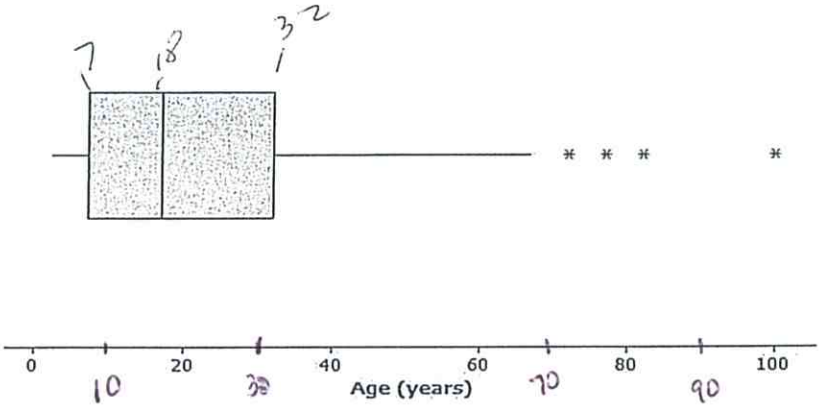
...oratory Challenge 3/Exercises 15–20: Outliers

...udents at Waldo High School are involved in a special project that involves communicating with people in Kenya. Consider a box plot of the ages of 200 randomly selected people from Kenya.

Box Plot of Ages for Kenya



A data distribution may contain extreme data (specific data values that are unusually large or unusually small relative to the median and the interquartile range). A box plot can be used to display extreme data values that are identified as **outliers.**

Each "*" in the box plot represents the ages of four people from this sample. Based on the sample, these four ages were considered outliers.

15. Estimate the values of the four ages represented by an *.

72, 77, 82, 100

An outlier is defined to be any data value that is more than $1.5 \times (IQR)$ away from the nearest quartile.

16. What is the median age of the sample of ages from Kenya? What are the approximate values of Q1 and Q3? What is the approximate IQR of this sample?

median = 18 $Q_1 = 7$ $Q_3 = 32$

$IQR = 32 - 7 = 25$ years

17. Multiply the IQR by 1.5. What value do you get?

$25(1.5) = 37.5$

18. Add $1.5 \times (IQR)$ to the 3rd quartile age (Q3). What do you notice about the four ages identified by an *?

$32 + 37.5 = 69.5$ They are all higher than 69.5

19. Are there any age values that are less than $Q1 - 1.5 \times (IQR)$? If so, these ages would also be considered outliers.

$7 - 37.5 = -30.5$ None

non outliers will be in ~~between~~ this interval $(-30.5, 69.5)$ but since you can't have neg ages, it's really $(0, 69.5)$ Any age above is an outlier

20. Explain why there is no * on the low side of the box plot for ages of the people in the sample from Kenya.

no negative ~~ages~~ can't have ages