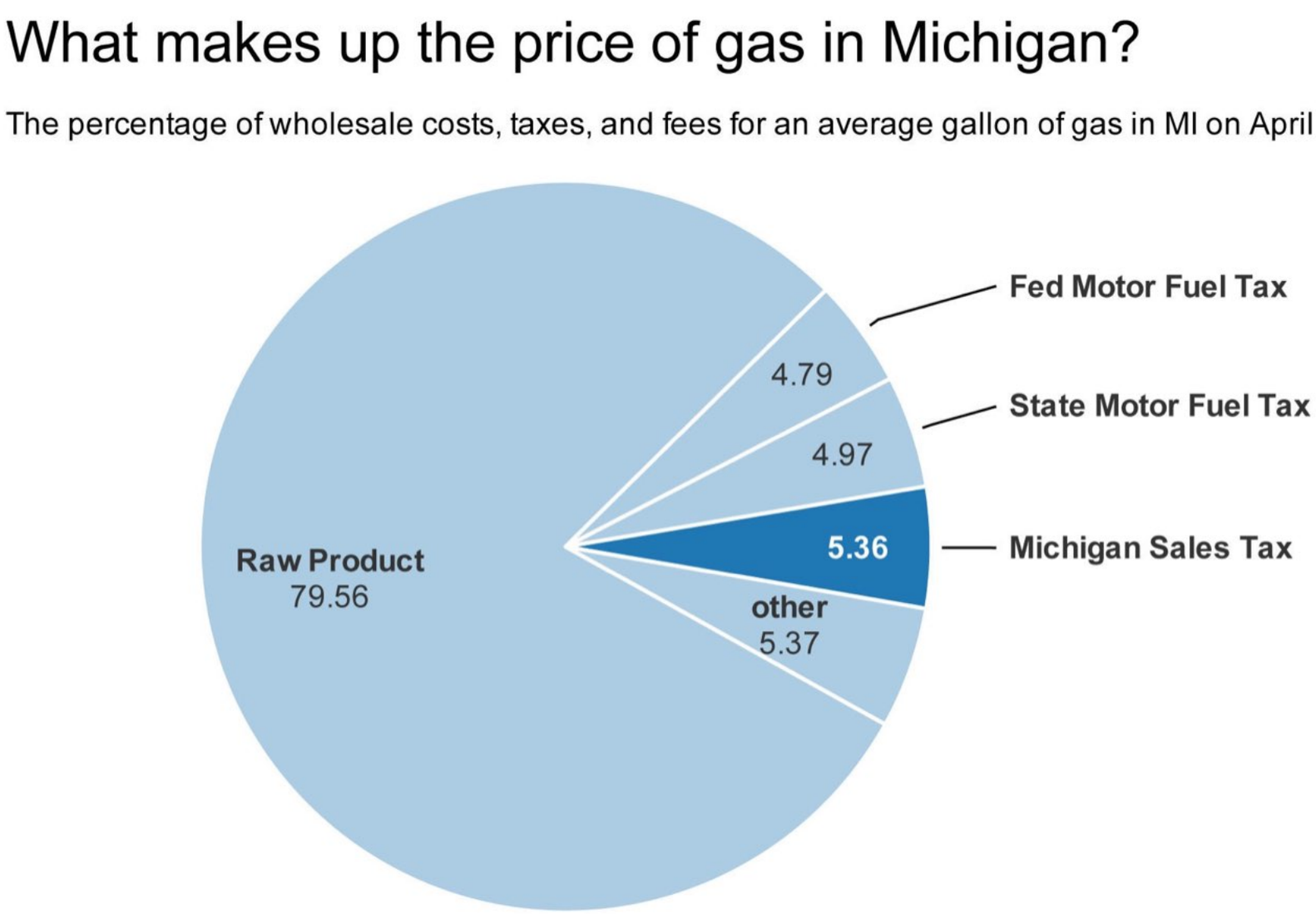
graphical visualisations

- one of the best ways to explore and understand

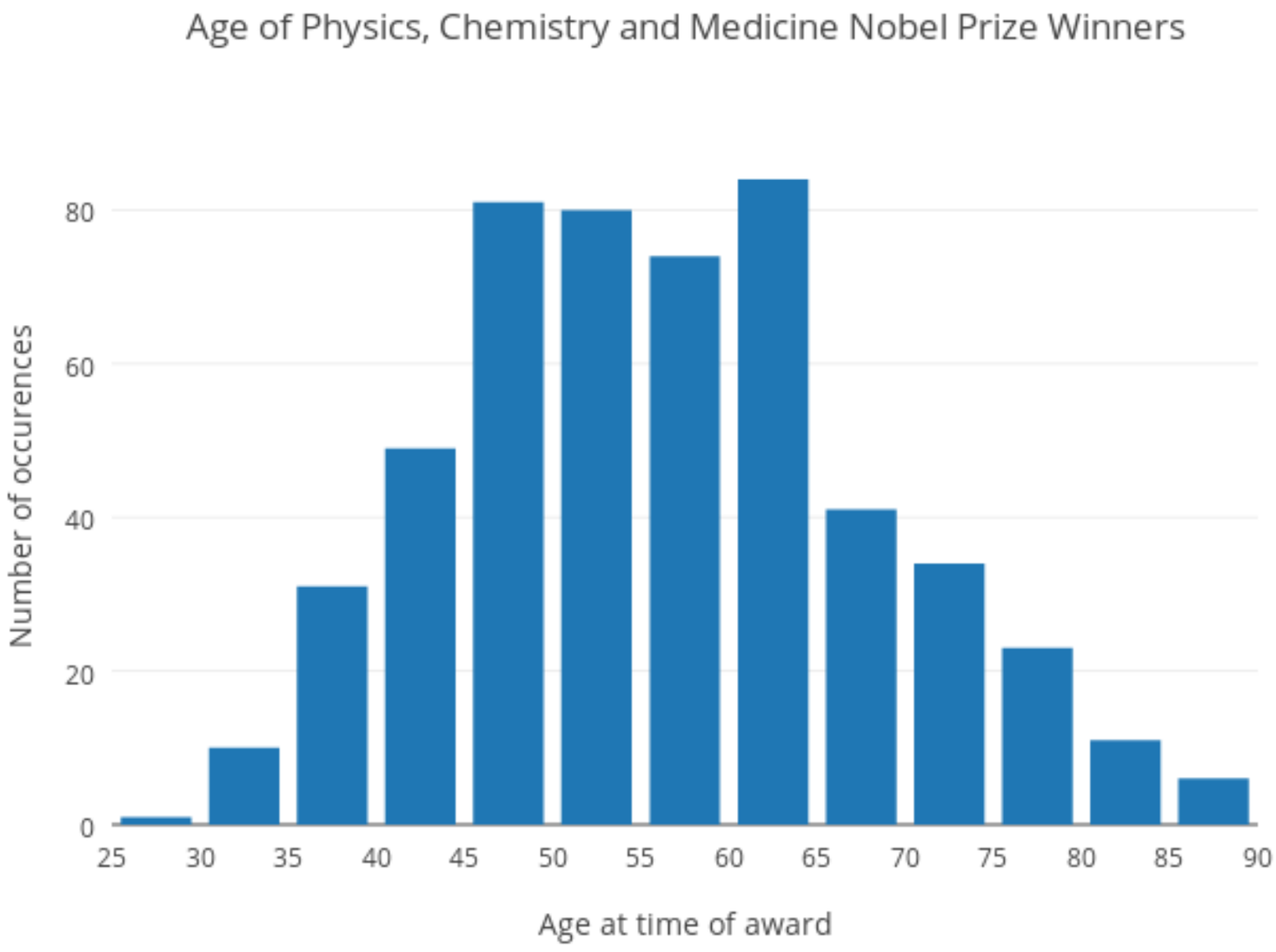
  large **data sets**

pie chart



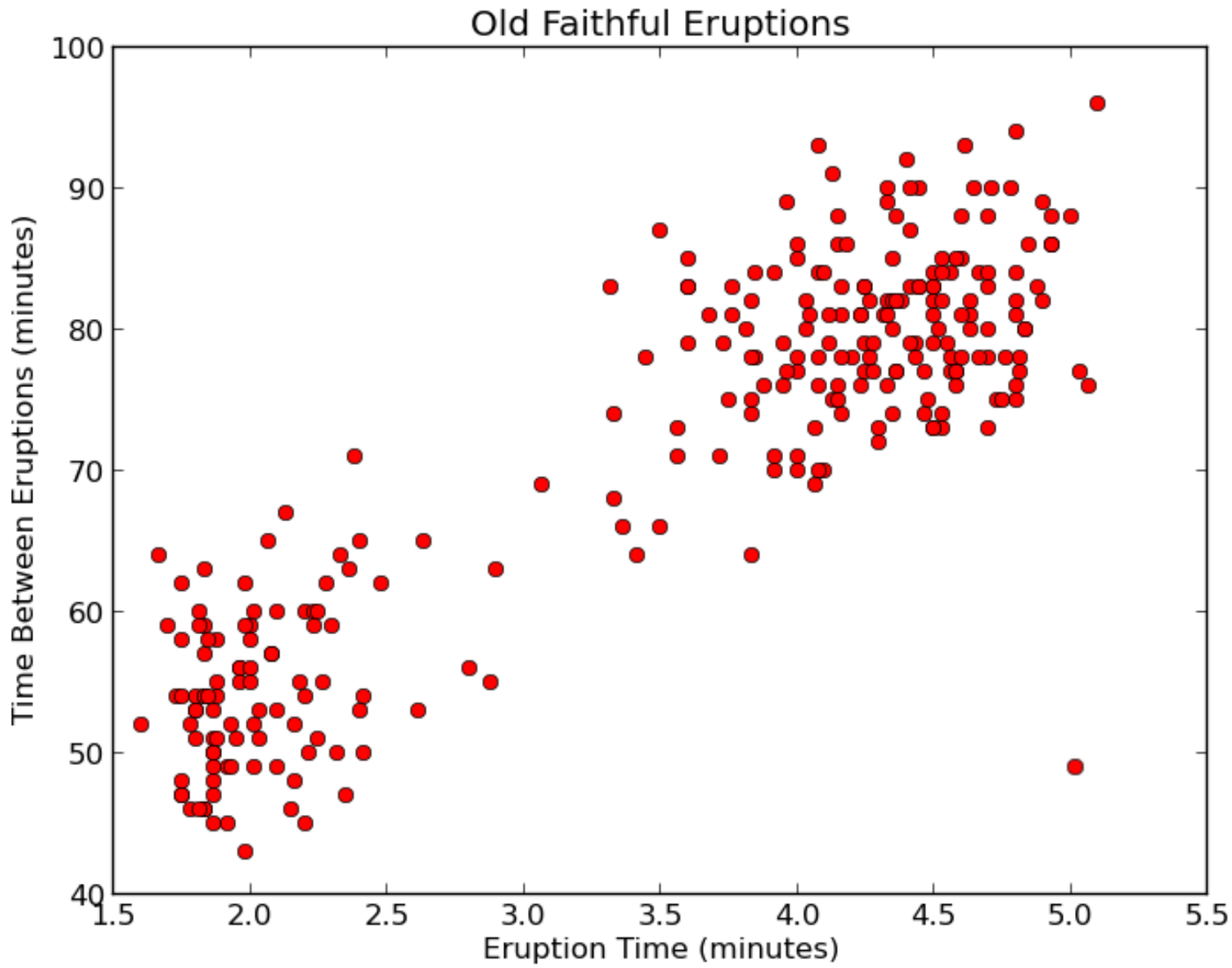
Note: Good for percentages divided into distinct classes. No more than 6.

histogram



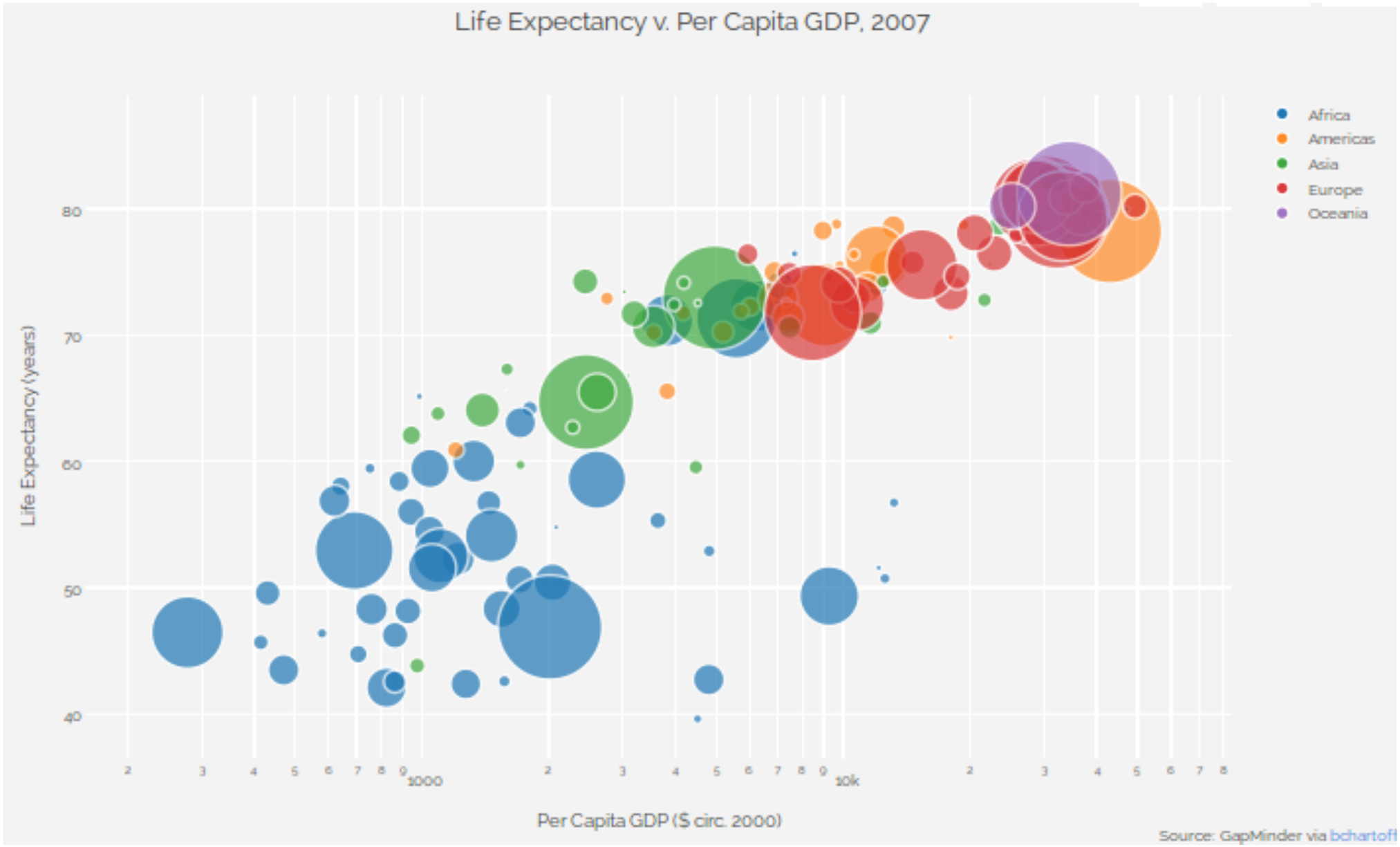
Note: Provides an overview of the distribution density of the samples from one variable.
Source: http://help.plot.ly/basic-statistics-mean-median-standard-deviation/

scatter plot



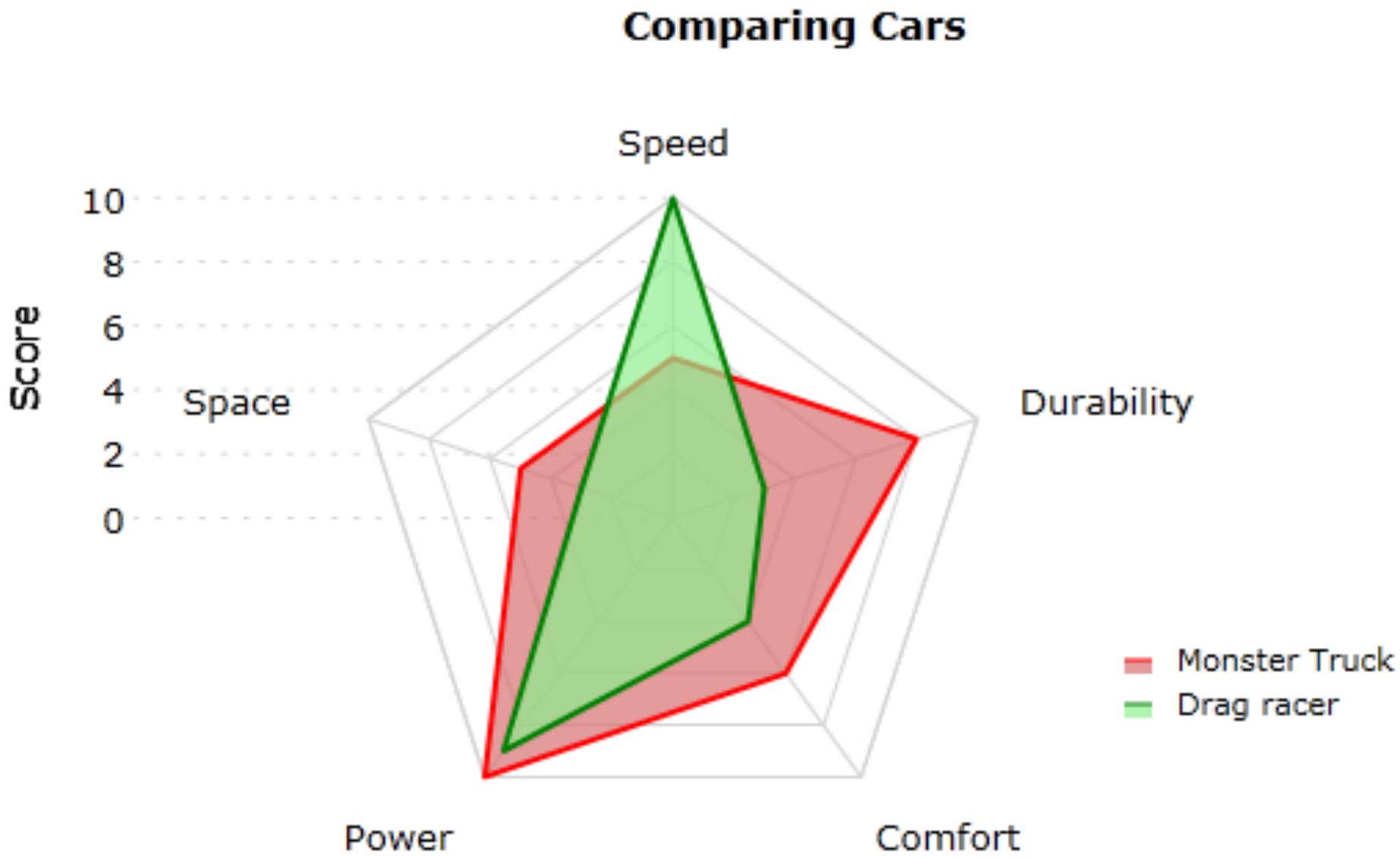
Note: Good for outlier and dependency assessment.

bubble chart



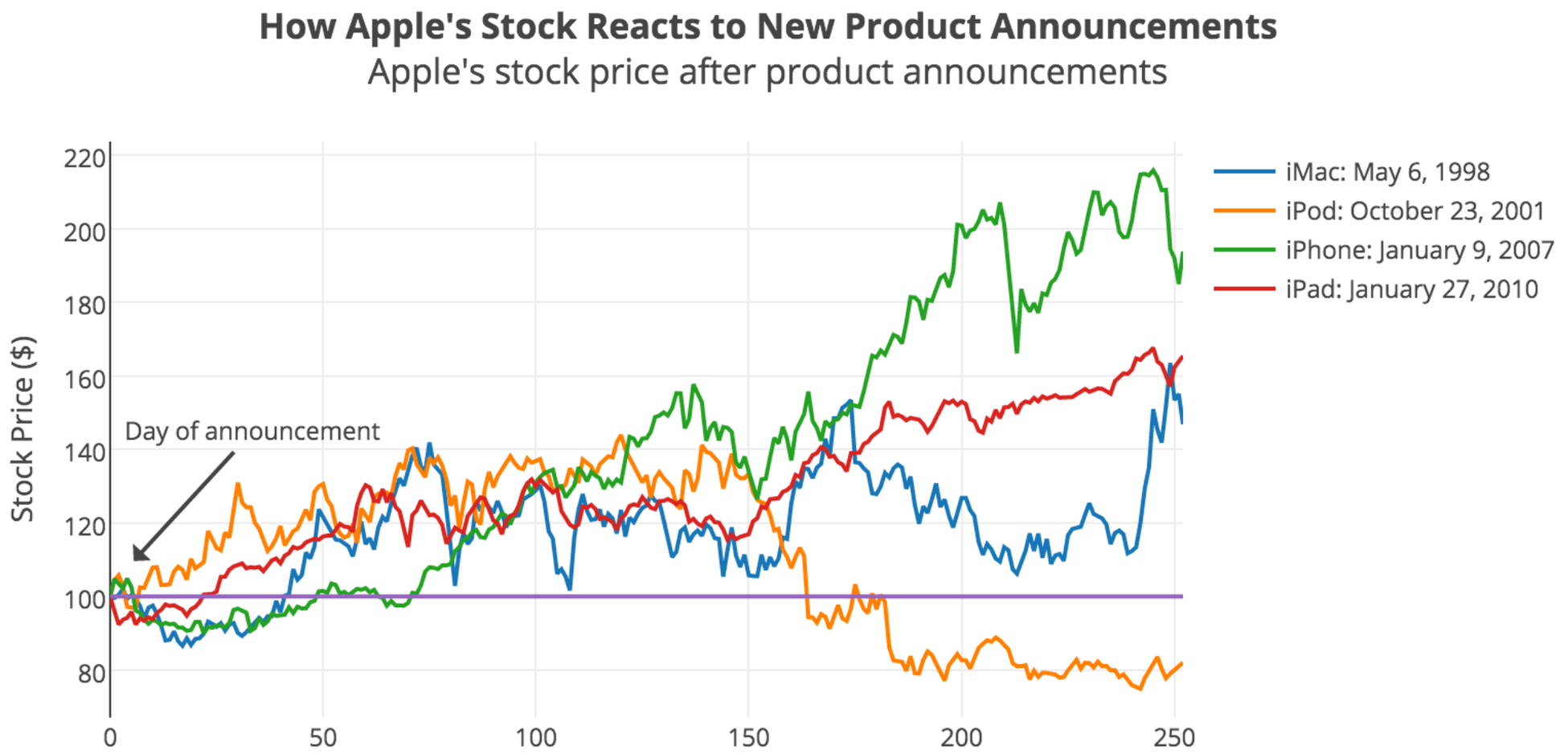
Note: Good to visualise data with more than 2 dimensions.
Source: https://plot.ly

radar chart



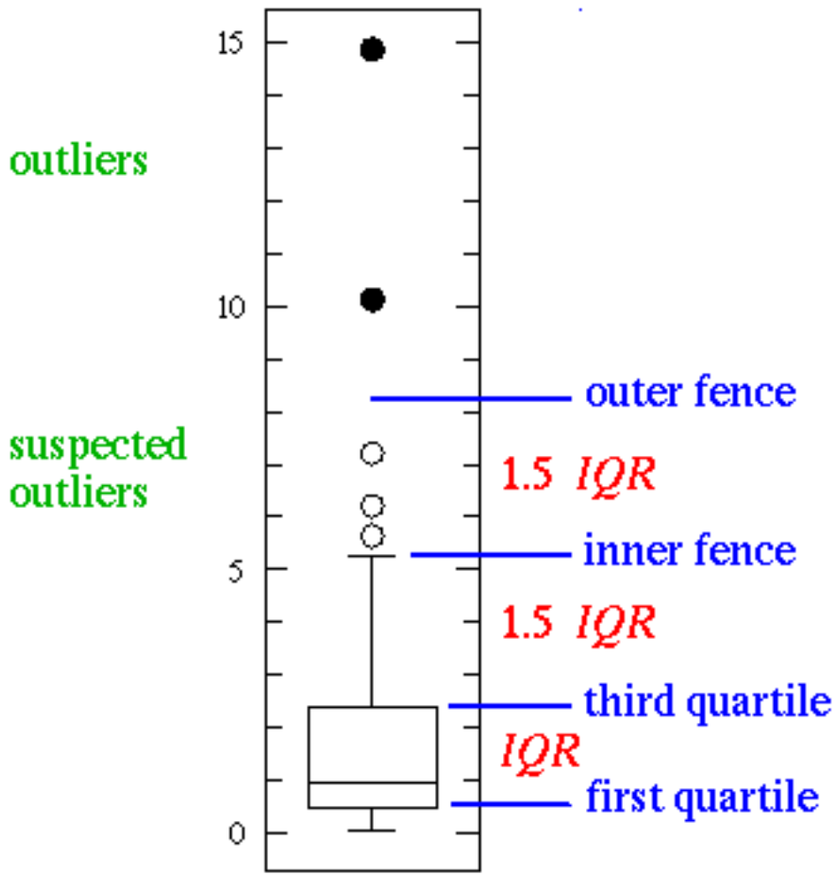
Note: Good for comparisons between samples with more than 2 dimensions.

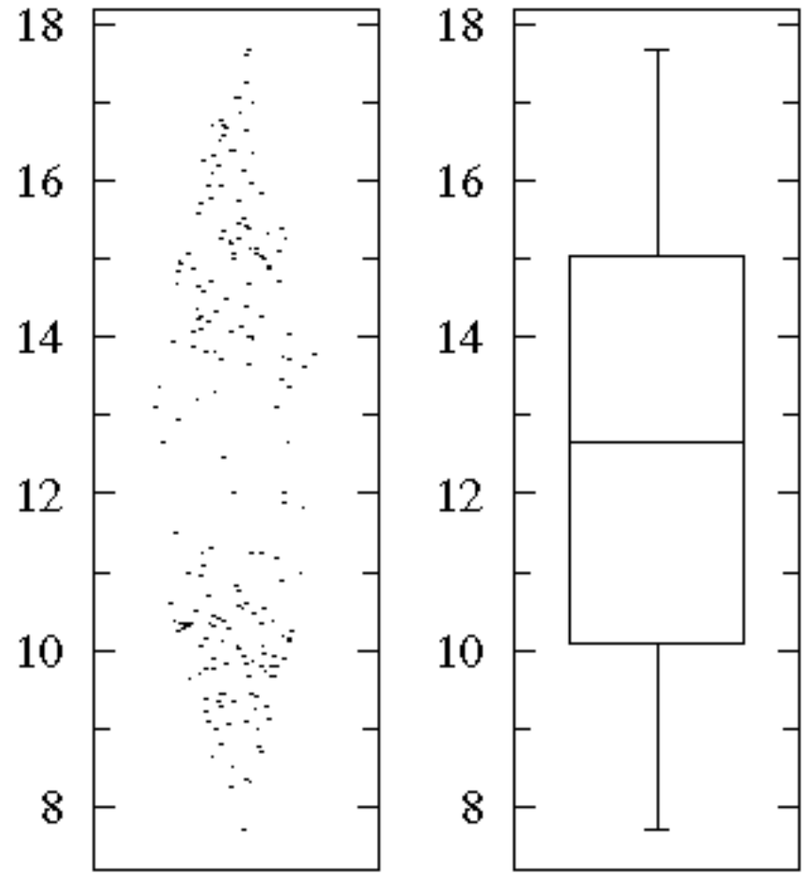
line chart



Note: Good to visualise and compare time series. No more than 7.
Source: https://plot.ly/609/~Dreamshot/

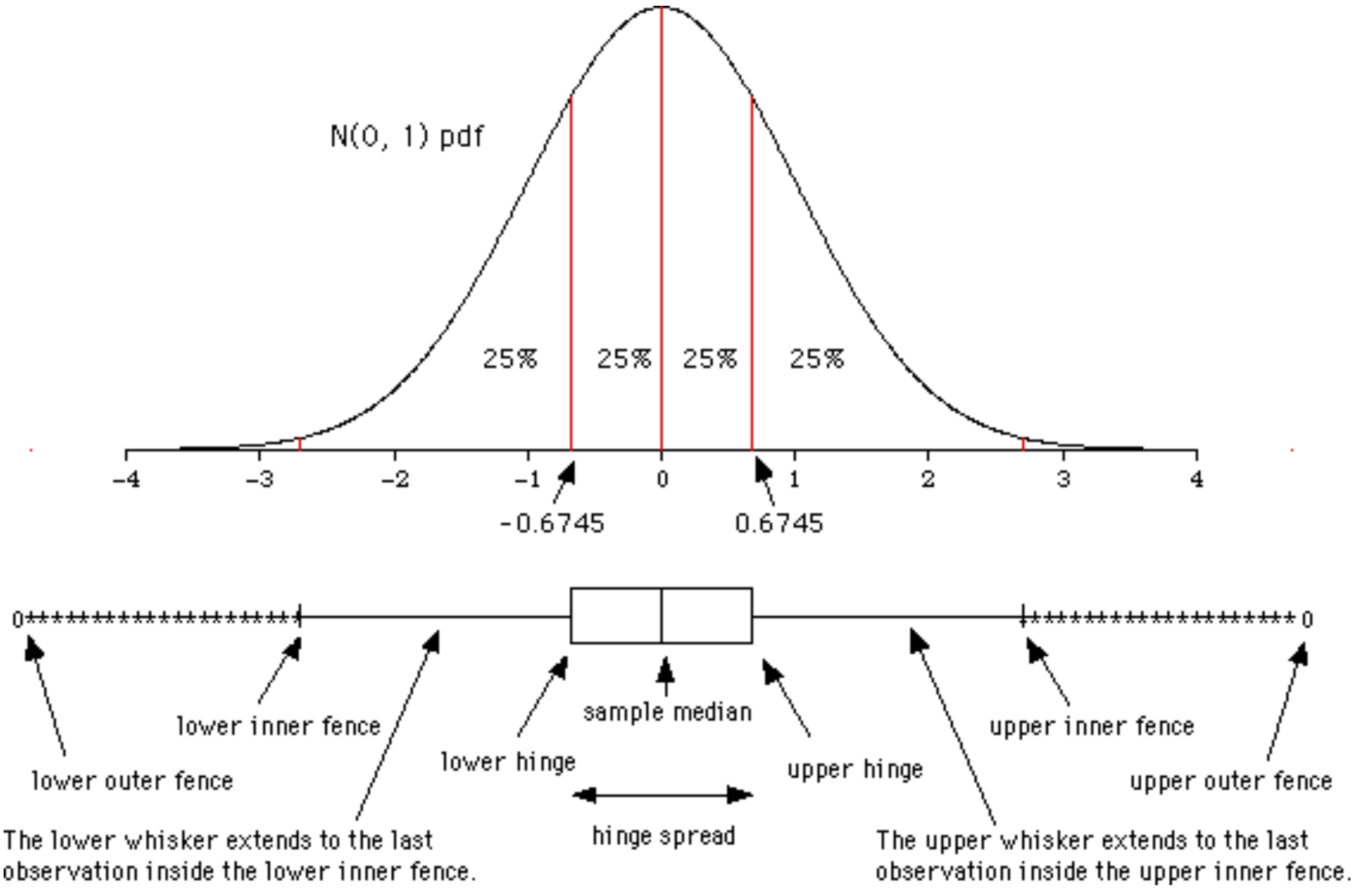
box plot





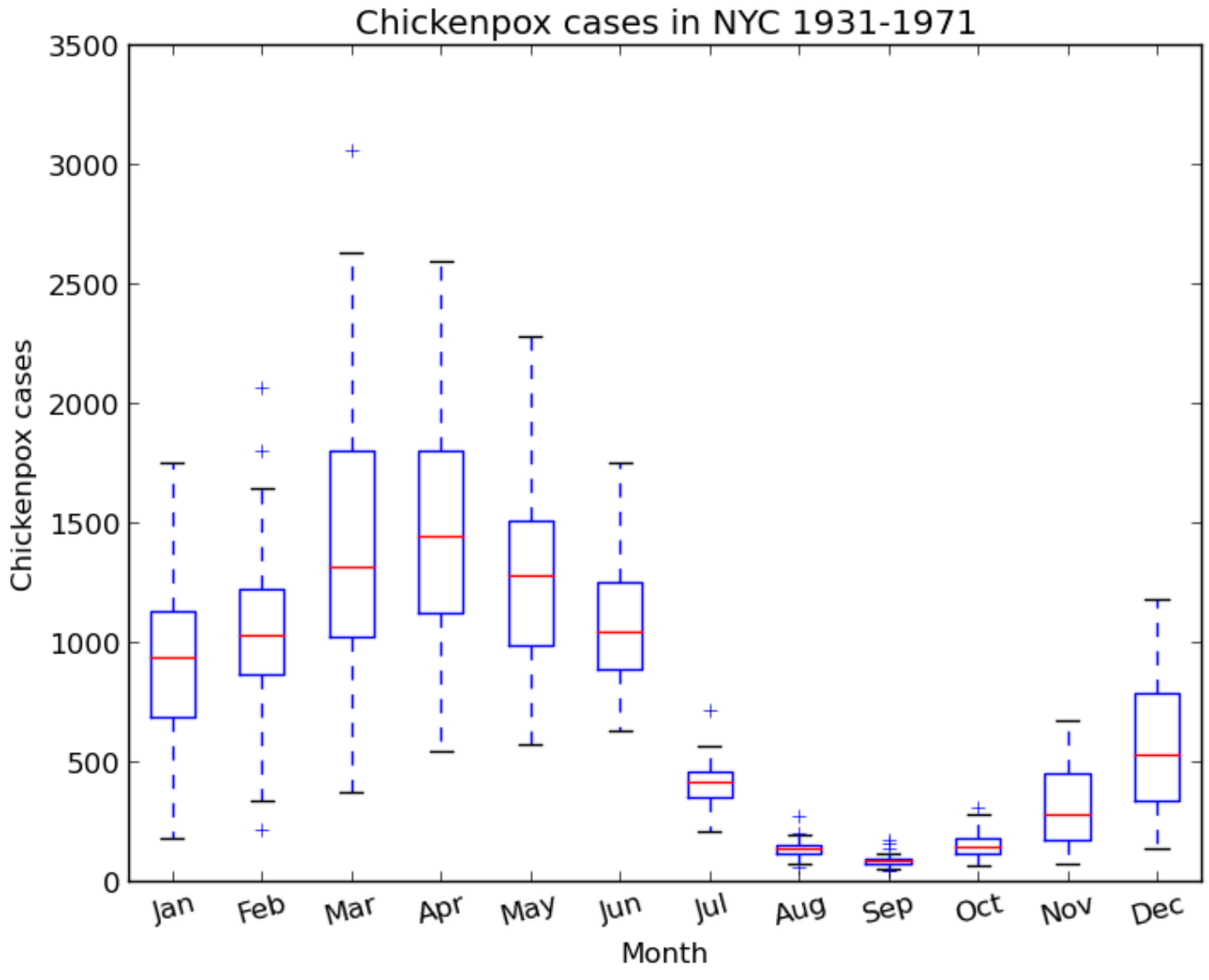
Source: http://www.physics.csbsju.edu/stats/box2.html

box plot



Source: http://ms.mcmaster.ca/peter/s2ma3/Images/norm_quart.gif

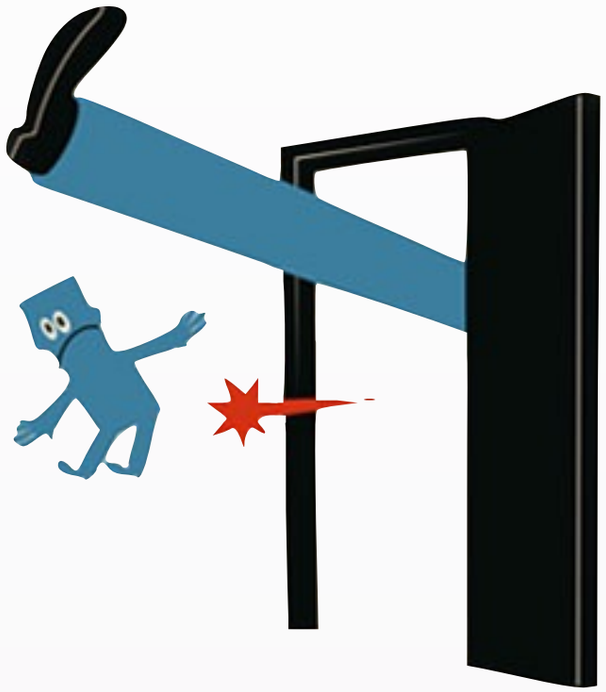
box plot



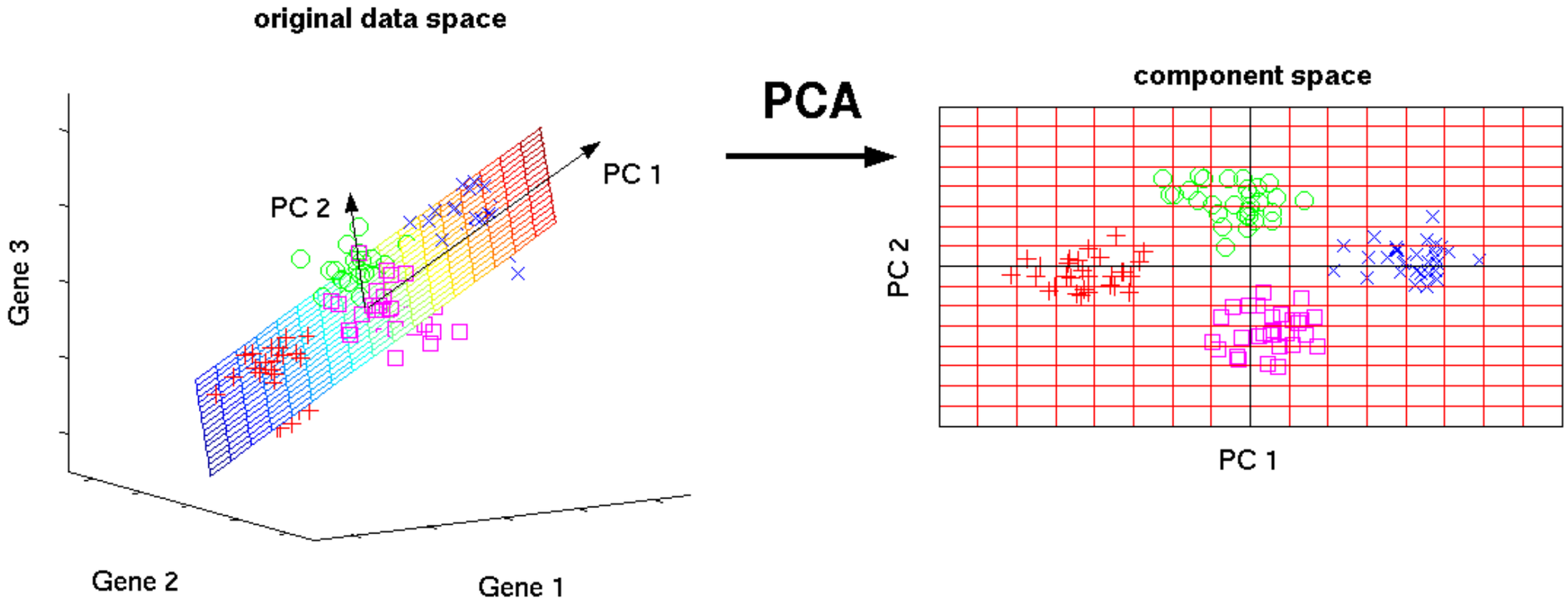
Note: Good for assessing dispersion and skewedness of samples.

dataset reduction

- the quality of the data is essential
- main sources of errors
  - systematic
  - outliers
- scatter plot, box plot
  - discard, correct or accept the suspicious samples



principal component analysis



Note: Used when the data are largely redundant.

hypothesis testing

- can we reject the H_0 hypothesis?
- needs to be taken into account AFTER the experiment has been planned
- plenty of tests around
- features
  - Input
  - Null hypothesis
  - Calculations
  - Criterion

the 3 fundamental questions

- parametric or non-parametric?
- 2 or more treatments?
- dependent or independent samples?

parametric vs. non-parametric

- parametric tests are more powerful because they use more information > require fewer data points
- scale types:
  - nominal or ordinal: non-parametric
  - interval or ratio: parametric
- distribution of your data:
  - close to normal: parametric
  - far from normal: non-parametric

Note: Skew and Kurtosis coefficients are about 1 for normal data.

joke: normality

Q How many statisticians does it take to change a lightbulb?

A This should not be determined using a non-parametric procedure, since statisticians are not normal.

Note: This is just a joke.

dependent vs. independent

subjects	treat. 1	treat. 2
1	x	x
2	x	x
3	x	x
4	x	x
5	x	x
6	x	x

subjects	treat. 1	treat. 2
1	x	
2		x
3		x
4	x	
5	x	
6		x

- dependent: if subjects are tested with the different treatments.

p-value

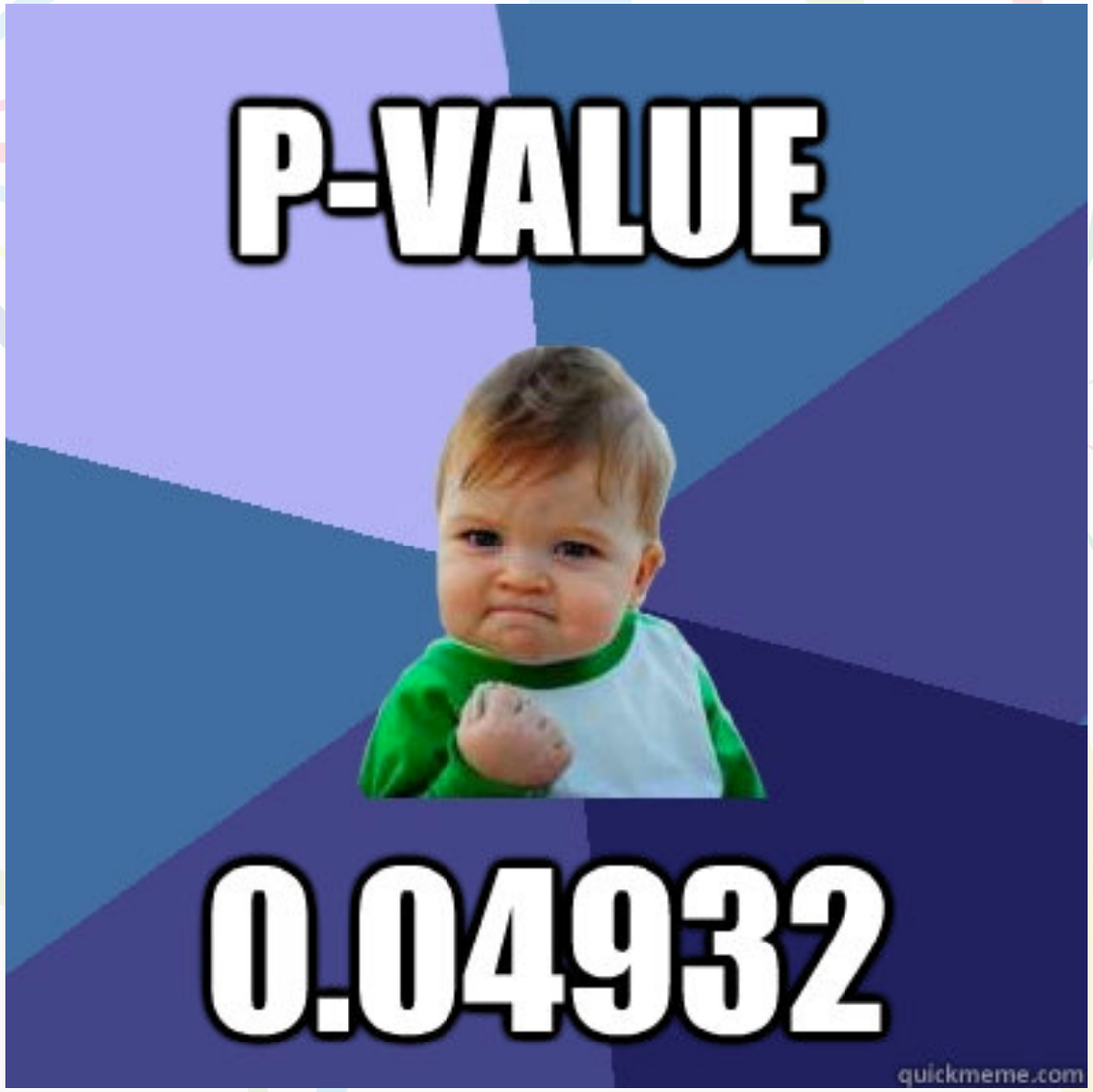
- indicates the amount of evidence to support the null hypothesis
- probability of mistakenly reject H_0
  - $p <= 0.05$: strong evidence against H_0
  - $p > 0.05$: weak evidence against H_0
- the lower the better
- every test provides a p-value

Note: The threshold may vary according to the field and community.

joke: p-value

Note: This is just a joke.

joke: p-value



Note: This is just a joke.

binomial test

Table 10.4 Binomial test

Item	Description
Input	Number of events counted for two different kind of events ($event_1$ and $event_2$)
H_0	$P(\text{event 1}) = P(\text{event 2})$
Calculations	Calculate $p = \frac{1}{2^N}\sum_{i=0}^{n}\binom{N}{i}$ where N is the total number of events, and n is the number of the most rare event
Criterion	Two sided ($H_1 : P(\text{event 1}) \neq P(\text{event 2})$): reject H_0 if $p < \alpha/2$ One sided ($H_1 : P(\text{event 1}) < P(\text{event 2})$): reject H_0 if $p < \alpha$ and event 1 is the most rare event in the sample

Source: Experimentation in Software Engineering, Springer, 2012

t-Test

Table 10.5 t-test

Item	Description
Input	Two independent samples: $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_m$
H_0	$\mu_x = \mu_y$, i.e. the expected mean values are the same
Calculations	Calculate $t_0 = \dfrac{\bar{x} - \bar{y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$ where $S_p = \sqrt{\dfrac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$ and, S_x^2 and S_y^2 are the individual sample variances
Criterion	Two sided ($H_1 : \mu_x \neq \mu_y$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha/2,n+m-2}$. Here, $t_{\alpha,f}$ is the upper α percentage point of the t distribution with f degrees of freedom, which is equal to $n+m-2$. The distribution is tabulated, for example, in Table B.1 and by Montgomery [125], and Marascuilo and Serlin [119]
	One sided ($H_1 : \mu_x > \mu_y$): reject H_0 if $t_0 > t_{\alpha,n+m-2}$

Source: Experimentation in Software Engineering, Springer, 2012

t-Test in MATLAB

- The defect density in two projects (A and B) has been compared.
  - A = {3.42, 2.71, 2.84, 1.85, 3.22, 3.48, 2.68, 4.30, 2.49, 1.59}
  - B = {3.44, 4.97, 4.76, 4.96, 4.10, 3.05, 4.09, 3.69, 4.21, 4.40, 3.49}

```
> A = [3.42 2.71 2.84 1.85 3.22 3.48 2.68 4.30 2.49 1.59];
> B = [3.44 4.97 4.76 4.96 4.10 3.05 4.09 3.69 4.21 4.40 3.49];
> [H, P] = ttest2(A,B)
    H = 1
    P = 8.4418e-04
```

F-test

Table 10.7 F-test

Item	Description
Input	Two independent samples: $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_m$
H_0	$\sigma_x^2 = \sigma_y^2$, i.e. the variances are equal
Calculations	Calculate $F_0 = \frac{\max(S_x^2, S_y^2)}{\min(S_x^2, S_y^2)}$, where S_x^2 and S_y^2 are the individual sample variances
Criterion	Two sided ($H_1 : \sigma_x^2 \neq \sigma_y^2$): reject H_0 if $F_0 > F_{\alpha/2, n_{max}-1, n_{min}-1}$, where n_{max} is the number of scores in the sample with maximum sample variance and n_{min} is the number of scores in the sample with minimum sample variance. $F_{\alpha/2, f_1, f_2}$ is the upper α percentage point of the F distribution with f_1 and f_2 degrees of freedom, which is tabulated, for example, in Table B.5 and by Montgomery [125], and Marascuilo and Serlin [119]
	One sided ($H_1 : \sigma_x^2 > \sigma_y^2$): reject H_0 if $F_0 > F_{\alpha, n_{max}-1, n_{min}-1}$, and $S_x^2 > S_y^2$

Source: Experimentation in Software Engineering, Springer, 2012

F-test in MATLAB

- The defect density in two projects (A and B) has been compared.
  - A = {3.42, 2.71, 2.84, 1.85, 3.22, 3.48, 2.68, 4.30, 2.49, 1.59}
  - B = {3.44, 4.97, 4.76, 4.96, 4.10, 3.05, 4.09, 3.69, 4.21, 4.40, 3.49}

```
> A = [3.42 2.71 2.84 1.85 3.22 3.48 2.68 4.30 2.49 1.59];
> B = [3.44 4.97 4.76 4.96 4.10 3.05 4.09 3.69 4.21 4.40 3.49];
> [H, P] = vartest2(A,B)
    H = 0
    P = 0.4845
```

joke: p-value

Note: This is just a joke.

Mann-Whitney test

Table 10.6 Mann-Whitney

Item	Description
Input	Two independent samples: $x_1, x_2, \ldots x_n$ and $y_1, y_2, \ldots y_m$
H_0	The two samples come from the same distribution
Calculations	Rank all samples and calculate $U = N_A N_B + \frac{N_A(N_A+1)}{2} - T$ and $U' = N_A N_B - U$, where $N_A = \min(n,m)$, $N_B = \max(n,m)$, and T is the sum of the ranks of the smallest sample
Criterion	Tables providing criterion for rejection of the null hypothesis based on the calculations are provided, for example, in Table B.3 and by Marascuilo and Serlin [119]
	Reject H_0 if $\min(U, U')$ is less than or equal to the value in Table B.3

Source: Experimentation in Software Engineering, Springer, 2012

Mann-Whitney in MATLAB

- The defect density in two projects (A and B) has been compared.
  - A = {3.42, 2.71, 2.84, 1.85, 3.22, 3.48, 2.68, 4.30, 2.49, 1.59}
  - B = {3.44, 4.97, 4.76, 4.96, 4.10, 3.05, 4.09, 3.69, 4.21, 4.40, 3.49}

```
> A = [3.42 2.71 2.84 1.85 3.22 3.48 2.68 4.30 2.49 1.59];
> B = [3.44 4.97 4.76 4.96 4.10 3.05 4.09 3.69 4.21 4.40 3.49];
> [P, H] = ranksum(A,B)
    P = 0.0022
    H = 1
```

paired t-Test

Table 10.8 Paired t-test

Item	Description
Input	Paired samples: $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$
H_0	$\mu_d = 0$, where $d_i = x_i - y_i$, i.e. the expected mean of the differences is 0
Calculations	Calculate $t_0 = \dfrac{\bar{d}}{S_d/(\sqrt{n})}$, where $S_d = \sqrt{\dfrac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$
Criterion	Two sided ($H_1 : \mu_d \neq 0$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha/2,n-1}$. Here, $t_{\alpha,f}$ is the upper α percentage point of the t distribution with f degrees of freedom. The distribution is tabulated, for example, in Table B.1 and by Montgomery [125], and Marascuilo and Serlin [119] One sided ($H_1 : \mu_d > 0$): reject H_0 if $\lvert t_0 \rvert > t_{\alpha,n-1}$

Source: Experimentation in Software Engineering, Springer, 2012

paired t-Test in MATLAB

- 10 programmers have independently developed 2 different programs, and measured the effort.
  - P = {105, 137, 124, 111, 151, 150, 168, 159, 104, 102}
  - Q = {86.1, 115, 175, 94.9, 174, 120, 153, 178, 71.3, 110}

```
> P = [105 137 124 111 151 150 168 159 104 102];
> Q = [86.1 115 175 94.9 174 120 153 178 71.3 110];
> [H, p] = ttest(P,Q)
    p = 0.7059
    H = 0
```

Wilcoxon test

Table 10.10 Wilcoxon

Item	Description
Input	Paired samples: $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$
H_0	If all differences $(d_i = x_i - y_i)$ are ranked $(1, 2, 3 \ldots)$ without considering the sign, then the sum of the ranks of the positive differences equals the sum of the ranks of the negative differences
Calculations	Calculate T^+ as the sum of the ranks of the positive d_i:s and T^- as the sum of the ranks of the negative d_i:s
Criterion	Tables that can be used to determine if H_0 can be rejected based on T^+, T^- and the number of pairs, n, are available. See for example Table B.4 or Siegel and Castellan [157], and Marascuilo and Serlin [119] Reject H_0 if $\min(T^+, T^-)$ is less than or equal to the value in Table B.4

Source: Experimentation in Software Engineering, Springer, 2012

Wilcoxon in MATLAB

- 10 programmers have independently developed 2 different programs, and measured the effort.
  - P = {105, 137, 124, 111, 151, 150, 168, 159, 104, 102}
  - Q = {86.1, 115, 175, 94.9, 174, 120, 153, 178, 71.3, 110}

```
> P = [105 137 124 111 151 150 168 159 104 102];
> Q = [86.1 115 175 94.9 174 120 153 178 71.3 110];
> [p, H] = signrank(P,Q)
    p = 0.6953
    H = 0
```

sign test

Table 10.11 Sign test

Item	Description
Input	Paired samples: $(x_1, y_1), (x_2, y_2) \ldots (x_N, y_N)$
H_0	$P(+) = P(-)$, where $+$ and $-$ represent the two events that $x_i > y_i$ and $x_i < y_i$
Calculations	Represent every positive differences $(d_i = x_i - y_i)$ by a $+$ and every negative difference by a $-$. Calculate $p = \frac{1}{2^N} \sum_{i=0}^{n} \binom{N}{i}$, where N is the total number of signs, and n is number of signs of the most rare signs
Criterion	Two sided $(H_1 : P(+) \neq P(-))$: reject H_0 if $p < \alpha/2$ One sided $(H_1 : P(+) < P(-))$: reject H_0 if $p < \alpha$ and the $+$ event is the most rare event in the sample

Source: Experimentation in Software Engineering, Springer, 2012

signed test in MATLAB

- 10 programmers have independently developed 2 different programs, and measured the effort.
  - P = {105, 137, 124, 111, 151, 150, 168, 159, 104, 102}
  - Q = {86.1, 115, 175, 94.9, 174, 120, 153, 178, 71.3, 110}

```
> P = [105 137 124 111 151 150 168 159 104 102];
> Q = [86.1 115 175 94.9 174 120 153 178 71.3 110];
> [p, H] = signtest(P,Q)
    p = 0.7539
    H = 0
```