Part 1: ‘Good enough’ data skills

In 2016, The Carpentries–a non-profit organization that teaches coding and data science skills to researchers–published an article titled, “Good Enough Practices for Scientific Computing”.

The article above has tons of great information, but it was aimed at “*researchers who are working alone or with a handful of collaborators on projects lasting a few days to several months*.” We thought the article’s information was too useful *not* to share with as many people as possible. This book essentially attempts to extend the advice in paper to analysts, journalists, grad students, and non-technical audiences.

Why ‘good enough’?

A constant theme that runs through this book is being able to do “good enough” work. We’re going to introduce you to a lot of technology in the next few pages, and we want you to set reasonable expectations for using these tools.

There are many courses, tutorials, and resources that promise you ‘expert training’ in all of the tools we’re covering. You don’t have time for that–you need enough knowledge to see cool stuff and recreate it. In the beginning, you need to focus on doing good enough work, and then sharing your work. The feedback and input you get from developing things in the open are what builds your skillset. Expertise is something you can attain with experience (if that’s your goal), but in many cases, being good enough will get lots of work done.

How to share your work

> *“Your work should speak for itself…” - author unknown*

There is a sea of information on the Internet, and that means everyone is competing for everyone else’s attention. You want to share whatever you’re doing (writing code, building graphs, creating apps, etc.) so it’s discoverable. This way, if a future collaborator, prospective employer, or up-and-coming analyst is looking around for cool stuff on the Internet, they’ll see what you’ve been doing.

Make cool sh!t and share it

We’ll introduce you to a few data scientists and journalists who are excellent communicators of their work. These examples use the Internet as a tool to engage with broader audiences, create better tools for doing science, document some of their daily struggles/successes.

Our first example, Lucy D’Agostino McGowan is a post-doc at Johns Hopkins Bloomberg School of Health. She maintains a blog, publishes ebooks, has online courses, and also attempts to create a real BB–8.

Ricardo Bion is the Data Science manager at Airbnb. He publishes papers on using R in their business setting, gives webinars on how to use modeling to make business decisions, and writes articles on workflow practices that contribute to success in their data science teams.

Or take Amber Thomas, the Senior Journalist-Engineer at the Pudding. She writes tutorials, maintains a blog, and spends her time “hanging upside down in aerial silks.”

All of these people have done two things very well:

1. **Created good work**: All created projects across multiple mediums, websites, and platforms
2. **Shared with as many people as possible:** these examples made their projects discoverable and collaborated with the data science community (by putting their work online for people to find)

Of course, they also had to know their subject areas, and have something worth sharing. But they didn't wait until their work was perfect, or until they were done with their careers to share and get feedback. They started engaging with people while they were working to show how their work gets done.

'Good enough' communication

It's important to remember that whenever you're trying to communicate something, you're convincing your audience that they should be paying attention to you instead of everything else. Today, there's a lot more 'everything else.'

Approaching communication this way puts you in the mind of your audience and keeps you asking, "*why would they want to know this?*" Try to keep the value of what you're saying visible to your audience.

Avoid technical jargon & acronyms

> *"You must learn to talk clearly. The jargon of scientific terminology which rolls off your tongues is mental garbage."*
> *- Martin H. Fischer*

The most substantial barrier to understanding new disciplines or technologies is getting a handle on their jargon. Because this book sits at the intersection of computer science, statistics, and web technologies, all the new vocabulary can often seem like learning a foreign language.

Wherever possible, we'll do our best to clear up or define any terms related to computer science, data management system, web technology, or statistics. To maximize the power of the tools in this text, it will help to know a little about their history, so we'll also cover some background.

Communication takes practice, but it's worth it!

No one is born with an ability to write well–it takes a lot of practice and feedback. The more you communicate with different audiences, the better you'll get at finding an ability to convey why what you have to say is essential.

When we were kids in math class, most teachers required us to show how we got the answers ("show your work"). Teachers told us 'show our work' so they could follow our thought processes through a problem, and see where our thinking was incomplete or mistaken.

If you're regularly sharing your work, people can follow your line of thinking as you progress throughout your project. More importantly, people who find your work will give you feedback and improve your ideas.

Get 'good enough,' then go for more if you need it

What is 'good enough'? We think being 'good enough' means reading about a tool or technology and being capable of distinguishing it from magic. Good enough also means seeing a chart or graph on the Internet and knowing how to evaluate its contents. Or imagining something that matters in your business or personal life, and then devising a way to count it.

Feedback

We sincerely hope you'll find this information useful and give us feedback at `mfrigaard@paradigmdata.io` or `pspangler@paradigmdata.io`.

Footnotes

- check out the Data Carpentry lessons for getting closer to expertise
- The author of the paper above, Greg Wilson, has an excellent blog on teaching technology and all things awesome.