Shinystan module structure

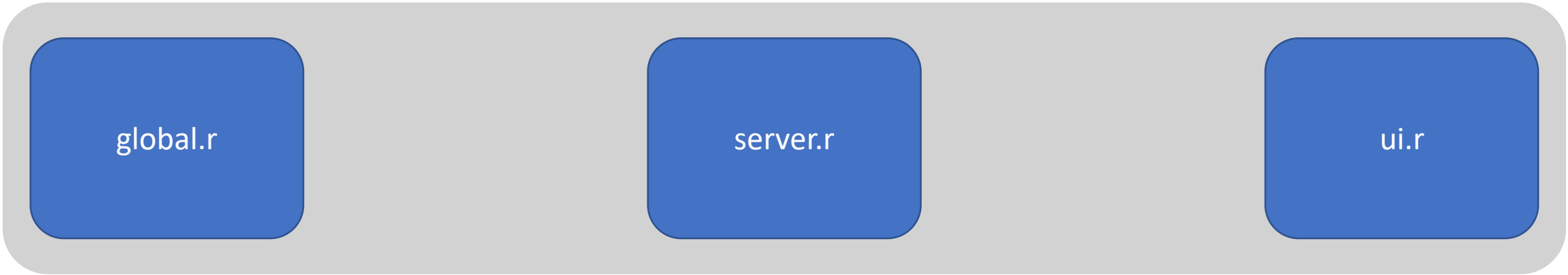
Outline

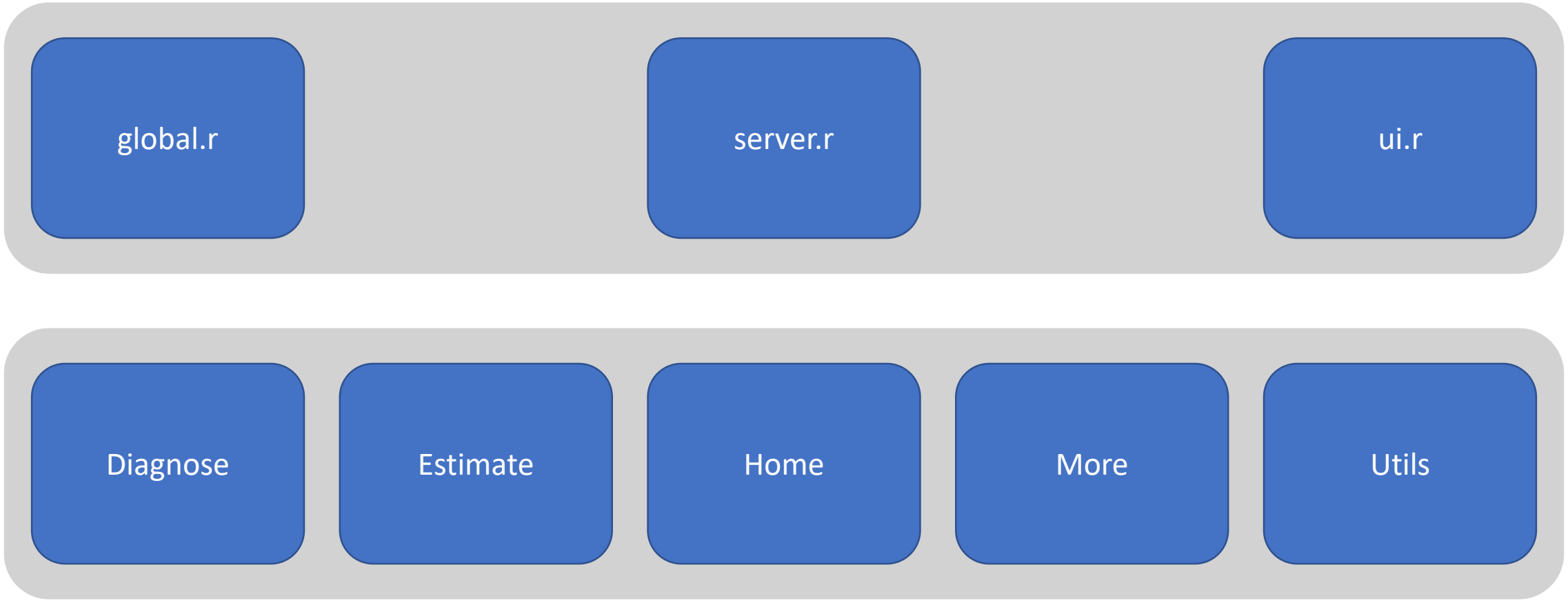
- Shinystan 3.0
  - Backend changes
  - Modules, why?
  - Report generation

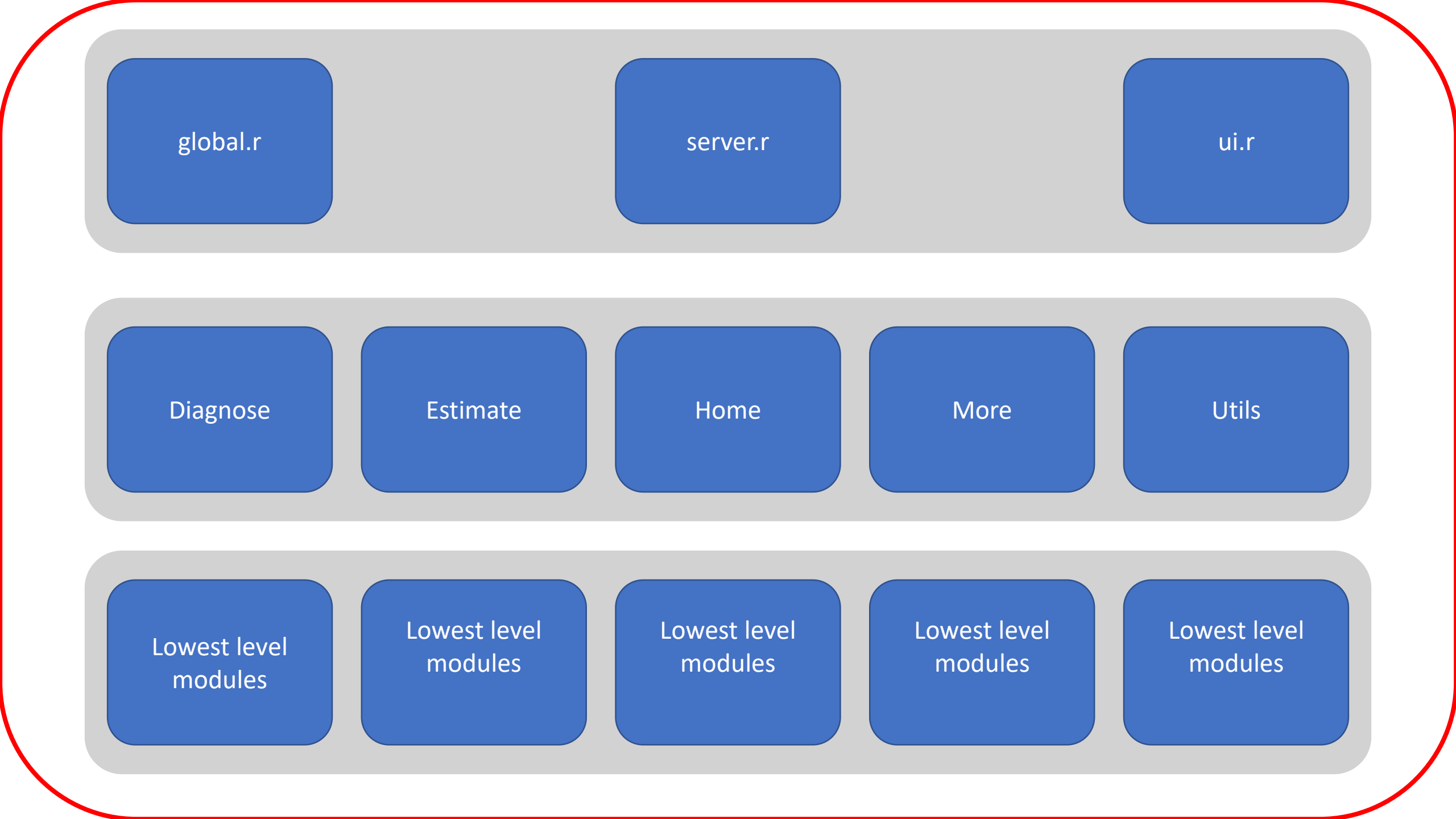
- On workflow
  - Where does shinystan fit in?
  - Adding apendices

Why?

- Make backend more stable / sustainable
- Easier to add / remove features
  - No naming issues due to different environments
- Base everything on bayesplot
- Add report option


global.r
server.r
ui.r


global.r
server.r
ui.r
Diagnose
Estimate
Home
More
Utils


global.r
server.r
ui.r
Diagnose
Estimate
Home
More
Utils
Lowest level modules
Lowest level modules
Lowest level modules
Lowest level modules
Lowest level modules

How do modules work?

- Write modules which have UI and Server part
- source("MODULES/DIAGNOSE/acceptance.r", local = TRUE)

UI

```
acceptanceUI <- function(id){
  ns <- NS(id)
  .....
  }


NESTED:
…
fluidRow(
        align = "right",
        plotOptionsUI(ns("options"))
      )
…
```

SERVER

```
acceptance <- function(input, output, session){
  …
  }
```

NESTED:

```
…
visualOptions <- callModule(plotOptions, "options")
…
```

What can you assume?

- Shinystan object
  - shinystan::as.shinystan()

model_name	posterior_sample	n_chain	user_model_info
param_names	summary	n_iter	model_code
param_dims	sampler_params	n_warmup	misc

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Model:

Demo (Eight Schools)

- 65 of 4000 iterations ended with a divergence (1.6%).
- 1 parameters have an effective sample size less than 10% of the total sample size.

DIAGNOSE ESTIMATE MORE

Close

SHINYSTAN | DIAGNOSE | ESTIMATE | MORE

Plots | Stats | Report

NUTS/HMC

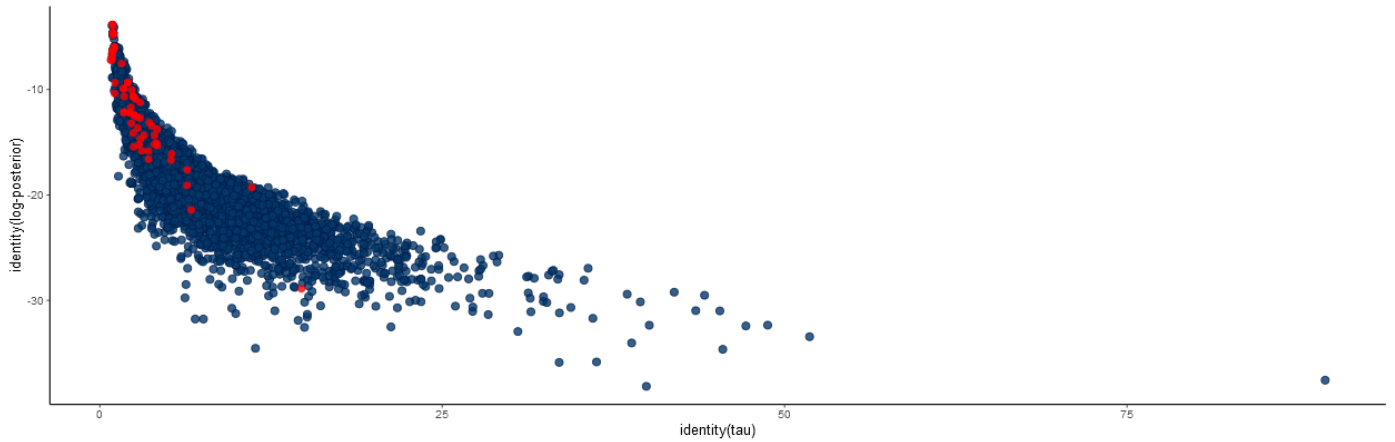
- **Divergent Scatter**
- Parallel Coordinates
- Pairs
- Divergence Information
- Energy Information
- Treedepth Information
- Step Size Information
- Acceptance Information

MCMC

- Autocorrelation
- Trace Plots
- Rank Plots
- $\hat{R}$, n_{eff}, se_{mean}

Parameter: tau, log-posterior | Transform X: identity | Transform Y: identity | All chains: 0

Show/Hide Options



Show/Hide Caption

Include in report?

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots | Stats | Report

NUTS/HMC

- Divergent Scatter
- Parallel Coordinates
- Pairs
- Divergence Information
- Energy Information
- Treedepth Information
- Step Size Information
- Acceptance Information

MCMC

- Autocorrelation
- **Trace Plots**
- Rank Plots
- $\hat{R}$, n_{eff}, se_{mean}

Parameter: tau, log-posterior

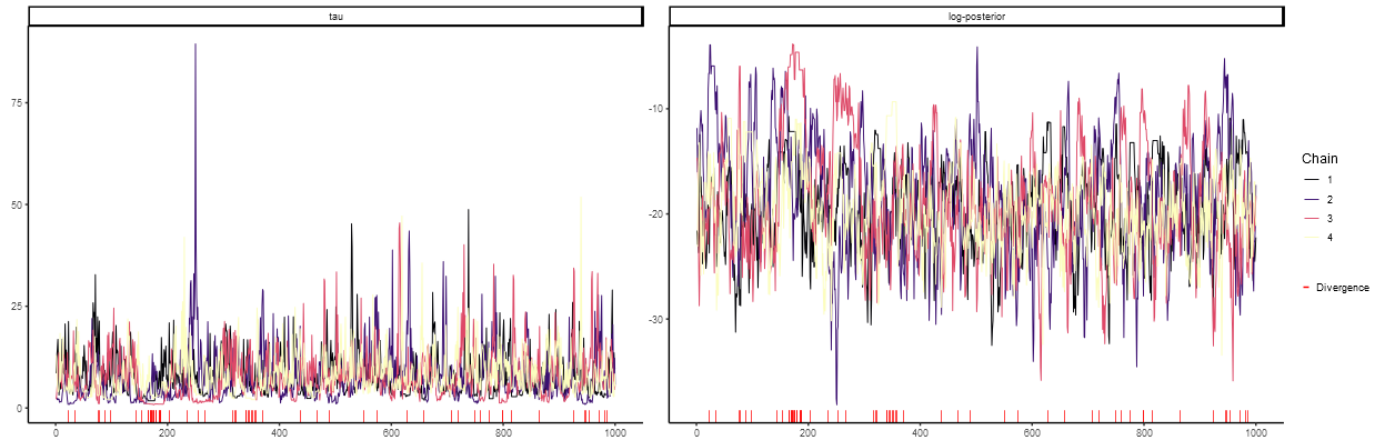
All chains: 0

Show/Hide Options

Select Theme: classic

Select Colors: viridisA

Select Divergent Color: red



Show/Hide Caption

Include in report?

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots Stats Report

NUTS/HMC

- Divergent Scatter
- Parallel Coordinates
- Pairs
- Divergence Information
- Energy Information
- Treedepth Information
- Step Size Information
- **Acceptance Information**

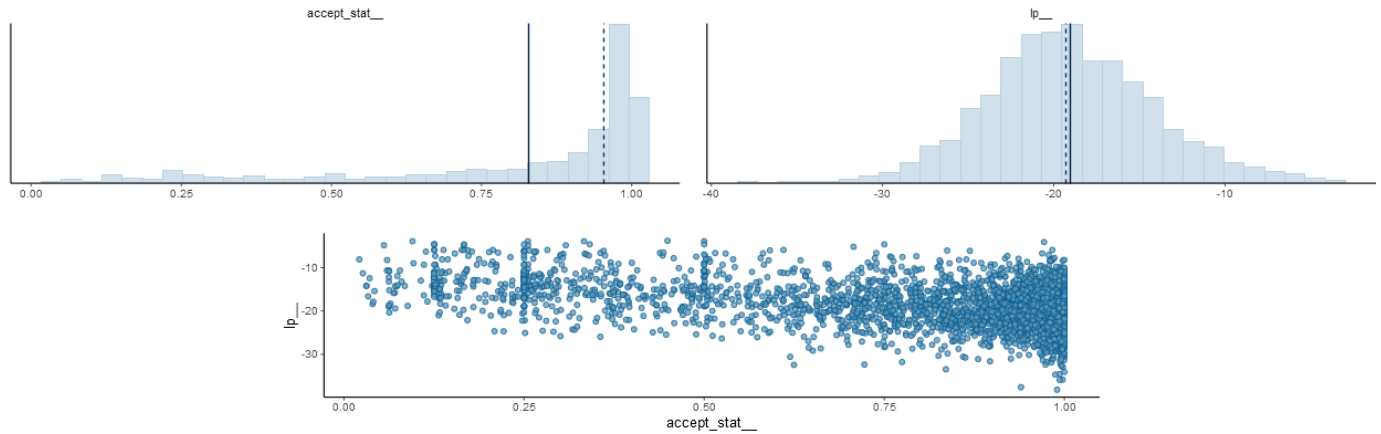
MCMC

- Autocorrelation
- Trace Plots
- Rank Plots
- $\hat{R}$, n_{eff}, se_{mean}

All chains

0

Show/Hide Options



Show/Hide Caption

These are plots of the *acceptance statistic* (top leftpanel), the *log-posterior* (top right panel), and, the *acceptance statistic* (x-axis bottom panel) against the *log-posterior* (y-axis bottom panel) for all chains. The vertical lines indicate the mean (solid line) and median (dashed line). A bad plot would show a relationship between the acceptance statistic and the log-posterior. This might be indicative of poor exploration of parts of the posterior which might be might be mitigated by reparametrization or adaptation of the step size. If many proposals are rejected the integrator step size might be too large and the posterior might not be fully explored. If the acceptance rate is very high this might be indicative of inefficient sampling. The target Metropolis acceptance rate can be set with the *adapt_delta* control option. For more information see https://mc-stan.org/docs/2_19/reference-manual/hmc-algorithm-parameters.html.

Include in report?

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots Stats Report

NUTS/HMC

All NUTS/HMC stats

MCMC

$\hat{R}$, n_{eff}, se_{mean}

Statistic

Mean SD Max Min

Warmup

Omit Include

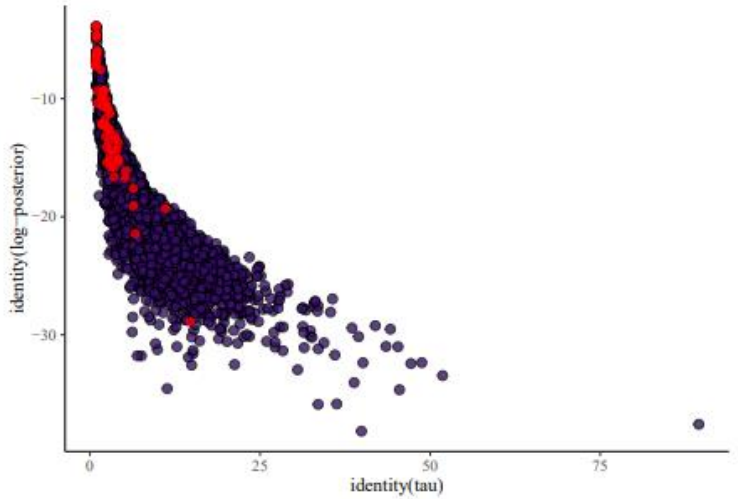
Decimals

2

	accept_stat	stepsize	treedepth	n_leapfrog	divergent	energy
All chains	0.83	0.24	3.64	17.44	0.02	24.00
chain1	0.74	0.34	3.12	10.89	0.02	24.37
chain2	0.93	0.12	4.36	27.25	0.00	23.40
chain3	0.90	0.16	3.97	20.90	0.02	23.69
chain4	0.73	0.35	3.08	10.73	0.02	24.54

Close
SHINYSTAN
DIAGNOSE
ESTIMATE
MORE
Plots
Stats
Report
ShinyStan
Download Plots
Download Report
Download RDS

ShinyStan Diagnostics Report



This is a plot of MCMC draws of tau (x-axis) against log-posterior (y-axis). The red colored draws represent, if present, divergent transitions. Divergent transitions can indicate problems for the validity of the results. A good plot would show no divergent transitions. A bad plot would show divergent transitions in a systematic patern. For more information see https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup.

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots Stats Report

Intervals

Areas

Scatter

Density

Histogram

Parameter

mu theta[1] theta[6] theta[7] theta[8] tau

theta

ALL (remaining) theta

theta[2]

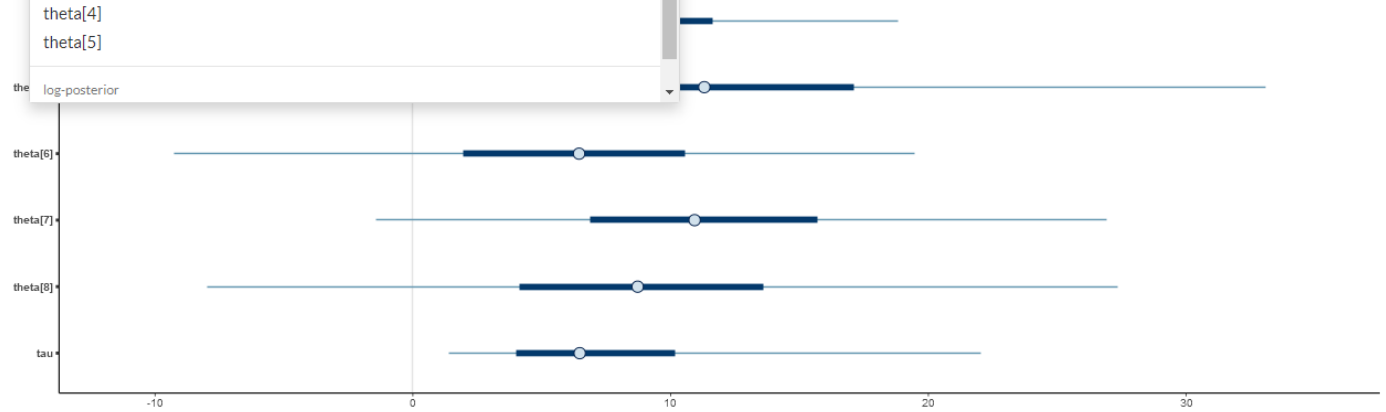
theta[3]

theta[4]

theta[5]

log-posterior

Show/Hide Options



Show/Hide Caption

Include in report?

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots Stats Report

Intervals
Areas
Scatter
Density
Histogram

Parameter

mu theta[1] theta[2] theta[3] theta[4] theta[5] theta[6] theta[7] theta[8] tau

Show/Hide Options

Select Theme

bayesplot default

Select Colors

blue

Point Estimate

Median Mean None

Plot Type

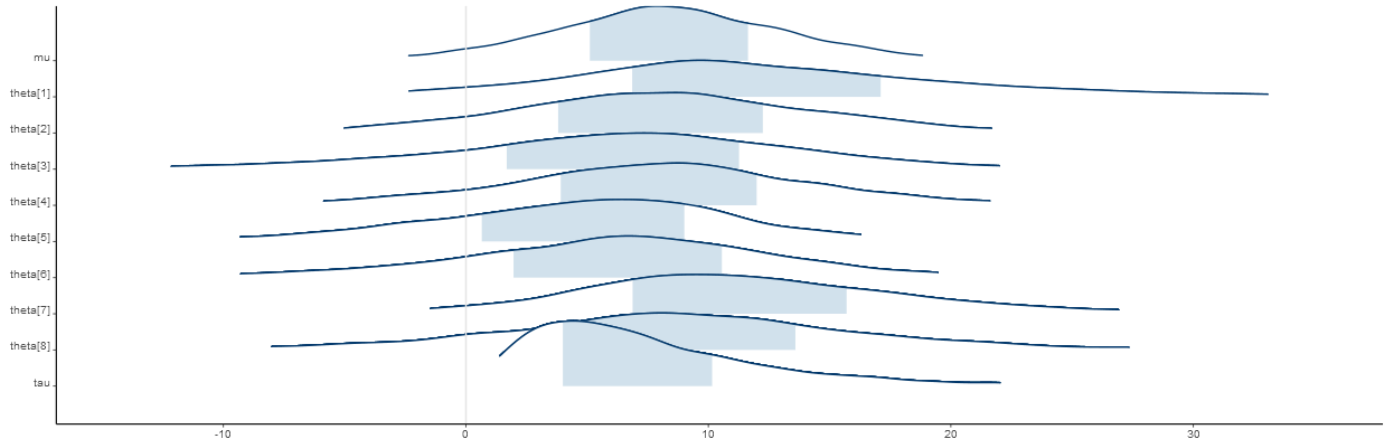
Areas Ridges

Posterior Interval (inner)

50% 95%

Posterior Interval (outer)

50% 95%



Show/Hide Caption

Close

SHINYSTAN DIAGNOSE ESTIMATE MORE

Plots Stats Report

Posterior Summary Statistics

Generate LaTeX table

Parameter

mu theta[1] theta[2] theta[3] theta[4] theta[5] theta[6] theta[7] theta[8] tau

Caption

Inlcude a caption.

Decimals

1

Columns

Posterior mean Monte Carlo error Posterior standard deviation
Quantile: 2.5% Quantile: 25% Quantile: 50% Quantile: 75%
Quantile: 97.5% Effective sample size Rhat

Packages

Booktabs
Longtable

```
% latex table generated in R 3.6.1 by xtable 1.8-3 package
% Mon Sep 16 12:19:37 2019
\begin{table}[ht]
\centering
\begin{tabular}{rlllll}
  \hline
 & mean & sd & 2.5\% & 50\% & 97.5\% & Rhat \\
  \hline
mu & 8.4 & 5.4 & -2.4 & 8.2 & 18.8 & 1.0 \\
  theta[1] & 12.5 & 8.7 & -2.3 & 11.3 & 33.1 & 1.0 \\
  theta[2] & 8.1 & 6.7 & -5.0 & 8.0 & 21.7 & 1.0 \\
  theta[3] & 6.2 & 8.4 & -12.1 & 6.6 & 22.0 & 1.0 \\
  theta[4] & 7.9 & 6.8 & -5.9 & 8.1 & 21.6 & 1.0 \\
  theta[5] & 4.7 & 6.4 & -9.3 & 5.2 & 16.3 & 1.0 \\
  theta[6] & 6.1 & 7.1 & -9.3 & 6.4 & 19.5 & 1.0 \\
  theta[7] & 11.5 & 7.0 & -1.4 & 10.9 & 26.9 & 1.0 \\
  theta[8] & 8.9 & 8.3 & -8.0 & 8.7 & 27.3 & 1.0 \\
  tau & 7.9 & 5.8 & 1.4 & 6.5 & 22.0 & 1.0 \\
   \hline
\end{tabular}
\caption{Inlcude a caption.}
\end{table}
```

Report Generation

- Just one line of code
- Report for N worst or all parameters of model
  - Or specific parameters
- On diagnostics, estimates or both
- PDF, Word or HTML
- `generate_report()`

generate_report()

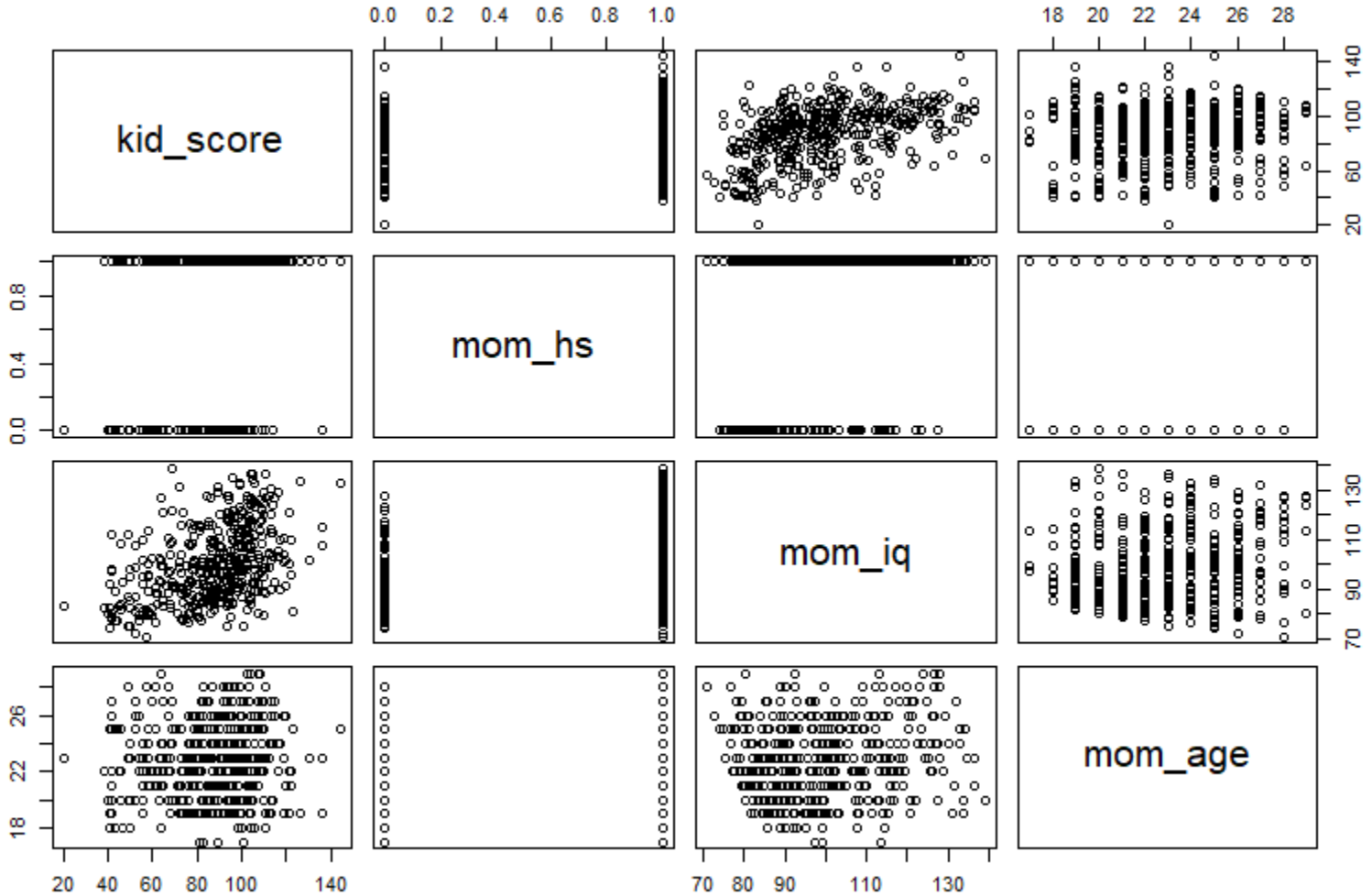
Arguments	
sso	A shinystan object.
n_param	On how many parameters do you want the report to be? To print the report for all parameters put 'n_param = Inf' or n_param = "all"'.
pars	An optional character vector of parameter names. If no names are specified the n_param count will be used. Arguments passed to pars take precedence over the n_param argument.
output_format	What type of report would you like? The options are 'html_document', 'pdf_document' and 'word_document'.
View	Do you want to open the report after it is generated?
report_type	What type of report would you like? The options are 'diagnose', 'estimate' and 'both'. The default is 'diagnose'.

Workflow

- Hopefully, live coding works
    - Or we have the next few slides

- We use a data set from rstanarm on IQ of kids.
    - rstanarm::kidiq

Visualize your data



Does it make sense?

IQ ~ Normal(100, sd = 15)

```
> kidiq$kid_score[order(kidiq$kid_score)]
[1] 20 38 40 41 41 41 42 42 42 42 42 42 42 43 43 44 45 46 46 46 47 49 49 49 50 50 50 50 50 52 52
[32] 52 52 52 54 54 54 56 56 56 56 56 56 56 56 57 57 58 58 58 58 58 58 58 58 59 60 60 60 61 61 61 61
[63] 62 62 63 63 64 64 64 64 64 64 65 65 65 65 65 66 67 67 67 67 68 68 68 68 68 69 69 69 69 69 69
[94] 70 70 71 72 72 72 72 73 73 73 73 73 73 74 74 74 74 75 75 75 75 76 76 76 76 76 76 76 76 77 77
[125] 77 77 78 78 78 78 78 78 78 78 79 79 79 79 79 80 80 80 80 80 81 81 81 81 81 82 82 82 82 83 83
[156] 83 83 83 83 83 83 83 83 84 84 84 84 84 84 84 85 85 85 85 85 85 86 86 86 86 86 86 86 86 86 86
[187] 87 87 87 87 87 87 87 87 87 87 87 87 87 88 88 88 88 88 88 89 89 89 89 89 89 90 90 90 90 90
[218] 90 90 90 90 91 91 91 91 91 92 92 92 92 92 92 92 92 92 92 93 93 93 93 93 94 94 94 94 94 94 94
[249] 94 94 94 94 94 94 94 94 94 94 94 95 95 95 95 95 95 95 95 95 95 95 96 96 96 96 96 96 97 97 97
[280] 97 97 97 97 97 97 98 98 98 98 98 98 98 98 98 98 98 99 99 99 99 99 99 99 99 99 100 100 100 100 100
[311] 100 100 100 100 100 100 100 101 101 101 101 101 101 102 102 102 102 102 102 102 102 102 102 102 102 102 102 103 103 103 103
[342] 104 104 104 104 104 104 104 104 104 104 104 104 104 104 104 105 105 105 105 105 106 106 106 106 106 106 106 106 106 106 107
[373] 107 107 107 107 107 108 108 108 108 109 109 109 109 109 109 109 110 110 110 110 110 110 110 111 111 111 111 112 112 112 112
[404] 113 113 113 113 113 113 114 114 114 115 115 116 116 116 117 117 119 119 119 120 120 121 121 122 122 123 126 130 136 136 144
> pnorm(20, 100, 15)
[1] 4.821303e-08
> pnorm(50, 100, 15)
[1] 0.0004290603
```

Does it make sense?

IQ ~ Normal(100, sd = 15)

```
> psych::describe(kidiq)
           vars   n   mean    sd median trimmed   mad   min    max  range  skew kurtosis   se
kid_score     1 434  86.80 20.41  90.00   87.93 19.27 20.00 144.00 124.00 -0.46    -0.19 0.98
mom_hs        2 434   0.79  0.41   1.00    0.86  0.00  0.00   1.00   1.00 -1.39    -0.07 0.02
mom_iq        3 434 100.00 15.00  97.92   99.11 15.89 71.04 138.89  67.86  0.47    -0.59 0.72
mom_age       4 434  22.79  2.70  23.00   22.71  2.97 17.00  29.00  12.00  0.18    -0.65 0.13
```

Decide on model

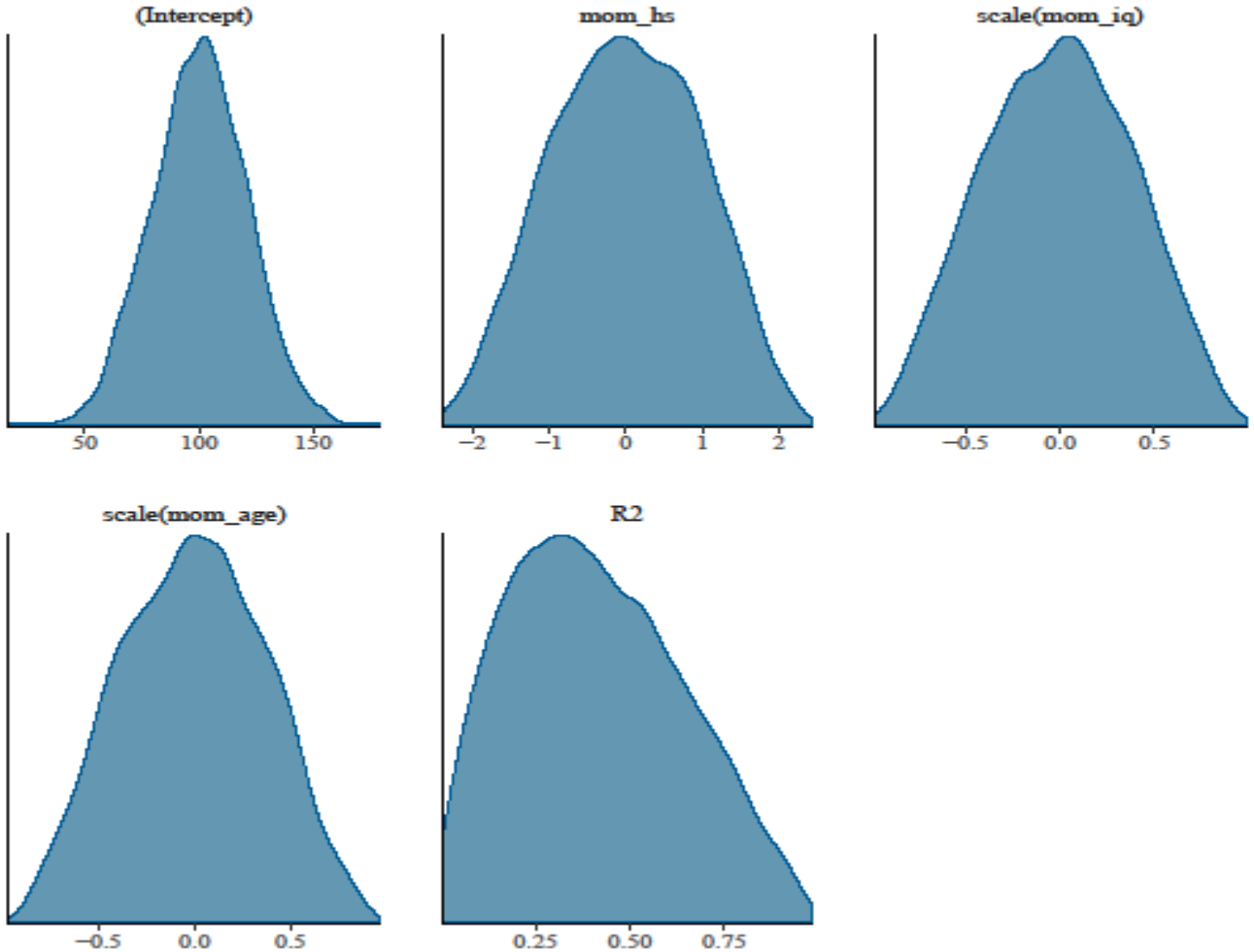
- Kidiq as outcome
- Simple regression
- Predictors
  - Did mom go to high school yes/no
  - IQ of mom – standardize?
  - Age of mom – standardize?
- We can use rstanarm!

Prior Predictive Distributions

```
# uniform prior on R^2 if NULL, .3 R^2 seems reasonable in social sciences

fit_priorPD <- stan_lm(kid_score ~ mom_hs + mom_iq + mom_age,
                       data = kidiq,
                       prior = R2(.3),
                       prior_intercept = normal(100, 20),
                       prior_PD = TRUE,
                       seed = 13031990)
```

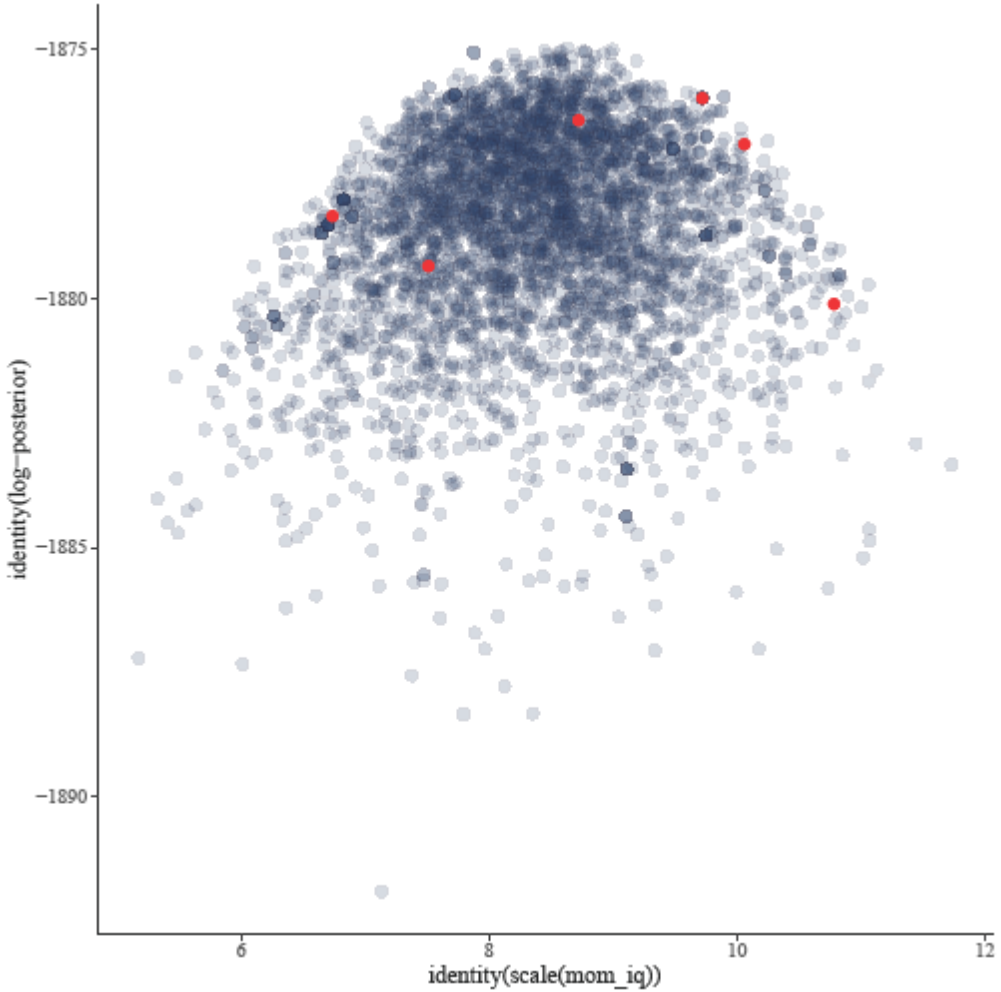
Prior Predictive Distributions



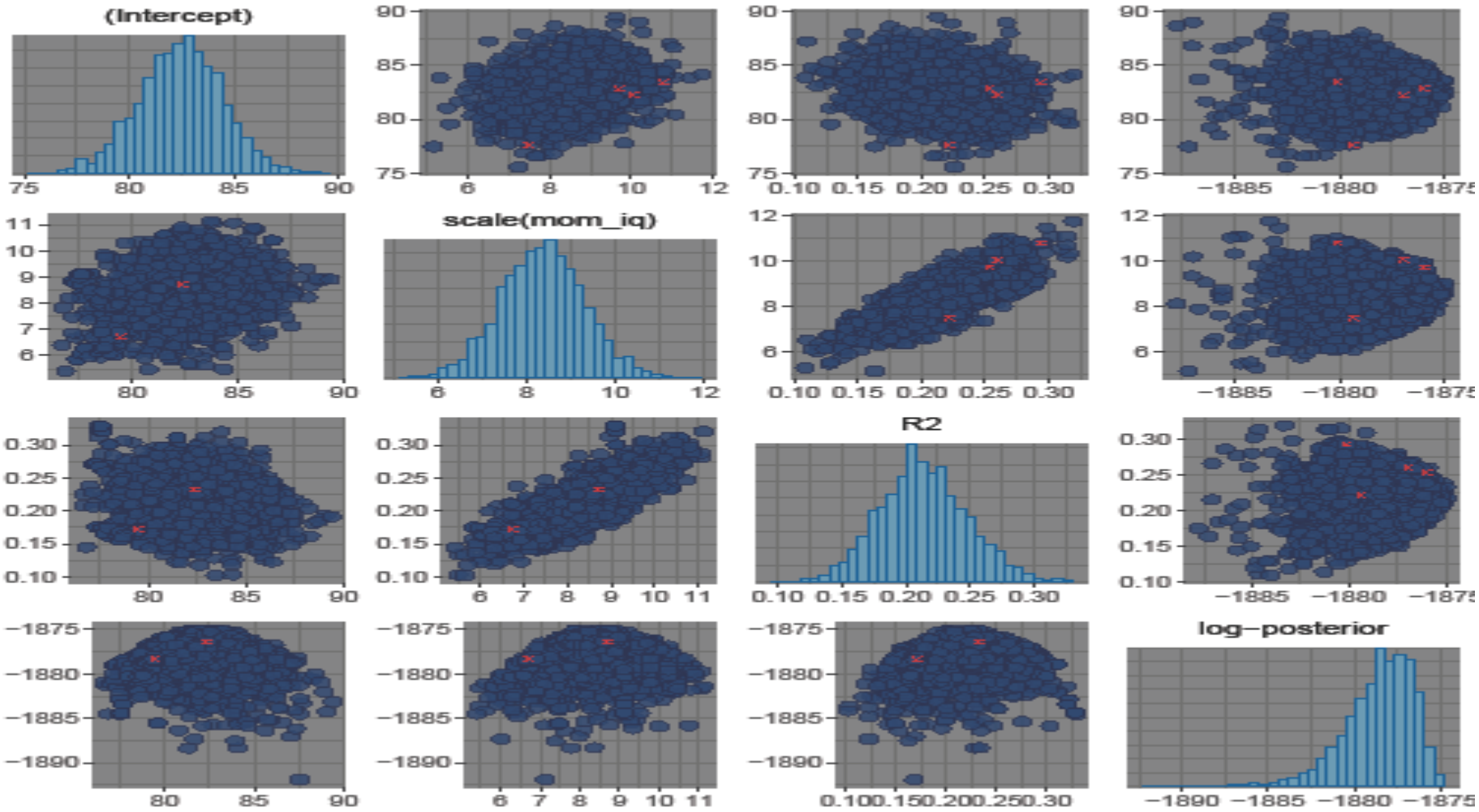
Estimate your model

```
fit <- stan_lm(kid_score ~ mom_hs + scale(mom_iq) + scale(mom_age),
               data = kidiq,
               prior = R2(.3), seed = 123456, iter = 2000, adapt_delta = .7,
               prior_intercept = normal(100, 20),
               prior_PD = FALSE)
```

Diagnose


−1875
−1880
−1885
−1890
identity(log−posterior)
6
8
10
12
identity(scale(mom_iq))

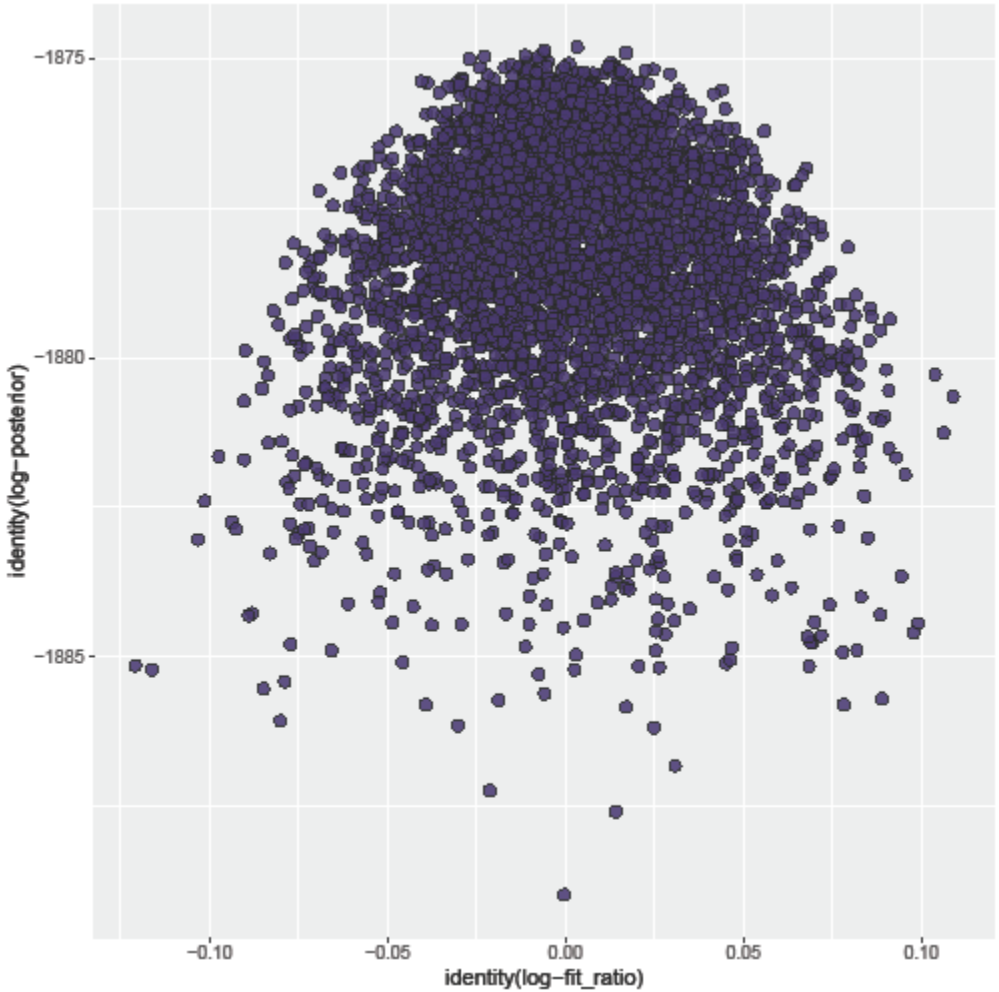
Diagnose



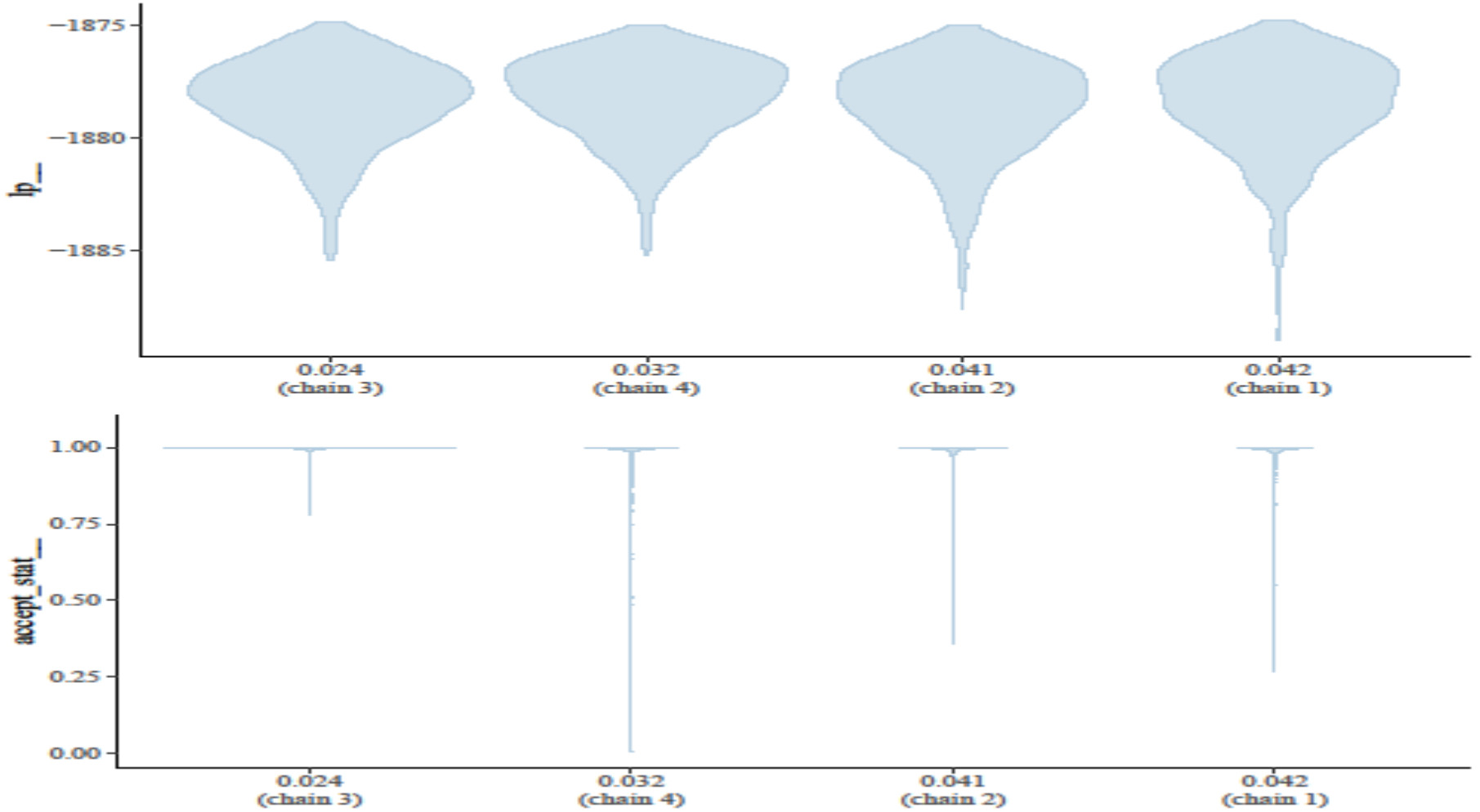
Run again

```
fit <- stan_lm(kid_score ~ mom_hs + scale(mom_iq) + scale(mom_age),
               data = kidiq,
               prior = R2(.3), seed = 123456, iter = 2000, adapt_delta = .99,
               prior_intercept = normal(100, 20),
               prior_PD = FALSE)
```

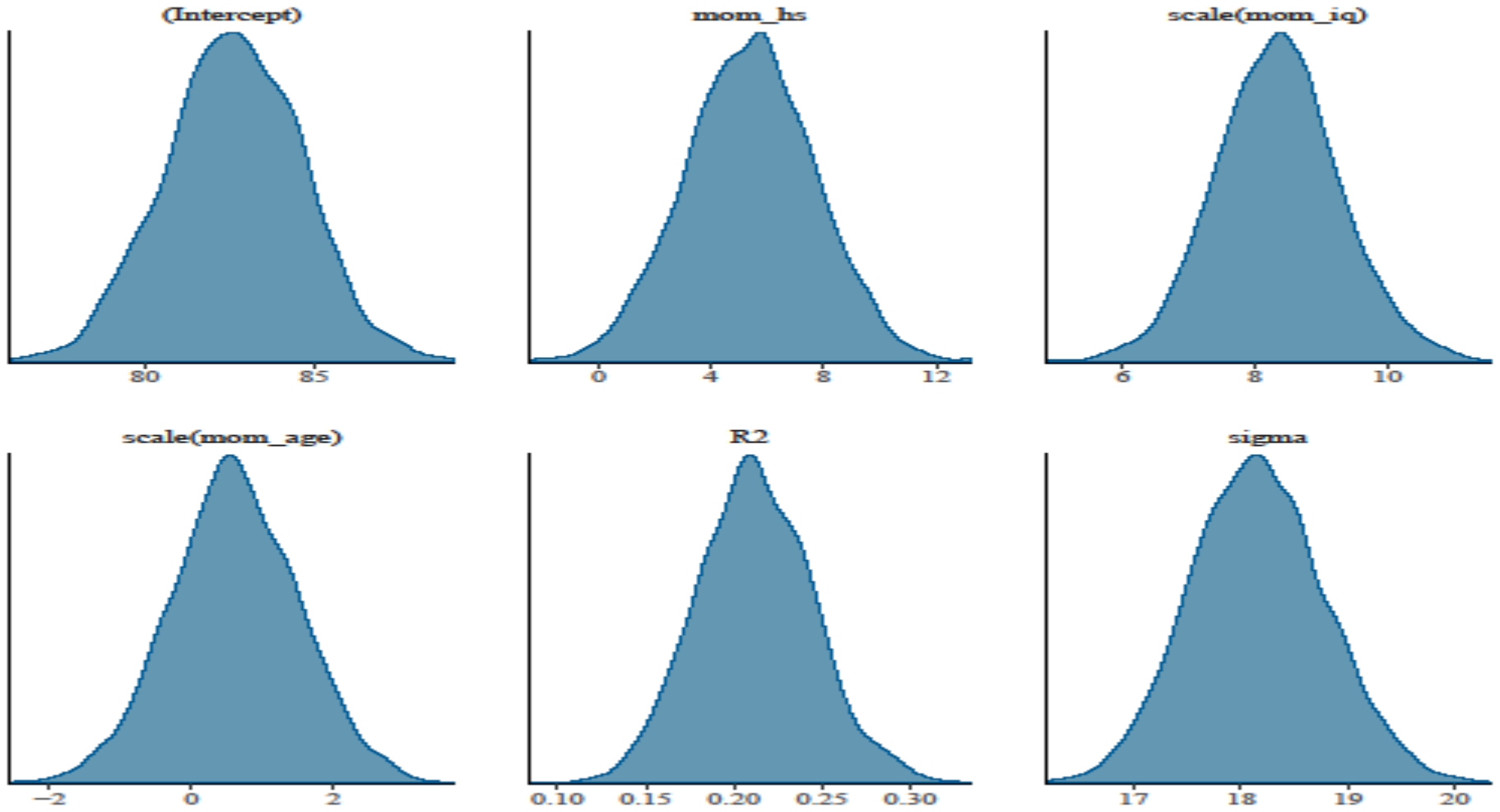
Diagnose again



Diagnose again



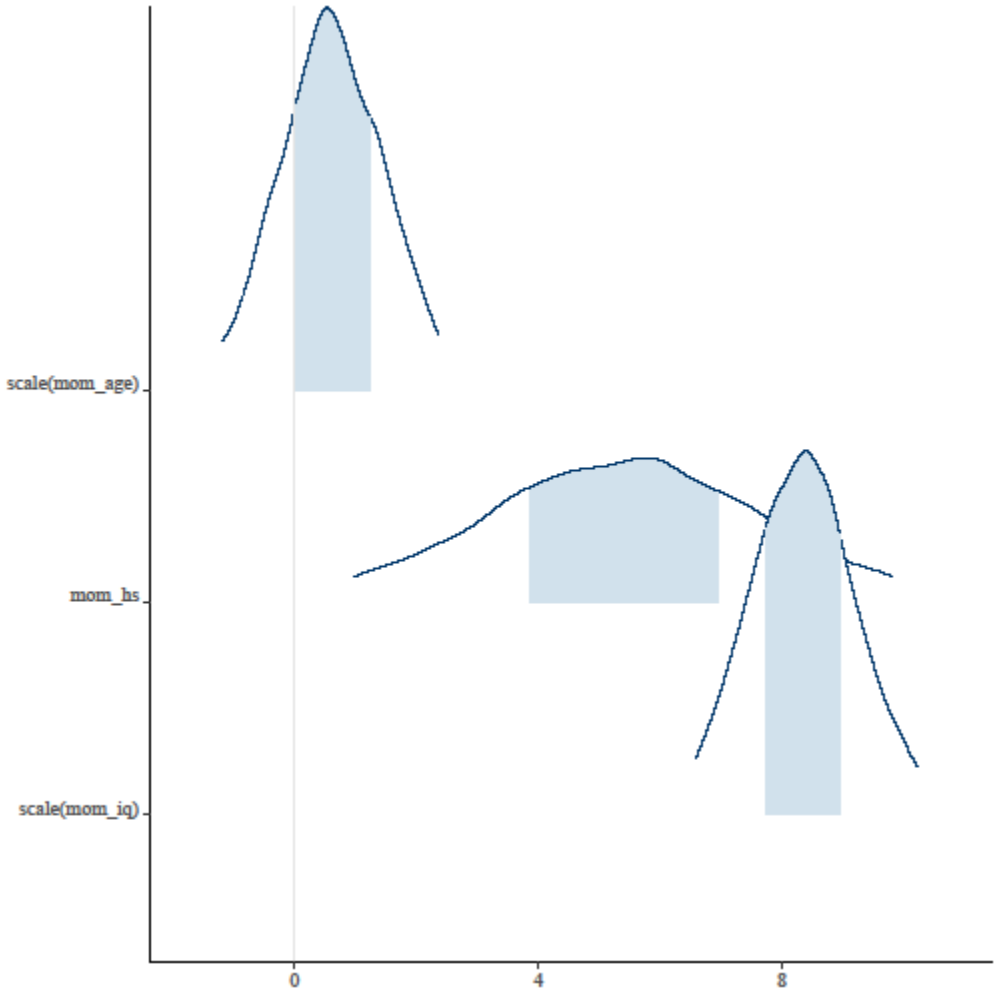
Look at estimtes



Look at estimtes

	mean	sd	2.5%	50%	97.5%	n_eff	Rhat
(Intercept)	82.62	1.99	78.73	82.63	86.41	2525.23	1.00
mom_hs	5.39	2.28	0.98	5.43	9.79	2554.19	1.00
scale(mom_iq)	8.34	0.92	6.57	8.34	10.23	2322.26	1.00
scale(mom_age)	0.61	0.90	-1.18	0.59	2.36	3330.14	1.00
sigma	18.16	0.62	16.98	18.15	19.41	3105.02	1.00

Look at estimtes



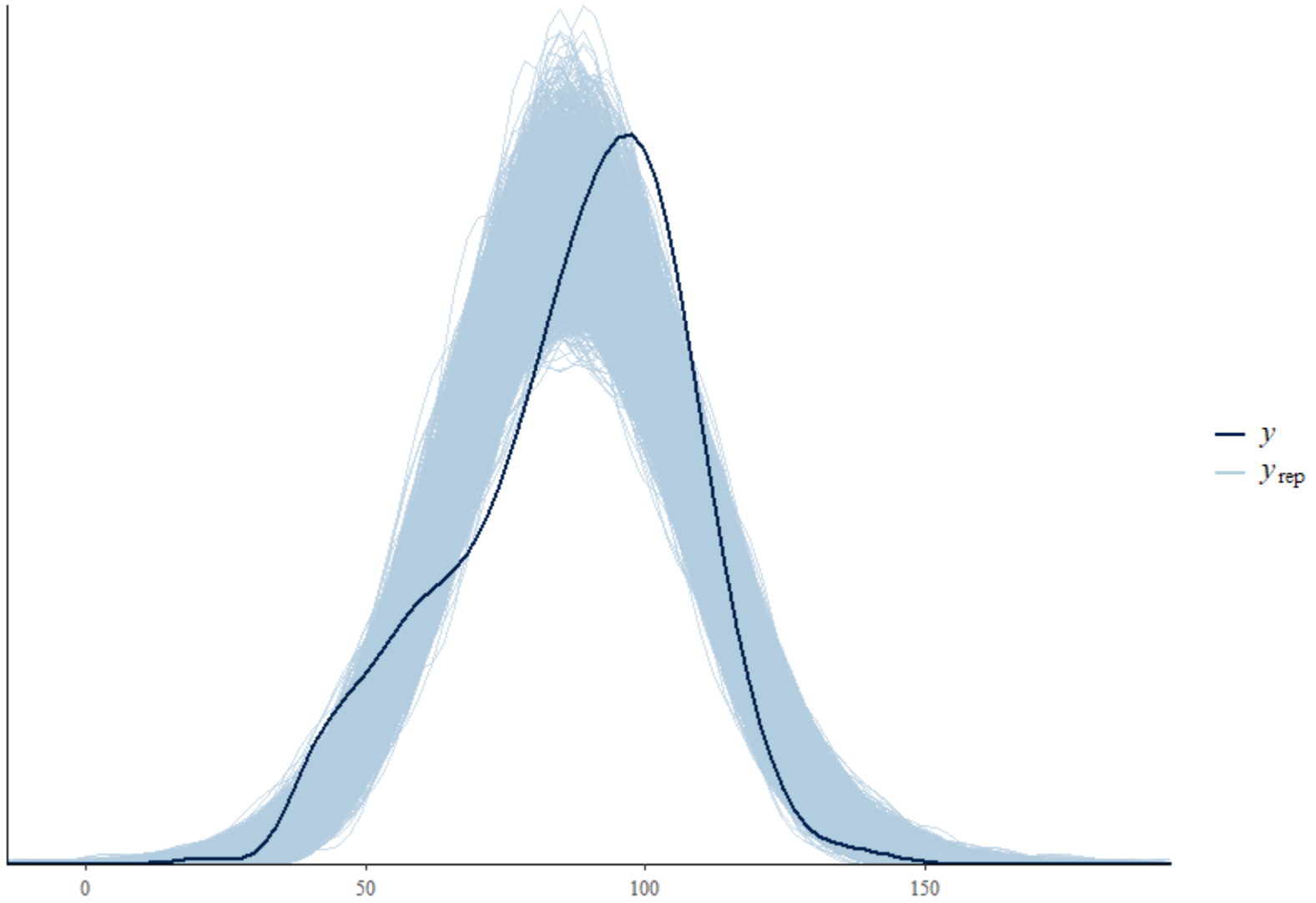
Posterior Predictive

- Not in shinystan 3.0 at the moment

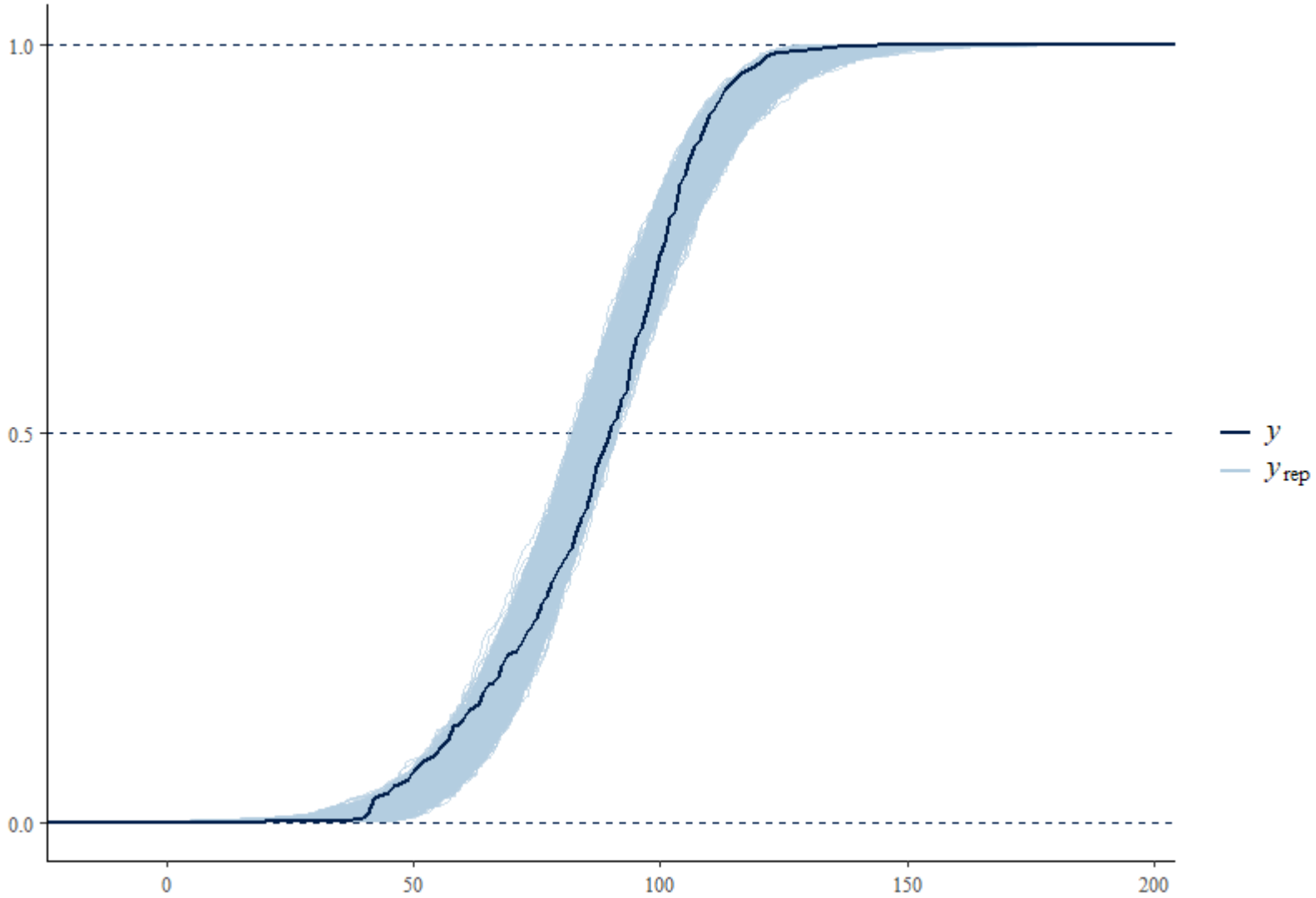
- Bayesplot very useful:
    - posterior_predict()
    - ppc_.....

```
fit_postPD <- posterior_predict(fit)
bayesplot::ppc_dens_overlay(kidiq$kid_score, fit_postPD, n_dens = 100)
```

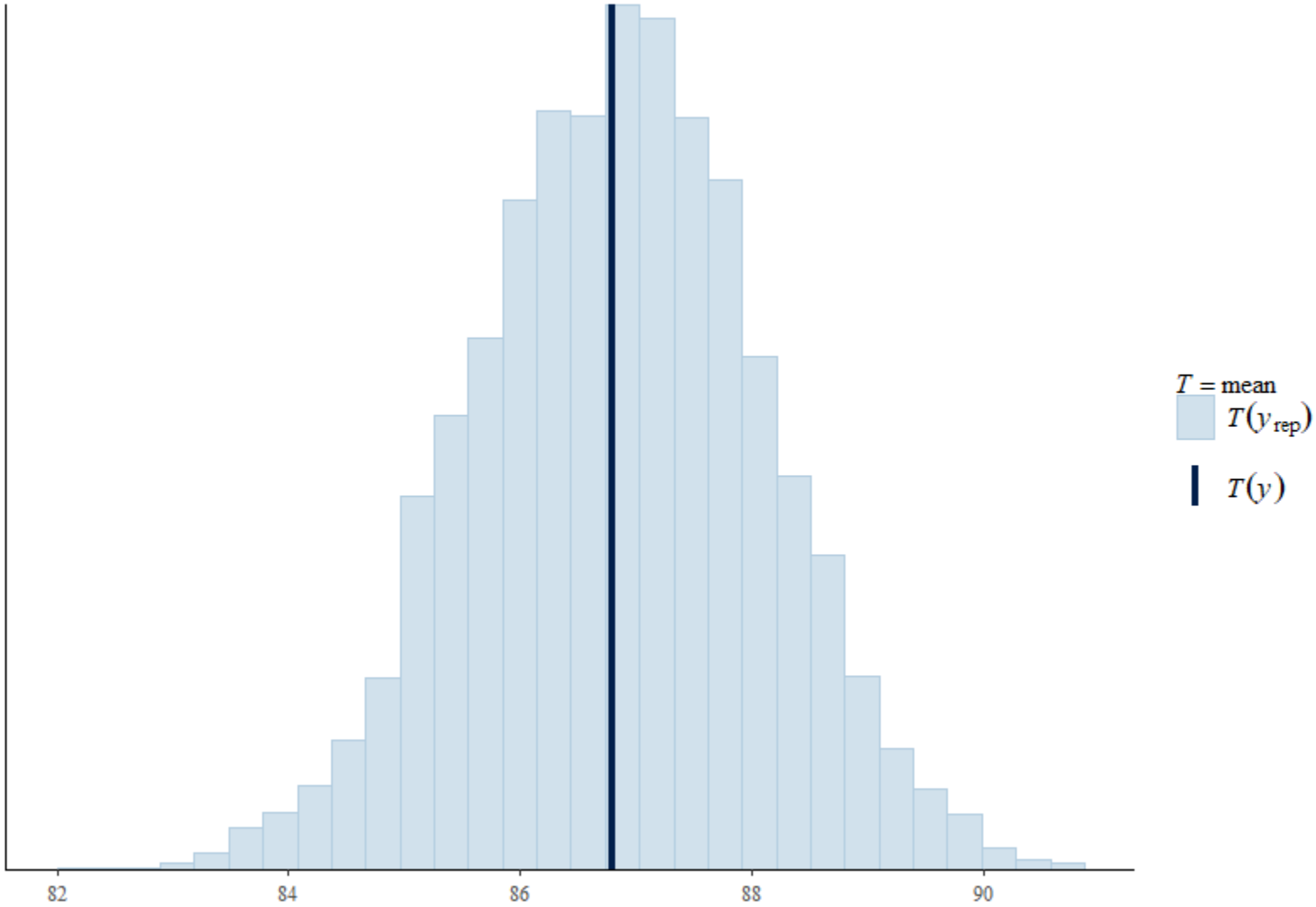
Posterior Predictive



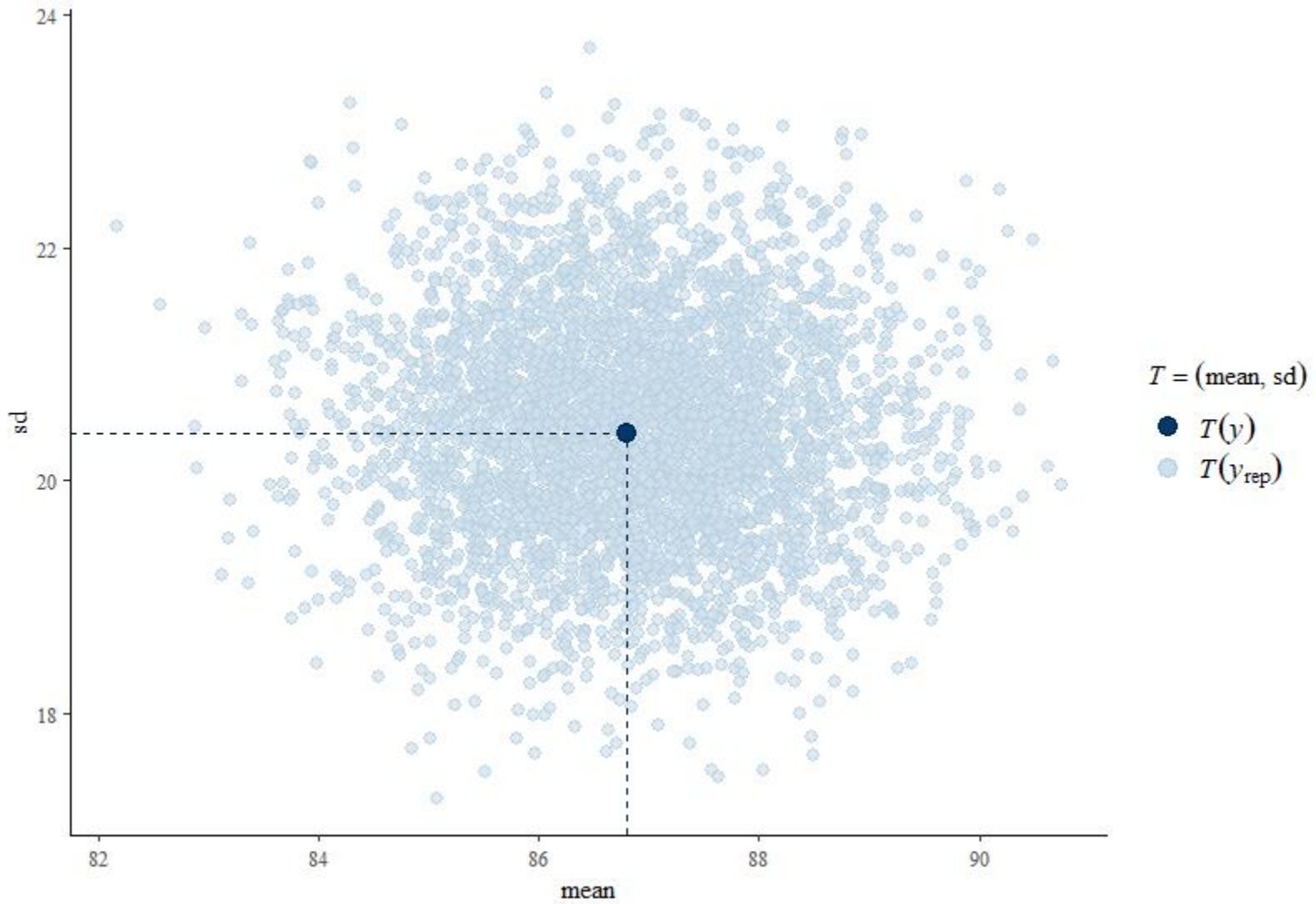
Posterior Predictive



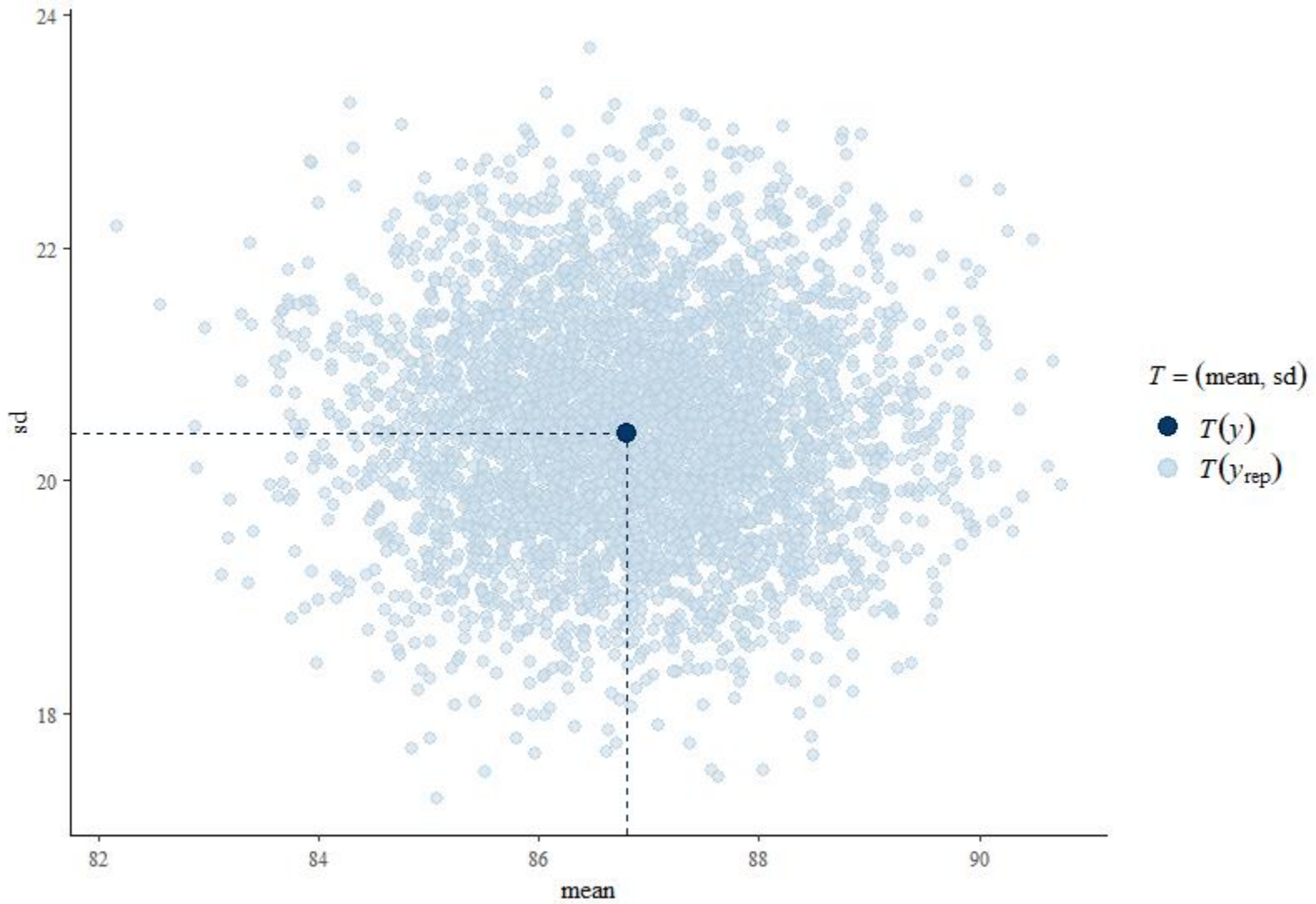
Posterior Predictive



Posterior Predictive



Posterior Predictive



Wishlist

- Get you all to break it….

Try it out:

- remove.pacakges("shinystan")
- devtools::install_github("stan-dev/shinystan@testing-modules", local = TRUE)
- shinystan::launch_shinystan_demo()