

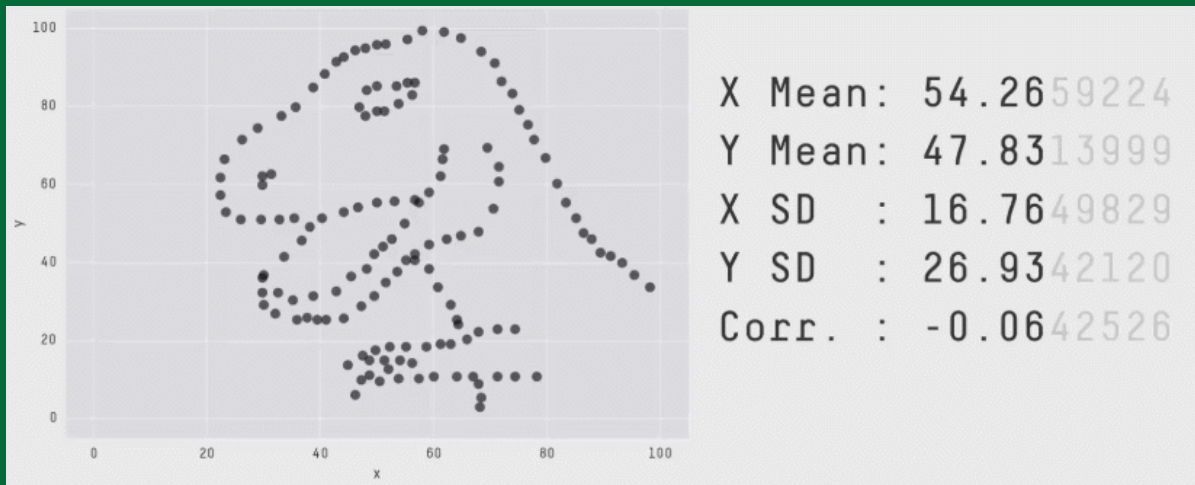
DSBA 5122: Visual Analytics

Class 5: Distributions and Uncertainty

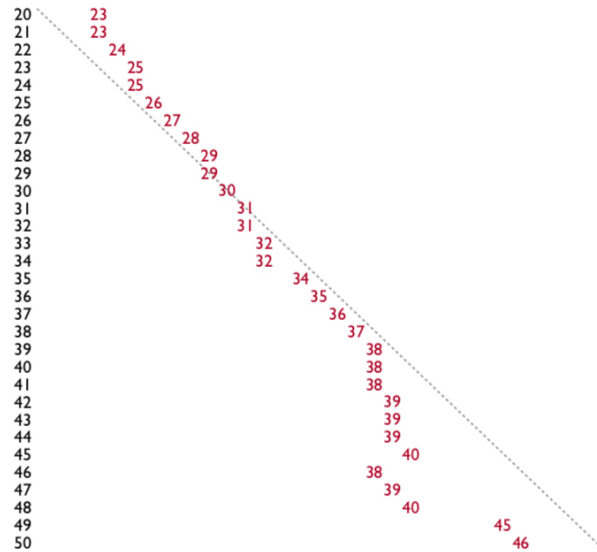
Ryan Wesslen

February 18, 2019

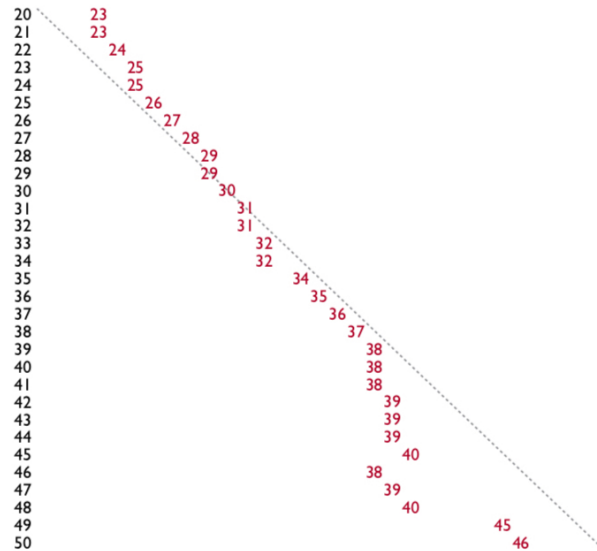
Why view Distributions? Cairo Ch. 7 & Wilke Ch. 7 - 9



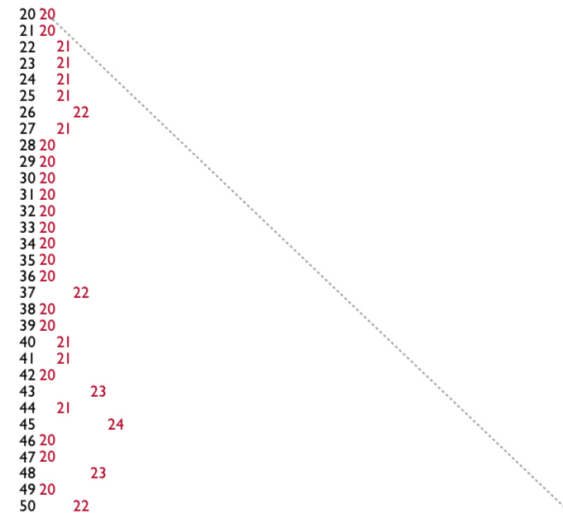
a woman's age vs. the age of the men who look best to her



a woman's age vs. the age of the men who look best to her



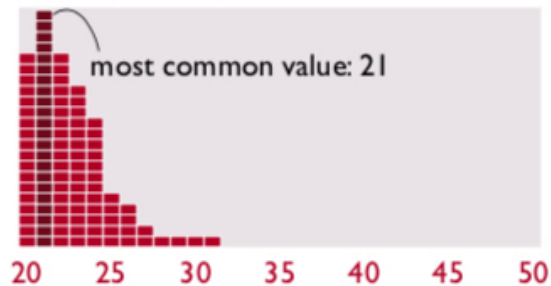
a man's age vs. the age of the women who look best to him



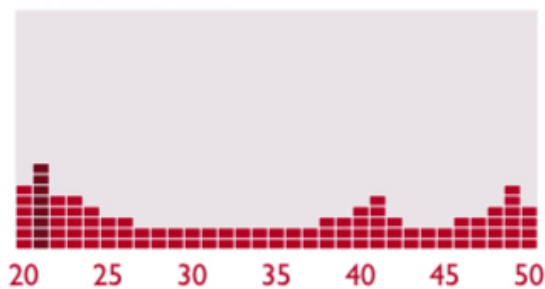
*men of 40 vs. the age
of the women who look
best to them*

■ = 1% of men

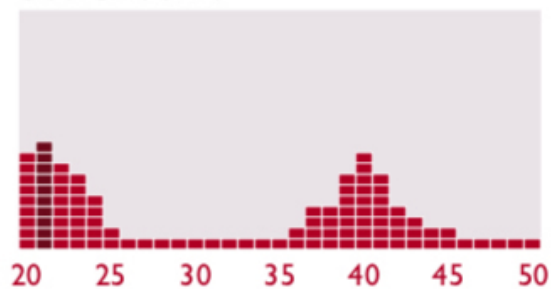
distribution 1



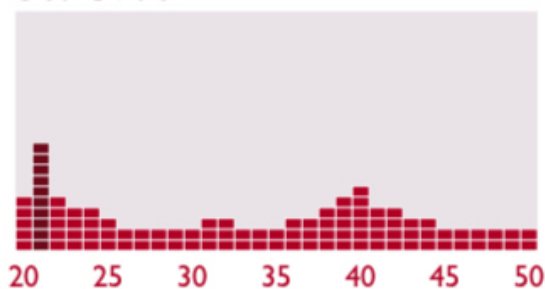
distribution 2



distribution 3



distribution 4



distribution 5

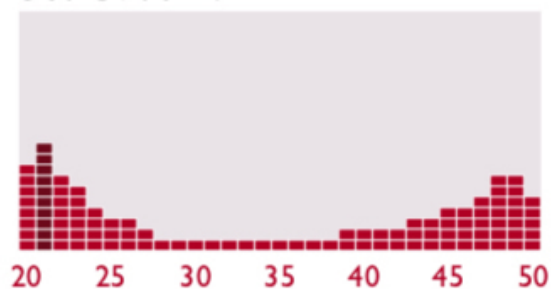
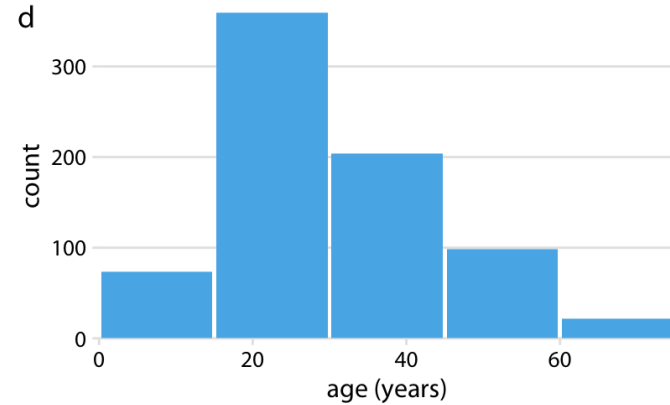
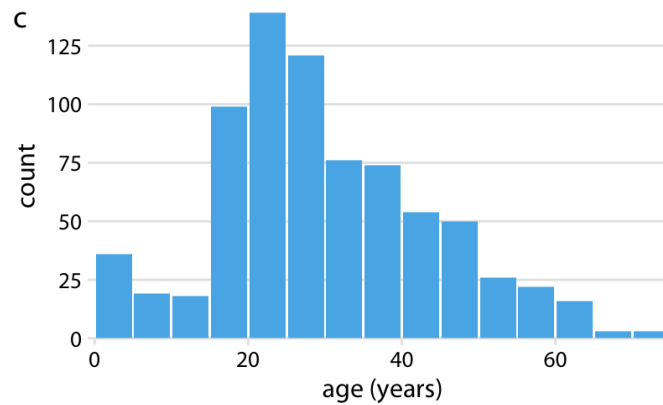
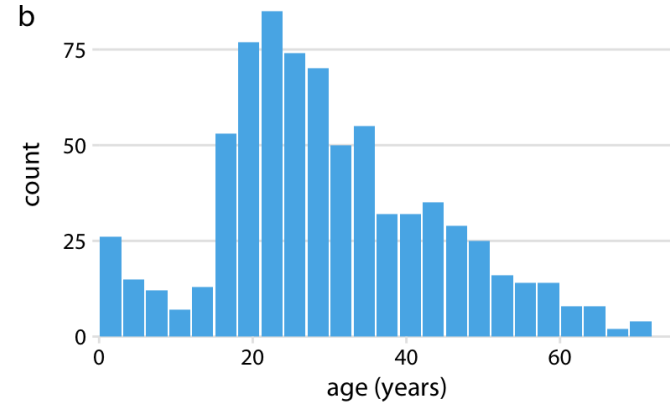
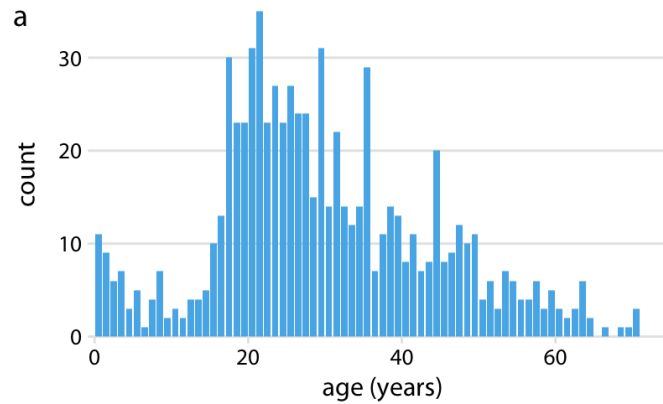
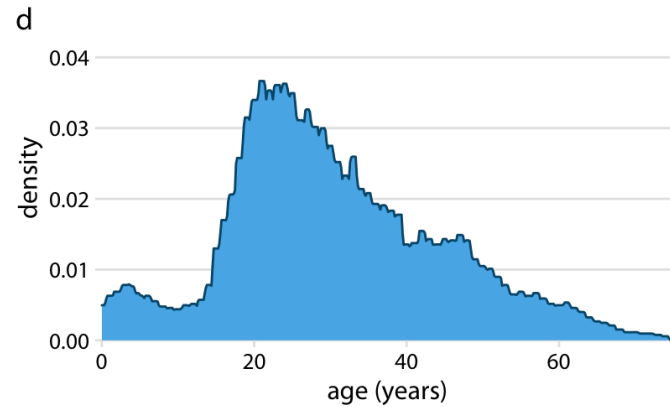
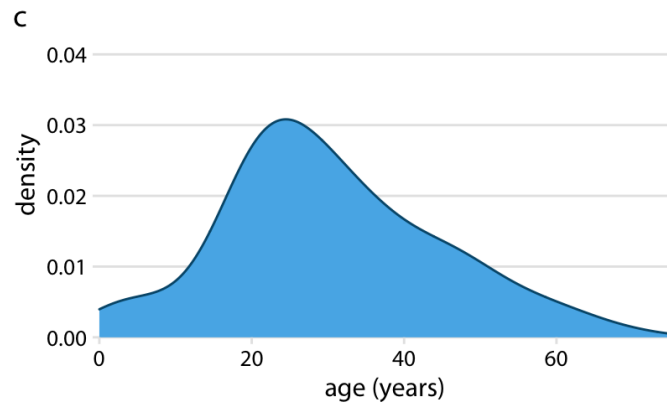
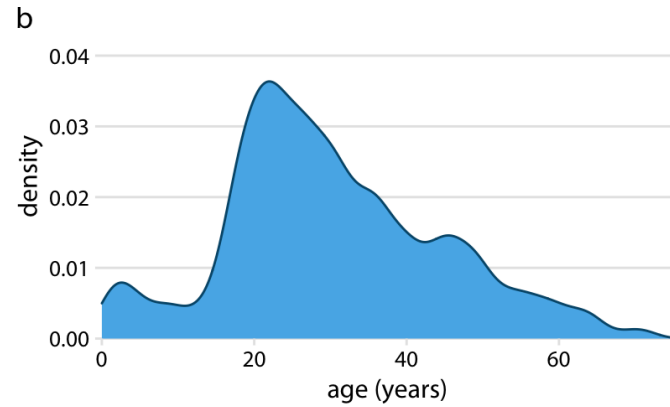
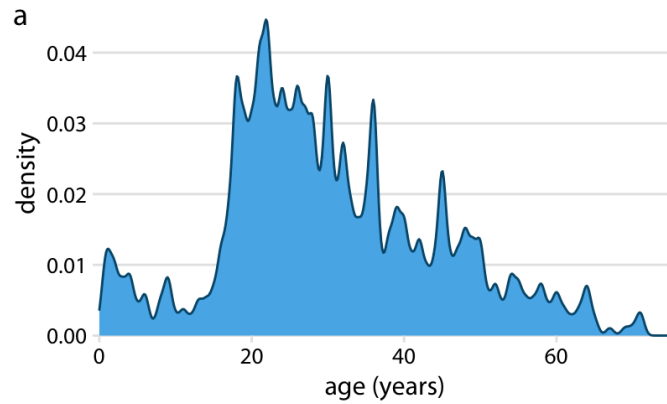
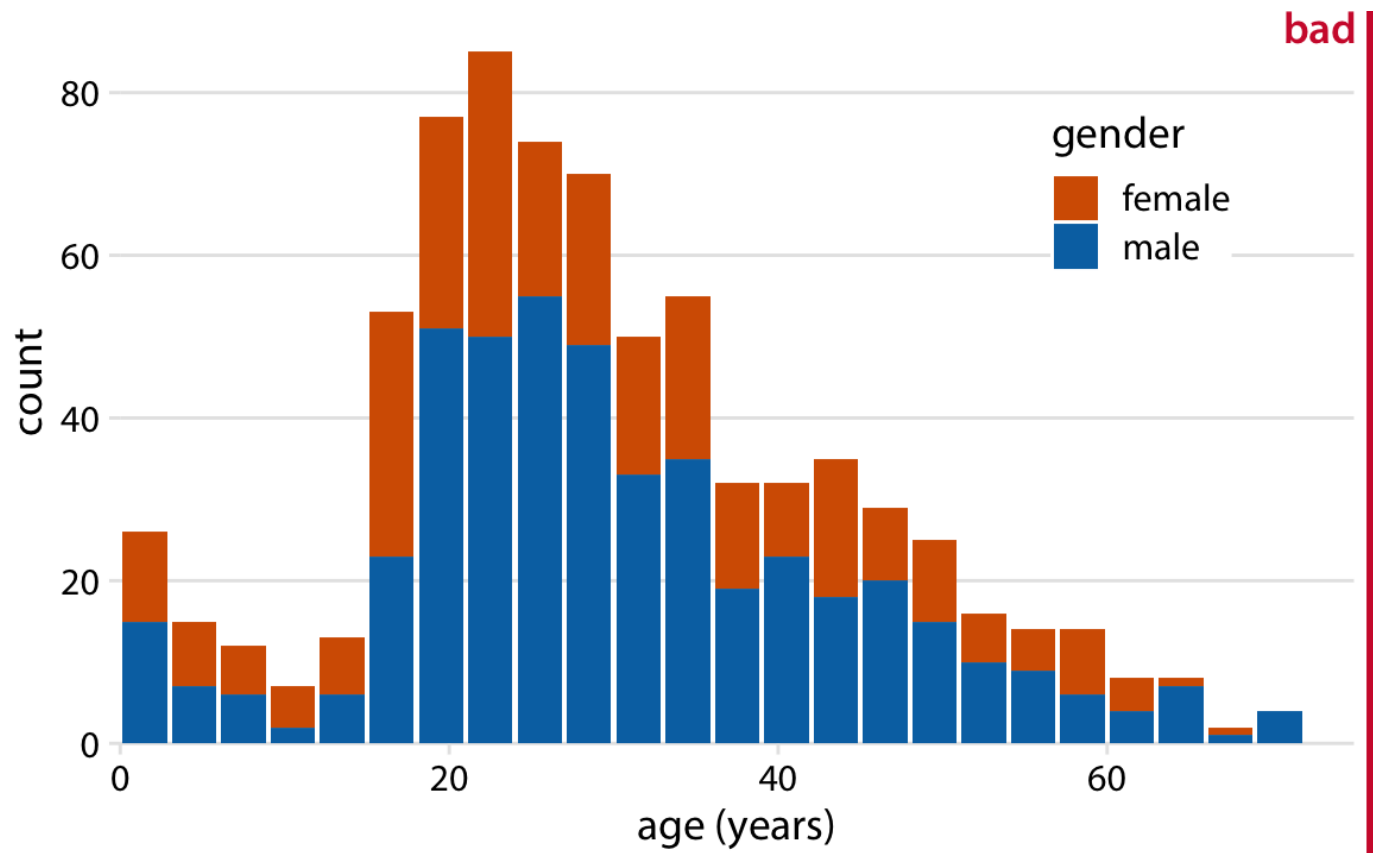
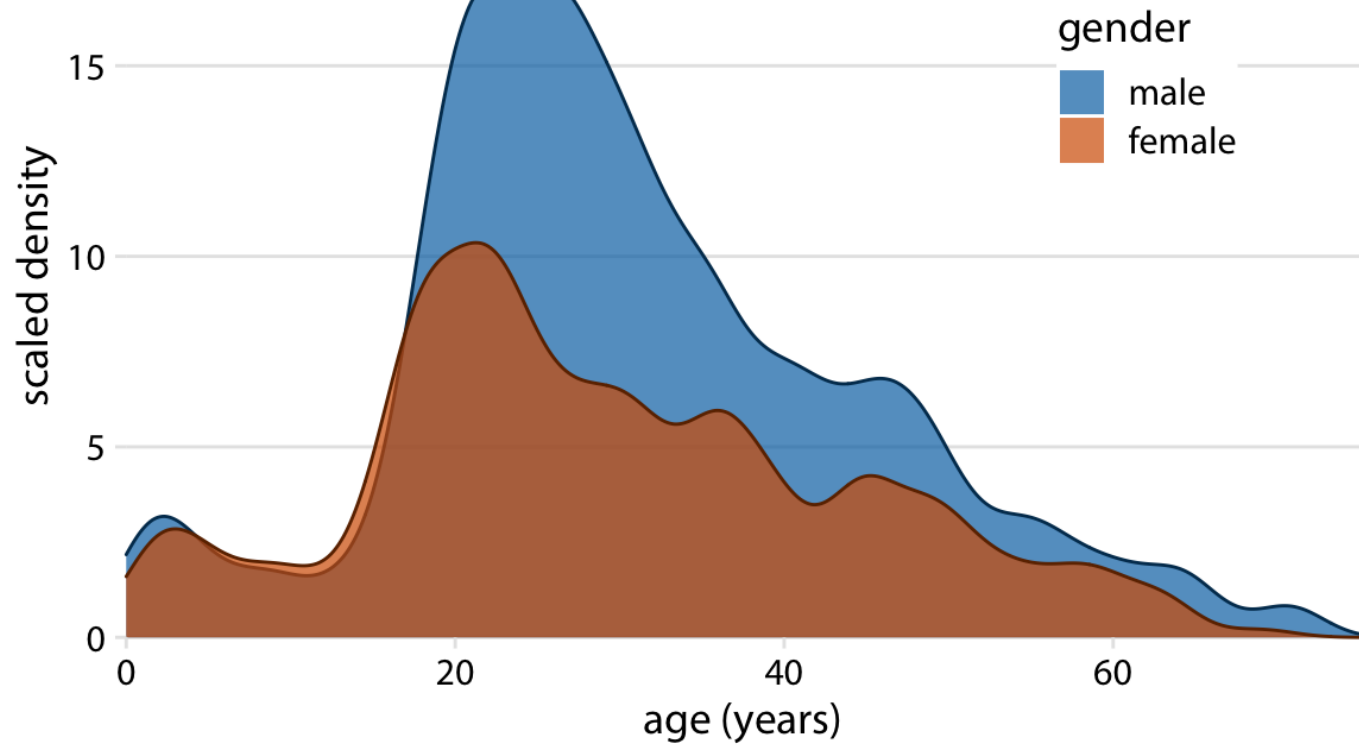


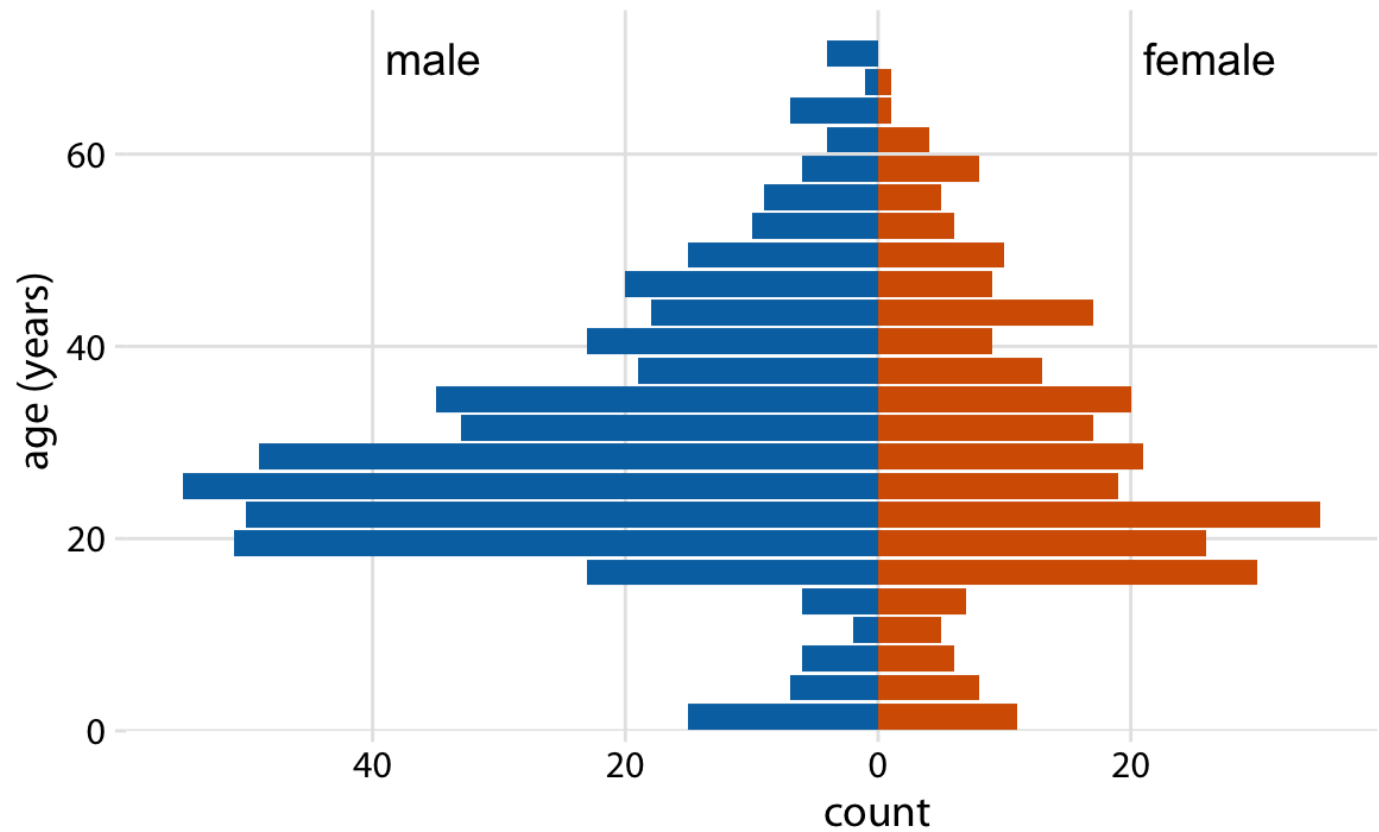
Figure 7.2 Five possible (and fictitious) distributions for the data corresponding to the preferences of men of 40. All of them have the same mode: 21.

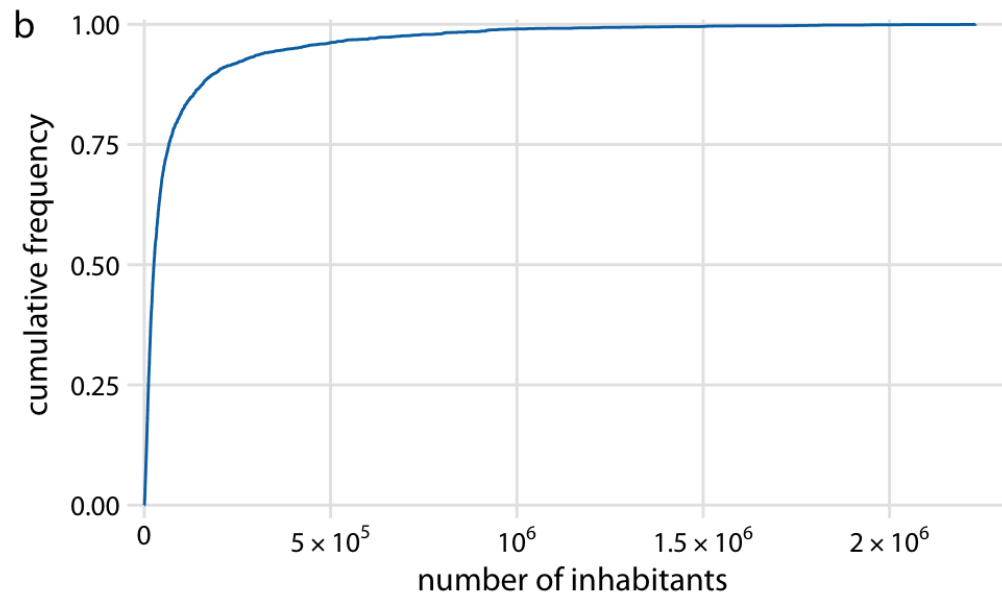
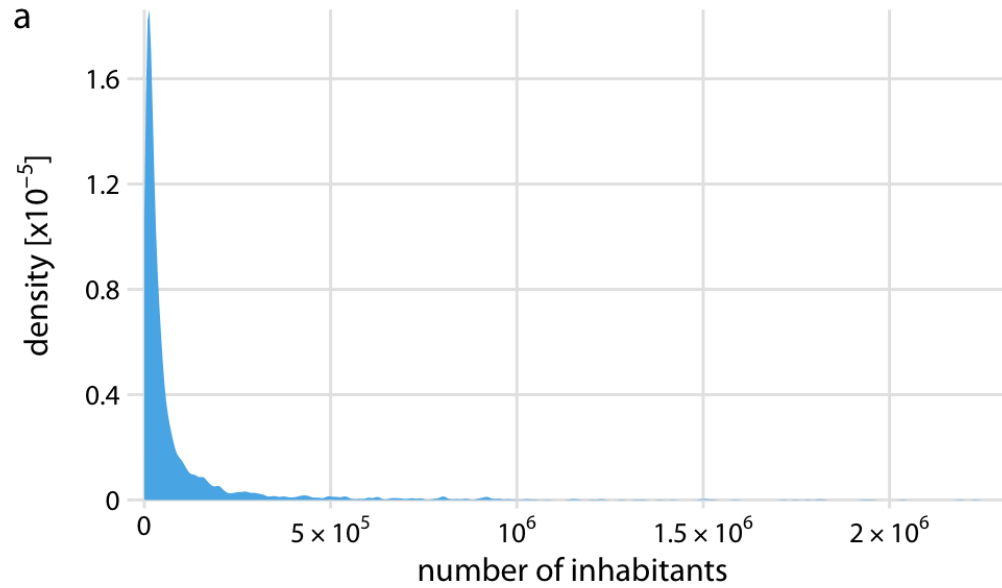


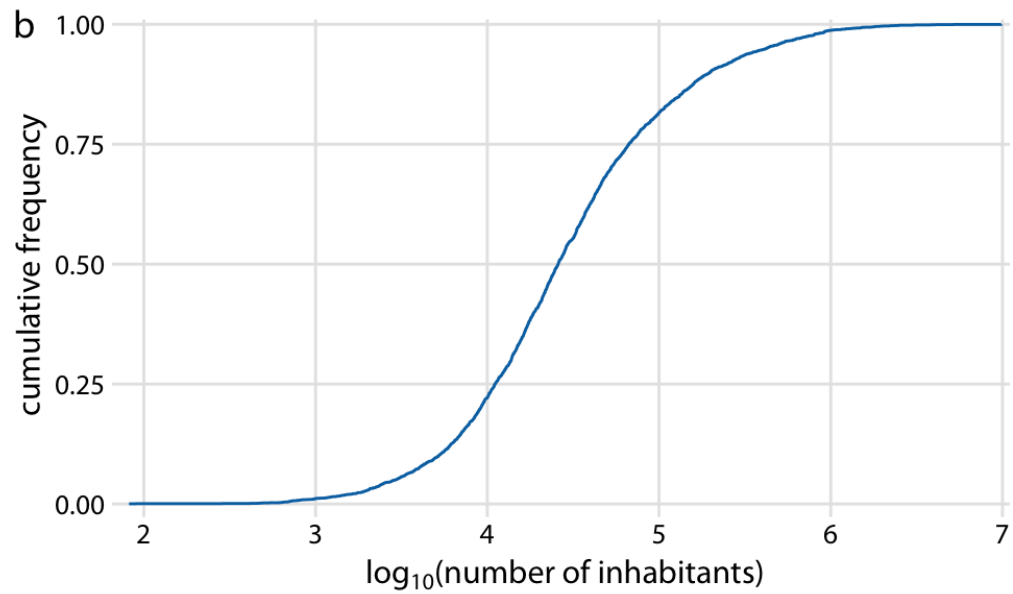
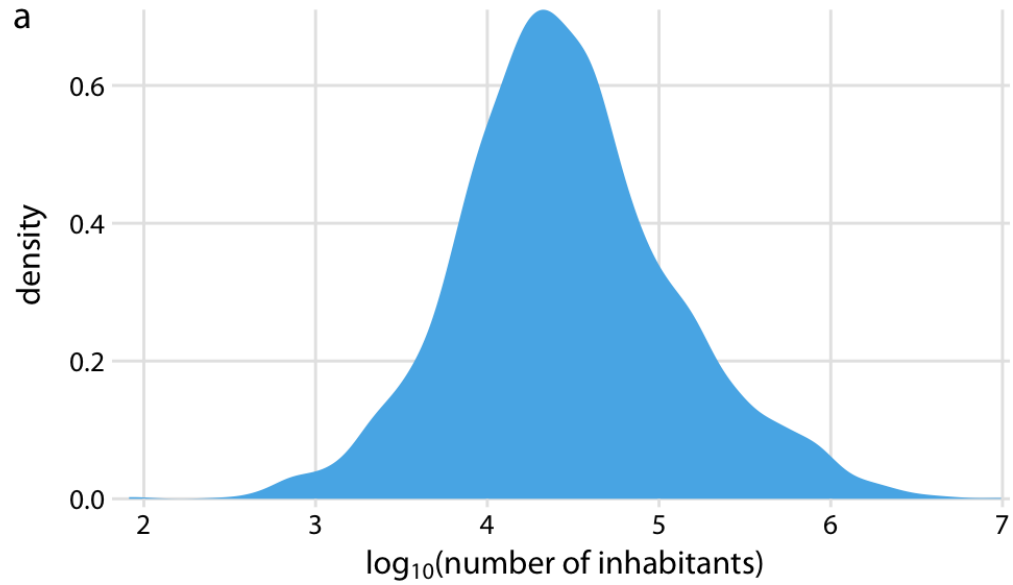


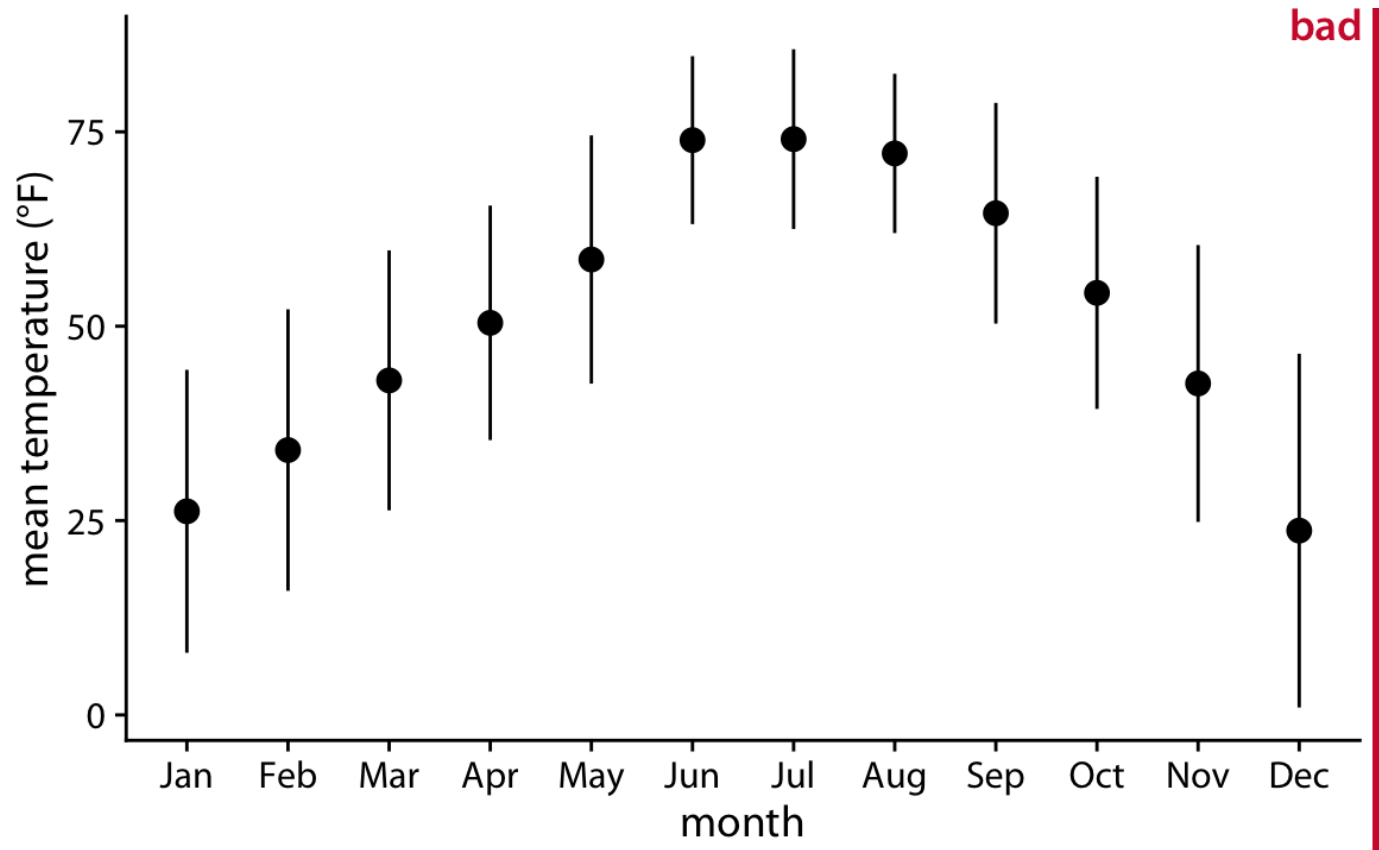




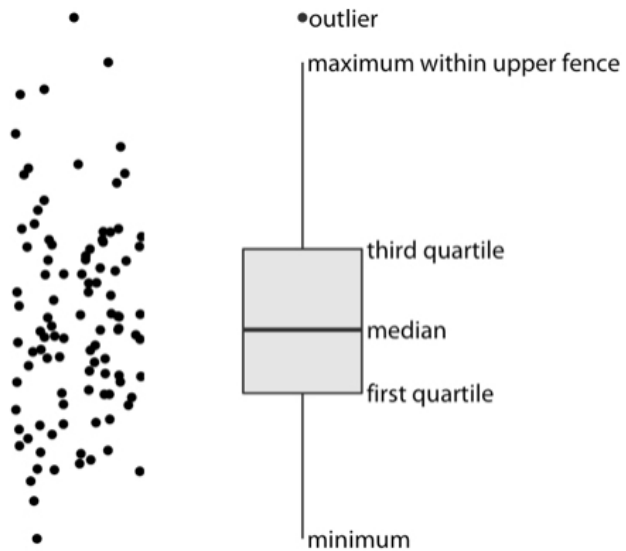




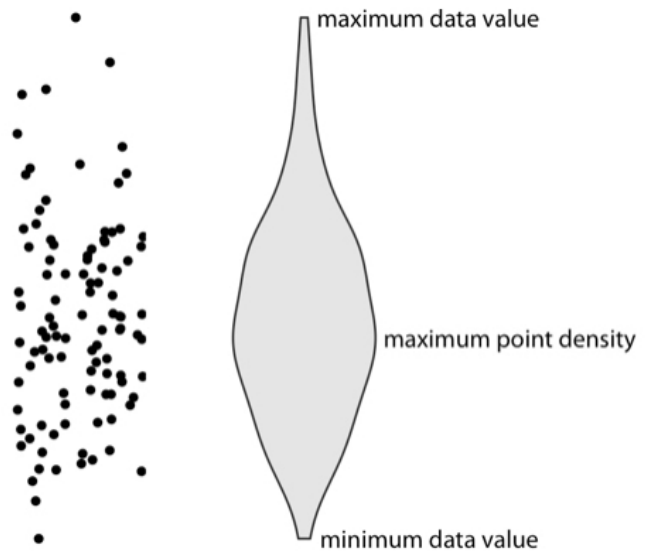


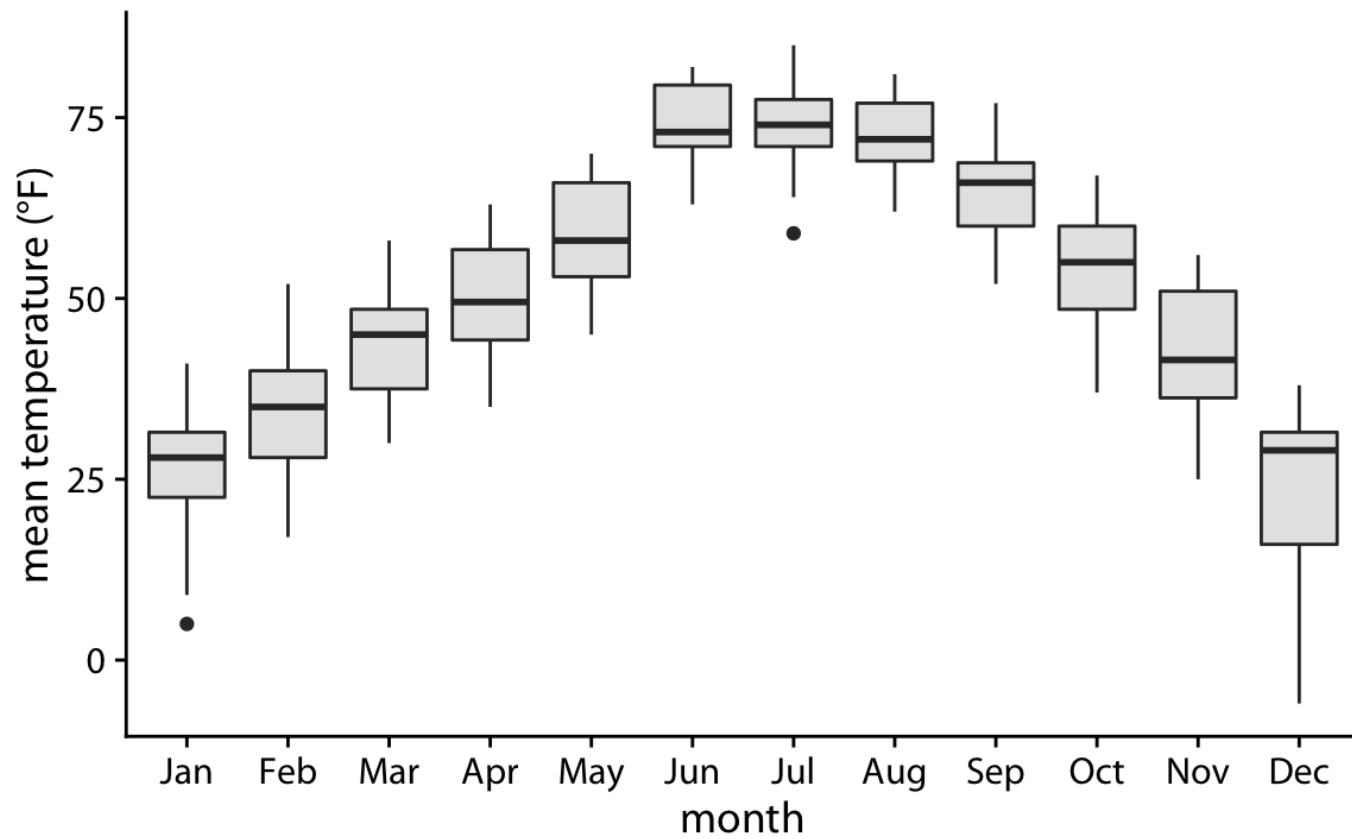


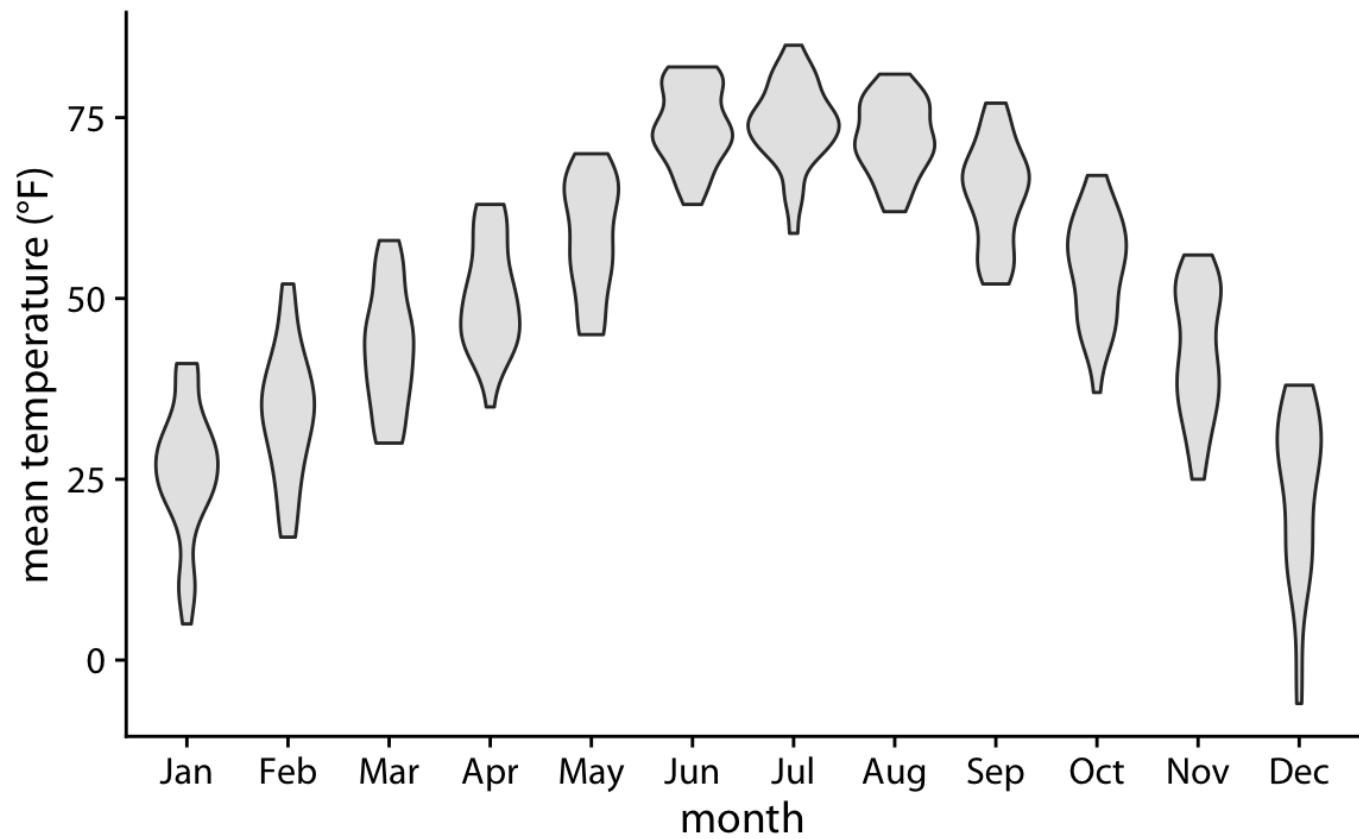
```
ggplot(df, aes(x, y)) + geom_boxplot()
```

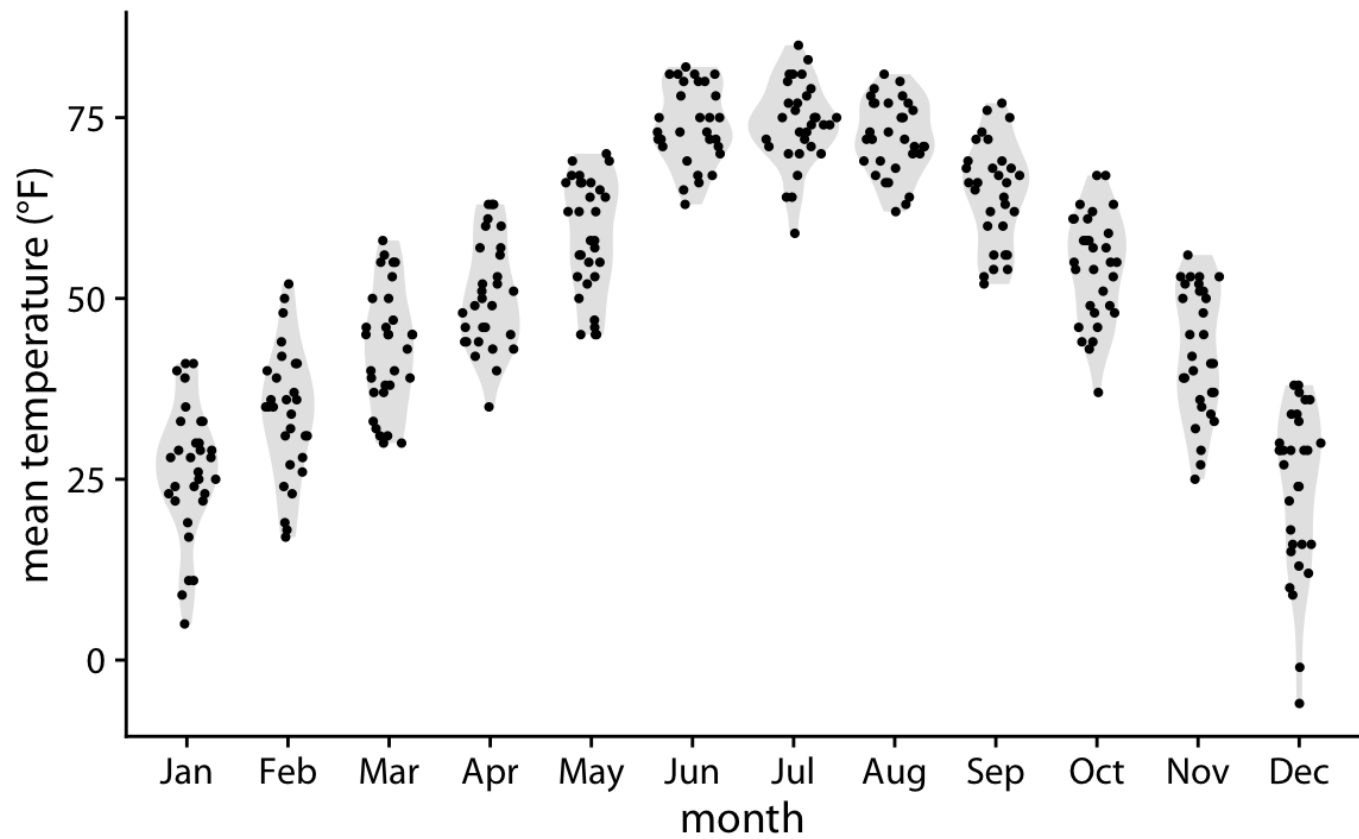


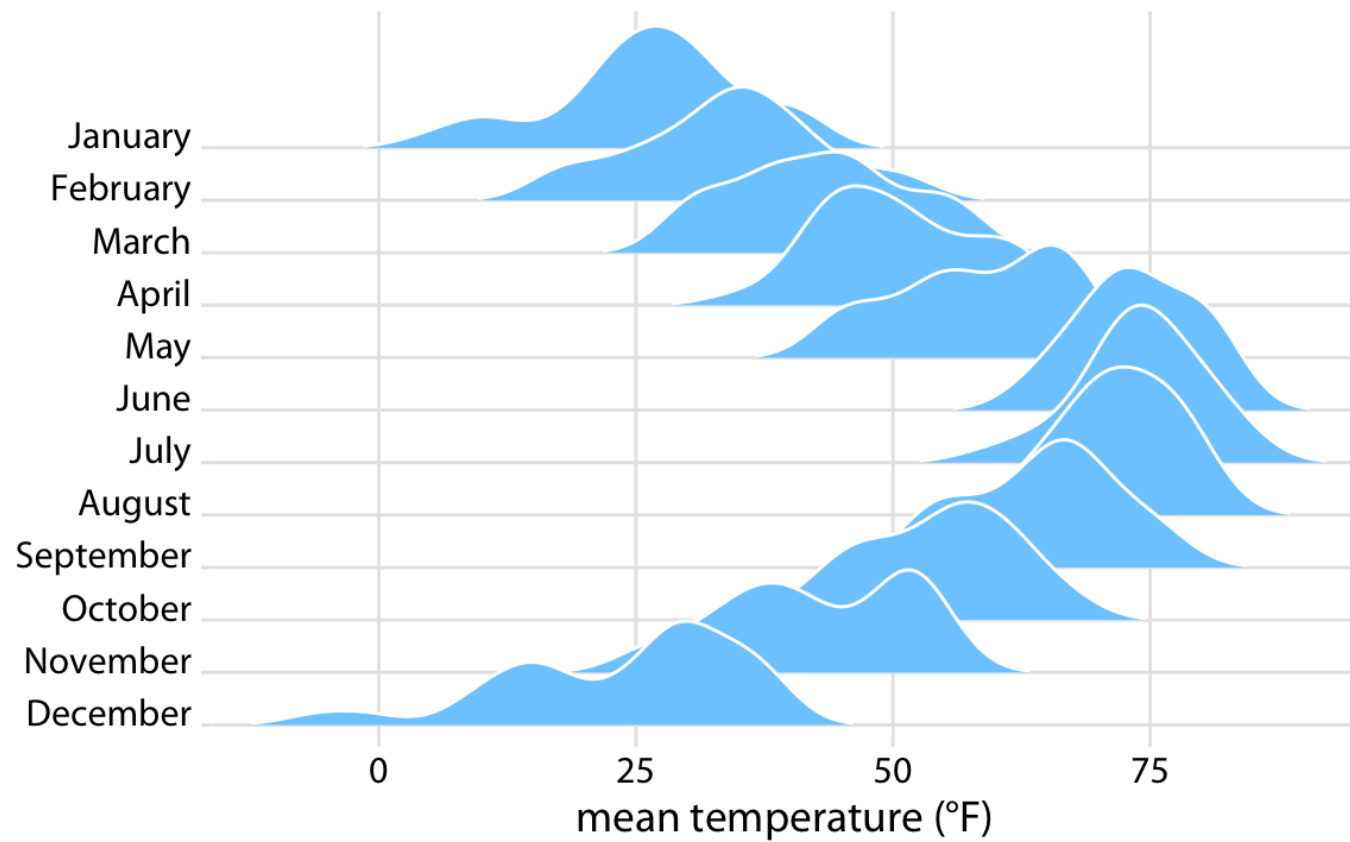
```
ggplot(df, aes(x, y)) + geom_violin()
```

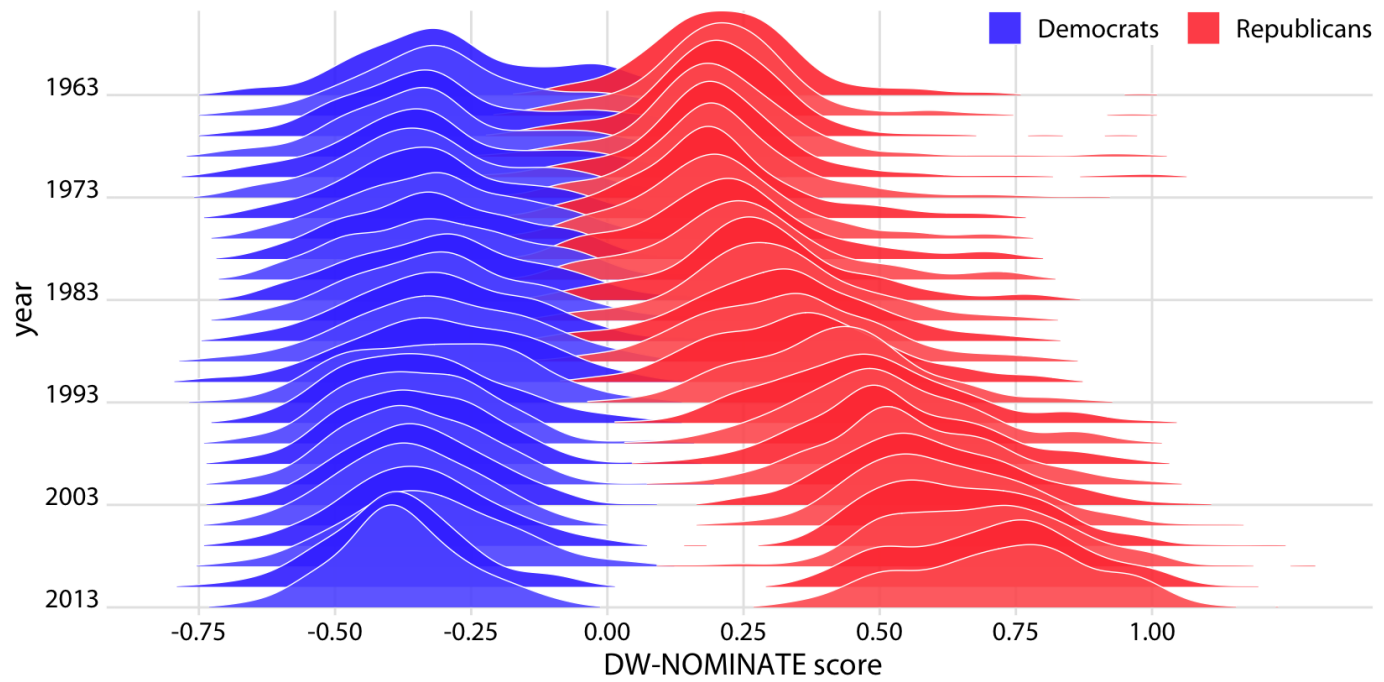




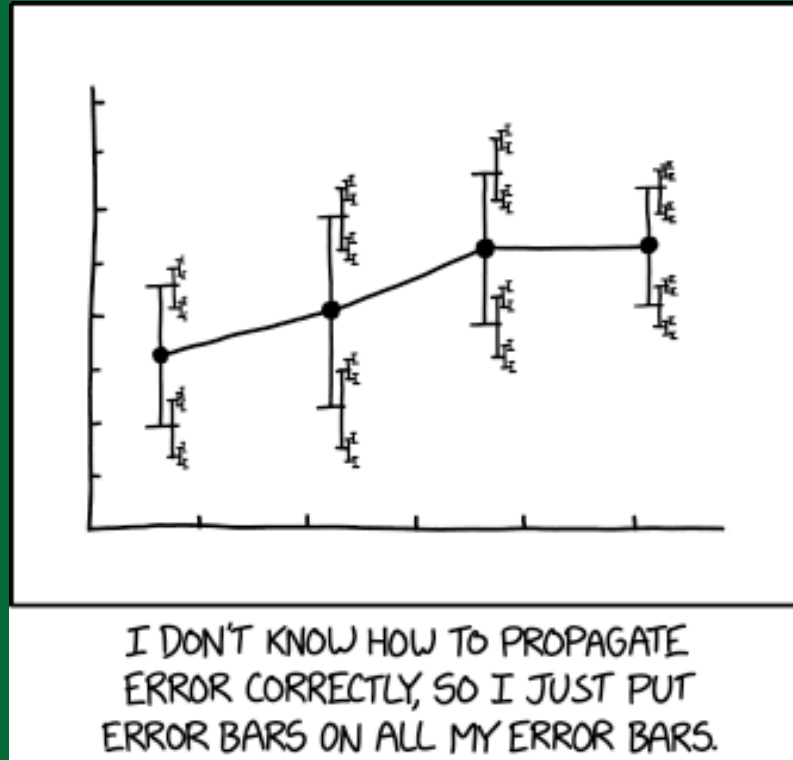








Uncertainty: Cairo Ch. 10 & Wilke Ch. 16



xkcd

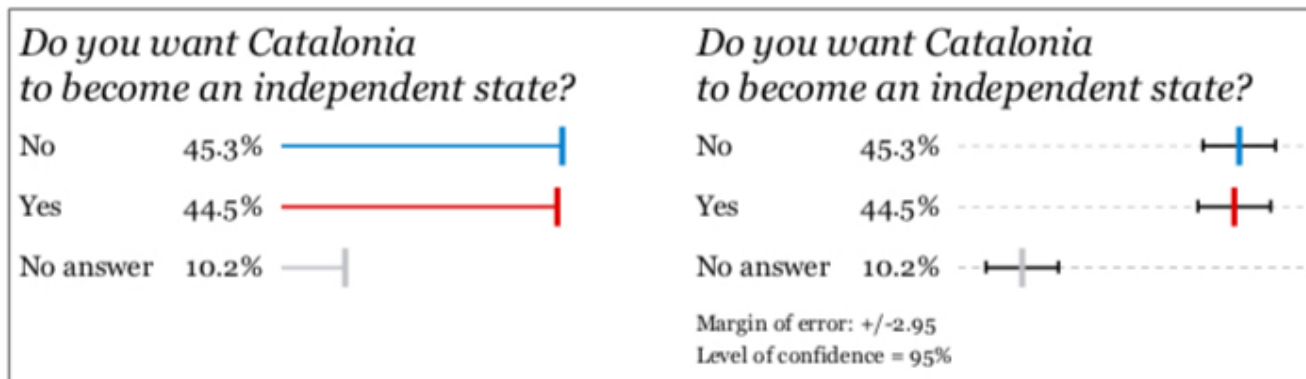


Figure 11.1 Displaying the margin of error can change your view of the data.

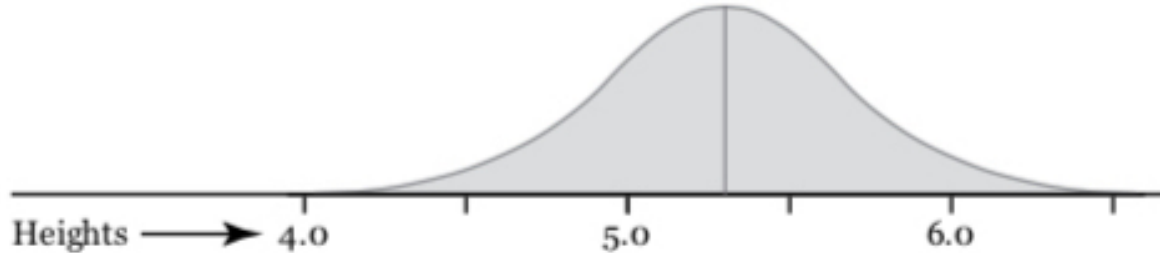
**POPULATION
distribution
(Unknown to you)**

I've drawn a normal distribution,
but the actual population distribution
could have a different shape



SAMPLE A

Mean: 5.3 feet



SAMPLE B

Mean: 5.25 feet

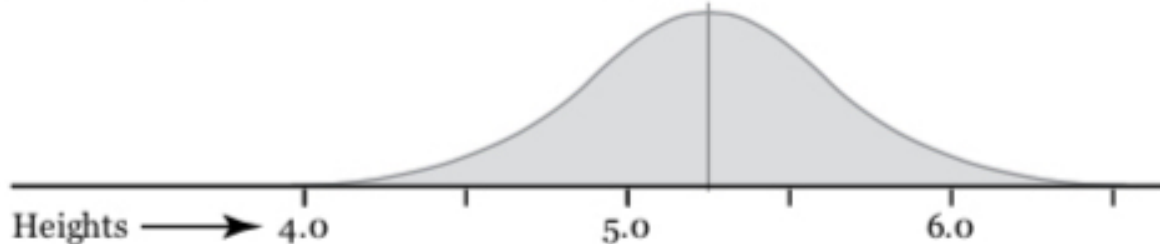
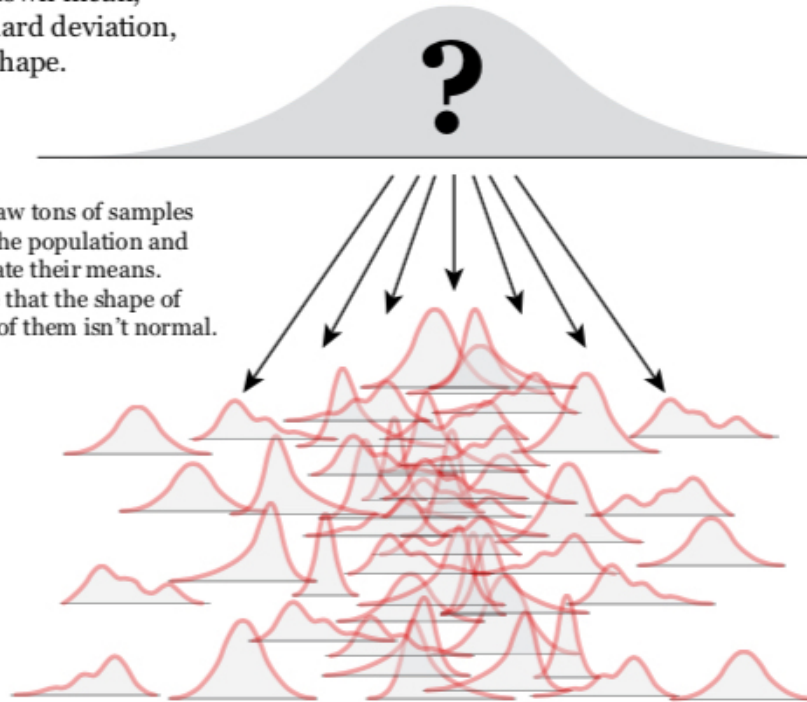


Figure 11.2 Population and samples.

Population:

Unknown mean,
standard deviation,
and shape.

We draw tons of samples
from the population and
calculate their means.
Notice that the shape of
many of them isn't normal.



We plot the distribution
of the means of the
samples above. The result
is approximately a **normal**
distribution of
sample means.

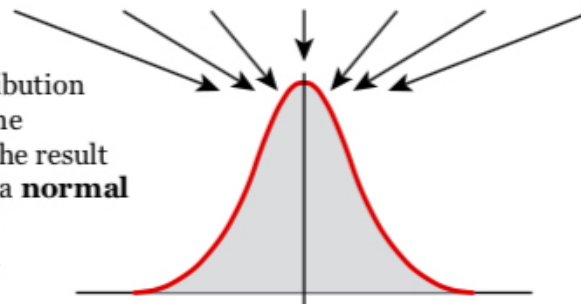
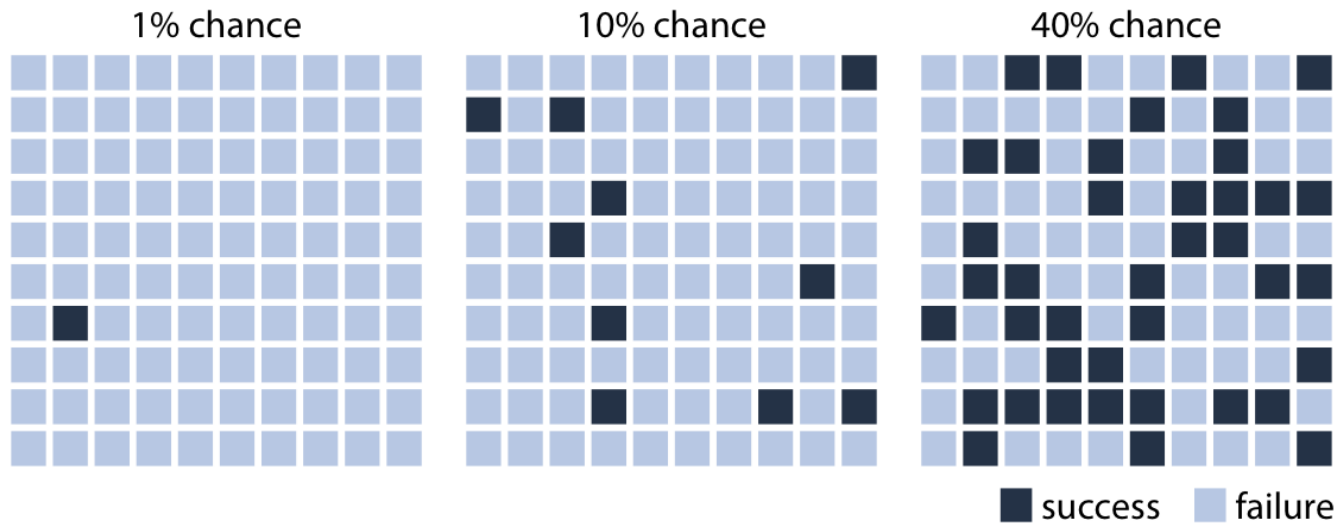
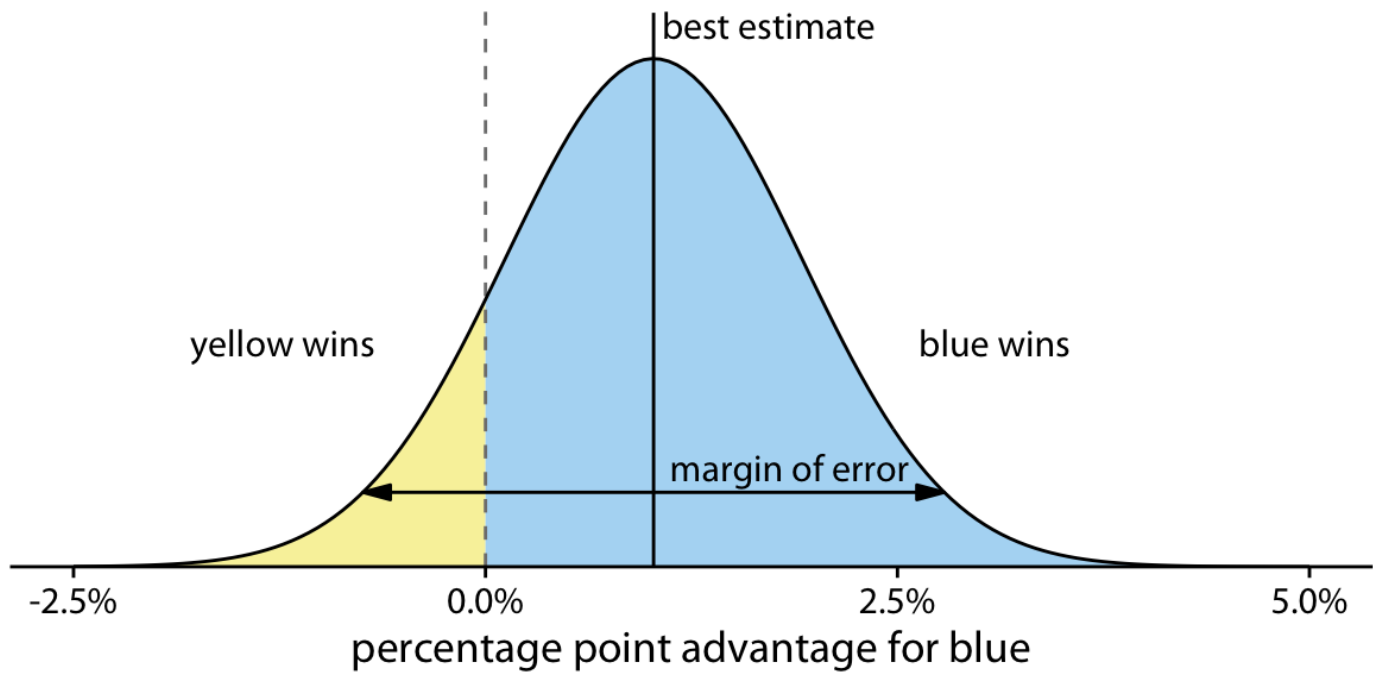


Figure 11.5 The distribution of (imaginary) sample means.

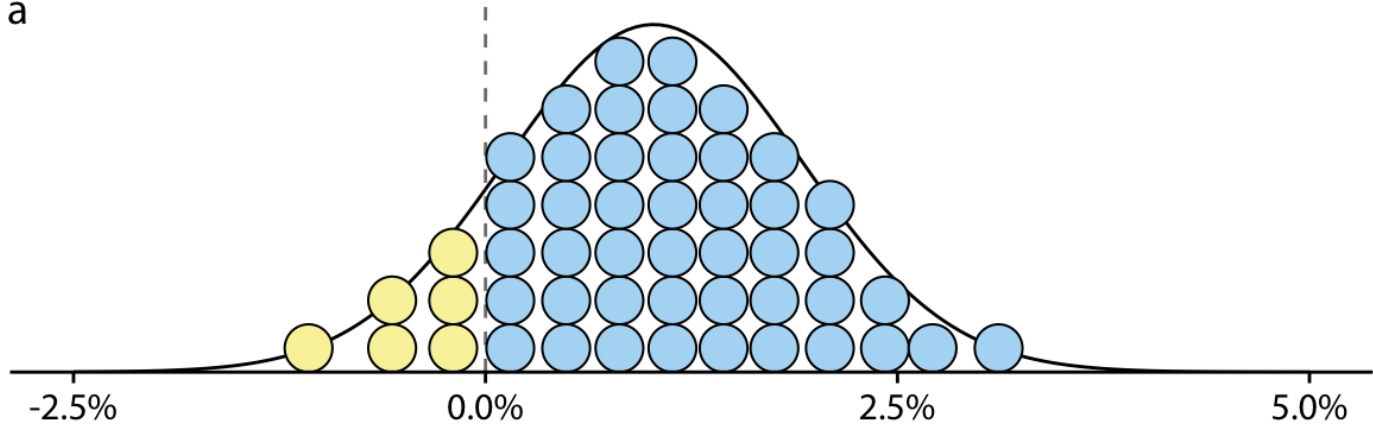
Why is visualizing uncertainty hard?

- Efficient encodings for uncertainty can be hard to find.
- Make sure people understand encodings (what does the plot mean?).
- Perceptual models of probability (e.g., quantile dot plot, HOP).
- Decisions under uncertainty (e.g., Gigerenzer et al or Monty Hall problem).
- Findings may not apply in all contexts.
- Plus, you still have to actually build it!

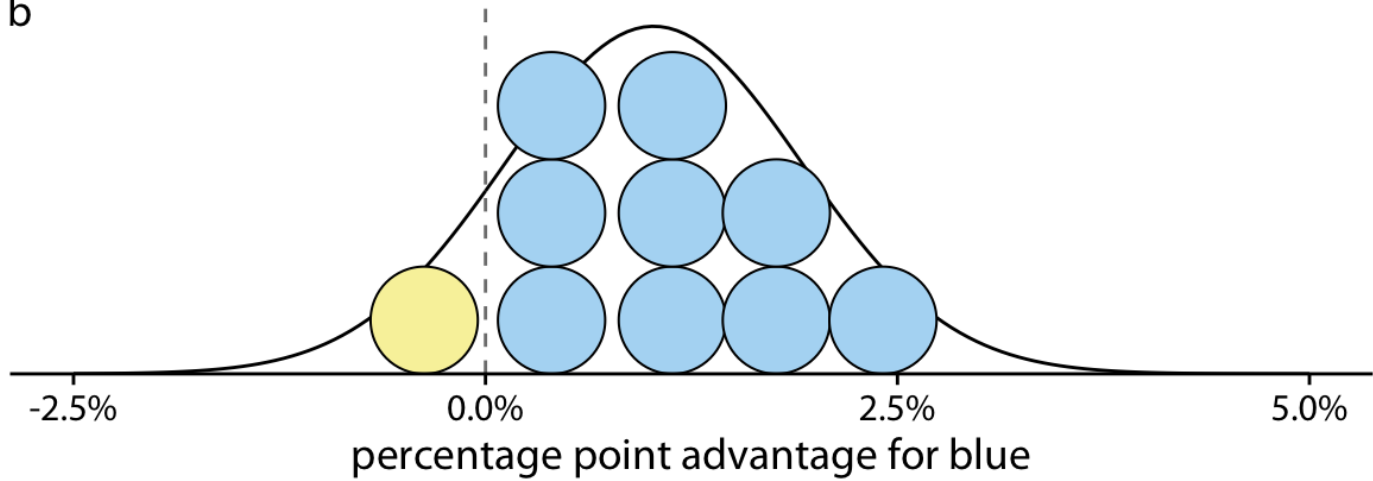




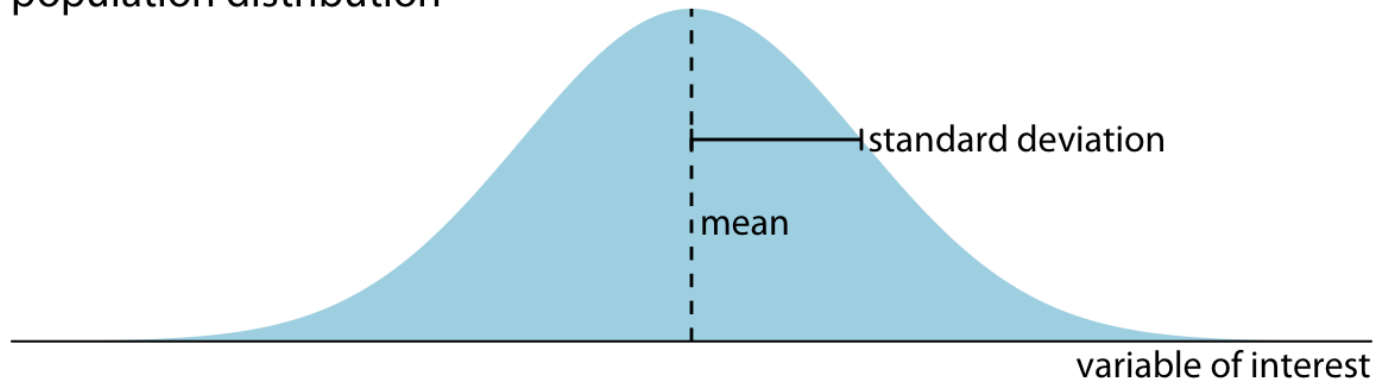
a



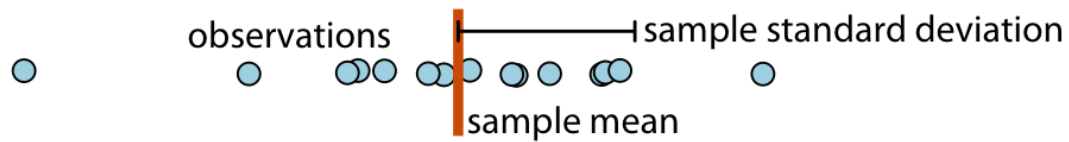
b



population distribution



sample



sampling distribution of the mean

