

Visualizing Uncertainty

Michael Friendly

Psych 6135

<http://euclid.psych.yorku.ca/www/psy6135/>

Topics

- Problems with uncertainty in visualization
- Visualizing distributions
- “Error bars”
- Uncertainty in fitted curves
- Hypothetical outcome plots
- Cartographic uncertainty

Sources of uncertainty

- Where does the uncertainty in statistics come from?
There are three main sources:
 - **Data:** data can contain random processes, or have missing entries.
 - **Assumptions:** model assumptions take plausible values with distributions.
 - **Models:** there is choice over the techniques and models we use.
 - Different analysts may choose different methods, yielding different estimates.

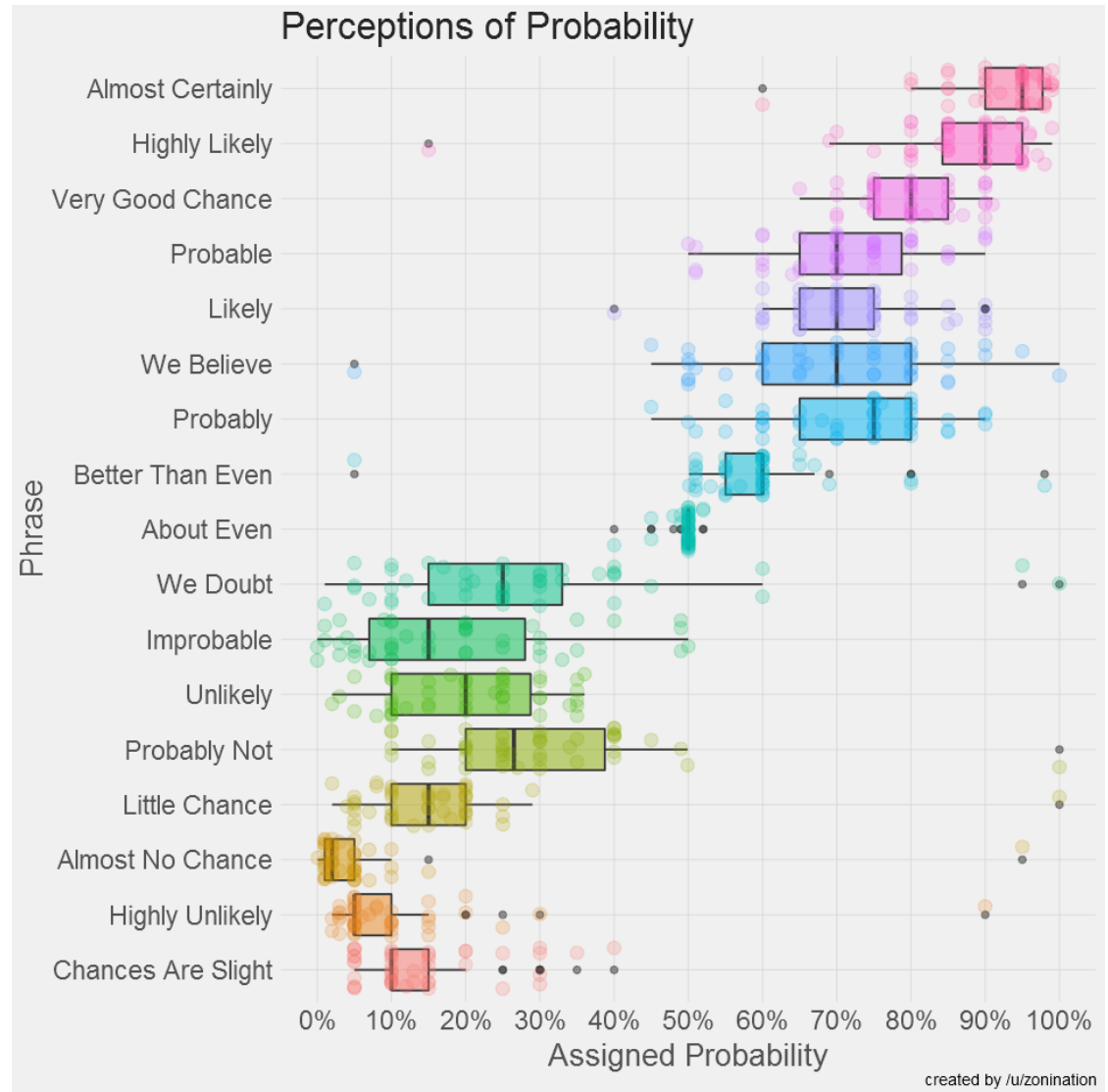
Problems

- Uncertainty is fundamental to data analysis & models
 - data: IQR, std dev., std error, ... (variation)
 - assumptions: we assume some distribution for errors, e.g., $\varepsilon \sim N(0, \sigma^2)$, independent with constant variance
 - models:
 - classical: confidence intervals, p-values;
 - Bayesian: credible intervals, posterior distributions
- In data graphics,
 - Easy to show “fit” – means, regression estimates, ...
 - Harder to show the uncertainty in these numbers

How people view “probability”

What makes this graph successful?

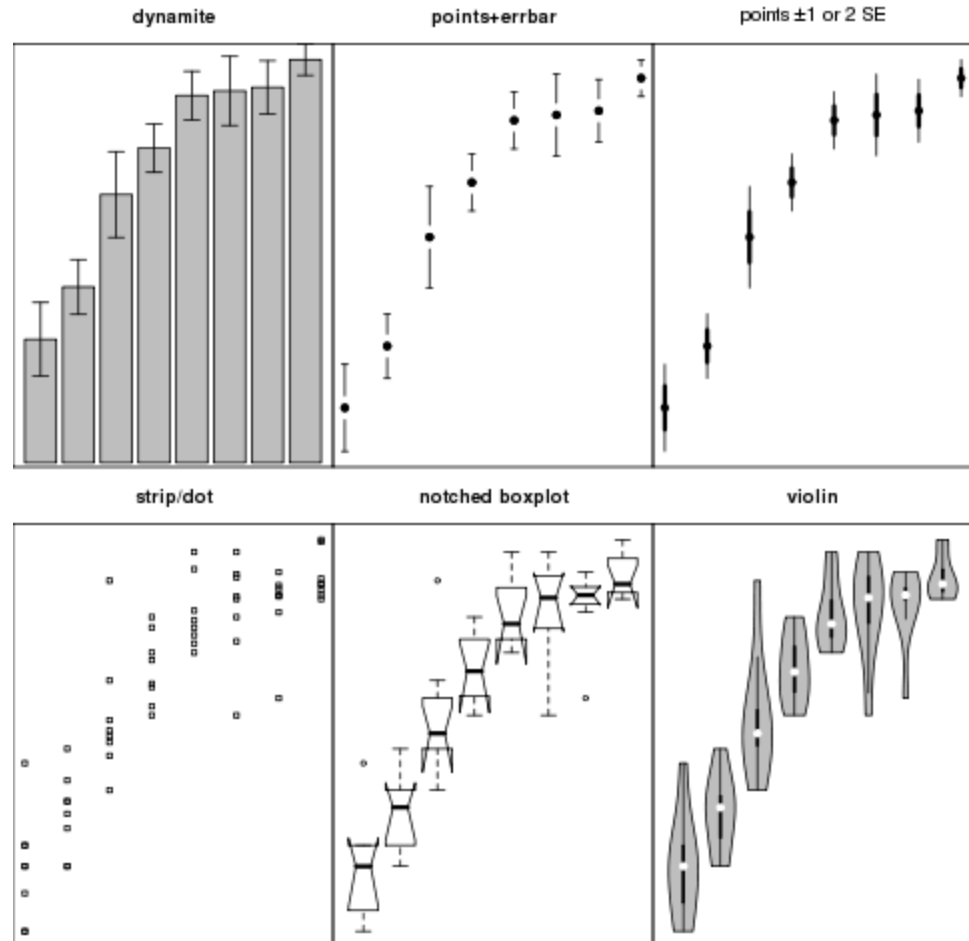
Note the wide range of variability (uncertainty) in the estimates:
“about even” vs. “we believe”



Comparing groups: Summary + Uncertainty

Six different graphs for comparing groups in a one-way design

- which group means differ?
- equal variability?
- distribution shape?
- what do error bars mean?
- unusual observations?



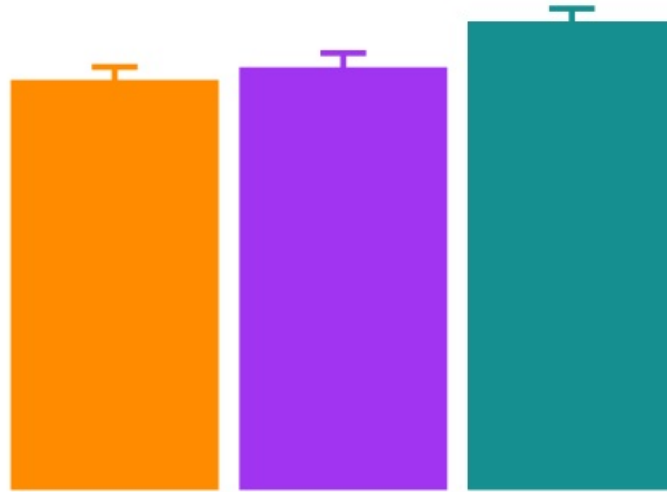
Never use dynamite plots

Always explain what error bars mean

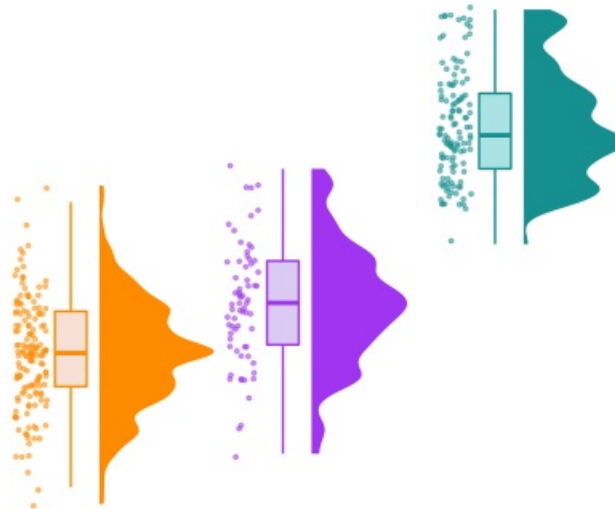
Consider tradeoff between
summarization & exposure

Don't dynamite me!

rage

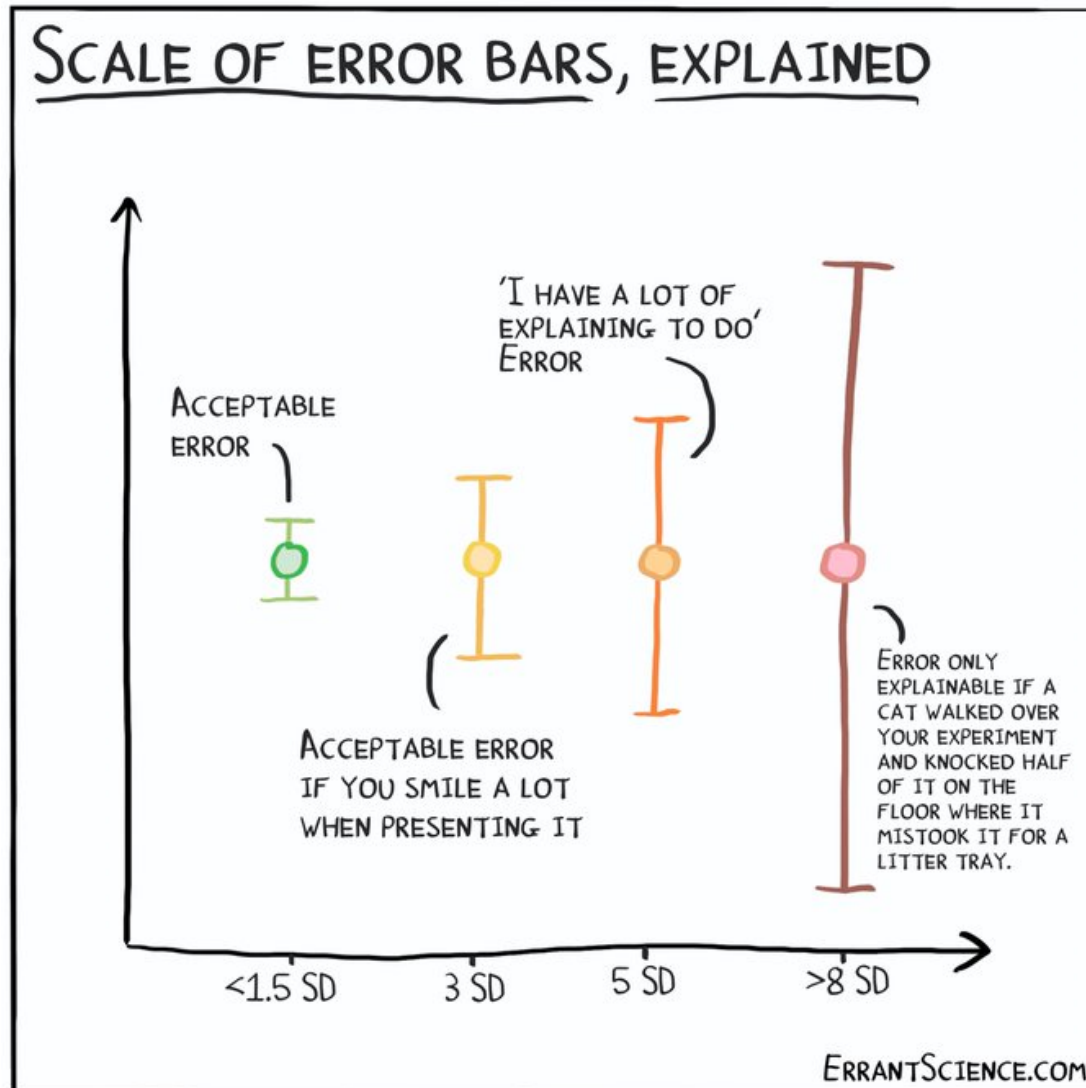


joy



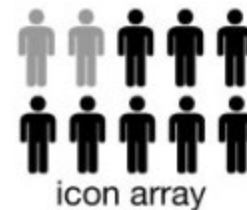
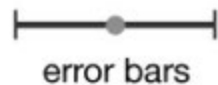
Graphic: Cédric Scherer (@CedScherer)

Error bars: When should I worry?

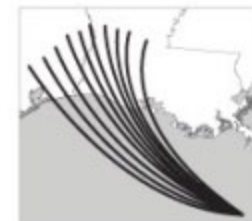
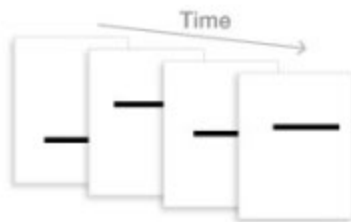
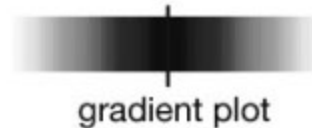


Graphical annotations for uncertainty

Intervals and Ratios



Distributions

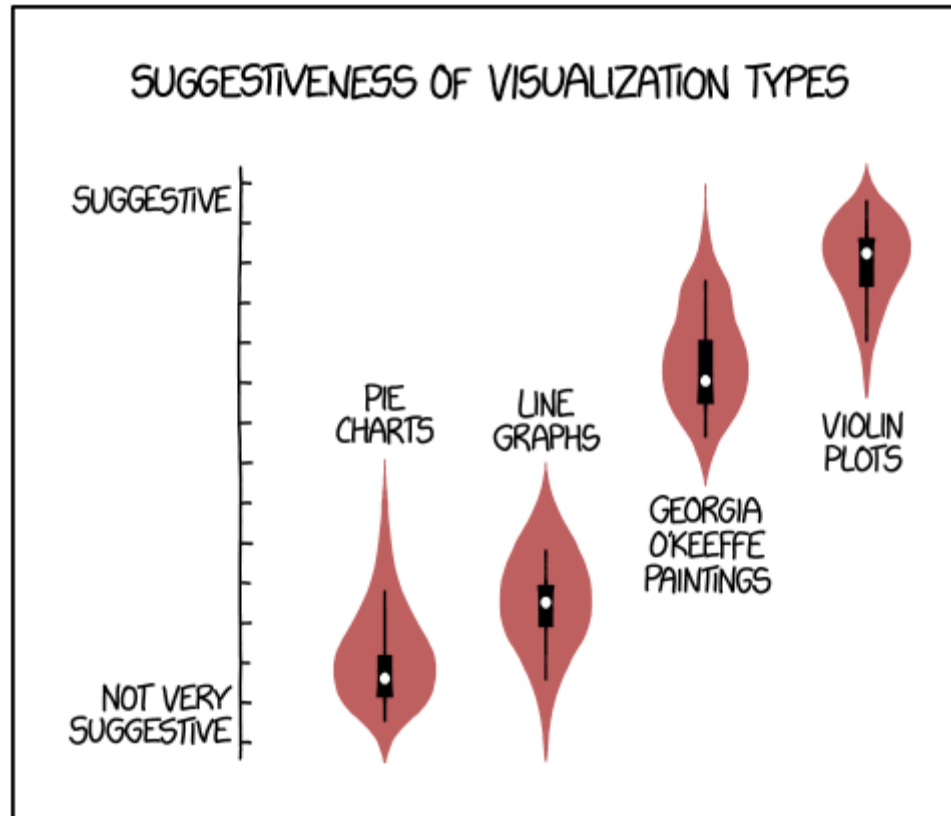


hypothetical outcome plot

quantile dot plot

ensemble plot

Visualizing distributions



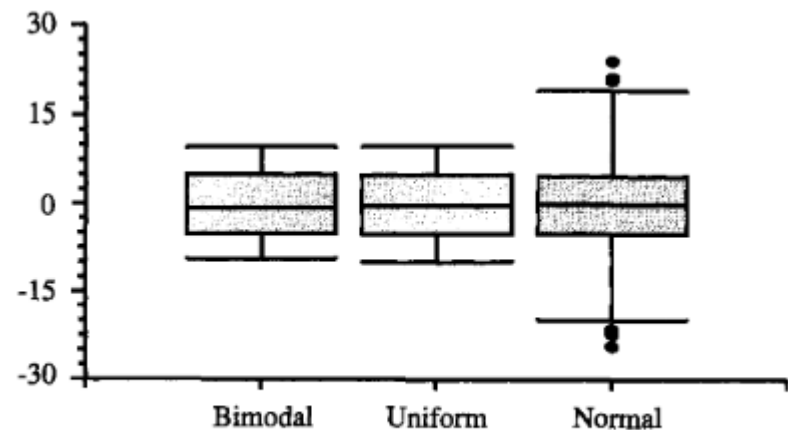
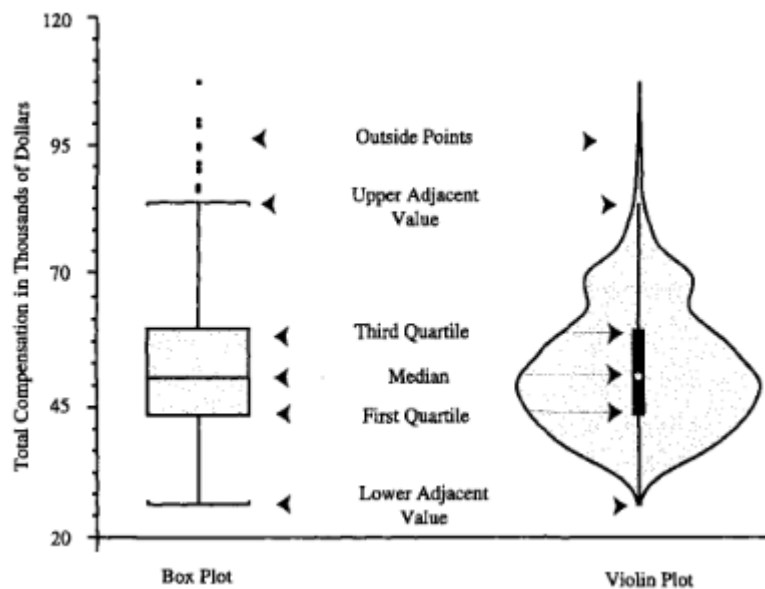
From: <https://xkcd.com/1967/>

Violin plots

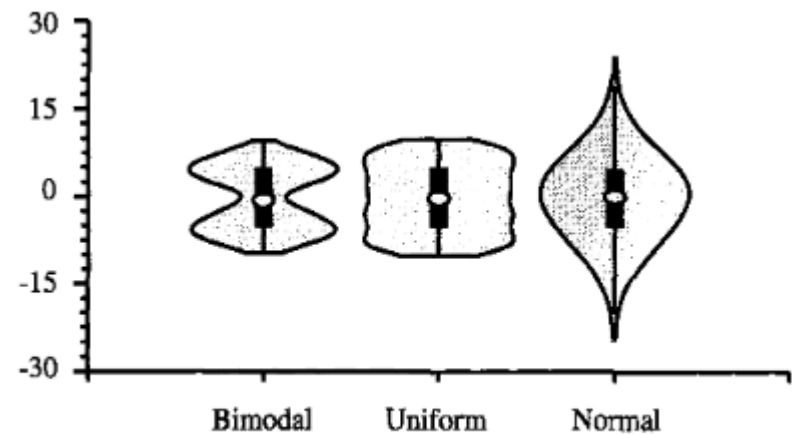
Boxplots are great for \sim normal data

- Shows center, spread, outliers

Violin plots add a (reflected) density curve to show the **shape** of the distribution



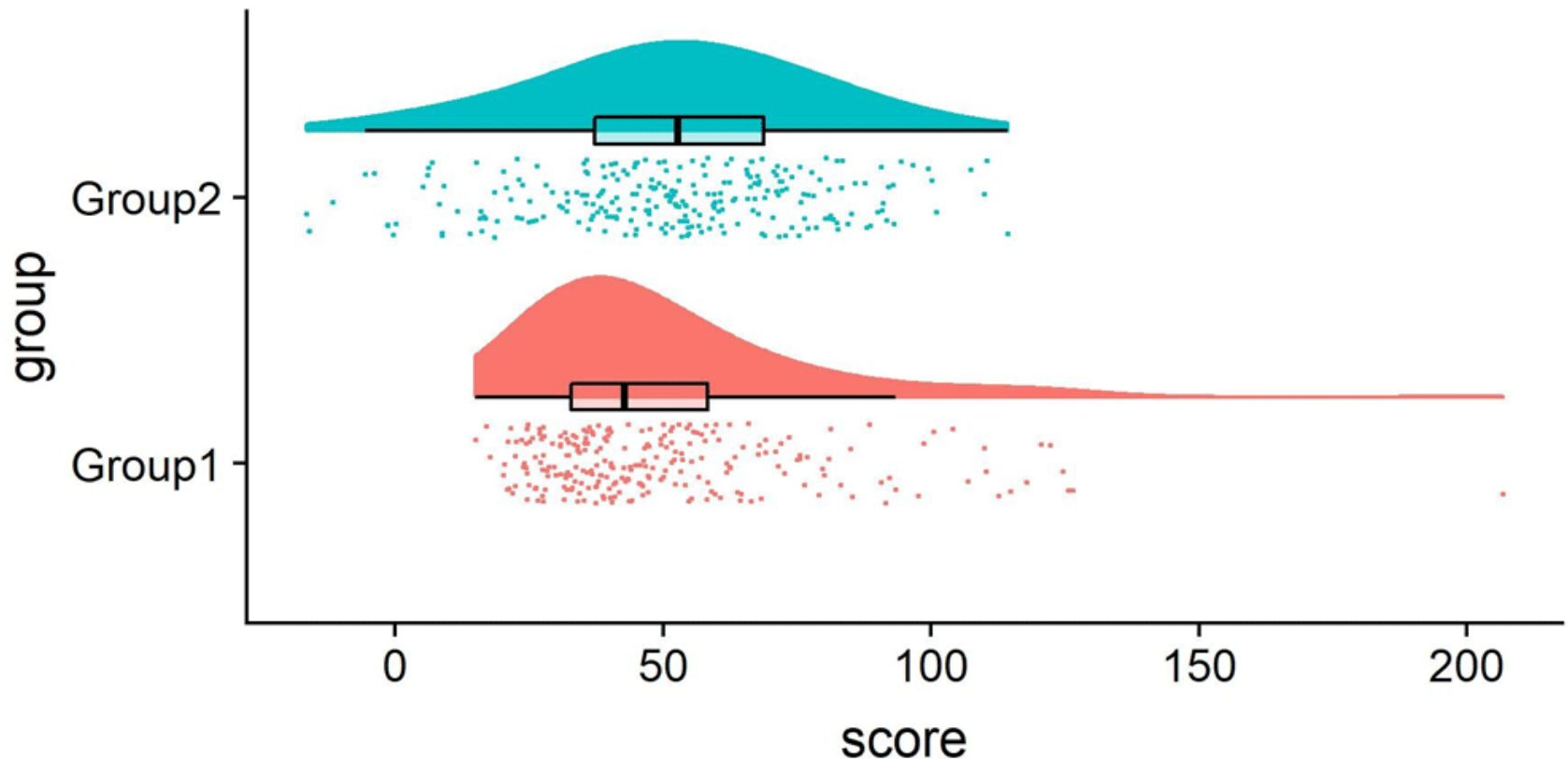
(a)



(b)

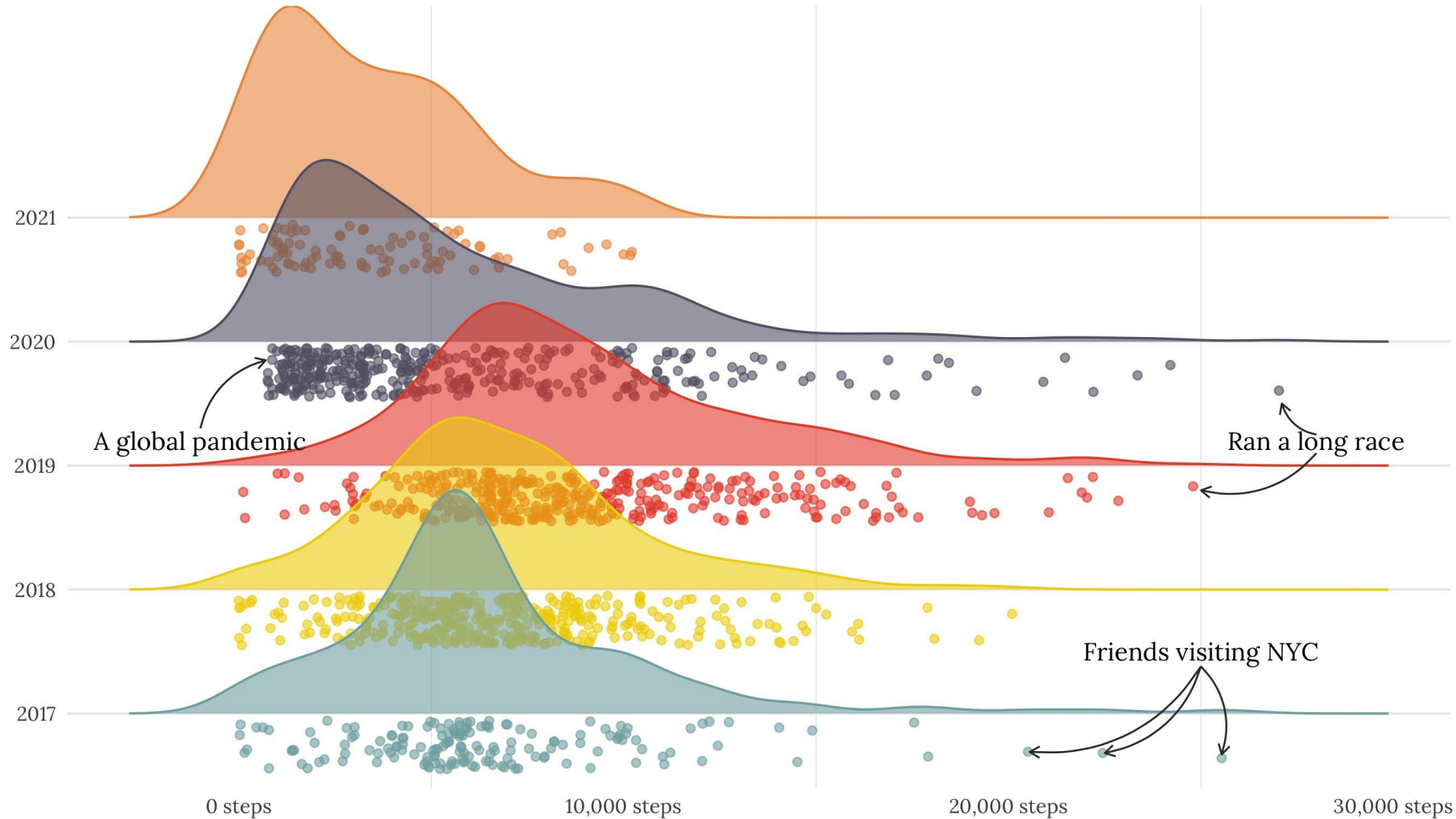
Raincloud plots

Raincloud plots are similar, but also show the observations as jittered points

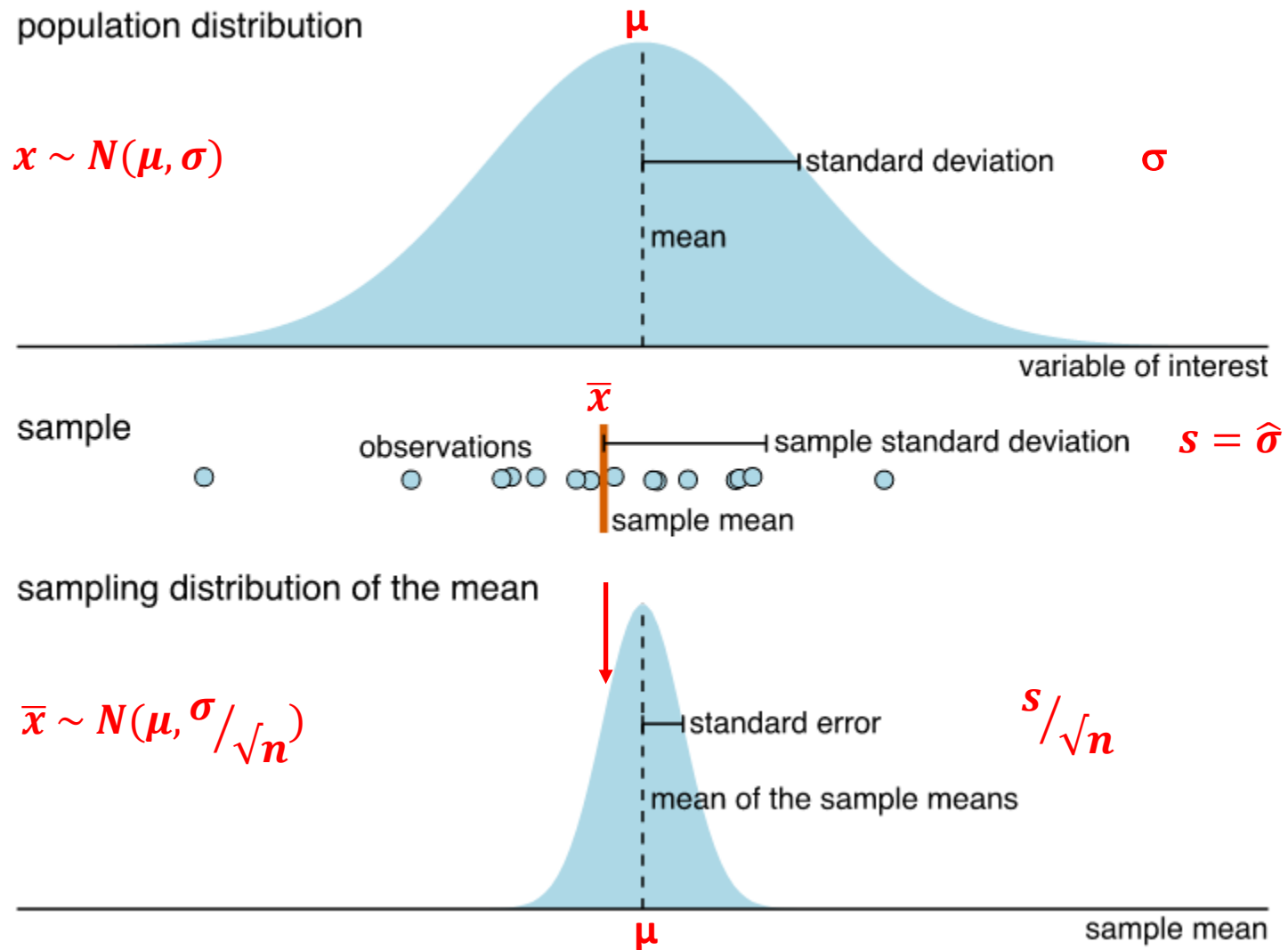


How many steps have I taken since 2017?

Since July 2017, I have tracked the number of steps I've taken (almost) every day. In a little over 4 years, I have taken **9,232,798** steps. This includes days spent walking around New York with visiting friends, running a half-marathon, and a pandemic that dropped my step count to nearly 0.

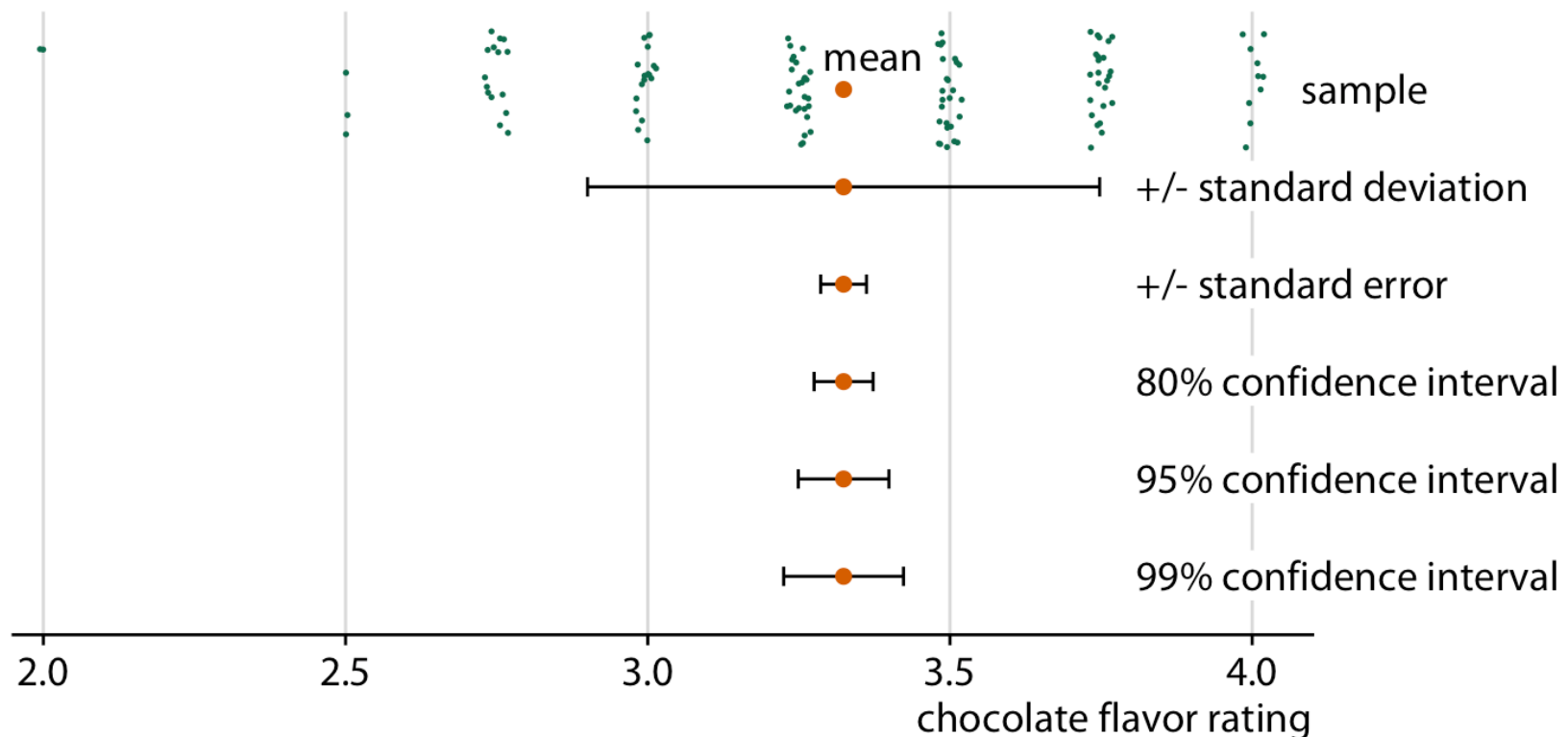


Key ideas of statistical sampling



Visualizing distributions: Error bars

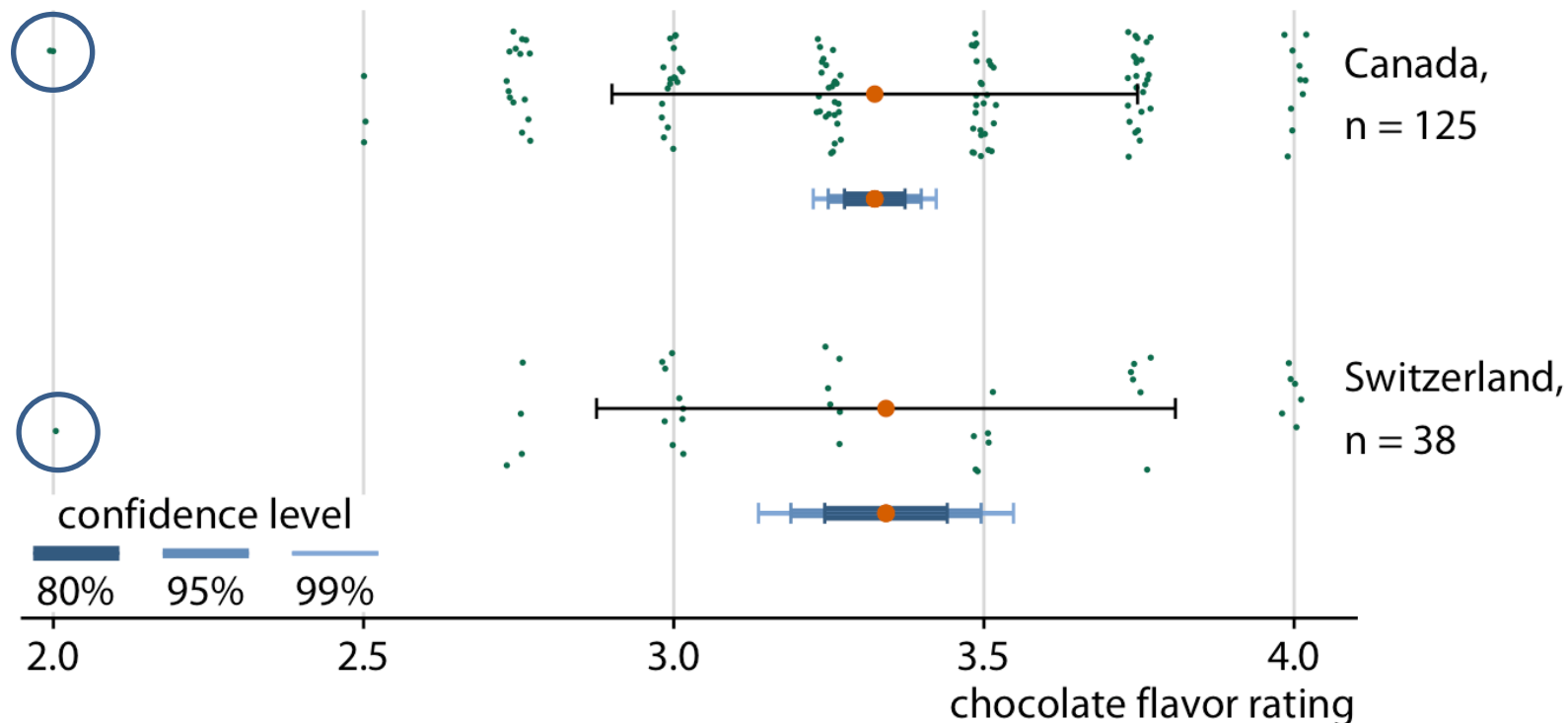
There are many ways to show variability in a single sample



Expert ratings of 125 chocolate bars manufactured in Canada

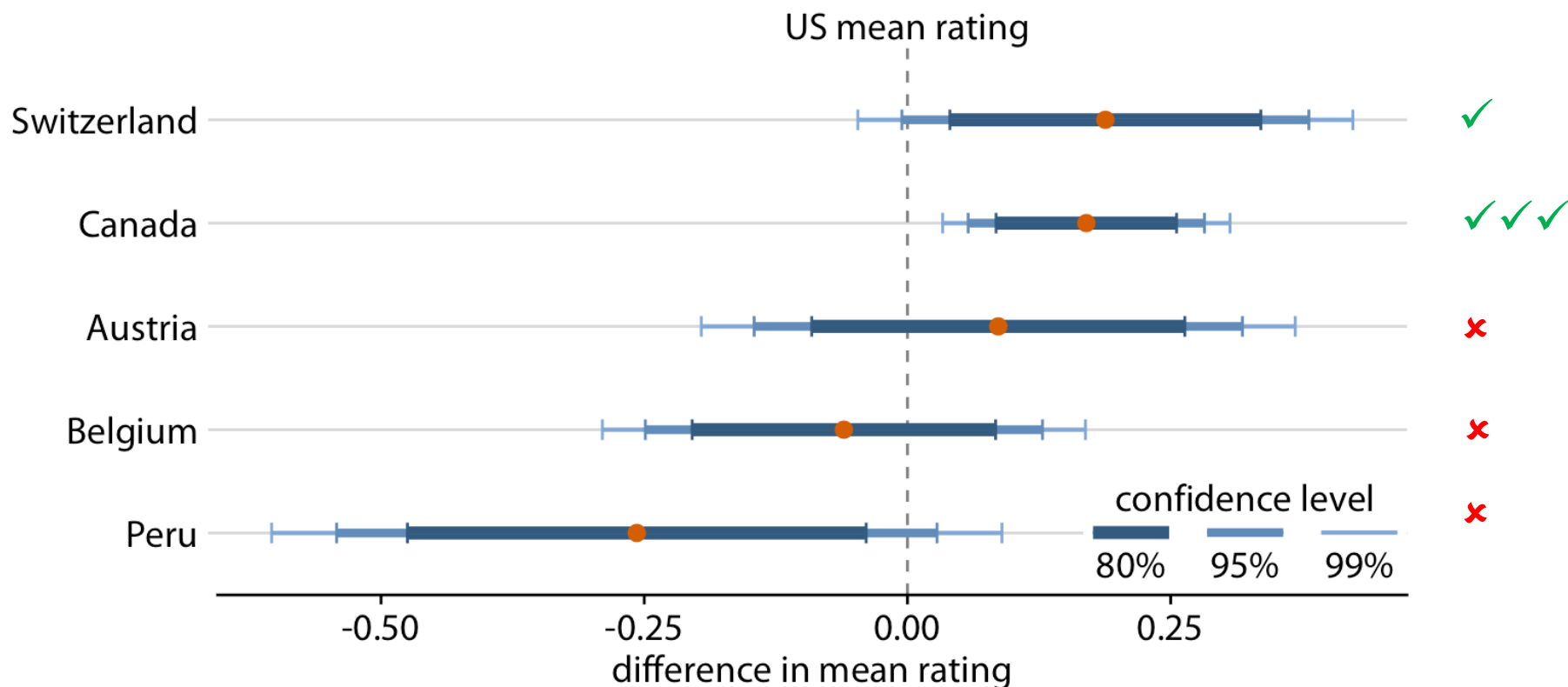
Comparing distributions: Sample size

- means and standard deviations are similar for Canada & Switzerland
- confidence interval widths $\sim 1/\sqrt{n}$
- can show different sized confidence bands together
- dots show the data: are there any outliers?



Comparing distributions: Contrasts

- For comparison of one group to all others, plot the **difference** directly
- Easy to see which differences exclude 0, at what confidence level



Intervals: Direct vs. Differences

The standard error for the difference between two means is always **larger than the standard error of either mean**

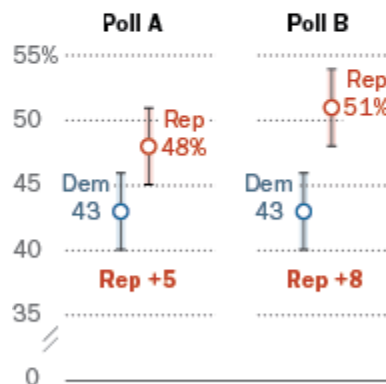
$$SE(\bar{x}) = \sqrt{s^2 / n}$$

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{s_1^2 / n_1 + s_2^2 / n_2}$$

For election polls, different measures of the race have different margins of error

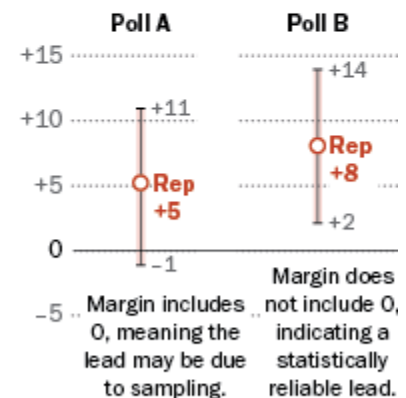
The margin of error reported for most polls applies to support for individual candidates ...

Margin of error for single candidate support
(MOE +/- 3 pct. points)



... while the margin of error for a candidate's lead is nearly twice as large.

Margin of error for difference between two candidates' level of support (%Rep - %Dem)
(MOE +/- 6 pct. points)



Source: Hypothetical polling results from a fictitious election.

PEW RESEARCH CENTER

What kind of intervals?

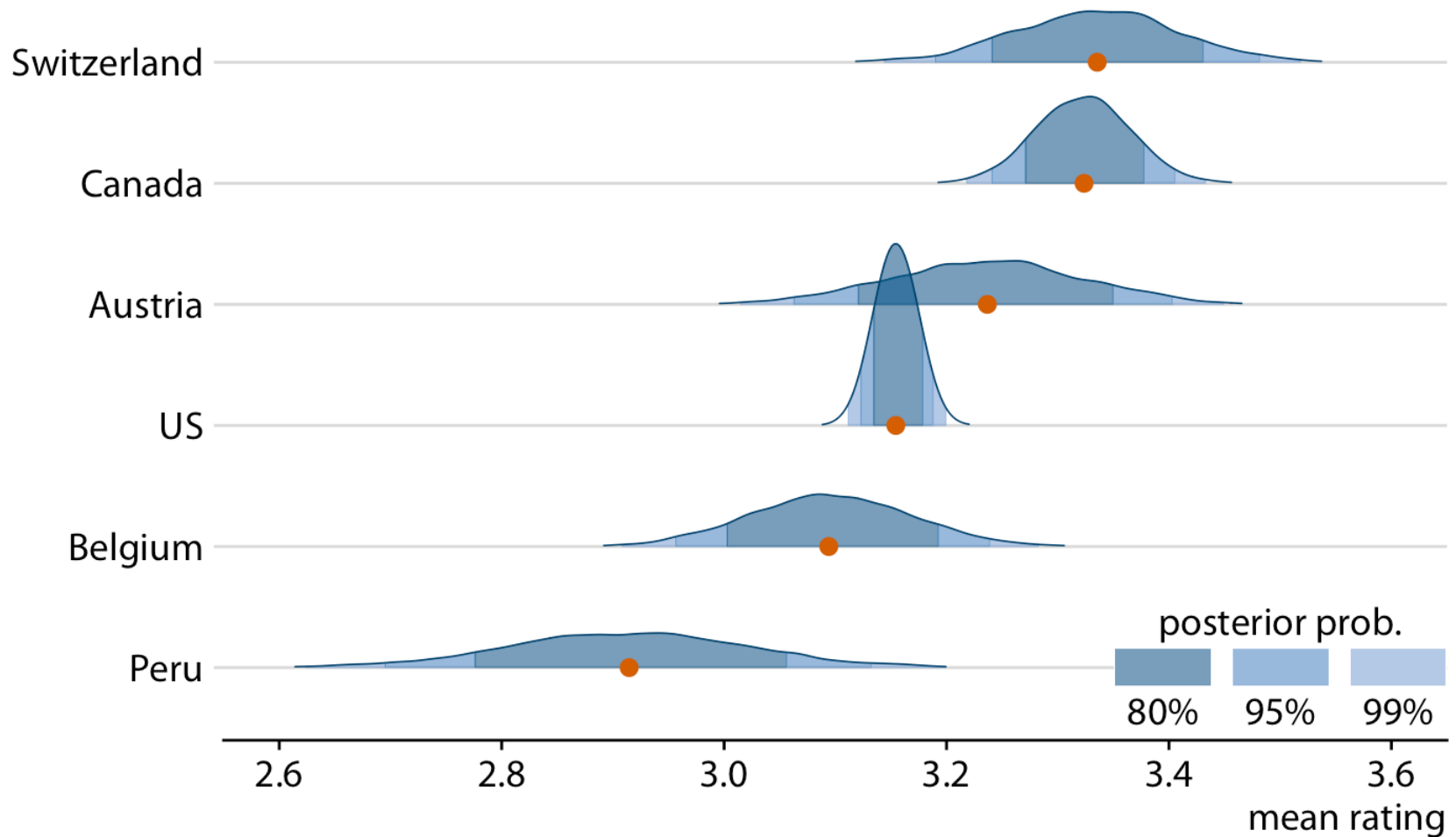
Frequentist

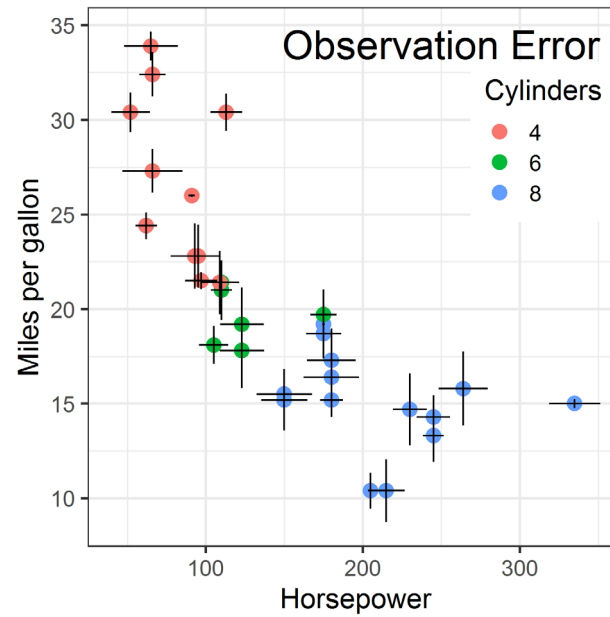
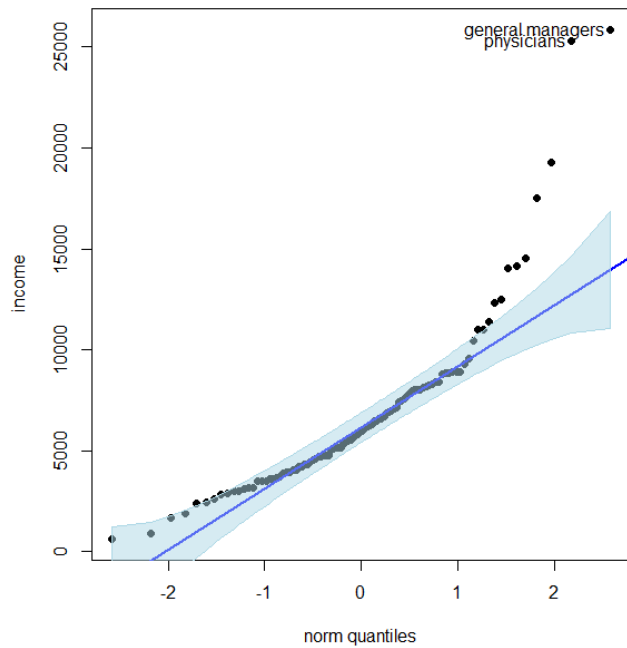
- Confidence interval
- Scope: repeated (hypothetical) samples
- Center: parameter estimate
 - $\mu \rightarrow \bar{x}; \beta \rightarrow \hat{\beta}$
- Width: \sim std. error = $\hat{\sigma}/\sqrt{n}$
- Interpretation: true parameter w/in this interval 1- α %

Bayesian

- Credibility interval
- Scope: repeated draws from the posterior distribution
- Center: median of posterior distribution
- Width: MAD sd of posterior
- Interpretation: Given prior, expect parameter w/in this interval 1- α %

Bayesian intervals





- QQplots
- Model fit plots

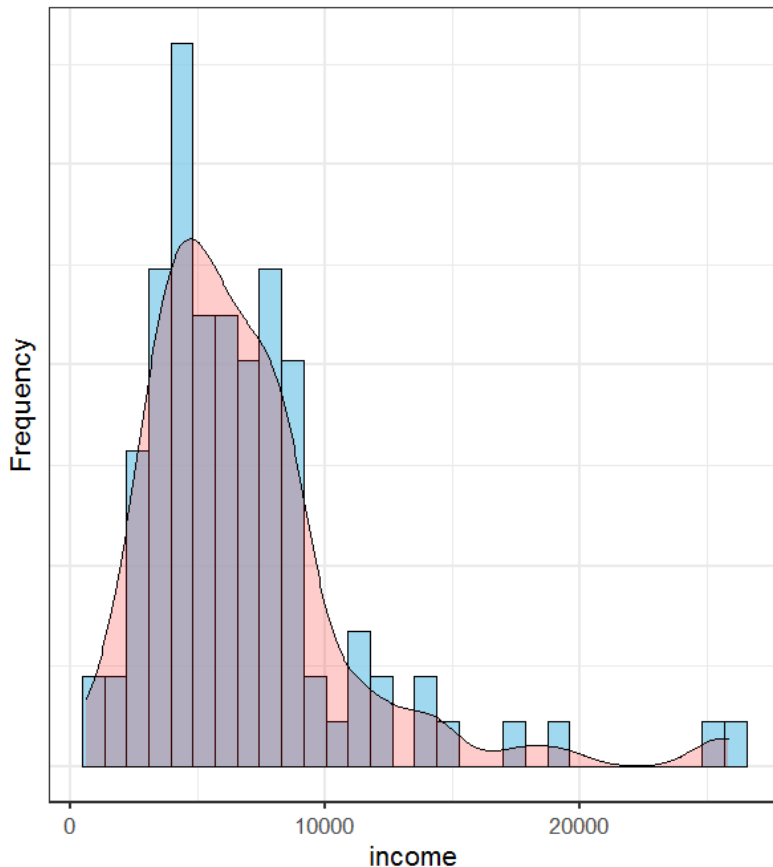
Uncertainty in fits & curves

QQ plots

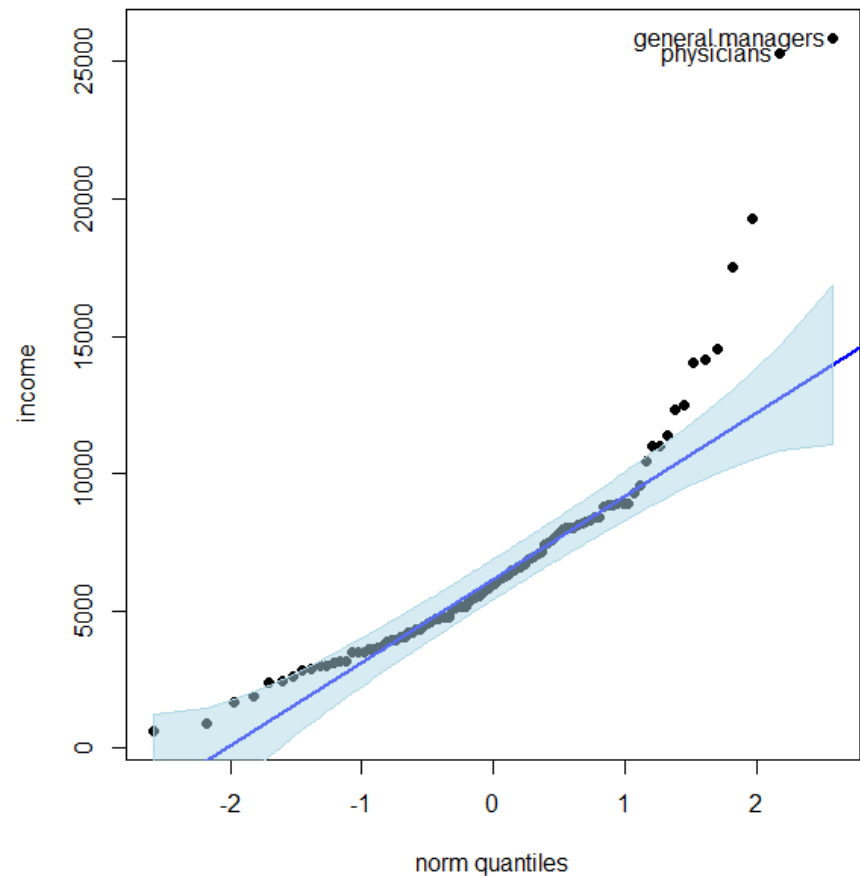
- How close is my data to a {Normal | exponential | χ^2 } distribution?
- There are lots of statistical tests, but these don't tell **why** or **where** a distribution is rejected.
- These tests are also overly sensitive to small departures
- Plot observed Quantiles vs. theoretical Quantiles
 - If observed \sim theoretical with slope = 1, OK
 - **Confidence bands** help to identify outliers
- Use cases:
 - Is a single variable reasonably normally distributed?
 - Are the residuals from my linear model Normal?
 - Outliers in multivariate data? $D^2 \sim \chi^2 \rightarrow$ chisq QQ plot

Prestige data: income

Income is clearly positively skewed.
(But normality is not required for predictors.)

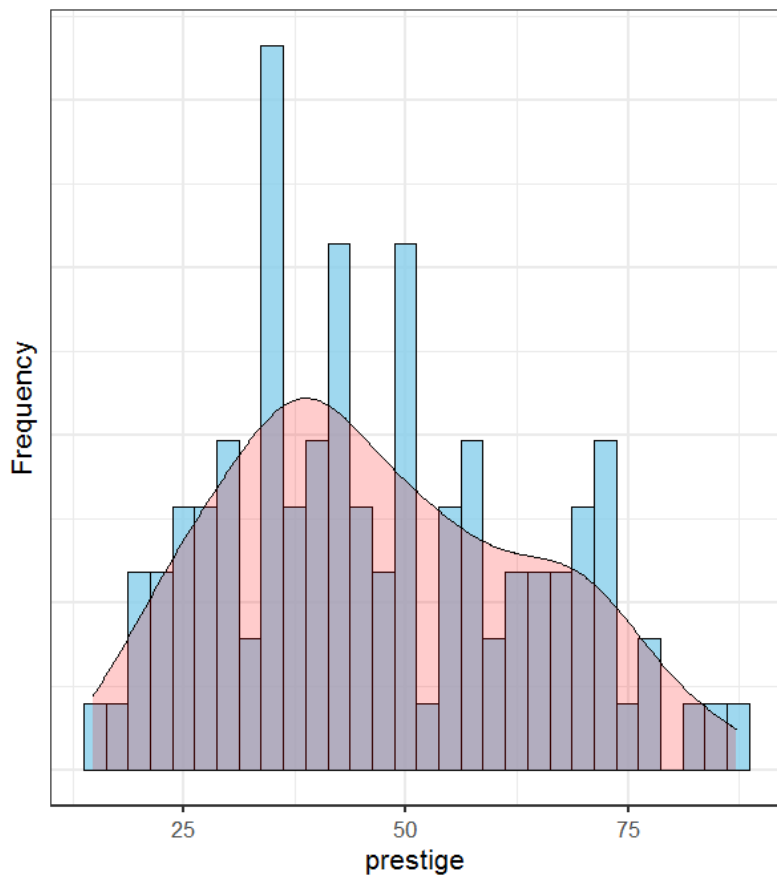


This shows up as a U-shaped pattern
The 95 % confidence band shows
greatest departure in the upper tail

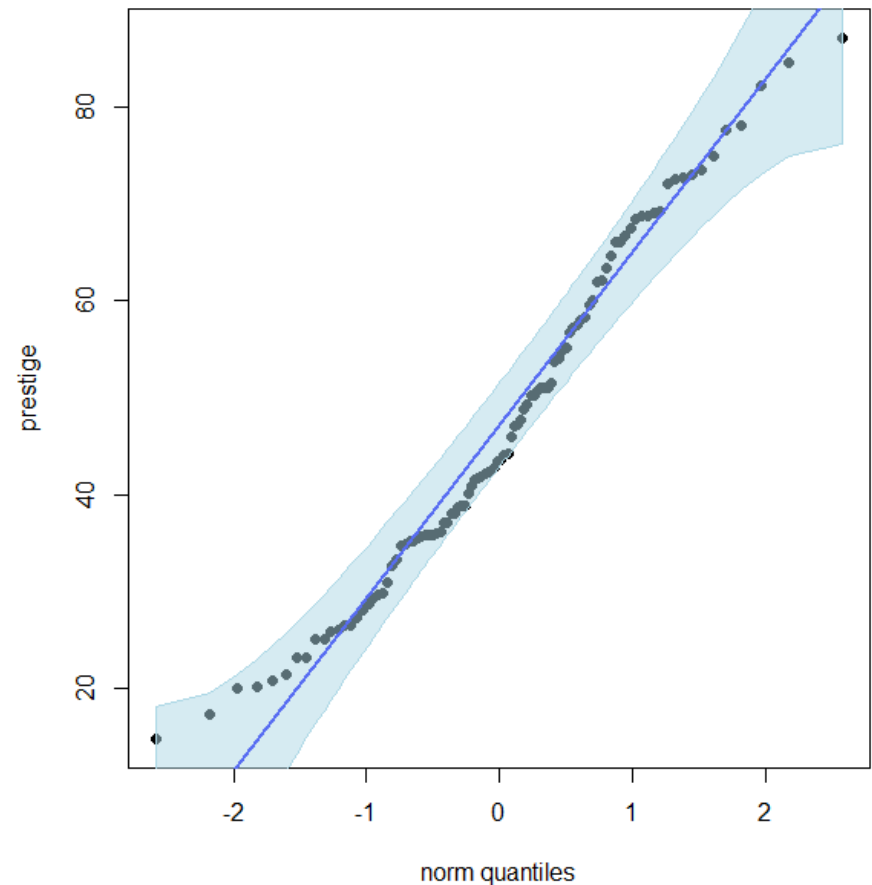


Prestige data: prestige

Occupational prestige doesn't look precisely normal, but not that bad.



The 95% confidence band includes all the observations



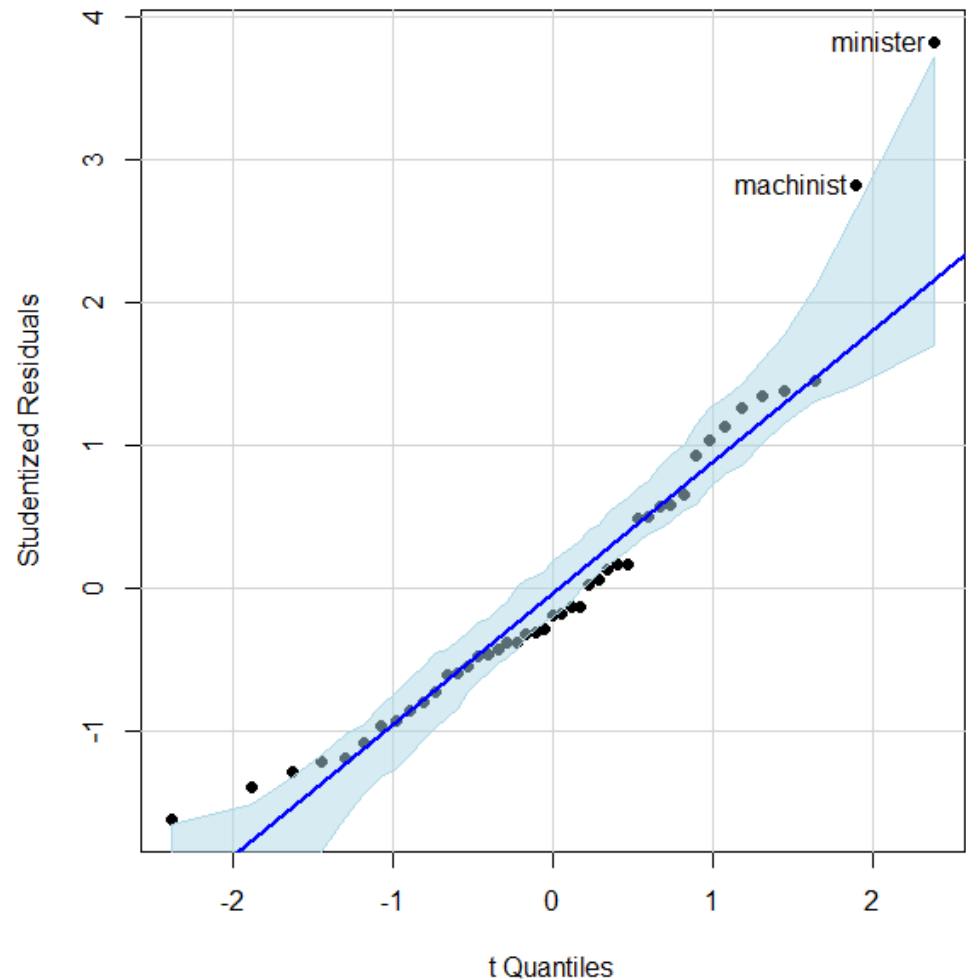
Prestige data: residuals

Normality of residuals is more important for linear models

Some small evidence of + skew

Confidence bands help to identify potential outliers – badly fitted pts

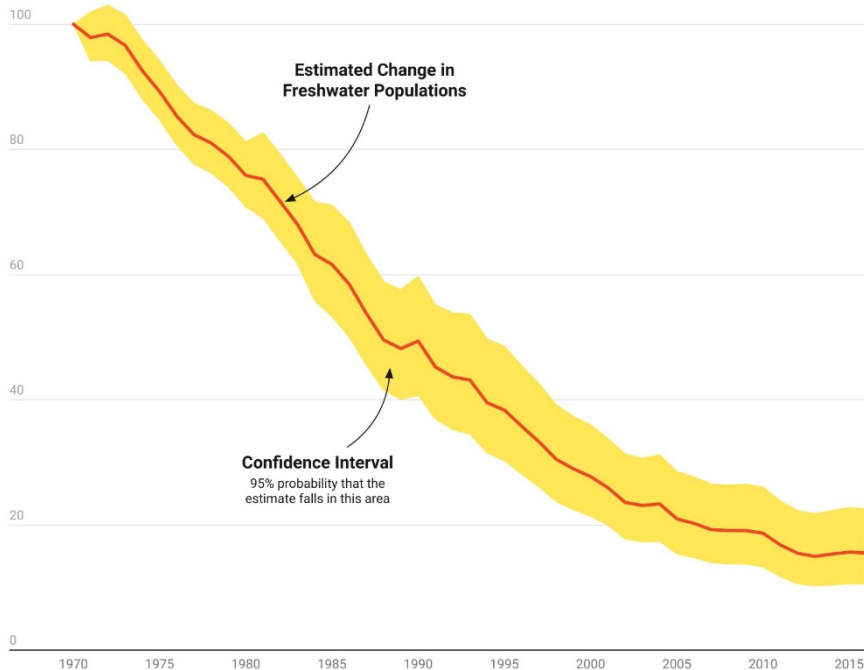
```
qqPlot(lm
```



Curves + Uncertainty

Humanity has wiped out 60% of animal populations since 1970 – and freshwater habitats are the worst hit with populations having collapsed by more than 80%

The Living Planet Index, produced for WWF by the Zoological Society of London, uses data on 16,704 populations of mammals, birds, fish, reptiles and amphibians to track the decline of wildlife. It underlines how the vast and growing consumption of food and resources by the global population is destroying the web of life upon which human society ultimately depends on.



#30DayChartChallenge 2021 | Day 8: Animals

Chart: Cédric Scherer • Source: World Wildlife Fund (WWF) and Zoological Society of London • Created with Datawrapper

Cederic Scherer used this graphic to argue about the decline of animal & freshwater populations.

Details aside, the confidence band gives visual evidence that the decline is systematic.

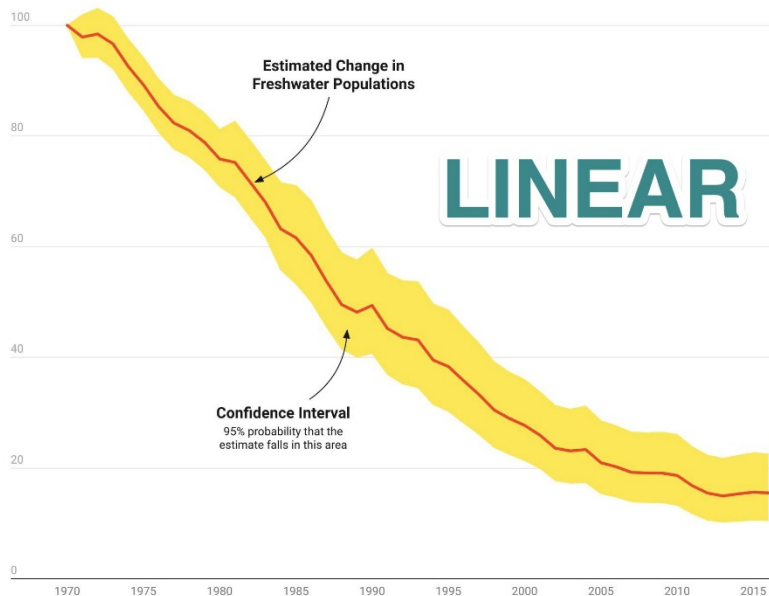
From: <https://twitter.com/CedScherer/status/1380211291466399744>

Curves + Uncertainty: Scale

Arguably, **percent** reduction in animal population should be viewed on a **log** scale. Transformed uncertainty intervals are here the logs of the Upper/Lower levels

Humanity has wiped out 60% of animal populations since 1970 – and freshwater habitats are the worst hit with populations having collapsed by more than 80%

The Living Planet Index, produced for WWF by the Zoological Society of London, uses data on 16,704 populations of mammals, birds, fish, reptiles and amphibians to track the decline of wildlife. It underlines how the vast and growing consumption of food and resources by the global population is destroying the web of life upon which human society ultimately depends on.

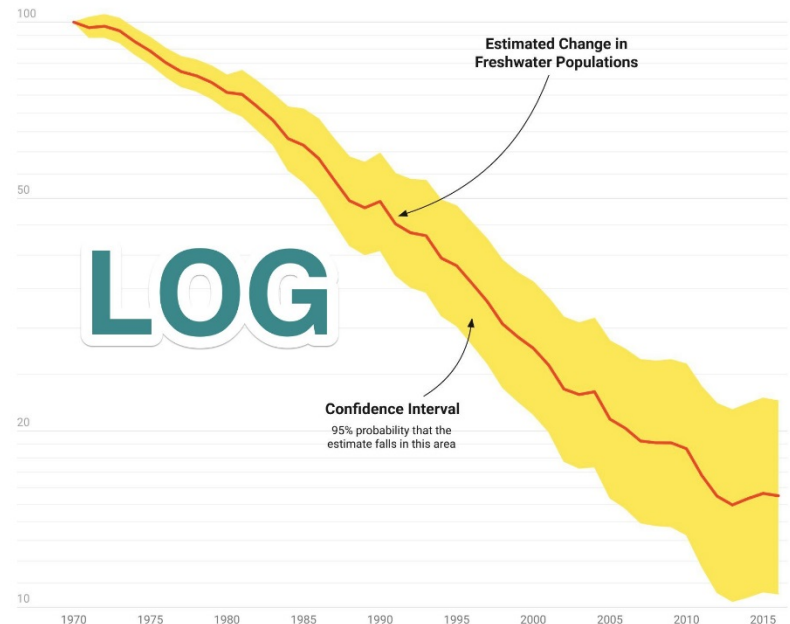


#30DayChartChallenge 2021 | Day 8: Animals

Chart: Cédric Scherer • Source: World Wildlife Fund (WWF) and Zoological Society of London • Created with Datawrapper

Humanity has wiped out 60% of animal populations since 1970 – and freshwater habitats are the worst hit with populations having collapsed by more than 80%

The Living Planet Index, produced for WWF by the Zoological Society of London, uses data on 16,704 populations of mammals, birds, fish, reptiles and amphibians to track the decline of wildlife. It underlines how the vast and growing consumption of food and resources by the global population is destroying the web of life upon which human society ultimately depends on.

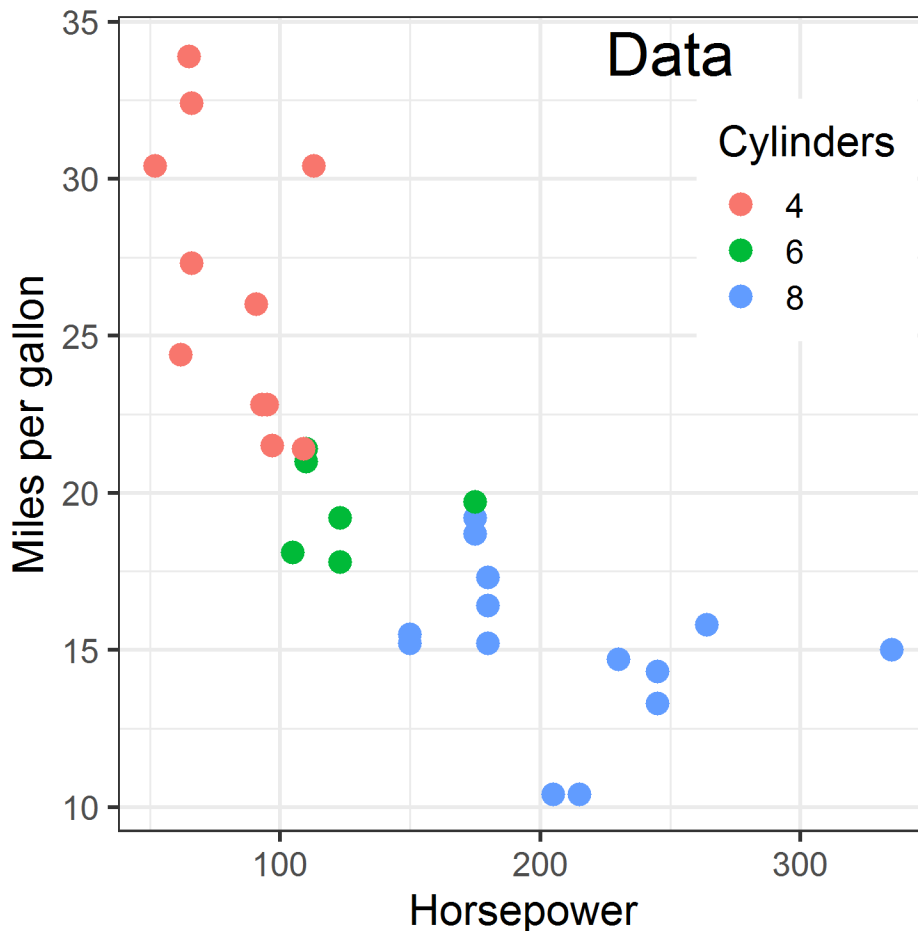


#30DayChartChallenge 2021 | Day 8: Animals

Chart: Cédric Scherer • Source: World Wildlife Fund (WWF) and Zoological Society of London • Created with Datawrapper

Fitted curves

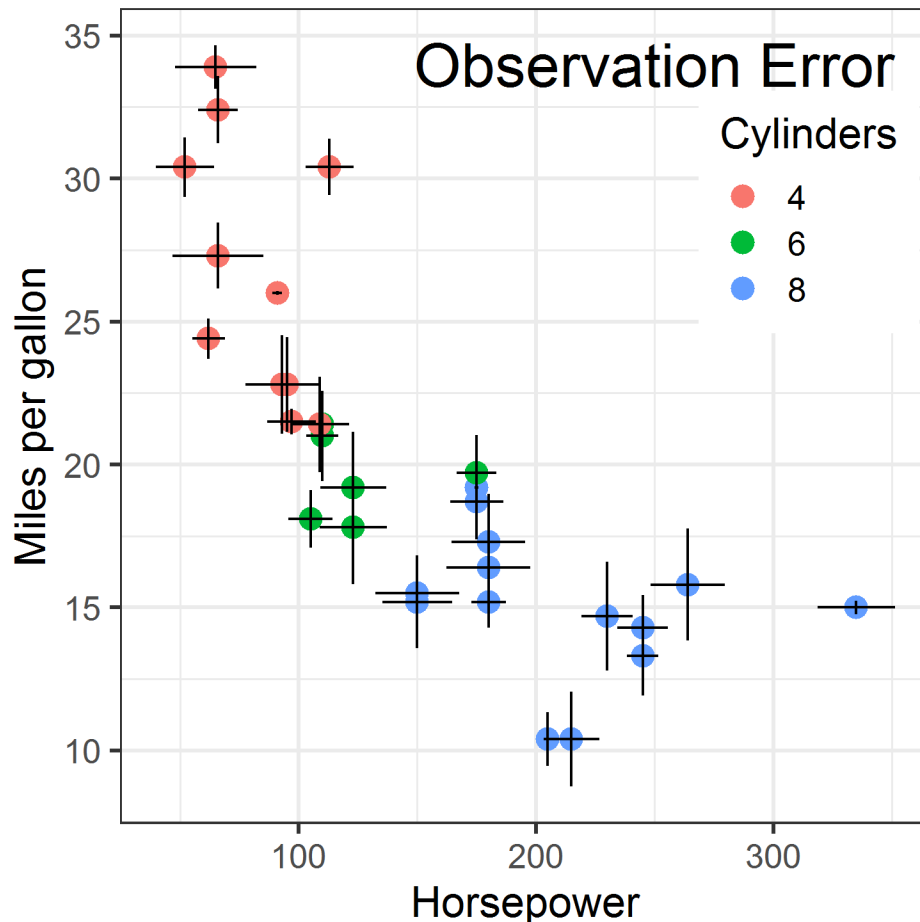
Data on gas mileage of *Motor Trend* 1974 cars



Sources of uncertainty:

- Observations: **measurement error** in MPG and/or HP?
- **Model form**: Linear? Quadratic? Interaction with cylinders
- **Model fit uncertainty**: normal theory CIs? Bootstrap? Bayesian?

Measurement uncertainty



Sometimes, we can quantify the uncertainty (“error”) in values of x and or y .

e.g., each point is the average of $n > 1$ cars.

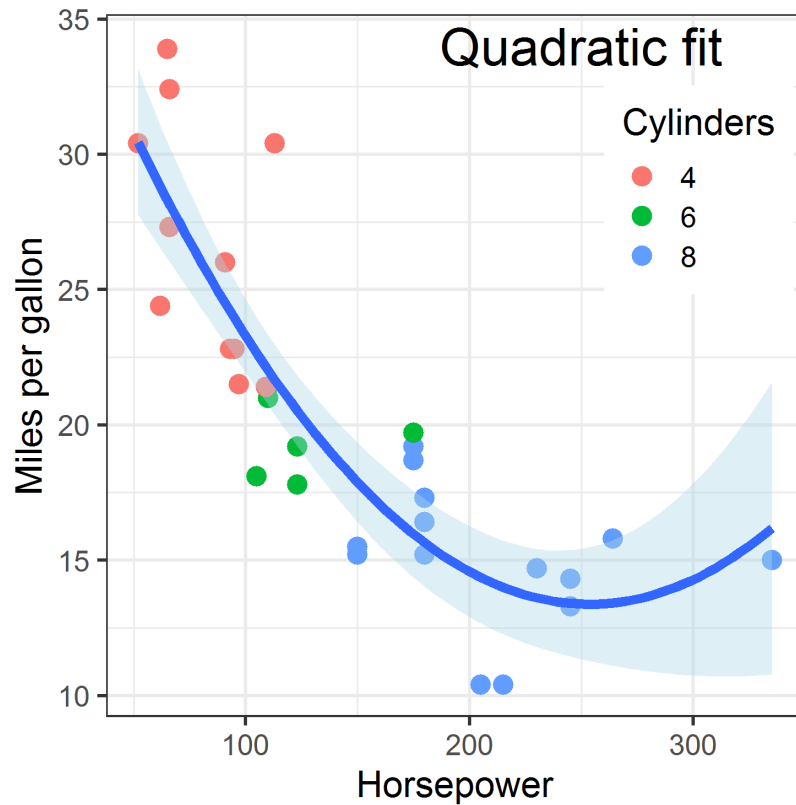
Fitted models allow for errors in y :
 $y = f(x) + \text{error}$
and find estimates to minimize error

Most fitted models assume x is measured w/o error.

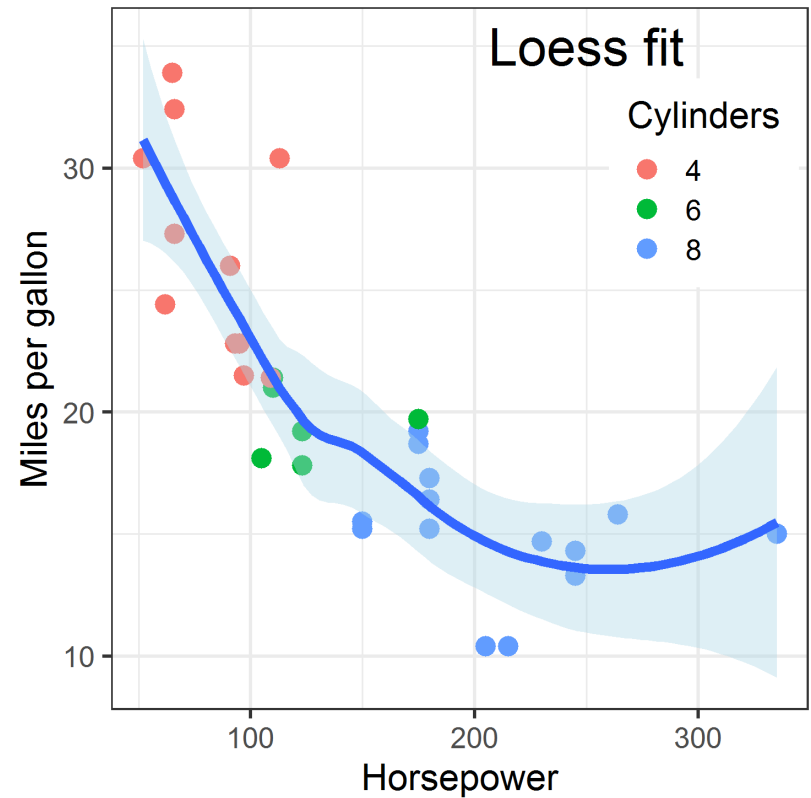
Big problem if $\text{error} \sim f(x, \text{other } x\text{s})$

Model forms: nonlinear fits

When a relation is clearly non-linear, we can fit alternative models.
The CI bands tell us where the data is too thin to rely on the predicted value.

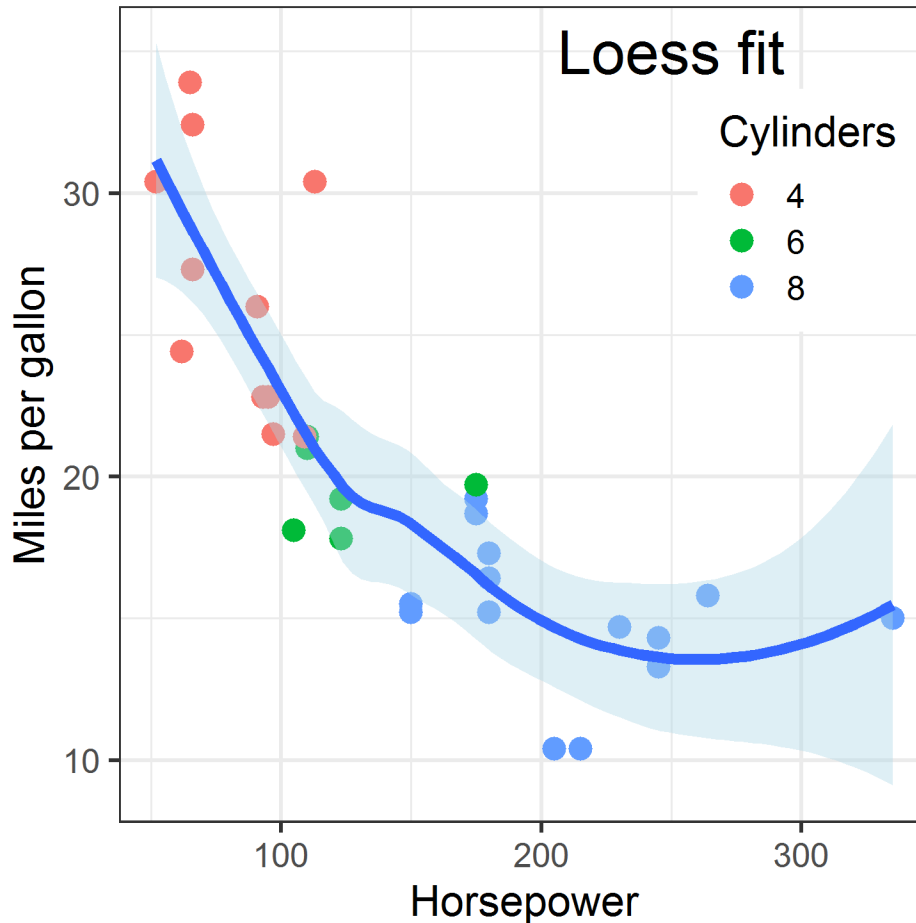


```
p1 + geom_smooth(method = lm, formula = y~poly(x,2), ...)
```



```
p1 + geom_smooth(method = loess, formula = y~x, ...)
```

Fitted curves: smoothers

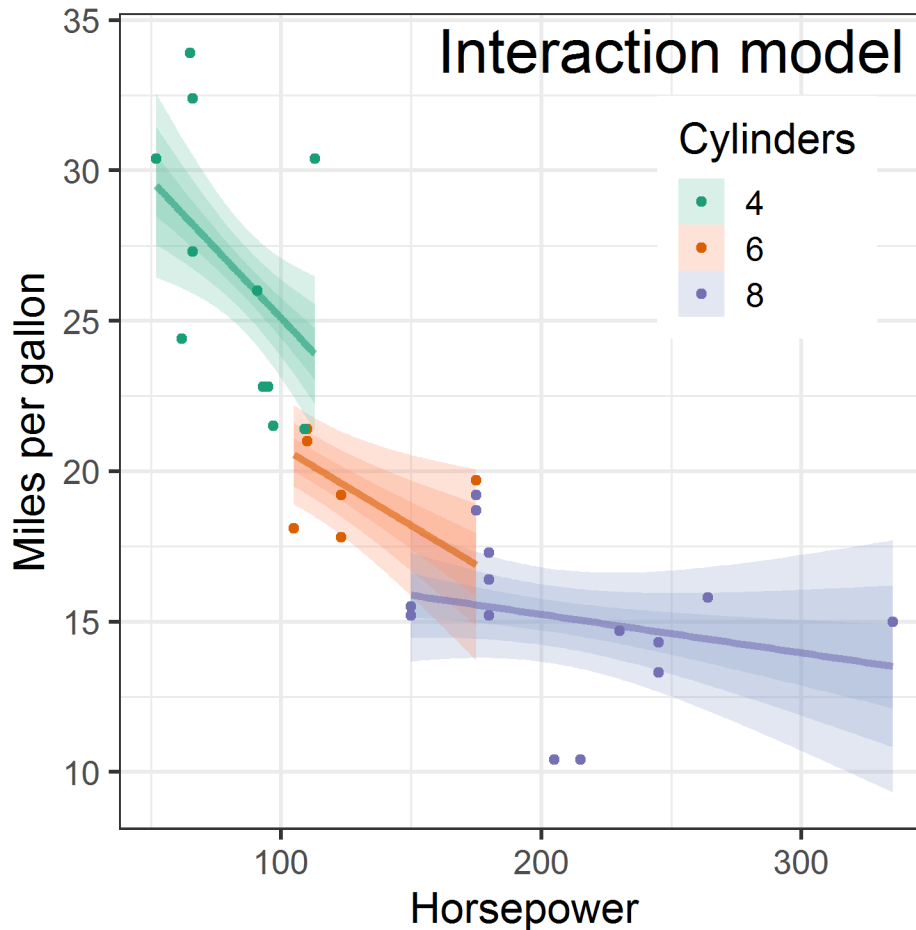


In each case, the confidence band gives visual evidence for uncertainty of the predicted values.

But, uncertainty may be expressed differently.

- a formula for std. error based on normal/large sample theory
- envelope of (normal) simulations
- Bayesian predictive distribution

Interaction models

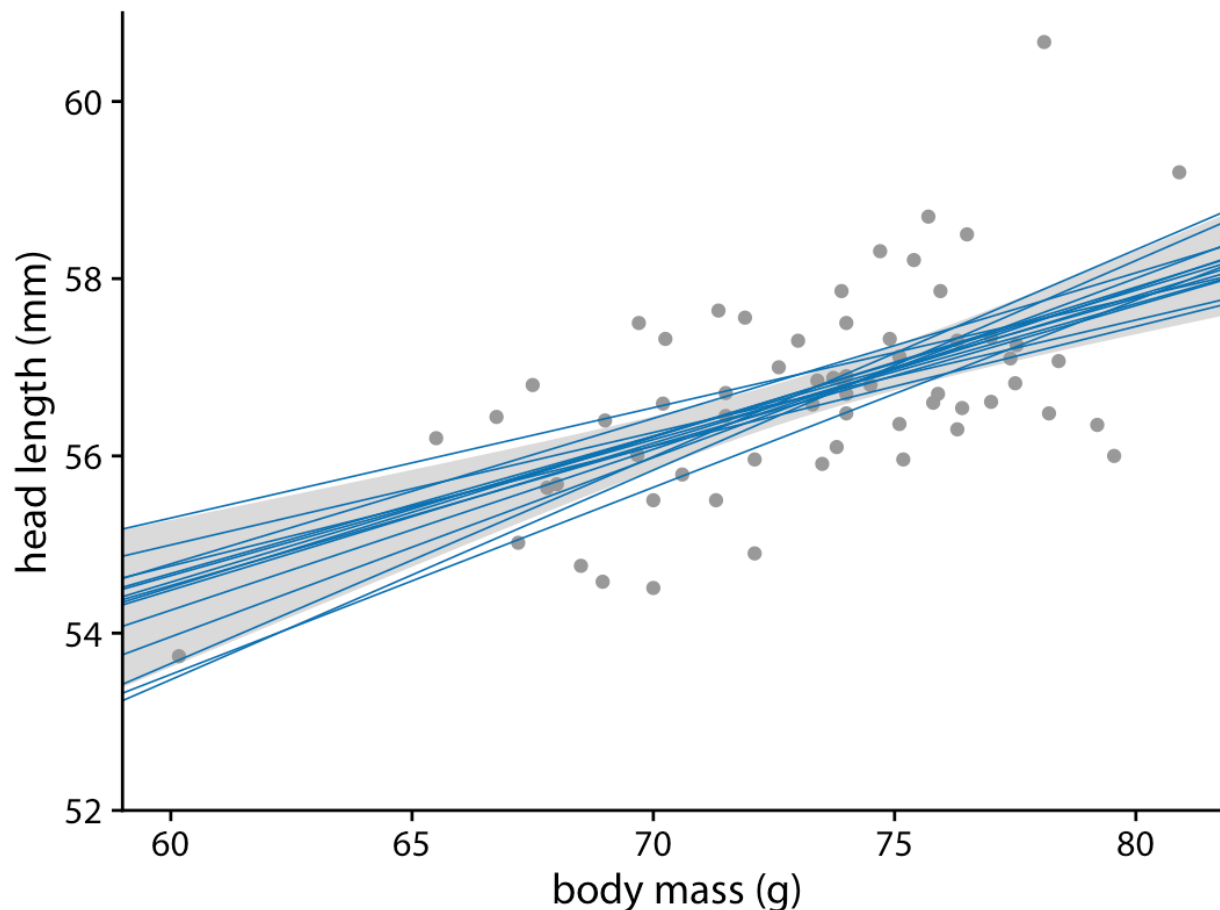


The non-linear relation between hp & mpg can (arguably) be better explained by a model that allows different slopes for 4, 6, 8 cylinders.

The graph shows normal theory CIs at 95%, 90%, and 80% for each cylinder level

Simulations to convey uncertainty

Simulating fits from the data (e.g., bootstrap, Bayesian estimation) shows the variability. Doesn't rely on classical, normal theory.

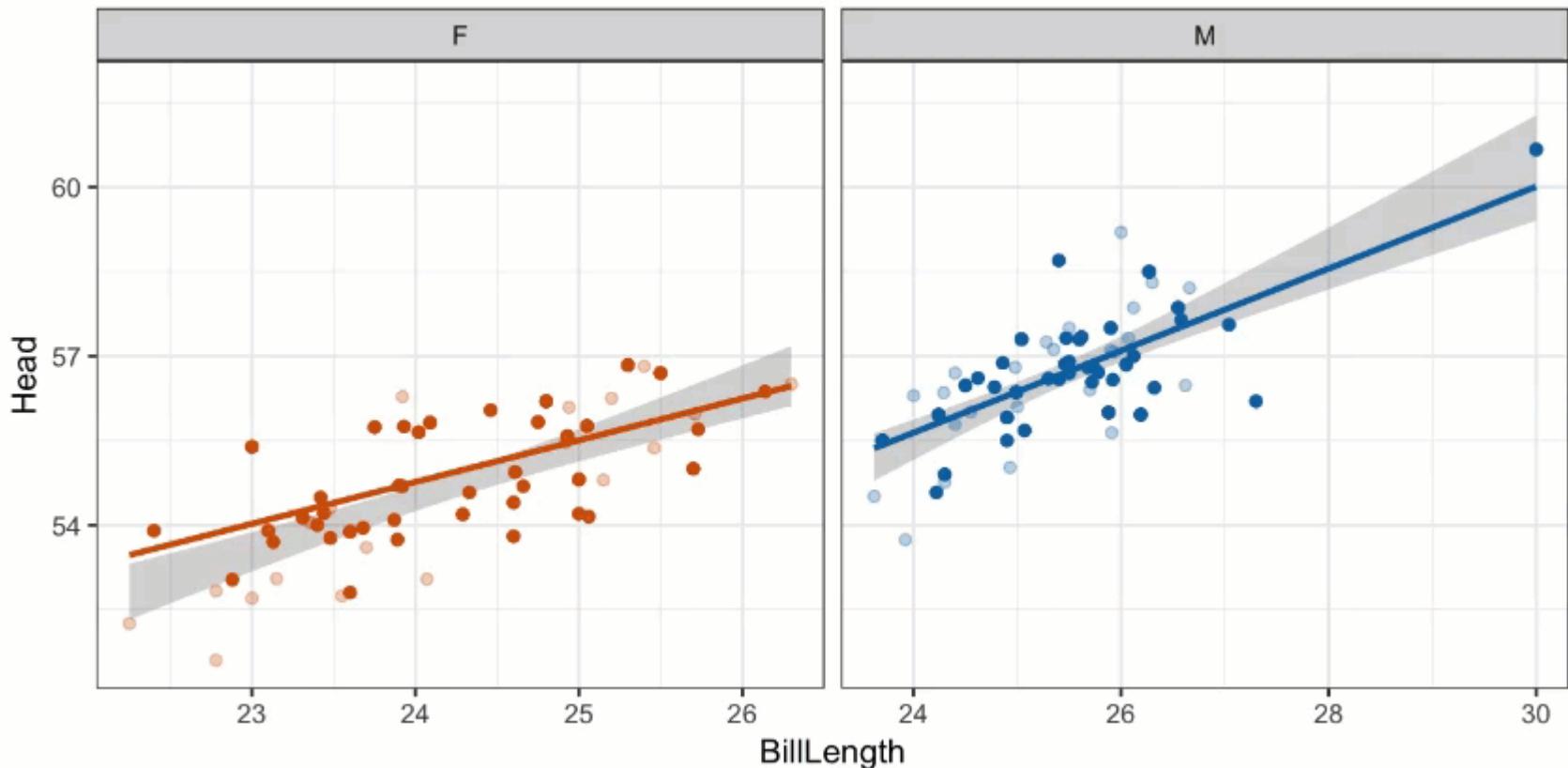


Animation to understand uncertainty

All assessments of uncertainty rely on a comparison: data vs. could have been

- Sampling distributions, simulations, Bayesian posterior distributions, ...

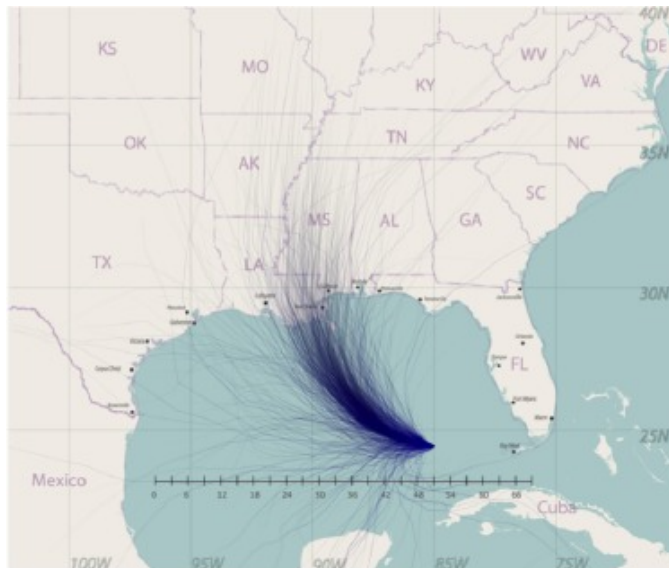
Sometimes useful to appreciate the variability with animated graphics



Geographic uncertainty

Predicting the path of hurricanes:

- Given what we can measure today (location, wind speed, direction, ...) where is **this** hurricane likely to be in 1 day, 3 days, 5 days?
- Most forecasts are based on an **ensemble of predictions**, representing the uncertainty in initial conditions, model physics, ...
- Often this is represented as a “cone of uncertainty”



(a) Storm path ensemble



(b) Uncertainty cone.

What is the Cone of Uncertainty?

As seen on TV:

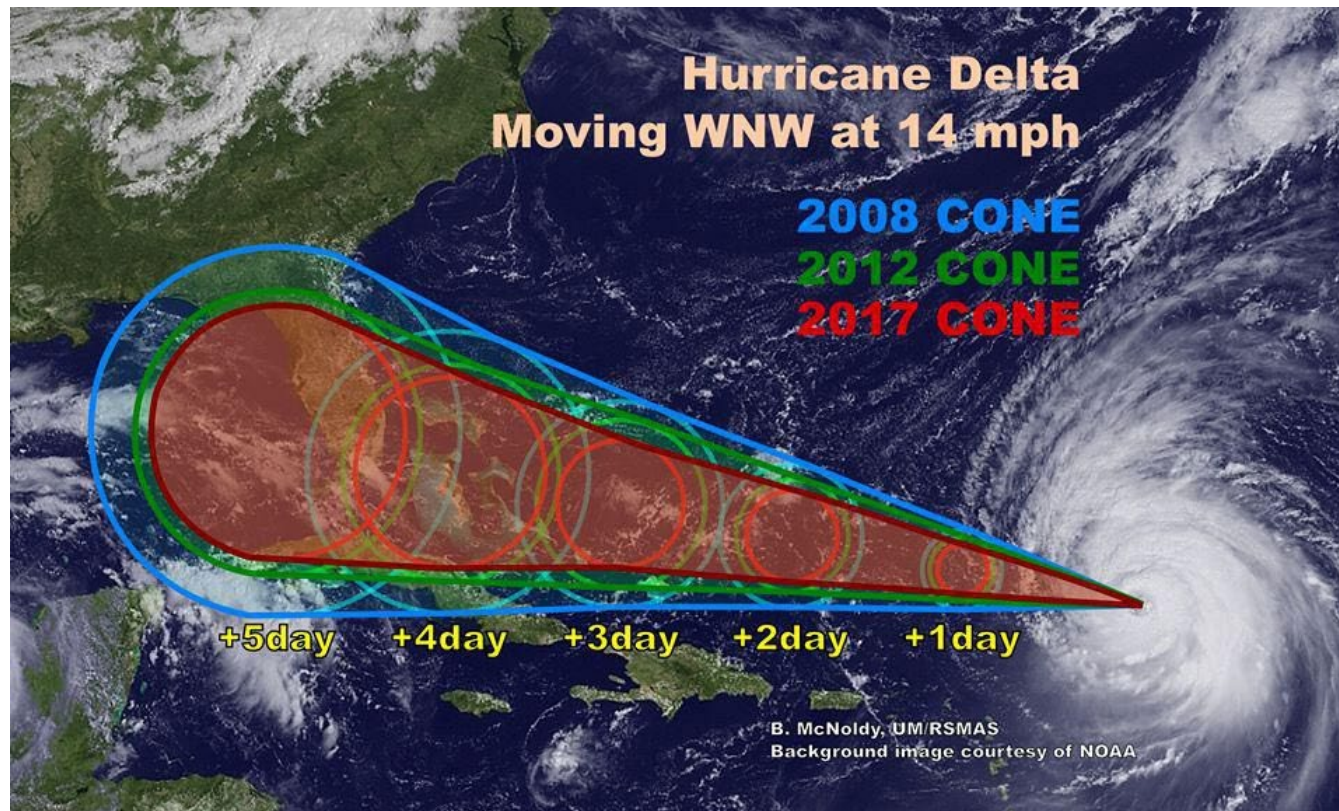
- The center is meant to track the average prediction, either over models or history
- The cone size generally represents some “2/3 confidence interval”
- Does this mean I am safe if I lived in Tallahassee FL ★ in 2005? 2020?



The Incredible Shrinking Cone

Changes in presumed accuracy are often shown as below

- The cone represents the probable track of the center of a tropical cyclone, formed by enclosing the area swept out by a set of circles along the forecast track (at 12, 24, 36 hours, etc).
- The size of each circle is set so that two-thirds of historical official forecast errors over a 5-year sample fall within the circle.



Sharpiegate

In Sept. 2019, Donald Trump went live with “extrapolated” predictions of the path of Hurricane Dorian.

- He had earlier predicted it would hit Alabama & Georgia.
- Let it be said, let it be written (with a sharpie)

