HOW DO I FIND DATA?
Select a topic and work toward defining your research question. When working on data-based projects, it’s helpful to ask four key questions:

- **What?**
- **Who?**
- **Where?**
- **When?**
For example, to investigate the topic of political polarization, you need to make a few decisions. What kind of polarization are you investigating?

Are you interested in polarization of the general public, polarization in selected red and blue states, or polarization in Congress?
Who collects the data you need?

Some candidates include government agencies, academic researchers, businesses, non-profits, and intergovernmental or non-governmental organizations.
What time period will you investigate?

You may be interested in finding out how long polarization has been an issue or you may want to look at a shorter period of time.
"When" can be tricky. It takes time to collect and publish data, so in some cases, the most current datasets can be several years old. Also, datasets gathered this year may not have been collected 10, 20, or 50 years ago. Investigate the datasets available in earlier years before proceeding.
What about geography?

Is there a geographic component to your research project? Are you looking for national, regional, state, or local data?
The Library's discipline-based Research Guides provide a good starting point for identifying resources.
Many Library Research Guides provide access to data and statistical resources. To start our investigation of political polarization in the U.S. population, we'll check the Data, Statistics, and Polls tab on the Political Science Guide.
Scholarly articles often contain reference to datasets and polls. In this library database search for materials on political polarization, we've added two terms data and polling to find scholarly articles with this type of content.

Note that we've typed data or polling to indicate that we're interested in articles on polarization with data or polling or both.
In this example, a scholarly article on political polarization references the General Social Survey, a recurring survey produced by the National Opinion Research Center.
Identify datasets from tables and charts

In this example taken from the Statistical Abstract of the United States, we’re directed to the American National Election Studies for additional detail.

Table 404. Democratic and Republican Percentages of Two-Party Presidential Vote by Selected Characteristics of Voters: 2004 and 2008

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>2004 Democratic</th>
<th>2004 Republican</th>
<th>2008 Democratic</th>
<th>2008 Republican</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of birth:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975 or later</td>
<td>66</td>
<td>54</td>
<td>65</td>
<td>56</td>
</tr>
<tr>
<td>1969 to 1974</td>
<td>45</td>
<td>55</td>
<td>56</td>
<td>44</td>
</tr>
<tr>
<td>1949 to 1958</td>
<td>44</td>
<td>56</td>
<td>54</td>
<td>46</td>
</tr>
<tr>
<td>1943 to 1942</td>
<td>51</td>
<td>49</td>
<td>41</td>
<td>59</td>
</tr>
<tr>
<td>1911 to 1926</td>
<td>52</td>
<td>48</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>1895 to 1910</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sex:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46</td>
<td>54</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>Female</td>
<td>53</td>
<td>47</td>
<td>57</td>
<td>46</td>
</tr>
</tbody>
</table>

= Represents zero. 1 Includes other characteristics, not shown separately.
Source: American National Election Studies, [http://www.electionstudies.org/].
Identify datasets through data archives

- ICPSR (Inter-university Consortium for Political and Social Research)
- Data.gov
- National Science Foundation Data
- Centers for Disease Control
- National Archive of Data on Arts and Culture
- United Nations Data
ASK A LIBRARIAN

library.northeastern.edu/ask