

Lightweight Provenance-Driven Exploration

Aran Lunzer*

Viewpoints Research Institute, Los Angeles

ABSTRACT

I'm interested in how interactive tools can encourage users to go beyond a single default visualisation for a data set or simulation, exploring and comparing alternatives in the hope of gaining a deeper understanding. Here I present two prototype interfaces that tackle this by reifying provenance: one gives the user explicit, instant access for revisiting parameter values that have been tried over the course of an exploration; the other ties provenance values to the visualisation marks themselves, to guide the user towards settings that lead to interesting alternative results. My hope is that the principles of these interfaces can be deployed more broadly, on general sensemaking tasks.

Keywords: Exploratory data analysis, comparisons, interaction history, subjunctive interfaces.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical User Interfaces (GUI);

1 INTRODUCTION

In scientific discovery, exploring parameter spaces has been recognised as a crucial task for which provenance support can help [1, 9]. Parameter spaces also feature strongly in my long-running research on subjunctive interfaces [5]: interfaces that make it easy for users to explore and compare alternatives, by setting them up as multiple parallel scenarios. Here I describe two prototype interfaces that have arisen from that research, which—although designed to support only simple visualisations, rather than general sensemaking—I believe could offer some useful directions for development of highly interactive sensemaking tools.

2 AN INTERACTIVE HISTORY OF PARAMETER SETTINGS

In recent work (reported at this year's InfoVis [6]) we have been investigating how to support users in exploring the impact of parameter settings on simple statistical charts. The parameters are controlled with direct-manipulation widgets such as buttons and sliders, allowing dynamic adjustment with immediate update of the chart, and all the adjustments—including intermediate values during a slider drag—are recorded in a visual 'history list'. The list is thus an accumulating record of a chart's provenance.

One of the facilities offered by the history list is what we call History Rewind: random-access revisiting of any point in the recorded sequence of interactions. An example is shown in Figure 1. During rewind, pointing to a value within a history item causes all chart parameters to revert to the state they had at the time when the user tried that value. A consequence is that if the user remembers seeing an interesting result but isn't sure what parameter combination caused it, it will always be possible to relocate that result by hunting through the history.

A second facility is History Override. Instead of rewinding history to an earlier state, this is an ephemeral modification of the latest state of the chart. In this operation, pointing to a history-item value overrides the final value that had been selected for that parameter,

*e-mail: aran@acm.org

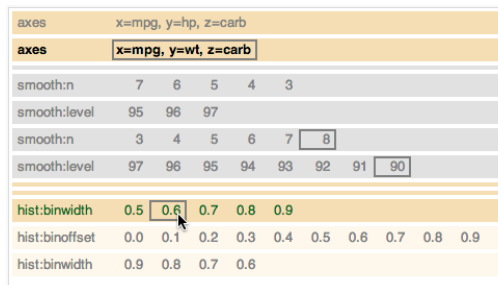


Figure 1: A history list is a record of the user's parameter adjustments. A single item represents one continuous manipulation of one parameter, possibly over many values. Items for parameters that are no longer contributing to the chart (in this case, the settings for the smoothed regression line, which has been switched off) are greyed out. Here a History Rewind is in progress, and the user is revisiting the first time the histogram bin width was set to 0.6. The chart—not shown here—is restored to the state it had at that time. Adjustments that are 'in the future' with respect to the probed time are ignored, and their appearance is faded.

providing a quick way to test a counterfactual: 'How would the chart look now if I'd released the bin-offset slider on 0.2 instead?'. Having the provenance record thus makes it easy to revisit parameters whose values were decided early in the exploration, but that may deserve another look given the subsequent choices made for other parameters.

3 GUIDING USERS TO OTHER NEARBY RESULTS

An interactive history supports exploration, but leaves the user to take the initiative in varying input-parameter settings. We have also investigated a style of interface that seeks to encourage exploration by guiding the user towards additional results, available through the use of alternative settings. For now, our examples assume that these settings and their results have been derived in advance, based on alternative parameter values in which the user has expressed an interest; an extension of the technique would be to derive potentially interesting alternatives automatically.

Our interfaces reveal information in a series of steps. When a user first indicates interest (e.g., with the mouse pointer) in some region of a visualisation for which alternative results are available, we display **hints** of the existence and nature of those alternatives. A user who follows up by interacting with a hint marker is then shown **hallmarks**—distinguishing features of the alternative scenario(s) represented by the hint. Interacting with a hallmark then leads to a **hijack**, in which the represented scenario takes over the display, with highlighting on all the values that have changed relative to the base scenario.

Figure 2 shows a chart—a line plot—derived from a spreadsheet-like calculation model driven by the values in seven parameter-setting cells. In this case the hints are faintly drawn alternative plot lines, and the hallmarks appear as colour-tagged values next to the changed parameter cells.

savings					
	A	B	C	D	E
1	inflation %/yr	3.00	return %/yr	6.00	
2	initial capital	100000.00			
3	retire after yr	57.00	income/yr	48000.00	
4					
year	inflation adj.	adj. saving/yr	adj. income/y	capital	
6	40.00	1.00	24000.00	27000	0.00
7	41.00	1.03	24720.00	30000	0.00
8	42.00	1.06	25461.60		0.00
9	43.00	1.09	26225.45		0.00
10	44.00	1.13	27012.21		0.00
11	45.00	1.16	27822.58		0.00
12	46.00	1.19	28657.26		0.00
13	47.00	1.23	29516.97		0.00
14	48.00	1.27	30402.48		0.00
15	49.00	1.30	31314.56		0.00
16	50.00	1.34	32253.99		0.00
17	51.00	1.38	33221.61		0.00
18	52.00	1.43	34218.26		0.00
19	53.00	1.47	35244.81		0.00
20	54.00	1.51	36302.15		0.00
21	55.00	1.56	37391.22		0.00
22	56.00	1.60	38512.95		0.00
23	57.00	1.65	39668.34		0.00
24	58.00	1.70	0.00	81716.79	1197659.53
25	59.00	1.75	0.00	84168.29	1185350.81
26	60.00	1.81	0.00	86693.34	1169778.52
27	61.00	1.86	0.00	89294.14	1150671.09
28	62.00	1.92	0.00	91972.96	1127738.39
29	63.00	1.97	0.00	94732.15	1100670.54
30	64.00	2.03	0.00	97574.12	1069136.66
31	65.00	2.09	0.00	100500.00	1033300.00

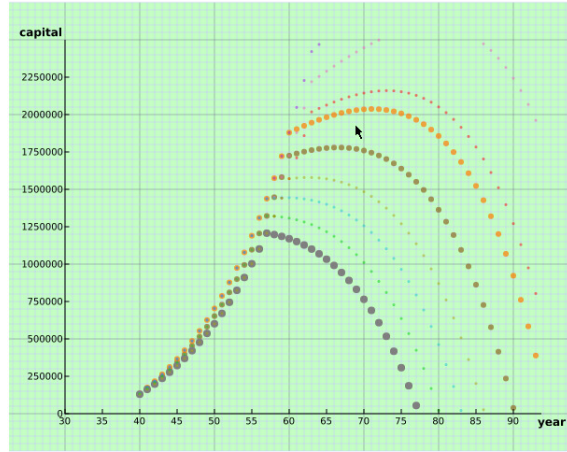


Figure 2: Hints and hallmarks, in a simple model of saving and spending for retirement. The cells with blue text are parameters to the model; the grey dotted plot line shows the result for the current settings. When the user mouses over the plot region, hints appear in the form of faint alternative plot lines. Here the user has moved close to two of those lines; the lines are given more prominence, and scenario hallmarks are shown in the form of colour-tagged values on the cells responsible for the lines' divergence from the default scenario. Mousing directly over a line would cause its values to take over (hijack) the whole table.

Liveability							
	A	B	C	D	E	F	
1	name	stability	healthcare	culture	education	infrastructure	
2	(weights)	25.00	20.00	25.00	10.00	20.00	
3	Melbourne	95.00	100.00	95.14	100.00	100.00	
4	Vienna			94.44	100.00	100.00	
5	Vancouver			94.44	100.00	100.00	
6	Toronto			94.44	100.00	100.00	
7	Adelaide			95.00	100.00	100.00	
8	Calgary			100.00	05	100.00	
9	Sydney			90.00	66	100.00	
10	Helsinki			100.00	99	100.00	
11	Perth			95.00		100.00	
12	Auckland			95.00		95.00	

Figure 3: From hints to a hijack, on alternative ways of weighting city characteristics to derive a ranking. The hints (markers containing orange and/or red blobs) show that every city seen in the base scenario appears in different places in at least some of the alternative scenarios. The first inset shows how mousing over the hint in cell A3 brings up a hallmarks display, which distinguishes the scenarios by showing (as a list) what values other than Melbourne could appear in this cell, and (using the orange and black markers) where else Melbourne could appear. Finally, mousing over the Toronto list entry hijacks the sheet with the scenario in which A3 is occupied by Toronto: cell values that are different from the base scenario are coloured orange.

Figure 3 shows a progression from hints to hijack in a sheet that ranks world cities according to weighted scores on various criteria¹. In this example, the hints are markers that are added to a cell to show when some scenarios lead to different values appearing in that cell. A cell's hallmarks answer 'What else?' and 'Where else?' questions [7]: what other values can appear in this cell, and where else can the cell's default-scenario value appear?

4 RESEARCH DIRECTIONS

There are many directions for extending the work mentioned here. To extend the interface to deal with a broader range of data and visualisation types, we are looking at general issues of visual comparison (see [3] for a survey), while generalising to richer patterns

¹The data here are from the Economist Intelligence Unit's Global Liveability Report (http://www.eiu.com/site_info.asp?info_name=The_Global_Liveability_Report).

of exploration will benefit from recent work on undo/redo models, such as [4]. A useful direction regarding the balance of initiative between user and system in choosing relevant scenarios is visual steering (e.g., [8]). Finally, even if a system makes it relatively easy for users to explore, there is still a question of whether they will expend the effort needed to do so; work on persuasive design [2] suggests ways of increasing the chance that they will.

REFERENCES

- [1] S. P. Callahan, J. Freire, E. Santos, C. E. Scheidegger, C. T. Silva, and H. T. Vo. Using provenance to streamline data exploration through visualization. Technical Report UUSCI-2006-016, SCI Institute, Univ. of Utah, 2006.
- [2] B. Fogg. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, pages 40:1–40:7, New York, NY, USA, 2009. ACM.
- [3] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, Oct. 2011.
- [4] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. Graphical histories for visualization: Supporting analysis, communication, and evaluation. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1189–1196, Nov 2008.
- [5] A. Lunzer and K. Hornbæk. Subjunctive interfaces: Extending applications to support parallel setup, viewing and control of alternative scenarios. *ACM Trans. Comput.-Hum. Interact.*, 14(4):1–44, 2008.
- [6] A. Lunzer and A. McNamara. It ain't necessarily so: Checking charts for robustness (poster). In *Proceedings of IEEE InfoVis 2014*. IEEE, 2014.
- [7] A. Lunzer and Y. Ohshima. What else and where else: two worthwhile questions for an information interface. In *Proc. 7th Nordic Conf. on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 815–816, New York, NY, USA, 2012. ACM.
- [8] K. Matkovic, D. Gracanin, M. Jelovic, and H. Hauser. Interactive visual steering - rapid visual prototyping of a common rail injection system. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1699–1706, Nov 2008.
- [9] C. Silva, J. Freire, E. Santos, and E. Anderson. Provenance-enabled data exploration and visualization with VisTrails. In *Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2010 23rd SIBGRAPI Conference on*, pages 1–9, Aug 2010.