Graphing Change Over Time

Andy Grogan-Kaylor

October 24, 2018

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. 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.\textsuperscript{4}

\begin{center}
\begin{tikzpicture}
  \begin{axis}[
    title={slopegraph},
    xlabel={time},
    ylabel={outcome},
  ]
    \addplot coordinates {
      (1,1)
      (2,5)
      (3,4)
      (4,3)
      (5,0)
    };
    \addplot coordinates {
      (1,2)
      (2,4)
      (3,2)
      (4,2)
      (5,1)
    };
    \addplot coordinates {
      (1,3)
      (2,3)
      (3,2)
      (4,4)
      (5,0)
    };
    \addplot coordinates {
      (1,4)
      (2,3)
      (3,4)
      (4,2)
      (5,1)
    };
    \addplot coordinates {
      (1,5)
      (2,5)
      (3,1)
      (4,1)
      (5,3)
    };
    \addplot coordinates {
      (1,5)
      (2,4)
      (3,2)
      (4,3)
      (5,4)
    };
  \end{axis}
\end{tikzpicture}
\end{center}

The Data Used In This Example Are Simulated.

Many data sets, but not all, are originally created in the \textit{wide} format—as shown below—where every row of data is an \textit{individual}, and an individual only has a \textit{single row}. Ideally, every row in \textit{wide} data is uniquely identified by an individual \textit{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 \textit{long} format, as shown in the margin. In \textit{long} data every row represents a particular \textit{measurement occasion} for a \textit{particular individual}. Each individual in the data set thus has \textit{multiple rows}. Ideally, every row in data in the \textit{long} format is uniquely identified by the combination of an \textit{id} number and a \textit{study wave}.

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

Long Data

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

\textsuperscript{4}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.

\begin{center}
\begin{tikzpicture}
  \begin{axis}[
    title={slopegraph},
    xlabel={time},
    ylabel={outcome},
  ]
    \addplot coordinates {
      (1,1)
      (2,5)
      (3,4)
      (4,3)
      (5,0)
    };
    \addplot coordinates {
      (1,2)
      (2,4)
      (3,2)
      (4,2)
      (5,1)
    };
    \addplot coordinates {
      (1,3)
      (2,3)
      (3,2)
      (4,4)
      (5,1)
    };
    \addplot coordinates {
      (1,4)
      (2,3)
      (3,4)
      (4,2)
      (5,1)
    };
    \addplot coordinates {
      (1,5)
      (2,5)
      (3,1)
      (4,1)
      (5,3)
    };
    \addplot coordinates {
      (1,5)
      (2,4)
      (3,2)
      (4,3)
      (5,4)
    };
  \end{axis}
\end{tikzpicture}
\end{center}

Graphics made with ggplot\textsuperscript{5} created by Hadley Wickham.

Designed using the Tufte-La\TeX \textsuperscript{6} variant of RMarkdown which is inspired by the works of Edward Tufte\textsuperscript{7}.

\textit{Graphing Change Over Time} by Andrew Grogan-Kaylor is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

\textsuperscript{5}http://ggplot2.org/
\textsuperscript{6}https://code.google.com/p/tufte-latex/
\textsuperscript{7}http://www.edwardtufte.com/tufte/books_be