US Colleges Analysis Report

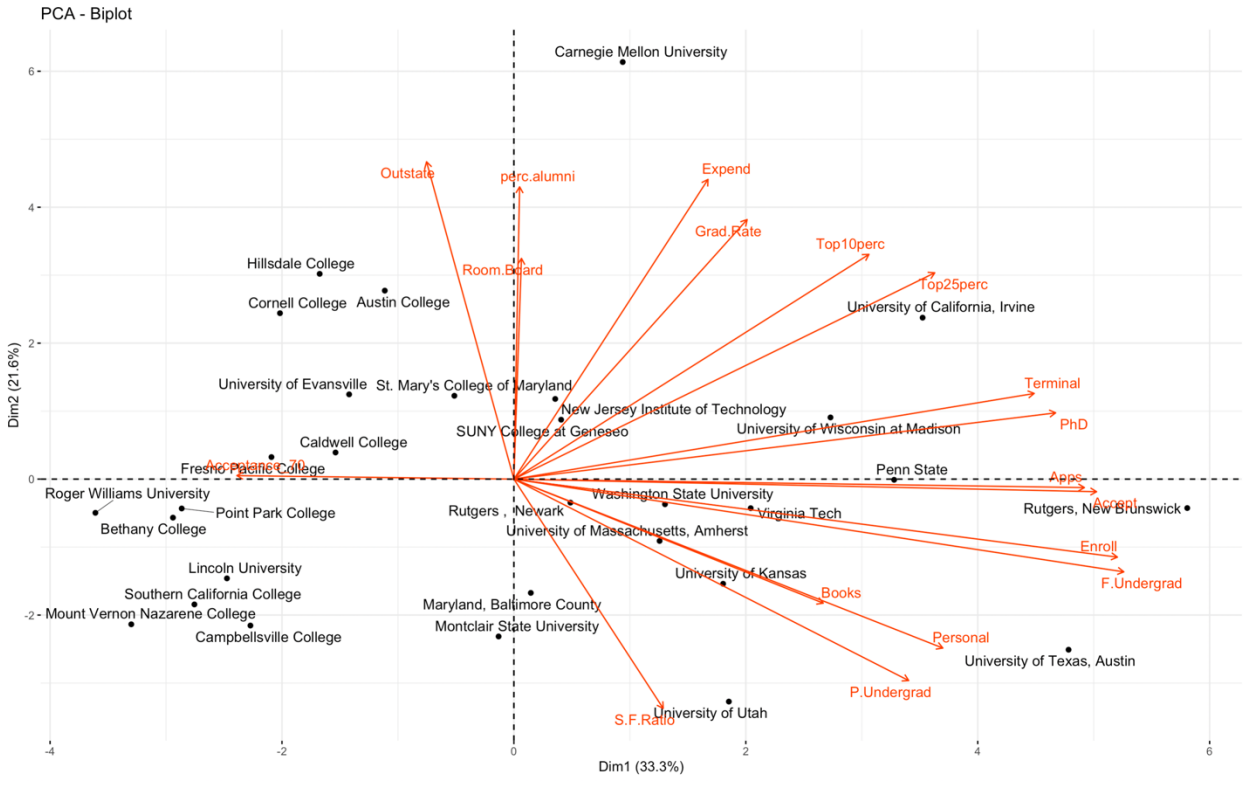
Detailed explanation on GitHub: https://github.com/ajeyvishnu/US_Colleges_Analysis

Hypotheses:

1. More students prefer Public universities over Private universities.
2. The costs associated with Public universities are less compared to Private universities.
3. Public universities have better acceptance rates compared to Private universities.
4. The differentiation between Public and Private institutes can be classified using Multivariate Analysis techniques.

Questions and answers based on analysis:

1. **Investigate the relationships between the universities based on given variables using Principal Component Analysis. Do we find any relations, and can we analyse if there is any difference between the Private and Public Universities?**

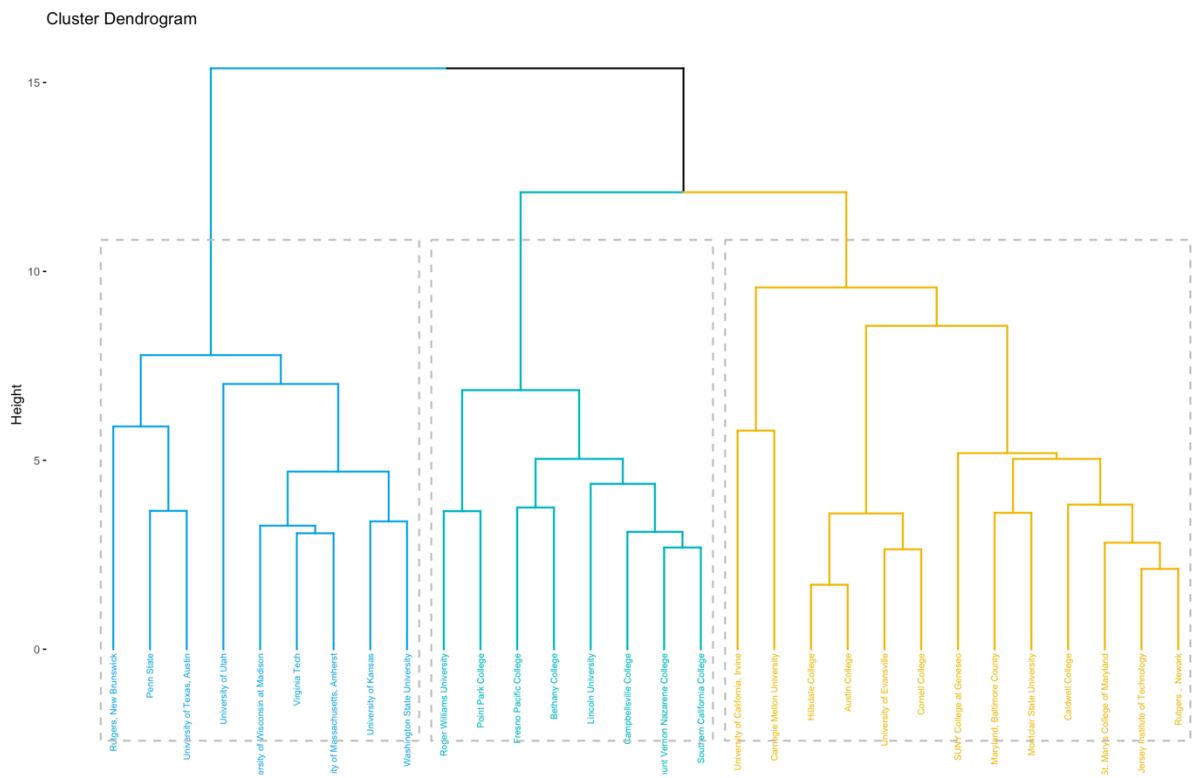


- The Principal component analysis has classified all the public universities to the right side of the Y axis and all the private ones to the left.
- Yes, we can classify the type of university based on the principal component analysis.

- We can also see that the number of applications, acceptances, and enrollments is more for Public universities, as per our hypothesis.
- But, contradicting our hypothesis, the costs incurred by students are more for Public universities compared to private ones.

2. Carry out cluster analysis to study the relationship between the universities and find out if there are any reasons behind the clustering. Is the clustering done based on the type of university or something different?

- We observe that all the universities in Cluster 3 (blue) are Private universities.
- Except two of the Private universities rest, all fell under Cluster 3.
- The Public Universities are placed in the remaining two clusters.
- The ones in Cluster 1 (Red) have many applications, acceptances and enrollments (in thousands).
- Cluster 2 (green) is also of Public universities but the ones with fewer applications, acceptances, and enrollments (in hundreds).

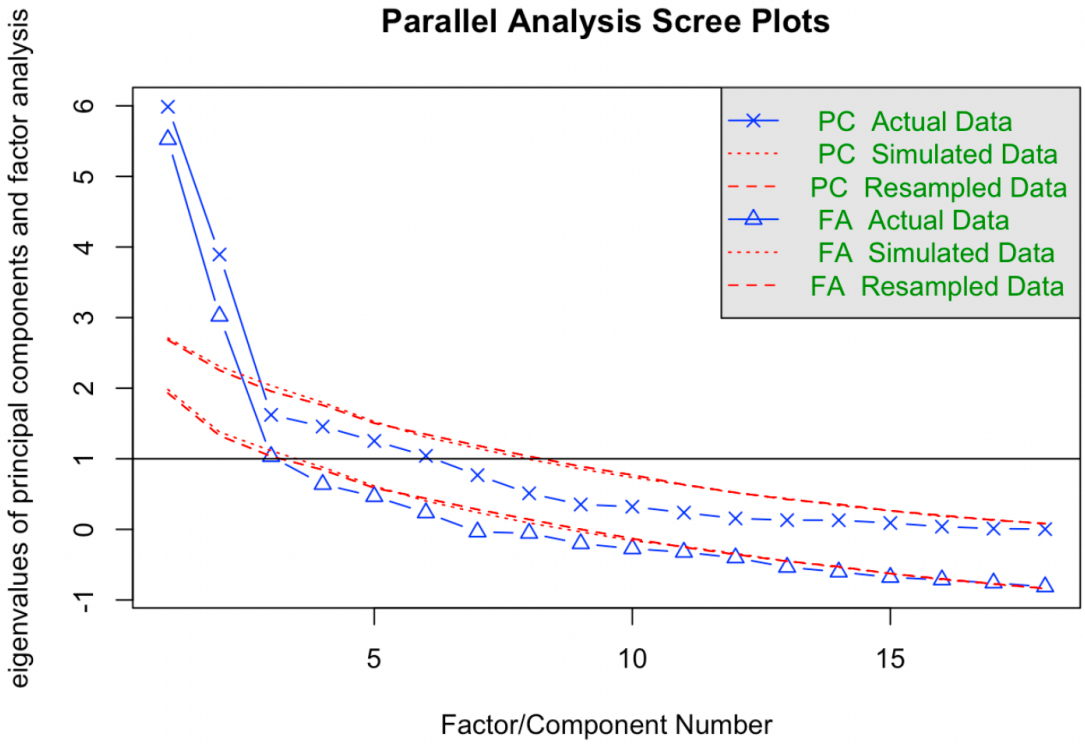


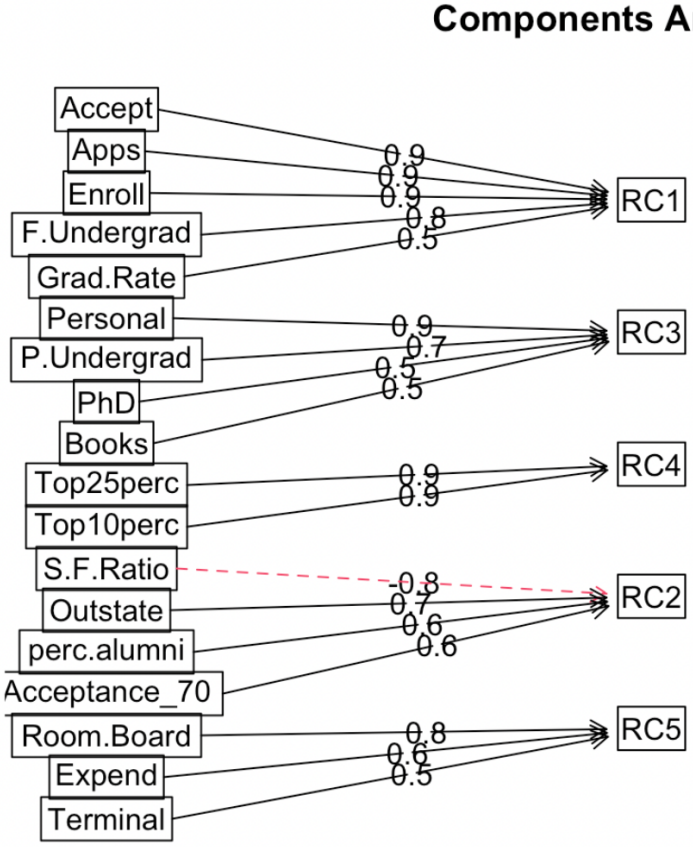
- In this clustering, all the universities in teal colour are Private universities.
- The universities in Blue are all Public, and the ones in the Yellow cluster are a mixture of public and private universities.
- Even using this method, we can somewhat differentiate between public and private.
- The Blue and Yellow differentiation mostly focuses on the number of applications, acceptances, and enrollments. The ones in blue have much higher numbers when compared to the ones in yellow.

3. Identify the important factors underlying the observed variables and examine the relationships between the universities for these factors. Check if the factors help us classify the type of university (Public/Private)

```
##                                          RC1         RC3         RC4
## Virginia Tech                     1.485468071  0.21131139 -0.21998370
## Maryland, Baltimore County       -0.481006334  0.76030425 -0.31548733
## University of Kansas              0.177350813  1.40427674  0.02317481
## Rutgers, New Brunswick            3.133346443 -0.54184707  0.10004650
## New Jersey Institute of Technology -0.698565381  0.57433793  0.29395115
## Penn State                        1.263617108  0.06342282  0.84411017
## University of California, Irvine  0.283893077  0.27706042  2.08221597
## Rutgers ,  Newark                -0.677051736  0.74194350  0.27258849
## University of Utah               -0.356647575  3.34936020 -1.18308501
## University of Texas, Austin       1.271969226  1.44186502  0.74986789
## SUNY College at Geneseo           0.008424103 -1.52037815  1.98538731
## University of Wisconsin at Madison  1.516380130  0.26320285 -0.09742845
## St. Mary's College of Maryland   -1.124429195  0.14350357  1.14223565
## Washington State University       0.185200681  0.95130029  0.31005570
## Lincoln University               -1.161856724  0.18943920 -0.60705986
## Montclair State University       -0.377018822 -0.01890837 -0.20013527
## University of Massachusetts, Amherst  1.576944277 -0.15158429 -1.40007992
## Mount Vernon Nazarene College    -0.364193850 -1.23343747 -0.82760015
## Carnegie Mellon University       -0.224847119 -0.25735884  1.07996340
## Campbellsville College           -1.032269951 -0.40964800  0.33895063
## Hillsdale College                -0.206749033 -1.13226255  0.14398325
## Austin College                   -0.377970174 -0.66656689  0.56118337
## Fresno Pacific College           -1.061532078 -0.09982983  0.80092587
## Roger Williams University         0.162940212 -1.47084476 -2.46329998
## Point Park College               -0.114030259 -0.81216742 -1.83109281
## Bethany College                  -0.819570894  0.05679587 -0.65195750
## Caldwell College                 -0.859520115 -0.02187551 -0.34922095
## Southern California College      -0.792310218 -0.03837150 -0.97211902
## University of Evansville          0.021108340 -1.20027859  0.28912388
## Cornell College                  -0.357073020 -0.85276482  0.10078590
```

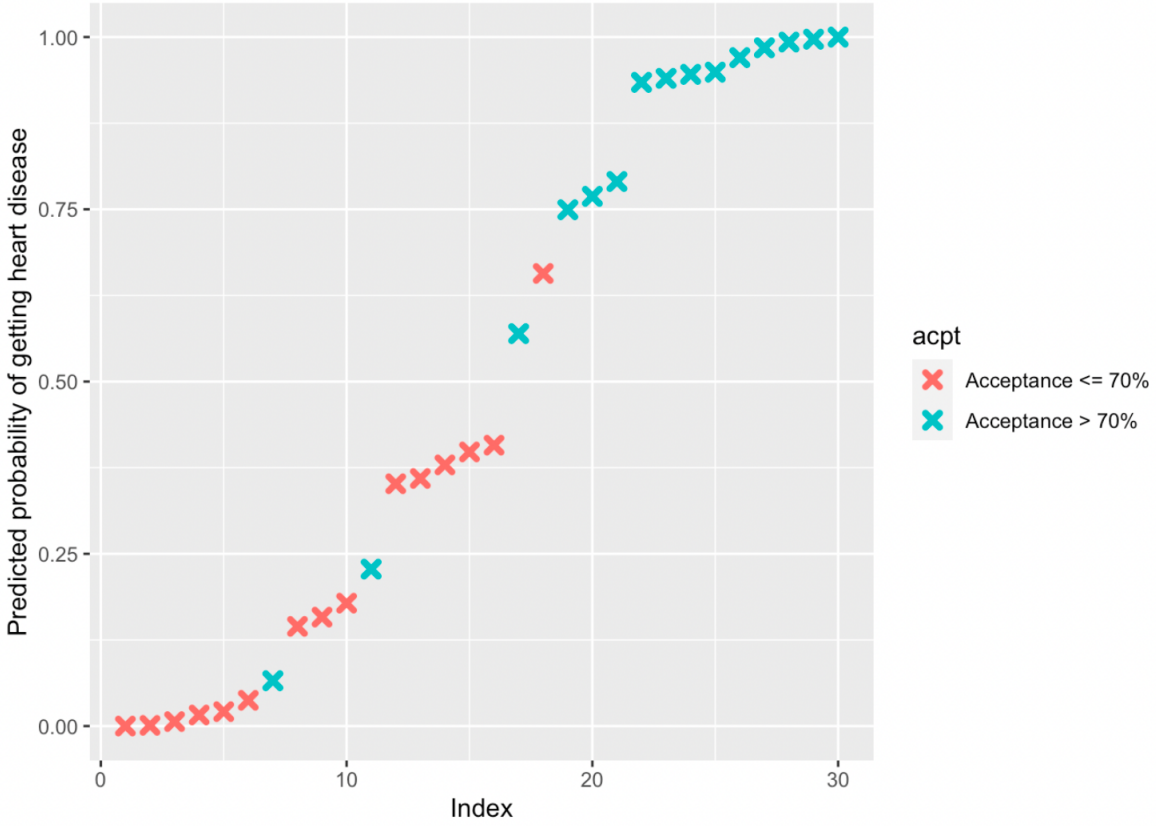
- From observation, all Private Universities had negative RC1 and RC4.
- Except for a few public universities, the rest have positive RC1 and RC4.
- We can further look at the fit.pc diagram to understand a better classification based on these values.

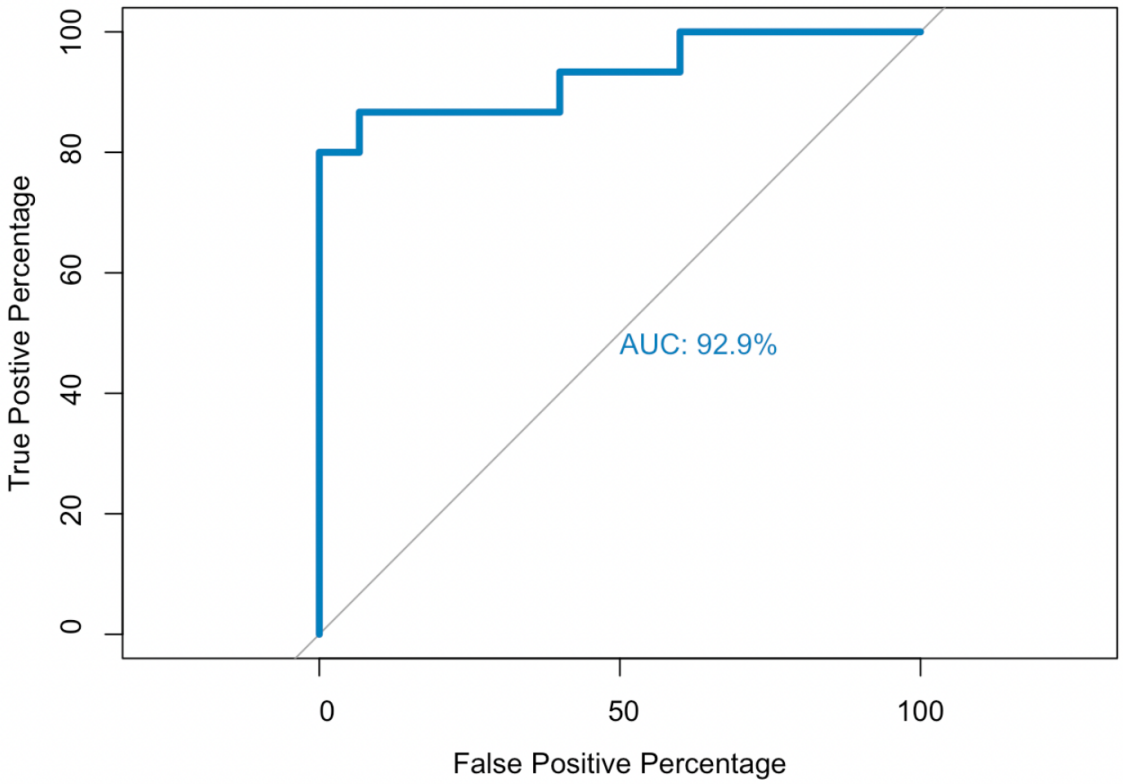

Parallel Analysis Scree Plots
eigenvalues of principal components and factor analysis
Factor/Component Number
PC Actual Data
PC Simulated Data
PC Resampled Data
FA Actual Data
FA Simulated Data
FA Resampled Data


Components Analysis
Accept
Apps
Enroll
F.Undergrad
Grad.Rate
Personal
P.Undergrad
PhD
Books
Top25perc
Top10perc
S.F.Ratio
Outstate
perc.alumni
Acceptance_70
Room.Board
Expend
Terminal
RC1
RC3
RC4
RC2
RC5

- The diagram shows that Applications, Acceptances, Enrollments, and Graduation rates come under RC1.
- This can be validated because all public universities had positive RC1, the exact inference we got in the PCA.
- The private column showed a strong negative value with RC4. It can be recollected that all Private colleges had negative RC4.
- Books and Personal expenses are related to RC3.
- This shows that even using EFA, we can classify between Public and Private universities.

4. Using regression, can we predict the acceptance rate of the universities based on all the factors provided for the universities? Which regression technique works best for this?





- The analysis shows that the Acceptance rate greater or less than 70% can be predicted using the college dataset variables.
- We have seen a 92.9% area under the curve which states the results are very good.
- And we prefer Logistics Regression, as the variable factors are only two greater or less than 70%.

SUMMARY

Hypotheses:

1. More students prefer Public universities over Private universities.
   Yes, the hypothesis is correct.
2. The costs associated with Public universities are less compared to Private universities.
   No, the hypothesis is proved wrong
3. Public universities have better acceptance rates compared to Private universities.
   Yes, the hypothesis is correct.

4. The differentiation between Public and Private institutes can be classified using Multivariate Analysis techniques.
   Yes, the hypothesis is correct.

Questions:

1. Investigate the relationships between the universities based on given variables using Principal Component Analysis. Do we find any relations, and can we analyse if there is any difference between the Private and Public Universities?
   Yes, we found the differences and were able to validate our hypothesis. And, we were able to differentiate between Public and Private universities.

2. Carry out cluster analysis to study the relationship between the universities and find out if there are any reasons behind the clustering. Is the clustering done based on the type of university or something different?
   Yes, we can classify Public and Private universities based on cluster analysis.

3. Identify the important factors underlying the observed variables and examine the relationships between the universities for these factors. Check if the factors help us classify the type of university (Public/Private)
   Important factors were identified, and we could classify Public and Private Universities.

4. Using regression, can we predict the acceptance rate of the universities based on all the factors provided for the universities? Which regression technique works best for this?
   Yes, we were able to predict. Logistic Regression works best in this case.