

FIT5147 Data exploration and visualisation TP2 2020

[Dashboard](#) / [My units](#) / [FIT5147_MO-TP2_2020](#) / [Assessments](#) / [Quiz 1 - Requires Respondus LockDown Browser + Webcam](#)

ch 2020, 10:28 PM

State Finished**Completed on** Sunday, 15 March 2020, 11:28 PM**Time taken** 59 mins 57 secs**Grade** 27.10 out of 36.00 (75%)[Print friendly format](#)

Question 1

Incorrect

Mark 0.00 out of
1.00Which of the following is an example of *ordinal* data?

Select one:

- ☒ a. weight in kilograms ✗ Incorrect: review Background topic of Module 1 in the eTextbook
- ☐ b. gender
- ☐ c. date of birth
- ☐ d. 5-point scale: 'disagree strongly', 'disagree', 'neutral', 'agree', 'agree strongly'
- ☐ e. property boundary coordinates

Your answer is incorrect.

The correct answer is: 5-point scale: 'disagree strongly', 'disagree', 'neutral', 'agree', 'agree strongly'

Question 2

Which of the following is/are correct about visual analytics? (check all that apply)

Select one or more:

- ☒ a. Visual analytics combines data visualisation with machine learning ✓ correct: review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook
- ☒ b. Tableau is a visual analytics tool ✓ correct: review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook
- ☒ c. Visual analytics supports human-in-the-loop analytics ✓ correct: review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook
- ☒ d. Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces ✓ correct: review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook
- ☐ e. The initial application of visual analytics is security analysis

Your answer is partially correct.

You have correctly selected 4.

The correct answers are: The initial application of visual analytics is security analysis, Visual analytics combines data visualisation with machine learning, Visual analytics supports human-in-the-loop analytics, Tableau is a visual analytics tool, Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces

Question 3

Incorrect

Mark 0.00 out of 1.00

Which of the following is **not** a role of visualisation in Data Science?

Select one:

- ☐ a. Looking for trends in a time series.
- ☐ b. Identifying missing data.
- ☐ c. Understanding the provenance of data.
- ☒ d. Presenting the results of a data analysis to government officials. ✖ Incorrect. Presentation of results is one of its main purposes.
Review **What is Data Visualisation Good For?** and try again.
- ☐ e. Checking that data distributions meet the assumptions of a statistics test.

Your answer is incorrect.

The correct answer is: Understanding the provenance of data.

Question 4

Correct

Mark 1.00 out of 1.00

What would be the best method for visualising the relationship between the pressure, temperature and volume of a gas?

Select one:

- ☐ a. Spider diagram
- ☒ b. 3D scatter plot ✔ Correct. This is the best choice
- ☐ c. MDS
- ☐ d. Chernoff faces
- ☐ e. PCA

Your answer is correct.

The correct answer is: 3D scatter plot

Question 5

Correct

Mark 1.00 out of 1.00

If you have wish to compare the sales figures for 5 different categories of products which would be the most appropriate graphic?.

Select one:

- ☐ a. Scatter plot
- ☒ b. Bar chart. ✔ Correct.
- ☐ c. Pie chart
- ☐ d. Gantt Chart
- ☐ e. Histogram

Your answer is correct.

The correct answer is: Bar chart.

Question 6

Mark 0.00 out of 1.00

Which are the typical tasks during the exploration of data? (check all that apply)

Select one or more:

- ☐ a. Computing clusters
- ☐ b. Sorting elements

- ☒ c. Filtering data ✓ review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook
- ☒ d. Visually identifying patterns ✓ review Visual analytics and the role of data visualization topic of Module 1 in the eTextbook

Your answer is partially correct.

You have correctly selected 2.

The correct answers are: Filtering data, Sorting elements, Computing clusters, Visually identifying patterns

Question 7

Correct

Mark 1.00 out of 1.00

Which of the following led to interactive data visualisation?

Select one:

- ☒ a. The introduction of computers. ✓ Correct. The introduction of computers led to interactive data visualisation.
- ☐ b. Need to understand big data sets.
- ☐ c. Invention of statistical charts.
- ☐ d. The Oculus Rift.
- ☐ e. Invention of visual analytics.

Your answer is correct.

The correct answer is: The introduction of computers.

Question 8

Correct

Mark 1.00 out of 1.00

Which of the following is **not** a method for visualising multi-dimensional tabular data? (select any that apply)

Select one or more:

- ☒ a. Box plot ✓ Correct.
- ☐ b. Scatter plot matrix
- ☒ c. Pie chart ✓ Correct.
- ☐ d. MDS
- ☐ e. Small multiples
- ☐ f. Chernoff faces
- ☐ g. Parallel coordinates
- ☐ h. Stacked bar chart

Your answer is correct.

The correct answers are: Box plot, Pie chart

Question 9

Correct

Mark 1.00 out of 1.00

Fill in the following blanks (in order) with the components/files of a Shiny app:

- 1: ____ is the script to assemble inputs into outputs and provide instructions on how to build the visual objects to show with the page.
- 2: ____ is the script to create the visual elements and controls the layout and appearance of the Shiny app

Select one:

- ☐ a. 1: ui.R ; 2: server.R
- ☒ b. 1: server.R ; 2: ui.R ✓ Correct.

Your answer is correct.

The correct answer is: 1: server.R ; 2: ui.R

Question **10**

Correct

Mark 1.00 out of 1.00

What does GIS stand for?

Select one:

- ☒ a. Geographic information systems. ✓ Correct. GIS stands for geographic information systems.
- ☐ b. Global information standards.
- ☐ c. General information services.
- ☐ d. Global indexing systems.
- ☐ e. Geological indexing standards.

Your answer is correct.

The correct answer is: Geographic information systems.

Question **11**

Correct

Mark 1.00 out of 1.00

Which of the following is **not** a kind of curve fitting?

Select one:

- ☐ a. Linear regression
- ☐ b. Total least squares fitting
- ☐ c. Locally weighted polynomial regression (LOESS)
- ☐ d. Multiple linear regression
- ☒ e. Principal components analysis (PCA) ✓ Correct

Your answer is correct.

The correct answer is: Principal components analysis (PCA)

Question **12**

Correct

Mark 1.00 out of 1.00

In ggplot2, the aesthetic properties for the visual elements (i.e., geoms) must be defined individually for each layer, rather than for the whole plot. This statement is:

Select one:

- ☐ True
- ☒ False ✓

Incorrect. Review Chapter 4 of Module 2 in the eTextbook

Aesthetics can be defined in general for the whole plot, or individually for a specific layer.

The correct answer is 'False'.

Question **13**

Correct

Mark 1.00 out of 1.00

Anscombe's Data Quartet ...

Select one:

- ☒ a. demonstrates that four different data sets can generate the same line of best-fit ✓ Correct.
- ☐ b. visualises data with four attributes using a directed graph
- ☐ c. clusters data in four categories
- ☐ d. determines the quartiles of the data - Q1, Q2, Q3

Your answer is correct.

The correct answer is: demonstrates that four different data sets can generate the same line of best-fit

Question **14**

Correct

Mark 1.00 out of 1.00

What does a dendrogram show?

Select one:

- ☒ a. Hierarchical clusters ✓ Correct
- ☐ b. Evolutionary relationships
- ☐ c. Set relationships
- ☐ d. Multivariate data
- ☐ e. Social networks

Your answer is correct.

The correct answer is: Hierarchical clusters

Question **15**

Correct

Mark 1.00 out of 1.00

When creating a data visualisation with the R package **shiny**, in which file would you typically place a call to **ggplot**?

Select one:

- ☒ a. server.R ✓ Correct
- ☐ b. vis.R
- ☐ c. ui.R
- ☐ d. output.R
- ☐ e. ggplot.R

Your answer is correct.

The correct answer is: server.R

Question **16**

Correct

Mark 1.00 out of 1.00

Which would be the best way to see if there was a linear correlation between any two variables in a data set with 4 dimensions?

Select one:

- ☐ a. Use small multiples
- ☐ b. Use a 2D scatter plot with colour and shape to show the other two dimensions
- ☐ c. Use a 3D scatter plot with animation to show the 4th dimension
- ☐ d. Multidimensional table
- ☒ e. Scatter plot matrix ✓ Correct.

Your answer is correct.

The correct answer is: Scatter plot matrix

Question **17**

Correct

Mark 1.00 out of 1.00

Which is a property of connectivity-based clustering methods?

Select one:

- ☐ a. They can only be used with spatial data
- ☒ b. They compute a hierarchy of clusters ✓ Correct

- ☐ c. They can only be used with a single data dimension
- ☐ d. They compute soft or fuzzy clusters
- ☐ e. They assign an item to exactly one cluster

Your answer is correct.

The correct answer is: They compute a hierarchy of clusters

Question **18**

Correct

Mark 1.00 out of 1.00

What does the aesthetic attribute **aes** in ggplot do?

Select one:

- ☐ a. Controls the arrangement of **layers**
- ☐ b. Map data attributes to visual or geometric attributes of a **facet**
- ☐ c. Controls the arrangement of **facets**
- ☐ d. Ensure **geoms** are not overlaid
- ☒ e. Map data attributes to visual or geometric attributes of a **geom** ✓ Correct

Your answer is correct.

The correct answer is: Map data attributes to visual or geometric attributes of a **geom**

Question **19**

Correct

Mark 1.00 out of 1.00

Which would be the best method for visualising the total number of sales at 5 different stores, broken down by category of product?

Select one:

- ☒ a. Stacked bar chart ✓ Correct. It allows ready comparison of the totals
- ☐ b. Compound bar chart
- ☐ c. Spider diagram
- ☐ d. Parallel coordinates for each store with axes for each product.
- ☐ e. MDS

Your answer is correct.

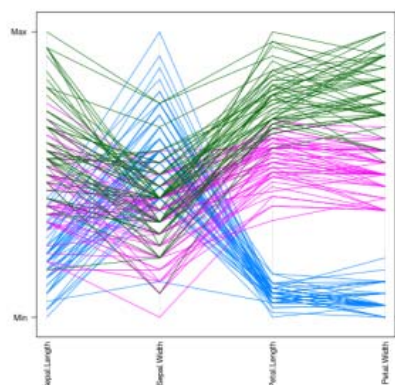
The correct answer is: Stacked bar chart

Question **20**

Correct

Mark 1.00 out of 1.00

What is the name of the following graphic?



Select one:

- ☐ a. Spider diagram

- ☒ b. Parallel coordinates ✓ Correct
- ☐ c. Scatter plot matrix (SPLM)
- ☐ d. Small multiples
- ☐ e. Chernoff faces

Your answer is correct.

The correct answer is: Parallel coordinates

Question 21

Correct

Mark 1.00 out of 1.00

Small multiple plots are produced using which keyword in **ggplot2**?

Select one:

- ☒ a. facet ✓ Correct.
- ☐ b. aes
- ☐ c. group
- ☐ d. multiple
- ☐ e. stat

Your answer is correct.

The correct answer is: facet

Question 22



In ggplot2, a graph is made up of a series of layers. Which of the following options could be considered as a layer? (select all that apply)

Select one or more:

- ☒ a. Lines ✓ correct.
- ☒ b. Axis Title ✓ correct.
- ☒ c. Bars ✓ correct.
- ☐ d. Symbols (circles, squares for data points)
- ☒ e. Legends ✓ correct.

Your answer is partially correct.

You have correctly selected 4.

The correct answers are: Axis Title, Lines, Bars, Symbols (circles, squares for data points), Legends

Question 23

Complete

Mark 1.50 out of 5.00

Consider the following example Shiny app that plots a histogram of R's faithful dataset with a configurable number of bins (i.e., specified by the user):

```
# ui.R

library(shiny)

shinyUI(fluidPage(

  titlePanel("Hello Shiny!"),

  sidebarLayout(

    sidebarPanel(

      sliderInput("bins",

        "Number of bins:",
```

```

        min = 1,
        max = 50,
        value = 30)
    ),
    mainPanel(
      plotOutput("histPlot")
    )
  )
})

# server.R
library(shiny)
shinyServer(function(input, output) {

  output$plot <- renderPlot({
    x    <- faithful[, 2] # Old Faithful Geyser data
    bins <- seq(min(x), max(x), length.out = input$binNum + 1)

    hist(x, breaks = binNum, col = 'darkgray', border = 'white')

  })
})

```

Answer the following questions:

- 1) What kind of output has been specified in ui.R to be displayed in the main panel? A value, a plot or a text field?
- 2) What visual element/s is/are specified by ui.R for gathering dynamic user input?
- 3) ui.R specifies an input variable to store the value of the user input. What is this input variable?
- 4) Assume the code contained in ui.R is correct. Point out possible errors in the corresponding server.R

1. plot (histogram)
2. A slider between 1 and 50 with a default value of 30
3. bins
4. binNum is undefined. It appears that binNum should be the one that is captured to calculate binNum to be used in the histogram.

1) Answer: A plot

2) Answer: Sidebar with a slider input

3) Answer: bins

4) Answer:

- The name of the reactive element "plot" (output\$plot) does not match that specified in ui.R "histPlot"
- The name of the variable "binNum" (input\$binNum) does not match the input variable specified in ui.R "bins"
- Parameter breaks for the hist function should be "bins" rather than "binNum", as otherwise will receive an object "Not found" error, i.e., binNum has not been defined.
- The statement "hist(x, breaks = binNum, col = 'darkgray', border = 'white')" to draw the histogram is outside the renderPlot function, and so will not be re-executed automatically in reaction to user input changes.

Comment:

In Q4, the answer should be more specific.

Question **24**

Correct

Mark 1.00 out of 1.00

Which best describes what multidimensional scaling (MDS) does?

Select one:

- ☐ a. Identifies clusters in multidimensional data
- ☐ b. Identifies the dimensions with least variance in multidimensional data
- ☐ c. Identifies outliers in multidimensional data
- ☒ d. Projects multidimensional data onto two or three dimensions ✓ Correct
- ☐ e. Identifies the dimensions with most variance in multidimensional data

Your answer is correct.

The correct answer is: Projects multidimensional data onto two or three dimensions

Question **25**

Correct

Mark 1.00 out of 1.00

Which of the following plots would you use to check that data has a normal distribution?

Select one:

- ☐ a. Scatter plot
- ☐ b. Bubble chart
- ☐ c. Residual plot
- ☐ d. Box plot
- ☒ e. QQ plot ✓ Correct

Your answer is correct.

The correct answer is: QQ plot

Question **26**

Correct

Mark 1.00 out of 1.00

If you have some data giving the total value of sales for each month of 2014, which would be the most appropriate graphic to visualise this data in order to understand trends?

Select one:

- ☒ a. Line graph ✓ Correct.
- ☐ b. Pie chart
- ☐ c. Scatter plot
- ☐ d. Histogram
- ☐ e. Heat map

Your answer is correct.

The correct answer is: Line graph

Question **27**

Incorrect

Mark 0.00 out of 1.00

Which **geom** would you use to fit a regression line with ggplot?

Select one:

- ☐ a. **geom_least squares**
- ☒ b. **geom_fit** ✗ Incorrect: review Activity: Advanced plots with R topic of Module 2 in the eTextbook
- ☐ c. **geom_smooth**
- ☐ d. **geom_line**
- ☐ e. **geom_errorbar**

Your answer is incorrect.

The correct answer is: **geom_smooth**

Question **28**

Incorrect

Mark 0.00 out of 1.00

Which of the following is/are **not** true? (check all that apply)

Select one or more:

- ☐ a. Tableau is used in business intelligence.
- ☐ b. Tableau allows you to create interactive graphics.
- ☐ c. Tableau uses a drag-and-drop interface to create visualisations.
- ☒ d. Tableau allows you to publish your graphics on the web. **✗** Incorrect. It does. Review **Tools for Data Exploration and Visualisation**.
- ☐ e. Tableau supports a wide variety of statistical tests.

Your answer is incorrect.

The correct answer is: Tableau supports a wide variety of statistical tests.

Question **29**

Correct

Mark 1.00 out of 1.00

Which of the following is **not** an important role for visualisation in data science?

Select one:

- ☒ a. Entertaining policy makers **✓** Correct
- ☐ b. Revealing patterns in the data
- ☐ c. Checking for data entry errors
- ☐ d. Checking that data satisfies assumptions of statistical tests
- ☐ e. Communicating results to stakeholders

Your answer is correct.

The correct answer is: Entertaining policy makers

Question **30**

Complete

Mark 2.50 out of 3.00

Consider the following sample data set "sample-data-for-r-plot3.txt". This data set contains data on the number of staff for a company in Australia between 2010 and 2014:

	A	B	C	D
1	State	Year	Sex	Number
2	ACT	2010	M	7
3	ACT	2010	F	10
4	NT	2010	M	35
5	NT	2010	F	38
6	TAS	2010	M	39
7	TAS	2010	F	40
8	SA	2010	M	45
9	SA	2010	F	50
10	WA	2010	M	50
11	WA	2010	F	46
12	QLD	2010	M	77
13	QLD	2010	F	80
14	VIC	2010	M	98
15	VIC	2010	F	90
16	NSW	2010	M	102
17	NSW	2010	F	100
18	ACT	2011	M	10
19	ACT	2011	F	12

Explain the following code highlighted in red:

```
#####
```

```
library(ggplot2)
```

```
myData <- read.csv('sample-data-for-r-plot3.txt')
```

```
ggplot(myData, aes(Year, Number)) +  
  geom_point(aes(shape = Sex, color = State))
```



plots mydata data set using point determined by the x axis being Year and the y axis being State, the shape variances is controlled/allocated by/using the values in sex and the color is controlled/allocated by/using the values of the state.

In this case, there are two different shapes for Sex (point and triangle) and 8 different colors allocated to different states.

It generates a plot with points from the data 'myData', using Year as the x-axis, and Number as the y-axis. The shapes specify different sexes (male and female), colour specifies the different states.

Comment:

y-axis is Number. It is not the State.

[◀ Quiz 1 - Sample questions](#)

[Jump to...](#)

[Assignment 1 | Project Proposal - Description ▶](#)