UNIVERSITY OF LONDON

GOLDSMITHS COLLEGE

Department of Computing

B. Sc. Examination 2013

IS52032A - Data Visualisation and the Web
IS53048A - Data Journalism

Duration: 2 hours 15 minutes

Date and time:

There are three questions in this paper. You should answer no more than TWO questions. Full marks will be awarded for complete answers to a total of TWO questions. Only two questions will be marked, even if you answer all three.

Each question carries 25 marks. The marks for each part of a question are indicated at the end of the part in [.] brackets. There are 50 marks available on this paper.

Where you are instructed to "label your answers", use the following convention:

> *Example question:*
> *Name two types of domestic pet. Label your answers (1) and (2).*
>
> *Example answer:*
> *(1) Dog*
> *(2) Cat*

THIS PAPER MUST NOT BE REMOVED FROM THE EXAMINATION ROOM

Question 1

Question 1 concerns visual perception and graphical encoding.

(a)
Research has demonstrated that visual perception is relative ("perception by differences"), rather than absolute, which has significant consequences for how data is encoded visually. Describe one example that illustrates this characteristic of visual perception.

[3 marks]

(b)
Cartographer Jacques Bertin identified a number of pre-attentive or "visual variables" which can be used to encode information graphically. Name two of these. Label these (1) and (2). Briefly describe how one of them might be used to encode data.

[3 marks]

(c)
Quantitative data always describes relationships. For categorical data, there are four relationships. Name and describe two of these. Label these (1) and (2). In each case, give a hypothetical example of data that illustrates the relationship.

[4 marks]

(d)
In what circumstances would you recommend using a pie chart to represent data and why? State three considerations that you must take into account to using them appropriately.

[5 marks]

(e)
Visualisation expert Edward Tufte introduced the concept of the *data-to-ink ratio*, related to printed statistical graphics. Explain this concept, identifying two ways to highlight data. Discuss why the data-to-ink ratio is relevant to on-screen data visualisations.

[10 marks]

Question 2

Question 2 concerns story craft, data collection, and data manipulation.

(a)
Where might you search online to find data about rates of knife crime in UK Local Authorities? Name one source and describe what format the data might be in.

[3 marks]

(b)
Name three types of data file format you might reasonably expect to be able to download from a formal data source, such as a national or international statistical organisation. Label these (1), (2), and (3).

[3 marks]

(c)
Search engines contain a variety of tools to aid users in refining their results. Name and describe four techniques for "strategic searching". Label your answers (1), (2), (3), and (4).

[4 marks]

(d)
What is "screen scraping"? Describe why you might need to use it. Describe a tool(s) or software language you could use.

[5 marks]

(e)
Journalist Paul Bradshaw describes 5 components of a Data Visualisation workflow. Name and describe each step of this process. Label these (1), (2), (3), (4), and (5).

[10 marks]

Question 3

Question 3 concerns type of visualisations and techniques.

(a)
What is a choropleth representation? Give an example of a data set that it might be used to visualise.

[3 marks]

(b)
Data Driven Documents (D3) is a software framework for creating visualisations on the web. It consists of three related technologies. HTML, CSS and the D3 JavaScript library. In terms of creating a visualisation, what job does each of these perform? Give short answer for each and label these (1), (2), and (3).

[3 marks]

(c)
Network visualisations are used to illustrate relationships. Name and describe two types of network diagram. Draw a picture approximating what the diagram looks like. Label these (1) and (2).

[4 marks]

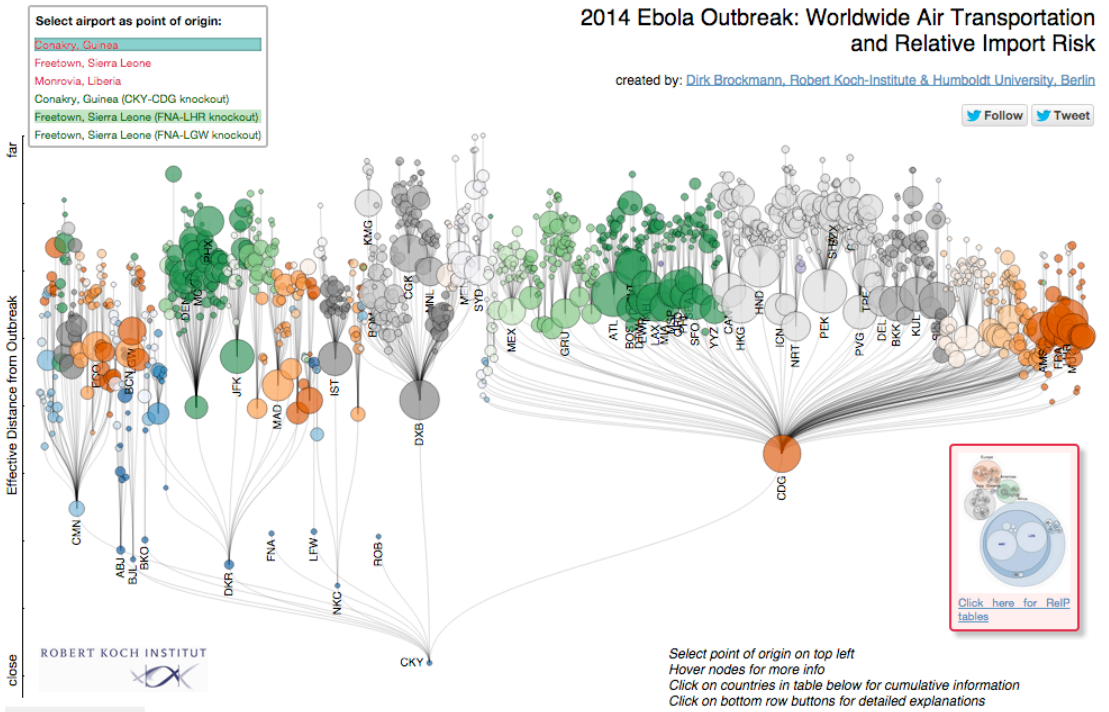
(d)
What is a *dot-distribution map* visualisation? Describe a hypothetical example of a data set that could be visualised with a *dot-distribution map*. What important considerations must designers take into account when considering the size and value of marks?

[5 marks]

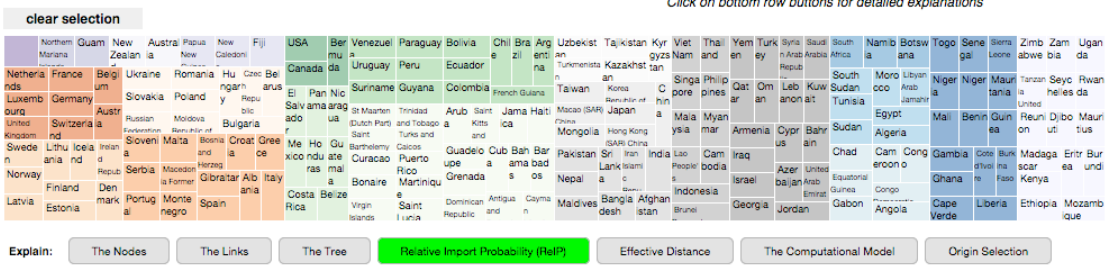
(e)
Examine the visualisation on the next page. It uses two visualisation techniques – on the top of the visualisation, (labelled **A**) and the other across the bottom (labelled **B**). Identify the two types of visualisation. Label these (1) and (2). Name three basic graphical elements used to encode the representation (there are many more than three), and the categories of data they encode. Label these (3), (4), (5). What is the biggest limitation visualisation (A)?

[10 marks]

A


Select airport as point of origin:
Conakry, Guinea
Freetown, Sierra Leone
Monrovia, Liberia
Conakry, Guinea (CKY-CDG knockout)
Freetown, Sierra Leone (FNA-LHR knockout)
Freetown, Sierra Leone (FNA-LGW knockout)
2014 Ebola Outbreak: Worldwide Air Transportation and Relative Import Risk
created by: Dirk Brockmann, Robert Koch-Institute & Humboldt University, Berlin
Follow
Tweet
far
close
Effective Distance from Outbreak
CKY
CDG
DXB
DKR
FNA
LFW
ROB
NKC
CMN
ABJ
BJL
BKO
Click here for RelP tables
ROBERT KOCH INSTITUT
Select point of origin on top left
Hover nodes for more info
Click on countries in table below for cumulative information
Click on bottom row buttons for detailed explanations

B


clear selection
Explain:
The Nodes
The Links
The Tree
Relative Import Probability (RelP)
Effective Distance
The Computational Model
Origin Selection