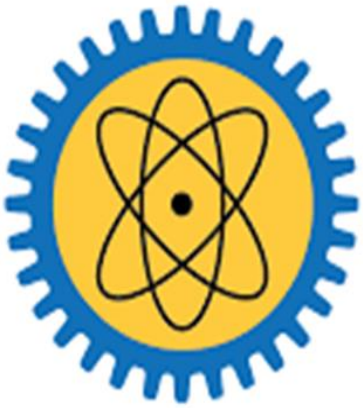


BUSITEMA UNIVERSITY

Pursuing excellence

FACULTY OF ENGINEERING AND TECHNOLOGY

A REPORT ON VISUALIZING PARAMETERS AND DATA INTO MATLAB TO SHOW TRENDS, PATTERNS OR RELATIONSHIPS WHILE INCLUDIN STATISTICAL CHARACTERISTICS

BY GROUP 19

COURSE UNIT: COMPUTOR PROGRAMING

LECTURER: Mr. MASERUKA BENEDICTO

ACKNOWLEDGEMENT

First and foremost, we would like to thank the Almighty God for giving us the strength to carry on with our research in Group 19. We would love to extend our gratitude to all the persons with whose help we managed to make it this far. The willingness of each one of us to invest time and provide constructive feedback has been immensely valuable in this assignment. We wish to extend our gratitude to our lecturer for his consistent guidance and valuable insights throughout this assignment. His teaching and encouragement made it possible for us to understand and practically apply concepts of data importing, organization, and storage in MATLAB.

We also thank our group members for their cooperation and contribution. Each member actively participated in research, coding, and report writing, which ensured the success of this work. Finally, we would like to express our gratitude to all the sources and references that have been cited in this report

ABSTRACT

We started our first meeting for research on 13th, September, 2025 in the university library out of which we were exposed to various concepts and applied a variety of data processing and visualization techniques. Through the use of tables, arrays, structural conversions, and MATLAB plotting functions, the dataset was explored to identify trends, patterns, and relationships among the variables. Different types of plots, including line graphs, bar charts, scatter plots, pie charts, Pareto charts, logarithmic plots, and 3D visualizations were generated.

The second part of the assignment focused on the group members' personal data. Individual attributes such as name, age, gender, course, region, religion, village, and personal interests were loaded into MATLAB then Statistical analysis was then performed to compute characteristics such as the average age or median age. The results were further illustrated using bar graphs, pie charts, and histograms to highlight the diversity within the group. We achieved this through division combined efforts, Different members were assigned different tasks in order to ease the work and to save time.

DEDICATION

We dedicate this report to all the individuals especially Group 19 members, who have been there with us in the process of formulating and compiling this report. To our lecturer Mr. Maseruka Benedicto whose guidance and expertise have been invaluable, your mentorship and insightful feedback have shaped our understanding.

DECLARATION

We hereby certify and confirm that the information in this report is out of our own efforts, research and it has never been submitted in any institution for any academic award.

STUDENT NAME	**REG.NO**	**COURSE**	**SIGNATURE**
NTALE JOASH KATEREGA	BU/UG/2024/2595	WAR	
WANYAMA JOSEPH EROGO	BU/UP/2024/1077	WAR	
KAKULU ALFRED	BU/UP/2024/0981	MEB	
OKORI DARIOUS	BU/UP/2024/1061	WAR	
ALITEMA VICTOR	BU/UP/2024/5098	WAR	
AMASO SUSAN	BU/UP/2025/5435	PTI	
MAATE PHILEMON	BU/UP/2024/0830	AMI	
SIKUKU BELIZER RUTH	BU/UP/2024/0846	AMI	
TWICHIRIZE FLORENCE	BU/UP/2022/1836	WAR	
OKWI NICHOLAS	BU/UP/2024/4457	WAR	

APPROVAL

We are presenting this report which has been written and produced under our efforts. We carried out research on visualizing our data into plots and graphs that are well labeled ready for easy interpretation by the final user.

DATE OF SUBMISSION: …….

SIGNATURE: …….

Table Of Contents

List OF ACRONYMS/ABBREVIATIONS.

MATLAB –Matrix Laboratory.

GUI – Graphics user interface

1 CHAPTER 1: INTRODUCTION

1.1 Background

Matrix Laboratory, or just MATLAB, is an interpreted, high-speed programming language and computational environment employed in technical and scientific contexts. It was first developed in the late 1970s by computer science professor, Cleve Moler, who desired to provide his students with access to sets of mathematical software without their having to learn to program in Fortran themselves.

1.2 Historical Development

Early Development: The initial release of MATLAB, in the latter 1970s, as an interactive matrix calculator, was in Fortran. It consisted of rudimentary matrix operations and was built upon two early numerical libraries, LINPACK, for linear algebra computations, and EISPACK, to solve eigenvalue problems.

Commercialization: The program entered commercial status in 1984, when Moler, in conjunction with Jack Little and Steve Bangert, began MathWorks. This release marked an extensive revision, as it was fully implemented in C and considerably increased in features, including user-defined functions, toolboxes, and graphical user interfaces, significantly broadening the ways in which it could be utilized.

Expansion through Toolboxes: Until the late 1980s, MATLAB had expanded considerably beyond its original limits. The introduction of toolboxes enabled having specialist applications in signal processing and control systems, and others. At this point, MathWorks also added Simulink, which also became a graphical environment to model and simulate in a dynamic state system.

Recent Advances: Since its past updates, MATLAB has also evolved to meet researchers', engineers', and educators' needs as it advances in this direction. New versions have added capabilities, including the Live Editor, which supports combining code, visualizations, and descriptive text in interactive documents. These advances demonstrate how MATLAB has been evolving to become an adaptable infrastructure supporting both research in academics and industrial practice.

2 CHAPTER 2: STUDY METHODOLOGY

2.1 Introduction

It is one of the core competencies of engineering and data science to not only process and organize data but also to interpret and effectively visualize it. It provides a suite of tools through which datasets can be transformed into relevant graphical representations for purposes of analysis and decision-making.

As part of number one for this assignment, we were provided a dataset downloaded from Kaggle. It consisted of multi-year water pollution and disease case data. The necessary condition here was to develop a set of various types of visualization, which could show patterns, distributions, and relationships in the dataset, such as bar charts, scatter plots, histograms, pie charts, Pareto charts, logarithmic plots, and 3D plots consisting of bar and surface plots. Each of these figures, after due labelling, was shown to improve clarity and ease of understanding.

The objectives were:

- To import and see the structure of dataset in MATLAB.
- In order to apply correct plotting functions for 2D and 3D visualization.
- To show frequency, trends, and cumulative effects by means of charts such as histograms and Pareto plots.

In part two, we utilized individual descriptive details of group members, stored in MATLAB as structure variables. This part of the assignment was to create statistics and graphical summary to describe the diversity of the group. Variables like religion, course, region, and interests were examined, and their distributions were represented in bar plots, pies, and heatmaps.

The Objectives were to:

- Load the previously saved variable into our current work/scrip.
- Convert it to table for ease of access of data.
- Describe some statistical characteristics like the median, and mean ages.
- And visualize some of our statistical information in different plots such bar charts, pie-charts, histograms and working heat maps.

This project made us acquainted with MATLAB's potential in statistical analysis and data visualization. The project showed us how unprocessed data, whether from datasets externally obtained or from an individual's records, can be converted to figures in an ordered and readable format. These skills provide an avenue for future works in engineering, research, and decision-making activities.

2.2 Design Process

1. Initial exploration was carried out on the data included checking variable types and ease of which one datatype may be related to another for meaningful plots. Columns such as Year, Region, Country, Contaminant Levels, Diarrheal Cases, GDP per Capita were identified as key for analysis.
2. We organized several meetings during our available time where we went through lecture notes and modules to come up with possible lines of code to put in our script.
3. We inquired from other groups about their progress and refined some of ideas from them.
4. The code for both numbers was written down.
5. Descriptive statistics such as means and sums were computed for numerical columns as well as ease of manipulation of data. Summaries were grouped by region and year to highlight seasonal and geographic patterns.
6. Under Visualization different plots were created to highlight data characteristics: like; Line plots, Scatter plots, Bar plots, pie and Pareto charts among others.
7. Saving Outputs was done to ensure compliance with the last part of the number that required us to save plots as images.
8. Debugging was done in the presence of all members that were available to get a better understanding of how it worked.

3 CHAPTER 3: METHODOLOGY

3.1 NUMBER ONE

Trends of possible data patterns, parameters, relationships between data in the Kaggle data set.

3.1.1 Extracting data from the table

```
wpmain = readtable("water_pollution_disease.csv",ReadVariableNames=true);
```

3.1.1.1 Separating data for years into different tables

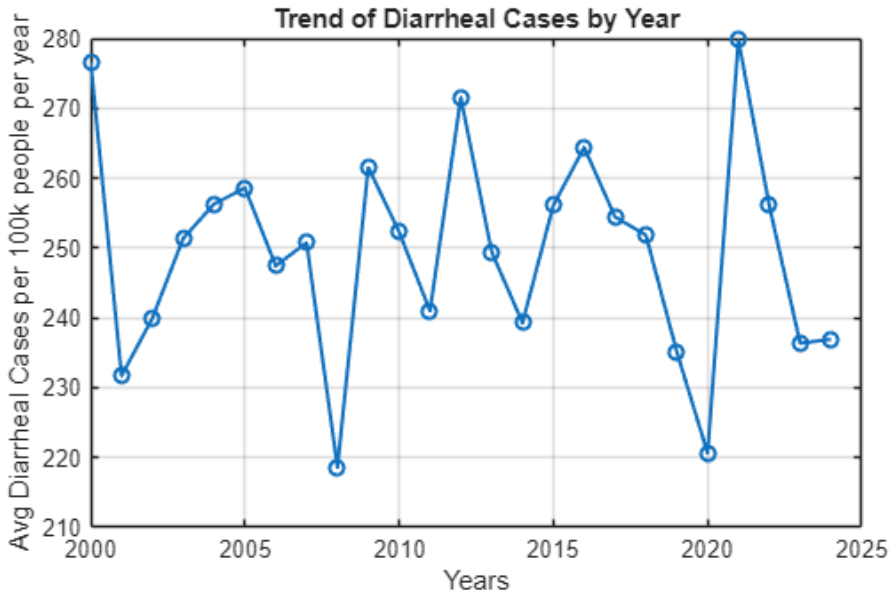
```
wp_2023 = wpmain(wpmain.Year == 2023, :);
wp_2024 = wpmain(wpmain.Year == 2024, :);
```

3.1.1.2 Summarizing data by Year and Including mean Diarrheal Cases

```
YearSummary =
groupsummary(wpmain,"Year","mean","DiarrhealCasesPer100_000People");
```

3.1.2 Line Graph Years by Diarrheal Cases

```
plot(YearSummary.Year, YearSummary.mean_DiarrhealCasesPer100_000People,'-
o',LineWidth = 1.5)
xlabel("Years"); ylabel("Avg Diarrheal Cases per 100k people per year")
title("Trend of Diarrheal Cases by Year")
grid on
saveas(gcf,'Trend of Diarrheal Cases line graph.png')
```



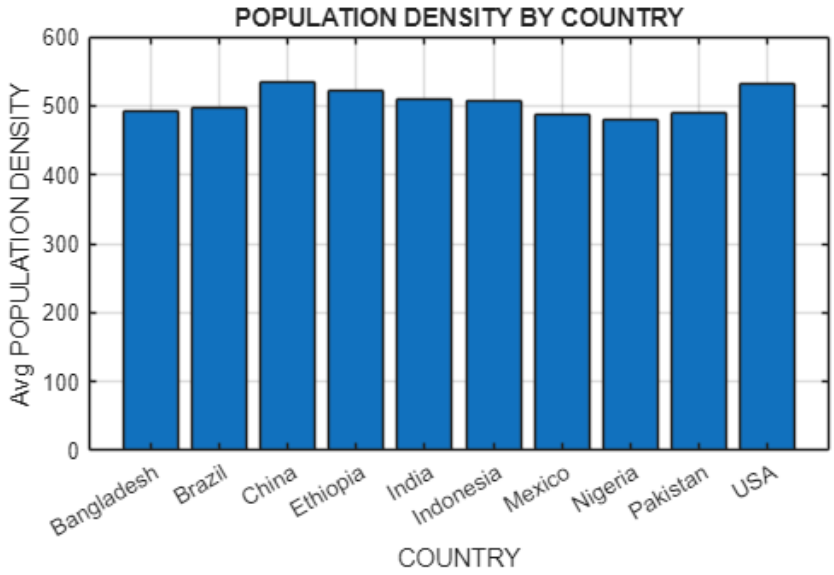
3.1.3 Bar Chart Showing Population Distribution by Country

3.1.3.1 Summarizing data by countries and including Mean Information

```
C_summary =
groupsummary(wpmain,"Country","mean",["PopulationDensity_peoplePerKm__","UrbanizationRate___","SanitationCoverage__OfPopulation_","DiarrhealCasesPer100_000People"]);
```

3.1.3.2 Plotting the graph

```
bar(categories(categorical(C_summary.Country)),C_summary.mean_PopulationDensity_peoplePerKm__)
xlabel("COUNTRY"); ylabel("Avg POPULATION DENSITY")
title("POPULATION DENSITY BY COUNTRY")
grid on
saveas(gcf,'POPULATION DENSITY BY COUNTRY BAR GRAPH.png')
```



3.1.4 Scatter Relationship between Contaminant Level & CHOLERA Cases

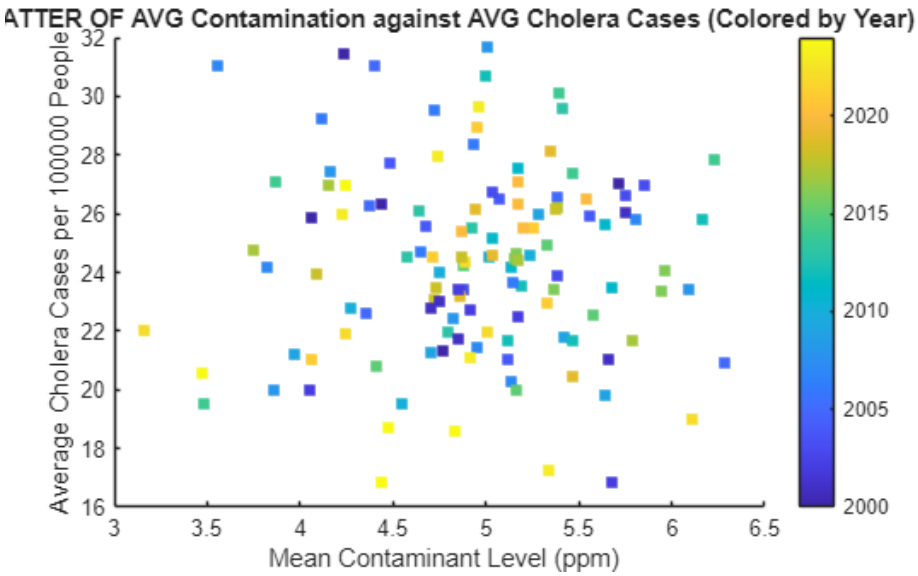
3.1.4.1 Summarizing data by Region and Year while Including Mean Information

```
regionYearSummary =
groupsummary(wpmain,["Region","Year"],"mean",["CholeraCasesPer100_000People","ContaminantLevel_ppm_","BacteriaCount_CFU_mL_","DiarrhealCasesPer100_000People"]);
```

3.1.4.2 Scatter Plot

```
CO = regionYearSummary.mean_ContaminantLevel_ppm_;
CH = regionYearSummary.mean_CholeraCasesPer100_000People;
scatter(CO,CH,40,regionYearSummary.Year,"s","filled")
```

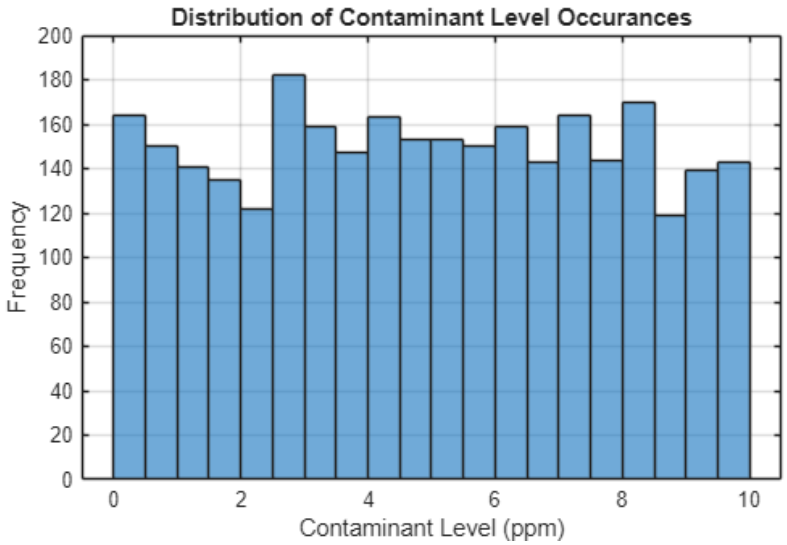
```
 colorbar
 xlabel(" Mean Contaminant Level (ppm)")
 ylabel("Average Cholera Cases per 100000 People")
 title("SCATTER OF AVG Contamination against AVG Cholera Cases (Colored by
Year)")
 saveas(gcf,'Contamination_Cholera Cases_ SCATTER PLOT.png')
```



3.1.5 Distribution of Contaminant Levels

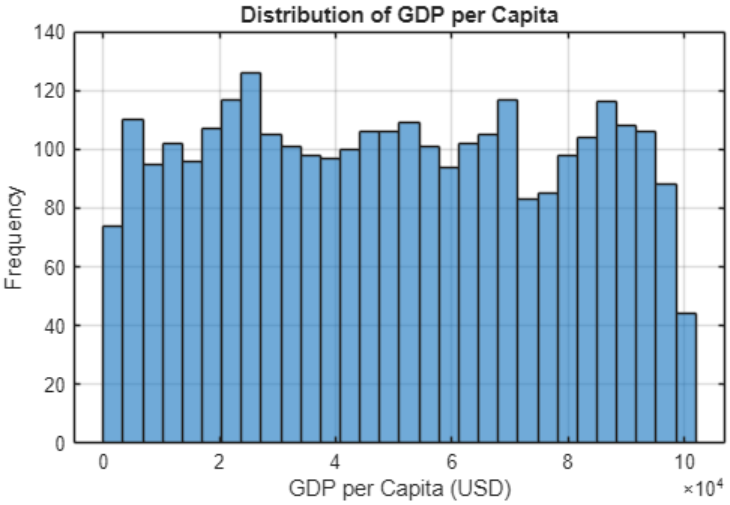
3.1.5.1 Histogram plot 1

```
 histogram(wpmain.ContaminantLevel_ppm_)
 xlabel("Contaminant Level (ppm)")
 ylabel("Frequency")
 title("Distribution of Contaminant Level Occurances")
 grid on
 saveas(gcf,'Distribution of Contaminant Level HISTOGRAM.png')
```



3.1.5.2 Histogram Plot 2

```
histogram(wpmain.GDPPerCapita_USD_, 30)
xlabel("GDP per Capita (USD)")
ylabel("Frequency")
title("Distribution of GDP per Capita")
grid on
saveas(gcf,'GDP HISTOGRAM.png')
```



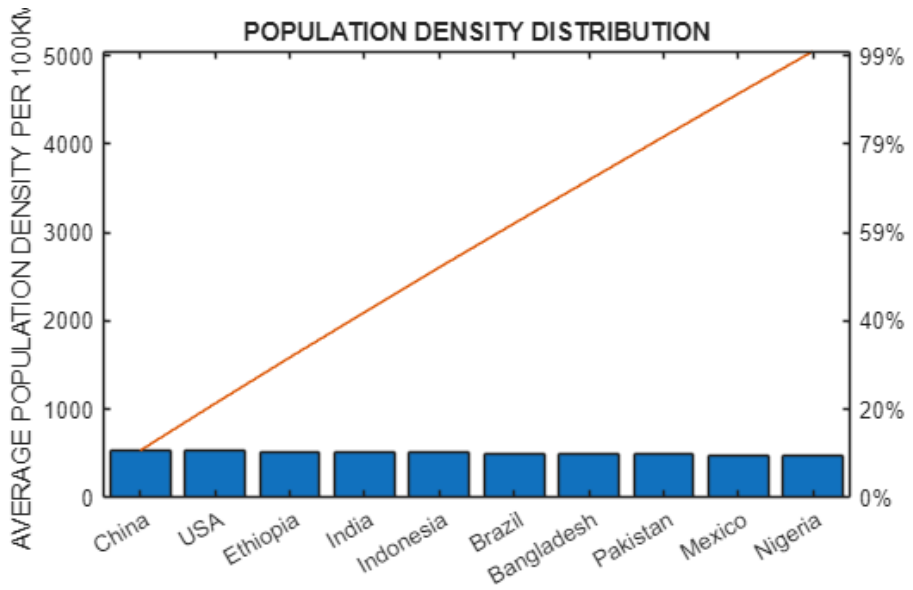
3.1.6 Pareto Chart of Population Density by Country

3.1.6.1 Summarizing data by Country and Including mean Population Density

```
cT= groupsummary(wpmain,"Country","mean","PopulationDensity_peoplePerKm__");
```

3.1.6.2 Pareto Chart Plot

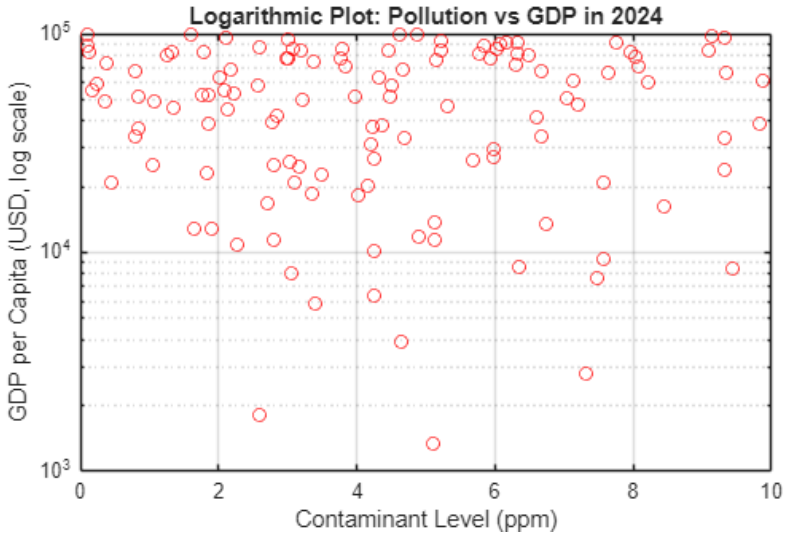
```
pareto(cT.mean_PopulationDensity_peoplePerKm__,cT.Country)
ylabel("AVERAGE POPULATION DENSITY PER 100KM")
title("POPULATION DENSITY DISTRIBUTION")
saveas(gcf,'Population Distribution PARETO CHART.png')
```



3.1.7 Relationship between Contaminant Levels and GDP per Capita

3.1.7.1 Logarithmic Plot

```
semilogy(wp_2024.ContaminantLevel_ppm_,wp_2024.GDPPerCapita_USD_,'r O')
xlabel("Contaminant Level (ppm)")
ylabel("GDP per Capita (USD, log scale)")
title("Logarithmic Plot: Pollution vs GDP in 2024")
grid on
saveas(gcf,'Contamination against GDP LOGARITHMIC PLOT.png')
```



3.1.8 Distribution of Diarrheal Cases by Year & Region (3D PLOTS)

3.1.8.1 Summarizing data by year and region showing mean diarrheal cases.

```
 regionYearSummary2 =
groupsummary(wpmain,["Year","Region"],"mean","DiarrhealCasesPer100_000People");
```

This line of code was used to delete the groupcount column that automatically appears when using groupsummary.

```
 regionYearSummary2.GroupCount = [];
```

We used to “unstack”, to convert the region column to a row.

```
 regionYearSummary3 = unstack(regionYearSummary2,
'mean_DiarrhealCasesPer100_000People','Region');
```

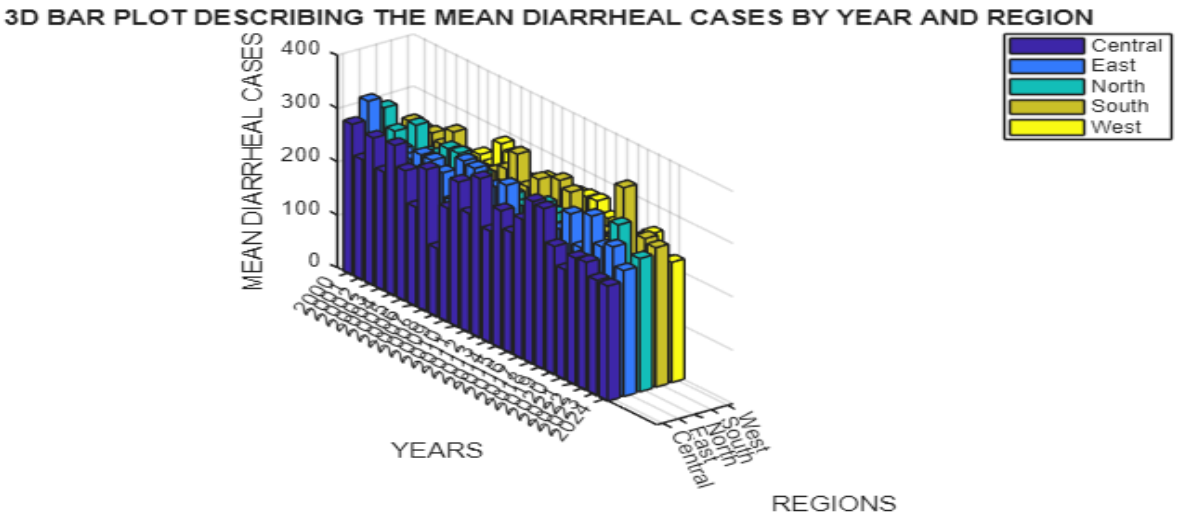
3.1.8.2 Convert the table into numerical array form

```
 t1 = table2array(regionYearSummary3(:, 2:end));
```

3.1.8.3 3D Bar plot

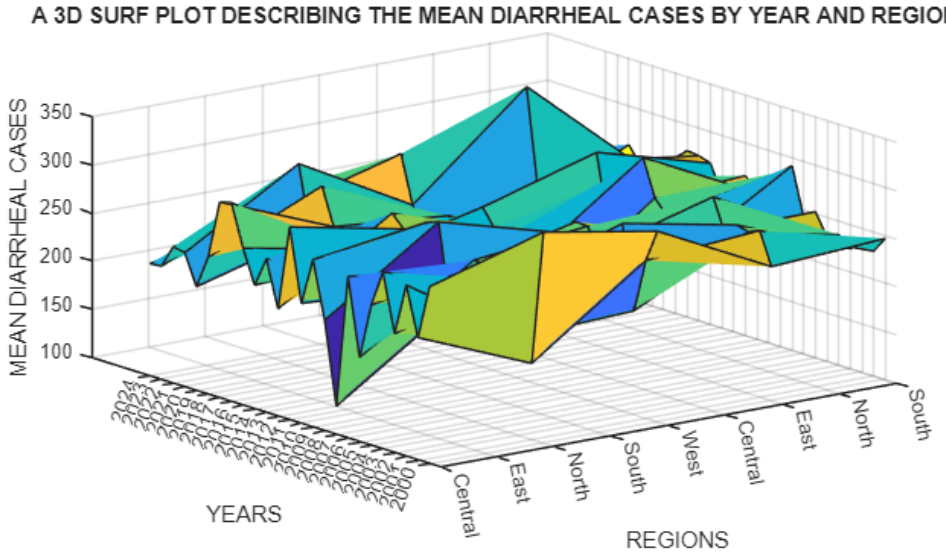
```
 figure(Position=[100,100,1200,650])
 bar3(t1)
 legend(regionYearSummary2.Region)
 xlabel("REGIONS")
 xticklabels(regionYearSummary2.Region)
 ylabel("YEARS")
 yticklabels(regionYearSummary3.Year)
 yticks(1:25)
```

```
zlabel("MEAN DIARRHEAL CASES")
title('A 3D BAR PLOT DESCRIBING THE MEAN DIARRHEAL CASES BY YEAR AND REGION')
saveas(gcf,"MEAN DIARRHEAL CASES BY YEAR AND REGION 3D BAR PLOT.png")
```



3.1.8.4 3D Surface Plot

```
figure(Position=[50,50,1200,650])
surf(t1)
xlabel("REGIONS")
xticklabels(regionYearSummary2.Region)
ylabel("YEARS")
yticklabels(regionYearSummary3.Year)
yticks(1:25)
zlabel("MEAN DIARRHEAL CASES")
title('A 3D SURF PLOT DESCRIBING THE MEAN DIARRHEAL CASES BY YEAR AND REGION')
saveas(gcf,"MEAN DIARRHEAL CASES BY YEAR AND REGION 3D SURF PLOT.png")
```



3.2 NUMBER TWO

STATISTICAL CHARACTERISTICS OF GROUP 19 DATA WITH VISUALISATIONS

3.2.1 Load the previously saved variable into MATLAB

```
load("ASSIGNMENT 1 STUDENT DATA VARIABLE.mat");
```

3.2.1.1 Converting the structural array into tabular format

```
Student_Table = struct2table(Member);
```

3.2.2 Statistical Data

3.2.2.1 Average age

```
Average_age = mean(Student_Table.Age)
```

```
Average_age = 21.8889
```

3.2.2.2 Median age

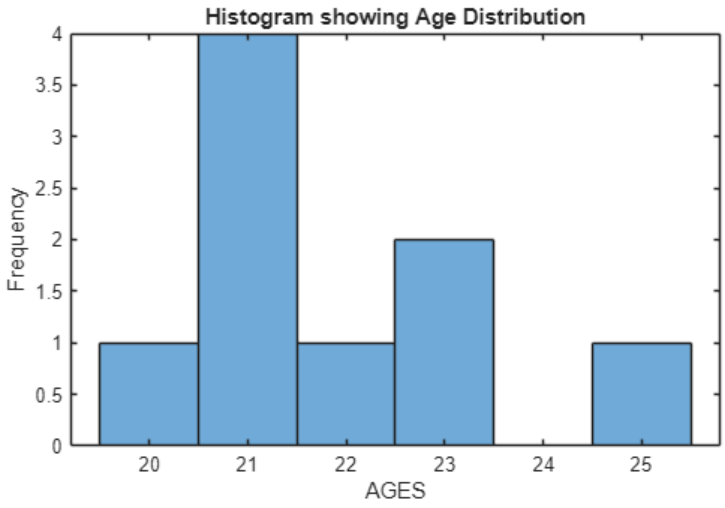
```
median_AGE = median(Student_Table.Age)
```

```
median_AGE = 21
```

3.2.3 Visualizations

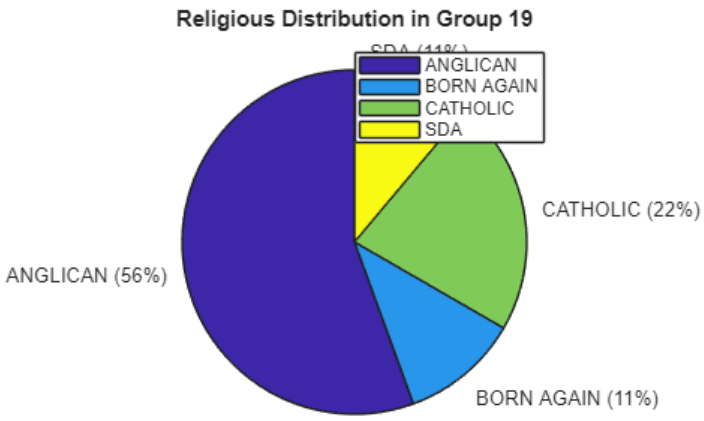
3.2.3.1 Histogram showing Age distribution amongst the group.

```
histogram(Student_Table.Age)
xlabel("AGES")
ylabel("Frequency")
title("Histogram showing Age Distribution")
```



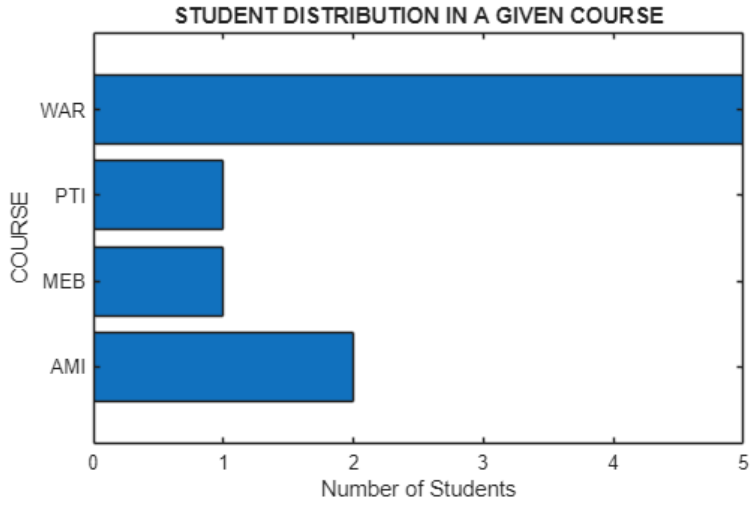
3.2.3.2 Pie chart showing Religion Distribution

```
pie(categorical(Student_Table.Religion))
legend
title("Religious Distribution in Group 19")
```



3.2.3.3 Horizontal Bar Chart showing Students per Course

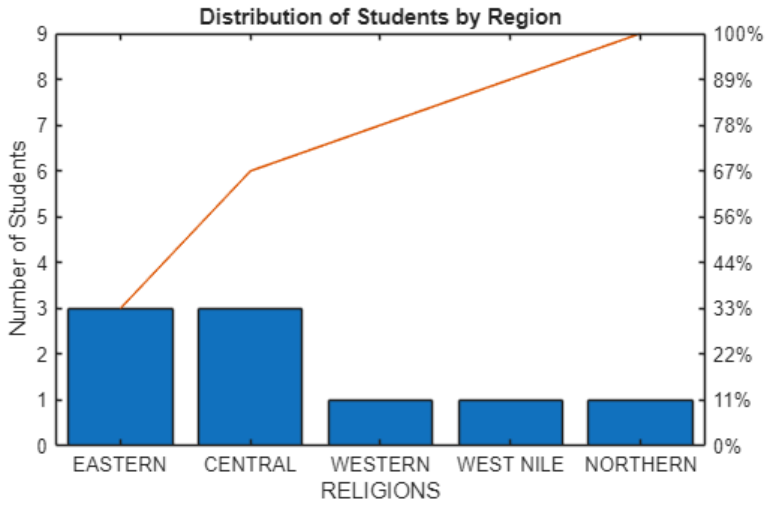
```
barh(categories(categorical(Student_Table.Course)),
countcats(categorical(Student_Table.Course)))
xlabel("Number of Students")
ylabel("COURSE")
title("STUDENT DISTRIBUTION IN A GIVEN COURSE")
```



3.2.3.4 Pareto chart showing regional distribution of students

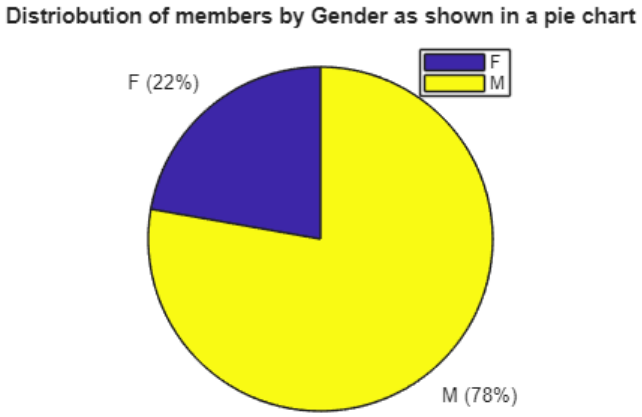
```
pareto(countcats(categorical(Student_Table.Region)),categories(categorical(Stud
ent_Table.Region)))
```

```
xlabel("RELIGIONS")
ylabel("Number of Students")
title("Distribution of Students by Region")
```



3.2.3.5 Pie Chart showing Distribution of Gender

```
pie(categorical(Student_Table.SEX))
legend
title("Distriobution of members by Gender as shown in a pie chart")
```



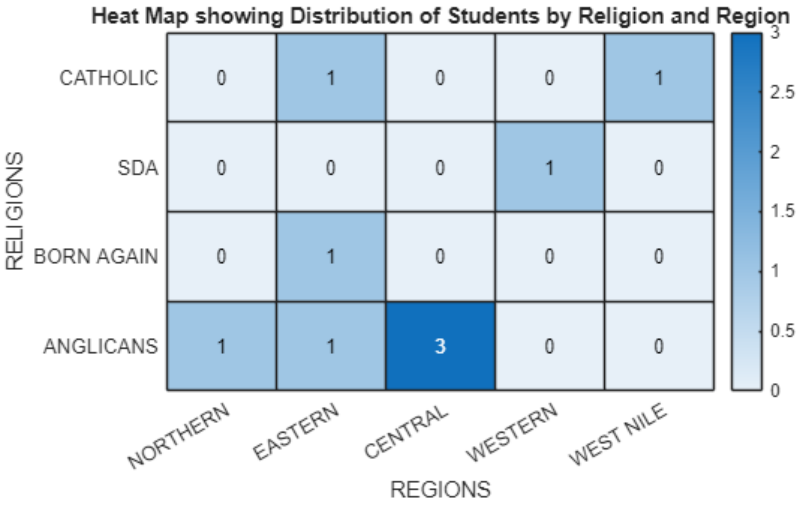
3.2.3.6 Heat Map showing Distribution of Students by Religion and Region

3.2.3.6.1 Construction of Matrices for Heat map plot

```
Rdata = [0 1 0 0 1; 0 0 0 1 0; 0 1 0 0 0; 1 1 3 0 0];
RELIGION = ["CATHOLIC", "SDA", "BORN AGAIN", "ANGLICANS"];
REGIONS = ["NORTHERN","EASTERN","CENTRAL","WESTERN","WEST NILE"];
```

3.2.3.6.2 Heat Map Plot

```
heatmap(REGIONS,RELIGION,Rdata)
xlabel("REGIONS")
ylabel("RELIGIONS")
title("Heat Map showing Distribution of Students by Religion and Region")
```



4 CONCLUSION

The project was successfully executed through collaboration and step-by-step implementing of ideas in MATLAB. In number one demonstrated the ability of transforming an enormous outside dataset on water pollution and disease into types of pictorial presentation, ranging from simple 2D graphs to complex 3D graphs. These graphics facilitated achieving better perception of the trends, distributions, and overall effect dominating the dataset.

In assignment two, personal descriptive details of our group were maintained systematically and programmatically assessed under MATLAB struct array data types. Age distribution, enrollment in courses, religion, and personal interests were analyzed and represented graphically. This assignment enabled us to comprehend how descriptive characteristics can be combined and compared.

We faced a major challenge on trying to summarize our work so as to have meaningful plots since we had a very large dataset of about 3000 rows. Luckily this is where we applied knowledge of here `groupsummary` function from MATLAB to combine repeated information.

5 REFERENCES

- *MATLAB Documentation* Retrieved from: https://www.mathworks.com/help/matlab/
- Kaggle Datasets. *water_pollution_disease Dataset*
- Kirani Singh, T., & Chanduri, B. B. *MATLAB Programming*. PHI Learning Pvt. Ltd.
- Course Notes of Modules 1-4: Computer Programming with MATLAB, Mr. Maseruka Benedicto