

King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

Examination Period	August 2020 (Period 3)
Module Code	7CCSMSDV
Module Title	Simulation and Data Visualization
Format of Examination	Project Based
Time Overall	FORTY EIGHT hours
Instructions	You are permitted to access any materials you wish, but this is not mandated and is not expected. You may use a calculator if you find this helpful.
Rubric	<p>ANSWER ALL QUESTIONS: Part 1, 2 and 3. Part 1 carries THIRTY marks. Part 2 carries THIRTY FIVE marks. Part 3 carries THIRTY FIVE marks.</p> <p>Submission instructions. You shall submit a zipped file named <i>StudentID.zip</i> (where <i>StudentID</i> shall be replaced by your student ID) and containing the following:</p> <ol style="list-style-type: none">1. The answers to Part 1 and Part 2 shall be submitted as a single report in .pdf format. Please label each section of the report clearly with Part 1, Part 2 headings.2. The answer to Part 3 shall be submitted as a folder containing your webapp implementation (accepted formats: a WebStorm project, or a self-contained .html page).
Submission Deadline	August 21, 2020 - 11pm - BST
Submission Process	Work must be submitted to the appropriate Submission link on the module (e.g. 7CCSMDV) KEATS page.

You should check that your work displays correctly after it has been uploaded.

ACADEMIC HONESTY AND INTEGRITY

Students at King's are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a College policy that students take responsibility for their work and comply with the university's standards and requirements. Online proctoring / invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarise themselves with the procedures on the College website

Important: Students should copy out the following statement and include it with their submission for each examination:

I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.

Project Based Exam

Introduction

This exam is divided into 3 parts: Analytics, Design and Prototyping, Implementation. Please read carefully the instructions related to each part.

You are given 2 working days to complete.

Fake news is a phenomena where news or stories are created to deliberately misinform or deceive readers. During the Covid-19 pandemic there has been a proliferation of online misinformation. The fake news phenomena has touched aspects such as: the disease itself, its origins, potential treatments, government responses etc.. False health information can be dangerous.

Even when fake news is not as consequential as it may be feared, it is important to be able to identify and analyse it.

We as Data Scientists are called to put our skills to help.

Data. We are providing you with:

- Two starting datasets (courtesy of [The GDELT Project](#)). The first dataset reports a compilation of URLs and brief snippets of worldwide English language news coverage mentioning Covid-19 and the words: “fake news” OR misinformation OR disinformation OR “false claim” OR “conspiracy” OR “falsehood” OR “rumor” . The second dataset reports locations mentioned in Covid-19 news coverage.

At least one of the two provided datasets should be used for the Exam.
[Dataset available on Keats Exam page]

- A series of links to Covid-19 related fake news datasets currently available. These are given as a starting point, it is not mandatory to use them and you are invited to look for any other datasets available on the web that you may find interesting. [List provided on Keats exam page]

Visualizations. We are providing you with links to existing visualizations, again these are given as a starting point and for inspiration, it is not mandatory to follow them and you are invited to look for other visualizations and visualization

platforms that are available on the web and that you may find inspirational. [List provided on Keats exam page]

Research questions. You are provided with an initial research question:

Q1: "Analyze the spread trend of tweets by falsehood score. What is the spread over time and by tweet?"

1. Analytics [30 marks total]

Based on the data provided in the Data section and/or other datasets you may have encountered or may find online, you are required to:

- a.** Propose two or more exploratory research questions (non-trivial questions) beyond Q1, label them as Q2, Q3, etc. [Note: You can propose questions that might require access to datasets beyond the one provided, these can be questions complementary to Q1, investigating other aspects than the ones investigated by Q1, or refinements/expansions of Q1, you are free to decide.].

[10 marks]

- b.** Explain what type of data, beyond that provided, could be used to answer Q1 and each one of the questions you proposed in **a.**. Assess the fitness of each dataset(s) you would potentially be using to answer the questions (these include the ones we provided or other resources you may have found, please provide links to the latter).

[10 marks]

- c.** Explain if and how datasets, described in **b.** and including the ones provided, could be correlated.

[10 marks]

2. Design and Discussion [35 marks total]

Based on visualization approaches surveyed in class and in recommended readings:

- a. Propose and design a minimum of 3 visualizations that would answer the research questions proposed in **Part 1** (creativity will be rewarded). By design we mean drawing/sketching a prototype. Design can be hand-drawn on paper or using a tool of your choice e.g. PowerPoint, Sketch, Illustrator, D3, Tableau, etc.

[20 marks]

- b. Each visualization should be accompanied by a maximum of 500 words describing the design rationale, which question(s) your design would help answer and if/how your design may improve upon existing examples. By **design rationale** we mean: the process and principles followed in choosing the specific visualization. You should provide a rigorous rationale for your design decisions, e.g. visual encodings used and why they are appropriate for the data. These decisions include the choice of visualization type, size, colour, scale, mark and channels and other visual elements, as well as the use of sorting or other data transformations. Consider how these decisions facilitate analysis and/or communication.

[15 marks]

Note: In this part we are only asking to design possible visual layouts not to implement them. If you are hand-drawing your designs take pictures and add them to your document as figures. If you are using a tool to develop your designs save them as images and add them to your document.

3. Implementation [35 marks]

Out of the visualizations proposed in **Part 2** implement one in D3 as a webapp. You shall use either the data provided or other data sources you foraged yourself.

Your visualization shall support answering one of the research questions, therefore:

- it shall be accompanied by a short description of how data are being processed (and acknowledgement of your data source(s)). [10 marks]
- it can include either a composition of linked/related simple visual layouts or a more sophisticated single visual layout. [15 marks]
- it shall allow some level of user interaction. [10 marks]

Note: You are allowed to use D3 Example code available on the web as long as it is adapted to your data and you explicitly acknowledge the source of the original code.