Which College is Best for You? Part 2

Michael L. Thompson

January 17, 2016, rev. Oct. 25, 2017

Contents

Introduction

This is a followup to my first script "Which College is Best for You?_".

Ben Hamner suggested in his comments to that first script that I write a script allowing anyone to specify his or her student profile right at the top. In this way the script might be more appealing for other folks to fork and play with it themselves. Good advice!

So here it is.

One thing though,. . . Using RStudio on this R Markdown file, you can click on "Chunks–>Run All" to save the variables in the global environment. Then it won't have to re-process the database every time you make a change to the student profile and re-run it each time with "Chunks–>Run All" to generate a new top schools list.

By the Way. . .

I've now added an R Shiny GUI and made it into a browser-based app: "Best Colleges for You"!

Specify Your Student Profile

```
Case <- list(studentProfile = list(
  dependent = TRUE,              # c(TRUE,FALSE)
```

```
    ethnicity = 'white',         # c('white','black','hispanic','asian')
    gender    = 'male',        # c('male','female')
    age       = 'le24',         # c('le24','gt24') (years of age)
    income    = 'gt30Kle110K',      # c('le30K','gt30Kle110K','gt110K') ($US)
    earnings  = 'gt110K',       # c('le30K','gt30Kle75K','gt75Kle110K','gt110K') ($US)
    sat       = 'gt1000le1200',     # c('le800','gt800le1000','gt1000le1200','gt1200le1400','gt1400')
    fafsa     = 'fsend_3',      # c('fsend_1','fsend_2','fsend_3','fsend_4','fsend_5')
    discipline= 'ComputerInformation',  # See below.
    region    = 'GreatLakes',       # See below for which states in which regions.
    locale    = 'CityMidsize',    # See below.
    traits    = list(           # See below.
      Risk      = 'M',
      Vision    = 'M',
      Breadth   = 'M',
      Challenge = 'M') # Must be in order; only 'L','M','H' valid.
)
)
```

(Note: The comments show the only valid values for each field. The `fafsa` field refers to how many schools
FAFSA form was sent to. Values `'fsend_1'` to `'fsend_4'` correspond to 1 to 4 schools, resp., and `'fsend_5'`
refers to 5 or more schools. Values for `income`, `earnings` and `sat` contain `"le"` meaning "less than or equal
to" and `"gt"` meaning "greater than". The `"K"` in the `income` and `earnings` values means "thousand". So
`'gt30Kle110K'` means "greater than $30,000 and less than or equal to $110,00".)

The fields with "See below." are explained in section ***"Explanation of Valid Entries for Student
Profile"*** at the end of the file.)

Setup the Data & Model

The next chunk loads the college database and defines necessary data structures and functions to implement
the model.

```
#===============================================================
# RUN THE SETUP FUNCTION:
DataSpec <- make_DataSpec()
DataSpec$utility <- utility


#===============================================================

# # Assign setup variables to objects in current environment:
# for( vnm in names(setupResult )){
#   eval(parse(text=sprintf('%s = setupResult$%s',vnm,vnm)))
# }
```

Here are *Your* Top Colleges

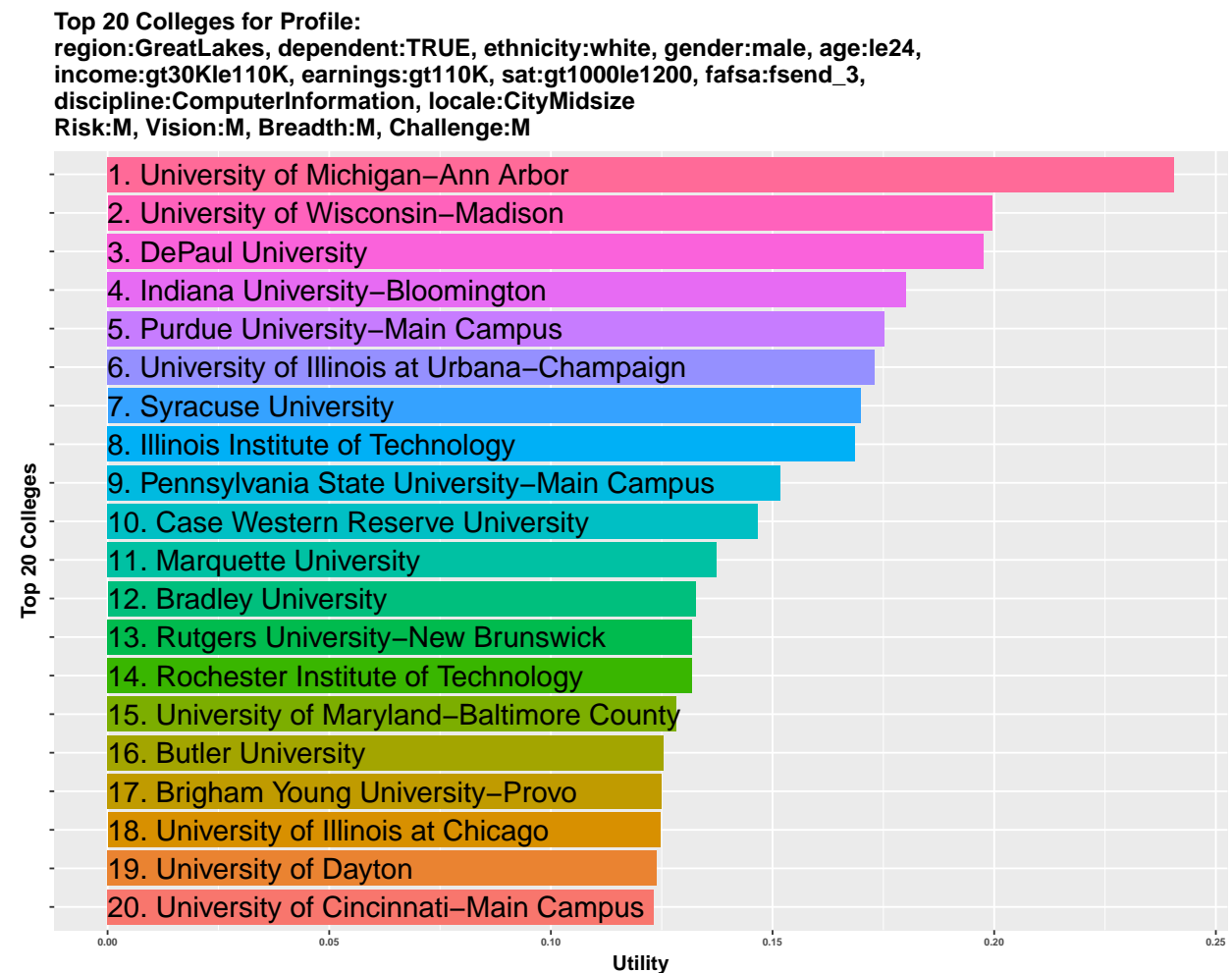Apply the model to get your `ntop` colleges:

```
# ENTER N, for top-N colleges:
Case$ntop <- 20
Case$studentProfile$beta <- getParameters(Case$studentProfile, DataSpec$propertyMap)
Case <- studentCaseStudy(Case, DataSpec, verbose = FALSE, ntop = Case$ntop)
```

```
# # Check-in on some select schools: selectSchools <- paste(c('Massachusetts
# Institute of Technology', 'California Institute of Technology',
# 'Princeton', 'Yale', 'Harvard', 'Stanford', 'Duke', 'Vanderbilt',
# 'Berkeley', 'Northwestern U', 'Princeton', 'Cornell U'),collapse='|')
# Case$BF %>% filter(grepl(selectSchools,labels)) %$% { tmp<-t(.[-1]);
# dimnames(tmp)<-list(names(.)[-1],gsub(' | University|nstitute of
# |echnology|ifornia|assachusetts|of','',labels)); tmp } %>% round(2)
```

Display the results:

```
plotTopN(Case,plot.it = TRUE)
```

Top 20 Colleges for Profile:
region:GreatLakes, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
**Risk:M, Vision:M, Breadth:M, Challenge:M**

Display a bit more output to help explain why those are the top schools.

The factor `prior` refers to the contribution due to size of school in terms of undergraduate enrollment (database field `"UGDS"`).

```
# Plot the partworth contributions to the utility for each of the top schools.
plotContributions(Case,plot.it = TRUE)
```

Utility Contributions of Factors
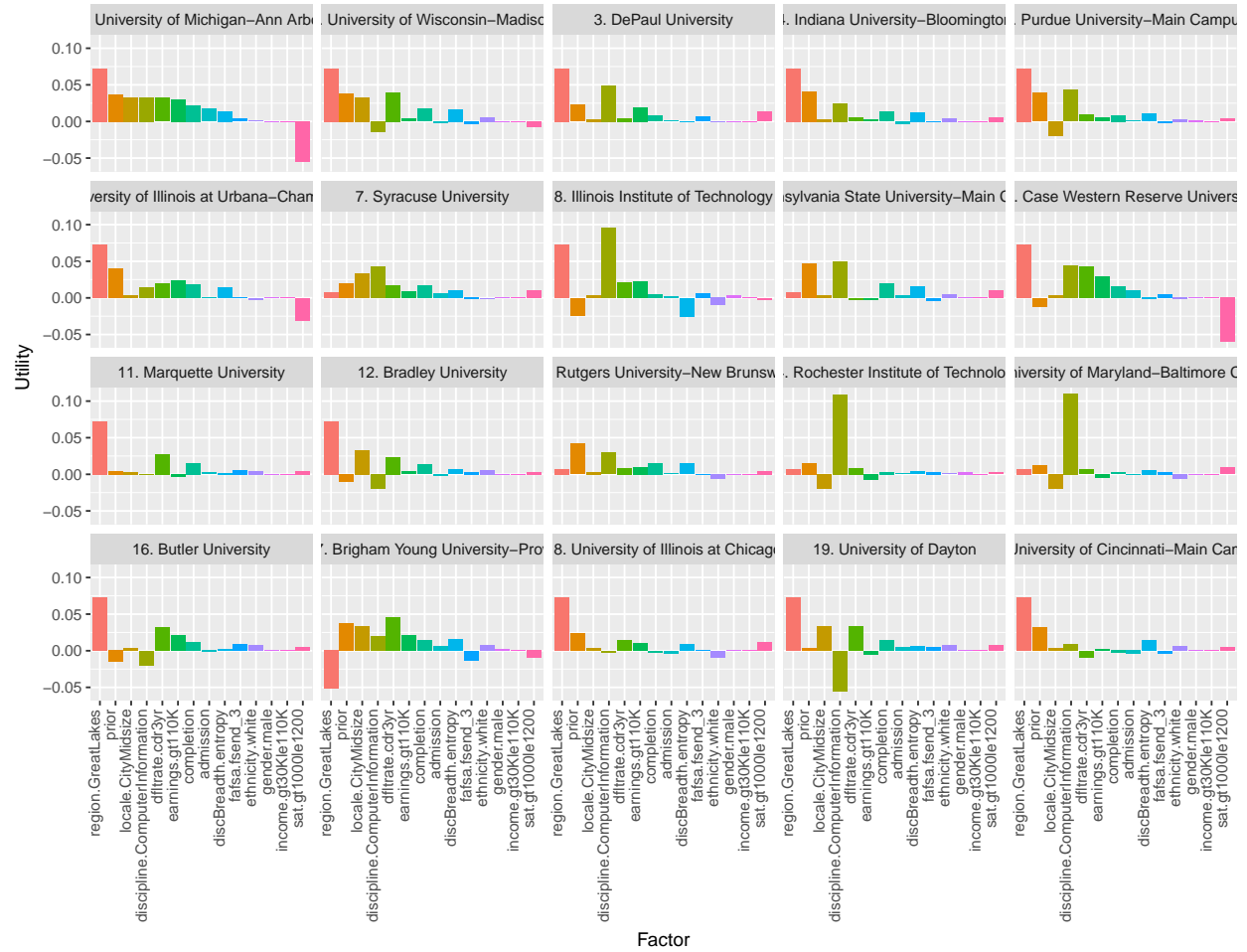
```
# Show the partworth (utility function parameters):
plotBeta(Case$studentProfile$beta,Case$studentProfile)
```

4

Utility Partworths for Risk:M, Vision:M, Breadth:M, Challenge:M
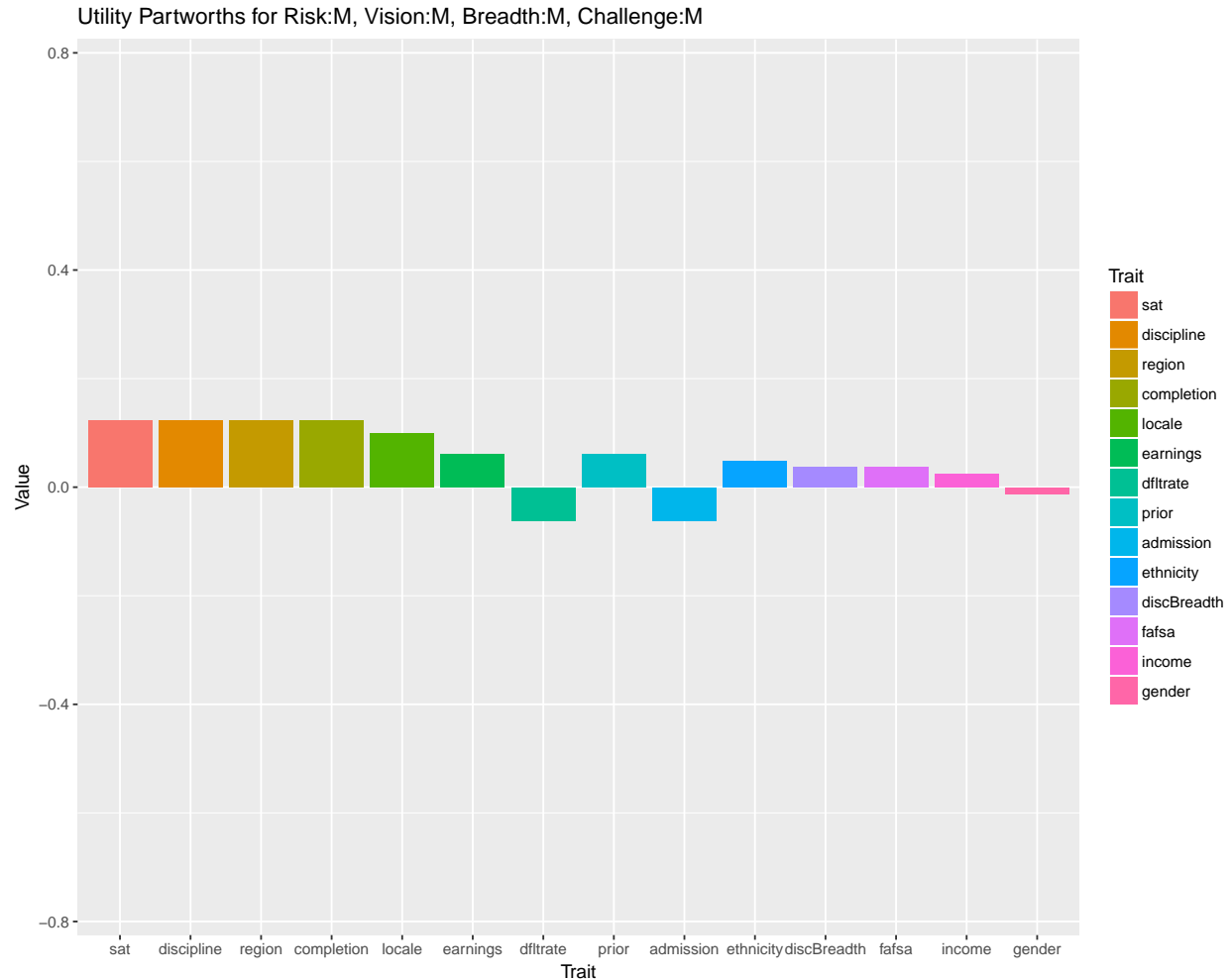
How About a Road-Trip to Your Top Schools?!

I've copied idea and code snips from Tad Dallas's script on Kaggle.com:

Sensitivity of Your Top Schools to Changes in Your Profile

Perform sensitivity analysis on your profile factors.

```
# Sweep region
sweepVariable <- "region"
case0 <- Case
stdtProf0 <- case0$studentProfile
pmap <- DataSpec$propertyMap[sweepVariable]
maxcnt <- length(pmap[[sweepVariable]])
sensResults <- sensitivity(stdtProf0, pmap, case0, DataSpec, maxcnt = maxcnt,
    plot.it = FALSE)
print(sensResults$grobList)
```
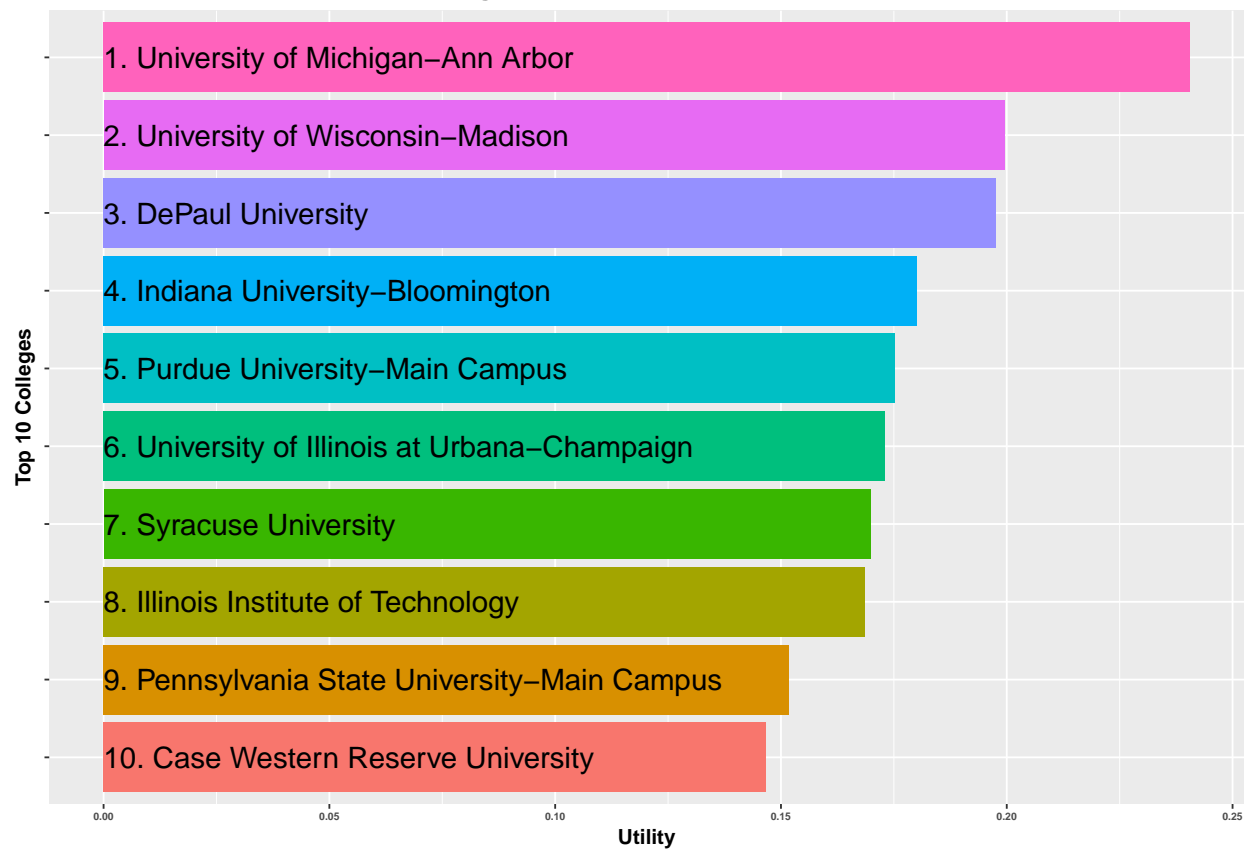
```
## [[1]]
```

Top 10 Colleges for Profile:
region:FarWest, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
## 
## [[2]]
```

Top 10 Colleges for Profile:
region:GreatLakes, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
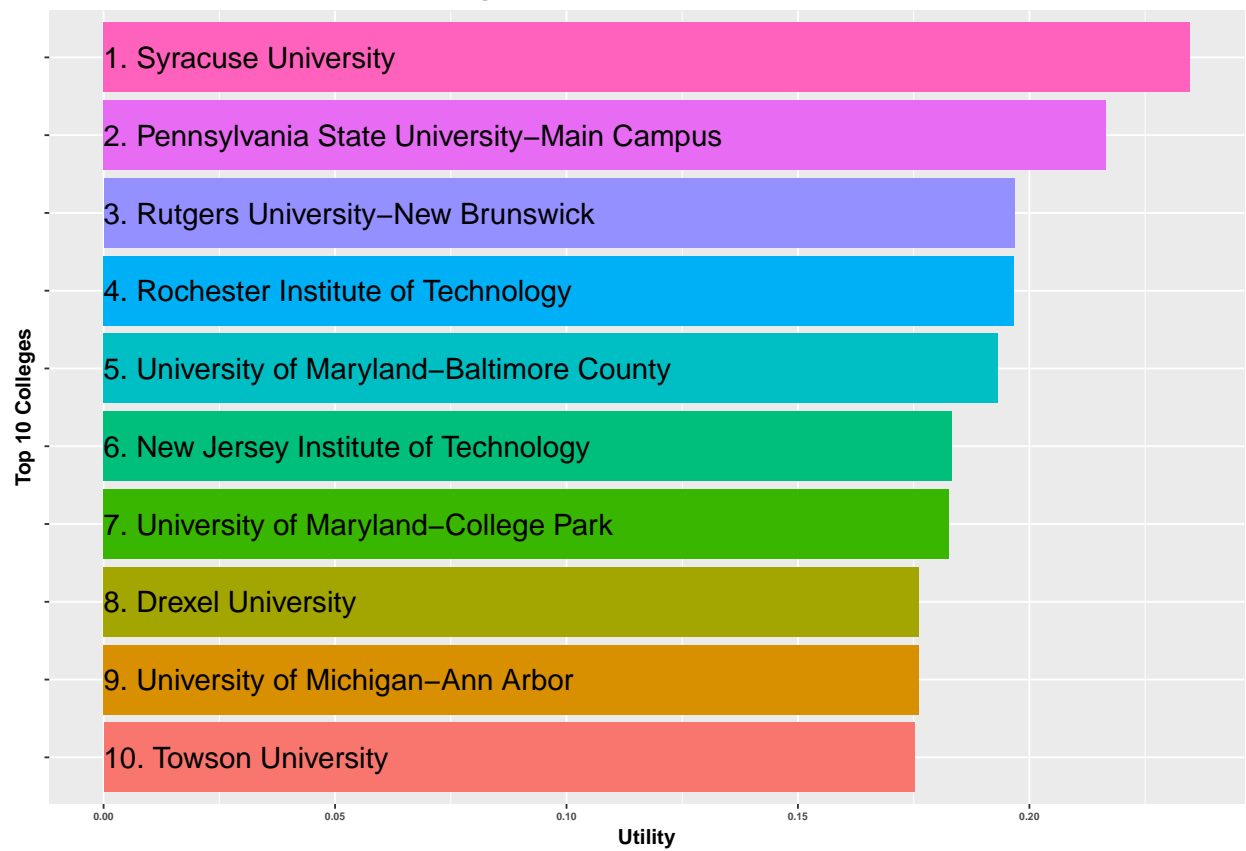**Risk:M, Vision:M, Breadth:M, Challenge:M**
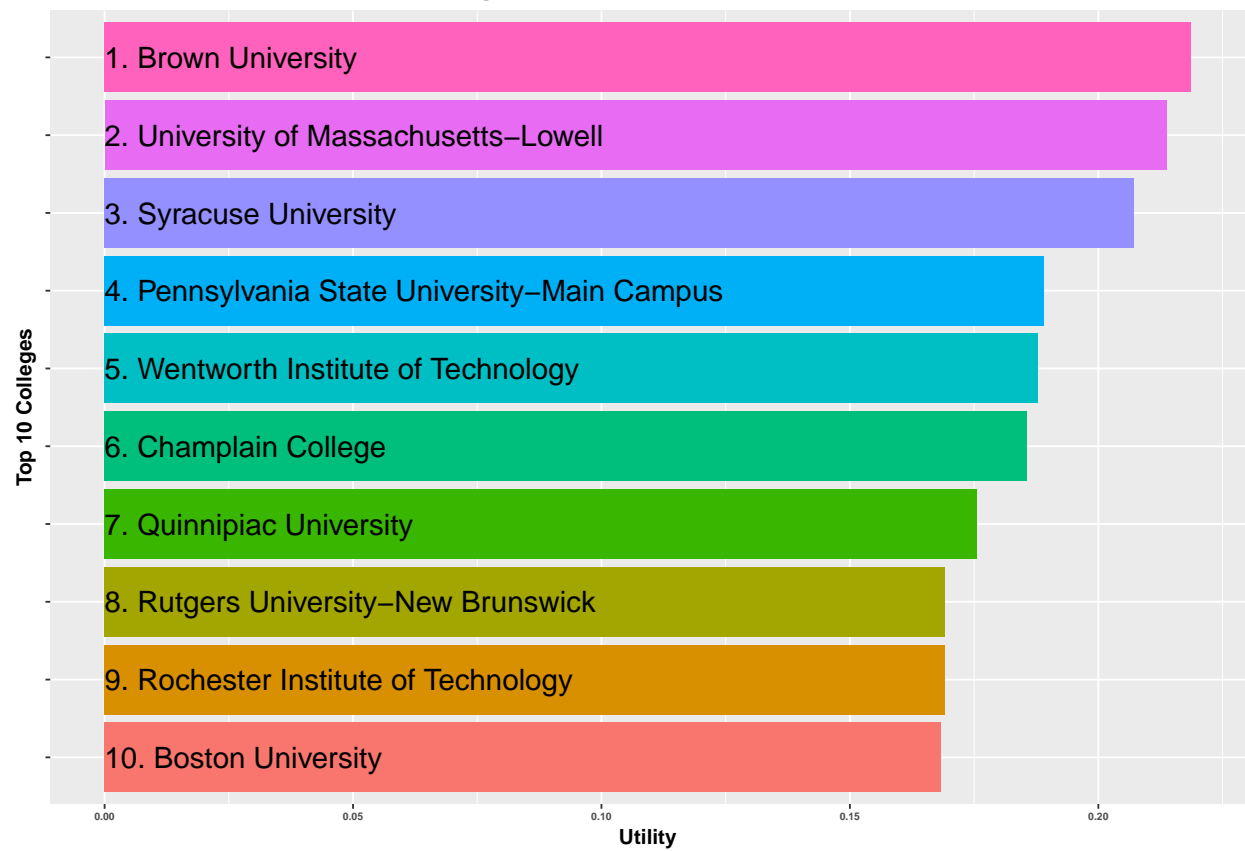


```
## 
## [[3]]
```
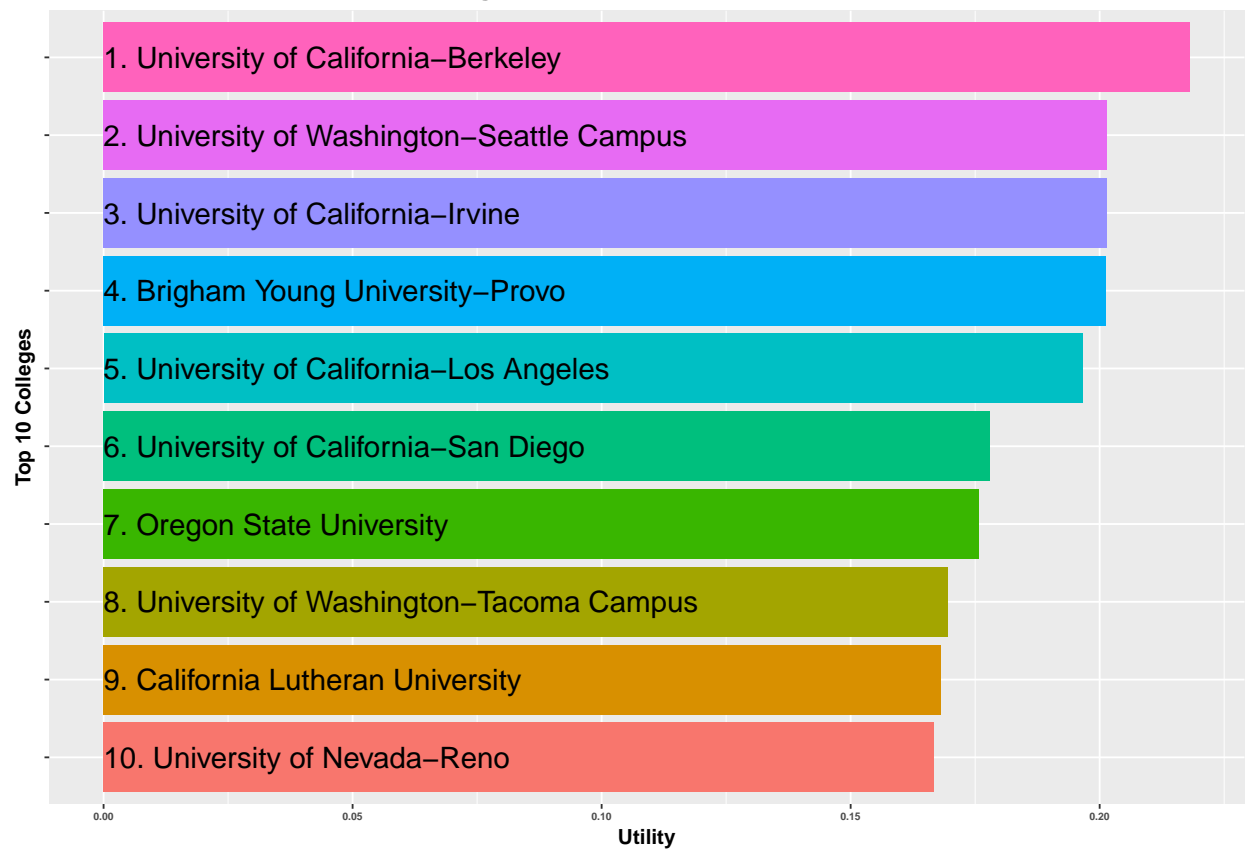
Top 10 Colleges for Profile:
region:MidEast, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
##
## [[4]]
```

Top 10 Colleges for Profile:
region:NewEngland, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
## 
## [[5]]
```

Top 10 Colleges for Profile:
region:Plains, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
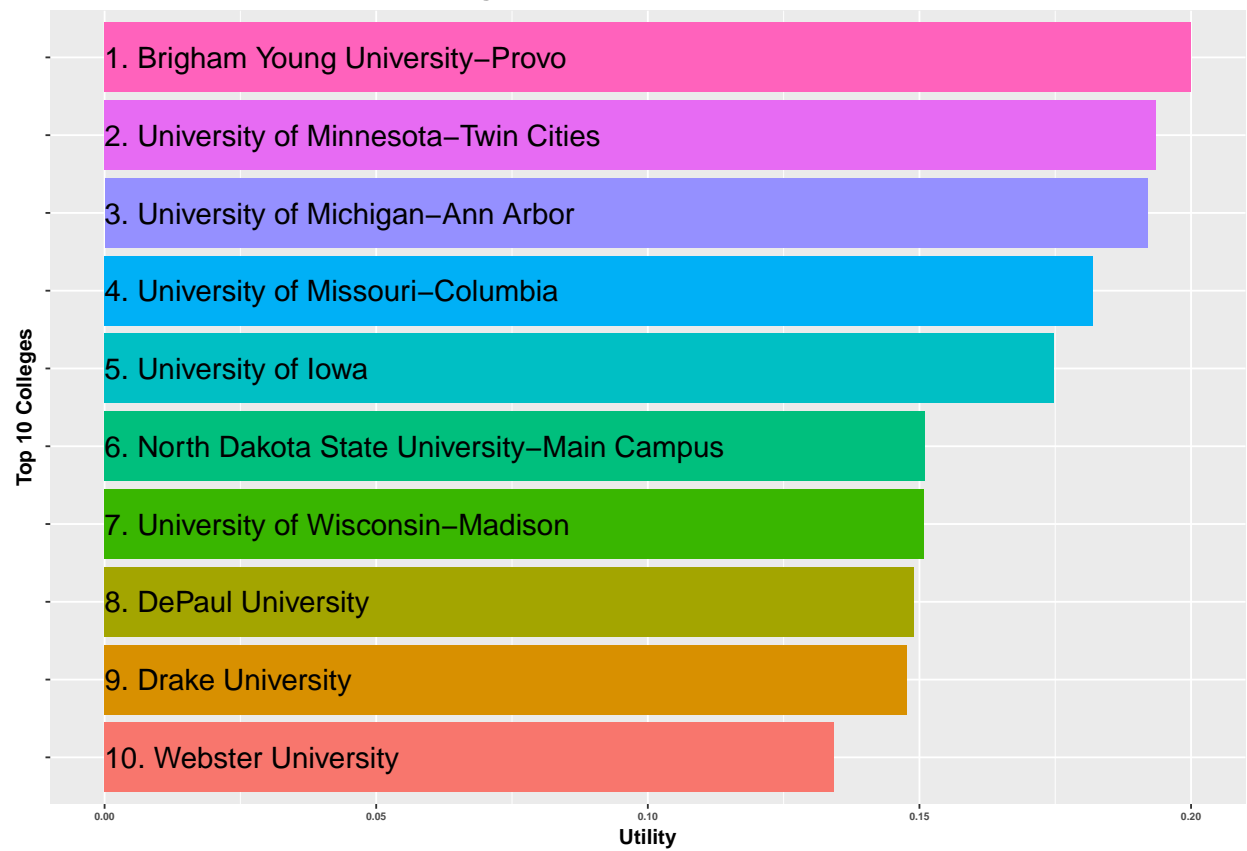**Risk:M, Vision:M, Breadth:M, Challenge:M**


```
##
## [[6]]
```

Top 10 Colleges for Profile:
region:RockyMountains, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
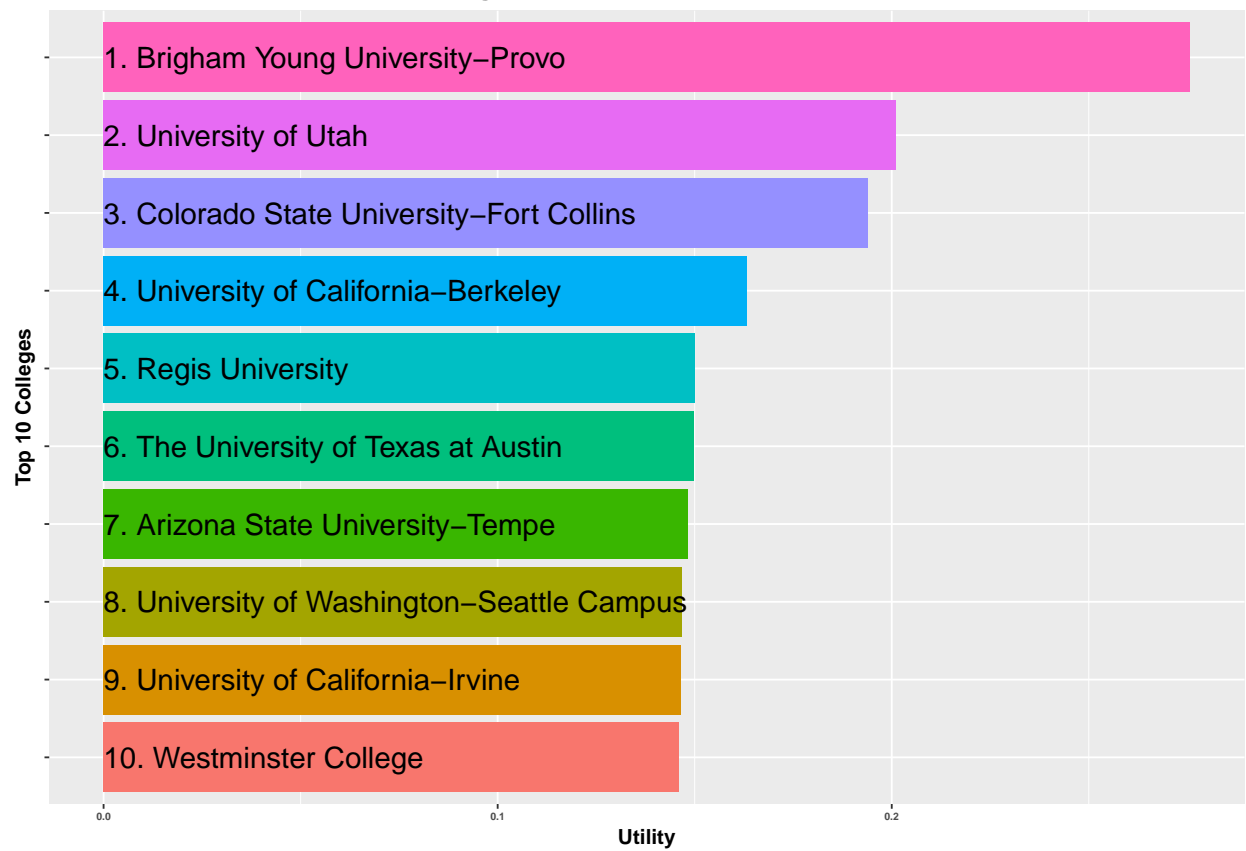**Risk:M, Vision:M, Breadth:M, Challenge:M**


```
## 
## [[7]]
```

Top 10 Colleges for Profile:
region:Southeast, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
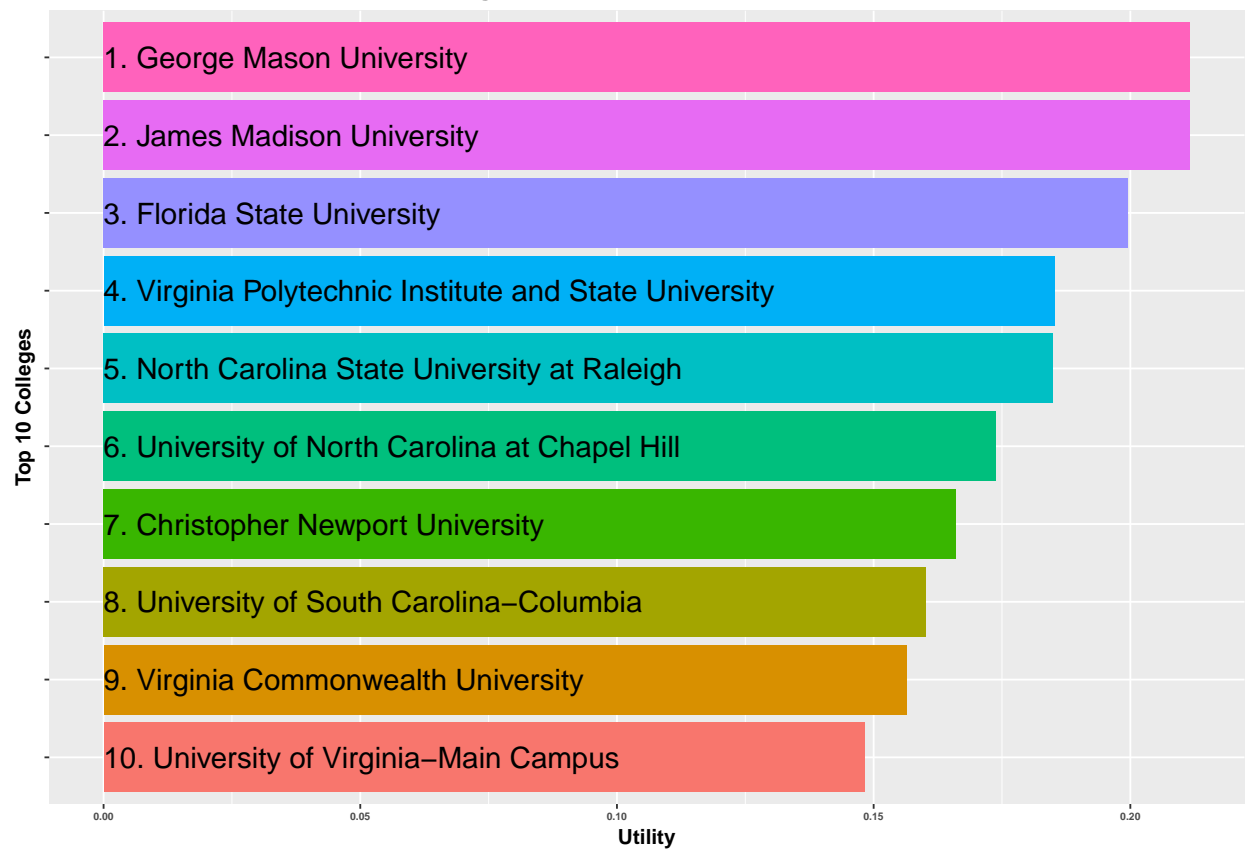**Risk:M, Vision:M, Breadth:M, Challenge:M**
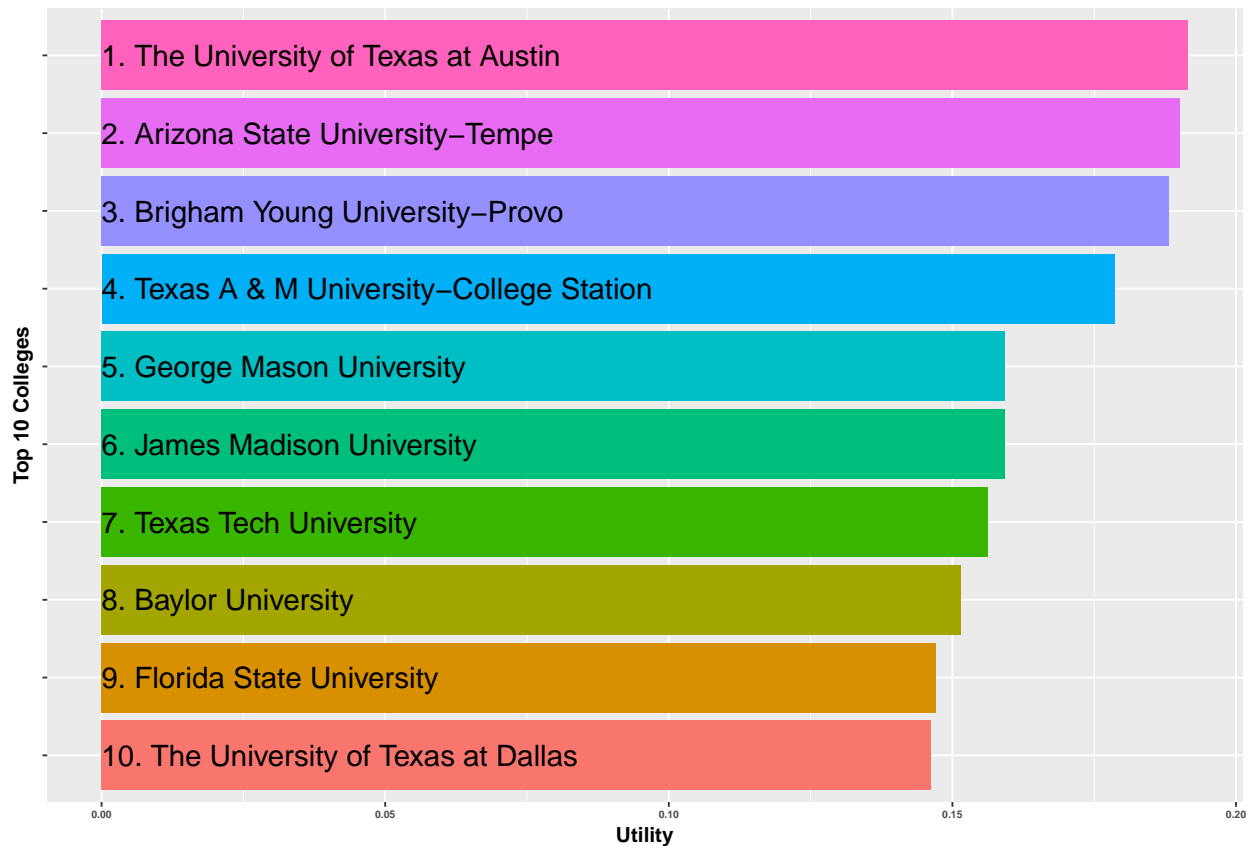


```
## 
## [[8]]
```

Top 10 Colleges for Profile:
region:Southwest, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, sat:gt1000le1200, fafsa:fsend_3,
discipline:ComputerInformation, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```r
# This is how to sweep over SAT:
sweepVariable <- "sat"
case0 <- Case
stdtProf0 <- case0$studentProfile
pmap <- DataSpec$propertyMap[sweepVariable]
maxcnt <- length(pmap[[sweepVariable]])
sensResults <- sensitivity(stdtProf0, pmap, case0, DataSpec, maxcnt = maxcnt,
    plot.it = FALSE)
print(sensResults$grobList)
```

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

Top 10 Colleges for Profile:
sat:le800, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, fafsa:fsend_3,
discipline:ComputerInformation, region:GreatLakes, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
##
## [[2]]
```
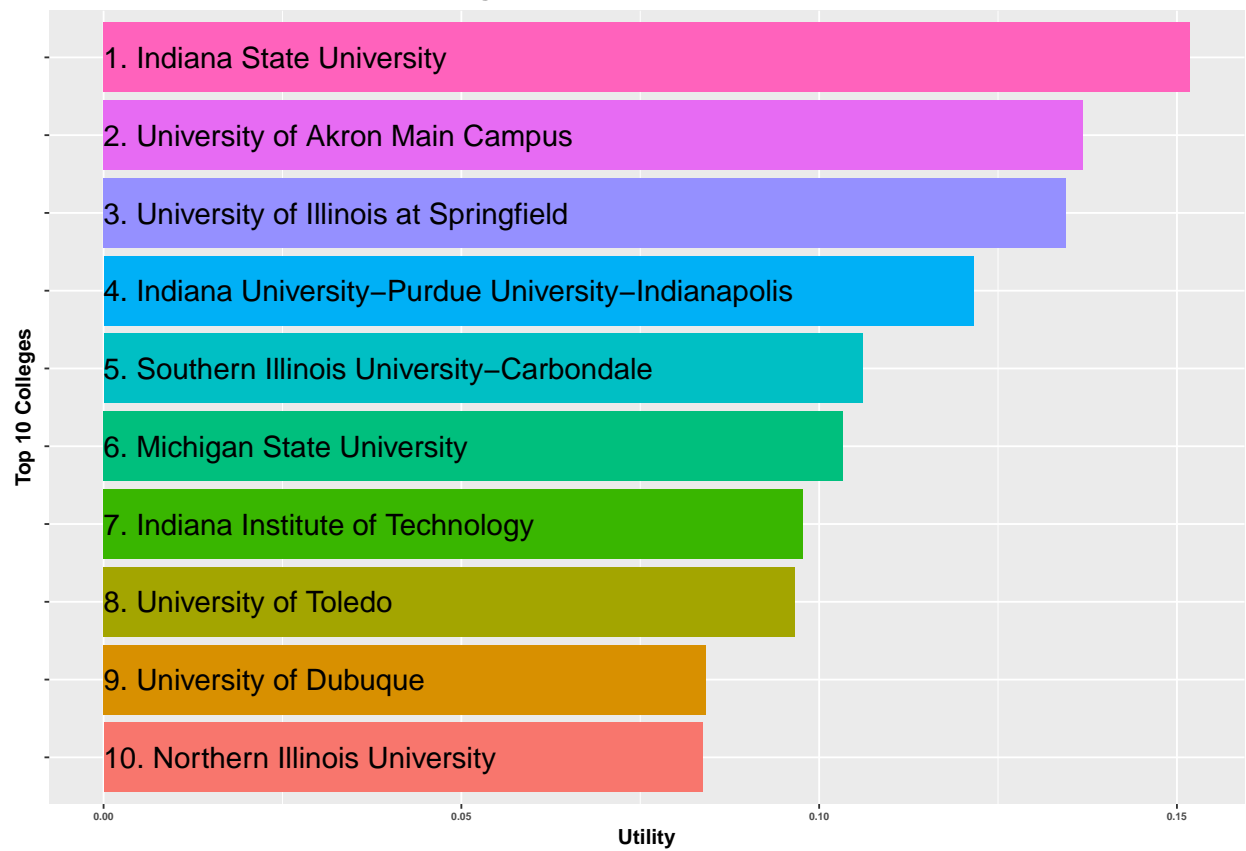
Top 10 Colleges for Profile:
sat:gt800le1000, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, fafsa:fsend_3,
discipline:ComputerInformation, region:GreatLakes, locale:CityMidsize
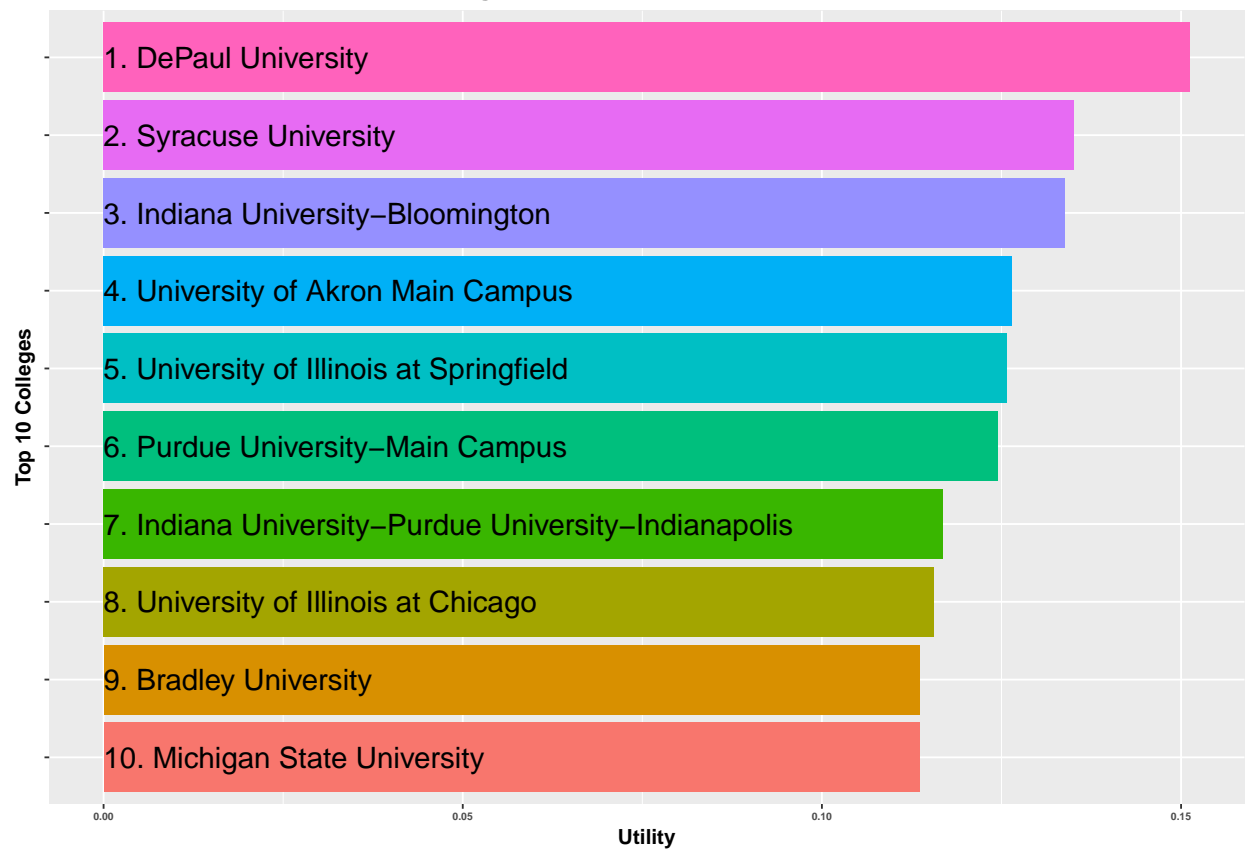**Risk:M, Vision:M, Breadth:M, Challenge:M**


```
##
## [[3]]
```

Top 10 Colleges for Profile:
sat:gt1000le1200, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, fafsa:fsend_3,
discipline:ComputerInformation, region:GreatLakes, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
##
## [[4]]
```

Top 10 Colleges for Profile:
sat:gt1200le1400, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, fafsa:fsend_3,
discipline:ComputerInformation, region:GreatLakes, locale:CityMidsize
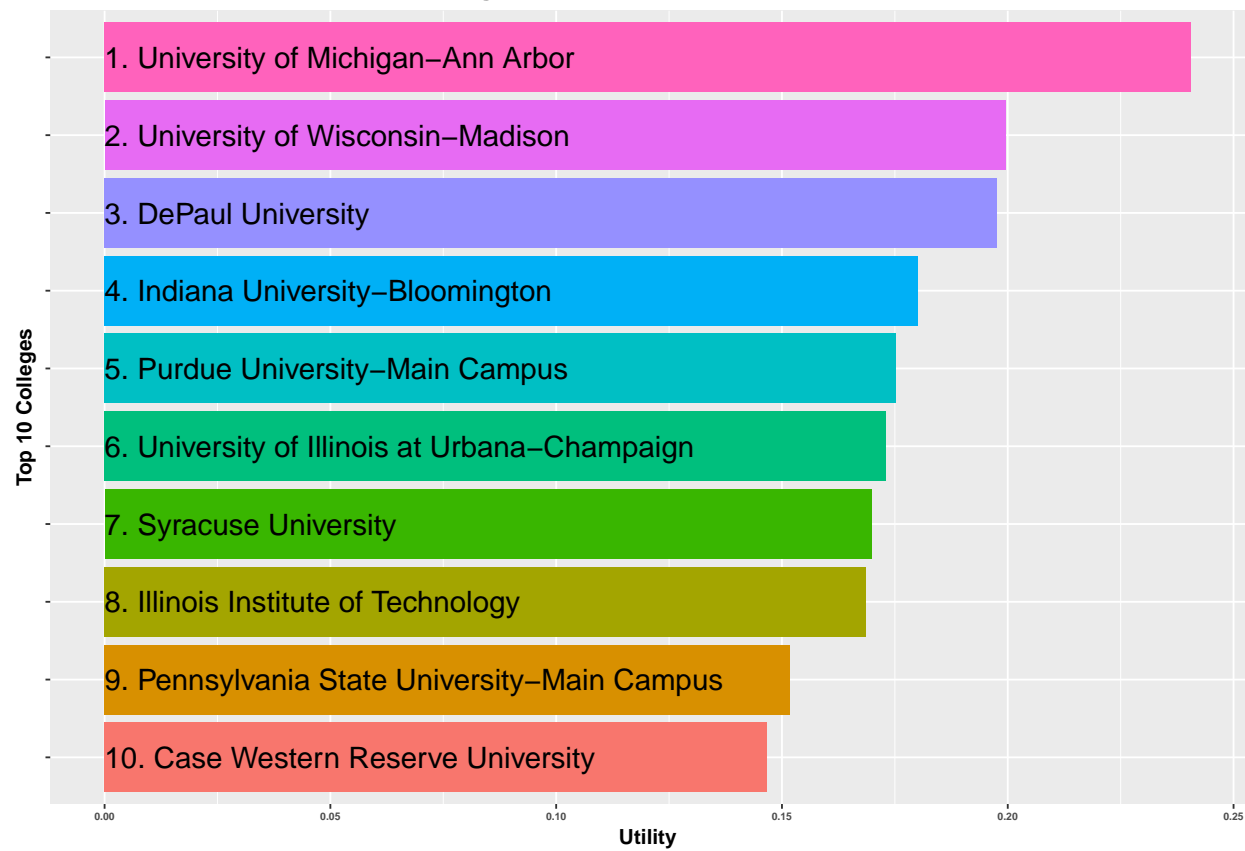**Risk:M, Vision:M, Breadth:M, Challenge:M**
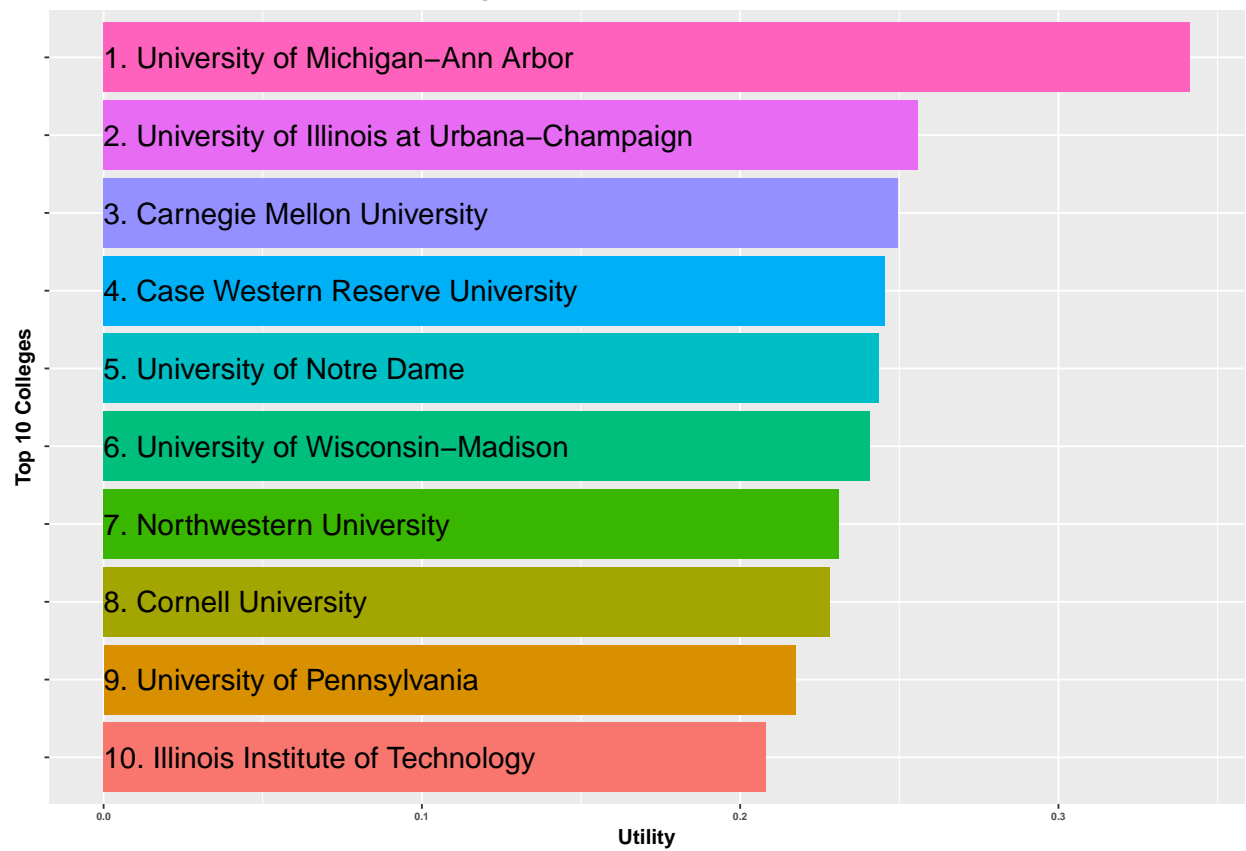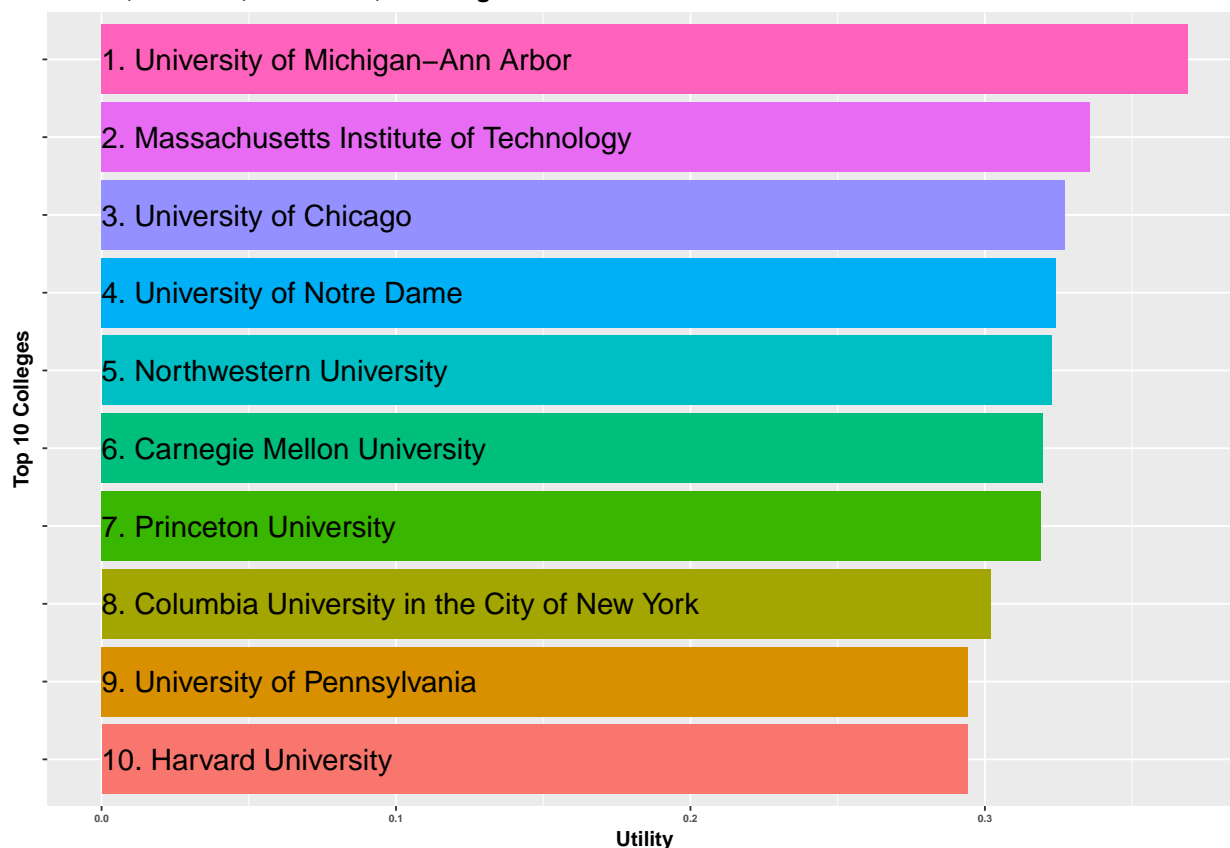


```
##
## [[5]]
```

Top 10 Colleges for Profile:
sat:gt1400, dependent:TRUE, ethnicity:white, gender:male, age:le24,
income:gt30Kle110K, earnings:gt110K, fafsa:fsend_3,
discipline:ComputerInformation, region:GreatLakes, locale:CityMidsize
Risk:M, Vision:M, Breadth:M, Challenge:M


```
# sweepVariable <- 'locale' stdtProf0 <- studentProfile pmap <-
# propertyMap[sweepVariable] maxcnt <- length(pmap[[sweepVariable]])
# sensResults <-
# sensitivity(stdtProf0,pmap,Case,DataSpec,maxcnt=maxcnt,plot.it=FALSE)
# print(sensResults$grobList)

# # This is how to sweep over a subset of disciplines: stdtProf0 <-
# Case$studentProfile pmap <- DataSpec$propertyMap[c('discipline')]
# pmap$discipline <-
# pmap$discipline[c('ComputerInformation','BiologicalBiomedical','HealthProfessions','PhysicalSciences'
# maxcnt <- pmin(7,length(pmap$discipline)) sensResults <-
# sensitivity(stdtProf0,pmap,Case,DataSpec,maxcnt=maxcnt,plot.it=FALSE)
# print(sensResults$grobList)

# # This is how to sweep over a behavioral trait: stdtProf0 <-
# Case$studentProfile ntop <- 10 for(lvl in c('L','M','H')){
# stdtProf0$traits$Risk <- lvl stdtProf0$beta <-
# getParameters(stdtProf0,DataSpec$propertyMap) case0 <-
# list(studentProfile=stdtProf0,ntop=ntop) case0 <-
# studentCaseStudy(case0,DataSpec,verbose=FALSE,ntop=ntop)
# plotTopN(caseResult = case0,plot.it = TRUE) }
```

Explanation of Valid Entries for Student Profile

The following selections show what entries are valid for the student's `region`, `locale`, `discipline` and behavioral traits.

Valid Regions

Valid region names to choose for your student profile `region` property are shown in column ***Region*** of the table below; and states in each region are listed in the ***States*** column:

Region	States
1	
FarWest	AK,CA,HI,NV,OR,WA
2	
GreatLakes	IL,IN,MI,OH,WI
3	
MidEast	DE,DC,MD,NJ,NY,PA
4	
NewEngland	CT,ME,MA,NH,RI,VT
5	
Plains	IA,KS,MN,MO,NE,ND,SD
6	
RockyMountains	CO,ID,MT,UT,WY
7	
Southeast	AL,AR,FL,GA,KY,LA,MS,NC,SC,TN,VA,WV
8	
Southwest	AZ,NM,OK,TX

Valid Locales

Valid locale names to choose for your student profile `locale` property are shown in column ***Locale*** of the table below:

Locale	Description
1	
Rural	Rural: Remote (rural territory more than 25 miles from an urbanized area and more than 10 miles from an urban cluster)
2	
Rural	Rural: Distant (rural territory more than 5 miles but up to 25 miles from an urbanized area or more than 2.5 and up to 10 miles from an urban cluster)
3	
Rural	Rural: Fringe (rural territory up to 5 miles from an urbanized area or up to 2.5 miles from an urban cluster)
4	
TownRemote	Town: Remote (in urban cluster more than 35 miles from an urbanized area)
5	
TownDistant	Town: Distant (in urban cluster more than 10 miles and up to 35 miles from an urbanized area)
6	
SuburbSmallMid	Town: Fringe (in urban cluster up to 10 miles from an urbanized area)
7	
SuburbSmallMid	Suburb: Small (outside principal city, in urbanized area with population less than 100,000)
8	
SuburbSmallMid	Suburb: Midsize (outside principal city, in urbanized area with population of at least 100,000 but less than 250,000)
9	
SuburbLarge	Suburb: Large (outside principal city, in urbanized area with population of 250,000 or more)
10	
CitySmall	

City: Small (population less than 100,000)

11

CityMidsize

City: Midsize (population of at least 100,000 but less than 250,000)

12

CityLarge

City: Large (population of 250,000 or more)

Valid Academic Disciplines

Valid academic disciplines to choose for your student profile `discipline` property are shown in column *discName* of the table below:

LABEL

discName

1

Agriculture, Agriculture Operations, and Related Sciences

AgricultureAgriculture

2

Architecture and Related Services

ArchitectureRelated

3

Area, Ethnic, Cultural, Gender, and Group Studies

AreaEthnic

4

Biological and Biomedical Sciences

BiologicalBiomedical

5

Business, Management, Marketing, and Related Support Services

BusinessManagement

6

Communication, Journalism, and Related Programs

CommunicationJournalism

7

Communications Technologies/Technicians and Support Services

CommunicationsTechnologies

8

Computer and Information Sciences and Support Services

ComputerInformation

9

Construction Trades

ConstructionTrades

10

Education

Education

11

Engineering

Engineering

12

Engineering Technologies and Engineering-Related Fields

EngineeringTechnologies

13

English Language and Literature/Letters

EnglishLanguage

14

Family and Consumer Sciences/Human Sciences

FamilyConsumer

15

Foreign Languages, Literatures, and Linguistics

ForeignLanguages

16

Health Professions and Related Programs

HealthProfessions

17

History

History

18

Homeland Security, Law Enforcement, Firefighting and Related Protective Services

HomelandSecurity

19

Legal Professions and Studies

LegalProfessions

20

Liberal Arts and Sciences, General Studies and Humanities

LiberalArts

21

Library Science

LibraryScience

22

Mathematics and Statistics

MathematicsStatistics

23

Mechanic and Repair Technologies/Technicians

MechanicRepair

24

Military Technologies and Applied Sciences

MilitaryTechnologies

25

Multi/Interdisciplinary Studies

MultiInterdisciplinary

26

Natural Resources and Conservation

NaturalResources

27

Parks, Recreation, Leisure, and Fitness Studies

ParksRecreation

28

Personal and Culinary Services

PersonalCulinary

29

Philosophy and Religious Studies

PhilosophyReligious

30

Physical Sciences

PhysicalSciences

31

Precision Production

PrecisionProduction

32

Psychology

Psychology

33

Public Administration and Social Service Professions

PublicAdministration

34

Science Technologies/Technicians

ScienceTechnologies

35

Social Sciences

SocialSciences

36

Theology and Religious Vocations

TheologyReligious

37

Transportation and Materials Moving

TransportationMaterials

38

Visual and Performing Arts

VisualPerforming

Explanation of Behavioral Traits

Trait level has 4 types with impact on properties as follows when each trait increases from low to medium to high:

1. Risk – Willingness to surround self with people & settings very different from self & origins;

- Weights homogeneity & sameness as self lower than other traits.

2. Vision – Willingness to look past short-term suitability in pursuit of future;

- Weights completion rates, future earnings, debt & debt repayment rates higher than current campus setting compatability.

3. Breadth – Willingness to entertain a variety of academic disciplines;

- Weights predominance of own major discipline lower than otherwise.

4. Challenge – Willingness to embrace highest academic rigor

- Weights admissions selectiveness, high SAT & higher than otherwise.

Traits Have Impact On Four Property Categories Listed in Rows:

	Risk	Vision	Breadth	Challenge	Properties
1. Self & Origins	same to differ	——	——-	—-	ethn,inc,1stgen,usbrn,pov

	Risk	Vision	Breadth	Challenge	Properties
2. Campus Setting & Folks	homog to diverse	——	——-	—-	reg,loc,siz
3. Academics	—-	——	narrow to broad	low to high sat/adm	sat,admrt,disc
4. After-College Prospects	—-	low to high dbt/ern	——-	—-	comprt,earn,rpy,dfltrt

Any feedback would be appreciated.

Thanks,

-Michael L. Thompson

(my info at LinkedIn)