Blogging for Data Science

Basic Fundamentals

Basic Principles

Required

Send your post to ebonique@budgetcollector.org

- Minimum post length is 500 words
- Attach a featured image
- List out of names who contributed to the article
- Create a title for your blog

Recommended

- Post length is at least 1000 words
- Write an Excerpt
- Create a subtitle
- Have at least three headlines, such as:
  - TLDR
  - Team Members
  - Why We Chose This Database
  - How Johann Wolfgang von Goethe Influenced Our Color System Choice
- Identify tags for the article
- Identify focus keyword
- Create a team name, otherwise the author will be Georgia Tech.
- Write a short bio for each team member, submit a headshot for team members, and provide a link to your linkedin, etc.
- Create five social media posts. Example below:
  - “Sorting through the data, our team learned this about @uichicago’s art”
  - “Our team decided to evaluate 3000 of these artworks at @aiartadviser’s database. This is why...”

Using Gist

Using gist is an easy way to embed readable code in your articles.

First, go to GitHub.com, login, and select New Gist. Then write or copy-paste your code. If you include a file extension, such as .py, gist will color your code nicely.

Before clicking the green Create button, be sure you toggle it to Create public gist.

Then take the link and paste it in your blog editor and it will display like the image to the right.

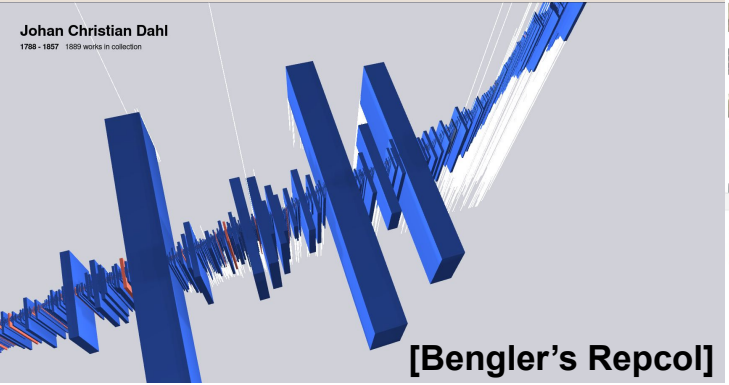
Recommended Blogs

- SmartData Collective https://www.smartdatacollective.com/
- Kaggle https://medium.com/kaggle-blog
- Inside BigData https://insidebigdata.com/

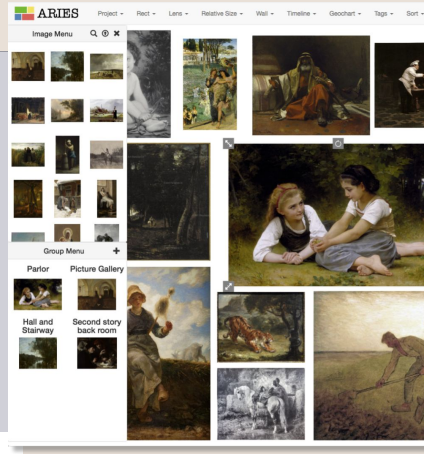
Data Visualization

Basic Fundamentals

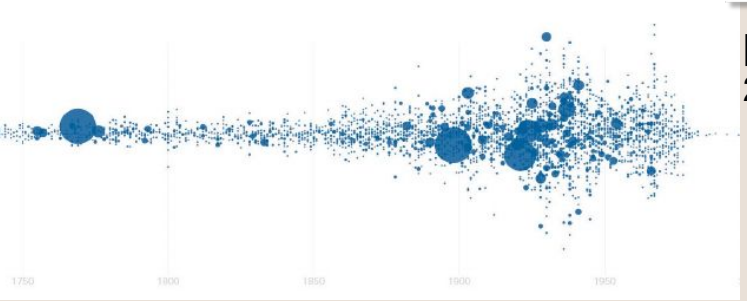
Example Uses in the Arts



[Bengler's Repcol]



[Crissaff et al, 2017]



[Kräutli, 2020]

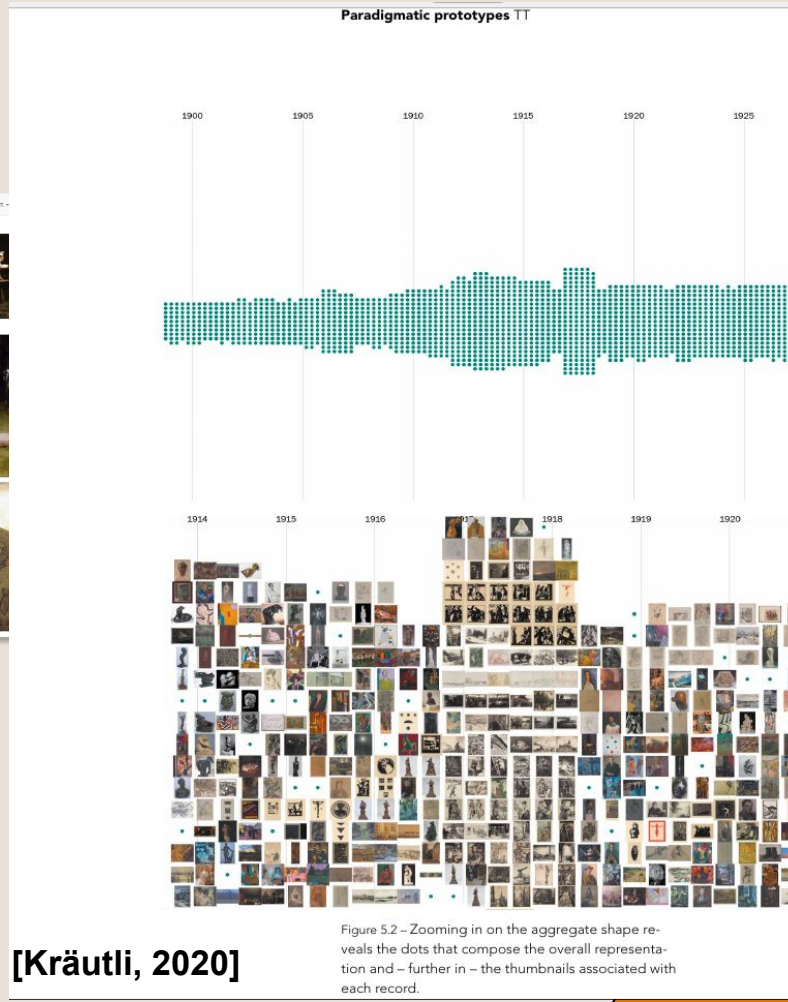


Figure 5.2 – Zooming in on the aggregate shape reveals the dots that compose the overall representation and – further in – the thumbnails associated with each record.

[Kräutli, 2020]

What is Data Visualization?

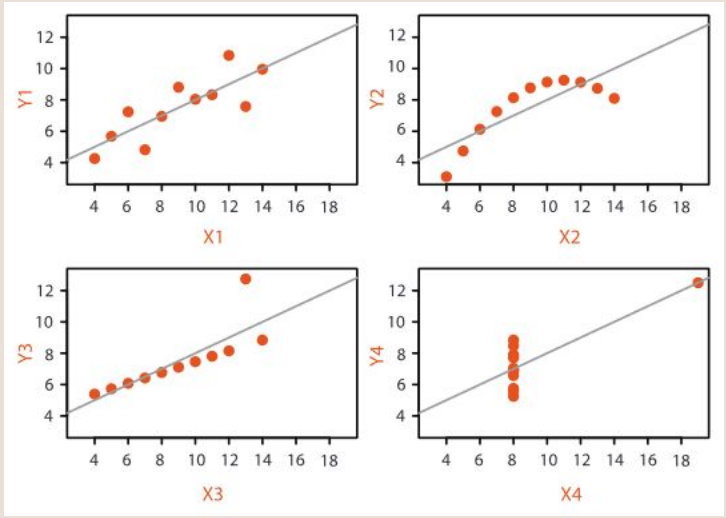
*“Data visualization is an interdisciplinary field that deals with the **graphic representation** of data and information. It is a particularly **efficient way of communicating**.”* —Wikipedia

Why is this important?

- Easier to remember
- Communicate ideas and insights efficiently
- Reveal patterns that would not have been obvious otherwise

Anscombe's Quartet: Raw Data

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



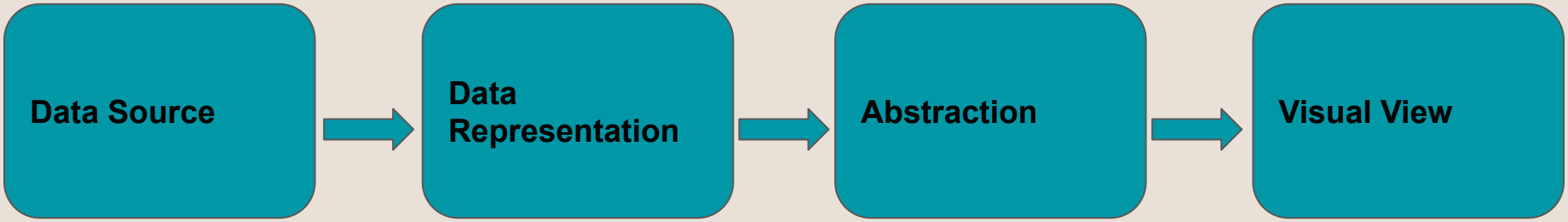
[Munzer, 2014]

Digital Humanities

“Digital humanities is not a unified field but **an array of convergent practices** that **explore** a universe in which print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated.”

“[It] is **qualitative, interpretive, experiential, even emotive.** It immerses the digital toolkit within what represents the very core strength of the Humanities: **complexity**.” — Digital Humanities Manifesto, 2008

Process of Visualization



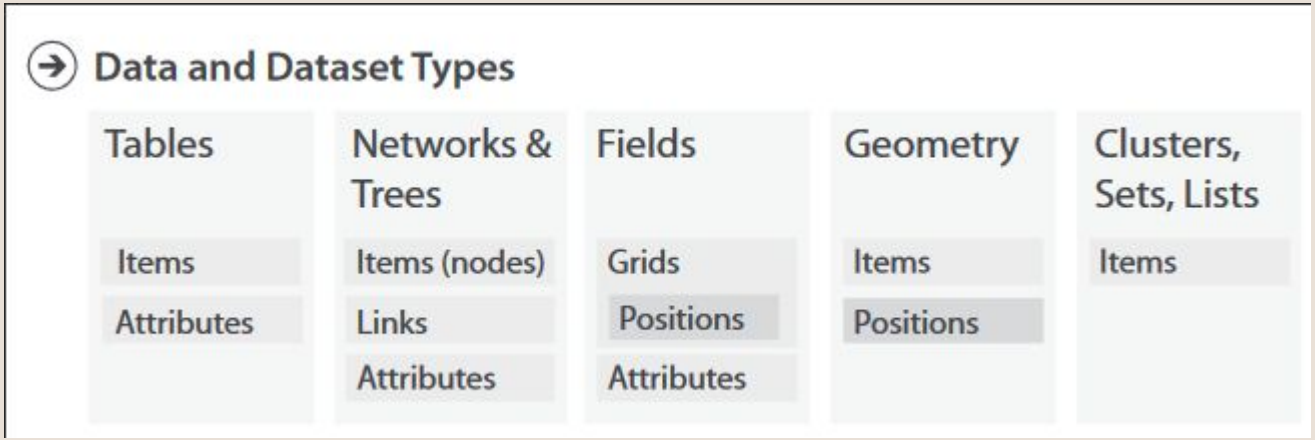
Building Blocks

- Principles
- Abstraction
- Design

Building Blocks

- Principles
- Abstraction
- Design

- What data does the user see?

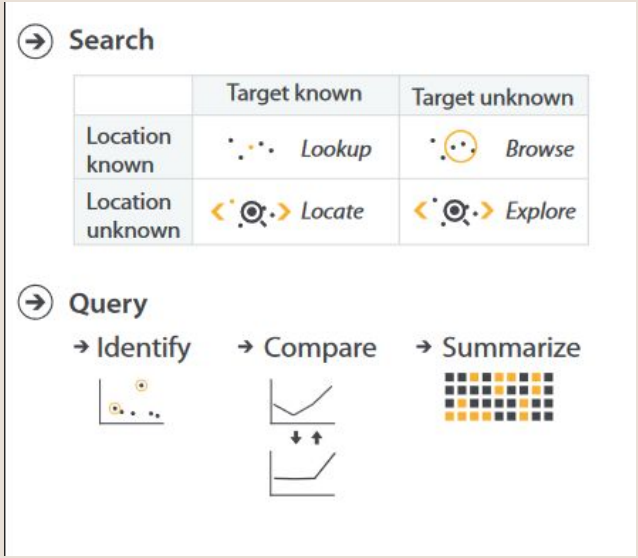


[Munzer, 2014]

Building Blocks

- Principles
- Abstraction
- Design

- What data does the user see?
- Why is the user looking at it?

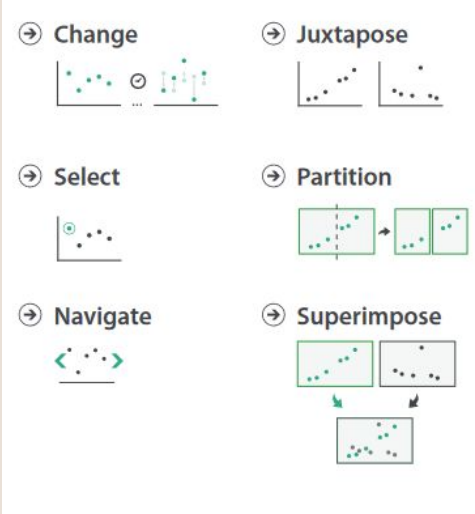


[Munzer, 2014]

Building Blocks

- Principles
- Abstraction
- Design

- What data does the user see?
- Why is the user looking at it?
- How will the user use the vis tool?



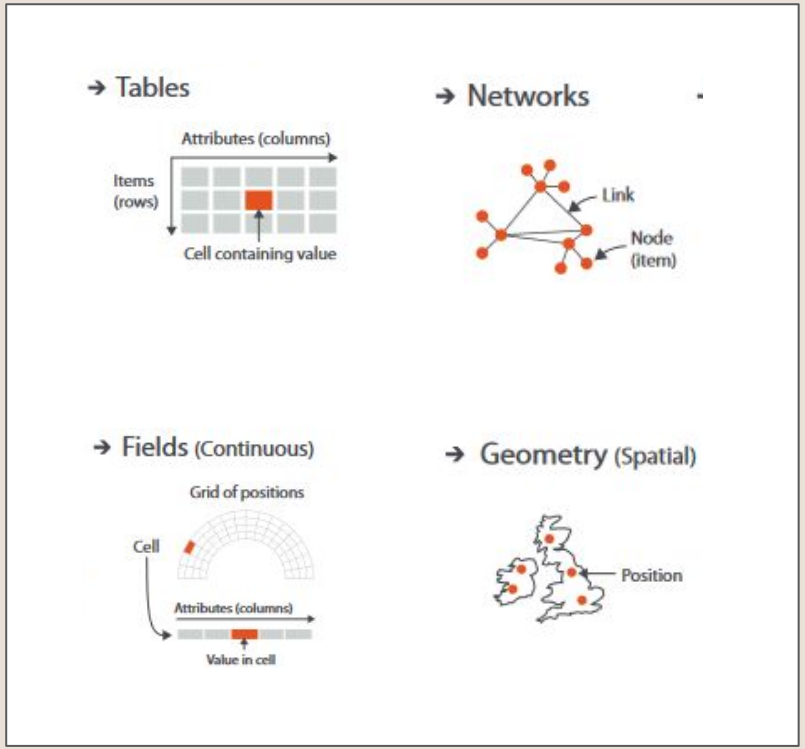
[Munzer, 2014]

The Principles

- Data
- Visual Marks
- Visual Channels
- Interactions

Principles: Data & Datasets

- Data Types
  - Items
  - Attributes
  - Links
  - Position
  - Grids
- Dataset Types
  - Tables
  - Networks & Trees
  - Fields
  - Geometry
  - Clusters, Sets, & List

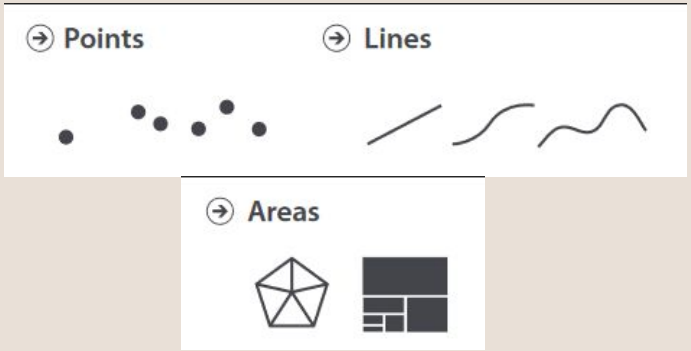


[Munzer, 2014]

Principles: Visual Mark & Visual Channels

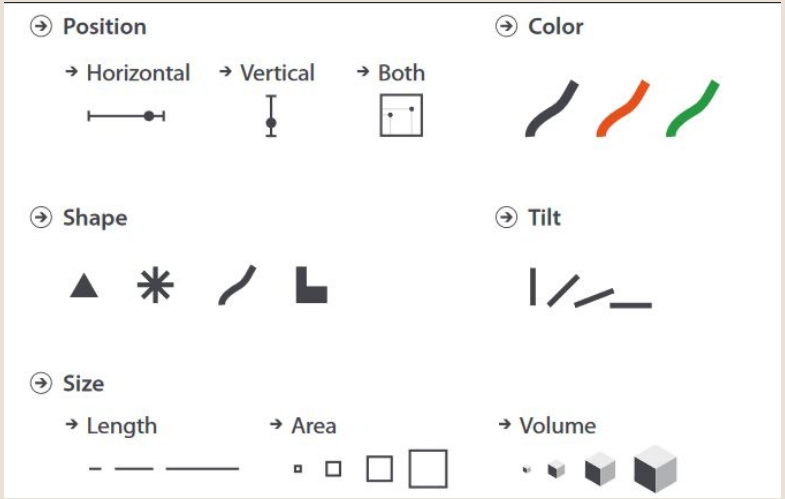
Visual Mark

- A geometrically primitive object



Visual Channel

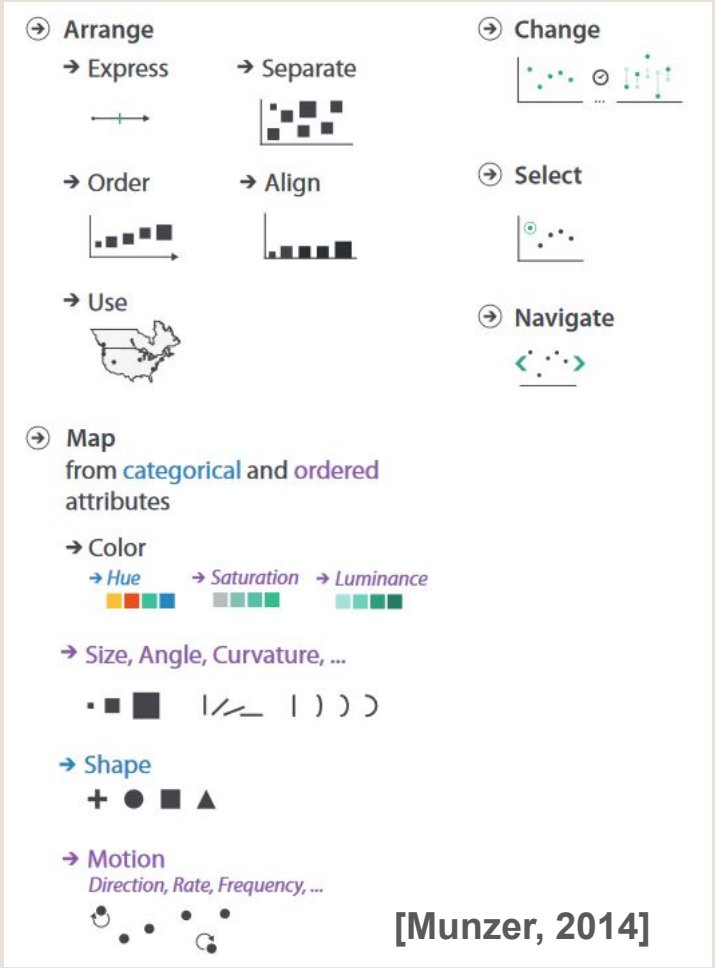
- Affect the appearance of a visual mark



[Munzer, 2014]

Principles: Interaction

- Arrange
- Map
- Change
- Select
- Navigate
- Juxtapose
- Partition
- Superimpose
- Filter
- Aggregate
- Embed



[Munzer, 2014]

Task Abstraction

- Given a **domain specific task**, translate it into **generic visualization task**
  - Domain task: “Are there similar colors between dark arts movement and neoclassicism?”
  - Generic task: Compare two groups.
- Given **domain specific data**, translate it into **generic visualization data**
  - Domain data: #6D8700, #0A6522, #9DB300, #BAC600
  - Generic data: yellowish green, green, yellowish green, yellowish green

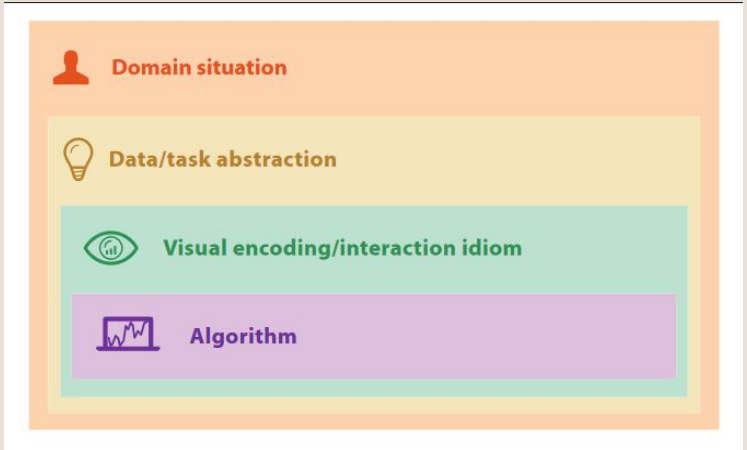
Design

Visual Component

- Charts: bar or pie
- Isosurfaces/slices
- Translucency
- Stereopsis (depth perception)
- Animation

Algorithmic Component

- Color quantization
- Searching algorithms
- Scheduling algorithms



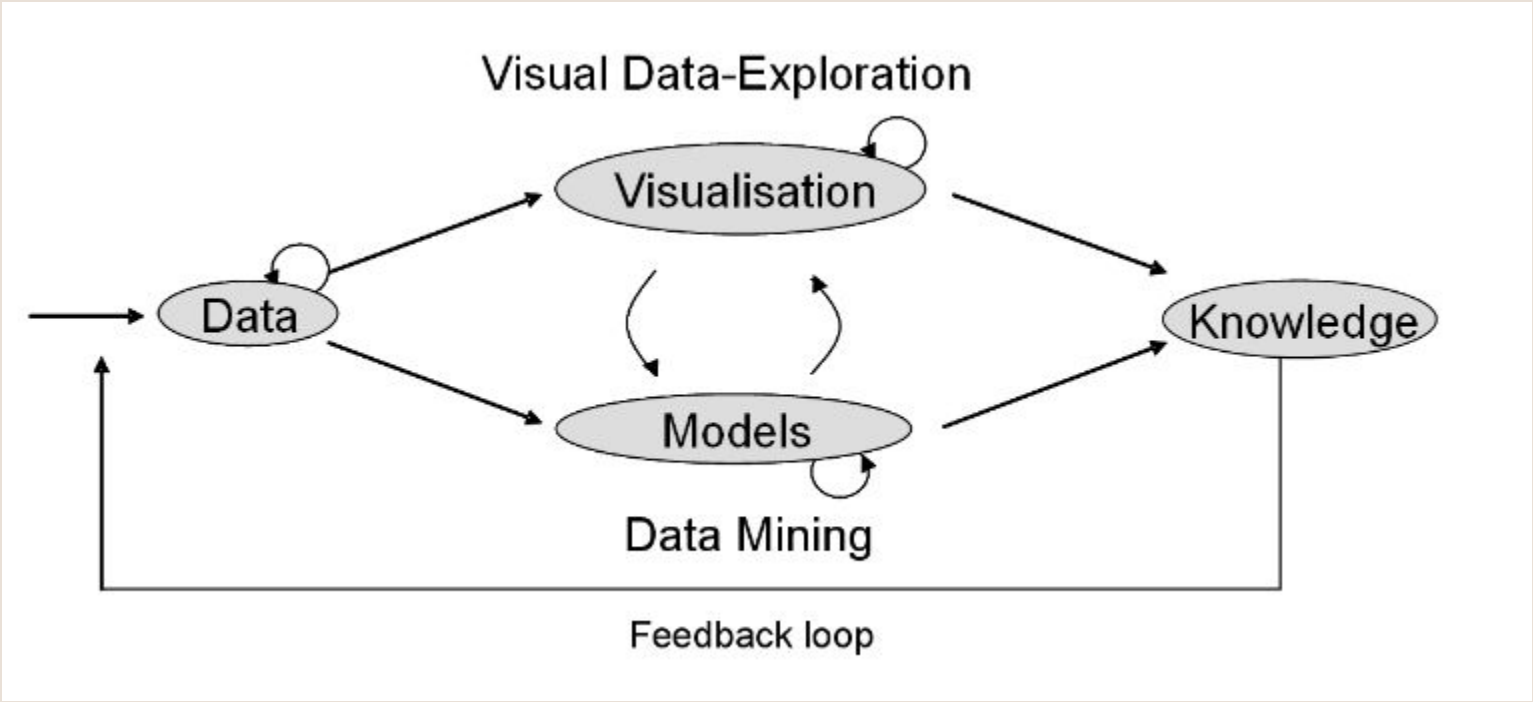
[Munzer, 2014]

Visual Analytics

“The overarching driving vision of visual analytics is to turn the information overload into an opportunity”

“Analyze first - show the important - zoom, filter and analyze further - details on demand” — D. Keim et al, 2008

Process of Visual Analytics



[D. Keim et al, 2008]

Possible Further Reading

- Tamara Munzner. Visualization Analysis and Design. A K Peters Visualization Series, CRC Press, 2014.
- Kräutli, F. (2016). Visualising cultural data. Thesis, Royal College of Art, London.
- S. Srabanti, C. V. de Souza, E. J. da Silva, M. Lage, N. Ferreira, F. Miranda. A Comparative Study of Methods for Visualization of Probability Distributions of Geographical Data. Multimodal Technologies and Interaction 6 (7), 53
-