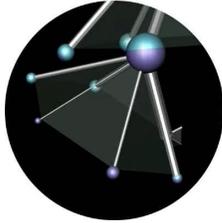Viant
Innovation
Center
 Project

ONLINE TRAFFIC VISUALIZATION

A tool for making group behavioral patterns visible

ABSTRACT

After a Web site launches, it is hard to know whether users are behaving as designers anticipated. This is particularly important for sites where designers have gone to the trouble and expense of personalization, tailoring different versions for different audiences. This paper describes a tool for analyzing site logs and translating that data into visualizations. Using these visualizations, designers can pinpoint more easily and effectively where their site succeeds and fails, and so make smarter decisions about how to retune it.

Viant is a leading Internet consulting firm that helps clients in a broad range of industries plan, build, and launch digital businesses. Established in 1996, Viant has become a trusted partner to Fortune 500 and emerging companies.

The Viant Innovation Center is a research group that imagines how people will use networks in the future, and helps clients profit from these insights. The Innovation Center is composed of an incredibly diverse team of people with skills in architecture, cinema, human factors, 3D animation, programming, music, design, and cognitive science. We explore themes inspired by clients including personalization, smart appliances, agent systems, and collaborative environments.

INTRODUCTION: WHY VISUALIZE?

When you get stuck in a traffic jam in real life, there's no way out. With luck, you might be able to sneak down a circuitous alternate route, but unfortunately you can't just change your mind about the journey and wish yourself back at home. A Web user can. At any moment, when he or she gets frustrated, a user can simply leave a site—the site that designers have spent countless hours working on, perhaps even personalizing it for different groups of users. The barriers to leaving are very low.

Digital businesses are aware of the importance of monitoring customer experience. But they typically track just a few key financial factors, such as conversion rate and percentage of repeat buyers. Sometimes they'll examine customer e-mail or conduct usability tests. These businesses need a way to observe customer behavior on their sites more closely.

In fact, most online businesses do have detailed information about user behavior at their fingertips. To better appreciate this, contrast urban traffic engineers. To gather data on traffic flow at different times, they have to go down the city streets and position boxes that electronically register every car passing by. Luckily, Web designers have site logs packed with information. The only problem is the difficulty of discerning patterns in this confusing mass of data.

Over thousands of years, we human beings have refined a visual cortex that allows us to see patterns quickly. We can process a mass of optical data in a split second. Walking down a busy street in New York, for example, we can pick out the one face we know from the crowd in an instant. Ironically, modern technology can help us use this ancient ability. A visualization tool allows us to "see" site-log data so that we can pick out patterns more easily. Urban planners have already learned to harness this primal part of our brains. In the last fifty years, they have become the experts in using visualization.

WHAT CAN WE LEARN FROM URBAN PLANNERS?

Urban planners understand that sometimes you need to step back in order to see the whole. For half a century, they have been developing a variety of sophisticated techniques for mapping cities and the traffic flowing through them. They have taken the bird's-eye view to levels the bird never dreamed of. A classic technique is thematic mapping (also called overlay mapping). Land-use planner Ian McCarg pioneered this method in the 1970s. Hired by the state to choose the best place to put a scenic parkway through Virginia, he gathered data on population spread, vegetation, and more. Then he layered the different data sets over a geographical map of the state, using colors to represent them. This "picture" of his problem made it easy for him to pick out the best route: the one where, simply put, there were the most flowers and the least people. Now urban planners use McCarg's method too.

BUT ISN'T THERE A DIFFERENCE BETWEEN THE ONLINE WORLD AND THE OFF-LINE ONE?

Of course, managing a digital business and planning a highway are two very different activities, and there are limits to this analogy. For one thing, city planners might be concerned about traffic jams, but Web designers certainly aren't. Unconstrained by space, they delight in user volume. Instead they fear wastelands and leaks (wastelands are tracts of the site that attract few visitors, and leaks take place when a person hits the back button, closes the window, or types in a new URL).

And life is easier for the Web designer. Urban problems are comparatively obvious—after all, you don't even need a visualization to see bumper-to-bumper traffic in New York—but they can take many years and millions of dollars to fix. In comparison, Web design problems are quick and cheap to solve. And with a visualization tool, they are easy to spot, too.

HOW DOES TRAFFIC VISUALIZATION WORK?
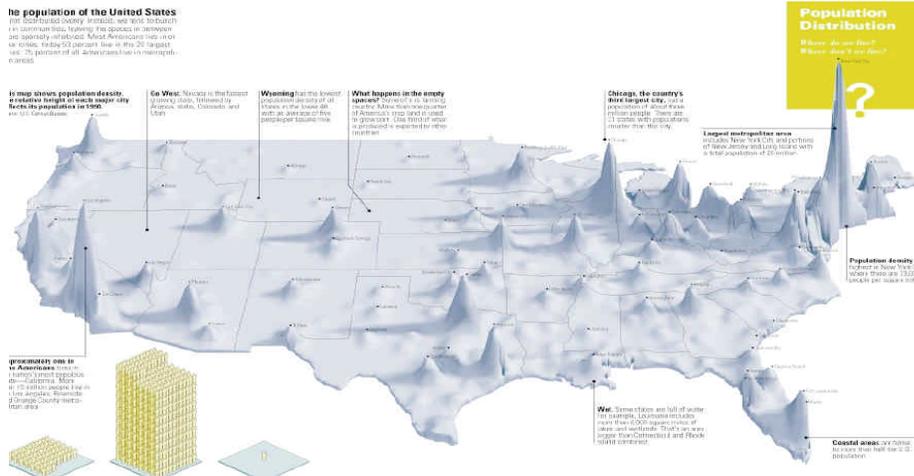
This classic example of thematic mapping layers population density over a map of the United States. Before you can create your visualization, you need to answer the question: what map will you use and what data will you layer over it? We'll start with the layer and talk about the map later. It may seem obvious that what the Web designer wants to layer over his map will be user data. But it's more complicated than that. Designers know much more about their users than just how many of them there are. This means that they can segment their audience and then trace the behavior of those individual segments onto their chosen map.

HOW DO I GATHER USER DATA?

A site log records the number of users passing through the site. It collects a huge amount of detail about them without the need for intrusive questionnaires. For example, a log registers a user's technology profile. It tells us whether his modem is fast or slow, what browser he is using, and whether he is visiting from a PC, Mac, or Palm. The site log also tracks a user's path over the Web: it records his previous URL, and it notes whether a user was directed to the site through a search engine or whether he typed in the address himself.
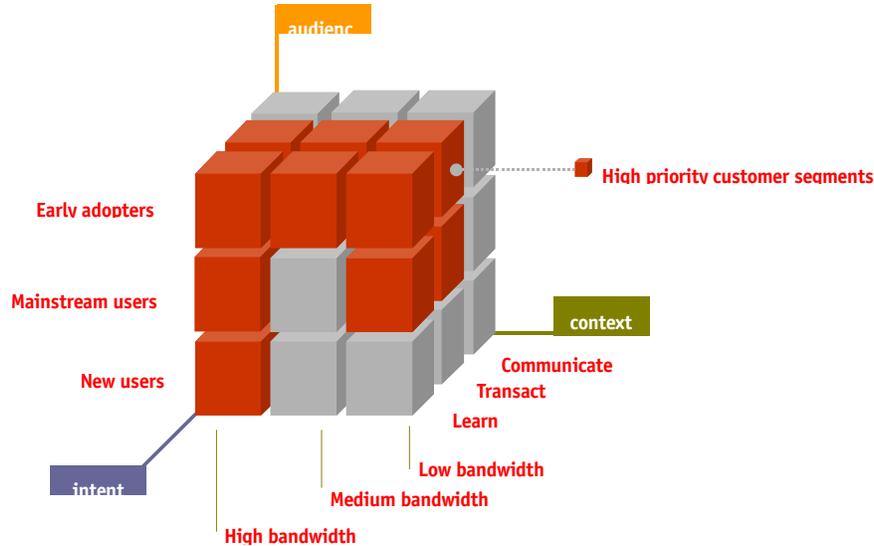
And most importantly, a site log follows a user's journey through the site: what pages he visited, what he did there, and how long it took.

There are also other, more controversial ways of gathering data on users. Companies like DoubleClick and Engage are sharing information across sites in order to build user profiles. And researchers, with one database as their Holy Grail, are working out how to consolidate information gathered across consumer touch-points. This means that customer call centers could use a person's phone number to add any information they have received from him to his record in the site log. With that end in mind, DoubleClick has allied itself with Abacus Direct, which gathers user data off-line.

And if you really want to know something, you can always use the old-fashioned method: just ask. We'll cover this in more detail later. First let's examine how you will use this data to better understand your users.

HOW DO YOU SEGMENT YOUR AUDIENCE?

We identified three broad factors that inform how we design different site experiences: Intent, Audience, and Context.   1

audienc

High priority customer segments

Early adopters

Mainstream users

context

New users

Communicate
Transact
Learn

Low bandwidth

intent

Medium bandwidth

High bandwidth

INTENT

Users might want to learn (for example, to compare prices, to get information about job vacancies), to communicate, or to get goods and services. Off-line retailers typically divide customers into "seekers" and "browsers" in order to calculate what kind of service to offer. If we can capture user intent, then we can meet needs more directly—helping seekers, for example, and leaving browsers to look about at leisure.

CONTEXT

This factor consists of environment in which the user is making his visit, and the technology platform he is using. The person glancing down at his Palm Pilot on a crowded train on his way to work might be more interested in doing a quick errand than in browsing. An airline passenger might be impatient with heavy pages, since plane connections are slow and expensive. In addition, no airline passenger wants to spend too long surfing the net on a tiny table designed to hold a beverage and not much else. In both cases, environment and technology platform combine to influence the user's needs.

AUDIENCE

This includes demographic data like age and gender. It might also include differences between attitudes and learning types: is the user a technophile, a technophobe, or somewhere in between? Is he a new visitor or an old hand? Is he an auditory or visual learner? "Audience" encompasses user characteristics that change slowly, if at all.

---

[1] 1 This segmentation method is also used in Experience Architecture™, Viant's method of personalizing Web sites. For more information, please contact innovation@viant.com.

To get an idea of how segmentation works in practice, let's examine an imaginary site, that of XYZ Services Corp. Call it xyz.com. The site log shows us that most of the visitors are students, clients, journalists, and investors. So we have an obvious way of segmenting users along the Audience axis. If we combine that with a Context axis—say, bandwidth, for example—we get the table below. For each intersection there is a clear business goal.

	Low bandwidth	High bandwidth
Students	*No students with low bandwidth*	Hire new employees
Press	Provide press releases	Provide streaming video press releases
Potential clients	Provide information on Viant, offer opportunity to sign up for mailing list	Provide streaming video of client stories
Investors	*No investors with low bandwidth*	Provide streaming video press releases and sign up to chat with C*X*O

Table showing business goals for four audiences and two contexts.

When they have segmented their users, designers identify business goals along with targeted content and functionality for each segment. Then they use the visualization tool to see if the site is helping them to meet those goals. If not, they can put their fingers on exactly how to do so.

WHEN SHOULD I INFER, AND WHEN SHOULD I ASK?

"I have answered three questions, and that is enough,"
Said his father; "don't give yourself airs!
Do you think I can listen all day to such stuff?

Be off, or I'll kick you down stairs!"
"You are old, Father William," from   *The Adventures of Alice in Wonderland*

You may not have to run the risk—like Father William's son—of asking too many questions and causing offense. Plenty of data can be discreetly harvested from site logs. You know a user's technology platform. And from a user's recorded path on the site you can sometimes divine his intent. It is on the topic of Audience that site logs are perhaps least informative. You may wish to ask your users about themselves directly.

There are many examples of data-gathering techniques—ranging from covert to overt, and from subtle to nosy. Some sites tempt users to enter sweepstakes and then pounce on their zip codes and other valuable pieces of information. Others require the user to enter his data on a registration form.

Whether you decide to use any of these techniques is up to you. Many sites decide that possible affronts to user privacy are not worth risking, and that you will make do with what you can find in the site log. Many sites now use cookies to determine when someone returns to a site without requiring a log in. But since users visit from different computers (at work, at home, at the cyber-café), cookies are not an foolproof way of identifying him or her.

WHICH SEGMENTS SHOULD I LOOK AT?

 Audience, Intent and Context encompass tremendous variety. You are going to be interested only in certain swaths of data—your high-priority customer segments, for example, or perhaps those that you are worried about. With a visualization tool, it is easy to illuminate only these target or problem segments—like flipping a switch in a vault of filing cabinets to reveal just the row you want to investigate.

Equipped with a Boolean search capability, a visualization tool can even track precisely defined segments. For example, we might ask the tool to show us the Hierarchy View for kids and adults (users segmented by age). With Boolean segmentation, we could ask it to show us age and bandwidth. This might enable us to perceive that kids

with low bandwidth are leaving, while adults with low bandwidth have the patience to wait while pages load. The solution might be to learn a lesson from the children's playground at McDonald's and build a Web equivalent—lighter pages for kids.

But what if you don't know what you want to know? You have a vague feeling that there might be some answers in your vault of data—only you're not yet sure what the questions are. No problem. With a visualization tool it is easy to trace several different user segmentations over a site map. Even without a specific suspicion in mind, you can simply monitor your site's health, like a doctor performing a checkup.

INTERPRETING THE VISUALIZATIONS

Now you have your data. The next step is to choose what you are going to map. Like Ian McCarg choosing his scenic route, or the urban planners following traffic, you may wish to begin with geography—or rather, virtual geography. You could layer user behavior onto a site map in order to observe it page by page, locating the hot spots, leaks, and wastelands. You could also layer data over a time line, in order to trace a site's popularity among different segments day by day. And finally you could place data on what you might call a psychological map.
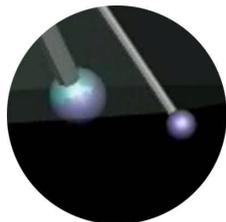
Size of Node
shows number of
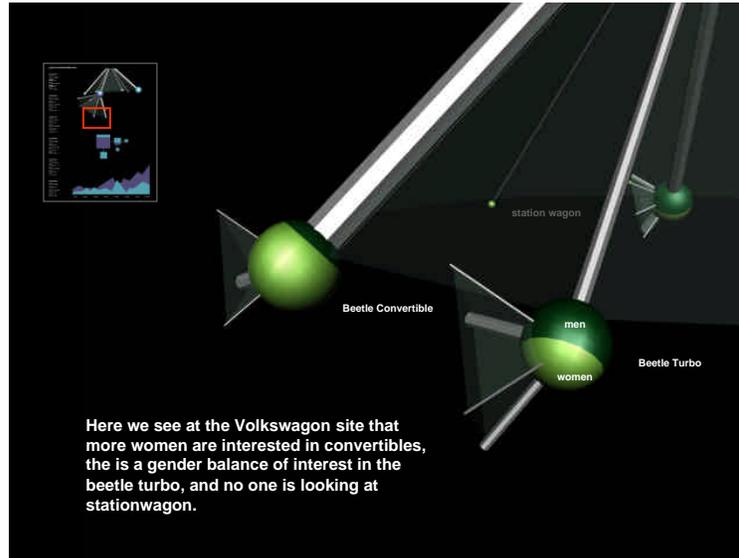*unique visitors*
that pass through
that section

Width of Spoke
reflects total
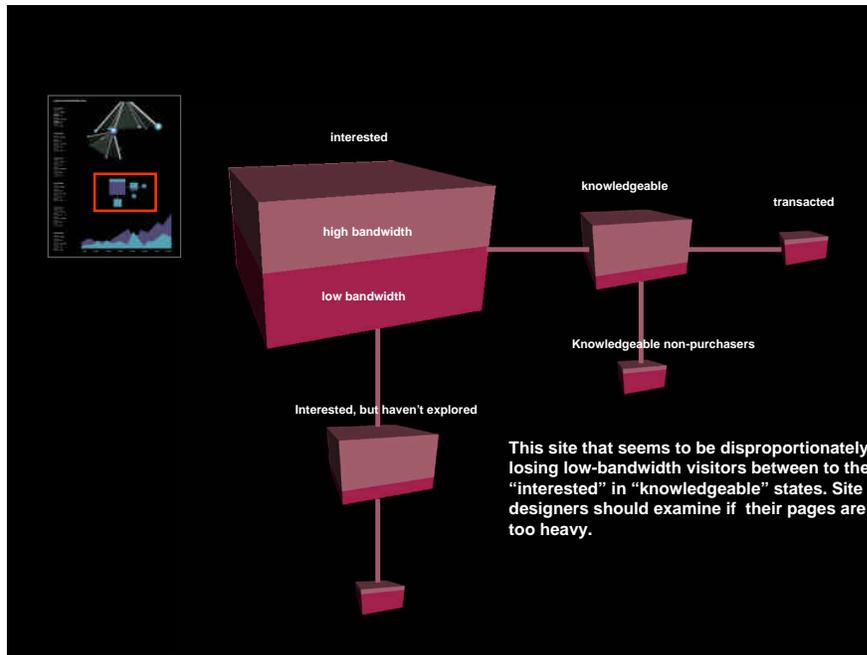*traffic to through*
that section
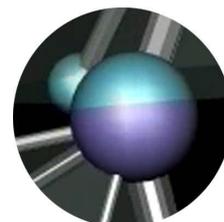
Colors
indicate
percentage of
selected audiences

Here we see at the Volkswagon site that more women are interested in convertibles, the is a gender balance of interest in the beetle turbo, and no one is looking at stationwagon.

Hierarchy View

In the Hierarchy View, each node represents a page or a section of a site. In the Volkswagon site pictured above, more women are interested in convertibles, men and women are more or less equally interested in the Beetle Turbo, and no one is interested in the station wagon. One useful response to the traffic pattern represented here might be to put a prominent link to the station wagon on the home page—maybe people are simply missing it.

The figure shows labels: interested, knowledgeable, transacted, high bandwidth, low bandwidth, Knowledgeable non-purchasers, Interested, but haven't explored, and the caption text:

This site that seems to be disproportionately losing low-bandwidth visitors between to the "interested" in "knowledgeable" states. Site designers should examine if their pages are too heavy.

Goal Path View

In the Goal Path View, each square represents a different step along the way to falling in love with your site. In square one, for example, you might group all users who have been to the site. Square two, the second step, might take several different forms. In this square you could group all users who have checked out five or more pages, stayed at the site for at least three minutes, or just visited it twice. The next square might represent users who have started filling a shopping cart, and the next step would be making a transaction. And the final step in your relationship with a site (the equivalent of marriage, if square one is first base) is becoming an evangelist. The evangelist spreads word of the site to others, and even explicitly recommends it (for example, he might e-mail an article or wish list to a friend).

The site represented above seems to be disproportionately losing low bandwidth visitors in the transition from "interested" to "knowledgeable." Site designers should consider whether their pages are too heavy.
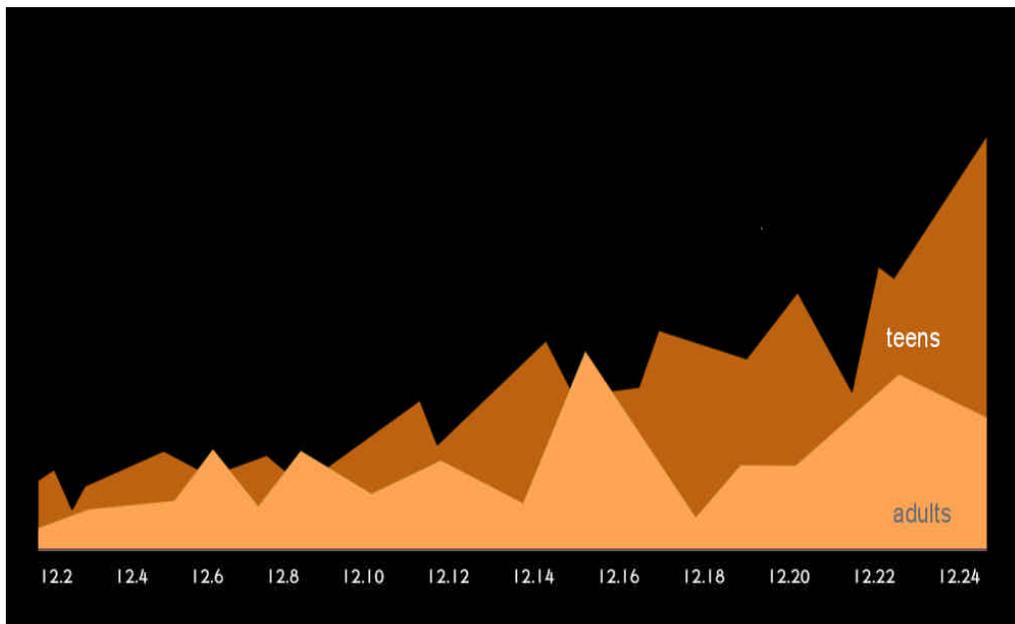
Time Slice View

In the Mini-disc campaign, pictured above, advertising impacted teens powerfully, but didn't generate much reaction from adults. One response might be to refocus the ad campaign to attract teens as well as adults.

WHERE NEXT?

A visualization tool could layer time spent per page over the site map, combining elements of the Hierarchy View and the Time Slice View. Designers would have to allow for a little distortion here, since some users leave a page open while doing something else, or hop from one browser window to another. Nonetheless, this view would be a good way of gauging confusion. If people spend a long time on the registration page, for example, designers might conclude that the form is too complicated. This view would also be a litmus test for boredom. If people are just skimming through pages, spending ten seconds or so on each one, then the designers need to add something that will catch the user's eye.

A visualization might also look beyond the site itself, recording the behavior of users on rival sites and linked sites, and so expanding the Hierarchy View. This could be done, like television's Nielsen ratings, by logging the behavior of a sample group. Designers could compare the behavior of users on competitors' sites. For example, they might discover that while people gather information on their site, they go to a rival site to make transactions. And designers could get an idea of the typical length of a visit to a rival's site and compare it with their own. In addition, monitoring linked sites would reveal what sort of users they are bringing to yours.

Finally, designers could create an ideal version of each View, and map the ideal over the real. That way, it would be a snap to spot the differences between user behavior and business goals and implement the necessary changes.

VISUALIZE WORLD PEACE—BUT FIRST, VISUALIZE WEB TRAFFIC

Remember the hardworking traffic engineers who have to get their data from electronic boxes that they distribute throughout the streets? Imagine how they would feel if they suddenly discovered that the boxes would not only register the number of cars passing by, but could also see the make of cars and their passengers. These visionary boxes would even register where the person was going and why. How much more precise traffic planning might be if engineers had such penetrative vision.

Web designers have the means to attain such precision in monitoring traffic. Like the engineers' electronic monitors, a site log gathers data. And like an X-ray, a visualization tool allows designers to pinpoint problems. Unfortunately, doctors can't reform bodies any more than urban planners can remodel cities. But Web designers can reshape their sites. With a visualization tool they pinpoint their site's weaknesses and see how to iterate it, personalizing for different audiences. With a visualization tool, they can also see their site's strengths, and so validate their labor and expense.