David Robinson, Chief Data Scientist at DataCamp. -- Putting stuff out in public is much more valuable than stuff sitting on your computer waiting to be polished.

Date: 9/17/2019 6:00 pm ~ 1h.

Build your career in data science, MEAP 2019.
Section 4.3: Interview with David Robinson on how he built his portfolio (p 59-61)

David Robinson is Chief Data Scientist at **DataCamp**, an education company for teaching data science through online interactive courses. He co-authored with Julia Silge the tidytext package in R and the O'Reilly book Text Mining with R. He previously worked as a data scientist at Stack Overflow and holds a PhD in Quantitative and Computational Biology from Princeton University. He writes about statistics, data analysis, education, and programming in R on his popular blog, varianceexplained.org

4.3.4 How has your view on the value of public work changed over time? (p. 60) -- add to my blog.html page.

The way I used to view projects is that you made steady progress as you kept working on something. In graduate school, an idea wasn't very worthwhile, but then it became some code, a draft, a finished draft, and finally a published paper. I thought that along the way my work was getting slowly more valuable.

Since then I realized I was thinking about it completely wrong. Anything that is still on your computer, however complete it is, is worthless. **If it's not out there in the world, it's been wasted so far, and anything that's out in the world is much more valuable.**

What made me
realize this is a few papers I developed in graduate school that I never published. I put a lot of work into them, but I kept feeling they weren't quite ready. Years later, I've forgotten what's in them, I can't find them, and they haven't added anything to the world. If along the way I'd written a couple of blog posts, done a couple of tweets, and maybe made a really simple open source package, all of those would have added value along the way.

4.3.5 How do you come up with ideas for your data analysis posts?

My advice is whenever you see the opportunity to analyze data, even if it's not in your current job or you think it might not be interesting to you, take a quick look and see what you can find in just a few minutes.

Pick a dataset, decide on a set amount of time, do all the analyses that you can, and then publish it. It might not be a fully polished post, and you might not find everything you're hoping to find and answer all the questions you wanted to answer. But by setting a goal of one dataset becoming one post you can start getting into this habit.

You should start
by getting very comfortable transforming and visualizing data, programming with a wide variety of packages, and using statistical techniques like hypothesis tests, classification, and regression. It's worth understanding these concepts and getting good at applying them before you start worrying about concepts at the cutting edge.

4.4 Summary
- Having a portfolio of data science projects shared on a GitHub and a blog can help you get a job.

- There are many places you can find good datasets for a side project; the most important thing is **it’s something interesting to you and a little bit unusual**.
- You don’t just have to blog about your side projects; you can also share tutorials or your personal experience with a bootcamp, conference, or online course.