University of California, Berkeley

Department of Statistics

International Students and the Trump Administration

Submitted by

Team Diversity

SeoHyeong Jeong, Jolina Yao, Katarina Palermo,

Jessica Chen, and Qi Ming Yeow

May, 2017

Table of Contents

I. Introduction

International students are becoming increasingly important for college and university institutions' financial stability (Ruiz, 2016)[1]. In addition to contributing a rich diversity to the student body of college campuses, international students are typically charged with higher tuition rates than state residents attending colleges in the US. Because international students pay more in tuition dollars, colleges are becoming more friendly to accepting international students due to the increase in profit.

President Donald Trump's rhetoric regarding immigration has created high tensions regarding US citizens and foreigners. The restrictions proposed by President Trump affect countries world-wide, however, the seven countries banned by executive order under the Trump administration are very sensitive to the changes that may occur throughout his presidency. Under the new Trump administration, many international students are fearful about their future in the US. Will they be able to stay to finish schooling? Will they be allowed to stay following the attainment of their degree? For prospective students abroad, will they even be allowed to get their education once they have been accepted to a college in the US?

All of these questions have led us to research the trends in college revenue, proportion of international students attending US colleges, and the changing political administrations in the US. Our aim is to determine whether these trends could help us predict how international students may be affected under the Trump Administration.

We had three main questions guiding our research:

1. How will the President, House of Representatives, and Senate affect international students, in terms of being granted visas to the US?
2. Is there a relationship between proportion of international students and college revenue?
3. How will international students from the seven countries recently banned under executive order be affected by the Trump administration?

In an effort to predict what may happen to international students and college revenue under the Trump Administration, we wanted to run two separate regressions. One to predict how proportion of international students across all US colleges may be affected and another to predict how students from the seven travel banned countries will be affected. This report takes you through our process of finding, processing, and visualizing the data, how we conducted our regression analyses, our interpretation of the data, and our conclusions.

[1] Ruiz, N. G. (2016, July 29). The Geography of Foreign Students in U.S. Higher Education: Origins and Destinations | Brookings Institution.

II. Data and Methods

Visa Data

Beginning this year, the new president of the United States, Donald Trump, signed an executive order that banned citizens of Iran, Iraq, Libya, Somalia, Sudan, Syria, and Yemen from entering the US for 90 consecutive days. Placing restrictions on certain travelers has occurred before, following the September 11th terrorist attacks in 2001. As the concern about people from predominantly Muslim countries grows, we decided to gather visa data[2] to research visa admission trends in these countries. Because of the way the data set is arranged, we focused on the admission of international student visas: F1, J1 and M1.

The visa data set is taken from the US Department of State, Bureau of Consular Affairs and contains visa admission information from 1997 to 2016 for each country in the world in an excel workbook. In order to make it tidy while keeping the country variable, we appended the data table vertically from 1997 to 2016 and made a new variable to indicate the year and whether the country is one of the seven travel ban countries. (See Appendix I.A.)

The variables of our interest are the total number of F1, J1, and M1 visas issued for each country[3]. F1 visas are issued to students who are attending an academic program or English Language Program. J1 is for students who need to obtain practical training, and M1 is for students who attend a non-academic or vocational school. To find the percentage of international student visa admissions in the US, we created an extra column indicating the statistics of the proportion of those visas.

Along with predicting the proportion of international students and college revenue under the new Trump administration, we wanted to carry out an analysis using linear regression model to see how the number of issued visas for international students is affected by the presidential party and year[4].

Delta Cost Project Database

The Delta Cost Project[5] is conducted by the American Institutes for Research, a non-profit organization that gathers publicly available data and conducts research on the social

---

[2] "Nonimmigrant Visa Statistics." U.S. Department of State. U.S. Department of State, n.d. Web. 7 May 2017. <https://travel.state.gov/content/visas/en/law-and-policy/statistics/non-immigrant-visas.html>.

[3] Note that how this can differ from the data obtained for international students.

[4] Further discussion in the next section on why and how did we picked such variable as an explanatory variable.

[5] "Home Page." Delta Cost Project. N.p., n.d. Web. 7 May 2017. <http://www.deltacostproject.org/>.

and behavioral sciences in health, education, and workforce productivity. The database we looked at is taken from the US Department of Education and contains information about each university on record from 1987 to 2013. The variables of interest are the number of international students enrolled, the number of total enrollment in the university, the amount of government funding the university receives, the revenue of the university, which includes direct income from the tuition of the students, as well as federal, state, and local fundings, and the academic year.

Methods used in relation to this data include using UNIX commands to clean the given dataset into workable size for RStudio. (See Appendix I.B.)

Congress Data

This data is available for public use from the Brookings Institution, a nonprofit public policy organization seeking to research and answer policy issues in the world. The data we obtained from this organization is *Political Parties of Senators and Representatives.*[6] The variables of interest from this data include chamber, which is either House of Representative or Senate, party status of each member, in which we focus only on the two major parties in the US - Democrat and Republican, seats, which is a numerical value that states how many members are in each category, and year. The time period covered in this data is from 1855 to 2015. (See Appendix I.C.)

Presidential Party Data

In order to find the data on each president's political party, we had to webscrape the information off of the web from "enchantedlearning.com".[7] We chose this website, because the tables are nicely formatted with well-formed XML syntax. This data contains information of each president's political party and the time period each president served. The variables of interest are the political party of the president and the year, which spans from 1841 to 2017. (See Appendix I.D.)

---

[6] "Vital Statistics on Congress | Brookings Institution." Brookings. Brookings, 29 Mar. 2017. Web. 07 May 2017. <https://www.brookings.edu/multi-chapter-report/vital-statistics-on-congress/>.

[7] The Presidents of the USA - EnchantedLearning.com. N.p., n.d. Web. 07 May 2017. <http://www.enchantedlearning.com/history/us/pres/list.shtml>.

III. Results and Discussion

Proportion of International Students in US by Years

Figure 3.1, 3.2 and 3.3 (on the following page) show the change in average percentage of international students in all universities across United States from 1987 to 2013. The gray, shaded areas indicate years of economic recession. The colors of the line of the plot, as indicated by the legend, shows the current political party in office that year. In order to compare the differences in party preferences of the three government institutions, the Presidential administration, the Senate, and the House of Representatives, we provide three separate graphs, colored by their respective party. As indicated by the legend, red portions of the line represent years with a Republican administration, whereas blue portions represent a Democratic administration. The significance of these party differences will be discussed later on, but we will first focus on major trends in the change in international students.

In terms of the general trend of the graph, one can easily see that the percentage of international students has been increasing stably from 1988 to 1998 by around 0.3% (despite the small dip in 1993), followed by a drop of 0.125%, then a sharp increase in 1999 from around 2.8% to 3.15%. The recession of 2001 coincides with a slight decrease of international student enrollment rate, but the rate then continues to increase until 2004, reaching past 3.5%. For the next seven years, the enrollment rate dropped back to around 3.2% in 2011, then surged to an all time high 3.7% by 2013, which is the last year of the data we were able to collect.

We cannot see much of an relationship between the international students enrollment rate and recession in this graph, but the political party in office follows closely with the international students enrollment rate. For all four sections of the graph, namely the first red, first blue, second red, and second blue sections, the ending point is always higher than the starting point. This shows that. In general, both parties are correlated with increase in international student rates. However, the strength of growth in international students enrollment rate during the period of time when Democratic president is in office is much higher than that when Republican president is in office.

Interestingly, the graph for *Proportion of International Students in US by Years* mirrors closely with the graph for *Proportion of College Revenue in GDP by Years*, notably, the sharp increase in both graph at 1999. From the similarity of both of these graph, we can see that the two variables, proportion of international students and the proportion of college revenue in GDP may be correlated with one another, thus we decided to further the investigation through another form of measure to calculate international students in the US - through visas issued.

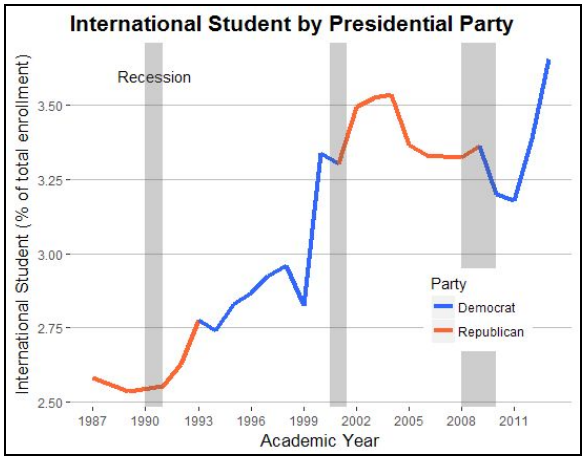


Figure 3.1

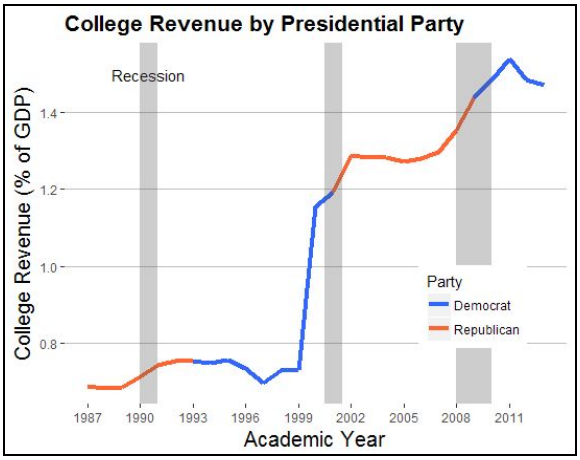


Figure 3.4

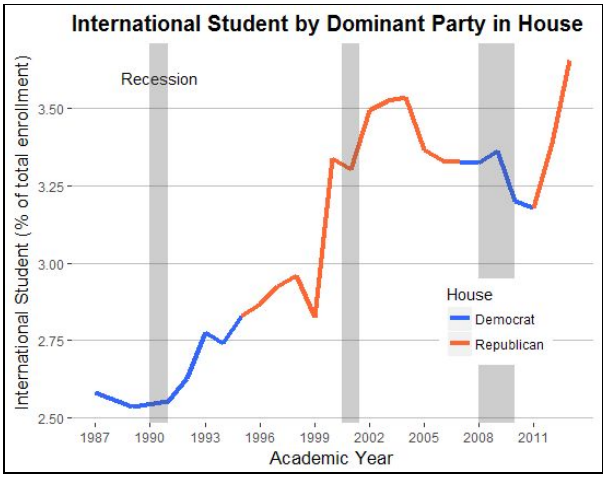


Figure 3.2

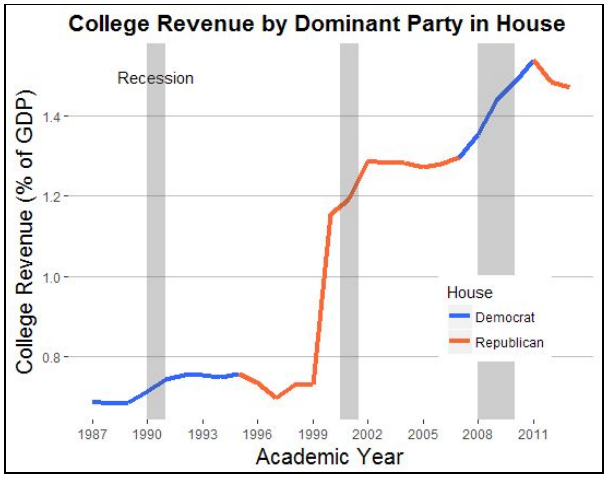


Figure 3.5

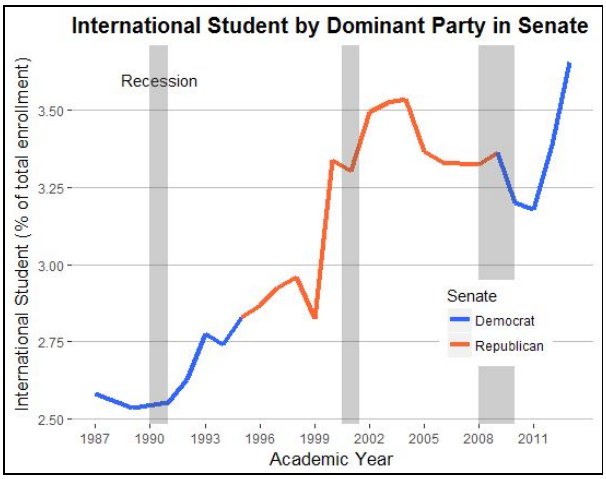


Figure 3.3

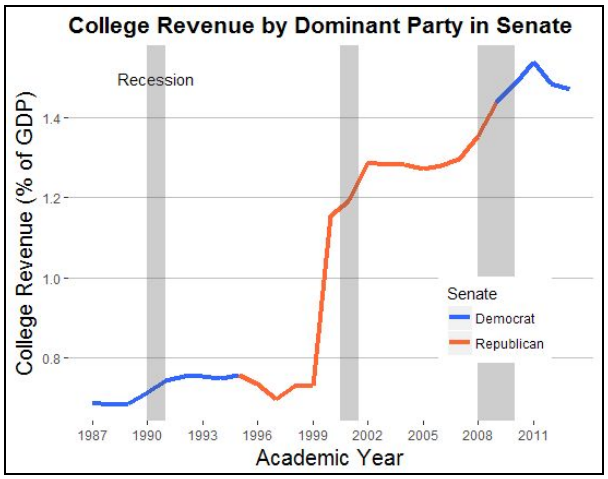


Figure 3.6

Finally, we can use our findings to predict the upcoming changes in the Trump administration. In the graphs above, we can observe that from 1987 to 2013, the two chambers of Congress were often dominated by the same political parties, while the President was under the opposite party. However, from 2001 to 2007, all three institutions held a Republican party preference. This scenario matches the current situation under Trump, where all three institutions are majority Republican as well.

Proportion of College Revenue in GDP by Years

Figure 3.4, 3.5, and 3.6 (on the previous page) show change in college revenue, as a percentage of national GDP, from 1987 to 2013. Again, the vertically shaded years represent economic recessions and the different graphs show the respective party preferences of the three government institutions, the President, the Senate, and the House of Representatives. In the upcoming regression analysis, we will focus on the years from 2001 to 2007, where all three institutions were from Republican which correspond to the current situation.

From 1987 to 2013, we can observe that overall, similar to the change in international students, college revenue relative to the proportion of the national GDP also significantly increased. From 1987 to 1995, growth in proportion of college revenue relative to GDP increases from ~0.7% to ~0.75%. Beginning in 1995, the proportion of GDP that constitutes college or university revenue dips, rises, and plateaus until 1999, when an enormous increase occurs. This increase goes from ~0.7% of the GDP to ~1.17% in the span of one year. The years following continue to increase but at a lower rate. By 2002, the proportion rose up to ~1.27%. This remained as the highest proportion for the following 3-4 years with a slow dip followed by an increase from 2005 to 2011. During this increase, the highest proportion of college revenue relative to GDP reached ~1.6% in 2011. Following this peak, a decline is observed until 2013.

Based off of these two sets of graphs, Figures 3.1 - 3.3 and Figures 3.4 - 3.6, we found that from 1987 to 2013, the proportion of college revenue relative to the GDP tended to exhibit two trends: a dramatic increase in 1999 and significant increases during periods of economic recession. In an effort to explain the 1999 increase of college revenue, we considered possible changes in government policies. After some researching, we discovered that federal and state funding for higher education decreased around that time (Schoen 2016)[8]. With lower funding, higher education institutions would attempt to recover their losses by increasing tuition fees, and this increase could possibly lead to increased revenue and profits.

Other than the 1999 increase, the greatest rates of college revenue increases occurred during the three economic recessions. In times of economic recession, it becomes more difficult

[8] Schoen, John W. "The Real Reasons a College Degree Costs so Much." *CNBC*. CNBC, 08 Dec. 2016. Web. 07 May 2017. <http://www.cnbc.com/2015/06/16/why-college-costs-are-so-high-and-rising.html>.

to secure jobs. Consequently, we predicted that those who lost their jobs or could not get hired resorted to returning to school for more education instead.

Proportion of International Students from the Seven Countries Targeted by the Travel Ban during Trump's Administration

The next part of our analysis focuses on J1, F1, M1 visas from the seven countries targeted by Trump's travel ban executive order. Since international student statistics in the dataset are not organized by country, we used an alternate dataset, the Visa Data. We found out that visa admission rate does not exactly correspond with the proportion of international students reported in our other dataset. With these factors in mind, we extracted F1, J1, and M1 visa admissions for each country and compared the admission rate of the seven countries with the other countries. Taking year and Presidential party as explanatory variables, we predicted the percentage of international student visa admissions by using the linear regression model we learned in Statistics 133 lecture.

Figure 3.7 shows the trend in average international student visa admission (% of total enrollment). Normalizing the value with respect to total US college student enrollment tells us how many students with those visas are in college. A downsloping trend is observable, however Presidential party does not seem to be a contributing factor. Shaded areas, which represent two economic recessions (in 2001 and 2009), affect the trend more than the presidential party. For each recession, there was a big drop observed in average international student visa admissions. Figure 3.7 can be divided into admission (% of total enrollment) in seven countries on the list of Trump's executive travel ban order and the others. The mint line in Figure 3.8 shows that the admission rate is more volatile in those seven countries, responding more sensitively to factors.

Looking at the absolute number of international students visa admission in the seven countries gives a clearer look at how the Presidential party influences these numbers (Figure 3.9). There were less F1, J1 and M1 visas issued when Republicans were in office. Year, which shows the downsloping trend, and Presidential party, which shows a difference in number of visas issued, seems to be suitable as explanatory variables in the prediction of F1, J1 and M1 visa admission for the seven countries. In addition to Presidential party, majority party in the Senate and House of Representatives were also considered explanatory variables, but Figure 3.11 and Figure 3.12 show that there is either not enough data or not a clear enough relation.

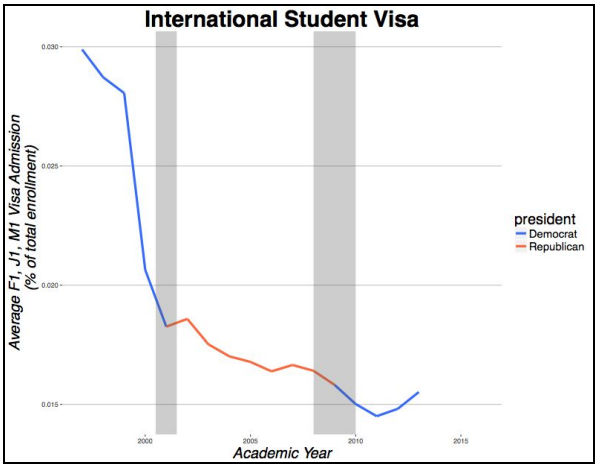


Figure 3.7

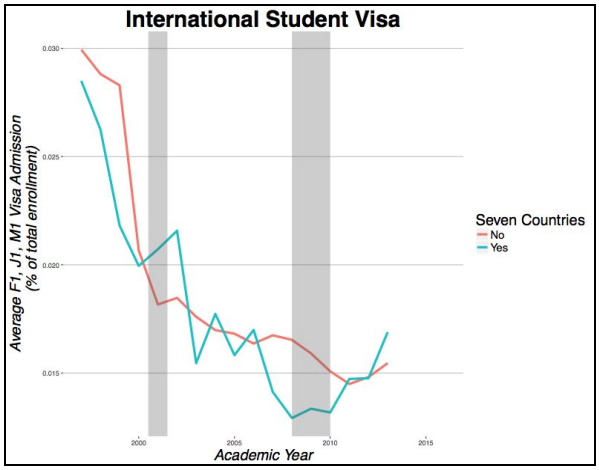


Figure 3.8

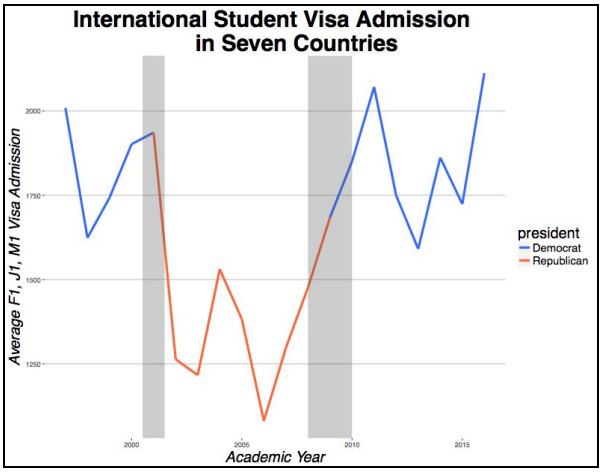


Figure 3.9

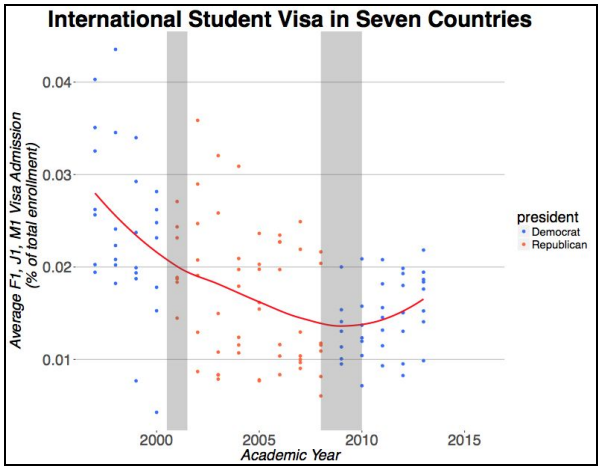


Figure 3.10

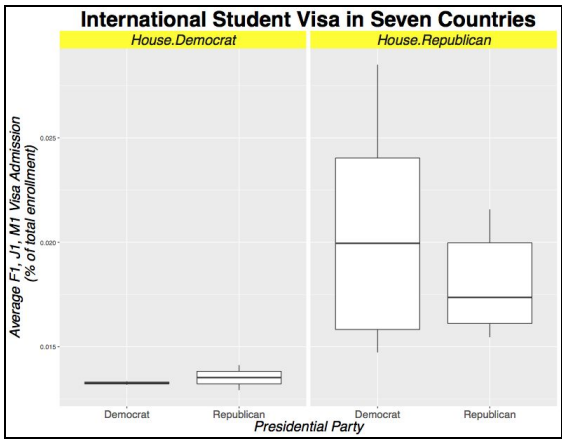


Figure 3.11

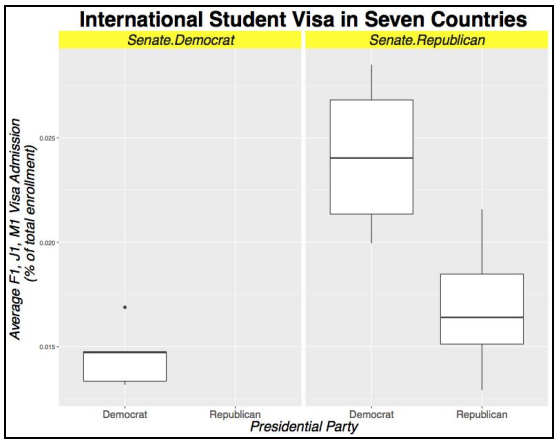


Figure 3.12

IV. Regression Analysis

Proportion of International Students and College Revenue Prediction

Our goal in this section is to create models to predict the proportion of international students and proportion of college revenue generated based on the existing data. First, we will use political parties (Presidential, House and Senate) to predict the proportion of international students. As political parties are categorical variables, we will use recursive partition to model our data, which can be described in the equation below. From the model, it appears that only House majority party plays a significant roles in determining the proportion of international students. The result shows that the proportion of international students to be 2.40% if the House party is majority Democrat and 3.20% if the majority is Republican. (See Appendix IV. A.)

$$rpart(Prop\_Intl\_Student \sim Party + Senate + House)$$

Evaluating the above model with current political parties; Presidential party being Republican, in addition to majority of House of Representatives and Senate being Republican, the model predicts that the proportion of international students is 3.234834%, which is lower as compared with the proportion of international students in 2013. Apart from that, we also evaluate the above model when all the political parties are democrat. The result shows that the proportion of international students is 2.421612%, which is even lower than the current predicted level.

Figure 4.1 shows a positive relationship between college revenue and proportion of international students over the years. From the graph, we conclude that linear regression is not suitable for modelling as the variables are not linearly correlated.[9] Hence, we use a polynomial regression.

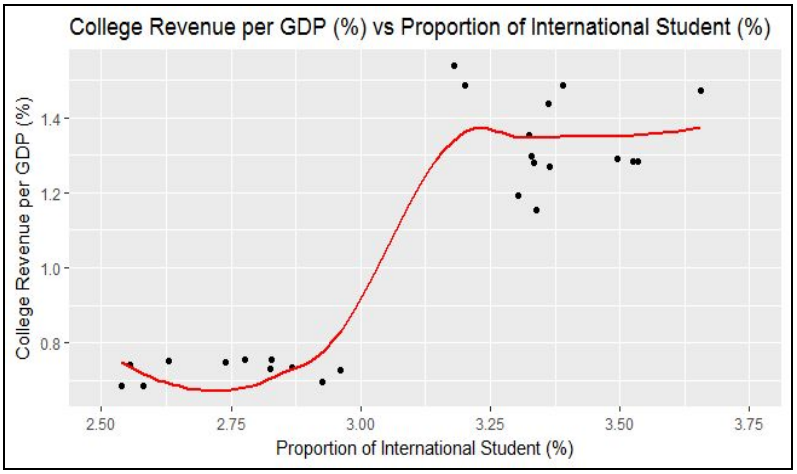


Figure 4.1

[9] We first included government funding as an explanatory variable. However, from the result we obtained, government funding turned out to be non-trivial.

The linear regression model that we designed is as follows:

$$lm(Income\_per\_GDP \sim poly(Prop\_Intl\_Student, 2))$$

We evaluated the regression model by plugging the estimate that we got from our previous prediction. At the level of 3.234834% of international students, the model predicts the college revenue to be 1.160418% ± 0.6063204% of annual GDP.

Prediction for the Proportion of International Students from the Seven Travel Ban Countries

Going back to Figure 3.9 again, plotting with year on the x-axis and average admission of F1, J1, M1 visas in seven countries on the y-axes, international student visa admissions show different trends under Democrat Presidents and Republican Presidents. Figure 3.10 is also plotted with year on the x-axis but instead of average admission of F1, J1, M1 visas, we normalized the admission visa rate by dividing the number by total US college student enrollment. Figure 3.10 shows the downsloping trend with respect to time line. Note that the second period of a Democrat President the percentage is increasing again despite an overall downsloping trend. We have not included extra dummy variable to indicate this pattern but it seems that Presidential party not only affects the absolute number, but also the proportion of international students with F1, J1 and M1 visas from the seven countries in the total US college student population.

$$lm(prop \sim president + Year)$$

Executing linear regression on the average proportion of international student visa admission in seven countries by presidential party and year gives following summary. Model p-Value and p-Value of individual predictor variables are less than 0.05 (see Appendix IV. B.). Hence, we can consider a linear model to be statistically significant. Using this model, we find that under the Trump administration (Republican), in 2020 the average proportion of international students from all of seven travel ban countries will be 0.004799% (with the standard deviation of 7.445756e-05%) of the total college enrollment in the US. Compared to the 2014 level, which is 0.01689229%, the model predicts the population of international students from those seven countries to decrease. Furthermore, comparing it to the prediction of 2020 under democrat president, which is 0.00744450%, prediction given by the model under republican president is lower.

V. Conclusion

Results Summary

Based on our analysis, we found a positive relationship between college revenue and proportion of international students over the years. Our first model predicted that under the Trump administration (with a majority Republican in both House and Senate), the proportion of international students will be 3.234834% (with a standard deviation of 7.445756e-05%). This value is well below 2013 levels pictured in the graphs earlier in the document. By plugging this number into our next regression equation, our model predicts that college revenue will be 1.160418% ± 0.6063204% of annual GDP, which is also much lower than 2013 levels. Our second analysis involved the seven countries targeted by Trump's travel ban executive order. Our model predicted that in 2020 the average proportion of international students from the seven travel ban countries will be 0.004799% ± 7.445756e-05% of the total college enrollment in the US. Compared to the 2014 level, which is 0.01689229%, the model predicts the population of international students from those seven countries to decrease significantly.

Overall, our predictions lead us to believe that under the new Republican Trump's Administration, international students and college revenue will decrease. In addition, students from the seven countries targeted by the travel ban will see a bigger decrease than other international students.

Limitations and Departing Thoughts

As we carried out our analysis, we came across some limitations that we find important to discuss. First, regarding our definition of college revenue, we think there is more depth in what the revenue of a college consists of other than our chosen variable that consists of federal, state, and local funding granted to colleges in addition to tuition fees. To explore this further, we believe that conducting analysis for each of the variable separately and/or other sources of funding for a college would be interesting to consider. By using this data, we believe we could find tighter relationships between certain types of colleges (public or private, liberal arts or research colleges), and how big of a role international students play in financing those colleges.

Second, we decided to normalize our data by averaging our table data. Although we managed to get compelling results with our method, we believe it would be helpful to explore different kinds of normalization techniques such as median, mode, or weighted values. This could have led us to more detailed conclusions about the trends we were observing by adding careful perspectives in our data processing.

Third, we found it difficult to determine whether or not the monetary data we received was adjusted for inflation. Although our analysis was observing trends in the last 20 years, we think adjusting for inflation could have been a helpful tool in getting an accurate picture of the data.

Lastly, time was our biggest limiting factor. If we wanted to go further with these research questions and include our previous suggestions, we would need much more time and resources to do a thorough analysis. This topic is interesting and very important for understanding how political parties influence change for international students. To do a complete analysis would take months of refining and would go beyond the scope of this course. Given the tools we learned in Statistics 133 and the time we had as full time students at UC Berkeley, we believe this project begins to answer the bigger questions about political party influence in the the United States, and how that influence affects visa admissions and international students.

Appendix

I. Data Cleaning

A. Visa Data

```
library(DataComputing)

V97 <- read.csv("/Users/jessicachen/Downloads/FY97.csv")
V98 <- read.csv("/Users/jessicachen/Downloads/FY98.csv")
V99 <- read.csv("/Users/jessicachen/Downloads/FY99.csv")
V00 <- read.csv("/Users/jessicachen/Downloads/FY00.csv")
V01 <- read.csv("/Users/jessicachen/Downloads/FY01.csv")
V02 <- read.csv("/Users/jessicachen/Downloads/FY02.csv")
V03 <- read.csv("/Users/jessicachen/Downloads/FY03.csv")
V04 <- read.csv("/Users/jessicachen/Downloads/FY04.csv")
V05 <- read.csv("/Users/jessicachen/Downloads/FY05.csv")
V06 <- read.csv("/Users/jessicachen/Downloads/FY06.csv")
V07 <- read.csv("/Users/jessicachen/Downloads/FY07.csv")
V08 <- read.csv("/Users/jessicachen/Downloads/FY08.csv")
V09 <- read.csv("/Users/jessicachen/Downloads/FY09.csv")
V10 <- read.csv("/Users/jessicachen/Downloads/FY10.csv")
V11 <- read.csv("/Users/jessicachen/Downloads/FY11.csv")
V12 <- read.csv("/Users/jessicachen/Downloads/FY12.csv")
V13 <- read.csv("/Users/jessicachen/Downloads/FY13.csv")
V14 <- read.csv("/Users/jessicachen/Downloads/FY14.csv")
V15 <- read.csv("/Users/jessicachen/Downloads/FY15.csv")
V16 <- read.csv("/Users/jessicachen/Downloads/FY16.csv")

Visas <-
list(V97,V98,V99,V00,V01,V02,V03,V04,V05,V06,V07,V08,V09,V10,V11,V12,V
13,V14,V15,V16)

F1 <- c()

for (i in 1:length(Visas)){
  names(Visas[[i]])[1] <- "Year"
  F1[[i]] <- Visas[[i]] %>% filter(Year == "Grand Totals") %>%
select(F.1) %>% mutate(Year = 1996 + i)
}

F1.Year <- F1[[1]]
for (i in 1:length(F1)){
  F1.Year <- full_join(F1.Year, F1[[i]])
}

F1.Year <- F1.Year[,c(2,1)]

write.csv(F1.Year, file = "F1.Year.csv", row.names = FALSE)
```

```
Total <- c()

for (i in 1:length(Visas)){
  names(Visas[[i]])[1] <- "Year"
  Total[[i]] <- Visas[[i]] %>% filter(Year == "Grand Totals") %>%
select(Grand.Total) %>% mutate(Year = 1996 + i)
}
Visas.Year <- Total[[1]]
for (i in 1:length(Total)){
  Visas.Year <- full_join(Visas.Year, Total[[i]])
}

Visas.Year <- Visas.Year[,c(2,1)]

write.csv(Visas.Year, file = "Visas.Year.csv", row.names = FALSE)

Comparison <- full_join(F1.Year,Visas.Year)
Comparison <- Comparison %>% mutate(F.1 = gsub(",","",F.1),
Grand.Total = gsub(",","",Grand.Total)) %>% mutate(F.1 =
as.numeric(F.1), Grand.Total = as.numeric(Grand.Total))

write.csv(Comparison, file = "Comparison.csv", row.names = FALSE)

options(scipen=999)

graph1 <- Comparison %>% ggplot(aes(Year)) + geom_line(aes(y=F.1,
group =1), col = "Blue") + geom_line(aes(y=Grand.Total, group = 1),
col = "Red") + scale_y_continuous(breaks = seq(0,15000000, by =
500000))
Graph1
```

B. Delta Cost Project Database

```
## Unix Data Cleaning
 mkdir 133final_project
 cd 133final_project
 less delta_public_release_00_13.csv
 cat delta_public_release_00_13.csv |
       cut -d “,” -f 2,6,55,81,330,338 > data_clean_00_13
 cat delta_public_release_87_99.csv |
       cut -d “,” -f 2,6,55,81,330,338 > data_clean_87_99

## Input the data into R
data_00_13 <-
read.csv("C:\\Users\\QiMing\\Downloads\\Delta_database_87_13_CSV\\data_c
lean_1.csv")
data_87_99 <-
read.csv("C:\\Users\\QiMing\\Downloads\\Delta_database_87_13_CSV\\data_c
lean_2.csv")

## Clean up and summarise the data table
```

```
data_0013_clean <- data_00_13 %>%
  group_by(academicyear) %>%
  summarise(Operating_Income = sum(stable_operating_rev, na.rm = TRUE),
            Total_Enrollment = sum(total_enrollment, na.rm = TRUE),
            Total_Intl_Enrollment = sum(total_enrollment_nonres_tot,
na.rm = TRUE))
data_8799_clean <- data_87_99 %>%
  group_by(academicyear) %>%
  summarise(Operating_Income = sum(stable_operating_rev, na.rm = TRUE),
            Total_Enrollment = sum(total_enrollment, na.rm = TRUE),
            Total_Intl_Enrollment = sum(total_enrollment_nonres_tot,
na.rm = TRUE))
education_budget <-
read.xlsx("C:\\Users\\QiMing\\Downloads\\edhistory.xls")
Government_funding_clean <- education_budget[43,] %>%
  gather(key, value, -NA.) %>%
  filter(grepl("X[0-9]{4}", key)) %>%
  select(-NA.) %>%
  mutate(Year = as.numeric(gsub("X", "", key)),
         Government_funding = (as.numeric(value) * 1000)) %>%
  select(Year, Government_funding)
data_clean <- rbind(data_0013_clean, data_8799_clean)
data_clean <- data_clean %>%
  arrange(academicyear)
```

C. Congress Data

```
congressRaw <- read.csv("vitalstats_ch1_tbl20.csv")

congress <- congressRaw %>%
  mutate(Years = as.character(Years)) %>%
  extractMatches("([0-9]*).", Years, "startYear"=1) %>%
  extractMatches("([0-9]*)$", Years, "endYear"=1) %>%
  mutate(startYear = as.numeric(as.character(startYear))) %>%
  mutate(endYear = as.numeric(as.character(endYear))) %>%
  select(startYear, endYear, Chamber, PartyStatus, Seats)

trial = congress
for (i in seq(1855,2017)) {
  tempDF <- trial %>%
    filter(startYear == i)
  if (nrow(tempDF) == 0) {
    newRow <- trial %>%
      filter(startYear == i-1) %>%
      mutate(startYear = i)
    trial <- rbind(trial, newRow)
    trial
  }
}
```

```
trial <- trial %>%
  arrange(startYear) %>%
  mutate(Year = startYear) %>%
  mutate(Seats = as.numeric(as.character(Seats))) %>%
  select(Year, Chamber, PartyStatus, Seats)

CongressParty <- trial %>%
  filter(Seats > 0) %>%
  filter(PartyStatus != "Other") %>%
  filter(PartyStatus != "Vacant")
CongressParty

#saving file as a .csv file for further use
write.csv(CongressParty, "CongressParty.csv", row.names=FALSE)

DominantCongressParty <- CongressParty %>%
  spread(PartyStatus, Seats) %>%
  mutate(DominantParty = ifelse(Democrat > Republican, "Democrat",
"Republican")) %>%
  select(Year, Chamber, DominantParty) %>%
  spread(Chamber, DominantParty)
DominantCongressParty

#saving file as a .csv file for further use
write.csv(DominantCongressParty, "DominantCongressParty.csv",
row.names=FALSE)
```

D. Presidential Party Data

```
library(rvest)

# Web scrape tables for political party by year
presidentRawNodes <-
read_html("http://www.enchantedlearning.com/history/us/pres/list.shtml
") %>%
  html_nodes(xpath = "//table") %>%
  html_table(fill=TRUE)
presidentRawNode <- presidentRawNodes[[9]]

# Turning raw table into tidy table
presidents <- presidentRawNode %>%
  extractMatches("([0-9]{4})\\-([[0-9]]*)", `Term as President`,
"startYear"=1, "endYear"=2) %>%
  mutate(startYear = as.numeric(as.character(startYear))) %>%
  mutate(endYear = as.numeric(as.character(endYear))) %>%
  select(Party, startYear, endYear) %>%
  filter(startYear >= 1841)

trial = presidents
for (i in seq(1841,2017)) {
  tempDF <- trial %>%
    filter(startYear == i)
```

```
  if (nrow(tempDF) == 0) {
    newRow <- trial %>%
      filter(startYear == i-1) %>%
      mutate(startYear = i)
    trial <- rbind(trial, newRow)
    trial
  }
}

Presidents <- trial %>%
  arrange(startYear) %>%
  mutate(Year = startYear) %>%
  select(Year, Party)
Presidents

#saving file as a .csv file for further use
write.csv(Presidents, "Presidents.csv")
```

II. Data Wrangling

```
## Input other data into R
gdp_clean <-
read.csv("C:\\Users\\QiMing\\Downloads\\Diversity-master\\Diversity-maste
r\\cleangdp.csv")
president_clean <-
read.csv("C:\\Users\\QiMing\\Downloads\\Diversity-master\\Diversity-maste
r\\Presidents\\Presidents.csv")
senate_houses_clean <-
read.csv("C:\\Users\\QiMing\\Downloads\\Diversity-master\\Diversity-maste
r\\CongressParty\\DominantCongressParty.csv")
visa <-
read.csv("C:\\Users\\QiMing\\Downloads\\Diversity-master\\Diversity-maste
r\\visa_all.csv")
total_visa <-
read.csv("C:\\Users\\QiMing\\Downloads\\Diversity-master\\Diversity-maste
r\\Visas\\Visas.Year.csv")
visa_clean <- visa %>%
  group_by(Year) %>%
  summarise(total_J1 = sum(J.1, na.rm = TRUE),
            total_M1 = sum(M.1, na.rm = TRUE),
            total_F1 = sum(F.1, na.rm = TRUE)) %>%
  mutate(Visa_for_Intl_Students = (total_J1 + total_M1 + total_F1)) %>%
  select(Visa_for_Intl_Students,Year
## Join up other variables into data_clean
data_clean <- data_clean %>%
  left_join(Government_funding_clean, by = c(academicyear = "Year")) %>%
  left_join(gdp_clean, by = c(academicyear = "year")) %>%
  left_join(president_clean, by = c(academicyear = "Year")) %>%
  left_join(visa_clean, by = c(academicyear = "Year")) %>%
```

```
  left_join(total_visa, by = c(academicyear = "Year"))
names(data_clean)[names(data_totclean) == "gdp.billion."] <-
"GDP.billion"
names(data_clean)[names(data_clean) == "academicyear"] <- "Academic_Year"
names(data_clean)[names(data_clean) == "Grand.Total"] <- "Total_Visa"
data_clean <- data_clean %>%
  select(-X.x, -X.y)
data_clean <- data_clean %>%
  mutate(Operating_Income.billion = (Operating_Income / 1000000000),
         Income_per_GDP = (Operating_Income.billion / GDP.billion * 100),
         Prop_Intl_Student = (Total_Intl_Enrollment / Total_Enrollment *
100),
         Government_funding_per_GDP = (Government_funding / (GDP.billion
* 1000000000) * 100),
         Group_Prop_Intl_Student = ifelse(Prop_Intl_Student > 3, 1, 0))

data_clean <- data_clean %>%
  left_join(senate_houses_clean, by = c(Academic_Year = "Year"))
write.csv(data_clean, file =
"C:\\Users\\QiMing\\Desktop\\Ming\\data_clean.csv")
```

III. Data Visualization

A. International Students and College Revenue vs. Government Institution Graphs

```
data_clean <- read.csv(file="C:\\Users\\jolina\\stat
133\\Diversity\\Diversity-master\\Diversity-master\\Ming\\data_clean.csv
", header=TRUE, sep=",")
visa <- read.csv(file="C:\\Users\\jolina\\stat
133\\Diversity\\Diversity-master\\Diversity-master\\Visas\\visa_party_en
rollment.csv", header=TRUE, sep=",")
visa <- visa %>%
  mutate(IS=ifelse(grepl("Iran|Iraq|Syria|Sudan|Libya|Yemen|Somalia",
Country), "Yes", "No")) %>%
  transform(F.1 = as.numeric(F.1),
            J.1 = as.numeric(J.1),
            M.1 = as.numeric(M.1),
            Total.Visas = as.numeric(Total.Visas))
# plots of college revenue/prop of international students and
# president/house/senate.
college_revenue_Pres <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Income_per_GDP, color = Party)) +
  geom_line(aes(color=Party, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=1.5, label= "Recession") +
  labs(title = "College Revenue by Presidential Party",
       x = "Academic Year",
       y = "College Revenue (% of GDP)") +
```

```
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
college_revenue_Pres

college_revenue_House <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Income_per_GDP, color = House)) +
  geom_line(aes(color=House, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=1.5, label= "Recession") +
  labs(title = "College Revenue by Dominant Party in House",
       x = "Academic Year",
       y = "College Revenue (% of GDP)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
college_revenue_House

college_revenue_Senate <- data_clean %>%
```

```
  ggplot(aes(x = Academic_Year, y = Income_per_GDP, color = Senate)) +
  geom_line(aes(color=Senate, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=1.5, label= "Recession") +
  labs(title = "College Revenue by Dominant Party in Senate",
       x = "Academic Year",
       y = "College Revenue (% of GDP)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
college_revenue_Senate

int_student_Pres <- data_clean %>%
  filter(Academic_Year != 1988, Academic_Year != 1990) %>% ## this line
is to exclude 1988 and 1990
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = Party)) +
  geom_line(aes(color=Party, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=3.6, label= "Recession") +
  labs(title = "International Student by Presidential Party",
       x = "Academic Year",
       y = "International Student (% of total enrollment)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 16, face = "bold"),
        axis.title = element_text(size = 13)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
```

```
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
int_student_Pres

int_student_House <- data_clean %>%
  filter(Academic_Year != 1988, Academic_Year != 1990) %>% ## this line
is to exclude 1988 and 1990
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = House)) +
  geom_line(aes(color=House, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=3.6, label= "Recession") +
  labs(title = "International Student by Dominant Party in House",
       x = "Academic Year",
       y = "International Student (% of total enrollment)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 15, face = "bold"),
        axis.title = element_text(size = 13)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
int_student_House

int_student_Senate <- data_clean %>%
  filter(Academic_Year != 1988, Academic_Year != 1990) %>% ## this line
is to exclude 1988 and 1990
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = Senate))
+
  geom_line(aes(color=Senate, group=1), size = 1.5) +
  annotate("text", x=1990.5, y=3.6, label= "Recession") +
  labs(title = "International Student by Dominant Party in Senate",
       x = "Academic Year",
       y = "International Student (% of total enrollment)") +
```

```
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(legend.position = c(0.9, 0.2),
        legend.justification = c(0.9, 0.2),
        panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 15, face = "bold"),
        axis.title = element_text(size = 13)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  scale_x_continuous(breaks = seq(min(data_clean$Academic_Year),
max(data_clean$Academic_Year), by = 3), "Academic Year")
int_student_Senate
```

B. College Revenue vs. Proportion of International Students Graphs

```
## Load the required library
library(DataComputing)
library(tidyr)
library(rpart)
library(statisticalModeling)
## install.packages("rpart.plot")
library(rpart.plot)
## install.packages("xlsx")
library(xlsx)

## ggplot
p1 <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Income_per_GDP)) +
  geom_line(aes(color = Party), size = 1.5) +
  labs(title = "Income per GDP from 1987 - 2013",
       x = "Academic Year",
       y = "Income per GDP") +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic")) +
```

```
        grid.arrange(arrangeGrob(ncol=1, nrow=2))
p2 <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Income_per_GDP)) +
  geom_line(size = 1.5, colour = "Blue") +
  facet_grid(. ~ Party) +
  labs(title = "Income per GDP from 1987 - 2013",
       x = "Academic Year",
       y = "Income per GDP") +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
p3 <- data_clean %>%
  ggplot(aes(x = Prop_Intl_Student, y = Income_per_GDP)) +
  geom_point() +
  geom_smooth( method = "loess", se = FALSE, col = "red") +
  scale_x_continuous(limits = c(2.5,3.75)) +
  labs(
    title = "Income per GDP (%) vs Proportion of International Student
(%)",
    x = "Proportion of International Student (%)",
    y = "Income per GDP (%)"
  )
p3.1 <- data_clean %>%
  ggplot(aes(x = Prop_Intl_Student, y = Income_per_GDP, group =
Group_Prop_Intl_Student)) +
  geom_point(aes(colour = Group_Prop_Intl_Student)) +
  geom_smooth(method = "lm",  se = FALSE, col = "red") +
  scale_x_continuous(limits = c(2.5,3.75))

p4 <- data_clean %>%
  ggplot(aes(x = Government_funding_per_GDP, y = Income_per_GDP)) +
  geom_point() +
  geom_smooth( method = "loess", se = FALSE, col="red")
correction <- c(Democrat = "Senate.Democrat", Republican =
"Senate.Republican")

p9 <- data_clean %>%
  ggplot(aes(x = Party, y = Prop_Intl_Student)) +
  geom_boxplot(color = "Brown") +
  facet_grid(. ~ Senate, labeller = labeller(Senate = correction)) +
  labs(title = "Proportion of International Student",
       x = "Presidential Party",
```

```
       y = "Proportion of International Student") +
  scale_x_discrete(labels = c("Presidential.Democrat",
"Presidential.Republican")) +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
        axis.text.x = element_text(size = 10))
p9.1 <- data_clean %>%
  ggplot(aes(x = House, y = Prop_Intl_Student)) +
  geom_boxplot(color = "Brown") +
  labs(title = "Proportion of International Student",
       x = "House Party",
       y = "Proportion of International Student") +
  scale_x_discrete(labels = c("House.Democrat", "House.Republican")) +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
        axis.text.x = element_text(size = 10))
p9.2 <- data_clean %>%
  ggplot(aes(x = Senate, y = Prop_Intl_Student)) +
  geom_boxplot(color = "Brown") +
  labs(title = "Proportion of International Student",
       x = "Senate Party",
       y = "Proportion of International Student") +
  scale_x_discrete(labels = c("Senate.Democrat", "Senate.Republican")) +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
        axis.text.x = element_text(size = 10))
p9.3 <- data_clean %>%
  ggplot(aes(x = Party, y = Prop_Intl_Student)) +
  geom_boxplot(color = "Brown") +
  labs(title = "Proportion of International Student",
       x = "Presidential Party",
       y = "Proportion of International Student") +
  scale_x_discrete(labels = c("Presidential.Democrat",
"Presidential.Republican")) +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
```

```
        axis.text.x = element_text(size = 10))
p10 <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = Party,
group = 1)) +
  geom_line(size = 1.5) +
  scale_y_continuous(limits = c(1,4)) +
  labs(title = "Proportion of International Student",
       x = "Academic_Year",
       y = "Proportion of International Student") +
  geom_smooth( method = "loess", se = FALSE, col="black") +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
p10.1 <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = Senate,
group = 1)) +
  geom_line(size = 1.5) +
  scale_y_continuous(limits = c(1,4)) +
  labs(title = "Proportion of International Student",
       x = "Academic_Year",
       y = "Proportion of International Student") +
  geom_smooth( method = "loess", se = FALSE, col="black") +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
p10.2 <- data_clean %>%
  ggplot(aes(x = Academic_Year, y = Prop_Intl_Student, color = House,
group = 1)) +
  geom_line(size = 1.5) +
  scale_y_continuous(limits = c(1,4)) +
  labs(title = "Proportion of International Student",
       x = "Academic_Year",
       y = "Proportion of International Student") +
  geom_smooth( method = "loess", se = FALSE, col="black") +
  theme(plot.title = element_text(size = 20, face = "bold", hjust =
0.5),
```

```
        axis.title = element_text(size = 15, face = "italic"),
        strip.text = element_text(size = 15)) +
  annotate("rect", xmin=1990, xmax=1991, ymin=-Inf, ymax=Inf, alpha=0.3)
+
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)
```

C. Visa Graphs

```
visa <-
read.csv(file="/Users/seohyeongjeong/Desktop/visa/visa_party_enrollment.
csv", header=TRUE, sep=",")

## travel ban order countries
visa <- visa %>%
  mutate(IS=ifelse(grepl("Iran|Iraq|Syria|Sudan|Libya|Yemen|Somalia",
Country), "Yes", "No")) %>%
  transform(F.1 = as.numeric(F.1),
            J.1 = as.numeric(J.1),
            M.1 = as.numeric(M.1),
            Total.Visas = as.numeric(Total.Visas))

visa %>%
  filter(IS=="Yes")%>%
  group_by(Country)%>%
  summarise(prop=(F.1+J.1+M.1)/Total_Enrollment)

## Figure 3.7
Avg_Int_Stu_Visa <- visa %>%
  group_by(Year, president, dominant_house, dominant_senate) %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  summarise(avgProp=mean(prop, na.rm=TRUE)) %>%
  arrange(Year) %>%
  ggplot(aes(x=Year, y=avgProp, color=president)) +
  geom_line(aes(color=president, group=1), size = 1.5) +
  labs(title = "International Student Visa",
       x = "Academic Year",
       y = "Average F1, J1, M1 Visa Admission
       (% of total enrollment)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
```

```
        axis.title = element_text(size = 20, face = "italic"),
        legend.text = element_text(size = 15),
        legend.title = element_text(size = 20)) +
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)

## Figure 3.8
is3 <- visa %>%
  group_by(IS, Year, president, dominant_house, dominant_senate) %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  summarise(avgProp=mean(prop, na.rm=TRUE)) %>%
  arrange(Year, IS)

names(is3)[names(is3) == "IS"] <- "Seven Countries"

IS3 <- is3 %>%
  ggplot(aes(x=Year, y=avgProp, color=`Seven Countries`)) +
  geom_line(size = 1.5) +
  labs(title = "International Student Visa",
       x = "Academic Year",
       y = "Average F1, J1, M1 Visa Admission
       (% of total enrollment)") +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 20, face = "italic"),
        legend.text = element_text(size = 15),
        legend.title = element_text(size = 20)) +
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)

## Figure 3.8
IS1 <- visa %>% filter(IS=="Yes") %>%
  group_by(Year, president, dominant_house, dominant_senate) %>%
  summarise(Avg=mean(Total.Visas), na.rm=TRUE) %>%
  arrange(Year) %>%
  ggplot(aes(x=Year, y=Avg, color=president, group=1)) +
  geom_line(size=1.5) +
  labs(title = "International Student Visa Admission
       in Seven Countries",
       x = "Academic Year",
       y = "Average F1, J1, M1 Visa Admission") +
```

```
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 20, face = "italic"),
        legend.text = element_text(size = 15),
        legend.title = element_text(size = 20)) +
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)

## Figure 3.9
point <- visa %>%
  filter(IS=="Yes") %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  ggplot(aes(x=Year, y=prop)) +
  geom_point(aes(color=president)) +
  geom_smooth(method="loess", se = FALSE, col="red") +
  labs(title = "International Student Visa in Seven Countries",
       x = "Academic Year",
       y = "Average F1, J1, M1 Visa Admission
       (% of total enrollment)") +
  scale_color_manual(values=c(Democrat="#3366FF", Republican="#FF6633"))
+
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_line(size = 0.5, color = "Grey"),
        panel.grid.major.x = element_blank(),
        panel.background = element_blank(),
        plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 20, face = "italic"),
        axis.text = element_text(size = 20),
        legend.text = element_text(size = 15),
        legend.title = element_text(size = 20)) +
  annotate("rect", xmin=2000.5, xmax=2001.5, ymin=-Inf, ymax=Inf,
alpha=0.3) +
  annotate("rect", xmin=2008, xmax=2010, ymin=-Inf, ymax=Inf, alpha=0.3)

## Figure 3.10
IS222 <- visa %>% filter(IS=="Yes") %>%
  group_by(Year, president, dominant_house, dominant_senate) %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  summarise(avgProp=mean(prop, na.rm=TRUE)) %>%
  arrange(president) %>%
```

```
  ggplot(aes(x = president, y = avgProp)) +
  geom_boxplot() +
  facet_grid(. ~ dominant_house, labeller = labeller(dominant_house =
c(Democrat = "House.Democrat", Republican = "House.Republican"))) +
  labs(title = "International Student Visa in Seven Countries",
       x = "Presidential Party",
       y = "Average F1, J1, M1 Visa Admission
       (% of total enrollment)") +
  scale_x_discrete(labels = c("Democrat", "Republican")) +
  theme(plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 20, face = "italic"),
        strip.text = element_text(size = 20, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
        axis.text.x = element_text(size = 15))

##Figure 3.11
IS22 <- visa %>% filter(IS=="Yes") %>%
  group_by(Year, president, dominant_house, dominant_senate) %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  summarise(avgProp=mean(prop, na.rm=TRUE)) %>%
  arrange(president) %>%
  ggplot(aes(x = president, y = avgProp)) +
  geom_boxplot() +
  facet_grid(. ~ dominant_senate, labeller = labeller(dominant_senate =
correction)) +
  labs(title = "International Student Visa in Seven Countries",
       x = "Presidential Party",
       y = "Average F1, J1, M1 Visa Admission
       (% of total enrollment)") +
  scale_x_discrete(labels = c("Democrat", "Republican")) +
  theme(plot.title = element_text(size = 30, face = "bold", hjust =
0.5),
        axis.title = element_text(size = 20, face = "italic"),
        strip.text = element_text(size = 20, face = "italic"),
        strip.background = element_rect(fill = "Yellow"),
        axis.text.x = element_text(size = 15))
```

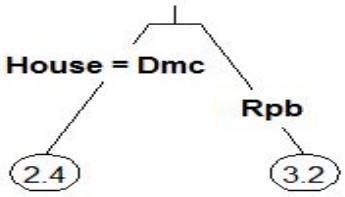
IV. Regression Analysis

A. Proportion of International Students and College Revenue Analysis

```
Model_1 <- rpart(Prop_Intl_Student ~ Party + House + Senate, data_clean)
%>% rpart.plot::prp(type=3)
summary(linear_model)
Model_1 %>%
```

```
  evaluate_model(at=list(Party = "Republican", House = "Republican",
Senate = "Republican"))
## Predicted Output = 3.234834

predict(linear_model, list(Party = "Republican", House = "Republican",
Senate = "Republican"))
linear_model %>%
  evaluate_model(at=list(Party = "Democrat", House = "Democrat", Senate
= "Democrat"))
## Predicted output = 2.421612
rpart(Prop_Intl_Student ~ House, data_clean) %>% fmodel()
## Income_per_GDP ~ Prop_Intl_Student
linear_model_1 <- lm(Income_per_GDP ~ Prop_Intl_Student, data_clean)
print(linear_model_1)
summary(linear_model_1)
## Result: Income_per_GDP = 0.2202(Prop_Intl_Student) + 0.4127
p6 <- fmodel(linear_model_1)
linear_model_1 %>%
  evaluate_model(at=list(Prop_Intl_Student = 3.234834))
linear_model_1 %>%
  evaluate_model(at=list(Prop_Intl_Student = 2.421612))
## Predicted Output = 1.125046 (Republican)
## Predicted Output = 0.9459594 (Democrat)
## Income_per_GDP ~ Prop_Intl_student + Government_funding_per_GDP
linear_model_2 <- lm(Income_per_GDP ~ Prop_Intl_Student +
Government_funding_per_GDP, data_clean)
print(linear_model_2)
summary(linear_model_2)
## Result: Income_per_GDP = 0.2098(Prop_Intl_Student) +
-1.3663(Government_funding_per_GDP)
p7 <- fmodel(linear_model_2)
linear_model_2 %>%
  evaluate_model(at=list(Prop_Intl_Student = 3.234834,
Government_funding_per_GDP = (23.91038/19284.99)*100))
## Budget for FY2017 - 23.91038billion on tertiary education, GDP for
FY2017 = 19287.99billion
## Predicted_output = 1.172496
linear_model_2 %>%
  evaluate_model(at=list(Prop_Intl_Student = 2.421612,
Government_funding_per_GDP = (23.91038/19284.99)*100))
## Predicted_output = 1.001874
## Recursive Partition
rpart(Income_per_GDP ~ Prop_Intl_Student + Government_funding_per_GDP,
data_clean) %>%
  rpart.plot::prp(type=3)
rpart(Operating_Income ~  Party + Visa_for_Intl_Students, data_clean)
%>%
  rpart.plot::prp(type=3)
```

```
model_1 <- lm(Income_per_GDP ~ poly(Prop_Intl_Student,3), data_clean)
model_1 %>%
  evaluate_model(at=list(Prop_Intl_Student = 3.234834))
fmodel(model_1)
summary(model_1)
## Predicted Output = 1.198933
model_2 <- lm(Income_per_GDP ~ poly(Prop_Intl_Student,3) +
Government_funding_per_GDP, data_clean)
model_2 %>%
  evaluate_model(at=list(Prop_Intl_Student = 3.234834,
Government_funding_per_GDP = (23.91038/19284.99)*100))
## Predicted Output = 1.186774
summary(model_2)
model_3 <- lm(Income_per_GDP ~ poly(Prop_Intl_Student,2), data =
data_clean)
model_3 %>%
  evaluate_model(at=list(Prop_Intl_Student = 3.234834))
plot(fitted(model_3),residuals(model_3))
```




```
Residuals:
     Min       1Q   Median       3Q      Max
-0.23736 -0.10453 -0.00899  0.08166  0.37077

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     0.9841     0.1262   7.797 9.01e-08 ***
poly(Prop_Intl_Student, 3)1     1.0129     0.1625   6.234 2.83e-06 ***
poly(Prop_Intl_Student, 3)2     1.1227     0.1712   6.556 1.36e-06 ***
poly(Prop_Intl_Student, 3)3    -0.1852     0.1604  -1.155    0.260
Government_funding_per_GDP      0.3799     0.7579   0.501    0.621
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 0.1603 on 22 degrees of freedom
Multiple R-squared:  0.7979,    Adjusted R-squared:  0.7612
F-statistic: 21.72 on 4 and 22 DF,  p-value: 2.241e-07
```

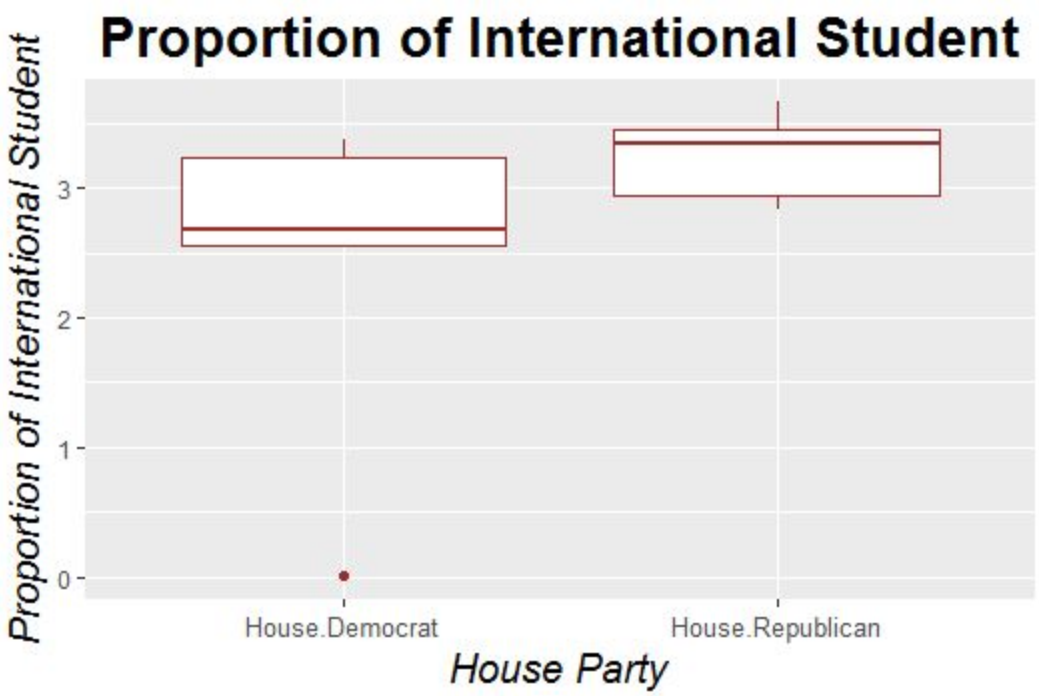
B. Visa Analysis

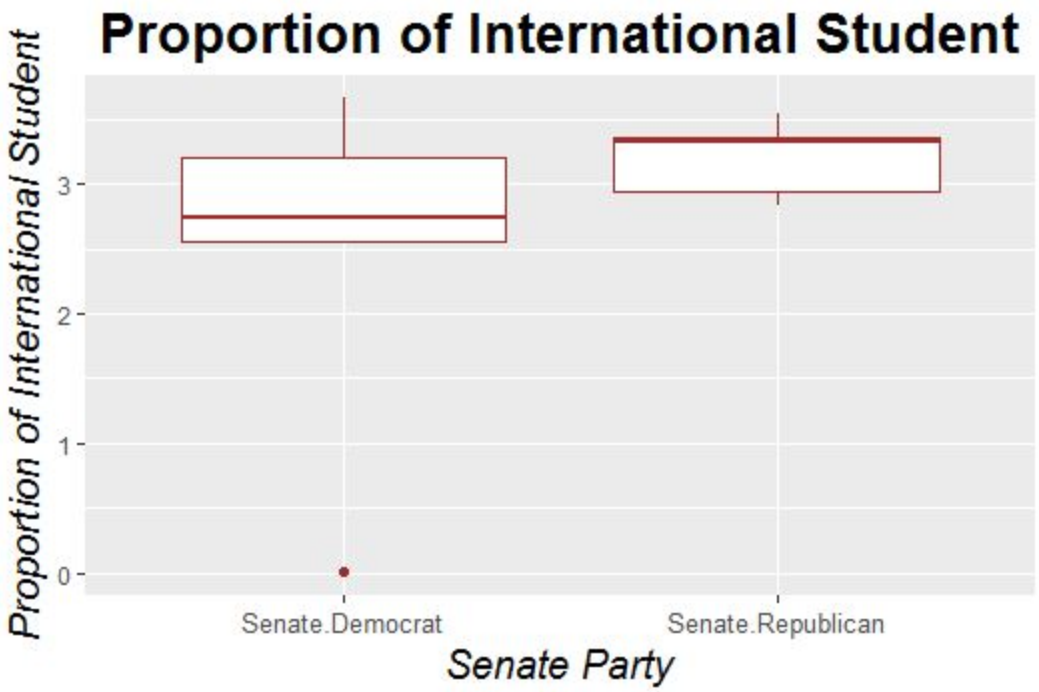
```
visa_prop <- visa %>% filter(IS=="Yes") %>%
  group_by(Year, president, dominant_house, dominant_senate) %>%
  mutate(prop=((F.1+J.1+M.1)*100/Total_Enrollment)) %>%
  summarise(avgProp=mean(prop, na.rm=TRUE))

## regression prop ~ president + Year
library(rpart)
library(statisticalModeling)
linear_model <- lm(avgProp ~ president + Year, data=visa_prop)
print(linear_model)
summary(linear_model)

## result: president=republican 0.01691987
```

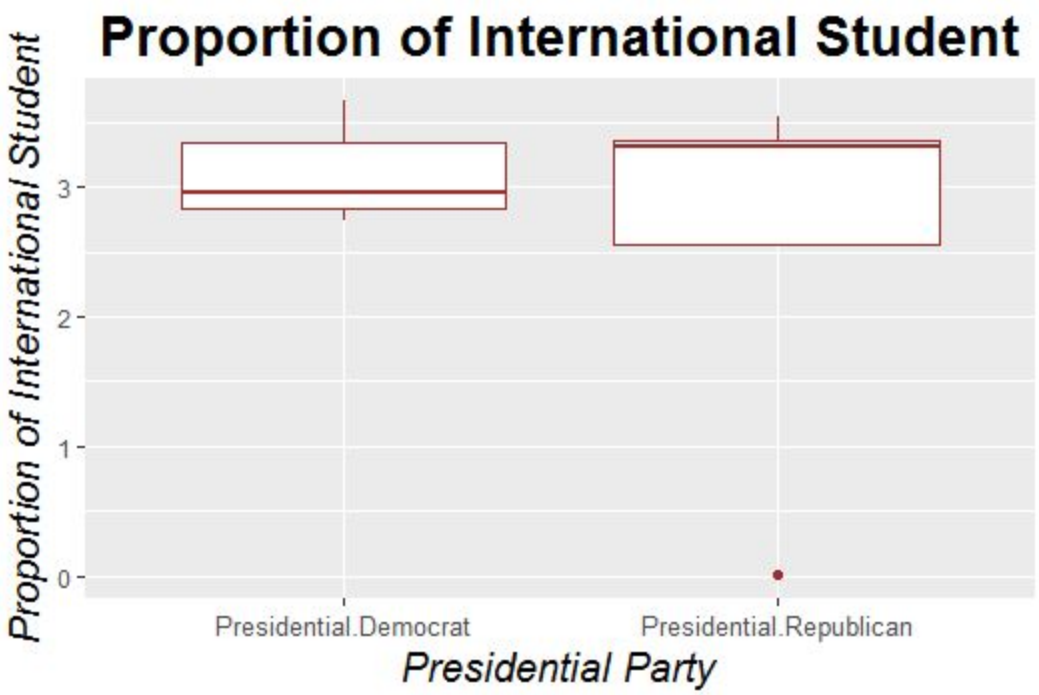
```
## result: president=democrat 0.01882654
linear_model %>%
  evaluate_model(at=list(president="Republican"))
linear_model %>%
  evaluate_model(at=list(president="Democrat"))
```

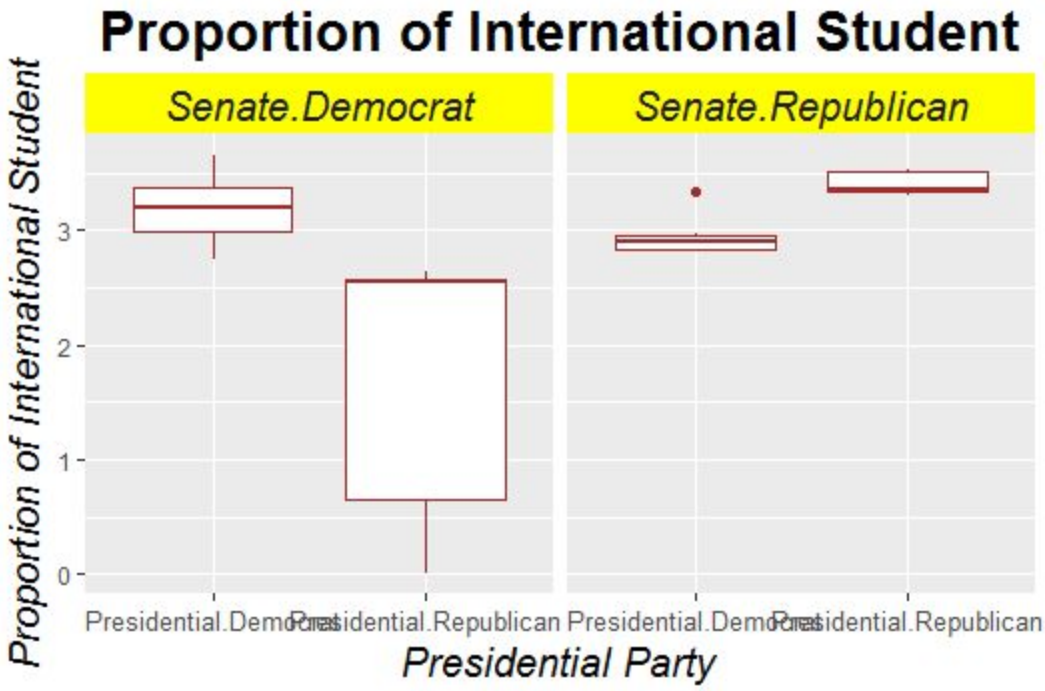
```
Residuals:
      Min        1Q    Median        3Q       Max
-0.003131 -0.002053  0.000246  0.001245  0.003974

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.5870277  0.2300347   6.899 7.33e-06 ***
presidentRepublican -0.0026452  0.0011258  -2.350    0.034 *
Year                -0.0007820  0.0001147  -6.817 8.36e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.002306 on 14 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.7791,    Adjusted R-squared:  0.7475
F-statistic: 24.69 on 2 and 14 DF,  p-value: 2.569e-05
```

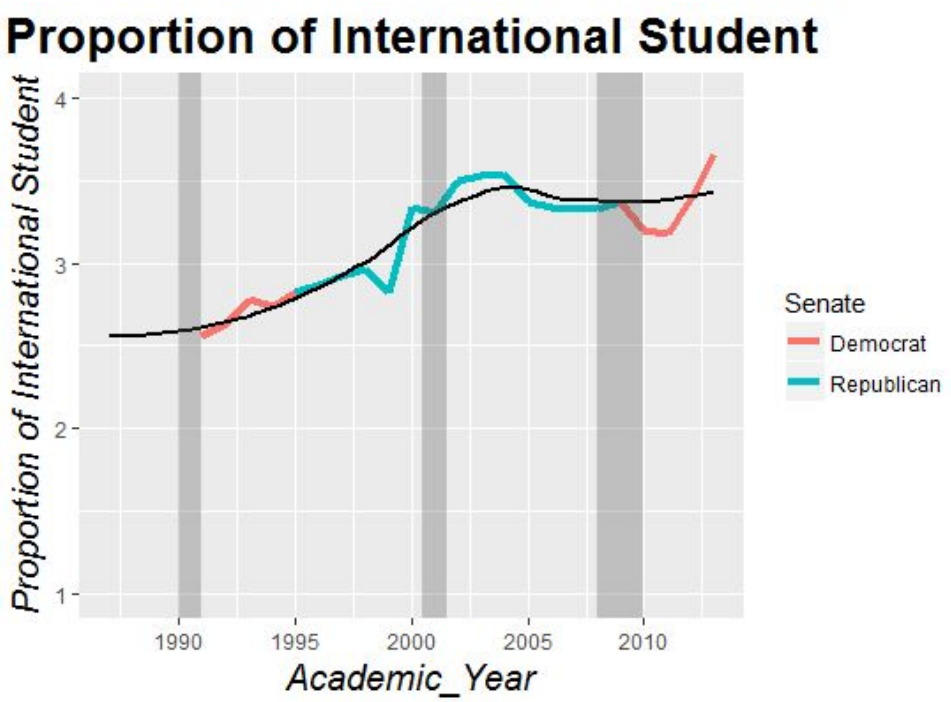
Additional Graphs

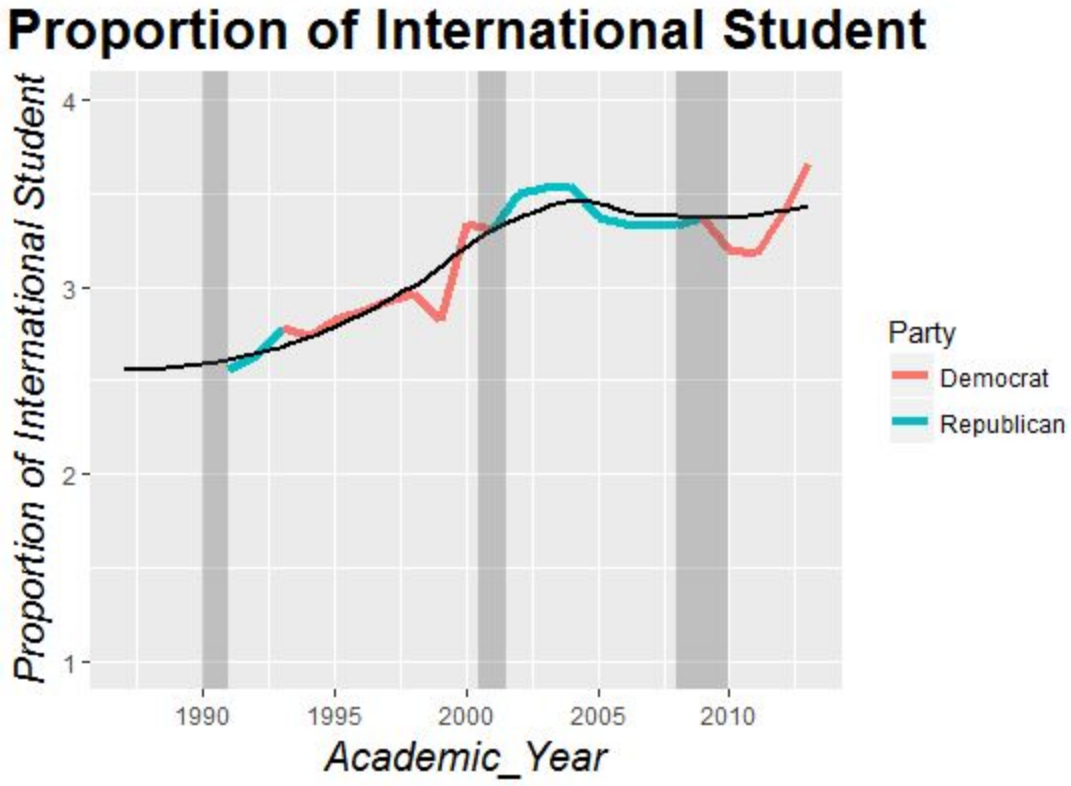


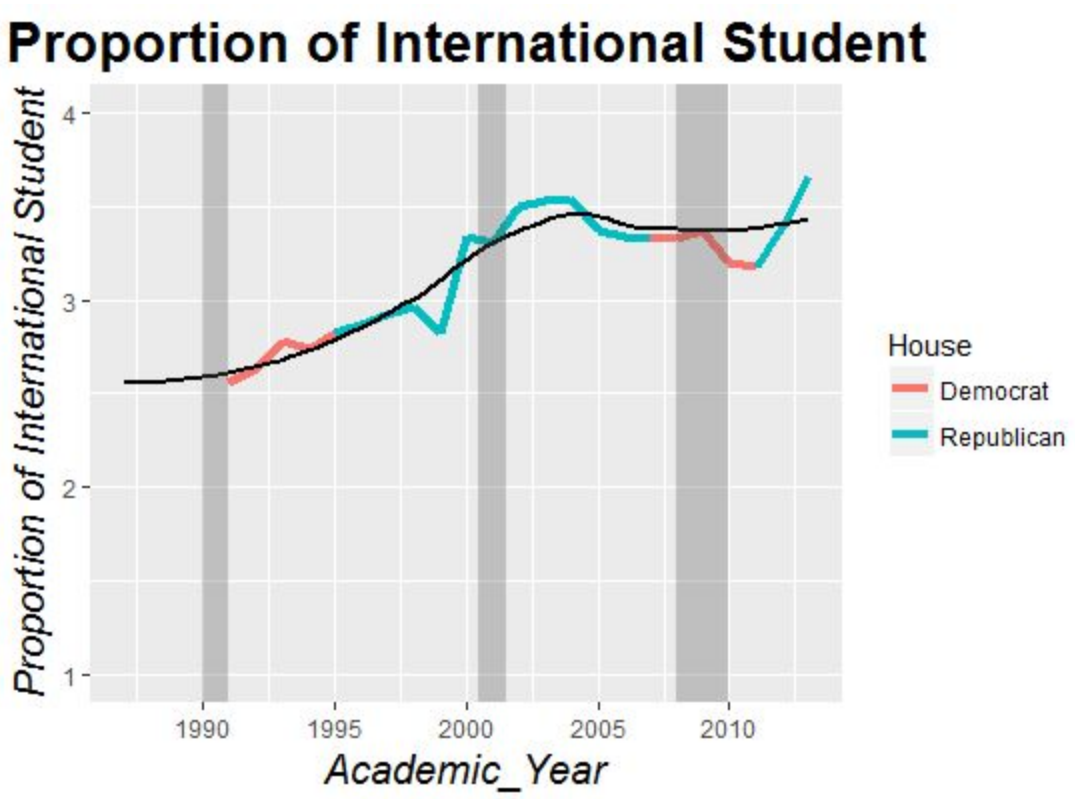



Proportion of International Student
Proportion of International Student
3
2
1
0
Presidential.Democrat
Presidential.Republican
Presidential Party


Proportion of International Student
Senate.Democrat
Senate.Republican
Proportion of International Student
3
2
1
0
Presidential.Democrat
Presidential.Republican
Presidential.Democrat
Presidential.Republican
Presidential Party


Proportion of International Student
Proportion of International Student
4
3
2
1
1990
1995
2000
2005
2010
Academic_Year
Senate
Democrat
Republican


Proportion of International Student
Proportion of International Student
4
3
2
1
1990
1995
2000
2005
2010
Academic_Year
Party
Democrat
Republican


Proportion of International Student
Proportion of International Student
4
3
2
1
1990
1995
2000
2005
2010
Academic_Year
House
Democrat
Republican