

UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2019

IS53048C

Data Visualisation and the Web

Duration: 2 hours 15 minutes

Date and time:

---

*This paper is in two parts: part A and part B. You should answer ALL questions from part A and TWO questions from part B. Part A carries 40 marks, and each question from part B carries 30 marks. The marks for each part of a question are indicated at the end of the part in [.] brackets.*

*There are 100 marks available on this paper.*

*Calculators are not permitted in this examination.*

**THIS PAPER MUST NOT BE REMOVED  
FROM THE EXAMINATION ROOM**

## **Part A**

**Please answer all questions**

**Question 1**

Explain the difference between *quantitative* and *qualitative* data and give one example of each. [4]

**Question 2**

List four real-world situations in which data visualisation can be used to inform decision making or influence behaviour. [4]

**Question 3**

Explain *exploratory data visualisation* and when it is used. [4]

**Question 4**

Explain *explanatory data visualisation* and when it is used. [4]

**Question 5**

Define *median*, and provide two examples of variables for which the median is an appropriate descriptive statistic. [4]

**Question 6**

Measures of central tendency (mean, median and mode) are important quantities for summarising data. Describe two other descriptive statistics and explain what these measures communicate about data. [4]

**Question 7**

Inclusivity and privacy are important concepts in designing surveys and communicating findings. Give an example of each concept and briefly explain its importance in relation to responsible data visualisation practice. [4]

**Question 8**

List four common mistakes when using pie charts. [4]

**Question 9**

Using diagrams, briefly explain the Gestalt laws of *proximity* and *continuity*. Give an example of how each law can apply in the context of data visualisation. [4]

**Question 10**

In scientific visualisation describe *scalar field*, and give an example. [4]

## **Part B**

**Please answer two questions**

**Question 11** The data visualisation process

- (a) The process of turning data into information, and ultimately into action, requires a number of steps. Using examples, explain each of the different stages of the information visualisation process. [15]
- (b) When searching for data to answer a particular research question, how do you assess whether:
- i. the source of the data is *trustworthy*;
  - ii. the findings of your analysis will be *valid* (i.e. provide accurate answers to your research question)? [10]
- (c) Describe the difference between *arbitrary* and *sensory* visual symbols, giving an example for each. State two benefits of using sensory symbols in data visualisation. [5]

**Question 12** Visualising distributions

- (a) Tables and grouped bar charts can be effective ways to visualise relationships between categorical variables.
- i. Construct a table showing the joint frequency distribution of two *nominal* variables, i.e. the table should show the cross-tabulation of one variable by the other. [5]
  - ii. Sketch a *grouped* bar chart visualising the cross-tabulation from a.(i). [10]
- (b) One of the first steps in exploratory data visualisation is to plot the distribution of each variable. For numerical variables, a box plot is often used. Draw an example of a box plot and explain in detail the statistical features used and how are they calculated. Explain the definition of a Tukey plot. Describe how and why box plots are useful in data analysis. [15]

**Question 13** Types of data visualisation and understanding correlation

- (a) Describe the characteristics and aims of *information visualisation* and *scientific visualisation*, and provide two examples of each. Briefly discuss what the two approaches share in common. [15]
- (b) The following figure is a visualisation of Francis Anscombe's *Quartet*. These data sets are special in that they all have near identical statistical properties (mean, variance, correlation coefficient, regression line).
- i. For each plot describe the relationship visualised between  $x$  and  $y$ , and comment of the goodness of fit of the linear regression line. [9]
  - ii. For each plot describe a real-world example that could likely generate data with such a relationship. [4]

iii. What point was Anscombe making with his quartet?

[2]

