Introduction to Visualisation

VU Visual Data Science
Johanna Schmidt

WS 2018/19
Definition

- Visualisation enables **insight into data**
- **Visual information** for humans
  - Easier to interpret
  - Allows faster analysis of the data

*The purpose of computing is insight, not numbers.*

[R. Hamming, 1962]
Definition

• Visualisation rather old discipline
• Comprises intuitive illustrations (e.g., pie/bar charts, scatter plots),
• But nowadays also more complex applications for data analysis
Visualisation disciplines

• Depending on the type of data that should be visualised
  • **Spatial data** -> Scientific Visualisation
  • **Abstract data** -> Information Visualisation
Visualisation disciplines

- Depending on the type of data that should be visualised:
  - Spatial data - Scientific Visualisation
  - Abstract data - Information Visualisation
Visualisation disciplines

• Depending on the type of data that should be visualised:
  • **Spatial data** -> Scientific Visualisation
  • **Abstract data** -> Information Visualisation
Visualisation disciplines

• **Scientific Visualisation**
  • Volume visualisation
  • Flow visualisation

• **Information visualisation**

3D

nD
Visualisation disciplines

• Nowadays, borders are **not that well defined** any more
  • Information visualisation may also comprise spatial data (e.g., geographic data)
  • Need for integration of abstract data in spatial data
Visualisation as a dynamic field

• Visualisation always driven by **data and tasks**
  • Many different domains (medicine, biology, archeology, astronomy, business analysis, ...)
  • Different types of datasets

• Uses knowledge from **different fields**
  • Human perception
  • Color theory
  • Geometry, morphology
Visualisation in computer science

• Area of **computer graphics**
  • Generation of visual representations from data
  • Usage of rendering techniques

• Strong need for **data analysis**
  • Computer-generated or -supported analysis

• **Data management**
  • Storage
  • Processing
Why Visualisation?

• **Human vision** provides high bandwidth which can be used
• Data getting increasingly **complex** (size / parameters)
• **Statistical analysis** alone may not transport the full picture
Why Visualisation?

• **Anscombe’s Quartet**
  • Developed in 1973 by the statistician Francis Anscombe
  • Demonstration to show strength of **visual data representation**

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>10,0</td>
<td>y</td>
<td>8,04</td>
<td></td>
<td>10,0</td>
<td>y</td>
<td>7,46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8,0</td>
<td></td>
<td>6,95</td>
<td></td>
<td>8,0</td>
<td></td>
<td>6,77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13,0</td>
<td></td>
<td>7,58</td>
<td></td>
<td>13,0</td>
<td></td>
<td>12,74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9,0</td>
<td></td>
<td>8,81</td>
<td></td>
<td>9,0</td>
<td></td>
<td>7,11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11,0</td>
<td></td>
<td>8,33</td>
<td></td>
<td>11,0</td>
<td></td>
<td>7,81</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14,0</td>
<td></td>
<td>9,96</td>
<td></td>
<td>14,0</td>
<td></td>
<td>8,84</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6,0</td>
<td></td>
<td>7,24</td>
<td></td>
<td>6,0</td>
<td></td>
<td>5,08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,0</td>
<td></td>
<td>4,26</td>
<td></td>
<td>4,0</td>
<td></td>
<td>5,30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12,0</td>
<td></td>
<td>10,84</td>
<td></td>
<td>12,0</td>
<td></td>
<td>8,15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7,0</td>
<td></td>
<td>4,82</td>
<td></td>
<td>7,0</td>
<td></td>
<td>6,42</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5,0</td>
<td></td>
<td>5,58</td>
<td></td>
<td>5,0</td>
<td></td>
<td>5,73</td>
<td></td>
</tr>
</tbody>
</table>

[6]
Why Visualisation?

- **Anscombe’s Quartet**
  - Four groups of numbers have identical statistical parameters

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of $x$</td>
<td>9</td>
<td>exact</td>
</tr>
<tr>
<td>Sample variance of $x$</td>
<td>11</td>
<td>exact</td>
</tr>
<tr>
<td>Mean of $y$</td>
<td>7.50</td>
<td>to 2 decimal places</td>
</tr>
<tr>
<td>Sample variance of $y$</td>
<td>4.125</td>
<td>±0.003</td>
</tr>
<tr>
<td>Correlation between $x$ and $y$</td>
<td>0.816</td>
<td>to 3 decimal places</td>
</tr>
<tr>
<td>Linear regression line</td>
<td>$y = 3.00 + 0.500x$</td>
<td>to 2 and 3 decimal places, respectively</td>
</tr>
<tr>
<td>Coefficient of determination of the linear regression</td>
<td>0.67</td>
<td>to 2 decimal places</td>
</tr>
</tbody>
</table>

### Anscombe’s Quartet Data

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>$y$</td>
<td>8.04</td>
<td>8.04</td>
<td>7.46</td>
<td>8.0</td>
</tr>
<tr>
<td>$x$</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
<td>8.0</td>
</tr>
<tr>
<td>$y$</td>
<td>6.95</td>
<td>6.95</td>
<td>6.77</td>
<td>6.58</td>
</tr>
<tr>
<td>$x$</td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
<td>13.0</td>
</tr>
<tr>
<td>$y$</td>
<td>7.58</td>
<td>7.58</td>
<td>12.74</td>
<td>7.71</td>
</tr>
<tr>
<td>$x$</td>
<td>9.0</td>
<td>9.0</td>
<td>9.0</td>
<td>9.0</td>
</tr>
<tr>
<td>$y$</td>
<td>8.81</td>
<td>8.81</td>
<td>7.11</td>
<td>8.84</td>
</tr>
<tr>
<td>$x$</td>
<td>11.0</td>
<td>11.0</td>
<td>11.0</td>
<td>11.0</td>
</tr>
<tr>
<td>$y$</td>
<td>8.33</td>
<td>8.33</td>
<td>7.81</td>
<td>8.47</td>
</tr>
<tr>
<td>$x$</td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
<td>14.0</td>
</tr>
<tr>
<td>$y$</td>
<td>9.96</td>
<td>9.96</td>
<td>8.84</td>
<td>7.04</td>
</tr>
<tr>
<td>$x$</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>$y$</td>
<td>7.24</td>
<td>7.24</td>
<td>5.08</td>
<td>5.25</td>
</tr>
<tr>
<td>$x$</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>$y$</td>
<td>4.26</td>
<td>4.26</td>
<td>5.30</td>
<td>10.0</td>
</tr>
<tr>
<td>$x$</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td>$y$</td>
<td>10.84</td>
<td>10.84</td>
<td>8.15</td>
<td>8.62</td>
</tr>
<tr>
<td>$x$</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>$y$</td>
<td>4.82</td>
<td>4.82</td>
<td>6.42</td>
<td>7.91</td>
</tr>
<tr>
<td>$x$</td>
<td>6.0</td>
<td>6.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>$y$</td>
<td>5.58</td>
<td>5.58</td>
<td>5.73</td>
<td>8.89</td>
</tr>
</tbody>
</table>
Why Visualisation?

• Anscombe’s Quartet
  • Visual representation of the data quite different
[Diagram showing scatter plots and data table]
Lecture Focus

• Information Visualisation
Lecture Focus

• Information Visualisation

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition

[Card et al., 1999]
Visualisation Use Cases

• Exploration
• Confirmation
• Presentation
Visualisation Use Cases

- **Exploration**
  - Searching and analysis
  - No or only minor *prior knowledge* about the data
  - Find *potentially useful* information

- **Confirmation**

- **Presentation**
Visualisation Use Cases

- Exploration
- **Confirmation**
  - Goal-oriented
  - Examination of a *prior defined* hypothesis
  - More prior knowledge of the data
- Presentation
Visualisation Use Cases

- Exploration
- Confirmation
- **Presentation**
  - Efficient *communication* of data features and findings
  - Clear definition of what to show
  - Often targeted towards *external people*
Visualisation Use Cases

• **Exploration**
  • Highly *interactive*
  • Need to visualize many **different aspects** of the data
  • **Challenges** for data storage/processing and rendering (visual clutter)

• Confirmation

• Presentation
Visualisation Use Cases

• Exploration

• **Confirmation**
  • Only parts of the data are needed, data can be **prepared** (e.g., filtered)
  • Visualisation **targeted** towards goal
  • **Interactive**

• Presentation
Visualisation Use Cases

• Exploration
• Confirmation

• **Presentation**
  • Only parts of the data are needed, data can be prepared (e.g., filtered)
  • **Simple/intuitive** visualisation techniques
  • No/minor interaction
Visualisation Techniques

• Scatter Plots
• Bubble Charts
• Parallel Coordinates
• Radar Charts
• Box Plots
• Violin Plots
• Venn Diagrams
• Node-Link Diagrams
Visualisation Techniques

• **Scatter Plots**
  • Bivariate data (2 dimensions)
  • Visual channel: position
  • Intuitive, easy to spot outlier/cluster, correlations, and to identify distributions
  • More dimensions may be added by using additional visual channels
  • For **Exploration, Confirmation**, and **Presentation**
Visualisation Techniques

• **Bubble Charts**
  • Scatter plot, but additional dimensions can be shown
  • Uses size for additional attribute -> 3 dimensions
  • Color used for categorical attribute
  • For **Exploration, Confirmation,** and **Presentation**
Multivariate Data

• Data in which analysis is based on **more than two variables**
• Analysis always needs to take **several dimensions** into account
• Usually variables have **different domains** (e.g., numbers, categories)
• **Challenges** for analysis and visualisation
Visualisation Techniques

• Parallel Coordinates
  • Invented probably 1885, but got popular in 70s (Alfred Inselberg)
  • Align data dimensions as vertical axes
  • Axes need to be scaled accordingly
Visualisation Techniques

• **Parallel Coordinates**
  • Can be used to identify statistical parameters:
    • **Parallel lines**: positive correlation
    • **X-shaped lines**: negative correlation
    • **Random**: no correlation
  • **Axes order** very important to spot correlations
Visualisation Techniques

- **Parallel Coordinates**
  - Splines instead of lines
Visualisation Techniques

• Parallel Coordinates
  • Examples
Visualisation Techniques

• Parallel Coordinates

Examples
Visualisation Techniques

• Parallel Coordinates
  • Examples
Visualisation Techniques

- Parallel Coordinates

Examples

[15]
Visualisation Techniques

• Parallel Coordinates
  • Examples
Votes in Donald Trump and Social data by state - Parallel coordinates
Visualisation Techniques

• Parallel Coordinates
  • Drawbacks
    • Axes ordering
Visualisation Techniques

- Parallel Coordinates
  - Drawbacks
    - Axes ordering
    - Overplotting
Visualisation Techniques

• **Parallel Coordinates**
  • Overplotting
    • Filtering
Visualisation Techniques

- **Parallel Coordinates**
  - Overplotting
    - Filtering
    - Clustering
Visualisation Techniques

- **Parallel Coordinates**
  - Overplotting
    - Filtering
    - Clustering
    - Brushing
Visualisation Techniques

• **Parallel Coordinates**
  • Visualisation technique for multivariate data
  • For *Exploration* and *Confirmation*
Visualisation Techniques

• **Radar Charts**
  - *Circular* alignment of axes
  - *Axes* need to be scaled
  - Harder to spot **correlations**
  - Rather used for **comparisons**
Visualisation Techniques

• **Radar Charts**
  • Possible to use different representations
Visualisation Techniques

• Radar Charts
  • Examples
Visualisation Techniques

• **Radar Charts**
  • Examples
Visualisation Techniques

• **Radar Charts**
  • **Interpretation** may be difficult, because of radial distortion
  • **Axes ordering** important
  • Should not be used for **linear** data, axes should be **independent**
  • For **Exploration, Confirmation, and Presentation**
Visualisation Techniques

• **Box Plots**
  
  • Visualisation of *statistical parameters*
    
    • Box: 50% of the data
    
    • Whisker: data area
    
    • Line: median
    
    • Circles: outlier
  
• For *(Exploration,)* Confirmation, and *Presentation*
Visualisation Techniques

• **Violin Plots**
  • Visualisation of **statistical parameters**
  • For (**Exploration,**) **Confirmation, and Presentation**
Visualisation Techniques

- **Venn Diagrams**
  - Visualisation of **categorical data**
  - Especially useful for **overlapping sets**
  - For (Exploration,) **Confirmation, and Presentation**

![Venn Diagram Example](image)
Visualisation Techniques

• **Node-Link Diagrams**
  • Visualisation of network data
    • Nodes: elements
    • Links: relations between the elements
  • For **Exploration** and **Confirmation**
Visualisation Techniques

• **Node-Link Diagrams**
  • Overplotting
Further Reading

• InfoVis Community
  • https://infovis-wiki.net/wiki/Main_Page

• Data Visualisation Catalogue
  • https://datavizcatalogue.com/

• VO & UE Informationsvisualisierung
  • https://www.cg.tuwien.ac.at/courses/InfoVis/

• VO & UE Informationsvisualisierung
  • https://tiss.tuwien.ac.at/course/courseDetails.xhtml?dswid=1832&dsrid=65&courseNr=188305&semester=2018W
Hans Rosling Ted Talk

https://www.youtube.com/watch?v=Z8t4k0Q8e8Y
Visualisation principles

• Interaction
• Usage of color
• Usage of shapes
Visualisation principles

- **Interaction**
- Usage of color
- Usage of shapes
Interaction

• Especially for Exploration and Confirmation, *interactive exploration* of the data is needed

*Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface*  
[Wong and Thomas 2004]
Interaction
Interaction

Overview first, details on demand
[Shneiderman, 1986]

• **Overview** needed, before being able to explore the data
• Exploration can then lead to **details**
Interaction

- Creating an **overview** can be challenging due to data size
  - Big Data
  - Data storage/processing
  - Overplotting
  - Limited screen space
  - Needs for aggregation
Interaction

• **Details** by interacting with the data
  • Selection techniques
  • Filtering algorithms
  • Data processing
Interaction

- **Details** by interacting with the data
  - Selection techniques
  - Filtering algorithms
  - Data processing

- **Focus+context**
  - While exploring details, context should be preserved
Interaction

• Details by interacting with the data
• Selection techniques
• Filtering algorithms
• Data processing

Focus+context

While exploring details, context should be preserved
Interaction

- **Details** by interacting with the data
  - Selection techniques
  - Filtering algorithms
  - Data processing

- **Focus+context**
  - While exploring details, context should be preserved

- **Linking and brushing**
  - Apply filtering to all views
References

[5] https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcShP_eFMGkGOhqtnwPUBLcOWAp9WpDdMVQaNNuU4d4Pc5x3IDQ
[14] https://bl.ocks.org/jasondavies/1341281
[16] https://socviz.co/
References

[22] https://docs.telerik.com/devtools/aspnet-ajax/controls/htmchart/chart-types/radar-chart
[25] https://de.wikipedia.org/wiki/Box-Plot
[26] https://datavizcatalogue.com/methods/violin_plot.html
[27] https://datavizcatalogue.com/methods/venn_diagram.html
[28] https://www.oreilly.com/library/view/learning-neo4j-3x/9781786466143/assets/a6c1bfc4d-06a-d066-49dd-a262-639f050bd075.png
[29] https://periscopic.com/content/2017/03/
[31] https://de.wikipedia.org/wiki/Fluggesellschaft
[33] https://hal-lirmm.ccsd.cnrs.fr/lirmm-01592614/document
[34] https://ieeexplore.ieee.org/document/1532821