



Data Storytelling

*With effective data visualization
techniques*

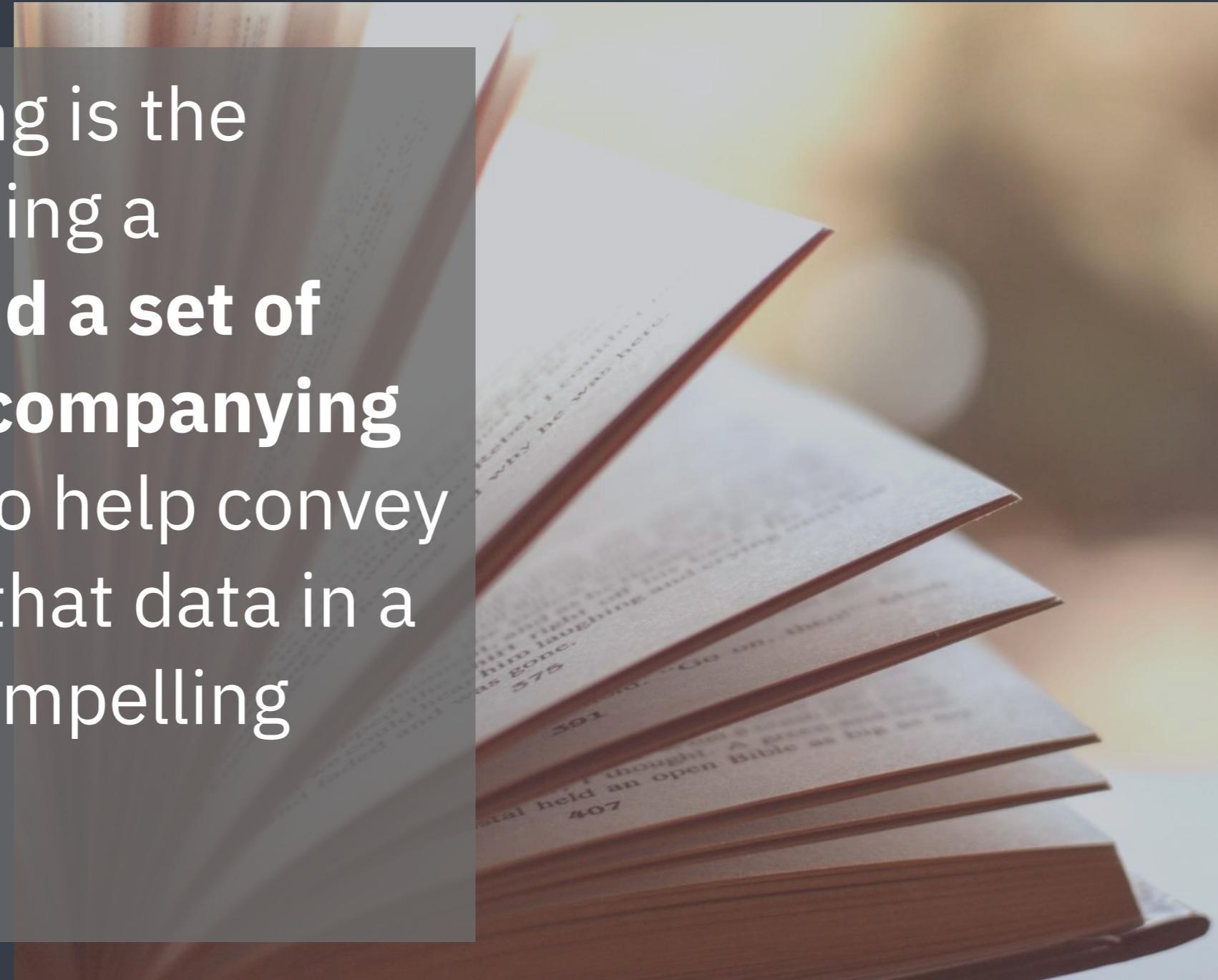
Laura Ellis

STSM, Analytics Architect for IBM Cloud

www.littlemissdata.com

“Data storytelling is the practice of building a narrative around a set of data and its accompanying visualizations to help convey the meaning of that data in a powerful and compelling fashion.”

- TDWI Glossary



WHY SHOULD YOU CARE ABOUT STORYTELLING?

You have invested a great deal of time into your work.

The intention is to have your work acted on / adopted / deployed.

Now you need to communicate the work you have done and convey its value.

a visualization.

1 Evaluate the data frame column types with the inspect_types() function

a) Evaluate the full data frame: allGrades

```
inspect_types(allGrades, show_plot = TRUE)
```



```
## # A tibble: 2 x 4
##   type     cnt  pcnt col_name
##   <chr>    <int> <dbl> <list>
## 1 character    11  64.7 <chr [11]>
## 2 numeric       6  35.3 <chr [6]>
```



“Purposeful
Storytelling isn’t
show business,
it’s good
business”

- Peter Guber

@littlemissdata

A supposed data scientist's time sheet

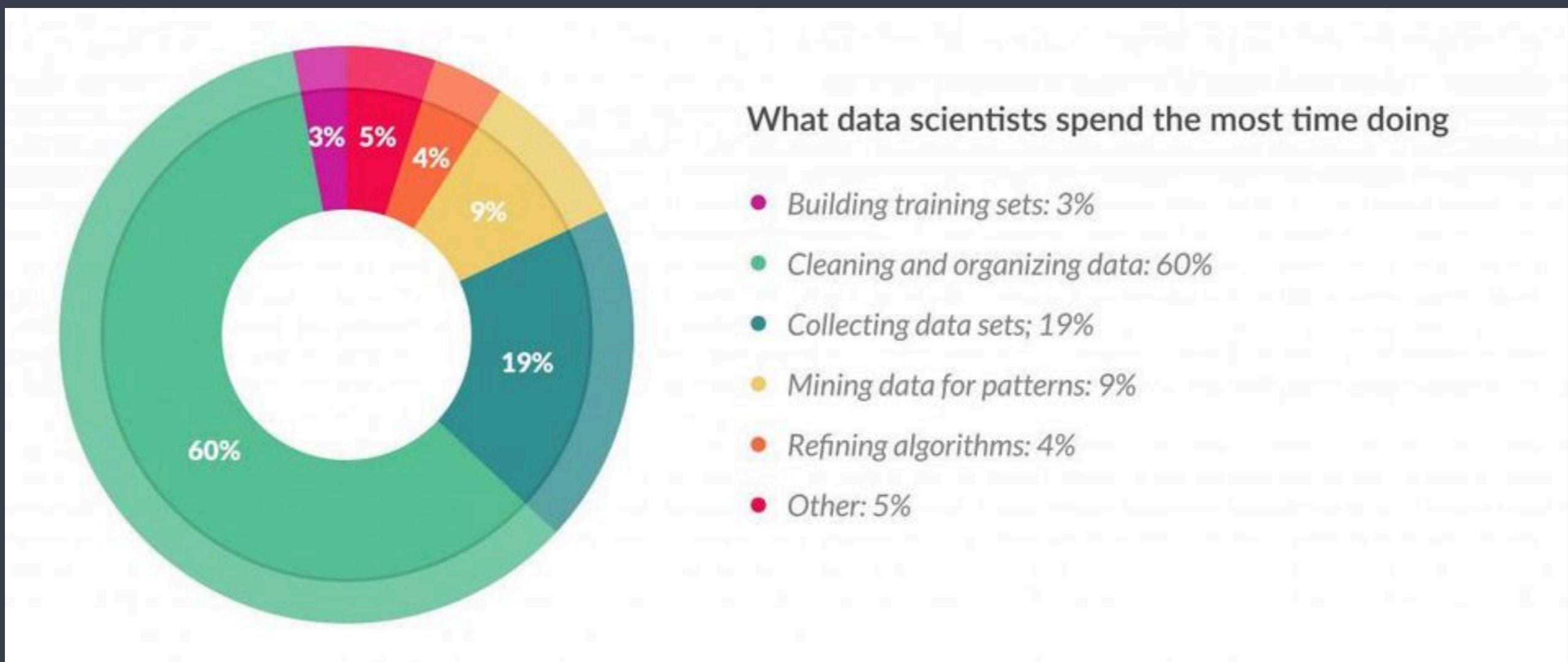


Image Source: forbes.com Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says

A supposed data scientist's time sheet

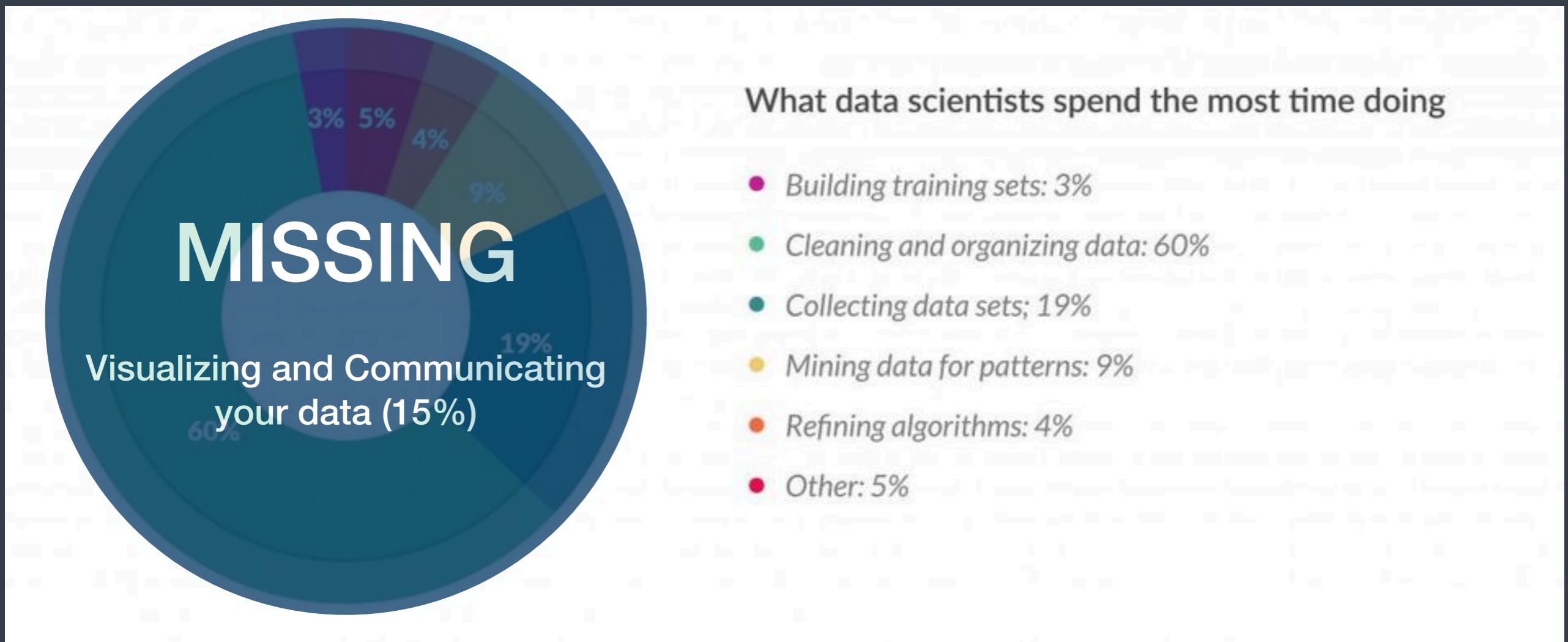


Image Source: forbes.com Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says

Storytelling is a tool in your toolbox

It can be used to further the investment in your data science work.



Image Source: [forbes.com](https://www.forbes.com/sites/forbestechcouncil/2018/01/18/data-storytelling-the-essential-data-science-skill-everyone-needs/) Data Storytelling: The Essential Data Science Skill Everyone Needs



Presentation Guidelines

BUILDING YOUR STORY

PURPOSE



STRUCTURE



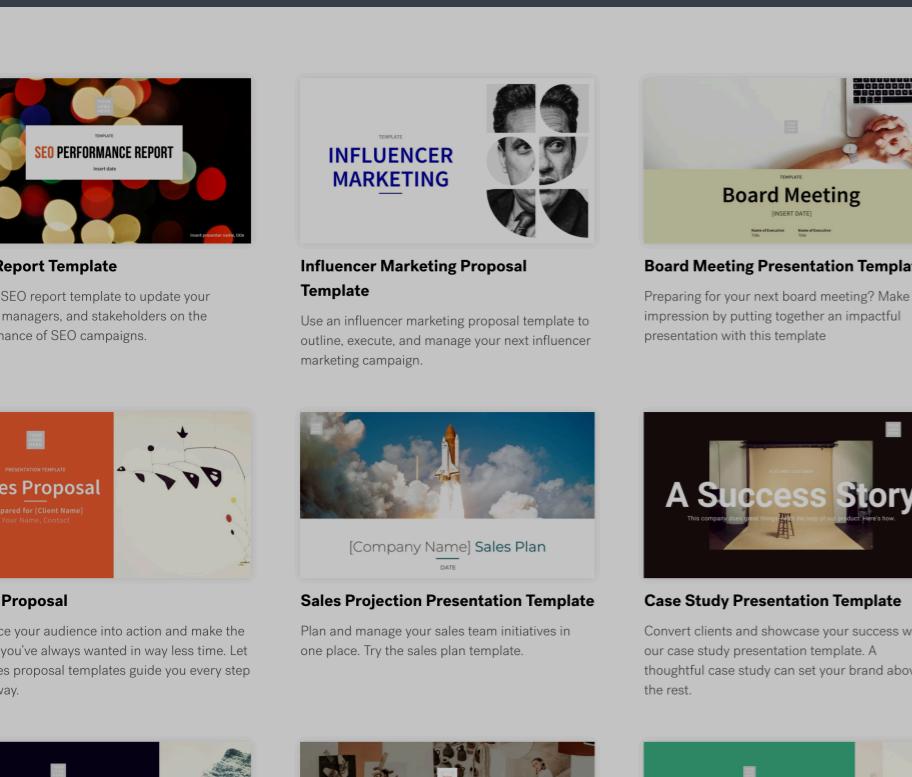
JOURNEY



- Why are we here? Why does audience care?
- What is your purpose and objective?
- What motivates your audience?
- What is your through line?
- Stories have a start middle and end
- Stories are relatable
- Help your audience understand the big picture as you transition across parts of your story
- Give the audience a way to continue the journey after the presentation

FUNDAMENTALS

BUILDING YOUR SLIDES



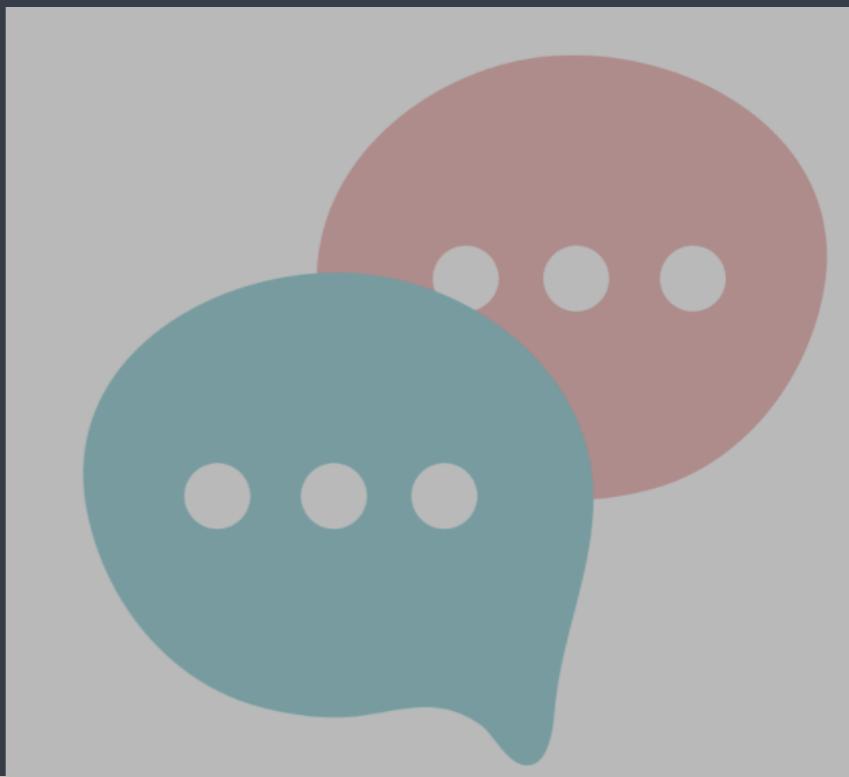
DETAILS MATTER

Theme
Spelling
Spacing



BUILD TRUST

Lineage
Consistency



CONTEXT

What
Why
So What?

@littlemissdata

BUILDING YOUR SKILLS



Practice with an audience

Offline content review



Check in with your audience during the presentation



Seek out more opportunities

Toastmasters

B E F O R E

D U R I N G

A F T E R



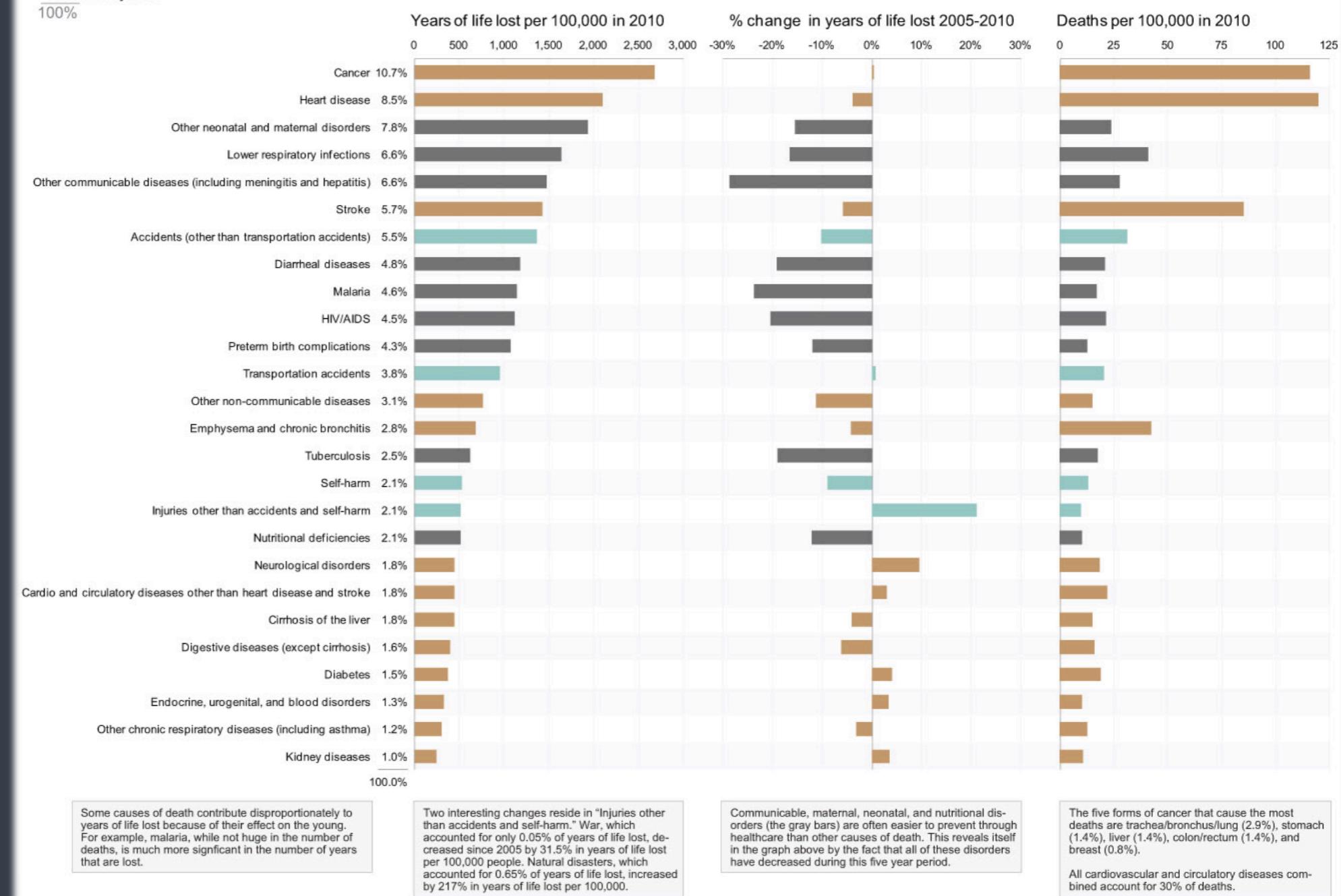
Viz Guidelines

It's OK to be complicated sometimes...

Global Causes of Lost Life

44% ■ Communicable, maternal, neonatal, and nutritional disorders
 43% ■ Non-communicable diseases
 13% ■ Injuries

Comparing the number of deaths alone, as shown in the right-most graph below, doesn't tell the entire story. Some causes of death have a greater effect on the young, which can be seen when comparing years of life lost in the leftmost graph.



Link: [Perceptual Edge](#)

@littlemissdata

... But keep it simple
when communicating
in a presentation

Let your data tell the story

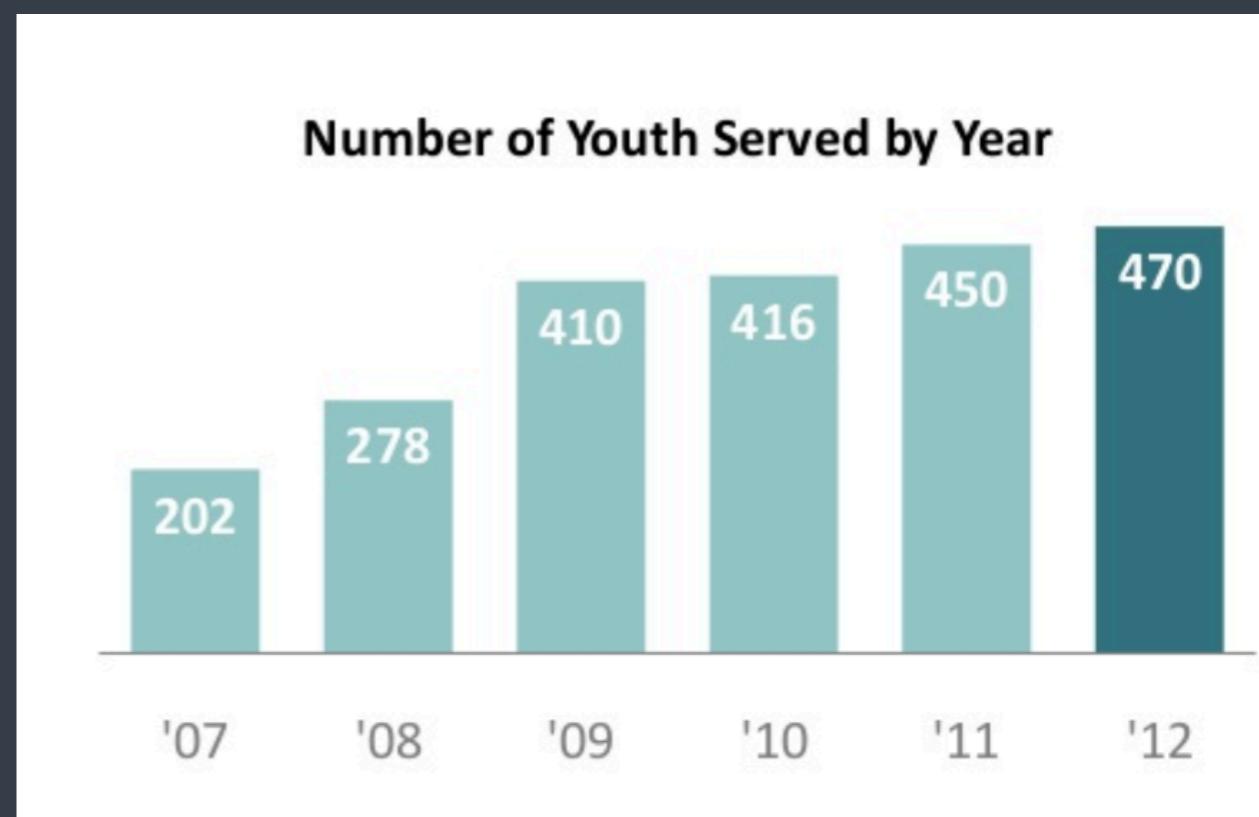
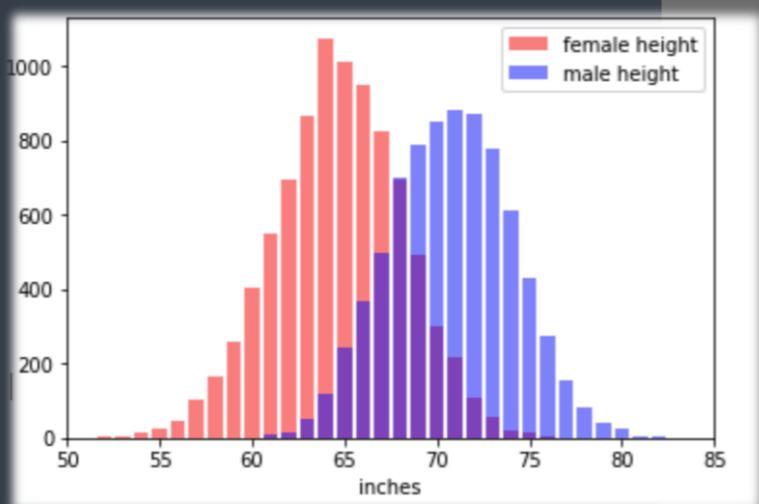
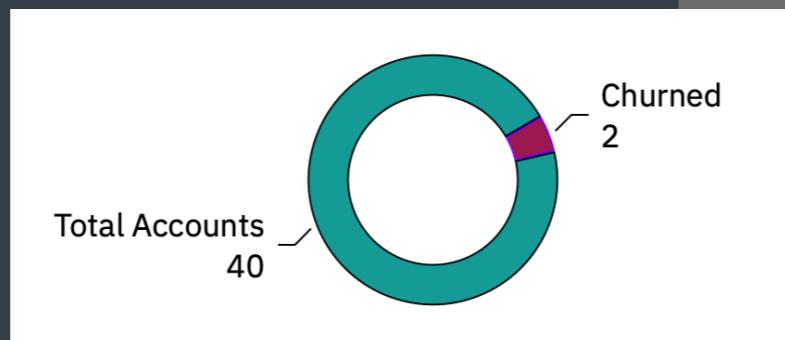
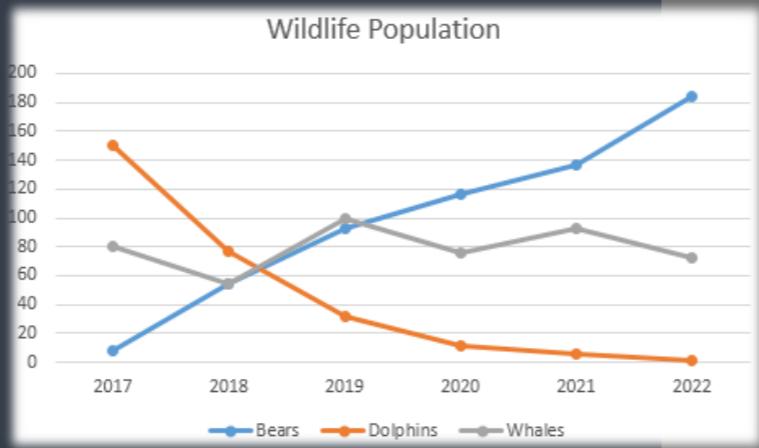


Image Source: [forbes.com](#) Data Storytelling: The Essential Data Science Skill Everyone Needs

Chart Types



Line Graphs

- Useful when you are graphing data over sequential dates.
- Especially helpful with limited number of categories

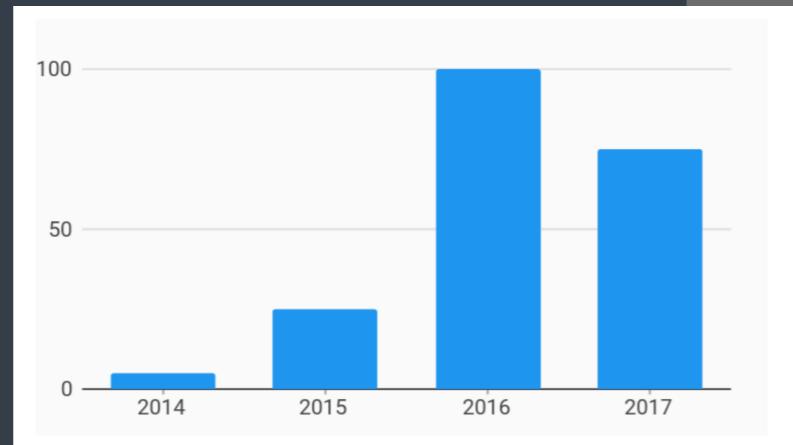
Pie Charts

- A simple and effective way to display percentages. Make sure that the values within the chart always add up to 100%
- Use sparingly. An often more impactful way would be to display a large single value next to the insight.

Histograms

- Use when showing a numeric distribution of accounts such as their spend or the number of services they use.

Chart Types



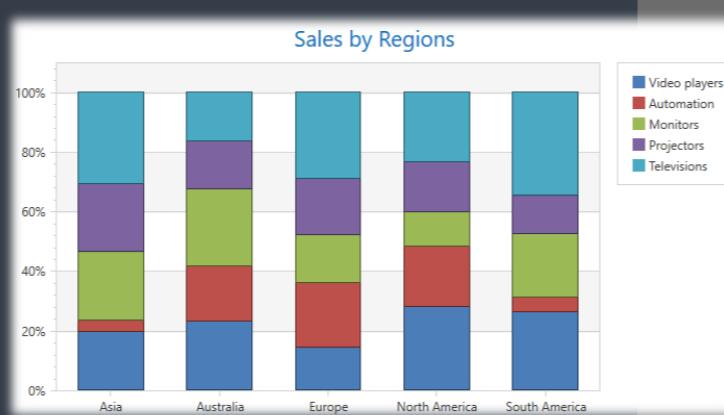
Bar Charts

- Use to display categorical data.
- Bars can be divided up by month, geographic region, product type and account plan type, etc.



Stacked Bar Charts

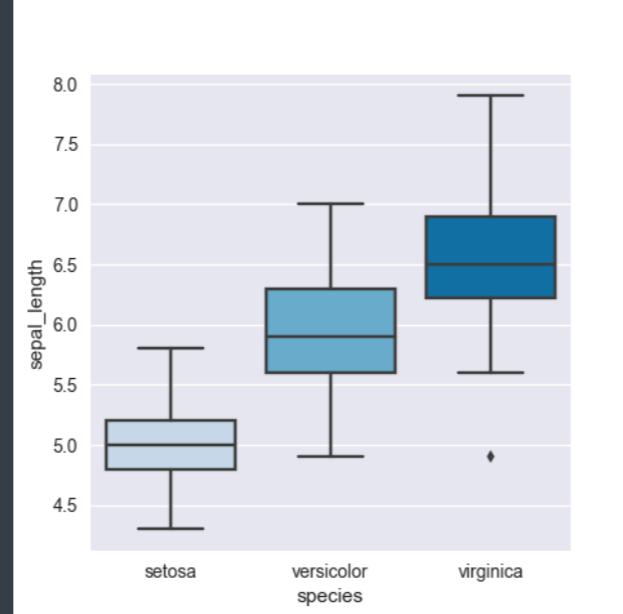
- A stacked bar chart shows 3 types of information per data point (2 axis, cuts within bar).
- These are great in their own purposes allowing the user to take their time and hover over each area to explore, compare and internalize.



100% Stacked Bar Charts

- Scales all bars to 100% allowing you to see the relative population of the categories

Chart Types



Indicator Name	2011	2012	2013	2014	2015	2016	Average	Improvement
Prevalence of Obesity	19.1	23.6	23.3	20.5	24.0	23.2	22.28	↓ -21.47
Prevalence of Tobacco Use	17.4	15.0	15.3	12.2	16.6	16.7	15.53	↑ 4.02
Prevalence of Cardiovascular Disease	5.0	4.9	1.5	4.4	4.9	6.2	4.48	↓ -24
Prevalence of Diabetes	8.0	7.2	9.3	7.2	7.5	10.4	8.27	↓ -30

\$457,885
Last Month's Sales

Boxplots

- Use to show the spread of a group of data.
- Highlights the minimum, median, maximum, 25th percentile and 75th percentile of a data set.
- Use sparingly

Tables

- Display a summary of the data in numerical format

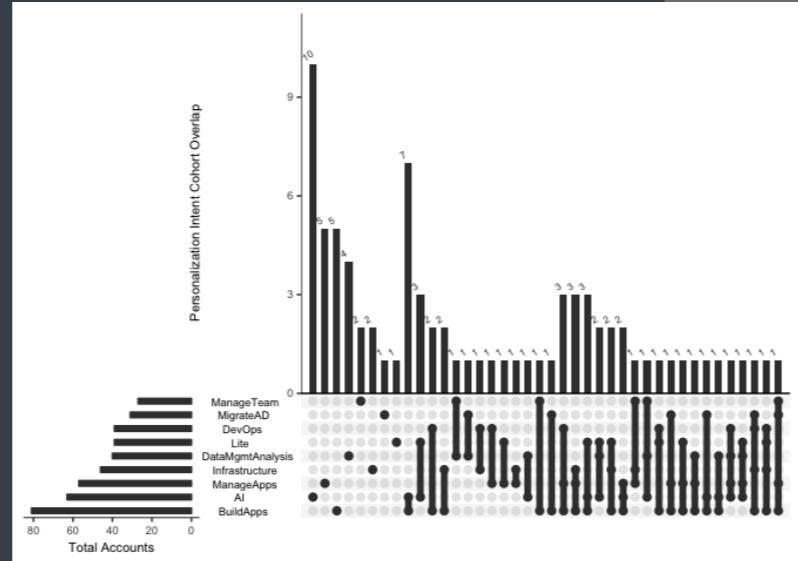
Single Values

- Display a single number metric

General

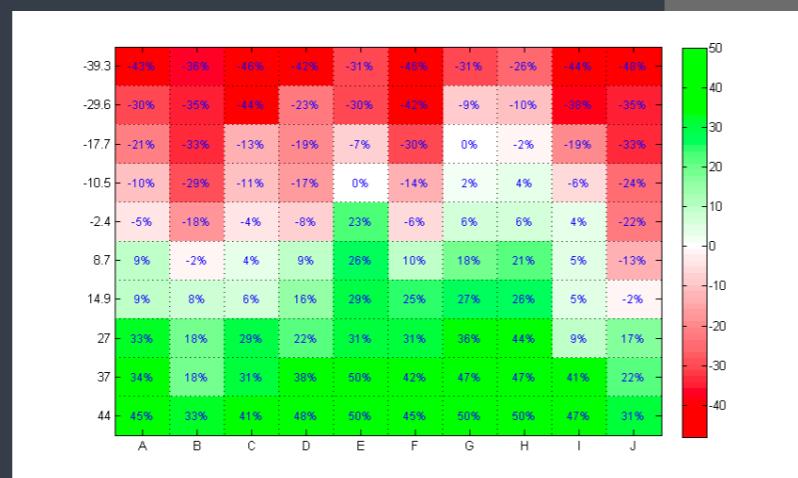
Less is more

- Everything on a slide should be added with intention.
- If a graph is complex, it needs to be there to represent inherently complex information.
- You need to give your audience a chance to digest the graph.
- Insight based graphs (i.e. look at the graph and understand the point immediately) should be incredibly simple.

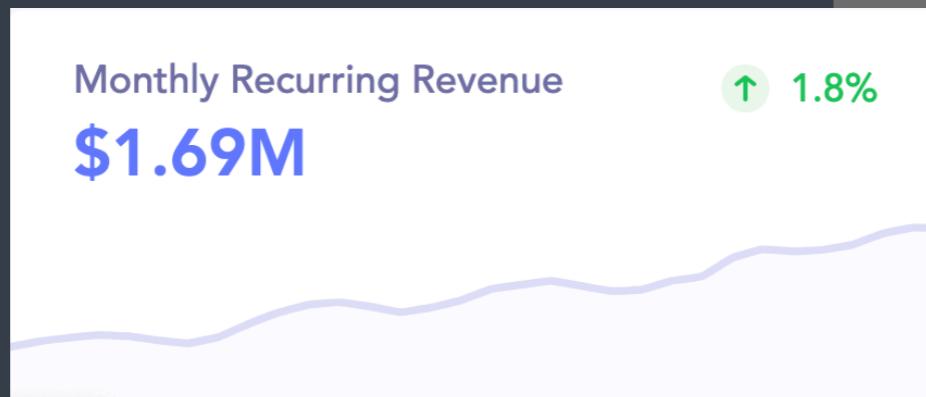
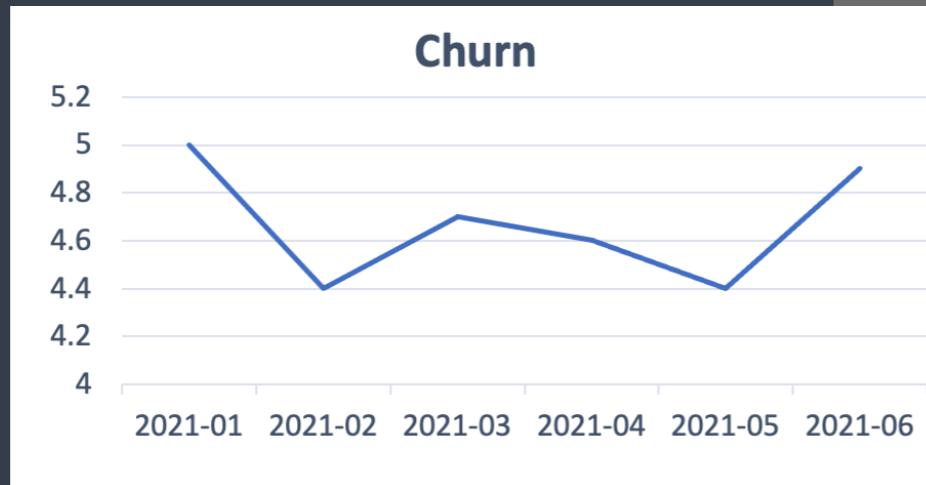


Colors

- Colors are great for easily displaying different categories.
- Use red and green appropriately. Typically red = bad, green = good, blue is often neutral.



General



Scales

- Generally start scales at zero to avoid deceiving your users as to the volume of the issue.
- Always use volume proportionately.

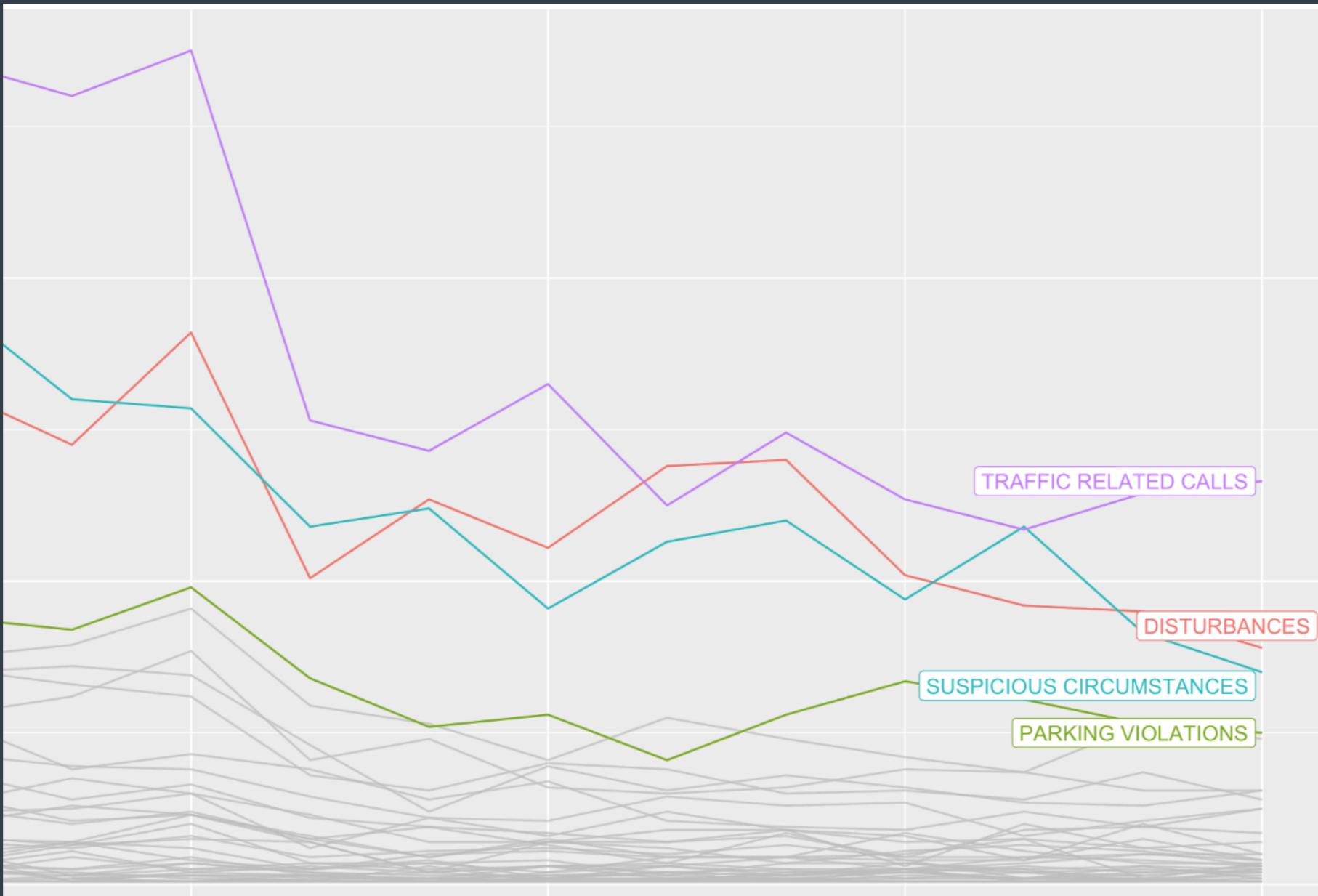
Labels

- The raw value is often not meaningful without the context of its trends.
- Where possible, each graph should display in some form of context (ie the actual/raw value or the relative trend value)



Tips & Tricks

Annotation: Highlight only the story

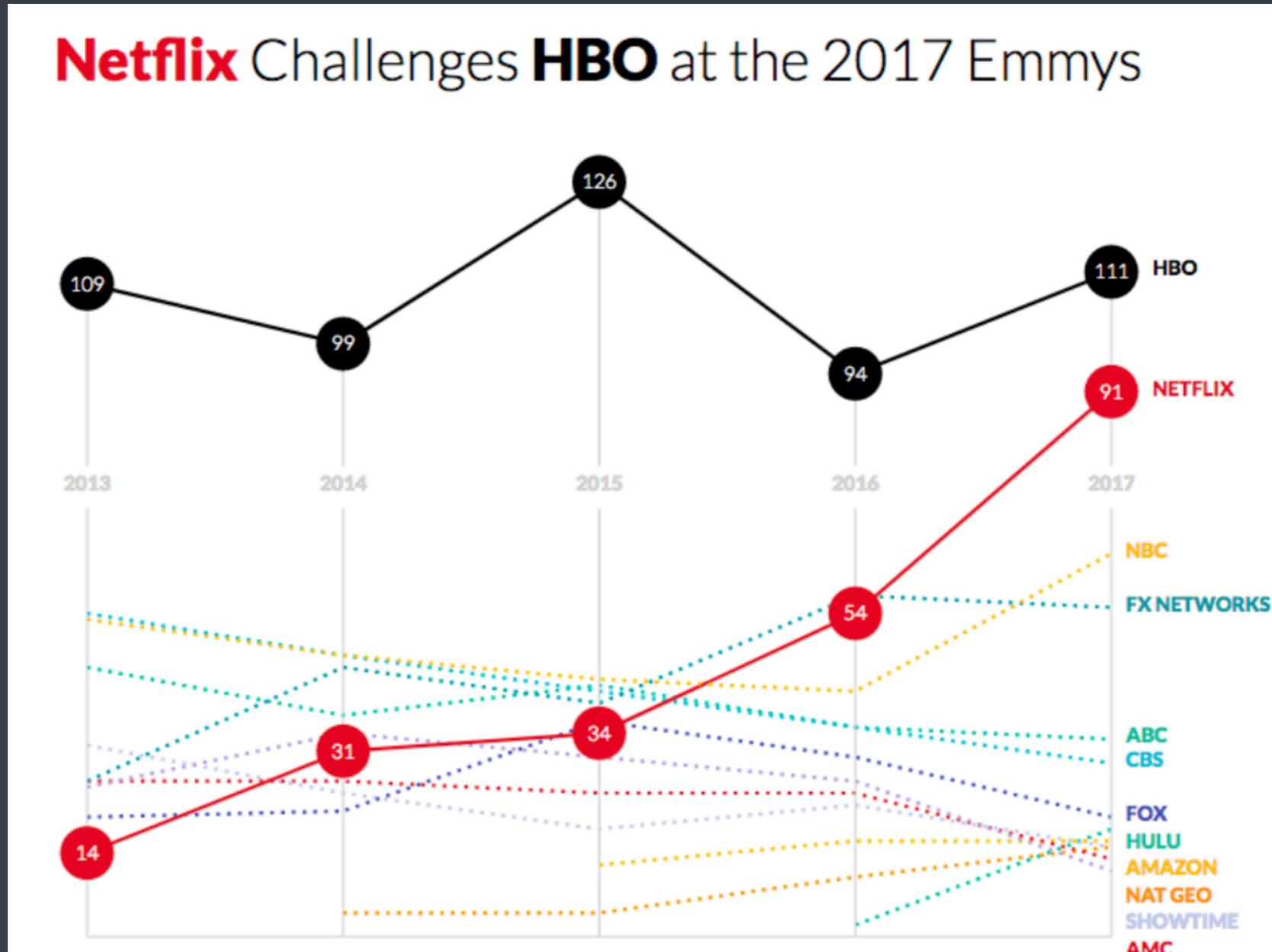


Link: [Little Miss Data Tutorial on Data Highlighting](#)

Focus on the parts of the story that add value

@littlemissdata

Annotation: Highlight with more context

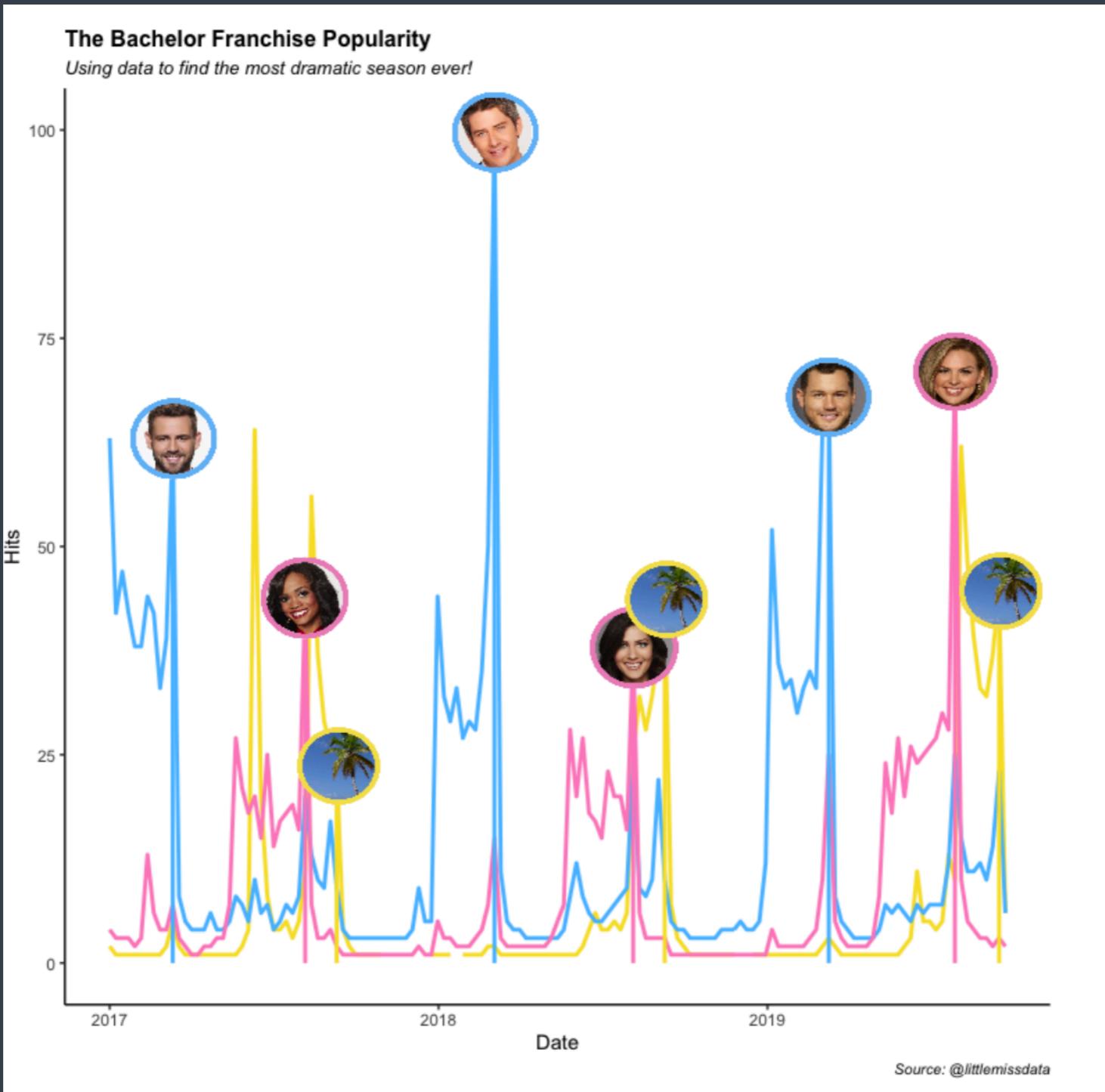


Link: [Susie Lu Emmy Award Nomination Viz](#)

Use annotation strategically to provide extra details

@littlemissdata

Annotation: Other visual cues

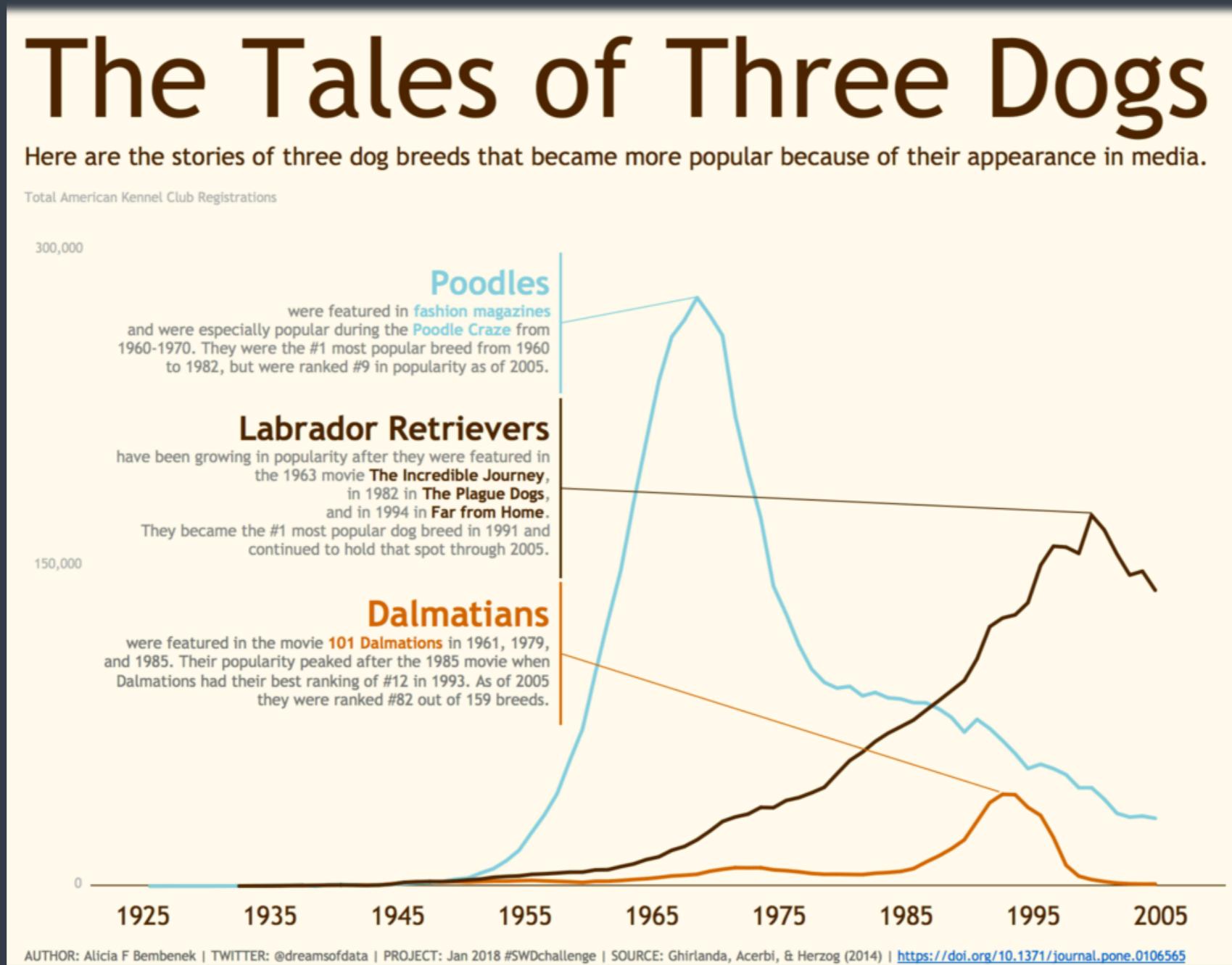


Link: [Little Miss Data GTrendsR + GGImage Tutorial](#)

Try using images as annotation

@littlemissdata

Annotation: Explaining Anomalies

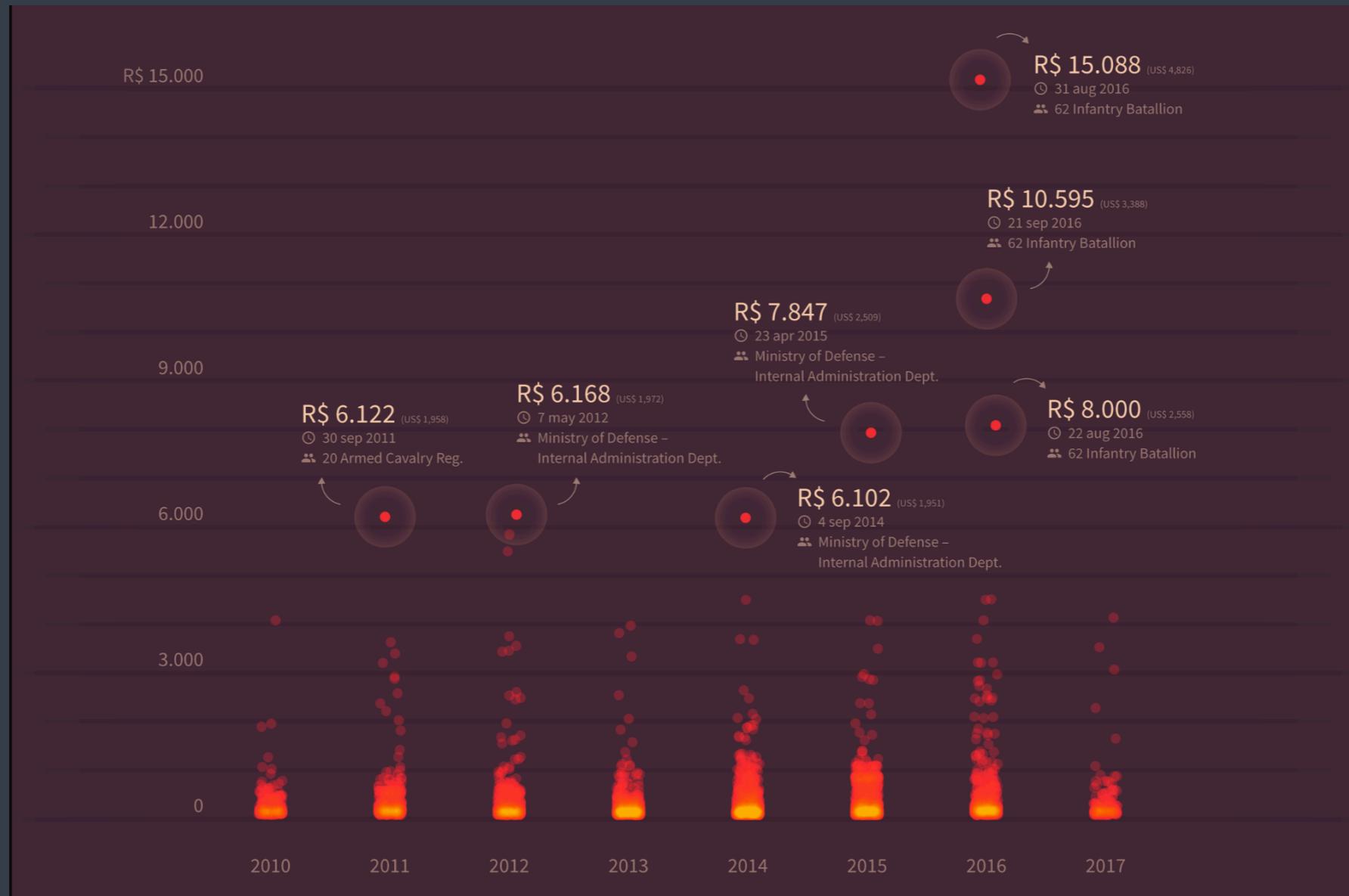


Link: [Storytelling with Data](#)

Explain anomalies through annotation

@littlemissdata

Show the forest through the trees ... and some trees

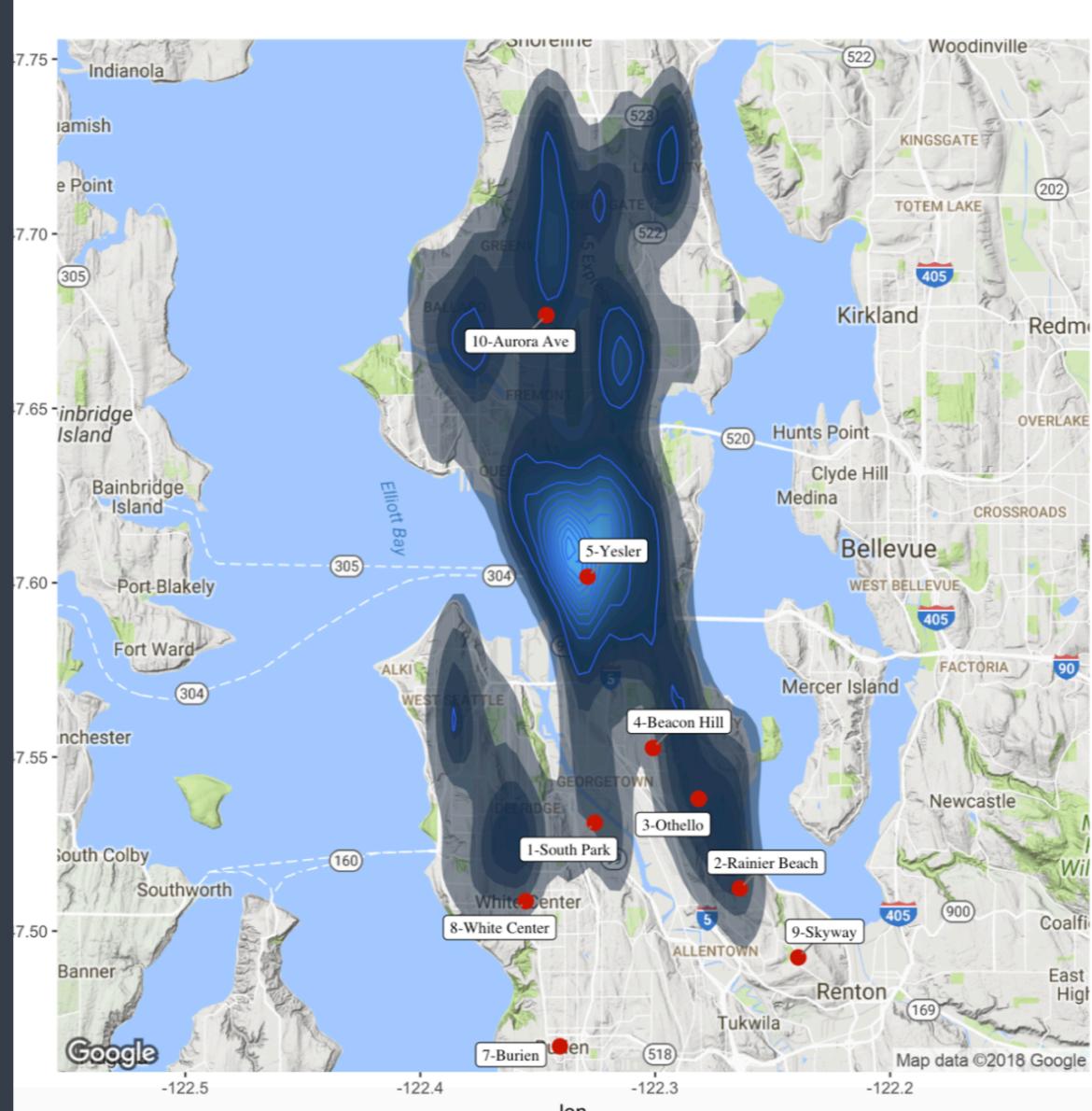


Link: [Data Viz Today Analysis of “All the Government Barbecues” by Rodrigo Menegat and Vinicius Sueiro](#)

Let consumers see the big picture and small picture

@littlemissdata

Make your charts relatable

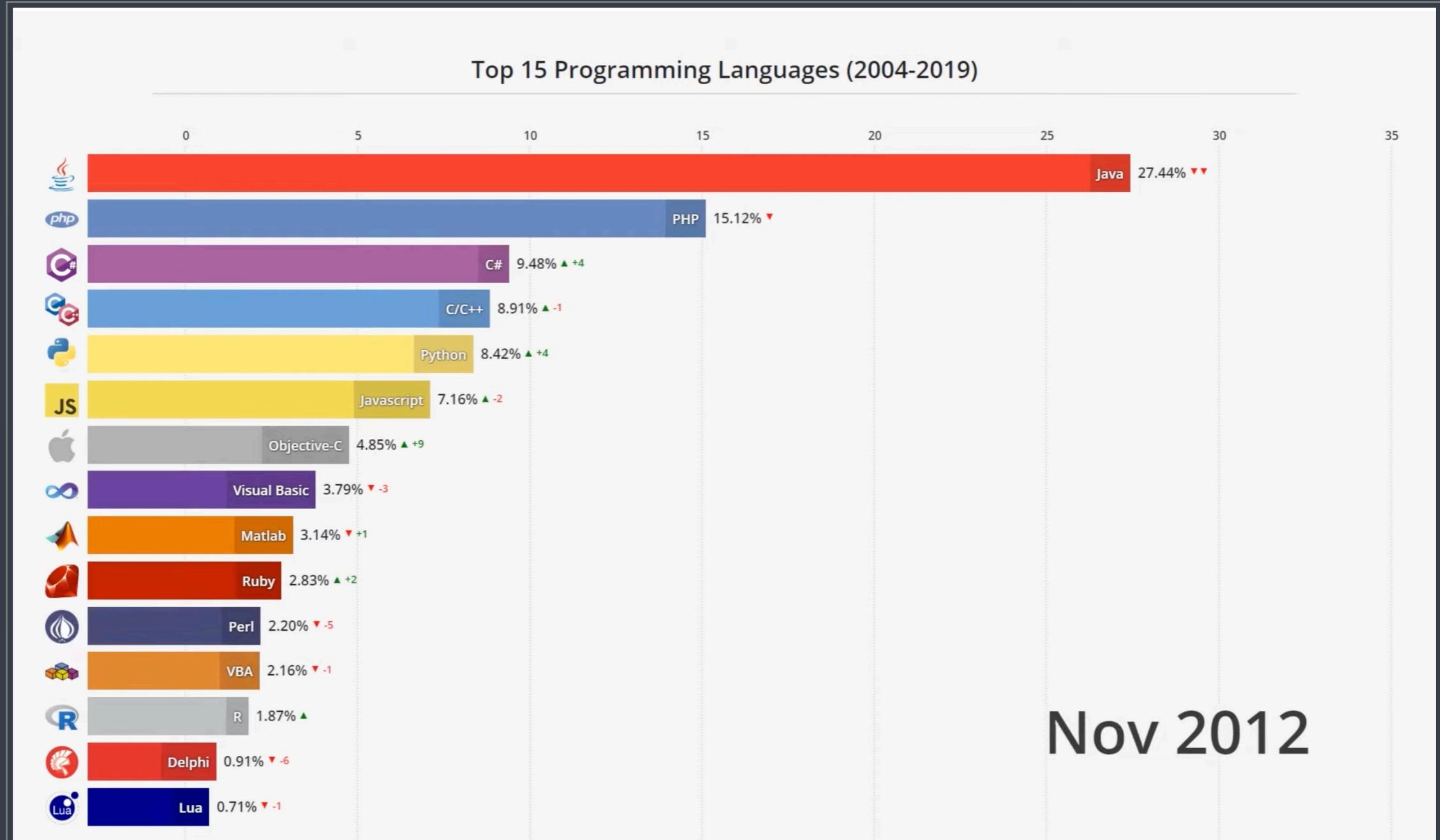


Link: [Little Miss Data Tutorial on Maps](#)

Relate to your consumer, make the content familiar

@littlemissdata

Use some movement



[Link: History in Numbers](#)

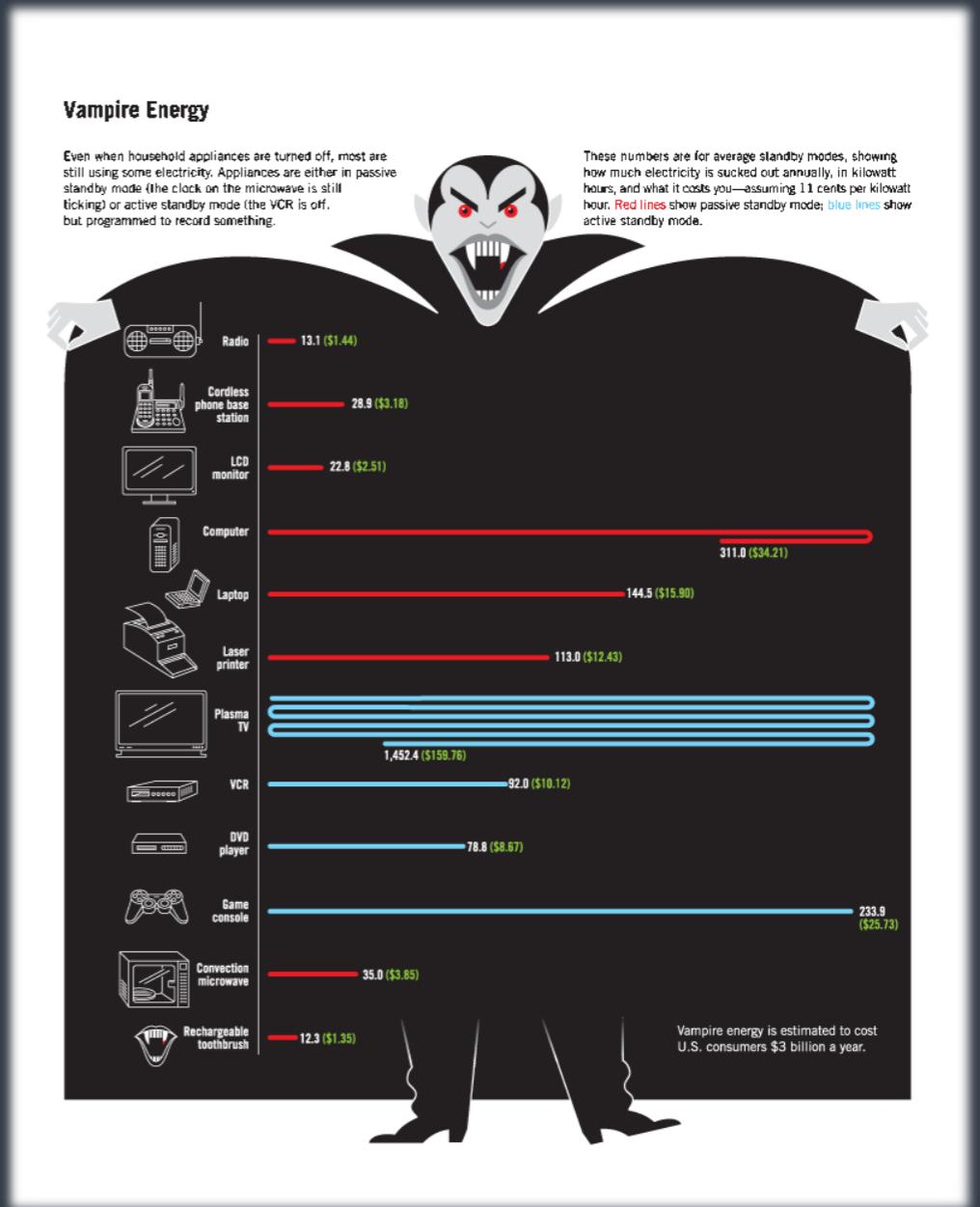
In certain contexts, animation can be powerful

@littlemissdata

It's OK to have fun



Link: [LittleMissData WordCloud R Tutorial](#)



Link: Nigel Holmes Gallery

Fun graphics can be memorable, but only in certain contexts

@littlemissdata



My favorite things

RAWGraphs

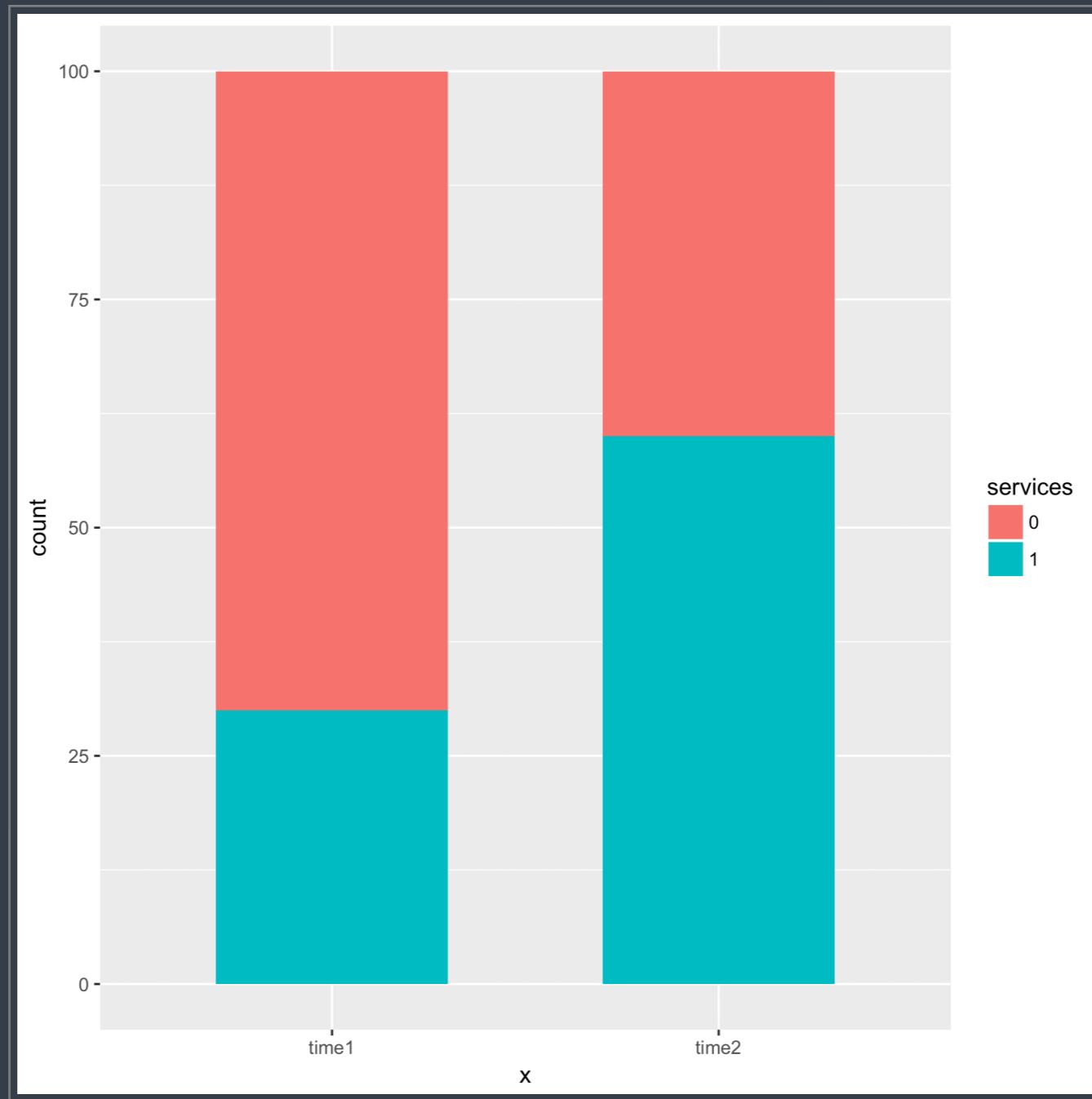
The screenshot shows the main interface of the RAWGraphs website. At the top, there is a navigation bar with links to "About", "Gallery", "Learning", "FAQs", "User survey", "Developer Guide", and a GitHub icon. Below the navigation bar, the text "Load your data" is displayed. There are three input methods: "Paste" (with a clipboard icon), "Upload a file" (with a folder icon), and "From URL" (with a cloud icon, which is highlighted in green). A text input field contains the placeholder "http://". Below the input fields, a note reads: "Enter a web address (URL) pointing to the data (e.g. a public Dropbox file, a public API, ...). Please, be sure the server is CORS-enabled." At the bottom of the main section, there is a dark footer bar containing project information: "RAW is an open source project by DensityDesign Lab and Calibro", developer names "Giorgio Caviglia, Michele Mauri, Giorgio Uboldi, Matteo Azzi", and a copyright notice "© 2013-2017 (Apache License 2.0)". To the right of the footer, there are social media links: an envelope icon for email ("hello@rawgraphs.io"), a Twitter icon for the Twitter handle ("rawgraphs"), a GitHub icon for the GitHub repository ("GitHub"), and a Google group icon for the Google group ("Google group").

Link: Little Miss Data Tutorial on RAWGraphs

“No code” easy access to high quality data visualizations

@littlemissdata

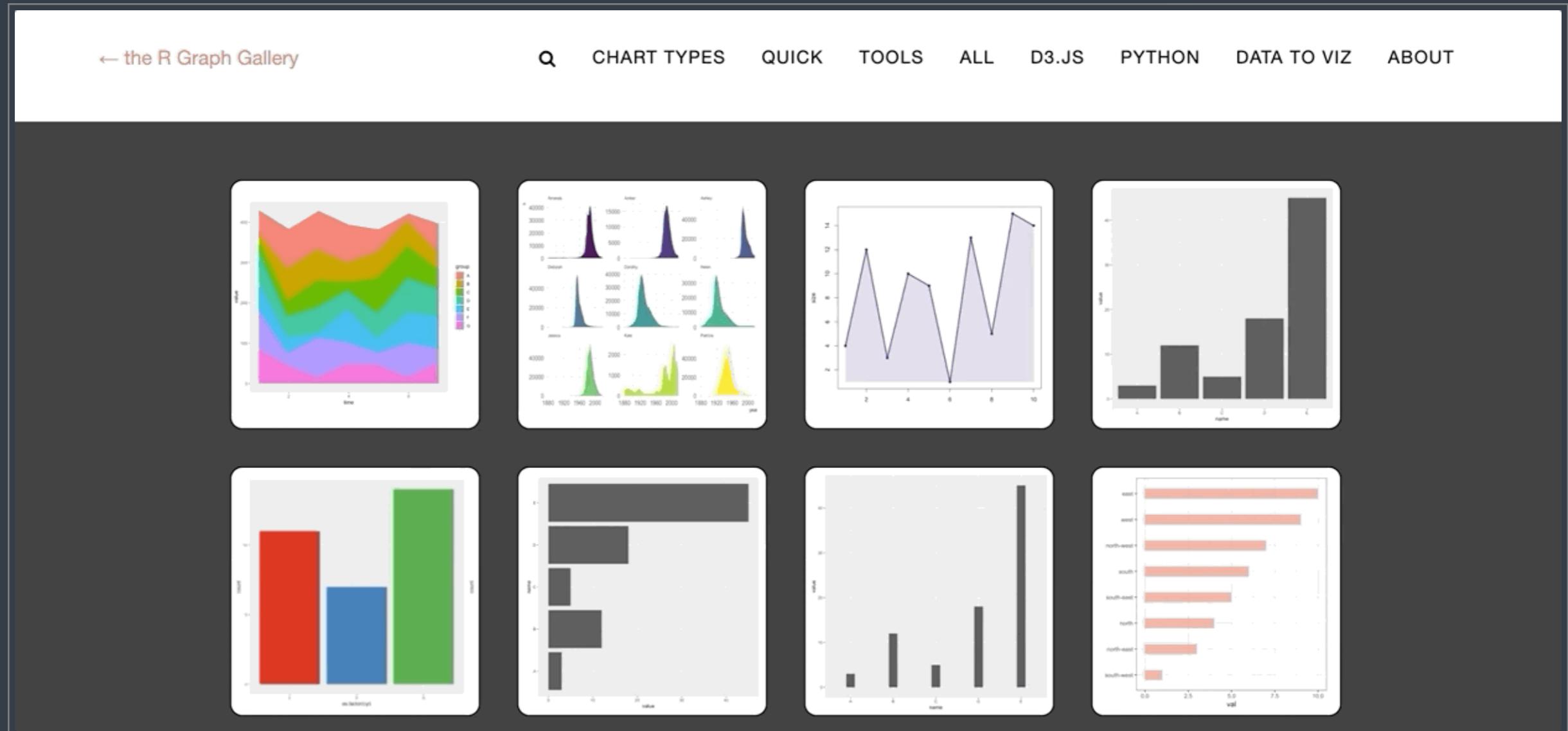
Take a Sad Graph and Make it Better



Understanding the impact a visually appealing graph can make

@littlemissdata

R Graph Gallery

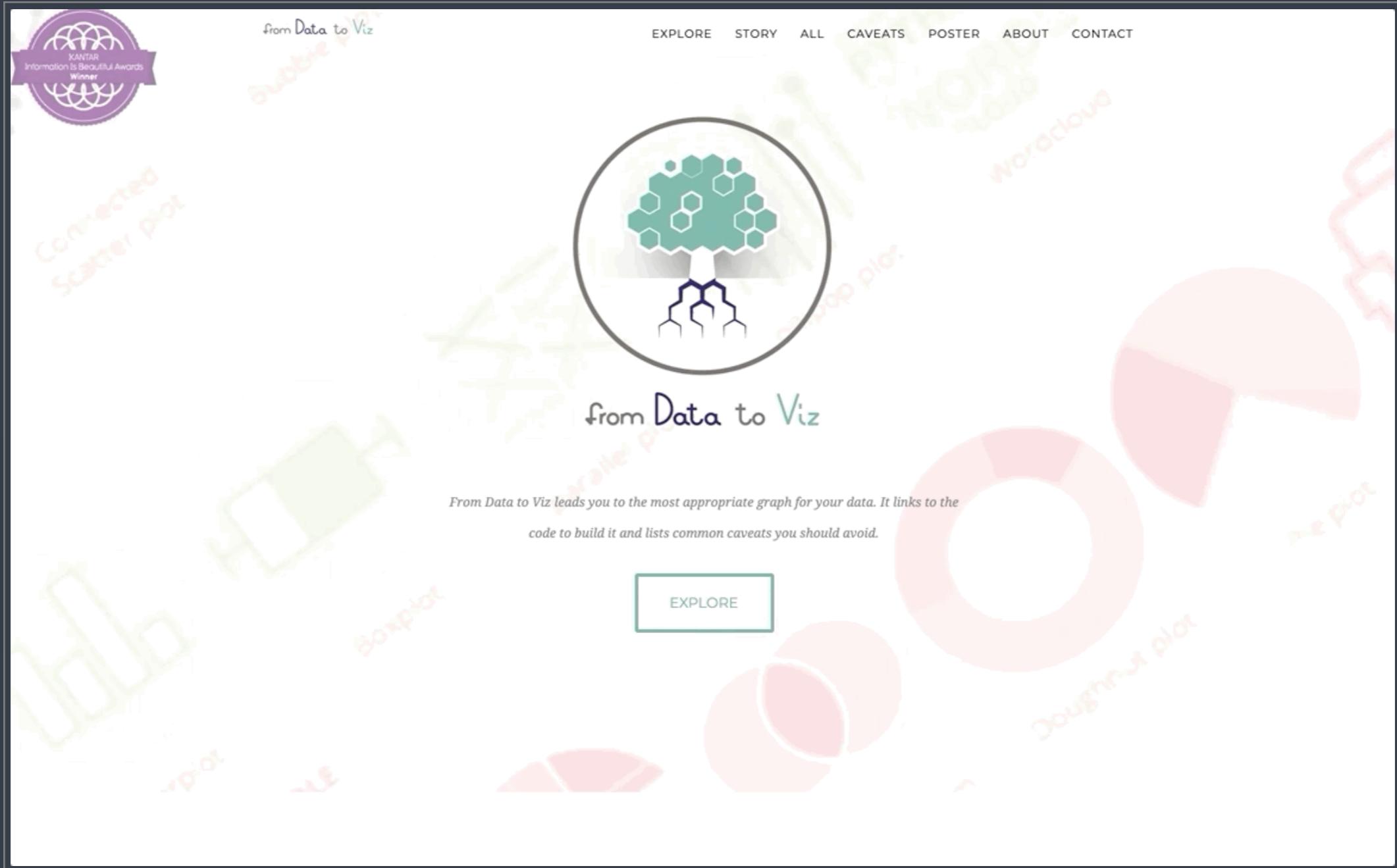


[Link: R Graph Gallery](#)

For your graphing inspiration

@littlemissdata

From Data to Viz



Link: [Data to Viz](#)

Understand graphing formats and jumpstart your graph creation

@littlemissdata

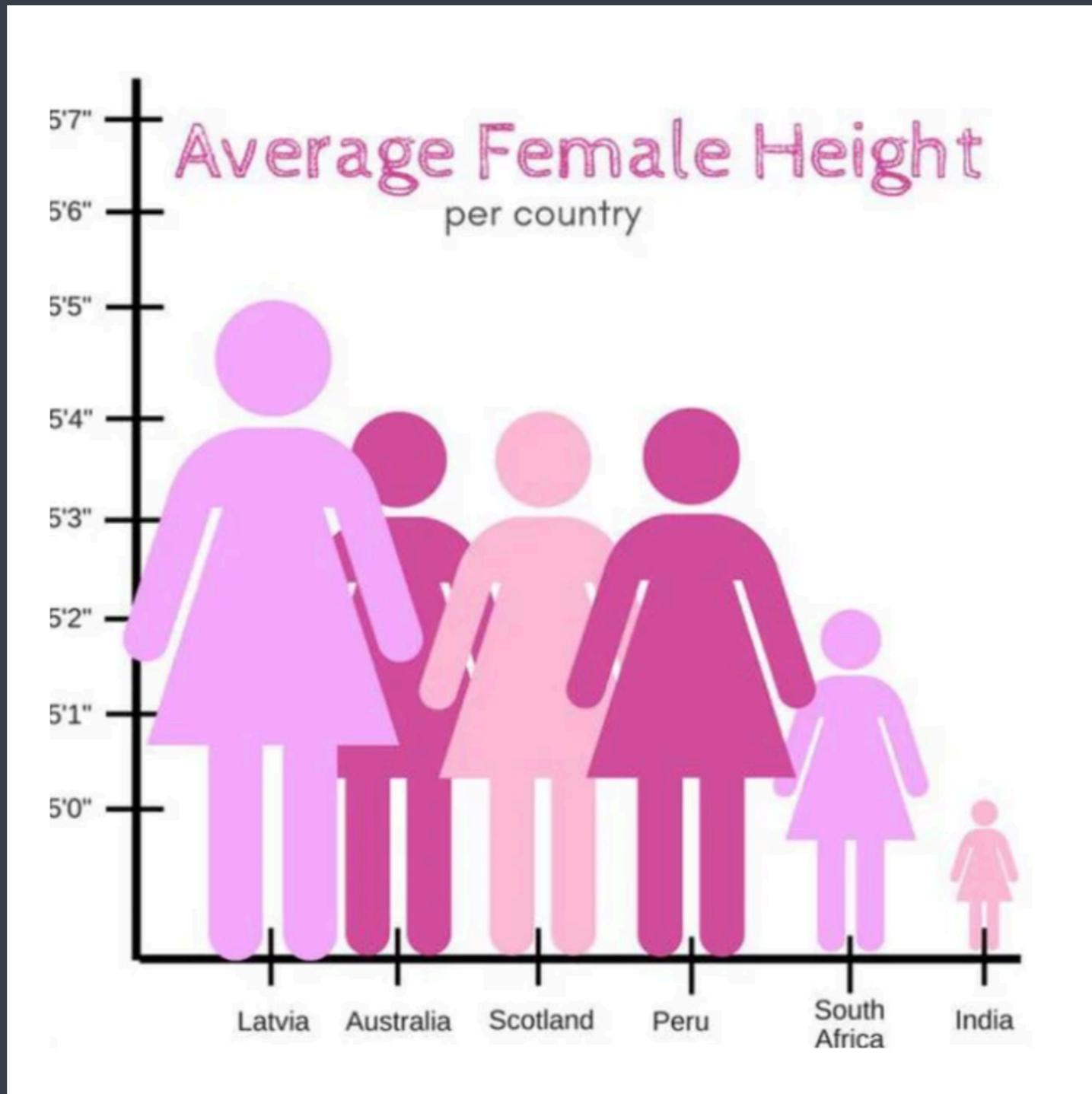
Information is Beautiful

The screenshot shows the homepage of the KANTAR Information is Beautiful Awards 2019. At the top, there are social media icons for Twitter, Facebook, and Pinterest, followed by 'Register' and 'Sign In' buttons. A search icon is also present. The main header features a purple circular logo with concentric lines and the text 'KANTAR Information is Beautiful Awards 2019'. Below the header, there are navigation links for 'About', 'News', 'Awards', 'Entry Showcase', and 'Sponsor'. The 'Awards' tab is selected, showing dropdown menus for years (All, 2019, 2018, 2017, 2016, 2015, 2014, 2013, 2012), categories (All, Winners, Silver, Gold, Shortlist, Longlist, Bronze, Rising Star, Impressive Individual, Best-Non-English-Language), and sub-categories (Outstanding Outfit, Student, Community, Most Beautiful). Below these are more detailed category filters: All, Arts, Entertainment & Culture, Humanitarian, Leisure, Games & Sport, Maps, Places & Spaces, News & Current Affairs, People, Language & Identity, Politics & Global, Visualization & Information Design, Science & Technology, and Unusual. Three award entries are displayed in cards at the bottom: 'The Last Stronghold' (A series of maps explain how President), 'Advantage distribution of school sectors' (2016), and 'Cartographers Of North Korea'.

[Link: Information is Beautiful](#)

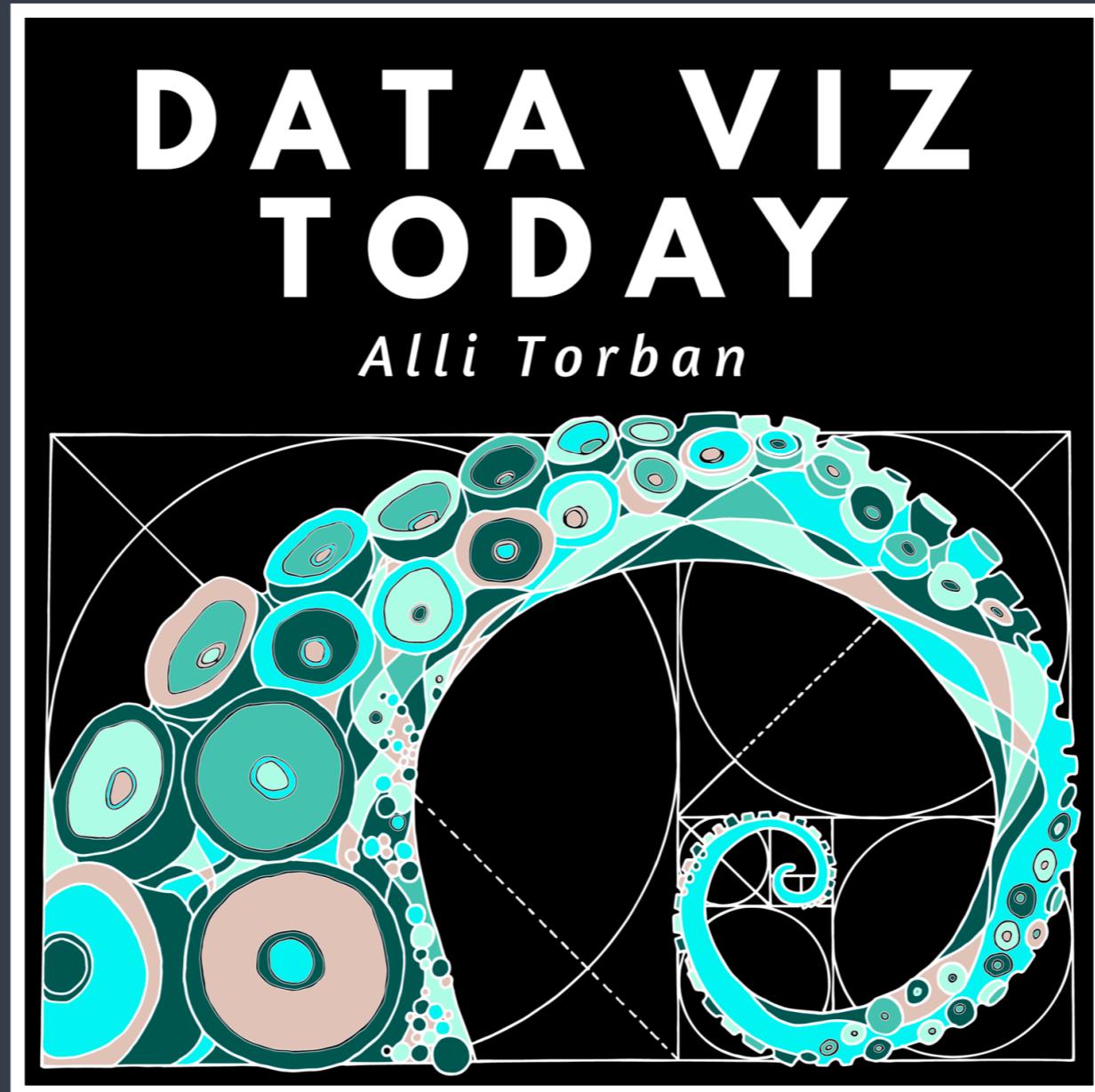
@littlemissdata

Data is Ugly



[Link: Data is Ugly Subreddit](#)

Data Viz Today



[Link: Data Viz Today](#)

Best DataViz Podcast!

@littlemissdata

So many books

Fundamentals of Data Visualization

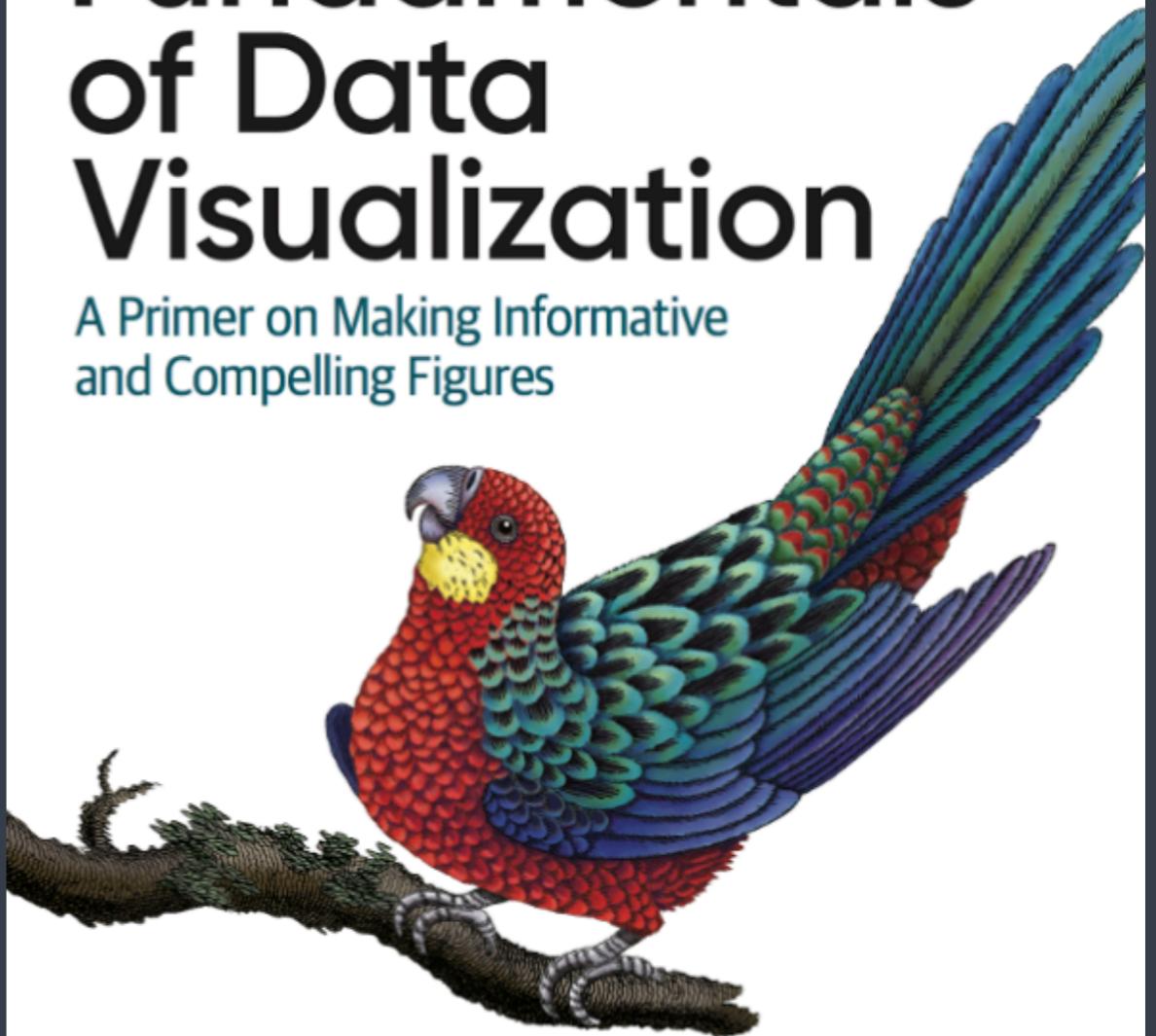
The Functional Art

Storytelling with Data

+ many more

Fundamentals of Data Visualization

A Primer on Making Informative
and Compelling Figures



Claus O. Wilke

Image Source :Fundamentals of Data Visualization

@littlemissdata

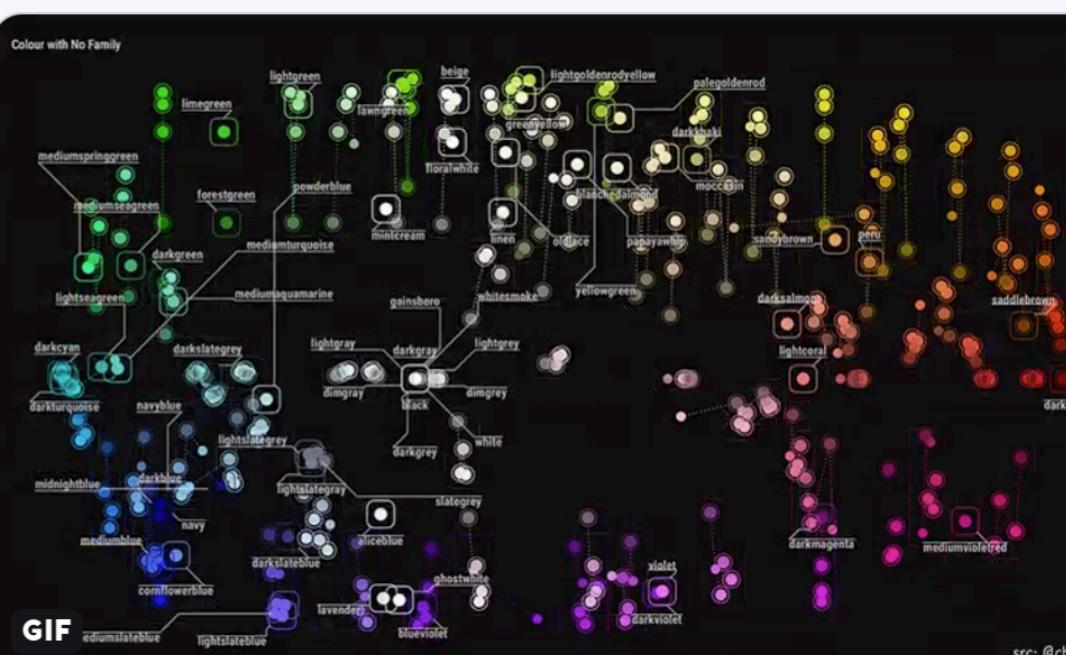
Social media

#dataviz on twitter and instagram

← Q #dataviz

Top Latest People Photos Videos

Mara Averick @dataandme · Sep 10
🌈 Fab use of {ggforce} for annotation!
"Built-in Colour Names in R" 🧑‍💻 @chisatini
buff.ly/2Nbm2kf #rstats #dataviz



src: @chisatini

GIF

3 26 142

Randy Olson @randal_olson · Sep 7
If you like those bar chart race #dataviz, this article shows how you can create one using matplotlib. #Python #programming #DataScience
towardsdatascience.com/bar-chart-race...

The most populous cities in the world from 1500 to 2018

@littlemissdata