

*Teaching Business Statistics and Analytics  
Using R*

***Welcome!***



**Leslie Hendrix**

***University of South Carolina***

# Teaching R to Business Students: Why and How?

Leslie Hendrix  
Darla Moore School of Business  
University of South Carolina



# Background

- Business students at the Darla Moore School of business (DMSB) take 2 statistics courses.
- First course is basic.
- Second course covers more advanced topics such as ANOVA, logistic regression, multiple regression, time series.

# Background

- First iteration used Excel for both courses.
- Piloted R in 2 sections Spring 2018.
- Once students got over the hurdle of learning basic coding, they had a much more solid grasp of the statistical concepts taught.
- Student evaluations liked learning R vs disliked it 5:1.

# Teaching R

- Department chair looked at student evaluations.
- Talked to dean.
- Learning R became mandatory for every section of MGSC 291 Fall 2018.

# Student Perception

*“I wish I could’ve learned more about R. When I took this class I was one of the lucky students to learn how to code in R. I feel that we didn’t have much time to develop this amazing skill in class.”*



# Student Perception

*“It helped at the career fair to differentiate me from other people. Also, I would love to take more data science courses, and I hope are offered.”*



# Student Perception

*“Even though sometimes I am not using R directly, the coding language was a good way to understand other coding, like SQL for example.”*





# Student Perception

On working with our baseball team

*“...we are currently working on analyzing what makes specific pitches effective (spin rate, vertical break, etc.), developing a more in depth post game box score with more insightful analytics, visualizing data to better explain data, and also preparing scouting reports for them.”*



# Why R or Python?



Find jobs

Company reviews

Find salaries

Upload your resume

What

Job title, keywords, or company

Where

City, state, or zip code

R Python



Find jobs

Advanced Job Se.

Tip: Enter your city or zip code in the "where" box to show results in your area.

Salary Estimate ▾

Job Type ▾

Location ▾

Company ▾

Experience Level ▾

R Python jobs

Sort by: **relevance** - [date](#)

Page 1 of 14,659 jobs



**AVP, AML Statistician - Independent  
Validation R, SAS, SQL**

Be the first to see new R Python jobs

My email:

Activate

# Why R or Python?



Find jobs

Company reviews

Find salaries

Upload y

What

Job title, keywords, or company

Where

City, state, or zip code

R Python \$70,000



Find jobs

\$70,000+ X

Job Type ▾

Location ▾

Company ▾

Experience Level ▾

R Python \$70,000 jobs

Sort by: **relevance** - date

Page 1 of 12,229 jobs

## Marketing Scientist

Valassis 3.6 ★

Morrisville, NC

- Prior experience with Python, R, and/or SQL.
- Use your knowledge in programming (such as SQL, R, and/or Python) to...
- ON ANY GIVEN DAY, YOU WILL...

Be the first to see new R Python \$70,000 jobs

My email:

Activate

By creating a job alert or receiving recommended jobs, you agree to our [Terms](#). You can change your consent settings at any time by unsubscribing or as detailed in our terms.

# Why R or Python?



Find jobs

Company reviews

Find salaries

What

Job title, keywords, or company

Where

City, state, or zip code

R Python \$70,000



Find jobs

\$70,000+ ✕

Job Type ▾

Location ▾

Company ▾

Entry Level ✕

R Python \$70,000 jobs

Sort by: **relevance** - date

Page 1 of 1,782 jobs



**\$65/hr | Data Scientist/R/Python/NoSQL |  
Nashville, TN**

Vaco 3.7 ★

Brentwood, TN 37027

**\$75 an hour**

Easily apply

R Python Databricks Scala Visualization - PowerBI or Tableau  
D3...

3 days ago · Save job

Be the first to see new R Python \$70,000 jobs

My email:

Activate

By creating a job alert or receiving recommendations, you agree to our [Terms](#). You can change your settings at any time by unsubscribing or adjusting our terms.

# Why R or Python?



*“Business analysts at all levels at Amazon ... will be much more insightful and productive if they learn how to pull, manipulate, and analyze data using scripting languages.”*

*“Anyone working with data will need to continue learning and updating their computational (and methodological) skills.”*

**-Matt Taddy**

**Amazon VP Economic Technology**

**Chief Economist North America**



# Teaching R

- Teach R in the first lecture



# Teaching R

- Incorporate R into lecture slides / videos.
- Demonstrate R in almost every class.
  - Always include R code in slides/notes.
  - Students encouraged to bring laptop to class.



# Teaching R

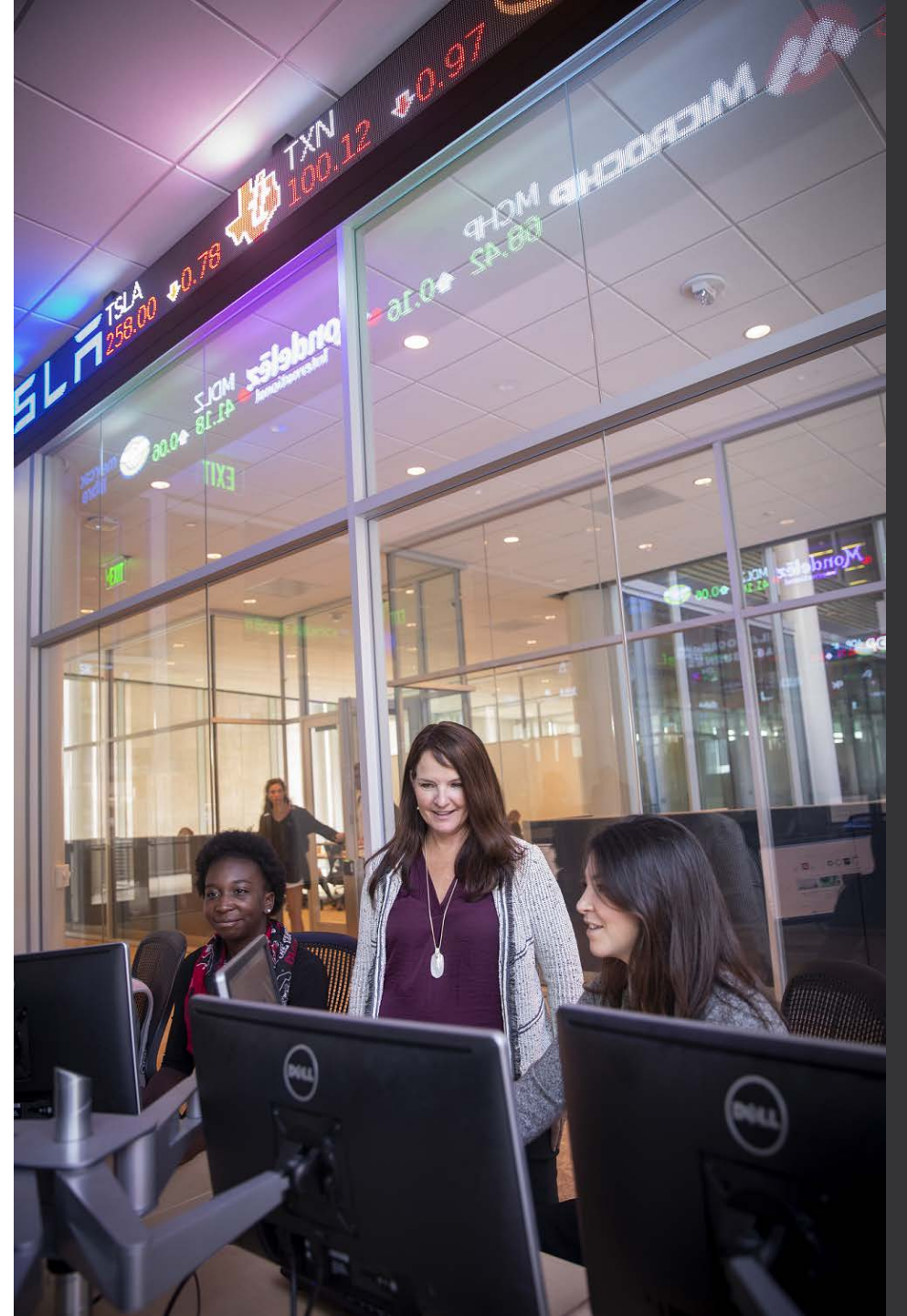
- 3 case studies and 3 projects with R
- Encourage “Googling”





# Issues

- Students need support outside of class.
- Students will struggle with transitioning from menu-driven Excel to the coding interface of R.
- Proper level and amount of support yields a much better student experience.



# Issues

- Need strategy for large class sizes
- Need strategy for the fully online courses





# Data Lab

- Fall 2018, MBA students were trained to mentor our students with R and course topics.
  - No dedicated space
  - Not enough hours for 700+ students learning R.
- Spring 2019
  - We have staff!
    - 7 undergraduates, 4 MBAs, 1 PhD



# Data Lab

- Fall 2019
  - We have space!
    - 5 undergraduates, 10 MBAs, 1 PhD, and 2 volunteer undergrads
    - Two undergraduate students from computer science





# Our Space



# Our Space





# Data Lab Staff



# What We are Doing Now

- First case study uses Excel
  - Reminds them of Excel functions/pivot tables
- Second case study uses R
  - Dataset small enough to check answers in Excel
- Projects and remaining case studies use R
  - Must use R
  - Dataset is too large to use Excel



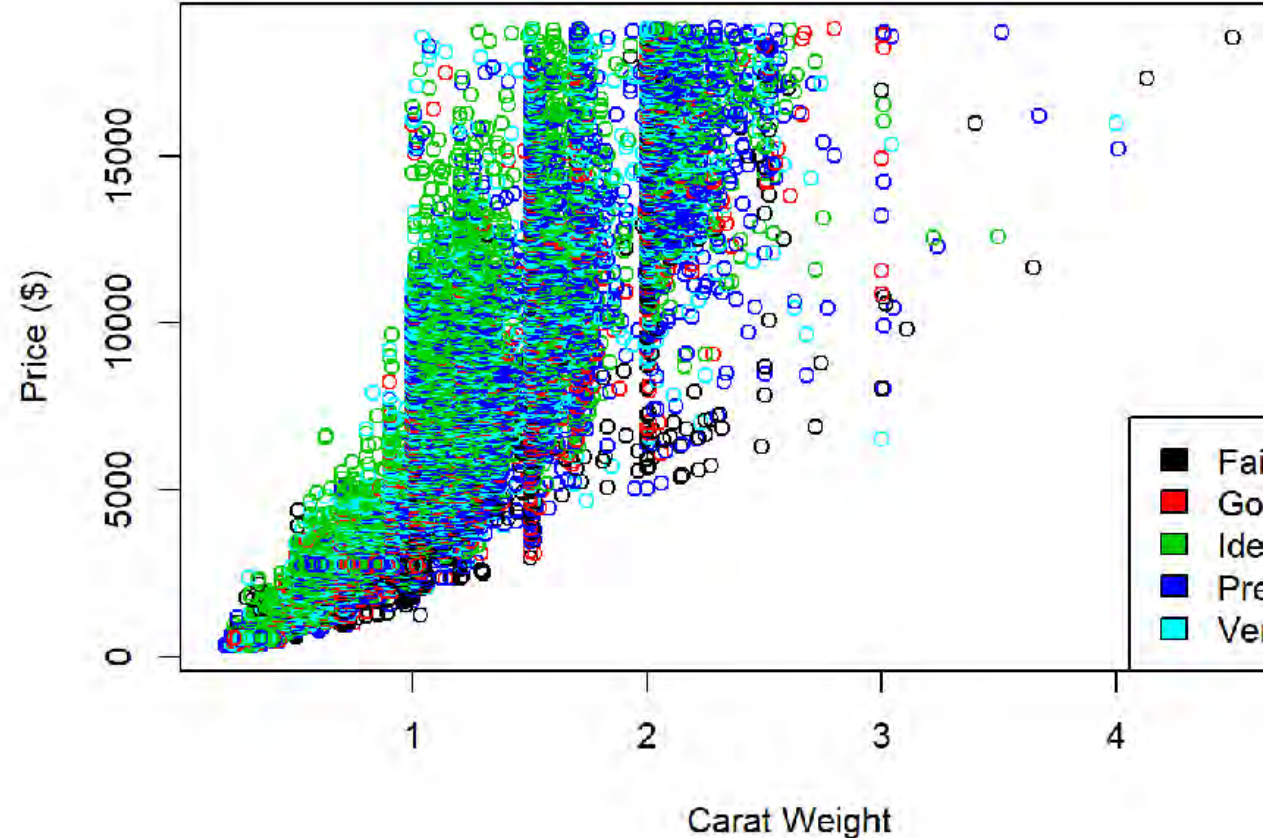
# What We are Doing Now

- Data Lab help pages
- R Help pages written using R script file with Markdown formatting
  - rmarkdown package renders the script as an html file

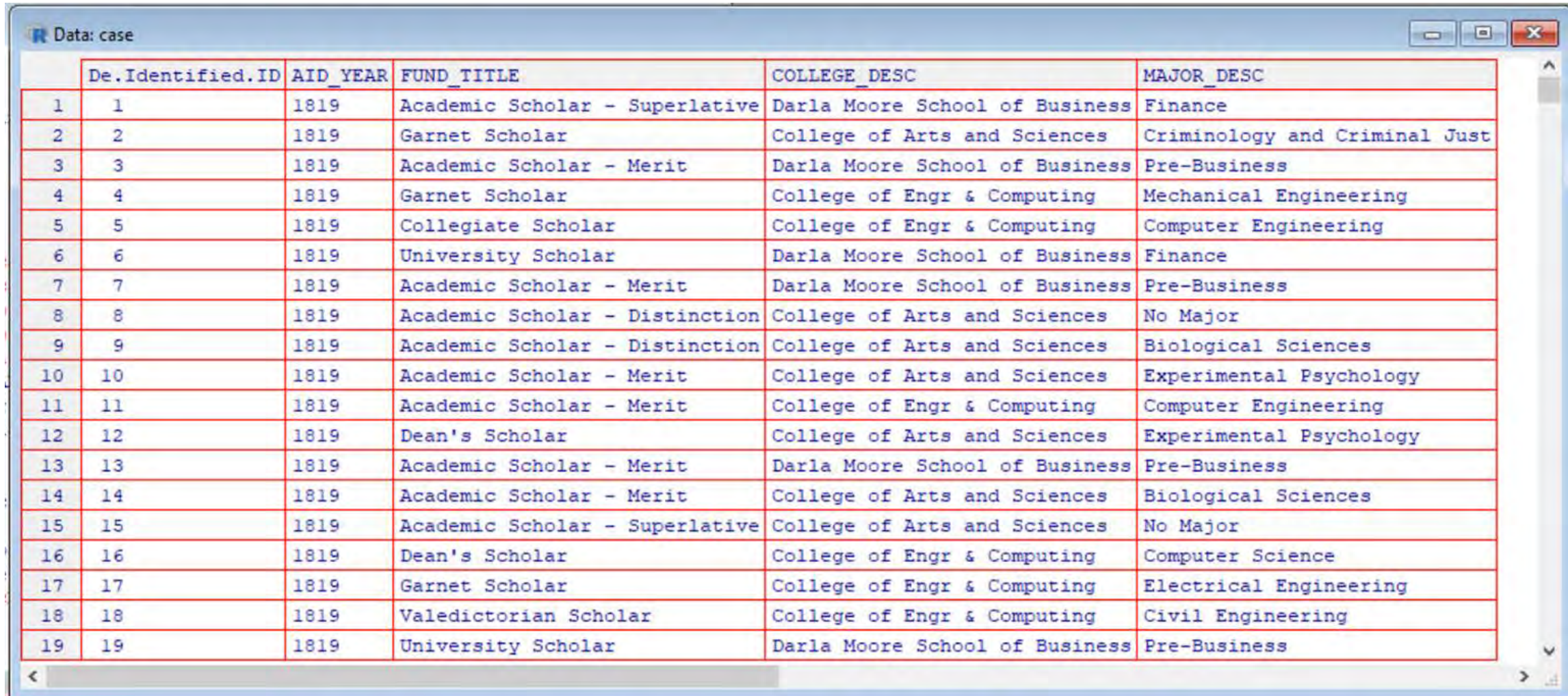
## Scatterplot

A scatterplot is a nice way to visualize the relationship between 2 quantitative variables. It se following scatterplot does this and adds a third variable, cut, to the graph.

```
plot(sparkly$carat, sparkly$price, col=sparkly$cut, xlab="Carat Weight", ylab="Pri  
legend("bottomright", fill=1:5, legend=levels(sparkly$cut))
```



# What We are Doing Now



The image shows a screenshot of an R Data Viewer window titled "Data: case". The window displays a table with 19 rows and 6 columns. The columns are labeled: De. Identified. ID, AID\_YEAR, FUND\_TITLE, COLLEGE\_DESC, and MAJOR\_DESC. The data is as follows:

	De. Identified. ID	AID_YEAR	FUND_TITLE	COLLEGE_DESC	MAJOR_DESC
1	1	1819	Academic Scholar - Superlative	Darla Moore School of Business	Finance
2	2	1819	Garnet Scholar	College of Arts and Sciences	Criminology and Criminal Just
3	3	1819	Academic Scholar - Merit	Darla Moore School of Business	Pre-Business
4	4	1819	Garnet Scholar	College of Engr & Computing	Mechanical Engineering
5	5	1819	Collegiate Scholar	College of Engr & Computing	Computer Engineering
6	6	1819	University Scholar	Darla Moore School of Business	Finance
7	7	1819	Academic Scholar - Merit	Darla Moore School of Business	Pre-Business
8	8	1819	Academic Scholar - Distinction	College of Arts and Sciences	No Major
9	9	1819	Academic Scholar - Distinction	College of Arts and Sciences	Biological Sciences
10	10	1819	Academic Scholar - Merit	College of Arts and Sciences	Experimental Psychology
11	11	1819	Academic Scholar - Merit	College of Engr & Computing	Computer Engineering
12	12	1819	Dean's Scholar	College of Arts and Sciences	Experimental Psychology
13	13	1819	Academic Scholar - Merit	Darla Moore School of Business	Pre-Business
14	14	1819	Academic Scholar - Merit	College of Arts and Sciences	Biological Sciences
15	15	1819	Academic Scholar - Superlative	College of Arts and Sciences	No Major
16	16	1819	Dean's Scholar	College of Engr & Computing	Computer Science
17	17	1819	Garnet Scholar	College of Engr & Computing	Electrical Engineering
18	18	1819	Valedictorian Scholar	College of Engr & Computing	Civil Engineering
19	19	1819	University Scholar	Darla Moore School of Business	Pre-Business

# What We are Doing Now



Question 6

10 pts

What is the average cumulative institutional GPA for computer science majors who are freshmen?

```
> case<-read.csv(file.choose(),strings=T)
> attach(case)
> mean(GPA[MAJOR=="Computer Science"&CLASS=="Freshman"])
```



# Multiple Choice Exams

## Question 2

3.33 pts

What does the following line of R code do?

```
myData <- read.csv("data.csv")
```

- Calls in a dataset and stores it as myData
- Calls in a dataset and stores it as myData.csv
- Calls in a dataset and stores it as data
- Calls in a dataset and stores it as data.csv

# Multiple Choice Exams

the probability that less than 105 will generate a profit in their first year

`dbinom(105,127,0.8)`

`pbinom(105,127,0.8)`

`dbinom(104,127,0.8)`

`pbinom(104,127,0.8)`

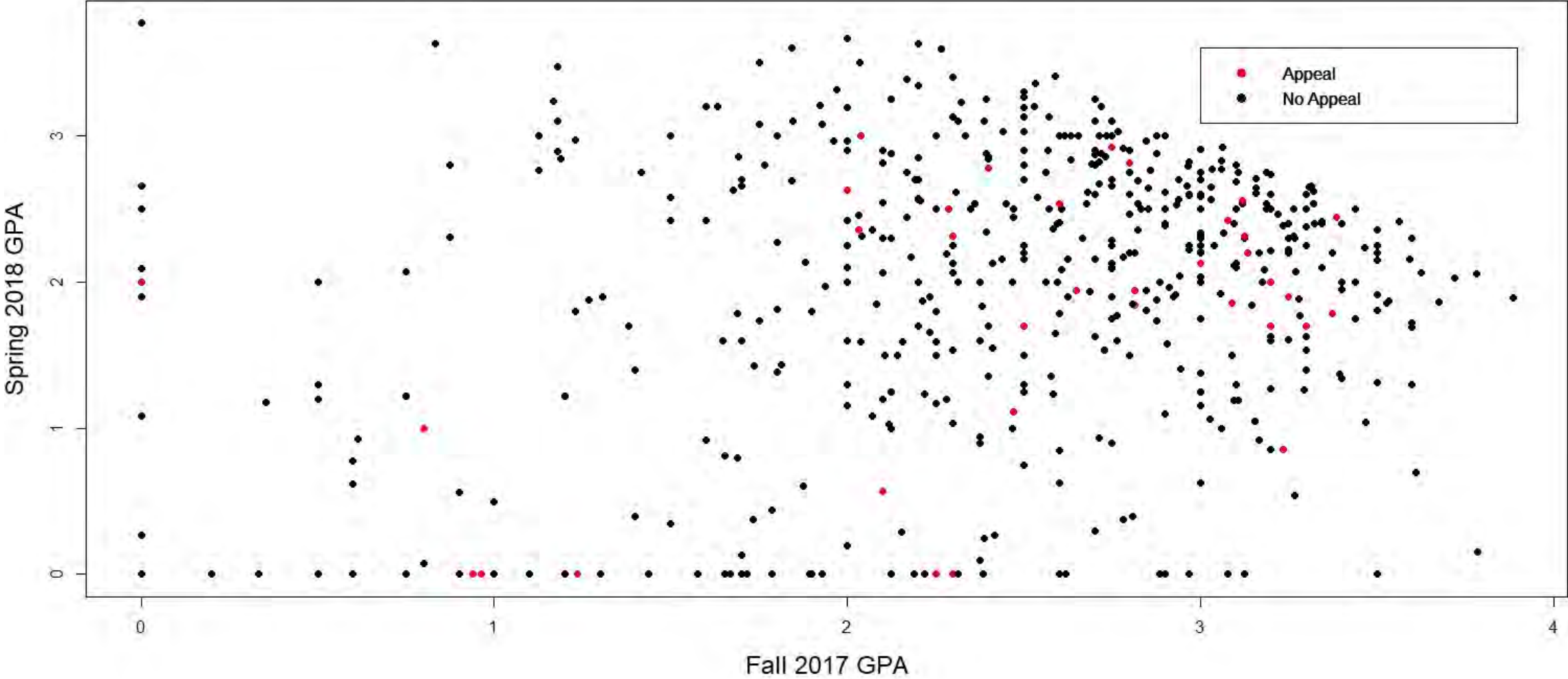
# Sample Student Work

## Hospital Charges Across the US



# Sample Student Work

Cumulative Institutional GPAs of Students with Suspended General University Scholarships (2017-2018 School Year)



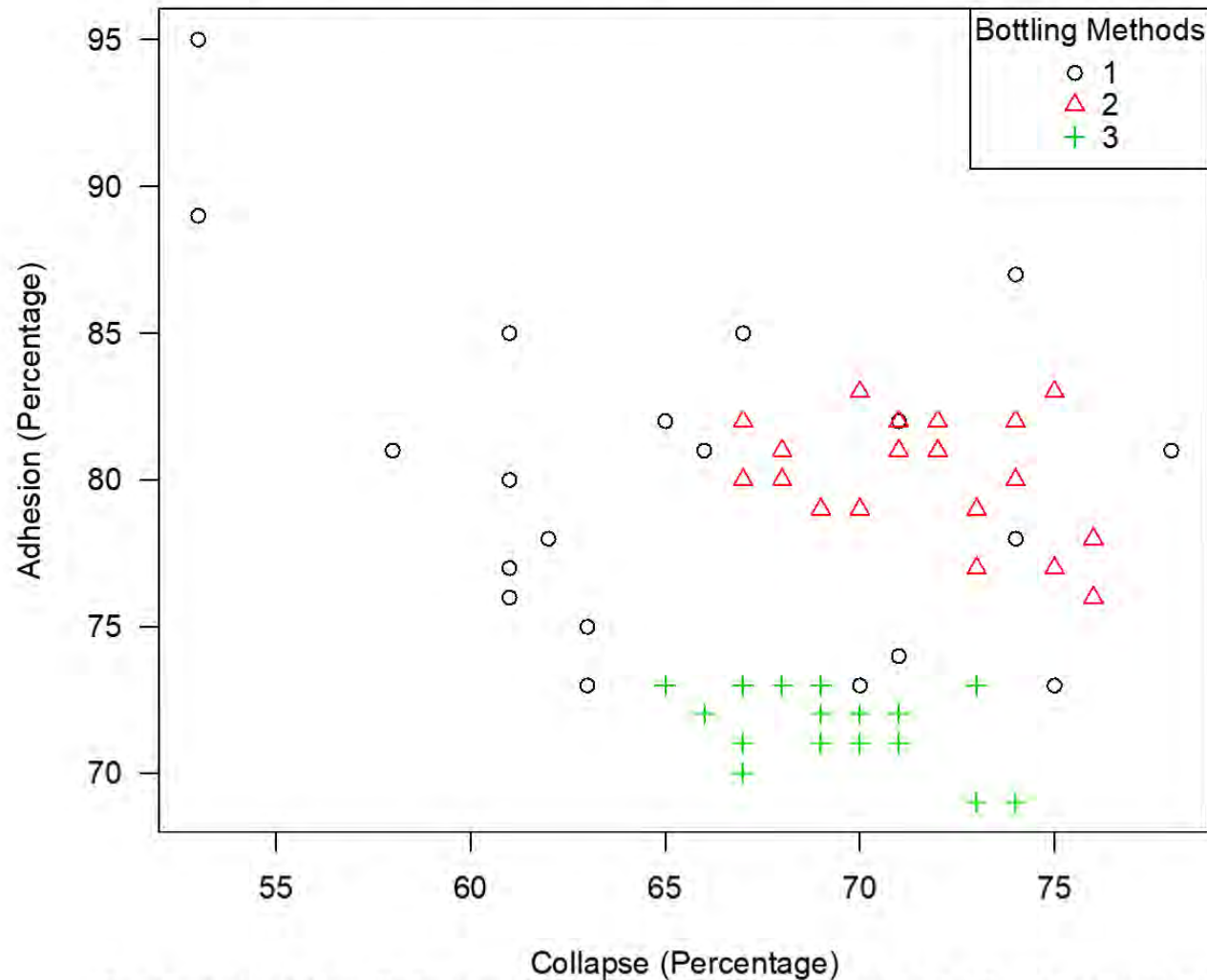
# Sample Student Work

```
plot(FALL_GPA,SPRING_GPA,pch=20,cex=1.5,col=ifelse(Appeal2018==1,"red","black"),xlab="Fall 2017 GPA",ylab="Spring 2018 GPA",main="Cumulative Institutional GPAs of Students with Suspended General University Scholarships (2017-2018 School Year)",cex.main=1.5,cex.lab=1.5)
```

```
legend(inset=0.1,x=3.0,y=3.6,legend=c("Appeal","No Appeal"),col=c(2,1),pch=c(20,20),bty="o",pt.cex=1.8,inset=0.1)
```



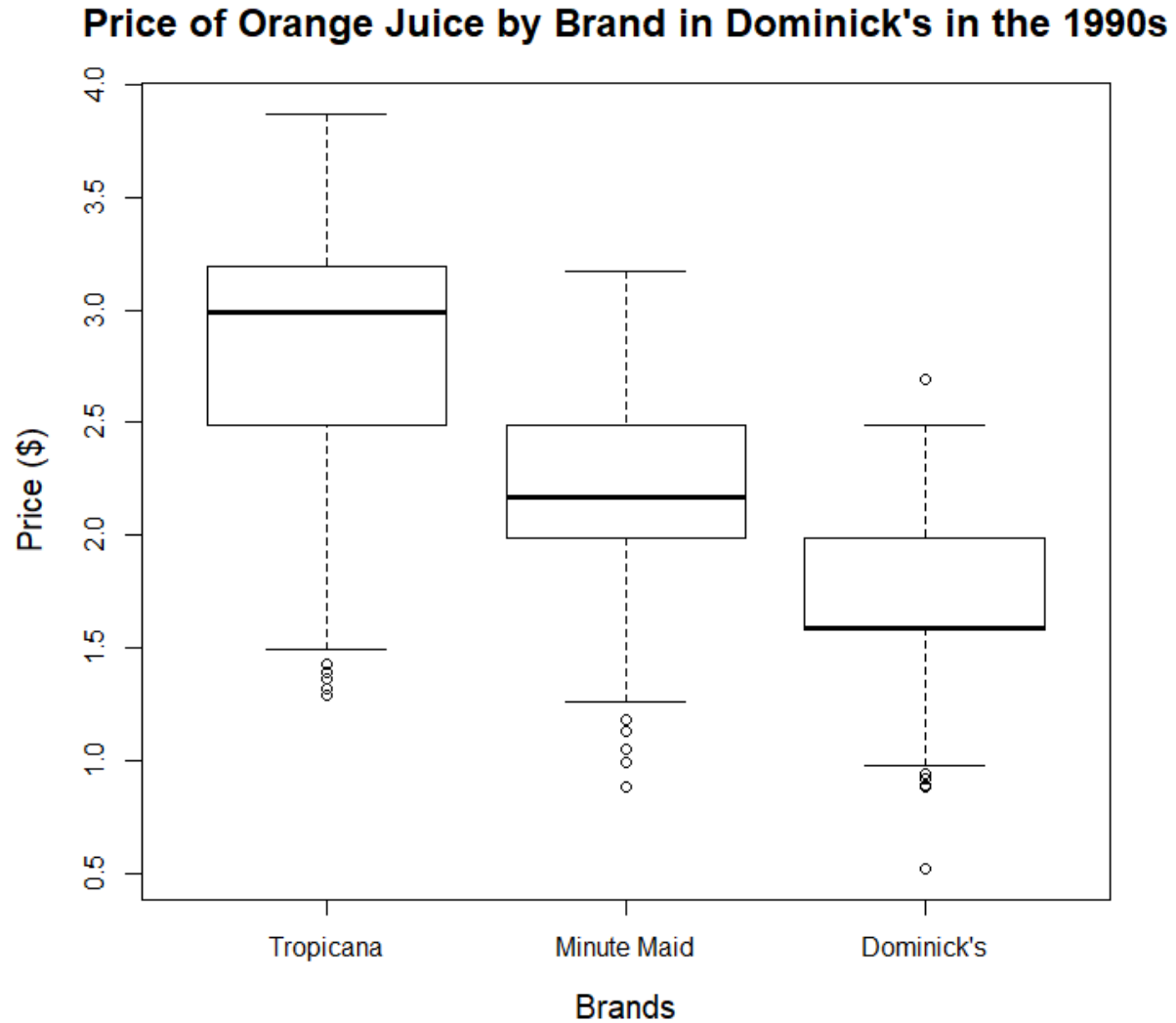
### Percentage of Collapse and Percentage of Adhesion of Beer Foam Shown in Different Bottling Methods



Data retrieved from R.G. Ault, E.J. Hudson, D.J. Linehan, and J.D. Woodward (1967)

# Sample Student Work

# Sample Student Work



# Challenges and Ideas

- Grading!
- Work on student perception
- Make R help pages better
- Training of Data Lab staff
  - Let Data Lab staff grade MGSC 291 projects



UNIVERSITY OF  
SOUTH CAROLINA

Walter R. Rouse Life Sciences Building  
Clemson University  
Data Lab

Data Lab