Creative Innovation: Happy, Healthy, Wealthy, or Wise?

Team 67: Marc Boulet, Minkwon Lee, Michael Migliacio, Ricardo Stamato, Lisa Walkosz-Migliacio, Catherine Yin

Introduction

As a concept, innovation requires tackling problems in new, creative ways. There are many discussions concerning the concepts of innovation and creativity in terms of breakthroughs in technology, disruption by the gig economy, advances in treating diseases, and more. However, how can innovation be defined in a quantitative manner, and are creative nations by definition also more innovative?

Problem Definition

Based on research by Lopez-Claros & Mata (8) as well as others, innovation as a concept is notoriously difficult to quantify accurately. Is innovation impacted by educational, personal, financial, and societal structures of a nation? How can all of these relationships be effectively compared in an understandable way? We intend to address those questions by determining innovation across the world utilizing a variety of measures (including creative, economic, human, and cultural), comparing our findings against an existing innovation index (19), and then craft a means to visually break down the “what”, “why”, and “how” of innovation at a country level in an interactive manner. Our visualization will also be able to contrast each country’s level of innovation to data concerning health (life expectancy), wealth (GDP), wisdom (education), and happiness (quality of life), or a combination thereof. This is an important question to explore because innovation is an essential component to a country’s development and has the potential to improve quality of life across the board. By understanding what factors contribute to innovation and recognizing its relationship to a nation’s development, we can seek to encourage those factors.

Research Survey

How is innovation quantified? Many measurements have been attempted up to today, including relating the quantity of innovation in a nation to available social capital (1), human capital (2), entrepreneurial orientation & cultural values (3), non-business factors in public and household sectors (4), creativity itself through the generation of a “bohemian” index based on data around performing arts, museums, and more (5), financial markets (6), and resources including investments and seed funding (7). These findings could prove useful to us when building out our innovation index as they provide a wide view into the kinds of factors that have been examined in the past. Others have taken a more comprehensive approach, generating innovation indices based on governmental policies and institutions of several nations (8), or combining a large number of innovation predictors across research, business, and economic sectors into a single synthetic variable (9). These studies are relevant to us as the indices described can be used to compare against our own index. Researchers also spent time focusing on entrepreneurship, which has also been shown to play a significant role in innovation - though the effect was impacted substantially by differences in cultural norms (3). Similarly, the presence of foreign owned companies in a nation (10) and as well as relationships built between subsidiaries of a foreign company and its host (11) also show demonstrable increases in the measure. While interesting, factors such as entrepreneurship, cultural norms, and foreign-owned investment are not a focus of our work. Researchers also discovered that country-specific factors play a role in

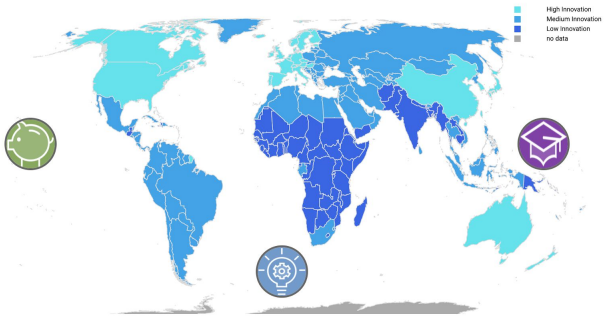
innovation - for example, it has been shown that lower income countries produce a wider variation of solutions to problems (12), and that the concept of “innovation” in general varies based on economic context when comparing OECD versus developing nations (13). These factors will certainly play a role in our work as we look to create an innovation index that is valid for all nations regardless of economics or level of development.

Quantifying creativity as a concept has been explored via analysis of a similarly wide set of parameters. Pratt & Jeffcutt, for example, found that a set of 13 discrete creative industries such as film and recorded music can be used as an organizing mechanism for predictive attributes (14), while others found that a single mechanism, such as export performance, had an impact on creativity (15). Looking at creative industries will be useful when computing our innovation index. Additionally, research has shown how certain metrics related to innovation (such as income and education) impact each other positively (16), and whether happiness is correlated to creativity (17). These two studies show relationships to measures of innovation and creativity, and therefore will prove useful to us due to our focus in those areas. Innovation is difficult to quantify; as research has demonstrated, there are many factors that can affect the measure, and they vary by nation. Starting with creativity/cultural factors is our way to lessen the impact of economics on developing nations. Through this seven week analysis, we plan to measure innovation through a number of factors tied to both economic and non-economic factors, and then try to identify predictors at a country level. Data will be gathered from a variety of sources and then reduced through the use of regression and graph analysis to compute an innovation index, which can then be used to compare countries in the common measures described below. New in our approach is combining a mix of sources including both creative fields such as film and music and economically-significant measures, like patents. Our visualization will allow users to view not only our index, but the related measures of life expectancy, GDP, quality of life, and education alone or combined. This approach will be extremely useful to government officials in developing countries as well as investors across the world as they look to build a global marketplace. If successful, a new way of computing innovation through creativity will be introduced. We will be able to validate our findings by comparing them with the Global Innovation Index (19). Through research, we discovered that many existing studies lacked recent data which could have a significant impact in how innovation is measured due to the global spread of technology, including Saint Paul & Verdier (16), Almeida & Phene (11), Dahkli & DeClercq (2), DiPietro & Anoruo (15), and Pratt & Jeffcutt (14), while other studies limited their focus to only a few usually wealthy and/or developed nations, including Dachs et al. (10), Rauch et al. (3), Williams & McGuire (18), and Griffiths et al. (7) which impacts the validity of the data on a worldwide scale. Others still compared variables that could introduce unexpected outliers in the data depending on the nation, including Fagerberg et al. (13), Lopez-Claros & Mata (8), Lee et al. (5), Crespo & Crespo (12), Roszko-Wójtowicz & Białek (9), Gault (4), and Polatcan et al. (1). Others focused on only one factor, which reduces the potential for identifying multiple factors contributing to findings, including Hsu et al. (6) and Ceci & Kumar (17).

Proposed Method

Due to the large number and diverse amount of datasets being analyzed for this effort, our method requires several different approaches. First, we plan to combine and clean and merge the datasets to generate a single large dataset organized by country

for years 1950 to 2020. We will utilize PCA/feature extraction to determine the most relevant predictors for innovation. A supervised learning approach (linear regression) will be implemented to compute a potential response variable; we will be utilizing the Global Innovation Index (19) as a benchmark in order to check our results against existing measures. Further checking will be done via an unsupervised model (k-means clustering), in which nations with similar levels of innovation as determined by our analysis will be grouped together. This will enable us to modify our index: if we find unexpected results, we will do further analysis, as needed, to minimize the gap between the results of the supervised and unsupervised models. In order to compensate for incomplete or missing information in datasets under test, we will be calculating and utilizing an adjacency matrix containing the nations of the world. This will allow us to make certain assumptions about countries for which our data is incomplete and interpolate the results.
The final results will be fed into a visualization that will group nations by level of innovation, allow users to compare results in both economical and cultural measures, and then easily deduce the differences between them.



In this sample visualization, icons in each of the circles can be pressed to refresh the map with respect to values of the corresponding icon's index (creativity & economical measures, with darker colors indicating larger values).

Innovations for this Methodology
This methodology will provide several benefits / distinct innovations to the existing body of research, including:

- Information is easily consumable and interactive
- A significantly large number of cultural and economic measures are taken into account when examining innovation from a wide variety of datasets in order to compensate for challenges across nations - most studies focus on a small number of variables rather than going for a "big picture" view
- Users can easily deduce which related measures (economic or cultural) are correlated most with innovation for a given nation, and also which measures contribute more significantly to innovation per nation

In terms of the value of our experiments, based on our research survey, many existing studies focused either on a small subset of nations, limiting the scope of the accuracy of their findings, or instead focused on a small subset of variables emphasizing either economics or culture - not both. We're hoping to address both in our analysis and experiments.

Testbed
Concerning our testbed, we will be utilizing Python for data cleaning and analysis, D3/JavaScript for visualization, and CSV files for data storage and access. We are combining elements from several publicly available datasets measuring economic and cultural factors in order to compute our findings; a full list is available in the References area of our document. In order to measure the results, we are investigating

a variety of metrics to assess the impact of innovation. This enables us to be flexible in approaches beyond linear regression. The application of clustering will serve as a check on our supervised model to determine if our results are sufficient at measuring innovation.
We are trying to measure predictive factors that could be used to signal "innovativeness" in a nation, as well as a numeric metric representing "innovativeness" itself. We will be using the Global Innovation Index (19) as our control in order to serve as a check on our findings.

Evaluation / Experiments & Results

- We intend to define a potential metric for innovativeness (through an aggregation of features).
- We will be attempting to compare the effects of including/excluding cultural indicators in a supervised model, with the potential innovativeness metric as the response variable, and use feature extraction to determine the most significant factors.
- Finally, we will use an unsupervised model (e.g. clustering) to group countries and determine the factors that should be used in the analysis. If these factors are correlated with innovation (based on our previous experiment), does the clustering analysis reflect those findings?
- We're looking to answer the following questions: on a per-nation basis, what factors contribute most to innovation? What makes a nation more "innovative" in general (cultural or economic factors)?

The data under analysis consists of a dataset with 46 features built off of 40 datasets from a variety of sources (the World Bank, the Global Innovation Index site, WHO, World Population Review, and UNESCO). Due to the differences in date ranges (1950-2020) and available data for less populated nations, certain challenges were encountered during cleaning and integration of data. However, we were able to overcome those challenges and eventually clean, transform, and merge the datasets organized by both year and country ISO. Rows are wide and include all features being examined across economic and cultural bounds.

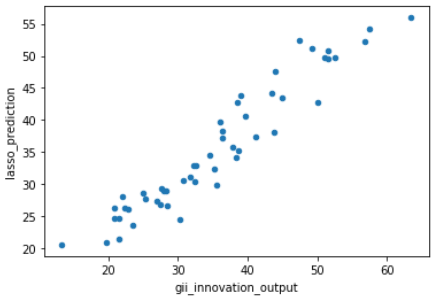
Supervised Linear Regression Models
Because the generation of a model in which the predicted value of our computed index could be statistically compared against existing research would be exceptionally helpful in drawing conclusions from our analysis, we decided to utilize several commonly used models of supervised learning (including several types of linear regressions) in order to explore the relationship between potential economic and cultural factors impacting innovation. However, generating supervised models based on our dataset provided several challenges for us. First, the large dataset containing all 46 features consists of information going back several decades, beginning in 1950. In order to account for changes over time and maximize the recent relevance of our data, we decided to compute our innovation index based on data beginning in 2015 the year. We're also limiting our predictions to the set of countries for which control data is available (in this case, from the Global Innovation Index (19)). This reduced the number of features used across all supervised models - including random forest (which we utilized for comparison purposes against our linear regressions).

	Mean absolute error	**r2_score**
random_forest	3.947963	0.711394

lr	5.615887	0.320886
lr_lasso	3.164182	0.814075
lr_rfe	3.461543	0.858872

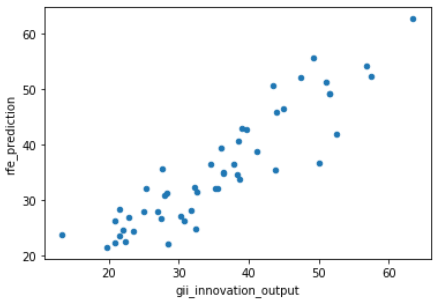
As shown in the table above, the two best performing models are the Lasso linear model and the linear regression using RFE when plotted against the Global Innovation Index. According to the Lasso model, the "ease of business" feature was shown to have a negative effect on generating the innovation index. However, the features used while generating the model include a mix of creativity, economic and educational factors. There isn't a specific subset of features that could be parsed out as having a more significant effect on results.

Lasso Model - Prediction vs. GII



In contrast, the linear regression model using RFE shown below was able to predict innovation within a reasonable accuracy using only two factors: number of patent families and broadband subscriptions.

RFE Model - Prediction vs. GII



Looking at the scatter plots and the final Mean Absolute Error (considering all countries), we were able to conclude that the first model, Lasso, actually performs a lot better.

Therefore, while we were able to build linear regression models that predict a country's innovation index with reasonable accuracy, we were not able to tease out specific features across both economic and cultural landscapes that contribute significantly to our results.

Unsupervised K-means Clustering Model

Pair-wise testing of our dataset using correlation matrices revealed a high level of correlation between 26 of the 45 features present.

As a result, unsupervised clustering was performed on two separate datasets: one containing all 46 features, and another curated dataset containing the feature set shown with the least amount of correlation. K-means clustering was used for unsupervised classification, with cluster sizes between 3 and 8. As a QC step, the results were compared using the Global Innovation Index as a target variable. Based on this analysis, the most robust results came from the 3 cluster model using all features. For simplicity in visualization, we have labeled them as "low", "medium", and "high" levels of innovation.

We did discover some interesting findings around differences between cultural and economic levels of innovation on a per-nation basis which are clearly described in our visualization. Our full set of graphs can be found on the site below (https://github.gatech.edu/lwalkosz3/Innovation/blob/master/data_analysis/unsupervised/04_final_results_comparison.ipynb)

The curated dataset contained a significant amount of "noise", which made it difficult to identify an appropriate number of clusters with which to generate the model. Even with this small number of clusters, only two obvious trends (low innovation and high innovation) could be identified in the curated dataset - we are anticipating this is

due to the sparse nature of the collected data. Our unsupervised model also provides up to 140 extra predictions over what the target variables do and all countries are covered - not just those present in the Global Innovation Index.

Distribution of Team Member Effort

All team members contributed an equal amount of effort.

Conclusions

As opposed to explicit health, wealth, wisdom, happiness, our visualization displays innovation in economic vs. creative measures. The reason for this decision was based on due to large amounts of missing data for specific years and/or nations that introduced a significant amount of noise in the calculations for our supervised model (linear regression). Continuing analysis showed the requirement for “filling in” missing data through averaging in order to obtain granular values around health, wealth, wisdom, and happiness against innovation.

Despite these challenges, we were able to obtain interesting results. While we examined both supervised and unsupervised models in order to visualize innovation, the K-means unsupervised model using 3 clusters (evaluating both creative and economic factors) produced the best assessment of the innovation level of a country.

Our results showed that economic and creative innovation, while seemingly correlated at first glance, have been shown to be independent of one another. Some nations with moderate levels of creative innovation - such as India - are nonetheless economically limited, while the reverse is true of nations like Guatemala that register significant economic innovation without many cultural indicators. One of the largest players on the world stage right now, China, registered limited economic innovation, though significant cultural factors boosted the nation’s overall score. Unfortunately, our analysis using supervised models could not tease out much information about which groups of features would lead to the most significant impact on determining innovation at a quantitative level. Further analysis could be done looking at a more constrained set of factors across economic and cultural boundaries within a recent period of time, and comparing the results - particularly in supervised models, where we stand to gain the most knowledge of how innovation should be quantified when factoring in both cultural and economic measures.

References & Appendix

1. Polatcan, M., & Balcı, A. (2019). Social Capital Wealth as a Predictor of Innovative Climate in Schools. *International Journal of Contemporary Educational Research*, *6*(1), 183-194.
2. Dakhli, M., & De Clercq, D. (2004). Human capital, social capital, and innovation: a multi-country study. *Entrepreneurship & regional development*, *16*(2), 107-128.
3. Rauch, A., Frese, M., Wang, Z. M., & Unger, J. (2010). National cultural values, firm's cultural orientations, innovation, and performance: testing cultural universals and specific contingencies across five countries. *Frontiers of Entrepreneurship Research*, *30*(15), 4.
4. Gault, F. (2018). Defining and measuring innovation in all sectors of the economy. *Research policy*, *47*(3), 617-622.
5. Lee, S. Y., Florida, R., & Gates, G. (2010). Innovation, human capital, and creativity. *International Review of Public Administration*, *14*(3), 13-24.
6. Hsu, P., Tian, X., & Xu, Y. (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics, 112*(1), 116-135.
7. Griffiths, M., Gundry, L., Kickul, J. and Muñoz Fernandez, A. (2009), "Innovation ecology as a precursor to entrepreneurial growth: A cross-country empirical investigation", Journal of Small Business and Enterprise Development, Vol. 16 No. 3, pp. 375-390.
8. Lopez-Claros, A., & Mata, Y. N. (2010). The innovation capacity index: factors, policies, and institutions driving country innovation. In *The Innovation for Development Report 2009–2010* (pp. 3-65). Palgrave Macmillan, London.
9. Roszko-Wójtowicz, E., & Białek, J. (2016). A multivariate approach in measuring innovation performance. *Zbornik radova Ekonomskog fakulteta u Rijeci, časopis za ekonomsku teoriju i praksu-Proceedings of Rijeka Faculty of Economics, Journal of Economics and Business*, *34*(2), 443-479.
10. Dachs, B., Ebersberger, B., & Lööf, H. (2008). The innovative performance of foreign-owned enterprises in small open economies. *The Journal of Technology Transfer*, *33*(4), 393-406.
11. Almeida, P., & Phene, A. (2004). Subsidiaries and Knowledge Creation: The Influence of the MNC and Host Country on Innovation. *Strategic Management Journal, 25*(8/9), 847-864. Retrieved February 22, 2020, from www.jstor.org/stable/20142164
12. Crespo, N. F., & Crespo, C. F. (2016). Global innovation index: Moving beyond the absolute value of ranking with a fuzzy-set analysis. *Journal of Business Research*, *69*(11), 5265-5271.
13. Fagerberg, J., Srholec, M., & Verspagen, B. (2010). Innovation and economic development. In *Handbook of the Economics of Innovation* (Vol. 2, pp. 833-872). North-Holland.
14. Pratt, A. C., & Jeffcutt, P. (2009). *Creativity, innovation and the cultural economy*. Routledge.
15. DiPietro, W. R., & Anoruo, E. (2006). Creativity, innovation, and export performance. *Journal of Policy Modeling*, *28*(2), 133-139.
16. Saint-Paul, G., & Verdier, T. (1993). Education, democracy and growth. *Journal of development Economics*, *42*(2), 399-407.
17. Ceci, M. W., & Kumar, V. K. (2016). A correlational study of creativity, happiness, motivation, and stress from creative pursuits. *Journal of Happiness Studies*, *17*(2), 609-626.
18. Williams, L.K., McGuire, S.J. Economic creativity and innovation implementation: the entrepreneurial drivers of growth? Evidence from 63 countries. Small Bus Econ 34, 391–412 (2010).
19. Cornell University, INSEAD and WIPO (2019). "Global Innovation Index Report 2019" (PDF). *www.globalinnovationindex.org/*. Fontainebleau, Ithaca, and Geneva.

Datasets

- Country adjacency (JSON formatted datasets of country

neighborhoods) Awile, Omar. (2013). Country adjacency. Retrieved from: https://github.com/P1sec/country_adjacency

- DataHub. (2018). Comprehensive country codes: ISO 3166, ITU, ISO 4217 currency codes and many more. Retrieved from: https://datahub.io/core/country-codes#resource-country-codes_zip
- Wikipedia contributors. (2019, November 16). Books published per country per year. In *Wikipedia, The Free Encyclopedia*. Retrieved 21:27, March 26, 2020, from https://en.wikipedia.org/w/index.php?title=Books_published_per_country_per_year&oldid=926419380
- Broadband subscriptions (fixed broadband subscriptions refers to fixed subscriptions to high-speed access to the public Internet (a TCP/IP connection), at downstream speeds equal to, or greater than, 256 kbit/s. This includes cable modem, DSL, fiber-to-the-home/building, other fixed (wired)-broadband subscriptions, satellite broadband and terrestrial fixed wireless broadband. This total is measured irrespective of the method of payment. It excludes subscriptions that have access to data communications (including the Internet) via mobile-cellular networks. It should include fixed WiMAX and any other fixed wireless technologies. It includes both residential subscriptions and subscriptions for organizations.) The World Bank. (2019). Fixed broadband subscriptions (per 100 people). https://data.worldbank.org/indicator/IT.NET.BBND.P2?view=chart
- Creative goods (Total value of creative goods exports, net of re-exports (current US$) over total trade. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Creative goods. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Creative services (Creative services exports (% of total exports) according to the Extended Balance of Payments Services Classification. EBOPS 2010. (2019 data). Creative services. Retrieved from: https://unctadstat.unctad.org/wds/TableViewer/tableView.aspx
- Cultural occupations (Number of persons who have a cultural occupation in cultural industries expressed as a percentage of total

number of persons employed) UNESCO Institute for Statistics. (2019). Creative occupations. Retrieved from: http://data.uis.unesco.org

- World Population Review. (2013). World's most and least ethnically diverse countries. Retrieved from: https://www.washingtonpost.com/news/worldviews/wp/2013/05/16/a-revealing-map-of-the-worlds-most-and-least-ethnically-diverse-countries/
- Ease of business (Ease of doing business ranks economies from 1 to 190, with first place being the best. A high ranking (a low numerical rank) means that the regulatory environment is conducive to business operation. The index averages the country's percentile rankings on 10 topics covered in the World Bank's Doing Business) The World Bank. (2019). Ease of business. Retrieved from: https://data.worldbank.org/indicator/IC.BUS.EASE.XQ?view=chart
- Electrification (Access to electricity (% of population)) The World Bank. (2019). Electrification. Retrieved from: https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS
- Feature films (Nationally-produced feature films: Nationally-produced feature film includes 100% nationally-financed feature films and international coproductions. It excludes production of foreign feature films of which nationally-registered producers do not have any ownership.) UNESCO Institute for Statistics. (2019). Feature films. Retrieved from: http://data.uis.unesco.org
- Foreign investment (Foreign direct investment, net inflows (% of GDP). Foreign direct investment are the net inflows of investment to acquire a lasting management interest (10 percent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the balance of payments. This series shows net inflows (new investment inflows less disinvestment) in the reporting economy from foreign investors, and is divided by GDP.) The World Bank. (2019). Foreign investment. Retrieved from: https://data.worldbank.org/indicator/BX.KLT.DINV.WD.GD.ZS?view=chart
- GDP (GDP per capita (current US$)) UNESCO

Institute for Statistics. (2019). GDP. Retrieved from: http://data.uis.unesco.org

- Domestic credit (Domestic credit to private sector refers to financial resources provided to the private sector by financial corporations, such as through loans, purchases of nonequity securities, and trade credits and other accounts receivable, that establish a claim for repayment. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Domestic credit. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Human capital (This pillar gauges the human capital of economies) Cornell Insead WIPO. (2019). Global Innovation Index: Human capital. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- ICT access (The ICT access index is a composite index that weights five ICT indicators (20% each). It is the first sub-index in ITU"s ICT Development Index (IDI). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: ICT Access. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- ICT services imports (Telecommunications, computer and information services (% of total trade) according to the Extended Balance of Payments Services Classification EBOPS 2010. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: ICT services imports. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Innovation output (2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Innovation output. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Institutions (2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Institutions. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Mobile apps (Global downloads of mobile apps, by origin of the HQ firm, scaled by PPP$ DGP (billions). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Mobile apps. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Patent applications (Number of resident patent applications filed at a given national or regional patent office (per billion PPP$ GDP). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Patent

applications. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator

- Patent families (Number of patent families filed by residents in at least two offices. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Patent families. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Research talent (Professionals engaged in the conception or creation of new knowledge, products, processes, methods, and systems, as well as in the management of these projects, broken down by the sectors in which they are employed. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Research talent. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Rule of law (2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Rule of law. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Scientific publications (Number of scientific and technical journal articles (per billion PPP$ GDP). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Scientific publications. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- STEM assessment (PISA average scales in reading, mathematics, and science. 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: STEM assessment. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Top-level domains (Generic top-level domains (gTLDs) (per thousand population 15–69 years old). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Top level domains. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- University-industry collaboration (Average answer to the survey question: In your country, to what extent do people collaborate and share ideas in between companies and universities/research institutions? [1 = not at all; 7 = to a great extent]) Cornell Insead WIPO. (2019). Global Innovation Index: University-industry collaboration. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Wikipedia edits (Wikipedia yearly page edits (per million population 15–69 years old). 2019 data) Cornell Insead WIPO. (2019). Global Innovation Index: Wikipedia

edits. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator

- Global Innovation Index (2013-2019 data). Cornell Insead WIPO. (2019). Global Innovation Index. Retrieved from: https://www.globalinnovationindex.org/analysis-indicator
- Hospitals (Total density per 100 000 population) World Health Organization. (2016). Health infrastructure. http://apps.who.int/gho/data/view.main.30000
- Life expectancy (Life expectancy at birth, total (years)) UNESCO Institute for Statistics. (2019). Life expectancy. Retrieved from: http://data.uis.unesco.org
- Literacy rate (Adult literacy rate is the percentage of people ages 15 and above who can both read and write with understanding a short simple statement about their everyday life.) The World Bank. (2019). Literacy rate. https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?view=chart
- Mobile subscriptions (Mobile cellular subscriptions (per 100 people)) The World Bank. (2019). Mobile subscriptions. https://data.worldbank.org/indicator/IT.CEL.SETS.P2?view=chart
- Mortality rate (Mortality rate, infant (per 1,000 live births)) UNESCO Institute for Statistics. (2019). Mortality Rate. Retrieved from: http://data.uis.unesco.org
- Population (Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. The values shown are midyear estimates.) The World Bank. (2019). Population. https://data.worldbank.org/indicator/SP.POP.TOTL
- Poverty ratio (Poverty headcount ratio at $1.90 a day (2011 PPP) (% of population)) The World Bank. (2019). Poverty ratio. https://data.worldbank.org/indicator/SI.POV.DDAY?view=chart
- R&D expenditures (gross domestic expenditure on R&D (GERD), GERD as a percentage of GDP) UNESCO Institute for Statistics. (2019). R&D in GDP. Retrieved from: http://data.uis.unesco.org
- Rural population (Rural population (% of total population)) UNESCO Institute for Statistics. (2019). Rural population. Retrieved from: http://data.uis.unesco.org
- School enrollment, tertiary (Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group

that officially corresponds to the level of education shown. Tertiary education, whether or not to an advanced research qualification, normally requires, as a minimum condition of admission, the successful completion of education at the secondary level.) The World Bank. (2019). School enrollment. https://data.worldbank.org/indicator/SE.TER.ENRR?view=chart

- Secure internet servers (Secure Internet servers (per 1 million people)) The World Bank. (2019). Secure internet servers. https://data.worldbank.org/indicator/IT.NET.SECR.P6?view=chart

Heilmeier Questions [HQ]

Proposal must answer all 9 of these, or we lose a ton of points. Michael (Miggs) will combine the answers to these as part of Project Summary and Research Survey.

1. What are you trying to do? Articulate your objectives using absolutely no jargon. *We are trying to discover through data analysis the definition of what makes a nation innovative, as well as common factors that could influence the innovativeness of nations around the world (economics, lifestyle, education, health). We do not know for sure if these links exist; we are looking to discover the answer through this project.*
2. How is it done today; what are the limits of current practice? *There are many ways in which innovation of a nation can be measured; we are hoping to utilize a variety of measures to produce a unique picture of what makes a nation innovative as well as identifying potential factors that could have contributed to that result (either correlative or causative).*
3. What's new in your approach? Why will it be successful? *We are utilizing a unique combination of datasets [EXAMPLES] to define the concept of what makes a nation innovative as well as investigating other factors in new ways. Our visualization will express this information in an easily-digestible format which will effectively communicate a large degree of information to users.*
4. Who cares? *Government officials and potential investors in developing areas of the world would find this data interesting in terms of understanding how and where to focus potential investment.*
5. If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)? *Our tool would be useful for research and investment purposes to identify characteristics of developing nations that could be primed for significant investment. (NOTE: We should look carefully at certain nations in Africa and SE Asia which are currently receiving large investments in infrastructure and technology and see how those changes are impacting those nations based on our innovation index). We*

could verify the data through identifying whether the increasing the related factors for innovation show an improvement in our innovation index. We could complement this further data analysis with user studies, etc.

6. What are the risks and payoffs? *Risks: causation versus correlation - are the measures we are taking actually responsible for innovation, or is innovation responsible for the increases in those measures (e.g. better innovation in medicine leads to better health outcomes)? Recency of studies. However, if a direct link can be proven out through the analysis of the provided data, this could provide a unique viewpoint into how and where to invest in nations around the world.*
7. How much will it cost? *We are using publicly available datasets and research in the preparation and development of our project; the only real cost is time.*
8. How long will it take? *Based on the deliverable timelines of our project, the maximum amount of time this analysis will take is approximately eight weeks.*
9. What are the midterm and final "exams" to check for success? How will progress be measured? *The first checkpoint will be building the innovation index itself and beginning to compare that number against other related datasets. The visualization which showcases innovation of nations around the world and the potential factors related to that finding will serve as the final deliverable.*

Paper Author	Main Idea	Useful or Not	Shortcoming	Category
s, Title, Year		& Why	s	
Polatcan, M., & Balcı, A. (2019). Social Capital Wealth as a Predictor of Innovative Climate in Schools. *International Journal of Contemporary Educational Research*, *6*(1), 183-194.	Relationship found between social capital and innovative climate	Interesting, potentially applicable in the measures for health; mental health in particular (social capital <-> mental health)	Measuring social capital at societal level is difficult	QI
Crespo, N. F., & Crespo, C. F. (2016)	Conditions for increased innovation are	Could be used to explain differences	Finding an appropriate criteria to divide countri	DF

. Global innovation index: Moving beyond the absolute value of ranking with a fuzzy-set analysis. *Journal of Business Research*, *69*(11), 5265-5271.	different between countries of varying incomes (low income has higher variation in solutions)	in findings between low and high income countries, may necessitate further breaking apart of results	es. Where is the 'cutoff'?	
Dakhli, M., & De Clercq, D. (2004). Human capital, social capital, and innovation:	Human capital versus innovation (measures like patents, expenditures, high-tech export	Human capital has been shown to affect innovation measures	Small dataset (59 countries)	QI

a multi-country study. *Entrepreneurship & regional development*, *16*(2), 107-128.	)			
Rauch, A., Frese, M., Wang, Z. M., & Unger, J. (2010). National cultural values, firm's cultural orientations, innovation, and performance:	Investigated entrepreneurial orientation and cultural values' relationship to innovation	Again, useful measures for looking at creativity and innovation	Another small dataset, but data uncovered is more detailed than most resources so far.	QI

testing cultural universals and specific contingencies across five countries. *Frontiers of Entrepreneurship Research*, *30*(15), 4.				
Pratt, A. C., & Jeffcutt, P. (2009). *Creativity, innovation and the cultural economy*. Routledge.	Cultural economy consists of creative industries that contribute to innovation]	13 creative industries cited as contributors to predicting innovation	Based on 2009 data, so may be limited in terms of 'big tech', may require updating	QC
Dachs	Relati	Measu	Small	QI

, B., Ebersberger, B., & Lööf, H. (2008). The innovative performance of foreign-owned enterprises in small open economies. *The Journal of Technology Transfer*, *33*(4), 393-406.	onship between foreign ownership and innovation.	rable predictive feature.	dataset (5 Nordic countries).	
Fagerberg, J., Srholec, M., & Verspagen, B. (2010). Innovation	Innovation means different things based on economic context (OECD vs	Can use different attributes to predict innovation	One model may not cover all cases	DF

and economic development. In *Handbook of the Economics of Innovation* (Vol. 2, pp. 833-872). North-Holland.	developing countries)			
William R. DiPietro, Emmanuel Anoruo, Creativity, innovation, and export performance, Journal of Policy Modeling, Volume 28,	Export performance has a positive relationship with creative activity	Relates creative activity to economic growth	A bit out of date - exports are digital more often now	QC

Issue 2, 2006, Pages 133-139, ISSN 0161-8938				
López-Claros A., Mata Y.N. (2010) The Innovation Capacity Index: Factors, Policies, and Institutions Driving Country Innovation. In: The Innovation for Development Report 2009–2010.	Innovation Capacity Index (ICI) computes innovation based on policies and institutions.	Well-defined framework of measuring innovation that we can use as a starting point.	Doesn't factor in all of what we're looking at. Policies and institutions vary based on nations.	QI

Palgrave Macmillan, London				
Sam Youl Lee, Richard Florida & Gary Gates (2010) Innovation, Human Capital, and Creativity, International Review of Public Administration, 14:3, 13-24	Innovation is a joint product of human capital and creativity.	Measures of creativity (bohemian index) and diversity (gay index) are generated.	Only focuses on human capital. What other factors are at play?	QI
Williams, L.K., McGuire, S.J. Economic creativity and innovation implementation: the entrepreneurial drivers of growth? Evidence from 63 countries. Small Bus Econ 34, 391–412 (2010).	NA	NA	NA	
Hsu, P., Tian, X., & Xu, Y. (2014). Financial development and innovation: Cross-country	Relationship between financial market and technological innovation	Economics is shown to influence tech innovation (patents & R+D expenses)	There may be other metrics that could impact these measures.	QI

evidence. *Journal of Financial Economics*, *112*(1), 116-135.				
Almeida, P., & Phene, A. (2004). Subsidiaries and Knowledge Creation: The Influence of the MNC and Host Country on Innovation. *Strategic Management Journal*, *25*(8/9), 847-864.	Innovation relationship with technological richness of int'l companies, subsidiary knowledge related to host country, technical diversity with host country	Maps back to papers showing relationship between level of innovation and foreign companies (foreign ownership)	Old data	QI

Griffiths, M., Gundry, L., Kickul, J. and Muñoz Fernandez, A. (2009), "Innovation ecology as a precursor to entrepreneurial growth: A cross-country empirical investigation", Journal of Small Business and Enterprise Development, Vol. 16 No. 3, pp. 375-390.	Governmental, economic, and technological factors in innovation (resources, human capital, seed funding)	New variables like seed funding introduced.	Old data, 1995-2005	QI
Saint-Paul, G., & Verdie	Paper gives a formul	We want to look at	Old data (90s and	

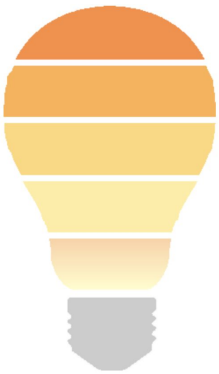
r, T. (1993). Education, democracy and growth. *Journal of development Economics*, *42*(2), 399-407.	a on how their statistics follow that voting, education, growth and income rise together.	relationships between wealth and education, this shows they rise together	before)	
Gault, F. (2018). Defining and measuring innovation in all sectors of the economy. *Research policy*, *47*(3), 617-622.	applies a systems approach to measure innovation across all economic sectors, not just the business sector	we can extend our analysis to other types of innovation, including public sector and household sector innovation. Especially important as developing countr	lack of consistent datasets across countries across sectors	QI

		ies have smaller business sectors.		
Roszko-Wójtowicz, E., & Białek, J. (2016). A multivariate approach in measuring innovation performance. *Zbornik radova Ekonomskog fakulteta u Rijeci, časopis za ekonomsku teoriju i praksu-Proceedi*	it is a good idea to combine similar innovation predictors into a synthetic variable through the use of cluster or factor analysis in order to avoid coliniarity.	this provides a methodology to derive the most impactful variables	this was done using European data, will attempt to extend this to other regions.	

ngs of Rijeka Faculty of Economics, Journal of Economics and Business, *34*(2), 443-479.				

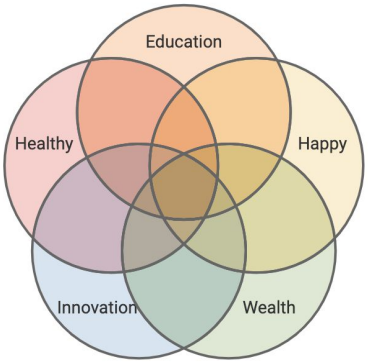
Visualization Concepts

These visual concepts are ideas that may be implemented given time and possibility.

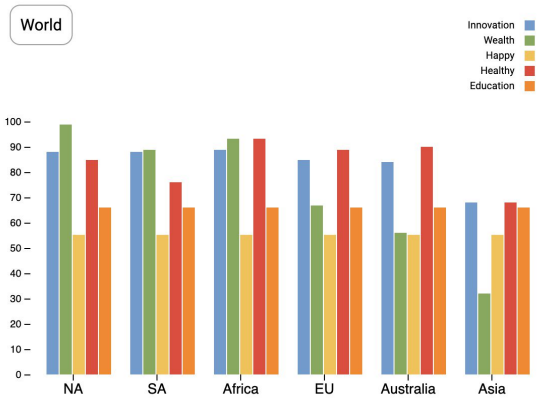


Innovation “pie chart”, show the top 5-10 countries and everyone else as a slice of the lightbulb. This gives the main point of innovation and who is doing the best at innovating in the world.

Finding the influx point of average income and level of education that jumps the country into “innovative”. If we find there’s a gulf between innovative and not, we can find the other values. Finding a statement that someone can compare to themself individually makes this more impactful.



Depending on what and how many indices we have, have a venn diagram show an example of each section in the graph that fits those overlaps. Above example with 5 circles. Having a graph like this points examples of countries at different levels on each of the indices.

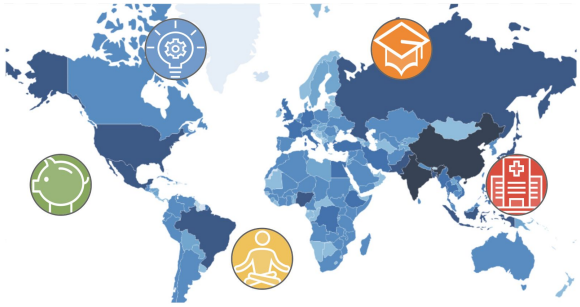


A bar chart for regions showing indices next to each other. Easy to combine their indices from each county located in the region to look at a glance. This is an example of a bar graph with multiple bars per region. When you click the region, it shows the countries in that region itself. When you click on the country itself it shows only that country information for the

five different bars (health, happy, wealth, education, innvoation).

Amazon Alexa Skill: Alexa, according to the Poloclub…

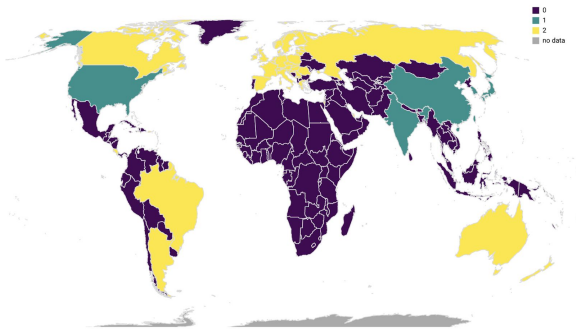
- Who is the world's most innovative country?
- What region of the world is most innovative?
- What country is happy, healthy, wealthy and wise?
    - Is that same country the most innovative?



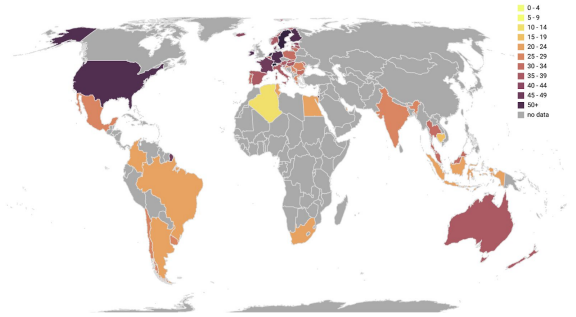
Icons in each of the circles that can be pressed to show values of their indices between wealth / education / health / mental health and innovation, each button is the information you turn on or off in the main graph. The main graph could be a map. This is the most interactive graph where all information would be available to the user.

Visualizations for Analysis

Maps will be used to find visually where the clustering is located geographically. The first graph is from the unsupervised model.

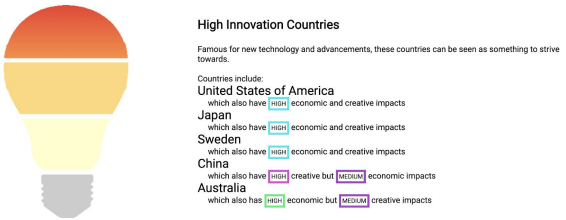


The next map is from the supervised model.

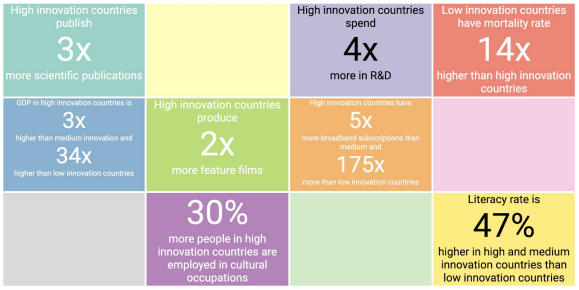


Visualizations for Conclusions

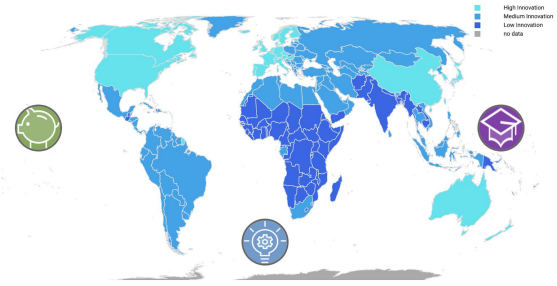
The first conclusion was from the three clusters of low, medium and high innovations. We found countries in each cluster and looked at their specific economic and creative impacts and saw if they were less or more than the overall innovation in a lightbulb split into three parts.



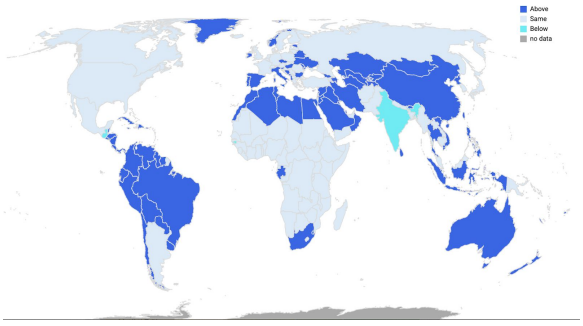
Next, we had a lot of data that was interesting to tell a story such as showing differences from low to high innovation and what properties they had with our data. These are sentences that put our data into context to a user. Each specific value is spinning until the user hovers over the number to reveal its value.



The two maps used in our conclusion are putting all data from the unsupervised model to show each low, medium, and high of innovation in overall combined, economic and creative. You may also filter out countries which are not in the cluster.



The last map is the difference between each of these low, medium and high clusters between each of the different categories, such as overall combined, economic and creative.



Testbed

Your testbed is how you are performing an experiment.

>> I'm assuming that each experiment will have its own testbed? One of the TAs listed these 6 items that a testbed should consist of. I put in values that, for the most part, can be used for each of our experiments.

1. Tools used: Python; D3; csv files
2. Data set \ subset of dataset used: TBD - based on what Minkwon, Marc, and Ricardo decide
3. What you are using to measure the results:
    - Adjusted R2 (assuming we're doing a regression): Does adjusted R2 increase as cultural indicators are added to a base model (which has economic indicators).
    - Clustering model: compare the results of the clusters with the results of the supervised model (if the supervised model returns numbers, then to categorize those). Serves as a rough check for supervised model and to determine if the response variable used for the model is sufficient in representing innovation.

4. What results you are trying to measure
    - Trying to determine the predictors/factors that could be used to indicate the innovativeness of a country
    - Determine/define a metric for innovativeness
5. Any controls in place
    - Not sure. If anything, we can use the model with economic/financial data only as one of our controls, and compare those models with those that add cultural indicators.
6. Any other details about how you are setting up a reproducible experiment:
    - Using publicly available data
    - Setting the randomness seed to return reproducible results from the models
    - Documenting the cleansing process (possibly making assumptions about

countries that we don't have much data for / interpolation) and the models

List of experiments:

- Define a potential metric for innovativeness (through a single factor or aggregation).
- Comparing the effects of including/excluding cultural indicators in a supervised model, with the potential innovativeness metric as the response variable. Use feature extraction to determine the best factors to use.
- Using an unsupervised model (clustering?) to group countries and determine the factors that should be used in the clustering analysis. If these factors are correlated with innovation, does the clustering analysis reflect that?

20200321 Meeting Notes

-PCA/feature extraction to determine most relevant predictor variables (probably tied to the supervised models)

-Split up economic and cultural datasets

- Potential benchmark -- global innovation indicator
>>Potential: Do cultural indicators add to the model -- do just economic and see if adding cultural adds value

1.Economic + cultural >> regression/supervised

- To deal with the response variable (a bunch of different factors; one or a aggregate of different variables)
- Determine if the variable is sufficient for response indicator for innovation or if need to be joined

2.Cultural + Economic >> clustering/unsupervised

- Serves as a rough check for the results of the supervised/to check the response variable
- Don't have a good innovation measure
    - Visualization is useful here

3.Country node distance >> making assumptions about countries that we don't have much data for / interpolation