CS35 Final Project Reflection

Alex Xiaotong Gui

May 2019

1 Overview

For my final project, I created an interactive visualization platform for African Studies Journal Review Database project, using a Python package Bokeh. The goal is to develop an open source data analysis platform for African Studies researchers and policy stakeholders to gain insights on trends and focuses of research done in the past ten years.

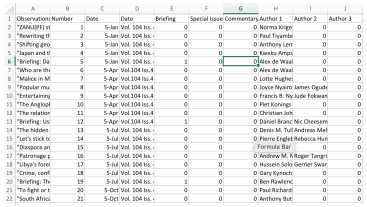
The source code of the project is on github: github: https://github.com/alexaaag.

2 Data Description

The data took a while to clean, since the data is high-dimensional. The categorical topic variables are all binary, so I spent some extra time gathering and aggregating them into data frames I want.

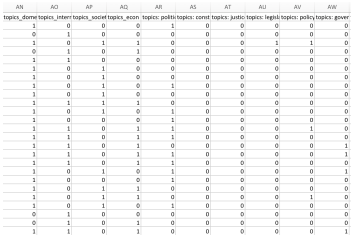
Here is what the data looks like:

For my data analysis, since I only focused on countries, topics and journals, I got rid of all the irrelevant columns, including those containing methodology information.



	A	B	C	D	E	F	G	H	I	J
1	Observation	Number	Date	Date	Briefing	Special Issue	Commentary	Author 1	Author 2	Author 3
2	"ZANU(PF) st	1	5-Jan	Vol. 104 Iss. 1	0	0	0	Norma Kriger	0	0
3	"Rewriting th	2	5-Jan	Vol. 104 Iss. 1	0	0	0	Paul Tiyambe	0	0
4	"Shifting geo	3	5-Jan	Vol. 104 Iss. 1	0	0	0	Anthony Lent	0	0
5	"Japan and t	4	5-Jan	Vol. 104 Iss. 1	0	0	0	Reeiku Ampiz	0	0
6	"Briefing: Da	5	5-Jan	Vol. 104 Iss. 1	1	0	0	Alex de Waal	0	0
7	"Who are thi	6	5-Apr	Vol.104 Iss.4	0	0	0	Alex de Waal	0	0
8	"Malice in M	7	5-Apr	Vol.104 Iss.4	0	0	0	Lotte Hughes	0	0
9	"Popular mu	8	5-Apr	Vol.104 Iss.4	0	0	0	Joyce Nyairo James Ogude		0
10	"Entertaining	9	5-Apr	Vol.104 Iss.4	0	0	0	Francis B. Ny Jude Fokwan		0
11	"The Anglopl	10	5-Apr	Vol.104 Iss.4	0	0	0	Piet Konings	0	0
12	"The relation	11	5-Apr	Vol.104 Iss.4	0	0	0	Christian Joh	0	0
13	"Briefing: Us	12	5-Apr	Vol.104 Iss.4	1	0	0	Daniel Branc Nic Cheesem		0
14	"The hidden	13	5-Jul	Vol. 104 Iss. 3	0	0	0	Denis M. Tull Andreas Mel		0
15	"Let's stick tc	14	5-Jul	Vol. 104 Iss. 3	0	0	0	Pierre Englet Rebecca Hun		0
16	"Diaspora an	15	5-Jul	Vol. 104 Iss. 3	0	0	0	Formula Bar		0
17	"Patronage p	16	5-Jul	Vol. 104 Iss. 3	0	0	0	Andrew M. R Roger Tange		0
18	"Libya's forei	17	5-Jul	Vol. 104 Iss. 3	0	0	0	Hussein Solo Gerrier Swar		0
19	"Crime, conf	18	5-Jul	Vol. 104 Iss. 3	0	0	0	Gary Kynoch	0	0
20	"Briefing: Th	19	5-Jul	Vol. 104 Iss. 3	1	0	0	Ben Rawlenc	0	0
21	"To fight or t	20	5-Oct	Vol. 104 Iss. 4	0	0	0	Paul Richard	0	0
22	"South Africa	21	5-Oct	Vol. 104 Iss. 4	0	0	0	Anthony But	0	0

Figure 1: Data Description, Park I



AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
topics_dome	topics_inter	topics_socie	topics_econ	topics: politi	topics: consi	topics: justic	topics: legisl	topics: policy	topics: gover
1	0	0	0	1	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	1	1	0
0	1	0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0
1	0	1	0	1	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0
1	0	1	0	1	0	0	0	0	0
1	0	0	0	1	0	0	0	0	0
1	1	0	1	1	0	0	0	0	0
1	1	0	1	1	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1
1	1	0	1	1	0	0	0	0	1
1	1	0	1	1	0	0	0	0	0
1	0	1	0	1	0	0	0	0	1
1	0	0	0	1	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0
1	0	1	1	0	0	0	0	1	0
1	0	1	0	1	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	0	0	0
1	1	0	1	0	0	0	0	0	1

Figure 2: Data Description, Park II

3 Final Analysis

In my exploratory analysis, I plotted many things based on countries, author countries, and topics. However, since I encountered many difficulties making Bokeh work, my final visualization included only major topics and major journals. I was able to add interactions, so users can choose what they want to visualize and they can move their mouse to see the exact number of the data.

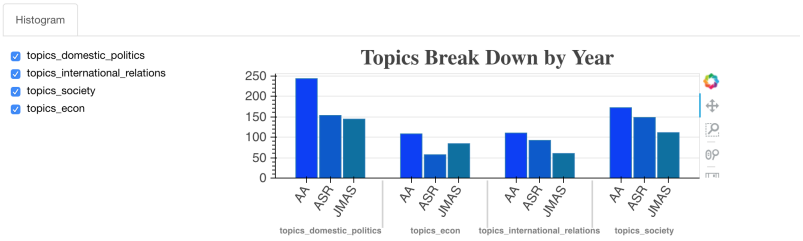


Figure 3: A screenshot of my dashboard

To run the app, go to the folder that contains bokeh_app and run the command bokeh serve –show bokeh_app'.

4 Conclusion

As I mentioned in my own presentation, I started with a very ambitious goal of creating multiple dashboards with a comprehensive range of features to be visualized. Unfortunately I have not met my goal. When I was learning about how to use Bokeh, I found out many shortcomings of the package. Bokeh plotting, compared to seaborn, matplotlib and pandas plotting tools, is fairly complicated to use. The syntax does not feel very intuitive, and the data structure that supports the ploting function is very different from pandas dataframe.

However, I think it's great learning experience for myself since I learned about an entire new package in a couple of weeks and made something out of it. I learned a lot about reading tutorials and documentations, which I believe are very important skills in computing in the future.

Since I spent an insane amount of time on debugging on interactions, I didn't develop as many features as I wanted to. Had I had more time, energy or a partner, I would definitely try to work out more features. But overall I think I did well on figuring out the overall structure. I developed a web-based visualization of topics in major journals and added interactions. I'm pretty happy about what I have achieved.