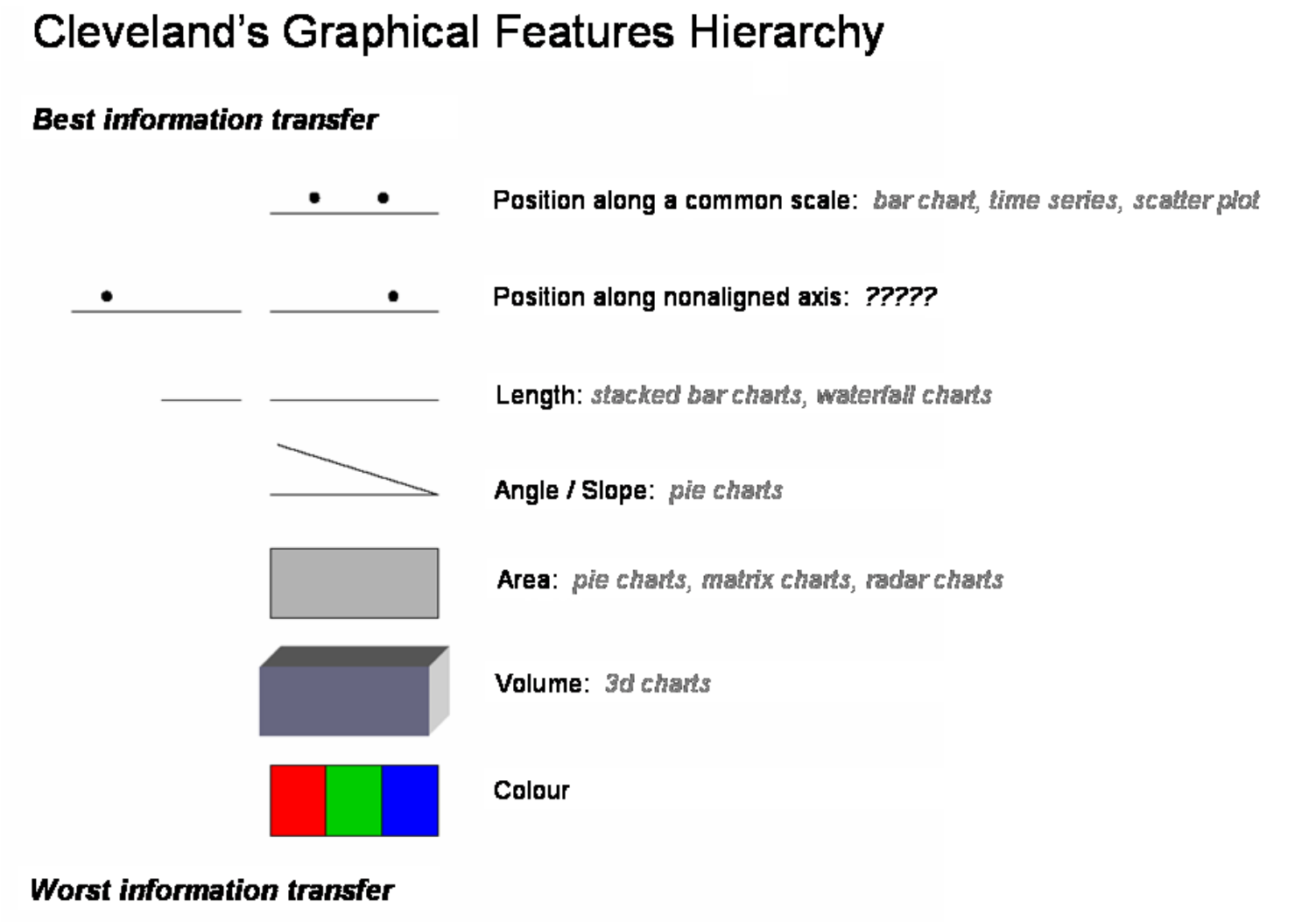
Cleveland's hierarchy



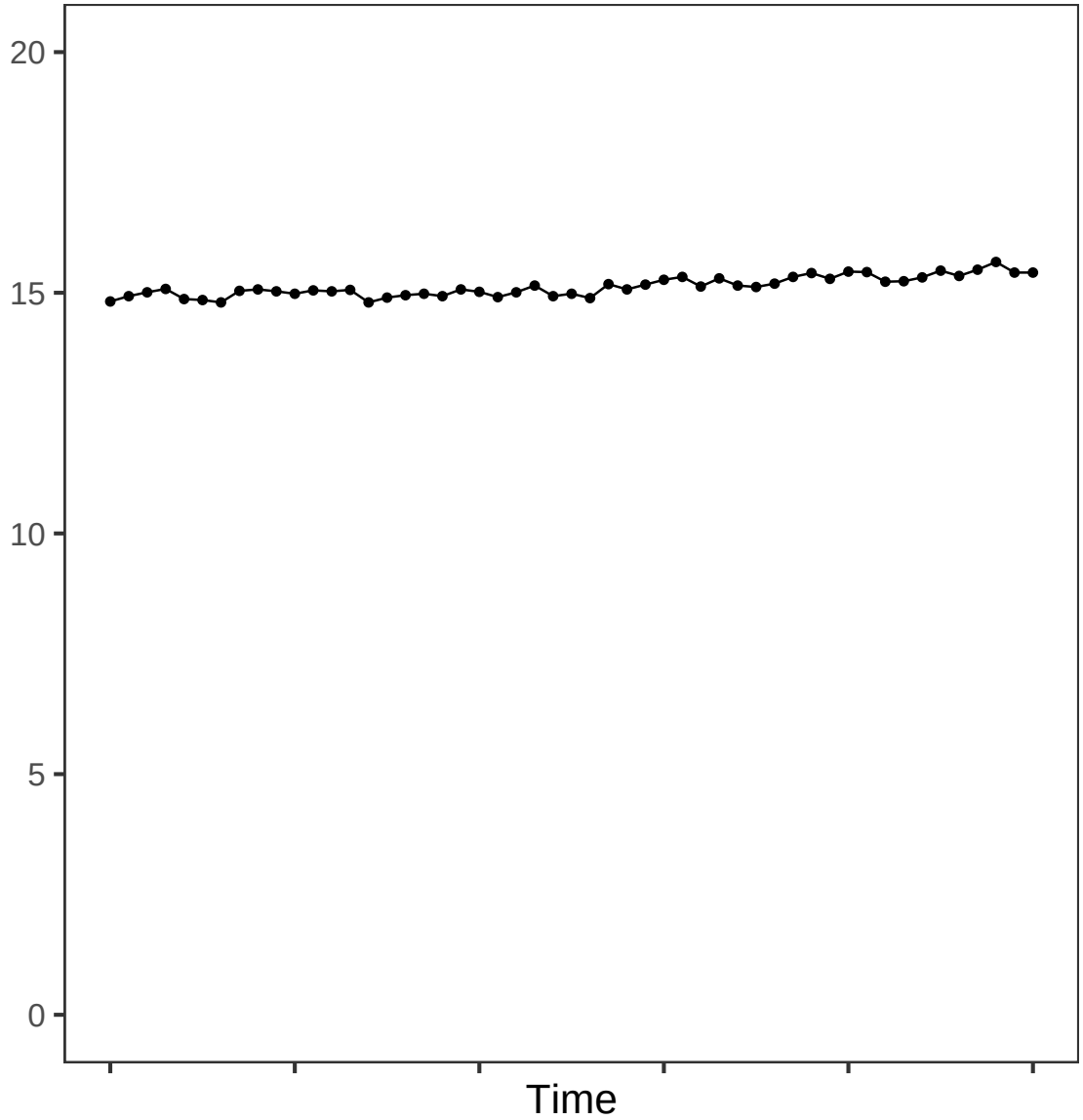
http://sfew.websitetoolbox.com/post/clevelands-graphical-features-hierarchy-4598555

Scales

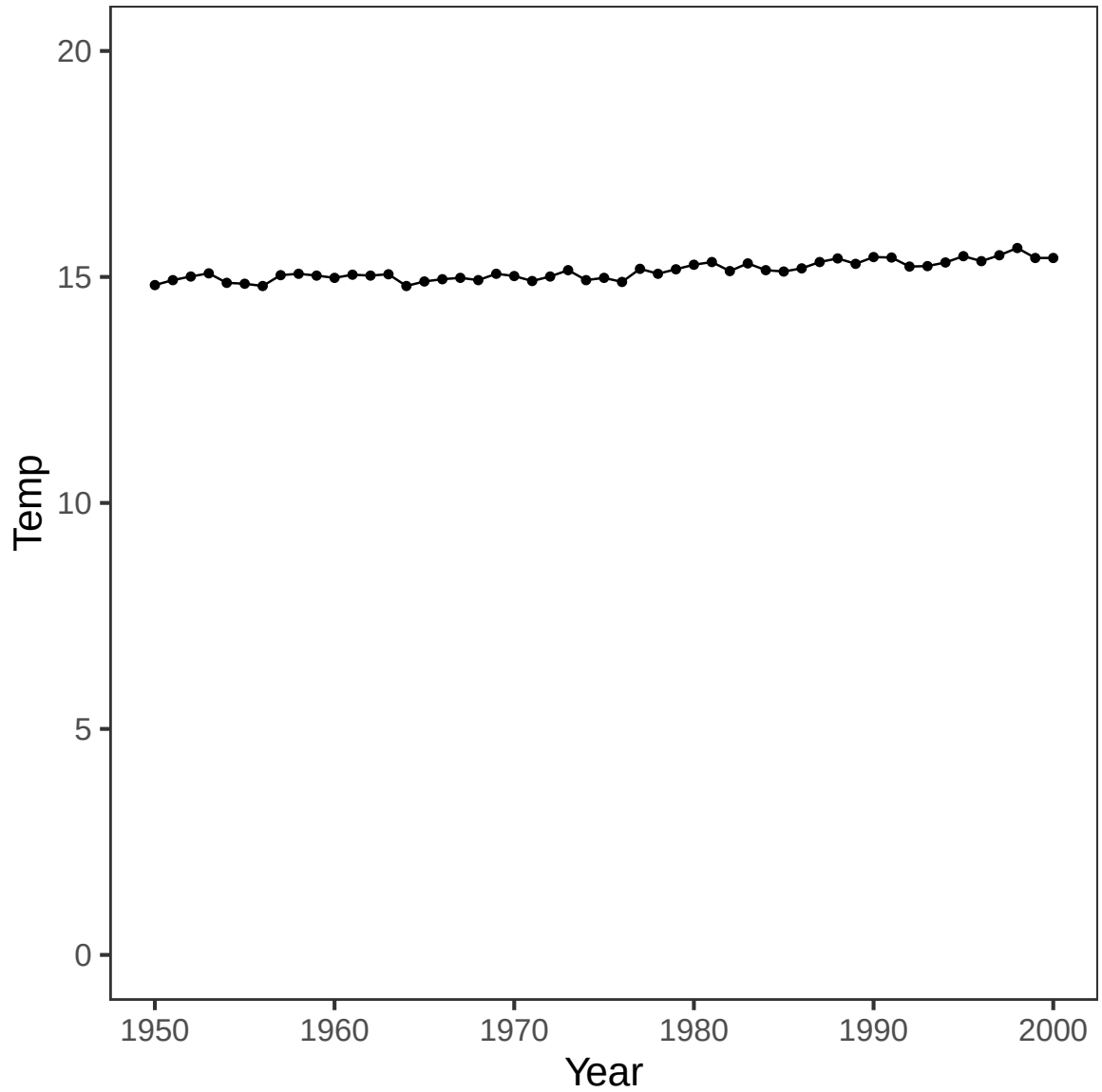
- The top of the hierarchy involves putting things on scales
- But what scale do we use?
  - Are our data anchored to zero?
    - If so, are we interested in differences or ratios?
  - Are they anchored somewhere else?

Outline

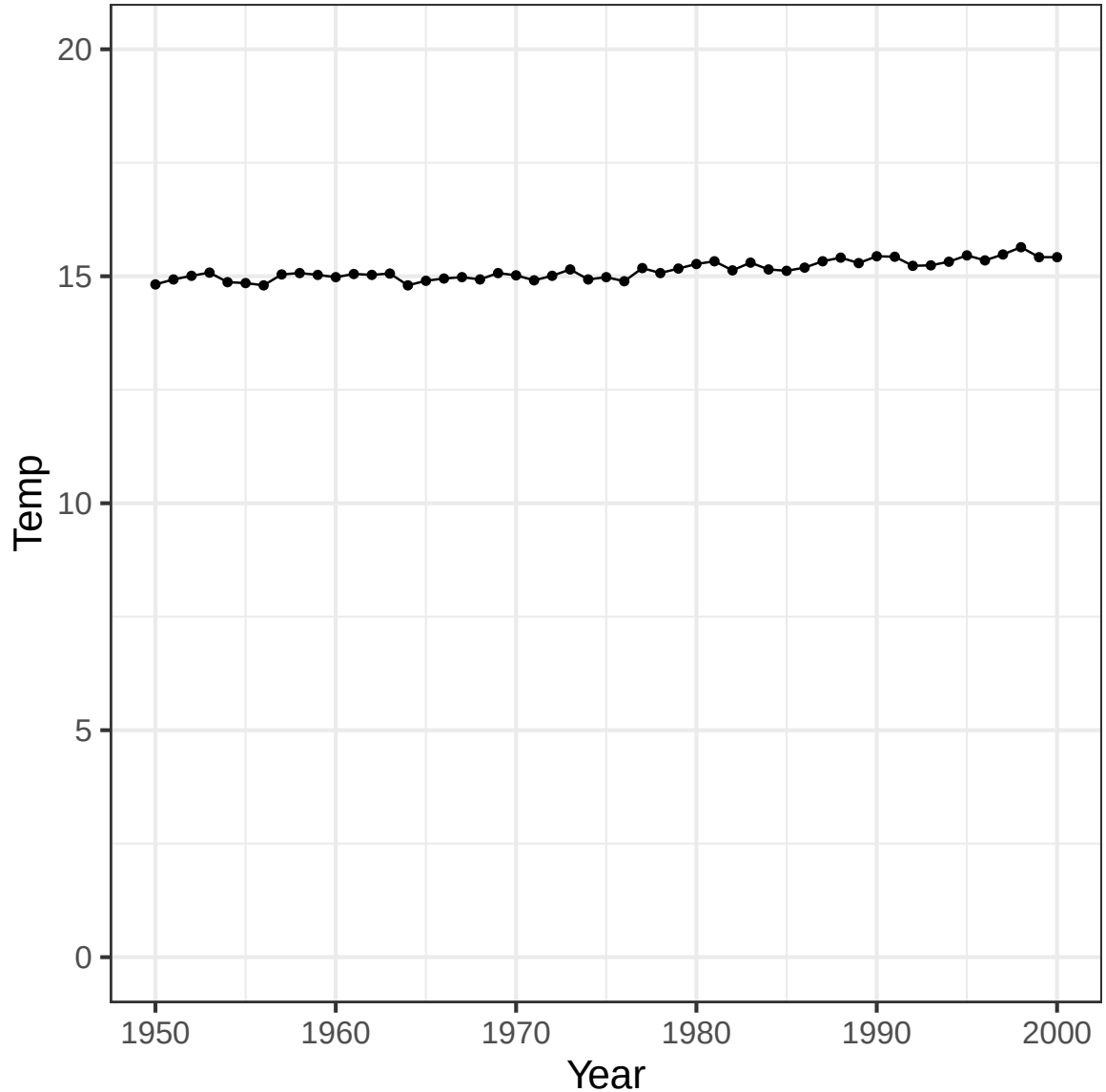
Golem bait call



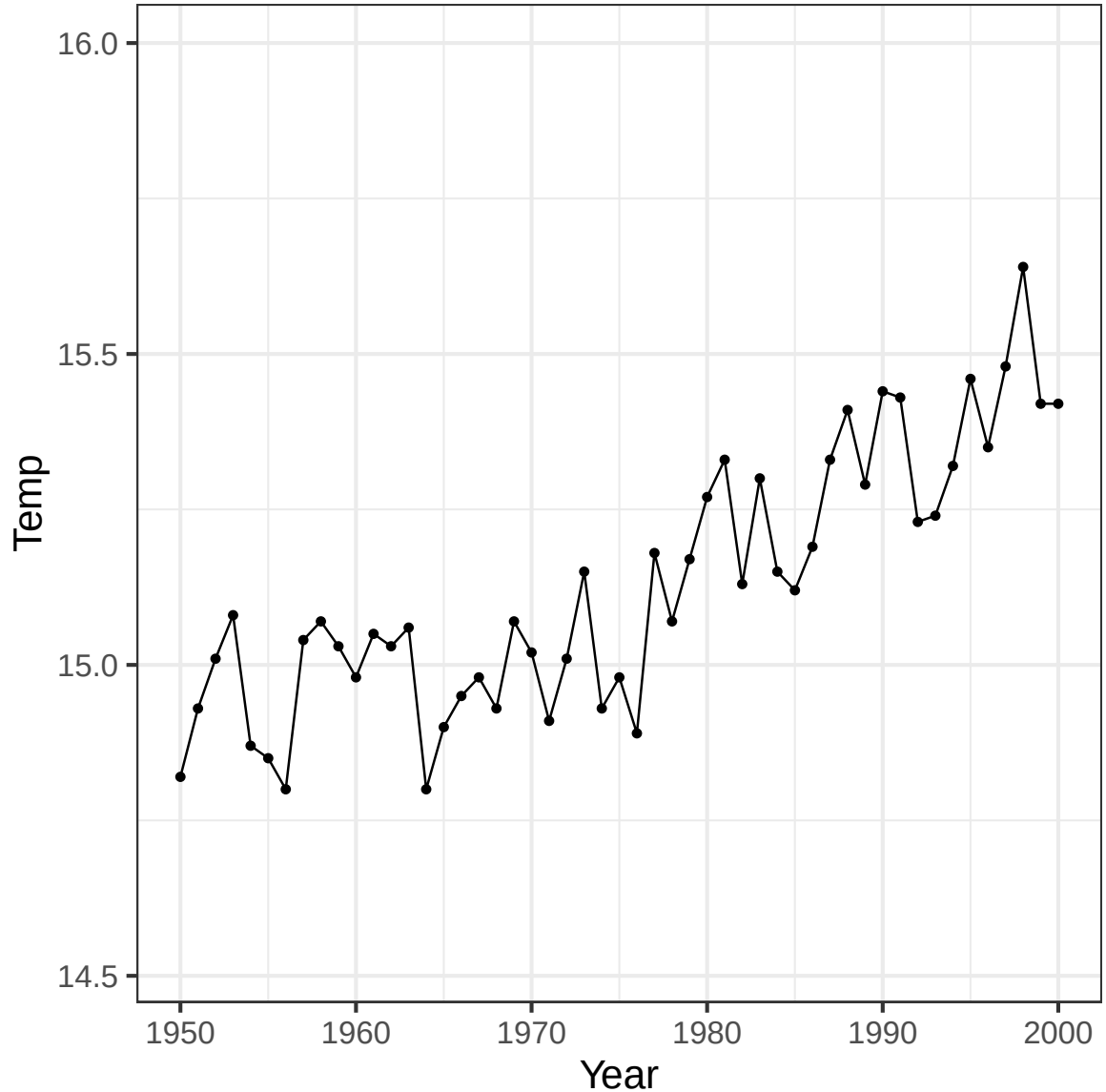
Global climate



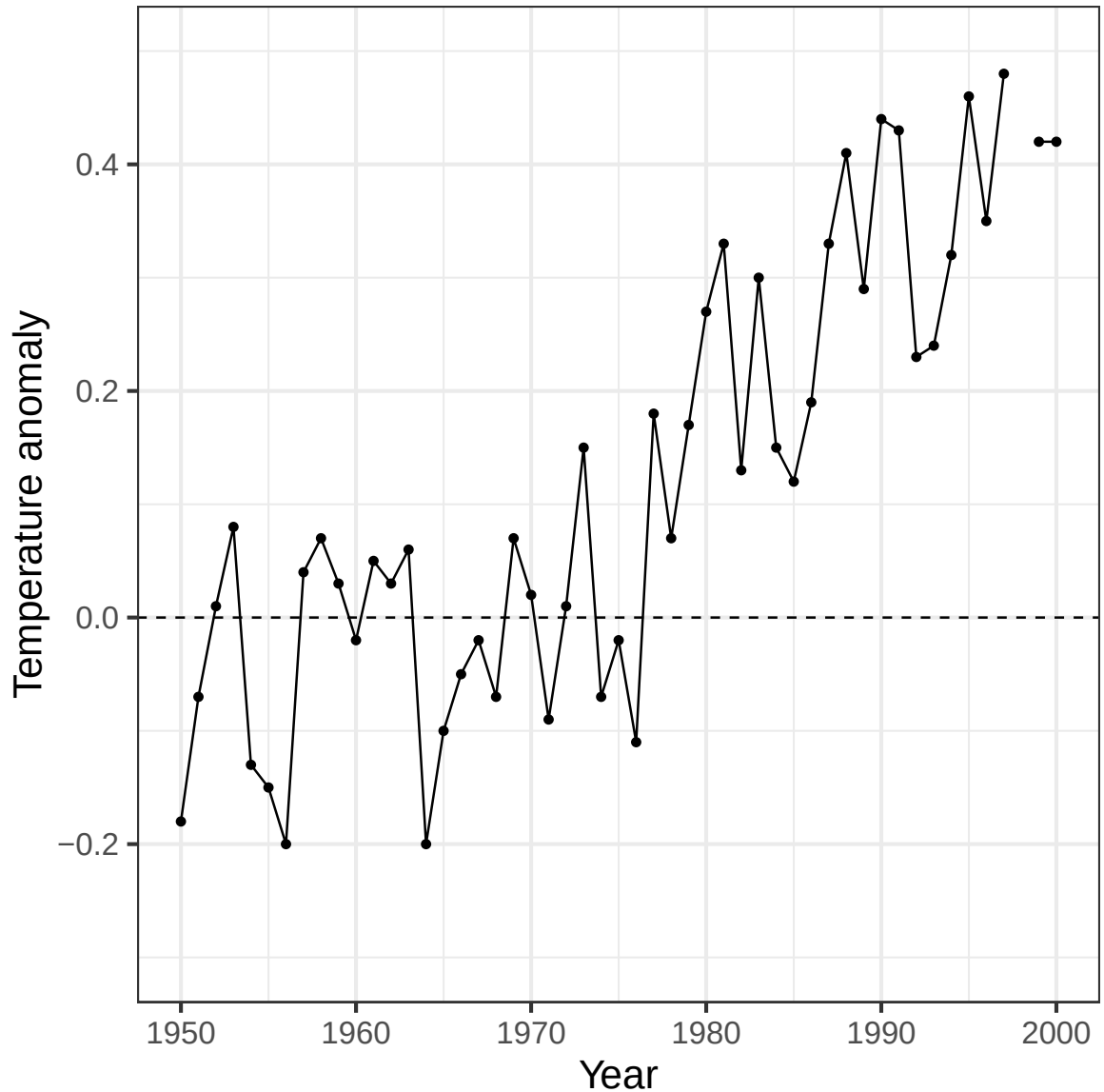
Global climate



Global climate



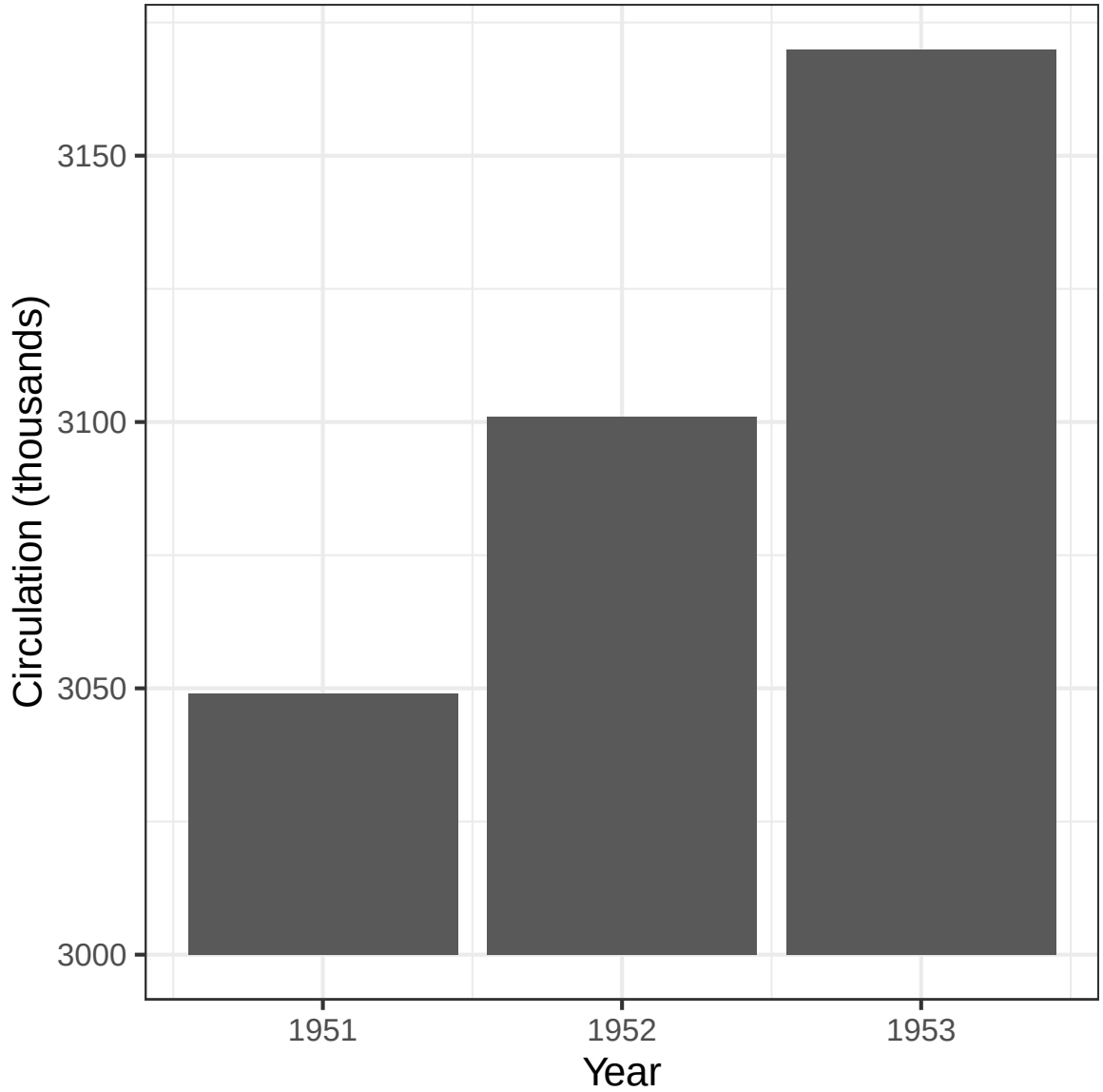
Global climate



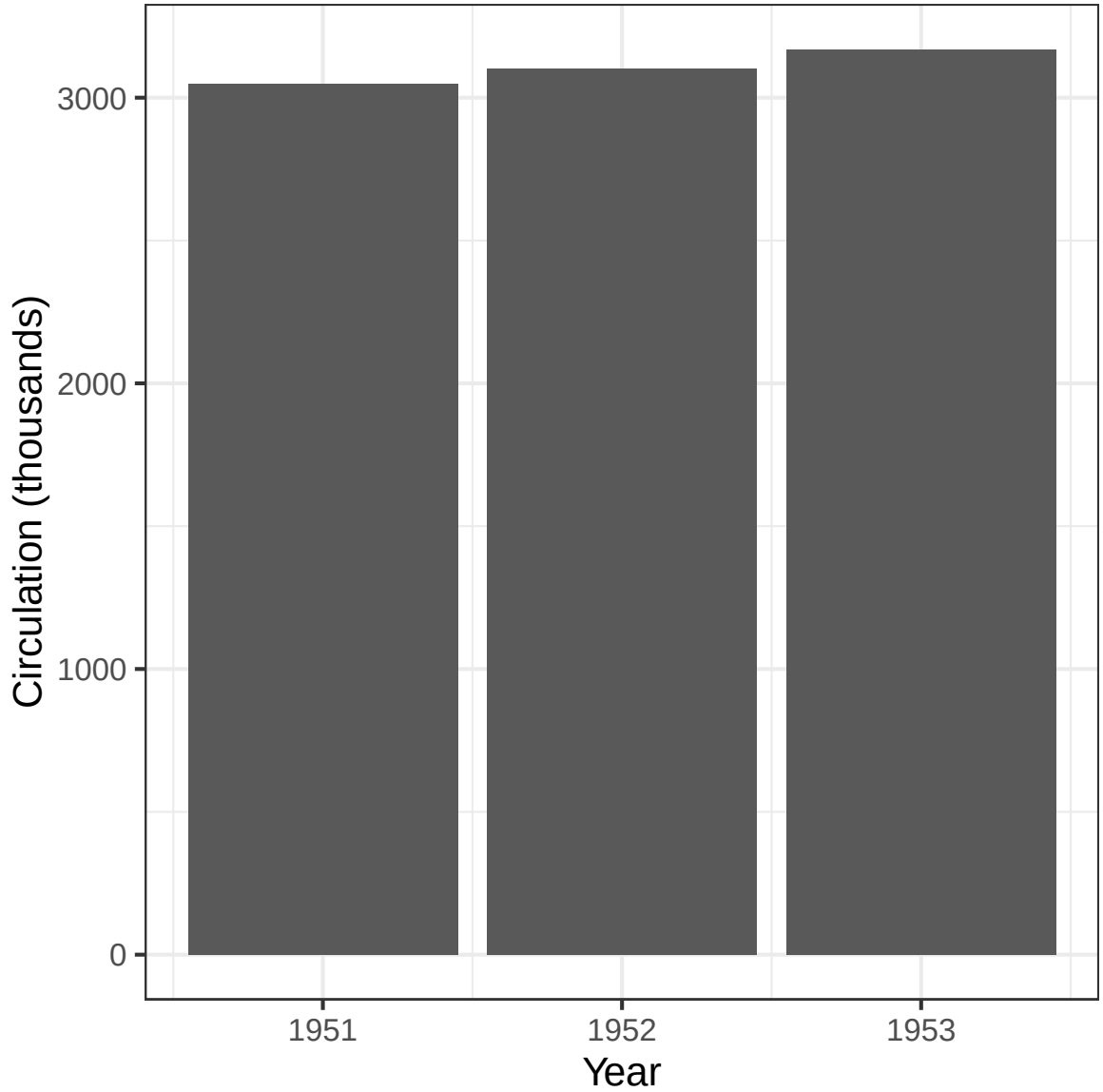
Climate lessons

- Choosing an anchor is a scientific decision
- Remember: graphic design is communication

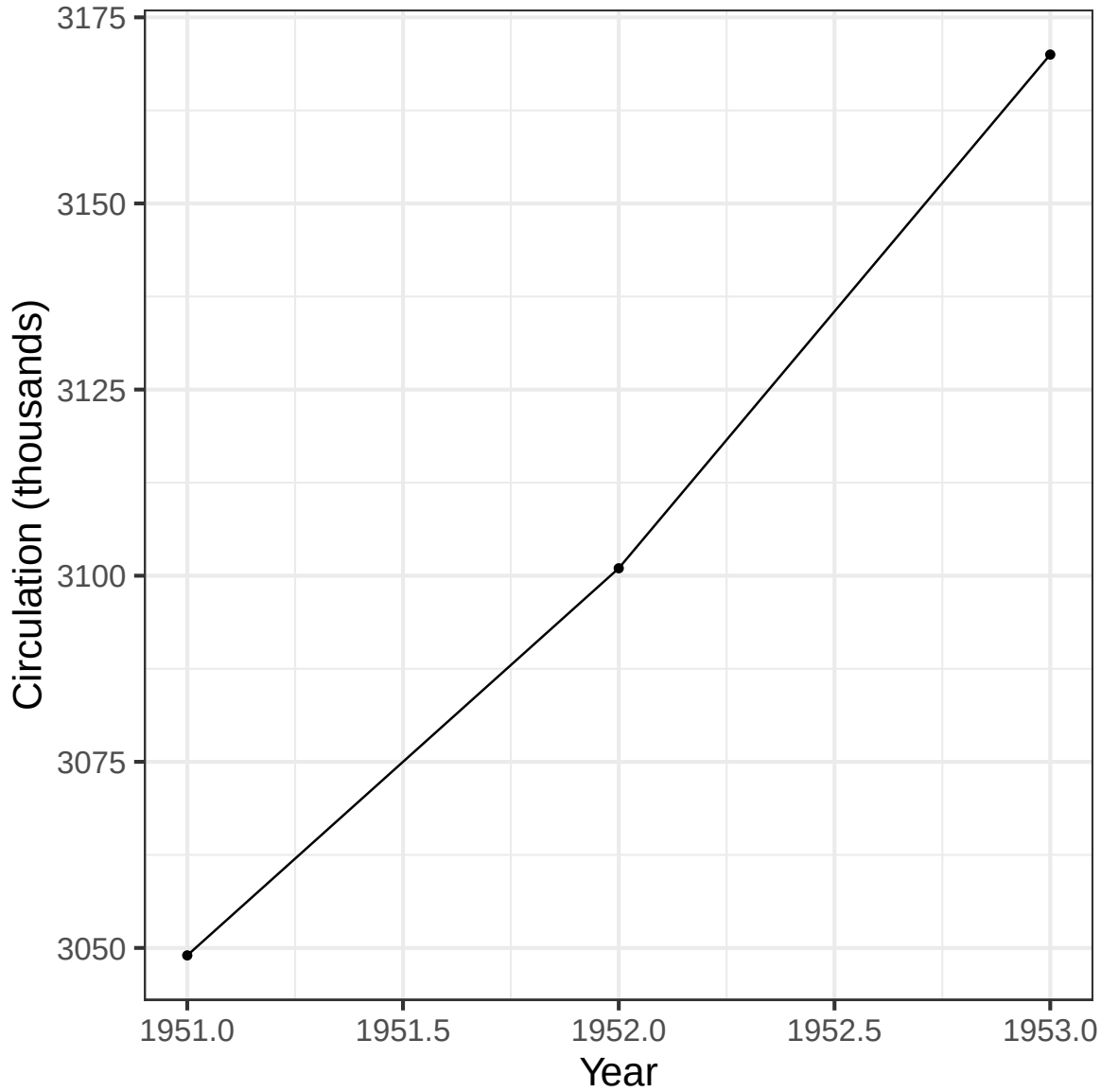
Magazine circulation (advertisement)



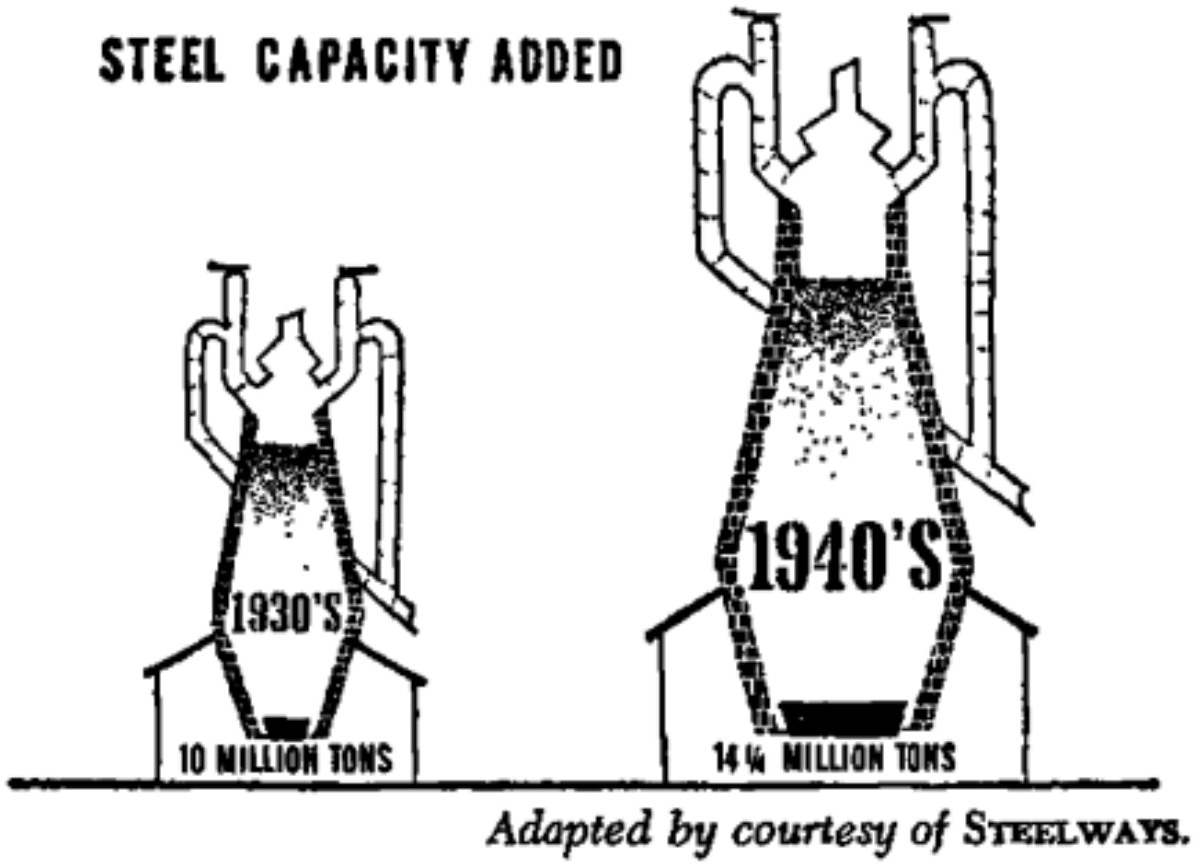
Magazine circulation (absolute amount)



Magazine circulation (trend)



Area and volume



How to Lie with Statistics

Advertisement lessons

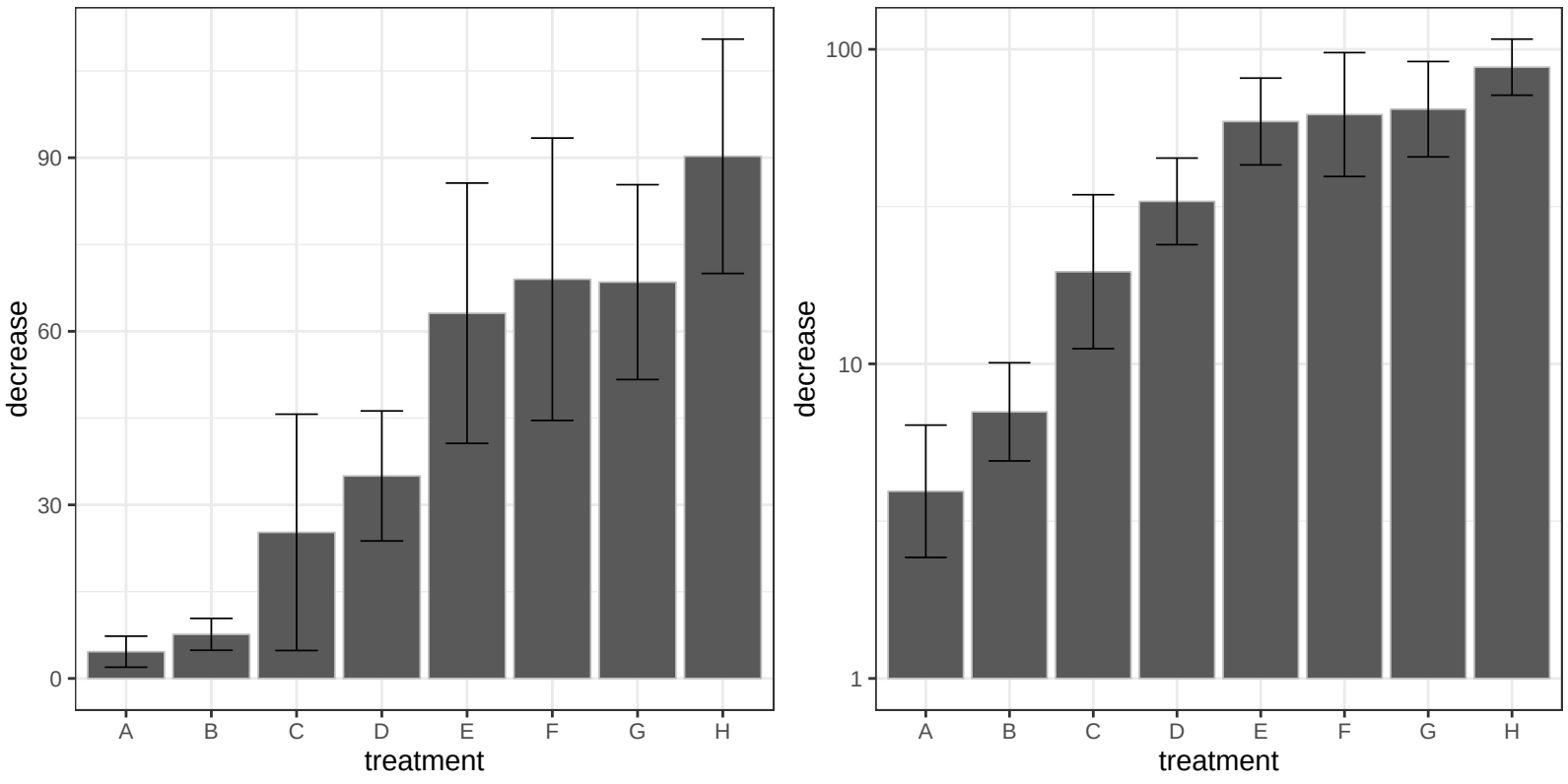
- Use area to indicate fair comparisons
  - On a physical scale
- Areas that can be compared linearly should be preferred
  - Depends on importance of feature
- Avoid using (or hinting at) volume

Outline

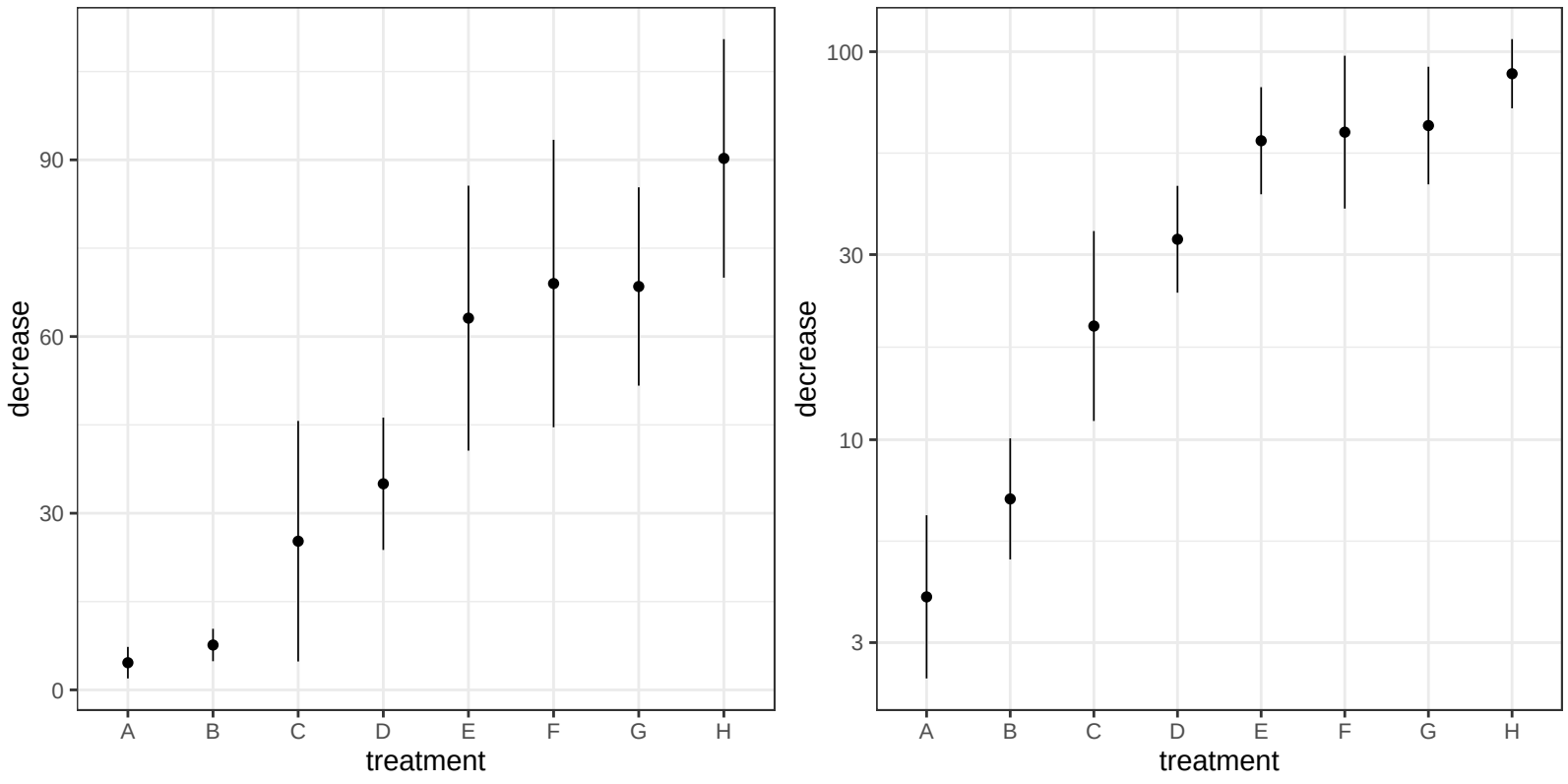
Physical quantities

- 1 is to 10 as 10 is to what?
  - * If you said 19, you are thinking on a linear scale
  - * If you said 100, you are thinking on a log scale
- The log scale is often good for physical quantities:
  - When zero means zero

Log vs. linear



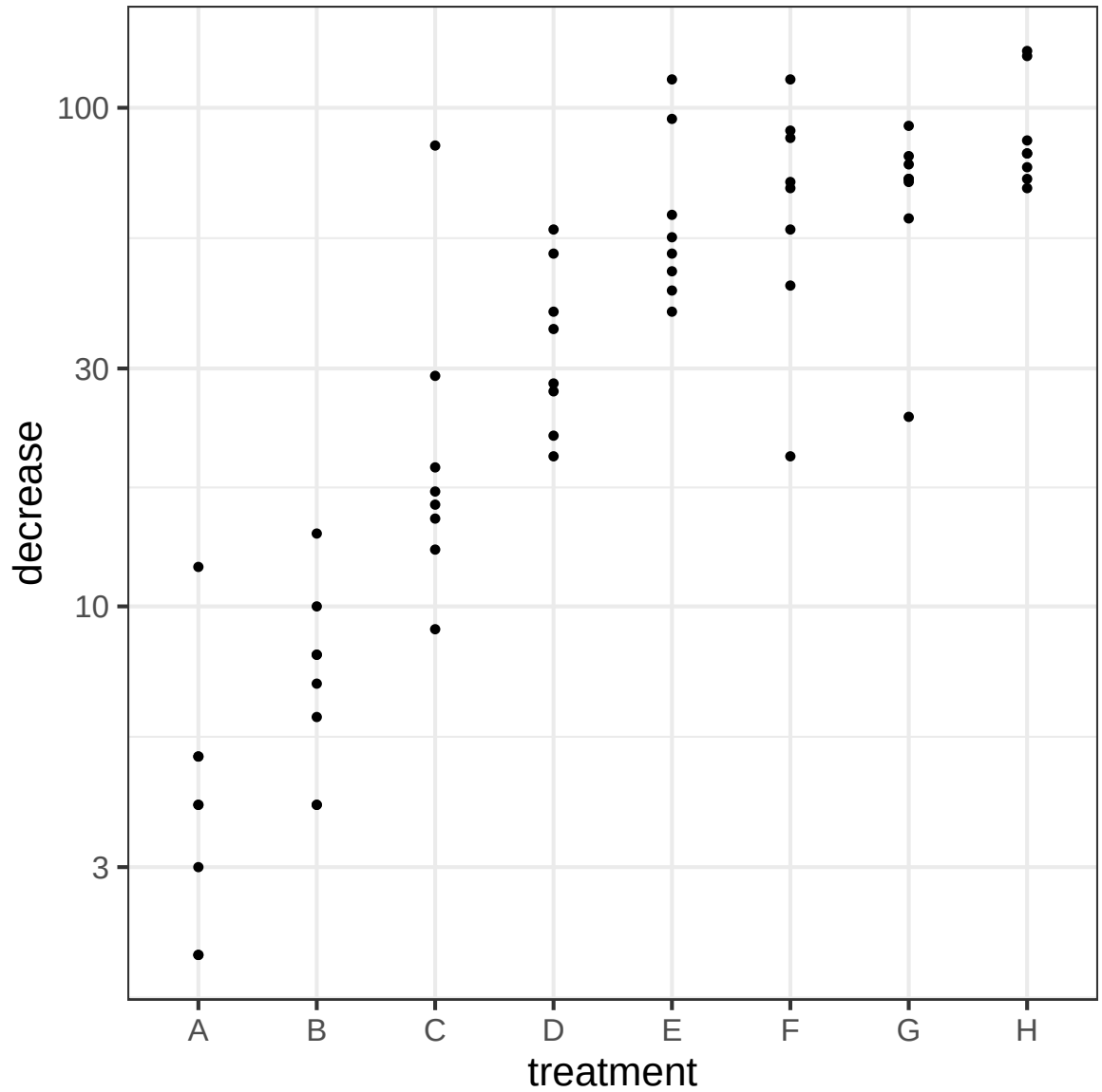
Making room


decrease
90
60
30
0
A
B
C
D
E
F
G
H
treatment
decrease
100
30
10
3
A
B
C
D
E
F
G
H
treatment

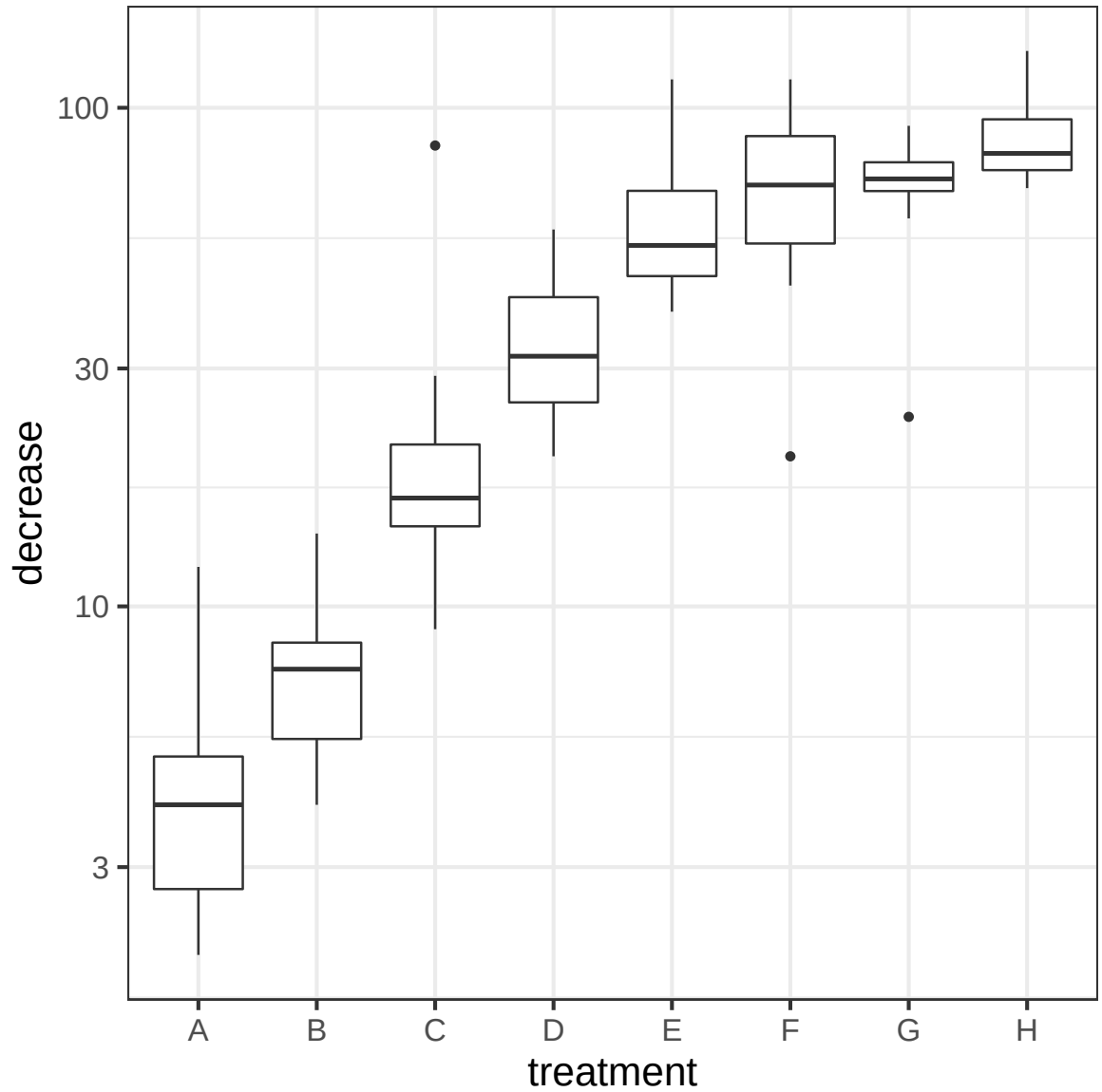
Data shape

- There are a lot of different ways to show data shape
- Choices will depend on your data set:
  - Overall size
  - Number of replicates
  - Number of levels, predictor variables, etc.

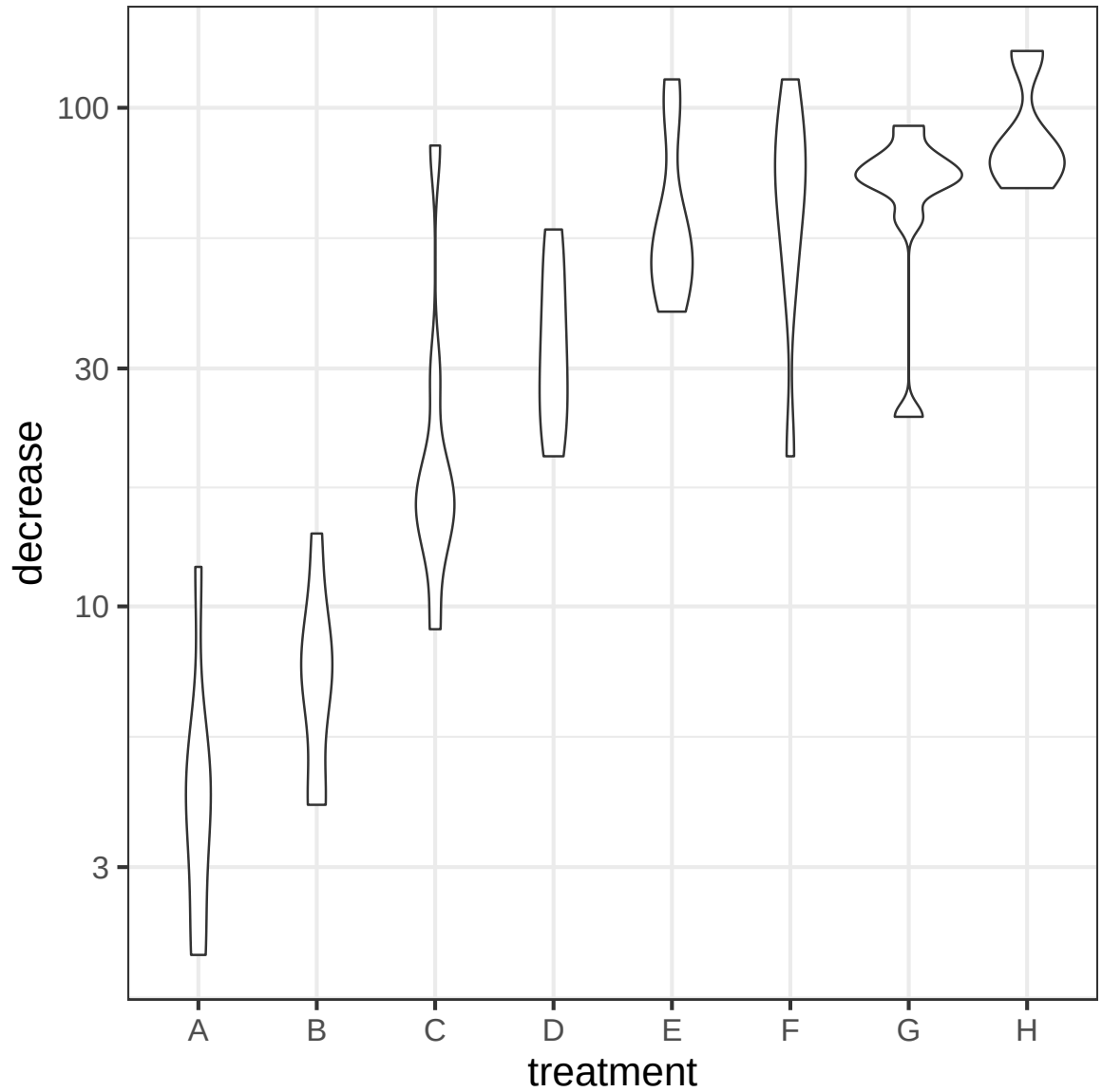
Showing points



Boxplot



Violin plot



Orchard lessons

- Choices about log vs. linear scale are scientific choices
  - Neither is more valid, or closer to the data
- You can also make choices about
  - sending a simple message
  - providing more information about shape
- Log scales are almost never physical
  - Don't mislead with area information on a log scale

Probabilities

- 1% is to 2% as 50% is to what?
  - * 51% is way too small
  - * 100% is way too large
- The natural distance to use on a probability scale is log odds
  - * 1% is to 2% as 50% is to 67%
  - * . . . as 2% is to 4%
  - * . . . as 98% is to 99%

Odds

- Odds are a ratio between the probability of something and the probability of its opposite:
  - $o = p/(1-p)$
- Log odds give a natural distance on probability space

Extreme values

- Our transformations take extreme values to infinity.
- Use link functions: this is like using estimated values instead of observed; they are rarely infinite
- Extend the scale (e.g., use $\log(1 + x)$ instead of $\log(x)$)
  - This usually involves arbitrary choices
  - Should often be *avoided* for analysis
  - Usually *OK* for visualization