Final Project

Where Do International Students in the U.S. Come From

And What Affects Their Enrollment?

Yerkezhan Medetbekova
Professor Nicole Rockweiler
STAT 228 – Spring 2025
May 1, 2025

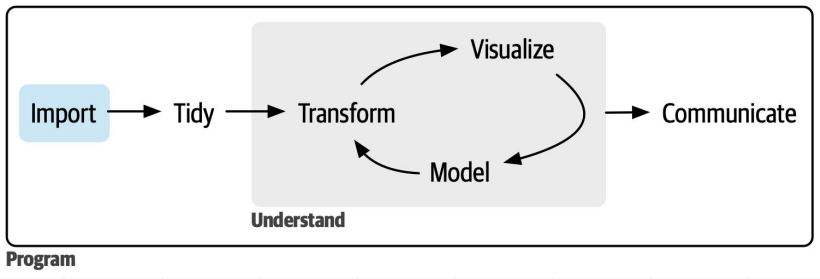
Data & Wrangling

Data & Preparation

- **Source**: Kaggle – International Student Demographics https://www.kaggle.com/datasets/webdevbadger/international-student-demographics

- **Files Used**: 3 out of 5 datasets: origin.csv, field_of_study.csv, status.csv
- **Wrangling**:

1. Cleaned year column
2. Filtered for Graduate & Undergraduate students
3. Created factor() variables (region, academic type)
4. Used group_by(), summarise() to prepare for plots and modeling



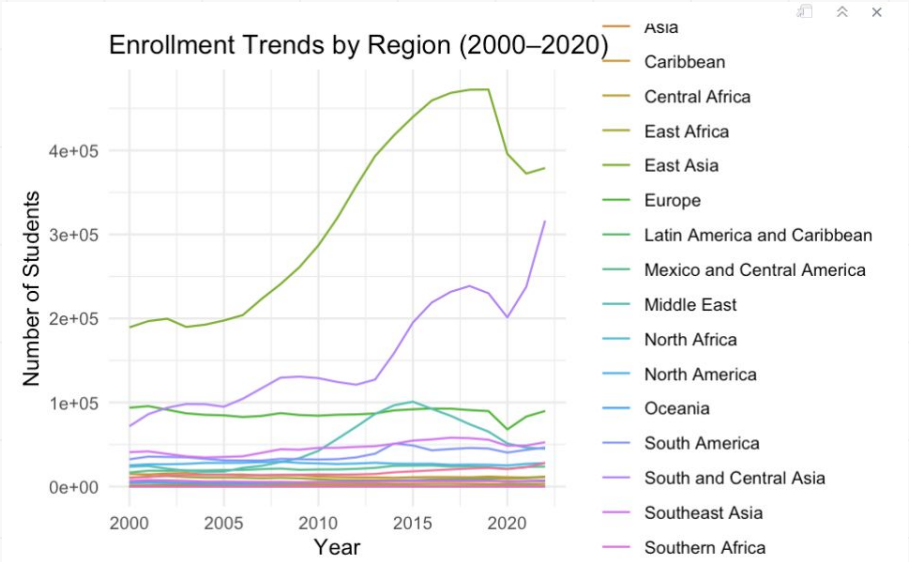

```{r}
glimpse(origin_clean)
```

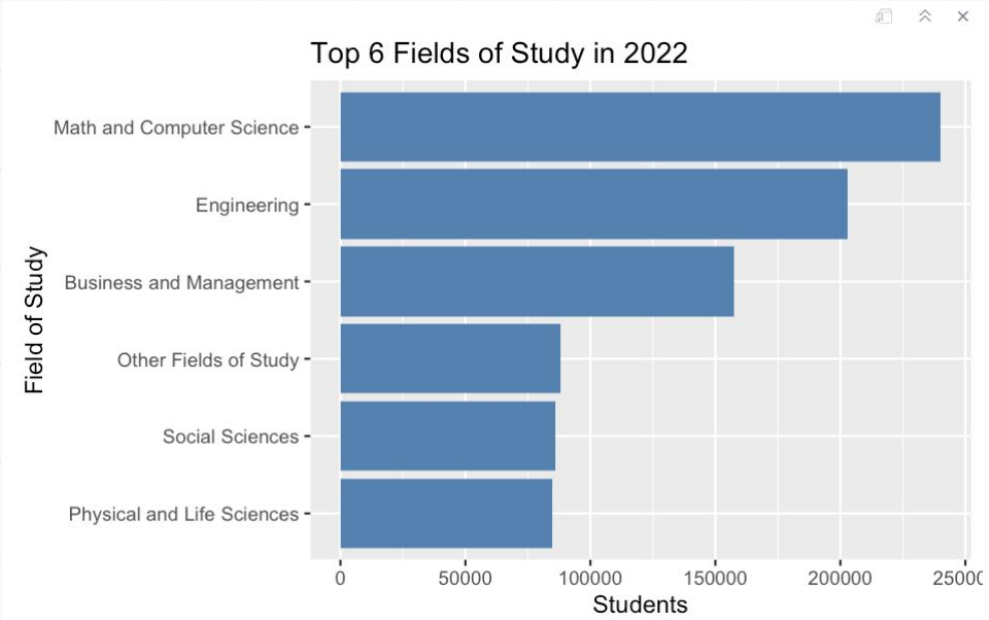
```
Rows: 20,411
Columns: 8
$ year             <chr> "2000/01", "2000/01", "2000/01", "20
$ start_year_text  <chr> "2000", "2000", "2000", "2000", "200
$ end_year         <chr> "01", "01", "01", "01", "01", "01",
$ origin_region    <fct> "Africa, Subsaharan", "Africa, Subsa
$ origin           <fct> "Africa, Subsaharan, Unspecified", "
$ academic_type    <fct> Graduate, Other, Undergraduate, Grad
$ students         <dbl> 2, 0, 6, 0, 6, 0, 5, 1, 59, 32, 4, 2
$ start_year       <dbl> 2000, 2000, 2000, 2000, 2000, 2000,
```

Visualizations & Trends

Exploring the Data: Trends & Patterns

- Regional enrollment trends: Asia leads, but other regions grew over time
- Math, Engineering, and Business were top fields in 2022





Modeling & Results

Modeling: What Explains Enrollment Patterns?

- Model used: Linear regression (lm)
- Response variable: Number of students
- Predictors: Region + Academic Type
- Key takeaway:
- **Graduate programs** have more students than undergraduate
- Some regions send significantly more students than others

Region: East Asia → +129,233
Region: Europe → +34,434
Academic Type: Graduate → +30,000 (vs. undergrad)

```
Residuals:
   Min     1Q Median     3Q    Max
-55759  -1403    -66   1097 130936

Coefficients:

                                          Estimate
(Intercept)                               -69.978
origin_regionAsia                           1.128
origin_regionCaribbean                   5745.109
origin_regionCentral Africa              1196.717
origin_regionEast Africa                 4358.478
origin_regionEast Asia                 129233.239
origin_regionEurope                     34433.609
origin_regionLatin America and Caribbean   -0.324
origin_regionMexico and Central America  9273.261
origin_regionMiddle East                20286.804
origin_regionNorth Africa                2325.326
origin_regionNorth America              12240.304
origin_regionOceania                     2329.674
origin_regionSouth America              15489.043
origin_regionSouth and Central Asia     58071.196
origin_regionSoutheast Asia             19843.326
origin_regionSouthern Africa             2733.891
origin_regionStateless                      5.152
origin_regionWest Africa                 7141.109
academic_typeUndergraduate                141.651
```

Conclusion & Limitations

What I Learned:

- Asia sends the most international students, but other regions are growing
- Graduate programs attract more students than undergraduate
- Fields like Math, Engineering, and Business are most common

🔭 Next Steps:

- Add economic variables (like GDP, tuition) in future analysis
- Explore classification models to predict visa type or academic level

⚠️ Limitations:

- Data does not account for COVID-era changes (post-2020)
- No information about country-specific policies or tuition costs
- Modeling only used region and academic type — other factors may matter

Thank you!

References:
https://www.kaggle.com/datasets/webdevbadger/international-student-demographics