DL836
Data Visualisation
Weighting 40%

Assignment 1 Static Visualisation

Creative Computing Year 4
Data Visualisation
Level 8

Instructions

Please place your Tableau and R files in the CA1 folder on sideshowbob on or before **Friday 23rd November 2018**. The Tableau files should be as Tableau packaged workbooks in **.twbx** format with one workbook per question. The R code should be included as an R Studio file. Please include a **screenshot** of all graphics created in the assessment in your report.

Please note the following important guidelines:

- The assessment should be submitted as a hard copy report using a word processing document.
- Report submissions by e-mail will not be accepted.
- Late submissions will incur penalties.
- The assessment must be based on your own work. Checks will be performed to ensure this important criterion is complied with.
- Please ensure that any sources e.g. books, articles etc. consulted are attributed to the author and appropriately referenced in your submitted document.

Creative Computing Year 4
Data Visualisation
Level 8

Question 1

The data in the Excel worksheet **Premium** in the file *ExercisesData(2018).xlsx* records the average premium for a sample of Irish motor insurance policyholders classified by Age and Sex.

i) State giving a reason if this data set is 1D, 2D or MD?

ii) List the name of each variable in the data set and state, giving a reason if they are discrete or continuous.

iii) Using Tableau compute the following graphics:
multiway box plot
multiway dot plot
where the x-axis lists age cohort and the y-axis the average premium.
(Note: gender is not required for this part of the question.)

iv) Using the results of part iii) or otherwise briefly summarise the distribution of premia by age cohort for the industry.

v) Using Tableau compute:

a) Trellis **bar plot** where the **sex** is on the x-axis of each panel with **average premium** on the y-axis while the grouping variable is **age cohort.**

b) A **heatmap** and a **highlight table** using variables listed in v) a).

vi) What is the principal difference between the plot computed in iii) and the plots in v)?

vii) The Directorate of Consumer Protection in the Central Bank want to present just one of the three plots computed in v) as part of their consumer awareness brief to increase pricing transparency and to encourage shopping around for financial products. Assuming the material is aimed at the general public, state, which graph you think the Bank should adopt giving your reason(s) why.

Creative Computing Year 4
Data Visualisation
Level 8

Question 2

The number of motor insurance policies for the top four companies in the Irish market by Cover and Company in 2017 is shown in the table below.

	Cover	
Company Code	**Third Party Fire & Theft**	**Comprehensive**
A	18,717	355,697
B	32,086	59,293
C	40,108	192,669
D	58,825	180,000

i) State giving a reason if this data set is 1D, 2D or MD?

ii) List the name of each variable in the data set and state if they are discrete or continuous.

iii) Using graph paper draw a **mosaic** plot by **hand** using graph paper for this data set placing **Company** on the x-axis.

iv) Using the **R** package **vcd** create a **mosaic** plot placing **Company** on the x-axis.

v) Using iii) summarise the principle features of insurance cover by company.

vi) From a visual inspection of this plot, state, giving a reason why, if you think the Cover is independent of Company?

Creative Computing Year 4
Data Visualisation
Level 8

Question 3

A sample of 822 patients diagnosed with **Influenza** during 2016 and 2017 in the North-East region of Ireland is provided in the Excel worksheet **influenza4D** in ExercisesData(2018).xls

i) List the name of each variable in the data set and state if they are discrete or continuous variables.

ii) Using the **R** package **ggplot2** compute a **Facet** plot using all the available variables in **influenza4D** where each panel of the Facet contains:

a) boxplot
b) violin plot
c) jitter plot

iii) Using ii) and/or otherwise summarise the principle features of influenza in Ireland during this time period.

Question 4

The data in the Excel worksheet cancer3D in ExerciseData(2018).xlsx records cancer type, age cohort and sex for over 50,000 patients diagnosed between 2012 and 2016 in Ireland.

i) List the name of each variable in the data set and state if they are discrete or continuous variables.

ii) State, giving a reason if you think this data set is in long or wide format?

iii) Using the **R** application **ggplot2** create a Facet chart for this data set using all available variables where each panel is a bar chart with age on the x-axis.

iv) List three other graphics that could be suitable for visualising this data configuration.

v) Using iii) summarise the principal features of cancer incidence in Ireland.

4

Question 5

Theme 10 in the CSO Census 2016 file (*CensusED2016.xls*) is based around Education. One of the variables lists citizens aged 15 years and over who have highest level of education recorded as *advanced certificate/completed apprentice*. The variable is called **T10_4_ACCAT** and is located in the excel file *CensusED2016.xls*.

i) Visualise the geographical distribution of this variable by **Electoral Division (ED)** by **joining** the shape file *Electoral_Divisions__CSO_Generalised_20M* with the CSO Census file *CensusED2016.xls*.

ii) Generate a calculated field to visualise the distribution of **T10_4_ACCAT** per 100 ED male county population.

iii) Using results of i) and ii) write a short note on the geospatial distribution of this variable.

Question 6

Select a variable from one of the themes that are of interest to you from the 2016 Census.

i) Visualise the geographical distribution of your chosen variable by joining the shape file: *Small_Areas__CSO_Generalised_20M* with the CSO Census file *CensusSA2016.xls*

ii) Create a visualisation of your chosen variable that is not map based.

iii) Using i) and ii) write a short note on the distribution of this variable highlighting any interesting observations.

Please note that the glossary of themes and variables from Census 2016 are provided in the glossary.xls file in the Mapping Folder.