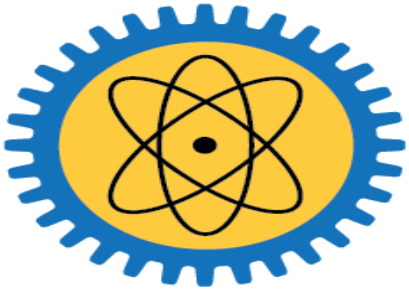


BUSITEMA UNIVERSITY

Pursuing excellence

FACULTY OF ENGINEERING & TECHNOLOGY

ASSIGNMENT REPORT ON APPLICATION OF KNOWLEDGE ACQUIRED FROM MODULES ONE TO FOUR OF COMPUTER PROGRAMMNIG USING MATLAB

PRESENTED TO
THE COMPUTER PROGRAMMING COURSE LECTURER
MR. MASERUKA BENDICTO

By GROUP 17

GROUP MEMBERS

	NAME	REG NUMBER	PROGRAM
1	AHAISIBWE CHRISTOPHER	BU/UP/2024/0824	AMI
2	GIFT EMMANUEL	BU/X/2024/3250	WAR
3	ADWONG CALEB	BU/UP/2024/3813	MEB
4	OWINO POSH IAN	BU/UP/2024/1067	WAR
5	AHEREZA FAITH	BU/UG/2024/2673	APE
6	MUWUBYA WYCLIFFE	BU/UP/2024/1045	WAR
7	NABUKEERA ANNET RIONAH	BU/UG/2024/2676	AMI
8	AKELLO BARBRA	BU/UP/2024/1001	WAR
9	ATATI EDWINE	BU/UP/2024/3730	AMI
10	OPOKA VINCENT	BU/UG/2024/1058	AMI

ABSTRACT

This report details a two-part MATLAB assignment focused on applying data handling and organization skills building from assignment one as we studied from module one to four during lecture time.

From question one of assignment one, we were to use that data to visualize the parameters, patterns trends and relationships saving the plot as an image and labelling it.

Also from question two of assignment one, we were to describe the different statistical characteristics in data and ensure they are visualized and the different attributes per individual are detailed enough to describe them.

ACKNOWLEGEMENT

We thank God Almighty for the gift of life, love and wisdom throughout this course.

Also our gratitude goes to our course lecturer, Mr. Maseruka Bendicto for guiding us in this course that is a necessity in our engineering professions.

And lastly, appreciation goes to all our group members for the commitment and team spirit which simplified work and made it easy for us to complete the task and come up with this report.

DECLARATION

We, Group 17 members hereby declare that report is to the best of our knowledge and has been developed and written by us. It includes the details of the assignment and it has never been presented by any individual.

Group 17

	NAME	REG NUMBER	PROGRAM	SUGNATURE
1	AHAISIBWE CHRISTOPHER	BU/UP/2024/0824	AMI	
2	GIFT EMMANUEL	BU/X/2024/3250	WAR	
3	ADWONG CALEB	BU/UP/2024/3813	MEB	
4	OWINO POSH IAN	BU/UP/2024/1067	WAR	
5	AHEREZA FAITH	BU/UG/2024/2673	APE	
6	MUWUBYA WYCLIFFE	BU/UP/2024/1045	WAR	
7	NABUKEERA ANNET RIONAH	BU/UG/2024/2676	AMI	
8	AKELLO BARBRA	BU/UP/2024/1001	WAR	
9	ATATI EDWINE	BU/UP/2024/3730	AMI	
10	OPOKA VINCENT	BU/UG/2024/1058	AMI	

APPROVAL

This is to confirm that this report has been written and presented by Group 17, giving details of the assignment carried out.

Name .....

Date .....

Signature .....

Table of Contents

Chapter 1:

Introduction

This report gives a detailed overview of the MATLAB assignment built from the two questions of assignment one showing use of the different MATLAB features from module one to module four. The first exercise focused on data manipulation, specifically the process of importing a dataset from Kaggle, structuring it for analysis, and exporting it to an organized Excel workbook. The second exercise involved a more focused task of structuring heterogeneous data by storing a group of members' personal attributes into a single, cohesive variable. In this assignment we were required to demonstrate fundamental skills in data handling, organization, MATLAB.

From question 1;

- ✓ Utilize all the knowledge obtained from module one to four to visualize the different parameters, patterns, trends and relationships.
- ✓ Ensure that each plot is saved as an image its well labelled.

From question 2;

- ✓ Utilize all the knowledge obtained from module one to module four to describe the different statistical characteristics in data.
- ✓ Ensure to visualize them.
- ✓ Ensure that the different attributes collected per individual are detailed enough to describe them.

Chapter 2: SOLUTIONS

From question one 1:

Using the Kaggle Dataset

The primary objective of this exercise was to create a MATLAB script suitable for handling an external dataset. Our approach was a three-step process:

Methodology

Data Acquisition and processing

Using the already cleaned dataset; Indian-Water data that we had used in the first assignment that was obtained from Kaggle inform of excel workbook.

Our task was to come up with a MATLAB code that would return different skills as required.

Below is a list of the different skills we were able to apply in question one

- 3D Visualization
- Bar chart
- Heat map
- Histogram
- Line plot
- Pie chart
- Scatter plot

Below is a code that returns all the above skills.

```
% Input file path
kaggle_data_file = "C:\Users\EDDY\Documents\sub\Indian_water_data.xlsx";
Table = readtable(kaggle_data_file);


% Tables of each year of data
```

```
T_2022 = Table(Table.Year == 2022, :);

T_2023 = Table(Table.Year == 2023, :);


% Process data for each year
struct_arrays = struct();

unique_years = unique(Table.Year);

for i = 1:length(unique_years)

    year = unique_years(i);

    year_data = Table(Table.Year == year, :);

    currentYear = sprintf('Year_%d', year);


    % Store data in struct and Excel
    S = table2struct(year_data);

    struct_arrays(i).(currentYear) = S;

    writetable(year_data,"output.xlsx",'Sheet',currentYear);

end
```

Data Visualization and plotting

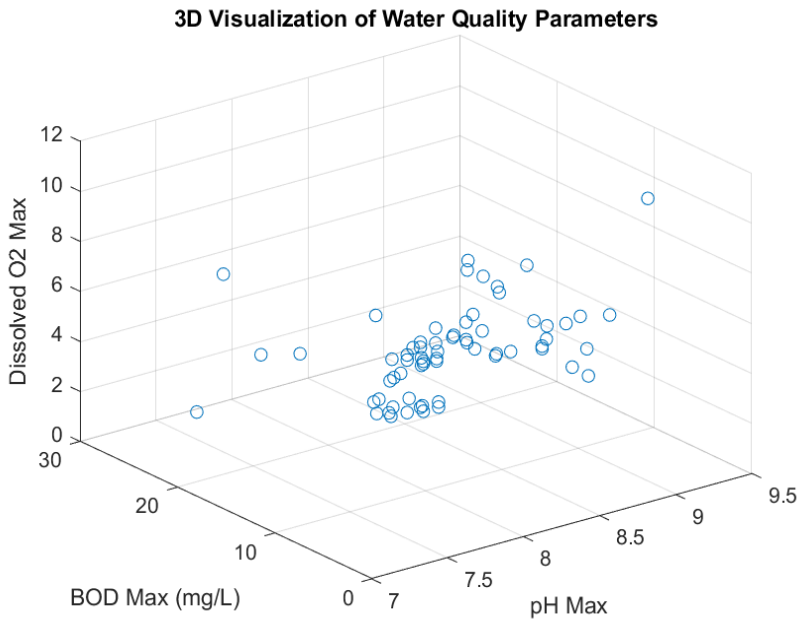
```
% Line plot: Year vs pH, Dissolved Oxygen, BOD

figure;

plot(Table.Year, Table{:,'pH_Max'}, '-o'); hold on;

plot(Table.Year, Table{:,'Dissolved_Max'}, '-s');

plot(Table.Year, Table{:,'BOD_mg_L__Max'}, '-^');

xlabel('Year'); ylabel('Value');

legend('pH Max','Dissolved O2 Max','BOD Max');

title('Trends of Key Water Quality Parameters Over Time');

grid on;

saveas(gcf,'Line_plot.png');


% Bar chart
figure;
```

```
avgState = Table.BOD_mg_L__Max;

bar(avgState);

xticklabels(Table.StateName); ylabel('BOD (mg/L)');

title('Average BOD by State');

xtickangle(45);

saveas(gcf,'Bar_chart.png');


% Histogram of pH values
figure;

histogram(Table.("pH_Max"), 15);

xlabel('pH'); ylabel('Frequency');

title('Distribution of pH (Max Values)');

saveas(gcf,'Histogram.png');


% Scatter plot Dissolved O2 vs BOD
figure;

scatter(Table.("Dissolved_Max"), Table.("BOD_mg_L__Max"), 'filled');

xlabel('Dissolved Oxygen (mg/L)');

ylabel('BOD (mg/L)');

title('Relationship between Dissolved Oxygen and BOD');

grid on;

lsline; % add regression line

saveas(gcf,'Scatter_plot.png');


% Mean values of pollutants
pollutants = [mean(Table.("BOD_mg_L__Max"),'omitnan'), ...

              mean(Table.("NitrateN_mg_L__Max"),'omitnan'), ...

              mean(Table.("TotalColiform_MPN_100ml__Max"),'omitnan')];


labels = {'BOD','Nitrate','Total Coliform'};
figure;
```

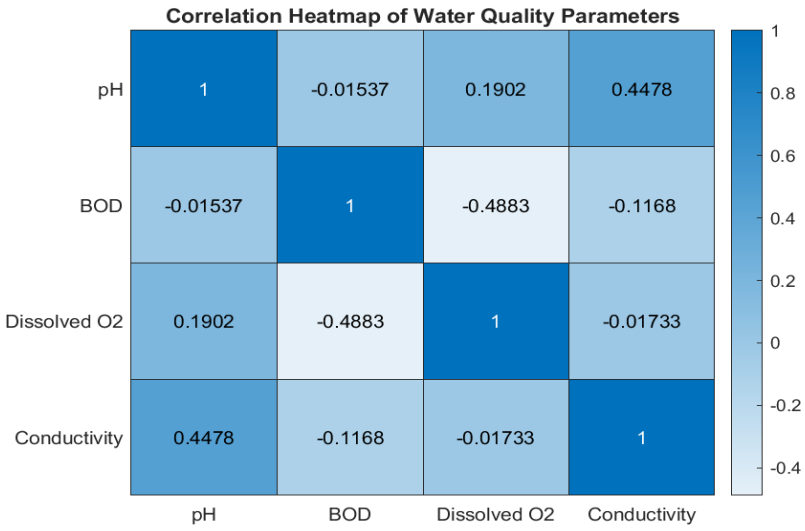
```
pie(pollutants, labels);

title('Proportion of Major Pollutants');

saveas(gcf,'Pie_chart.png');


% Correlation matrix
vars = Table{:,
{'pH_Max','BOD_mg_L__Max','Dissolved_Max','Conductivity___mho_cm__Max'}}; %
Select reliable numeric variables

corrMatrix = corr(vars, 'Rows','complete');

figure;

heatmap({'pH','BOD','Dissolved O2','Conductivity'}, ...

        {'pH','BOD','Dissolved O2','Conductivity'}, ...

        corrMatrix);

title('Correlation Heatmap of Water Quality Parameters');

saveas(gcf,'Heat_map.png');
```

The figures below show a sample of plots from question one.

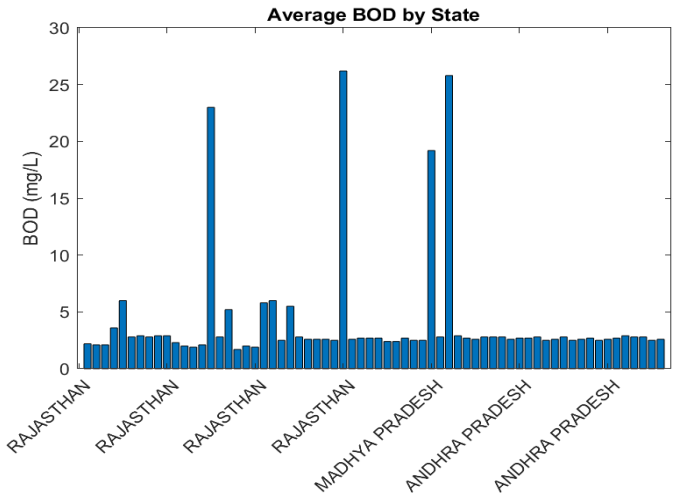
```
 % Scatter plot Dissolved O2 vs BOD
figure;

scatter(Table.("Dissolved_Max"), Table.("BOD_mg_L__Max"), 'filled');

xlabel('Dissolved Oxygen (mg/L)');

ylabel('BOD (mg/L)');

title('Relationship between Dissolved Oxygen and BOD');

grid on;

lsline; % add regression line

saveas(gcf,'Scatter_plot.png');
```




```
% Correlation matrix
vars = Table{:,
{'pH_Max','BOD_mg_L__Max','Dissolved_Max','Conductivity___mho_cm__Max'}}; %
Select reliable numeric variables

corrMatrix = corr(vars, 'Rows','complete');

figure;

heatmap({'pH','BOD','Dissolved O2','Conductivity'}, ...

        {'pH','BOD','Dissolved O2','Conductivity'}, ...

        corrMatrix);

title('Correlation Heatmap of Water Quality Parameters');

saveas(gcf,'Heat_map.png');
```

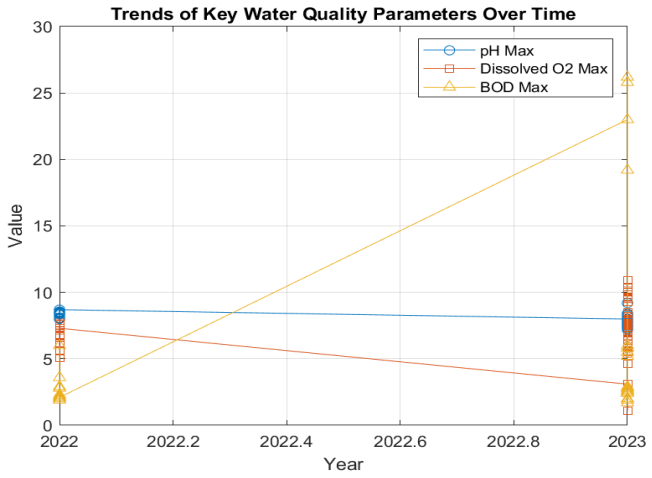



```
% Bar chart
figure;

avgState = Table.BOD_mg_L__Max;

bar(avgState);

xticklabels(Table.StateName); ylabel('BOD (mg/L)');

title('Average BOD by State');

xtickangle(45);

saveas(gcf,'Bar_chart.png');
```

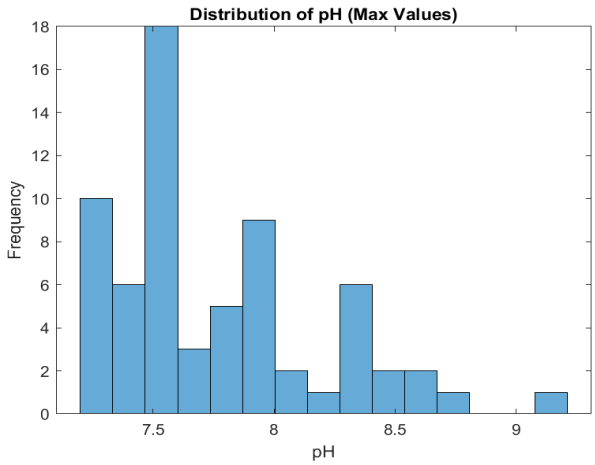



```
% Line plot: Year vs pH, Dissolved Oxygen, BOD
```

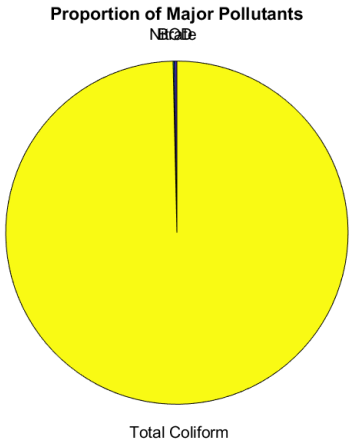
```
figure;
plot(Table.Year, Table{:,'pH_Max'}, '-o'); hold on;
plot(Table.Year, Table{:,'Dissolved_Max'}, '-s');
plot(Table.Year, Table{:,'BOD_mg_L__Max'}, '-^');
xlabel('Year'); ylabel('Value');
legend('pH Max','Dissolved O2 Max','BOD Max');
title('Trends of Key Water Quality Parameters Over Time');
grid on;
saveas(gcf,'Line_plot.png');
```




```
% Histogram of pH values
figure;
histogram(Table.("pH_Max"), 15);
xlabel('pH'); ylabel('Frequency');
title('Distribution of pH (Max Values)');
saveas(gcf,'Histogram.png');
```




```
pie(pollutants, labels);
title('Proportion of Major Pollutants');
saveas(gcf,'Pie_chart.png');
```



From question two:

Using the member attribute data.

This exercise focused on a different aspect of data management: structuring varied data types into a single variable. The task was to take a set of attributes for each group member including home district, religion, tribe, interests, age, name, and a description for facial representation and store them collectively in the first assignment.

In assignment two, using this very data that we collected, we also applied different skills that were acquired from module one to module four,

Below is a list of different skills that were applied to question two of the second assignment by group 17

- 2D bar graph
- 2D pie chart
- 3D bar graph
- 3D pie chart
- Fill plot
- Polar histogram
- Step response plot

Below is a code that returns all the above skills in MATLAB.

```
fprintf('\n=== GROUP 17 MEMBERS ATTRIBUTES ===\n');

% Each group members attributes stored and saved into a single variable

NAME =
["Babra";"wycliff";"Gift";"Chris";"Vincent";"Caleb";"Edwine";"Faith";"Rhionah";"P
osh"];

AGE = [20;25;31;23;22;34;26;21;19;28];

TRIBE =
["Atesot";"Musoga";"Kakwa";"Muganda";"Kejong";"Jap";"Mutoro";"Mukiga";"Munyoro";"
Samia"];

HOME_DISTRICT =
["Amuria";"Kamuli";"Serori";"Mukono";"Moroto";"Tororo";"Kagadi";"Kabale";"Hoima";
"Busia"];
```

```
VILLAGE =
["Acowa";"Achai";"Timo";"Mbalala";"Akwak";"Nagongera";"Busesa";"Kayoli";"Kiriatet
e";"Shaule"];

RELIGION =
["Catholic";"Pentecostal";"Anglican";"Catholic";"Catholic";"Anglican";"Muslim";"A
nglican";"Catholic";"SDA"]

COURSE = ["WAR";"WAR";"WAR";"AMI";"AMI";"MEB";"AMI";"APE";"AMI";"WAR"];

INTEREST =
["Reading";"Praying";"Research";"Reading";"Football";"Bako";"Praying";"Swimming";
"Dancing";"Bako"];

FACIAL_REPRESENTATION = ["brown"; "dark brown"; "dark"; "dark brown"; "fair
dark"; "fair dark"; "optimum brown"; "fair brown"; "brown" ; "fair brown";];



TABLE =
table(NAME,AGE,TRIBE,HOME_DISTRICT,VILLAGE,RELIGION,COURSE,INTEREST,FACIAL_REPRES
ENTATION);
disp(TABLE); % displays table

writetable(TABLE,'Group 17 member attributes.xlsx'); %creates an excell
workbook to save the attributes
```

2. Different statistical characteristic analysis of group data

```
% Age Statistics

fprintf('AGE Statistics: '); %Displays age statistics

fprintf('Mean AGE: %.2f years', mean(AGE)); %calculating the mean age of
10members and fixes it to 2 decimal places

fprintf('Median AGE: %.2f years', median(AGE)); %calculates the median age of
10members and fixes it 2 decimal places

fprintf('AGE Range: %d - %d years', min(AGE), max(AGE)); %minimum and maximum
age

fprintf('Standard Deviation: %.2f years', std(AGE)); %standard deviation of the
ages

%Upper and lower quartiles

quartiles = quantile(AGE, [0.25, 0.75]);

fprintf('Lower quartile: %.2f years, Upper quartile: %.2f years', quartiles(1),
quartiles(2));
```

a). Categorical data distribution

```
% Count unique values for district frequency distribution

fprintf('District Distribution: ');

unique_districts = unique(HOME_DISTRICT); %removing dublicated values leaving
only unique elements

district_counts = groupcounts(HOME_DISTRICT); %how many times a unique value
appears in the data

for i = 1:length(unique_districts) % creates a vector that goes from 1 to the
length of unique districts

    fprintf('  %s: %d members(s)\n', unique_districts{i}, district_counts(i)); %
Prints out unique district with the number of members in each district

end



% Count unique values for course frequency distribution
fprintf('Course distribution: ');

unique_courses = unique(COURSE);

course_count = groupcounts(COURSE);

for i =1:length(unique_courses)

    fprintf(' %s: %d member(s)\n', unique_courses{i}, course_count(i));

end



% Count unique values for religion frequency distribution
fprintf('Religion distribution: ');

unique_religion = unique(RELIGION);

religion_count = groupcounts(RELIGION);

for i =1:length(unique_religion)

    fprintf(' %s: %d member(s)\n', unique_religion{i}, religion_count(i));

end



% Count unique values for facial representation frequency distribution
fprintf('Facial representation distribution: ');
```

```
unique_facial_rep = unique(FACIAL_REPRESENTATION);

facial_rep_count = groupcounts(FACIAL_REPRESENTATION);

for i =1:length(unique_facial_rep)

    fprintf(' %s: %d member(s)\n', unique_facial_rep{i}, facial_rep_count(i));

end


% Count unique values for village frequency distribution
fprintf('Religion distribution: ');

unique_village = unique(VILLAGE);

village_count = groupcounts(VILLAGE);

for i =1:length(unique_village)

    fprintf(' %s: %d member(s)\n', unique_village{i}, village_count(i));

end
```

b). Data Visualization and plotting

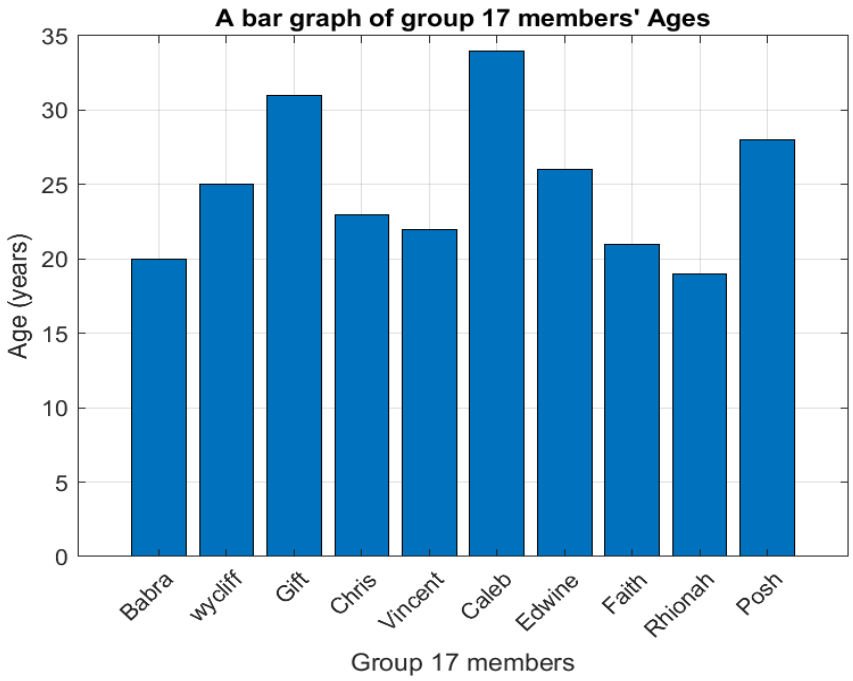
```
%2D bar graph for members' ages

figure; %creates new graphical window for plotting

bar(AGE);

title("A bar graph of group 17 members' Ages");

xlabel('Group 17 members');

ylabel('Age (years)');

xticklabels(NAME); %labels the horizotal axis with the name variables

xtickangle(45); %angle for the horizotal axis name variables

grid on; % enables gridlines for easy visibility

saveas(gcf,'2D_bar_graph.png'); %saves the bar graph as an image in png format


%3D bar plots
figure;

bar3(course_count, 'magenta');

title('3D bar graph of Courses offered');
```

```
zlabel('No_ of members');

yticklabels(unique_courses);

ylabel("Courses");

grid on;

saveas(gcf,'3D_bar_graph.png');


% 2D Exploded pie chart
figure;

exploded = [2,0,1,0]; %pulls AMI with a factor of 2 and MEB with a factor of 1

pie(course_count, exploded,unique_courses);

title('Course distribution');

saveas(gcf,'2Dpie_chart.png');


% 3D exploded pie chart
figure;

expld = [0,0,0,0,0,0,1,0,0,0];

pie3(AGE,expld,NAME);

title('Age distribution');

saveas(gcf,'3Dpie_chart.png');


%fill plot
figure;

fill(1:10, AGE, 'g'); %fills up the area inside the plot with green

title("Group members' Age Fill Plot");

ylabel('Age (Years)');

xlabel('Number of members');

grid on;

saveas(gcf,'fill_plot.png');


%Step respose plot
figure;
```

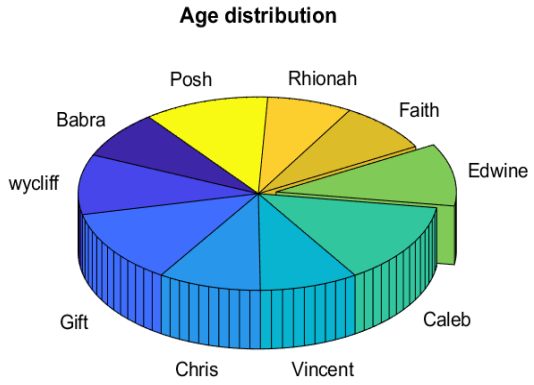
```
plot(AGE);

yline(mean(AGE), 'r--', sprintf('Mean: %.1f',mean(AGE))); %draws a red dotted
line showing the mean age

yline(quartiles(1), 'b--', sprintf('Lower quatile: %.1f',quartiles(1))); %draws
a blue dotted line showing the the lower quatile

yline(quartiles(2), 'm--', sprintf('Upper quatile: %.1f',quartiles(2))); %draws
a magenta dotted line showing the upper quatile

title('Age Data step respose');

xlabel('Number of members');

ylabel('Age (Years)')

grid on;

saveas(gcf,'Step_respose_plot.png');


%polar histogram
figure;

polarhistogram(AGE, 5, 'Facecolor','r');

title('Polar histogram of Ages');

saveas(gcf,'polar_histogram.png');
```

Below are some of the plots from question two

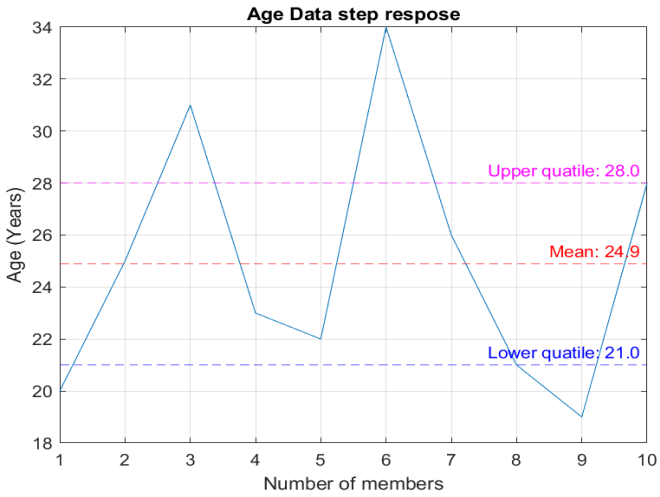
```
%2D bar graph for members' ages

figure; %creates new graphical window for plotting

bar(AGE);

title("A bar graph of group 17 members' Ages");

xlabel('Group 17 members');

ylabel('Age (years)');

xticklabels(NAME); %labels the horizotal axis with the name variables

xtickangle(45); %angle for the horizotal axis name variables

grid on; % enables gridlines for easy visibility

saveas(gcf,'2D_bar_graph.png'); %saves the bar graph as an image in png format
```



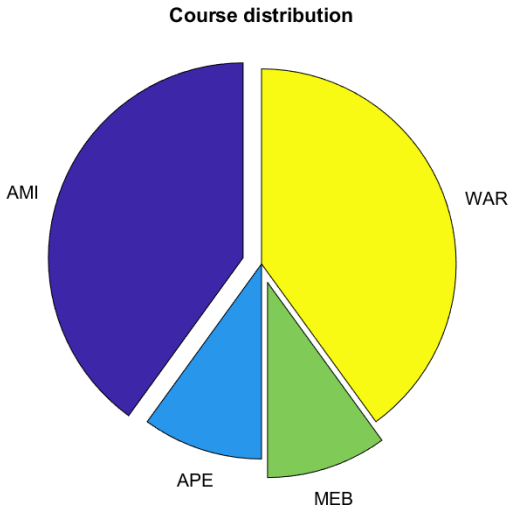

```
% 3D exploded pie chart
figure;

expld = [0,0,0,0,0,0,1,0,0,0];

pie3(AGE,expld,NAME);

title('Age distribution');

saveas(gcf,'3Dpie_chart.png');
```



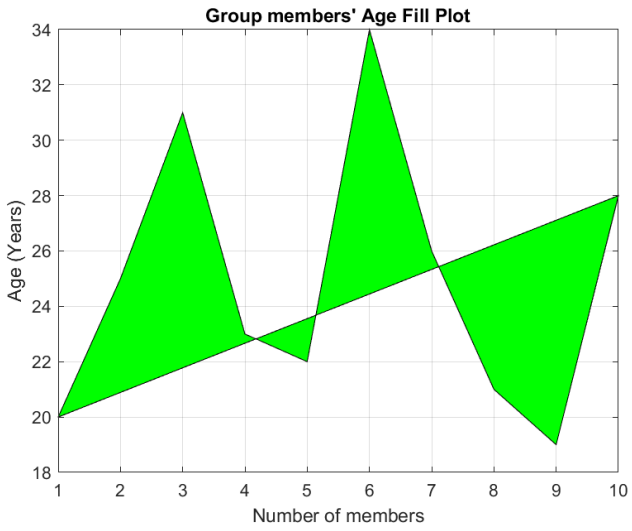

```
%Step respose plot
figure;

plot(AGE);

yline(mean(AGE), 'r--', sprintf('Mean: %.1f',mean(AGE))); %draws a red dotted
line showing the mean age

yline(quartiles(1), 'b--', sprintf('Lower quatile: %.1f',quartiles(1))); %draws
a blue dotted line showing the the lower quatile

yline(quartiles(2), 'm--', sprintf('Upper quatile: %.1f',quartiles(2))); %draws
a magenta dotted line showing the upper quatile

title('Age Data step respose');

xlabel('Number of members');

ylabel('Age (Years)')

grid on;

saveas(gcf,'Step_respose_plot.png');
```



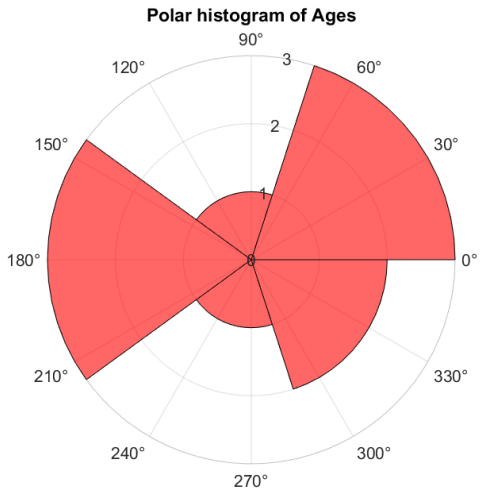

```
% 2D Exploded pie chart
figure;

exploded = [2,0,1,0]; %pulls AMI with a factor of 2 and MEB with a factor of 1

pie(course_count, exploded,unique_courses);

title('Course distribution');

saveas(gcf,'2Dpie_chart.png');
```



```
%fill plot
figure;
fill(1:10, AGE, 'g'); %fills up the area inside the plot with green
title("Group members' Age Fill Plot");
ylabel('Age (Years)');
xlabel('Number of members');
grid on;
saveas(gcf,'fill_plot.png');
```

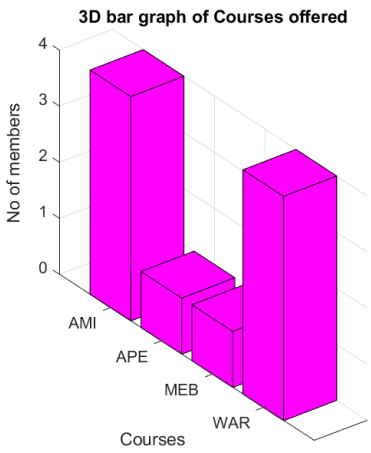


```
%polar histogram
figure;
polarhistogram(AGE, 5, 'Facecolor','r');
title('Polar histogram of Ages');
saveas(gcf,'polar_histogram.png');
```



3D BAR PLOT

```
%3D bar plots
figure;
bar3(course_count, 'magenta');
title('3D bar graph of Courses offered');
zlabel('No_ of members');
yticklabels(unique_courses);
ylabel("Courses");
grid on;
saveas(gcf,'3D_bar_graph.png');
```


3D bar graph of Courses offered
No of members
0
1
2
3
4
AMI
APE
MEB
WAR
Courses

Chapter 3:

Conclusion and Learning Experience

This assignment has helped us to acquire knowledge and experience that helped us understand MATLAB programming concepts and gave us experience with the foundations we had acquired from Modules one to four. We also understood how to handle real-world data given we used the Kaggle dataset.

Generally, this assignment provided a practical foundation in data management and problem-solving within a computing environment.

Chapter 4:
References and Resources

- kaggle.com - source for the dataset used in question one.
- MATLAB Documentation - Used for syntax and function guidance on readtable(), struct(), and writetable().
- Microsoft Excel Cleaning techniques from Microsoft community forums.
- Computer programming lecture notes.