FIT3179 Data Visualisation

**Sample Questions for Test 1**

Test 1 will cover the content of weeks 1 and 2 on Moodle: pre-recorded lectures, live lectures, textbook chapters listed under “Required Readings” and 4 scanned sections from Kirk 2019 “Data Visualisation” listed under “Lecture Resources” of week 2 on Moodle.

Test 1 will consist of 12 to 15 multiple choice questions and one or two open answer questions. The duration of Test 1 will be 40 minutes.

The correct answer to the sample questions below is indicated in capital letters or highlighted in yellow.

# What is the most likely dataset type of this visualisation?

Chart, surface chart

Description automatically generated

A) 3D Table.

B) 2D Field.

C) 3D Field.

D) 3D Network.

E) 2D Network.

*ANSWER: C*

# Which of the following is not a data attribute type?

A) Qualitative.

B) Ordinal.

C) Ratio.

D) Diverging.

E) Categorical.

*ANSWER: D*

# Let’s think about the size of T-shirts (XS, S, M, L, XL, etc.) as a variable. This variable is best described as which one of the following?

A) categorical attribute

B) ordinal attribute

C) quantitative attribute

D) numeric attribute

E) nominal attribute

*ANSWER: B*

# Which of the following channels is the most effective to encode qualitative data attributes?

A) Length

B) Shape

C) Area

D) Colour hue

E) Colour saturation

*ANSWER: D*

# The bubble chart below presents an overview of different countries in 2019. Each bubble represents a country, and the visual encodings used in the graph include:

\* x-axis position: income

\* y-axis position: life expectancy

\* area size: population

\* colour hue: continent

\* animation: year

# Which of the following list all the attributes that are represented with magnitude channels in this visualisation?

A picture containing chart

Description automatically generated

[Source: https://www.gapminder.org/tools/#$chart-type=bubbles]

A) income, life expectancy & population

B) income, life expectancy & continent

C) income & continent

D) population, continent & year

E) income, life expectancy

*ANSWER: A*

# Consider the following bubble chart from Gapminder [[link](https://www.gapminder.org/tools/#$state$time$value=2012;&marker$axis_x$which=mean_years_in_school_women_percent_men_25_to_34_years&domainMin:null&domainMax:null&zoomedMin:null&zoomedMax:null&scaleType=linear&spaceRef:null;;;&chart-type=bubbles)] which compares different countries in a few aspects (population, life expectancy, etc.). Select all the channels used in this visualisation.

(+1 mark per correct channel identified, -1 mark per incorrect channel selected).

Chart, bubble chart

Description automatically generated

A) Length

B) Position

C) 1D Size

D) Colour hue

E) Colour Saturation

F) Colour luminance

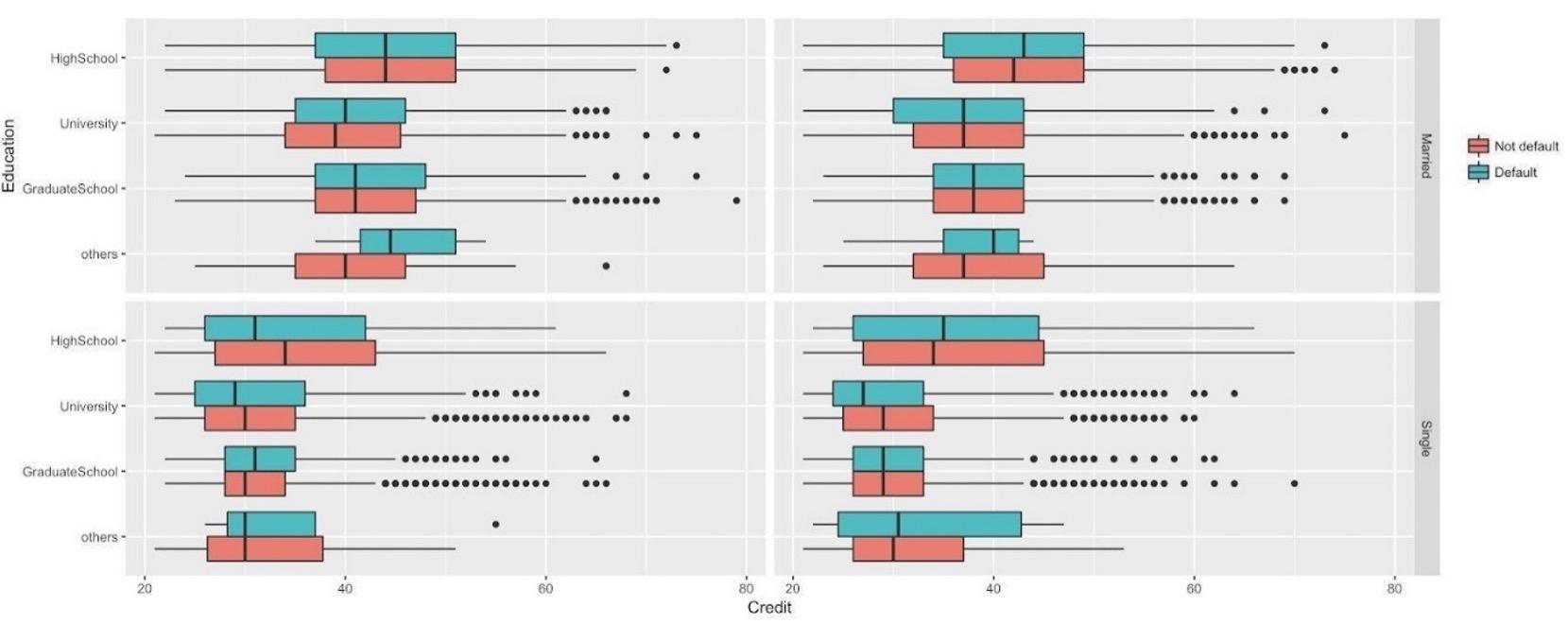
G) Area

H) Shape

*ANSWERS: B, D, G*

# The following boxplot explores factors that may be associated with defaulting on loan repayment. List all the channels used in this visualisation.

(+1 mark per correct channel identified, -1 mark per incorrect channel selected).



A) Length

B) Position

C) Area

D) 2D Size

E) Colour hue

F) Colour Saturation

G) Colour luminance

H) Shape

*ANSWERS: A, B, E*

# Which of the following statements is not true regarding line charts?

A) The line chart idiom is suitable to show how values develop over time.

B) You can use colours, line width and line dashes in a line chart to make your most important values stick out.

C) Using annotations might make your line chart more interesting to read.

D) A line chart is more effective to show how values differ in different categories compared to a (stacked) bar chart.

E) It is a good idea to show each line in a line chart with a distinctive colour.

*ANSWER: D*

# What is an appropriate idiom for a table dataset with two quantitative value attributes?

A) Scatter plot

B) Bar chart

C) Pie Chart

D) Area Chart

E) Line chart

*ANSWER: A*

# Short answer question example 1: Consider the following bubble chart from Gapminder [[link](https://www.gapminder.org/tools/#$state$time$value=2012;&marker$axis_x$which=mean_years_in_school_women_percent_men_25_to_34_years&domainMin:null&domainMax:null&zoomedMin:null&zoomedMax:null&scaleType=linear&spaceRef:null;;;&chart-type=bubbles)] and answer the following questions.

Chart, bubble chart

Description automatically generated

1). What is the dataset type of this visualisation? (1 mark)

2). List all the attributes visualised in the graph and describe the attribute type (1 mark)

3). Identify the “Why?” Part of the visualisation based on the framework by Munzner. (1 mark)

4). Identify the marks and channels used in this visualisation. (3 marks)

5). For the channels that you have identified, which of them are magnitude channels? Based on the effectiveness rank of the magnitude channels, which variable(s) are more important in this visualisation? (1 mark)

***Sample answers:***

*1). Dataset type: table*

*2). Categorical: "continent/world regions" and "country"; Ordinal or quantitative: "year"; Quantitative: "life expectancy", "gender ratio of mean years", and "population".*

*3). Find trends, outliers, distribution, and correlation; locate clusters (you need to write sentences to explain the tasks)*

*4). Marks are points. Channels are colour hue (continent/world regions), vertical (Life expectancy) and horizontal (gender ratio) positions, and size (2D) (population).*

*5). Magnitude channels: positions on a common scale and size (2D). Based on the effectiveness rank, position channels on a common scale are more effective, so the life expectancy and gender ratio are more important in this vis.*

# Short answer question example 2: The two visualisations both visualise the covid-19 cases by continent. They were created with the same dataset and based on the same data attributes (date, new cases, continent).

Chart, line chart

Description automatically generated

Figure 1. Covid-19 cases by Continent (Multiple line chart)

Chart, bar chart

Description automatically generated

Figure 2. Covid-19 cases by Continent (Stacked bar chart)

1) List the marks and channels used in the two visualisations. (2 marks)

2) Based on these two visualization examples, identify and explain the advantages of multiple line charts compared to stacked bar charts. (2 marks)

3) Based on these two visualization examples, identify and explain the advantages of stacked bar charts compared to multiple line charts. (2 marks)

***Sample answers:***

*1)*

*Line chart*

* *Marks: point, connection*
* *Channels: position (on a common scale), position (on a common scale), colour hue*

*Stacked bar chart:*

* *Marks: line or bar*
* *Channels: position (on a common scale), length or height, colour hue*

*2) if you want to compare the weekly new cases of each continent, then a line chart is better. In the line chart, you can clearly see the weekly new cases of Asia surpassed North America during the week of 1 Jun 2020; however, such information is not that intuitive in the stacked bar chart.*

*Reason: position is a more effective magnitude channel compared to length.*

*3). If your task is to present the global new cases as well as the continenbreakdownwn, then a stacked bar chart is a better choice. You need to manually sum up the cases in each continent if you use a line chart for this task.*

*Or: if the data is very close, it might get very messy in the line chart.*