Investigations and Findings on Visualizing Multilevel Modeling Data

Ni Bei, Yawen Chiu, Linghui Chu, Matthew J. Davidson,
Melissa Gasser, Wm Salt Hale, Ussama Ahmad Khan,
Youngwon Kim, Anwesha Pan & Megan Ramaiya

Introduction

Multilevel Modeling (MLM) is commonly used in education, psychology, sociology and other fields to deal with data collected from groups or multiple observations, sharing features or demographic characteristics. MLM focus on analysis of data with complex patterns of variability as well as nested sources of such variability (e.g., longitudinal measurements of subjects, students in classes, classes in schools) (Snijders, T. A., & Bosker, R. J., 2011). Other names including hierarchical linear modeling, mixed linear modeling, growth-curve modeling also belong to methods in purpose of analyzing multilevel data (Dedrick, R., Ferron, J., Hess, M., Hogarty, K., Kromrey, J., Lang, T., . . . Lee, R.,2009).

MLM can handle various types of hierarchical structures in one unified framework, with factor either fixed (e.g. gender difference) or random (sample to population). The design of MLM is dependent on data, can be cross-sectional or with longitudinal structure if there is time nested within an individual. According Glaser, D., & Hastings, R. (2011), handling missing data and imbalanced designs is one distinct advantage of MLM. Currently, the most common way to represent MLM data is a small multiples approach. Often this is accomplished via a series of faceted panels. This can become unwieldy and difficult to interpret, especially when there are a large number of categories for more than one of the levels.

Figure 1. A public health example (by Chris Adolph, homicidePres - slide46.png "small multiple facets example 2")

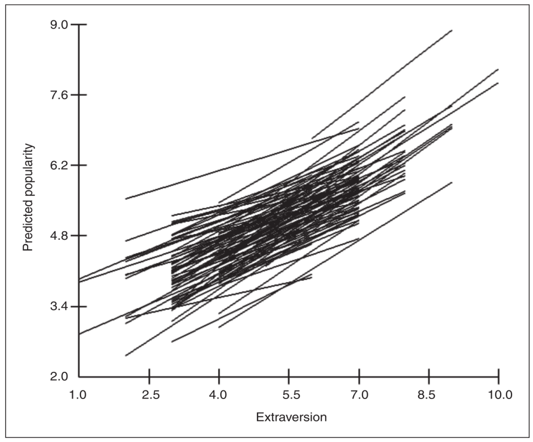

Figure 13.6 Plot of the 100 class regression slopes for pupil extraversion.

Figure 2. R Graphics-CRC Press (2018) (page10.png "Small multiple facets example 1")

Another would be to plot a series of regression lines on the same graph, but this, too, can quickly become visually overwhelming. These are occasionally referred to as "spaghetti" plots due to the difficulty of excessive overlap.

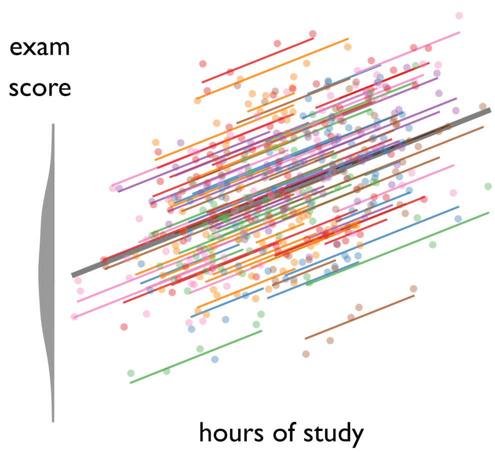

Figure 3. Basic concepts for hierarchical_panel data models (by Chris Adolph, with example of HLM visuals from Gelman et al - topic5 - slide41.png "Regression lines example 1")

Using different glyphs, colors, sizes, or similar features can help to differentiate levels, but these can often cause additional cognitive load and require further explanation.

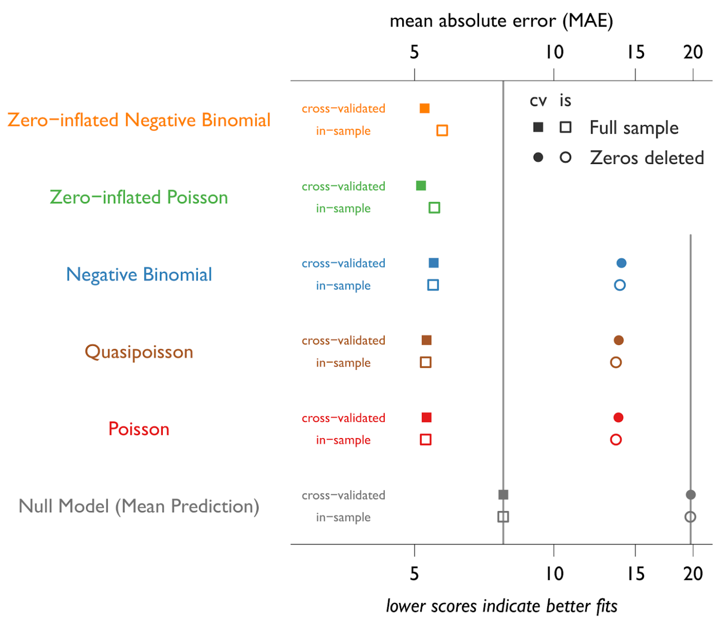

Figure 4. Visualization ideas for zero-inflated models (by Chris Adolph, topic6.p - slide190.png "Different features example 1")

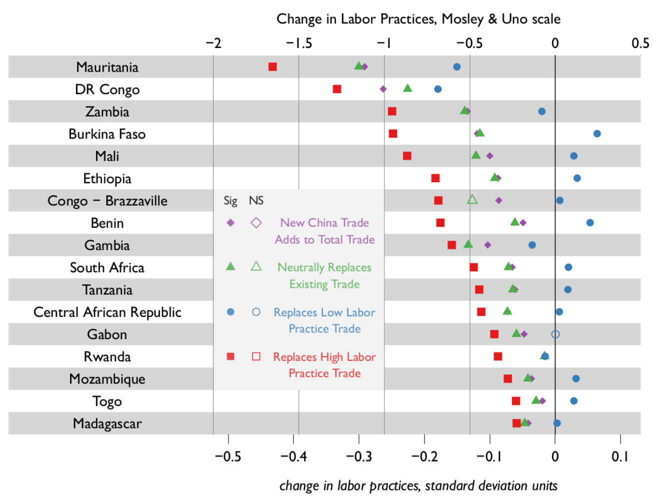

Figure 5. Visuals simulated from a hierarchical panel setup (by Chris Adolph, Paper - aqpAfrica - page32.png "Different features example 2"

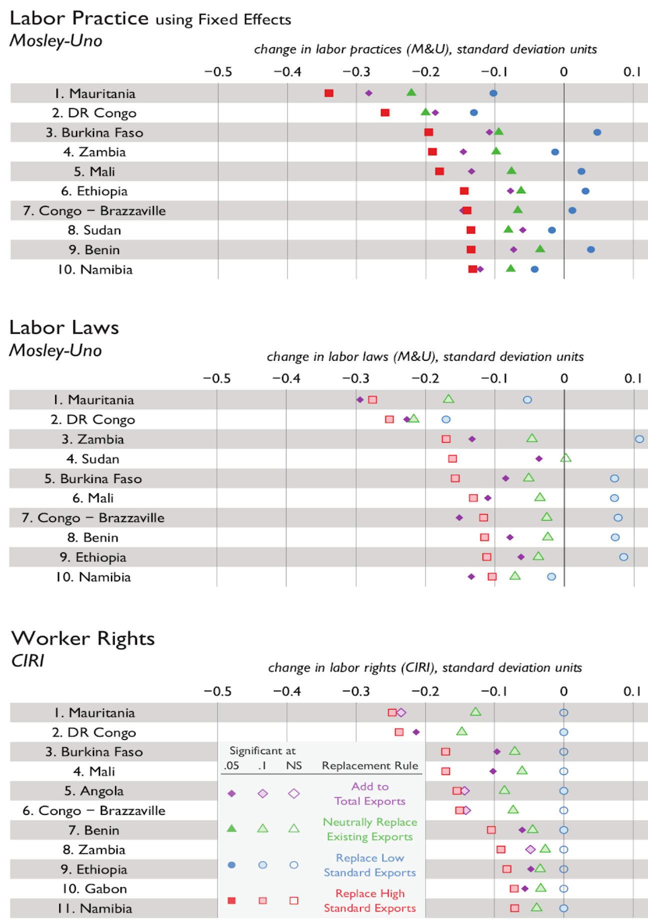

Figure 6. Visuals simulated from a hierarchical panel setup (by Chris Adolph, Paper - aqpAfrica - page32.png "Different features example 3"

There are a number of other presentations and the [Centre for Multilevel Modelling (CMM)](http://www.bristol.ac.uk/cmm/) had provided an excellent search tool (https://www.cmm.bris.ac.uk/gallery/) to discover papers that have utilized MLM techniques.

Lessons Learned

Purpose of Visualization

The purpose of visualization is to help us better understand the patterns of the data, as well as the results of our analysis. Instead of plotting all the information we have, it is important to know what we want to show in the plot. Consider grouping or aggregating objects or variables according to research interests and questions could help to simplify the plot. Gelman et al. (2005)'s paper provides some examples of visualizing the result of a multilevel logistic regression to better explain the results. They fitted a multilevel logistic regression to study the relationship of income to individual vote preferences. According to the results, richer voters support Republicans within states and overall; though there are differences between states. To better visualize the confusing pattern they found, they constructed a graph that simultaneously displays variation within and between states. As shown in their figure 3, they assumed intercepts vary by state. In figure 4, they assumed slopes and intercepts to vary across the states. From figure 4, we could see a systematic pattern of the within-state slope between states. That is, Mississippi (poorest state) has steepest slope and Connecticut (richest state) with a shallowest slope, representing the income is a strong predictor of vote preference in Mississippi, and a relatively weaker predictor in Connecticut. The authors clearly explained their findings through the plots and provided us examples of how to include variation within and between states through the plots.

On the other hand, visualizing confidence interval in line plots allows us to assess the uncertainty of the patterns we found. To visualize the confidence interval in a multilevel context, one way is to simulate the expected values and plot the confidence interval using “simcf” and “tile” or “ggplot2” packages in R. Another way is to use “ggpredict” function in R. “ggpredict” helps to computes predicted values for the outcome variable. However, there might be some differences based on these two ways of generating the plots.

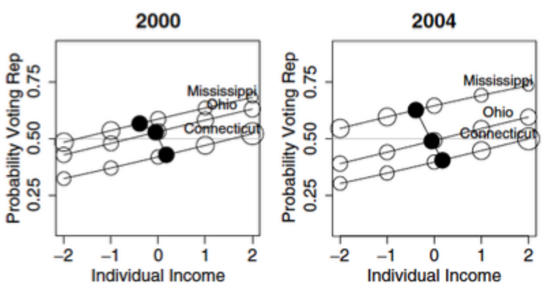

Figure 3

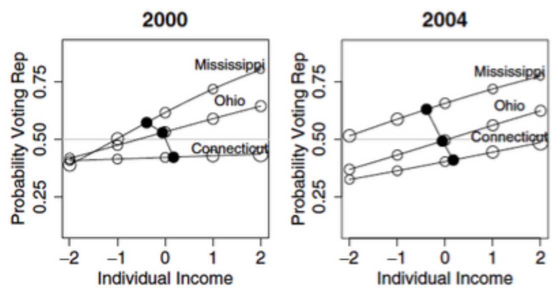

Figure 4

Figure 7. Visualizing confidence intervals in R

Zero-Inflated Negative Binomial (ZINB) Regression Models

A special case occurs when creating visualizations of zero-inflated negative binomial (ZINB) regression models. The need for this approach often arises in health data, such as individuals’ typical number of drinks per week or heavy episodic drinking episodes per month. Though we will be discussing these recommendations using count data examples such as these, our guidance may be applied more broadly. ZINB models can be understood as two parts: a logistic/zero-inflated component and a counts component. Through considering these two parts as drawn from different distributions, such as individuals who never drink or never engage in heavy episodic drinking, we may choose to model the findings through two figures representing these respective components. Whether or not to include both of these components in the final product for distribution, however, should be determined by the central research questions. In this example, the research question may concern risk of alcohol use disorders and thus the individuals who typically drink at all in a week are the focus as in Figure A (Lindgren et al., 2015). Alternatively, we may be interested in displaying the sampled population in its entirety. When considering the added dimension of MLMs, identifying, simplifying, and potentially reducing the number of visualizations is even more vital to efficiently and effectively communicate findings.

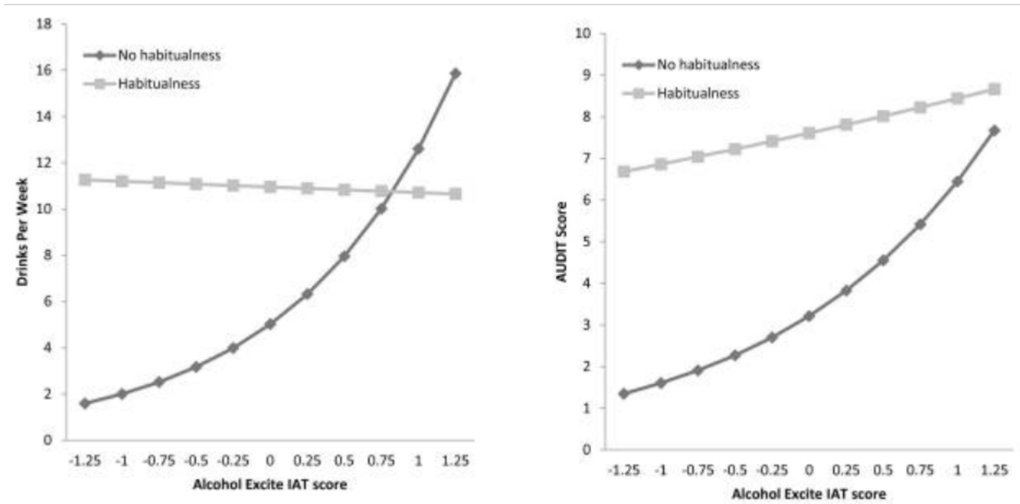

Figure 8. Alcohol Excite IAT x Drinks Per Week image (Lindgren et al., 2015)

Visualizing Results and/or Model Predictions

There are a number of possible approaches to visualizing model results and model predictions for multilevel models. Perhaps unsurprisingly, many of these approaches are the same ones you might use for a single-level model. As in the single-level case, it's important to be clear what you actually need to visualize: are the model coefficients of particular interest to your readers? Could model-based predictions communicate the results effectively? What aspect(s) of the model directly address your research questions? For example, is the predictive utility of the entire set of covariates of interest, or is the effect of a small set of covariates on the DV of more interest? Depending on your answers to the previous questions, you may choose to use a visualization that does not even require the reader to know that you used a multilevel model!

A commonly used visualization in multilevel growth models is the “spaghetti plot”, where a line is plotted for each case in the dataset, illustrating that intercepts and/or slopes are random, as shown in Figure 9. These plots show the model predictions for many cases in the dataset, but do not facilitate comparisons between interesting quantities. Possible alternative approaches: choose a small set of cases that illustrate bigger trends in the data; look for latent classes in the data based on slopes/intercepts, and describe those; construct interesting counterfactual scenarios (e.g., what if the value of covariate X changed by one unit?) and show how the model predictions would change.

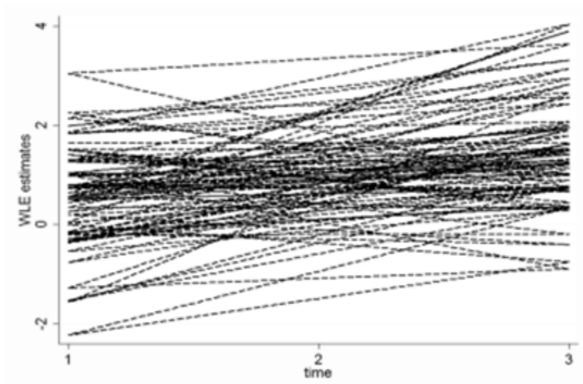

Figure 5. Individual Cognitive Linear Growth Trajectories
Note: This plot shows a random sample of 50 linear growth trajectories plotted using the WLE estimates for the baseline and linear growth parameters from the cognitive growth example.

Figure 9. Spaghetti plot of individual cognitive growth trajectories

Visualizing Multilevel Dyadic Data

Dyadic data have been used in couple's psychology research, but are increasingly being utilized to understand interpersonal emotion dynamics relevant across a broad range of psychopathology (e.g., my own research interests are in understanding how and when maladaptive interpersonal emotion transactions increase risk for suicide over time). Though a number of statistical guides are available to help researchers estimate dyadic models, there is no consensus on how to represent and/or visualize results. One issue is how to account for intra-individual variability (i.e., changes in person 1's affect over time) as well as inter-individual variability (i.e., changes in person 2's affect due to person 1's behavior) in a way that tells a coherent narrative about the underlying dyadic system. One lesson I encountered in my reading was to follow a staged approach to visualizing multilevel dyadic dyad: first, we can use time series plots to visualize the data before building the rest of one's statistical model.

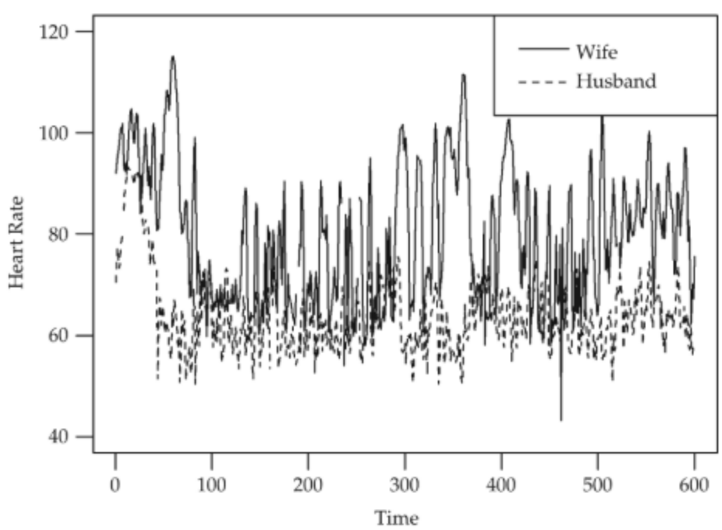

Figure 10. Time series plot of husband and wife's heart rate every second during a laboratory task

The plot above serves a similar function to a histogram plot in standard regression models. We can then model intra-individual dynamics by using scatterplots to visualize first-order discrete change in both partner's behaviors or physiology (e.g., heart rate). Such plots would help determine potential linearity of associations (e.g., plotting change in heart rate on the Y axis, and current heart rate on the X axis for both partners in two separate plots). Finally, one way of modeling inter-individual variability is through kernel density plots (KDPs). In the below KDP, higher "peaks" indicate more likely data values. We see a peak at ~60 BPM for the husband (X axis) and ~67.5 BPM for the wife (Y axis), and may represent a "set point" for the couple.

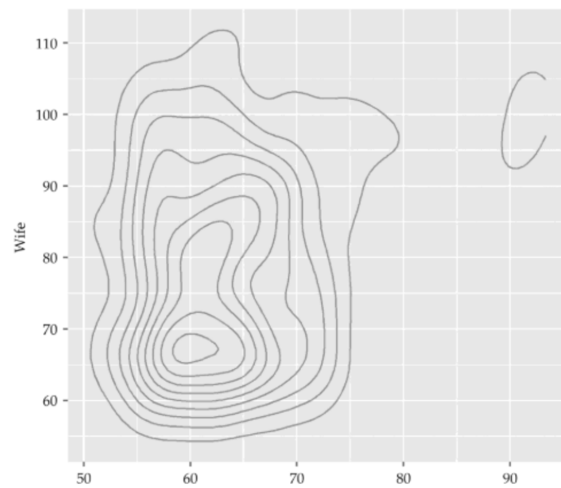

Figure 11. KDP of husband and wife heart rate in a dyadic model

Visualizing Longitudinal Data with Alluvial Diagrams

Visualizing longitudinal data can be difficult, and the key decision is how to incorporate time into the visualization. Spaghetti plots can be one option, but as seen above if there are many observations the spaghettis become clogged easily, making it very difficult to comprehend the trends. For example, suppose we wish to visualize the change of depression levels (a continuous variable) of individuals over time, if we draw one line per individual, then if there are more than 100 individuals then the lines get jammed together easily.

One way to visualize longitudinal change is that instead of individual lines, we can first categorize the continuous outcome, then make channels that represent changes between time points. This is shown as the "alluvial diagram" in Figure 10. As seen, the X axis is time, while the Y axis represents depression levels. The width of the channel represent proportions that correspond to changes/non-changes between time points. The grey channels are cases that never changed their depression status across time points, while the red channels represent cases that changed their status at least once. The transformation here is that rather than a continuous outcome of depression, depression is categorized into a binary outcome. However, one can envision that the plot could be extended into ordinal categorizations.

The key part for visualizations is that we are interested in the number of individuals that change, and the channel visualization provides such information. Instead of individual lines, the channels nicely summarize the proportion of changes/non-changes, which is akin to providing a summary statistic that aggregates individuals. Here the diagram is plotted using the "alluvial" package in R.

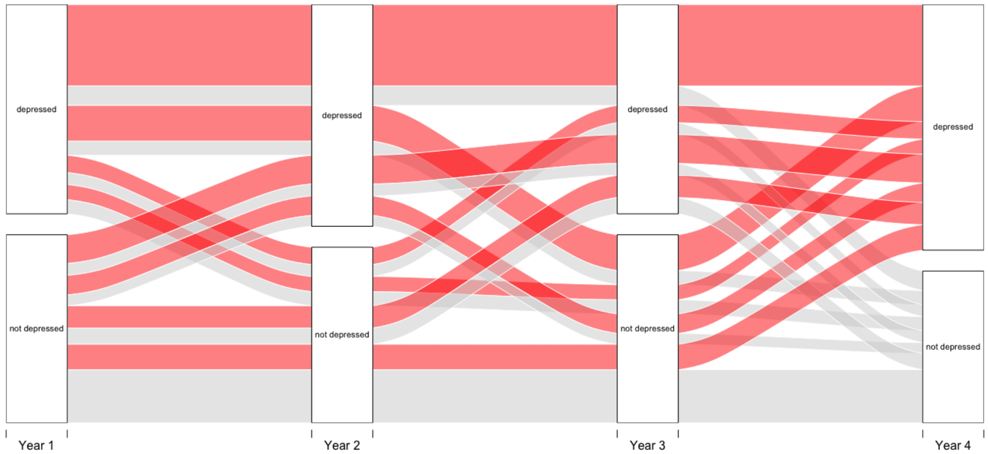

Figure 12. Alluvial Diagram of Depression change over time

Visualizing Daily Diary Data

A growing number of studies in the field of management use daily diary studies to capture phenomena that vary within individuals across multiple days. For instance, researchers have looked at whether charismatic or abusive leadership behaviors vary within and across days. I have a dataset that looks at whether an increase in ambient temperature leads to supervisors behaving more abusively towards their subordinates. It's essentially a repeated measured design where I surveyed the same supervisor-subordinate dyads over ten consecutive work days. For temperature, we used archival weather station data.

One of the challenges with repeated measures design is to try to find trends within the data that can disentangle the relationship between variables over time. Some variables might behave one way while others might behave in a different way during the same time period. I think one useful thing to do here is to make several loess/spline plots to get a sense of how each variable relates to time. Once that is established, one might be able to use GLM models within a Generalized Additive Model (GAM) framework to allow for certain variables to be smoothed in the model. For instance, as shown in the following image, I made several GAM plots which allowed me to get a sense of how temperature relates to abusive supervision. I was able to find that when temperature got greater than 90F, there was a bump in abusive behavior:

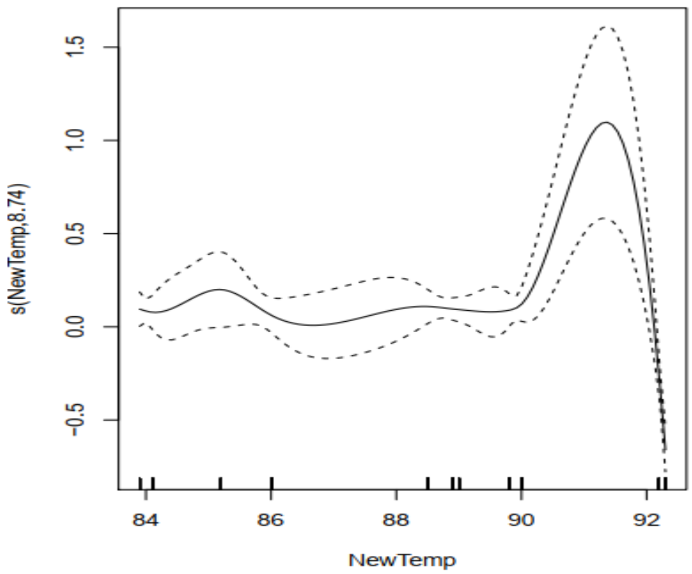

Figure 13. Relationships between temperature and abusive supervision using generalized additive models

Another useful lesson that we learned is that displaying MLMs is not necessarily different from displaying single-level models. In fact, the exploratory data analysis and the model results can be visualized in such a way that the reader doesn't even have to know that the model is multilevel.

Finally, another lesson that we learned (which is not necessarily about multilevel nature of the data) is that if the outcome variable has a low base-rate and has a narrow range, it might be more useful to treat it like a categorical variable and do ordered probit. The challenge here is that ordered probit doesn't nicely fit with multilevel data so one solution is to do ordered probit and do cluster-robust standard errors rather than doing a standard multilevel regression.

Visualizing Multilevel Logistic Regression

Multilevel logistic regression is more difficult than ordinary multilevel regression in that due to binary outcome variables, the log odds of the outcomes are modeled. Data visualization can help readers to understand multilevel logistic regression regardless of the lack of a deep understanding of complex mathematical statistics and algorithms.

We usually use scatter plots to check random effects. However, if we use ordinary scatterplots, this may mislead people for a data set. Because the dependent variables are binary, there is obviously a large degree of overlap. There are other different ways to show random effects using various ways. For example, we can use LOWESS smoothing (Bell, 2001). The dataset related to the below chart (see Figure 12) consists of 1,129 students nested in 19 schools. The outcome variable is whether a student passes an English test or and the x-axis is Ravens test in year 1 which is an ability measure. From the right plot using LOWESS smoothing, we can see not only that students with higher RAVENS test scores are more likely to pass the English test, but also that a random slopes model might be appropriate.

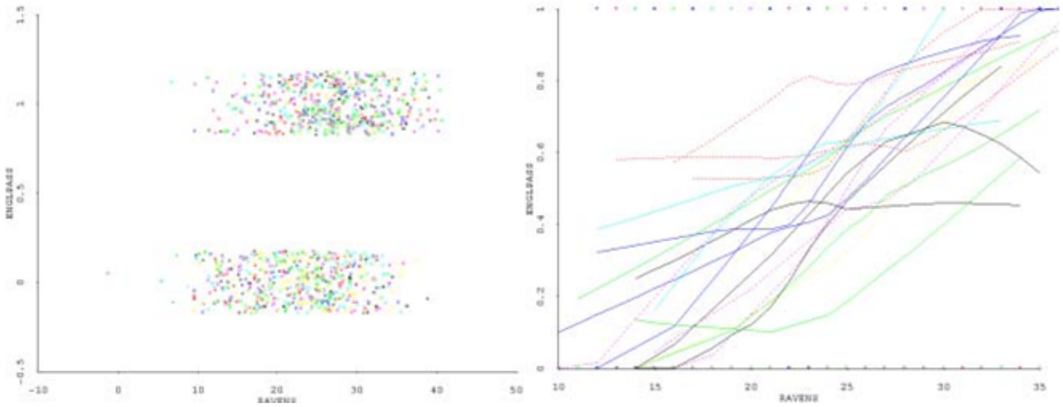

Figure 14. Comparing an ordinary scatter plot with a scatter plot using LOWESS smoothing

In addition, we can use a caterpillar plot with error bars below (UCLA statistical consulting, n.a). The example dataset consists of 8,525 pupils and nested in 407 doctors. The outcome variable is whether a patient's lung cancer goes into after treatment and the model (remission ~ Age + LengthofStay + FamilyHx + IL6 + CRP + CancerStage +Experience + (1 | DID) + 1(1 | HID)) included patient's age, family health status, IL6, and CRP as patient level continuous predictors, CancerStage as a patient level categorical predictor (I, II, III, or IV), Experience as a doctor level continuous predictor, and a random intercept by DID, doctor ID and HID hospital ID. Each blue dot is each remission mean of a doctor and a hospital and error bars are 95% confidence intervals. From the left graph (see Figure 13), we can see that many error bars deviate from 0, indicating that there are random effects of this model. However, we can see hospital effects on remission from the right graph (see Figure 13).

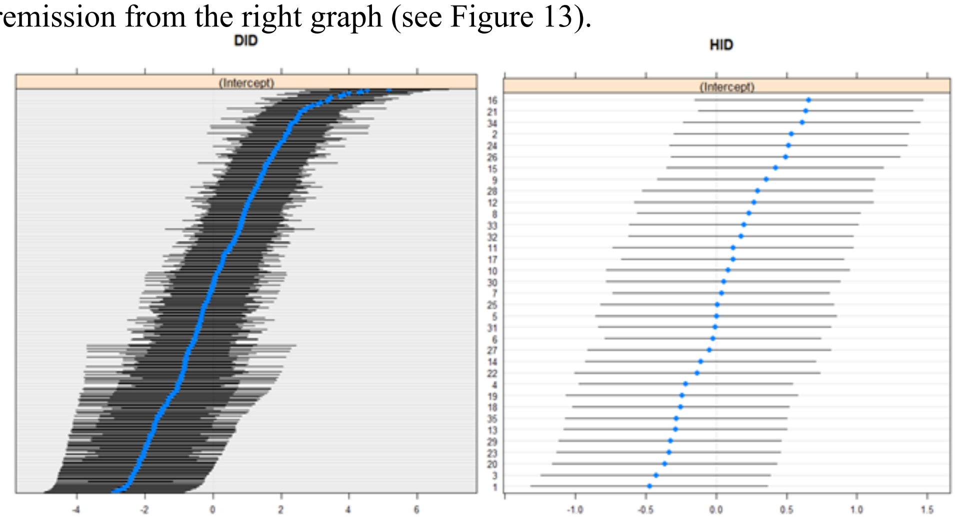

Figure 15. Random effects using a caterpillar plot with error bars

For the fixed effects, we usually see the result of a model in a table or a text of a research manuscript. However, they are very difficult to interpret because they are log odds, odd ratios, or probabilities. The logit is applicable for regression model because it is linearized, indicating that every 1 unit increase in a independent variable results in a coefficient unit increase in the outcome variable holding all else constant. However, it is hard for readers to interpret the numbers. Although the probabilities gives us better understanding intuitively, they are not linear.

From the study (Dolšak, Adolph, & Prakash, 2020) I learned that one of the reasonable ways to show the result of fixed effects of logistic models is to use a ropeladder plot. This can be a very efficient and effective way to show logistics regression model' results because the ropeladder plot provides an immediate and accurate impression and understanding of the statistical precision of results. Solid circle symbols indicate effects significantly different from zero at the 0.05 level and open symbols indicate non-significant results. 95% confidence intervals are expressed as horizontal lines. The upper row is the outcome variable. Every estimate obtained from a model ordered based on value of probit. From the graph below (see Figure XX), we can see that there is a significant effect only for the Mitigation, which increases the percentage of respondents strongly supporting carbon taxes by around 6%. However, other treatments do not have significant effects because 95 confidence intervals include 0%.

These visualizations clearly show the effects of random and fixed effects of the multilevel logistic model. This can contribute to the reader's understanding of multilevel logistic model's results.

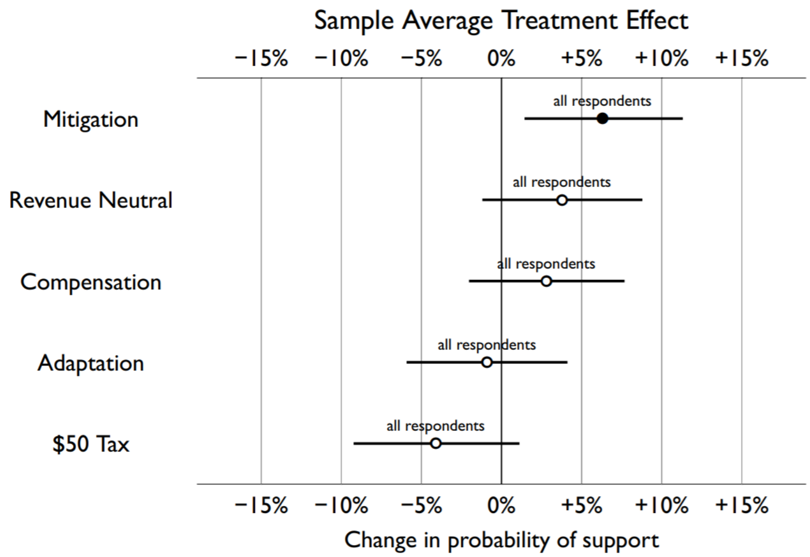

Figure 16. Fixed effects using a rope ladder plot with error bars

Visualizing Hierarchical Models

Visuals of the hierarchical data can provide us a better understanding of the structure of a hierarchy and the distribution of data within the context of the structure and to provide a summary of the data into the aggregated data in social sciences, biology, and public health domain. The main challenge of presenting hierarchical data is to avoid the information overload.

To present the visuals of a hierarchical model - we need to keep one thing in our mind that how well we are presenting our scientific story. When we are dealing with multiple levels it is important to present the effect of that level on the outcome variable to the readers.

Below I have presented an example of a plot which I created to visualize the school effect on maths score of the students in the USA. Here I have used the dataset egsingle which is a subset of the mathematics scores from the U.S. Sustaining Effects Study. The subset consists of information on 1721 students from 60 schools. I made this plot for my homework in the hierarchical modeling course.

Below the figure-1 shows the effects of school in eight different linear and hierarchical models. The vertical lines around each point estimates in the figure represent the 95% confidence intervals of the model and when these lines cross y=0 that is an indication of school effect on students' math score is statistically indistinguishable from zero (which indicates the relationship between school effect and students' math scores was not statistically significant). The plot suggests that the effects of school do not differ hugely among these eight different models. However, it should be noted that the confidence intervals for the hierarchical models are wider than the linear models. This may be because of the correlated nature of the multilevel observations in different levels.

This figure clearly highlights the effect of the school on students' maths score comparing the linear models. Such a presentation can lead us to have a clear view about the effect of schools, districts, states, countries on the outcome variable and it will fulfill one of the purposes of visualizing hierarchical models.

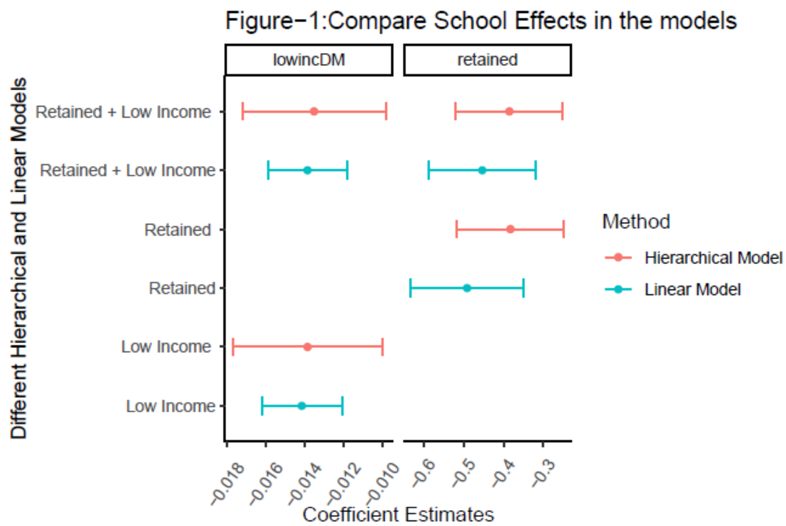

Figure 17. Visualising Multilevel Models: The Initial Analysis of Data and Policy Design and Public Support for Carbon Tax: Evidence from a 2018 U.S. National Online Survey Experiment

Conclusion

Multilevel modeling (MLM) is complex, both from a specification as well as visualization perspective. They are used to analyze data with complex variability patterns, with an emphasis on forms of variability that are nested (e.g., students within classrooms, observations within person). Our findings show that, although MLM specification may require additional statistical nuance, many visualizing approaches you may consider using may actually overlap with those used to visualize single-level models. Attention to your specific research questions is key in determining what visualization approach to use (e.g., are the model coefficients important to highlight? Could model-based predictions communicate the results effectively?), particularly because it can be easy to get carried away in creating an excess number of figures and diagrams for these complex models. Our "Lessons Learned" indicate a move away from spaghetti plots (they are visually taxing on the reader if observations are manifold). Alternative approaches include selecting a small number of representative cases that illustrate larger data trends, searching for latent classes in the data, and visualizing counterfactuals that highlight how model predictions would change. We also highlighted a number of ways to visualize changes over time (e.g., in longitudinal models), including use of alluvial or kernel density plots (KDPs). We can also use visuals in MLM in a similar way we'd use them to guide model building in standard linear regression (e.g., conducting exploratory plots to examine the underlying structure of the data and test model assumptions, use of smoothing techniques). Visualization can thus serve important initial guidance on the MLM process.

A few remaining issues uncovered include the following:

(1) How do we handle a large number of categories within a level from a visualization perspective? If we choose to visualize "representative" data points (vs. all data points), how do we select them?
(2) What are other ways of visualizing categorical data (e.g., criterion variables) in MLM?
(3) How do we handle (and visualize) non-linear associations in MLM, particularly in a way that's easy to interpret for readers?

References

Dedrick, R., Ferron, J., Hess, M., Hogarty, K., Kromrey, J., Lang, T., . . . Lee, R. (2009). *Multilevel Modeling: A Review of Methodological Issues and Applications. Review of Educational Research, 79*(1), 69-102.

Snijders, T. A., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

Glaser, D., & Hastings, R. (2011). An introduction to multilevel modeling for anesthesiologists. *Anesthesia and Analgesia, 113*(4), 877-887.

Gelman, A., Shor, B., Bafumi, J., & Park, D. (2005). Rich state, poor state, red state, blue state: What's the matter with Connecticut? *Quarterly Journal of Political Science*, 2, 345–367.

Lindgren, K. P., Neighbors, C., Teachman, B. A., Gasser, M. L., Kaysen, D., Norris, J., & Wiers, R. W. (2015). Habit doesn't make the predictions stronger: Implicit alcohol associations and habitualness predict drinking uniquely. *Addictive Behaviors, 45*, 139-145.

Randall, A. K., & Schoebi, D. (Eds.). (2018). *Interpersonal Emotion Dynamics in Close Relationships*. Cambridge University Press.

Wilson, M., Zheng, X., & McGuire, L. (2013). Formulating Latent Growth Using an Explanatory Item Response Model Approach. *Journal of Applied Measurement 13*(1), 1-22.

Bell, J. F. (2001, March). Visualising multilevel models: the initial analysis of data. In Third International Conference on Multilevel Analysis, Amsterdam, April.

Dolšak, N., Adolph, C., & Prakash, A. (2020). Policy Design and Public Support for Carbon Tax: Evidence from a 2018 US National Online Survey Experiment. *Public Administration*.

UCLA statistical consulting. Mixed effects logistic regression: R data analysis examples. Retrieved from https://stats.idre.ucla.edu/r/dae/mixed-effects-logistic-regression/