CSCE5320 Scientific Data Visualization

The below task should be completed as Activity 8 and Activity 9. The topics includes Seaborn, Matplotlib, PCA, t-SNE, network visualization, and color theory. The questions include a mix of programming and descriptive tasks.

Questions

Question 1: Data Visualization with Seaborn

Programming Task:
Using Seaborn, create a pair plot for the Iris dataset. Explain how the pair plot helps in understanding the relationships between different features in the dataset.

Requirements:

- Load the Iris dataset using seaborn.load_dataset('iris').
- Create a pair plot using seaborn.pairplot().
- Include a brief interpretation of the visual output, focusing on the relationships between species and features.

Question 2: Customizing Matplotlib Plots

Programming Task:
Generate a bar plot using Matplotlib to visualize the average petal length of each species in the Iris dataset. Customize the plot with titles, labels, and colors.

Requirements:

- Use matplotlib.pyplot to create a bar plot.
- Calculate the average petal length for each species.
- Customize the plot with appropriate titles and axis labels, and choose a color palette.

Question 3: Principal Component Analysis (PCA)

Descriptive Task:
Explain the concept of PCA and its application in data visualization. Implement PCA on the Iris dataset and visualize the first two principal components using a scatter plot.

Requirements:

- Describe how PCA reduces dimensionality and retains variance.
- Perform PCA using sklearn.decomposition.PCA.
- Plot the first two principal components with labels indicating the species.

Question 4: t-Distributed Stochastic Neighbor Embedding (t-SNE)

Programming Task:
Perform t-SNE on the Iris dataset and visualize the result. Compare the t-SNE plot with the PCA plot from the previous question. Discuss the differences.

Requirements:

- Use sklearn.manifold.TSNE for the t-SNE implementation.
- Create a scatter plot of the t-SNE results.
- Discuss how t-SNE differs from PCA in terms of visualization, especially regarding cluster separation.

Question 5: Network Visualization with NetworkX

Programming Task:
Create a simple network graph using NetworkX. Visualize the graph with Matplotlib and customize the node colors based on degree centrality.

Requirements:

- Create a random graph using networkx.erdos_renyi_graph().

- Calculate degree centrality and use it to color the nodes.
- Visualize the network and describe how the node color indicates their connectivity.

Question 6: Color Theory in Visualization

Descriptive Task:
Discuss the importance of color selection in data visualization. Provide examples of good and bad color choices, and describe how color blindness considerations can impact design.

Requirements:

- Explain concepts like color contrast, harmony, and accessibility.
- Include examples (images or plots) that illustrate effective and ineffective color use in visualizations.

Question 7: Combining Techniques for Advanced Visualization

Programming Task:
Using the Titanic dataset, create a visualization that combines multiple techniques (e.g., a heatmap for correlations, followed by a scatter plot of age vs. fare). Discuss the insights gained from this multi-faceted visualization.

Requirements:

- Load the Titanic dataset using Seaborn.
- Create a correlation heatmap using seaborn.heatmap().
- Follow with a scatter plot of age vs. fare using Matplotlib.
- Discuss how the combination of these plots provides deeper insights into the data.

Question 8: Evaluating Visualizations

Descriptive Task:
Select a visualization from a recent scientific publication (or provide one). Critically evaluate its effectiveness in conveying information. What aspects could be improved? Provide suggestions based on visualization best practices.

Requirements:

- Provide a link or image of the chosen visualization.
- Discuss clarity, accuracy, and design choices.
- Suggest specific improvements based on principles of effective visualization.

Instructions for Submission

Please complete all tasks in Google Colab. Follow the rubrics provided below to complete the two activities. The first four questions are for Activity 8, and the last four questions are for Activity 9.

You should submit the tasks separately:

- **Activity 8**: Complete the first four questions in Google Colab and submit them as Activity 8.
- **Activity 9**: Complete questions 5 to 8 and submit them in the designated tab for Activity 9 in Google Colab.

Make sure to label your submissions clearly.

Rubrics:

Question	Criteria	Excellent (5 points)	Satisfactory (3 points)	Unsatisfactory (0 points)
1: Data Visualization with Seaborn	Insights	Provides comprehensive insights from the visualization.	Basic insights provided but lacks depth.	No insights provided.
	Code Quality	Code is clean, organized, and follows best practices.	Code is somewhat organized but has issues.	Code is poorly organized and hard to follow.
Total		**10 points**		
2: Customizing Matplotlib Plots	Creativity	Innovative customization techniques are applied.	Some creative elements are present.	No creativity in customization.
Total		**10 points**		
3: PCA Visualization	Clarity	PCA plot is exceptionally clear and easy to interpret.	PCA plot is understandable but could be clearer.	PCA plot is confusing and difficult to interpret.
Total		**10 points**		
4: t-SNE Visualization	Clarity	t-SNE plot provides valuable insights and is well labeled.	t-SNE plot provides some insights but lacks clarity.	t-SNE plot is unclear and not informative.
Total		**10 points**		

Total Points for Activity 8: 100 points

Question	Criteria	Excellent (5 points)	Satisfactory (3 points)	Unsatisfactory (0 points)
1: Network Quality	Clarity	Network visualization is exceptionally clear and informative.	Network visualization is understandable but lacks clarity.	Network visualization is confusing and difficult to interpret.
Total		**10 points**		
2: Color Theory Application	Creativity	Excellent application of color theory in the visualizations.	Some application of color theory but lacks creativity.	No application of color theory.
Total		**10 points**		
3: Evaluation of Visualizations	Thoroughness	Evaluation is thorough and covers all aspects of visualization.	Evaluation is basic and covers some aspects.	Evaluation is lacking or incomplete.
Total		**10 points**		
4: Overall Presentation	Clarity	Overall presentation is professional and well-organized.	Presentation is somewhat organized but has issues.	Presentation is poorly organized.
Total		**10 points**		

Total Points for Activity 9: 100 points