Title

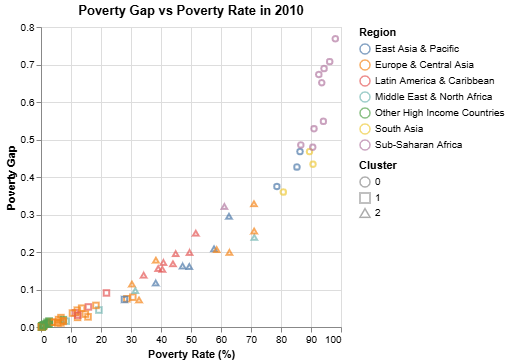
**Analyzing the trends in individuals pursuing higher education internationally and identifying the socioeconomic factors that correlate with these changes.**

# Team 3: Brynn Hare, Nora Manolescu, Sam Muehlebach, Kaitlyn Clements, Kyle Johnson

## **Introduction**

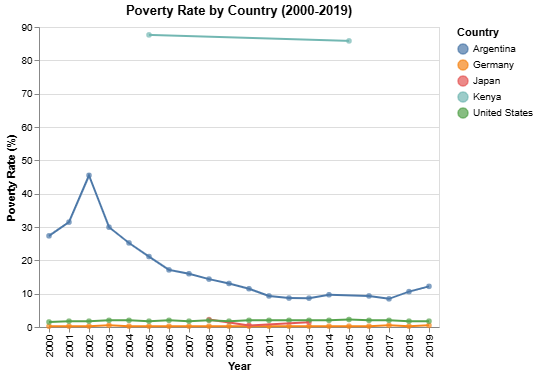
This report explores and analyzes relationships between higher education enrollments and graduates with socioeconomic factors of different countries and regions such as the poverty rate, GDP rate, literacy rates, government spending on education, internet access per country, attendance ratios, enrollment ratios, and graduation rates. It aims to look at how different variables and factors influence higher education across different countries and within specific year ranges.

## **Insights and Visualizations**



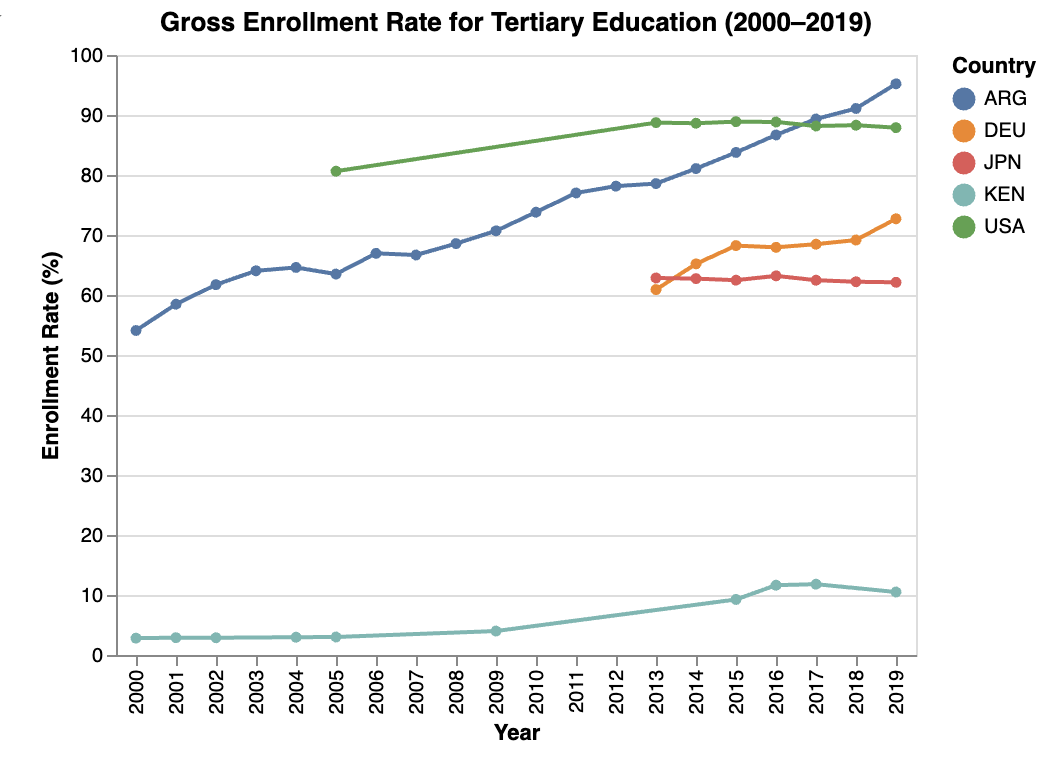
**Fig. 1: Chart displaying the poverty rate percentage compared to the poverty gap worldwide from the year 2010.**

Figure 1 shows the positive correlation between the poverty rate percentage and the poverty gap for different countries across the world. The poverty rate percentage conveys the number of people within the country that are below the poverty line. The poverty gap conveys the severity of the poverty. In general from the dataset, when more people are below the poverty line in a country, the impacts of poverty are more severe. The chart is color coded based on the regions of the countries from the dataset. The color coding reveals that countries in similar regions tend to have similar poverty rate percentages and poverty gap values. Clustering was utilized in order to group countries by these poverty characteristics and accurately predict the cluster of a country that was not in the original dataset to estimate what their poverty rate percentage and poverty gap would be.



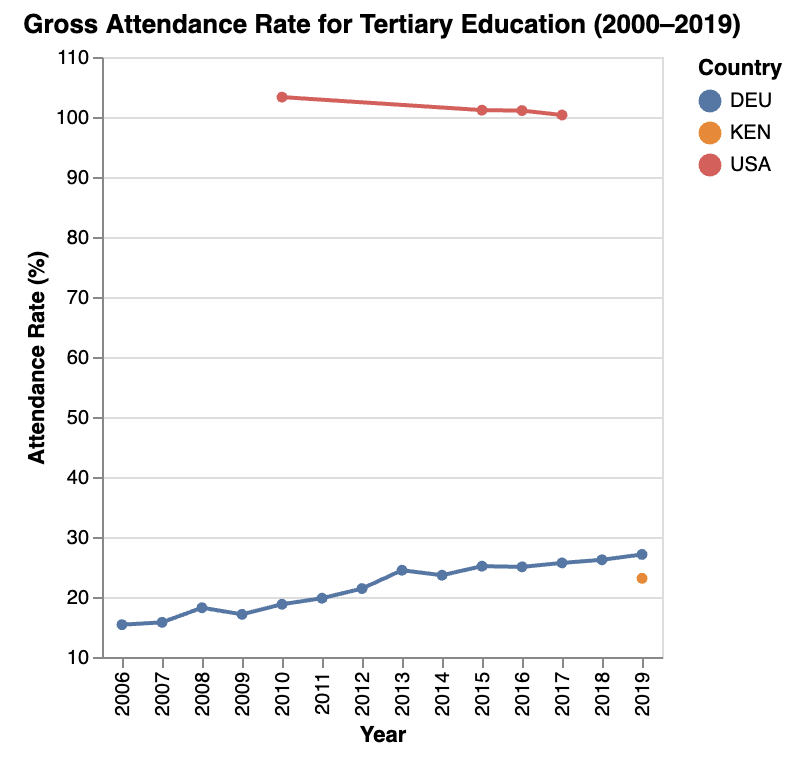
**Fig. 2: Graph showing the poverty rate percentage over the years of 2000 - 2019 in the following countries- Argentina, Germany, Japan, Kenya, United States of America.**

The poverty rate percentage is calculated with the poverty line of $6.85. The poverty rate percentage represents the percentage of the population living on less than $6.85 per person per day. The line chart reveals that Kenya has a significantly higher poverty rate percentage compared to Germany, Japan, and the United States. These poverty rates reflect the enrollment rates for tertiary education in an inverse way where Kenya has far lower enrollment rates than Germany, Japan, and the United States.



**Fig. 3: Graph showing the enrollment rate (%) over the years of 2000 - 2019 in the following countries- Argentina, Germany, Japan, Kenya, United States of America.**

As seen from the visual, most countries experienced an overall increase in their enrollment rate from 2000 - 2019.

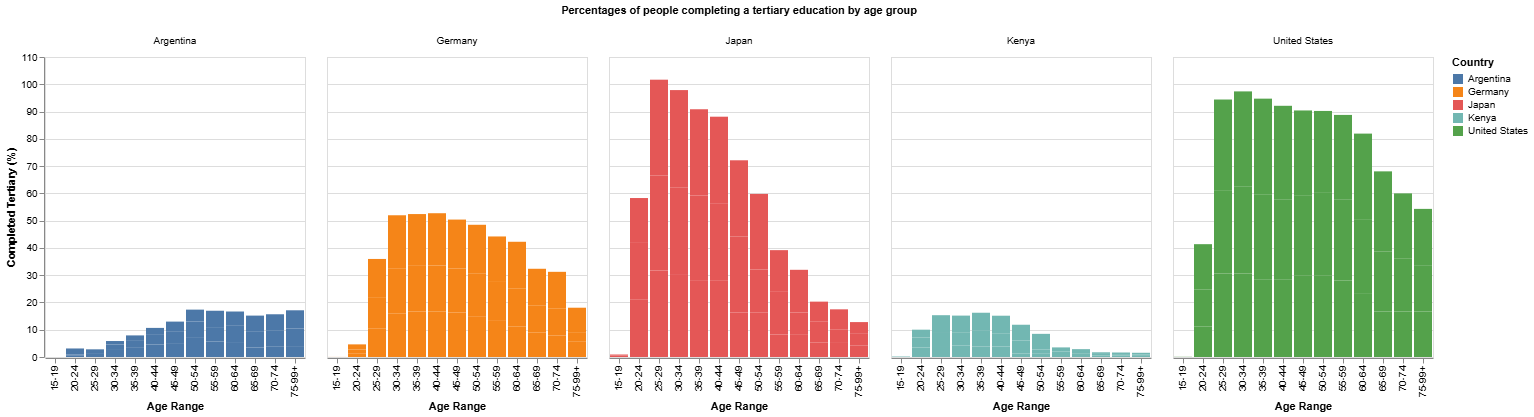


**Fig. 4: Graph showing the attendance rate (%) over the years of 2006 - 2019 in the following countries- Germany,, Kenya, United States of America**

**\* Note that Argentina and Japan did not report this data and are excluded in the visualization. There also was no data present from 2000 - 2005 for any country, so this visualization begins in 2006.**

A rate of over 100% attendance was seen in the United States for the data reported. This may be due to students attending multiple schools at the same time. E.g., one student attending the University of Kansas while also attending some classes at Johnson Community College would lead to an attendance rate of 200% for the population of the one student.

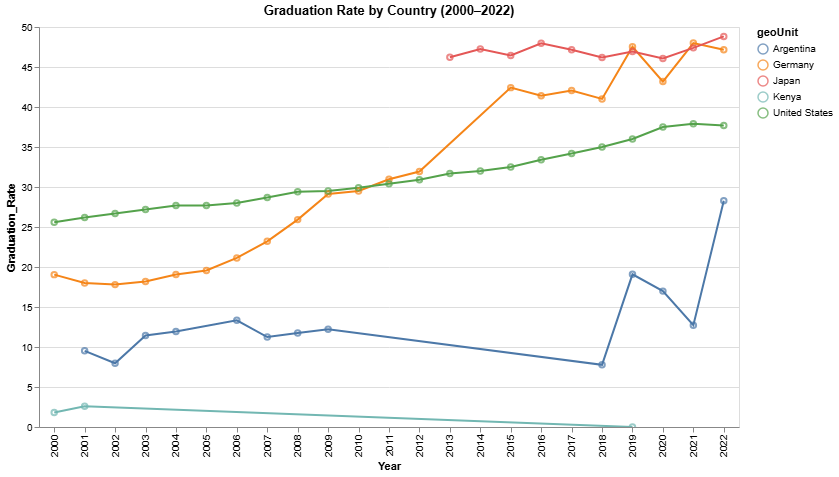
As seen from the visual, Germany’s reported attendance rate is significantly lower than the country’s reported enrollment rate. The United States and Kenya were within a much closer range for the two observations.



**Fig. 5: Charts displaying the age at which a tertiary education is obtained for Argentina, Germany, Japan, Kenya, and the United States from the year 2010.**

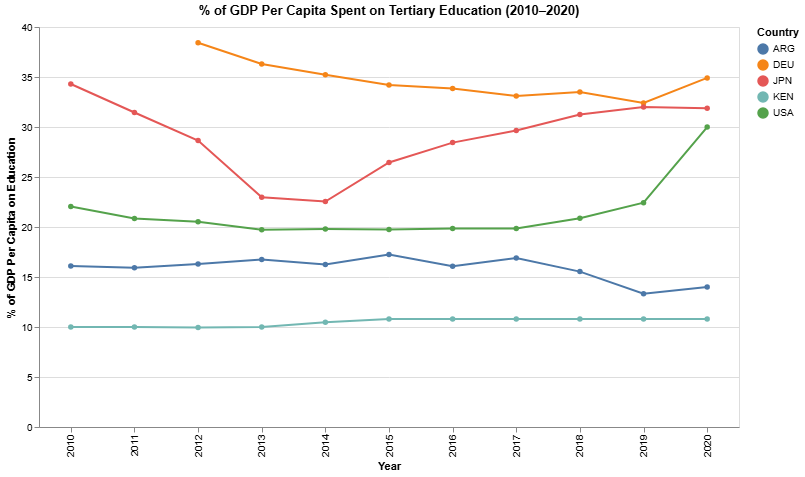
The bar charts display the age at which a tertiary education is completed for people in Argentina, Germany, Japan, Kenya, and the United States. The charts trend downwards as the age increases past roughly age 40 reflecting the decline in overall population numbers of people in those age ranges. Japan and the United States have a drastic increase in people getting their tertiary education from the ages of 20 to 29 while Argentina, Germany, and Kenya remain steady over all of the ages. The charts also reveal that more people in Japan and the United States are obtaining a tertiary education compared to people in Argentina and Kenya.

After exploring national graduation rate trends across Argentina, Germany, Japan, Kenya, and the United States, it became important to understand whether these outcomes can be explained or predicted by how much each country invests in education. To investigate this, we conducted a linear regression analysis using education spending (as a percentage of GDP per capita) to predict graduation rates in the following year (N+1 model). This approach tests whether an increase in funding precedes improvements in graduation rates, showing insight into how educational outcomes may be to economic input.



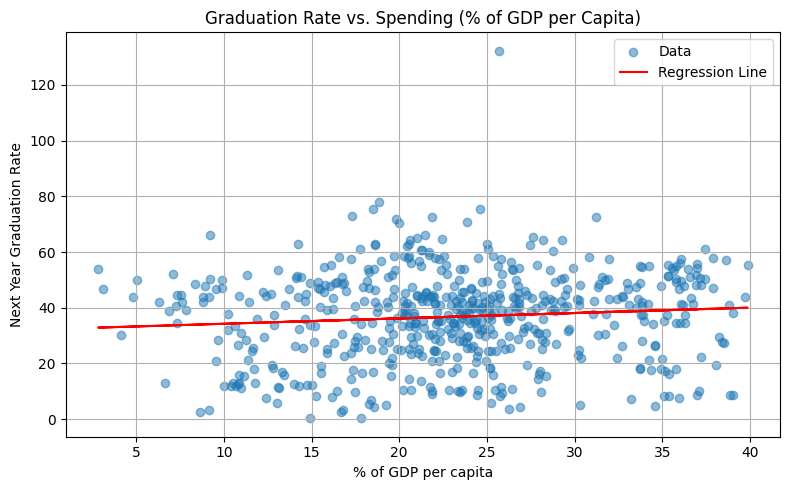
**Fig. 6: Graduation Rate Trends by Country (2000–2022)**

Graduation rate trends show different disparities across various countries. Germany and Japan always have high and rising graduation rates, while the United States shows a steady rise. The trend in Argentina is more volatile, and Kenya has the lowest graduation rates for most of the time period considered.

****

**Fig. 7: Education Spending Trends as % of GDP Per Capita (2010–2020)**

This figure shows the percentage of GDP per capita each country spends on education. Germany and Japan consistently outpace others in investment, followed by the U.S. and Argentina. Kenya’s consistently low investment aligns with its lower graduation rates, though the relationship is not strictly proportional when viewed across all countries. Together, these findings reinforce that while spending is important, graduation outcomes depend on a broader ecosystem.

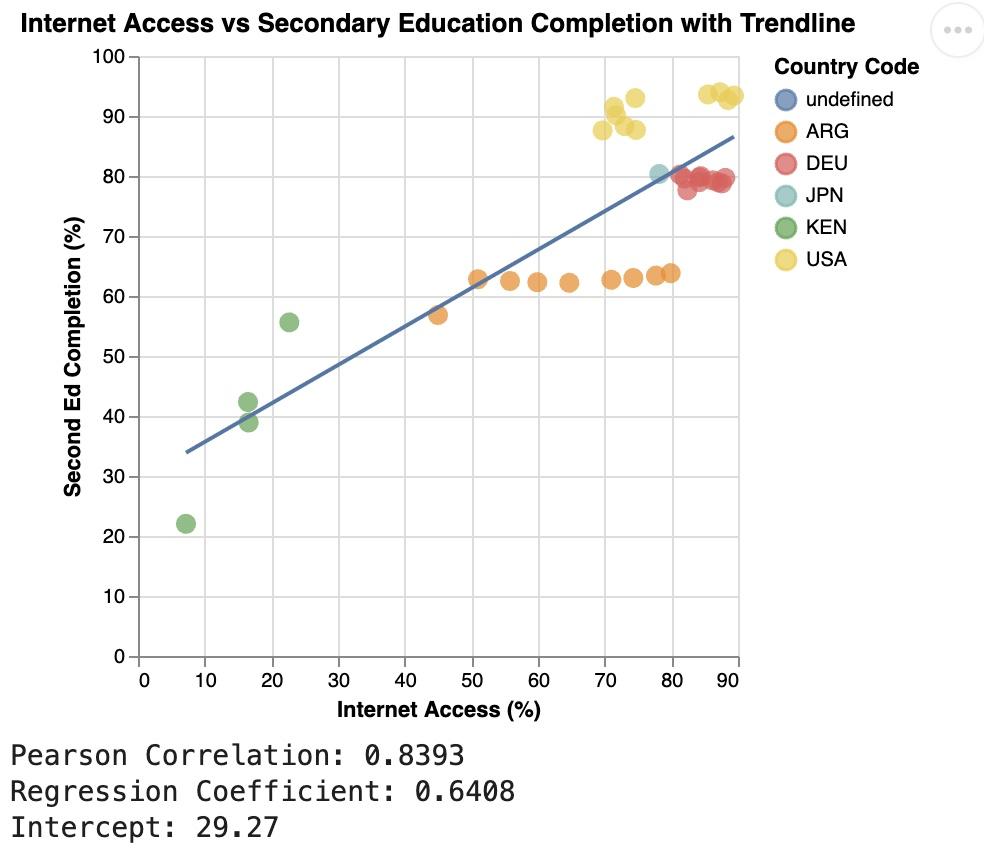
****

**Fig. 8: Relationship Between Education Spending and Graduation Rates Across Five Countries (N+1 Model)**

This figure examines the relationship between education spending in one year and graduation rates in the following year, across five countries. Despite the hypothesis that more spending might yield better outcomes, the regression model reveals an extremely weak relationship: R² = 0.009, Coefficient = 0.193, Intercept = 32.33, Mean Squared Error (MSE) = 252.79 While the trend line is slightly upward, indicating a mild positive association, the model explains less than 1% of the variance in graduation rates. This suggests that although financial investment is a critical input, other structural or policy driven factors are likely more impactful in determining graduation success.

This section focuses on the correlation between internet access and upper secondary education completion across the five selected countries from 2010 to 2019. For this, three DS practices were used: regression to quantify the relationships, clustering to determine groupings with similar completion and access, and statistical inference to test the impact of the differences in the data.

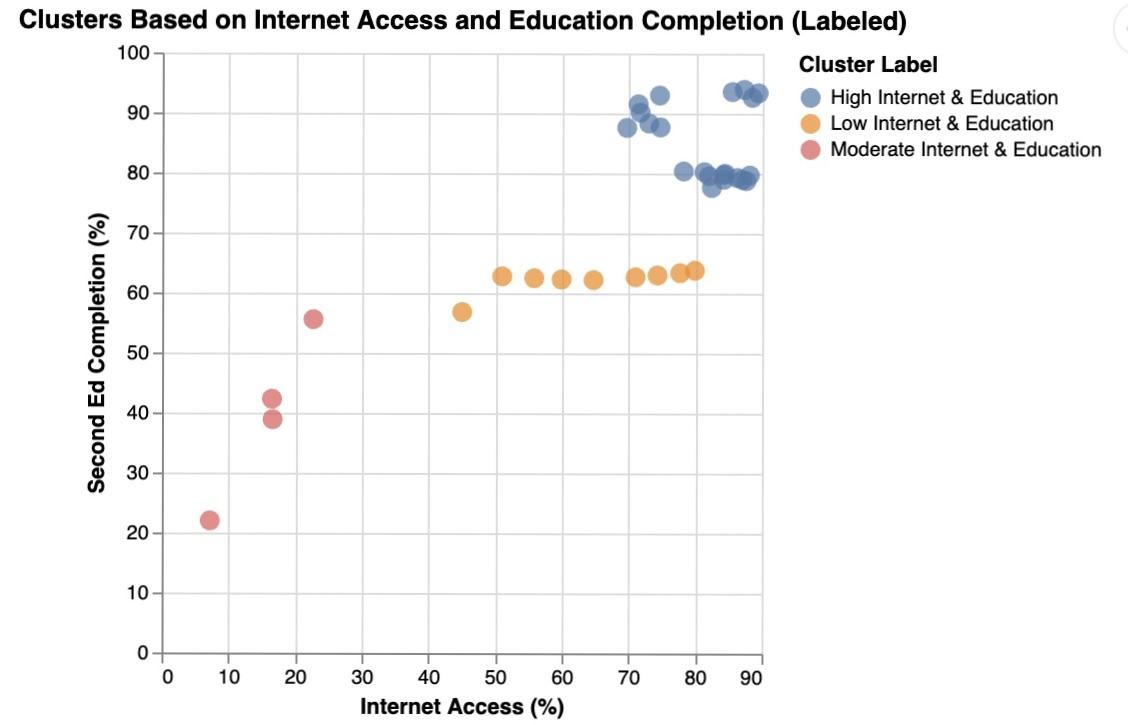
The first is regression analysis. Linear regression was used to determine the magnitude of the correlation between internet access and secondary education completion. This was done by using the LinearRegression library from sklearn, creating and fitting a model where internet access was the independent variable and education completion as the dependent variable.



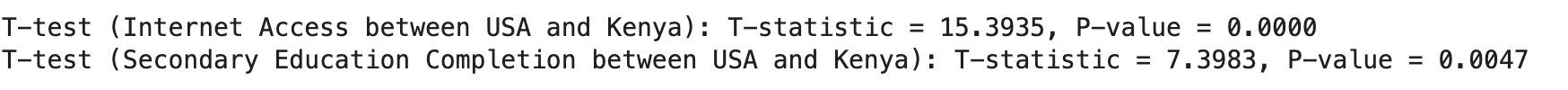
**Fig 9: Scatter Plot showing relationship between internet access and secondary completion in Argentina, Germany, Japan, Kenya, and the United States of America according to data from 2010 to 2019**

The model’s Pearson correlation and Regression Coefficient values show that there is a strong positive correlation between these two datasets. Countries with more access to the internet tend to have more higher secondary education graduates.

The second is clustering. In order to group countries with similar internet access and secondary education completion rates, K-Means clustering was used via the KMeans library from sklearn. StandardScalar wsa also used to ensure that all input features were weighted equally to avoid inaccurate conclusions. This resulted in three cluster labels: High Internet & Education, Moderate Internet & Education, and Low Internet & Education:

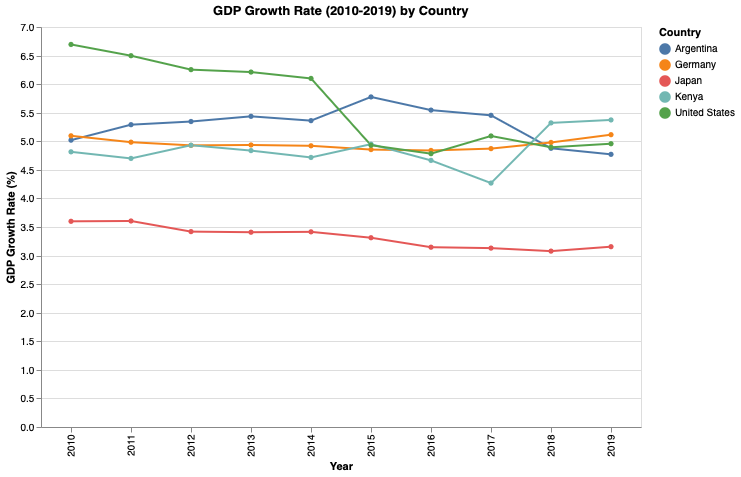


**Fig 10: Scatter Plot showing country-year data groups through K-means clustering. This was split into three categories: High, Moderate, and Low Internet and Education**

The last is statistical inferencing. To determine if the differences in secondary education completion and internet access were statistically significant, independent sample T-tests were performed on two countries with opposite cluster labels: USA and Kenya. This was done using ttest\_ind and scipy.stats, which outputted the following:  
  


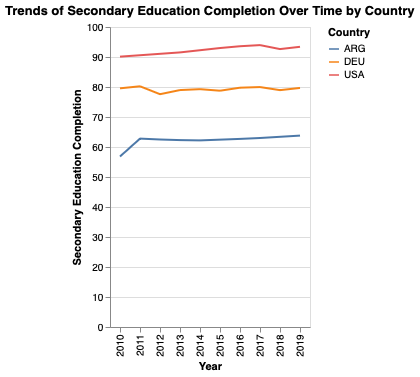
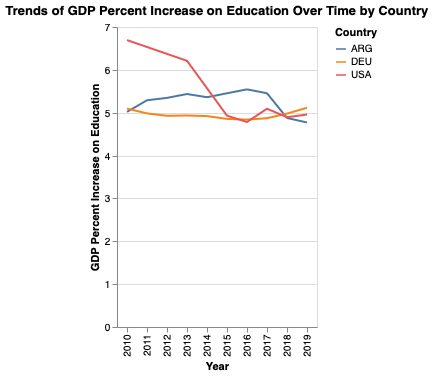
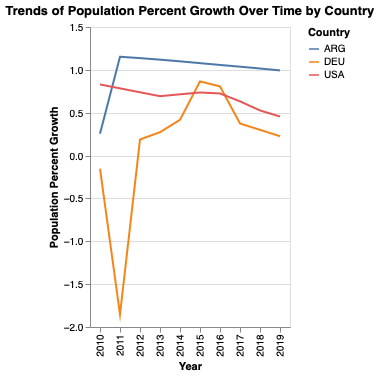
Both p-values fall below .05, which shows that the differences in internet access and secondary education completion in USA and Kenya are statistically significant, rather than being caused by random variation

This data science analysis using Python helped us conclude that there is a strong relationship between internet access and secondary education completion. Countries with more internet access were shown to have higher secondary education completion rates than countries with less access. These methods show that investing in digital infrastructure, such as the internet, could strongly impact a student's ability to qualify as a college student.



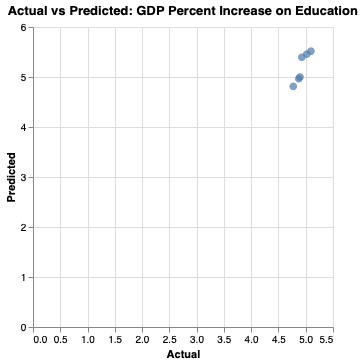
**Fig 11: Line plot showing government expenditure on education as a GDP % from 2010-2019**

In this section, we will be using multivariable regression to predict government expenditure on education as a GDP % for a given year and country. We will use variables based on data and findings discussed earlier to form an accurate, useful model for this. The above plot, Fig X, depicts the actual data on government expenditure as a GDP % from 2010-2019 for three out of our five countries of interest: Argentina, Germany, and the United States. The selection for these countries was based on their differences in values from methods above, and the data availability. We will use this information for testing and understanding the accuracy of our model.

**Fig 12: Individual line plots for government expenditure on education as GDP percent increase and our variables of interest, secondary education completion and population percent growth.**

After generating simple line plots for Argentina, Germany, and the United States for government expenditure on education as GDP percent, secondary education completion, and population percent growth, we can observe possible patterns visually before creating our model to understand what we are working with. We will be creating a model to predict the expenditure on education as a GDP percent. As we can see, it appears that the United States has a clear, sharp decline on expenditure on education from 2010-2019. There is also a decline in the later years within that range for Argentina, with an increase in the earlier years. Finally, it seems that of the three countries, Germany has had the most consistency and has been slowly increasing in the later years.

When observing secondary education, all three countries are shown to have a slight, consistent increase, with the United States having higher numbers overall, and Argentina having the lowest. And finally, for population growth, all three countries also share a decrease within this range of years, with Germany having a sharp increase and decrease earlier and later.



**Fig 13: Resulting plot using our multivariable regression model to predict GDP percent increase on education using features population growth, secondary education completion, and year for each country.**

We built our model using government expenditure on education as a GDP percent as the target variable, and feature variables population percent growth, secondary education completion, and year. The above scatterplot shows the actual vs. predicted values based on our model and the data points from the actual dataset for our target variable. We can see there are limited points due to the limitations of our datasets, such as the few countries we used and the small range of years. Additionally, we ended up with a very weak value for the correlation coefficient. This is likely because of the features we selected. This is an expected outcome given the plots we generated in the beginning for each of our feature variables. Therefore, we can determine creating a linear model would not be the best way to correlate these variables. In the future, another form of regularization or even a tree-model should instead be used for connecting these variables to GDP percent increase.

## **Conclusion**

This analysis reveals that different socioeconomic factors influence higher education variables for individuals differently across countries and time periods. For example, countries with high poverty rates such as Kenya and Argentina, also have lower amounts of individuals completing secondary education across different age ranges, whereas those with lower poverty rates such as the United States, Japan, and Germany, have much higher amounts of individuals completing secondary education across different age ranges. Based on our multiple linear regression model, factors such as population growth do not seem to correlate with government spending on education. However, we see a strong relationship with completion of secondary education and internet access. Therefore, factors regarding resource availability such as poverty and internet access seem to have the most significant impact on individuals pursuing higher education. This would make sense as actual access to higher education would come from access and completion of secondary education, and having the resources and environment to allow individuals the time and education to pursue degrees.