Identifying and Querying Local and Large-Scale Structures Using L^∞ Norms in Visual Analytics [*]

Anushka Anand[†] Robert Grossman[‡] Matthew Handley[§] Rajmonda Sulo[¶] Leland Wilkinson[‖]

University of Illinois at Chicago Laboratory for Advanced Computing

Abstract

We introduce a new visual analytic framework based on the L^∞ norm. This framework involves a 3-operator algebra on hyper-rectangles. The We have developed a visual analytic system that generates set-wise rules from simple gestures in an exploratory visual GUI. Logging these rules allows us to apply our analysis to a new sample or batch of data so that we can assess the validity and reliability of our visual model. As an example of this technology, we present an interactive application for visually detecting conflict Wikipedia. We chose the Wikipedia database to illustrate this application because of its feature richness and relatively large scale.

CR Categories: H.5.2 [User Interfaces]: Graphical User Interfaces—Visualization; I.3.6 [Computing Methodologies]: Computer Graphics—Methodology and Techniques;

Keywords: visualization, statistical graphics

1 Introduction

Illuminating the Path [20] defines visual analytics as “the science of analytical reasoning facilitated by interactive visual interfaces.” Many have pointed out the unique strength of visual analytics for this purpose – they exploit the highly evolved pattern detection capabilities of the eye. Visual pattern detection involves a rich environment of features loosely coupled to a diverse set of rules.

The strengths of visual analytics expose their weaknesses, however. It is difficult to construct a visual tool that can be used to identify similar structure in different datasets. When data arrive over time in batches (credit inquiries, purchases), or when data arrive in a continuous stream (phone and web logs, remote sensing data), or when we wish to generalize to a population (experiments, surveys), we cannot visually examine every new piece of evidence.

To cope with this situation, we must find rules we can apply algorithmically to a new sample. Classical and Bayesian statisticians employ parametric or distribution-free models for this purpose. Most statistical packages help with this process. They use scripting languages to estimate models and apply them to new samples of data.

The research presented here explores a new approach that marries visual analytics and inference. We have constructed an environment that facilitates interactive visual exploration and generates a set of rules from visually-oriented actions that can be applied to new data. We are visually identifying structure and formally defining it so that we can query unseen data for similar structure. Our method is neither model nor distribution based.

The basic idea is to design an analytic system around rectangular description regions. The composition of these regions (using three operators) can be used to define local and large-scale structures and provides the basis for a formal description of structures suitable for visual analytics. Users may interactively select outliers, clusters, trends, and other configurations of multidimensional data points and apply their selections to new sets of points in a single click. We discuss tools and architecture that contribute to a familiar user experience while facilitating data exploration and model generation.

2 Local and Large-Scale Structures in Visual Analytics

While the union of open scherical balls is used for defining a basis for the the L^2 Euclidean metric topology, we can alternately use the union of open hypercubes to define the L^∞ metric topology. In this paper, we employ the L^∞ or sup metric:

$$||x||_\infty = \sup(|x_1|, |x_2|, \ldots |x_n|)$$

when we search for nearest neighbors in an L^∞ space. In this search, we are looking for all neighbors of a point at the center of a hypercube of fixed size in a vector space. We can also devise analytics in this metric. Statisticians have employed the L^∞ norm for density estimation. See [24], for example.

The following three definitions are central to this paper:

A *rectangular description region* is the collection of points a fixed distance from a single point (called the center) using the weighted L^∞ norm:

$$||x||_\infty = \sup(w_1|x_1|, w_2|x_2|, \ldots w_n|x_n|).$$

These are also called *hyper-rectangles*, or, more simply, *rectangles*. We usually define these weights locally, so that different points in a high-dimensional space can have different weights defining their rectangles. This approach is similar to locally weighted statistical models that specify different variances in different regions of space.

By a *local structure*, we mean the points defined by a single rectangle. By a *large-scale structure*, we mean the points defined by two or more rectangles under the operations of union, intersection and set-theoretic complement.

Rectangles defined by L^∞ metrics have the following three properties which are fundamental for this paper:

1. Rectangles are naturally closed under the operations of union, intersection, and set-theoretic complement.

2. Complex structures in data are well approximated by collections of rectangles under these three operations.

3. These three operations can be provided with simple, intuitive user interfaces.

[*] Steve Eick provided valuable suggestions.

[†] e-mail:aanand2@uic.edu

[‡] e-mail:grossman@cs.uic.edu

[§] email:mhandley@uic.edu

[¶] e-mail:rsulo1@uic.edu

[‖] e-mail:leland.wilkinson@gmail.com

What benefits do we get from this alternative view? First, it simplifies the specification of neighborhoods because they are product sets of intervals. Our specifications can be expressed in a basic algebra on intervals. In this way, we answer the question posed by Wilhelm [23]: how can we express visual brushing operations in a simple set of rules?

Second, it allows us to specify in simple expressions relatively complex geometric objects through the union of hyper-rectangles. Figure 1 shows how this works for a 2D object. The union of the red rectangles is a fairly good cover for the set of points. Gao and Ester [14] discuss this property further in the context of cluster analysis. We note, of course, that some shapes can require quite a few rectangles for a good cover.

To summarize, rectangles can be used to define local structures in data, as usual. Combining rectangles using the operations of union, intersection, and set-theoretic difference can also be used to define large-scale structures in data.

In this paper, we describe a system in which users can easily define local and large-scale structures in data visually and then query unseen data for similar structures.

Although combinations of rectangles have been used recently to define clusters in data mining [14], this is the first paper that we are aware of that uses this technique in visual analytics.

3 Related work

Perhaps the most widespread use of rectangular description regions is in recursive partitioning trees [5] [16]. These methods partition a space into nested rectangular regions that are relatively homogenous over the values of a predicted variable. Their popularity is due in large measure to the two strengths we have mentioned in this section. Our approach differs from these models, however, because it is not restricted to a partitioning. Our description regions need not be disjoint and exhaustive.

The article that introduced brushing [3] used a square brush. This has become the common usage, even though a circular brush shape is more consistent with the geometric assumptions used in classical statistics. The original motivation for the square shape was computational efficiency, because the developers of the Bell Labs system wanted the user to be able to move the brush rapidly through the cells of a SPLOM. Interval selection for set inclusion is faster than Euclidean distance-based selection. Our motivation for choosing rectangular description regions is somewhat different. Using intervals enables us to simplify the set-wise algebra on the original variables. An additional benefit for us is that they enable SQL queries to a database for SELECT queries on new data.

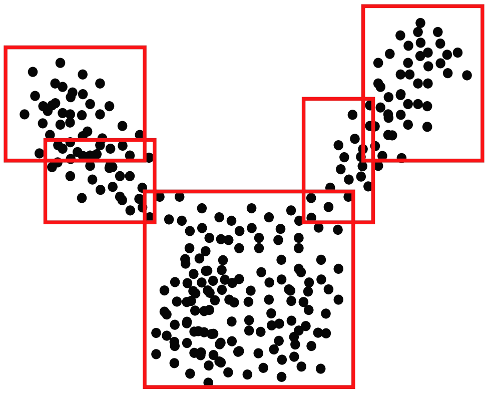


Figure 1: Rectangular cover of 2D point configuration

4 Using the L^∞ Norm to Analyze Wikipedia

We will explain the design of a system to implement L^∞ norm visual analytics by telling a story. Our story has the goal of identifying controversial Wikipedia pages. We visually derive a structure from a sample of pages. Then we apply that structure to another sample of pages.

4.1 Wikipedia Data

We began by downloading the complete English Wikipedia up to November 2006. The database dump contains about 1.8 million pages and the total size of the expanded MySQL database is close to a terabyte. The tables relevant to our analysis were Page, Revision and Text. These enabled us to investigate the revisions of a page, corresponding editors, and the associated text.

We next selected eight features for measuring conflict in Wikipedia Talk pages. We chose these from similar measures in Spertus [18], who developed a classifier for hostile email messages. We also consulted an online journal devoted to malefication [1]. And we consulted [17] [21] [19] [22] for measures peculiar to Wikipedia.

1. **bold**: Number of bold characters in document.
2. **upper**: Number of upper case characters in document.
3. **italic**: Number of italic characters in document.
4. **dirty**: Number of dirty words in document.
5. **indent**: Maximum level of indentation (replies to replies to replies ...) in document.
6. **strikeout**: Number of strikeout characters in document.
7. **editors**: Number of unique editors (Wikipedians).
8. **edits**: Frequency of edits per month.

In order to extract the first six measures, we built a rule engine in Python using a regular expression for each feature. We extracted the last two measures by using a one-month time window. The first six measures were normalized by the number of words in the page.

We now describe the components of the visual system. We will use these components to search for a region of the high-dimensional space of Wiki features that characterizes highly controversial pages. We begin with a single page that we guess will be controversial, we look for L^∞ neighbors, and then use set-wise operations to modify our selection. Finally, we apply our rules to a new sample.

4.2 The Scatterplot Matrix (SPLOM) Window

Figure 2 shows the scatterplot matrix of features. We are looking at a sample of 32,000 Wiki pages in one SPLOM. We implemented a SPLOM display for our eight features and colored the diagonal elements by a set of contrasting hues following rules outlined in [6]. Because we need to handle possibly millions of records, we used hexagonal binning to populate the SPLOM [7].

The SPLOM has five controllers. The first controller allows the user to select a single scatterplot cell by clicking with the mouse in a cell. The remaining controllers are implemented in buttons at the top of the SPLOM window. They allow the user to clear selections, save rules, apply rules to new set of data, and show the text log of the current rules.

We decided that the italic-uppercase window was interesting because it had outliers for both variables that could indicate a high level of controversy. By clicking on this cell, we launched the Scatterplot window described in the next section.

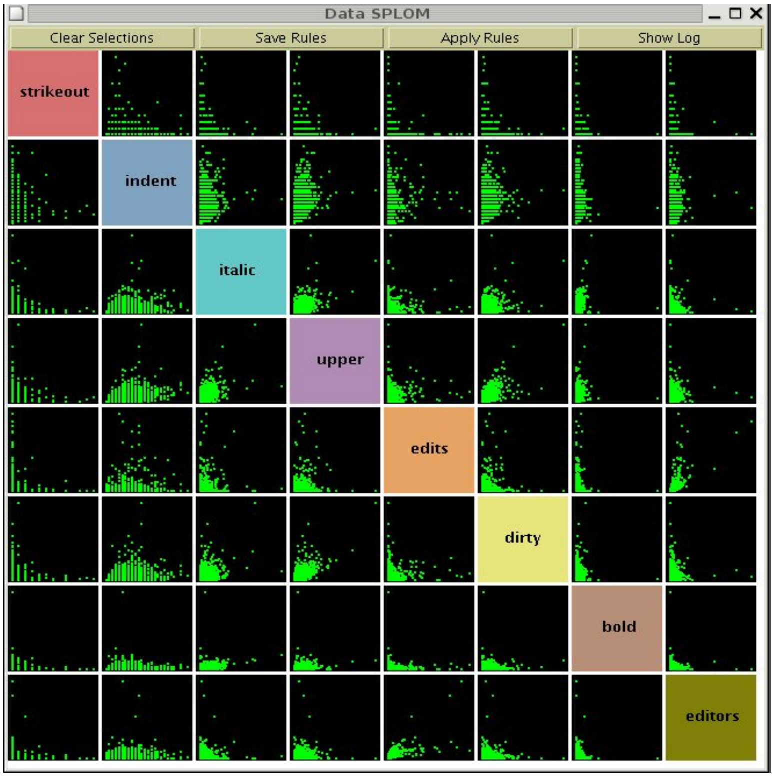


Figure 2: Scatterplot matrix view

4.3 The Scatterplot Window

Figure 3 shows the scatterplot window. The scatterplot window has five controllers. The first controller is a mouse-down rectangle selection widget. This allows the user to define a selection rectangle by dragging the mouse. The last four controllers are implemented in buttons at the top of the window. Find Nearest allows the user to select similar points (documents). The number of nearest neighbors selected or the size of the selection hypercube are controllable in a preference dialog. The remaining buttons implement the setwise operators for the L^∞ algebra. Add implements the union operator by adding points selected in the rectangle to the collection. Remove implements the difference operator by removing selected points from the collection. Restrict implements the intersection operator by retaining points common to overlapping rectangles.

The figure shows our next action. We have selected an outlying point and requested ten near neighbors of this point. Note that the SPLOM and Scatterplot windows are linked so that the selected points are immediately highlighted in red in both windows.

At this point, we notice that several L^∞ neighbors are not outlying in this scatterplot. So we remove them by drawing a new rectangle with the Remove button. Figure 4 shows this action.

Figure 5 shows our next action. We add in some points we think are interesting. They have high indentation and a lot of dirty words. The effects we see are that the number of editors stays consistently low and a few pages with low italic count are selected in the SPLOM.

4.4 The Table Window and the Text Window

The table window contains a grid of the eight features (columns) and all the documents (rows). This window has three controllers. The first is a sort method controller, implemented by clicking on a feature column. The second is a partition controller, activated by clicking on the first column, that moves all selected documents (highlighted in yellow) to the top of the editor. The third controller is activated by clicking on a row. It causes a text window to pop up for the selected document.

The text window is a Java browser programmed to highlight text in the colors used to designate features in the SPLOM window (following [11]). This makes it easy to locate blocks of text that correspond to the features (italic, bold, capital, etc.).

Figure 6 shows both windows. The table view is linked to the SPLOM view and the scatterplot view. Red points in the plots are shown as checked rows in the table view.

Since we began our investigation with points that had high upper count and italics, we now look at the corresponding selection in the table. After sorting by the upper case count, and then sorting by italic count, we pick a point with high values for both measures. We show a snapshot of the corresponding page with a lot of italics highlighted. The page is titled *Creationism*. We see that the talk page has long discussions where editors use italics to identify the sections that they are responding to or discrediting.

4.5 Applying our Rules to a New Dataset

Figure 7 shows the SPLOM window populated with a new sample of 32,000 Wikipedia pages. Alongside this window is the Log window containing the rules generated from our session. Notice that the rules processor has highlighted points in the same relative regions of each scatterplot cell as in the SPLOM for the first dataset.

To check the validity of our rules, we looked at several highlighted points and examined the corresponding pages in the Text window. We found them to contain significant controversial sections. Figure 8 shows a typical page from this set. It is titled *The Illuminatus! Trilogy*, which is a series of novels that describe a drug-, sex- and magic-laden trek through a number of conspiracy

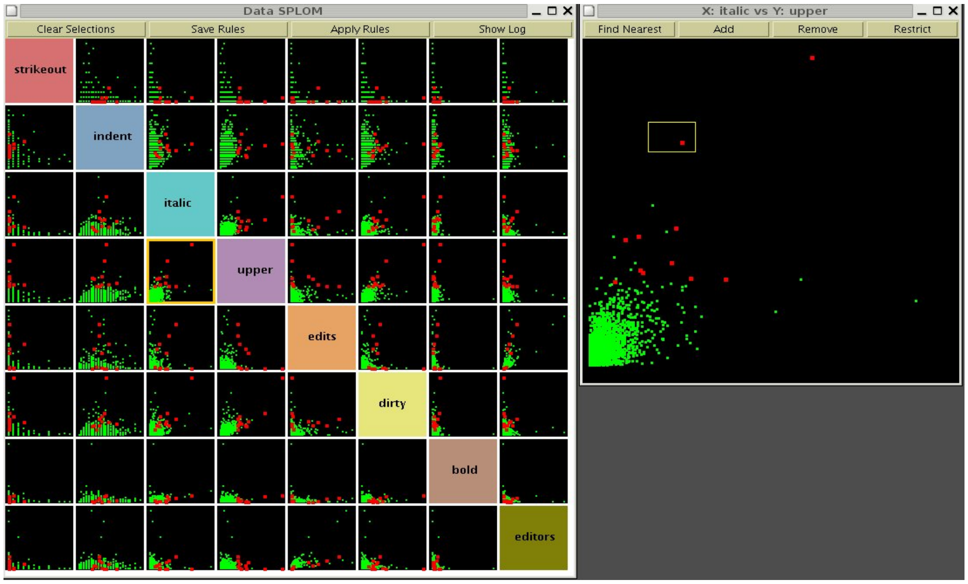


Figure 3: Wiki pages scatterplot view with near neighbor selection

theories, both historical and imaginary. Editors used strikeouts to express their dissatisfaction with the way things were phrased in the article.

5 Conclusion

We have presented the advantages of employing the L^∞ norm for visual data analysis through the user management of rectangular description regions. We conclude by mentioning a few alternative approaches. We have already noted that rectangular covers can be unparsimonious for describing non-rectangular shapes. One alternative is to give the user a 2D brush tool that can specify any connected region. MacSpin [10], for example, implemented a lasso to define a selection region in scatterplots. After the user forms such a region, we could automatically generate a set of rectangles as a cover. This approach breaks the link between the user specification and the actual cover, however. We believe users are familiar with rectangular selection regions and are comfortable composing them.

Another approach is to allow rotation before selection. The first dynamic graphics visualization program, Prim9 [12] added projection pursuit [13] and rotation controls to a selection tool (what they called a *mask*). In that paper, the authors showed a nonlinear discrimination problem that could be managed with a combination of rotation and linear masking. We could extend our method to allow rotation, although this would be problematic for data embedded in more than a few dimensions. Point rotation is not easy to manage in more than three dimensions. Projection pursuit and the Grand Tour [2] can be useful as guides, but the additional complexity has other drawbacks.

We have concentrated on introducing L^∞ visual analytics in this paper. There is more to do, however. Our next step will be to assess formally out-of-sample performance on a variety of artificial and real datasets in a supervised learning task. To evaluate our approach further, we plan to compare our method to traditional automated classification and prediction algorithms on real datasets. We anticipate that a trained visual analyst can rival traditional data mining methods – not necessarily in the training sample, but more likely in the validation sample. We expect that a well-trained analyst can spot exceptions, outliers, unusual distributions, and other anomalies that can thwart distance-based algorithmic models.

References

[1] R. Aman. http://sonic.net/maledicta/.

[2] D. Asimov. The grand tour: A tool for viewing multidimensional data,. *Siam Journal on Scientific and Statistical Computing*, 6:128–143, 1985.

[3] R. A. Becker and W. S. Cleveland. Brushing Scatterplots. *Technometrics*, 29:127–142, 1987.

[4] U. Brandes, D. Fleischer, and J. Lerner. Summarizing dynamic bipolar conflict structures. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1486–1499, 2006.

[5] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.

[6] C. A. Brewer. Guidelines for selecting colors for diverging schemes on maps. *The Cartographic Journal*, 33:79–86, 1996.

[7] D. B. Carr, R. J. Littlefield, W. L. Nicholson, and J. S. Littlefield. Scatterplot matrix techniques for large n. *Journal of the American Statistical Association*, 82:424–436, 1987.

[8] C. Chen, F. Ibekwe-Sanjuan, E. San Juan, and C. Weaver. Visual analysis of conflicting opinions. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*, Baltimore, October 2006.

[9] J. Donath, K. Karahalios, and F. B. Viégas. Visualizing conversation. In *HICSS '99: Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 2*, page 2023, Washington, DC, USA, 1999. IEEE Computer Society.

[10] A. W. Donoho, D. L. Donoho, and M. Gasko. Macspin: Dynamic graphics on a desktop computer. *IEEE Computer Graphics and Applications*, 8(4):51–58, 1988.

[11] S. Eick, J. Mauger, and A. Ratner. Visualizing the performance of computational linguistics algorithms. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*, Baltimore, October 2006.

[12] M. A. Fisherkeller, J. H. Friedman, and J. W. Tukey. Prim9: An interactive multidimensional data display and analysis system. In W. S. Cleveland and M. E. McGill, editors, *Dynamic Graphics for Statistics*. Wadsworth and Brooks/Cole, Pacific Grove, CA, 1988.

[13] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers*, C-23:881–890, 1974.

[14] B. J. Gao and M. Ester. Turning clusters into patterns: Rectangle-based discriminative data description. In *ICDM '06: Proceedings of the Sixth International Conference on Data Mining*, pages 200–211,

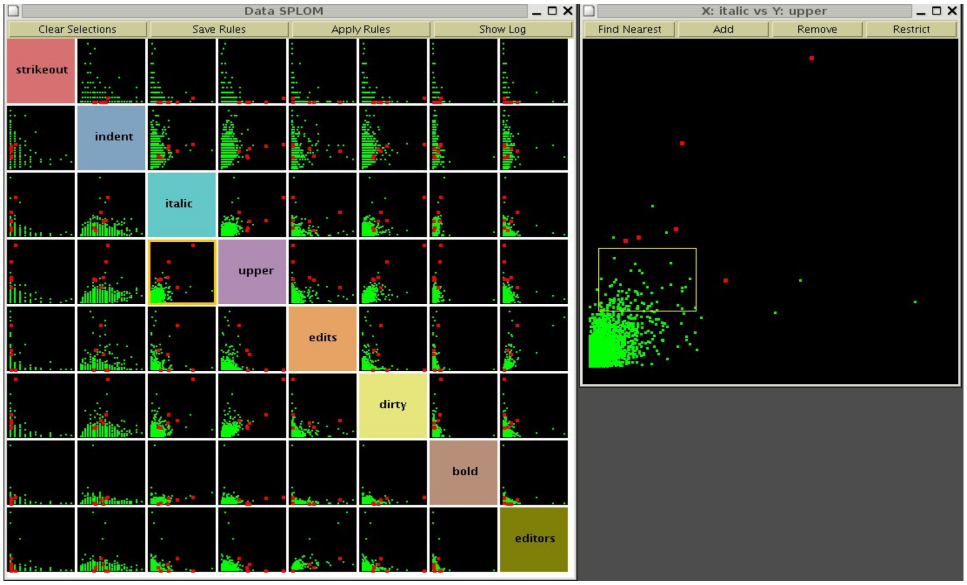


Figure 4: Wiki pages scatterplot view with remove rectangle

Washington, DC, USA, 2006. IEEE Computer Society.

[15] N. Henry and J. D. Fekete. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, 2006.

[16] R. J. Quinlan. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, 1993.

[17] M. A. Smith and A. T. Fiore. Visualization components for persistent conversations. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 136–143, New York, NY, USA, 2001. ACM Press.

[18] E. Spertus. Smokey: Automatic recognition of hostile messages. In *AAAI/IAAI*, pages 1058–1065, 1997.

[19] B. Stvilia, M. B. Twidale, L. C. Smith, and L. Gasser. Assessing information quality of a community-based encyclopedia. In *Proceedings of the International Conference on Information Quality - ICIQ 2005*, pages 442–454, Cambridge, MA, 2005.

[20] James J. Thomas and Kristin A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, chapter 2. August 2005.

[21] F. B. Viégas and M. Smith. Newsgroup crowds and authorlines: Visualizing the activity of individuals in conversational cyberspaces. In *HICSS '04: Proceedings of the Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04) - Track 4*, page 40109.2, Washington, DC, USA, 2004. IEEE Computer Society.

[22] F. B. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations, 2004.

[23] A. Wilhelm. User interaction at various levels of data displays. *Computational Statistics and Data Analysis*, 43(4):471–494, 2003.

[24] B. Yu. Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. *The Annals of Statistics*, 21:711–735, 1993.

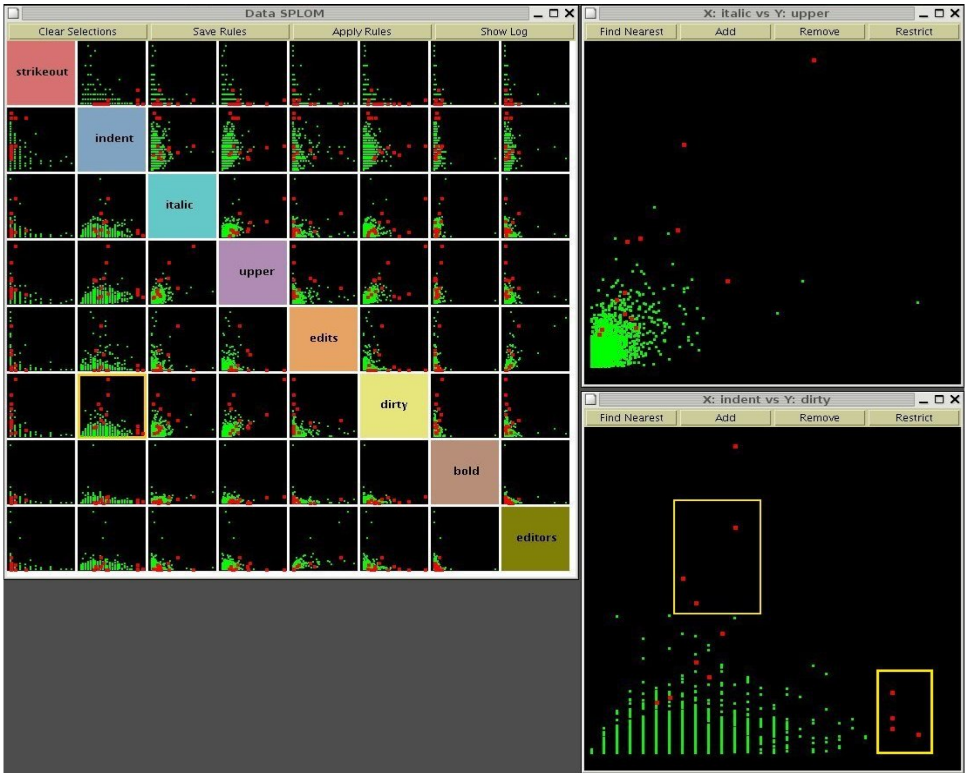
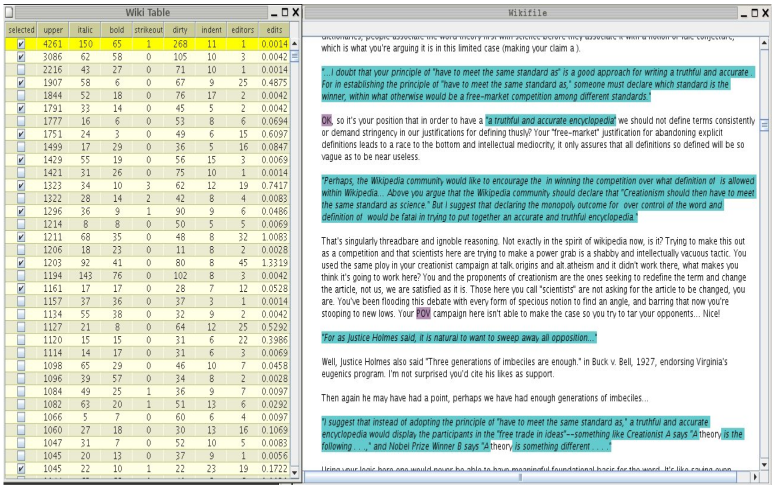


Figure 5: Add operation over different SPLOM cells



selected	upper	italic	bold	strikeout	dirty	indent	editors	edits
☑	4261	150	65	1	268	11	1	0.0014
☑	3086	62	58	0	105	10	3	0.0042
☐	2216	43	27	0	71	10	1	0.0014
☑	1907	58	6	0	67	9	25	0.4875
☐	1844	52	18	0	76	17	2	0.0042
☑	1791	33	14	0	45	5	2	0.0042
☐	1777	16	6	0	53	8	6	0.0694
☑	1751	24	3	0	49	6	15	0.6097
☐	1499	17	29	0	36	5	16	0.0847
☑	1429	55	19	0	56	15	3	0.0069
☐	1421	31	26	0	75	10	1	0.0014
☑	1323	34	10	3	62	12	19	0.7417
☐	1322	28	14	2	42	8	4	0.0083
☑	1296	36	9	1	90	9	6	0.0486
☐	1214	8	8	0	50	5	5	0.0069
☑	1211	68	35	0	48	8	32	1.0083
☐	1206	18	23	0	11	8	2	0.0028
☑	1203	92	41	0	80	8	45	1.3319
☐	1194	143	76	0	102	8	3	0.0042
☑	1161	17	17	0	28	7	12	0.0528
☐	1157	37	36	0	37	3	1	0.0014
☐	1134	55	38	0	32	9	2	0.0042
☐	1127	21	8	0	64	12	25	0.5292
☐	1120	15	15	0	31	6	22	0.3986
☐	1114	14	17	0	31	6	3	0.0069
☐	1098	65	29	0	46	10	7	0.0458
☐	1096	39	57	0	34	8	2	0.0028
☐	1084	49	25	1	36	9	7	0.0097
☐	1082	63	20	1	51	13	6	0.0292
☐	1066	5	7	0	60	6	4	0.0097
☐	1060	27	18	0	30	13	16	0.1069
☐	1047	31	7	0	52	10	5	0.0083
☐	1045	20	13	0	37	9	1	0.0056
☑	1045	22	10	1	22	23	19	0.1722

dictionaries, people associate the word theory first with science before they associate it with a notion of idle conjecture, which is what you're arguing it is in this limited case (making your claim a).

"...I doubt that your principle of "have to meet the same standard as" is a good approach for writing a truthful and accurate . For in establishing the principle of "have to meet the same standard as," someone must declare which standard is the winner, within what otherwise would be a free-market competition among different standards."

OK, so it's your position that in order to have a *"a truthful and accurate encyclopedia"* we should not define terms consistently or demand stringency in our justifications for defining thusly? Your "free-market" justification for abandoning explicit definitions leads to a race to the bottom and intellectual mediocrity; it only assures that all definitions so defined will be so vague as to be near useless.

"Perhaps, the Wikipedia community would like to encourage the in winning the competition over what definition of is allowed within Wikipedia... Above you argue that the Wikipedia community should declare that "Creationism should then have to meet the same standard as science." But I suggest that declaring the monopoly outcome for over control of the word and definition of would be fatal in trying to put together an accurate and truthful encyclopedia."

That's singularly threadbare and ignoble reasoning. Not exactly in the spirit of wikipedia now, is it? Trying to make this out as a competition and that scientists here are trying to make a power grab is a shabby and intellectually vacuous tactic. You used the same ploy in your creationist campaign at talk.origins and alt.atheism and it didn't work there, what makes you think it's going to work here? You and the proponents of creationism are the ones seeking to redefine the term and change the article, not us, we are satisfied as it is. Those here you call "scientists" are not asking for the article to be changed, you are. You've been flooding this debate with every form of specious notion to find an angle, and barring that now you're stooping to new lows. Your POV campaign here isn't able to make the case so you try to tar your opponents... Nice!

"For as Justice Holmes said, it is natural to want to sweep away all opposition..."

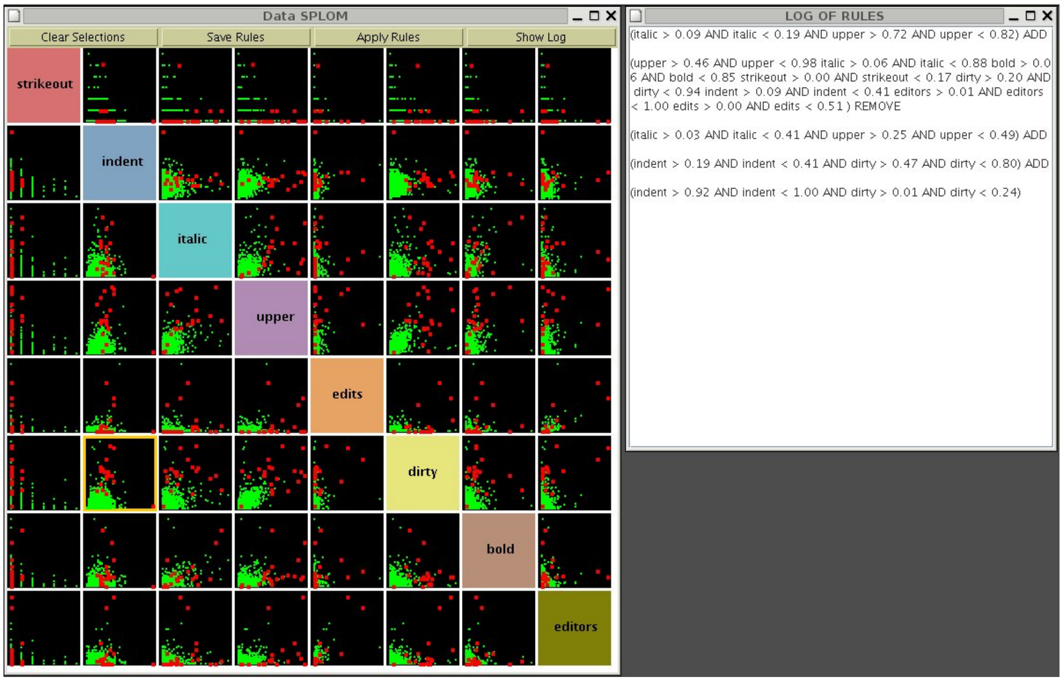
Well, Justice Holmes also said "Three generations of imbeciles are enough." in Buck v. Bell, 1927, endorsing Virginia's eugenics program. I'm not surprised you'd cite his likes as support.

Then again he may have had a point, perhaps we have had enough generations of imbeciles...

"I suggest that instead of adopting the principle of "have to meet the same standard as," a truthful and accurate encyclopedia would display the participants in the "free trade in ideas"--something like Creationist A says "A theory *is the following . . .," and Nobel Prize Winner B says "A* theory *is something different . . . ."*

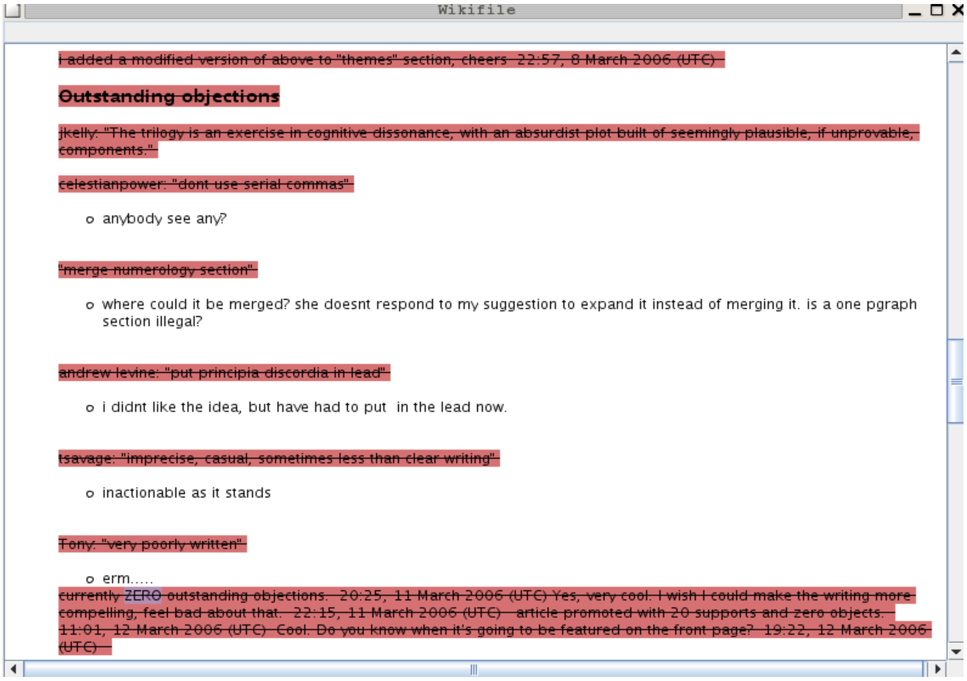
Using your logic here one would never be able to have meaningful foundational basis for the word. It's like saying even

Figure 6: Table view and selected controversial Wiki page



```
(italic > 0.09 AND italic < 0.19 AND upper > 0.72 AND upper < 0.82) ADD

(upper > 0.46 AND upper < 0.98 italic > 0.06 AND italic < 0.88 bold > 0.0
6 AND bold < 0.85 strikeout > 0.00 AND strikeout < 0.17 dirty > 0.20 AND
 dirty < 0.94 indent > 0.09 AND indent < 0.41 editors > 0.01 AND editors
< 1.00 edits > 0.00 AND edits < 0.51 ) REMOVE

(italic > 0.03 AND italic < 0.41 AND upper > 0.25 AND upper < 0.49) ADD

(indent > 0.19 AND indent < 0.41 AND dirty > 0.47 AND dirty < 0.80) ADD

(indent > 0.92 AND indent < 1.00 AND dirty > 0.01 AND dirty < 0.24)
```

Figure 7: SPLOM populated with new dataset and log of rules determining highlighted points in SPLOM



~~i added a modified version of above to "themes" section, cheers 22:57, 8 March 2006 (UTC)~~

~~**Outstanding objections**~~

~~jkelly: "The trilogy is an exercise in cognitive dissonance, with an absurdist plot built of seemingly plausible, if unprovable, components."~~

~~celestianpower: "dont use serial commas"~~

- anybody see any?

~~"merge numerology section"~~

- where could it be merged? she doesnt respond to my suggestion to expand it instead of merging it. is a one pgraph section illegal?

~~andrew levine: "put principia discordia in lead"~~

- i didnt like the idea, but have had to put in the lead now.

~~tsavage: "imprecise, casual, sometimes less than clear writing"~~

- inactionable as it stands

~~Tony: "very poorly written"~~

- erm.....

~~currently ZERO outstanding objections. 20:25, 11 March 2006 (UTC) Yes, very cool. I wish I could make the writing more compelling, feel bad about that. 22:15, 11 March 2006 (UTC) article promoted with 20 supports and zero objects. 11:01, 12 March 2006 (UTC) Cool. Do you know when it's going to be featured on the front page? 19:22, 12 March 2006 (UTC)~~

Figure 8: Controversial page found by applying saved rules to new dataset