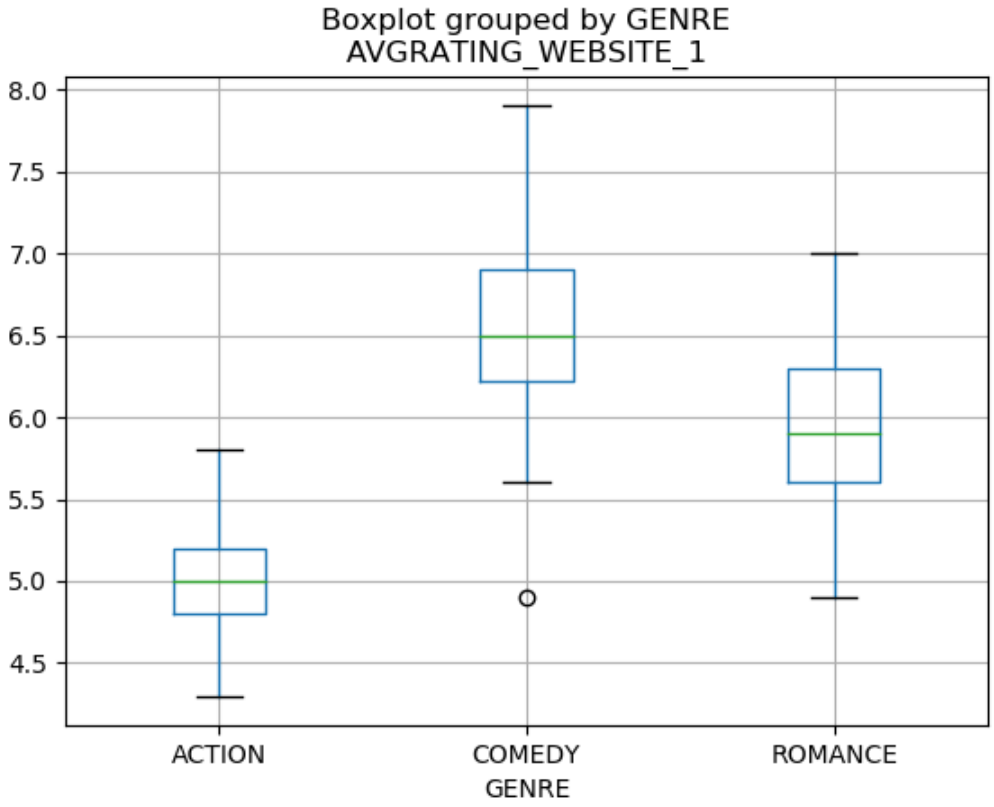
Kailee Madden
HW 1 – Programming

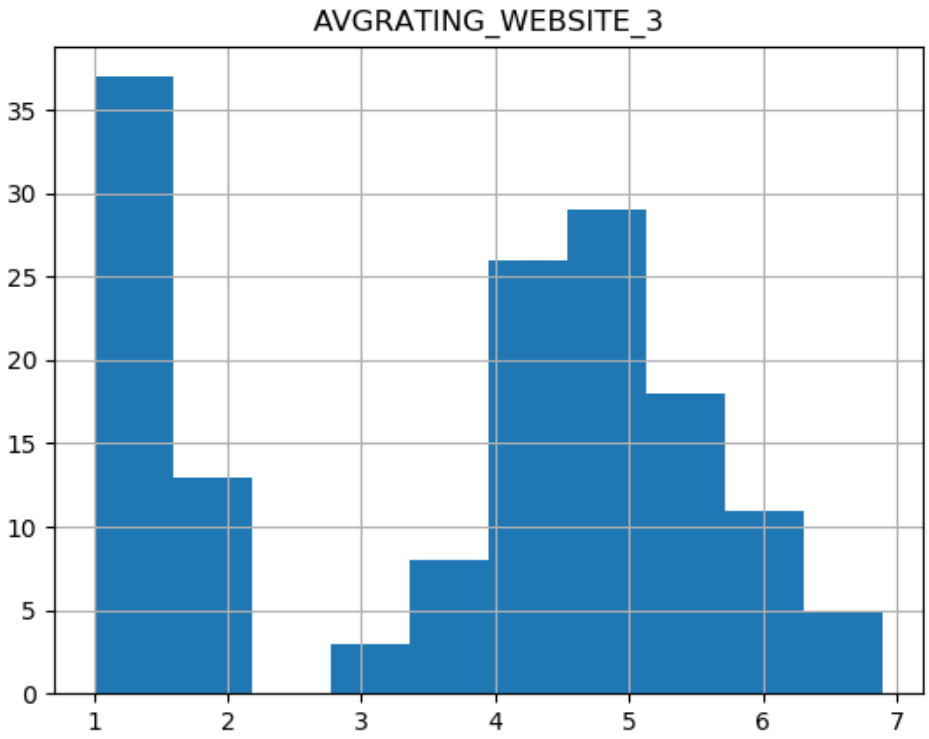
Data Description
Maximum z-scores (in order from website 1 to website 4):
2.49202, 3.09078, 1.78583, 1.7131
Minimum z-scores (in order from website 1 to website 4):
-1.87002, -2.43395, -1.56758, -1.44955

Data Visualization

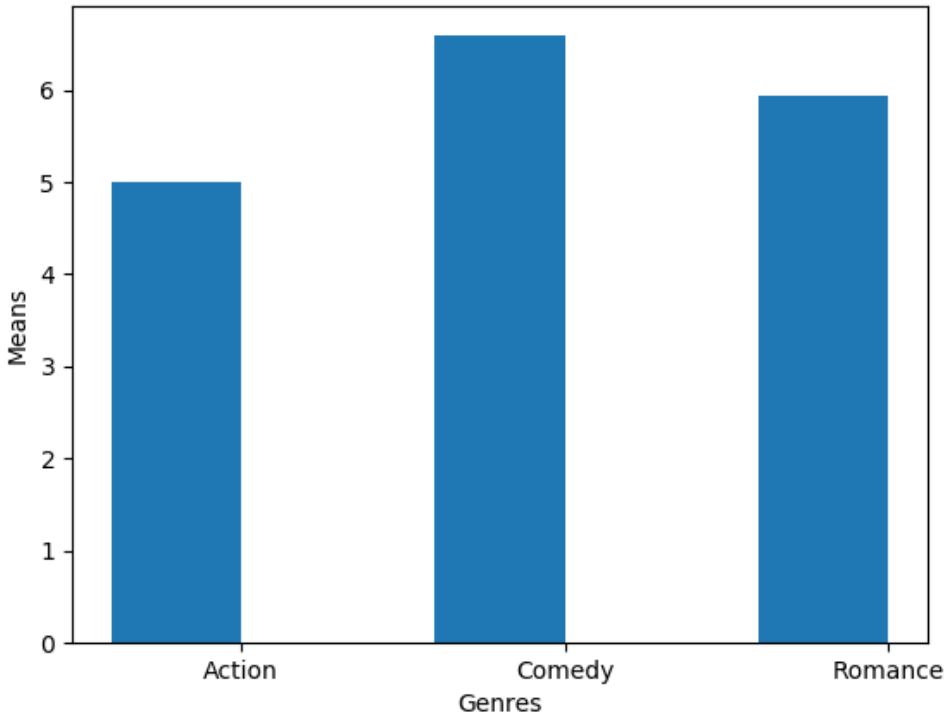
Boxplot for AVGRATING_WEBSITE_1



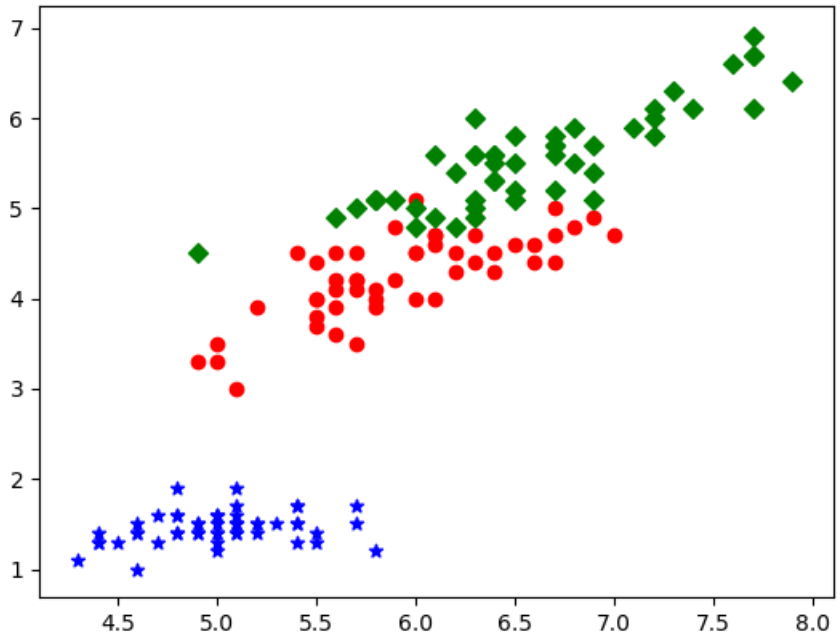
Histogram for AVGRATING_WEBSITE_3



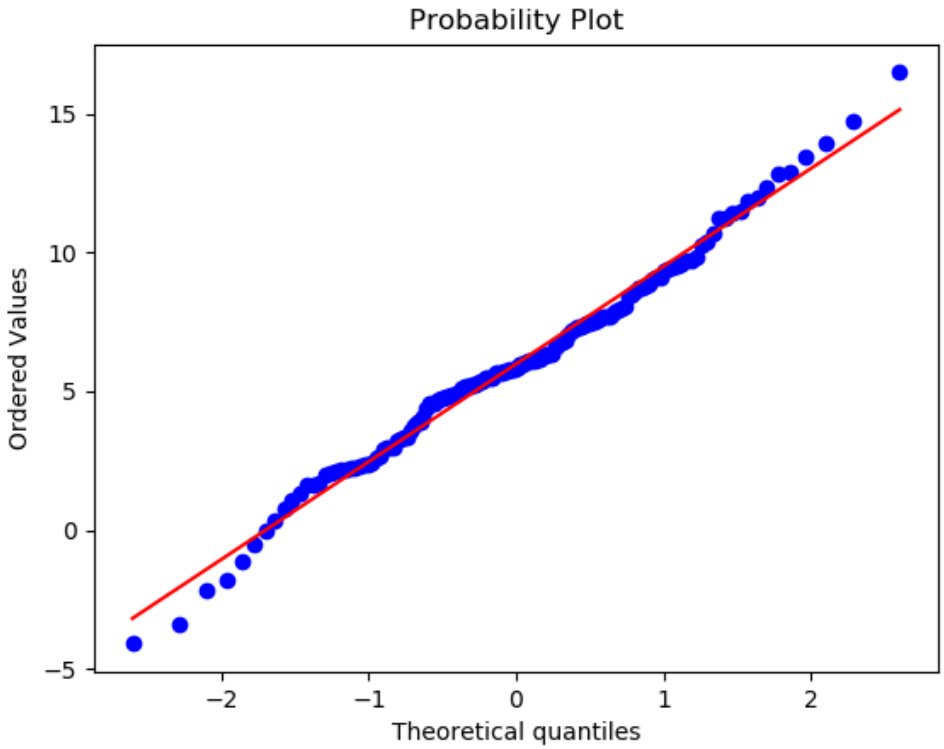
Bar chart for AVGRATING_WEBSITE_1



Scatter plot for AVGRATING_WEBSITE_1 and AVGRATING_WEBSITE_3



Q-Q plot for AVGRATING_WEBSITE_1 and AVGRATING_WEBSITE_3



KL Divergence:
0.872373121237761
1.3285877172613967

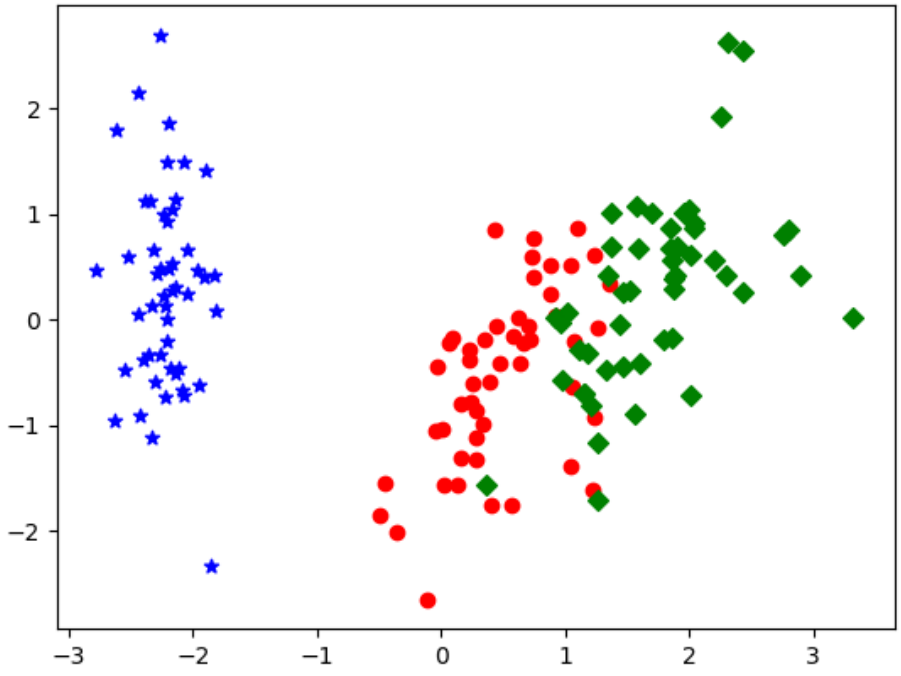
Data Cleaning and Integration

Using covariance analysis the correlation coefficients are

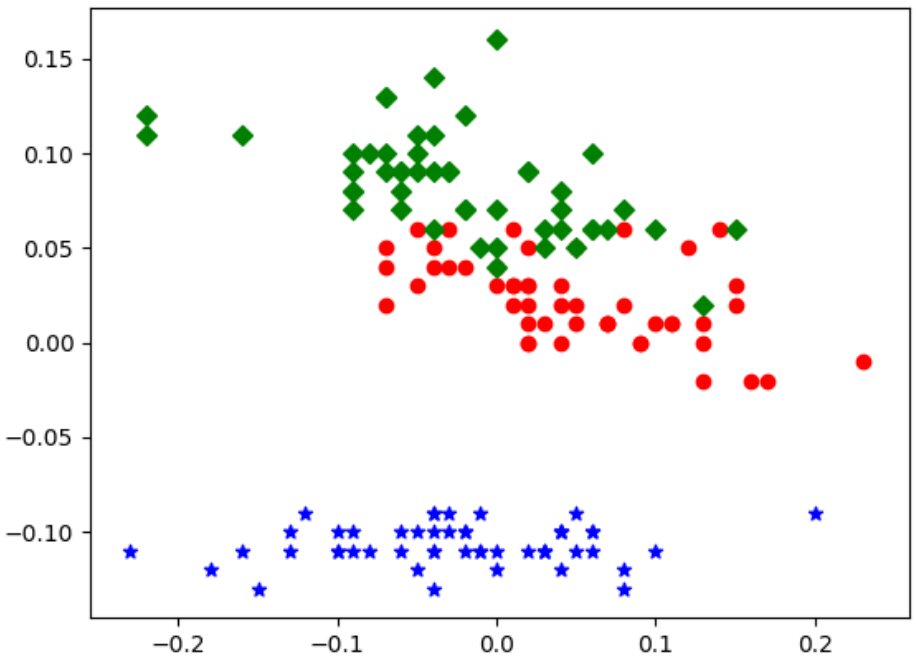
	Website 1	Website 2
Website 1		
Website 2		

The above results are not the same. In D the data has not been normalized, whereas in A the data has been normalized (mean-zeroed). Generally speaking, correlation standardizes covariance by some measure of variability in the data and thus produces a quantity that has intuitive interpretations and a consistent scale. However, in this case, we used covariance analysis rather than a correlation measure, such as Spearman correlation, so since only one of our data matrices was normalized, it makes sense that since we used covariance analysis the values would differ. If we were to use a correlation method though, then the values ought to be the same, since D would become normalized, and A would have already been normalized. In fact, since A is already normalized, we can perform covariance analysis on A and correlation analysis on D and get the same result, since when the data is already normalized then the covariance and correlation matrices will be the same.

Data Reduction
Scatter plot of the first two principal components



Scatter plot on the first two left-singular vectors



Top 3 eigenvalues in PCA – the eigenvalues represent the variance in the direction of the eigenvector, working backwards from the eigenvectors, it is possible to obtain the eigenv alues. Alternatively, one can restart and do the pca by hand, multiplying the transpose of t he matrix with the matrix, then using the linear algebra library to obtain the eigenvalues a nd eigenvectors. There is also a function in the PCA class that provides these eigenvalues , called explain_variance_ [2.93779398 0.92025136 0.14793596]

Top 3 singular values in SVD: [4.69 11.71 20.92]

Propogation-based method: Using this method the matrices do not align properly so it is neither the eigenvectors or the singular values.