EGH-404 portfolio 1 Semester 2 2020

By James Norman

Introduction:

this portfolio passement revolves around data and the various ways in can be displayed and manipulated using MATLAB. There were 11 tasks that were to be completed and all 11 will be shown in this document with the code, an explanation on the code and any related graphs or tables.

Question 1:

Question 1 was straightforward and just asked to import the data sheet downloaded on the website to be imported into MATLAB. This was done using the read table function in MATLAB which reads the specified data into a table as shown in figure 1.

```
%%loads the data set into matlab as a table
T = readtable('n10256989.csv');
```

Figure 1. MATLAB code to import a table

Question 2:

Question 2 asked us to remove any outliers from the data set loaded above, using three standard derivations from the mean as the cut off. This task was preformed once again by using an inbuilt MATLAB function by the name of “rmoutliers” which simply removes the outliers from the data set and also takes the a variable mean so as that it makes sure to do it from three standard derivations (figure 2.).

```
%%removes the outliers from the data set using more than three standard deviations from the mean.
B = rmoutliers(T(:,[2 3 4 5 6]),'mean');
```

Figure 2. function to remove outliers

This function is also only used on columns 2-6 because the locations cannot be outliers due to being strings.

Question 3:

Question 3 asked for a table that contained the minimum, max and mean of columns 2-6 this was done by using the min, max and mean functions in matlab. Regex was also used to make sure the sensor B value was not a NAN.

```
%finding the min values
min_vals= min(B{:,1:5});
%finding the max values
max_vals = max(B{:,1:5});
%finding the average values
avg_vals= mean(B{:,1:5});
```

Figure 3. functions to find min, max and mean

Type	temperature (c)	Rainfall (mm)	No. people	Sensor A	Sensor B
Min	25	0	35	-9	0
max	36	110	350	134.6	13.6
mean	30.5	55.1	192.5	62.8	6.154

These functions then create individual tables that can be added together to join all the values in the one table.

Question 4.

In question 4 the task was to create a box and whisker plot for both column three and four this was done by using the boxplot () function in MATLAB (the x and y axis were labelled accordingly)

```
%box and whisker plot for column 3 (rainfall)
figure (1)
L=boxplot(B{:,2});
ylabel('Rainfall (mm)')
xlabel('point locations')
title ('rainfall box and whisker plot')
%box and whisker plot for column 4 (number of people)
figure (2)
M=boxplot(B{:,3});
ylabel('number of people in location')
xlabel('point locations')
title ('people in location box and whisker plot')
```

Figure 4. function to plot box and whisker plots

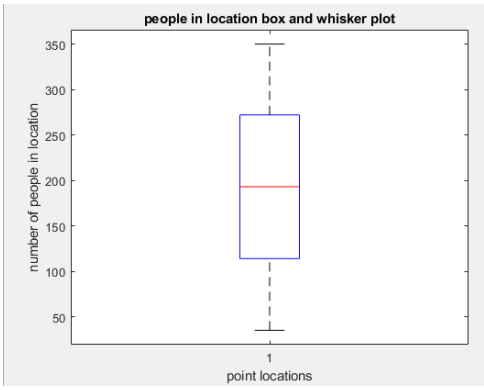

Figure 5. box and whisker plot of location

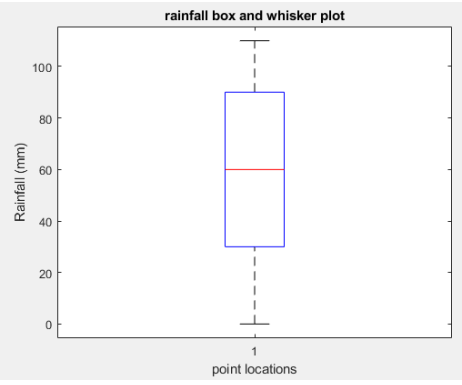

Figure 6. box and whisker plot of rainfall

Question 5.

Question 5's task was to create a scatter plot to look for correlation between column 3 and 6, this was done using the scatter function and column 3 and 6 as variables (plots have also been labelled appropriately)

```
%scatter plot to discern correlations between rainfall(column3) and sensor
%value (column 6)
figure (3)
scatter((B{:,2}),(B{:,5}))
ylabel('Sensor value')
xlabel('Rainfall (mm)')
title ('scatter plot of rainfall depending on sensor value')
```

Figure 7. code to plot scatter graph of column 3 and 6

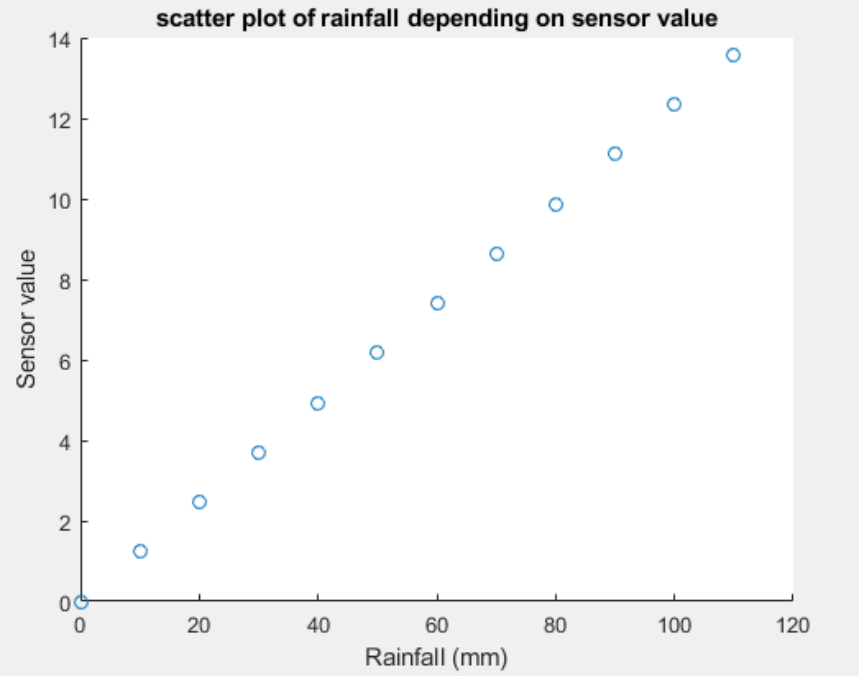

Figure 8. scatter plot graph of column 3 and 6

Question 6.

Question 6 asked that I create a histogram for sensor value A using 10 equal bin sizes, this was done by using the histogram function in MATLAB which also lets you specify the number of bins you want.

```
%histogram of sensor A using 10 equal bins
figure(4)
histogram((B{:,4}),10)
ylabel('Sensor value')
title ('Histogram of values from sensor A')
```

Figure 9. Code to create a histogram of sensor values with 10 equal bins

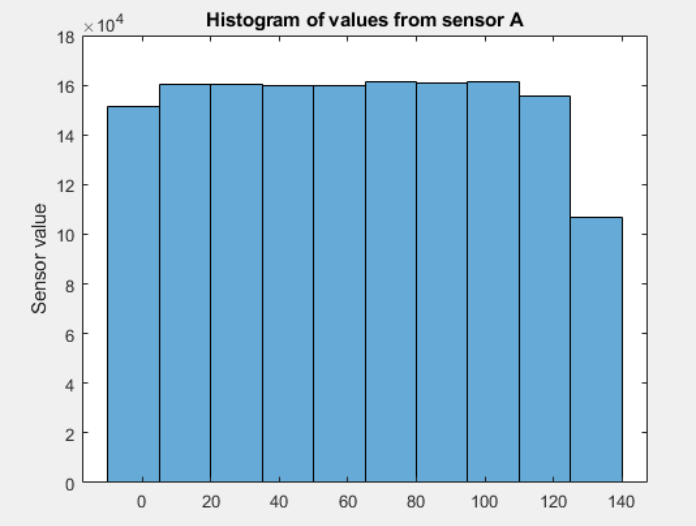

Figure 10. Histogram of sensor values with 10 equal bins

Question 7:

Question 7 wants me to determine the average number of people at each location and then plot the top 100 in descending line graph. This task was a bit trickier than the others but started by grouping the data in the table together to get the locations lined up and then using grpstats() . then after grpstats averaged the values out it was as easy as getting the list into descending order and plotting the top 100 on a graph against a linspace() of 0-100.

```
%Q.7 Determine the average number of people for all locations then plot top 100
%on a descending line graph
Groups = T(:,{'Var1','Var2','Var3','Var4','Var5','Var6'});
Summary = grpstats(Groups,'Var1',{'mean','min','max'});
LocPeople = Summary(:,[1,4,5,9]);
DesLocPeople = sortrows(LocPeople,4,'descend');
Top100People = DesLocPeople([1:100],4);
Ranking = linspace(1,100,100);
figure(5)
plot(Ranking,Top100People{:,1});
ylabel('Number of people')
xlabel('Rank')
title ('Descending line graph of top 100 average people per place')
```

Figure 11. Code to find and plot the average number of people at a place by top100

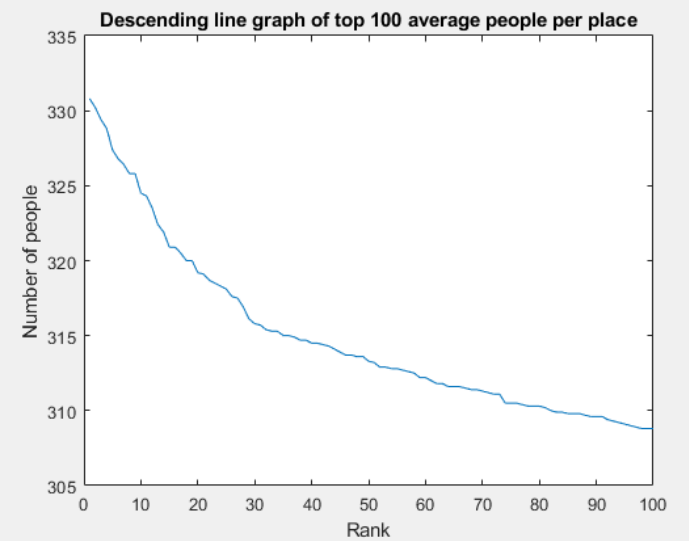

Figure 12. Descending line graph of average people ranked 1-100

Question 8:

In question 8 it was asked that we determine the min and max temperature values for the points in the question above and plot them on a line graph

```
%Q.8 Determine the max and min temperatures and plot for the values for the
%locations from 6 as line graphs
Ranking = linspace(1,100,100);
MinTemp = DesLocPeople([1:100],2);
MaxTemp = DesLocPeople([1:100],3);

figure(6)
plot(Ranking,MinTemp{:,1});
ylabel('Min Temp (c)')
xlabel('Rank')
title('Line graph of minimum temeprature for top 100 locations')

figure(7)
plot(Ranking,MaxTemp{:,1});
ylabel('Max Temperature (c)')
xlabel('Rank')
title('Line graph of maximum temperature for top 100 locations')
```

Figure 13. Code to find and plot the min and max temperatures of the top 100 places

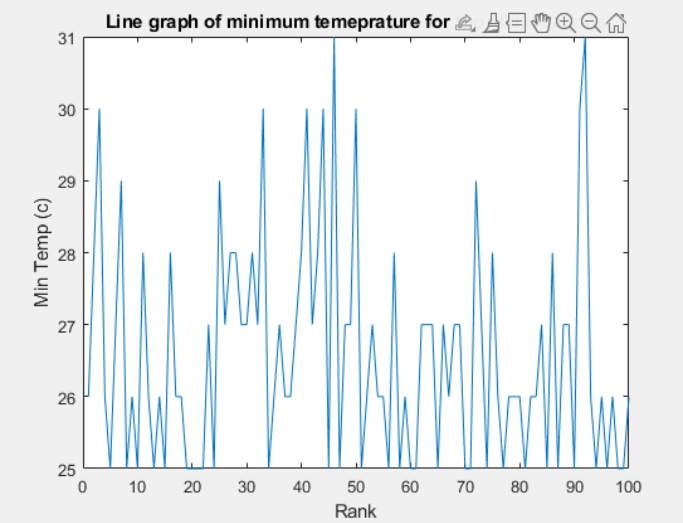

Figure 14. line graph of min temperatures for the top 100

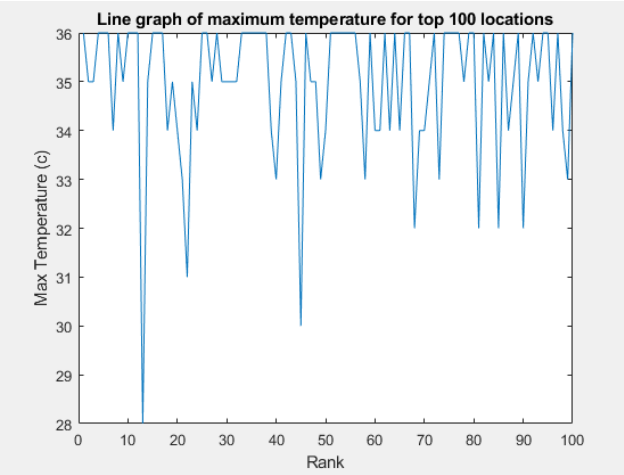

Figure 15. line graph of max temperatures for the top 100

Question 9:

Question 9 asks if there is any correlation between temperature and rainfall, this was done by creating a scatter plot that plotted temperature against rainfall using the scatter function in MATLAB

```
%Q.9 is there correlation between rainfall and temperature explain
figure (5)
scatter((B{:,2}),(B{:,1}))
ylabel('Temperature value (c)')
xlabel('Rainfall (mm)')
title ('scatter plot of rainfalls correlation with Temprature')
```

Figure 16. Code to make a scatter plot of rainfall and temperature

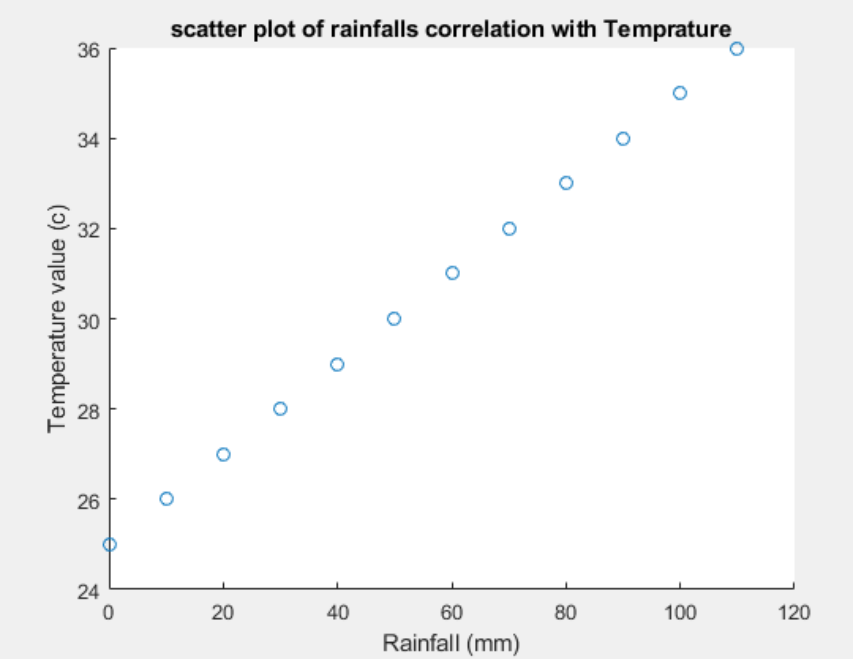

Figure 17. scatter plot of rainfall and temperature

The task also asks for a written explanation which goes as. The scatter plot above shows that the relationship between rain fall and temperature is a linear one. Meaning that when the temperature increases the value of the rainfall also increases and vice versa.

Question 10:

Question 10 once again asks us to find the correlation between two pieces of data except this time it's sensor value B and rainfall, and once again I shall use a scatter plot to display this with the following code.

```
% Q.10 find out if there is correlation between rainfall and sensor value B explain
figure (6)
scatter((B{:,2}),(B{:,5}))
ylabel('Sensor value')
xlabel('Rainfall (mm)')
title ('scatter plot of rainfalls correlation with sensor value B')
```

Figure 18. code to create a scatter plot of rainfall and Sensor value B

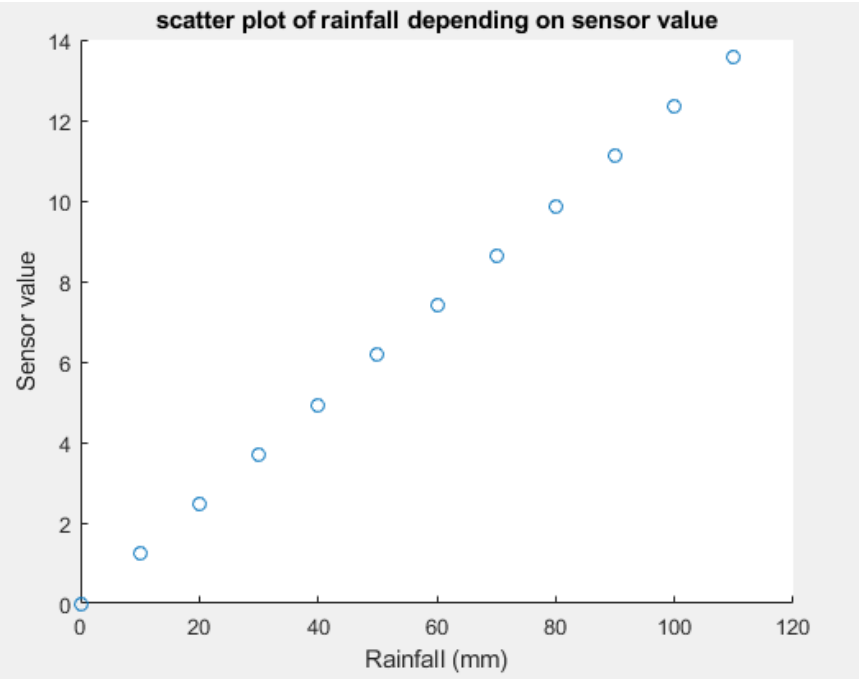

Figure 19. scatter plot of rainfall and Sensor value B

This question also asked for a short explanation on the correlation which goes as such: The scatter plot above shows that the relationship between rain fall and sensor value B is a linear one. Meaning that when the sensor value increases the value of the rainfall also increases and vice versa.

Question 11:

The final questions asks what the expected temperature would be when the rainfall is between 15-25mm, this has been achieved by creating a logical array for when the rainfall is between 15 and 25mm and then comparing that against the data set to determine the average value for when rainfall is at the criteria.

```
%Q.11 what is the expected temprature when rainfall is between 15 and 25mm
L = table2array(varfun(@(x)((x>=15) & (x<=25)), B(:,2)));
Temp_Rain = B(L,:);
avg_temp= mean(Temp_Rain{:,1});
```

Figure 20. Code to determine average temperature when rainfall is between 15 &25 mm

The variable avg_temp returns 27 degrees making it the answer.

Conclusion:

To conclude all the above tasks have been completed using the data set given by the egh404 website and have been documented in both MATLAB code and in this document.