VISUALISATION (BUSI97273)
Mock Exam

Part 1 – MCQ (30%)
The system for multiple choice marking is as follows:
- You score 1 point for each correct answer
- A quarter point is deducted for each incorrect answer
- You score zero for each blank answer
- You must choose only ONE answer per question

Note that the actual exam will contain 10 multiple choice questions each worth 3 points.

1.1 Scatterplot matrices are able to visualise the
a) composition of multiple numerical attributes
b) pairwise correlations between multiple numerical attributes
c) correlation between two categorical attributes
d) distribution of multiple numerical attributes

1.2 Inattentional blindness is a perceptual effect where
a) a person fails to notice a stimulus in plain sight
b) a person is unable to notice the difference between colours
c) a person cannot remember details across separate scenes
d) a person does not pay attention to hidden stimuli

1.3 The three dimensions of colour are
a) luminance, brightness, and hue
b) luminance, saturation, and cognition
c) hue, luminance, and brightness
d) hue, saturation, and brightness

1.4 An adjacency matrix view is well suited for
a) path-related tasks
b) neighbourhood-related tasks
c) exploring topology
d) encoding another attribute with size of each cell

…

Part 2 – Open-ended Questions (40%)

2.1   What is *Anscombe's Quartet* and for which purpose is it used in the context of data visualisation?

2.2   Describe what *multiple coordinated views* are and which different types exist. What is *brushing* in this context?.

2.3   What do *Gestalt laws* describe? Name four specific Gestalt laws together with one example of a visualisation technique that makes use of the law.

2.4   Explain the difference between *geometric* and *semantic zooming*. Give one example for each zooming technique. What other navigation techniques do you know?

Part 3 – Practical Question (30%)

Dataset
The well-known *Iris* flower data sets consists of three different types of irises' (Setosa, Versicolour, and Virginica) sepal/petal length and width. "iris" is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. "iris" data frame comes a part of the datasets package (datasets::iris).

Tasks
3.1   Sketch how you would visualise the dataset to compare the distribution of the three species of flowers. The visualisation should consist of two juxtaposed plots where the plot on the left side should compare the distribution of the petal length and the plot on the right side should compare the distribution of the sepal length.
3.2   Explain what the alternative ways of achieving this task are and why you think that your suggested visualisation is the most effective one.
3.3   Create the visualisation you suggested in 3.1 using ggplot2 within an RMarkdown document. Note that you can make use of the ggplot2 cheatsheet that is provided as PDF (https://github.com/rstudio/cheatsheets/raw/master/data-visualization-2.1.pdf).
3.4   Briefly describe what differences in the distribution can be seen when looking at your visualisation.