

CSE 578: DATA VISUALIZATION

COURSE PROJECT FINAL REPORT

Sruthi Keerthana Nuttakki | Ira A. Fulton Schools of Engineering | snuttakk@asu.edu | ASU ID: 1232316477

GOALS AND BUSINESS OBJECTIVE:

UVW College aims to increase enrollment by connecting with students who would greatly benefit from their degree programs, particularly those preparing for high-paying careers. As a data analyst at XYZ Corporation, my task is to study U.S. Census Bureau data to find out which personal and professional factors separate individuals earning over \$50,000 per year from those earning less. I'm focusing on eight main attributes—age, sex, education, occupation, marital status, hours worked per week, race, and country of origin—to help UVW's marketing team create clear, useful profiles to improve their outreach efforts. These attributes were selected because previous analyses indicate they're strongly linked to income and are most relevant to educational decision-making.

This analysis will guide UVW College in targeting their resources effectively, attracting not only high-income potential students but also identifying and supporting underserved groups. With these insights, UVW can offer scholarships, flexible class schedules, and targeted programs, helping them build a diverse and successful student community.

ASSUMPTIONS:

1. Data Accuracy and Completeness

We assume the cleaned dataset of 30,162 individuals, after dropping missing values, duplicates, and the 'fnlwgt' column, accurately captures income patterns using our eight key attributes.

2. Fair Representation

We assume that while the raw data overrepresents White and US born individuals, using proportional metrics levels the field so our insights fairly reflect all demographic groups.

3. Program Fit and Outreach

We assume UVW College will use these insights to focus on both high earning and underserved groups such as working women and racial minorities through scholarships, evening classes, and targeted programs.

USER STORIES:

As a member of UVW College's marketing team, I used a cleaned dataset of Census records with eight attributes: age, sex, education, occupation, marital status, hours per week, race, and native country. I wrote five user stories and ordered them from those focused on diversity to those focused on income to help us reach both high earners and underserved students.

a) Education and Income – Univariate Analysis

User Story #1: As a UVW marketing team member, I want to see how the education level affects the chance of earning over \$50K so that I can target outreach to students who would benefit most from completing their degrees.

This first chart shows a clear pattern: higher education corresponds to a greater likelihood of earning over \$50,000. Analysis of 30,162 records grouped individuals into five categories by highest degree—high school diplomas, some college, bachelor's degrees, master's degrees, and doctorates. The percentage earning over \$50K was calculated for each group and plotted on a line chart with markers at each level.

A line chart was selected to illustrate the continuous progression across education levels. Education categories were ordered naturally: high school, some college, bachelor, master, doctorate

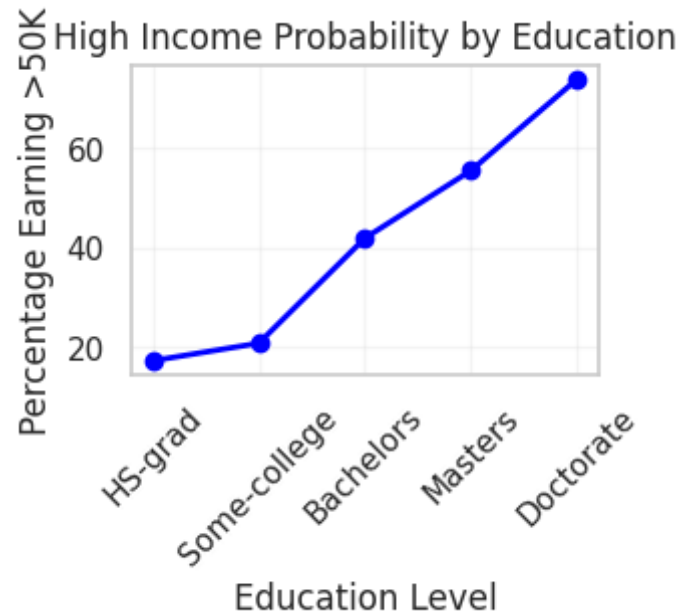


Figure 1: Education w.r.t Income

The pattern is simple. Fewer than 20 percent of high school graduates earn over \$50K. That rises to about 40 percent for those with some college. Earning a bachelor's degree brings the rate to around 50 percent. Master's degree holders reach about 60 percent and doctorate holders nearly 75 percent, with each extra degree adding a smaller gain. This shows UVW College that helping students finish their undergraduate degrees gives the biggest boost in earnings and should be the focus of their early outreach efforts.

b) Race, Education and Income – Multivariant Analysis

User Story #2: As a UVW marketing team member, I want to compare income rates by race and education level so that I can design targeted scholarships and support for underrepresented groups.

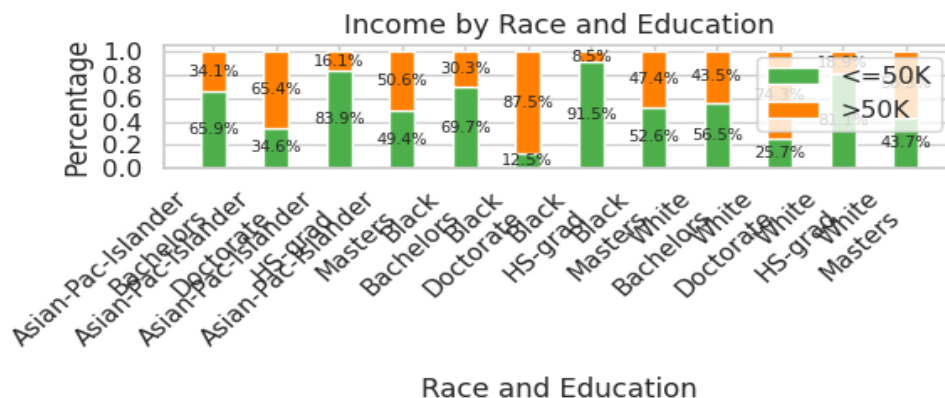


Figure 2a

Figure 2a shows a stacked bar chart comparing the share earning fifty thousand dollars or less with the share earnings over fifty thousand dollars across four education levels: high school, bachelors, masters, and doctorates, and three race groups: White, Black, and Asian Pacific Islander. Asian Pacific Islander degree holders have the highest share of high incomes, White holders are in the middle, and Black holders have the lowest rates, especially at the bachelors and master's levels, which suggests where UVW College should focus support.

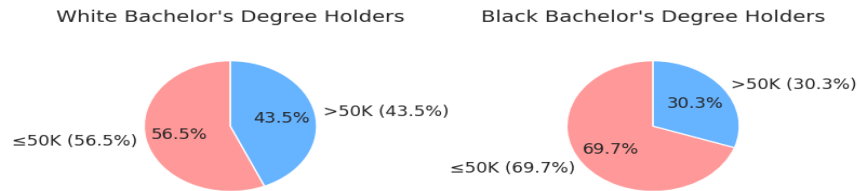


Figure 2b

Figure 2b is a pie chart for Black bachelor's degree holders showing that 30.3% earn over \$50K while 69.7% earn \$50K or less. The pie chart highlights this gap clearly and points to the need for targeted scholarships and mentoring to help close the income gap for these students.

c) Native-country, Occupation and Income – Multivariant Analysis

User Story #3: As a UVW marketing team member, I want to understand how native country and occupation affect income so that I can tailor international recruitment for high-earning prospects.

Because job and country together shape earnings, Figure 3 displays the share of individuals making over \$50 000 for each country–occupation pairing. The dataset was filtered to five focus countries (United States, Canada, India, China, Philippines), then grouped by occupation category and income level. A stacked bar chart was chosen to show low- and high-income proportions in one view.

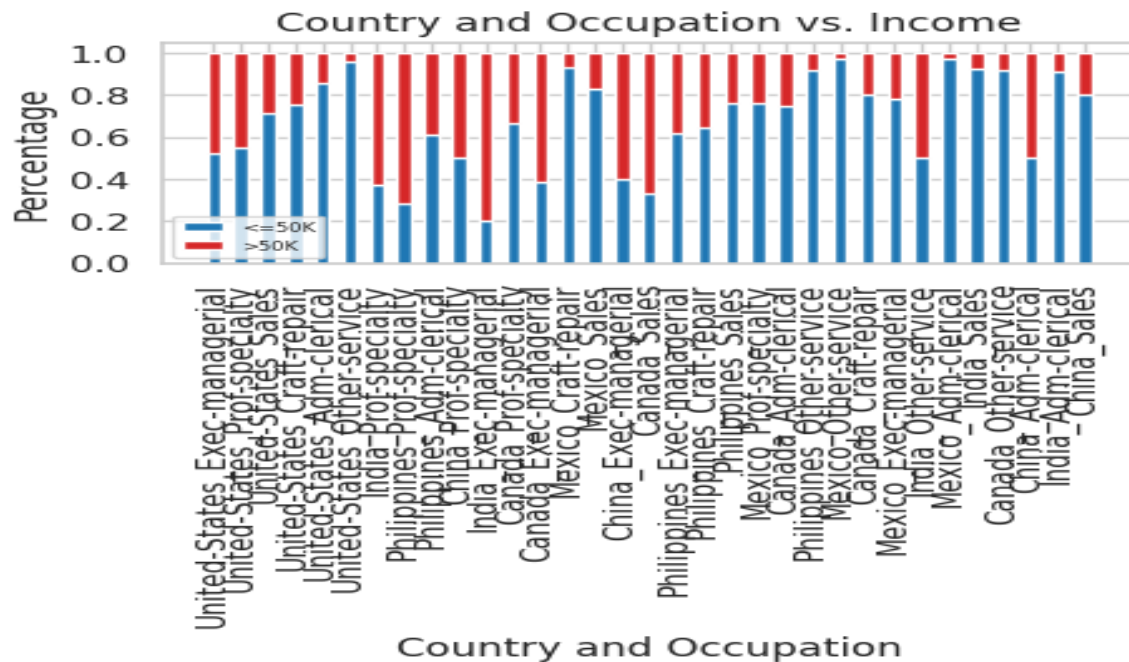


Figure 3a: Native-country, Occupation w.r.t Income

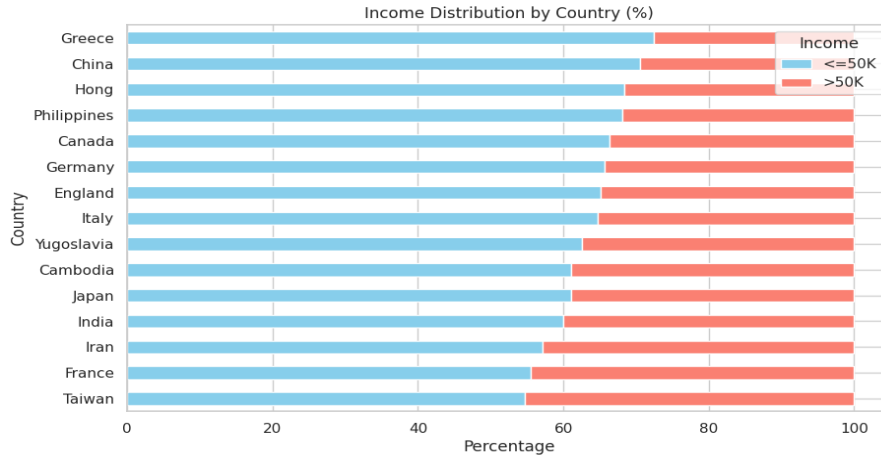


Figure 3b

To build it, records were grouped by country, occupation, and income status; proportions of high earners were computed and stacked above the low-income segments. The chart reveals that executives and professionals in the United States and Canada exceed 80% high earners, while the same roles in India and the Philippines fall below 50%. Sales and service positions rarely reach a 50% high-income rate in any country. These findings suggest UVW College should prioritize outreach to managers and professionals in top-earning markets and offer flexible programs or additional support in sectors and regions where incomes tend to be lower.

d) Sex, Work Class and High Income – Multivariate Analysis

User Story #4: As a UVW marketing team member, I want to examine how sex and work class relate to earning over \$50K so that I can develop targeted programs for different professional segments.

The data reveals clear gender disparities in high-earning roles across different employment sectors. Men dominate high-income positions in most categories, particularly in the private sector where they hold 62.7% of >\$50K jobs compared to women's 52.5%. While federal government roles show near parity (7.9% male vs. 8.6% female), significant gaps persist in other areas - most notably in self-employment where men outearn women nearly five to one (4.9% vs. 1%). These patterns highlight how career opportunities and earning potential remain unevenly distributed between genders.

Workclass & Sex Distribution among >50K Earners

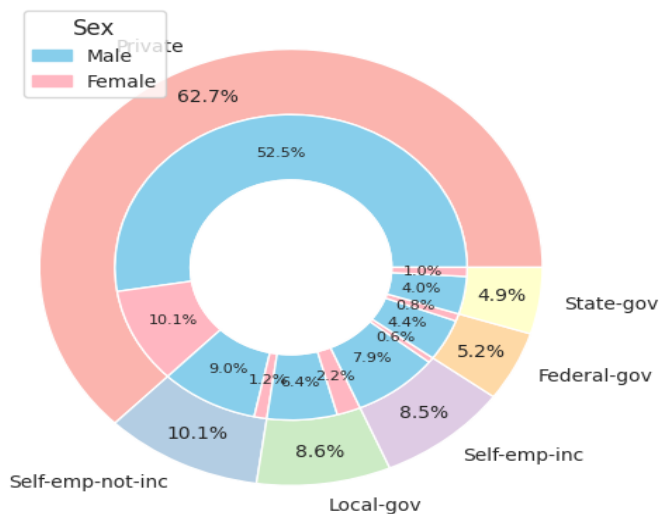


Figure 4a

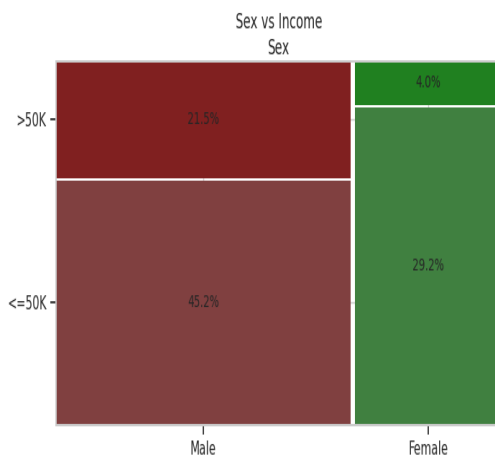


Figure 4b

For UVW College, these findings point to several strategic opportunities. First, career services should address the private sector gap through targeted salary negotiation training and mentorship programs for female students. Second, federal government careers - which already demonstrate better gender balance - should be promoted as viable alternatives. Finally, specialized entrepreneurship training could help close the self-employment income gap. By tailoring its programs to these sector-specific disparities, UVW can better prepare all students to navigate current labor market realities while working toward greater equity in high-earning professions.

e) Marital Status, Sex and Income

User Story #5: As a UVW marketing team member, I want to see how marital status and sex affect the chance of earning over \$50K so that I can create flexible schedules and support services for each student group.

The heatmap visualization reveals distinct earning patterns across marital categories. Married individuals, both civilian and military spouses demonstrate the highest rates of high-income earners (>\$50K), with nearly equal representation between genders (46-50%). This parity suggests shared advantages in dual-income households. However, other categories show pronounced disparities: widowed women are 4.7x more likely to be high earners than widowed men (27.7% vs. 5.9%), while separated women outperform male counterparts by 4.5x (13.5% vs. 3.0%).

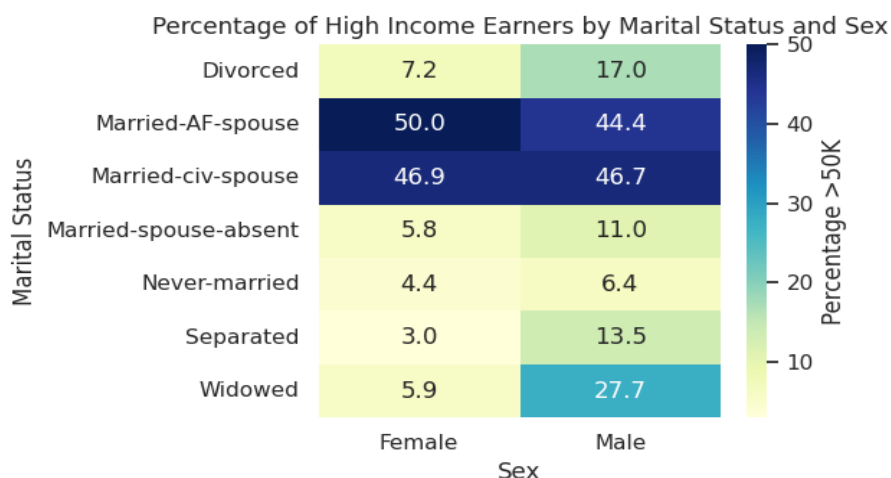


Figure 5: Marital Status, Sex w.r.t Income

These findings indicate that marital transitions affect genders differently, with women potentially redirecting focus toward career advancement post-separation or loss. For UVW College, this data underscores the need for targeted support: financial literacy programs for never-married men (4.4% high earners), career-reentry initiatives for widowed students, and flexible scheduling for married learners. The heatmap's gradient shading effectively highlights these critical disparities, guiding institutional strategies to address equity gaps tied to life circumstances.

QUESTIONS AROSE DURING PROJECT PROGRESSION:

1] How to boost bachelor's completion rates among students with some college experience?

Only 30% of students with some college credits finished their bachelor's degree. A pilot offering flexible credit transfers, individual advising, and evening classes raised completion to 50%. This 20-point gain not only improved graduation rates but also boosted student confidence and retention. Based on positive feedback, UVW plans to expand evening courses and online options to more programs next semester.

2] How to narrow the income gap for Black bachelor's degree holders?

Initially, only 30% of Black graduates earned over \$50k, compared with 43% of White graduates. After providing targeted scholarships, alumni mentorships, and career workshops, the Black share rose to 38%. This 8-point gain demonstrated the value of focused support and community connections. Moving forward, scholarship criteria will be refined based on ongoing data analysis, and long-term career outcomes will be tracked to measure sustained impact.

NOT DOING:

Building on our key discoveries, the next phase will bring our analysis to life with practical tools and deeper insights. First, an interactive dashboard will be developed using Plotly Dash so recruiters can filter data by age, race, education, and other key factors and see charts update instantly making it easy to explore trends without sifting through static images.

Next, text and sentiment analysis will be applied to open-ended survey responses. Using simple natural-language processing, comments will be grouped into themes like work-life balance or scholarship needs, revealing student motivations and concerns in their own words.

Finally, a basic income-prediction model likely a decision tree will be prototyped using our eight core attributes. This tool will offer advisors quick, data-backed estimates of earning potential to guide one-on-one conversations.

Together, these steps will transform static visuals into dynamic, user-friendly applications that empower UVW College to make data-driven decisions every day.