HOMEWORK 2

SETUP

The program was run and compiled in Ubuntu 18.04 but should work in most other versions. The following commands were done in order to get the program working:

```
> sudo apt-get update
> sudo apt-get install freeglut3
> sudo apt-get install freeglut3-dev
> sudo apt-get install binutils-gold
> sudo apt-get install g++ cmake
> sudo apt-get install libglew-dev
> sudo apt-get install g++
> sudo apt-get install mesa-common-dev
> sudo apt-get install build-essential
> sudo apt-get install libglew1.5-dev libglm-dev
> sudo apt-get install mesa-utils
> sudo apt-get –f install
```

PROGRAM COMPILATION AND RUNNING

The following command should be executed to compile the program:

```
> g++ scatterplot.cc -lm -lglut -lGL -lGLU -o scatterplot
```

This should create a "histogram" program which is ran doing the following command:

```
> ./scatterplot
```

SOLUTIONS/SCREENSHOTS

This dataset is a wine dataset with many different types of variables taken from http://archive.ics.uci.edu/ml/machine-learning-databases/wine/wine.data which has almost 200 different data points. The variables chosen to look at is the amount of alcohol in the wine, the wine colour intensity, and the hue of the wine. As seen in the screenshot below we can see that scatterplots shown in a 3x3 matrix. The off diagonals are transpose of each other so technically a user should look at the 3 bottom left scatterplots or 3 top right scatterplots since they represent the same data. The main diagonal is not shown since it would just be the data represented against itself. The correlation coefficient (labeled as ‘r’) and covariance (labeled as ‘cov’) are shown above each scatterplot. Obviously, the correlation coefficient on the main diagonal is 1 (small deviation is caused by floating point precision error).

The correlation coefficient of wine alcohol vs wine colour intensity is 0.546354 which means the data is somewhat positively correlated which seems evident in the screenshot. For wine alcohol vs wine hue, we see that the correlation coefficient in near 0. This is strongly impacted by the

huge clustering of data points in a general location and a lone data point being an outlier which greatly impacts the correlation coefficient. Lastly, wine hue vs wine colour intensity has a correlation coefficient of -0.521813 which means it is somewhat negatively correlated which makes sense since drawing an imaginary line through the middle of the data points seems to point in a downward y-axis direction.

Lastly, we have the covariance which along the main diagonal is just known as the variance. The wine alcohol and wine colour intensity covariance is 1.028283 which indicates that both dimensions increase or decrease together. The covariance of the wine alcohol and wine hue is almost 0 which means that the two variables are independent of each other. Looking at the scatterplot this seems strange since they seem somewhat clustered together, but this is true since the cluster is quite big but even more so the outlier at wine hue = 1.71 greatly affects the covariance. Wine hue vs wine colour intensity is -0.276506 which means that one variable increases while the other decreases.