

# Thesis

# Data visualisation

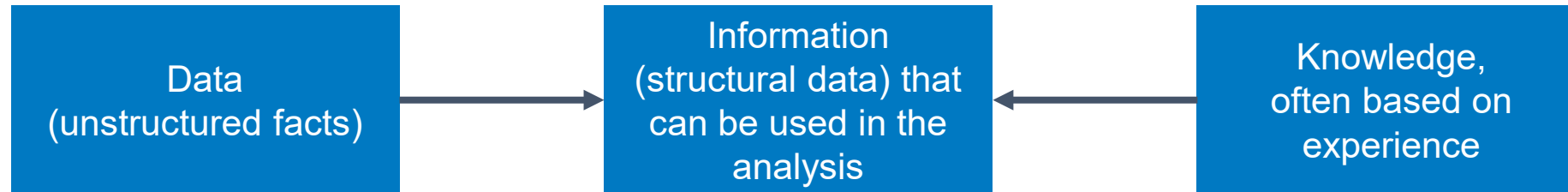
Teija Aarnio  
Development manager, D.Sc. (Tech.), M.Sc.

2023



# What is data?

## - Three levels



# What is data?

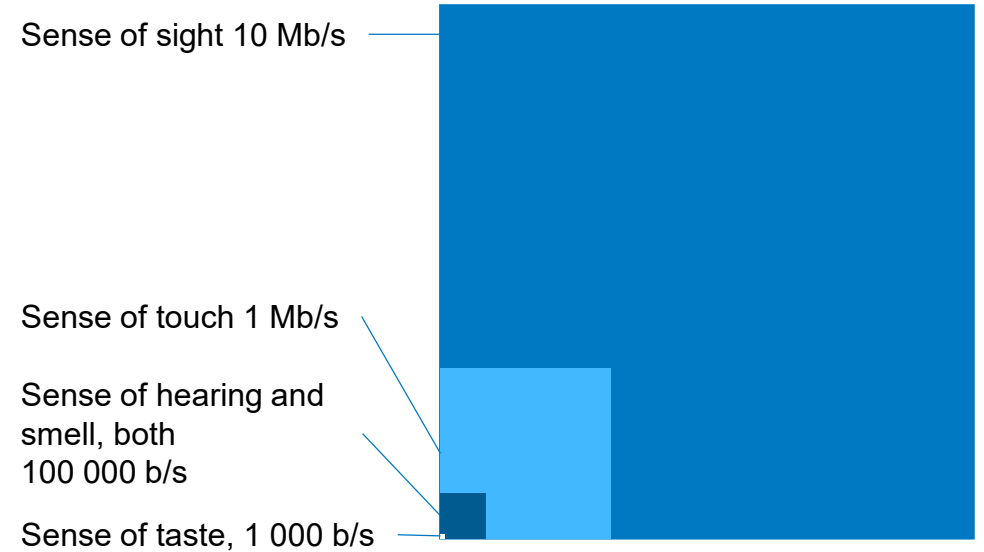
## - Qualitative and quantitative data

- **Qualitative data** tend to describe the topic. They seek to find out people's motivations, mindsets, and attitudes, such as perceptions, opinions, and perspectives. This makes the results more difficult to analyze. Qualitative data are not as structured as quantitative ones.
- **Quantitative data** are figures. They are structured, accurate, computational and often statistical in nature. They help to see the bigger picture and the results are clear to analyze.
- Qualitative data cannot answer the question of why. **Information on correlation, where quantitative information is needed, is always needed to verify the cause-and-effect relationship.**

# What is visualisation?

Definition of visualization researcher Robert Kosara

- Visualization is based on (non-visual) **data**,
  - The visualization process results in a **picture**
  - Which must be **interpretable and identifiable**.
- 
- Why visualisation?
    - The sense of sight is superior to other senses.
    - Our visual sense transmits information to our brain at all times, approximately eight times more than the other senses combined.



Lähde: Juuso Koponen, Jonatan Hildén, Tapio Vapaasalo: Tieto näkyväksi – Informaatiomuotoilun perusteet

# Data visualisation

## Infographics

- Exploratory graphics supporting communication.
- The task is to pass information from one person to another.
- Options
  - Static
  - Animated
- Often describe numerical data or other information that can be easily converted to such.
- Can also describe qualitative information, such as event flows and principles.
- **Tells the story to the reader**

Information  
(structural data)  
that can be used  
in the analysis

## Visualisation

- Exploratory new graphics revealing features of the material.
- The task is to find new information, not just to communicate information.
- Serves as a tool that allows the reader to find interesting features in the material.
- Most often implemented on a computer and are often interactive.
- The underlying material is always numerical information or at least information that can be converted to numbers.
- **The reader can find their own story in the material.**

Lähde: Juuso Koponen, Jonatan Hildén, Tapio Vapaasalo: Tieto näkyväksi – Informaatiomuotoilun perusteet

# Data visualisation in thesis

- Helps both the author of the thesis and its reader to understand the data.  
"One picture speaks more than a thousand words."
- Creates a common understanding.
- Prevents misunderstandings.



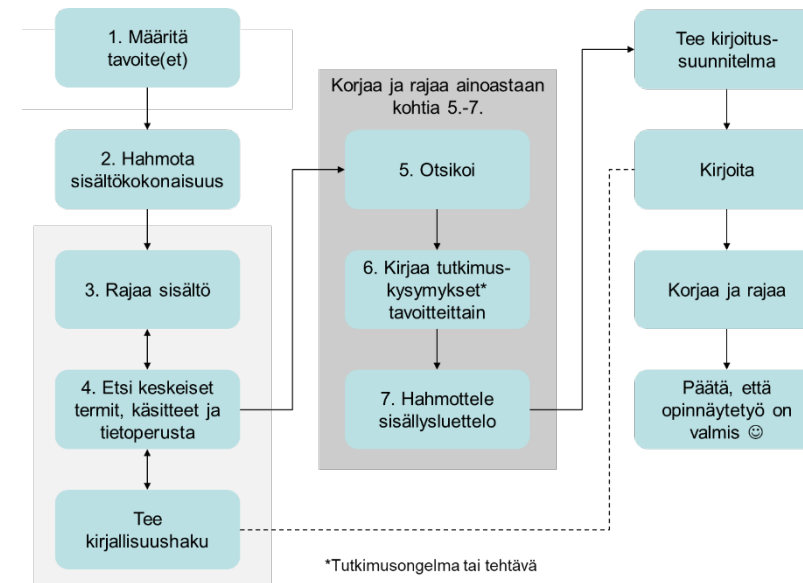
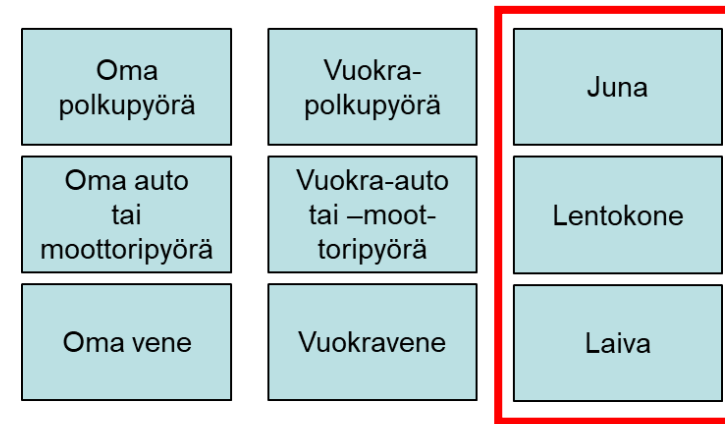
# 1. principle: Appropriate for purpose

- Whatever the mode of visuality, the visualisation must be appropriate for purpose.
- Appropriate for the purpose means that the visualization gives the reader (1) the information the reader needs (2) from the reader's perspective.
- Use
  - a picture to indicate scoping or principle.
  - a table when there is a lot of numerical or verbal data.

Use a picture, when you desire to present scoping

Use a picture, when you desire to present scoping

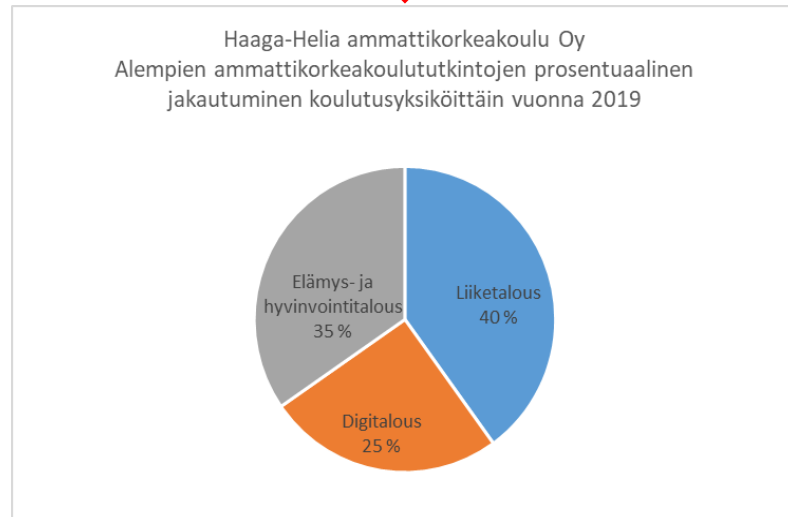
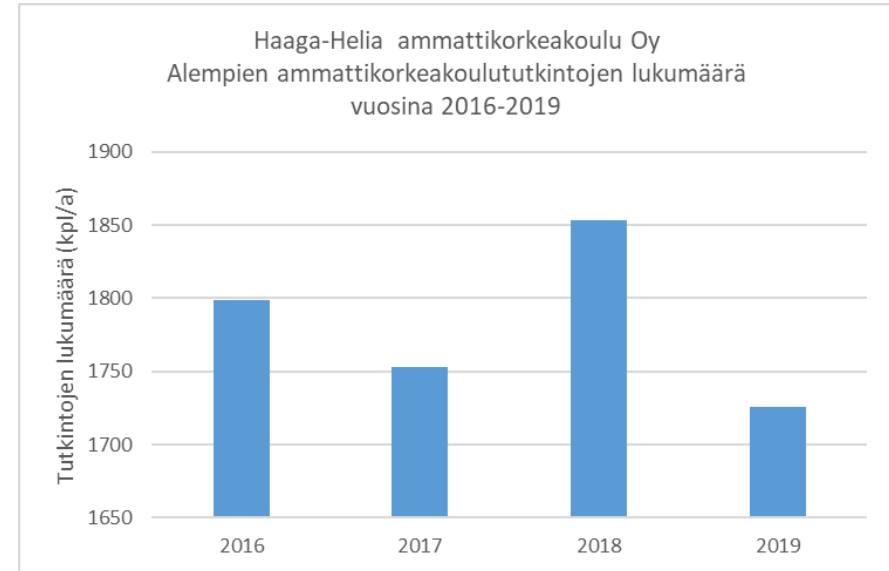
## Matkustusvaihtoehtoja





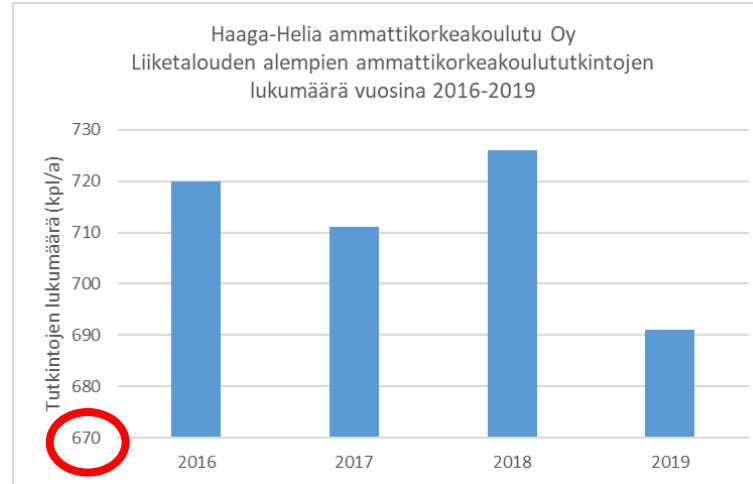
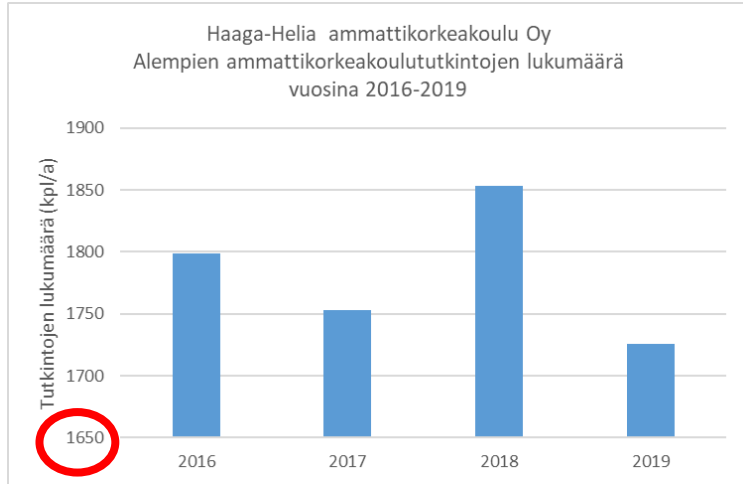
# Example

Vuositain suoritettut alemmat ammattikorkeakoulututkinnot				
	Vuodet			
Tutkintojen lukumäärä (kpl)	2016	2017	2018	2019
Yhteensä	1799	1753	1853	1726
Liiketalous	720	711	726	691
Digitalous	456	448	504	480
Elämys- ja hyvinvointitalous	623	594	627	555



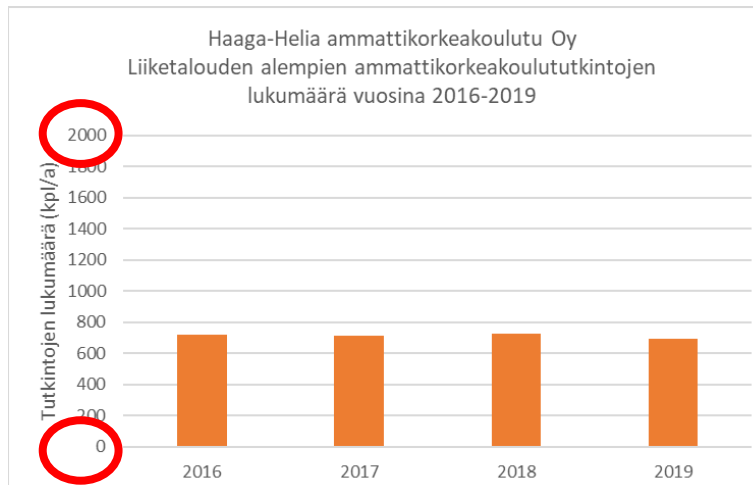
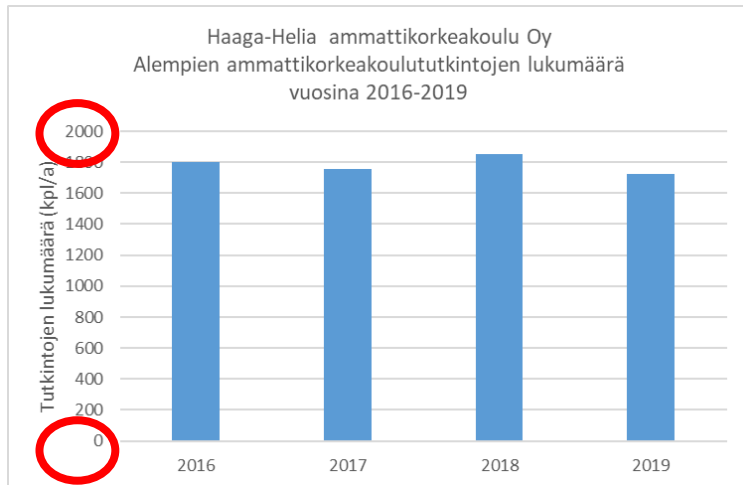
Same data,  
but different visualisation  
for different purposes

# Example



## Automatic

The differences between the columns are visible when the y-axis does not start from zero



## Modified manually

The comparison is easier when both charts have the same y-axis

and when the columns have different colors.

# Example

	Number of observations by option 1-5					Average of observations
	1	2	3	4	5	
Case 1	14	32	55	48	45	38,80
Case 2	8	24	49	55	58	38,80

	Number of observations by option 1-5					Number of observations	Average	Standard deviation	Standard deviation (%)
	1	2	3	4	5				
Case 1	14	32	55	48	45	194	38,80	16,177	42 %
Case 2	8	24	49	55	58	194	38,80	21,811	56 %

When you do statistical analysis and you report the average,

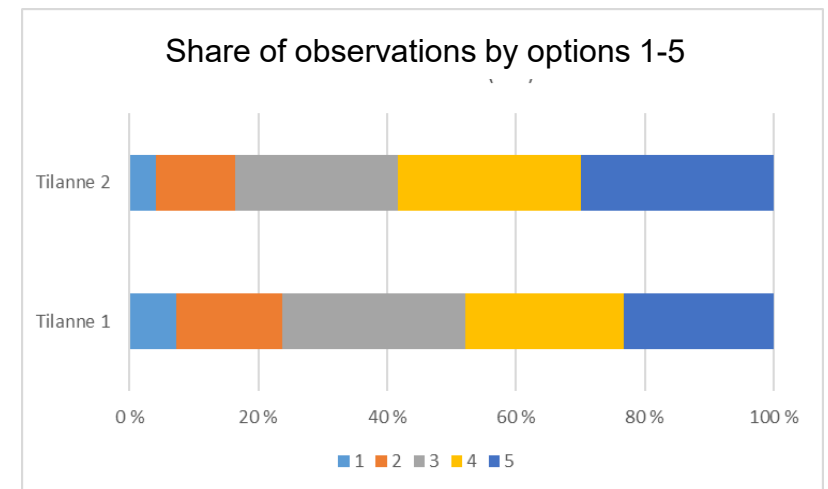
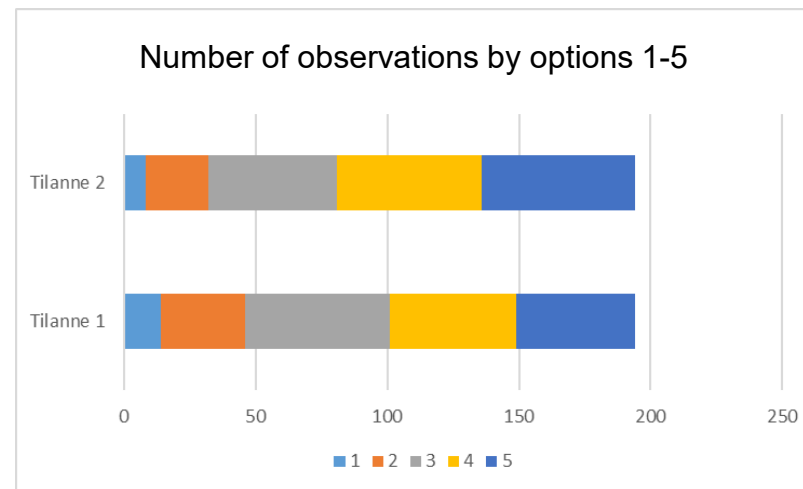
- Always indicate the total number and the standard deviation and standard deviation as a percentage.
- Discuss all four values in the text. For example, "The standard deviation is between 42 and 56%, which describes the dispersion of responses."

# Example

		Number of observations by options 1-5					Number of observations	Average of observations
		1	2	3	4	5		
Case 1		14	32	55	48	45	194	38,80
Case 2		8	24	49	55	58	194	38,80

A useful alternative to the average is for example a stacked horizontal column according to either the number of responses or their percentages.

The graph shows the distribution of observations better than the average alone.



# Average and its options

- **The arithmetic mean** is the most common of the distribution means. To obtain the average, the measured values of all observations are summed and divided by the number of observations. The arithmetic mean is only suitable for describing distributions described on a ratio scale (e.g., money, weight, length) or on a distance scale (e.g., temperature, index).
- **The median** should be used on an order scale (for example, level of education). The median indicates the point that divides the distribution into two parts, with half of the observations in both parts. The median is not as sensitive to outliers as the arithmetic mean.
- When the distribution is based on quality differences, the median cannot be used either, but the only possible average is the type value, i.e. the **mode**. It is the value that is observed most often. The same distribution can have more than one mode class.

<https://tilastokoulu.stat.fi>

# Example

	Havainnot vaihtoehtoisin (1-5)								
	1	2	3	4	5	Havaintojen lukumäärä (kpl)	Havaintojen keskiarvo	Havaintojen keskihajonta	Keskihajonta (%)
Tilanne 1	14	32	55	48	45	194	38,80	16,177	42 %
Tilanne 2	8	24	49	55	58	194	38,80	21,811	56 %

38,8      16,2      42  
                  21,8      56  
                  three      three      two  
                  (two)      (two)  
                  significant figures

When presenting numbers, remember the significant figures:

[https://en.wikipedia.org/wiki/Significant\\_figures](https://en.wikipedia.org/wiki/Significant_figures)

The significant figures does not take into account zeros at the end of the integer and zeros at the beginning of the decimal number. Other numbers are significant. For example, 0.0220 (three significant figures) and 5300 (two significant figures).

Be aware of program presets presenting usually two decimal places.

Usually two or three significant figures are sufficient. Use the same significant figures in the reading series for example, the standard deviation of the answers.

## 2. principle: Self-explanatory

- Whatever the mode of visibility, visualization must be self-explanatory.
- Self-explanatory means that all the information needed to interpret the visualization is found in the visualization or its caption.

# Example

Vaihtoehto	Selite	Myynti (€) <sup>1</sup>	Asiakkaiden lukumäärä (kpl) <sup>1</sup>
Uusmyynti	Palvelun myynti uudelle asiakkaalle <sup>2</sup>	+	+
Lisämyynti	Uuden palvelun myynti nykyiselle asiakkaalle <sup>3</sup>	+	±
Sopimusmuutos	Asiakkaan nykyisen palvelusopimuksen sisällön muuttaminen kesken sopimuskauden <sup>4</sup>	+ tai -	±
Sopimusuusinta	Asiakkaan palvelu-sopimuksen uusiminen sopimuskauden päättyessä <sup>5</sup>	+ tai ± tai -	±

<sup>1</sup> + kasvaa, ± pysyy ennallaan, - laskee

<sup>2</sup> Uusi asiakas tarkoittaa asiakasta, (a) joka on ollut, mutta ei ole enää organisaation asiakas tai (a) joka ei ole ollut koskaan aikaisemmin organisaatio asiakas (ei voimassa olevaa asiakassuhdetta).

<sup>3</sup> Nykyinen asiakas tarkoittaa asiakasta, joka ostaa tällä hetkellä yritykseltä vähintään yhtä palvelua (voimassa oleva asiakassuhde).

<sup>4</sup> Sopimuksen sisältö joko laajenee tai supistuu.

<sup>5</sup> Jatkosopimus on usein sisällöltään sama tai lähes sama kuin edellinen sopimus.

- Provide relevant information.
- Add explanations, definitions, and additional information as a footnote.
- Add a unit to the title.



# Example

- Explain all other abbreviations and symbols except SI abbreviations and prefixes ([https://en.wikipedia.org/wiki/International\\_System\\_of\\_Units](https://en.wikipedia.org/wiki/International_System_of_Units)).
- Mark the euros correctly (€ million, € 1,000)
  - t = time, K = kelvin, thermodynamic temperature
  - k = kilo- ( $10^3$ ), M = mega- ( $10^6$ )

# Summary

1. Find a **perspective**: what the visualization says, what the information is used for, and who uses it.
2. Determine which **data** is appropriate for this purpose. Limit the data to be presented.
3. Choose **the appropriate visualisation method**.
4. Make a **self-explanatory** visualization presenting the data and its explanations.

Practice visualizing your data because it is a good and useful skill.

