Data Visualization: The Underrated Data Skill

Soham Kudtarkar for Statistics 133 (12/3/2017)

Introduction

What comes to mind when statisticians, computer scientists, or data scientists describe data analysis is the high level abstract models that are meant to cut through huge amounts of data to produce useful results. Unfortunately, the visualization of the data or the results come only as an afterthought. In this paper, I will argue for why data visualization not only produces both useful and beautiful results, I will ultimately aim to show why it should be an essential tool in any data analyst's skill set.

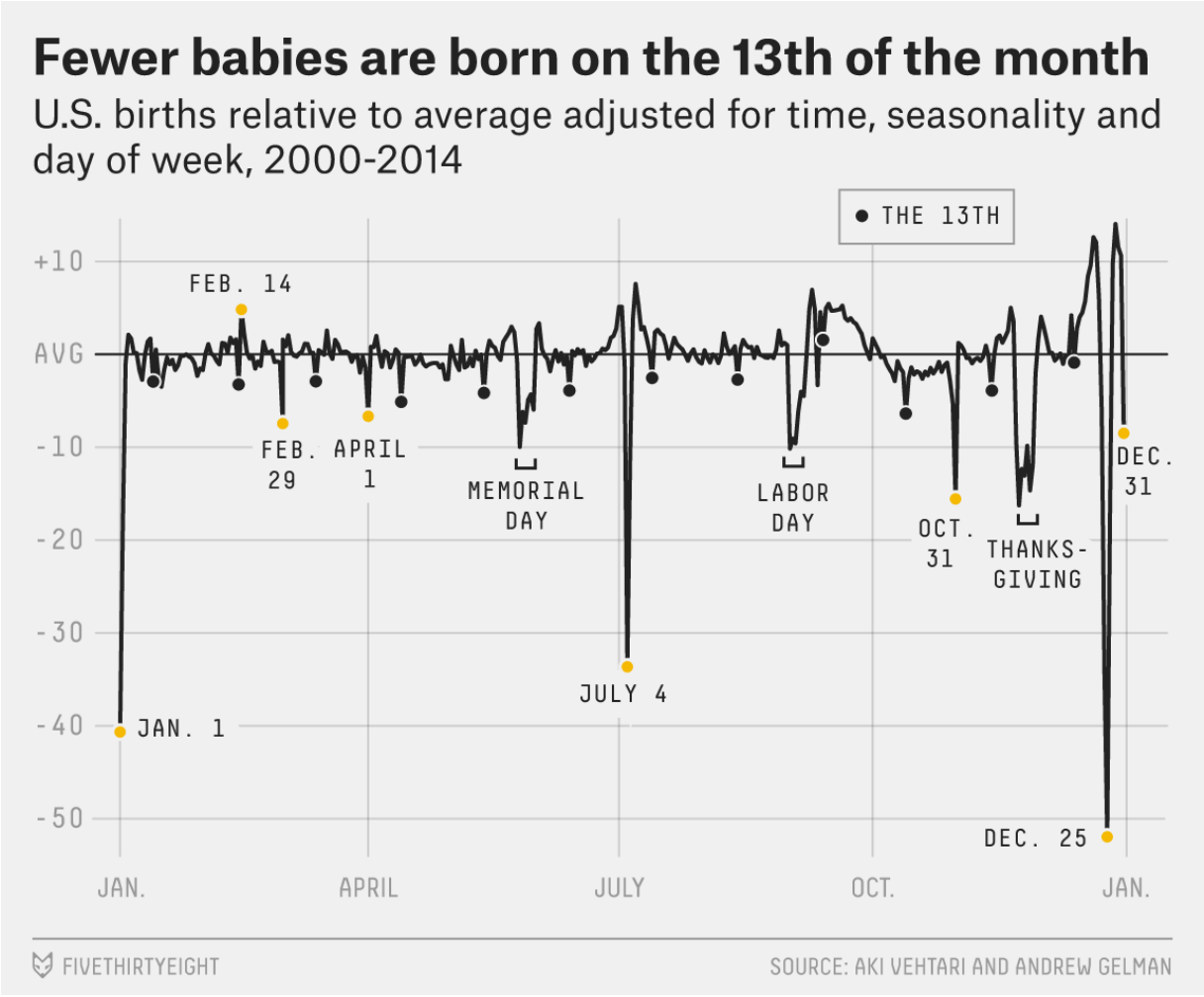
The outline of this post will be as follows:

- Introduction
- What can we do with data visualization?
- Why is data visualization important?
- Visualization demonstration
- Conclusion
- References

What Can We Do With Data Visualization?

There are two main reasons why a data analyst would visualize data.

The first is obvious: visualizing data gives us a new perspective on the data. We are able to see information in a way that we might not even be able to conceptualize. Visualizing data puts us into a frame of reference that allows us to glean information that we might not have been able to find simply with the many hundreds or thousands of rows we have in front of us. Take birth rates for instance.



A visualization of birth tendencies over a year (see reference A).

This visualization clearly gives us productive value from the dataset since we can glean that mothers-to-be are far less likely to have babies on holidays like the New Year, Christmas, and Thanksgiving. As the graphic also points out, fewer babies are born on Friday 13th. While we may have be able to arrive at these conclusions computationally (perhaps by a min-centered function), we have arrived at useful conclusions in a far more intuitive manner through this visualization.

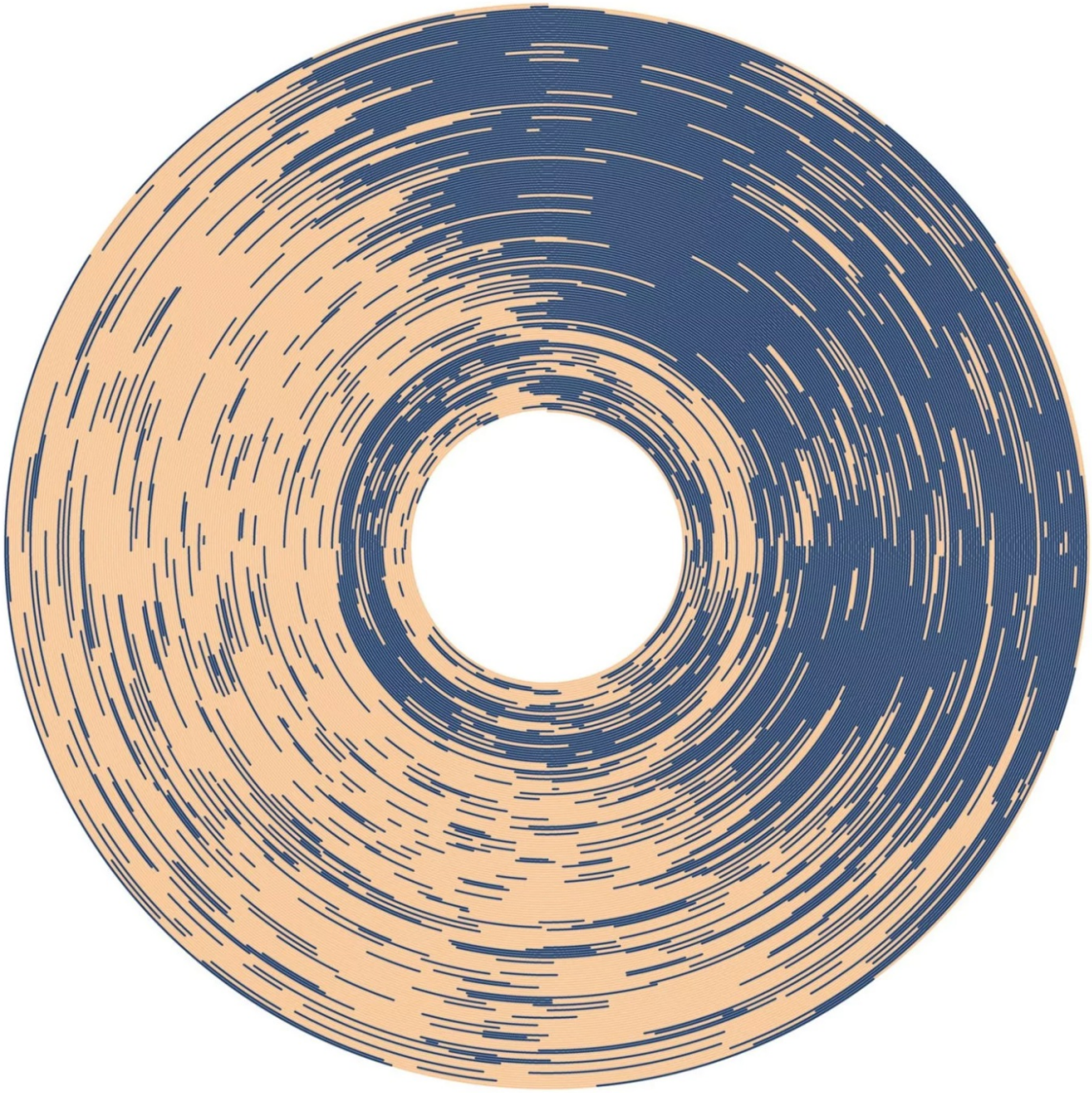
Here is what simply the raw data would have looked like [B].

```
dat <- read.csv("US_births_1994-2003_CDC_NCHS.csv")
head(dat)
```

```
##   year month date_of_month day_of_week births
## 1 1994     1             1           6   8096
## 2 1994     1             2           7   7772
## 3 1994     1             3           1  10142
## 4 1994     1             4           2  11248
## 5 1994     1             5           3  11053
## 6 1994     1             6           4  11406
```

Clearly, the method of visualization helped make the insight found here.

The second reason is a bit less thought of: data can be beautiful. And data visualization is the method by which we extract that beauty from our multitude of rows. Allow me to demonstrate what I mean.



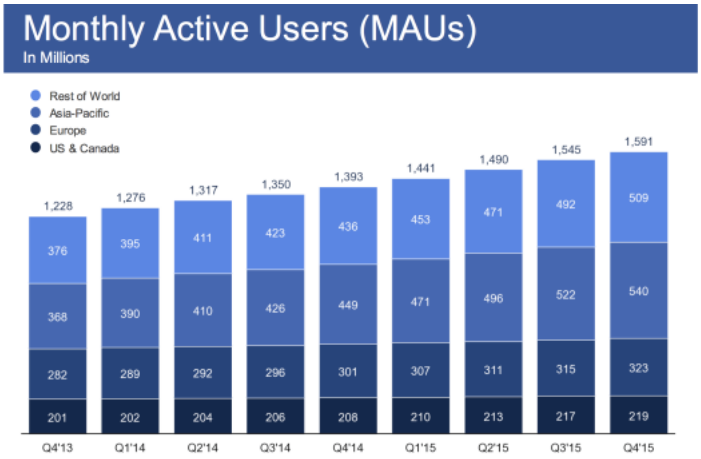
Baby sleep cycle visualization (see reference C).

Above is pictured what *The Washington Post* calls “the most beautiful data visualization of all time, according to Reddit.” [C] Created by an Australian Redditor who goes by the handle “andrew_elliott,” it is a visualization of the creator’s newborn daughter’s sleeping patterns. Although on first glance, it is just a group of circles, we are able to glean beauty from the visualization through both the aesthetically pleasing mix of colors as well as through the underlying beauty in the understanding that the representation is that of a newly awoken life taking in its first experiences of the world. The beauty is breathtaking, and many seem to agree.

Why Is Data Visualization Important?

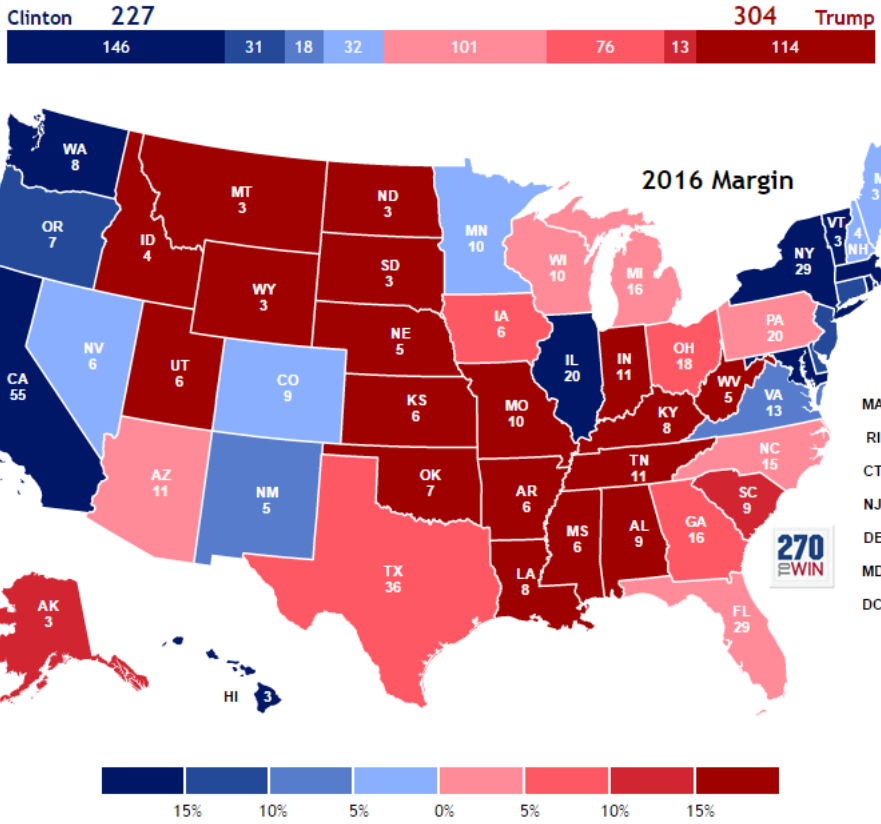
Data visualization is a key skill of a data scientist for one critical reason: it allows the wielder of the skill to relay important information is a useful and rapidly absorbable way. Beyond just professional data analysts, this ability is key for businesspeople, journalists, politicians, and anyone who at any point in their career will need to convince others of a claim using data. The fact that the skill goes underused in data analysis is a testament to the potential that data analysts have to reach new levels in educating others in what they have found. This is important both on micro and macro levels.

Let’s start on the micro level. Suppose we have a data analyst who has moved into a business position where crunching out numbers is not as key. Even then, having the ability to demonstrate what the former analyst has accomplished and what they hope to accomplish will be a key signal to their manager.



An example of a data visualization used in a business setting (see reference D).

A macro level example would include the ability for journalists to show the public exactly which states went red and which went blue during a presidential election. Conveying that level of complex information in any way other than data visualization would be wildly inefficient.



An example of a data visualization used in a mass public setting (see reference E).

Conclusion

In this post, I have gone over what we can do with data visualization and why it is important. The big take-home message is this: regardless of if you want to end up in a computational field or not, the ability to visualize data is absolutely critical for a myriad of reasons, ranging from providing useful information to your boss to graphing out the sleep cycle of a newborn to informing the public of key election outcomes. Data visualization is an underrated skill and the information will be far better off when many more individuals invest the time to communicate their data in more useful and beautiful ways through data visualization.

Further Reading

Of course, data visualization is certainly not the end of the road. Data visualization inevitably encounters ideas and methodologies like design and the iterative process. If this has interested you at all, please take a look at these data visualization best practices [F] and data visualization tools [G].

References

[A] https://fivethirtyeight.com/features/some-people-are-too-superstitious-to-have-a-baby-on-friday-the-13th/

[B] https://github.com/fivethirtyeight/data/blob/master/births/US_births_1994-2003_CDC_NCHS.csv

[C] https://www.washingtonpost.com/news/wonk/wp/2017/01/05/what-its-like-to-sleep-like-a-baby-visualized-by-a-dad/?utm_term=.0c419d10c048

[D] https://techcrunch.com/2016/01/27/facebook-earnings-q4-2015/

[E] https://www.270towin.com/2016_Election/

[F] https://unilytics.com/top-7-data-visualization-best-practices/

[G] https://thenextweb.com/dd/2015/04/21/the-14-best-data-visualization-tools/