

8 Effective Storytelling Visuals

In a professional setting, the stories of interest will most likely be **data stories**.

In this chapter, we discuss how **visuals** mesh with stories, how we can **evolve** a simple chart into a data story, and how to build a dashboard with a message that **easily understood** by (and that will stay with) its intended audience.

8.1 Stories and Illustrations

In the seminal *Making Comics* [47], graphic novelist Scott McCloud makes an interesting point regarding the **flexibility** of use of visuals in storytelling:



He is specifically talking about comics, but we believe the idea also applies to more general storytelling formats: **any principle is at best a guideline**, and it is not too difficult to imagine situations where they can be violated in the interest of getting the message across (see previous section).

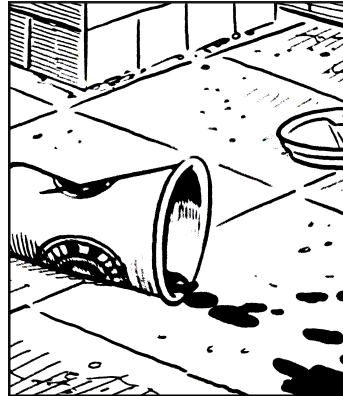
That being said, it does help to have a few ideas about how one may incorporate visuals into data presentations: we will discuss, in particular, how we can combine **words and images** to create impactful data stories, and what **various visual storytelling choices** can do to improve the audience's comprehension of the (data) story.

8.1 Stories and Illustrations . . .	149
Words and Images	150
Visual Storytelling Choices	151
8.2 Data and Stories	154
Storytelling Stages	156
Falsification	157
Data Stories Revisited . . .	158
8.3 Evolving a Story Chart . . .	160
Data Storytelling Tropes . .	162
Examples	165
8.4 Anatomy of a Dashboard . .	167
Dashboard Types	168
Audiences (Reprise)	170
Presentation Requirements	171
Storyboarding	171
Dashboard Narratives . . .	173
Putting it All Together . . .	173

Words and Images

It is said that a picture is worth a thousand words. How so? “Words bring an **unparalleled level of specificity** to all kinds of situations: there is no image so vague that words cannot lock it into a **desired meaning**.”

McCloud presents the following example [47]:



What are we seeing here? A coffee cup, certainly. Thrown to the ground by an apathetic city dweller who could not be bothered to place it for a refuse bin? Or perhaps the coffee was too hot and scalded a customer, who dropped it to the ground in agony? Or any of a variety of other scenarios?

Notice how the caption below collapses all interpretation to a single one, leaving no doubt as to what it is that the image is conveying.

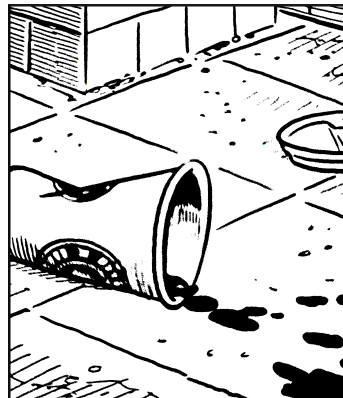


Figure 8.1: "On the bright side, I got my caffeine. On the not-so-bright side, we got mugged on the way home." [47]

Additionally, some specific concepts and names can only be **clearly** expressed through words. Who could ever correctly guess what is going on in Figure 8.2 without the caption?

Data story audiences are not always privy to the ins-and-outs of the analytical process that lead to a chart; what may appear obvious to the analysts may be opaque to stakeholders. In general, it is preferable to err on the side of caution and use text wisely, leaving **no room for ambiguity**. Coyness has no room in data stories.



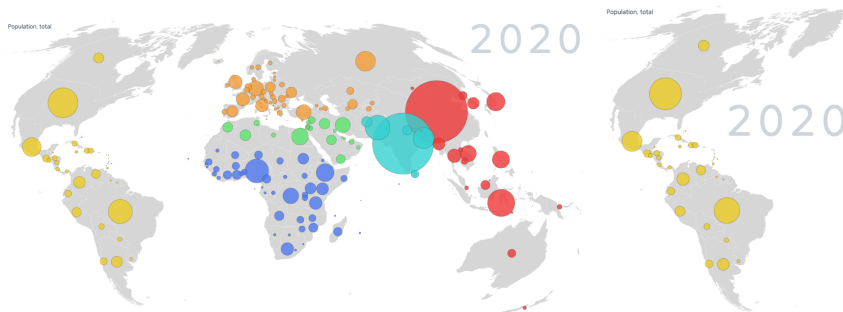
Figure 8.2: "Look! It's Kelly Donovan, twin brother of the guy who played Xander on *Buffy the Vampire Slayer*, plus Humprey Bogart wearing a Freddy Mercury mask, and a robot duplicate of former U.N. Secretary-General Boutros Boutros-Ghali!" [47]

Visual Storytelling Choices

In Chapter 4 (*The Mechanics of Visual Perception*), we introduced the *Gestalt principles*, with the main objective being to reduce the cognitive load associated with **reading** (or parsing) **charts**.

The visual storytelling choices we discuss here are mostly of a different nature: they are concerned with **clarity in communication**.¹ Communicating with **clarity** means that audience comprehension remains the **ultimate goal**.

- The choice of **moment** helps “connect the dots”, which is to say that we should show **only** what matters to the story, as in Figure 8.3 – if we are interested in the distribution of population in the Americas, the chart on the left contains irrelevant information.²



1: Our treatment is based on an idea from [47].

2: Although who knows if it might not become relevant at some point in the future?

Figure 8.3: Population distribution in the Americas in 2020 [106].

- The choice of **frame** helps create and direct the audience's focus, as in Figure 8.4. In the chart on the left, we do not easily see what we should be focusing on; in the chart on the right, we quickly understand that we should turn our attention to the month of May.

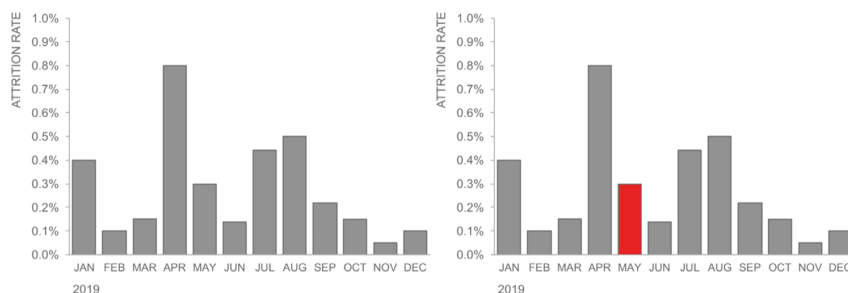


Figure 8.4: Monthly attrition rate in May 2019 [4].

- The choice of **image** helps select the right charts for the story, with emphasis on simplicity and an ability to convey the message, as illustrated in Figure 8.5, where the two charts are different but built from the same data – depending on the context and/or on the audience, either one of those could be preferable to the other one.

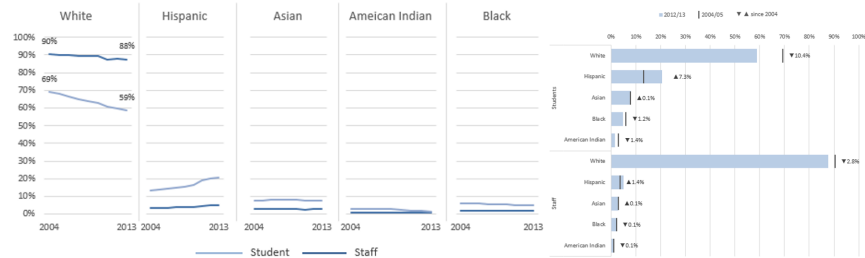


Figure 8.5: Washington State University; percentage of staff and students by ethnicity, 2004-2013 [author unknown].

- The choice of **word** (text) helps communicate ideas clearly and persuasively, in seamless combination with the charts, as in Figure 8.6 – without the text (as we had seen in Figure 8.4, it is not entirely clear what we are supposed to get from the chart without a choice of frame or word).

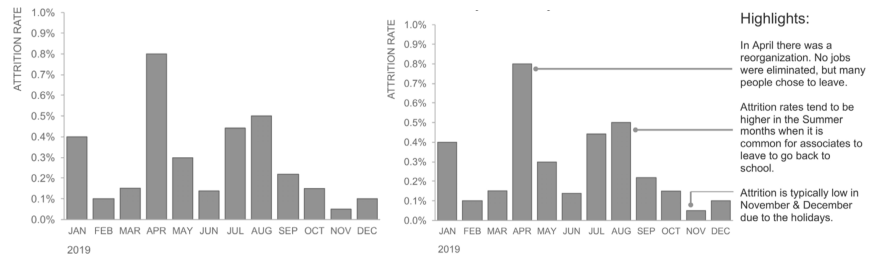


Figure 8.6: Monthly attrition rates, revisited (2019) [4].

- The choice of **flow** helps guide the audience from one chart to the next, from one page to the next, and creates a transparent and intuitive “reading” experience by arranging the pages in a dashboard, the charts on a page, and the elements within charts intelligently.

Decisions having to do with *moments*, *frames*, and *flow* are likely to be made in the **dashboard planning stages**, while *images* and *words* decisions are usually being made right up to the finish line.³

For the most part, data stories are built according to one of three general approaches:

- we start with a **rough sketch** of the dashboard (moment, frame, flow), then come up with the **narrative** (word), and finally populate the dashboard with **charts** (image);
- we start with a **full ‘script’/storyboard** (moment, word), then use that to do a **rough layout** of the dashboard (frame, flow), then populate the dashboard with **charts** (image), or
- we create a **finished chart** (moment, frame, image, word) with no idea as to what else will show up on the dashboard until we create **another chart** (flow), and **so on**, and **so on**.

3: “The best-laid plans of mice and men oft go astray” definitely holds for data stories – it is going to take more time to get it right than you think it is going to take, even when you take into account that it is going to take more time than you think it is going to take (see “Hofstadter’s Law”, [107]).

Needless to say, we do not recommend the latter approach; nonetheless, it is probably the most commonly-used; it leads to analysts and stakeholders trying to **shoehorn a story to the available charts**, the exact opposite of what storytelling with data should be.⁴

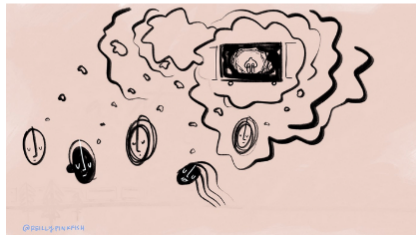
Visual storytelling, then, sinks or swims based on how **text** and **charts** are integrated into the final product.

- In **text-specific** combinations, the text provides all that is needed to know while the charts illustrate some aspects of the story that is described, as in Figure 8.7:



I have a story I'd like to tell you. It's about a train, and a group of people who live on that train and know of nothing else.

This train has been moving since anyone can remember. The people on the train can't imagine a time when the train wasn't moving, and when they were not on the train. Everyone works to keep the train moving. The train never stops.



It never stops. It cannot stop.

People on the train live in constant churn. The work to keep the train moving is hard, and inhumane. On the train, people are treated with cruelty and oppression. Some are treated worse than others. But nobody is truly living.



Sometimes they get breaks, but it is hard.

One day, a fire breaks out in one of the carriages of the train.



There is panic. The fire spreads throughout the whole train... Without getting off the train everyone is going to die.

Then the impossible happens.



The brakes no-one believed existed start to work. In the emergency, no-one notices how extraordinary it is that the train is stopping. They're too focused on the fire. The old rules go out the window.

For years on the train, the "worker class" of people have been dying from the awful conditions of the work they have to do on the train. They sleep in the aisles and sometimes have nowhere to sleep at all.

Suddenly, there are orders to house them and treat their ailments.

The train stops, and people begin to get off. Apart from the sound of the fire, suddenly there is a great silence.

4: We understand why that happens in practice: analysts are often overworked and burdened with unrealistic timelines and expectations. Nevertheless, this is not a sustainable strategy and the harm caused by this "fly-by-night" approach has long-lasting consequences for organizations that rely on these stories for strategic decision-making.

Figure 8.7: Extract from *COVID-19: The Story of the Impossible Train* (Illustrated) [108].

- In **chart-specific** combinations, the charts provide all that is needed to know while the text accentuates some aspects of the story that is shown, as in Figure 8.8.
- In **duo-specific** combinations, both text and charts telling roughly the same story, as in Figure 8.9.
- In **intersecting** combinations, text and charts work together in some respects but also contribute to the story independently.
- In **interdependent** combinations, text and charts combine to convey an aspect of the story that neither could convey alone.
- In **parallel** combinations, text and charts follow seemingly different storylines, without intersecting.

Cumulative vaccination doses administered in Israel, UAE, UK and US
Cumulative doses administered per 100 residents • Data last updated 24 Feb

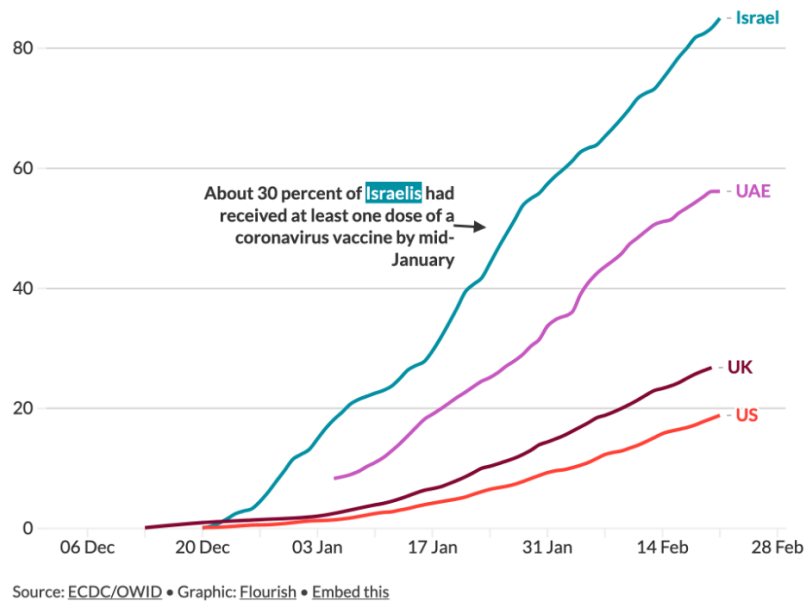


Figure 8.8: COVID-19 vaccination rates in 4 countries [68].

5: Analysts and storytellers of the future: take note!

The last three combinations are infrequently used in data storytelling;⁵

In the best data stories, text and charts are like partners in a dance and each one takes turns leading. When both partners try to lead, the **competition** can subvert the overall goals (e.g., clarity); when each partner knows their roles and **supports the other's strengths** (which may differ from one dashboard page to the next, or from one story to the next), data presentation dashboards become fantastic storytelling media (paraphrased from [48]).

8.2 Data and Stories

There are very few constraints associated with storytelling, in general: **all** story ideas, stylistic choices, or delivery modes are **potentially in play**, assuming that the choices are appropriate to the story function (education, entertainment, etc.).

Even when executed flawlessly, such stories are not necessarily good or compelling; they still have to be shared at the right time and the right place to find an appreciative audience, which could prove to be short-lived and/or quite small.⁶

But as long as a tale is recognized as an honest **attempt** at a story (using whatever metric the appraiser thinks is appropriate to do so), humans agree that they are dealing with a story.

Not so for data stories: we may only tell stories that are **supported by the data**. No flight of fancy, no faking data for the sake of the cause, no ignoring contradictory observations because it “makes for a better story”.

6: It is far from a perfect science: J.K. Toole's modern classic *A Confederacy of Dunces* only found a publisher posthumously, 11 years after his early death at the age of 31; **cubism** was widely derided as a movement when it first came to prominence, but it is firmly established as “real art” in the current era; **achromatic** music is still finding its footing and feels like it might never really outgrow its experimental/novelty status.

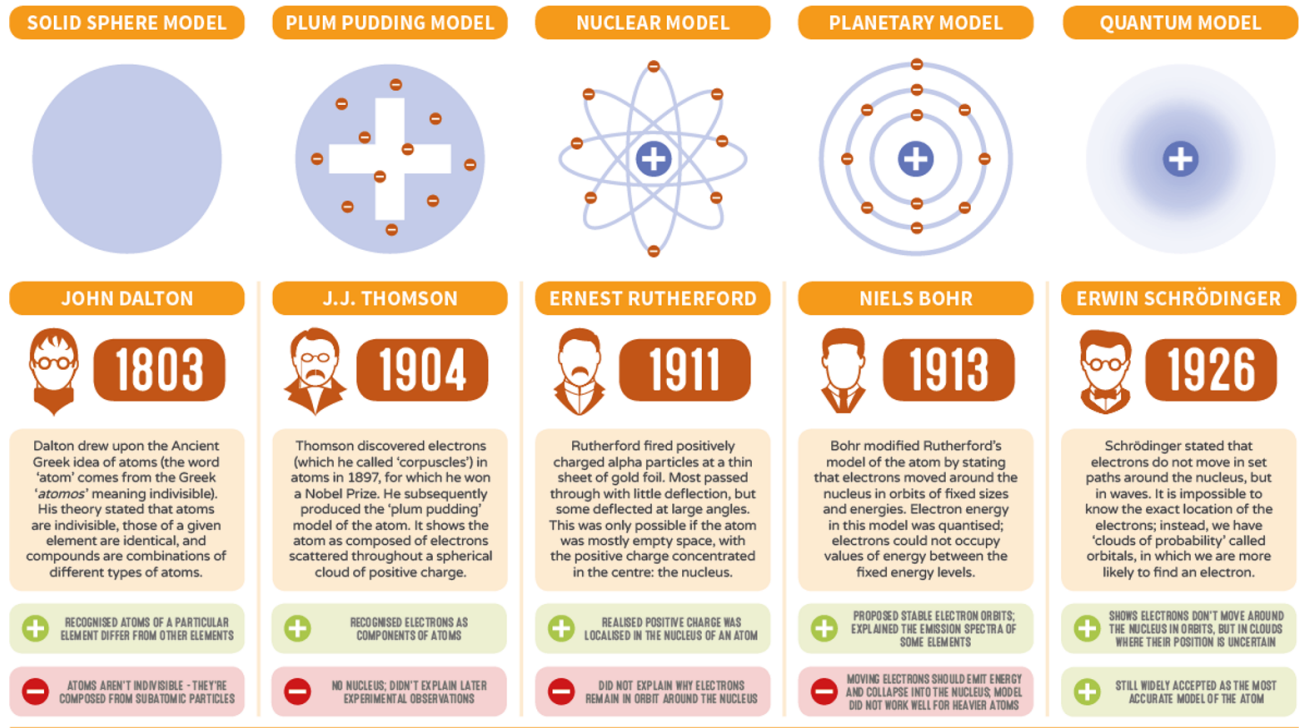


Figure 8.9: A history of the atom: theories and models. How have our ideas about atoms changed over the years? This graphic looks at atomic models and how they developed [Compound Interest 2016](#) .

As we have discussed above (see Figure 7.10), there are different ways to be “supported by the data”. Indeed, data analysts have **agency** – they select:

- the question(s) to answer;
- what data to collect;
- how to clean that data;
- which analytical method(s) to use;
- which feature(s) to include in the analysis;
- on what part(s) of the data to focus;
- what visualization software to use to display the data, etc.

These all have an effect on the stories that can be told with data, on the stories that could be told about the situations and events **represented by the data**.

Data analysts would do well to remember that they are human, that they come to the process with a whole slew of pre-conceived notions and cultural baggage. This does not preclude them having **preferences** about what the results will show or what the story **should** be about, of course, but this is another reason why having a **tsarina of common sense** and **diverse teams** around can act as safety checks.

As an illustration of the impact that data analysis agency can have on data stories, consider the example of Figure 8.10.

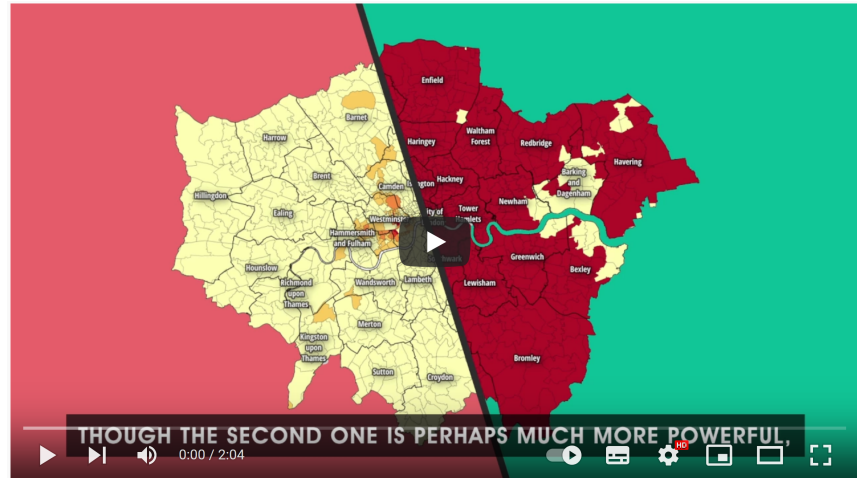


Figure 8.10: Why Numbers Matter, Episode 2, *Do Maps Lie?*, by the Sheffield Methods Institute.

Scoping vs. Exploration vs. Explanation vs. Persuasion

When working with data, we create visualizations at **multiple stages** in the process; the degree of **polish** we expect from such charts, as well as their **number**, depend on the specific stage. This is not without reminding one of the general process underlying **investigative journalism**:

1. initially, we **scope** out the area of investigation (objectives, data collection, story); at this stage, data charts usually show up in the form of conceptual doodles and proofs-of-concepts using the most basic of visualization software capabilities (Excel, base R, `matplotlib`, etc.);
2. then we **explore** the situation's *phase space* and the data we have collected about it; many more charts are created than will be used in any final deliverable (perhaps at a ratio of 10-to-1?), some of them **entirely at random** (to provide a baseline and some protection against pre-conceived notions and implicit biases); little effort is made on the aesthetic side of things as chart quantity and diversity are more valuable than chart quality at this stage; automatic and batch visualization capabilities is preferred (base R, `matplotlib`, etc.);
3. we may use the outcome of this exploration to to answer the original analytical questions/objectives and **explain** the situation to our satisfaction; perhaps 1 in 5 exploration charts will be selected for **beautification** and **decluttering** *via* the Gestalt principles guidelines, based on relevance and on the planned data story's outline and priorities; the *tsarina* of common sense comes into action at this stage, both in the selection process and as a critique of aesthetic choices (`ggplot2`, `seaborn`, etc.);
4. finally, the most appropriate charts will be incorporated into a data presentation dashboard and woven into a data story used to **persuade** the audience and stakeholders into a suggested course of action to be taken with respect to the situation; the layout and construction of the final product is often left in the hands of **graphic designers** who use specialized software (Power BI, Tableau, Photoshop, video editing software, etc.) to create vivid stories.

Falsification

In [109], Karl Popper differentiated **science** and **pseudo-science** by introducing the notion that scientific theories had to be **falsifiable**: this does not mean that theories had to be false, but that there should be some way to determine if they are false, namely that they should make predictions not just about what would be observed if they were true, but **also if they were false**.

For instance, “every person on Earth dies at most 500 years after they were born” is a falsifiable statement: all that is needed to contradict the statement is to exhibit a human being that has been alive for more than 500 years ago; “everybody dies at some point”, although almost certainly true, is not a falsifiable statement: any human being exhibited as contradictory evidence might simply not have died *yet*.

But this approach is **prescriptive**, suggesting that scientists should eschew **inductive processes**⁷ in favour of **deductive processes**:⁸

“no matter how many observations are made which confirm a theory there is always the possibility that a future observation could refute it. Induction cannot yield certainty. [...] Science progresses when a theory is shown to be wrong and a new theory is introduced which better explains the phenomena. Scientists should attempt to disprove their theory rather than attempt to continually prove it.” [110]

In this view of the discipline, the scientific method is a set of tools that help us get progressively closer to “the truth” (assuming “perfect” experiments and measurements), but which we can never be certain has been reached.

This approach has been criticized by philosophers:

- the **Duhem-Quine thesis** [↗](#) argues that since a theory is usually a complex collection of statements, the falsification of a single statement of secondary significance is not sufficient to reject the **full theory** [111];⁹
- Thomas Kuhn argues that far from progressing gradually in small increments, science knows long periods of status quo, shaken rather frequently by seismic **paradigm shifts** [↗](#) [112].¹⁰

Be that as it may, “Popperism” can also be applied to data storytelling: it should in principle be possible for analysts and storytellers to **imagine some type of data that could falsify the story they are telling**. If this cannot be done, then the story and the data are not really connected.

As an example, in the 2014 Bill Nye-Ken Ham debate around “**Is Creation a Viable Model of Origins?** [↗](#)”, the participants were asked what would make them change their minds on creationism: Ham said that he was committed to his views, being a Christian, whereas Nye’s response was that a single piece of evidence to support creationism (rock layers can form in just 4,000 years, say) would cause him to change his mind immediately. One of them has the potential to be a data storyteller, the other, *emphatically* does not. We leave you to connect the dots and determine which is which.

7: observation > pattern > hypothesis > theory

8: theory > hypothesis > observations > confirmation

9: **Common descent** (all living organisms on Earth descend from a common ancestor [E. Darwin]) is often packaged with the theory of evolution by natural selection; common descent implies evolution by natural selection, but it is secondary to it. The former could be disproven by finding life that does not use DNA and RNA for information storage and retrieval (or by finding extra-terrestrial life, see **Rational Wiki** [↗](#)); while this would reduce the **plausibility** of the latter, it would not challenge its validity.

10: Going from Newtonian mechanics to quantum/relativity theory requires a giant leap, not a series of small ones (fixed maximum velocity c , the quantum realm running on stochastic processes, etc.); moving from Lamarckian biology to Darwinian natural selection requires abandoning the former model entirely, etc.

Data Stories Revisited

In popular culture, stories can be **fictional**, **real**, or a **blend of both**. In the workplace, we classify stories as **active** or **potential**:

- **active stories** are **continuously happening**, and we need to understand their context to understand if they are important (if they are worth telling/sharing) – some already have outcomes (“We got funded!”), while some outcomes have yet to happen (“Will we get funded? What happens if we don’t?”).
- **potential stories** are those we **want to see happen** – we craft these stories to attempt to convince the various **actors** to reach a specific conclusion (“Please give us some funding!”).

In professional activities, we are typically:

- **uncovering active stories** that are unfolding in real-time (“My project just got cancelled!”);
- **crafting potential stories** to achieve/obtain an outcome (create and present a Treasury Board submission to obtain funding), and
- **working out what possible outcomes** an in-process, active story might have (“Is there a higher likelihood of success if we do A instead of B?”).

In all cases, we are attempting to track the context, actors, events, and outcomes which will help us to identify and obtain the data needed to create data stories which, in turn, help us tell the actual story.¹¹

We can revisit the map of the sky, say, and find **patterns** (see Figure 7.2) around which we can build stories (active and potential), but we can also improve our understanding of our **place in the Universe** with astrophysical data, as in Figure 8.11.

11: **Data stories** help us explain active stories, or to articulate potential stories. Note however that the data required to be successful in these endeavours might not be accessible to the data storytellers.

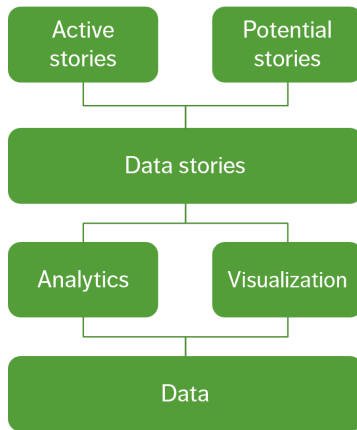
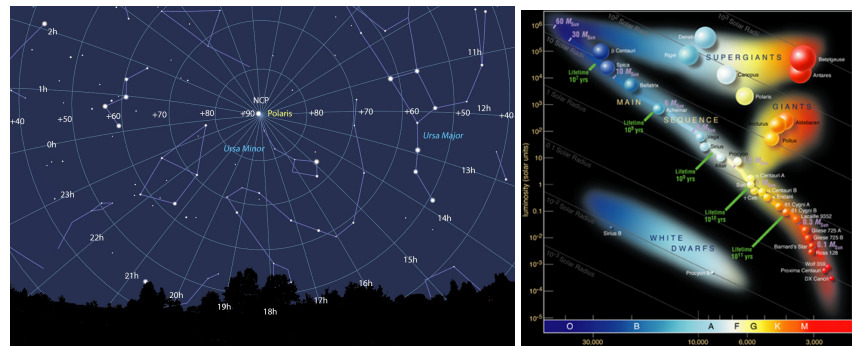


Figure 8.11: Constellation and coordinates ☞ (stories, left) and Hertzsprung-Russell ☞ diagram (data story, right).



12: A shame, really.

When we analyze the data, it becomes obvious that the fantastical interpretation of the stars is **meaningless**¹² but other **meaningful** stories begin to emerge based on the analysis and interpretation of the data.

It is entirely possible (likely, even) that the Universe is too strange to be understood entirely *via* monkey stories (see *The Story Trap*, p. 126), but it is hard to imagine how we could have reached our current understanding (and any future understanding) without first telling stories about Cassiopeia and her myth sisters all those years ago.¹³

13: Stories lead to insight, although the impact is not always immediate.

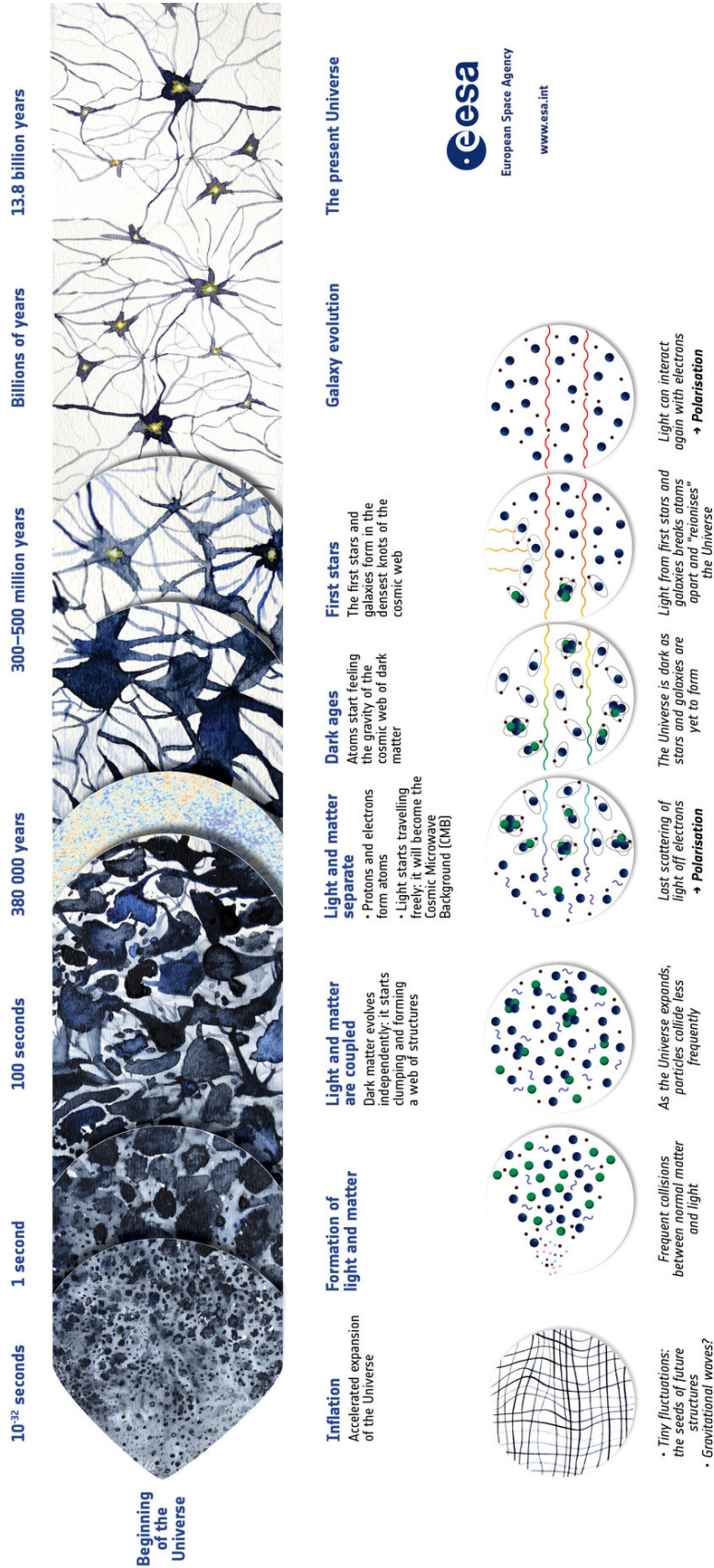


Figure 8.12: Cosmic history: the natural evolution of night sky stories? [European Space Agency ↗]

8.3 Evolving a Storytelling Chart

While **great graphics** are not sufficient for **great data stories**, they are at the very least a requirement. Following the work of [4, 113], we refer to the process that takes raw numbers and analytical results and progressively transforms them into charts worthy of data stories as “**evolving a storytelling chart**”:

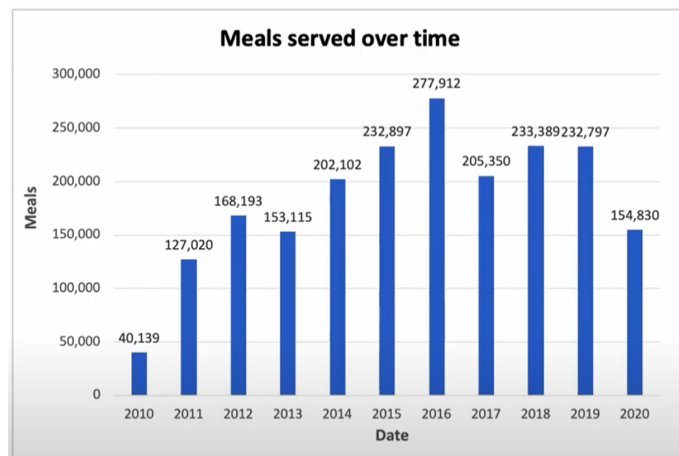
1. **numbers and tables** are *visualized* into ...
2. **ugly graphs**, which are then *decluttered* into ...
3. **simple graphs**, which are further *enhanced* into ...
4. **appealing graphs**, which are finally *integrated* into ...
5. compelling **data stories**.

We base the following illustrative example on [4]: a charity organization records the number of meals they serve to the needy. From 2010 to 2020, the **counts** are as in the table below:

Meals served over time

Campaign Year	Meals Served
2010	40,139
2011	127,020
2012	168,193
2013	153,115
2014	202,102
2015	232,897
2016	277,912
2017	205,350
2018	233,389
2019	232,797
2020	154,830

Most people will unsurprisingly fail to identify *anything* from these numbers: they are, after all, just numbers. A downright **unappealing**) chart shows that there is structure in the table:

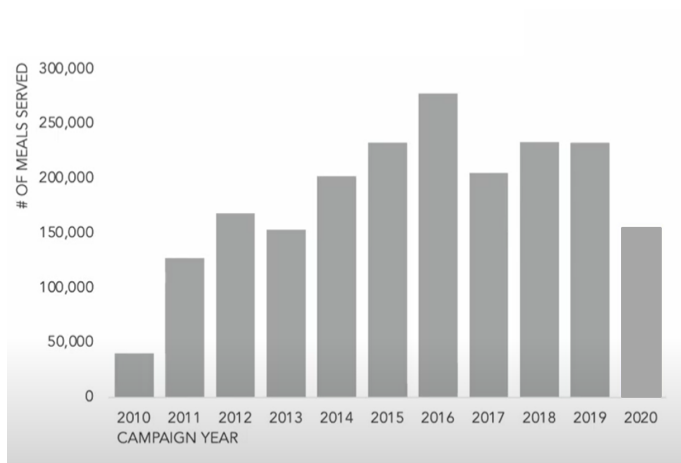


We see that the number of meals served increased roughly linearly from 2010 to 2016, and that there was significant drops between 2016 and 2017, and 2019 and 2020, for instance: visualizing the data can provide some ideas for any eventual data story.

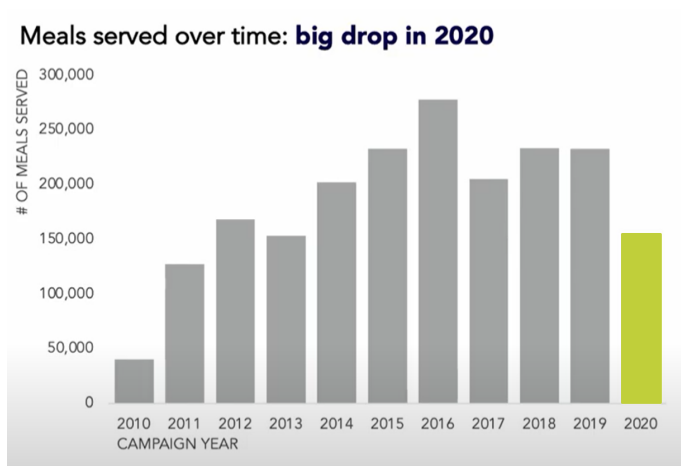
The chart is decluttered by:

- removing the bounding box,
- increasing the width of the bars,
- reducing the number of colours,
- renaming the axes, and
- eliminating the number labels on the bars.

This leads to a **simple** and **clean** chart:

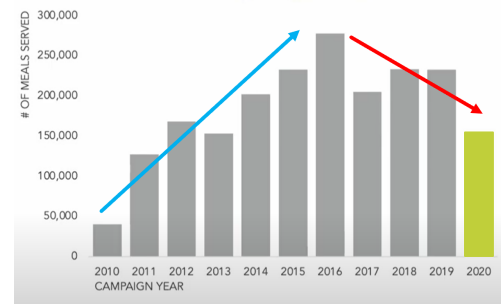


Let us assume that we have chosen to focus on the 2020 drop in number of meals served as the primary data story: we use the Gestalt principle of the focal point to highlight the observation of interest, and add a brief title and description of the insight, which triggers (we hope) long-term memory in the audience, as in the image below.



The **appealing** chart that results might be sufficient if it only need to appear in an internal memo, for instance. But if it is going to be part of a concerted communication strategy with the organizations that bankroll the charity (or some other stakeholders), some additional aspects of storytelling might help paint a more thorough picture, as in the **data story** below.

Meals served over time: **big drop in 2020**



Even though it might seem obvious that there would be a **big drop in meals served in 2020** due to the pandemic, note that the **2017-2019 numbers** were already going against the **2010-2016 trend** – we should not be planning for a return to 2016 levels without first understanding what happened in 2017-2019.

While the specifics of the process may change depending on the data, the charts, the audience, and/or the story objectives, the general path is basically consistent.

Data Storytelling Tropes

The step that transforms simple charts into appealing charts is often accomplished with the help of the **Gestalt principles** or with the help of **data storytelling tropes**, which is to say, data visualization patterns and strategies that have become so familiar as to not need explanation (icons, conventions: see Section 7.3, *Forms and Structures*).

Some examples include:

- the inclusion of a **trend line** with any scatter plot, indicating the direction of the correlation between two variables (positive: going up; negative: going down), or the absence of any such correlation;

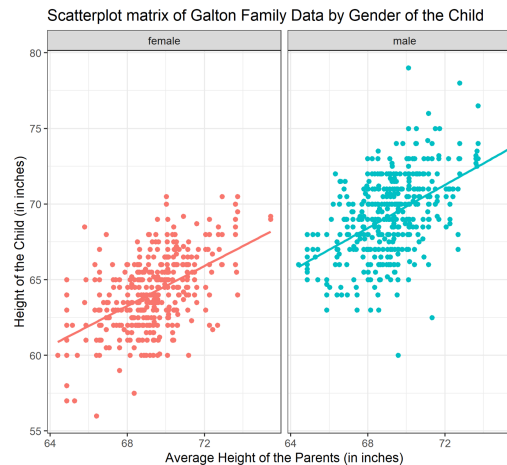


Figure 8.13: Based on Galton's original data [↗](#).

- using a **cluster bar chart** with two categories where one is always lower than the other to showcase a “significant” difference between the two categories;

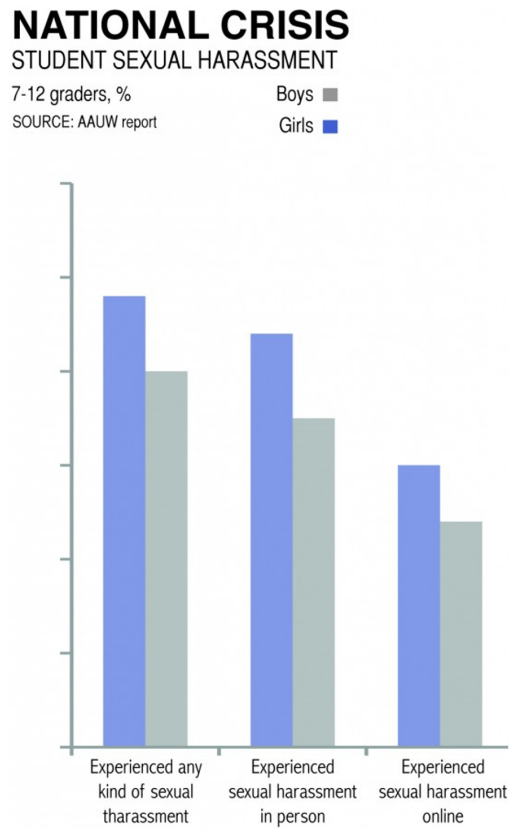


Figure 8.14: Image taken from [114].

- a point of interest located at the **intersection** of two curves;

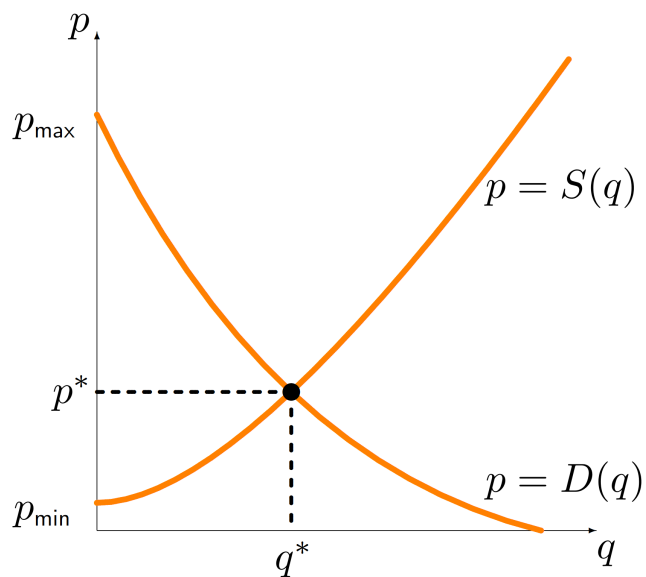
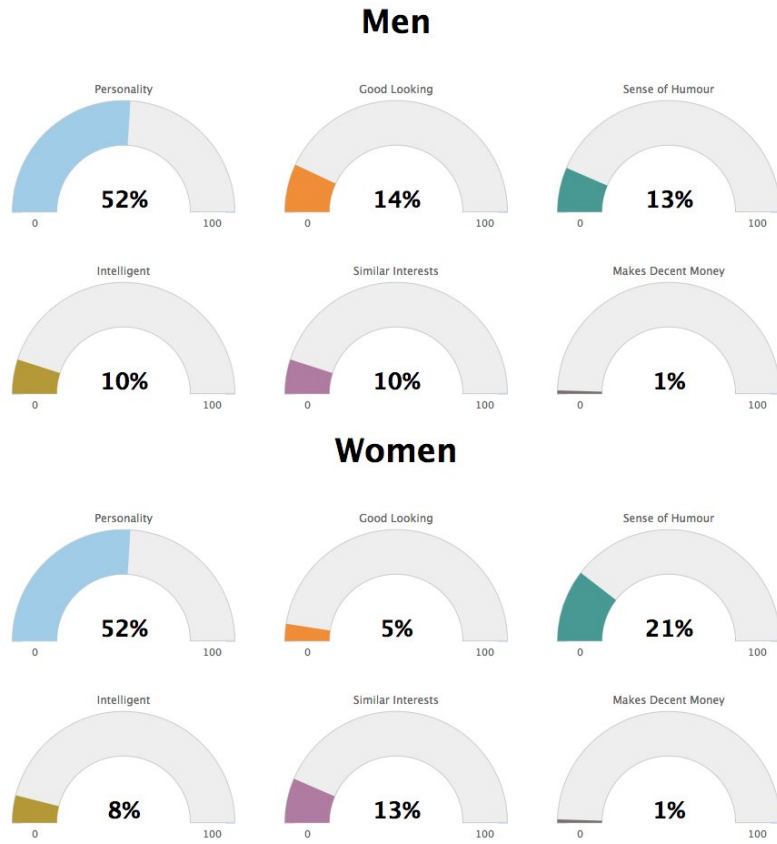


Figure 8.15: Equilibrium point at the intersection of supply and demand curves.

- using gauge charts, pie charts, and doughnut charts in all manners of dashboards;

What characteristics are most important to British men and women?

Men and women both rate personality as the most important characteristic in a romantic partner. However, men rate good looks as the second most important characteristic while women prefer a good sense of humour.



Source: YouGov

Figure 8.16: Circular charts from [115].

- the use of **colour coding** to represent political positions in maps and general charts (in the U.S.: red for right and right-leaning, blue for left and left-leaning; in the rest of the world, these are mostly reversed);

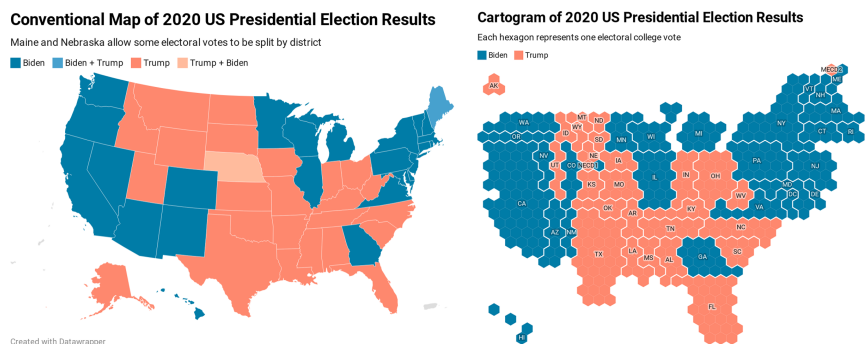


Figure 8.17: United States political cartogram colours [116].

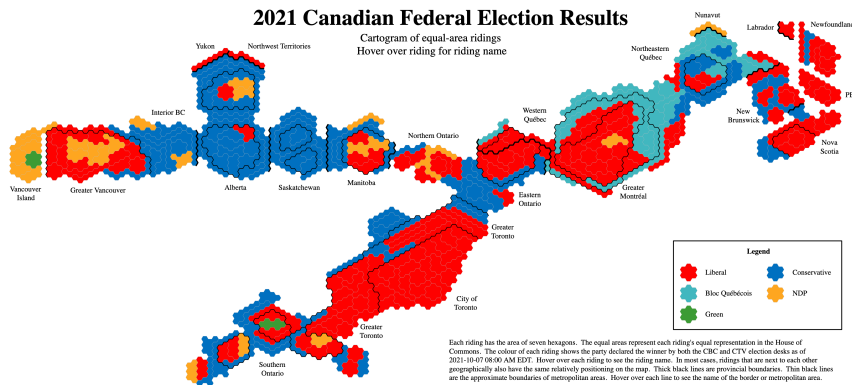


Figure 8.18: Canadian reversal of political cartogram colours [Mark Gargul].

- using **broken** axes to exaggerate the data story effect;

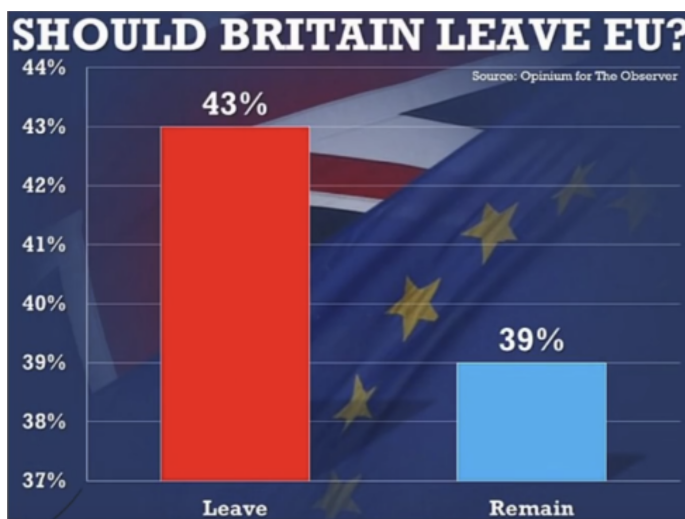


Figure 8.19: Chart created by Wikipedia user 'Vasy1 10'.

Some data storytelling tropes are **discipline-dependent** and may not be recognized as such by practitioners in different fields, so the need to educate the audience may arise¹⁴. We will no doubt continue to identify and construct more of them as data visualization and data stories become commonplace.

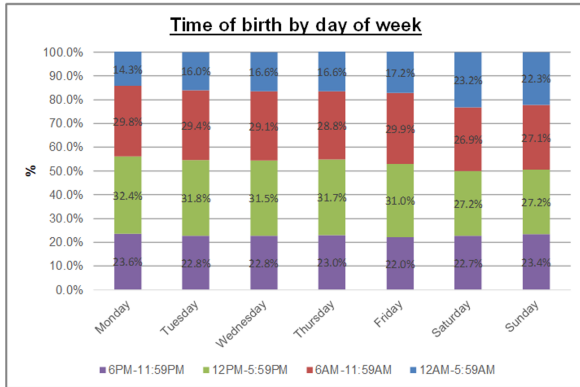
14: This is yet another area where the tsarina of common sense could come in handy.

Examples

In the previous example involving the charity serving meals, the use of a bar chart for the initial (“ugly”) visualization **informed** the look and feel of the final data story. It is conceivable, however, that upon first visualization the data into an ugly chart, an **intermediate transformation step** could be required before we can create an appropriate simple chart.

The following BEFORE/AFTER charts provide additional examples of “ugly” charts **evolving** into compelling data stories. Note that there is no **unique** to do this; for each of them, might there be other ways to present the story?

BEFORE



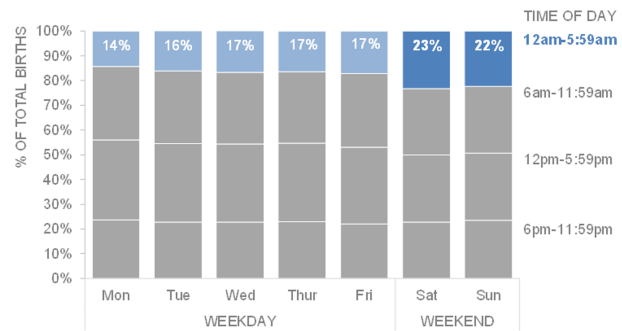
Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2018)

Figure 8.20: Evolving a storytelling chart [117].

AFTER

When babies are born

Weekend deliveries are more likely to be in early morning, compared to weekdays



Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2018)

BEFORE

TICKET TREND

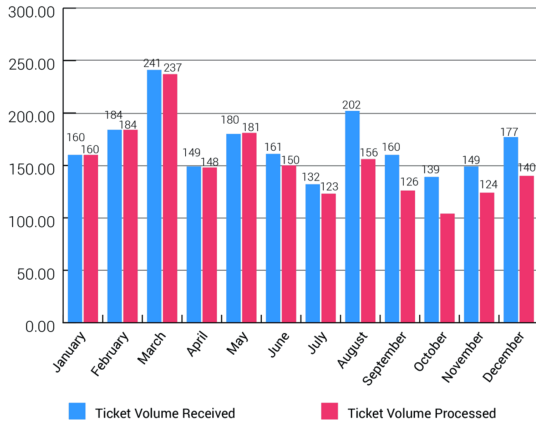
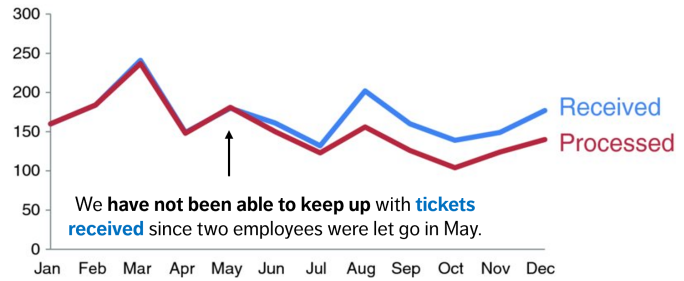


Figure 8.21: Evolving a storytelling chart, modified from [4].

AFTER

Lag in Tickets Processed Since May Layoffs



BEFORE

Survey Results

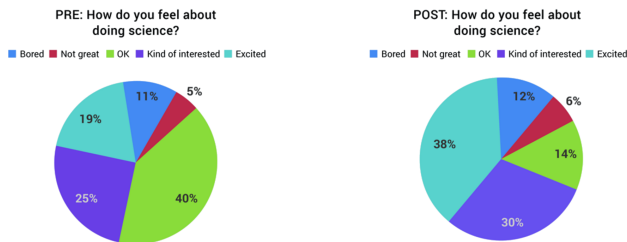
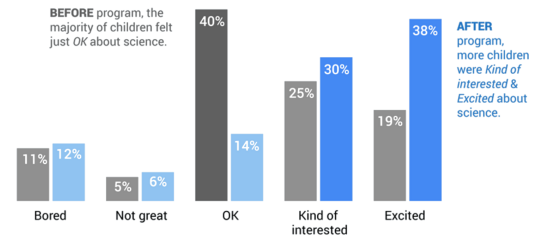


Figure 8.22: Evolving a storytelling chart [4].

AFTER

Pilot program was a success

How do you feel about science?



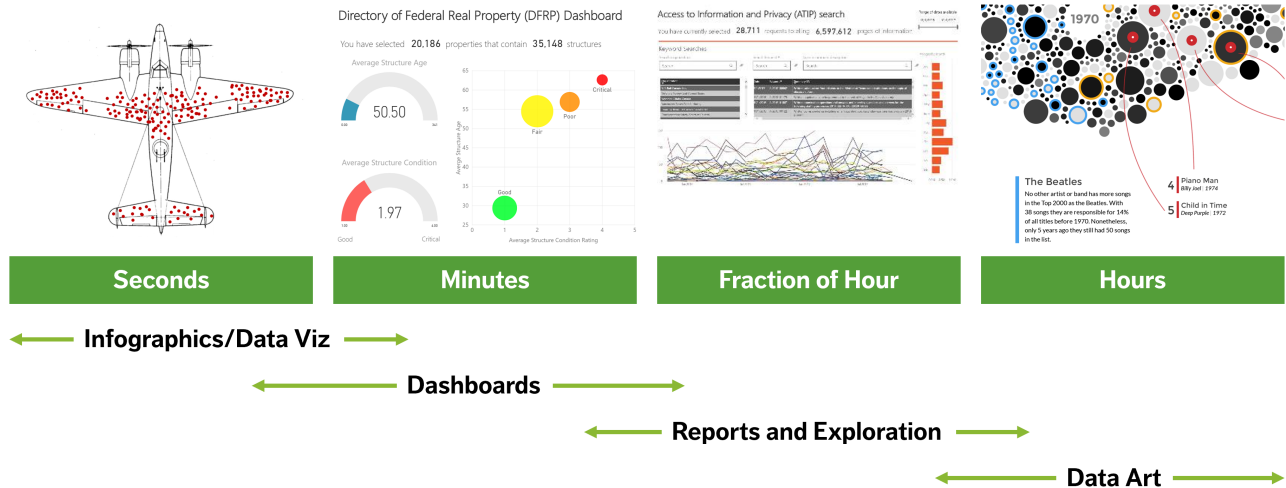


Figure 8.23: Defining the visualization context.

8.4 Anatomy of a Storytelling Dashboard

Non-data stories have their own presentation conventions, which depend of course on the story format and genre, and on the audience expectations.

For data visualization, the exact nature of the presentation depends on the overall context and the amount of time the audience is willing or able to put into “consuming the story” (this is presented graphically in Figure 8.23).

The choice of “product” is **prescriptive**: we expect the audience to spend *at most* a few minutes (5, perhaps?) on a simple, non-evolved chart; conversely, if we know that our audience only has a few minutes to spare, it would be beneficial to produce a single **simple visualization**.

The same principle applies to the other types of presentations:

- **dashboards** (at most 12-15 minutes),
- **reports** (at most 30-45 minutes), and
- **data art** (no time limit).

Data stories are usually presented as single (evolved) **charts**, **infographics**, or **storytelling dashboards**.¹⁵

In the last section, we discussed how to evolve data into storytelling charts and infographics; in this section, we briefly discuss the **composition** of a (storytelling) dashboard, which takes into account various elements:

- the **audience** and its expectations;
- the storytelling **goals** and the available data and analysis results;
- the dashboard’s **narrative** approach, and
- how that narrative is presented (its **logic**).

Note that, practically speaking, the **layout** is informed by visual memory considerations (see Section 5.1, *Visualization and Memory*).

15: Some reports can take on a distinctly lyrical quality, and could be perceived as data stories, were it not for their prohibitive lengths.

Exploration, Situational Awareness, Storybook Dashboards

Dashboards come in many flavours, although the distinctions are sometimes lost on users. In short,

- **exploratory dashboard** focus on understanding the **DATA**;
- **situational awareness dashboards**, on keeping tabs on a **SITUATION** of interest, and
- **storybook dashboards** on communicating a **STORY**.

Exploration Dashboards Exploration dashboards use visualizations as a tool to **explore the data**. They come with a **high level of interactivity** (filters, sliders, drill-down options, etc.) and offer **high levels of detail**: all aspects of data should be represented (tables, columns, calculations, etc.), not just summaries. Since such dashboards are usually prepared for small internal audiences with an in-depth understanding of the context from which the data arises, **annotations** (or explanations) **are neither required nor useful**, for the most part.

Exploration dashboards capture as much interesting information about the data as possible, in as short a time as possible: the focus is on automatic chart generation, not on aesthetic considerations.

Financial Data Exploration

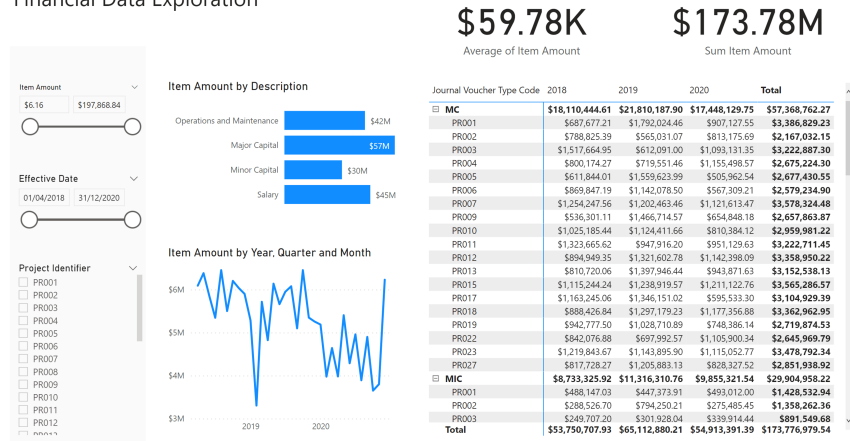


Figure 8.24: A dashboard exploring financial data for a fictitious organization.

Situational Awareness Dashboards Situational awareness dashboards use visualizations as a tool to **provide a “real-time” snapshot of the data**. They are implemented with a **medium level of interactivity**, allowing for the focus to be temporarily directed to different departments or processes. These dashboards are not **“scripted”** as the focus is determining whether certain key performance indicators (KPI) are staying or trending above (or below) some pre-determined **“warning”** or **“emergency”** thresholds.

As the dashboards may be regularly updated with new data, it is crucial that they be **well organized** and **decluttered** so as not to hinder the organization’s reaction speed, when needed. Situational awareness dashboards contain data summaries tables and charts and may feature anomaly detection.

Financial Snapshot

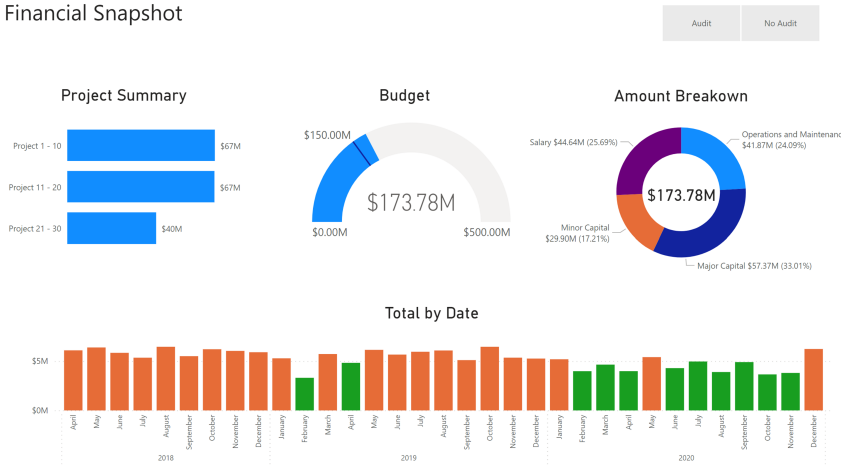
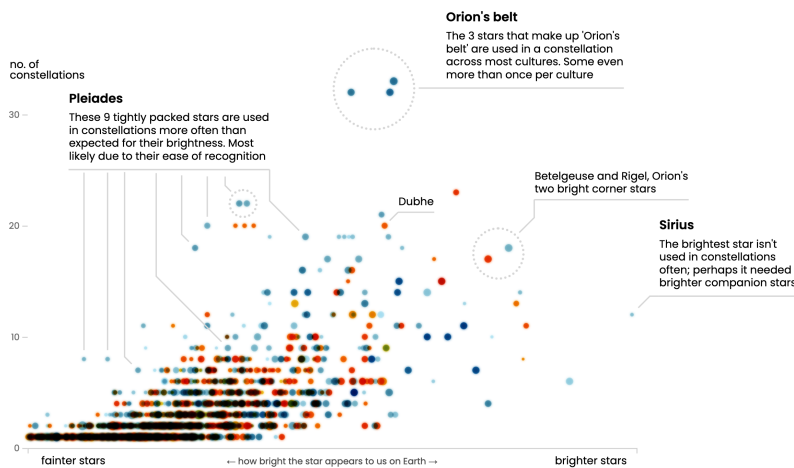


Figure 8.25: A dashboard providing situational awareness of financial data for a fictitious organization.

Storybook Dashboards Storybook dashboards use visualizations as a tool to explain the data. They are accompanied by a **low level of interactivity**, in part to ensure the audience does not trip over itself by modifying the charts and losing the story in the process (if some interactivity is provided, a reset option should also be present). Storybooks are characterized by **low levels of detail**: typically, numbers are rounded or altogether removed from the presentation, and only the **key aspects** of the data and analytical results are represented.

As the target audience may have little data and domain expertise, **annotations** used to drive the story and to ensure that there is no ambiguity about what is being communicated. Storybooks present the results of quantitative data analysis as qualitative insights.



NOTE | Star colors, based on their actual visible colors, have been exaggerated/saturated for better visibility. The sizes of the stars are (not-linearly) scaled to the actual brightness of the stars (called absolute magnitude); how bright they would be if they were all placed the same distance from Earth. Because stars are all at widely different distances from the Earth, how bright a star appears to us is called the apparent magnitude.

Sirius is about 25 times more luminous than the Sun, but Canopus is ~10,000 times more luminous!

Take Sirius, the brightest star, it appears almost twice as bright to us as Canopus, the next brightest star. However, compared to the other stars we see at night it's actually not exceptionally bright. The reason it appears so bright to us, is because it's one of our closest neighboring stars (at 8.6 light-years).

Figure 8.26: Storybook for stars and cultural constellations [70].

Audiences (Reprise)

We have already discussed the importance of the **audience** to storytelling, and that it is preferable to avoid preparing a data story that will be all things to all people all the time.

Instead, we suggest addressing **lines of business**, such as Finance, Engineering, Marketing, HR, your immediate supervisors and team members, the minister's office, etc. Doing so will allow the storyteller to **identify the decision-makers** and the role played by various audience members, which will, in turn, inform the storytelling process: if a minister is reputed to put a lot of faith in raw numbers when making decisions, it could be useful for a few such figures to appear on the dashboard.

This is not to say that analysts and storytellers must always cave in to their audience's wishes. While we say that we want to make data-driven decisions and that we are data-friendly, the reality is that this is typically only the case as long as the data keeps supporting our beliefs and positions. It is good practice, then, to throw the audience a bone and help soften any blows which could lead to audience push-back and reject conclusions that are not to their liking. This is best achieved, however, when the audience is **known** by the data storytellers.

But **knowing who** the audience is is not the same as **knowing** the audience, and what what relationship analysts and dashboard designers have with them. In particular, it is crucial to figure out how they **perceive data** and **data analysis**, and how **trust** and **credibility** can be established with them (heeding the lessons from the boy who cried wolf, presumably).

Another thing that might change from audience to audience is what they need the data storytelling for, what they need it to do for them. To get a sense for this, we need to know how the results will be used and what **actions** the decision-makers are likely to take:

- what decisions are people going to make from the analysis?
- how often are they going to be looking at the data?
- how often do they expect the data to be refreshed?

We also need to know **what the audience needs to know**:¹⁶

- about data availability,
- data cleanliness,
- data governance and accessibility, and
- whether the data it is being "massaged" or used to paint a rosy picture?

Finally, the audience's need/wishes to **interact** with the charts will inform the type of dashboards that is produced: will they be passive receptacles for the dashboard? Are they expecting to be able to slice, dice, and filter the data and the charts?

All of this is part of the legwork (soft-skills) required before the story can be told, no matter what format that will take: finding the answers requires **questions to be asked** and **answers to be gotten**, which is definitely much easier to do when the audience is known.

16: What the audience **needs to know** is not the same thing as what it **wants**: Stephen *wants* an Astin-Martin roadster, but he *needs* a Dodge Caravan.

Identifying and Gathering Presentation Requirements

The **requirements** for a dashboard, report, or presentation are driven by the **primary consumers**, the stakeholders that will primarily be getting “value” from using the product.

It is a (sadly) very common mistake to **cast the net too wide** and to build something for too many consumer types at once, if only because the answers to the previous section’s questions might then be all over the place and attempts to exhibit a uniformly coherent visualization deliverable will be stymied by the competing audience priorities.

Once the group of primary consumers has been identified, we suggest following a **formal process** to gather the visualization requirements as accurately as possible; these can be obtained by surveying the stakeholders and obtaining answers to questions such as (but not necessarily limited to):

- what is the proposed name of the product?
- who are the target data consumer(s)?
- what are the product high-level objectives?
- when does it need to be published/made available?
- with what frequency is the data updated?
- what kind of business decisions will be made by the target consumer group?
- what are the sources of data?
- is the data/information duplicated anywhere else (e.g., by a 3rd party)?
- what is the sensitivity level of the source data?
- what is the sensitivity level of the final product?
- how is the source data gathered?
- what quality assurance is performed on it?
- etc.

This might seem like unnecessary work to add onto already busy analysts and dashboard designers, who mostly want to focus on the presentation and its **contents**, but removing ambiguities and ensuring that the visualization product is aligned with stakeholders’ needs will reduce the risk that the deliverables fail to align with stakeholders needs.¹⁷

Storyboarding

Once we have a set of well defined requirements we are in a position to perform a **storyboarding** exercise. Storyboarding is a way to summarize the flow of information into a **coherent whole**, before we start working with data and software (see *Narrative Structures*, p. 137).

This helps the design team determine **how many pages/elements per page** might be needed to create an impactful data story. Note that this is **NOT the same** as designing the **dashboard layout**: storyboarding is used to **define the story** and the dashboard’s **content**.

17: Thankfully, identifying and gathering presentation requirements can be conducted simultaneously with the get-to-know-your-audience fact finding from the previous section.

18: This is another occasion where the external eye of a tsarnina of common sense can provide support.

In the age of multiple available software options for every aspect of the data analysis pipeline ranging from data collection to publication, it may seem counter-intuitive (if not downright retrograde) to eschew digital tools; however, storyboarding is a **pen-and-paper** exercise that may require a heavy time commitment in order to be conducted properly.¹⁸

Example: a government department has an aggressive hiring goal for the year, imposed by new legislative requirements. Funding for the hiring process has been approved, however the new budget will severely impact hiring managers and other corporate service groups (IM/IT, Real Property, etc.). Could a data story help with the situation?

What might the narrative structure of such a story look like?

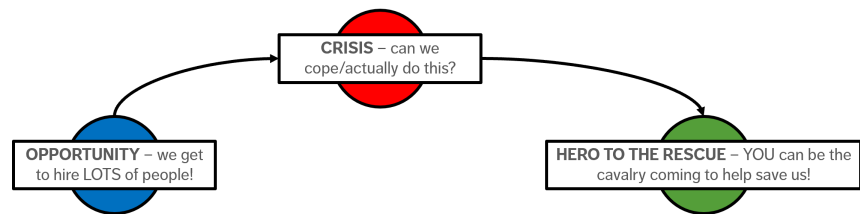


Figure 8.27: Simple narrative structure to inform an eventual storyboard.

The simple path of Figure 8.27 is unlikely to be the foundation for whatever reports and/or dashboards the above situation calls for.

But it may inform the associated storyboard by **breaking the narrative down** into easier-to-parse steps, each of which gets placed in a **box**, not unlike what one would find in a cartoon strip.¹⁹

19: If we were making a movie, we would also use images – this could actually prove to be quite useful for data stories.

Such a storyboard is shown in Figure 8.28.

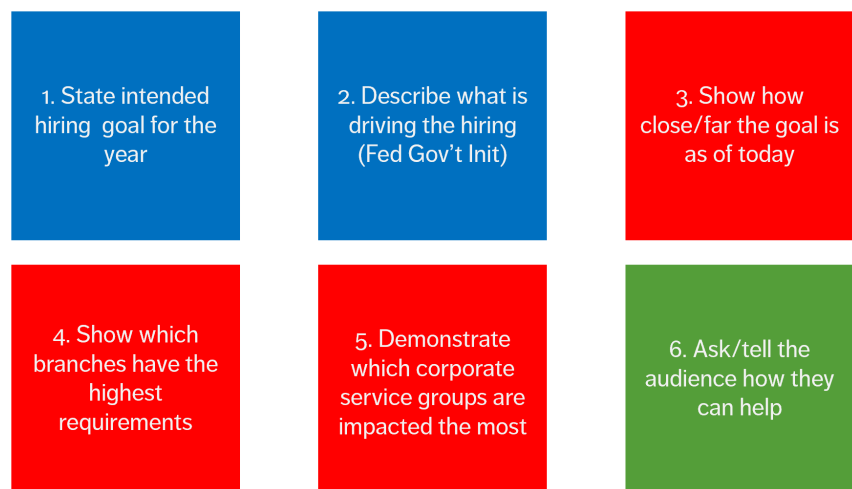


Figure 8.28: A simple storyboard for the hiring situation, colour-coded as in Figure 8.27.

At no point of this process does data enter into the equation; no charts are produced, no software is used.

Dashboard Narratives

We have discussed narrative structures as they apply to general stories in Section 7.3 (*Forms and Structures*); what does that mean in dashboard terms?

There are a number of ways of **constructing a narrative**, including:

- all events being presented in chronological order;
- the most/least important event is presented first/last;
- the first/last occurring event is presented first/last;
- the most/least successful event is prepared first/last;
- the worst/best news are presented first/last, etc.

When its applicable, we try to tell data stories in a number of different ways so as to maximize buy-in and minimize push-back.

A dashboard's **logic** determines its format:

- a **horizontal logic dashboard** is akin to a *saga* or an *epic*,²⁰ in which the chapter/episode/book titles, when laid out in succession, provide a **summary of the story** – in this analogy, the chapters/episodes/books correspond to the dashboard's pages.²¹
- a **vertical logic dashboard** is closer to an *anthology* on a theme,²² in which individual episodes/stories/essays are self-contained and **logically linked to one another** – in this analogy, the theme represents the data, and the various pages, different aspects of the data stories.

A hallmark of horizontal logic dashboard is that the **order of its pages is crucial** and must be respected when the story is presented: a **missed page** (or a page read out of order) changes the story entirely and makes it altogether unintelligible; the order of the vertical dashboards, on the other hand, is **not important**: pages may be omitted without “crashing” the overall dashboard story since each page stands as a coherent story in its own right.²³

Putting it All Together

All of these come together to suggest a storytelling dashboard's **guidelines**:

- an **executive summary page**, which provides the story outline and contents and addresses issues related to long-term memory retention by combining text and visuals;
- at most 7 or 8 (and quite often much fewer than that) **regular pages**, arranged according to context and dashboard logic considerations;
- each of which containing at most 4-5 **decluttered, evolved, and annotated charts**, employing pre-attentive attributes to attract the eye (iconic memory), a small number of charts (due to short-term memory limits), and text to drill the visuals into long-term memory, and
- as a final reminder, **less is more**.²⁴

We present two examples in Figures 8.29 and 8.30 that combine all of these concepts together.²⁵

20: [Lord of the Rings](#), [Babylon 5](#), or the [Anne of Green Gables](#) series.

21: Horizontal logic can be reinforced with an **executive summary** or report placed at the presentation's beginning.

22: [Black Mirror](#), [Dropped Threads](#), [Billions and Billions](#).

23: In practice, dashboards may **combine** horizontal and vertical logic: for instance, there could be three vertical segments, each of which consisting of a horizontal sequence of two pages each, say.

24: A good general principle for storybook dashboards is that we should cut from them as much as is necessary to convey the story, and once that is done, to go re-visit and cut again as there is likely still too many elements on the dashboard.

25: They are entirely too busy to constitute anything other than a first pass at a storybook; we hope that you will see them for the pedagogical spirit they provide.

WEEKLY number of boats sold (20X6) – Store #16

2869408609876
 9348586748676
 2967303986739
 3967496749674

Yearly goal: **290**
 20X6 total: **307**
 Do these numbers look **reasonable?**

2869408609876
 9348586748676
 2967303986739
 3967496749674

Most frequent weekly number of boats sold: **6 (11 times)**
 Occurred: **randomly** (as expected)

2869408609876
 9348586748676
 2967303986739
 3967496749674

Another frequent weekly number of boats sold: **8**
 Occurred: **5 times immediately before a 6 (out of 7)** (surprising)

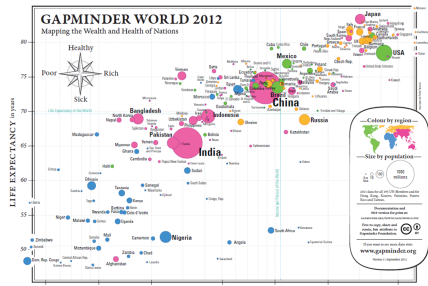
2869408609876
 9348586748676
 2967303986739
 3967496749674

Another frequent weekly number of boats sold: **7**
 Occurred: **7 times immediately before a 6 (out of 8)** (surprising)

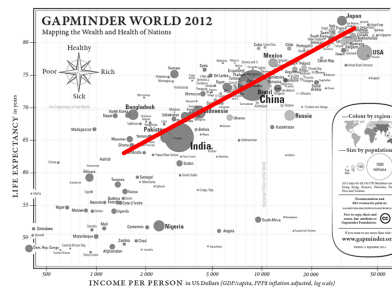
VERDICT: The two last charts suggest that the weekly sale numbers **are not random**, and that they may have been falsified. We recommend **performing an audit** of sales for store #16.

Figure 8.29: Storytelling dashboard based on the sales numbers example from [4].

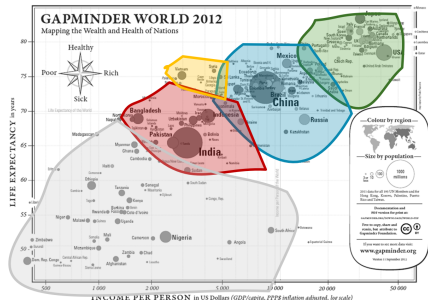
The Health and Wealth of Nations (2012) – Gapminder



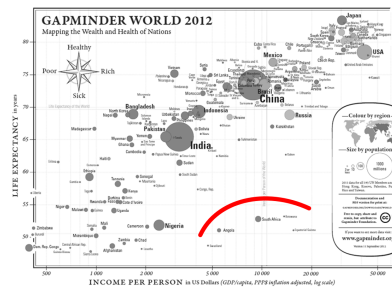
2012 **life expectancy** of nations (health, vert.) against the **log of the GDP per capita** (wealth, hor.); colour represents **world regions**, size is **population**.



The relationship between **health** and **wealth** is **roughly linear**, at least when they are both high enough – is it a **causal relationship?**



We see roughly **5 nations groups (clusters)**, when we only focus on health and wealth. Would these groups be stable when using **other variables?**



There are **outliers** in the bottom right quadrant (**wealthy, but not healthy**); 4 of them are in southern Africa – a manifestation of **apartheid?**

NOTES: The relation between health, wealth, and region can clearly be seen in the charts, but the big surprise might be that life expectancy is as high as it is across the board. Can we get more insight from other variables?

Figure 8.30: Storytelling dashboard based on the 2012 Gapminder Health and Wealth of Nations chart [106].