June 7, 2019

Video 2: Three Strategies for Finding Data

Jylisa Doney
University of Idaho, Social Sciences Librarian
jylisadoney@uidaho.edu

Transcript

00:00 Hi everyone, welcome to Video 2.
00:02 In this video, I'm going to share
00:04 three strategies for finding data
00:06 and highlight a few data resources.
00:08 Strategy 1 is to brainstorm
00:11 potential producers and creators.
00:13 Some of the main producers of data
00:15 on a variety of topics are government agencies.
00:18 One top data creator in the United States
00:21 is the Census Bureau.
00:22 The Census Bureau conducts more than
00:24 100 surveys across the United States every year
00:27 to collect data about people and the economy.

[United States Census Bureau: https://www.census.gov]

00:30 In Video 3, I'm going to demonstrate how we can use
00:34 data.census.gov, a Census Bureau tool, to find data.
00:38 Another example is the Bureau of Labor Statistics.
00:42 The BLS collects, analyzes, and disseminates
00:45 data related to labor market activity,
00:48 working conditions, and price changes in the economy.

[Bureau of Labor Statistics: https://www.bls.gov]

00:51 Non-governmental organizations,
00:53 non-profits, research centers,
00:55 and specific colleges and universities
00:58 also produce and collect data.
01:00 A few examples include
01:02 NORC at the University of Chicago,
01:04 which is one of the largest
01:05 independent social research organizations
01:08 in the United States.
01:09 NORC is responsible for the General Social Survey.

[NORC at the University of Chicago - The General Social Survey: https://gss.norc.org]

01:12 Pew Research Center, which conducts research
01:15 related to social issues, public opinion,
01:17 and demographic trends.

[The Pew Research Center: https://www.pewresearch.org]

01:19 And World Bank, which collects
01:21 and makes available global development data.

[The World Bank: https://data.worldbank.org]

01:24 Lastly, we may be able to identify
01:27 top researchers in an area of interest
01:29 who conduct similar research
01:30 and make their data available.
01:32 This strategy would be most helpful
01:34 if our topic was specific enough to warrant
01:37 reusing a data set created by other researchers.
01:40 A good example could be
01:41 if you wanted to validate and replicate
01:43 previous research findings.
01:45 If we are unable to brainstorm
01:47 relevant data creators or producers,
01:49 we can often use statistics to identify these creators,
01:53 which in turn can lead us to the raw data.
01:56 When using this strategy, it is important to understand
01:59 how statistics are different from raw data.
02:02 Statistics are the result of
02:03 data analysis, interpretation, or evaluation.
02:07 They often answer "how much" or "how many."
02:09 This first example
02:10 clearly represents a statistical visualization
02:13 as it shares the outcome of some type
02:15 of data analysis and interpretation.
02:18 This next example is still statistics,
02:20 but the information has undergone
02:22 less transformation than the previous example.
02:25 Since this table includes the actual
02:27 percentages associated with each variable,
02:29 we could potentially transform this information
02:32 into a different type of visualization.
02:35 In comparison, raw data is the information
02:38 created or collected during research.
02:40 Data can be analyzed and interpreted
02:42 to create new knowledge
02:44 and is the information from which
02:45 statistics are created.
02:47 And remember, as we talked about in Video 1,
02:50 raw data can appear as aggregate data or microdata.
02:54 Compared to the statistics screenshots,
02:57 this image is definitely microdata
02:59 as it presents individual records.
03:02 However, we can also see that these data
03:04 have been simplified or cleaned
03:06 because we can understand the column headings
03:08 and cell values at first glance.

03:11 If we compiled and analyzed this data,
03:13 we could create a statistical output,
03:16 such as a chart or visualization.
03:18 This last screenshot
03:19 is also an example of raw microdata
03:21 and includes the exact same data
03:23 that we saw on the previous slide.
03:25 However, we would need to find
03:27 more information about the variables
03:29 and potential answer options
03:31 to fully understand this data.
03:33 This type of information is usually available
03:35 in the codebook or data dictionary.
03:37 Getting back to Strategy 2,
03:40 a few ways that you can use statistics
03:42 to identify data producers include,
03:44 reviewing the statistics and charts
03:46 included in publications and using the
03:48 associated citations to identify the creators,
03:51 investigating recent media coverage of a topic
03:54 and determining where they found this information,
03:56 or conducting a Google Image search for your topic
03:59 to identify potentially relevant visualizations on the web.
04:02 The third strategy is to identify relevant
04:05 data archives or repositories.
04:07 Data archives and repositories are
04:09 platforms that host data.
04:11 Usually the platform is not the
04:12 party that collected the data,
04:14 they are just making it available for others.
04:17 Some repositories host data
04:19 across disciplines and institutions,
04:21 while others only include data sets
04:23 that are related to each other in some way.
04:25 For example, the data may be from researchers
04:28 or projects in a specific discipline,
04:30 like political science.
04:31 While some universities
04:33 may even have a data archive or repository
04:35 that only includes data sets
04:37 from their own faculty or students.
04:39 Most often, data archives and repositories
04:42 can be searched by anyone,
04:44 but some do require visitors to create
04:46 a free account download data.
04:48 However, some data archives
04:50 and repositories restrict access
04:52 and only allow members of specific groups
04:55 or those who have paid for access
04:56 to search for and download data.
04:58 One repository example is data.gov,

05:01 which provides public access to data sets
05:04 created by the government
05:05 at the federal, state, local, or tribal levels.

[Data.gov: https://www.data.gov]

05:08 Another example is Harvard Dataverse,
05:10 which is an open repository of research data
05:13 across disciplines and institutions.

[Harvard Dataverse: https://dataverse.harvard.edu]

05:15 Both data.gov and Harvard Dataverse are free,
05:19 and can be searched by anyone,
05:20 but they may restrict who can download sensitive data
05:23 or require visitors to create an account
05:26 A unique example is IPUMS,
05:29 which has various sub-repositories
05:30 based on different topics
05:32 and provides access to census and survey data
05:34 from around the world.

[IPUMS: https://www.ipums.org]

05:36 IPUMS is free and can be searched by anyone,
05:39 but does require visitors
05:40 to create an account to download the data.
05:43 In Video 4, I'm going to demonstrate
05:45 how we can use IPUMS-NHGIS
05:48 to find aggregate geographic data.
05:51 If these repositories don't include
05:52 the data you're looking for,
05:54 you can also visit two websites,
05:56 which link to archives and repositories
05:58 from a variety of disciplines.
06:00 These websites are,
06:01 the Open Access Directory and re3data.

[Open Access Directory – Data Repositories: http://oad.simmons.edu/oadwiki/Data_repositories; re3data: https://www.re3data.org/]

06:04 Each archive or repository linked to
06:07 within these resources
06:08 will have different requirements for visitors
06:10 to search for and download data.
06:13 Coming up in Video 3, I'm going to demonstrate
06:16 how to use one of the Census Bureau's tools
06:18 to find microdata.