Build Your Skills in Data Analysis!

Roxanne Johnson
john3718@umn.edu
@roxLjohnson
roxjohnson.com

a little zine adventure

About

There are so many cool things you can do to obtain, clean, analyze, explore, and visualize data if you know some programming. I've been a research analyst for about five years, and actively trying to learn Python and R for data analysis for about two years. I've found it extremely challenging and I've met many others with similar experiences. This zine is based on the true story of my own learning experience and aims to highlight some of the challenges I continue to face:

- Not knowing where to start or what the big picture looks like; no clear view of what's possible or a map of how to get there
- The vast number of resources is daunting; it is difficult to find resources that are both at my skill level and relevant to what I want to learn
- Not knowing what I don't know- lacking the awareness that I don't understand a core concept, and not being able to articulate questions or search terms

It's 9am on a Monday and you are late for a meeting. “Who schedules a meeting for 9am on a Monday?” you think to yourself as you grab some coffee. Your colleagues are talking about how the organization wants the public to understand an important new issue. You're half listening and half planning your week when your colleague mentions an existing public data set, which grabs your attention.

As the only research analyst at your organization, it's not every day that you start a project with a nice, clean csv file like this. It means that you can easily open it in Excel and get what you need. “Or “maybe,” you think to yourself, “I can use this opportunity to learn some new tools to make my work easier, better, and less error-prone!”

If you proceed to do the analysis using Excel like you always do, turn to page 2

If you think there might be a better way, turn to page 3

- Many resources are intended for people who are or want to become developers or professional programmers, which isn't me
- I don't feel like I fit anywhere- other researchers may not know about programming, people who regularly work with data in Python sometimes Excel-shame me, and programmers tell me that learning Python is easy

Reflecting on my experience and talking to other data people about their challenges, I decided to create a tool that will help data people start to explore the world of programming for data analysis. The tool starts with an assessment of why you want to learn programming, what your current skills are, and how you like to learn. It will then recommend ways to find relevant resources and direct you to some examples of actual resources.

I am looking for stories of other data people working to learn programming, and also feedback on the development of my tool! If you are interested, let me know!

Excel is up to the task of working with this csv. Using filters, pivot tables, and some formulas, you are able to complete your project.

However, you have to check for errors you may have made and you have to fight with Excel because it did something weird to your data. Then you spend some time trying to make a graph that doesn't look like the ugly Excel defaults, wishing that you could customize it differently.

You got it done and your colleagues think that what you do is magic. As long as you always get small and clean csv files, you've got this!

-End-

Congratulations! You finished your analysis on time, and you learned something along the way! You now have a better understanding of what kinds of tools are out there that can help you do your data analysis work.

You might have learned about what skills are needed to use those tools and how to find the resources to help you actually start learning them. Maybe you've brainstormed some new analysis projects that you couldn't have dreamed up before! And maybe you've connected with others who are looking to learn the same types of things, or who are already doing those things.

In any case, this is probably only the beginning of your adventure in building skills for data analysis!

-End-

Tableau is pretty cool; it allows you to do really great visualizations and the interface is fairly easy to use with your csv. You got your analysis done, made some neat charts, and learned about a new thing!

Unfortunately for the chart you really want to make, you realize you will have to reformat your data. Also, you are using Tableau Public because it's free- if your data was proprietary, you couldn't use that version because it publishes your data to the internet.

If you were empowered by your success with Tableau and want to try learning some programming, turn to page 5

If you are happy with your progress and consider yourself done, turn to page 15

4

You stick with trying to read the documentation for the programming language you want to learn. It takes you a while to figure out how to read the examples that show how to use the arguments of functions. This seems to be written for "real programmers" who can speak this language.

With a lot of messing around, you can get a function to work! Unfortunately, this was only the first step. As you dig more deeply into how to work on your project in this language, you become convinced that for the task you want to do, everyone else uses a specific library that you didn't know about before. This is going to require some significant effort.

Turn to page 11

13

You googled how to do your data analysis problem using a programming language. You get 100,000 ways to do the thing you want to do! How do you decide what's good?

If you try another learning tool, turn to page 5

If you try and wade through the results, turn to page 12

If you give up on using programming for your analysis, turn to page 2

6

You have learned some programming and have made progress towards completing your task using the method you chose. You still have time to keep working with what you have but you know that choice might be difficult and frustrating. You also have time to try other tools, but you're not sure how long those might take to learn either.

If you keep going until you've learned enough to complete your analysis, turn to page 15

To try another learning tool, turn to page 5

If you give up on using programming for your analysis, turn to page 2

11

You have convinced someone who knows more than you to meet with you about your project. At first it was frustrating for both of you; you didn't understand and your mentor couldn't figure out why. Once she realized that you didn't understand a core concept of programming and she could explain it to you, you had a breakthrough and could see what she was talking about! High five!

She also showed you which libraries, modules, and packages are the ones to use for this type of project, and the basics of how to install them and use their functions.

You probably couldn't have come up with this code on your own, but your mentor helped you get what you need to do your task, and you even understand it! (Mostly)

Turn to page 11

You've been using Excel for years. It works great for some things, but lately you've started to suspect that there might be other tools you could use. You've run into some tasks that are tedious in Excel, or just not that easy to do. You think there might be other ways. In fact, some data analysts you have met tell you that they would never let Excel touch their data. But no one in your organization uses anything but Excel.

You have an idea: maybe you could learn some new skills! But you still have to get your project done on time. Maybe there's a user-friendly tool available; you remember hearing something about Tableau. Or maybe you were inspired at a hackathon recently and you've got it in your head you should learn to code!

If you check out Tableau, turn to page 4

If you try to learn some programming, turn to page 5

You try to wade through the google results; you know that it's likely someone has done what you're trying to do. You realize that it may be to your advantage to figure out which sites tend to have useful results, so you spend some time evaluating your hits. Stack Overflow seems good, and there are usually some random people working on something similar to your project but they use strange libraries, packages, and modules that are unfamiliar and confusing. There are a few data bloggers that you decide you like but their posts are only relevant if they are working on the same type of project as you.

Turn to page 11

You have decided to give programming a try. It looks cool, and you would be cool if you could do it! Maybe you could automate tedious tasks, customize every aspect of your visualizations, or quickly restructure data. But where do you start?

If you google it, turn to page 6

If you dive right into the documentation for the language you want to learn, turn to page 7

If you sign up for an online class, turn to page 8

If you join a local user group, turn to page 9

If you attend a hackathon to learn programming, turn to page 10

If you find a personal tutor, turn to page 14

You signed up for a free online class through Coursera. You realize looking at the syllabus that this won't help you get your current analysis done on time so you pull it together in Excel at the last minute.

The first class you try is way too hard and you can't even figure out how to read the data into the program, so you are unable to move forward. The second one you try is way too easy. The third one you try is not about data analysis at all but gives you a great overview of the programming language you want to learn. You only make it halfway through the course because it was really time intensive and you still have to go to your job. Maybe you'll try to finish that class one day! Or maybe you'll go back to learning programming for data analysis to use on your next project.

-End-

You joined a local user group for a programming language you want to learn. You get so excited about the group that you become an organizer! You quickly realize that you won't get your current analysis project done this way and now you're busy organizing, so you pull together your project at the last minute using Excel. At least you got it done.

You spend a year planning three large events and become burned out from organizing. You still haven't learned much programming because you spend most of your time event planning. Many people tell you that your events are really great and that they enjoyed the learning opportunity. You're grateful to now know so many people trying to learn programming for data analysis, and also experts you might be able to ask for help. Part of you wishes that you could have attended these great events instead of planning them, though.

-End-

You're at a hackathon, thinking that maybe if you hang out with programmers that you'll pick something up. Unfortunately, they all seem to know how to work together seamlessly, and they're throwing around words you've never heard. You can't figure out how html, css, and javascript work together on websites and you also don't care because you don't want to be a web developer. You have a GitHub account but you've only done a pull request once and it was with some pretty intense handholding.

They're trying to be inclusive and show you how to get into their IRC channel, but at this point you don't want to slow them down so you offer to do some research, or maybe prepare the presentation for the end. They seem relieved and think that's a great idea. The event was cool and you did participate- you even finally made a cool prezi- but you didn't learn much about programming.

To try another learning tool, turn to page 5

If you give up on using programming for your analysis, turn to page 2

You dove right into the documentation for the language you want to learn. At first glance, it reads like an instruction manual written in a foreign language.

If you try another learning tool, turn to page 5

If you stick with the documentation, turn to page 13

If you give up on using programming for your analysis, turn to page 2