ITERATING OVER STATISTICAL MODELS

NCAA TOURNAMENT EDITION

WHY ARE YOU LISTENING TO ME?

- Stan developer
  http://mc-stan.org

- Researcher at Columbia

- Co-founder of Stan Group
  training / statistical support / consulting

- email: bearlee@alum.mit.edu
  twitter: @djsyclik / @mcmc_stan
  web: http://syclik.com

MARCH MADNESS

WHAT IS MARCH MADNESS?

WHAT IS MARCH MADNESS?

CHAMPIONSHIP GAME:
VILLANOVA VS NORTH CAROLINA

CHAMPIONSHIP GAME:
VILLANOVA VS NORTH CAROLINA

P(VILLANOVA WIN)?

VILLANOVA WILDCATS VS. NORTH CAROLINA TAR HEELS

- 1:00. 70 - 69
- 0:35. 72 - 69
- 0:23. 72 - 71

VILLANOVA WILDCATS VS. NORTH CAROLINA TAR HEELS

- 1:00. 70 - 69
- 0:35. 72 - 69
- 0:23. 72 - 71
- 0:13. 74 - 71

VILLANOVA WILDCATS VS. NORTH CAROLINA TAR HEELS

- 1:00. 70 - 69
- 0:35. 72 - 69
- 0:23. 72 - 71
- 0:13. 74 - 71
- 0:06. **74 - 74**

VILLANOVA WILDCATS VS. NORTH CAROLINA TAR HEELS

- 1:00. 70 - 69
- 0:35. 72 - 69
- 0:23. 72 - 71
- 0:13. 74 - 71
- 0:06. 74 - 74
- 0:00. **77 - 74. Villanova wins.**

SBNATION

FL NBA MLB NHL CFB RECRUITING CBB UFC SOCCER SHOP MORE ALL 311 BLOGS

COLLEGE BASKETBALL | MARCH MADNESS 2016: NCAA TOURNAMENT COVERAGE

The national championship game had the greatest ending in NCAA Tournament history

By Rodger Sherman @rodger_sherman on Apr 5, 2016, 12:09a 14

The national championship game ended on a buzzer-beater — but it had more than one shining moment.

Armour: Villanova win vs. UNC will be remembered as best title game ever

Nancy Armour
18 Hours Ago

USA TODAY

FTW! | NCAAB

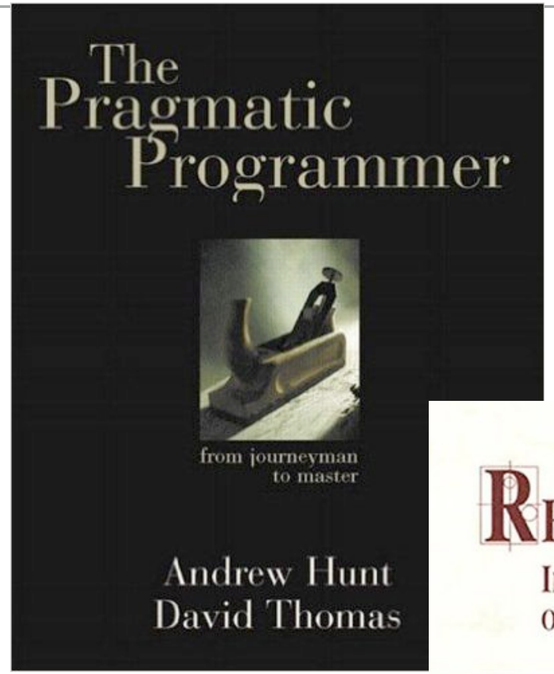
Villanova-UNC was the best NCAA championship game ever

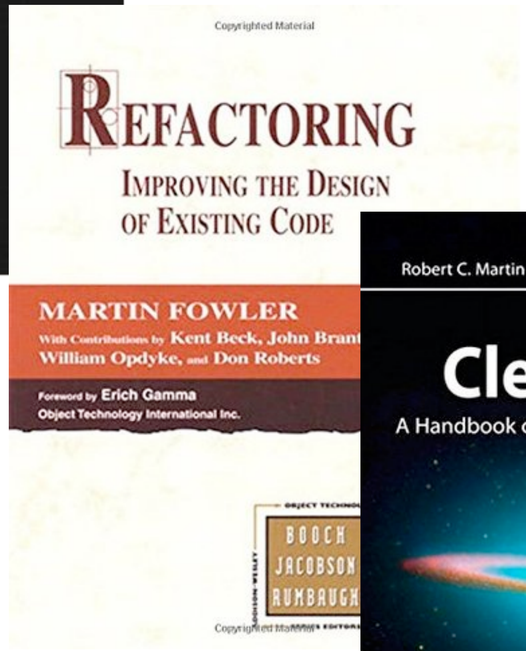
TREAT

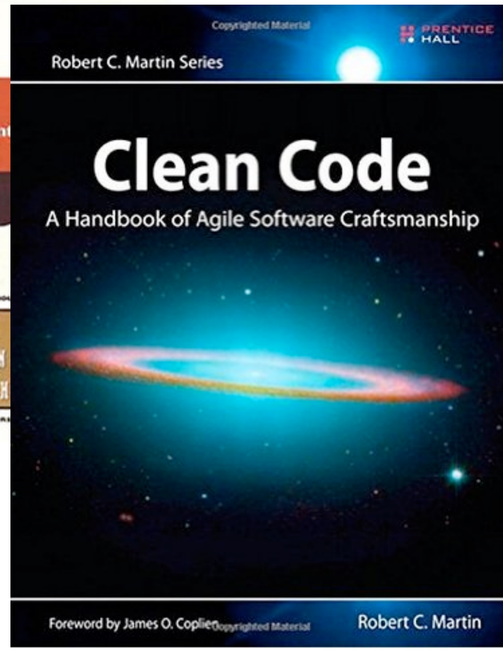
STATISTICAL MODELS

AS CODE

WRITING CODE IS A DISCIPLINE

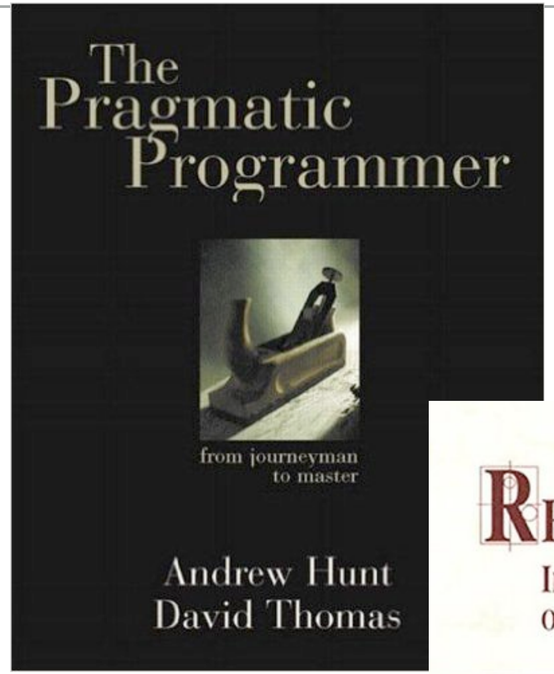


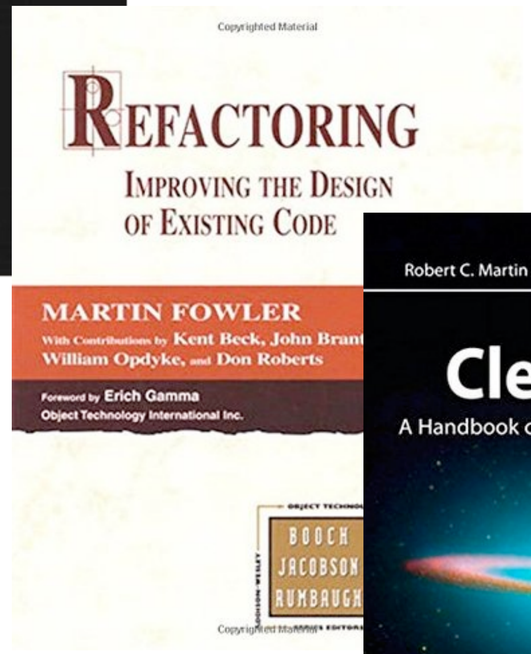


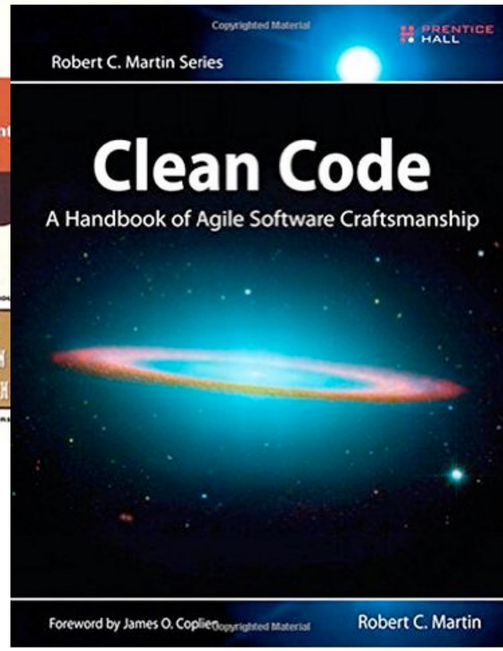


WRITING CODE ~~IS~~ CAN BE A DISCIPLINE

- design patterns
- testing
- code review
- maintenance
- modularity
- collaboration

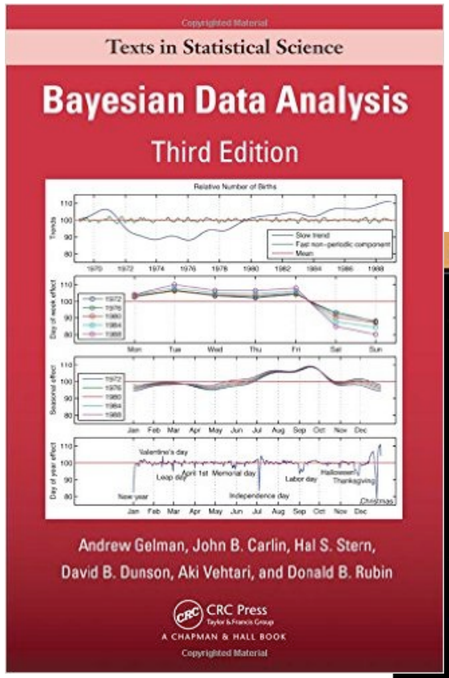


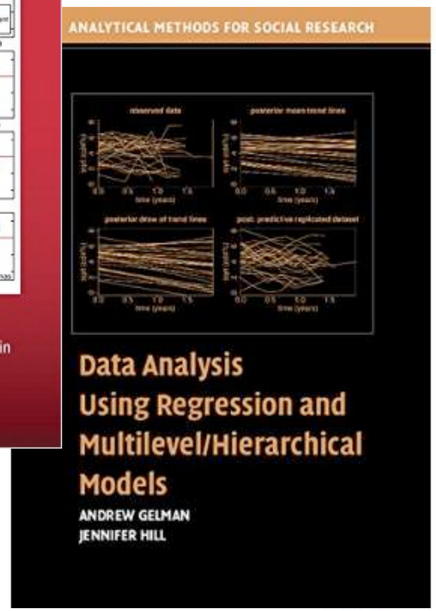


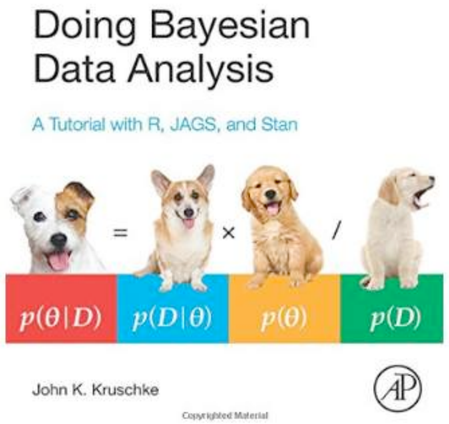


IS STATISTICAL MODELING A DISCIPLINE?

- Art or science?
- Models have names
- Statistical model vs implementation
- Collaboration on the statistical model?







WHAT DO WE NEED TO DO

- Elevate statistical models: **first class entity**
  - Modularize
  - Language
  - Discuss subtle details
- Collaborate

STAN GETS US CLOSE

- statistical modeling language
  - domain-specific language
    has its own grammar; not R or BUGS!
  - rstan
    shinystan, RStudio integration, rstanarm
  - open-source
    core libraries are new BSD

- Stan program
  - plain text
  - plays nicely with source repositories
  - imperative language

Note: Stan isn't the only thing you can do this with

MODELING BASKETBALL

BASKETBALL HISTORY

- 1891
- Dr. James Naismith. Springfield, MA
- Non-contact conditioning
- 13 simple rules
- Peach basket

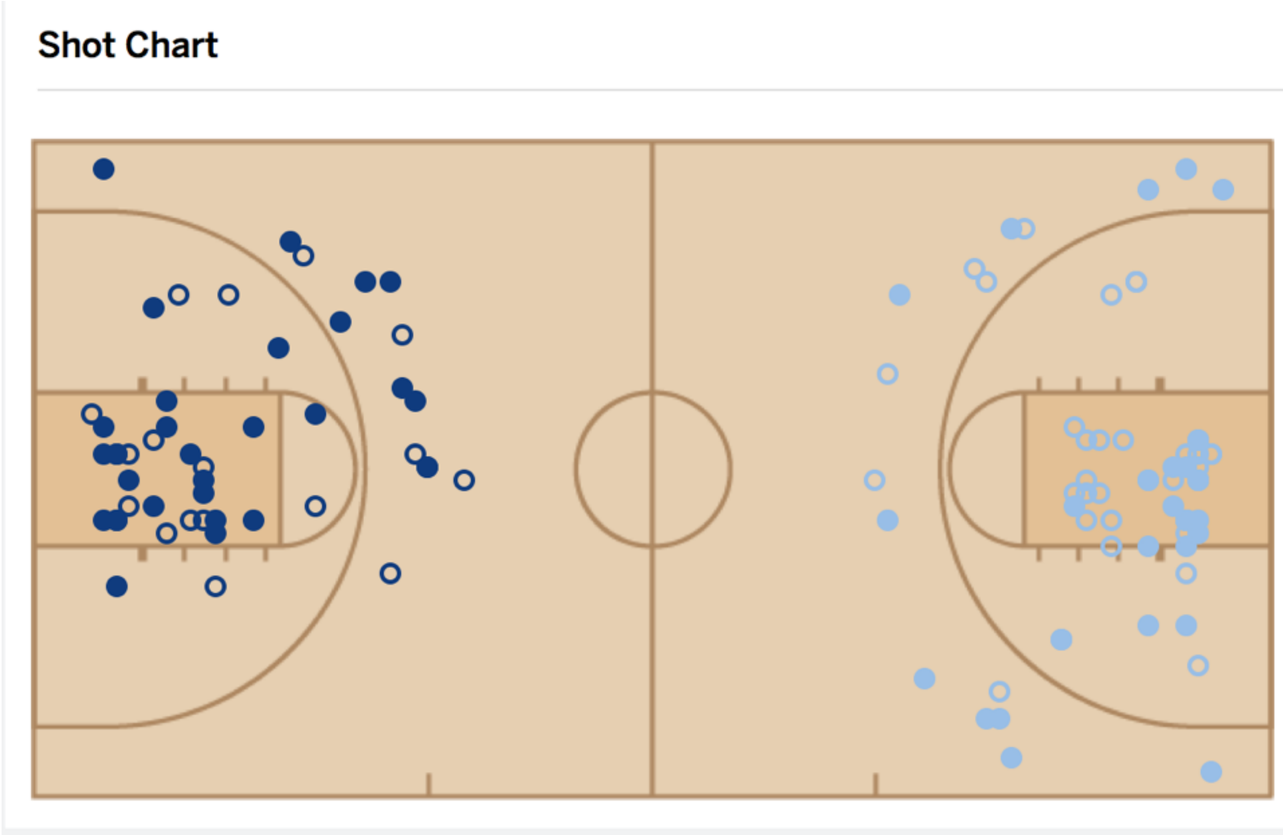
BASKETBALL HISTORY

- 1891
- Dr. James Naismith. Springfield, MA
- Non-contact conditioning
- 13 simple rules
- Peach basket

10. **The umpire shall be judge of the men** and shall note the fouls and notify the referee when three consecutive fouls have been made. He shall have power to disqualify men according to Rule 5.

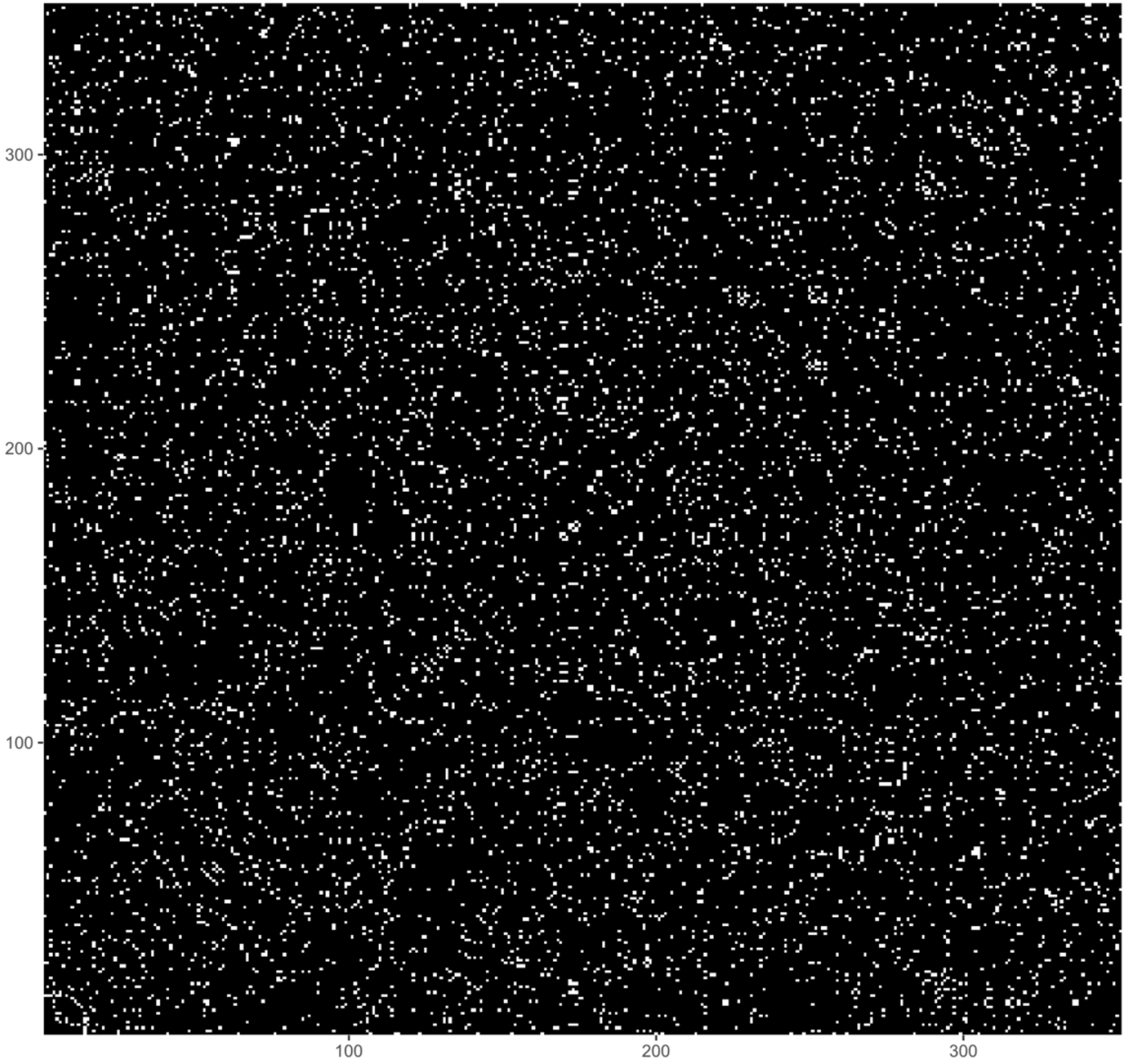
BASKETBALL NOW

- 2 x 20 min half
- Increasing score
- Points increment by 2, 3, and 1
- 5 players, unlimited substitutions
- player DQ: 5th foul
- Bonus: 7 team fouls
  Double bonus: 10 team fouls



DATA

- 351 NCAA Division 1 Men’s basketball teams
- 33 conferences
- 5421 games
- 24 - 35 games per team
- Max 3 observations


300
200
100
100
200
300

IS THIS BIG DATA?

TALL DATA VS WIDE DATA

- Tall data
  lots of replications

- Wide data
  lots of fields

  day, home, score, ot, fgm, fga, 3pm, 3pa, 3a, fta, ftm, or, dr, ast, to, stl, blk, pf

THREE STEPS OF BAYESIAN DATA ANALYSIS

THE THREE STEPS OF BAYESIAN DATA ANALYSIS

1. Set up full probability model

2. Condition on observed data

3. Evaluate the fit of the model

THE THREE STEPS OF BAYESIAN DATA ANALYSIS

1. Set up full probability model
   Write a Stan program

2. Condition on observed data

3. Evaluate the fit of the model

THE THREE STEPS OF BAYESIAN DATA ANALYSIS

1. Set up full probability model
   Write a Stan program

2. Condition on observed data
   Run RStan

3. Evaluate the fit of the model

THE THREE STEPS OF BAYESIAN DATA ANALYSIS

1. Set up full probability model
   Write a Stan program

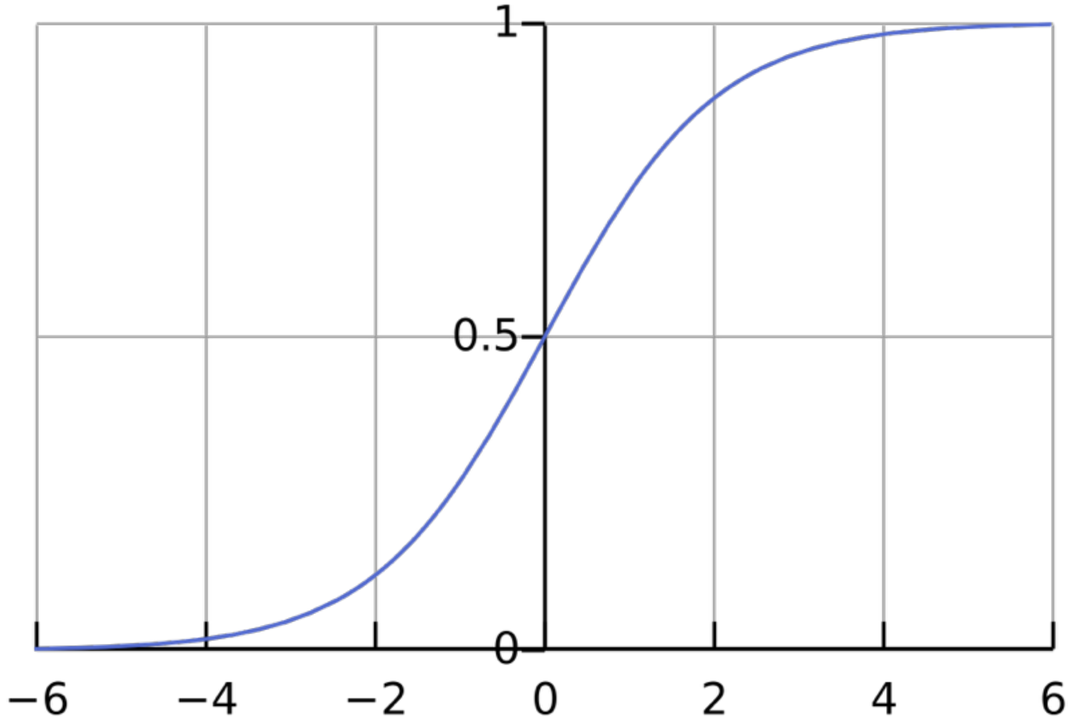
2. Condition on observed data
   Run RStan

3. Evaluate the fit of the model
   R, ShinyStan, posterior predictive checks

ITERATING OVER MODELS

STATISTICAL MODEL #1

- Only 2015-2016 matters
- Teams have a latent ability
- “logistic regression”
- “Bradley-Terry model”



$$y \sim \text{bernoulli}(\text{logit}^{-1}(\theta_1 - \theta_2))$$

```
data {
  int N_games;
  int W_id[N_games];
  int L_id[N_games];
}
transformed data {
  int<lower=1> win[N_games];
  win <- rep_array(1, N_games);
}
parameters {
  vector[351] team_ability;
}
model {
  team_ability ~ normal(0, 1);
  win ~ bernoulli_logit(team_ability[W_id] - team_ability[L_id]);
}
```

```
source("dat_2015.dat.R")
fit <- stan("ncaa.stan")
print(fit, pars=c("team_ability[326]", "team_ability[205]"))



Inference for Stan model: model1.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                   mean se_mean   sd 2.5%  25%  50%  75% 97.5% n_eff Rhat
team_ability[326] 2.82    0.01 0.48 1.93 2.50 2.80 3.13  3.80  4000    1
team_ability[205] 1.82    0.01 0.39 1.05 1.55 1.82 2.07  2.59  4000    1

Samples were drawn using NUTS(diag_e) at Sat Apr  9 01:12:53 2016.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

$$P(\text{VILLANOVA} > \text{UNC} \mid \text{DATA}) = 0.73$$

Save & Close

SHINYSTAN DIAGNOSE ESTIMATE EXPLORE MORE

Parameters plot | Posterior summary statistics | Generate LaTeX table

Select or enter parameter names

team_ability[326] team_ability[205]

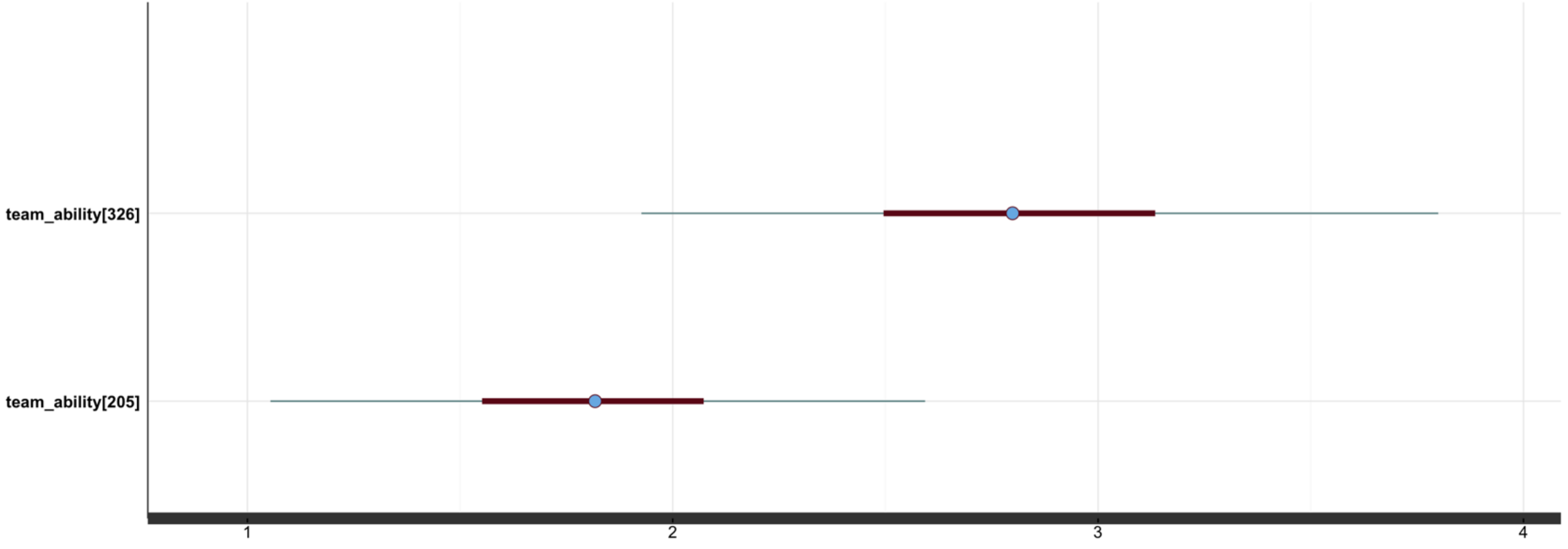
Search

Add parameters by regex search

Credible interval

50% 95%

Show/Hide Options



Save & Close

SHINYSTAN DIAGNOSE ESTIMATE EXPLORE MORE

Select parameter

team_ability[326]

Rhat	n_eff	mean	sd	2.5%	50%	97.5%
1	4000	2.82	0.48	1.93	2.8	3.8

Multiview

Bivariate

Trivariate

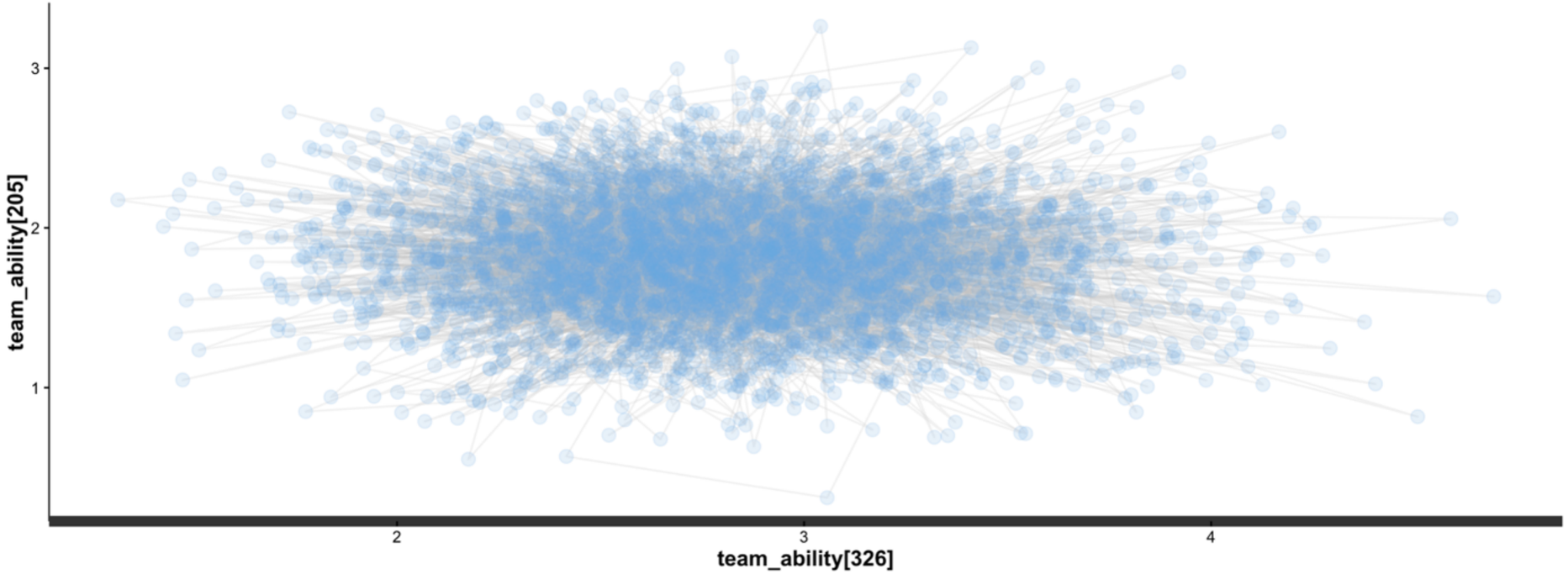
Density

Histogram

y-axis

team_ability[205]

Show/Hide Options



For Stan models using the NUTS algorithm, red points indicate iterations that encountered a divergent transition. Yellow points indicate a transition that hit the maximum treedepth rather than terminated its evolution normally.

ggplot2 pdf

TREAT STATISTICAL MODELS AS CODE

- Statistical model in a separate file
- Git
- Testing
  - Inspection of fit
  - Backtest on historical data
- Priors
- “Model #1”

IN WRITING, YOU MUST KILL ALL YOUR DARLINGS.

William Faulkner

STATISTICAL MODEL #2

- Home court advantage!

- Teams have a latent ability
- “logistic regression”
- “Bradley-Terry model”

$$y \sim \text{bernoulli}(\text{logit}^{-1}(\alpha + \theta_1 - \theta_2))$$

```
data {
  int N_games;
  int W_id[N_games];
  int L_id[N_games];
  int W_h_ind[N_games];
}
transformed data {
  int<lower=1> win[N_games];
  win <- rep_array(1, N_games);
}
parameters {
  vector[351] team_ability;
  vector[3] court_advantage;
}
model {
  team_ability ~ normal(0, 1);
  court_advantage ~ normal(0, 1);
  win ~ bernoulli_logit(court_advantage[W_h_ind]
                        + team_ability[W_id] - team_ability[L_id]);
}
```

```
Inference for Stan model: model2.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                    mean se_mean   sd  2.5%   25%  50%  75% 97.5% n_eff Rhat
team_ability[326] 0.13    0.02 0.97 -1.68 -0.52 0.10 0.76  2.05  4000    1
team_ability[205] 0.06    0.01 0.94 -1.77 -0.58 0.07 0.69  1.93  4000    1

Samples were drawn using NUTS(diag_e) at Sat Apr  9 01:33:33 2016.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

$$P(\text{VILLANOVA} > \text{UNC} \mid \text{DATA}) = 0.52$$

```
Inference for Stan model: model2.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                    mean se_mean   sd  2.5%   25%  50%  75% 97.5% n_eff Rhat
court_advantage[1] 6.46    0.01 0.40  5.74  6.18 6.44 6.72  7.30  4000    1
court_advantage[2] 7.01    0.01 0.38  6.31  6.74 7.00 7.27  7.81  4000    1
court_advantage[3] 5.48    0.01 0.44  4.69  5.18 5.45 5.77  6.41  4000    1
team_ability[326]  0.13    0.02 0.97 -1.68 -0.52 0.10 0.76  2.05  4000    1
team_ability[205]  0.06    0.01 0.94 -1.77 -0.58 0.07 0.69  1.93  4000    1

Samples were drawn using NUTS(diag_e) at Sat Apr  9 01:33:33 2016.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

STATISTICAL MODEL #3

- Assumptions:
  - Only 2015-2016 matters
  - Teams have a latent ability
  - Model points
  - Add a home court advantage

```
data {
  int N_games;
  vector<lower=0>[N_games] WL_spread;
  int W_id[N_games];
  int L_id[N_games];
  int W_h_ind[N_games];
}
parameters {
  vector[351] team_ability;
  real<lower=0> sigma;
  vector[3] court_advantage;
}
model {
  team_ability ~ normal(0, 1);
  court_advantage ~ normal(0, 1);
  sigma ~ normal(0, 1);
  WL_spread ~ normal(court_advantage[W_h_ind]
                     + team_ability[W_id] - team_ability[L_id],
                     sigma);
}
generated quantities {
  real championship_game;
  real villanova_wins;
  championship_game <- normal_rng(team_ability[326] - team_ability[205],
                                  sigma);
  villanova_wins <- if_else(championship_game > 0, 1, 0);
}
```

```
Inference for Stan model: model3.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

                    mean se_mean   sd   2.5%   25%   50%  75% 97.5% n_eff Rhat
sigma               8.33    0.00 0.08   8.16  8.27  8.33 8.38  8.49  4000    1
team_ability[326]   2.21    0.01 0.80   0.61  1.67  2.20 2.75  3.77  4000    1
team_ability[205]   2.22    0.01 0.79   0.70  1.68  2.21 2.76  3.78  4000    1
championship_game  -0.24    0.13 8.35 -17.13 -5.87 -0.15 5.37 16.42  4000    1
villanova_wins      0.49    0.01 0.50   0.00  0.00  0.00 1.00  1.00  3616    1

Samples were drawn using NUTS(diag_e) at Sat Apr  9 02:07:34 2016.
For each parameter, n_eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor on split chains (at
convergence, Rhat=1).
```

Save & Close

SHINYSTAN DIAGNOSE ESTIMATE EXPLORE MORE

Parameters plot | Posterior summary statistics | Generate LaTeX table

Select or enter parameter names

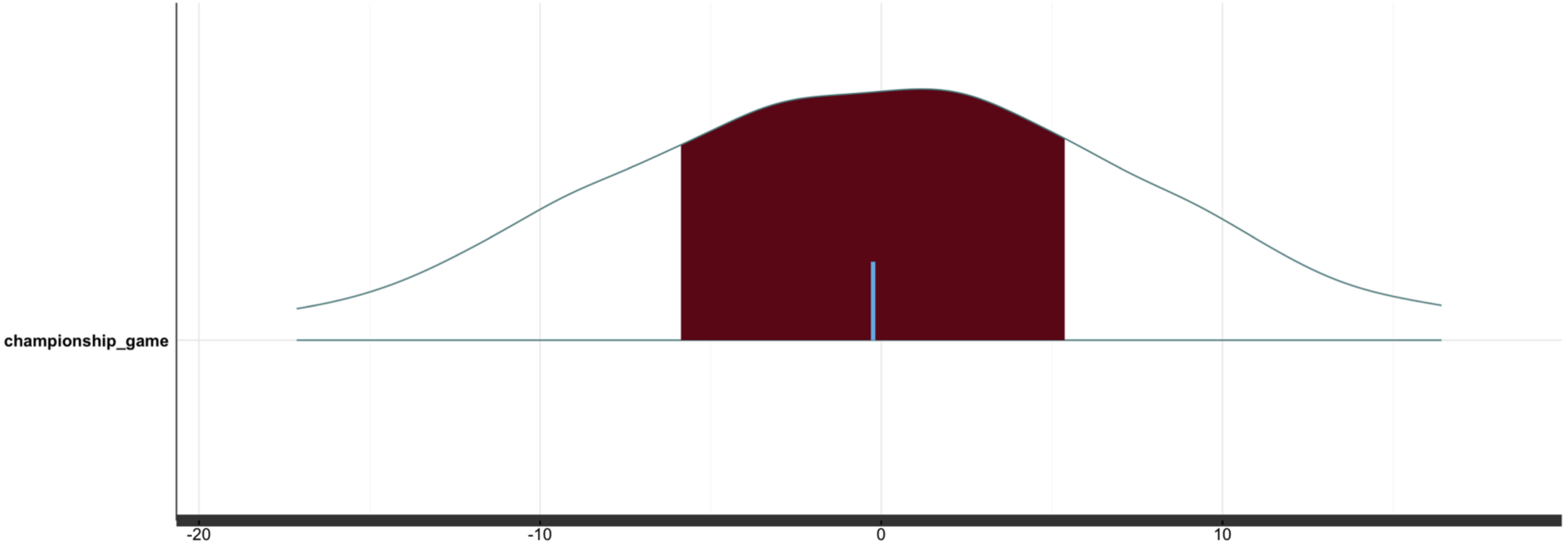
championship_game

Search

Credible interval

50% 95%

Show/Hide Options



HOW DID WE DO?

- Kaggle: 87 / 608

- What went wrong?
- What's next?

TREAT

STATISTICAL MODELS

AS CODE

THANKS

- Collaborative effort
  - NCAA modeling:
    Rob Trangucci
  - Stan team
    Andrew Gelman, Bob Carpenter, Matt Hoffman, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, Allen Riddell, Marco Ignacio, Jeff Arnold, Mitzi Morris, Rob Goedman, Brian Lau, Jonah Gabry, Alp Kucukelbir, Robert Grant, Dustin Tran, Krzysztof Sakrejda, Aki Vehtari, Rayleigh Lei, Sebastian Weber

HELP

- http://mc-stan.org
- stan-users mailing list
- Stan Group Inc.
  http://stan.fit
  training / statistical support / consulting
- bearlee@alum.mit.edu / @djsyclik / @mcmc_stan

A Fiscally Sponsored Project of
NUMFOCUS
OPEN CODE = BETTER SCIENCE