Hi, I'm Nathan Yau from FlowingData, and I help people understand data through visualization. Ask me anything.

flowingD [1] and r/Science AMAs[1]

[1]Affiliation not available

April 17, 2023

Abstract

Hi everyone, Nathan Yau here. I run FlowingData, a blog on visualization, statistics, and information design. I started it on a whim as a statistics graduate student, but now it's my full-time job. My PhD research was on how visualization could help non-experts understand their personal data better, and that spilled over to more general sorts of visualization. I've written two books, Visualize This and Data Points, and I write a lot of practical how-tos. I also work on random data projects, some more traditional and others more experimental. Recently, I remade the Statistical Atlas of the United States from 1870 with modern data, brewed beer based on county demographics, and illustrated famous movie quotes as charts. Here's proof that it's me. I'll be back at 1:30 PM ET to answer your questions. Ask Me Anything! Update: Away we go. Update: And still going. I'll answer as many more as I can before I break for lunch. You know those Snickers commercials with the cranky, hungry celebrities? Those are about me. Update: Calling it. Thanks for all the questions, everyone. It was fun.

Hi, I'm Nathan Yau from FlowingData, and I help people understand data through visualization. Ask me anything.

FLOWINGD [R/SCIENCE](#)

ABSTRACT

Hi everyone, Nathan Yau here.
I run [FlowingData](#), a blog on visualization, statistics, and information design. I started it on a whim as a statistics graduate student, but now it's my full-time job. My PhD research was on how visualization could help non-experts understand their personal data better, and that spilled over to more general sorts of visualization.
I've written [two books](#), *Visualize This* and *Data Points*, and I write a lot of practical [how-tos](#). I also work on random data projects, some more traditional and others more experimental. Recently, I remade the [Statistical Atlas of the United States](#) from 1870 with modern data, [brewed beer](#) based on county demographics, and [illustrated famous movie quotes as charts](#).
[Here's proof that it's me.](#)
I'll be back at 1:30 PM ET to answer your questions.
Ask Me Anything!
Update: Away we go.
Update: And still going. I'll answer as many more as I can before I break for lunch. You know those Snickers commercials with the cranky, hungry celebrities? Those are about me.
Update: Calling it. Thanks for all the questions, everyone. It was fun.

[👁 **READ REVIEWS**](#)

[✏ **WRITE A REVIEW**](#)

I'm sure we'll see a lot of questions about visualization tools and whatnot today, but I have a deeper question for you: Where do you get your inspiration for a new data visualization? Where do you get your ideas, where do you find the data to implement those ideas, and how do you know when you've come across a good idea for a data visualization? If you ask me, this is one of the most important skills for any visual journalist, yet it's so rarely talked about.

[rhiever](#)
I follow your blog. Maybe you should be answering this question.

For me, I tack on "Could this be answered with data?" with a lot of my curiosities. If not, could it at least be informative?

For example, before my son was born, my wife and I had to pick a name. That led me to a punch of digging into the name data from the Social Security Administration, like this:

http://flowingdata.com/2013/09/25/the-most-unisex-names-in-us-history/

Initial explorations often aren't fruitful, but then the questions that branch off that initial jump seem to be pretty interesting.

Can you remember a time where the use of statistics dramatically changed your opinion on something? A scenario where the stats disproved many of your preconceived notions about a

topic?

rhiever

Two come immediately to mind, and both were during the early part of graduate school, when I was really learning the depths of data.

I got a class assignment to look at a dataset from a study that was published in a prominent scientific journal. The prof just told us to analyze the set that week, write up what we found, and then compare it to the results of the article. Basically, the data didn't support the conclusions even remotely. Up until then, I always thought of data and statistics as this really hard and concrete thing. Facts. I realized it was much more open for interpretation and based on experience. I think that feeds into how I approach visualization.

The second. So like I said, my dissertation was personal data collection. The quantified self and stuff like that. I found that I pee way more often than I thought and poo much less often that I thought. DATA.

What is your favorite statistical anomaly?

rhiever
My son.

Hey Nathan,

What would be your go to starting point for someone looking to break out of their standard bar / pie chart visualisation into something more complex?

I work with lots of data in my job (digital marketing / web analysis) and have been looking to do more visualisation work. Currently I'm mainly creating charts and graphs using pages and excel, but I've always wanted to move into more diverse and complex methods of displaying data. So far I've dabbled in a little bit of d3 but am far from competent in js. Thanks!

Volny

Find work you like (there's a ton of great stuff out there), and do your best to mimic it. It will suck at first but you will improve quickly. Do this for the mechanics, and you eventually will develop your own style.

Naturally, I was in the same situation years ago. I only made charts in R for analytical reports, and they looked and read that way. It was just default stuff. Then I interned for the New York Times graphics desk. I had to learn their style quickly and pick up software I hadn't used before, all on a deadline.

Don't overwhelm yourself with super advanced stuff right away though. You have to work up to it, so if you're working with D3.js, learn the basics–the mechanics—and work your way up.

Nathan, thanks for doing this AMA!

As a new(ish) dad myself, I've always been impressed by how much you manage to do. Running FlowingData while finishing your PhD, writing books, and publishing journal articles is a lot on its own - but to do so while balancing family life is super impressive.

What advice would you have for others - especially graduate students - in being so productive? It seems academia is especially challenging in regards to a healthy work-life balance.

Geographist
The books, PhD, and all my academic work was finished before my son was born :). I was finishing up my dissertation and second book when we found out the little guy was on the way. I kicked everything up to full gear to get those things done before my life started to revolved around someone else's

schedule.

These days, it's all about getting work done when I can. If there's downtime, like when my son's taking a nap, I work. I also have three guaranteed full work days per week when he's in daycare. So efficiency I guess is key. When I work, I work. When I'm with my family, I try to keep the phone and computer away.

I don't get to bike, brew, or play with LEGOs nearly as much, but now I find value in other things.

What's your opinion on Edward Tufte?

starfish_warrior
Ha. It's been an evolution.

Like almost everyone these days, I got the Tufte books in the beginnings of learning about visualization. I treated them like sacred texts or something. Then I got to NYT, and it was like, okay, that reading wasn't all that useful for practical purposes.

So it always make me chuckle when I see people quote his books like I did in the early stages. it's a dead giveaway for where you're at in the visualization development program.

I don't Tufte personally, and I've never been to a workshop, but I'd say his books are great as introductory text. Mainly his first one. Just gotta make sure to keep going after that. Make things.

Hey Nathan!
I just wanted to say thank you for creating your.flowingdata!
I use it literally daily ever since I stumbled across it some four and a half years ago.
I'm (mis)using it as an online 'diary' for my morning workout routine and it helped me develop discipline because I always enjoyed having my workout sessions visualized - it gave me a sense of achievement.
Basically, I am fit thanks to you!
So again thank you very much!

the_exiled_one
Awesome. You're using it exactly as I intended it to be (for myself).

Most personal data collection is about improving the self in some way, getting actionable results and insights, etc. I'm more interested in how it ties into the everyday like a diary. That's pretty much what I said in the opening chapter of my dissertation.

What does the data look like 10 or 15 years from now? That's what interests me the most about the quantified self stuff.

Have you ever considered doing a show called "Nathan for Yau"?
http://www.cc.com/shows/nathan-for-you

hoopladude
I have. Right when I saw the show on Netflix. Alas, I don't think I'm cut out for showbiz.

Your dissertation was on personal data collection and how we can use visualization in an everyday context. What are some examples of personal data collection + visualization that you think more people should do? What could they learn or gain from those examples?

rhiever

More everyday formats. Like lists and calendars used a visualization formats with colors or styling. That initial familiar bump is huge to get people moving towards more in depth data exploration.

Beyond your 2 books which are excellent, what would you say the next top 5 resources would be for those wanting to expand their creativity with data visualization?

TheWarDoctor
Thanks!

From a practicing perspective...

visualization with d3.js by Murry functional art by Cairo R graphics by Murrell

Get through all that, and you should be good. Practice after that.

Is there any data that's especially difficult for you to convey meaningfully? Do you only deal with clean data, as in absent of confounding variables?

zod_bitches
Uncertainty. People ask me how to include standard error and confidence intervals a lot, and I still don't have a great answer for them. One problem is that we often try to tack on uncertainty to an existing visualization type, but it ends up confusing and cluttering up the place.

The main problem, though, I think comes from the other side. Most people don't get the concept of uncertainty or distributions, so we have to do extra leg work to help others understand the concept before they can even see it.

what is the best data visualization you have encountered so far?

Helium002
Tough question. I pick my favorites every year, and ranking those is even a challenge, which is why last year I resorted to just putting up an amorphous blob collection of greatness instead of ranking.

Just wanted to say thanks. Flowing Data has been in my RSS feed forever and I love it.

HotKarl_Marx
nice. so nice. thanks for reading.

You like to use R as a visualization tool. Practically speaking, how much potential does it have for everyday users of Excel/PowerPoint/Office? When should one be used over the other?

spilled_fishguts
In the words of Amanda Cox, there's nothing special about R really, other than it is the greatest language in the world.

Getting into R from the click-and-point arena of analysis can be tough, I think. But the jump is worth it for a lot of people, especially those looking to move up in the analysis working world. It seems to be a more common job requirement.

More important though, people should develop analysis skills. Learn how to *really* analyze data, outside of hypothesis tests, bell curves, and robot-computed standard errors.

After that, use the software you want. If you know it well enough, you make it do what you want.

How much money do you make with your blog? How much with consulting or specific work?

yardightsure

I won't go into specifics but I make enough to justify not getting a "real job." I do very little consulting these days, mainly because it typically requires that I travel away from home.

I'd have to do the math, but the breakdown is maybe 45% sponsorship, 45% membership, and the rest from random things.

Outside of your books, which I will purchase now that I know that you and them exist, what resources would you recommend to someone looking to express information visually that may be difficult to comprehend or inefficiently delivered through the written word? To give you an idea of how I''m approaching the subject, I've read the Age of the Image by Steven Apkon and I'm in the middle of reading Resonate by Nancy Duarte. Those are both about the visual presentation as a more effective medium for conveying information. I've also read Thinking Fast & Slow by Daniel Kahneman which provided some insight to what sort of shortcuts the brain takes with visual information and information in general.

zod_bitches

Functional Art by Cairo is a good place to go.

Have you done any experimentation with video and gifs? Would you? What do you think of them as mediums as opposed to the still image?

zod_bitches

I've only done a little bit with gifs and no video. I've done some animation.

I think they're worthwhile mediums to explore further, especially animation for transitions between different views. Like that piece by Gregor Aisch and Amanda Cox. Really good.

There's also that paper by Heer and Robertson about animated transitions.

That said, I still think we can say a lot with static graphics and words.

What do you think the best way is to introduce students(young and college+) into making visualizations outside of excel? Would you introdfuce them to R first? or whatever else.

sarahbotts

Yeah. It seems clear that R is going to be around for a while, so it'll be useful in the long-run. For interactive and the web though, I'd go with D3.js. Start with fun examples to show what's possible and to get the students excited, and they'll take it from there.

In your opinion, what is the most overused visualization, relative to its usefulness?

ihazaredditz

Voronoi all the things.

I guess it's not used a *ton*, but it gets used more than it should because it looks neat. It's sometimes useful with maps, and good for interaction though.

Also, the individual person icons to show counts and take up an oddly large amount of space. Moderation, people.

the WINNOWER

HI, I'M NATHAN YAU FROM FLOWINGDATA, AND I HELP PEOPLE UNDERSTAND DATA THROUGH VISUALIZATION. ASK ME ANYTHING. : **REDDIT**

Hi Nathan,

Thanks for doing this AMA! I am currently a Junior Data Scientist, and in love with statistics and probability. I have asked questions on 4 topics below - but you can answer whatever you feel like, dont want to take up much of your time.

1. **Do you think visualization capabilities will play a key role / be a key hurdle in making sentient machines? <Because in my opinion we humans derive so much out of visualizations, we dont just use it as sensors to avoid obstacles, but also to find patterns in things and derive conclusions from them.> Also what is your thought in general on the future of A.I?)**

2. **You mentioned you were a stats grad student, and later you did a PhD. How did you get proficient in programming? What are some of the tools you admire (have used them or plan to use them in future)? And what are some tips you can give to stats grad students currently?**

3. **What are your ideas on representing / visualizing high dimensional data? For example if we think about curse of dimensionality and k nearest neighbors, even a small percentage of similar data gets scattered far away. So if we want to look for multiple features in a high dimensional space - can such problems be visualized efficiently?**

4. **From your educational and professional experience, what innovation do you think is required in statistics? What questions are unanswered in this field? What one way do you think this field can be different than it is currently?**

Whoa thats too many questions. It'll be awesome if you can answer any of them!

bwwaahhaahaa

1. Um, yes? Wait, no, I take it back. It'll be the statistics that make that happen. Statistics. Then visualization understanding.

2. I majored in electrical engineering and computer science, so the programming experience was kind of there. The weird thing is that I left CS to get away from programming but now I do it all the time (and it's fun). For current grad students, learn to read documentation. It will take you places.

3. Subset the heck out of it.

4. Hm, innovation? In some ways I'm an outsider looking in, but it always feels like stat is falling behind in tech. Not understanding how to use computers quite well enough.

Andrew Gelman has frequently commented on "bad" visualizations that would include many of the types of things frequently found on this sub. Basically, his argument is that many of these are good in the sense that they make people think about numbers they may have previously ignored, but can be bad for many technical reasons. I think there's truth to his argument that much of the appeal of these visuals is the "puzzle" effect--the satisfaction of deciphering them.

What kind of questions do you ask yourself about a visual to strike a balance between technical precision and visual appeal?

all_your_bayes

Gelman and I seem to disagree on many things. He's written a few papers on it.

My main thing is that for you, the maker, to understand that data as deeply and as detailed as you can. That interestingness comes across in the visualization.

Has there been anything you've found particularly difficult to visualise?

sweetchilichicken
Uncertainty and all things related to that.

Hi Nathan! Much thanks for doing this AMA! I was wondering about your thoughts on how data viz has changed since you started your blog in 2007. Any recent developments that you're particularly excited about?

tungs
Sooo much better now. It was about big flashy spammy initially. People put more statistical thought into it these days. Or rather, there are more people with statistical knowledge who work on or collaborate on interesting graphics and interactives.

Hey there! Thanks for doing this ama.

I've gotten really interested in data analysis, even going to far as interning at a company for nine months to gain experience. How could this turn into a career for someone who likes this kind of stuff?

Bonus question: what is the most beautiful/pretty/significant visualization you have seen or made?

Brayzure
Lots and lots and lots of job opportunities for people who know their schtuff. Anywhere analyzing data – research, tech companies, journalism – can use someone who knows visualization, and they're pretty aggressively searching.

I wrote a short thing on this way back in 2008 (I feel old now.). Still applies.

I'm so excited to see this AMA here, Nathan. I've been following your work since about 2011 and love flowingdata. It's a huge inspiration as someone obsessed with clean visualizations. I work with a ton of data via GIS and excel and take pride in my graphs and work so thank you for making data so *accessible* for people. In my opinion you're working towards revolutionizing how data is accepted and used in our lives.

For my question; how do you normally tackle *bad* data visualization? Or rather, what do you think most people do in error when creating their own data sets and how do you normally work towards correcting them?

derpaderp1
Thanks!

Generally speaking, I think people go with default options too much, don't iterate enough, and don't take the time to analyze and understand their data before publishing.

Have you been able to conquer the map-territory paradox when it comes to relaying data? That is to ask, have you found that conveying data visually has resulting in a loss of nuance, information, context, or resulted in misunderstandings a significant portion of the time? Are you doing any tracking on that?

zod_bitches

I might be misunderstanding your question, but my thought is that visualization is a complement to traditional analysis. One informs the other. So it always make me kind of uncomfortable to see visualization treated as the end-all cure-all. See something in the visualization? Go back to the numbers and analyze. Find something interesting in the analyze. Go look at the visual for verification for explore the details further.

Hi Nathan,

***Visualize This* was a recommended read in a class I took on SAS Visual Analytics! It was a lot of fun to read. After reading your book I got the sense that dataviz is currently more art than science, even with all the tools available right now in software, just because a lot of these tools are so new.**

I was wondering if you think that visualization is heading into a more "scientific" path recently, whereby users can follow specific guidance or learn a best-practice kind of procedure in order to make the most effective visualizations. As someone being asked by my job to develop visualizations with really rudimentary tools like Microsoft Excel's charts because my company won't buy other software, I'm really hoping there's some way to figure out how to do this and eliminate a lot of guesswork.

[caulfield45](#)

Oh for sure. I mean there's a whole research side to visualization. People meet every year at VisWeek to talk about best colors, angles, sizes, shapes, annotation, and animation to use, etc.

Check out Martin Wattenberg and Fernanda Viegas' work. They're an excellent bridge between practice and research.

I guess you would agree that we have not only to train journalists on how to visualise data and convey statistical information but also train people on how to be critical when presented data. In that context, how would you describe the current state of education in terms of data visualisation/understanding? And how do you think it could be improved?

[PhJulien](#)

Yeah, data literacy from all angles could use improvement. If people are more familiar with *data* and *statistics*, the path to visualization understanding is much shorter and easier to travel.

Have you thought about a serious side project analyzing politics through visualizations? Nate Silver made his name mainstream by applying statistical analysis to election cycles (though FiveThirtyEight is far less objective in its apporach now than it used to be). I have to think that there are a lot of political facts and figures that would resonate for readers if only they could be visualized.

[CarrollQuigley](#)

Visualization-wise, I feel politics (especially election) is best left to the big news groups. The New York Times in particular does *great* work.

What are you biggest no-no's when creating a data vis... And what are your best quick go-to's?

Cheers

[Stats_Sexy](#)

Boom, wrote about this last week:

http://flowingdata.com/2015/08/11/real-chart-rules-to-follow/

Thank you for doing this!

Could you give me some career advice?

I have an intense passion for visualization and presentation. I am a BI analyst and would love to move to the data science side of data. I don't have a degree and cannot code (yet). Is a degree 100% necessary for data science?

Thanks again!

icameforthemusic
It depends on the area, I guess. But with the internet, a lot more is possible now. I'd check out the John Hopkins Data Science track on coursera (I think). The group of profs who run that are top notch.

Just wanted to say I have your book, Data Points, and loved it. It gave structure to something that is largely in the realm of art and I use it at work often

shortcake_minus_cake
thanks so much

Hi Nathan, I work mostly with R for data visualization. Was hoping to ask two questions today:

1) So, a lot of the figures in scientific publications are shit. Could you highlight what you see as some common problems with data visualization in science suggest some tools or ideas that would help improve this?

2) I'm very interested in creating more interactive figures for my work (and for fun). For example, I've been using web-based tools like cartodb, plotly and even making animated gifs. What are your preferred tools to produce iterative plots for the web? Do you use e.g., d3 or Bokeh? Thanks!

jiujitsulab
1. Number one tip is to get off the default train. In R you can customize everything, andit's easy to do (especially since you're working with R already).

2. d3.js for interactive. Can't go wrong with it. Great community and lots of examples to work from.

You should really meet Mike Bostock, creator of D3JS (see below). The two of you would end up improving each others work, and probably make something jaw-droppingly awesome.

D3JS is a charting library written in javascript. Some examples here :

https://github.com/mbostock/d3/wiki/Gallery

Ob101010
haha. Bostock is legend. I am mere mortal.

Do you have any plans on coming up with new visualization techniques using emerging VR technologies such as Oculus Rift or Samsung Gear VR?

Hexorg
I'll leave that to Aaron Koblin and his crew.

Hi,

Thanks for AMA. I have 3 questions.

**1) I am new to data visualization but not new to data analysis which I do in Stata and sometimes
in R or Python. Would you recommend that I learn Processing or D3.js? 2) Are there any books
on aesthetics in data visualization that you would recommend? 3) Are there any good free / very
inexpensive online courses on data visualization that you think are worthwhile?**

Thanks!

polished_iconoclast
For the web? D3.js. If not, a crapshoot.

You get access to a four-week course on visualization in R with FlowingData membership. I heard it's
pretty awesome. https://flowingdata.com/membership/

**Hi Nathan. I've been interested in learning R and data visualization for a while, but finding time
is difficult. I plan to ask my boss soon for dedicated hours to teach myself R (or maybe Stata). I
work in the health research field on a project that will start receiving data soon. We have a
statistician who will do his job, but I don't think he will 'make it pretty' so to speak.**

**How can I convince my boss to let me dedicate hours (and pay, your books may be bought!) for
my training?**

Thanks!

kylecajones
It'll more than pay for itself once you've learned R. Your work will be better, faster, bigger, and make
your boss look good.

**What is your view on the "data is" vs. "data are" usage debate (i.e. whether "data" should be
treated as singular or plural)? Do you think this debate will settle down anytime soon?**

meltingintoice
Semantics. It doesn't change the analysis or visualization.

Hi, Nathan

**I've been interested in data vis for the last four years, reading you, Cairo, Tufte, Few, etc.
Currently I'm working as the "infographics guy" in a market research company, but contrary to
what anybody might think, I cannot really apply any of the principles and knowledge of data vis.
I'm dictated what to do by either the client or the boss, meaning the type of charts to use (yeah,
lots and lots of pie charts, they just cannot get enough of them), the colors to apply, the
number of points/categories to show, cutting out the y axis in column charts to amplify the
differences, and some more terrible things.**

**This happens because society in general lacks a minimum understanding about data vis,
specially in market research business, but since that is not going to change in the near future
and leaving the company is not an option, what do you recommend me and people like me to
do? I'm sure we are quite a lot.**

Thanks!

mikelowski

I'm familiar with that feeling. Incremental change. All those little things add up, and no one will be the wiser.

Hi. What's your feeling on Hans Rosling and Gapminder?

comment_moderately
Amazing presenter.

What's your favorite reporting tool?

For someone whose inspiring to be a data scientist, what would you advise them? Learn Python and R? Get a masters in Databases?

huginnatwork
R all day. Learn statistics. Have a beer and relax.

Hi Nathan,

When I'm looking for new/interesting work, your site is one of the first I check. It's a great hub. Do you have a list of go to sites that you pull from, or do you just keep your eyes open and stumble onto stuff?

MildRedSalsa
I do. Slightly dated, but still valid mostly.

What do you do when you get stuck on a problem? How do you get around it?

MurphysLab
Lay on the floor, with my face buried in the carpet.

I love that your site isn't covered in ads, yet you are able to do it full time. I take it the membership route has been successful? How else did you try to monetize before ending up with that business model?

Thanks!!

SaltwaterShane
I was a grad student with a meager research assistant salary for the first half of FlowingData's life, so that was sort of a supplement. Honestly, I had a hard time picturing FlowingData as any more than a side project until several years in.

Hi, Nathan. Fan of your projects.

Question: what tools do you use for building your blog, statistical inferences, and data viz?

Thanks

Neocruiser
WordPress for the blog, R for static graphics and analysis, D3.js for interactive web stuff.

Do you have siblings?

If you do, what do they do for a living?

Are you the favorite?

[Blactam](#)
I have two sisters who are way cooler than me.

I just wanted to say thanks for doing what you do. I discovered your blog early on in my career and it was a major influence that led me to become a Data Scientist. Keep up the good work.

[shaggorama](#)
so great to hear.

This is Estevan from DMA|UCLA. Friends with Casey Alt and in his class of 08. Did we have a class together? I don't recall.

[rpeg](#)
Maybe? I took one class with DMA. Database Aesthetics with Mark Hansen.

If you do, what do they do for a living?

Are you the favorite?