G53FIV: Fundamentals of Information Visualization

Lecture 14: Review

Ke Zhou
School of Computer Science
Ke.Zhou@nottingham.ac.uk

https://moodle.nottingham.ac.uk/course/view.php?id=68644

A Bit About the Exam

What to Learn

- Lecture slides
- Selected chapters from the core text books and additional reading papers (all available on Moodle)
  - The Visual Display of Quantitative Information (2nd Edition). E. Tufte. Graphics Press, 2001 [available in the library].
  - R Graphics Cookbook, Winston Chang, O'Reilly Media, 2013 [you can find it online by googling].
  - Paper handouts.

Lecture Schedule

Week	Topic	Topic
1 (w19)	Introduction	The Value of Visualization
2 (w20)	Data and Image Models	Graphs and Charts
3 (w21)	Multivariate Data Visualization	Visualization with R - Fundamentals
4 (w22)	Visualization with R - Advanced	Visualization Tools and Visual Perception
5 (w24)	Interaction	Evaluation
6 (w25)	Visualizing Text and Documents	Visualizing Time Series, Trees and Graphs
7 (w26)	Recap of Fundamentals	Review
Break		
8 (w33)	Demo	Demo

Exam Format

- Two hours
- The written exam accounts for 75% of the whole module
- 4 questions relating to different aspects of information visualization (learned within the module)
  - Different sub-questions
- Bring your pen

Question Types

- Examples
  - Describe the definition or the key concept
  - Compare and assess different information visualizations
  - How to manipulate data
  - And others…

Past Paper

- The past paper is available on Moodle.
- You should practice with it after the revision.

Practice Questions

Describe a Key Concept

- Describe the three basic data types. Assess each column of the table on the corresponding data type.

	Student 1	Student 2	Student 3	Student 4
Name	Tom	Jim	Mary	Jane
Age	20	19	22	21
Grade	A	B	A-	B+
Course	Math	Math	Art	Sport
Entry Year	1997	1998	1995	1996

Nominal, Ordinal & Quantitative

- N - Nominal (labels or categories)
  - Operations: =, ≠

  e.g. math, art (course)

- O – Ordered
  - Operations: =, ≠, <, >

  e.g. A, A-, B+, B (grade)

- Q - Interval (location of zero arbitrary)
  - Operations: =, ≠, <, >, -
  - Can measure distances or spans

  e.g. (3.23, -1.2) (GPS)

- Q - Ratio (zero fixed)
  - Operations: =, ≠, <, >, -, %
  - Can measure ratios or proportions

  e.g. 20, 19, 22, 21 (age)

Expected Answer

- There are three basic data types: nominal (N), ordinal (O) and quantitative (Q).
- With respect to the data in the table, each row represents a data case. The column “name” denotes nominal (N) data; “age” represents quantitative (Q) data; “grade” denotes ordinal (O) data; “course” represents nominal (N) data; and “entry year” denotes quantitative (Q) data.

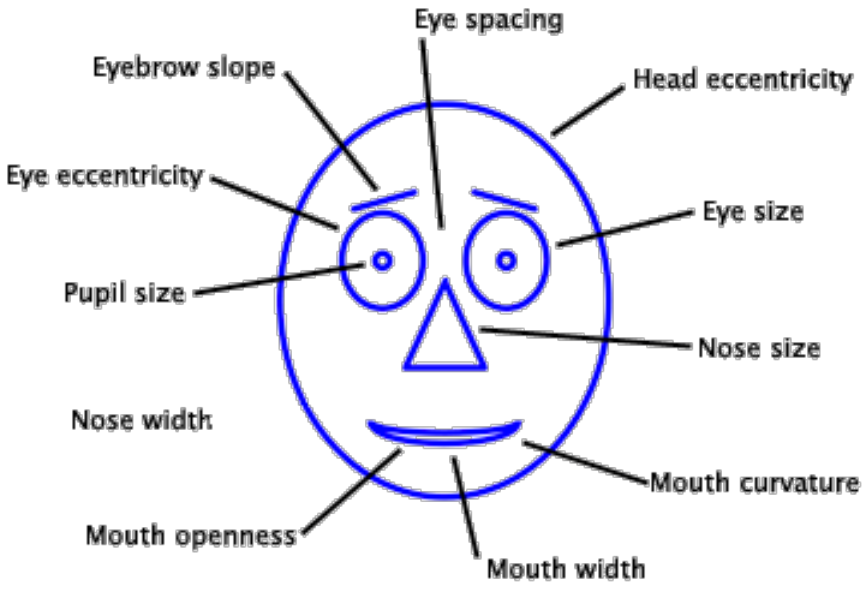
	Student 1	Student 2	Student 3	Student 4
Name (N)	Tom	Jim	Mary	Jane
Age (Q)	20	19	22	21
Grade (O)	A	B	A-	B+
Course (N)	Math	Math	Art	Sport
Entry Year (Q)	1997	1998	1995	1996

Compare Different Visualizations

- Compare and contrast two common techniques for visualizing multivariate data: Chernoff Faces and Parallel coordinates.
  - Explain Chernoff Faces and Parallel coordinates.
  - Identify the strengths and weaknesses in terms of "Find value of data case"

Chernoff Faces

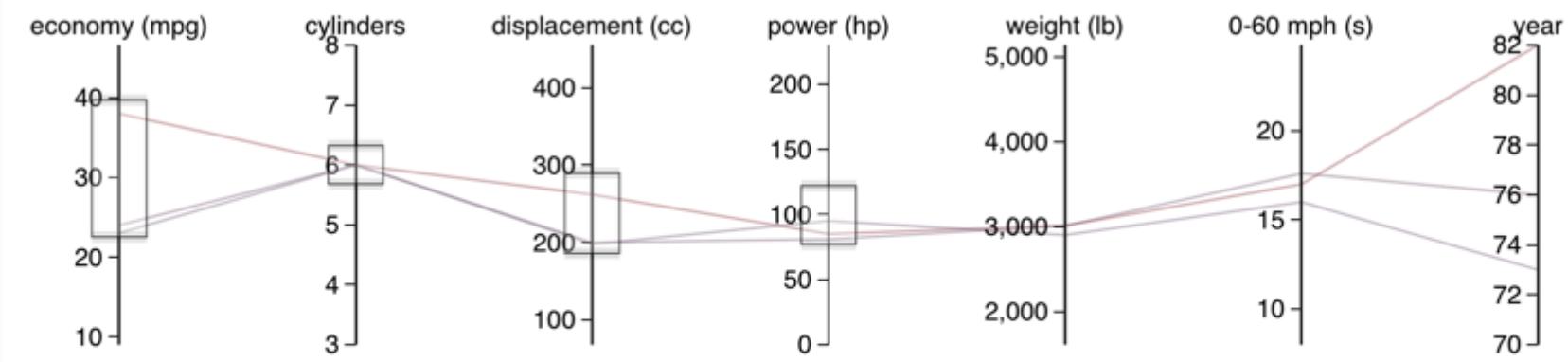
- Observation: We have evolved a sophisticated ability to interpret faces.
- Idea: Encode different variables' values in characteristics of human face



Chernoff, Herman. "The use of faces to represent points in k-dimensional space graphically." Journal of the American Statistical Association 68.342 (1973): 361-368.

Parallel Coordinates

- Encode variables along a horizontal row
- Vertical line specifies different values that variable can take
- Data point represented as a polyline



name	economy (mpg)	cylinders	displacement (cc)	power (hp)	weight (lb)	0-60 mph (s)	year
Ford Maverick	24	6	200	81	3012	17.6	76
Oldsmobile Cutlass Ciera (Diesel)	38	6	262	85	3015	17	82
Plymouth Duster	23	6	198	95	2904	16	73

A. Inselberg, Multidimensional Detective, Proc. IEEE InfoVis, 1997

Dr. Ke Zhou (http://www.cs.nott.ac.uk/~pszkz/)

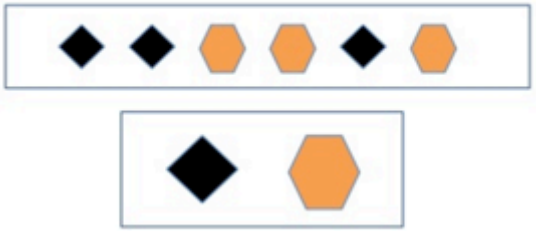
Expected Answer

- Explain Chernoff Faces and Parallel coordinates
  - Chernoff faces exploits the individual parts, such as eyes, ears, and nose of the face to represent values of the variables.
  - In a parallel coordinates plot, the axes are placed in parallel and each data point is represented as a series of line segments intersecting the axes at the corresponding values.
- Find value of data case
  - Parallel coordinates are more suitable for finding value of data case when the data is of high dimension;
  - It is more difficult to find value in Chernoff faces, but it is easier to recognize differences between data cases.

Data Manipulations

- List and describe the five most common data manipulation techniques.

5 Basic Verbs

- FILTERows 
- SELECT Column Types 
- ArRange Rows (SORT) 
- Mutate (into something new) 
- **Summarize** by Groups 

dplyr

- dplyr takes the %>% operator and uses it to great effect for manipulating data frames
  - Works only with data frames
  - 5 basic "verbs" work for 90% of data manipulations

Verbs	What does it do?
`filter()`	Select a subset of ROWS by conditions
`arrange()`	Reorders ROWS in a data frame
`select()`	Select the COLUMNS of interest
`mutate()`	Create new columns based on existing columns (mutations!)
`summarise()`	Aggregate values for each group, reduces to single value

Expected Answer

- Filter: select a subset of data cases by a given condition.
- Arrange: reorder the data cases.
- Select: select a subset of the variables of interest.
- Mutate: create new variables of interest based on existing variables.
- Summarize: aggregate values for each group, reducing to single value.
- (Other answers may be correct as well, such as joining, etc.)

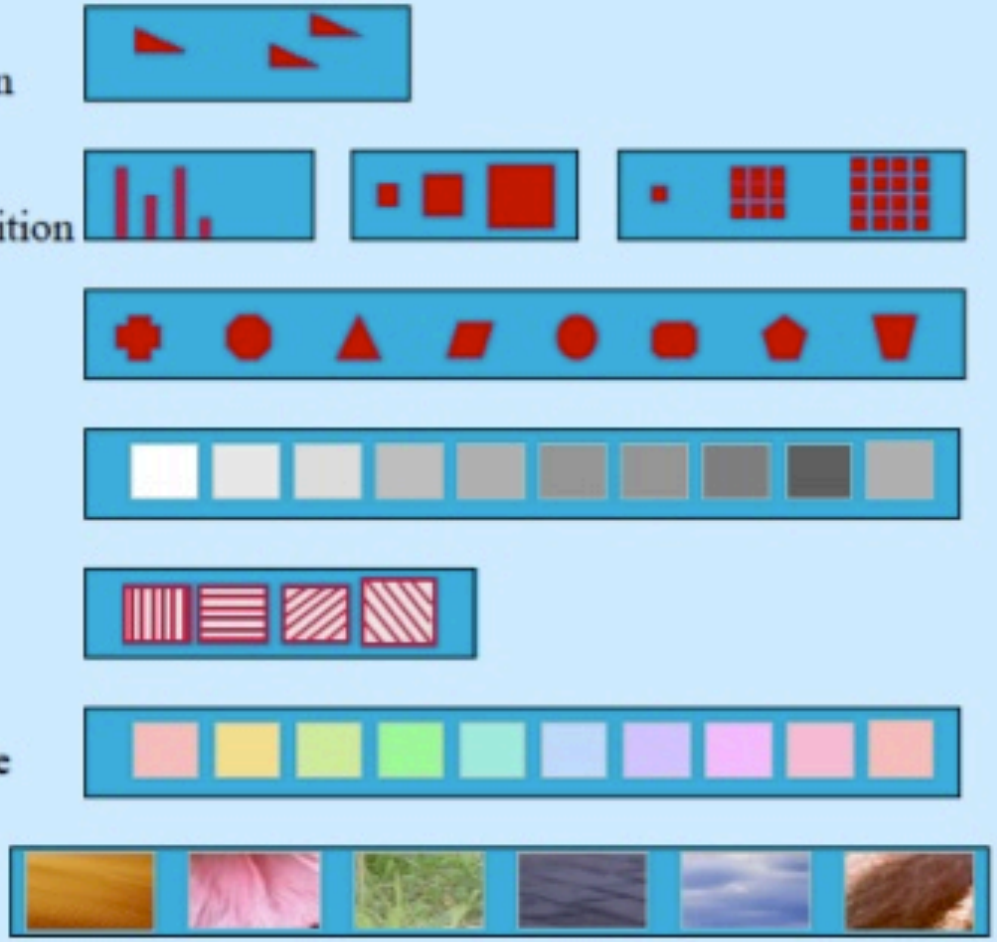
Visualization

- List and explain the visual encodings and their corresponding data types in this visualization.

Visual Encoding Variables

Bertin's Semiology of Graphics (1967)

- **position**
  - changes in the x, y, (z) location
- **size**
  - change in length, area or repetition
- **shape**
  - infinite number of shapes
- **value**
  - changes from light to dark
- **orientation**
  - changes in alignment
- **colour**
  - changes in hue at a given value
- **texture**
  - variation in pattern
- **motion**



Graphic by: Sheelagh Carpendale

Levels of Organization

	Nominal	Ordinal	Quantitative
Position	✓	✓	✓
Size	✓	✓	~
(Grey)Value	✓	✓	~
Texture	✓	~	✗
Color	✓	✗	✗
Orientation	✓	✗	✗
Shape	✓	✗	✗

✓ = Good
~ = OK
✗ = Bad

Expected Answer

- Two visual encodings are used in this visualization: position and color.
- The x and y positions represent respectively the school (nominal data) and annual salary (quantitative data).
- The color Hue demonstrates different gender, which is of nominal data type.

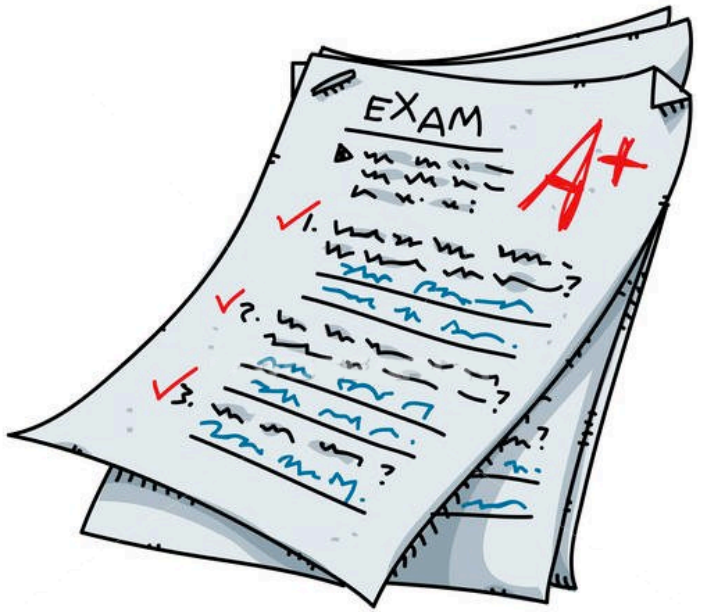
Review Tips

- You can find most of the key concepts or visualizations in the “recap of fundamentals” slides (Lecture 13).
  - A quick overview

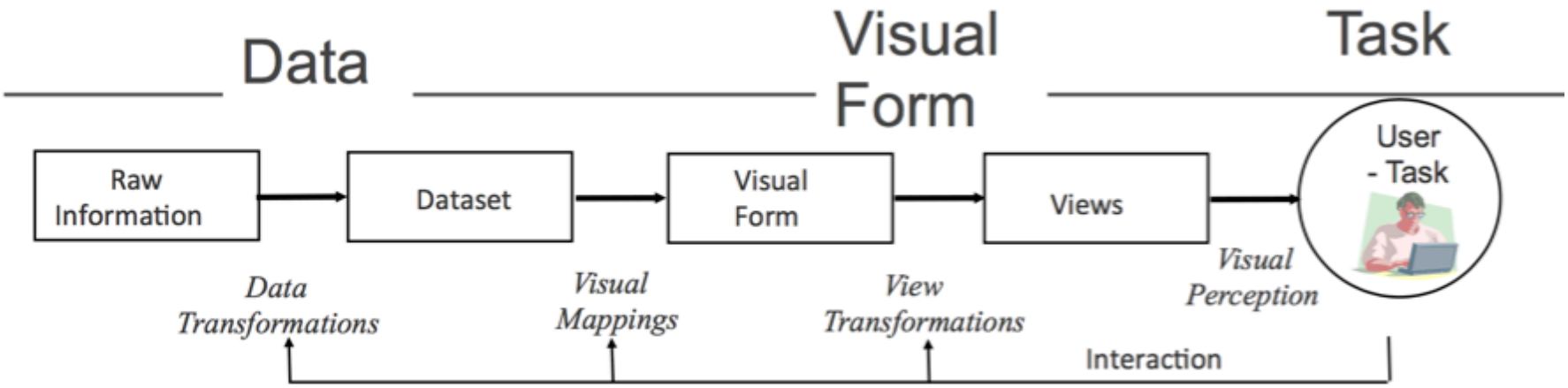
- Review the lecture slides, the core texts and paper handouts.

Past Paper

- The past paper is available on Moodle.
- Let us go over it to see how you can get 100%.



Information Visualization



- Fundamental understanding on how visualizations convey information and how humans perceive
- Master an essential set of visualization techniques
- Practical experience in visualizing real-world data

SET/SEM Survey

SET/SEM Survey

- Official campus course evaluation
- Distributed and completed online. Your opinion is valued!

Module	Survey Type
Fundamentals of Information Visualisation	SET
Fundamentals of Information Visualisation	SEM

- Thanks for a great semester!

Next Lecture

- Topic:
  - Demo

- The Monday on 13 May
  - 12:00 - 14:00
  - A25, Business South, Jubilee Campus