An a journey through uncertain lands MVE080/MMG640 Lecture 5

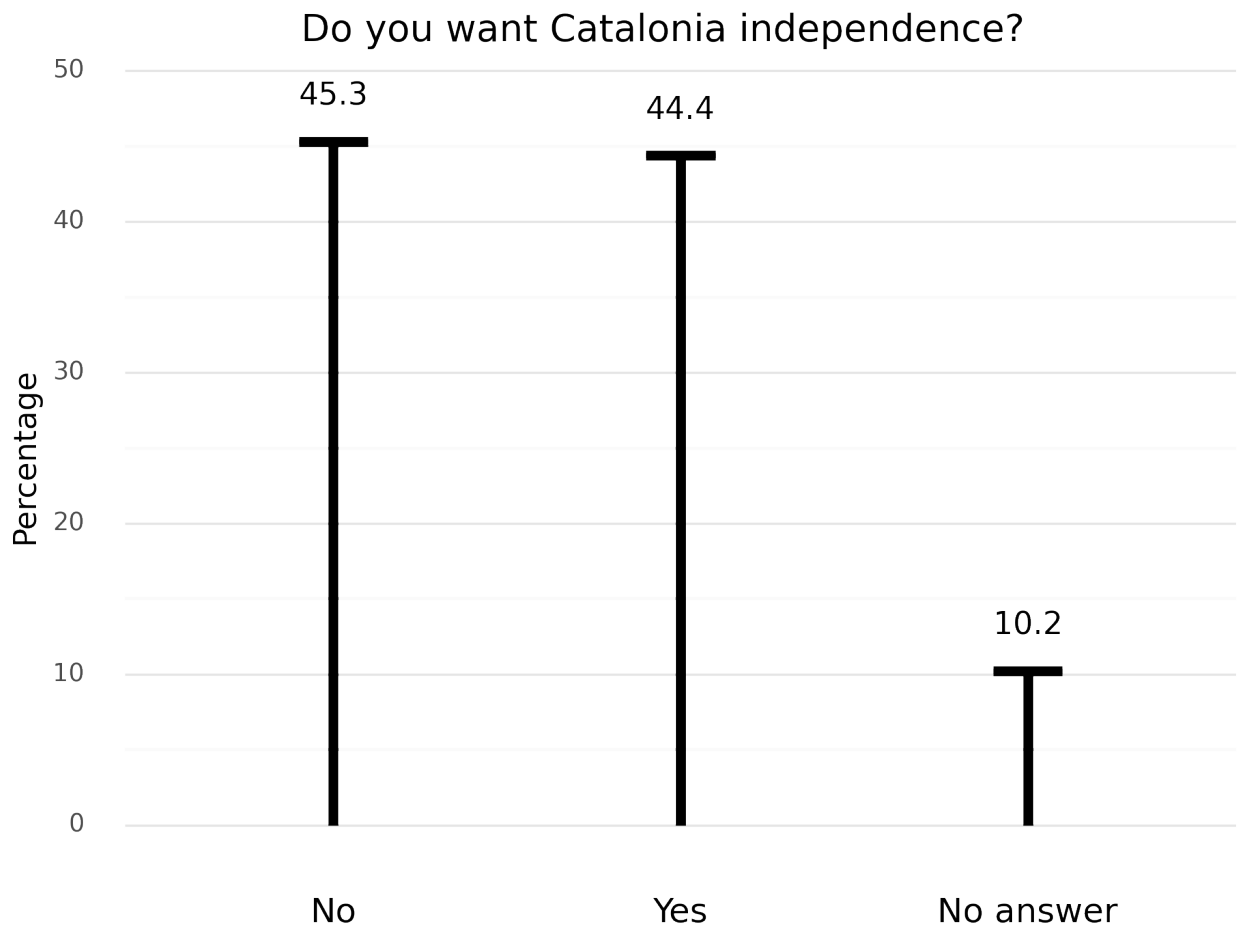
Sebastian Persson
sebpe@chalmers.se

November 15, 2022

Uncertainty

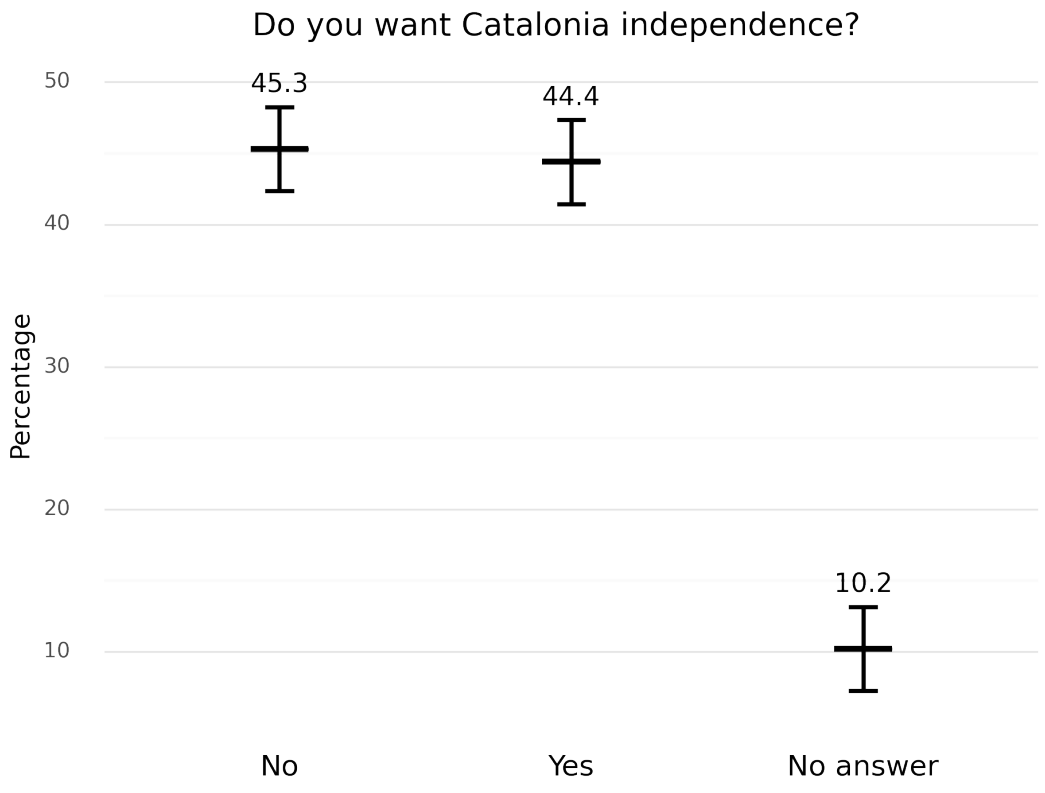
From last lecture ...

- Why is this a bad visual?



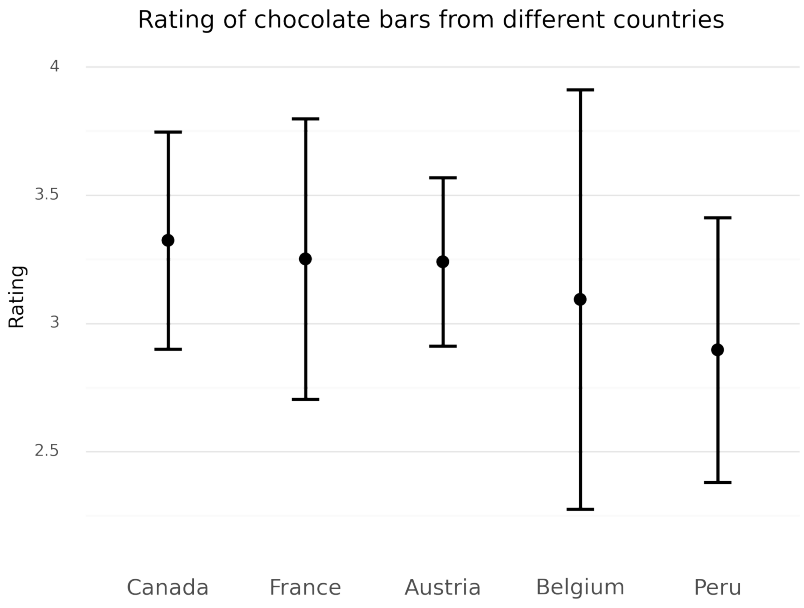
Always report uncertainty for point estimates

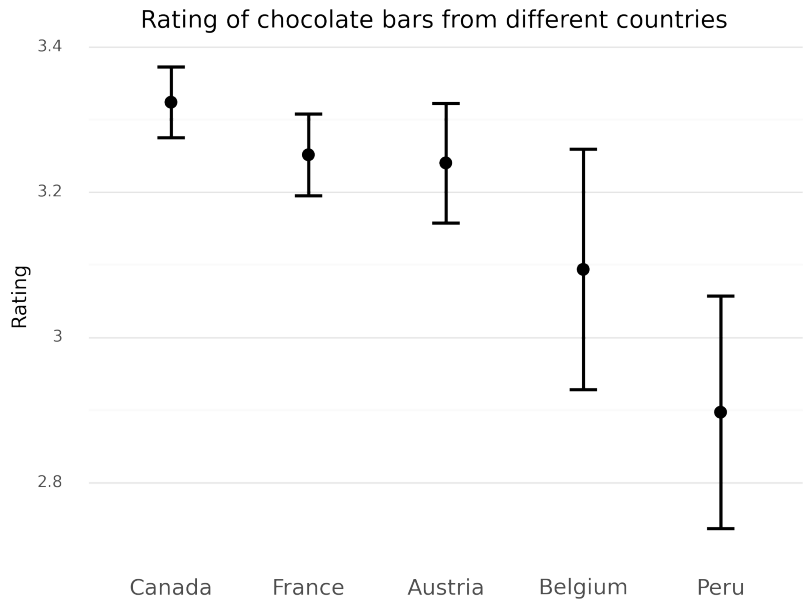
Often we cannot observe the entire population. This means we must from a finite sample estimate a property (e.g party sympathy). Such estimates have an associated uncertainty.



Error bars : Curse or blessing?

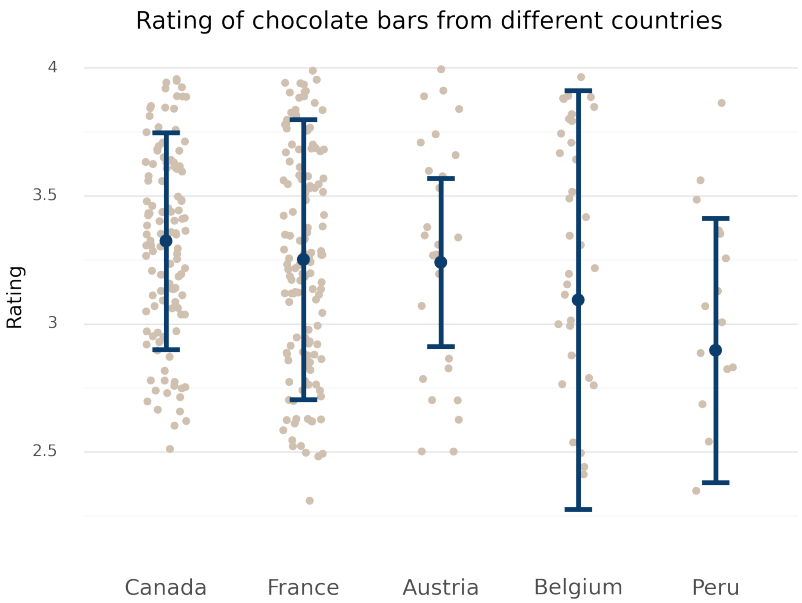
- Would you rather have chocolate from Belgium or Canada?

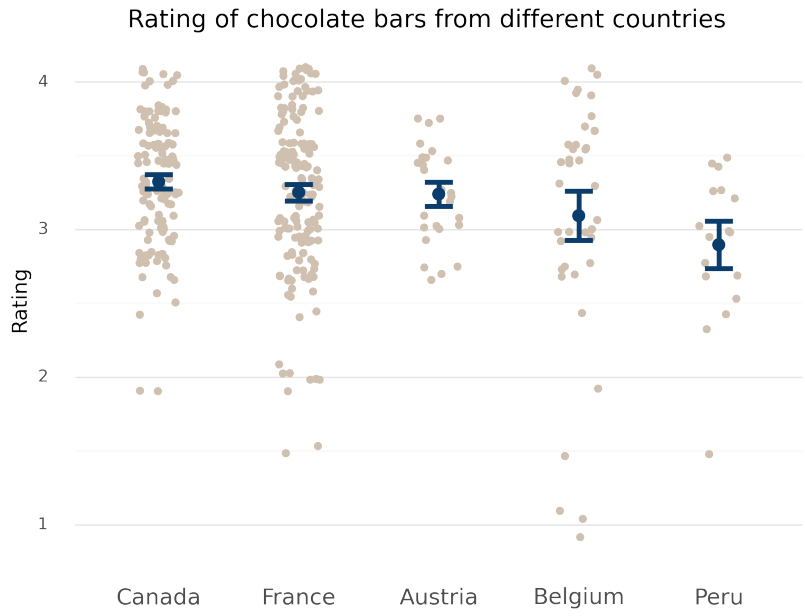




Error bars : Curse or blessing?

- Would you rather have chocolate from Belgium or Canada?
- Always state what the error-bars are showing
  - However, even with captions hard to interpret





Standard error and standard deviation

- Given a dataset $x_1, \ldots, x_n$ where we assume $X_i \sim \mathcal{N}(\mu, \sigma^2)$ with sample mean

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Sample standard deviation;

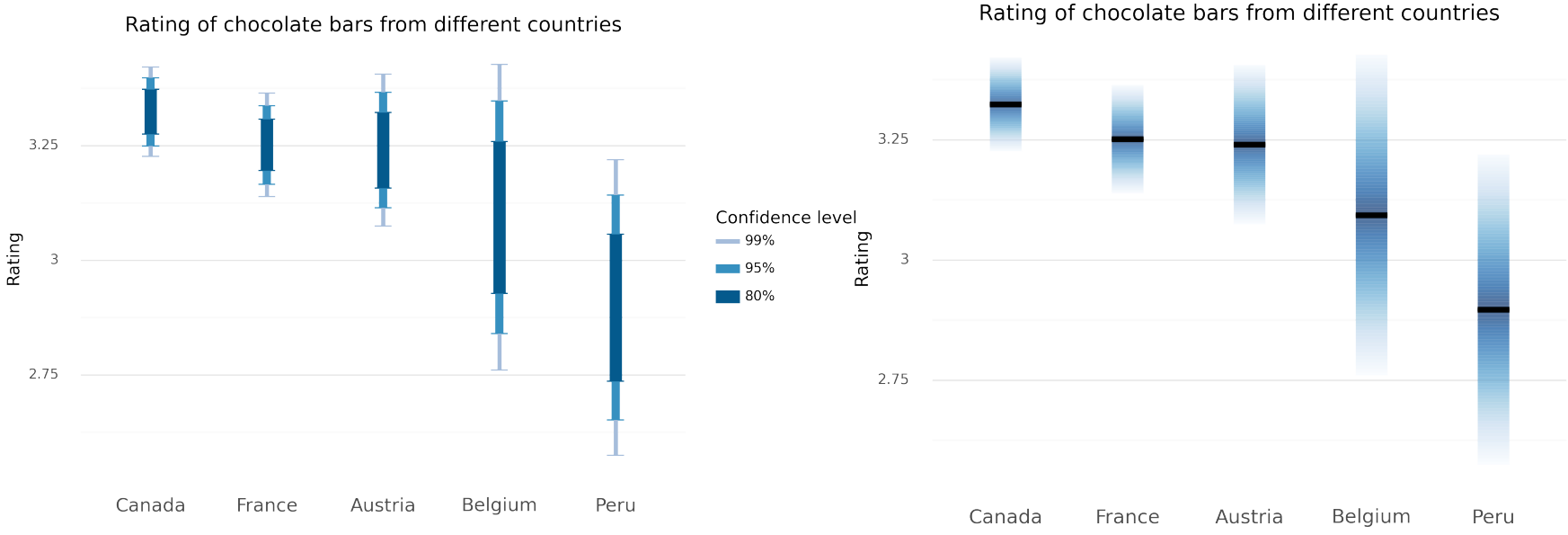
$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{\mu})^2}$$

- Standard error for $\hat{\mu}$ (estimate of μ)

$$\hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{n}}$$

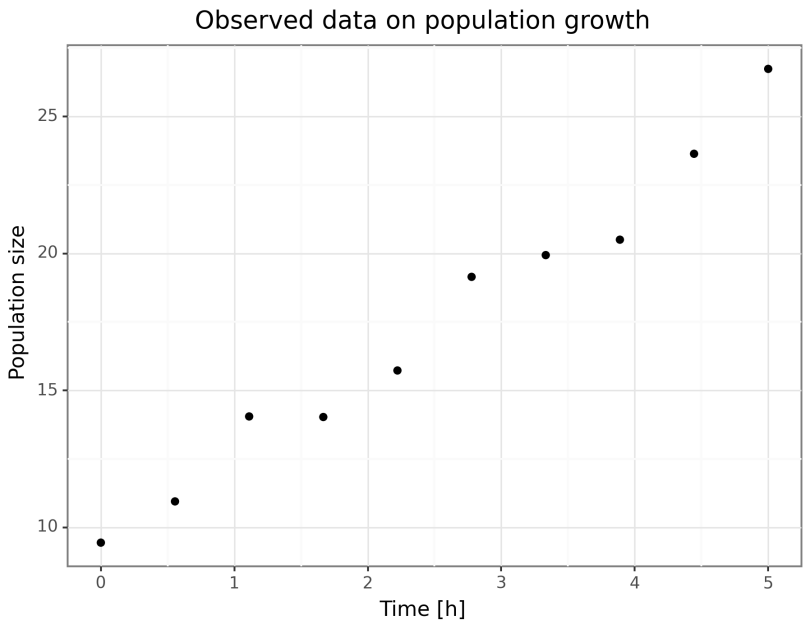
Can we improve error-bars?

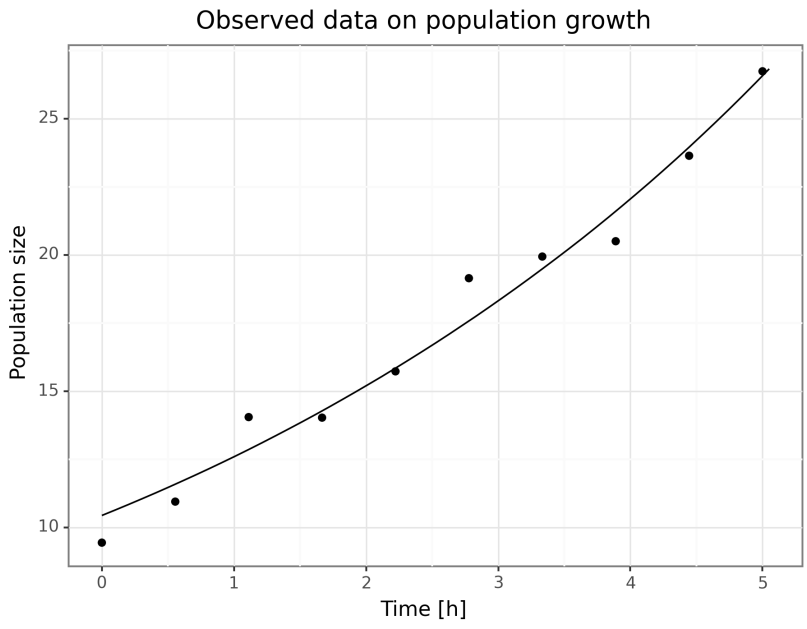
- Add several levels of confidence to remind the viewer of the uncertainty in the estimate
- Fuzziness to trigger intuitive sense of uncertainty
  - Might be better for lay audience



Uncertainty and curve fits

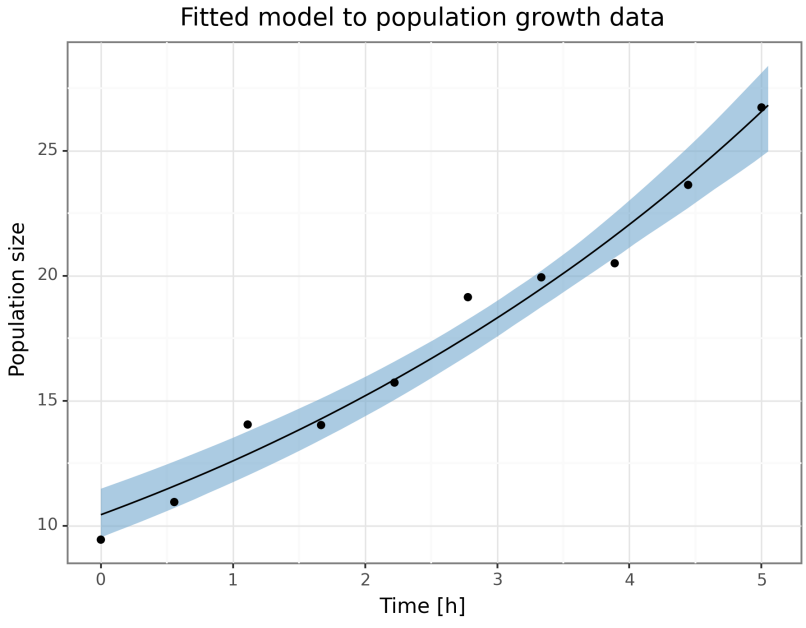
- Lets say we want want to fit some growth data to the model $N = N_0 \exp(rt)$ in order to predict the number of bacteria 5 hours from now

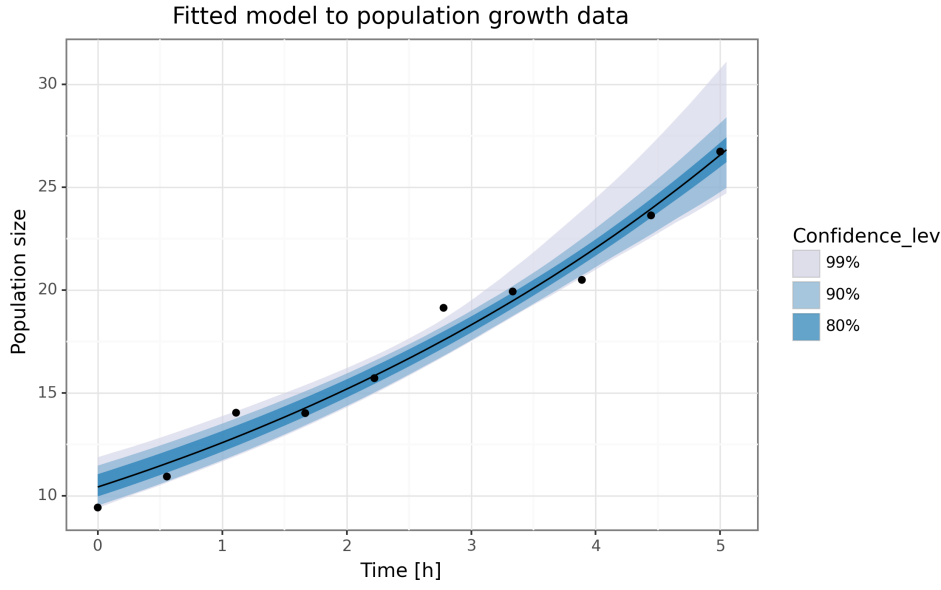




Uncertainty and curve fits

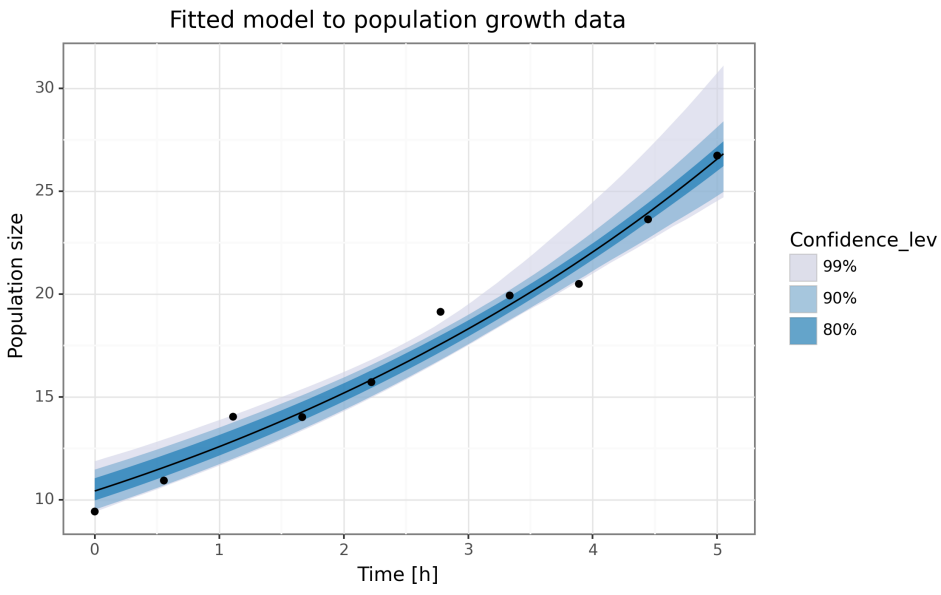
- Confidence bands have similar problems as error-bars
- Adding several layers can help

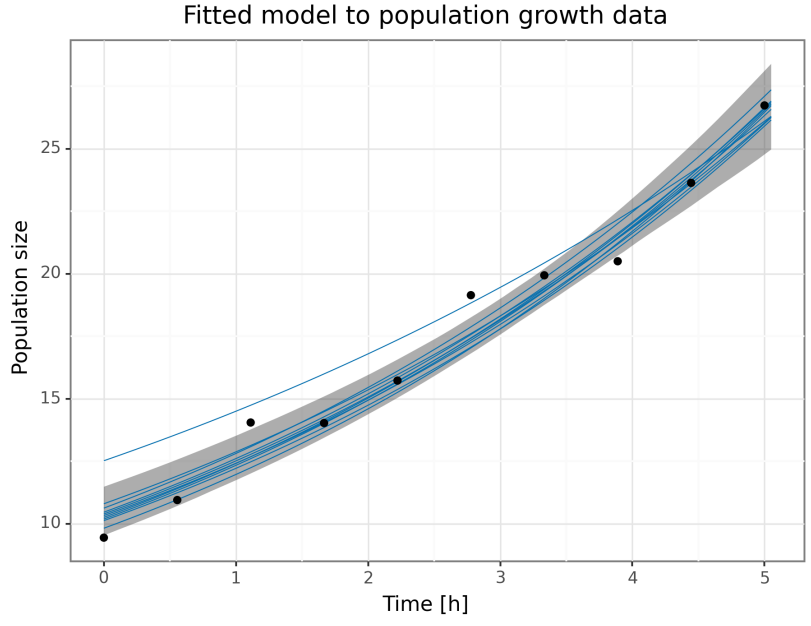




Uncertainty and curve fits

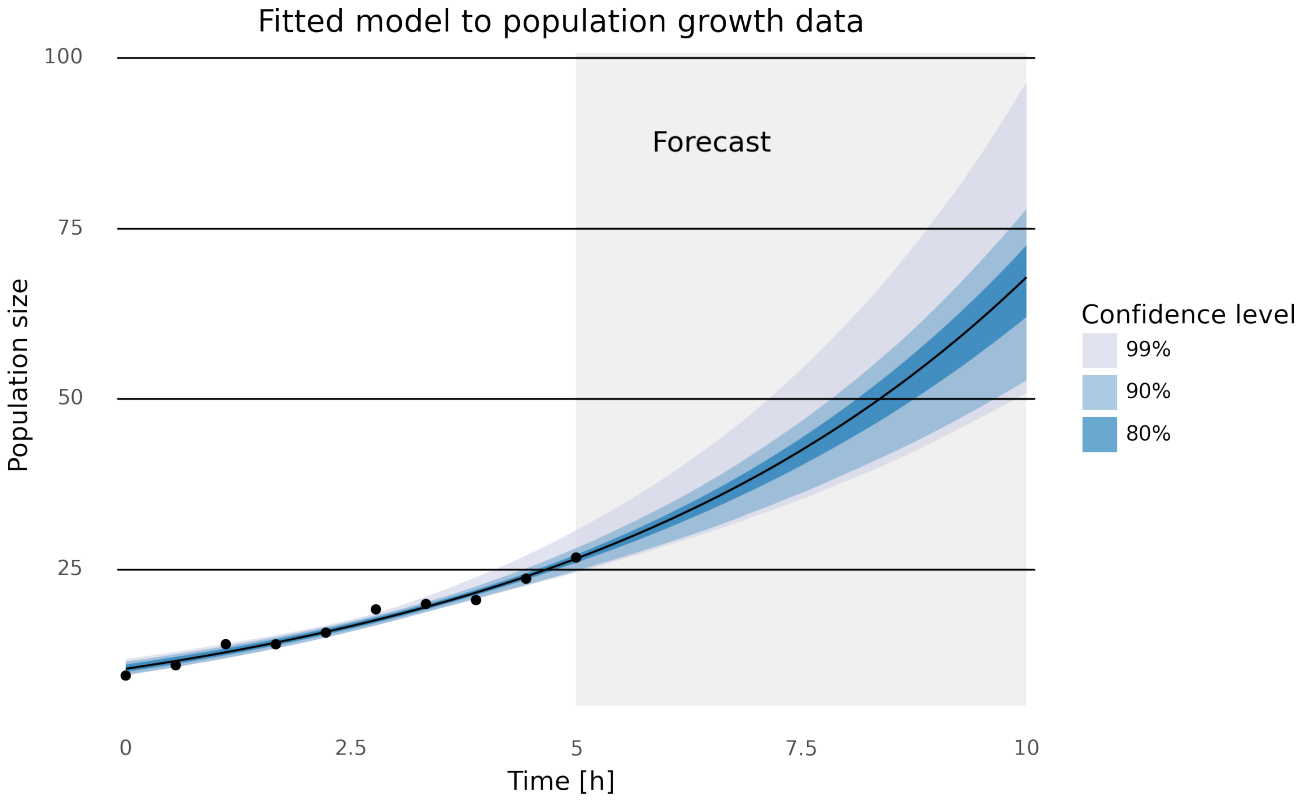
- Adding several layers can help
- Alternatively show several representative fits





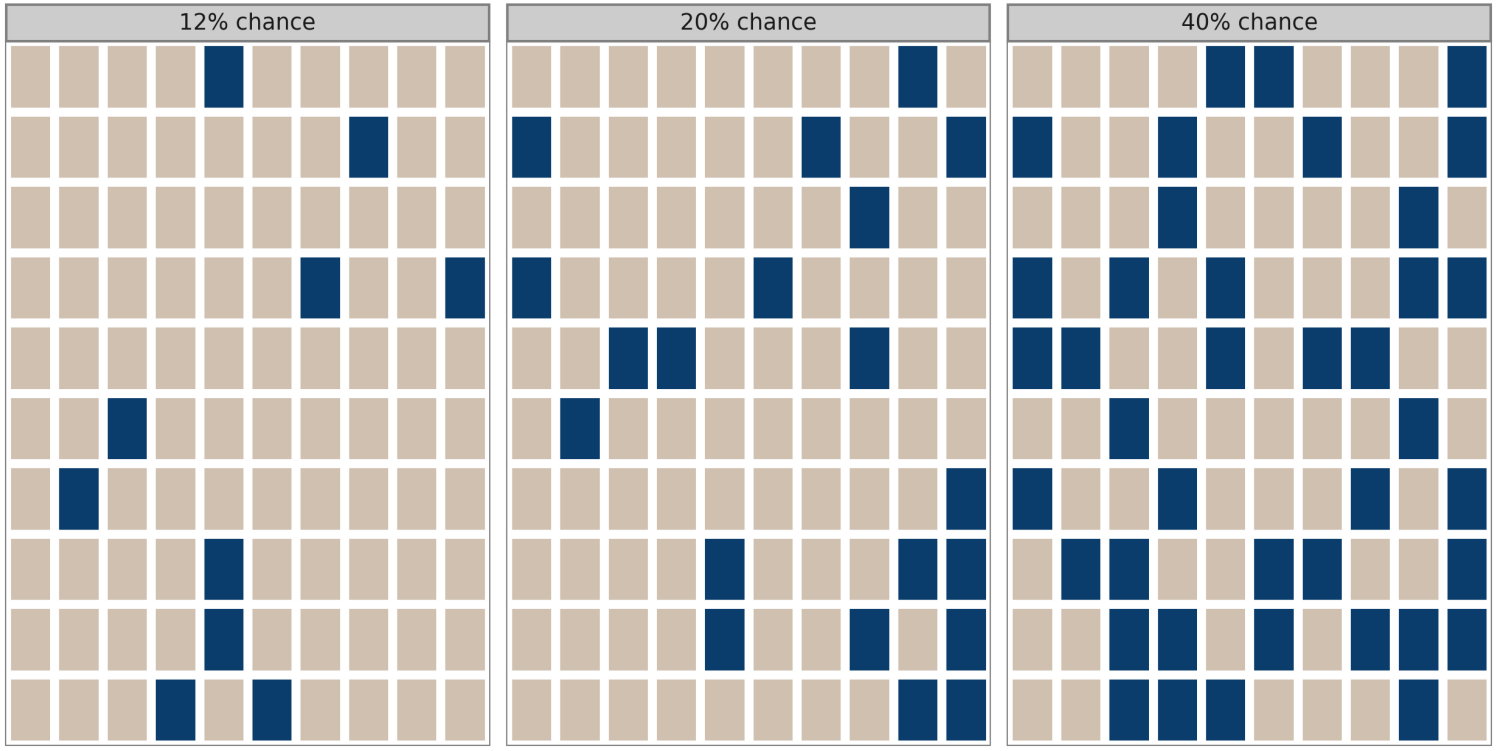
When making model predictions

Try to always make it clear what is a model prediction



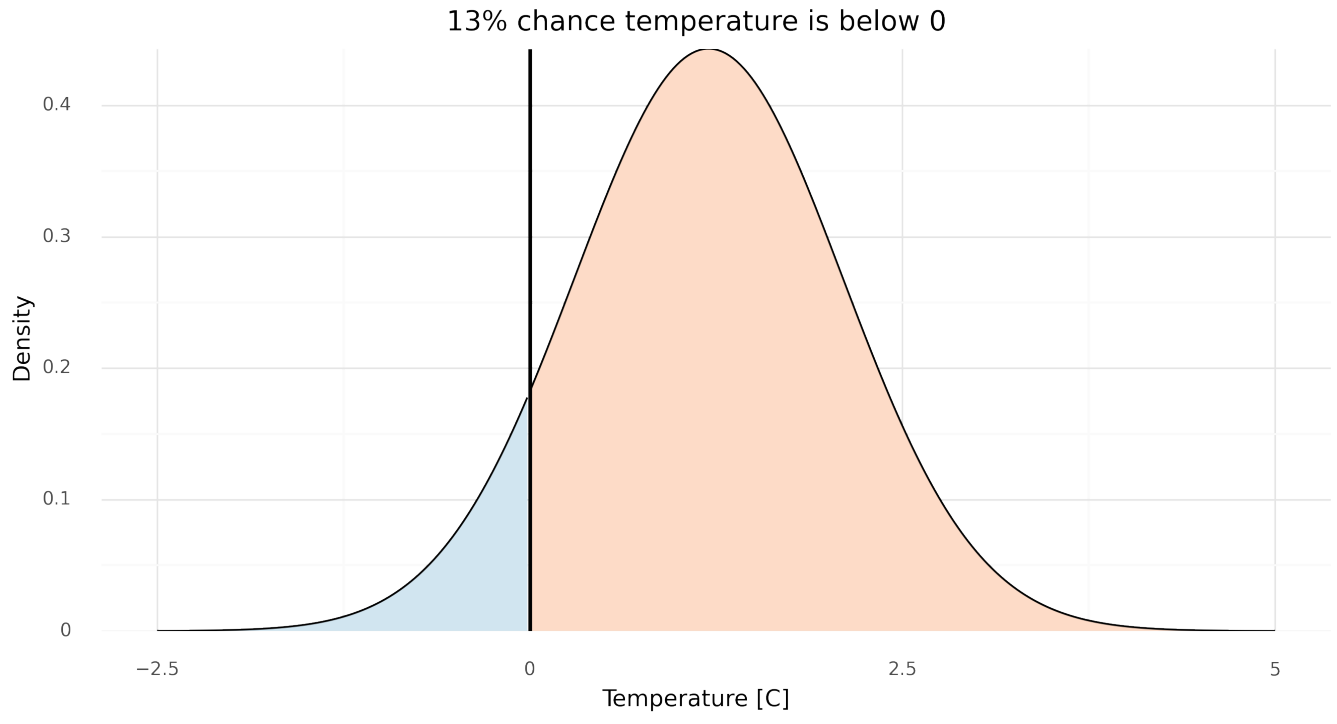
How do we visualise probabilities?

- Cancer1 : 12 out of 100 dies
- Cancer2 : 200 out of 1000 dies
- We are good at discerning differences when counting



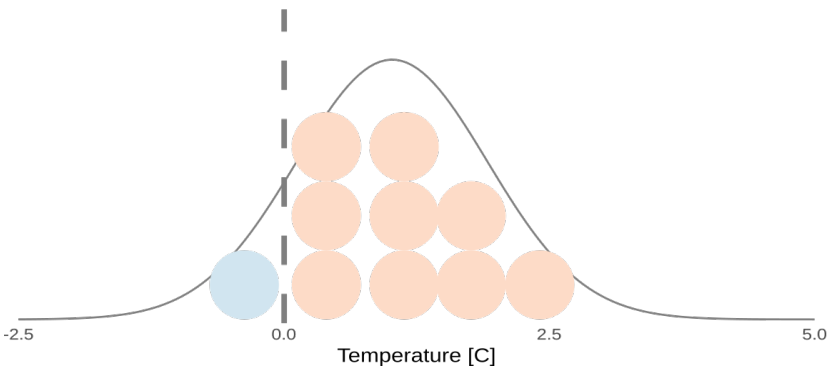
How do we visualise densities?

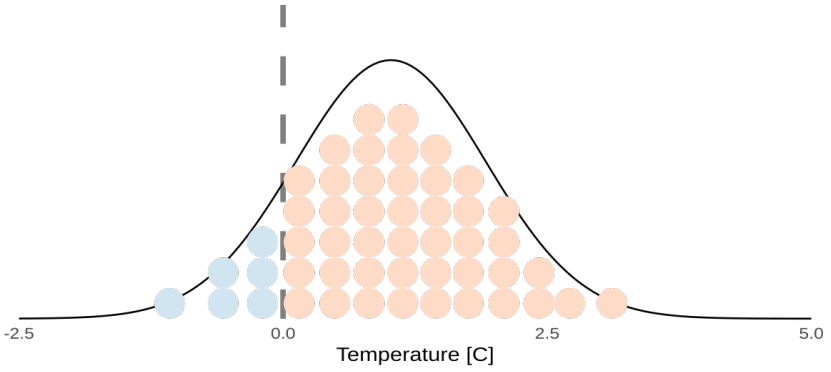
- There is a 13% chance that the temperature goes below zero.
- Densites are great when we also want to see magnitudes.



How do we visualise densities?

- We are better at counting than seeing area
  - Sacrificing precision might be worth it



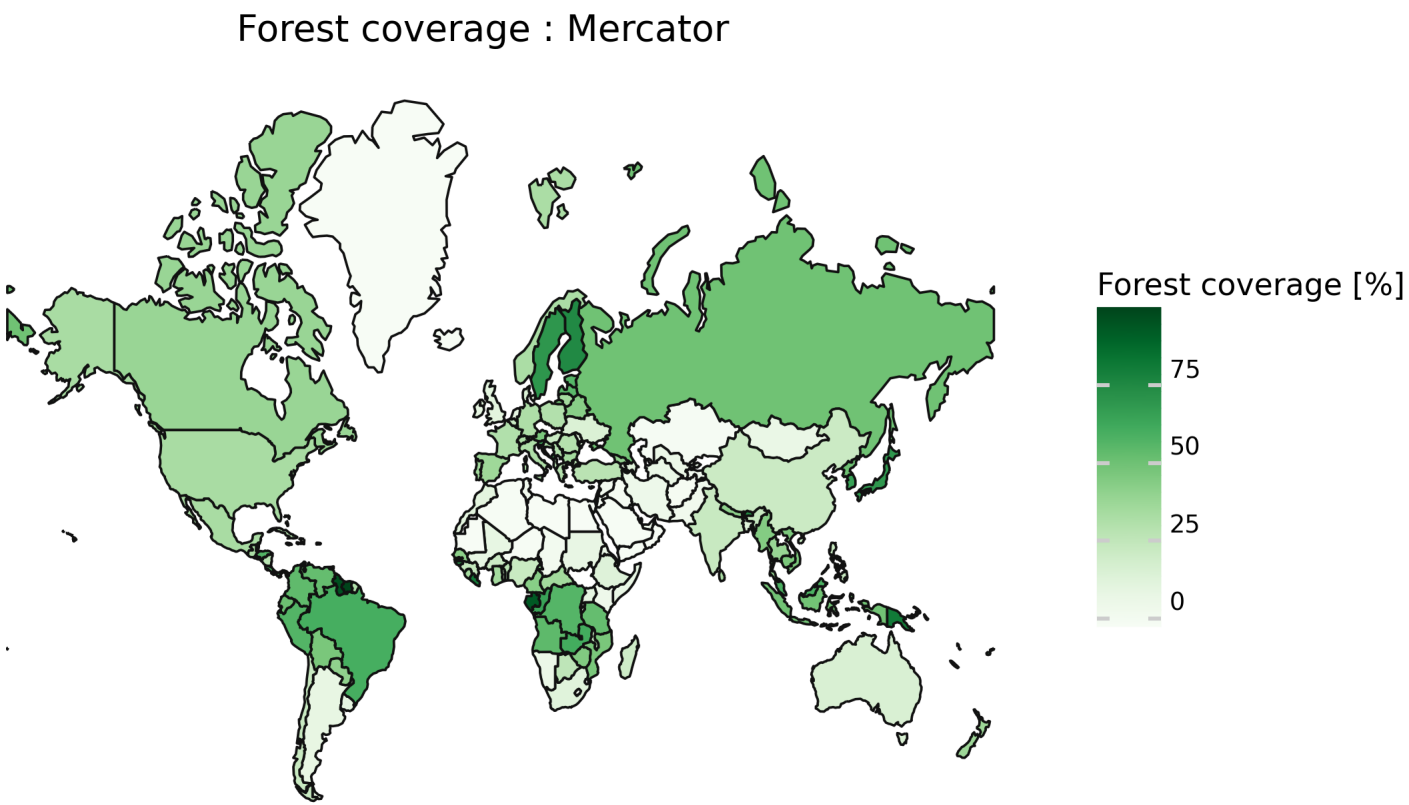


Geospatial data

Earth is a Möbius strip

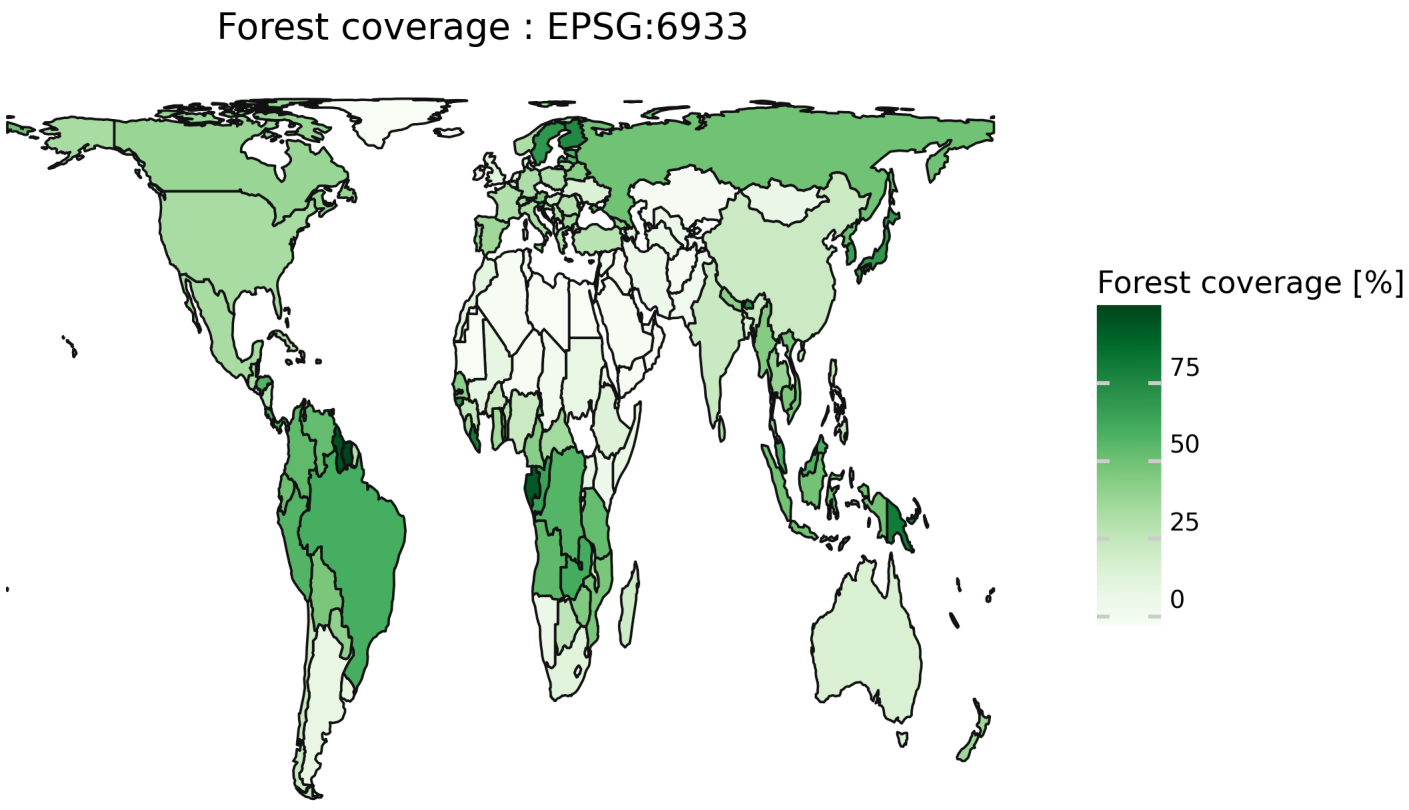
Every map is a projection

- Can preserve either angles or area
- Mercator
  - Preservers local directions and shapes (Google Maps)
  - Exaggerates size with increasing latidude



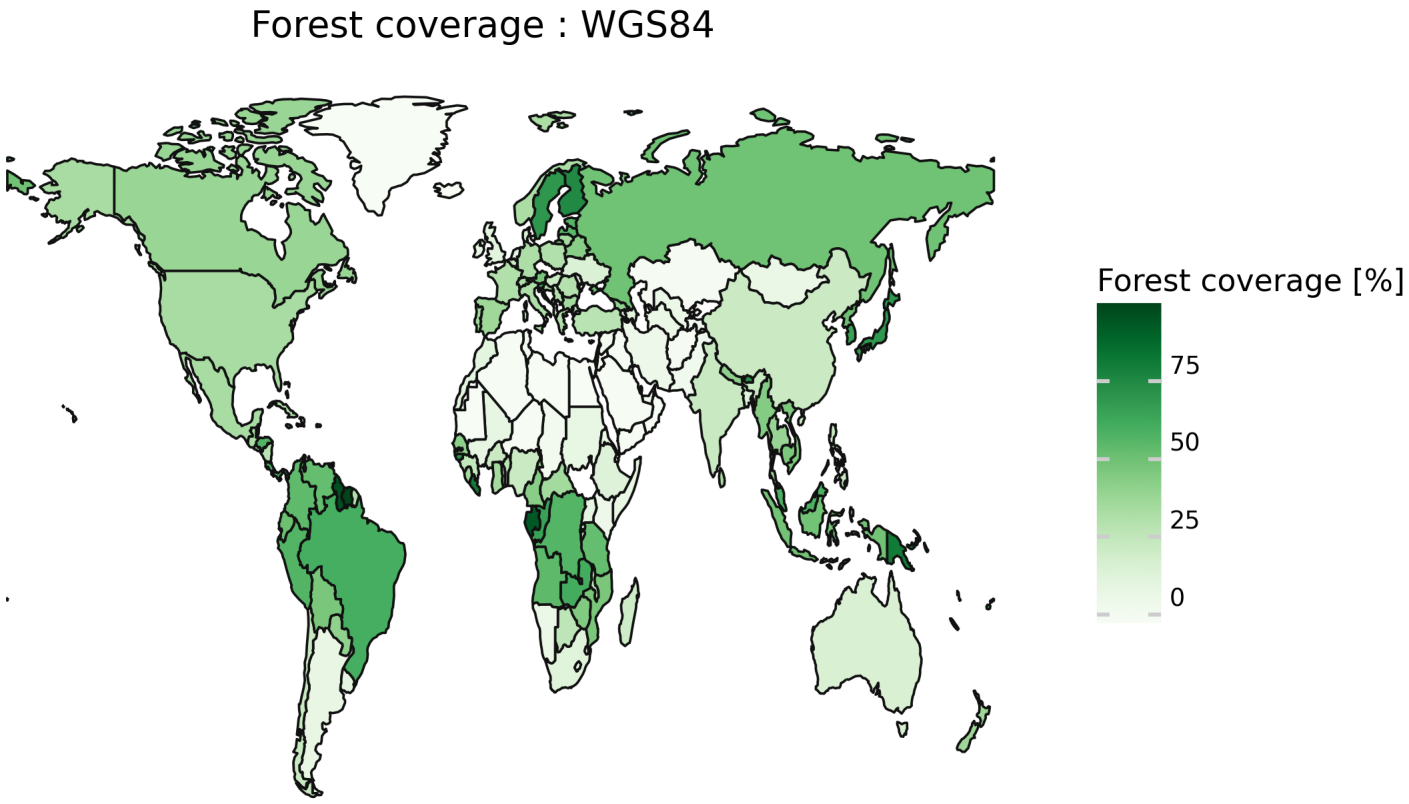
Every map is a projection

- EPSG:6933
  - Preservers area
  - Heavily distorts shapes



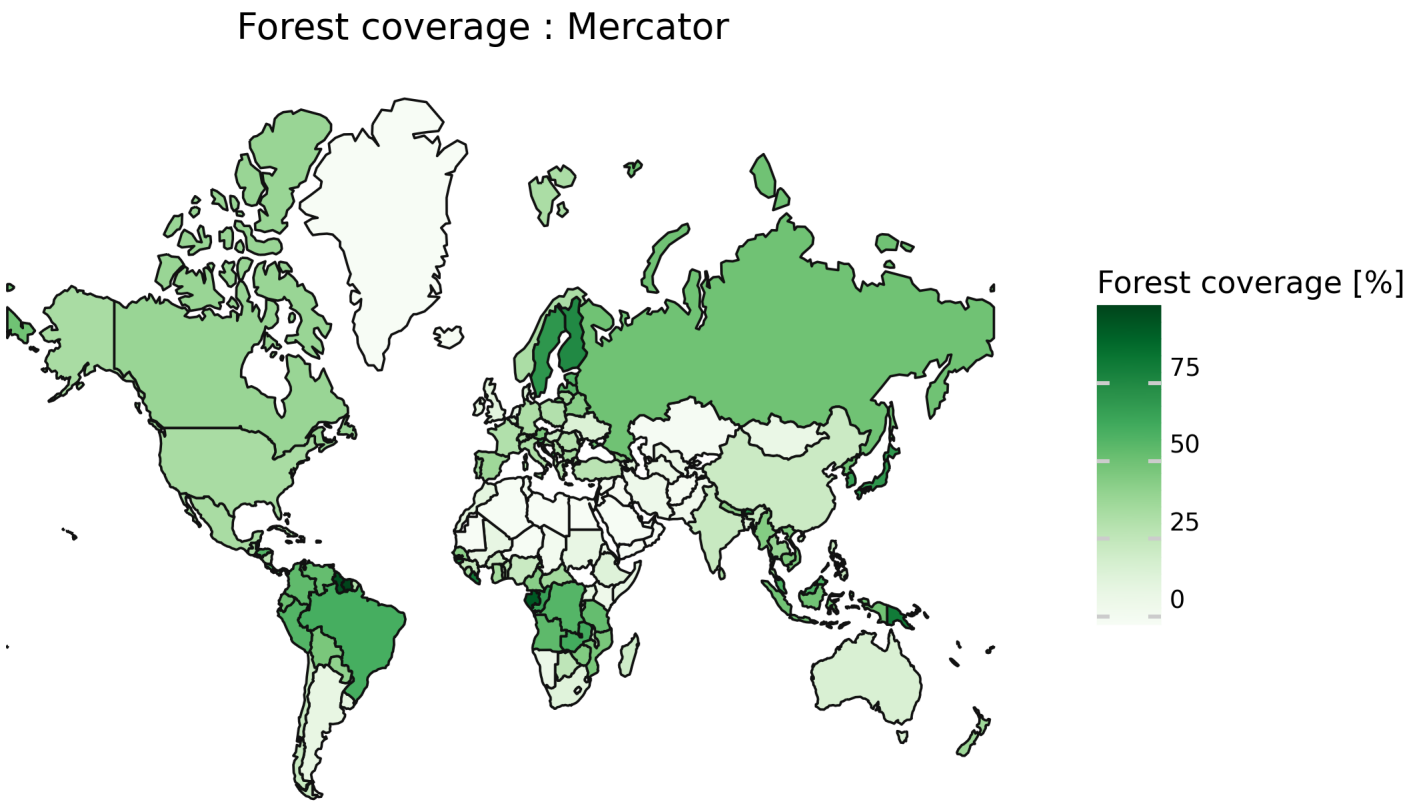
Every map is a projection

- Plate Carrée
  - Neither preserves angles or area (compromise)
  - Often used for rasterised images (not to wrong in any direction)



Which projection to use?

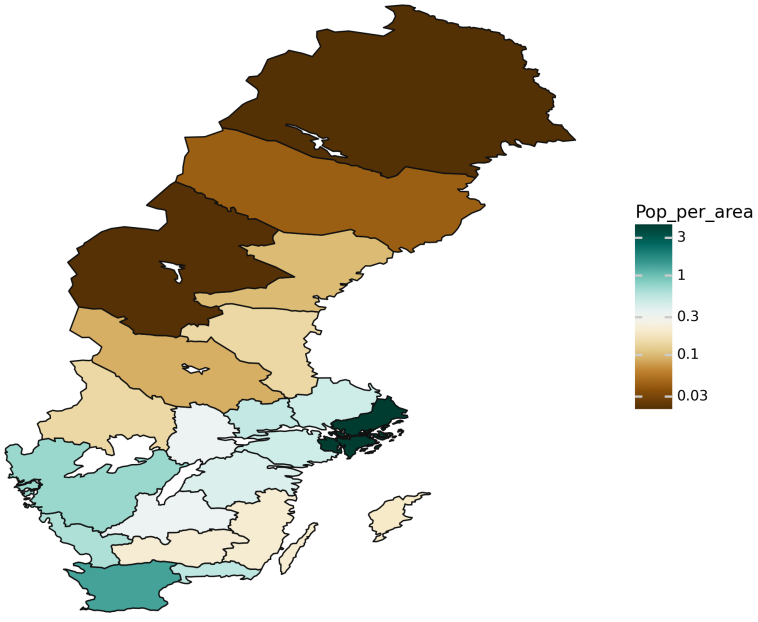
- Choose according to task
  - If size matter (e.g parts of world covered in forests) do not use Mercator
- *https://www.thetruesize.com/*



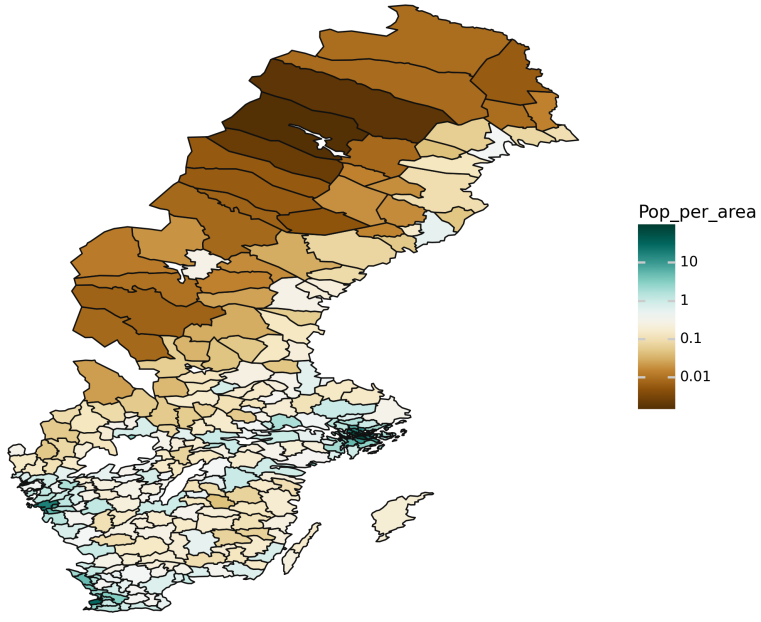
Choropleths - easy to misinterpret

- Color parts of map according to criteria
- Divide by area (we perceive larger area as larger effect)
- Choose appropriate granularity

Sweden län population density

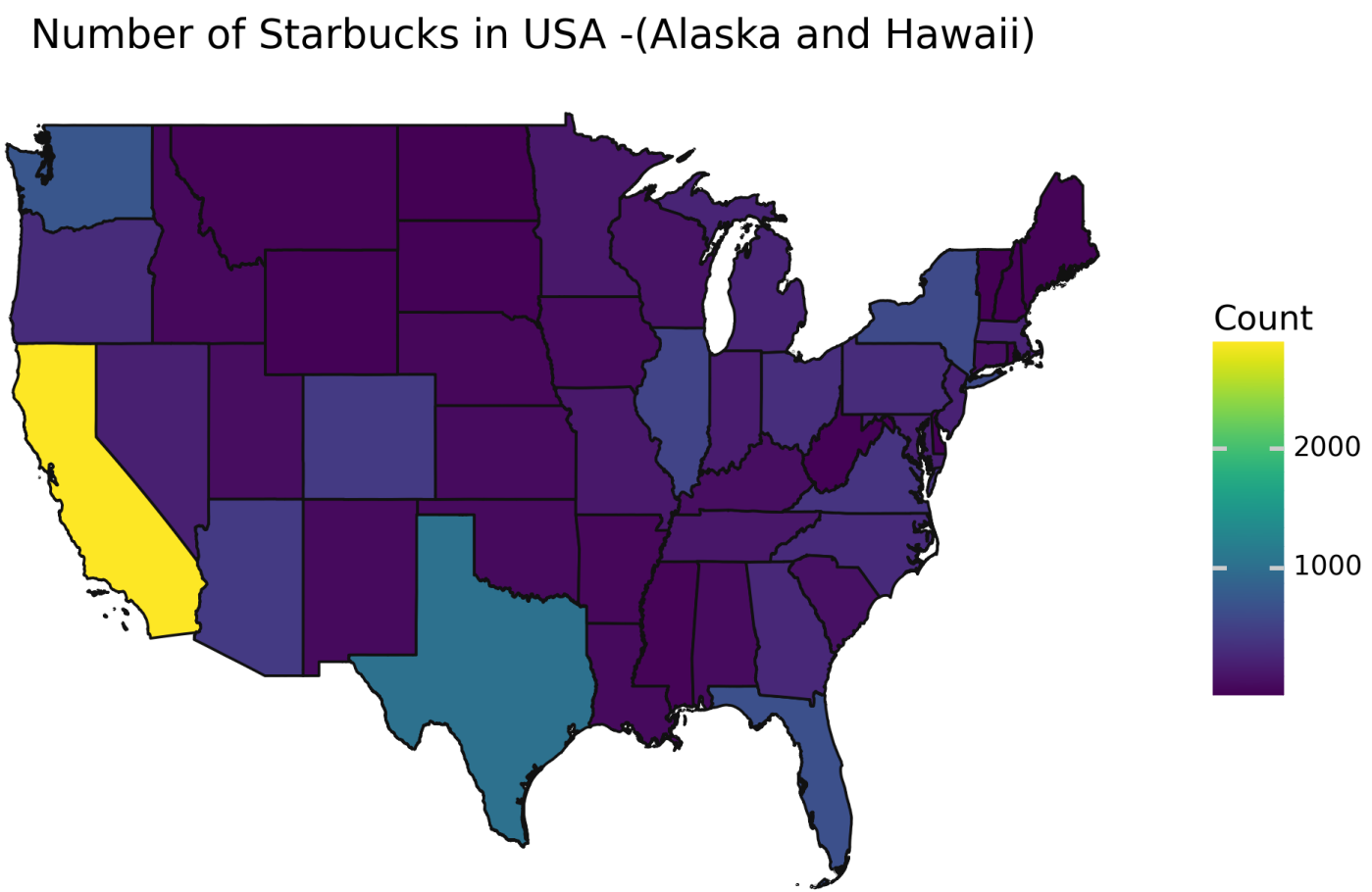


Sweden municipal population density



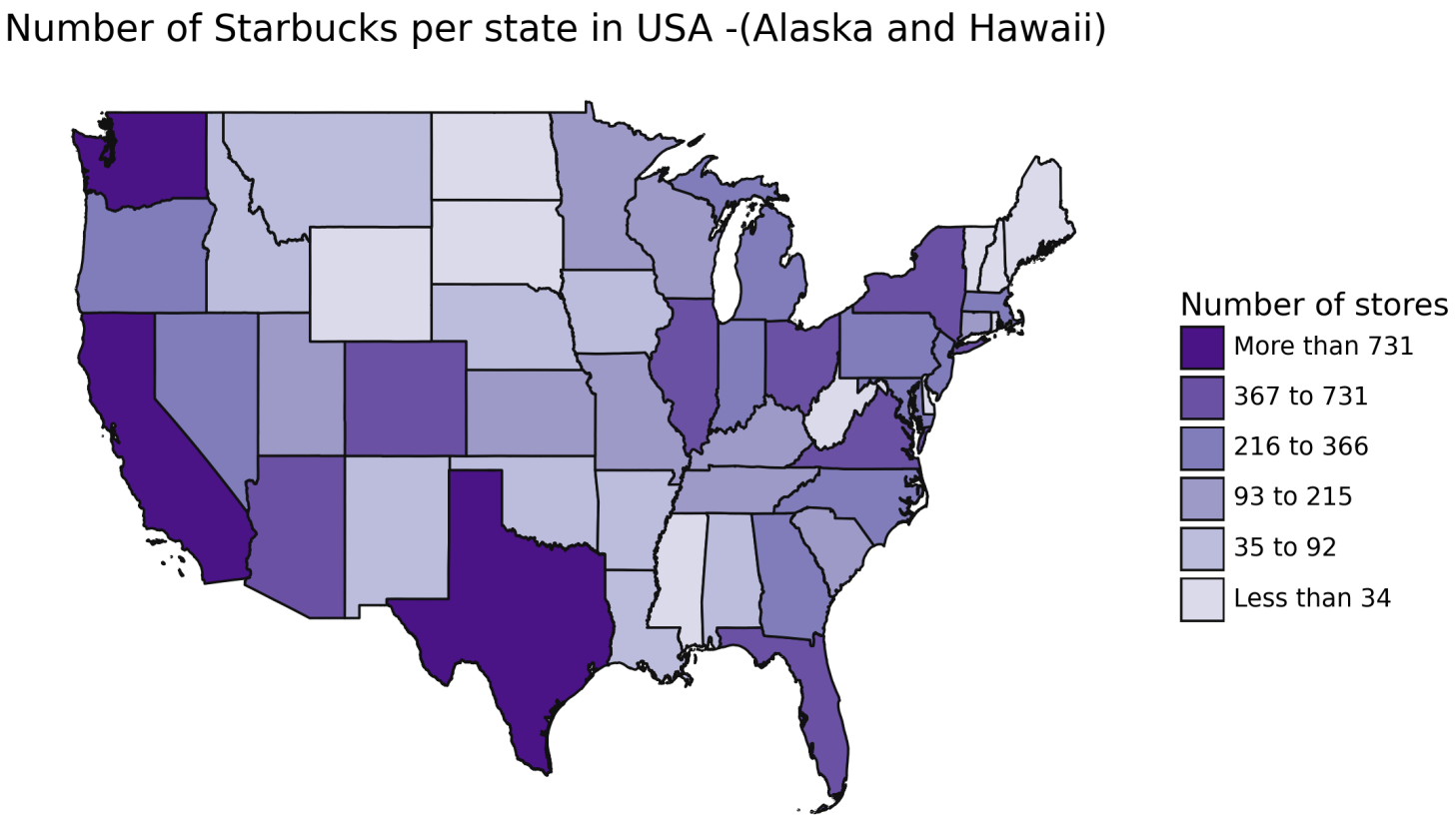
Choropleths - easy to misinterpret

- Choose appropriate granularity
- Are there many Starbucks in Seattle?
  - Calibrating the colour-scale for a map is an art



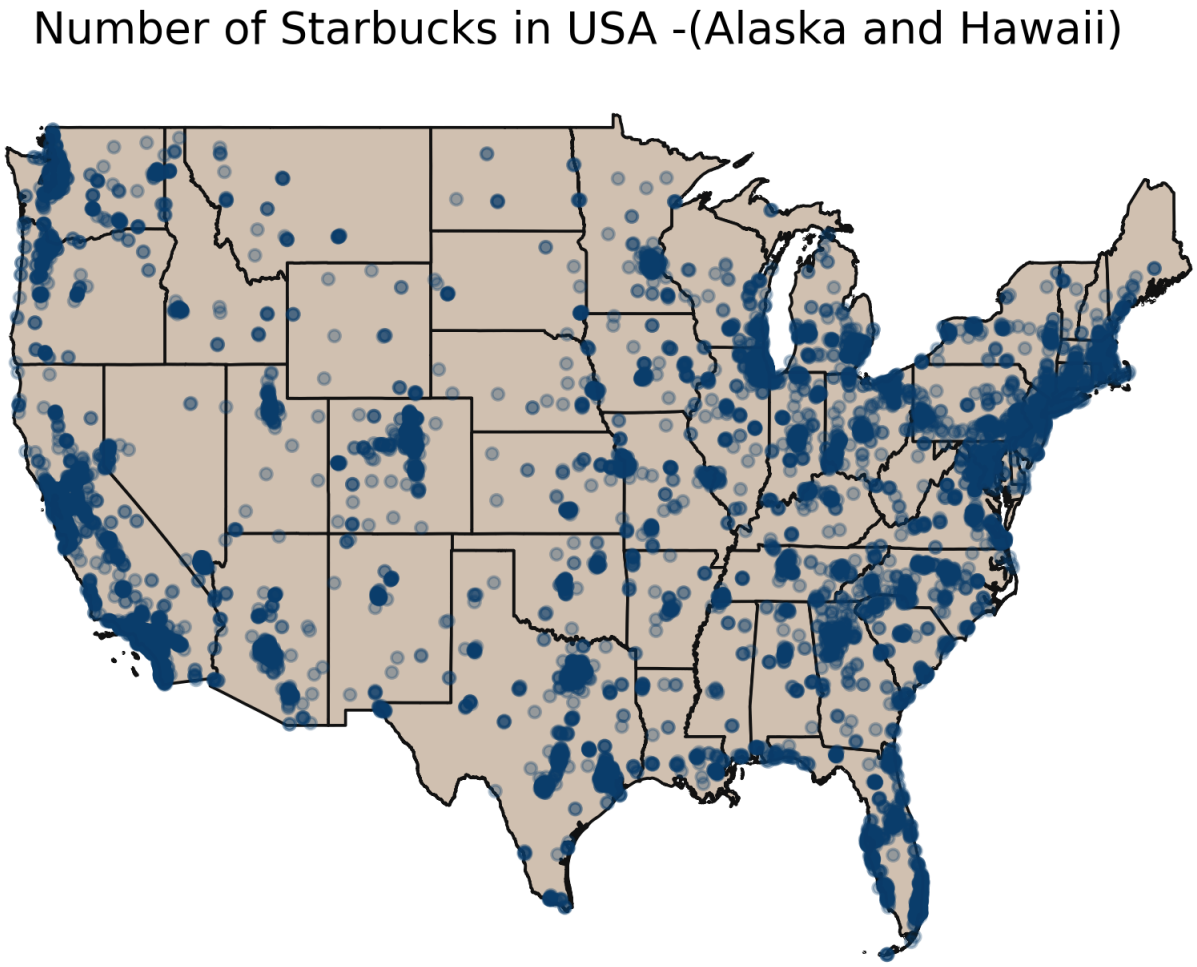
Choropleths - easy to misinterpret

- Calibrating the colour-scale for a map is an art
  - Sometimes binning (e.g based on quantiles) helps



Choropleths - easy to misinterpret

- Dots can sometimes be used instead of color
- A lot of lessons from Lecture 3 also apply for maps



Take home messages

- For point estimates uzing ”fuzzy” bars or several level error-bars can help highlight uncertainty. Similar applies for curve fits.
- We have easier time to reason about probability if we can count (e.g compared to area or numbers).
- Choose map-projection based on the task
- Choropleths typically lack details. Can be misgiving if you do not choose granularity and colour-scaling appropriately.

For next lecture...

Why is this a bad visual?