School of Computing Science and Engineering

Lab exercise-6

Code/Course	**:**	**CSE3020 – Data Visualisation**	**Date**	**:**	**16/01/2018**
Lab Experiments		**Construct the subplot , scatterplot, using ggplot in python**	**Slot**	**:**	**L37+L38**

Pre-requisite: Moderately familiar with basic concepts in python

1. For the given airport dataset do the following
   a. **Pre-Processing : Assign column headers**

Sample output for airport

	id	name	city	country	code	icao	latitude	longitude	altitude	offset	dst	timezone
0	1	Goroka	Goroka	Papua New Guinea	GKA	AYGA	-6.081689	145.391881	5282	10	U	Pacific/Port_Moresby
1	2	Madang	Madang	Papua New Guinea	MAG	AYMD	-5.207083	145.788700	20	10	U	Pacific/Port_Moresby

Sample output for airlines

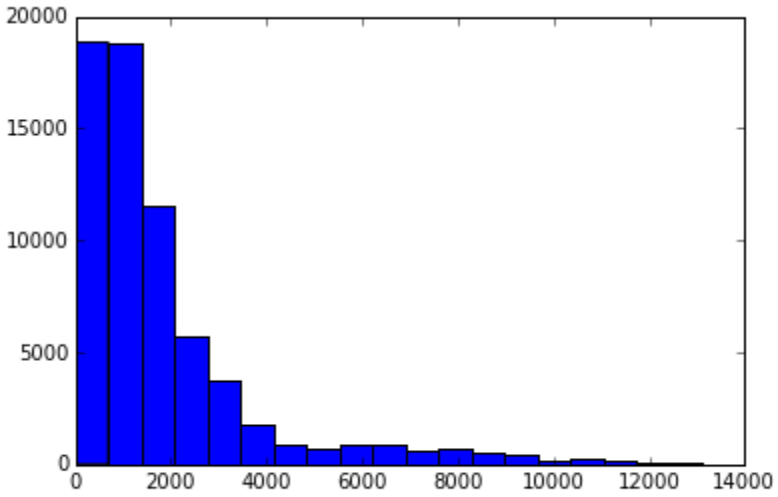
id	name	alias	iata	icao	callsign	country	active	
0	1	Private flight	\N	-	NaN	NaN	NaN	Y

Sample output for route

	airline	airline_id	source	source_id	dest	dest_id	codeshare	stops	equipment
0	2B	410	AER	2965	KZN	2990	NaN	0	CR2
1	2B	410	ASF	2966	KZN	2990	NaN	0	CR2

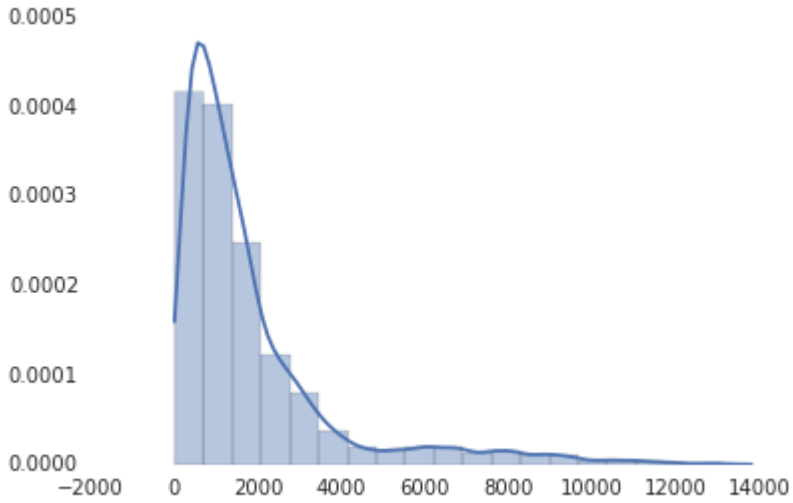
2. **Make histogram for route length, bin the values into ranges and count how many routes fall into each range**

   Sample output



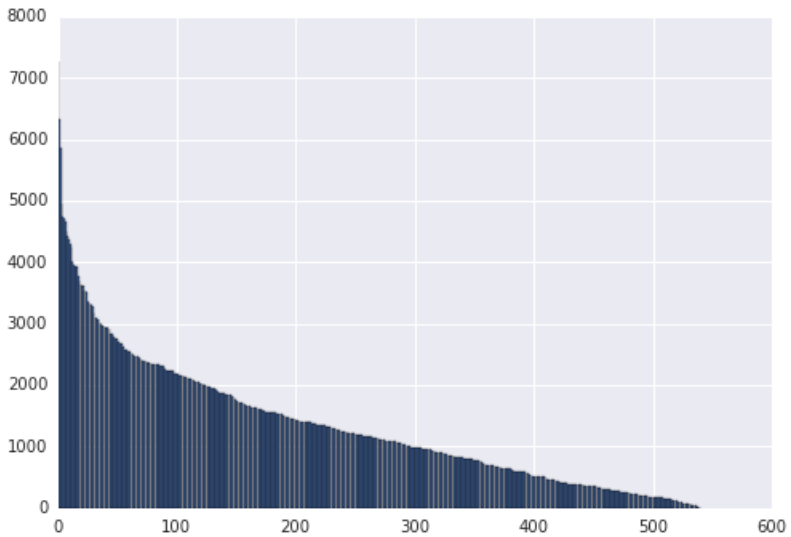
   a. **Use seaborn for route dataset (route Length)**

      Sample output

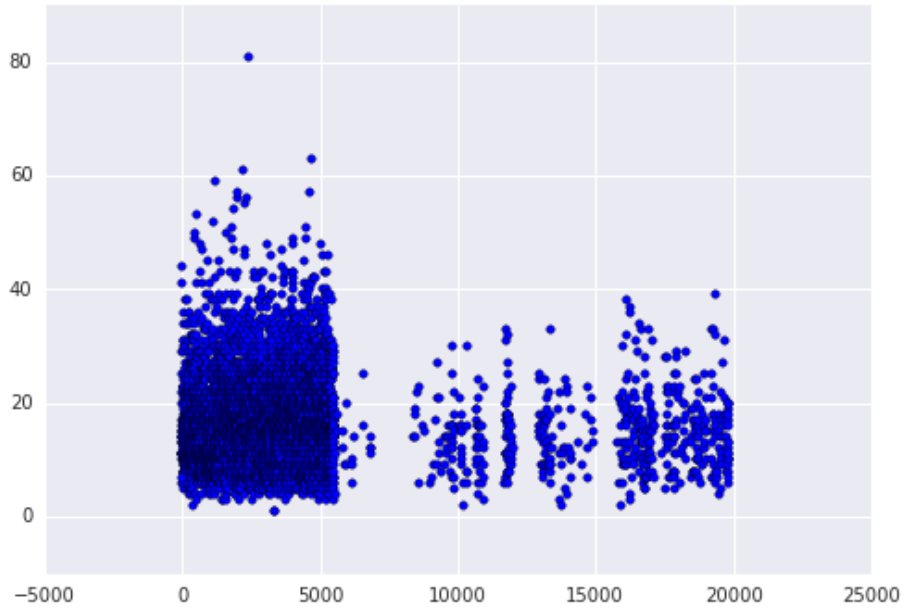


b. Bar chart - plot each airline against the average route length each airline flies

Sample output



c. Create a scatter plot comparing the airline ids to the name lengths


80
60
40
20
0
−5000
0
5000
10000
15000
20000
25000