College Index

2024-08-14

```
install.packages("ggplot2", repos = "http://cran.us.r-project.org")
```

```
## Installing package into '/Users/swain/Library/R/arm64/4.4/library'
## (as 'lib' is unspecified)
```

```
##
## The downloaded binary packages are in
##  /var/folders/0s/0md080fx6xxg3fvxdh43mm_r0000gn/T//RtmpSqiDVE/downloaded_packages
```

```
install.packages("dplyr", repos = "http://cran.us.r-project.org")
```

```
## Installing package into '/Users/swain/Library/R/arm64/4.4/library'
## (as 'lib' is unspecified)
```

```
##
## The downloaded binary packages are in
##  /var/folders/0s/0md080fx6xxg3fvxdh43mm_r0000gn/T//RtmpSqiDVE/downloaded_packages
```

```
install.packages("readr", repos = "http://cran.us.r-project.org")
```

```
## Installing package into '/Users/swain/Library/R/arm64/4.4/library'
## (as 'lib' is unspecified)
```

```
##
## The downloaded binary packages are in
##  /var/folders/0s/0md080fx6xxg3fvxdh43mm_r0000gn/T//RtmpSqiDVE/downloaded_packages
```

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
```

```
acc <- read_csv("acc_10_20.csv")
```

```
## Rows: 165 Columns: 26
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (6): instnm, stabbr, sector, region, locale, conference
## dbl (20): unitid, year, adm_rate, satvrmid, satmtmid, sat_avg, ugds, ugds_wh...
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
bigten <- read_csv("bigten_10_20.csv")
```

```
## Rows: 153 Columns: 26
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (6): instnm, stabbr, sector, region, locale, conference
## dbl (20): unitid, year, adm_rate, satvrmid, satmtmid, sat_avg, ugds, ugds_wh...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
outcomes <- read_csv("outcomes_10_20.csv")
```

```
## Rows: 318 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (8): unitid, year, c100_4, c150_4, c200_4, debt_mdn, cdr3, md_earn_wne_p6
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
setequal(names(acc), names(bigten))
```

```
## [1] TRUE
```

```
# Horizontally and vertically combining bigten, acc, and outcomes
college_data <- acc %>% bind_rows(bigten)

college_data <- college_data %>%
  inner_join(outcomes, by = c("year" = "year", "unitid" = "unitid"))
```

```
# Filtering out Wake Forest so we can use sat_avg
college_data <- college_data %>% filter(unitid != 199847)

# We will create an index of selectivity based on average SAT and admission rate
college_data %>% select(instnm, sat_avg, adm_rate)
```

```
## # A tibble: 307 x 3
##    instnm                   sat_avg adm_rate
##    <chr>                      <dbl>    <dbl>
##  1 Florida State University    1192    0.595
##  2 Florida State University    1203    0.585
##  3 Florida State University    1214    0.537
##  4 Florida State University    1212    0.568
##  5 Florida State University    1228    0.554
##  6 Florida State University    1215    0.559
##  7 Florida State University    1214    0.580
##  8 Florida State University    1304    0.492
##  9 Florida State University    1289    0.368
## 10 Florida State University    1284    0.360
## # i 297 more rows
```

```
# Adding standardized variables for average SAT and admission rate
college_data <- college_data %>% group_by(year) %>% mutate(across(c(sat_avg, adm_rate), scale, .names =
```

```r
# Adding a variable for a selectivity index where higher SATs and lower admission rates produce a highe
college_data <- college_data %>%
  mutate(selec_index = 1*sat_avg_stnd - 1*adm_rate_stnd)

# Showing average SAT, admission rate, and selectivity index for 2020
college_data  %>%
  filter(year == 2020) %>%
  select(instnm, sat_avg, adm_rate, selec_index) %>%
  arrange(desc(selec_index))
```

```
## Adding missing grouping variables: `year`
```

```
## # A tibble: 27 x 5
## # Groups:   year [1]
##     year instnm                                sat_avg adm_rate selec_index[,1]
##    <dbl> <chr>                                   <dbl>    <dbl>           <dbl>
##  1  2020 Duke University                          1530   0.0774           3.75
##  2  2020 Northwestern University                  1505   0.0931           3.41
##  3  2020 University of Notre Dame                 1489   0.190            2.83
##  4  2020 Georgia Institute of Technology-Main ~   1465   0.213            2.46
##  5  2020 University of Michigan-Ann Arbor         1445   0.261            2.04
##  6  2020 University of Virginia-Main Campus       1426   0.226            1.98
##  7  2020 Boston College                           1433   0.264            1.90
##  8  2020 University of North Carolina at Chape~   1393   0.25             1.52
##  9  2020 University of Miami                      1354   0.331            0.748
## 10  2020 University of Maryland-College Park      1391   0.510            0.401
## # i 17 more rows
```

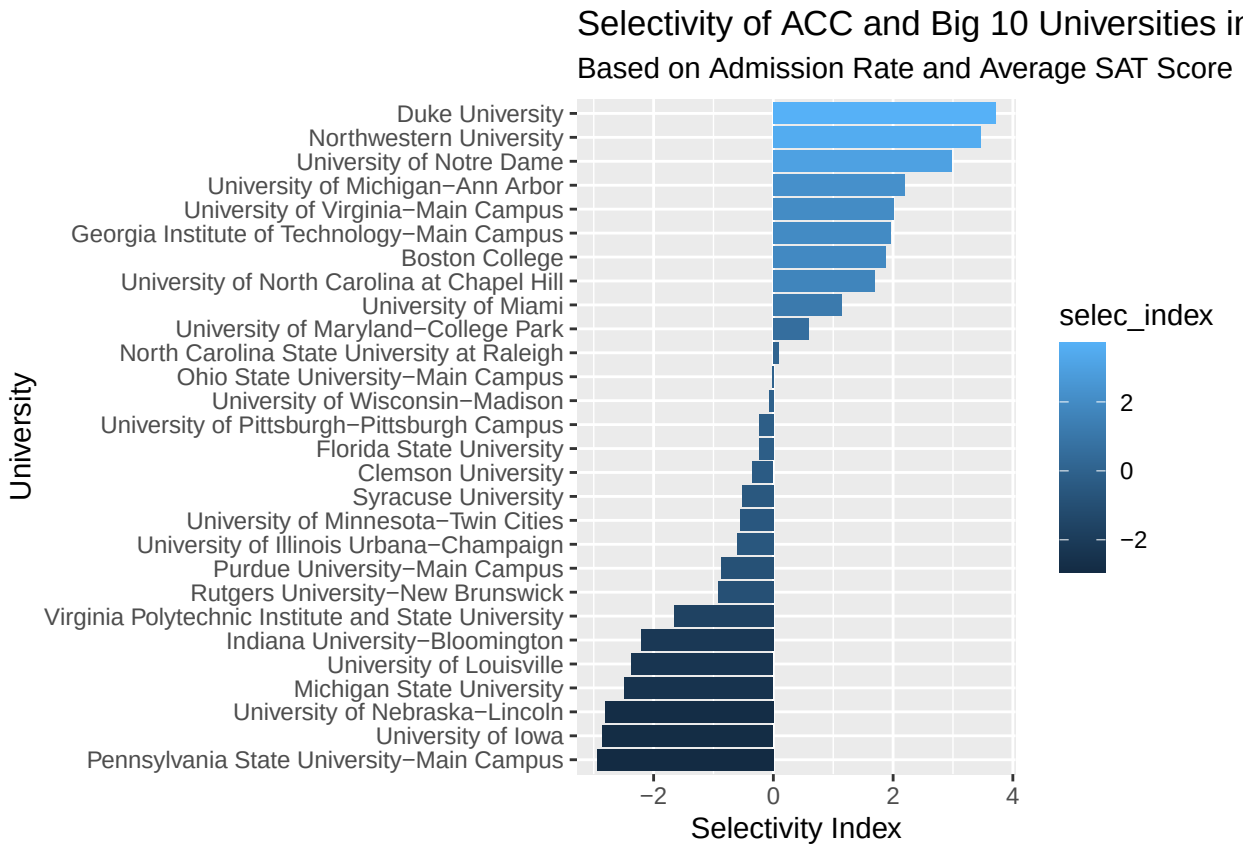
```r
# Creating a selectivity ranking variable
college_data <- college_data %>% group_by(year) %>% mutate(selec_rank = dense_rank(desc(selec_index)))

# Showing the ranking with its components
college_data %>% filter(year == 2019) %>% select(instnm, sat_avg,adm_rate, selec_index, selec_rank) %>%
```

```
## Adding missing grouping variables: `year`
```

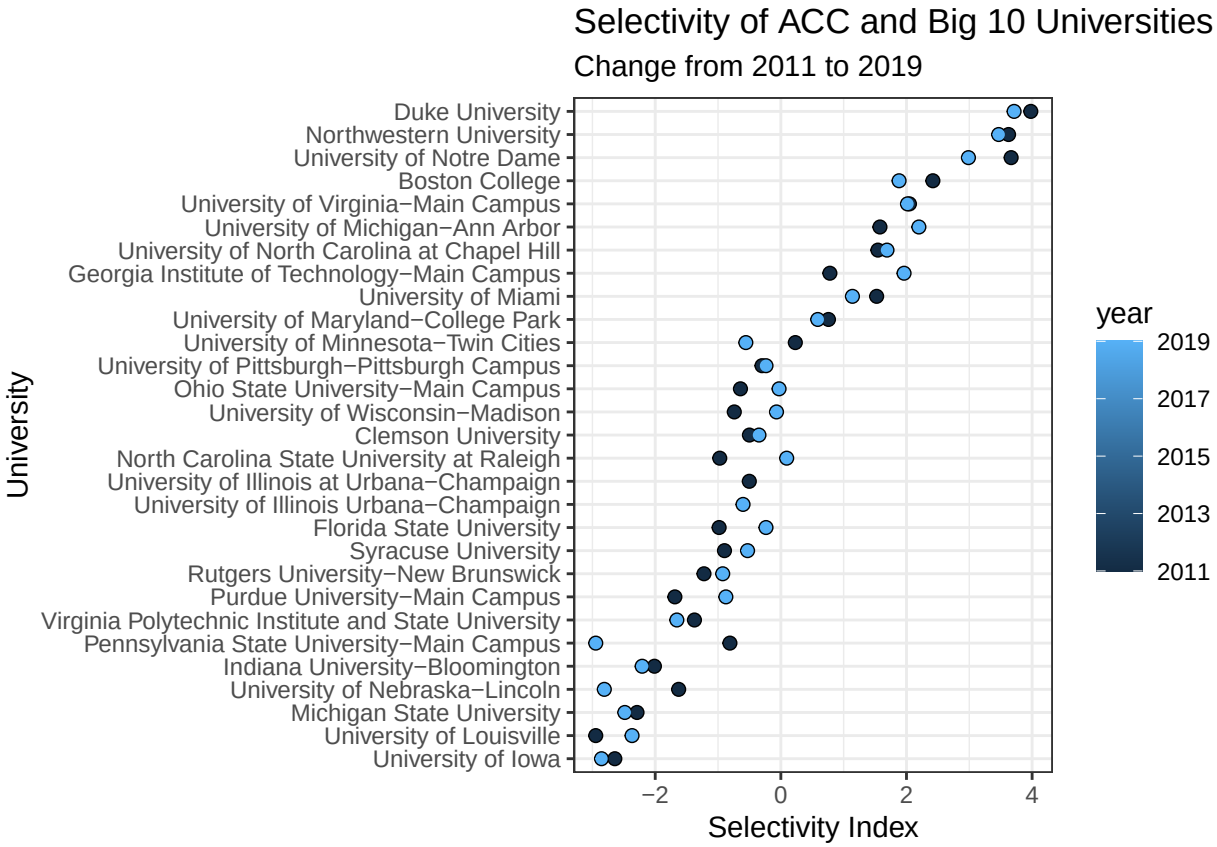
```
## # A tibble: 28 x 6
## # Groups:   year [1]
##     year instnm                          sat_avg adm_rate selec_index[,1] selec_rank
##    <dbl> <chr>                             <dbl>    <dbl>           <dbl>      <int>
##  1  2019 Duke University                    1522   0.076            3.71           1
##  2  2019 Northwestern University            1506   0.0905           3.47           2
##  3  2019 University of Notre Dame           1490   0.158            2.99           3
##  4  2019 University of Michigan-Ann~        1448   0.229            2.20           4
##  5  2019 University of Virginia-Mai~        1436   0.239            2.02           5
##  6  2019 Georgia Institute of Techn~        1418   0.206            1.96           6
##  7  2019 Boston College                     1437   0.272            1.88           7
##  8  2019 University of North Caroli~        1402   0.226            1.69           8
##  9  2019 University of Miami                1371   0.271            1.14           9
## 10  2019 University of Maryland-Col~        1389   0.442            0.588         10
## # i 18 more rows
```

```r
# Let's use the University of Minnesota because they have the best name, the Golden Gophers

# Bar graph -- I chose a different year because Syracuse didn't publish SAT data for 2020
ggplot(data=filter(college_data, year == 2019), aes(x = reorder(instnm,selec_index), y=selec_index, fil
```

```
  geom_bar(stat="identity") + coord_flip() + labs(x = "University", y = "Selectivity Index", title = "Se
```




```
# Again changing 2020 to 2019 to accomodate Syracuse

ggplot(data=filter(college_data, year==2011 | year == 2019), aes(x = reorder(instnm,selec_index), y = se
```


Selectivity of ACC and Big 10 Universities
Change from 2011 to 2019
Duke University
Northwestern University
University of Notre Dame
Boston College
University of Virginia−Main Campus
University of Michigan−Ann Arbor
University of North Carolina at Chapel Hill
Georgia Institute of Technology−Main Campus
University of Miami
University of Maryland−College Park
University of Minnesota−Twin Cities
University of Pittsburgh−Pittsburgh Campus
Ohio State University−Main Campus
University of Wisconsin−Madison
Clemson University
North Carolina State University at Raleigh
University of Illinois at Urbana−Champaign
University of Illinois Urbana−Champaign
Florida State University
Syracuse University
Rutgers University−New Brunswick
Purdue University−Main Campus
Virginia Polytechnic Institute and State University
Pennsylvania State University−Main Campus
Indiana University−Bloomington
University of Nebraska−Lincoln
Michigan State University
University of Louisville
University of Iowa
University
−2
0
2
4
Selectivity Index
year
2019
2017
2015
2013
2011