Data Visualisation

CMP020L013A

Week 8: Visual Analytics & Decision-Making

Mohammad Javaheri
(Dr Mohammad Ali Javaheri Javid)

Agenda

- Visual Analytics
- Using a dataset to ask the right questions
- Using Visualisation to answer questions

Visual Analytics

- "The science of analytical reasoning facilitated by interactive visual interfaces",
  - from Illuminating the Path – the Research and Development Agenda for Visual Analytics, J. Thomas and K. Cook (eds.), 2005
- More than information visualisation or visual data mining, it involves technology to support all aspects of the analysis and reasoning processes.

Data

- Data are collections of facts and information.
- Primary:
  - Data that have been collected from an original source for a specific purpose.
- Secondary:
  - Data that are not originally collected for a specific purpose, such as domain knowledge or literature.

What is Decision-Making?

- Only after collecting data (information), we can make a decision.
- When we make a decision, we use the collected data/information as evidence to make an informed decision.
- Data-driven decision-making is the practice of making decisions based on data analysis rather than intuition, guesswork, or estimation.
- Visualisation helps decision-makers understand the significance of data by placing them in a visual context
- It enables decision-makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns.

Analytical Approaches

- Explore
  - attempt to develop initial rough description or understanding
- Describe
  - provide a detailed account or precise measurement and reporting of the characteristics of some population, group or phenomenon
- Explain
  - establish the elements, factors or mechanisms that are responsible for producing the state of, or regularities in, a social phenomenon

Analytical Approaches

- Understand
  - establish reasons for particular social action, the occurrence of an event or the course of a social episode
- Predict
  - use some established understanding or explanation of a phenomenon to postulate certain outcomes under particular conditions

Asking The Right Questions

- Research questions
- Focus the researcher's attention
- Influence the scope and depth of the research
- Point towards particular strategies and methods of data collection and analysis

Three Types of Research Questions

- What
  - questions seek descriptions
- Why
  - questions seek explanations (reason)
- How
  - questions seek intervention for change (process)

Sample questions in the context of a dataset

- What variables does the dataset contain?
- How are they distributed?
- Are there any notable data quality issues?
- Are there any surprising relationships among the variables?

Example: Recycling

- Topic
  - recycling behaviour
- Question
  - what is the extent of recycling behaviour among university students?
- Motivation
  - reveal demographic trends or attitudes towards recycling, encourage more recycling

Information and Relationships

Ratio between highest and lowest paid employee	**Money** (salaries high/low)
Units of a product sold per region	**Sales** (frequency count) related to **geography**
Percentage of CO_2 emissions caused by cars	**Pollution** (proportion) related to **mode of transport**
Healthcare spending per region per quarter	**Money** (total cost) related to **geography** and **time**
Typical price of a meal in a 5-star rated restaurant in London	**Money** (median price) related to **ratings** and **geography**

Data in Quantitative Stories

- Numerical variables
  - measurable quantities
    - e.g. frequency counts, height, profit, cost, speed
  - relationships between numerical quantities
    - e.g. tips and total bill
- Categorical variables
  - divide information into useful groups or *factors*
    - e.g. geographical locations, companies, months
  - relationships between categorical and numerical
    - e.g. cost of a Big Mac across countries

Data in Quantitative Stories

- Quantitative stories always feature relationships
  - e.g. simple associations between quantitative and categorical variables
    - *number of dog versus cat owners*
- More complex associations among multiple sets of quantitative and/or categorical variables
    - *average income and life expectancy per country*

Display Quantitative Information

- Table or graph? colour? x and y axis?
- What is the story?
- Quantitative stories are always about relationships
- Turn numbers into stories using their relationships

Variables and Concepts

- Data consists of variables
  - quantitative
  - categorical
- Concept
  - an abstract idea in the domain of enquiry
- Variables derive from concepts

Concepts → Variables → Data

- 1. Define all relevant concepts in the study
- education
  - participation in a programme of learning provided by a recognised teaching institutions, typically associated with receiving formal qualifications
- age
  - years since birth
- gender
  - socially constructed categories of identity including female, male, trans, non-binary…
- use standard meanings in the field whenever possible
- simplifies relating results from different studies

Visual Analytics Questions

- Visual Coding
  - Representing the questions and answers visually
  - Converting the finding to graphs
- Missing data
  - how to deal with NAs?
  - Context knowledge
- What questions or problems are trying to be addressed?
- What do you need to know?

Visual Analytics Questions

- There are many questions that a dataset can answer
- Choose one initial question and write it down.
- As new questions emerge, continue to write these down, and only after you've answered the initial question should you come back to the "new" questions
- Do I have enough data?
- Do I trust the data that I have?

Visual Analytics Questions

- Who collected these data?
- How is it collected?
- When is it collected?
- Communicate (clarification and justification):
  - how much data or what kind of data were used
  - how you arrived at an answer, and
  - what were the limitations of the data that were used
  - how the data is used to answer questions or solve your problem (method).