Problem 1.1

Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative. If you consider that some cases may have more than one interpretation, briefly indicate your explanation.

ATTRIBUTES	Bin/Con/Disc	Qual/Quan	Nom/Rat/Ord
AM & PM	Binary	Qualitative	Ordinal
Can be considered binary because the chose is between AM & PM. AM and PM time can be ordered, so the attribute can be classified as ordinal.			
Degrees between 0◦ and 360◦	Continuous	Quantitative	Ratio
Degrees are in $\mathbf{0} < \boldsymbol{R} < \mathbf{360}$ so they can be defined continuous. They can be rescaled in radiant or gradians.			
Bronze, Silver, and Gold medals	Discrete	Qualitative	Ordinal
It can be defined discrete because is a finite set of medal types. Ordinal because they rank the performance of the athletes.			
Opaque, Translucent, Transparent	Discrete	Qualitative	Ordinal
Discrete because there are only 3 possible values. You can also give an order between the three attributes. In fact, the light can pass better in opaque transparent, than translucent and then opaque.			

Problem 1.3

Consider the following data set:

$$D = \begin{array}{cc} & \begin{array}{cc} A_1 & A_2 \end{array} \\ \begin{array}{c} {a_1}^T \\ {a_2}^T \\ {a_3}^T \end{array} & \begin{pmatrix} 2 & 0,8 \\ 5 & 2,4 \\ 8 & 5,5 \end{pmatrix} \end{array}$$

1. **Calculate mean vector** μ

$$\mu_{A_1} = \frac{2+5+8}{3} = 5$$

$$\mu_{A_2} = \frac{0,8+2,4+5,5}{3} = 2,9$$

2. **Center the data, i.e., all features should be zero mean.**

$$Z = \begin{array}{cc} & \begin{array}{cc} \widetilde{A_1} & \widetilde{A_2} \end{array} \\ \begin{array}{c} {a_1}^T \\ {a_2}^T \\ {a_3}^T \end{array} & \begin{pmatrix} -3 & -2,1 \\ 0 & -0,5 \\ 3 & 2,6 \end{pmatrix} \end{array}$$

3. **Show that the data centering can be a matrix manipulation.** $\boldsymbol{Z} = \boldsymbol{D} - \mathbf{1}\mu^{\boldsymbol{T}}$

$$Z = \begin{pmatrix} 2 & 0,8 \\ 5 & 2,4 \\ 8 & 5,5 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} * (5 \quad 2,9) = \begin{pmatrix} 2 & 0,8 \\ 5 & 2,4 \\ 8 & 5,5 \end{pmatrix} - \begin{pmatrix} 5 & 2,9 \\ 5 & 2,9 \\ 5 & 2,9 \end{pmatrix} = \begin{pmatrix} -3 & -2,1 \\ 0 & -0,5 \\ 3 & 2,6 \end{pmatrix}$$

4. Calculate covariance matrix

The covariance matrix (C) can be calculating as follow

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \frac{1}{N-1} * \begin{pmatrix} \sum_{i=1}^{N} (A_{i1} - \mu_{A_1}) * (A_{i1} - \mu_{A_1}) & \sum_{i=1}^{N} (A_{i1} - \mu_{A_1}) * (A_{i2} - \mu_{A_2}) \\ \sum_{i=1}^{N} (A_{i2} - \mu_{A_2}) * (A_{i1} - \mu_{A_1}) & \sum_{i=1}^{N} (A_{i2} - \mu_{A_2}) * (A_{i2} - \mu_{A_2}) \end{pmatrix}$$

With N number of rows of the dataset matrix.

$$c_{11} = \frac{1}{2} * (-3 * -3) + 0 + (3 * 3) = 9$$

$$c_{12} = c_{21} = \frac{1}{2} * (-3) * (-2{,}1) + (0 * -0{,}5) + (3 * 2{,}6) = 7{,}05$$

$$c_{22} = \frac{1}{2} * (-2{,}1 * -2{,}1) + (-0{,}5 * -0{,}5) + (2{,}6 * 2{,}6) = 5{,}71$$

$$C = \begin{pmatrix} 9 & 7{,}05 \\ 7{,}05 & 5{,}71 \end{pmatrix}$$