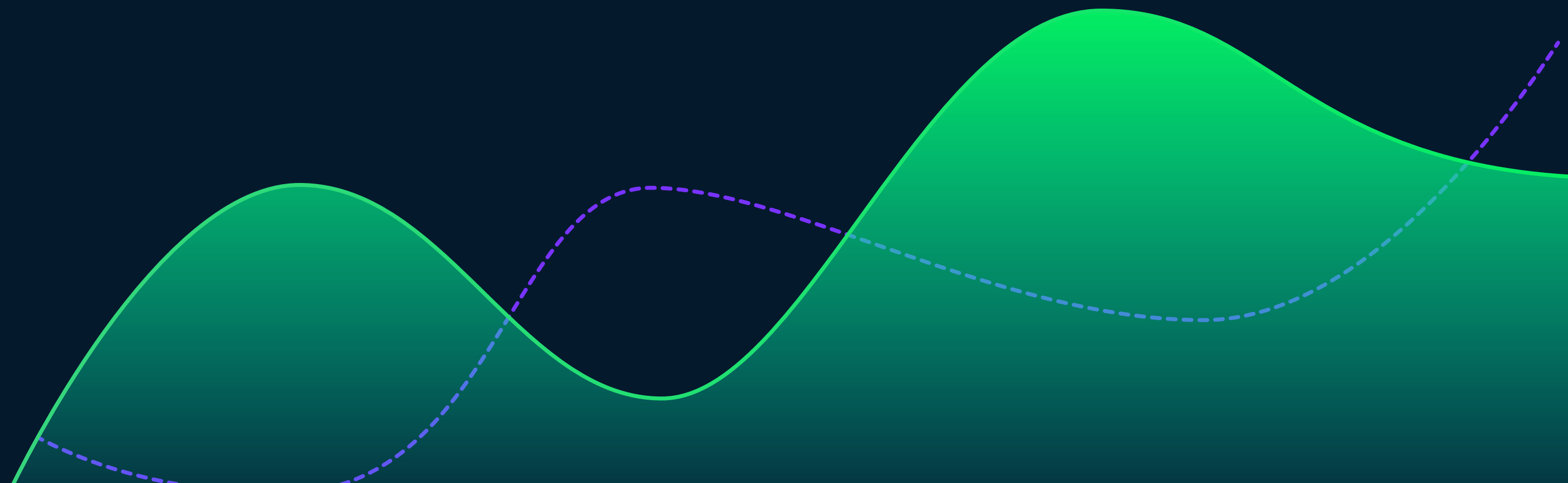
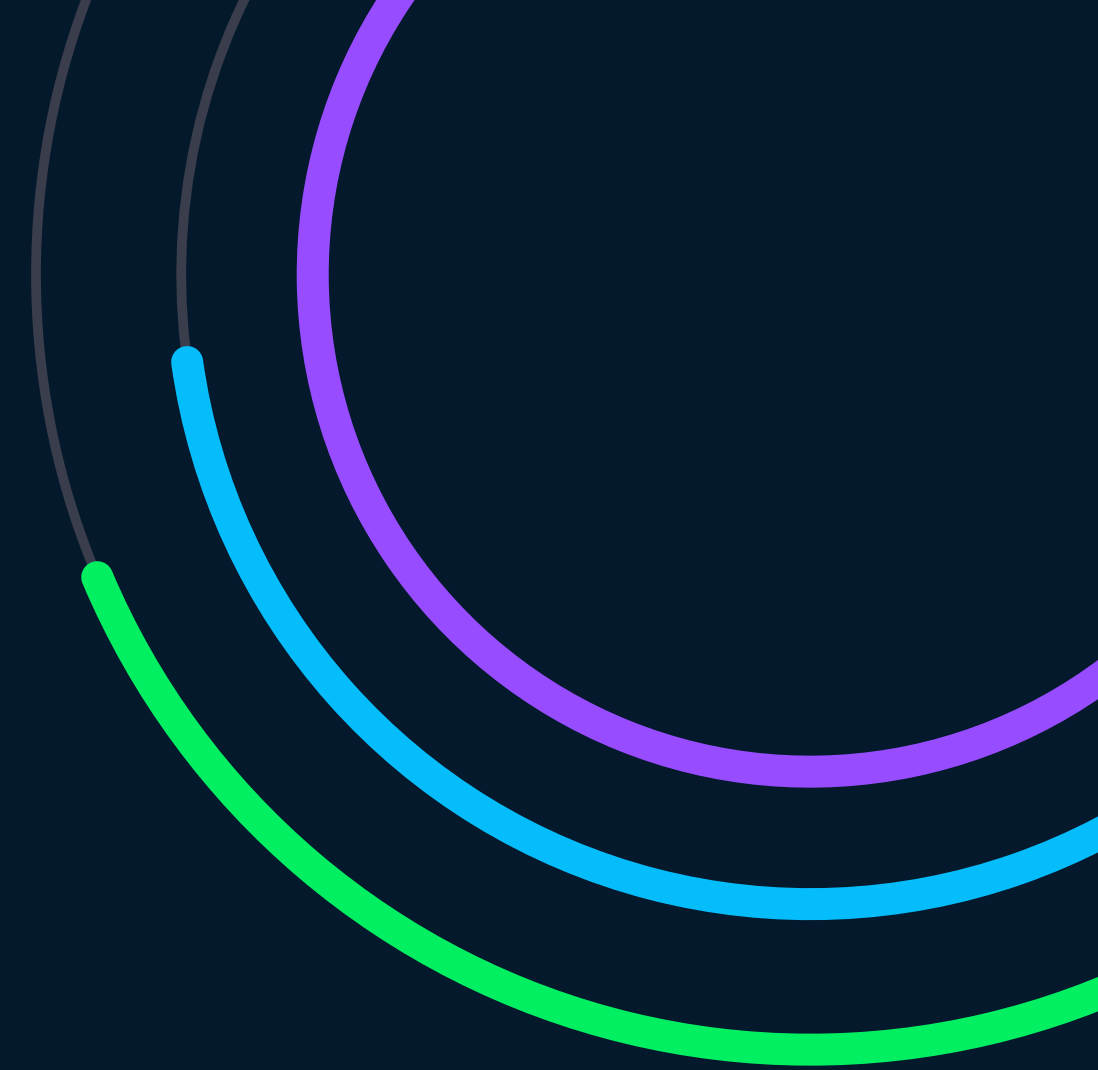




# 8 rules for better data storytelling



It was the end of 2020 when social media platforms faced an influx of Spotify Wrapped content—a summary of the user’s most-streamed music in the past year. Those given the title of “top 1% fans worldwide” of their favorite artists proudly paraded their title like a badge of honor, while others shared their Most Played Genre as a declaration of their musical preferences. In 2020 alone, [90 million users engaged with Spotify Wrapped](#). Clearly, users [enjoyed the data stories their streaming statistics told](#) about them.

Spotify Wrapped is only one example that illustrates the increasing use of data visualizations in our daily lives. As the world watched the Covid-19 pandemic unfold, data visualizations like [John Hopkin’s Dashboard](#) played a crucial role in informing the public on the extent of the pandemic. Data stories like National Geographic’s coverage on [500,000 Covid-19 deaths](#) elicited awareness of the deadliness of the virus. All these examples were in line with DataCamp’s prediction of data visualization (and data stories) becoming mainstream in our [2021 Trends Report](#)

The difference between data visualizations and data stories is nuanced but important. Put simply, data visualizations are good at articulating what has happened, but not necessarily why something has happened. This is where data storytelling comes into the picture.

[Data storytelling](#) is a methodology for communicating information with a compelling narrative. It is made up of three components—the data, visuals, and narratives. Together, these components help data storytellers engage the audience, make data more memorable, and be more persuasive.

## Why we need data stories

### Data stories make the data more memorable

You might find it much easier to recall a funny story than the tenth digit of pi.

Stories serve as a framework to connect disparate pieces of information coherently and elicit an emotional response.

Humankind has used storytelling to pass down culture and traditions. In a paper [Memory, Imagination and Learning: Connected by the Story](#), it was discovered that storytelling is the most effective technique for oral cultures to faithfully memorize their knowledge. In a sense, humans are hardcoded to recall stories better than cold hard facts.



Data stories help us communicate better stories facilitate communication between the storyteller and the listener. In fact, neuroscientist Uri Hasson observed that the brain activities of the listeners of a story are similar to the brain activities of storytellers. This phenomenon, called the “speaker-listener neural coupling.” indicates the achievement of [successful communication](#). This demonstrates the potential of stories in eliciting mutual understanding.

### Data stories drive actions

When data are presented intuitively in the form of stories, they can be [understood and acted upon quickly](#). Data stories also serve as a medium to communicate data insights and [inspire collaborative action](#) in a way that a regular visualization or dashboard on its own could not.

## Who should learn data storytelling?

Data storytelling is a valuable skill set for both technical and non-technical professionals. As companies build up self-serve capabilities in analytics, more non-technical professionals will get access to a treasure trove of data. However, the potential of that untapped data remains locked unless employees are capable of telling compelling data stories to communicate insights and drive action.

On the other hand, data practitioners who craft audience-specific data stories can [effectively convey the impact of their data science projects](#) in the language of the audience. This technique can be applied in the scoping, implementation, and evaluation stages of the projects to alleviate the skepticism of non-technical users. This paves the way to getting the stamp of approval from decision makers and end users.

With the growing importance of data storytelling in the upcoming decade, all employees must be equipped with the fundamentals of data storytelling. In this white paper, we outline eight rules for better data storytelling that will help anyone craft impactful data-driven narratives.

# Table of contents

<b>Know your audience</b> .....	4
<b>Begin with the goal in mind</b> .....	6
<b>Choose the best visualization for your story</b> .....	7
<b>Do not lie with data stories</b> .....	8
<b>Keep your visualizations minimal and avoid clutter</b> .....	10
<b>Add color to your stories</b> .....	11
<b>Use texts appropriately</b> .....	12
<b>Develop a narrative around your data</b> .....	13

# 1. Know your audience

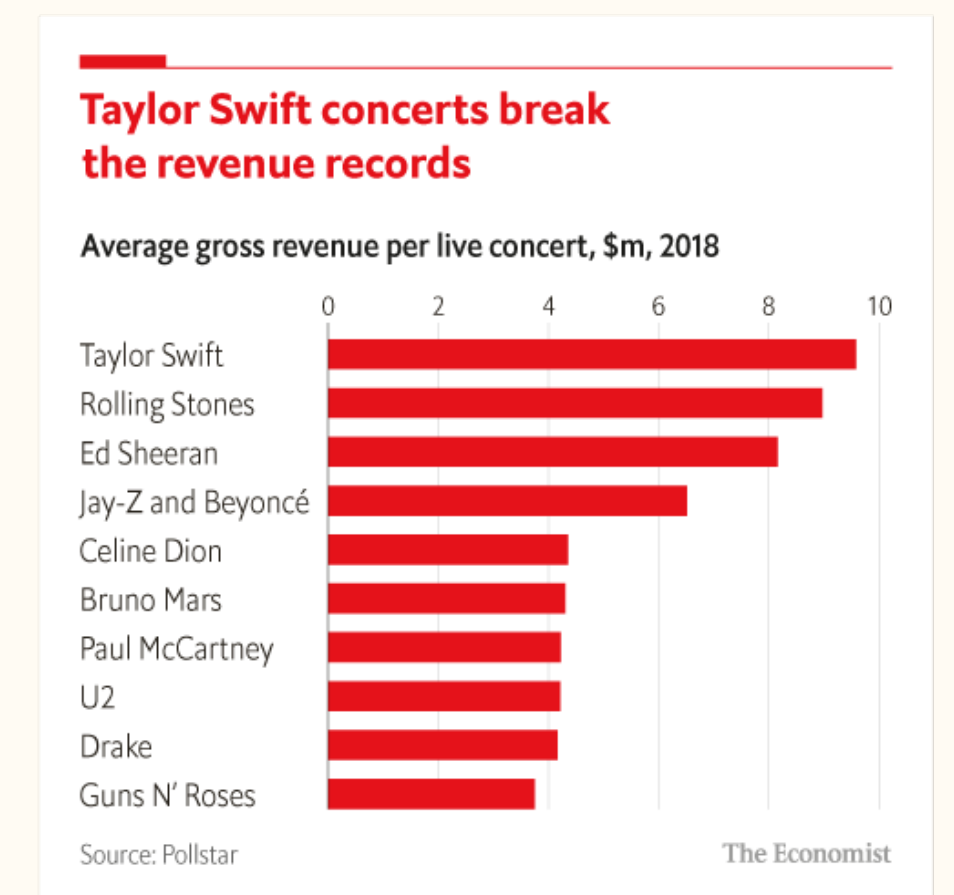
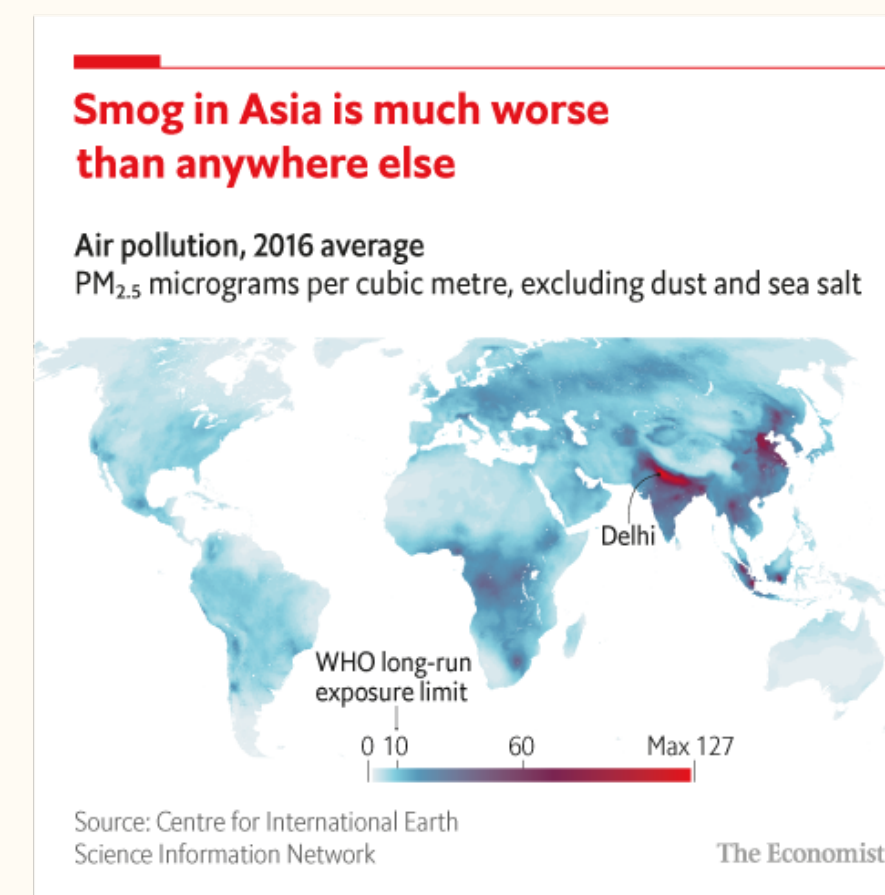
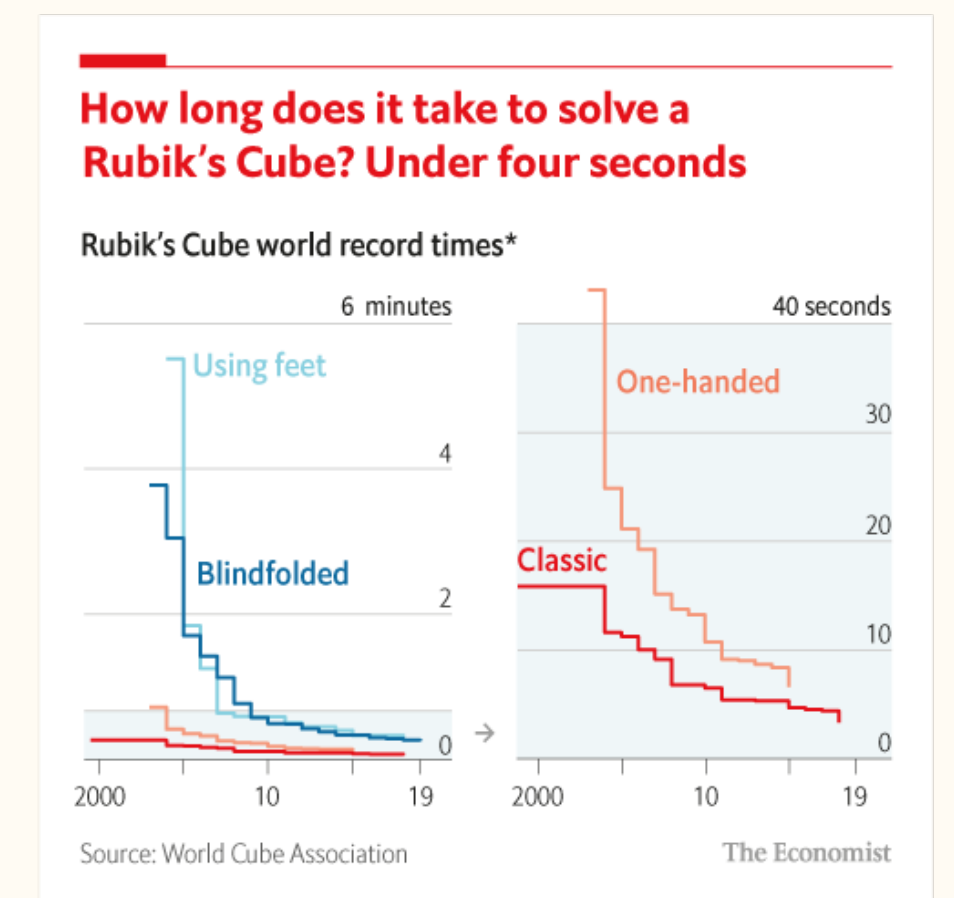
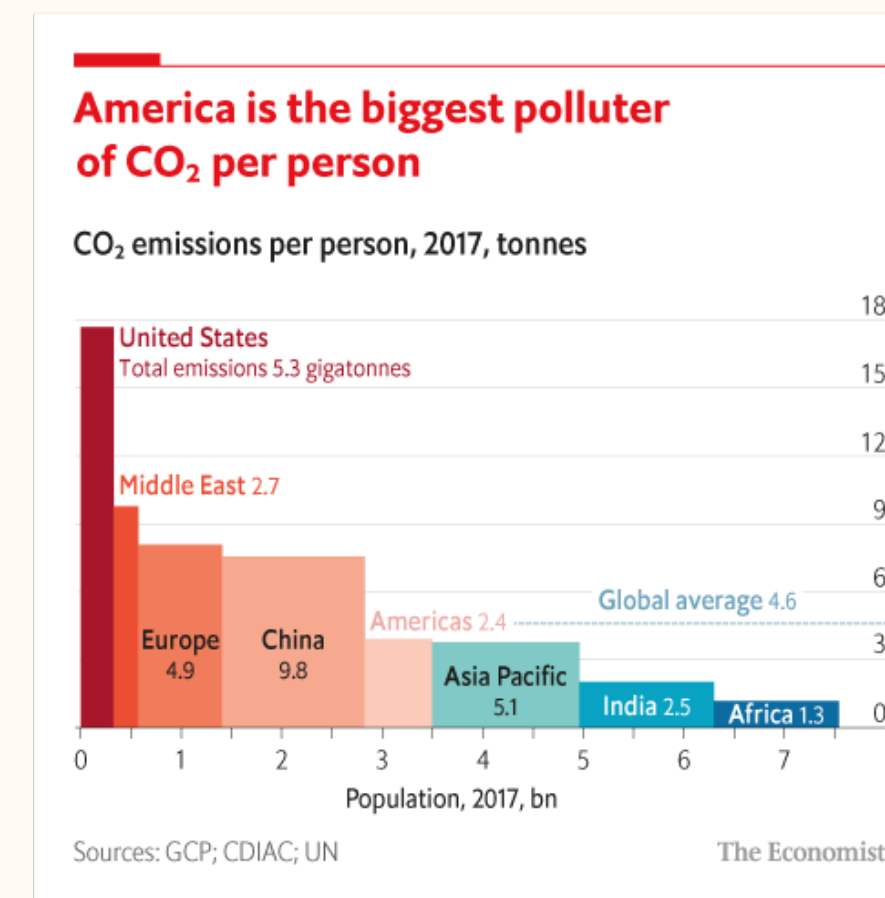
The Economist, a print international newspaper with an emphasis on data journalism and analysis, recently ventured into Instagram. In charting this new territory, it constantly redesigns charts for its audience instead of simply reusing graphs meant for the magazine. The Economist's feat in garnering millions of Instagram followers within a few years attests to its ability to cater to its audience. Here, we distill a few learning points we gathered from the Economist.

## Know what is the priority of the audience

Data storytellers should know what matters to the audience, and present data stories that are of importance to the audience. Otherwise, one runs the risk of losing interest or wasting the time of the audience.

Particularly, a non-technical audience cares much more about the business impact of a data science project rather than a 1% incremental improvement in accuracy. As such, data practitioners should be mindful that the improvement of business metrics has a higher priority than that of technical metrics for business stakeholders. They can also benefit from data storytelling techniques in conveying how AI projects can help with business metrics—whether it is an improvement in customer experience, an increase in retention rates, or a decrease in cost.

An example of this in action is from [The Economist and how it leverages Instagram](#) as a medium to an audience that it could not with its print version. Knowing that its Instagram audience is younger than those of its regular readers, it highlighted stories of interest to its younger audience, running the gamut from climate change to celebrity news.



**Fig 1:** The Economist published data stories that are relevant to its Instagram audience

## Empathize with your audience

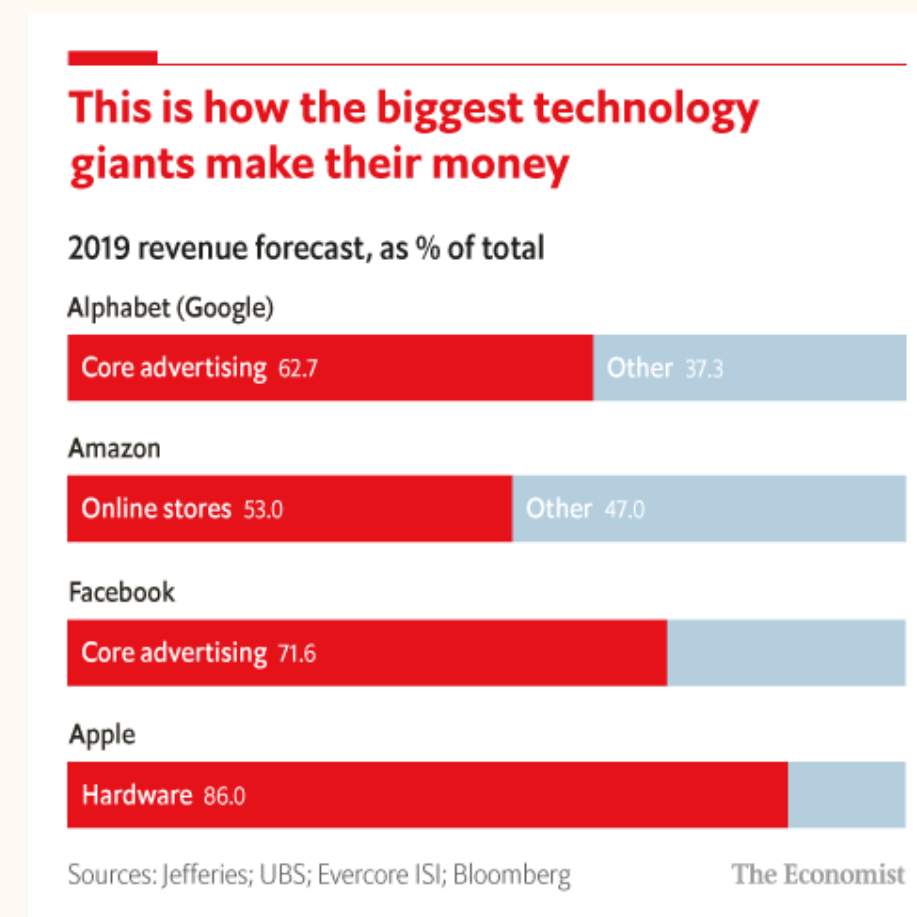
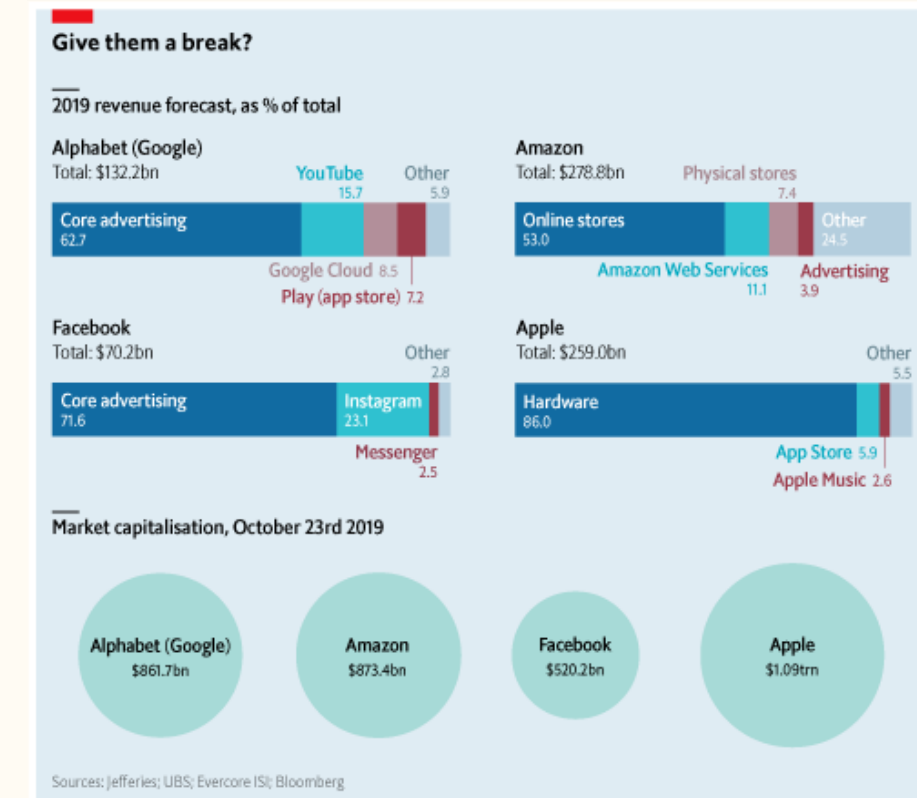
Data stories crafted with the audience in mind have a higher chance of retaining the audience's attention. Empathizing with your audience might influence the content of the data story, the tone of the presentation, the length of the document, and even the rate of speech. Here are some questions that might help you understand and empathize with the audience better.

- Does the audience have the necessary prerequisite knowledge to understand a particular metric?
- How much time does the audience have to consume this data story?
- What is the medium of presentation (written/oral) that the audience prefers?

## Know their level of data literacy

The Economist demonstrated empathy when designing their data visualizations on Instagram. On one occasion, it decided not to publish a chart (initially meant for print) because it required niche prerequisite knowledge that an online audience is unlikely to have. In another instance, it distilled only the main takeaway of the chart and did away with the superfluous because a reader on Instagram is likely to expect immediate gratuity than one on print media.

The Economist is not afraid to present complex charts that require high data literacy to understand since it is targeted at highly educated business professionals. Data storytellers should be aware of the data literacy of their audience when designing charts to avoid underwhelming or overwhelming them. Whenever possible, data practitioners should avoid technical jargons that might confuse non-technical stakeholders, and instead use business metrics to convey the impact of a data project.



**Fig 2:** How the Economist simplified its chart for Instagram



## 2. Begin with the goal in mind

“Success in data visualization does not start with data visualization,” proclaimed Cole Knaflic, the author of the influential book [Storytelling with Data](#). In the book, Knaflic emphasized the importance of understanding the end goal and the context of a data story when starting a data visualization.

Before creating your data story, you might find it useful to ask yourself the following questions:

- **Who are you communicating with?** The more specific is the target audience, the easier it is for the analyst to build a targeted data story that resonates with and meets the needs of the audience. As mentioned previously, understanding the target audience also helps the analyst decide on the level of complexity of the data story.
- **What do you want the audience to know or do?** The main goal of building a data story is to communicate a point and/or to give recommendations. Knowing what exactly the point or recommendation is can help the analyst stay focused on the main point instead of beating around the bush. A good data story drives insights, which in turn drives decisions and calls for actions.
- **How can you use your data to convey your point?** With an understanding of the audience and the intended action, we can start gathering the necessary data evidence that supports the story arc.

The answers to these questions are ingredients for a coherent data story. Once that is done, we can start building data visualizations that support the story.



**“When it comes to the form and function of our data visualizations, we first want to think about what it is we want our audience to be able to do with the data (function) and create a visualization (form) that will allow for this with ease.”**

— Cole Nussbaumer Knaflic, Author of [Storytelling with Data: A Data Visualization Guide for Business Professionals](#)

# 3. Choose the best visualization for your story

The process of choosing the correct type of visualization for a particular insight requires an understanding of the different chart types. Here are the four most common types of charts.

- **Bar charts** are best suited to show comparisons of different categories.
- **Line plots** are useful to show the changes of a variable over time.
- **Histograms** show the distribution of one variable. This tells us how frequently a particular value occurs relative to others.
- **Scatterplots** can be used to show the relationship between two variables.

In general, data visualizations can be used to show comparison, relationship, composition, and distribution. [The following mindmap](#) is a guide for determining a suitable chart based on its purpose.

	Description	Examples
 <b>Bar charts</b>	Best suited to show comparisons of different categories.	<ul style="list-style-type: none"> <li>• A bar chart showing the reasons for users to unsubscribe from the company's newsletter. (Fig 3A)</li> <li>• A bar chart showing the breakdown of company cost into different sources (marketing, operations, IT, etc)</li> </ul>
 <b>Line plots</b>	Useful to show the changes of a variable over time.	<ul style="list-style-type: none"> <li>• A line plot of the company revenue over time. (Fig 3B)</li> <li>• A line plot of the app rating or user satisfaction score over time.</li> </ul>
 <b>Histograms</b>	Show the distribution of one variable. This tells us how frequently a particular value occurs relative to others.	<ul style="list-style-type: none"> <li>• A histogram showing the distribution of user spending in the past quarter. (Fig 3C)</li> <li>• A histogram of the distribution of employee satisfaction (score of 1 to 5) from an HR survey.</li> </ul>
 <b>Scatterplots</b>	Used to show the relationship between two variables.	<ul style="list-style-type: none"> <li>• A scatterplot that shows the relationship between the time a user spends on the app and the customer lifetime value. (Fig 3D)</li> <li>• A scatterplot that shows the relationship between the length of a video advertisement and the clickthrough rate.</li> </ul>

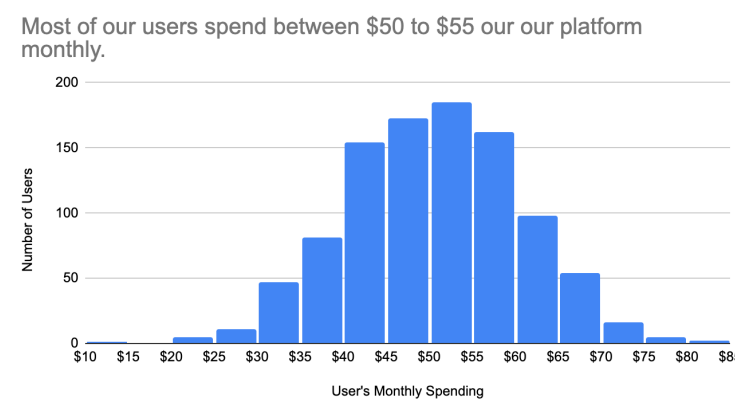


Fig 3A: A bar chart

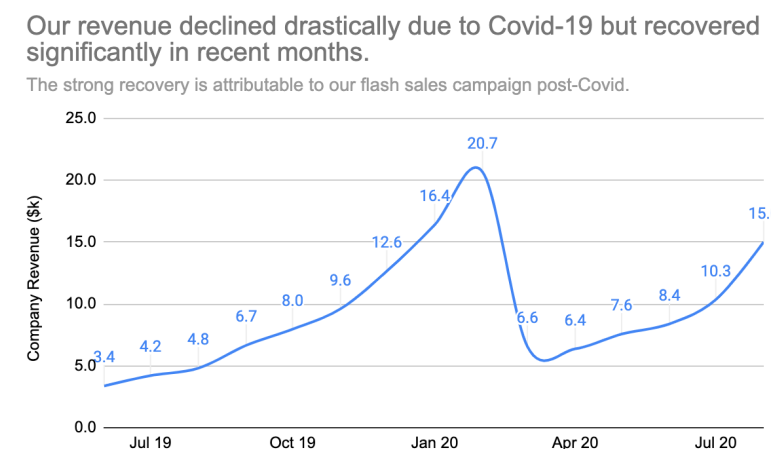


Fig 3B: A line plot

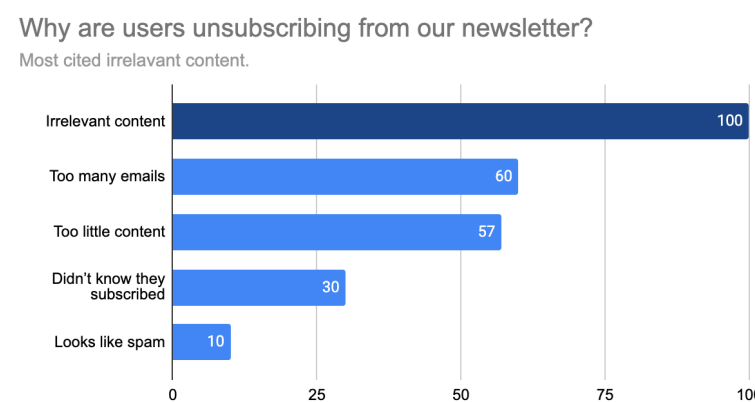


Fig 3C: A histogram

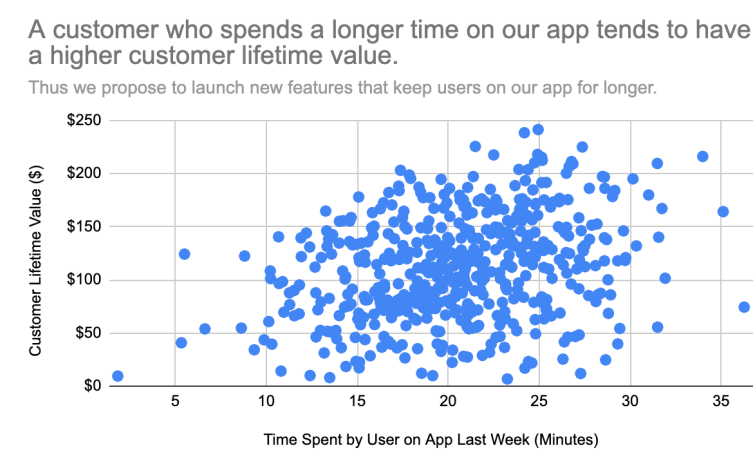


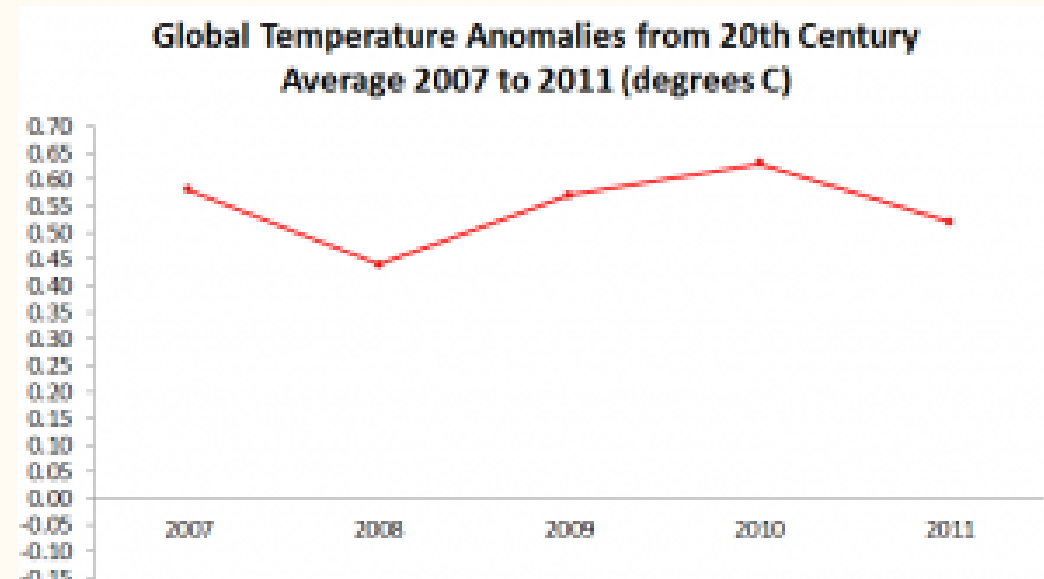
Fig 3D: A scatterplot



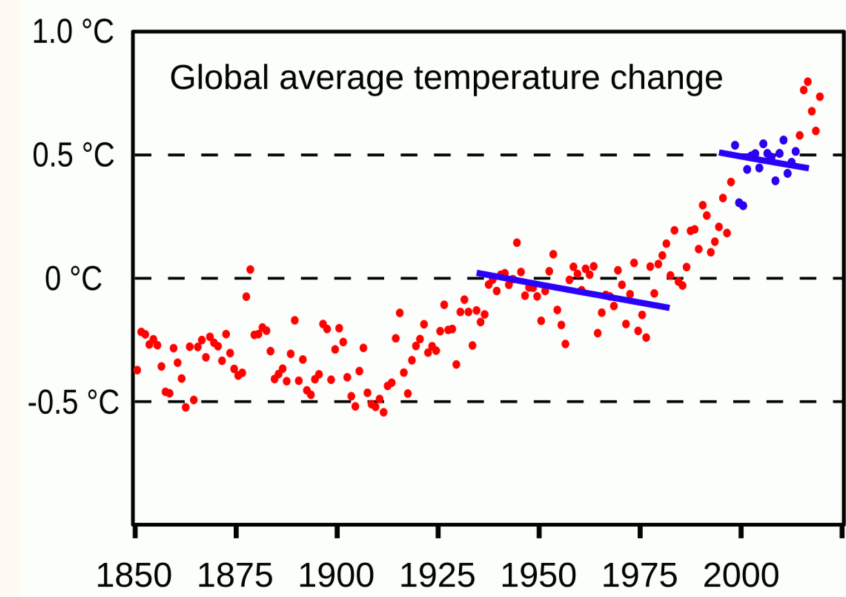
## 4. Do not lie with data stories

As the saying goes, “if you torture the data long enough, it will confess to anything.” Statistics can be manipulated to support any conclusion, but that certainly doesn’t mean that it should be.

For example, multiple research projects corroborate the extent of climate change. Yet even a phenomenon as irrefutable as climate change can be dismissed by climate change deniers with malicious data manipulation. Presenting cherry-picked data, they pass pseudo-evidence as facts while discrediting greening efforts as alarmists. “Global warming is not real because global temperatures dropped in the years 2008 and 2011,” they claim. An observer who sees the chart for the first time might agree.



**Fig 4:** A cherry-picked chart on global temperature



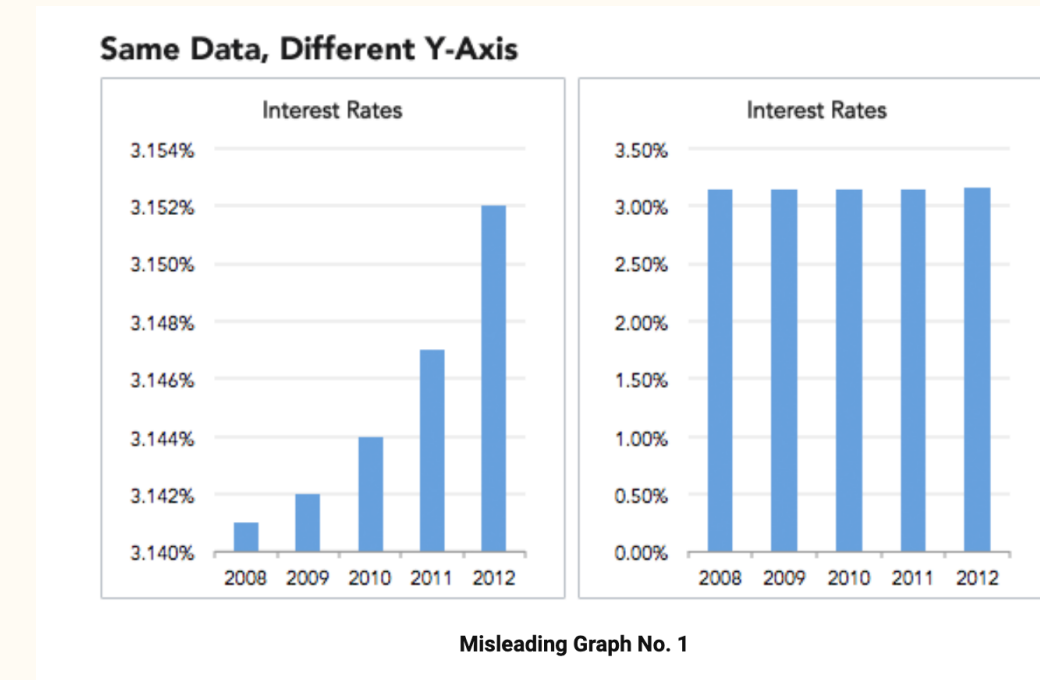
**Fig 5:** The complete chart on global temperature

Such a claim ignores the fact that global temperatures naturally fluctuate from time to time due to events like El Ninos, volcanic activities, and ocean conditions. When such effects are taken into account, and data are presented from the 19th century to now, it is clear that the earth is warming.

Presenting cherry-picked data is fallacious and arguably unethical. Conceivably, analysts who commit such an error run the risk of losing their hard-earned credibility. Conversely, presenting the complete story, warts and all, is essential in gaining the trust of the audience.

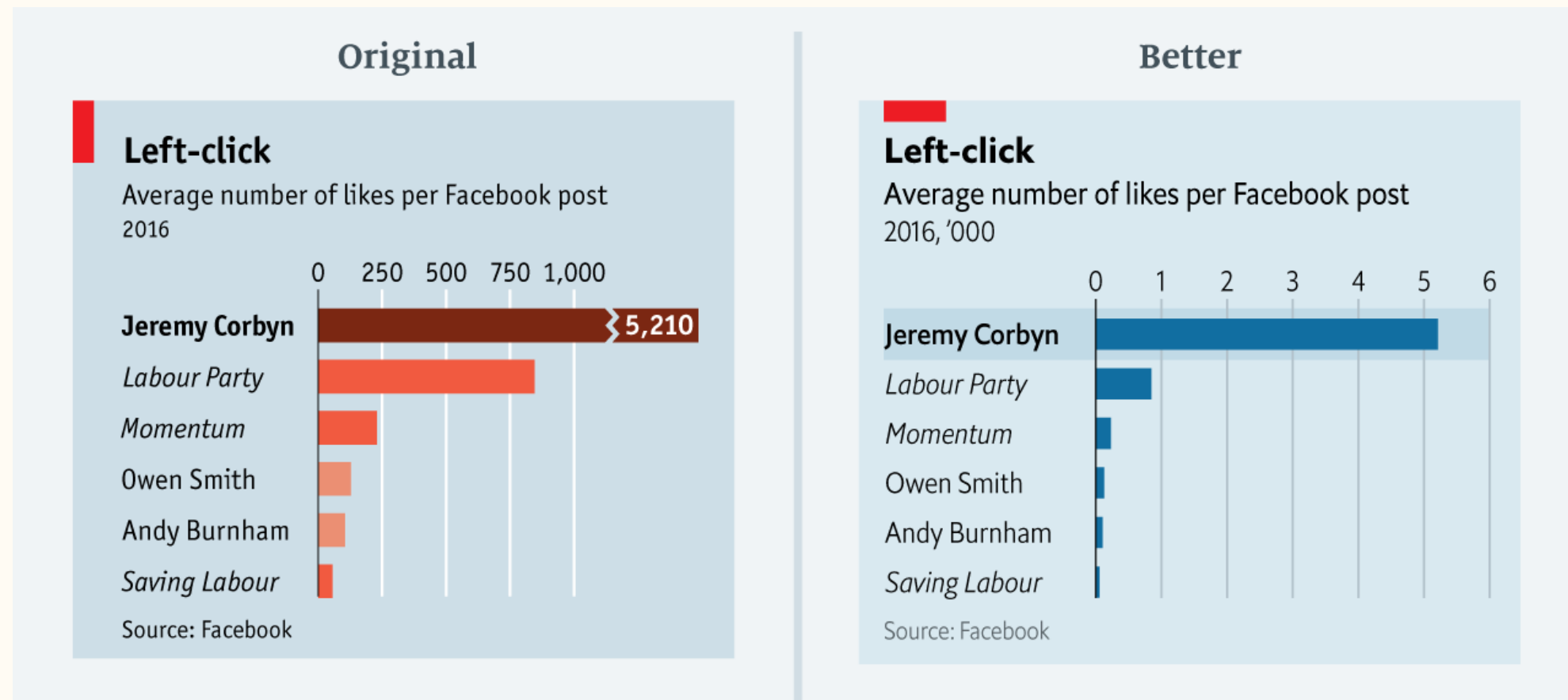
It is the responsibility of the data storyteller to represent the data truthfully. Apart from avoiding cherry-picking, data storytellers should strive to avoid data misrepresentation. To do so, the book *How to Lie with Statistics* offers a few pieces of advice.

- **Ensure that axes start with zero since manipulating axes range can result in misleading correlations.**



**Fig 6:** The same data points can convey dramatically different results

- Ensure that the axes scales are appropriate



**Fig 7:** By changing the scales of the y-axes, a forced correlation emerges

- When needed to select a sample for a data story, ensure that the sample is representative of the population. A sample that is too small can yield extreme (very large or very small) results, whereas an unrepresentative sample can yield incorrect inference about the population.
- Use mean, median, and mode appropriately to ensure that the average is truthful to the population especially when dealing with a skewed dataset.

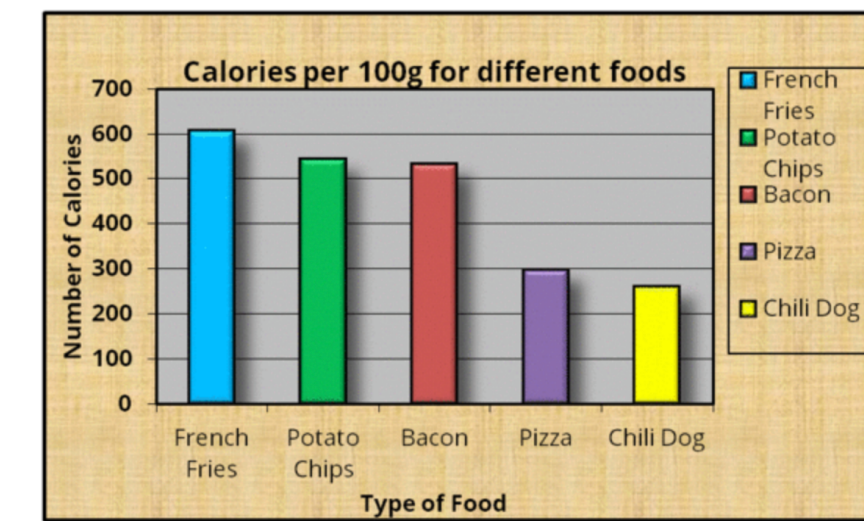
# 5. Keep your visualizations minimal and avoid clutter

Clutter in visualizations takes up the cognitive load while not providing additional value. Removing distractions from a visualization focuses the attention of the audience on the core message of the chart.

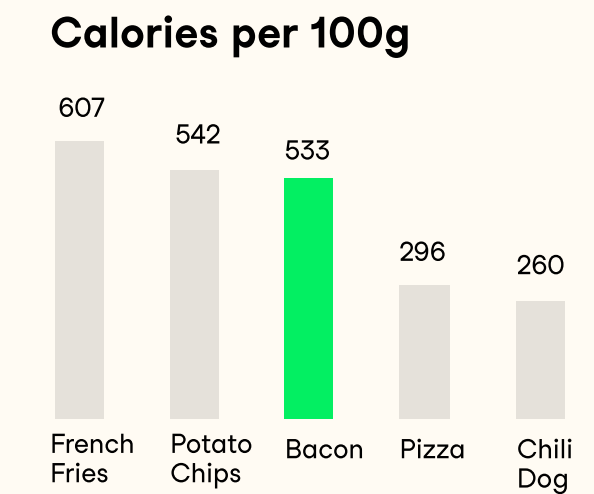
To quantify the amount of unnecessary data in a chart, Edward Tufte presented an influential concept called the data-ink ratio, which is the proportion of ink used to present the actual data to the total ink used to print the graphic. Good graphics should have a high data-ink ratio.

To maximize the data-ink ratio, analysts should strive to minimize superfluous words, colors, and lines without sacrificing clarity. In particular, here are some ways one can keep visualizations minimal yet effective.

- Use white spaces
- Remove chart borders
- Remove gridlines or axes
- Clean up axis labels
- Label data directly (as opposed to using a legend)
- Remove data markers
- Use special effects (bold, underline, italics, shadows) sparingly



**Fig 8A:** A cluttered chart



**Fig 8B:** The same chart, decluttered



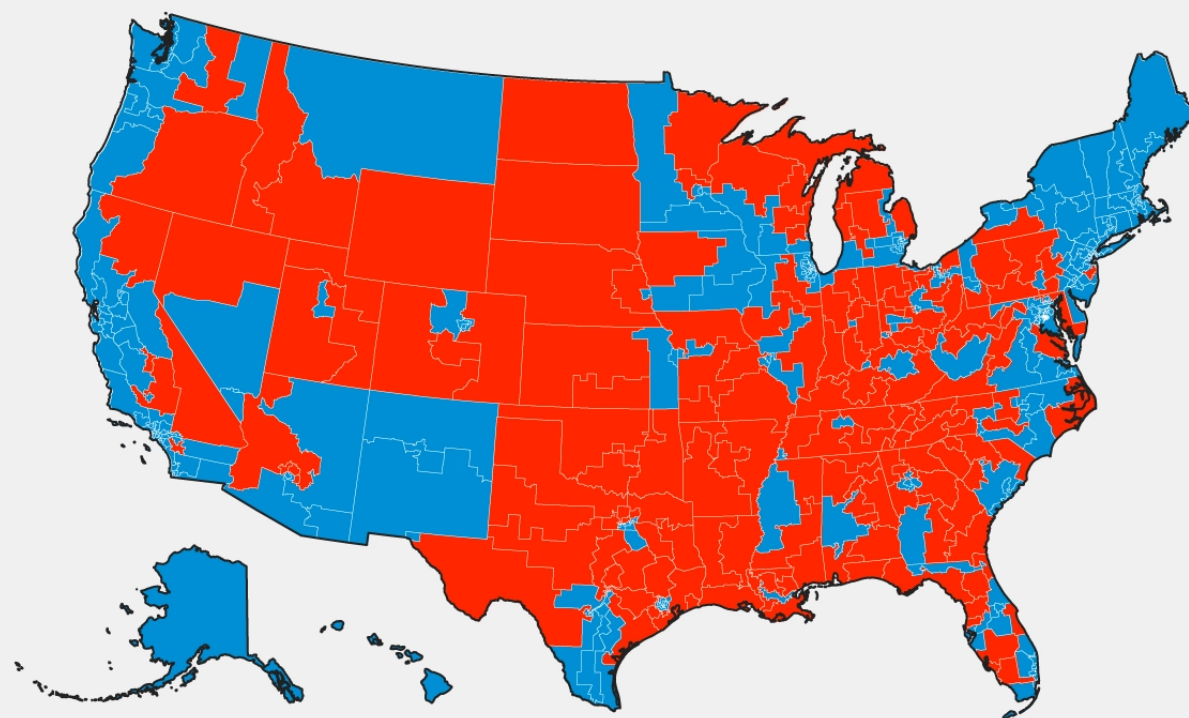
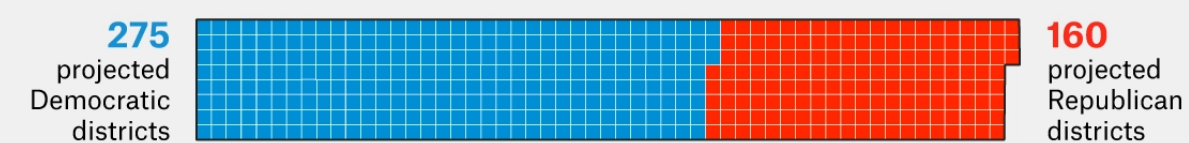
# 6. Add color to your stories

Colors serve three fundamental use cases in data storytelling. We illustrate these use cases using FiveThirtyEight's colorful visualizations.

First, it is used to distinguish between groups that do not have an intrinsic order. A prime example is the use of blue for Democrats and red for Republicans as in Figure 10 below.

## What if only women voted?

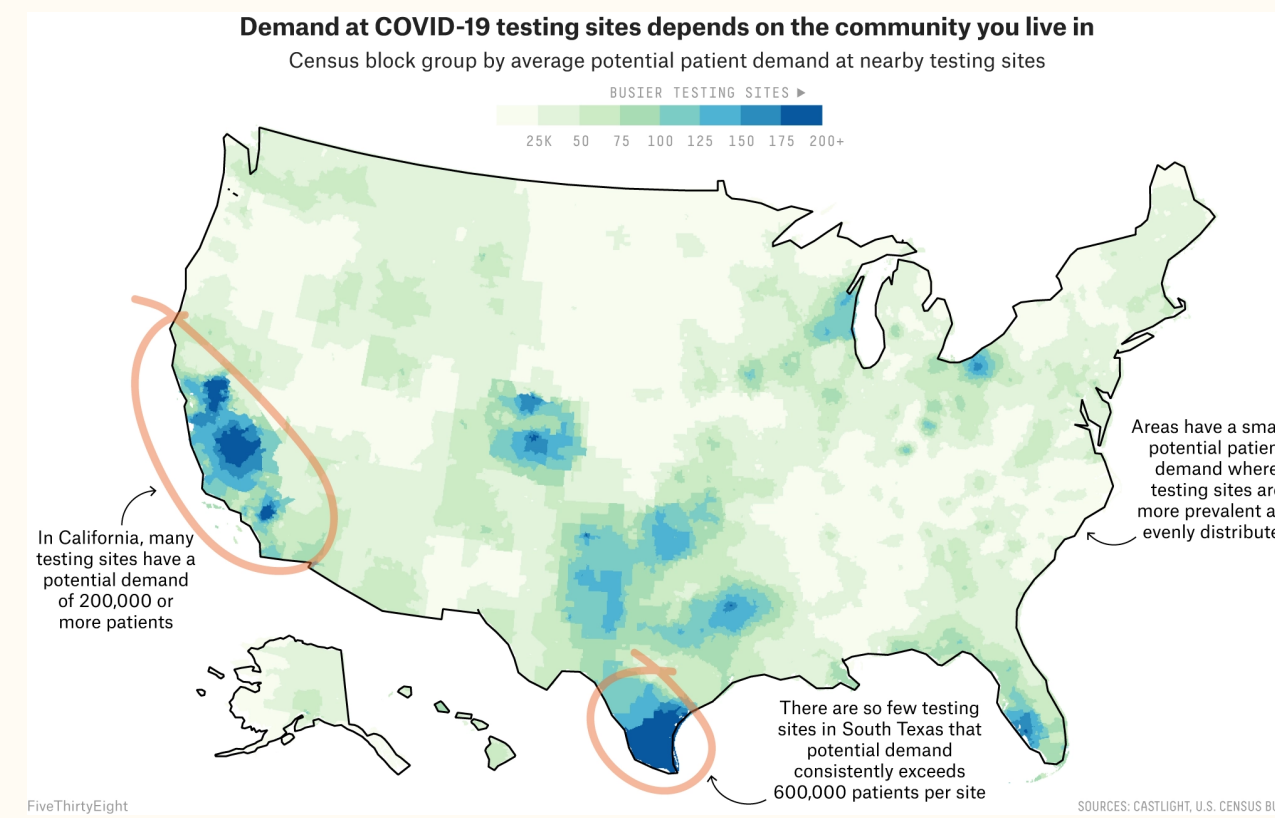
Projected results for the 2018 midterms based on polling patterns and FiveThirtyEight's Lite forecast on Oct. 24



FiveThirtyEight SOURCE: VARIOUS POLLSTERS

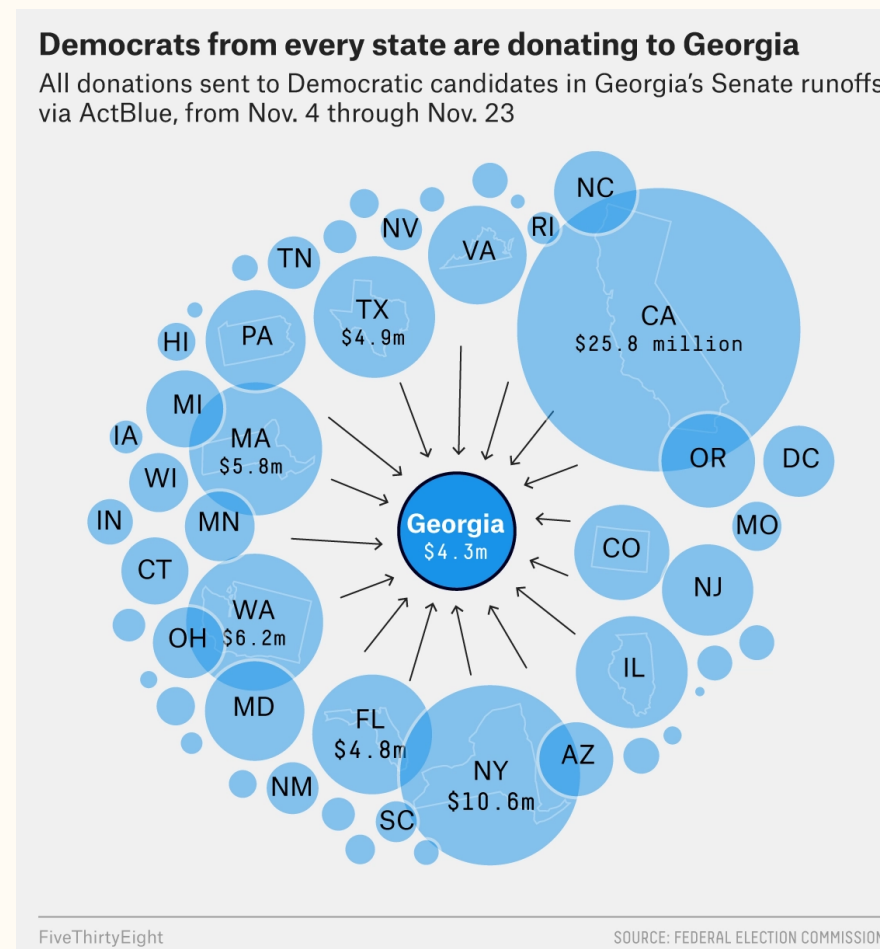
**Fig 9** FiveThirtyEight's use of colors to distinguish between republicans and democrats

Second, colors are used to represent data values. Here, different shades of similar colors represent the varying magnitude of a metric. In the following example, we can immediately identify areas with darker colors as locations of busier Covid-19 testing sites.



**Fig 10** Use of color intensity to highlight prevalence of COVID-19 testing sites

Third, it is used to highlight specific elements in the data. For instance, a dark color can be used to highlight amongst other elements of lighter color.



**Fig 11:** Georgia is highlighted since it is the main character of the data story

Colors are a double-edged sword of data storytelling. While they can inject life into data stories, they can also potentially kill the interest of the audience. To avoid this pitfall, Datawrapper suggests the following best practices.

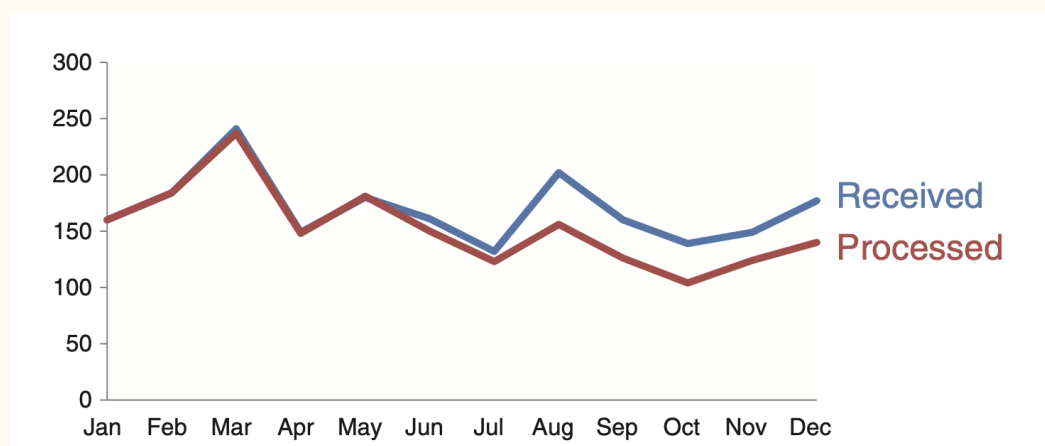
- Explain what your colors encode
- Use colors sparingly
- Use light shades for low values and dark shades for high values
- Use intuitive colors (e.g. red for bad, green for good)
- Use gray generously for less important data
- Cater to color-blind audiences

# 7. Use texts appropriately

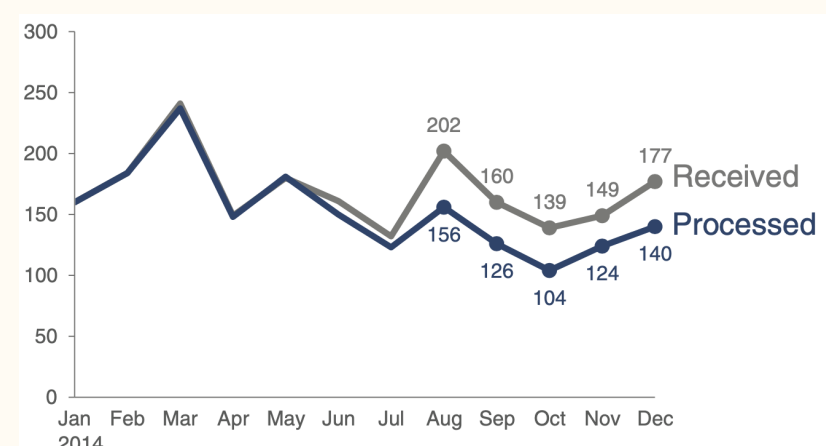
Texts clarify the meaning of charts and make them accessible. Too much text adds clutter while too little text causes confusion. Here are some general rules for using texts in your data stories.

## Label data points when necessary

At times, it is useful to label important points directly on a graph to highlight a certain point. In the chart below, we see that the points are labeled to emphasize the diverging pattern of tickets received and processed.



**Fig 12:** Data points not labeled



**Fig 13:** Data points labelled to convey a point

## Label axes and titles for clarity

Without the axes and the titles, the audience is left guessing the meaning of the chart, as is the case for Figure 14. Thus, it is crucial to add texts to name the metrics being plotted to avoid miscommunication.

When the goal of the visualization is to drive a certain action, using actionable chart titles can quickly call the audience to action. It is also helpful to add notes to the footnote section of the chart for further details.

## Please approve the hire of 2 FTEs

to backfill those who quit in the past year

### Ticket volume over time



Data source: XYZ Dashboard, as of 12/31/2014 | A detailed analysis on tickets processed per person and time to resolve issues was undertaken to inform this request and can be provided if needed.

**Fig 14:** Use actionable titles

## Choose different font sizes

You may also notice the differences in font sizes in the charts above. In general, font sizes need to be legible, and the difference in sizes depends on the relative importance of the texts in conveying the message. In the chart above, the call to action is the most prominent while the footnote is the least thanks to the choice of word sizes.



# 8. Develop a narrative around your data

[Cognitive psychologist Jerome Bruner](#) suggests we are 22 times more likely to remember a fact when it has been wrapped in a story. A narrative connects the dots between facts and makes them more memorable.

The first step to data storytelling, according to [Tableau's whitepaper](#), is to find the data story. To do so, one needs to identify the core elements of the story.

- Who is the protagonist?
- What is the challenge?
- What should the audience do at the end of the story?

With these elements identified, one can write out the storyboard, which helps the data storytelling in identifying the key visualizations that support the storyline. Metaphors and anecdotes are also particularly helpful in making the story more personal and memorable.

Hans Rosling, a global health expert and the founder of [GapMinder](#), developed a narrative around global health data to effectively debunk myths about the developing world in his influential TED Talk [Stats that Reshape Your Worldview](#).

In this data story, the developing countries (the protagonists) overcame the HIV epidemic and overpopulation (the challenge) with public health measures before eventually reaching life expectancies on par with developed countries (the ending). As the audience engaged with the visual narrative, Rosling is one step closer to Gapminder's goal of overcoming systematic misconceptions about global trends.



**"The power of data storytelling is if we combine the right data with the right narrative and the right visuals, we have something that's very powerful that can really drive change and alter people's perspectives."**

— Brent Dykes, Effective Data Storytelling  
Author of [Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals](#)

# Everyone will be a data storyteller

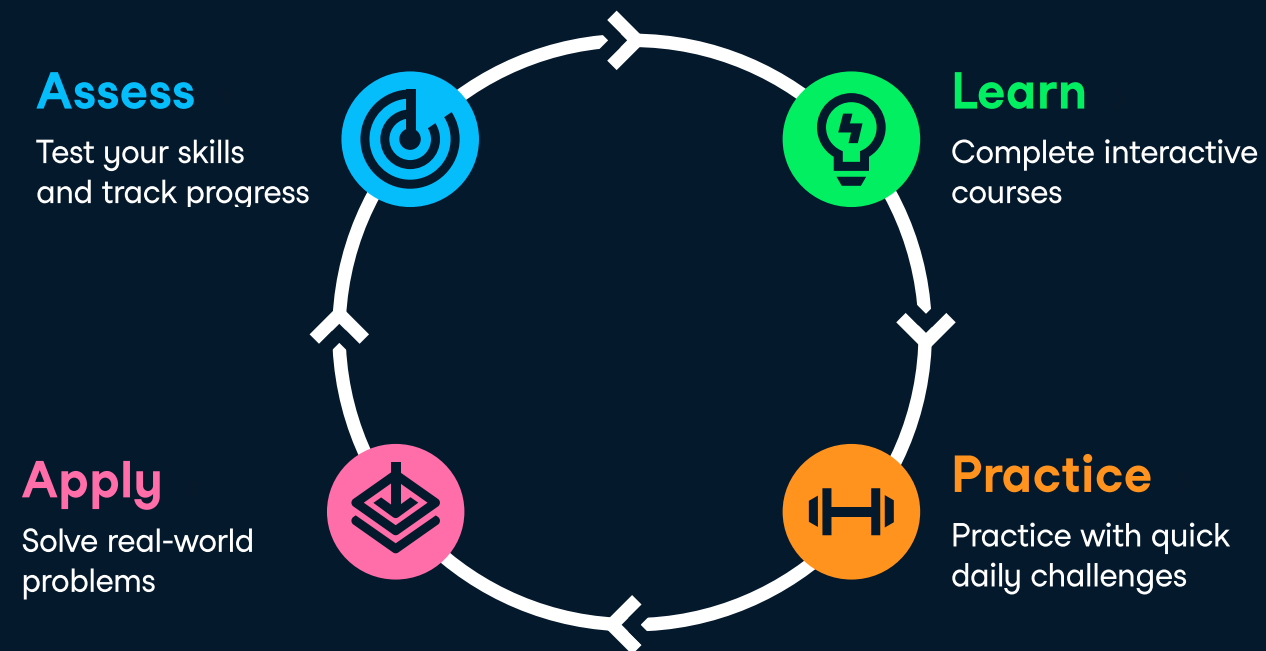
Pablo Picasso once espoused, “learn the rules like a pro, so you can break them like an artist.” While these eight rules serve as a general guideline for data storytellers, they are by no means the commandments of data storytelling. Data storytellers should recognize that there are always exceptions to the rule, much like how the Economist [consciously broke the rules in data visualization](#) in creating masterful data stories.

One should also note that these eight rules to data storytelling are by no means exhaustive. Instead of aiming to learn every single theoretical rule, one should aim to start crafting and telling their first data stories as soon as possible. The experience gained from data storytelling will pay dividends as [data storytelling becomes a necessary skill](#) set for all in the coming decade, and data insights become a cornerstone of organizations.

# Build a Team of Data Visualization Experts with DataCamp

Take your data storytelling to the next level with Data Visualization best practices

DataCamp's extensive data visualization curriculum can help the data storytellers of the future hone one of the key elements of data storytelling: DataCamp's proven learning methodology provides a cyclical process for learning and retention. This learning methodology enables learners across the data literacy spectrum to assess their skills and identify gaps, develop a learning plan based on these gaps, practice skills, and apply them in a real-world setting. Practitioners of any skill level can upskill on the latest data tools, techniques, and concepts.

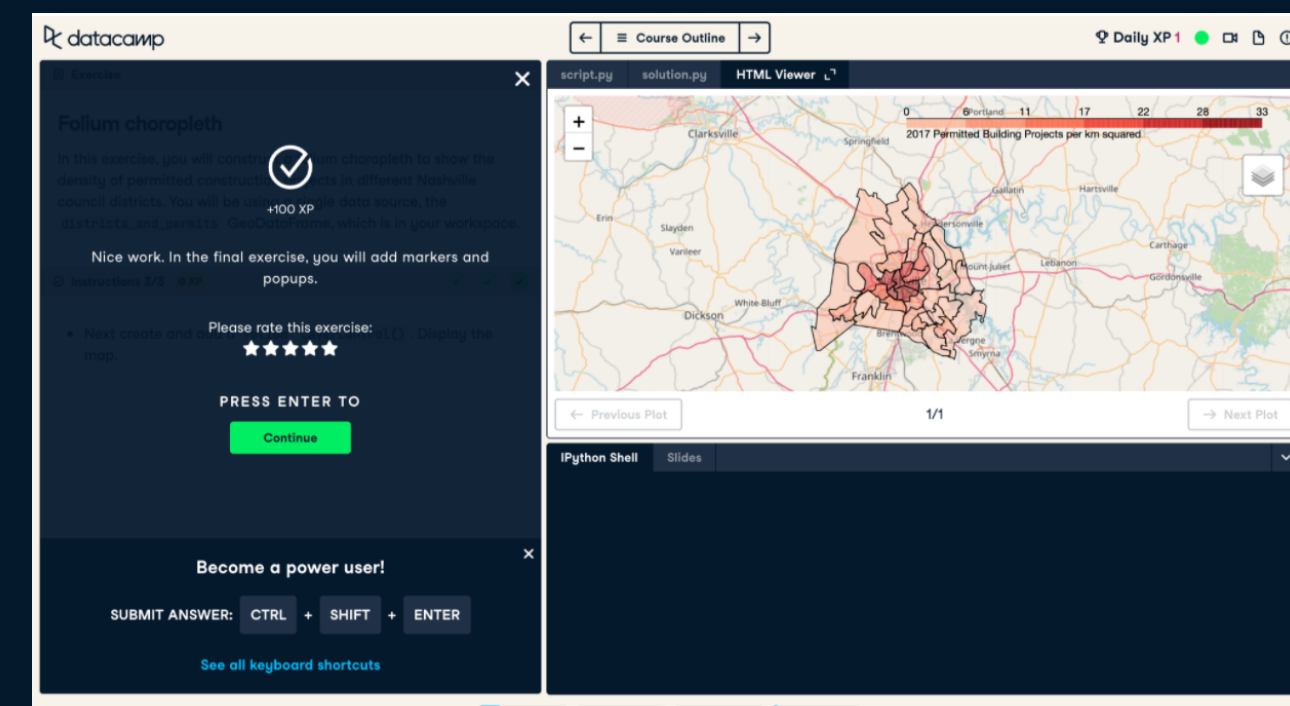
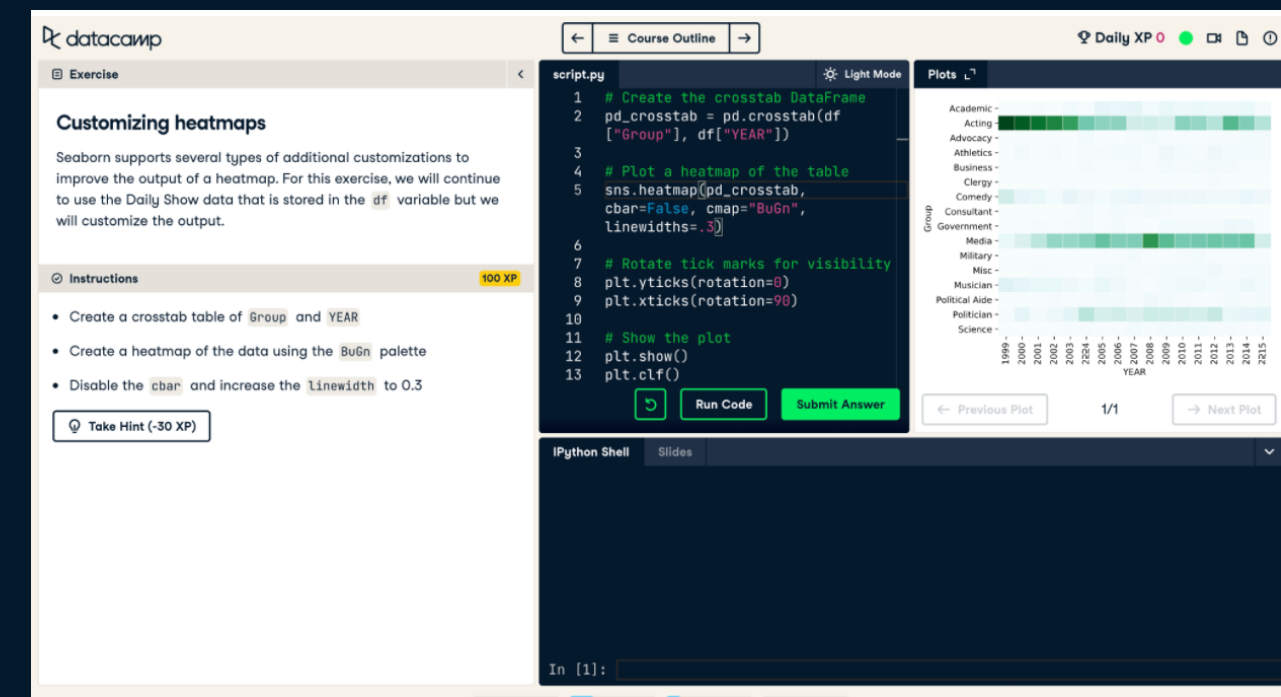


Speak with a learning expert

# Data Visualization Courses at Your Disposal

## Data Visualization with Python

DataCamp provides a host of [data visualization](#) courses across the most popular Python packages. Whether it's matplotlib, seaborn, bokeh, plotly, or dash, you'll be able to develop succinct data visualizations and dashboards that are deeply customizable for your data stories.





## Data Visualization with R

If you're looking to sharpen up your [data visualization](#) skills in R, we also have you covered. R is considered one of the best tools for data visualization and reporting. Whether using ggplot2, leaflet, or plotly for data visualization, or the shiny package for developing interactive dashboards, you'll be able to design better visualizations that complement your data story.

The screenshot shows a DataCamp exercise titled "Bubble charts". The interface includes a "script.R" editor with the following R code:

```
1 # Change the sizemode so that size
2 # refers to the diameter of the points
3 us_economy %>%
4 filter(year == 2017) %>%
5 plot_ly(x = ~gdp, y = ~house_price)
6 %>%
7 add_markers(size = ~population,
8             color = ~region,
9             marker = list(sizemode
10                          = "diameter"))
```

The "HTML Viewer" displays a bubble chart where the x-axis is "gdp" and the y-axis is "house\_price". Bubbles are colored by region: Midwest (green), Northeast (orange), South (blue), and West (pink). The size of each bubble represents the population. The "R Console" is empty, and the "Instructions" section shows progress at 3/3 with 0 XP.

## Tell your Data Stories with Tableau and Power BI

Business intelligence tools were made with data storytellers in mind. Powering the analysts of the future, business intelligence tools like [Tableau](#) and [Power BI](#) provide easy drag-and-drop interfaces to design and deploy dashboards that walk consumers through a data story.

The screenshot shows a Tableau dashboard titled "Side by side dashboards in Story Points". The dashboard is designed to compare customer and subscriber activity. It features a line chart titled "Who rides Divvy bikes?" showing activity by time of day and day of week. The chart is split into two panels: "Customer" and "Subscriber". The "Customer" panel shows a total of 282,749 rides. The "Subscriber" panel shows a total of 6,666 rides. The dashboard also includes a heatmap showing activity by day of the week and time of day. The "Instructions" section shows progress at 4/5 with 100 XP.

# Upskill your team

## Create Custom Tracks

DataCamp makes it easy for you to create bespoke learning paths and assignments to meet the needs of all your roles, teams, and departments.

The image shows two columns of course cards. The left column is for 'Python for Finance' and the right is for 'Market Basket Analysis in Python'. Each column contains three cards: a course introduction, an intermediate course, and a machine learning project. The cards include course titles, brief descriptions, and action buttons like 'Learn', 'Assess', and 'Project'.

## Set Assignments

Assignments are a great way to set clear, time-sensitive learning goals. On average, courses assigned by Enterprise customers have completion rates that are twice as high as unassigned courses.

A grid of six assignment options, each with an icon and a label: Gain XP (XP icon), Course (book icon), Chapter (book icon), Custom Track (network icon), Assessment (power icon), and Project (download icon).

## Track skills development with skill matrix

Track the data skills your team has today and map a path to the skills they need tomorrow. Using the Skill Matrix, admin users can easily filter to identify individuals with the skills you need to take on specific projects or teams with low use or data skills gaps. They can then create and assign custom tracks to help bridge these gaps and report on skill development.

	200	200 ↗	200	200	200	169 ↘
	179 ↗	135 ↗	96	68 ↘	128 ↗	96
	86 ↗	139	112	95 ↘	197 ↗	127
	6 ↗	104	28	58 ↗	7 ↘	19

## Certify your people with certifications

Upskill with an industry recognized credential from the leader in data science education\*. Our Data Scientist Professional Certification lets businesses recognize qualified talent, both internally as well as during the hiring process.



\*Based on competitive analysis and current as of June 2021:  
[datacamp.com/why-datacamp](https://datacamp.com/why-datacamp)

Take your team's data literacy to the next level -  
with customized, tailored data training

Speak with a learning expert