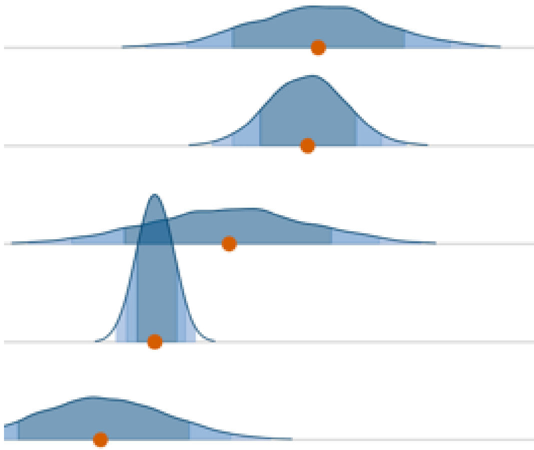
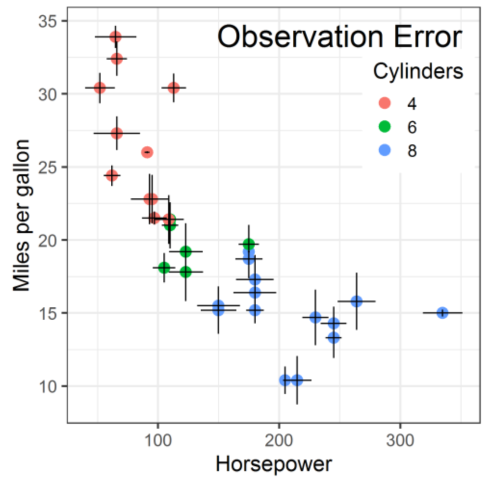




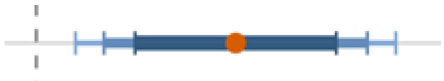
Visualizing Uncertainty

Michael Friendly

Psych 6135

https://friendly.github.io/6135/

Topics

- Uncertainty in statistics & visualization
- Visualizing distributions
- "Error bars" 
- Bayesian uncertainty
- 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: data, models, graphics

- Uncertainty is fundamental to data analysis & models
  - data: IQR, std dev., std error, … (variation)
  - assumptions: we assume some distribution for errors, e.g., $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, 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

P-values, significance & uncertainty

ASA President's Task Force on Statistical Significance

- "Much of the controversy surrounding statistical significance can be dispelled by better understanding of uncertainty, variability, multiplicity & replicability"
- "Different measures of uncertainty can complement each other; no single measure serves all purposes"
- "Controlling and accounting for uncertainty begins with the design of the study"
- "The theoretical basis of statistical science offers general strategies for dealing with uncertainty"
  - **Frequentist** approach: p-values, confidence intervals & prediction intervals
  - **Bayesian** approach: Bayes factors, posterior probability distributions, credible intervals

Model fits as uncertainties

Table 2 from Stevens (2006): Determinants of Authoritarian Aggression

Variable	Coefficient (Standard Error)
Constant	.41 (.93)
Countries	
Argentina	1.31 (.33)### B,M
Chile	.93 (.32)### B,M
Colombia	1.46 (.32) ### B,M
Mexico	.07 (.32)A,CH,CO,V
Venezuela	$\boxed{\text{.96 (.37)\#\# B,M}}$
Threat	
Retrospective egocentric economic perceptions	.20 (.13)
Prospective egocentric economic perceptions	.22 (.12)#
Retrospective sociotropic economic perceptions	-.21 (.12)#
Prospective sociotropic economic perceptions	-.32 (.12)##
Ideological Distance from president	
Ideology	
Ideology	.23 (.07) ###
Individual Differences	
Age	.00 (.01)
Female	-.03 (.21)
Education	.13 (.14)
Academic Sector	.15 (.29)
Business Sector	.31 (.25)
Government Sector	-.10 (.27)
R^2	.15
Adjusted R^2	.12
n	500

###p < .01, ##p < .05, #p < .10 (two-tailed)

A Coefficient is significantly different from Argentina's at p < .05;

B Coefficient is significantly different from Brazil's at p < .05;

CH Coefficient is significantly different from Chile's at p < .05;

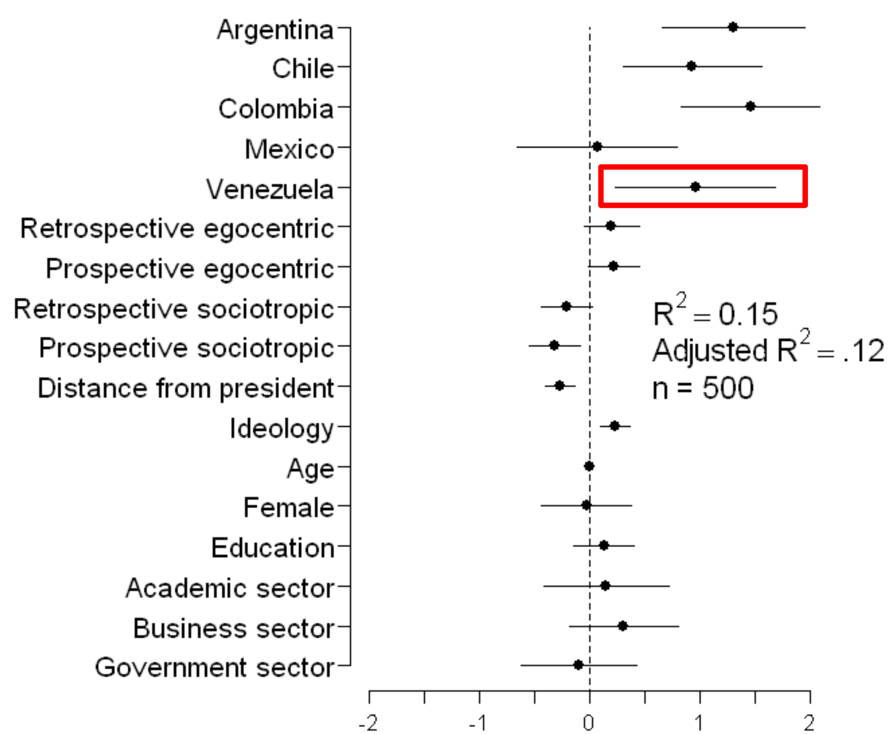
CO Coefficient is significantly different from Colombia's at p < .05;

M Coefficient is significantly different from Mexico's at p < .05;

V Coefficient is significantly different from Venzeluela's at p < .05

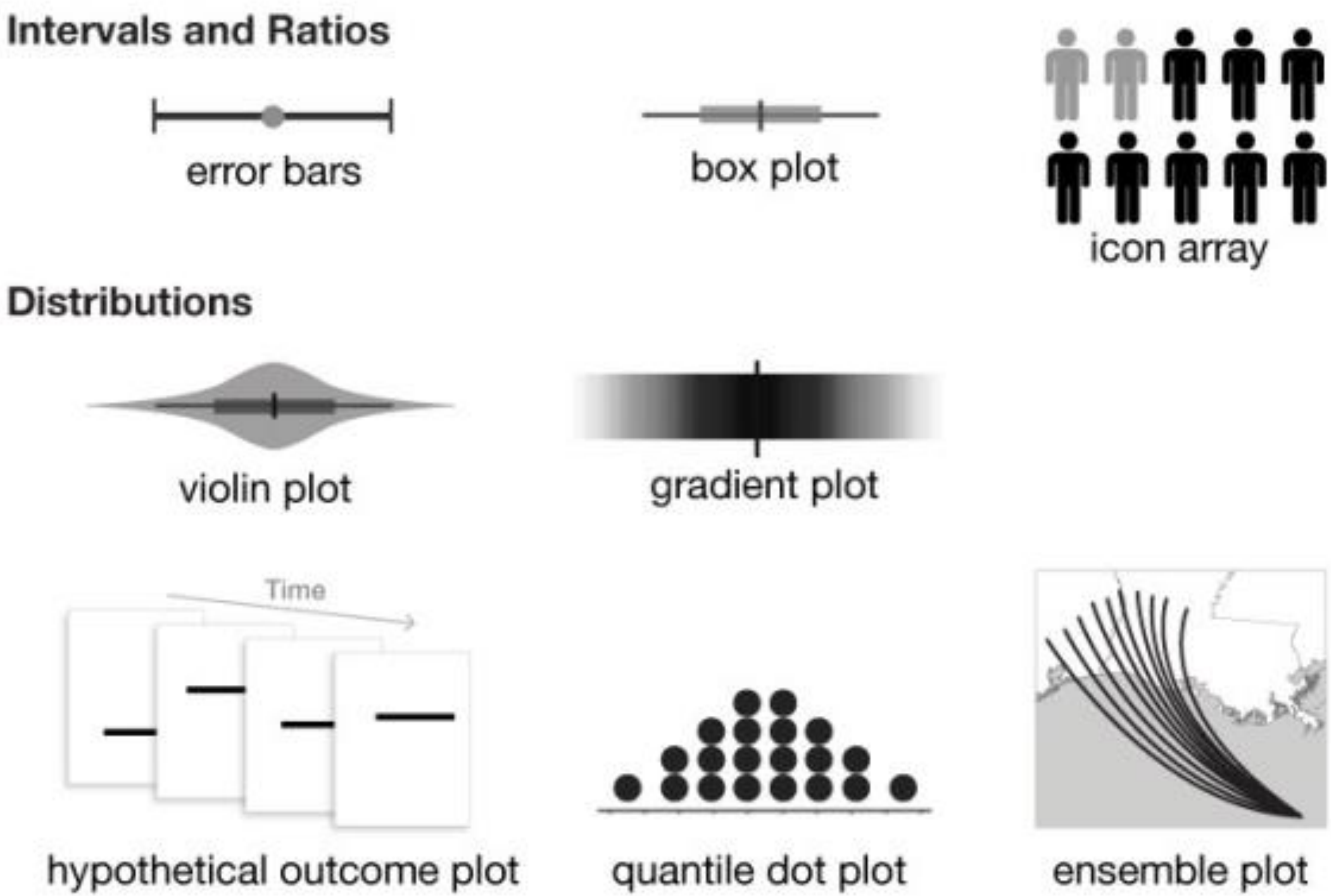
Coefficients** & std errors express uncertainty

Can we do better?



Source: tables2graphs.com

Graphical annotations for uncertainty



From: Padilla, Kay & Hullman (2021), *Uncertainty Visualization*, DOI: 10.1002/9781118445112.stat08296

Visualizing distributions

- The basics:
  - Histograms
  - Density plots
  - Boxplots
- Doing better:
  - violin plots
  - rainclouds
  - {ggdist}: data, distribution, interval

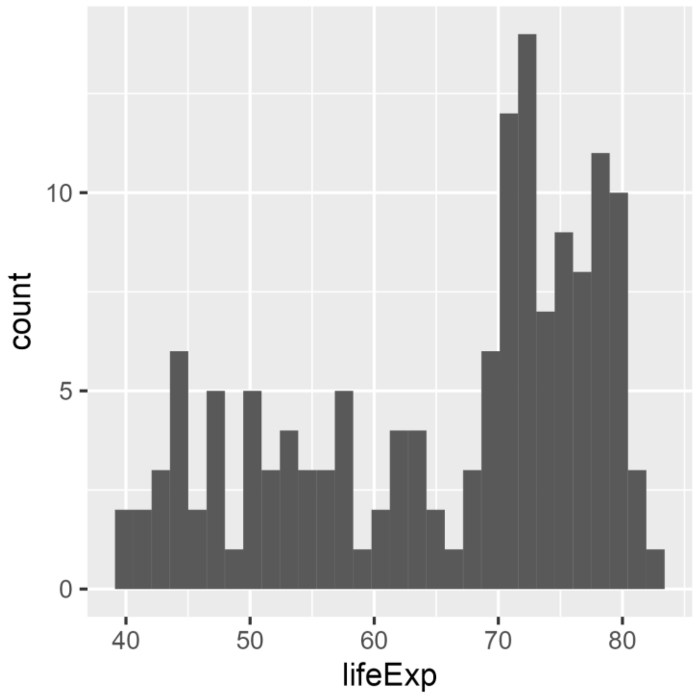
Histograms

- Perhaps the simplest display
  - divide the data into bins
  - bar plot of the frequencies: length ~ frequency

```
library(gapminder)
gapminder_2002 <- gapminder %>%
  filter(year == 2002)

ggplot(gapminder_2002,
       aes(x = lifeExp)) +
  geom_histogram()
```

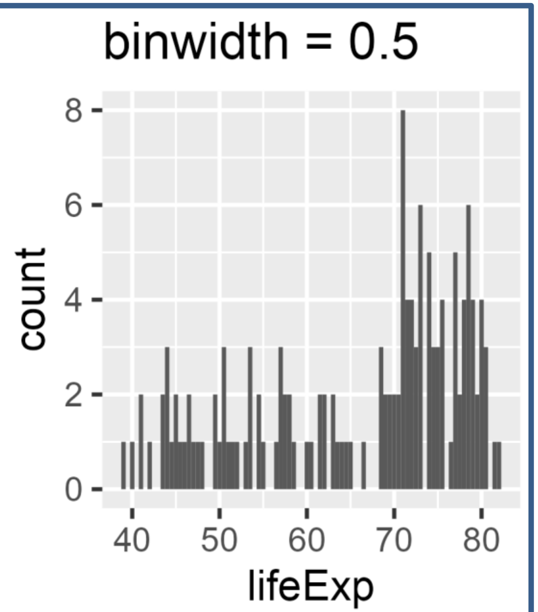
```
`stat_bin()` using `bins = 30`. Pick better value
with `binwidth`.
```



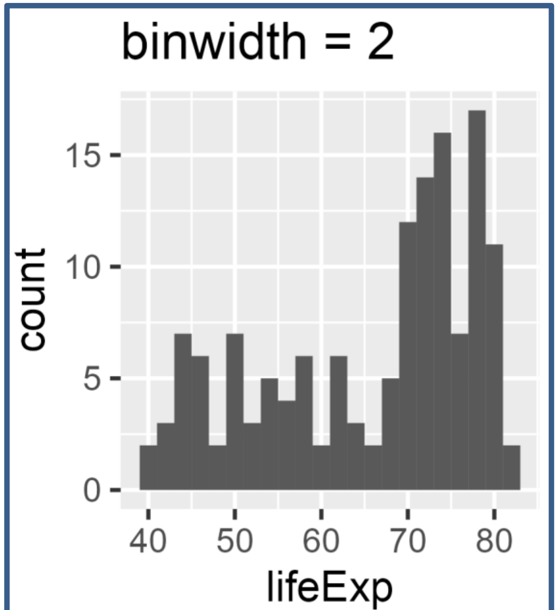
Histograms: bin width

- Explicitly selecting the binwidth shows:
  - the Goldilocks principle
  - the default is often OK, but optimal "best" is harder to define

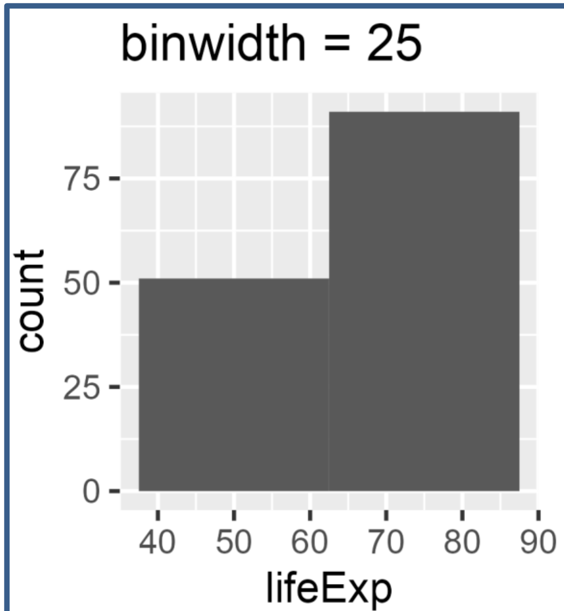
```
ggplot(gapminder_2002, aes(x = lifeExp)) + geom_histogram(binwidth = )
```



too small



about right

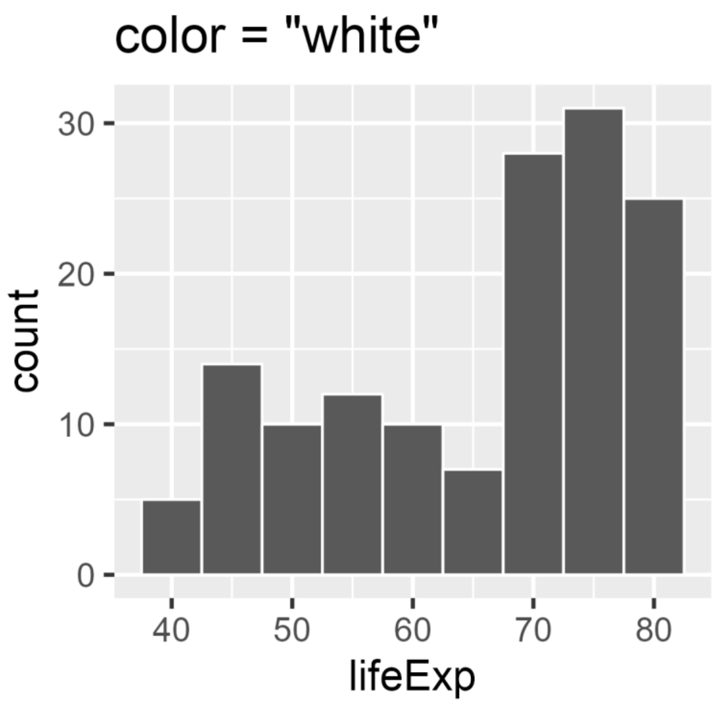


too large

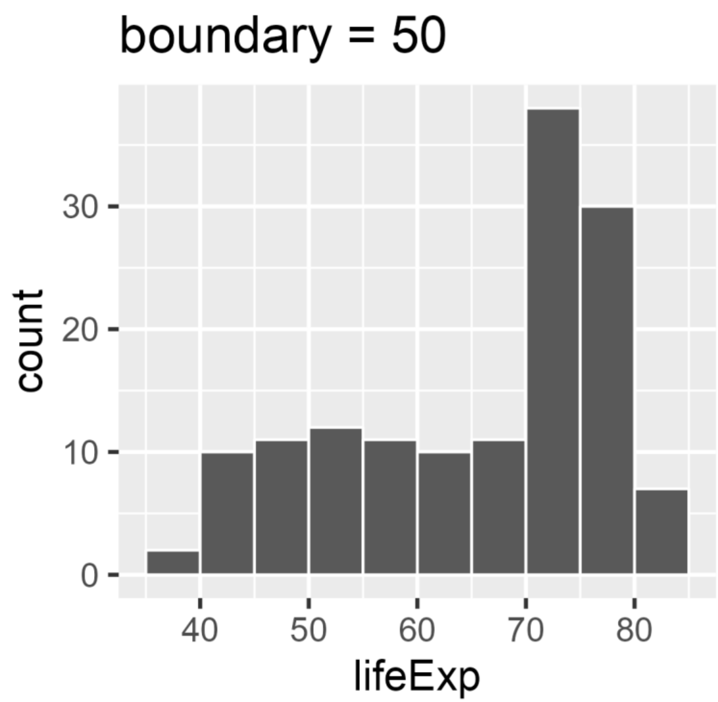
Histograms: other properties

- Pay attention to graphic details
  - border color to make bars distinct
  - set bar boundaries: to edges? – it can make a difference

geom_histogram(..., color = “white”)

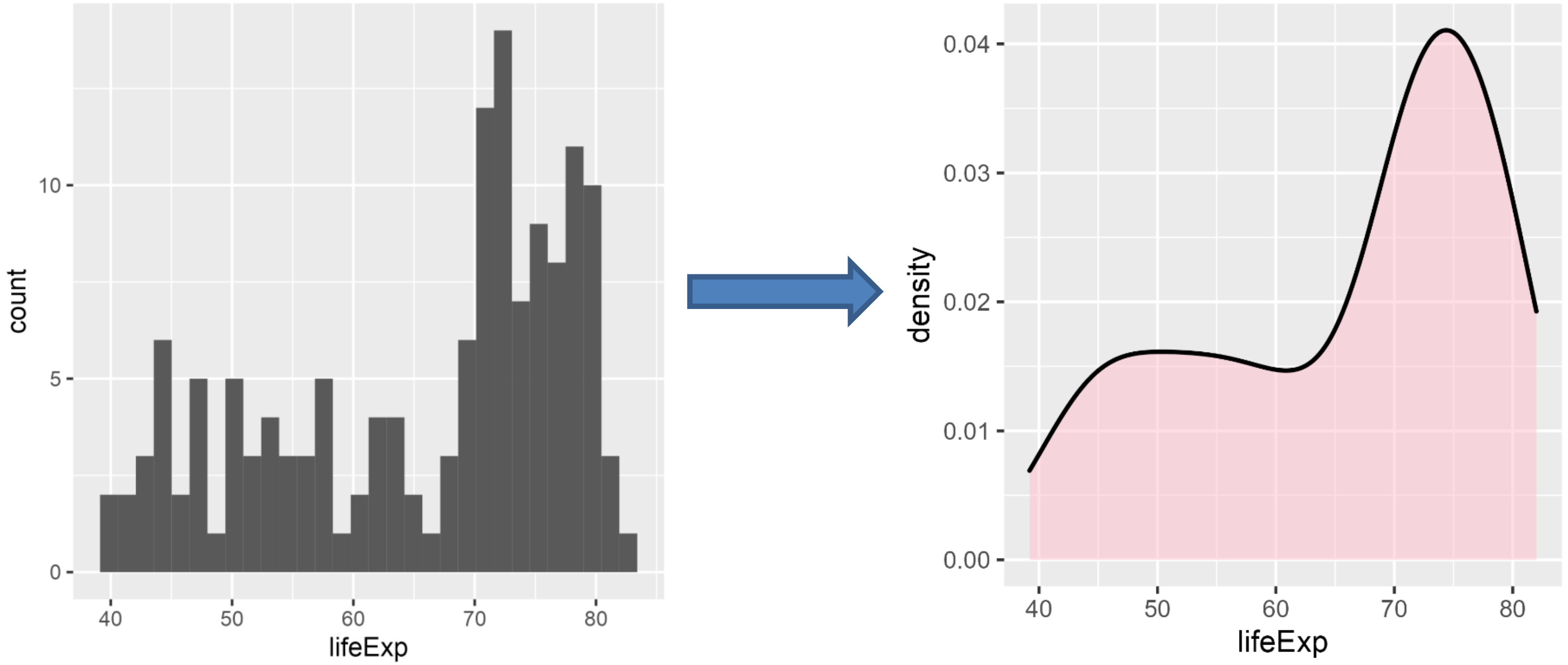


geom_histogram(..., boundary=50)



Density plots

- Basic idea: Smooth the distribution to avoid artifacts of discrete bins and bin centers
    - Uses a "kernel", e.g, gaussian, averaged over a moving window

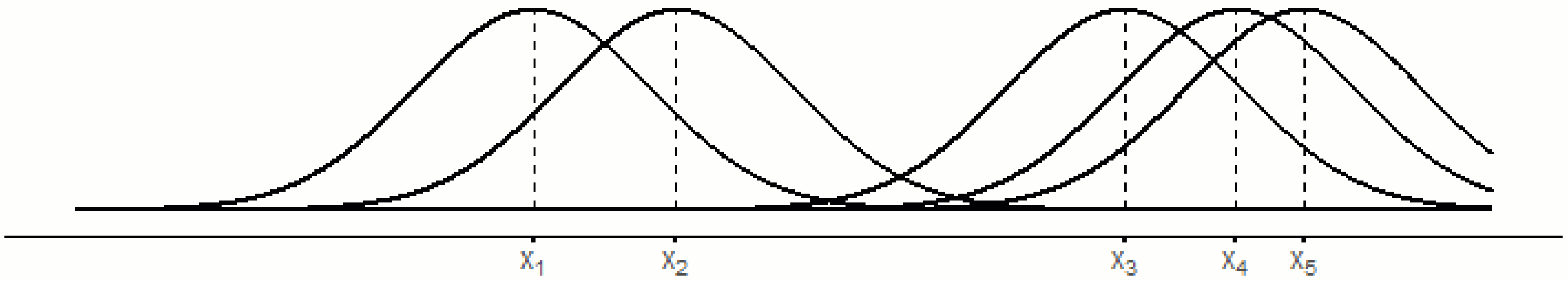


Kernel density estimation

Imagine a distribution of potential density centered at each X_i, w/ sd = h (bandwidth)

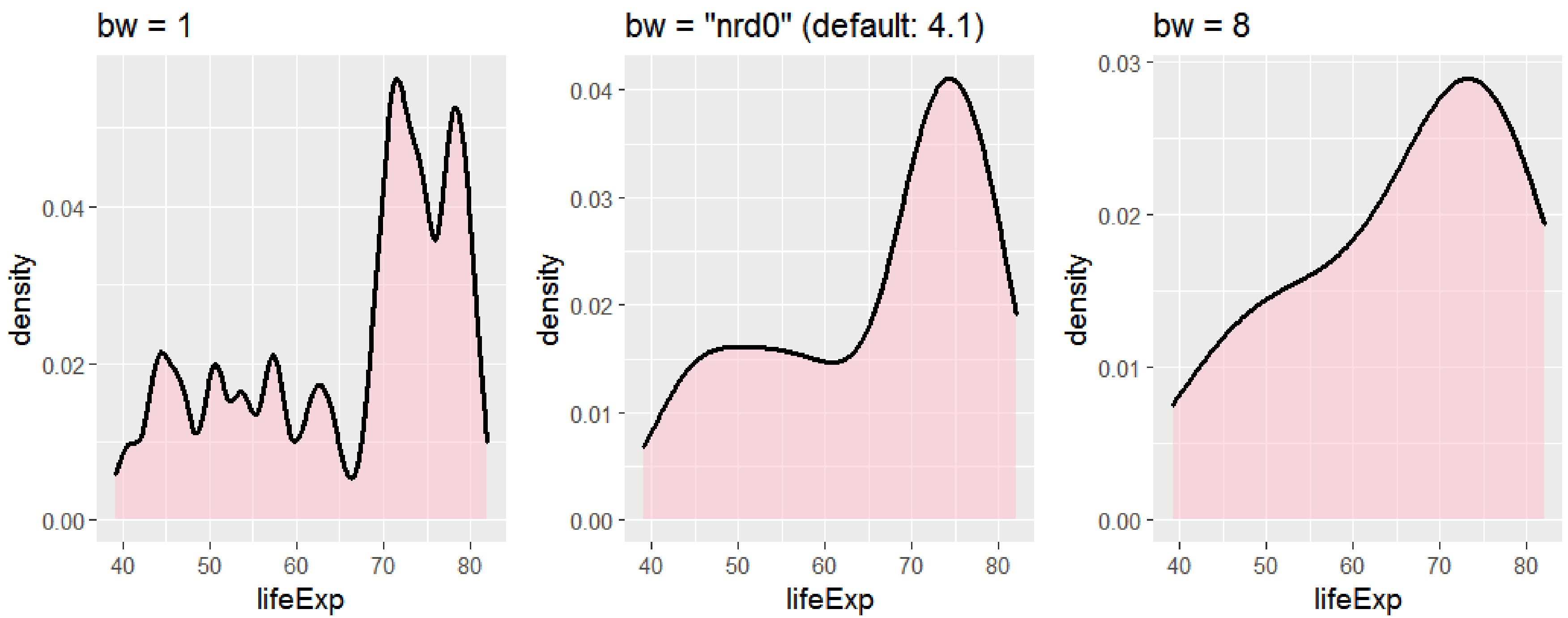
$$x \sim \mathcal{N}(\mu=X_i, \sigma=h)$$

A moving window sweeps across, averaging the density for all observations



Density plots: bandwidth

- The result depends on the width of the moving window – bandwidth
  - The default calculation is usually reasonable, but beware of weird data

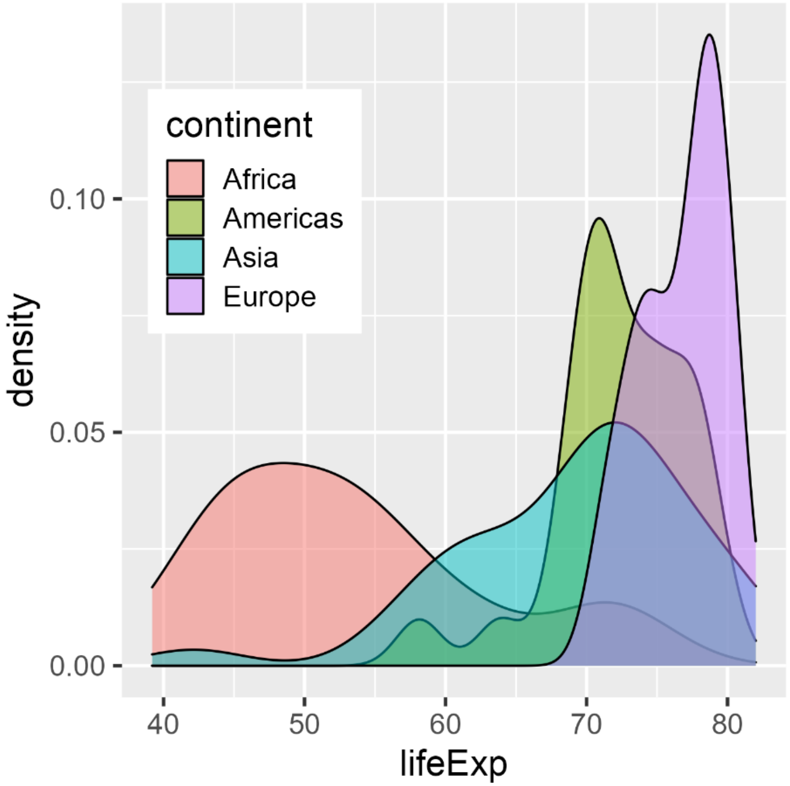


Comparing groups

For multiple groups, using the fill aesthetic → overlaid curves -- is a decent start
But even with transparency it may be hard to see the separate curves

```
gap_2002c <-
  gapminder_2002 %>%
  filter(continent != "Oceania")

ggplot(gap_2002c,
      aes(x = lifeExp,
             fill = continent)) +
  geom_density(alpha = 0.5) +
  theme(legend.position = c(.2, .7))
```

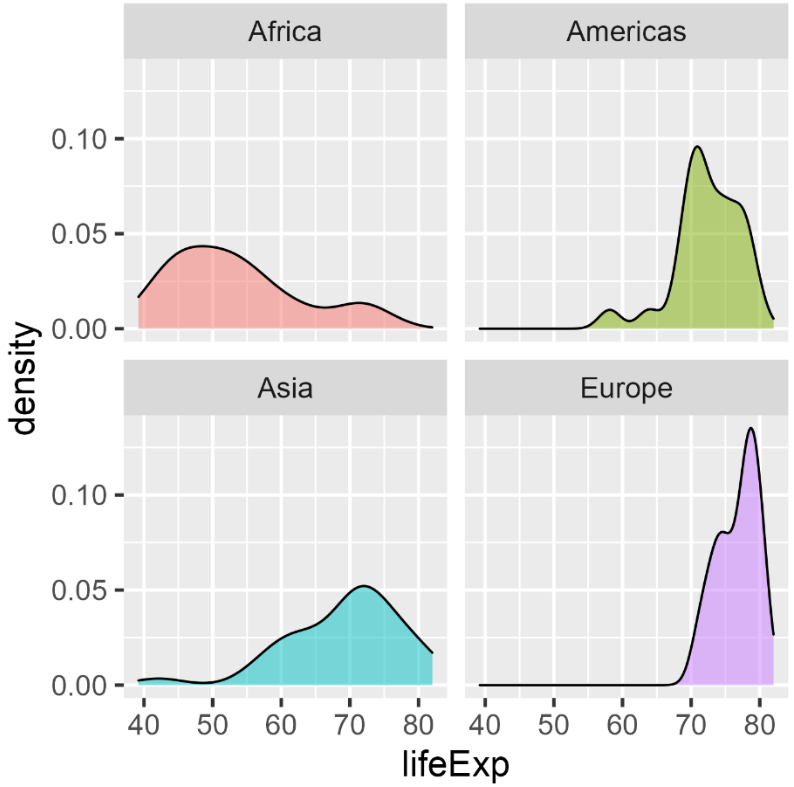


NB: ggplot picks a joint bandwidth, here: 2.52

Comparing groups: Facets

Faceting solves the overlap problem, but the eye has to move from panel to panel to make comparisons.

```
ggplot(gap_2002c,
       aes(x = lifeExp,
           fill = continent)) +
  geom_density(alpha = 0.5) +
  facet_wrap(~ continent) +
  theme(legend.position = "none")
```

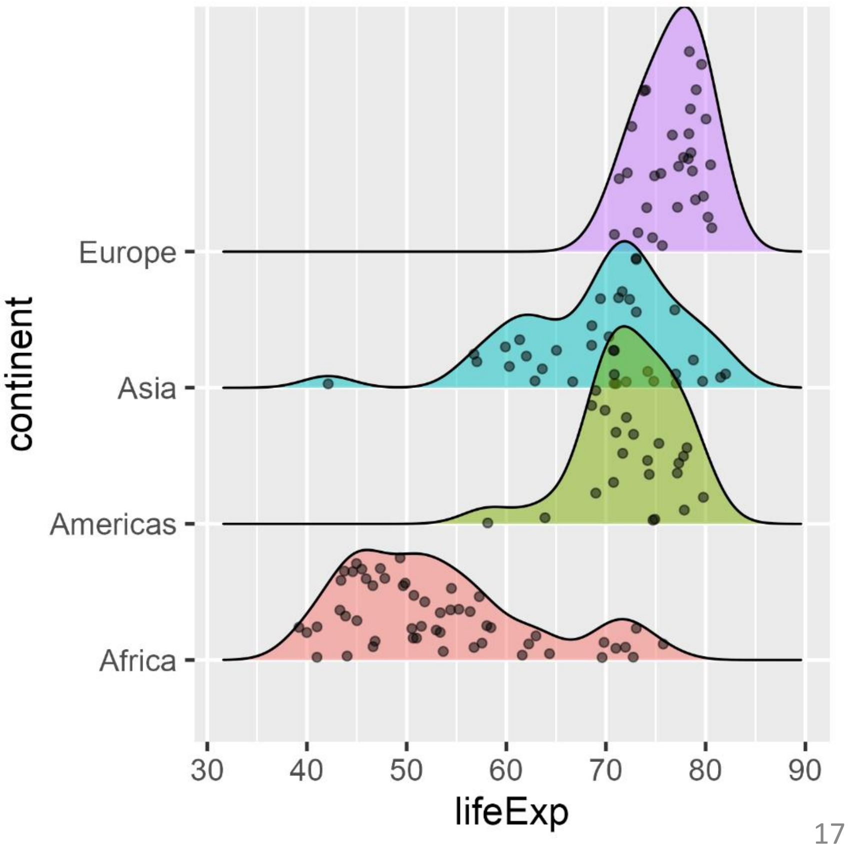


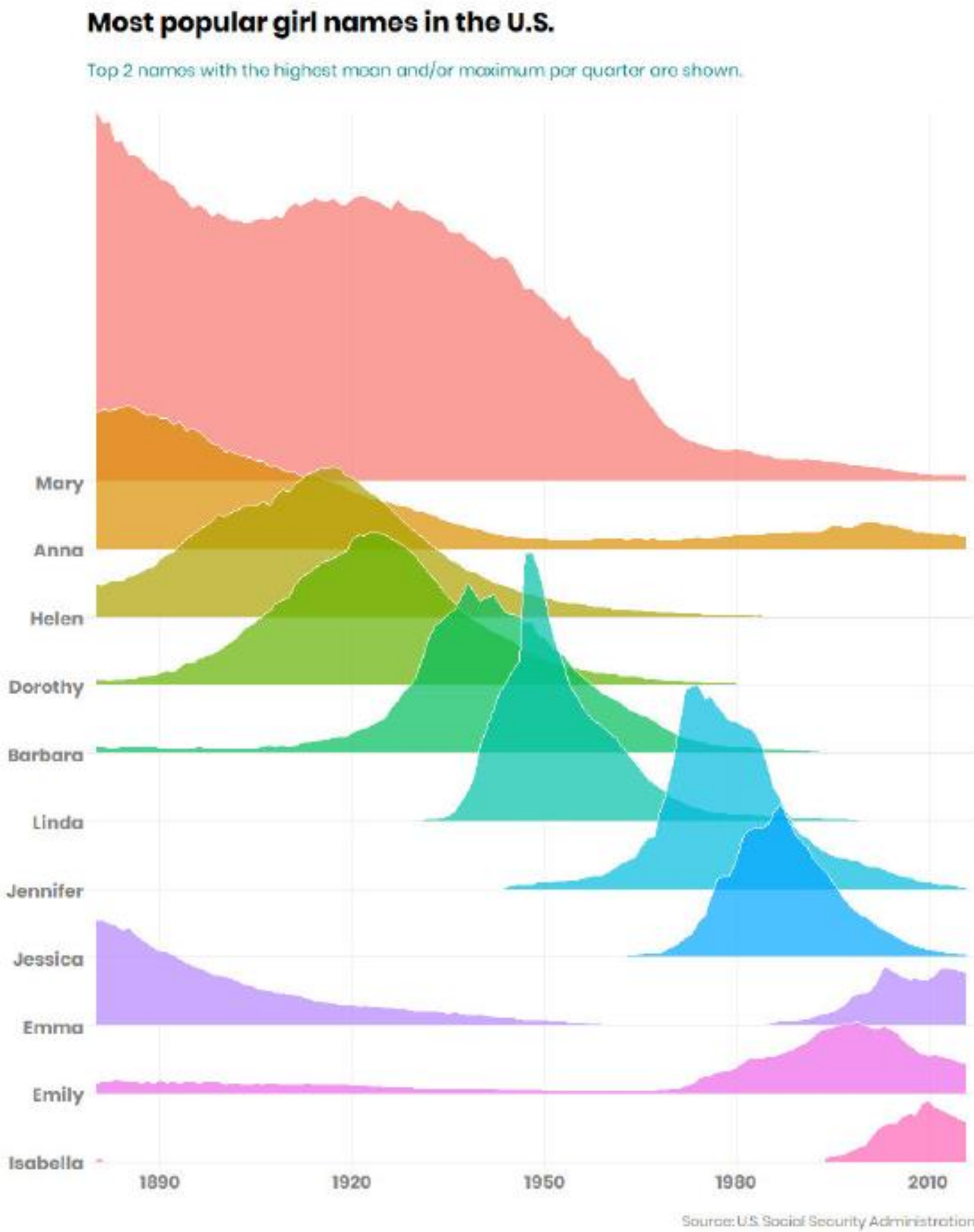
{ggridges}: Ridgeline plots

Ridgeline plots are partially overlapping density plots, suggesting a mountain range.

- Useful for comparing distributions over time or circumstances
- Adding jittered points helps to show where the data are

```
library(ggridges)
ggplot(gap_2002c,
      aes(x=lifeExp,
          y=continent,
          fill=continent)) +
  geom_density_ridges(
      alpha = 0.5,
      jittered_points=TRUE) +
  theme(legend.position = "none")
```





Ridgeline plots are particularly effective with more than a few categories, and when the distributions differ in shape as well as central location

Which names stand out from the rest?

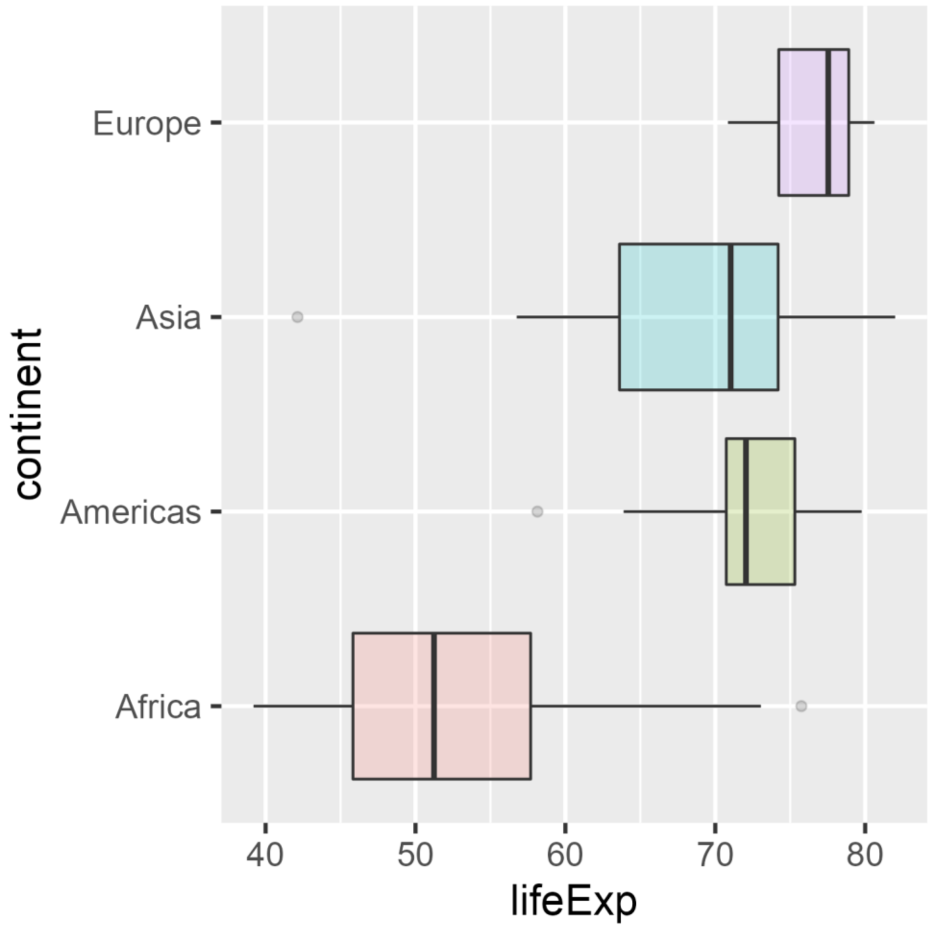
What is the role of color here?

Note the subtle use of white to outline each distribution

Boxplots

Boxplots give a more schematic summary of a dataset–
median, quartiles, whiskers & outliers

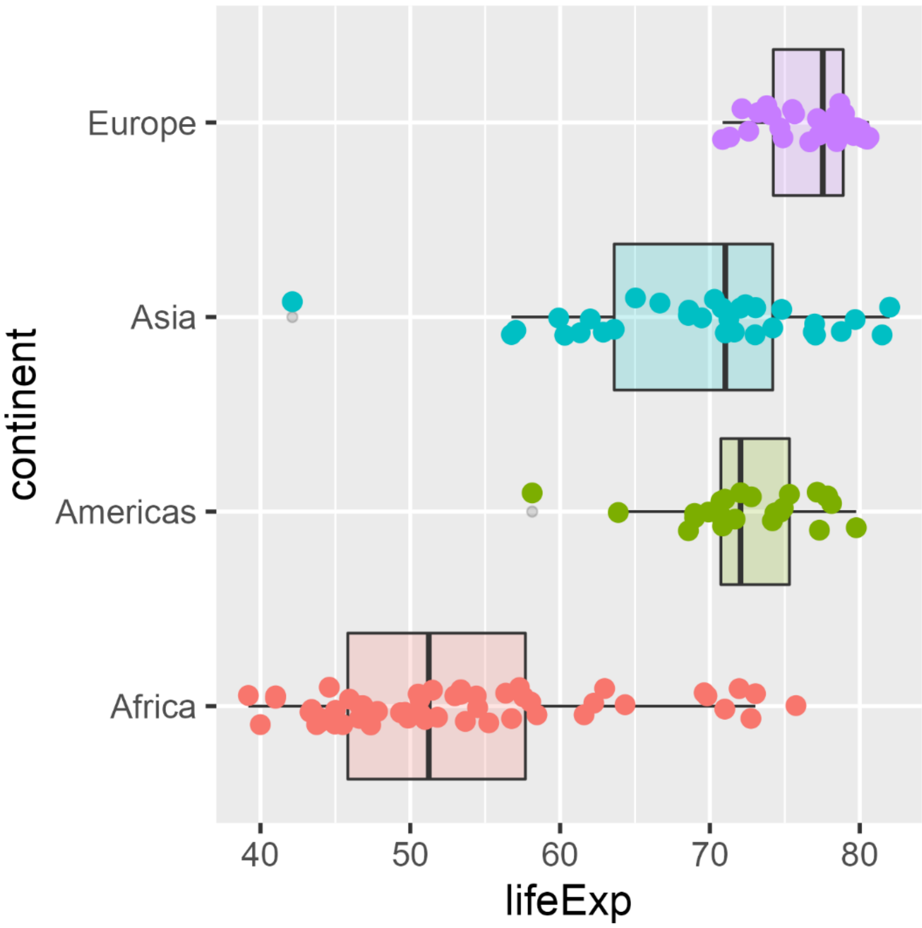
```
ggplot(gap_2002c,
      aes(x=lifeExp,
          y=continent,
          fill=continent)) +
  geom_boxplot(alpha = 0.2) +
  theme(legend.position = "none")
```



Boxplots

But perhaps too schematic– it sometimes helps to see the data as jittered points

```
ggplot(gap_2002c,
      aes(x=lifeExp,
            y=continent,
            fill=continent)) +
  geom_boxplot(alpha = 0.2) +
  geom_point(aes(color = continent),
    position =
          position_jitter(height=0.1)) +
  theme(legend.position = "none")
```

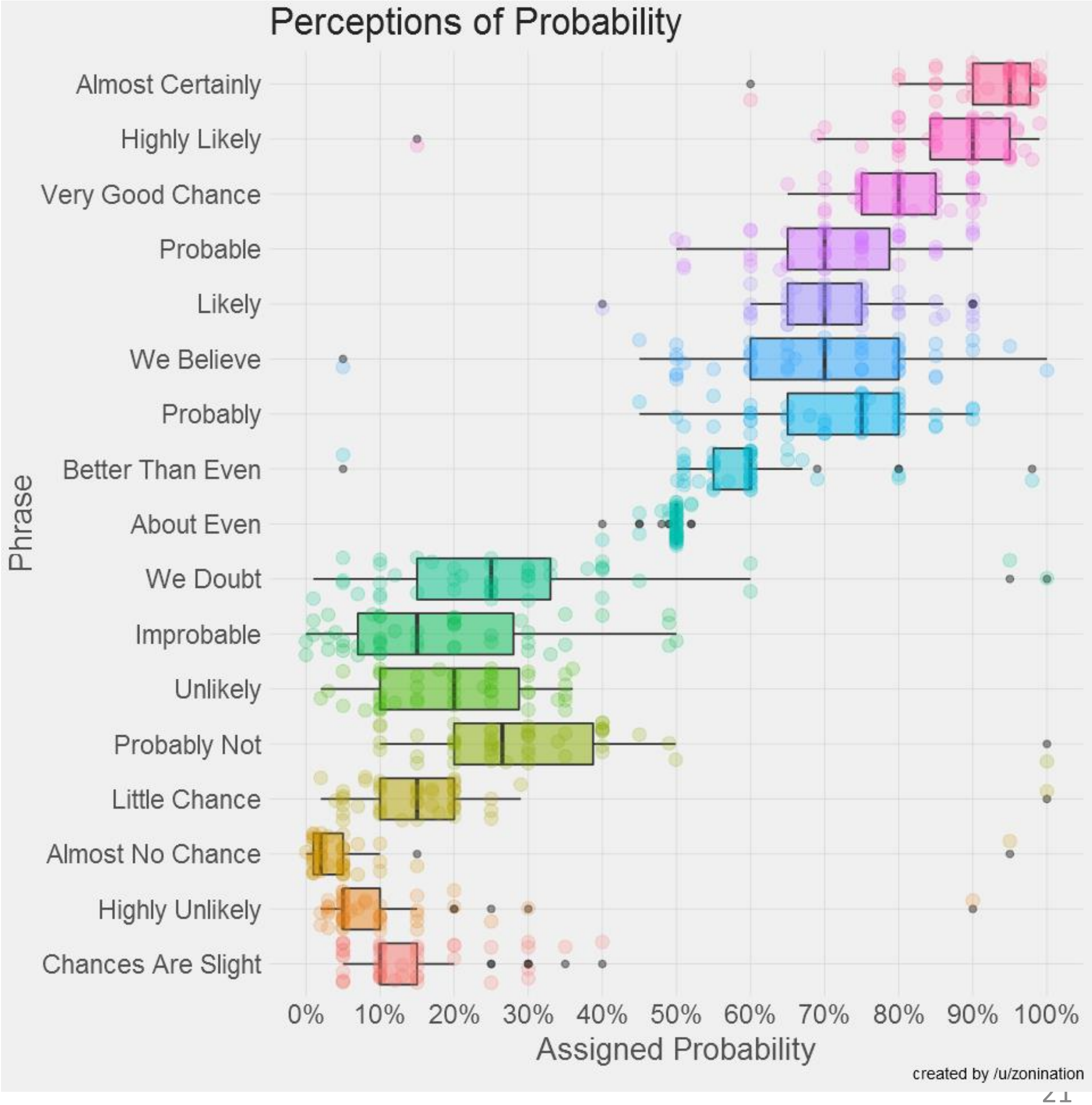


How people view “probability”

What makes this graph successful?

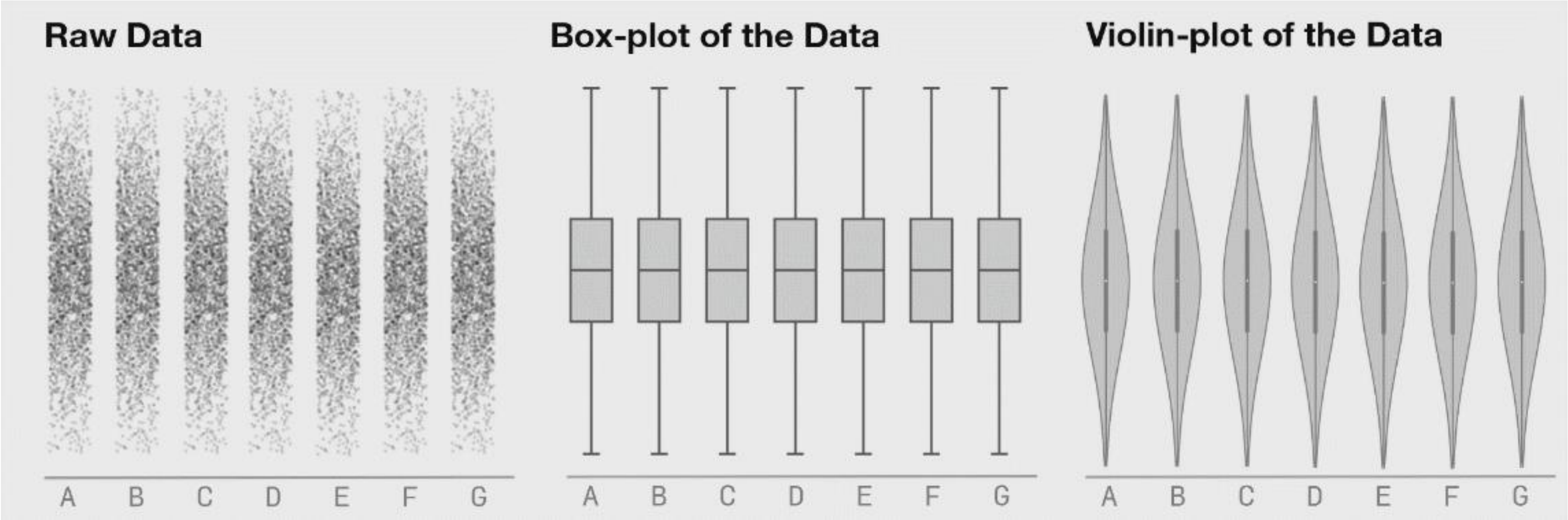
Note the wide range of variability (uncertainty) in the estimates:
“about even” vs. “we believe”

Outliers: individuals who misunderstood instructions?



Problems with boxplots revealed

Boxplots are fine for unimodal distributions – well summarized by Q1, Median, Q3
They are insensitive to multi-modal data

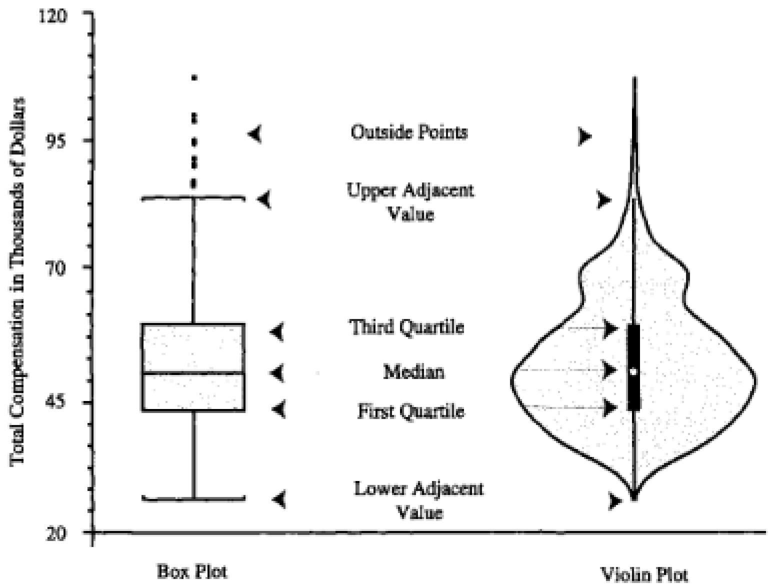


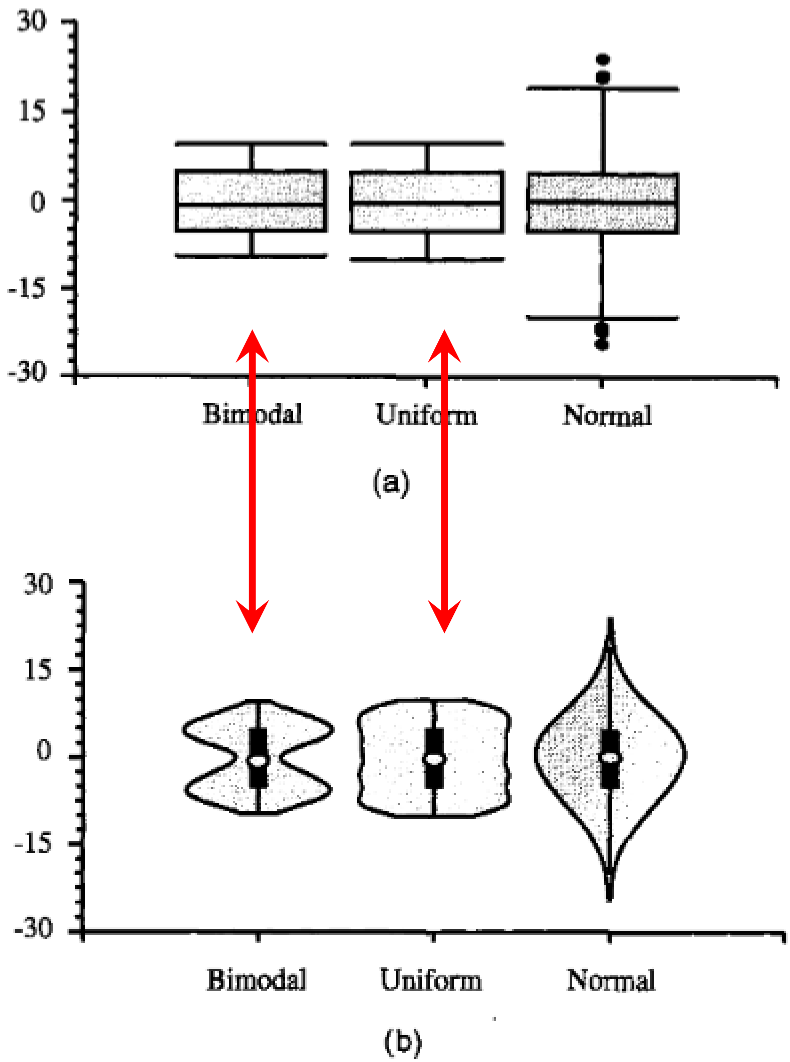
Violin plots

Boxplots are great for ~ normal data

- Shows center, spread, outliers

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

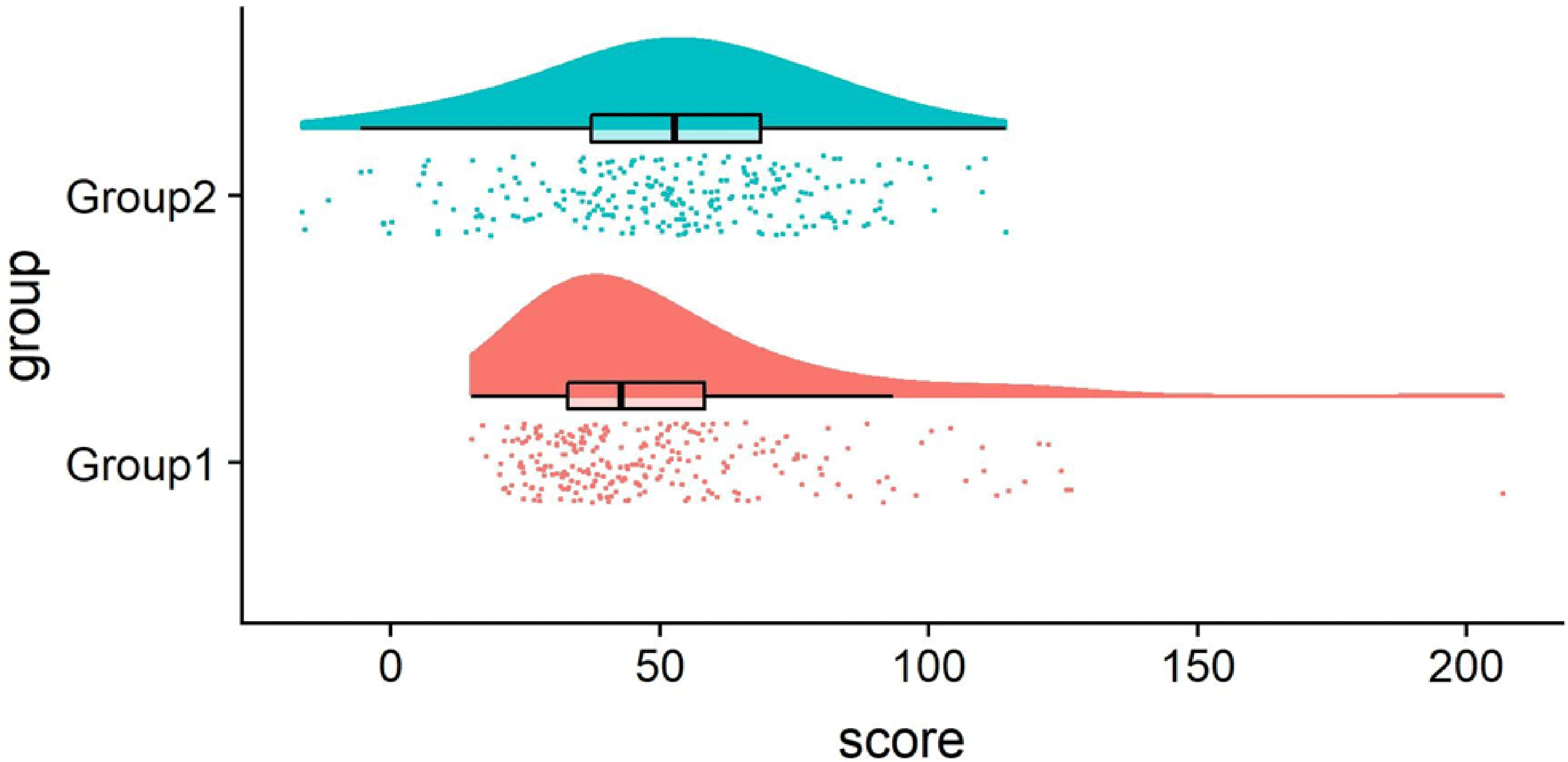




Hintze & Nelson (1998), *American Statistician*, 52:2, 181-184

Raincloud plots

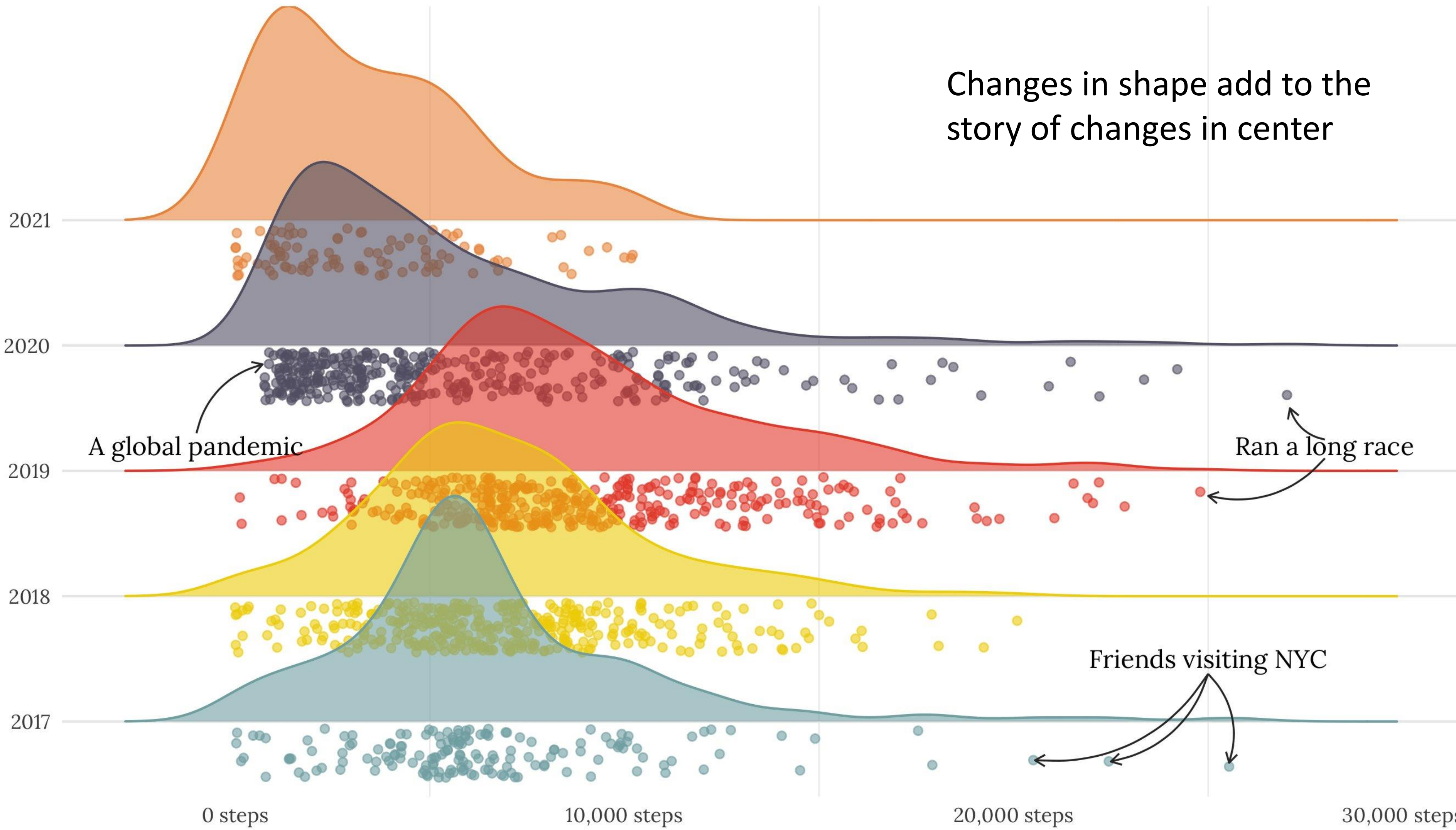
Raincloud plots combine density curve & boxplot, but also show the observations as jittered points



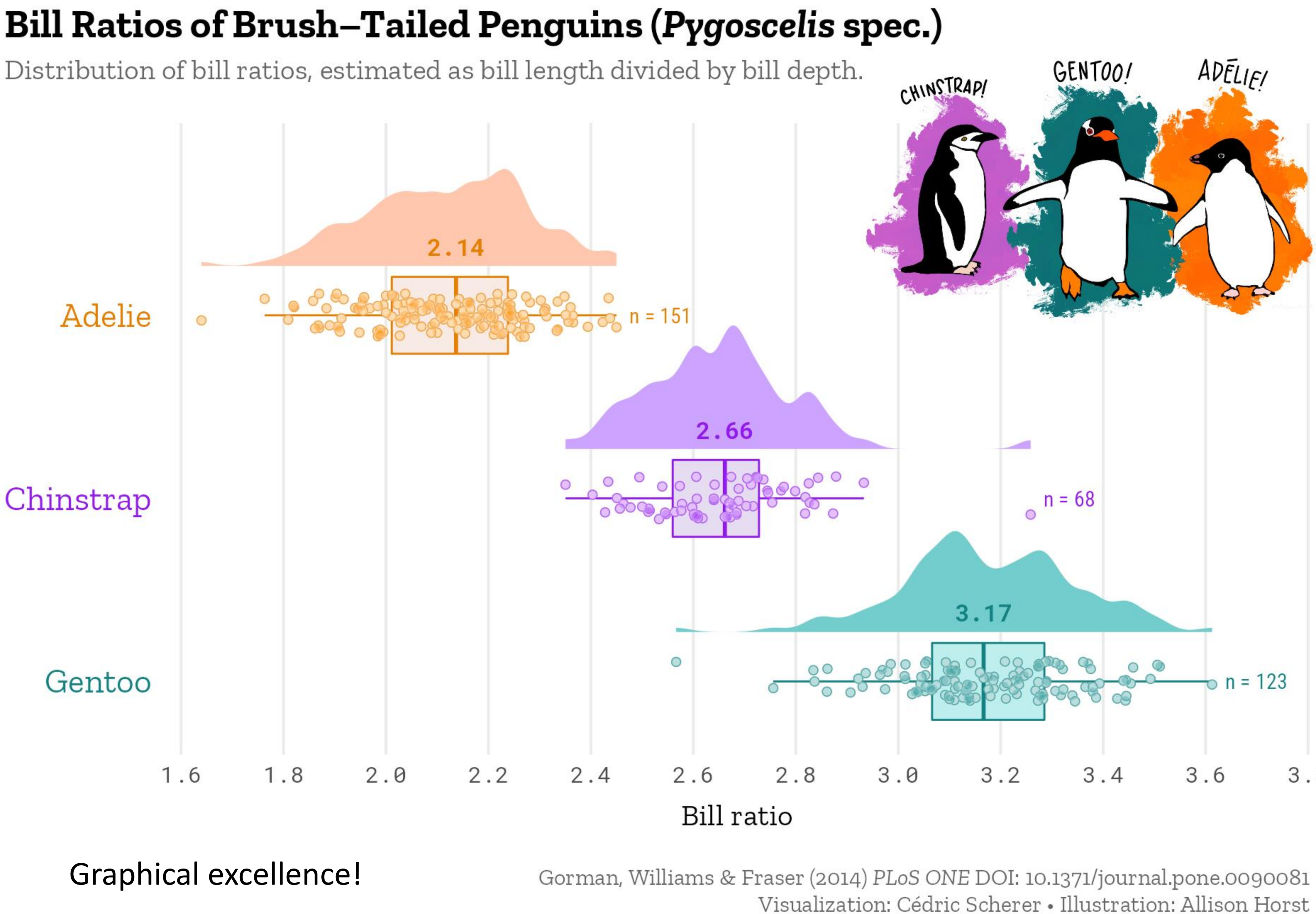
Allen M, Poggiali D, Whitaker K et al. Raincloud plots: a multi-platform tool for robust data visualization [version 2]. Wellcome Open Res 2021, 4:63 (doi: 10.12688/wellcomeopenres.15191.2)

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.



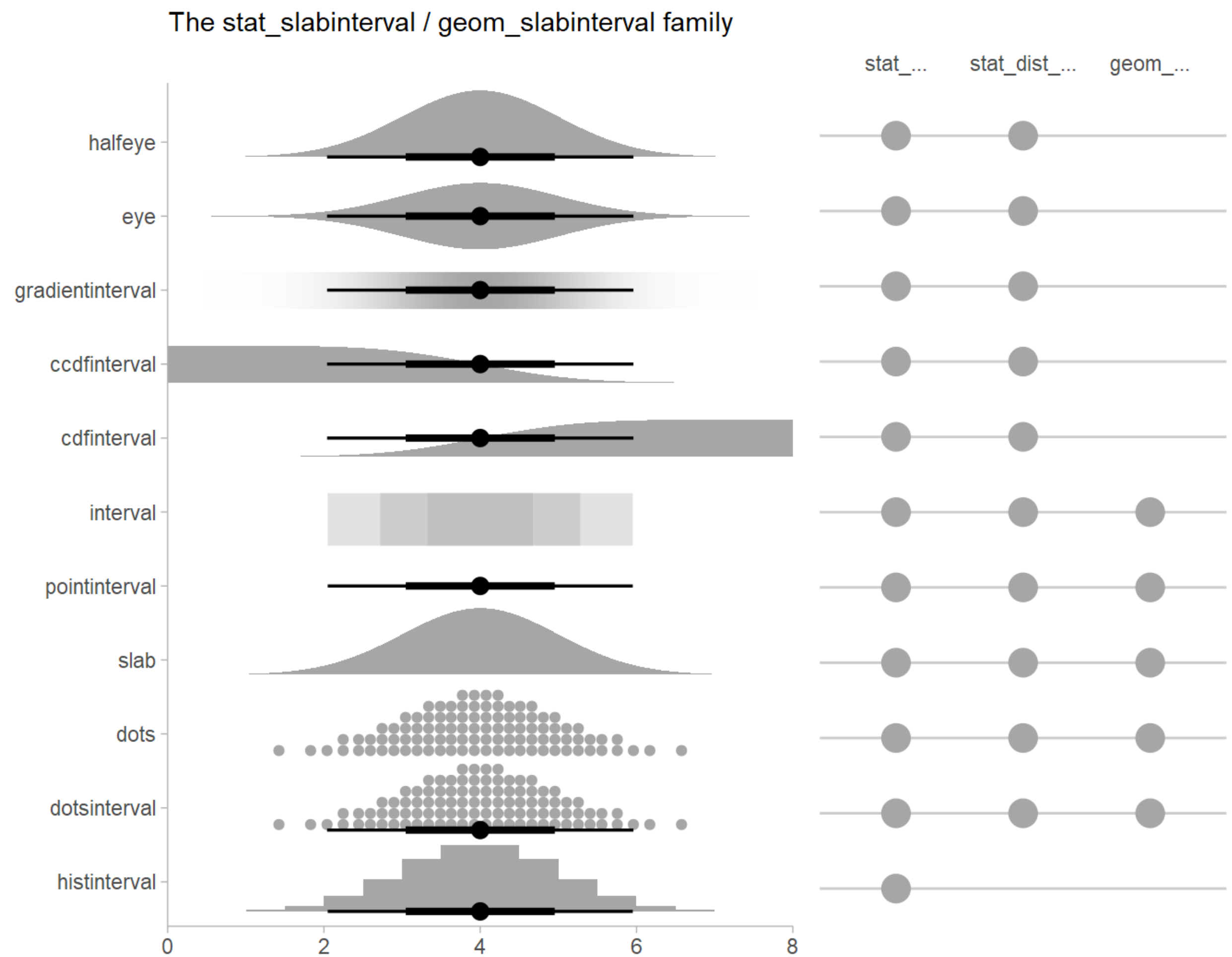
Created by @kllycttn | Data from personal health data

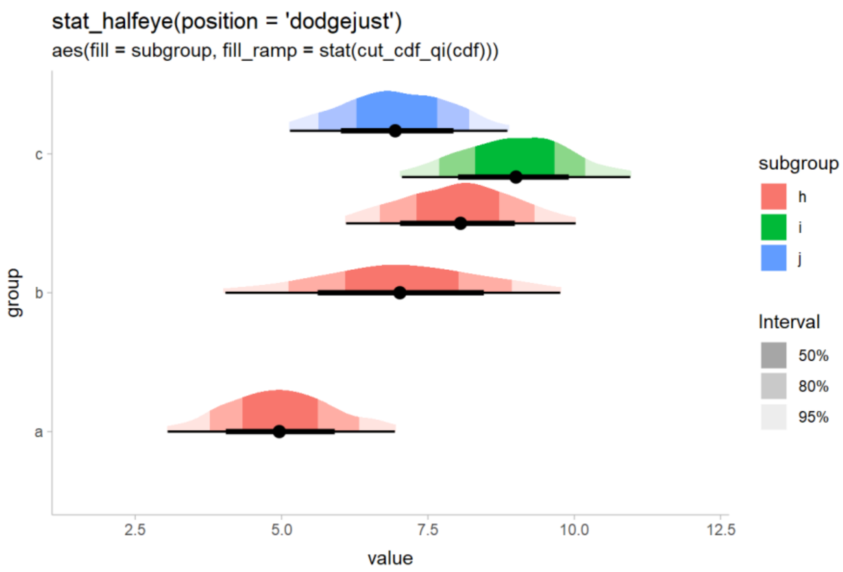


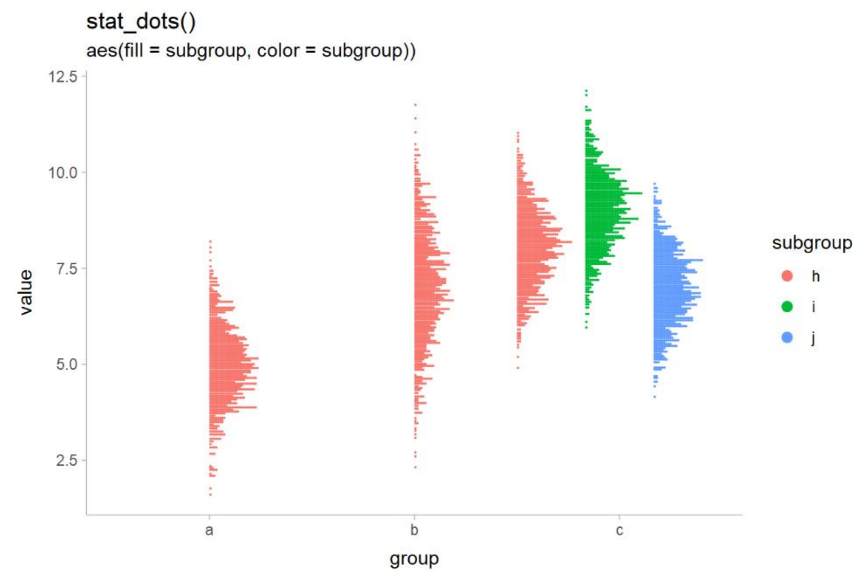
Graphical excellence!

{ggdist} package

The {ggdist} package provides a wide variety of ggplot stats to display distributions & intervals

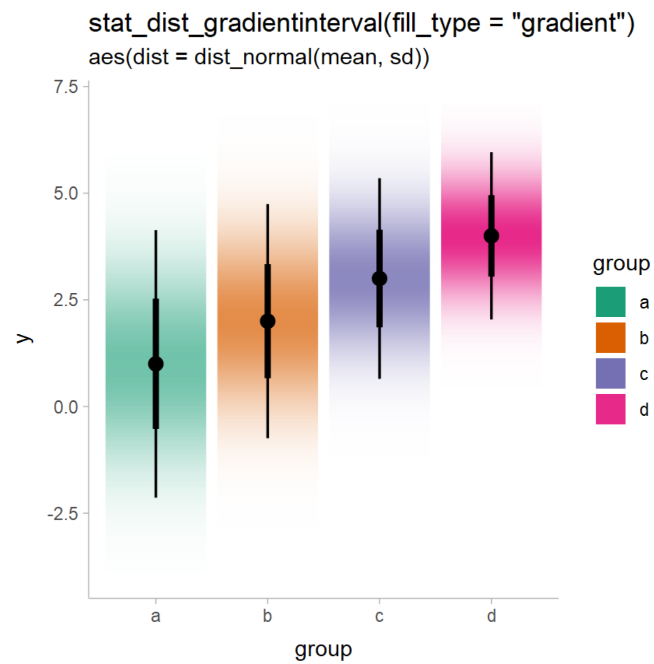






The design of {ggdist} makes it easy to combine two or more graphical representations –

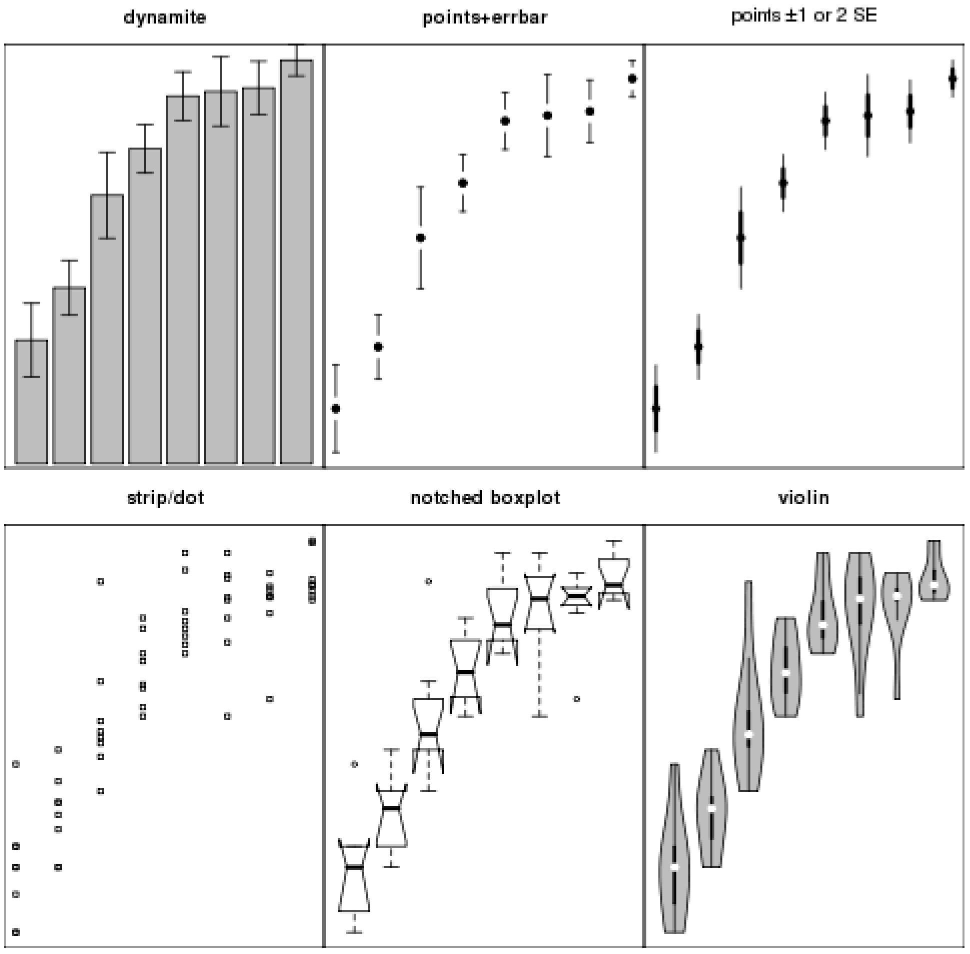
data + distribution + interval



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

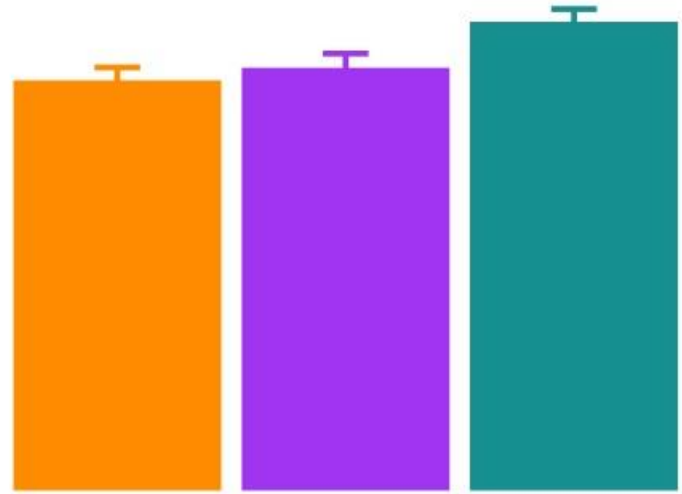
Always explain what error bars mean

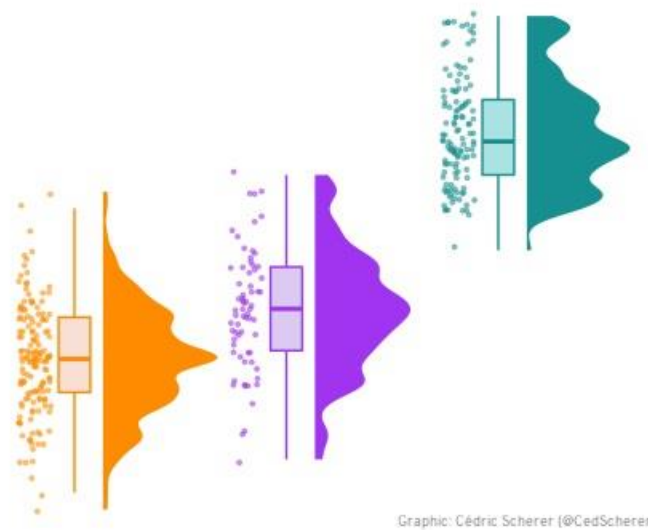
Consider tradeoff between summarization & exposure

Barplot memes: Don’t dynamite me!

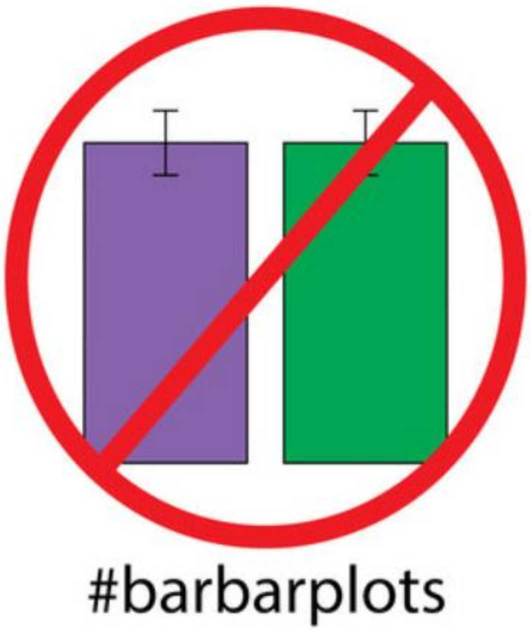
rage

joy





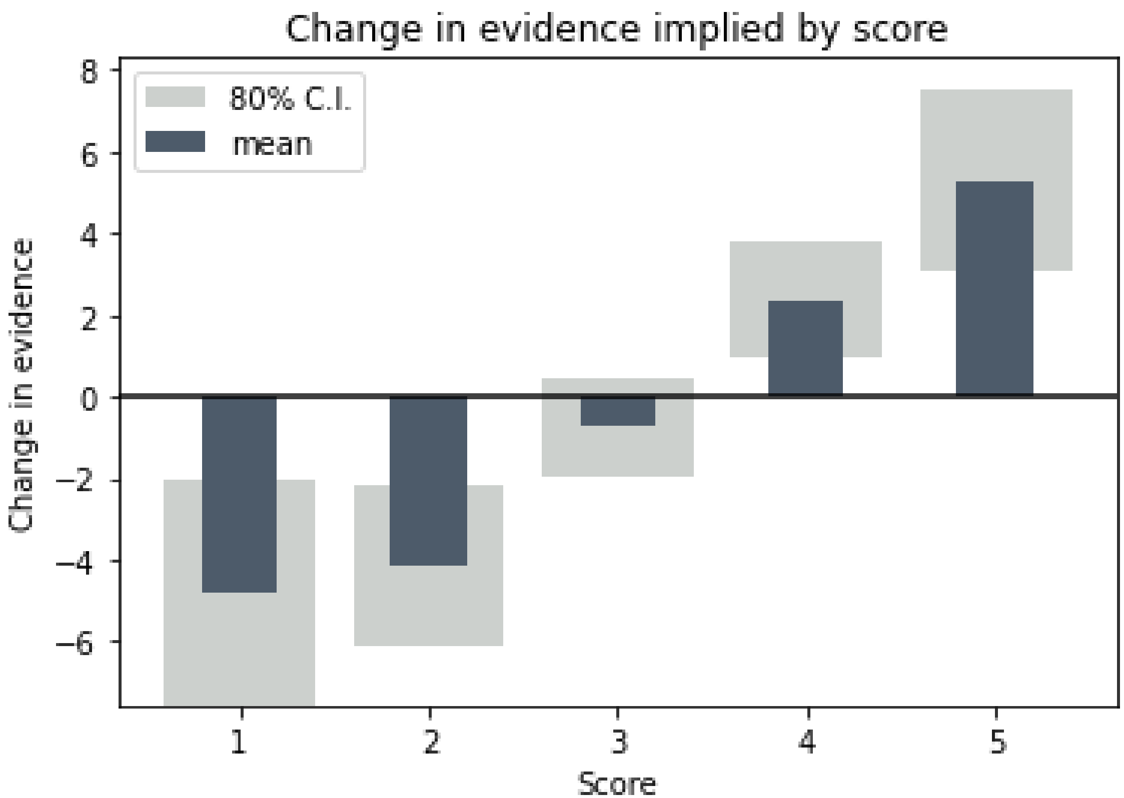
De-fusing the barplot



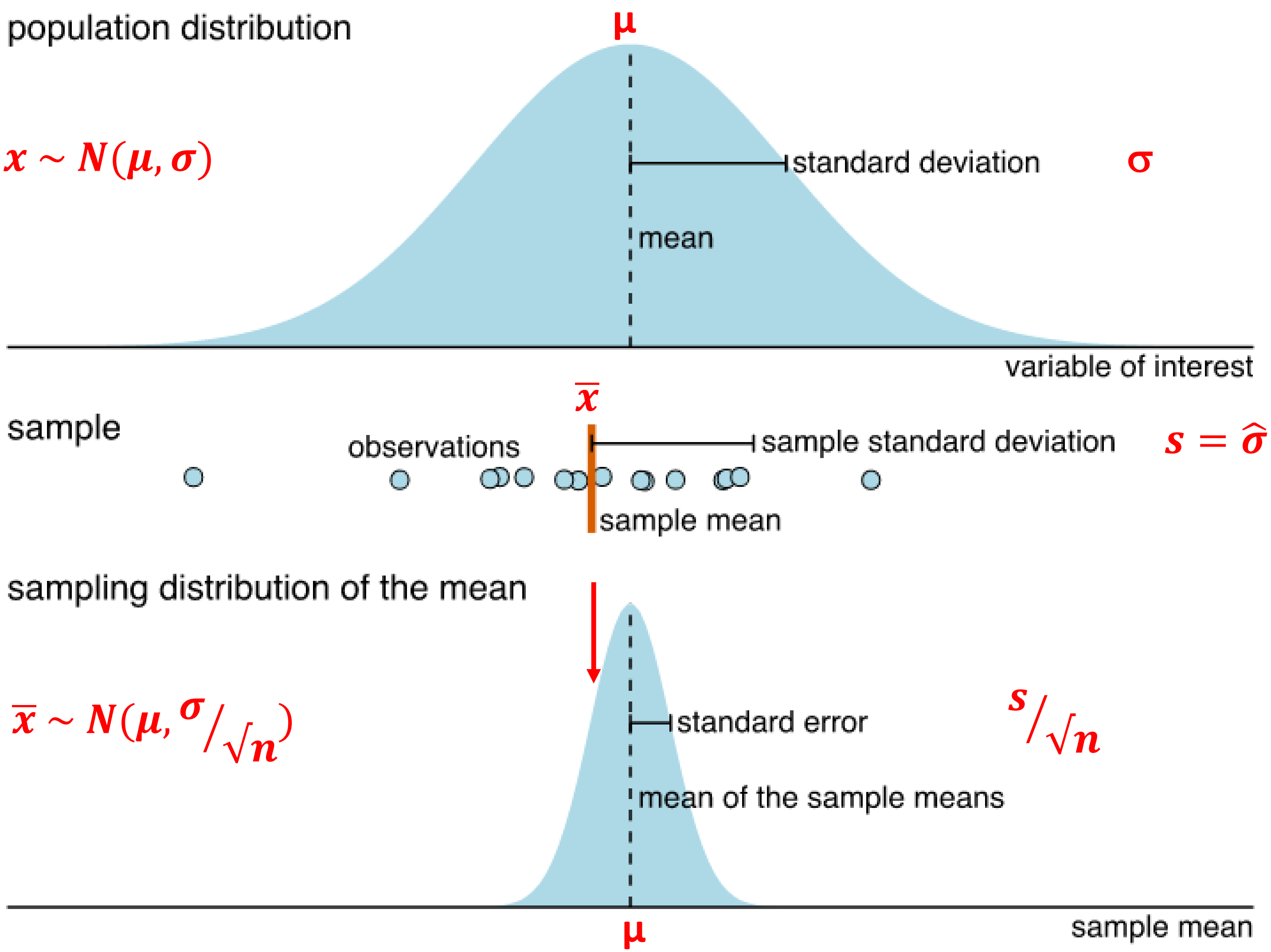
“Friends don’t let friends make barplots” (video)

https://barbarplots.github.io/

If you insist on bars, use a better visual representation of uncertainty or CI



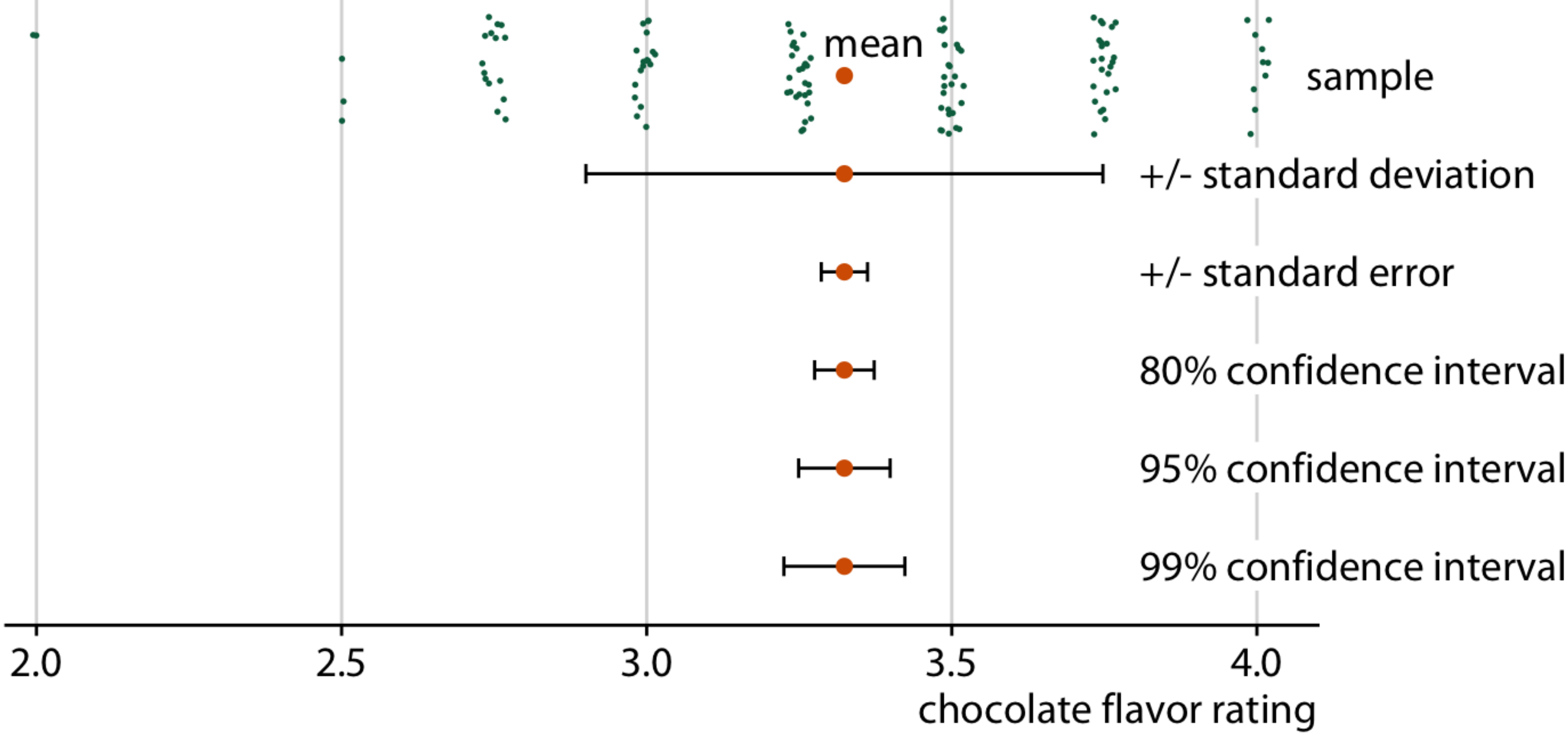
Key ideas of σtatistical σampling



From: Claus Wilke (2021), *Fundamentals of Data Visualization*, https://clauswilke.com/dataviz/ , Ch 16

Visualizing distributions: Error bars

There are many ways to show variability in a single sample

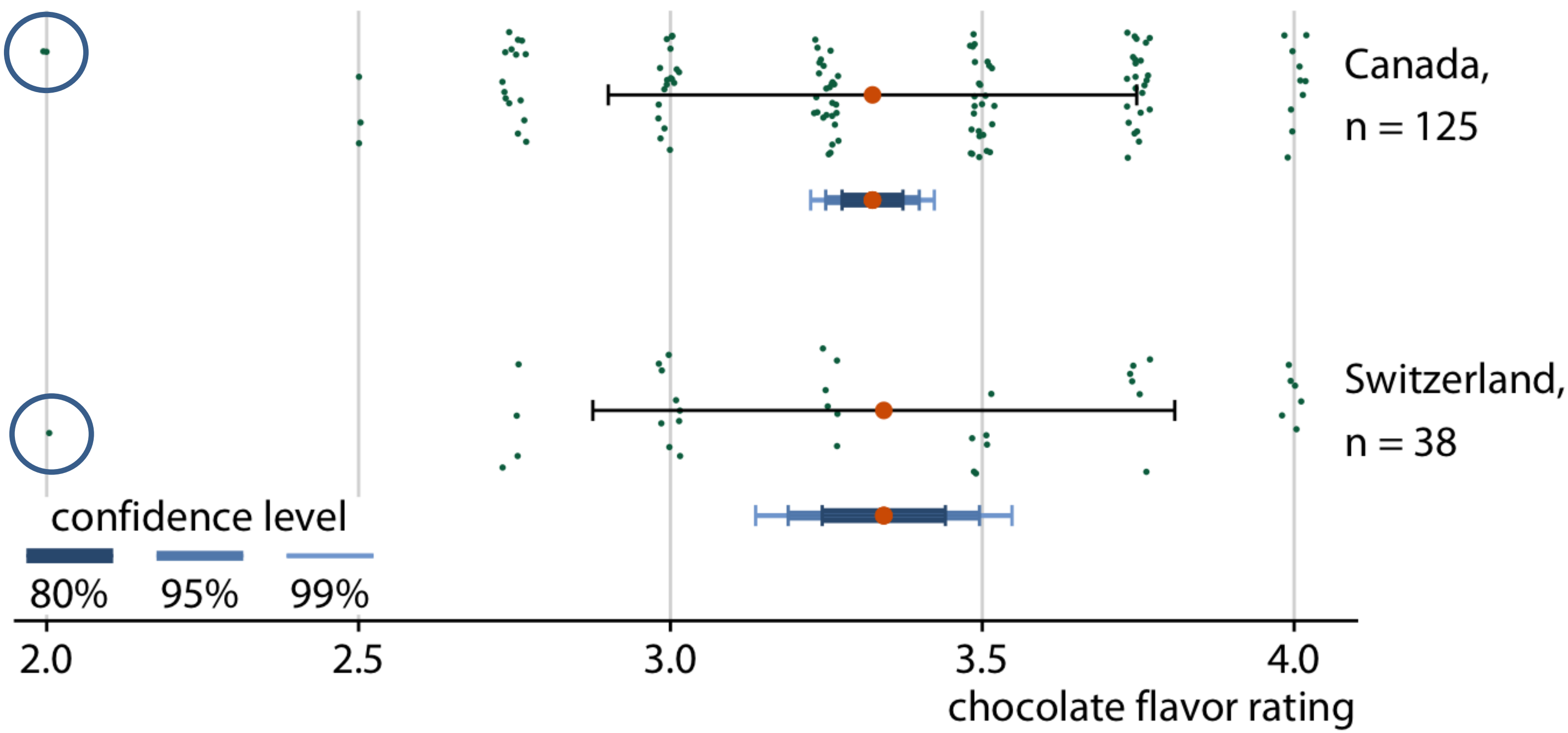


Expert ratings of 125 chocolate bars manufactured in Canada

From: Claus Wilke (2021), *Fundamentals of Data Visualization*, https://clauswilke.com/dataviz/ , Ch 16

Comparing distributions: Sample size

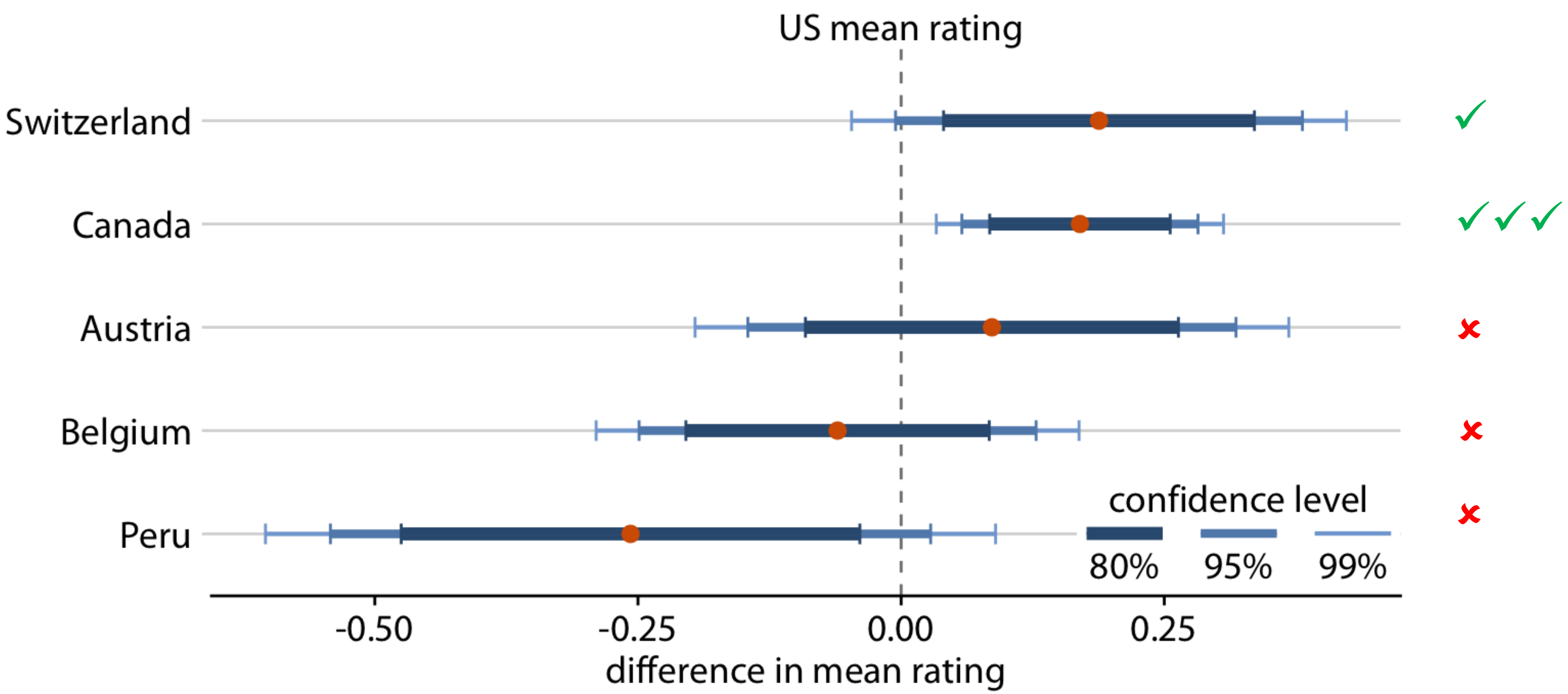
- means and standard deviations are similar for Canada & Switzerland
- confidence interval widths ~ $1/\sqrt{n}$
- can show different sized confidence bands together 
- dots show the data: sample size & are there any outliers?



From: Claus Wilke (2021), *Fundamentals of Data Visualization*, https://clauswilke.com/dataviz/ , Ch 16

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



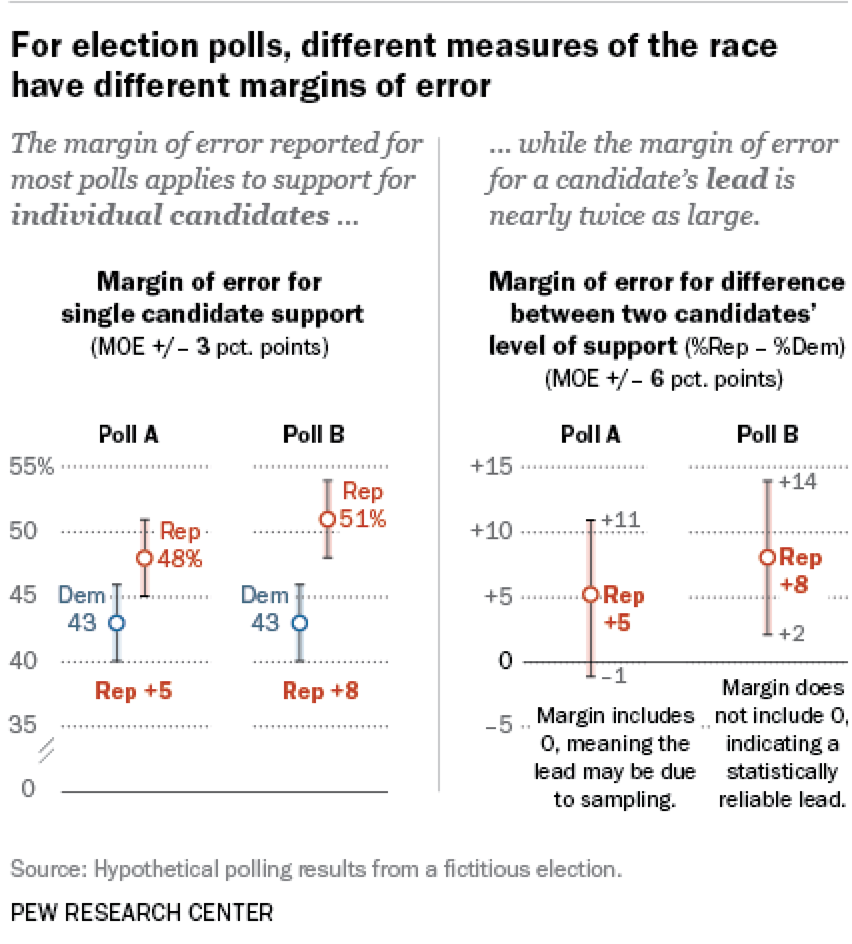
From: Claus Wilke (2021), *Fundamentals of Data Visualization*, https://clauswilke.com/dataviz/ , Ch 16

Intervals: Direct vs. Differences

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

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

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



What kind of intervals?

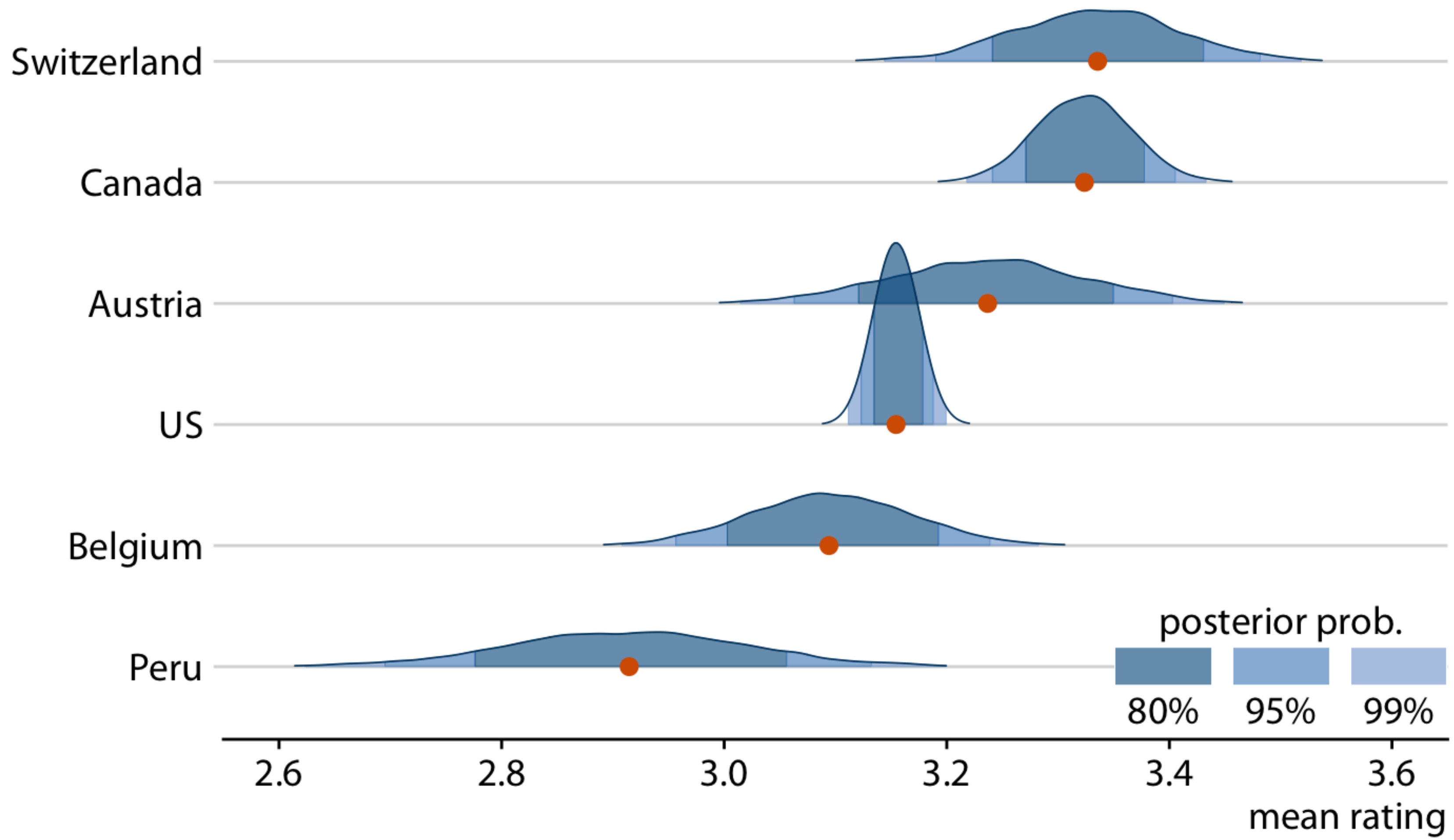
Frequentist

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

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

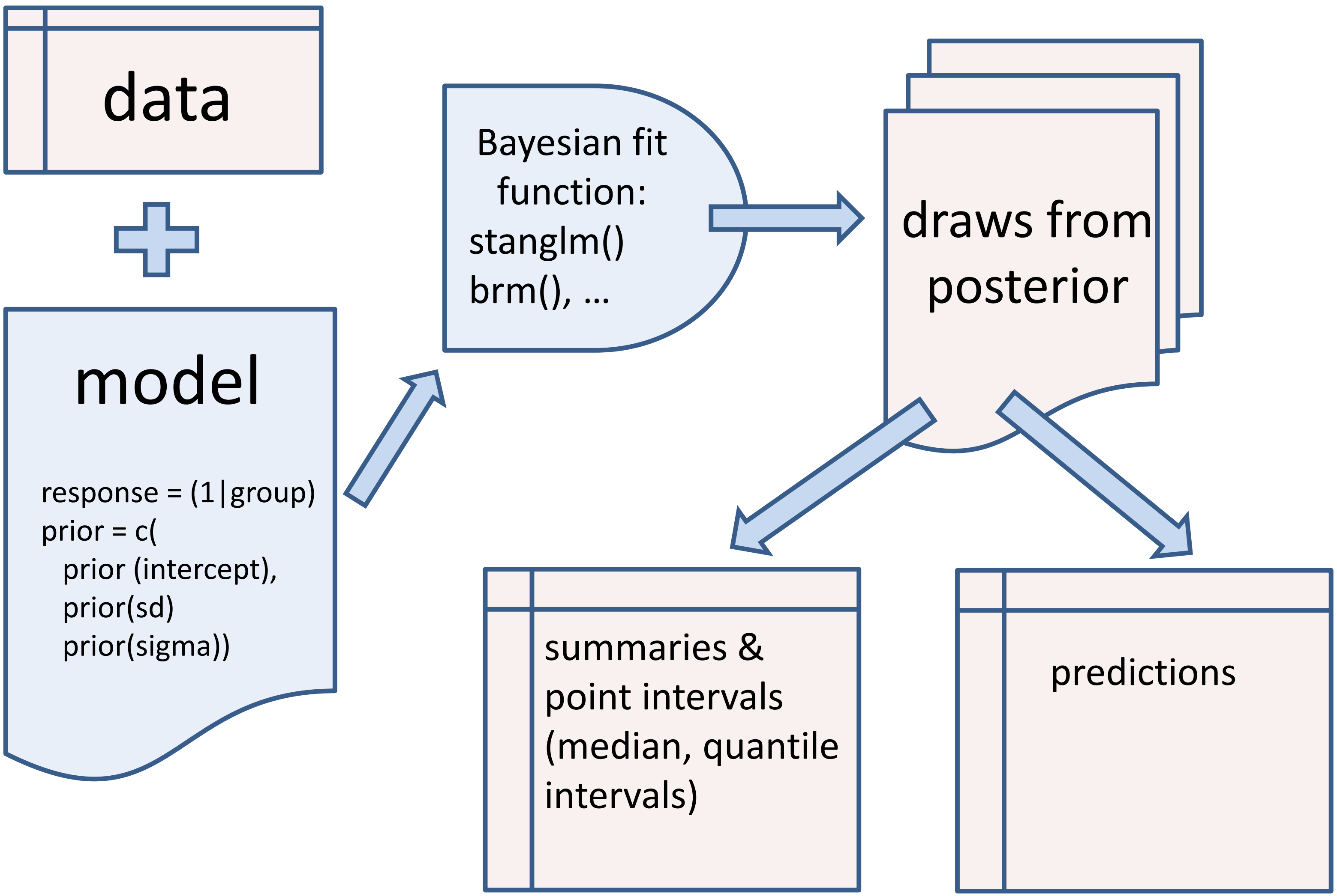


From: Claus Wilke (2021), *Fundamentals of Data Visualization*, https://clauswilke.com/dataviz/ , Ch 16

tidybayes: Bayesian analysis + tidy data + geoms

- The {tidybayes} package makes it easier to combine Bayesian analysis with insightful ggplot visualization
  - Bayesian packages: JAGS, Stan (rstanarm), brms
  - Inputs: data, model specifications aren't tidy
    - Need to translate data into forms these packages expect
  - Outputs: Posterior draws, distributions aren't tidy
    - Need to translate these into form suitable for summaries & plotting
    - → Extract tidy fits and predictions from models
    - → Summarize posterior distributions
    - → Visualize priors and posteriors

The Bayesian process



Posterior = Prior * Likelihood

We have: Data, some model, some parameter(s) of interest, θ

Can calculate likelihood, $p(Data|\theta)$

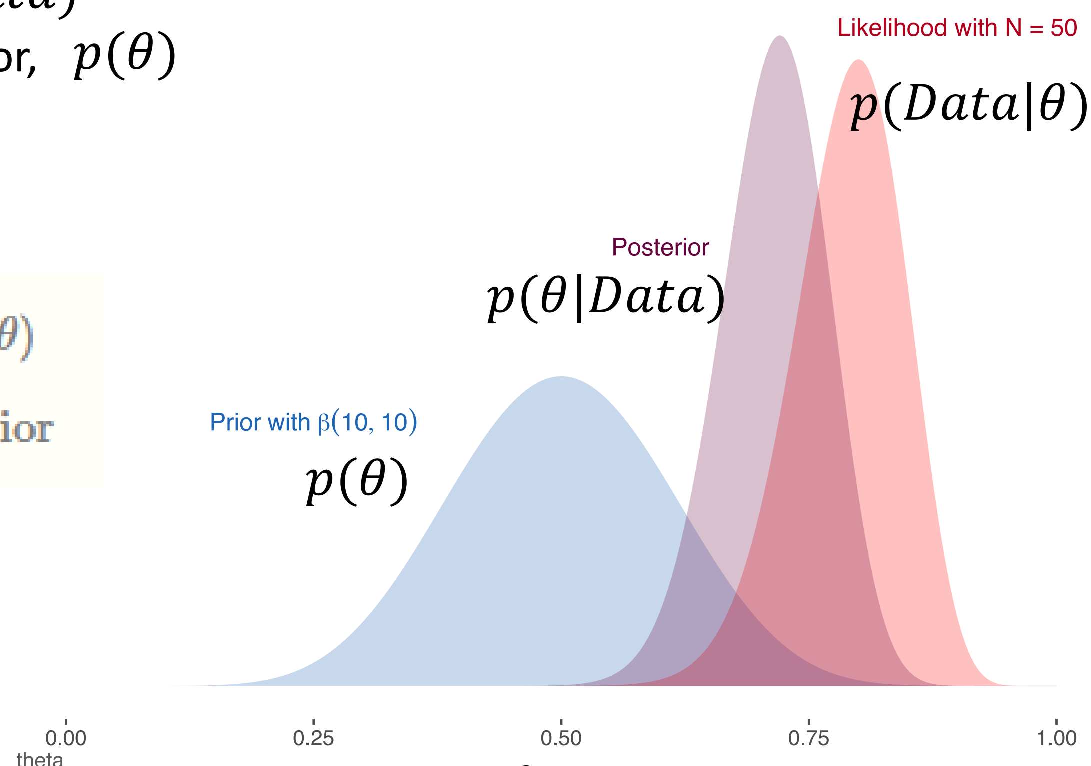
Want: posterior: $p(\theta|Data)$

Previous research: some prior, $p(\theta)$

Bayes theorem:

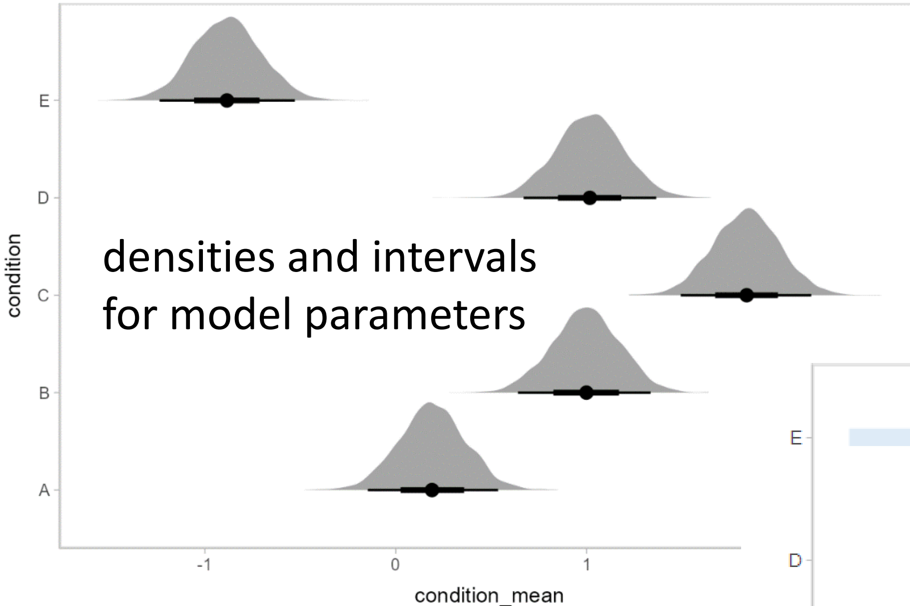
$$p(\theta|Data) \propto p(Data|\theta) \cdot p(\theta)$$

$$\text{posterior} \propto \text{likelihood} \cdot \text{prior}$$



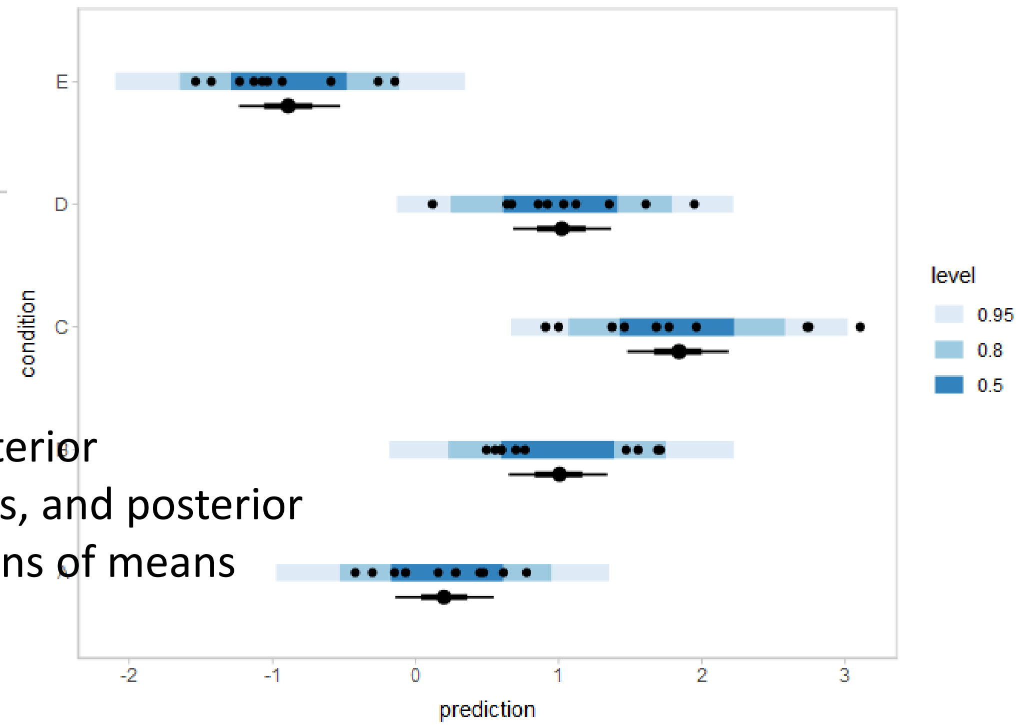
θ (mean, slope, …)

tidybayes plots

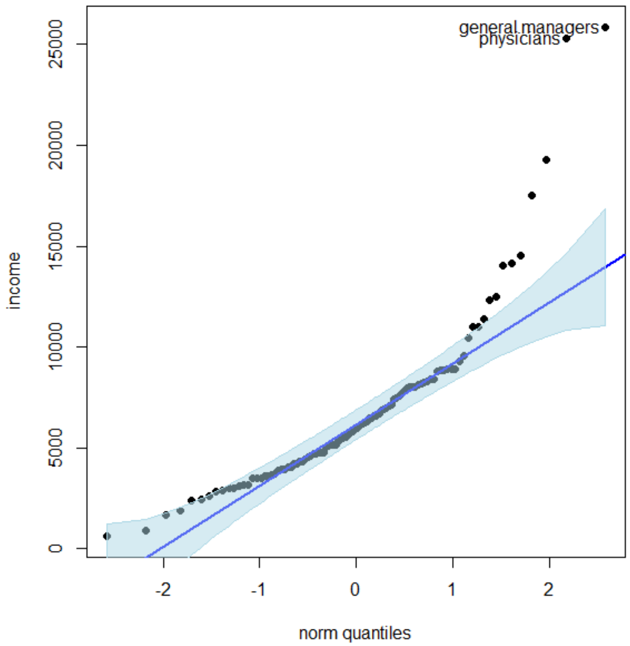


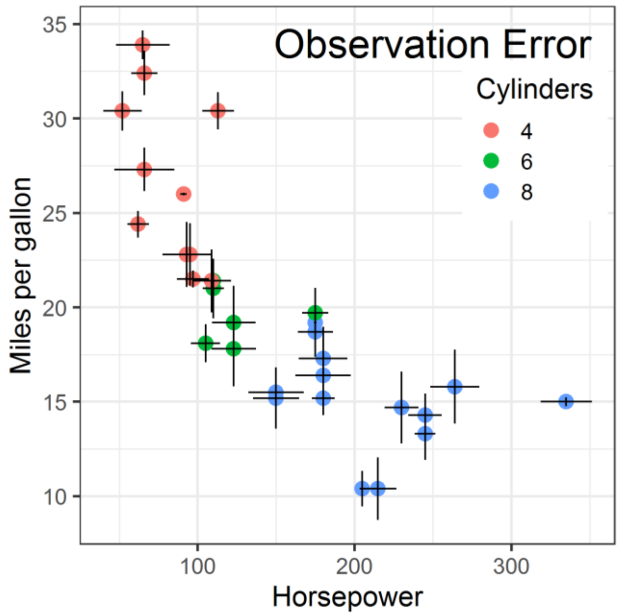
densities and intervals
for model parameters

Everything we can do for data with {ggdist} we can do for Bayesian models with {tidybayes}



data, posterior
predictions, and posterior
distributions of means





- QQplots
- Model fit plots

Uncertainty in fits & curves

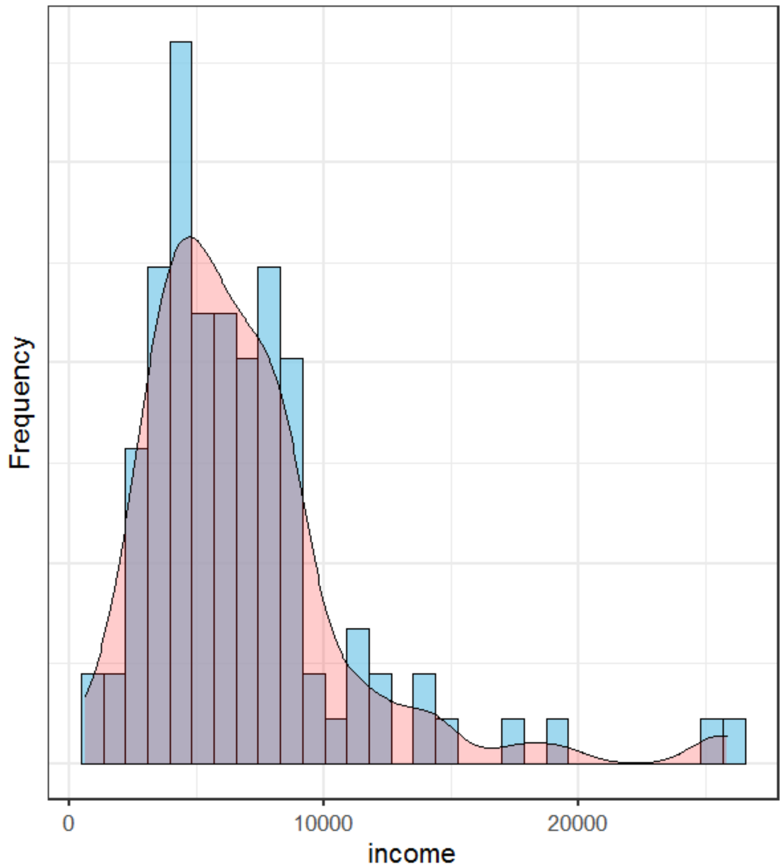
QQ plots

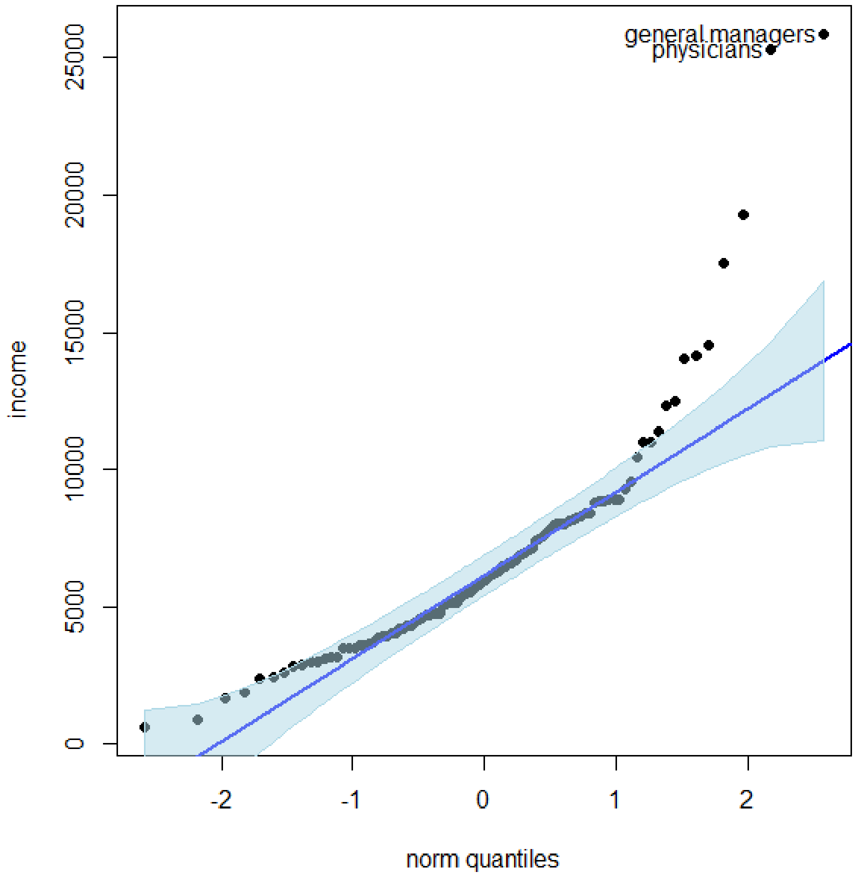
- How close is my data to a {Normal | exponential | $\chi 2$} distribution?
- There are lots of statistical tests, but these don't tell why or where a distribution is rejected.
- These tests are also overly sensitive to small departures
- Plot observed Quantiles vs. theoretical Quantiles
  - If observed ~ theoretical with slope = 1, OK
  - Confidence bands help to identify deviation from model & 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.)

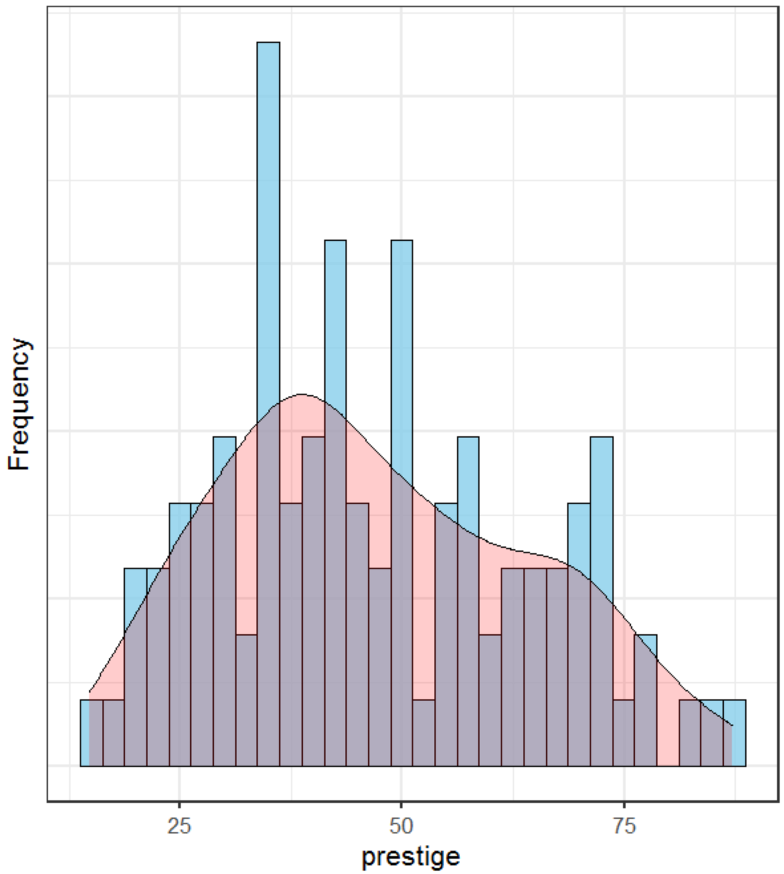
This shows up as a U-shaped pattern
The 95 % confidence band shows greatest departure in the upper tail



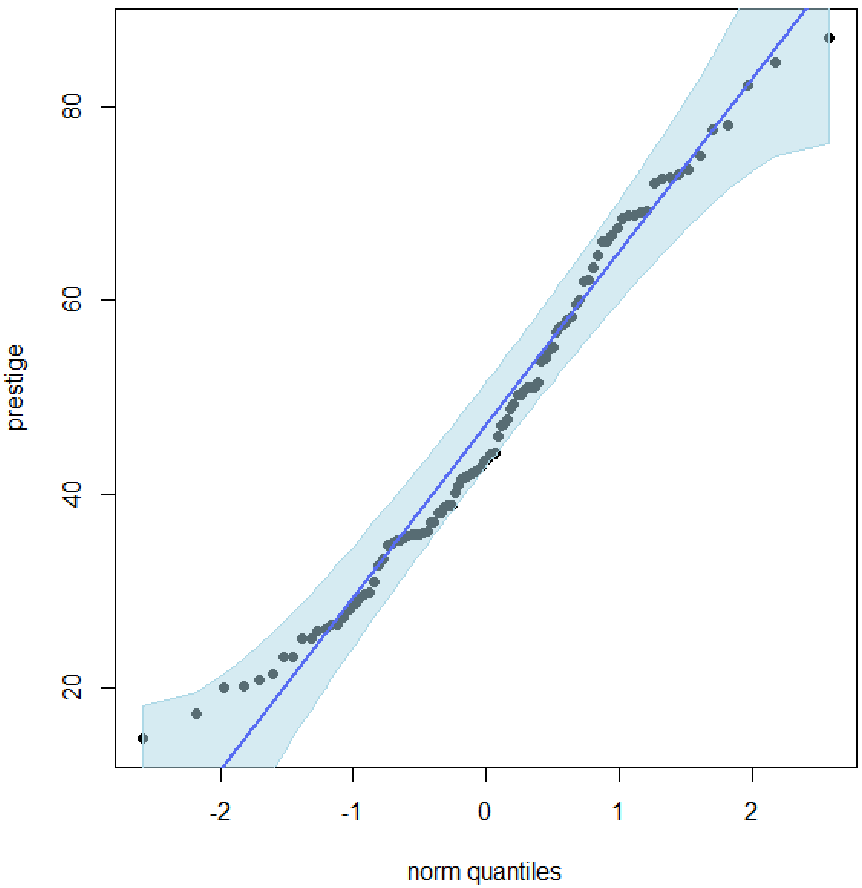


Prestige data: prestige

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



The 95% confidence band includes all the observations



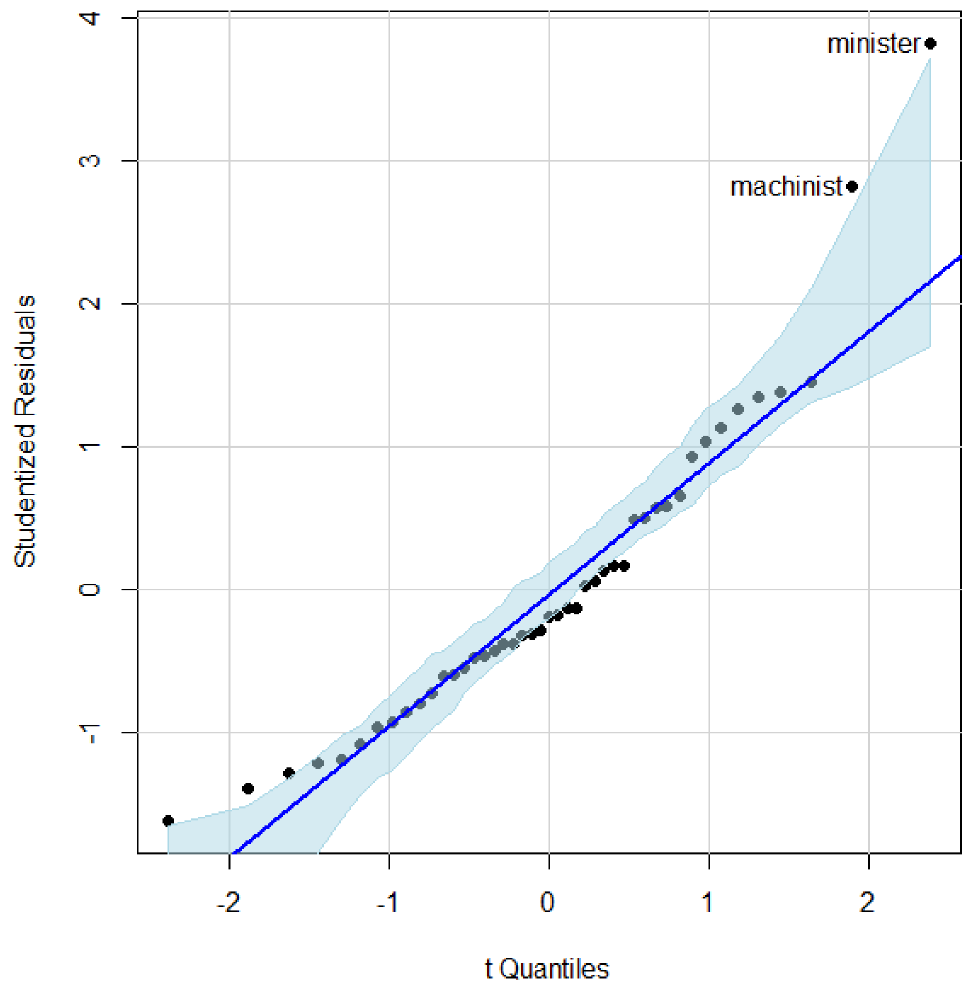
Prestige data: residuals

Normality of residuals is more important for linear models

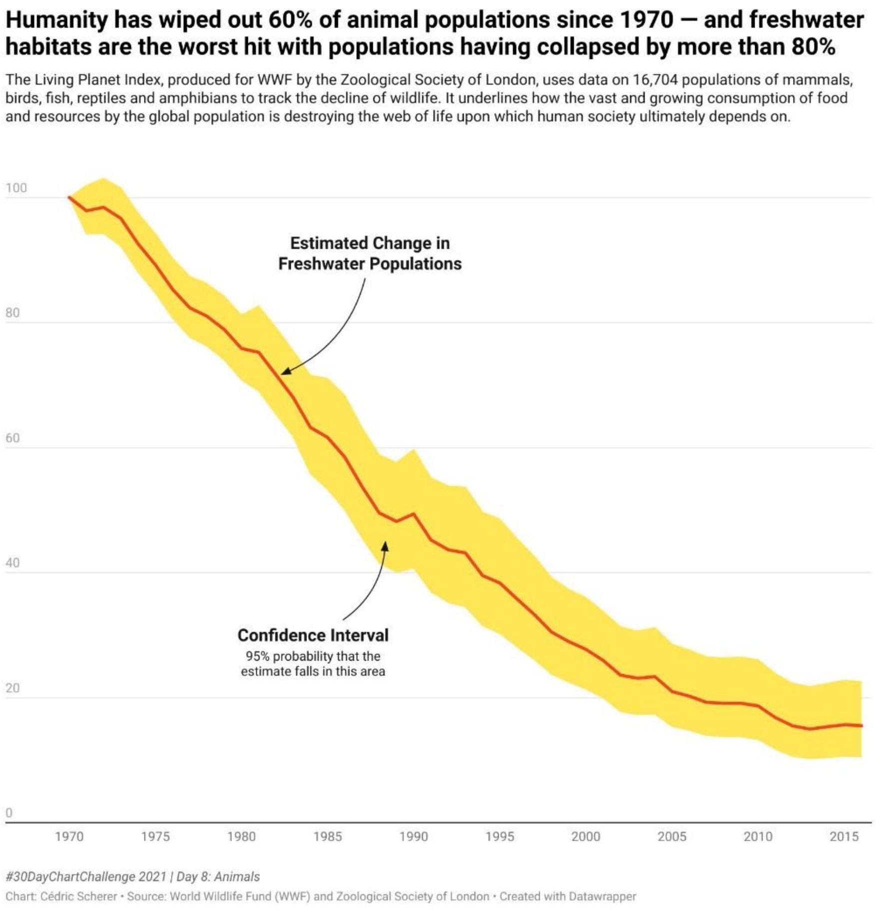
Some small evidence of + skew

Confidence bands help to identify potential outliers – badly fitted pts

```
qqPlot(lm(prestige ~ income + education + type, data=Duncan))
```



Curves + Uncertainty



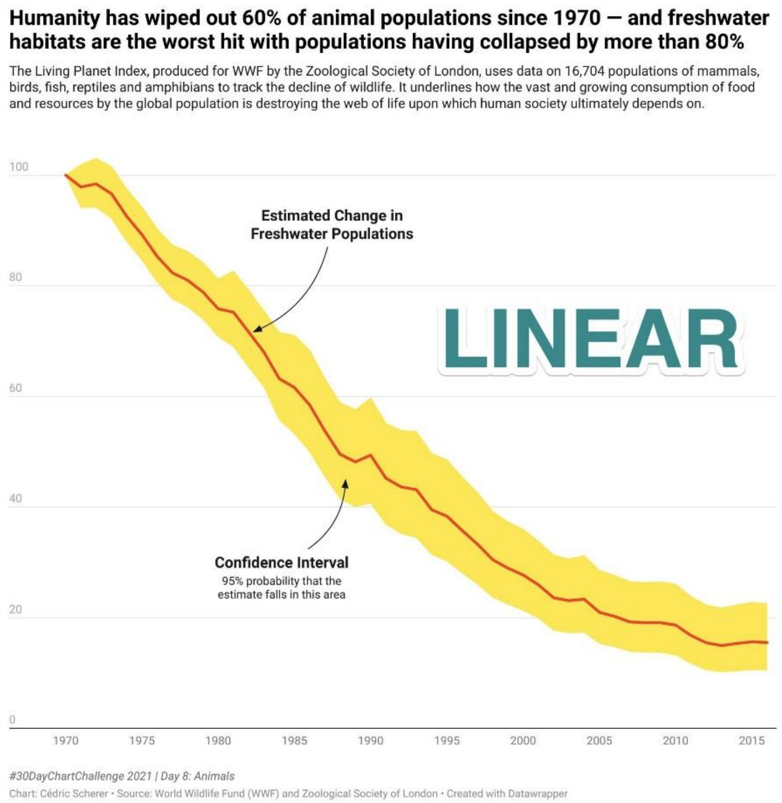
Cederic Scherer used this graphic to argue about the decline of animal & freshwater populations.

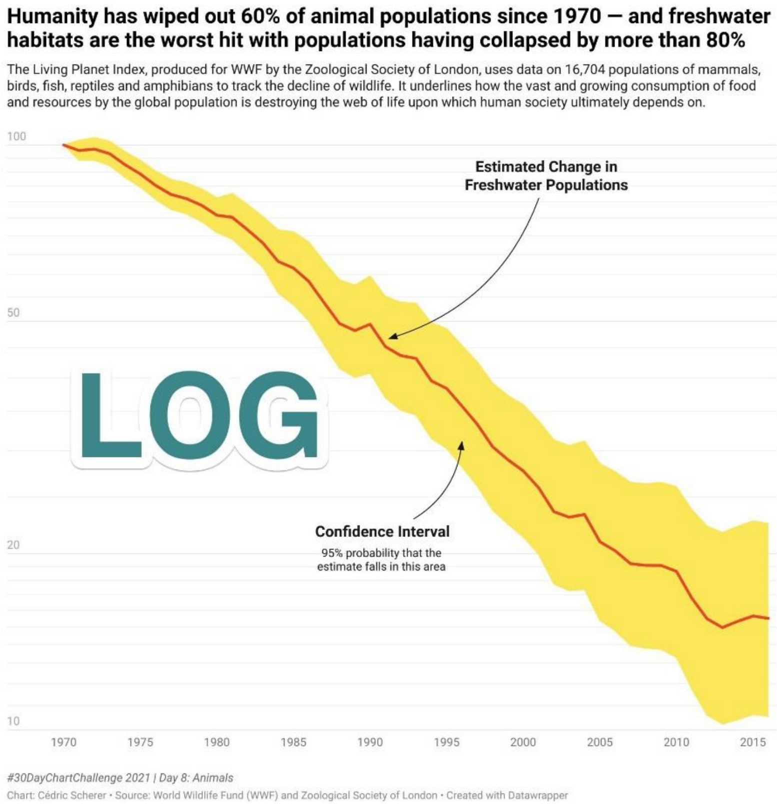
Details aside, the confidence band gives visual evidence that the decline is systematic.

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

Curves + Uncertainty: Scale

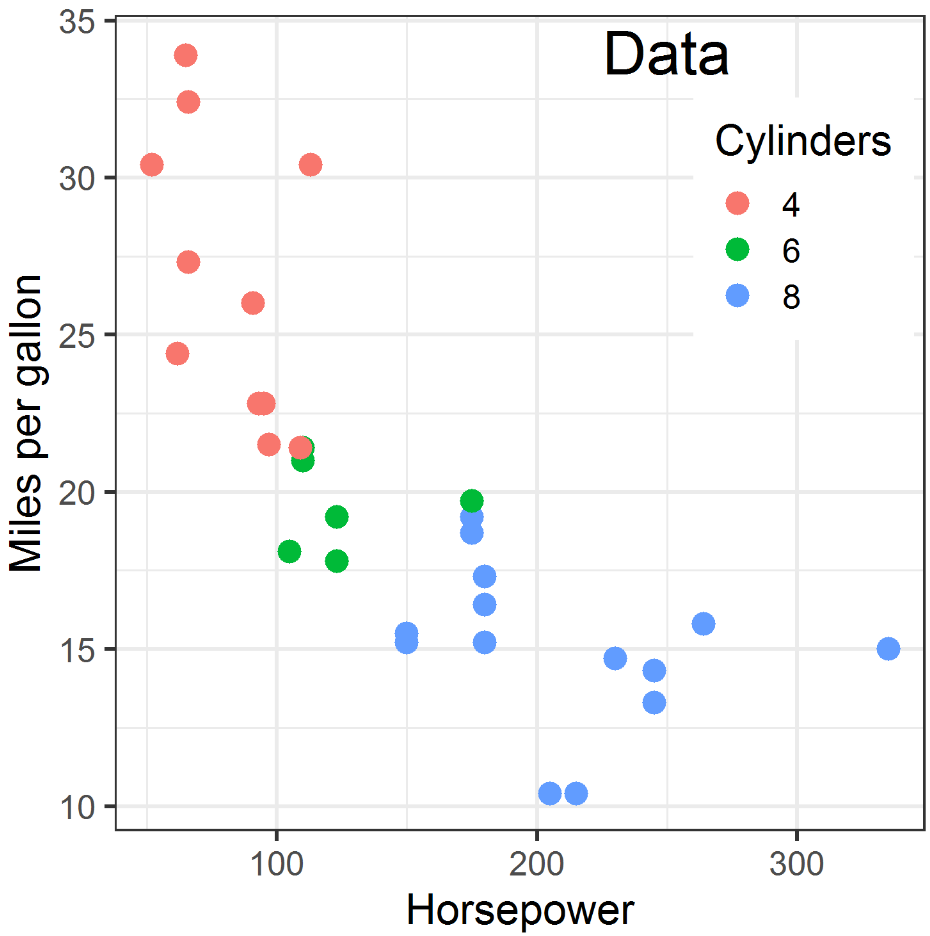
Arguably, percent reduction in animal population should be viewed on a log scale. Transformed uncertainty intervals are here the logs of the Upper/Lower levels





Fitted curves

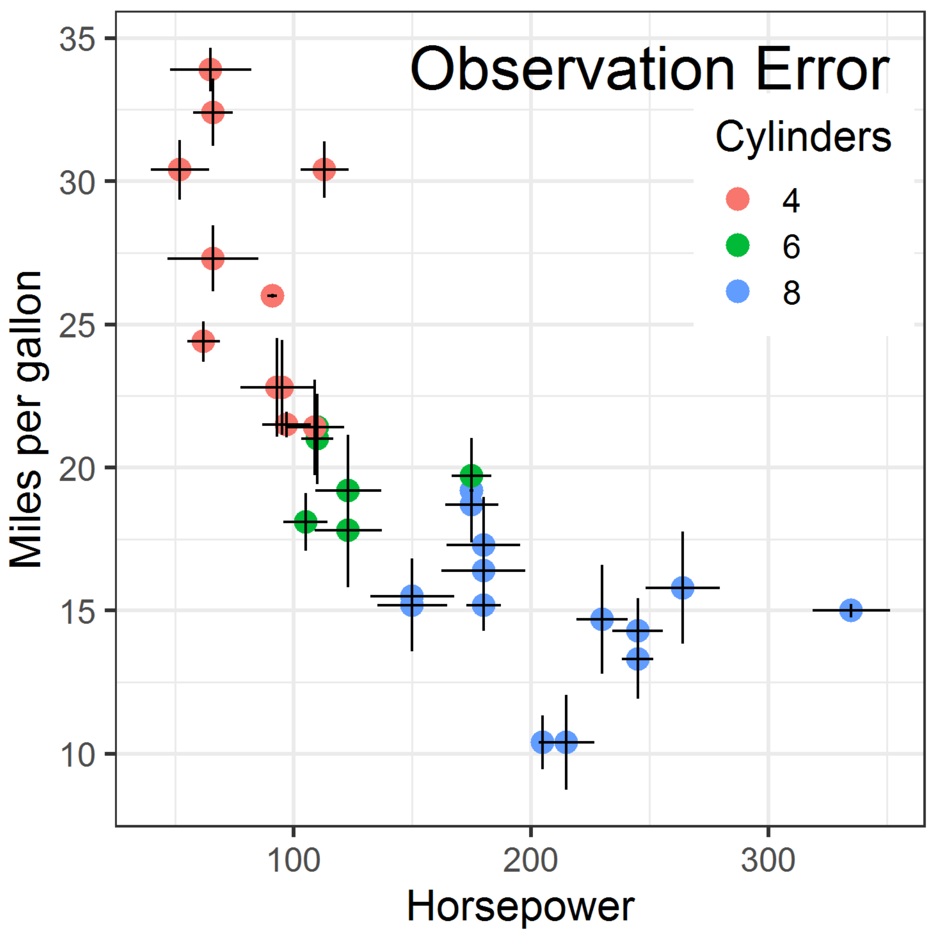
Data on gas mileage of *Motor Trend* 1974 cars



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.

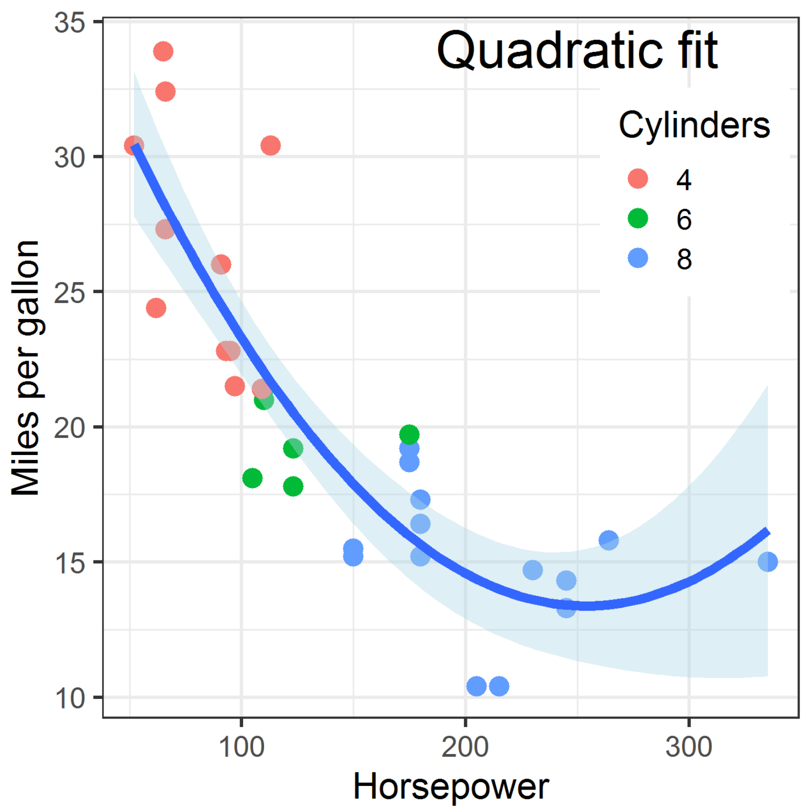
Fitted models allow for errors in *y*:
y = f(*x*) + error
and find estimates to minimize error

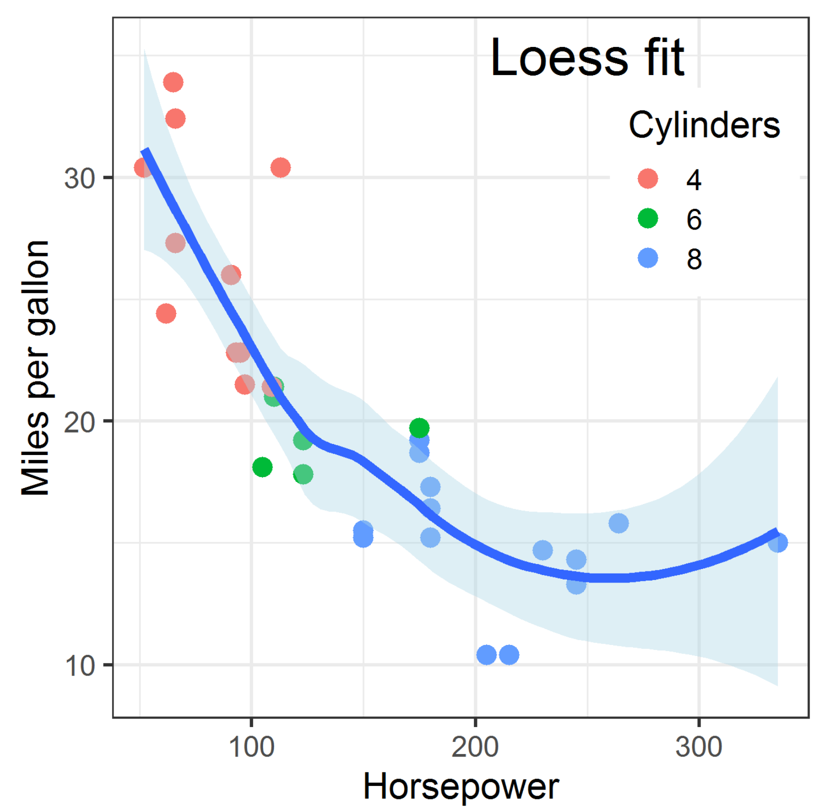
Most fitted models assume *x* is measured w/o error.

Big problem if error ~ f(x, other xs)

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.

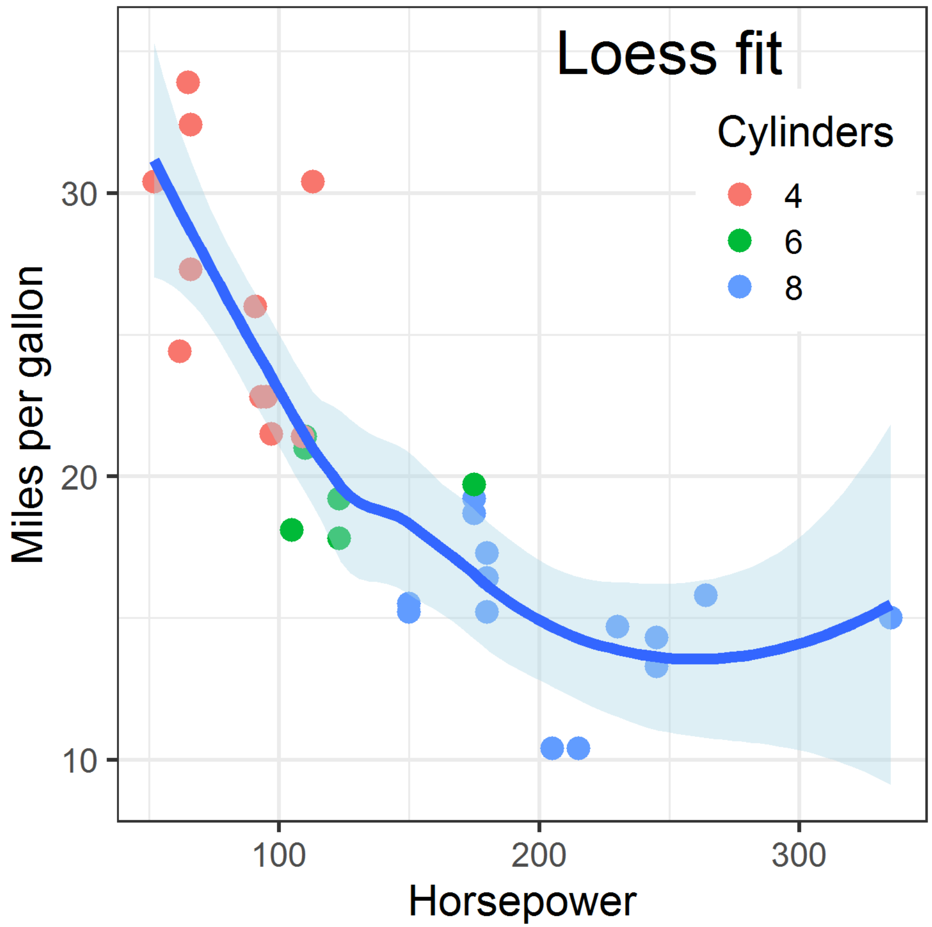





```
p1 + geom_smooth(method = lm, formula = y~poly(x,2), …)
```

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

Fitted curves: smoothers

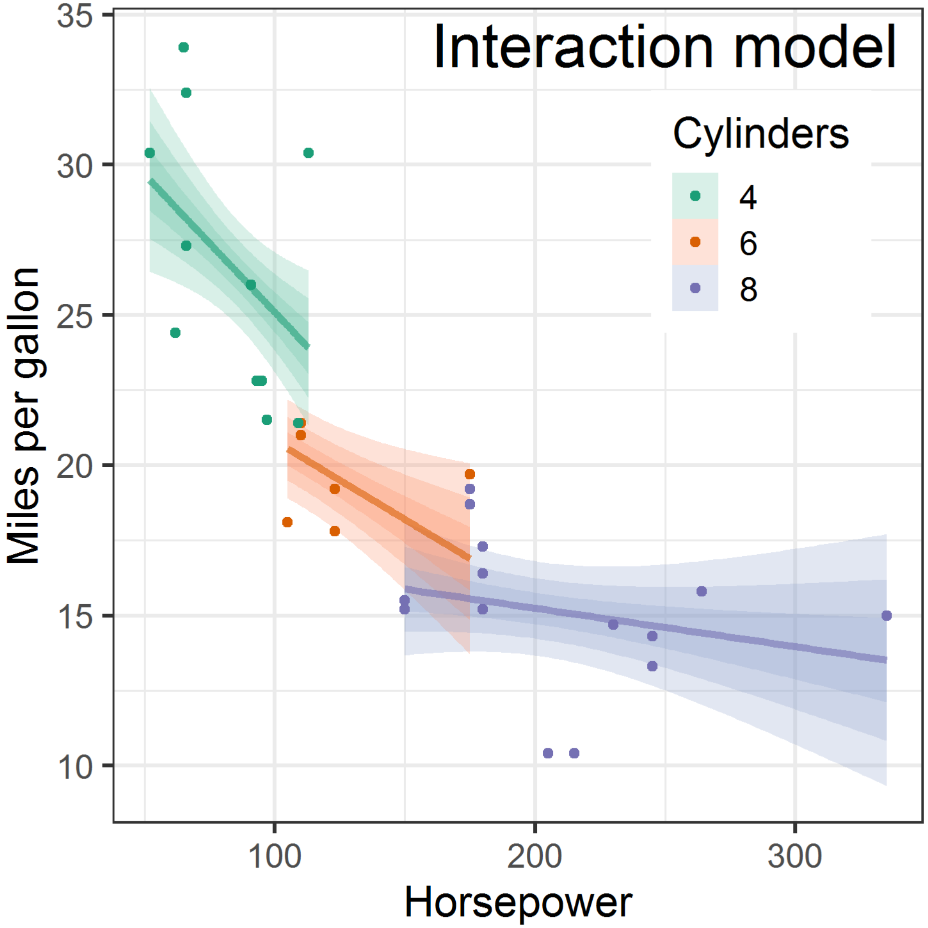


In each case, the confidence band gives visual evidence for uncertainty of the predicted values.

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

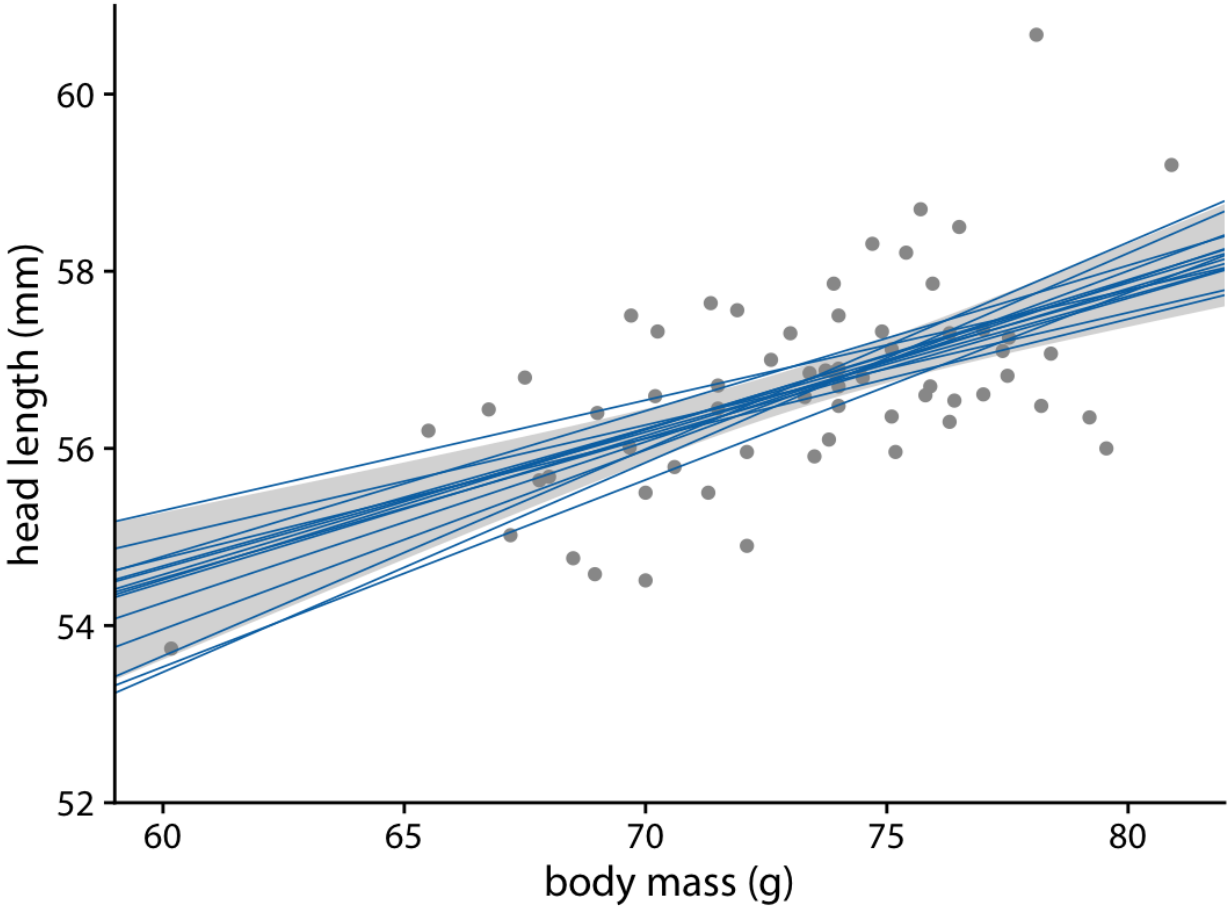


The non-linear relation between hp & mpg can (arguably) be better explained by a model that allows different slopes for 4, 6, 8 cylinders.

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.

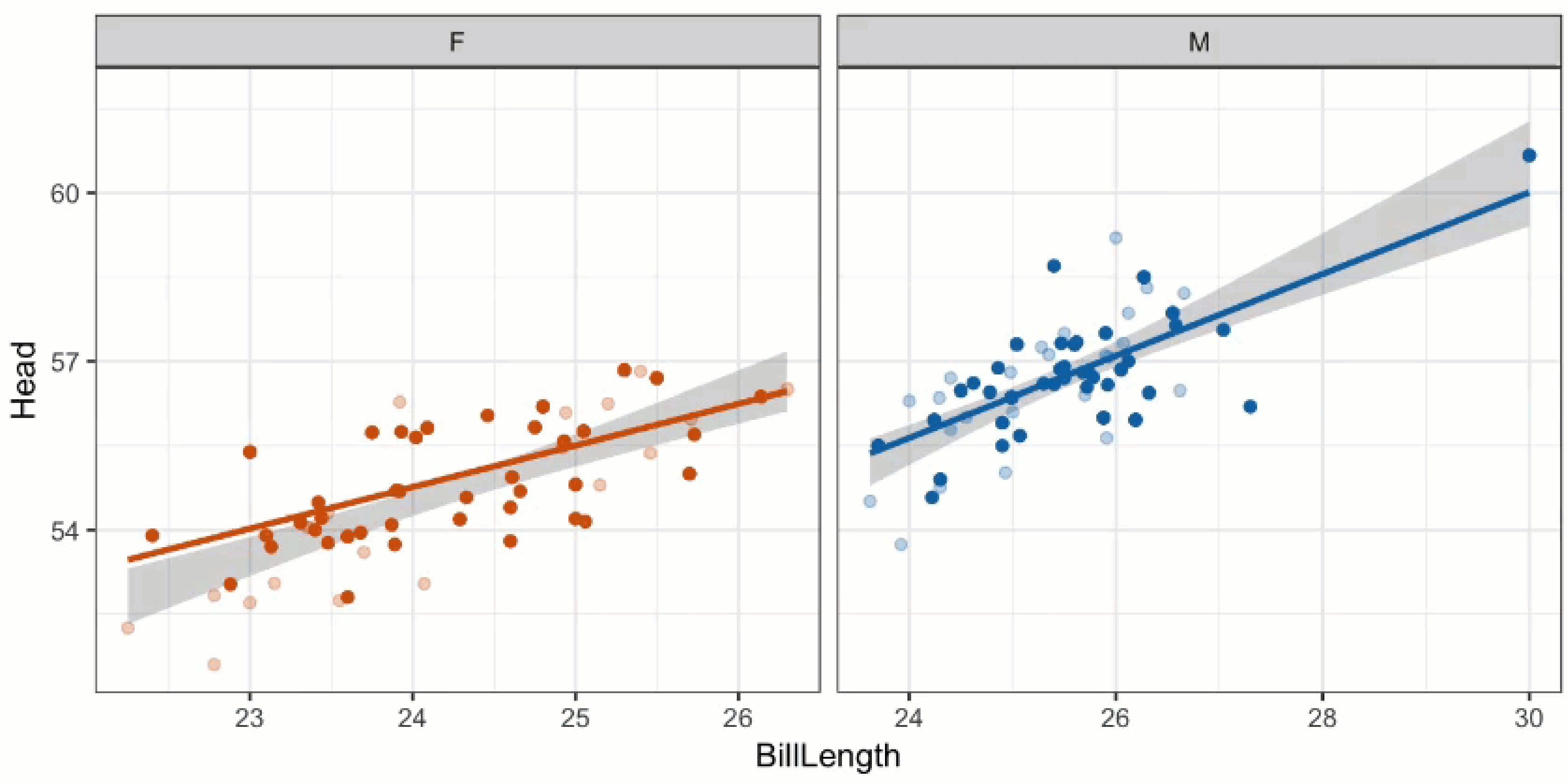


Animation to understand uncertainty

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

- Sampling distributions, simulations, Bayesian posterior distributions, …

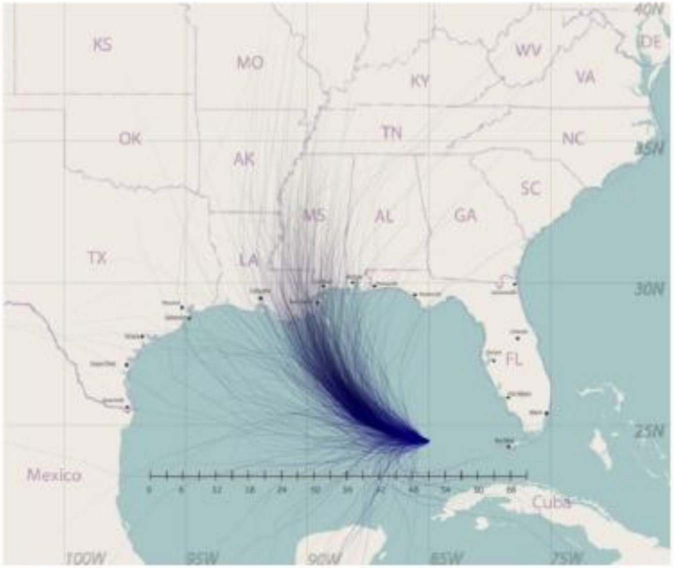
Sometimes useful to appreciate the variability with animated graphics



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"



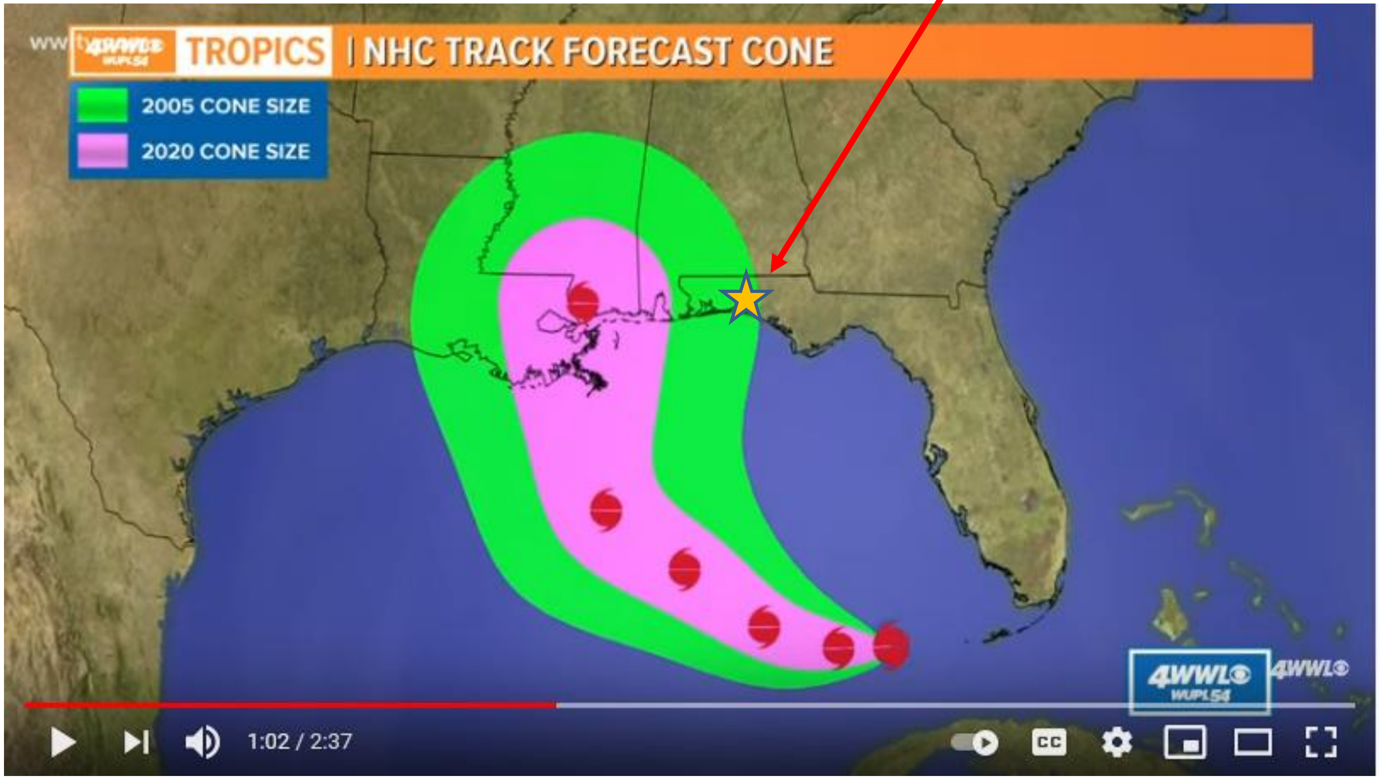
(a) Storm path ensemble

(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?

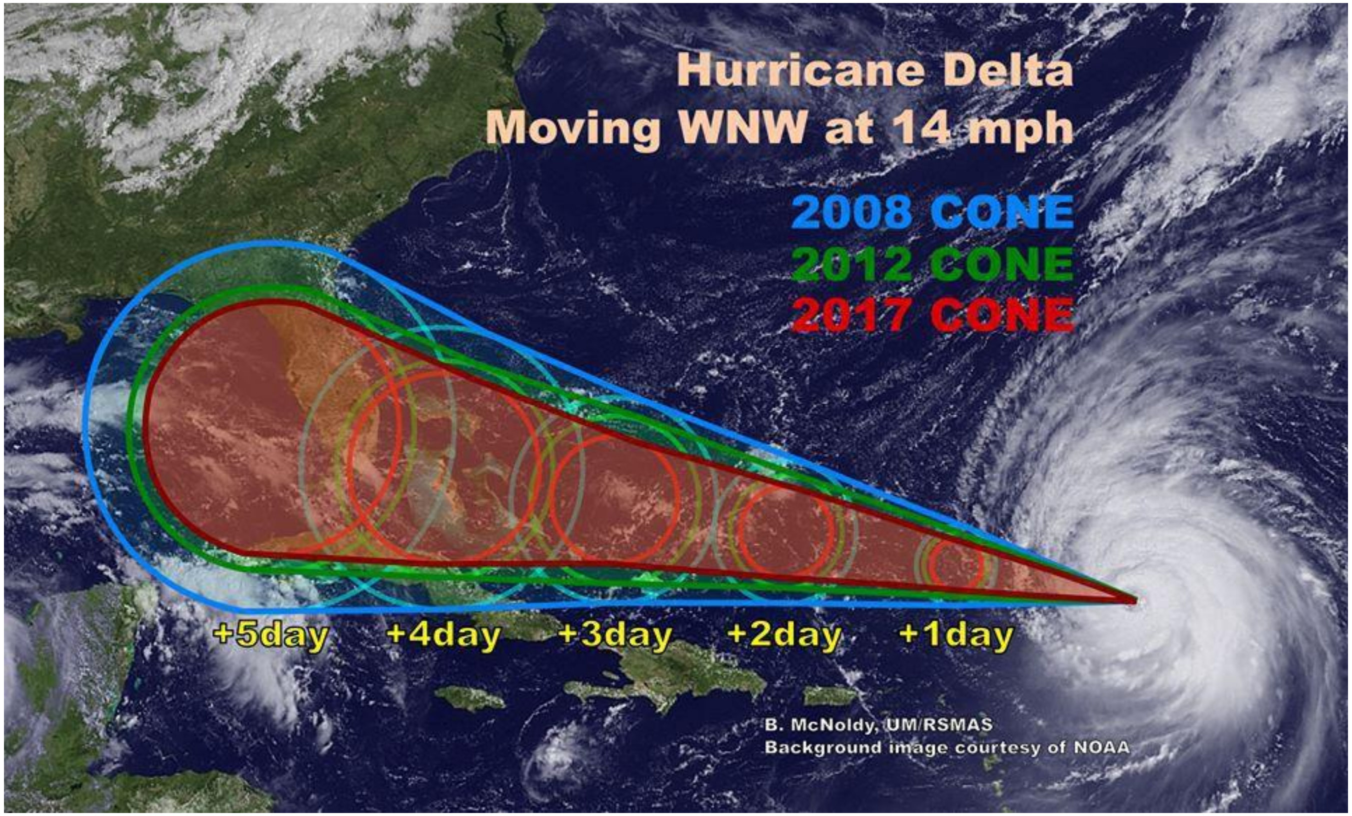


From: https://www.youtube.com/watch?v=nE6XaHtpm04

The Incredible Shrinking Cone

Changes in presumed accuracy are often shown as below

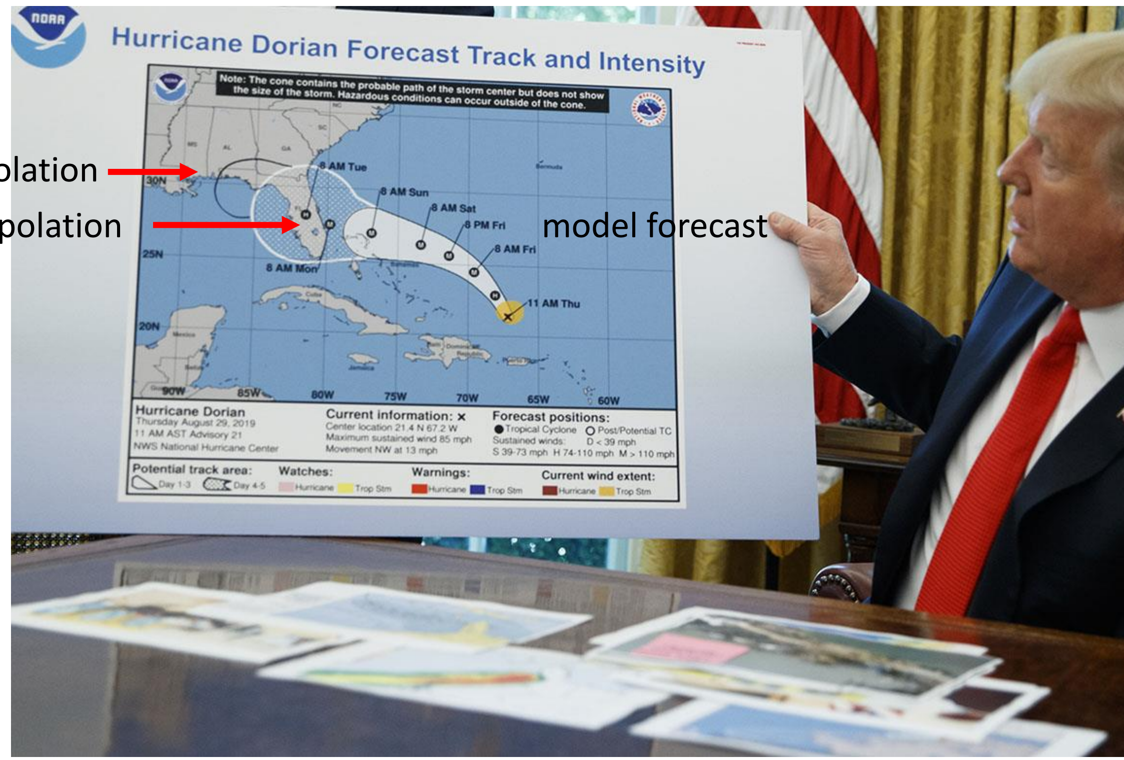
- The cone represents the probable track of the center of a tropical cyclone, formed by enclosing the area swept out by a set of circles along the forecast track (at 12, 24, 36 hours, etc).
- The size of each circle is set so that two-thirds of historical official forecast errors over a 5-year sample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



Sharpiegate

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

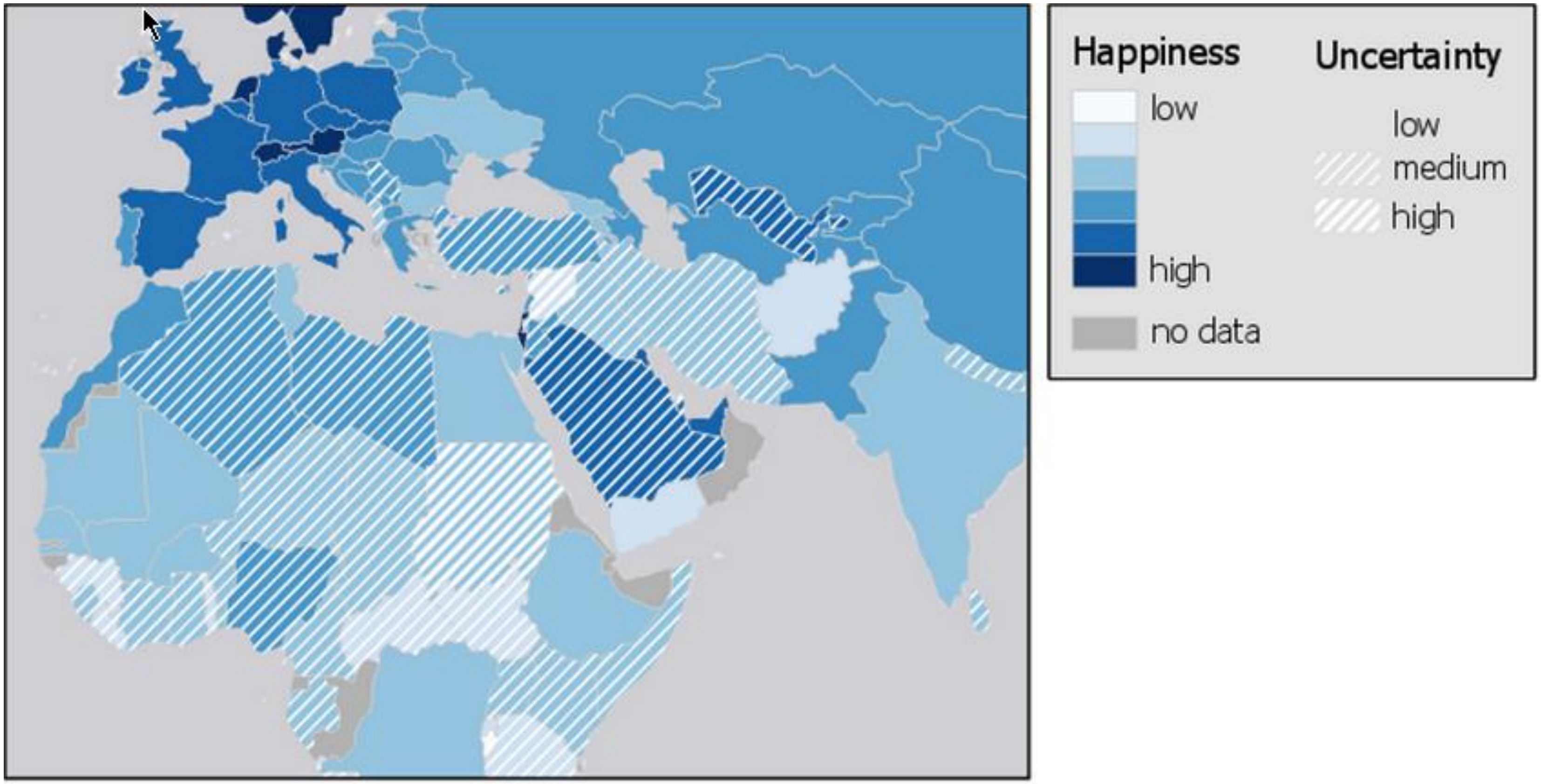
- He had earlier predicted it would hit Alabama & Georgia.
- Let it be said, let it be written (with a sharpie)



From: https://www.theguardian.com/us-news/2019/sep/05/trump-hurricane-dorian-alabama-map-sharpiegate

Coding maps for uncertainty

In choropleth maps we can show uncertainty with another attribute



From: https://www.e-education.psu.edu/geog486/node/693

Summary

- Uncertainty is fundamental to data analysis & models
- Showing variation in distributions a basic problem
  - histograms, density plots, boxplots
  - Better: violin, raincloud, …
  - {ggdist} offers many alternatives
- Error bars: many flavors; can show multiple intervals
- Bayesian methods, bootstrap, simulation
  - Different methods, but similar ways to show uncertainty
- Geographic data: need to be careful about interpretations