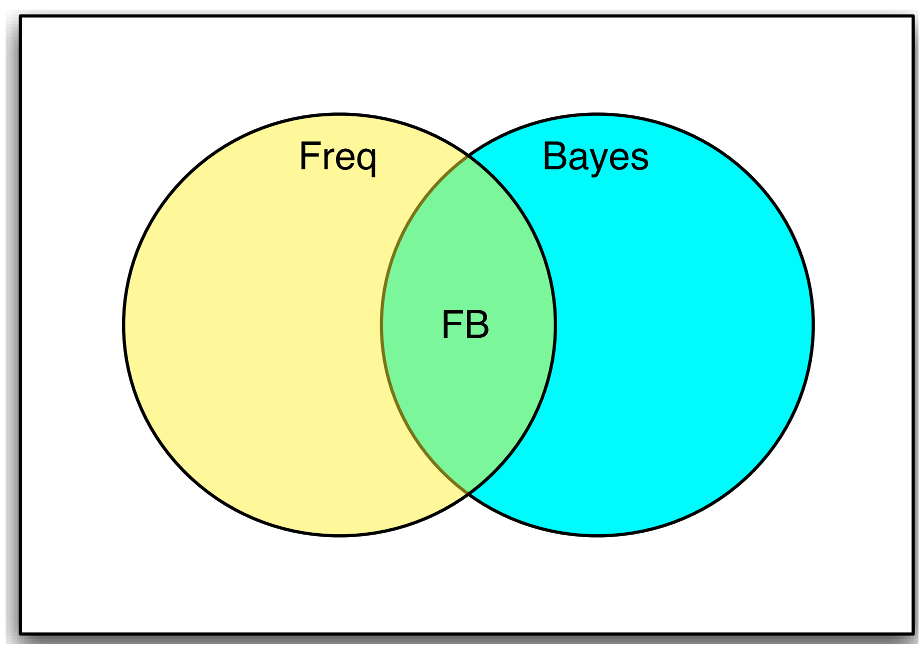
CS109/Stat121/AC209/E-109
Data Science
Bayesian Methods Continued

Hanspeter Pfister, Joe Blitzstein, Verena Kaynig



This Week

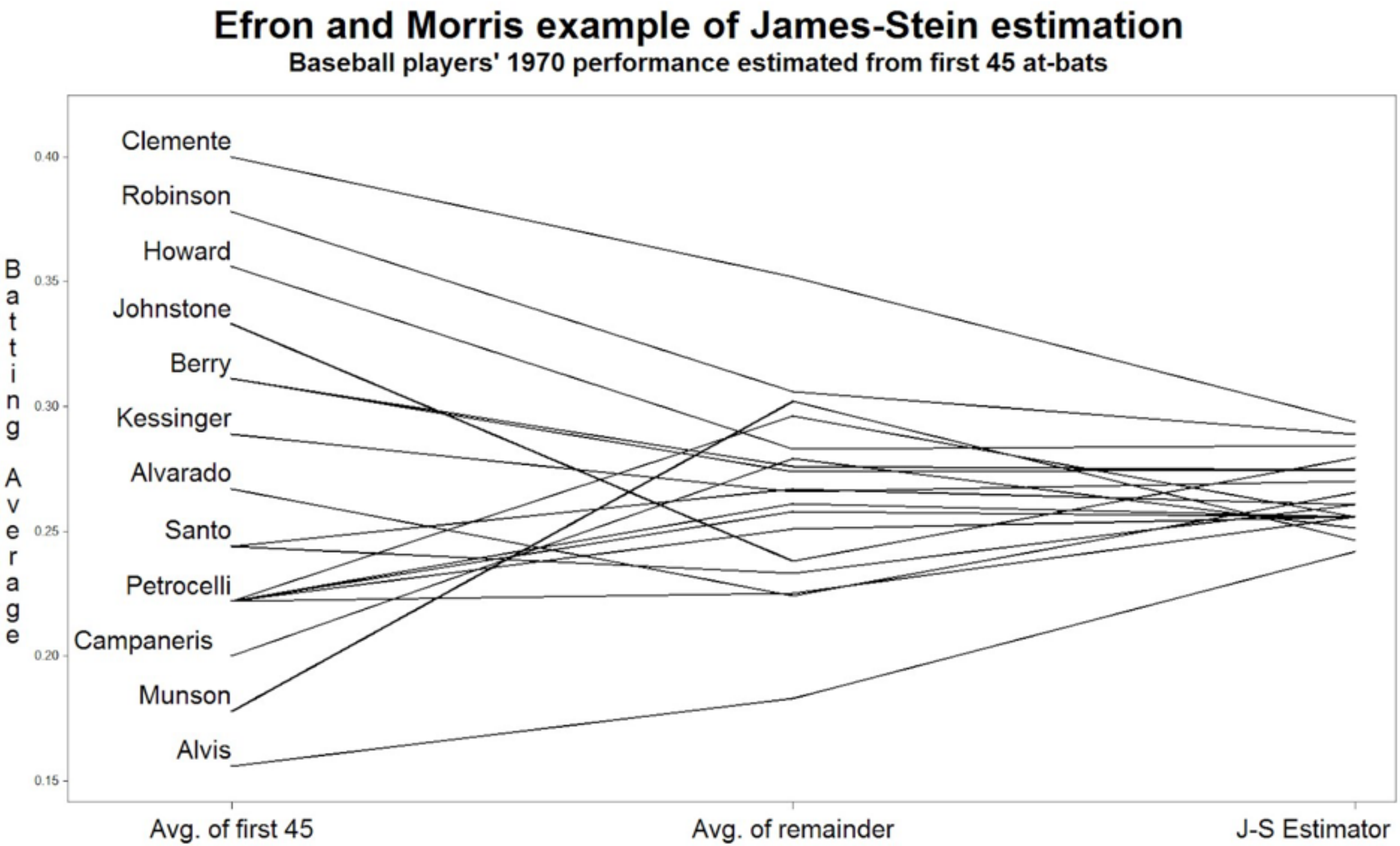
- Make sure HW3 Google presentation is *public* ASAP if you haven't already. See Rahul's Piazza post on this (@1159).
- Form project team if you haven't already. Try your best to have your team formed by next Tuesday, but in any case it is required to fill in the Google form by Tuesday Nov 3, 11:59 pm: http://goo.gl/forms/CzVRIuCZk6
- HW4 is due Thursday Nov 5, 11:59 pm.

Which proportion is bigger?
A: 1 out of 2
B: 40 out of 100

Which baseball player would you prefer to have on your team?
Player A: 1 hit out of 2 at-bats
Player B: 40 hits out of 100 at-bats

Which cab driver would you prefer to have drive you somewhere?
Driver A: 1 of 2 reaching destination
Driver B: 40 out of 100 reaching destination

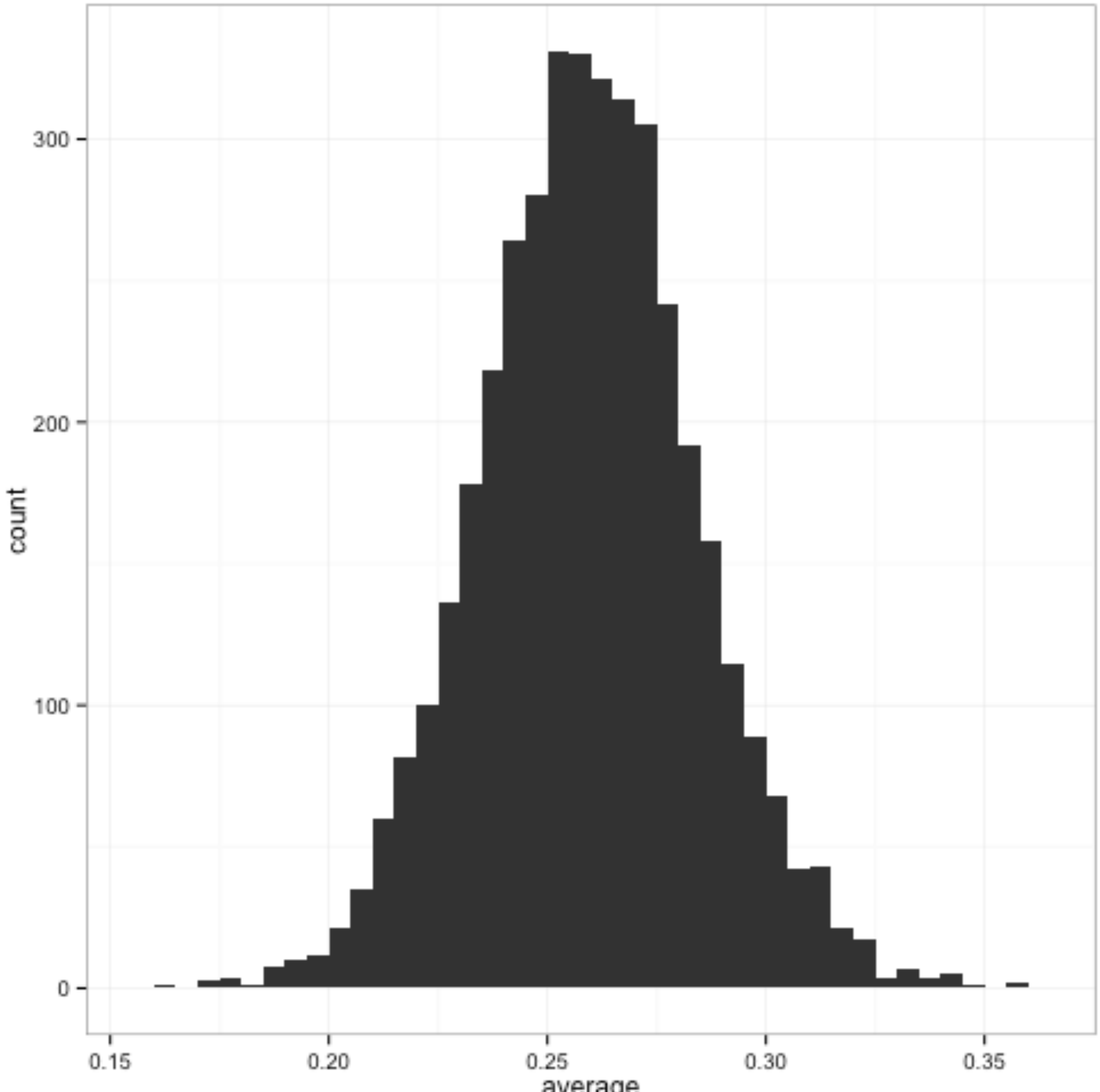
Bayes and Shrinkage Estimation



http://sas-and-r.blogspot.com/2012/04/example-927-baseball-and-shrinkage.html

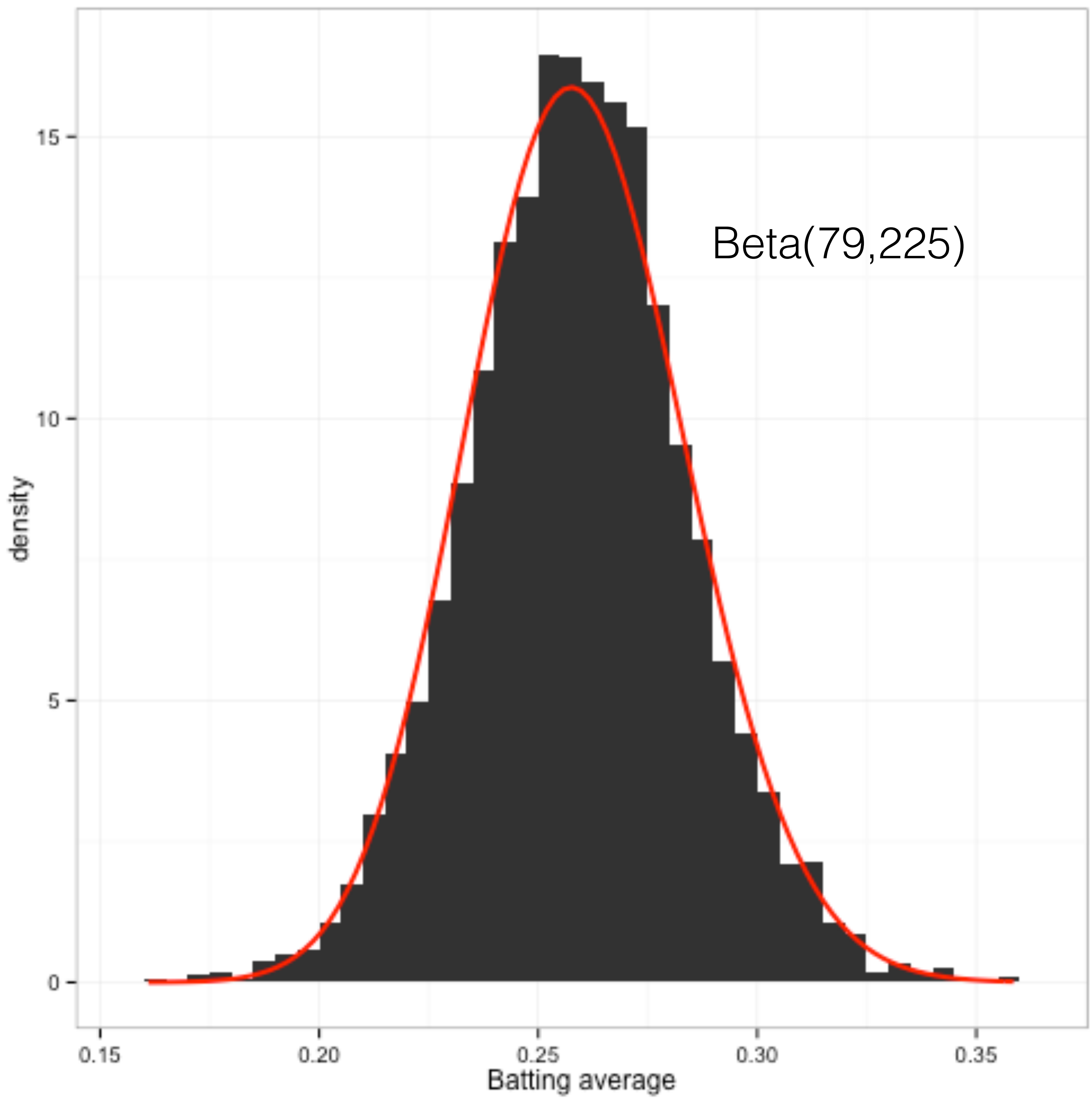
Empirical Bayes

Consider overall distribution across baseball players, estimate the prior from the data.



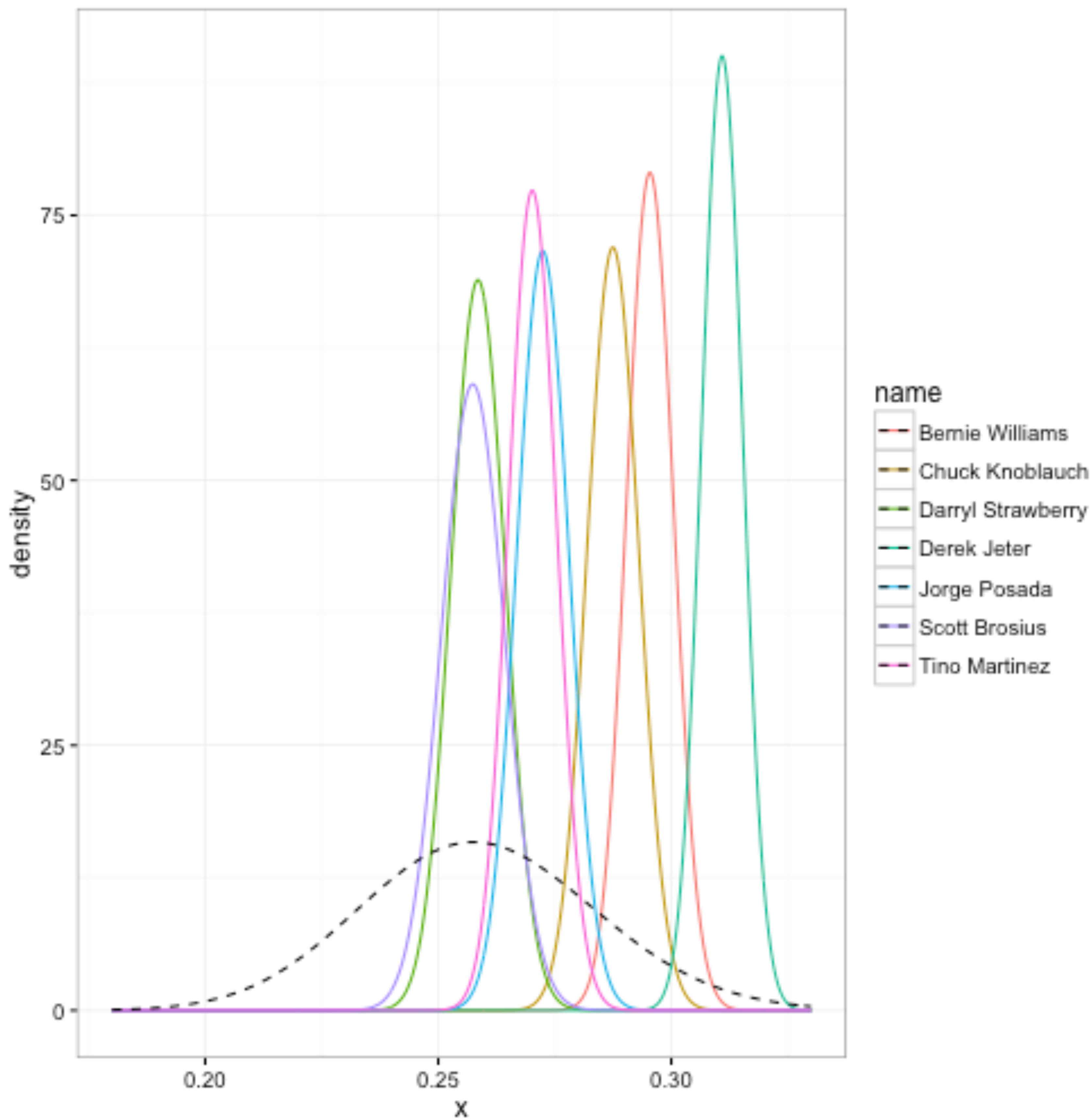
http://varianceexplained.org/r/empirical_bayes_baseball/

Empirical Bayes



http://varianceexplained.org/r/empirical_bayes_baseball/

Empirical Bayes



http://varianceexplained.org/r/credible_intervals_baseball/

Empirical Bayes

Best batters according to MLE

name	H	AB	average
Jeff Banister	1	1	1
Doc Bass	1	1	1
Steve Biras	2	2	1
C. B. Burns	1	1	1
Jackie Gallagher	1	1	1

http://varianceexplained.org/r/empirical_bayes_baseball/

Empirical Bayes

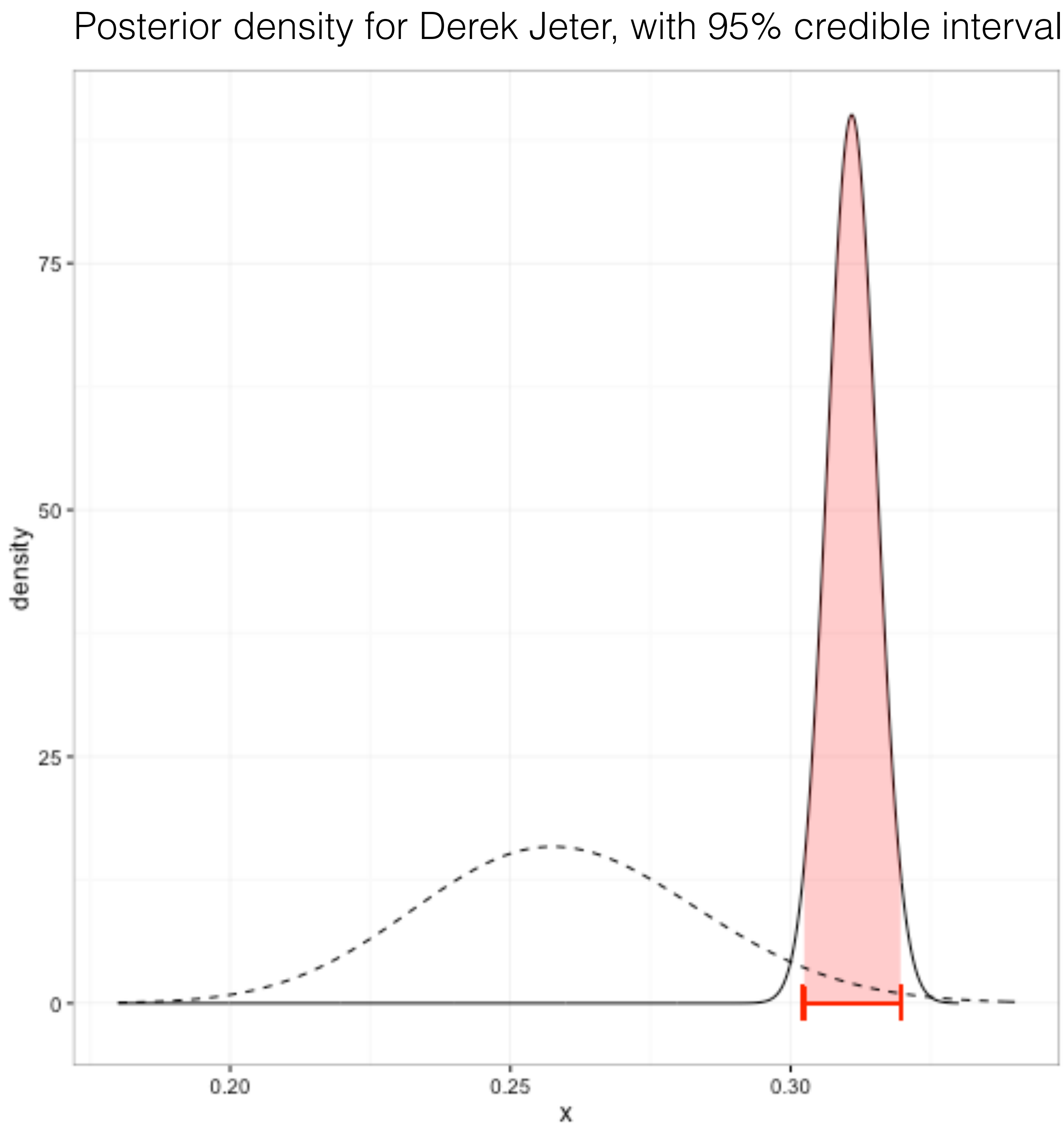
Best batters according to EB estimates.

name	H	AB	average	eb_estimate
Rogers Hornsby	2930	8173	0.358	0.355
Shoeless Joe Jackson	1772	4981	0.356	0.350
Ed Delahanty	2596	7505	0.346	0.343
Billy Hamilton	2158	6268	0.344	0.340
Harry Heilmann	2660	7787	0.342	0.339

```
// We hired a Data Scientist to analyze our Big Data
// and all we got was this lousy line of code.
float estimate = (successes + 78.7) / (total + 303.5);
```

http://varianceexplained.org/r/empirical_bayes_baseball/

Empirical Bayes



http://varianceexplained.org/r/credible_intervals_baseball/

Confidence Intervals vs. Credible Intervals

95% confidence interval: $P(a(Y) \leq \theta \leq b(Y)) = 0.95$

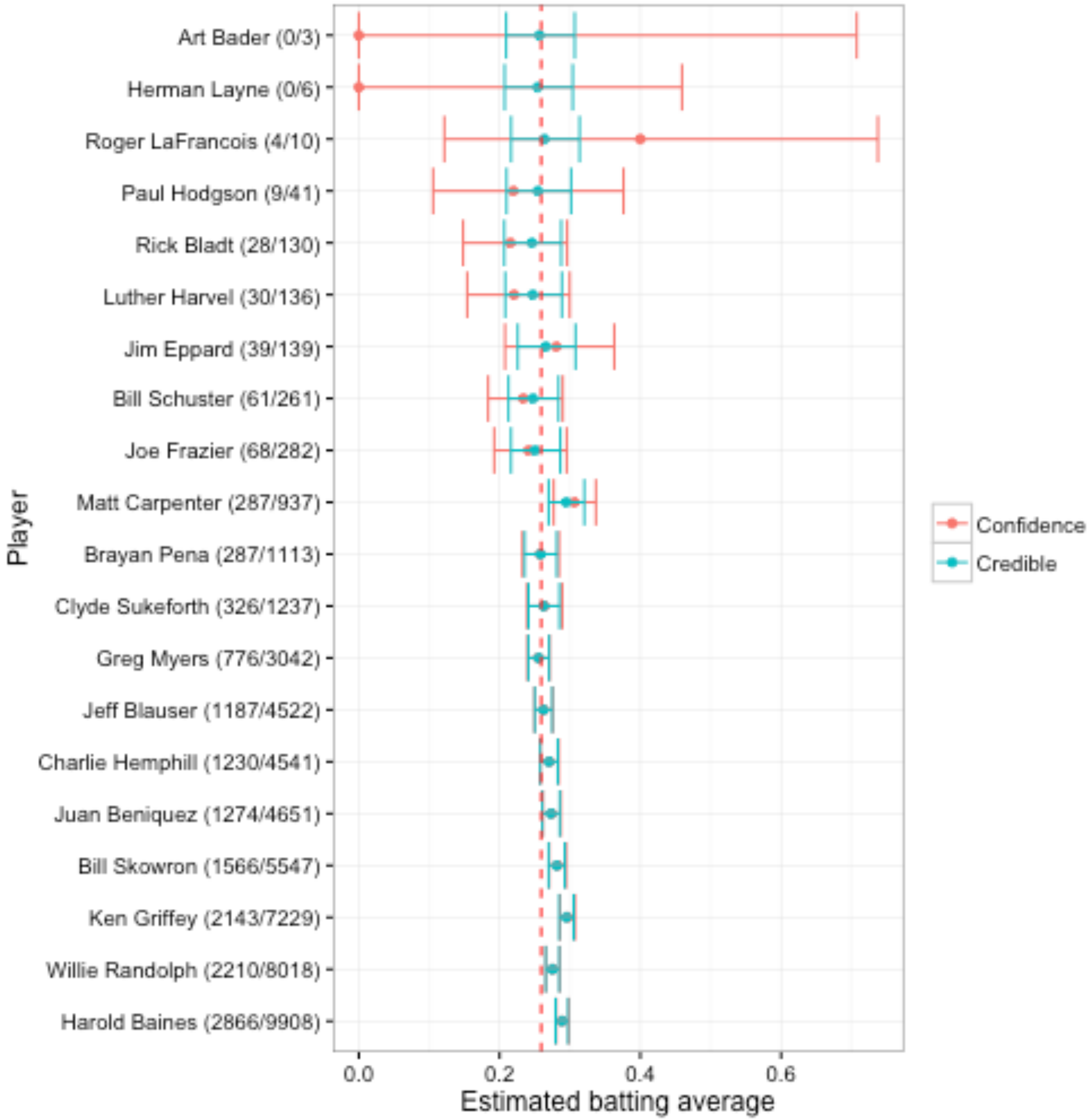
parameter is fixed, data is random

$P(3 \leq \theta \leq 7)$ is 0 or 1 (we just don't know which), from non-Bayesian perspective

95% credible interval: $P(a(Y) \leq \theta \leq b(Y) | Y) = 0.95$

parameter is random, data is fixed

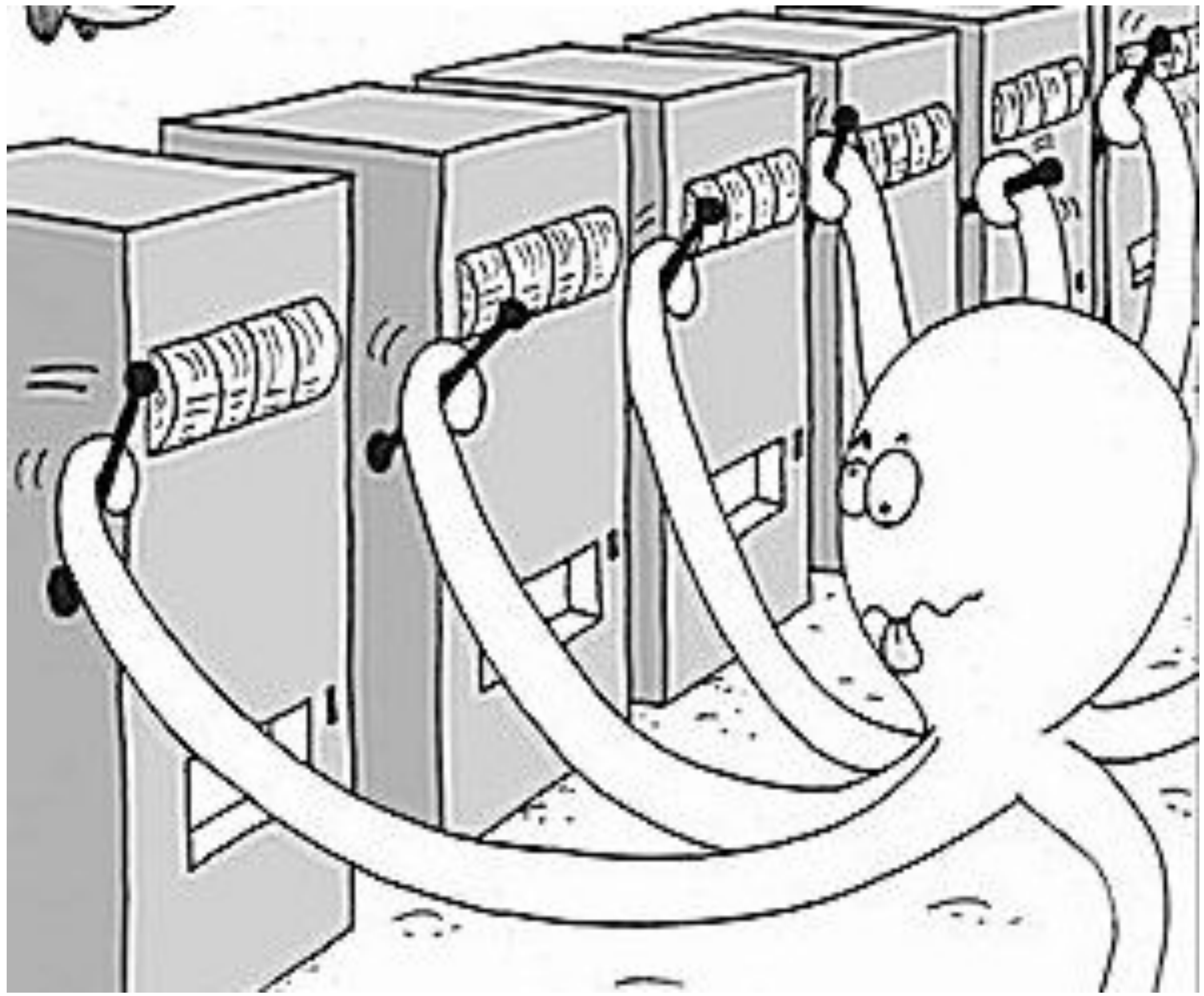
But we can often get the best of both worlds: we can study the coverage probabilities of credible intervals; we can study the frequentist performance of Bayesian methods.



http://varianceexplained.org/r/credible_intervals_baseball/

Bayesian Bandits

Example from Probabilistic Programming and Bayesian Methods for Hackers



http://research.microsoft.com/en-us/projects/bandits/

N slot machines, each with its own unknown probability distribution for rewards. Exploration-exploitation tradeoff.

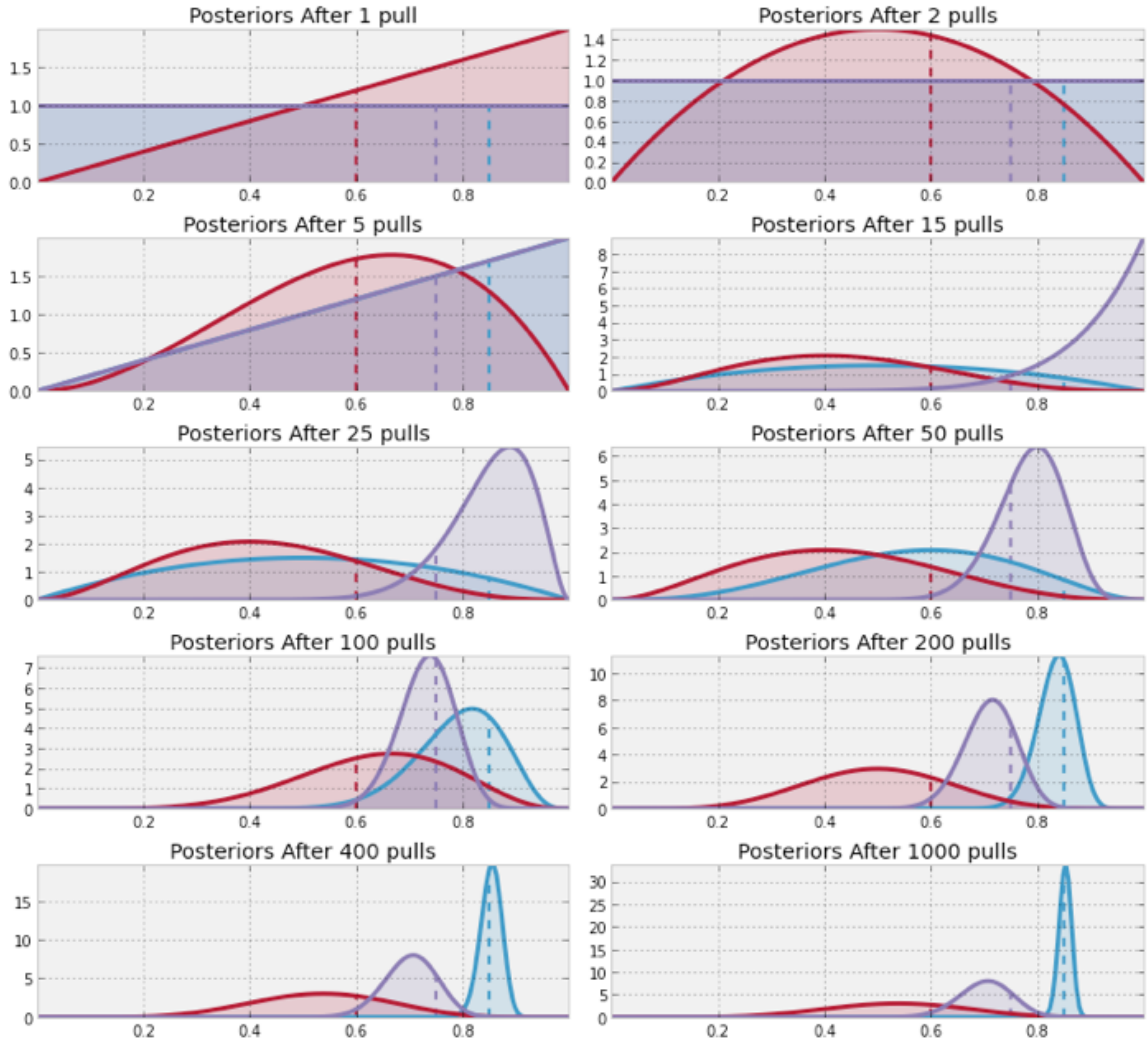
Bayesian Bandits

Example from Probabilistic Programming and Bayesian Methods for Hackers

A fast, simple Bayesian algorithm:

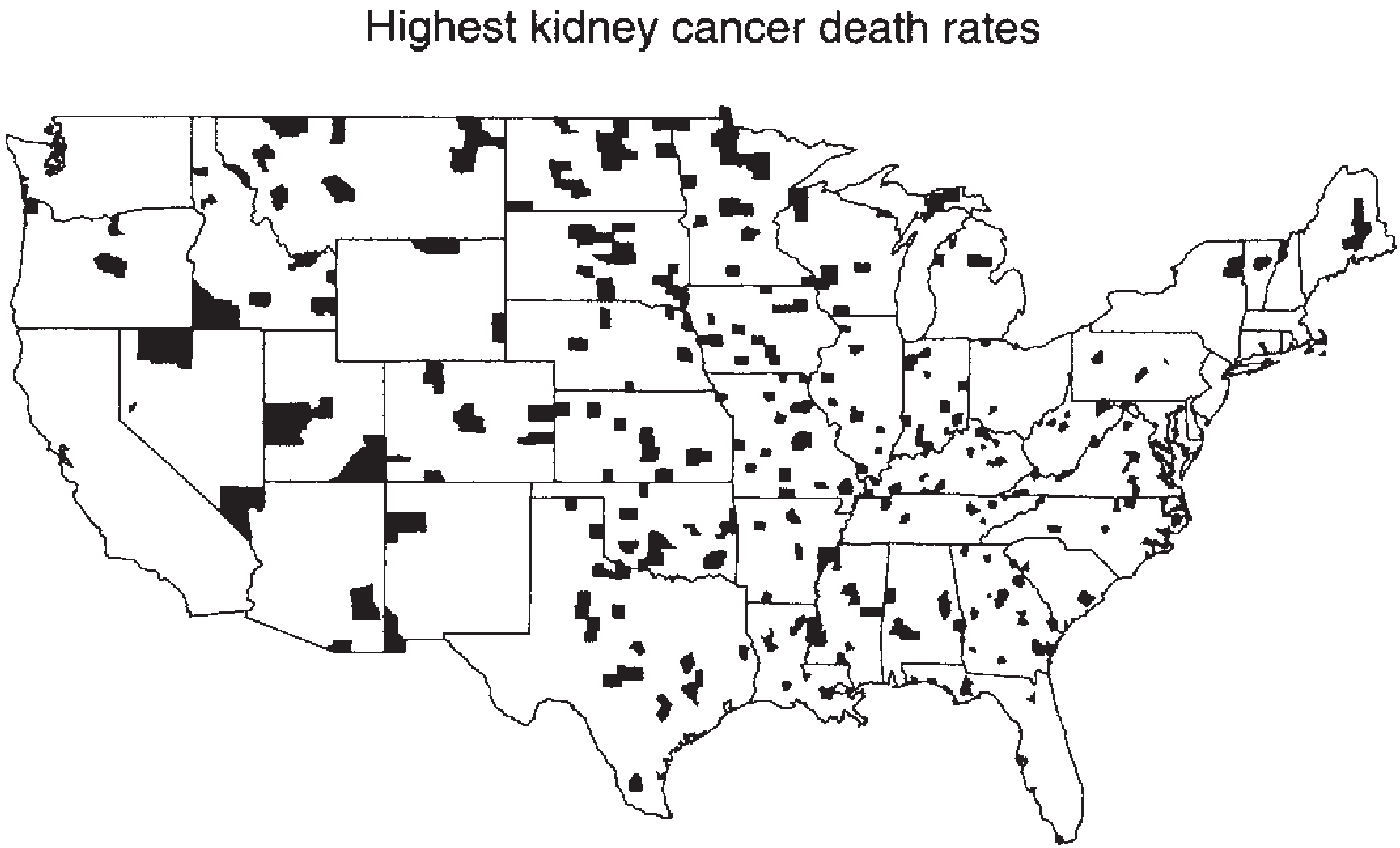
1. sample from the prior of each bandit
2. select the bandit with the largest sampled value
3. update the prior for that bandit (the posterior becomes the new prior)
4. repeat.

Bayesian Bandits



http://nbviewer.ipython.org/urls/raw.github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/master/Chapter6_Priorities/Priors.ipynb

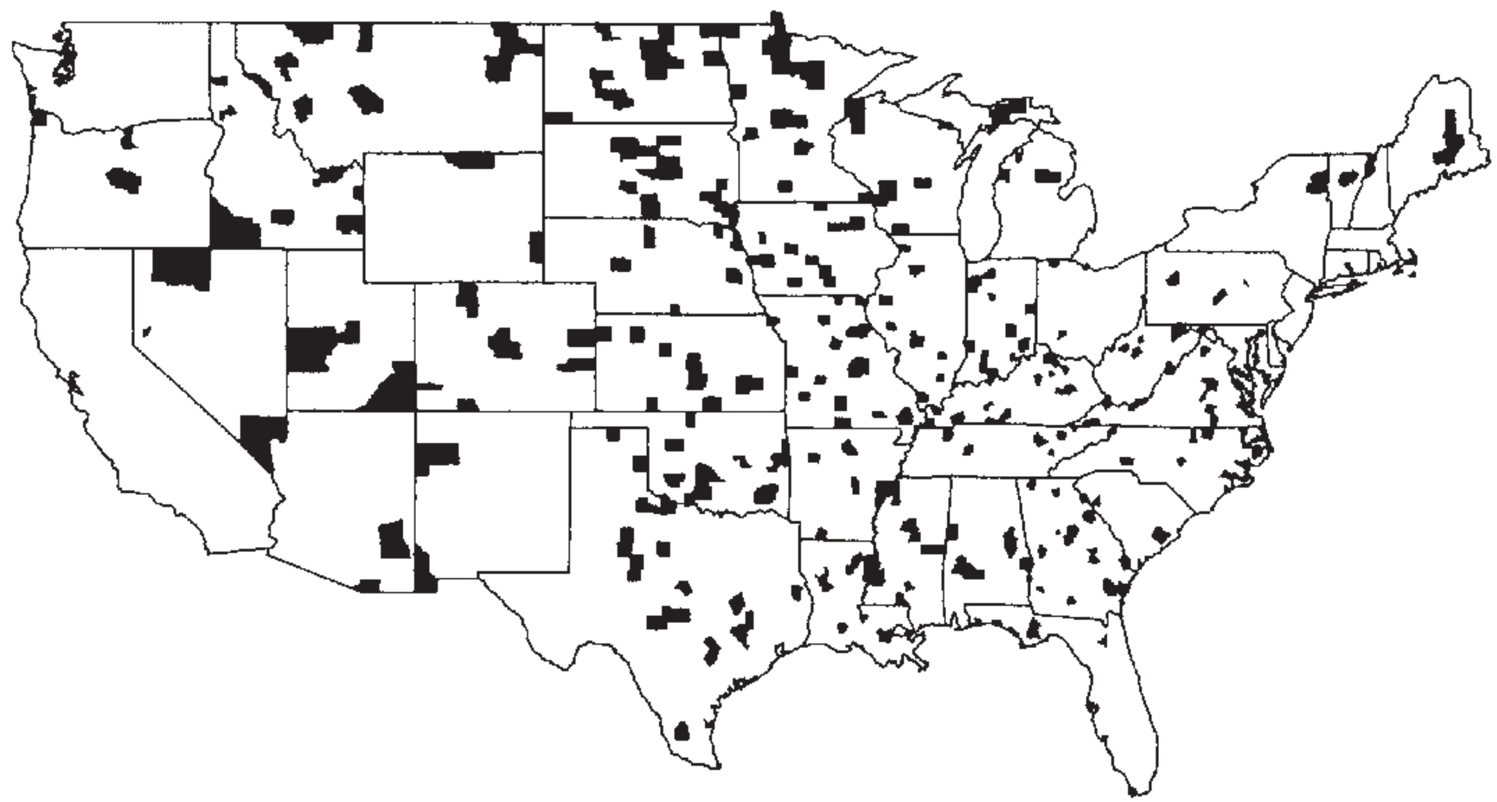
Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)



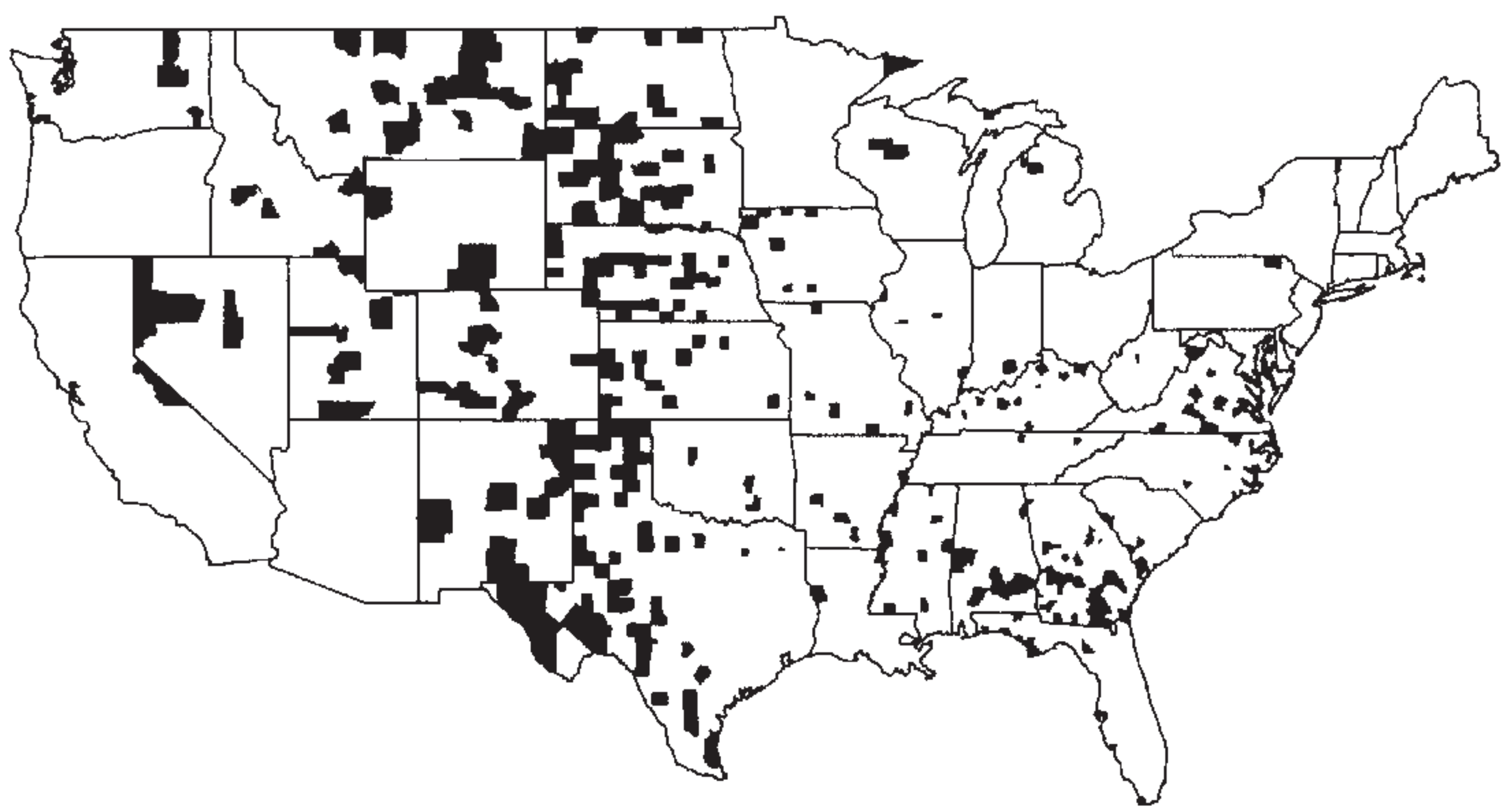
U.S. counties with the highest 10% of kidney cancer death rates (age-adjusted)

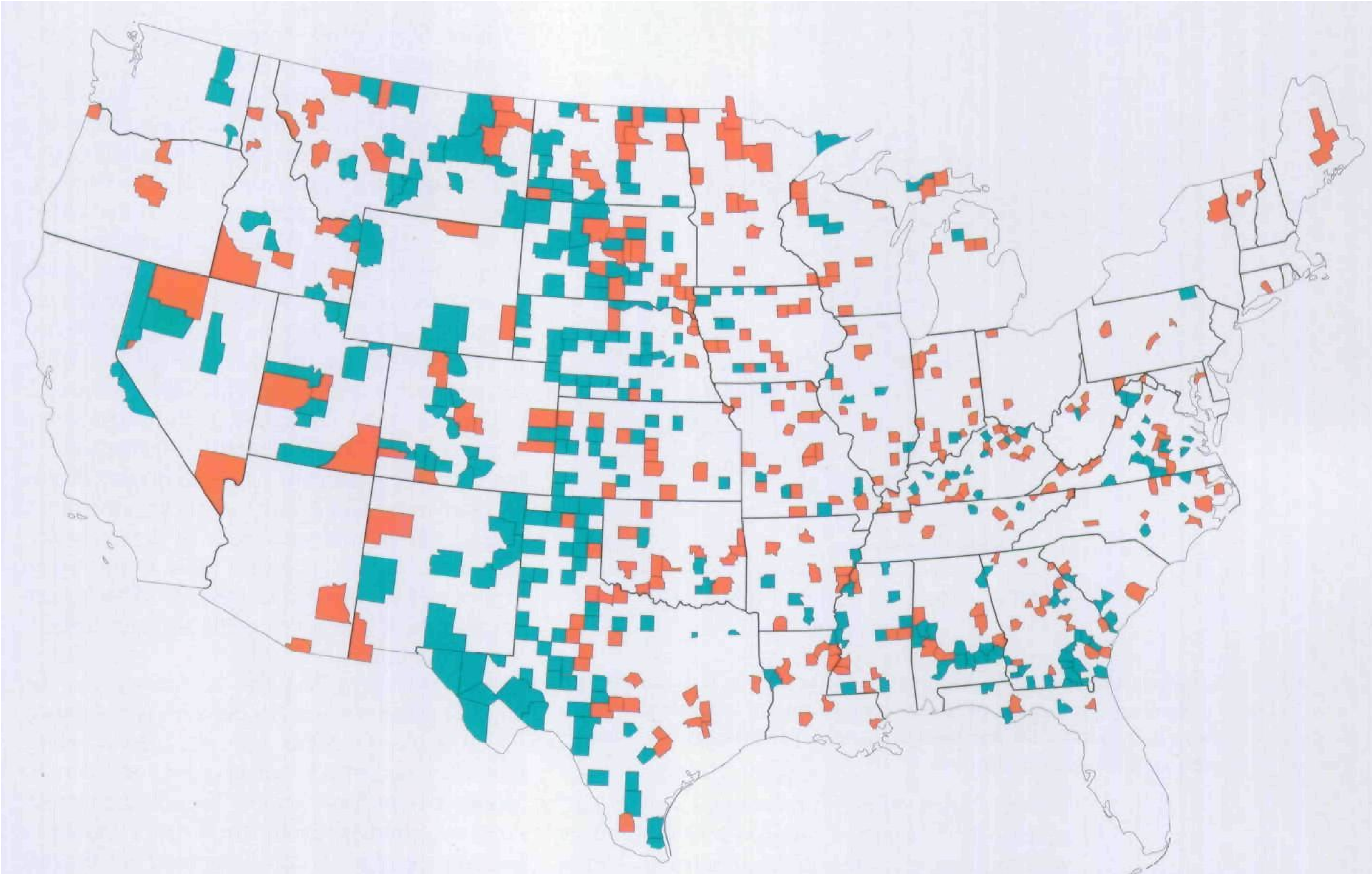
Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)

Highest kidney cancer death rates



Lowest kidney cancer death rates

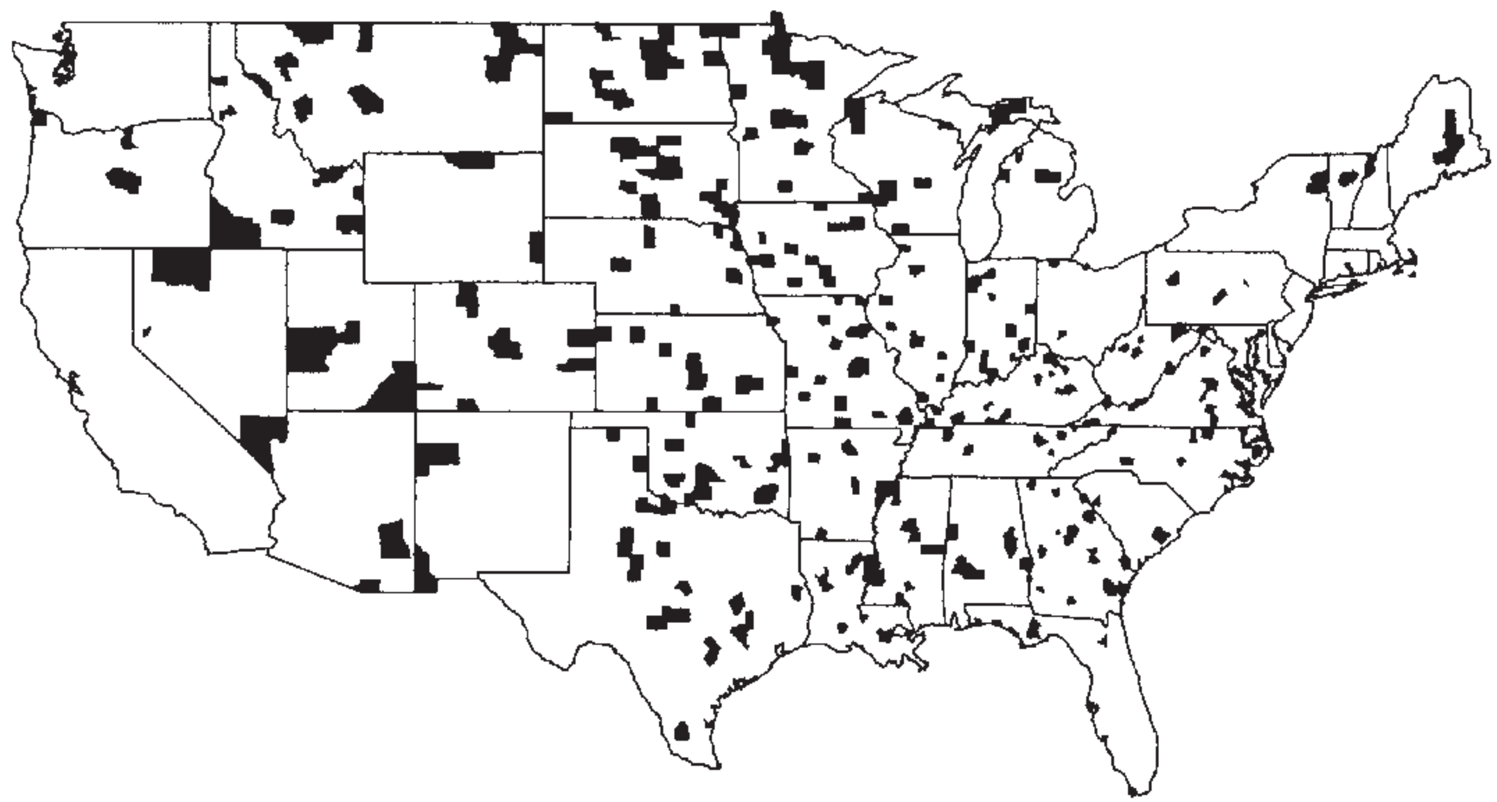




lowest rates: blue
highest rates: orange

H. Wainer, The Most Dangerous Equation

Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)



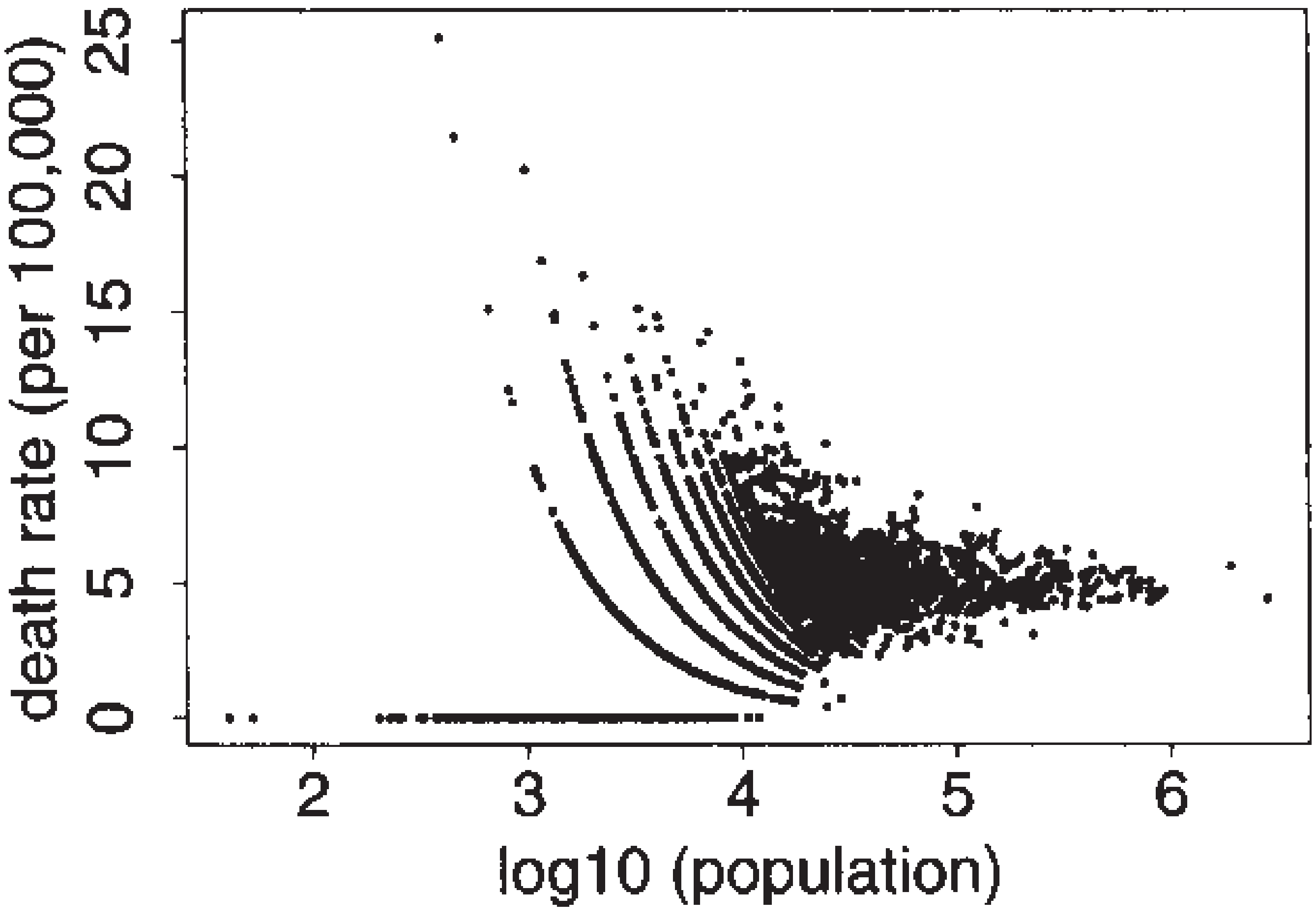
simple model: $$y_j \sim \mathrm{Pois}(10 n_j \theta_j)$$

$$\theta_j \sim \mathrm{Gamma}(\alpha, \lambda)$$

$$E(\theta_j | y_j) = w \frac{y_j}{10 n_j} + (1 - w) E(\theta_j)$$

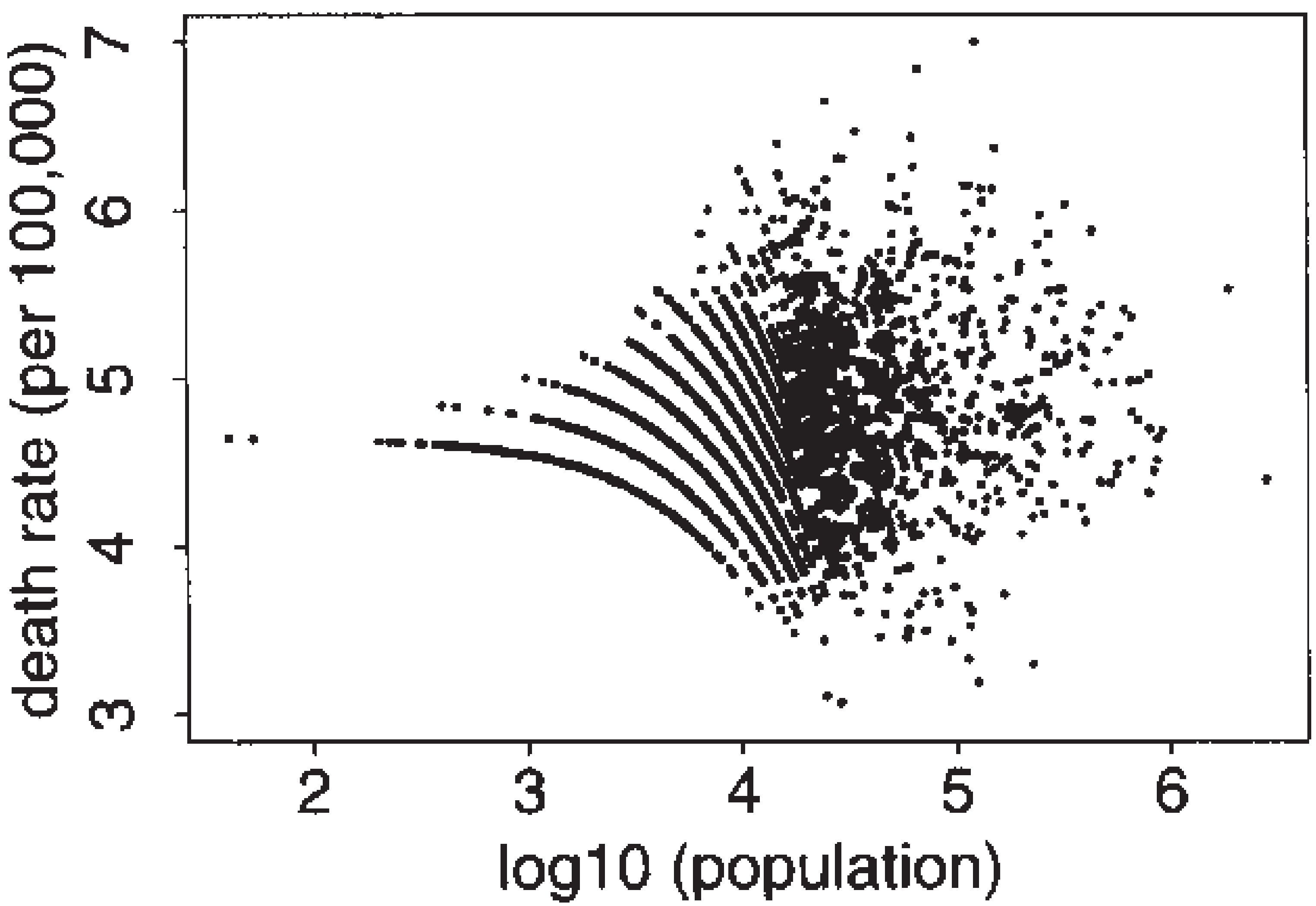
weighted combination of the data and the prior mean

Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)



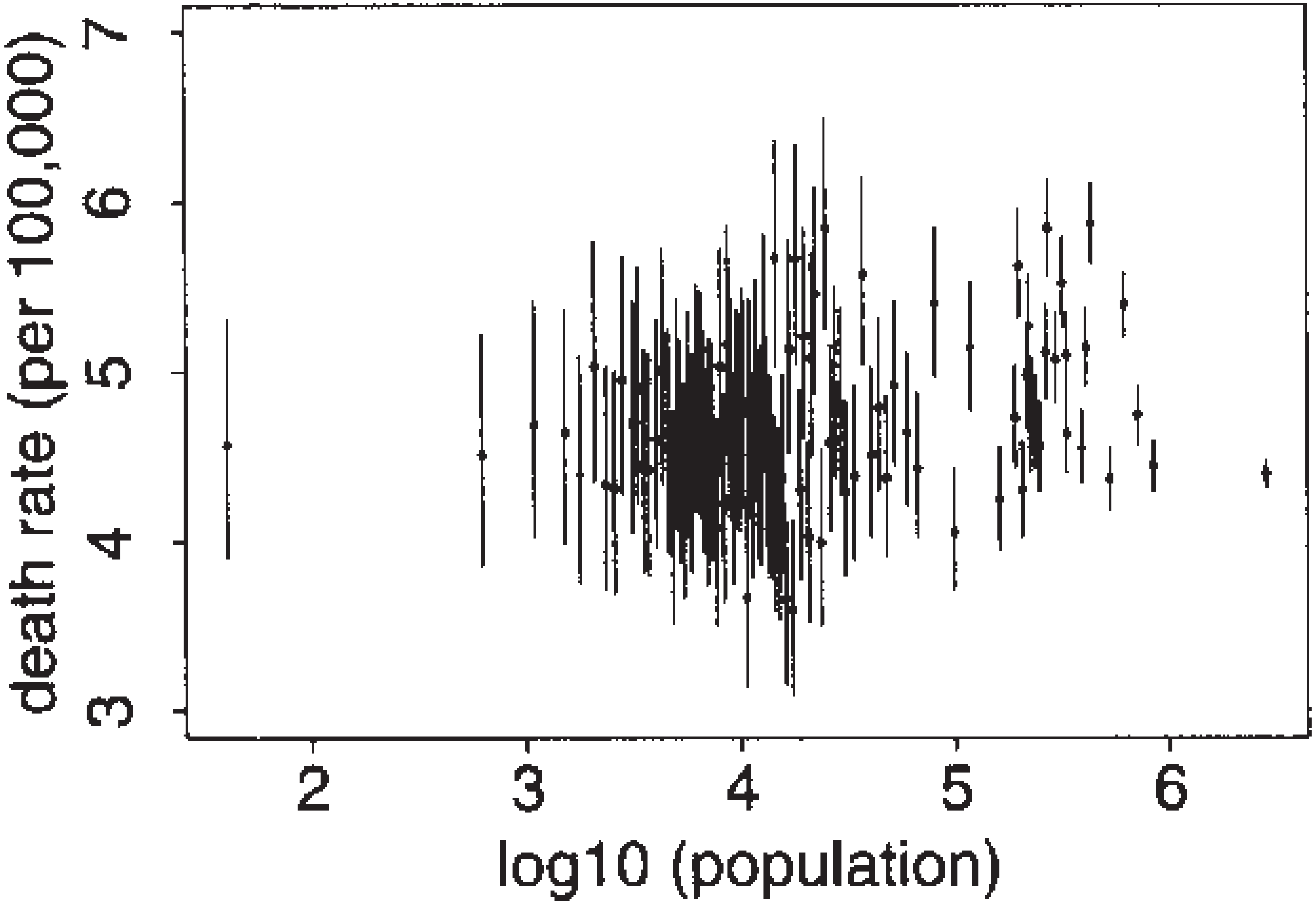
raw data: small counties account for almost all of the high and low death rates

Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)



Bayes estimates: automatically accounts for regression toward the mean

Kidney Cancer Example from Bayesian Data Analysis (Gelman et al)



Bayesian posterior medians and 50% probability intervals

Decision Theory: Nature vs. Data Scientist

Nature picks the parameter, data scientist gets data and then chooses an action (estimate the parameter, predict a future observation, give an interval estimate, accept or reject a hypothesis, ….)

Then some loss is incurred, based on a *loss function*.

Decision Theory: Nature vs. Data Scientist

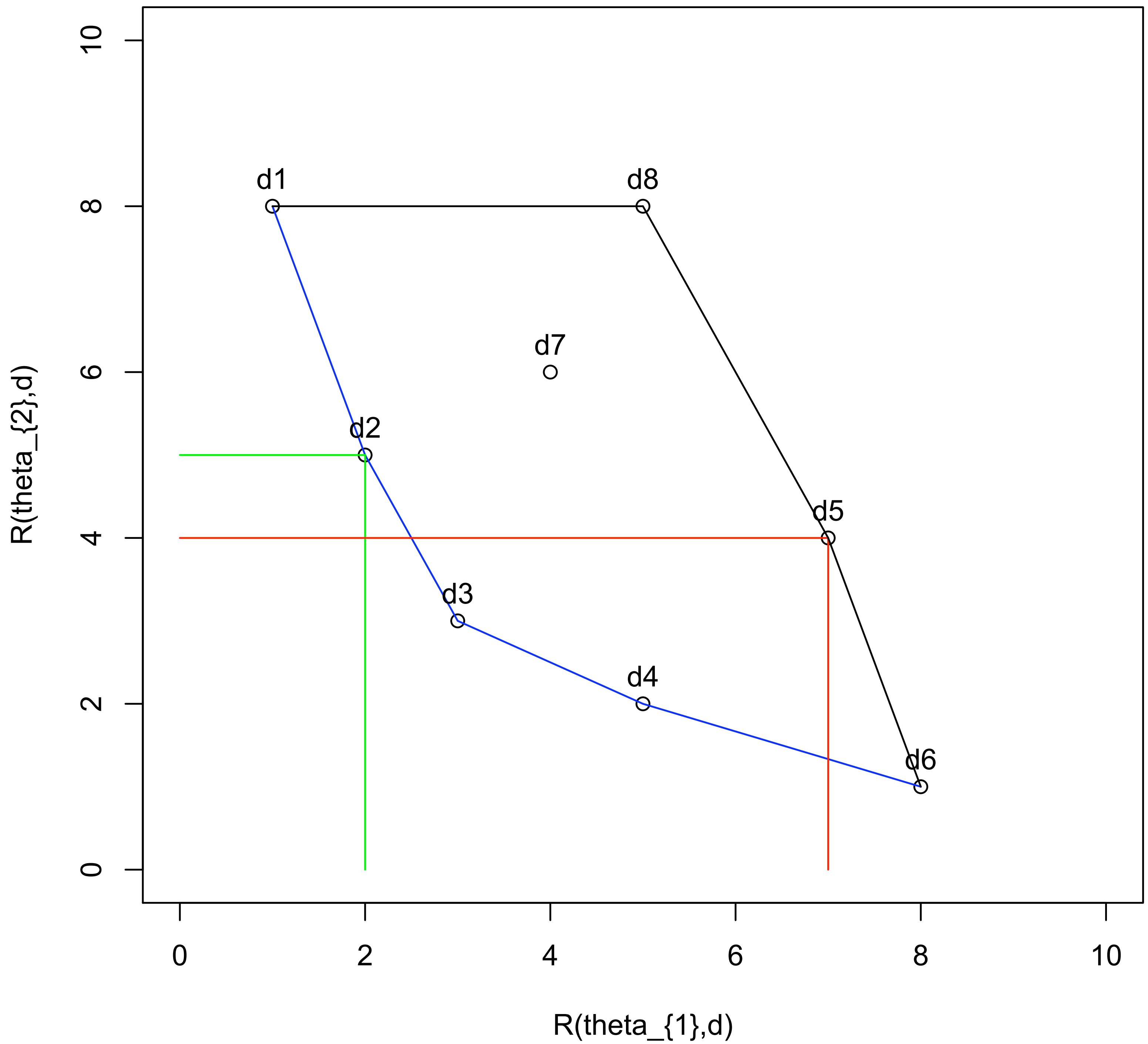
Most common loss functions for estimation:

$$L_2 \text{ (mean square error)} : L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$$

$$L_1 \text{ (mean absolute error)} : L(\theta, \hat{\theta}) = |\hat{\theta} - \theta|$$

Bayesian can try to minimize the expected risk, given the data.
Mean square error: use posterior mean
Mean absolute error: use posterior median

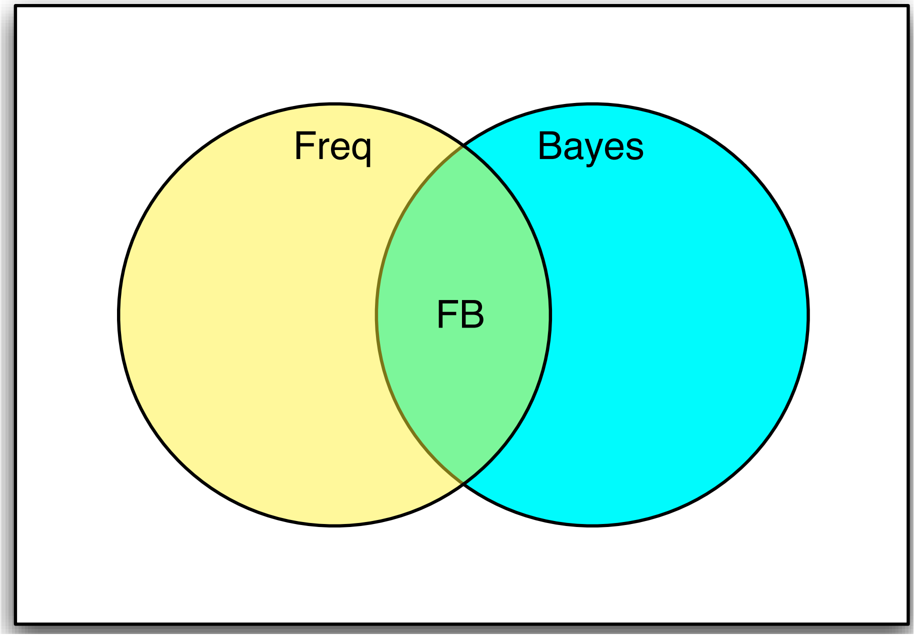
Decision Theory: Geometry of Admissibility



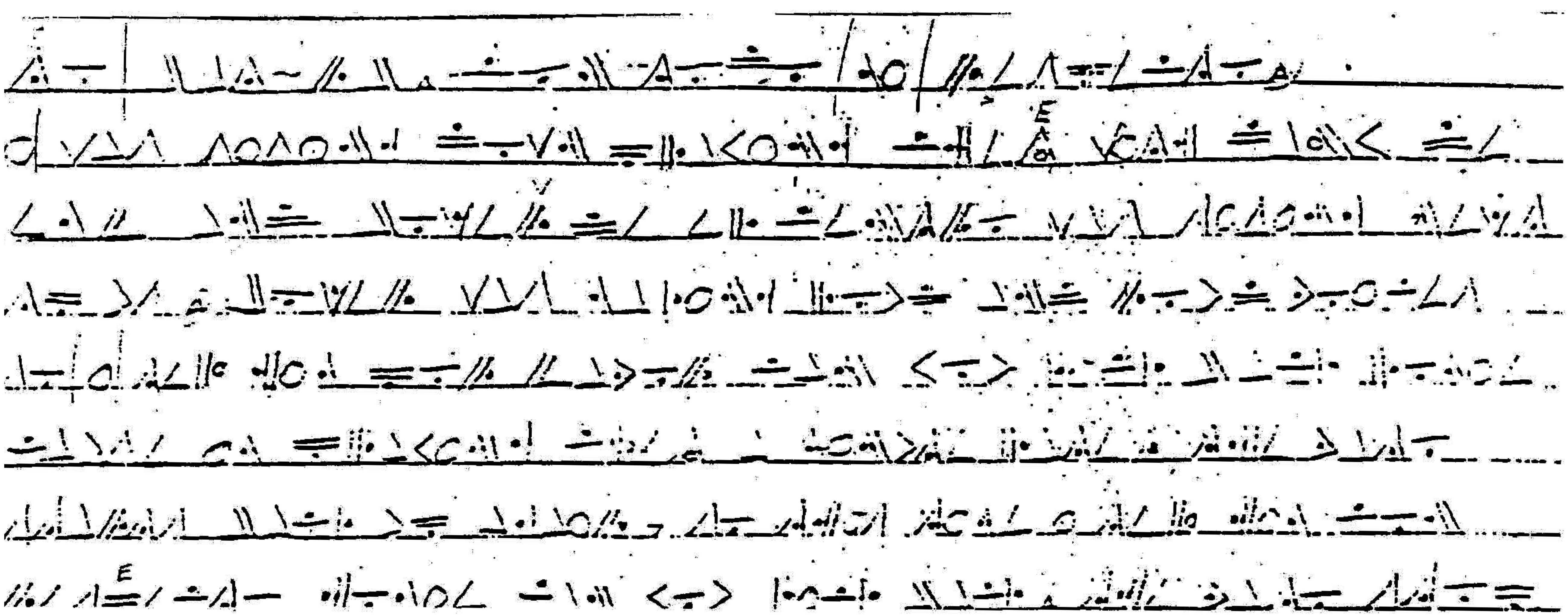
Complete Class Theorem

Under mild technical conditions:

Any Bayesian procedure is admissible.
Any admissible procedure is Bayesian (or a limit of such).



Markov Chain Monte Carlo (MCMC): Diaconis-Coram Example



MCMCryptography

Get a transition matrix M(x,y) for English (the probability of going from letter x to letter y)

Define "plausibility"

$$\mathrm{Pl}(f) = \prod_i M\left(f(s_i), f(s_{i+1})\right),$$

"Try" to swap two random letters in the decoding, based on the ratio of plausibilities.

ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER IN THE MIND TO SUFFER THE SLINGS AND ARROWS OF OUTRAGEOUS FORTUNE OR TO TAKE ARMS AGAINST A SEA OF TROUBLES AND BY OPPOSING END

```
 100 ER ENOHDLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
 200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDEL
 300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOFHE HELO ITOSUVDEL
 400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
 500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOSR O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODIO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFIO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHETHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHETHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBLER
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
```