

Google Summer of Code Proposal: bayesplot

1 About

In May, I'll finish a Master of Applied Data Science at the University of Michigan; the curriculum has been Python-centric and essentially acted as a crash course in vanilla machine learning, but provided some foundations in causal inference, uncertainty, and math that have carried over to my personal study of Bayesian stats. I'm motivated by the prospect of making the Bayesian workflow accessible to anyone who's interested in applying it, and think bayesplot is a really key component of both learning and maintaining a principled way of doing Bayesian inference in the real world.

- [Github](#)
- [Linkedin](#)
- [Email](#)

2 Abstract

Bayesplot is an R package that provides an extensive library of plotting functions for use after fitting Bayesian models (typically with MCMC). The plots created by **bayesplot** are ggplot objects, which means that after a plot is created it can be further customized using various functions from the **ggplot2** package.

Currently **bayesplot** offers a variety of plots of posterior draws, visual MCMC diagnostics, graphical posterior (or prior) predictive checking, and general plots of posterior (or prior) predictive distributions.

The goal of this project will be to implement the ensemble of visual predictive checks proposed by [Säilynoja et al.](#) in bayesplot, with a particular focus on adding visualizations for predictive checks of discrete and categorical outcomes.

Visualizations to be added include:

- Overlaid KDE Plots
- Discrete Rootograms
- PAV-Adjusted Calibration Plots
- PAV-Adjusted Residual Plots

Other project outcomes include documentation updates.

3 Previous Contributions to Stan

3.1 Projpred

- [PR #509](#)
 - Add contribution guidelines to README.Rmd

4 List of Deliverables

Editable [here](#).

deliverable	deadline	notes
Overlaid KDE Plots	NA	NA
Discrete Rootogram	NA	NA
PAV-Adjusted Calibration Plots	NA	NA
PAV-Adjusted Residual PlotsTimeline	NA	NA

5 Timeline

Editable [here](#).

date	task
Thu, May 1, 2025	Project start
Fri, May 16, 2025	Proposal revised. Project deliverables list finalized.
Mon, Jun 2, 2025	Timeline finalized. Work on deliverable 1 begun.
Fri, Jun 13, 2025	NA
Fri, Jul 4, 2025	NA
Fri, Jul 18, 2025	NA
Fri, Aug 1, 2025	NA
Sun, Aug 17, 2025	NA

6 Contributor/Project Fit

6.1 Motivation

The Stan team has made a huge imprint on my work and how I see the world, and I'd love to contribute to the project in any way that I can. I consider GSOC an opportunity to learn a ton and help get Stan to more people. I'm motivated by the prospect of making the Bayesian workflow accessible to anyone who's interested in applying it, and think bayesplot is a really key component of both learning and maintaining a principled way of doing Bayesian inference in the real world.

I also think working on a GSOC project with the Stan team represents an invaluable educational experience. I'm excited to learn new ways to solve difficult problems and expand my capabilities as a data scientist and software engineer.

6.2 General Competencies

6.2.1 R

- Used for personal and professional data projects since 2020. Lots of reps with packages in the tidyverse and tidymodels ecosystems, web scraping, and medium-big data processing tools like DuckDB/Arrow.
- I write R code every day to do data science, but there's a lot of room for growth in my ability to produce code in production.

6.2.2 Stan

- I started learning Stan syntax within the past year. I read Stan models every day, and can write simple models. Most of my exposure to Stan is via brms; the more complex models I've built have been with brms.
- I have a solid understanding of the theory behind prior, retrodictive, and predictive checks in the Bayesian workflow, and experience using bayesplot, tidybayes, ggdist, marginaleffects, and vmc to analyze/visualize models.
- I've recently incorporated sbc into my workflow and have read the papers by Talts, Mondrak, and Bürkner. I get the theoretical need for sbc, can run sbc and begin to interpret its outputs, and am actively developing my understanding of the associated math.

6.2.3 Bayes

- I started learning probability theory and Bayesian inference about a year ago, primarily using Statistical Rethinking, BDA3, the Bayesian Workflow paper, and Michael Betancourt's writing as resources. I took a one-day course on hierarchical modeling from Betancourt earlier this year. I have a solid grasp of the steps in the Bayesian workflow and their justification. I understand and can implement a justifiable version of the workflow to solve problems using observational data.

6.2.4 Git

- I use git to organize all my personal and academic projects. I'm familiar with git best practices and can collaborate with others on a shared repository.
- I have opened issues on open source projects, but these cases have been limited to fixing typos in documentation.

7 Availability

I'm available through the entire GSOC period, and plan on working on this project at least 20 hours each week. I'll be happy to work with the Stan team to adjust the total working hours as needed.

8 Other GSOC Applications

I'm not applying for any other GSOC projects.

9 After GSOC

I hope to continue contributing to the broader Stan project as I figure out the next steps in my professional or academic life.