



Diversity in Higher Education Using Python

Dana Lucas

HON 350 (Using coursework from PHY 299 and HON 300)



Description of the Project

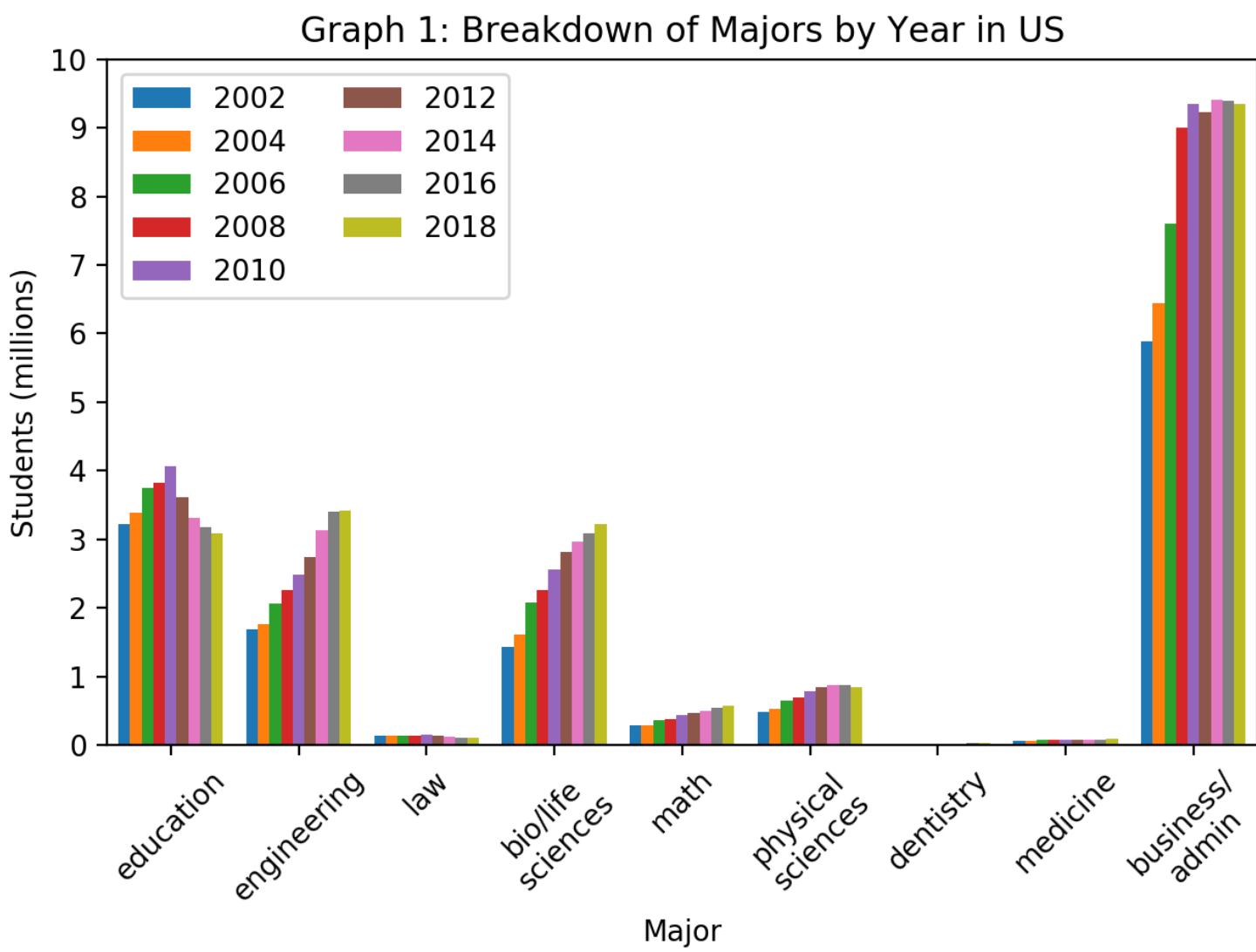
This project examines data taken from the National Center for Education Statistics (NCES). NCES provides data from 6858 schools across the US on the number of students enrolled in certain college majors over the years ranging from 2002 to 2018 and the racial background of these students. Code was developed to process this large amount of data to analyze statistics such as how students' chosen majors have changed over the past decade and how this is related to factors such as their racial background and choice of school. This project allowed me to efficiently use Python to process large amounts of data and apply some of the programming techniques learned in PHY 299. Upon processing the data, I expected to see that majors in the STEM field and health profession show a steady increase over time as technology and the need for scientists becomes a priority for society. Additionally, as minorities are given more opportunities to succeed in society, I expected diversity in college statistics to increase as well over the years. The results of the processed data were interpreted, and the process used to gather these results was explained. Some of the problems encountered and troubleshooted in the process were also recorded.

Code

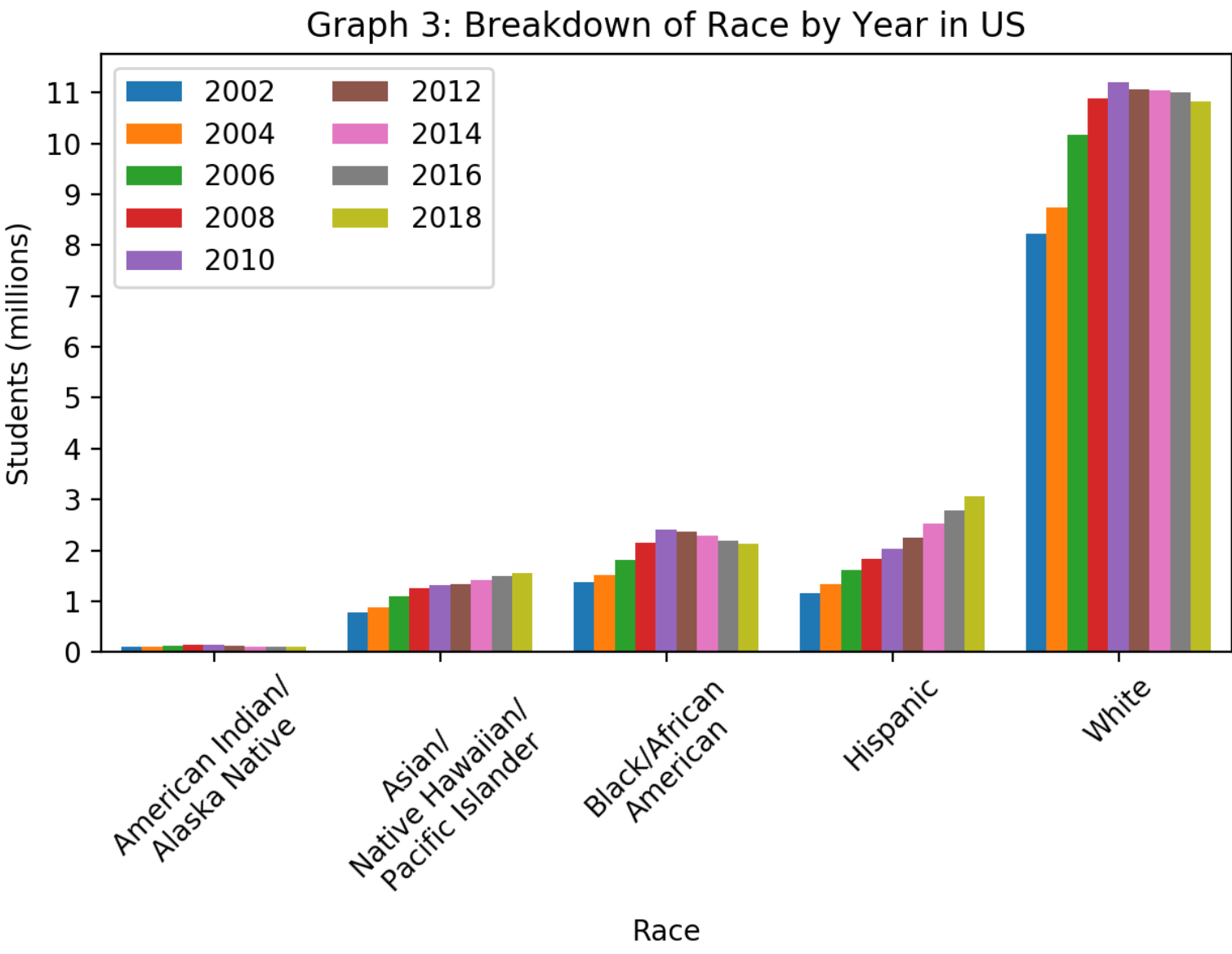
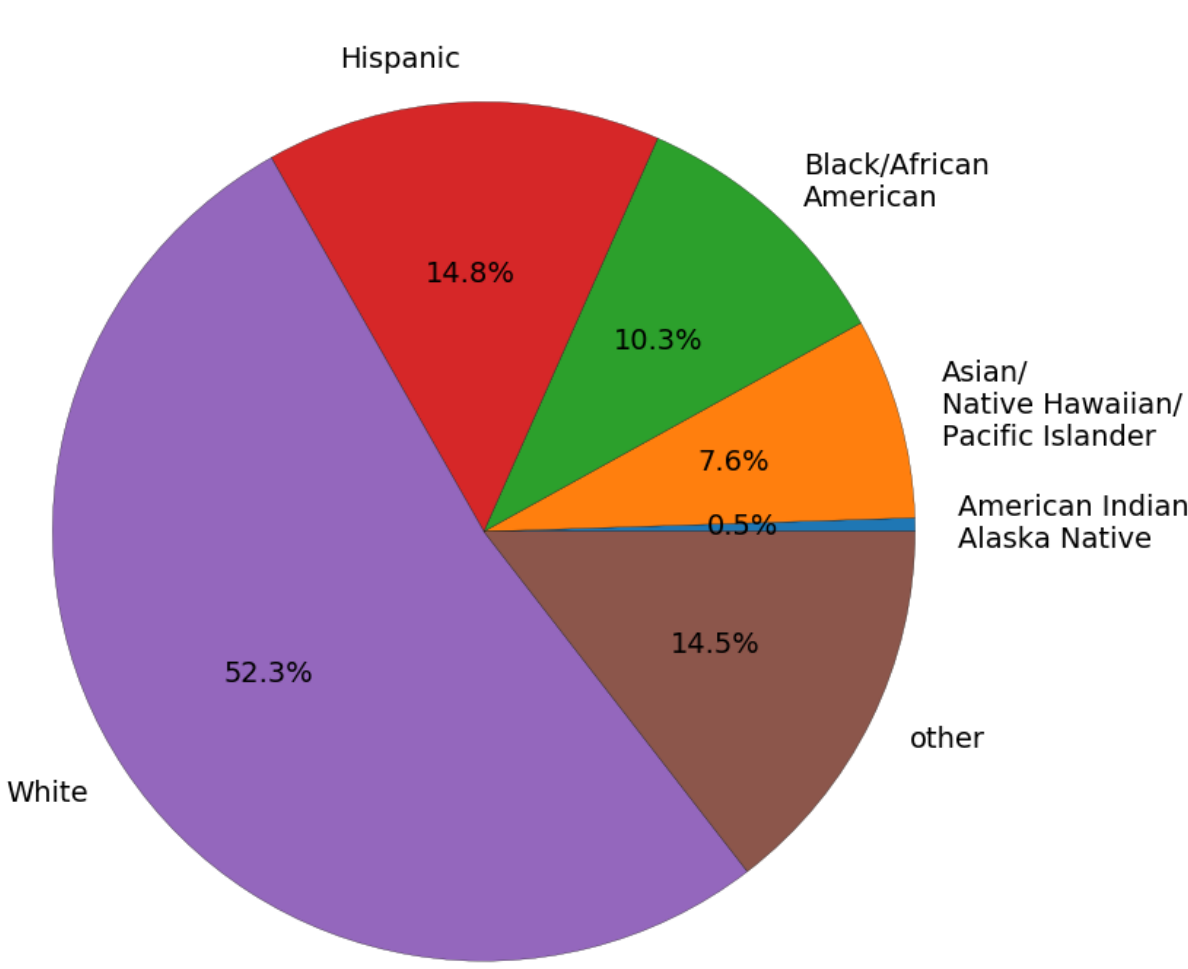
A portion of the code used to process this data is shown below. A full version of the code, the Excel files [3], and additional graphs can be found at [this link](#) [2].

```
33
34
35     if frame in frame_list1:
36         # Don't read the first line, which has the labels
37         if s[0] == 'UNITID': continue
38
39     else:
40         modifiedline = []
41
42         # These are the columns that I wish to extract
43         # data from; all the other columns are useless
44         # and unnecessary so I'll leave them out
45         listofcolumns = [0,2,7,9,11,13,25,31,43]
46
47         # These are all strings by default; convert to
48         # numbers so the statistics can be worked with
49         s[4] = int(s[4])
50         s[19] = int(s[19])
51         s[37] = int(s[37])
52         for num in listofcolumns:
53             if num == 2:
54                 s[num] = float(s[num])
55             else:
56                 s[num] = int(s[num])
57
58         # Reduce the needed lines down to
59         # only the needed columns
60         if s[4] == 3:
61             modifiedline.append(s[num])
62
63         # This format does not combine Asian and
64         # Pacific Islanders, so do it manually
65         modifiedline.append(s[19]+s[37])
66
67         # Form a matrix that contains only the
68         # desired rows and columns
69         if s[4] == 3:
70             matrix.append(modifiedline)
71
72
187
188 # Calculate yearly stats of majors
189 total += D[frame][j][2]
190 mc = D[frame][j][1]
191 st = D[frame][j][2]
192 if mc == 13: t_edu += st
193 if mc == 14: t_ege += st
194 if mc == 22.0101: t_law += st
195 if mc == 26: t_bio += st
196 if mc == 27: t_mth += st
197 if mc == 40: t_phy += st
198 if mc == 51.0401: t_dty += st
199 if mc == 51.1201: t_med += st
200 if mc == 52: t_bus += st
201
202 # Calculate breakdown of race among business majors per year
203 if D[frame][j][1] == 52:
204     bus_aminid += D[frame][j][5]
205     bus_aa += D[frame][j][6]
206     bus_his += D[frame][j][7]
207     bus_white += D[frame][j][8]
208     bus_ashawpi += D[frame][j][9]
209     bus_male += D[frame][j][3]
210     bus_female += D[frame][j][4]
```

Results



Graph 2: Race Breakdown of US Students in 2018



Scientific Modeling Analysis

Python Skills Used in Code [1]

- Python commands such as For Loops and If Statements were used to access only the necessary rows and columns.
- Each spreadsheet lists the data for one year. The glob module was used to extract all the Excel spreadsheets in a given range of years. This allowed for simplification of the code and allowed for multiple years worth of data to be easily analyzed at once.
- Dictionaries, which can organize data in Python, were used to allow for any piece of data to be accessed quickly at any time. A hierarchy of three dictionaries were used to describe the year the data was coming from, the line from the Excel spreadsheet that was being accessed and the data value of the race or major on that specific line.
- Lists of data such as the race of students in each major and the majors of all students enrolled in higher education were used to create plots (bar graphs and pie charts) in order to visualize the data. Titles, labels and legends were added to each graph, and the font size and location was modified to enhance readability.

Troubleshooting

- The format of some of the Excel files varied throughout the years, so modifications had to be made to account for the changes. Some of these adjustments included using a different column number if the columns were switched around or adding two columns together if a format separated two values that should have been together.
- Because each spreadsheet contained a massive amount of rows and columns (typically around 160,000 by 65), the numbers used to label each row and column had to be carefully tracked. At one point, a value of 2 was used rather than a value of 1 to pick out all the business majors, which caused the total number of business students to be only a few thousand rather than a few million.
- When the program was run, an error occasionally appeared that pointed out that a string (text) could not be converted to a float (a number with a decimal point). This issue was resolved by redownloading and replacing the Excel spreadsheet.

Bibliography

1. Hill, Christian. Learning Scientific Programming With Python. Cambridge University Press, 2015.
2. Lucas, Dana. Lucas HON 350 (2020). https://mountunion-my.sharepoint.com/:f/g/personal/lucasdm2022_mountunion_edu/EmgL64ZHUEplg4dJinfrP_8BliGF3lIscnv-jYjO5bTlug?e=UBmrSb
3. National Center for Education Statistics. Major field of study, race/ethnicity, gender, attendance status, and level of student: Fall 2018 (n.d.). Retrieved January 27, 2020, from <https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx>
4. United States Census Bureau. QuickFacts (2019). Retrieved March 16, 2020, from <https://www.census.gov/quickfacts/fact/table/US/IPE120218>

Diversity Analysis

Diversity in Major

- As technology advances, there is a greater focus on STEM education. In my data, it was expected to see an increase in STEM majors over the past decade and for all other majors to level out to a constant value. By analyzing Graph 1 it is seen that there was a steady, slightly quadratic increase of engineering majors and linear increase in other STEM majors throughout the past two decades.
- Business majors showed a high increase in the 2000s but then leveled out in the 2010s. Education majors peaked in 2010 and are now decreasing at a decreasing rate, suggesting that they will level out within the next couple years. Both of these majors are still highly important to society, which explains why the students enrolled in them has begun to level out rather than continue to decrease. The societal focus on STEM, however, explains why interest has not increased.

Diversity in Race

- In 2018, around half of students enrolled in higher education across the US were white. According to the US Census Bureau, race demographics show that 76.5% of the US population was of the white race [4]. Graph 3 shows that the amount of minority students, Hispanics in particular, have been increasing in the past decade while the amount of whites slowly decreased. Therefore, there is more diversity among students pursuing higher education than there is diversity in the United States.
- Within the engineering major, the amount of Asian and Hispanic students have been exponentially increasing while the amount of African American and white students have increased linearly. Within education, the amount of Asian and Hispanic students have been increasing while African American and white students peaked in 2010 and have been decreasing. Within business, the amount of white students enrolled seems to have leveled out in the past decade and African American students peaked and are now decreasing. Asian and Hispanic students show a steady increase.

Conclusions

- ❑ This amount of data would be extremely difficult to process by hand. The code, once created, allows all the information to be processed in about 20 seconds.
- ❑ STEM majors have been increasing in the past decade, showing society's increased focus on this area.
- ❑ Hispanics have been showing an increased presence within all majors in the past 16 years, allowing for more diversity within higher education institutions.
- ❑ The amount of students enrolled in business and the presence of whites in higher education institutions have leveled out over the past decade, showing that diversity is increasing and that students are taking greater interest in STEM disciplines.