

Data Analysis and Visualization

Topic 01: Introduction, data analysis 101, and Tableau crash course

Course practicalities

Course topics, agenda and outline

Weekly routine (new topic intro + assignment workshops), project work (introduced later)

Working with the assignments (exercises, solving, submitting, peer-reviews, marking progress points)

Completing the course

Moodle, MS Teams and study groups

First assignment

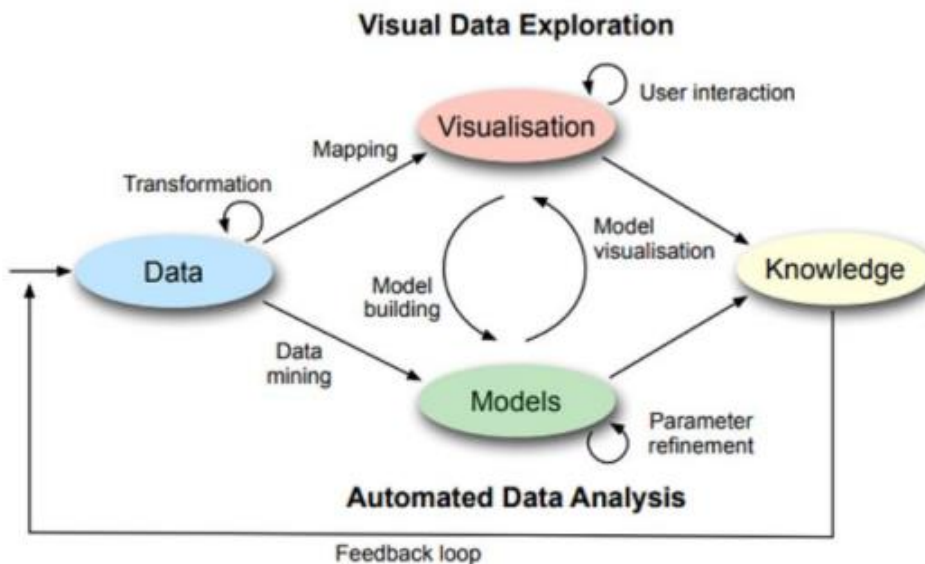
What is data analysis and visualization?

Big picture: From data to business (revenue) via CRISP-DM (CRoss Industry Standard Process for Data Mining)



Source: <https://www.datascience-pm.com/crisp-dm-2/>

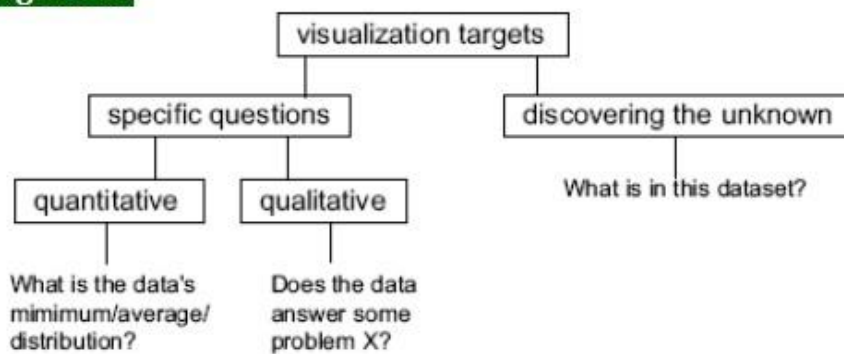
Detailed picture: Visual data analytics process (note that in this course, we restrict scope to the upper branch in the path)



Source: <https://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf>

Fundamental forms of data analysis and visualization questions

Figure 1.1



Types of questions targeted by the visualization process.

Source: https://andor.tuni.fi/permalink/358FIN_TAMPO/1j3mh4m/alma9910687952005973

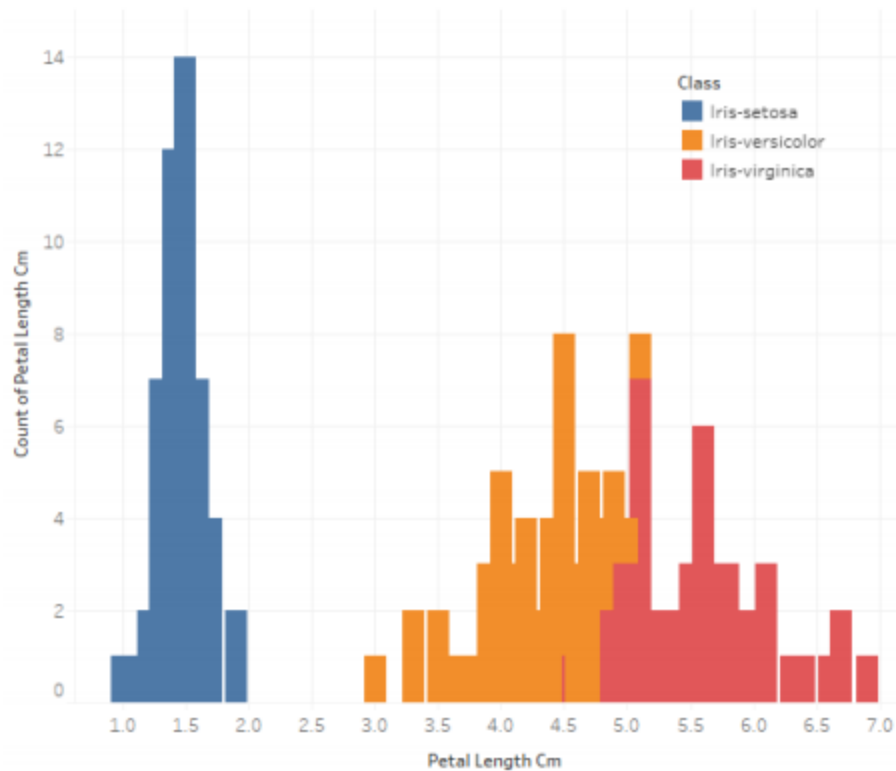
In practice, the objective often lies in capturing actionable insights from data, in forms of various kinds of statistics and the related visualizations, e.g.:



	sepalLength	sepalWidth	petalLength	petalWidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000





Flower image source: https://en.wikipedia.org/wiki/Iris_flower_data_set

Analytics objectives may include, e.g.:

- Describing things, identifying objects, concepts or trends (e.g., statistics, classification, and regression)
- Clustering data
- Detecting anomalies

- Identification of (associative) rules
- Making predictions (e.g. ,classification and regression)

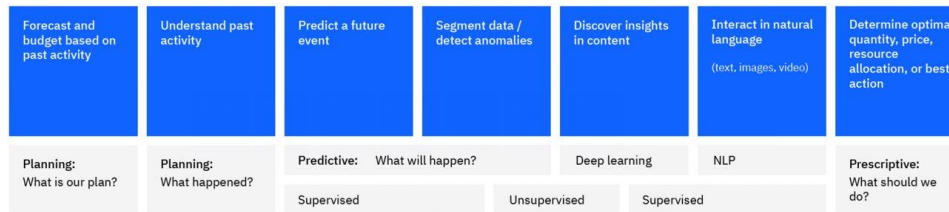
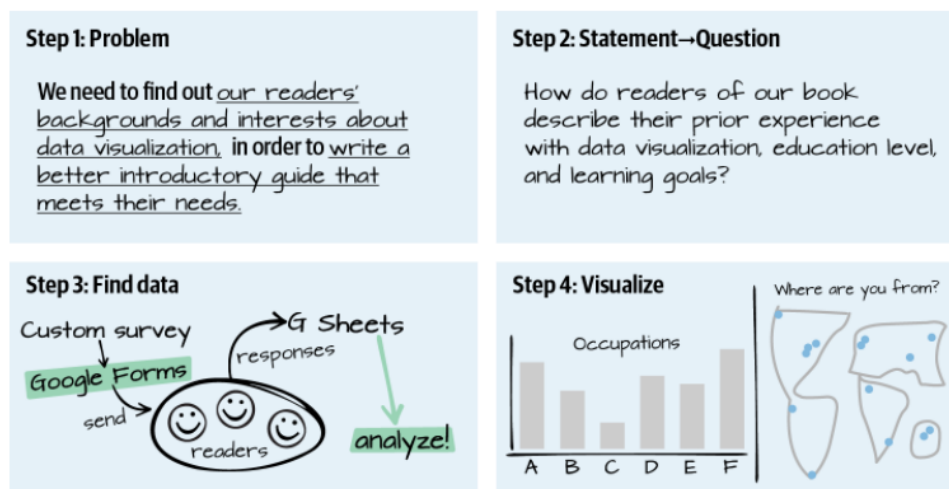


Image source: <https://www.ibm.com/garage/method/practices/think/data-needs-for-ai-data-science/>

Sketching a data story in practice (book planning example):



Source: https://andor.tuni.fi/permalink/358FIN_TAMPO/1j3mh4m/alma9911194976405973

Tableau Public demonstration

Example of using specific data visualization tool: Tableau Public

Sample analysis with the Iris dataset (see https://en.wikipedia.org/wiki/Iris_flower_data_set and <https://archive.ics.uci.edu/ml/datasets/iris>)

Potential questions: What's in the data? What kind of flowers are these? How to distinguish the flower classes?

Structure of a nice and clean dataset

Building simple visualizations: Connecting to data, adding a worksheet, compiling a dashboard and telling a story

Assignment 01