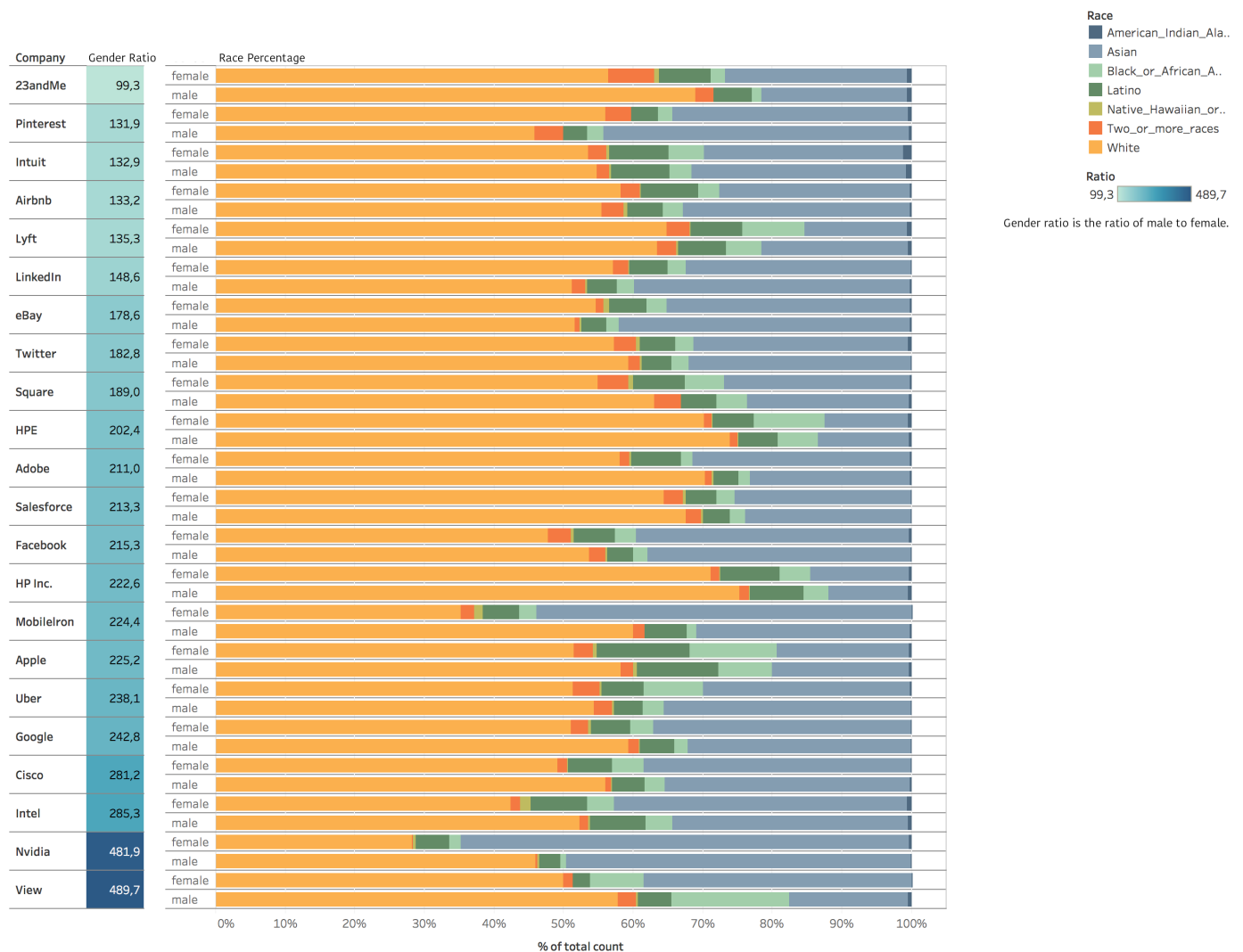


Diversity data of 22 Tech companies in Silicon Valley (2016)

The diversity in Silicon Valley - How are the race percentage and gender ratio of 22 tech companies in Silicon Valley in 2016?



Finding

The gender ratio in the visualization shows that there are far more male employees than female employees in these companies. Among the 22 companies, only 23andMe has a balanced gender ratio, other 21 tech companies are all having a significant difference between male employees and female employees.

This evidence suggests that employees in tech companies tend to be male instead of female, and the gap between male and female is very significant. Among them, View and Nvidia have the highest gender ration with the number of men is almost five times the number of women.

The bar chart on the right clearly shows that these 21 tech companies in Silicon Valley have more white employees and Asian employees, especially white employees. Light orange shows that white employees

account for more than half of the employees in these companies. And then the Asian is the second large ethnic group following by black and Latino.

Visualization Design

This graph shows the diversity data (gender and race) of 22 tech companies in Silicon Valley in 2016. The original dataset comes from EEO-1 reports filed by Silicon Valley tech companies in 2016. In the graph, the names of the 22 companies are plotted in the first column of Y-axis, and it's ordered by the gender ration in the second column from the lowest on the top to the highest on the bottom.

Gender ration is the ratio of male to female, calculating the number of males per 100 females. The higher the number is, the higher the male percentage of the total employment. The number 100 in gender ratio means having an equal count of male and female. The race data is shown in the bar chart on the right with different colors representing different races. The X-axis is scaled by the percentage which shows the different race in the percentage.

Document Notebook of the Exploratory Analysis Process

● Dataset Overview

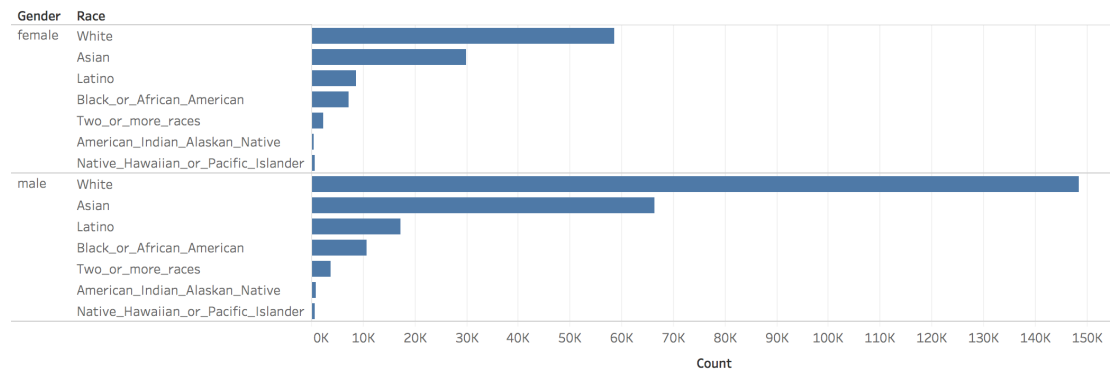
1. Title: Silicon Valley Diversity Data
2. Resource: <https://www.kaggle.com/rtatman/silicon-valley-diversity-data>
3. Description: This database contains EEO-1 reports filed by Silicon Valley tech companies in 2016. It includes 22 tech companies and there are six columns in this dataset: (The title of the website mention 23 companies but there's actually only 22 companies' data in the dataset)
 - **company**: Company name
 - **year**: For now, 2016 only
 - **race**: Possible values: "American_Indian_Alaskan_Native", "Asian", "Black_or_African_American", "Latino", "Native_Hawaiian_or_Pacific_Islander", "Two_or_more_races", "White", "Overall_totals"
 - **gender**: Possible values: "male", "female". Non-binary gender is not counted in EEO-1 reports.
 - **job_category**: Possible values: "Administrative support", "Craft workers", "Executive/Senior officials & Mgrs", "First/Mid officials & Mgrs", "laborers and helpers", "operatives", "Professionals", "Sales workers", "Service workers", "Technicians", "Previous_totals", "Totals"
 - **count**: Mostly integer values, but contains "na" for a no-data variable.
4. Acknowledgements: The EEO-1 database is licensed under the Open Database License (ODbL) by Reveal from The Center for Investigative Reporting.
<https://www.revealnews.org/article/hidden-figures-how-silicon-valley-keeps-diversity-data-secret/>

● Initial Research Question

What's diversity (race and gender) like for 22 tech companies in Silicon Valley in 2016?

● Initial Plot

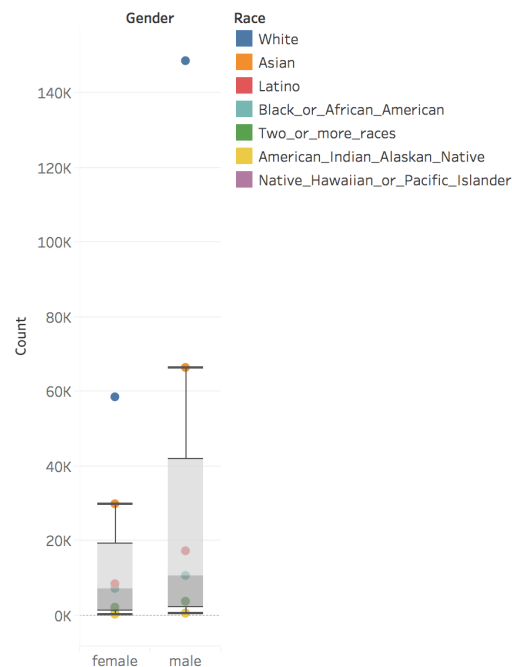
Gender & Race



The bar chart shows the overall total distribution of the gender and race in 22 companies.

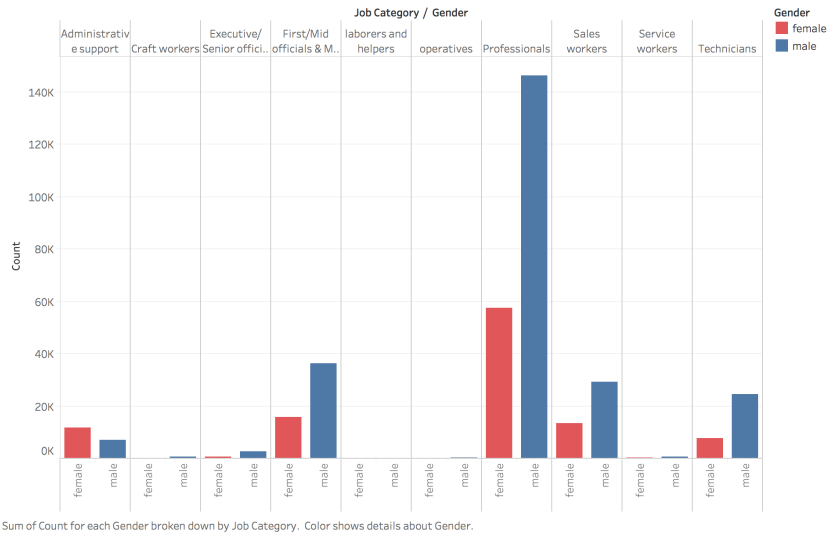
Gender & Race

Boxplot

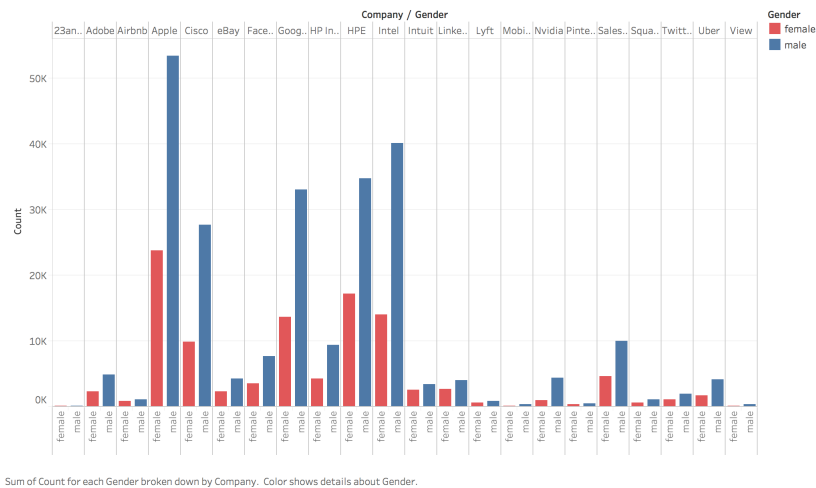


Sum of Count for each Gender.
Color shows details about Race.

The boxplot shows the Max, min and quantile of both female and male in those companies, which present more detail about the difference between male and female.



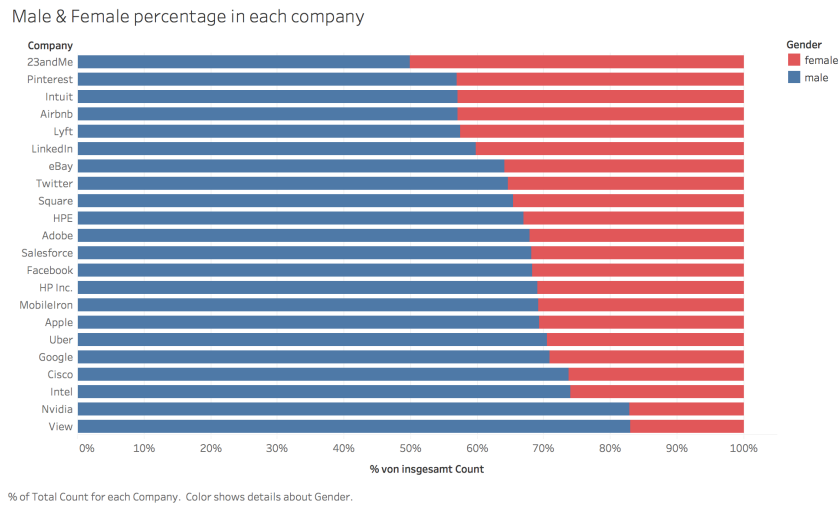
This bar chart shows the difference between gender in 10 job categories. It also depicts the distribution of 10 job categories. What is worth to mention is that only administrative support job has more female than male.



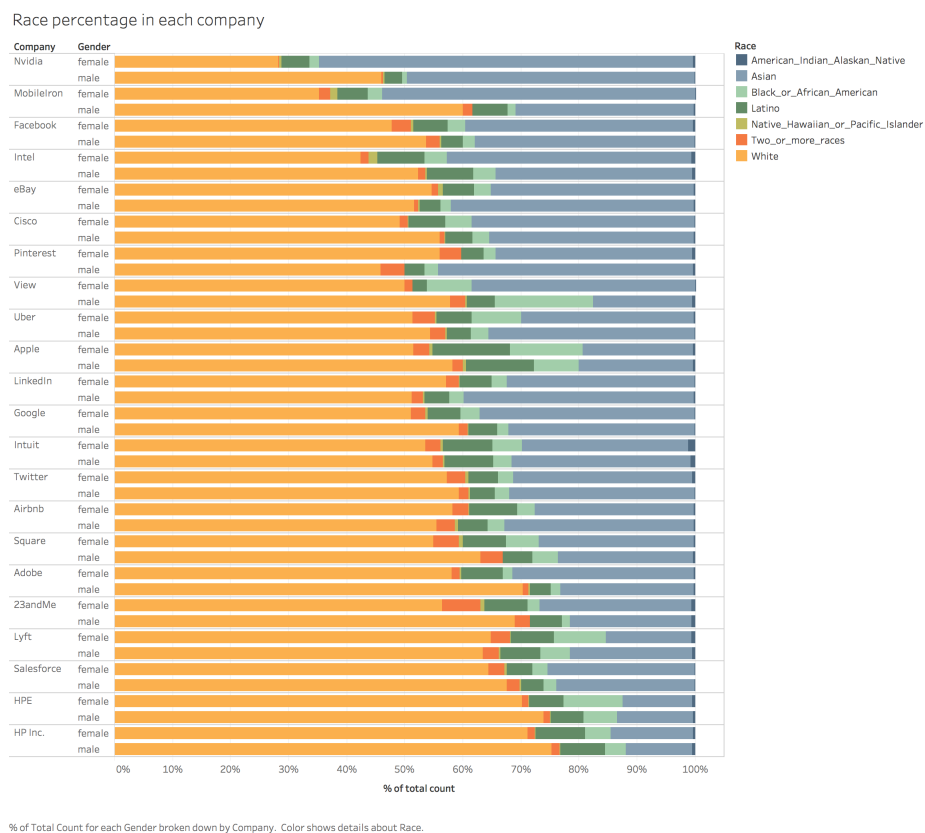
The prior plots are putting all companies' data together, but it's also important to show the difference between companies. With this plot, we can find that the scales of the companies vary a lot, so it would be more meaningful to exclude the effect of the scale of the companies.

● New Research Question

How are the proportion of different race and gender of 22 tech companies in Silicon Valley in 2016?



To eliminate the effect of scale, this bar chart plot the percentage of total count as X-axis and order by the percentage so that it's clear enough to compare.



By using the same way as plotting gender percentage in each company, this graph show the race percentage in each company.

The graph would be very messy if directly put the prior two bar chart together. To make the final graph more readable, I transform the gender data by calculating gender ratio of each company. This make it possible to show the diversity data of all companies in a more intuitive approach.

● Final research question & plot

How are the race percentage and gender ratio of 22 tech companies in Silicon Valley in 2016?

