Visual Complements to p-values
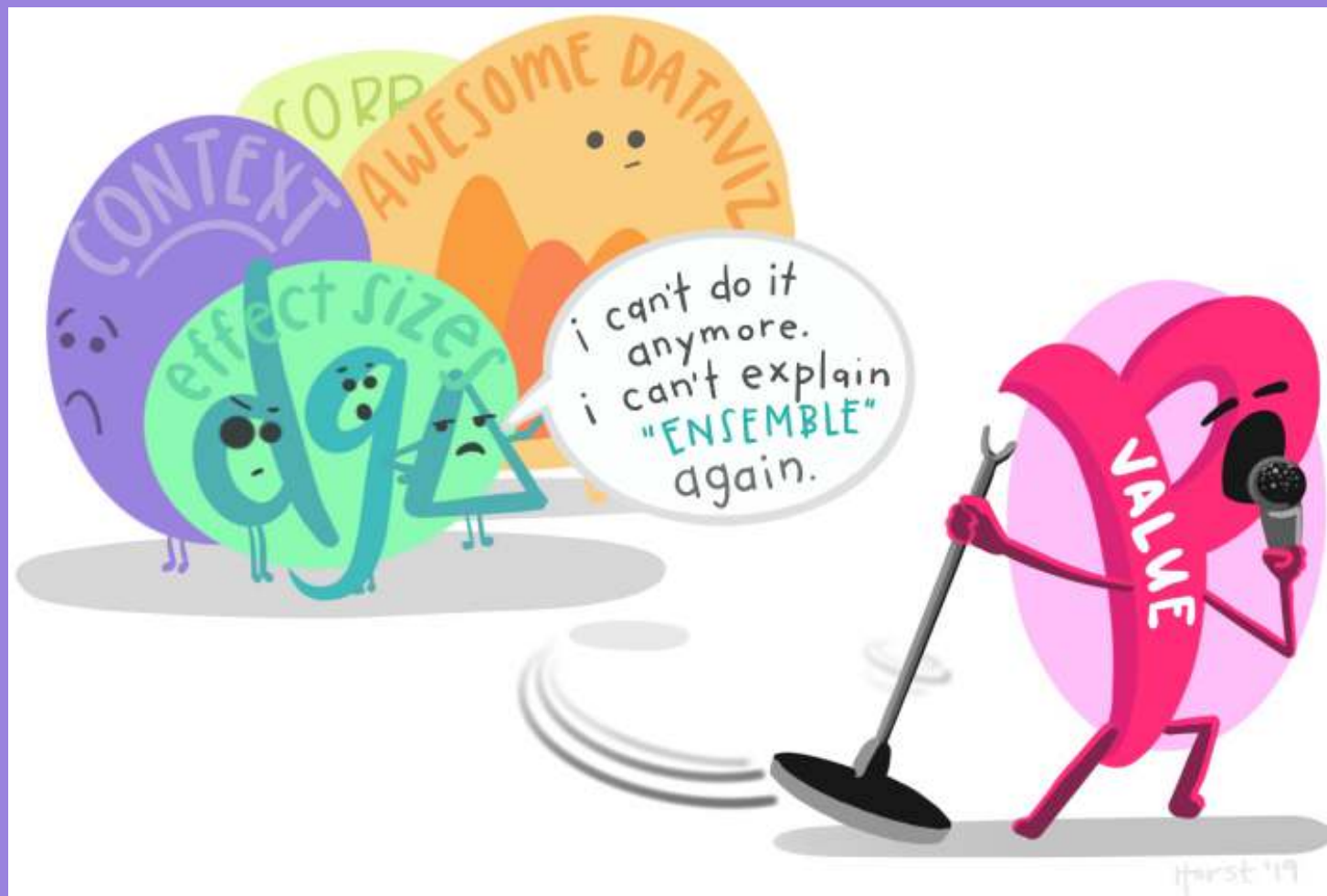
FCSM 2020

14 Sep. 2020

Samantha C. Tyner, Ph.D.

sctyner
sctyner
sctyner.me

Slides available at bit.ly/viz-fcsm

Image credit: Allison Horst

Case Study: Early use of "p-values"

John Arbuthnot. Source: Wikipedia

John Arbuthnot, "An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes" [1]

- Used probability ("p-value") to "prove" divine intervention in the human birth ratio (males are more common than females)
- Probability calculation that today would be the p-value in a sign test. (`binom.test()` in R)

Case Study: Early use of "p-values"

John Arbuthnot. Source: Wikipedia

Summary

⊙ Theory: By "divine provenance" there is a 1:1 ratio of men to women on Earth.

⊙ Assume $Pr(male) = 0.5$. In a year, there's a 50% chance that more males than females are born (and vice versa).

⊙ In 82 years of London births, more males than females are born. The probability of this is $(0.5)^{82} = 2.1 \times 10^{-25}$.

⊙ This number is so small, could not have happened by chance. Thus, "divine provenance" causes more males to be born.

🤨 Despite the imbalanced ratio at birth, it remains balanced in adulthood, because more males die than females once born (due to more deaths from hunting for food and going outside the home). The even ratio is designed by "divine provenance" and thus polygamy is unnatural. 🤔

Analyzing the Arbuthnot Data

Image by Allison Horst

Resulting p-value[1] is incredibly small. How to understand it's smallness?

[1] To arrive at the same p-value as Arbuthnot, do this test in R: `binom.test(x = 82, n = 82, p = .5, alternative = "greater")`

Analyzing the Arbuthnot Data

Image by Allison Horst

Resulting p-value[1] is incredibly small. How to understand it's smallness?

3 visual complements for 3 types of analyses:

- Simulation
- Frequentist model
- Bayesian parameter estimation

Notation:

- i: year {1629, 1630, ..., 1710}
- n_i: number of recorded births in year i
- y_i: number of male births in year i
- p_i: y_i/n_i

[1] To arrive at the same p-value as Arbuthnot, do this test in R: `binom.test(x = 82, n = 82, p = .5, alternative = "greater")`
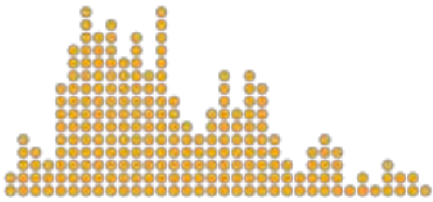
Simulation Study

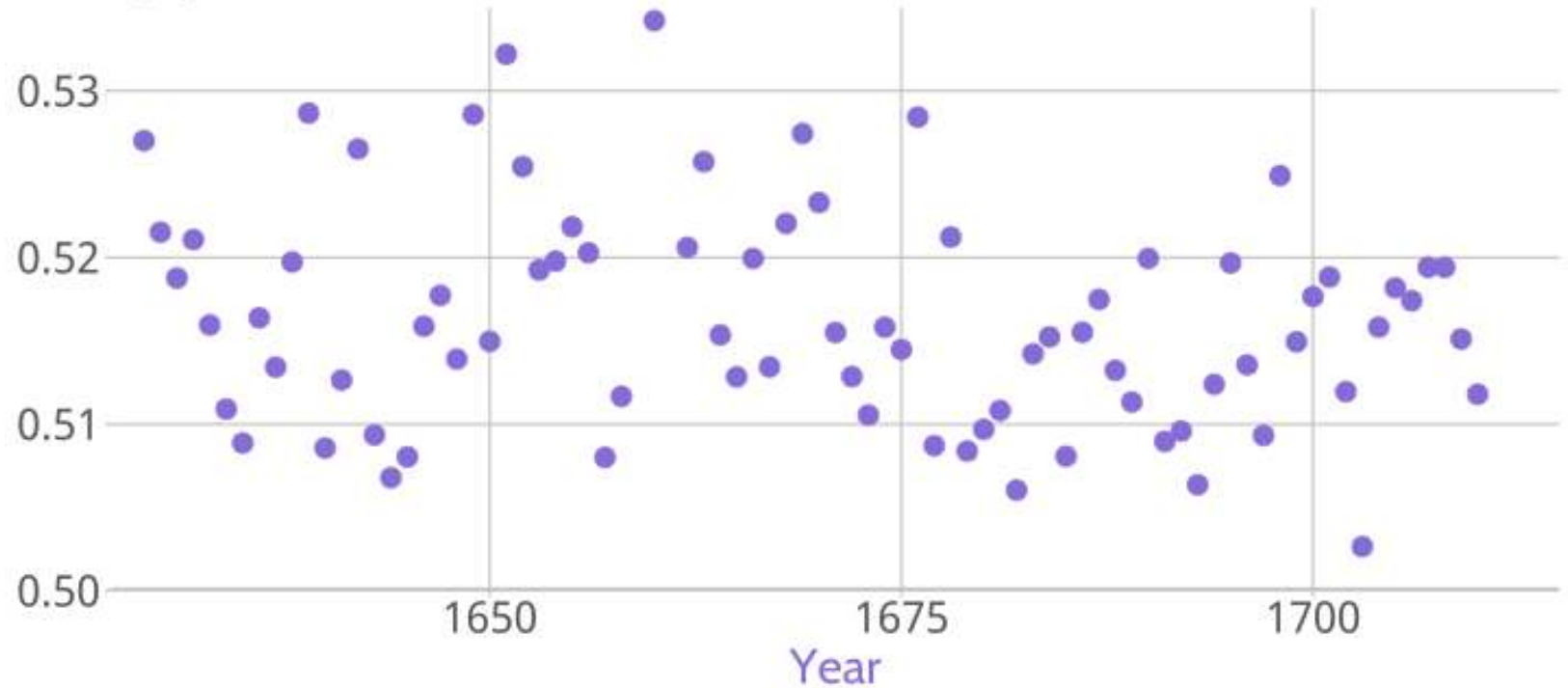Image by Allison Horst

For some large B:

1. For each year i, draw B samples: $y_{ib}^* \sim Binom(n_i, 0.5)$ for b from 1 to B.
2. For each year i, compute the proportion of male births in each simulation:
   $p_{ib}^* = y_{ib}^*/n_i$
3. For each year i, compute the proportion of simulations in which the proportion of males is greater than or equal to the observed proportion of males in Arbuthnot's data: $P_i^* = \frac{1}{B} \sum_{b=1}^B \mathbb{I}(p_{ib}^* \geq p_I)$.
4. Compute the average: $P^* = \frac{1}{82} \sum_{i=1629}^{1710} P_i^*$.
5. Assuming years are independent, compute the joint probability of observing values as or more extreme than Arbuthnot's data: $\hat{p} = \prod_{i=1629}^{1710} P_i^*$.

Results

Average probability,
P^*: 0.012

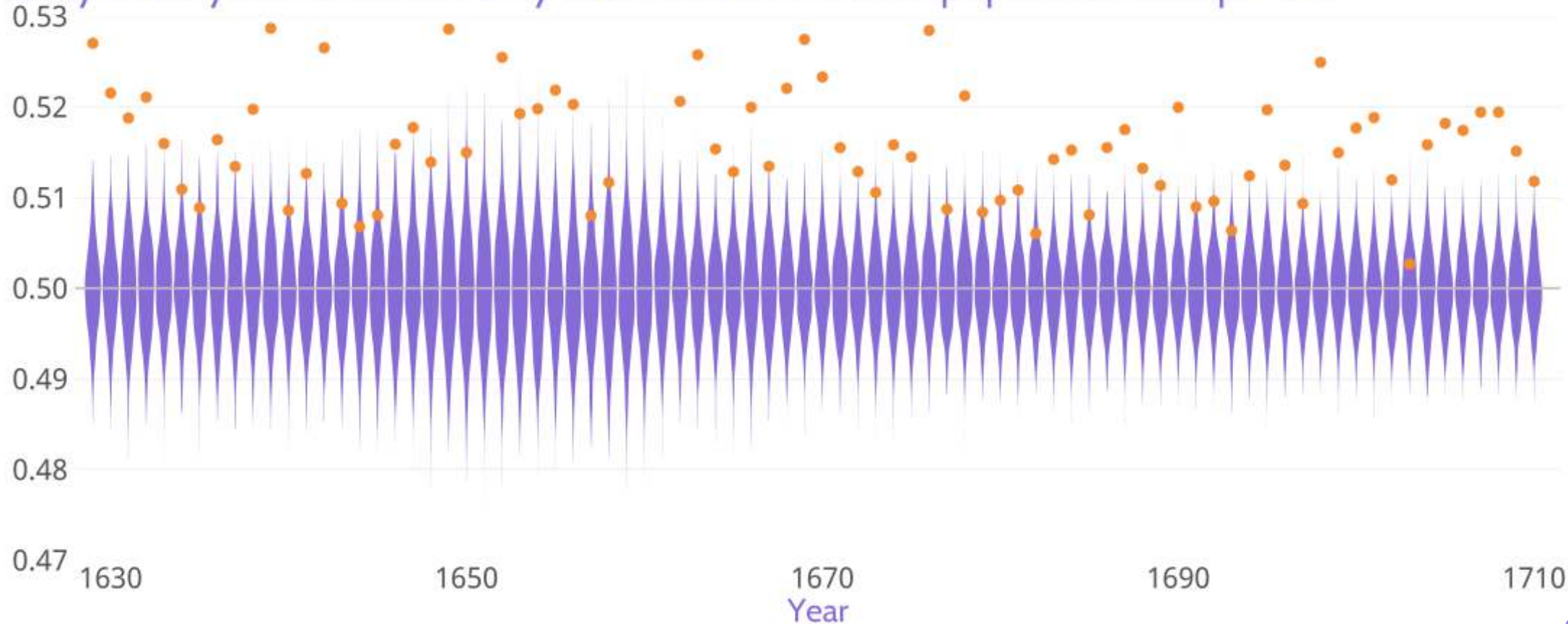Joint probability, $\hat{p}$:
$< 10^{-100}$

Every year in Arbuthnot's data had over 50% male births

Visualization

Arbuthnot's data rarely overlap with simulations when p=0.5
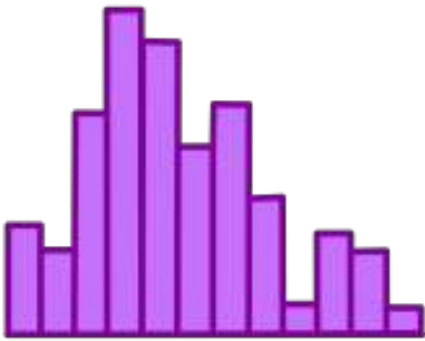Only a few years seem like they could've come from a population with p = 0.5

Frequentist model

Image by Allison Horst

Generalized Linear Mixed-Effects Model:

$$
\begin{aligned}
y_i &\sim Binom(n_i, p_i) \\
logit(p_i) &= \mu + a_i \\
a_i &\sim N(0, \sigma^2)
\end{aligned}
$$

Use `lme4::glmer()` in R to fit the model to Arbuthnot's data.

Results

Table: Parameter estimates

term	estimate	std.error	statistic	p.value
mu	0.0657	0.003	21.9947	0
sigma	0.0190	NA	NA	NA

Table: 95% Confidence Intervals

	2.5 %	97.5 %
sigma	0.0132	0.0255
mu	0.0598	0.0716

Fitted Model:

$$
\begin{aligned}
y_i &\sim Binom(n_i, p_i) \\
logit(p_i) &= 0.06565 + a_i \\
a_i &\sim N(0, 3.6149 \times 10^{-4})
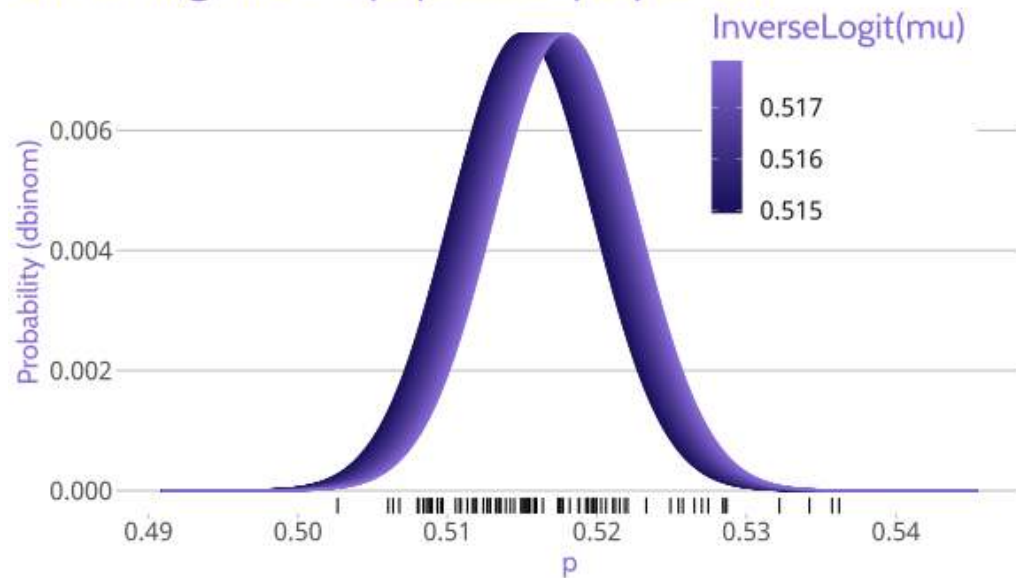\end{aligned}
$$

Estimate of population proportion:

$$logit^{-1}(\hat{\mu}) = logit^{-1}(0.06565) = 0.51641$$

Table: Fit metrics

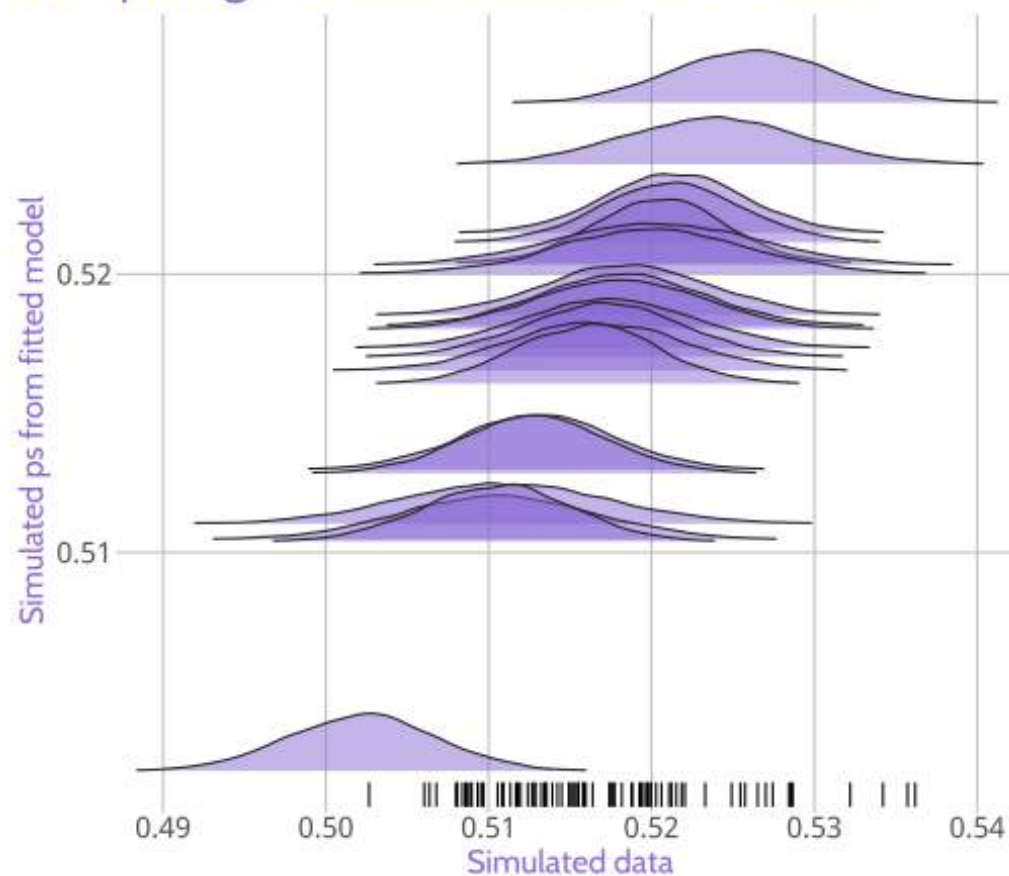| AIC	BIC	logLik	deviance	df.resid
946.31	951.12	-471.15	942.31	80

Visualization

Bayesian approach

Image by Allison Horst

Bayesian Hierarchical Model:

$$
\begin{aligned}
y_i &\sim Binom(n_i, p_i) \\
p_i &\sim Beta(\alpha, \beta) \\
\mu = \frac{\alpha}{\alpha+\beta} &\qquad \eta = \alpha + \beta \\
\mu &\sim Beta(1, 1) \\
\log(\eta) &\sim Logistic(\log(\bar{n}), 1)
\end{aligned}
$$

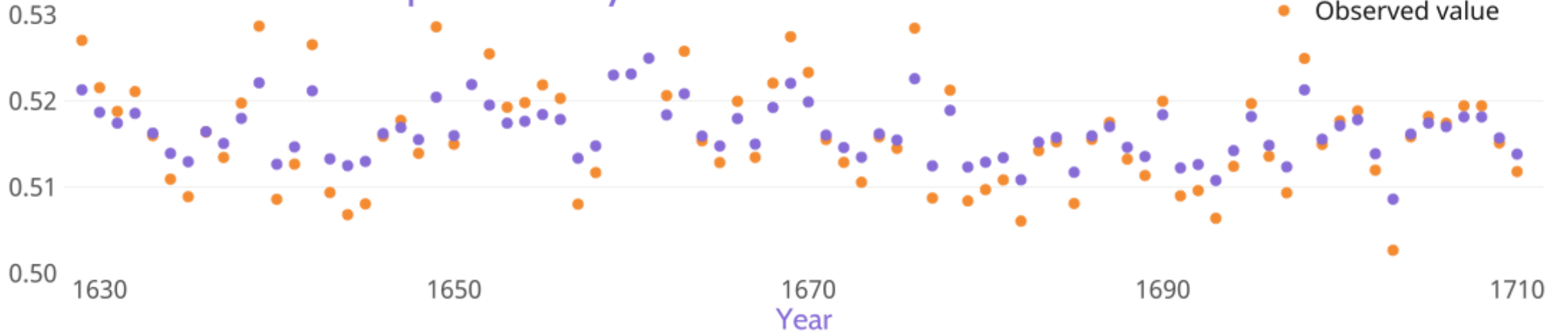Fit to Arbuthnot's data with the `runjags` package in R.

Results

Table: Summary of posterior draws of α, β
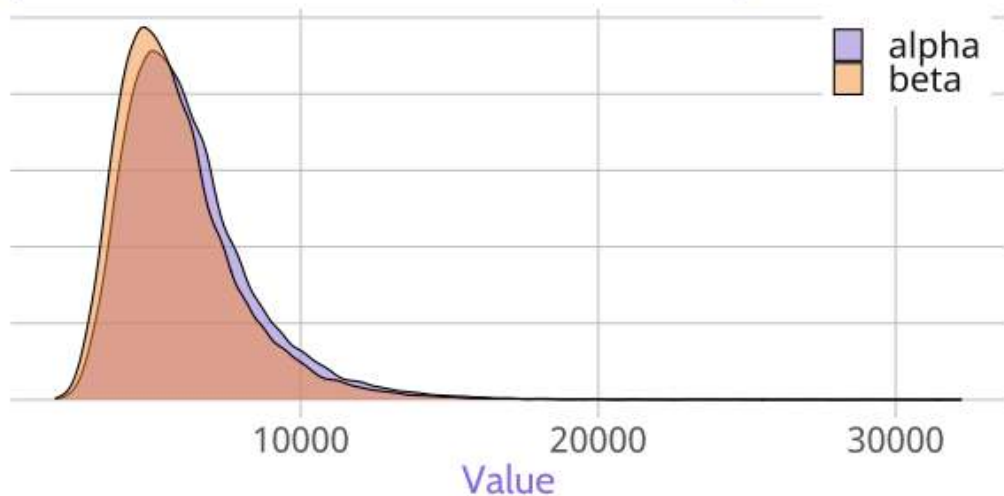
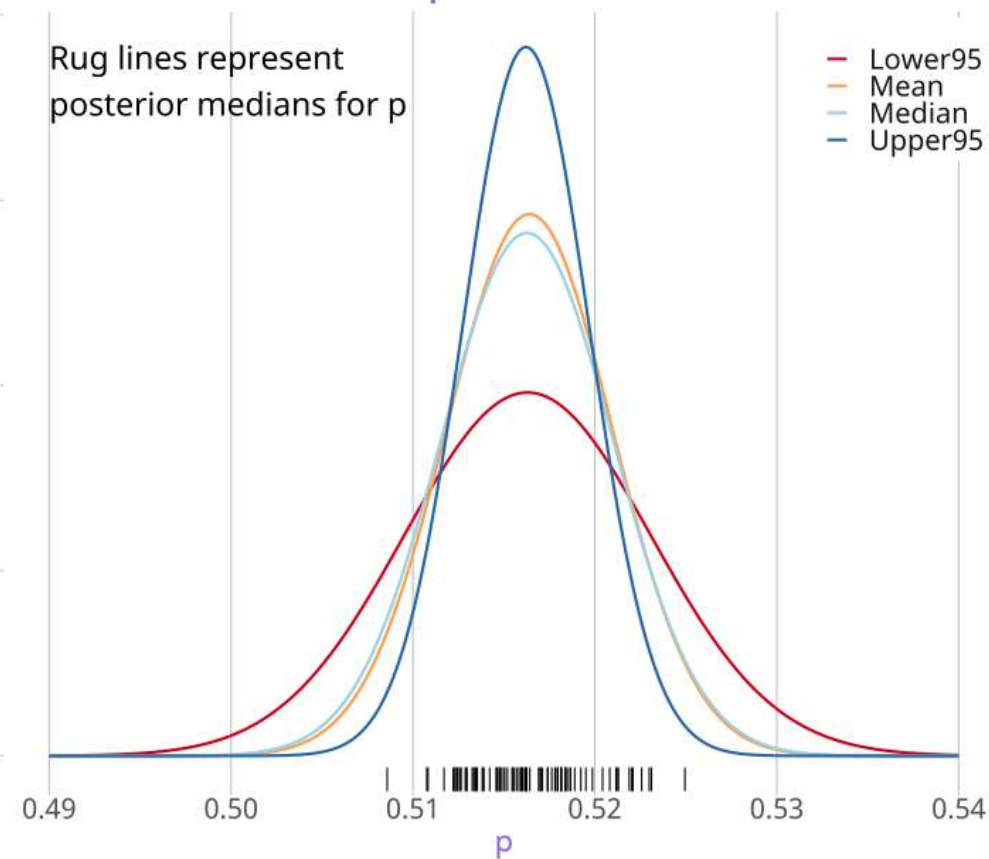| param	Lower95	Median	Upper95	Mean	SD
alpha	2805.40	5808.95	10673.32	6233.47	2246.01
beta	2628.34	5442.88	10003.01	5837.60	2104.54

Posterior Medians of p for each year

Visualization

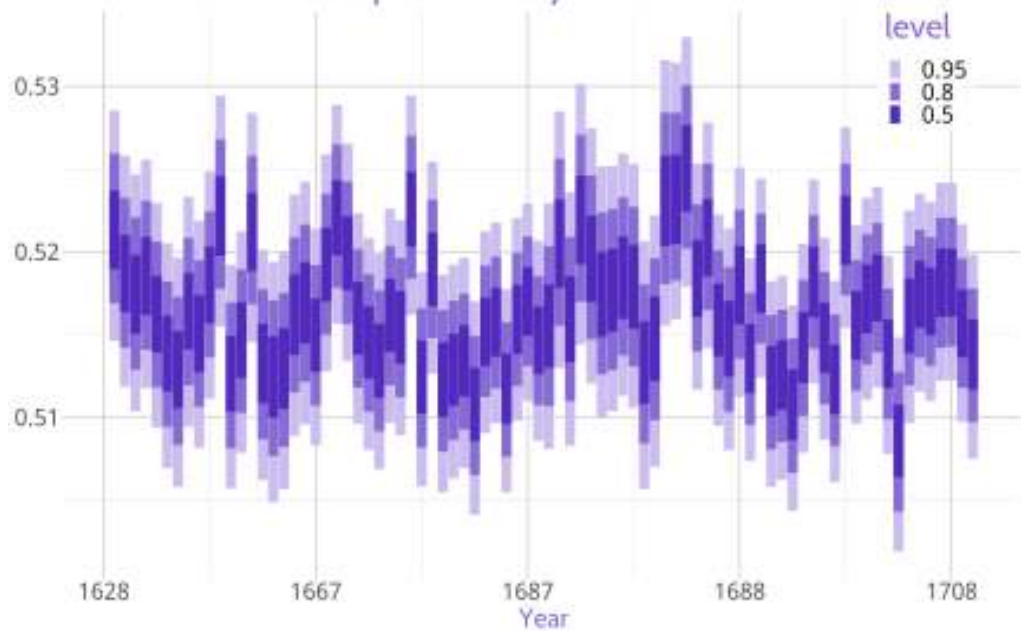Posterior distributions nearly overlap
(Mode of beta distribution near 0.5)

alpha
beta

Beta distributions for p

Rug lines represent posterior medians for p

— Lower95
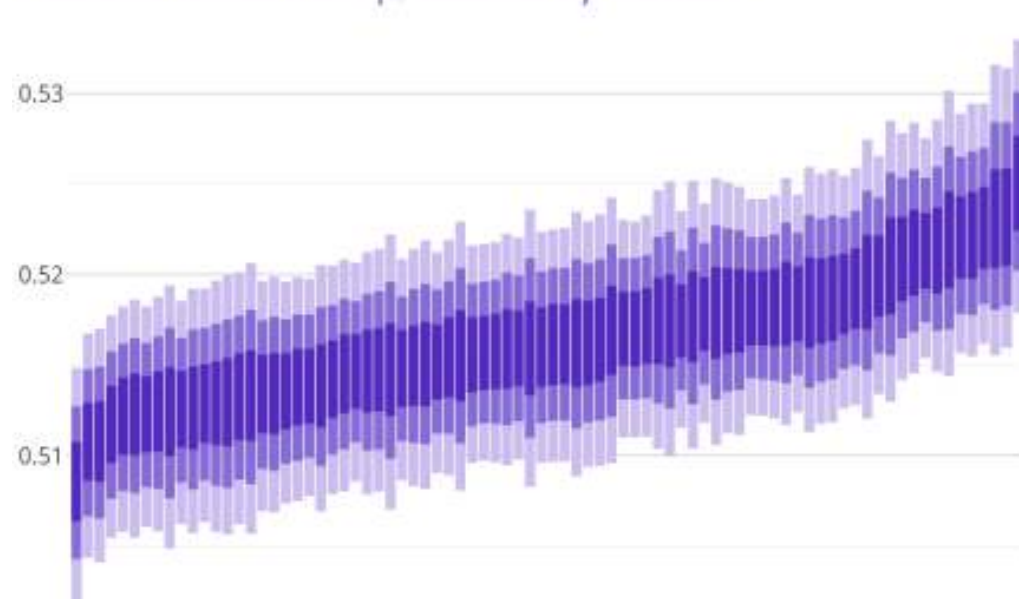— Mean
— Median
— Upper95

Visualization

Posterior draws from p for each year

Posterior draws from p, ordered by median

Takeaways

Image by Allison Horst

Dig deep into your model results

- Think about how you can you show probabilities
  - What other ways are there to model your data?
  - What other parameter values could result in your data?
- Simulation is your friend
- Take advantage of existing tools (in R or elsewhere)
  - `ggplot2` extensions: `ggridges`, `ggdist`
- Further reading: *Model Visualization*, *Statistics on Street Corners* [2; 3]

References

[1] J. Arbuthnot. "An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes". In: *Philosophical Transactions of the Royal Society of London* 27.4 (1710), pp. 186-190. DOI: https://doi.org/10.1098/rstl.1710.0011.

[2] H. Wickham, D. Cook, and H. Hofmann. "Visualizing statistical models: Removing the blindfold". In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 8.4 (Jul. 2015), pp. 203-225. DOI: https://doi.org/10.1002/sam.11271.

[3] A. Loy, H. Hofmann, and D. Cook. "Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners". In: *Journal of Computational and Graphical Statistics* 26.3 (2017), pp. 478-492. DOI: 10.1080/10618600.2017.1330207.

R Packages Used

tidyverse (ggplot2, dplyr, tidyr), lme4, runjags, tidybayes, ggdist, ggridges

Thank you!

Contact:

✉ sctyner90@gmail.com
🐦 sctyner
🐙 sctyner
🌐 sctyner.me

Slides available at bit.ly/viz-fcsm