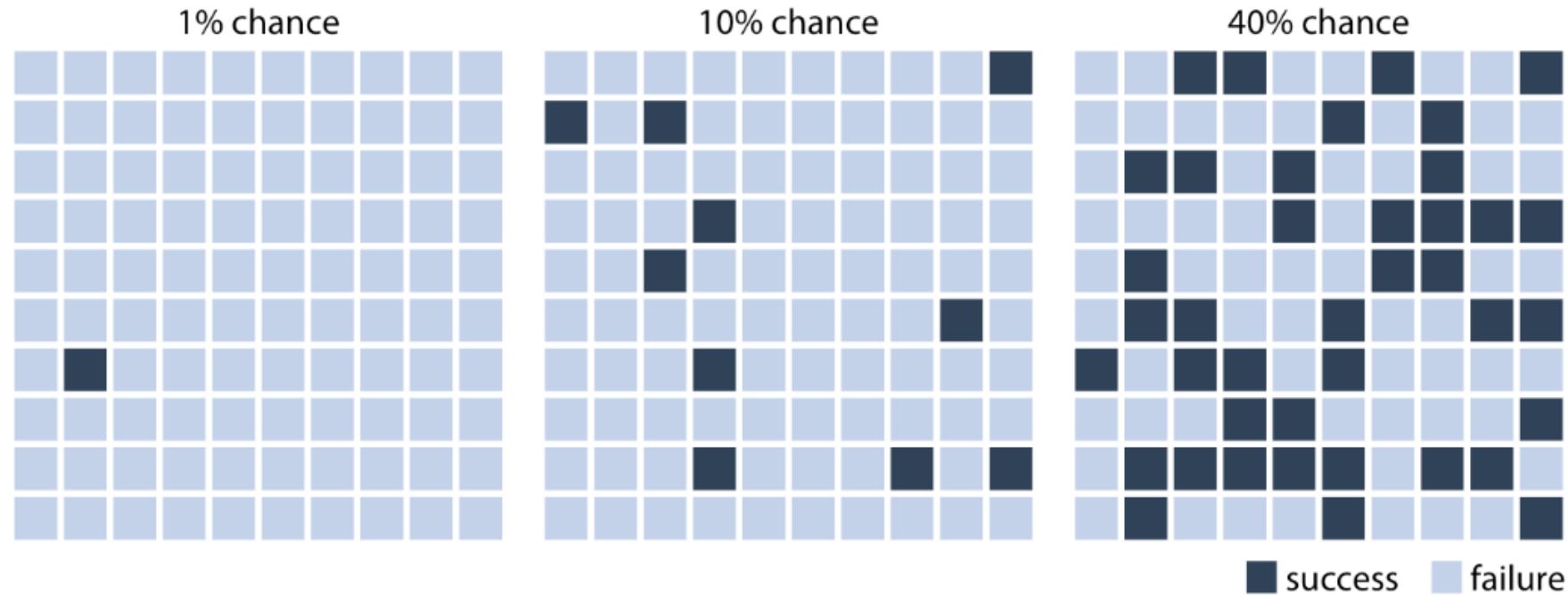# Visualizing Uncertainty

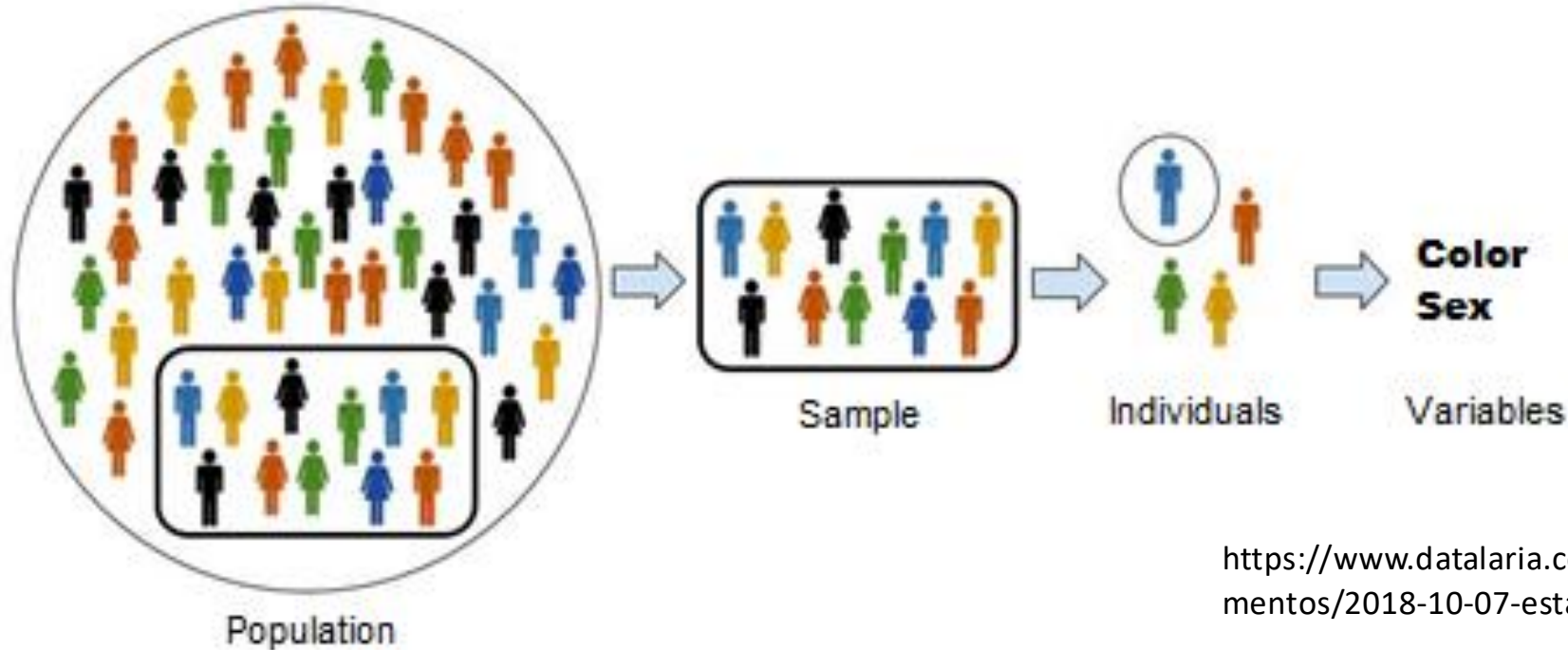HUGEN 2073

Genomic Data Visualization and Integration

Slides borrowed/adapted from H. J. Park with permission

Visualizing uncertainties is visualizing probability

- Visualizing uncertainty to a binary outcome, survival or mortality
- e. g., how certain is a breast cancer patient to survive?

- OK since 2 factors to define a probability
1. Sample size
2. Individuals with event (case or control)

- Not OK since it doesn't show stochasticity?

1% chance   10% chance   40% chance

■ success   ☐ failure

Population, samples, individuals and variables

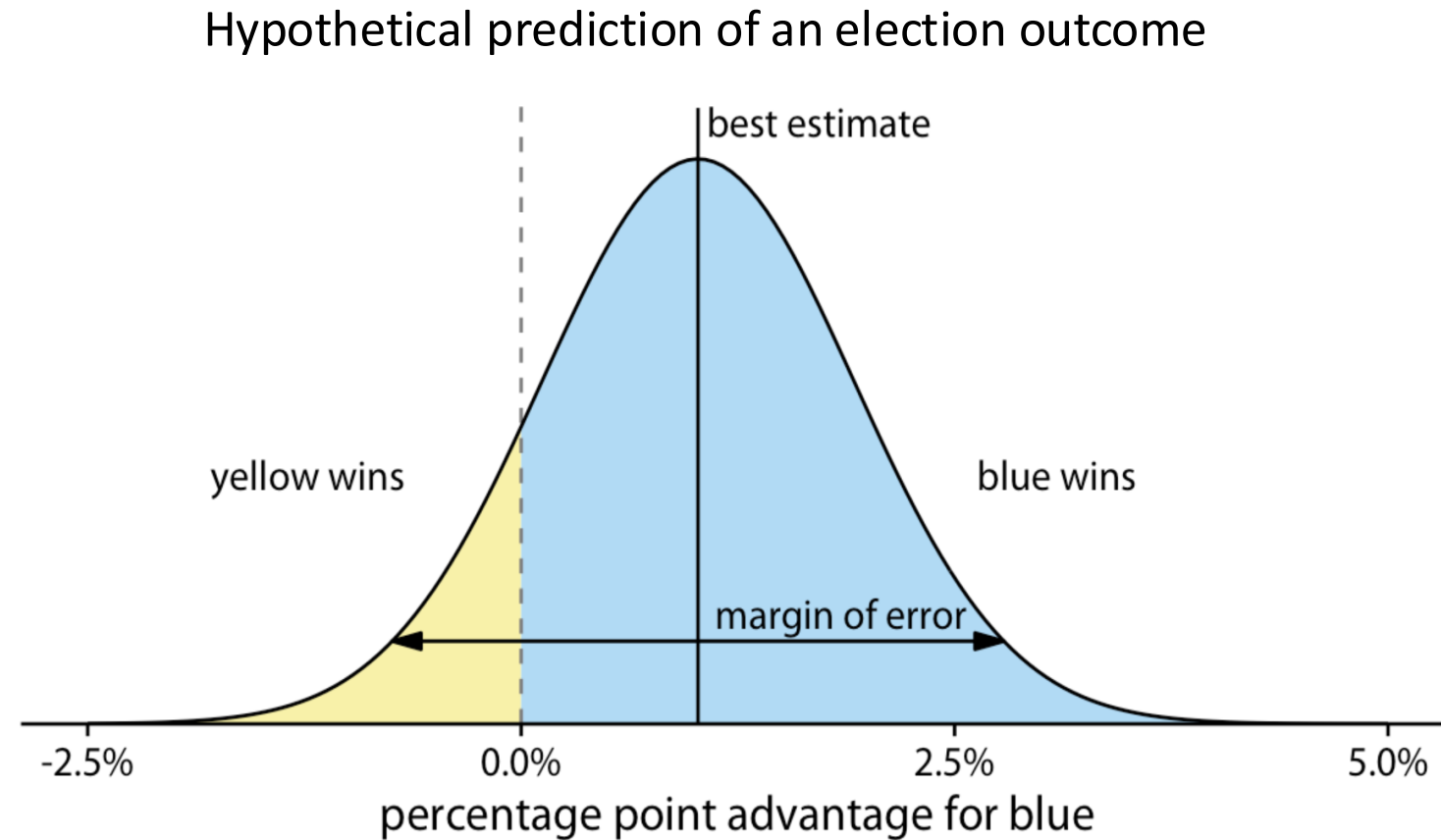
https://www.datalaria.com/en/post/funda mentos/2018-10-07-estadistica-descriptiva/

- **Population:** The set of all the elements under study (e. g., all breast cancer patients)
- **Sample:** A subset of elements of the population (assumed to be representative of the population) (e. g., a breast cancer cohort)
- **Individuals:** Individual element in the population (e. g., a breast cancer patient)
- **Variables:** Characteristics of the individuals (clinical, molecular, socioeconomic features)
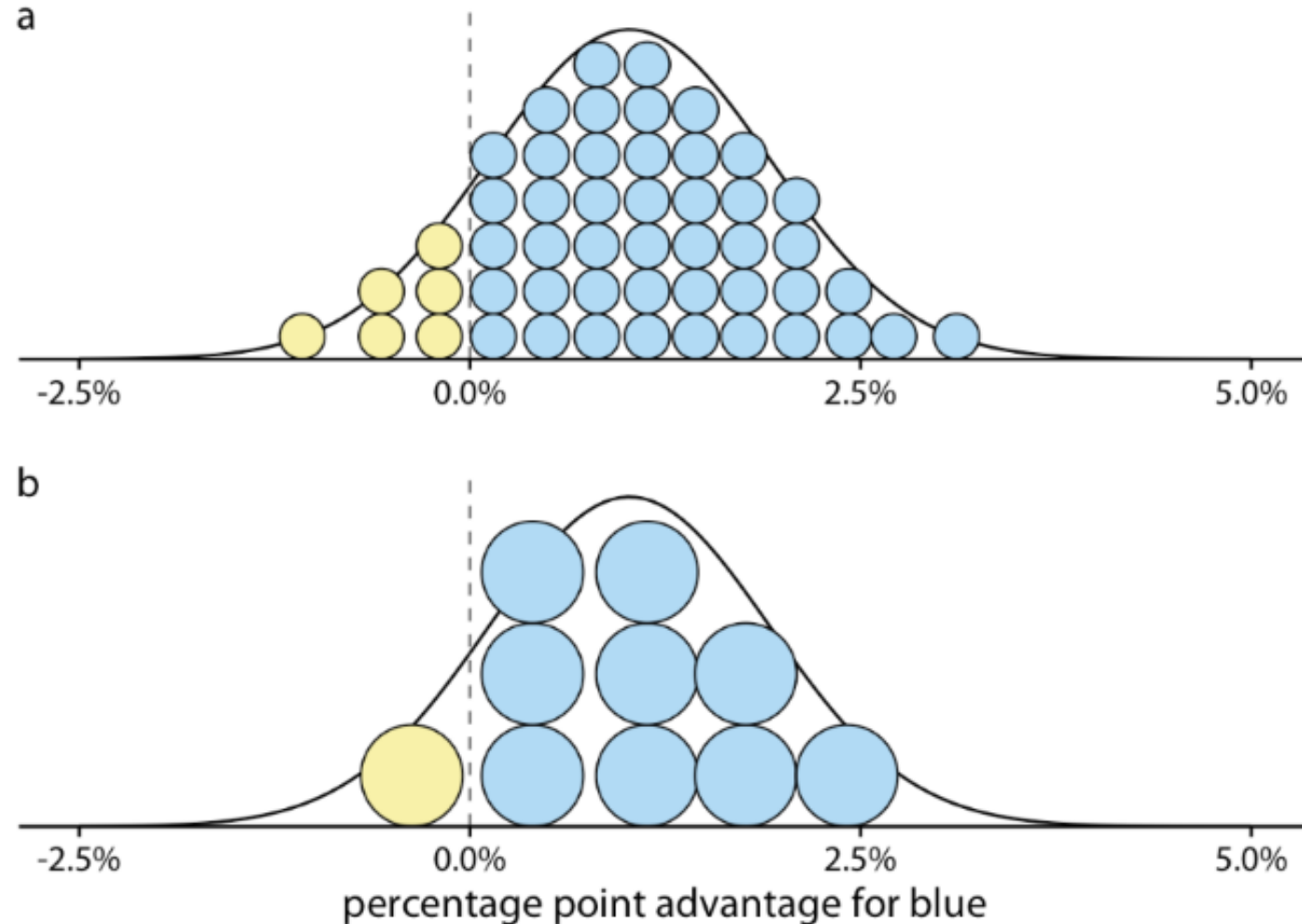
Outcome prediction as probability

- E. g., an upcoming election with two parties, the yellow party and the blue party
- On radio, the blue party is predicted to have a 1% advantage with a margin of error of 1.76 % points
- The range of possible outcomes with their associated likelihoods in curve

Hypothetical prediction of an election outcome

best estimate

yellow wins

blue wins

margin of error

-2.5% 0.0% 2.5% 5.0%

percentage point advantage for blue

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Quantile dot plots to represent probability

- Visualizing that the chance of the yellow party winning is 12.9%
- Too many dots: have it perceived as a continuum
- Too few dots: make the estimation inaccurate
- How many dots would work?
  - Manual inspection
  - Who's your audience?
- Similar for bin width, bandwidth, window size in the sliding window
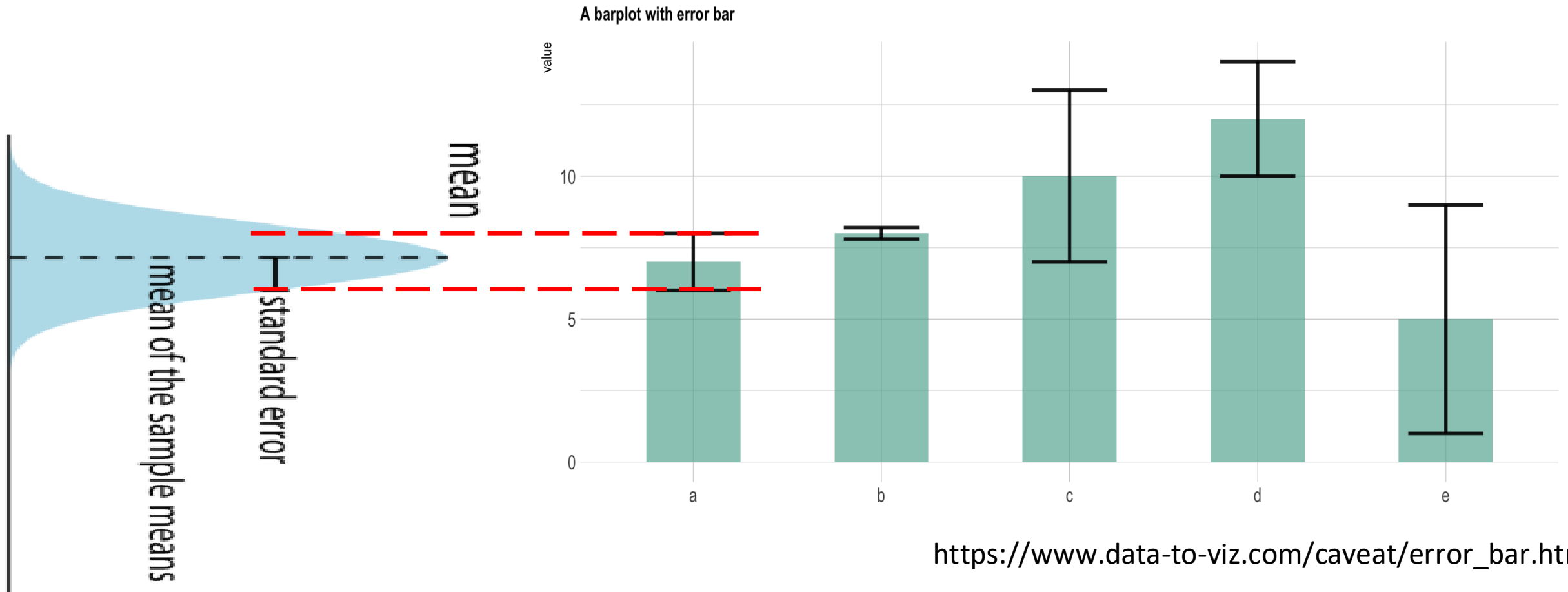
Hypothetical prediction of an election outcome

percentage point advantage for blue

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Standard deviation and standard error of the mean to measure uncertainty

Term	Symbol	Meaning
Standard deviation (of population)	σ	• Variability in population • Typically unknown (so estimate it)
Sample standard deviation	• $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ • $\hat{\sigma}$	• Variability of a particular sample • Also an estimator of σ
• Standard error (of sample mean) • SE or SEM	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	• Variability of sample mean (think of $\bar{x}$ as a random variable) • Think of central limit theorem • Helps measure likely distance between sample mean $\bar{x}$ and true mean μ • Typically unknown (since σ is unknown), so estimate it
Estimator of standard error	$\widehat{\sigma_{\bar{x}}} = \frac{\hat{\sigma}}{\sqrt{n}}$	• Confusingly also called standard error/SE/SEM • Ingredient for error bars • Ingredient for confidence intervals

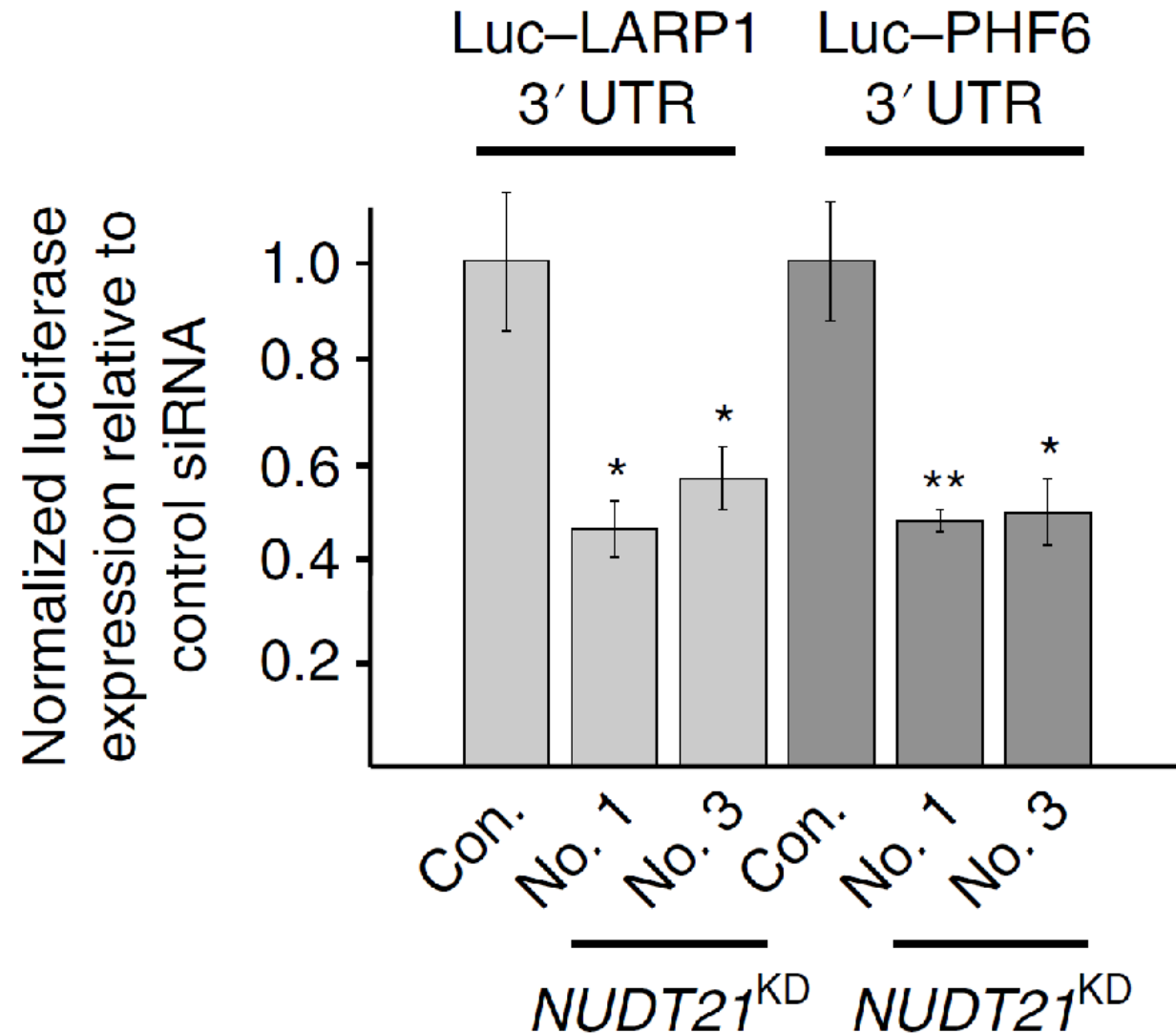Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Error bars visualize uncertainty

- Five samplings from five populations
- Showing the sample distributions with confidence
- Error bars usually are +/- SE (not always – could be a confidence interval)
- Always define error bar!

A barplot with error bar
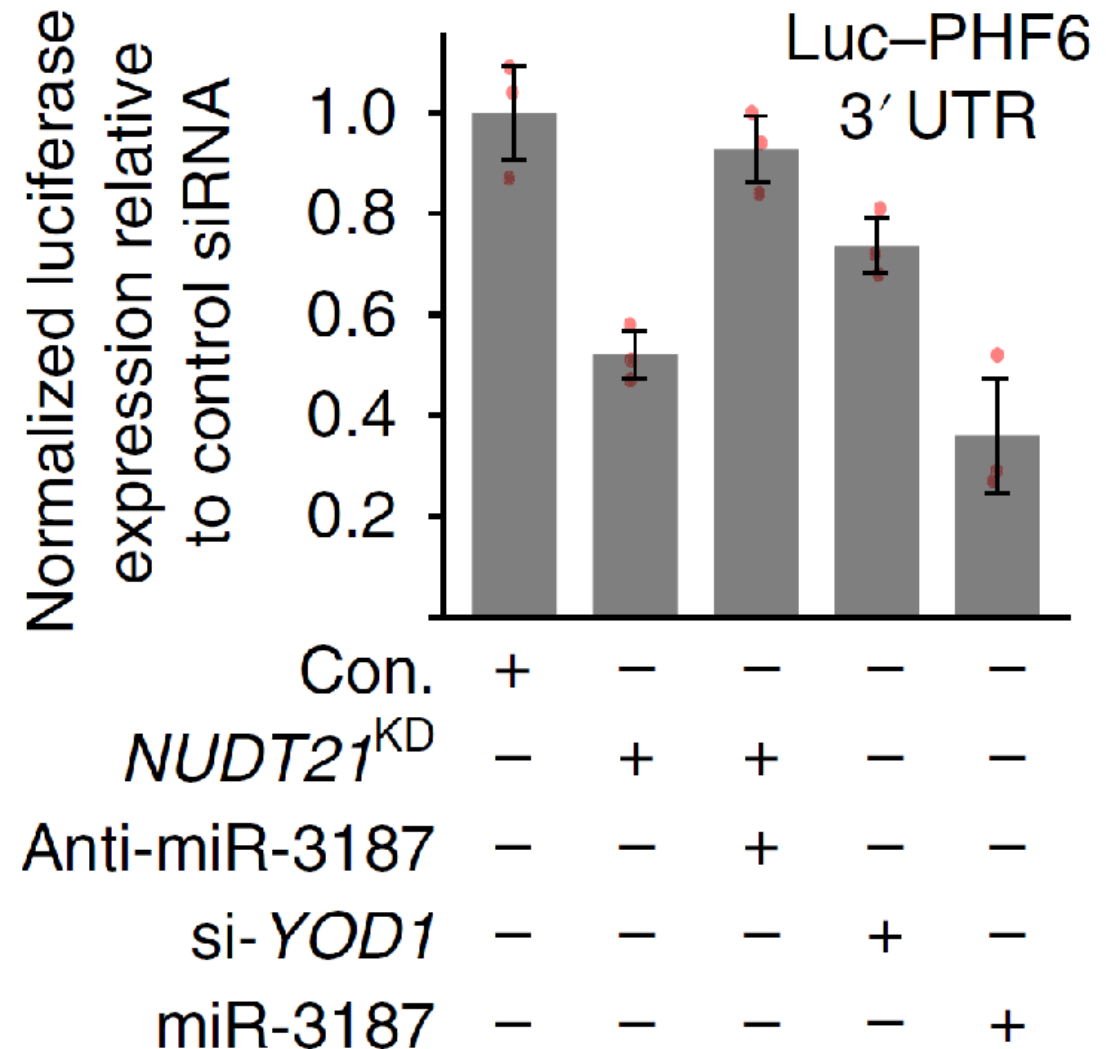
https://www.data-to-viz.com/caveat/error_bar.html

Show significance whenever possible

- Different numbers of stars to represent significance of difference (usually from control)
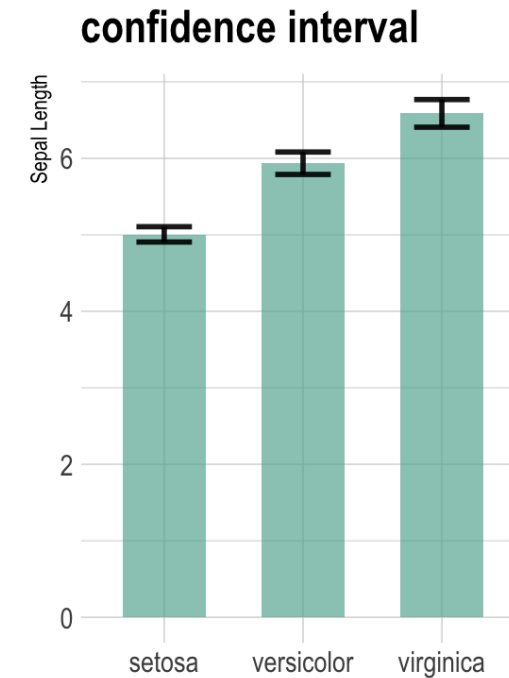
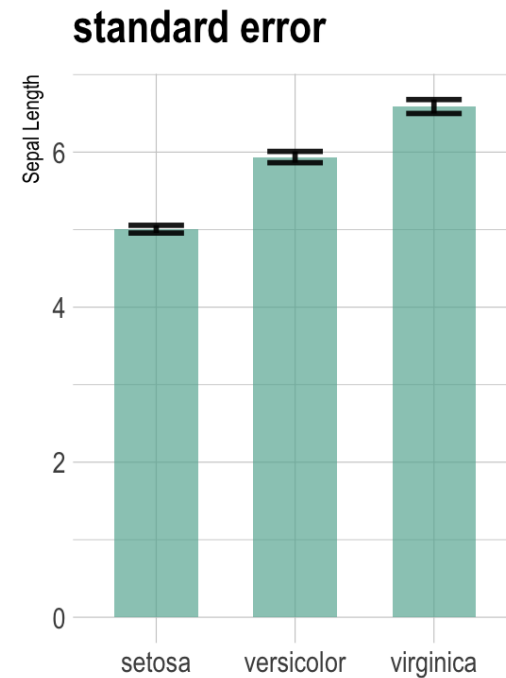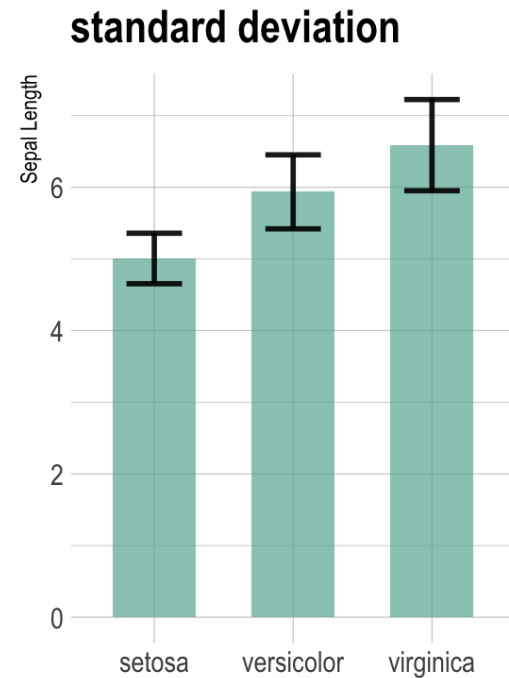# Show real data points whenever possible

- Error bars to represent uncertainty
- Stars to represent significance of difference (usually from control)
- Dots to represent real values
- X-axis label can be a table

Con.	+	−	−	−	−
$NUDT21^{KD}$	−	+	+	−	−
Anti-miR-3187	−	−	+	−	−
si-$YOD1$	−	−	−	+	−
miR-3187	−	−	−	−	+

Luc–PHF6 3′ UTR

Normalized luciferase expression relative to control siRNA

Park et al. Nature Genetics, 2018

Error bars can indicate different types of variations
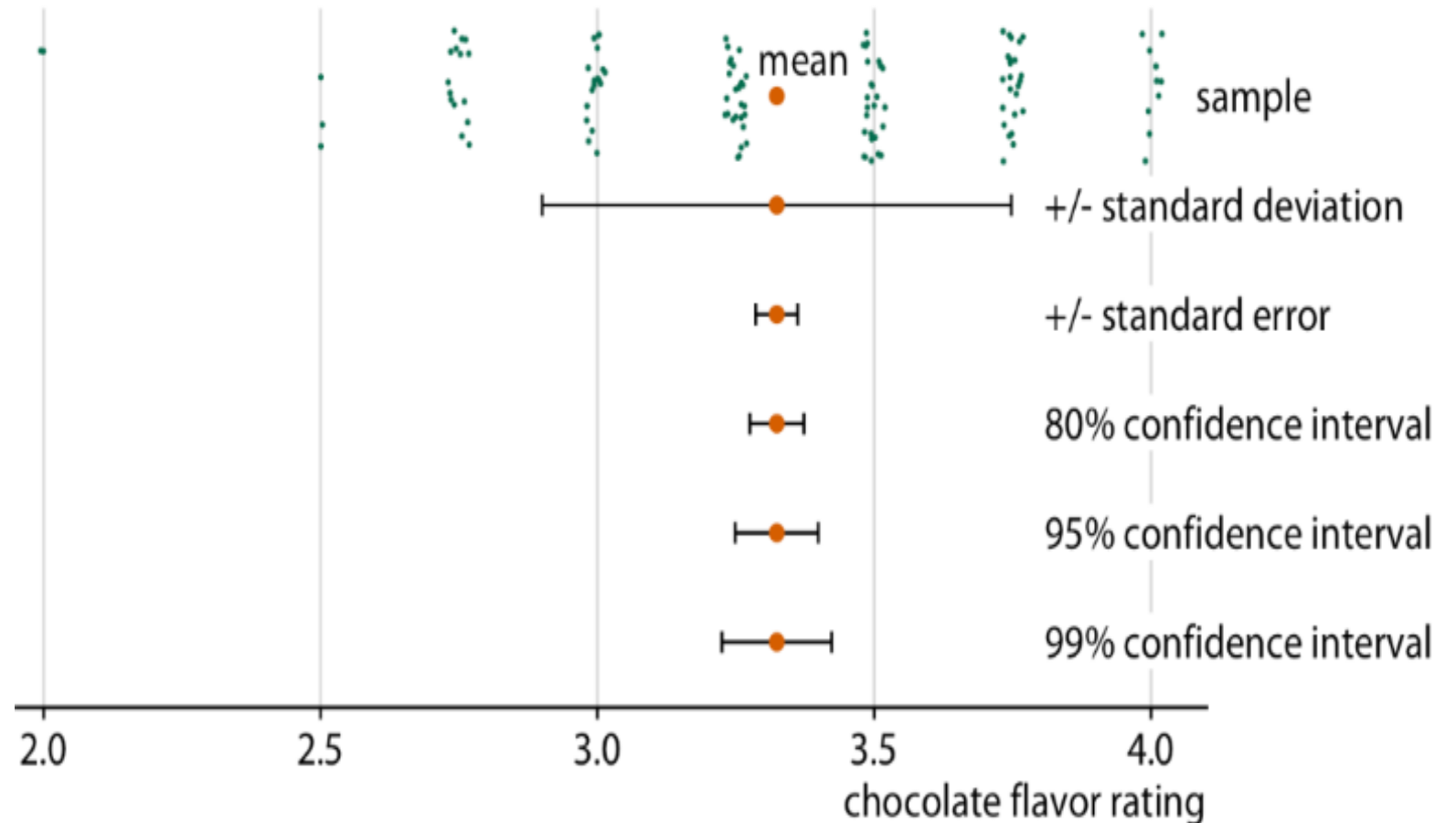
State clearly if yours is
- Standard Deviation (SD): variability within a sample,
- Standard Error (SE): variability across samples of a population, or
- A confidence interval (CI): a range that a value is usually expected to lie in
- Credible interval (Bayesian-flavored confidence interval)

https://www.data-to-viz.com/caveat/error_bar.html

Visualizing uncertainty with error bars

Expert ratings of chocolate bars

- Specify what quantity and/or confidence level the error bars represent

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Various visualization styles for error bars

Mean chocolate flavor rating by country (vs. US chocolate rating)

- More information through multiple boxes

E. g., 1 ≈ 2>…>3 ≈ 4

Various visualization styles for error bars

Mean chocolate flavor rating by country (vs. US chocolate rating)
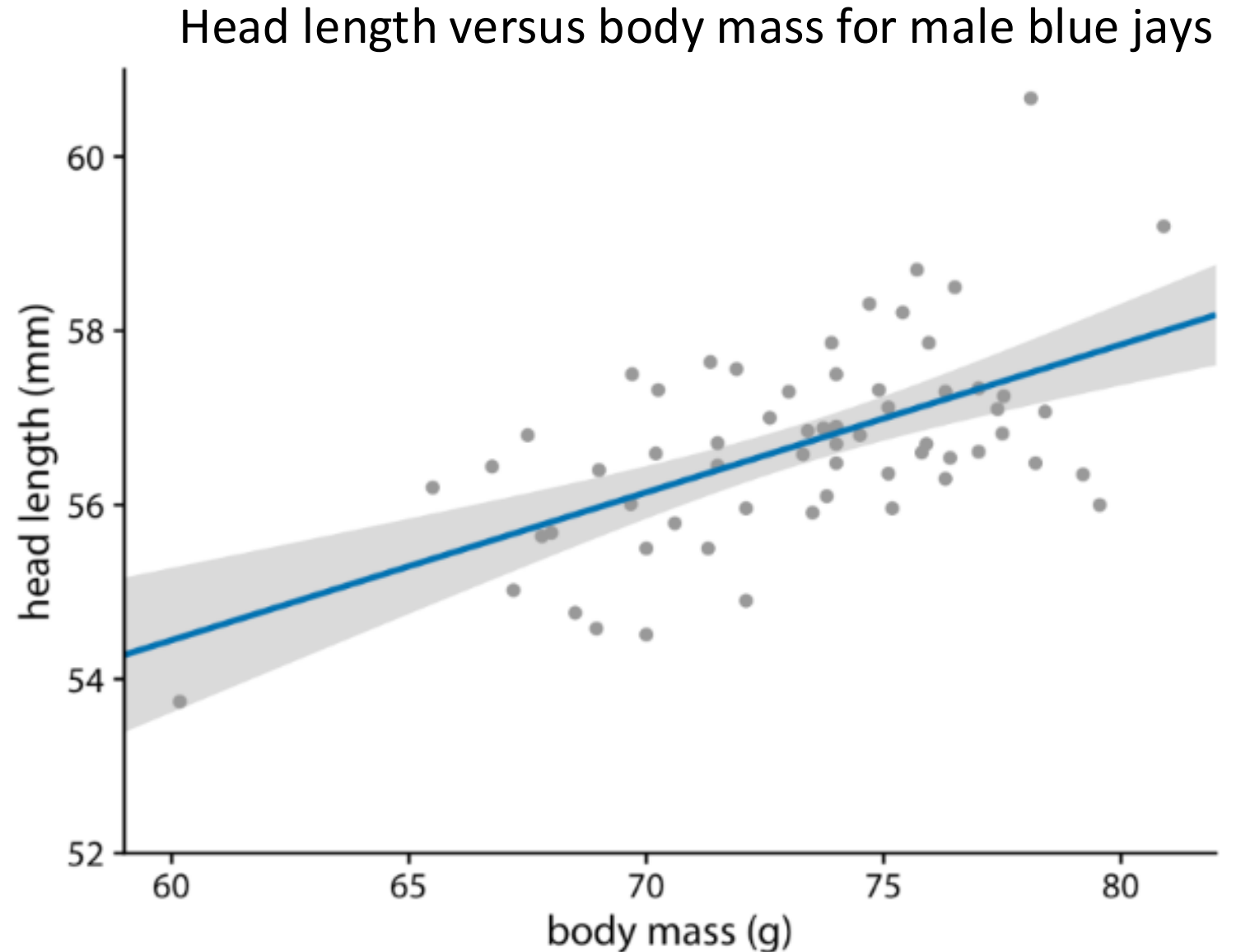
- Often, styles obstruct scientific communication
  - Confidence strip, e.g. Is the mean rating of Peruvian chocolate significantly lower than that of the US?
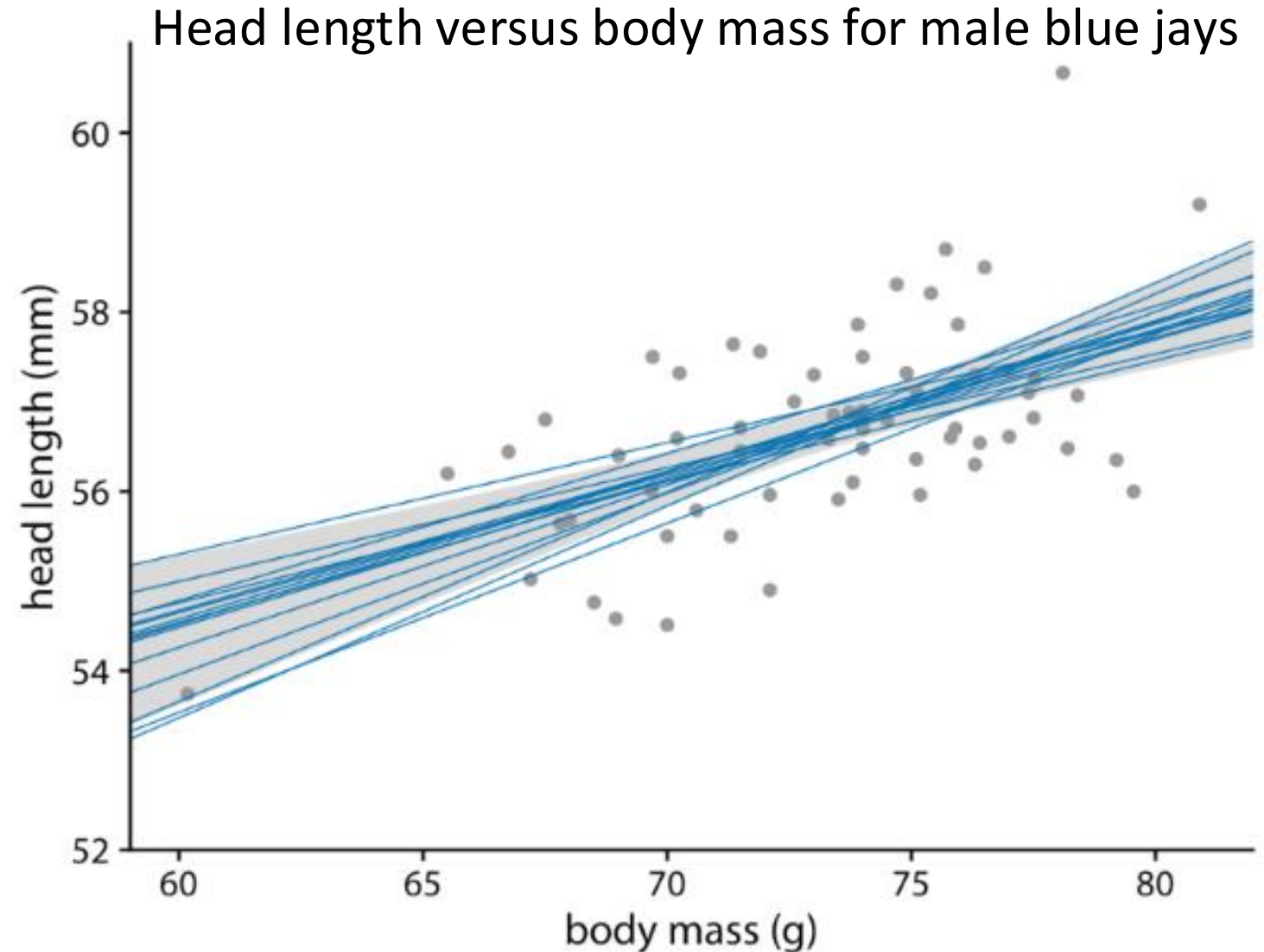 - Ridgeline style

Confidence band to show uncertainty of a linear model

- Trend in a real line and the uncertainty in a trend line with a confidence band
- Need to be very clear about what quantity the uncertainty is in

Head length versus body mass for male blue jays

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Confidence band to show uncertainty of a linear model
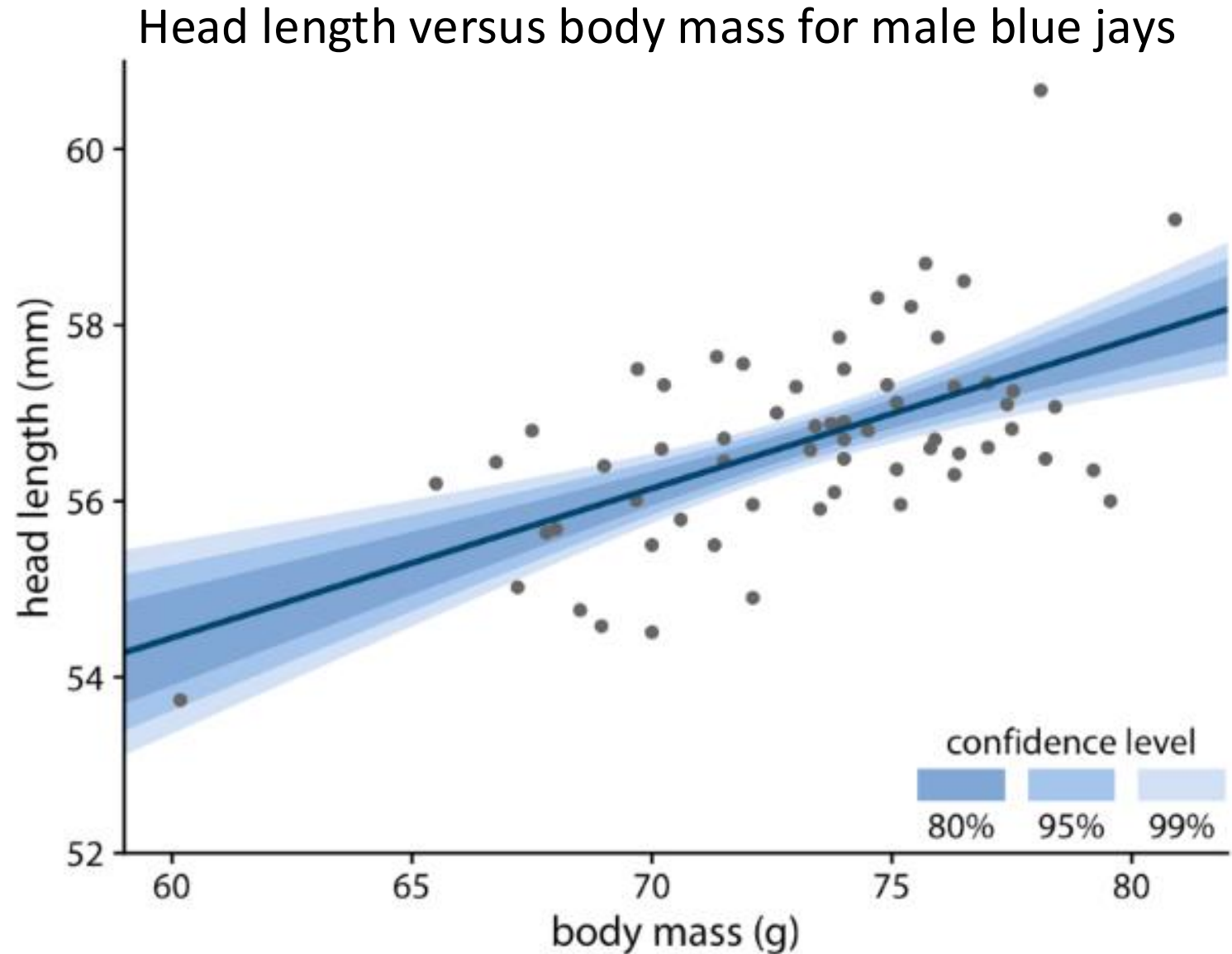
- Different intercepts and slopes of line fits create the curvature
- 15 randomly chosen alternative fits from the confidence band

Head length versus body mass for male blue jays

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

Confidence band to show uncertainty of a linear model
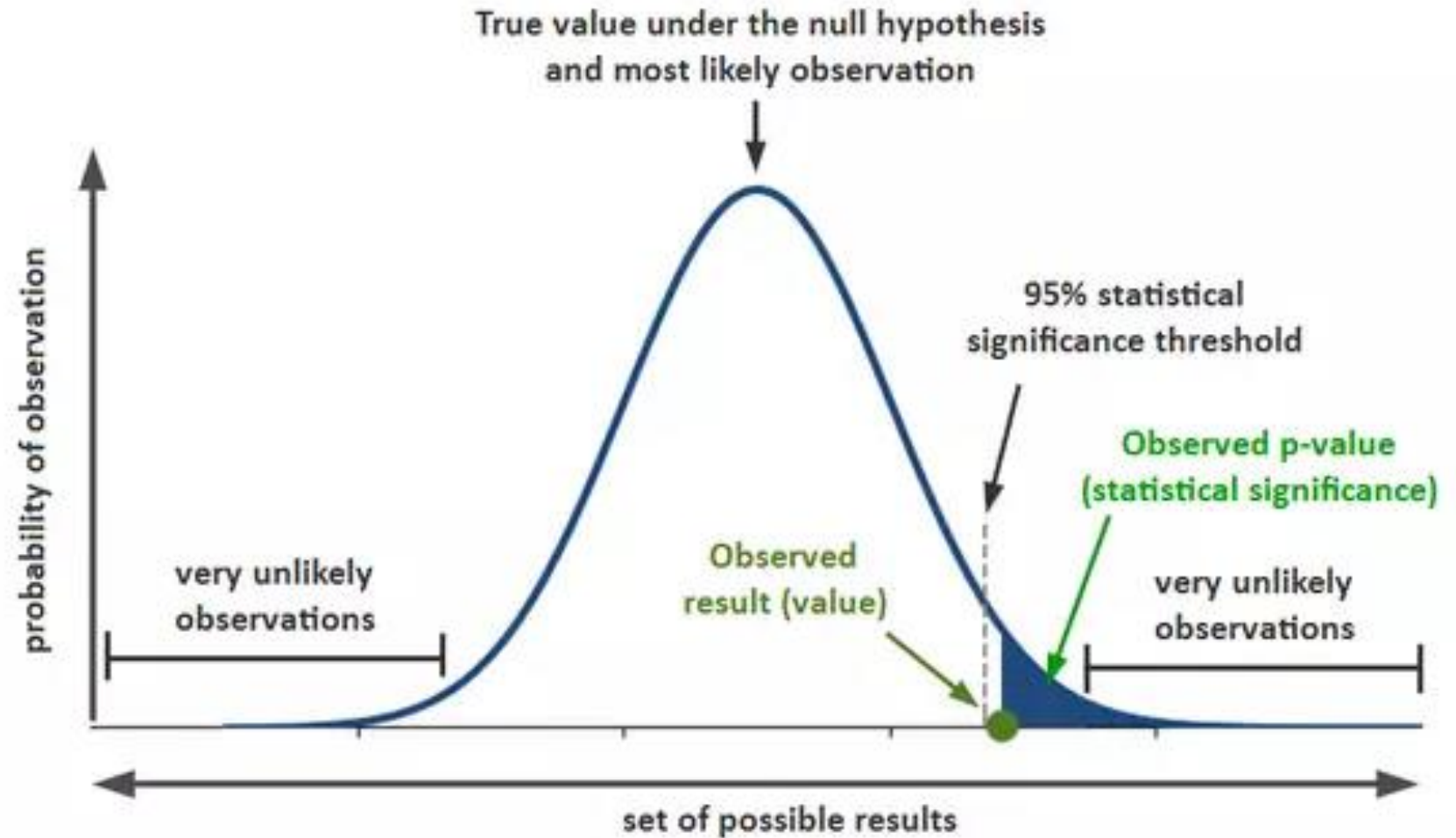
A graded confidence band
- Highlights different levels of confidence
- Not only to show the distribution across the confidence levels
- Enhances the sense of uncertainty in the reader, and
- Forces the reader to confront the possibility that the data might support different alternative trend lines

Head length versus body mass for male blue jays

Fundamentals of Data Visualization, Wilke, O'Reilly, 1st Ed.

P-value to represent (un)certainty of a hypothesis

P-value
- The probability of obtaining the observation (or something more extreme) under the null hypothesis or null distribution
- The lower, the more certain to reject null hypothesis and to accept alternative hypothesis
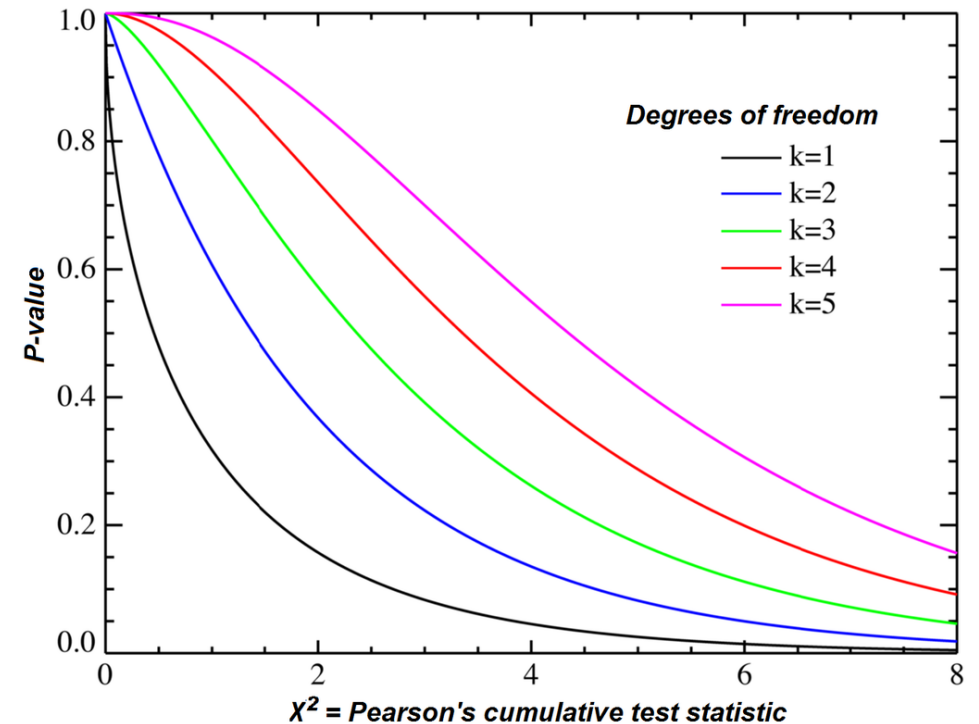

https://www.simplypsychology.org/p-value.html

P-value to represent (un)certainty of a hypothesis

Now consider testing many similar hypotheses
- Null hypothesis: the number of tumor suppressors and oncogenes in each group, adding up to 40, is balanced
- Can use chi-square with df=1
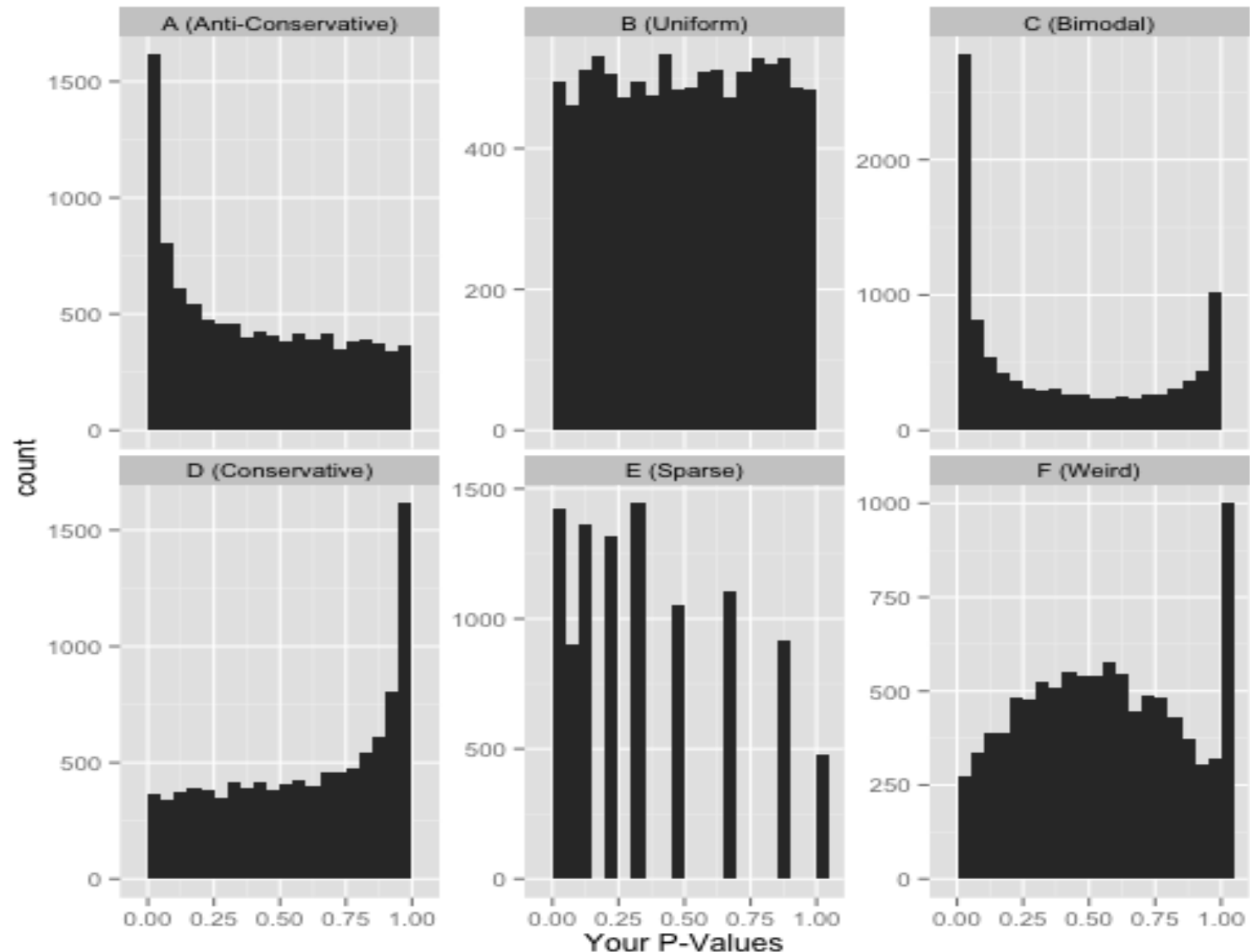- Imagine we test this across many gene groups (several thousand)

	# tumor suppressors	# oncogenes	P-value
Expected number	20	20	NA
Gene group 1	9	31	..
Gene group 2	21	19	..
…	…	…	…

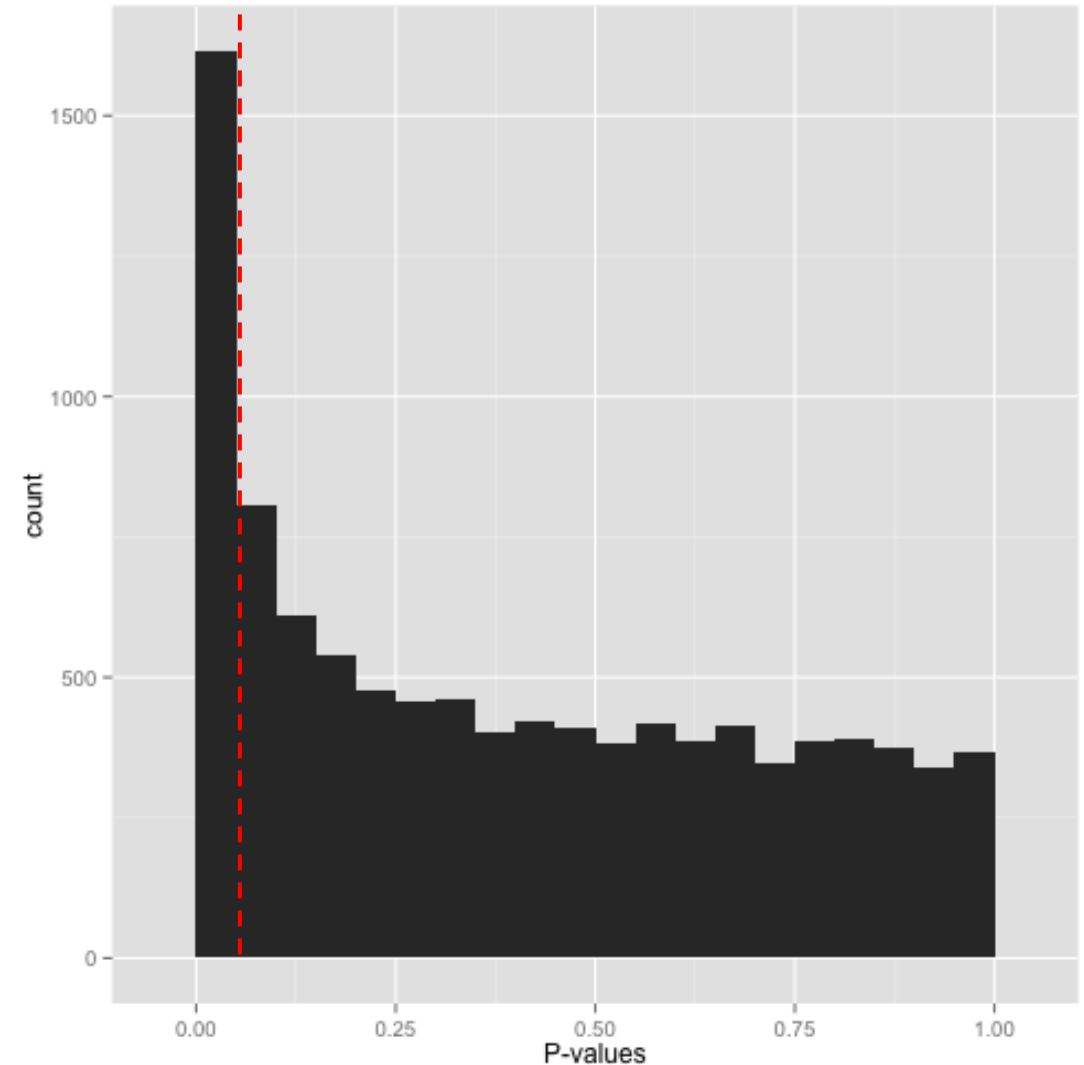https://en.wikipedia.org/wiki/Chi-squared_distribution

How to interpret P-value distribution?

- If we have many gene groups to test, we will have many p values, one for each gene group
- The distribution of the p values provides insights to your experimental design

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Anti-conservative p-values ("Hooray!")

- Example:
  - ~1,700 hypotheses (tests or gene groups) with p-value < 0.05
  - ~780 hypotheses with p-value b/w 0.05 and 0.1
- One may think that all ~1,700 hypotheses under p-value < 0.05 are truly unbalanced gene groups (true positives), but not all of them are.
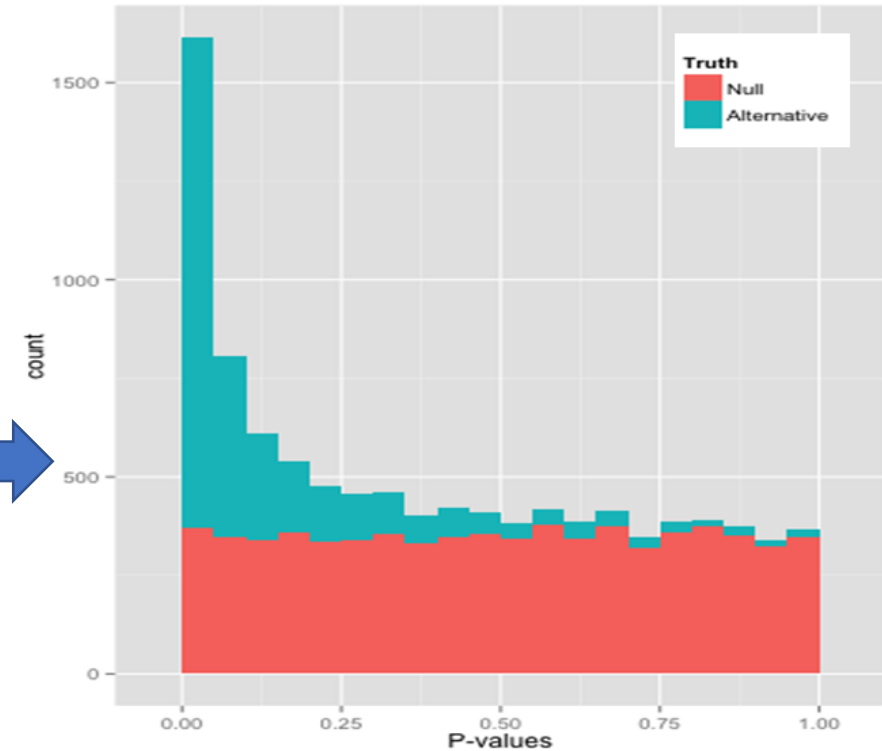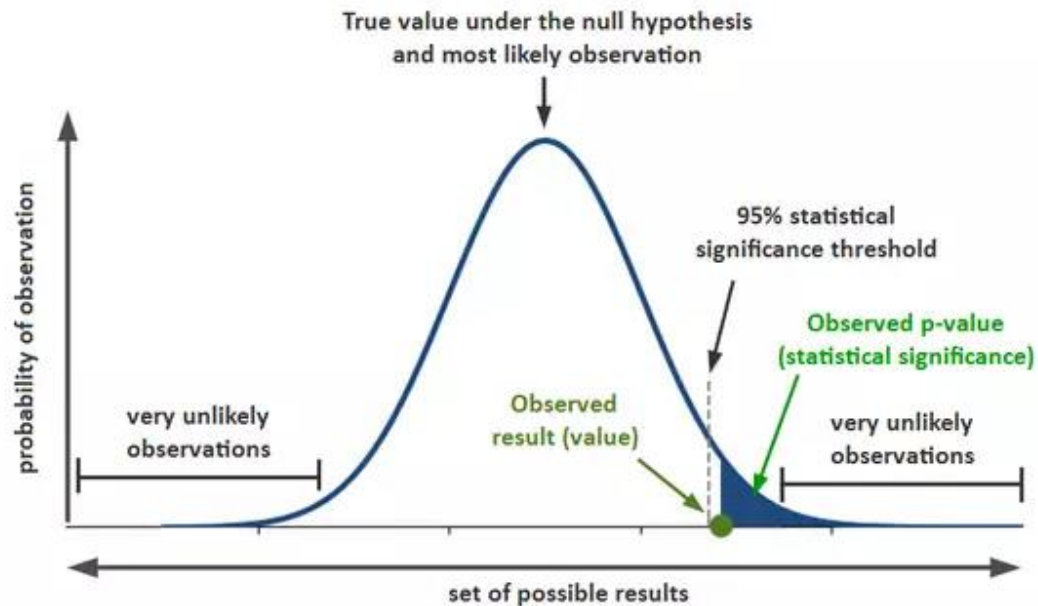

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Anti-conservative p-values ("Hooray!")

1,700 gene groups with p-value < 0.05 consist of
- unbalanced gene groups (true positives)
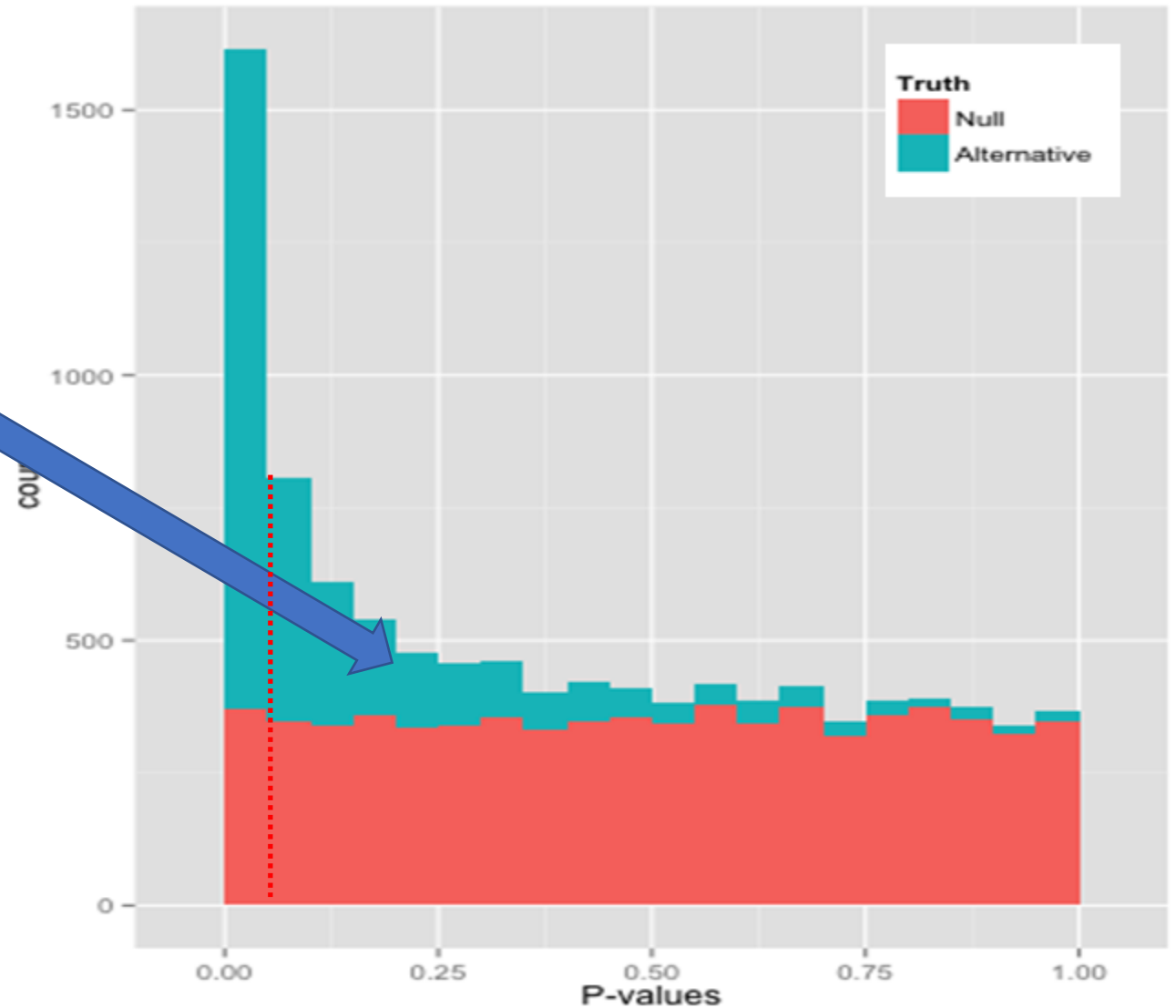- balanced gene groups (false positives)

Because p-values are distributed uniformly under H_0 (and the null is usually true)

Anti-conservative p-values ("Hooray!")

Some positives are not identifiable to your test (all the blue rectangles to the right of p=0.05).

Anti-conservative p-values ("Hooray!")
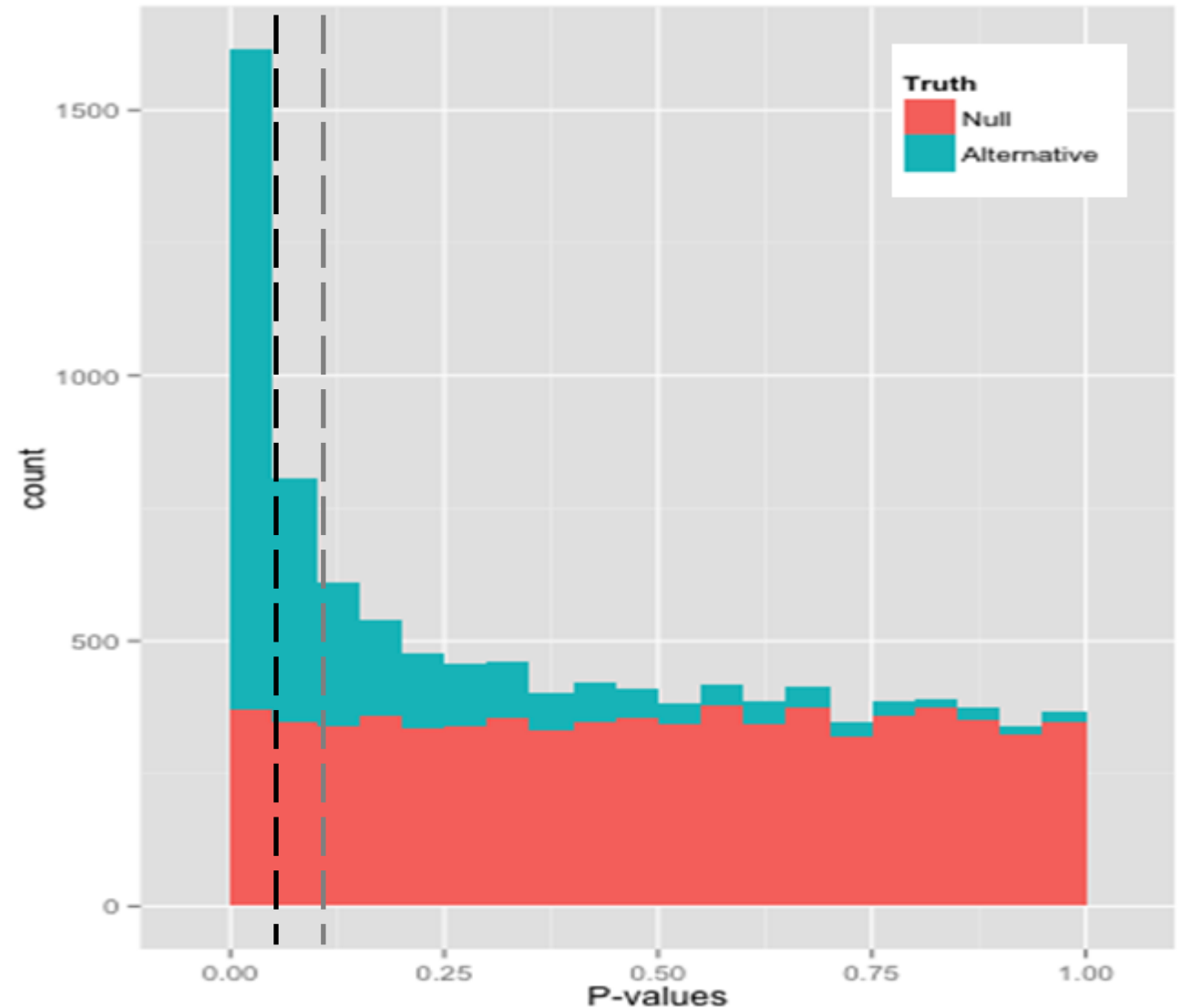
Consider the significance thresholds:
< 0.1 (gray dotted line) and
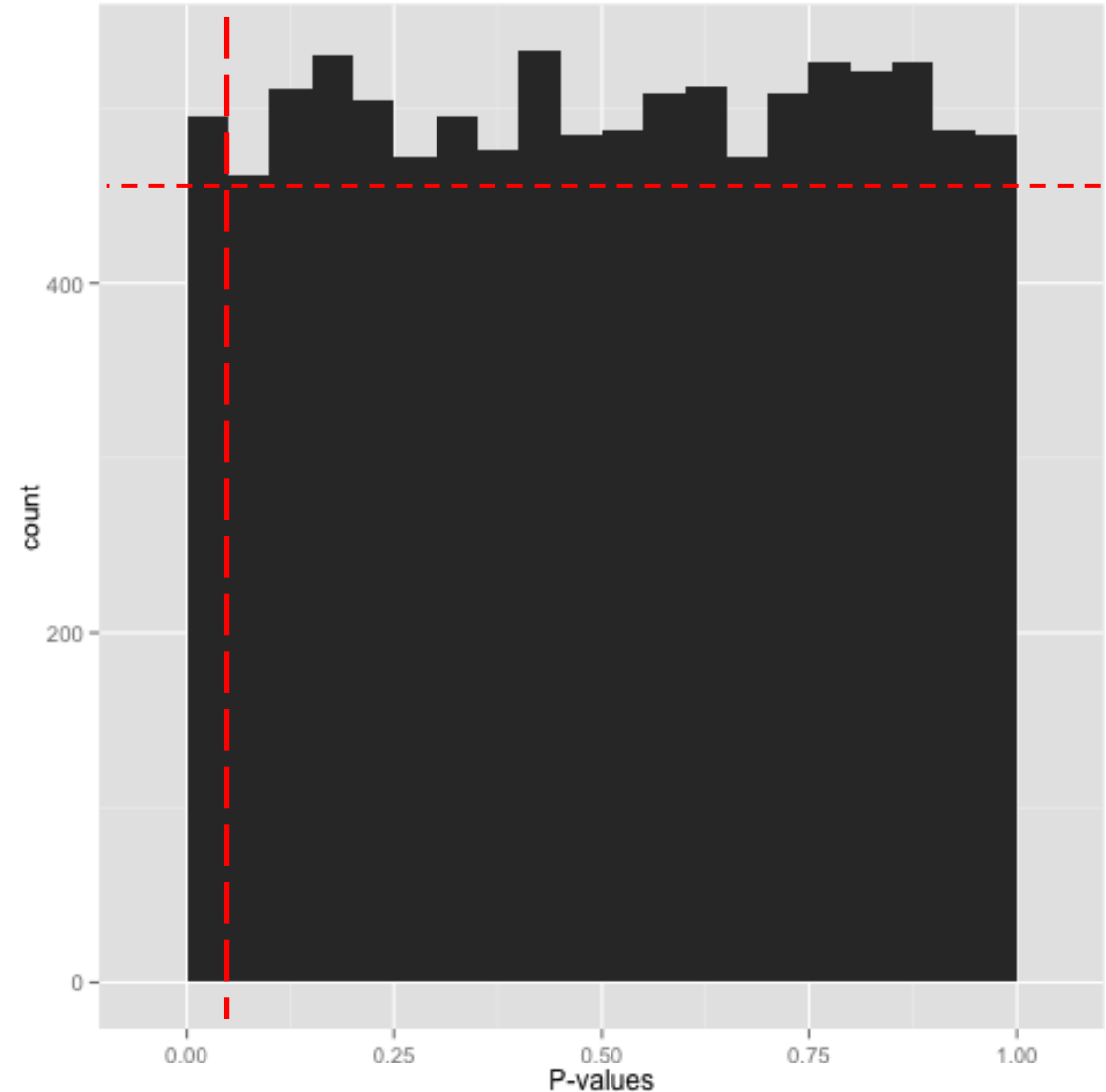< 0.05 (black dotted line)
What would you pick if you want to
- to increase the *number* of true positives

    vs.

- to increase the *ratio* of true positives

Let's not consider multiple testing issue

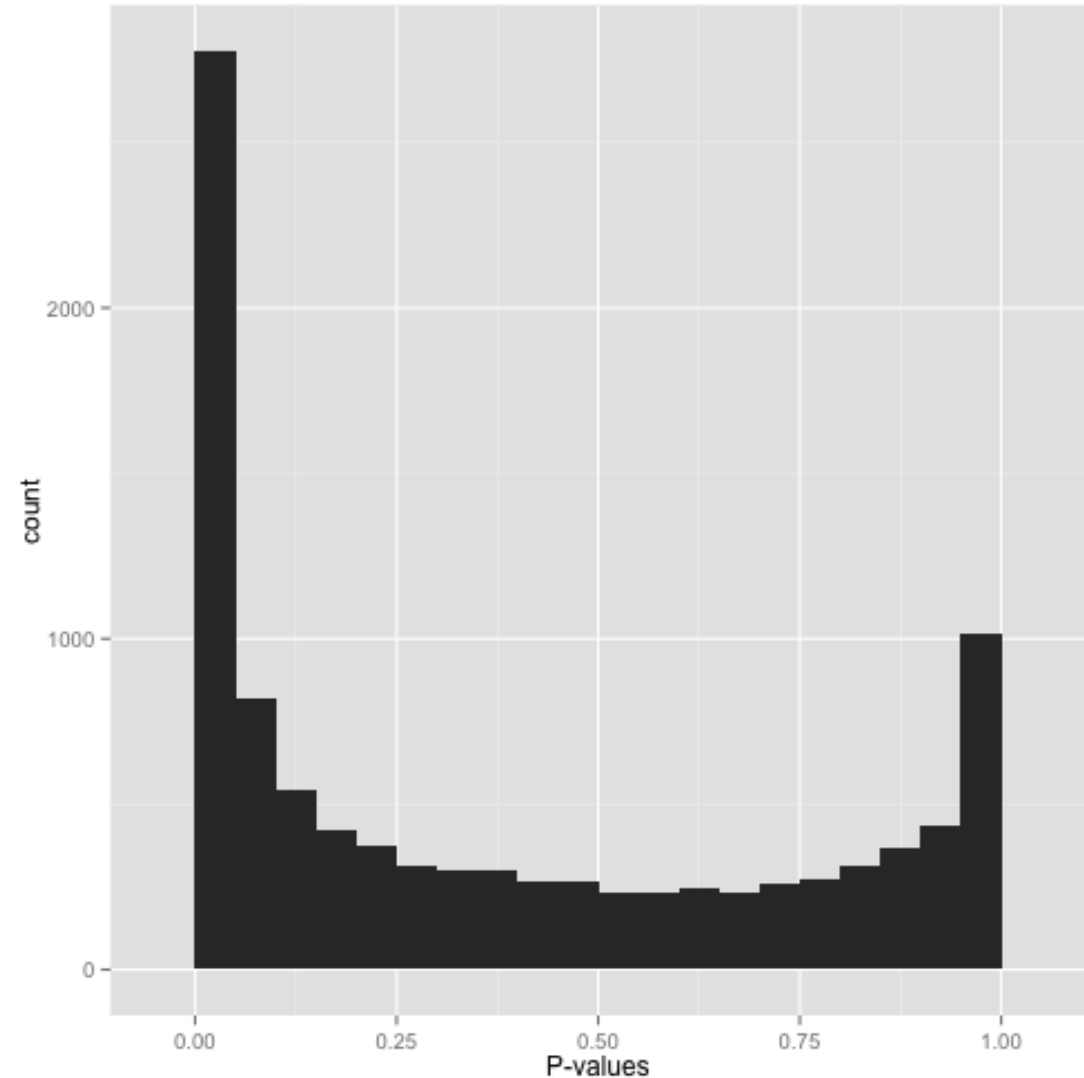http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Uniform p-values ("Awww…")

- P ~ U(0,1) under null
- The picture is consistent with an appropriate model, where H_0 is generally false
- There could be some very significant p-values in the first bin (there just are not a whole lot of significant gene groups)
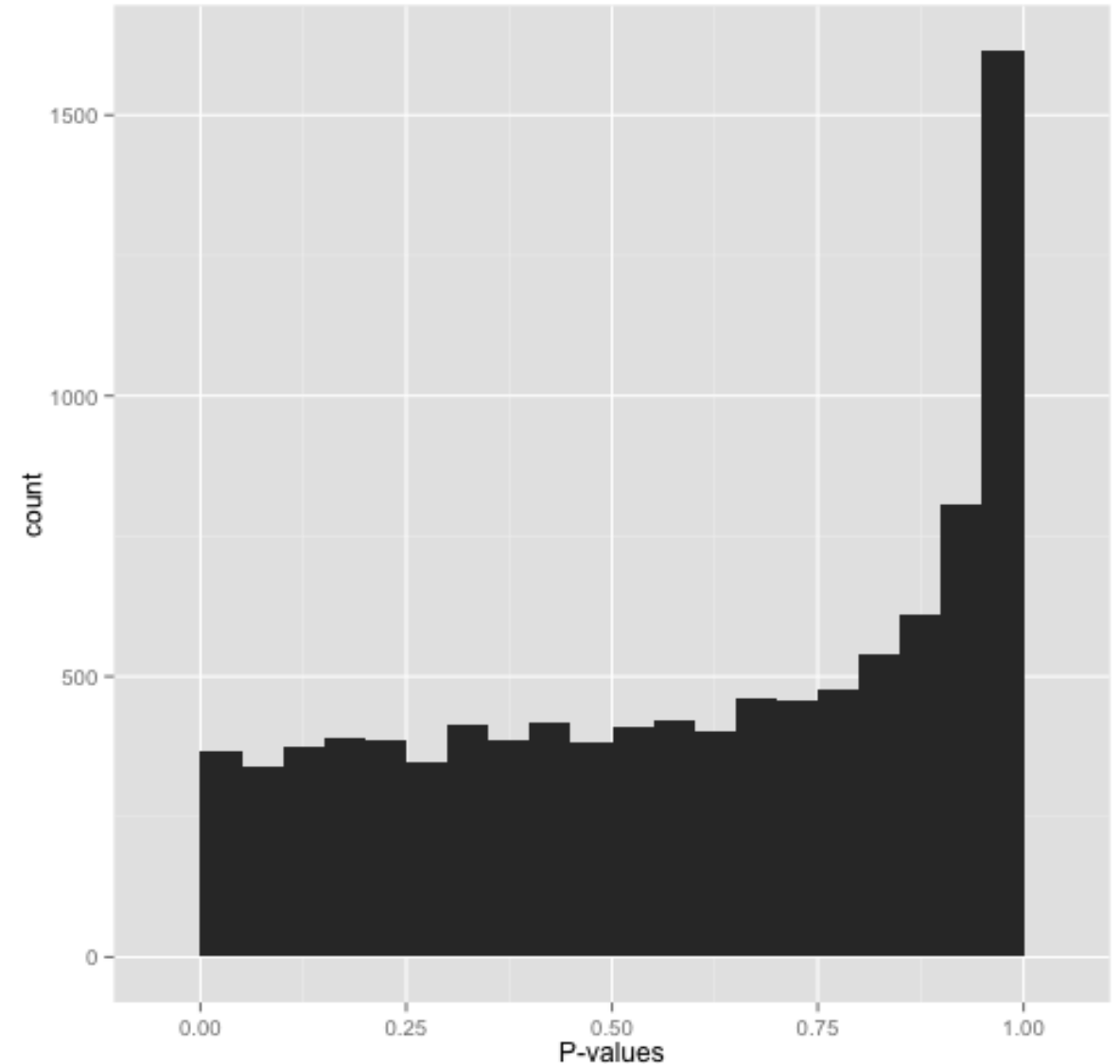

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Bimodal p-values ("Hmmm…")

- Don't proceed just yet (why?)
- Possible reasons:
  - A wrong handling of exceptions in the hypothesis testing (e. g., your method may throw p-value 1 for undefined cases)?
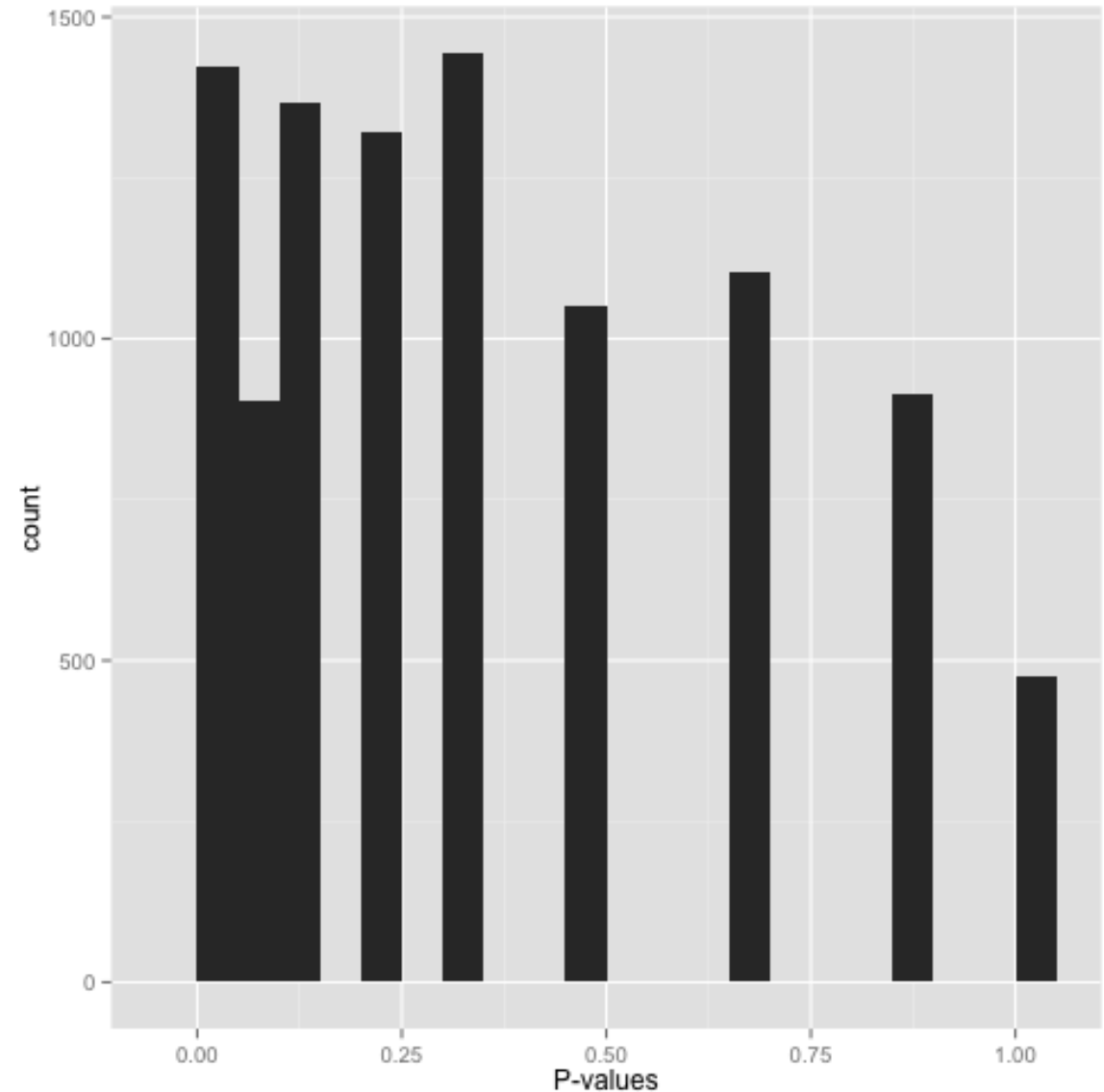 - Did you use a two-tailed test when it should've been one-tailed?

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Conservative p-values ("Whoops...")

- Don't just assume nothing is significant!
- May signify that your test
  - Assumes a wrong distribution/model mis-specfication
    - Designed for normally-distributed data but your data is not?
    - Designed for continuous data but your data are discrete?
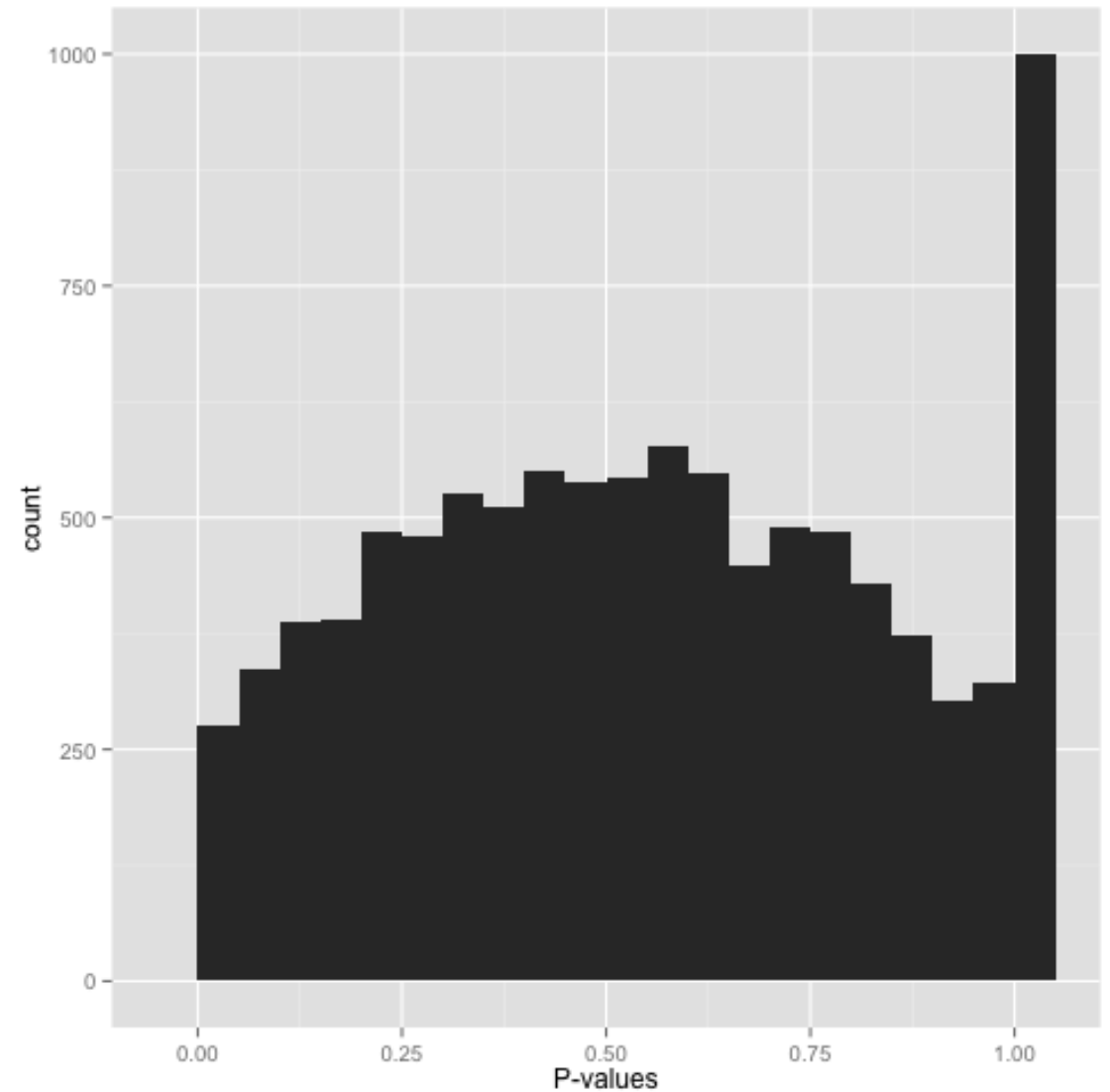 - Already multiple testing corrected?

Sparse p-values ("Hold on...")

- Don't proceed just yet
- Only a small number of distinct p-values
- Did you run
  - A bootstrap or permutation with too few iterations?
  - Non-parametric test (e. g., the Wilcoxon rank-sum test or Spearman correlation) on a small sample size data?
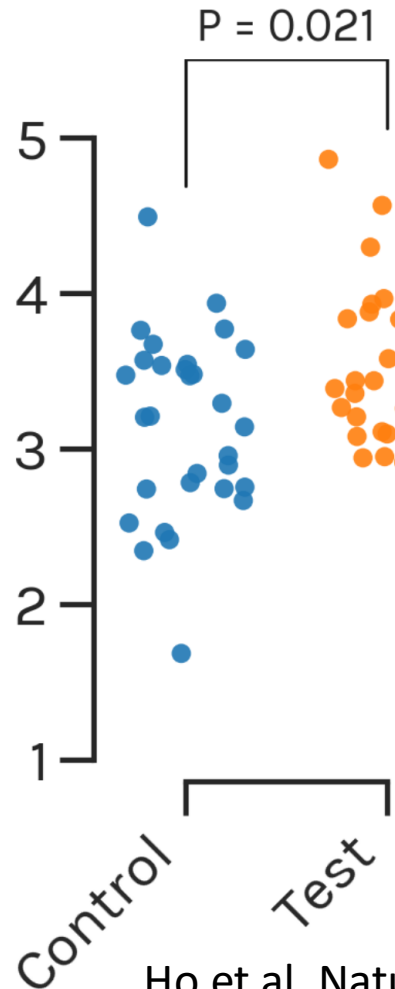

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Something even weirder ("What the…?!?")

- Stop whatever you're doing and find a statistician

http://varianceexplained.org/statistics/interpreting-pvalue-histogram/

Recent criticism on significance testing (p-value)

Let's say an experiment with 30 control subjects, and 30 test subjects

P = 0.021

Ho et al. Nature Methods, 2019

- Observed values shown
- Underlying distribution not shown
- Effect size (the mean difference in the observed samples) not shown
- Confidence and likelihood of effect size not shown

Recent criticism on significance testing (p-value)
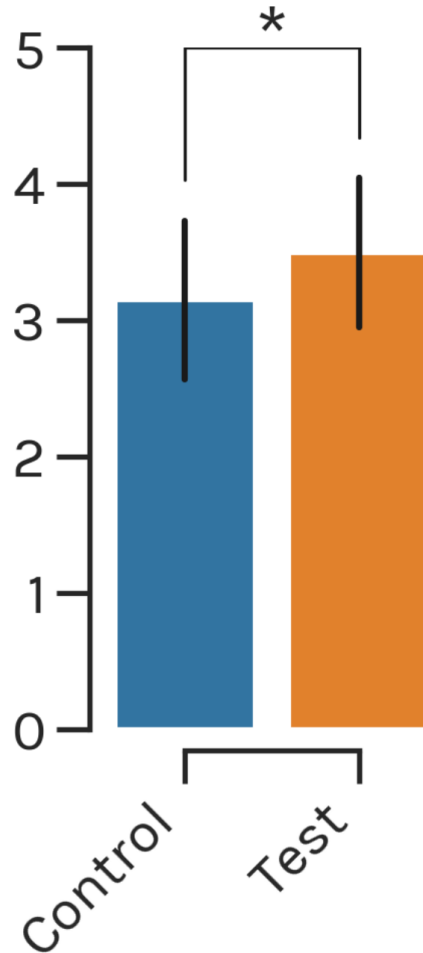
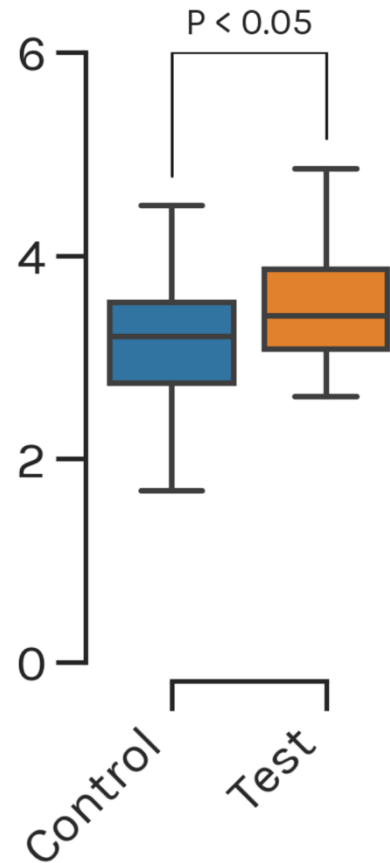Let's say an experiment has 30 control subjects, and 30 test subjects

- Observed values not shown
- Underlying distribution not shown
- Effect size not shown
- Confidence and likelihood of effect size not shown

Ho et al. Nature Methods, 2019
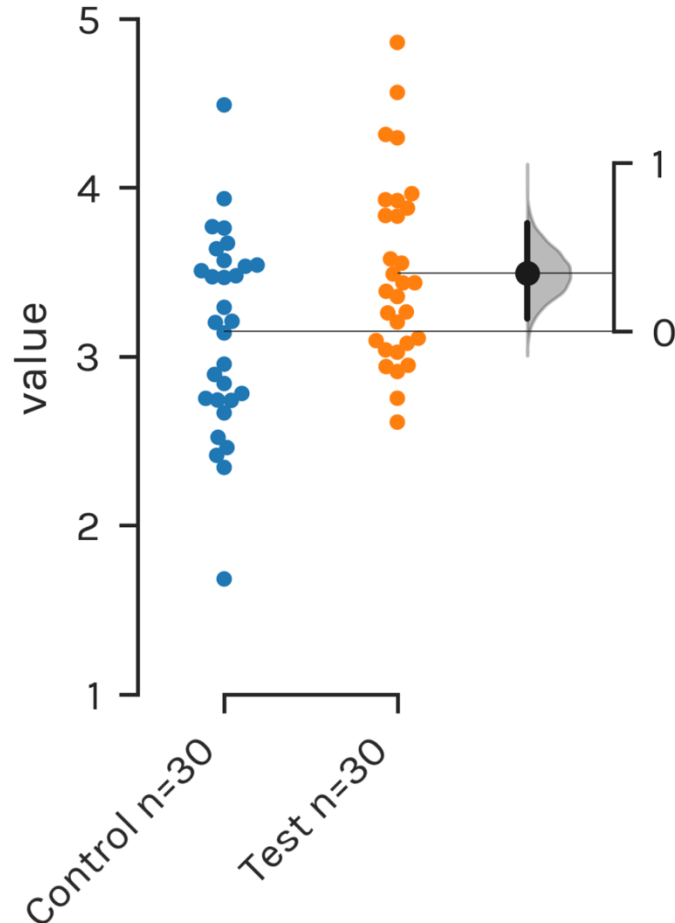
Recent criticism on significance testing (p-value)

Let's say an experiment has 30 control subjects, and 30 test subjects

- Medians, quartiles, minima, and maxima shown
- Observed values not shown
- Effect size not shown
- Confidence and likelihood of effect size not shown

Ho et al. Nature Methods, 2019
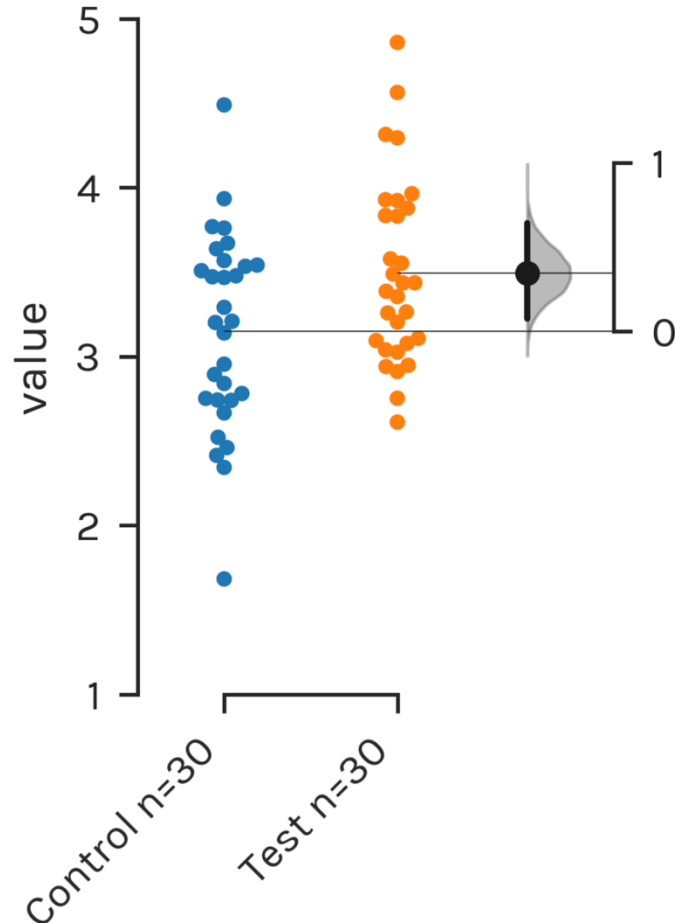
Introducing estimation statistics

Let's say an experiment has 30 control subjects, and 30 test subjects

- Data points in swarm plot
- Effect size on a separate but aligned axes as a bootstrap 95% confidence interval (CI)
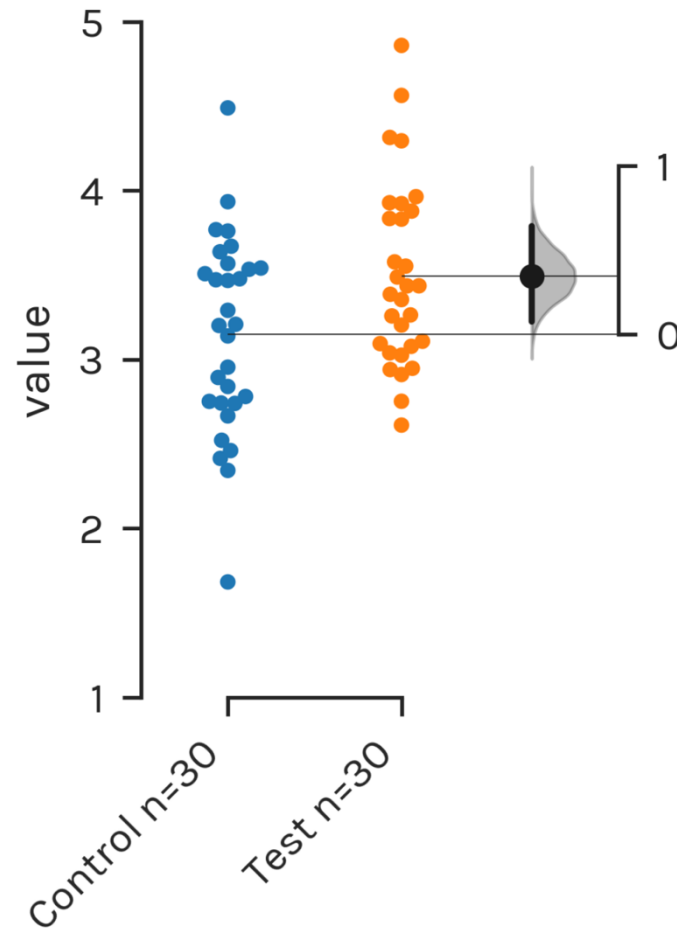
Ho et al. Nature Methods, 2019

Introducing estimation statistics

Let's say an experiment has 30 control subjects, and 30 test subjects

- Observed values shown
- Effect size shown in the curve
- Precision of effect size shown
- Confidence and likelihood of effect size shown

Ho et al. Nature Methods, 2019

Estimation plot vs. barchart or boxplot (P-value based)

	Bars-and-Stars	Boxplot & *P*	Estimation Plot
Avoid false dichotomy	✗	✗	✓
Display all observed values	✗	✗	✓
Focus on intervention effect size	✗	✗	✓
Visualize estimate precision	✗	✗	✓
Show mean difference distribution	✗	✗	✓

Ho et al. Nature Methods, 2019