Module: Data Visualisation

Problem Set 3

Topics Covered: *Stem and leaf plots, skewed data, box plots*

Exercise 1. The approximate GDP of the Republic of Ireland from 2006 to 2015, in billions of US$, is given by the following data table

2006	**2007**	**2008**	**2009**	**2010**
232	270	275	236	221
2011l	**2012**	**2013**	**2014**	**2015**
240	226	239	256	283

Source: countryeconomy.com

Using this data set, answer the following:

(i) Identify the data source type

(ii) Use a stem and leaf table to represent the data

(iii) Find the median of the data set

(iv) Find the first and third quartiles (Q_1 and Q_3)

(v) Determine if the data set has any extreme outliers

(vi) Determine if the data set has any mild outliers

(vii) Identify the fences for the data set

(viii) Use a box plot the represent this data set

Exercise 2. The fuel consumption of a Volkswagen Golf is measured in litres per 100 kilometer with the following data obained

6.7	5.0	5.0	5.4	5.4	4.5	4.5
5.8	5.8	5.4	6.0	5.5	6.4	5.9
6.9	5.9	5.9	6.3	6.3	7.1	7.1
7.3	7.3	6.7	7.6	7.6	7.6	7.7
7.9	7.9	8.2	8.2	9.7	10.7	

Source: carfuelconsumption.com

Using the data given in this table, answer the following

(i) Identify the data source type

(ii) Use a stem and leaf table to represent the data

(iii) From the stem and leaf plot, determine if the data is skewed or centred

(iv) Find the median of the data set

(v) Find the first and third quartiles (Q_1 and Q_3)

(vi) Determine if the data set has any extreme outlieres

(vii) Determine if the data set has any mild outliers

(viii) Identify the fences for the data set

(ix) Use a box plot the represent this data set

Exercise 3. The hourly lifetime of a sample of 32 light bulbs were measured, with the following data obtained

8	12	22	39	43	43	44	44
45	46	46	47	50	52	54	54
55	56	59	61	63	64	64	68
68	70	71	72	72	76	80	85

Using this data, answer the following:

(i) Represent the data on a stem and leaf plot

(ii) Using this plot determine if the data is central or skewed

(iii) What should be used to measure the centrality and spread of the data?

(iv) What is the interquartile range of the data?

(v) Does the data have outliers?

(vi) Draw a box-plot to represent this data