Graph-Based Visualization of Stochastic Dominance in Statistical Comparisons

Gurdal Ertek, Gul Tokdemir, Mohamad Mustafa Hammoudi

Graph-Based Visualization of Stochastic Dominance in Statistical Comparisons

Gurdal Ertek
College of Business and Economics
United Arab Emirates University
Al Ain, Abu Dhabi, UAE
gurdal@uaeu.ac.ae
https://orcid.org/0000-0002-5167-4836

Gul Tokdemir
Department of Computer Engineering
Cankaya University
Ankara, Turkey
gtokdemir@gmail.com
https://orcid.org/0000-0003-2441-3056

Mohamad Mustafa Hammoudi
College of Arts and Sciences
Abu Dhabi University
Abu Dhabi, UAE
hammoudi_mh@hotmail.com
https://orcid.org/0000-0001-5546-2841

***Abstract*—In this paper, a graph visualization scheme and methodology is proposed for representing, understanding, and interpreting the statistical comparison of means and the resulting stochastic dominance. The practicality and applicability of the visualization scheme and the methodology is illustrated through a case study, with data coming from higher education institutes in the United States of America (U.S.A.). The objective of the research is to make statistical results more accessible and readable, enabling the visual derivation of actionable insights.**

***Keywords*— graph visualization; graph drawing; stochastic dominance; hypothesis testing; higher education.**

I. INTRODUCTION

Various statistical and machine learning techniques can be applied in industry analytics projects and in applied research studies. These range from unsupervised learning techniques, such as clustering [1] and association mining [2], to supervised learning techniques, such as regression [3] and classification [4]. Specifically, in survey-based research, more in-depth analyses, including dimensionality reduction, regression, Structural Equation Modeling (SEM), are applied as the preferred primary method [5,6].

Yet, basic summary statistics and statistical comparison of means are critical results, especially in the early-stages of data analysis [4,7,8,9]. Thus, in analytics projects, the comparison of means is a fundamental statistical analysis.

When the distribution and central tendency of observations belonging to two or more independent datasets (populations or samples) are to be compared, statistical tests for comparison of means are applied. The choice of the specific test, either parametric or non-parametric, depends on whether the populations or samples follow or violate the assumptions of normality. If there are two datasets, either the parametric t-test or the non-parametric Mann-Whitney test is applied. When there are multiple datasets to be compared, if the normality assumptions hold for all the datasets, parametric ANOVA is applied. If any one of the datasets is not normally distributed, one has to resort to the non-parametric tests, typically the Kruskal-Wallis H test, which tests whether the datasets are coming from the same distribution.

In practical projects, knowing whether or not multiple datasets come from the same distribution is a useful but not a sufficient insight. What is needed more is to determine in particular which datasets have larger means than which others, and at which level of statistical significance. To this end, a pairwise comparison of stochastic dominance is needed. If the mean of a sample A is higher than that of another sample B with a statistical significance, the former is said to "have a second-order stochastic dominance" over the later sample [10]. The statistical test for such pairwise comparisons is one of Dunn's test [11], pairwise Mann-Whitney tests without Bonferroni correction, or the Conover-Iman test.

The Dunn's test takes as input a collection of datasets, each dataset representing a population or sample, compares each dataset pair, and presents within its output the following information for each pair: the compared pair, z-value, unadjusted p-value, and the adjusted p-value. For example, Table I illustrates sample results for the Dunn's test, for the data used in the case study. Each sample dataset comes from a different state in Northern U.S.A., e.g., MA, NY, etc., and contains observations of earnings of alumni of higher education institutes in that state. For example, the first row in Table I provides results regarding the mean earnings in Colorado (CO) vs. Iowa (IA). From summary statistics, the sample mean for CO is known to be 40,345 and the sample mean for IA is 42,559. The adjusted p-value in Table I, corresponding to this particular pairwise comparison is 1.000000e+00, meaning that the difference in mean earnings between the two states is not statistically significant, at all. The threshold p-value determines which pairs have differences in means with statistical significance. A statistically significant difference is between Colorado (CO) and Massachusetts (MA): The mean earning of alumni who graduated from higher education institutions in MA is larger than the mean earnings of those who graduated from CO.

Given the importance and ubiquitous utilization of pairwise statistical comparison of means, in this paper, a graph visualization scheme and methodology is proposed for representing, understanding, and interpreting the statistical comparison of means and the resulting stochastic dominance. The visualization scheme is illustrated with a real-world dataset from the education sector, which serves as an example of how the proposed methodology can be applied in practice.

II. Related Work

The visualization of sample means can significantly improve the description and the understanding of stochastic dominance. In this section, various visualization schemes that have been used in the past are reviewed.

A common statistical approach in visualizing distributions of data from multiple samples is through the use of box-and-whisker plot [12]. The box-and-whisker plot is a statistical graph that displays the median as a point-estimate of central tendency, skewness, spread and distribution of observations, and outliers in a dataset, where each box and the outlier points along that box represents the data points in a sample [13,14]. The box-and-whisker plot is graphed by finding the three quartiles of the data set, as well as the lowest and largest values between the two inner fences. This type of plot is a very helpful statistical graph that can be used to observe the distribution of values and to detect outliers in a dataset. There are many variants of the box-and-whisker plot, and there are many closely related types of plots, such as violin plots, strip charts, and dot plots [15].

TABLE I. SAMPLE RESULTS OF THE DUNN'S TEST

Comparison	*z-value*	*p-unadjusted*	*p-adjusted*
CO - IA	-1.705690826	8.806566e-02	1.000000e+00
CO - IL	-0.143199762	8.861324e-01	1.000000e+00
IA - IL	1.888214571	5.899715e-02	1.000000e+00
CO - IN	-1.393611282	1.634350e-01	1.000000e+00
IA - IN	0.190000952	8.493084e-01	1.000000e+00
IL - IN	-1.485002563	1.375431e-01	1.000000e+00
CO - KS	-0.456310417	6.481668e-01	1.000000e+00
IA - KS	1.294031462	1.956546e-01	1.000000e+00
IL - KS	-0.401938997	6.877289e-01	1.000000e+00
IN - KS	1.003068192	3.158280e-01	1.000000e+00
CO - MA	-4.513333838	6.381645e-06	**5.807297e-04**
IA - MA	-2.511416416	1.202478e-02	1.000000e+00
IL - MA	-5.637810808	1.722256e-08	**1.567253e-06**
IN - MA	-2.496984870	1.252543e-02	1.000000e+00
KS – MA	-4.133510183	3.572645e-05	**3.251107e-03**
…	…	…	…

When drawing the box-and-whisker plot visualization for the same higher education dataset discussed earlier, each box shows the distribution of (mean) alumni earnings who graduated from a particular institution. For enabling an easier comparison between the samples, the box-and-whisker plots can be positioned in the order of decreasing mean values. Figure 1 illustrates a sample ordered box-and-whisker plot, in the center of the figure. While the box-and-whisker plot is easier to grasp and interpret compared to the Dunn's test results in Table I, it still does not clearly display the stochastic dominance relations between the pairs of samples.

Alternative visualization schemes for comparing distributions and central tendencies of samples include the density plot, histogram plot [15], and the frequency trail [16]. In these schemes, multiple histograms or density plots for the samples are plotted together in one plot, either completely or partially superimposed.

Other alternatives, such as the empirical cumulative density function and Q-Q plots [15] illustrate stochastic dominance to some degree, both of them are hard to interpret in the case of increased number of samples and do not directly reflect the Dunn's test results. In summary, none of the mentioned types of visualizations provide direct actionable information on the existence of stochastic dominance, as obtained from the results of the Dunn's test.

While enabling convenient visual comparisons, the aforementioned visualization schemes do not explicitly show the statistical significance of the differences. A tabular visualization scheme suggested in [17] does display stochastic dominance and is useful to observe the stochastic dominance between a particular pair of samples. However, the visualization in [17] does not enable the comparison of all samples at once and does not make the associations between these comparisons and the hierarchical ordering of these associations conveniently visible.

The schemes listed above for the comparison of distributions and means are not based on graphs (networks, with nodes and arcs). The only graph-based scheme for visualizing the results of the Dunn's test can be found within the SPSS software, where each sample is represented as a node, and statistically significant differences are shown through colored arcs [18]. However, the visualization in SPSS does not exhibit the hierarchical nature of the comparisons and the display may become quite crowded and confusing when the number of comparisons increases (with the square of number of samples).

The Dunn's test executes all pairwise comparisons and provides as result the essential statistics. The challenge that is encountered inherently with the Dunn's test is the plethora of results which is not easy to grasp through the results table. For example, Table I does provide which pairs of datasets have statistically significant differences, but this information has to be compiled and studied to enable the deduction of actionable insights regarding the complete set of samples. For example, Table I does not provide the answer to the question of which institutions could be prioritized by a candidate, from a holistic perspective. The candidate would need to go through the complete table and compile the results. Thus, the tabular presentation of results in text is hard to interpret jointly and does not provide an overall picture of stochastic dominance.

In summary, the results generated by the Dunn's test constitute data in that need to be cleaned and analyzed for insights. To the best of the authors' knowledge, there does not yet exist a visualization scheme for displaying stochastic dominance in a single plot, with all the dominance relations being presented. This is the particular challenge that motivated this presented research.

In the presented research, the challenge of making Dunn's test results more accessible and easily interpretable is addressed. To this end, a novel graph-based visualization scheme and an underlying methodology for generating the mentioned visualization, are presented.

III. Methodology

The methodology proposed in this paper aims at making Dunn's test results more understandable and interpretable through graph visualization [19].

Graph visualization can enable the discovery of hidden patterns in data and the generation of novel and actionable insights, in any domain of application [1,20,21]. Graph visualization, also referred to as graph drawing and graph layout (with subtle differences in these terms), is a field of research and practice within the greater field of information visualization.

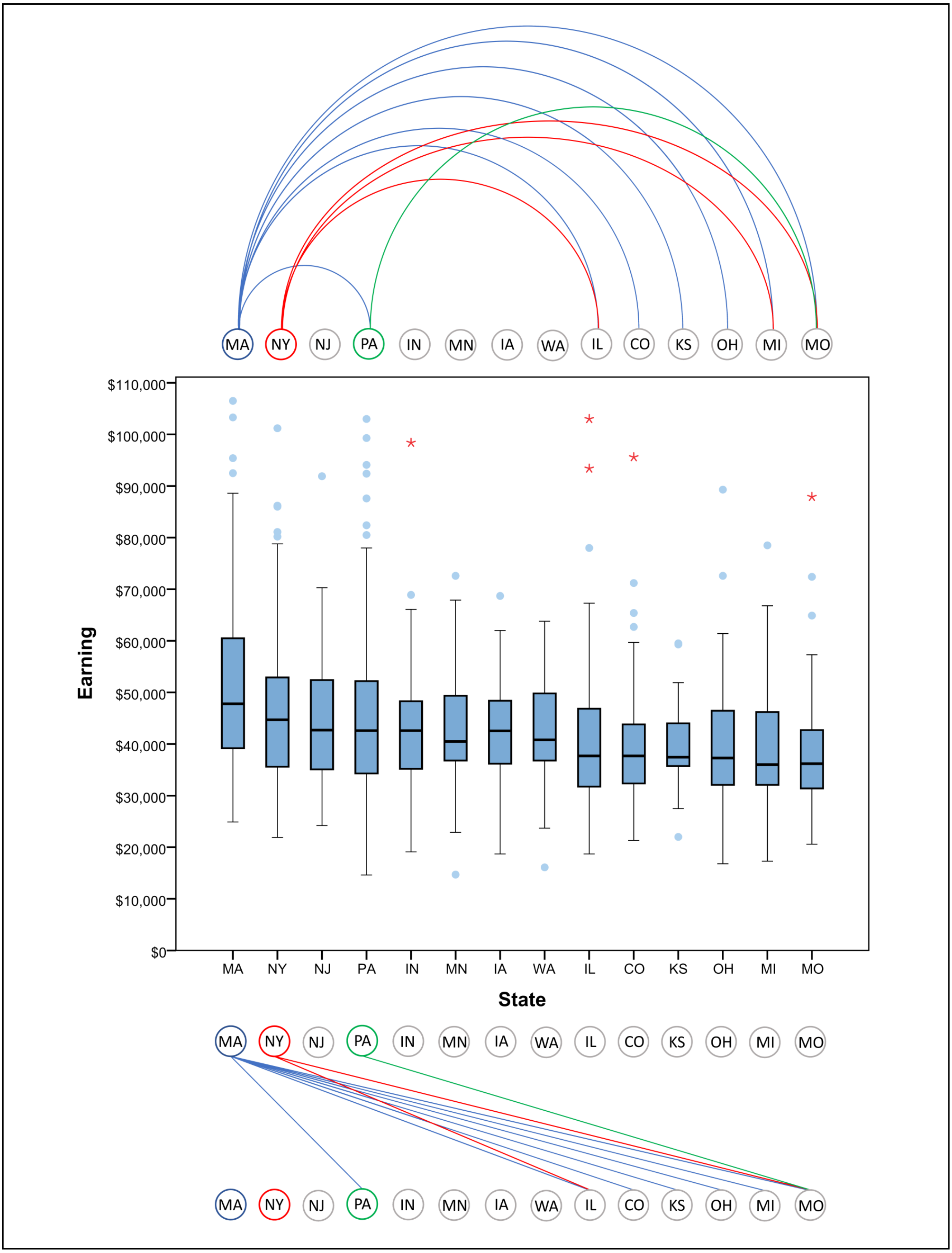

Fig. 1. Arc diagram and slope chart, super-imposed with the sorted box-and-whisker plot

.

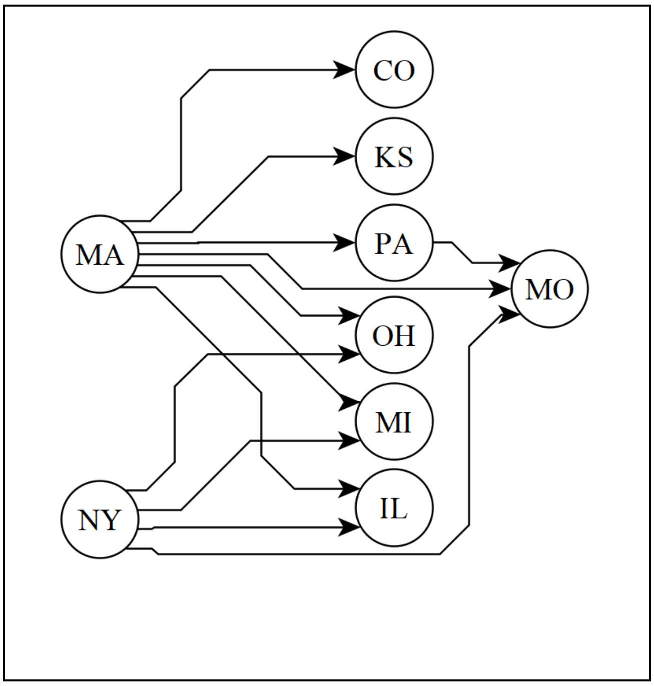

Fig. 2. Hierarchical (layered) graph visualization of stochastic dominance.

The discipline of data and information visualization derives from a multitude of fields, including Human Computer Interaction (HCI) [22], data mining [23], statistics, visual arts, graphic design, and cognitive psychology. The objective is to increase the understanding and interpretation of data, and to derive actionable insights from data. Especially with large datasets, interpreting diagrammatic representations is more efficient with respect to textual alternatives [24]. The primary task in information visualization is to display data in two or three-dimensions, in static or dynamic format, and with or without user interaction, to achieve this objective.

With the advent of information technologies, many novel visualization schemes and techniques are being developed, with applications in new areas. The field of data and information visualization has advanced to the point where dedicated research journals have emerged [25]. To organize and summarize the various visualization alternatives, [26] has developed a "Periodic Table of Visualizations", [27] has developed a "Graph Gallery", and [28] has developed a "Chart Suggestions". The motivation for the methodology was the realization that, in order to visualize the results of the Dunn's test, a scheme suitable for making comparisons needs to be adopted. To this end, an online source [29], which provides a catalogue and list of possible visualization schemes for data with different structures, proved beneficial.

Many novel visualization schemes have been developed in recent decades, to illustrate distribution, correlation, rankings, part of a whole, evolution, maps, flow, and other aspects of data [27].

The visualization schemes selected to visualize stochastic dominance between samples are arc diagram, slope chart, and hierarchical (layered) graph.

In graph theory and graph analysis, the fundamental building blocks are nodes (vertices) and arcs (edges). The proposed graph visualization scheme builds on nodes and arcs; each sample that is being compared is represented as a node and stochastic dominance between a pair of samples is represented as an arc. The tabular results from the Dunn's test are thus mapped onto a graph and subsequently visualized to reveal the hierarchical structure within the graph.

These selected schemes specifically provide a perspective on the hierarchical structure of stochastic dominance between samples. The selected schemes can scale up in the case of increased number of statistically significant relations, hence their selection.

Arc diagram

An arc diagram is a graph where the nodes, which show the entities, are aligned along a single axis (typically the x-axis) and edges, which show the relationships, are drawn as (typically semi-circular) arcs [30]. An example of an arc-diagram, that displays the Dunn's test results in this paper, is given in the top-level plot in Figure 1.

Slope chart

Slope chart is a type of chart which consists of two parallel lines with points aligned on these lines [31]. The slope chart is ideal to compare the change in the values of distinct entities over time [32]. In this presented research, the slope chart has positions of entities (nodes) also aligned, and the lines display stochastic dominance from nodes on the upper line to those in the lower line. An example of a slope chart, that displays the Dunn's test results in this paper, is given in the bottom-level plot in Figure 1.

Hierarchical (layered) graph

Hierarchical (layered) graph drawing (layout) is a scheme for visualizing a graph, achieved through graph drawing (layout) algorithms, such that the nodes are placed along successive layers [33]. Algorithms for generating graph layouts, including hierarchical layouts, is discussed in [34]. An example of a hierarchical (layered) graph, that displays the results in this paper, is given in Figure 2.

Methodology

The following are the specific steps of the developed methodology:

Step 1. Collect, clean, and prepare data, with one categorical (nominal) input variable (that denotes the samples) and one numerical output variable (target attribute).
Step 2. Obtain sample means and other summary statistics.
Step 3. Draw the box-and-whisker plot and/or similar plot and identify any outliers to remove. Remove the selected outliers from the dataset.
Step 4. Conduct the non-parametric Kruskal-Wallis H test to examine whether the samples are coming from the same distribution or not.
Step 5. If all samples are *not* coming from the same distribution (or, in other words, if the p-value for the Kruskal-Wallis H test is below the threshold value, rejecting the null hypothesis) then continue to Step 6; otherwise, terminate the analysis.
Step 6. Conduct the Dunn's test for the pairwise comparison of sample means, for all samples. Store the test results in a table.

Step 7. Using the sample means from Step 2 and the pairwise comparison results from Step 5, construct the hierarchical (layered) graph.

Step 8. Apply a hierarchical layout algorithm for visualizing the dominance graph. Change the parameters of layout algorithm and run the algorithm under each parameter setting (minimum distances between edges, minimum distances between arcs, minimum slope, edge routing style, layer assignment policy), selecting the particular graph visualization which provides the most insights with most ease. Where needed, manually position nodes and edit arcs to increase interpretability. While this step can be conducted exhaustively for every combination of parameter settings, experienced users can also obtain valuable insights with much less work, through the interactive selection of parameter settings.

Step 9. Construct the other two stochastic dominance graphs, namely the arc diagram and the slope chart, where nodes are sorted with respect to decreasing sample means. An arc denotes stochastic dominance, as suggested by a statistically significant difference in means. A different color can be used for arcs emanating from each different node.

Step 10. Construct the ordered box-and-whisker plot, where the boxes are ordered in decreasing order with respect to sample means.

Step 11. Stitch together the arc diagram, slope diagram, and the box-and-whisker plot, such that nodes are vertically aligned.

Step 12. Analyze the obtained graph visualizations and observe any stochastic dominance relationship between the samples. Develop strategies to apply the observed insights for the improvement of the system.

IV. CASE STUDY

In this section, the developed graph visualization methodology is applied and demonstrated within the education sector. The dataset used, College Scorecard Data, has been compiled by the U.S.A. Department of Education [35], and contains extensive data on university and colleges in the U.S.A. The scorecard was designed to increase transparency and empower the public through being informed about institutions of higher education in the country. The data that contains hundreds of data fields for each institution, is regularly updated (with the last update on October 30, 2018), and is documented systematically with clear explanations [36]. The following data fields, as defined in the data dictionary [37], were included in the current analysis:

- **INSTNM:**
  Institution name
- **STABBR:**
  State postcode (renamed as "**State**" in the analysis)
- **MN_EARN_WNE_P10:**
  Mean earnings of students working and not enrolled 10 years after entry (renamed as "**Earning**" in the analysis)
- **COUNT_WNE_P10:**
  Number of students working and not enrolled 10 years after entry.

Among the mentioned four data fields, the values under "State" (categorical variable) and "Earning" (numerical output variable) were used for the statistical tests and visualizations.

For the analysis of the differences in mean earning, universities in the Northern states of U.S.A., as defined by the U.S.A. Census Bureau [37] were selected. Among the Northern states, only states with at least 30 universities were included. For each included state, only institutions with earnings data available were included. Seven outlier universities, with average earnings of greater than 110K USD, were excluded from the box-and-whisker plot and further analyses. Overall, 1,196 universities from 14 states were included in the exploratory and statistical analysis, as well as for the application of the proposed methodology.

The mean values and boxes in the box-and-whisker plot are sorted according to the decreasing values of earning, as MA, NY, NJ, PA, IN, MN, IA, WA, IL, CO, KS, OH, MI, and MO.

A tabular dataset with two columns, namely state and earning, was compiled and the Kruskal-Wallis H test and the Dunn's pairwise comparison test were applied. The p-value for the Kruskal-Wallis test was obtained as 1.28241395705066E-11, which clearly suggests that the sample means are not all same. Next, the Dunn's test was applied, resulting in pairwise comparisons, z-values, p-values, and adjusted p-values (Table I). Then, graph visualizations were created.

The arc diagram and the slope chart, stitched together with the sorted box-and-whisker plot shown in Figure 1, display stochastic dominance among the samples. Figure 1 illustrates how the earnings in MA are higher than PA, IL, and certain other states. While there is statistically significant difference between the earnings in MA and these states, the differences are not statistically significant when it comes to MA vs. NJ, IN, MN, IA, and WA. The earnings from NY are higher than those in IL, MI, and MO, with statistical significance. The only state, which has stochastic dominance over another state is PA, which has higher earnings than MO, with a statistical significance.

The hierarchical graph, shown in Figure 2, displays the layers of hierarchy for stochastic dominance. The dominance relations of Figure 1 can be observed in this figure, as well. However, this figure also shows the layers of states with respect to stochastic dominance. So, the state that is dominated the most (in terms of mean earnings) is MO, which has lower mean earnings than all the other states, and has lower mean earnings than MA, NY, and PA, with statistical significance. For the federal government, such insights can be beneficial in devising educational policies.

V. DISCUSSION

The main contribution of the presented research emanates from the prevalent usage of statistical comparison of means in the scientific literature. While the proposed visualizations (arc diagram, slope chard, hierarchical graph) have been applied using data from various domains with success, to the best of the authors' knowledge, they have not been applied before for the statistical comparison of means, with data coming from multiple samples. It is hoped that the adoption of this and other visualization schemes for interpreting statistical results will contribute to the conduct, efficiency, and effectiveness of scientific research, across all disciplines.

One challenge to the validation of the proposed methodology can be its reliance on p-values. From a theoretical point of view, it is not perfectly reliable to use only p-values for comparing the magnitude and significance of differences. This is because the z-values in the pairwise comparisons of the Dunn's test, which in turn affect the p-values, are dependent not only on the differences, but also the sample sizes. Therefore, the proposed methodology can be refined to accommodate sample sizes into the analysis and thus resolve this challenge.

VI. Conclusions and Future Work

A critical innovation of the proposed methodology is the incorporation and application of graphs and graph layout algorithms for visualizing stochastic dominance.

The research can be advanced in various ways:

The benefits of the proposed visualization schemes can be measured through usability tests.

Other visualization schemes that are used for comparisons, as reviewed by [38], can be adopted in order to visualize stochastic dominance.

On the arc diagram, arcs and related nodes can be highlighted when the analyst holds the cursor on or clicks a node, as illustrated in [39].

Arcs on the arc diagram can display additional data mapped onto thickness, color, arrow type, and other attributes [40].

While many schemes exist for visualizing the distribution and central tendency of values in multiple samples, this paper presents for the first time, schemes that display the statistical significance of differences and thus the stochastic dominance relations, all in a single connected graph. Statistical comparison of means is a relatively simple yet fundamental analysis, universally used in all areas of applied sciences. The presented visualization schemes can accompany statistical comparison tests and can be used to facilitate and enhance their application in any domain.

Acknowledgment

The authors thank Eda Ulgen Aydın for her support throughout the project. The authors also thank the developers of the open source software and the freely available commercial software used throughout the study.

Appendix

Throughout the study, the following software tools have been used for the mentioned tasks:

- *Summary Statistics and Kruskal-Wallis H test:*
  rstats R package [41].
- *Dunn's pairwise comparison test:*
  dunn.test R package [42].
- *Sorted box-and-whisker plot:*
  IBM SPSS 24 software [43].
- *Hierarchical (layered) graph:*
  yEd graph drawing software [44] and the hierarchical layout algorithm.
- *Arc diagram and slope chart:*
  MS PowerPoint was used together with image editing software. While there exist several R packages [45] and other software for drawing both the arc diagram and the slope chart, a more convenient tool was selected in this paper, to implement a minimalist design and facilitate cognition by the analyst.

References

[1] G. Ertek, X. Chi, and A. N. Zhang, "A framework for mining RFID data from schedule-based systems". IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 11, pp. 2967-2984, 2016.

[2] A. Demiriz, "Using pairwise associations for multi-item markdown optimization". International Journal of Systems Science: Operations & Logistics, vol. 4, no. 2, pp. 108-121, 2017.

[3] G. Tokdemir, and N. E. Cagiltay, "Investigating the Relationship between SLOC and Logical Database Measures to Improve the Early Estimation of Software Cost". International Journal of Software Engineering and Knowledge Engineering, vol. 29, no. 3, pp. 401-413, 2019.

[4] S. Asian, G. Ertek, C. Haksoz, S. Pakter, and S. Ulun, "Wind turbine accidents: A data mining study". IEEE Systems Journal, vol. 11, no. 3, pp. 1567-1578, 2016.

[5] D. Lu, Y. Ding, S. Asian, and S. K. Paul, "From supply chain integration to operational performance: The moderating effect of market uncertainty". Global Journal of Flexible Systems Management, vol. 19, no, 1, pp. 3-20, 2018.

[6] J. K. Pool, S. Asian, S. M. Arabzad, H. B. Jamkhaneh, and J. K. Lashaki, "Development of a model to analyse the factors affecting RFID technology acceptance in small and medium-sized enterprises". International Journal of Services and Operations Management, vol. 28, no. 4, pp. 468-494, 2017.

[7] G. Ertek, N. Kasap, S. Tokman, Ö.Bilgin, and M. İnanoğlu, "Competitive pattern-based strategies under complexity: the case of Turkish managers". In A Systemic Perspective to Managing Complexity with Enterprise Architecture. IGI Global, pp. 301-330, 2014.

[8] D. Lu, S. Asian, G. Ertek, and M. Sevinc, "Mind the perception gap: An integrative performance management framework for service supply chains". International Journal of Physical Distribution & Logistics Management, vol. 49, no. 1, pp. 33-51, 2019.

[9] N. Nistor, A. Göğüş, and T. Lerche, "Educational technology acceptance across national and professional cultures: a European study". Educational Technology Research and Development, vol. 61, no. 4, pp. 733-749, 2013.

[10] J. Hadar, and W. Russell, "Rules for ordering uncertain prospects". American Economic Review, vol. 59, no. 1, pp. 25–34, 1969.

[11] O. J. Dunn, "Multiple comparisons using rank sums". Technometrics, vol. 6, no. 3, pp. 241–252, 1964.

[12] R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots". The American Statistician, vol. 32, no. 1, pp.12–16, 1978.

[13] FlowingData. "Length of the average master's thesis," 2015 [Online]. Avaiable: https://flowingdata.com/2015/06/09/length-of-the-average-masters-thesis/. Short URL: https://is.gd/g0NqdU.

[14] G. J. Matthews, Olympics Boxplot, 2012 [Online]. Available: https://statsinthewild.com/2012/07/09/olympics-boxplot/. Short URL: https://is.gd/QIluMO.

[15] A. Kassambara, "Facilitating exploratory data visualization: Application to TCGA genomic data," 2017 [Online]. Available: http://www.sthda.com/english/articles/24-ggpubr-publication-ready-plots/77-facilitating-exploratory-data-visualization-application-to-tcga-genomic-data/. Short URL: https://is.gd/myaEdu.

[16] M. Smith, "How good is "good"?," 2018 [Online]. Available: https://yougov.co.uk/topics/lifestyle/articles-reports/2018/10/02/how-good-good. Short URL: https://is.gd/Y3gssk.

[17] A. Emery, "How to visualize statistically significant p-values with squares," 2018 [Online]. Available: https://depictdatastudio.com/statistically-significant-p-values/. Short URL: https://is.gd/dfPCoR.

[18] A. Ghoordjani, "Why should I use a Kruskal Wallis Test?," 2016 [Online]. Available: https://www.researchgate.net/publication/305911256_Why_should_I_use_a_Kruskal_Wallis_Test. Short URL: https://is.gd/gPvode.

[19] H. Gibson, J. Faith, and P. Vickers, "A survey of two-dimensional graph layout techniques for information visualisation". Information Visualization, vol. 12, no. 3-4, pp. 324-357, 2013.

[20] G. Ertek, G. Tokdemir, M. Sevinç, and M. M. Tunç, "New knowledge in strategic management through visually mining semantic networks". Information Systems Frontiers, vol. 19, no. 1, pp. 165-185, 2017.

[21] N. Cagiltay, G. Tokdemir, Ö. Kılıç, D. Topallı. "Performing and Analyzing non-Formal Inspections of Entity Relationship Diagram (ERD)". Journal of Systems and Software, vol. 86, no. 8, pp. 2184-2195, 2013. DOI: 10.1016/j.jss.2013.03.106.

[22] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, "Designing the user interface: strategies for effective human-computer interaction". Pearson, 2016.

[23] G. Shmueli, P. C. Bruce, I. Yahav, N. R. Patel, and K. C. Lichtendahl Jr, "Data mining for business analytics: concepts, techniques, and applications in R." John Wiley & Sons, 2017.

[24] O. Kumar, A Goyal. "Visualization: a novel approach for big data analytics". In 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT). pp. 121-124. IEEE. 2016.

[25] Information Visualization. 2019 [Online]. Available: https://journals.sagepub.com/home/ivi/. Short URL: https://is.gd/x2Dsit.

[26] R. Lengler, and M. J. Eppler, "Towards a periodic table of visualization methods for management". In IASTED Proceedings of the Conference on Graphics and Visualization in Engineering, Florida, USA, 2007.

[27] The R Graph Gallery. 2019 [Online]. Available: https://www.r-graph-gallery.com/. Short URL: https://is.gd/NzEYHU.

[28] A. Abela, "Choosing a good chart," 2006 [Online]. Available: https://extremepresentation.typepad.com/blog/2006/09/choosing_a_good.html. Short URL: https://is.gd/gn9PJL.

[29] Data Viz Project, A Project in Beta by Ferdio. 2019 [Online]. Available: https://datavizproject.com/. Short URL: https://is.gd/m1DH31.

[30] K. Sugiyama, S. Tagawa, and M. Toda, "Methods for visual understanding of hierarchical system structures". IEEE Transactions on Systems, Man, and Cybernetics, vol. 11, no. 2, pp. 109-125, 1981.

[31] P. Healy, and N. Nikolov, "Hierarchical Graph Drawing", in Tamassia, Roberto (ed.), Handbook of Graph Drawing and Visualization, CRC Press, pp. 409–453, 2014.

[32] M. Wattenberg, "Arc diagrams: Visualizing structure in strings". In IEEE Symposium on Information Visualization, INFOVIS, pp. 110-116, 2002.

[33] E. R. Tufte, "The visual display of quantitative information". Cheshire, CT: Graphics press, vol. 2, 2001.

[34] Slope Chart. Data Viz Project, A Project in Beta by Ferdio. 2019 [Online]. Available: https://datavizproject.com/data-type/slope-chart/. Short URL: https://is.gd/7CET7T.

[35] College Scorecard Data. U.S. Department of Education. 2019 [Online]. Available: https://collegescorecard.ed.gov/data/documentation/. Short URL: https://is.gd/Twm3fL.

[36] Data Documentation for College Scorecard. U.S. Department of Education. 2018 [Online]. Available: https://collegescorecard.ed.gov/assets/FullDataDocumentation.pdf. Short URL: https://is.gd/Wtes50.

[37] Census Regions and Divisions of the United States. U.S. Census Bureau. 2016 [Online]. Available: https://www2.census.gov/geo/pdfs/reference/ua/Defining_Rural.pdf. Short URL: https://is.gd/BkJ6zo.

[38] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization". Information Visualization, vol. 10, no. 4, pp. 289-309, 2011.

[39] D3.js Graph Gallery. Arc Diagram. 2019 [Online]. Available: https://www.d3-graph-gallery.com/arc.html. Short URL: https://is.gd/tAPYQk.

[40] Arc Diagram. Data Viz Project, A Project in Beta by Ferdio. 2019 [Online]. Available: https://datavizproject.com/data-type/arc-diagram/. Short URL: https://is.gd/jP8LBK.

[41] rstats R Package. 2019 [Online]. Available: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html. Short URL: https://is.gd/62kPco.

[42] dunn.test R Package. 2019 [Online]. Available: https://cran.r-project.org/web/packages/dunn.test/index.html. Short URL: https://is.gd/hjl45m.

[43] IBM SPSS 24. 2019 [Online]. Available: https://www.ibm.com/products/spss-statistics. Short URL: https://is.gd/Yyyg0G.

[44] yEd. yWork, The Diagramming Company. 2019 [Online]. Available: https://www.yworks.com/products/yed. Short URL: https://is.gd/j0pP7T.

[45] arcdiagram R Package. 2019 [Online]. Available: https://github.com/gastonstat/arcdiagram/. Short URL: https://is.gd/BUB6aY.