CA682 Data on Management and Visualisation

December 16,2018

Ashish Vijay Lakamale (18210227), Dublin City University
M.Sc. in Computing (Data Analytics), MCM
ashish.lakamale2@mail.dcu.ie

Disclaimer:

An essay submitted to Dublin City University, School of Computing for module CA640 Professional and Research Practice, 2018/2019.

I understand that the University regards breaches of academic integrity and plagiarism as grave and serious. I have read and understood the DCU Academic Integrity and Plagiarism Policy. I accept the penalties that may be imposed should I engage in practice or practices that breach this policy.

I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations, paraphrasing, discussion of ideas from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged, and the sources cited are identified in the assignment references.

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

By signing this form or by submitting this material online I confirm that this assignment, or any part of it, has not been previously submitted by me or any other person for assessment
on this or any other course of study. By signing this form or by submitting material for assessment online I confirm that I have read and understood DCU Academic Integrity and Plagiarism Policy (available at: http://www.dcu.ie/registry/examinations/index.shtml)

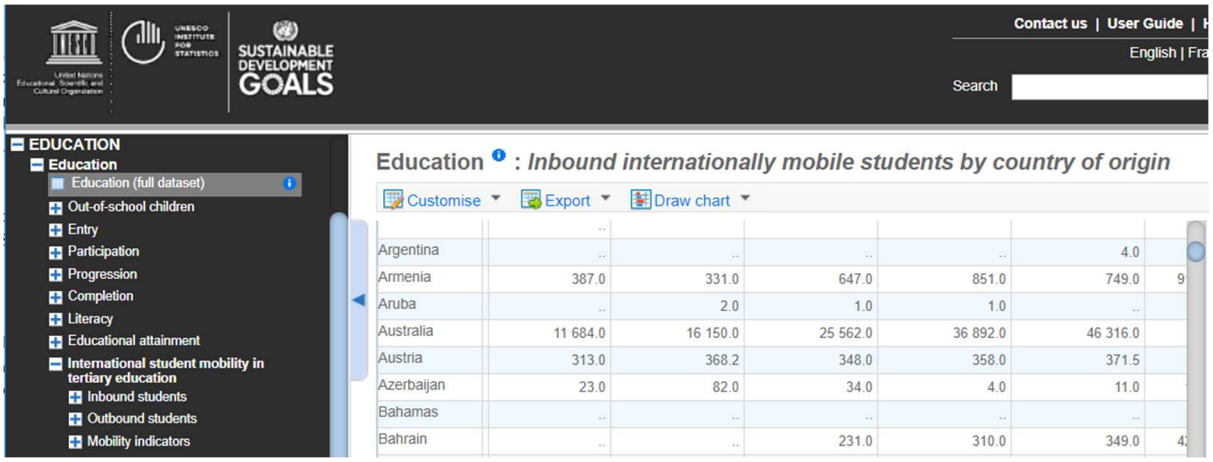
Name: Ashish Vijay Lakamale
Date: 16/12/2018

Global Flow Of Tertiary Level Students:

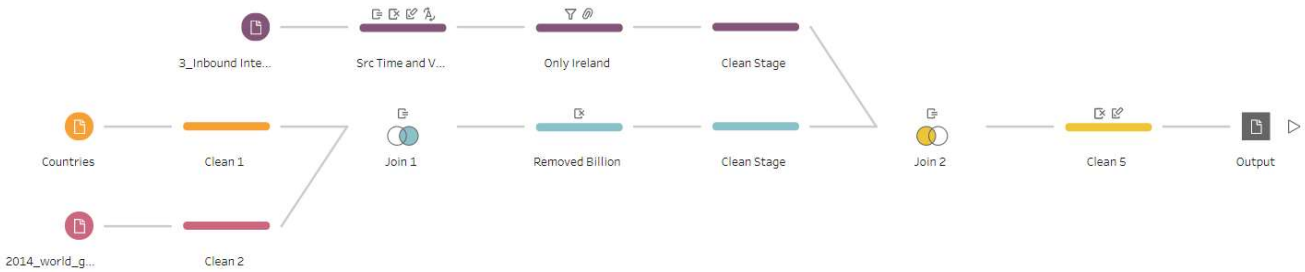
The mobility of students and academics across borders has increased tremendously in the last decade. By enrolling in a study abroad program, one gets the chance to see a side of your major that you may not have been exposed to at home. You'll find that completely immersing yourself in the education system of your host country is a great way to really experience and understand the people, its traditions, and its culture. Education is the centrepiece of any study abroad, after all, a study abroad program—and choosing the right school is a very important factor. When you finish your study abroad program one absorbs new perspective on culture, language skills, a great education, and a willingness to learn. All of these are very attractive to future employers.
In 2016, there were nearly 5.1 million internationally mobile students (i.e. 2.3% of all tertiary students), up from 2.1 million in 2016. The United States and the United Kingdom attracted one-third of all international students in 2016.

Dataset:

The **United Nations Educational, Scientific and Cultural Organization** (**UNESCO**) is one of the most trusted popular platform for data visualisation and analytics. UIS data on the mobility of students shed light on the shifting demand for higher education, particularly in the developing world. This data contains over 12500 rows and 7 different fields.



Data cleaning process and converting it into data frames is done using **Tableau Prep** and **Pandas library** in Python. Datasets of all the country maps are merged together to form one single large dataset. This dataset is then used for visualization. There are 3 different datasets are used i) International Student Flow, ii) Countries with code names and iii) World map with latitude and longitudes. The data set **cleaning flow** is given below,



The head of the dataset is given below.

In [10]:
```
import plotly.offline as py
import pandas as pd
from plotly import graph_objs as go

df_aus = pd.read_csv('D:\Visualisation\Tableau\TertiaryInternationalStudents\DateWise\Canada_2016.csv')
df_aus=df_aus.fillna(0)
sorted(df_aus)
df_aus.head()
```

Out[10]:

	SRC COUNTRY	LATITUDE	LONGITUDE	name	COUNTRY	CODE	DST COUNTRY	TIME	Value
0	Macedonia	41.608635	21.745275	Macedonia	Macedonia	MKD	Canada	01-01-2016	21.0
1	Macedonia	41.608635	21.745275	Macedonia	Macedonia	MKD	Canada	01-01-2015	18.0
2	Macedonia	41.608635	21.745275	Macedonia	Macedonia	MKD	Canada	01-01-2014	15.0
3	Macedonia	41.608635	21.745275	Macedonia	Macedonia	MKD	Canada	01-01-2013	12.0
4	Macedonia	41.608635	21.745275	Macedonia	Macedonia	MKD	Canada	01-01-2012	156.0

PROCESS:

The objective is to determine top 8 countries where International students prefer to go for study and to identify the cause which led these countries to host the most number of International students for tertiary level education. Following graphs and libraries are used for visualisation.

1. Geographical Plot

Here we are using a Geographical plot to compare and describe, where do students go to study? And where they come from? using the world map. This plot will give the audience an overall perspective of international student's choices for tertiary education. For this Python's Plotly library is used to create interactive maps and data exploration. Plotly is a powerful Python library that helps to create a different kind of world maps.

A **choropleth map** is a thematic map in which areas are shaded or patterned in proportion to the measurement of the international student count variable. Choropleth maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region. A heat map is similar but does not use geographic boundaries. The countries are represented using choropleth and are displayed using latitude and longitude. Also, the colour-bar give the range of student count for that country, located right of the world map. A drop-down menu is given for interaction with the map. Menu is created with the help of plotly. On Selecting one of the given country, it will highlight other major countries from where students prefer to come to the selected country. Hover info is used to show the exact count of students travelling for education purposes.

2. Bar Chart

A bar graph used to show how top 8 countries around the world host international students. Along the vertical axis, or axis Y, the maker of the graph would plot a quantitative or numerical scale i.e. top 8 host countries. On the horizontal axis, or axis X, the graph maker might plot a category, i.e. the number of universities from the top 500. In this way, viewers can easily see how which country has more number of universities that ranks in the top 500 universities around the world.

3. Line Plot

Line graphs are drawn so that the independent data are on the horizontal x-axis and the dependent data are on the vertical y-axis. They can also be used to display several dependent variables against one independent variable.

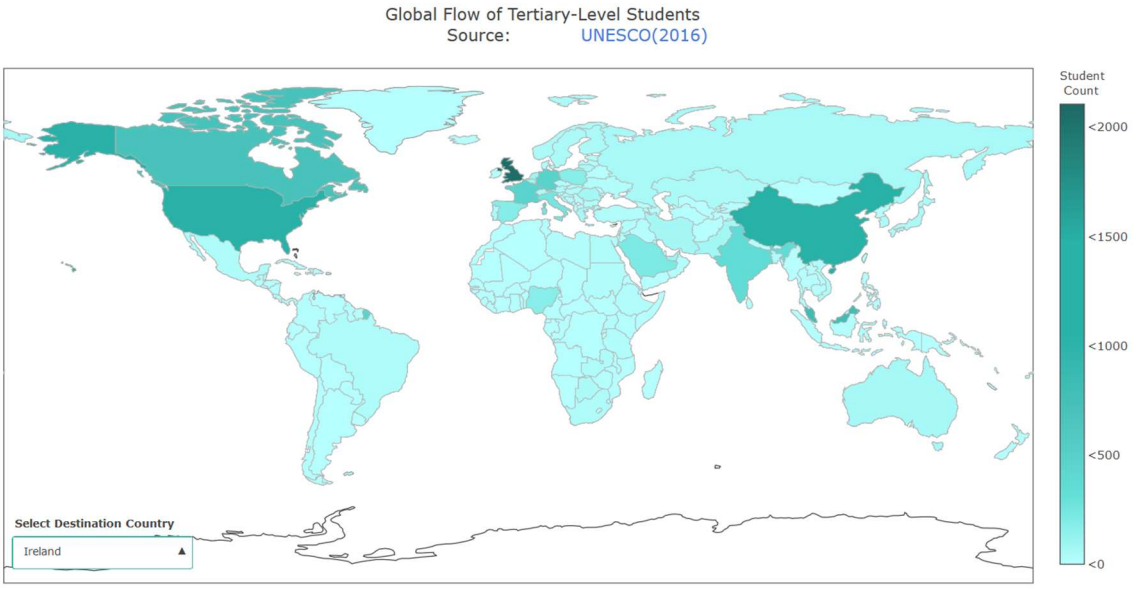
The other option of Bar chart was neglected as it is difficult to differentiate between two categories if their values are close to each other. In contrast, the line graph clearly helps in differentiating between the count of students from different countries.
Here we are comparing the student flow in different countries for the year 2016. Since we are plotting one quantity against its category, a line chart is a suitable option. The line chart is created with the help of the Plotly library. Along the vertical axis, or axis Y, the maker of the graph would plot a quantitative or numerical scale i.e. top 8 host countries. On the horizontal axis, or axis X, the graph maker might plot a category, i.e. the total number of student flow. In this way, viewers can easily see which countries are hosting more number of international students and compare the capacity of different countries to accommodate the student flow.

Result:

The graphical representation helps the audience to quickly know which countries have more share in international students. Since people are used to seeing such representation of maps in Google apps, they quickly relate to it. This helps in the quick understanding of the context on display. The user can zoom in to the map if he/she wishes to have a look at the number of students travelling for study.

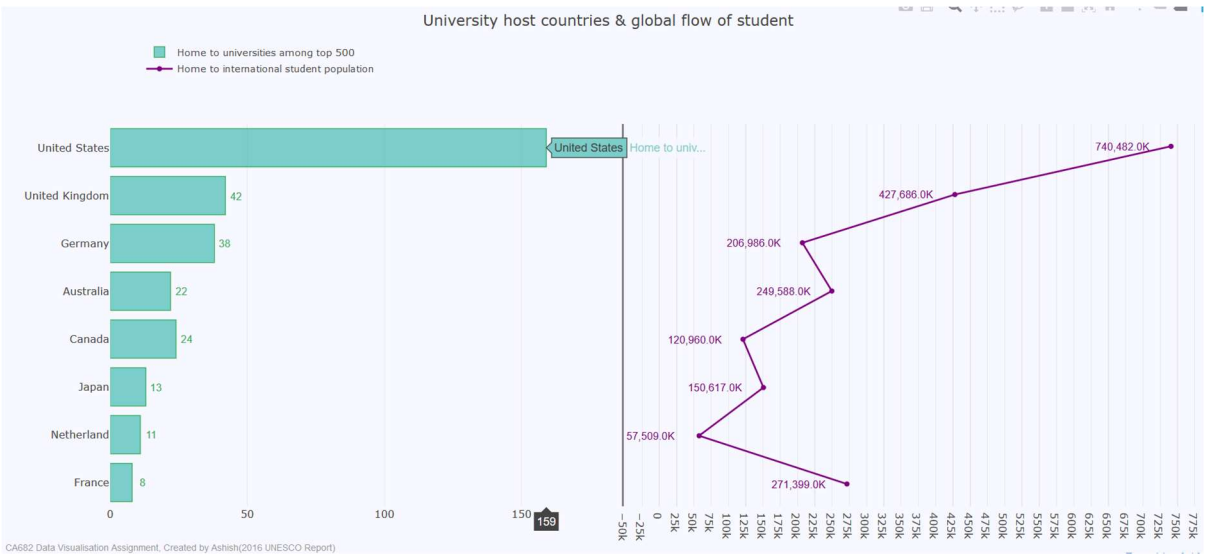
i) The choropleth highlights major countries from where students prefer to come to a selected country. Hover info is used to show the exact count of students travelling for education purposes. Here we can visualise the concentration of colours from where more students come i.e. China, India, South Korea, Saudi Arabia and Canada are few countries of origin of international students.

```
In [2]: py.plot(fig, validate=False, filename='d3-world-map.html')
Out[2]: 'file://C:\\Users\\alakamale\\d3-world-map.html'
```



ii) Both the plots are adjacent to each other, clearly showing the university and student flow comparison. Here it describes the trend of student flow and what causes students to select these countries over others. We can conclude that since USA and UK are countries hosting more universities that fall under the top 500 universities around the world which led students to choose these university and country as well over others. Well, there are other factors too involved in why students travel to these countries, but education is a major one.

```
In [4]: py.plot(fig, filename='University_Global_Student.html')
Out[4]: 'file://C:\\Users\\alakamale\\University_Global_Student.html'
```



Conclusion:

Above visualisation of global student, flow dataset tells us that with more number of prestigious universities attract a higher number of international students.

Refrences and Resources:

1. UNESCO, Sustainable development goals, Outbound internationally mobile students by host region. http://data.uis.unesco.org/Index.aspx?queryid=172#
2. Open Source Graphing library, Plotly. https://plot.ly/python/reference/#layout-geo-projection
3. Tableau Prep, for cleaning and refining of data.
4. Python, for creating graphical maps.
5. Slide Presentation at https://prezi.com/view/4XRTECjj2fEYLLfCbBl0/