Who are our users?
Gathering data on web visitors

Sherri Berger, California Digital Library
SCA Annual General Meeting
April 13, 2013 - Berkeley, CA
What do users like about the sites?
What’s frustrating for them?
Which new features should we pursue?
How do they prefer to find materials?
How can we reach out to them better?
Are they finding what they need?
Who *are* these people anyway?
No user sign-in
Privacy policy
How to collect data?
One week
Every visitor
At entrance
<table>
<thead>
<tr>
<th>Event Action</th>
<th>Total Events</th>
<th>Unique Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>declined</td>
<td>7,848</td>
<td>7,344</td>
</tr>
<tr>
<td>college/grad-student</td>
<td>454</td>
<td>444</td>
</tr>
<tr>
<td>k12-student</td>
<td>280</td>
<td>280</td>
</tr>
<tr>
<td>other</td>
<td>225</td>
<td>220</td>
</tr>
<tr>
<td>k12-teacher-librarian</td>
<td>134</td>
<td>133</td>
</tr>
<tr>
<td>faculty-researcher</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>genealogist-researcher</td>
<td>71</td>
<td>68</td>
</tr>
<tr>
<td>archivist/librarian</td>
<td>61</td>
<td>55</td>
</tr>
<tr>
<td>student, coll</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Brian’s code:

Response Rate

21% for OAC (2,162 total)
15% for Calisphere (1,283 total)
K-12 teacher 3%
K-12 student 3%
College or grad student 25%
Faculty or academic researcher 14%
Archivist or librarian 17%
Genealogist or family researcher 11%
Other 27%
others
user data magic
More questions
How **accurate** is our data?
What if that was an unusual week?
Can we get solid data without surveying all visitors?
Ongoing

Every 1,000 pages viewed
Bayesian Statistics

A **model** for analyzing data:

1. Establish prior knowledge
2. Estimate probability, predict outcomes
3. Gather more data
4. Refine estimate/prediction
What we want to estimate:

- K-12 student: 22%
- College or grad student: 35%

“Credible Interval”
Original data ("shotgun")

New data: 1-10 every day

Bayesian Formula

\[(\theta_1, \ldots, \theta_k) \mid (x_1, \ldots, x_k) \sim \text{Dirichlet}(x_1 + 1, \ldots, x_k + 1)\]

credible intervals
Questions & feedback?

sherri.berger@ucop.edu