CLiC Workshop, "Weeding Your Website with Basic Google Analytics," Grand Junction CO, 21 March 2017

[slide 1]

My name is Babi Hammond. I work for the State Library in Denver as their Digital

Experience Consultant. Mostly what I do is build and administer websites for library-

related organizations, but I also do some research and consulting to help people make

their websites better organized and easier to use.

Thanks for coming to my talk. I expected four people to show up for this, because I'm

their ride home. It's nice to see a few more than that in the audience.

It's especially nice because I know that the topic of this talk may not seem at first all that

exciting. Words like "analytics" and "weeding" might conjure up thoughts of chores—dull

things you have to do. But I think analytics is pretty cool, and weeding is a worthwhile

activity. And I promise you that during this presentation we'll get to talk about cool things

like

[slide 2]

spiders and

[slide 3]

'bots. And

[slide 4]

cookies! So stick with me.

This talk is mostly about how Google Analytics works, and some ways you can be sure you're getting the best possible data out of it to help you improve your website. But I want to start by making the case for *why* you need to weed your website, before I move on to the how.

[slide 5]

This is picture of a drawer in my kitchen. Let's say we're in my kitchen right now, and I pointed to the drawer and said, "hand me the potato peeler, please?" Could you find it? For me, this is a trivial task. I probably look into this drawer at least three times a day, and I know exactly what the potato peeler looks like. And I know that it's usually somewhere in the back for some reason. So for me, this kind of clutter and disorganization is not really a problem. Also, I don't have a lot of visitors to my kitchen, and very seldom do I ask anyone to peel potatoes for me.

But the situation is different if we're talking about a website—where, ideally, you have a lot of visitors, many of them new to the site, and they might not really know exactly what they're looking for. So you can't afford to have a website like my kitchen drawer—you need to know what the most important things are on your site, make them easy to find, and weed out the things that are not important.

[slide 6]

So how do you tell which things on your site are not important, or are less important? Well, by itself Google Analytics won't really tell you what's most important. What's important on your site grows out of a clear sense of your library's mission, and a determination to make your site as efficient as possible in serving that mission. At its

best analytics can tell you what your patrons are looking at on your website. And that is one way of seeing what is important to *them*. There are other ways as well, like what patrons are asking for at the front desk, maybe, or the results of satisfaction surveys or other kinds of research. There are many kinds of data, judgements, and even stakeholder politics that usually go into determining the look, structure, and content of a typical library website.

[slide 7]

Still, since analytics gives you information that no other source can, it should be an important input into your decision making. So you should understand the basics of how it works, what the numbers it gives you really represent, and how to ensure that you're getting the highest quality data possible from your analytics system. And those are the things that the rest of this talk will deal with.

So let's turn now to Google Analytics, and talk a bit about how it works.

[slide 8]

This is what you typically see when you log in to a Google Analytics account—the Audience Overview report. For the selected time frame

[slide 9]

(shown in the upper right there) you get numbers right away for Sessions, Users, and Pageviews for the site as a whole.

[slide 10]

And if you want to look at numbers for a particular page,

you can click on Behavior in the left-hand menu, then Site Content, and All Pages to see the Pages report.

[slide 11]

The table at the bottom gives you the pages with the most pageviews.

[slide 12]

You can filter the table for a particular URL with this search box here, to display data for the page or pages you're interested in. So, if we accept these numbers at face value at least, we can now see how popular particular pages are; we can just export the list and start weeding out the least popular pages.

But what do these terms—Pageviews, Unique Pageviews, and Sessions really mean? How accurate are these numbers, and how well do they equate with how many actual patrons use these pages?

The terminology of Google Analytics gets confusing pretty fast, so let's try and make this a little more concrete by looking at what happens when a single person visits a library website. So this person, let's call him Fred,

[slide 13]

has recently moved to Grand Junction

[slide 14]

and he wants to see what the local library has to offer. Maybe he'd like some books on how to quit smoking, or how to recover from poor Photoshopping. In any case, he gets on his computer and

[slide 15]

Googles "Grand Junction public library", and then clicks on the link for the Mesa County Libraries website, which takes him to the Mesa County homepage.

[slide 16]

From here he clicks on Hours and Locations

[slide 17]

 to find a branch near him. And then he clicks on the Central Library,

[slide 18]

and looks at the map to make sure he knows where he's going. Then, just because he's curious, he goes back to the homepage

[slide 19]

and looks around for a while. Then he closes his browser and sets out for the library.

[slide 20]

We can reconstruct Fred's path on the web like this—he went to Google, then the Mesa County homepage, then Hours & Locations, to the Central Library, then back to the homepage, and then closed his browser. Let's translate this into the language of Google Analytics, and see how Fred's trip to the site would show up in our analytics data.

[slide 21]

Our Google Analytics account is completely unaware of Fred until he first enters the site. Once he does, since he hasn't visited the site before he becomes a New User. Each page he looks at in the site generates a Pageview, and together all those pageviews—from the moment he enters the site to the moment he leaves the site or closes his browser—constitute a single session. Also note that when he returns to the

homepage in the last step of his session he generates another pageview for that page, but not a unique pageview, since he had already visited that page once before during the session. A unique pageview is generated only once per session. If Fred had visited the Mesa County homepage, and then refreshed the page six times, that would generate seven pageviews but only one unique pageview.

[slide 22]

Google Analytics will also tell you something about how Fred got to the site. In this case his use of Google would show up in the Acquisition Channels report as "Organic Search." *Organic* in this context just means that Fred got to the site from regular old natural Google search results, not a *paid* search placement. Basic Google Analytics will not tell you anything about where Fred went when he left your site—whether he went to some other site, closed his browser, or what. You can edit the Google Analytics code to get this information, but that's beyond the scope of this presentation.

[slide 23]

So these are the basic units of your Google Analytics—the User, the Session, and the Pageview. You can visualize how they relate to each other a little more schematically by thinking of it as kind of like a Venn diagram. Users can have multiple sessions, and sessions are made up of at least one, and usually multiple pageviews.

[slide 24]

Or, even more schematically, you can think of it as a pyramid. The basic unit is the pageview down here at the bottom. A sequence of pageviews constitutes a session, and sessions are associated with a particular user.

[slide 25]

So these different numbers answer different questions. If you just want to see how many times a page has been looked at, then probably you want to look at pageviews.

[slide 26]

Maybe some visits to the site viewed that page numerous times, but you want to know who many visits involved at least one look at the page. In that case, you want to look at unique pageviews, which equate to the number of sessions that included at least one visit to that page.

[slide 27]

But maybe you want to know how many individual people looked at the site, regardless of how many times each person looked at it. In that case you'd want to look at Users. (You can also look at users who viewed a particular page, but it's a bit complicated.)

[slide 28]

So, when you come to a site, how does Google know what user you are? In our previous example, how did Google know that Fred was a New User instead of a Returning User? And when does one Session end and another one begin?

Perhaps in the not-to-distant future user identification will be done by facial recognition software gazing at you through your computer's webcam, but until that blissful day Google Analytics uses—

[slide 29]

Cookies! (We're getting to 'bots and spiders, I promise.) I have to admit that the cookies pictured here—not to mention the furry blue monster—are not really the kind of cookies we're talking about. The cookies that Google Analytics gives out are just tiny text files

that the browser saves on your computer's drive so that it can remember things about what you were up to while you were browsing.

[slide 30]

These User and Session cookies have expiration dates, though they are renewed whenever their is new activity. (And really Google Analytics now uses just one cookie— the information about the User and Session is stored in different bits of data in that one cookie. But it's more fun to say "cookies" than "bits of data," so just bear with me.) User cookies expire 2 years after the end of the last activity.

[slide 31]

So when Fred visited the Mesa County Libraries website for the first time, the Google Analytics code on the site checked his computer's hard drive for a User cookie. It didn't find one, so it wrote a new one and counted Fred as a New User, and assigned him a Client ID. The next time Fred visits the website, he will be counted as a Returning User, and his user cookie will be renewed—it will always expire 2 years after the end of his last visit to the site. Google doesn't really know who Fred is—it just looks for the User Cookie and a Client ID. When the facial recognition thing gets started that may change, but for now that's it.

[slide 32]

The session cookie is by default set to expire 30 minutes after the last activity.

[slide 33]

So if Fred visits our website, the analytics code will check to see if there is a current session cookie already there. When Fred leaves the site, the session is ended. Or, if Fred leaves the site open but doesn't visit any other page for the next 30 minutes, the

session will expire. When he returns to the site, everything will be counted as a new session.

[slide 34]

Okay—so you've got these three basic sets of data to tell you how your site is being used, and you know a little more about how they work.

But how accurate are they? How well do they reflect what your patrons are actually doing on the site? The answer is that they are almost surely inaccurate to some degree. So it helps to understand why they might be inaccurate, and know what steps you can take to minimize their inaccuracy.

[slide 35]

Let's start with some sources of inaccuracy that we can't do much about—we just have to be aware of them and try to factor that into our decisions somehow. The big ones here, I think, are cookie problems. As I mentioned, Google Analytics doesn't (at present) have any way to identify your site's users and sessions except through cookies. Nor for most sites is there a reliable way to unify data for users who use different devices or different browsers.

[slide 36]

So here's Fred again. When he first visited our site to become a New User, he was using his home desktop computer.

[slide 37]

But he also uses these other devices. If he visits the site on any of these, he's counted again as a New User. Plus, he might use more than one browser on some devices. When he looks at your site in Chrome, he'll get a User Cookie. When he looks at your

site in Firefox he'll get *another* user cookie. So there is probably some degree of device and browser inflation in your user numbers.

[slide 38]

Plus, there's this problem. How many people have their browsers set up to block cookies, or otherwise delete cookies at some regular interval? On my browser, I usually allow cookies from the sites I visit, but all my cookies are deleted when I close the browser. So I'm a New User almost every time I start my browsing a site. There is no definitive answer that I've found for how many people routinely delete or block cookies. The commonly-cited figure is that about 30% of internet users delete their cookies at least once a month.[1] Perhaps as common today is the use of ad-blocking software, which can sometimes block Google Analytics altogether. Studies have shown that somewhere between 8% and 20% of internet users use ad-blockers, and these users may not show up in your analytics at all.

[slide 39]

So one user probably does not equal one patron. But there's no way at present, at least without significantly changing your website's setup, to reliably correct for these sources of error. I tend to deal with these problems by not putting too much emphasis on New and Returning Users, and assuming that over most time periods I'll be looking at, these sources of error are more-or-less constant background noise, affecting various months and various pages about equally. So they don't really factor into comparisons from month to month or page to page.

---

[1] ComScore (2007) "Cookie-Based Counting Overstates Size of Web Site Audiences": http://www.comscore.com/chi/Insights/Blog/Cookie-Deletion-Rates-and-the-Impact-on-Unique-Visitor-Counts

[slide 40]

That was all kind of a downer, wasn't it, since there isn't really anything we can do to correct those errors? But now lets look at some errors we can correct, at least to some degree. And this is filtering out from your analytics data "junk traffic"—or traffic from sources that are almost surely not from the audience your site is designed to serve. This might be from…

[slide 41]

Spiders! Spiders are computer programs that crawl the web to index its content for search engines. Most are benign, or even beneficial. You probably want spiders from reputable search engines, such as Google or Bing, to crawl through your site's pages so they can be included in search results. These crawlers won't show up in your Google Analytics—they just sort through the code of your pages. But some spiders are less polite, and they can pollute your analytics data with bogus pageviews and sessions. Other junk traffic on your site might be from

[slide 42]

'Bots! 'Bots are also crawling all over the web, but they're performing some other information-gathering task, not indexing your pages for search engines. Again, some of these tasks are benign. But some 'bots are looking for security vulnerabilities in your site, trying to harvest personal information, or looking to do some other nefarious deed. In the meantime they can also pollute your analytics numbers.

[slide 43]

Luckily there is a simple way to eliminate a lot of this bogus traffic. In your View Settings you should check this box to "Exclude all hits from known bots and spiders." This will

likely not completely eliminate bots and spiders from your data, but it will filter out the largest, most annoying ones.

[slide 44]

You should also take a look at the Location report to get an idea of how much of your traffic is coming from outside the United States. This traffic may be legitimate in the sense that it is actual humans looking at your web pages. But in some cases this traffic can seriously skew your numbers, if what you're primarily interested in is what content your actual patrons are using.

For example, this is the Location report from a site I administer, the Colorado Virtual Library, from the first half of 2015. One of my first projects at the State Library was to redesign this site, and as part of the redesign project I spent a fair amount of time looking at the site's analytics. On this map, the darker the color of the country, the higher the number of sessions that originated from that country. Do you notice anything odd about the map?

There is a lot of blue everywhere, which is not that unusual, but India is pretty dark blue. It turned out that less than half of the site's sessions originated from the United States.

[slide 45]

To see data for users in the US I added a segment that looked only at sessions from the United States, and this was the result. A segment is a way to define a sub-set of your analytics data that meets some criteria you specify. If you don't know how to create one,

don't worry—it's really easy. But I'm not going to explain it here—it's important for the moment just that you're aware of the function and what it can do.

With the segment for only US users in place, you can see that only about 43% of the site's users were from the United States. While I'm glad that people all over the world, and particular in India, found some of the site's content valuable, they weren't really the audience the site was intended to reach. So I had to filter out that traffic to try to get a better picture of what the users we most wanted to reach were doing on the site.

Google's Location reports will let you filter your data down to the city level. I feel that even for a local library there might be legitimate reasons why someone from out of state might want to use your website. And some of your local patrons may be using internet connections that identify themselves to Google as originating in another state. But I encourage you to take a look at your data and see if you detect anything odd; then you can apply filters accordingly.

[slide 46]

So just to try and round things back up here: I've tried to make the following points to help you in weeding your website:

First, that analytics by itself doesn't tell you what content is important—but it can tell you what your patrons are actually using, and that's an important thing to know.

[slide 47]

We talked about how pageviews, sessions, and users relate to each other,

[slide 48]

How Google Analytics uses Cookies to determine users and sessions, and some of the ways this can create errors in your analytics data that you can't really correct.

[slide 49]

And I talked about a couple of simple ways to improve the quality of the analytics data you use to weed your site, by filtering out known bots and spiders and using geographic location filters to help ensure you're looking at traffic generated by your actual patrons, and not at some internet cafe in New Delhi or somewhere.

There are other ways to use Analytics to see not only what pages your patrons are looking at, but also how they get to your site and what they're looking for, but I will have to leave that to leave that for some other presentation.

[slide 50]

But before I end I want to return to the idea of weeding and analytics as a routine chore. Every analytics account is different, and I think it's important that you regularly spend time looking at your analytics data. But the obvious question is, how often is "regular"? I think it's a good idea to spend at least an hour a month looking at your data. But approach that hour as an exploration of the data, to get more familiar with the regular patterns, the ebbs and flows of your users. This will help you become more familiar with Google Analytics, and help you spot things that seem out of the ordinary, which might signal a problem or an opportunity for your library.

[slide 51]

But don't sit down each month to make decisions about what content goes or stays; a month's worth of analytics data is really not enough of a baseline for decision making. And although you want to maintain an up-to-date and efficiently organized website, you

probably don't want to be cutting content every month. Take more time to make those decisions.

[slide 52]

Weed, but weed wisely. Don't approach your analytics data as some kind of oracle, from which answers will arise. Instead you should come to your data with some focused questions, informed by your knowledge of how analytics works, and what kind of conclusions the data might support. The amount of data available in your analytics account is staggering, and you can spend weeks sifting through it and slicing and dicing the numbers in various ways. That's way too inefficient. Instead look at the data already knowing what kind of information you're looking for, and how what you find will help you decide one course of action over another.

[slide 53]

Those are my thoughts about Google Analytics and weeding your website. Here are some recommendations for other resources you can look at that might reinforce the information in this presentation, or allow you to dig a bit deeper into Google Analytics. Unfortunately, I have not found the one source to go to for comprehensive information about Google Analytics. The best thing I can recommend is Google's own Analytics Academy. They have tons of support documents, some free online courses, and videos to help you learn more about your analytics. Google is constantly tinkering with how its analytics system works, and Google's own pages, while not always completely up-to-date, are probably your best bet as a starting point.

[slide 54]

There's a lot of good info out there on the web, but it tends to be scattered about different sites and indifferently sourced. One of the best basic overviews of how Google Analytics works that I found was this article by Kristi Hines. It starts out with how to register for and deploy an analytics account, so in some ways is even more basic than this talk. But it's a good thing to bookmark as a general reference, at least until Google changes its reporting system again.

[slide 55]

Finally, if you search for things about analytics on the web, you will find a ton of stuff from companies that want to sell you their search engine optimization services. Some of these articles are useful, but many are really out-of-date, so be sure you're searching for recent articles. Also, much of this advice is geared toward businesses, not libraries, so things like adwords campaigns and monetization strategies are not really going to apply to you.

[slide 56]

It's kind of surprising how difficult it can be to answer some seemingly basic questions about what your'e seeing in your analytics reports. So if you find yourself baffled by something related to analytics, or if any questions occur to you after this talk is over, please feel free to email me or call me. I will do my best to help in any way I can.

Thank you very much.