DL836
Data Visualisation
Weighting 60%

Creative Computing Year 4
Data Visualisation
Level 8

Assignment 2 Dynamic and Interactive Visualisation

Instructions

Please place your Tableau, DataDesk and R files in the CA2 folder on sideshowbob on or before **Friday March 14th 2019**. The Tableau files should be as Tableau packaged workbooks in **.twbx** format with one file per question. The R code should be included in an R Studio file.

Please note the following important guidelines:

- The assessment is to 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 criteria 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

Theme 14 in the CSO Census 2016 examines employment by industry. The variable **T14_1_MIT** records the number of persons in occupations described as Manufacturing industries. The variable is located in the Excel file CensusSA2016.xls.

i)   Visualise the geographical distribution of **T14_1_MIT** by **small area** by joining the shape file **Small_Areas__Generalised_20M** with the CSO Census file **T14_1_MIT**.

ii)   Create a calculated field of **T14_1_MIT** per 100 small area population using the variable **T14_1_TT** and encode the map created in i) by colour using this calculated field.

iii)   Design some interactive features into your map by:

   a) Allowing users to visualise the name of the electoral division the small area is located in, as they mouse around the map.
   b) Creating a county filter that zooms to the selected county when selected by the user.

iv)   a) Create a new worksheet and using the selection tools of Tableau create a group that includes approximately the Greater Dublin Area, Cork City, Waterford City, Limerick City and Galway City.
      Label the group **urbanregions** and label the filter values Dub, Cork, Wford, Lmk and Gal.

   b) Create a filter of **urbanregions** that zooms to each chosen region on selection.

   c) Create a new worksheet with a trellis plot where each panel represents a suitable graphic visualisation of **urbanregions** for **T14_1_MIT** per 100 small area population.

v)   Create a dashboard which includes the plots created in iv). Include the screenshot of your dashboard in your written report.

vi)   Summarise the distribution of **T14_1_MIT** per 100 small area population for the five regions.

Creative Computing Year 4
Data Visualisation
Level 8

Question 2

i) Create the rotating plot shown below (include a copy of the
    plot in your report) using the variables ageDriver, ageFrontPass
    and ageReatrPass provided in the worksheet **accidentdata** in
    ExerciseData(2019). The variables
    correspond to the ages of drivers, front seat and rear seat
    passengers in road accidents involving private cars in Ireland.

ii) Use the lasso tool or otherwise create three selector variables by
    recording approximately the three circled areas shown below. Call
    the recorded selections **Centre**, **LowerLeft** and
    LowerRight.
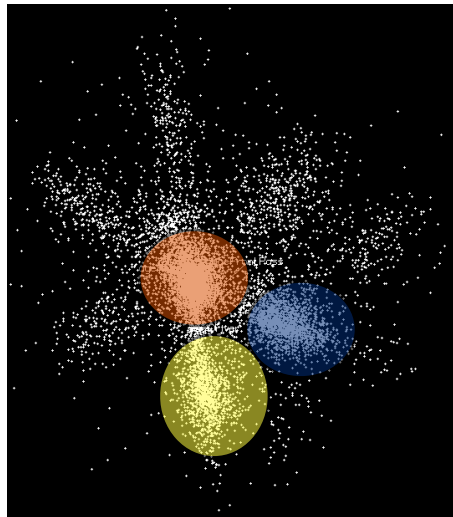
iii)   Using the frequency breakdown table of Segment investigate if there
    are any differences in the patterns associated with each of the three
    clusters by selecting one cluster at a time (by selecting the
    number 1 in each of the frequency breakdown tables for the three
     clusters) and observing the patterns in linked
    visualisations of the following variables provided in
    accidentdata:

ageDriver	ageFrontPass	ageRearPass
GenderDriver	GenderFrontPass	GenderRearPass
prcoltypR	weekdayR	hourR

Contd. Overleaf

Question 2 (Contd)

iv)   Import the graphical displays created in iii) into your report

v)   Write a report on your observations of the road user profile of the three clusters identifying if possible the profile of each cluster.

Question 3

The data in the Excel worksheet **ecoli** in *ExerciseData(2019)* records location, gender, age and diagnosis of three diseases *Cryptosporidiosis*, *Verotoxigenic Escherichia coli infection* and *Giardiasis* during 2016 and 2017 in the Greater Dublin Area. The location data has been jittered by adding noise to the longitude and latitude of each location in the form of a standard normal probability distribution. Please note that the data set is for illustrative purposes only as it excludes some locations for which the longitude and latitude were not located.

a)   Create three worksheets in Tableau for this data set that visualise:

i) geospatial distribution of ecoli in the Greater Dublin Area

ii) Incidence of ecoli over time

iii) gender and age distribution associated with ecoli

b)   Create a dashboard using the worksheets created in a).
Please save your Tableau file as a **.twbx** file.

c)   Using the results of a) and b) summarises the incidence of ecoli in the Greater Dublin Area.

Question 4

The number of patients that experienced Catheter Associated Neoecomial Infection from monthly samples of 50 patients in a large hospital are contained in the worksheet **SPC** in *ExerciseData(2019).*

i)   Compute an nP chart for this data.

ii)   From the plot in i) do you think the monthly changes in infection numbers are explained by random variation only?

iii)   In reaching your conclusion in ii) list what assumptions are you making ?

iv)   State, giving a reason(s) if you think the assumptions in iii) are realistic?

v)   What additional information do you think would be useful in drawing inferences from the SPC plot created in i)

Question 5

The data in the Excel worksheet origindestination in *ExerciseData(2019)* contains the results of an origin destination study of cars conducted on the M50 motorway by Transport Infrastructure Ireland (TII). A total of eight motorway entry/exit junctions were stopped and each driver was given a questionnaire which they then completed at a later date and returned to the TII. One of the questions on the survey was the origin and destination of their journey. This information was further analysed to obtain the postal code of the trip origin and destination as shown in the extract in the table below.

	Postal Code	POSTAL CODE 1	POSTAL CODE 2	POSTAL CODE 3	POSTAL CODE 4	POSTA
1						
2	POSTAL CODE 1	3	0	0	0	
3	POSTAL CODE 2	0	2	0	0	
4	POSTAL CODE 3	0	0	6	0	
5	POSTAL CODE 4	0	0	0	1	
6	POSTAL CODE 5	1	1	0	0	
7	POSTAL CODE 6	2	2	0	3	
8	POSTAL CODE 7	0	1	1	0	
9	POSTAL CODE 8	0	0	0	0	
10	POSTAL CODE 9	4	1	1	1	
11	POSTAL CODE 10	1	0	0	0	
12	POSTAL CODE 11	1	1	0	0	
13	POSTAL CODE 12	1	1	1	0	
14	POSTAL CODE 13	1	2	1	1	

Note rows represent origin postcode and columns the destination postcode. For example, 2 private cars had an origin postal code of 6 and a destination postal code of 1. Zeros represent trip destinations outside of the Dublin postal code area.

i) Create a network visualisation of this data set using the libraries VisNetwork and igraph.

ii) Identify the main flow patterns in the graph created in i).

iii) What additional features could improve the visualisation created in in i)

iv) What additional variables not included in the worksheet origindestination do you think could provide transport planners with a greater understanding of traffic flow rates for vehicles using the M50 as part of their journey.

Question 6

The data in the Excel worksheet tourism in *ExerciseData(2019)* contains seven variables with associated values (shown below) for overseas visitors to Ireland in 2018.

Variable Name	Values of Variable
Purpose	Holiday Trips, Business Trips, VFR Trips
Town	Name of town visited
Longitude	
Latitude	
Country of Residence	British, North American
Brand	Wild Atlantic Way, Irelands Ancient East and Dublin - Breath of Fresh Air
Number	Number of tourists visiting each town

a) Using leaflet create a shiny app that visualises the geospatial distribution of visitors to Ireland. The app should contain the following features:

i) Interactive filters that allow the users to select values from Brand, Country of Residence and Purpose variables.

iii) The name of the town (from the Town variable) when a user mouses over a circle marker in the map interface.

Contd. Overleaf

Question 6 (Contd)

iv)   Circle markers that adjust to the number of visitors. In this regard note that AddCircleMarker in leaflet has an option that allows for the radius of the circles to be assigned a data variable.

b)   The marketing department of Failte Ireland want to develop more targeted advertising campaigns aimed at smaller regional areas. Using the app created in a) filter your view to the *Wild Atlantic Way* brand and identify any Holiday visitor cold spots (i.e. areas that record few visitors) which could be targeted in future advertising campaigns. Include a plot of the locations from the app in your answer.

c)   Failte Ireland are concerned about the UK holiday and UK VFR market arising from Brexit and have received governmental approval for the establishment of a special fund to offset the adverse impacts of Brexit. Using the app created in a) identify geospatial areas that Failte Ireland should prioritise for fund allocation.