Answers to five questions about the salaries dataset

1. Which industries are the top 10% of earners associated with?
   - Financial
   - Tech
   - Government
   - Health care
   - Manufacturing

2. How would you characterize the top 10% of earners in each industry?

Financial
   - Education: evenly split between people with college and Master's degrees
   - Age: Between 25 and 44
   - Experience in the field: 5-10 years
   - Gender make up: evenly split between men and women
   - Race: entirely white
   - Locations: FL and MD

Tech
   - Education: range from high school education to PhD with the overwhelming majority having a college degree
   - Age: Between 18 and 64 with the largest group being 35-44
   - Experience in the field: 2-20 years with largest group having 11-20 years of experience
   - Gender make up: men significantly outnumber women and nonbinary people
   - Race: Nearly two thirds more white people than non-white people
   - Locations: WA, VI, TX, RI, PA, OR, NC, NY, MO, MN, MI, MA, IL, GA, FL, DC, CT, CO, CA, and AZ with CA having the largest number of respondents

Government
   - Education: evenly split between people with high school, college, and professional degrees
   - Age: Between 35 and 44
   - Experience in the field: 8-30 years
   - Gender make up: twice as many men as women
   - Race: majority nonwhite
   - Locations: CA and NY
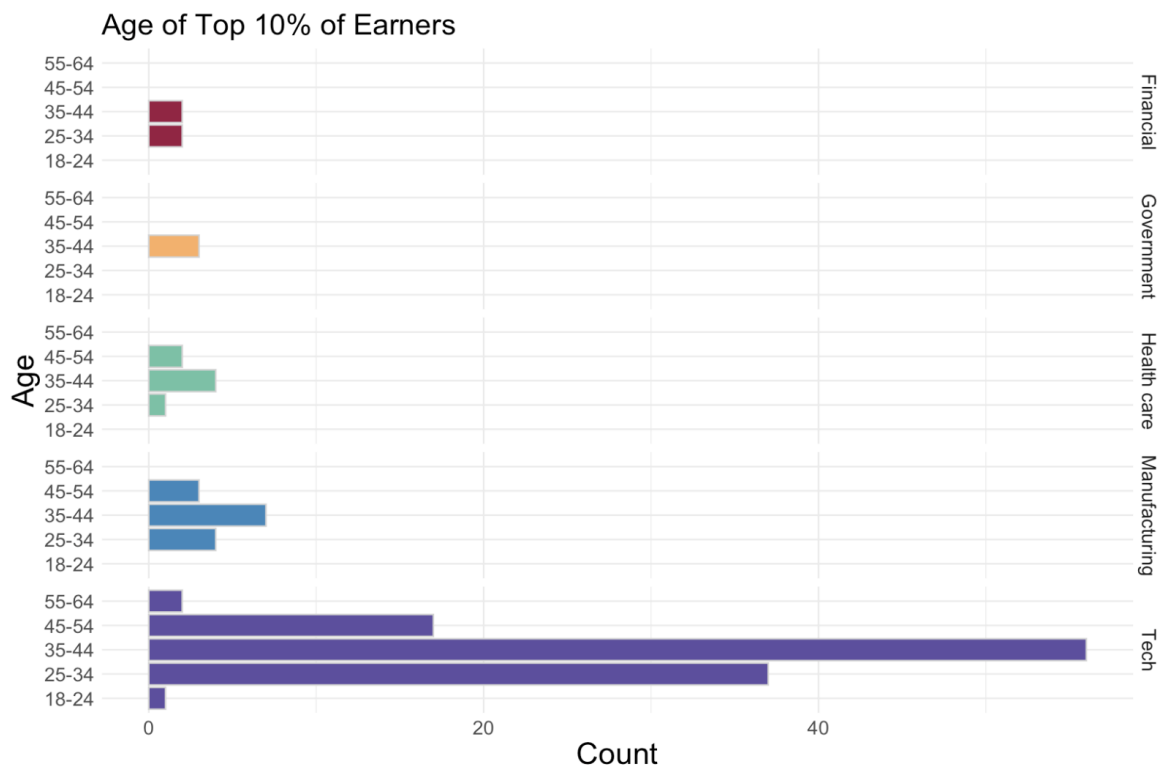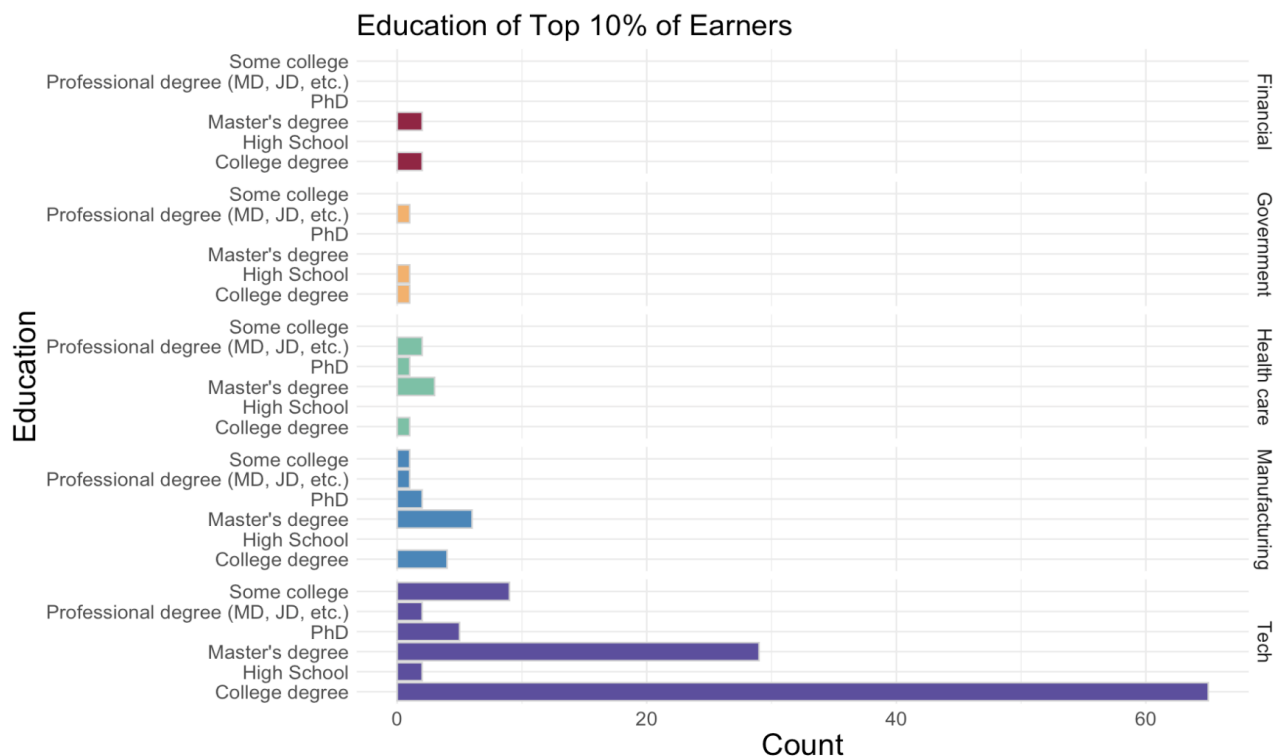
Healthcare
   - Education: college degrees, Master's degrees, professional degrees, and PhDs with the largest group having Master's degrees
   - Age: Between 25 and 54 with the largest group being 35-44
   - Experience in the field: 5-20 years with the vast majority having 11-20 years of experience
   - Gender make up: entirely women and non-binary people with the majority being women
   - Race: six times as many white people as nonwhite people
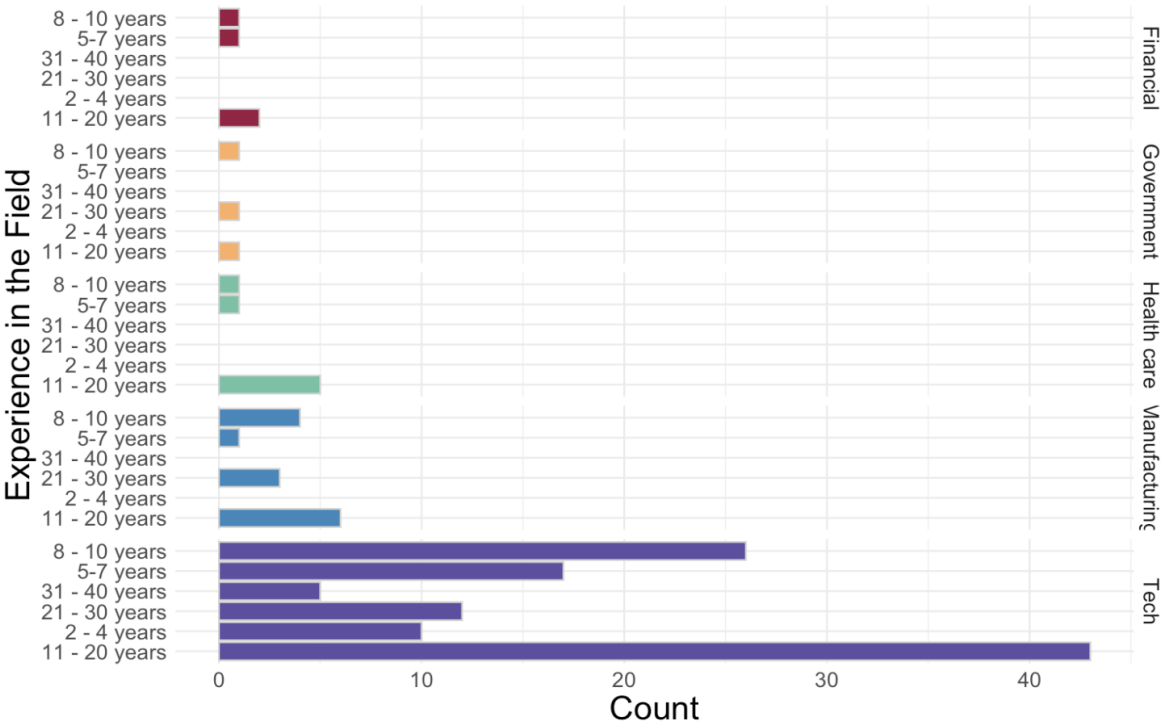   - Locations: UT, TX, OK, MN, and GA

Manufacturing
   - Education: some college, college degrees, Master's degrees, professional degrees, and PhDs with the largest group having Master's degrees
   - Age: Between 25 and 54 with the largest group being 35-44

- Experience in the field: 5-30 years with the vast majority having 11-20 years of experience
- Gender make up: men, women, and nonbinary people with the majority being men
- Race: majority white
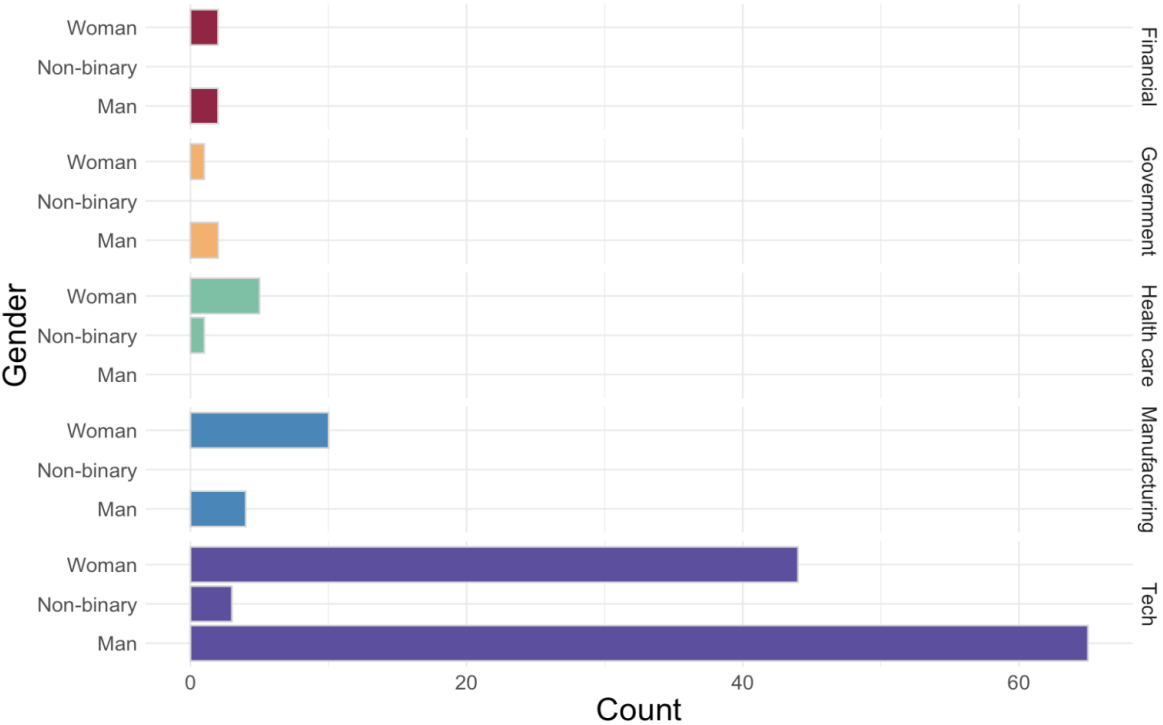- Locations: WA, TX, OH, NJ, MA, CA, and AL
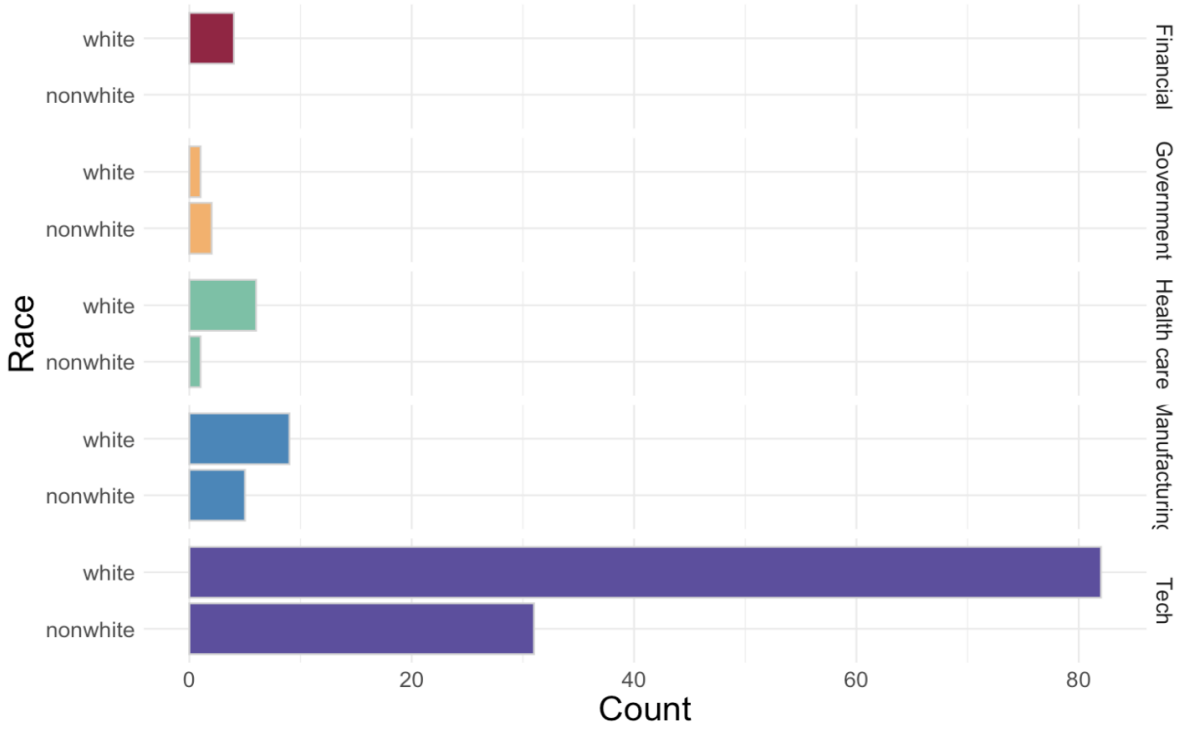
Summary Visualizations
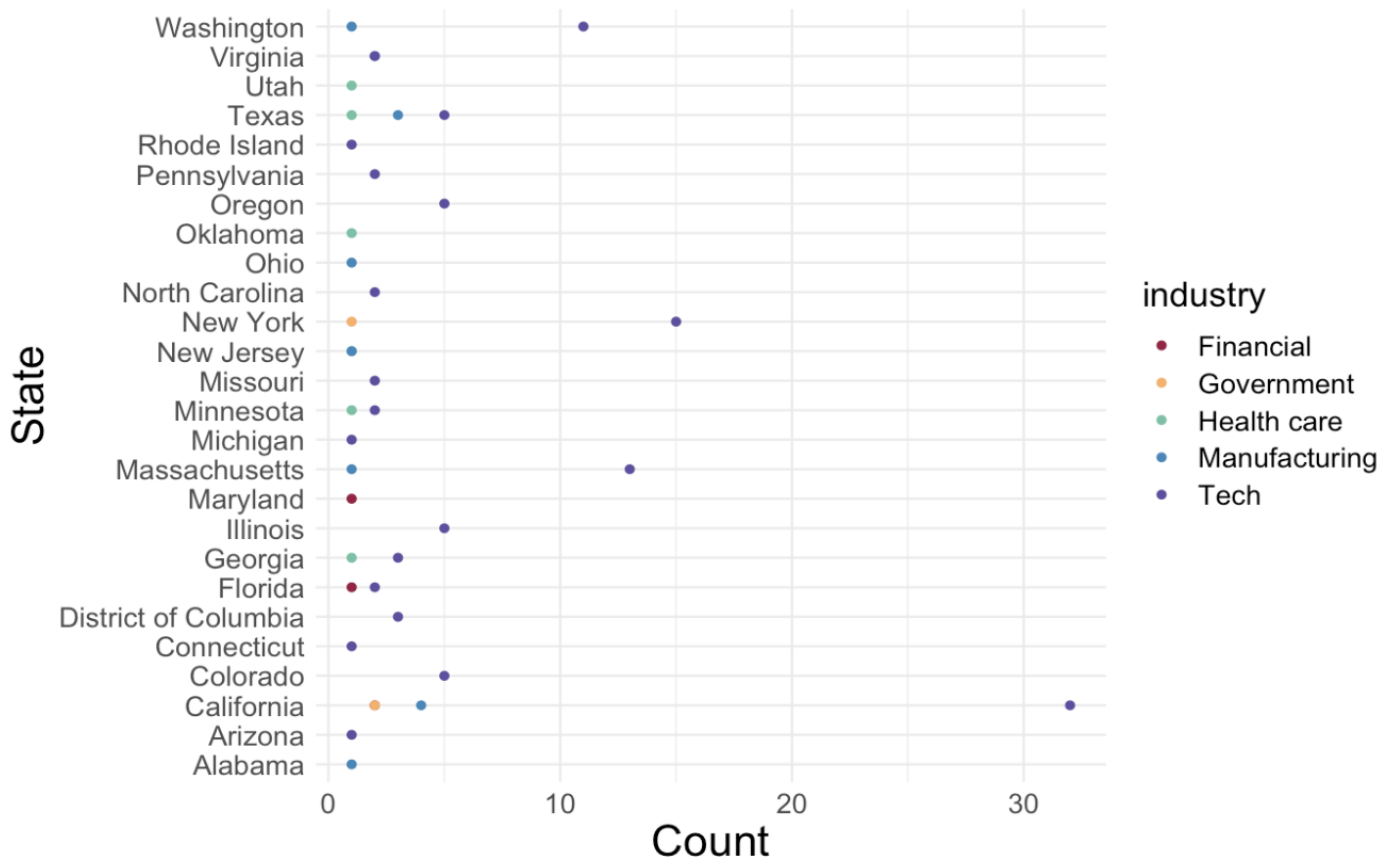

Education of Top 10% of Earners

Age of Top 10% of Earners

Experience of Top 10% of Earners

Experience in the Field (y-axis)

Count (x-axis): 0, 10, 20, 30, 40

Financial: 8 - 10 years, 5-7 years, 31 - 40 years, 21 - 30 years, 2 - 4 years, 11 - 20 years

Government: 8 - 10 years, 5-7 years, 31 - 40 years, 21 - 30 years, 2 - 4 years, 11 - 20 years

Health care: 8 - 10 years, 5-7 years, 31 - 40 years, 21 - 30 years, 2 - 4 years, 11 - 20 years

Manufacturing: 8 - 10 years, 5-7 years, 31 - 40 years, 21 - 30 years, 2 - 4 years, 11 - 20 years

Tech: 8 - 10 years, 5-7 years, 31 - 40 years, 21 - 30 years, 2 - 4 years, 11 - 20 years

Gender of Top 10% of Earners

Gender (y-axis)

Count (x-axis): 0, 20, 40, 60

Financial: Woman, Non-binary, Man

Government: Woman, Non-binary, Man

Health care: Woman, Non-binary, Man

Manufacturing: Woman, Non-binary, Man
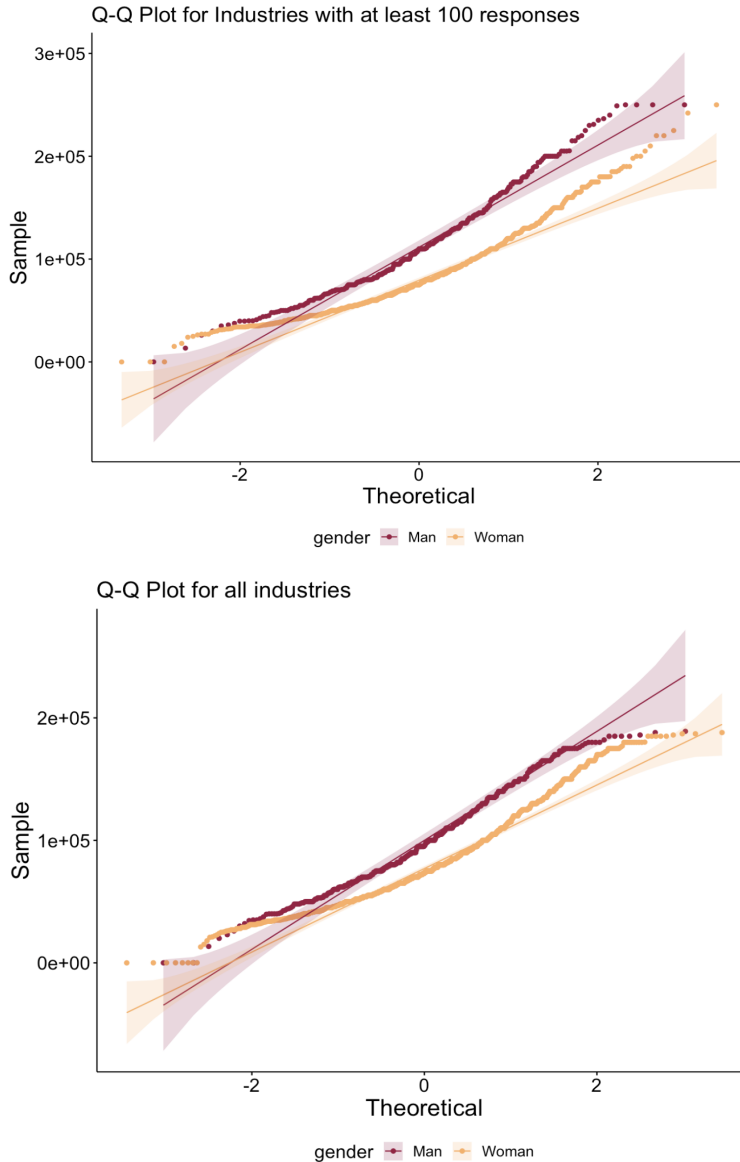
Tech: Woman, Non-binary, Man

Whiteness Among Top 10% of Earners

States of Top 10% of Earners

3. Are there statistically significant annual salary differences between genders in each industry? Overall?

Q-Q Plot for Industries with at least 100 responses


Q-Q Plot for all industries

The first Q-Q plot to the left shows only industries with at least 100 respondents while the second includes the full dataset.

I intended to run sets of two-sample t-tests to determine whether or not the pay differences observed in the dataset are statistically significant, however, in order to use the t-test the data would need to be normally distributed and there would need to be no significant difference in variance between the two comparison groups. In addition to running Shapiro-Wilks tests and F-tests, I created quantile-quantile plots (Q-Q plots) of the data to test for normality. Both the plots and the statistical tests showed that normality could not be assumed and that there was significant difference in variance between the two comparison groups. As a result, I could not use two-sample t-tests to assess the significance of the differences seen in the dataset and we are limited in our ability to attribute any observable differences to gender from a statistical standpoint.

We are also limited in our ability to control for years of experience and age because unfortunately the dataset includes only year ranges (categorical variables) and it's not clear that those ranges are representative of the thresholds at which you would be able to detect a significant difference using regression techniques. For instance, does pay start to be impacted at 13 years of experience or 17? We only have a range "11-20 years" so we cannot determine this. The same is true with the age ranges. These limitations would also be relevant in a granular examination of race or education within this dataset.

Despite these limitations I was still able to generate comparison visualizations and descriptive statistics to represent the gender pay gap seen within these industries and within the full dataset.

The below plot shows that men have higher salaries on average than women but the distribution of salaries between the two groups is similar within each industry. The largest pay disparity is found in the Tech industry.

Pay Differences by Industry
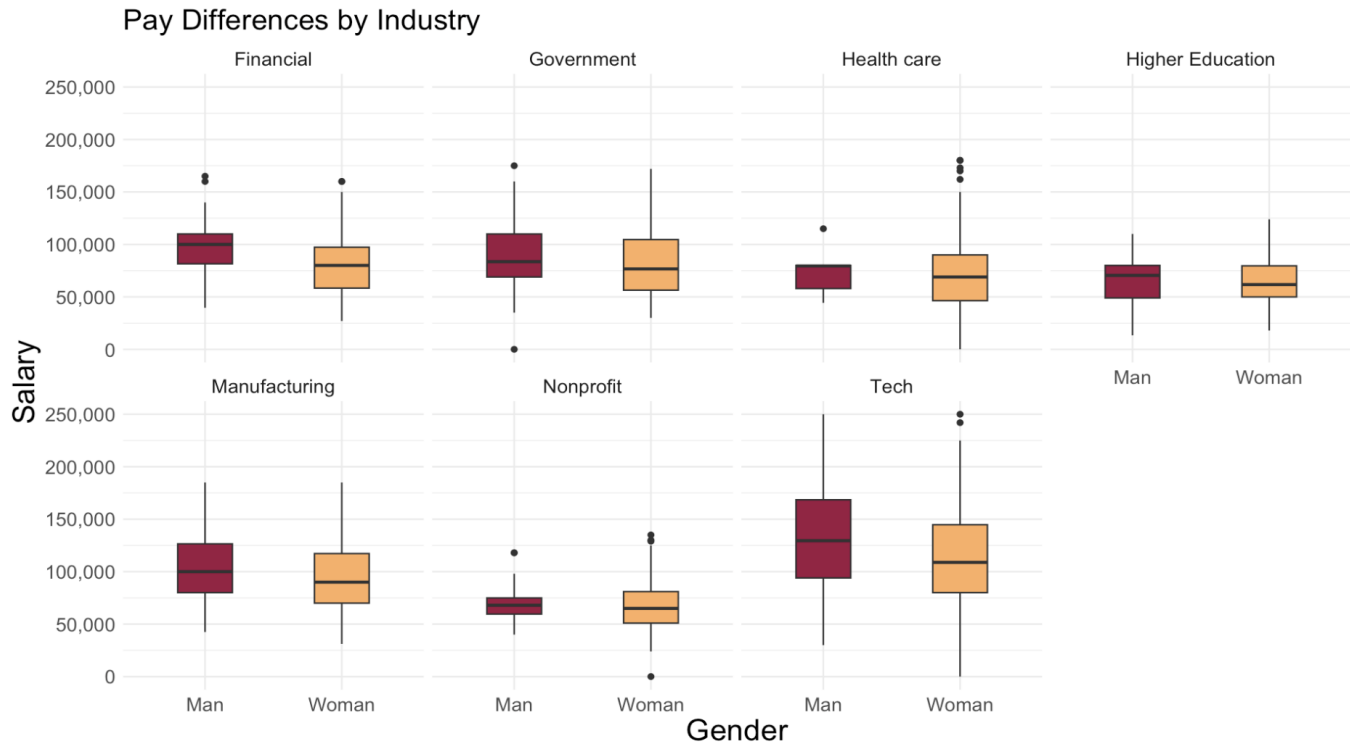
Industry	Gender	Mean Salary
Financial	Man	$99,421.07
Financial	Woman	$81,311.93
Government	Man	$87,377.27
Government	Woman	$82,309.44
Healthcare	Man	$74,880.83
Healthcare	Woman	$73,954.14
Higher Education	Man	$67,268.00
Higher Education	Woman	$65,438.85
Manufacturing	Man	$105,322.86
Manufacturing	Woman	$94,814.13
Nonprofit	Man	$72,176.47
Nonprofit	Woman	$67,674.13
Tech	Man	$133,547.42
Tech	Woman	$113,505.27

Industry	Gap Between Men and Women
Financial	-$18,109.14
Government	-$5,067.83
Healthcare	-$926.69
Higher Education	-$1,829.15
Manufacturing	-$10,508.73
Nonprofit	-$4,502.34
Tech	-$20,042.15

The final boxplot examining the relationship between gender and salary includes data for all industries within the original dataset.
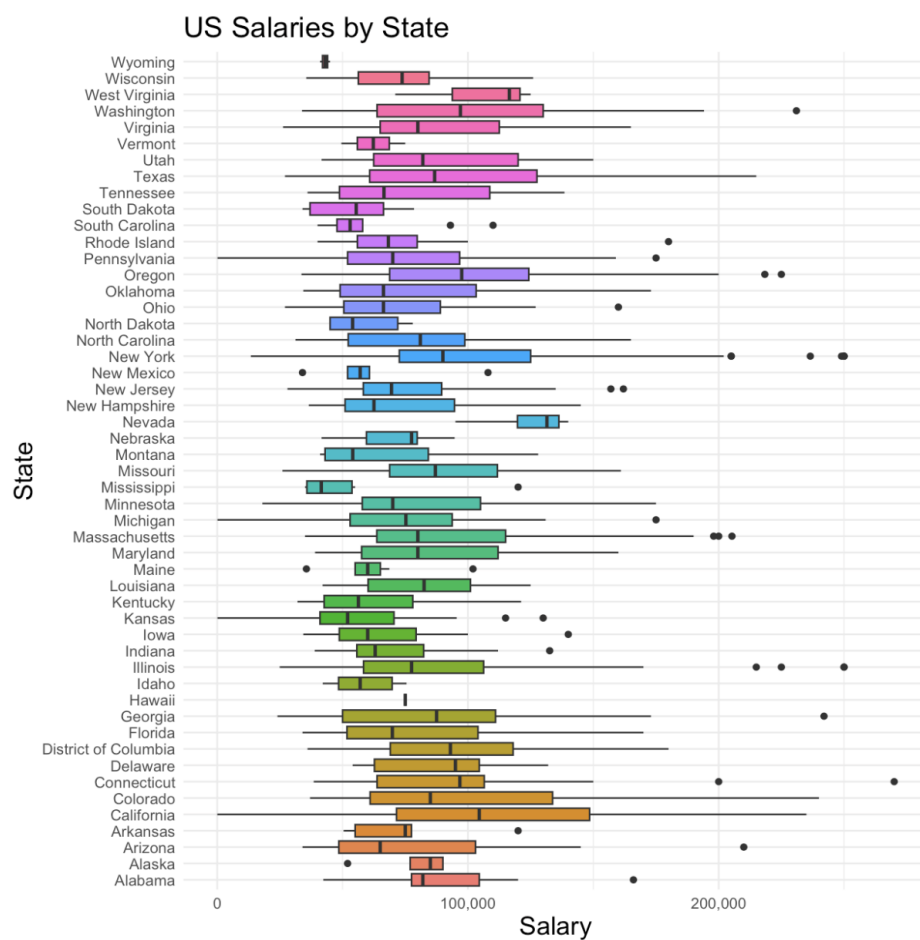
US Salaries by Gender
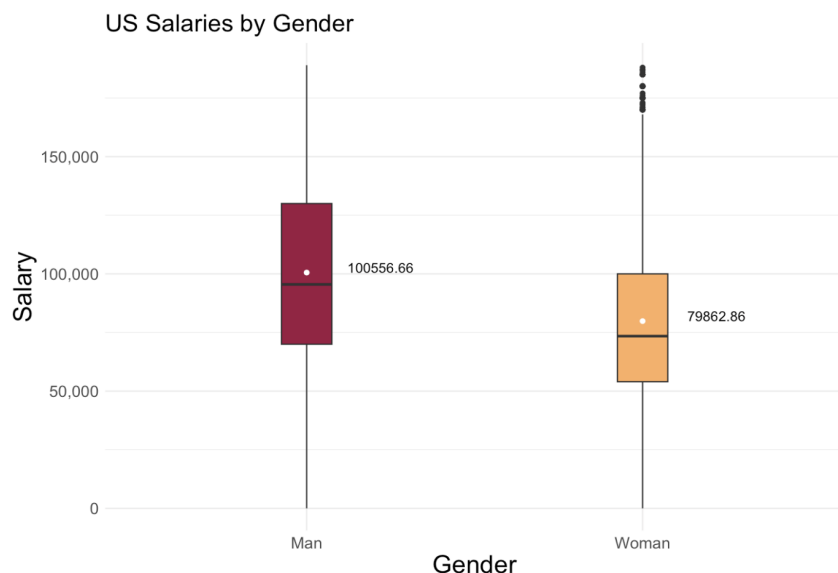
Average Pay Gap Across All Industries

-$20,693.80

4. Which industries pay the highest at entry level? Select for the top 10% of all salaries.
   ● Tech
   ● Financial
   ● Healthcare

5. Which states have the highest salaries? What about just among nonprofits? Select for the top 50 highest paying jobs.

US Salaries by State

States with the highest salaries across all industries:
   ● Connecticut
   ● New York
   ● Illinois
   ● Georgia
   ● Colorado
   ● California
   ● Washington
   ● Oregon
   ● Texas
   ● Arizona
   ● Massachusetts

US Nonprofit Salaries by State
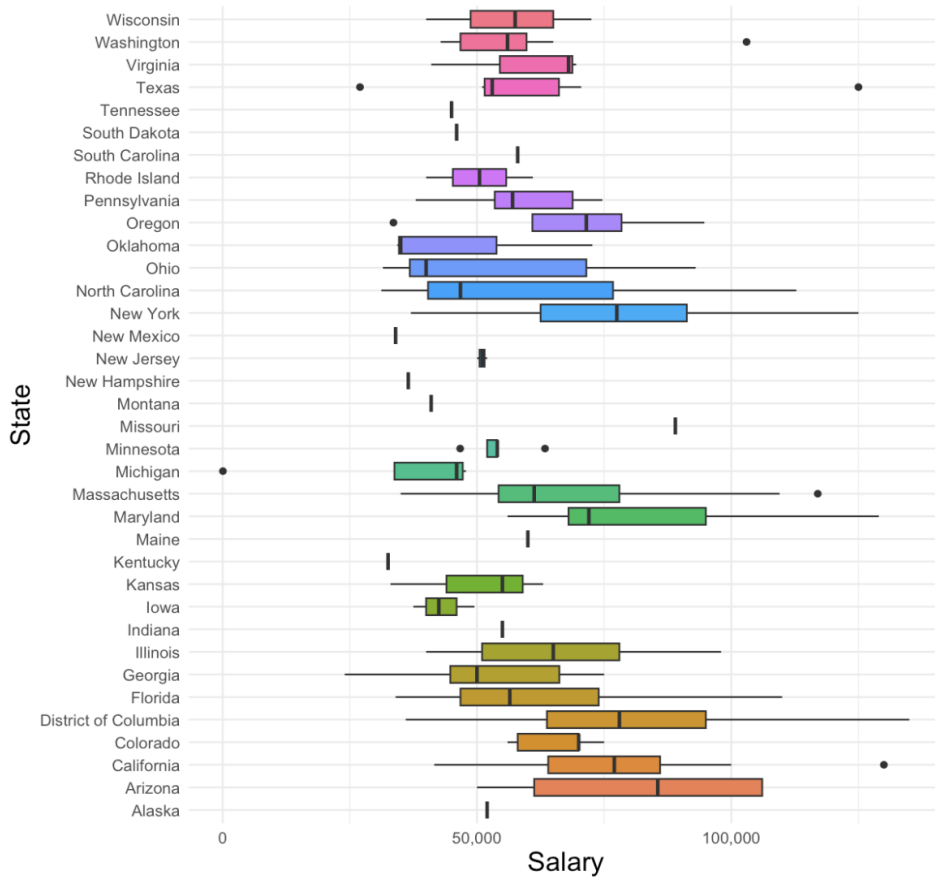
States with the highest nonprofit salaries:
- District of Columbia
- California
- Maryland
- New York
- Texas
- Massachusetts
- North Carolina
- Florida
- Arizona
- Washington
- Illinois
- Oregon
- Ohio
- Missouri