Lab 5: Visualizing PCA Exercise

Objective:
Apply Principal Component Analysis (PCA) to a dataset of your choice (not Iris) to gain hands-on experience with covariance computation, eigen-analysis, and multidimensional visualization.

Assignment Description:

1. **Dataset Selection:**
   - Choose any real-world dataset you have used (or find a new one) except the Iris dataset.
   - The dataset should have at least four numeric features.
2. **Visualize Covariance Matrix with Heatmap:**
   - Compute the covariance matrix of your selected features (after centering or standardizing as needed).
   - Visualize the covariance matrix as a heatmap.
   - Use a diverging or sequential color palette.
   - Title and axis labels must be in Times New Roman, size 14.
3. **Eigenvalues & Eigenvectors Visualization:**
   - Compute the full set of eigenvalues and eigenvectors of the covariance matrix.
   - Visualize your results in a clear, interpretable format (e.g., bar chart of eigenvalues, table or heatmap of eigenvectors).
   - All text (titles, labels, legends) must be in Times New Roman, size 14.
4. **Produce three more visualizations that leverage PCA skills:**
   - For example:
     - Scree Plot showing cumulative variance explained.
     - 2D Scatterplot of your data projected onto the first two principal components, colored by a relevant category or continuous variable.
     - Biplot overlaying original feature vectors on the 2D projection.
     - 3D Plot of the first three principal components.
     - Or any other PCA-driven map or diagram that reveals structure in your data.
   - Formatting: All titles, labels, and annotations in Times New Roman, size 14.
5. **Code Documentation and Detailed Explanations:**
   - Include comprehensive comments throughout your code to explain your logic, data processing steps, and visualization choices.
   - Provide detailed explanations in Markdown cells (if using Jupyter Notebook) or as part of your code documentation. Explain:
     - How the data is loaded and preprocessed.
     - Any assumptions or decisions made during your analysis.
   - For each visualization, write a brief summary (2–4 sentences) describing what you observe and how it informs your understanding of the data's structure or relationships.
6. **Saving Visualizations:**
   Save each of your visualizations as an image file. Name each image using the following format: `<Lab5+YourName+Title>.png` (for example, if your name is JaneDoe and your

first visualization is titled “ScatterAnalysis,” then name the file `Lab5_JaneDoe_ScatterAnalysis.png`). This naming convention will help keep your files organized and easily identifiable.

Submission Instructions:

- Upload all your deliverables—including your Jupyter Notebook with the visualizations results and statistical analyses—to a well-organized GitHub repository.
- Submit the GitHub repository link via Canvas.