Visualizing a Bayesian model

BAYESIAN REGRESSION MODELING WITH RSTANARM

Jake Thompson
Psychometrician, ATLAS, University of Kansas

```
stan_model <- stan_glm(kid_score ~ mom_iq, data = kidiq)
tidy(stan_model)
```

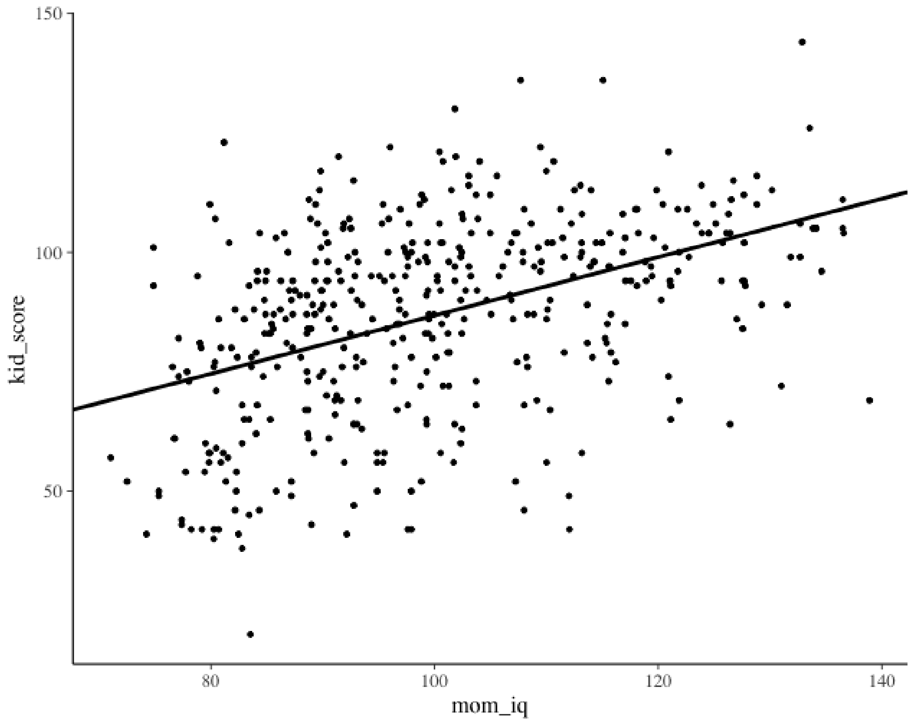
```
# A tibble: 2 x 3
  term        estimate std.error
  <chr>          <dbl>     <dbl>
1 (Intercept)   25.7      5.92
2 mom_iq         0.611    0.0590
```

```
tidy_coef <- tidy(stan_model)
model_intercept <- tidy_coef$estimate[1]
model_intercept
model_slope <- tidy_coef$estimate[2]
model_slope
```

```
25.67857
0.6110473
```

Creating a plot

```
ggplot(kidiq, aes(x = mom_iq, y = kid_score)) +
  geom_point() +
  geom_abline(intercept = model_intercept, slope = model_slope)
```



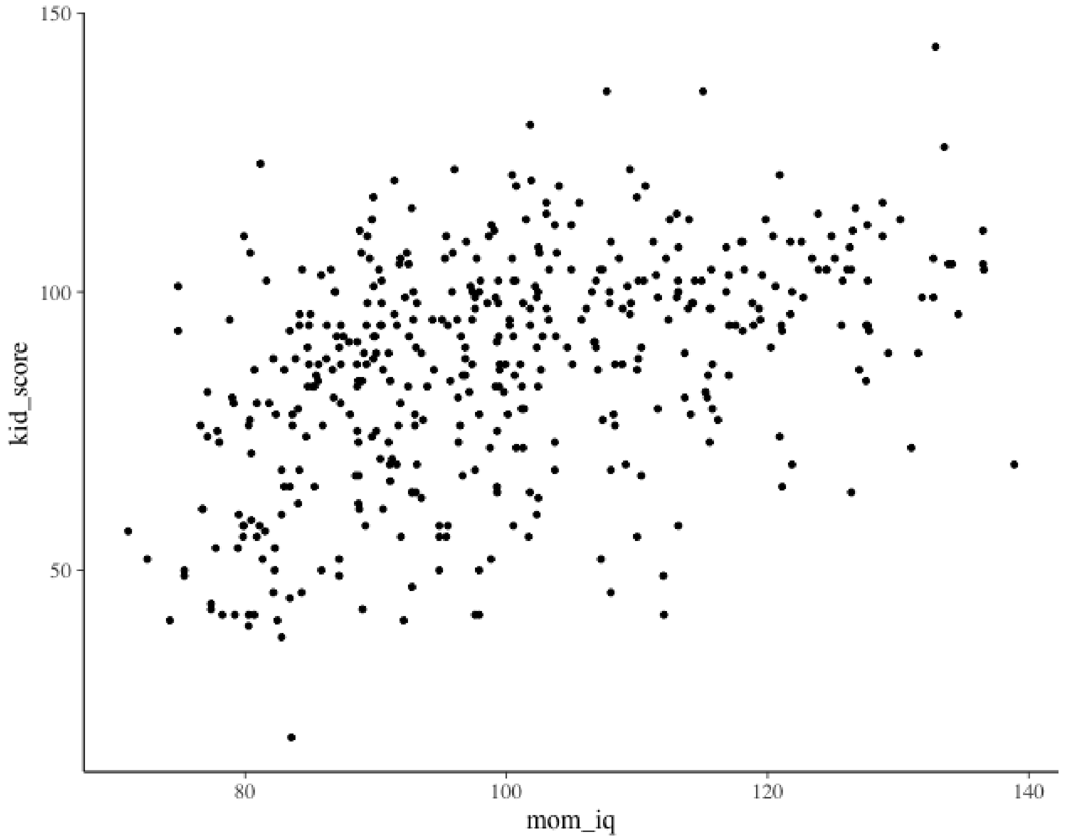
Plotting uncertainty

```
draws <- spread_draws(stan_model, `(Intercept)`, mom_iq)
draws
```

```
# A tibble: 4,000 x 5
   .chain .iteration .draw `(Intercept)` mom_iq
    <int>      <int> <int>         <dbl>  <dbl>
 1      1          1     1          28.2  0.586
 2      1          2     2          28.7  0.593
 3      1          3     3          13.5  0.735
 4      1          4     4          30.3  0.564
 5      1          5     5          34.5  0.522
 6      1          6     6          19.2  0.669
 7      1          7     7          34.8  0.523
 8      1          8     8          16.3  0.707
# ... with 3,992 more rows
```

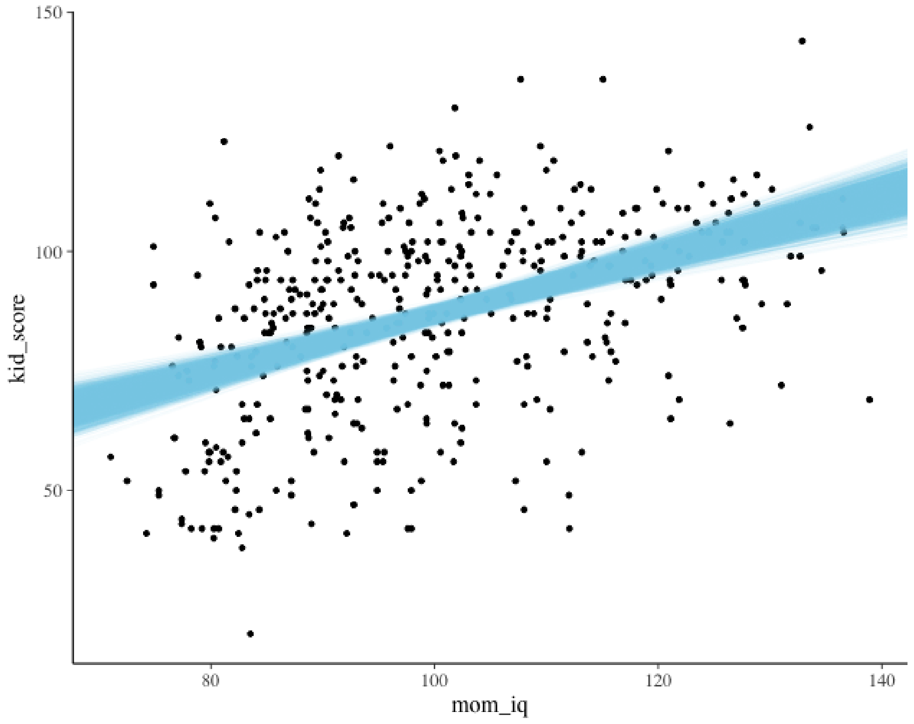
Plotting uncertainty

```
ggplot(kidiq, aes(x = mom_iq, y = kid_score)) +
  geom_point()
```



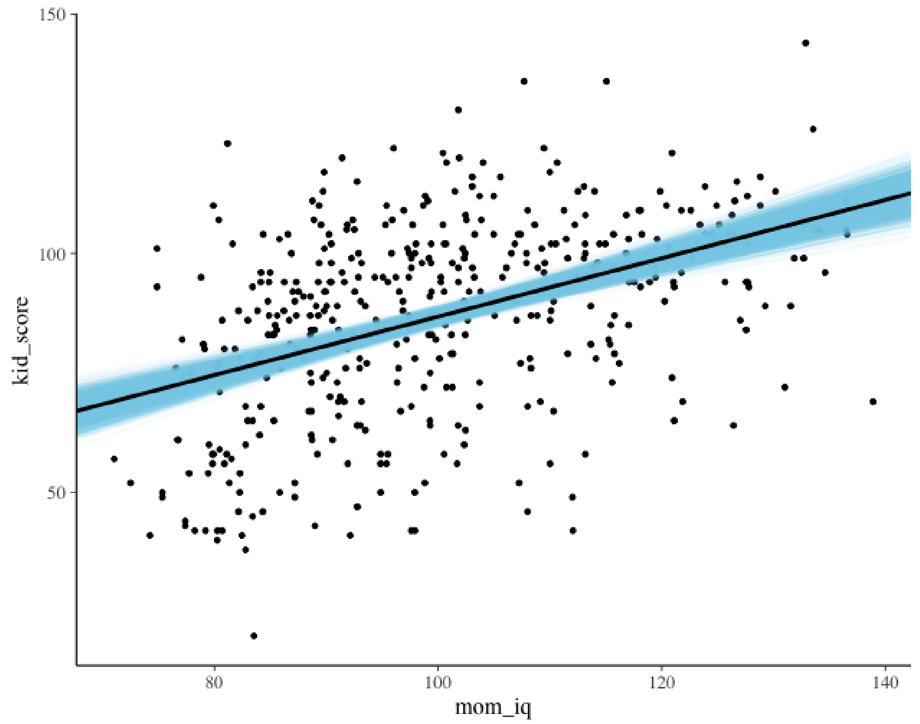
Plotting uncertainty

```
ggplot(kidiq, aes(x = mom_iq, y = kid_score)) +
  geom_point()
  geom_abline(data = draws, aes(intercept = `(Intercept)`, slope = mom_iq),
    size = 0.2, alpha = 0.1, color = "skyblue")
```



Plotting uncertainty

```
ggplot(kidiq, aes(x = mom_iq, y = kid_score)) +
  geom_point()
  geom_abline(data = draws, aes(intercept = `(Intercept)`, slope = mom_iq),
    size = 0.2, alpha = 0.1, color = "skyblue") +
  geom_abline(intercept = model_intercept, slope = model_slope)
```



Let's practice

BAYESIAN REGRESSION MODELING WITH RSTANARM

Making predictions

BAYESIAN REGRESSION MODELING WITH RSTANARM

Jake Thompson
Psychometrician, ATLAS, University of Kansas

Making predictions for observed data

```r
stan_model <- stan_glm(kid_score ~ mom_iq + mom_hs, data = kidiq)
posteriors <- posterior_predict(stan_model)
posteriors[1:10, 1:5]
```

```
              1         2         3         4         5
 [1,]  61.08989  58.57298  80.68946 101.00810  76.37946
 [2,] 111.52704  49.92284  99.09657  97.33291  72.98906
 [3,]  83.36793  81.35768  94.16414 101.73570  64.69375
 [4,] 118.15092  74.00476 107.28852  75.75912  91.93991
 [5,] 103.95042  58.98491 128.40312 121.42753  62.70008
 [6,] 102.29874 127.74050  84.10661  67.94056  82.02546
 [7,]  91.39445  88.49029  75.05702  94.48594 102.50331
 [8,]  93.33446  84.99589 101.49261  66.74698  68.26968
 [9,] 101.85065  91.46998 123.43011  76.53226  74.93288
[10,]  79.61489 101.29745 105.97636  97.48332  99.80582
```

Making predictions for new data

```
predict_data <- data.frame(
  mom_iq = 110,
  mom_hs = c(0, 1))

predict_data
```

```
  mom_iq mom_hs
1    110      0
2    110      1
```

Making predictions for new data

```
new_predictions <- posterior_predict(stan_model,
                                     newdata = predict_data)
new_predictions[1:10,]
```

```
              1         2
 [1,]  90.90581 107.75710
 [2,]  78.72466 139.86677
 [3,]  80.67743  88.81523
 [4,]  83.47852  74.06063
 [5,]  69.07708  87.81177
 [6,]  40.46229  85.45969
 [7,]  79.41597  64.19011
 [8,] 107.93867 117.49345
 [9,]  95.31493  82.51476
[10,]  91.18056  94.22732
```

```
summary(new_predictions[, 1])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
20.90    75.26   87.64   87.68  100.02  156.00
```

```
summary(new_predictions[, 2])
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
34.78    81.32   93.49   93.66  105.62  159.82
```

Let's practice

BAYESIAN REGRESSION MODELING WITH RSTANARM

Visualizing predictions

BAYESIAN REGRESSION MODELING WITH RSTANARM

Jake Thompson
Psychometrician, ATLAS, University of Kansas

Plotting new predictions

```
stan_model <- stan_glm(kid_score ~ mom_iq + mom_hs, data = kidiq)
predict_data <- data.frame(mom_iq = 110, mom_hs = c(0, 1))
posterior <- posterior_predict(stan_model, newdata = predict_data)
posterior[1:10,]
```

```
              1         2
 [1,]  76.75484  96.26407
 [2,]  74.39001 100.38898
 [3,]  90.90370  70.00591
 [4,]  70.43835 120.82787
 [5,] 113.98411  82.40497
 [6,]  56.15829 121.84269
 [7,]  90.46640  92.77966
 [8,]  98.56337 110.17948
 [9,] 108.86147 123.67762
[10,]  94.29429  83.77102
```

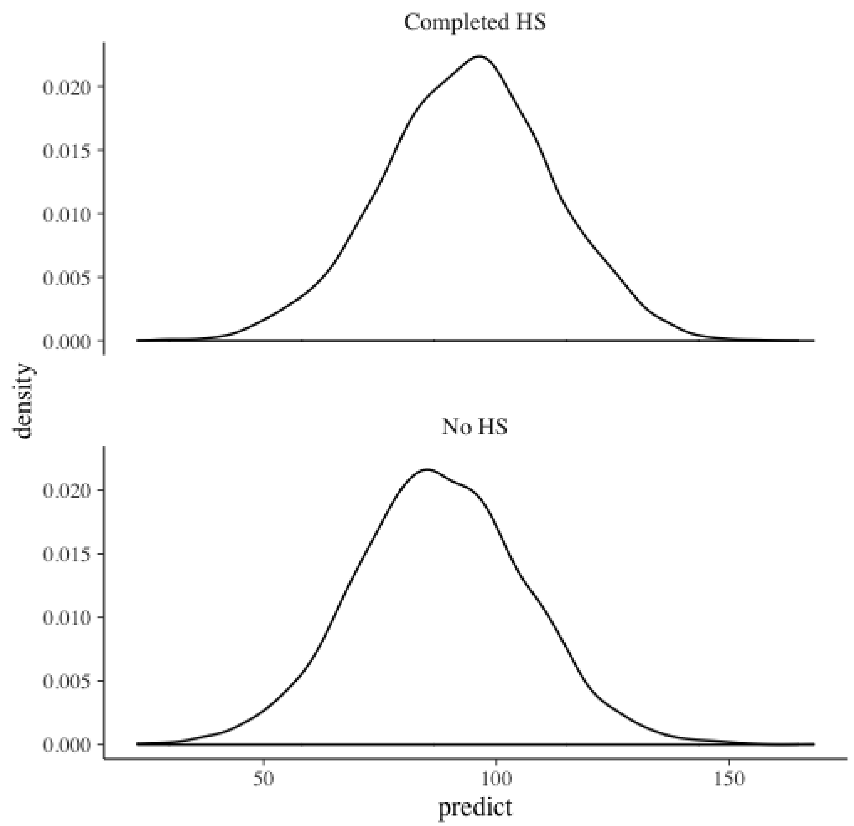
Formatting the data

```
posterior <- as.data.frame(posterior)
colnames(posterior) <- c("No HS", "Completed HS")
plot_posterior <- gather(posterior, key = "HS", value = "predict")
head(plot_posterior)
```

```
     HS   predict
1 No HS  76.75484
2 No HS  74.39001
3 No HS  90.90370
4 No HS  70.43835
5 No HS 113.98411
6 No HS  56.15829
```

Creating the plot

```
ggplot(plot_posterior, aes(x = predict)) +
  facet_wrap(~ HS, ncol = 1) +
  geom_density()
```



Let's practice

BAYESIAN REGRESSION MODELING WITH RSTANARM

Conclusion

BAYESIAN REGRESSION MODELING WITH RSTANARM

Jake Thompson
Psychometrician, ATLAS, University of Kansas

What we've learned

- How to estimate a Bayesian regression model
  - Differences betweens frequentist and Bayesian approaches
  - Importance of making correct inferences
- Modifying a Bayesian model
  - Size of the posterior distribution
  - Prior distributions
  - Estimation algorithm

What we've learned

- Evaluate model fit
  - R-squared
  - Posterior predictive model checks
  - Model comparisons
- Using the model
  - Model visualizations
  - Predictions

What we've missed

- Math behind posterior calculations and LOO approximation
- Choosing a prior distribution
- Causes of estimation errors

What comes next?

- More DataCamp courses
  - Bayesian Modeling with RJAGS
- **rstanarm** documentation
  - **mc-stan.org/rstanarm**
- *Bayesian Data Analysis*, Gelman et al., (2013)

Thank you!

BAYESIAN REGRESSION MODELING WITH RSTANARM