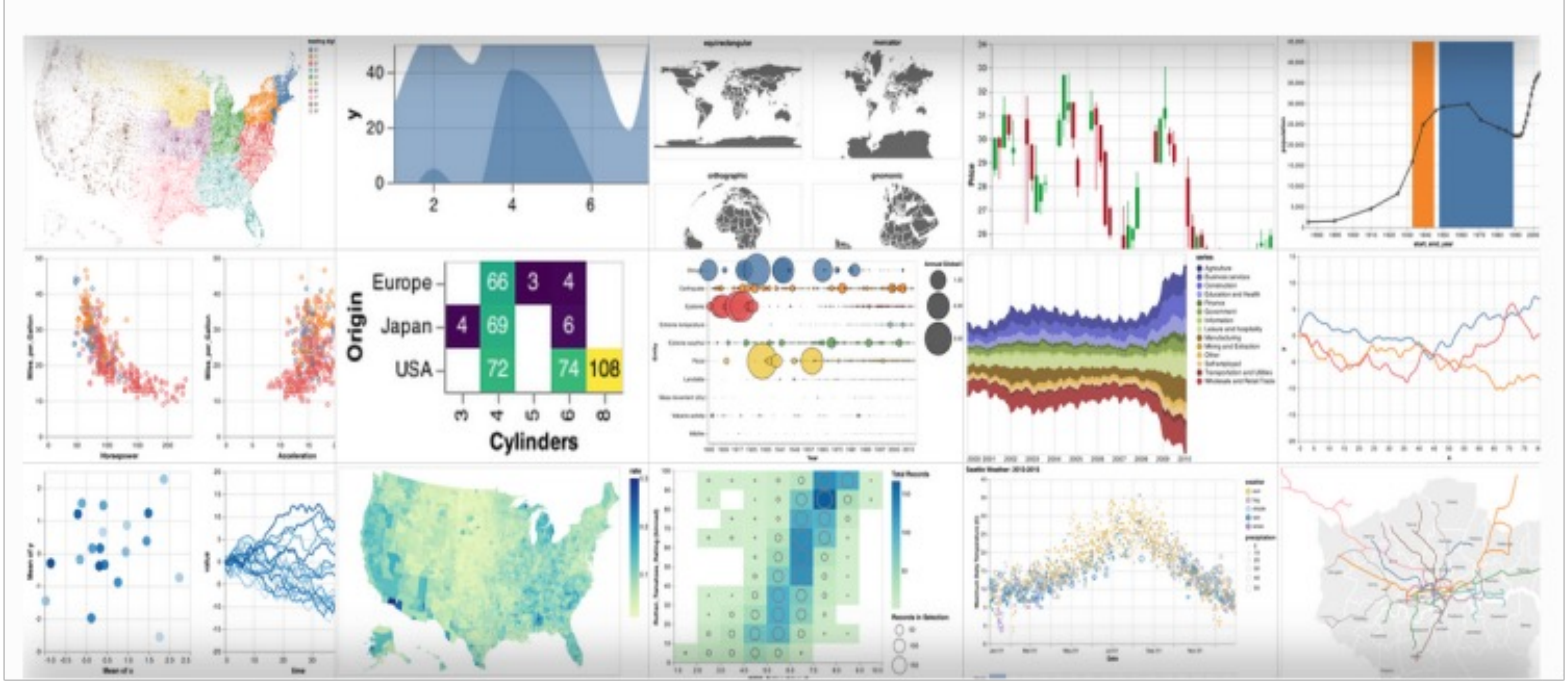


Data 227
Data Visualization & Communication

Beta distributions for communicating uncertainty in proportions

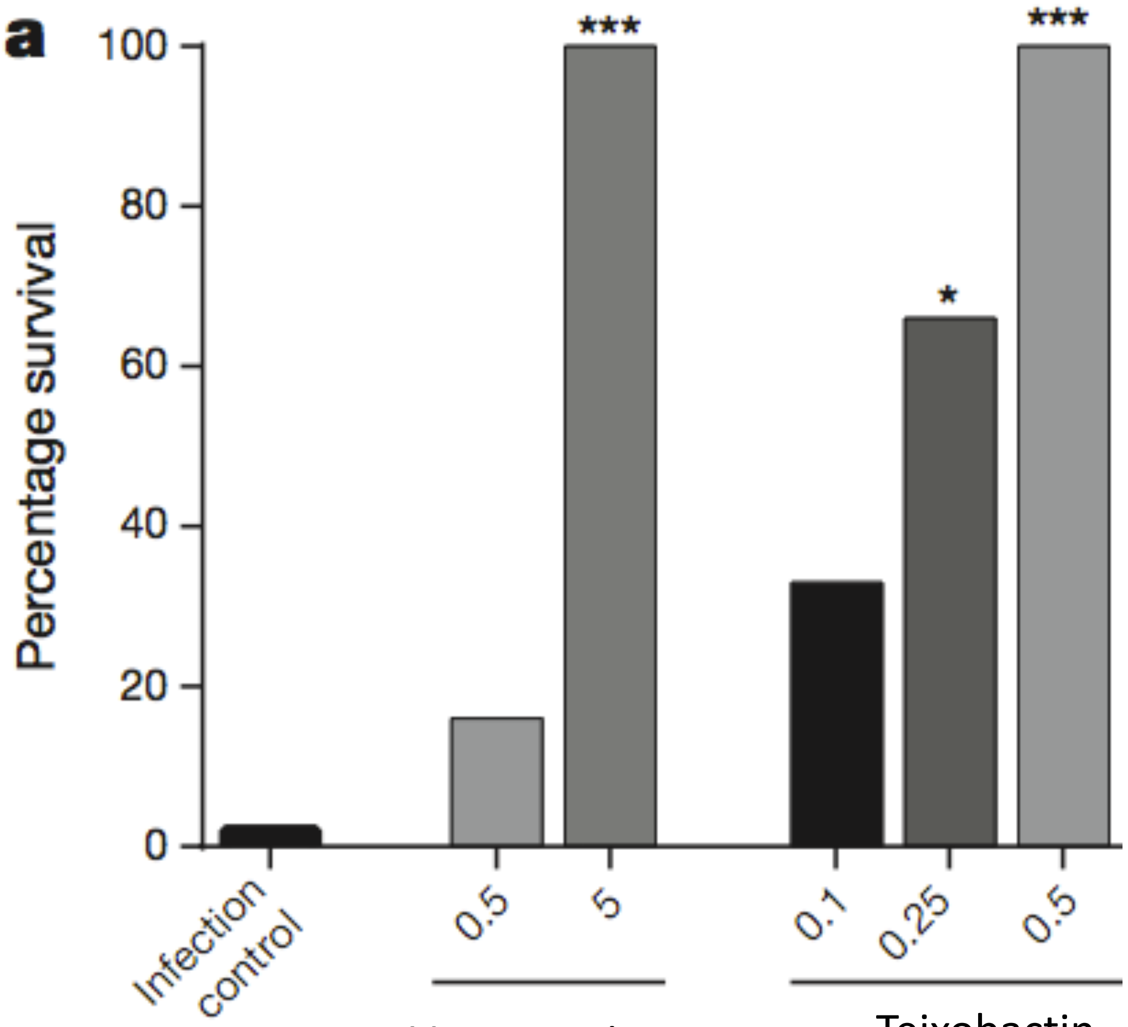
A new antibiotic

Ling et al. Nature 517:455 (2015)

The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

Mice were given a lethal infection, and then treated with two antibiotics at different doses.

Something about the graph is funny.



Vancomycin (mg per kg)

Teixobactin (mg per kg)

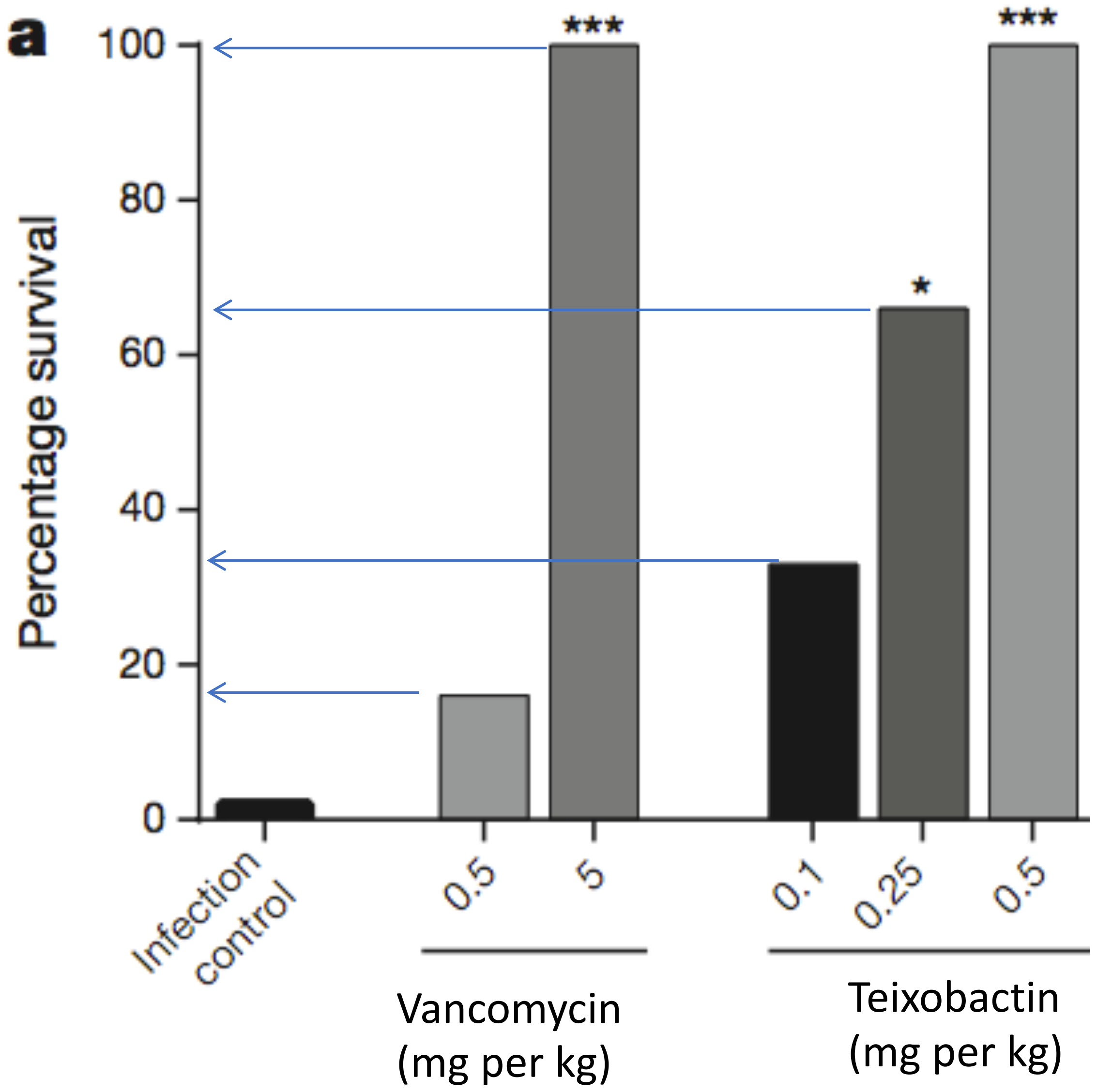
A new antibiotic

Ling et al. Nature 517:455 (2015)

The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

Mice were given a lethal infection, and then treated with two antibiotics at different doses.

Something about the graph is funny.



A new antibiotic

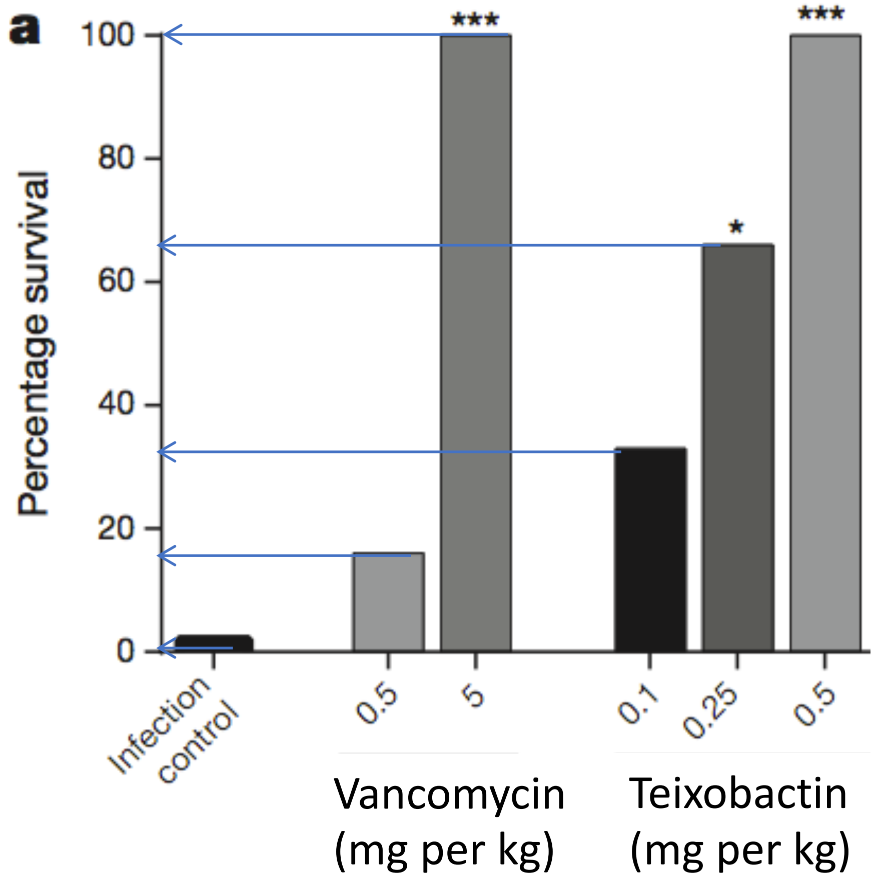
The paper announced the discovery of a new polypeptide antibiotic (tiexobactin).

Mice were given a lethal infection, and then treated with two antibiotics at different doses.

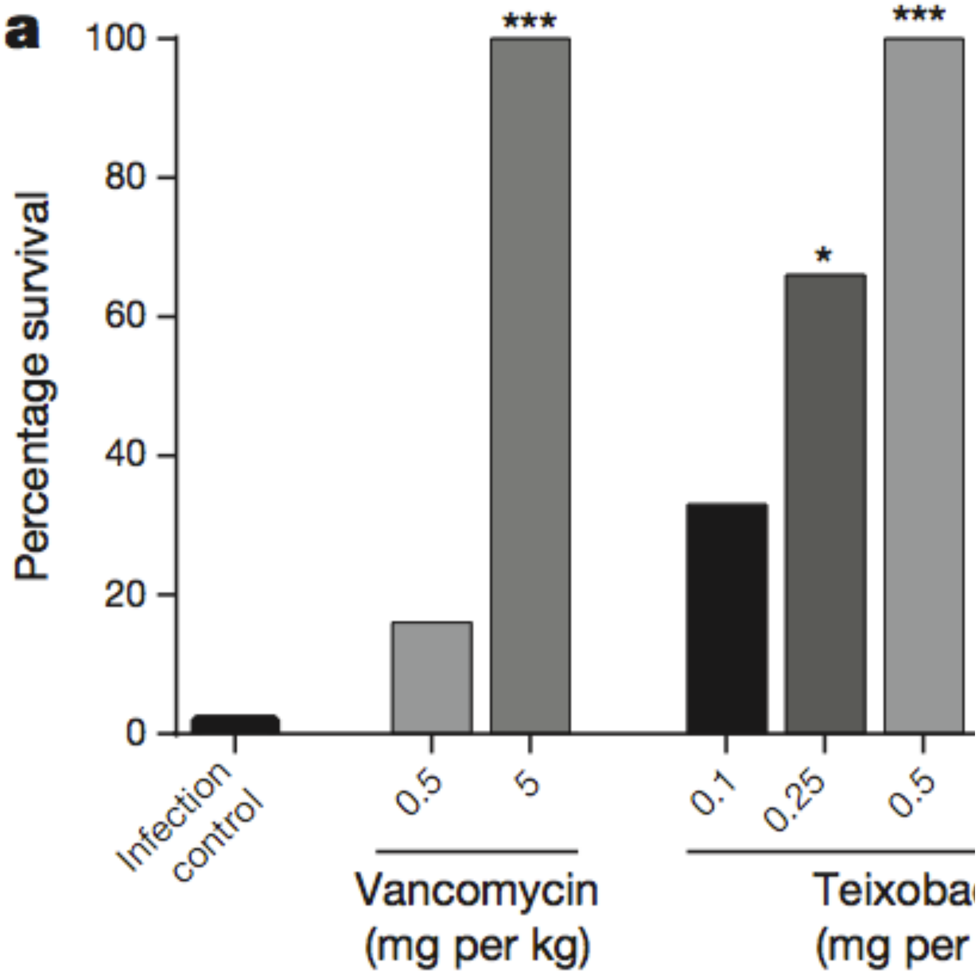
Something about the graph is funny.

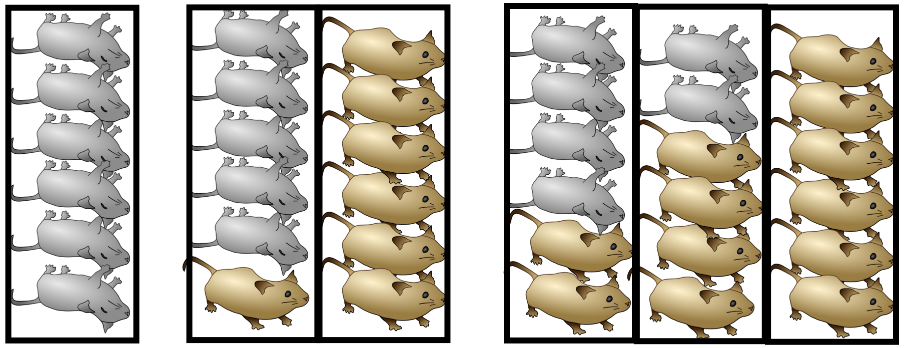
All of the bars are multiples of 1/6.
Read the fine print..

Ling et al. Nature 517:455 (2015)



This graph summarizes the fates of thirty-six mice, six cages of six mice each.





Binomial likelihood

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ {}_{n}C_{\mathrm{r}}\, p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ \binom{n}{k} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\mathrm{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ \frac{n!}{k!(n-k!)} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

Binomial likelihood

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \;{}_{n}C_{\mathrm{r}}\, p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \;\binom{n}{k} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\, \mathrm{n},\, \mathrm{p}) = \;\frac{n!}{k!(n-k!)} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

Binomial likelihood

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ {}_{n}C_{\mathrm{r}}\ p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ \binom{n}{k} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

$$P_{\text{Binomial}}\,(\mathrm{k};\,\mathrm{n},\,\mathrm{p}) = \ \frac{n!}{k!(n-k!)} p^{\mathrm{k}}(1-p)^{n-\mathrm{k}}$$

This is the number of outcomes

This is the probability of each outcome

Binomial likelihood

$$P_{\text{Binomial}}(p;\, n, k) = \; f(n,k)\; p^{k}(1-p)^{n-k}$$

This is the number of outcomes

This is the probability of each outcome

Binomial log - likelihood

$$\log P_{\text{Binomial}}(\mathrm{p};\, \mathrm{n}, \mathrm{k}) = \log(\, f(n,k)\,) \;+\; k\log(p) \;+ (\mathrm{n}\text{-}\mathrm{k}) \log\,(1-p)$$

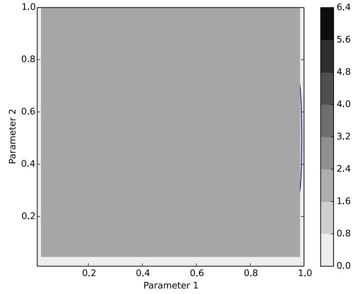
Term that depends on n, k but independent of p

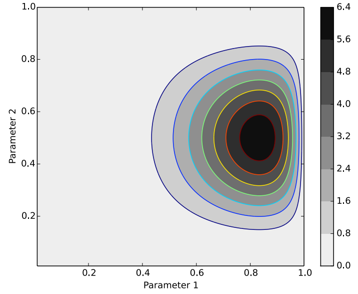
Terms that depend on p

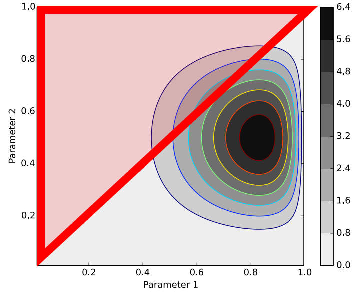
All Bayesian problems are the same.

- Construct prior distribution consistent with ignorance of the parameters
  - Laplace (uniform) prior = beta(x;1,1) and
  - Jeffreys prior = beta(x;0.5,0.5)

  are the only serious contenders
- Evaluate posterior distribution in light of the data (joint binomial)
- Integrate posterior distribution to get parameters of interest.
  (Beta inequality)







Binomial log - likelihood

$$\log P_{\text{Binomial}}(\text{p};\, \text{n},\, \text{k}) = \log(\, f(n,k)\,) + k\log(p) + (\text{n-k})\log\,(1-p)$$

Term that depends on n, k but independent of p

Terms that depend on p

Sometimes we use log-likelihood because it turns multiplication into addition and this makes some inferences easier.

Other times we use log-likelihood for numerical reasons; we may want to take the ratio of two likelihoods that are smaller than 10^{-16}

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

a and b are number
of successes and number
of failures (almost)
k, (n-k)

This is a number

Two terms with powers of
x and *1-x*
x is probability of success

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1}$$

The late David MacKay pointed out that probability distributions are often named for their normalization constants.

"Normalization constant" – jargon for "the number you have to multiply the math by to make the probability sum to 1"

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1}(1-x)^{b-1}$$

You can think of this as the binomial probability reorganized to add up to 1 when integrated over x instead of summed over k.

The Univariate Beta distribution

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}$$

Expectation value

$$\mu = \frac{a}{a+b}$$

Standard deviation

$$\sigma = \frac{ab}{(a+b)^2(a+b+1)}$$

Mode

$\frac{a-1}{a+b-2}$ for a, b > 1

any value in [0,1] for α, β = 1

{0, 1} (bimodal) for α, β < 1

0 for $\alpha \leq 1$, $\beta > 1$

1 for $\alpha > 1$, $\beta \leq 1$

The math

Univariate Beta distribution:

$$f(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

$$p(x;0/3) = 4(1-x)^3$$

$$p(x;1/3) = 12x(1-x)^2$$

$$p(x;2/3) = 12x^2(1-x)$$

$$p(x;3/3) = 4x^3$$

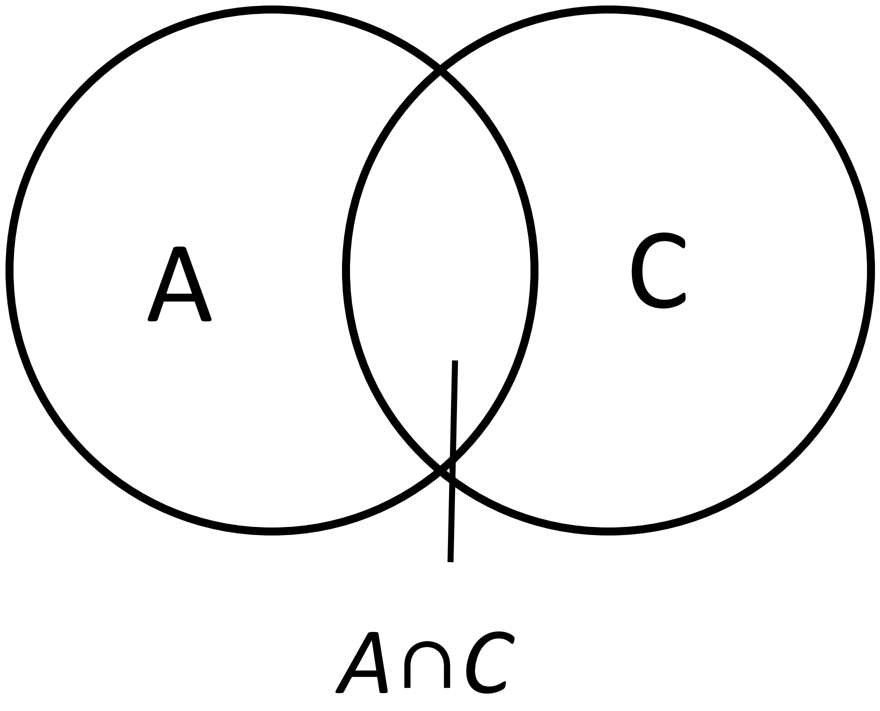
The posterior distributions for 0, 1, 2, and 3 successes out of 3 are “polynomial functions” known as middle- and high-school torture instruments.

Bayesian inference

Event A: unknown state of nature

Event C: experiment

$$P(A \mid C) = P(C|A)\, P(A) / P(C)$$



Bayesian inference

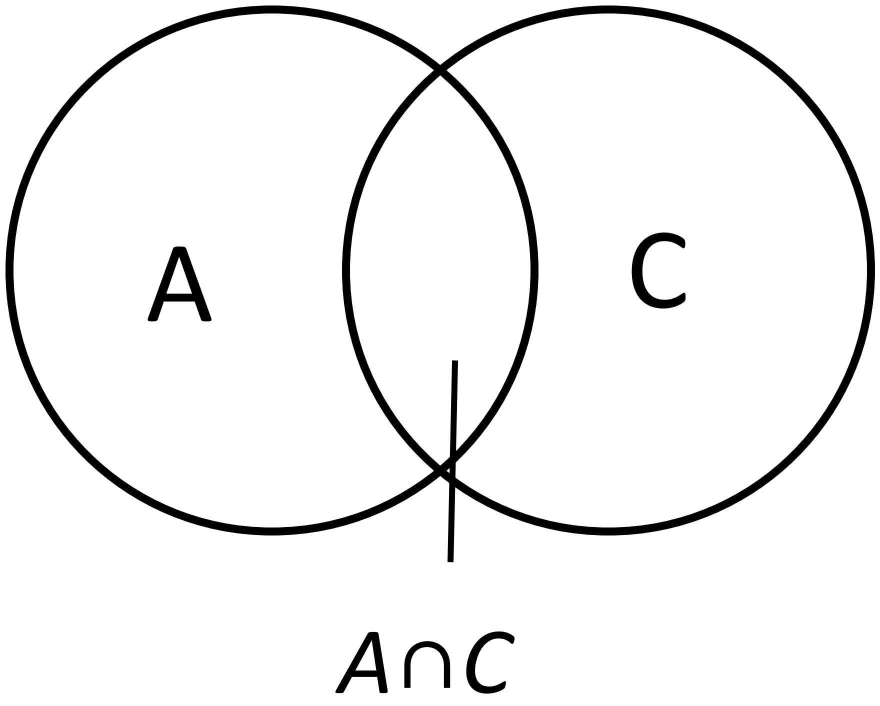
Event A: am I infected?

Event C: experiment (test result)

$$P(A \mid C) = P(C|A)\, P(A) / P(C)$$

The thing I want to know

The thing the FDA wants know



Bayesian inference

Event A: unknown state of nature

Event C: experiment

$$P(A \mid C) = P(C|A)\, P(A) \,/\, P(C)$$

Life's persistent question "What is death rate?"

Term that depends on experiment
in this case binomial

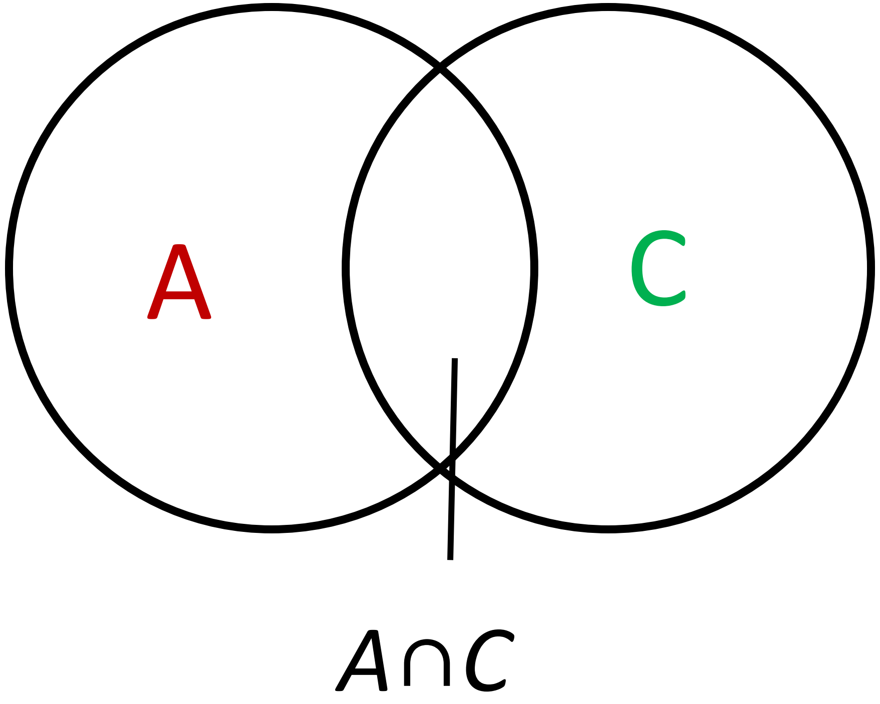
likelihood

Term with knowledge and bias about A

prior

Probability for event that is no longer uncertain

evidence



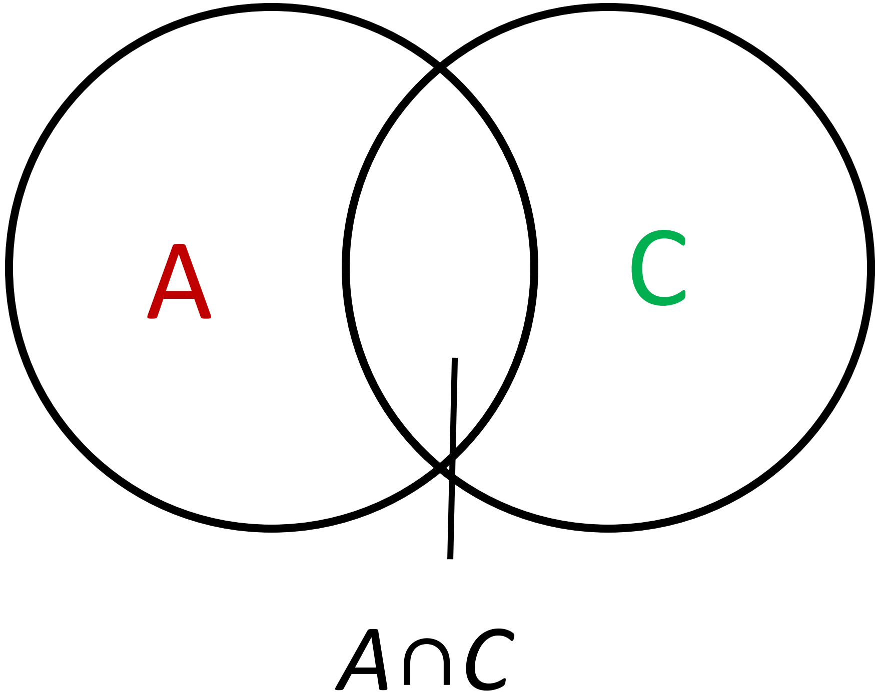
Bayesian inference

Event A: unknown state of nature

Event C: experiment

$$P(A \mid C) = P(C|A)\,P(A) \,/\, P(C)$$

$P(A \mid C)$	$P(C\|A)$	$P(A)$	$P(C)$
Desired probability	We know this from practice with H_0	Population prevalence	Doesn't matter for us
	likelihood	*prior*	*evidence*



Bayesian inference for proportions

Event A: unknown state of nature true proportion

Event C: experiment sample proportion

$$P(A \mid C) = P(C|A)\,P(A)\,/\,P(C)$$

Posterior probability density for proportion

Binomial

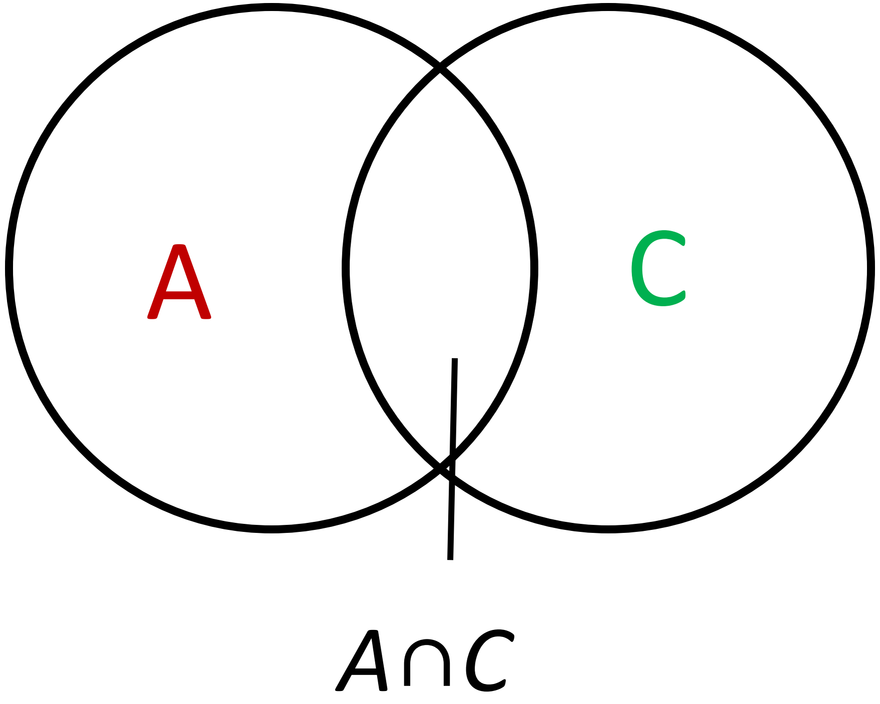
likelihood

Prior dist. for prop.

prior

Doesn't matter for us

evidence



posterior density = prior density * likelihood (function of the data)

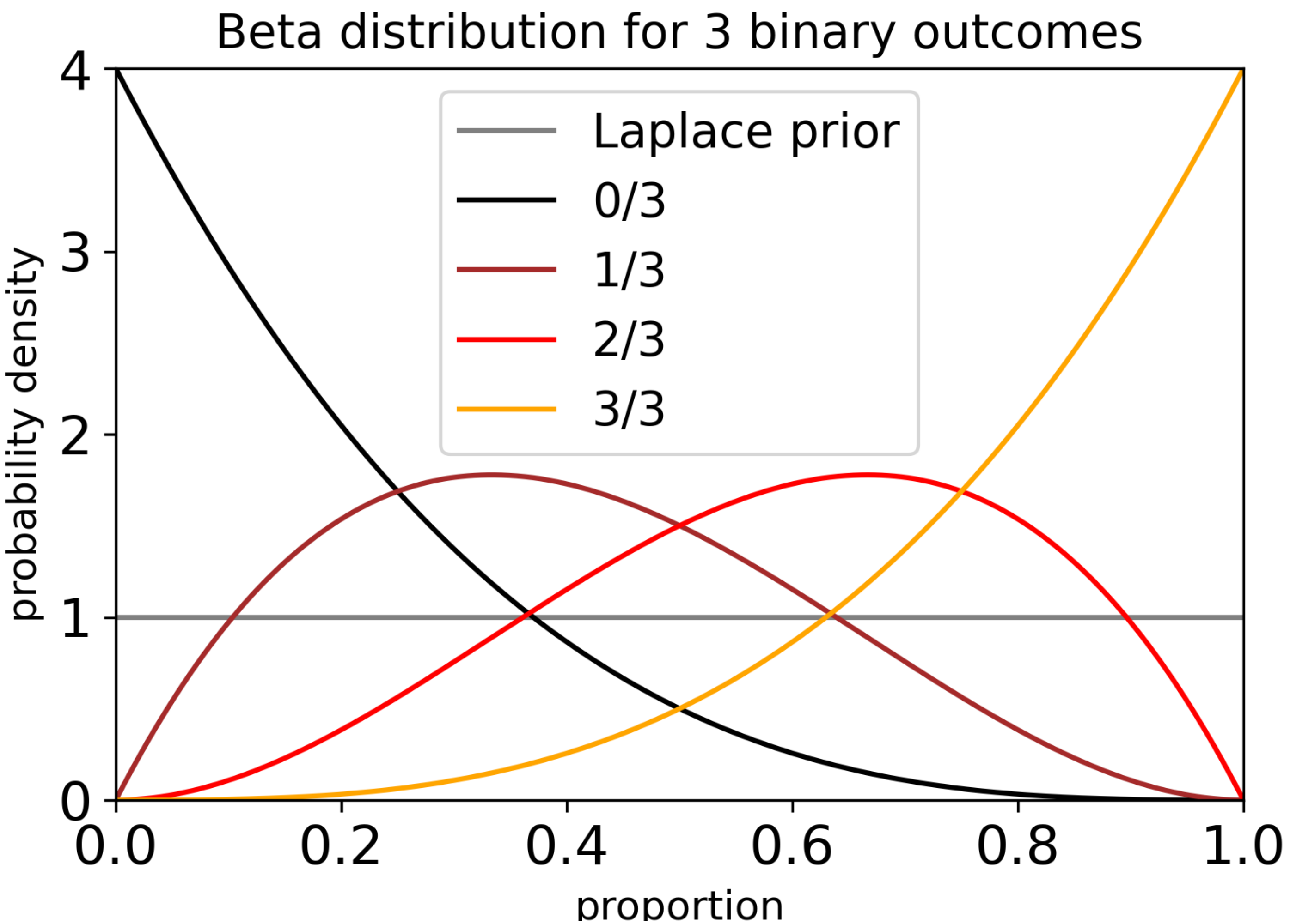
Beta is a conjugate prior for the binomial distribution:
Beta priors * binomial likelihoods = Beta posteriors

Beta (α, α) prior
x Binomial likelihood with m successes and n failures
$= \text{Beta}(\alpha + m, \alpha + n)$

As long as we write a paragraph justifying our choice of prior, we can easily get exact confidence intervals for parameters known by binomial sampling.

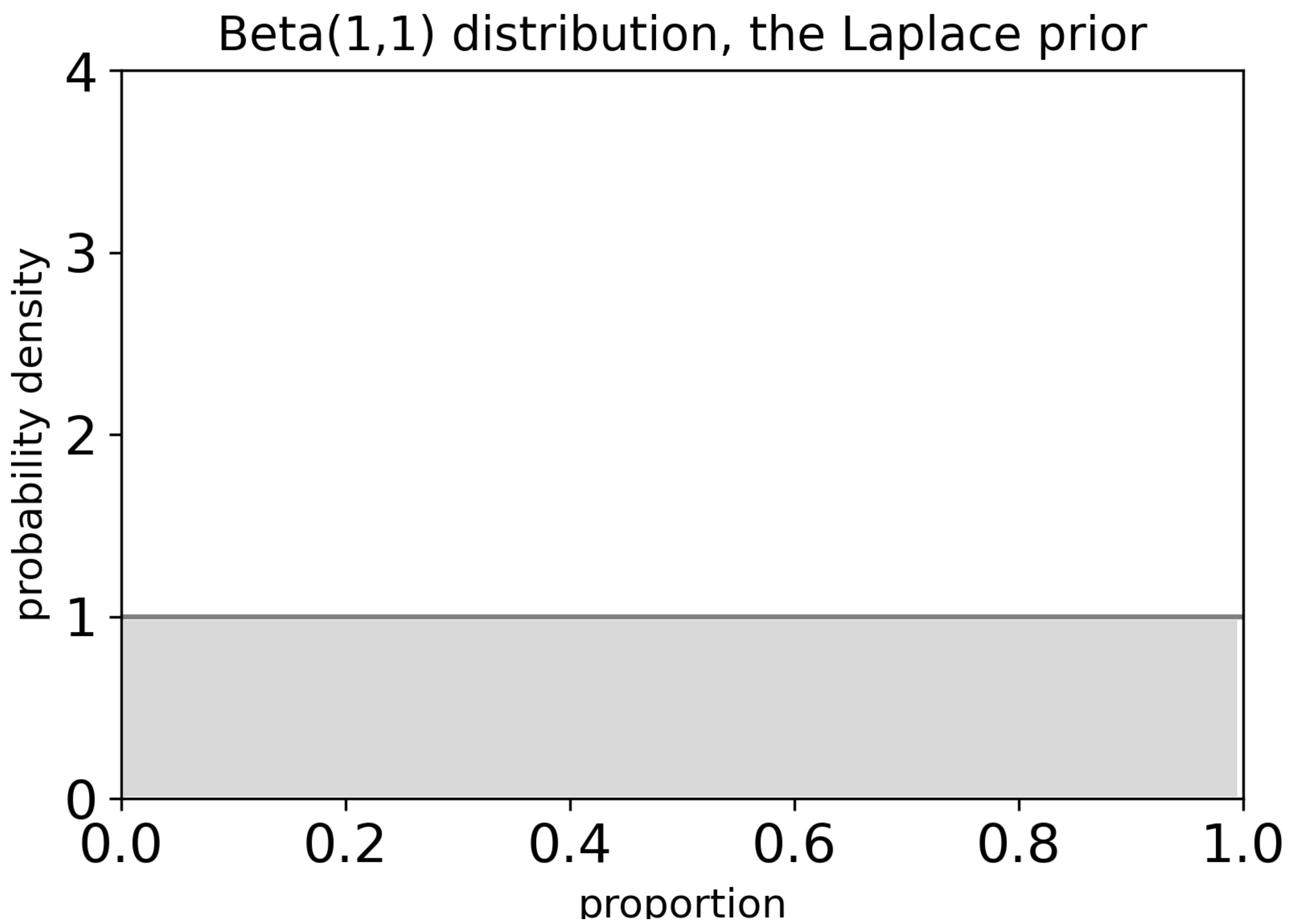
Beta distributions for all possible outcomes of an n=3 binomial experiment

$$p(x; 0/3) = 4(1-x)^3$$
$$p(x; 1/3) = 12x(1-x)^2$$
$$p(x; 2/3) = 12x^2(1-x)$$
$$p(x; 3/3) = 4x^3$$



What does ignorance look like?

Before I do the experiment,
I don't know much about
the proportion.
But we're grownups, so I
have to put numbers on
my ignorance.



The oldest choice is the "flat" prior

Beta(a=1, b=1) = $x^0 (1-x)^0 = 1$

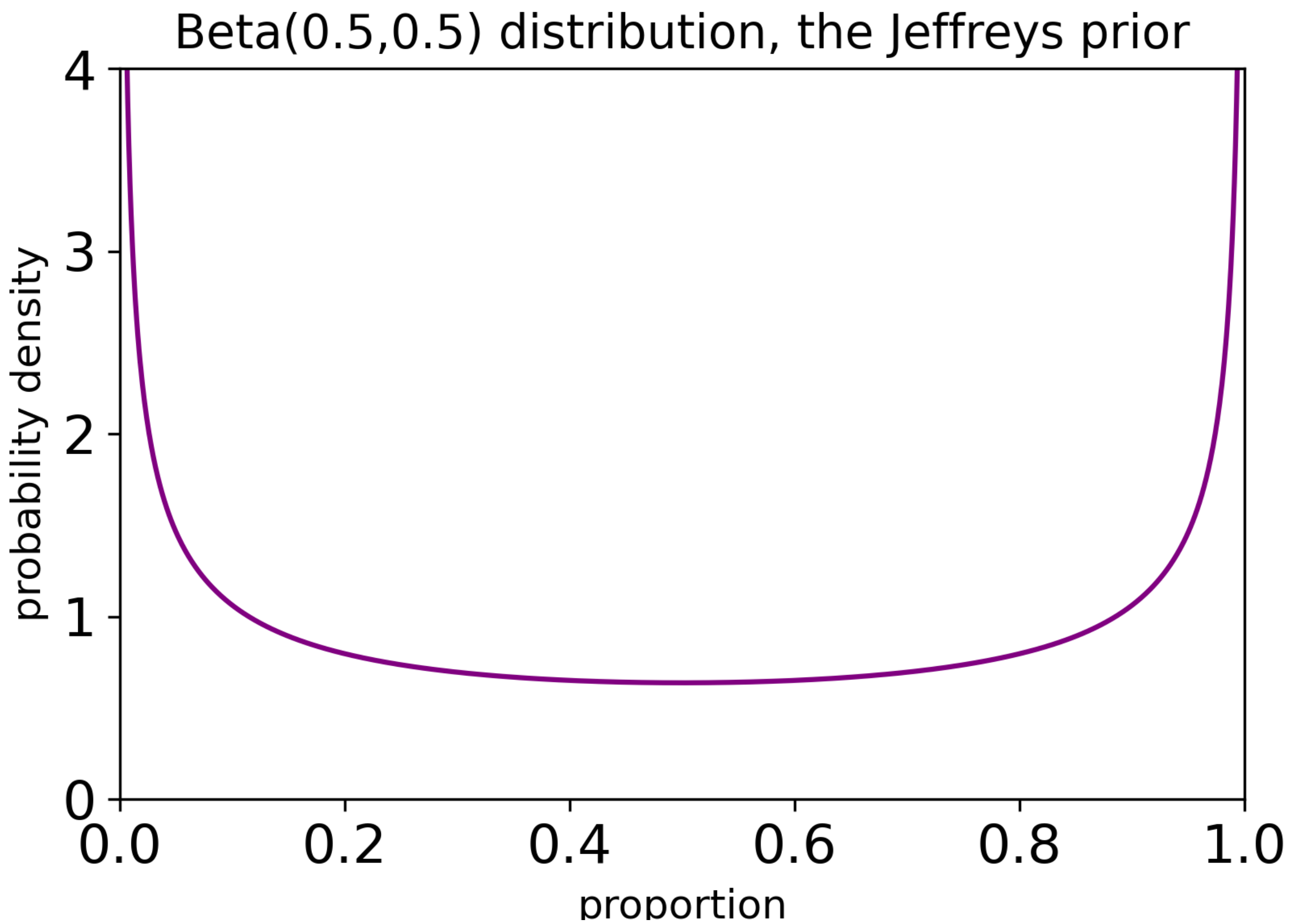
This prior introduces "bias" — it has an expectation value of 0.5 !!!

What does ignorance look like?

Perhaps there is reason to prefer the extremes?

Jeffreys prior does that.

Note the symmetry: if a = b, your prior does not prefer successes or failures.

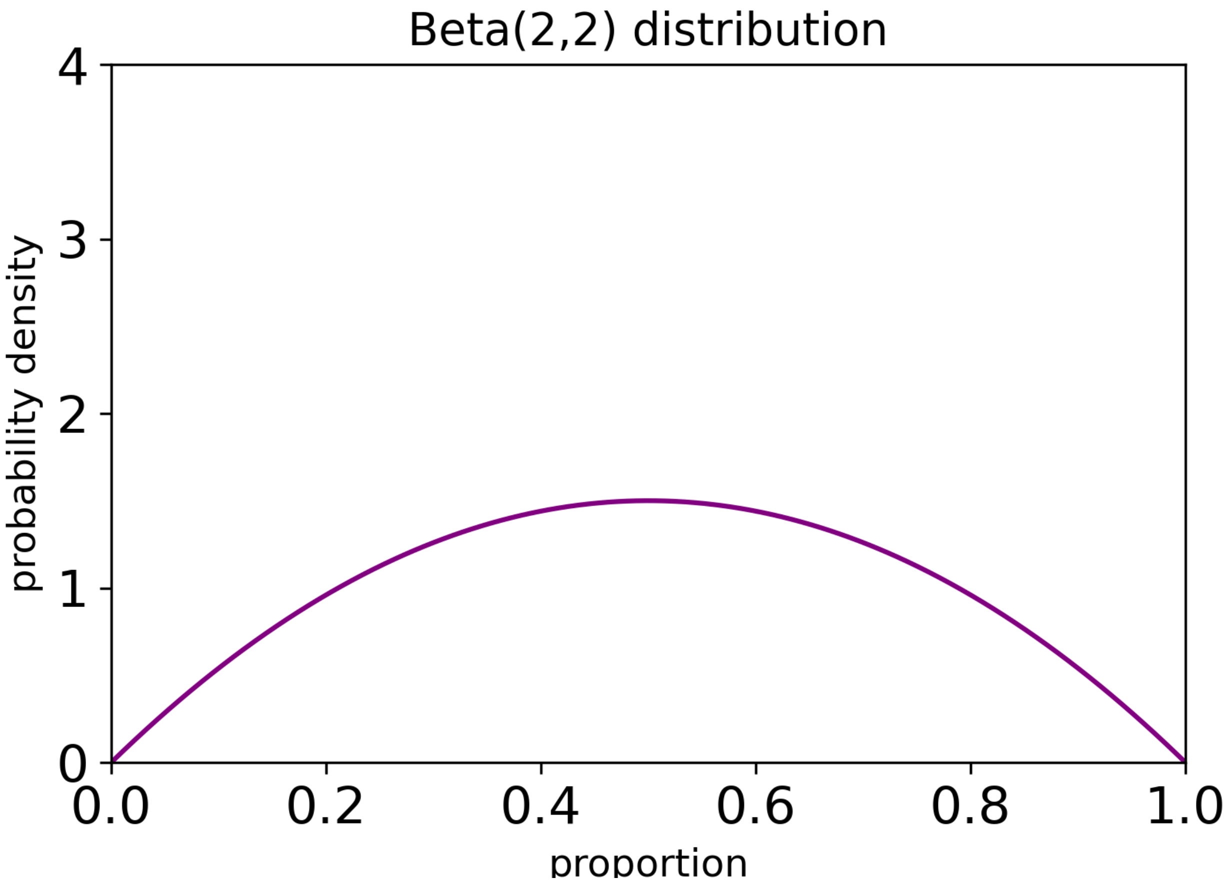


Beta(a=0.5, b=0.5) = c $x^{-0.5}$ $(1\text{-}x)^{-0.5} = \dfrac{c}{\sqrt{p(1-p)}}$

What does ignorance look like?

Note the symmetry: if a = b,
the prior does not prefer successes
or failures.

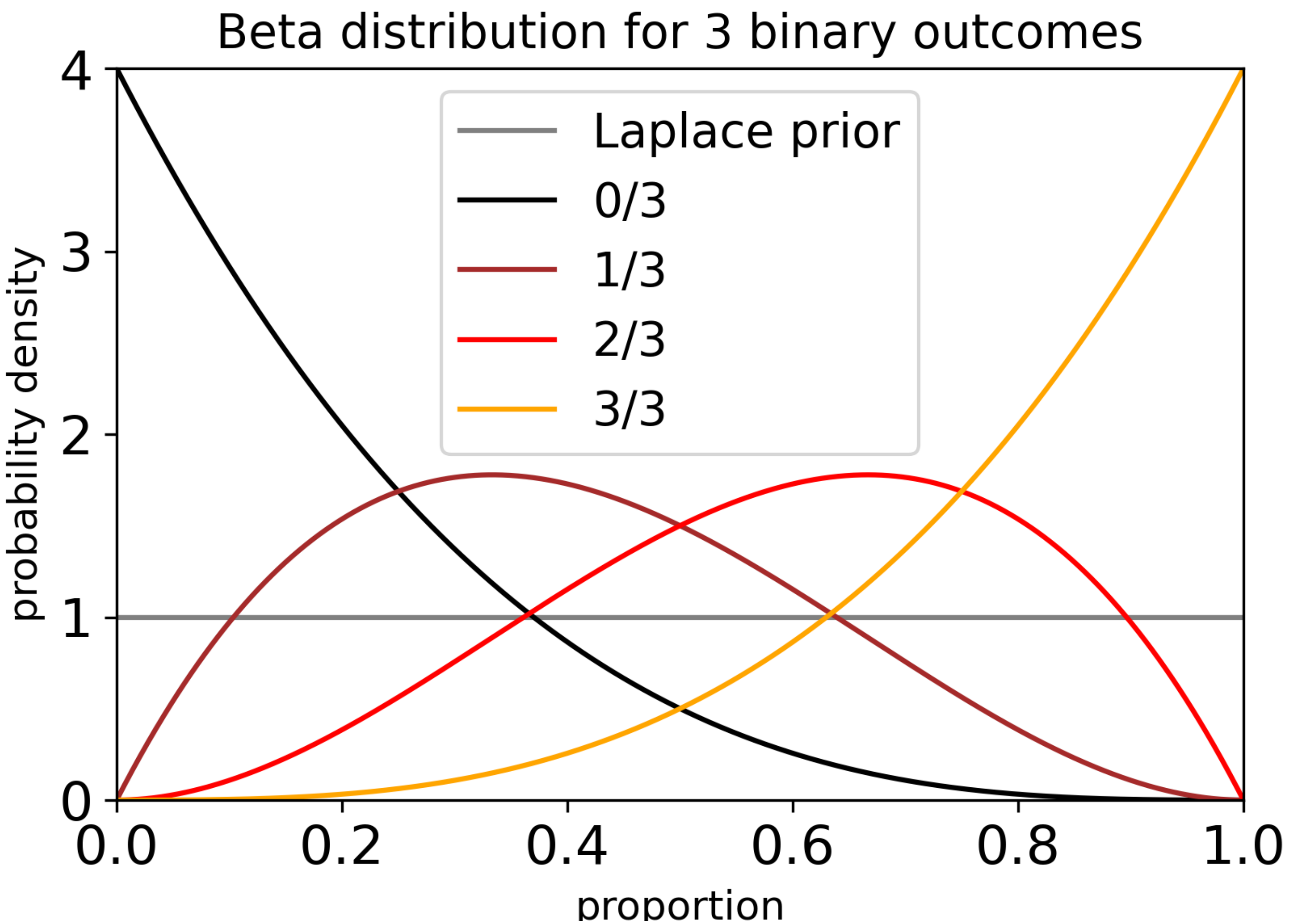
a = b = 2 is reasonable if we are certain
that both successes and failures
are possible (p cannot be exactly 0 or
exactly 1)



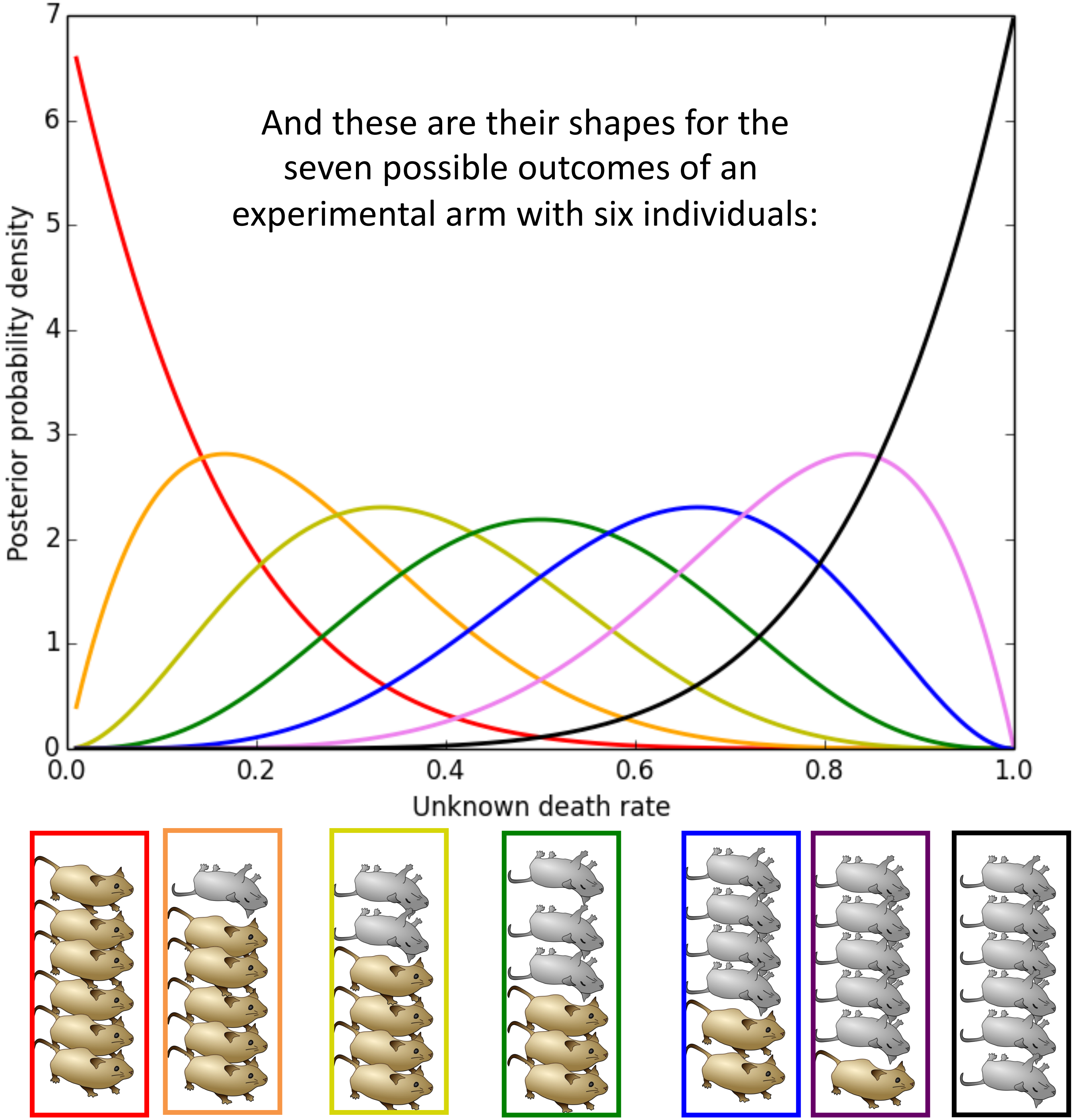
Beta(a=2, b=2) = c x (1-x)

Beta for n=3 binomial trial

$$p(x; 0/3) = 4(1-x)^3$$
$$p(x; 1/3) = 12x(1-x)^2$$
$$p(x; 2/3) = 12x^2(1-x)$$
$$p(x; 3/3) = 4x^3$$



And these are their shapes for the seven possible outcomes of an experimental arm with six individuals:



The biased top

200 spins of a four-sided top.

Outcome A: 1

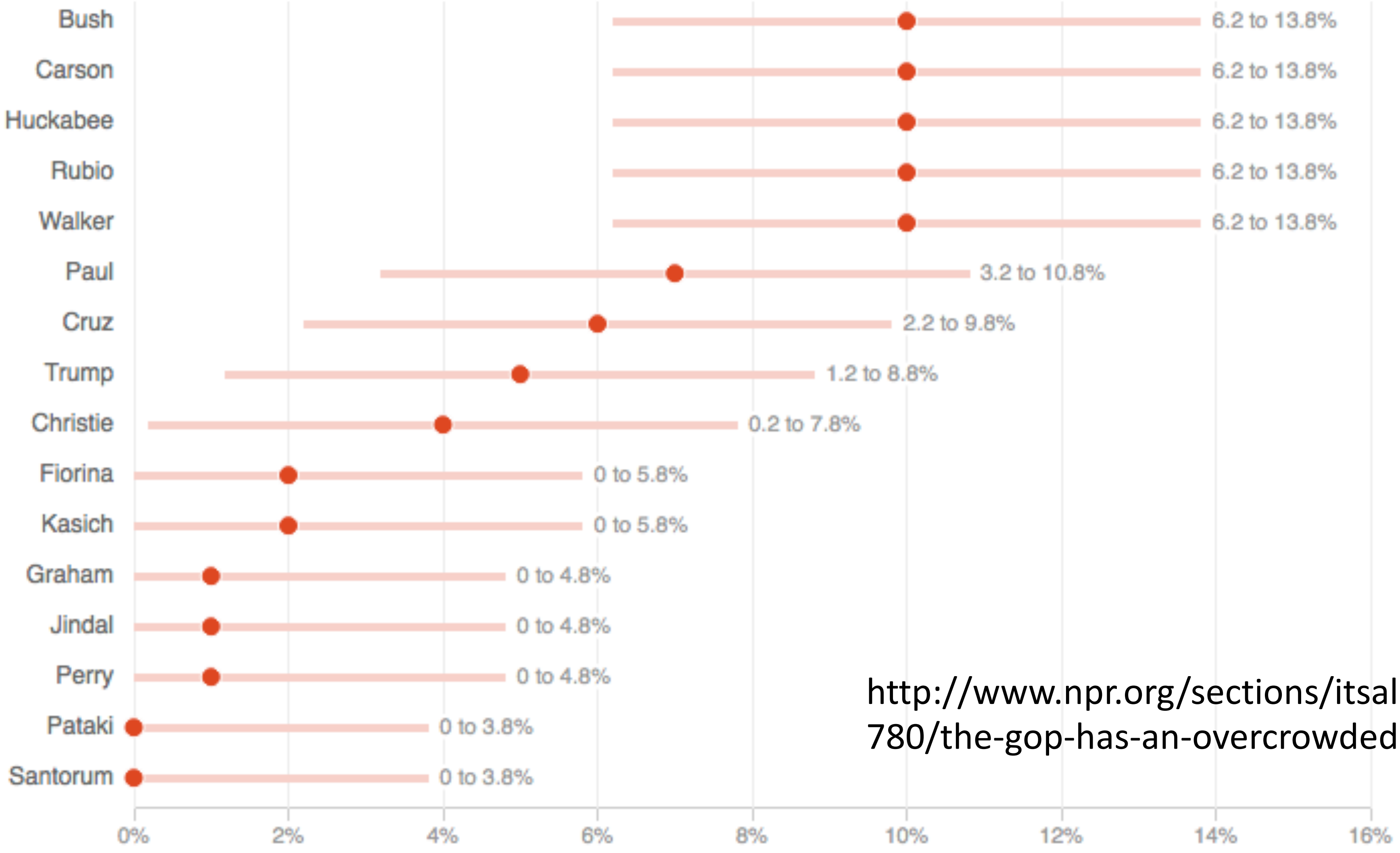
Outcome B: 34

Outcome C: 87

Outcome D: 78

Republican Candidate Support, Factoring In The Margin Of Error

Among respondents who said they were Republican or leaning Republican



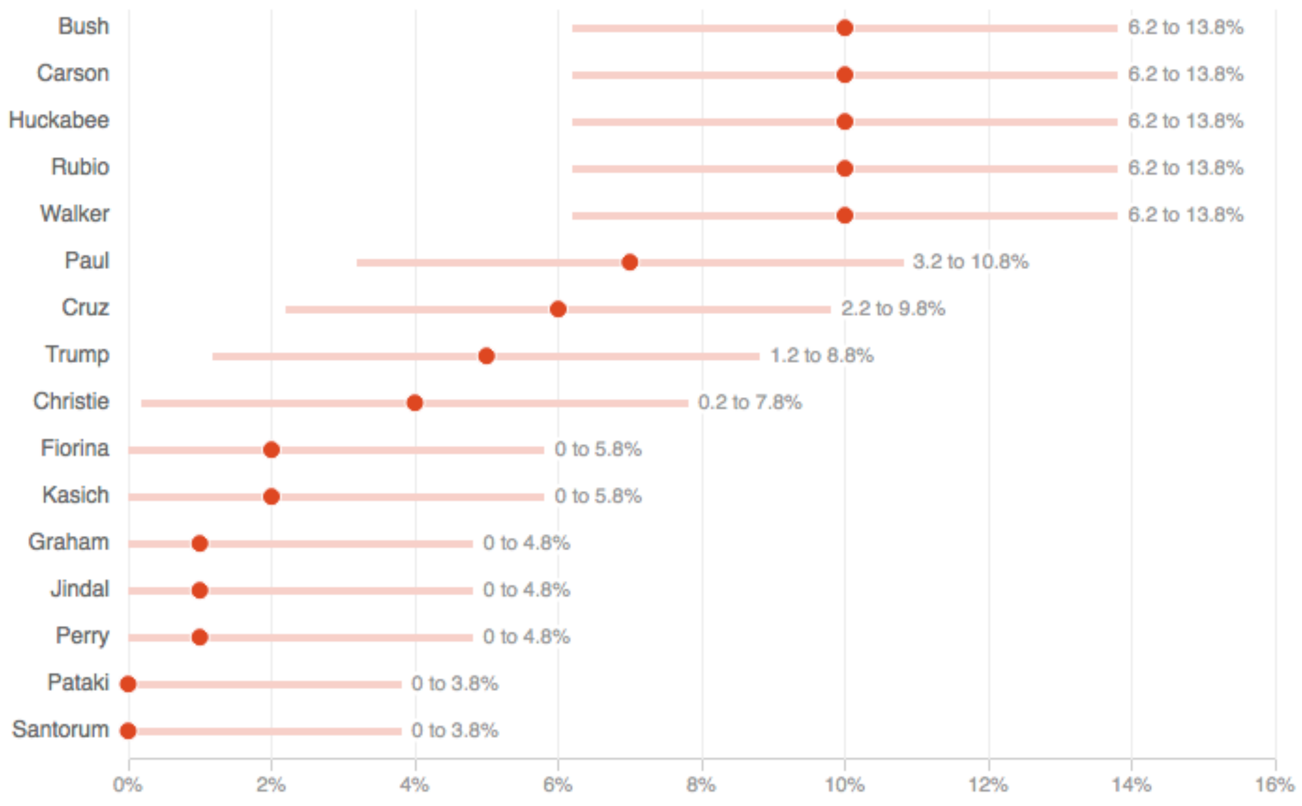
http://www.npr.org/sections/itsallpolitics/2015/05/29/410524780/the-gop-has-an-overcrowded-debate-problem

Source: Quinnipiac University poll taken May 19-26. The survey included 679 Republicans, with a margin of error of +/- 3.8 percentage points.

Credit: Alyson Hurt and Danielle Kurtzleben/NPR

Republican Candidate Support, Factoring In The Margin Of Error

Among respondents who said they were Republican or leaning Republican



Source: Quinnipiac University poll taken May 19-26. The survey included 679 Republicans, with a margin of error of +/- 3.8 percentage points.

Credit: Alyson Hurt and Danielle Kurtzleben/NPR

- Proportions have been rounded.
- Error bars are symmetrical, clipped at zero
- Error bars are independent of point estimates!

In essence, the error bars are inappropriate.

The biased top

200 spins of a four-sided top.

Outcome A:	1	Here be dragons
Outcome B:	34	0.17± 1.96 * S.E.M.
Outcome C:	87	0.435 ±1.96 *S.E.M.
Outcome D:	78	0.39 ± 1.96*S.E.M.
Outcome NOT A :	199/200	(can’t use Standard Error on the Mean!)

The biased top

200 spins of a four-sided top.

1
34
87
78