CTT227 - Data Visualization

Project

DATA VISUALIZATION USING D3.JS

1. General Information

Assignment ID	PROJECT
Estimated duration:	10 weeks
Submission deadline:	12/04/2025
Assignment type:	Student Group
Submission chanel:	Moodle
Teachers:	Nguyễn Ngọc Minh Châu
Contacts:	nnmchau@fit.hcmus.edu.vn

2. Learning Outcomes

This exercise is designed to help students achieve the following competencies:

- Demonstrate the ability to clean, and preprocess datasets from various origins, ensuring data integrity and readiness for visualization.
- Apply foundational principles of data visualization, such as selecting appropriate chart types and utilizing visual encodings, to effectively represent complex data.

- Exhibit competence in using D3.js to create dynamic and interactive visualizations, including the implementation of scales, axes, and transitions.
- Articulate data-driven insights through visualizations, enabling audiences to comprehend and interpret the underlying data narratives.

3. Project Assignment

3.1. Dataset

Using the Heart Disease dataset given via this link.

Dataset description:

#	Field name	Description
1	Age	The individual's age.
2	Gender	The individual's gender (Male or Female)
3	Blood Pressure	The individual's blood pressure (systolic)
4	Cholesterol Level	The individual's total cholesterol level.
5	Exercise Habits	The individual's exercise habits (Low, Medium, High).
6	Smoking	Whether the individual smokes or not (Yes or No).
7	Family Heart Disease	Whether there is a family history of heart disease (Yes or No).
8	BMI	The individual's body mass index
9	Alcohol Consumption	The individual's alcohol consumption level (None, Low, Medium, High)

10	Stress Level	The individual's stress level (Low, Medium, High)

3.2. Mandatory requirements

The number of Task required for each group = MAX(3, number of members * 2)

For example:

- Group A has 1 person: This group is expected to complete at least 3 tasks of their choice.
- Group B has 3 people: This group is expected to complete 6 tasks of their choice.
- Group C has 4 people: This group is expected to complete 8 tasks.

Task 1: What is the distribution of heart disease status across different age groups?

Task 2: How does gender correlate with heart disease prevalence?

Task 3: Is there a relationship between smoking status and heart disease occurrence?

Task 4: How do exercise habits influence heart disease status?

Task 5: How does cholesterol level vary between individuals with and without heart disease?

Task 6: Is there a correlation between BMI and heart disease status?

Task 7: Does a family history of heart disease increase the risk?

Task 8: How does the distribution of cholesterol levels differ between males and females

3.3. Bonus point requirements

*You can choose **one** of the following Bonus requirements for extra points.*

Bonus 1: How do lifestyle factors such as smoking, alcohol consumption, and exercise habits collectively influence the risk of heart disease across different age groups?
Bonus 2: What is the relationship between blood pressure and BMI across different stress levels, and how does this relationship differ between individuals with and without heart disease?

Bonus 3: Combine the (all or some) visualizations to form a dashboard.

4. Assessment

- Form: Oral defense
- The teacher will give the total scores, the team should discuss to have the point for each individual based on the contribution to team work

5. Submission rules

Submit a compressed file which includes:

- Report. Format: .pdf
- Source Code
- Repository to team version control such as Github, Gitlab

Cheating will result in 0, no exception