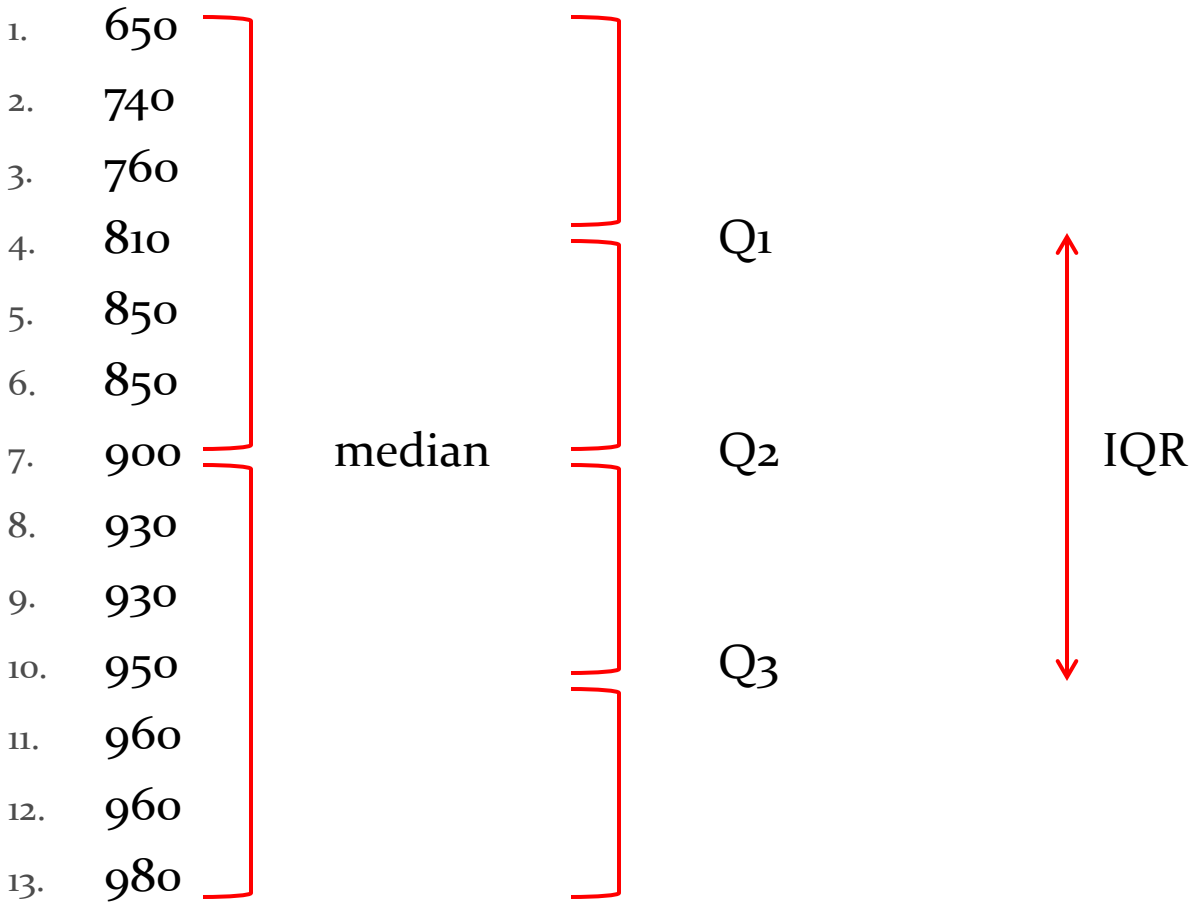
Basic statistics (continued)

Interquartile range

- like the mean, the standard deviation and variance are sensitive to erroneous measurements
- the interquartile range is a more robust measure of dispersion or spread
  - quartiles
    - the three points that divide a sorted set of measurements into four equal groups
    - Q_2: 2nd quartile = median
    - Q_1: 1st quartile is the measurement that is midway between the smallest measurement and the median
    - Q_3: 3rd quartile is the measurement that is midway between the median and the largest measurement
- interquartile range $IQR = Q_3 - Q_1$

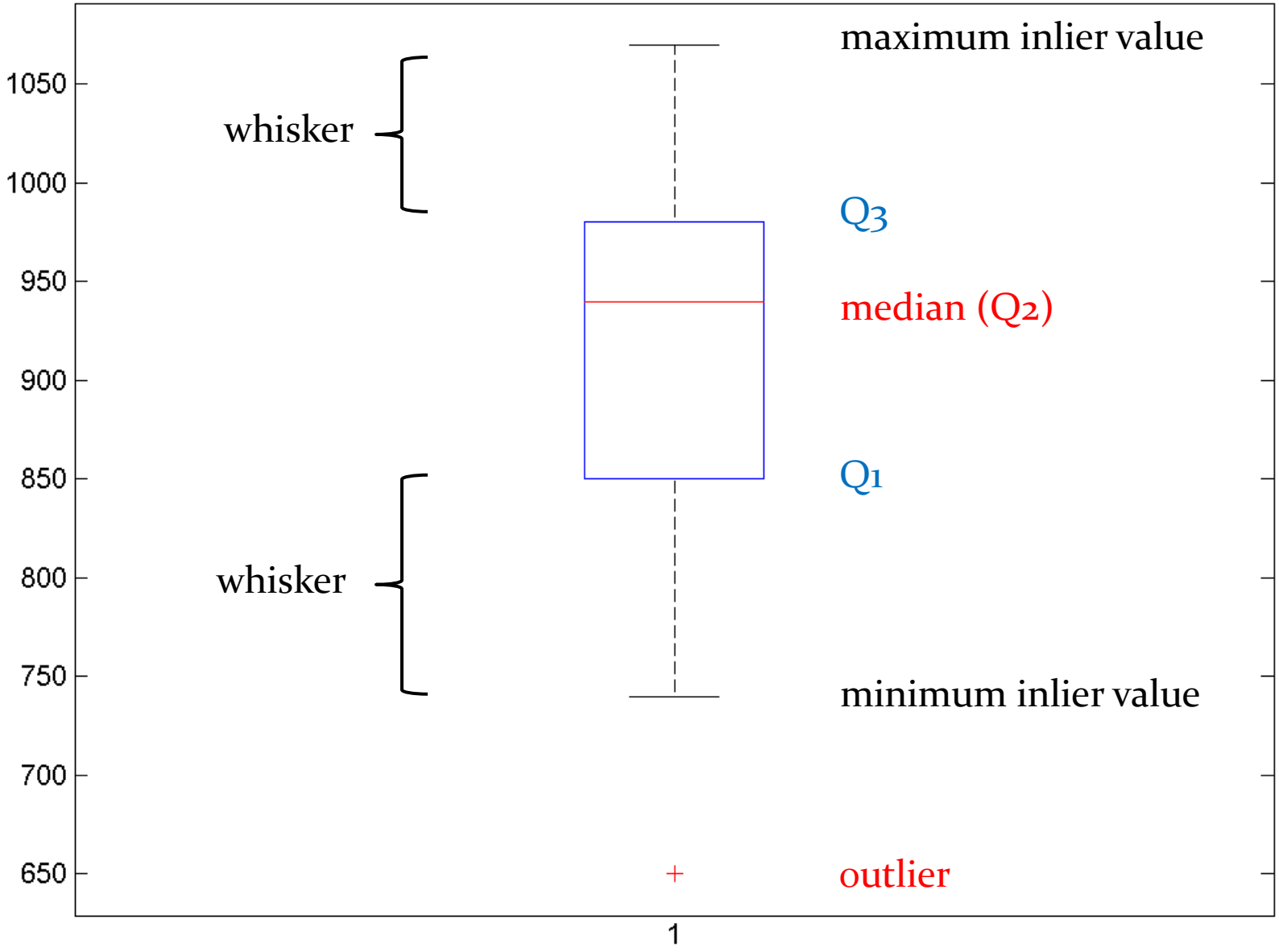
Interquartile range

- measurements in sorted order



Boxplots

- the box plot provides a graphical summary of the quartiles and extreme values of a group of measurements
- closely related to the *five-number summary*
  1. sample minimum
  2. first quartile
  3. median
  4. third quartile
  5. sample maximum


maximum inlier value
whisker
Q3
median (Q2)
Q1
whisker
minimum inlier value
outlier
1050
1000
950
900
850
800
750
700
650
1

Measuring the speed of light

- the first accurate measurements of the velocity of light in air were performed by Albert Michelson and Simon Newcomb in 1879-1882
  - based on methods developed by
    - Armand Hippolyte Louis Fizeau (1849)
      - 313,300 kilometres per second
    - Jean Bernard Léon Foucault (1862)
      - 298,000 kilometres per second

    - http://en.wikipedia.org/wiki/Fizeau-Foucault_apparatus
    - http://www.pas.rochester.edu/~pavone/particle-www/teachers/demonstrations/FoucaultDemonstration.htm

Michelson's data

Data set 12	Data set 13	Data set 14	Data set 15	Data set 16	Data set 17
850	960	880	890	890	883
740	940	880	810	840	816
900	960	880	810	780	778
1070	940	860	820	810	796
930	880	720	800	760	682
850	800	720	770	810	711
950	850	620	760	790	611
980	880	860	740	810	599
980	900	970	750	820	1051
880	840	950	760	850	781
1000	830	880	910	870	578
980	790	910	920	870	796
930	810	850	890	810	774
650	880	870	860	740	820
760	880	840	880	810	772
810	830	840	720	940	696
1000	800	850	840	950	573
1000	790	840	850	800	748
960	760	840	850	810	748
960	800	840	780	870	797
					851
					809
					723

measured speed of light in air = 299,000 + number in table

Michelson's data

- Microsoft Excel spreadsheet here
- MATLAB can read an Excel spreadsheet using **`xlsread`**

```
>> X = xlsread('michelson.xlsx');
>> boxplot(X)
```

Michelson's data

- compute some summary statistics of Michelson's data (watch out for NaNs!)
  - min
  - max
  - mean (compare to modern value of 299,705 km/s)
  - median
  - variance
  - standard deviation
- produce a box plot of Michelson's data

Percentiles

- the quartiles are specific examples of *percentiles*
- a percentile is the value below which a given percentage of observations fall
  - e.g., if 80% of test scores are below yours then your score would be in the 80^{th} percentile
- Q_1 is the 25^{th} percentile
- Q_2 is the 50^{th} percentile
- Q_3 is the 75^{th} percentile

Percentiles

- there is no single method for computing percentiles for discrete distributions (finite numbers of measurements)
  - e.g., what is the 33rd percentile of 10 measurements?
- MATLAB computes percentiles in a particular way that may be different than other commonly used software

```
>> y = prctile(x, 95)           % the 95th percentile of x

>> y = prctile(x, [25 50 75])   % the quartiles of x
```

Summary statistics are not enough

- Anscombe provides 4 data sets where the mean and variance of each set are (almost) the same
  - Microsoft Excel spreadsheet here
- compute some summary statistics for Anscombe's data
- draw a box plot for Anscombe's data
- plot Anscombe's data