Stat 407 Quiz 1 Name _____

1. *(5 pts)* For the following data matrix,

$$\mathbf{X} = \begin{bmatrix} 11 & 32 \\ 31 & -5 \\ -2 & 17 \\ -8 & 5 \\ -42 & 39 \end{bmatrix}$$

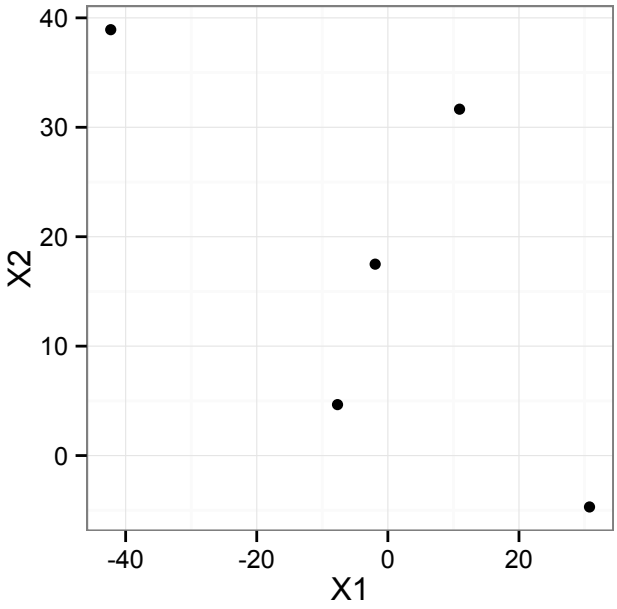
   (a) *(1)* How many cases are there? ____

   (b) *(1)* How many variables are there? ____

   (c) *(1)* Compute the multivariate mean, and write it in vector notation.

   (d) *(1)* Compute the Euclidean distance between cases 1 an 2? ____

   (e) *(1)* A scatterplot of the data is drawn below. Plot the mean of the data on this graph.



2. *(3 pts)* **T** or **F**.

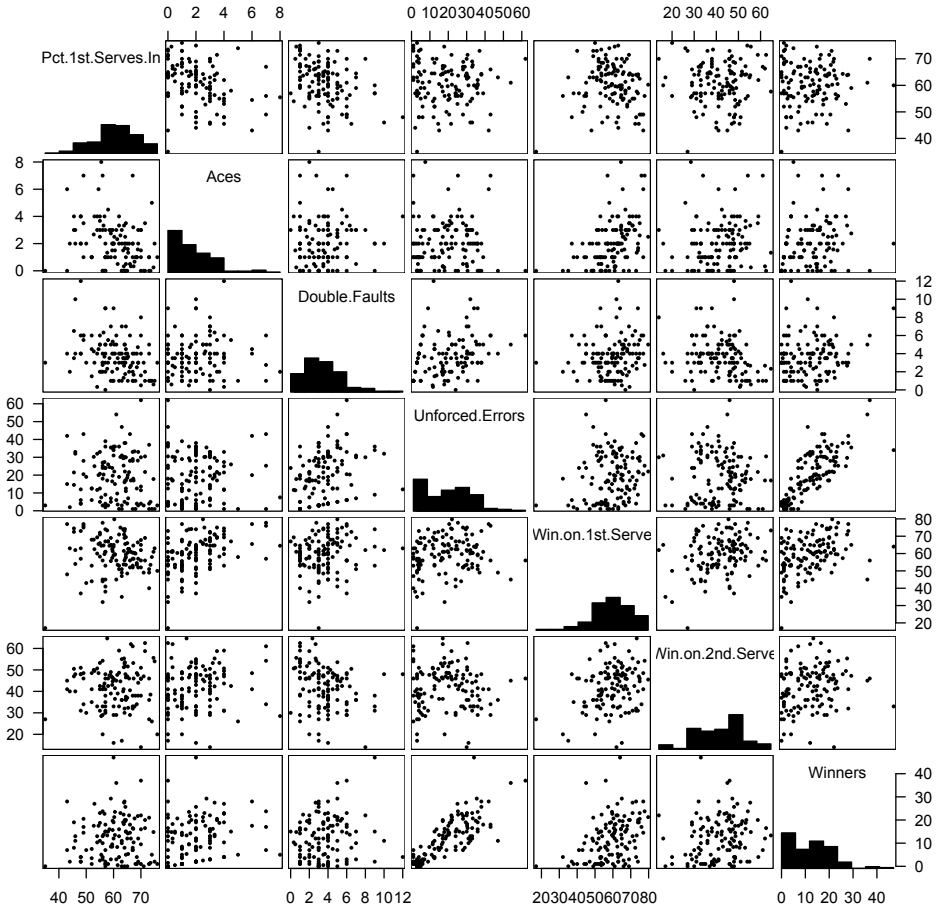
   (a) *(1)* If the variables in a data matrix are standardized the correlation matrix will change. ____

   (b) *(1)* If the covariance is 0, then all of the data values of the two variables are exactly the same. ____

   (c) *(1)* When there are values missing for a pair of variables, using the mean of the complete cases to impute the missing values will probably reduce the correlation between the two variables. ____

3. *(2pts)* Below there are two pieces, from an analysis of tennis statistics, a correlation matrix and a scatterplot matrix. The variables are in the same order for each summary.

$$\mathbf{R} = \begin{bmatrix} 1.00 & -0.28 & -0.32 & -0.03 & -0.06 & 0.02 & 0.03 \\ -0.28 & 1.00 & 0.05 & 0.06 & 0.45 & 0.13 & 0.21 \\ -0.32 & 0.05 & 1.00 & 0.29 & 0.15 & -0.23 & 0.17 \\ -0.03 & 0.06 & 0.29 & 1.00 & 0.17 & -0.03 & 0.81 \\ -0.06 & 0.45 & 0.15 & 0.17 & 1.00 & 0.28 & 0.38 \\ 0.02 & 0.13 & -0.23 & -0.03 & 0.28 & 1.00 & 0.17 \\ 0.03 & 0.21 & 0.17 & 0.81 & 0.38 & 0.17 & 1.00 \end{bmatrix}$$



(a) *(1)* Which two variables have the strongest correlation? _____ _____

(b) *(1)* What is the correlation between Aces and Double.Faults? _____