Graphing Change Over Time
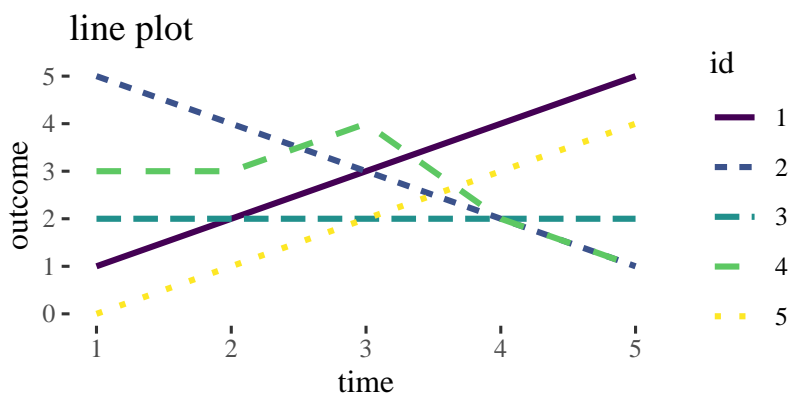
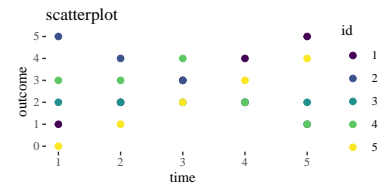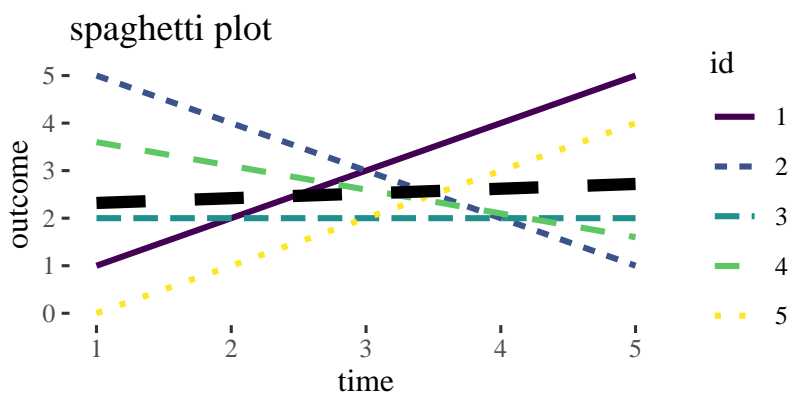*Andy Grogan-Kaylor*

December 09, 2019

Graphs

We start in thinking about graphing change over time with a scatterplot.[1] [2]

A natural next step is to connect the dots of a scatterplot with straight line segments to form a line plot. [3]

scatterplot

[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.
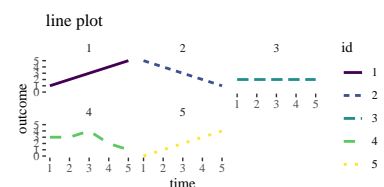
line plot

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*.
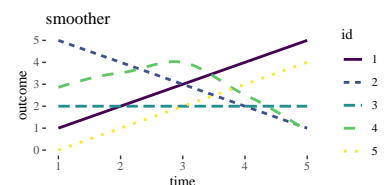
line plot

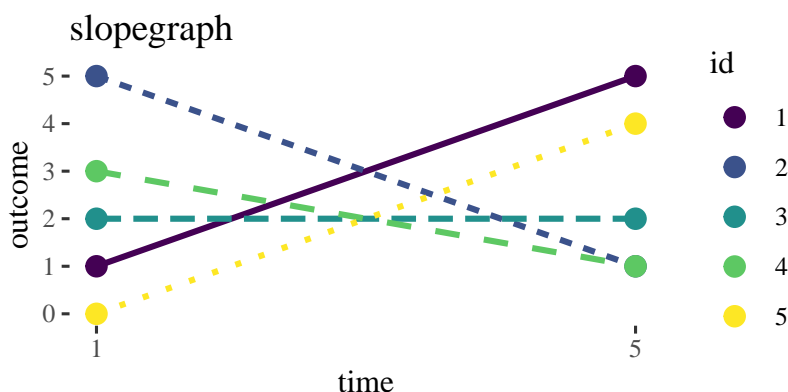Alternatively, rather than connecting observations with straight lines, or estimating an overall straight line trajectory for each individual, it may be useful to "smooth" the trajectories by drawing curved lines between individual observations.
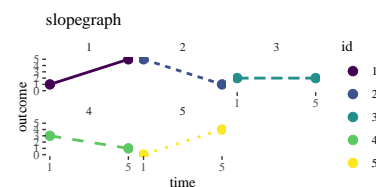
smoother

spaghetti plot

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.[4]

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.

id	outcome.1	outcome.2	outcome.3	outcome.4	outcome.5
1	1	2	3	4	5
2	5	4	3	2	1
3	2	2	2	2	2
4	3	3	4	2	1
5	0	1	2	3	4

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[5] created by Hadley Wickham.

Designed using the Tufte-LaTeX [6] variant of RMarkdown which is inspired by the works of Edward Tufte[7].

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

[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.

Long Data

id	t	outcome
1	1	1
1	2	2
1	3	3
1	4	4
1	5	5
2	1	5
2	2	4
2	3	3
2	4	2
2	5	1
3	1	2
3	2	2
3	3	2
3	4	2
3	5	2
4	1	3
4	2	3
4	3	4
4	4	2
4	5	1
5	1	0
5	2	1
5	3	2
5	4	3
5	5	4

[5] http://ggplot2.org/

[6] https://code.google.com/p/tufte-latex/

[7] http://www.edwardtufte.com/tufte/books_be