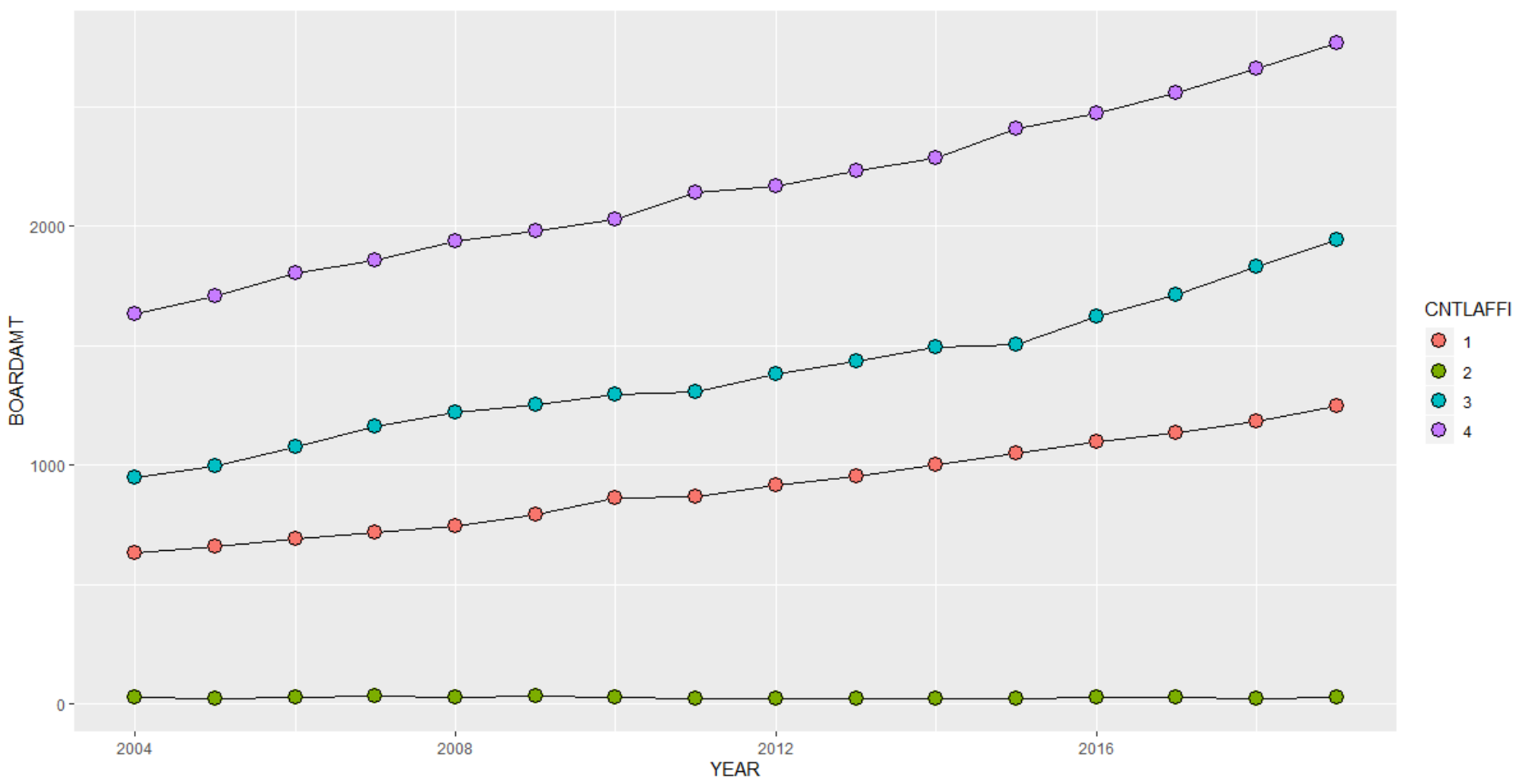
**On Institutional Research**

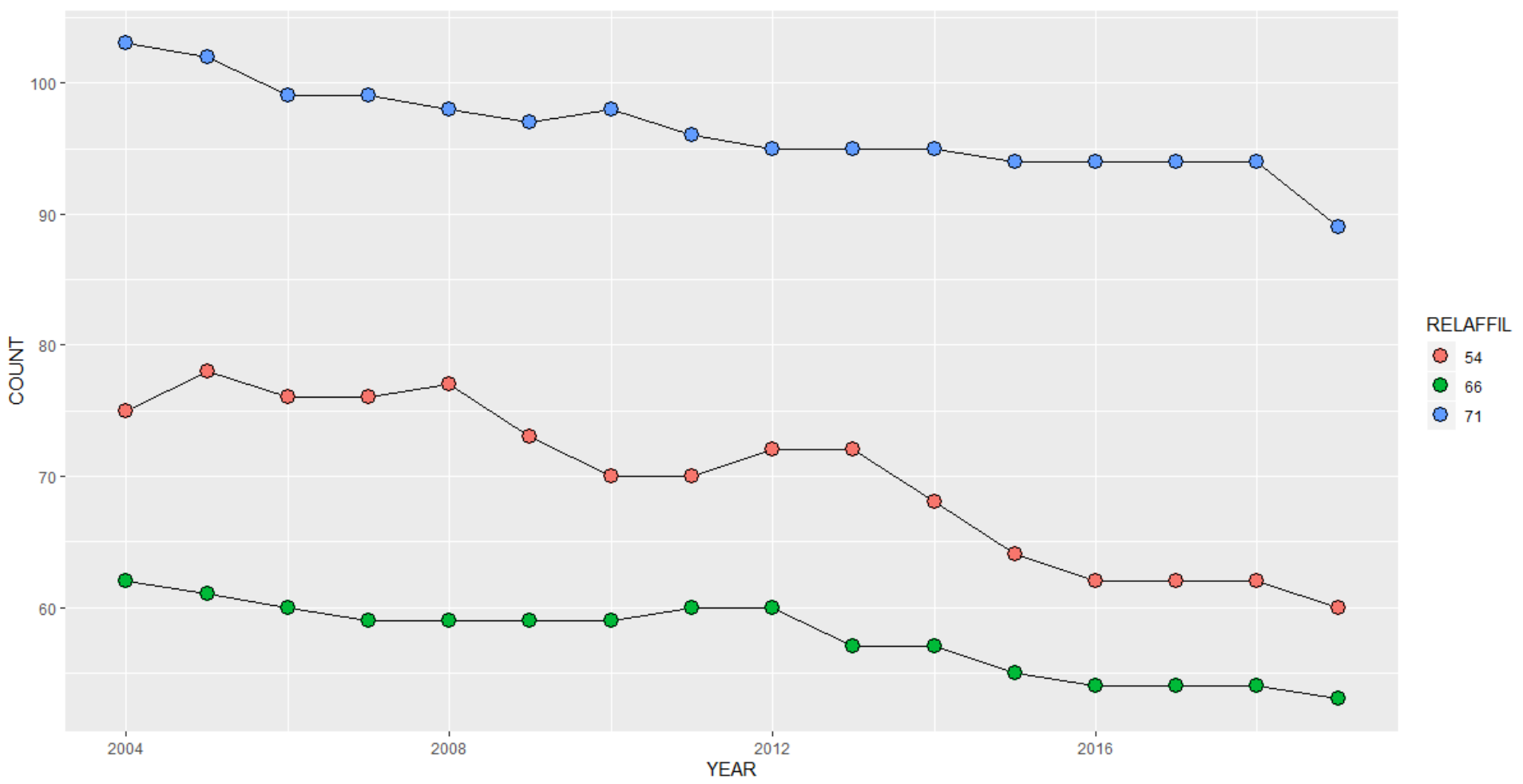
In this brief and hastily written post, I will seek to acquaint the reader with a data source I have recently made use of. Namely the IPEDS database. The field of IR research has been a growing one in recent years. It’s kind of like the overachieving peer of measuring our public-school performance here in America.

In some senses, it is very much about education, extending value, and tracking demographics and audiences. In another, it is about meeting a demand of working professionals, and/or upwardly mobile white-collar families.

That is not to say it is a bad thing. In fact, many of the faculty staffing these institutions are stellar members of society delivering an invaluable service to those benefitting from their expertise. Only IR is like the dirty politics behind it all, the bureaucracy, administration, and profit motives linked to running any business.

In that sense, education is cheapened when looked at through this lens. But it comes with the territory of keeping an organization alive and thriving. After all, many tenured university professors at top 50 colleges are in some of the top earning brackets throughout the USA. It is only reasonable then for them to do their research to maintain this advantage, to some degree, without falling prey to unscrupulous behavior.





The main purpose of this post is to share portions of my workflow.

From graphs above, you will notice I have not put much thought or effort into relabeling my charts currently. However, here is a brief description of how I went above this project.

1. On a good day, I would like to start with a question. On a bad day, I start with a topic.
2. On a great day, I begin with a topic, and questions.
3. So, given that I began with the idea of higher ed research in mind, I navigated to the department of ed’s database portal.
4. After selecting the most generic survey, I decided to loop through it, and see if I could weave together a dataset. (Fun secret, there is a certain pride in putting together a clean dataset itself)
5. Once I had data enough to test out a chart or two, I let myself be surprise with some of the results I saw!

And that is the joy of working with data on the side, and expanding your skillsets, and tooling for your future data science careers. To be honest, I plan to expand on this week’s post. As a rough point, I plan to ask questions such as –

* How has funding for college attendees changed over the years?
* How does that vary by state, income group, and race?
* What are some of the majors that have grown in popularity over the years? How does this differ by gender?
* Importantly, to any education researcher – what are rates of completion currently looking like?
* What factors influence completion rates?
* And of course, the converse of completion rates, how are enrolments varying in recent years?
* Lastly, I would end on the same question that I started with. That of funding – only for universities instead of students. How are endowments for universities changing over the years?

Above we examined the cute and fun trend of religious affiliations in colleges, however, these do not seem to be a major dividing factor. Management – private or public – meanwhile seems to be the major dividing line. In addition to some of the steps outlined above, to my workflow I would like to add:

1. I selected variables of choice by going through the data dictionary and codebook, and formulated my questions accordingly. (ideally, I would be looking for variables after formulating a question 😊)
2. And lastly, I used some ingenuity and string parsing to locate what the university id’s for IPEDS data linked to. That is, which school was represented by which id. And included that in my dataset.

Such enhancing of a dataset is the pride and joy/job of a budding data scientist. Then tacking on any additional insights that can be gleaned from the enhanced dataset.