

Visualizing trends and clusters in ranked time-series data

Michael B. Gousie^a, John Grady^b, and Melissa Branagan^a

^aDepartment of Math & Computer Science, Wheaton College, Norton, MA, USA;

^bDepartment of Sociology, Wheaton College, Norton, MA, USA

ABSTRACT

There are many systems that provide visualizations for time-oriented data. Of those, few provide the means of finding patterns in time-series data in which rankings are also important. Fewer still have the fine granularity necessary to visually follow individual data points through time. We propose the Ranking Timeline, a novel visualization method for modestly-sized multivariate data sets that include the top ten rankings over time. The system includes two main visualization components: a ranking over time and a cluster analysis. The ranking visualization, loosely based on line plots, allows the user to track individual data points so as to facilitate comparisons within a given time frame. Glyphs represent additional attributes within the framework of the overall system. The user has control over many aspects of the visualization, including viewing a subset of the data and/or focusing on a desired time frame. The cluster analysis tool shows the relative importance of individual items in conjunction with a visualization showing the connection(s) to other, similar items, while maintaining the aforementioned glyphs and user interaction. The user controls the clustering according to a similarity threshold. The system has been implemented as a Web application, and has been tested with data showing the top ten actors/actresses from 1929-2010. The experiments have revealed patterns in the data heretofore not explored.

Keywords: Visualization, time-series, clusters, Web application

1. INTRODUCTION

Time series data analysis is an important aspect in many applications. A variety of systems include complex visualizations for finding patterns within huge data sets (e.g., parallel coordinates and the like). However, there are applications in which finer-grained visualizations are required. For example, we wish to explore relationships among a few thousand points. Furthermore, these data points reflect several attributes, each of which we might want to investigate individually and within the context of the entire data set (focus+context). It is also desirable to be able to follow one individual over the course of a given time period. Most systems do not allow for this kind of fine-grain exploration.

We have implemented the Ranking Timeline, a prototype Web-based system for visualizing relationships, in the form of patterns and clusters in ranked data over time. The tool has two main components that allow the user to view ranking changes over time as well as generate clusters according to a given similarity percentage. Although the system has been developed with a database containing information describing actors/actresses, we show how the tool can be extended to any data that contains rankings.

The system incorporates the following contributions:

- an overall regular-grid framework for the visualization of time-series rankings and cluster analysis;
- an improved line plot-like visualization method for coarse- and fine-grain exploration of top ten-ranked time-series data, including various filters for highlighting or culling chosen attributes;
- a cluster analysis visualization of time-series data that displays results in the context of a timeline;
- in the cluster visualization, a novel notion of a *king* that defines the centroid of each cluster;

Further author information: (Send correspondence to M.B.G.)

M.B.G.: E-mail: mgousie@wheatoncollege.edu

J.G.: E-mail: jgrady@wheatoncollege.edu

- options for viewing various attributes are available in both visualizations and can be enabled/disabled identically;
- the system is implemented as an easily accessible Web application, and uses a browser’s scrollbar to enhance scalability of the visualizations.

2. MOTIVATION

The Ranking Timeline could be used for many data sets for which there is a top ten ranking over time, such as baseball player statistics, school or individual test performance, country GDP, and the like. However, the motivation for this work stems from an investigation of the poorly understood process of collective behavior from the viewpoint of sociology. In particular, we wish to determine why certain actors/actresses (henceforth the term “actor” will be used) capture the public’s imagination, and what the relationships between different people may be. Furthermore, these actors, or stars, form a symbolic community, which is stratified hierarchically. It may be illuminating to determine which star shines the brightest during different time periods.

In order to understand the hierarchy and the relationships among actors, a visualization system is required that shows the top ten stars each year. Variables such as gender, race/ethnicity, and the age at which an actor first appeared should be included as well. The visualization should show the interactions among the top ten actors over all of the years that they appear. Finally, a cluster analysis and visualization would show the relative importance of actors over time.

There is a database that contains information that could be used to identify just who is in the hierarchy of stars at any given year, their persistence over time, and the symbolic relationships that exist between them. The Quigley poll¹ was created in 1932 to identify the top money-making stars. These were the actors that a survey of distributors chose, in rank order, as the people who filled their theaters in the year of the poll. Since 1929, some 206 movie actors have appeared in the poll. Film scholars routinely use the Quigley poll to discuss the popularity of a given star or to identify those stars who were most popular during a given decade.² But rarely do they attempt to discuss any other patterns that a rank ordered poll of this sort might make possible. In particular, it is difficult to identify natural clusters of stars, which might illuminate links between different kinds of personas. Finding such clusters might give clues as to the psyche of American society at large during the same time period.

3. RELATED WORK

Time-Series Visualization. There are myriad systems that display time-series data. A comprehensive survey can be found in Aigner et al.³ derived from an earlier paper;⁴ the TimeViz Browser, an excellent visualization of all the systems discussed in the book can be found at www.timeviz.net. A very common method of displaying time-oriented data is through the simple line plot.⁵ TimeSearcher⁶ shows multiple time-series and also allows the specification of a time period through its *timebox*. The Braided Graph⁷ allows one to superimpose silhouette graphs to show multivariate data. The Horizon Graph⁸ uses two-tone coloring techniques to visualize multivariate data. In both cases, scaling is a problem.

Another general method for displaying time-series data is through stacked graphs.⁹ The Layer Area Graph¹⁰ builds on this when the data contains the same unit that can be summed up. Havre et al.’s ThemeRiver¹¹ shows the importance of a topic by increasing/decreasing the thickness of its flow through time. This has been made quite popular by the online Baby Name Wizard (www.babynamewizard.com);¹² the design does not handle crossing lines well, however, nor starts-and-stops in the data.⁴ A combination of ThemeRiver and parallel coordinates is described in Butkiewicz et al.¹³ Also extending the ThemeRiver is the more recent RankExplorer,¹⁴ which provides level-of-detail views for value and ranking changes.

The aforementioned systems are basically linear, with time flowing from left-to-right. Other systems exist that follow a completely different model. One example uses a radial layout; specifically, the TimeWheel¹⁵ can display many variables in a time-series. Another is KronoMiner,¹⁶ which also uses a circular arrangement of data. While scaling is typically an issue with such a format, the system allows the user to navigate and focus on various elements.

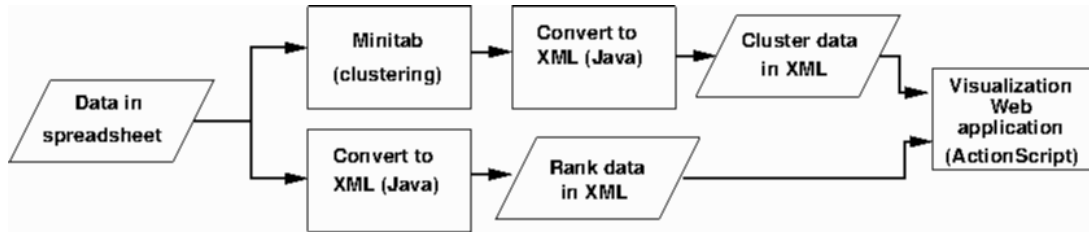


Figure 1. Flowchart showing system components.

Cluster Visualization. Systems for displaying clusters are many and varied. For example, cluster size may be represented by color in a heat map.¹⁷ A nicer visualization of a dendrogram turned on its side, and thus can “grow” downwards, is described by Wang et al.¹⁸ Another approach is HOV^{3,19} an improved method to display clusters based on Star Coordinates.²⁰ However, these systems do not handle the notion of time independently from the clustering; there are no provisions for observing the clusters within the context of time.

4. SYSTEM OVERVIEW

We have implemented the Ranking Timeline as a Web application. It is written in ActionScript 3.0 using Adobe Flash CS6, and is available with the full movie actors database and a small test data set of major league baseball player batting statistics at <http://cs.wheatoncollege.edu/~mbranagan/>. The system encompasses two main functions: a visualization of actor rankings over the years 1929 to 2010 and a cluster visualization for the same time period. Data, originally in spreadsheet format, is converted and stored in XML files. Using XML is a convenient format for extending this tool to any top ten-ranked time-series data. All computations for the actor rankings visualization portion are done on-line. Computations for the clustering are currently completed off-line using Minitab,²¹ a statistical software package. The output from Minitab is converted to XML using a short Java program. Figure 1 shows the various components of the system. See Section 6 for a discussion regarding extending the tool for other data files and for being completely self-contained.

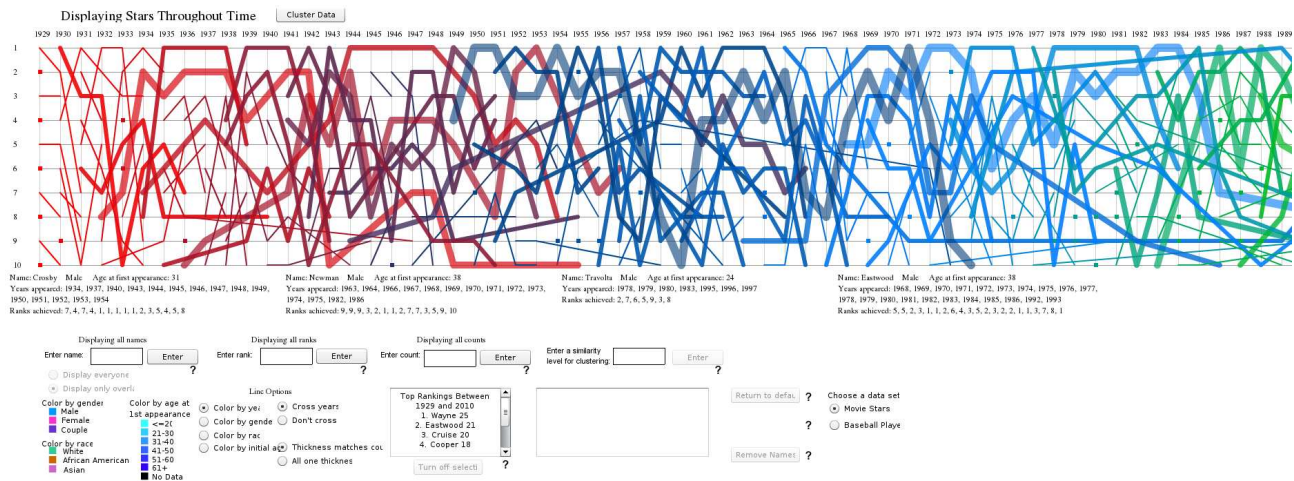


Figure 2. The Ranking Timeline initial screen.

The Web application comprises both the ranking and the cluster visualizations. Each tool uses the same framework of a regular grid, where the x -axis represents the year and the rankings $\{1..10\}$ on the y -axis. Since the visualizations run on a browser, the x -axis may grow to almost any length, yielding good scalability; the scrollbar allows the user to view any portion of the data. The initial screen of the Ranking Timeline is shown in Figure 2; a view of the cluster tool is shown in Figure 3. In both cases, the system embodies the seven tasks as

described in Shneiderman’s seminal paper:²² *overview*, as seen in the previous figures; *zoom*; *filter*; *details-on-demand*; *relate*; *history*, insofar as the browser’s back and forward buttons may be used; and *extract*. The details of the system as it relates to these tasks are covered in the following sections.

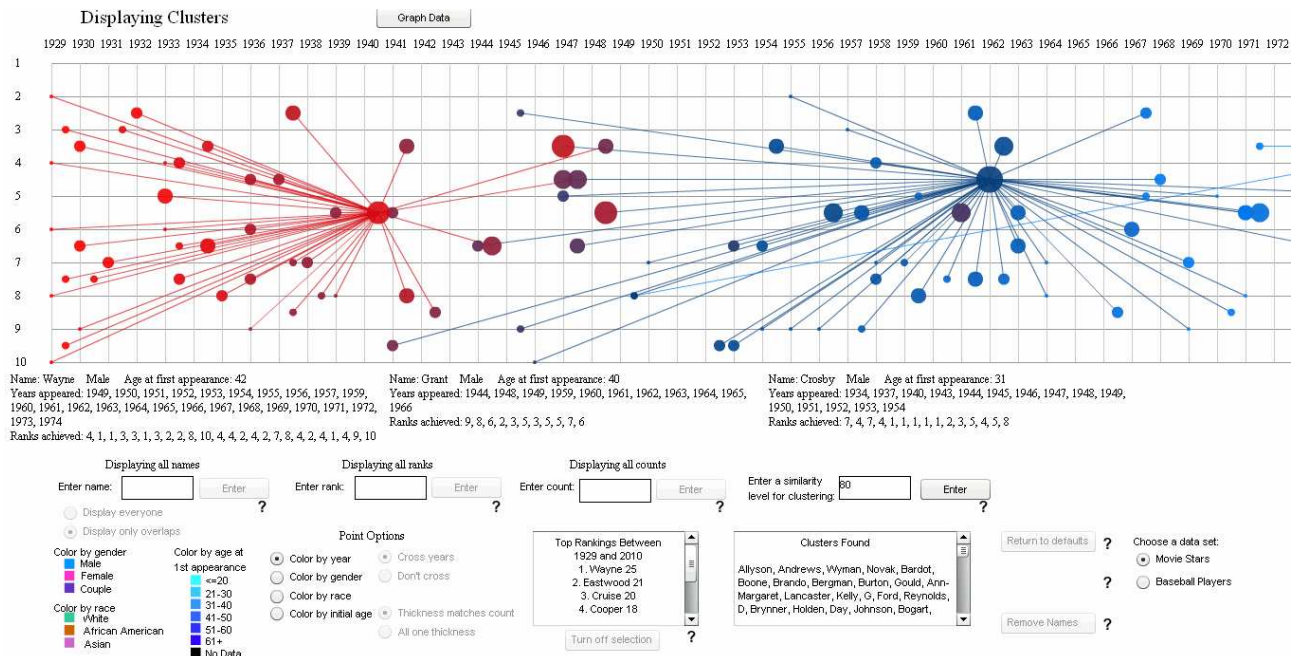


Figure 3. Overview of cluster visualization for years 1929-1972.

4.1 Rankings Over Time

Referring to Figure 2, the initial configuration of the tool shows the top ten actors from 1929 through 2010 displayed on a regular grid. Each actor is represented by a colored square at the intersection of the rank and year. If the star appears in the rankings again, a line is drawn between occurrences. The more times the actor appears, the thicker the line. Initially, the color of the line represents the year in which the actor first appeared in the list; thus, one can track an individual star easily in both directions while at the same time observing all of the interactions that occur. With these simple visual aids, one can easily see how influential John Wayne was by observing the thick blue-gray line starting at 1949 and continuing until 1974. Details for an individual actor are retrieved by clicking on a square or a line, and are shown in plain text below the plot. The information includes the actor’s name, gender, age at first appearance, years appearing in the top ten, and the ranks achieved during those years. Finally, the ten stars that appear the most often in the list are shown in a scrollable text box.

Options. There are myriad options for the user to investigate the data. While the Ranking Timeline is generally similar to the basic time-series visualization⁵ and line plots as used in TimeSearcher,⁶ there are important differences. One factor is that actors come and go from the top ten, so visualizations that assume the data continues from start to end are not viable. Thus, there are no lines that span the entire range of years. This is also the impetus for the design decision of coloring the lines according to the year an actor first appeared in the rankings so as to better see when (s)he first appeared and when (s)he left the list. The varying-thickness lines may help find relationships, but may also be distracting. Thus, the user may opt to turn the thickness glyph off, making all of the lines a uniform size, creating a neater view of the visualization. Furthermore, an important feature is that, initially, the line for a particular actor begins when (s)he first appeared in the rankings and continues until the last appearance. This means that a line may span years in which the actor actually did not appear in the top ten. Therefore, another option allows the user to choose whether the lines “cross” or not. A crossing line means that it crosses over years in which the actor was not in the rankings. If the “Don’t cross”

option is chosen, the lines connect only *contiguous* appearances in the list (see Section 5 for examples). Another feature is that the lines may be colored by gender, race/ethnicity, or the actor’s initial age, rather than the year of the star’s first appearance. Filtering options include choosing an individual actor; in this case, his or her line is highlighted in orange while all of the other stars whose appearance intersects with that actor’s are shown in light gray. This gives a clear indication not only of that specific actor’s rise and/or fall but also of the star’s cohort during those years. In this mode, other actors may be chosen individually, and their lines will revert back to their original color. Displaying only those whose rank is greater than or equal to a given value and showing only those who appeared at least a given number of years are additional options. These allow the user to view additional sub-categories of actors.

In many studies, researchers wish to investigate a certain decade or other time frame. The Ranking Timeline offers a feature that allows the user to choose a time interval by clicking a start and end year the top of the plot. The time frame is highlighted by yellow boundaries; the plot to the left and right of the area of interest is still shown but with a light gray mask so as to lessen distraction but yet still provide context. This is similar to the Timebox in TimeSearcher.⁶

Finally, a zoom feature allows the user to see the entire plot, ostensibly to find global patterns. Although this needs improvement, Figure 4 shows this feature with the gender option turned on. It clearly shows how actresses dominated the box office in the early years of film compared to their male counterparts. Women (Doris Day and Elizabeth Taylor) again were the top draws from 1960-1964; however, women have been less prominent since then. Further examples using this data set are covered in the case studies; see Section 5.1.



Figure 4. Entire timeline of actors showing gender option.

4.2 Clustering

We are concerned with grouping “similar” actors into classes, or clusters. There are many and varied methods of computing clusters; in our case, we wish to find clusters in which the distance between individual actors is minimized, which in turn means that they are similar. We also wish to find the *king*, or most influential actor within a cluster. K-means clustering²³ is a technique that finds a user-specified number of clusters K which are represented by their centroids. However, because the number of clusters must be known *a-priori*, and other problems such as not handling outliers well, we did not employ this method. Instead, we used the agglomerative hierarchical clustering technique.²³ In this method, the distance d between two data items must be calculated. A common way to do this is by calculating the Euclidean distance between two observations:

$$d = \|a - b\|_2 = \sqrt{\sum_i (a_i - b_i)^2} \tag{1}$$

Clusters are formed by linking observations or sets of observations together. We used the average linkage method, which finds distance between sets of observations, A and B :

$$\frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b) \tag{2}$$

The two data points whose average linkage is closest form a pair. Note that variables are non-standardized; this is because we want to cluster partially by year, so this attribute *should* have a more pronounced affect on the results.

In order to compute the distance, the most appropriate attributes (a_i and b_i above) must be chosen for a particular study; choosing a different set of attributes can have a profound effect on the resulting clusters. In this case, we chose three attributes we thought most important to defining the relative importance of an actor: the median of all years in the top ten (this places the actor within a particular era), the median rank (stars at similar ranks ought to be clustered together), and the mean degree. The degree is defined as the number of other stars, s , that are in the top ten at least once during the years y that a given actor is in the list. In other words, $\frac{s}{y}$ is a measure of the number of stars that intersect with a given actor. This variable, then, is a representation of the actor's cohort.

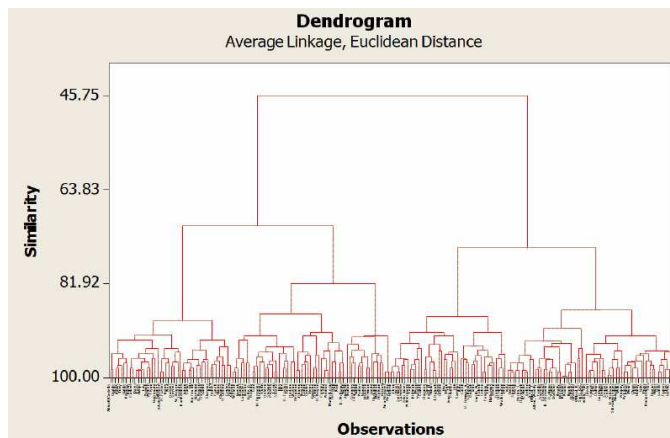


Figure 5. Dendrogram for all actors.

The typical method of displaying clusters is through a dendrogram, as shown in Figure 5. The bottom of the chart shows all the actors; each connection represents a pair of actors or group of actors who are close in distance as defined above. At the bottom, individuals have a similarity measure of 100%, and as one moves up the chart, the similarity percentage diminishes. A dendrogram, although useful in showing overall patterns, is not ideal for viewing small clusters; that is, a cluster at a particular similarity. The chart does not scale well, as is obvious in the figure. It is also not possible to determine which *one* actor is most similar to the rest within a given cluster. Instead, referring back to Figure 3, the cluster tool provides a visualization for both of these items. In this case, the similarity level is set to 80%. The colored circles represent individual actors; as in the Ranking Timeline, the color indicates the first year the actor appeared in the top ten. The x -axis placement of the data point is at the median year of all years in the top ten, and the y -axis placement is at the median rank. The size of the circle is scaled according to the number of years the actor appeared in the rankings. A text box lists all of the actors in each cluster. Most of the options that are available in the ranking visualization, such as changing the color-coding, choosing a time frame, etc., are still accessible and have the same user interface.

In the figure, which shows the years between 1929 and 1972, two clusters are clearly visible. The cluster on the right has John Wayne as its central figure. The large size of his blue circle and the text below the chart shows that he appeared in the top ten 25 times starting in 1949. Just below and to the left is a circle colored dark violet, indicating that this actor appeared in the rankings much earlier; indeed, it is Cary Grant whose first rank of nine occurred in 1944. However, Grant's proximity to Wayne indicates the former was in the rankings for a quite a long time period (1944-1966). One might also note the large red circle in 1947 that is connected to Wayne; although important, according to the clustering algorithm, Bing Crosby is connected to John Wayne. What is unique to this visualization and that is not represented in the dendrogram is that Wayne is, in fact, the central figure of the cluster. Because high rankings and longevity are the most important measures in this data set, the *king* of a cluster is found by finding the maximum k -value of all the actors:

$$king = \max\{k_1, k_2, \dots, k_n\} \tag{3}$$

where n is the total number of actors. Furthermore,

$$k_i = \sum_{j=1}^{years} \frac{l}{rank_j}, \forall i \in \{1..n\} \quad (4)$$

where *years* represent the total number of years an actor appeared in the rankings and $l = 10$ is the lowest possible rank. Thus, the clustering and the dendrogram show that Wayne and Crosby are connected, but our method shows Wayne is king because $k_{Wayne} > k_{Crosby}$ even though Crosby’s mean ranking is higher than Wayne’s.

5. CASE STUDIES

In this section, we illustrate how the system can be used to analyze two different ranked time-series databases: Actors from 1929 to 2010 and a smaller data set of baseball players from 1990 to 2010.

5.1 Case Study 1: Actors

Suppose one wished to study the decade of the 1970s. Which actor had the biggest influence in that decade? By finding this, one could match the actor’s films to the politics and ideals of the decade to compare the affects of society on popular culture and vice-versa.

Immediately it is clear that the dendrogram of Figure 5 is not helpful, as it provides no provisions for viewing the years. Looking at the Ranking Timeline for the years 1955 to 1995 (Figure 7) is somewhat daunting. However, we can choose the time frame we wish, the 1970s, as shown in Figure 8. This highlights the decade allowing the user to focus on a much smaller set of lines while keeping the years before and after in context. The visualization was cleaned up a bit by choosing the “Don’t cross” option; this reduces clutter by eliminating lines between non-consecutive years in the top ten. By clicking on the line that is at the number one ranking from 1974-1976, we find it is Robert Redford. Note that name labels and other actor information appear under the visualization and are not shown in the figures so as to save space; see Figure 2 for a view of this information. Typing in Redford highlights his rankings in orange while changing the color of his cohorts to gray, as shown in Figure 9. This also eliminates all actors from the view except those whose appearance in the rankings intersect with Redford’s. This affords an even simpler view, allowing us to click on two additional major players (those actors with at least two top ten rankings in the decade) to highlight those lines, which turn back from gray to the original shade of blue. These two actors are Clint Eastwood and Burt Reynolds. Notice how Eastwood’s line begins before 1970 and ends far beyond, in 1985; Reynold’s, by contrast, begins in 1973 but also continues until 1984. Redford’s rankings are almost completely within the decade, save for 1980 (ranked 2nd) and 1984 (7th). With the “Cross years” option turned on, these latter two rankings are easily visible; in the figure, these two points are shown as small orange squares. Thus, these options must be used carefully to glean the information one desires.

While the Ranking Timeline’s options give us good insight as to which actors are important in the decade, that tool does not yield any quantitative results. One option is to look at the clusters using the dendrogram. Figure 6 shows the portion of the dendrogram that contains all three actors. In fact, Eastwood and Reynolds are clustered together *immediately*, signifying that they are quite similar. However, who is “more important?” We can now turn to the cluster visualization tool, which initially shows the view shown in Figure 10. The largest and second largest top-ranked circles within the decade are Eastwood and Redford, respectively. Interestingly, another large circle, between 1971 and 1972, is Paul Newman, who had not figured into the investigation using the Ranking Timeline alone. Reynolds is the circle to the immediate right of Eastwood, between 1979 and 1980. Clearly, Reynolds impacted the decade of the 80s more than the 70s.

The clusters are displayed according to the similarity percentage given by the user. Figure 11 shows the clusters at the level of 80%. Not only are two main clusters shown, but they are anchored by a *king*. The cluster at 1978 is centered about Eastwood (the centroid at 1962 is Wayne), making him the leading candidate for being the most influential in the decade. Indeed, Redford is connected to Eastwood, as is Reynolds. However, because the visualization shows not only the clusters but also the relationship between actors and the years, one notes that Eastwood’s cluster is mostly *outside* of the 1970s. By increasing the similarity percentage to 85%, we get the

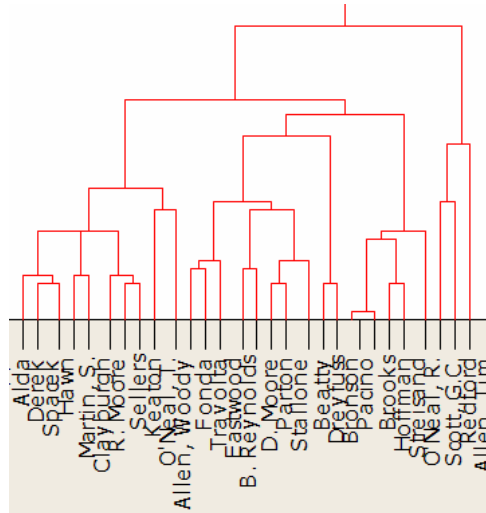


Figure 6. Dendrogram zoomed in on section containing Redford, Eastwood, and Reynolds.



Figure 7. Initial view of all actors between 1955 and 1995.

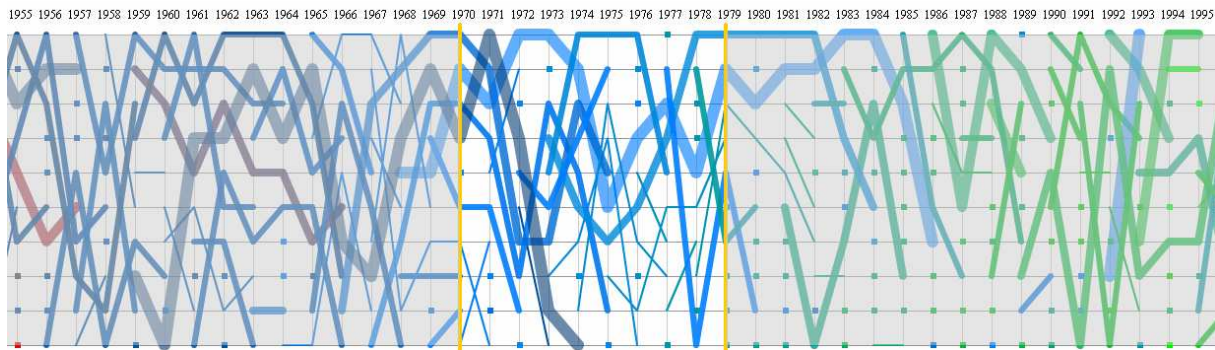


Figure 8. Highlighting the 1970s with crossed lines turned off.

visualization shown in Figure 12. Now Newman is a centroid of a cluster; Redford remains part of Eastwood's cluster. Interestingly, there is no central figure in the middle of the decade, so one might conclude that the 1970s were a time of transition, from the sensibilities of Redford to the more macho image of Eastwood. Sociologists could now match this information with other political, economic, and other social data to succinctly describe the psyche of the population living in that decade. A more comprehensive analysis of the actors data using this tool was given by author Grady at the Conference of the International Visual Sociology Association (IVSA) in July, 2013, which supports the efficacy of the system.

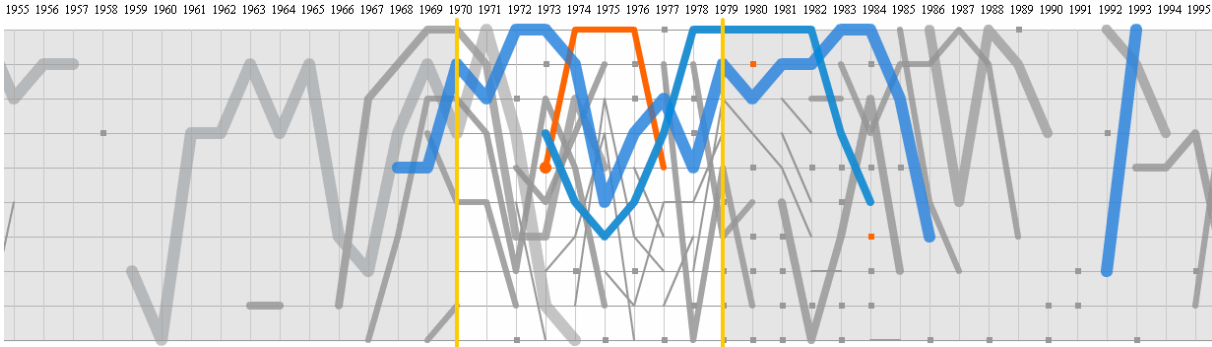


Figure 9. Highlighting Redford (orange), Eastwood, and Reynolds.

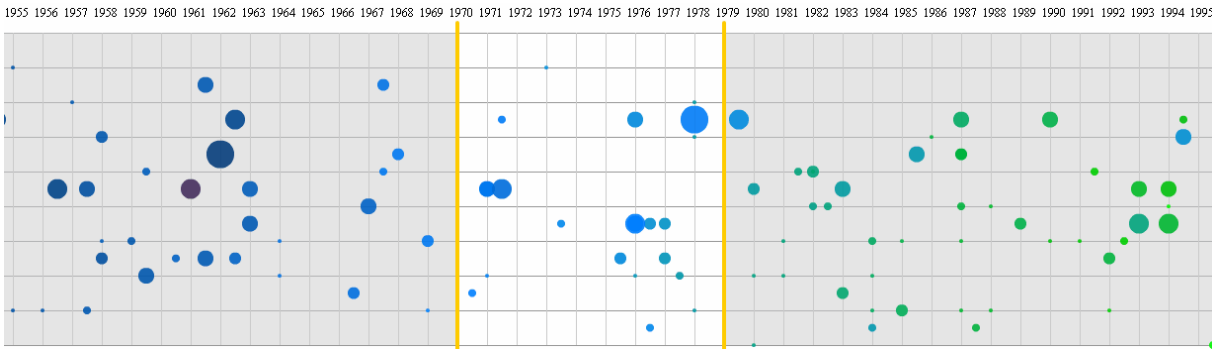


Figure 10. Viewing the relative importance of actors with emphasis on the 1970s.

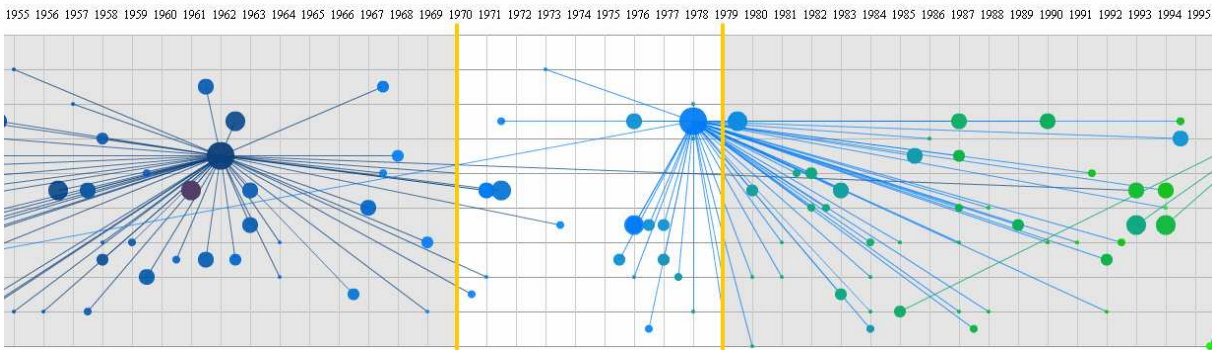


Figure 11. Clusters at the 80% similarity level.

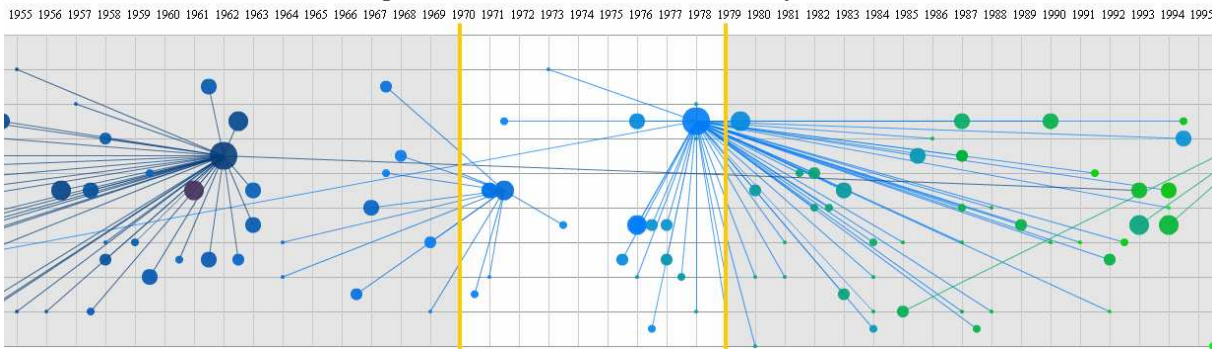


Figure 12. Clusters at the 85% similarity level.

5.2 Case Study 2: Baseball players

Although the Ranking Timeline was initially built for a sociological study of the popularity of actors and actresses through time, it is clear that the tools can be used for other ranked time-series data. To test the current system,

we chose a small database of major league baseball players between 1990 and 2010. The data contains the players with top ten batting averages for each year. This test is a proof of concept; no additional data, such as gender, age, etc., was included.

Figure 13 shows all of the players with the years 2003-2008 highlighted. This is the period in which the Boston Red Sox were a major force in the league; they won the World Series in 2004 and again in 2007. Manny Ramirez was one of Boston’s top sluggers; Figure 14 shows Ramirez’s rankings. It is interesting that he did not appear in the top ten in batting average in 2004, the year the team won the World Series. Furthermore, in the second World Series year, 2007, his rank was only 9th, although the 10th ranked player was teammate David Ortiz. Finally, Figure 15 shows the entire database, this time with only consecutive years in the rankings connected. The difference between this data set and the actors is striking; it seems to be much more difficult to achieve a top-ranked batting average year after year compared to box office success.

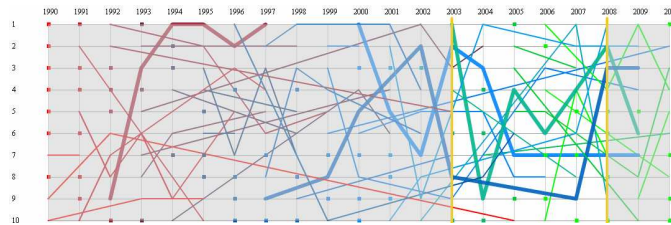


Figure 13. Top ten batting averages from 1990 to 2010, highlighting 2003-2008.

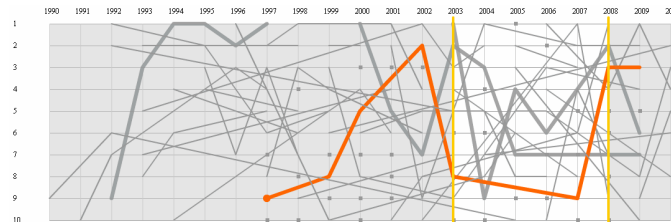


Figure 14. Highlighting Manny Ramirez.

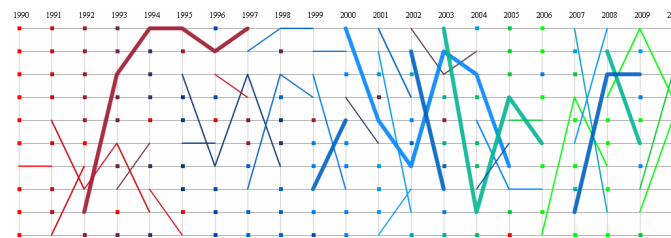


Figure 15. View of all players connecting only consecutive years.

6. CONCLUSION AND FUTURE WORK

We have presented the Ranking Timeline, a visualization that incorporates ranked time-series data. The system extends line plots in several ways, always in the context of a regular grid that shows the rank and date. One part of the system shows relationships among actors in the top ten rankings; a second portion shows clusters using the agglomerative hierarchical clustering technique in the context of the overall system timeline. The system has been implemented and is available on the Web. The case studies show how the user can interact with the system to follow individual or groups of actors and to tease out useful patterns and information, which can then be used for further sociological study.

An advantage of the system is that because of the use of a regular grid in both visualizations, the user may switch easily from one view to the other with minimal distraction. Color and line thickness show the various

attributes of the data without introducing complicated glyphs. Data can be viewed in different scales in time and rank: the entire timeline or a desired portion; all or just certain ranks; and all actors or chosen individuals. All of these are viewable in the context of the rankings and dates. Another advantage is that additional data may be added without scaling problems because the visualization is implemented as a Web application, and therefore the horizontal dimension can grow and be accessed by the browser's scrollbar. Finally, a quantitative method for calculating the most important data point in a cluster yields a visualization with actors connected to a central figure. The system does suffer some limitations. As currently implemented, the system only works with the actors database (and some limited baseball statistics). Furthermore, it was designed specifically for attributes relating to actors; changing these will require extensive work. In the timeline visualization, there is the problem of visual clutter. Some users feel daunted when seeing the visualization for the first time; however, this may not be more problematic than many previous systems, such as parallel coordinates or treemaps, for example, which took some time before universal acceptance. More importantly, however, is the problem of attribute scaling; at this point, there are no provisions for providing visualizations for additional attributes. Also related to scaling issues, while adding additional years to the database is not a problem (because a sufficient number of years can always be seen on the screen), changing the rankings from top ten to top N could cause issues. In particular, a grid of more than 15 or 20 rankings would not comfortably fit in a browser window, necessitating a scrollbar in the y dimension. This may not be viable from the user's point of view. However, it is our view that there are many data sets for which the top ten rankings are sufficient. Finally, the clustering computations currently are accomplished off-line, and only one version of the agglomerative hierarchical clustering method has been tested.

Some of the limitations listed above can be ameliorated in future versions of the system. One planned improvement is to incorporate the clustering computations in the on-line system. This would allow the user to choose the clustering method and to set parameters (what type of linking to use, for example). More importantly, this would be a necessary step in making the system as a whole work with any arbitrary set of ranked time-series data. The ultimate goal of the system is to allow users to input their own data sets, including a description of attributes. The desired result would be a visualization that has the properties of the current system but that would incorporate custom labels, descriptions, etc. that conform to the user's definitions in the data set. These changes would lend the system ultimate flexibility.

REFERENCES

- [1] Quigley, W. (2013). http://www.quigleypublishing.com/MPalmanac/Top10/Top10_lists.html.
- [2] Burr, T., [*Gods Like Us*], Pantheon (2012).
- [3] Aigner, W., Miksch, S., Schumann, H., and Tominski, C., [*Visualization of Time-Oriented Data*], Human-Computer Interaction Series, Springer (2011).
- [4] Aigner, W., Miksch, S., Müller, W., Schumann, H., and Tominski, C., "Visual methods for analyzing time-oriented data," *IEEE Transactions on Visualization and Computer Graphics* **14**(1), 47–60 (2008).
- [5] Spence, R., [*Information Visualization: Design for Interaction*], Pearson (2007).
- [6] Hochheiser, H. and Shneiderman, B., "Dynamic query tools for time series data sets: Timebox widgets for interactive exploration," *Information Visualization* **3**(1), 1–18 (2004).
- [7] Javed, W., McDonnell, B., and Elmqvist, N., "Graphical perception of multiple time series," *IEEE Transactions on Visualization and Computer Graphics* **16**(6), 927–934 (2010).
- [8] Heer, J., Kong, N., and Agrawala, M., "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations," in [*CHI '09: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*], 1303–1312, ACM (2009).
- [9] Byron, L. and Wattenberg, M., "Stacked graphs - geometry & aesthetics," *IEEE Transactions on Visualization and Computer Graphics* **14**(6), 1245–1252 (2008).
- [10] Harris, R. L., [*Information Graphics: A Comprehensive Illustrated Reference*], Oxford University Press (1999).
- [11] Havre, S., Hetzler, B., and Nowell, L., "Themeriver: Visualizing theme changes over time," in [*2000 IEEE Symposium on Information Visualization (InfoVis 2000)*], 115–123 (2000).
- [12] Wattenberg, M., "Baby names, visualization, and social data analysis," in [*IEEE Symposium on Information Visualization (InfoVis 2005)*], 1–7 (2005). <http://www.babynamewizard.com>.

- [13] Butkiewicz, M., Butkiewicz, T., Ribarsky, W., and Chang, R., “Integrating time-series visualizations within parallel coordinates for exploratory analysis of incident databases,” in [*Visual Analytics for Homeland Defense and Security*], Tolong, W. and Ribarsky, W., eds., **7346**, SPIE (2009).
- [14] Shi, C., Cui, W., Liu, S., Chen, W., and Qu, H., “Rankexplorer: Visualization of ranking changes in large time series data,” *IEEE Transactions on Visualization and Computer Graphics* **18**(12), 2669–2678 (2012).
- [15] Tominski, C., Abello, J., and Schumann, H., “Axes-based visualizations with radial layouts,” in [*SAC ’04: Proceedings of the 2004 ACM Symposium on Applied Computing*], 1242–1247, ACM (2004).
- [16] Zhao, J., Chevalier, F., and Balakrishnan, R., “Kronominer: Using multi-foci navigation for the visual exploration of time-series data,” in [*Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*], *CHI ’11*, 1737–1746, ACM (2011).
- [17] Islam, S., Krinke, J., and Binkley, D., “Dependence cluster visualization,” in [*SOFTVIS ’10: Proceedings of the 5th international symposium on Software Visualization*], 93–102, ACM (2010).
- [18] Wang, H., Zheng, H., and Azuaje, F., “Poisson-based self-organizing feature maps and hierarchical clustering for serial analysis of gene expression data,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **4**, 163–175 (April 2007).
- [19] Zhang, K.-B., Orgun, M. A., Zhang, K., and Zhang, Y., “Hypothesis oriented cluster analysis in data mining by visualization,” in [*AVI ’06: Proceedings of the working conference on Advanced Visual Interfaces*], 254–257, ACM (May 2006).
- [20] Kandogan, E., “Visualizing multi-dimensional clusters, trends, and outliers using star coordinates,” in [*Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’01)*], ACM (2001).
- [21] “Minitab,” (2013). www.minitab.com.
- [22] Shneiderman, B., “The eyes have it: A task by data type taxonomy for information visualizations,” in [*Proceedings of the IEEE Symposium on Visual Languages*], 336–343, IEEE (September 1996).
- [23] Tan, P.-N., Steinbach, M., and Kumar, V., [*Introduction to Data Mining*], Addison-Wesley (2005).