How You Can Engage Patrons with Data (and Learn to Love Data Yourself)
ABOUT ME

• I’m one of two Research & Data Coordinators at the NNLM-PNR*

• I’ve worked as:
  • a public health epidemiologist (for Public Health – Seattle & King County)
  • the medical librarian for Group Health (now Kaiser Permanente – Washington)
  • a public library consumer health librarian (at The Seattle Public Library)

*National Network of Libraries of Medicine, Pacific Northwest Region (one of 8 outreach arms of the National Library of Medicine)
WHAT'S TODAY'S AGENDA? *

• How data are created and maintained
• How data are analyzed
• How data are visualized
• Crossover skills from librarianship to data wizardry
• How to be a critical data consumer (data’s pitfalls for the unwary)
• Tricks for finding data (or at least answers)
• Best practices around data literacy
• Learning opportunities

*I promise it will be more fun than it looks!!!
BUT FIRST, WHY DATA, WHY NOW?

• Increasingly, we encounter data in everyday life (we need it for understanding the news, engaging in public discourse, assessing information about weather patterns, etc.)

• Many data publishers, governments in particular, are making their data accessible to the public

• With more accessibility and more encounters come more patron requests for help in understanding and using data for a wide range of purposes
TODAY’S FOCUS: GOVERNMENT AND PUBLIC DATA

Small vs big data
Open vs closed data

TODAY WE WILL TALK ABOUT SMALLER DATA, LIKELY TO BE OPEN

(but feel free to ask about other types!)
In Washington State, almost every birth has a certificate filled out. So, this can be considered a data source that covers an entire population (newborns). However, the data aren’t always the best quality… what can go wrong?

This and next slide:
### Mother's Statistical Information

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>34. Mother's Medical Record Number</td>
<td></td>
</tr>
<tr>
<td>35. Mother's Prepregnancy Weight (Pounds)</td>
<td></td>
</tr>
<tr>
<td>36. Mother's Weight at Delivery (Pounds)</td>
<td></td>
</tr>
<tr>
<td>37. Mother's height</td>
<td></td>
</tr>
<tr>
<td>Feet: Inches:</td>
<td></td>
</tr>
<tr>
<td>38. Did Mother get WC food for herself during pregnancy?</td>
<td>☐ Yes ☐ No</td>
</tr>
<tr>
<td>39. Cigarette Smoking Before and During Pregnancy</td>
<td></td>
</tr>
<tr>
<td>Average number of cigarettes or packs per day:</td>
<td></td>
</tr>
<tr>
<td>Three months before pregnancy</td>
<td></td>
</tr>
<tr>
<td>First three months of pregnancy</td>
<td></td>
</tr>
<tr>
<td>Second three months of pregnancy</td>
<td></td>
</tr>
<tr>
<td>Last three months of pregnancy</td>
<td></td>
</tr>
<tr>
<td>40a. Number of Previous Live Births (Do not include this child)</td>
<td></td>
</tr>
<tr>
<td>Number Now Living</td>
<td>☐ None</td>
</tr>
<tr>
<td>Number Now Dead</td>
<td>☐ None</td>
</tr>
<tr>
<td>41a. Number of Other Pregnancy Outcomes (Spontaneous or induced losses or ectopic pregnancies)</td>
<td>☐ None</td>
</tr>
<tr>
<td>41b. Date of Last Other Pregnancy Outcome (MM/DD/YYYY)</td>
<td></td>
</tr>
<tr>
<td>42a. Date of First Prenatal Care Visit (MM/DD/YYYY)</td>
<td>☐ No Prenatal Care</td>
</tr>
<tr>
<td>42b. Date of Last Prenatal Care Visit (MM/DD/YYYY)</td>
<td></td>
</tr>
<tr>
<td>43. Total Number of Prenatal Visits for this Pregnancy</td>
<td></td>
</tr>
<tr>
<td>(if none, enter '0')</td>
<td></td>
</tr>
<tr>
<td>44. Date Last Normal Menses Began (MM/DD/YYYY)</td>
<td></td>
</tr>
<tr>
<td>45. Was mother transferred to higher level care for maternal medical or fetal indications for delivery?</td>
<td>☐ Yes ☐ No</td>
</tr>
<tr>
<td>If yes, name of facility mother was transferred from:</td>
<td></td>
</tr>
<tr>
<td>46. Principal Source of Payment for this Delivery</td>
<td>☐ Medicaid ☐ Self Pay ☐ Private Insurance ☐ Indian Health ☐ CHAMPUS ☐ Other Gov’t ☐ Other (Specify)</td>
</tr>
<tr>
<td>47. Newborn Medical Record Number</td>
<td></td>
</tr>
<tr>
<td>48. Birth Weight</td>
<td></td>
</tr>
<tr>
<td>lbs: ozes: grams:</td>
<td></td>
</tr>
<tr>
<td>49. Infant Head Circumference (cm)</td>
<td></td>
</tr>
<tr>
<td>50. Obstetric Estimate of Gestation (Completed weeks)</td>
<td></td>
</tr>
<tr>
<td>51. Apgar score at 5 minutes: If score is less than 6, score at 10 minutes</td>
<td></td>
</tr>
<tr>
<td>52. Plurality – Single, Twin, Triplet, etc. (Specify)</td>
<td></td>
</tr>
<tr>
<td>53. If not single birth – Born 1st, 2nd, 3rd, etc. (Specify)</td>
<td></td>
</tr>
<tr>
<td>54. Was infant transferred within 24 hours of delivery?</td>
<td>☐ Yes ☐ No</td>
</tr>
<tr>
<td>If yes, name of facility infant was transferred to:</td>
<td></td>
</tr>
<tr>
<td>55. Is infant living at the time of report?</td>
<td>☐ Yes ☐ No</td>
</tr>
<tr>
<td>56. Is infant being breastfed?</td>
<td>☐ Yes ☐ No</td>
</tr>
</tbody>
</table>

### Newborn's Statistical Information

### Medical and Health Information
WHO FILLS OUT THE FORM?

• Nurse at mom’s bedside?
• Hospital registrar using mom’s medical record?
• Midwife after home birth?
• Mom herself?
• Dad or relative?

AND...

• Can any of them remember the exact dates of things like prenatal visits?
• Does the hospital insist on completion of that field/other fields?
POTENTIAL DATA ISSUES WITH PRENATAL CARE

• Misremembering dates
• Not wanting to look like they started too late, so, fudging dates
• Filling out dates incompletely
• Skipping the question even if they remember month and year, since they can’t remember the day
• Human error if using medical record
• Cultural differences (European dates reported differently)
• No one wants to fill out the “No” box
• If hospital doesn’t insist on good quality data for that item, hospital staff may ignore it
AND ONCE THE CERTIFICATES ARE FILLED OUT

• Data has to be entered into a machine readable format—data entry errors?
• Data are vulnerable to technical glitches and file corruption
• Data must be “cleaned”—what to do with a record where mom’s age is 199?
• Coding errors must be fixed—human race is not the same as other race
• Certificates have to be traded across borders (certificates for Oregon moms having babies in Washington are sent to Oregon, and vice versa)
• ALL THESE THINGS TAKE TIME
HOW DOES ALL THIS PLAY OUT IN MAINTAINING AND USING THE DATA?

• **Missing data**
  
  "What if a large local hospital accounting for a major proportion of county/state births was not filling out the prenatal care question? Along with missing data from other hospitals, the resulting prenatal care info would be very poor quality!"

• The data that there are can take longer to release if extensive cleaning needs to be done

• The data that there need to be reviewed for quality control—if they are poor quality, it has big impacts on programs and services, and our understanding of what people are actually doing

• Or, the data are trusted too much, given all of the pitfalls—THIS IS THE PART WHICH WE CAN REALLY HELP WITH!
DATA CREATION AND MAINTENANCE: RECAP!

• Data don’t come out of nowhere—they represent a phenomenon, and have to be collected! (either by humans or machines, and machines aren’t perfect)

• There are many points at which the data can be “compromised”—there’s no perfect source, you just have to know the caveats!

• This is why documentation is so important! Also, user forums, and hands-on experience

• Maintenance is a big deal too—most users want data at regular intervals, in consistent formats, and with consistent fields (questions). This takes time and money and the will to continue!

• Often older data sets are not offered online, but usually the owner can supply them (at least in government)
HOW DATA ARE ANALYZED

• I hate to say it, but, it depends! On what?
  • The data file
  • The question being answered
  • The analyst

• Usually structured data (that falls into rows and columns) is analyzed using some software; Excel can do a lot but statistical packages can do more

• In research, generally there is a starting hypothesis that is tested, but, you can just explore the data too (do descriptive work rather than statistical analysis)
QUANTITATIVE VS QUALITATIVE

• The two work together- qualitative can inform quantitative
• Qualitative data is stronger than anecdote, and can help you persuade someone that an anecdote is true
• Can be directed at specific groups and questions
• Can get at issues that quantitative data doesn’t
• Can incorporate cultural difference
• Combining stories with quantitative data gives a human face to the issue
• Qualitative data analysis methods are very sophisticated!
POPULATION DATA ARE IDEAL, BUT WHAT ABOUT SURVEYS?

• Surveys let us use a sample of the population to represent the whole---
• This saves money!
• Data can be collected more quickly!
• Since time and money are saved, can ask more detailed questions
• Often can be more customized (languages, geographic areas covered, etc.)
• Data can be collected about questions of immediate interest, trends
• BUT—Survey data analysis is REALLY complex; you can’t just throw it in Excel! For large data sets, you pretty much need training and data savvy to use it in its raw form
VisuAlization: the next frontier

- This is a huge topic. And I only have two slides.
- Visualization is a way to understand the data better, and show patterns so that others do too.
- There are different types of visuals that are best for showing different aspects of the data:
  - Line graphs are great for timelines.
  - Pie charts can be good if the percentages add up to 100.
  - Bar charts can be good for comparing groups.
  - Maps are good if there is a spatial component.
- Colors, labels and other aspects help make points as well!

#1 Draw a Venn diagram

It doesn’t matter if your Venn diagram is wildly inaccurate, in fact, the more inaccurate the better.

What we don’t know

What we know

the sweet spot

Even before you’ve put that marker down, your colleagues will begin fighting about what exactly the labels should be and how big the circles should be, etc.

At this point, you can slink back to your chair and go back to playing Candy Crush.

TheCooperReview.com, used with permission
• Check out this site: https://eazybi.com/blog/data_visualization_and_chart_types/

• Check out the work of Edward Tufte, if you haven’t already

• Check out some of the sites that let you make infographics, like Piktochart—you can combine different visualizations to make a story

• Pay attention to visualizations in publications that are great at them—the New York Times, the Wall Street Journal, and others

• LISTEN TO SALLY GORE! http://tinyurl.com/sallygoredataviz
ANSWERING PATRON QUESTIONS AROUND DATA: YOU ALREADY KNOW HOW TO DO THIS!

• We already answer questions using data! Many tools on the internet that deliver answers are doing data analysis behind the scenes.

• We are comfortable with classification and ordering of information—this is a fundamental data skill.

• The ways we attack problems can translate neatly to data questions:
  • We do a careful reference interview
  • We evaluate sources for currency, bias, authority and reliability
  • We know how to search to find out what’s already been done rather than recreating the wheel.

#3 ENCOURAGE EVERYONE TO “TAKE A STEP BACK”

There comes a point in most meetings where everyone is chiming in, except you. This is a great point to go, “Guys, guys, guys, can we take a step back here?”

Follow it up with a quick, “What problem are we really trying to solve?” and, boom! You’ve bought yourself another hour of looking smart.

TheCooperReview.com, used with permission
DATA VS. STATISTICS

• Data is the raw information (such as, an Excel file with the information from all the birth certificates).

• Statistics are what comes out when an analysis/data interpretation has been done (such as, a report on birth weights by county, perhaps in a visualization such as a bar chart)

• Consider whether you and/or your patron really need to mess around with the data, or, can you find reports or tables online that are ready-made?

• This relates to the question of primary and secondary data—it’s just like primary and secondary sources! Do you really have to collect data yourself, or has someone done it already?
DATA PITFALLS, OR, HOW TO BE A CRITICAL DATA CONSUMER

• Look for limitations; for example, there may be bias if not everyone is represented and the lack of representation might be systematic:
  • People of color
  • People whose first language is not English
  • Low income people
  • Homeless people
  • People without landlines (at least in the olden days)

• It can be hard to get enough respondents to look at small geographic areas (the “small numbers” issue; some statistical methods can help, but not totally!)

• Geographies can be tricky (zip codes don’t match census tracts, and more!)
MORE PITFALLS TO AVOID!

(Note: These are for social science data, but the same critical eye is needed for any data!)

Look at the survey instrument
- Wording of questions makes a big difference
- Order of questions makes a big difference
- Certain types of questions are problematic (example—"recall bias")

Remember the birth certificate example about how people answer questions
- They want to please the interviewer
- They don’t like sharing sensitive information
- Women and older people are more likely to say yes even among randomly chosen people
- Many people are not comfortable answering questions online
FINDING THE PERFECT DATA SET (OR NOT?)

• Look for published statistics and/or secondary data
• Ask “who cares?” What organization might have already collected data?
• Have a sense of which data are open and which are proprietary
• Look in the literature for similar instances of dealing with the topic; what data sets did they use?
• Check out LibGuides, user forums, and other places data might be discussed
• Consider data archives and repositories (if you are game to analyze data once you get it)
• Ask a data librarian! 😊
SOME TRICKS

• Do a REEEEEELY good reference interview, so that you are looking for the right thing! Need a refresher? Check this out: https://libguides.lib.fit.edu/ReferenceTraining

• It’s OK to ask people to hone in! http://www.bpl.org/contact/examples.htm

• Remember your own databases (ReferenceUSA, others related to finance/business, weather, environment, education/social welfare, health care, etc.)

• Maximize Google! You know Site:.gov; now try http://www.powersearchingwithgoogle.com/ and https://www.process.st/google-search-tips/

• RIP, Internet Public Library 😞 See this page http://www.ipl.org/div/faq/ for lists of most-asked reference questions, with links to potential answers

• “How to Find Data & Statistics: Finding Data”—LibGuide from Michigan State University (primary data focused, but some great suggestions)

• Who cares? Use association pages and list-servs!
WHAT IF YOU AREN’T FINDING DATA ON THE TOPIC?

• Give data for the larger area, then show proportions that apply to your neighborhood/city/smaller area
• Use data for a larger group, and then compare your group to the larger group
• Use substitutions (such as, teen births if teen pregnancies aren’t available)
• Internet (start with Google, but don’t stop there!)
• Advocate for better collection of data sources of interest
• Consider gathering qualitative data
• Consult experts (in fact, they may do a special data run for you!)
BEST PRACTICES AS YOU USE DATA (WITH PATRONS OR NOT!)

• ALWAYS look at the documentation!
• Compare what you find to published sources where possible
• Consult with others who have used the source, or the software
• Consult an expert (data requests!)
• Prioritizing—try to focus on 2 or 3 crucial facts to make an argument
• Comparing—consistent units, correct frame of reference, eliminate bias
• Use maps and graphics AS APPROPRIATE (it isn’t always!)
• Build a case with several data points—tell a story (infographics can be good here)
• Watch out for technical aspects such as size of files (open data files can get BIG)
BEST PRACTICES FOR ENGAGING PATRONS

• Narrate your process as you work
• Explain why you need each piece of information from them
• Emphasize that you are exploring together
• Explain why some data are not available openly
• Suggest adjustments to their question that may make data easier to find (warning—this often does not go over well!)
• See if they can use estimates (may be easier to find in other sources online!)
• Encourage them to follow best practices such as consulting the data dictionary, comparing with published sources
• Suggest tutorials as a refresher if it’s been a while since they’ve done data work
• Know when to refer to an expert
WHAT ABOUT COMPUTER SKILLS, NUMERACY, CHALLENGES OF COMPLEX DATA ANALYSES?

• There’s no way to make this easy. Data can be tough to work with, no matter what your level. The main goal is avoiding shaming, and offering support as possible
• You can’t assume the patron’s level of competency or understanding—just keep checking in on how they are receiving the information; are they repeating it back to you or sitting silently? Asking for clarification? Etc.
• Suggest other library or online classes that may be helpful either before or while they delve into the data; Excel is an important tool
• Offer ongoing assistance, as possible
• Offer options for professional assistance (This also does not go over well! But if someone wants a complex analysis of raw data for market research, and they can’t perform it, they may have to pay for it!)
WHY A PATRON MIGHT ASK FOR OPEN DATA

• They know there’s an issue in their community but need to prove it
• To get funding
• To affect policy
• To advocate for their community or organization
• To question what someone else is saying (fake news)
• To decide what programs they need to plan
• To direct you to next steps (remember- data can create more questions!)
QUESTIONS (REAL!) PATRONS MAY ASK

• How many times on average were fake news links shared compared with real news?
• How can I find affordable hospitals? Health care providers?
• How can I find out about education levels by neighborhood?
• How many grants are available for women business owners?
• What is the average price of X product/service globally/locally?
• What were the voting patterns by county in 2016?
• Which cities are dedicated to funding services that are important to me? (schools, library, public transit, etc.)
• What is the participant usage of my city’s bikeshare program in a given month?
• Where is the best location to look for a job in X field?
• Which nursing homes in my area have passed audits?
• How can I get information about first time home buyers?
• I want to know about universities’ financial aid and graduation data.
CONNECTIONS TO EXPLORE (PROGRAM IDEAS?)

• Between data and numeracy
• Between data and fake news
• Between data and advocacy
• Between data and coding (Hour of Code, Girls Who Code)
• Between data and business reference
• Between data and intellectual freedom/privacy
• Between the data you wanted (but isn’t available) and other data sources you can use (finding data is easier for large areas!)
• Between your library assessment data and available open data
TRAINING FOR THE DATA-SAVVIER

• School of Data https://schoolofdata.org/
• Data Journalism Handbook http://datajournalismhandbook.org/1.0/en/index.html
• Open Data Institute https://theodi.org/courses
• Open Data Handbook http://opendatahandbook.org/resources/
• Watch for “Data Equity for Main Street” curricula, coming in 2018! https://ocio.wa.gov/open-data/open-data-your-library

(Also there are many online courses related to research data management—contact me if you want to know more)
TRAINING FOR THE DATA BEGINNER

• Coursera, edX and other MOOCs (the levels may vary!)
• Lynda.com if available through your library
• May work best to find basic statistics or even numeracy/math courses rather than starting with data, such as https://onlinecourses.science.psu.edu/statprogram/review_of_basic_statistics or https://www.ipracticemath.com/learn/basicmath
• Or look for resources for specific topic areas, such as Basic Data Analysis for Health Programs https://www.measureevaluation.org/resources/training/capacity-building-resources/basic-data-analysis-for-health-programs
• And again, the “Data Equity for Main Street” curriculum is coming soon! 😊
WASHINGTON-SPECIFIC OPEN DATA RESOURCES

• NW Data https://nwdata.org/#!data
• State of Washington Open Data https://data.wa.gov/ (Will Saunders!)
• OFM https://ofm.wa.gov/washington-data-research
• OSPI http://www.k12.wa.us/DataAdmin/default.aspx
• Department of Revenue Business License Lookup https://secure.dor.wa.gov/gteunauth/_
• Washington State GIS Data https://wagda.lib.washington.edu/data/geography/wa_state/
• MSRC http://mrsc.org/Home/Explore-Topics.aspx
• Puget Sound Regional Council https://www.psrc.org/data-and-resources/data-psrc
• And so many more!
SHAMELESS SELF-PROMOTION

If you want a lot of this content in article format, check out my two articles in Alki: The Journal of the Washington Library Association (http://www.wla.org/alki-archives):


• “Safety in Numbers: Helping People with Health Numeracy Challenges (Which is All of Us)” (November, 2016)
QUESTIONS?
CONTACT ME!

Ann Glusker
Research & Data Coordinator
National Network of Libraries of Medicine, Pacific Northwest Region
nnlm.gov/pnr

Glusker@uw.edu
206-543-5112