

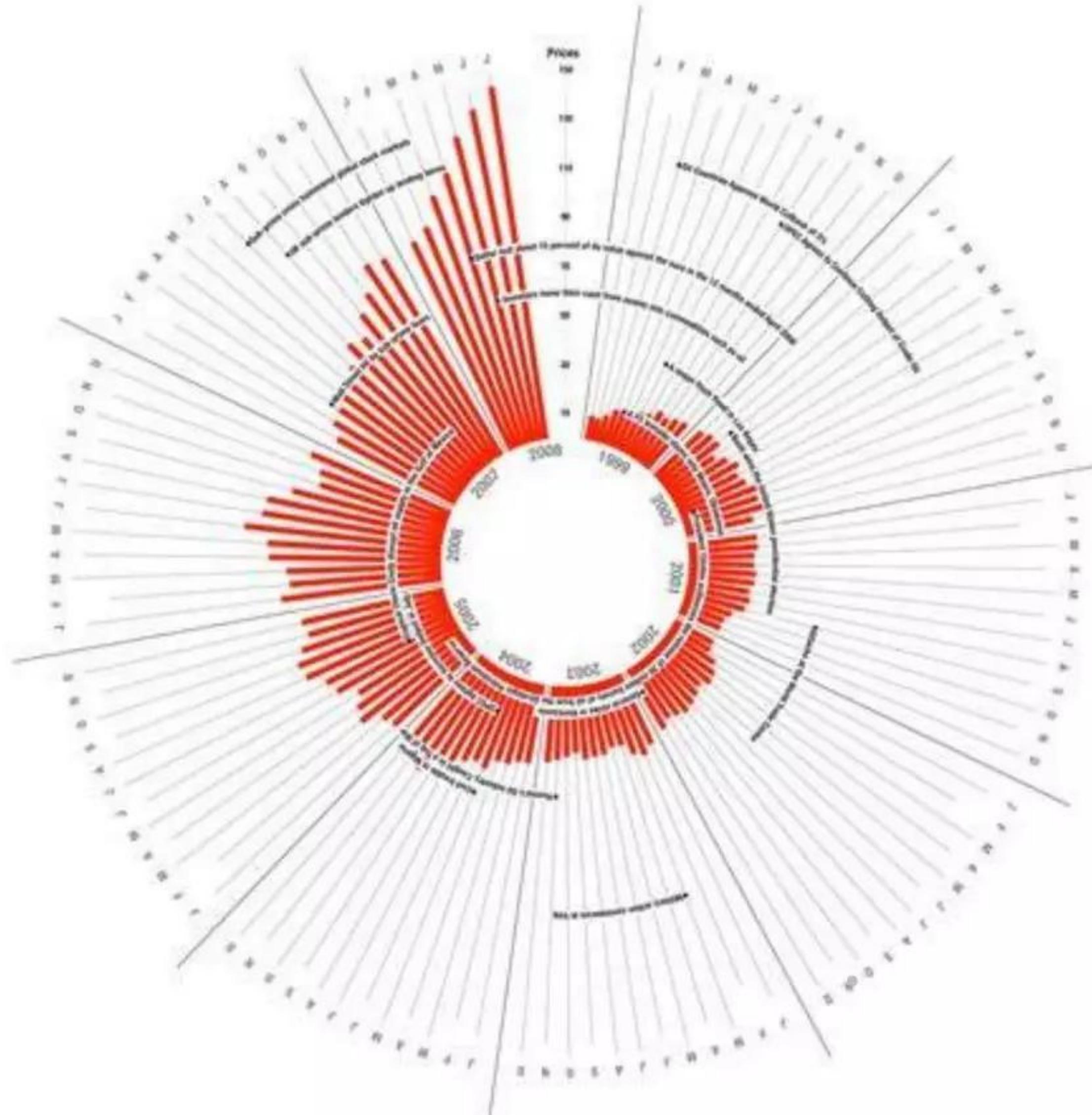
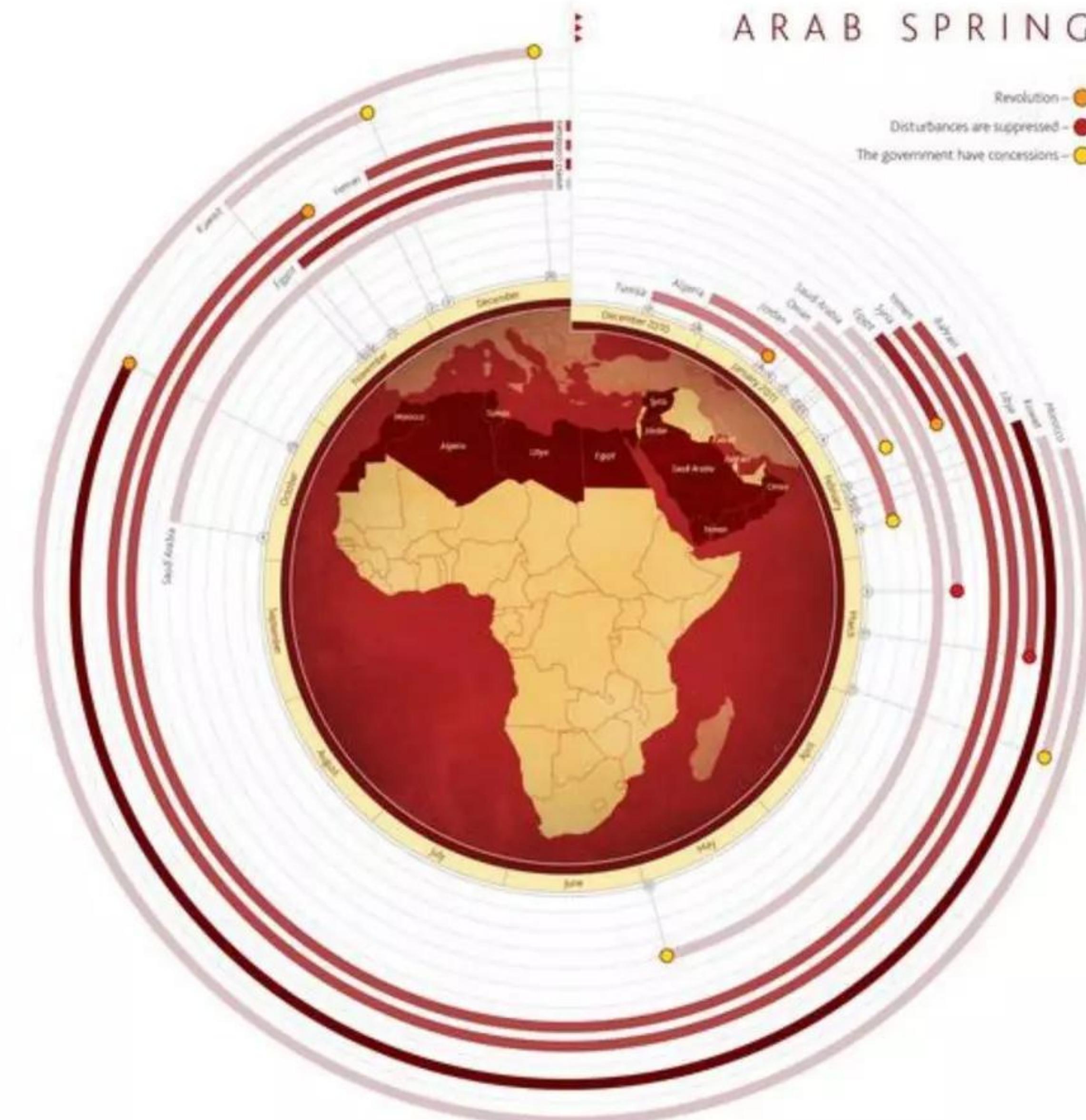
data visualization

Stephen Tracy

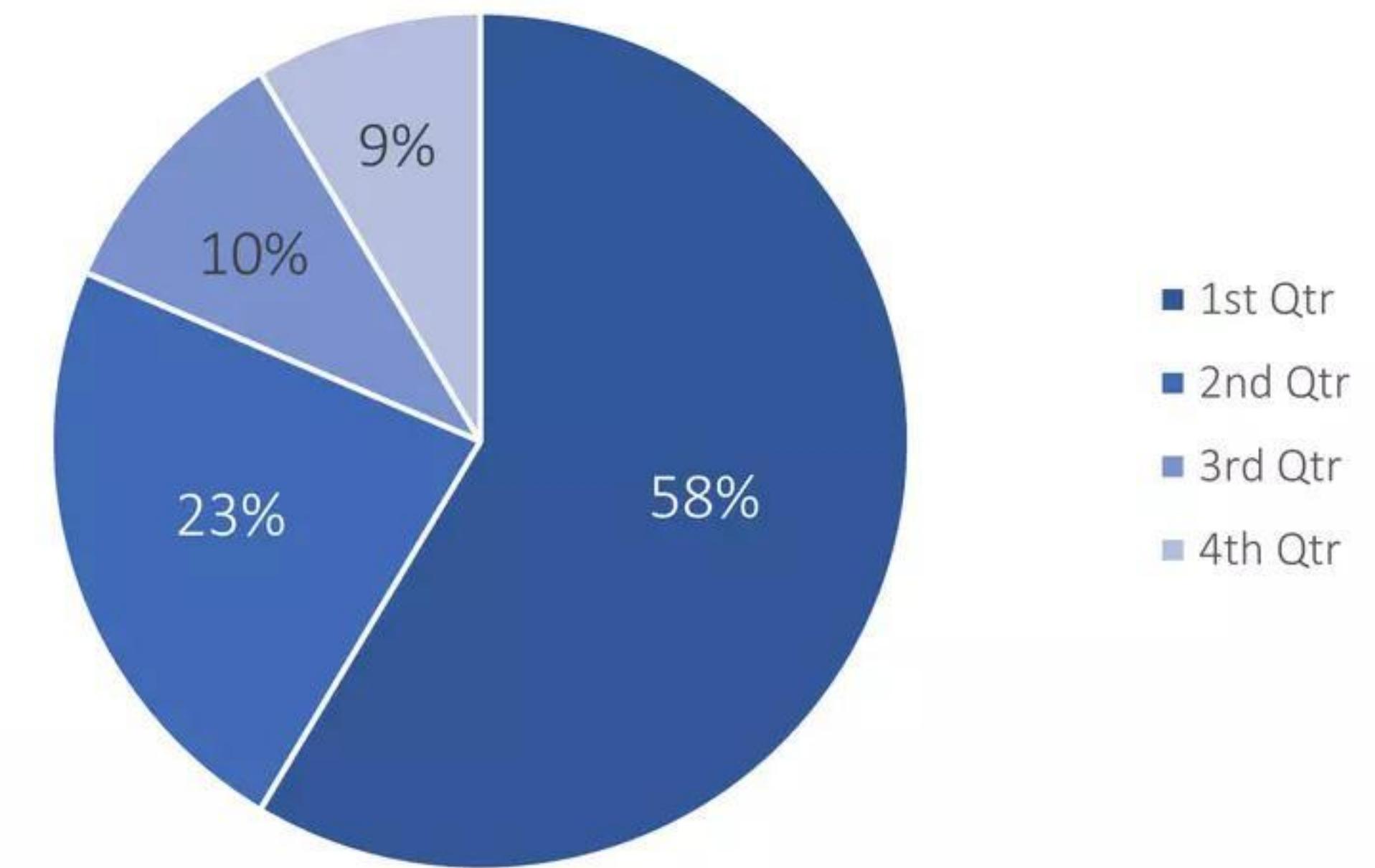
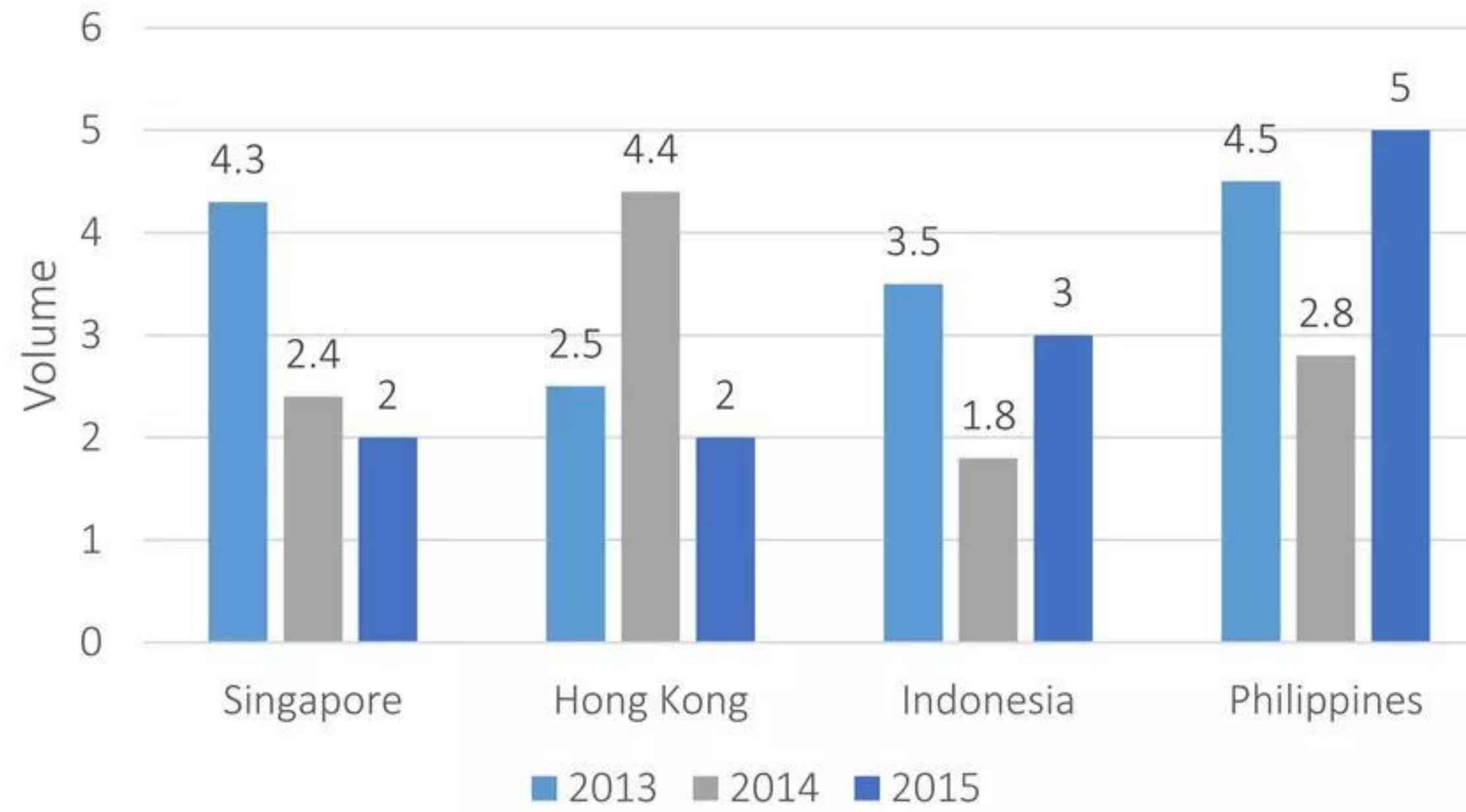
Data & Insight Lead, SEA



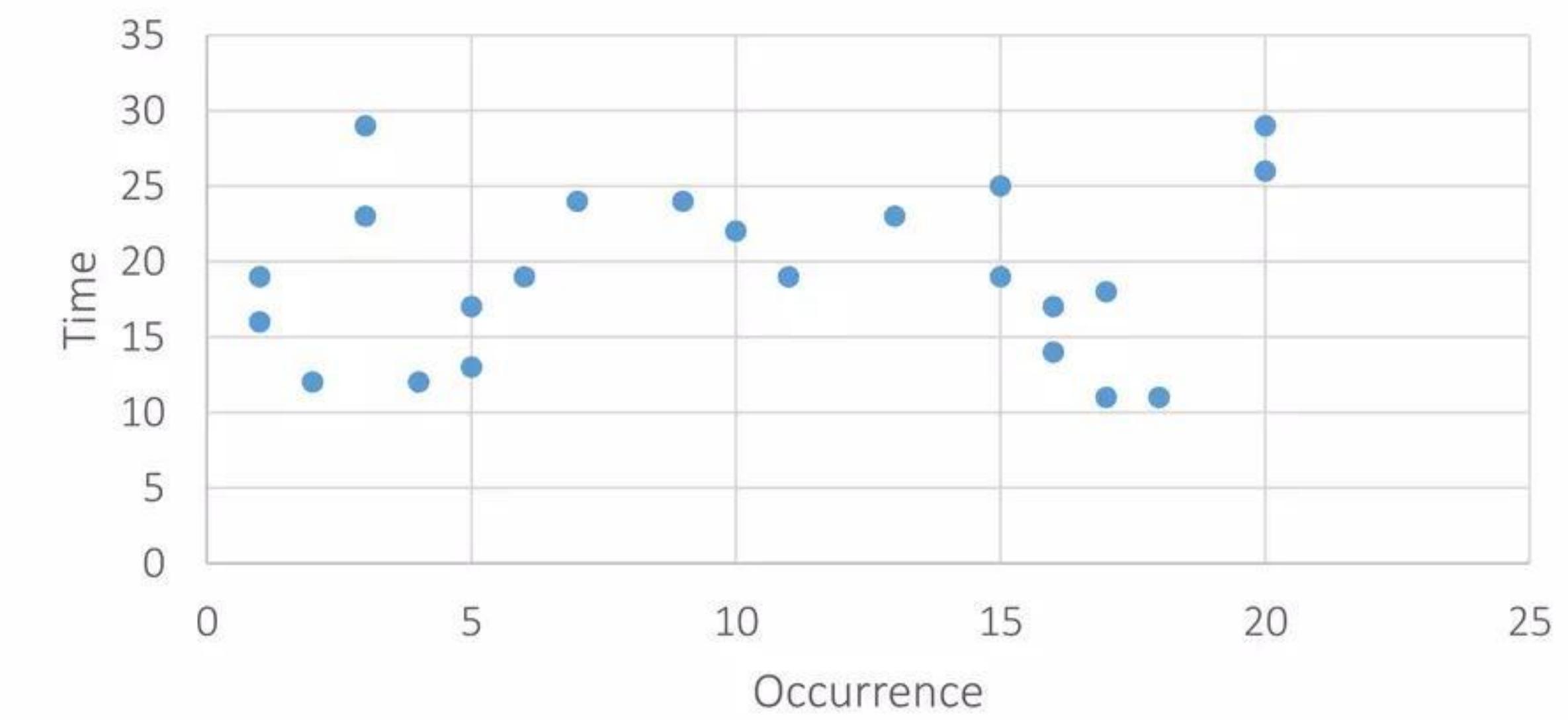
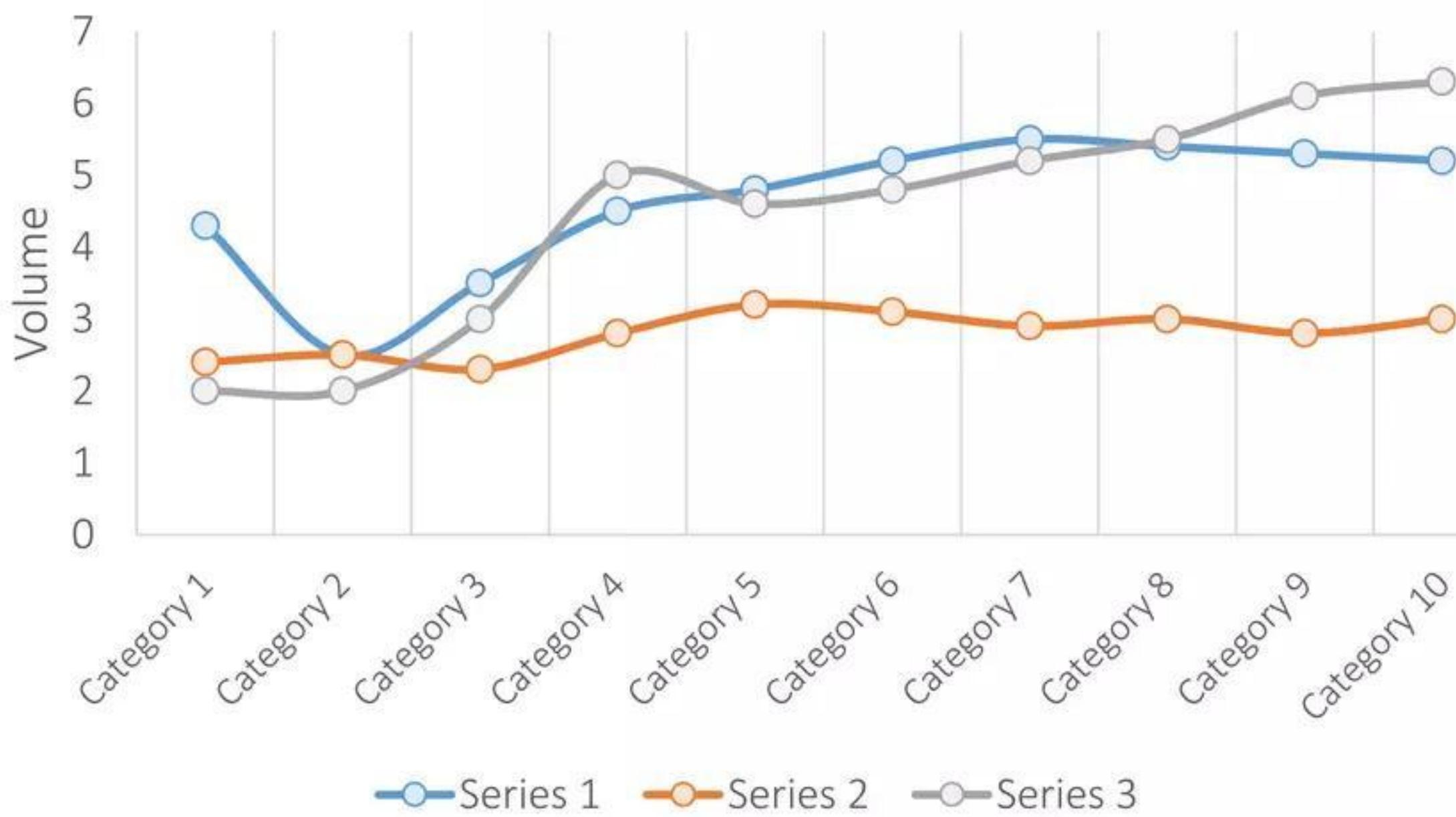
ARAB SPRING



when most people think about dataviz, they think about this



but data visualization starts with the basics



A wide-angle photograph of a rugged mountain range. In the foreground, a waterfall cascades down a dark, rocky cliff face. The middle ground shows another rocky outcrop with sparse evergreen trees. The background is heavily shrouded in thick, white mist or low-hanging clouds, obscuring the sky. A rectangular text box with a thin black border is positioned in the upper left area of the image. Inside the box, the words "some history" are written in a large, serif font.

some history

data visualization.

the visual representation of data

a primary goal of data visualization is to communicate information **clearly** and **efficiently** to users via the statistical graphics, plots, information graphics, tables, and charts selected

“the purpose of visualization is insight, not pictures”

- Ben Shneiderman, computer scientist

charles joseph minard.

■ Napoleon's Russian campaign of 1812

■ Produced in 1869

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussees en retraite

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en tracets des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent Oroscha et Witebsk, avaient toujours marché avec l'armée.

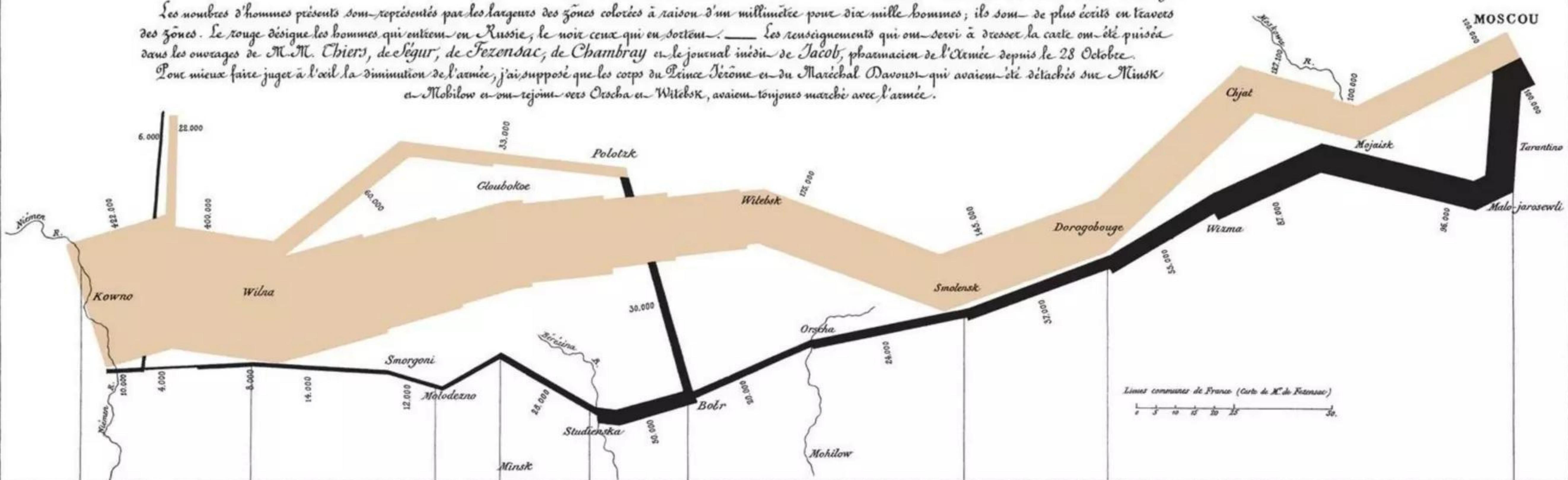
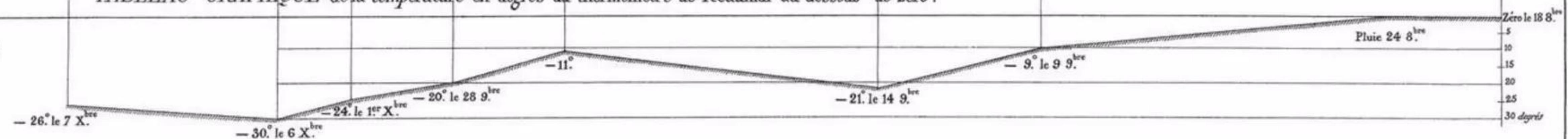


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

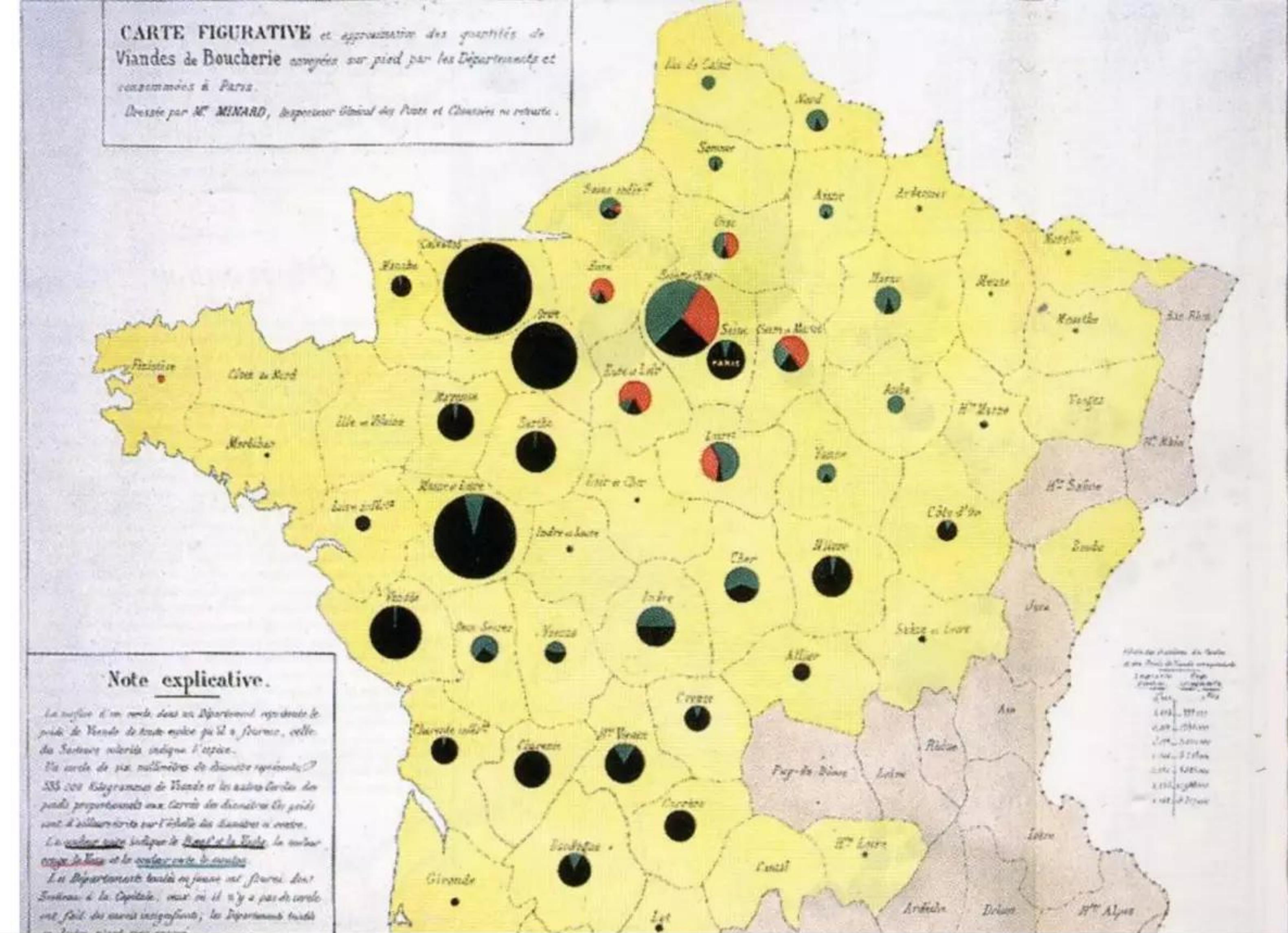
Les Cosaques passent au galop
le Niémen gelé.



Minard's map is notable for plotting 6 different data points on a single graphic

Charles joseph minard.

- Cattle distribution and consumption in France
- Produced in 1858



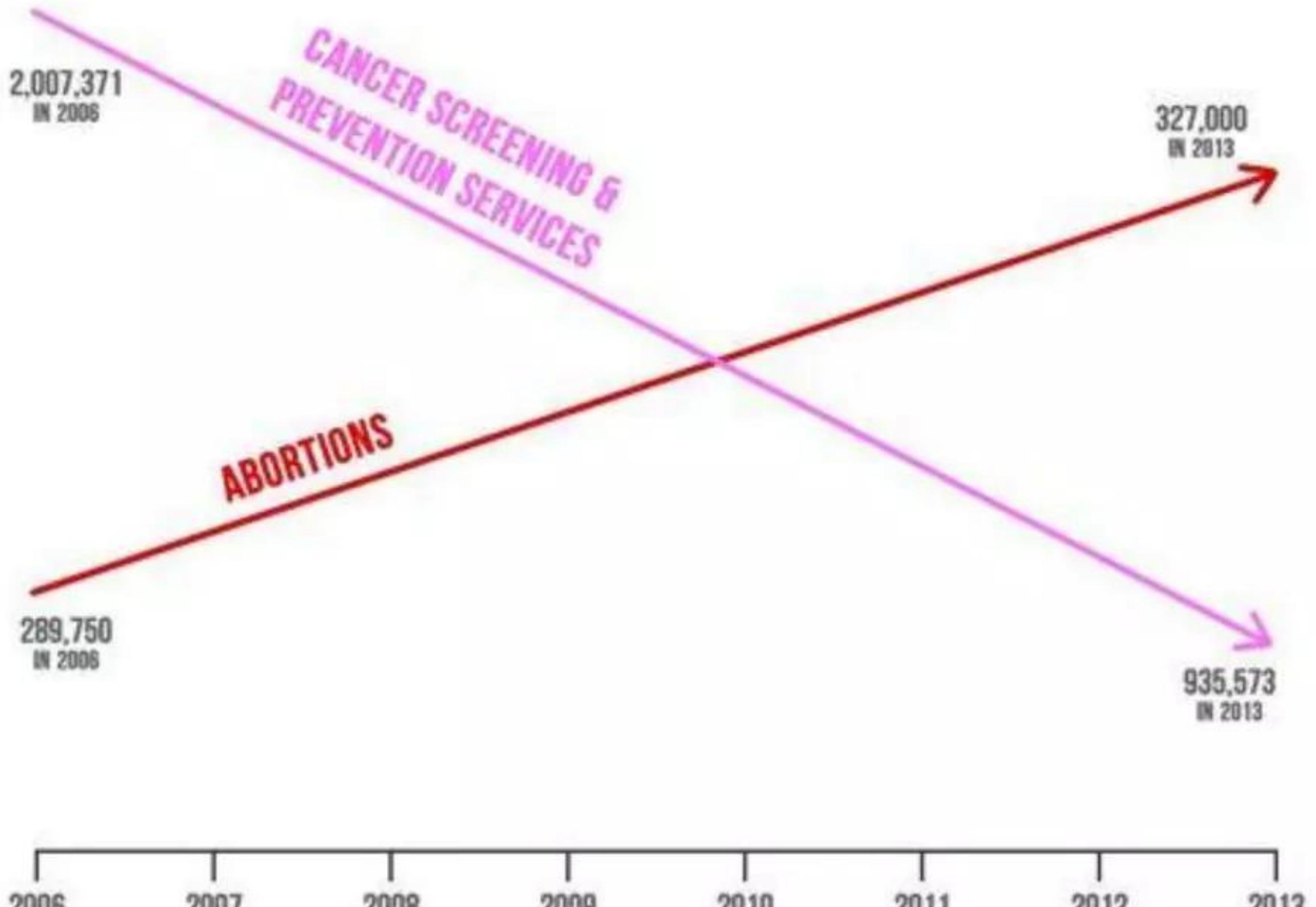
although this wasn't the first use of the pie chart, Minard popularized it's use with his graphic depiction of cattle volumes and distribution throughout France. Remember, this was hand drawn!

A wide-angle photograph of a forested hillside. The foreground shows dark green, low-lying vegetation and shrubs. In the middle ground, a dense forest of tall evergreen trees rises up the hill. The sky is filled with thick, white clouds that hang low over the treetops, creating a misty and atmospheric scene.

some not so good examples

Example 1 - Lying with data

PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP – LIFE-SAVING PROCEDURES DOWN



SOURCE: AMERICANS UNITED FOR LIFE

what is it?

- Chart shown by Rep Jason Chaffetz during Planned Parenthood hearing
- Watch it here - <http://bit.ly/1UhTYoU>

what's wrong with it?

- No Y axis
- Multi-axis not declared and labelled
- X-axis spans 8 years, but chart only includes 2 years (2006 and 2013)

Example 1 - Lying with data

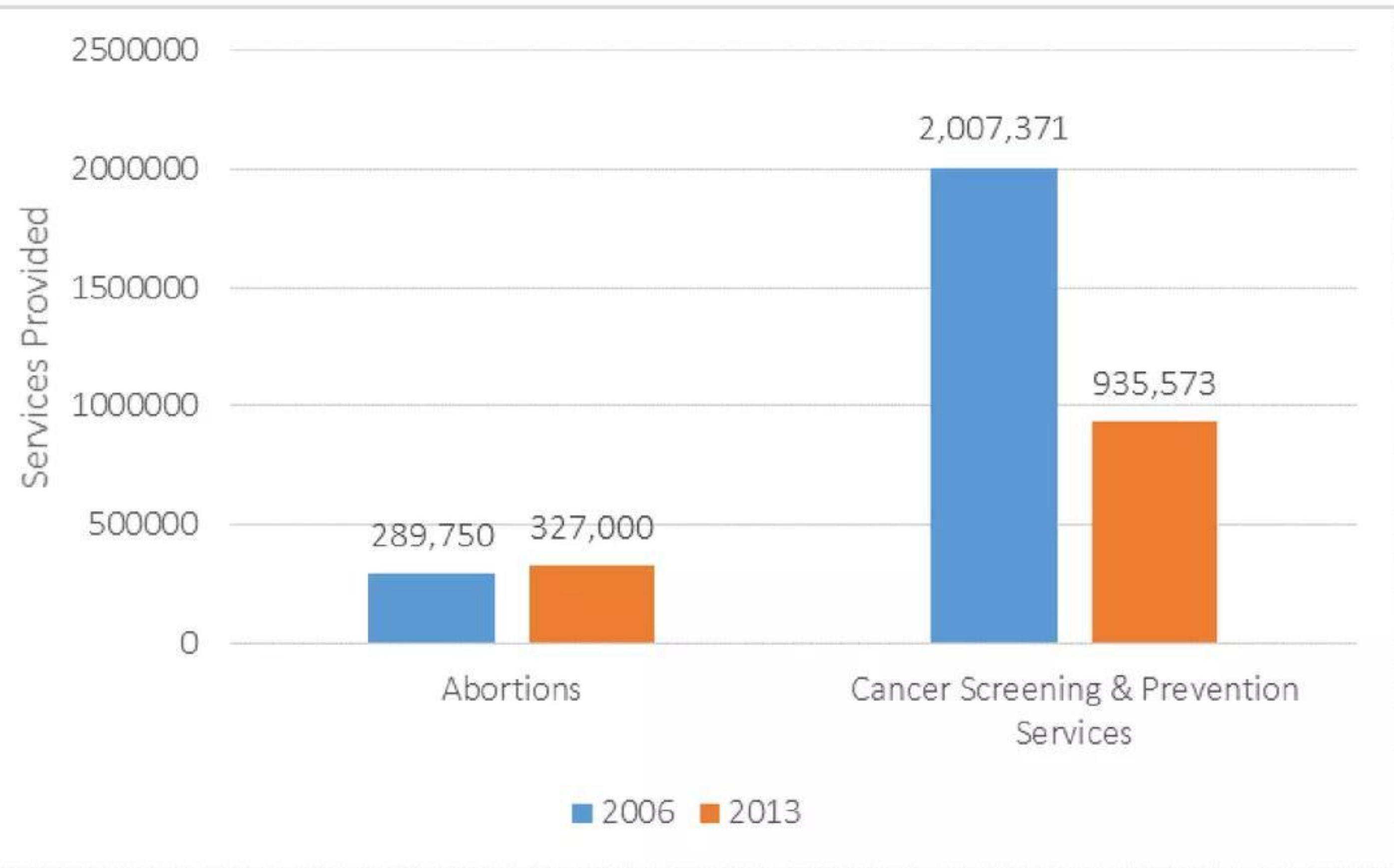


Figure 1. Planned Parenthood – Abortions vs Cancer Screening Services



Bar Chart | Single Axis | Vertical

how to fix it?

- | if you are comparing data between 2 non-adjacent years a bar chart would be a better choice
- | here's the same data in a bar chart, and on a single axis

Example 1 - Lying with data

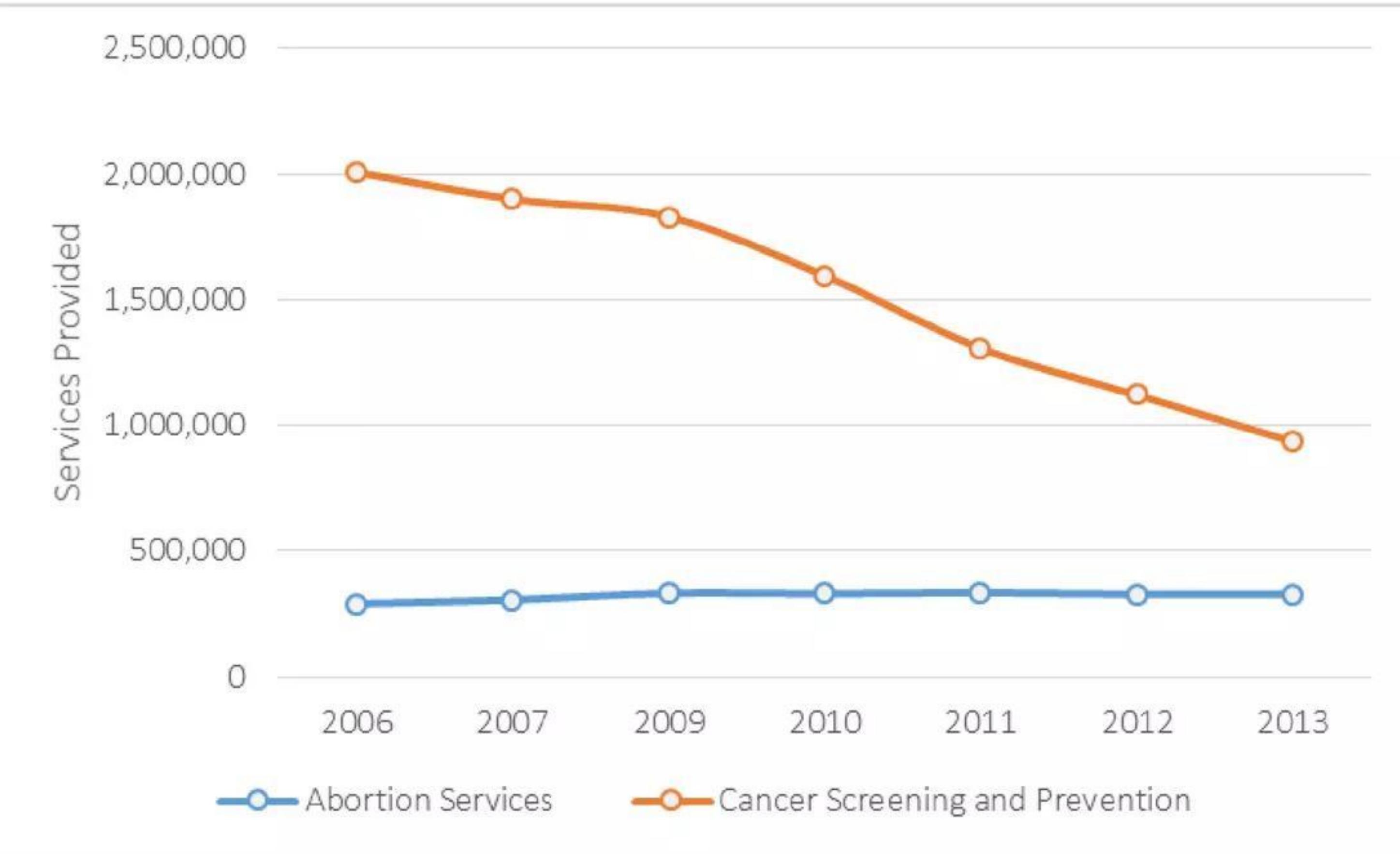
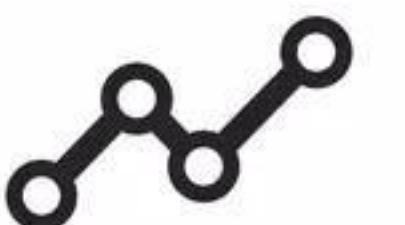


Figure 2. Planned Parenthood – Abortions vs Cancer Screening Services



Line Chart | Single Axis

how to fix it?

Better yet, if you retrieve the data for the missing years you can use a line chart to show the change over time

here's the same data in a line chart with the missing years (2007 - 2012) included.

*Note, 2008 is missing as I couldn't find the report for that year

Example 1 - Lying with data

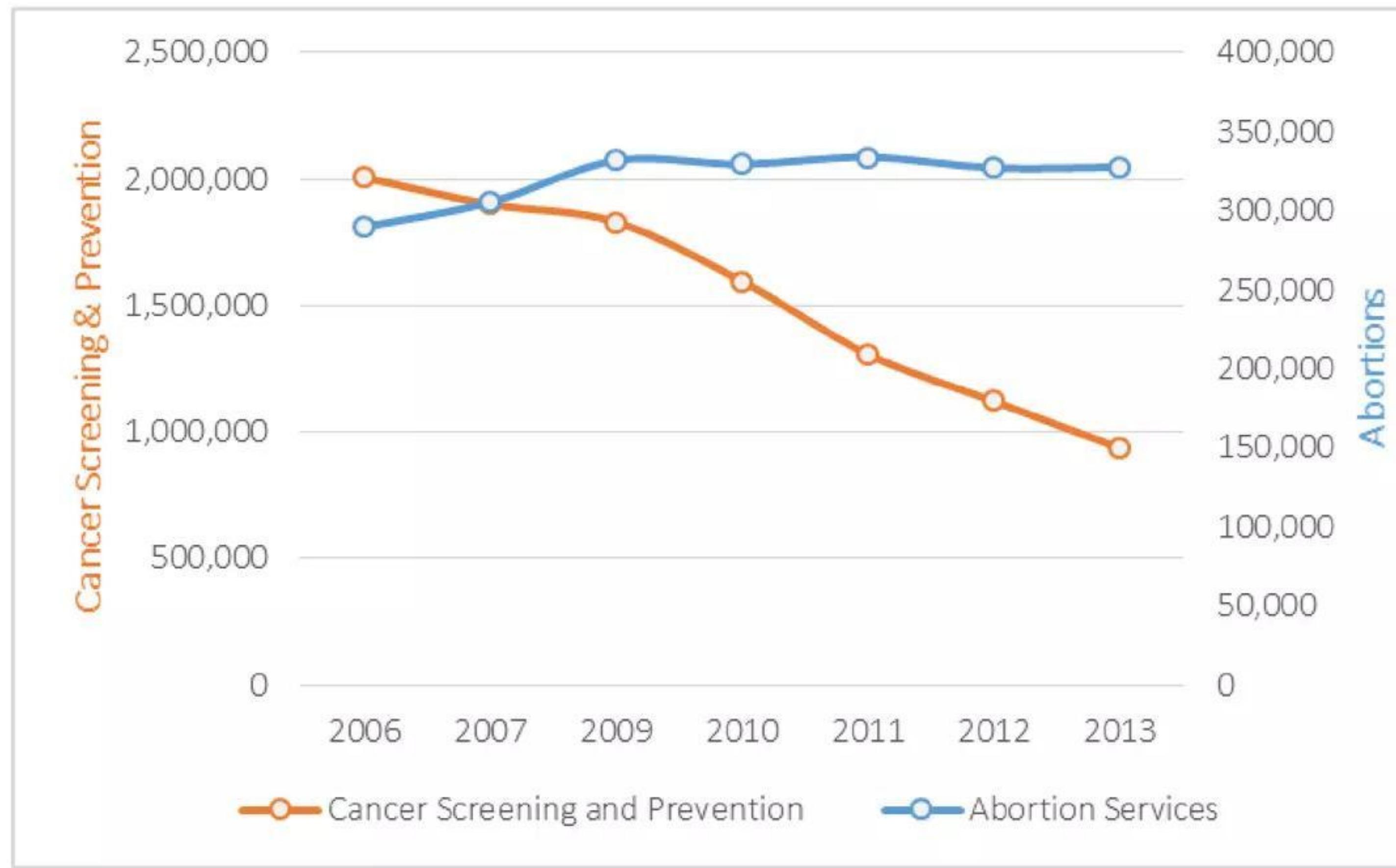
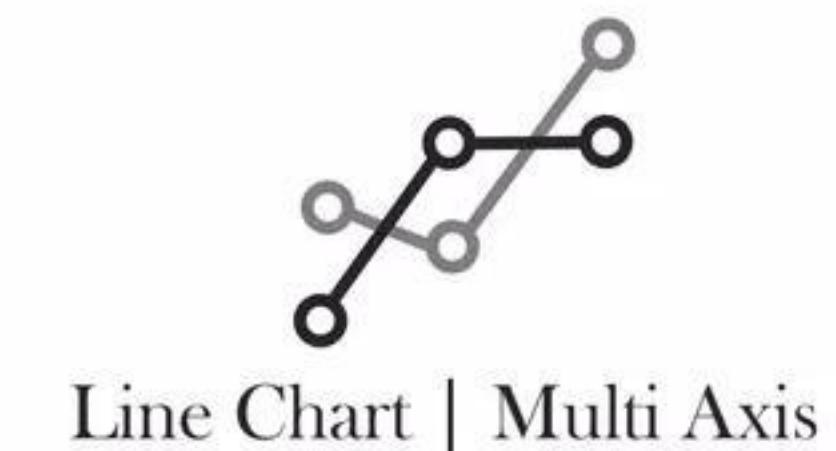


Figure 3. Planned Parenthood – Abortions vs Cancer Screening Services



how to fix it?

And for good measure, here's the same data represented using a multi axis chart, and with the axis' properly labelled

Example 1 - Lying with data

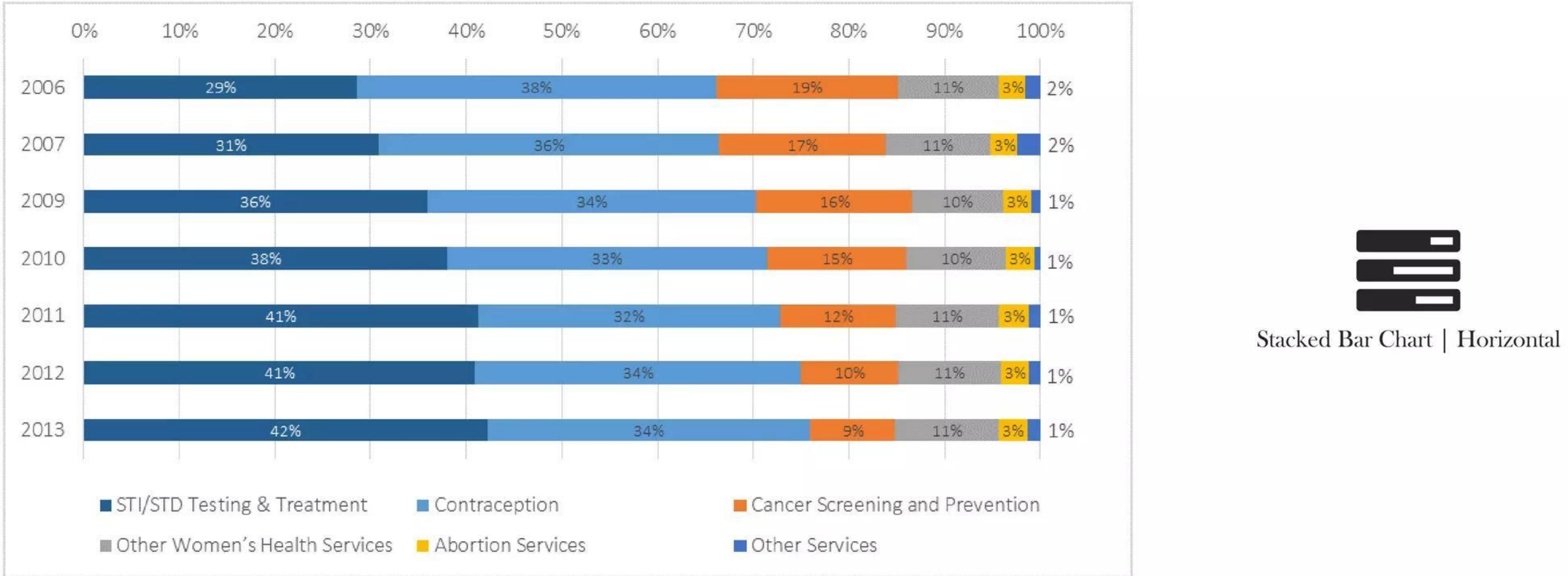


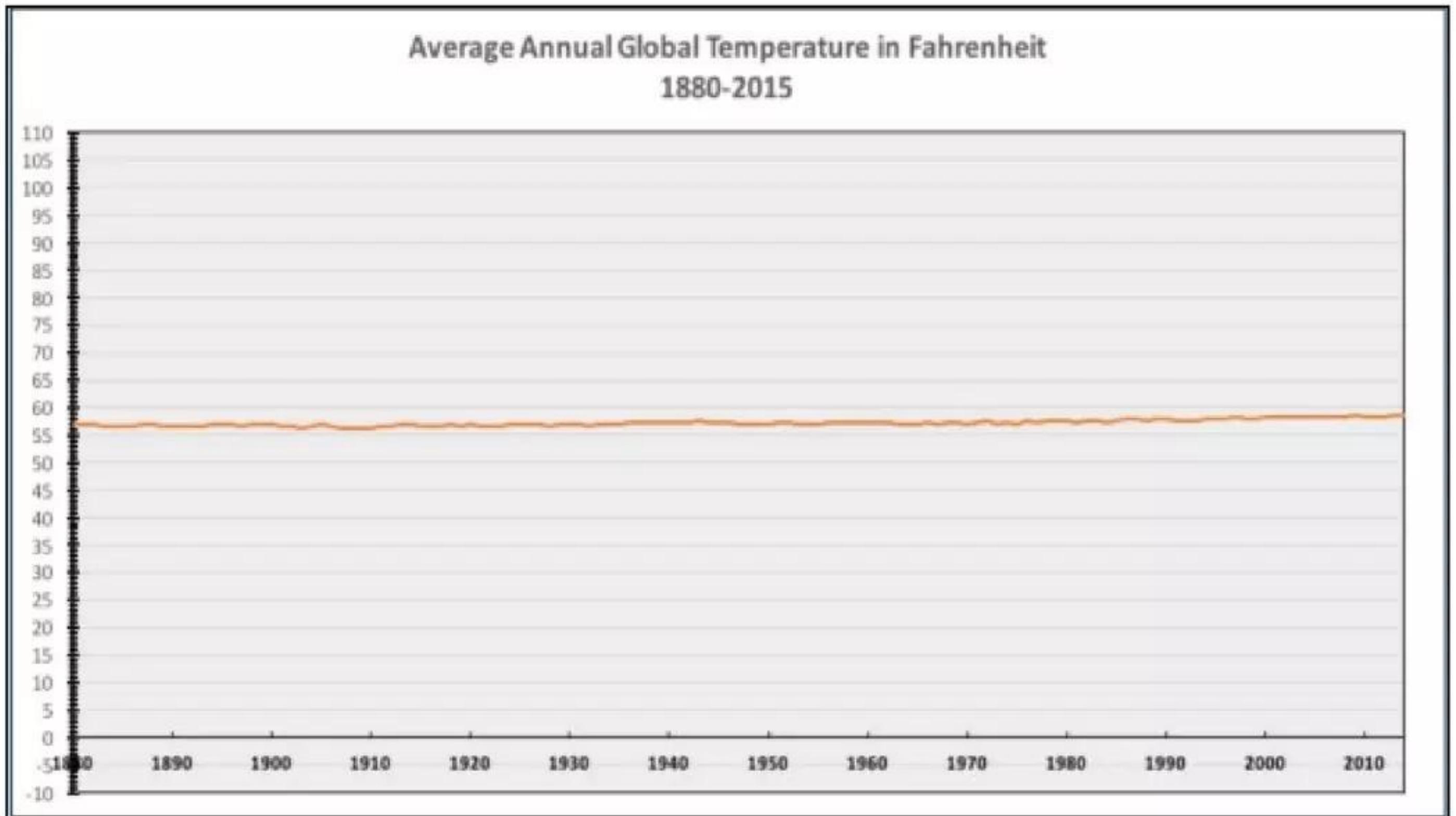
Figure 4. Planned Parenthood – Breakdown of Services (%)

how to fix it?

Here's a slightly different way to look at this data. The above is a stacked bar chart, which shows the % share of all services PP offers, not just Abortions and Cancer Screening. Notice that, as a % of total services, Abortions is constant at 3% every year.

Read more about this chart on my blog - <http://bit.ly/1nrbi0x>

Example 2 - Lying with data



what is it?

- national review (@NRO) chart depicting change in global average temperature over time
- See it here - <http://bit.ly/23klTus>

what's wrong with it?

- Y axis inflated, removes reader's ability to see meaningful change in data
- Chart does not convey context. The increase in global average temp of just 2 degrees can have significant effects on our planet, so using 120 point scale removes important context and makes this chart impossible to read

Example 2 - Lying with data

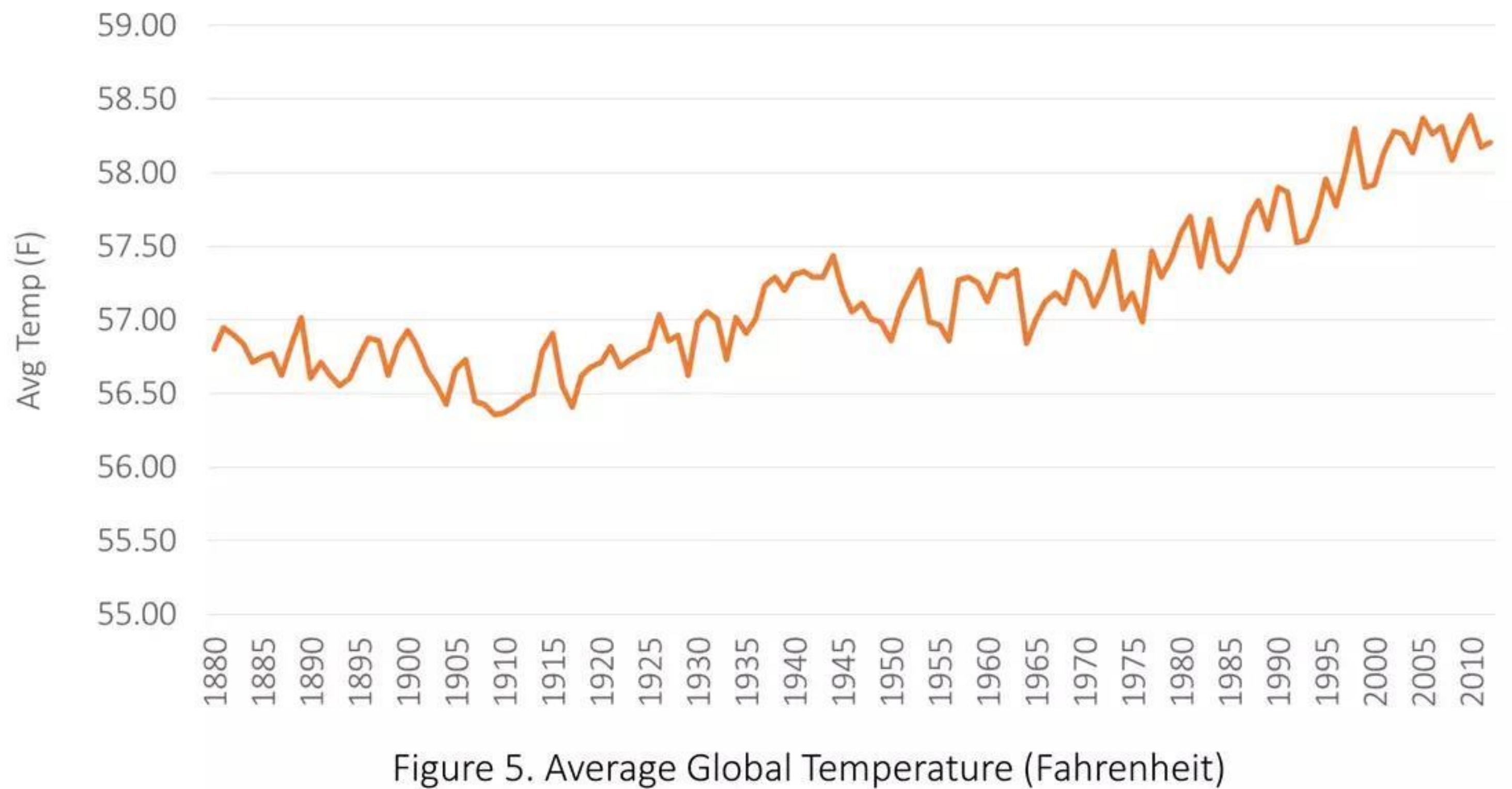


Figure 5. Average Global Temperature (Fahrenheit)



Line Chart | Single Axis

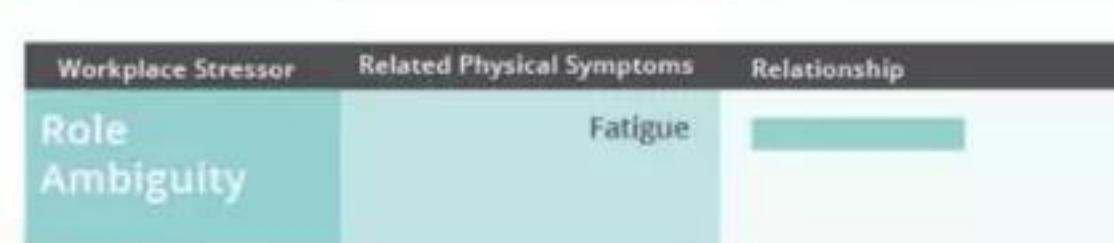
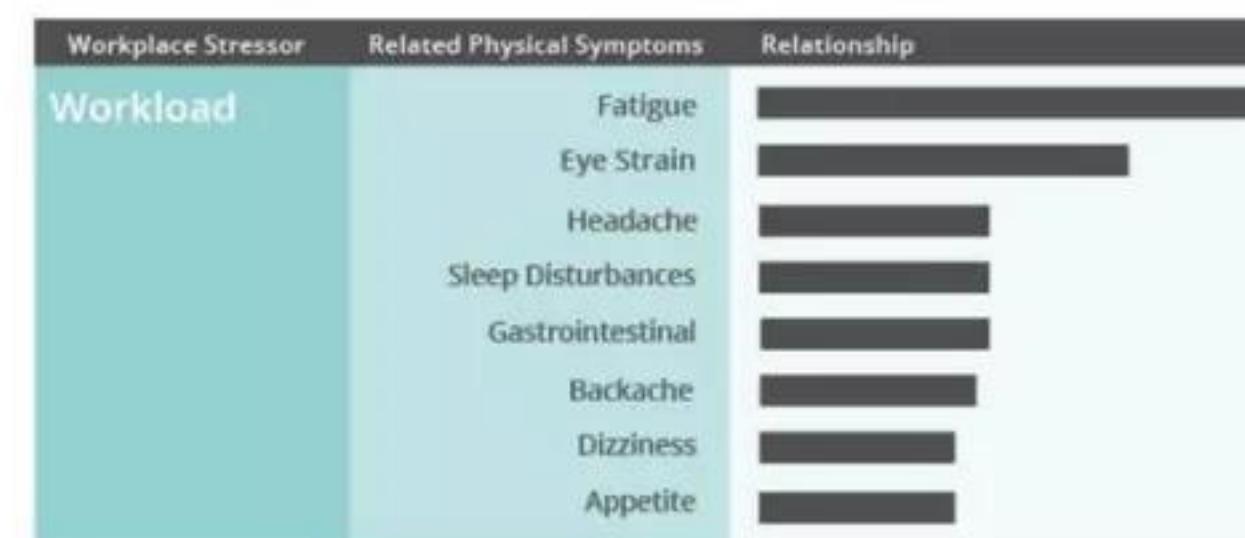
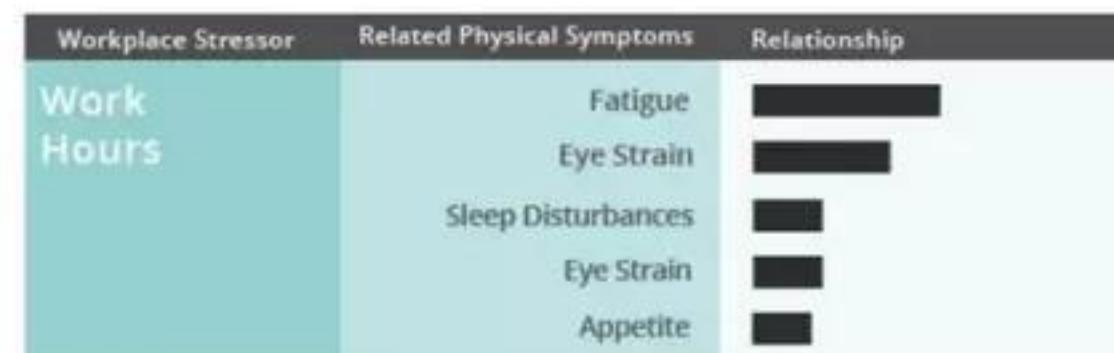
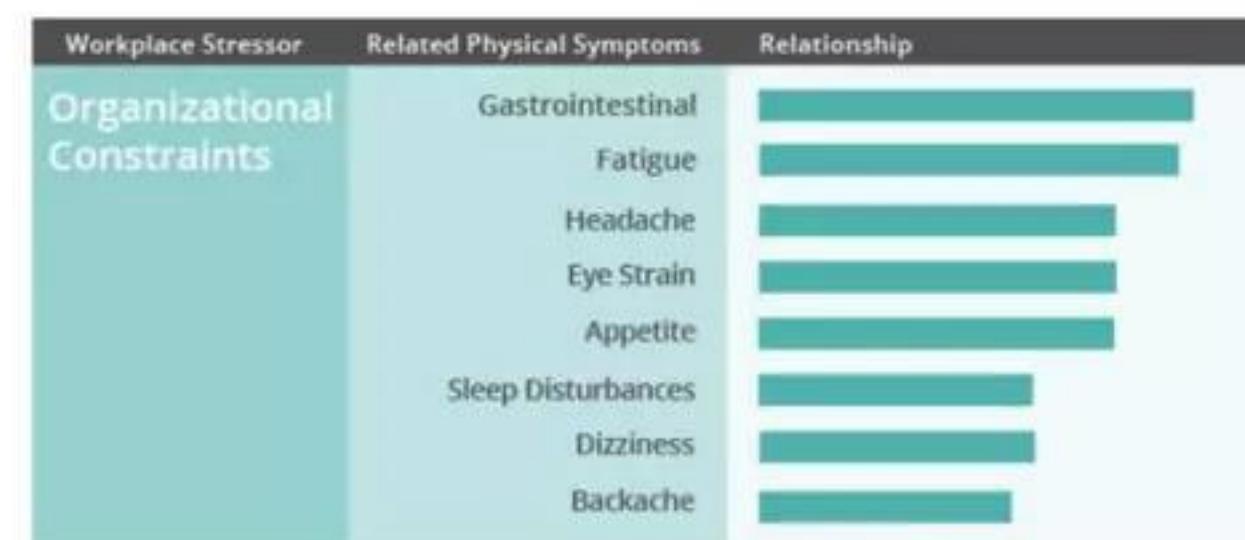
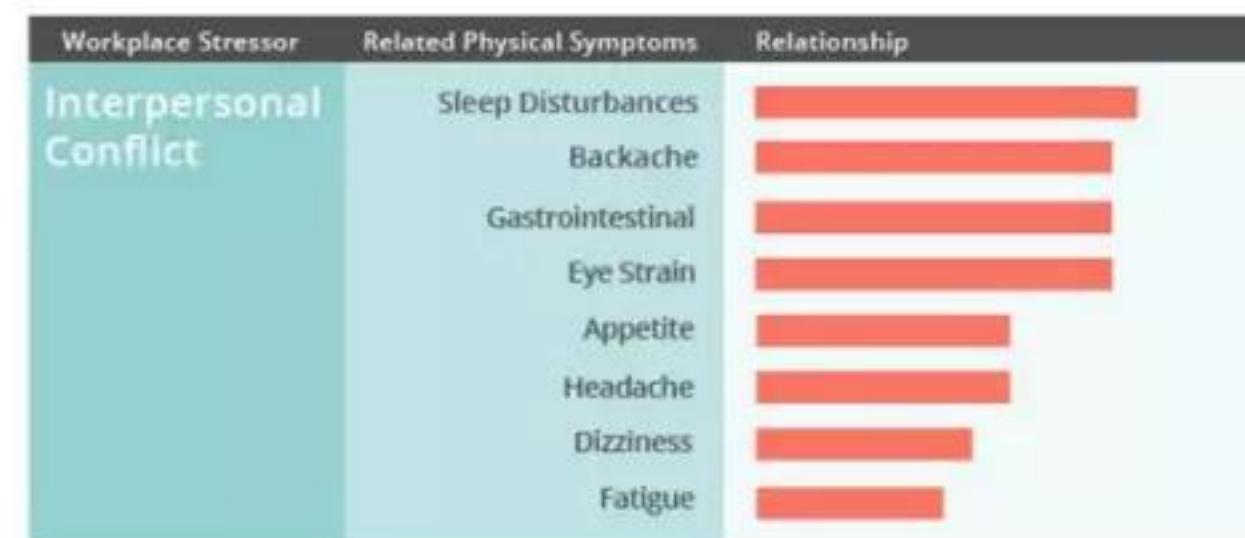
how to fix it?

There is much debate over whether y-axis baselines should always be zero, and this is a good example of why that isn't an unbreakable rule. In this case, the context of global temperature is that the average annual temp has never in our history dropped below 56 degrees, and very small change in this unit of measure can yield significant outcomes. Therefore, a tighter scale range and non-zero baseline would help make this chart clearer to the reader

Example 3 - Misleading labels and headings

WORK PLACE STRESSORS AND THEIR RELATIONSHIP TO PHYSICAL SYMPTOMS:

A comparison of workplaces stressors and what health symptoms they cause.



Bar Chart | Single Axis | Horizontal

what's wrong with it?

- | misleading column label
- | doesn't clearly communicate the unit of measure

how to fix it?

- | clearly and meaningfully label charts and axis
- | Read more about how to fix this chart:
<http://bit.ly/1nxzbZ8t>

Example 4 - Charts within charts and lack of focus



Pie Chart

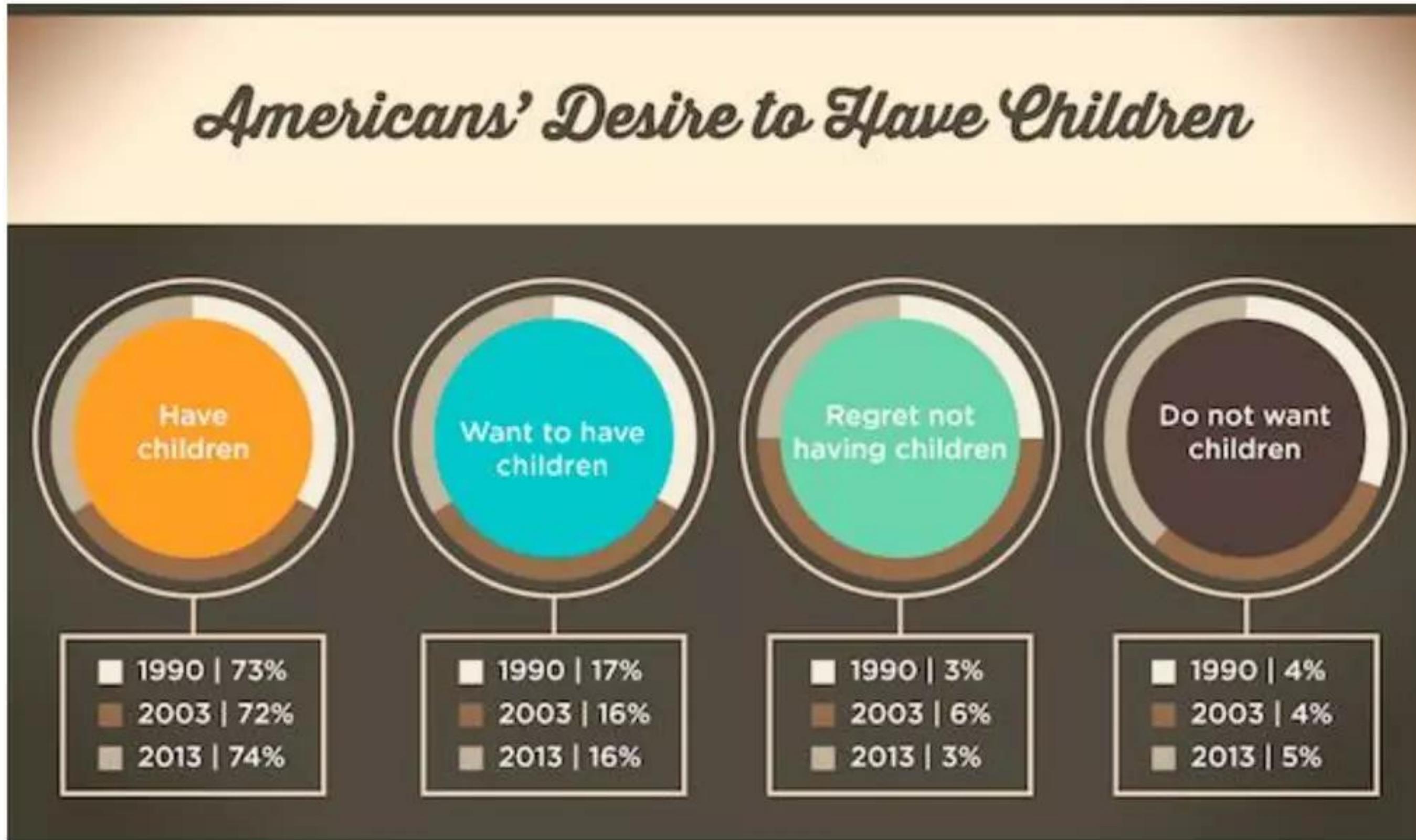
what's wrong with it

- | charts within charts
- | convoluted datapoints and interpretation

how to fix it?

- | don't embed charts within charts
- | don't make data needlessly confusing. If it's difficult to interpret your not finished
- | Read more about how to fix this chart:
<http://bit.ly/1nxbz8t>

Example 5 - The parts don't add up to a whole



what's wrong with it

- improper use of pie chart
- data presented in chart is different from the legend

how to fix it?

- understand which chart to use based on your data and what you want to communicate
- Read more about how to fix this chart:
<http://bit.ly/1nxbz8t>

Example 5 - Meaningless visualization



what's wrong with it

- ─ bubble size has no relation to value inside
- ─ unintentional Venn diagram gives data new meaning

how to fix it?

- ─ Always think about how your design choices impact the interpretation of your data and design with purpose
- ─ Read more about how to fix this chart:
<http://bit.ly/1nxbz8t>

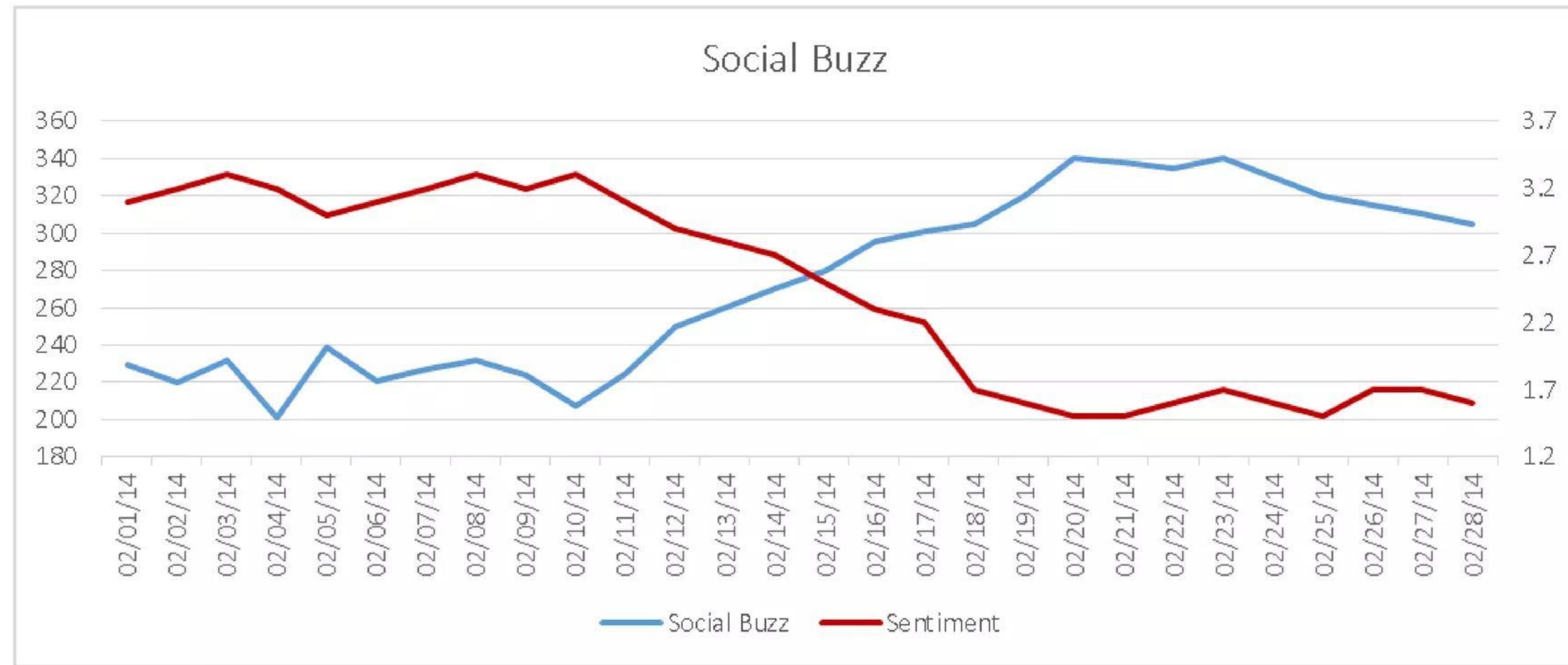
A landscape photograph showing a winding asphalt road curving through a dense forest of green trees and shrubs. The lighting suggests either early morning or late afternoon, with long shadows and a warm glow from the sun. In the upper portion of the image, there is a large, bright, overexposed area representing a clear sky.

some lessons

1

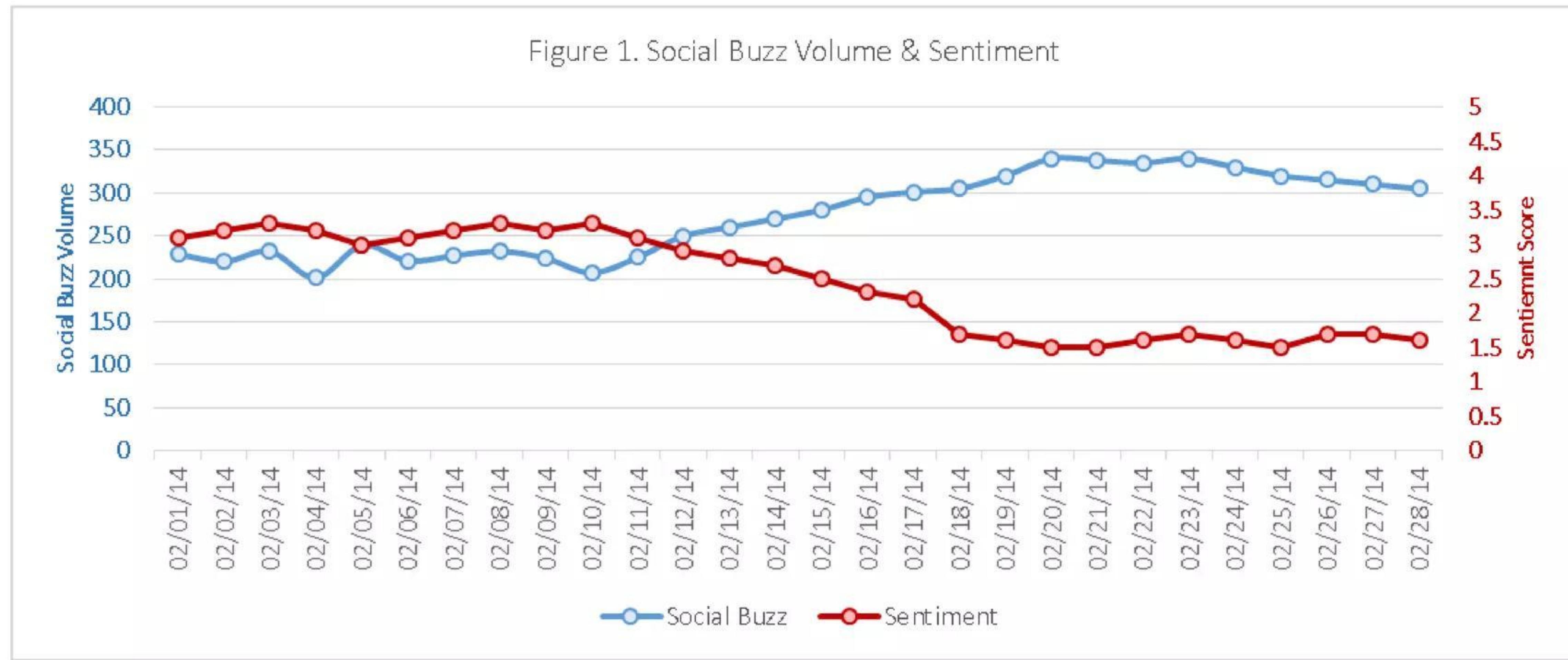
follow basic chart formatting rules

follow basic chart formatting rules



bad 😞

follow basic chart formatting rules

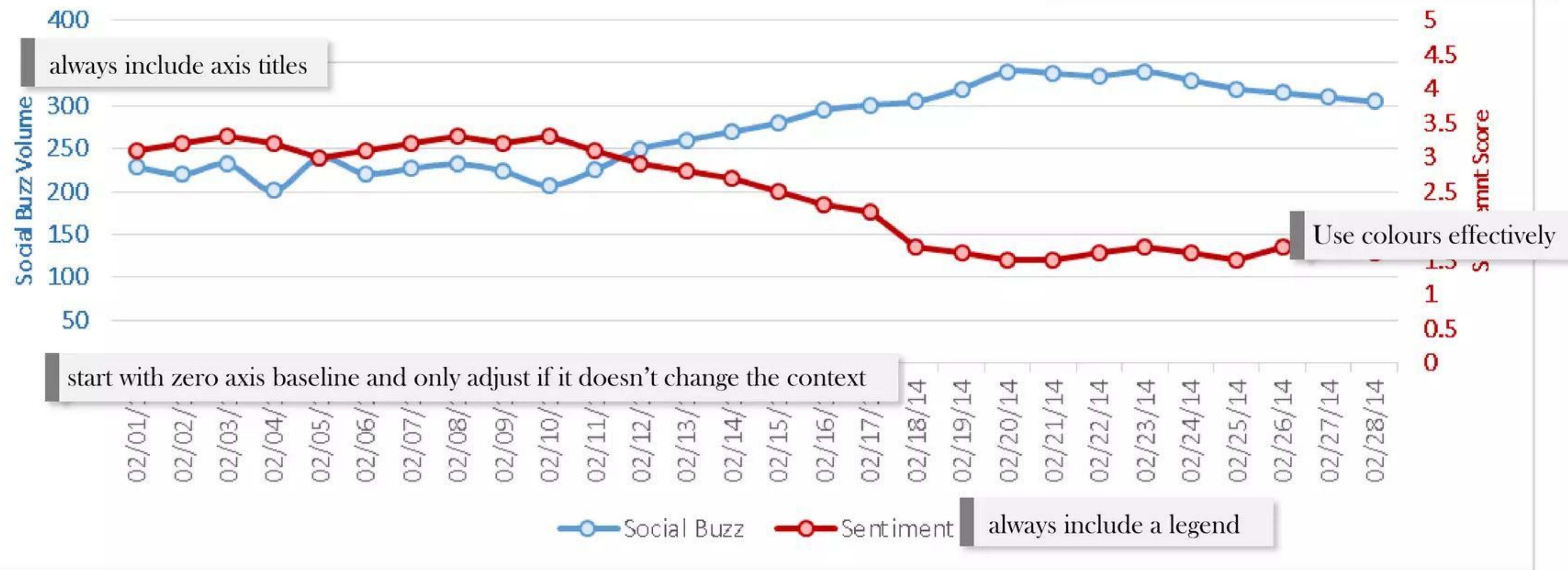


good 😊

follow basic chart formatting rules

Figure 1. Social Buzz Volume & Sentiment

add figure labels and use a descriptive title



good 😊

2

context is everything

context is everything



would you buy
this stock?

Blackberry (BBRY) Stock - 3 Month Performance

context is everything



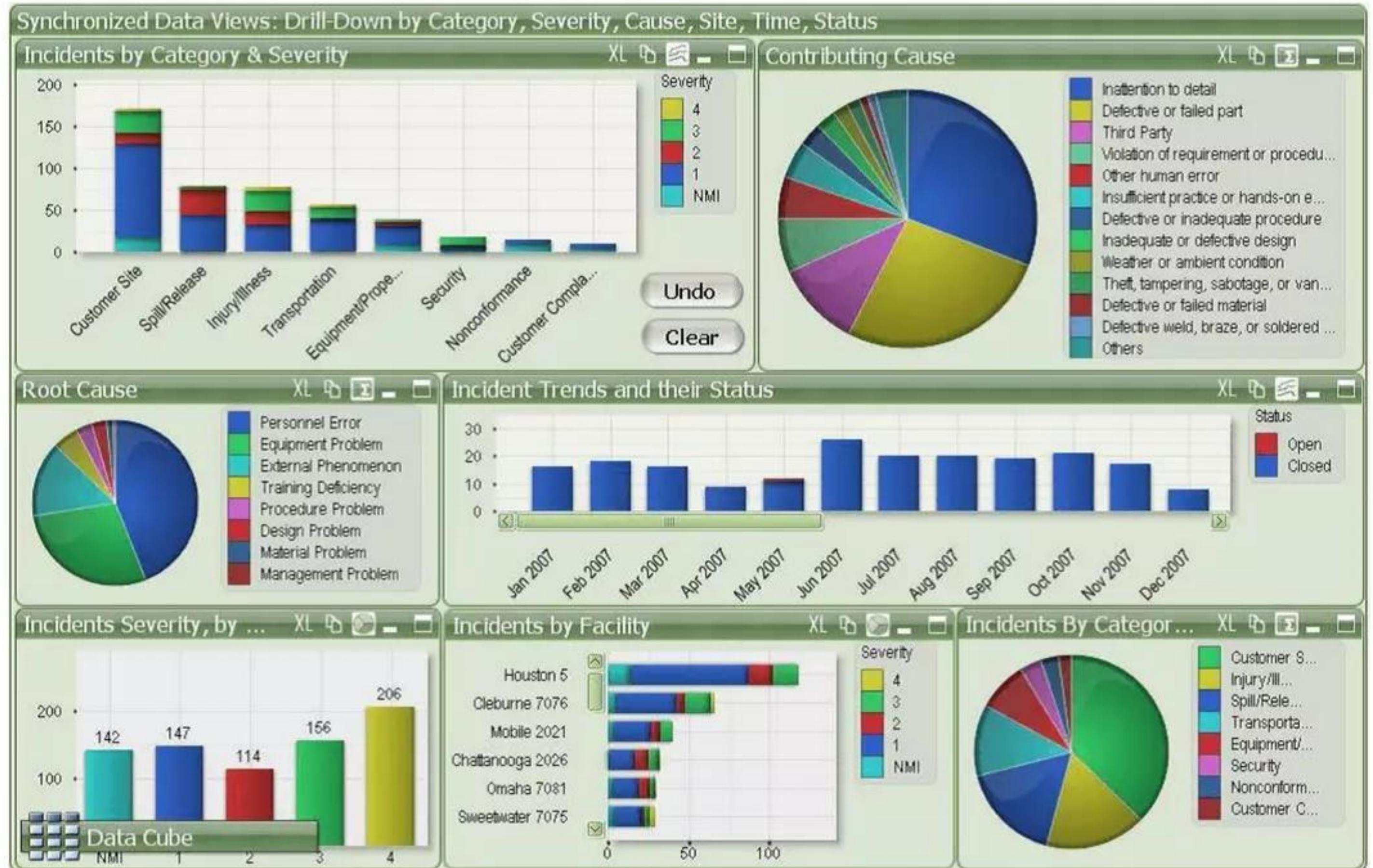
how about now?

Blackberry (BBRY) Stock - 5 Year Performance

3

design with purpose

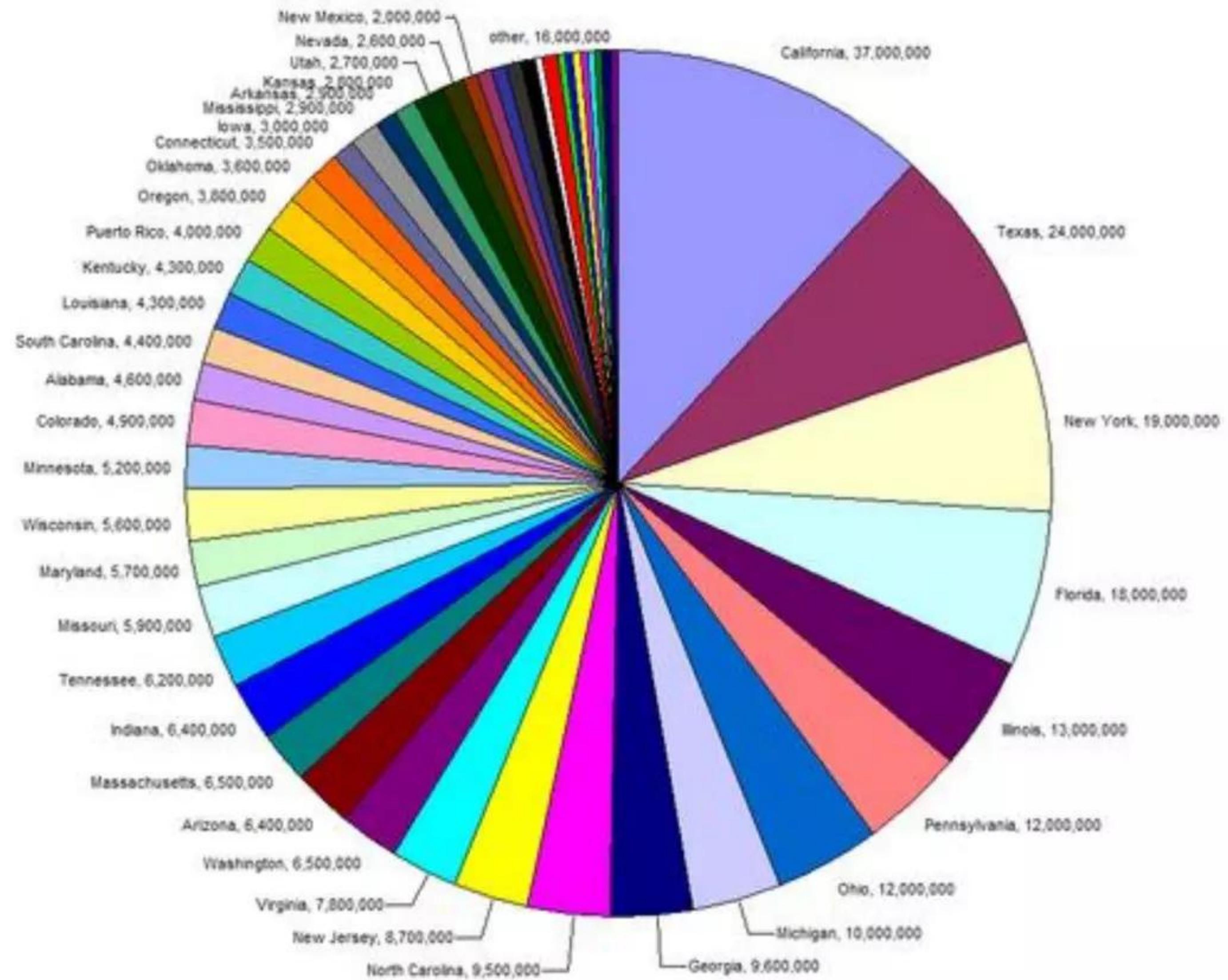
design with purpose



Don't just fill a dashboard or report with meaningless charts. Every chart and graphic should have a specific purpose

4

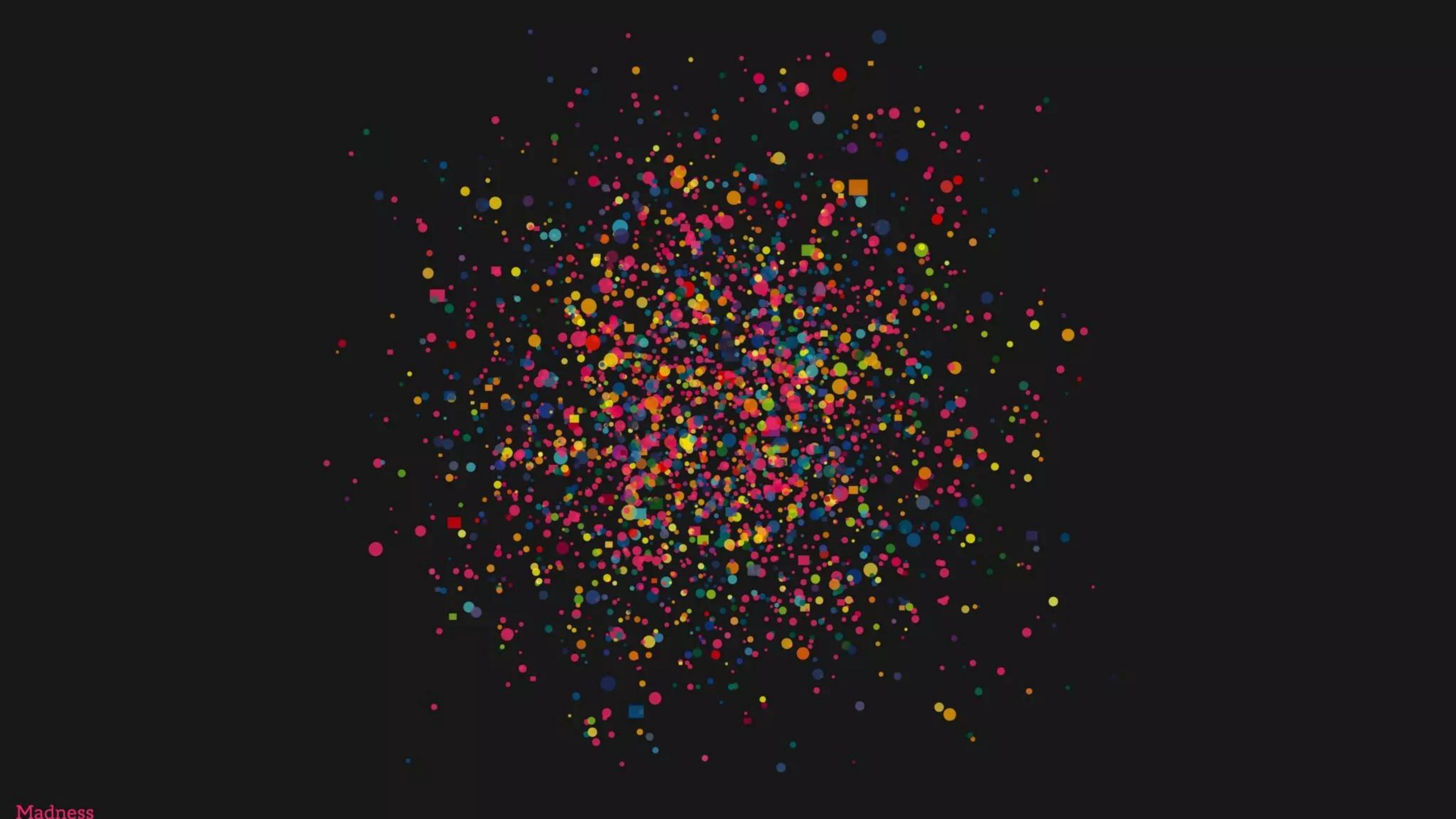
strive to make the complex simple



When it comes to visualizing data, your goal should always be to make the complex simple. If your chart looks like this, you have failed

5

make it beautiful



Madness

some resources.

- | Visage: Data Viz 101 - <http://bit.ly/1ncO4dK>
- | Flowing Data - Basic Rules for Making Charts - <http://bit.ly/1nqd3v7>
- | National Geographic - Spotting Charts that Lie - <http://bit.ly/1Guozac>
- | Analythical - My Blog! - analythical.com

thank you

 analythical.com

 stracy@sapient.com

 [@stephen_tracy](https://twitter.com/@stephen_tracy)

 [linkedin.com/in/tracystephen](https://www.linkedin.com/in/tracystephen)

