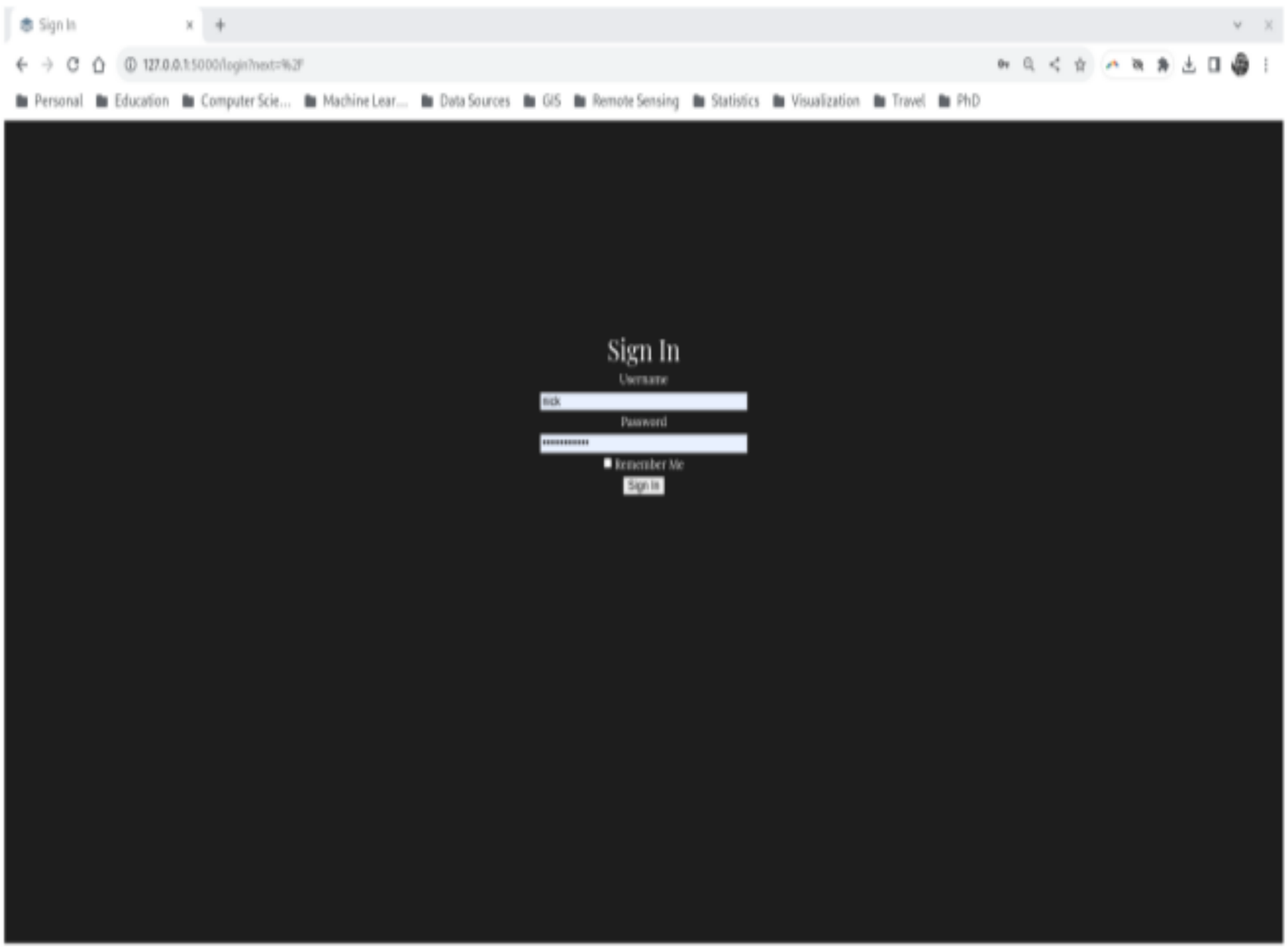
Nicholas Grokhowsky
October 27, 2023
Professor Helena Mitasova

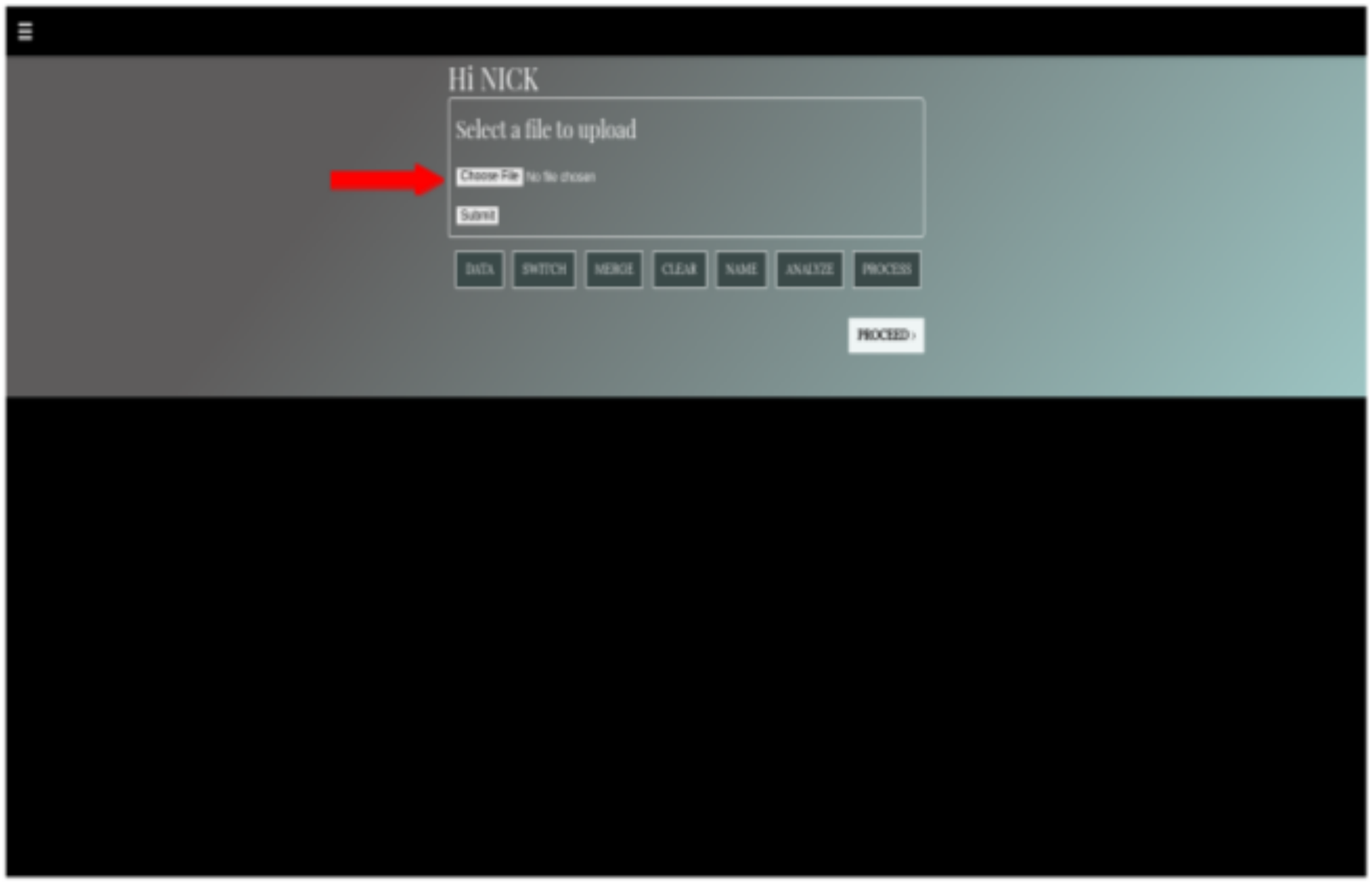
You have started to design a web application to improve the efficiency of literature reviews and their analysis. Design and incorporate a simple mapping tool into your web application and write a tutorial on how to use your web application with special focus on geospatial analysis in geohealth domain. You can use data from your publications or from your webapp demo (COVID-19). Include a section "known issues" into your tutorial to indicate what fixes you plan to address in the current version of your application.

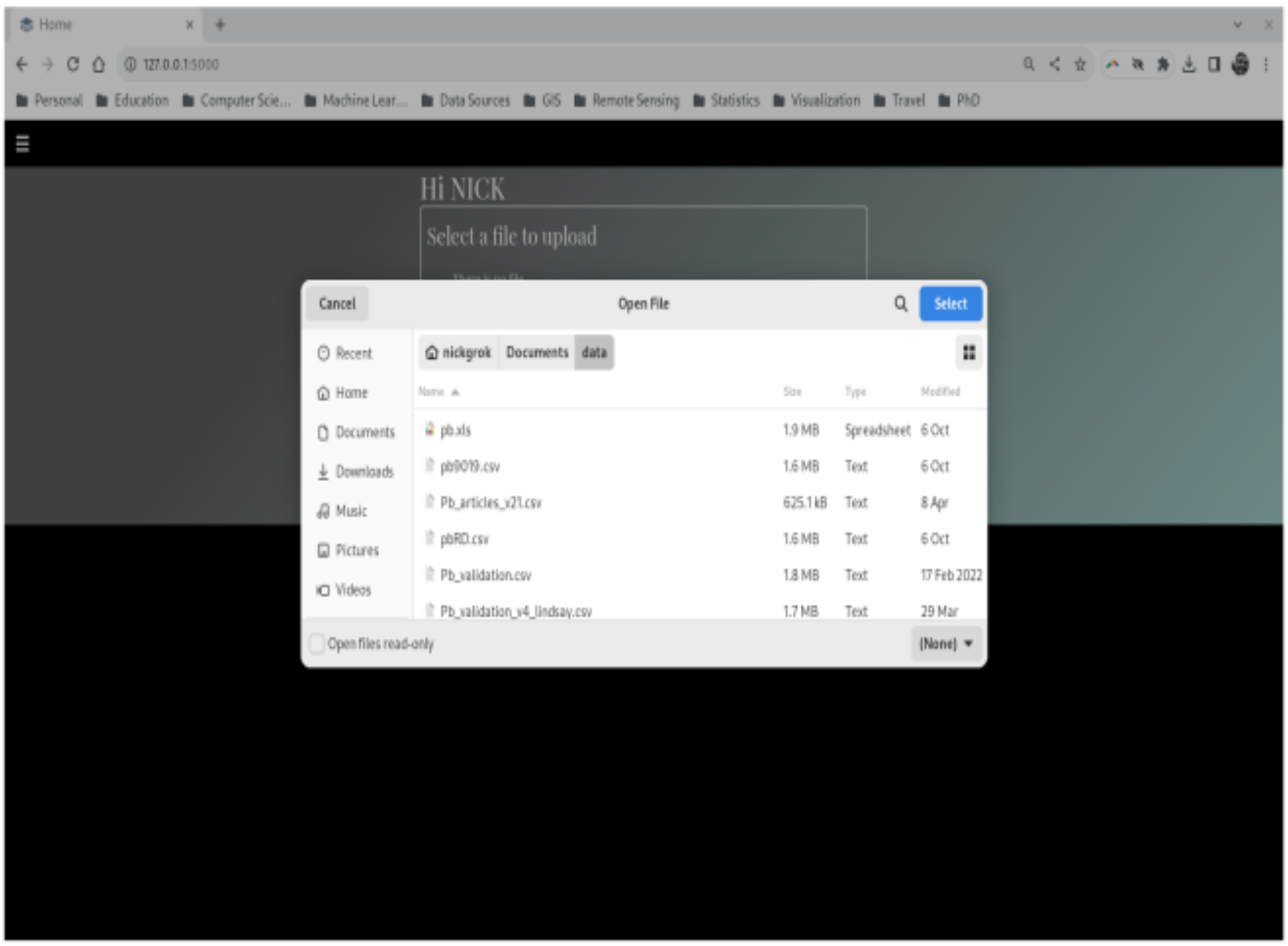
TUTORIAL

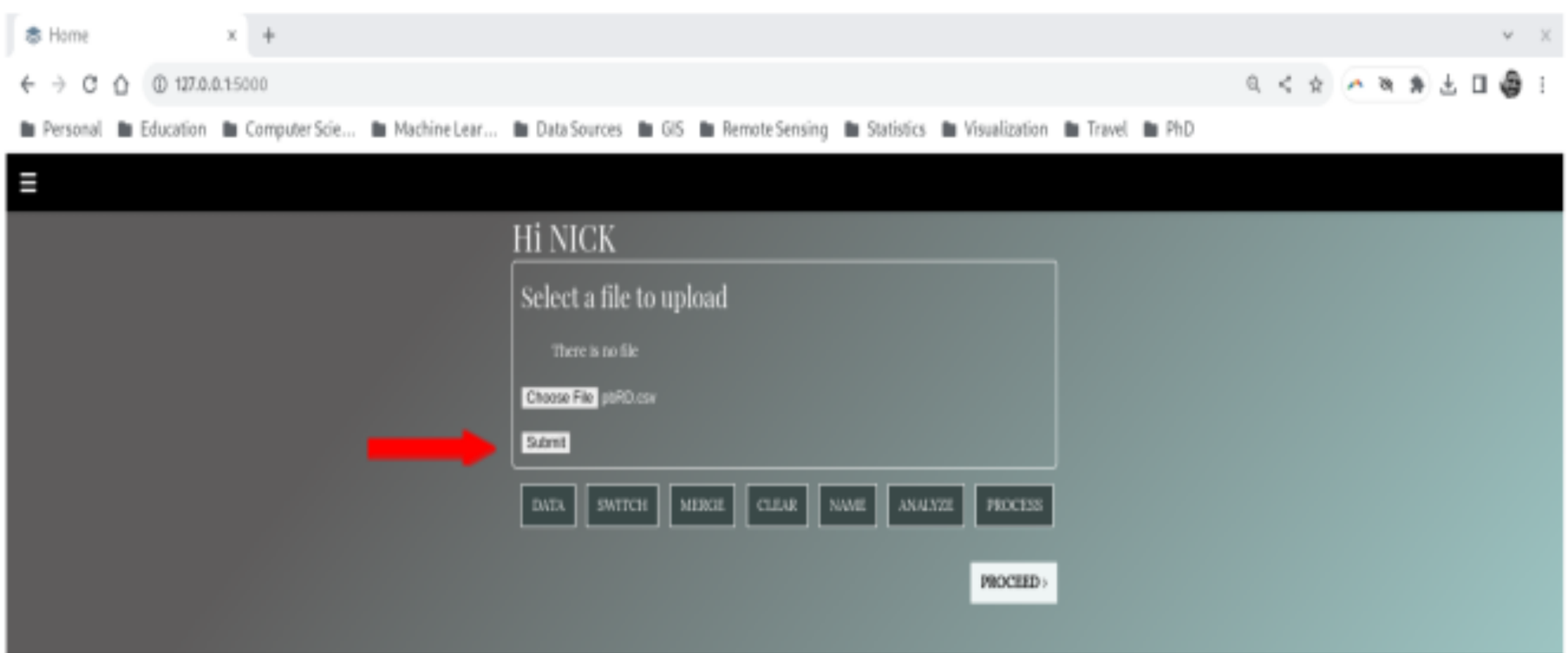
1. To begin this protype application a login screen is displayed. To gain access to the application the administrative user(s) must assign a user name and password to the intended user. If the incorrect user name and/or password are given the application will deny access and refresh the login screen. Providing a registered username and password, and clicking the 'Sign In' button allows access to the applications welcome screen.

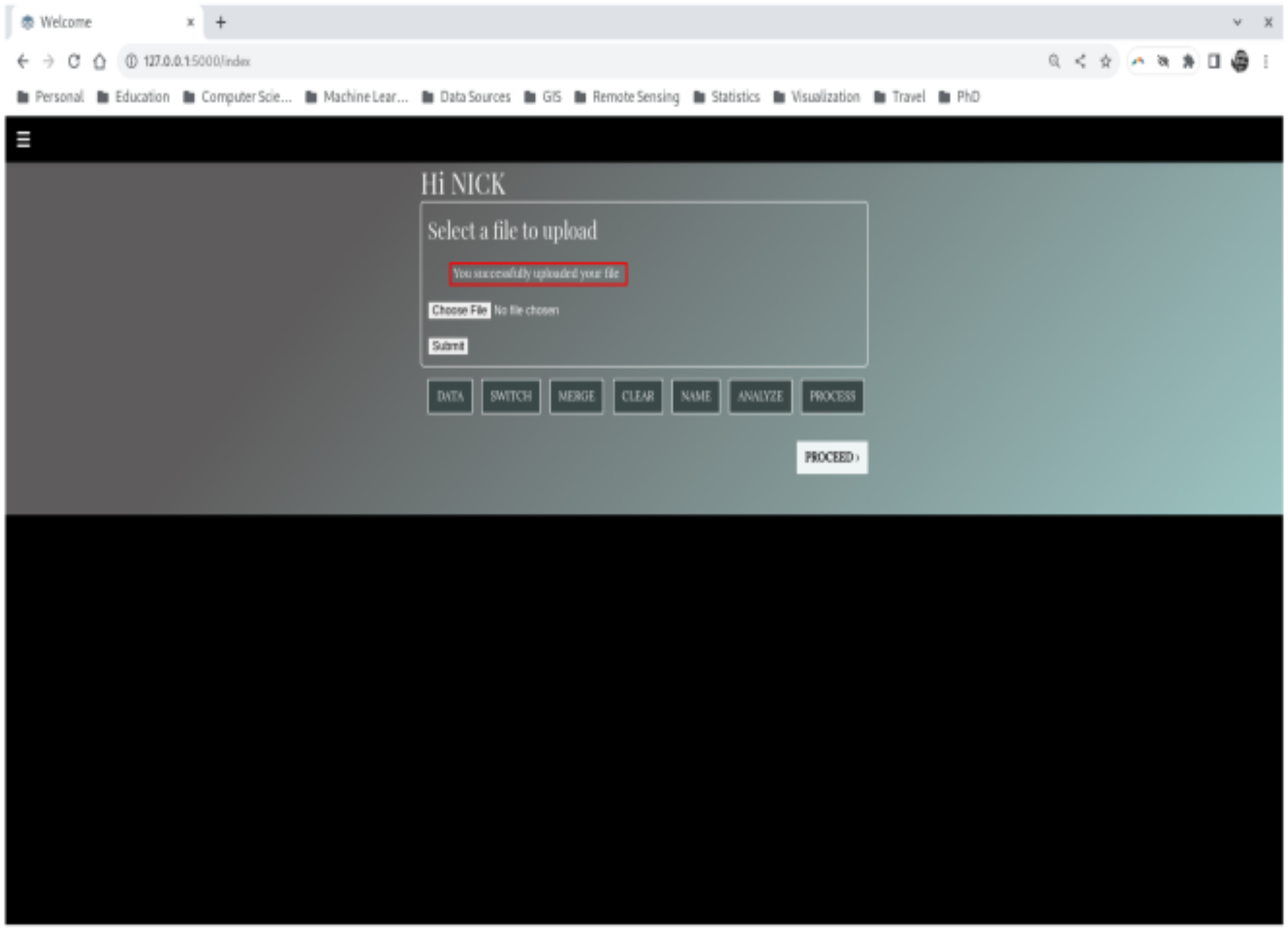


2. Next, the login screen is displayed. The user is given a series of options below a file upload section. To use the options a file must be uploaded. First the user must select 'Choose File'. Upon this selection a pop-up box will appear to allow the user to select a file from the local file tree. Once selected the user must click 'Submit' to store the file in the application and access the data in memory.

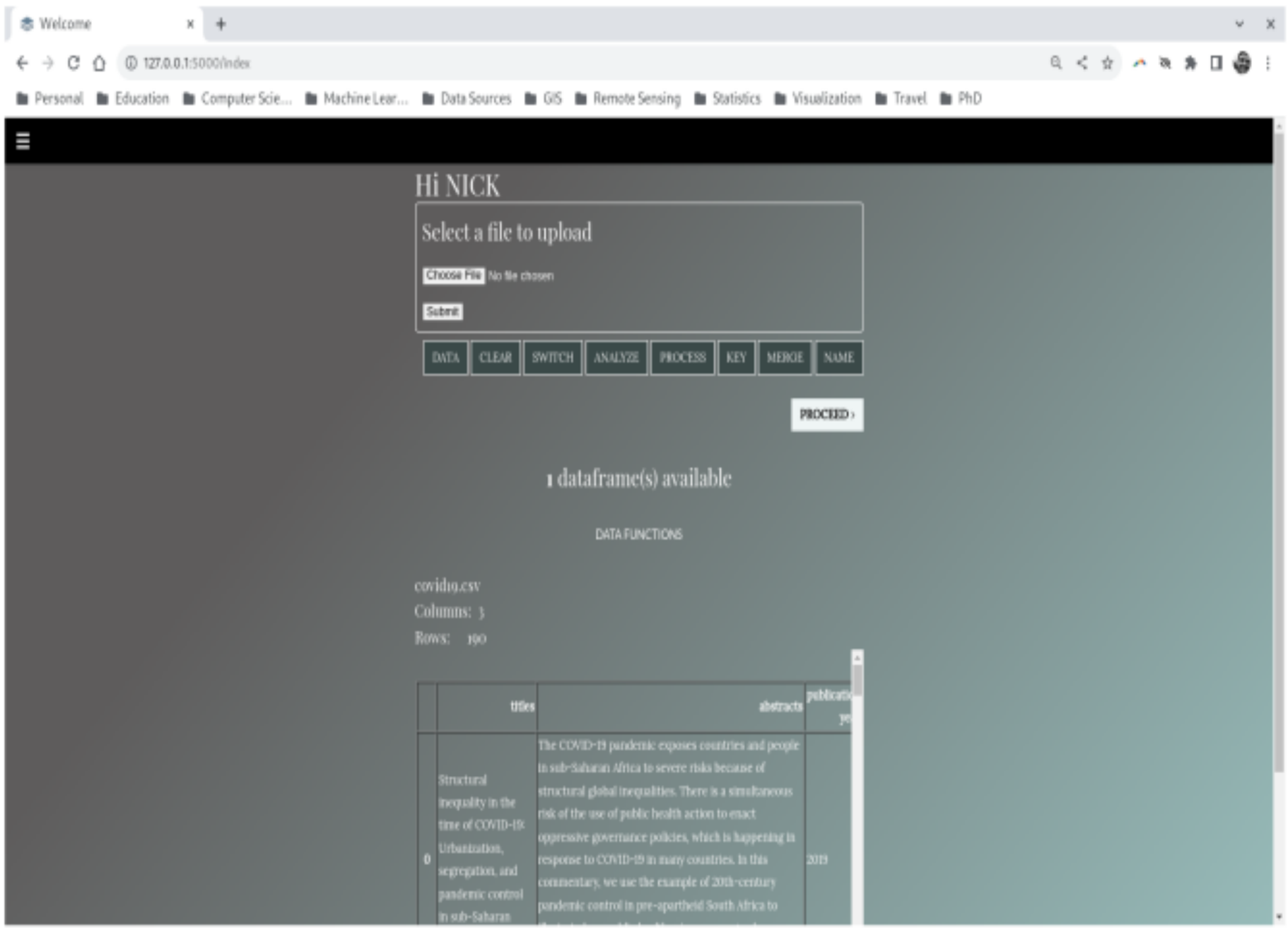

Hi NICK
Select a file to upload
Choose File
No file chosen
Submit
DATA
SWITCH
MERGE
CLEAR
NAME
ANALYZE
PROCESS
PROCEED ›


Home
127.0.0.1:5000
Personal
Education
Computer Scie...
Machine Lear...
Data Sources
GIS
Remote Sensing
Statistics
Visualization
Travel
PhD
Hi NICK
Select a file to upload
Cancel
Open File
Select
Recent
Home
Documents
Downloads
Music
Pictures
Videos
nickgrok
Documents
data
Name
Size
Type
Modified
pb.xls 1.9 MB Spreadsheet 6 Oct
pb9019.csv 1.6 MB Text 6 Oct
Pb_articles_v21.csv 625.1 kB Text 8 Apr
pbRD.csv 1.6 MB Text 6 Oct
Pb_validation.csv 1.8 MB Text 17 Feb 2022
Pb_validation_v4_lindsay.csv 1.7 MB Text 29 Mar
Open files read-only
(None)

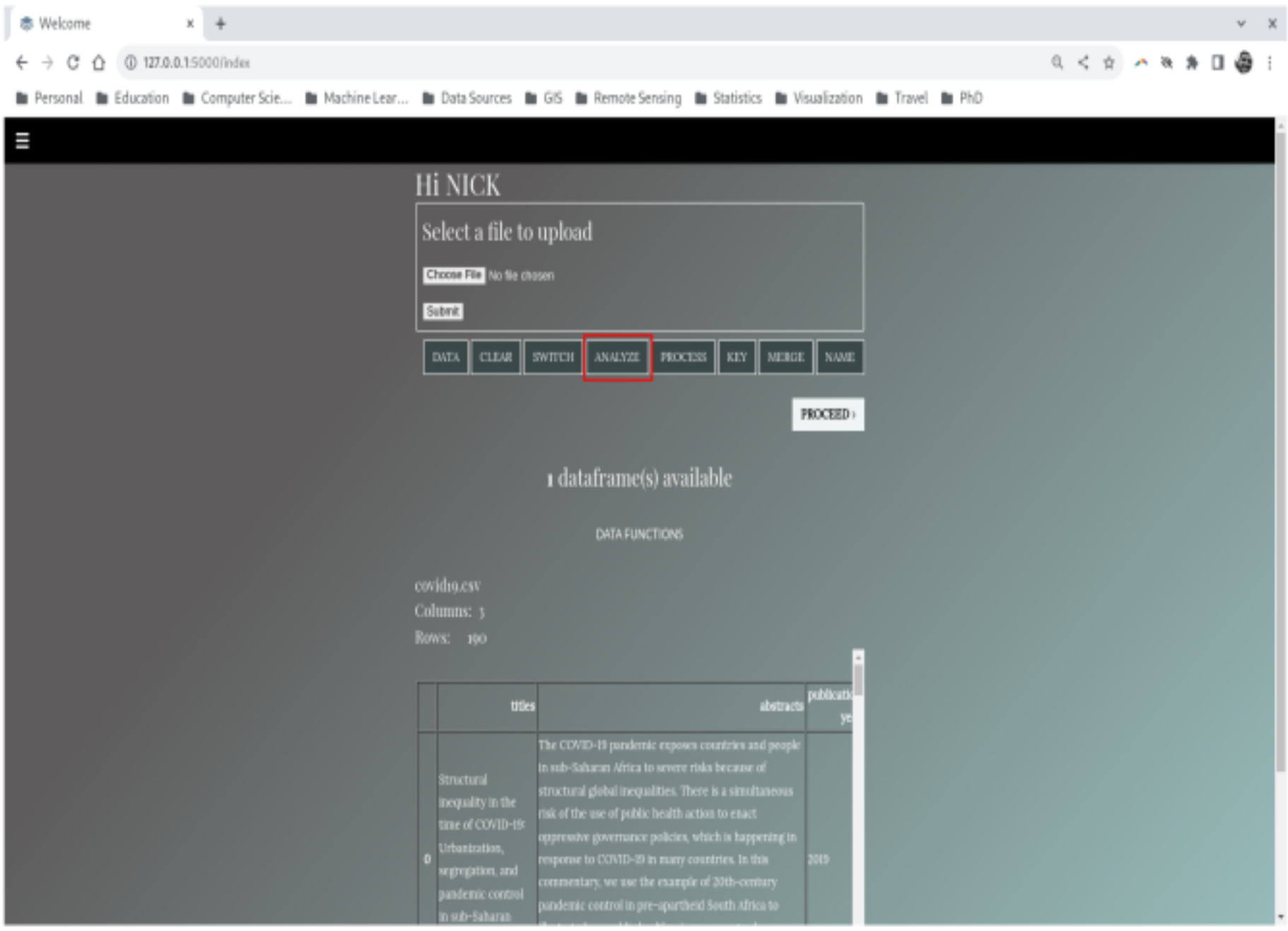

Home
127.0.0.1:5000
Personal
Education
Computer Scie...
Machine Lear...
Data Sources
GIS
Remote Sensing
Statistics
Visualization
Travel
PhD
Hi NICK
Select a file to upload
There is no file
Choose File
pbRD.csv
Submit
DATA
SWITCH
MERGE
CLEAR
NAME
ANALYZE
PROCESS
PROCEED ›



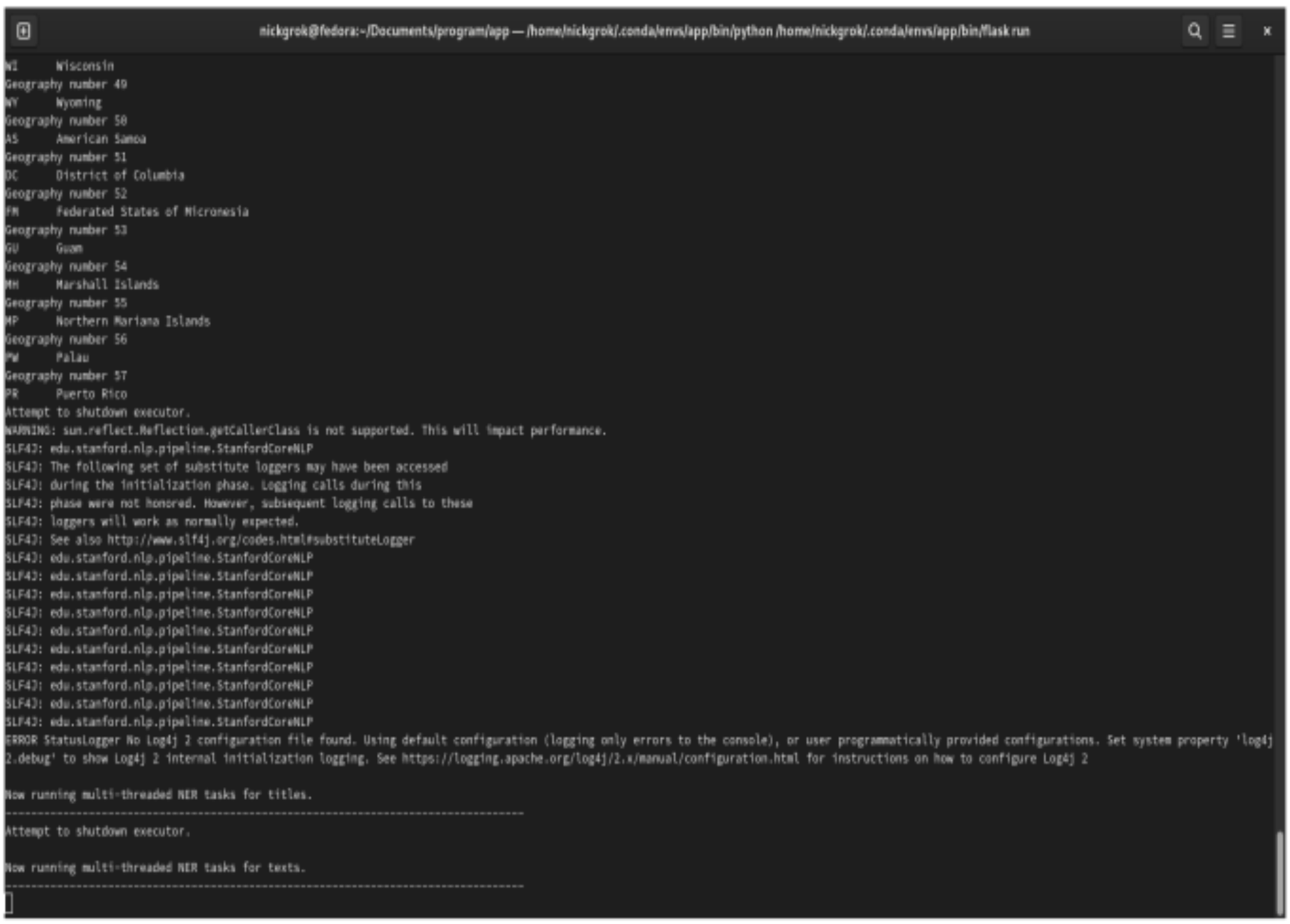
3. With a successful upload, a success message is flashed to notify the user. Now the user is able to look at the data frame uploded by clicking the 'DATA' button. The 'DATA' menu allows the user to see the entire data frame.

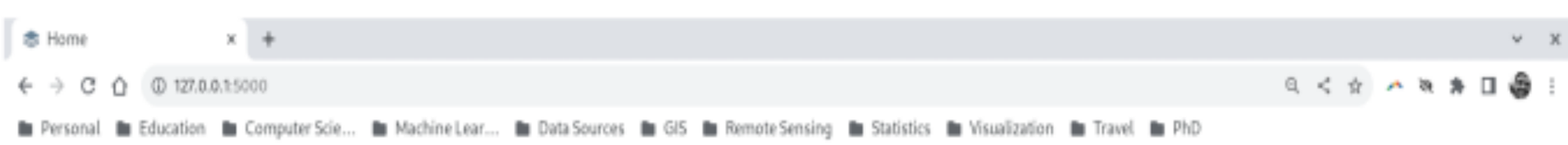


4. If a user is interested in geoparsing and sub-topic clustering this data frame they must use the 'ANALYZE' button in the main set of options below the file upload feature. By clicking the 'ANALYZE' button, the user is submitting the data to the geoparser and sub-topic clustering application. This requires the data to be uploaded in the correct format, which is three columns of 'titles', 'abstracts', and 'publication year'.

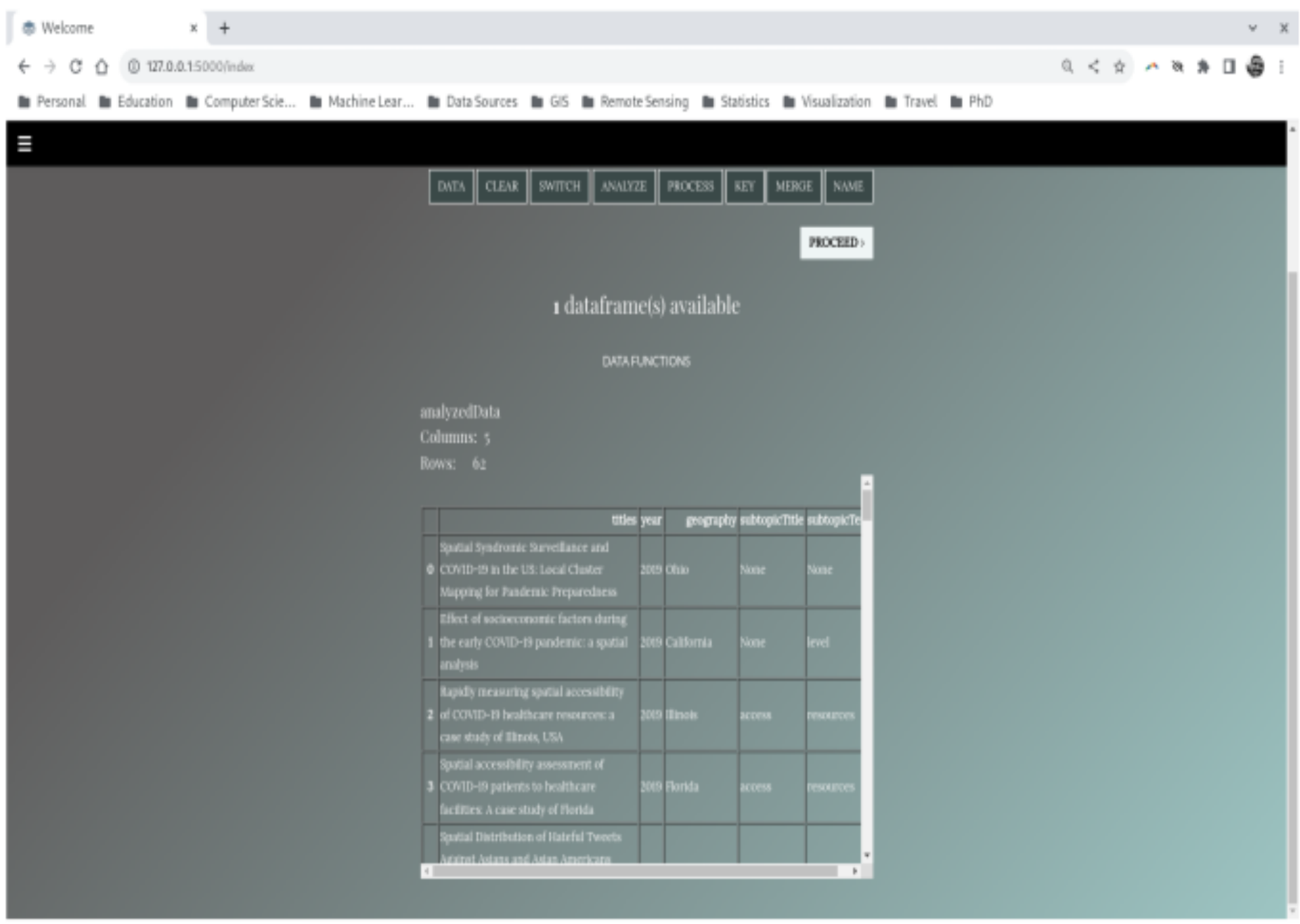


5. Upon clicking 'ANALYZE', the geoparser and sub-topic clustering application runs in the background. Once it is complete the front-end application flashes a success message.

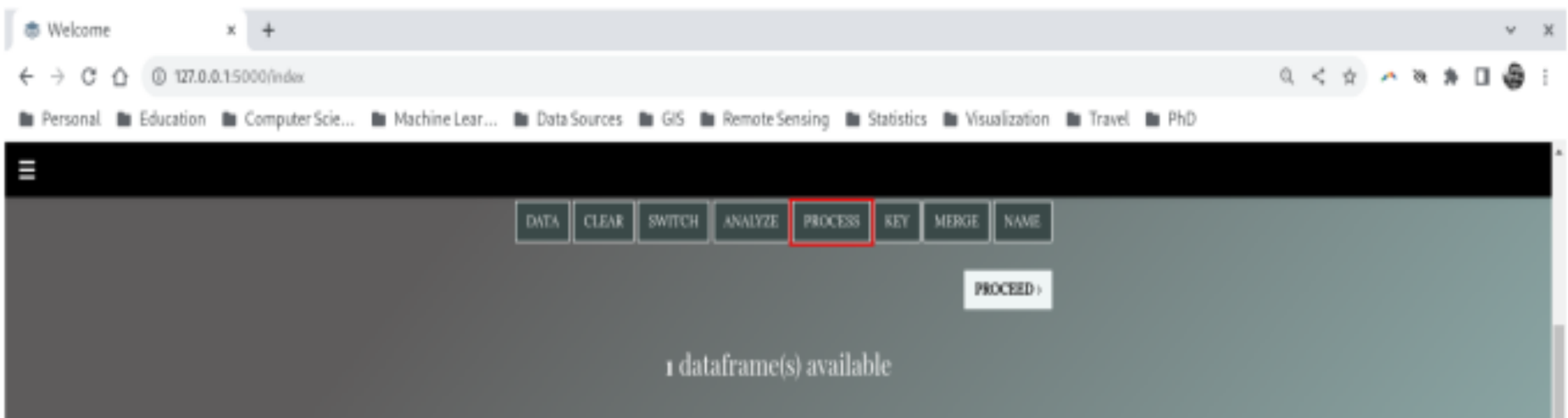




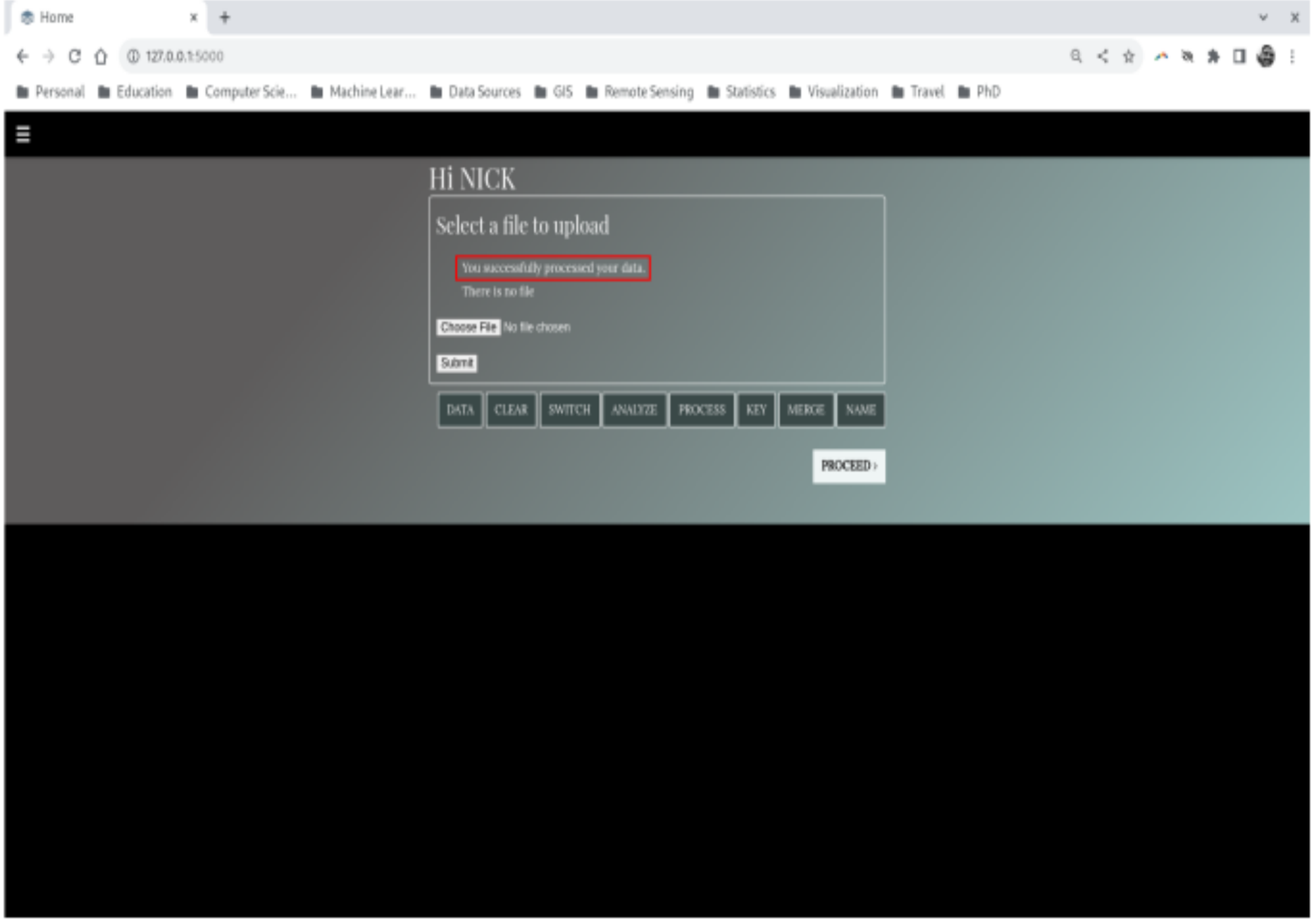
6. When the 'DATA' button is clicked, the application now shows the geoparsed and classified data. Three new columns are added to the data frame: 1) geography; 2) subtopicTitle; 3) subtopicText. Also, the data frame name has been changed to 'analyzedData'.

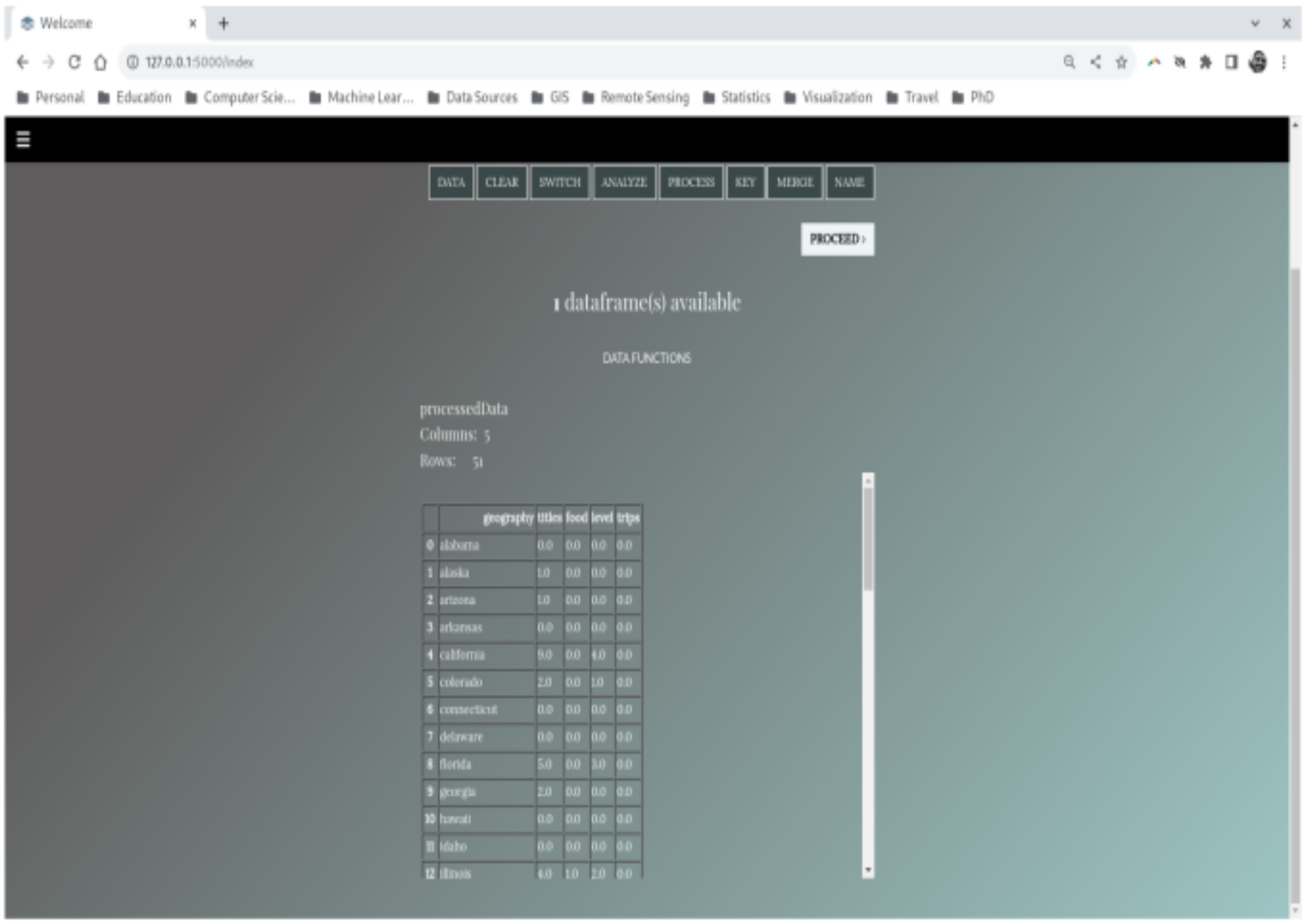


7. Next, the 'ANALYZE' button is used to group the data frame by the geography column. Additionally, any missing geographies are included in the new data frame. The output of this process returns a success message and a new data frame in place of the previous one.

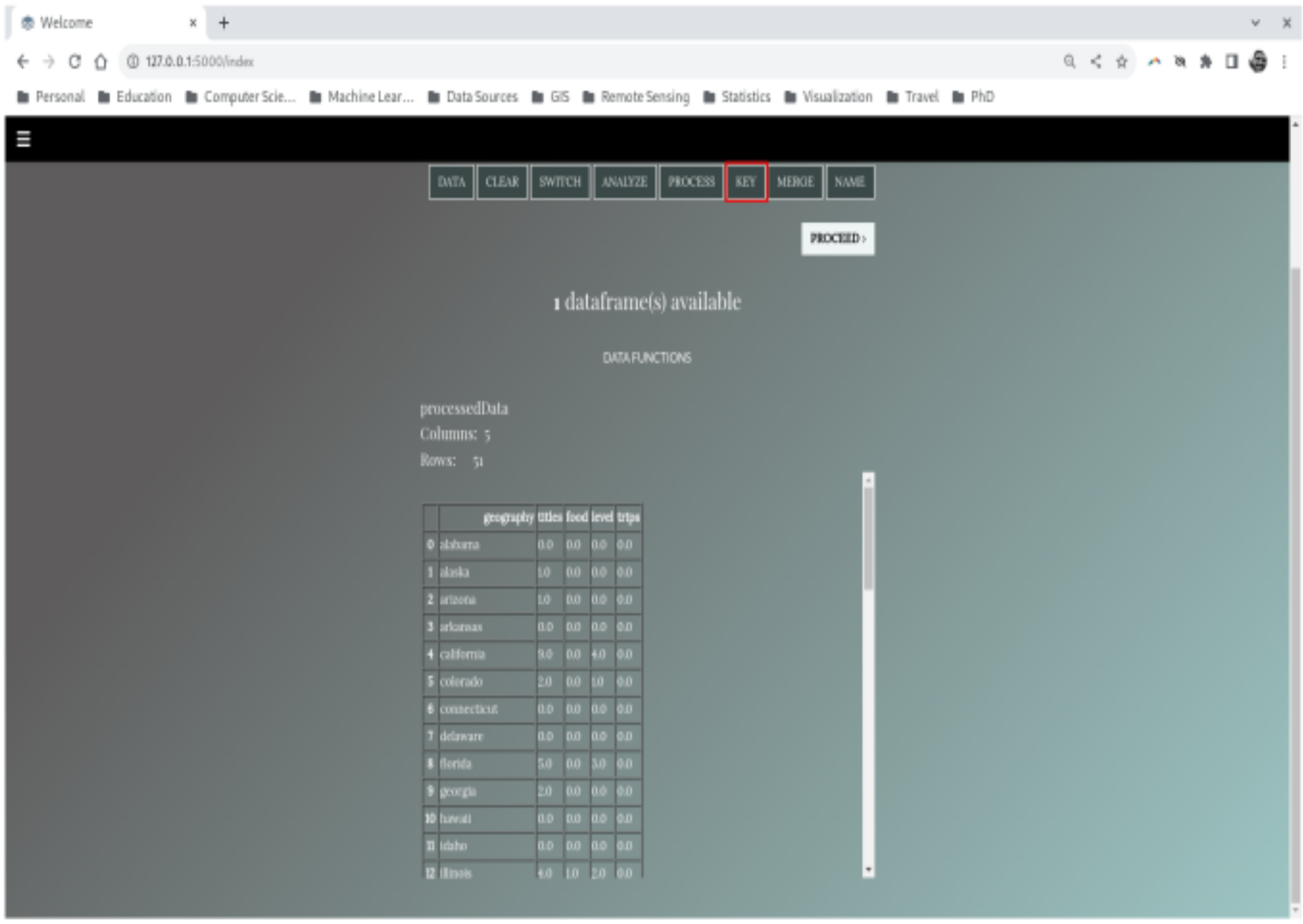


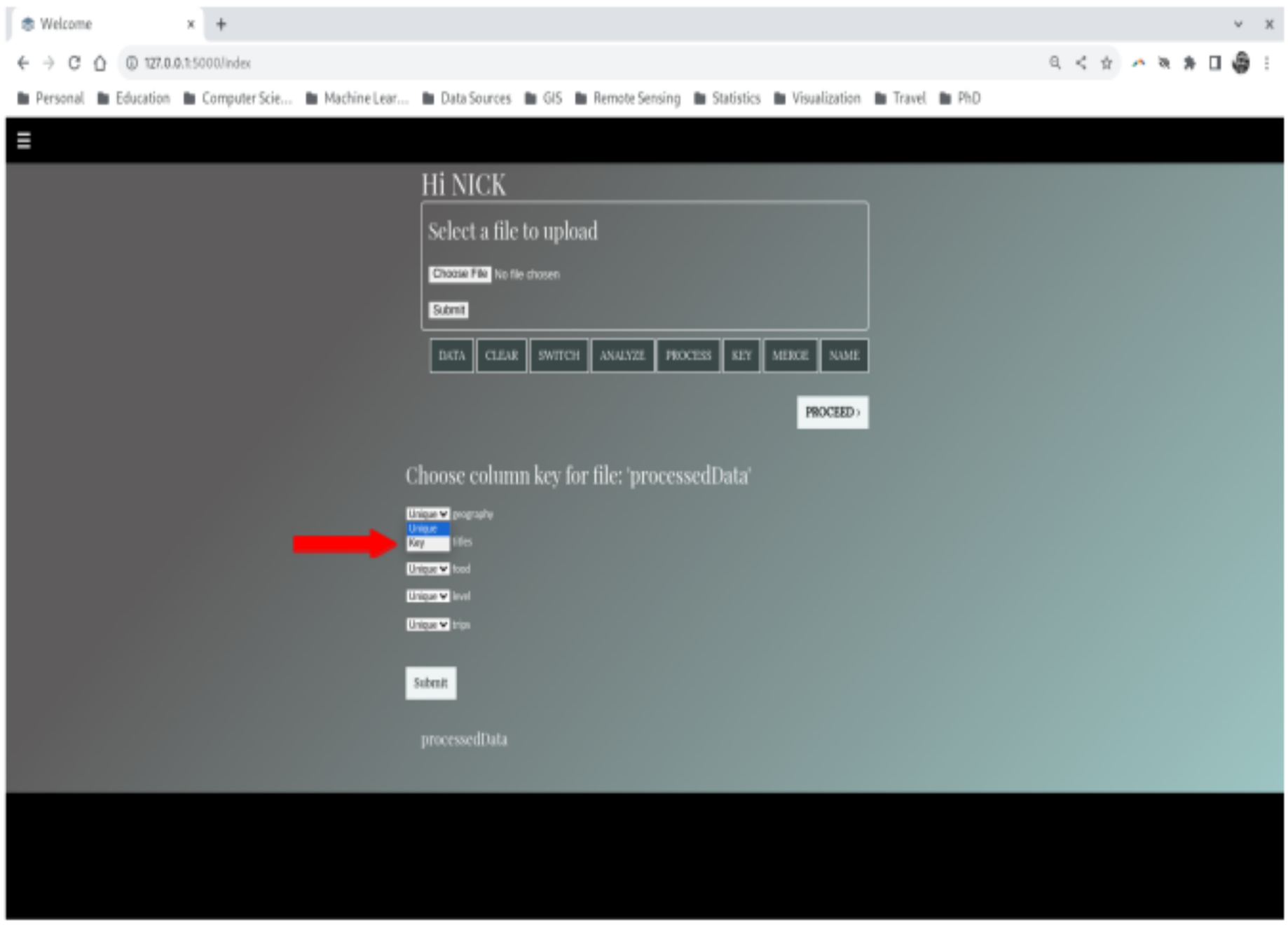
DATA FUNCTIONS
analyzedData
Columns: 5
Rows: 62
titles year geography subtopicTitle subtopicTe
0 Spatial Syndromic Surveillance and COVID-19 in the US: Local Cluster Mapping for Pandemic Preparedness 2019 Ohio None None
1 Effect of socioeconomic factors during the early COVID-19 pandemic: a spatial analysis 2019 California None level
2 Rapidly measuring spatial accessibility of COVID-19 healthcare resources: a case study of Illinois, USA 2019 Illinois access resources
3 Spatial accessibility assessment of COVID-19 patients to healthcare facilities: A case study of Florida 2019 Florida access resources
Spatial Distribution of Hateful Tweets Against Asians and Asian Americans


Home
127.0.0.1:5000
Personal Education Computer Scie... Machine Lear... Data Sources GIS Remote Sensing Statistics Visualization Travel PhD
Hi NICK
Select a file to upload
You successfully processed your data.
There is no file
Choose File No file chosen
Submit
DATA CLEAR SWITCH ANALYZE PROCESS KEY MERGE NAME
PROCEED ›


Welcome
127.0.0.1:5000/index
Personal Education Computer Scie... Machine Lear... Data Sources GIS Remote Sensing Statistics Visualization Travel PhD
DATA CLEAR SWITCH ANALYZE PROCESS KEY MERGE NAME
PROCEED ›
1 dataframe(s) available
DATA FUNCTIONS
processedData
Columns: 5
Rows: 51
geography titles food level trips
0 alabama 0.0 0.0 0.0 0.0
1 alaska 1.0 0.0 0.0 0.0
2 arizona 1.0 0.0 0.0 0.0
3 arkansas 0.0 0.0 0.0 0.0
4 california 9.0 0.0 4.0 0.0
5 colorado 2.0 0.0 1.0 0.0
6 connecticut 0.0 0.0 0.0 0.0
7 delaware 0.0 0.0 0.0 0.0
8 florida 5.0 0.0 3.0 0.0
9 georgia 2.0 0.0 0.0 0.0
10 hawaii 0.0 0.0 0.0 0.0
11 idaho 0.0 0.0 0.0 0.0
12 illinois 4.0 1.0 2.0 0.0

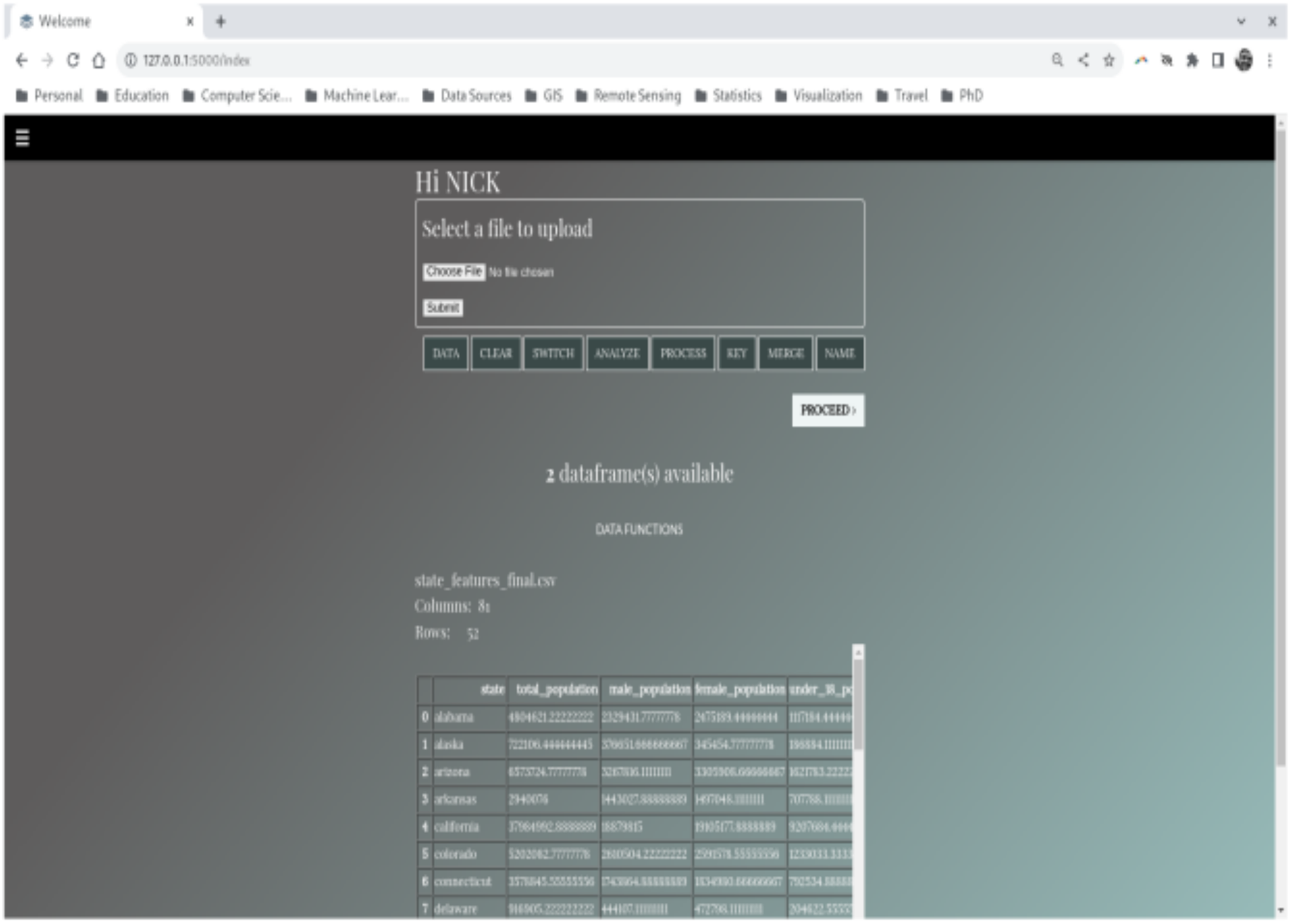
8. Finally, for this data frame, the 'KEY' button is used to select the column that holds the key values for a merge with the feature variable data. The 'KEY' button is pressed and then the 'key' value is selected next to the appropriate column name in the menu that follows. The 'Submit' button submits the selection to the application.



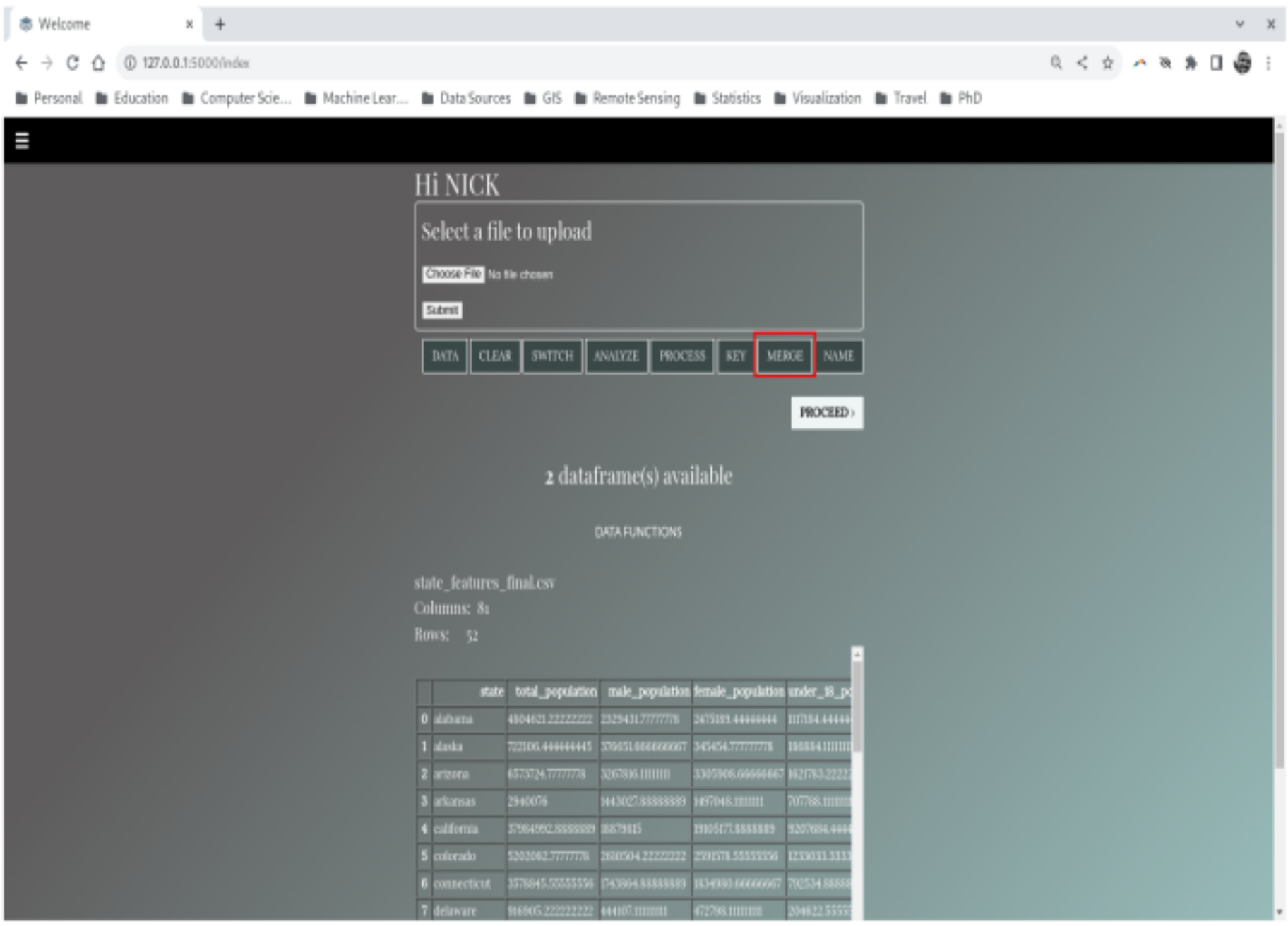


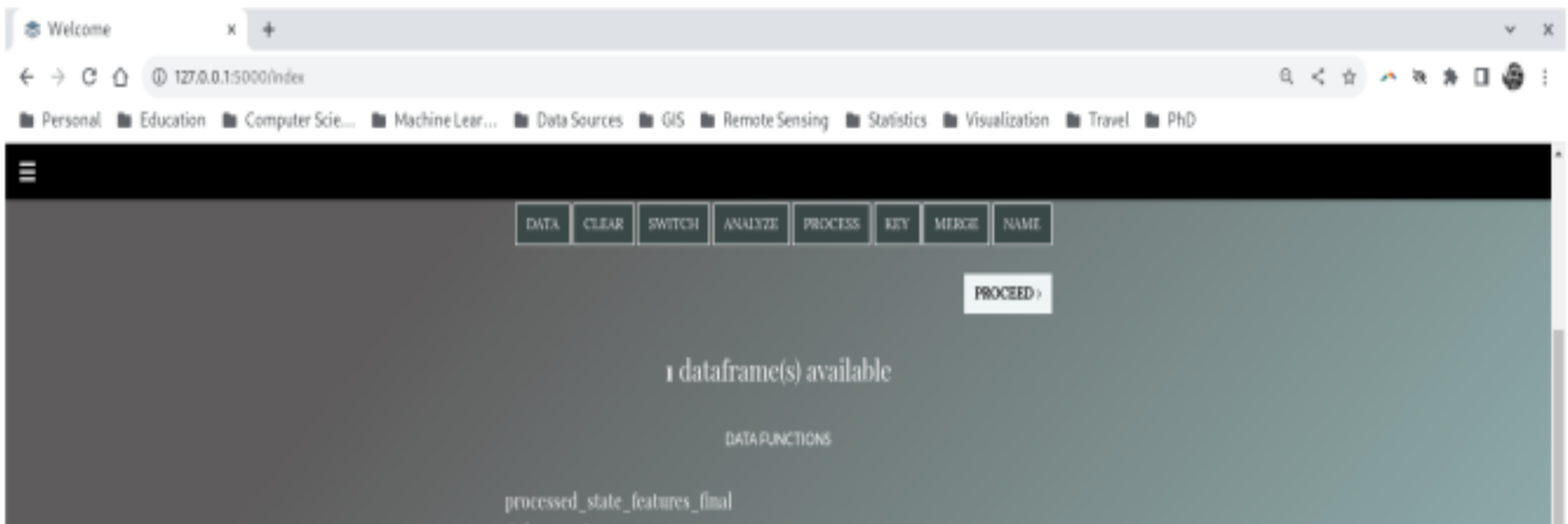
9. For the feature variable data frame, the same upload process is used as the first data frame was uploaded. Once this is uploaded the 'SWITCH' button can be used to switch between data frams. The program will store two data frames at one time. Also, the 'KEY' button must be used to select the merge key while the feature variable data frame is selected. This is the

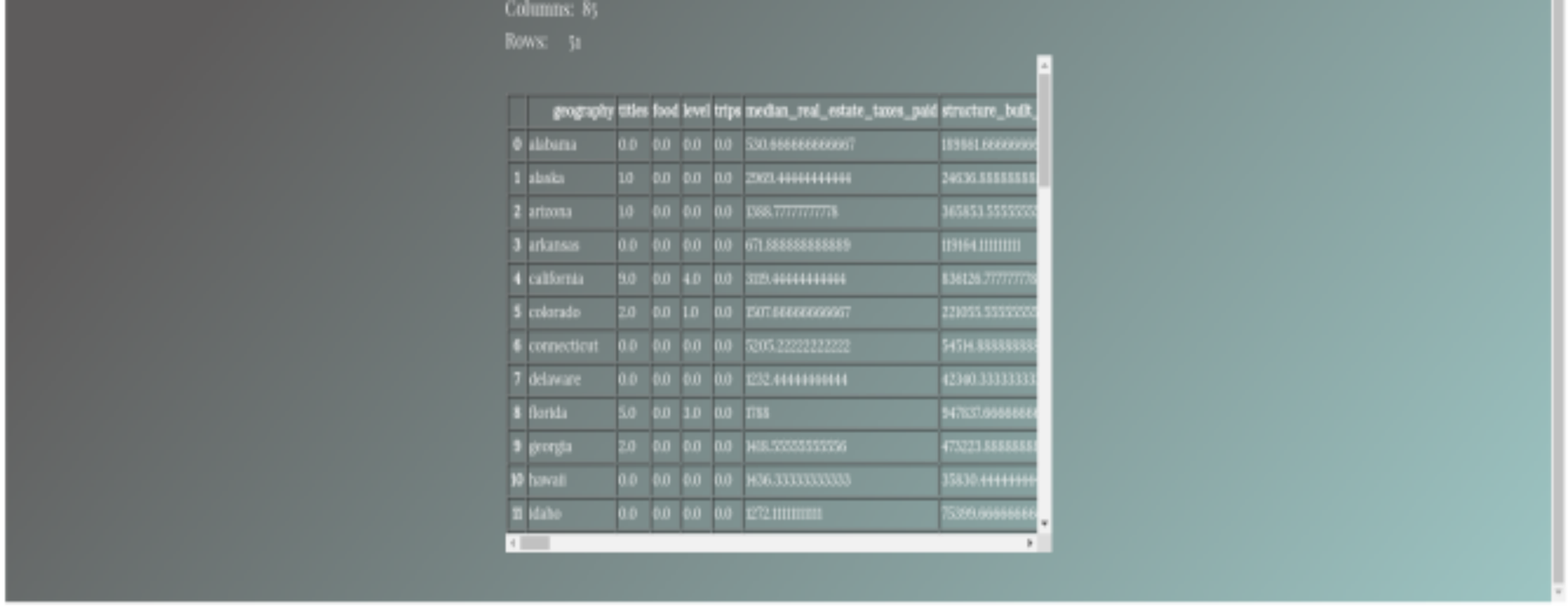
same process as the first data frame.



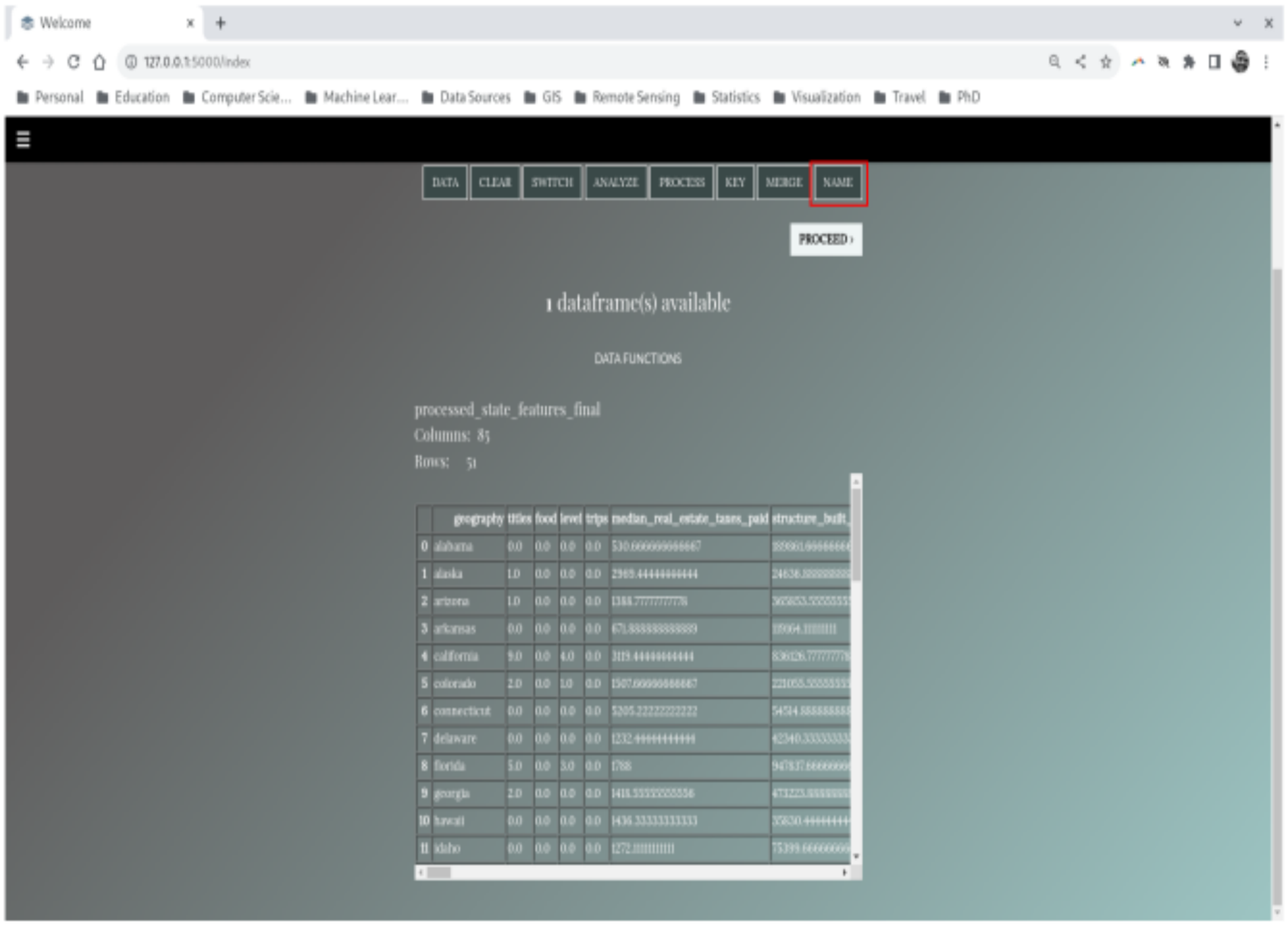
10. After the feature variable data frame key has been selected the 'MERGE' button can be pressed to merge the two data frames.

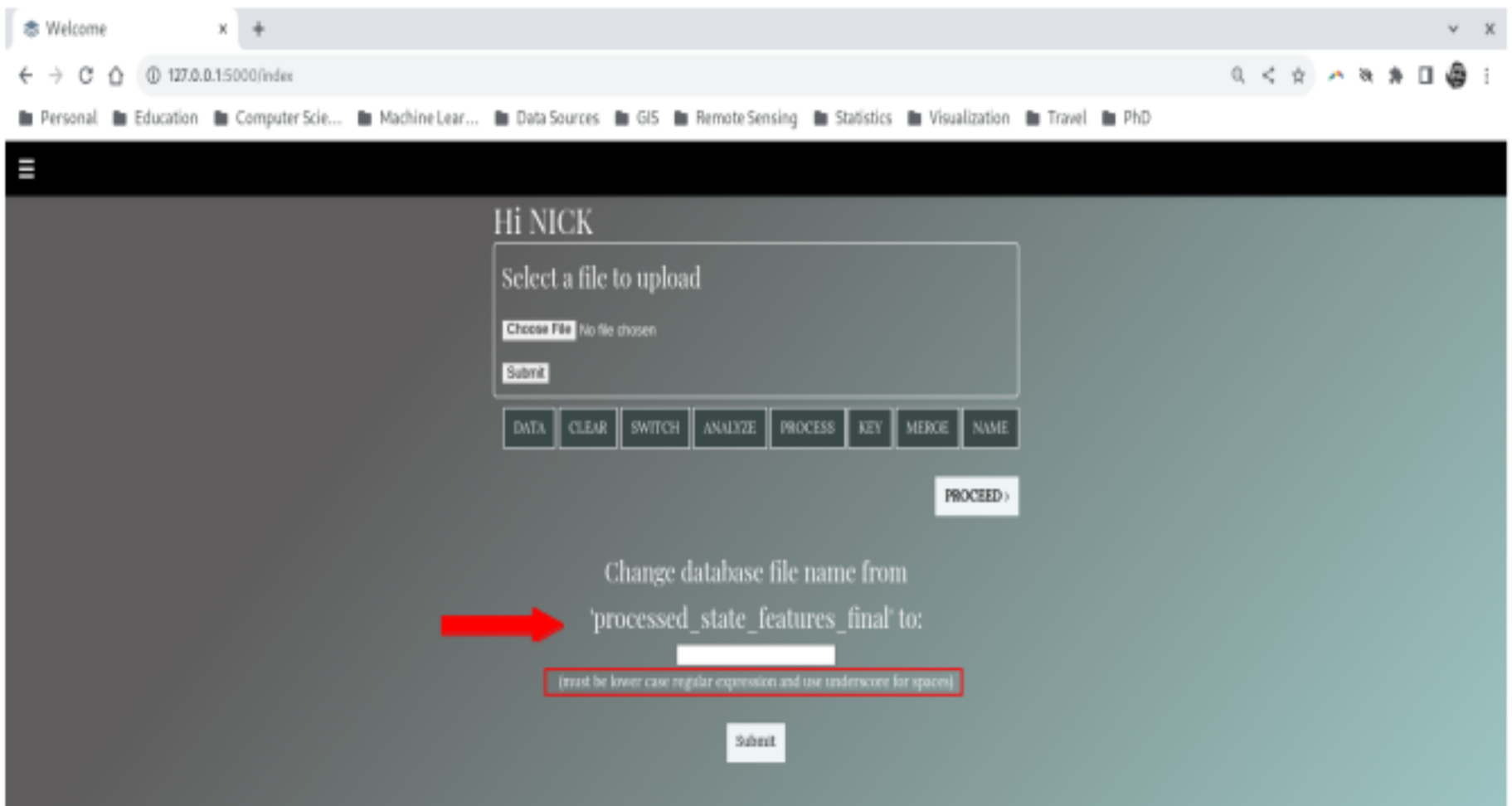


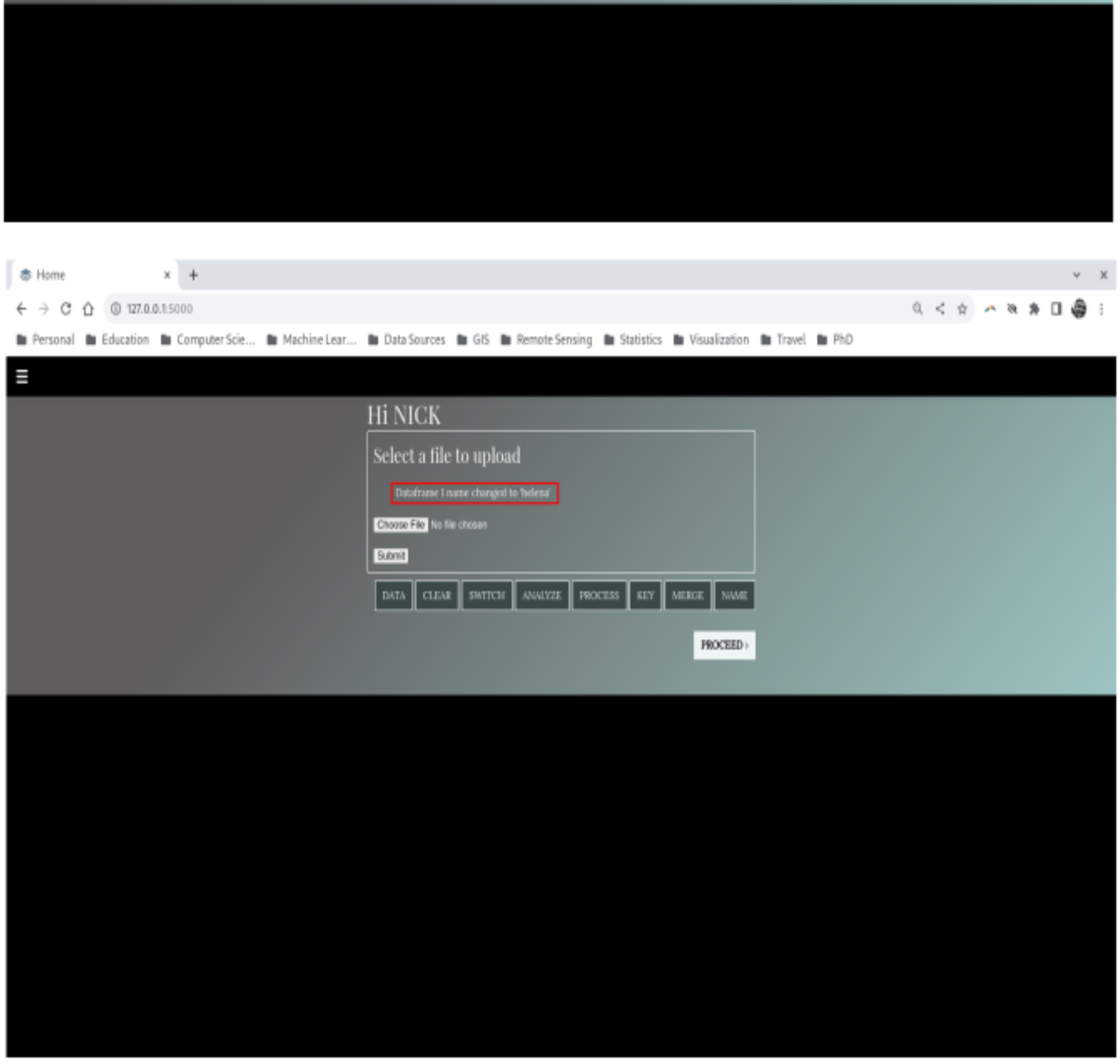




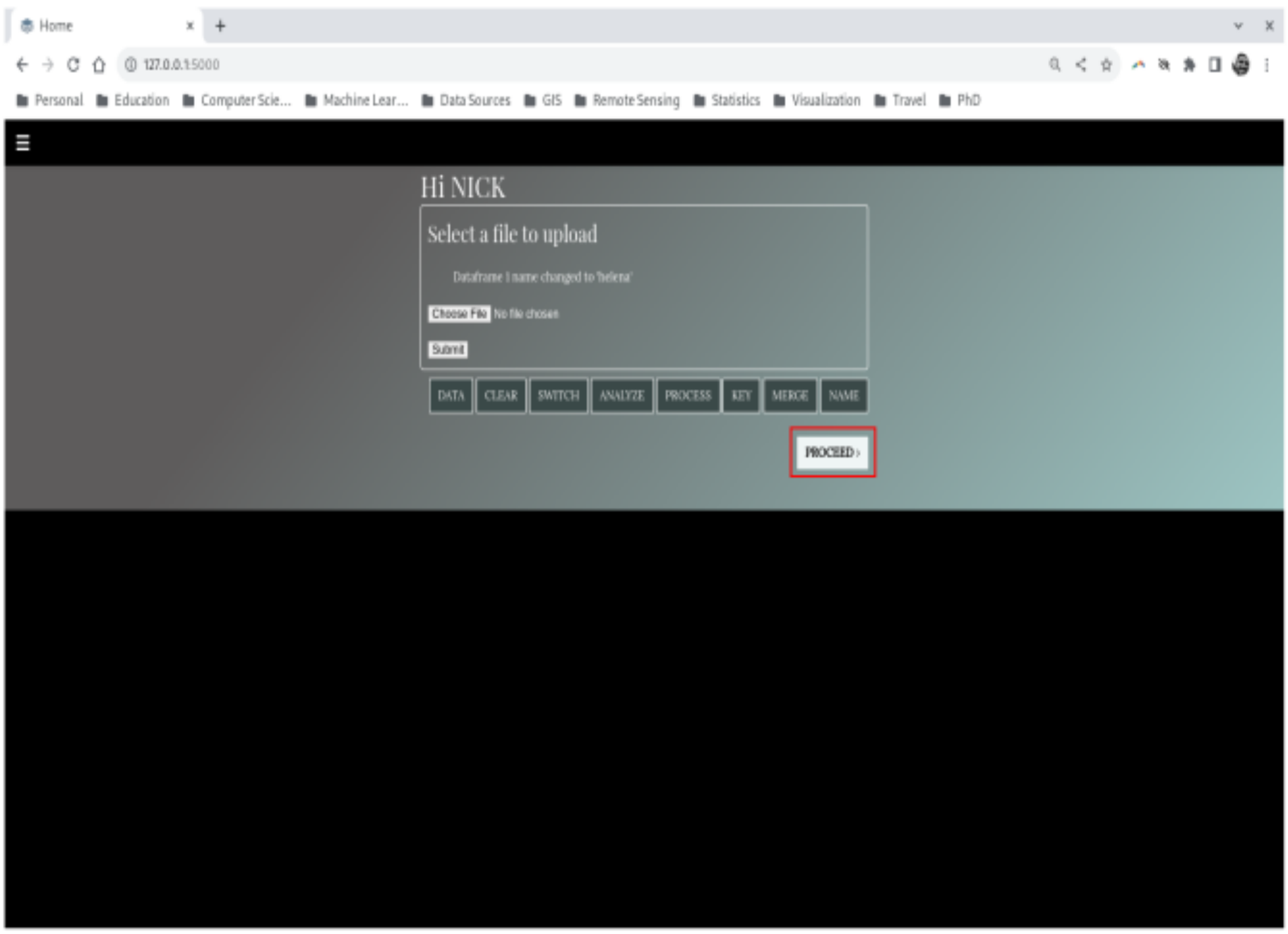
11. Next, the 'NAME' button is used to name the final database that will store this data. A list of rules for the naming convention is displayed. The 'Submit' button will save the new name to the data frame, and a success message will be flashed.

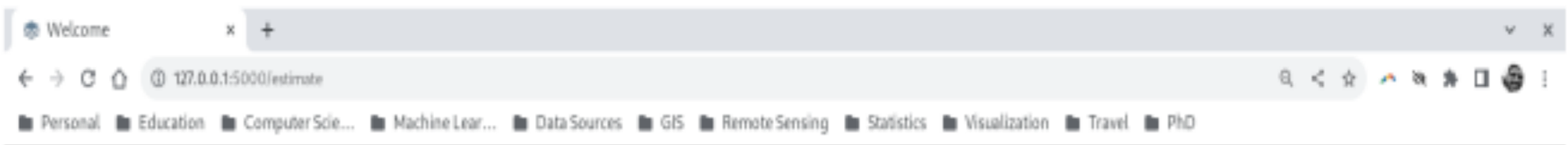


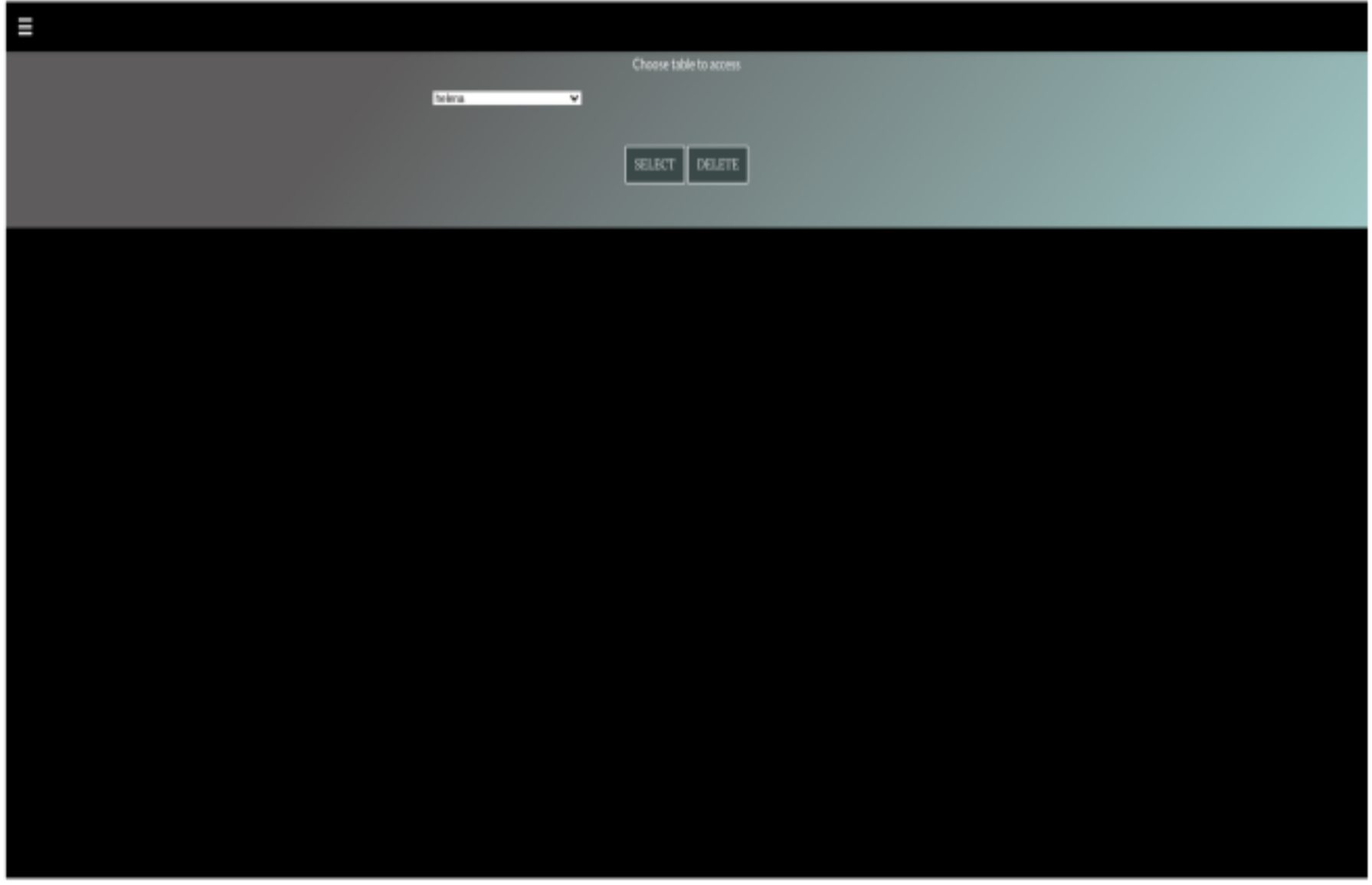




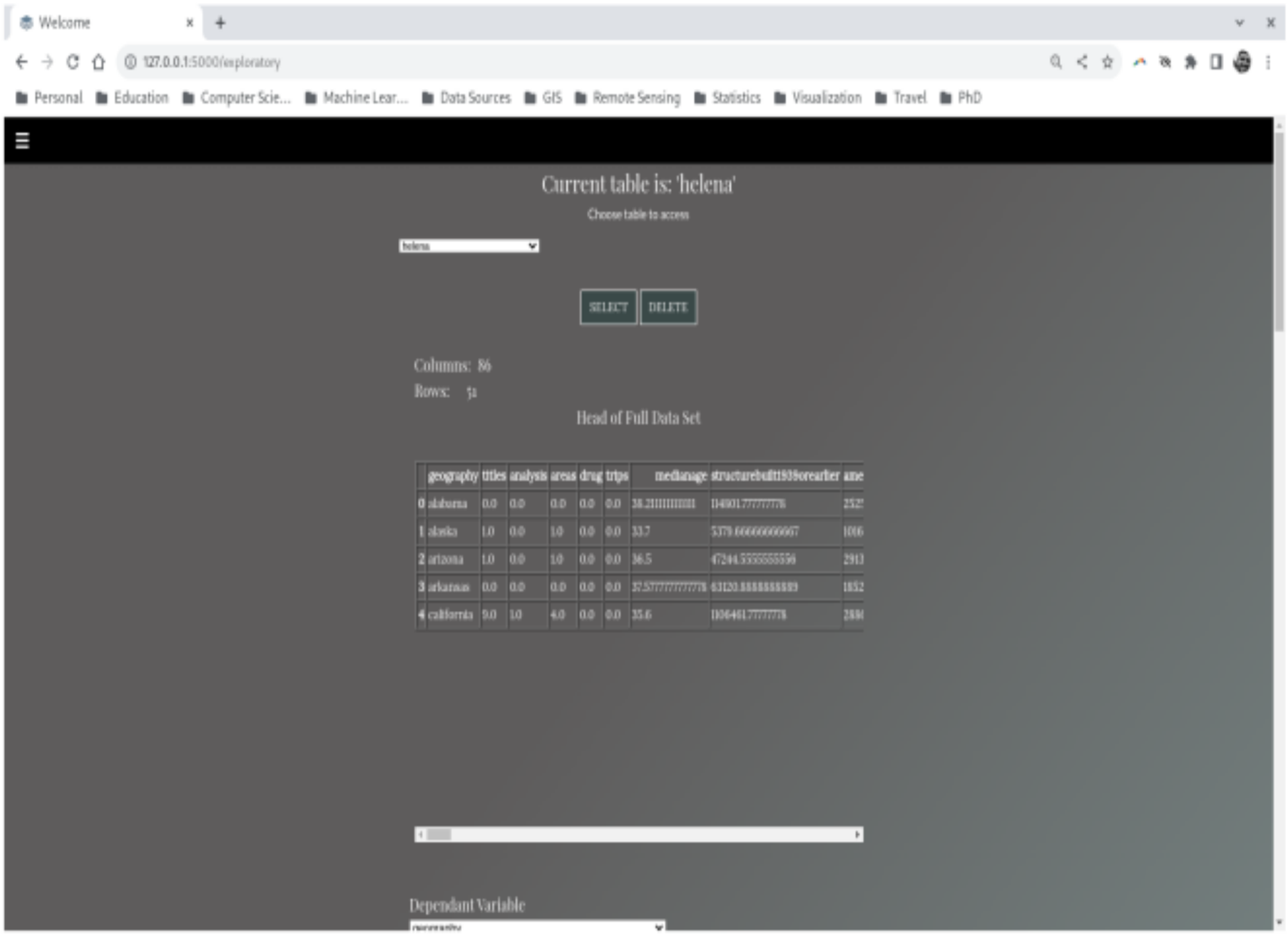
12. Pressing the 'PROCEED' button will save the data frame as a database table named after the name selected in the previous step. It will also bring the user to a new web page that allows them to choose which database table to work with.

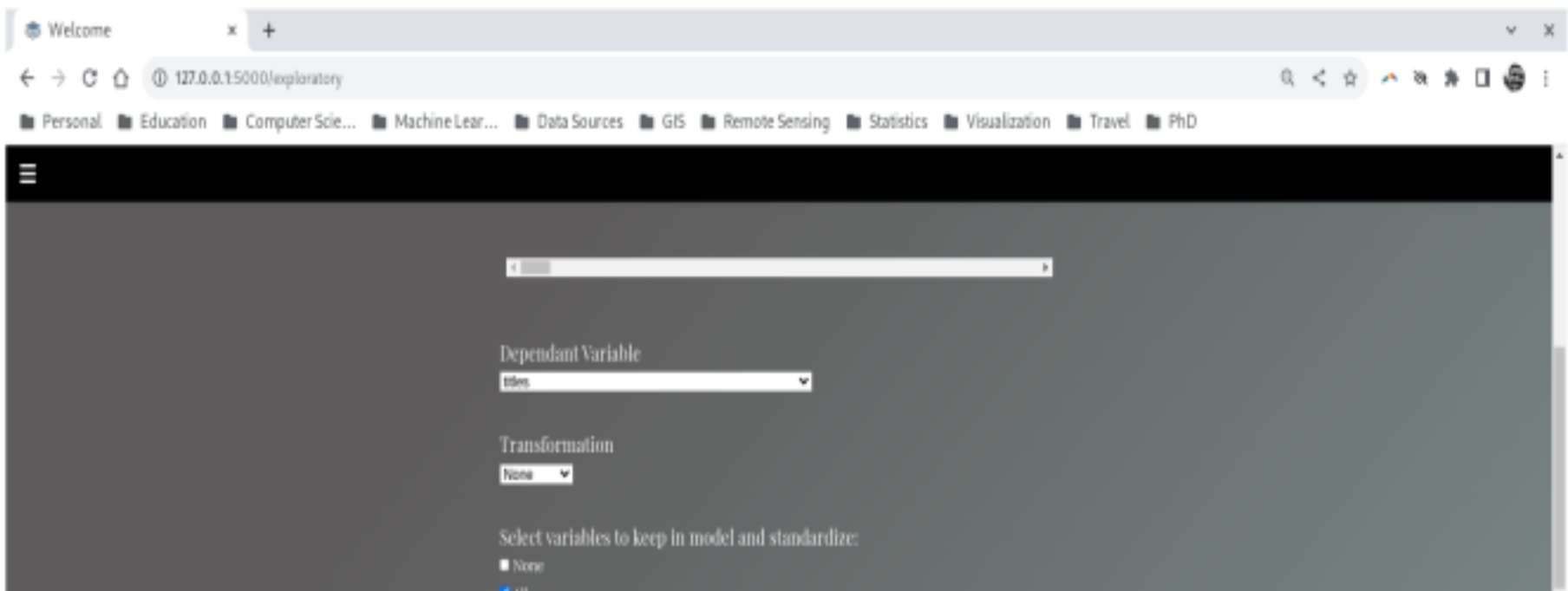


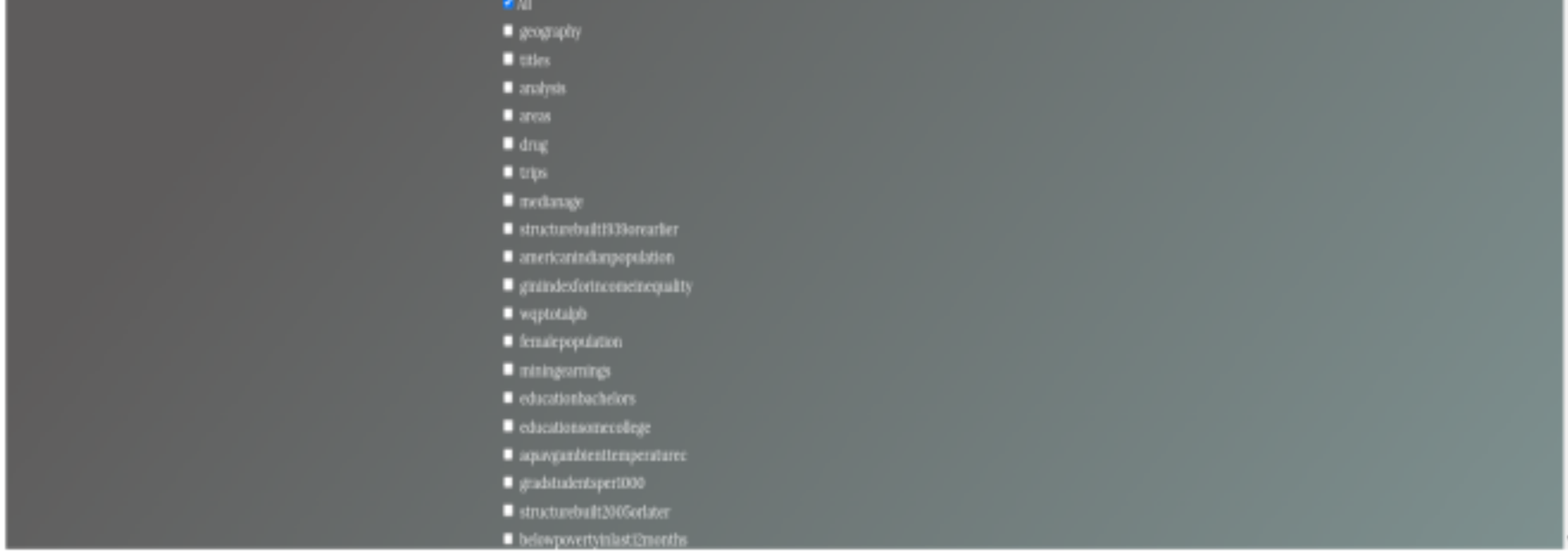


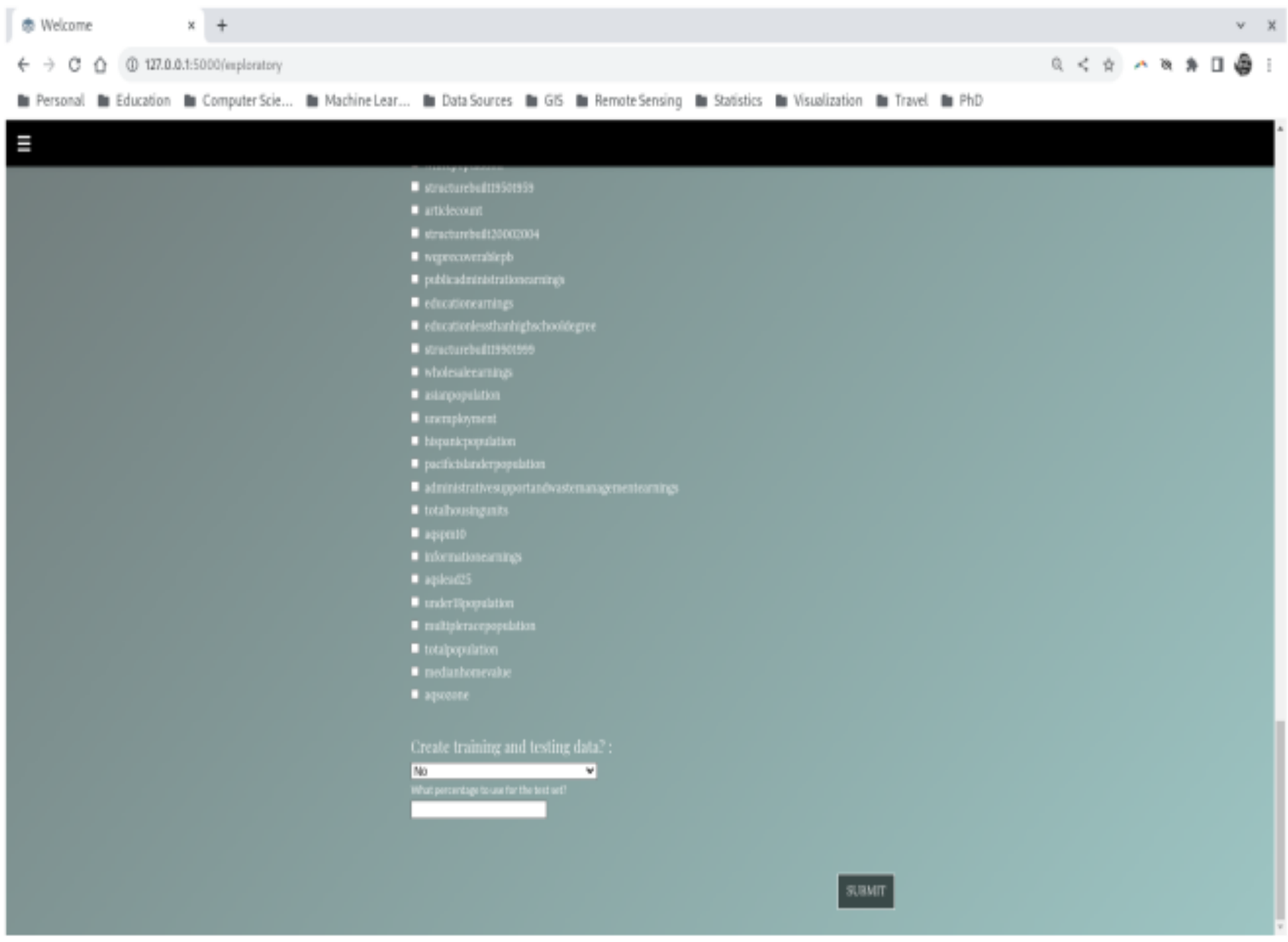


13. After selecting the database table of interest, the user can press 'SELECT" to see the head of the database table. The user can also scroll down to select the dependent variable, any transformations, and which independent variables to use. Selecting 'All' independent variables will disable the mapping function because it is designed as a hypothesis test. This allows the user to re-run the analysis with only the independent variables of interest.

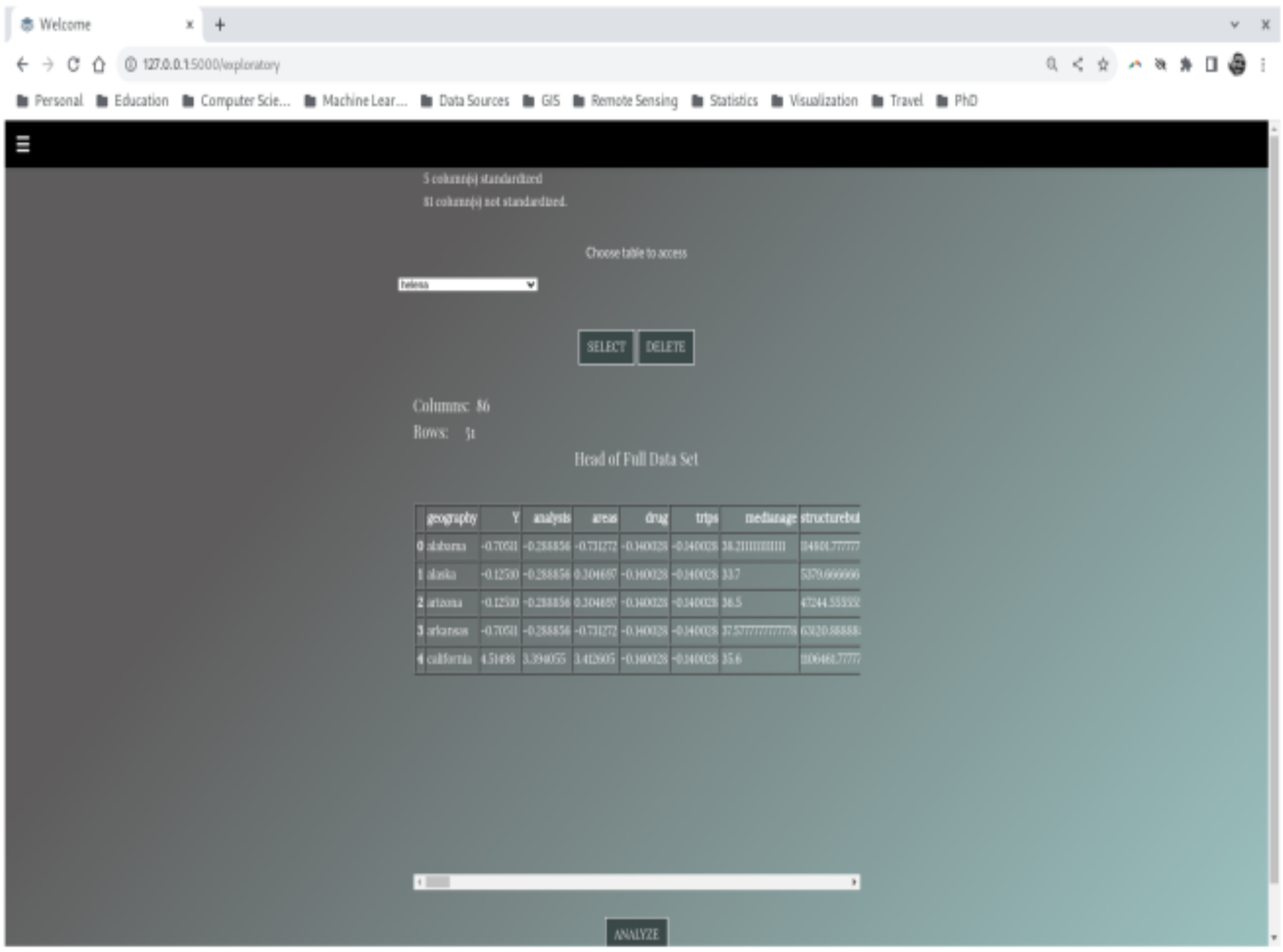




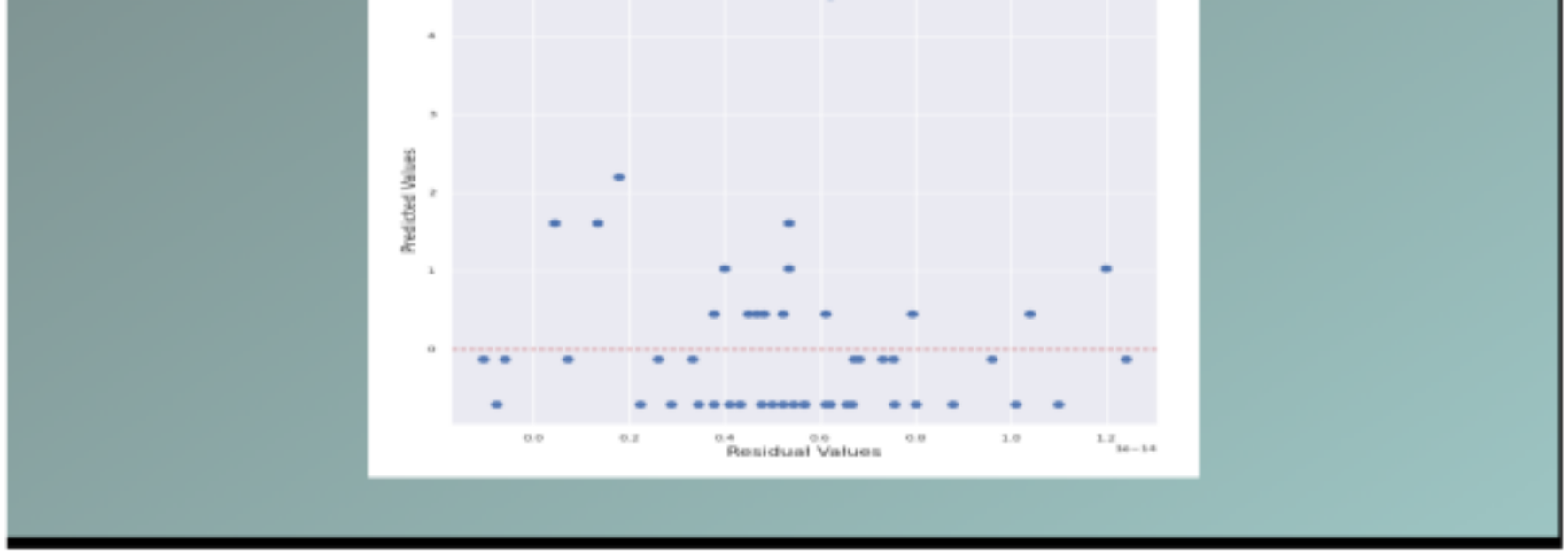




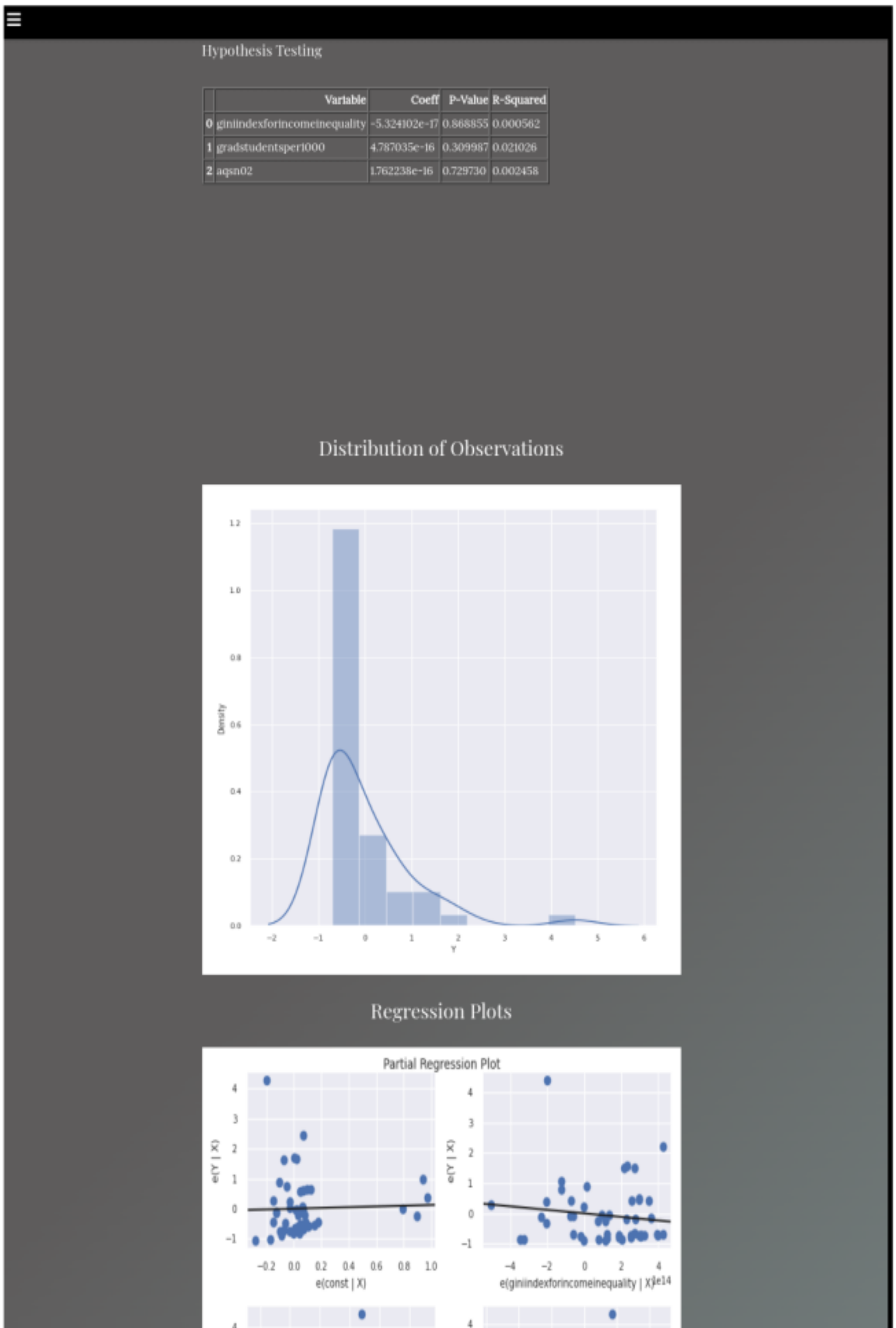
14. A verification screen is next displayed to the user. This allows the user to determine that the variables selected and the transformations selected are correct. The 'Submit' button proceeds to the full analysis visualization page.



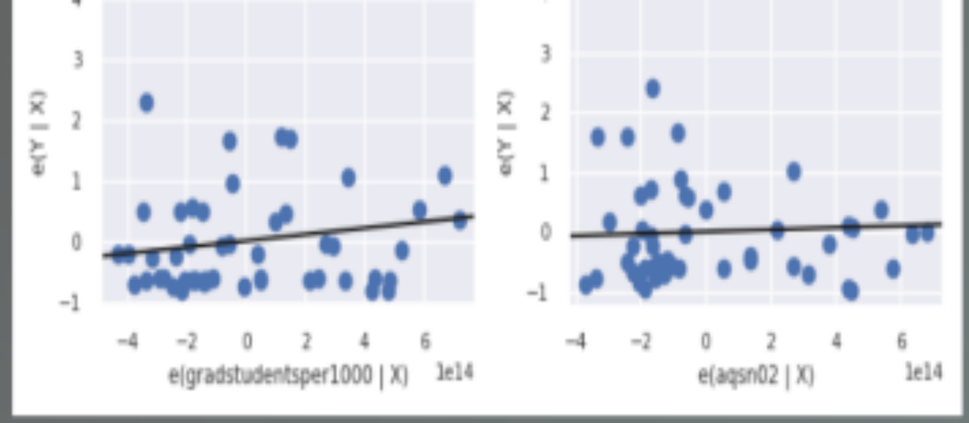
15. The analysis visualization page displays a hypothesis test table, the distribution of the observations, all of the regression plots, and residual distribution plots. Too many variables will disable the map and make the regression plots impossible to read.



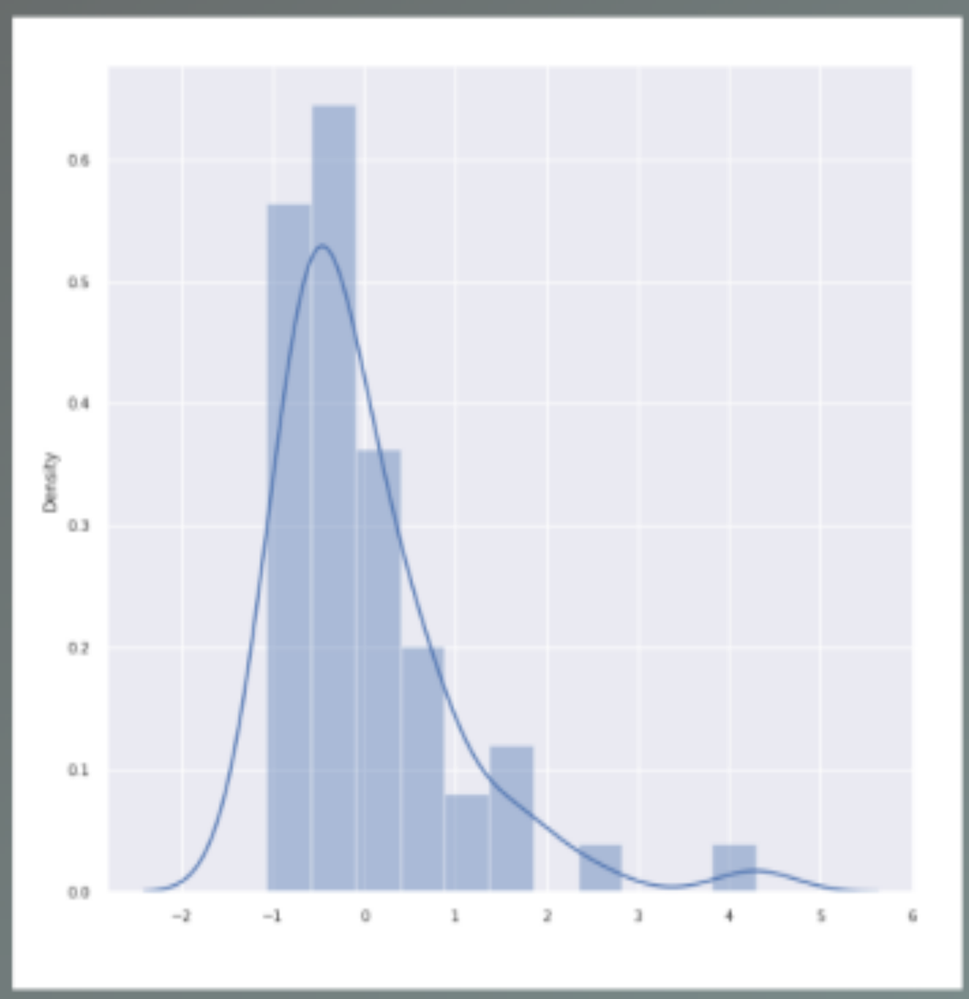
16. Finally, re-running the anlalysis with less variables will display a readable regression plot and a map of the estimated values. These are in addition to the other plots shown in the previous analysis visualization page.



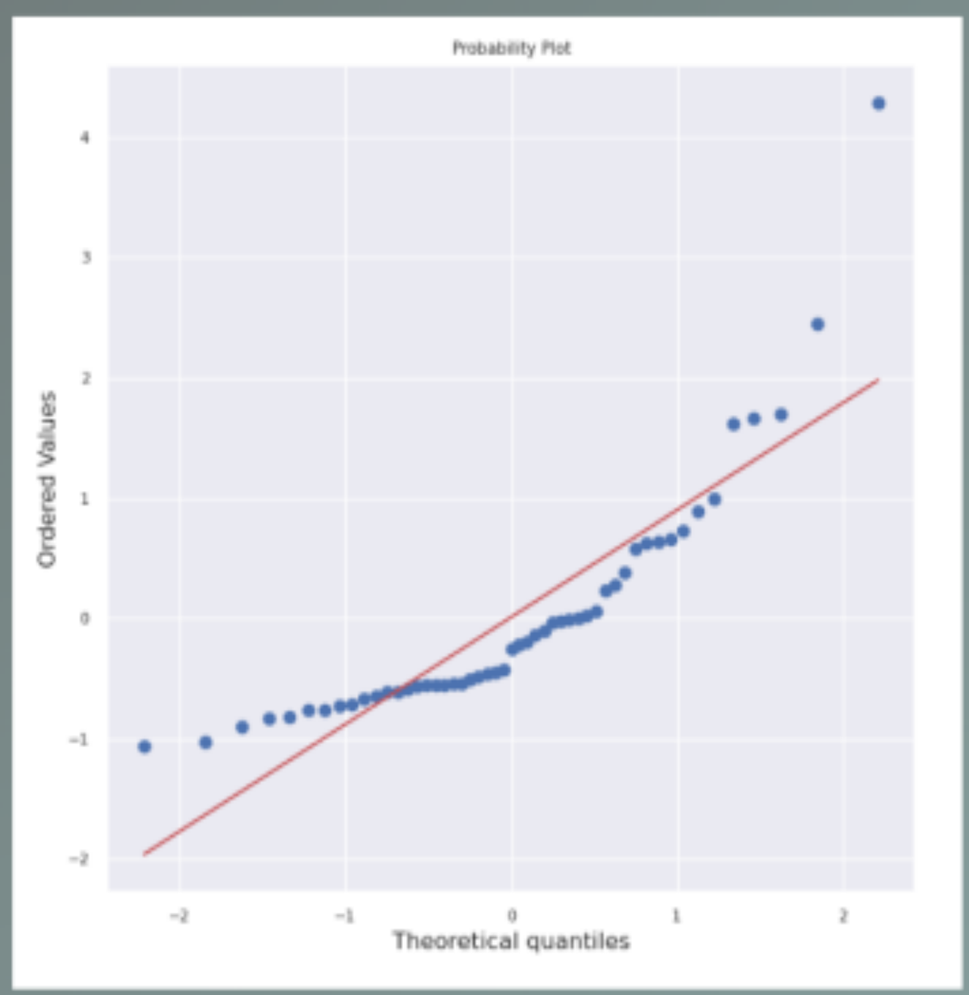
	Variable	Coeff	P-Value	R-Squared
0	giniindexforincomeinequality	-5.324102e-17	0.868855	0.000562
1	gradstudentsperi000	4.787035e-16	0.309987	0.021026
2	aqsn02	1.762238e-16	0.729730	0.002458



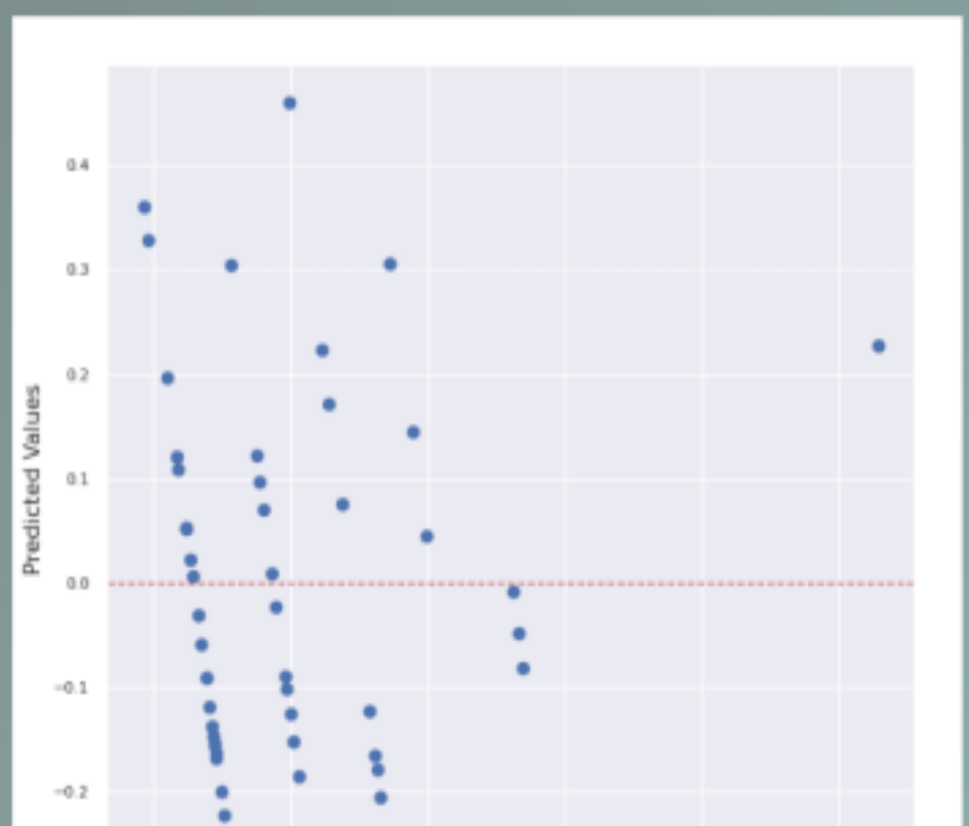
Residual Distribution

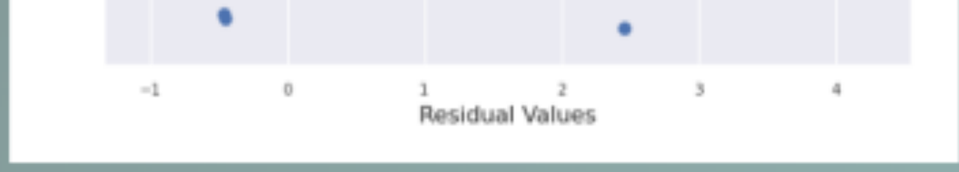


Quantile-Quantile Plot

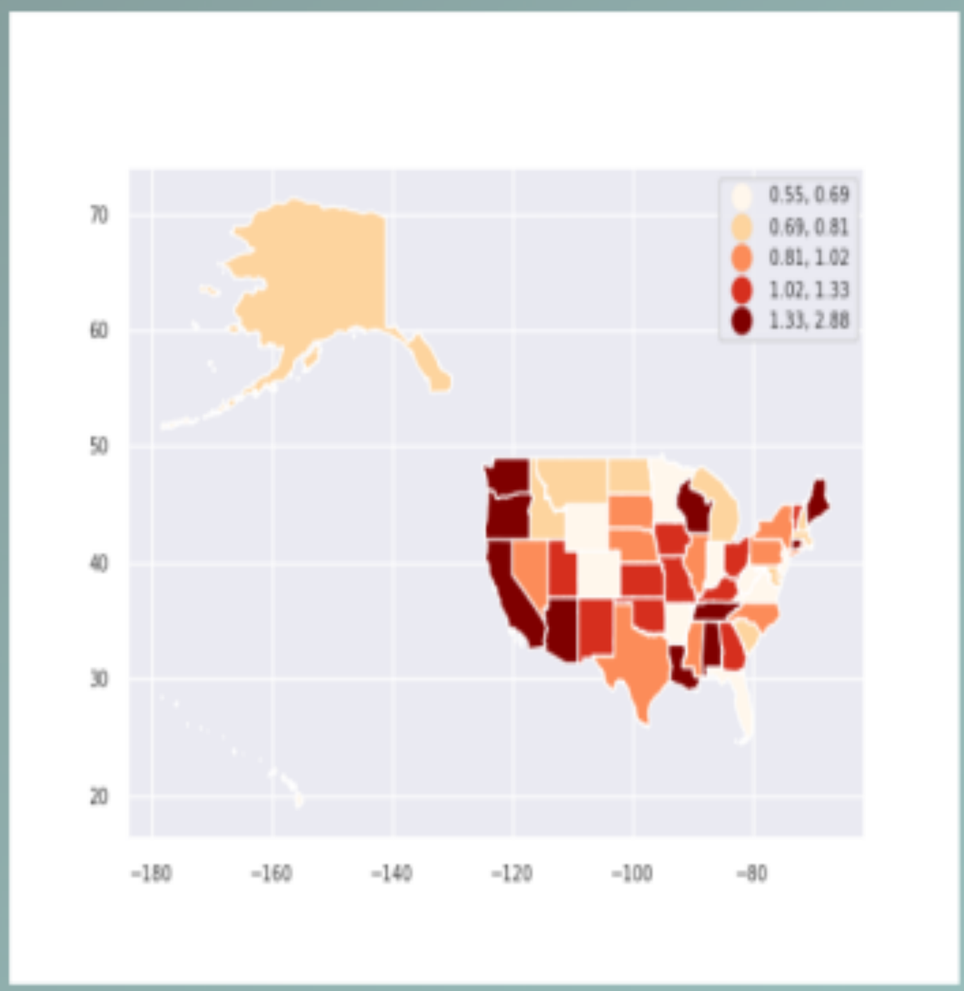


Predicted Values Versus Residual Values





Shapefile Plot



Known Issues:

Known issues include 1) reliance on the user to upload data in a specific format; 2) initial processing relies on the user inputs; 3) stopword selection is not available; 4) database issues; 5) only a single model (OLS) is available; 6) total factor productivity and spatial autocorrelation steps are not included; 7) improvements on the graphics are needed. However, the prototype is a functioning and performing application as specified, and it can be extremely beneficial to research and decision making, the field of innovation, science, and written communication.

Of the listed issues, the stopword selection and the database issues are the two most egregious problems. The stopword selection will require an additional step that allows the user to upload a stopword list. An alternative is for users to type in stopwords and then append the user entered stopwords to the generic stopword list provided by the geoparser and sub-topic clustering application. Without this step, users are required to use a generic stopword list that isn't specific to their topic of interest. The implications affect the sub-topic clustering piece of the analysis. The database issues are a recent development. The Python provided SQLite database was used for storing merged files at the end of the process between the 'Welcome Screen' and the 'Estimate Screen'. However, the program creates a new, unique table for each database submitted. Each table, being unique, has a different structure due to different features being analyzed. There are a few possible solutions. One is to create a form using the WTForms library, but the problem of table structure persists. Another solution is to do away with the database and keep the database in memory until the analysis is complete, but this will require users to go through the time consuming steps of geoparsing, topic clustering, and analysis multiple times for multiple analyses. As of now, this creates an occasional systematic error when the 'Proceed' button, on the 'Welcome Screen', is pressed. Both of these issues are significant, but they do not impede the program's performance as a prototype application.

Application Location:

Please e-mail me at nickgrok@gmail.com or call (561) 305-9239 so that I may make the application live for your use. Currently, the application is kept dormant for security reasons.

https://42f1-136-56-95-23.ngrok-free.app/login
Or, you can go to nickgrok.com, click the hamburger menu, and select the 'Application' button