Our Changing World Over Time

Pranav Bhagat, Cal Poly, San Luis Obispo

Abstract

Over time, each country has developed in different ways and at different paces from one another. There have been various contributing factors which have led to these developments. One such factor is technology which has had a profound impact in our society. Technologies include artificial intelligence and the Internet which has positively affected the world. Although these advancements have improved our world, the speed at which each region has grown varies from each other.

Introduction/Motivation

In this paper, we'll look into the development of different regions over time starting around the 1960's. The 1960's was when relevant data appeared and could be used. Also, this would be a good time period as a starting point because it was approximately twenty years after the start of the The Electronic Age. Therefore, more countries would have that time to advance and implement the technology, decreasing variability and bias in the data. In order to investigate further, we will analyze three different variables which include GDP, literacy rates, and life expectancy. GDP can be used to determine the growth of each country economically, literacy rates can help show the improvements in education, and life expectancy can be used to determine the improvements of health. I am curious to find out how these three variables impact the growth of different areas in the world and to observe various trends over time.

GDP Analysis

GDP stands for Growth Domestic Product which is the total value of goods and services produced in a country. GDP is measured over specific time frames, such as a quarter or a year. GDP is used as an economic indicator to show the economic state of a country. A higher GDP rate means that the country is prospering whereas a low GDP means that the country is struggling economically. A great way to compare these regions is to use the confidence intervals and data from ANOVA to come to our conclusions. The data is as follows:

Confidence levels for mean gdp by region

The MEANS Procedure

Analysis Variable : gdp			
four_regions	N Obs	Lower 95% CL for Mean	Upper 95% CL for Mean
africa	2581	1825.40	2025.87
americas	1919	7383.80	8198.54
asia	2324	9109.97	10381.50
europe	1910	23668.90	25486.51

Table of 95%CI for mean of Continents

Clearly, if we simply look at the mean of the region of the highest GDP, we can conclude that Europe has the highest average GDP whereas Africa has the lowest average GDP. More specifically, we can compare the average GDP of each continent which is as follows:

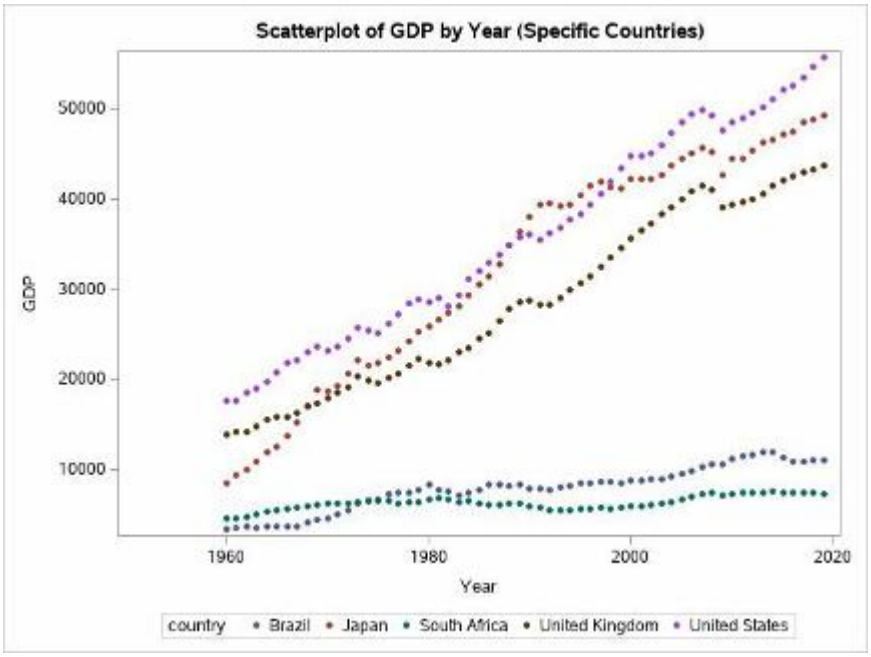
Comparisons significant at the 0.05 level are indicated by ***.				
four_regions Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
europe - asia	14832.0	13782.6	15881.4	***
europe - americas	16786.5	15688.3	17884.8	***
europe - africa	22652.1	21626.5	23677.6	***
asia - europe	-14832.0	-15881.4	-13782.6	***
asia - americas	1954.6	906.5	3002.6	***
asia - africa	7820.1	6848.4	8791.8	***
americas - europe	-16786.5	-17884.8	-15688.3	***
americas - asia	-1954.6	-3002.6	-906.5	***
americas - africa	5865.5	4841.3	6889.7	***
africa - europe	-22652.1	-23677.6	-21626.5	***
africa - asia	-7820.1	-8791.8	-6848.4	***
africa - americas	-5865.5	-6889.7	-4841.3	***

Table which compares confidence means by region/continent

In the table above, we compared the average GDP of each country to each other using Tukey's test. The dots on the right side indicates that it's statistically significant at the .05 alpha level which means that the variation is not due to random chance. We also know it is significant because 0 is not in any of intervals which is good since it makes the data more reliable. Using both the previous charts, the highest GDP in order is Europe, Asia, the Americas, and lastly Africa.

However, are there any specific outlying countries in each continent? For instance, it

could be true that a country in Africa is doing very well economically compared to other top producing GDP countries in other continents. Perhaps outside factors, such as other country's slow economic development in the same continent, could be degrading and underrepresenting the GDP of that specific country in the average data. In order to limit misleading data, we should look at the trends of the top producing country in each continent:

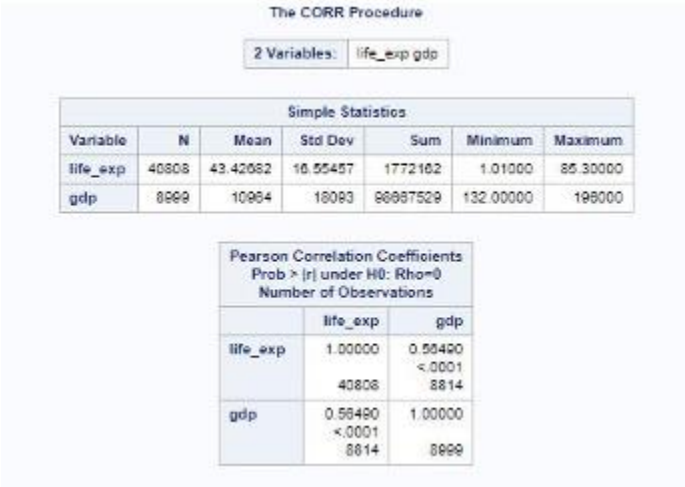


In the scatterplot above, I took the top producing GDP countries in each continent to see if it follows the overall average GDP of each continent which was Europe, then Asia, the Americas, and lastly Africa. Starting from the bottom, we can see that South Africa is at the bottom which is accurate with the overall average GDP since Africa was the lowest. As for the Americas, we can see that the United States has the highest GDP. However, Brazil is near the bottom so that lowered the position of the Americas and they were third in average GDP. Asia is second in overall average GDP; however, the United Kingdom (from Europe) has a higher peak GDP compared to Japan (from Asia) which doesn't follow the order of the average overall GDP per continent. This data and scatter plot helps us understand that a specific country's GDP should not be predicted by the average GDP of its continent. A better alternative would be to look at the trend of a specific country and compare it to others.

A possible explanation for the low GDP countries in the chart could be due to lack of technology in those places. This would hurt their efficiency and therefore their GDP compared to more advanced places like Europe and the Americas.

Life Expectancy

An important factor in looking at the quality of life is by observing the life expectancy in that region. This data can help us correlate the advancements of healthcare in each region. The longer the life expectancy, the more advanced the region should be. A good way to assess life expectancy is to correlate our previous variable, GDP, and life expectancy. I predict that the life expectancy will be longer if the GDP is higher and for the data to be significant. This is because more advanced regions would have more advanced medicine and healthcare. The Proc Corr method is a great way to see this. The results are as follows:



The CORR Procedure

2 Variables: life_exp gdp

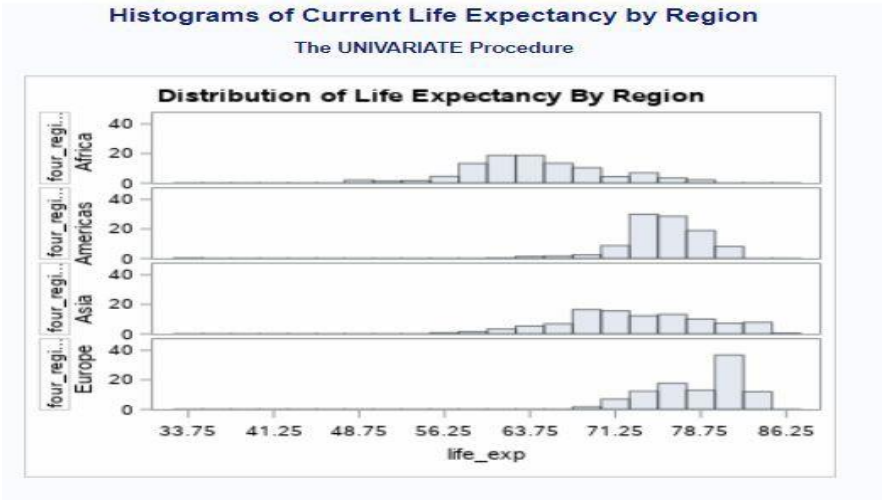
Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
life_exp	40808	43.42682	16.55457	1772162	1.01000	85.30000
gdp	8999	10964	18093	98687529	132.00000	196000

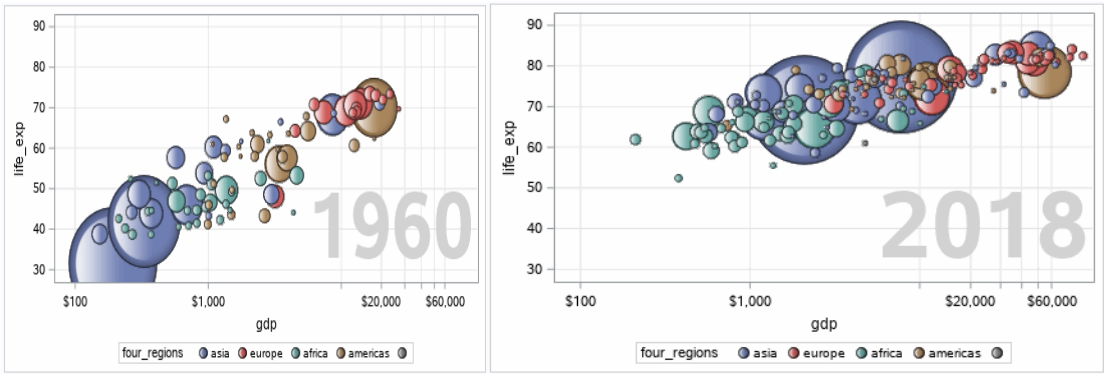
Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	life_exp	gdp
life_exp	1.00000 40808	0.56490 <.0001 8814
gdp	0.56490 <.0001 8814	1.00000 8999

Based on the Proc Corr results above, we can tell that there is a strong correlation between life expectancy and GDP. The correlation value is 0.56490 which is strong. Furthermore, my prediction was correct as the p-value is very small (less than 0.0001) which means that the data is significant. Now that we know that they correlate, let's see how the life expectancy has changed over the year by continent:



If we look at Africa, it has a large spread and has a lower life expectancy compared to the other continents. The second lowest life expectancy is Asia which is also spread out but looks to have a higher average. The second highest life expectancy is in America and the highest is Europe. The data from the histogram correlates with the data from the previous graph. Thus, we can confidently conclude that the highest life expectancy is Europe, followed by the Americas, then Asia, and lastly Africa. Now, let's take a look on how GDP and life expectancy are related by region through the GIF I made below:

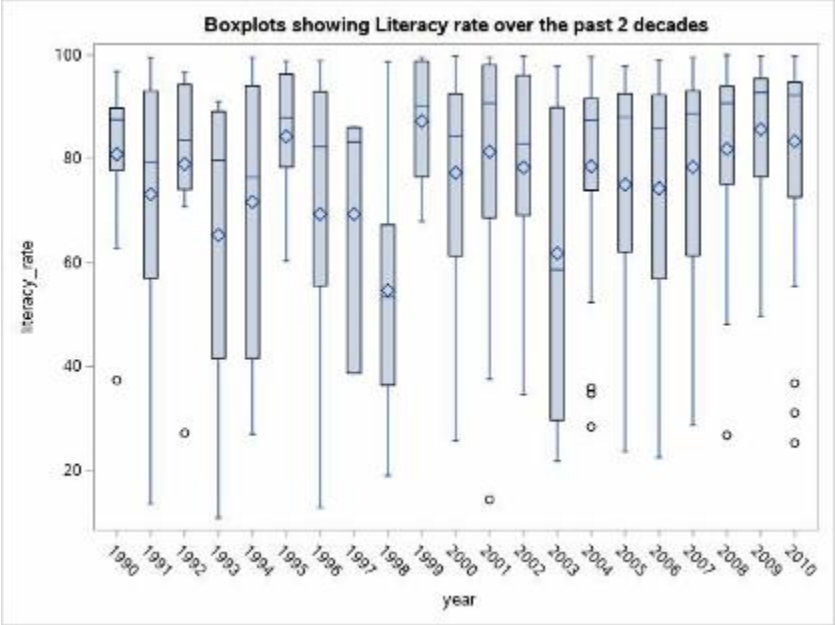


The size of the bubble indicates the size of the population of each region compared to each other. From this graph, we can tell that life expectancy has increased over time as GDP increases. The graph was higher in 2018 compared to 1960.

Literacy Rates

The last factor we should look at is literacy rate. After research, studies have shown that a better educated society correlates better with a stronger society (see references). Therefore, I thought it would be interesting to compare the literacy rates of different

regions across the globe. Let's first observe a macro perspective of the data throughout the whole world:

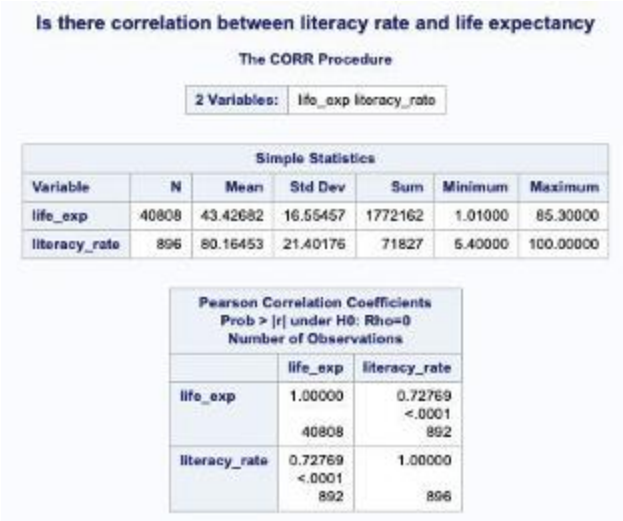


From these boxplots, we can conclude that toward recent years, there has been a more consistent trend with the maximum value and third quartile increasing for the most part. Moreover, the mean is either at or lower than the median which indicates that there are a few low outliers meaning that certain countries have a low literacy rate. The minimum value somewhat increases over the years which is good for the literacy rate.

Now that we have a global view of the literacy rate, let's deep diver into each continent from 2000-2010. A line plot of the data is as follows (next page):

Historically, we can tell that Africa has and still does have the lowest literacy rate. Europe and America have high literacy rates with Europe being slightly ahead. Asia is below them but has had a great increase of literacy starting in around 2012. They are very close to the Americas in terms of literacy rate.

Another interesting way we can determine whether or not education and literacy rates are impactful is to correlate it with life expectancy. I believe that a strong and positive literacy rate would lead to a higher life expectancy rate. This is because I predict that an education will decrease poverty rates and lead to an increase in jobs for people. We can use the proc corr function as we did earlier in the paper to determine whether there is a strong correlation or not. The data is as follows:



Is there correlation between literacy rate and life expectancy

The CORR Procedure

2 Variables:	life_exp literacy_rate

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
life_exp	40808	43.42682	16.55457	1772162	1.01000	85.30000
literacy_rate	896	80.16453	21.40176	71827	5.40000	100.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	life_exp	literacy_rate
life_exp	1.00000 40808	0.72769 <.0001 892
literacy_rate	0.72769 <.0001 892	1.00000 896

From the data, we can tell that there is a strong and positive correlation between life expectancy and literacy rate. We have a high value of 0.728 and a low p-value which is good since it means that the data did not come from random chance alone. Thus, my prediction was correct and literacy rate and life expectancy are correlated with each other. Causation would be hard to conclude since the sampling method is not included and there could be possible confounding variables affecting the data.

Conclusion:

In conclusion, we can say that Europe has the best overall quality of life. It has the highest GDP which means it excels economically and has the highest life expectancy showing that it is advanced medically and safe to live in. It also has a superior education compared to the rest of the continents. To further ensure our conclusion is solid, we compared the variables with one another. In order to limit any lurking variables or confounding variables, we compared the GDP and life expectancy in each continent and did the same with life expectancy and literacy rates. The results were consistent with what we expected. Asia and the Americas have also improved their GDP and literacy rates which is a positive sign. After analyzing these results, the order of the best quality life is Europe, followed by the Americas, then Asia, and lastly Africa.

References:

Gapminder.org

https://www.emerald.com/insight/content/doi/10.1108/IJSHE-07-2017-0114/full/html

Appendix

```
*First I created a library statement to access the dataset;

libname midterm "/folders/myfolders/My_data";


 /* Data step for observations including the gdp and life expectancy. */
 data countries;
 set midterm.gapminder_data;

 /* GDP Analysis*/
 /* Doing confidence intervals for the mean of gdp by region */
 title "Confidence levels for mean gdp by region";
 proc means data = countries clm;
            class regions;
 run;

 /*Doing anova and multiple comparisons  */
 proc anova data = countries;
 class four_regions;
 model gdp = regions;
 means regions / hovtest tukey;
 quit;

  /*Scatterplot of advanced countries GDP by continent   */
  proc sgplot data = specific_countries;
```

```
 where year >= 1950;

                    scatter y=gdp x=year / group=country
                    markerattrs=(symbol=CircleFilled size=5)
                    colormodel=TwoColorRamp;
 xaxis label="Year";

 yaxis label="GDP";

 title "Scatterplot of GDP by Year (Specific Countries)";

 run;

/*Life Expectancy */
proc corr data=midterm.gapminder_data;
var life_exp gdp;
run;
*Generate the correlation and coefficient between Life Expectancy and GDP;
proc sgplot data=midterm.gapminder_data;
scatter x=GDP y=life_exp;
run;

 /* Histogram using proc univariate */
 Title "Histograms of Current Life Expectancy by Region";
 proc univariate data=current;
   class four_regions;
   var life_exp;
```

```
  histogram life_exp / nrows=4 odstitle="Distribution of Life Expectancy By Region"
vaxislabel= "Europe  Asia     Americas         Africa";
  ods select histogram;
run;

/* GIF Code */
%macro GapMinder(start=, end=, incr=); %do year=&start %to &end %by &incr;
proc sgplot data=midterm.gapminder_data
noautolegend;
format gdp dollar12.0;
where year=&year;
format country $country.;
bubble x=gdp y=life_exp
size=population / group=four_regions
  dataskin=gloss
 bradiusmin=2 bradiusmax=50;
    keylegend;

inset "&year" / position=bottomright textattrs=(size=60 weight=bold color=lightgray);
 xaxis type=log offsetmin=0.05 offsetmax=0.08 logstyle=linear
 grid min=100 max=60000;
 yaxis offsetmin=0.05 offsetmax=0.05 grid min=30 max=90;
run;
```

```
 %end;
 %mend GapMinder;
  options papersize=('5 in', '3 in')
 printerpath=gif
 animation=start animduration=0.1 animloop=yes noanimoverlay;
 ods printer file= "/folders/myfolders/sasuser.v94/MidtermProject.gif";
 ods graphics / width=5in height=3in imagefmt=GIF;


 %GapMinder(start=1960, end=2020, incr=2);
 options printerpath=gif animation=stop;
 ods printer close;


/*Literacy Rate */

*Graph of boxplots of literacy rates;
  proc sgplot data = midterm.gapminder_data;
  title "Boxplots showing Literacy rate over the past 2 decades";

  where year >= 1990 and year <= 2010;
          vbox literacy_rate / category = year;
  run;

*The line plot of literacy rates of each region from 1960-2020;
data recent;
set midterm.gapminder_data;
if year >= 1960 and year <= 2020;
```

```
run;

proc sgplot data = recent;
vline year / response=literacy_rate group=four_regions stat=mean;
yaxis label = "Avg Region Literacy Rate";
xaxis value = (1960 to 2020 by 4);
run;

 *The numeric correlation between life expectancy and literacy rate;
   proc corr data = midterm.gapminder_data;
   title "Is there correlation between literacy rate and life expectancy";
   var life_exp literacy_rate;
  run;
```