

Univariate visualization

- interval and ratio variables

A bit of statistics

Categorical variables

Counts (frequencies)

- Relative (percentage)
- Absolute (number as such)

Cardinal (interval and ratio) variables

Too many different values

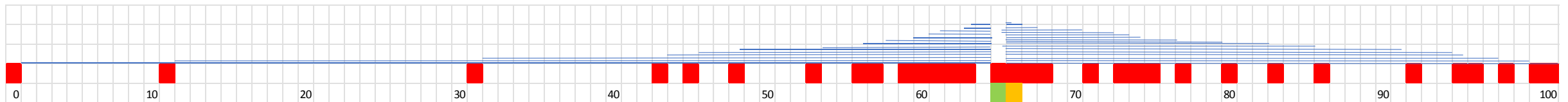
- Problems with showing just counts

What can be good quantity to show

- Central value : Average, Median
- Variation: standard deviation
- Other descriptive statistics: minimum, maximum

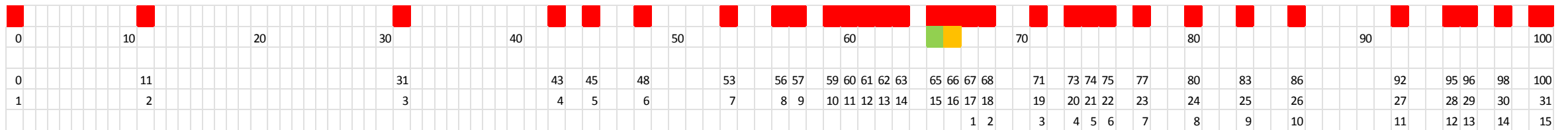
Average

- The „geographical centre“ of data
- sum of distances to lower values = sum of distances to higher values
- movie rating 0-100 points
- Red dots: users evaluation
- Green dot: average value



Median

- Value standing in the middle
- Half of data are lower, half higher



Relation between mean and median

Can be the same

- symmetric distribution
- Normal
- U-shape
- Low variance

Can be different

- Skewed distribution
- E.g. income
- Germany: household wage in 2022: 42,192 € 45,457 €
- Why is it different?

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$$

X - The Value in the data distribution

μ - The population Mean

N - Total Number of Observations

Standard deviation

- How far are data from average
- Average speed 50 KMPH (30 miles)
 - Because the car went whole time exactly 50 (sd=0)
 - Because car went half of journey 30 and other half 70 (sd=20)
 - Because car spent one hour in traffic jam and half hour went 150 (sd=70)
- higher deviation means higher variance

Other descriptive stats



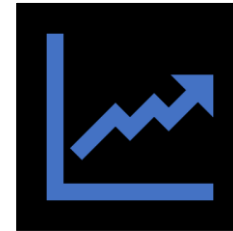
Minimum, maximum (+ range)

World records, temperatures,



Quartiles

Way how to group values

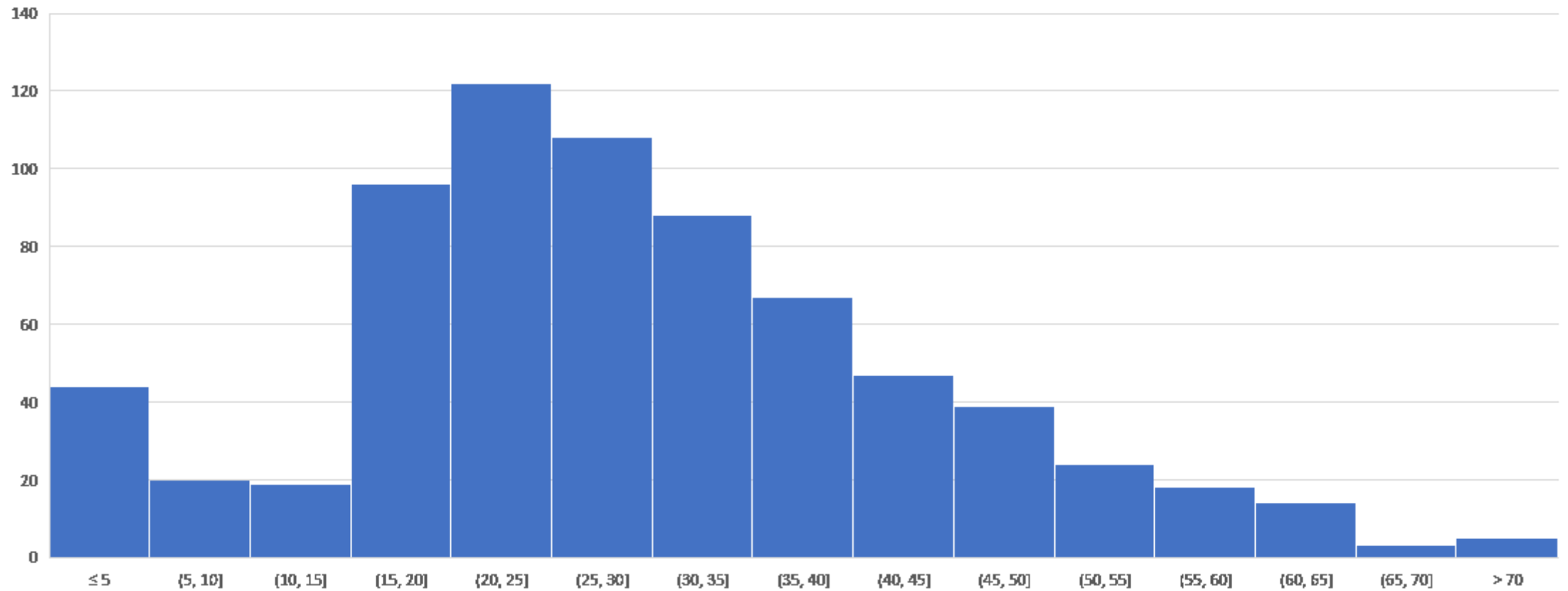


Gini and variation coef.

Measures of concentration

Histogram

Age of Titanic passengers

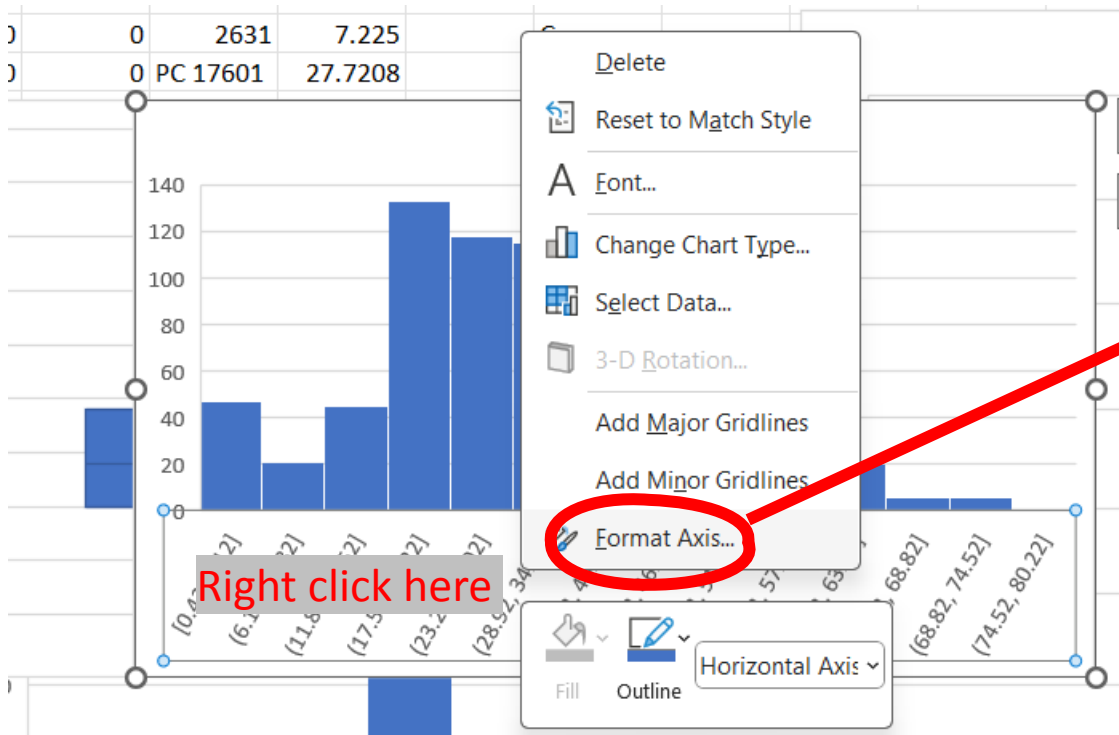


How to make histogram

- No need for any computation!
- Select column and click on histogram

The screenshot shows the Microsoft Excel interface. The 'Insert' tab is selected and highlighted with a red circle. In the 'Charts' group, the 'Histogram' icon is also highlighted with a red circle. A tooltip for the 'Histogram' chart type is displayed, showing a histogram icon and the text: 'Histogram', 'Use this chart type to:', and '• Show the distribution of the data grouped into bins.' Below the tooltip is a link for 'More Statistical Charts...'. The background shows a spreadsheet with the 'Age' column selected.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
2	1	1	Cumings, Mrs. James	female	38	1	0	PC 17599	71.2834
10	1	2	Nasser, Mr. Nasser	female	14	1	0	237736	31.2708
20	1	3	Masselmani, Mrs. Ines	female		0	0	2649	51.8642
27	0	3	Emir, Mr. Abdol	male		0	0	2631	53.1
31	0	1	Uruchurtu, Mr. Raulo	male	40	0	0	PC 17601	51.6642



Format Axis

Axis Options ▼ Text Options

Axis Options

Bins

- By Category
- Automatic
- Bin width 1
- Number of bins 14
- Overflow bin 73.0 Aut
- Underflow bin -14.0 Aut

Tick Marks

Major type None

Minor type None

Number

Category General

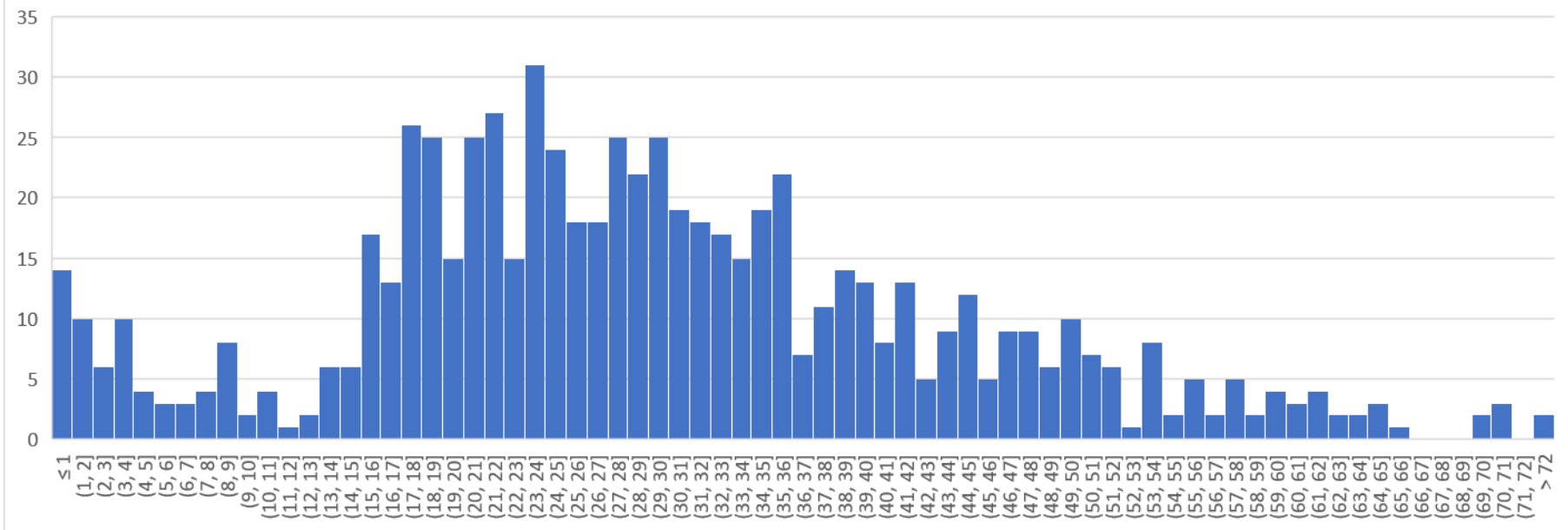
Format Code všeobecný

Overflow bin 72.0

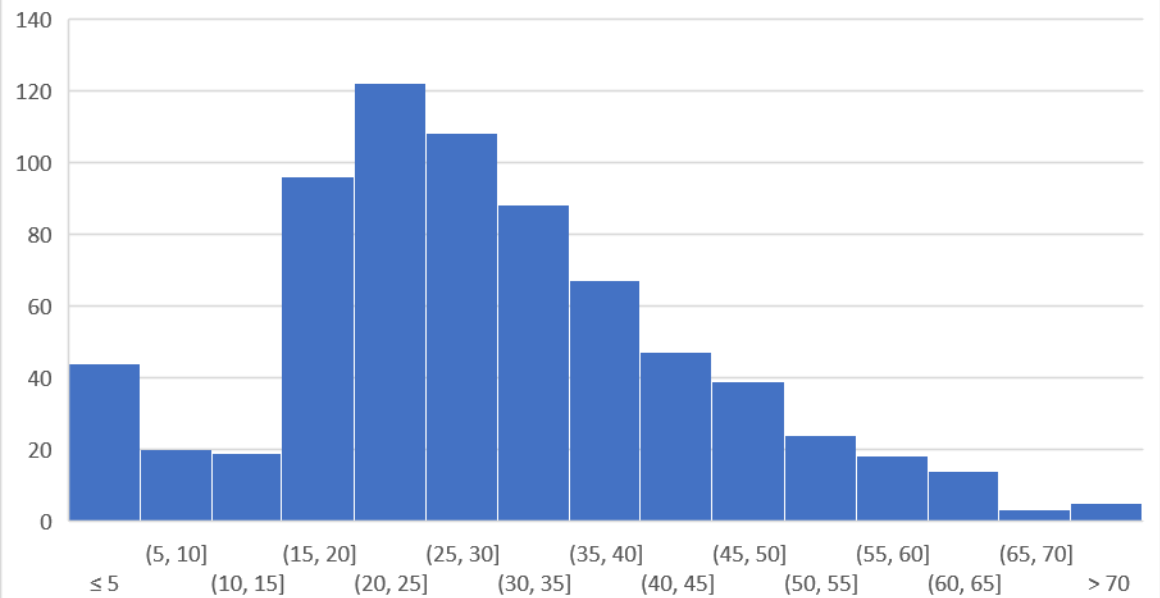
Underflow bin 1.0

Keep it in line with width

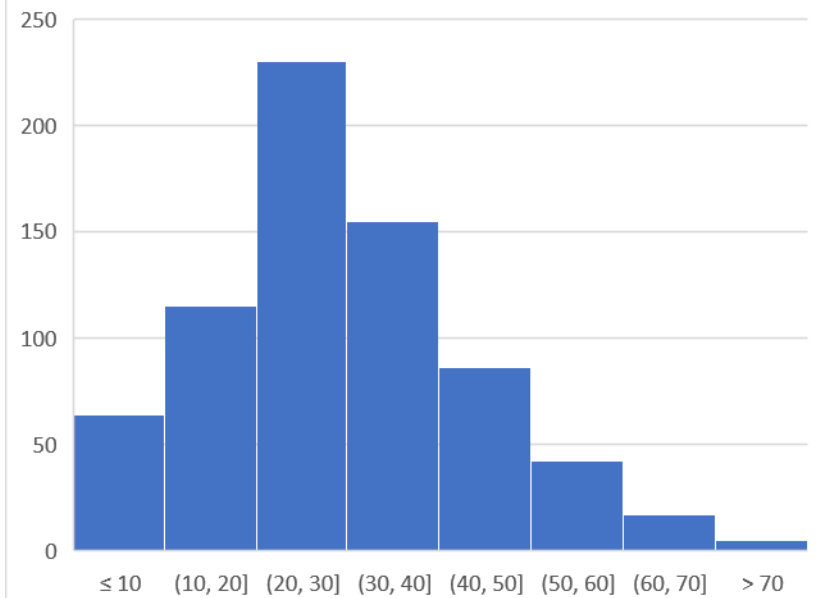
Bin width = 1



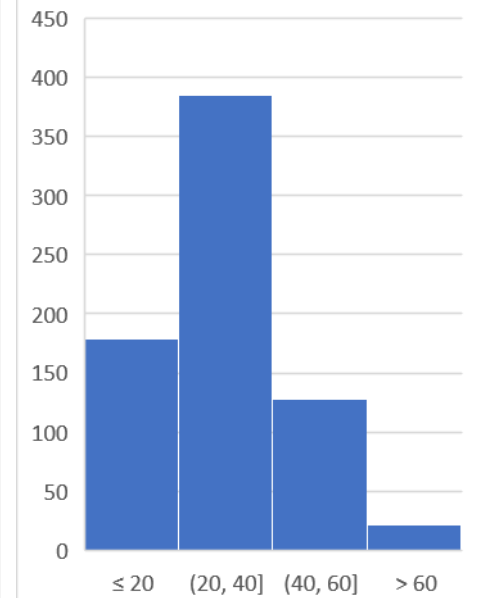
Bin width = 5



Bin width = 10

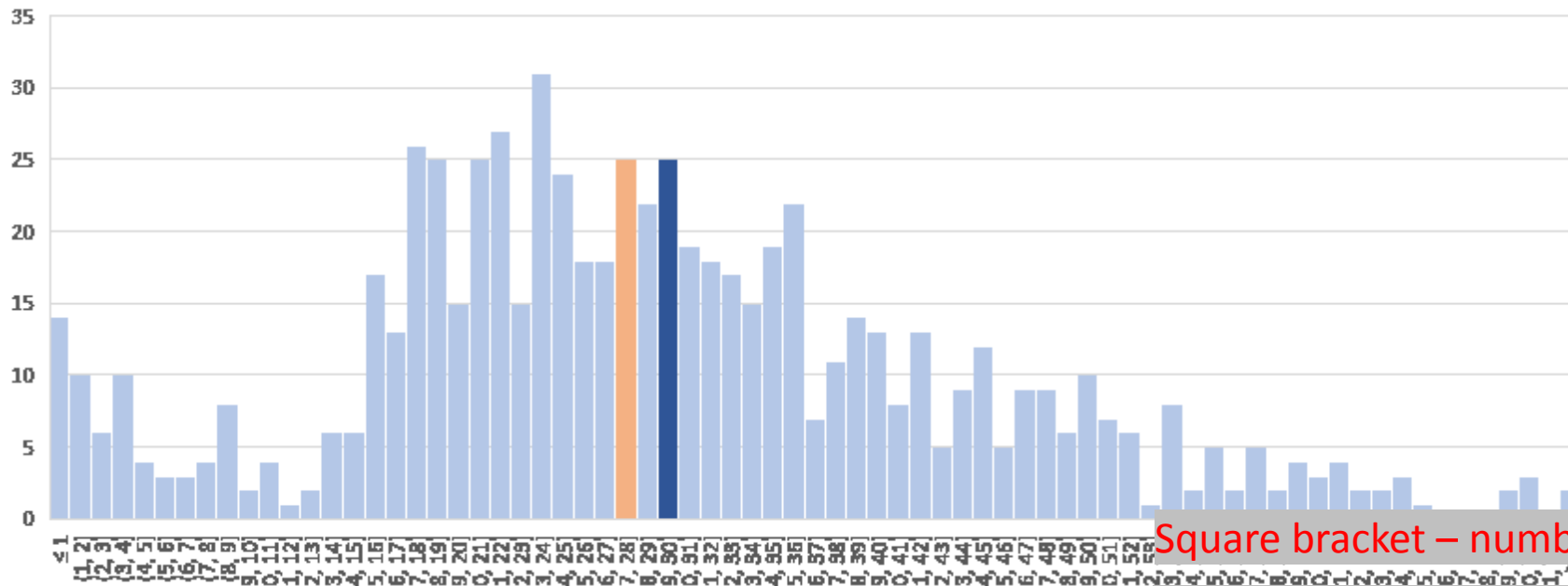


Bin width = 20



Highlighting mean

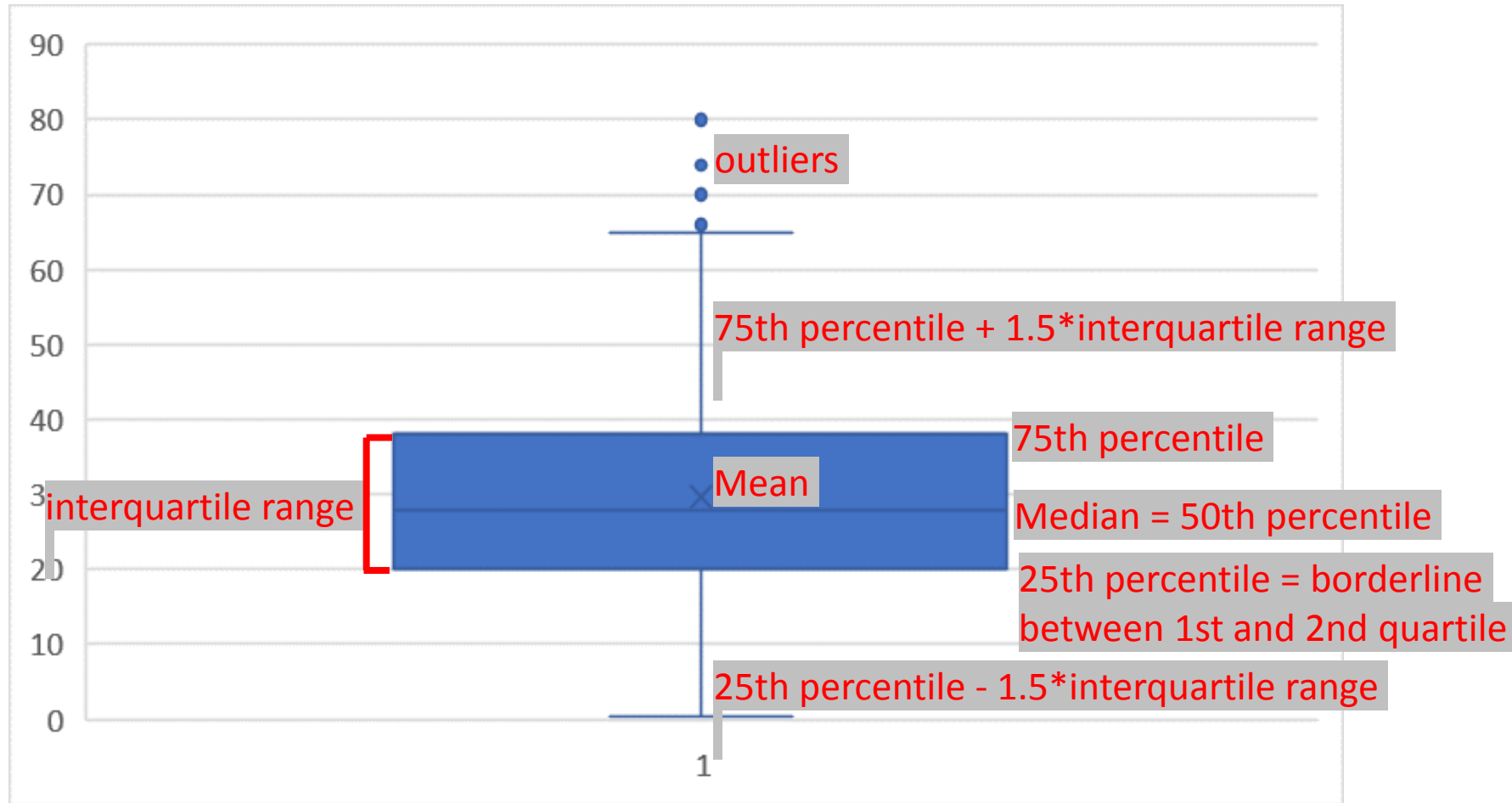
- =AVERAGE(*column*)
- = MEDIAN(*column*)
- Change of color mave to be done manually



Square bracket – number belongs to interval

Round bracket – number does not belong to interval

Box plot



- Avoid the usage of 3D versions
 - They can be very misleading

- Be careful with ratio aspect

- Always make titles, subtitles and labels as parsimonous as possible
 - (parsimony means to be maximally simplistic and maximally informative simultaneously)

