Applying a Sunburst Visualization to Summarize User Navigation Sequences

Kerry Rodden • Google

How are users navigating around our website? Many businesses ask this, but providing a succinct answer remains difficult. Typically, the necessary data is relatively easy to collect and aggregate from standard usage logs, but there's no straightforward way to analyze and visualize it.

Funnel diagrams can summarize user progress toward a single important goal on the site (such as making a purchase). However, analysts can consider only one path at a time, rather than getting an overview of all possible paths. Also, many sites and Web applications have important business goals that relate to overall usage patterns (for example, user engagement with site content) rather than progress toward a single outcome.

For example, at YouTube (which is owned by Google), we're interested in maximizing user engagement with videos. So, we need to understand how users are discovering videos to watch and how this discovery process evolves throughout a visit to the site. E-commerce site owners might have the same questions about their products, news site editors might want to understand this about their articles, and so on.

To aid with this analysis, I applied a sunburst visualization, after simplifying the aggregated navigation data into a hierarchical format (see Figure 1). The resulting visualization has been successful in YouTube and is frequently referenced and accessed.

Simplifying Navigation Data

In working with site navigation data for YouTube, I used my knowledge of the data (and my goals in visualizing it) to make appropriate simplifications, producing a more tractable visualization design. Because the YouTube data is confidential, I'll use the running example of a fictional e-commerce site to illustrate the process.

Categorizing Page Types

First, I categorized the many possible URLs into a small set of page types, after filtering out those that didn't correspond to a user-initiated page transition. For example, a search results page can have a range of URL parameters and possible content (corresponding to different search queries), but all such pages are fundamentally the same type. Other types could include the site's home page, a product details page, and a user account page. Using a small set of types minimizes the number of possible permutations of those types in navigation sequences. I iterated on this set of page types as I created the visualization and gained a better understanding of the underlying data—for example, collapsing categories that were mostly too small to see.

Focusing on the Beginning of a Visit

Then, I included only those sequences that oc-
occurred at the beginning of a user's visit to the site. I use the traditional Web analytics definition of visit: a continuous sequence of activity on the site with a gap of no more than 30 minutes. From a business perspective, it's important to understand how users are arriving at the site and whether they continue to discover new content and remain engaged. Given the widespread use of search engines and social link sharing, users could land on any indexed page, not only the home page.

To further reduce the number of possible permutations, I truncated the sequences. Initially, I truncated them after 10 events. After some initial experiments with visualizing the data, I realized that six steps were enough to capture the main patterns at the beginning of a visit. I used an explicit end marker to more clearly distinguish the end of a visit in the original data (for example, if the user left the site after viewing fewer than six pages) from a visit that I had truncated to six events.

So, for example, if the user's session consisted of two visits, ABCDEFGH and CDEFABCD, only ABCDEF and CDEFAB would be in the final visualization. I counted the subsequences (for example, BCD or CDEF) only as part of the longer main sequences, and I ignored the repetition of the CD subsequence in the second visit.

Including Only Direct Transitions
My final simplification step was to count only direct transitions between two page types. So, I counted AB and ACB as different sequences, even though they both involved a transition from A to B. This is an acceptable (and even desirable) result in this application area, provided that the page types have been carefully chosen and less meaningful pages have been filtered out.

For example, on an e-commerce site, home → product is likely to be a different type of path from home → search → product. In the first case, the user is in more of a browsing mode and selects a product from the home page. In the second, the user knows what he or she is looking for and searches for it.
Applying the Sunburst Visualization

Next, I applied the sunburst visualization, which has a compact radial format like the Disk Tree (for more on the Disk Tree, see the sidebar). Unlike the Disk Tree, it uses a space-filling presentation to explicitly show the proportion of the total value represented by each node, which in this case corresponds to the number of visits (see Figure 1). The sunburst was originally designed for hierarchical data—for example, visualizing an entire file system according to the number of bytes that each file and directory consumed.

Given that today’s websites consist mostly of dynamically generated content, not static URLs, a usage-based approach is the only one that makes sense. Sites might have an infinite number of possible “pages,” with many ways for users to navigate from one page to another. This requires a design solution that simplifies the aggregated navigation data’s complexity. Flow diagrams represent the best current solution in commercial Web analytics. However, they’re limited regarding the number of paths they can represent in a small amount of screen space.

Implementation and Outcomes

In mid-2011 I used D3 (Data-Driven Documents) to implement a sunburst visualization summarizing aggregated user paths on YouTube, and I made it available on an internal site. The visualization was shared extensively inside YouTube and the wider Google organization and was used by product and engineering leadership to illustrate and motivate strategy presentations. Almost three years later, it’s still widely known and referenced and is used in presentations to new employees. The internal short link has been followed more than 6,000 times; this usage continues to increase by approximately 40 to 50 accesses per week.

It’s interesting to reflect on what might have made this visualization stand out from the hundreds of graphs generated daily at Google, to extract lessons to apply to future projects. I believe that three main aspects were helpful.

• It’s an attractive, engaging format, encouraging users to want to understand what it’s showing.
• The compact representation fits easily onto a single screen, for immediate comprehension.
• The radial space-filling approach is a useful way to represent proportion, for users who are familiar with pie and donut charts (despite those graph types’ known limitations).

The first was the sunburst visualization itself:

• The visualization answered a question in which all YouTube employees were interested: how are users navigating our site?
• It made visible some usage patterns that were informally known but had never been displayed in such a concrete, easy-to-understand way. One product manager commented that looking at the visualization was like watching thousands of users navigating the site in the usability lab, but all at the same time.

The second was the data in the visualization:

• The visualization answered a question in which all YouTube employees were interested: how are users navigating our site?
• It made visible some usage patterns that were informally known but had never been displayed in such a concrete, easy-to-understand way. One product manager commented that looking at the visualization was like watching thousands of users navigating the site in the usability lab, but all at the same time.

The third was the user experience design, research, and branding:

• Despite the data’s complexity, the visualization’s interactive elements make it easy to understand...
when users first encounter it.
- Before releasing the visualization widely, I did several informal user studies, watching people use it for the first time. I then made design changes based on what confused them.
- I gave the visualization an easy-to-remember internal URL so that people could revisit it when they needed the information.

Following this visualization’s success at YouTube, I created a modified version that was easy to reuse with different data. Many teams at Google took advantage of it. For example, the original visualization had covered only desktop and laptop use of YouTube, so the teams working on the mobile and TV versions of YouTube quickly created their own visualizations, for comparison.

The YouTube TV application provides a good example of an actionable insight that was gained from the visualization. On the TV platform, when a video has finished playing, YouTube automatically plays the next video—for example, a suggested video. When the YouTube TV team looked at their sequence visualization, they were surprised to see a repeated pattern of users doing a search, watching a video, doing another search, watching another video, and so on. Although the number of instances of this pattern was small, it was still bigger than the team expected, given that typing is difficult with a TV. This pattern indicated that users sometimes didn’t want to watch the automatically cued-up video and were choosing to search for something else. This issue wasn’t exposed by the team’s existing metrics, which were all very positive. In contrast, in the visualization, the pattern stood out clearly and memorably.

In response, the team implemented a feature to give users more choices of videos to watch next, before autoplay starts. This feature reduced the incidence of the repeated search–watch sequences and led to improved metrics for the product overall—that is, users watching more minutes of video.

From my experiences using this visualization at Google, two clear areas for future research have emerged.

The first is comparisons. It’s natural to want to compare a sequence visualization to others related to it—for example, comparing YouTube on a desktop or laptop to YouTube on a mobile device. An area for future work is to facilitate such comparisons because people can have difficulty drawing conclusions when seeing two sunbursts side-by-side. This is especially true if the sunbursts are very similar (such as in a comparison of the sequences of users in an experimental condition and a control condition).

The second area is reverse application. In some cases, it’s of interest to understand the sequences leading to the user’s end state or a particular goal, rather than the user’s behavior at the beginning of the visit. I’ve experimented with versions of the visualization in which the center represents the goal achieved, to understand the sequences that most commonly led to that goal. However, these versions require additional work to make them as easily understandable as the original version.

You can try out the visualization at http://bl.ocks.org/kerryrodden/7090426. The full code and data for it are at https://gist.github.com/kerryrodden/7090426.

Acknowledgments
Thank you to everyone at YouTube who gave me feedback and encouragement while I worked on this visualization, especially Josh Sassoon for user experience design help, Yannet Interian for discussions that led
Kerry Rodden is a quantitative user experience researcher at Google, analyzing large-scale usage data to inform product design. She has worked on products including Search, Gmail, and YouTube and is interested in using data visualization to create compelling deliverables from user experience research. She has a PhD in human–computer interaction from the University of Cambridge. Contact her at krodden@google.com.

References

Figure 3. The visualization from Figure 1, showing the interactive elements that make it easier for the user to understand the relationship between the sunburst and the original sequences. When the user’s mouse hovers over a segment, the visualization highlights the full path to that segment and desaturates all other segments. The path also appears at the top of the screen.

Figure 3. The visualization from Figure 1, showing the interactive elements that make it easier for the user to understand the relationship between the sunburst and the original sequences. When the user’s mouse hovers over a segment, the visualization highlights the full path to that segment and desaturates all other segments. The path also appears at the top of the screen.