Towards an Interactive Learn-to-Rank System for Economic Competitiveness Understanding

Caitlin Kuhlman
Worcester Polytechnic Institute
cakuhlman@wpi.edu

Elke Rundensteiner
Worcester Polytechnic Institute
rundenst@wpi.edu

ABSTRACT

Ranking models are useful tools often employed to aid in decision making. In fields such as economics, the development of indicators to rank economies or regions are typically dictated by *expert opinion*. With the increased availability of high fidelity open data, better tools for developing and understanding rankings can provide valuable insight into social and economic questions. This paper presents a preliminary foray into the development of such tools. We introduce a vision for leveraging state-of-the-art algorithms from the Information Retrieval field to design interactive learn-to-rank tools. Incorporated into data analytics systems via plug-and-play components, such tools hold the potential to better evaluate the comparative merits of different regions and to interactively assess the impact of different features on the final rankings. The *MyRanker* paradigm is applied in the context of MATTERS, a public system for evaluating the economic competitiveness of US states. A preliminary analysis and discussion of the system highlight its promise for ranking analysis.

KEYWORDS

Learning algorithms, Interactive Visualizations, Rankings, Machine Learning

ACM Reference format:
Caitlin Kuhlman and Elke Rundensteiner. 2017. Towards an Interactive Learn-to-Rank System for Economic Competitiveness Understanding. In *Proceedings of KDD 2017 Workshop on Interactive Data Exploration and Analytics (IDEA'17), Halifax, Nova Scotia, Canada, August 14th, 2017,* 9 pages.
DOI:

1 INTRODUCTION

Rankings are a fundamental tool used in many applications to help people understand the relative merit of objects or choices. They are commonly employed to simplify decision making when the number of factors impacting choice is large. In economics, a variety of rankings or indicators are used to gauge the relative performance of countries and regions [5, 17, 18]. These indicators are formed by combining various statistics such as tax codes, GDP, population, and so on with the aim to measure a relative ranking among objects according to economic principles. Another example of ranking for decision making is the use of college rankings [8, 20], which are designed to help potential students choose which school to attend by accounting for factors such as student demographics, location, and salary outcomes for graduates. In fact aggregated statistics are used widely for everything from valuing the stock market via indices such as the S&P 500 and NASDAQ to ranking restaurants and other businesses via social media sites such as Yelp.

Machine learning techniques to automatically rank objects have been extensively researched in the context of Information Retrieval [13]. Indeed, efficient web search is facilitated by learning-to-rank algorithms which evaluate the relevance of online documents to a given user query. This type of user-driven ranking has been on the front lines of the information revolution, broadly providing high fidelity results over data resources otherwise too large to browse.

With advances in data science and the proliferation of high quality open data come opportunities to leverage these machine learning methods to answer social and economic questions. Interest in data science and knowledge discovery techniques as applied to social sciences and economics is growing [11, 18]. The development of robust tools is needed to enable non-expert users to better understand the explanatory power of machine learning models for socioeconomic outcomes. Such tools can ultimately greatly increase the utility of open data for social good.

In this work we explore the use of machine learning to aid in the construction and understanding of ranking models. Powered by learning-to-rank machine learning [13], we introduce a new paradigm for interactive exploration to aid in the understanding of existing rankings as well as facilitate the automatic construction of user-driven rankings. Components are incorporated into a plug-and-play framework. We demonstrate how the framework can be applied to the problem of measuring and analyzing economic competitiveness through the development of a prototype data analytics system. Interactive learn-to-rank tools extend the capabilities of MATTERS [1], an online platform designed to evaluate the relative competitive advantages of US states.

1.1 Motivation: Rankings and Their Pitfalls

One of the most common applications of data to social science and policy is in the field of economics, where measurements such as rankings serve a critical role in evaluating the economic health of regions and shaping policy. For example, the so-called "Misery Index" [2] is a historical measure of economic performance developed by economist Arthur Okun. This simple index is computed as the sum of the seasonally adjusted unemployment rate to the annual inflation rate. Many such indicators have been designed to measure past economic performance or predict future economic conditions.

[1]http://matters.mhtc.org/
[2]http://www.miseryindex.us/

KDD 2017 Workshop on Interactive Data Exploration and Analytics (IDEA'17), Halifax, Nova Scotia, Canada

DOI:

Regional rankings are often employed to quantify relative economic competitiveness. For instance, the World Economic Forum publishes an annual Global Competitiveness Report [17] containing a ranking of countries around the world computed from numerous metrics and survey data. In the US, other indices compare the competitiveness of different states, such as the Milken Institute's State Technology and Science Index [5] or CNBC's Top States for Business [3] ranking.

The design of such indices depends heavily on "expert opinion", which drives the selection of the metrics and weightings used in their construction. While extremely valuable, the dissemination of expert knowledge through such rankings is somewhat limited. One, it may be that the actual formula used to compute the ranking is not made public. Or, though published, the model is unlikely to be accessed by the average consumer. By only considering the final ordering given by the ranking, consumers are limited in their ability to gain further insight into the implications of the ranking. Another consideration in the design of such ranking models is confirmation bias. Even with the best of intentions, evaluations originating in expert opinion may succumb to this common phenomena [15].

Furthermore, latent factors incidentally measured by a ranking are not always made explicit. The data used in the design of the ranking may be serving as a proxy for undesirable metrics. For instance, when considering the economic competitiveness of different regions, evaluating the quality of the available talent pool based on race or gender would be highly undesirable. For this reason, a ranking model should be evaluated in the context of all available data, and compared with rankings based on demographic information to ascertain whether it is reflecting inherent bias in the underlying datasets.

At times, the designers of ranking systems may struggle to avoid these pitfalls, and to determine the most useful and fair data upon which to base rankings. The incorporation of learning-to-rank algorithms into highly usable interfaces is crucial to help users better understand existing rankings, and to aid them in the creation of ranking models which reflect their intuition and value system. In this work we thus propose new interactive tools to explore the construction of such indices.

1.2 MyRanker: Our Proposed Interactive Learn-to-Rank Paradigm

In this work, we investigate the question: *Are there data-driven approaches to the construction of rankings that would allow designers to gain more insight into the concepts they attempt to model?* To address this question, we propose the design of an interactive exploratory paradigm for both the construction of rankings, as well as better understanding of existing ranking models, here referred to as MyRanker. Our easy-to-use rank construction tools in MyRanker allow stakeholders to drive the ranking process. They can either manually specify criteria for their preferred rankings via a visual interface, or leverage rank learning algorithms from the machine learning field [13] to automatically derive rankings based on partial information based on their domain knowledge or priorities. Intuitive visual interfaces allow users to design rankings which reflect their intuition or meet their objective simply by indicating partial preferences over objects in the dataset.

Further, we study the question: *Can we more closely couple the underlying data and the resulting ranking, so that users better understand the impact of data on the relative ranking of objects?* For this, visual rank exploration tools in MyRanker are offered to interactively explore, compare, and analyze these newly constructed rankings. Multiple visual displays are closely interlinked – visualizing both the relative ordering and the detailed description of the objects being ranked. Direct display of any adjustment of criteria on the resulting ranking can bring insights into the relative impact of particular data on a particular ranking – with promise of a high value return.

This seamless integration between rank construction tools and rank exploration tools into one single easy-to-use analytics system empowers users to gain insights into the differences between ranking models. Incorporated into a plug-and-play framework, our MyRanker paradigm enables users to better understand the interplay of the data and their rankings through integrated data visualizations and interactions.

1.3 Contributions

The contributions of this work including the following:

(1) This paper introduces the notion of an interactive paradigm for learn-to-rank tools supporting the process of exploring and understanding rankings. This considers both the ease of the specification of rankings via visual support as well as the display of ranking results.
(2) We present a plug-and-play framework, called MyRanker, to support stakeholders to interact with and thus understand ranking models. MyRanker integrates rank learning components into a data analytics system.
(3) We describe a demonstration of our interactive learn-to-rank tool, MyRanker, applied for economic competitiveness evaluation as part of the Massachusetts Technology, Talent, and Economic Reporting System (MATTERS). Integrated components for user interaction and rank learning are seamlessly incorporated into this analytics platform.

2 LEARNING-TO-RANK BACKGROUND

In the Information Retrieval (IR) field, a number of methodologies for learning-to-rank have been developed. In [13], Lui et al. categorize 3 different supervised learning approaches: pointwise, pairwise, and listwise. The pointwise approach reduces to a regression analysis, where a model is trained on instances which each have either a numeric or ordinal score assigned to them. A ranking of unseen data is determined based on the regression values given by the model. Listwise approaches learn based on entire sets of ordered objects. That is, the training set consists of multiple ordered lists with corresponding rankings, and the model assigns an ordering over an entire previously unseen list [4, 19].

In the pairwise approach, training data is composed of pairs of objects. Labels are assigned to each pair which indicate a preference between them. For instance, given a pair (a, b) it is assigned label 1 is a is preferred to b, label -1 if b is preferred to a, and 0 if the two instances are equally preferred. Datasets which consist of individual

[3] http://www.cnbc.com/americas-top-states-for-business/

objects and labels can be transformed to this pairwise format by forming every possible ordered pair and comparing their labels. In [7], the pairwise approach is shown to reduce the learning-to-rank problem to a binary classification problem. Given this, any classification model can be employed to learn a ranking. Proposed classification models, to just name a few, include SVM [7], neural networks [3], and regularized least-squares [16].

The techniques developed in IR are intended to present only the most relevant results ranked based on user queries to a search engine. This task has a few distinguishing features from ranking in other contexts. For one, the amount of data in this problem is large. Thousands of documents may be returned for a single query, and prohibitively many features could be extracted from each. Models can be trained on a huge corpus of text. For data at this scale, certain methods may have an advantage over others. For instance, neural networks can leverage and in fact improve when applied to huge training sets [3]. For many other problems, such as social or economic evaluations, the data is likely to not be so large. While much public data are available, often for analytics tasks it is cleaned and preprocessed and only a small relevant set of data is used to determine the final outcome. In such cases models which require large amounts of training data will not perform well.

Another consideration is that in the search context, often the task is to find only the top results, not necessarily a complete ordering of all objects. For a thousand documents returned for a query, a user will likely only be interested in a handful of the most relevant results. Therefore a number of evaluation metrics have been developed which favor correct results at the top of a list and are forgiving of mis-ranked data toward the end of the list [9]. Such measures are not appropriate in other contexts, where the position of all objects in the ranking is of keen interest.

3 USER-DRIVEN RANKING: THE MYRANKER FRAMEWORK

Since the design of rankings is usually intended to capture some quality of the objects being ranked that cannot be directly measured, it is by its nature a difficult task for users to perform. Our MyRanker interactive learn-to-rank framework thus alleviates the need for the user to assign an explicit value as rank to each object in the dataset. Instead, by asking users to assign preferences between a select subset of objects (possibly those objects that they are personally familiar with), the system learns from their mental model of the problem. This empowers the users to construct rankings automatically using only partial information.

MyRanker solicits hints from the user in the form of a preference assignment between pairs of objects. This approach lets users tap into their expertise or intuition about the concept they are trying to capture in their ranking – yet without the need to manually construct an entire ordering. We employ the pairwise learning-to-rank approach to facilitate user-guided ranking analysis. Users assign priorities to pairs of objects via sample pairs, e.g., (object1 > object2). The system then learns a pairwise model which determines both a continuous value for each object and a resultant ranking.

The interactive learn-to-rank system is realized as a plug-and-play framework for ranking analytics, i,e., modules such as the learning algorithm as well as displays can be easily switched out.

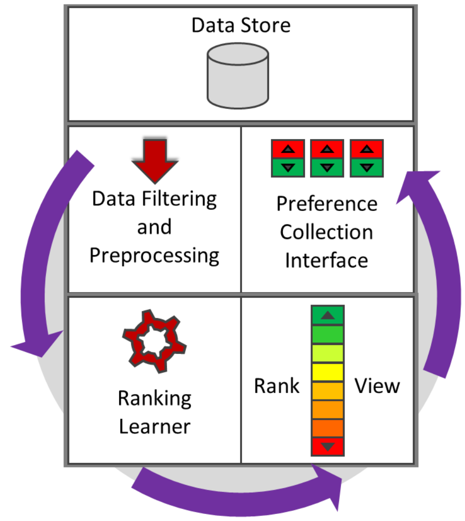


Figure 1: MyRanker Framework

We target applications where the data may be relatively small and is available from a data warehouse. Figure 1 sketches the system architecture of the MyRanker framework. It consists of a data repository, data filtering and preprocessing module, preference collection module, rank learning model, as well as a visual analytics interface. The visual analytics interface presents results, both rankings as well as their underlying data, to users through a number of interlinked displays. Upon visual inspection, users can manually update their preference and metric selections or individual weightings of metrics to refine the learned model in an iterative fashion.

4 MYRANKER APPLIED TO ECONOMIC COMPETITIVENESS ANALYSIS

The Massachusetts Technology, Talent, and Economic Reporting System (MATTERS) is an online public tool developed at Worcester Polytechnic Institute under the guidance of the Massachusetts High Technology Council[4] – in partnership with numerous stakeholders and domain experts. The goal of MATTERS [14] is to better understand and measure the economic competitiveness of US states using open data. To achieve this goal, MATTERS consolidates a rich collection of publicly available socioeconomic datasets. By making over 50 datasets available in one place for the first time, the system empowers decision makers from government officials to company executives to evaluate the economic conditions in their state in contrast to other states.

Developed with experts from high technology industry, research organizations, and higher education institutions, MATTERS provides descriptive analytics for the data in the system. In addition, automated web extraction tools and administrative easy-to-use data curation tools have been developed by WPI [6]. These data curation tools have now been made available to external partners, namely, teams of students at Brandeis University for continued data curation into the MATTERS warehouse twice a year. MATTERS

[4] http://mhtc.org

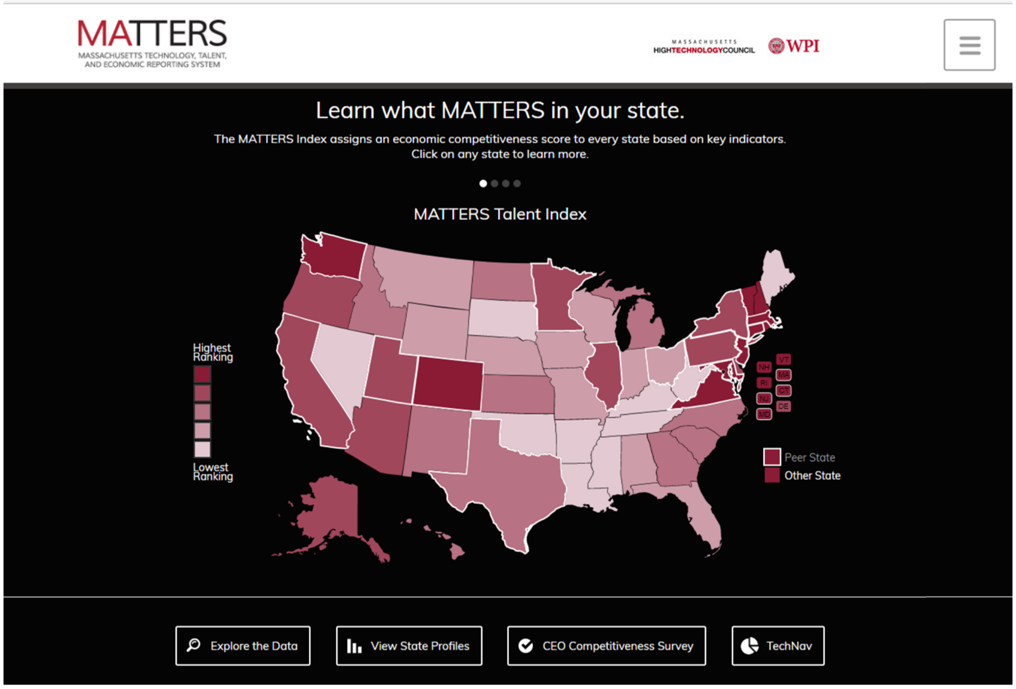


Figure 2: The Massachusetts Technology, Talent, and Economic Reporting System (http://matters.mhtc.org)

also provides a public-facing API to facilitate data reuse by other researchers. [5]

MATTERS is not designed to be prescriptive in scope, rather this open data repository includes a suite of tools which allow for user-driven data exploration [1]. Therefore, the MATTERS system provides the perfect test-bed for new ranking analysis features. The MyRanker framework applied in this context provides additional exploratory tools to aid in the understanding of existing rankings, as well as to facilitate the automatic construction of new user-driven ranking models for states. Seamless integration of MyRanker into MATTERS capitalizes on the customized visualizations in the MATTERS system, which are designed to provide insights specific to the spatio-temporal nature of the data. This allows for comparison of rankings to discover correlations, or observe changes in rankings of certain regional areas or over time.

4.1 Rank Specification Tools

Economic indices represent cumulative data over different related datasets. Beyond carefully crafted indices based on expert knowledge, the MATTERS rank specification tools now also empower users to combine data and perform complex analysis themselves in an interactive fashion. The data in the system is comprised of existing rankings (including 4 MATTERS Indices), as well as a collection of other (raw) data related to measuring economic competitiveness. The data is organized under 4 categories: Talent, Cost of Doing Business, Tax Climate and Quality of Life. Easy to use interfaces allow for the selection of individual metrics of interest to the user. Rank specification tools create new rankings based on these selections, via manual weightings of individual metrics (Figure 3) or pair preferences (Figure 4).

Custom rankings are stored as numeric formulas retrievable through user accounts. This way, the rankings are kept up-to-date as new data becomes available. The system regenerates data according to the user-defined rules each time the index is requested. For use in custom rankings, the following strategies are used to clean and standardize the data in order to produce meaningful results:

Missing Values. The MATTERS system contains data for multiple years, and the availability of each dataset may vary. A state ranking is computed for each year that at least one selected metric is available. If a value for some metric is missing for a given year, the closest previous value is used. If there is no previous value, the closest possible value is used.

Normalization. The datasets in the system vary greatly. Some, such as tax rates, are percentages which vary only by a few tenths of a point, while others are numbers in the millions representing state populations, or GDP. The data must be normalized so that large values do not dominate. Therefore the data in the system is standardized by setting the mean of each metric to 0 and the standard deviation to 1.

Inverted Trends. Typically when looking at trends, high values are considered better than low values. However, for some data the opposite is true, as in a low unemployment rate being preferred to a high rate. For the manual construction of metrics in the Rank Builder tool, negative coefficients are used for data with inverted trends.

[5] http://matters.mhtc.org/api

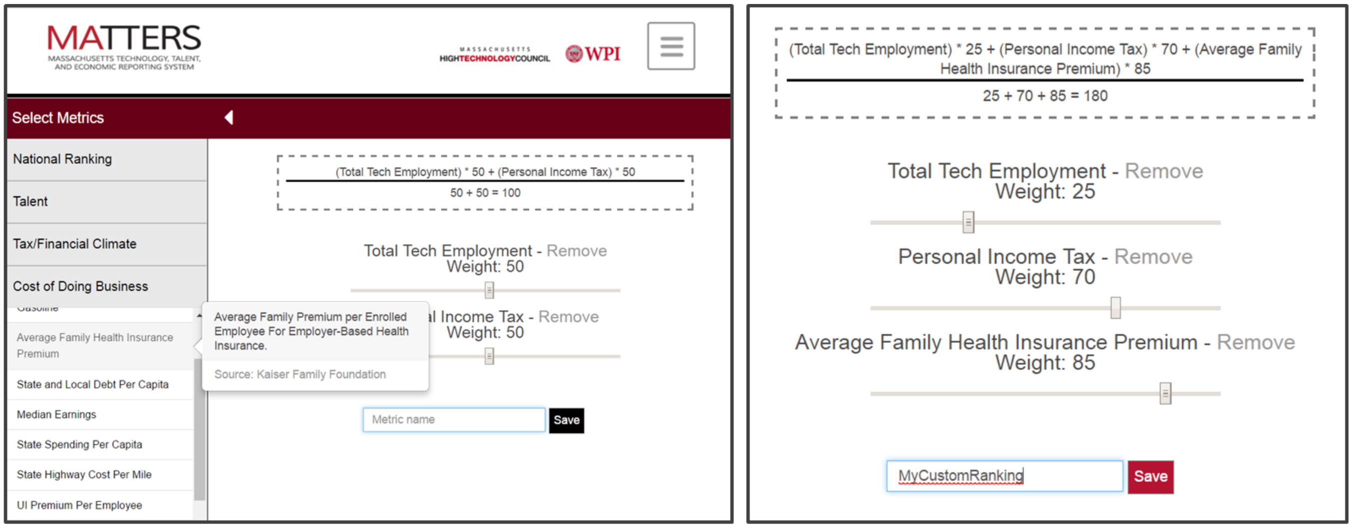


Figure 3: MATTERS manual Rank Builder tool.

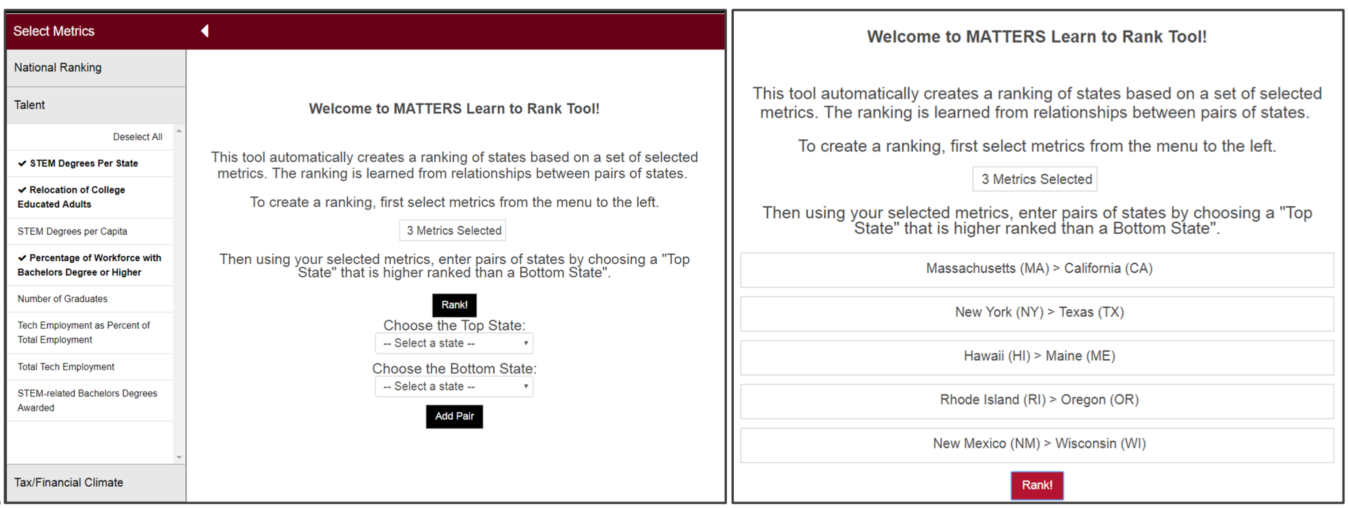
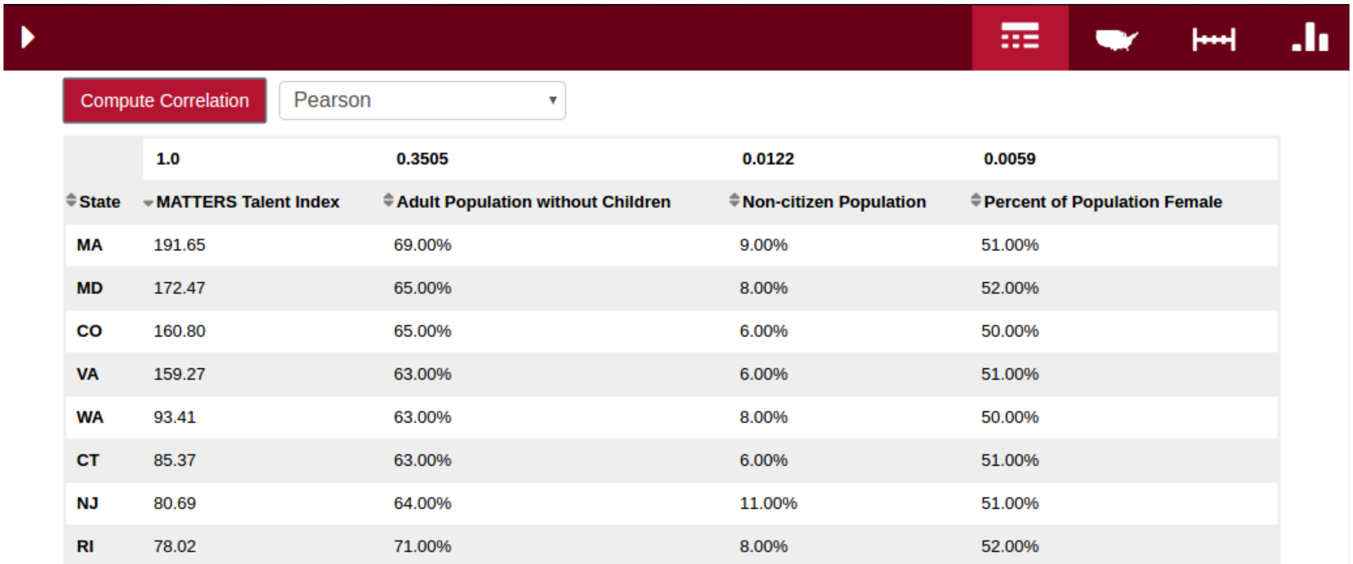


Figure 4: MATTERS pairwise Learn-to-Rank tool.

4.1.1 Rank Builder Tool. This novel interactive metric-creation tool was developed as an instrument for users to visually define their own indices. Users determine which datasets are of most interest and specify their relative importance to compose a compound index. Figure 3 shows the interface for the Rank Builder tool. Users can select metrics from the menu on the left, which lists datasets organized by category in collapsible menus. A description and source link are displayed when the user hovers over each dataset in the menu. Selected metrics are each displayed along with a slider tool. The screen on the right of Figure 3 shows how users can specify a weighting for each metric to indicate its impact on the desired state ranking. The formula for the resulting weighted average is shown at the top of the screen.

4.1.2 Learn-to-Rank Tool. This feature implements the pairwise rank learning component of the MyRanker framework. Figure 4 shows the interface to the MATTERS Learn-to-Rank tool. Metrics to be used in the automatic construction of a ranking are selected from the left-hand menu. Then, via the dropdown menus shown in the center screen, users can enter pairs of states. Preference is indicated by selecting a “Top State” and a “Bottom State”. Once the user has entered a series of state pairs (as shown on the right of Figure 4), the automatic learn-to-rank engine is run using the “Rank!” button. A global ordering of all states is automatically learned based on this partial input from the user.

Upon naming and saving it, users can view their resulting ranking model in the rank builder view described above. This allows them to examine the weightings learned for each underlying metric and the resulting overall formula for the ranking. Users can then further customize the model by adjusting the weights manually through the rank builder interface if desired.



	1.0	0.3505	0.0122	0.0059
State	MATTERS Talent Index	Adult Population without Children	Non-citizen Population	Percent of Population Female
MA	191.65	69.00%	9.00%	51.00%
MD	172.47	65.00%	8.00%	52.00%
CO	160.80	65.00%	6.00%	50.00%
VA	159.27	63.00%	6.00%	51.00%
WA	93.41	63.00%	8.00%	50.00%
CT	85.37	63.00%	6.00%	51.00%
NJ	80.69	64.00%	11.00%	51.00%
RI	78.02	71.00%	8.00%	52.00%

Figure 5: The MATTERS Talent ranking is compared with demographic data in the MATTERS table view.

4.2 Rank Views

Multiple visual displays in the MATTERS system provide further opportunity for evaluation and understanding of custom rankings. Displays offer descriptive analytics not only for all newly defined user rankings, but the entirety of the data in the MATTERS collection. This enables easy comparison between created rankings as well as between rankings and other data in the system. Customized views provide insight into the spatio-temporal nature of rankings created from data in the MATTERS system. Comparisons can be made across states and over time using the MATTERS Data Explorer feature. Figures 5 and 6 show some of the views available in the MATTERS system.

4.2.1 Table View. Custom rankings can be viewed in a table alongside other data in the MATTERS system. Data can be displayed in a number of configurations, varying the number of states, metrics, or years to be displayed. To further aid in understanding the relationships between data and rankings, correlation analysis is provided in the table view. A measure of the correlation between the first metric in the table with each of the other selected metrics is computed with the click of a button. Users can select a number of useful correlation measures including the Pearson Correlation Coefficient[12] and the Kendall Tau Coefficient [10] (Fig 5).

4.2.2 Timeline View. Time series analysis is provided for all MATTERS data including custom rankings in the timeline view. Users can select a set of states to view in the chart and compare their relative performance and changing trends over time (Fig 6a).

4.2.3 Map View. A choropleth map view allows users to compare rank values across all states at once. This view paints a picture of the distribution of rankings throughout the country at a glance. A sequential color scheme [2] allows users to easily discern and evaluate the relative performance of states and their geographic neighbors. A comparison of maps highlights the regional differences in rankings (Fig.6b).

4.2.4 State Profile View. Finally, MATTERS provides individual State Profile views. These displays show the values for each underlying metric contributing to the rank of an individual state. The colors red, yellow, and green indicate the performance of each metric as compared to the rest of the states, and trend arrows show whether the state has been improving or declining over time. This view can help users gain insight into the impact of each metric on the composite score for an individual state. Comparisons between profile views expose the differences in individual metrics which impact the relative performance of states (Fig. 6c).

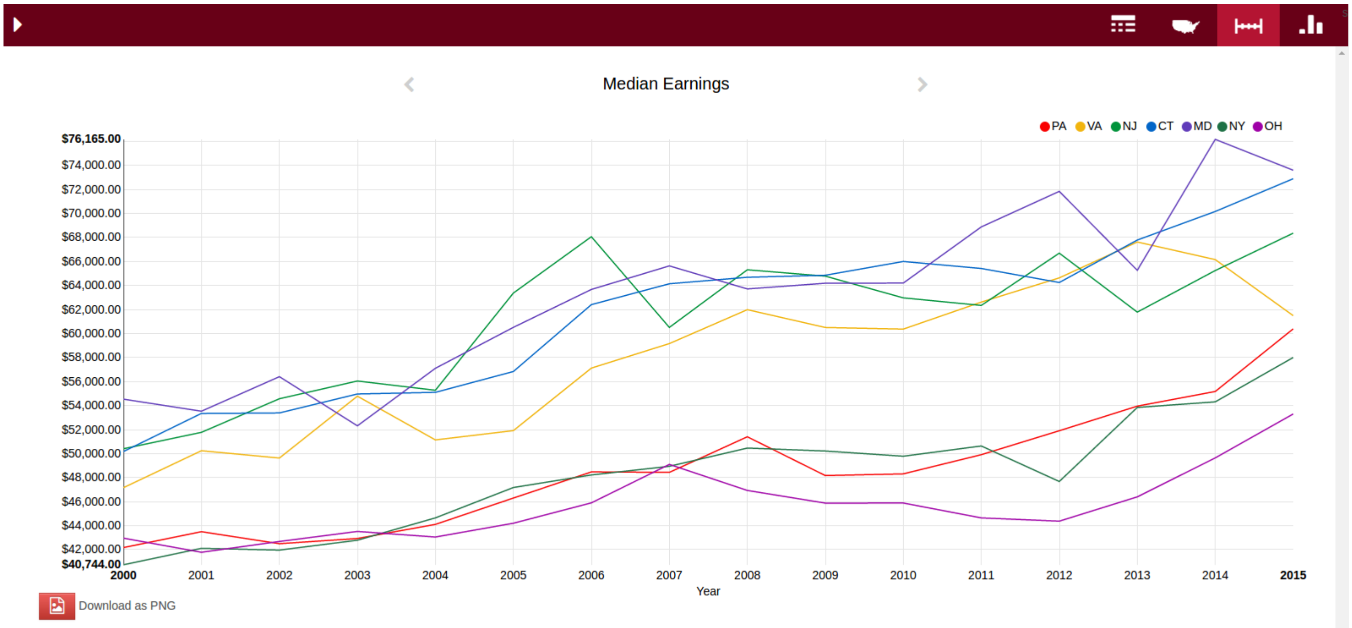
5 PRELIMINARY EVALUATION

The design of tools for interactive ranking exploration poses a number of challenges and open research questions. The MyRanker paradigm introduces questions regarding whether we can successfully learn rankings based on partial user input, and how to evaluate the "goodness" of such a ranking. Will the rankings we learn achieve the goal of the end user (economic or otherwise motivated)? Or might they contain bias and possibly put certain players into an unfair disadvantage. To begin to answer these questions we evaluate the utility of the MyRanker framework applied to the MATTERS system.

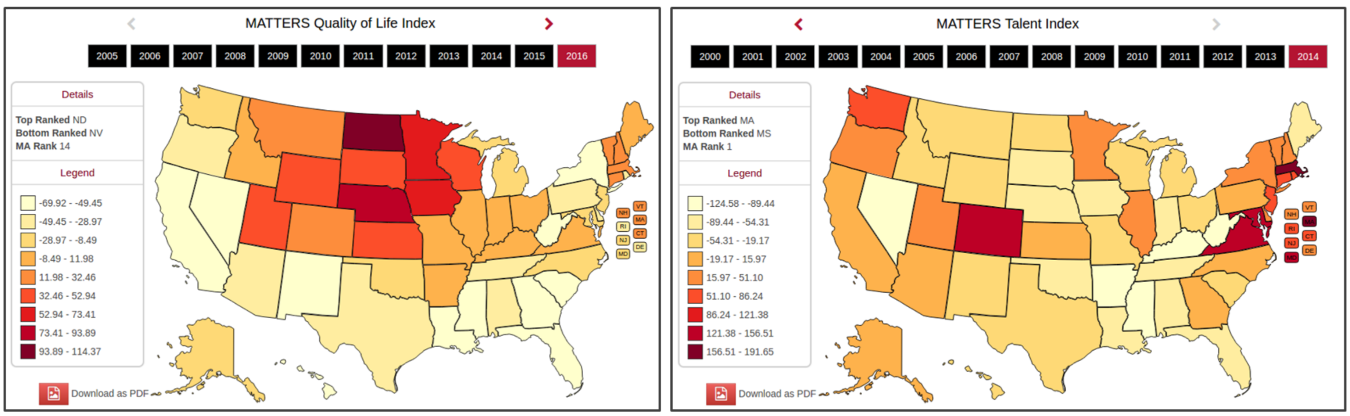
5.1 MyRanker Learning Module Evaluation

The MyRanker Framework provides a plug-and-play strategy wherein any pairwise learning-to-rank method can be easily incorporated into the ranking learner module. In the MATTERS interactive learn-to-rank tool we employ the method of regularized least squares given in [16] called RankRLS. The authors have provided a public software package for learning-to-rank. [6]

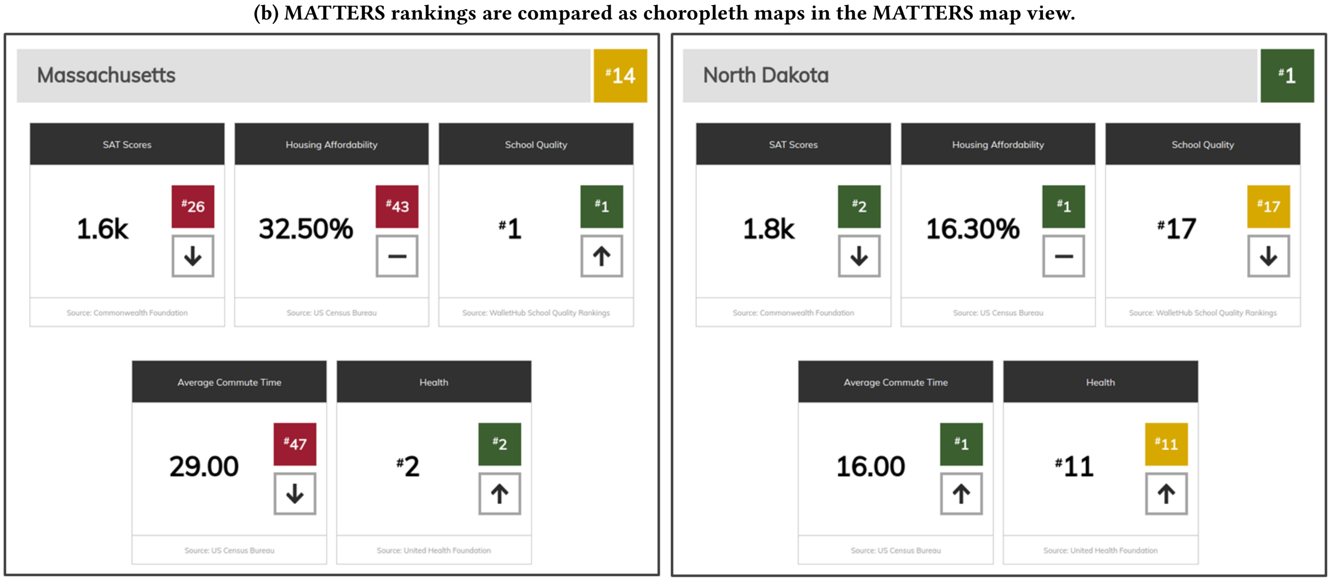
[6] http://staff.cs.utu.fi/~aatapa/software/RLScore/index.html



(a) The MATTERS timeline view shows how earnings data changes over time for a set of states.



(b) MATTERS rankings are compared as choropleth maps in the MATTERS map view.



(c) The metric values which make up a custom ranking are compared using the MATTERS State Profile.

Figure 6: Ranking views in MATTERS.

Number of States	Number of Pairs	cindex	tau
2	2	0.50	0.00
3	6	0.64	0.28
4	12	0.68	0.37
6	30	0.80	0.60
8	56	0.83	0.66
12	132	0.86	0.73
16	240	0.88	0.77
24	552	0.92	0.84

Table 1: The impact of the number of training pairs of states used to predict the MATTERS Cost index using RankRLS.

Number of Metrics	cindex	tau
0	0.95	0.91
2	0.94	0.88
4	0.93	0.87
8	0.92	0.84
12	0.90	0.80
16	0.88	0.77
18	0.87	0.75

Table 2: The impact of the number of metrics used to predict the MATTERS Cost index using RankRLS.

To evaluate the performance of this pairwise model on the data in the MATTERS warehouse, we would like to determine first how well RankRLS can learn an existing ranking of states, and second how much information is necessary to collect from users for a high quality ranking. To perform a preliminary assessment of the performance of the learn-to-rank tool we take advantage of rankings in our system which have been designed by experts from the Massachusetts High Tech Council. Four indices to measure the economic competitiveness of states have been constructed using the manual MATTERS Rank Builder interface – one for each of the data categories in the MATTERS system: Talent, Tax Climate, Cost of Doing Business and Quality of Life. Each metric consists of a weighted combination of metrics in the MATTERS system. We choose one, the 2014 MATTERS Cost of Doing Business Index, and evaluate how well RankRLS can learn this ranking.

Evaluation Metrics. Two ranking correlation measures are used to evaluate the RankRLS model. The first measure denoted "cindex" is a simple measure of pair concordance between the predicted ranking and the true ranking. A pair (a, b) is concordant if the predicted rank of a is greater that the predicted rank of b and this corresponds to a true rank of $a > b$. The cindex score is computed as the fraction of concordant pairs out of all pairs. This measure yields a number between 0 and 1, with 0.5 indicating random performance. We also employ the well-know statistical Kendall Tau [10] correlation coefficient. This measures the ordinal association between two rankings. A tau score of 1 indicates a perfect match, -1 indicates that one ranking is the reverse of the other, and 0 that the two rankings are independent.

The Cost of Doing Business ranking is computed using the metrics *Retail Price of Electricity, Median Earnings, Average Family Health Insurance Premium, and UI Premium per Employee.* Each metric is weighted evenly. Training on all possible pairs of states and the data from the 4 underlying metrics in the ranking yields a cindex of 0.95 and Kendall Tau score of 0.91. This was determined using 5-fold cross validation repeated over 10 randomized trials and taking the average score.

To evaluate the impact of the number of training pairs on the ranking we randomly selected subsets of states of varying sizes and formed all pairs to train on. Then the model was tested on the rest of the data in a cross-validation manner. The average of 10 randomized trials were taken. Table 1 shows the impact of the number of pairs on the accuracy of the learned rankings. As expected, performance increases with the amount of training data.

To evaluate the impact of noise from other underlying metrics, Table 2 shows the performance of the RankRLS model when trained on collections of metrics of increasing size. For each trial we include the 4 underlying metrics which make up the Cost of Doing Business ranking, and then randomly select additional datasets in the MATTERS system to train on as well. In the experiment given in the first row of Table 2 we train only on the 4 metrics, then in the next row 2 additional metrics are added, and so on. 10 randomized trials are averaged for each experiment. We can see that the pairwise learning model is impacted by noise from other data. This suggests the MyRanker framework could benefit from the addition of a feature selection or regularization step.

5.2 Use Case in Ranking for Talent Understanding

In addition to constructing rankings, the goal of an interactive ranking paradigm is to help users better interpret and understand ranking models. In the MATTERS system, the interplay of rank specification tools and data visualizations facilitates this understanding. Ranking models can be inspected using the rank builder interface (Section 4.1.1), and rankings can be compared with underlying raw data using the many MATTERS views (Sec 4.2).

Here we give an example of how MATTERS can be used to perform this type of analysis, by considering the MATTERS Talent Index. It is easy to look a state like California with its hub of innovation in Silicon Valley and observe that science and technology can drive prosperity with astounding impact. Policy makers and business leaders in other states may wonder how to foster similar drivers of economic growth in their own states. When the Massachusetts High Tech Council was developing the MATTERS system, they identified the ability to attract and maintain a highly skilled talent pool as one key to success in this area.

MATTERS provides a custom Talent Index designed by domain experts based on 4 metrics: *STEM Degrees Per Capita, Relocation of College Educated Adults, Bachelor's Degree Holders in Workforce, and Tech Employment as Percent of Total Employment.* The views and comparison tools in MATTERS can easily provide insight into this ranking, ensuring that users do not simply have to accept the index at face value. They may be interested in evaluating a number of concerns. Perhaps this ranking could contain implicit negative bias, measuring not just the underlying metrics, but also a trend based on race, gender or another undesirable measure. Or users might wonder how much additional insight this ranking really provides compared to other measures based on different metrics.

Figure 5 shows the MATTERS Talent Index displayed in the table view along with three demographic datasets. A highly undesirable ranking might reflect such information. However, using the correlation button, the Pearson Correlation Coefficient for the Talent Index compared to each demographic dataset is displayed in the top row of the table. We learn that none of these datasets are highly correlated with the Talent Index.

In another view we can observe how the Talent Index ranking compares to the other MATTERS Indices. It could be that the states with the highest quality of life attract the most talent, and therefore the Talent ranking might not be a unique measure. However, when we compare the choropleth maps shown in Figure 6b we can observe that the Talent ranking clearly has a different distribution across states from that of the MATTERS Quality of Life Index. Therefor the Talent Index is indeed measuring different phenomena. A deep dive into the factors contributing to individual state scores could provide additional insight using the State Profile view (Figure 6c).

With these strategies, the ranking model can be evaluated in the context of all available data. Compared with demographic information and other rankings, users can ascertain whether it is providing a meaningful measure of states.

6 DISCUSSION AND FUTURE WORK

Traditional approaches to the design of economic indices are based in theory and start with expert opinion. New approaches to data-driven analysis may provide previously unseen insights by leveraging open data and state-of-the-art machine learning techniques. In this work we introduce a vision for interactive learn-to-rank tools to facilitate the creation, exploration, and understanding of rankings. Our plug-and-play MyRanker framework provides interlinked components to achieve this goal in a data analytics system. We demonstrate the power of this paradigm for economic competitiveness evaluation as part of the MATTERS dashboard.

The brief evaluation presented here indicates potential for the construction of high quality rankings using partial knowledge specified by a user. Additional evaluation of pairwise learning based on user preferences is required to understand the trade-offs between user experience and accuracy. Further, highly usable interfaces are required to bring the power of the ranking algorithm to a non-expert audience. User studies are required to evaluate the utility of the learn-to-rank tools proposed in this paper. A simple to understand graphical display and intuitive interactions must capture the intent of the user, and facilitate their understanding and trust in the learned model. While alternate views are feasible, their relative utility must be formally studied for a given application context and user group targeted. Continued study and subsequent refinement of this new class of interactive learn-to-rank tools is planned.

Finally, to aid in the interpretation of learned rankings, a number of useful extensions to the MyRanker paradigm are being explored. We have shown that comparisons between underlying metrics and rankings provide useful insights regarding the explanatory power of ranking models. Tools to automatically learn those relationships would greatly aid in this type of analysis. Regularization and feature reduction techniques could provide an integral piece of the data analysis pipeline to further this understanding. Reducing the original input feature space to select only data which have the greatest impact on the position of individual objects can increase both user understanding and thus acceptance of constructed rankings.

7 ACKNOWLEDGEMENTS

The authors would like to thank the Massachusetts High Tech Council for their guidance and collaboration on the MATTERS project and partial funding support. We also appreciate the contributions of students who have worked developing the MATTERS system, especially Dmytro Bogatov and Jillian Hennessy for their work on the Rank Builder Tool. Student contributers are listed on the project website at http://davis.wpi.edu/dsrg/PROJECTS/MATTERS. We also acknowledge partial funding support from the Student Research Participation Program at NSRDEC administered by ORISE through an interagency agreement with the U.S. DOE.

REFERENCES

[1] Dmytro Bogatov and Jillian R Hennessy (advisor: Elke Rundensteiner; sponsor: Massachusetts High Technology Council). Data matters: Customizing economic indices to measure state competitiveness. 2016. URL https://web.wpi.edu/Pubs/E-project/Available/E-project-042616-011522/.

[2] Cynthia A Brewer, Mark Harrower, B Sheesley, A Woodruff, and D Heyman. Colorbrewer 2.0. Accessed June 2017. URL http://www.ColorBrewer.org/.

[3] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning*, pages 89–96. ACM, 2005.

[4] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning*, pages 129–136. ACM, 2007.

[5] Ross DeVol, Joe Lee, and Minoli Ratnatunga. 2016 State Technology and Science Index: Sustaining americafis innovation economy. 2016.

[6] Alex Fortier, Westley Russell, Kevin Mee, and Long Nguyen Duc (advisor: Elke Rundensteiner; sponsor: Massachusetts High Technology Council). Advancing matters. 2015. URL https://web.wpi.edu/Pubs/E-project/Available/E-project-042915-094249/.

[7] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Support vector learning for ordinal regression. 1999.

[8] White House. College scorecard. 2013. URL https://collegescorecard.ed.gov/data/.

[9] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[10] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2): 81–93, 1938.

[11] Caitlin Kuhlman, Karthikeyan Natesan Ramamurthy, Prasanna Sattigeri, Aurelie C Lozano, Lei Cao, Chandra Reddy, Kush R Varshney, and Aleksandra Mojsilovic. How to foster innovation: a data-driven approach to measuring economic competitiveness. *IBM Journal of Research and Development*, in press.

[12] Joseph Lee Rodgers and W Alan Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[13] Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225–331, 2009.

[14] Rodica Neamtu, Ramoza Ahsan, Jeff Stokes, Armend Hoxha, Jialiang Bao, Stefan Gvozdenovic, Ted Meyer, Nilesh Patel, Raghu Rangan, Yumou Wang, Dongyun Zhang, and Elke A. Rundensteiner. Massachusetts Economy and Technology Index System. In *Proceedings of the International Workshop on Data Science for Macro-Modeling*, DSMM'14, pages 11:1–11:2. ACM, 2014.

[15] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175, 1998.

[16] Tapio Pahikkala, Evgeni Tsivtsivadze, Antti Airola, Jorma Boberg, and Tapio Salakoski. Learning to rank with pairwise regularized least-squares. In *SIGIR 2007 Workshop on Learning to Rank for Information Retrieval*, volume 80, pages 27–33. Citeseer, 2007.

[17] Klaus Schwab, Xavier Sala-i Martin, et al. The global competitiveness report 2016-2017. World Economic Forum, 2017.

[18] Hal R Varian. Big data: New tricks for econometrics. *The Journal of Economic Perspectives*, 28(2):3–27, 2014.

[19] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. Listwise approach to learning to rank: theory and algorithm. In *Proceedings of the 25th International Conference on Machine Learning*, pages 1192–1199. ACM, 2008.

[20] Li Zhou. Obama's new college scorecard flips the focus of rankings. *The Atlantic*, 2015. URL https://www.theatlantic.com/education/archive/2015/09/obamas-new-college-scorecard-flips-the-focus-of-rankings/405379/.