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

Caitlin Kuhlman   Elke A. Rundensteiner

Worcester Polytechnic Institute
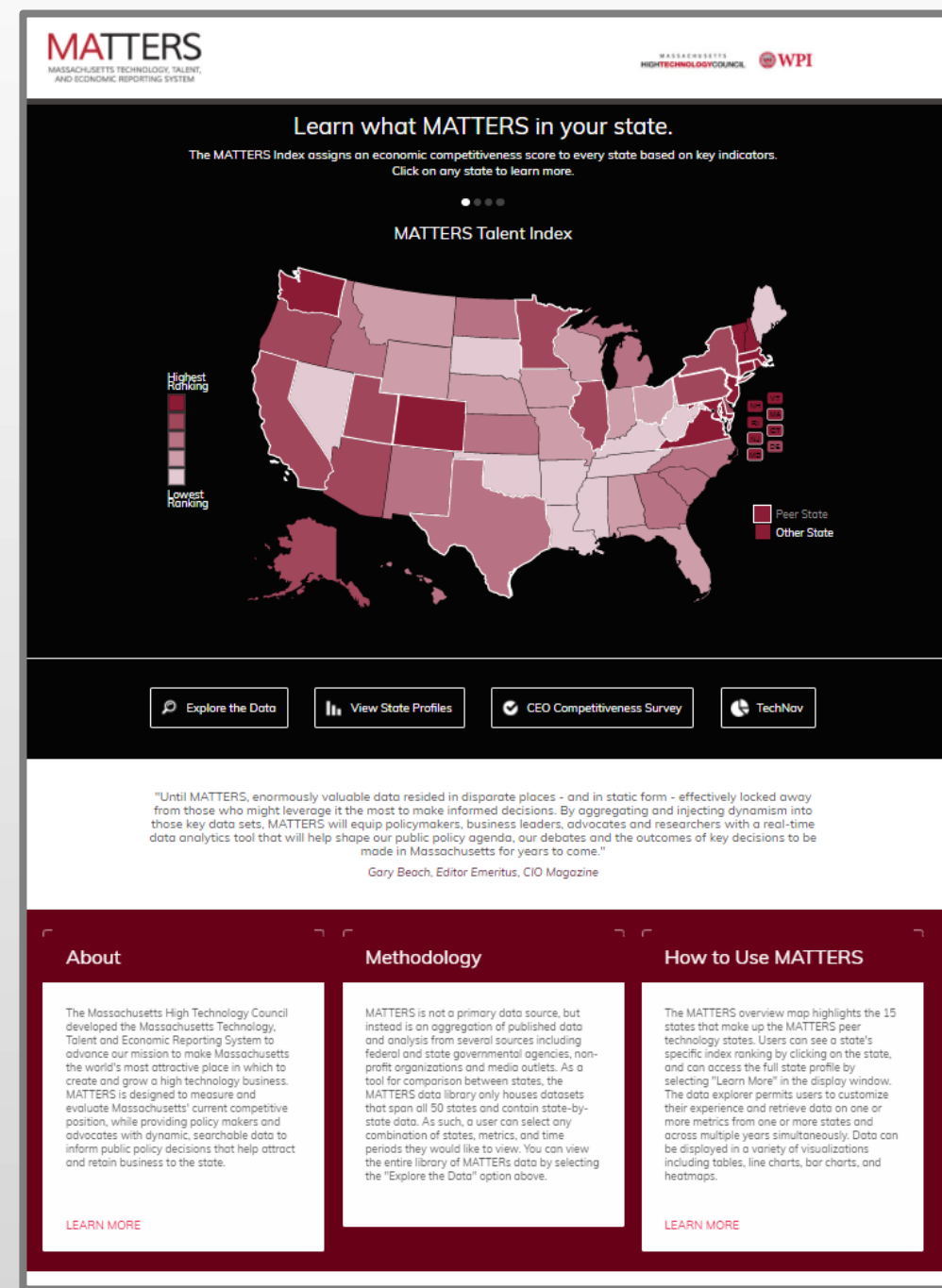
Rankings and their pitfalls

- Rankings are a fundamental tool used to help people understand the relative merit of objects and to simplify decision making when the number of factors impacting choice is large.

- **Examples:** ranking countries or regions according to economic principles, ranking institutions such as colleges or hospitals, ranking search results for user queries.

- Ranking models are often hand-designed in ad hoc manner. May succumb to several shortcomings:
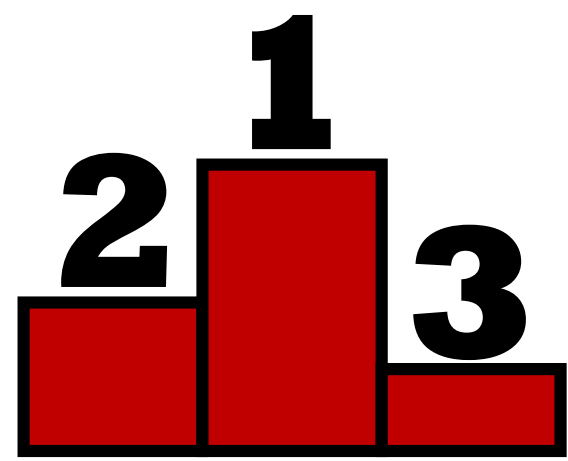
    - Depend heavily on expert opinion.
    - Formulation not always transparent.
    - May contain non-discoverable bias.

MyRanker Framework:
Plug and play system for interactive learn-to-rank

Data Store

Data Filtering and Preprocessing

Preference Collection Interface

Ranking Learner

Rank   View

Goals

- Introduce an interactive paradigm for automatic learn-to-rank tools supporting exploration and understanding of rankings.

- Aid users in the creation of ranking models which reflect their intuition and value system, through the incorporation of learning-to-rank algorithms [1] into highly usable interfaces.

- Provide interlinked visual displays allowing users interact with ranking models to gain deeper insights.

- Demonstrate the power of interactive ranking analysis applied for economic competitiveness evaluation.

Use Case: MATTERS Analytics Dashboard

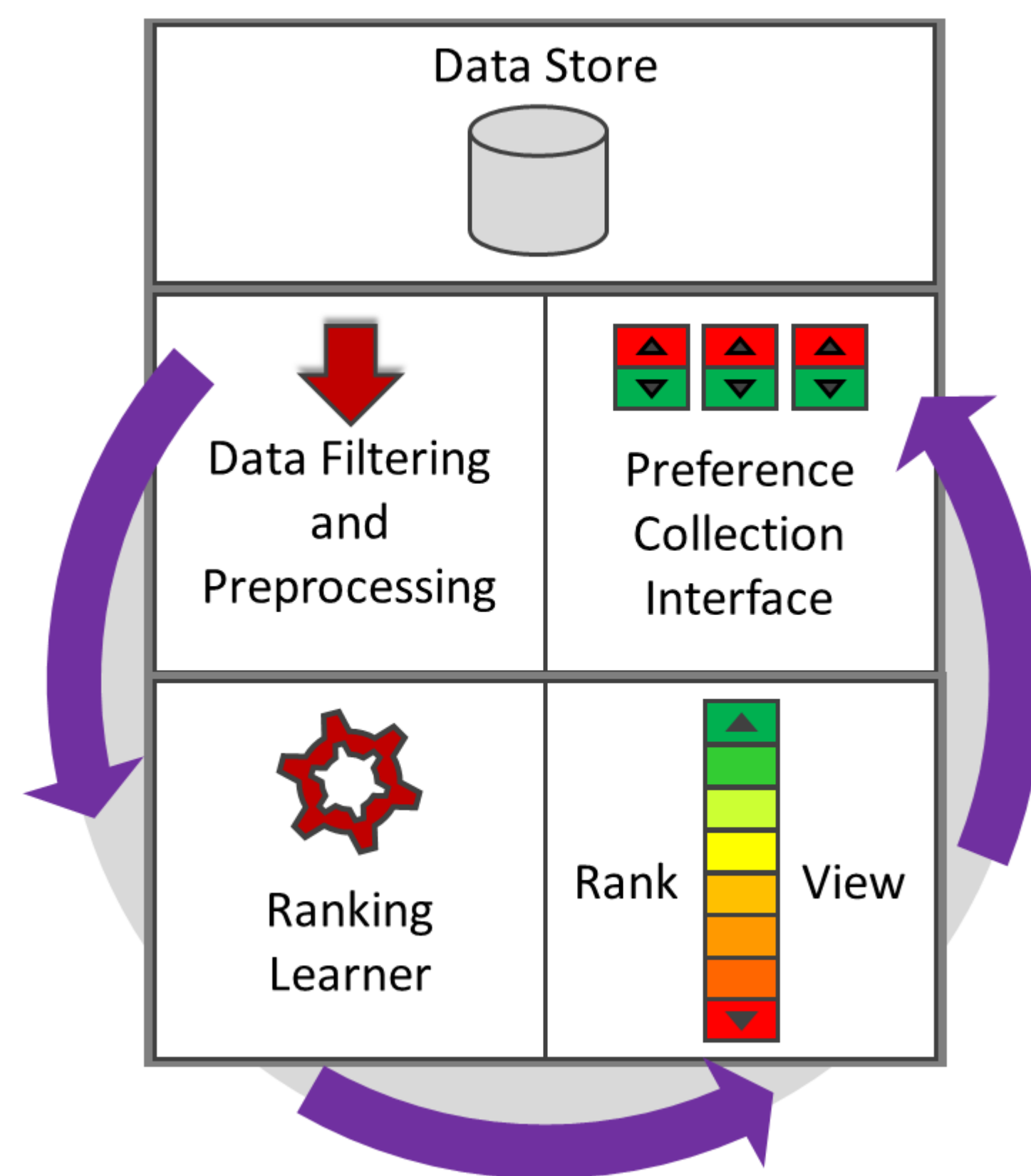**Massachusetts Technology, Talent, and Economic Reporting System**

- Online public tool which measures the economic competitiveness of US states using open data [2].

- MATTERS consolidates over 50 publicly available socioeconomic datasets.

- Publish 4 rankings annually to measure economic competitiveness according to Talent, Tax Climate, Cost of Doing Business, and Quality of Life.

MATTERS™
MASSACHUSETTS TECHNOLOGY, TALENT, AND ECONOMIC REPORTING SYSTEM

Ranking Specification Tools

Manual Rank Builder Tool
Explicitly construct ranking formulas and adjust learned weightings

(Tech Employment as Percent of Total Employment) * 75 + (Housing Affordability) * 40 + (Average Commute Time) * 25
75 + 40 + 25 = 140

Tech Employment as Percent of Total Employment - Remove
Weight: 75

Housing Affordability - Remove
Weight: 40

Average Commute Time - Remove
Weight: 25

My Custom Metric   Save

Pairwise Learn-to-Rank Tool
Indicate preferences amongst objects. Rankings are learned automatically.

3 Metrics Selected

Massachusetts (MA) > New York (NY)
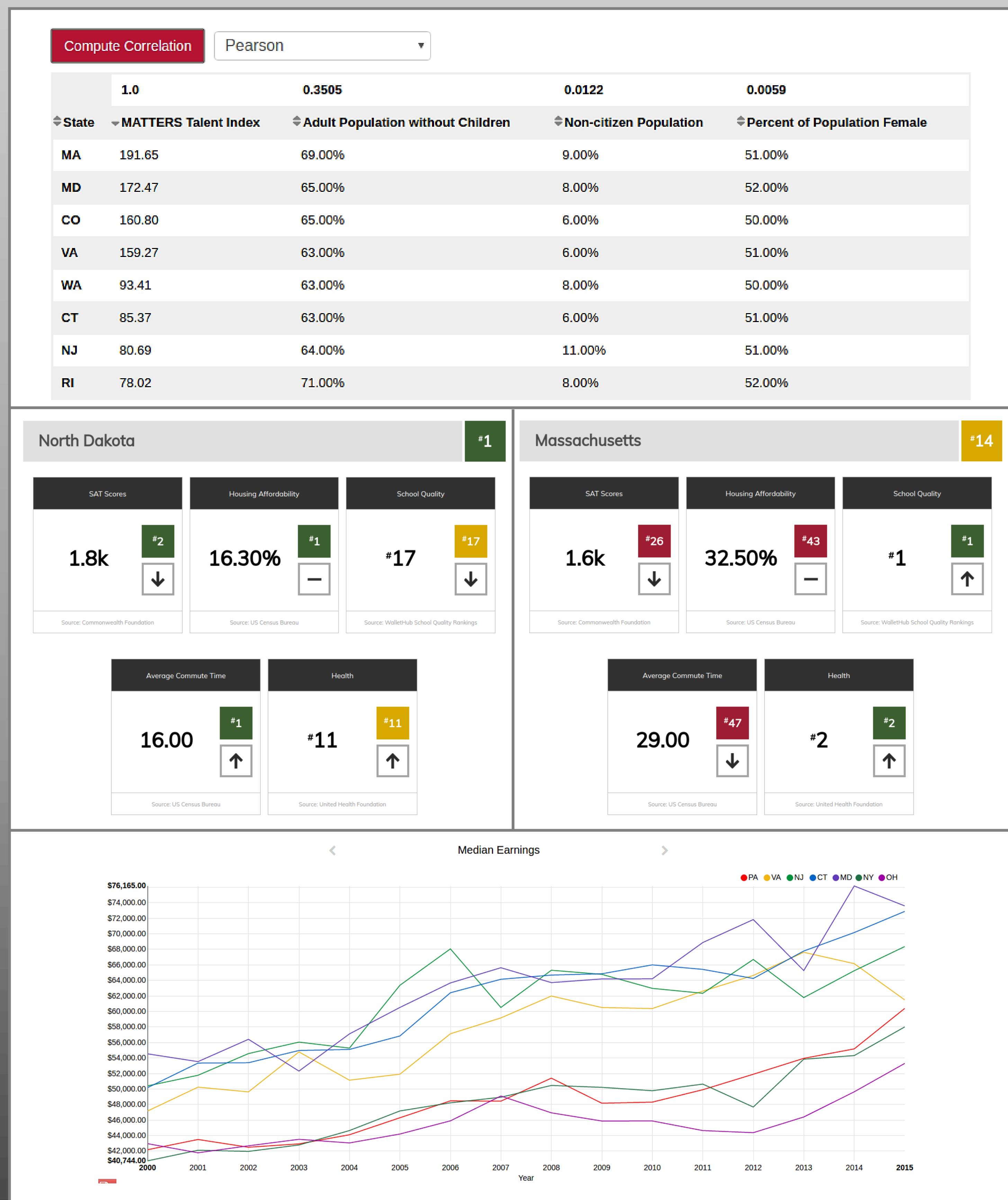
California (CA) > Texas (TX)

Rank!

Choose the Top State:
Hawaii (HI)

Choose the Bottom State:
North Dakota (ND)

Add Pair

Ranking Views: Compare, Explore, Interact

Compute Correlation   Pearson

	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%

North Dakota #1

SAT Scores 1.8k #2   Housing Affordability 16.30% #1   School Quality #17 #17

Average Commute Time 16.00 #1   Health #11 #11

Massachusetts *14

SAT Scores 1.6k #26   Housing Affordability 32.50% #43   School Quality #1 #1

Average Commute Time 29.00 #47   Health #2 #2

Median Earnings

PA VA NJ CT MD NY OH

Multiple views allow users to inspect and compare rankings and their components.
(Shown clockwise from the top left)

- **Table view:** correlation analysis provided to understand the relationships between data and rankings.
- **State Profile view:** displays the metric values which make up a custom rankings for individual states.
- **Timeline view:** shows how data changes over time for a set of states.
- **Choropleth Map view:** allows users to compare the distribution of ranking values across states.

MATTERS Talent Index

MATTERS Quality of Life Index
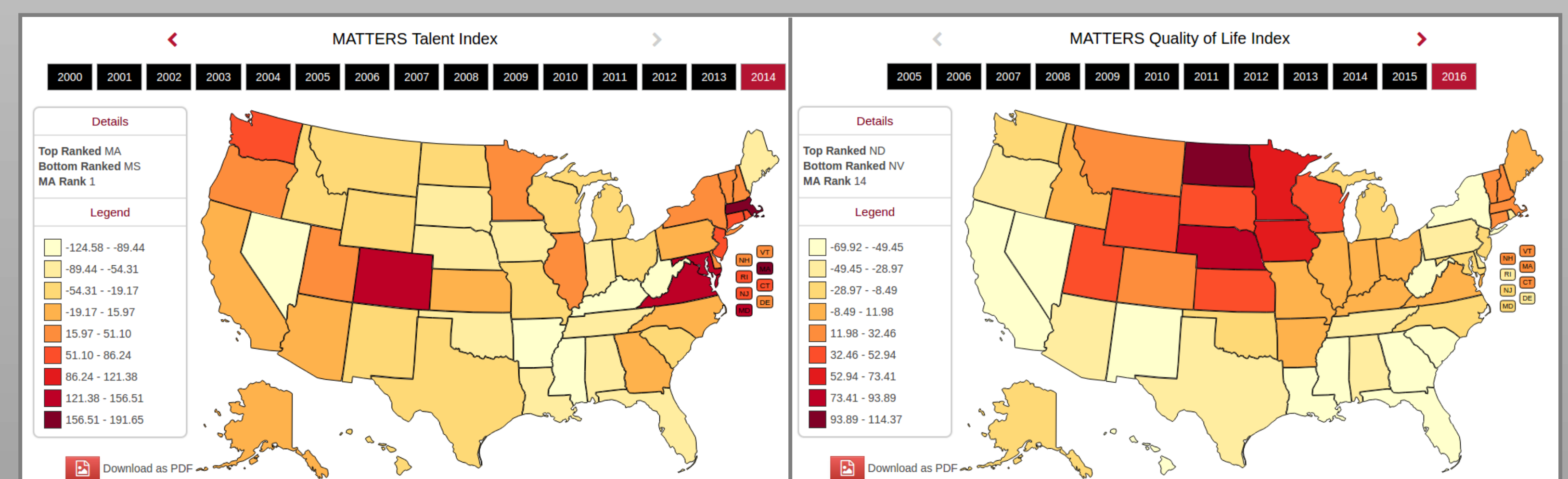
Evaluation

Preliminary evaluation of the pairwise Learn-to-Rank tool on the data in the MATTERS warehouse aims to answer key questions:

- How well can the ranking algorithm [3] can learn an existing ranking?
- How much information is necessary to collect from users for a high quality ranking?

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

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

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

[1] Liu, Tie-Yan. "Learning to rank for information retrieval." Foundations and Trends® in Information Retrieval 3.3 (2009): 225-331.
[2] Ahsan, Ramoza, et al. "Massachusetts Economy and Technology Index System." Proceedings of the International Workshop on Data Science for Macro-Modeling. ACM, 2014.
[3] Pahikkala, Tapio, et al. "Learning to rank with pairwise regularized least-squares." SIGIR 2007 workshop on learning to rank for information retrieval. Vol. 80. 2007.

MASSACHUSETTS HIGH TECHNOLOGY COUNCIL
Dedicated to Growth... Committed to Action