Telling Stories With Data: Graphing Change Over Time

Andy Grogan-Kaylor

April 08, 2020

1 Graphs

We start in thinking about graphing change over time with a scatterplot. A natural next step is to connect the dots of a scatterplot with straight line segments to form a line plot.

Instead of simply connecting the observations, one may estimate an individual linear trajectory. In multilevel modeling these line plots showing individual estimated linear trajectories are sometimes called spaghetti plots.

Alternatively, rather than connecting observations with straight lines, or estimating an overall straight line trajectory for each individual, it may be useful to smooth the trajectories by drawing curved lines between individual observations.

1 Scatterplots show every data point. However, with many data points, scatterplots may become overcomplicated, and difficult to interpret. Points may even be plotted over other data points.

2 Note that we are using color and line type to distinguish different individuals. This may not always be possible, especially when there are a large number of individuals in the data.

3 With any of the options discussed, one may consider small multiples where each individual trajectory is placed in its own sub-graph.

4 One needs to be careful, however, as the smoothed trajectories may give the impression of having more data points than one actually has.
An increasingly popular option is a slope graph.\(^5\)

In order to be clear and effective, a slope graph may often only show the outcome at the beginning point, and at the end point. A slope graph may be less satisfactory when there are multiple timepoints. The small multiple idea works with a slopegraph as well.

\(^5\) The Data Used In This Example Are Simulated. Many data sets, but not all, are originally created in the wide format— as shown below— where every row of data is an individual, and an individual only has a single row. Ideally, every row in wide data is uniquely identified by an individual id number.

<table>
<thead>
<tr>
<th>id</th>
<th>outcome.1</th>
<th>outcome.2</th>
<th>outcome.3</th>
<th>outcome.4</th>
<th>outcome.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Generally, for graphing change over time, it is most appropriate to have data that are in a long format, as shown in the margin. In long data every row represents a particular measurement occasion for a particular individual. Each individual in the data set thus has multiple rows. Ideally, every row in data in the long format is uniquely identified by the combination of an id number and a study wave.

Data can be reshaped from wide to long format, and vice versa. Two straightforward options are the reshape command, as available in both Stata and R.

Graphics made with ggplot2 created by Hadley Wickham. "Graphing Change Over Time" by Andrew Grogan-Kaylor is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.