CS ELEC 3C – Data Visualization Project Guidelines

Project Title: Visualizing World Development Indicators (WDI) - Addressing a UN Sustainable Development Goal (SDG)

Project Overview

In this project, students will analyze and visualize data from the **World Development Indicators (WDI)** provided by the **World Bank**, with the goal of addressing one of the **United Nations Sustainable Development Goals (SDGs)**. Students will choose a global development question aligned with an SDG and use data visualization techniques to explore trends, patterns, and disparities across time and geography.

---

1. Project Background

Introduce the nature and importance of your project.

- Describe the **global issue** your project will explore.
- State **which UN SDG** your project addresses (e.g., SDG 4: Quality Education, SDG 3: Good Health and Well-being, SDG 13: Climate Action).
- Explain the importance of this topic in the context of global development and policy.

Example: "This project focuses on SDG 6: Clean Water and Sanitation. We aim to understand how access to clean water has changed over time across income groups and regions."

---

2. Statement of the Problem

Define the research questions, objectives, scope, and limitations.

- Clearly articulate the **problem statement**: What questions are you trying to answer?
- Provide **specific objectives** (e.g., compare education levels across regions, analyze CO_2 emissions vs. GDP).
- Define the **scope** (countries, regions, time frame, indicators).
- Mention any **limitations** (data availability, missing years, indicator granularity).

Example Problem Statement: "How has access to primary education evolved across low-income and high-income countries over the past 30 years, and what correlations exist between education access and GDP per capita?"

---

3. Background on the Dataset

Provide context and technical details about the dataset.

- Name: **World Development Indicators (WDI)**
- Source: **World Bank Open Data** (https://data.worldbank.org/indicator)
- Describe the dataset:
    - Time span (typically 1960–present)
    - Coverage (~200 countries)
    - Over 1,000 indicators grouped by themes: education, health, environment, economy, etc.
- Explain how the dataset is organized:
    - Columns: Country, Indicator Name, Indicator Code, Year, Value
- Mention any secondary datasets used (e.g., country regions, income levels).
- Note: Data can be accessed via the **World Bank API**, **CSV downloads**, or **Kaggle datasets**.

---

4. Literature Review

Ground your project in existing work.

- Summarize **studies, articles, or dashboards** that explore similar topics or use WDI data.
- Explain how these works inform or inspire your approach.
- Highlight gaps or limitations in their visualizations or analyses that your project might address.

---

5. Methodology

Break this section into subcomponents:

a. Data Set

- Specify the indicators used (e.g., literacy rate, CO_2 emissions, mortality rate).
- Justify why they are relevant to your SDG and project question.

b. Data Preparation

- Describe preprocessing steps:
    - Filtering countries, years, or regions
    - Handling missing values
    - Creating new derived variables if needed
    - Normalizing/scaling for comparability

c. Exploratory Data Analysis (EDA)

- Provide basic statistics, distributions, and data summaries.
- Use **3–4 preliminary visualizations** to understand patterns (e.g., line plots, boxplots, histograms).

d. Data Visualization

- Create **4–6 polished and informative visualizations**.
- Use appropriate chart types:
    - **Time series** for trends
    - **Choropleth maps** for geographic variation
    - **Scatter plots** for correlations
    - **Heatmaps** for indicator relationships
    - **Interactive dashboards** (e.g., Tableau, Plotly) encouraged
- Ensure clarity: label axes, titles, legends, and sources.

---

6. Data Analysis

Interpret your visualizations and extract insights.

- Answer your research questions using your charts.
- Describe trends, correlations, differences between regions/income levels, and unexpected findings.
- Relate your findings to the SDG and global development priorities.

---

7. Conclusion

Summarize findings and reflect on their implications.

- Restate your key insights and what they reveal about your chosen SDG.
- Discuss data or methodological limitations.
- Suggest policy or research recommendations.

- Offer ideas for future exploration (e.g., including more indicators or using machine learning).

Deliverables

A Team is composed of 2-3 students.

Each team must submit the following:

1. **Final Report** (8-10 pages, ACM format)
    - Contains all sections
    - Embedded and well-labeled visualizations
    - Clear narrative linking data to insights
2. **Code and Data Repository** (GitHub or ZIP)
    - Include all scripts for data processing and visualization
    - README file with setup instructions

Each Team will have a 10-min Presentation

- Slides and dashboard demo
- Summarize goals, visualizations, insights, and conclusions

Grading Rubric

CRITERIA	EXCELLENT	VERY GOOD	GOOD	NEEDS IMPROVEMENT	SCORE
Project Background & SDG Rationale (10)	Provides a thorough and compelling background with strong alignment to a specific Sustainable Development Goal (SDG); clearly articulates the significance and relevance of the project.	Provides a clear background and links to an SDG; rationale is mostly well-explained and relevant.	Background is adequate but may lack depth or strong SDG alignment; rationale is somewhat general.	Background is vague or missing; SDG rationale is unclear or absent.	___ /10
Statement of the Problem (10)	Clearly and precisely defines the problem; shows deep understanding of the context and implications.	Clearly states the problem with some contextual understanding.	Problem is stated but lacks depth or clarity.	Problem is vague, incomplete, or missing.	___ /10
Dataset Description & Preparation (10)	Dataset(s) are clearly described (source, features, relevance); thorough data cleaning and preprocessing with justification for each step.	Dataset is described with minor omissions; data preparation is mostly complete and appropriate.	Basic dataset description; limited data preparation with minimal explanation.	Dataset description is poor or missing; little or no evidence of data preparation.	___ /10
Literature Review (10)	Provides a well-organized, critical review of multiple relevant sources; clearly connects previous work to the current project.	Covers relevant sources with some connection to the project; shows understanding of background work.	Includes some relevant sources; limited analysis or connection to current project.	Few or no relevant sources; lacks depth or analysis.	___ /10
Exploratory Data Analysis (EDA) (15)	Conducts in-depth EDA with appropriate techniques; identifies patterns, trends, and outliers with insightful commentary.	Conducts EDA with relevant techniques; observations are sound though some details may be missing.	Basic EDA is performed; analysis is somewhat superficial or lacks depth.	Minimal or no EDA; lacks interpretation or misuses techniques.	___ /15
Visualization Techniques (15)	Uses a variety of effective, well-labeled, and appropriate visualizations to enhance understanding; excellent design and clarity.	Uses suitable visualizations that are mostly clear and informative; minor issues in labeling or design.	Limited variety or effectiveness in visualizations; some may lack clarity or purpose.	Visualizations are poorly chosen, unclear, or absent.	___ /15
Analysis and Insights (20)	Demonstrates deep analysis leading to insightful and original conclusions; strongly supported by data.	Good analysis with logical insights; mostly supported by data.	Analysis is basic; insights are general or only partially supported by data.	Analysis is weak, incorrect, or unsupported; lacks meaningful insights.	___ /20
Conclusion / Clarity (10)	Provides a strong, clear conclusion that summarizes key findings; reflects on limitations and suggests future directions.	Conclusion is clear and summarizes findings; some mention of limitations or next steps.	Basic conclusion that somewhat summarizes findings; limited reflection.	Conclusion is unclear, missing, or does not summarize the project.	___ /10