1a) Q1 = 82, Median = 89, Q3 = 95

1b) Mean = 87.011

1c) Mode = 95

1d) The data is positively skewed. This is because the mean < median < mode. 87.011 < 89 < 95.

2a)



2b) A = (3, 1, 2) B = (-1, 0, 8)

2b1)



S









2b2)









2b3)











2c) The Euclidean distance is a direct line between 2 points in space. The Manhattan distance is the distance between 2 points in space if the path traveled from A to B is taken at right angles. If we look at these two distances in a 2-d plane, the Manhattan distance will always be the 2 shorter sides of a right triangle, whereas the Euclidean distance will be the hypotenuse. This is all assuming the points are not already on a 90 degree axis with each other, which would mean the Manhattan and Euclidean distances would be equal.

2d1) h = 2: 412.941

2d2) h = 3: 216.448

3a)

**Before**

Mean = 76.814

Variance = 171.396

**After**

Mean = 0

Variance = 1

3b)

Original Value = 90

Z-Score = 1.007

4a)





Correlation Coefficient = .985. The 2 vectors in the data set are positively correlated, meaning as X changes, Y changes in the same direction with a similar magnitude.

4b) PCA will help to reduce the data size because the correlation coefficient is high. The higher the coefficient, the more redundancy in a data set. Because there is so much redundancy, PCA will be effective in reducing the data set size.

4c) Covariance Matrix =

4d) There are 2 principal components in the data set. They are...