UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2020

IS52032A
Data Journalism and Visualisation

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.

THIS PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Part A

Question 1 What type of data are each of the following:

(a) Height: 60m

(b) Notes: "Police stated they were looking for more than one assailant"

(c) Favourite snack: Snickers

(d) Period: 3rd Quarter

[4]

Question 2 Write the following percentages as natural frequency phrases:

(a) 20%

(b) 33%

(c) 75%

(d) 15%

[4]

Question 3 Suppose you want to search all government websites in the UK for any Powerpoint presentation files they may have. What would you type into Google?

[4]

Question 4 List two common mistakes made when using pie charts.

[4]

Question 5 What type of relationship does a time-series analysis describe? What is the type of chart most suitable to reflect this relationship? Explain what would be your x and y axes.

[4]

Question 6 Describe the difference between a population and a sample. Name two important considerations when choosing a sample.

[4]

Question 7 What is a cartogram? When is it commonly used in data journalism?

[4]

Question 8 Briefly describe why the median is often a more useful measure of centre than the mean. Give one example of how the median is commonly used in journalism.

[4]

Question 9 What is a CSV file? Give two reasons why this file format is common in data journalism.

[4]

Question 10 List four summary statistics used by data journalists.

[4]

Part B

Question 11 Scraping and statistics

(a) What is web scraping? Discuss the legality of web scraping. [9]

(b) Give three examples of how web scraping is used to produce data sets that can be used by journalists. [9]

(c) Briefly explain the difference between descriptive and predictive statistics. When would a journalist use predictive statistics? [6]

(d) What does the standard deviation tell you about a data set? Use diagrams if necessary. What should you do to understand how the data is distributed? [6]

Question 12 Data journalism

(a) Data journalism is often described in terms of a 5-step workflow. List the 5 steps. [5]

(b) Describe the typical processes and at least one tool involved at each of the steps listed in part (a). [10]

(c) Looking for stories in data is often called interviewing the data. Name three techniques for interviewing data using Excel. Briefly describe how each technique is used to find data to use in stories. [9]

(d) Data scientist Hadley Wickham called data that was in the correct form 'tidy data'. List the three rules of tidy data. [6]

Question 13 Digital mapping

(a) What are the two common coordinate systems used to represent a point on a map? [4]

(b) What is a heat map? Give one example of how a heat map might be used in journalism. Describe one common problem with heat maps. [6]

(c) What is geocoding? Describe how geocoding is used by data journalists. [8]

(d) Name three common geographical units of analysis used in the UK and give two examples of data commonly used in stories by data journalists for each unit. [12]