Summarizing Quantitative Data: Part 4

- Topics: Box Plots & Outliers
- Objective: Students will be able to interpret data using box plots and be able to identify outliers of a given set of data.
- Standards: AP Stats: UNC 1 (EU), UNC 1.L (LO), UNC 1.L.1 (EK), UNC 1.L.2 (EK)
  CCSS Math: 6.SP.B.5, 6.SP.B.5c

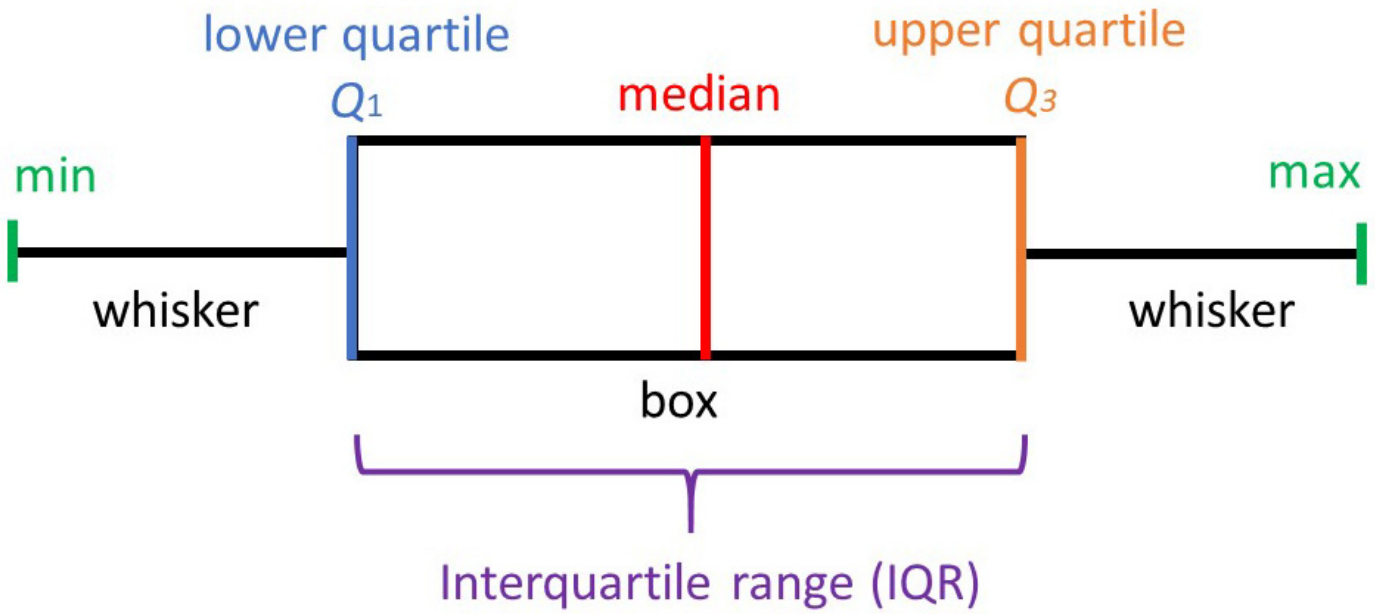
What is a box and whisker plot?

A box and whisker plot—also called a box plot—displays the five-number summary of a set of data. The five-number summary is the minimum, first quartile, median, third quartile, and maximum.

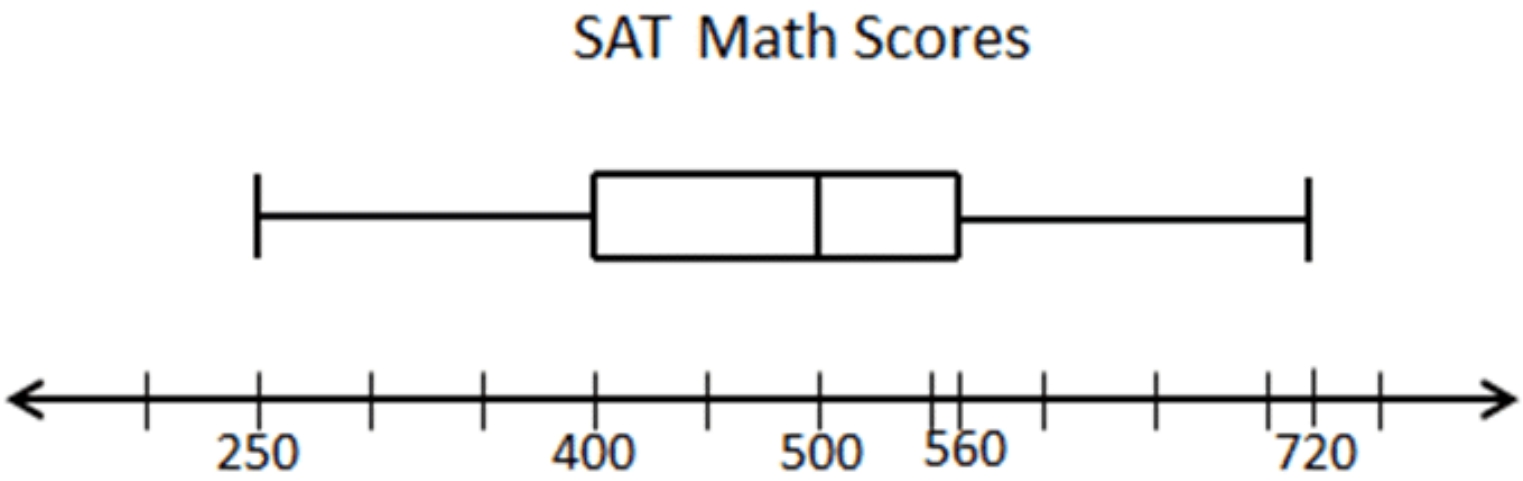
In a box plot, we draw a box from the first quartile to the third quartile. A vertical line goes through the box at the median. The whiskers go from each quartile to the minimum or maximum.

$IQR = Q3 - Q1$

Box Plots


lower quartile
Q_1
median
upper quartile
Q_3
min
max
whisker
whisker
box
Interquartile range (IQR)

Box Plots



Creating Box Plots

- Example 1: The data below represents the number of runners on each cross country team in the Northern Conference.

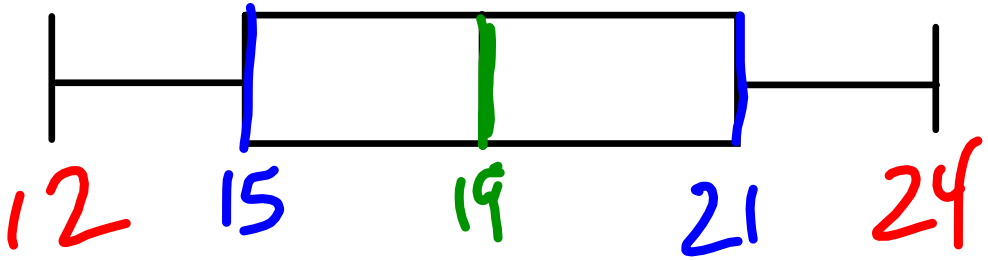
odd: there's a median
even: calculate median

12, 15, 15, 17, 19, 19, 20, 22, 24

median: 19
Q1: 15
Q3: 21

12 15 21

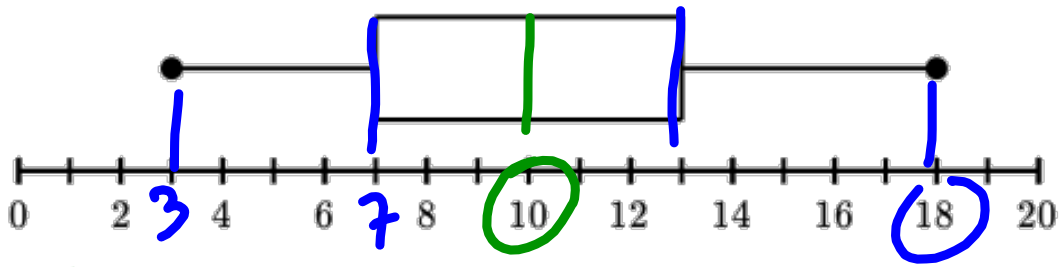
Create a box plot for the data.



Creating Box Plots

Which data set could be represented by the box plot shown below?

- Example 2:

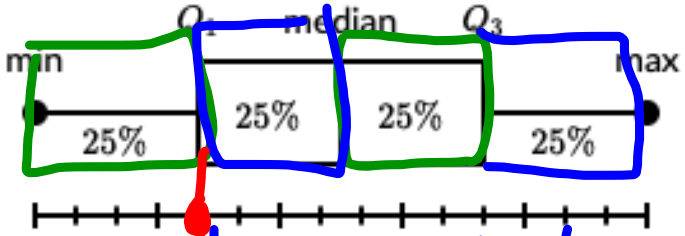


- 3, 4, 8, 9, 9, 12, 12, 13, 13, 16, 18
- 2, 4, 7, 9, 9, 10, 12, 13, 13, 16, 18
- 3, 4, 8, 9, 9, 10, 12, 13, 13, 16, 18
- 3, 4, 7, 9, 9, 10, 12, 13, 13, 16, 18

min 3
Q1 7
med 10
Q3 13
max 18

Interpreting quartiles

The five-number summary divides the data into sections that each contain approximately 25% of the data in that set.



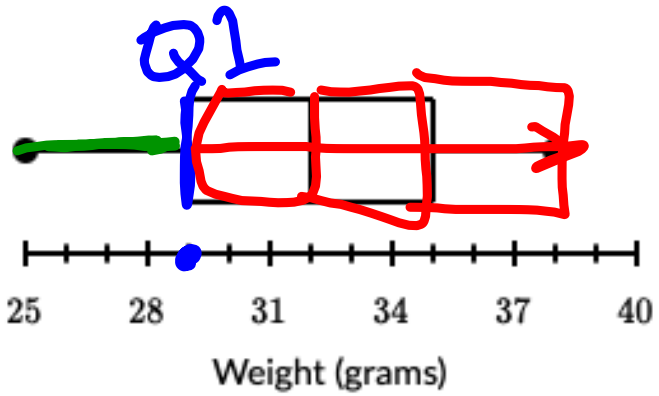
Quartile = 25%

25% less than / fewer than

75% more/greater than

Example: Interpreting quartiles

About what percent of the boxes of raisins weighed more than 29 grams?



75%

Outliers and How to Find Them...

An outlier is a data point that lies outside the overall pattern in a distribution.

The distribution below shows the scores on a driver's test for 19 applicants. How many outliers do you see?

0 - 1

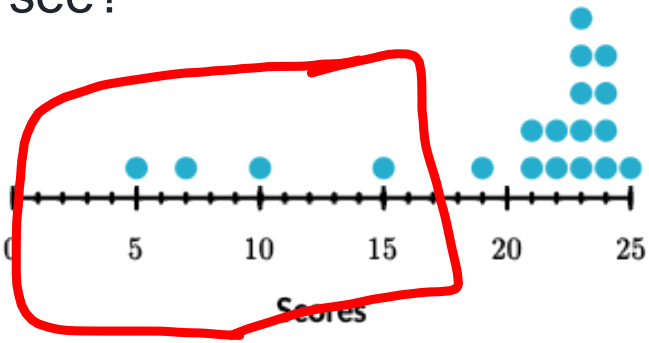
1 - 0

2 - 3

3 - 2 ✓

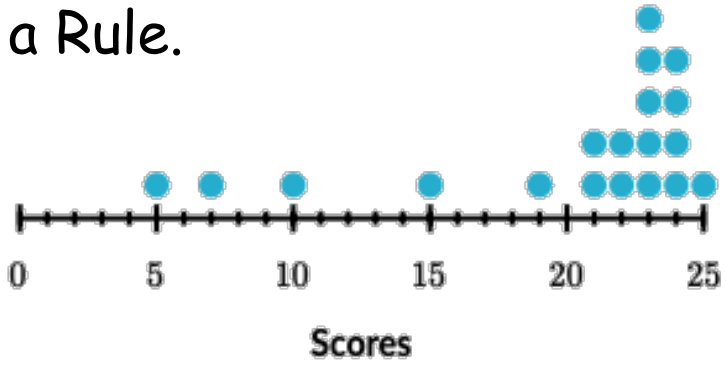
4 - 4.5

5 - .5



Outliers and How to Find Them...

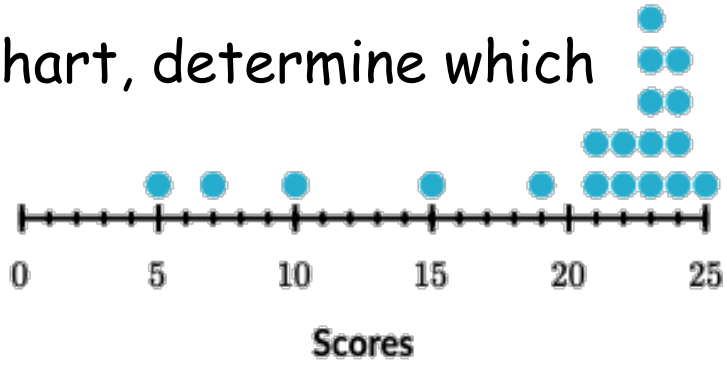
How to tell, For Real...there's a Rule.



- The 1.5*IQR Rule:
  - > Calculate or determine Q1, Q3, & the IQR.
    - Multiply 1.5 * IQR
      1. Subtract the number from Q1
      2. Add the number to Q3
      - « Any datum below #1, or above #2 is an outlier.

Outliers and How to Find Them...

Example: Given the data in the chart, determine which data points are outliers.



5, 7, 10, 15, 19, 21, 21, 22, 22, 23, 23, 23, 23, 23, 24, 24, 24, 24, 26

Q3 - 24

- Q1 - 19

IQR: 5

5(1.5) = 7.5

Q3 + 7.5

24 + 7.5 = 31.5

Q1 - 7.5

19 - 7.5 = 12.5

Mean Absolute Deviation (MAD)

- The mean absolute deviation of a dataset is the average distance between each data point and the mean. It gives us an idea about the variability in a dataset.

- How to calculate Mean Absolute Deviation (MAD)
  1. Calculate the mean $\bar{x}$ x-bar
  2. Calculate how far away each data point is from the mean using positive distances. These are called absolute deviations.
  3. Add those deviations together.
  4. Divide the sum by the number of data points. n

Mean Absolute Deviation (MAD)

- Example: Erica enjoys posting pictures of her cat online. Here's how many "likes" the past 6 pictures each received: 10, 15, 15, 17, 18, 21

$\bar{x} = 16 \quad n = 6$

$10 - 16 \rightarrow |-6| = 6$

$15 - 16 \rightarrow |-1| = 1$

$15 - 16 \rightarrow |-1| = 1$

$17 - 16 \rightarrow |1| = 1$

$18 - 16 \rightarrow |2| = 2$

$21 - 16 \rightarrow |5| = 5$

16

$\frac{16}{6} = 2.67$

2.7

MAD = $8/3$

Displaying and Comparing Quantitative Data

You should be working on the following skills:

1. Creating Box Plots
2. Reading Box Plots
3. Interpreting Quartiles
4. Identifying Outliers
5. Mean Absolute Deviation (MAD)