Computer Science Principles
Performance Task: Investigate — Bits to Information to Knowledge

Computing enables and empowers new methods of information processing that have led to monumental change across disciplines, from art to business to science. Managing and interpreting an overwhelming amount of raw data is part of the foundation of our information society and economy. People use computers and computation to translate, process, and visualize raw data, enabling new understandings that in turn contribute new knowledge to the world.

For this task, you and a partner will research and work with data that is of interest to you. You will use appropriate computational tools and techniques to transform the raw data into information and knowledge.

General Requirements

- You must develop and submit one collaborative document.
- You must develop and submit two individual documents.
- You must work in pairs on the collaborative portion of this project. A group of three is only allowed if there are an odd number of students in the class.
- Your team must identify an area of focus, and both the collaborative and individual submissions must share that same area of focus.
- You must work alone on the individual portions of this task. You must write the individual report and reflection on your own.

Prepare and submit the following:

Collaborative Submission
For this submission, each group member will submit identical documents.
- Along with a partner, identify and describe a set of questions you will investigate to gain insight and knowledge from publicly available data. See the Collaborative Investigation section (**A**) for details.
- Submit a report that describes the details of your investigation and communicates your findings. See the Collaborative Report section (**B**) for details.

Individual Submission
- Produce a detailed description and justification of the tools and techniques your group used in the investigation, including a description of how your group achieved its results. See the Individual Report section (**C**) for details.
- Submit a written description and analysis on the collaborative aspects of the investigation. See the Individual Reflection section (**D**) for details.

A. Collaborative Investigation

1. Develop a set of three to five non-trivial questions that will provide the focus of the investigation. These questions should reflect the rationale for your investigation and will determine the overall direction of your research; it is therefore important for you to think carefully about them. Your questions should enable you to investigate different attributes of the data you select, not just *different properties of the same data attribute*. For example, asking for the minimum, maximum, and median value of temperatures would not count as different attributes, since the questions all refer to temperatures.

   You will be evaluated on the quality and richness of your questions. The answers to the questions you ask should provide insight or knowledge about the area being investigated.

   Note: It is highly unlikely that questions that simply count data items will lead to great insight. Furthermore, questions that yield only Yes/No answers can often bias the answer toward a particular direction and overlook the subtlety of the data itself. For example, if the answer is mostly *Yes*, but sometimes *No*, it might be very important *when* it's *No*. You may have to consider factors that led to the *No* answer and whether a *No* answer has additional implications.

2. Find one or more large data set that will allow you to obtain answers to your questions. You must use a data set that begins with at least 1000 data values, where a data value is considered a cell in a table. You may use more than one data set. You must use a publicly available, permanent data set or sets from a credible source that supports a meaningful investigation.

3. Apply computational tools and techniques to answer your questions (e.g., by finding patterns in the data, by transforming or translating data, by creating visualizations, or by finding connections between the data and other sources of knowledge).

4. Apply computational tools and techniques to create non-textual representations that communicate the results of your investigation. Use tables, diagrams, charts, graphs, visualizations, or any other appropriate representations to communicate insight and knowledge gained from data.

Your technique(s) should be able to handle data that is 10, 100, or 1000 times larger than your original dataset. Avoid the simpler approaches of looking at the raw data, manually performing the analysis, or manually creating the non-textual representations.

B. Collaborative Report

As a group, respond directly to each of the following prompts. This report will provide the details of your investigation and should include non-textual representations that communicate the results of your investigation. These representations could include charts, tables, graphs, visualizations, or other appropriate resources that augment or clarify your questions, analysis, and answers.

1. Describe the area of focus you and your partner chose to explore and explain why you chose this area. Your answer should be no more than 200 words.

2. List the set of three to five questions you sought to answer.

3. Identify each permanent data set you used. **For each set,** you must:
   a. Provide the permanent URL.
   b. Briefly describe the data set. Each description should be no more than 100 words.
   c. Provide the date on which you accessed each data set.

4. Clearly present the answer to each of the questions you investigated. Each answer should be no more than 200 words. Attach and reference any appropriate non-textual representations. Include a one sentence description of each non-textual attachment.

5. Explain how your questions and answers contribute to the development of insight and knowledge in your chosen area of inquiry. Your explanation should be no more than 300 words. Attach (as pdfs) and reference any appropriate non-textual representations. Include a one-sentence description of each non-textual attachment.

C. Individual Report

On your own, respond directly to each of the following prompts.
1. Explain why your questions are rich enough and your data set(s) are large enough to require the use of computation to analyze the data and answer your questions. Your explanation should be no more than 200 words.

2. Explain why your group selected the data set(s) used in your investigation. Your explanation should be no more than 200 words.

3. Describe and justify the specific computational tools (visualization software, spreadsheet, statistical package, etc.) and techniques (filtering, clustering, visualizing, analyzing, etc.) your group used. Provide a detailed description of how your group processed the information in the data set(s) to conduct the investigation and how this process enabled you to meet your objective of gaining insight and knowledge. This description should be sufficiently detailed to make it clear that you could conduct the investigation on your own. It should also include sufficient detail to allow a reasonably skilled reader to use the tools and techniques to replicate the investigation and verify the results. Your description should be no more than 700 words; the following questions may help guide your writing.
   a. What challenges did you face in finding the data sources?
   b. Did the data require any filtering or cleaning? If so, how did you process the data?
   c. Did you do any clustering or classifying of your data? If so, how?
   d. Did you translate or transform the data to determine patterns? If so, how?
   e. Did you use any tools to visualize your data? If so, which tools? How did you use them?

f.   Was the size of your data set a challenge? If so, how did you address the challenge?

D. Individual Reflection

Write a brief essay in which you reflect on the collaborative process your group used to do the research. Answer the following questions in your essay, keeping your answers at a strategic level rather than an operational level. For example, talk about how and why you divided the tasks, but not what time and for how long you met with your group. Your answer should be no more than 300 words. Include answers to the following questions:
   a.   How did you share or divide the work?
   b.   Why did you choose to divide the work in this way?
   c.   How did you coordinate your efforts?
   d.   What worked well, and what did not work well, in your collaboration?
   e.   How did you benefit from working with a partner (or partners)?

Learning Objectives
The **Investigate — Bits to Information to Knowledge** Performance Task addresses the following Computer Science Principles Learning Objectives (LOs):
LO 1.1.1 Use computing tools and techniques to create artifacts. [P2]
LO 1.1.2 Collaborate in the creation of computational artifacts. [P6]
LO 1.2.1 Use computing tools and techniques for creative expression. [P2]
LO 2.3.1 Use models and simulations to raise and answer questions. [P3]
LO 3.1.1 Use computers to process information to gain insight and knowledge. [P1]
LO 3.1.2 Collaborate when processing information to gain insight and knowledge. [P6]
LO 3.1.3 Communicate insight and knowledge gained from using computer programs to process information. [P5]
LO 3.2.1 Use computing to facilitate exploration and the discovery of connections in information. [P1]
LO 3.2.2 Use large datasets to explore and discover information and knowledge. [P3]
LO 3.3.1 Analyze the considerations involved in the computational manipulation of information. [P4]