VINUNIVERSITY

COLLEGE OF ENGINEERING AND COMPUTER SCIENCE

VinUniversity

COMP4010 - Data Visualization

Project 2 Proposal
Inclusive Data Visualization: A Comprehensive Guide for Accessibility in R with ggplot2 and Quarto.

Professor Le Duy, Dzung
Teaching Assistants Hoang Tien, Vu and Luong Ha Tri, Nhan

Nguyen Trong, Nhan SIS ID: V202000224

April 25th, 2024

1. TD;DR

In this project, I plan to create a comprehensive guide for accessibility in R using ggplot2 and the Quarto library for data visualization.

2. Goals

The project "Inclusive Data Visualization: A Comprehensive Guide for Accessibility in R with ggplot2 and Quarto" aims to equip participants with the necessary knowledge and skills to create accessible and inclusive visualizations using the ggplot2 and Quarto packages within the R ecosystem. With a clear understanding of the diverse data visualization tools available in ggplot2 and Quarto, participants will learn how to design visualizations that prioritize accessibility for individuals with diverse needs, including those with disabilities. By mastering techniques for creating accessible visualizations, participants will contribute to making data-driven insights more widely accessible and usable for all audiences. Through practical exercises, hands-on projects, and guided instruction, the project seeks to empower participants to create visualizations that adhere to accessibility best practices and promote inclusivity in data visualization practices.

3. Motivation

- Create visualizations that reach a wider audience, including those with disabilities
- Enhance my professional portfolio by showcasing the necessary skill set in accessible data visualization
- Contribute to making information more accessible and equitable for everyone.
- Increase the effectiveness of your visualizations by making them clearer and more understandable for all users.

4. Dataset collection

Two commonly used datasets that offer a wealth of visualization opportunities are:

1. Iris Dataset: The Iris dataset is a classic dataset in the field of data science and is often used for introductory data visualization exercises. It contains measurements of various attributes of three species of iris flowers (setosa, versicolor, and virginica), including sepal length, sepal width, petal length, and petal width. With its simple structure and clear class separation, the Iris dataset is ideal for creating scatter plots, histograms, box plots, and other types of visualizations to explore patterns and relationships between variables.

2. Titanic Dataset: The Titanic dataset is another popular dataset frequently used for data visualization and machine learning tutorials. It contains information about passengers aboard the RMS Titanic, including their survival status, age, sex, ticket class, fare, and cabin. This dataset offers numerous opportunities for visualization, such as survival rate by passenger class, age distribution of passengers, survival

rate by gender, and more complex analyses like survival prediction models. Visualizations created with the Titanic dataset can help uncover insights about factors influencing survival rates and contribute to a deeper understanding of the Titanic disaster.

Both the Iris and Titanic datasets are widely available and well-documented, making them excellent choices for practicing data visualization techniques in R with ggplot2, Quarto, and other visualization tools.

5. Weekly plan

Week 3	Understanding Accessibility Principles / Dataset Preparation
Week 4	Practical Application with ggplot2
Week 5	Creating Accessible Visualizations with Quarto
Week 6	Advanced Data Visualization Techniques / Data Visualization for Storytelling
Week 7	Designing Interactive Dashboards
Week 8	Project Report & Presentation