This question paper consists of 3 printed pages each of which is identified by the Code Number (COMP373601)

This is an open book examination. Any written or printed material is permitted.

School of Computing

January 2018

COMP373601

Information Visualization

Answer all 2 questions

Time allowed: 1 hour 30 minutes

Question 1

(a) Imagine that you have been provided with the following data for every student who has graduated from the School of Computing during the past 5 years and is currently working in the UK.

- The student's year of graduation
- The student's current salary
- The student's final year project mark (out of 100)
- Whether or not the student did an industrial placement

(i) For each variable, classify the data type (categorical, ordinal, or numerical). **[2 marks]**

(ii) You have been asked to create one chart that shows all four variables. Illustrate the chart that you would choose with a neat diagram. **[4 marks]**

(iii) With reference to both Mackinlay's rankings and the characteristics of the data, explain your choice for each variable's visual encoding. **[6 marks]**

(b) Imagine that you have developed a novel visualization tool that engineers can use to analyse data from sensors that have been installed in some equipment. Two methods for evaluating the visualization tool are:

- Analysing logfiles from the tool
- Interviewing the engineers after they have used the tool

(i) State the purpose of recording logfiles, provide an example of an insight that you may gain about the engineers' usage of the tool, and state how that insight would be exhibited in the logfile data. **[3 marks]**

(ii) State one way in which interviews can complement logfile analysis, and explain why that is useful. **[3 marks]**

(c) Image that a climate scientist tells you that they work with big data because their model uses a fine-grained spatial grid and short time steps. Which of the '5 Vs' is most applicable, and why? **[2 marks]**

[question 1 total: 20 marks]

Question 2

(a) Spreadsheet software allows you to create tiny charts (heat maps, line charts and bar charts) within the individual cells of a spreadsheet, using functionality such as conditional formatting and sparklines.

Imagine that you have a dataset that contains 50 categorical variables (examples are a person's name, gender and occupation) and one million records. How would you use those tiny charts to provide an overview of the dataset, with each variable limited to one row of a spreadsheet? You may find it useful to illustrate your answer with a diagram. **[6 marks]**

(b) Imagine that you have data about the % of unemployed people in each county of the UK, for each of three years. Compare small multiples vs. animation for visualizing the data. **[2 marks]**

(c) Now imagine that you have 100 years of data for the scenario that is described in (b). Explain two advantages of using animation rather than small multiples to visualize the data. **[4 marks]**

(d) The University of Leeds employs 7861 staff. Assume that each member of staff works for either one academic department (e.g., the School of Computing) or one administration unit (e.g., Human Resources). There are approximately 150 academic departments and administration units. Within each department/unit there is a hierarchy, which extends from the Head of Department/Unit, through several management layers, to staff who have no line management responsibilities.

What type of network visualization method would you use to show every member of staff within a single organisation hierarchy? Illustrate your answer with a neat diagram that indicates how academic departments, administration units, management layers and individual staff are portrayed. Your visualization should also show whether each member of staff works part-time or full-time. **[8 marks]**

[question 2 total: 20 marks]

END