



DataCamp

Learning by doing

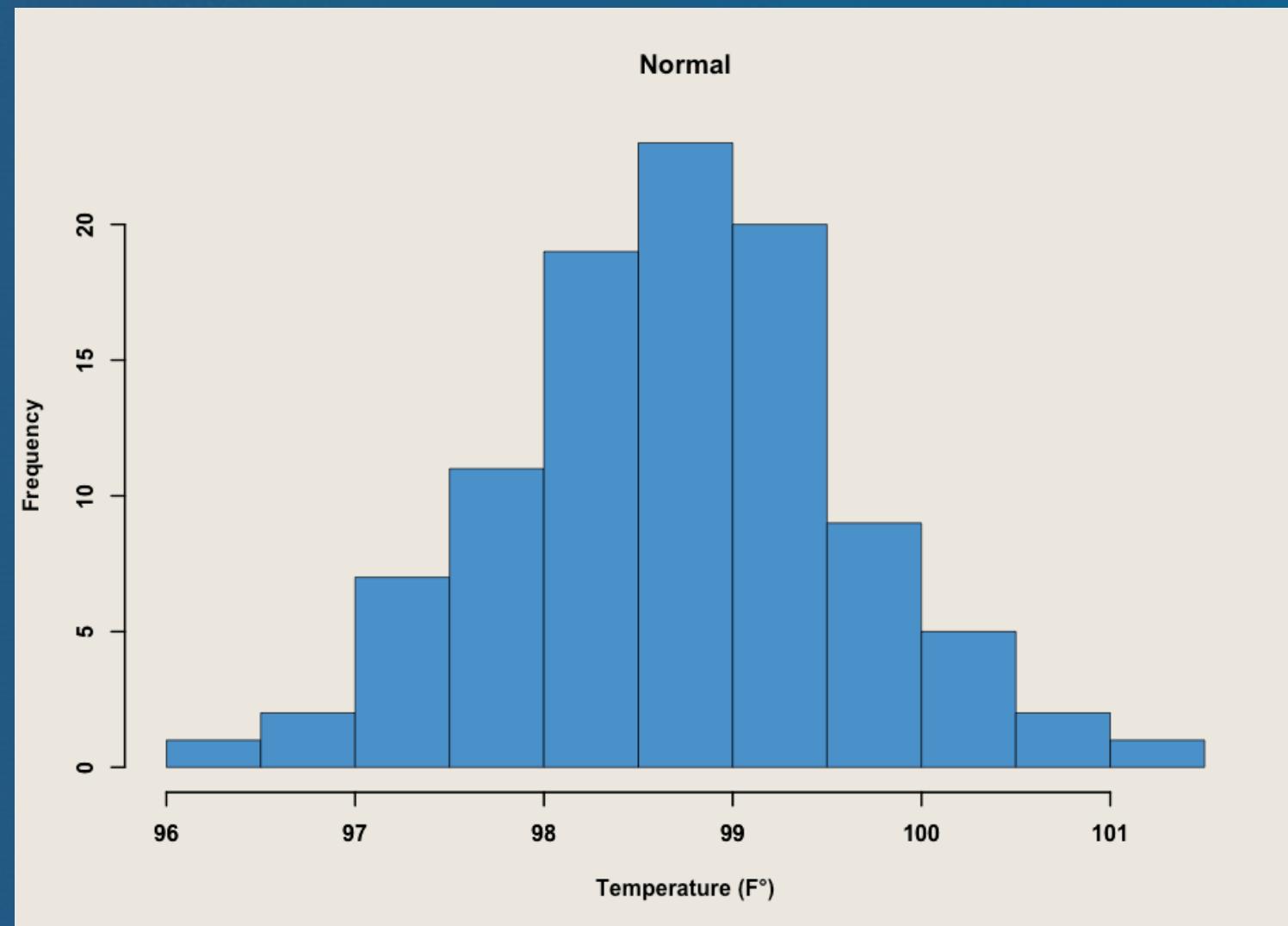
Histograms

- A histogram is a type of graph used to display a distribution

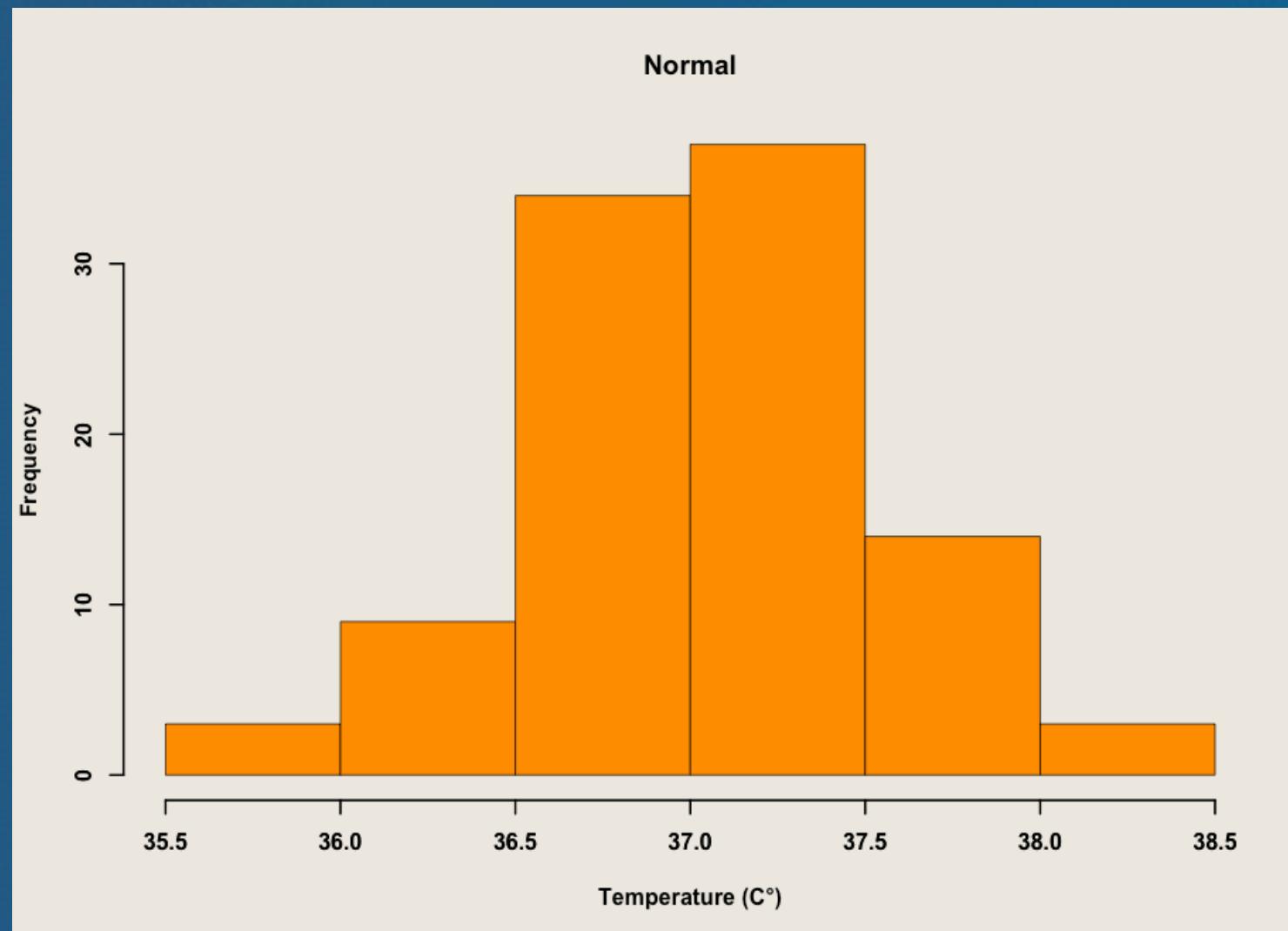
Histograms

- Why start with histograms?
 - To overcome the natural tendency to rely upon summary information, such as an average

An example: Body temperature



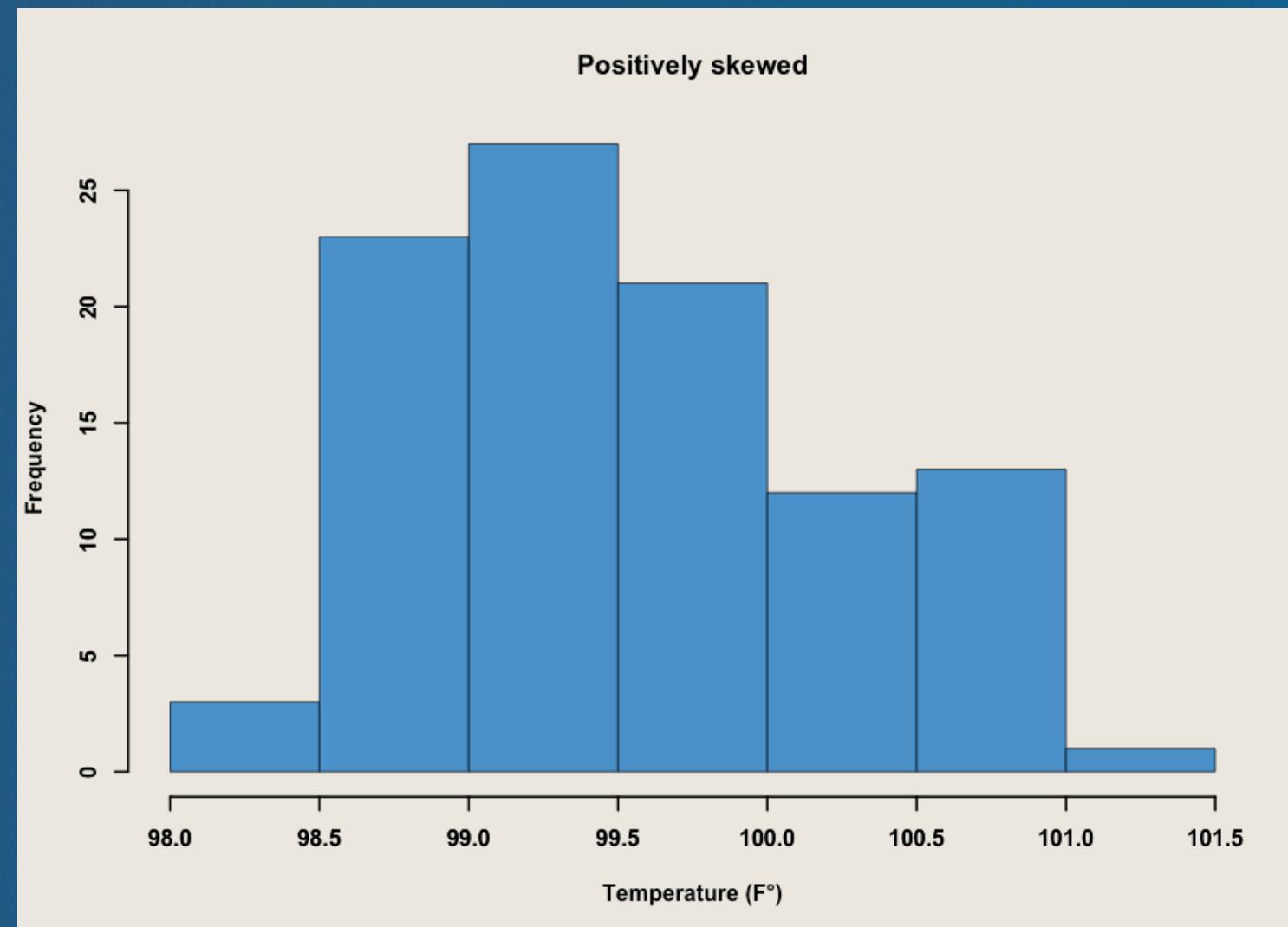
An example: Body temperature



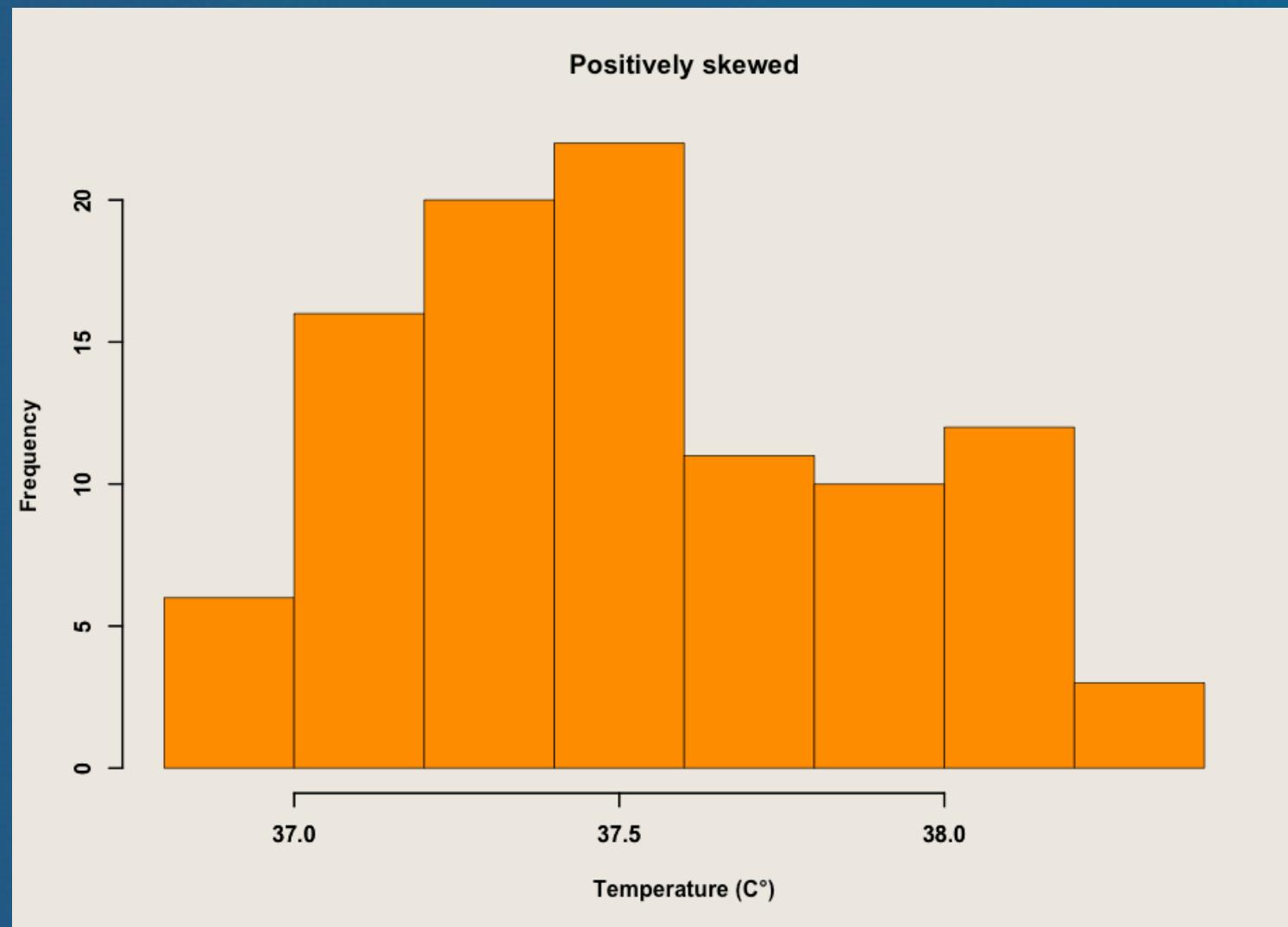
Histograms

- Histograms can reveal information not captured by summary statistics
 - Suppose a few children in a school are sick with influenza (flu) and have a high temperature
 - The distribution will be positively skewed

An example: Body temperature



An example: Body temperature

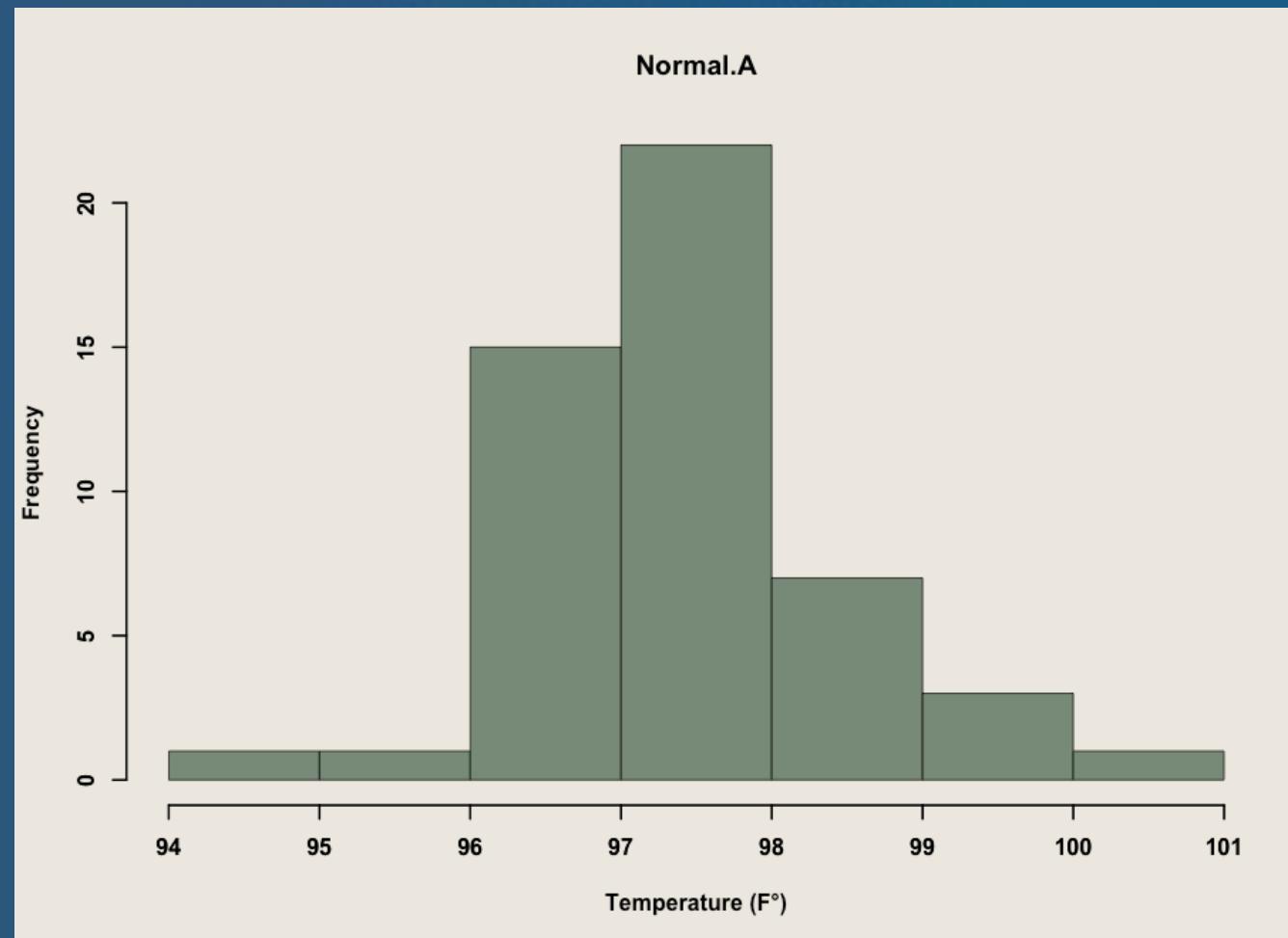


Histograms

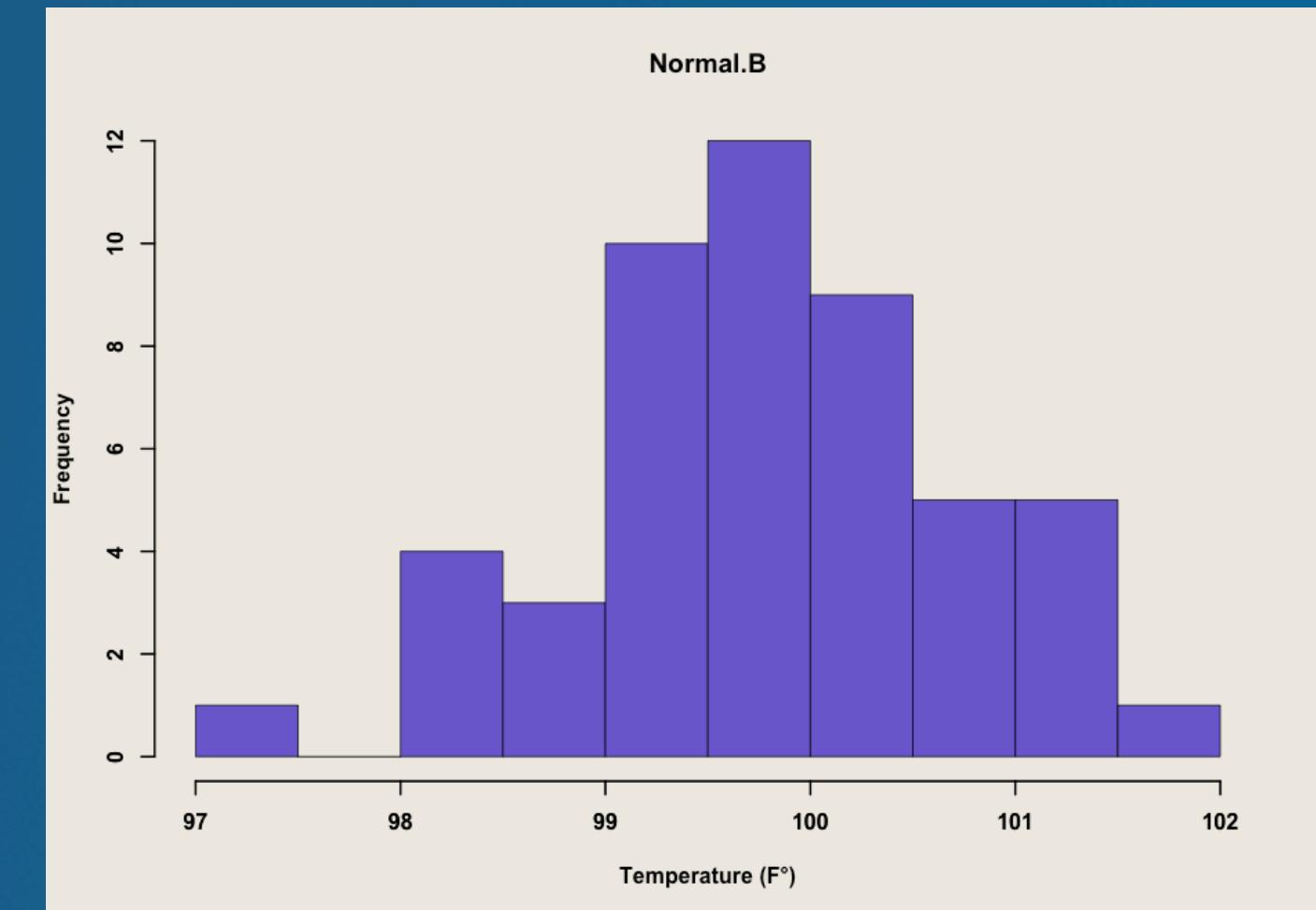
- Not all distributions are normal
 - Suppose one group of children had the flu a week prior to a second sick group of children
 - Assume the first group received antibiotics, which temporarily caused their body temperatures to be slightly below normal, while the second group was still above normal

An example: Body temperature

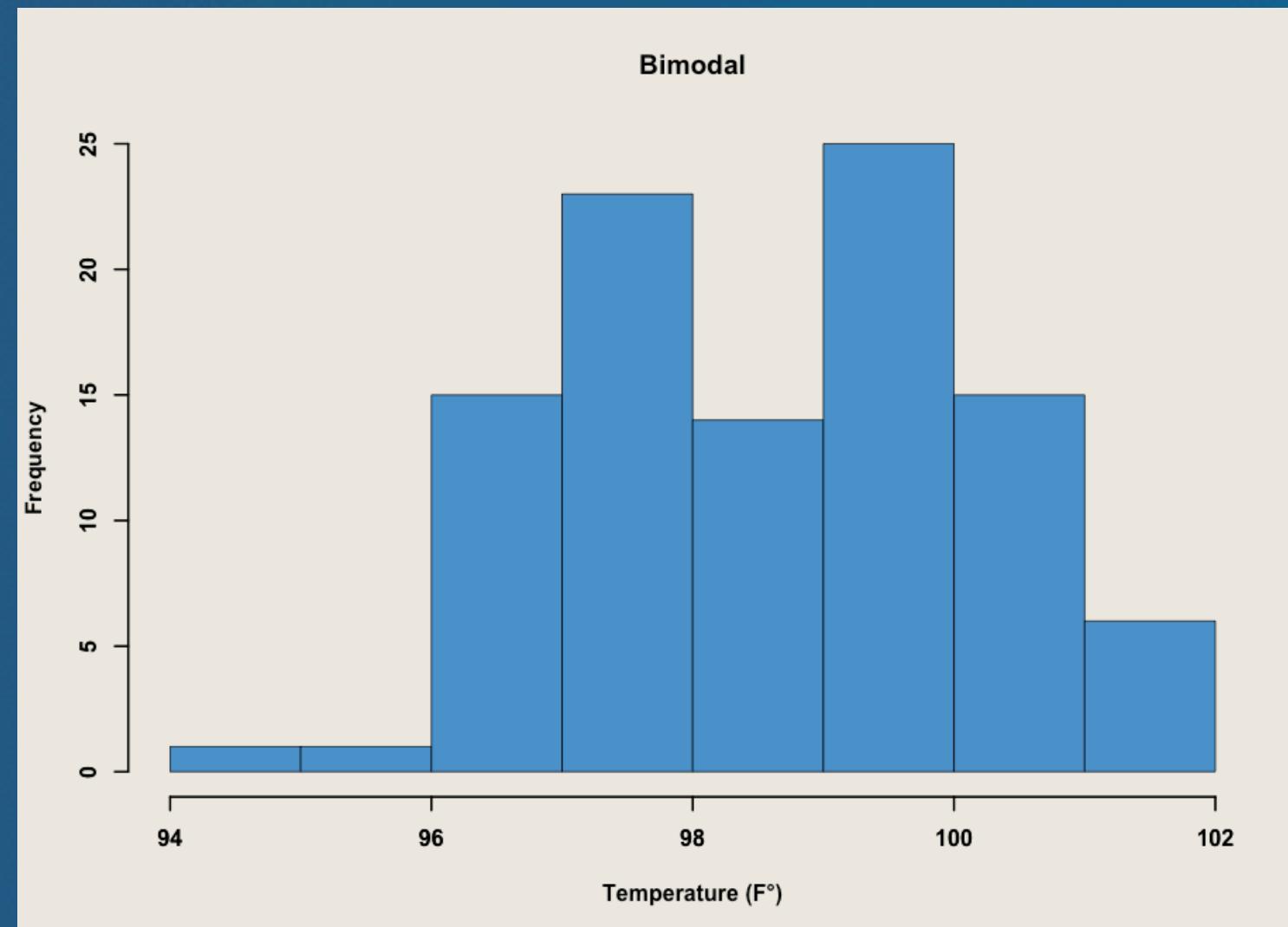
Normal, below average



Normal, above average

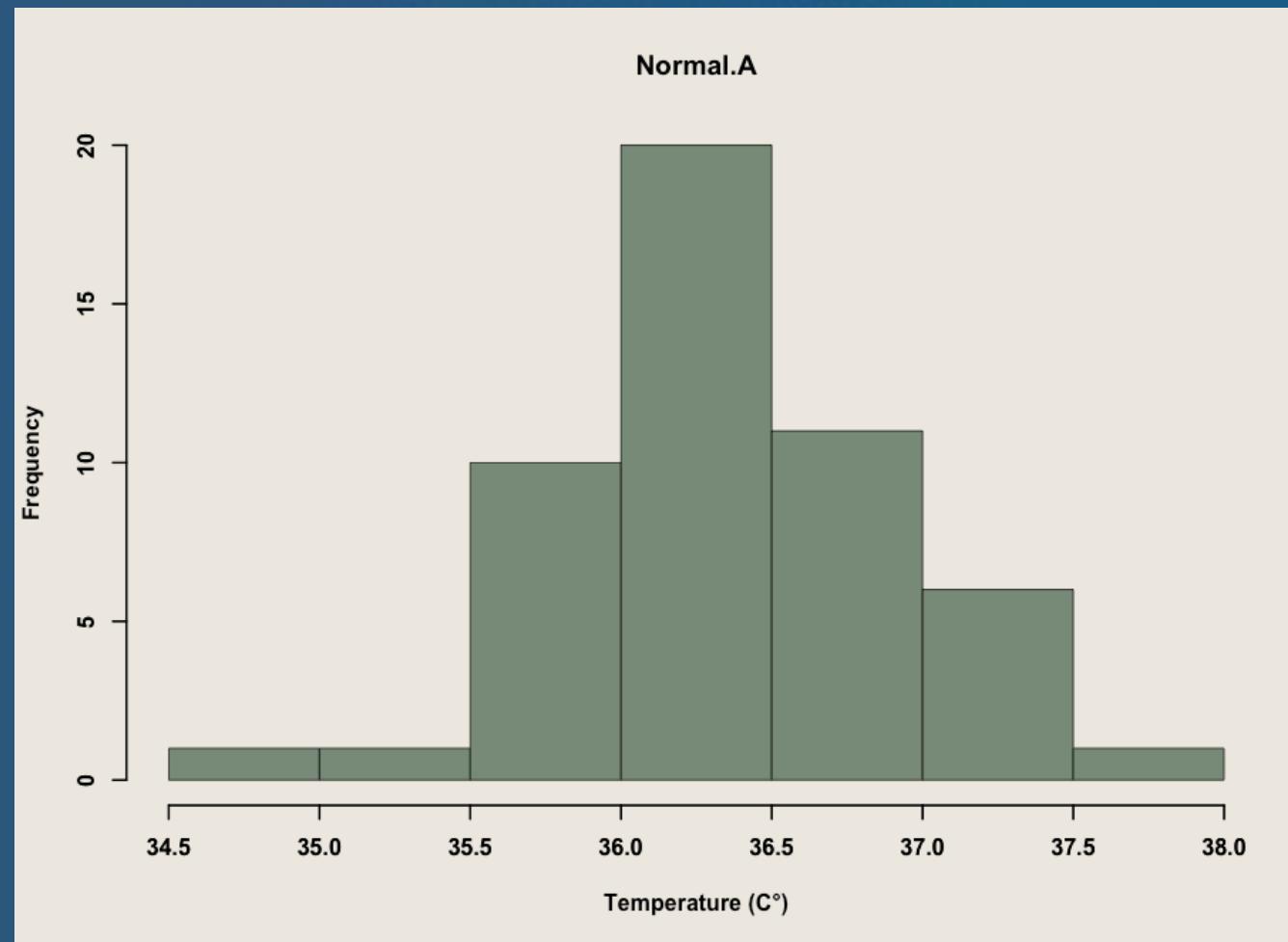


An example: Body temperature

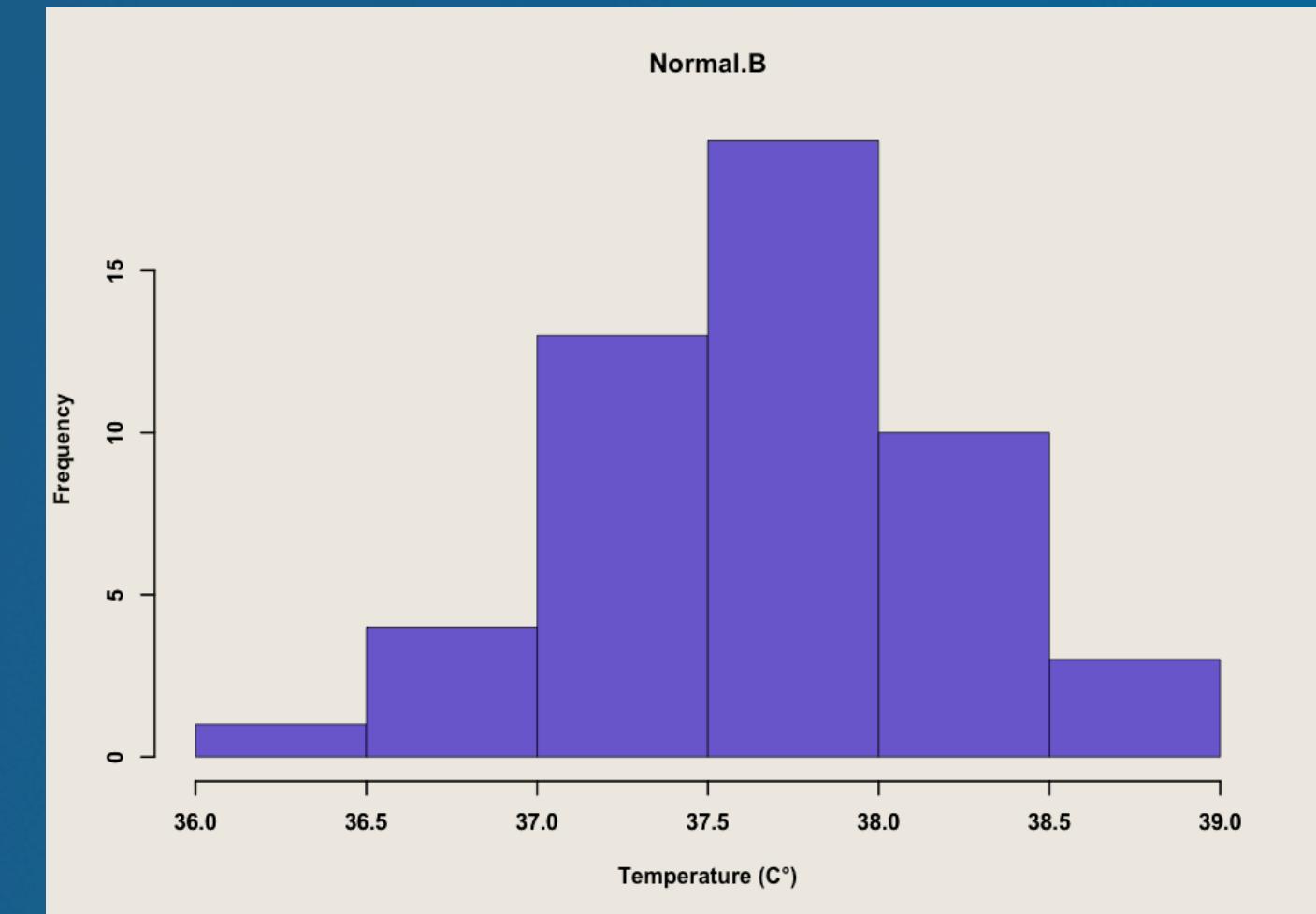


An example: Body temperature

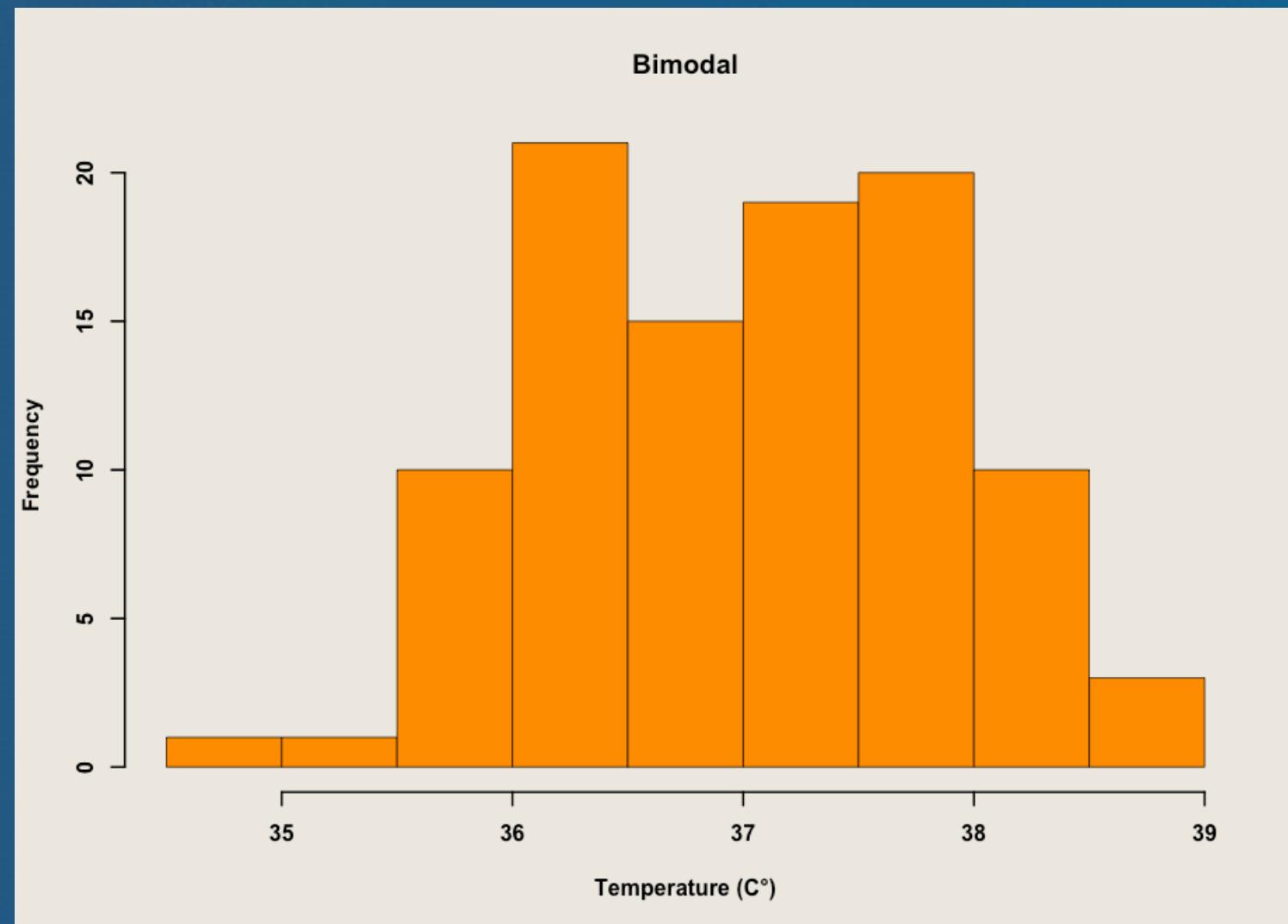
Normal, below average



Normal, above average



An example: Body temperature



Histograms

- Not all distributions are normal
 - Simply viewing a histogram often reveals whether a distribution is normal or not normal
 - However, sometimes it is hard to determine
 - Summary statistics help in such cases

Histograms

- Not all distributions are normal
 - As you view more and more distributions you will get a better sense of what is normal and what is not normal
 - So, let's look at more distributions

Wine tasting!



An example: Wine ratings

- Suppose that 100 wine experts rated the overall quality of 8 different wines on a scale of 1 to 100
 - Higher scores indicate higher quality

An example: Wine ratings

- Suppose four countries submitted two wines each, one red and one white
 - Argentina
 - Australia
 - France
 - USA

An example: Wine ratings

Malbec & Chardonnay

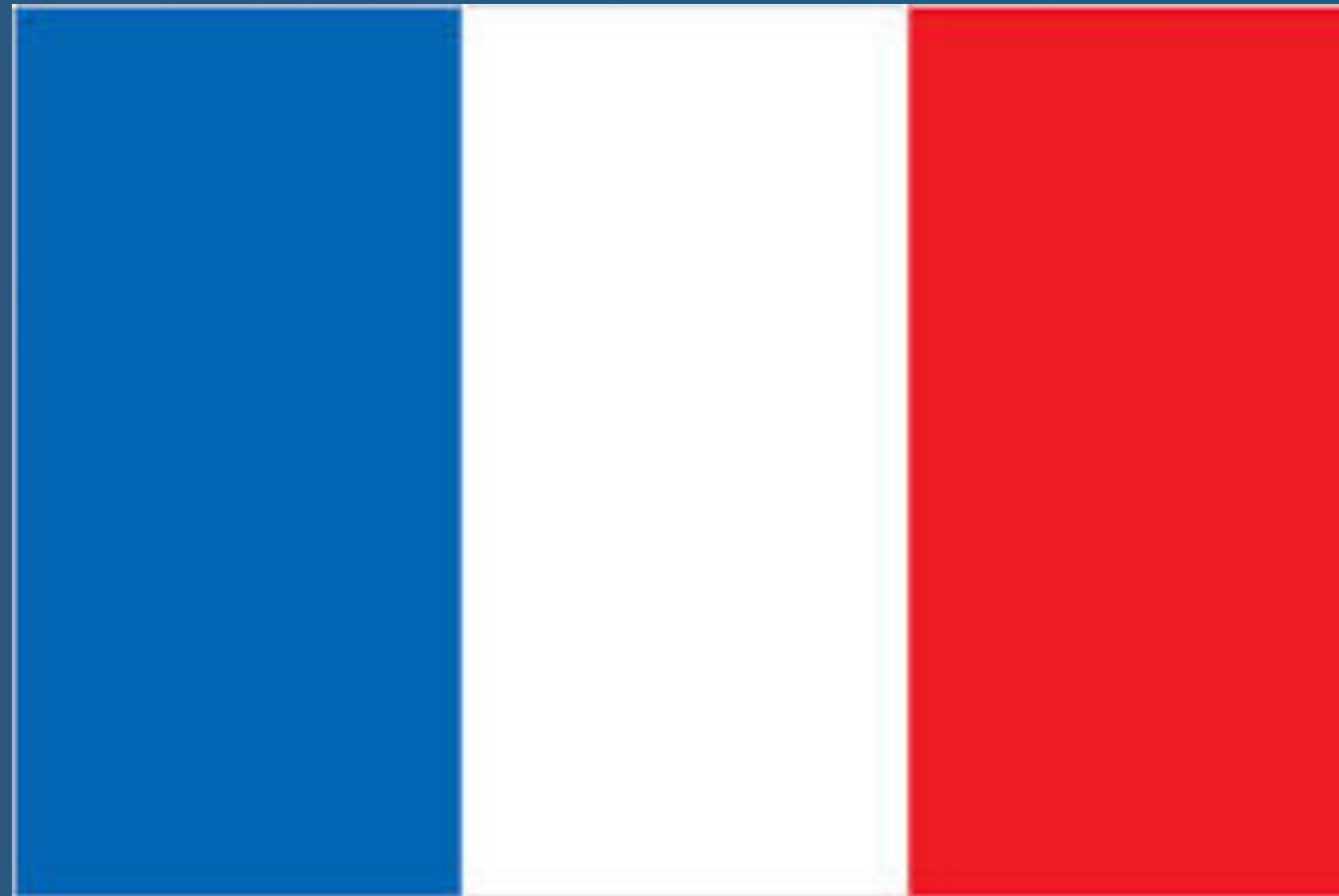


Shiraz & Pinot Grigio



An example: Wine ratings

Bourdeaux & Sauvignon Blanc



Cabarnet & Reisling



An example: Wine ratings

- Preview
 - The ratings of the red wines are normal
 - The ratings of the whites are not normal

Segment summary

- Histograms are used to display distributions
- Many distributions are normal

Segment summary

- Some distributions are not normal, for example:
 - Bi-modal
 - Positively skewed
 - Negatively skewed
 - Uniform (platykurtic)
 - Leptokurtic

Advanced graphs

