

UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2019

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** Suppose you want to search all police websites in the UK for any Excel files they may have. What would you type into Google?

[4]

**Question 2** What type of data are each of the following:

- (a) Temperature: 17.4C
- (b) Notes: "The council stated the roadworks should be completed by February"
- (c) Best selling album 2017: Pink - Beautiful Trauma
- (d) Reporting period: 2nd Quarter

[4]

**Question 3** Write the following percentages as natural frequency phrases:

- (a) 25%
- (b) 33%
- (c) 75%
- (d) 8%

[4]

**Question 4** 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 5** Describe the difference between a population and a sample. Name two important considerations when choosing a sample.

[4]

**Question 6** 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 7** List the four categories of preattentive properties of visualisations.

[4]

**Question 8** What does a large standard deviation indicate in a normally distributed data set?

[4]

**Question 9** Explain briefly why a journalist might want to use annualised change in a story? Give an example.

[4]

**Question 10** List four descriptive statistics commonly used by data journalists.

[4]

## Part B

**Question 11** 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) What are the three most common types of chart used by journalists to present data in stories? What general types of data are these charts used to present and give an example of a data set that might be used in each case. [9]
- (d) Give two examples of where using small multiple charts is better than using one large chart. Explain the advantages of small multiples in each case. [6]

**Question 12** Scraping and statistics

- (a) What is web scraping? Briefly discuss the legality of web scraping. What does a robots.txt file do? [6]
- (b) Give three examples of how web scraping is used to produce data sets that can be used by journalists. [6]
- (c) Explain, using two examples, of how relative risk and absolute risk are used in journalism. [9]
- (d) What are the common problems with trying to communicate data to the public? What techniques and strategies can you employ to make your writing more understandable when dealing with numbers? [9]

**Question 13** Digital mapping

- (a) What is a choropleth map? Give one example of how a choropleth map might be used in journalism. Describe one common problem with choropleth maps. [6]
- (b) Name three common geographical units of analysis used in the UK and give two examples of data commonly used by data journalists for each unit. [6]
- (c) Name four types of visualisation most commonly used by journalists to show geographical data and for each type give an example of a data set that might be used with it. [8]
- (d) Write a short essay describing how digital maps and satellite imagery are used in online journalism. Give examples to support your points. What are some of the advantages and common problems with using digital maps and satellite imagery? [10]