Analytics Resources

Paul Oldham

4 May 2017

At the validation workshop quite a few participants expressed an interest in learning more about the analytics tools. This document provides some resources for self-learning. At a later stage we will hopefully be able to run an analytics training workshop for those interested across the different authorities (and learn from each other as well).

### GitHub

GitHub is a more powerful version of Dropbox that is used for storing and versioning software code, the publication of websites and electronic books. It is a very powerful free tool for storing and collaborating on code and writing. I use it all the time for publishing and hosting quick responsive websites.

Sign up for an account here: <https://github.com/>

Then download GitHub desktop for your machine.

You can copy repositories (folders) to your machine from Github by opening the repository and choosing Clone.

When you have registered and downloaded Github Desktop try Cloning this repository: <https://github.com/poldham/kenya_validation>

You can also raise issues on Github and this is very very close to the internal notification system we discussed at the workshop.

At the link above go to Issues. Click on Create an Issue. Type Hello Paul and then Submit New Issue. That will store the issue and send an email to me.

If you would like to become a collaborator on a project (say the Kenya\_validation folder) then Submit New Issue and request to make a collaborator. As the repo owner I will then add you (but you must have a Github Account).

### The WIPO Manual on Open Source Patent Analytics

The WIPO Manual provides a set of practical walkthroughs for common analytics tasks such as data cleaning, dashboarding and network visualization. The Manual is used in training workshops for patent analytics around the world.

The GitHub home for the training materials is here: <https://github.com/wipo-analytics> The electronic book version of the Manual is here: <https://wipo-analytics.github.io/> You can download any of the repositories (folders) by logging in to Github selecting a repository and choosing CLONE. The files can then be opened in RStudio (below) The chapters you will find interesting are:

1. Tableau – for dashboarding. See this chapter to learn how to do it: <https://wipo-analytics.github.io/tableau-public-2.html> For self-learning you can download any of my workbooks (including data for Kenya) here: <https://public.tableau.com/profile/poldham>.

For other examples try the Gallery <https://public.tableau.com/en-us/s/gallery>

1. Network analysis with Gephi. This chapter needs updating as it uses an older (v8) version of Gephi (now version 9). Version 9 lacked key features but has now been fixed. So try following with version 9 and I will update during May. <https://wipo-analytics.github.io/gephi-1.html>
2. Data Cleaning, You can do a lot of data cleaning with Open Refine. <https://wipo-analytics.github.io/open-refine.html>
3. Infographics Now very popular for reports etc… try this walkthrough: <https://wipo-analytics.github.io/patent-infographics-with-r.html>

My WIPO colleague Irene Kitsara now leads on infographics and is much better than I am at it. Irene’s training presentation is here: <https://github.com/wipo-analytics/presentations/blob/master/Topic_15%20Creating%20Infographics.pdf>

### The Handbook on ABS Monitoring

This is a work in progress using data for Kenya. The Handbook Website is here: <https://poldham.github.io/abs/> The repository for the website is here: <https://github.com/poldham/abs>

The main focus of the Manual so far has been on mapping taxonomic data for Kenya and creating small Apps using the Shiny Framework. There is significant room for improvement but what is there is a start.

### The ABS Permit System Model

The model for the permit system is here: <http://abspermits.net/> The Github repository for the website is here: <https://github.com/poldham/abs_permits>

### RStudio

Everything I do is written in RStudio in rmarkdown and then published on Github directly from RStudio with accompanying R code chunks. Once you have got used to it you can publish websites and electronic books with code chunks, visuals and maps. RStudio rocks.

A quick guide to installing R and RStudio is here: <https://poldham.github.io/abs/gettingstarted.html>

Remember to install R first and the RStudio. If you have RStudio and R installed and download one of my repositories you can open this in RStudio by finding the blue Rproject file in the folder. The files in the folder are written in rmarkdown and all code chunks in them should run. Let me know if they have broken.

Resources for learning to do things in R include

1. Data Camp: <https://www.datacamp.com/>
2. [Coursera R programming course](https://www.coursera.org/specializations/jhu-data-science?utm_source=gg&utm_medium=sem&utm_campaign=r_programming_intl&campaignid=426374217&device=c&keyword=learn%20r%20programming%20online&matchtype=b&network=g&devicemodel=&adpostion=1t1&hide_mobile_promo&gclid=CI2QwImZ1tMCFQ0R0wodHXIBRg) (individual courses are free, specializations cost)
3. R-bloggers is great for walkthroughs: <https://www.r-bloggers.com/>
4. [Rstudio](https://www.rstudio.com/) Webinars, cheatsheets and online learning under Resources.