UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2018

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

List four important uses of data visualisation. [4]

Question 2

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

Question 3

Define and give one example of each of the following variable types. [4]

- ordinal
- interval

Question 4

List the mathematical operations that are applicable to ratio level variables. [4]

Question 5

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

Question 6

For each of the following scenarios state whether they are examples of *descriptive* or *inferential* statistics. Briefly justify your answer in each case. [4]

(a) A library produces a table of the top 10 most requested items of the year from its online catalogue.

(b) A university claims that five years after obtaining a degree, graduates from their institution will earn on average 12% higher salaries than graduates from other institutions.

(c) A farmer calculates the variability in wheat yields, measured in tonnes per hectare, across all fields of their farm.

(d) A sociologist publishes findings that 90% of Londoners are satisfied with the quality of coffee in the city.

Question 7

What is the *68–95–99.7* rule? When analysing standard deviation how is this rule useful? [4]

Question 8

List four advantages of bar graphs over pie charts. [4]

Question 9

Matplotlib provides a number of standard colourmaps suitable for different kinds of data. Explain the characteristics of *sequential* colourmaps and give an example of when they should be used. [4]

Question 10

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

Part B

Please answer two questions

Question 11   The data visualisation process

(a) List four reasons why it is important to articulate clear research questions when
undertaking data analysis.   [4]

(b) Describe the characteristics of *exploratory* and *explanatory* information visuali-
sation. Provide an example of each. The examples should clearly illustrate the
differences between the two visualisation processes.   [15]

(c) Describe three distinct contexts, other than conventional research, in which data
visualisation can play a key role.   [6]

(d) More powerful data visualisations can reveal *relationships* between variables. Using
a graph type of your choice, draw an example graph that visualises a relationship
between at least two variables. Explain or annotate the features of your graph.   [5]

Question 12   Distribution plots and visual perception

(a) Both histograms and box plots (Tukey plots) are used to visualise the distribution
of variables. Draw an example of each, annotating the important features. Briefly
explain the advantages of each representation.   [15]

(b) Explain the three-stage visual perception model discussed during the course: what
are the functions and characteristics of each stage?   [9]

(c) List three visual features that are processed *pre-attentively* by the visual system.
Give an example of how each feature can be used in effective visual communication.   [6]

Question 13   Data collection and scientific visualisation

(a) Describe the main differences between using an application programming interface
(API) to retrieve data from an online service and using web scraping.   [6]

(b) When designing an online survey, list four important design considerations that
will help ensure the collection of good data.   [4]

(c) In scientific visualisation describe *scalar field* and *vector field*. Give two examples
each.   [10]

(d) In scientific visualisation describe an isoline (contour line) and an isosurface and
their applications.   [10]