



AUGUST/RESIT EXAMINATIONS 2018/2019

MODULE: CA682A - Data Management and Visualisation

PROGRAMME(S):
MCM M.Sc. in Computing

YEAR OF STUDY: 2

EXAMINER(S):

Suzanne Little	(Internal)	(Ext:6052)
Dr. Brian MacNamee	(External)	External
Dr. Dan Emanoil Grigoras	(External)	External

TIME ALLOWED: 3 Hours

INSTRUCTIONS: Answer **four (4)** questions from the 5 available.

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE INSTRUCTED TO DO SO.

The use of programmable or text storing calculators is expressly forbidden.

Please note that where a candidate answers more than the required number of questions, the examiner will mark all questions attempted and then select the highest scoring ones.

There are no additional requirements for this paper.

QUESTION 1**[TOTAL MARKS: 25]****Q 1(a)****[7 Marks]**

Given the following brief to design a system for a data collection task, list three (3) important questions you would ask your client and suggest a type of database system to use, giving a reason for your choice.

“We are collecting data to use for a marketing campaign by DCU to increase public transport use when travelling to campus. Data sources include public surveys, records from Transport Ireland and information from DCU estates.”

Q 1(b)**[7 Marks]**

- (i) Why is it useful to categorise data?
- (ii) You have data from the last 5 track and field meets run by Athletics Ireland. Identify some different category descriptions for the following pieces of data:
 - A. Athlete profile
 - B. List of races
 - C. Gold, Silver and Bronze winners
 - D. Times of the winners from the sprint races

Q 1(c)**[6 Marks]**

Give two (2) advantages & two (2) disadvantages of non-relational databases and give an example of when a non-relational database would be useful.

Q 1(d)**[5 Marks]**

Given a generic data analytics pipeline – Gathering, Processing, Analysing, Presenting, Preserving – describe in 1 or 2 sentences the activities that can occur at each stage.

[End of Question1]

QUESTION 2**[TOTAL MARKS: 25]****Q 2(a)****[8 Marks]**

- (i) What is metadata?
- (ii) Name three (3) things that metadata is used for.
- (iii) Give examples of different types of metadata.

Q 2(b)**[7 Marks]**

Give simple example metadata describing the module CA682. Why would a standard be useful for this type of metadata and identify one (1) problem with enforcing a standard.

Q 2(c)**[6 Marks]**

Big data is characterised by its Volume, Velocity, Variety and recently Veracity. Veracity refers to how reliable the data is.

- (i) Why does big data have a particular issue with data uncertainty?
- (ii) Give one (1) method for checking the veracity of a big data source.

Q 2(d)**[4 Marks]**

Describe the process of scraping data from a website. Give two (2) rules that you should remember when using this as data source.

[End of Question2]

QUESTION 3**[TOTAL MARKS: 25]****Q 3(a)****[8 Marks]**

You are collecting data to for a marketing campaign to increase public transport use to travel to DCU. You have the following data sources:

- A. Survey of current transport patterns of staff and students.
- B. Access logs from Transport Ireland app filtered by GPS location for the Glasnevin Campus.
- C. Map of transport options for the DCU campuses.
- D. Medical research data showing improved heart health from regular walking based on user's wearing fitbit sensors.

For **each** of these sources, identify one (1) possible cause and consequence of poor quality data.

Q 3(b)**[6 Marks]**

Pick one of the data sources listed in Q3(a).

- (i) Give an example of an approach to cleaning data that you could use.
- (ii) Give an example of how you could enforce better data quality.

Q 3(c)**[6 Marks]**

(i) What are **constraints** with respect to Data Quality? Ensure you define and distinguish between static and dynamic constraints.

(ii) What are two (2) potential problems with enforcing constraints to improve data quality?

Q 3(d)**[5 Marks]**

Open datasets are made freely available for all people to access. Identify and explain two (2) potential problems that may arise in making data open or using open data.

[End of Question3]

QUESTION 4**[TOTAL MARKS: 25]****Q 4(a)****[8 Marks]**

Given the following visualisation tasks, suggest an appropriate graph type (specific chart type not just the category) for each to display the information and give a brief justification.

- A. Summary of voter choices in the Irish Election 2016
- B. Annual income for Computer Science students grouped by university
- C. Population trend for Ireland over the last decade
- D. Average rental prices for each Irish County in 2016

Q 4(b)**[7 Marks]**

In the appendix, Figure 1 shows a graph. Identify three (3) problems with the design and suggest a better method for showing the information, giving a specific chart type that could be used. You do not need to view the appendix in colour.

Q 4(c)**[6 Marks]**

Explain what D3.js is and give 2 examples of things it doesn't do. Explain the difference between rules and selectors (CSS) and how they are used in D3.js.

Q 4(d)**[4 Marks]**

How can design rules help to make better data visualisations? Give an example.

[End of Question4]

QUESTION 5**[TOTAL MARKS: 25]****Q 5(a)****[7 Marks]**

In the appendix, Figure 2 shows a graphic.

What visual communication goals are evident?

Identify two (2) design principles and explain how the graphic applies them to fulfil the communication goals.

Q 5(b)**[6 Marks]**

Using Figure 2, explain and give examples of two (2) gestalt principles and how they are used.

Q 5(c)**[6 Marks]**

In visualisation it is important to understand pre-attentive processing. Explain what a pre-attentive feature is and describe an experiment to determine if a feature is pre-attentive or not.

Q 5(d)**[6 Marks]**

Correctly match the following depth cues for human vision.

Occlusion	A. moving the head slightly to create differences in the sensed images
Convergence	B. difference in direction of our eyes when looking at closer objects
Accommodation	C. images sensed by our two eyes are slightly different and this difference is used to determine depth
Aerial Haze	D. blocking of more distant objects by closer objects
Binocular Disparity	E. objects on the far horizon look hazy due to particles in the air
Motion Parallax	F. muscle tension from re-focussing the eye

Which cue is (mostly) used to create the 3D effect in movies?

[End of Question5]

APPENDICES

Figure 1

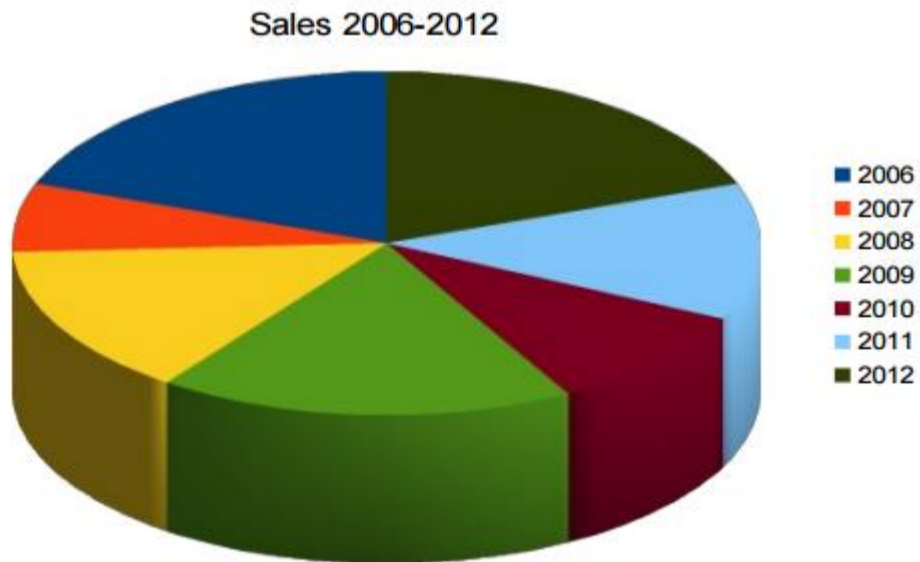
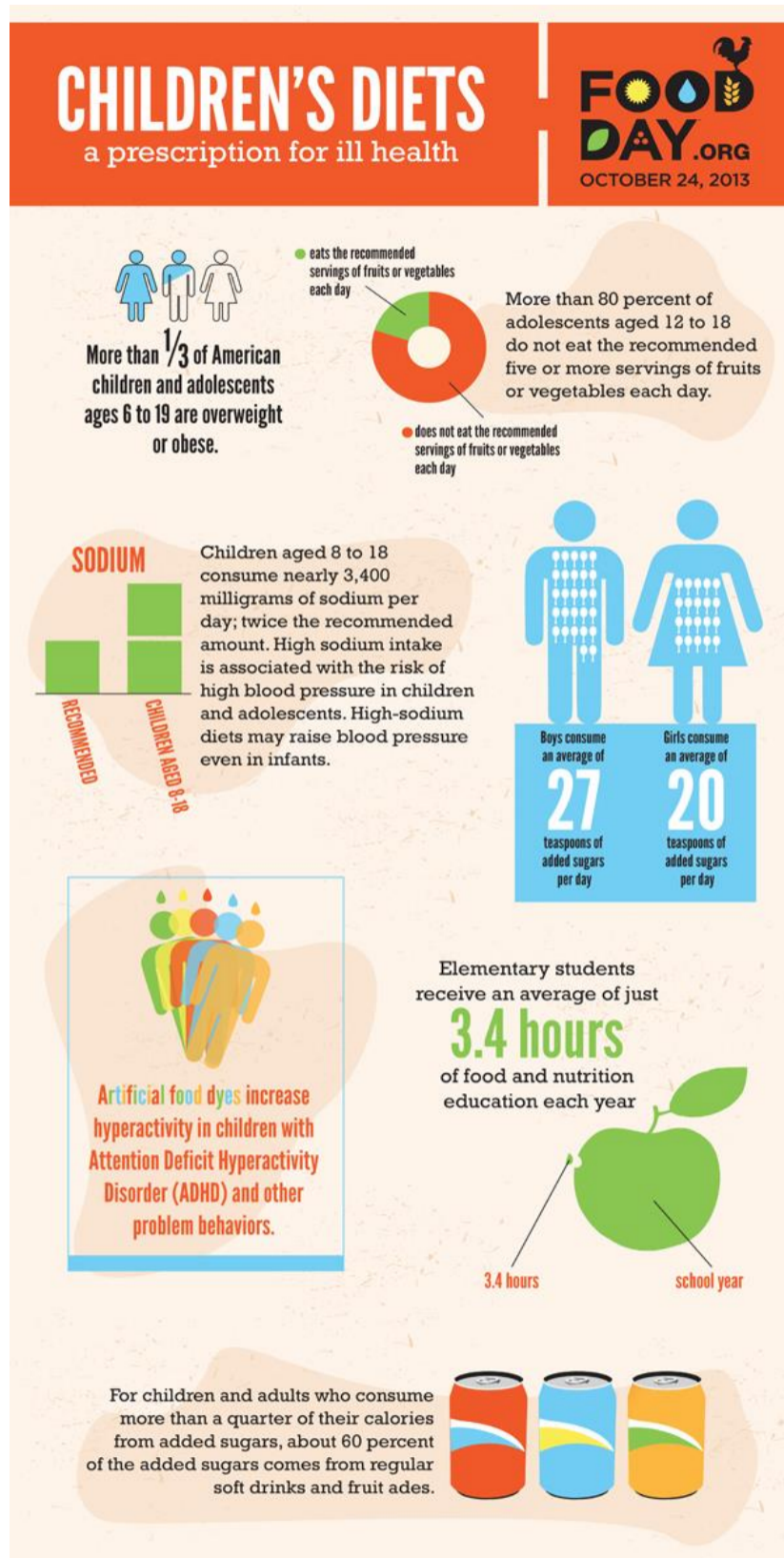


Figure 1 – a 3d pie-chart

Figure 2



[END OF APPENDICES]

[END OF EXAM]