EN685.622: Homework 1

B.A. Johnson | A.K. Saeed
bjohn159@jhu.edu | asaeed7@jhu.edu

Johns Hopkins Whiting School of Engineering — January 16, 2025

Introduction

Data Patterns and Representations, fundamentally, constitute the realm of data visualization within the domain of data science. Data visualization pertains to the sophisticated practice of encoding information in a manner that facilitates its easy and accurate interpretation by an audience. The extraordinary capabilities of human visual perception significantly influence our interaction with the world, echoing the adage, "a picture is worth a thousand words." The intellectual pursuit of integrating quantitative analysis with human visual comprehension has evolved remarkably over time.

This journey commenced with the pioneering work of William Playfair in the 1780s, followed by the formulation of Gestalt Principles in the 1920s, which emphasized the human ability to perceive patterns and wholes. Subsequently, the contributions of William S. Cleveland in the 1980s and Colin Ware in the 2000s further advanced our understanding of information perception.

In the current era, the proliferation of data visualization software, coupled with advanced user interface and experience design methodologies, is unparalleled in history. However, a critical inquiry remains: Are our contemporary practices in data visualization effectively grounded in the extensive body of research accumulated over the past centuries? This question beckons a reflective examination of our current methodologies in light of the rich legacy of data visualization research.

Info: The phrase "a picture is worth a thousand words" is an English idiom implying that a complex idea can be conveyed with just a single still image more effectively than a description does. Its origins are somewhat unclear, but it's often attributed to Frederick R. Barnard, who published a piece commending the effectiveness of graphics in advertising with the title "One look is worth a thousand words," in Printer's Ink, December 1921. However, he attributed this statement to a Japanese philosopher and later used "Chinese proverb" in 1927, likely to give it more weight.

1 Human Centered Design

In this task, you will focus on building a human-centered AutoML application using Streamlit. The application will allow users to input a dataset and automatically run multiple machine learning models (a minimum of 3), providing clear and concise results. The objective is to ensure that the application is easy to use, visually appealing, and addresses the user's needs in a human-centered way. You will complete this task in two parts:

Question 1

Journey Mapping and Storyboarding (20 points)

1. **Storyboard the Application Flow:** Create a visual storyboard that outlines how a user will interact with your application from start to finish. Think about key touch points, the sequence of interactions, and how the application will guide the user through the process of uploading their dataset, running multiple models, and viewing the output.
2. **Journey Mapping:** Map out the user's journey, emphasizing their needs, pain points, and goals at each stage. How will your design address these factors in a human-centered way? Think about how to minimize cognitive load and provide clear guidance to the user at every step.
3. **Incorporate Visual Design Elements:** Think about how you will integrate visual design elements to enhance clarity and user experience. Use mock-ups or sketches to show where elements like buttons, instructions, and outputs will be placed. Ensure that your design is intuitive and focused on solving the user's overarching problem.

Info: Visit the Streamlit documentation `https://docs.streamlit.io/` for more information on building apps.

Question 2

AutoML Application in Streamlit (40 points)

1. **Building** Once your storyboard is complete, the next step is to build the application itself. Your application should focus on one modality of data (e.g., tabular data, text, or images). The user will input their dataset, and your application will run at least three different machine learning models (e.g., linear regression, random forest, and a neural network) and provide the results. Your application must include:
2. **User-friendly Data Input:** A simple, intuitive interface for users to upload their datasets.
3. **Model Selection and Execution:** The application should automatically run a minimum of three machine learning models and provide the results of each.
4. **Clear and Concise Output:** Present the model results in a clear, easy-to-understand format, using visual design elements where appropriate. Ensure that the user can easily interpret the outputs, such as accuracy scores, visualizations of model performance, or other relevant metrics.
5. **Human-Centered Design:** Incorporate the principles of Human-Centered Design throughout the application to ensure that it is tailored to meet the needs of the user. Consider usability, visual clarity, and how the user interacts with the data and results.

By completing both parts of this question, you will demonstrate your ability to storyboard and journey map a human-centered application, as well as build a functional AutoML application in Streamlit that processes and outputs meaningful results for the user.

2 Principles of Data Visualization Analysis

You have learned the history of visual perception and the Gestalt principles. Let's see how well you can put your new learning to the test.

Question 3

Visual Perception (40 points)

Search the internet, articles, or books for a graphical visualization. Using all of the foundational information we have talked about thus far in the course, give an analysis of the visualization and pay specific attention to the following areas:

(a) Pre-Attentive Processing Analysis

(b) Visual Perception Techniques

(c) Attributes of Form

Your answer should be in mini-essay form (300-500 words). Grading will be based off analysis. Grammatical mistakes will not be punished severely, unless overly egregious to standard English language conventions.