How Does Access to Education Look Around the World?

Victoria Suarez

MS in Data Science, University of San Francisco

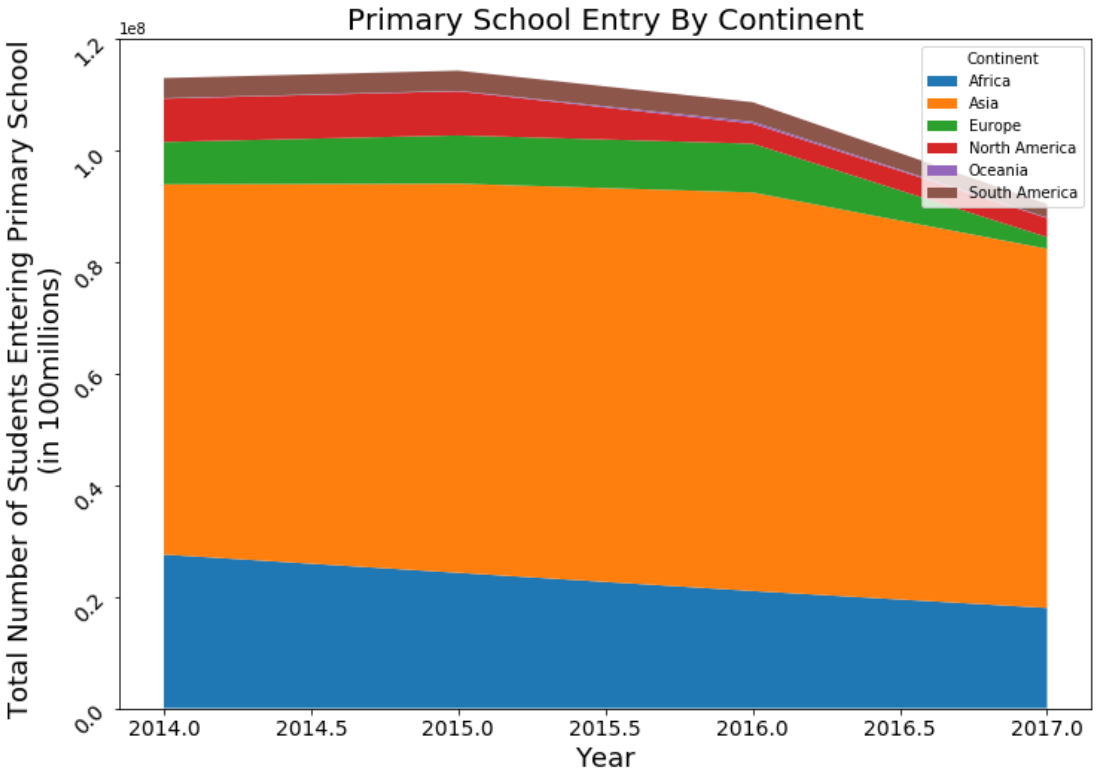
Introduction

Access to education is something many of us take for granted. Studied by The United Nations Educational, Scientific and Cultural Organization, UNESCO, have shown that 263 million children are out of school and girls are still more likely to never go to school in the first place [1]. These shocking statistics made me wonder why this is our current state for equal opportunity to education around the world. UNESCO has a vast amount of data on education broken down by country and by gender. Their datasets include: counts on primary school entry, rates of primary school completion, rates on youth literacy and illiteracy, and expenditure on education – just to name a few. For my analysis, I used data from the mentioned datasets that was collected from 2014 to 2017. Through my visualizations, I hope to show where access to education is most needed, where inequality to education is most prevalent, and potential solutions to address these issues.

Summary Graphs

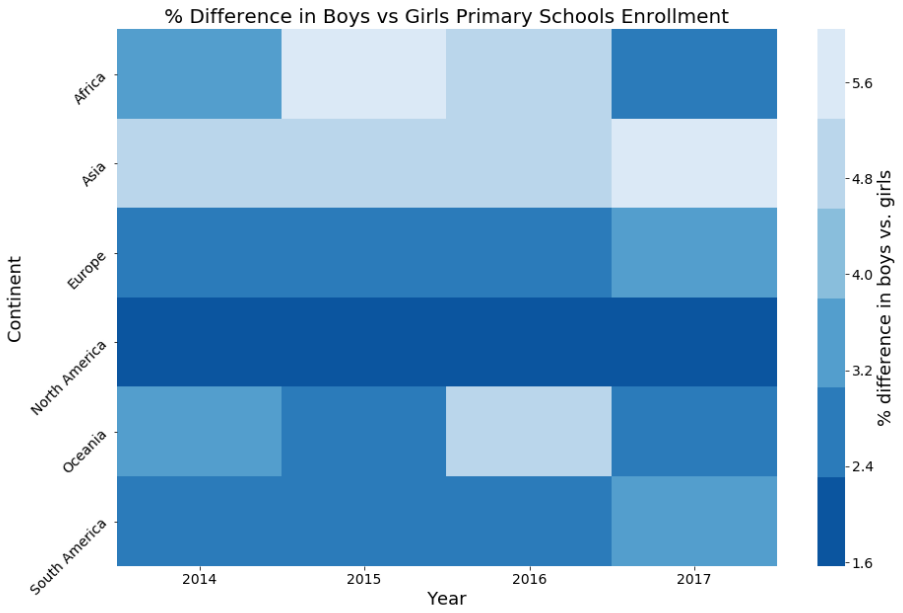
Stack Area Graph

This graph shows the total number of students entering primary school from 2014 to 2017 broken down by continent.



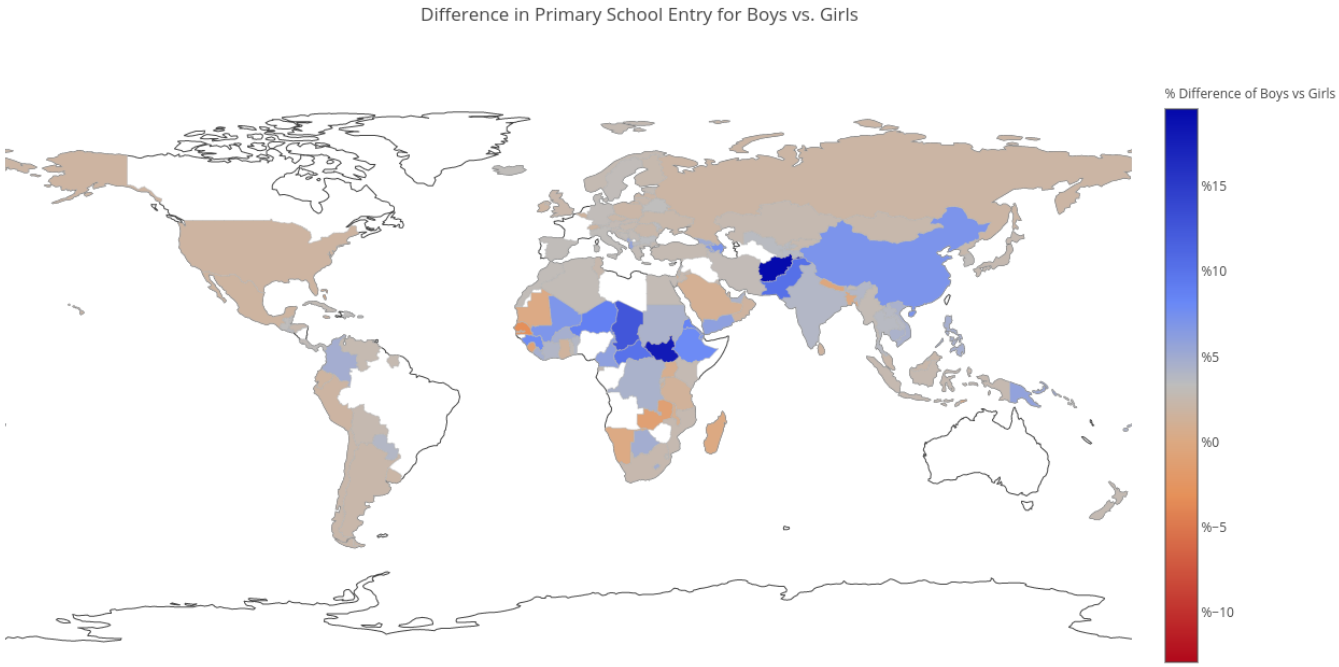
Heat Map

This heat map shows the difference between the percentage of boys vs the percentage of girls entering primary school over time by continent. The difference is calculated by the % of boys – the % of girls. Ideally, we would like for this difference to be approximately 0% – meaning just as many girls are entering primary school. The lighter the box, the larger the gap is between the number of boys vs the number of girls entering primary school. Luckily, most squares are darker shades of blue and most countries become darker shades of blue over time – meaning the trend shows us that inequality in the number of boys vs girls entering into primary school is decreasing.



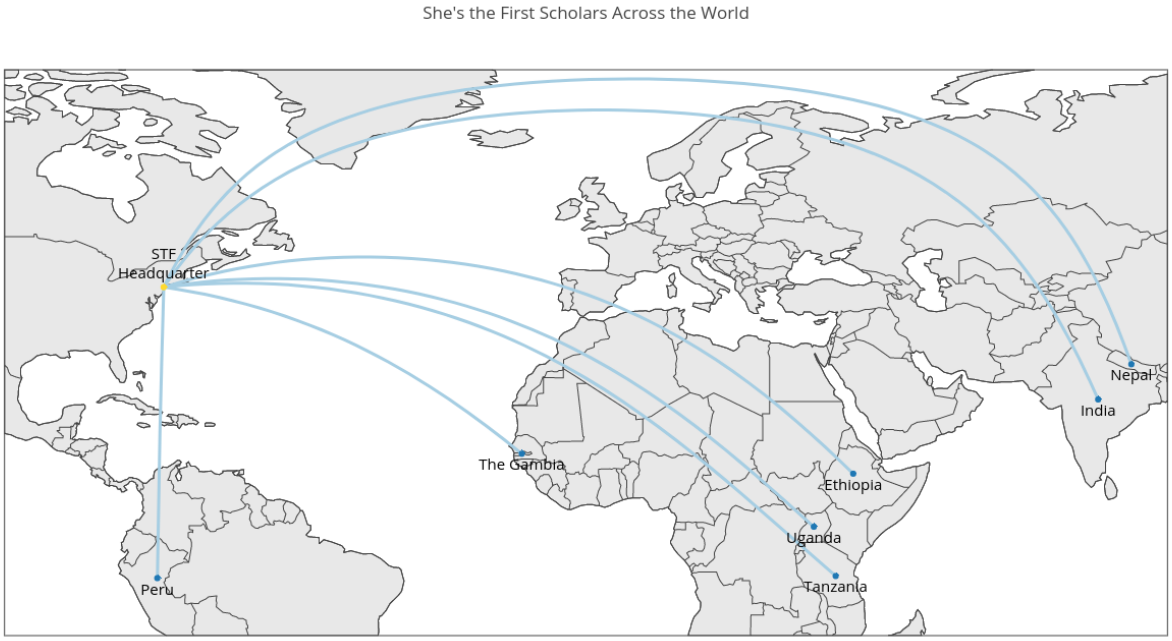
Choropleth Map

This map shows the difference between the percentage of boys vs the percentage of girls entering into primary school by country. The difference is calculated by % of boys – % of girls. The color bar represents how large the gap is.



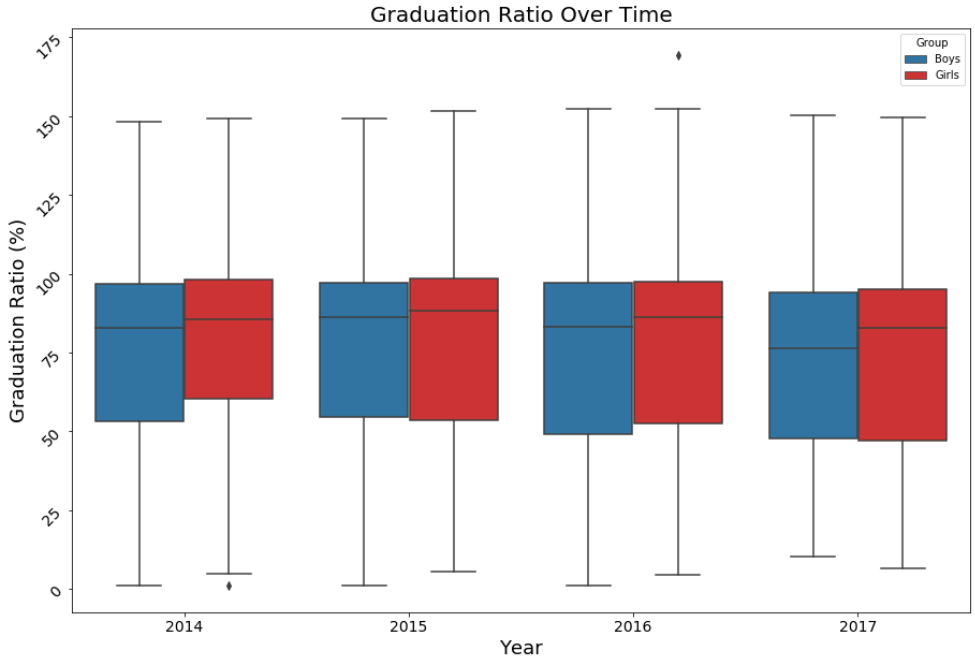
Connection Map

She's the First, STF, is a national non-profit that sponsors girls' education in developing countries. They are based in New York City and have 250+ scholars across 7 counties shown below.



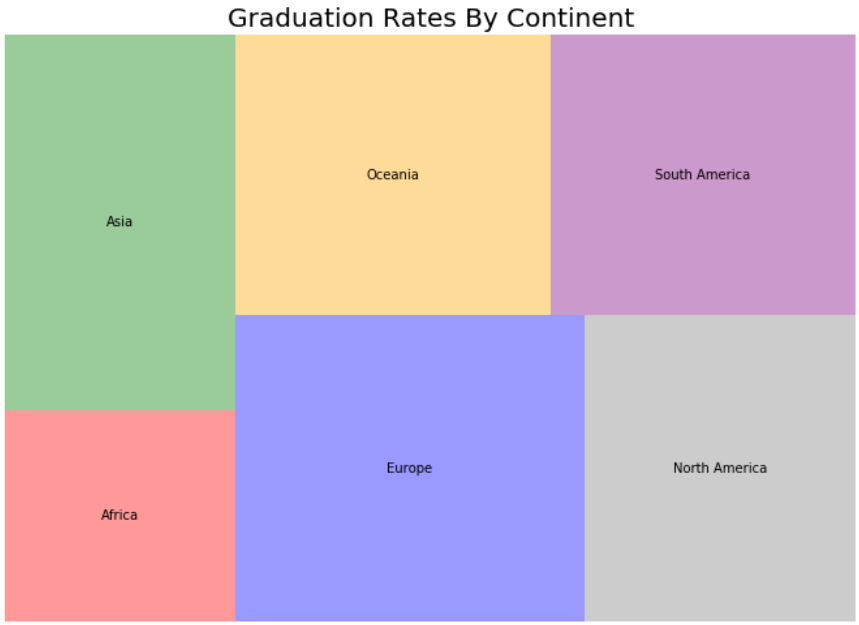
Box Plot

The box plot shows the distribution of countries' graduation ratios for girls and boys from 2014 to 2017. The Graduation ratio can be larger than 100% because UNESCO calculates the graduation ratio by the total number of students graduating divided by the number of students that theoretically should be graduating in that year; students may fall behind one year and graduate the next year – making the ratio larger than 100%.



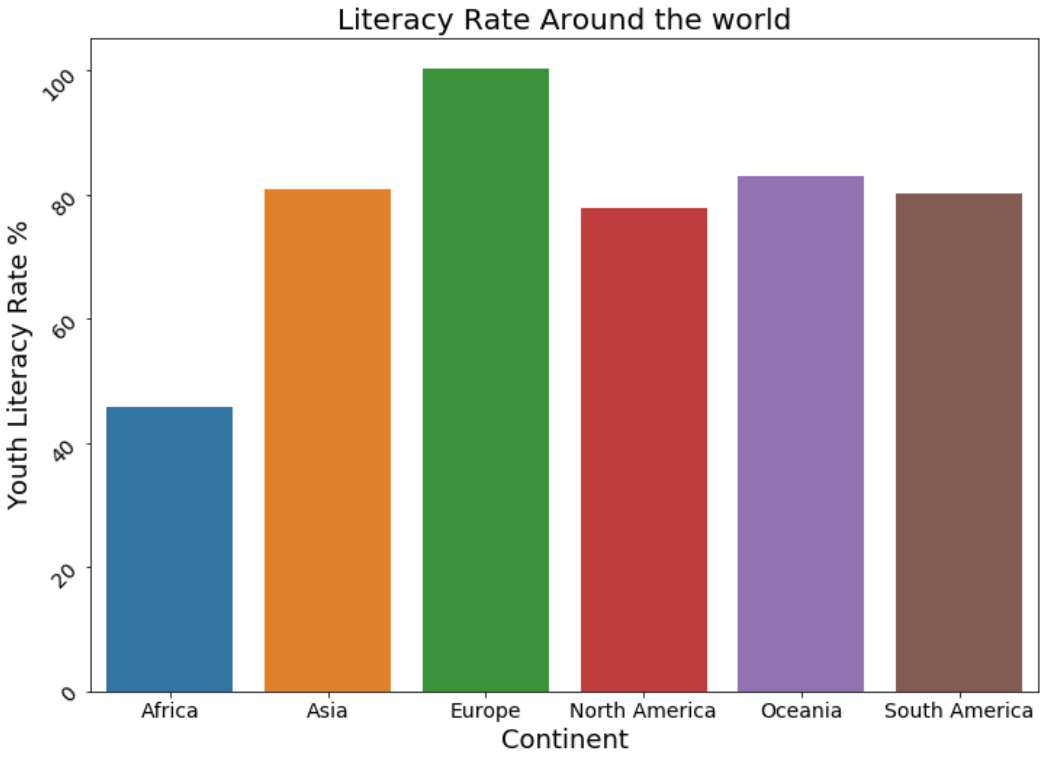
Tree Map

This tree map gives us a comparable and scaled-view of the average graduation rate for each continent. Europe has the largest graduation ratio and Africa has the smallest graduation ratio.



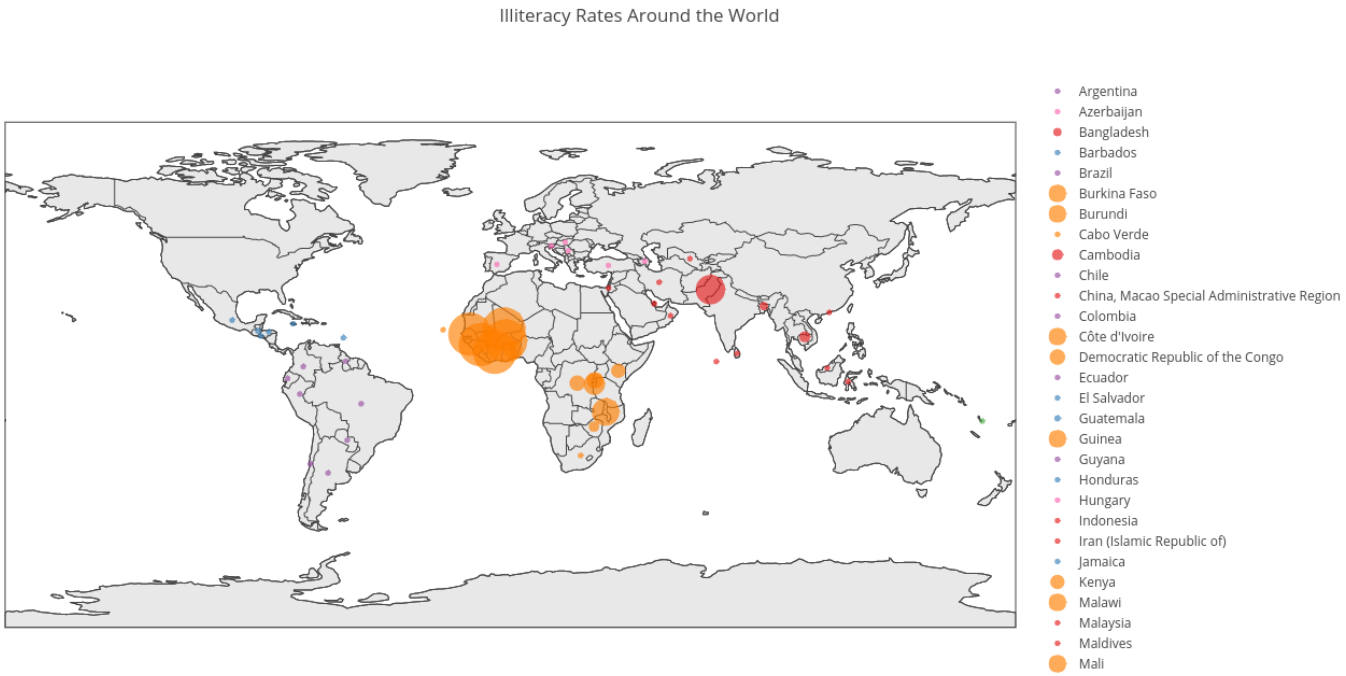
Bar Plot

This bar plot shows the percent of the youth population (anyone under 18) which are literate broken down by continent.



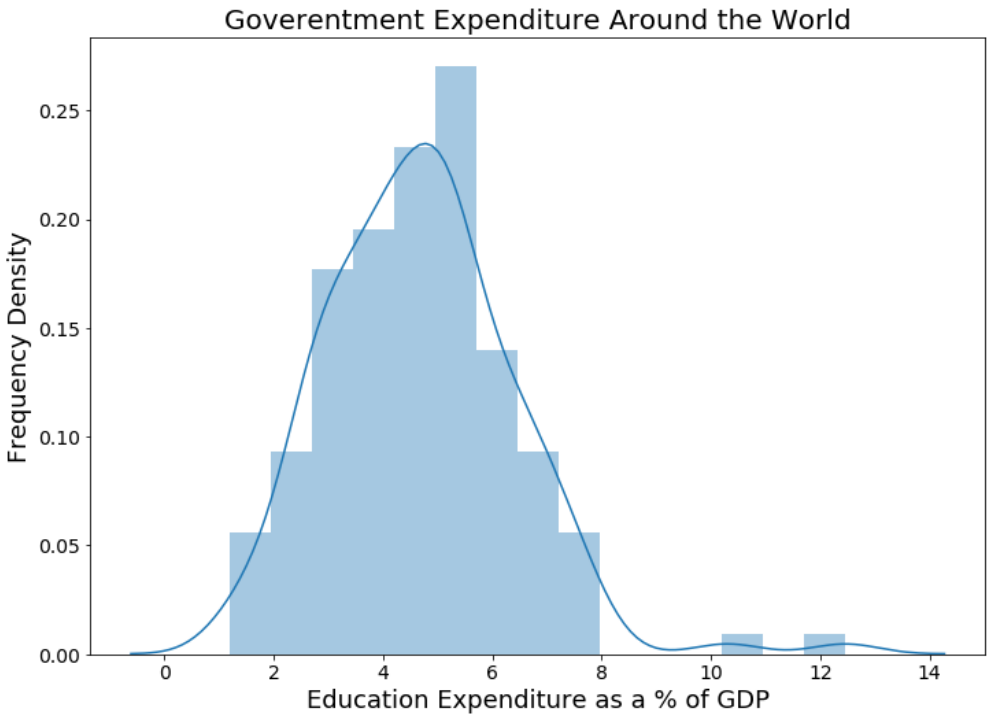
Bubble Map & Interactive Plot

This bubble map shows which countries have the largest illiteracy youth population. The Plotly version of the map is interactive and you can hover over the bubbles to see the country's name and their illiteracy rate.



Histogram

This histogram shows the distribution of how much countries' government spends on education as a % of their GDP.

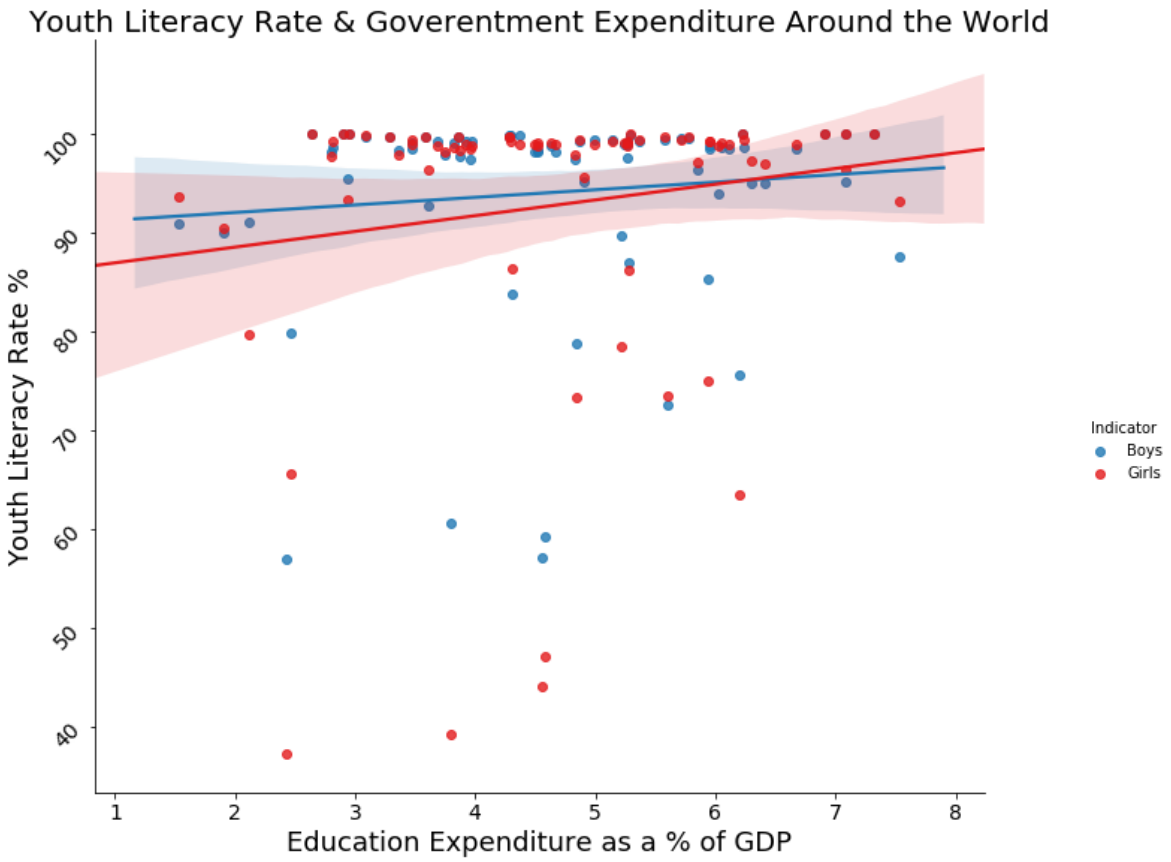


Story

Scatter Plot + Regression

So far, we have seen the inequality on access to education around the world and how education and literacy is around the world. For my final plot, I want to show the benefit a country may reap if their government increase their total investment in their education system. This scatter plot shows a country's youth literacy rate explained by their government's education expenditure (as a % of their GDP).

Poverty is often a primary reason for the discrepancy between the number of boys and girls enrolled in school; when families struggle financially, they tend to prioritize their boys when sending their children to school [2]. On top of the scatter plot is two linear regressions, one to explain boys' literacy rate and one to explain girls' literacy rate. Based on the two regression, we can see that both boys and girls benefit from an increase in education expenditure, but girls benefit more than boys – this is shown because the slope is larger on the girl regression. I think this difference in benefit is because when a country is thriving economically, they can spend more on education and families can afford to send both their boys and girls to school. This plot can be used to convince governments that are interested in increasing their literacy rate to invest more in their education system.



Conclusion

We have seen what the current world looks like in terms of access to education. We have also seen that the gap for girls vs boys entering primary school has shorten over the years. We have also seen where in the world exist issue with literacy and graduation rates exist the most. Lastly, we have seen that a country's education expenditure can lead to an increase in literacy rate in both boys and girls, while discussing the potential reasons why girls benefit more from an increase in education expenditure. Overall, I think the state of the world in terms of education has been improving over the years and as long as we have organizations such as UNESCO and She's the First actively pursuing to increase access to education and close the gap for girls' vs boys' education, we will achieve global education equality.

Appendix:

Below is the code by plot for all plots in this report.

Stack Area Graph

```
# data
x = list(total_entry_by_continent.reset_index().Time.unique())
y = list()
for c in total_entry_by_continent.Continent.unique():
    val = list(total_entry_by_continent[total_entry_by_continent.Continent == c].Total.values)
    y.append(val)

# plot
plt.stackplot(x,y, labels=list(total_entry_by_continent.Continent.unique()))
plt.legend(loc='best')

# size
plt.gcf().set_size_inches(11.7, 8.27)

# title
plt.title(f'Primary School Entry By Continent',fontsize=20)

#axis
plt.ylabel('Total Number of Students Entering Primary School \n (in 100millions)', fontsize=18)
plt.xlabel('Year', fontsize=18)
plt.tick_params(axis='y', labelrotation=45)
plt.tick_params(labelsize=14)

#legend
plt.legend(title='Continent')

# plt.show()
plt.savefig('StackArea')
```

Heat Map

```
# data to plot
matrix = entry_by_continent.reset_index().pivot(index='Continent', columns= 'Time', values='Dif%')

# plot size
plt.figure(figsize=(15,10))

# title
plt.title('% Difference in Boys vs Girls Primary Schools Enrollment',
          fontsize=20)

# data
sns.heatmap(matrix, cmap=sns.color_palette("Blues_r"),
            cbar_kws=dict(use_gridspec=False,location="right", label='% difference in boys vs. girls'))

# color bar
cax = plt.gcf().axes[-1]
cax.tick_params(labelsize=14)
cax.yaxis.label.set_size(18)

# axis
plt.xlabel('Year', fontsize=18)
plt.ylabel('Continent', fontsize=18)
plt.tick_params(axis='y', labelrotation=45)
plt.tick_params(labelsize=14)

# plt.show()
plt.savefig('HeatMap')
```

Choropleth Map

```
data = [go.Choropleth(
    locations = entry_by_contry['LOCATION'],
    z = entry_by_contry['Dif%'],
    text = entry_by_contry['Country'],
    autocolorscale = True,
    reversescale = True,
    marker = go.choropleth.Marker(
        line = go.choropleth.marker.Line(
            color = 'rgb(180,180,180)',
            width = 0.5
        )),
    colorbar = go.choropleth.ColorBar(
        tickprefix = '%',
        title = '% Difference of Boys vs Girls'),
)]

layout = go.Layout(
    title = go.layout.Title(
        text = 'Difference in Primary School Entry for Boys vs. Girls'
    ),
    geo = go.layout.Geo(
        showframe = False,
        showcoastlines = True,
        projection = go.layout.geo.Projection(
            type = 'equirectangular'
        )
    ),
)

fig = go.Figure(data = data, layout = layout)
py.iplot(fig, filename = 'ChoroplethMap')
```

Connection Map

```
# layout of plot
layout = go.Layout(
    title = go.layout.Title(text = "She's the First Scholars Across the World"),
    showlegend = False,
    geo = go.layout.Geo(
          resolution = 110,
          showcountries = True,
          landcolor = "#e8e8e8",
          showland = True,
          coastlinewidth = 1,
        # plot default view
            lataxis = go.layout.geo.Lataxis(
            range = [-15, 75]),
            lonaxis = go.layout.geo.Lonaxis(
            range = [-100, 90])))

# objects to be ploted
countries = list()
connections = list()
for scholar, location in scholars.items():
    lat, lon = location
    connect = [go.Scattergeo(
                lat = [40.7127, lat],
                lon = [-74.0059, lon],
                mode = 'lines',
                line = go.scattergeo.Line(
                       width = 3,
                       color = '#a6cee3'))]
    country = [go.Scattergeo(
                lat = [lat],
                lon = [lon],
                # marker
                mode = 'markers+text',
                marker = go.scattergeo.Marker(
                         size = 6,
                         color = '#1f78b4'),
                # annotation
                text = scholar,
                textposition='bottom center',
                textfont = dict(size=15, color = 'black'))]

    connections += connect
    countries += country

nyc = [go.Scattergeo(
           lat = [40.7127],
           lon = [-74.0059],
           # marker
           mode = 'markers+text',
           marker = go.scattergeo.Marker(
                    size = 6,
                    color = '#ffd92f'),
           # annotation
           text = "STF<br>Headquarter",
           textposition='top center',
           textfont = dict(size=15, color = 'black'))]

# plot
fig = go.Figure(data = connections + countries + nyc, layout = layout)
py.iplot(fig, filename = 'ConnectionMap')
```

Box Plot

```
# size
plt.figure(figsize=(15,10))

# title
plt.title('Graduation Ratio Over Time',
          fontsize=20)

# data
sns.boxplot(x="Time", y="Value",
            hue="Indicator", palette=["#1f78b4", '#e41a1c'],
            data=grad_by_country)

#axis
plt.xlabel('Year', fontsize=18)
plt.ylabel('Graduation Ratio (%)', fontsize=18)
plt.tick_params(axis='y', labelrotation=45)
plt.tick_params(labelsize=14)

#legend
plt.legend(title='Group')

# plt.show()
plt.savefig('BoxPlot')
```

Tree Map

```
squarify.plot(sizes=grad_ratio_by_country.Value.values,
              label=grad_ratio_by_country.Continent.values,
              color=["red","green","blue", "grey", "orange", "purple"], alpha=.4 )

# size
plt.gcf().set_size_inches(11.7, 8.27)

# title
plt.title(f'Graduation Rates By Continent',fontsize=20)

#axis
plt.axis('off')

# plt.show()
plt.savefig('Treemap')
```

Bar Plot

```
sns.barplot(x="Continent", y="Value", data=mean_grad_ratio.reset_index())

# size
plt.gcf().set_size_inches(11.7, 8.27)

# title
plt.title(f'Literacy Rate Around the world',fontsize=20)

#axis
plt.ylabel('Youth Literacy Rate %', fontsize=18)
plt.xlabel('Continent', fontsize=18)
plt.tick_params(axis='y', labelrotation=45)
plt.tick_params(labelsize=14)

# plt.show()
plt.savefig('BarPlot')
```

Bubble Plot & Interactive Plot

```
# color pallet
color = {'Asia':'#e41a1c', 'North America':'#377eb8',
        'Europe':'#FF69B4', 'South America':'#984ea3',
        'Africa':'#ff7f00', 'Oceania':'#4daf4a'}
```

```
# layout of plot
layout = go.Layout(
        title = go.layout.Title(
            text = 'Illiteracy Rates Around the World'
        ),
        showlegend = True,
        geo = go.layout.Geo(
          resolution = 110,
          showcountries = True,
          landcolor = "#e8e8e8",
          showland = True,
          coastlinewidth = 1
    ))

# objects to be ploted
countries = list()
for c, row in literacy_rate_by_country.iterrows():
    lat, lon = get_lat_lon(c[0])
    if lat is None:
        continue
    val = (100 - row['Value'])
    if val < 5:  # too samll, make a little bigger
        val = 5
    elif val > 30:  # too large, make a little smaller
        val = 40
    country = go.Scattergeo(
                lat = [lat],
                lon = [lon],
                text = 100 - row['Value'],
                mode = 'markers',
                marker = go.scattergeo.Marker(
                         size = val,
                        opacity = 0.65,
                         color = color[c[2]],
                sizemode = 'area'),
                name = c[0])

    countries.append(country)

# plot
fig = go.Figure(data=countries, layout=layout)
py.iplot(fig, filename='BubbleMap')
```

Histogram

```
sns.distplot(mean_education_expend['Expenditure%'].dropna().values)

# size
plt.gcf().set_size_inches(11.7, 8.27)

# title
plt.title(f'Goverentment Expenditure Around the World',fontsize=20)

#axis
plt.xlabel('Education Expenditure as a % of GDP', fontsize=18)
plt.ylabel('Frequency Density', fontsize=18)
plt.tick_params(labelsize=14)

# plt.show()
plt.savefig('Histogram')
```

Scatter Plot + Regression

```
# data
sns.lmplot(y="Value", x="Expenditure%", data=joined,
           hue='Indicator', palette=["#1f78b4", '#e41a1c'],
           truncate=False)

# size
plt.gcf().set_size_inches(11.7, 8.27)

# title
plt.title(f'Youth Literacy Rate & Goverentment Expenditure Around the World',fontsize=20)

#axis
plt.ylabel('Youth Literacy Rate %', fontsize=18)
plt.xlabel('Education Expenditure as a % of GDP', fontsize=18)
plt.tick_params(axis='y', labelrotation=45)
plt.tick_params(labelsize=14)

# plt.show()
plt.savefig('ScatterPlot')
```

GitHub

https://github.com/vasuarez/Data_Visualization

Citations

Sources:

[1]: http://uis.unesco.org/en/news/263-million-children-and-youth-are-out-school

[2]: https://borgenproject.org/top-10-facts-about-girls-education-in-developing-countries/

Datasets:

All datasets were exported from UNESCO's open data API: http://data.uis.unesco.org

The datasets used were:

- New entrants to primary education
- Gross graduation ratio
- Literacy rate
- Government expenditure on education as a percentage of GDP