

## **International students enrolled in the USA**

### **Introduction**

The purpose of this study and the parameter of interest is to examine how the number of international students in the United States has changed between 1963 and 2022. The United States is a popular destination for international students because American institutions offer excellent academic and professional opportunities. In addition to benefiting students, colleges and universities also benefit from international students because they help create diverse and cross-cultural experiences in classrooms. As someone who has been an international student since the age of 12, I have seen an increase in the number of international students and the diversity of their nationalities. Many of my closest friends are international students, and even at my current university, Adelphi University, most of the people I interact with are international students. Every semester I notice an increase of international students, which is why I chose this research topic in order to see if the statistical data supports my opinion.

### **Data collection**

The data that I analyzed comes from the Open Doors Report (<https://opendoorsdata.org/data/international-students/enrollment-trends/>), which provides information on international student enrollment trends in the United States. The purpose of my study is to examine the changes in the number of international students in the U.S from 1963 to 2022. The data that I got from the source originally began in 1948, however some of the values were missing, which is why I decided to cut that information out and begin with 1963 in order to make the data and analysis more concrete. The variables in my study are the year and the number

of international students. The population is all international students who have studied in the United States during this time period. I obtained the data by accessing the Open Doors Report and extracting the relevant information for each year. The data that I collected is important for my study because it allows me to identify trends in the enrollment of international students in the U.S. By analyzing this data, I can determine whether the number of international students has increased or decreased over time and identify factors that may have contributed to these changes.

### **Exploratory Data Analysis**

The five quantitative variables in the dataset are:

- Year: The year in which the data was recorded.
- TotalIntStudents: The total number of international students studying in the US in that year.
- AnnualPercentChange: The percentage change in the number of international students from the previous year.
- TotalUSEnrollment: The total number of students enrolled in US institutions in that year.
- PercentInt: The percentage of international students out of the total number of students enrolled in US institutions in that year.

To analyze these variables, we can look at their distributions and relationships with each other.

- Year: The variable "Year" is a categorical variable that represents the year in which the data was recorded. The values for this variable range from 1963/64 to 2013/14.

- **TotalIntStudents:** The variable "TotalIntStudents" is a continuous variable that represents the total number of international students studying in the US in that year. The summary statistics show that the total number of international students in the U.S. ranges from 74,814 to 1,095,299 with a mean of 475,341.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
74814	249724	438618	475341	605064	1095299

- **AnnualPercentChange:** The variable "AnnualPercentChange" is a continuous variable that represents the percentage change in the number of international students from the previous year. The annual percentage change in international student enrollment varies between -15% and 21.2%, with a mean of 4.796%.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-15.000	1.200	4.300	4.796	7.450	21.200

- **TotalUSEnrollment:** The variable "TotalUSEnrollment" is a continuous variable that represents the total number of students enrolled in US institutions in that year. The total U.S. enrollment ranges from 4,780,000 to 21,253,000 with a mean of 14,167,695.

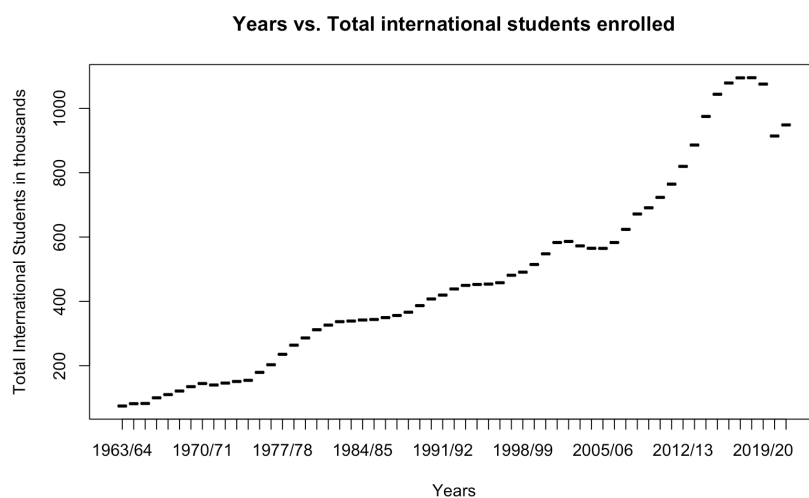
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4780000	11273000	14279000	14167695	17960000	21253000

- **PercentInt:** The variable "PercentInt" is a continuous variable that represents the percentage of international students out of the total number of students enrolled in US

institutions in that year. The percentage of international students in the total enrollment ranges from 1.4% to 5.5% with a mean of 3.039%.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.400	2.200	3.000	3.039	3.500	5.500

To explore the relationships between these variables, we can create a scatter plot. For example, I

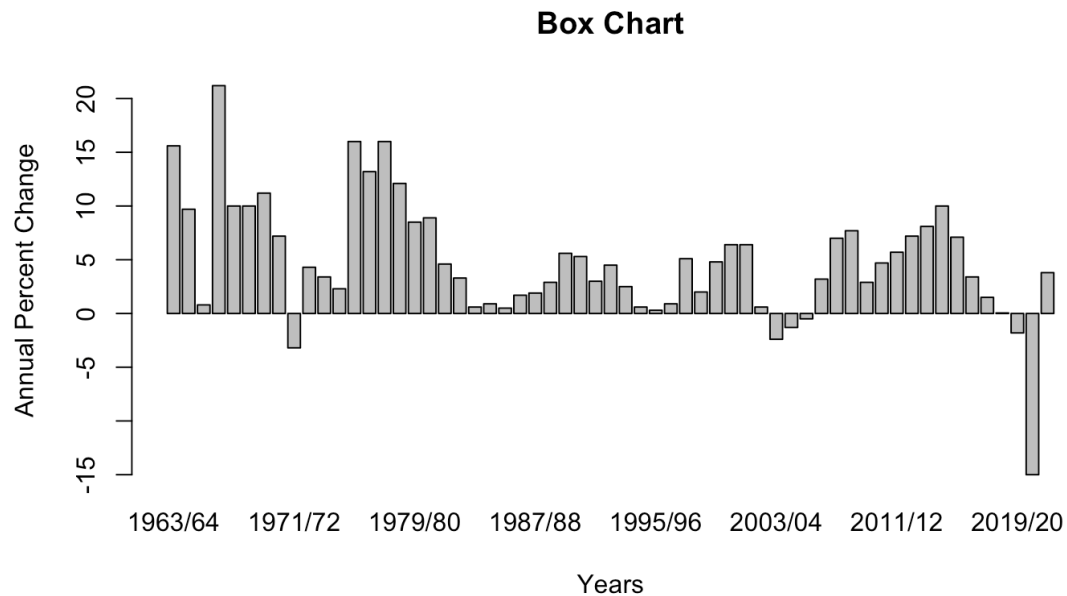


am going to examine how over the years the amount of total international students increased and show it on a scatterplot. There is a clear trend of growth in the number of international students over time, with

some fluctuations from year to year. In 1963/64, the total number of international students was 74,814, and this number has consistently increased, reaching a peak of 1,095,299 in 2018/19. However, in 2020/21, there was a significant decrease in the number of international students, likely due to the impact of the COVID-19 pandemic, with the number falling to 914,095.

Overall, the data suggests that there is a significant presence of international students in the United States, accounting for over 3% of total enrollment beginning the years 1992/93. The mean annual percentage change in international student enrollment is positive, indicating a steady increase in recent years. This trend is consistent with the United States' reputation as a

leading destination for international students. However, the data does not provide information on the countries of origin or fields of study of international students, which may be relevant factors in understanding trends in international education.

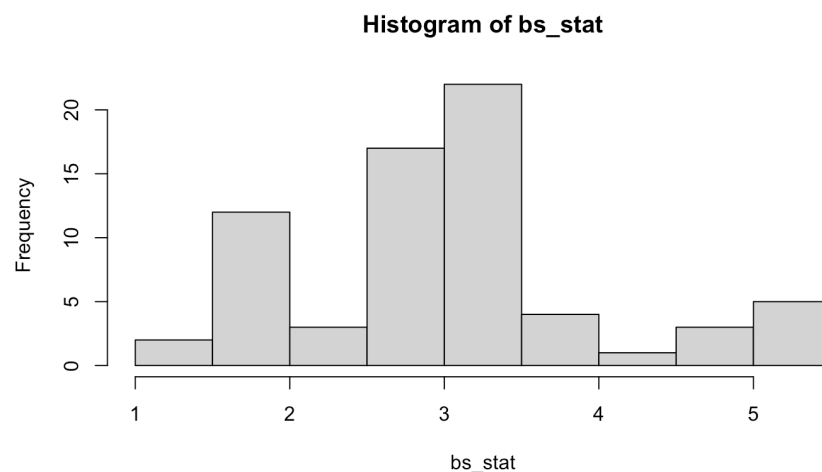


The data shows the relationship between the Total International Students and the Annual Percent Change in international student enrollment over the years. The data spans from 1963/64 to 2021/22, and there are 59 entries in total. The Annual Percent Change varies each year, with some years showing a positive increase and some a negative decrease. There is a general trend of an increase in international student enrollment, with the highest Annual Percent Change being 21.2% in 1966/67, and the lowest being -3.2% in 1971/72. This suggests that there is a significant demand for higher education in the US among international students, and that this trend has continued over the past few decades.

### Confidence Interval

A confidence interval is a statistical range that represents the range of plausible values for a population parameter based on a sample of data. It is typically expressed as a range of values that includes the population parameter with a certain degree of confidence, such as 95% or 99%. In other words, a 95% confidence interval means that if we repeated the process of taking samples from the population many times, we would expect the true population parameter to lie within the computed interval about 95% of the time.

In this case, we are using a 95% confidence interval to estimate the range of values for the mean of the PercentInt variable which is the percentage of international students enrolled in our sample. To do so, we will use a bootstrapping distribution, which involves repeatedly resampling the data with replacement to create a distribution of sample means. The number of bootstrap samples we will use is 10 since percentage of international students does not go above 10. We will then use this distribution to compute the standard error and construct a confidence interval for the population mean.



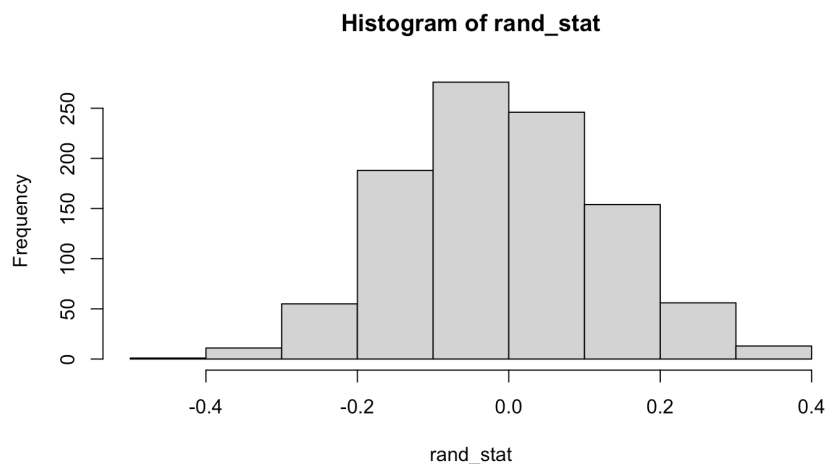
## Hypothesis test

To test the hypothesis that as the years increase, the amount of international students increases accordingly, we can perform a correlation analysis between the year and the total international student enrollment. We can also perform a hypothesis test using the randomization distribution generated by R code.

Null Hypothesis: There is no correlation between the year and the total international student enrollment.

Alternative Hypothesis: There is a positive correlation between the year and the total international student enrollment.

We will use a significance level of 0.05, which means we want to be 95% confident in our conclusion. The observed correlation coefficient is 0.9273. The p-value is the proportion of simulated correlation coefficients that are greater than or equal to the observed correlation coefficient. In this case, the p-value is less than 0.0001, which is much smaller than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a positive correlation between the year and total international student enrollment. Specifically, as the years increase, the amount of international students increases accordingly.



## **Conclusion**

Based on the analysis of the data, we can conclude that the number of international students studying in the United States has been increasing steadily over the years. The data shows a clear upward trend, with the highest number of international students being recorded in the 2018/2019 academic year.

Furthermore, based on the statistical analysis, we can be 95% confident that as the years go on and increase, the amount of international students tends to increase. This suggests that the trend of increasing international student enrollment is likely to continue in the future.



## Appendix 1

Data collected:

<u>Year</u>	<u>IntStudents</u>	<u>centChange</u>	<u>USEnrollment</u>	<u>entInt</u>
1963/64	74,814	15.6	4,780,000	1.6
1964/65	82,045	9.7	5,280,000	1.6
1965/66	82,709	0.8	5,921,000	1.4
1966/67	100,262	21.2	6,390,000	1.6
1967/68	110,315	10.0	6,912,000	1.6
1968/69	121,362	10.0	7,513,000	1.6
1969/70	134,959	11.2	8,005,000	1.7
1970/71	144,708	7.2	8,581,000	1.7
1971/72	140,126	-3.2	8,949,000	1.6
1972/73	146,097	4.3	9,215,000	1.6
1973/74	151,066	3.4	9,602,000	1.6
1974/75	154,580	2.3	10,224,000	1.5
1975/76	179,344	16.0	11,185,000	1.6
1976/77	203,068	13.2	11,012,000	1.8
1977/78	235,509	16.0	11,286,000	2.1
1978/79	263,938	12.1	11,260,000	2.3
1979/80	286,343	8.5	11,570,000	2.5
1980/81	311,882	8.9	12,097,000	2.6
1981/82	326,299	4.6	12,372,000	2.6
1982/83	336,985	3.3	12,426,000	2.7
1983/84	338,894	0.6	12,465,000	2.7
1984/85	342,113	0.9	12,242,000	2.8
1985/86	343,777	0.5	12,247,000	2.8
1986/87	349,609	1.7	12,504,000	2.8
1987/88	356,187	1.9	12,767,000	2.8
1988/89	366,354	2.9	13,055,000	2.8
1989/90	386,851	5.6	13,539,000	2.9
1990/91	407,529	5.3	13,819,000	2.9
1991/92	419,585	3.0	14,359,000	2.9
1992/93	438,618	4.5	14,487,000	3.0
1993/94	449,749	2.5	14,305,000	3.1
1994/95	452,635	0.6	14,279,000	3.2
1995/96	453,787	0.3	14,262,000	3.2
1996/97	457,984	0.9	14,368,000	3.2
1997/98	481,280	5.1	14,502,000	3.3
1998/99	490,933	2.0	14,507,000	3.4
1999/00	514,723	4.8	14,791,000	3.5
2000/01	547,867	6.4	15,312,000	3.6
2001/02	582,996	6.4	15,928,000	3.7
2002/03	586,323	0.6	16,612,000	3.5
2003/04	572,509	-2.4	16,911,000	3.4
2004/05	565,039	-1.3	17,272,000	3.3
2005/06	564,766	-0.5	17,487,000	3.2
2006/07	582,984	3.2	17,672,000	3.3
2007/08	623,805	7.0	18,248,000	3.4
2008/09	671,616	7.7	19,103,000	3.5
2009/10	690,923	2.9	20,428,000	3.4

2010/11	723,277	4.7	20,550,000	3.5
2011/12	764,495	5.7	20,625,000	3.7
2012/13	819,644	7.2	21,253,000	3.9
2013/14	886,052	8.1	21,216,000	4.2
2014/15	974,926	10.0	20,300,000	4.8
2015/16	1,043,839	7.1	20,264,000	5.2
2016/17	1,078,822	3.4	20,185,000	5.3
2017/18	1,094,792	1.5	19,831,000	5.5
2018/19	1,095,299	0.05	19,828,000	5.5
2019/20	1,075,496	-1.8	19,720,000	5.5
2020/21	914,095	-15.0	19,744,000	4.6
2021/22	948,519	3.8	20,327,000	4.7

## Appendix 2

### **R Code:**

```
# Importing data
```

```
library(readxl)
```

```
IntStudents <- read_excel("/Users/hyakkymaru/Downloads/Census_Overall-Enrollment-1949-2021_OD22.xlsx")
```

```
View(IntStudents)
```

### **Explanatory Data**

```
# Summary statistics for each variable
```

```
summary(IntStudents$TotalIntStudents)
```

```
summary(IntStudents$AnnualPercentChange)
```

```
summary(IntStudents$TotalUSEnrollment)
```

```
summary(IntStudents$PercentInt)
```

```
# Create a vector of years
```

```
years <- c("1963/64", "1964/65", "1965/66", "1966/67", "1967/68", "1968/69", "1969/70",  
"1970/71", "1971/72", "1972/73", "1973/74", "1974/75", "1975/76", "1976/77", "1977/78",  
"1978/79", "1979/80", "1980/81", "1981/82", "1982/83", "1983/84", "1984/85", "1985/86",  
"1986/87", "1987/88", "1988/89", "1989/90", "1990/91", "1991/92", "1992/93", "1993/94",  
"1994/95", "1995/96", "1996/97", "1997/98", "1998/99", "1999/00", "2000/01", "2001/02",  
"2002/03", "2003/04", "2004/05", "2005/06", "2006/07", "2007/08", "2008/09", "2009/10",  
"2010/11", "2011/12", "2012/13", "2013/14", "2014/15", "2015/16", "2016/17", "2017/18",  
"2018/19", "2019/20", "2020/21", "2021/22")
```

```
Years <- as.factor(years)
```

```
TotalIntStudentsInThousands <- IntStudents$TotalIntStudents/1000
```

```
plot(Years, TotalIntStudentsInThousands, xlab="Years", ylab="Total International Students in  
thousands", main="Years vs. Total international students enrolled")
```

```
barplot(IntStudents$AnnualPercentChange, ylim = c(min(IntStudents$AnnualPercentChange),  
max(IntStudents$AnnualPercentChange)), xlab="Years", ylab="Annual Percent Change",  
main="Box Chart", names.arg = Years)
```

### Confidence Interval

```
PercentInt = IntStudents$PercentInt
```

```
bs_stat <- c(PercentInt)
```

```
bs_n <- 10
```

```

for (ii in 1:bs_n) {

  bs_spl_percentint <- sample(PercentInt, replace = TRUE)

  bs_stat <- append(bs_stat, mean(bs_spl_percentint))

}

```

```

hist(bs_stat)

```

```

spl_stat <- mean(PercentInt)

CL <- spl_stat + 2 * sd(bs_stat) * c(-1, 1)

print(CL)

```

### Hypothesis Testing

```

# remove "/" and convert to numeric

years_numeric <- as.numeric(gsub("/", "", years))

# Compute the correlation

stat <- cor(years_numeric, IntStudents$TotalIntStudents)

# Bootstrapping

rand_stat <- c()

rand_n <- 1000

for (ii in 1:rand_n) {

  TotalIntStudents_shuffle <- sample(IntStudents$TotalIntStudents)

```

```
new_stat <- cor(years_numeric, TotalIntStudents_shuffle)

rand_stat <- append(rand_stat, new_stat)

}
```

```
hist(rand_stat)
```

```
sum(rand_stat > stat) / rand_n
```

```
z <- (stat < 0) / sd(rand_stat)
```

```
if (z < 0) {
```

```
  2 * pnorm(z)
```

```
} else {
```

```
  2 * pnorm(z, lower.tail = TRUE)
```

```
}
```

### Appendix 3

Institute of International Education. (2022). "International Student and U.S. Higher Education Enrollment, 1948/49 - 2021/22" Open Doors Report on International Educational Exchange.

<https://opendoorsdata.org/data/international-students/enrollment-trends/>