Visualizing Multilevel Models

Andy Grogan-Kaylor

2023-11-17

Table of contents

1 Introduction **2**

2 Organizing Questions **2**

3 Setup **3**

4 Get Data **3**

5 Scatterplots (`twoway scatter y x`) **3**

6 Simple Linear Fit (`twoway lfit y x`) **4**

7 Linear Fit With Confidence Interval (`twoway lfitci y x`) **4**

8 Combine Scatterplot and Linear Fit (`twoway (scatter y x) (lfit y x)`) **5**

9 Spaghetti Plots (`spagplot y x, id(group)`) **7**

10 Small Multiples (`twoway y x, by(group)`) **7**

11 Taking A Random Sample **8**

12 Multivariate (Predicted) Relationships **9**

12.1 Using Predicted Values (`predict`) 9

12.1.1 Estimate The Model 10

12.1.2 Generate Predicted Values 10

12.1.3 Graph With `twoway` Syntax 11

12.1.4 Spaghetti Plot With Predicted Values 12

12.2 `margins` and `marginsplot` 12

12.2.1 Estimate The Model 13

12.2.2 Generate Predicted Values *At Specified Values* With `margins` 13
12.2.3 Graph With `marginsplot` 14

13 Scatterplot With Linear Fit and Marginal Density Plots (`twoway ...`) 14
13.1 Manually Generate The Densities To Plot Them Below (`kdensity ...`) 15
13.2 Rescale The Densities So They Plot Well 15
13.3 Make The Graph (`twoway ...`) 16
13.4 Spaghetti Plot With Linear Fit and Marginal Density Plots 16

14 Curvilinear and Linear Fits 16

15 Random Effects 16

1 Introduction

An evolving set of notes on visualizing results from multilevel models.

The examples below use the `simulated_multilevel_data.dta` file from *Multilevel Thinking*. Here is a direct link to download the data.

2 Organizing Questions

Try to think about some of the advantages and disadvantages of different approaches to visualizing multilevel models. In multilevel models, we don't want to just *control for* variation, but to start to *explore* the variation. Put concretely:

- Some approaches use *dots*. Some approaches use *lines*. Some approaches use *dots and lines*.
- Some approaches use the *raw unadjusted* data. Other approaches use *adjusted or model predicted* data.
- Some approaches attempt to show the *Level 2 specific regression lines*; some approaches only show an *average regression line*.
- What approaches might work well with *large numbers* of Level 2 units? What approaches might work well with *smaller numbers* of Level 2 units?

What approach(es) do you prefer?

3 Setup

I am not terrifically fond of the default `s2color` graph scheme in earlier versions of Stata. Here I make use of the `michigan` graph scheme available at: https://agrogan1.github.io/Stata/michigan-graph-scheme/

```
set scheme michigan
```

Stata's `s1color` scheme would also would be an option as would be Asjad Naqvi's incredible `schemepack`: https://github.com/asjadnaqvi/stata-schemepack.

Throughout the tutorial, I make frequent use of the `mcolor(%30)` option to add some visual interest to scatterplots by adding transparency to the markers.

4 Get Data

```
use "https://github.com/agrogan1/multilevel-thinking/raw/main/simulate-and-analyze-multile
```

5 Scatterplots (`twoway scatter y x`)

```
twoway scatter outcome warmth, mcolor(%30)

graph export myscatter.png, width(1500) replace
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

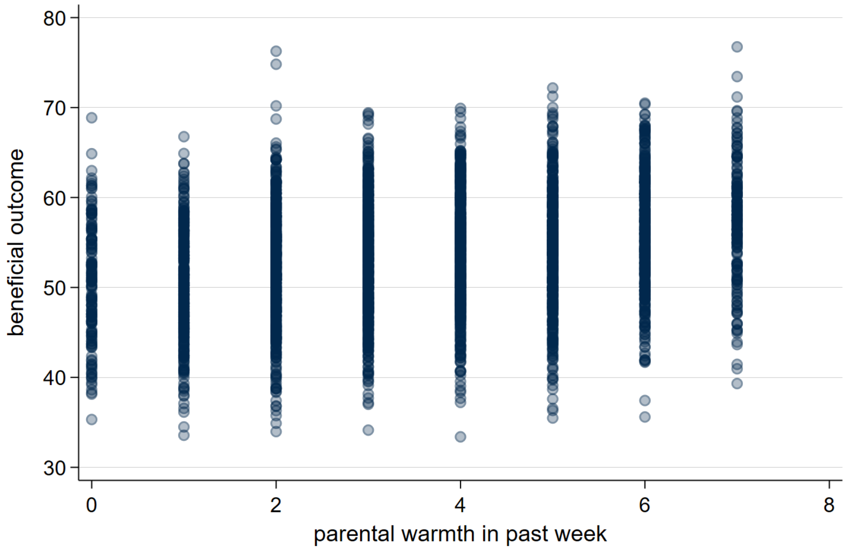

Figure 1: Scatterplot

6 Simple Linear Fit (`twoway lfit y x`)

```
twoway lfit outcome warmth

graph export mylinear.png, width(1500) replace
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

7 Linear Fit With Confidence Interval (`twoway lfitci y x`)

```
twoway lfitci outcome warmth

graph export mylfitci.png, width(1500) replace
```

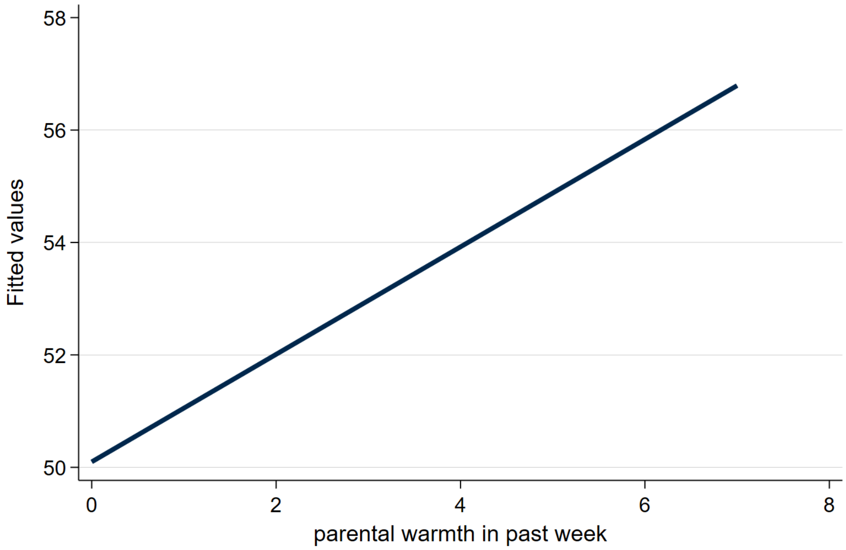

Figure 2: Linear Fit

```
variable outcome not found
r(111);

end of do-file
r(111);
```

8 Combine Scatterplot and Linear Fit (`twoway (scatter y x) (lfit y x)`)

```
twoway (scatter outcome warmth, mcolor(%30)) (lfit outcome warmth)

graph export myscatterlinear.png, width(1500) replace
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

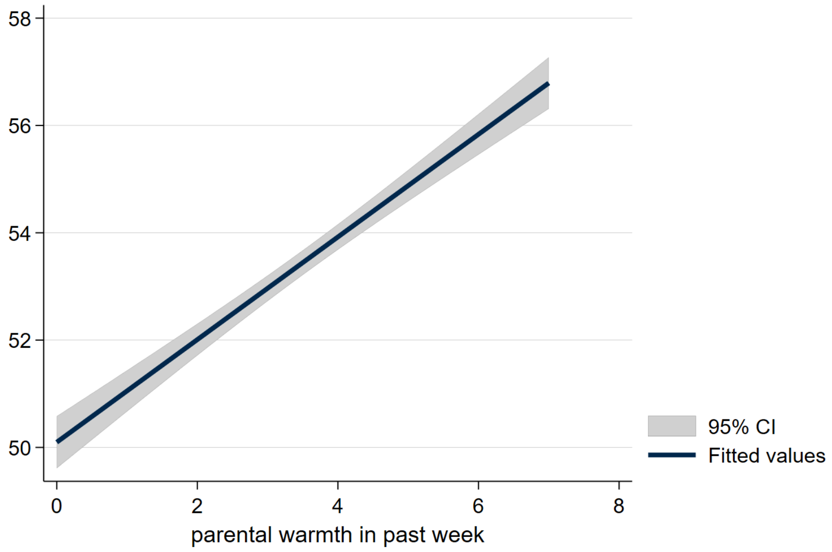

Figure 3: Linear Fit With Confidence Interval

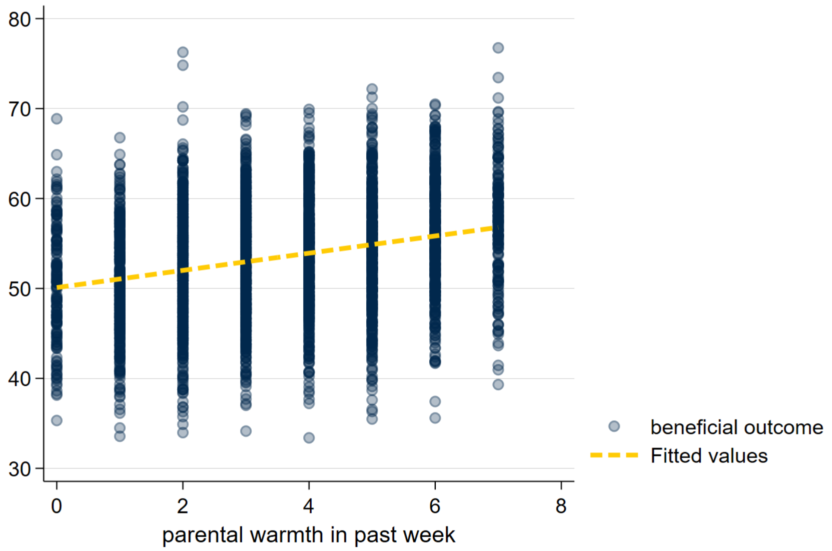

Figure 4: Scatterplot and Linear Fit

9 Spaghetti Plots (`spagplot y x, id(group)`)

```
spagplot outcome warmth, id(country)

graph export myspaghetti.png, width(1500) replace
```

```
no variables defined
r(111);

end of do-file
r(111);
```

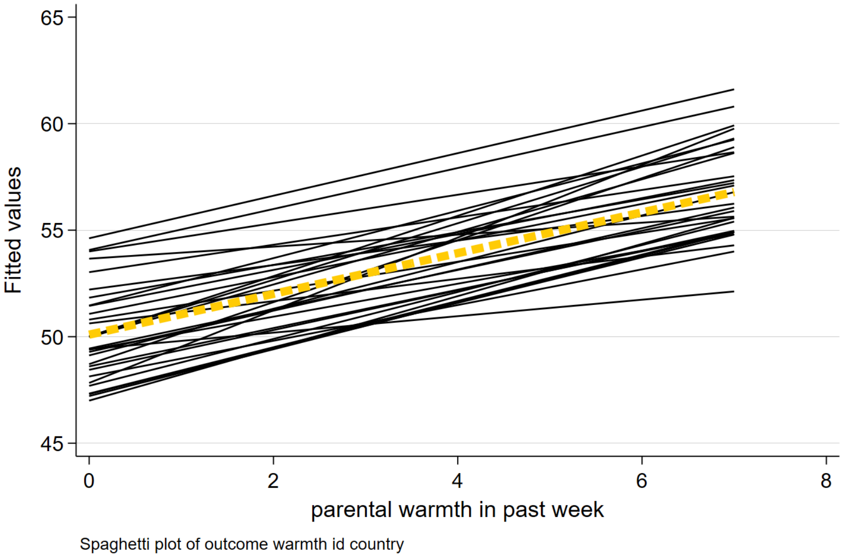

Figure 5: Spaghetti Plot

10 Small Multiples (`twoway y x, by(group)`)

Small Multiples, showing a separate graph for each group in the data, are an increasingly popular data visualization technique. Below, I build a small multiples graph using the `by` option in Stata. I use the `aspect` option to adjust the *aspect ratio* of the graph for better visual presentation.

```
twoway (scatter outcome warmth, mcolor(%30)) ///
(lfit outcome warmth), ///
by(country) aspect(1)

graph export mysmallmultiples.png, width(1500) replace
```

```
no variables defined
r(111);

end of do-file
r(111);
```

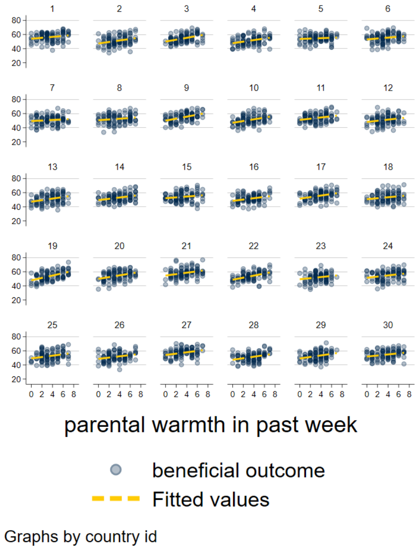

Figure 6: Small Multiples

11 Taking A Random Sample

At times, we may have *too many* Level 2 units to effectively display them on a *spaghetti plot*, or using *small multiples*. If this is the case, we may need to *randomly sample* Level 2 units. This can be difficult to accomplish as our standard `sample` command operates on each row, or on Level 1 units.

We can accomplish random sampling at Level 2, with a little bit of code.

```
set seed 3846 // random seed for reproducibility

gen randomid = runiform() // generate a random id variable

* by country (i.e. by Level 2 unit) replace the randomid
* with the first randomid for that country (Level 2 unit)
* so that every person in that country has the same random id

bysort country: replace randomid = randomid[1]

summarize randomid // descriptive statistics for random id

twoway (scatter outcome warmth, mcolor(%30)) /// scatterplot
(lfit outcome warmth) /// linear fit
if randomid < .5, /// only use a subset of randomids
by(country) aspect(1) // by country

graph export mysmallmultiples2.png, width(1500) replace
```

```
variable country not found
r(111);

end of do-file
r(111);
```

12 Multivariate (Predicted) Relationships

A sometimes unacknowledged point is that graphs–unless we take steps to correct this–reflect *unadjusted*, or *bivariate* associations. We may sometimes wish to develop a graphs that reflect the *adjusted* or *predicted* estimates from our models.

In multilevel models, *prediction* is a complex question. Prediction may–or may not–incorporate the information from the random effects. The procedures below outline graphs that incorporate predictions using the random effects, by using the `predict ..., fitted` syntax.

12.1 Using Predicted Values (`predict`)

`predict` generates a predicted value for *every observation in the data*.

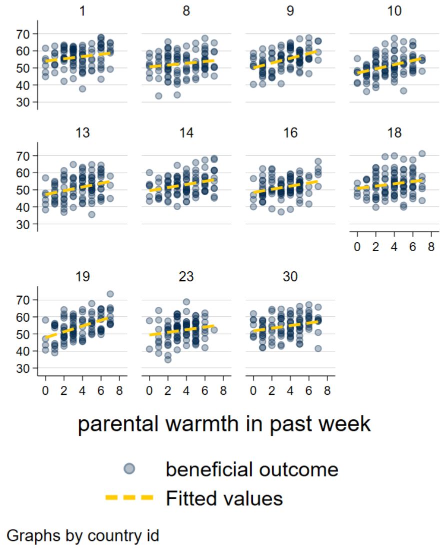

Figure 7: Small Multiples With A Random Sample Of Countries

12.1.1 Estimate The Model

```
mixed outcome warmth physical_punishment i.group || country: // estimate MLM
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

12.1.2 Generate Predicted Values

```
predict outcome_hat, fitted // predict yhat (`fitted` uses fixed AND random effects)
```

```
last estimates not found
r(301);

end of do-file
r(301);
```

12.1.3 Graph With twoway Syntax

```
twoway (scatter outcome_hat warmth, mcolor(%30)) (lfit outcome_hat warmth)

graph export mypredictedvalues.png, width(1500) replace

twoway (lfit outcome_hat warmth)

graph export mypredictedvalues2.png, width(1500) replace
```

```
variable outcome_hat not found
r(111);

end of do-file
r(111);
```

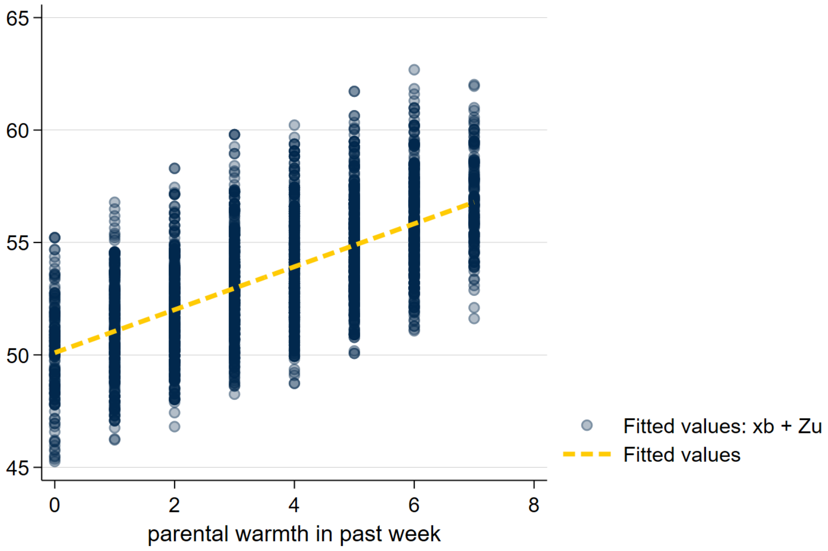

Figure 8: Predicted Values From predict

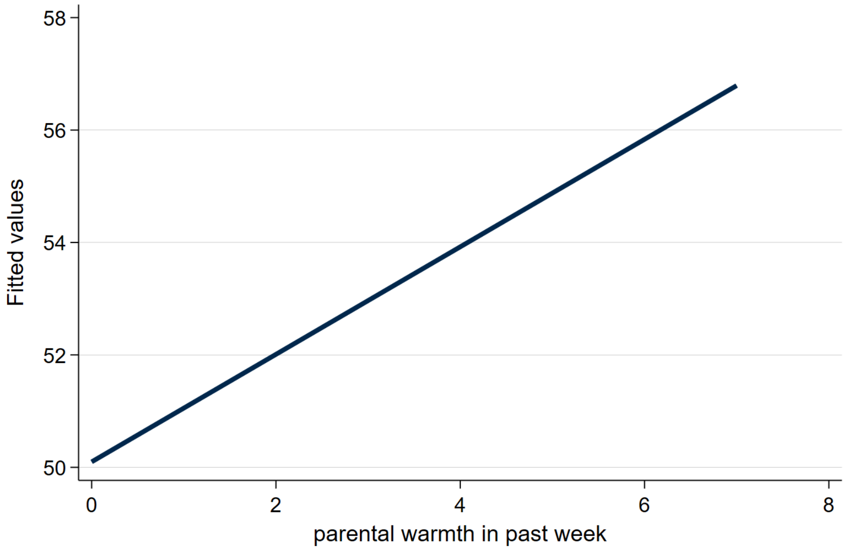

Figure 9: Predicted Values From `predict` With Only Linear Fit

12.1.4 Spaghetti Plot With Predicted Values

```
spagplot outcome_hat warmth, id(country)

graph export myspaghetti2.png, width(1500) replace
```

```
no variables defined
r(111);

end of do-file
r(111);
```

12.2 `margins` and `marginsplot`

In contrast to `predict`, which generates a predicted value for *every observation in the data*, `margins` generates predicted values at *specific values of certain variables.*

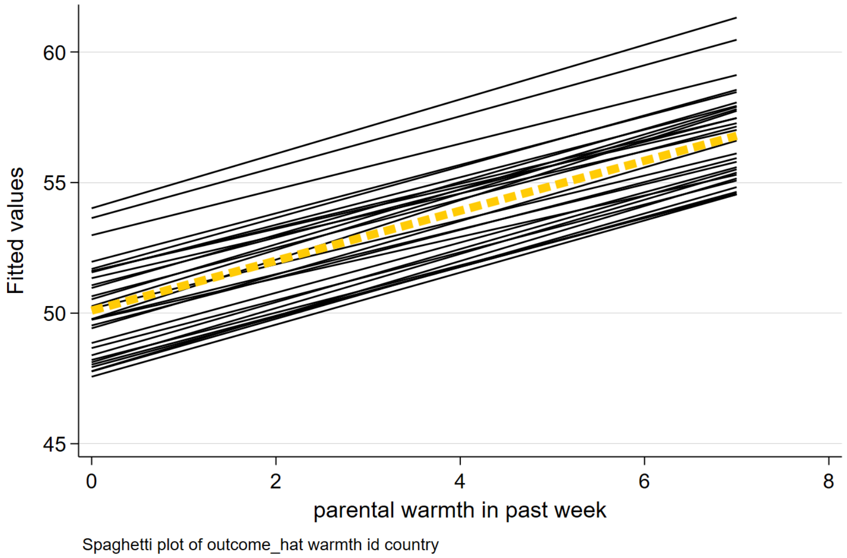

Figure 10: Spaghetti Plot With Predicted Values

12.2.1 Estimate The Model

```
mixed outcome warmth physical_punishment i.group || country: // estimate MLM
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

12.2.2 Generate Predicted Values *At Specified Values* With `margins`

```
margins group, at(warmth = (1 2 3 4 5 6 7)) // predictive *margins*
```

```
last estimates not found
r(301);

end of do-file
r(301);
```

12.2.3 Graph With `marginsplot`

```
marginsplot // plot of predicted values

graph export mymarginsplot.png, width(1500) replace
```

```
previous command was not margins
r(301);

end of do-file
r(301);
```

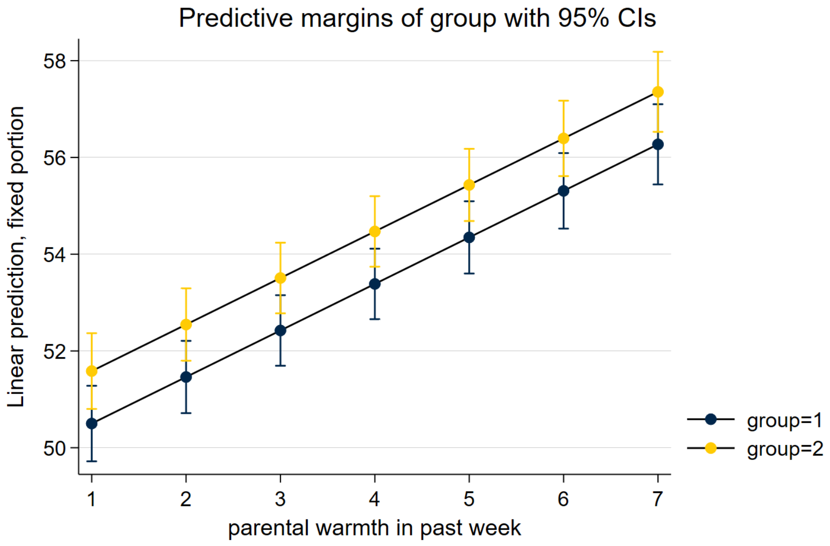

Figure 11: Predicted Values From `margins` and `marginsplot`

13 Scatterplot With Linear Fit and Marginal Density Plots (`twoway ...`)

As another possibility, we may wish to show more of the variation, by showing the variation in the *independent* variable and the *dependent* variable along with a *scatterplot* and *linear fit*. This is a complex graph and requires a little bit of manual programming in Stata.

You could also investigate the user written program binscatterhist (ssc install binscatterhist) which produces a similar looking graph, and automates much of this work.

13.1 Manually Generate The Densities To Plot Them Below (kdensity ...)

We generate the density for *warmth* at only a few points (n(8)) since this variable has relatively few categories.

```
kdensity warmth, generate(warmth_x warmth_d) n(8) // manually generate outcome densities

kdensity outcome, generate(outcome_y outcome_d) // manually generate outcome densities
```

```
variable warmth not found
r(111);

end of do-file
r(111);
```

13.2 Rescale The Densities So They Plot Well

You may have to experiment with the scaling and moving factors.

```
replace warmth_d = 100 * warmth_d // rescale the density so it plots well

replace outcome_d = 5 * outcome_d - .5 // rescale AND MOVE the density so it plots well

label variable outcome_y "density: beneficial outcome" // relabel y variable
```

```
no variables defined
r(111);

end of do-file
r(111);
```

13.3 Make The Graph (`twoway ...`)

You may have to experiment with whether scatterplots or line plots work best for displaying the x and y densities.

```
twoway (scatter outcome warmth, mcolor(%10)) /// scatterplot w some transparency
(lfit outcome warmth) /// linear fit
(line warmth_d warmth_x) /// line plot of x density
(line outcome_y outcome_d), /// line plot of y density (note flipped order)
title("Outcome by Warmth") /// title
ytitle("beneficial outcome") /// manual ytitle
xtitle("parental warmth") /// manual xtitle
legend(position(6) rows(2) ) /// legend at bottom; 2 rows
xlabel(0 1 2 3 4 5 6 7) /// manual x labels
name(mynewscatter, replace)

graph export mynewscatter.png, width(1500) replace
```

```
variable outcome not found
r(111);

end of do-file
r(111);
```

13.4 Spaghetti Plot With Linear Fit and Marginal Density Plots

14 Curvilinear and Linear Fits

15 Random Effects

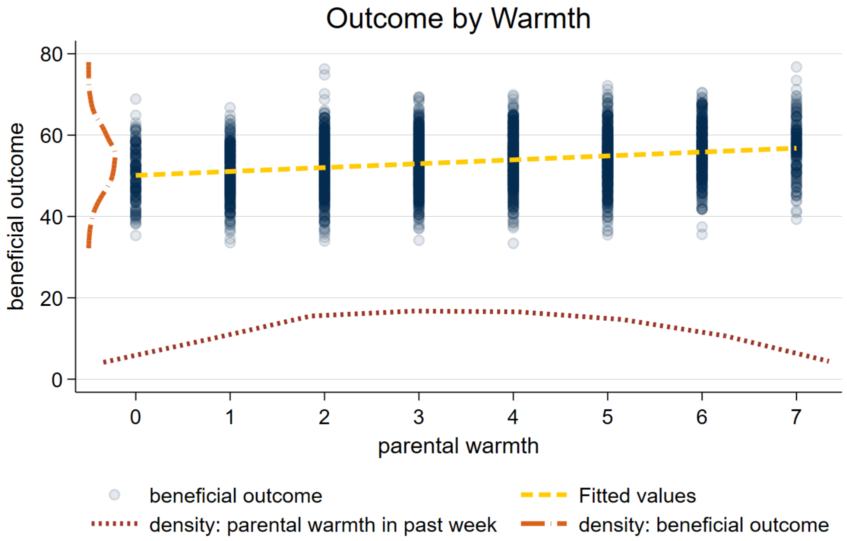

Figure 12: Scatterplot and Linear Fit With Marginal Density Plots