5   Visual Design and Data Charts

When it comes to data charts, however, the Gestalt principles are not the entire story; various other factors enter the **visual design** picture.

5.1   Visualization and Memory

The first thing to realize is that human brains are engaged with different **types of memories** when they are faced with visual information and storytelling:

- **iconic** memory directs the eye (pre-attentive processes);
- **short-term** memory limits how many charts one may reasonably expect to encounter on a report/dashboard page or in infographics, say, and
- **long-term** memory helps the audience retain the visual message and information presented to them.

Iconic Memory

Iconic memory is the visual sensory memory register relating to the visual domain; it is a fast-decaying, high-capacity store of visual information. Most importantly, iconic memory is **exceedingly brief**: it provides a **coherent representation of our entire visual perception** in a heartbeat. We think it evolved as a survival mechanism attuned to pre-attentive attributes (see Figure 4.6), the subconscious accumulation of information from the environment.

In a nutshell iconic memory draws our eyes to anything in our field of vision that seems out of place (and potentially represents a danger); once the situation has been "resolved" to our satisfaction, the triggering stimulii are absorbed by our regular memory and become nearly invisible.[1]

Visual thinking seeks patterns. In the search for such patterns, **pre-attentive processes** are fast, instinctive, efficient, and compatible with multitasking. They gather information and build patterns, identifying features first, and then seeing the big-picture objects.

We can easily locate Waldo (in the image at the top of the next page) using only pre-attentive attributes; the task is so easily accomplished that the image

1: We can get used to many things that do not quickly kill us outright: it took one of us (Patrick) about 30 minutes to become comfortable driving on the left on a trip to New Zealand after 20 years of driving on the right in Canada, but he has not been able to shake off his French Canadian accent after nearly 30 years of speaking English.

can become **uninteresting** in short order: we have seen what there was to see and are ready to move on to the next order of business.
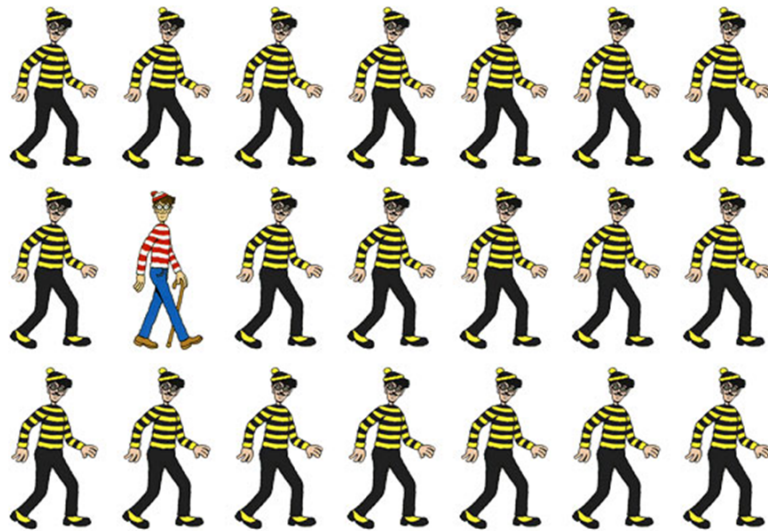

Figure 5.1: Using pre-attentive processes to easily locate Waldo, the observation of interest (from a brilliant idea in [43]).

Attentive processes, on the other hand, are slow, deliberate, and focused; they require time and concentration (cognitive energy). They identify objects first, then drill on down to the features.

Any serious attempt to locate Waldo in the image below requires shedding pre-attentive processes; the task can easily become engrossing (explaining the books' popularity, presumably).
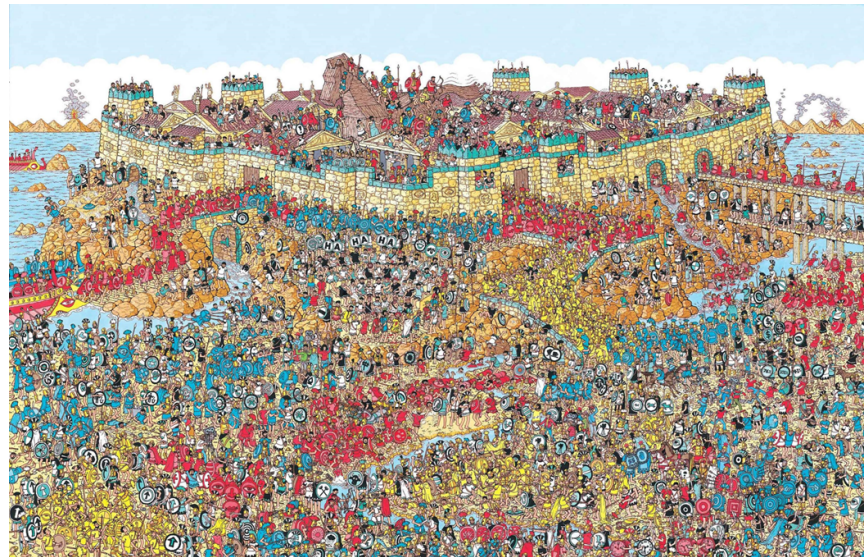

Figure 5.2: Failure of pre-attentive processes to easily locate Waldo, the observation of interest (again, from [43]).

2: In fact, this often proves counterproductive in data storytelling, where the main principle is to clearly **show and tell**.

3: Sorry about that ...

Most data presentations do not ask for such a commitment on the part of their audience;[2] ; how long would you be willing to search before concluding that the exercise was a charade and that Waldo is nowhere to be found in the picture?[3]

The distinction between the two is perhaps best illustrated by the images in Figure 5.3 – we can find the object of interest (red disk) in both cases, but the task is much more difficult to accomplish in one of the instances.
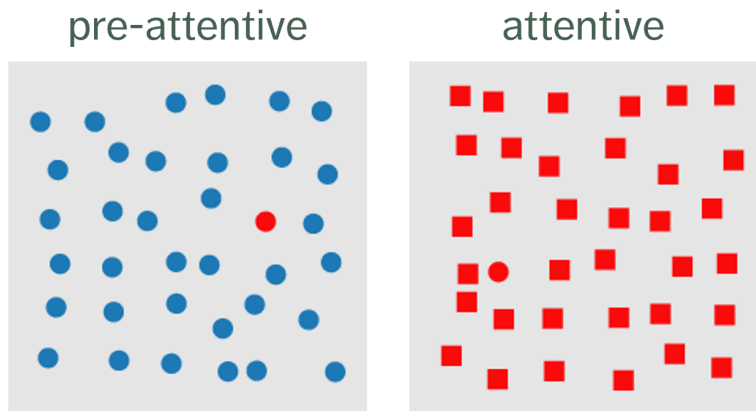

Figure 5.3: Classical illustration of the difference between a pre-attentive (left) and an attentive search (right) imposed by design element choices: we can still find the observation of interest on the right, but it requires work – not so on the left [original author unknown].

As we have seen in Figure 4.6, we can use a number of pre-attentive features to help identify markers (there are equivalent features for line segments [44]) – their relative effectiveness depends on the context.

Short-Term Memory

Short term memory provides a cap to how many visual information chunks we can hold in our brains simultaneously, namely about 4, for the most part (there are exceptions, of course, but they are rare).[4]

When presented with more information (such as a fifth chart on a dashboard page, say), some of the chunks currently held in the short term memory register must be processed to make room for the new information. The old chunks become invisible (until they are processed in again, at the expense of other charts).

This continuous short term memory **back and forth** can become quite distracting and interfere with the audience's "willing suspension of disbelief", which is to say, their willingness to play along with the data presentation. Quite simply, keeping to a low number of visual chunks on a given report or dashboard page makes it easier for the audience to understand the insights that are being conveyed by the visuals.

But this not to say that we can only read charts with 4 or fewer markers (be they lines, bars, dots, and the likes); in practice, we try to build charts so that they form **focused hierarchies** of visual parts – the Gestalt principles can help in that regard.[5]

Good visual design uses pre-attentive processes to draw in the audience and short term memory limitations to avoid overpopulating the display pages, while making the charts interesting enough to reward the audience that **sticks around for a longer look** and increasing the odds that the message is retained. This is where attentive processes and long term memory come in.

4: Note that there is conflicting evidence that suggests that the optimal short term memory chunk number may actually be 2, 3, 4, or even 5 [45, 46]. Due to layout consideration (see Section 5.2), we will stick with the convention and use "4" as our **magic number**.

5: What we mean here is that there may be various hierarchical levels in a chart. Think of the universe being composed of galactic groupings, themselves composed of galaxies, which are composed of stars, which may or may not have attendant planets and other bodies circling them. The equivalent structure might be a dashboard composed of pages, each consisting of a certain number of charts, which may hold clusters of markers; effective charts try to maintain a small number of components in each hierarchical level (it may not always possible to do so, of course).

Long-Term Memory

Long term memory is built up over a lifetime of experiences and forms the basis of **pattern recognition** and general **cognitive processing**; it is an aggregate of visual memory and verbal memory.

The combination of simple images with simple text helps us trigger long-term memory, making the story "stick". Design and text choices can also provide different contexts (and thus, different long-term memories).

As an example, consider the following text cards accompanying an unseen chart related to *Access to Information and Privacy* (ATIP) requests:

| You have currently selected 28,711 ATIP requests totaling 6,597,612 pages of information | VS | **ATIP Requests** | | |
| | | 30K requests | 6.6M pages | 230 pages/request |

The option on the left might be the **chart description** provided as an **accessibility** aid (see Section 6.4, *Practical Suggestions*). We will remember the **topic** (ATIP requests) and that there were **large numbers** (with different magnitudes), but it is a safe bet that the **vast majority** of the audience will not remember the exact numbers: there is really little point to reporting to the nearest unit with quantities in the ten thousands or the millions.

The text on the right, on the other hand, brings us a step closer to **conveying a data story**, with its focus on **general trends** and meaningful approximations, relative magnitudes, and the added statistical summary. It is likely that some audience members would remember 30K requests, 6.6M pages, and 230 pages/request when all is said and done.

We will say more on the topic in Chapter 8 (*Effective Storytelling Visuals*).

5.2 Aesthetic Considerations

Practitioners often end up spending quite a lot of time on considerations related to colour, size, and layout.[6] What is all the fuss about?

Colour

Colour theory is an old and complicated topic, mixing artistic, psychological, cultural, and scientific viewpoints, and it is not without its controversies;[7] we will not dwell on these topics, but here is a little start ⬀ .

From our perspective, it will suffice to know that there are three primary colours: blue, red, and yellow.[8] Additional colours can be built by mixing the primary colours in various combinations: purple is an equal mixture of red and blue, orange of yellow and red, and green of blue and yellow.[9] This information is often presented as **colour wheels**.

6: Overly much, perhaps, but that might simply be because they are fairly easy to control with modern tools.

7: Do colours have emotional meanings, for instance?

8: That is the **additive colour model**, which is not the only such model (see CMYK, [47, 48]).

9: That last one we would have a harder time believing had we not seen it time and time again on kids' drawings.. perhaps it has to do with there not being a green rod in our eyes?
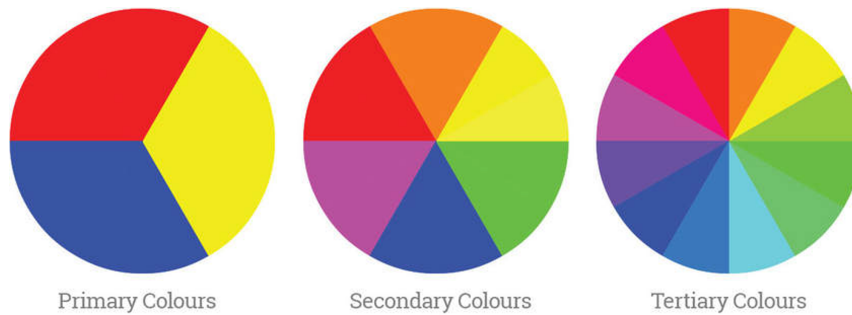
Primary Colours   Secondary Colours   Tertiary Colours

Figure 5.4: The main additive colour wheels (see above link).

The colour wheels provide a foundation for **colour schemes**, of which there are many:

- **achromatic** (colourless, using only blacks, whites and grays);
- **monochromatic** (1-colour schemes);
- **complementary** (colours directly across from each other on the colour wheel);
- **split complementary** (2 of the 3 colours are adjacent; 1 of the colours is opposite);
- **split-left** and **split-right complementary** ("split" colours are either to the left or right of the complementary colour);
- **analogous** (any 3 adjacent primary, secondary, or tertiary colours on the colour wheel);
- **colour diad** (2 colours that are 2 colours apart on the color wheel);
- **colour triad** (3 colours, equally distant from each other on the colour wheel), and
- **colour tetrad** (4 or more colors on the colour wheel).

Achromatic   Monochromatic   Complementary

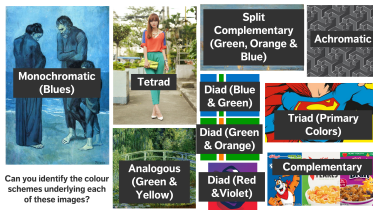Split Complementary   Left-Split Complementary   Analogous

Diad   Triad   Tetrad

Can you identify the colour schemes underlying each of these images?

Colour schemes are used in a variety of settings: can you recognize them in the images below?[10]

Can you identify the colour schemes underlying each of these images?

When it comes to colour, the advice is simple: **less is more**: it should be used sparingly.[11] But based on the Gestalt principles, **monochrome schemes** can be particularly effective: select a few base colours and build monochrome schemes from them (with an achromatic scheme thrown in for good measure).

When appropriate, the job of selecting a colour scheme can be made easier by using a **corporate identity**: this maximizes buy in, but be wary of schemes that run afoul of well-established conventions (especially those involving heavy use of reds) and accessibility considerations (in particular, schemes that ignore issues related to colourblindness and low-visibility).

Whatever colour scheme is settled on, it is good practice to take the time to create a **template**, share it around, and **stick to it**.

11: Graphic designers are taught to "get it right, in black and white".

Colour Palettes

Designers have built palettes with a variety of properties and made them available to practitioners.[12] Here is a sample of pre-built colour palettes.

Zeileis, Hornik, and Murrell

12: Various data visualization software also have commonly-used default palettes, see Top R Color Palettes to Know for Great Data Visualization ⬈ and Choosing a Color Palette For Your Power BI Report ⬈ , for instance.

Figure 5.5: Zeileis, Hornik, and Murrell's 24 distinct colour palette [49].

**Zeileis, Hornik & Murrell
24 Distinct Colours**

Hex	RGB	Display	Group	Hex	RGB	Display	Group
#023FA5	(2,63,165)		1	#11C638	(17,198,56)		3
#7D87B9	(125,135,185)		1	#8DD593	(141,213,147)		3
#BEC1D4	(190,193,212)		1	#C6DEC7	(198,222,199)		3
#D6BCC0	(214,188,192)		1	#EAD3C6	(234,211,198)		3
#BB7784	(187,119,132)		1	#F0B98D	(240,185,141)		3
#8E063B	(142,6,59)		1	#EF9708	(239,151,8)		3
#4A6FE3	(74,111,227)		2	#0FCFC0	(15,207,192)		4
#8595E1	(133,149,225)		2	#9CDED6	(156,222,214)		4
#B5BBE3	(181,187,227)		2	#D5EAE7	(213,234,231)		4
#E6AFB9	(230,175,185)		2	#F3E1EB	(243,225,235)		4
#E07B91	(224,123,145)		2	#F6C4E1	(246,196,225)		4
#D33F6A	(211,63,106)		2	#F79CD4	(247,156,212)		4

"Statistical graphics are often augmented by the use of color coding information contained in some variable. When this involves the shading of areas (and not only points or lines) – e.g., as in bar plots, pie charts, mosaic displays or heatmaps—it is important that the colors are perceptually based and do not introduce optical illusions or systematic bias. Based on the perceptually-based Hue-Chroma-Luminance (HCL) color space suitable color palettes are derived for coding categorical data (qualitative palettes) and numerical variables (sequential and diverging palettes)." [49]

Paul Tol

**Paul Tol
14 Colour
Rainbow Scheme**

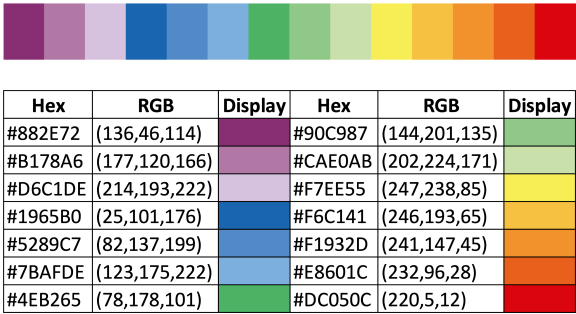| Hex	RGB	Display	Hex	RGB	Display
#882E72	(136,46,114)		#90C987	(144,201,135)	
#B178A6	(177,120,166)		#CAE0AB	(202,224,171)	
#D6C1DE	(214,193,222)		#F7EE55	(247,238,85)	
#1965B0	(25,101,176)		#F6C141	(246,193,65)	
#5289C7	(82,137,199)		#F1932D	(241,147,45)	
#7BAFDE	(123,175,222)		#E8601C	(232,96,28)	
#4EB265	(78,178,101)		#DC050C	(220,5,12)	

Figure 5.6: Tol's 14-colour rainbow scheme [50].

"Graphics with scientific data become clearer when the colours are chosen carefully. It is convenient to have good default schemes ready for each type of data, with colours that are: distinct for all people, including colour-blind readers; distinct from black and white; distinct on screen and paper; matching well together. This document shows such schemes, developed with the help of mathematical descriptions of colour differences and the two main types of colour-blind vision" [50]
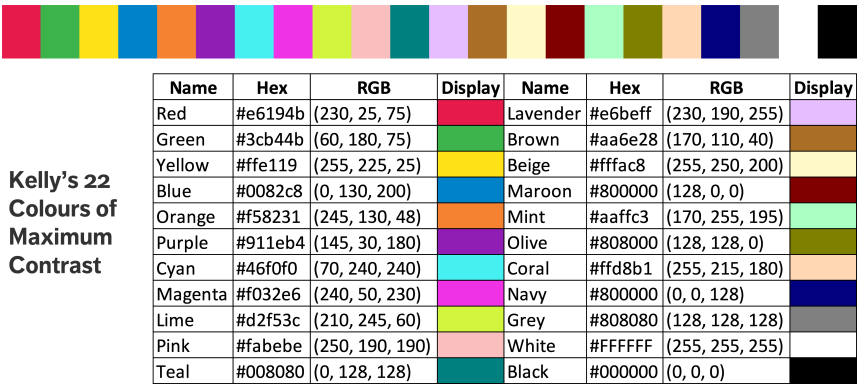
Kelly

**Kelly's 22
Colours of
Maximum
Contrast**

Name	Hex	RGB	Display	Name	Hex	RGB	Display
Red	#e6194b	(230, 25, 75)		Lavender	#e6beff	(230, 190, 255)	
Green	#3cb44b	(60, 180, 75)		Brown	#aa6e28	(170, 110, 40)	
Yellow	#ffe119	(255, 225, 25)		Beige	#fffac8	(255, 250, 200)	
Blue	#0082c8	(0, 130, 200)		Maroon	#800000	(128, 0, 0)	
Orange	#f58231	(245, 130, 48)		Mint	#aaffc3	(170, 255, 195)	
Purple	#911eb4	(145, 30, 180)		Olive	#808000	(128, 128, 0)	
Cyan	#46f0f0	(70, 240, 240)		Coral	#ffd8b1	(255, 215, 180)	
Magenta	#f032e6	(240, 50, 230)		Navy	#800000	(0, 0, 128)	
Lime	#d2f53c	(210, 245, 60)		Grey	#808080	(128, 128, 128)	
Pink	#fabebe	(250, 190, 190)		White	#FFFFFF	(255, 255, 255)	
Teal	#008080	(0, 128, 128)		Black	#000000	(0, 0, 0)	

Figure 5.7: Kelly's 22 colours of maximum contrast [51].

"Requests for sets of colours which would be maximally different for use in color coding, providing maximum contrast for those with deficient color vision, have resultes in the selection of 22 shades from the ICSS-NBS Centroid Colors." [51]

Carter and Carter

Carter & Carter
26 Colour High Contrast

Hex	RGB	Display	Hex	RGB	Display
#F0A3FF	(240,163,255)		#003380	(0,51,128)	
#0075DC	(0,117,220)		#FFA405	(255,164,5)	
#993F00	(153,63,0)		#FFA8BB	(255,168,187)	
#4C005C	(76,0,92)		#426600	(66,102,0)	
#191919	(25,25,25)		#FF0010	(255,0,16)	
#005C31	(0,92,49)		#5EF1F2	(94,241,242)	
#2BCE48	(43,206,72)		#00998F	(0,153,143)	
#FFCC99	(255,204,153)		#E0FF66	(224,255,102)	
#808080	(128,128,128)		#740AFF	(116,10,255)	
#94FFB5	(148,255,181)		#990000	(153,0,0)	
#8F7C00	(143,124,0)		#FFFF80	(255,255,128)	
#9DCC00	(157,204,0)		#FFFF00	(255,255,0)	
#C20088	(194,0,136)		#FF5005	(255,80,5)	

Figure 5.8: Carter and Carter's 26 colours of high contrast [52].

"A method is presented for choosing high-contrast sets of colors for additive color mixers (e.g., CRT). The method is based on data about target-location performance of human observers and adapts the color sets to the gamut of the color processor in use. The method produces any specified number of colors spread as far from each other as possible in color space to maximize contrast. Applications of high-contrast sets of colors are indicated, illustrative results are presented and discussed, and variations of the method are suggested." [52]

There is a French idiomatic expression that does not seem to have a direct equivalent in English (and which is quite *a propos*, given the topic at hand): "*des goûts et des couleurs, on ne discute pas*".[13] We must definitely invest time and effort to think critically about colour use and palettes for data presentations, but it is important to keep in mind that no amount of effort will please **every stakeholder and audience member**.

13: The nearest expression might be "beauty is in the eye of the beholder", but it does not quite convey the same meaning.

Case in point, when we ask our workshop participants which of the schemes of Figure 5.9 they prefer most (or least), we **never** reach a consensus. It is preferable not to over-think this.
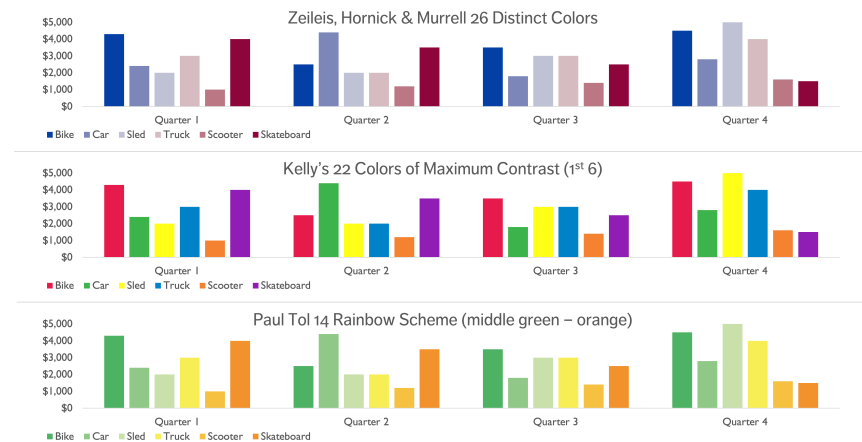

Figure 5.9: Competing colour palettes.

Sizing

We tend to view bigger charts as being more important than smaller ones. Once all charts have been decluttered, **things of equal importance are sized similarly**, and everything else scales to importance.
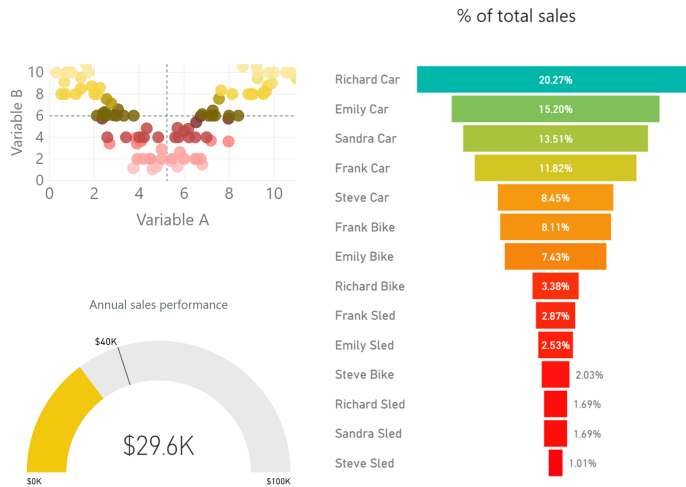

Figure 5.10: Sizing considerations.

In Figure 5.10, we would expect the audience to spend about half the time on the funnel chart, and the remaining time more or less split evenly on the other two charts.

Layout

If a page contains the short term memory limit of 4 charts, how should they be placed? The standard answer, in the Western world at least, is to place them according to the **reading flow**, zig-zagging from the top-left to the bottom-right, as below. This layout is also supported by the Gestalt principle of continuity as it allows for easy and natural alignment of the charts.
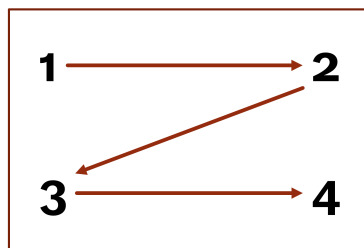
Figure 5.11: The standard 4-chart layout.

Our concentration depletes itself as we scan (and resets itself upon turning to a new page, whether in a book or a dashboard). Charts may be ordered according to some preference, but in order not to overtax the audience, they should get progressively simpler as we approach the 4th slot.

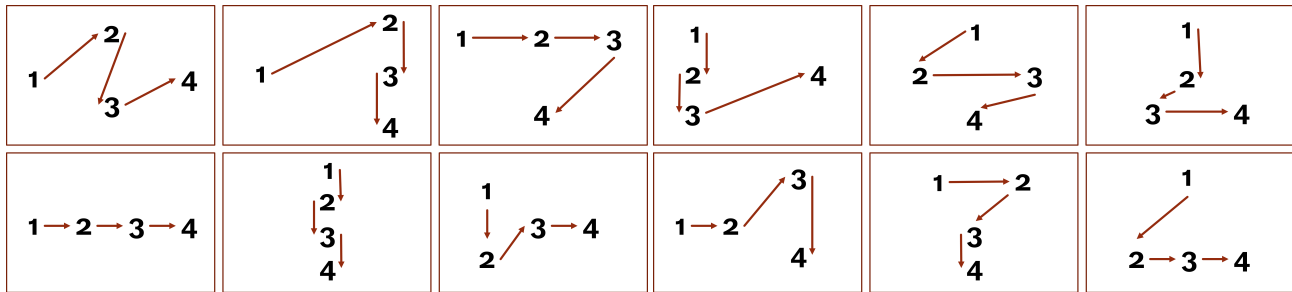This also applies to alternative layouts, such as in Figure 5.12.

Figure 5.12: Alternative 4-chart layouts.

5.3 Decluttering

In the world of charts and storytelling with data, **less is more** (one more time, for good measure): clutter gets in the way of clarity.

Every element on a chart or dashboard page adds to the **cognitive load** required to "read" the presentation; our brains are like all other organs in that they require energy to function[14] – the lower the energy requirement, the better (as a general evolutionary strategy).

So **decluttering** a chart or a dashboard page is simple: all we need to do is **identify** and **remove** anything that is not adding value to the graphics.[15]

Tufte refers to the **data-to-ink ratio** –– "the larger the share of a graphic's ink devoted to data, the better" [2], while Duarte refers to this as "**maximizing the signal-to-noise ratio**", where the signal is the information or the story we want to communicate [53]: the chart should first and foremost be about the data AND the story.

We use the Gestalt principles to **organize/highlight** data in the chart (see Section 4.3, *Gestalt Principles*). In particular, we should:

- align **all elements** (graphs, text, lines, etc.), without relying on the eye (use position boxes and values when possible;
- remove **chart borders**, **gridlines**, **data markers** (unless the marker is the only signal in the chart), and **useless annotations** (ditto);
- clean the **axis labels**;
- label the data **directly** (instead of using a legend box);
- use consistent **fonts**, **font size**, **colour**, and **alignment** (L.C. Muth discusses font choices here ⧉);
- keep the number of colours used in the chart to a **minimum**;
- avoid **rotated text** – everything should either be horizontal or vertical, but not slanted (even then, there is incompatibility with how vertical text is rendered: from top to bottom, or from bottom to top?);[16]
- use a canvas colour, a "main" chart colour, and a text colour that are in **strong contrast** with one another, and
- use **white space** (or empty space, if the canvas is not white) and keep clear margins free of text and visuals – we basically think of white space as a **border** for the charts.

14: Think of cognitive load as the mental effort required to process information.

15: To be fair, this may be easier said than done... especially if we have laboured to bring a chart to fruition. "Everything that we have added was needed!", we might scream in despair. Ask around, when in doubt – and leave the ego at the coat check.

16: In French, such text usually reads bottom to top; in English, it is the opposite. Go figure...

We will revisit this as part of Section 8.3 (*Evolving a Storytelling Chart*).

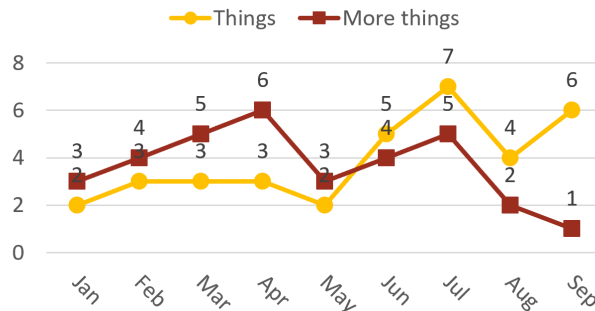As an example, the chart below is clunky and awkward.

Figure 5.13: A cluttered chart.

It is obviously a chart, but it is not clear what we are meant to focus on – there are too many distractions. Compare with the decluttered version of the same chart:
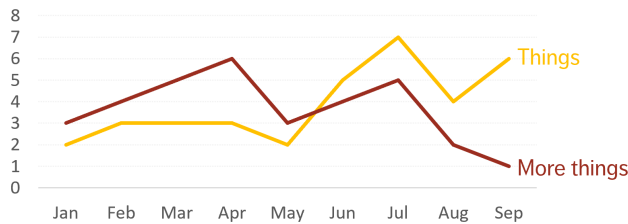

Figure 5.14: A decluttered chart.

Here, we are free to focus on the story: "More Things" used to dominate over "Things", but the roles were reversed in May. Why is that so? The chart does not say, presumably because this is not part of the story being told.[17]

17: Although it easily could have been, in a different context.

5.4 The Layered Grammar of Graphics

It is one thing to recognize when charts are **effective** (see Section 3, *Hall-of-Fame/Hall-of-Shame*), when their aesthetics make them **easy to read** (see Section 4.2, *Gestalt Principles*), and when they are laid out in a dashboard which tells a **compelling visual story** (see Section 8.4, *Anatomy of Storytelling Dashboards*). We can also fairly easily recognize when they **fail to do so** and **provide recommendations**, as needed (see Sections 5.3, *Decluttering*, and 8.3, *Evolving a Storytelling Chart*).

But it is another thing altogether to start with raw data and to **build** such charts in the first place – **storyboarding**, for instance (see Section 17), makes do entirely without charts in the planning stages, while dashboard layout guidelines and Gestalt principles considerations explicitly assume that base charts have already been produced.

The problem is often compounded by software choices made by clients or employers without practitioners' input, which can leave analysts grasping at straws and blindly tossing charts together.

The **grammar of graphics** [54, 55] provides a reliable tool-agnostic path out of the wilderness.[18]

18: Why "grammar"? D.J. Sarkar eluci-dates: "A grammar is defined as a set of structural rules which helps define and establish the components of a language. A language's system/structure usually con-sists of syntax and semantics. A grammar of graphics is a framework which follows a layered approach to describe and con-struct visualizations or graphics in a struc-tured manner. The layered grammar of graphics uses pre-defined components to build charts (instead of random trials and errors)." [56]

Naturally, the **layered grammar of graphics** is best described with the help of a chart; there are multiple versions in circulation, but we like the one found in the QCBS R Workshop Series ☐ (slightly modified, see Figure 5.15).

Figure 5.15: The layered grammar of graphics (modified from QCBS R Work-shop Series ☐.)

There are 7 layers in this version of the grammar of graphics:

1. **data** (required): the plotting observations are found in rows, the vari-ables in columns;
2. **aesthetics** (required): the mapping of the dataset's variables to the chart's design elements (position, shape, size, colour, etc.);
3. **geometry** (required): the type of chart on which the data is represented (bar, line, scatterplot, pie, etc.);
4. **facets** (optional): the subsets of the data represented on the chart (variable levels);
5. **statistics** (optional): the measures that help provide chart context (centrality, dispersion, trend, etc.);
6. **coordinates** (required): the chart's plotting space (axes, scale, etc.), and
7. **themes** (required): the design choices that are used to create a visual identity (fonts, colours, etc.).

In practice, layers 2 and 3 are often decided on simultaneously; and while layers 6 and 7 are required, they are often left unspecified in favour of default settings in applications (the most famous of which is without doubt Wickham's celebrated `ggplot2`, see Section 12, *ggplot2 Visualizations in* `R`).

Considering the layered grammar of graphics **prior to** embarking on analysis and visualization can help analysts save precious process time; a good way to practice (and thus reduce the required time) is to get into the habit of **deconstructing** charts using the grammar (which is the opposite of its habitual use in **building** charts).
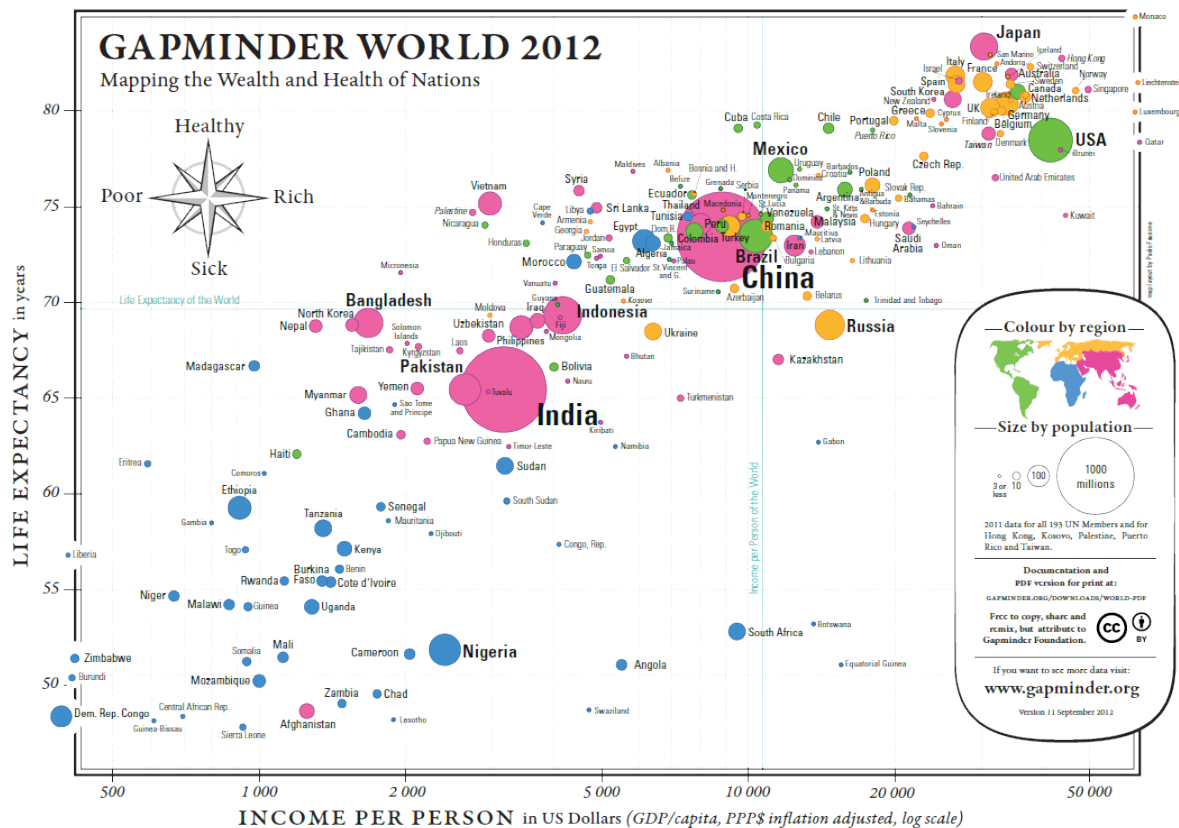
Figure 5.16: Health and Wealth of Nations (2012), Gapminder Foundation ⬈ .

Deconstructing a Chart Using the Layered Grammar of Graphics

We have already studied the chart in Figure 5.16, from the perspective of the fundamental principles of analytical design (see Section 1.3). What can be said about it using of the layered grammar of graphics?

It uses a subset of the 2012 Gapminder ⬈ **data**,[19] with a bubble chart **geometry**, allowing for the use of 4 design elements: position (2), size, and fill. The **aesthetics** mapping between the geometry and the data is:

19: Which contains socio-demographic information (upwards of 500 variables) for the planet's nations, for years ranging from 1800 to 2020.

- position → income per person (horizontal) and life expectancy (vertical)
- fill → region, size → population

This chart admits no **facet** as all observations are plotted in the same space (unique set of axes). Two **statistics** are added to the canvas: the world life expectancy (equator) and the world income per person (prime meridian).

On the **coordinates** front, life expectancy is displayed linearly (as this quantity is distributed more or less uniformly on its range), whereas income per person and population are displayed logarithmically (due to the presence of long tails in the positive direction for both quantities). Finally, the chart is displayed using an old Gapminder World **theme**, providing a nice twist on the traditional cartographer's map.
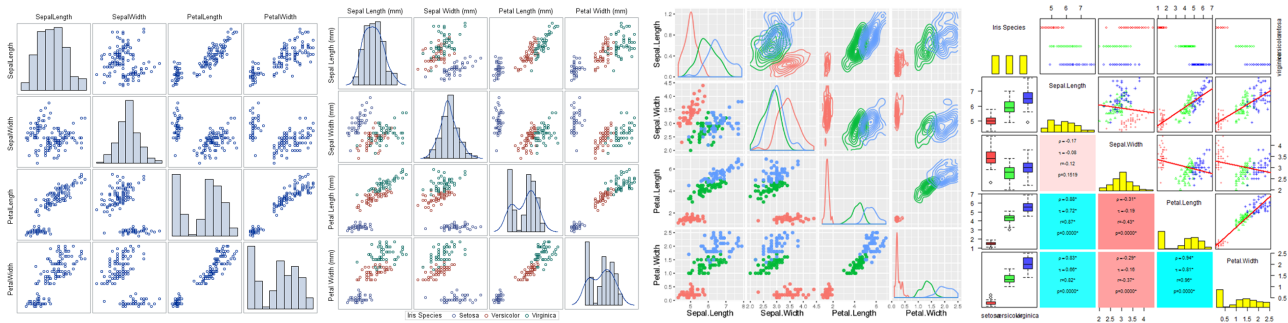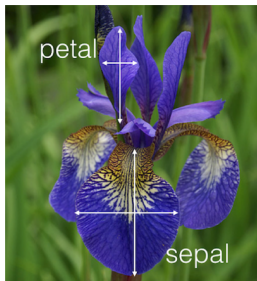
Figure 5.17: Scatterplot matrices of the iris dataset; the clutter level increases from the first to the last chart (the first three images were made with SAS and R; the last one is from an unkown author).

5.5 Additional Examples

Decluttering

20: The dataset's attributes are: **petal width**, **petal length**, **sepal width**, **sepal length**, and **species** (virginica, versicolor, setosa). A "description" of these features is provided by the picture below (from Wikipedia):

Iris is a genus of plants with showy flowers. The `iris` dataset contains 150 observations of 5 attributes for specimens collected by Anderson, mostly from a Gaspé peninsula's pasture in the 1930s [57].[20] This dataset has become synonymous with data analysis,[21] being used to showcase just about every algorithm under the sun. [22]

Figure 5.17 shows various scatterplot displays of the data, ranging from quite simple and uncluttered (left-most) to chock full of information in a chaotic soup of clutter (right-most). The first is probably too sparse to be of use, but the last one will drive anybody unlucky enough to gaze upon it mad.[23] The perfect visual would probably combine the best features of the second and third versions.

The next example shows how to declutter a chart sequentially (from [4]).

21: To the point that the standard joke is that "it's not necessary to be a gardener to become a data analyst, but it helps".

22: It is with a certain amount of sadness that we also elect to use it in this section. But note that the iris dataset has started being phased out in favour of the penguin dataset [58], for reasons that do not solely have to do with its overuse (take a look at the name of the journal that published Fisher's paper).

23: MAD!!

Figure 5.18: Decluttering a chart: cluttered chart (top); remove border and gridlines (middle left); remove markers (middle right); clean-up axis labels and legend (bottom left); colour code the lines to get a decluttered chart (bottom right) [4].

The Grammar of Graphics

Here is a brief grammar of graphics deconstruction of six additional charts built from the Gapminder dataset.[24]

24: These are obviously not the *only* charts that could be displayed.
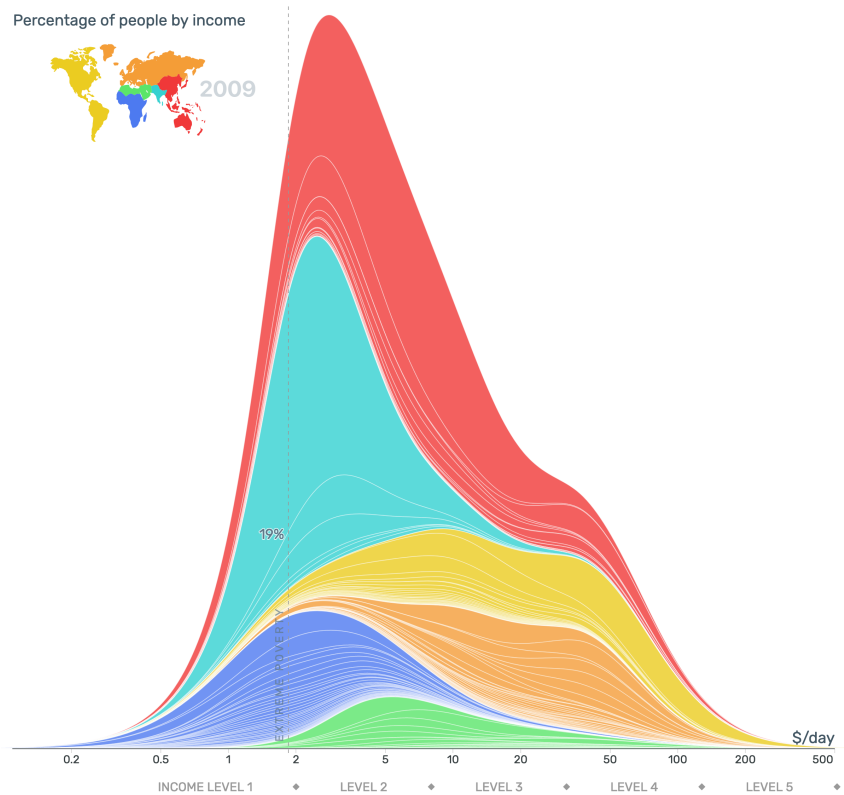

Figure 5.19: Proportion of population in income levels, in 2009 Gapminder Foundation ☐ .

Data: Gapminder countries, 2009

Geometry: stacked density chart

Aesthetics:

- x: daily income
- (y: percentage per country)
- fill: region

Facets: none

Statistics: extreme poverty proportion

Coordinates: logarithmic (x)

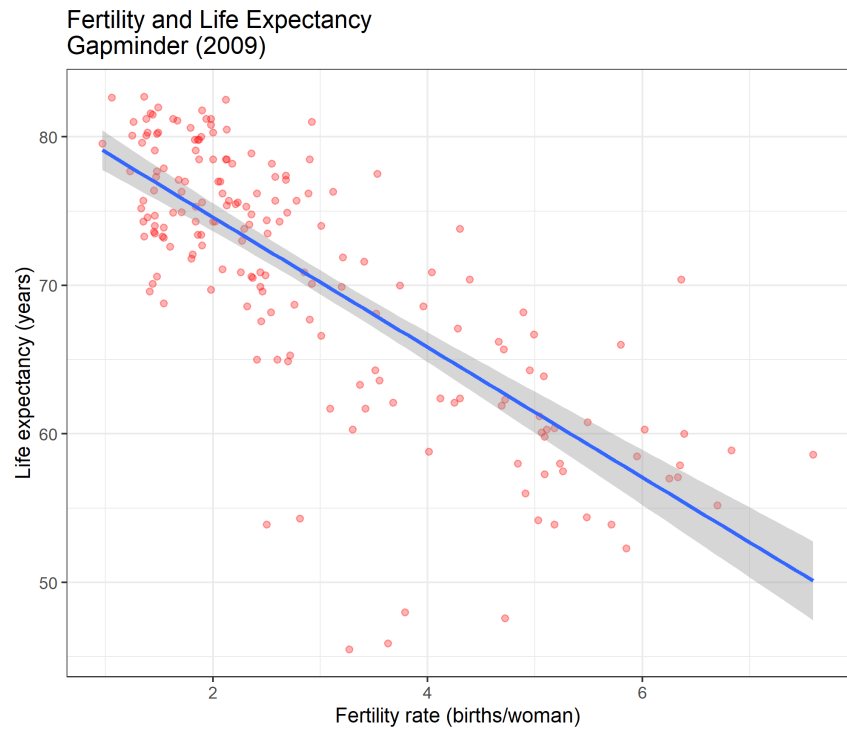**Theme:** Gapminder Tools; adornment (extreme poverty vertical line)

Figure 5.20: Life expectancy relative to fertility rates, in 2009 [personal file].

Data: Gapminder countries, 2009

Geometry: scatterplot chart

Aesthetics:

- x: total fertility
- y: life expectancy

Facets: none

Statistics: line of best fit, confidence interval for mean response
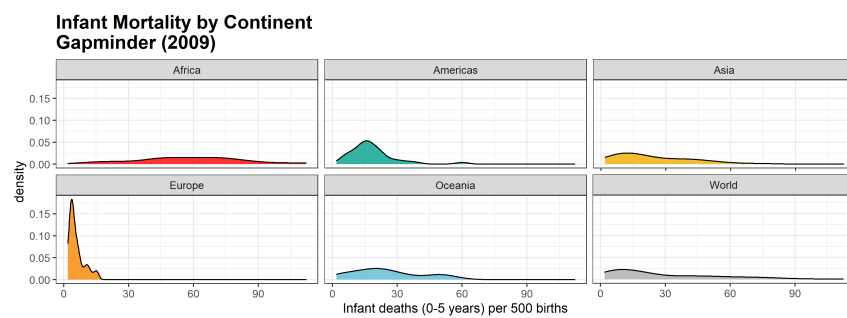
Coordinates: linear (x, y)

Theme: ggplot2 default

**Infant Mortality by Continent
Gapminder (2009)**

Figure 5.21: Infant mortality by continent, in 2009 [personal file].

Data: Gapminder countries, 2009

Geometry: density chart

Aesthetics:

- x: infant mortality
- fill: continent

Facets: continent

Statistics: none

Coordinates: linear (x)

Theme: Darjeeling1

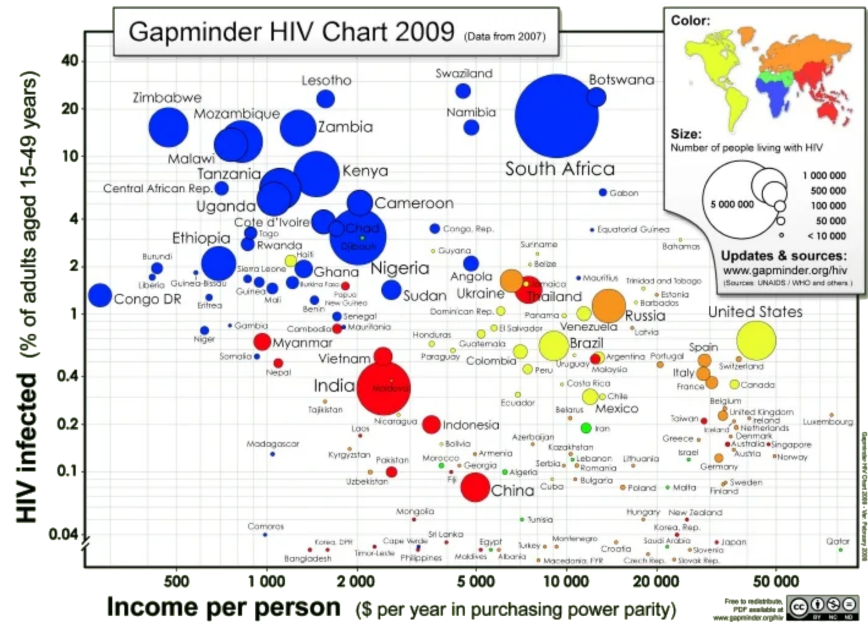**Figure 5.22:** HIV infection rates relative to income per capita, in 2007 Gapminder Foundation ↗ .

Data: Gapminder countries, 2007

Geometry: bubble chart

Aesthetics:

- x: income per person
- y: HIV infection rate
- fill: WHO region
- size: HIV infected population

Facets: none

Statistics: none

Coordinates: logarithmic (x, y, size)

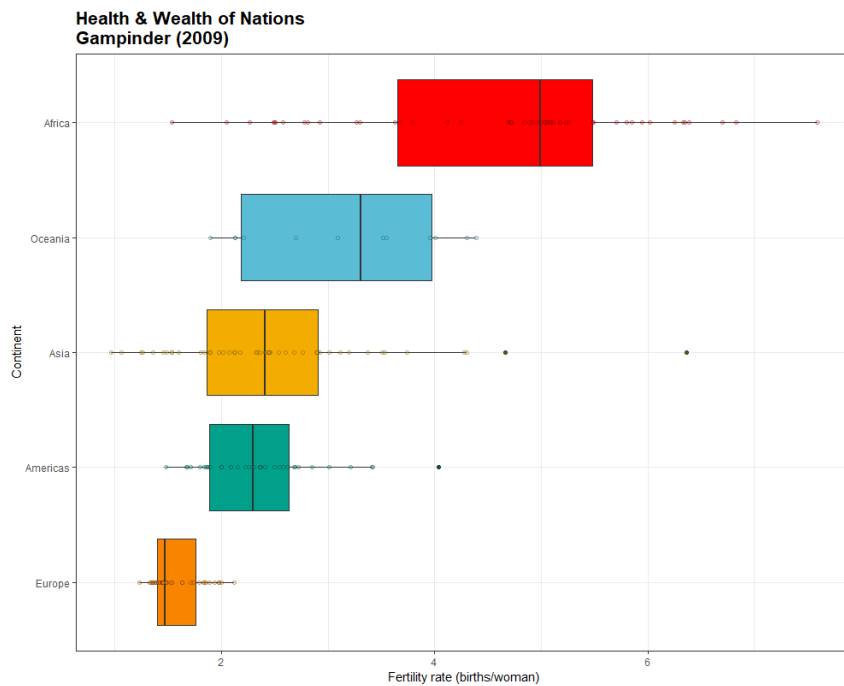**Theme:** old Gapminder World

Figure 5.23: Fertility rates by continent, in 2009 [personal file].

Data: Gapminder countries, 2009

Geometry: boxplot chart

Aesthetics:

- x: fertility rate
- y: continent
- fill: continent

"Facets": continent (?)

Statistics: 5-pt summary (built-in)

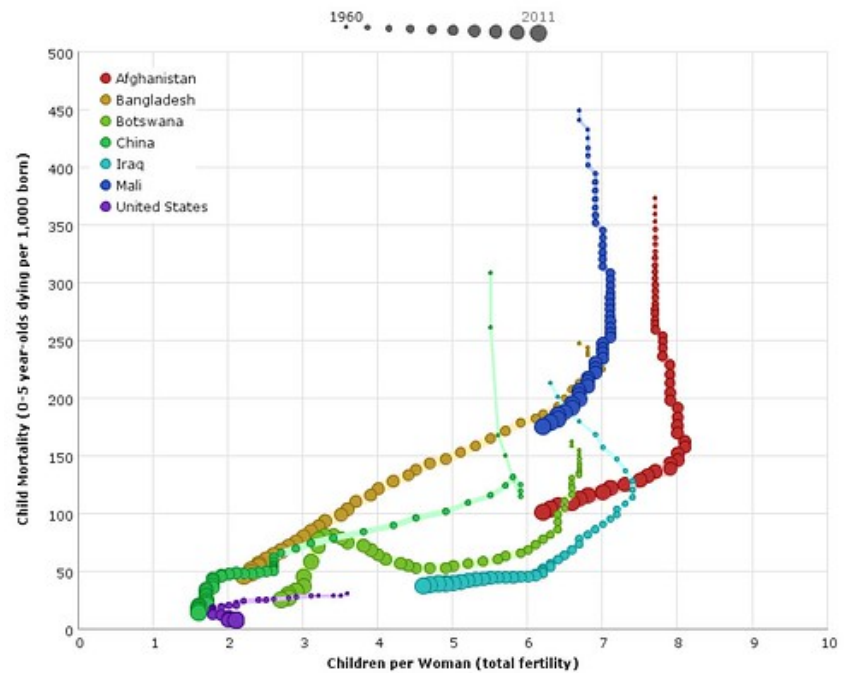**Coordinates:** linear (x)

Theme: Darjeeling1

Figure 5.24: infant mortality relative to fertility rates [author unknown].

Data: selected Gapminder countries, 1960-2011

Geometry: line chart

Aesthetics:

- x: total fertility
- y: percentage per country
- colour: country
- size: year

Facets: none

Statistics: none

Coordinates: linear (x, y, size)

Theme: custom (author unknown)