Visualisation, Data and Tasks

Interactive Data Visualisation
Lecture 1

Dr. Pierre Le Bras

Lecture Summary

- Why do we visualise data and bother with interaction?
- How to abstract data and its semantics
- How to abstract user tasks

Data Visualisation

Data Visualisation

What is your experience with it?

How would you define it?

A Definition

“The representation of information in the form of a chart, diagram, picture, etc.”

Oxford Dictionary

A Definition

“The representation of information in the form of a chart, diagram, picture, etc.”

Oxford Dictionary

*“The **purpose** of visualization is **insight**, not pictures”*

Ben Shneiderman

Another Definition

*"**Computer**-based visualisation systems provide visual **representations of datasets** designed to help **people** carry out **tasks** more effectively."*

Tamara Munzner

An Analysis Cooperation Tool

Data Visualisation enables the analysis of data

It relies on the **cooperation** between **humans** and **machines**

- A machine can compute and draw large datasets
- A human can make decisions, analyse and interpret patterns

Why Visualisation?

	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								
Variance								
Correlation								

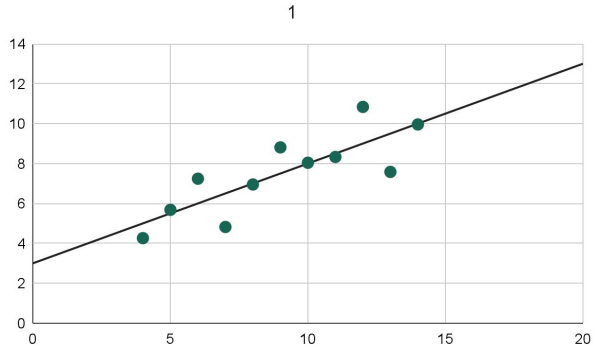
Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, *27*(1), 17-21.

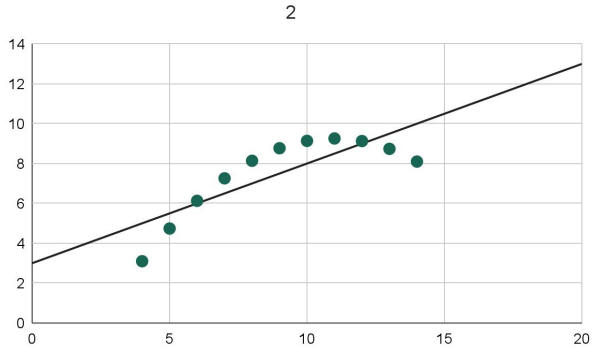
Why Visualisation?

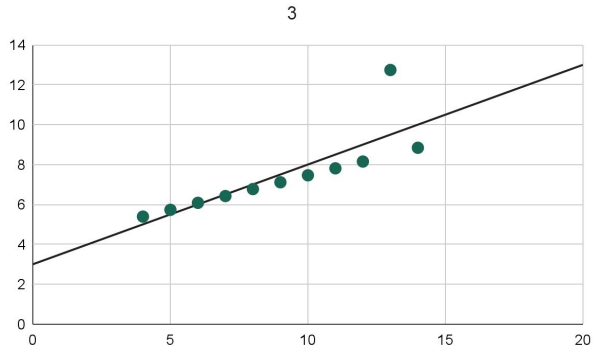
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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**	

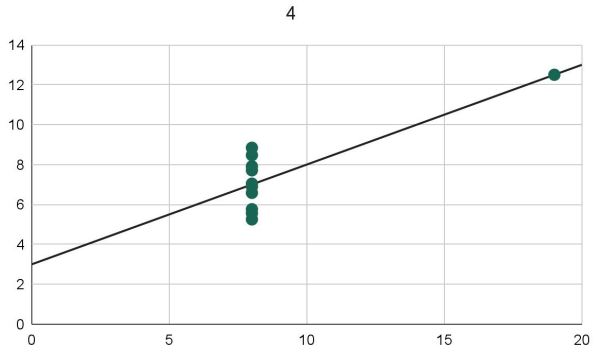
Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, *27*(1), 17-21.

Why Visualisation?









Anscombe, F. J. (1973). Graphs in statistical analysis. *The american statistician*, *27*(1), 17-21.

Why Visualisation?

The analysis of patterns sometimes requires the whole data

- Statistical summaries tend to loose details and important features

To facilitate human analysis, we must preserve cognitive efforts

- **External representations** allow the offload of working memory

Vision maximises perception

- Other senses have technological limitations (touch, smell, taste, …)
- Data sonification is possible, but distinguishing sounds is less accurate
- Vision is a **high-bandwidth channel**, with significant processing capabilities
- There are still accessibility concerns (e.g., Kim et al., 2021)

Kim, N. W., Joyner, S. C., Riegelhuth, A., & Kim, Y. (2021). Accessible visualization: Design space, opportunities, and challenges. In Computer Graphics Forum (Vol. 40, No. 3, pp. 173-188).

Why Interactivity?

Humans, computer displays and static views have limited capabilities

- Memory and attention (humans)
- Screen estate (displays)

Large datasets cannot be presented so easily

Why Interactivity?

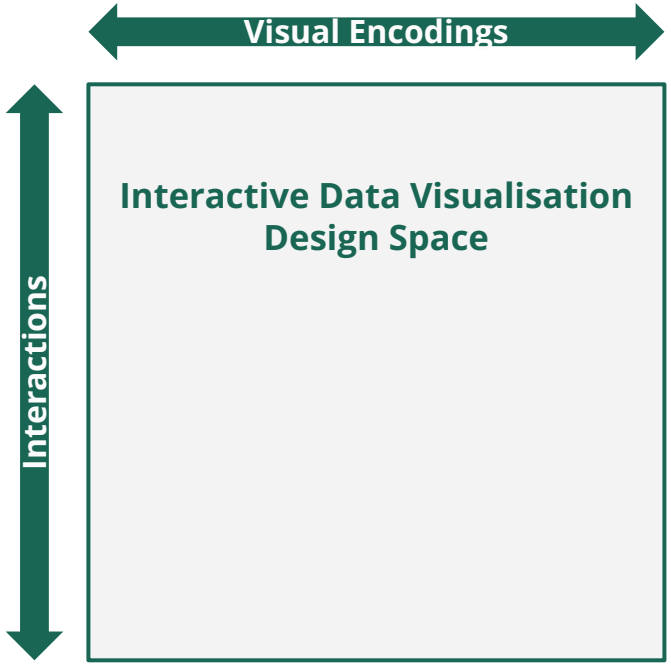
Interactivity enables the user to **adapt** the view to their analysis' needs

- Query different data subsets
- Change between levels of abstractions
- Find connections between parts of the dataset

Design Space

There are many **Visual Encoding** possibilities

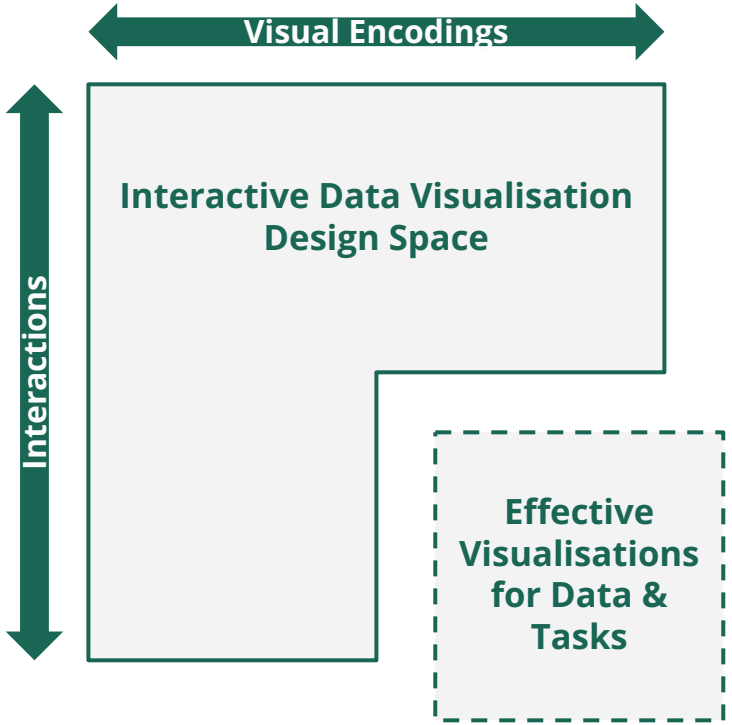
Adding **Interaction**, this makes a huge Design Space



Design Space

How we construct interactive data visualisations needs to be analysed with:

- **What** is shown: Data
- **Why** the visualisation is needed: Task



Data & Semantics

What type of data do you know of?

What semantics can you attach to it?

Data and Semantics

Understanding what is the data we are trying to represent will limit the possibilities for visualisation

The type of data and its semantics drive our choice for its visual arrangement

Positioning points *based on* ***geographical coordinates***

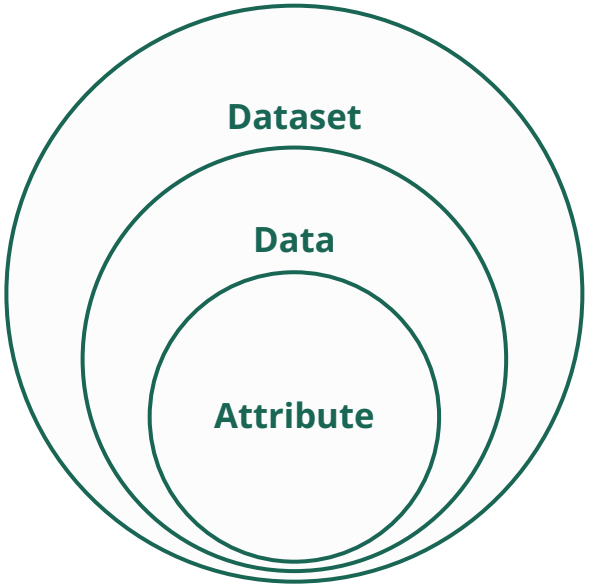
vs.

Sizing areas *based on* ***categorical proportions***

Types of Data

We can define types of data at three levels

- **Attribute** level
- **Data** level
- **Dataset** level



Attribute Types

There are typically 4 types of attributes:

Nominal - Qualitative

- Categorical - categories
- Ordinal - ordered categories

Numerical - Quantitative

- Interval - numbers
- Ratio - numbers with a true 0

Attribute Types

Focusing on visualisation design, we can see attribute types differently:

- Categorical: different categories, with **no order**
- Ordered:
  - Ordinal: **ordered categories**, with no arithmetic
  - Numerical: measures that **support arithmetic**
  - Order types: sequential, diverging, cyclic

Data Types

- **Attribute** - measured property
- **Item** - individual entity
  - One item is characterised by a set of attributes
- **Link** - relation between items
  - Can be directed or undirected
  - One link is also characterised by a set of attributes

- For *spatial* visualisations
  - Grid: sampling strategy in continuous space
  - Position: explicit location in 2D or 3D space

Dataset Types

- **Tables** - items as rows, attributes as columns
  - Multidimensional tables
- **Networks** (graphs) & **Trees** - items (nodes) connected by links (edges)
- Clusters / Sets / Lists - collections of items

- For *spatial* visualisations:
  - Fields - continuous domain
  - Geometry - shape of spatial regions

Semantics

Attribute and data types will drive design decision...

... but so will semantics!

Semantics is the **meaning** we attach to these attributes and data

Semantics

An attributes could be a **key** or a **value**

- Keys are used to uniquely identify an item
- Multidimensional tables require 2 or more keys
- Keys can sometimes be made of several attributes

Semantics

Categorical attributes can have a **hierarchical** relationship

Quantitative attributes can be a **measure**, a **moment** or a **position**

A link can signify **similarity**, **hierarchy** or **dependency**

...

User Tasks

Why do you use Data Visualisations?

User Tasks

Thinking about why a user needs an interactive visualisation drastically changes the way we design it

A tool designed and effective for one task, may not be suitable for another

Comparing correlations *between multiple dimensions*

vs.

Understanding *the relative* ***proportion*** *of one category in the dataset*

Actions and Targets

There are two aspects to tasks:

- **Actions**: what the user does
  e.g., compare, explore, present
- **Targets**: which aspect of the data interests the user
  e.g., trends, similarity, features

Action: Analyse

To **inspect** a visualisation in order to **interpret** and **explain** the data

Discover - find new knowledge
Present - communicate information

Annotate - add new temporary information
Record - extract persistent artifacts
Derive - create new or transformed data

Action: Search

To **find** elements of interest (targets) in the visualisation

	Location known	*Location unknown*
Target known	**Lookup**	**Locate**
Target unknown	**Browse**	**Explore**

Action: Query

Once found, **examine** the targets to:

Identify the attributes of a single target

Compare attributes between two targets

Summarise the attributes of all targets

Targets: Dataset Level

Targets can be over the whole dataset, on a high-level

- Trends or patterns in the dataset
- Outlier data points
- Domain-specific features

Targets: Attribute Level

Targets can be specific attributes

Distribution and extremes of one attribute

Correlation and similarity between multiple attributes

Data-specific structures

- Network topology
- Spatial shapes

Conclusion

Key Points

- Data visualisation enables the **cooperation** between humans and computers to **analyse** datasets
- Interactivity bypasses limitations and facilitates the analysis of large datasets
- Designing visualisations and interactions needs to take into account:
  - The **data**: types of attributes, data and dataset with their semantics
  - The **user tasks**: why a user might want to interact with the visualisation