***Group Name:*** College\_Score\_Card\_E139-Fall2015

***Group Members:***

|  |  |
| --- | --- |
| Himanshu Dave | hdave@g.harvard.edu |
| Michael Prassanna Antony Raj | michael83@gmail.com |
| Sway Ambikash | swayambikash@gmail.com |
| Bin Xiao | bin.xiao@wustl.edu |

***Github Link:*** [*https://github.com/aspdave/College\_Score\_Card\_E139-Fall2015*](https://github.com/aspdave/College_Score_Card_E139-Fall2015)

***Project Summary :***

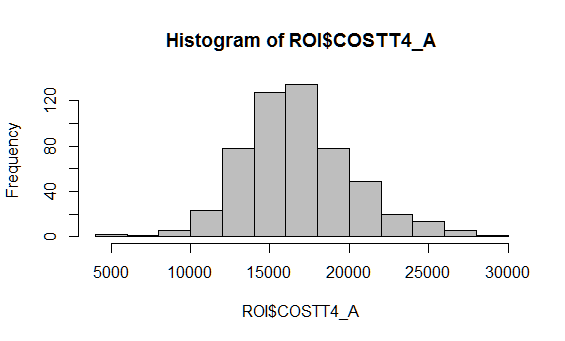
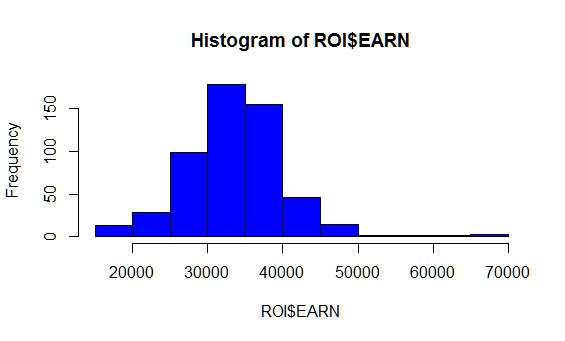
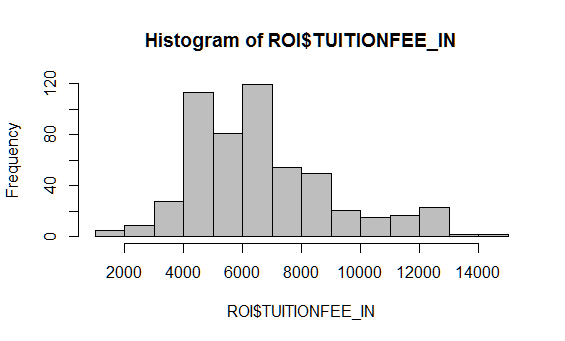
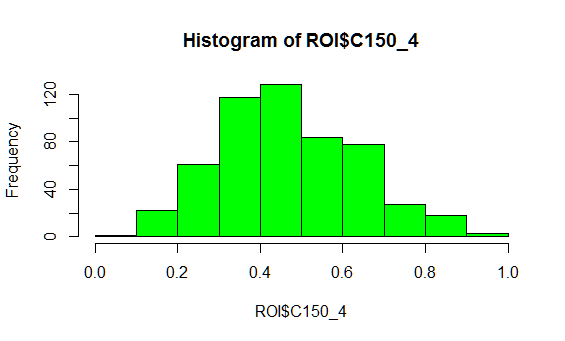
Project is to provide deep insights into relationship between college ROI and its impact factors so that students can select colleges that are best fit for their needs.

ROI for college education is defines for purpose of project is Investment in terms of Tuition and Cost of attending the 4 year degree. Returns are consider as Earning you make 6 years after completing college. Risk is Completion rate for the Particular college.

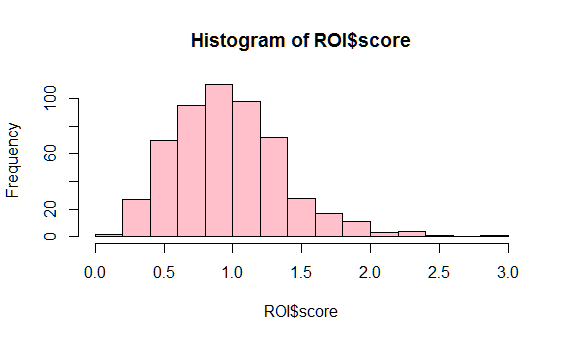
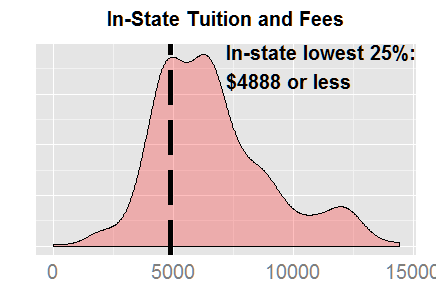
We created Score card for first 25 Value colleges for 4 year degree Based on Lowest In state Tuition Fees, low Cost of Tuition over all , high Rate of completion and high Earning Possibility . There is Quick Regression model ran with the variable above and Evaluated.

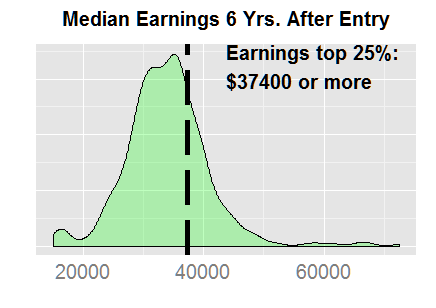
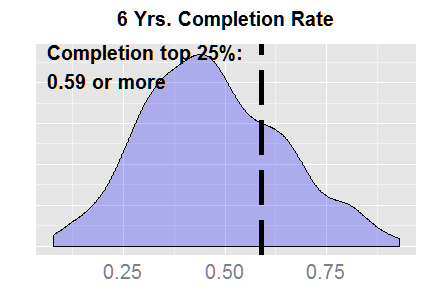
Our investigation is limited by Federal dataset available online. We choose Year 2009 data set that has quite a few Predictors available. We do have to pulled in right Variables , Clean , join in order to have dataset ready to be worked on . Code , Dataset and references are in Appendix.

First Look at the Variables: Earning , Cost of Attending , In state Tuition Cost and Completion rate.

After we ran the Score card for Total 539 colleges , Out score seems to be Normally distributed. TOP ROI list of 25 colleges are listed under . Score is Ratio of EARNING to COST \* Completion Rate. Following is the distribution of Top Value colleges looks like.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Institute Name** | **Tution\_IN** | **Cost** | **Completion Rate** | **EARN** | **Score** |
| Georgia Institute of Technology-Main Campus | 7506 | 16052 | 0.7918 | 60200 | 2.969497 |
| California Maritime Academy | 5090 | 18466 | 0.6867 | 66600 | 2.476672 |
| University of North Carolina at Chapel Hill | 5625 | 16194 | 0.8488 | 42900 | 2.248581 |
| University of Virginia-Main Campus | 9872 | 20488 | 0.93 | 49000 | 2.224229 |
| Maine Maritime Academy | 10105 | 20205 | 0.6667 | 67300 | 2.220683 |
| University of Florida | 4373 | 15395 | 0.8246 | 41300 | 2.212146 |
| CUNY Bernard M Baruch College | 4970 | 12261 | 0.6034 | 44000 | 2.16537 |
| Colorado School of Mines | 12244 | 21389 | 0.6725 | 65600 | 2.062556 |
| SUNY at Binghamton | 6761 | 17956 | 0.8028 | 45300 | 2.025331 |
| Missouri University of Science and Technology | 8488 | 18769 | 0.6333 | 57200 | 1.930031 |
| James Madison University | 7244 | 18702 | 0.8131 | 44200 | 1.921667 |
| Virginia Polytechnic Institute and State University | 8605 | 19538 | 0.7993 | 46600 | 1.906407 |
| University of Maryland-College Park | 8053 | 20831 | 0.8168 | 48500 | 1.901723 |
| University of Michigan-Ann Arbor | 11659 | 22606 | 0.894 | 47600 | 1.882438 |

**Complete List in APPENDIX I**

**Regression Model and Interpretation** :Model Ran with 3 variable and found Rate of completion is non-significant enough also interaction with Others . Luckily no Transformation needed for this model .

Predictors were found independent, Symmetric and with Constant Variance as Diagnostics plots show. Model is able to explain 42 % of Earning relates to cost of college and tuition fees paid by in state

students. Fees and cost is significant contributor factor to Earning as shown.

For no colleges education Earning will be $ 10,840 (Intercept), while Tuition Fees

Responsible by factor of around 3 and Cost of attending up to 0.9 . There is little interaction between

Variable and can be ignored for practical purposes.

Limitation of this discussion is related to Year 2009 Dataset, There are possibility of more refined

Results if numbers of Variables and couple of few years are jointly analyze.

Final model is as below.

modave=lm(EARN~(COSTT4\_A+TUITIONFEE\_IN+C150\_4)^2,data=ROI)

> summary(modave)

Call:

lm(formula = EARN ~ (COSTT4\_A + TUITIONFEE\_IN + C150\_4)^2, data = ROI)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.084e+04 3.455e+03 3.138 0.00180 \*\*

COSTT4\_A 9.603e-01 3.185e-01 3.015 0.00269 \*\*

TUITIONFEE\_IN 2.989e+00 6.174e-01 4.841 1.69e-06 \*\*\*

C150\_4 -8.006e+03 8.124e+03 -0.986 0.32483

COSTT4\_A:TUITIONFEE\_IN -1.785e-04 2.963e-05 -6.025 3.16e-09 \*\*\*

COSTT4\_A:C150\_4 6.781e-01 6.327e-01 1.072 0.28434

TUITIONFEE\_IN:C150\_4 1.920e+00 1.021e+00 1.880 0.06072 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5290 on 532 degrees of freedom

**Multiple R-squared: 0.4119, Adjusted R-squared: 0.4052**

F-statistic: 62.09 on 6 and 532 DF, p-value: < 2.2e-16

==============End of First Section==================