Diversity in Tech Final Report

Vivian Omondi, Priyanka Dasgupta, Merry Li

Project Goals **2**

Related Work **2**
- Understand the Background 2
- Set Common Language and Appropriate Benchmarks 3
- Explore Deeper: What are the Overall Numbers Hiding? 5
- Direct Inspiration for Our Visualizations 6
  - Overview of Tech Industry Representation 6
  - Gender Balance: Which tech companies best represent the US population? 7
  - Racial Diversity in Tech Companies 8
  - Comparing the % of People of Color vs Women at the Top 10 Tech Companies 9
  - Diversity Across Different Roles at Netflix 10
  - Predicted Dates for Gender Parity at Top 10 Tech Companies 11

Visualization Description **12**
- Current State Assessment of Tech Industry 12
  - Overview of Tech Industry Representation 12
  - Gender Balance: Which tech companies best represent the US population? 13
  - Racial Diversity in Tech Companies 14
- Top 10 Tech Companies 15
  - Market Landscape 15
  - Comparing the % of People of Color vs Female at the Top 10 Tech Companies 16
  - Diversity Across Different Roles at Netflix 17
- Looking to the Future 17
  - Predicted Dates for Gender Parity at Top 10 Companies 17
- Our Recommendations 18
  - Recommendations 18
- What You Can Do to Make a Change 19
  - What You Can Do Cards 19

Approach **20**
- Data 20
- Tools 21

User Testing and Results **21**
- Overview 21
- Test Setup 22
- Result Analysis 22
- Overall Impression and Usability of Visualization 24

Link to Code and Visualization **24**

Team Contributions Breakdown **25**

Appendix **26**

Project Goals

A 2017 Pew Research Center study found that 90% of U.S. households have at least one connected device such as a smartphone or computer (Source). As such, there is an increasing need to design for everyone, with top tech companies like Google focused on developing for *The Next Billion Users,* with the lofty value proposition of "Building for everyone, everywhere" (Source).

Yet, while inclusive design has become a hot buzzword in the industry today, how effectively can tech deliver on these goals when the demographics of these companies are starkly different from that of its user base? Much has been promised and said about diversity in tech, but progress is slow and the same gaps still remain.

Thus, our motivation is to uncover the extent to which these high tech companies have not addressed these DE&I issues, specifically taking an intersectional approach that looks at both race and gender breakdown. In examining each company closely, we hope to hold them accountable and prompt further diversity efforts that are much needed in this space.

Related Work

Understand the Background

Because we intend for our website to convey a cohesive story that generates empathy in our users, we conducted background research to understand the context and reason for why diversity matters in tech.

To this end, we drew from the below sources:

- Social Issues of Information class at UC Berkeley (INFO 203): The linked class curriculum was extremely helpful in providing case studies and content as to the negative impacts of tech design if it is not carefully deliberated and informed by diverse perspectives.
- Coded Bias (2020): This linked Netflix movie in popular culture explores racial and gender bias in our modern technological systems. It points towards a need for individuals who are impacted by factors overlooked in tech design and, consequently, can contribute a different view that lends itself to more inclusive design. These diverse individuals should be involved in a tech product at its inception so that unintended negative consequences can be alleviated.
- News articles: The linked CNBC article (June 2020) provides a current view of the tech industry, allowing us to orient ourselves in the limited progress companies have made up to the present day. Articles like this also make us aware of the shortcomings behind companies' self-reported data as they can include non-corporate or non-technical employees to inflate their own performance.

Set Common Language and Appropriate Benchmarks

After gaining awareness that tech companies can self report data to “look good” to the public, we sought to understand how a few sample companies did so. We found that across company websites in this small sample, there was no standard as to how the data was presented and benchmarked.

In the screenshots below captured from Amazon, Salesforce, and HP, various metrics are used to measure “success” and different groups are randomly lumped together to create an overall positive statistic:

Exhibit Amazon: Amazon shows the gender breakdown of corporate employees but juxtaposes this graph against one detailing the averages for corporate in the U.S. This makes it seem as if not much progress is needed because Amazon is on track. One also does not know what underlying data makes up the U.S. corporate average (i.e., who performed the study, which industries were included).

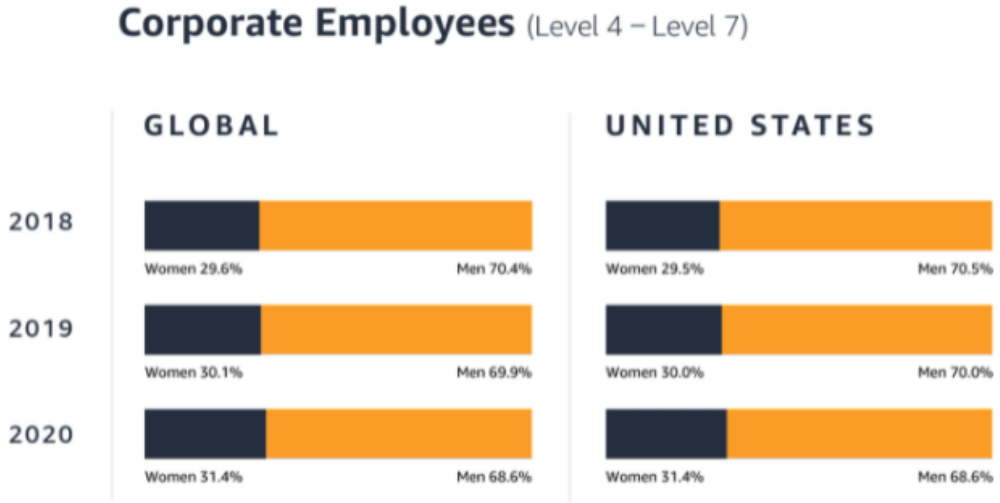


Exhibit Salesforce: Salesforce defines underrepresented groups as Women, Black, Latinx, Indigenous, Multiracial, LGBTQ+ employees, People with Disabilities, and Veterans. With such a wide net and benchmark for 50% representation of all these groups, Salesforce at 47.4% in 2020 seems to be very close to “building a workplace that reflects society,” as advertised on their Equality website.

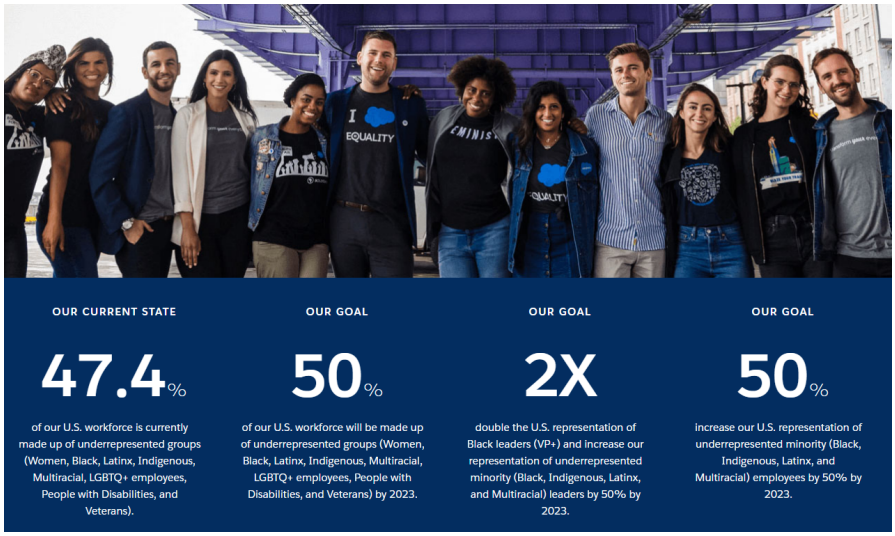
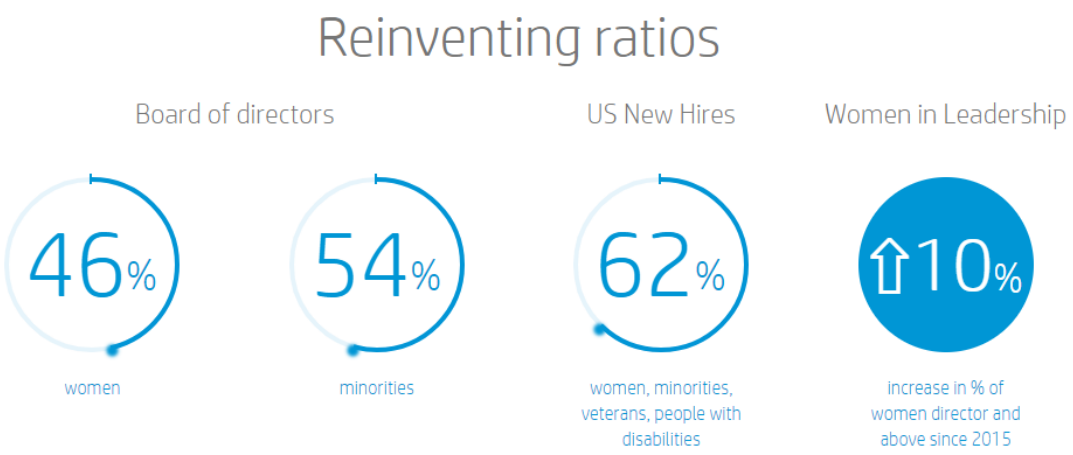


Exhibit HP: The groups that HP lumped together are very similar to Salesforce, with the key difference being that they do not define “minorities.” This leaves the viewer to guess if “minorities” also include the Asian group which could inflate the numbers.



Thus, we identified that each company uses a different language and metrics to represent their diversity efforts. Yet, in order to make a fair comparison between companies, we needed to set the same metrics and common language which would be used equally across all tech companies in a standardized manner. The following outlines the standards we defined and adhered to in all of our visualizations:

- **Focus on 5 main groups (Hispanic, Asian, Black, Native American, Female)**
    - Reason: While tech companies must also consider representation of other racial minorities, LGBTQ+ individuals, veterans, and people with disabilities, we tried to simplify the already complex narrative so as not to increase cognitive overload in the viewers of our website. We chose to focus on these 5 groups at most for the purposes of our project as they have the largest percentage representation in the U.S. population and are the most debated statistics in current media and culture.
        - Note that some visualizations do not include Native American because data was not available.

- **Evaluate each group separately in order to understand progress made towards their individual representations in tech**
    - Reason: When groups are lumped together as the companies above had done, it is difficult to tell an accurate story of diversity efforts as one group's progress or lack thereof can be eclipsed by another group's.
- **Compare the percentage makeup of each group against their percentage makeup in the U.S. population**
    - Reason: Our goal is to push for the demographics of technology creators to be the same or at least similar to the demographics of their users. Thus, comparisons to other U.S. averages in the tech industry or corporate is irrelevant to our project's purpose and can be disregarded.
- **Find a standardized template that requires all tech companies to report their workforce demographics in the same manner and leverage that data to make comparisons**
    - Reason: When companies self-report their data, they will include or not include certain categories of employees to inflate their statistics. From our research, we found that all companies are required to submit data in the same format to the federal government via EEO-1 reports and so used these (discussed in more detail under the Data section).

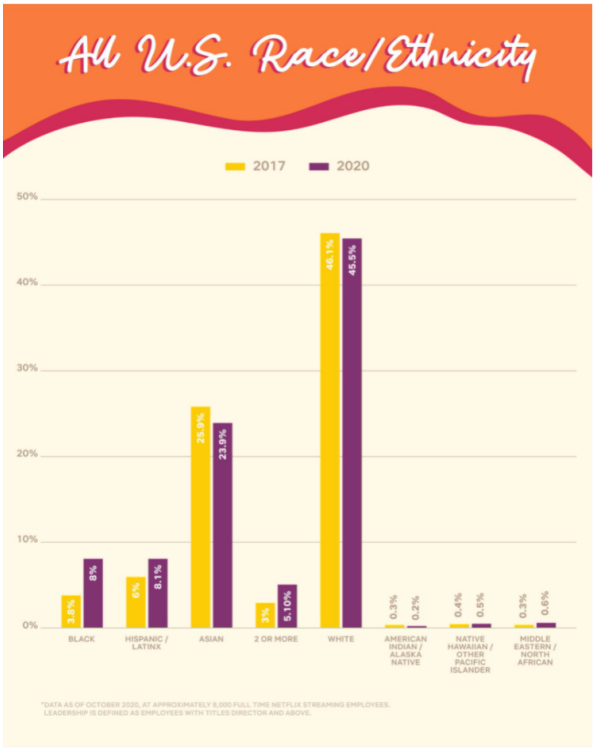
Explore Deeper: What are the Overall Numbers Hiding?

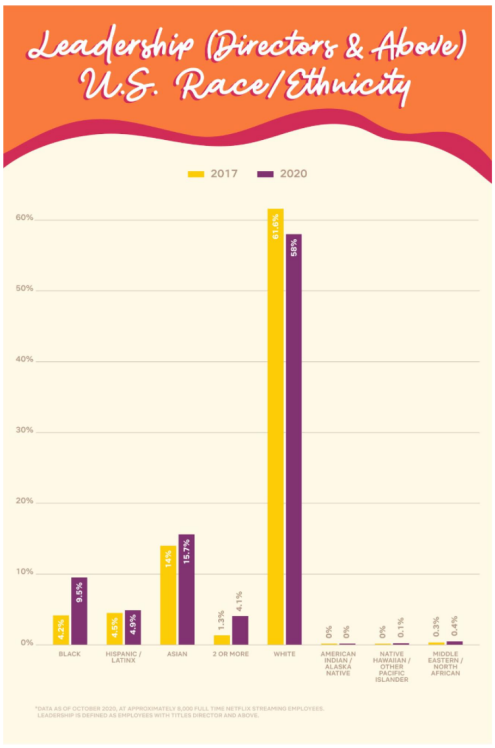
The overall numbers do give a high-level view that allows us to compare tech companies to each other in terms of representation in each of the 5 groups. However, we found graphs like the two below from Netflix which show that overall representation across all roles is higher than representation at leadership levels. Thus, overall numbers cannot give an accurate gauge of internal company culture and retention practices that are either inclusive or not inclusive to the individual as they progress in their careers.

Key Takeaways from the Graphs:

As an individual grows from an entry level employee to director and above, the below details what occurs as an aggregate:

- White representation increases by 27%.
- Black representation increases by 19%.
- Asian representation decreases by 34%.
- Hispanic representation decreases by 40%.





Because of this, we included visualizations on our website that allowed users to dig deeper past the aggregate totals and explore representation across race, gender, and career levels to explore what overall numbers may be hiding (Diversity Across Different Roles at Netflix viz).

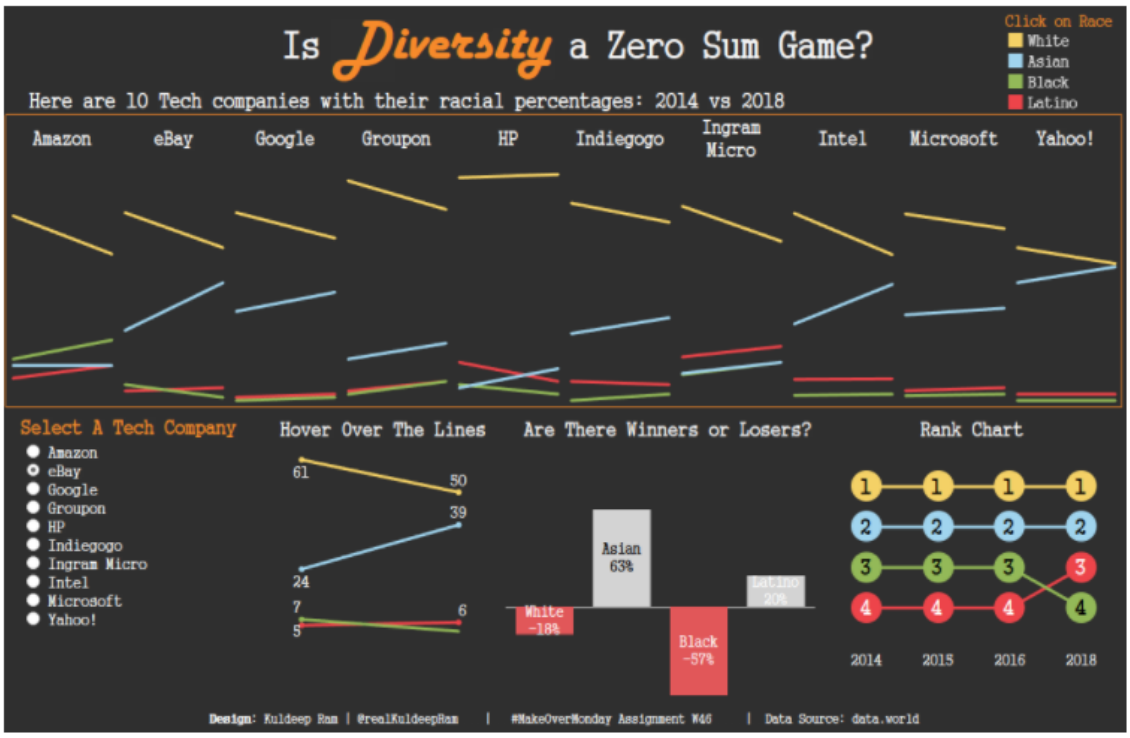
Inspiration for Our Visualizations

This section is divided based on the visualizations we created for our website. If any work was consulted for inspiration in the creation of that visualization, we will include it here.

Overview of Tech Industry Representation

As part of the narrative for our website, we needed to elicit empathy from the user and draw them into a story that will unveil more and more layers. Therefore, we needed an engaging visualization at the beginning of our website which at-a-glance could provide a snapshot of the current statistics for diversity in the tech industry.

One example that provides an overview is the graph “Is Diversity a Zero Sum Game?”

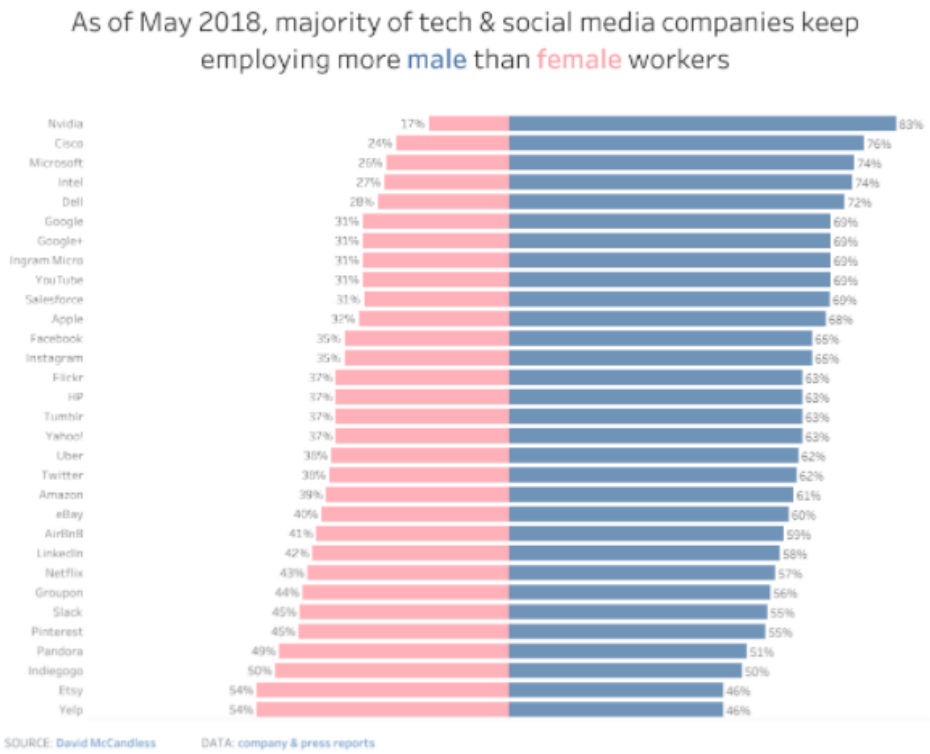


Source: Kuldeep Ram

The author picks 10 companies to compare against each other and uses slope graphs to show trends from 2014 to 2018. However, because the author selected 10 companies only, we found that this viz would not provide a robust snapshot of the industry and does not have reasoning for which companies were included or not. Also, the viz did not have any axises so we could not identify the percentage makeup of each race in the company. Lastly, while this viz does compare companies against each other, it does not address the main goal of our website which is to explore how they compare against the U.S. population. Thus, this viz gave us a starting point as we considered which aspects we would like to leverage and which to improve upon in our iteration of an overview of tech companies.

Gender Balance: Which tech companies best represent the US population?

We looked at graphs depicting the gender gap like the one below. The blue bar (for Males) in every tech company extended far away in the positive direction from the pink bar (for Females). Because we had studied data at these companies, we knew it meant that males are disproportionately represented in the tech workforce compared to females. We also knew the place where the two bars met in the middle was the gender parity line, an approximation being 50%. However, for a viewer without this context, the x-axis not being labelled in the graph below is a huge impediment as the viewer must guess what the percentages actually mean. Additionally, using light blue and pink to represent the genders serves to perpetuate existing stereotypes of female vs. male.

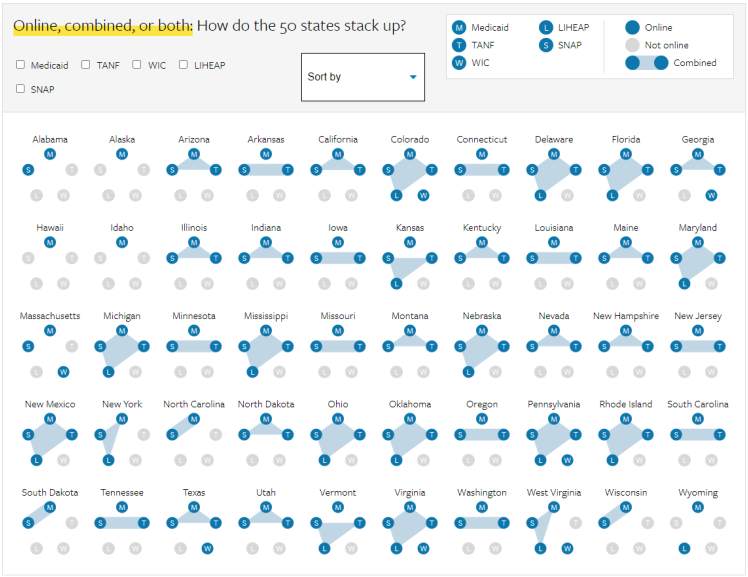


Source: Anna Ivanova

To this end, we did find a visualization that represented the gender gap very well and used it as a template (Source: Diego Parker). We built on top of it by including the U.S. population as a benchmark so companies that do better or worse can be immediately perceived. Also, we highlighted only the companies that are doing the best because it is such a small number in comparison to the number of companies in the entire industry.

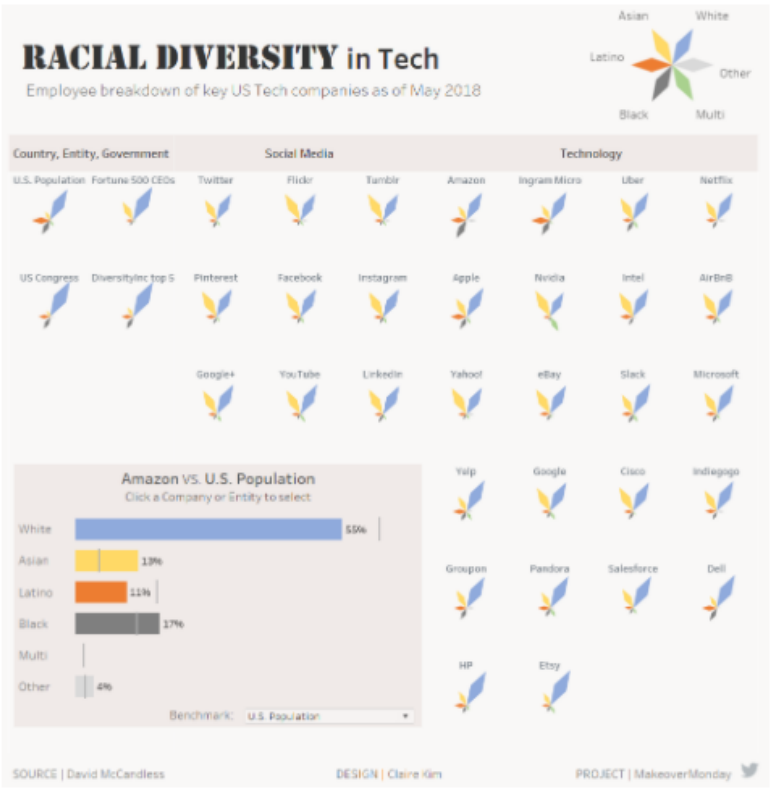
Racial Diversity in Tech Companies

In guest lectures for this class, we discovered that many entities can be represented in the same view without clutter because principles of consistency are followed. In Kerry's graph depicting the different benefits offered online by the 50 U.S. states (below), one can quickly see which state offers what online service. After initial familiarity with the structure, the user can carry that mental schema forward to subsequent states because they all follow the same design.



[Source: Kerry Rodden]

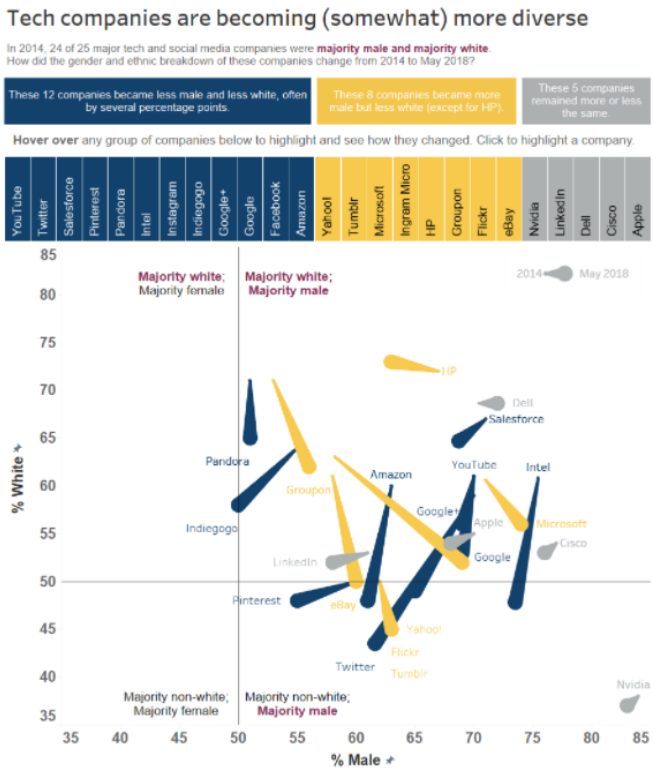
We combined Kerry's idea with the graph "Racial Diversity in Tech" below. Like Kerry's graph, this graph represents all tech companies in our dataset and keeps consistent standards throughout. If the percentages represented by the petal were not clear, the user could hover over the company to find a bar chart with the exact percentages for each race. Yet, we did feel that extraneous information (like Fortune 500 CEOs, U.S. Congress, and DiversityIncTop50) would not be needed for our site, as our purpose is to compare company stats to that of the U.S. We also found the partition between Social Media and Technology to be unnecessary since Social Media companies are a subset of Technology companies. To call them out separately would cause confusion as users would wrongly think they cannot be assessed together.



[Source: Claire Kim]

Comparing the % of People of Color vs Women at the Top 10 Tech Companies

In our viz, we wanted to compare the top 10 companies to each other and identify clear top performers. To do this, we needed a viz that could consolidate all of the diversity data (for both gender and race) in one view. In planning our design, we came across the graph below “Tech companies are becoming (somewhat) more diverse.”



Source: Steve Gesuale

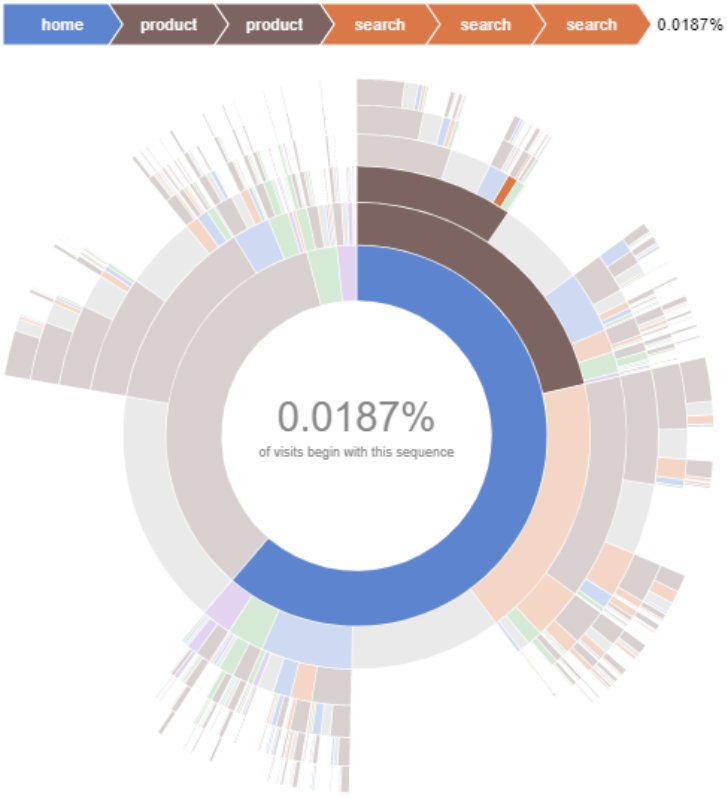
At a glance, this graph tells a powerful story. In dividing the view (based on % White and % Male at companies) into 4 quadrants and plotting all the tech companies in our dataset based on their representations, the author clearly conveys that the tech industry is “Majority white; Majority male” (greater than 50% male and 50% white) as most tech companies fall within this category. We wanted to leverage this idea about quadrants in our viz as well, yet take a different approach with the x and y axis being % Female and % POC so that we can look at diversity directly instead of indirectly as this graph does. We also liked that the companies were each plotted in this graph so we can track their individual progress as well as compare them to others. Yet, because the lines are in close proximity to each other, it was very hard for us to use preattentive attributes to find data that stood out from the crowd. For our own viz, we would find a way to distinctly represent companies and make them relatable to the user.

Diversity Across Different Roles at Netflix

To visualize the data as part of this case study, we wanted the user to be able to explore the complete breakdown of diversity data by not only race and gender but also by career levels (professional, manager, and executive). This required a viz with interactivity that could represent all of these components on their own as well as part of a whole.

In class, we were introduced to Kerry Rodden’s sunburst sequence diagram that visualized data as a whole but then also enabled the user to delve deeper into individual branches. As Kerry provided open source code, we repurposed the diagram with the Netflix 2018 EEO-1 data loaded in. We could have also represented our data using an icicle chart as it has the same root

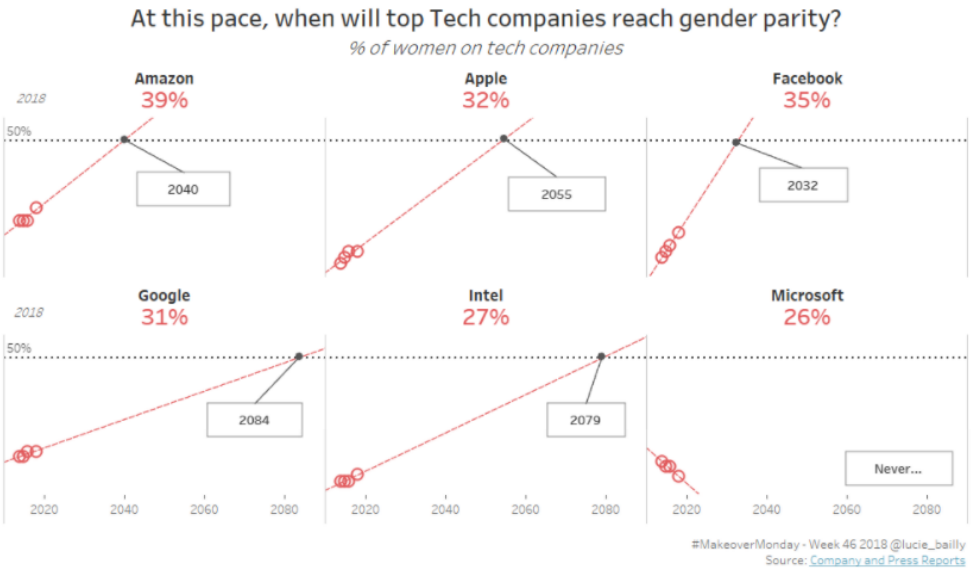
and branches, yet we liked the concept of the circle. The circle represents that all data is included within it and any segments result from slicing and dicing the original data. We can also quickly see which paths and sections are larger in comparison to others. We liked that Kerry's code included the breadcrumbs at the top, which allow the user to locate the path they took without having to memorize names and colors from a separate legend.



Source: Kerry Rodden

Predicted Dates for Gender Parity at Top 10 Tech Companies

We liked the concept of using trendlines to predict when select tech companies will reach gender parity (defined as 50% female and 50% male). While we leveraged the use of trendlines to estimate dates that we used in our final timeline for the top 10 tech companies, we found other parts of this design to not be beneficial to the story we were telling. Because 60 years is kept on such a short x axis, the message that the author is trying to convey loses its impact because in judging by distance in the graph alone, it does not appear that 2084, for example, is so far away from the present day. Also, laypeople who do not know about the concept of trendlines will not understand the purpose of the red circles and dotted lines and thus perform extra mental math when not required.



Source: Lucie Bailly

Visualization Description

This section follows the format of our website. We will list each website section and associated visualizations used by title for better navigation.

Current State Assessment of Tech Industry

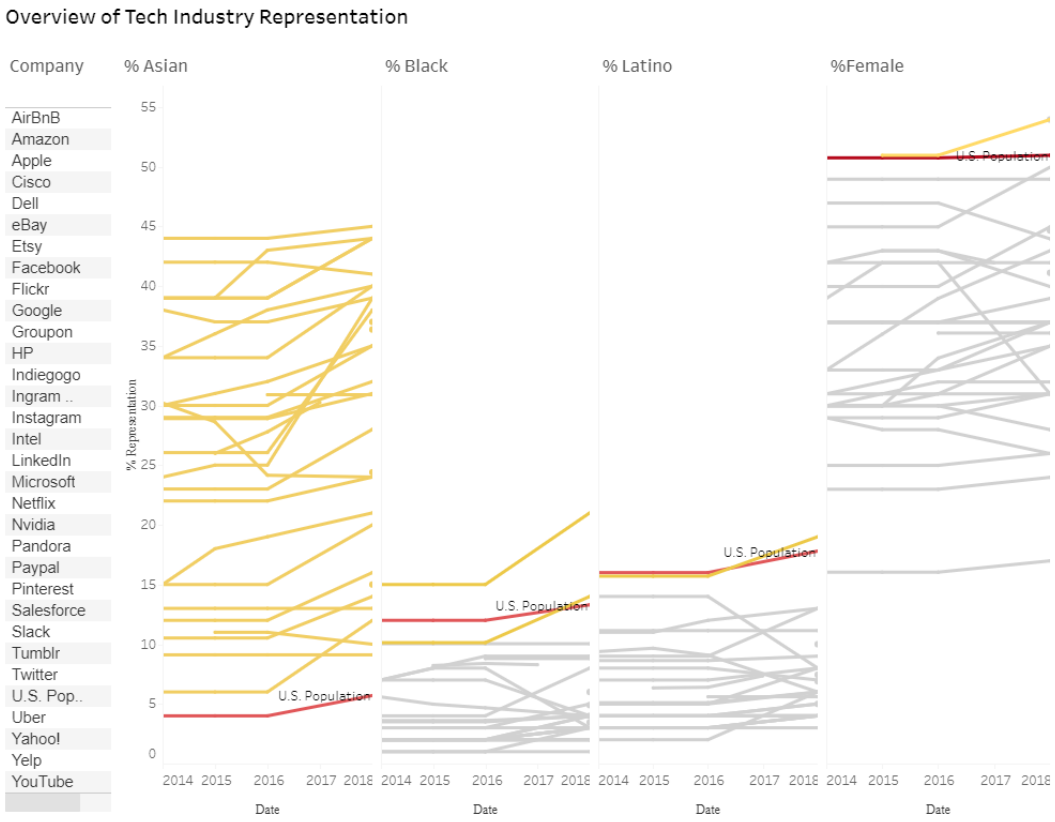
Overview of Tech Industry Representation

To begin our website and user journey, we wanted to provide an overview of the tech industry with as many companies as possible. We found the most comprehensive data set on data.world (refer to Data section). We wanted to hone in on not just racial and gender percentages but also on trends and changes in these two areas from 2014 to 2018. We experimented with many designs like bar graphs but ultimately decided that line graphs would best represent these two areas. We plotted company line graphs for each race or gender individually but then brought it all together by lining up the four graphs horizontally and ensuring that they shared the same axes range so that comparisons can be quickly made across all.

While we realize that we could have taken an extremely detailed intersectional approach in plotting gender+race (i.e., Black Female, Black Male, Asian Female, Asian Male, etc), we decided to only list the 4 underrepresented groups that are most cited and reported on by tech companies and media sources. The Female, Black, Asian, and Latino narrative is widely discussed and would be familiar to our intended audience of tech students and professionals. This familiarity would allow them to grasp our overview graph quickly and uncover overall insights in how each tech company is performing in these areas.

One suggestion that was made by several participants during our usability study was that there were too many labelled lines (representing companies) on the graphs. To get around this, we

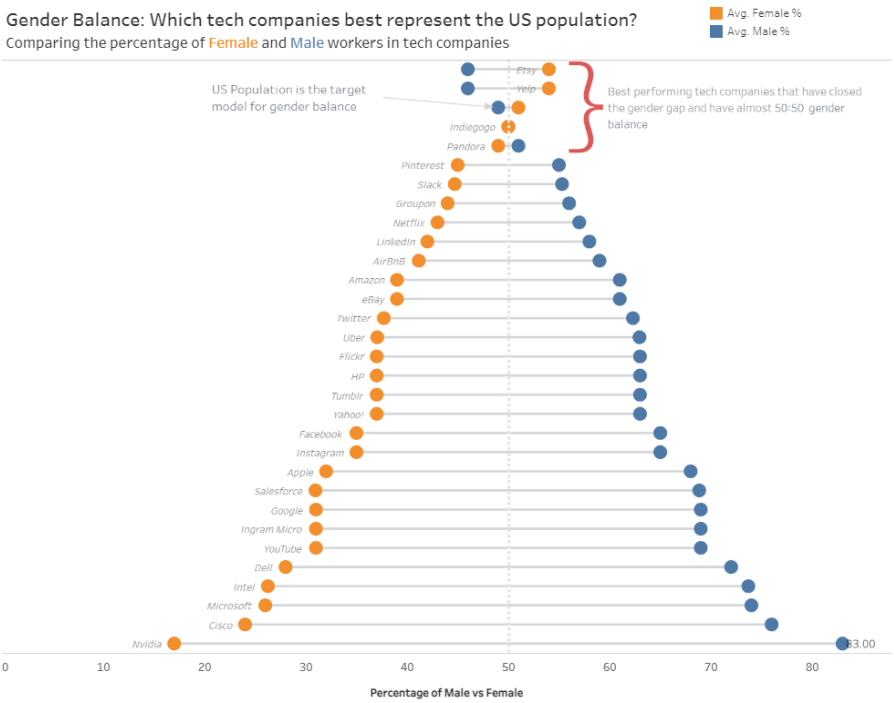
only labelled select lines, specifically U.S. population and companies which had better representation than the U.S. population benchmark. We also facilitated users exploring company performance across all graphs by using linking functionality offered in Tableau; thus, as users hover over Facebook, for example, either from our company list or in a graph, all other lines that represent Facebook are highlighted as well.



Overview of Tech Industry Line Graph

Gender Balance: Which tech companies best represent the US population?

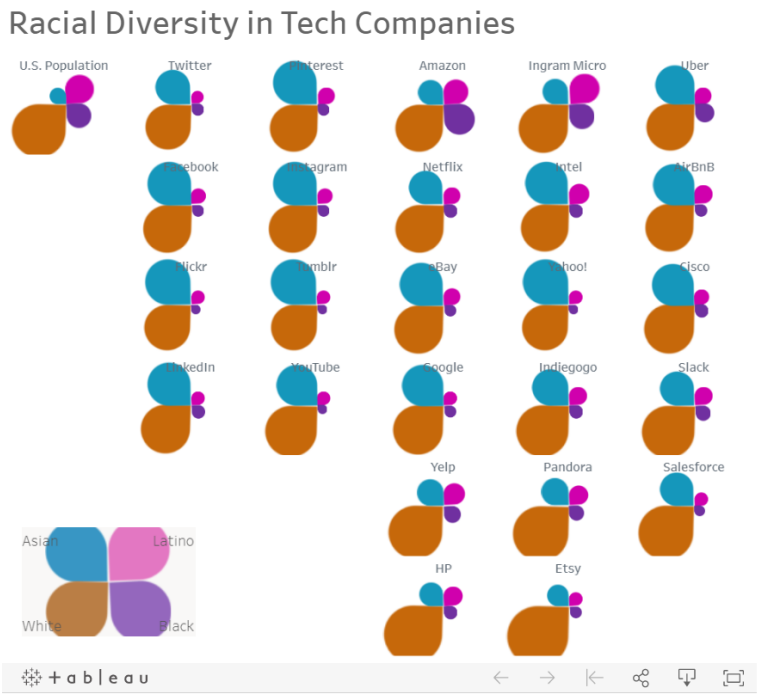
For this visualization, we wanted to dive deeper into data from 2018 (the most recent data that is available) and compare the percentages distribution of male to female employees at each company with the US population gender distribution. For the US population, there is almost a 50:50 gender ratio between male and female and so we used this as a benchmark comparison. Gender gap is defined as the difference between total percentage of male employees minus total percentage of female employees. From this visualization, the viewers can immediately see how most companies have a wide gender gap, with their being fewer percentage of female employees compared to male employees. Furthermore, this visualization makes it easy to see just how big this gender gap is for each company and the viewer can see the companies which have a smaller gender gap and can also see companies that have more female employees than men.



Gender Balance Visualization

Racial Diversity in Tech Companies

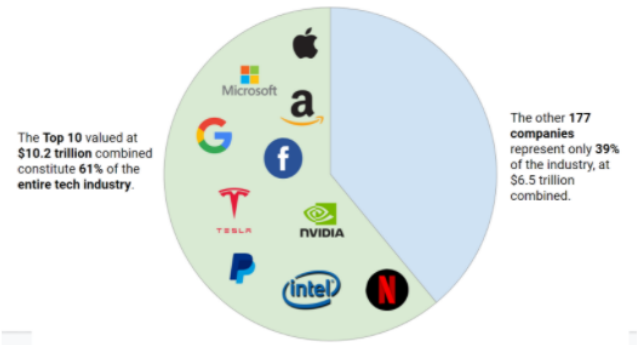
In this visualization we wanted to showcase the racial diversity amongst different companies and compare that with the racial distribution for the US population. We used different color flower petals to denote each race. The size of the petal changes based on the total percentage amount of employees at the company who identify as that specific race. Our goal with this visualization was for the user to immediately see how certain races are disproportionately represented at these tech companies and compare companies with each other and the US racial distribution. If the user wants more granular information, they can hover over the center of each flower and will see a tooltip with more specific details about the racial breakdown for each company.



Top 10 Tech Companies

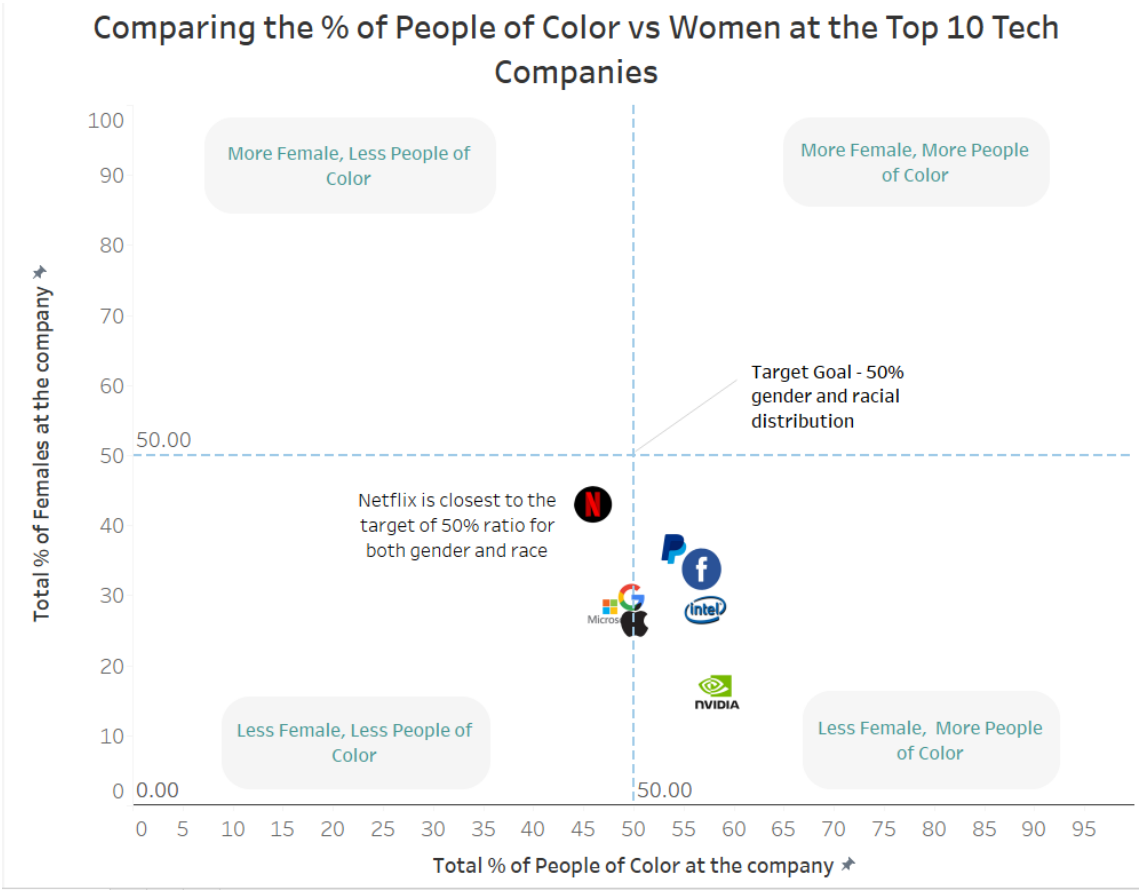
Market Landscape

This visualization serves as more of an infographic showing in simple terms the wide reach and power that the top 10 companies (by market cap) wield in the entire tech industry. While we learned in class that pie charts are not ideal in accurately representing proportions, our pie chart in this case can effectively relay the message that the top 10 companies take up well over half of the industry's total valuation, leaving a much smaller chunk in comparison that the other 177 tech companies contribute to. In this visualization, we also wanted to bring these top 10 companies to the forefront of attention as we are casting a light on them in the "Top 10 Tech Companies" and "Looking to the Future" sections of our website. Thus, we chose well-known logos to represent each company because they resonate quickly with most viewers. This type of iconography is used in the Olympics for instance, where instead of writing each country's names, news and reports use the country's flag. Leveraging this idea in our website makes tech company performance more like a competition and brings an element of fun to see which company ends up on top.



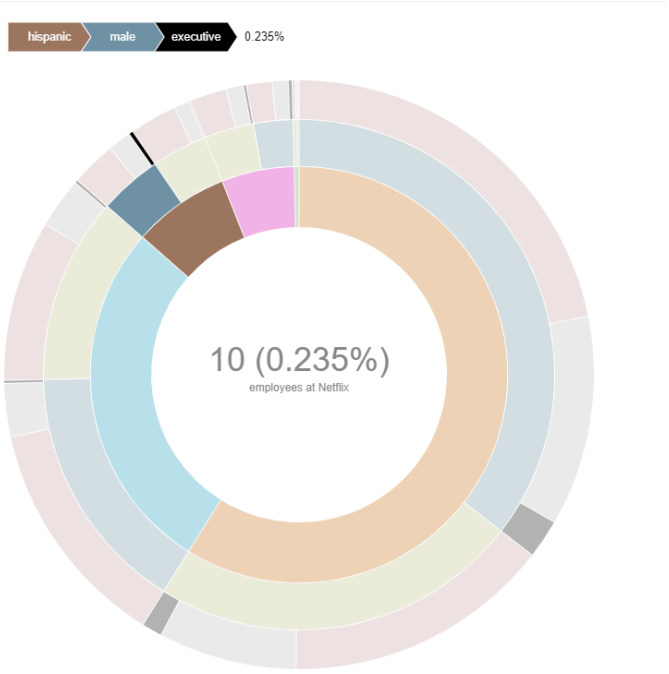
Comparing the % of People of Color vs Female at the Top 10 Tech Companies

For this visualization, we wanted to compare the top 10 tech companies with each other in terms of gender and racial diversity. Amazon and Tesla have not been included in this visualization as they did not release their EEO-1 reports to the public, so we had no standardized data to include for those two companies. Our goal with this visualization was to see if there were any companies that did better or worse than the others in terms of their racial and gender distribution. We divided the graph into 4 different quadrants based on the percentages of female employees and people of color at the company. We calculated the total percentages of women employees at each company and the total percentage of people of color employees and plotted them on the x and y axis accordingly. We determined 50/50 to be the "target" distribution for race and gender. This would mean a 50/50 goal for male vs female employees. For race, this would mean a goal of 50% White and 50% People of Color. However, we understand race is a lot more nuanced, especially considering the fact that within the category "people of color," some races are underrepresented compared to others. Therefore, we acknowledge that a 50/50 target for race is overly simplistic but for the scope of this project we thought this was a good starting point.



Diversity Across Different Roles at Netflix

From our scatterplot, we found that Netflix was closest to the 50:50 target we had set for racial and gender diversity. We wanted to dig deeper into the racial and gender distribution for Netflix across different roles such as executives, managers and professionals. We were inspired by the sunburst visualization Kerry Rodden showed us. We thought this would be an interactive way for the user to engage with the content. Our goal with this visualization was to show that although Netflix seems to be doing well in terms of racial and gender diversity, the higher one goes up the corporate ladder, the less diverse it becomes. As underrepresented individuals (specifically Hispanic, Black and Native American employees) progress through their careers, fewer and fewer make it through the ranks of manager and executive roles. Thus, retention is just as important to consider as hiring. For this visualization, we also included the whole numbers in addition to the percentages in the middle of the sun dial. We thought this would give the users a more concrete picture of how disproportionately certain groups are represented compared to others rather than solely relying on percentages.



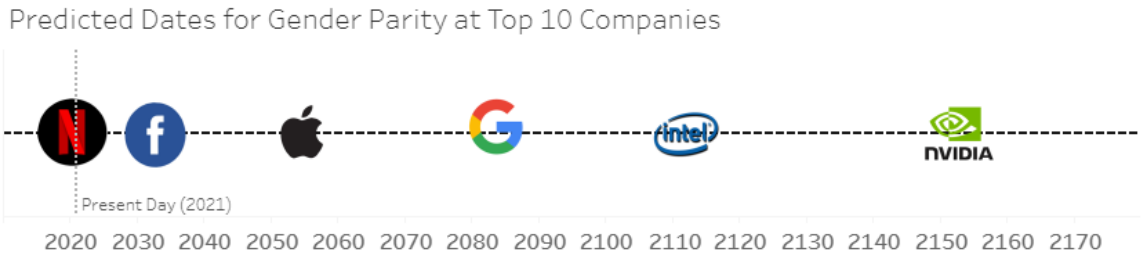
Looking to the Future

Predicted Dates for Gender Parity at Top 10 Companies

From class readings on *Storytelling with Data* (Knaflic), we learned that to deliver effective and memorable viz experiences, we must tell a coherent story backed by the data and generate user empathy and interest. On our website, we provided other visualizations to allow the user to explore gender and racial distribution in the tech industry; however, for our final interactive visualization, we wanted to bring the narrative closer to home in a way all viewers can relate. Thus, we came up with this idea of using dates based on forecasts of when certain

top 10 tech companies will reach gender parity. This will prompt the user to perform their own mental math and realize just how far away we are in this modern world from equal representation.

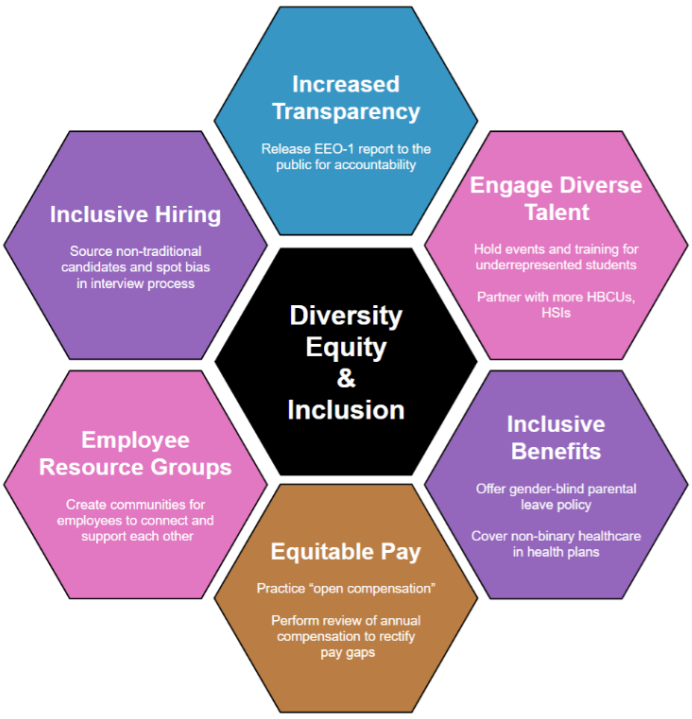
We decided to focus on gender and not racial parity because most companies are closest to gender parity while rather far from racial parity of underrepresented minorities. Also, racial representation is constantly evolving as minority makeup in the U.S. is only projected to increase over the years, with studies estimating that minorities will become the U.S. majority by 2042 (Source). Using gender parity estimates only, we then plotted all companies on one timeline because they are on the same journey together. Lastly, we kept consistent use of Top 10 company logos because we had already set the precedent for the user to be familiar with these icons and what they represent in previous sections of our website. Of the companies that were not included in the timeline, we explained why in the text--either they would never reach parity or they did not report their EEO-1 data.



Our Recommendations

Recommendations

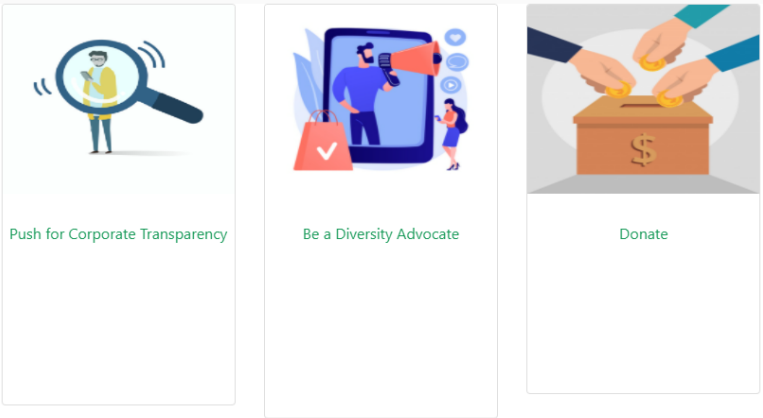
We initially researched DE&I initiatives and included them in a bullet point list. However, we realized that format as a big wall of text increased viewer mental load as they would have to slow down and read every word. As a result, we decided that this section would also be great for an infographic. For the infographic, we used hexagons that fit together in the middle around the words “Diversity, Inclusion, and & Equity.” It was important for these shapes to fit together like a puzzle because we wanted to convey that each piece is crucial to a holistic approach. All surrounding hexagons and text represented actions that have been and can be taken to improve diversity at companies. Because we wanted to convey our findings at a glance to the viewer, we summed up the action in 2--at most 3--words, with short phrases below in smaller text that provide more detail. We also applied class concepts of managing color. Because the DE&I hexagon is the focal point upon which all of the other hexagons are based, we used contrasting color from the rest to make it pop. Additionally, in keeping with previous visualization themes, we used colors from the “Racial Diversity Breakdown” visualization to color the other hexagons.



What You Can Do to Make a Change

What You Can Do Cards

A concept from a class exercise remained with us in which users are more likely to remember graph data if they had to perform an action like guessing it (Source). For this very last section of our website, we definitely wanted viewers to remember what we conveyed and take it upon themselves to take action. While we did not ask our viewers to plot data, we did want to engage them one last time through flipping over virtual cards to get more information about how to help diversity efforts. Because users must play an active (have desire to flip over card to receive more information) rather than passive (glance over an infographic), our hope is that our audience's minds were in an open state to receive and remember our parting words.



Approach

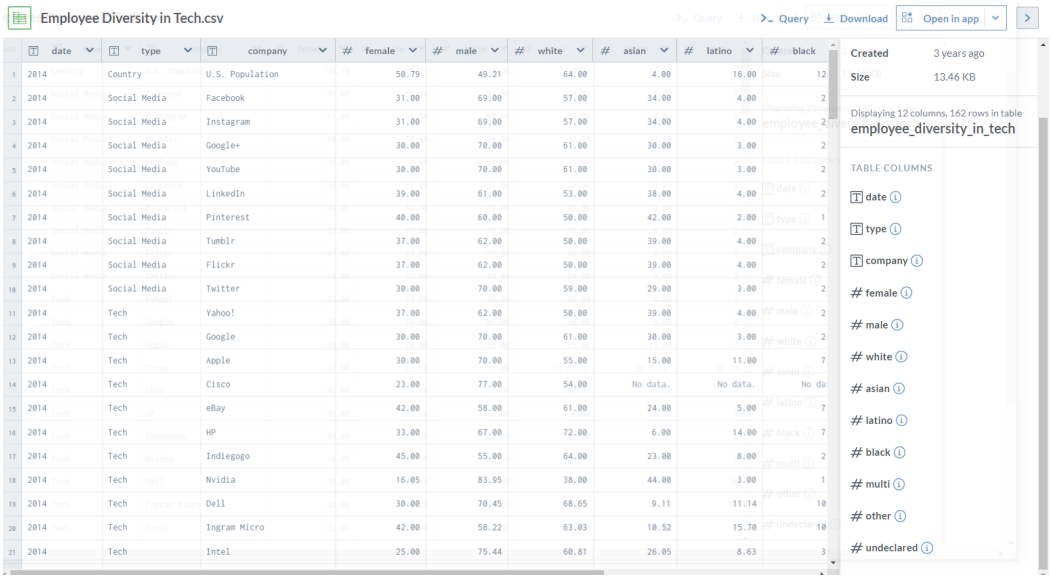
Data

We used two different datasets for our visualizations:

1. Diversity in Tech dataset from data.world
2. EEO-1 Reports for the Top 10 Tech Companies

Data.world: Diversity in Tech dataset

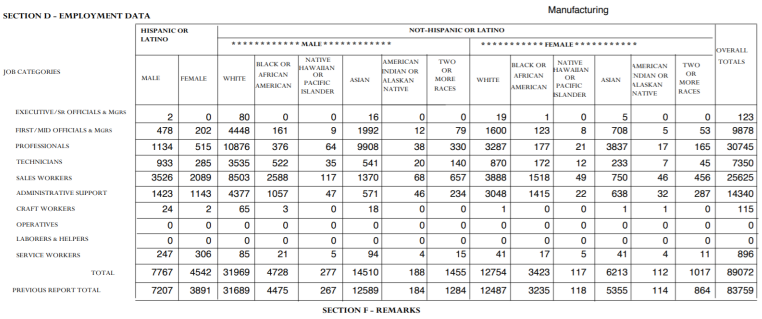
We found the most comprehensive dataset of tech companies on data.world, consisting of 12 columns and 162 rows. Curators of this dataset had manually visited each tech company's website and consolidated their self-reported data. This included percentage breakdowns of gender (Female, Male) and Race (White, Asian, Latino, Black, Multi, Other, Undeclared) for the years 2014, 2015, 2016, and 2018. While we realize the shortcomings of self-reported data, the reality is that many tech companies do not release their official employment statistics to the public (discussed in more detail in the EEO-1 data section below); thus, we had to rely on what we could find.



Screenshot view of dataset (Link Here)

EEO-1 Data

EEO-1 is a mandatory federal report that all companies in the U.S. with more than 50 employees are required to submit. The EEO-1 report contains information about the race & gender breakdown of employees across various roles in the company.



Manufacturing

SECTION D - EMPLOYMENT DATA

JOB CATEGORIES	HISPANIC OR LATINO		NOT-HISPANIC OR LATINO												OVERALL TOTALS
	********** MALE **********								********** FEMALE **********						
	MALE	FEMALE	WHITE	BLACK OR AFRICAN AMERICAN	NATIVE HAWAIIAN OR PACIFIC ISLANDER	ASIAN	AMERICAN INDIAN OR ALASKAN NATIVE	TWO OR MORE RACES	WHITE	BLACK OR AFRICAN AMERICAN	NATIVE HAWAIIAN OR PACIFIC ISLANDER	ASIAN	AMERICAN NDIAN OR ALASKAN NATIVE	TWO OR MORE RACES	
EXECUTIVE/SR OFFICIALS & MGRS	2	0	80	0	0	16	0	0	19	1	0	5	0	0	123
FIRST/MID OFFICIALS & MGRS	478	202	4448	161	9	1992	12	79	1600	123	8	708	5	53	9878
PROFESSIONALS	1134	515	10876	376	64	9908	38	330	3287	177	21	3837	17	165	30745
TECHNICIANS	933	285	3535	522	35	541	20	140	870	172	12	233	7	45	7350
SALES WORKERS	3526	2089	8503	2588	117	1370	68	657	3888	1518	49	750	46	456	25625
ADMINISTRATIVE SUPPORT	1423	1143	4377	1057	47	571	46	234	3048	1415	22	638	32	287	14340
CRAFT WORKERS	24	2	65	3	0	18	0	0	1	0	0	1	1	0	115
OPERATIVES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LABORERS & HELPERS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SERVICE WORKERS	247	306	85	21	5	94	4	15	41	17	5	41	4	11	896
TOTAL	7767	4542	31969	4728	277	14510	188	1455	12754	3423	117	6213	112	1017	89072
PREVIOUS REPORT TOTAL	7207	3891	31689	4475	267	12589	184	1284	12487	3235	118	5355	114	864	83759

SECTION F - REMARKS

Example of an EEO-1 Report

We had to manually search for the EEO-1 reports for the top 10 tech companies. While searching, we noticed that Amazon and Tesla are the only two companies in the top 10 that do not release these reports to the public. Intel on the other hand, is the only company in the top 10 who publicly releases information about the pay rates for different roles within the company (Source).

We compiled the EEO-1 reports for the 8 companies with the most market cap from 2018 (Google, Microsoft, Apple, Netflix, PayPal, Intel, Facebook and Nvidia) into one dataset. We used data from 2018 as that is the most recent EEO-1 data available. The reports for 2019 and 2020 were delayed due to the COVID-19 pandemic.

The EEO-1 report contains the race and gender breakdown for the various roles within the company such as: Executives, Managers, Professionals, Technicians, Sales Workers, Administrative Support, Craft Workers, Operatives, Laborers and Helpers & Service Workers. For our visualizations, we mainly considered the data about Executives, Managers and Professionals because we wanted to focus on the technical professionals and the corporate office.

Tools

For most of our visualizations, we used Tableau. For the sundial visualization, we used the D3 template on Observable that Kerry Rodden shared with us during class and modified the code according to our needs. We used Affinity Designer to modify the company logos and illustrations we found on freepik.com. For our website, we relied on a template provided by Github pages and used some open source code as listed in the Appendix.

User Testing and Results

Overview

We conducted a usability study on our project's website. The goal of this website is to convey a cohesive story with our visualizations about the diversity issues across the tech industry in the United States. We want to showcase how the creators of these technologies are

not completely representative of the actual users of the technologies, with specific groups being severely underrepresented.

For this usability study, we were interested in testing the following:

- If our visualizations and accompanying text helped to create a cohesive story/narrative.
- Whether our users were able to understand the main takeaways/goals of our visualizations.
- Whether our users were able to complete low level viz tasks that we had assigned.
- If the accompanying text provided better context and understanding of our visualizations.
- If our users enjoyed viewing our visualizations and interacting with them.

Test Setup

For our usability study, we used a combination of structured-interviews and a survey to get both qualitative and quantitative feedback about our designs. We created a script for the study which included all the tasks and questions we wanted to ask our participants (Link in Appendix). For each study, one team member facilitated the study and one took notes.

Study-set up:

- Introductions:
  - Brief introduction about the study and context for the designs. Participants signed the consent form.

- Tasks Execution:
  - Participants were instructed to share their screen on Zoom. Research facilitator walked the participants through tasks. For each task, the participants were encouraged to verbalize their thought process while viewing our visualizations. The note-taker took extensive notes about the participant's feedback and behavior while completing the tasks.

- Wrap Up:
  - Towards the end of the study, participants filled out a survey to gather more quantitative information about their experience viewing our website

Result Analysis

Visualization	Task Completion	Qualitative Feedback
Overview of Tech Industry Representation	All 3 of the participants were able to accurately describe the high level task and identify the main takeaway from the graphs. However, only ⅔ were able to accurately complete the low	• All participants were confused about the red and yellow lines and it took them a few minutes to understand what those lines meant.

	level viz task and identify the extrema.	• There were many grey lines which made it difficult to read and understand the visualization. • Noticed a discrepancy where the 4 graphs did not all have the same companies represented.
Gender Balance: Which tech companies best represent the U.S. population?	Most of the participants struggled to complete the high level task and identify the key takeaways for the visualizations. ⅔ participants were able to complete the low level task, but it took them several minutes to figure it out.	• Unclear what the term “gender gap” was representing • Not sure whether gender gap was calculated by male % - female % or vice versa • Confused about the US population data point and its significance • Not intuitive design. • Title unclear. • Supporting text is unclear and too long.
Racial Diversity in Tech Companies	All of our participants were able to accurately complete the high level and low level viz tasks.	• Color on the bar charts on the tooltip did not match the petals • Unclear why there are 3 categorizations • Unclear what “diversityinc50” indicates • Accompanying text is too long
Market Landscape	⅔ participants were able to complete the high level task but were uncertain about the relation of this visualization to the rest of the study.	• Microsoft and Amazon look the biggest but Apple has more market cap. • Unsure what the label “EEO-1” means on Tesla and Amazon. • Believes this visualization detracts from the narrative in being irrelevant.
Comparing the % of People of Color vs Women at the Top 10 Tech Companies	⅔ of the participants were able to complete the high and low level tasks.	• Different quadrants are confusing and unsure what they represent. • Accompanying text is too long and not engaging. • Axes and labels are not clear and are too small.

		• Scales are confusing - y axis hacking.
Diversity Across Different Roles at Netflix	All 3 of our participants were able to complete both the high level and low level tasks.	• Initially did not realize they had to hover over the visualization to see the data. • Unclear what the terms “professionals”, “managers” and “executives” mean. • Noticed color inconsistencies in some areas.
Predicted Dates for Gender Parity at Top 10 Companies	All of our participants were able to complete the high level viz task and understood the main takeaway for this visualization. However, it took them all a while to understand the visualization as it was not intuitive.	• Title is not clear • Too much of a gap between the points and the years. • Accompanying text is not interesting. Could ask the user a question to make it more engaging.

Overall Impression and Usability of Visualization

Overall, we learned the participants enjoyed our visualizations and found them interesting. All were able to articulate the main objective of our website and narrative as we had intended. However, we found that there are a few recurring changes we need to make across all sections and visualizations in our website as follows:

- Reduce the amount of text describing the visualization. Make the sentences more concise and engaging in generating empathy with the viewer. Highlight keywords.
- Clearly label the axes of the visualizations and make the text bigger.
- Use more intuitive titles for the visualizations.
- Appropriate scales for the axes - make sure they all start at 0.
- Label quadrants wherever needed.
- Annotate specific points more clearly.
- Use consistent colors throughout the visualizations.
- Clearer explanation of technical terms (i.e EEO-1, definition of “manager”, “executive”, “professional” etc)

Link to Code and Visualization

Our visualizations exist on this website: Diversity in Tech site.

Code is on github: GitHub Project Page.

Team Contributions Breakdown

Low Effort (0-24)	Medium Effort (25-49)	High Effort (50-100)

Project Phase	Task	Priyanka	Merry	Vivian
Exploratory Data Analysis	Perform individual EDA assignments based on the Data section	33%	33%	33%
Visualizations	Overview of Tech Industry Representation	5%	90%	5%
	Gender Balance: Which tech companies best represent the U.S. population?	5%	5%	90%
	Racial Diversity in Tech Companies	5%	5%	90%
	Market Landscape	5%	90%	5%
	Comparing the % of People of Color vs Female at the Top 10 Tech Companies	90%	5%	5%
	Diversity Across Different Roles at Netflix	90%	5%	5%
	Projected Dates for Gender Parity at Top 10 Companies	5%	90%	5%
	Recommendations	35%	55%	10%
	What You Can Do Card	15%	10%	75%
Usability Test	Write Script	80%	10%	10%
	Conduct Sessions	33%	33%	33%
	Synthesize Findings	33%	33%	33%
	Final Write Up	75%	12%	12%
Website	Narrative + Text	40%	40%	20%
	Design	25%	25%	50%

	Coding (includes transferring visualizations + text to site and refining presentation and structure of site)	10%	10%	80%
Project Management Activities	Attend weekly meeting to discuss project implementation and receive feedback on visualizations	33%	33%	33%
Final Write Up	Contribute to Final Write Up per agreed upon structure	25%	50%	25%

Appendix

- Direct links to team visualizations (stored on Tableau Public)
    - Overview of Tech Industry Representation
    - Gender Balance: Which tech companies best represent the U.S. population?
    - Racial Diversity in Tech Companies
    - Comparing the % of People of Color vs Women at the Top 10 Tech Companies
    - Diversity Across Different Roles at Netflix
    - Predicted Dates for Gender Parity at Top 10 Companies
    - Market Landscape, Recommendations, What You Can Do Cards
        - Code is included in team Github repository
- Image under first section (link here)
- Some images for the site from Freepik
- Used shapes from Slidesgo
- Flip cards open source code
- Visualization Templates
    - Sunburst
    - Gender Balance