best?
- Chart value sorting: consider the most meaningful sorting arrangement for your data and editorial focus, based on the LATCH (Location, Alphabet, Time, Category or Hierarchy) acronym.

Influencing Factors and Considerations

- Formulating the brief: what space have you got to work within?
- Working with data: what is the shape and size of your data and how might this affect your chart design architecture?
- Establishing your editorial thinking: how many different angles (charts) might you need to include? Is there any specific focus for theses angles that might influence a sequence of hierarchy between them?
- Data representation: any chart has a spatial consequence - different charts have different structures that will create different dimensions that will need to be accomodated.

Composition

Influencing Factors and Considerations

- Trustworthy design: the integrity and meaning of your chart scale, chart dimensions, and (for mapping) your projection choices are paramount.
- Accessible design: remember that good design is unobstrusive - if you want to facilitate comparisons between different chart displays these ideally need to be presented within a simulataneous view.
- Elegant design: unity of arrangement is another of the figer-tip sense judgements but will be something achieved by careful thinking about the relationships between all components of your work.

Some notes in Page 310.

Visualisation Literacy

Viewing: Learning to See

Before You Begin

- Setting: is the situation you are in conducive to the task of consuming a s visualisation? In a rush? Travelling?
- Visual appeal: are you sufficiently attracted to the appearance of the work?
- Relevance: do you have an interest or a need to engage with topic?
- Initial scan: quickly orientate yourself around the page or screen, and allow yourself a brief moment to be drawn to certain features.

Viewing: Learning to See

Outside the Chart

- The proposition: what task awaits? What format, function, shape and size of visualisation have you got to work with?
- What's the project about?: look at the titles, source, and read through any introductory explanantions.
- What data?: look for information about where the data has originated from and what might have been done to it.
- What interactive functions exist?: if it is a digital solution browse quickly and asquaint yourself with the range of interactive devices.

Viewing: Learning to See

Inside the Chart

- Perceiving: what does it show?
- Interpreting: what does it mean?
- Comprehensing: what does it mean to me?

Viewing: Learning to See

Becoming a More Sophisticated Consumer

- Appreciation of context: what circumstances might the visualiser have been faced with that are hidden from you as a viewer?
- Overciew first, details if provided: accept that sometimes a project only aims to (or maybe only can) provide a big-picture gist of the data, rather than precise details.
- False consciousness: don't be too quick to determine that you like a visualisation. Challenge yourself, do you really like it? Do you really gain understanding from it?
- Curiosity answered, curiosities not answered: just because it does not answer your curiosity, it might answer those of plenty of others.

Creating: The Capabilities of the Visualiser

The Seven Hats of Data Visualisation Design

- Project Manager: the coordinator - oversees the project.
- Communicator: the broker - manages the people relationships.
- Scientist: the thinker - provides scientific rigour.
- Data analysist: the wrangler - handles all the data work.
- Journalist: the reporter - pursues the scent of enquiry.
- Designer: the conceiver - provides creative direction.
- Technologist: the developer - constructs the solution.

Creating: The Capabilities of the Visualiser

Assessing and Developing Your Capabilities

The importance of reflective learning: evaluating the outcome of the work you have created and assesssing your own performance during its production.

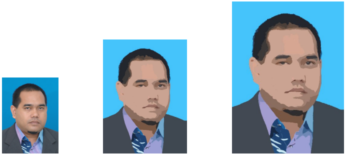
Some notes in Page 334.

Lecture 8: Show Me the Numbers

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Practice in Graph Design

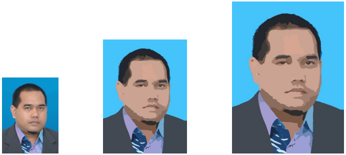
You've come far in your expedition into the world of graph design. It's now time for some pratice to pull together and reinforce all that you've learned. Expert graph design requires that you adapt and apply what you've learned to a variety of real-wolrd communication probelms.

Lecture 9: Save the Pies for Dessert

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

10 & 11 July 2018



Refs:

- Spence, I. (2005). No Humble Pie: The Origins and Usage of a Statistical Chart. Journaal of Educational adn Behaviour Statistics. Winter 2005, Vol. 30, No. 4, 353 - 368.
- Few, S. (2007). Save the Pies for Dessert. Perceptual Edge, Visual Business Intelligence Newsletter, August. 1- 14.

The Pie

Of all the graphs that play major roles in the lexicon of quantitative communation, however, the pie chart is by far the least effective. Its colorful voice is often heard, but rarely understood. It mumbles when it talks.

The Pie

The primary strength of a pie is the fact that the message "part-to-whole relationship" is built right nto it in an obvious way. We can learn fractions by looking at pies sliced in various ways and decoding the ratio (quarter, half, three quarters, etc.) of each slice.

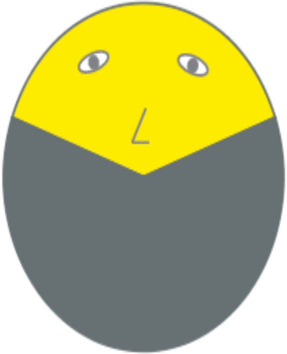
The Bar Chart

Despite the obvious nature of a pie charts message, bar graphs provide a much better means to compare the magnitudes of each part. Pie charts only make it easy to judge the magnitude of a slice when it is close to 0%, 25%, 50%, 75% or 100%. Any percentages other than these are difficult to dicern in a pie chart, but can be accurately discerned in a bar graph, thanks to the quantitative scale. Illustrations.

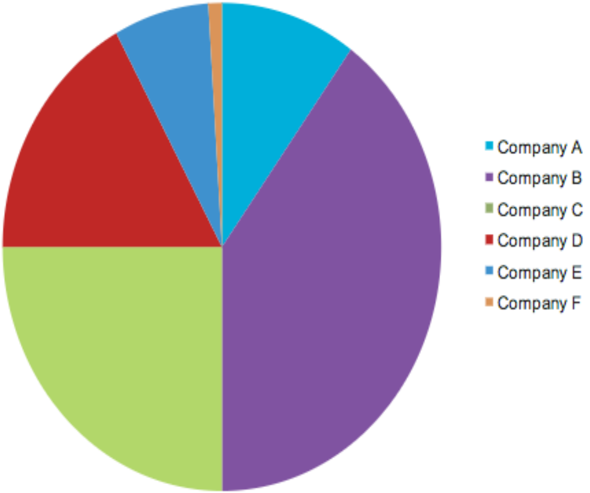
Positioning the Pies

Making it is hard to recognized the portion of the pies. You might argue that this problem can be easily solved by labelling the values of each slice. Do you realize what we have just done? Becaue the pie chart was difficult to read, we added values so we wouldn't have to compare the sizes of the slices and we added direct labels so we wouldn't have to rely on the legend. We turned the pie chart into an awkwardly arranged equivalent of a table of labels and values.

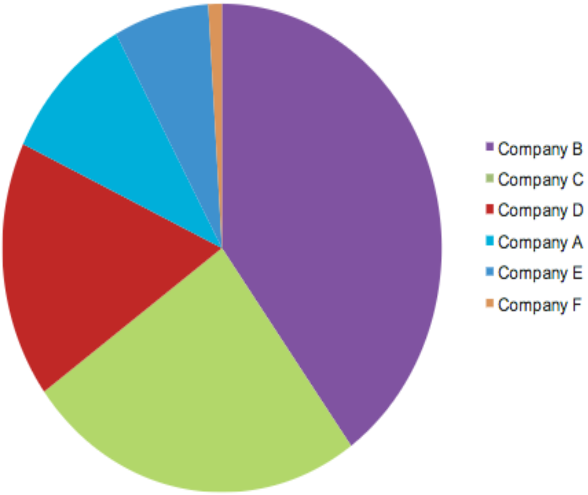
Save the Pies for Dessert



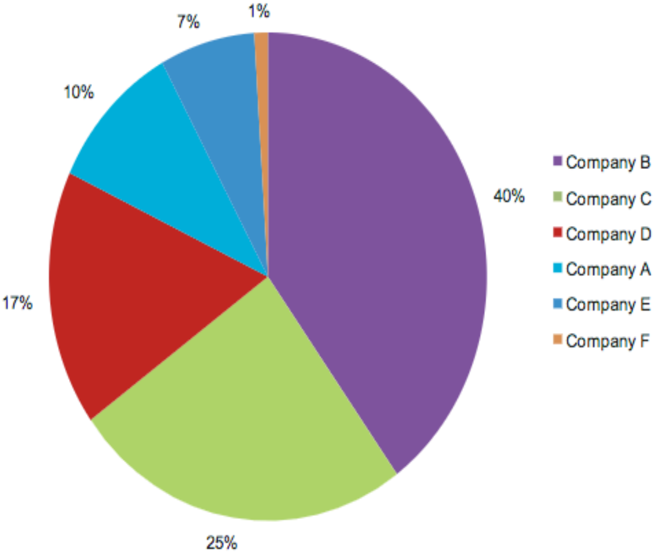
Save the Pies for Dessert



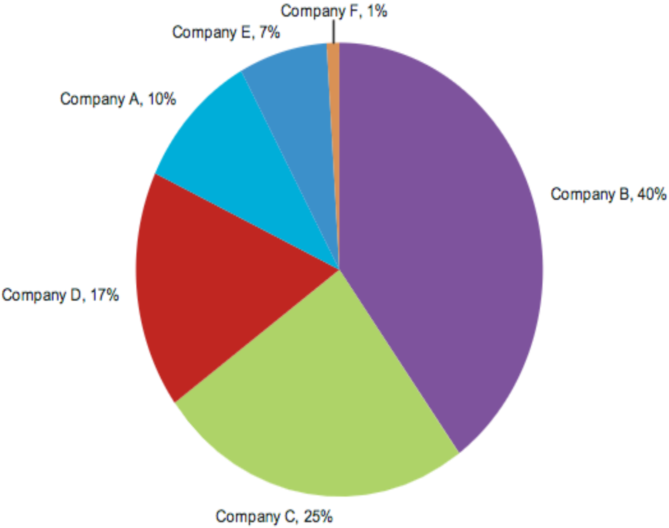
Save the Pies for Dessert



Save the Pies for Dessert



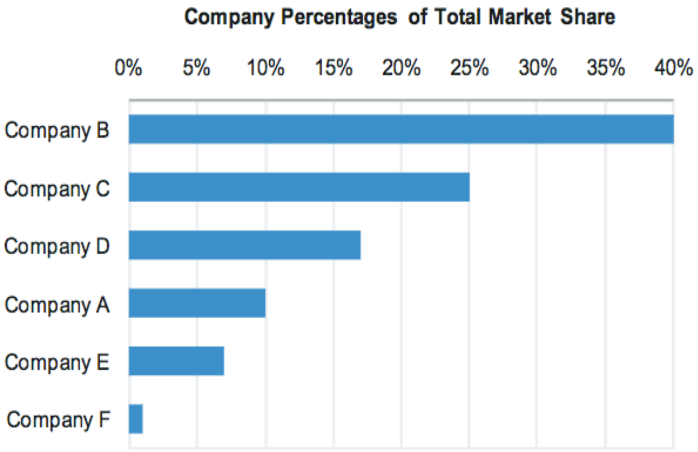
Save the Pies for Dessert



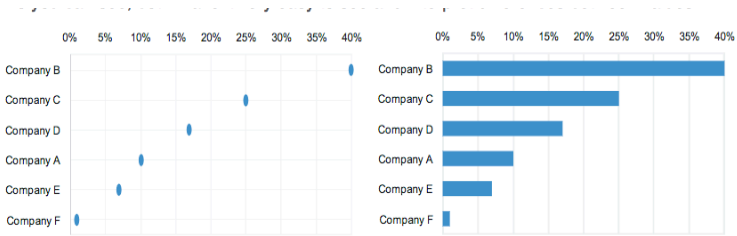
Save the Pies for Dessert

Companies	Percentage
Company B	40%
Company C	25%
Company D	17%
Company A	10%
Company E	7%
Company F	1%
Total	100%

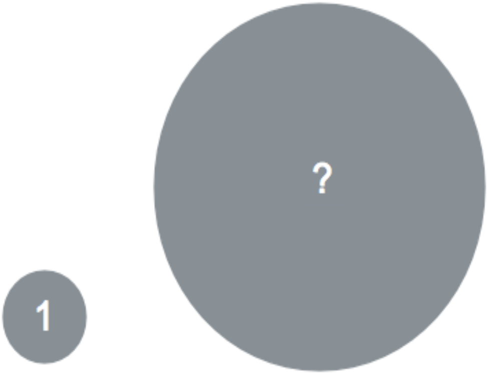
Save the Pies for Dessert



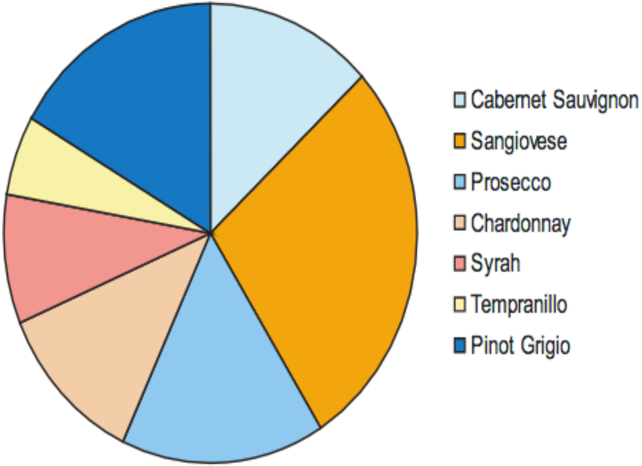
Save the Pies for Dessert



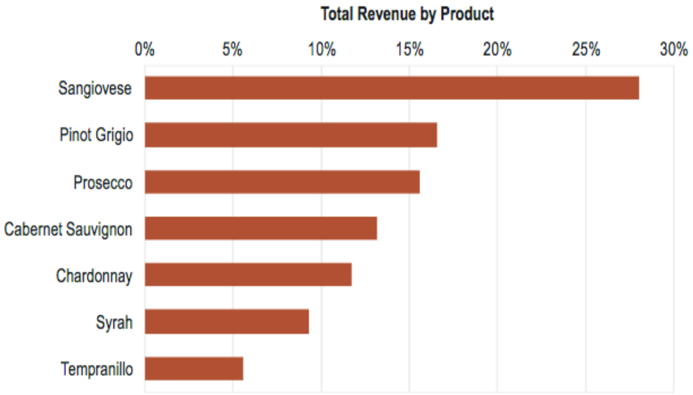
Save the Pies for Dessert



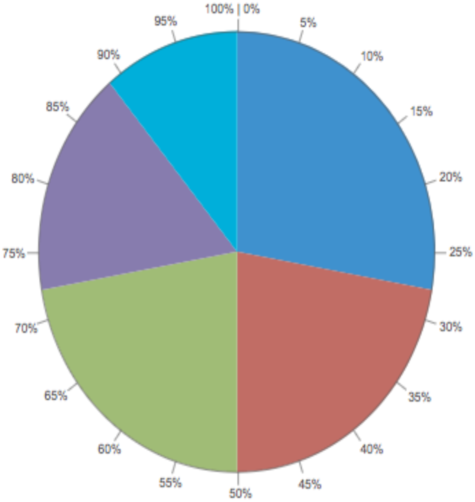
Save the Pies for Dessert



Save the Pies for Dessert

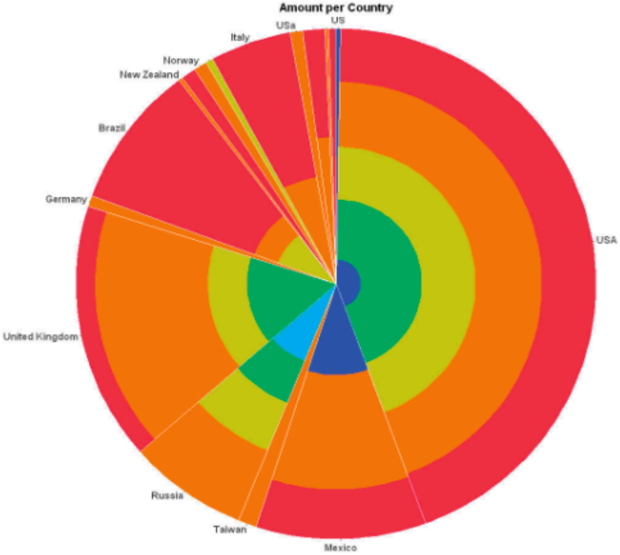


Save the Pies for Dessert

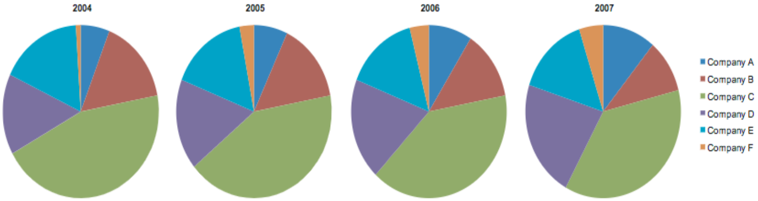


Save the Pies for Dessert

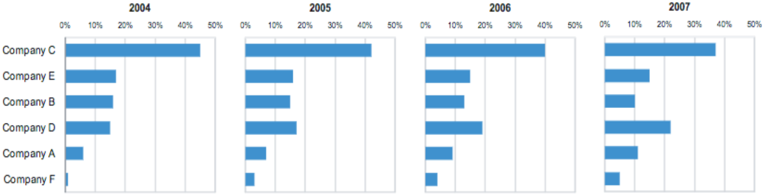
Save the Pies for Dessert



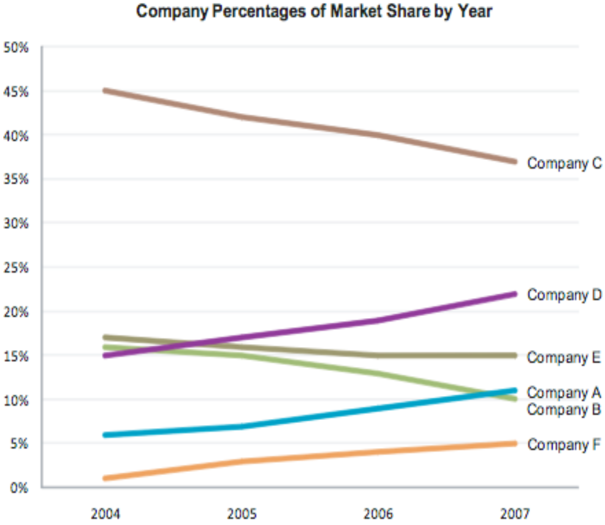
Save the Pies for Dessert



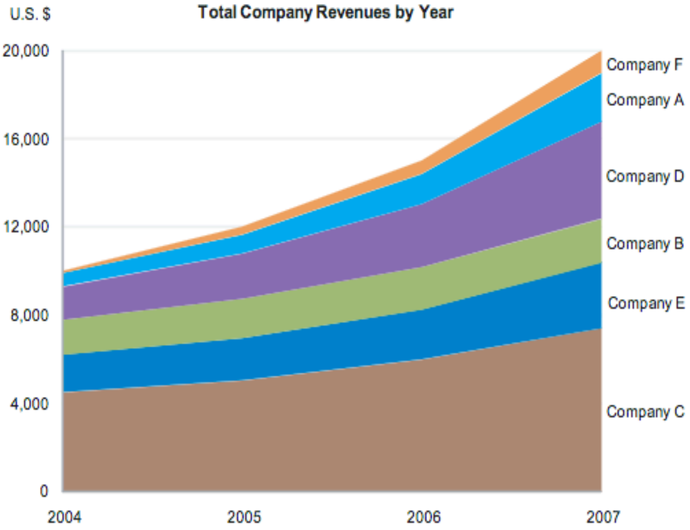
Save the Pies for Dessert



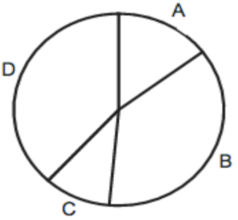
Save the Pies for Dessert

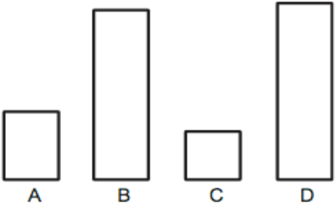


Save the Pies for Dessert



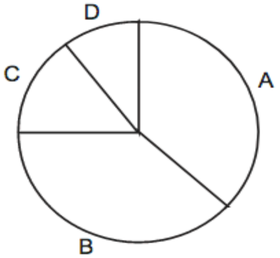
Save the Pies for Dessert

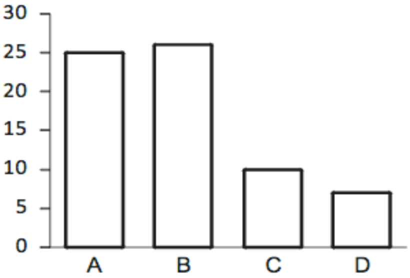




A	B	C	D
10	25	7	26

Save the Pies for Dessert





مُهَمَّدْ بَكْرِ آدَمْ
وْاَسَّلَمْ

Thank You for Your Participation in this Lecture Series From One Statistician Perspective.

Mohd Bakri Adam, PhD (Lancaster)

Laboratory of Statistical and Computing Services, INSPEM, UPM

11 July 2018