UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2020

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

Describe the characteristics and aims of *information visualisation* and provide two examples. [4]

Question 3

Describe the characteristics and aims of *scientific visualisation* and provide two examples. [4]

Question 4

What data type are each of the following? [4]

(a) How many siblings do you have?

(b) How did you travel to university today?

☐ bus

☐ cycle

☐ train/tube

☐ walk

(c) How would you rate the quality of your meal today?
⋆
⋆⋆
⋆⋆⋆

(d) What time did you wake up this morning?

Question 5

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

Question 6

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

Question 7

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

Question 8

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 9

List four situations when pre-processing and visualising data in which it would be reasonable to exclude an observation from a dataset. [4]

Question 10

Describe the difference between *arbitrary* and *sensory* visual symbols, giving an example for each. [4]

Part B

Please answer two questions

Question 11   The data visualisation process

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

(b) List five reasons why it is important to articulate clear research questions when undertaking data analysis. [5]

(c) When searching for data to answer a particular research question, how do you assess whether:

    i. a data source is *trustworthy*; [5]

    ii. the findings of your analysis will be *valid* (i.e. provide accurate answers to your research question)? [5]

Question 12   Visualising distributions

(a) Both histograms and box plots (Tukey plots) are used to visualise distributions.

    i. Draw an example of a histogram and annotate the important features. [4]

    ii. Draw an example of a box plot (Tukey plot) and annotate the important features. [9]

    iii. Briefly explain an advantage of each representation. [2]

(b) 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 *stacked* bar chart visualising the cross-tabulation from b.(i). [10]

Question 13   Perception, correlation and data collection

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

(b) Using diagrams, briefly explain the following Gestalt laws and give an example of how each can apply in the context of data visualisation. [6]

    i. *similarity*

    ii. *connectedness*

    iii. *proximity*

(c) Draw an example scatter plot, including regression line, for each of the following sample Pearson correlation coefficients. [9]

      i. $r = 0$

     ii. $r = 0.9$

   iii. $r = -0.4$

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