Telling Stories With Data: Graphing Change Over Time

Andy Grogan-Kaylor

2023-11-22

Table of contents

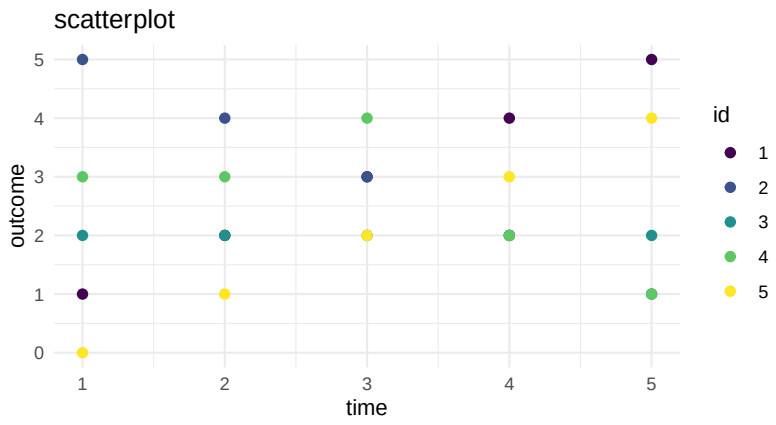
1 Graphs

1.1 Scatterplot

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

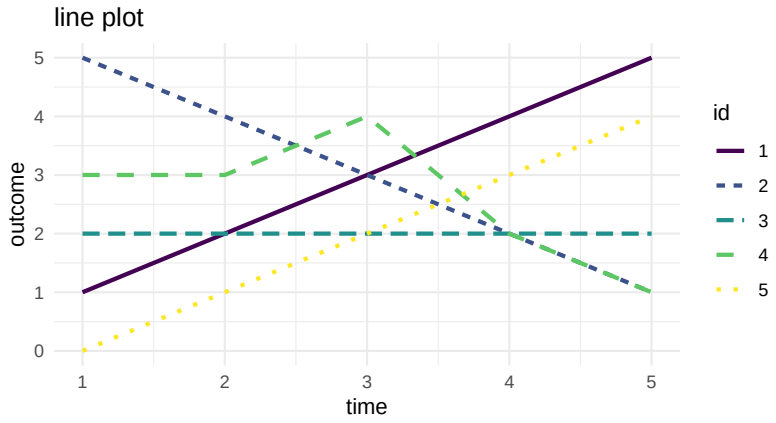
[1] Scatterplots show every data point. However, with many data points, scatterplots may become overcomplicated, and difficult to interpret. Points may even be plotted over other data points.

[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.

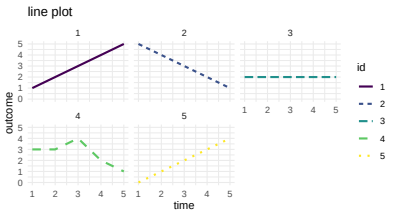


1.2 Line Plot

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

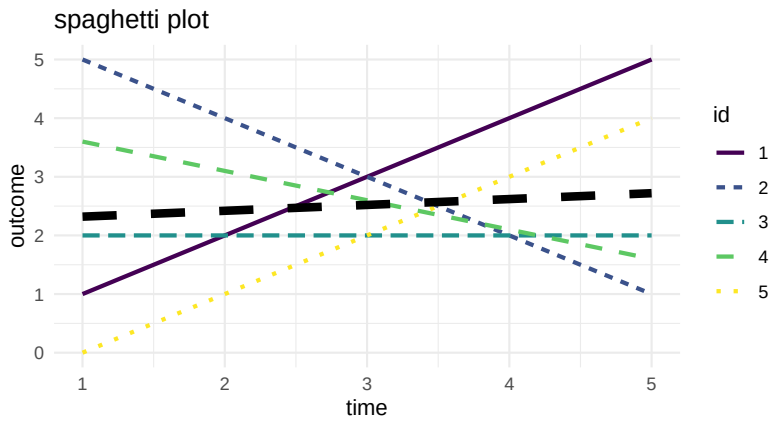


[3] With any of the options discussed, one may consider *small multiples* where each individual trajectory is placed in its own sub-graph.



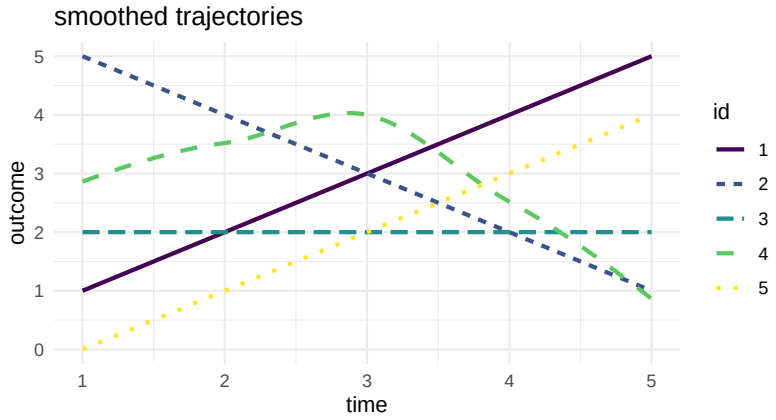
1.3 Spaghetti Plot

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*.

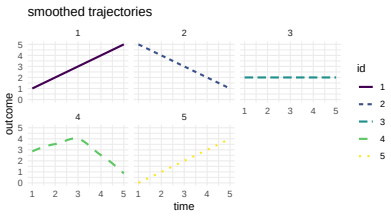


1.4 Smoothed Trajectories

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.[4]

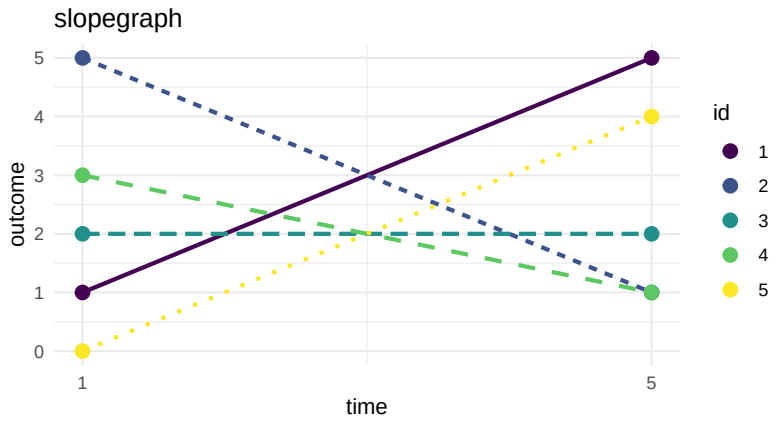


[4] One needs to be careful, however, as the smoothed trajectories may give the impression of having more data points than one actually has.

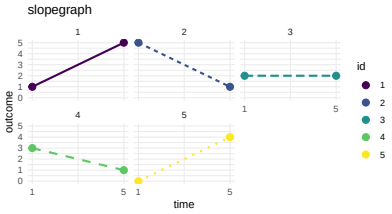


1.5 Slopegraph

An increasingly popular option is a slope graph.[5]



[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.



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

Figure 1: Long Data

2 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 ggplot2created by Hadley Wickham.