Boxplots
Boxplots

0:00 / 0:00 1.0x

Start of transcript. Skip to the end.

RAFAEL IRIZARRY: To introduce box plots, we will go back to the US murders data.

Suppose we want to summarize the murder rate distribution.

Using the data visualization techniques we have learned,

we notice that the normal approximation does not

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Textbook link

This video corresponds to the textbook section on boxplots.

Key points

- When data do not follow a normal distribution and cannot be succinctly summarized by only the mean and standard deviation, an alternative is to report a five-number summary: range (ignoring outliers) and the quartiles (25th, 50th, 75th percentile).
- In a *boxplot*, the box is defined by the 25th and 75th percentiles and the median is a horizontal line through the box. The whiskers show the range excluding outliers, and outliers are plotted separately as individual points.
- The *interquartile* range is the distance between the 25th and 75th percentiles.
- Boxplots are particularly useful when comparing multiple distributions.
- We discuss outliers in a later video.