Ira A. Fulton Schools of Engineering
Arizona State University

CSE 578: DATA VISUALIZATION

Course Project Progress Report

SRUTHI KEERTHANA NUTTAKKI

snuttakk@asu.edu| 1232316477

Problem Statement

As a data analyst at XYZ Corporation, I am assisting UVW College in enhancing its enrollment strategies by identifying key demographic factors associated with individuals earning over $50,000 annually. Using U.S. Census Bureau data, this project explores correlations between income and variables such as age, education, occupation, and marital status. The insights will inform targeted marketing campaigns for UVW's degree programs. This report summarizes my progress, key findings, challenges encountered, and next steps.

Tasks Completed

I loaded the adult.csv dataset, which starts with 32,561 rows and 15 columns, into Google Colab, using Pandas for cleaning and Seaborn/Matplotlib for visualizations. I cleaned it by removing the 'fnlwgt' column as it's not relevant, dropping 2,399 rows (7.4%) with missing values (marked "?"), and eliminating 3,258 duplicates. This left a solid dataset of 30,162 rows.

I selected 8 variables to focus on: age, sex, education, occupation, marital-status, hours-per-week, race, and native-country. These seem most likely to shape income and help UVW find their target students. I've created visualizations to compare these variables with income, narrowing down to a few clear ones that show patterns, like how age or sex relate to earning >50K. I also tested other combinations, such as race or occupation, to understand the data better and plan my next steps.

My visualizations revealed key patterns across the 8 variables:

- **Age**: Older adults (35–55) are more likely to earn >50K than younger ones.
- **Sex**: Men have a higher share of >50K earners than women.
- **Education**: Bachelor's degrees or higher strongly predict >50K incomes.
- **Occupation**: Professional and managerial roles dominate high earners.
- **Marital-Status**: Married individuals often earn more than singles.

- **Hours-per-week**: Working 50+ hours boosts chances of >50K.
- **Race**: White and Asian groups have more high earners than others.
- **Native-Country**: U.S.-born individuals lead in >50K earnings.

User Stories:

These insights suggest UVW can target high-earning professionals and underserved groups through three focused user stories built on key variable combinations:

- **Race, Education, Income:** I want to visualize race, education level, and income to identify income gaps for groups like Black or Indigenous adults with less education. UVW can offer scholarships to attract these students and support diversity.
- **Native-Country, Occupation, Income:** I want to see native country, occupation, and income together to understand if professionals from countries like India or China earn over $50K in key roles. UVW can design programs that appeal to this international talent.
- **Sex, Marital-Status, Hours-per-week, Income:** I want a chart of sex, marital status, hours worked, and income to see if married women working long hours earn less. UVW can offer flexible learning options like evening classes to support their advancement.

Challenges and Solutions:

The dataset's imbalance posed a challenge—White and U.S.-born individuals dominate, skewing early results for race and native-country. I tackled this by using proportions (e.g., % >$50K per group) instead of raw counts, ensuring fair comparisons. Another hurdle was making multi-variable visualizations clear; my first attempts were too crowded for UVW's team to follow. I switched to simpler formats, like grouped bars and heatmaps, testing them to keep insights accessible.

Next Steps:

Next, I'll complete the remaining visualizations to bring these user stories to life, creating clear charts that show how education, race, and income intersect, how international professionals perform in key occupations, and how work-life balance affects earnings. I'll also calculate concrete statistical relationships - like how strongly age and work hours correlate with income - to strengthen our findings. All of this will culminate in practical recommendations UVW can immediately use to shape their recruitment strategies, develop targeted programs, and ultimately grow their enrollment in meaningful, data-driven ways.