Exercises

1. Find examples of data presentations that you consider to be particularly insightful and/or powerful. Discuss their strengths and weaknesses.
2. Find examples of data presentations that you consider to be particularly misleading and/or useless. Discuss their strengths and weaknesses.
3. How do you think new technologies (e.g. virtual or augmented reality, 3D-printing, wearable computing) will influence data presentations?
4. Consider the following datasets:
   - `GlobalCitiesPBI.csv`
   - `2016collisionsfinal.csv`
   - `polls_us_election_2016.csv`
   - `HR_2016_Census_simple.xlsx`

   a) Create a data dictionary for these datasets. Establish a list of variables that you think are crucial to understand them well. Justify your choices.
   b) Create (at least) 5 bivariate/univariate visualizations that can help you understand the datasets.
   c) Produce (at least) 3 "definitive" visualizations for each dataset. Use the principles discussed in class (including documentation, legends, annotations, Multiple I's, etc.). Emphasis should be placed on content AND on presentation (suggestions: consider creating a reasonably high number of charts using a random selection of a random number of variables in order to minimize the odds of missing out on useful information).
5. Repeat the previous question with any dataset of your liking.
6. Identify a scenario for which a dashboard could prove useful. Determine specific questions that the dashboard could help answer or insights that it could provide. Identify data sources and data elements that could be fed into your dashboard. Design a display (with pen and paper) with mock charts. What are the strengths and limitations of your dashboard? Is it functional? Elegant?
7. The remaining exercises use the Gapminder Tools (there is also an offline version).

   a) At what point in the data science workflow do you think that visualizations of this nature could be useful?
   b) What are the ways in which observations could be anomalous? Have you found any such anomalies? Do you have explanations

for them? In particular, consider the case of South Africa in 2012, which appears to be a clear outlier. Follow the path of the South African bubble from 1975 to 2020, in relation to the general pattern. Does the apartheid/income inequity explanation suggested in the text still make sense?

c) Pick 2+ "definitive" visualizations (methods, variables, etc.) other than the default configuration. What are some important insights?

d) How would you describe the insights of step 3 without resorting to visual vocabulary?

e) Can you think of ways in which the data of interest to you in your day-to-day activities could benefit from the same treatment? What situations could you explore in such a scenario? How would that help your team better understand the system under consideration?

8. Consider the Australian population figures, by state (in 1000s):

a) Graph the New South Wales (NSW) population with all defaults using `plot()`. Redo the graph by adding a title, a line to connect the points, and some colour.

b) Compare the population of New South Wales (NSW) and the Australian Capital Territory (ACT) by using the functions `plot()` and `lines()`, then add a legend to display your graph.

c) Use a bar chart to graph the population of Queensland (QLD), add an appropriate title to your graph, and display the years from 1917 to 2017 on the appropriate bars.

d) Create a light blue histogram for the population of South Australia (SA).

Year	NSW	Vic.	Qld	SA	WA	Tas.	NT	ACT	Aust.
1917	1904	1409	683	440	306	193	5	3	4941
1927	2402	1727	873	565	392	211	4	8	6182
1937	2693	1853	993	589	457	233	6	11	6836
1947	2985	2055	1106	646	502	257	11	17	7579
1957	3625	2656	1413	873	688	326	21	38	9640
1967	4295	3274	1700	1110	879	375	62	103	11799
1977	5002	3837	2130	1286	1204	415	104	214	14192
1987	5617	4210	2675	1393	1496	449	158	265	16264
1997	6274	4605	3401	1480	1798	474	187	310	18532
2007	6889	5205	4182	1585	2106	493	215	340	21017
2017	7861	6324	4928	1723	2580	521	246	410	24599

9. Create a useful dashboard out of the `Gapminder.csv` dataset.
10. Use the "Accounting Transactions" set-up of Chapter 13.

a) Create a measure to show the ratio of Minor to Major Capital.

b) Create a line chart using this measure; use "Effective Date" for the "X Axis".

c) Create a slicer that uses "Effective Date", and explore its effect on the chart's shape.

d) Add the slicer to some of the previous charts from the section and explore its effects on their shape.

11. Use the "Accounting Transactions" set-up of Chapter 13.

    a) Create a new bar chart.
    b) Use "Journal Voucher Type Code" as the Axis.
    c) Use "Item Amount" as the Value.
    d) Change the chart so that it is sorted using a "Custom Sort Order" column (ordering according to some custom order).

12. Use the "Accounting Transactions" set-up of Chapter 13.

    a) Create a chart that contains values from two or more of the tables with "Project Identifier" as an axis.
    b) Go into any one of the data files and add a new project on the bottom with a NEW Project Identifier that is not on the list.
    c) Save the Excel file and then update the Power BI data model. What happens?

13. Determine if the following are stories, as discussed in Chapters 7 and 8.

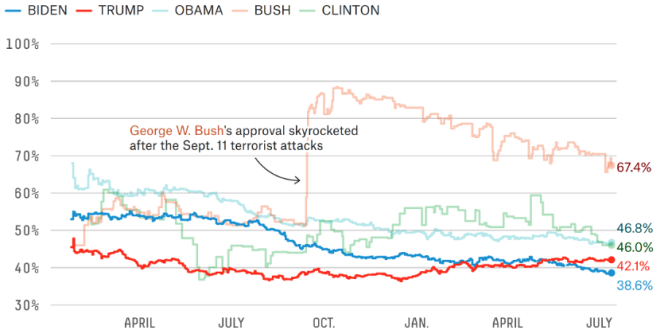
    a) Two identical infants lay in a cradle. "One you bore, the other is a Changeling. Choose wisely," the Fae's voice echoed from the shadow. "I'm taking both my children," the mother said defiantly.
    b) Solomon was required to decide which of two women was the mother of a baby, when each of them claimed parenthood. Both had recently given birth, but one child had died. Solomon announced that the child should be cut in two, so that each mother should have half. The real mother, unable to bear her son being killed, immediately offered it to the other woman, to save the child's life, whereas the other agreed to the proposal. The false mother was thus exposed, and Solomon returned the living child to its real mother.
    c) Yankel and Moishe ride a train across Poland, thinking about their brides – whom they're about to meet for the first time . Suddenly, Yankel stands up and says, "I'm not ready for marriage. I'm not getting married!" He grabs his suitcase and runs off the train at the next stop. Moishe watches him go. A day later he reaches his destination, where the two mothers of the prospective brides are shocked to discover there's only one groom on the train. "He's mine!" "Not on your life! He'll marry my daughter!" Moishe lets his prospective mothers-in-law argue over him for a while, then suggests they all go to the Rebbe for a ruling. "The solution is plain, per the wisdom of King Solomon. Cut the boy in two, and each of you take half." The first mother looks shocked. The second mother says, "Well, Rebbe, I didn't think that you had it in you. Yes! Cut him in half." The Rebbe points to the second mother and says, "That's the real mother-in-law!"
    d) For sale: baby shoes. Never worn.
    e) Doctors think that they may have improved the diagnosis of liver disease by 1%.
    f) Scientists claims cure for cancer.
    g) Spiritualist medium claims cure for cancer.
    h) Sens rally after blowing lead; beat Leafs to gain on Habs.

i) Macbeth and his wife
Want to become the royals
So they kill 'em all.

j) The following chart from FiveThirtyEight:

Biden could have the lowest midterm approval rating

FiveThirtyEight's historical presidential approval ratings for Biden and the four most recent presidents in their first 18 months in office, 1993-2022



The first data point for each president reflects when there was enough polling data to produce an average. All data is current as of July 13, 2022, at 5 p.m. Eastern.

FiveThirtyEight

k) The following news report, from the Associated Press:

Associated Press Feb 19, 2017

TORONTO -- The Ottawa Senators have the Atlantic Division lead in their sights.

Mark Stone had a goal and four assists, Derick Brassard scored twice in the third period and the Senators recovered after blowing a two-goal lead to beat the Toronto Maple Leafs 6-3 on Saturday night.

The Senators pulled within two points of Montreal for first place in the Atlantic Division with three games in hand.

"We like where we're at. We're in a good spot," Stone said. "But there's a little bit more that we want. Obviously, there's teams coming and we want to try and create separation, so the only way to do that is keep winning hockey games."

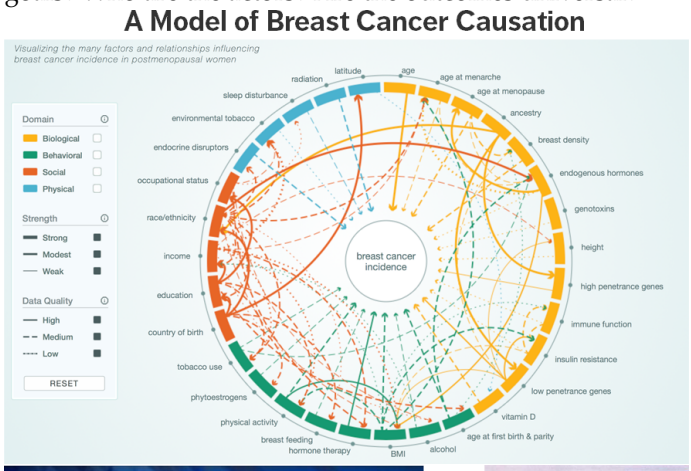
Ottawa led 2-0 after one period but trailed 3-2 in the third before getting a tying goal from Mike Hoffman and a power-play goal from Brassard. Stone and Brassard added empty-netters, and Chris Wideman and Ryan Dzingel also scored for the Senators.

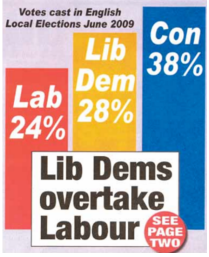
14. Find examples of stories that are about individuals; about organizations; about cultures/societies.
15. Find examples of stories that are used to persuade; educate; entertain.
16. Find examples of data stories that are about individuals; about organizations; about cultures/societies.
17. Find examples of data stories that are used to persuade; educate; entertain.
18. In your organization, who makes up the audience? Is there only one audience? What are the storytelling goals? Is the storytelling context clear? Constant? Universal?

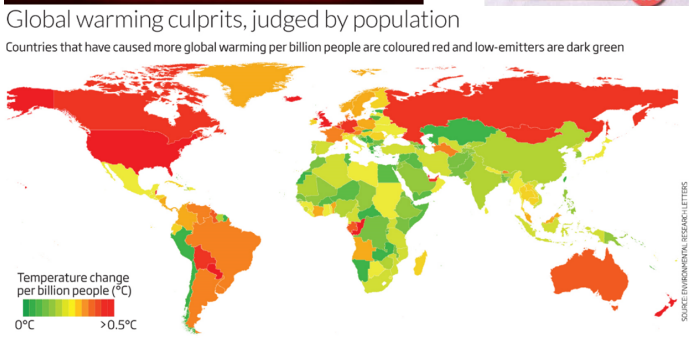
19. **The Death of the Author:** Isaac Asimov once sat anonymously in a class where the topic of discussion was one of his stories. He introduced himself afterwards to the teacher, saying that he had found his interpretation of the story interesting, but it wasn't really what he had meant at all. The teacher's response was "Just because you wrote it, what makes you think you have the slightest idea what it's about?" How could this come into play when telling stories with data?
20. Guess the story:

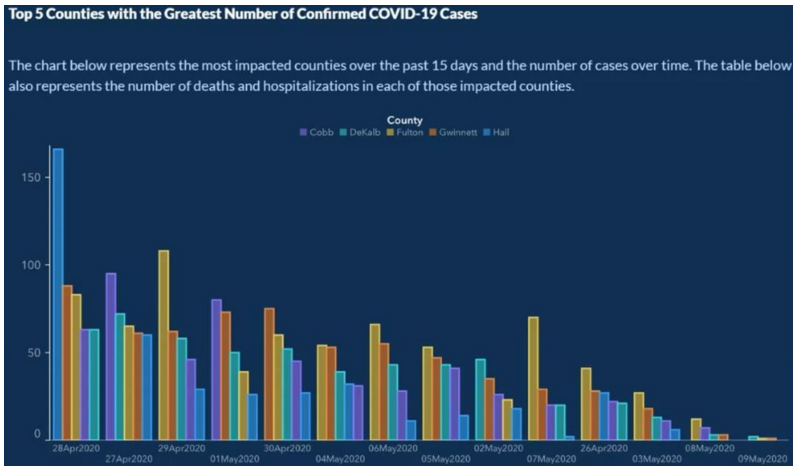
    a. A group of friends spends 9 hours returning jewellery.
    b. A talking frog convinces a son to kill his father.
    c. A young woman talks to furniture and marries her kidnapper.
    d. A depressed, widowed father teams up with an unwell woman to find his disabled son.

21. In the following charts, who is the intended audience? What are the goals? Who are the actors? Are the outcomes universal?





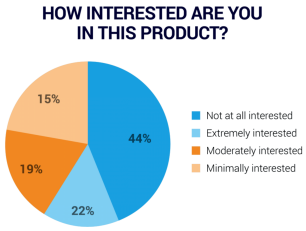




22. Select a few stories of your liking (from all genres, media, formats).
    a) Build the corresponding story spines.
    b) Identify some of their storytelling tropes.
    c) Build the corresponding story molecules.
23. How would you fix the following stories, using the concepts presented in the book?
    a) One day I woke up. I went outside and there was weather. I came back inside and did things.
    b) Once upon a time there was a hero. She left on an adventure and fought a giant dragon. Then she fought another giant dragon. Then she fought another giant dragon. Then she fought another dragon. Then she went home, victorious.
    c) The rain was bucketing down in immense impenetrable sheets of torrential freezing water. “Oh woe is me!”, the brave but sobbing child screamed. How shall I ever manage to dash as quickly as humanly possible into the terrible wreck of an ancient schoobus without completely mangling my magnificent hair. Indeed, the desolate child failed. She was the laughingstock of all of the pompous fools on the school bus.
24. Start by taking a few moments to think of a very short story you can tell (context, events, outcome).
    a) Pair up and tell the story (one of you will be the storyteller, the other person is the audience).
    b) Pair up again and tell the story (the storytellers should become the audience for someone who was an audience of another group in part a.)
    c) Pair up a third time and repeat parts a. and b., with different teammates if possible.
    d) How did the story change the second time you told it? What made you change it?
25. Think of a work story. Create a sketch that could illustrate this work story. What visual storytelling choices and combinations would you consider using? Would accessibility considerations change the way in which the story is presented to the audience?

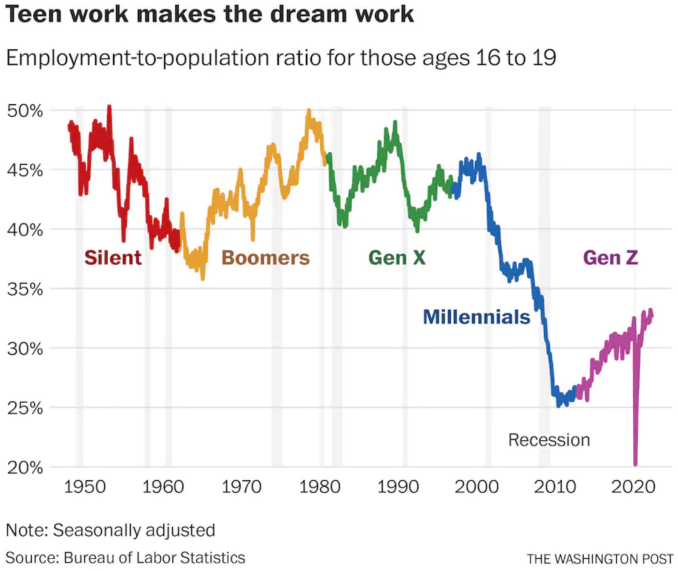
26. Re-cast the data stories presented in in Chapters 7 and 8 (or other data stories, as required) using different visual storytelling choices and combinations.
27. Find good candidates for the type of chart that could be used with the following:
    a) A dataset with information about financial transactions throughout the year.
    b) A dataset with the results of a survey of employee work satisfaction.
    c) A dataset showing levels of regulatory compliance with a set of government regulations (e.g., regulations relating to environmental reporting).
    d) Any other dataset of your choosing.
28. Identify instances of scoping, exploration, explanation, persuasion among the dashboards and charts presented in Chapters 7 and 8 (or other data stories, as required).
    a) What are the underlying dataset structure and its limitations?
    b) What analytical and data focus choices are at play?
    c) Are the charts falsifiable?
29. What is missing to turn this evolved data chart into a data story?

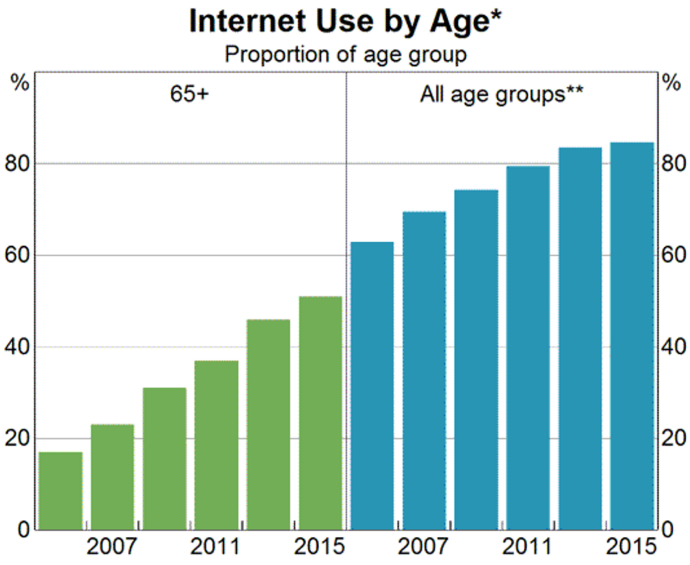
BEFORE



AFTER

30. Evolve the following charts into data stories. Focus on the message and how to avoid misleading the audience. Use data storytelling tropes as needed.

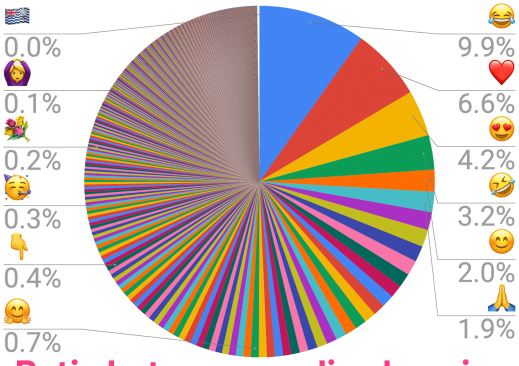




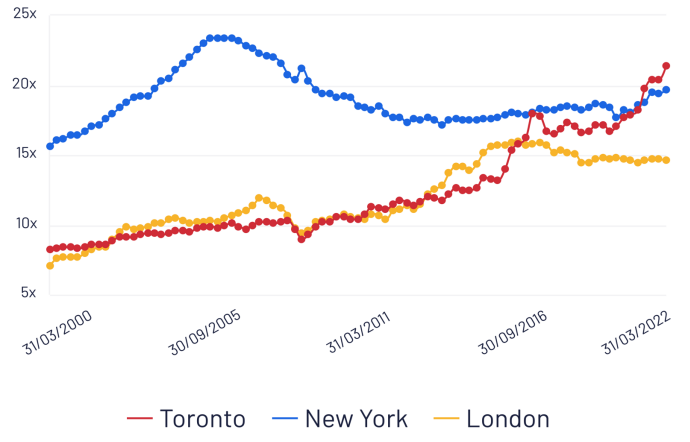
* Accessed the internet at any site over the year to June

** Persons aged 15 years and over

Source: ABS



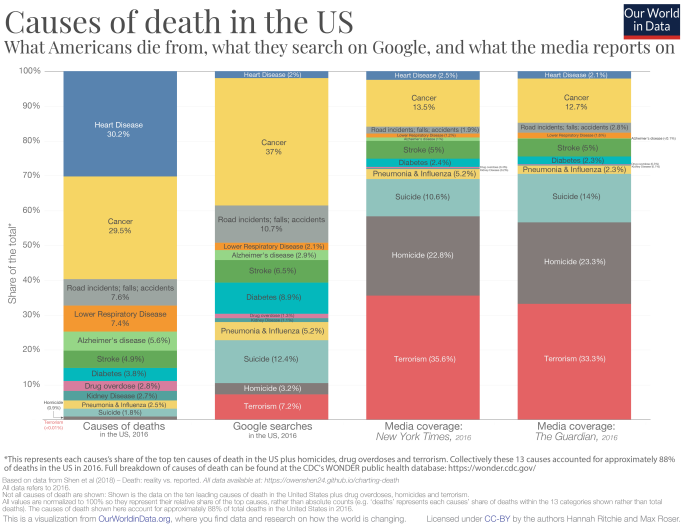
Ratio between median housing price and median annual salary



Methodology: Median housing price in each city divided by median pre-tax annual salary
Source: Bloomberg

More charts at (link in bio): genuineimpact.substack.com

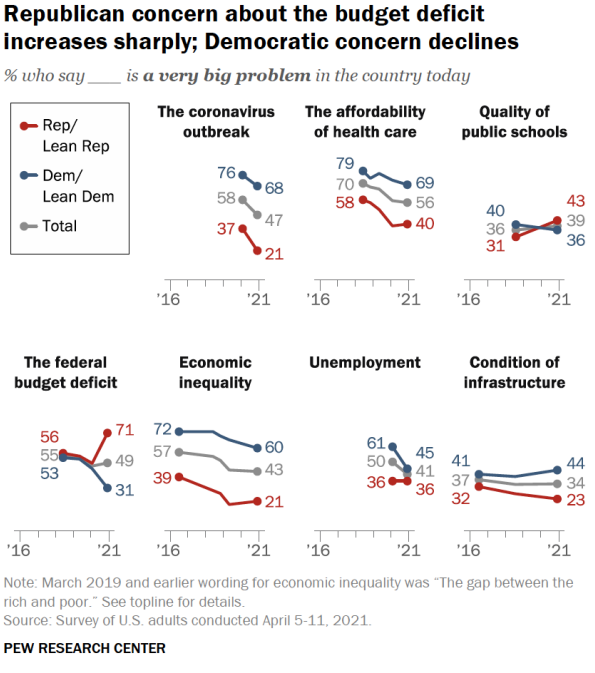
Created by genuine impact

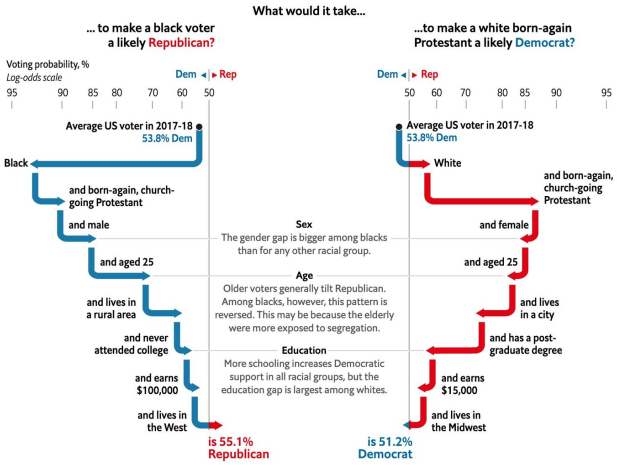


31. Consider a data question of interest to you personally, your organization, or your society.

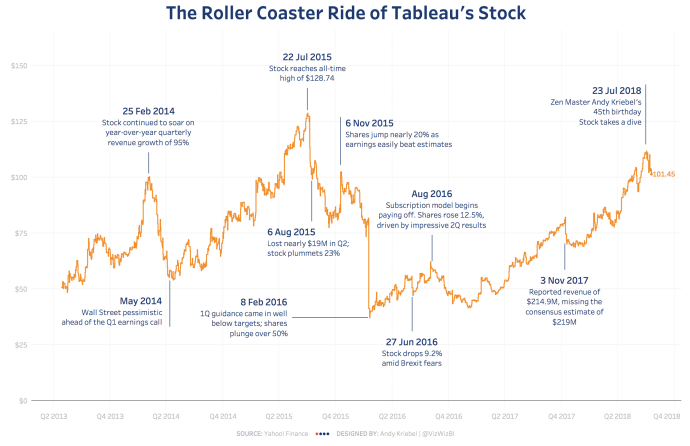
    a) Identify the target audience and the goals for your dashboard.
    b) Do you require an exploration dashboard? A storybook? A situational awareness dashboard? Some combination of the above?
    c) Identify the presentation requirements for your dashboard.
    d) Create a storyboard for your dashboard.
    e) What type of narrative and logic do you think would best serve your needs?

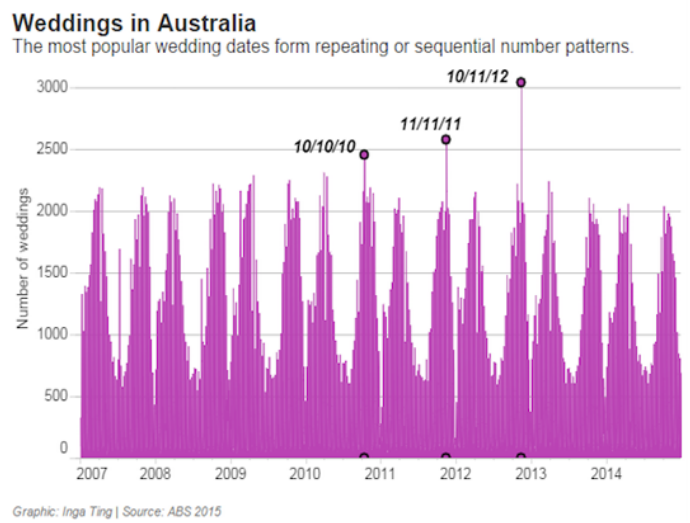
32. Comment on the aesthetics of the following charts, according to the Gestalt principles, their use of colours, clutter (or lack thereof), size and positioning, etc. Provide suggestions for improvements.





Sources: YouGov; *The Economist*





Graphic: Inga Ting | Source: ABS 2015

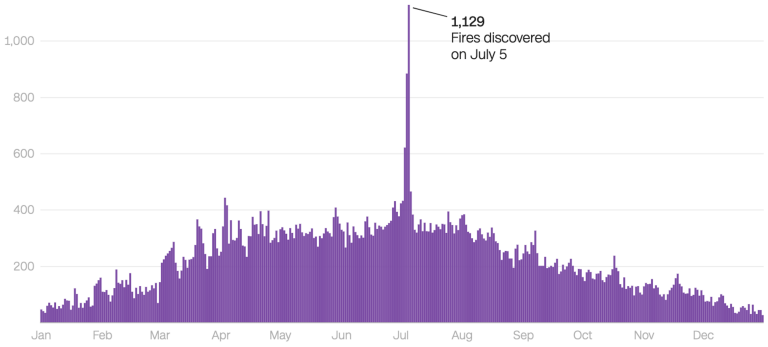
33. Consider the following examples of charts found in the wild. Are they examples of exploration, storytelling, situational awareness? Are they data stories? If not, how would you turn them into stories? If so, are they good stories? Bad ones? Ugly ones? If they are not good stories,

how would you improve them?

Wildfires spike around July 4 holiday

Human-caused wildfires in the United States jump around Independence Day.

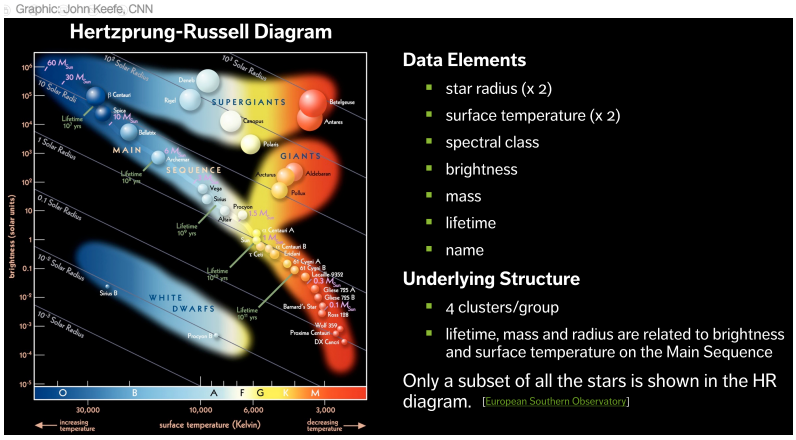
Total wildfires discovered each day of the year since 2014



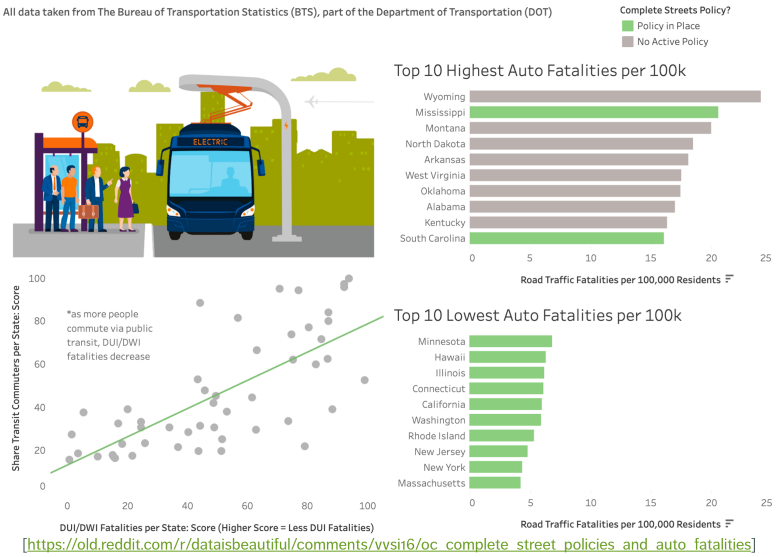
Human-caused fires, excluding prescribed fires. 2022 fires included through June 30. All incident times Eastern.

Sources: CNN analysis of data from the National Interagency Fire Center

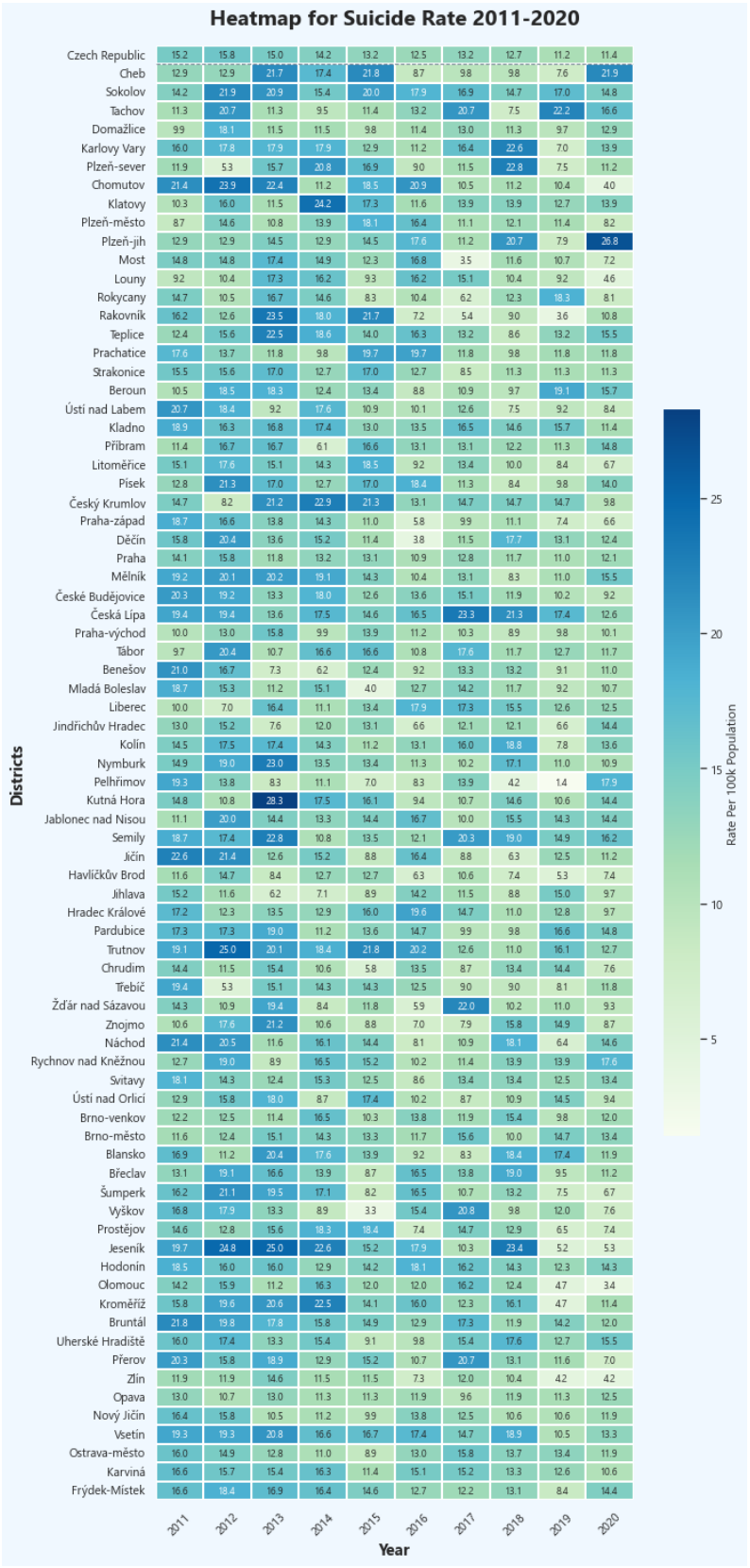
Graphic: John Keefe, CNN

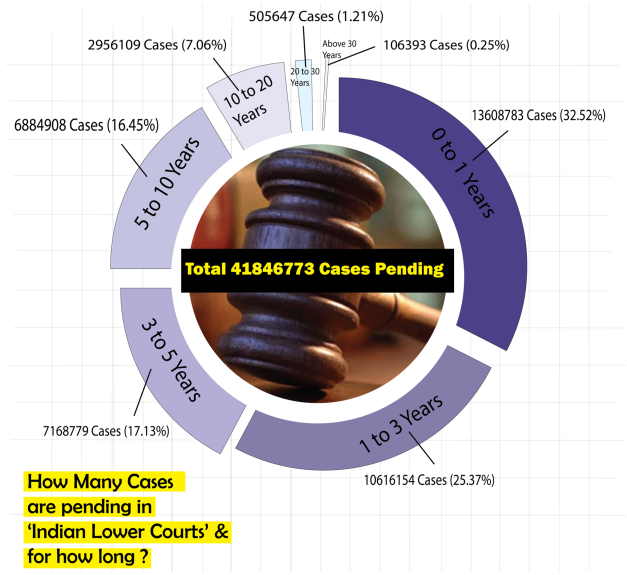


Public Transit and Complete Streets — How Do They Relate to Safety?

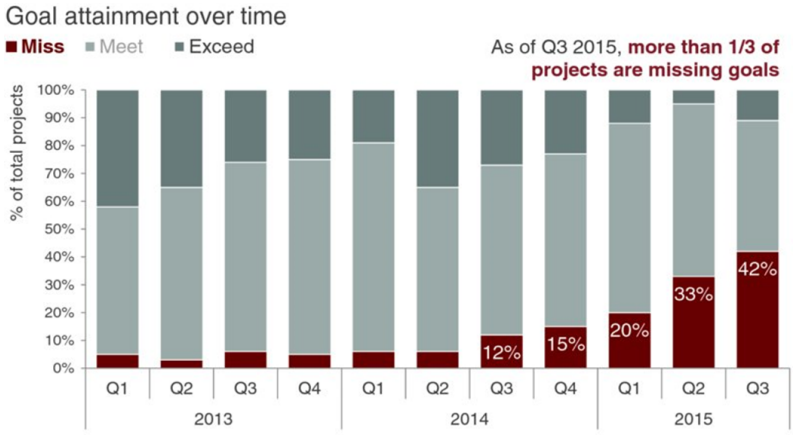


[https://old.reddit.com/r/dataisbeautiful/comments/vvsit6/oc_complete_street_policies_and_auto_fatalities]

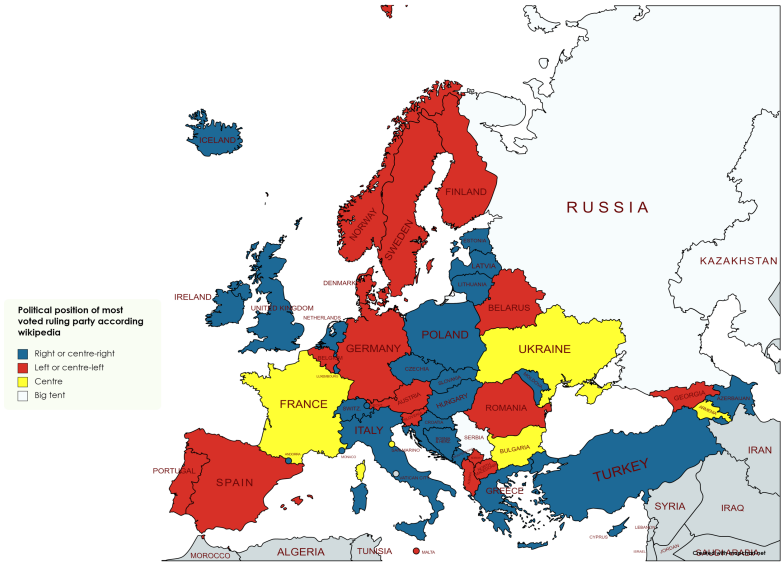

Heatmap for Suicide Rate 2011-2020
Districts
Year
Rate Per 100k Population
2011 2012 2013 2014 2015 2016 2017 2018 2019 2020
Czech Republic 15.2 15.8 15.0 14.2 13.2 12.5 13.2 12.7 11.2 11.4
Cheb 12.9 12.9 21.7 17.4 21.8 8.7 9.8 9.8 7.6 21.9
Sokolov 14.2 21.9 20.9 15.4 20.0 17.9 16.9 14.7 17.0 14.8
Tachov 11.3 20.7 11.3 9.5 11.4 13.2 20.7 7.5 22.2 16.6
Domažlice 9.9 18.1 11.5 11.5 9.8 11.4 13.0 11.3 9.7 12.9
Karlovy Vary 16.0 17.8 17.9 17.9 12.9 11.2 16.4 22.6 7.0 13.9
Plzeň-sever 11.9 5.3 15.7 20.8 16.9 9.0 11.5 22.8 7.5 11.2
Chomutov 21.4 23.9 22.4 11.2 18.5 20.9 10.5 11.2 10.4 4.0
Klatovy 10.3 16.0 11.5 24.2 17.3 11.6 13.9 13.9 12.7 13.9
Plzeň-město 8.7 14.6 10.8 13.9 18.1 16.4 11.1 12.1 11.4 8.2
Plzeň-jih 12.9 12.9 14.5 12.9 14.5 17.6 11.2 20.7 7.9 26.8
Most 14.8 14.8 17.4 14.9 12.3 16.8 3.5 11.6 10.7 7.2
Louny 9.2 10.4 17.3 16.2 9.3 16.2 15.1 10.4 9.2 4.6
Rokycany 14.7 10.5 16.7 14.6 8.3 10.4 6.2 12.3 18.3 8.1
Rakovník 16.2 12.6 23.5 18.0 21.7 7.2 5.4 9.0 3.6 10.8
Teplice 12.4 15.6 22.5 18.6 14.0 16.3 13.2 8.6 13.2 15.5
Prachatice 17.6 13.7 11.8 9.8 19.7 19.7 11.8 9.8 11.8 11.8
Strakonice 15.5 15.6 17.0 12.7 17.0 12.7 8.5 11.3 11.3 11.3
Beroun 10.5 18.5 18.3 12.4 13.4 8.8 10.9 9.7 19.1 15.7
Ústí nad Labem 20.7 18.4 9.2 17.6 10.9 10.1 12.6 7.5 9.2 8.4
Kladno 18.9 16.3 16.8 17.4 13.0 13.5 16.5 14.6 15.7 11.4
Příbram 11.4 16.7 16.7 6.1 16.6 13.1 13.1 12.2 11.3 14.8
Litoměřice 15.1 17.6 15.1 14.3 18.5 9.2 13.4 10.0 8.4 6.7
Písek 12.8 21.3 17.0 12.7 17.0 18.4 11.3 8.4 9.8 14.0
Český Krumlov 14.7 8.2 21.2 22.9 21.3 13.1 14.7 14.7 14.7 9.8
Praha-západ 18.7 16.6 13.8 14.3 11.0 5.8 9.9 11.1 7.4 6.6
Děčín 15.8 20.4 13.6 15.2 11.4 3.8 11.5 17.7 13.1 12.4
Praha 14.1 15.8 11.8 13.2 13.1 10.9 12.8 11.7 11.0 12.1
Mělník 19.2 20.1 20.2 19.1 14.3 10.4 13.1 8.3 11.0 15.5
České Budějovice 20.3 19.2 13.3 18.0 12.6 13.6 15.1 11.9 10.2 9.2
Česká Lípa 19.4 19.4 13.6 17.5 14.6 16.5 23.3 21.3 17.4 12.6
Praha-východ 10.0 13.0 15.8 9.9 13.9 11.2 10.3 8.9 9.8 10.1
Tábor 9.7 20.4 10.7 16.6 16.6 10.8 17.6 11.7 12.7 11.7
Benešov 21.0 16.7 7.3 6.2 12.4 9.2 13.3 13.2 9.1 11.0
Mladá Boleslav 18.7 15.3 11.2 15.1 4.0 12.7 14.2 11.7 9.2 10.7
Liberec 10.0 7.0 16.4 11.1 13.4 17.9 17.3 15.5 12.6 12.5
Jindřichův Hradec 13.0 15.2 7.6 12.0 13.1 6.6 12.1 12.1 6.6 14.4
Kolín 14.5 17.5 17.4 14.3 11.2 13.1 16.0 18.8 7.8 13.6
Nymburk 14.9 19.0 23.0 13.5 13.4 11.3 10.2 17.1 11.0 10.9
Pelhřimov 19.3 13.8 8.3 11.1 7.0 8.3 13.9 4.2 1.4 17.9
Kutná Hora 14.8 10.8 28.3 17.5 16.1 9.4 10.7 14.6 10.6 14.4
Jablonec nad Nisou 11.1 20.0 14.4 13.3 14.4 16.7 10.0 15.5 14.3 14.4
Semily 18.7 17.4 22.8 10.8 13.5 12.1 20.3 19.0 14.9 16.2
Jičín 22.6 21.4 12.6 15.2 8.8 16.4 8.8 6.3 12.5 11.2
Havlíčkův Brod 11.6 14.7 8.4 12.7 12.7 6.3 10.6 7.4 5.3 7.4
Jihlava 15.2 11.6 6.2 7.1 8.9 14.2 11.5 8.8 15.0 9.7
Hradec Králové 17.2 12.3 13.5 12.9 16.0 19.6 14.7 11.0 12.8 9.7
Pardubice 17.3 17.3 19.0 11.2 13.6 14.7 9.9 9.8 16.6 14.8
Trutnov 19.1 25.0 20.1 18.4 21.8 20.2 12.6 11.0 16.1 12.7
Chrudim 14.4 11.5 15.4 10.6 5.8 13.5 8.7 13.4 14.4 7.6
Třebíč 19.4 5.3 15.1 14.3 14.3 12.5 9.0 9.0 8.1 11.8
Žďár nad Sázavou 14.3 10.9 19.4 8.4 11.8 5.9 22.0 10.2 11.0 9.3
Znojmo 10.6 17.6 21.2 10.6 8.8 7.0 7.9 15.8 14.9 8.7
Náchod 21.4 20.5 11.6 16.1 14.4 8.1 10.9 18.1 6.4 14.6
Rychnov nad Kněžnou 12.7 19.0 8.9 16.5 15.2 10.2 11.4 13.9 13.9 17.6
Svitavy 18.1 14.3 12.4 15.3 12.5 8.6 13.4 13.4 12.5 13.4
Ústí nad Orlicí 12.9 15.8 18.0 8.7 17.4 10.2 8.7 10.9 14.5 9.4
Brno-venkov 12.2 12.5 11.4 16.5 10.3 13.8 11.9 15.4 9.8 12.0
Brno-město 11.6 12.4 15.1 14.3 13.3 11.7 15.6 10.0 14.7 13.4
Blansko 16.9 11.2 20.4 17.6 13.9 9.2 8.3 18.4 17.4 11.9
Břeclav 13.1 19.1 16.6 13.9 8.7 16.5 13.8 19.0 9.5 11.2
Šumperk 16.2 21.1 19.5 17.1 8.2 16.5 10.7 13.2 7.5 6.7
Vyškov 16.8 17.9 13.3 8.9 3.3 15.4 20.8 9.8 12.0 7.6
Prostějov 14.6 12.8 15.6 18.3 18.4 7.4 14.7 12.9 6.5 7.4
Jeseník 19.7 24.8 25.0 22.6 15.2 17.9 10.3 23.4 5.2 5.3
Hodonín 18.5 16.0 16.0 12.9 14.2 18.1 16.2 14.3 12.3 14.3
Olomouc 14.2 15.9 11.2 16.3 12.0 12.0 16.2 12.4 4.7 3.4
Kroměříž 15.8 19.6 20.6 22.5 14.1 16.0 12.3 16.1 4.7 11.4
Bruntál 21.8 19.8 17.8 15.8 14.9 12.9 17.3 11.9 14.2 12.0
Uherské Hradiště 16.0 17.4 13.3 15.4 9.1 9.8 15.4 17.6 12.7 15.5
Přerov 20.3 15.8 18.9 12.9 15.2 10.7 20.7 13.1 11.6 7.0
Zlín 11.9 11.9 14.6 11.5 11.5 7.3 12.0 10.4 4.2 4.2
Opava 13.0 10.7 13.0 11.3 11.3 11.9 9.6 11.9 11.3 12.5
Nový Jičín 16.4 15.8 10.5 11.2 9.9 13.8 12.5 10.6 10.6 11.9
Vsetín 19.3 19.3 20.8 16.6 16.7 17.4 14.7 18.9 10.5 13.3
Ostrava-město 16.0 14.9 12.8 11.0 8.9 13.0 15.8 13.7 13.4 11.9
Karviná 16.6 15.7 15.4 16.3 11.4 15.1 15.2 13.3 12.6 10.6
Frýdek-Místek 16.6 18.4 16.9 16.4 14.6 12.7 12.2 13.1 8.4 14.4
5
10
15
20
25



[https://old.reddit.com/r/dataisbeautiful/comments/vuma6b/how_many_court_cases_are_pending_in_lower_courts]

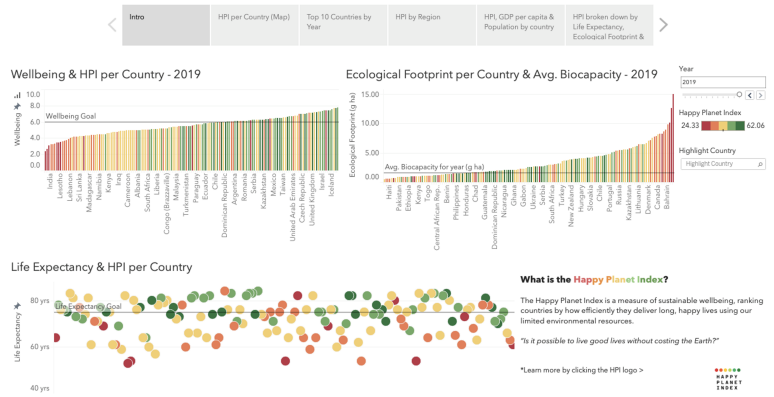


Data source: XYZ Dashboard; the total number of projects has increased over time from 230 in early 2013 to nearly 270 in Q3 2015.



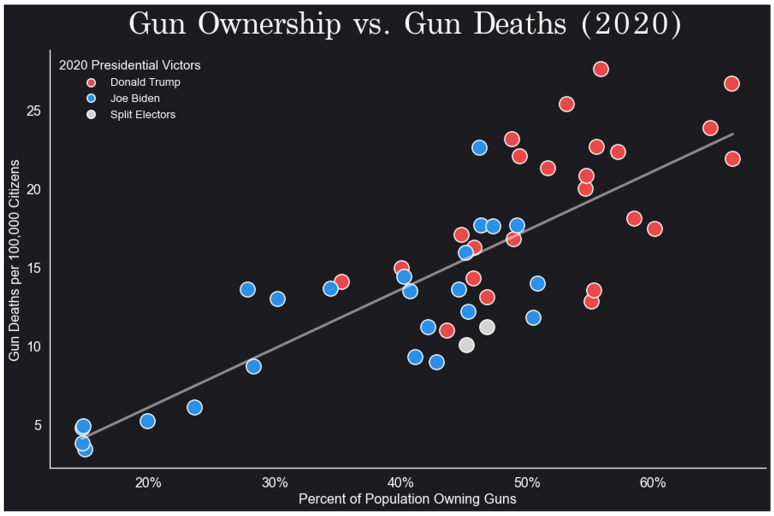
[https://old.reddit.com/r/dataisbeautiful/comments/vu8rm8/oc_political_position_of_most_voted_ruling]

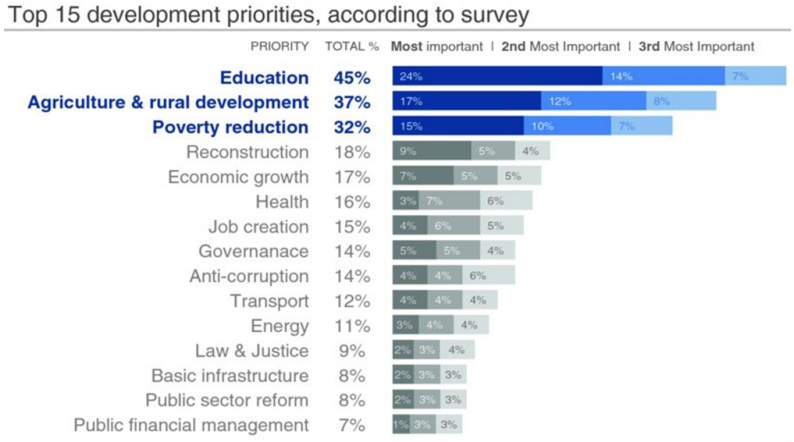
Happy Planet Index Vizualisation Story



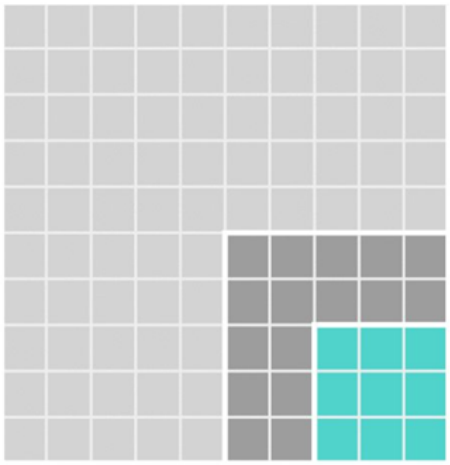
[https://public.tableau.com/app/profile/antonio2517/viz/HappyPlanetIndexVizualizationStory/HappyPlanetIndexVisualizationStory]

[https://old.reddit.com/r/dataisbeautiful/comments/vuefii/oc_gun_ownership_vs_gun_deaths_colorcoded_by_2020]





Interview breakdown

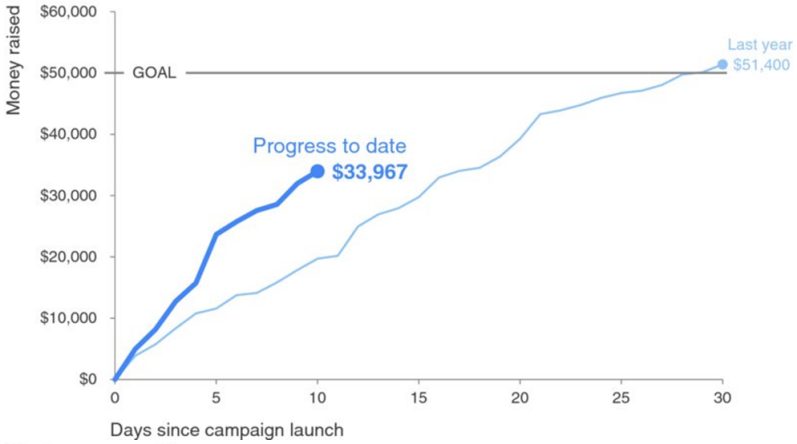


Out of every **100 phone screens**...

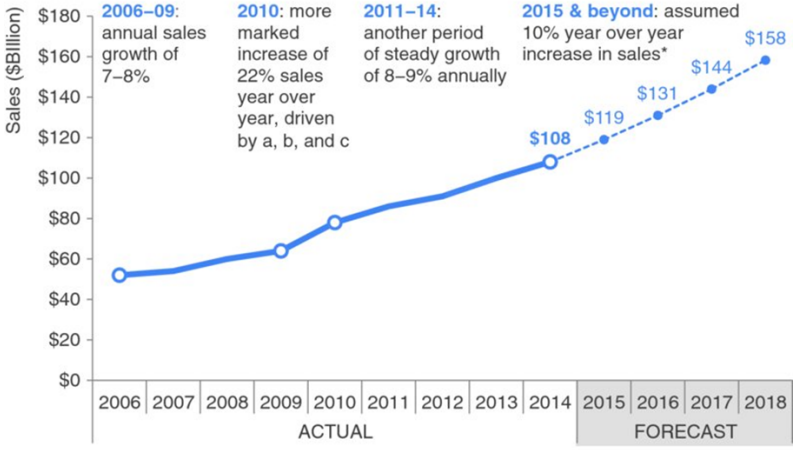
we bring **25 candidates onsite** for interviews...

and **extend 9 offers.**

Annual giving campaign progress

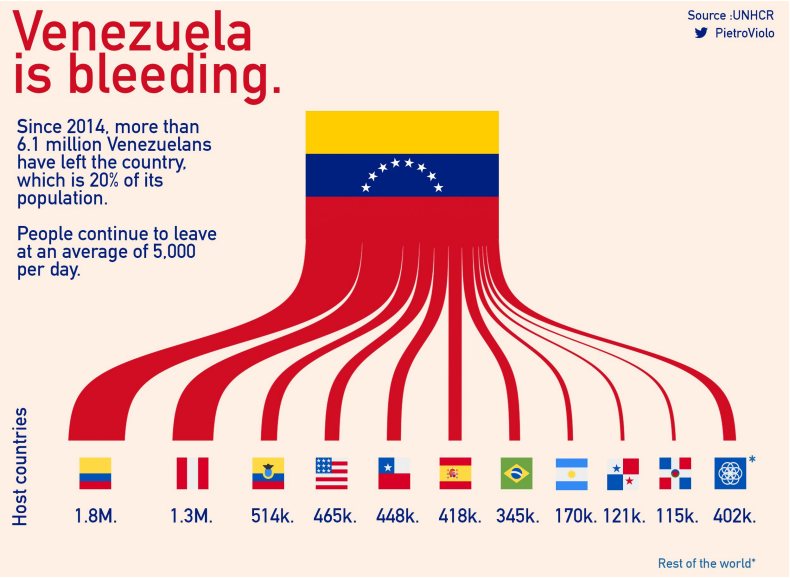


Sales over time

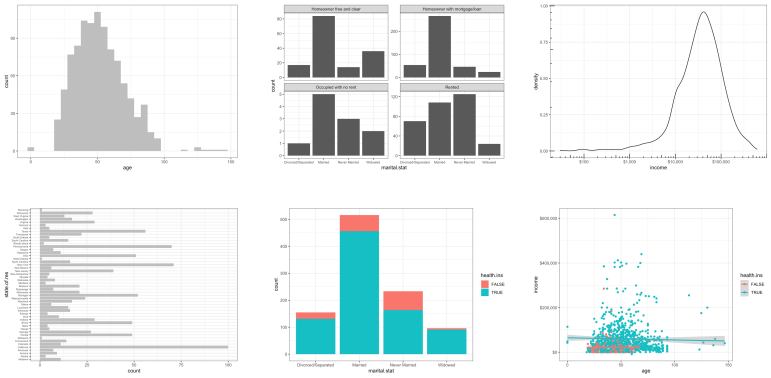


Data source: Sales Dashboard; annual figures are as of 12/31 of the given year.

*Use this footnote to explain what is driving the 10% annual growth forecast assumption.



34. Identify uses of the Gestalt principles in the charts presented in the previous question.
35. Deconstruct the charts introduced in the book in terms of the grammar of graphics. What do some of the most effective charts have in common? What about the least effective ones? Does that suggest a strategy?
36. Create some simple ggplot2 visualizations with data available in R. The emphasis should on becoming familiar with various geometries, their aesthetics, and the use of facets. You may use the examples found in the text as the basis of your work.[1]
37. The `custdata.tsv` file is derived from U.S. Census PUMS data. This synthetic dataset contains customer information for individuals whose health insurance status is known. Recreate the following ggplot2 charts (ignore missing values in this example).



38. Recreate the images of Section 2.1 using `ggplot2` and the `algae.blooms` dataset.

1: The available built-in datasets are obtained by running `data()` at the R command prompt.