Fordham University - Gabelli School of Business
Business Analytics for Managers
Prof. RP Raghupathi
BYGB 7975
Spring 2021

PROJECT IMPLEMENTATION- FINAL REPORT

Group Name:
JEAN GREY

Group Member Names:
Tahsin Rakib Himi
Kristel Kalm
Nicholas Sansone
Olha Sverdel

Figure #	Hypothesis	Question	Analytic type	Variable(s)	Chart	Conclusion
1	The larger the GDP, the higher the Venture Capital availability	Is there a positive association between the countries GDP and venture capital availability?	Descriptive	Countries (GDP), Year, Venture Capital Availability, Country (grouped by continent)	Bubble Chart	The hypothesis was rejected
2	The higher the Availability of latest technologies the higher the Innovation score	Is there a positive association between availability of latest technology and the innovation score?	Predictive	Availability of latest technologies, Innovation Score	Scatter Plot	The hypothesis was proven to be correct
3	There is a positive correlation between the Percentage of patents approved and Year	Has the percentage of patents approved increased over time?	Descriptive	Year, Percentage of patents approved	Bar chart	The hypothesis was proven to be partly correct- there was a weak positive correlation
4	There is a positive correlation between Countries GDP and the Gross domestic expenditure on R&D as a percentage of GDP	Is there a positive association between countries GDP and the Gross domestic expenditure on R&D as a percentage of GDP?	Descriptive	Country (grouped by continent), Gross domestic expenditure on R&D as a percentage of GDP	Line chart	The hypothesis was proven to be correct
5	There is a positive correlation between Venture Capital availability, Availability of latest technologies and Year	Is there a positive association between venture capital availability and time?	Descriptive	Year, Venture Capital availability, Availability of latest technologies	Dual line chart	The hypothesis was rejected

6	There is a positive correlation between availability of the latest technology and innovation efficiency ratio	How is the availability of the latest technology related to innovation efficiency ratio?	Predictive	Availability of latest technology, Innovation efficiency ratio, Year, Country (grouped by continent)	Scatter plot with trend and forecasting lines	The hypothesis originally was not supported, but as the years increased the data supports the hypothesis
7	The higher the Governance Indicators, the higher the Innovation Score	Is there a positive association between Governance Indicators and Innovation Score?	Descriptive	Governance Indicators: (Control of Corruption, Government Effectiveness, Political Stability and absence of Violence/Terrorism, Regulation Quality, Rule of Law, and Efficiency of Legal Framework in Settling disputes.), Innovation Score	Combination of Area Graph and Scatter Plot	The hypothesis was proven to be right
8	The higher the Intellectual Property Protection the higher the spending on Research & Development	Is there a positive association between Intellectual Property Protection and spending on Research & Development as a % of GDP?	Descriptive	Intellectual Property Protection Index, Research & Development as a percentage of GDP	Combination of Geographical Heat Map and Bubble Chart	The hypothesis was proven to be right
9	The higher the percentage of patents approved the higher the innovation score of the country	Is there a positive association between the percentage of patents approved and the innovation score?	Descriptive	Percentage of patents approved, Innovation Score, Country (grouped by geographical area)	Bubble chart	The hypothesis was not supported

10	Innovation score and Innovation Efficiency ratio are positively related to the development status of the country (developed or developing)	Is there a positive association between Innovation and the development status of the country?	Descriptive	Innovation Score, Innovation Efficiency, Country (grouped by the development status)	Geographical map combination with the bubble heat map	The hypothesis was proven to be right
11	The higher the country's availability of latest technology the higher the percentage of patents approved	Is there a positive correlation between availability of latest technology and the percentage of patents approved?	Descriptive	Availability of Latest Technology, Percentage of Patents approved, Year	Area graph and line chart	The hypothesis was not supported as there was not much of a correlation found
12	There is a positive correlation between the efficiency of legal system in settling disputes and the percentage of patents approved	How does a country's legal system efficiency relate to their patent approval efficiency?	Descriptive	Efficiency of the legal system in settling disputes, Percentage of patents approved, Country (grouped by geographical area)	Column chart	A positive correlation was found, therefore the hypothesis was supported
13	There is a positive correlation between the Percentage of patents approved and a Country's GDP	How is GDP for a country associated with their patent approval efficiency?	Descriptive	Percentage of patents approved, Countries (GDP)	Area Chart	The hypothesis was rejected
14	There is a positive correlation between Availability of latest technology and the year	Is there a general increase in Availability of latest technologies as the years increase?	Predictive	Availability of latest technology, Year, Country (grouped by geographical area)	Waterfall chart	The hypothesis was rejected as a very strong negative correlation was found

15	There is a positive correlation between Percentage of patents approved and Innovation Efficiency Ratio	How does a country's patent approval efficiency relate to their innovation efficiency ratio?	Predictive	Percentage of patents approved, Innovation Efficiency Ratio, Country (grouped by geographical area)	Scatter plot with Trend Lines	The hypothesis was not supported as there was no correlation found
16	There is a correlation between venture capital availability and a nation's regulatory quality score	How does a nation's ability to regulate policy impact the availability of venture capital for innovation?	Predictive	Regulatory Quality: Estimate, Availability of latest technologies	Bar Chart	There exists a positive correlation between the input variables
17	There is a positive correlation between the control of corruption and innovation score	How does a country score in terms of innovation, based on how they control corruption?	Descriptive	Control of Corruption: Estimate, Innovation Score	Line Chart	Based on figure 17 it has a positive slope.
18	The higher the regulatory quality of a nation, the higher the nation's innovation score	Is there a positive correlation between the regulatory quality and innovation score?	Pedictive	Regulatory Quality: Estimate, Innovation Score	Column Chart	The hypothesis was supported
19	The lower the political stability the lower the innovation score	Is there a negative correlation between lower political stability and higher innovation score?	Predictive	Political Stability and Absence of Violence/Terrorism: Estimate, Innovation Score	Line Chart	There is a positive correlation between the variables
20	There is a positive correlation between Efficiency of the legal system in settling disputes and innovation efficiency	Is there a positive correlation between Efficiency of the legal system in settling disputes and percentage of patent approved?	Predictive	Efficiency of the legal system in settling disputes and innovation ratio	Scatter Plot	Based on the chart the hypothesis holds true

Bubble Chart of the countries average GDP and average Venture Capital Availability for each country

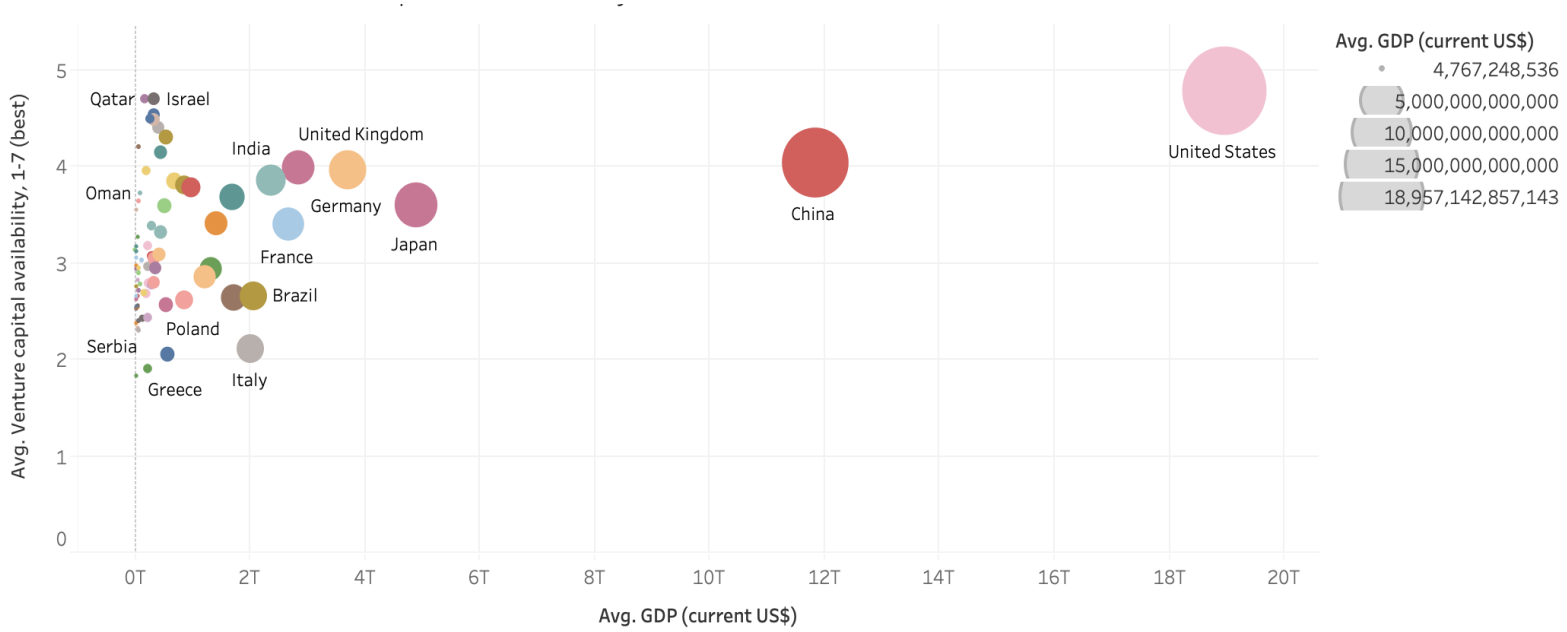

Figure 1. Bubble Chart of the average GDP and average Venture Capital Availability for each country - Kristel Kalm

The bubble chart describes the average countries GDP and the average Venture Capital availability. The bubble size describes the average GDP, where the size grows as the GDP increases. The color of the bubble stands for each of the countries included.

The chart shows that countries like the United States and China, which have the largest GDP, also have a high venture capital availability; however, it also shows that smaller countries with smaller GDP can have the same venture capital availability.

In conclusion, the GDP alone does not have enough evidence to define the venture capital availability of a country.

Scatterplot between Availability of Latest Technologies and the Average Innovation Score

Figure 2. Scatterplot between Availability of Latest Technologies and the Average Innovation Score - Kristel Kalm

The scatter plot chart describes the relationship between the availability of latest technologies and the average innovation score. The circle size describes the availability of latest technologies, where the size grows as the availability of the latest technologies increases. The shading represents the average innovation score, where dark green represents lower average innovation score and dark blue represent higher average innovation score.

The chart shows that there is a positive correlation between the availability of the latest technologies and the average innovation score.

In conclusion, when the availability of the latest technology increases, the average innovation score increases as well.

Bar chart between Average Percentage of patents approved and Year, which ranges from 2013 to 2019

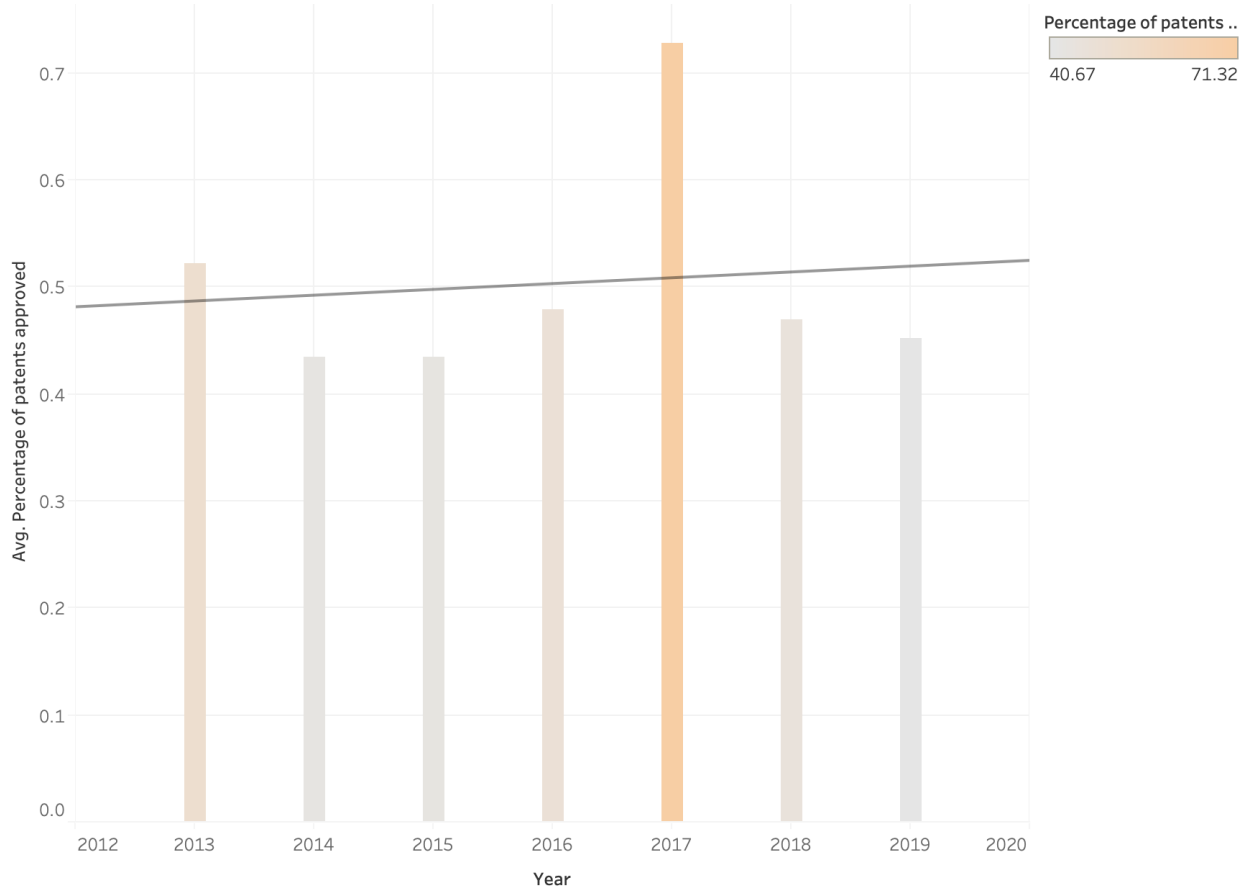

Figure 3. Bar chart between Average Percentage of patents approved and Year, which ranges from 2013 to 2019 - Kristel Kalm

The bar chart describes the average percentage of patents approved over time. Shading shows the average of percentage of patents approved, where the darker the color portrays a higher average percentage of patents approved. The view is filtered on year, which ranges from 2013 to 2019.

The chart shows that there is a weak positive correlation between the average percentage of patents approved over time.

In conclusion, there has not been an constant increase in average patents approved over time, but on average it has been a slight increase overall even if the past couple of years it shows a decrease in number of patents approved.

Average Availability of latest technologies plotted against average Research & Development as percentage of GDP and average Innovation Score

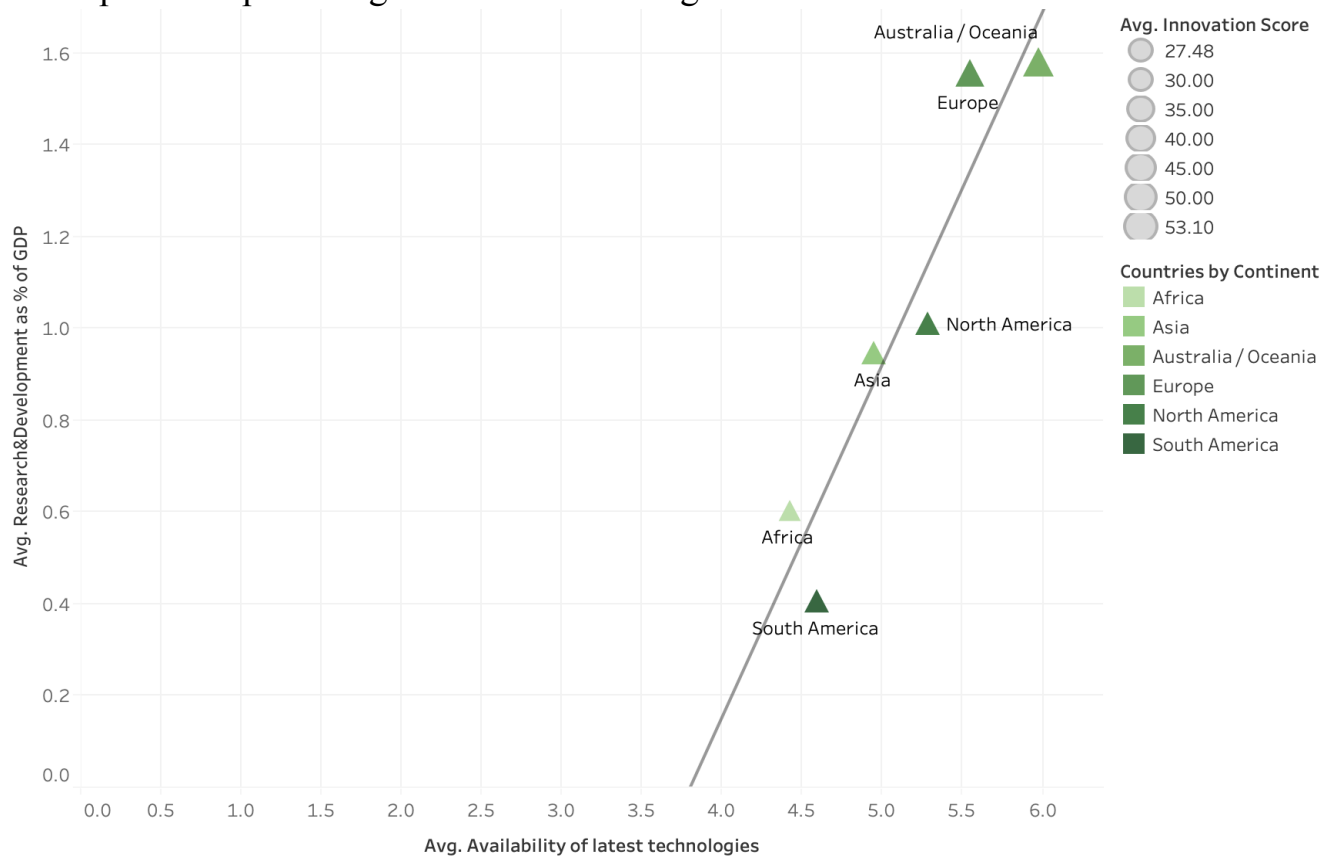

Figure 4. Average Availability of latest technologies plotted against average Research & Development as percentage of GDP - Kristel Kalm

The line chart describes the average availability of latest technologies vs. research & development as % of GDP. Shading shows details about countries by continent. Size shows average of innovation score. The marks are labeled by the countries by continent.

The chart shows that there is a positive correlation between the average availability of latest technologies vs. research & development as % of GDP.

In conclusion, when the average research & development as % of GDP increases, the average availability of the latest technology increases as well. The chart also shows that the higher both of these two variables the higher the innovation score. That is shown on continents like Europe and Australia/Oceania, which have the highest average availability of latest technologies and average research & development as % of GDP and also shows the highest innovation score which means that they are all associated with each other.

Line chart between Average Research & Development as % of GDP, Average Venture Capital Availability, Innovation Score and Year

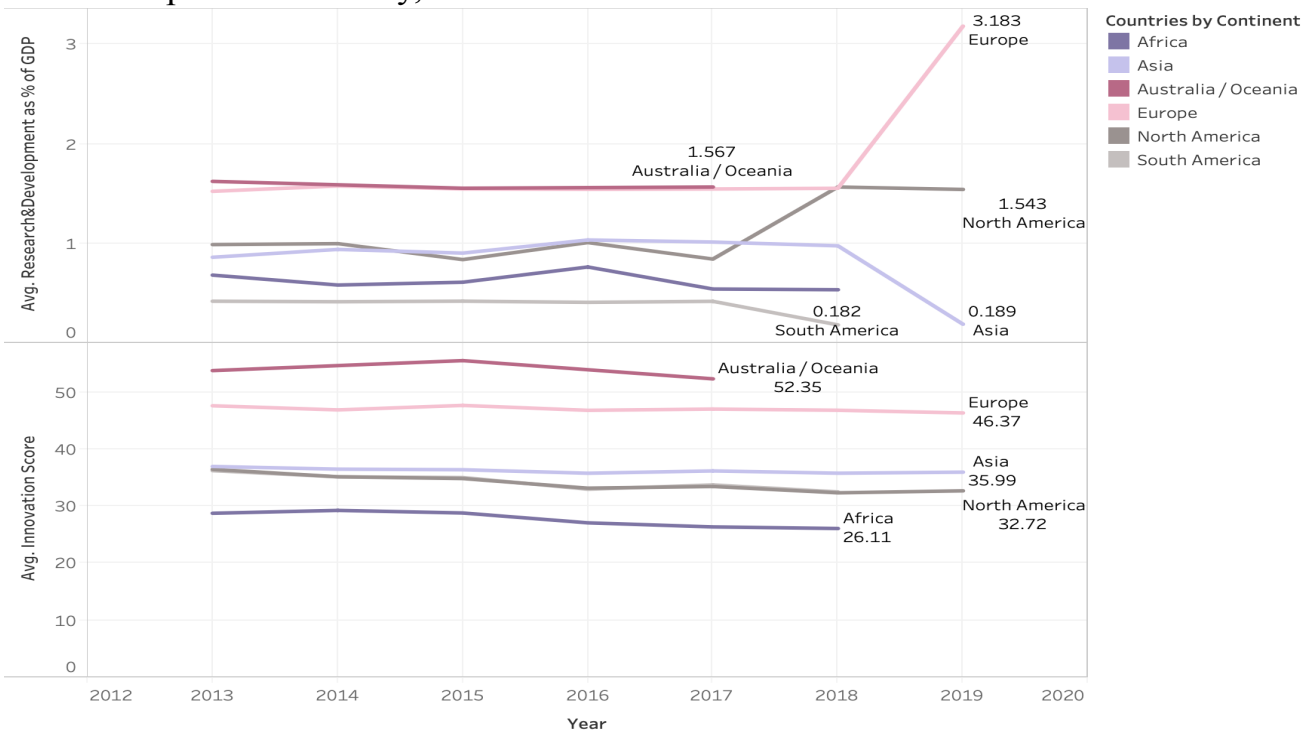

Figure 5. Line chart between Avg Research & Development as % of GDP, Average Innovation score, Country by Continent and Year, which ranges from 2013 to 2019 - Kristel Kalm

The line chart describes the average availability of latest technologies vs. research & development as % of GDP and the innovation score of each continent. Color shows details about countries by continent. Size shows average of innovation score. The marks are labeled by the average of research & development as % of GDP and countries by continent.

The chart shows that the average innovation score by continent has not had any drastic changes, however the average research & development as % of GDP did show us considerable changes between years 2017 and 2019.

In conclusion, there has not been drastic changes over time for innovation score for each continent and more changes to the average research & development as % of GDP, where the continent who had the biggest change was Europe that had a spike up between 2018 and 2019 and at the same time period there was a drastic decrease in Asia for the average research and development as % of GDP.

Scatterplots of Average Availability of Latest Technology plotted against Average Innovation Ratio, grouped by Continent

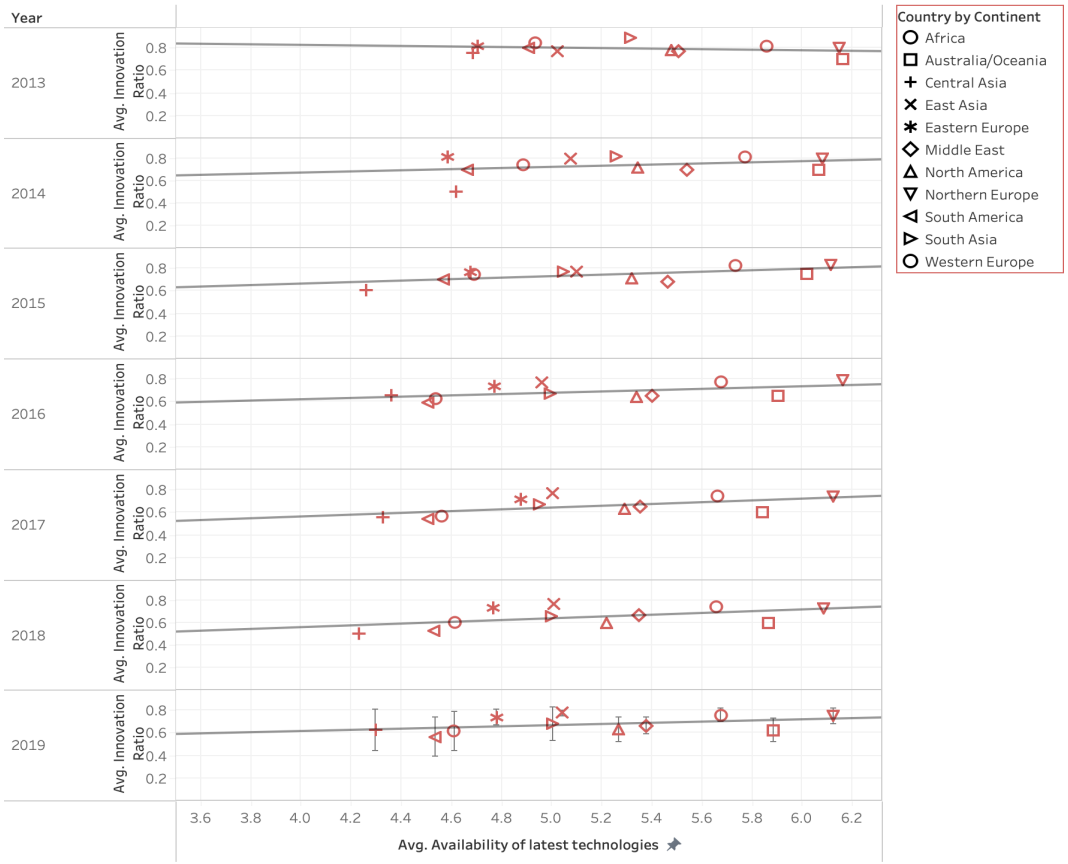

Figure 6. Scatterplots of Average Availability of latest Technology plotted against Average Innovation Ratio - Nick Sansone

This chart illustrates the average availability of the latest technologies for each continent compared to their average innovation ratio. Each year of data (2013-2019) is represented by its own graph that plots these two variables against each other, with each continent represented by a different shape.

In 2013, there exists a negative correlation, but the trend line shifts towards a positive slope as the years progress, with the steepest positive slope in 2019.

Analysis of this chart reveals that although the steepness of the trend line positively increases, the average innovation ratios for continents is actually decreasing as availability of latest technology remains relatively stagnant. This implies that the innovation input score is increasing in comparison to the innovation output score across continents.

Area Graph and Scatter Plot showing Average Governance Indicators plotted against average Innovation Score for each continent

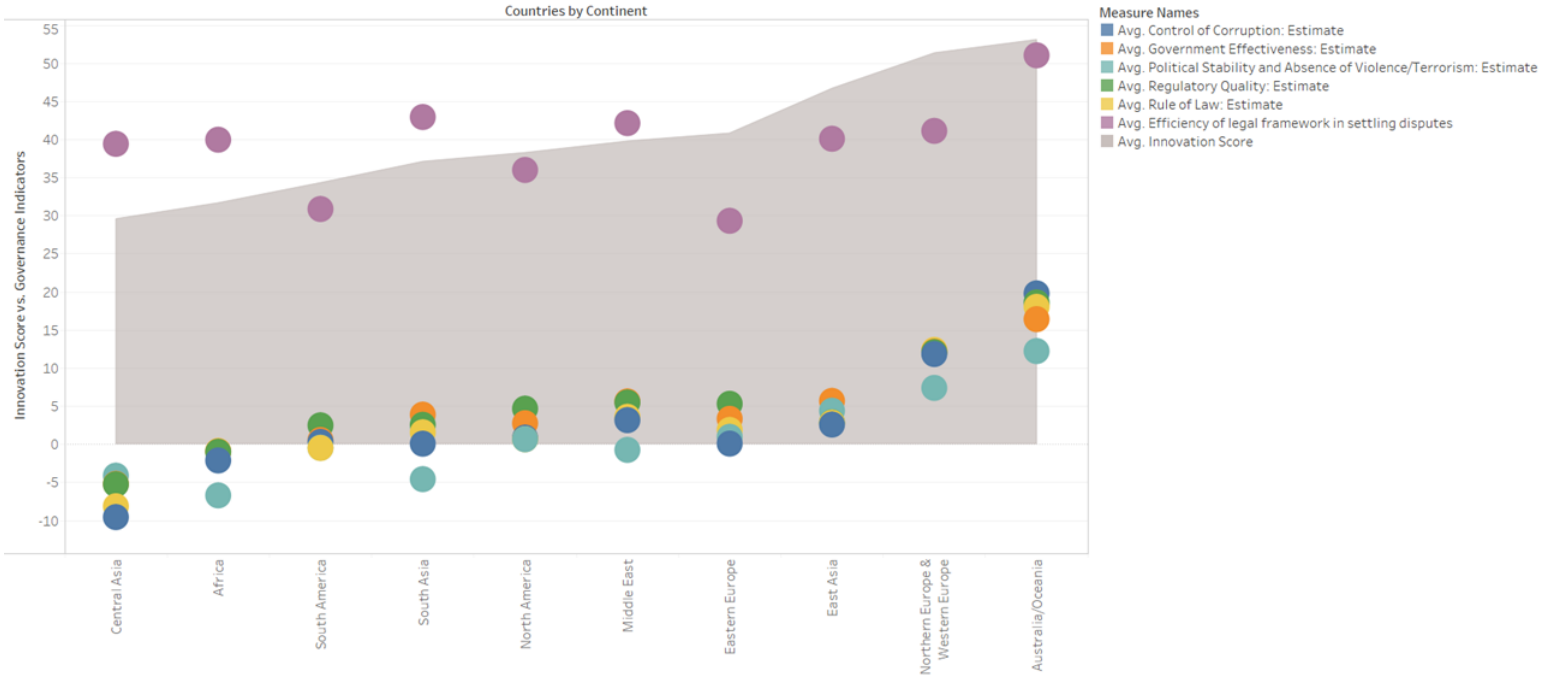

Figure 7. Average Governance Indicators plotted against average Innovation Score for each continent - Olha Sverdel

The chart graphs the average governance indicators and average innovation score over six years (2013 - 2018) for countries that are grouped by 10 geographical areas. Governance Indicators include Control of Corruption Estimate, Government Effectiveness Estimate, Political Stability and absence of Violence/Terrorism Estimate, Regulation Quality Estimate, Rule of Law Estimate, and Efficiency of Legal Framework in Settling disputes. Governance Indicators are marked as circles on the graph, and Innovation Score is marked as a linear area.

The graph shows if countries in a specific continent have higher Governance Indicators, they also have higher Innovation Score. Out of six Governance Indicators, Efficiency of Legal Framework in Settling Disputes, did not show the same behavior as other governance indicators.

The implication of the graph indicates a positive association between 5 out of 6 Governance Indicators and the Innovation Score. Countries might be able to improve their innovation Score by improving their governance indicators.

Geographical Heat Map and Bubble Chart showing Average Intellectual property protection plotted against average Research & Development as a % of GDP for each country

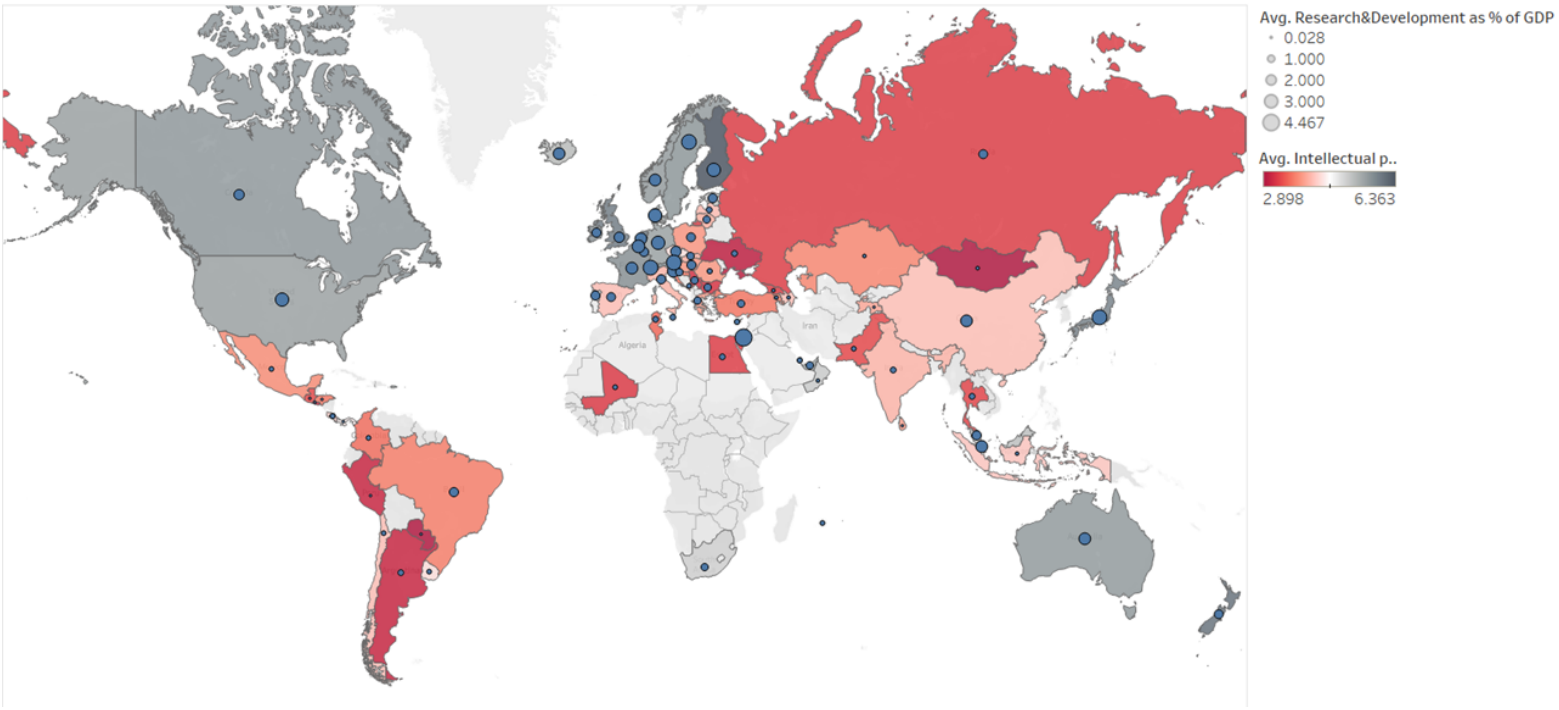

Figure 8. Average Intellectual property protection plotted against average Research & Development as a % of GDP for each country - Olha Sverdel

The chart shows the Geographical Heat Map that indicates the strengths of Intellectual Property Protection Index (red being the lowest and dark grey being the highest) and Research & Development as a percentage of GDP (the bigger the circle the higher the percentage) for each country taking the average of each indicator over six years (2013 - 2018).

The graph shows that in general, those countries that have higher Intellectual Property Protection Index, spend more on Research & Development.

The implication of the graph might be that those countries with better Intellectual Property Protection might be more motivated to spend more on Research & Development.

Heat bubble chart showing average percentage of patents approved plotted against the average innovation score of the country

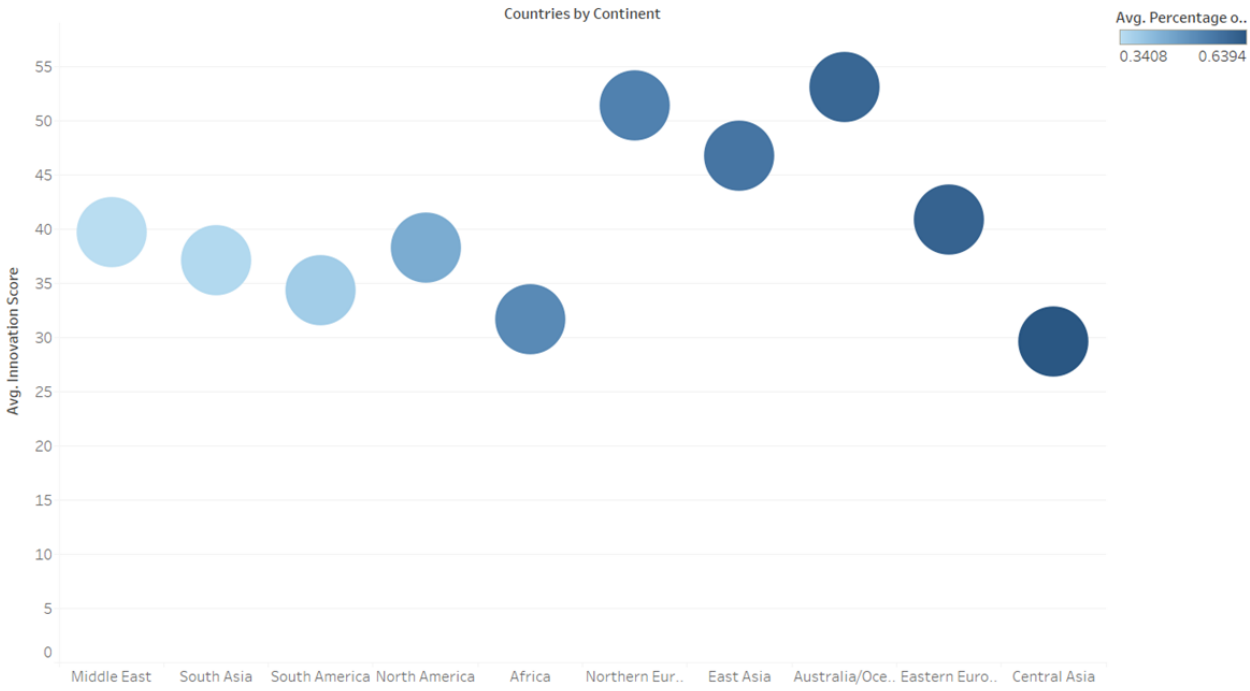

Figure 9. Average percentage of patents approved plotted against the average innovation score of the country - Olha Sverdel

The chart shows the average innovation score over six years (2013 - 2018) along the y axis and the average percentage of patents approved for each continent (darker circle means higher percentage).

The graph shows no significant relationship between the percentage of patents approved and the innovation score.

The implication of the graph indicates that on average the innovation score of the country does not depend on the amount of patents approved by that country, therefore the increase of the approved patents will not promote innovation in the country.

Bar Graph showing Innovation score and Innovation Efficiency ratio plotted against the income level (development status of a country)

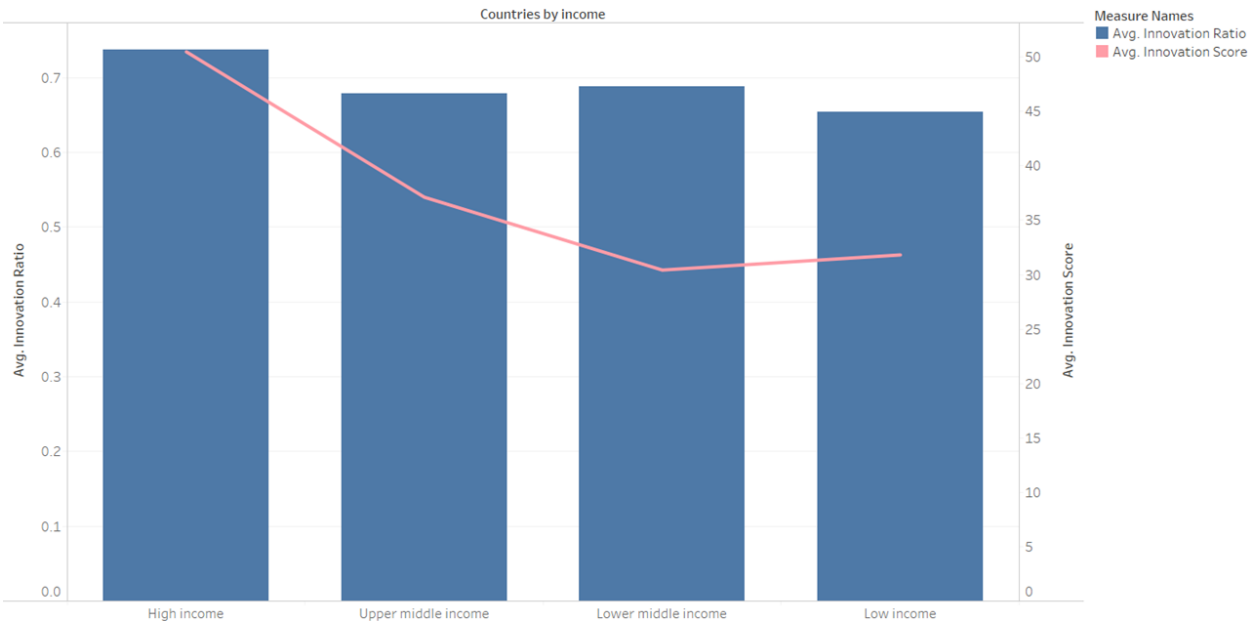

Figure 10. Innovation score and Innovation Efficiency ratio plotted against the income level (development status of a country) - Olha Sverdel

The chart shows the average Innovation Score and the average Innovation Ratio of a country depending to which income level it belongs to.

The graph shows a positive relationship between the income level (development status) of a country and the average innovation score. However, it also shows a weak positive relationship between the average Innovation Raio and the Income level (development status) of a country.

The implication of the graph is that it helps us understand that the income of the country impacts the innovation level of the country.

Area Graph of Availability of Latest Technologies plotted against Percentage of Patents Approved for the period 2013-2019, colored by Year

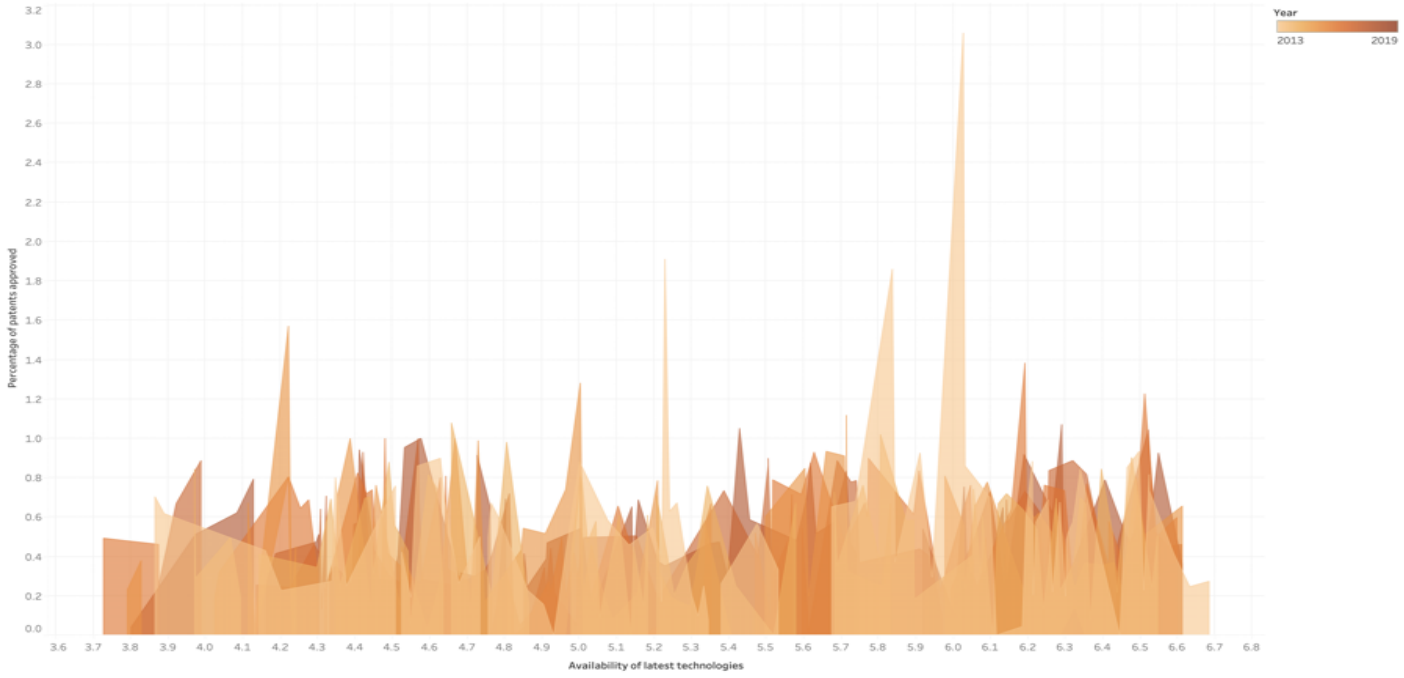

Figure 11. Area Graph of Availability of Latest Technologies plotted against Percentage of Patents Approved for the period 2013-2019 - Nick Sansone

This chart is visualizing scores for the Availability of latest technologies plotted against the Percentage of patents approved over the time period 2013-2019. Each country is plotted as a point in order to illustrate the overall relationship between these variables in our data.

As the score for availability of latest technologies increases, the percentage of patents approved does not have a significant increase except for a few outliers which exist as the spikes in Figure 11.

These insights imply that the variables do not have a strong correlation with each other. The shading of the area graph also suggests that the year does not play much of a role in the relationship either as there is no trend within the shading.

Stacked Bar Chart of Average Availability of Latest Technologies and Average Efficiency of Legal Framework in Settling Disputes for the period 2013-2019, grouped by Continent

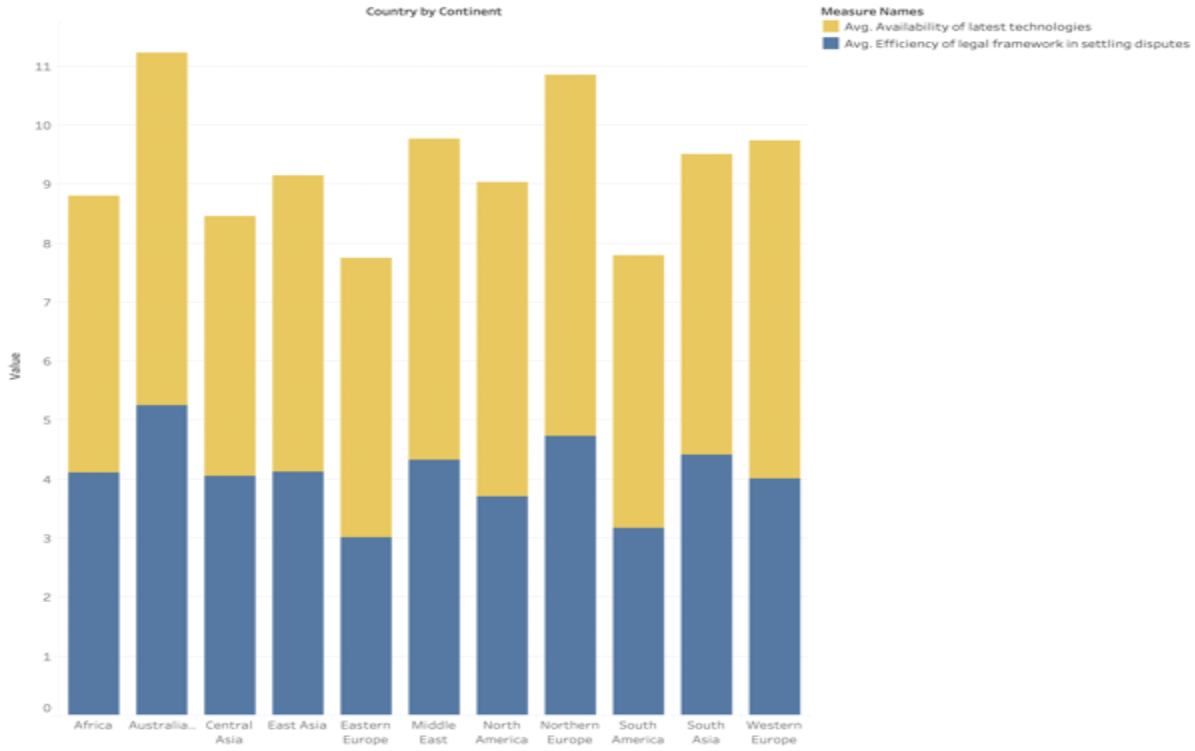

Figure 12. Stacked Bar Chart of Average Availability of Latest Technologies and Average Efficiency of Legal Framework in Settling Disputes by Year for the period 2013-2019 - Nick Sansone

This image depicts the average availability of latest technologies with the average efficiency of legal framework in settling disputes for each country, grouped together by continent and geographical location. The bottom portion of the bar represents the latter, with the top half representing the average availability of latest technologies.

As the bars increase in height, both portions of the bar increase proportionally displaying that as one variable increases, the other does as well.

This relationship implies that there is a positive correlation between the two input variables for each of the countries grouped together by continent.

Area chart showing the percentage of patents approved plotted against Country's GDP

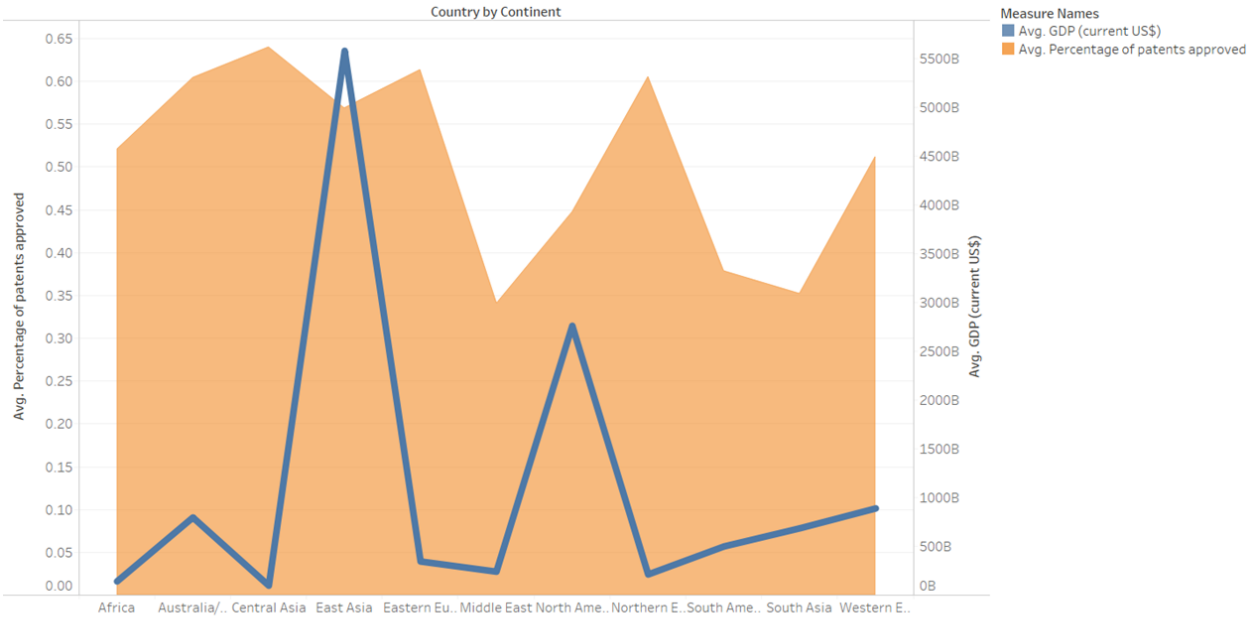

Figure 13: Percentage of patents approved plotted against Country's GDP - Olha Sverdel

The area chart combined with a linear chart shows the average percentage of patents approved and the average GDP of a country grouped by continents over six years (2013 - 2018).

The graph shows no significant relationship between the country's GDP and the percentage of patents approved.

The implication of the graph indicates that the GDP of a country is not driven by the amount of patents approved

Waterfall Chart of Difference in Average Availability of Latest Technologies over the Years 2013-2019, grouped by Continent

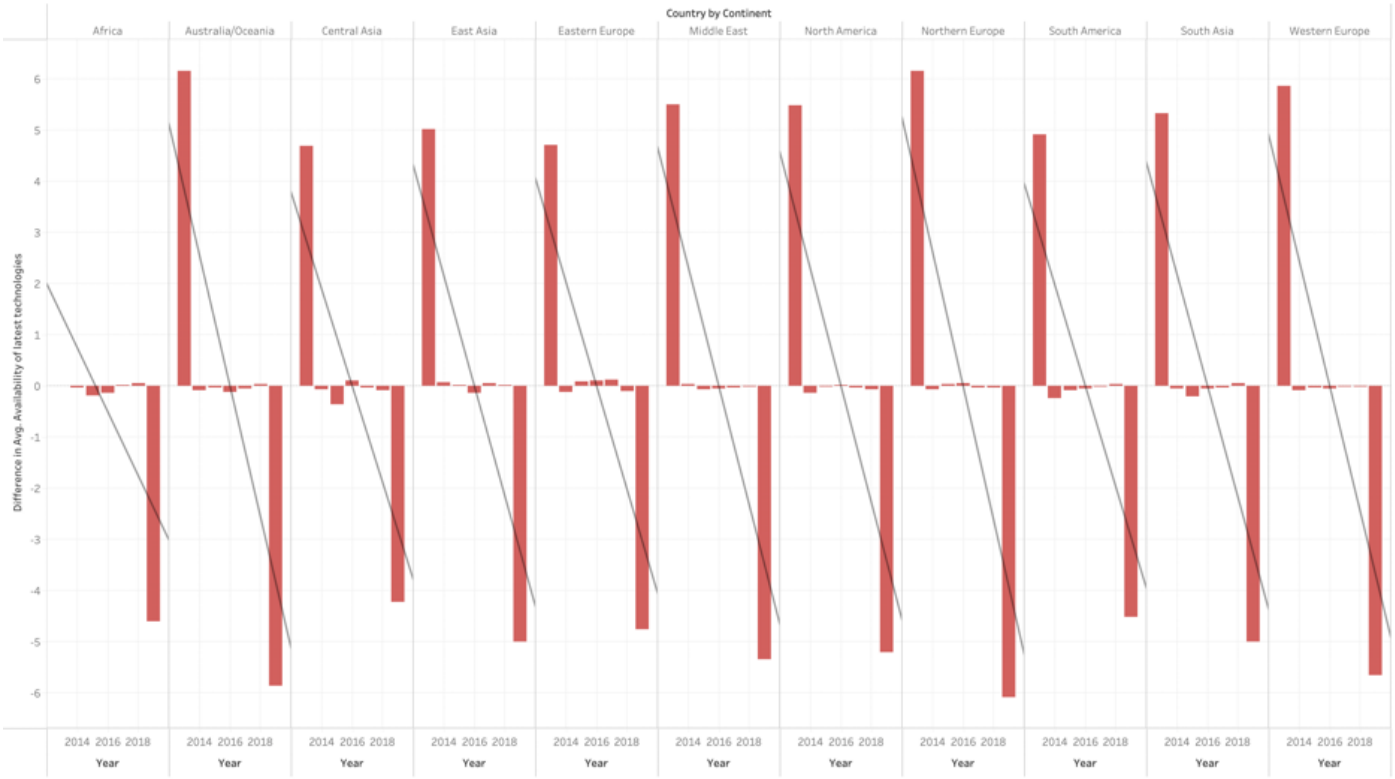

Figure 14. Waterfall Chart of Difference in Average Availability of Latest Technologies over the Years 2013-2019 - Nick Sansone

This waterfall diagram illustrates the difference in average availability of latest technologies over the years for each continent. This portrayal of the data depicts the trend of how the input variable changes throughout the time period 2013-2019.

There is a clear decrease in the scores for availability of latest technologies from 2013-2019 for all of the continents, with the largest decrease being from 2018-2019.

This negative correlation suggests that when grouped together, no continent dominates the market with the latest technology. Although some countries individually might, there are no geographical locations that are innovatively ahead of everywhere else.

Scatterplots with Trend Lines of Innovation Ratio plotted against Percentage of Patents Approved for the period 2013-2019, grouped by Continent

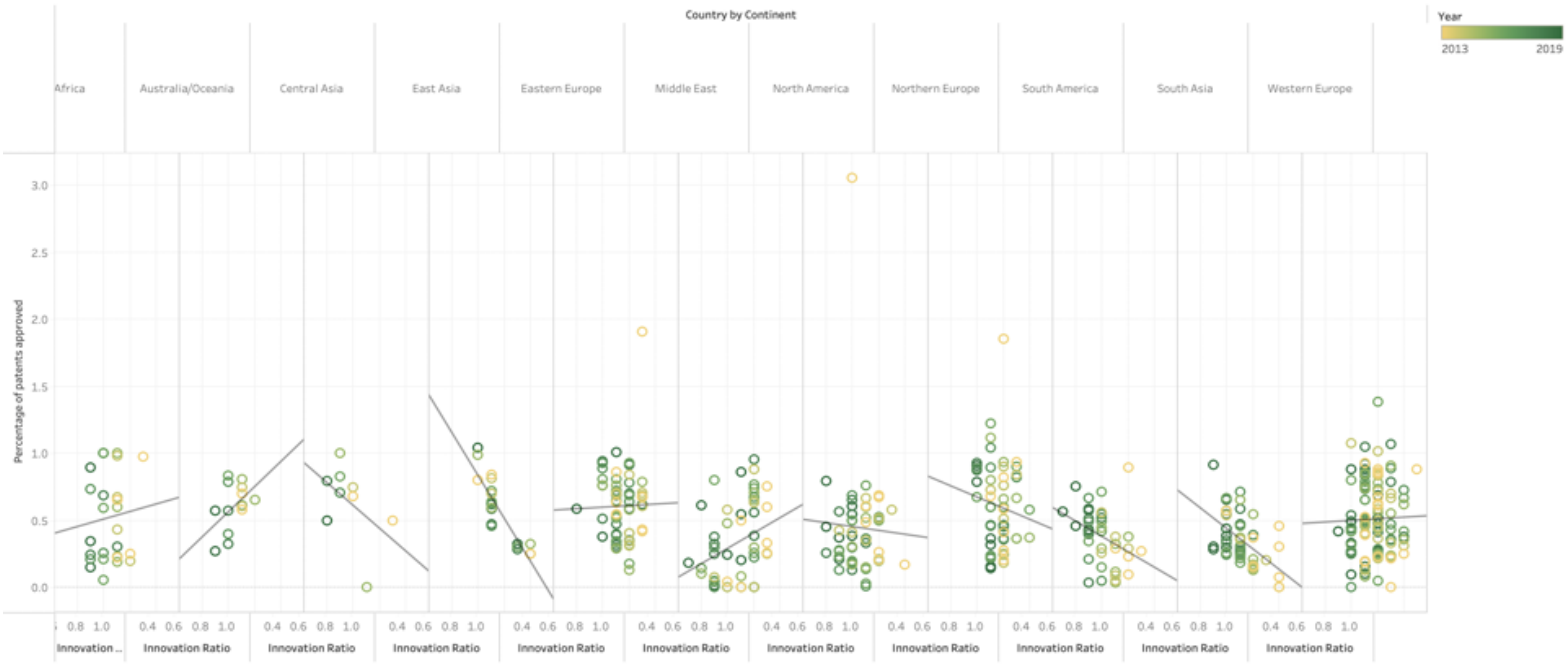

Figure 15. Scatterplots with Trend Lines of Innovation Ratio plotted against Percentage of Patents Approved for the period 2013-2019 - Nick Sansone

This visualization shows the Innovation Ratio for each continent plotted against their percentage of patents approved. The data points are shaded with lighter colors being earlier and darker colors being later, in the time period 2013-2019.

The plots represent that the percentage of patents approved increases as innovation ratio increases for some countries and decreases for others. The shading of the data points illustrates that as the years increase in the range 2013-2019, the innovation ratio tends to decrease for countries.

The above implies that there is not a strong correlation between the input variables as there is no clear trend displayed by Figure 15. It does however illustrate that countries have become less innovative over the course of 2013-2019, as there is a negative correlation between the innovation ratio and the year.

Bar chart of the countries average venture capital availability and the nations regulatory quality score for each year

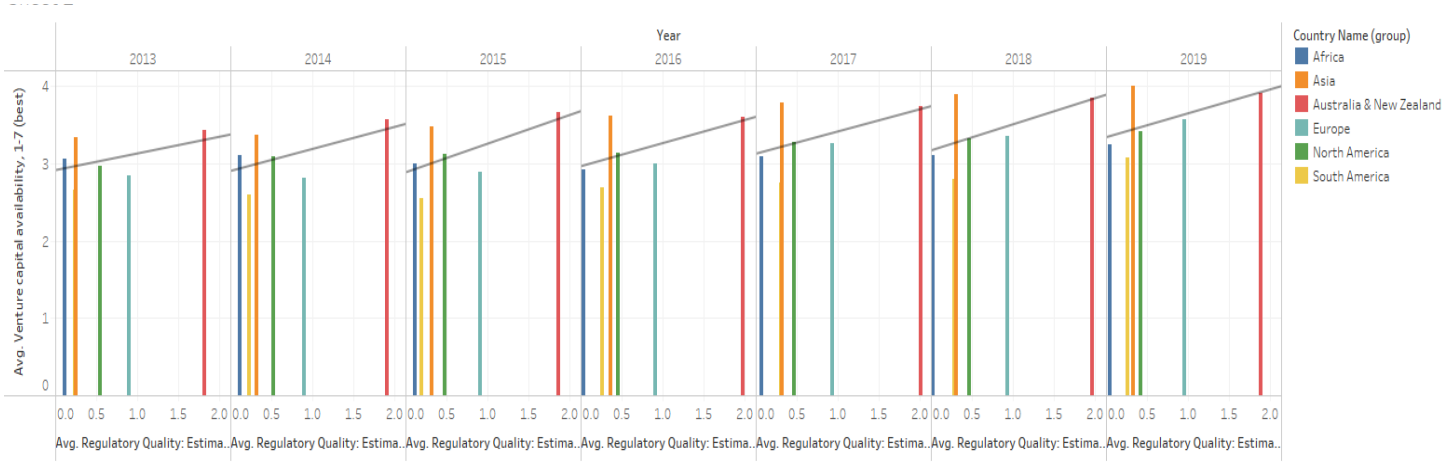

Figure 16. Bar chart of the countries average venture capital availability and the nations regulatory quality score for each country - Tahsin Rakib Himi

The graphs in figure 16 shows the average Regulatory Quality Estimate for each continent plotted against their average venture capital availability broken down by each year. The data is shown from 2013 to 2019. The bar charts each indicate the average for each continent.

The plots represent the average venture capital availability as the dependent variable vs. the Regulatory Quality estimate as the independent variable. From the graph's trend lines it can be shown that there is a positive correlation between the two variables.

The figure gives us the understanding of how the variables relate to each other. From figure 16 it can be seen that there is a positive correlation between average regulatory quality estimate and average venture capital availability. However when the P values are studied it can be seen that given that the P values is greater than 0.05, the results are not statistically significant.

Line chart with trend line shows the average control of corruption: Estimate plotted against average innovation score for each country

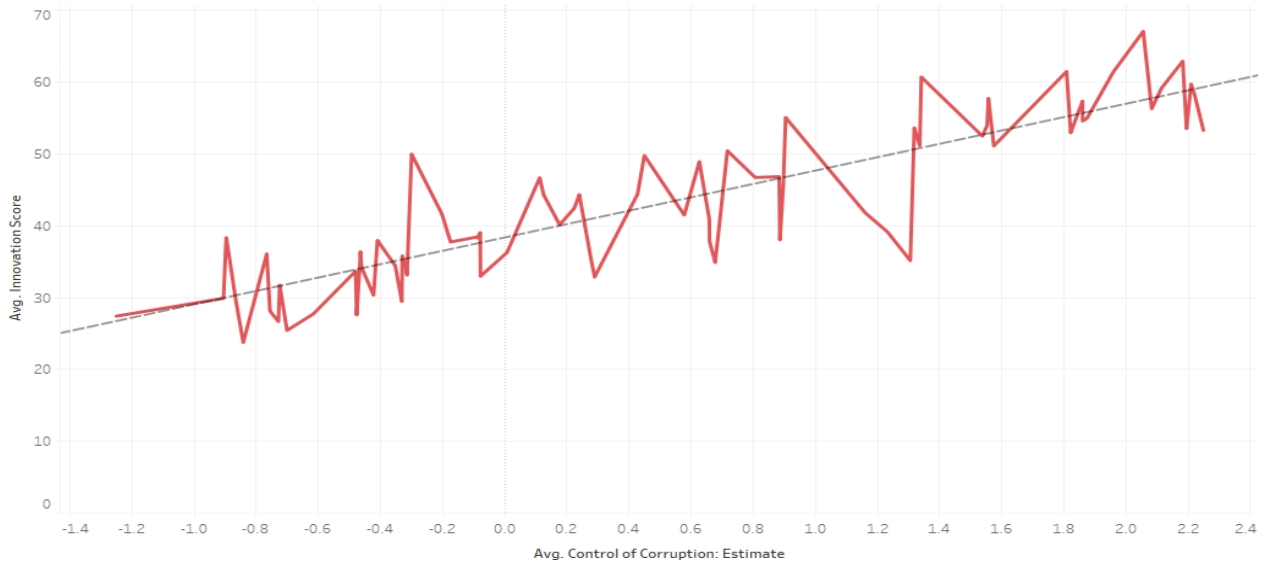

Figure 17. Line chart with trend line shows the average control of corruption: Estimate plotted against average innovation score for each country - Tahsin Rakib Himi

Figure 17 shows us a line chart showing the average control of corruption score as the independent variable vs average innovation score as the dependent variable plotted for each country. This data depicts how the nation's control of corruption can relate to the innovation score for each nation.

There is a clear increase in the innovation score as the control of corruption score increases for all the countries, with 1.3 control of corruption being an outlier.

There is a positive correlation between the variables indicated by the trend line for each country.

Bar chart of average regulatory quality score and average innovation score from 2013-2019 grouped by continent

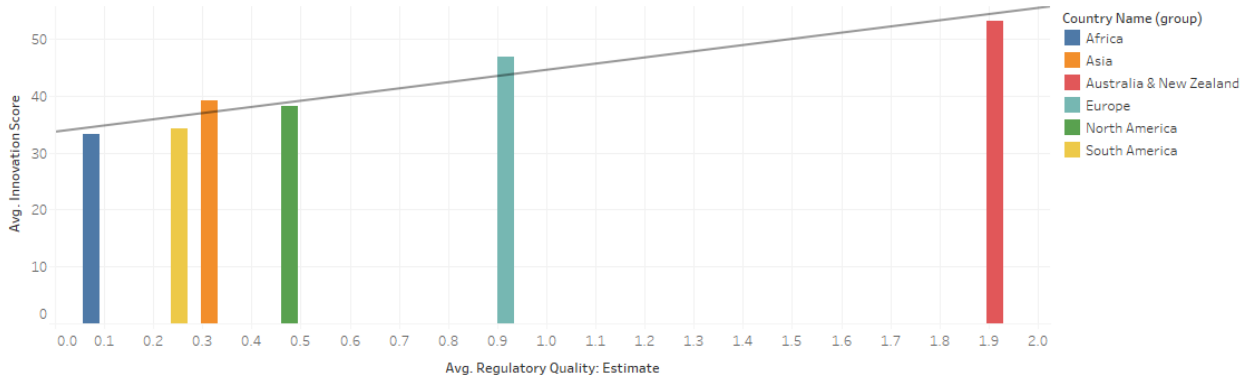

Figure 18. Bar chart of average regulatory quality score and average innovation score from 2013-2019 grouped by continent - Tahsin Rakib Himi

This image depicts the average regulatory quality score with average innovation score, grouped together by continent. The graph indicates that Australia as a continent has the highest score and highest average regulatory quality score.

The plots represent the average Regulatory quality as the dependent variable vs Average innovation score as the independent variable. From the graph's trend lines it can be shown that there is a positive correlation between the two variables.

In conclusion, based on the graph it can be inferred that there is a positive correlation between the variables. With Africa having the lowest innovation score and Australia having the highest innovation score.

Line chart between average innovation score and average Political stability, grouped by continent

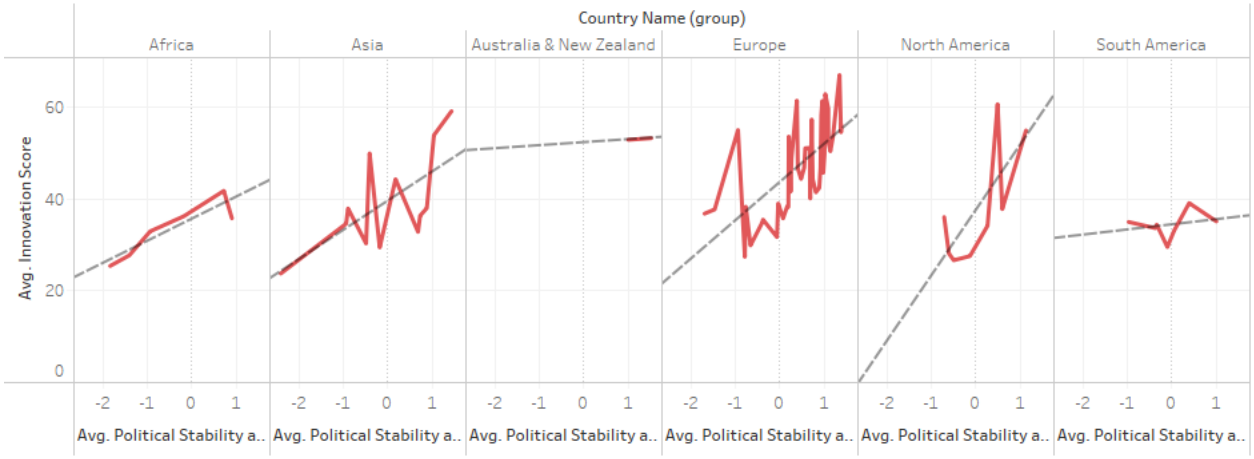

Figure 19. Line chart between average innovation score and average Political stability, grouped by continent - Tahsin Rakib Himi

The line chart describes the average political stability score for each country vs average innovation score grouped by continents. The graphs also contain trend lines to understand the relationship between the variables.

The chart shows that there is a positive correlation between average political stability and average innovation score.

In conclusion, there is a positive correlation graphically between average political stability and average innovation score grouped by continents. From the graph it seems that there is a strong positive correlation for North America. Looking at the P values for the trend lines, it can be given that besides North America the rest of the continents are not statistically significant. For North America the P value is less than 0.05 therefore that result is statistically significant.

Average Percentage of patents approved vs Average Efficiency of Legal Framework grouped by continent

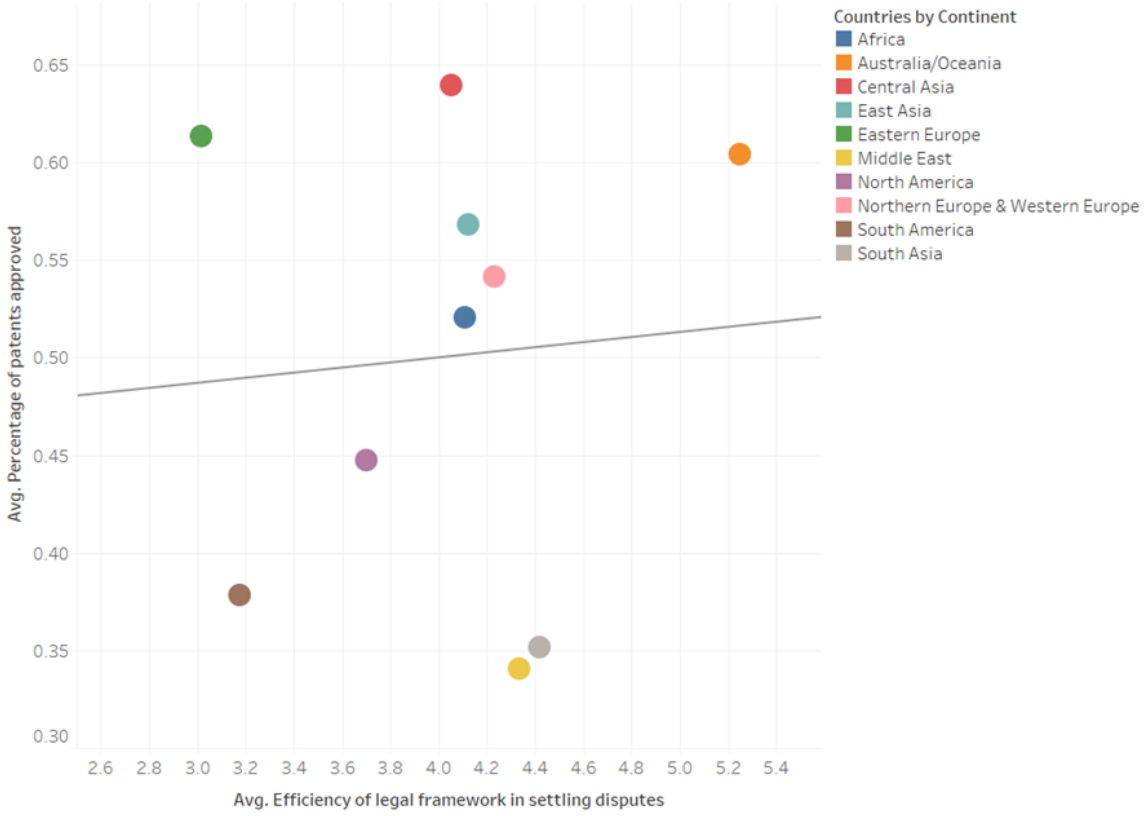

Figure 20. Average Percentage of patents approved vs Average Efficiency of Legal Framework grouped by continent - Tahsin Rakib Himi

This chart showcases the average score of the efficiency of the legal framework as the independent variable and average percentage of patents approved as the dependent variable.

Our hypothesis suggests that if a country has a higher score i.e. a higher efficiency of the legal system then they are more likely to have a higher innovation ratio. Based on the trendline it seems that our hypothesis holds true, that there is a positive correlation between the two variables.

The implication suggests that if a country has an efficient legal system then there is likely going to be a higher percentage of patents that get approved.

Machine Learning Model: Decision Tree

Decision Tree algorithm will be used in the project utilizing IBM SPSS tool. The independent variables of the model will be (1) Intellectual property protection (rated on a scale of 1 to 7); (2) Governance Indicators (5 variables); (7) Efficiency of the legal system in settling disputes, (1-7); (8) Availability of latest technologies (rated on a scale of 1 to 7); (11) percentage of patents approved; (12) GDP of the country. The dependent variables in the algorithm will be Gross Domestic Expenditure on R&D as a % of GDP and the Innovation Score. Using Decision Tree, all the independent variables will be utilized to explain the different ranges of the amount of the R&D spent in a country as well as its level of the Innovation Score. Decision Tree will help understand to what extent a country spends on R&D depending on the values of the given independent variables.

Using the variables described above as inputs, 2 different decision tree algorithms were run for each of the target variables. For the Innovation Score, it was hypothesized that Government Effectiveness plays a key role in determining the Innovation level of a country. For the Gross Domestic Expenditure on R&D as a % of GDP, we proposed that the GDP was the most critical input variable.

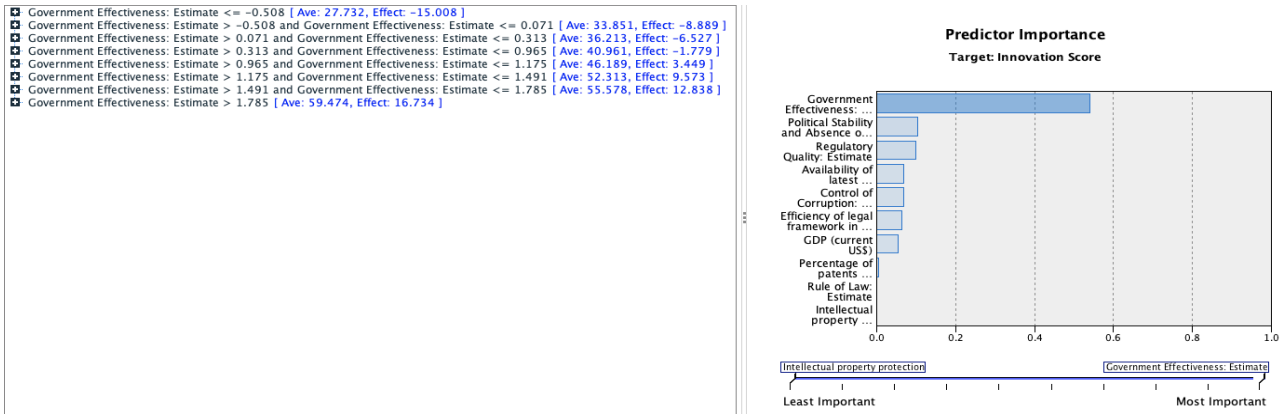

Figure 21: Decision Tree Model predicting Innovation Score as Target Variable

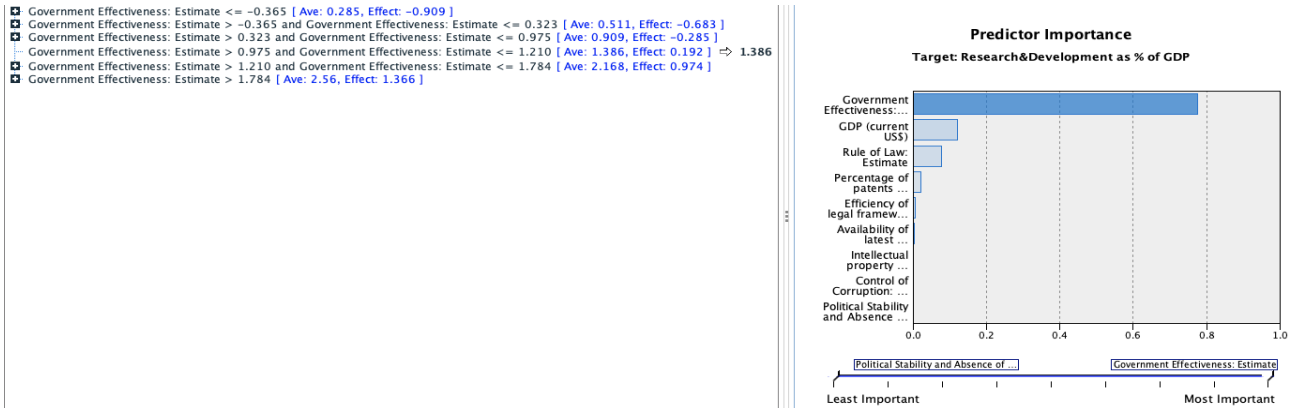

Figure 22: Decision Tree Model predicting Expenditure on R&D as a % of GDP as Target Variable - Nick Sansone

The above visualizations show the results of a decision tree algorithm run through SPSS Modeller. Using the input variables described, the top visual displays the output using innovation score as the target variable and the bottom shows the output using expenditure on R&D as a % of GDP as the target variable.

The government effectiveness (estimate) is by far the most significant factor in predicting both of our target variables, having a stronger predictor importance score for the expenditure on R&D as a % of GDP. For innovation score, the other input variables are very similar in predictor importance score (~0.10), whereas the GDP (current US$) is a clear second and percentage of patents approved being third for the other model.

The results of this model imply that a country's government effectiveness is crucial in its ability to invest into their own research and development. Originally, it was thought that their GDP would be the most pivotal variable in the expenditure on R&D as a % of GDP as a country with more money to spend would be willing to invest more on their own advancement- however, our decision tree models reveal that the ability for a government to work cohesively and be effective is by far the most influential for both target variables.

Statistical model: Two Way ANOVA

ANOVA Table result Statistical Modeling

Anova Table						
Source of Variation	**SS**	**Df**	**MS**	**F**	**P-Value**	**F Crit**
Sample	**678.7581**	**76**	**8.931027**	**498.5091**	**0**	**1.296958**
Columns	**48.624**	**1**	**48.624**	**2714.078**	**3.2E-277**	**3.85154**
Interaction	**84.89248**	**76**	**1.117006**	**62.34869**	**0**	**1.296958**
Within	**16.5539**	**924**	**0.017915**			
Total	**828.8285**	**1077**				

Table 1: ANOVA Table result Statistical Modeling - Tahsin Rakib Himi

For the statistical model, the study utilised two way ANOVA analysis for the Government effectiveness and Political stability and absence of violence/terrorism. Purpose for ANOVA analysis is to study the statistical significance between two independent variables. Table 2 shows the Anova score obtained from the analysis performed in EXCEL. The sample for this study were the countries with n being 7 for each year.

Based on the table 2 the F values are 498.5 for the samples and the for the columns the score is 2714.1. The table shows that the P score for each is less than 0.05.

From that table it can be inferred that the variables are statistically significant and therefore will play a role in the result for the dependent variable.