# CSC475 Fall 2018: Journal of Progress Reports

Mary Donahoo

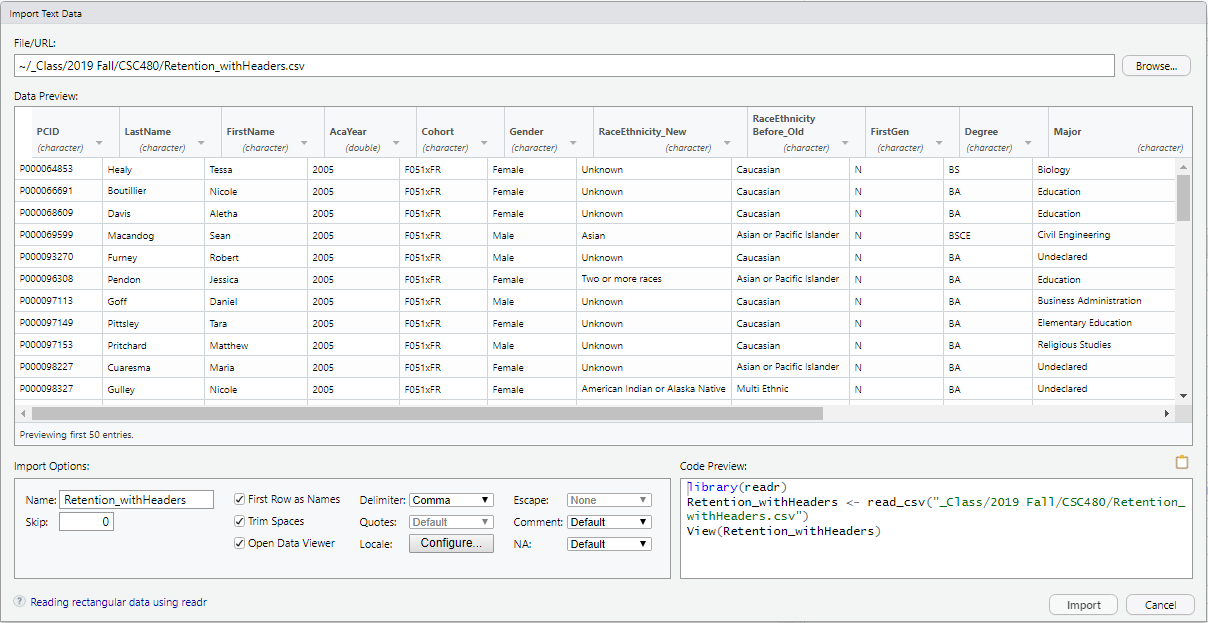
## Progress Report 1

### Team Meetings

Met with Kim on 9/2, 9/4, and 9/9. We shared our ideas and perspectives on what to do with the retention data. Kim leans more toward automating the initial prep of the data while I’d like to implement some kind of interactive visualization of the data in Shiny. I’d also like to build a predictive data model.

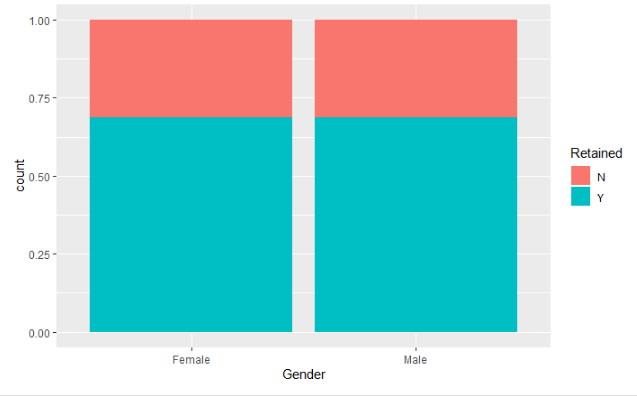
### Data

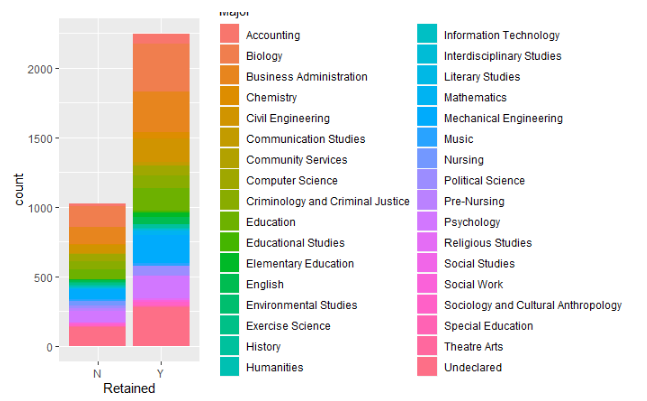
Kim built a query to pull retention data from the PowerCampus database. He is planning to include additional data items in the query. I used the initial query to produce a dataset and import it into R just to get started.



Right now I’m in the process of trying to do some initial data visualization. So far, I’m mostly just playing with the data as a way of trying to get my brain to form an interesting question. I don’t have anything yet. In fact, I’m doing something really wrong right now so I need to figure out what that is. I’m going to try stepping back from the computer and see if I can sketch something meaningful on paper which I can then implement in R.

Below are two examples of useless visualizations I’ve created:





## Progress Report 2

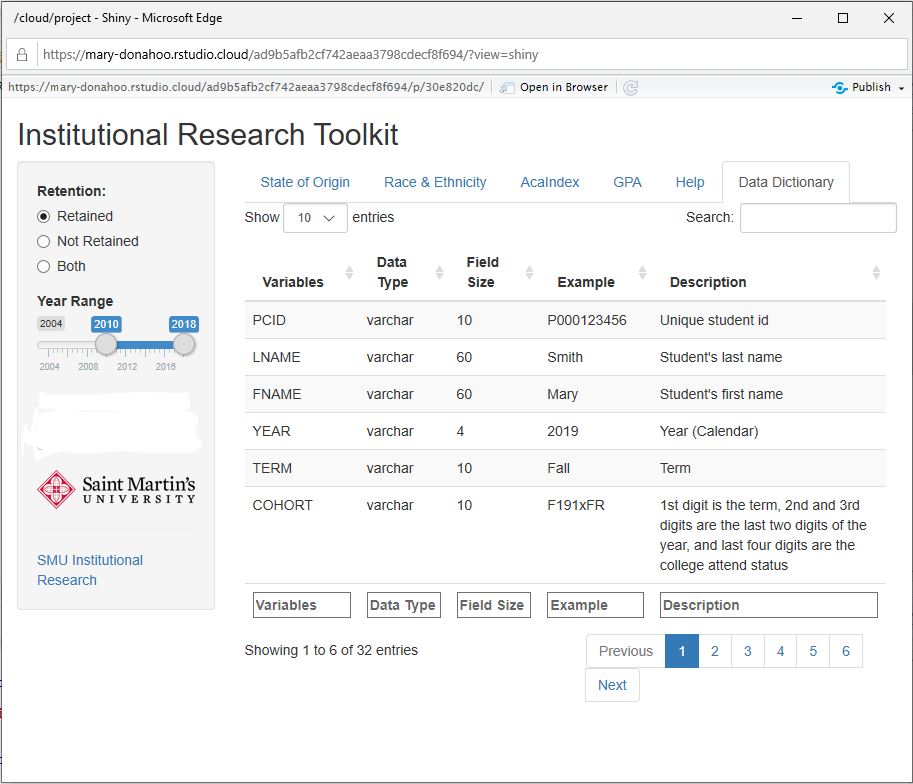
Continued to work on refining and clarifying the focus of our project.  Instead of delivering a completed retention report and/or analysis, we will provide a set of tools, methods and basic examples/templates to assist an IR person with the technical aspects of retrieving, preparing and presenting data in response to specific institutional requests.  Below is a laundry list of toolkit components and high-level to-do’s.  We will continue to refine this list as we add detail to the tasks.

|  |  |  |
| --- | --- | --- |
|  | Item | Description |
| 1. | SQL Script | Produce data set for:   * Import into R * Create a data dictionary that describes the values produced by the SQL query including the longitudinal changes in data (race/ethnicity, SAT revisions) (Kim) |
| 2. | Dataset prep for standard IR reports | *Kim – I only have a vague idea about what’s needed here.  I’m happy to help if I can.  Hopefully we can nail down some specific, discrete steps.*   * Using R Studio in the cloud (Kim) * Describe data types from output (Kim) |
| 3. | R Studio | Setting up an R environment  ·         Installing R and R Studio on a laptop  ·         Using R Studio in the cloud |
| 4. | R import guide | Step by step guide to importing the SQL data set into R |
| 5. | R Pubs | ·         Maybe setup R Pubs account for a generic IR mailbox (@stmartin.edu)?  This would be another option for deploying IR reports and analysis |
| 6. | R Visualizations | ·         Create some separate R scripts for different types of R visualizations.  ·         Links to R resources and cheat sheets |
| 7. | Shiny App | ·         Build a Shiny App with a tab set of some example visualizations  ·         Include instructions for how to make the data and other resources (like images) available to the published Shiny app.  ·         Include the SMU logo and contact information in the left panel – include instructions for updating image and info.  ·         Include a tab for the data dictionary  ·         Include instructions for |
| 8. | Machine Learning | How to build predictive Model in R, including examples/templates.  *If we go ahead with the Toolkit idea, probably won’t have time to do the Machine Learning piece since that would be a whole other set of data prep and how-to’s.* |

## Progress Report 3

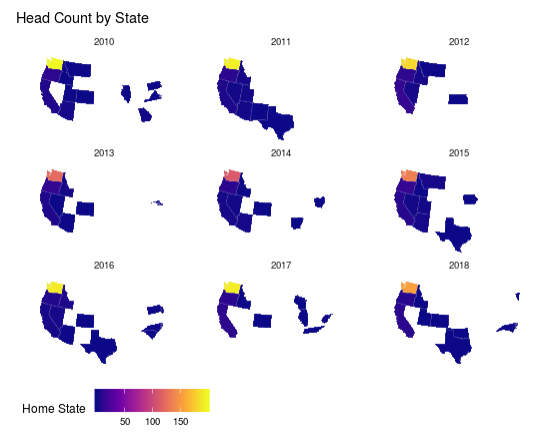
* Continued to refine SQL command to pull retention data from PowerCampus database
* Continued work on visualizations
* Began work on Shiny App:  Implemented tab set; radio button for retained/not retained; slider for cohort year range; data dictionary for retention data set.

The radio button and slider input isn't tied to the visualizations yet since those are still a work in progress.

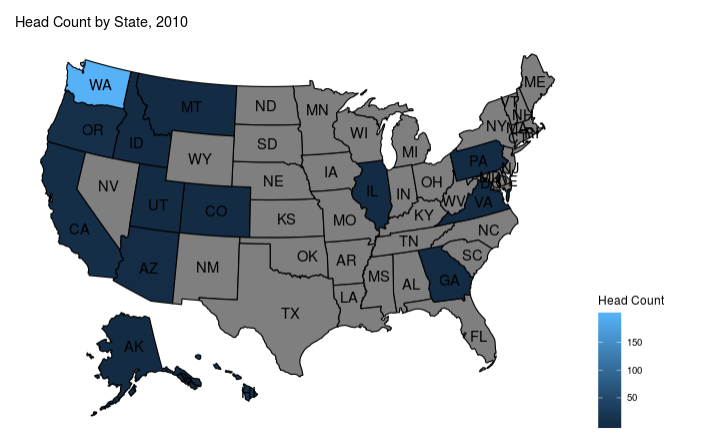


Examples of visualization work in progress

Facet of home State built using map\_data function from maps package but I ended up switching to a different package because the maps package doesn't show Alaska and Hawaii.



Started working with the usmap package since it includes Alaska and Hawaii.

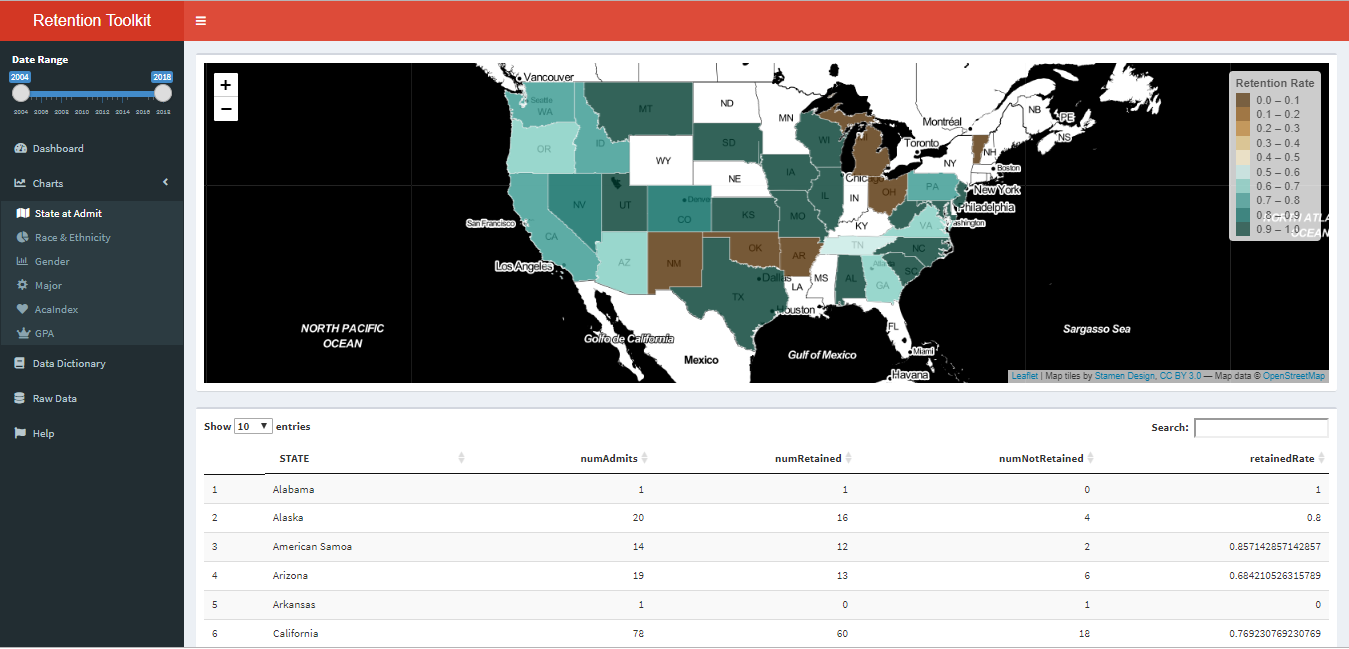


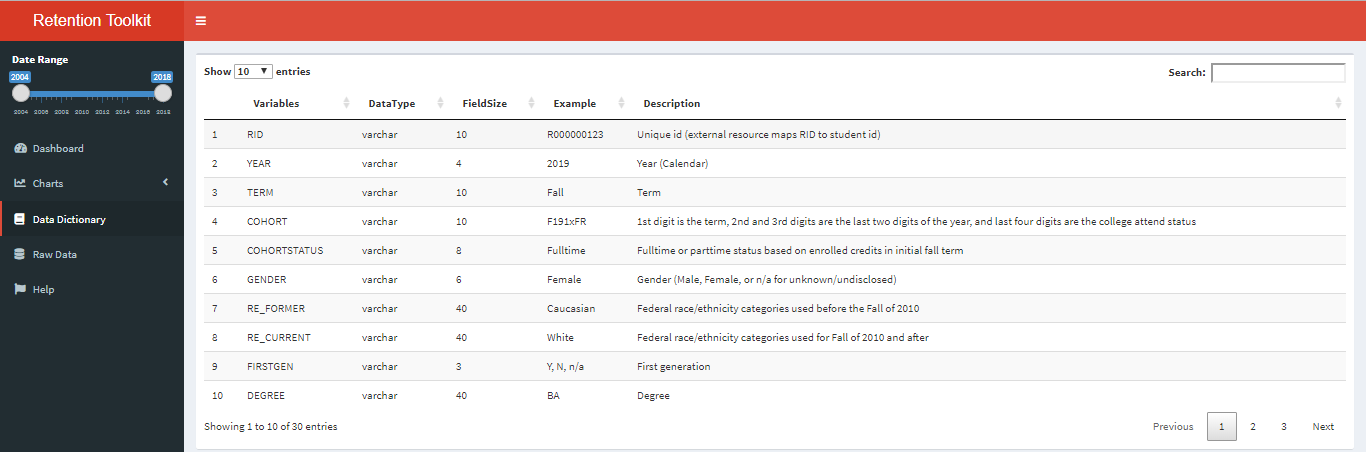
## Progress Report 4

I was running into various issues using RStudio Cloud and the shiny package.  I ended up reinstalling R, R Tools and RStudio on my laptop and then switched over to using the shinydashboard package.  I have published a working version of the app.

<https://mary-donahoo.shinyapps.io/Retention/>







## Progress Report 5

·         Continuing to work on visualizations and tying them into the shiny widgets (for example, select box and slider range)

·         Working on Help documentation

·         Need to resolve issue with raw data columns being cut off.  If you zoom way out, you can see all the columns but the font is very small.  Will look for some kind of mechanism to enable horizontal scrolling or something similar

## Progress Report 6

Continued working on visualizations and published a new version of the Shiny dashboard app:

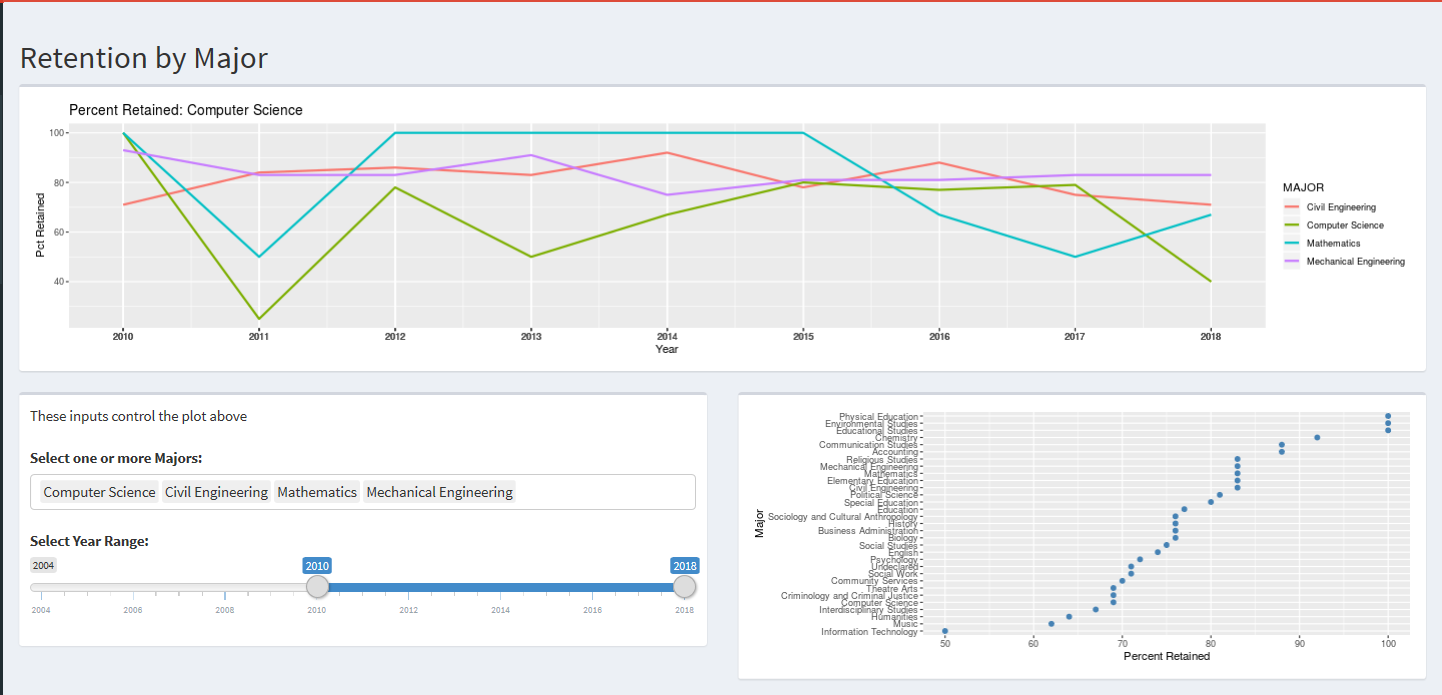
<https://mary-donahoo.shinyapps.io/Retention/>

Retention by Major

·         Modified Select widget to allow for more than Major to be selected

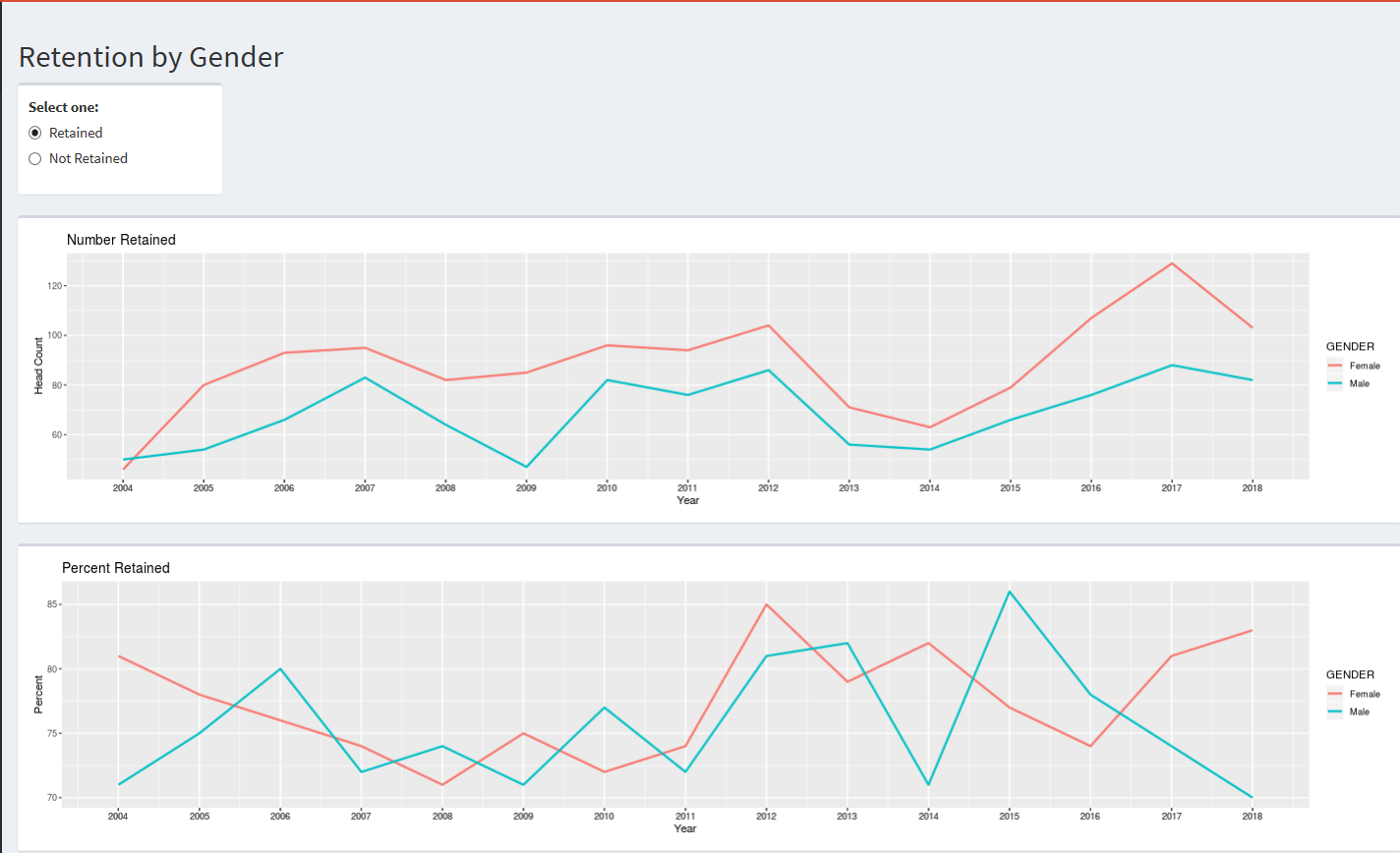
·         Tied date range slider to plot

·         **To do**:  Need to change the plot title to something generic now that I’m allowing more than one major to be selected.  The title is currently just showing the first major in the list



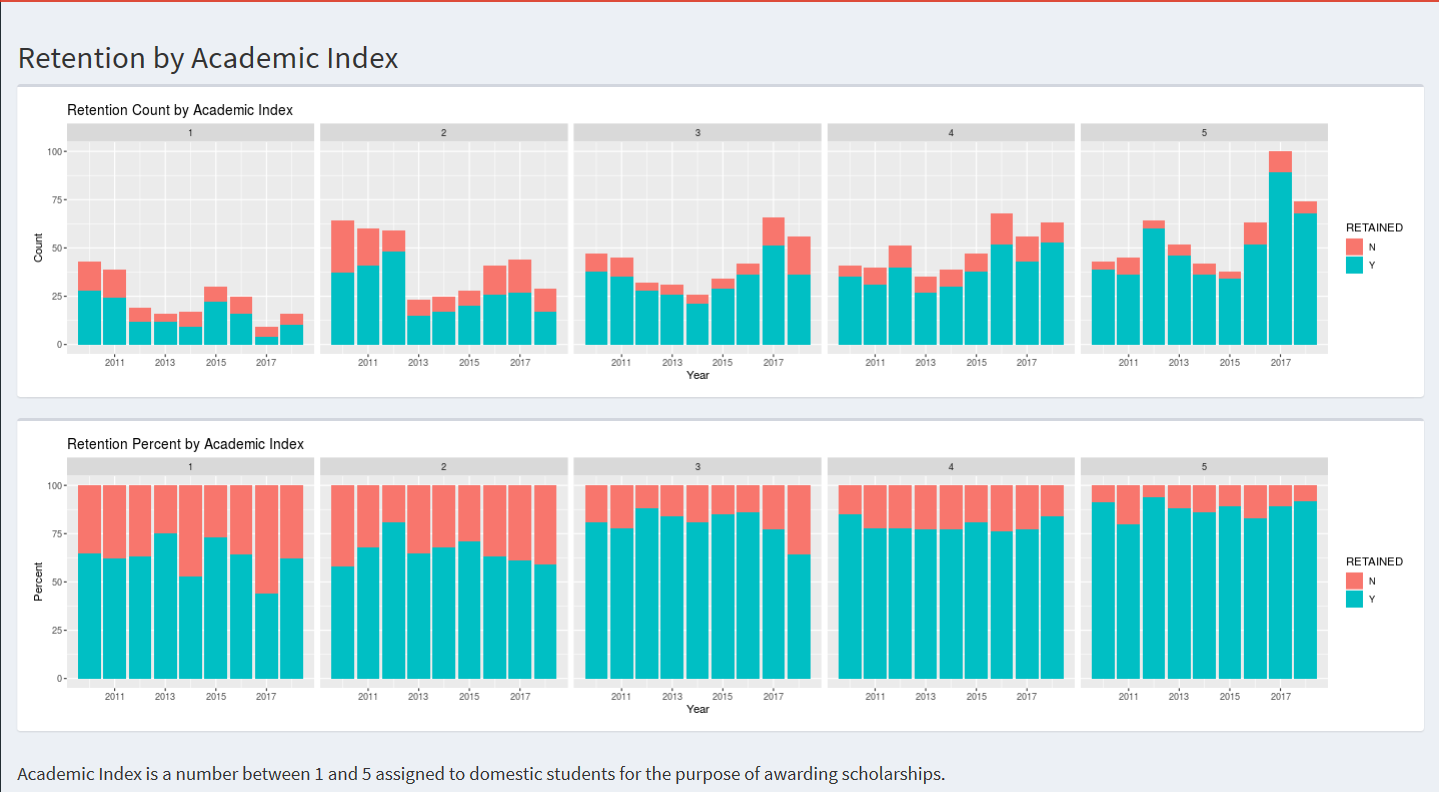
Retention by Gender

·         This page shows an example of using Radio Buttons to select Retained or Not Retained



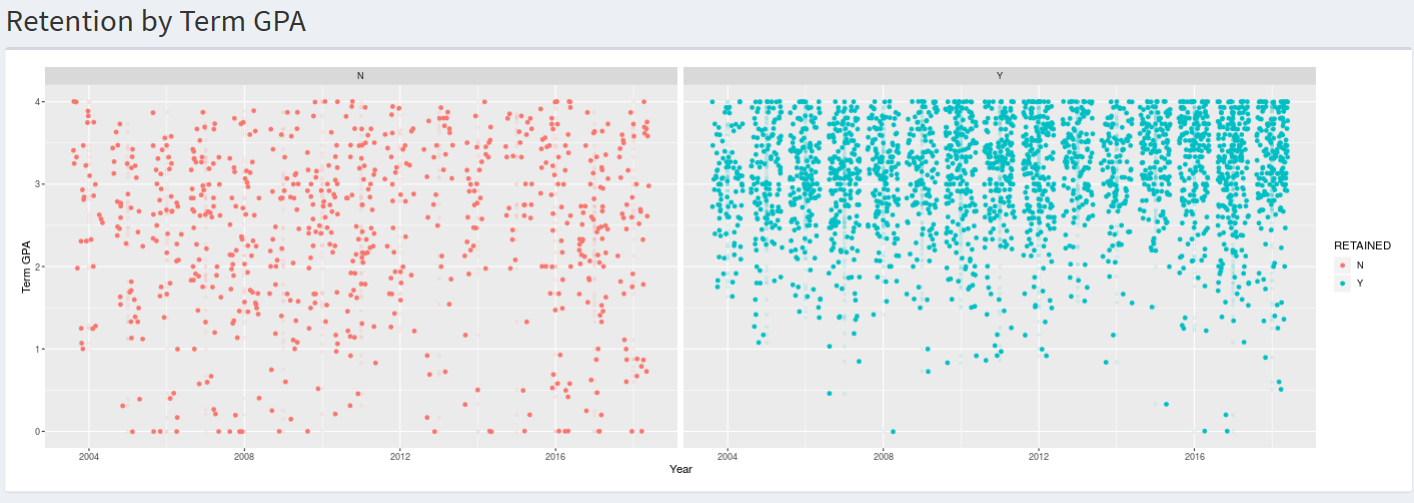
Retention by Academic Index

·         This page shows and example of using facet\_wrap to in two separate plots:  Retention Count and Retention Percent



Retention by Term GPA

·         Shows and example of a scatter plot with jitter



Working on next:

·         Need to come up with a plot for Race & Ethnicity

·         Need to work on the Help Document