Chapter 1: Summerize the Data with Median and Quartiles

- The median and the quartiles provide another way to describe center and spread.

- The median M is the midpoint of a distribution, the number such that half of the observations are smaller and the other half are larger. To find the median of a distribution
  - Arrange all observations in order of size, from smallest to largest.
  - If the number of observations n is odd, the median M is the center of observation in the ordered list. Find the location of median by counting $\frac{n+1}{2}$ obervations up from the bottom of the list.
  - If the number of observations n is even, the median M is the average of the two center observations in the ordered list. The location of the median is between $\frac{n}{2}$ and $\frac{n}{2}+1$ from the bottom of the list

Chapter 1: Summerize the Data with Median and Quartiles

- We use the first and the third quartiles to mark off the middle for each half of the observations.

- The first quartile Q_1 is the median of the observations whose position in the ordered list is to the left of the location of the overall median. The overall median is not included in the observations considered to be to the left of the overall median.

- The third quartile Q_3 is the median of the observations whose position in the ordered list is to the right of the location of the overall median. The overall median is not included in the observations considered to be to the left of the overall median.

Odd Number Observations Example

Hank Aaron's 23 home run counts are

10	12	13	20	24	26	27	29	30	32	34	34
38	39	39	40	40	44	44	44	44	45	47	

Even Number Observations Example

Barry Bonds's 22 home run counts are

5	**16**	**19**	**24**	**25**	**25**	**26**	**28**	**33**	**33**	**34**
34	**37**	**37**	**40**	**42**	**45**	**45**	**46**	**46**	**49**	**73**

The Five-number summary

The five-number summary of a distribution consists of the smallest observation, the first quartile, the median, the third quartile, and the largest observation, written in order from smallest to largest. In symbols, the five-number summary is

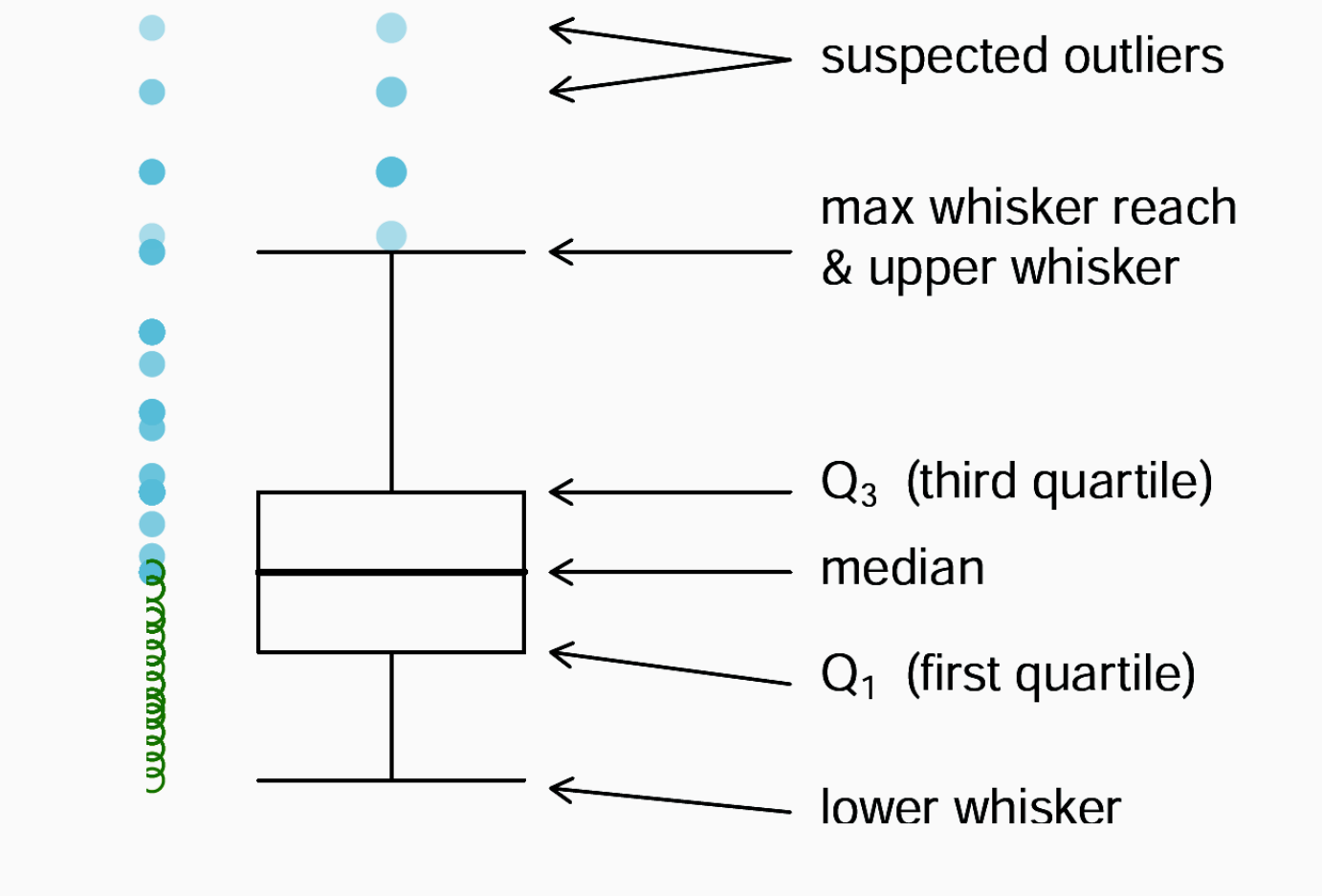
Minimum Q_1 M Q_3 Maximum

Whiskers Location in Boxplot

- **The upper whisker is located at the value of an observation which is the greatest value in the sample that is less than or equal to** $Q_3 + 1.5IQR$

- **The lower whisker is located at the value of an observation which is the smallest value in the sample that is greater than or equal to** $Q_1 - 1.5IQR$

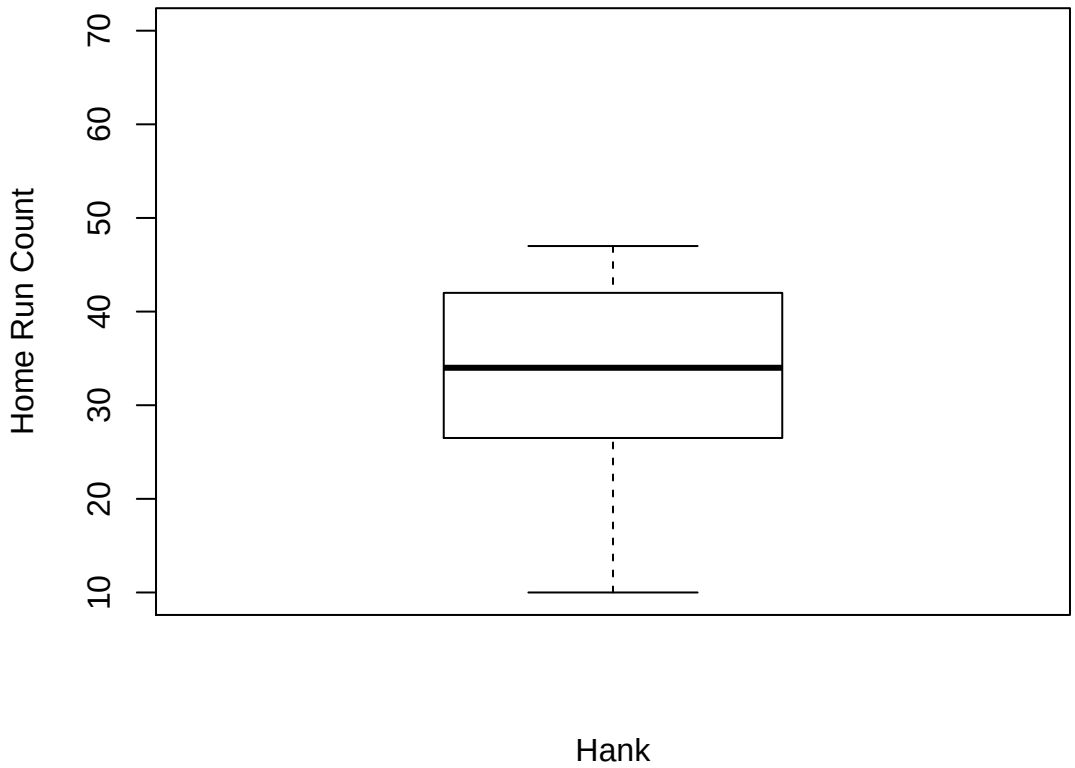
Anatomy of a box plot



Boxplot Example

Hank Aaron's 23 home run counts are

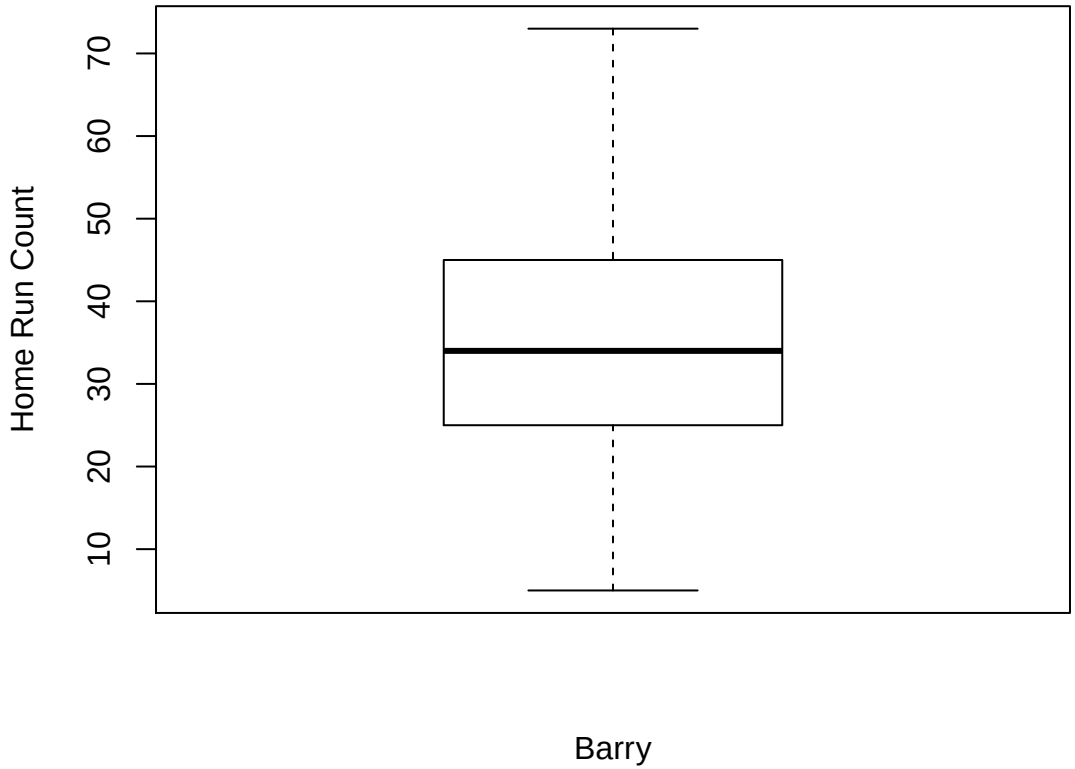
10	12	13	20	24	26	27	29	30	32	34	34
38	39	39	40	40	44	44	44	44	45	47	



Boxplot Example

Barry Bonds's 22 home run counts are

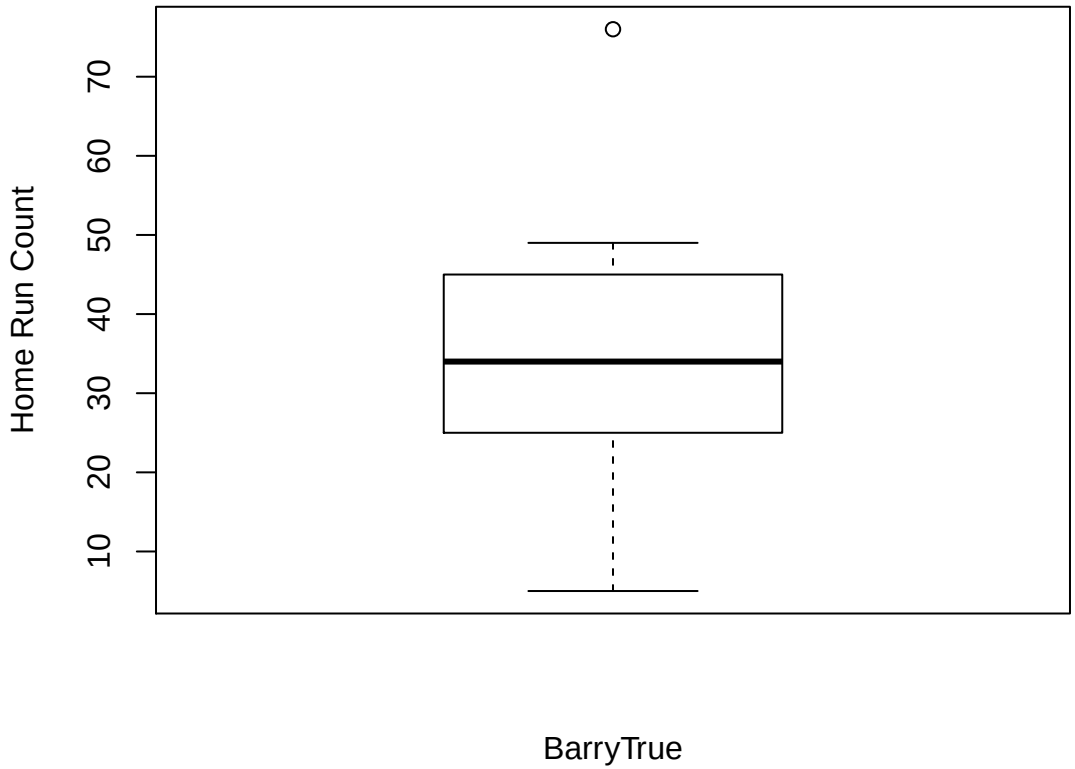
5	16	19	24	25	25	26	28	33	33	34
34	37	37	40	42	45	45	46	46	49	73



Boxplot Example

Suppose Barry Bonds's one of home run counts 73 is a typo, the true home run counts is 76. Then Barry Bonds's 22 home run counts called BarryTrue are

5	16	19	24	25	25	26	28	33	33	34
34	37	37	40	42	45	45	46	46	49	76



Sid-by-side Boxplot Example

Now we plot the boxplot side-by-side for Hank, Barry and BarryTrue