

Summary. Modern Bayesian inference is highly computational but commonly proceeds without reference to modern developments in statistical graphics. This should change. Visualization has two important roles to play in Bayesian data analysis:

(1) For model checking, graphs of data or functions of data and estimated model (for example, residual plots) can be visually compared to corresponding graphs of replicated data sets. This is the fully-model-based version of exploratory data analysis. The goal is to use graphical tools to explore aspects of the data not captured by the fitted model.

(2) For model understanding, graphs of inferences can be used to summarize estimates and uncertainties about batches of parameters in hierarchical and other structured models. Traditional tools of summarizing models (such as looking at coefficients and analytical relationships) are too crude to usefully summarize the multiple levels of variation and uncertainty that arise in Bayesian hierarchical models.

Our discussion is limited to relatively simple static displays. Having set up conceptual structures for model checking and model understanding, we anticipate that researchers in interactive graphics will develop appropriate tools for these tasks.

Visualization in Bayesian Data Analysis

Jouni Kerman¹, Andrew Gelman², Tian Zheng³, and Yuejing Ding⁴

¹ Columbia University kerman@stat.columbia.edu

² Columbia University gelman@stat.columbia.edu

³ Columbia University tzheng@stat.columbia.edu

⁴ Columbia University yding@stat.columbia.edu

1 Introduction

Modern Bayesian statistical science commonly proceeds without reference to statistical graphics; both involve computation, but they are rarely considered connected. The traditional views of the usage of Bayesian statistics and statistical graphics have a certain clash of attitudes. Bayesians might do some exploratory data analysis (EDA) to start, but once the model or class of models is specified, analysis continues with the fitting of the data; graphs are then typically used to check convergence of simulations, or used as teaching aids or as presentations – but not as part of the data analysis. Exploratory data analysis seems to have no formal place in Bayesian statistics once a model has actually been fit. With this extreme view, some people would see the connection between Bayesian inference and graphics only through convergence plots of Markov chain simulations, and histograms and kernel density plots of the resulting estimates of scalar parameters.

On the other hand, the traditional attitude of users of statistical graphics is that “all models are wrong”; we are supposed to get as close to data as possible without reference to a model, which just incorporates undesired components of subjectivity and parametric assumptions into preliminary analysis. In a true Tukey tradition, even if a graphical method can be derived from a probability model (e.g., rootograms from the Poisson distribution), we still don’t mention the model, because the graph should stand or fall on its own.

Given these seemingly incompatible attitudes, how can we then integrate the inherently model-based Bayesian inference with the (apparently) inherently model-averse nature of statistical graphics? Our attitude is a synthesis of ideas adopted from statistical graphics and Bayesian data analysis. The fundamental idea is that we consider all statistical graphs to be *implicit or explicit comparisons* to a reference distribution, that is, to a model. This idea is introduced in [8]; the article proposes an approach to unify EDA with formal Bayesian statistical methods. The connection between EDA and goodness-of-fit testing is discussed in [7]. These two articles formalize the graphical

model-checking ideas presented in [5, 4, 3] which have been applied informally in various applied contexts for a long time (e.g. [6, 15]).

EDA and its role in model understanding and model checking

Exploratory data analysis, aided by graphs, is done to look for patterns in the data. Here the reference distributions are typically implicit, but are always there in the mind of the modeler. In an early article on EDA by Tukey [18], he focused on “graphs intended to let us see what may be happening over and above what we have already described,” which suggests that these graphs can be built upon existing models. After all, to look for the unexpected is to look for something that differs from something that we were expecting – the reference model. For example, even simple time series plots are viewed implicitly as comparisons to zero, a horizontal line, linearity, or monotonicity, and so forth. Looking at two-way scatterplots imply usually a reference to an assumed model of independence. Before looking at a histogram, we have certain baselines of comparison (symmetric distribution, bimodal, skewed) in our minds. In the Bayesian sense, looking at inferences and deciding whether they “make sense” can be interpreted to be a comparison of the estimates to our prior knowledge, that is, to a prior model.

The ideas that EDA gives us can be more powerful than before if used with sophisticated models. Even if one believes that graphical methods should be model-free, it can still be useful to have provisional models to make EDA more effective. EDA can be thought of as an implicit comparison to a multilevel model; in addition, EDA can be applied to inferences as well as to raw data.

In Bayesian probability model understanding and model checking, the reference distribution can be formally obtained by computing the predictive distribution of the observables, also called the replication distribution. Draws from the posterior predictive distribution represent our posterior knowledge of the (marginal) distribution of the observables. Model fit can be assessed by comparing the observed values with posterior predictive draws; discrepancies represent departures from the model. Comparisons are usually best done by graphs, since the models for the observables are usually complex. However, depending on the complexity of the model, often very sophisticated graphical checks need to be devised and tailored to the model. In this article, we review the principles, show examples on how to apply them to a data analysis, and discuss potential extensions.

Comparable non-Bayesian approaches

Our Bayesian data visualization tools make use of posterior uncertainty as summarized by simulation draws of parameters and replicated data. A similar non-Bayesian analysis might compute a point estimate of parameters and then

simulate data using a parametric bootstrap. This reduces to (Bayesian) posterior predictive checking if the parameter estimates are estimated precisely (if the point estimate has no posterior variance).

A confidence interval (point estimate plus minus standard errors) summarizes approximately the posterior uncertainty about a parameter. In a multi-level model, a common non-Bayesian approach is to compute point estimates for the hyperparameters and then simulate the modeled parameters.

The visualization tools described in this article should also work in these non-Bayesian settings.

2 Visualization for Understanding and Checking Models

The key to Bayesian inference is its unified treatment of uncertainty and variability; we would like to use this in data visualization (e.g., [19], chapter 15) as well as in data analysis in general [13].

Using statistical graphics in model-based data analysis

EDA is the search for unanticipated areas of model misfit. Confirmatory data analysis (CDA), on the other hand, quantifies the extent to which these discrepancies could be expected to occur by chance. We would like to apply the same principles to the more complex models that can be fit today using methods such as Bayesian inference and nonparametric statistics. Complex modeling makes EDA more effective in the sense of being able to capture more subtle patterns in data. Conversely, when complex models are being used, graphical checks are more necessary than ever to detect areas of model misfit.

We, like other statisticians, do statistical modeling in an iterative fashion, exploring our modeling options, starting with simple models, and expanding the models into more complex, more realistic models, putting in as much structure as possible, at each stage trying to find deficiencies in our model, building new models, and iterating the process until we are satisfied. Then, we use simulation-based model checks (comparisons of observed data to replications under the model); to find patterns that represent deviations from the current model. Moreover, we apply the methods and ideas of EDA to structures other than raw data, such as plots of parameter inferences, latent data, completed data [10]; Figure 1 illustrates this.

At a theoretical level, we look at the model, identify different sorts of graphical displays with different symmetries or invariances in an explicit or implicit reference distribution of test variables. This serves two purposes: to put some theoretical structure on graphics and EDA, so that graphical methods can be interpreted in terms of implicit models and to give guidelines on how most effectively to express a model check as a graphical procedure.

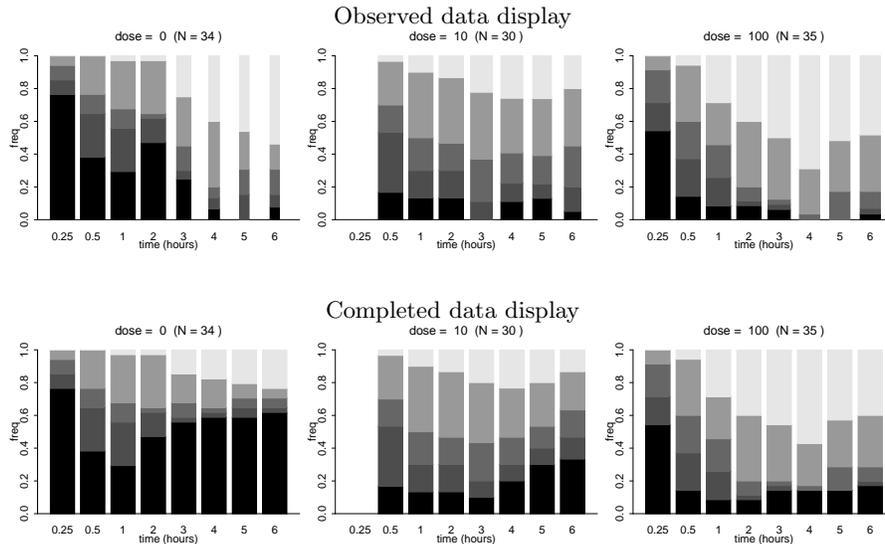


Fig. 1. Summary of pain relief responses over time under different doses from a clinical trial with nonignorable dropout. In each summary bar, the shadings from bottom to top indicate “no pain relief” and intermediate levels up to “complete pain relief.” The graphs in the top row include only the persons who have not dropped out (with the width of the histogram bars proportional to the number of subjects remaining at each time point). The graphs in the bottom row include all persons, with imputed responses for the dropouts. The bottom row of plots—which are based on completed data sets—are much more directly interpretable than the observed-data plots on the top row

Bayesian exploratory data analysis

Bayesian inference has the advantage that the reference distribution – the predictive distribution for data that could have been seen – arises directly from the posterior simulations (which are typically obtained using iterative simulation). Consequently, we can draw from this distribution and use these simulations to produce a graph comparing the predictions to the observed data. Such graphs can be customized to exploit symmetries in the underlying model to help interpretation. Inclusion of imputed missing and latent data can allow more understandable completed-data exploratory plots.

The existence of an explicit underlying model has the benefit of suggesting exploratory plots that address the fit of data to whatever model that is being fit. Putting EDA within the general theory of model checking allows the potential for graphical methods to become a more automatic presence in statistical modeling.

Our approach is to use statistical graphics in all stages of data analysis: model building, model fitting, model checking and model understanding. In all stages of data analysis, we need model checking tools so we can see the faults

and shortcomings in our model. Understanding complex probability models often requires complex tools. Simple test statistics and p -values just do not suffice, so we need graphs to aid us. More often than not we need to customize the graphs to the problem; we are not usually satisfied by looking only at a standard graph. However, new and better standard methods of graphical display should be developed.

Considering that graphs are equivalent to model checks, “exploratory” data analysis is not done only in the beginning of the modeling process; re-exploring the model is done at each stage after model fitting. Building a new, complex model may bring unforeseen problems in the fit; in anticipation of these, we explore the fitted model with standard and customized statistical graphs that are then used to attempt to falsify the model and to find ways to improve it, or to discard it completely and start all over again. It is also standard for our practice not to do model averaging or concentrate on the predictor selection in regression problems; our models usually evolve to more and more complex ones.

The key idea in Bayesian statistics – as opposed to simply “statistical modeling” or “estimation” – is working with posterior uncertainty in inferences. At the theoretical level, with random variables; at a more practical level, with simulations representing draws from the joint posterior distribution. This is seen most clearly in hierarchical modeling. Figure 2 shows an example of visualizing posterior uncertainty in a hierarchical logistic regression model [11].

Hierarchical models and parameter naming conventions

In hierarchical and other structured models, rather than to display individual coefficients, we wish to compare the values within batches of parameters. For example, we might want to compare group-level means along with their uncertainty intervals together. Posterior intervals are easily derived from the matrix of posterior simulations. Traditional tools of summarizing models (such as looking at coefficients and analytical relationships) are too crude to usefully summarize the multiple levels of variation and uncertainty that arise in such models. These can be thought of as corresponding to the “sources of variation” in an ANOVA table.

Hierarchical models feature multiple levels of variation, and hence feature multiple levels of batches of parameters. Hence, the choice of label for the batches is also important: parameters with similar names can be compared to each other. In this way, naming can be thought of as a structure analogous to hierarchical modeling. Instead of using generic $\theta_1, \dots, \theta_k$ for all scalar parameters, we would, for example, name the individual-level regression coefficients $\beta = (\beta_1, \dots, \beta_n)$, and the group-level coefficients $\alpha = (\alpha_1, \dots, \alpha_J)$, and the intercept μ . Figure 4 shows an example of why this works: parameters with similar names can be compared to each other. Rather than plotting posterior histograms or kernel density estimates of the parameters, we usually summa-

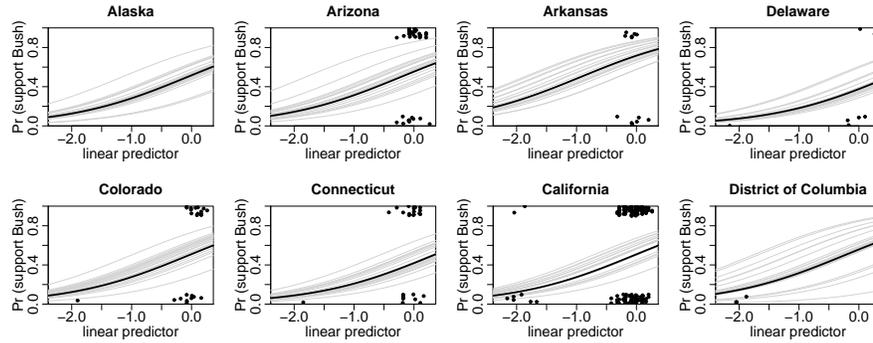


Fig. 2. Displaying a fitted hierarchical logistic regression model, along with inferential uncertainty. Eight states are shown from a voting preference model $\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + X_i\beta)$ including all 50 U.S. states. The solid black line is the median estimate for the probability of a survey respondent in that state supporting George Bush for President in 1988. The gray lines represent random draws from the posterior distribution of the logistic regression coefficients. The dots shown are the observed data (zeros and ones), vertically jittered to make them more distinguishable.

This figure demonstrates several principles of Bayesian visualization: (1) small multiples: parallel plots displaying the hierarchical structure of the data and model; (2) graphs used to understand the fitted model; (3) fitted model and the data on the same graph; (4) graphical display of posterior uncertainty. Also a principle of Bayesian inference is illustrated: in Alaska we have no data but nevertheless we were able to draw predictions from the model.

General principles of “good graphics” are used: common scale for all graphs with bounds at 0 and 1; clear axis labeling; jittering (which works for the moderate sample sizes of this example); thin lines and small dots; ordering of the small plots by some meaningful criterion (here by decreasing support for Republicans) rather than alphabetically.

The distribution of the linear predictor is skewed because the most important single predictor by far was indicator for African-American, which has expected value 0.12. The common scaling of the axes means that we do not actually need to label the axes on each graph; however, in this case we find the repeated labels to be convenient. Labeling only some graphs (as is done with Trellis plots) saves space but, to our eyes, makes the graph more of a puzzle to read, especially as presentation graphics

size the parameters (at least, for a first look of the inferences) by plotting their posterior intervals.

Model checking

As stated earlier, we view statistical graphics as implicit or explicit model checks. Conversely, we view model checking as comparison of data to repli-

cated data under the model, including both exploratory graphics and confirmatory calculations such as p -values. Our goal is *not* the classical goal of identifying whether the model fits or not – and certainly *not* the goal of classifying models into correct and incorrect, which is the focus of the Neyman-Pearson theory of Type 1 and Type 2 errors. We rather seek to understand in what ways the data depart from the fitted model. From this perspective, the two key components of an exploratory model check are (1) the graphical display and (2) the reference distribution to which the data are compared.

The appropriate display depends on the aspects of the model being checked, but in general, graphs of data or functions of data and estimated models (for example, residual plots) can be visually compared to corresponding graphs of replicated data sets. This is the fully-model-based version of exploratory data analysis. The goal is to use graphical tools to explore aspects of the data not captured by the fitted model.

3 Example: A Hierarchical Model of Structure in Social Networks

As an example, we take the problem of estimating the size of social networks [20]. The model uses a negative-binomial model with an overdispersion parameter:

$$y_{ik} \sim \text{Negative-binomial}(\text{mean} = a_i b_k, \text{overdispersion} = \omega_k),$$

where the groups (subpopulations) are indexed by k ($k = 1, \dots, K$) and the respondents are indexed by i ($i = 1, \dots, n$). In this study, $n = 1370$ and $K = 32$. Each respondent is asked how many people he or she knows in each of the K subpopulations. The subpopulations are identified by names (people called Nicole, Michael, Stephanie, etc.), and by certain characteristics (airline pilots, people with HIV, in prison, etc.).

Without going into the details, we remark that a_i is an individual-level parameter indicating the propensity of the person i to know persons in other groups (we call this a “gregariousness” parameter); it is modeled to be $a_i = e^{\alpha_i}$ where $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$; similarly, b_k is a group-level prevalence (or group size) parameter with a model $b_k = e^{\beta_k}$ where $\beta_k \sim N(\mu_\beta, \sigma_\beta^2)$. The overdispersion parameter vector $\omega = (\omega_1, \dots, \omega_K)$ and the hyperparameters are assigned noninformative (or, weakly informative) prior distributions.

The model is fit using a combination of Gibbs and Metropolis algorithms, so our inferences for the modeled parameters (a, b, ω) , and the hyperparameters, $(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta)$, are obtained as simulated draws from their joint posterior distribution.

Model-informed exploratory data analysis

Figure 3 displays a small portion of an exploratory data analysis, with histograms of responses for two of the survey questions, along with simulations

of what could appear under three fitted models. The last of the models is the one we used; as the EDA shows, the fit is still not perfect.

A first look at the estimates

We might summarize the estimated posterior distributions of the scalar parameters as histograms, but in most cases we find displaying intervals a more concise way to display the inferences; our goal here is not just the estimates, but also the comparison of the parameters within each batch.

We display the parameter estimates with their 50% and 95% posterior intervals as shown in Figure 4. Along with the estimates, the graph summarizes the convergence statistic graphically. Since there are over 1300 quantities in the model, all of them cannot be displayed on one sheet. For smaller models, this graph provides a quick summary of the results – but of course this is just a starting point.

We are usually satisfied with the convergence of the algorithm if the values of the \hat{R} convergence statistic [9] are below 1.1 for all scalar parameters. A value of \hat{R} close to 1.0 implies good mixing of the chains; if however $\hat{R} > 1.1$ for some of the parameters, we let the sampler iterate more. By using a scalar summary rather than looking at trace plots, we are able to quickly assess the mixing of all the parameters in a model.

Distribution of social network sizes a_i

We proceed to summarize the estimates of the 1370 parameters a_i . A table of numbers would be useless unless we want to find the numerical posterior estimates for a certain person in the study; our goal is rather to visualize the posterior distribution of the a_i , so a histogram is a much more appropriate summary. It is interesting to see how men and women differ in their perceived “gregariousness”; we therefore display the posterior mean estimates of a_i as two separate histograms by dividing a_i into men’s and women’s estimates. See Figure 5.

A histogram derived from the posterior means of hundreds of parameters is still but a point estimate; we also want to visualize the posterior *uncertainty* in the histogram. In the case of one scalar, we always draw the posterior intervals that account for the uncertainty in estimation; in the case of a vector that is shown as a histogram, we similarly want to display the uncertainty in the histogram estimate. To do this, we sample (say, twenty) vectors from the matrix of simulations and overlay the new twenty histogram estimates as lines on the average histograms. This gives us a rough idea how good an estimate the histogram is. As a rule of thumb, we don’t plot the “theta-hats” (point estimates), we plot the “thetas” (posterior draws representing the random variables) themselves.

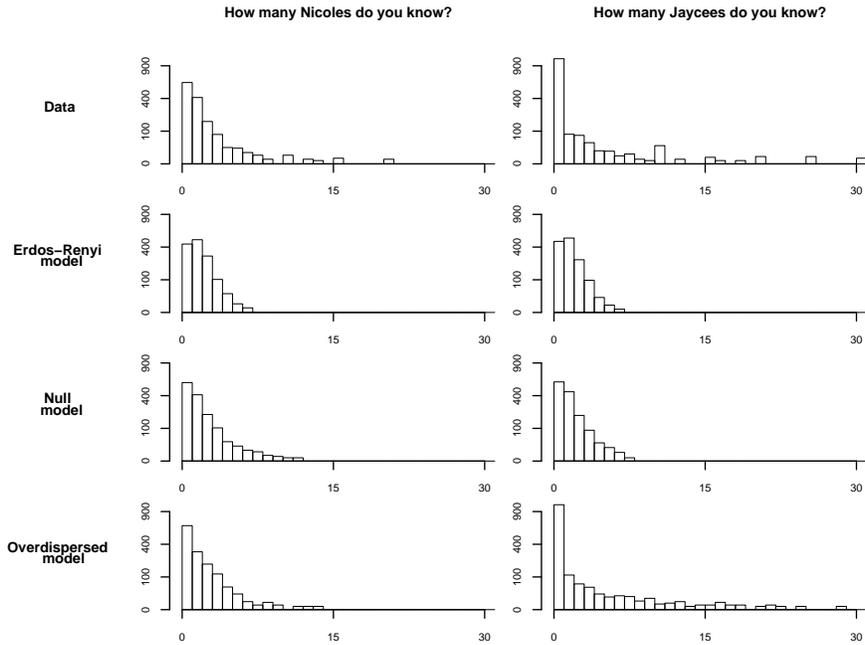


Fig. 3. Histograms (on the square-root scale) of responses to “How many persons do you know named Nicole?” and “How many Jaycees do you know?” from the data and from random simulations under three fitted models: the Erdős-Renyi model (completely random links), our null model (some people more gregarious than others, but uniform relative propensities for people to form ties to all groups), and our overdispersed model (variation in gregariousness and variation in propensities to form ties to different groups). Each model shows more dispersion than the one above, with the overdispersed model fitting the data reasonably well. The propensities to form ties to Jaycees show much more variation than the propensities to form ties to Nicoles, and hence the Jaycees counts are much more overdispersed. (The data also show minor idiosyncrasies such as small peaks at the responses 10, 15, 20, and 25. All values greater than 30 have been truncated at 30.) We display on square-root scale to more clearly reveal patterns in the tails

Extending the model by imposing a regression structure

We also fit an extended model, with an individual-level regression model, $\log(a_i) \sim N(X_i\psi, \sigma_\alpha^2)$. The predictors include the indicators of female, non-white, income > \$80000, income < \$20000, employment, education (high-school or more).

Figure 4 shows an example on how to visually compare regression coefficients ψ_{ik} on explanatory variables (characteristic of survey respondent, i) for different groups (k). Whenever our goal is to compare estimates, we

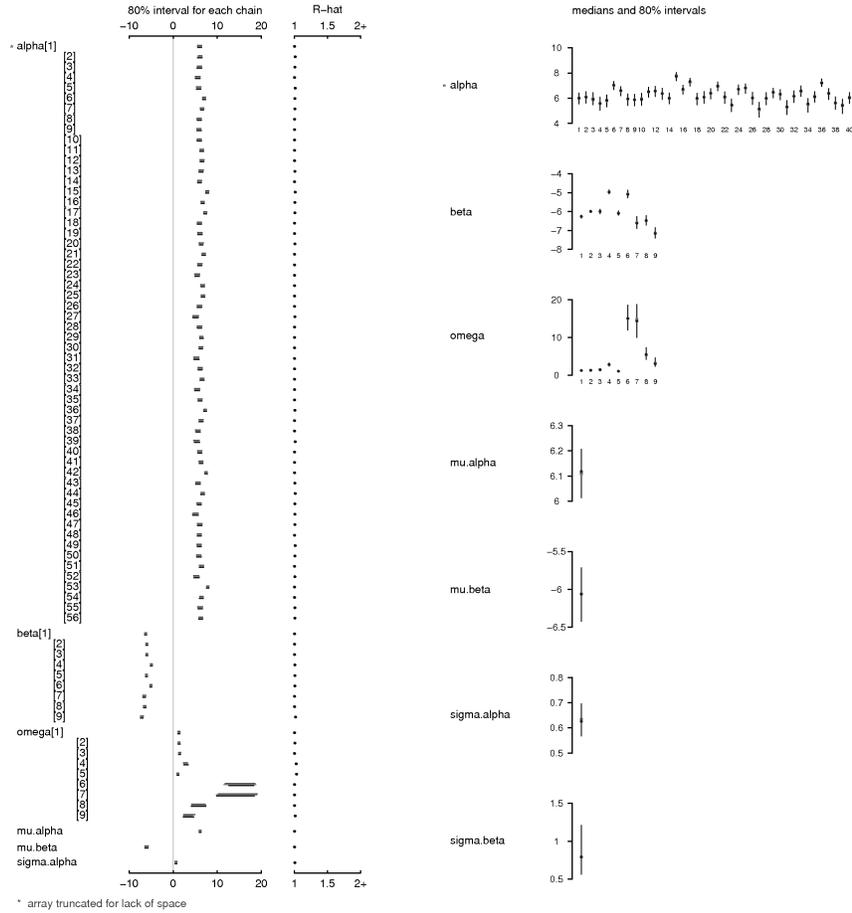


Fig. 4. A graphical summary (produced automatically by the R package R2WinBUGS [16]) of the estimated scalar parameters in the social networks model. Another alternative is density estimates or histograms of individual scalar parameters; these can be useful, but are difficult to apply to vector parameters. In contrast, the right side of our display here allows immediate understanding of inferences for the vectors α, β, ω (another option would be to use parallel coordinate plots for each vector parameter.)

For the purpose of displaying inferences, this graph is inefficient: given that approximately convergence has been reached, only the right side of the display is necessary. However, we include it as an example of quick inferential summary which, for all its simplicity, is still far more informative than a table of parameter estimates and standard errors

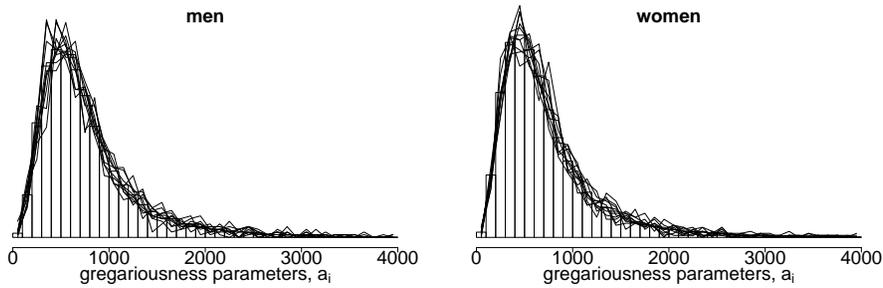


Fig. 5. Estimated distributions of the “gregariousness parameters” α_i for women and men. The overlain lines are posterior simulation draws indicating inferential uncertainty in the histograms. Our primary goal here is to display inferences for the distribution of gregariousness within each sex, not to compare averages between sexes (which could be done, for example, using parallel boxplots.) We compare groups using a regression model as in Figure 6

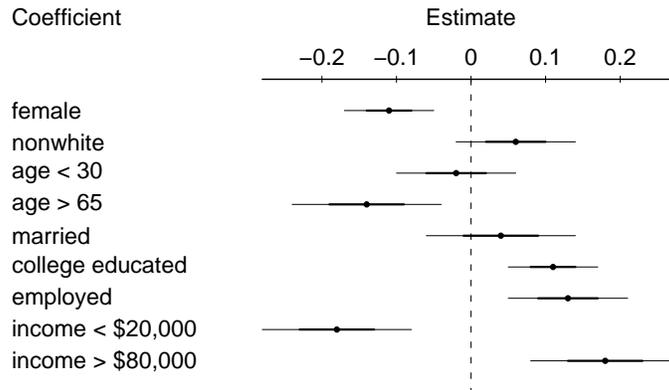


Fig. 6. Coefficients (and ± 1 standard error and ± 2 standard error intervals) of the regression of estimated log gregariousness parameters α_i on personal characteristics. Because the regression is on the logarithmic scale, the coefficients (with the exception of the constant term) can be interpreted as proportional differences: thus, with all else held constant women have social network sizes 11% smaller than men, persons over 65 have social network sizes 14% lower than others, and so forth

think of a graph first. Figure 7 could have been equivalently summarized by the uncertainty interval endpoints and the posterior median estimate, but it would not have been an efficient tool to visualize the coefficient estimates. For comparison, draw a graph; for looking up specific numbers, construct a table.

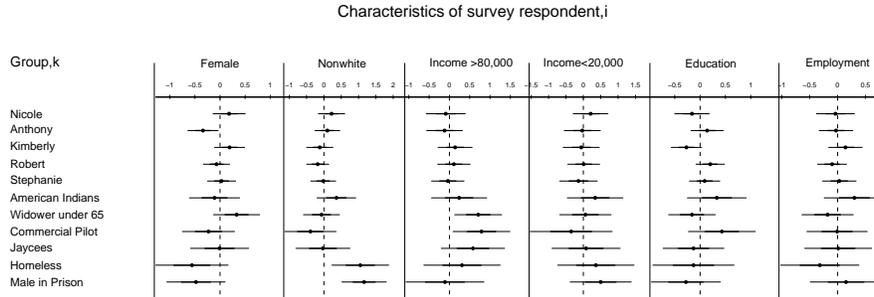


Fig. 7. A graph comparing the estimates from a more complex version of the social networks model, using individual-level regression predictors: $\log(a_i) \sim N(X_i\psi, \sigma_\alpha^2)$. The rows (Nicole, Anthony, etc.) refer to the groups and the columns refer to the various predictors. Comparison is efficient when coefficients are rearranged into “table of graphs” like this

Posterior predictive checks

A natural way to assess the fit of a Bayesian model is to look at how well the predictions from the model are consistent with the observed data [9]. We do this by comparing the posterior predictive simulations with the data. In our social networks example, we create a set of predictive simulations by sampling new data independently from the negative binomial distributions given the parameter vectors a , b , ω drawn from the posterior simulations already calculated. We draw, say, L simulations, $y_{ik}^{(1)}, \dots, y_{ik}^{(L)}$ for each i, k . Each set of simulations $\{y_{ik}^{(\ell)}\}_{\ell=1}^L$ is an approximation of the marginal posterior distribution of y_{ik} , denoted by y_{ik}^{rep} , where “rep” stands for the “replication distribution”; y^{rep} stands for the $n \times K$ replicated observation matrix.

It is possible to find a numerical summary (that is, a test statistic) for some feature of the data (such as standard deviation, mean, etc.) and then compare it to the corresponding summary of y^{rep} but in addition we prefer to draw *graphical test statistics* since a few single numbers rarely can catch the complexity of the data. In general notation, for some suitable test function T , which may thus be a graph, we compare $T(y)$ with $T(y^{\text{rep}})$.

Plots of data compared with replicated data

For the social networks model, we choose to compare the data and the predictions by plotting the observed versus expected proportions of responses y_{ik} . For each subpopulation k , we compute the proportion of the 1370 respondents for which y_{ik} equals 0, 1, 3, 5, 9, 10, and finally those with $y_{ik} \geq 13$. These

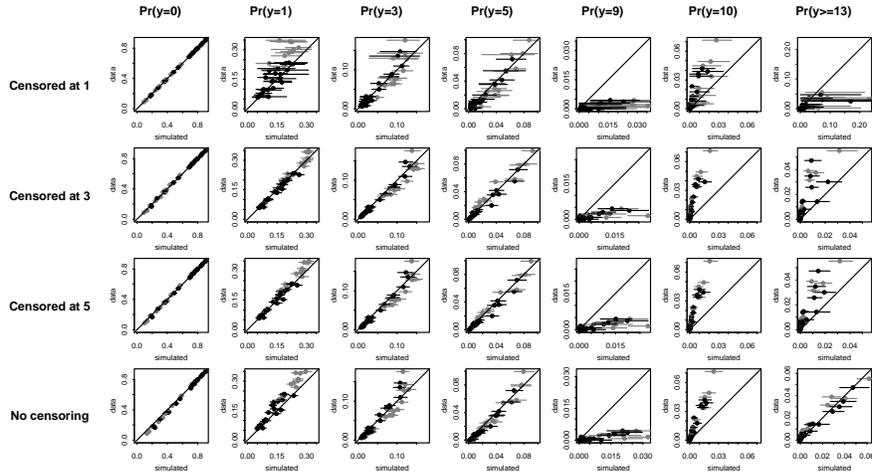


Fig. 8. Model checking graphs: observed vs. expected proportions of responses y_{ik} of 0, 1, 3, 5, 9, 10, and ≥ 13 . Each row of plots compares actual data to the estimate from one of four fitted models. The bottom row shows our main model, and the top three rows show models fit censoring the data at 1, 3, and 5. In each plot, each dot represents a subpopulation, with names in gray, non-names in black, and 95% posterior intervals indicated by horizontal lines

values are then compared to posterior predictive simulations under the model. Naturally, we plot the uncertainty intervals of $\Pr_k(y_{ik} = m)$ instead of their point estimates.

The bottom row of Figure 8 shows the plots. On the whole, the model fits the aggregate counts fairly well, but tends to under-predict the proportion of respondents who know exactly one person in a category. In addition, the data and predicted values for $y = 9$ and $y = 10$ show the artifact that people are more likely to answer with round numbers.

The three first rows of Figure 8 shows the plots for three alternative models [20]. This plot illustrates one of our main principles: whenever we need to compare a series of graphs, we plot them side by side on the same page so visual comparison is efficient [1, 17]. There is no advantage in scattering closely related graphs over several pages.

4 Challenges of Graphical Display of Bayesian Inferences

We expect that the quality of statistical analyses would benefit greatly if graphs were more routinely used as part of the data analysis. Exploratory data analysis would be more effective if it could be implemented as a part

of software for complex modeling. To some extent this is done with residual plots in regression models, but for complex models there is potential for much more progress.

As discussed in detail in [8], we anticipate four challenges: (1) Integrating the automatic generation of replication distributions into the computing environment; (2) Choosing the replication distribution – it is not an automatic task since the task requires selecting which parameters to resample and which to keep constant; (3) Choosing the test variables; (4) Displaying test variables as graphs. In the near future, automatic features for simulating replication distributions and performing standard model checks should be possible.

Integrating graphics and Bayesian modeling

We fit Bayesian models routinely with such software as BUGS [2], bringing the simulations over to R [14] using R2WinBUGS [16]. Summarizing simulations in R can also be done in a more natural way by converting the simulation matrices into vectors of random variable objects [13]. BUGS has also its limitations; we also fit complex models in R using the “Universal Markov chain sampler” Umacs [12].

We are currently investigating how to define an integrated Bayesian computing environment where modeling, fitting, and automatic generation of replications for model checking is possible. It requires further effort to develop standardized graphical displays for Bayesian model checking and understanding. An integrated computing environment is nevertheless the necessary starting point, since functions that generate such complex graphical displays should have full access to the models, the data, and predictions.

The environment must also take into account the fact that we fit multiple models with possibly different sets of observations; without a system that can distinguish multiple models (and the inferences associated with them) it is not possible to do comparison across them.

5 Summary

Exploratory data analysis and modeling can work well together: in our applied research, graphs are used for model understanding and model checking. In the initial phase we create plots to reveal to us how the models work, and then plot data as compared to the model to see where more work is needed.

We gain insight to the shortcomings of the model by doing graphical model checks. Graphs are most often drawn for comparison to an implicit reference distributions (e.g., Poisson model for rootograms, independence-with-mean-zero for residual plots, or normality for quantile-quantile plots), but we would also include more general comparisons; for example, a time series plot is implicitly compared to a constant line. In Bayesian data analysis, the reference

distribution can be formally obtained by computing the replication distribution of the observables; the observed quantities can be plotted against draws from the replication distribution to compare the fit of the model.

We aim to make graphical displays an integrated and an automatic part of our data analysis. Standardized graphical tests must be developed and these should be routinely generated by the modeling and model-fitting environment.

6 Acknowledgements

We thank Matt Salganik for collaboration with the social networks project, and the National Science Foundation for financial support.

References

1. J. Bertin. *Semiology of Graphics*. University of Wisconsin Press, 1967/1983. Translation by W. J. Berg.
2. BUGS Project. BUGS: Bayesian Inference Using Gibbs Sampling. <http://www.mrc-bsu.cam.ac.uk/bugs/>, 2004.
3. A. Buja, D. Asimov, C. Hurley, and J. A. McDonald. Elements of a viewing pipeline for data analysis. In W. S. Cleveland and M. E. McGill, editors, *Dynamic Graphics for Statistics*, pages 277–308, Belmont, CA, 1988. Wadsworth.
4. A. Buja and D. Cook. Inference for data visualization. Talk given at Joint Statistical Meetings 1999. Available at <http://www-stat.wharton.upenn.edu/~buja/PAPERS/jsm99.ps.gz>, 1999.
5. A. Buja, D. Cook, and D. F. Swayne. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, 5:78–99, 1996.
6. R. R. Bush and F. Mosteller. *Stochastic Models for Learning*. Wiley, New York, 1955.
7. A. Gelman. A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review*, 71:369–382, 2003.
8. A. Gelman. Exploratory data analysis for complex models (with discussion). *Journal of Computational and Graphical Statistics*, 13:755–779, 2004.
9. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, London, 2nd edition, 2003.
10. A. Gelman, I. V. Mechelen, G. Verbeke, D. F. Heitjan, and M. Meulders. Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics*, 61:74–85, 2005.
11. A. Gelman, B. Shor, J. Bafumi, and D. Park. Rich state, poor state, red state, blue state: What’s the matter with Connecticut? Technical report, Columbia University, Department of Political Science, New York, 2005.
12. J. Kerman. Umacs: A Universal Markov Chain Sampler. Technical report, Department of Statistics, Columbia University, 2006.
13. J. Kerman and A. Gelman. Manipulating and summarizing posterior simulations using random variable objects. Technical report, Department of Statistics, Columbia University, 2005.

14. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004.
15. B. D. Ripley. *Statistical Inference for Spatial Processes*. Cambridge University Press, New York, 1988.
16. S. Sturtz, U. Ligges, and A. Gelman. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, 12(3):1–16, 2005.
17. E. R. Tufte. *Envisioning Information*. Graphics Press., Cheshire, Conn., 1990.
18. J. W. Tukey. Some graphic and semigraphic displays. In T. A. Bancroft, editor, *Statistical Papers in Honor of George W. Snedecor*, Ames, IA, 1972. Iowa State University Press.
19. L. Wilkinson. *The Grammar of Graphics*. Springer, New York, 2nd edition, 2005.
20. T. Zheng, M. J. Salganik, and A. Gelman. How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, 2006. To appear.