

Aleksey Bilogur  
@residentmario

# Learning by Doing with Data Viz Blogging



## David Robinson

*Data Scientist at Stack Overflow, works in R and Python.*

✉ Email

🐦 Twitter

🔗 Github

📖 Stack Overflow

### Subscribe

Subscribe to this blog

### Recommended Blogs

- DataCamp
- R Bloggers
- RStudio Blog
- R4Stats
- Simply Statistics

## Advice to aspiring data scientists: start a blog

Last week I shared a thought on Twitter:



Ironically, this tweet hints at a piece of advice I've given at least 3 *dozen* times, but haven't yet written a post about. I've given this advice to almost every aspiring data scientist who asked me what they should do to find a job: **start a blog, and write about data science.**<sup>1</sup>

What could you write about if you're not yet working as a data scientist? Here are some possible topics (each attached to examples from my own blog):

- Analyses of datasets you find interesting ([example](#), [example](#))
- Intuitive explanations of concepts you've recently mastered ([example](#), [example](#))
- Explorations of how to make a specific piece of code faster ([example](#))
- Announcements about open source projects you've released ([example](#))
- Links to interactive applications you've built ([example](#), [example](#))
- Sharing a writeup of conferences or meetups you've attended ([example](#))
- Expressing opinions about data science or educational practice ([example](#))

In a future post I'd like to share advice about the **how** of data science blogging (such as how to choose a topic, how to structure a post, and how to publish with blogdown). But here I'll focus on the **why**. If you're in the middle of a job search, you're probably very busy (especially if you're currently employed), and blogging is a substantial commitment. So here I'll lay out three reasons that a data science blog is well worth your time.

### Practice analyzing data and communicating about it

If you're hoping to be a data scientist, you're (presumably) not one yet. *A blog is your chance to practice the relevant skills.*

“I’ve given this advice to almost every aspiring data scientist who asked me what they should do to find a job: **start a blog, and write about data science.**”

– William Gibson. Pattern Recognition

[illegible]

<http://residentmar.io>

seancurry1 38 points • 1 year ago

True story.

I've been in NYC about seven years. First four years of that I just worked any odd job I could, and one of those was dog walking for a while.

I'm walking a dog near Columbus Circle (SW corner of Central Park) one day when an obviously touristy couple approaches me. In a thick southern accent, they ask me if I know anywhere to get a cup of coffee.

"You kidding me? You're on the Upper West Side! There's amazing coffee everywhere. My favorite is the Sensuous Bean, 70th and Columbus. About five blocks up."

They looked unsatisfied. "Well, we're really looking more for tea."

"Oh, they've got tea! Plus there's Alice's Tea Cup three blocks up, cute little tea house."

They look at each other, then back to me.

"Well actually, uh... We were just looking for a Starbucks..."

I just stared at them, then walked away. I pointed in a direction - any direction it didn't matter - and said, "Walk two blocks."

Infuriating.

[More comments](#)

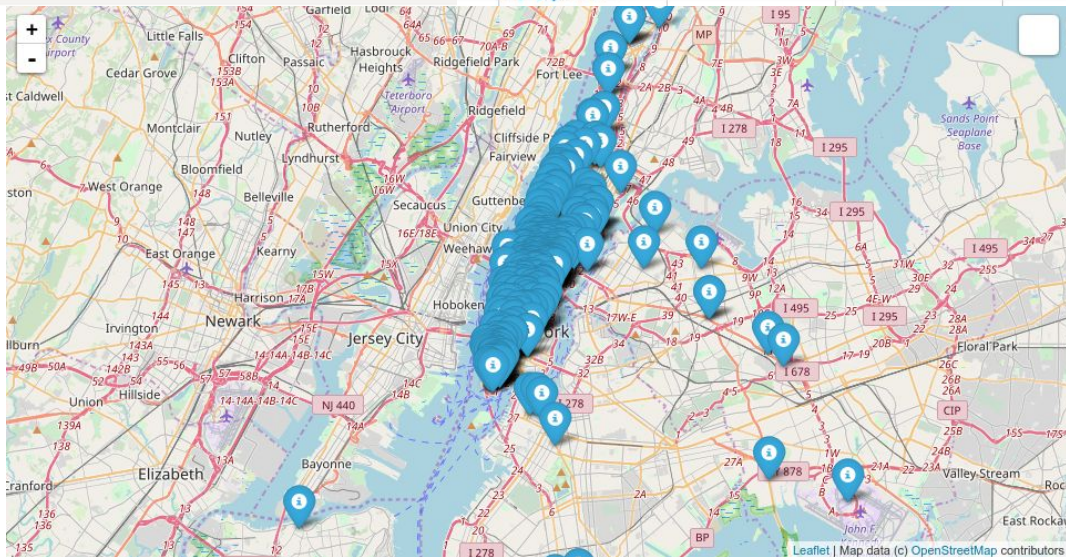
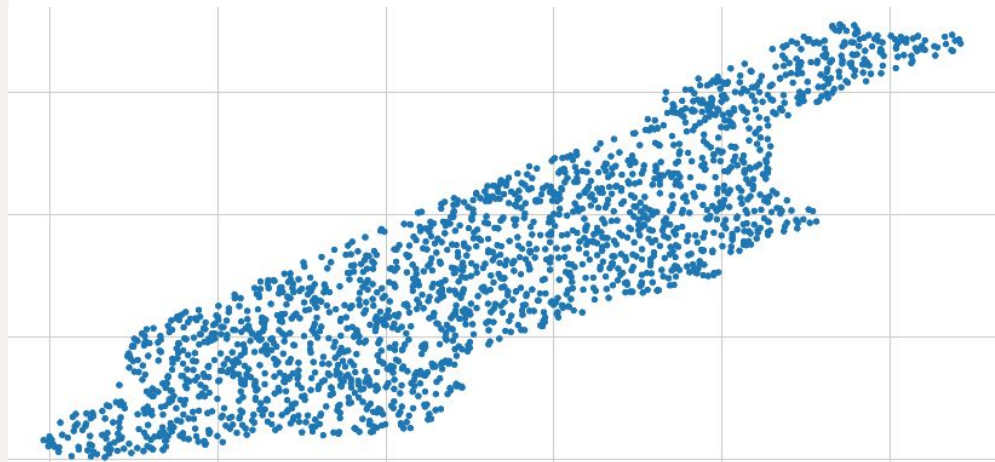


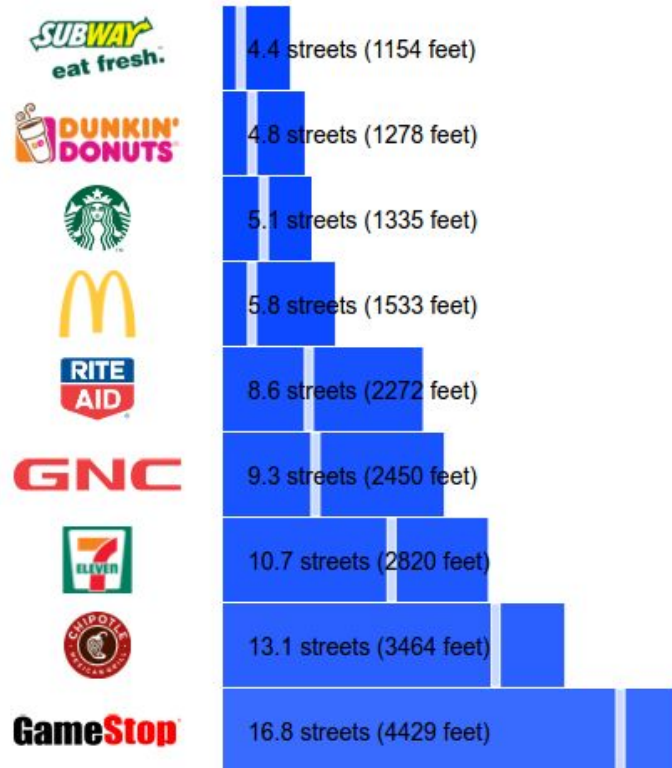
This is a \$9+ cheeseburger from BRGR on 61st and 3rd, posted on [/r/mildlyinfuriating](#). This is why I ignore Yelp skeptics. How are places like this even staying open in midtown?

[r/nyc](#) • 761 points • 281 comments



```
def fetch_businesses(name, area="New York", manual_override=0):
    area = area.lower().replace(' ', '-')
    name = name.lower().replace(' ', '-')
    i = 2
    # Run the first one through by hand.
    try:
        responses = [client.get_business("{}-{}".format(name, area))]
    # This can happen, and did, in the Dunkin' Donuts case.
    except BusinessUnavailable:
        responses = []
    pass
    # The rest are handled by a loop.
    while True:
        bus_id = "{}-{}-{}".format(name, area, i)
        try:
            response = client.get_business(bus_id)
        except BusinessUnavailable:
            # We manually check trouble spots.
            # But see the TODO.
            if requests.get('http://www.yelp.com/biz/' + bus_id).status_code != requests.codes.ok:
                break
            else:
                # Increment the counter but don't include the troubled ID.
                i += 1
                continue
        responses += [response]
        i += 1
    print("Ended 'fetch_businesses()' on:", "http://www.yelp.com/biz/" + bus_id)
    return responses
```





**Question(s)**



**Data**



**Analysis**



**Answer**



**Explanation**



# 5.5 Million Journeys at NYC Bike Share

🕒 2 April 2014 📌 Bike Share

[Tweet](#) [Like 4](#) [Share](#)

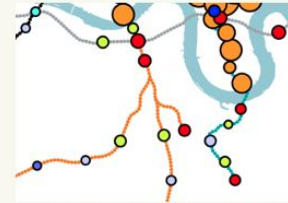
## 5.5 Million Journeys at NYC Bike Share

Based on origin-destination data released by NYC Bike Share for journeys between July 2013 and February 2014. Idealised route assumed, using OpenStreetMap data.



Created by Oliver O'Brien (@oobrien) using Routino and QGIS, at the Centre for Advanced Spatial Analysis at University College London. Further information about the map available at <http://oobrien.com/>

[shop.oobrien.com](http://shop.oobrien.com)



A small shop for my London data art.

### Popular Posts

[Google Maps 2013 – A Few Steps Forward, A Few Steps Back](#)

[named](#)

[Reworking Booth: Geodemographics of Housing](#)

[Tube Tongues](#)

[Bike Share Route Fluxes](#)

[Seeing Red: 15 Ways the Boris Bikes of London Could be Better](#)

### A Google Ad



**Google  
Pixelbook**



## 5.5 Million Journeys at NYC Bike Share

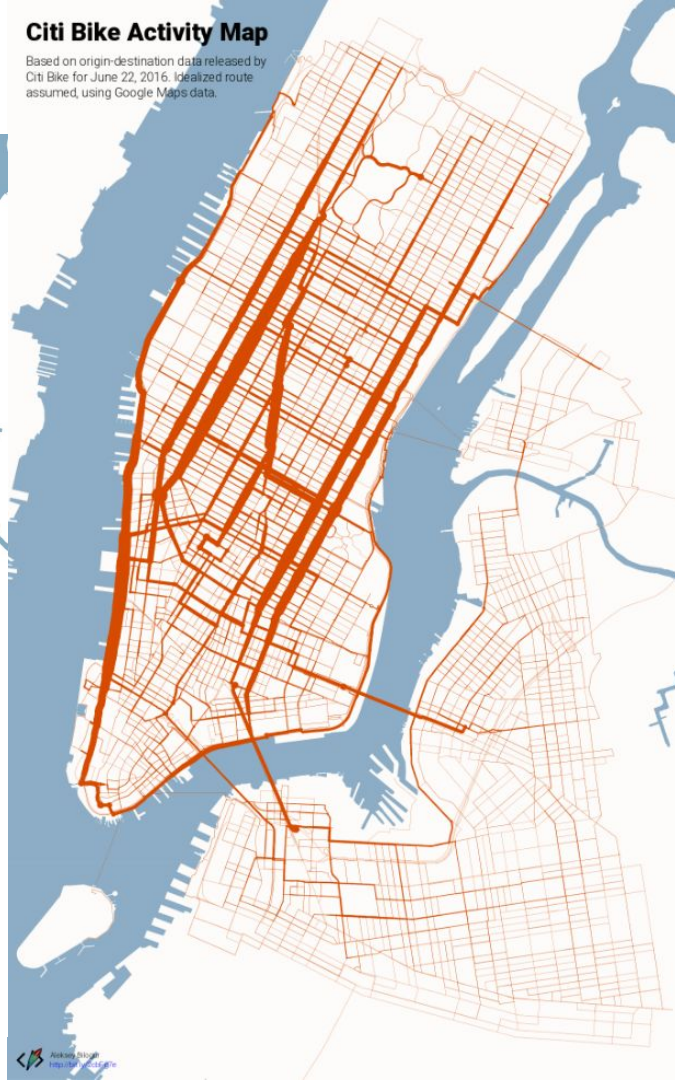
Based on origin-destination data released by NYC Bike Share for journeys between July 2013 and February 2014. Idealised route assumed, using OpenStreetMap data.



Created by Oliver O'Brien (@oobrien) using Routino and QGIS, at the Centre for Advanced Spatial Analysis at University College London. Further information about the map available at <http://oobrien.com/>

## Citi Bike Activity Map

Based on origin-destination data released by Citi Bike for June 22, 2016. Idealized route assumed, using Google Maps data.



**Comprehend**



**Codify**



**Reproduce**



**Present**

# Why?

- Learn skills and techniques.

# Why?

- Learn skills and techniques.
- Learn how to present your work to an audience.



# Why?

- Learn skills and techniques.
- Learn how to present your work to an audience.
- Build a portfolio.

# Further reading

- [Advice to aspiring data scientists: start a blog](#)
- [Building Earth: An Animated Map of Global Weather](#)
- [The New York Times Weather chart redux](#)
- [How to Remake Historical Data Visualization and Why You Should](#)



**David Robinson**

@drob

Following



When you've written the same code 3 times,  
write a function

When you've given the same in-person  
advice 3 times, write a blog post

5:22 PM - 8 Nov 2017

1,339 Retweets 3,853 Likes



45



1.3K



3.9K





**Questions?**