
Accentuating Visualization Parameters to Guide Exploration

Marian Dörk
Culture Lab
Newcastle University
Newcastle upon Tyne, UK
marian.doerk@ncl.ac.uk

Heidi Lam
Google
Mountain View, CA, USA
hllam@google.com

Omar Benjelloun
Google
New York, NY, USA
benjello@google.com

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2013 Extended Abstracts, April 27–May 2, 2013, Paris, France.

Copyright 2013 ACM 978-1-4503-1952-2/13/04...\$15.00.

Abstract

We present a new method for displaying visualization parameters to guide casual data exploration. When visualizing datasets with large parameter spaces it can be difficult to move between data views. Building on social navigation and degree-of-interest visualization, we propose the concept of accentuation as the selection and emphasis of visualization parameters based on social and semantic signals. We describe how we designed an accentuated visualization interface, and discuss open challenges and directions for future research.

Author Keywords

Information visualization; exploration; social navigation.

ACM Classification Keywords

H.5.2 [Information Interfaces and Presentation]: Misc.

Introduction

Navigating in real-life datasets can be frustrating due to the large number of parameters. For example, public data is an interesting domain where datasets typically come with long lists of indicators and countries that can be overwhelming to grasp and explore. Exploring public data in a visualization application may be frustrating due to the sparsity of interesting views and also missing data points across all parameter options. While visualization research has typically focused on providing optimal displays for a

given data view, the issue of navigating between visualization states has received relatively little attention. The goal of this work is to fill this gap: to make visualization of datasets with many parameters more accessible and inviting for casual exploration, i.e., an open-ended, leisurely activity not part of a specific task.

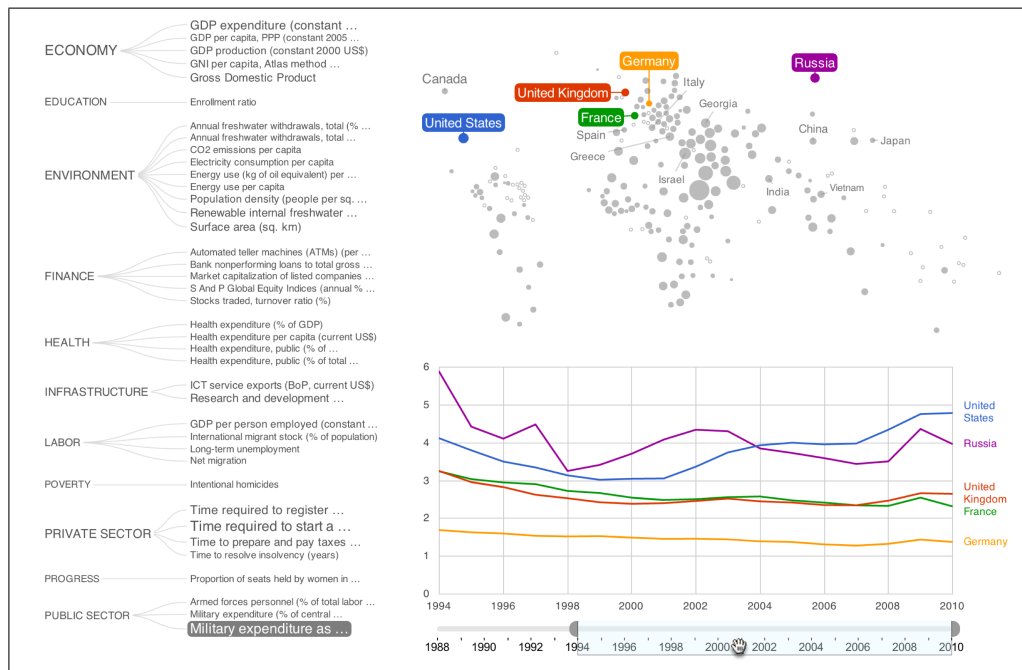


Figure 1: Accentuated visualization interface for exploring temporal data along indicators (left) and countries (top right).

To support casual visualization use, we propose using social and semantic signals to guide data explorers. With 'social' signals we refer to aggregate use statistics of the visualization or any activity outside of the visualization that could inform its use. 'Semantic' signals refer to

interest indications derived directly from the data. There has been significant research interest in using social signals to guide people's use of information structures such as documents and web pages [1, 3, 7, 9]. However, only a few attempts have been made to employ such signals to guide the use of information visualizations [6, 10]. So far the focus was either on explicit collaboration or on interfaces with few parameter options. Yet, especially when interacting with datasets that have comprehensive parameter spaces—typically shown as long lists of options—there is a need to support the viewer to find those options that lead to interesting visualization states.

We introduce the concept of *accentuation* as the system's selection and emphasis of visualization parameters based on social and semantic signals in relation to the current parameter settings. As shown in Figure 1, we present a reference design of a data exploration interface that uses accentuated displays of indicators and countries to provide access to thousands of time series of public statistics. We end with a discussion of open questions for future work.

Related work

There has been a large body of work applying social signals to guide navigation by emphasizing spatial, social, or semantic relationships in the content using layout [4]. Dynamic layouts based on social signals can be beneficial when the layout reflects actual information practices [1], provides social awareness, and relevant views [3, 7, 9]. However, spatial reconfiguration cannot provide a stable reference point to anchor navigation. While there are a few attempts to apply social signals in information visualizations, the work either assumes a small parameter space or emphasizes social collaboration. For example, interface widgets can be enhanced by minimalistic visualizations to guide a person towards more popular

interface options [10]. Furthermore, collaborative data analysis can be supported by displaying comments and annotations within and besides visualizations [6]. In these examples, the social signals are not explicitly used to select and emphasize visualization parameters.

In addition to spatial arrangement and visual highlights, another approach to emphasize information elements is based on the *degree-of-interest* (DOI) as established by a person's interaction with an information structure. DOI has been applied to trees [2], where tree segments that are far from a given selection are hidden to make room for the more 'interesting' elements. Instead of hiding less related elements, it is also possible to gradually decrease the font sizes for elements further away from the active element [8]. While hiding and shrinking can be useful to focus on a subset of data, the former lacks visual differentiation among the shown elements and the latter leads to poor readability of many elements. To convey differences in DOI and maintain readability, we use a combination of parameter selection and sizing.

The concept of accentuation

Accentuation takes inspiration from the degree-of-interest approach and the use of social and semantic signals as a means to highlight interesting visualization parameters. In contrast to previous efforts, our design can support casual exploration of comprehensive datasets while preserving the overall structure of the parameter space. Based on aggregate and anonymized usage data of the Google Public Data interface [5], which includes dwelling time and parameter changes, we noticed that some visitors must have had difficulties to find interesting parameter combinations and navigate between different views. To address these problems we suggest a shift towards casual data exploration with two concrete design goals:

1. *Reduce complexity.* Offer an inviting overview of the parameters by using lightweight visualizations that represent the overall structure of the parameter space, i.e., the topical nesting of indicators and spatial distribution of countries.
2. *Guide exploration.* Highlight parameters that lead to interesting views by promoting options with greater social and semantic signals and demoting dead-ends due to missing data. With changing visualization states, gradually change the parameter displays while keeping their overall structure stable.

To address these goals, we propose accentuation as a method for selecting and emphasizing visualization parameters based on diverse signals and the explorer's parameter settings. Accentuation shares the spirit of recommendations à la "customers who bought this also bought that" and adopts it for visualization. To accentuate visualization parameters, social and semantic signals need to be sourced, correlated, and displayed.

Besides social signals such as prior use, we consider other types of signal such as news stories and data availability to highlight views that are timely and backed with data. Because visualization states are typically defined by multiple parameter options, correlation computations need to consider combinations of parameter options. In contrast to most recommendation-based interfaces that display related items with little or no indication of correlation strengths, we wish to expose the differences in relevance by adjusting font sizes of parameter labels based on their DOI. Next we detail how we realize these steps towards accentuation in a prototype interface that provides access to time-series visualizations of public data.

An accentuated interface for public data

We use public data as a use case to create a navigational visualization interface for comprehensive datasets. We focus on two parameter types: indicators and countries. Indicators are numeric metrics that quantify the economic, social, and environmental condition of a country, such as population, unemployment rate, or energy use. International organizations that collect diverse statistics about countries have data for hundreds of indicators. For example, the World Bank’s World Development Indicators that we use for our prototype have 485 indicators categorized by 11 topics, such as economy, education, and environment. Most public data visualizations display indicators and countries as alphabetical lists.

Deriving Signals

While many signals can be considered, we wished to emphasize parameter options leading to relevant views based on their popularity (usage logs), timeliness (news articles), and availability (data points).

Usage logs are the signal source employed most for social navigation [3]. For visualization, a signal based on prior use can be described as: “viewers who selected this parameter, also selected those.” We used six months of anonymized usage logs from Google’s Public Data Explorer [5] that includes two basic parameters used in public datasets: indicators and countries. Occurrence of pairwise parameter combinations in scatterplots serves as evidence for viewers’ interests in both of them.

News articles that mention entities corresponding to visualization parameters are the second source for social signals. For the case of public data, using news activity as a signal is especially promising as countries and indicators are regular topics of news stories. Given a parameter option, a simplified reading of this signal is: “news articles

related to this parameter, also mentioned those parameters”. In order to get the correlated activity of parameter options in news stories we take the number of results on Google News about combinations of parameters (e.g., “France GDP”) over the last month.

Data availability can be derived directly from the visualized data. Considering the number of data points can help avoiding views lacking sufficient data. Given a particular indicator, the assumed meaning of this signal is the following: “data series for this indicator are more comprehensive for these countries”—or vice versa for indicators of a specific country. To retrieve this signal we count the data points for a given country and indicator.

Computing Relevance

With the signals from the above three sources, we can compute the relevance of a parameter option for a given parameter setting. To limit the complexity of the relevance computation, we first consider pairwise signals between parameters, i.e., signals for these combinations: country-indicator, indicator-country, country-country, and indicator-indicator. The formula is based on tf-idf, a common measure used in information retrieval to determine how relevant a result is for a given search term. The selected parameter A can be viewed as a search term and the remaining parameter options B_i as the search results. Given a selected parameter A , the relevance of B_i is then determined by this formula:

$$rel(A, B_i) = \frac{|A \cap B_i|}{|B_i|} \times \log \frac{N}{|A|}$$

With, $|A|$ as the occurrences of a given parameter, $|B_i|$ as the occurrences of another parameter option, $|A \cap B_i|$ as the joined occurrences of both parameters, and N as the total occurrences of all parameter options.

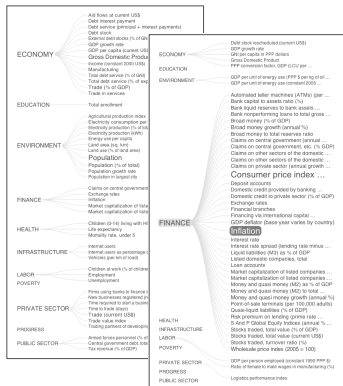


Figure 2: The visualization of indicators conveys their topical grouping and relevance. Without a parameter selected, indicators are displayed on the basis of their total frequency in the signals (left). When at least one parameter is selected, the DOI values are used to select and size indicators; selecting also a topic, e.g., 'Finance', will focus on the contained indicators (right).

Following the notion of tf-idf, the resulting value represents how relevant a parameter B is for the given parameter A according to their joined and separate occurrences in relation to the sum of occurrences of all parameters. To generate DOI values for visualization states with multiple parameters, we sum up the respective relevance scores and normalize them in order to have equal influence of country and indicator parameters.

Accentuating Parameters

The interface features three coordinated visualizations showing indicators, countries, and time series (see Figure 1). The DOI values are used in the interface to select and size indicators and countries to emphasize parameter options that are likely to lead to interesting data views. In addition, we use metadata for the topical nesting of indicators and positioning of countries, and actual data values for sizing country circles.

Indicators are visualized in a two-level hierarchy with subtle curves connecting topics and indicators (see Figure 2). Topics and indicators are consistently ordered alphabetically, whereas the indicators are broken up by their topics. The visualization allocates space for the topics and indicators to reflect the accentuation as defined by the DOI values. Depending on the number of shown indicators for a given topic, its label size represents the aggregated DOI values of the contained indicators. While the interface displays parameters that have an increased likelihood of relevance, there is interactive access to all indicators either by selectively focusing on individual topics or by full-text search.

For the display of the countries, our goal is to provide interactive access to all countries, while using DOI values for accentuating via the display of labels and public data values for the display of the circles. Countries are

displayed on a blank map representing each country as a circle placed by its geographic centre point. The circle sizes represent the currently selected indicator value. Prioritizing interactive access over precise location, an iterative algorithm avoids circle overlaps, which results in a subtle stretching of regions according to their indicator values of these countries. Country labels are selected and sized according to the DOI values in correspondence with the current parameter setting. The labels are placed so that they do not overlap circles or each other.

The *time-series* visualization is a conventional line-chart that represents the temporal development of an indicator for the selected countries (see Figure 1). To match countries in the map and their time lines, the colours are shared between these views.

Discussion

Our prototype is an initial attempt to realize accentuation. During our own uses of the interface we encountered some limitations that raise questions for future research.

Convey pruning. To reduce parameter complexity, we prune them and show a manageable number of items. While there is interactive access to the remaining elements via search and mouse interaction, one may assume that the displayed items are the only ones that are available. We will investigate this potential usability problem.

Explain accentuation. While the interface offers guidance to interesting data views by selecting and emphasizing parameter items, data explorers may not understand the system's accentuation. One attempt of making the system more transparent is a legend that details the signals being used, namely news activity, prior use, and data availability. However, it remains an open question if this information is sufficient for the explorers to understand the accentuation.

Context for signals. One could argue that the interface should provide more detailed context to interpret accentuated data views. For example, if a topic depicted in a parameter is mentioned frequently in recent news articles, the interface may need to alert the explorers of the news items for them to adequately interpret the data. Similarly in the case of the usage signal, one may require more details to understand why a view may be interesting.

Choice of signals. The initial signals we used were in part linked to the type of data being visualized. It is an open question what type of signals are more or less useful for encouraging casual data exploration. It would be interesting to study what type of signals are more beneficial for certain types of explorations. One could let people select the signals used as the basis for accentuation to study which they would prefer for which purposes.

Support storytelling. Our system could be extended to let explorers save visualization states as waypoints, annotate them, and share them among fellow explorers. Allowing explorers to tell stories found in the data turns them from passive data consumers to active data curators, which may in turn increase their levels of engagement in data exploration. Furthermore, data stories and waypoints can generate powerful social signal that can further be used.

In this paper, we introduced accentuation as a new method for displaying visualization parameters in order to guide casual data exploration. Illustrating the potential of accentuation, our prototype of a visual interface for public data uses news activity, usage logs, and data availability for selecting and emphasizing countries and indicators.

Acknowledgements

We thank our colleagues at Google's Public Data and Big Picture teams for their help and feedback, in particular Henrik Lindahl, Martin Wattenberg, and Fernanda Viegas.

References

- [1] Brusilovsky, P. Social information access: the other side of the social web. *SOFSEM 2008: Theory and Practice of Computer Science* (2008), 5–22.
- [2] Card, S., and Nation, D. Degree-of-interest trees: A component of an attention-reactive user interface. In *AVI 2002: Advanced Visual Interfaces*, ACM (2002), 231–245.
- [3] Dieberger, A., Dourish, P., Höök, K., Resnick, P., and Wexelblat, A. Social navigation: techniques for building more usable systems. *interactions* 7, 6 (2000), 36–45.
- [4] Dourish, P., and Chalmers, M. Running out of space: models of information navigation. In *Short paper presented at HCI*, vol. 94 (1994), 23–26.
- [5] Google Public Data Explorer, 2012. <http://www.google.com/publicdata>.
- [6] Heer, J., Viégas, F. B., and Wattenberg, M. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *CHI '07: Proc. of SIGCHI Conference on Human Factors in Computing Systems*, ACM (2007), 1029–1038.
- [7] Hill, W., Hollan, J., Wroblewski, D., and McCandless, T. Edit wear and read wear. In *CHI '92*, ACM (1992), 3–9.
- [8] Tominski, C., Abello, J., van Ham, F., and Schumann, H. Fisheye tree views and lenses for graph visualization. In *InfoVis 2006: Symposium on Information Visualization*, IEEE (2006), 17–24.
- [9] Wexelblat, A., and Maes, P. Footprints: history-rich tools for information foraging. In *CHI '99*, ACM (1999), 270–277.
- [10] Willett, W., Heer, J., and Agrawala, M. Scented widgets: Improving navigation cues with embedded visualizations. *TVCG: Trans. on Visualization and Computer Graphics* 13, 6 (2007), 1129–1136.