1214 Reporting quantitative data

November 4, 2008

(with thanks to Gary Kass and Edward Tufte)
Why graphs?

Ex: The Harris Poll #100, Oct 15, 2007

American adults have their choice of many different cuisines ... From pasta to burritos, the choices are almost endless. But what do Americans choose when eating out? American food, of course!

Over one-quarter (28%) of U.S. adults say American food is what they are most likely to choose if they had the choice to go out to a restaurant and eat one type of food. Just under one-quarter (22%) say that they would most likely choose Italian while 17 percent would choose Mexican while 16 percent would choose Chinese if they had the choice to go out to a restaurant. Japanese is the next choice as seven percent say they would choose this type of cuisine. Much further down the list are Indian (2% say they would choose) and then French and Middle Eastern cuisine (1% would choose each). Finally, four percent say they would choose another type of food.
TABLE 1
TYPE OF FOOD FOR EATING OUT

"Thinking of food now, if you had the choice to go out to a restaurant and eat one type of food, which of these are you most likely to choose?"

Base: All Adults

<table>
<thead>
<tr>
<th>Region</th>
<th>Total</th>
<th>East %</th>
<th>Midwest %</th>
<th>South %</th>
<th>West %</th>
<th>Echo Boomers (18-30) %</th>
<th>Gen Xers (31-42) %</th>
<th>Baby Boomers (43-61) %</th>
<th>Matures (62+) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>American</td>
<td>28</td>
<td>29</td>
<td>27</td>
<td>33</td>
<td>21</td>
<td></td>
<td>21</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>Italian</td>
<td>22</td>
<td>31</td>
<td>21</td>
<td>18</td>
<td>20</td>
<td></td>
<td>23</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Mexican</td>
<td>17</td>
<td>9</td>
<td>19</td>
<td>16</td>
<td>27</td>
<td></td>
<td>17</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>Chinese</td>
<td>16</td>
<td>14</td>
<td>21</td>
<td>15</td>
<td>17</td>
<td></td>
<td>18</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Japanese</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td>6</td>
<td></td>
<td>12</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Indian</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td></td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Middle-Eastern</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>*</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
<td>3</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>None of these</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>*</td>
<td>2</td>
<td>1</td>
<td>*</td>
</tr>
</tbody>
</table>

Note: Percentages may not add up to 100% due to rounding.

Note: * = less than 0.5%
SE9. How confident do you feel about your own searching abilities when using a search engine to find information online?

**Based on Internet users who have a used a search engine**

\[ N = 1,165 \]

<table>
<thead>
<tr>
<th>%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>Very confident</td>
</tr>
<tr>
<td>44</td>
<td>Somewhat confident</td>
</tr>
<tr>
<td>6</td>
<td>Not too confident</td>
</tr>
<tr>
<td>2</td>
<td>Not confident at all</td>
</tr>
<tr>
<td>*</td>
<td>Don’t know/Refused</td>
</tr>
</tbody>
</table>

http://www.pewinternet.org/pdfs/PIP_Search_Questions.pdf
### Tabular displays

Comparing dependent to independent variables

<table>
<thead>
<tr>
<th></th>
<th>Undergrads %</th>
<th>Grads %</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=120</td>
<td>n=200</td>
<td>n=320</td>
</tr>
<tr>
<td>Satisfied</td>
<td>60</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>N=71</td>
<td>N=25</td>
<td>N=96</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>40</td>
<td>87</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>N=47</td>
<td>N=165</td>
<td>N=127</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>n = 118</td>
<td>n = 190</td>
<td>n = 308</td>
</tr>
<tr>
<td>No ans.</td>
<td>n = 2</td>
<td>n = 10</td>
<td>n = 12</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grads</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>No ans.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Good information graphics

**Efficient**
minimizing non-data ink

**Meaningful**
displaying information relevant to the argument

**Unambiguous**
Self-explanatory, or at least difficult to misread


5 principles from E. Tufte

Above all else show the data
Maximize the data-ink ratio
Erase non-data-ink, within reason
Erase redundant data-ink
Revise and edit
Components of a chart

**labeling**
- Title
- axis titles and labels
- legends
- notes

**scales** of X, Y axis

**graphic representations** of data
Area chart

How the dependent relationship of parts of a whole varies over an independent variable.

Percentage: sum of the two always represents 100%

Cumulative: sum varies according to elements
Column or bar

Change within small data sets that vary over a nominal (ie, category) or interval scale

Column: vertical rectangles

Bar: horizontal
Segmented column

A column chart that represents the part-whole relationships of elements

II. SURVEY RESULTS

1. Internet Use: What Percentage, How Much, When, Where, What Activities and with Whom?

1a. Percentage: Among 4,384 respondents 18 to 64 years old, 33.6% used the Internet on the day before being surveyed. This result is an increase of 2.2 percentage points compared to last year. Over the past three years, the percentage of Internet users has been fairly stable (around 32%). (See Graph 1)

Frequency polygon; Histogram

Overall shape represents distribution of elements

Probability of Democrats Controlling At Least This Many Senate Seats**

**Counting Sanders (I-VT), Lieberman (I-CT)
Line charts

Used to interpolate between data points and extrapolate future values
Trends or change over time
Looking for correlations between variables

Lines: Connect "real" data points
Curves: represent functional relations between data points or to interpolate data
X must have numerical values!

Pie chart

Good for displaying approximate percentages at a single point in time

Caution:
- Makes exact comparisons difficult
- You cannot necessarily compare pie charts to each other
- Maybe use a bar chart instead?
Scatterplot

**One dimensional**
Show measurements over time

**Two-dimensional**
Convey an overall impression of the relation between two variables

Some caution:
Not good for following trends
Should not be used with more than one independent variable
### Analysis

Looking for **significant** differences across **important** groups

Looking for **correlation** – inferring causality
Maps

Placing data on physical space

World Within New York

World Within New York shows how different neighborhoods reach out to the rest of the world via the AT&T telephone network. The city is divided into a grid of square pixels where each pixel is colored according to the regions of the world wherein the top connecting cities are located. The heights of the color bars represent the proportion of world regions in contact with each neighborhood. Encoded within each pixel is also a list of the top ranking world cities that account for the communications with that particular area of New York.
Representing change with small multiples

Edward Tufte

Small multiples
“Graphical depictions of variable information that share context, but not content.”

Edward Tufte


graphic: www.uxmatters.com/MT/archives/000044.php
Table 1: Opinion on food
Are you satisfied or dissatisfied with the bagels at the Free Speech Movement Café?

<table>
<thead>
<tr>
<th></th>
<th>Undergrads</th>
<th>Grads</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td>n=120</td>
<td>n=200</td>
<td>n=320</td>
</tr>
<tr>
<td>Satisfied</td>
<td>60</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>N=71</td>
<td>N=25</td>
<td>N=96</td>
</tr>
<tr>
<td>Dissatisfied</td>
<td>40</td>
<td>87</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>N=47</td>
<td>N=165</td>
<td>N=127</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>n = 118</td>
<td>n = 190</td>
<td>n = 308</td>
</tr>
<tr>
<td>No ans.</td>
<td>n = 2</td>
<td>n = 10</td>
<td>n = 12</td>
</tr>
</tbody>
</table>

The reader should be able to reconstruct what you did.

Show the question

Include all data ie, percentages and numbers of respondents

Needs and Usability Assessment 11.04.08
Reporting: last words

Graphics are visual metaphors.
They are rhetoric.

Do you know the story you want to tell?

...the story your audience is seeing?
Appendix A:
Good places to look for example reports

Pew Internet
http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf pp. i-iv
http://www.pewinternet.org/pdfs/PIP_Search_Questions.pdf
"What Do Americans Do on the Internet?“ Appendix B

Harris Interactive Poll
http://www.harrisinteractive.com/harris_poll/

Stanford Time Study
http://www.stanford.edu/group/siqss/SIQSS_Time_Study_04.pdf