Question 1.

i) $\widehat{\mathbb{E}[\mathbf{X}]} = (2.5, 2)^T$

ii) $\widehat{Var[X_1]} = 8.75$, $\widehat{Var[X_2]} = 3.5$

iii) $\widehat{Cov[X_1, X_2]} = 4$

Question 2. Notice that the expected value for each data dimension is zero. So, Using the matrix

$$\chi = \begin{pmatrix} -3 & 2 & 6 & -5 & -3 & 2 & 4 & -3 \\ -7 & -1 & -2 & 16 & -6 & -7 & -1 & 8 \\ -1 & -6 & 6 & -1 & 4 & 2 & -2 & -2 \end{pmatrix},$$

We see that

$$\widehat{Cov[\mathbf{X}]} = \frac{1}{8}\chi\chi^T = \begin{pmatrix} 14 & -12.125 & 2.75 \\ -12.125 & 57.5 & -8.375 \\ 2.75 & -8.375 & 12.75 \end{pmatrix}.$$

Question 3. The shape of the data cloud appears symmetric about the x_2 axis. The variables X_1 and X_2 show a different covariance for each side of the x_2 axis. Regardless of the parity of $(X_1 - \mathbb{E}(X_1))$, $(X_2 - \mathbb{E}(X_2))$ will be both positive and negative "with equal measure", resulting in $Cov[X_1, X_2] \approx 0$. So, we expect that

$$Cov[\mathbf{X}] = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

where $\lambda_1 > \lambda_2$ (we assume equal scale for x_1 and x_2 in the figure).

Question 4. The (normalised) eigenvectors values of $\mathbf{A}$ are

$$\begin{aligned} \mathbf{v}_1 &= (0.639, 0.468, 0.611)^T \\ \mathbf{v}_2 &= (0.839, -0.317, -0.443)^T \\ \mathbf{v}_3 &= (-0.0105, -0.780, 0.625)^T \end{aligned}$$

The eigenvalues of $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ are 8.71, -3.15 and 0.437 respectively.

Question 5. We see that

$$\begin{aligned} Cov[\widetilde{X}_1, \widetilde{X}_2] &= \mathbb{E}[(\widetilde{X}_1 - \mathbb{E}[\widetilde{X}_1])(\widetilde{X}_2 - \mathbb{E}[\widetilde{X}_2])] \\ &= \mathbb{E}\left[\left(\frac{X_1 - \mu_1}{\sigma_1} - \mathbb{E}\left[\frac{X_1 - \mu_1}{\sigma_1}\right]\right)\left(\frac{X_2 - \mu_2}{\sigma_2} - \mathbb{E}\left[\frac{X_2 - \mu_2}{\sigma_2}\right]\right)\right] \\ &= \mathbb{E}\left[\left(\frac{X_1 - \mu_1 - \mathbb{E}[X_1] + \mu_1}{\sigma_1}\right)\left(\frac{X_2 - \mu_2 - \mathbb{E}[X_2] + \mu_2}{\sigma_2}\right)\right] \\ &= \frac{1}{\sigma_1\sigma_2}\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] \\ &= Corr(X_1, X_2), \text{ by definition.} \end{aligned}$$

Question 6.

i) When we want to preserve a proportion of the variability in the data, say ρ, we pick the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i$ to be our new axes for the data, where $i \in \{1, 2, \ldots, d\}$ is the smallest possible integer such that

$$\frac{\sum_{n=1}^{i} \lambda_n}{\sum_{m=1}^{d} \lambda_m} \geq \rho.$$

ii) Notice that

$$\frac{\lambda_1 + \lambda_2 + \lambda_3}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4} = \frac{4.78}{4.98} \approx 0.96.$$

Therefore we choose $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ for our new axes.

Question 7.

i) $\mathbf{C} = \begin{pmatrix} 4.06 & 0.446 \\ 0.446 & 3.25 \end{pmatrix}$

ii) The normalised eigenvectors of $\mathbf{C}$ and corresponding eigenvalues are

$$\mathbf{v}_1 = (0.914, 0.406)^T,\ \lambda_1 = 4.26$$
$$\mathbf{v}_2 = (-0.406, 0.914)^T,\ \lambda_2 = 3.05$$

Therefore, we can decompose $\mathbf{C}$ as

$$\mathbf{C} = \begin{pmatrix} 0.914 & -0.406 \\ 0.406 & 0.914 \end{pmatrix} \cdot \begin{pmatrix} 4.26 & 0 \\ 0 & 3.05 \end{pmatrix} \cdot \begin{pmatrix} 0.914 & 0.406 \\ -0.406 & 0.914 \end{pmatrix}.$$

iii) Since $\lambda_1 > \lambda_2$, we pick $\mathbf{v}_1$ as our new axis. If we let $\widetilde{\mathcal{D}}$ be our project data set, then

$$\begin{aligned} \widetilde{\mathcal{D}} = &\{-0.611, -0.812, 0.305, -1.12, 0.711, -0.203, 2.84, \\ &0.914, 2.74, 3.15, 2.13, 4.42, 4.87, 4.98\} \end{aligned}.$$

Question 8.

i) Consider the $(i, j)^{th}$ element of $\mathbf{C}$:

$$(\mathbf{C})_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathbb{E}[X_i X_j] - \mu_i \mu_j. \quad (1)$$

Verify that this is indeed true!

So, using (1), we can write

$$\mathbf{C} = \begin{pmatrix} \mathbb{E}[X_1X_1] & ... & \mathbb{E}[X_1X_d] \\ \vdots & \ddots & \vdots \\ \mathbb{E}[X_dX_1] & ... & \mathbb{E}[X_dX_d] \end{pmatrix} - \begin{pmatrix} \mu_1\mu_1 & ... & \mu_1\mu_d \\ \vdots & \ddots & \vdots \\ \mu_d\mu_1 & ... & \mu_d\mu_d \end{pmatrix}$$

$$= \mathbb{E}\left[(X_1, \dots, X_d)^T \cdot (X_1, \dots, X_d)\right] - (\mu_1, \dots, \mu_d)^T \cdot (\mu_1, \dots, \mu_d)$$

$$= \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

ii) If we want to show that the eigenvalues of $\mathbf{C}$ are always positive, we first need to prove that $\mathbf{C}$ is a positive-definite matrix. This means that for any non-zero column vector $\mathbf{z}$ of dimension d, the scalar $\mathbf{z}^T\mathbf{C}\mathbf{z}$ will always be a positive value.

Let $\mathbf{z}$ be a non-zero column vector. We have:

$$\begin{aligned} \mathbf{z}^T\mathbf{C}\mathbf{z} &= \mathbf{z}^T \cdot (\mathbb{E}[\mathbf{X}\mathbf{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T) \cdot \mathbf{z} \\ &= \mathbf{z}^T \cdot \mathbb{E}[\mathbf{X}\mathbf{X}^T - \boldsymbol{\mu}\boldsymbol{\mu}^T] \cdot \mathbf{z} \\ &= \mathbf{z}^T \cdot \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] \cdot \mathbf{z} \\ &= \mathbb{E}[\mathbf{z}^T(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z}] \\ &= \mathbb{E}[((\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z})^T(\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z}] \\ &= \mathbb{E}[((\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z})^2] \\ &\geq 0. \end{aligned}$$

The sixth equality follows from the fact that $(\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z}$ is a dot product (resulting in a scalar), hence $(\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z} = ((\mathbf{X} - \boldsymbol{\mu})^T\mathbf{z})^T$. Furthermore, $\mathbf{C}$ is positive definite if and only if $\mathbf{V}\widetilde{\mathbf{C}}\mathbf{V}^T$ is positive definite. However, notice that the column vectors of $\mathbf{V}$ form a basis and

$$\mathbf{z}^T\mathbf{C}\mathbf{z} \geq 0 \Leftrightarrow \mathbf{z}^T\mathbf{V}\widetilde{\mathbf{C}}\mathbf{V}^T\mathbf{z} \geq 0 \Leftrightarrow (\mathbf{V}^T\mathbf{z})^T\widetilde{\mathbf{C}}\mathbf{V}^T\mathbf{z} \geq 0 \Leftrightarrow \mathbf{y}^T\widetilde{\mathbf{C}}\mathbf{y} \geq 0. \qquad (2)$$

Since $\widetilde{\mathbf{C}}$ is a diagonal matrix, the last inequality in (2) holds only if the diagonal entries (variances) are all positive, proving our claim.