

Introduction and Motivation

Hypothesis

The Data Set

Graphs

Analysis/Inference

Conclusions

Future Directions

Works Cited

Analysis on the Effect of Students' Financial Contribution to Education on Class Attendance and Overall GPA

Code ▼

Rebecca Luo

27 April 2020

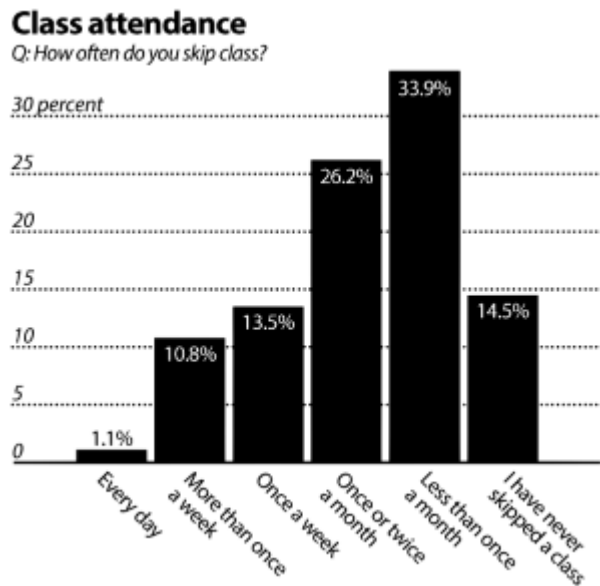


Introduction and Motivation

In a private institution such as Emory University, there is a large diversity of economic classes amongst students. Some students participate in federal work-study or other part-time jobs in order to contribute to the financial burdens of a higher education institution whereas others' tuition is paid for by a third-party, most commonly that student's parents. A study conducted by Laura Hamilton, a sociology professor at University of California, Merced, suggests that parental investments create a disincentive for student achievement. In other words, parental aid decreases students' GPA as students with parental funding often is associated with lower academic efforts.

As of 2020, tuition alone for Emory University is \$53,804 with around 56 percent of students receiving some sort of financial aid ("Types of Financial Aid"). Assuming a student takes 30 credits for the entire school year, each credit is worth around \$1793 and thus a 3-credit course is worth around \$5380. Even with such high costs for classes however, many college

students opt to skip classes frequently. According to the most recent survey by Class120 in 2015, the average college student skips 240 classes by the time he or she graduates. For students at private schools, it is \$24,960 over the course of 4 years.



Shown above, about one fourth of the student body skips class at least once a week, according to a Herald poll conducted March 3-4 (Dee, 2014).

I have always wondered if there exists such correlation between the number of classes a student would skip and if he or she is paying for part of their tuition. There is correlation suggesting that class attendance significantly improves student performance in terms of exam scores from a study by the Economics Department at UC Santa Cruz. Thus, in my project, I seek to analyze if there exists a correlation between whether if a student contributes financially to his/her tuition is dependent on financial aid, which would affect their tendency to skip class, and consequently their overall GPA.

Hypothesis

I believe that students who are on financial aid are more likely to pay a portion of their tuition. There exists a positive correlation between class attendance and overall average GPA. Students who contribute to their tuition and students on financial aid would skip less classes as opposed to those who do not and therefore have higher GPAs.

The Data Set

Data Gathering

The data used for this study was obtained from a class survey of 108 students. The questions in the survey that will be used are as follows:

- What is your GPA?
- Are you helping to pay for at least a portion of your Emory tuition?
- Are you on financial aid?
- What proportion of classes do you attend?

[Hide](#)

```
# upload data
load("C:/Users/12255/Box Sync/Econ 220 lab/Econ220DataS20_ano.Rdata")
# rename with a simple name
Prodata<-Econ220DataS20_ano
```

Data Cleaning

From the dataset, 4 new columns or variables were created. 2 categorical and 2 numerical.

NewGPA (Numerical)

GPA from the dataset shows the average GPA of each student in college. Answers provided should be rounded to two decimals. Data was modified to be numerical.

[Hide](#)

```
# view data for potenital changes
class(Prodata$GPA)
data.frame(table(Prodata$GPA))

#Change to numeric
Prodata$NewGPA <- as.numeric(Prodata$GPA)
#Prodata %>% select(NewGPA)%>% na.omit()
summary(Prodata$NewGPA)
data.frame(table(Prodata$NewGPA))
class(Prodata$NewGPA)
```

Tuition (Categorical)

Tuition (q194) examines whether a student is helping to pay for a portion of his/her tuition. Answers are yes or no. Data was modified to convert the class from character to factor.

[Hide](#)

```
# view data for potenital changes
class(Prodata$q194)
data.frame(table(Prodata$q194 ))

#Convert character to factor
Prodata$Tuition <- as.factor(Prodata$q194)
summary (Prodata$Tuition)
data.frame(table(Prodata$Tuition ))
class(Prodata$Tuition )
```

Aid (Categorical)

Financial aid (q96) examines whether a student is on financial aid or not. Answers are yes or no. Data was modified to change class from character to factor.

[Hide](#)

```
# view data for potenital changes
class(Prodata$q96)
data.frame(table(Prodata$q96))
#Convert character to factor

Prodata$Aid <- as.factor(Prodata$q96)
summary(Prodata$Aid)
data.frame(table(Prodata$Aid))
class(Prodata$Aid)
```

Classes (Numerical)

Class attendance (q121) tells the proportion of class attended by credit on average for each student. Answers were reported by credit: (i.e. I miss one class of chemistry (1.5 credits) and take 20 credits. I would report 18.5/20). Data was modified to change all answers to a proportion between 0 and 1. For more details view code below.

[Hide](#)

```
# view data for potenital changes
class(Prodata$q121)
data.frame(table(Prodata$q121))
Prodata$Classes <- Prodata$q121

#Turn 0.6 and 0.9 to fraction strings for later cleaning
Prodata$Classes <- ifelse(Prodata$Classes == "0.9", "9/10", Prodata$Classes)
Prodata$Classes <- ifelse(Prodata$Classes == "0.6", "6/10", Prodata$Classes)

#If string does not contain "/" (greater than 1) or if string == 0 make it 1 as well since students most likely meant th
ey skipped no classes
Prodata$Classes<- ifelse(!str_detect(Prodata$Classes, "/") | Prodata$Classes == "0", "1",Prodata$Classes)

#A student entered 0/24, if a student is taking 24 credits,they would most likely not skip all classes. Make this = 1 as
well.
Prodata$Classes<- ifelse(Prodata$Classes == "0/24", "1", Prodata$Classes)

#Assume N/A also means 1
Prodata$Classes <- ifelse(Prodata$Classes == "N/A", "1", Prodata$Classes)

#Change Classes from character to numeric and round to 2 decimal place

Prodata$Classes <- round(as.numeric(gsub("(\\d+)/((\\d+))", "\\1", Prodata$Classes, perl=T) ) / as.numeric(gsub("(\\d+)/((\\d+))", "\\2", Prodata$Classes, perl=T) ), digits = 2)
data.frame (table(Prodata$Classes))
```

Graphs

Tuition & Aid

GPA & Attendance

GPA & Tuition & Aid

Attendance & Tuition & Aid

First, let us look at tuition contribution status and financial aid status of all students.

Figure 1: Representation of students who are not on financial aid to see if they are paying for a portion of their tuition.

Hide

```
Prodata %>% filter(!is.na(Tuition)) %>%
  filter(Aid == "No")%>%
  group_by(Tuition)%>%
  summarise(Freq = n())

NewStats1 <- data.frame(Group = c("Not Pay Tuition", "Pay Tuition"),value = c(67,12))

# Compute percentages

NewStats1$fraction <- NewStats1$value / sum(NewStats1$value)
NewStats1$percentage <- round((NewStats1$value / sum(NewStats1$value))*100, digits = 2)

# Compute the cumulative percentages (top of each rectangle)
NewStats1$ymax <- cumsum(NewStats1$fraction)

# Compute the bottom of each rectangle
NewStats1$ymin <- c(0, head(NewStats1$ymax, n=-1))

# Compute label position
NewStats1$labelPosition <- (NewStats1$ymax + NewStats1$ymin) / 2

# Compute a good label
NewStats1$label <- paste0(NewStats1$percentage, "% ")

# Make the plot
ggplot(NewStats1, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=Group)) +
  geom_rect() +
  geom_label( x=3.5, aes(y=labelPosition, label=label), size=4) +
  scale_fill_brewer(palette=1) +
  coord_polar(theta="y") +
  xlim(c(2, 4)) + labs(title = "Non-Financial Aid Students on Tuition Contribution")+
  theme_void()
```


Non-Financial Aid Students on Tuition Contribution

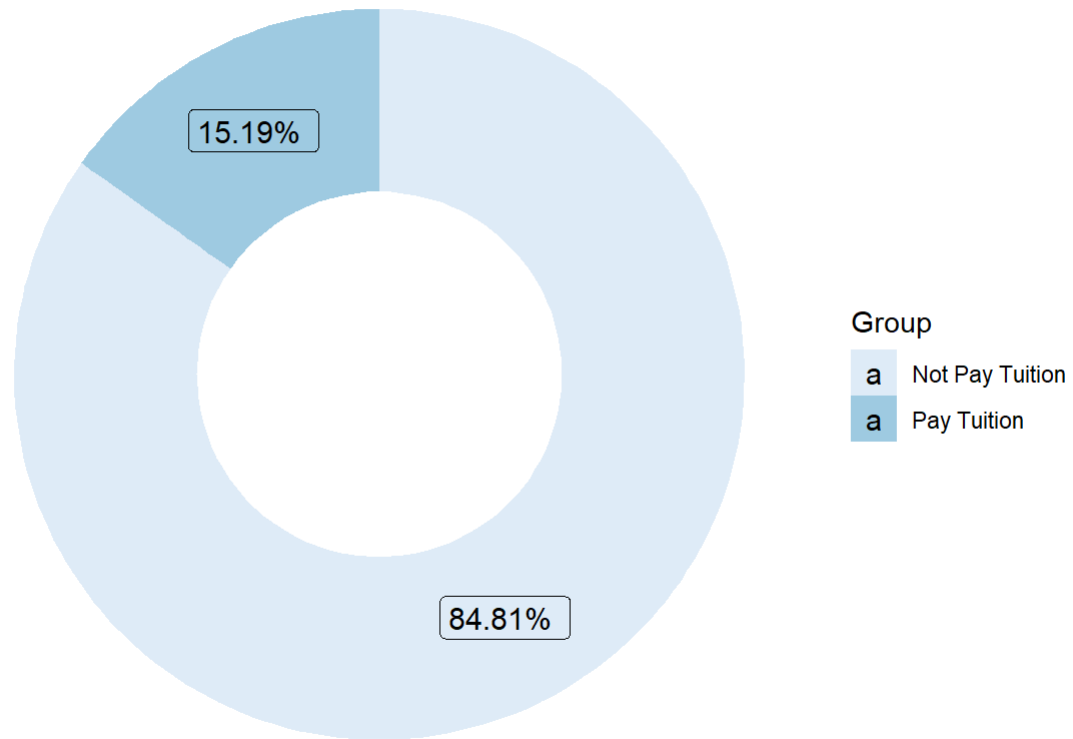


Figure 2: Representation of students who are on financial aid to see if they are paying for a portion of their tuition.

Hide

```
Prodata %>% filter(!is.na(Tuition)) %>%
  filter(Aid == "Yes")%>%
  group_by(Tuition)%>%
  summarise(Freq = n())

NewStats2 <- data.frame(Group = c("Not Pay Tuition", "Pay Tuition"),value = c(16, 11))

# Compute percentages

NewStats2$fraction <- NewStats2$value / sum(NewStats2$value)
NewStats2$percentage <- round((NewStats2$value / sum(NewStats2$value))*100, digits = 2)

# Compute the cumulative percentages (top of each rectangle)
NewStats2$ymax <- cumsum(NewStats2$fraction)

# Compute the bottom of each rectangle
NewStats2$ymin <- c(0, head(NewStats2$ymax, n=-1))

# Compute label position
NewStats2$labelPosition <- (NewStats2$ymax + NewStats2$ymin) / 2

# Compute a good label
NewStats2$label <- paste0(NewStats2$percentage, "% ")

# Make the plot
ggplot(NewStats2, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=Group)) +
  geom_rect() +
  geom_label( x=3.5, aes(y=labelPosition, label=label), size=4) +
  scale_fill_brewer(palette=1) +
  coord_polar(theta="y") +
  xlim(c(2, 4)) + labs(title = "Financial Aid Students on Tuition Contribution")+
  theme_void()
```

Financial Aid Students on Tuition Contribution

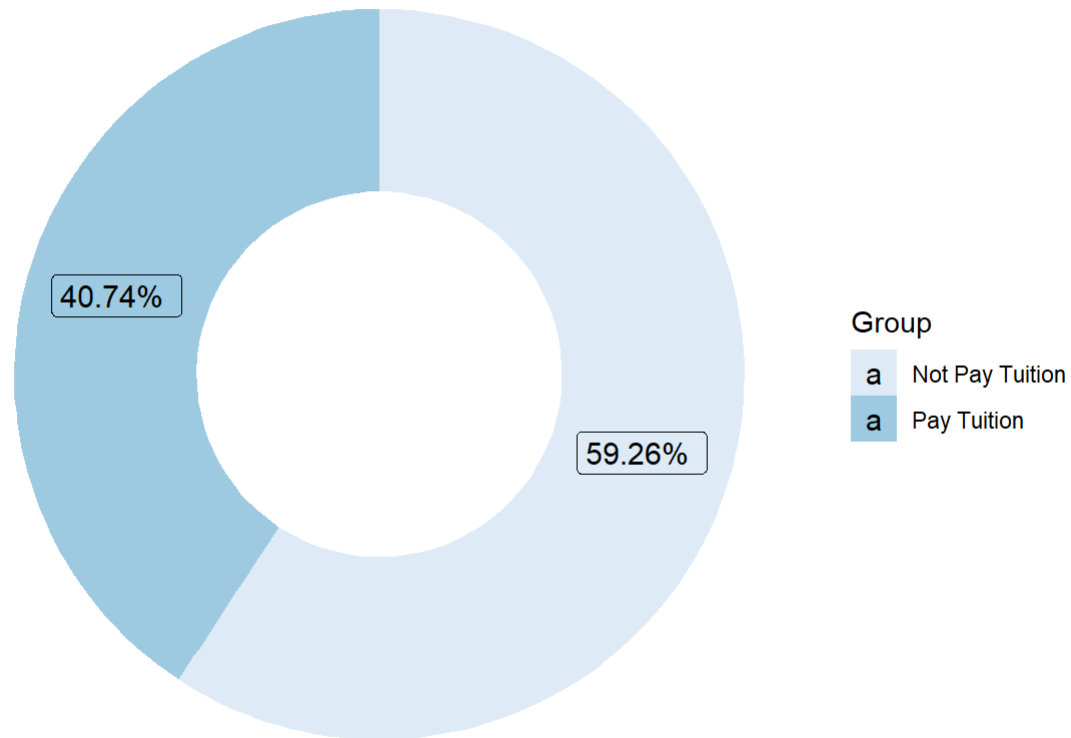


Figure 3: Representation of entire student population based on whether or not they are on financial aid and if they are paying for a portion of their tuition.

[Hide](#)

```
NewStats <- data.frame(Group = c("No Aid/ Not Pay Tuition", " No Aid/ Pay Tuition", " Aid/ Not Pay Tuition", "Aid/ Pay T
uition"),value = c(67, 12, 16, 11))

# Compute percentages

NewStats$fraction <- NewStats$value / sum(NewStats$value)
NewStats$percentage <- round((NewStats$value / sum(NewStats$value))*100, digits = 2)

# Compute the cumulative percentages (top of each rectangle)
NewStats$ymax <- cumsum(NewStats$fraction)

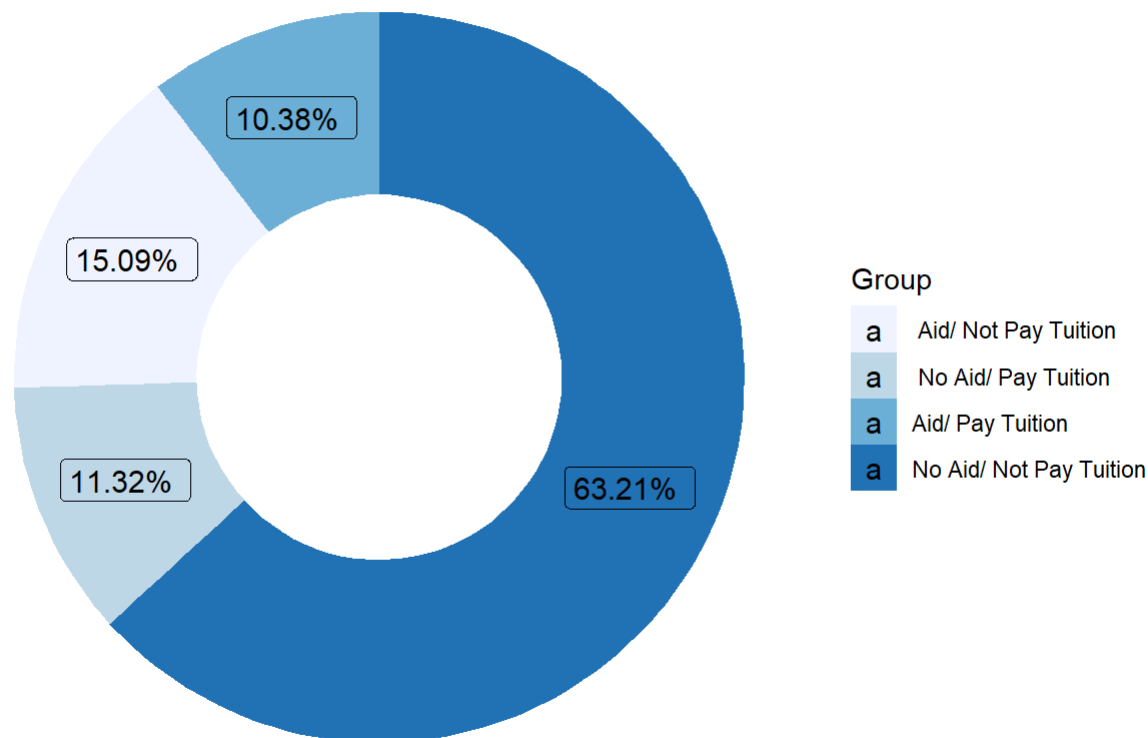
# Compute the bottom of each rectangle
NewStats$ymin <- c(0, head(NewStats$ymax, n=-1))

# Compute label position
NewStats$labelPosition <- (NewStats$ymax + NewStats$ymin) / 2

# Compute a good label
NewStats$label <- paste0(NewStats$percentage, "% ")

# Make the plot
ggplot(NewStats, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=Group)) +
  geom_rect() +
  geom_label( x=3.5, aes(y=labelPosition, label=label), size=4) +
  scale_fill_brewer(palette=1) +
  coord_polar(theta="y") + labs(title = "Students on Tuition Contribution and Financial Aid Status")+
  xlim(c(2, 4)) +
  theme_void()
```

Students on Tuition Contribution and Financial Aid Status



The table below summarizes the percentages of students in 4 categories concerning financial aid and contribution to their own tuition

Table 1:

Hide

```
PropStats <- data.frame(Group = c("No Aid/ Not Pay Tuition", " No Aid/ Pay Tuition", " Aid/ Not Pay Tuition", "Aid/ Pay Tuition"), Percentage = NewStats$percentage)

kable(PropStats, digits = 3, col.names = c("Groups of Students", "Percentage")) %>% kable_styling(bootstrap_options = c("striped", "hover"), full_width = F)
```

Groups of Students	Percentage
No Aid/ Not Pay Tuition	63.21
No Aid/ Pay Tuition	11.32
Aid/ Not Pay Tuition	15.09
Aid/ Pay Tuition	10.38

Analysis/Inference

The tables below show a summary of all students' class attendance and GPA based on tuition contribution status and financial aid status.

Table 2:

[Hide](#)

```
Prodata %>% filter(!is.na(Tuition)) %>% filter(!is.na(NewGPA)) %>% group_by(Tuition) %>% summarize(AvgClass=mean(Classess), AvgGPA = mean(NewGPA)) %>% kable(digits = 3, col.names = c("Tuition Contribution Status", "Average Proportion of Class Attended", "Average GPA"))%>% kable_styling(bootstrap_options = c("striped", "hover"))
```

Tuition Contribution Status	Average Proportion of Class Attended	Average GPA
No	0.918	3.577
Yes	0.947	3.644

Table 3:

[Hide](#)

```
Prodata %>% filter(!is.na(Aid)) %>% filter(!is.na(NewGPA)) %>% group_by(Aid) %>% summarize( AvgClass=mean(Classess), AvgGPA = mean(NewGPA)) %>% kable(digits = 3, col.names = c("Financial Aid Status", "Average Proportion of Class Attended", "Average GPA"))%>% kable_styling(bootstrap_options = c("striped", "hover"))
```

Financial Aid Status	Average Proportion of Class Attended	Average GPA
No	0.922	3.571
Yes	0.926	3.648

In order to test for statistical significance, I will look into the relationships of these variables in closer detail in the tabbed sections below.

[Tuition and Aid](#)
[GPA & Attendance](#)
[GPA & Tuition & Aid](#)
[Attendance & Tuition & Aid](#)

Null Hypothesis: No correlation between a student's tuition contribution status and financial aid status.

In order to test the correlation between a student's tuition contribution status and financial aid status, we will run the chi-squared test and the resulting p-value here can be seen as a measure of correlation between these two variables.

[Hide](#)

```
tbl = matrix(data=c(67, 12, 16, 11), nrow=2, ncol=2, byrow=T)
dimnames(tbl) = list(Tuition=c('N', 'Y'), Aid=c('N', 'Y'))
#calculate p-value
chi2 = chisq.test(tbl, correct=F)
c(chi2$statistic, chi2$p.value)
```

```
## X-squared
## 7.732182322 0.005424515
```

[Hide](#)

```
# calculate V
sqrt(chi2$statistic / sum(tbl))
```

```
## X-squared
## 0.2700835
```

A p-value of 0.005 and Crammer's V of 0.27 was obtained. Since is p-value is small enough, we can reject the null hypothesis of independence and conclude that tuition contribution status is depedent on financial aid status.

Conclusions

Through this analysis, it is shown that first, a student's tuition contribution status is highly dependent on if a student is receiving financial aid. 40.74% of financial aid students are paying for a portion of their tuition whereas only 15.19% of non-financial aid do the same. This can be attributed to perhaps financial aid students coming from a lower income level family and needing to work to have more flexibility in spending.

It is also shown that, as hypothesized, increasing class attendance is correlated to higher GPA. Although the correlation is not strong within the data we have collected, it is nevertheless positive and perhaps if the population sample were to be expanded to a more diverse group of students, the correlation would increase.

Although as seen from Table 2 and 3, students who contribute to their tuition has a higher class attendance and GPA (0.947, 3.644) than those who do not (0.918, 3.577). However, the difference is not significant in the population sample we used. Likewise, students who are on financial aid also has a higher class attendance and GPA (0.926, 3.648) than those who do not (0.922, 3.571) even though the difference is once again, insignificant. The inability to reject the null hypothesis can be attributed to either the null hypothesis is true or insufficient sampling in our data. Thus, more analysis is needed to fully conclude the effect of students' financial contribution to education on class attendance and overall GPA.

Future Directions

In our dataset, there are 63 percent of students who are not on financial aid and did not contribute to their tuition costs. This shows a potential bias in the data which could affect results. In order to fully conclude if school performance is associated with tuition contribution, the following can be improved in the future:

1. Analyze the results based on economic income brackets for each student as well.
2. Look into the actual amount of money students are contributing to their tuition.
3. Sample a more diverse population with students from different schools and majors to have a sample with equal distribution of students in terms of financial aid status and tuition contribution status.



Works Cited

Dee, Gabrielle. "Busy Schedules, Boring Lectures Drive Students to Skip Classes." Brown Daily Herald, 22 Apr. 2014, www.browndailyherald.com/2014/04/17/busy-schedules-boring-lectures-drive-students-skip-classes/.

Dignan, Sara. "The Cost of Skipping Class, by the Numbers." USA Today, Gannett Satellite Information Network, 26 Feb. 2016, www.usatoday.com/story/college/2016/02/26/the-cost-of-skipping-class-by-the-numbers/37413317/.

Hamilton, Laura T. "More Is More or More Is Less? Parental Financial Investments during College." American Sociological Review, vol. 78, no. 1, Feb. 2013, pp. 70–95, doi:10.1177/0003122412472680 (doi:10.1177/0003122412472680).

“Types of Financial Aid: Emory University: Atlanta GA.” Emory University | Atlanta GA, apply.emory.edu/financial-aid/types-of-aid/index.html.