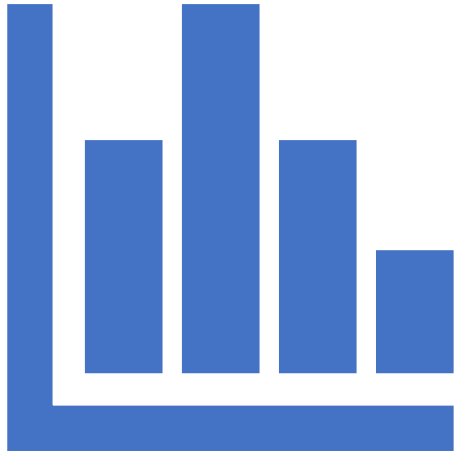
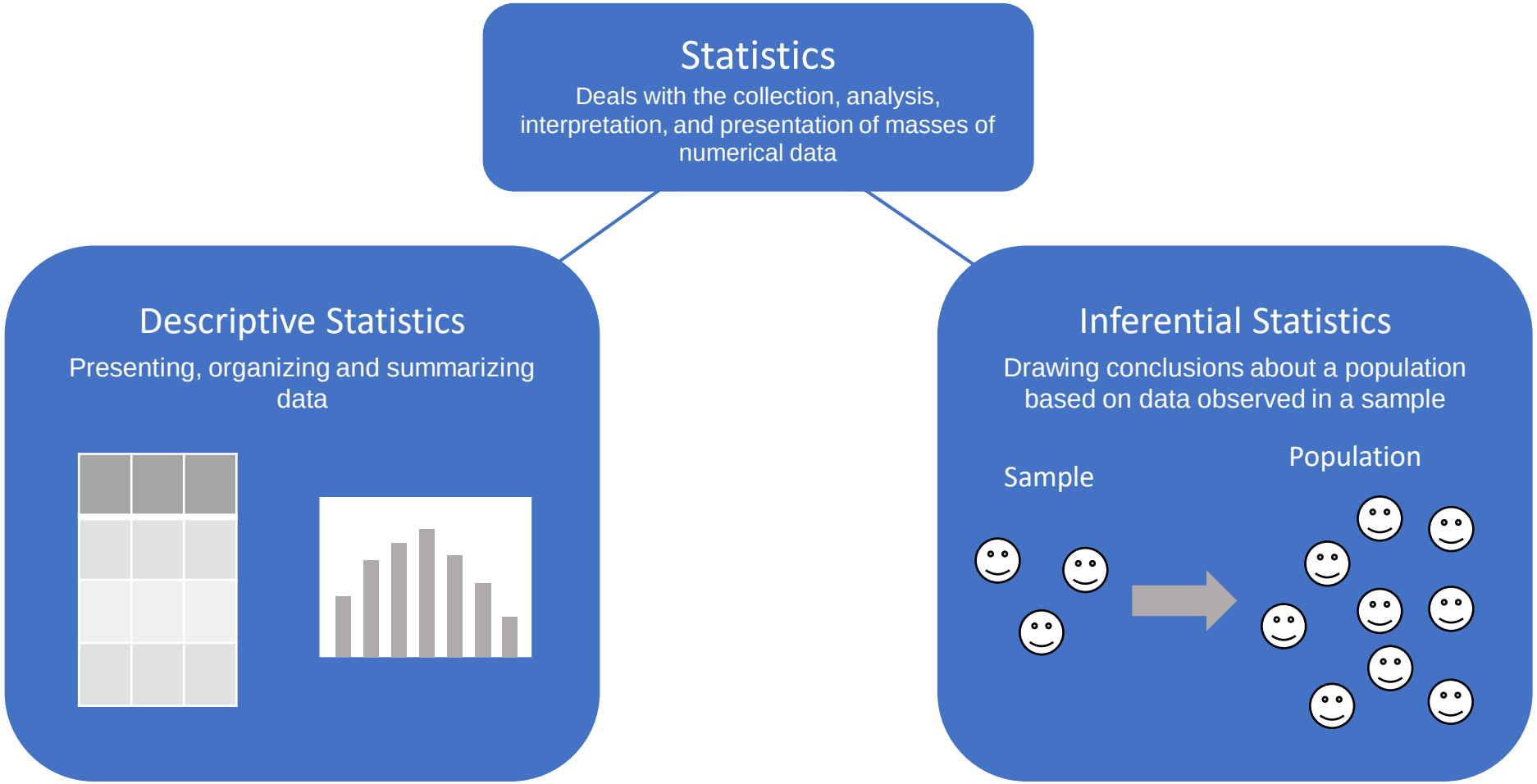
Descriptive Statistics and Diagrams

Lena Schielke, Ghafiq Alkurdi, Chris Vizcarrondo

Presentation Contents

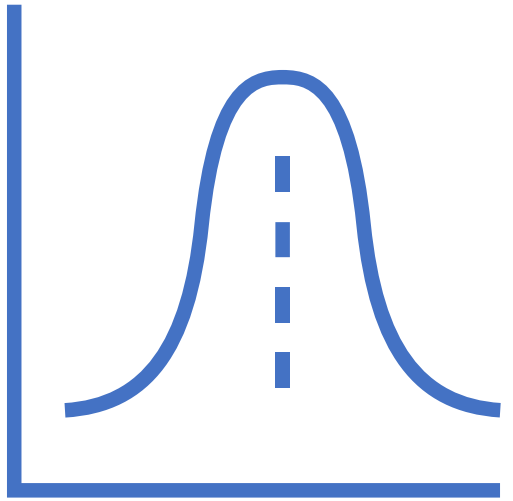


1 Defining Statistics



1 Descriptive Statistics in Detail

- Focuses on the sample
- Determines if we can compare a sample to a larger population
- Determines if the sample is "normally distributed" and is visualized on a bell curve, which is required most of the time
- The statistics are displayed in tables, charts, percentages, frequency distributions, and measures of central tendency



2 Frequencies

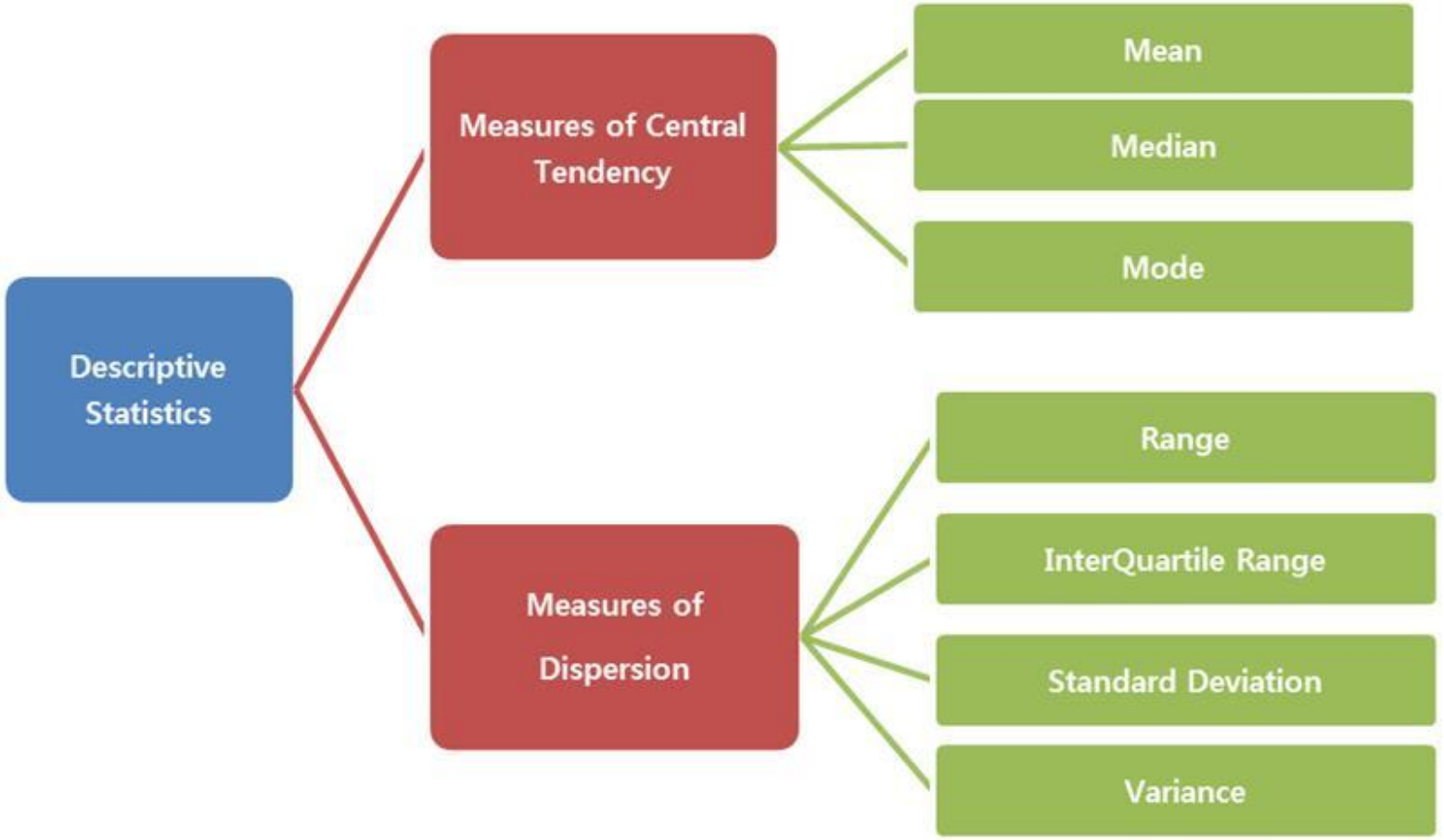
- Absolute Frequency: number of occurrences
- Relative Frequency: absolute frequency divided by the total number of occurrences

Age	Absolute Frequency	Relative Frequency	Percent in %
22	1	0.067	6.6
23	3	0.2	20
24	4	0.267	26.7
25	5	0.333	33.3
26	2	0.133	13.3
Total	15	1	100

Example: Calculate Relative Frequency
F(23) = 1/15 = 0.067
F(24) = 4/15 = 0.267

Example: Calculate Percent
P(23) = 0.2 x 100 = 20
P(24) = 0.267 x 100 = 26.7

3 Overview of the Measures



https://www.whatissixsigma.net/basic-statistics/

3.1 Measures of Central Tendency

Measures of central tendency are described by...

- **Mean**- the average
- **Median**- the midpoint
- **Mode**- value that appears most frequently

3.1 Measures of Central Tendency- Mean

Mean (average)

The mean is sensitive to extreme scores (outliers) in the sample

- The formula for samples
  - n is number of data items in sample
- Formulas for population
  - N is number of data items in population
- μ is used to present the population mean
- $\bar{x}$ is used to present a sample mean

Sample Mean

$$\bar{x} = \frac{\Sigma x}{n}$$

where Σx is sum of all data values

N is number of data items in population

n is number of data items in sample

Participant	Age
1	23
2	22
3	24
4	23
5	26
6	25
7	24
8	26
9	24
10	25
11	25
12	23
13	25
14	25
15	24

$$\bar{x} = \frac{23 + 22 + 24 + 23 + 26 + 25 + 24 + 26 + 24 + 25 + 25 + 23 + 25 + 25 + 24}{15} = 24.27$$

3.1 Measures of Central Tendency- Median

Median (midpoint)

The median is a better representative of a sample when scores are extreme, not sensitive to extreme scores

- Put numbers in order
  - If N is odd, the middle score is the media
  - If N is even, take the average of the two middle scores

22, 23, 23, 23, 24, 24, 24, 24, 25, 25, 25, 25, 25, 26, 26

Median = 24

Age	Absolute Frequency
22	1
23	3
24	4
25	5
26	2
Total	15

3.1 Measures of Central Tendency- Mode

Mode is the value that appears most frequently in a data set

- You can calculate the mode for everything
  - It is the only way to measure nominal variables such as gender, race, grades, etc.
    - These should be reported as numbers and percentages
- *No mode*: if no two categories are the same, there is no mode
- *Several modes*: More than one mode is possible
- *Visual depictions of modes*: Single mode, bi modal, multi-modal (shows varying peaks- 1, 2 or more peaks) depicting variables on a graph

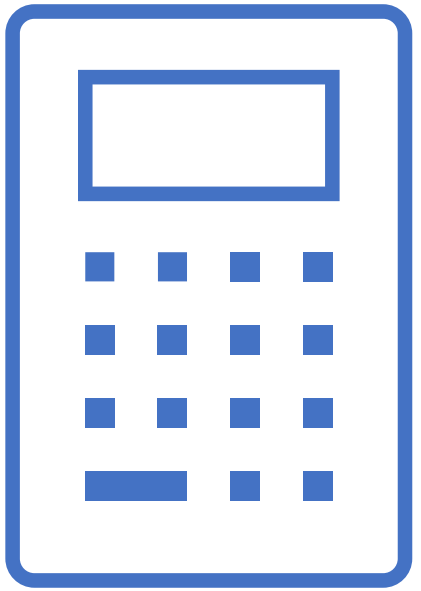
Age	Absolute Frequency
22	1
23	3
24	4
25	5
26	2
Total	15

Mode = 25

3.1 Measures of Dispersion

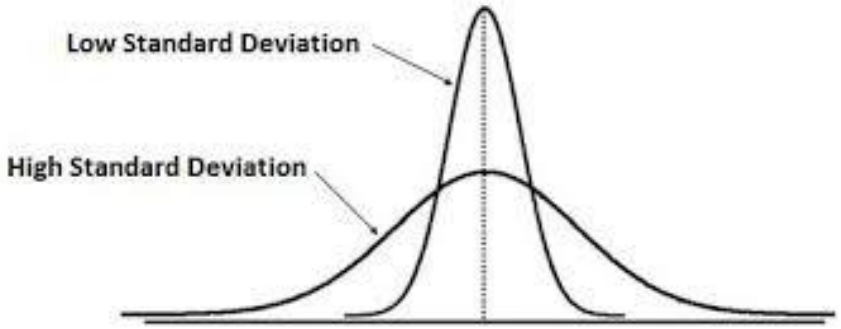
Measures of dispersion "is the degree to which data is distributed around this central tendency" (Yartsev, 2017)

- Standard deviation
- Variance



3.2 Standard Deviation

- Describes how much variation exists from average/mean
- This measures variability and shows how spread-out numbers are
- Standard deviation is the most frequently used measure of dispersion or variability and is sensitive to extreme values
    - Symbolized by the Greek letter sigma: σ
    - Is depicted as a bell curve with the center showing the normal distribution that is 1 standard deviation of the mean
    - Low standard deviation means the data is in a cluster, high standard deviation means the data is more spread out and varied



3.2 Standard Deviation

The formula includes each score, the mean, number of values, and sum across the values. It is the square root of the variance.

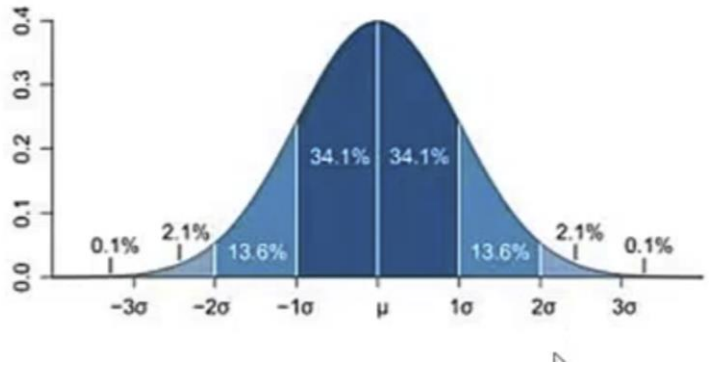
$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{(n - 1)}}$$

where:

X = each score
$\bar{X}$ = the mean or average
n = the number of values
Σ means we sum across the values

Standard Deviation (s) - the average distance of scores from the mean

Visual Representation of Standard Deviation- 68% of the population in a normal distribution is within 1 standard deviation of the mean.



3.2 Example: Calculate Standard Deviation

$$S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{(n - 1)}}$$

where:

X = each score
$\bar{X}$ = the mean or average
n = the number of values
Σ means we sum across the values

Standard Deviation (s) - the average distance of scores from the mean

Age	Absolute Frequency
22	1
23	3
24	4
25	5
26	2
Total	15

Mean = 24.27

$$S = \sqrt{\frac{(22-24.27)^2 + (23-24.27)^2 + (23-24.27)^2 + (23-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (26-24.27)^2 + (26-24.27)^2}{(15-1)}} = 1.16$$

3.2 Variance

- Describes dispersion of data points from the mean value
- Is the square of the standard deviation
  - Variance- the average of the squared differences from the mean
    - Steps for determining the variance
      1. Find the mean
      2. For each number, subtract the mean and square the result (squared difference)
      3. Find the average of the squared differences

Age	Absolute Frequency
22	1
23	3
24	4
25	5
26	2
Total	15

$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

$$S^2 = \frac{(22-24.27)^2 + (23-24.27)^2 + (23-24.27)^2 + (23-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (24-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (25-24.27)^2 + (26-24.27)^2 + (26-24.27)^2}{(15-1)} = 1.35$$

3.2 Difference between Standard Deviation and Variance

	Standard Deviation	Variance
Definition	Standard deviation looks at how spread out a group of numbers is from the average	The variance is a measure of variability of values from the average
Purpose	Describes variability of data of figures that are quite large; more meaningful than variance	Useful when doing math
Calculation	Root of the average of the squared differences from the mean	Average of the squared differences from the mean
Unit	Same unit as values	Squared units

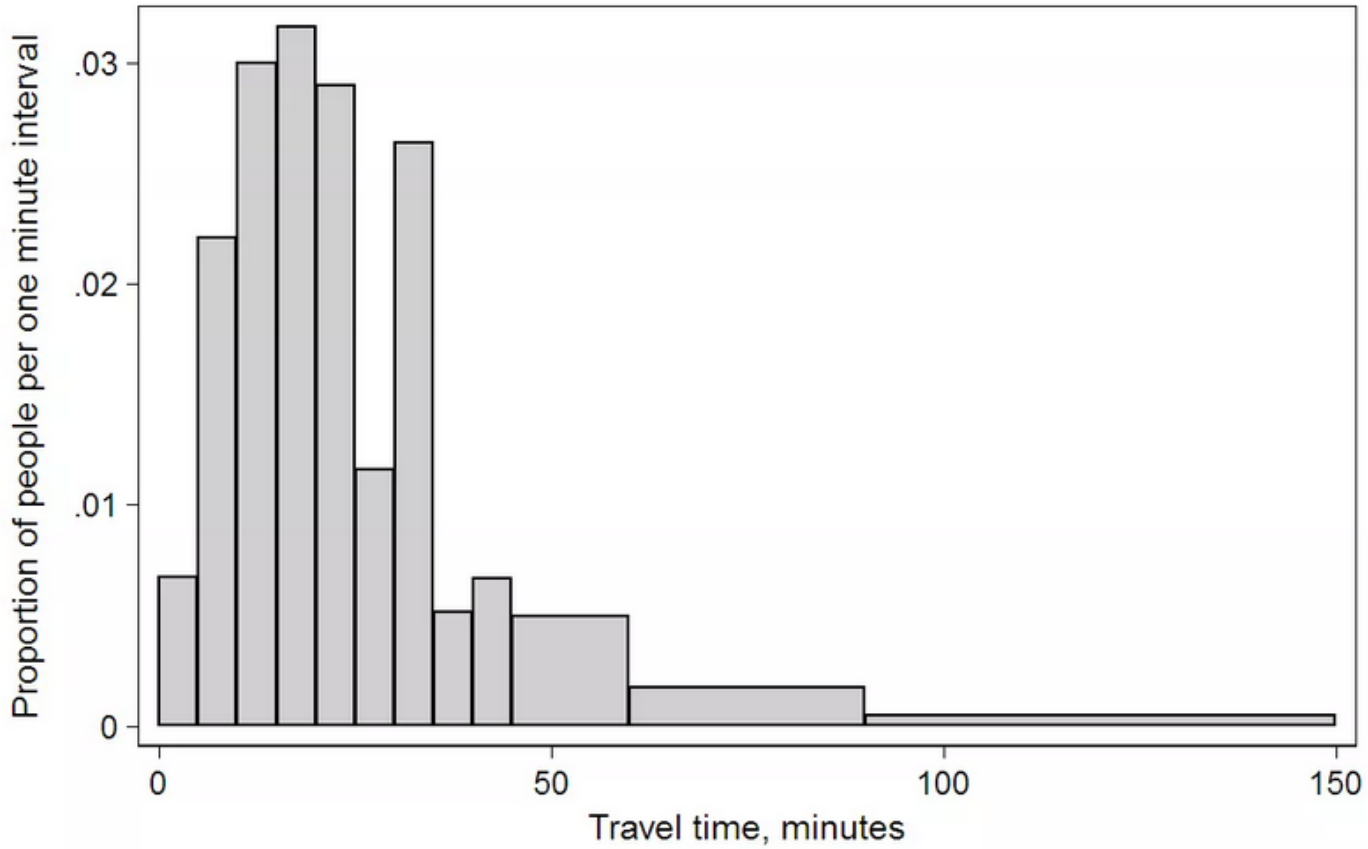
4 Diagrams - Bar Graph

- Divided into different categories
- Shows amount in each category
- Categorical data

https://keydifferences.com/difference-between-histogram-and-bar-graph.html

4.1 Histogram

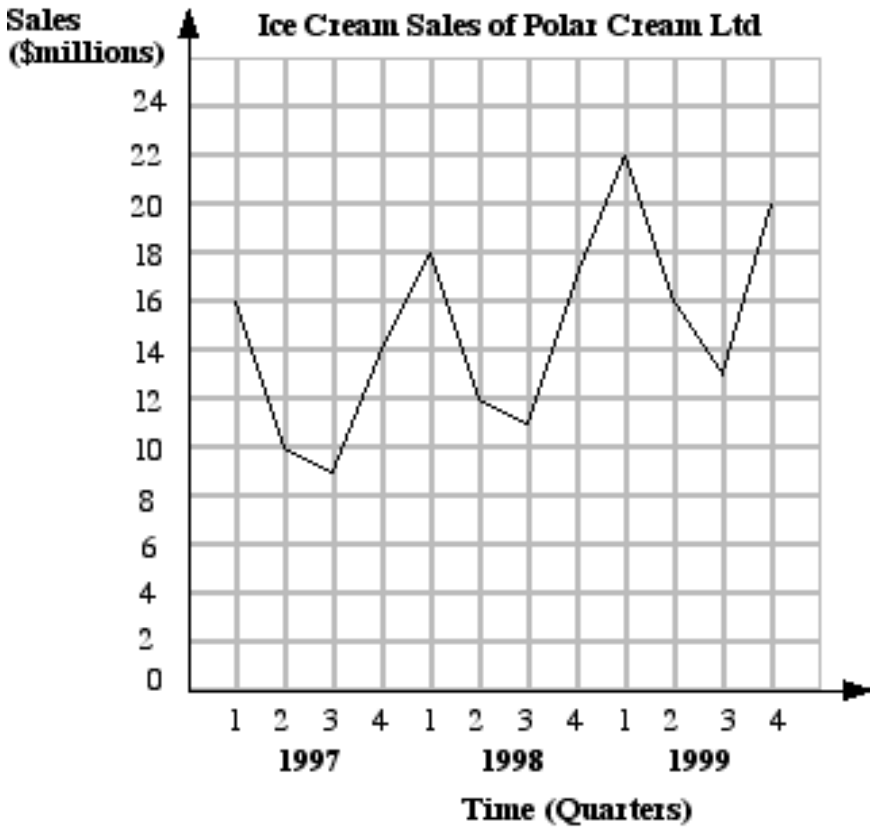
- Different from a bar graph because of the level of measurement of the data



https://www.thoughtco.com/frequently-used-statistics-graphs-4158380

4.2 Line Graph

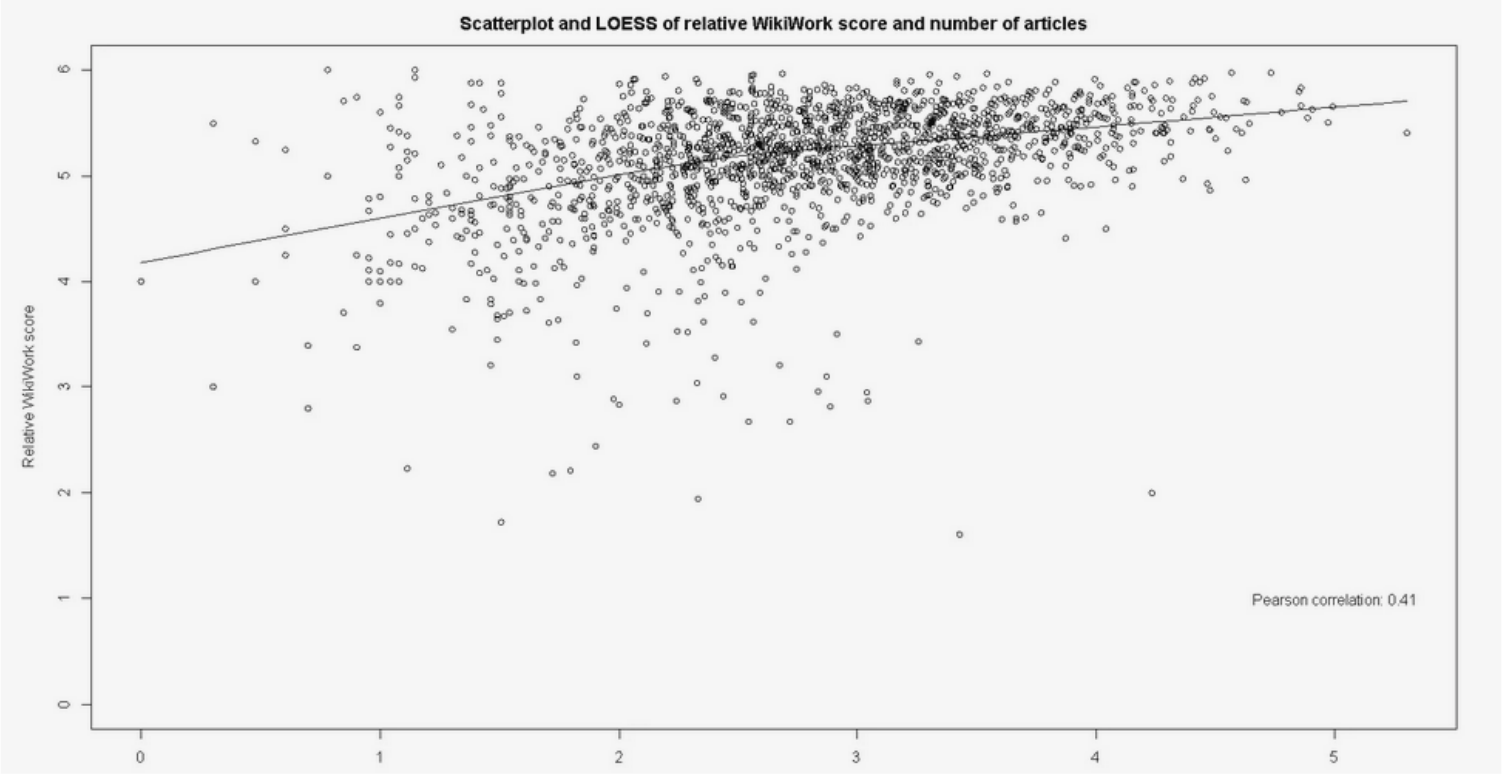
- Graph shows data on a line
- Demonstrates trends or numbers that are interconnected



https://www.thoughtco.com/frequently-used-statistics-graphs-4158380

4.3 Scatterplot

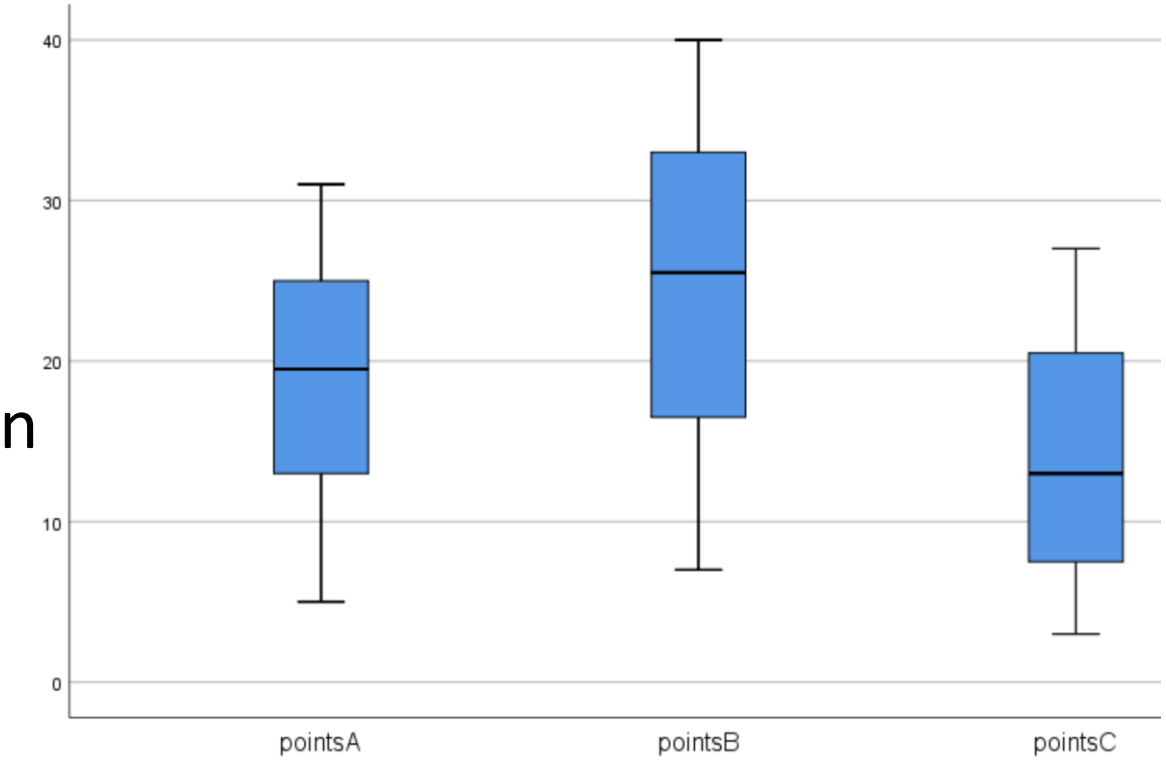
- Displays data as dots that are scattered across the page
- Helps to uncover overall trends among variables and outliers



https://www.thoughtco.com/frequently-used-statistics-graphs-4158380

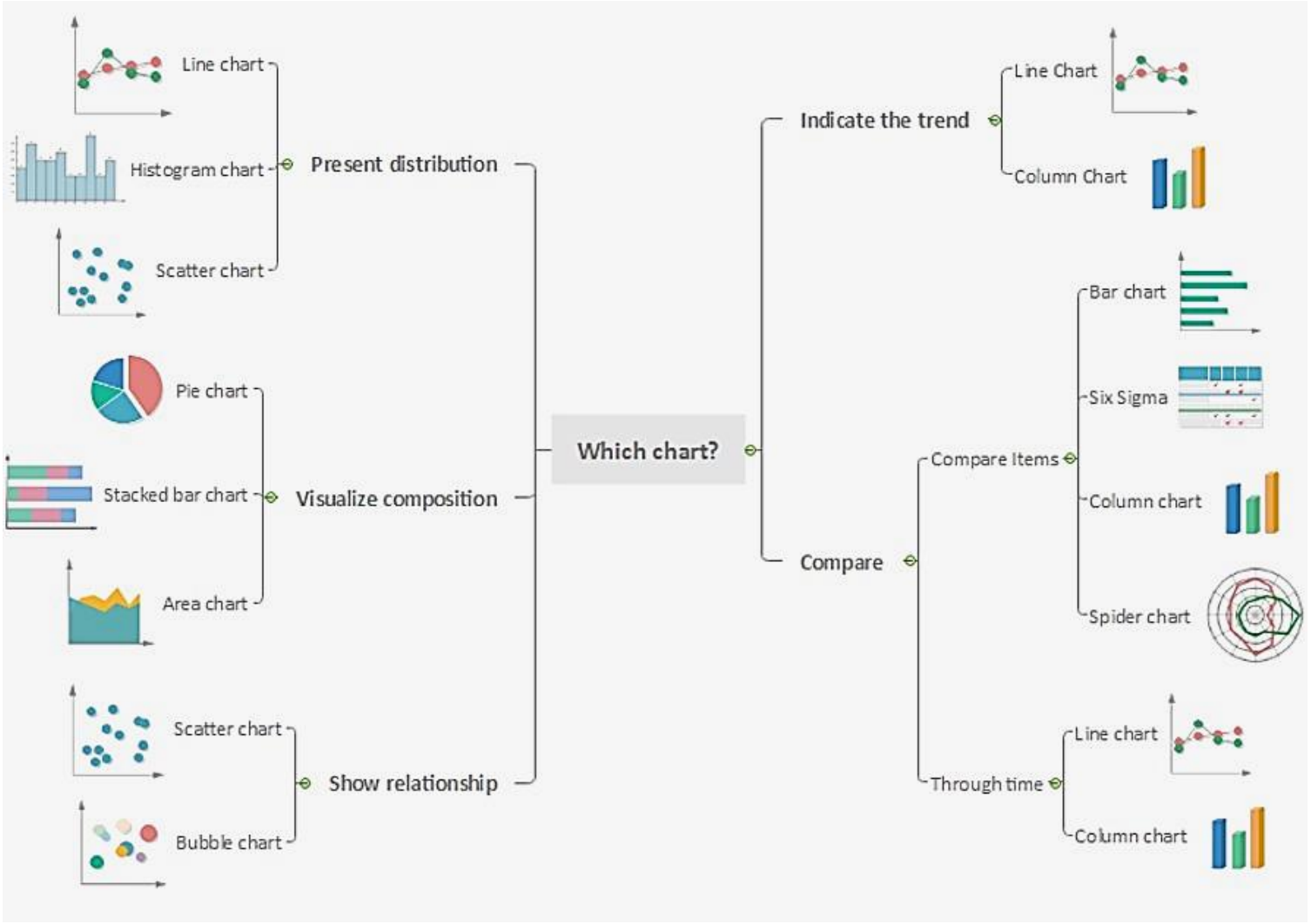
4.4 Boxplot

- The boxplot shows how data is distributed.
- It shows the spread and the center/average
- Measure of the spread is shown in interquartile ranges
  - Minimum- lowest data
  - Quartile 1- bottom 25%
  - Quartile 2- mean/average
  - Quartile 3 - top 25%
  - Maximum - highest data



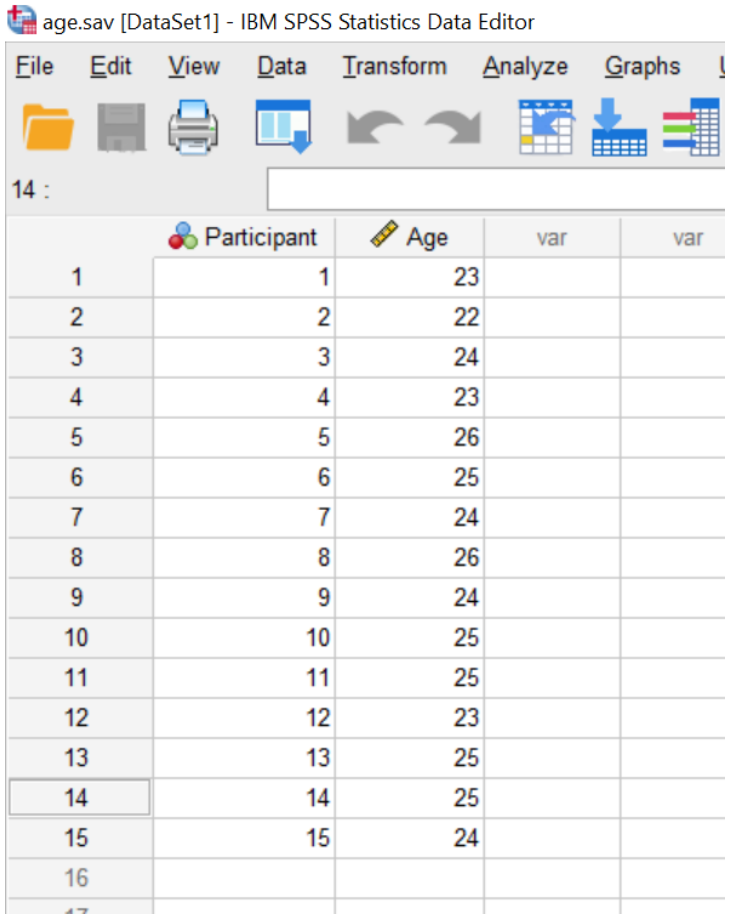
https://statisticsbyjim.com/basics/graph-groups-boxplots-individual-values/

4.5 Overview



https://www.edrawsoft.com/chart/choose-right-chart.html

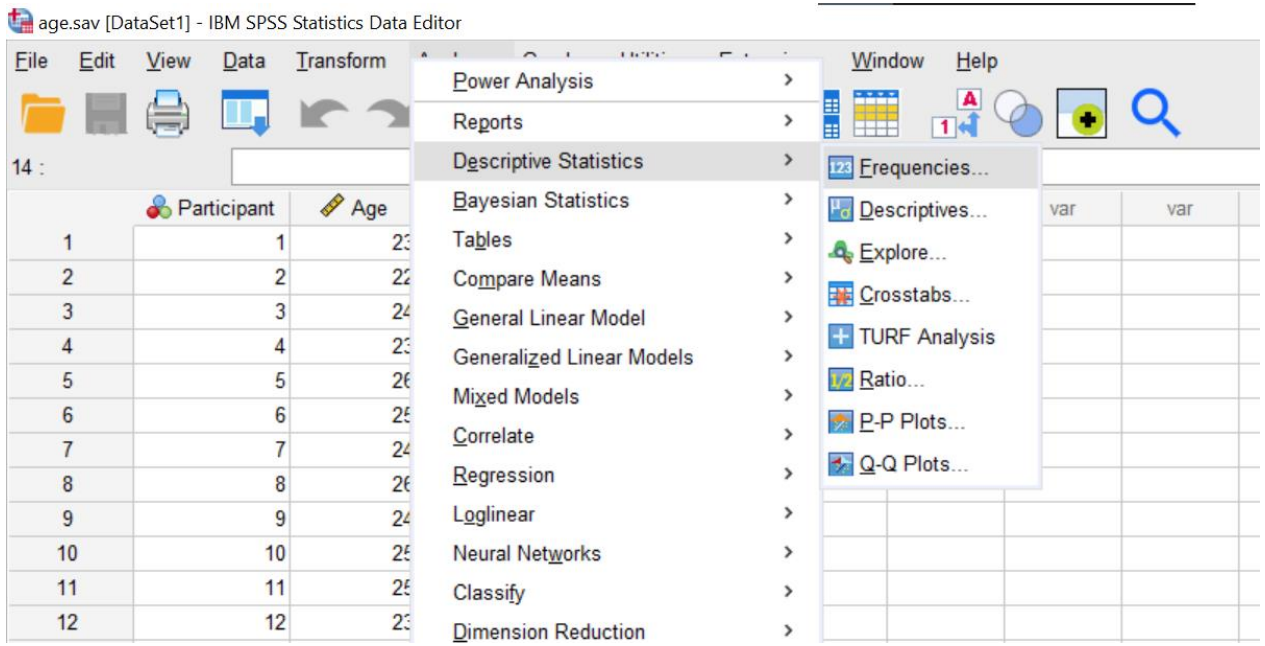
5 How to calculate with SPSS



	Participant	Age	var	var
1	1	23		
2	2	22		
3	3	24		
4	4	23		
5	5	26		
6	6	25		
7	7	24		
8	8	26		
9	9	24		
10	10	25		
11	11	25		
12	12	23		
13	13	25		
14	14	25		
15	15	24		
16				

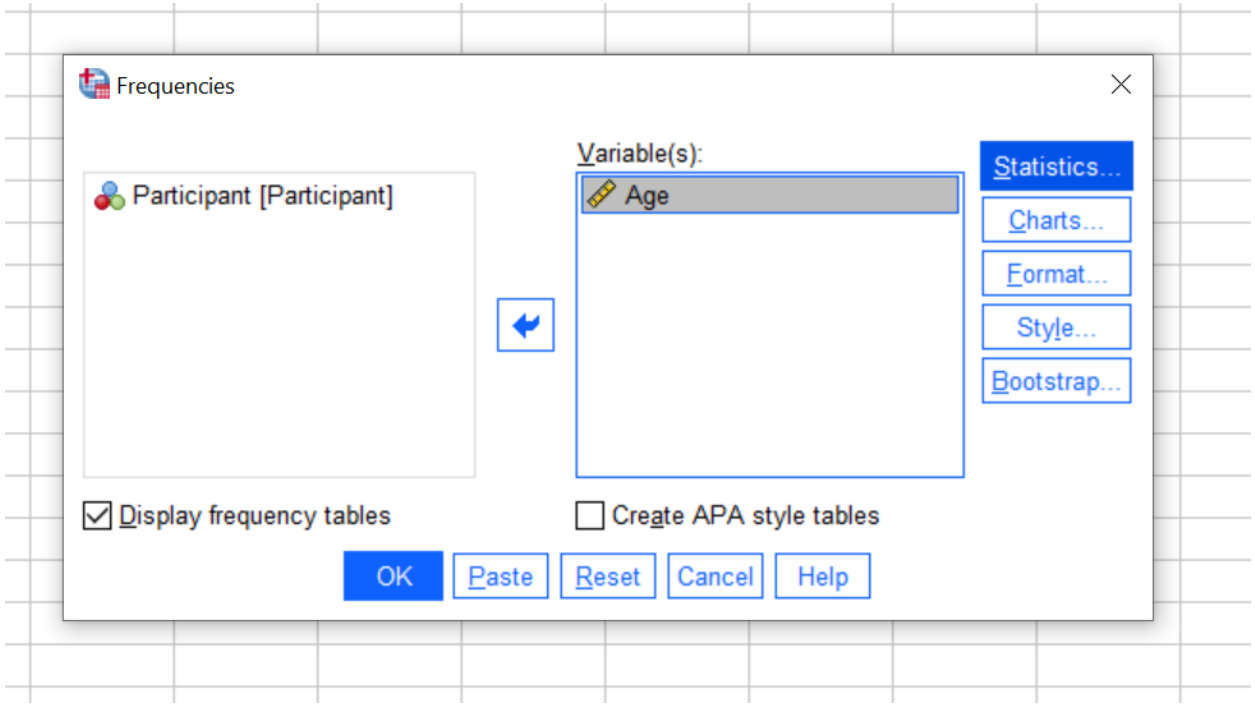
5.1 Frequencies – Step 1

- Select *Analyze* > *Descriptive Statistics* > Frequencies...



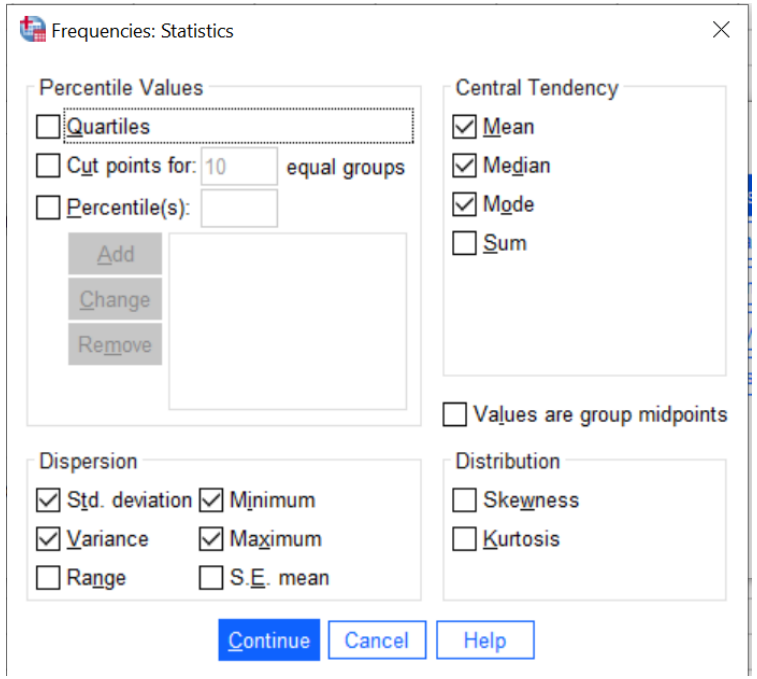
5.1 Frequencies – Step 2

- Select the variable and click on *Statistics*



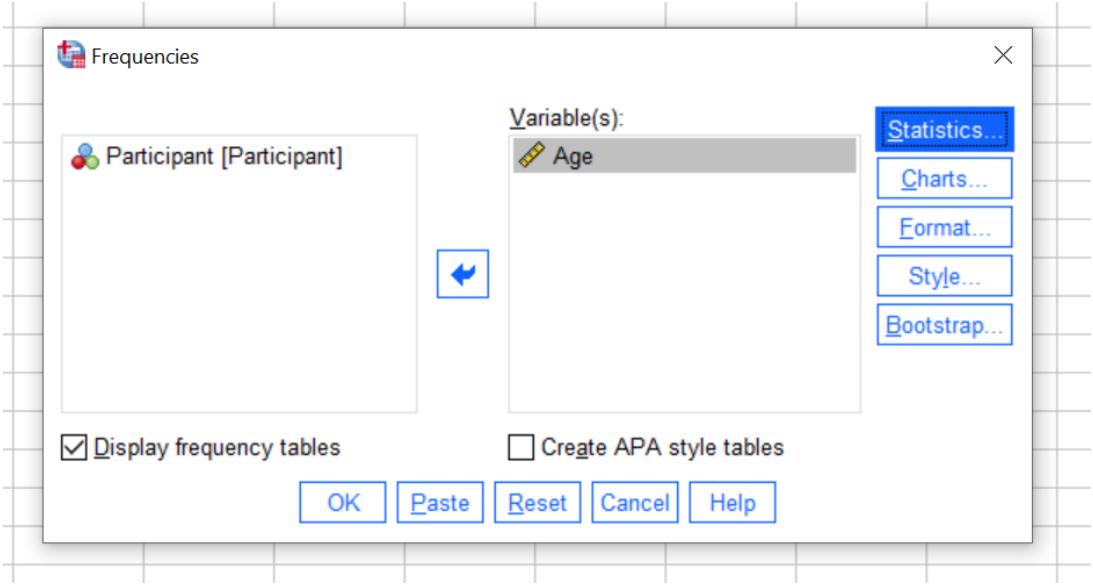
5.1 Frequencies – Step 3

- Select all values you would like to calculate and click on *Continue*



5.1 Frequencies – Step 4

- Click on *OK*



5.1 Frequencies – End

- The values are calculated.

Frequencies

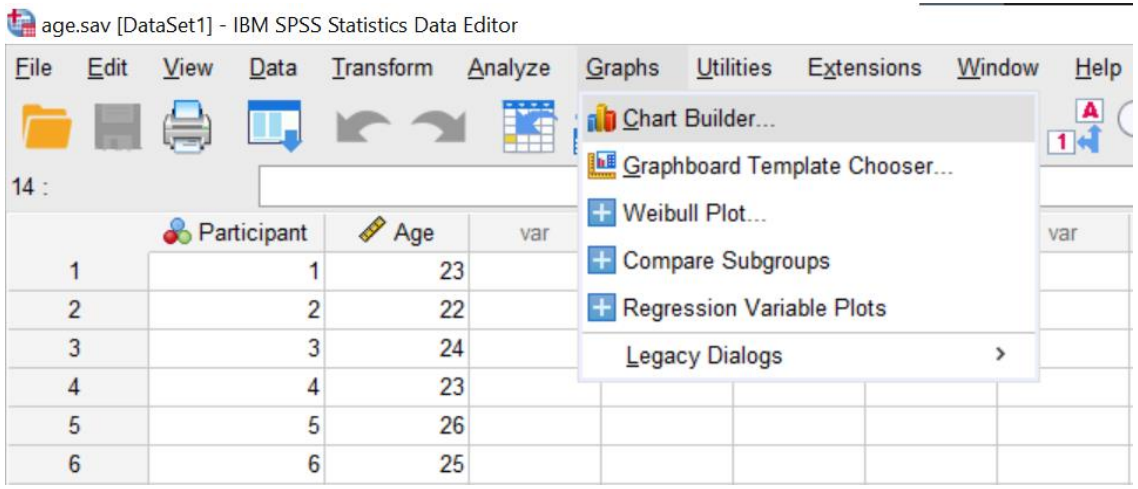
Statistics

Age

N	Valid	15
	Missing	0
Mean		24,2667
Median		24,0000
Mode		25,00
Std. Deviation		1,16292
Variance		1,352

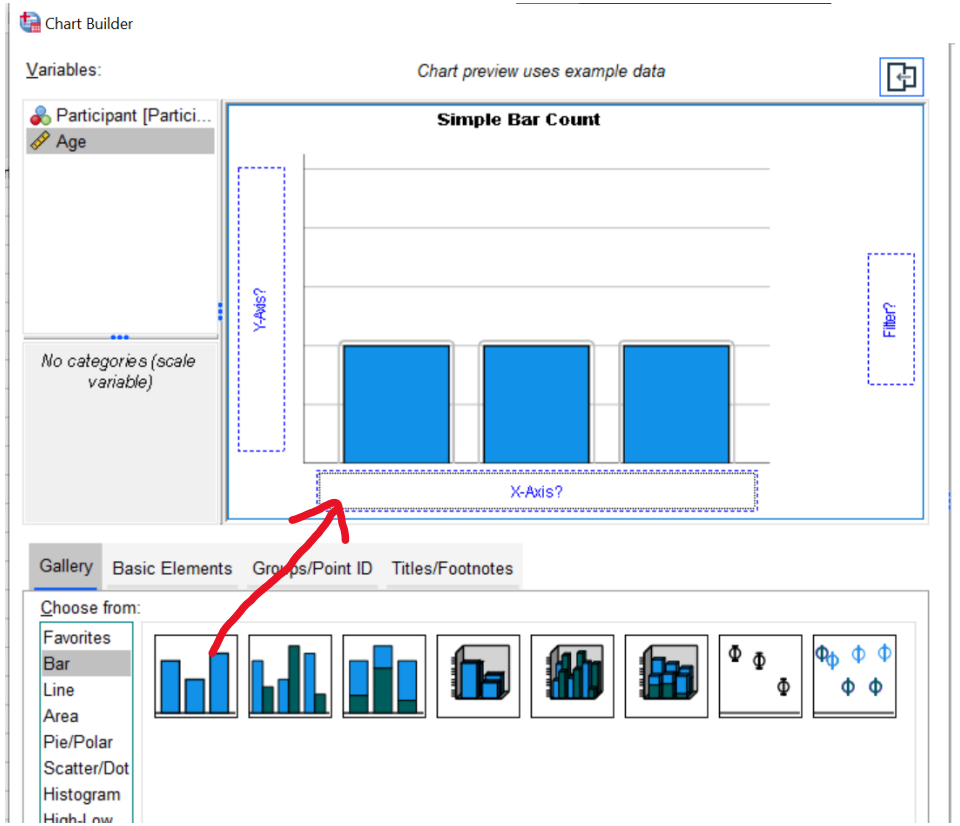
5.2 Graphs – Step 1

- Click on *Graphs* > *Chart Builder*



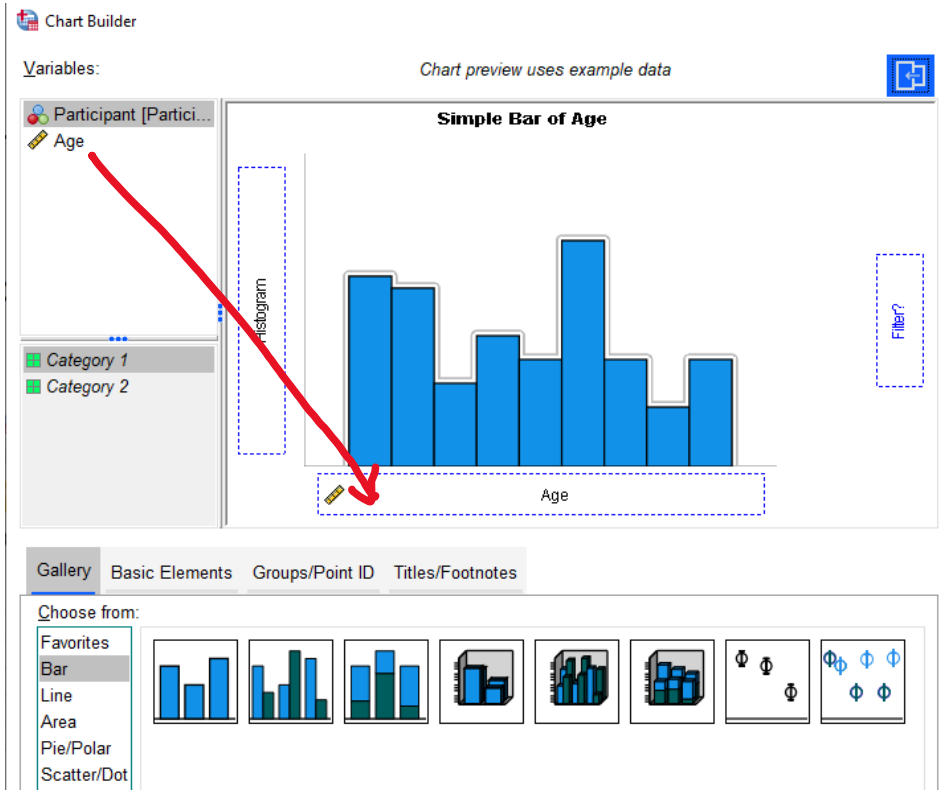
5.2 Graphs – Step 2

- Drag and drop the desired type of graph into the field



5.2 Graphs – Step 3

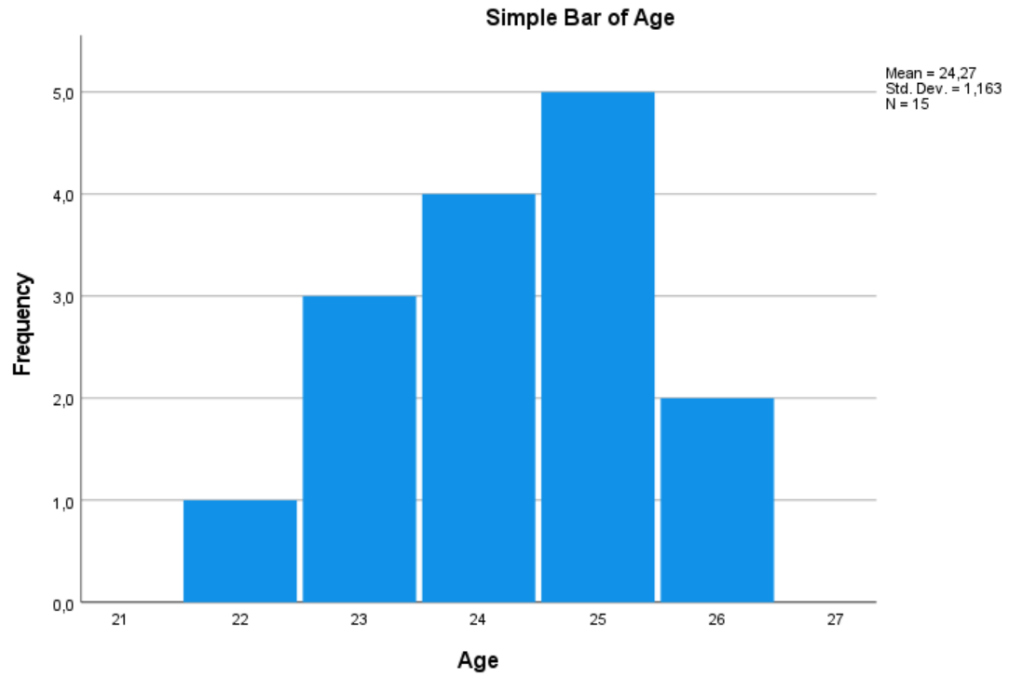
- Drag and drop the variable on one of the axes



5.2 Graphs – End

- A diagram is created.

GGraph



References

Basic Statistics. (2019). WhatisSixSigma.net. https://www.whatissixsigma.net/basic-statistics/

Benjamin, C. (2018, August 17). *How to describe charts, graphs, and diagrams in the presentation*

https://preply.com/en/blog/2018/08/17/charts-graphs-and-diagrams-in-the-presentation/#scroll-to-heading-0

Between Descriptive and Inferential Statistics. (2017, July 6). DifferenceBetween.net. http://www.differencebetween.net/language/words-language/difference-between-descriptive-and-inferential-statistics/

Freeman, James. (2019, November 26). *How to Choose the Right Chart - A Complete Chart Comparison.* Edraw. https://www.edrawsoft.com/chart/choose-right-chart.html

Johnson, Teresa. [Teresa Johnson]. (2013, June 13). *Introduction to Descriptive Statistics* [Video]. YouTube https://youtu.be/QoQbR4lVLrs

stikpet. [stikpet]. (2006, April 24). S02E01 - *Absolute Frequency and Relative Frequency* [Video]. YouTube https://www.youtube.com/watch?v=DPmwWxYYCp4

Surbhi, S. (April 25, 2016). *Difference Between Variance and Standard Deviation*. Key Differences. https://keydifferences.com/difference-between-variance-and-standard-deviation.html#comments

Yartsev, (2017, July 24). *Variability, dispersion and central tendency.* Deranged Psychology. https://derangedphysiology.com/main/cicm-primary-exam/required-reading/research-methods-and-statistics/Chapter%203.0.2/variability-dispersion-and-central-tendency