Minorities in the Movies

Kate Carey, Shelley Huang, Courtney Lett, Hannah Rosen

PROCESS BOOK

Visualization in HCI
May 3, 2017

Table of Contents

Project Proposal 1
Motivation
Objectives
Data
Features

Data Collection 3
Movies to investigate
Methods
Cleaning

Narrative 5

Iterations and Reasoning 6
Sketches of Visualizations
Iterations
Hi-fi Mockups
Final Visuals Descriptions

Building the Visualization 15
Sharing the Code
Progress

Final Product 23

Motivation

We believe there is inherent value in representing a diverse set of backgrounds in media and entertainment. As members of minority groups, we want to visualize for filmmakers and studios that more diversity in films is a valuable goal

.

Objectives

Visualize how minority representation in American film correlates to various measures of success of the film.

Data

Gathered data on film production staff from IMDb. Gathered character word count data directly from movie scripts.

Features

On camera representation of minorities

- Indicate women, ethnic groups, sexuality
- On camera
- Role type
- Movie genre
- Year released

Off camera representation of minorities (optional)

- Indicate women, ethnic groups, sexuality
- Movie genre
- Year released

Film success

- Critical acclaim
- Box office stats
- Total gross
- Public opinion of film (optional)

Societal outcomes (optional)

- Youth self confidence measures
- Public opinion of minority group

Movies to Investigate

We chose 10 movies released in 2016 for their box office gross, critical acclaim, and variety of genres represented. The movies we investigated further are:

- Arrival
- Batman vs Superman
- Captain America: Civil War
- Hidden Figures
- Jackie
- La La Land
- Lion
- Manchester By The Sea
- Moonlight
- Rogue One

Movie	Name	Gender	Race	LGBTQ?
Moonlight	Adele Romanski	Female	White	Straight
Moonlight	Jeremy Kleiner	Male	White	Straight
Moonlight	Dede Gardner	Female	White	Straight
La La Land	Jordan Horowitz	Male	White	Straight
La La Land	Fred Berger	Male	White	
La La Land	Marc Platt	Male	White	
La La Land	Gary Gilbert	Male	White	
Lion	Iain Canning	Male	White	
Lion	Angie Fielder	Female	White	
Lion	Emile Sherman	Male	White	
Hidden Figures	Pharrell Williams	Male	Black	Straight
Hidden Figures	Theodore Melfi	Male	White	Straight
Hidden Figures	Donna Gigliotti	Female	White	
Hidden Figures	Peter Chernin	Male	White	Straight
Hidden Figures	Jenno Topping	Female	White	Straight
Manchester by the Sea	Matt Damon	Male	White	Straight
Manchester by the Sea	Kimberly Steward	Female	Black	
Manchester by the Sea	Kevin J. Walsh	Male	White	
Manchester by the Sea	Chris Moore	Male	White	Straight

Methods

In order to gather the custom data we needed to tell our story, we combed through the scripts of the 10 movies we chose and counted up the number of words each character spoke. We made note of the character's gender, race, and sexual orientation, if available.

We also gathered film maker data for each of the movies from IMDb and other sources of information about the film. We looked specifically at the gender, race, and sexual orientation of the films producers, directors, and writers,

Cleaning

Once all the script data was gathered, we had to clean it to be sure we had the most useful set possible. We eliminated unnamed characters to narrow the data set. We also unified categories and capitalization across the data, to make it easier to prototype and code visualizations.

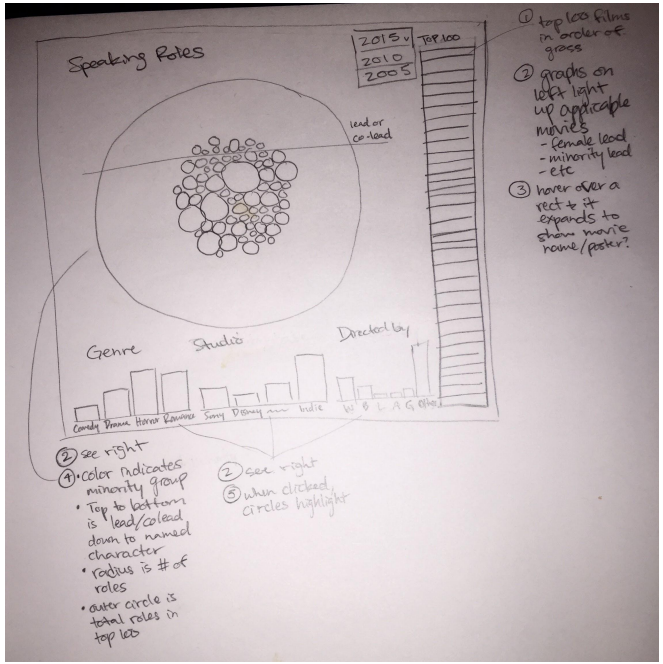
Structuring the Story

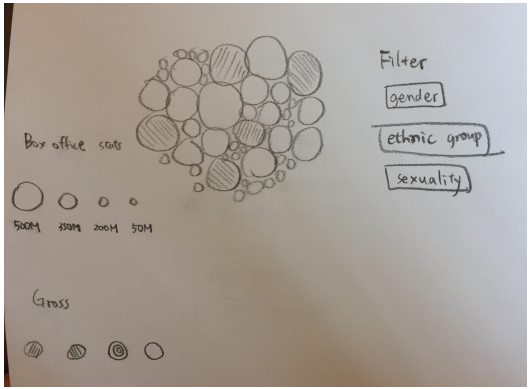
Film executives, stop patting yourselves on the back

The state of diversity in film and the business case for more diversity

1. 2016 seemed like a great year for diversity in film
    a. Upward trend of women making films
    b. Celebrated films like Moonlight, Hidden Figures, Arrivals
2. Pipeline film production > film cast > film viewership
    a. Is not representative
3. How much do people care about seeing themselves represented
    a. Higher median box office gross and ROI for film with casts that have representative ratio of minorities (compared to US pop)
    b. Token casting isn't enough
    c. General public outcry for more diverse films
4. Look deeper into 8 movies from 2016
    a. Words spoken as representation of women vs men - more men
    b. Words spoken as representation of races - few minorities
    c. Sexual orientation?
    d. What does it take for a female character to have a name? - fewer female peripheral characters with names
    e. Call out specific differences we find interesting
        i. Hidden Figures white male boss vs. black female lead

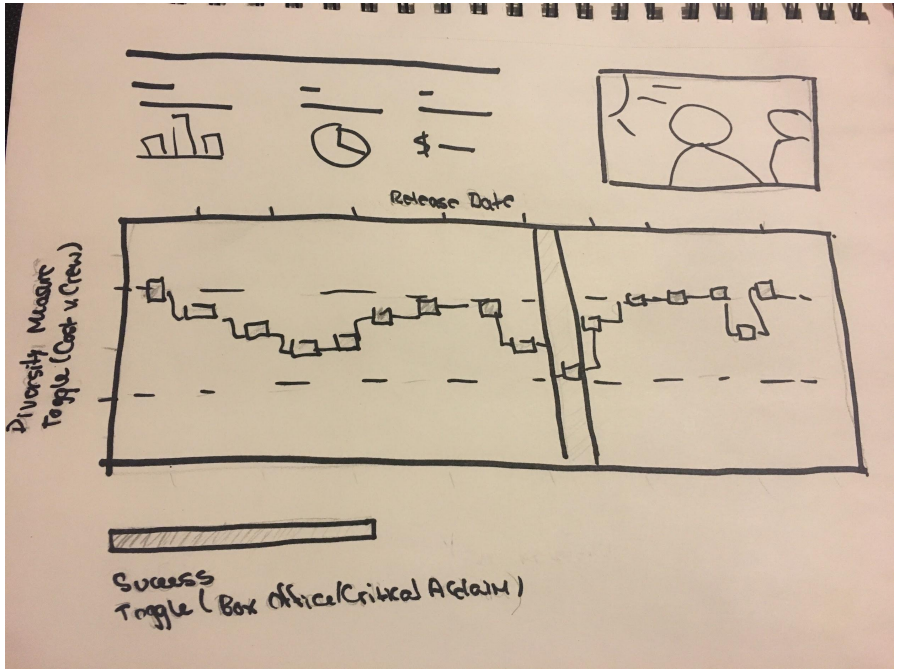
Initial Sketches





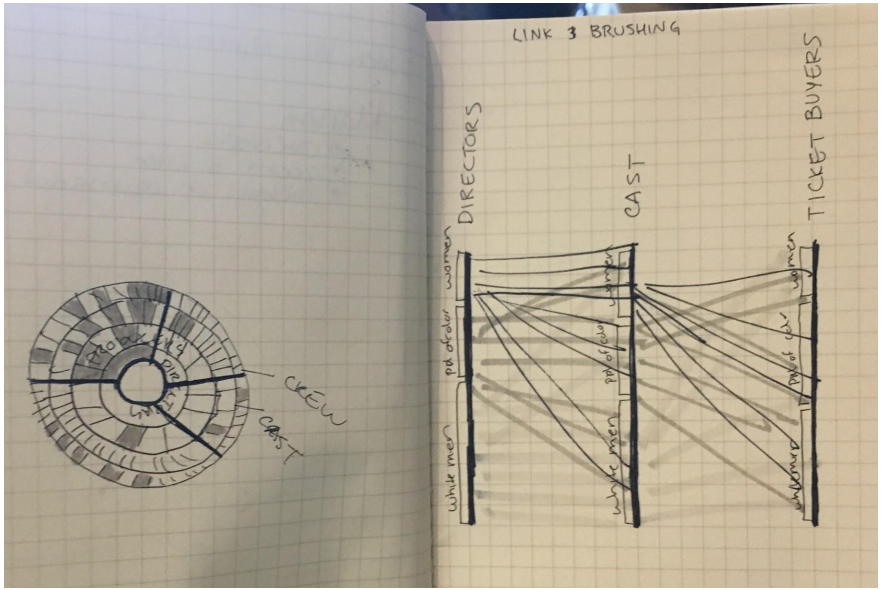
This interactive visualization shows the proportion of speaking roles held by different minority groups. The bubble graph is connected to the bar graphs on the bottom and the full list of the top 100 films of 2015 on the right. The top 100 list can also be explored by hovering and clicking.

ITERATIONS & REASONING



This visualization depicts movies organized by release date, a diversity measure, and success. The two axes (diversity and success) can toggle between two measures: cast/crew and gross/critical acclaim. When a film is highlighted, the more statistics about the film can be seen at the top. This visualization can also show the statistics for one movie, using the metrics of time on screen speaking for minority and women actors.

ITERATIONS & REASONING



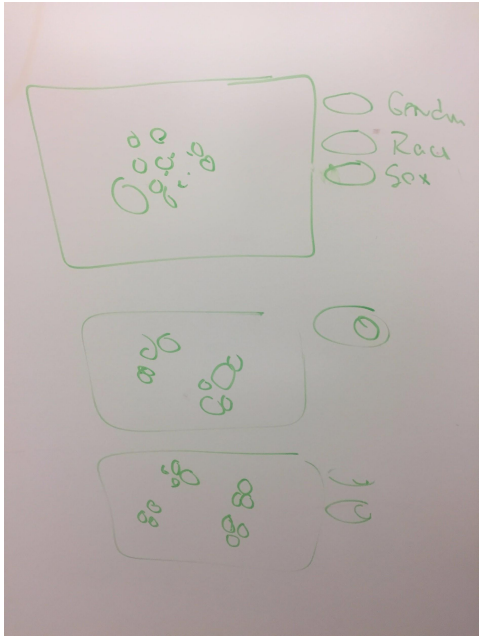
On the left, is a sunburst diagram with directors at the centre, radiating out to producers, cast and crew data. The different demographic factors we are interesting in looking at, such as gender and race, will be colour coded into the sections to see how the more powerful people at the centre impact the people further out in the diagram.

On the right is a parallel coordinates plot. The furthest left axis represents the demographic information about the director, the middle represents the same about the cast, and the left represents the same about the people who purchased tickets to see the movie in theatres. Through linking and brushing you could highlight how each demographic category on each axis impacts the others.

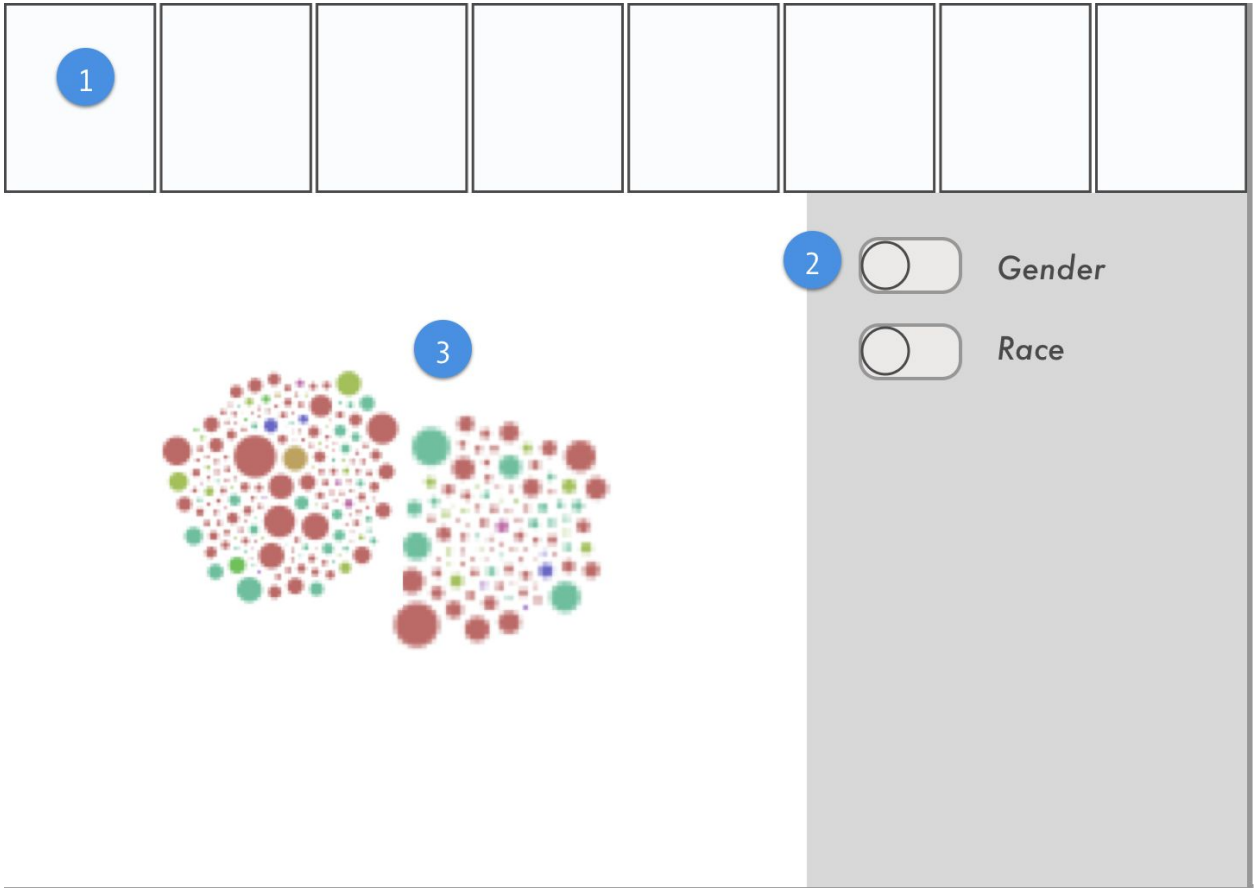
Iterations

So after we gathered our data and discussed about the initial graphs, we collect things we like and would love to work on:

1. Showing all the character data on one graph and allowing filtering
2. Bubble graphs to show size of bubble = words spoken
3. Filtering by film
4. Filtering by gender and race - simultaneously or one at a time



Hi-Fi mockups



Default view

1. Movie posters for each of the 8 movies.

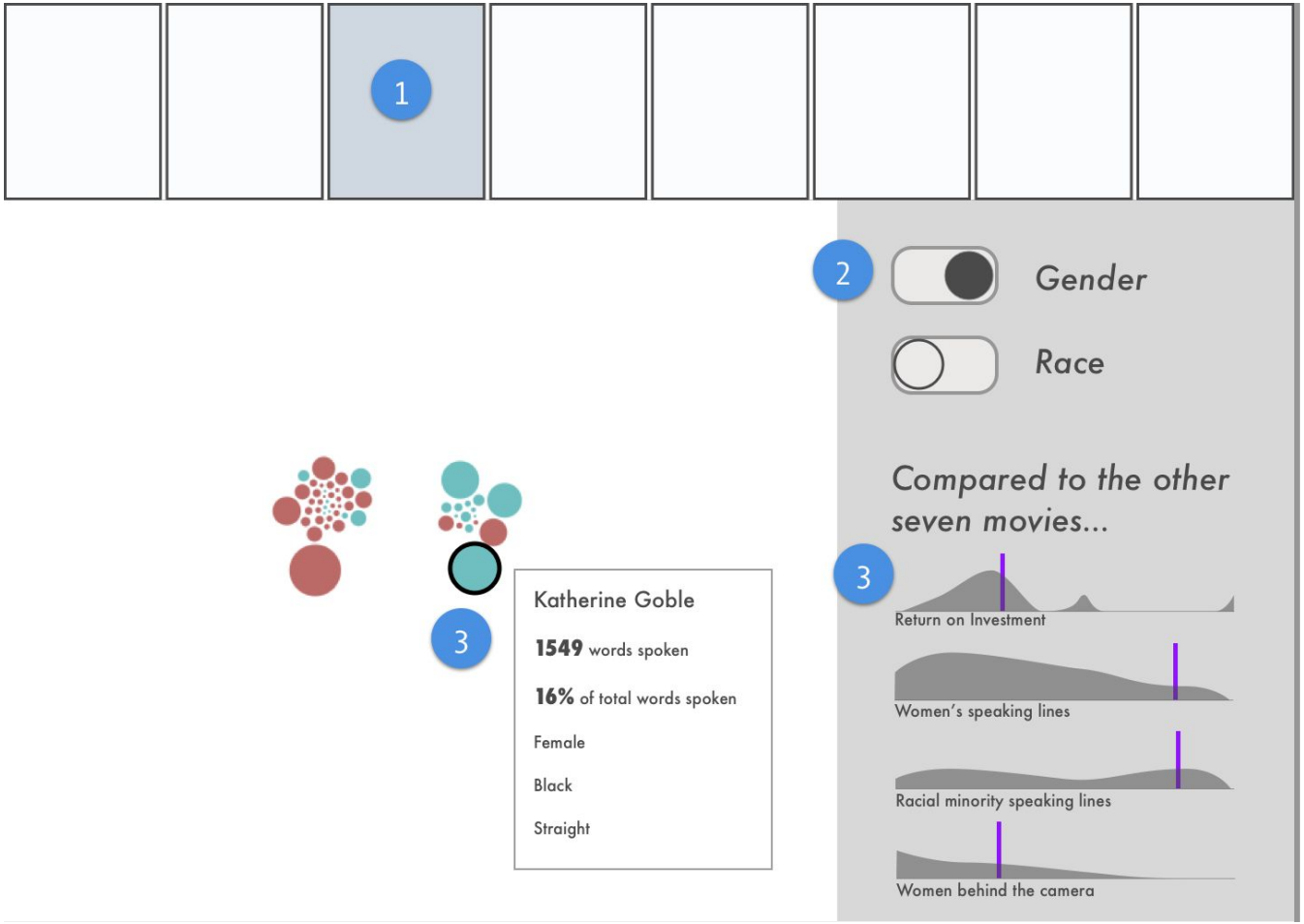
2. Switches to group bubbles differently - default both off

On click, groups bubbles by feature

Can have both switches on or off

3. Bubbles represent each character:
- Size = total words spoken
- Color = race

ITERATIONS & REASONING



Default view

1. Hidden Figures clicked. One movie selected at a time

On click, filters bubbles to only characters from that film.

On click, individual movie data appears below switches.

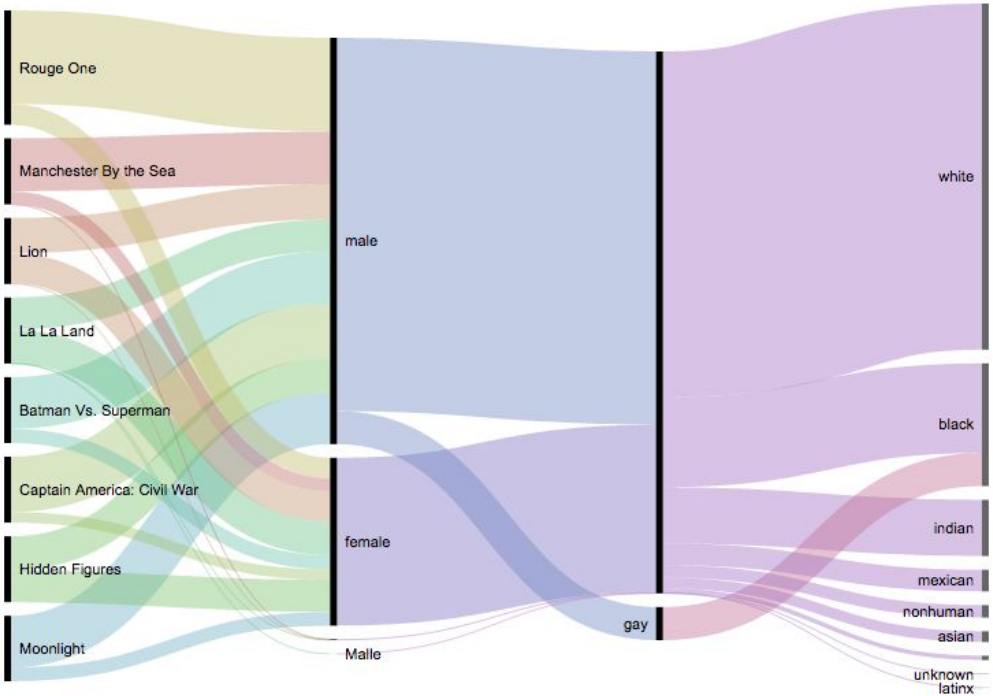
2. Switch on for Gender grouping, shows bubbles grouped by gender

3. Histograms comparing distribution of different factors across all 8 movies.

Shows where currently selected movie falls.

4. On hover over individual button, show character name, total word count, gender, race, sexual orientation

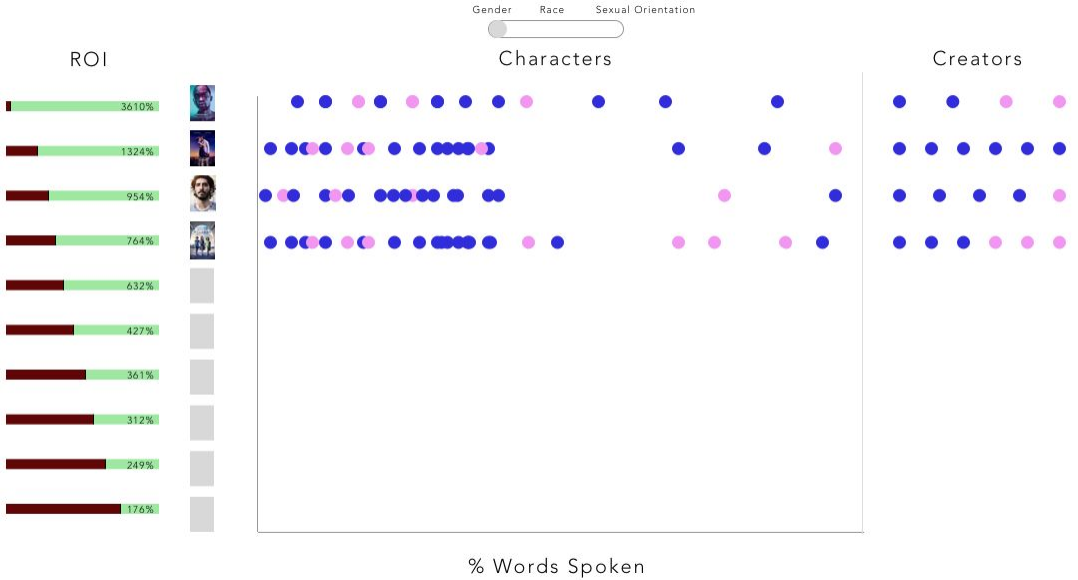
ITERATIONS & REASONING



We tried bubble chart and it ended up not helping with details we'd love to deliver. Therefore we also tried alluvial diagram, using our real data on RAW. We did user testing with our peers and got the following feedback:

1. Hard to connect percentages to particular character
2. Hard to know what groups each character belongs to

ITERATIONS & REASONING



Another alternative we came up with is this visualization. Films are organized by ROI, then characters and creators are shown on that line.

We also showed this to peers and got the following feedback:

1. Doesn't understand ROI graphs - make them 36x instead of percentages?
2. Creators looks like it's part of the character data - use different visual encoding?
3. Hard to draw conclusions from data on your own - add annotations?

Final Visuals

ROI
Return on Investment (ROI) represents how much money the movie made compared to what it cost to make. The higher the better! Moonlight, the only film on this list with a black director and all black cast, had a 3610% ROI. Compare that to Batman vs Superman, which had only a 249% ROI. In general, films with more diverse creators and casts garner higher ROI.

Character's Percent of Words Spoken
Each circle represents a named character and the percent of words they spoke from the total words spoken in the film. Despite the diversity of the 2016 films we chose, the named characters with more words spoken are predominantly white and male. Even in a groundbreaking film like Hidden Figures, the character who spoke the most words is not the main character, the black, female scientist Katherine Goble, but her white, male boss, Al Harrison.

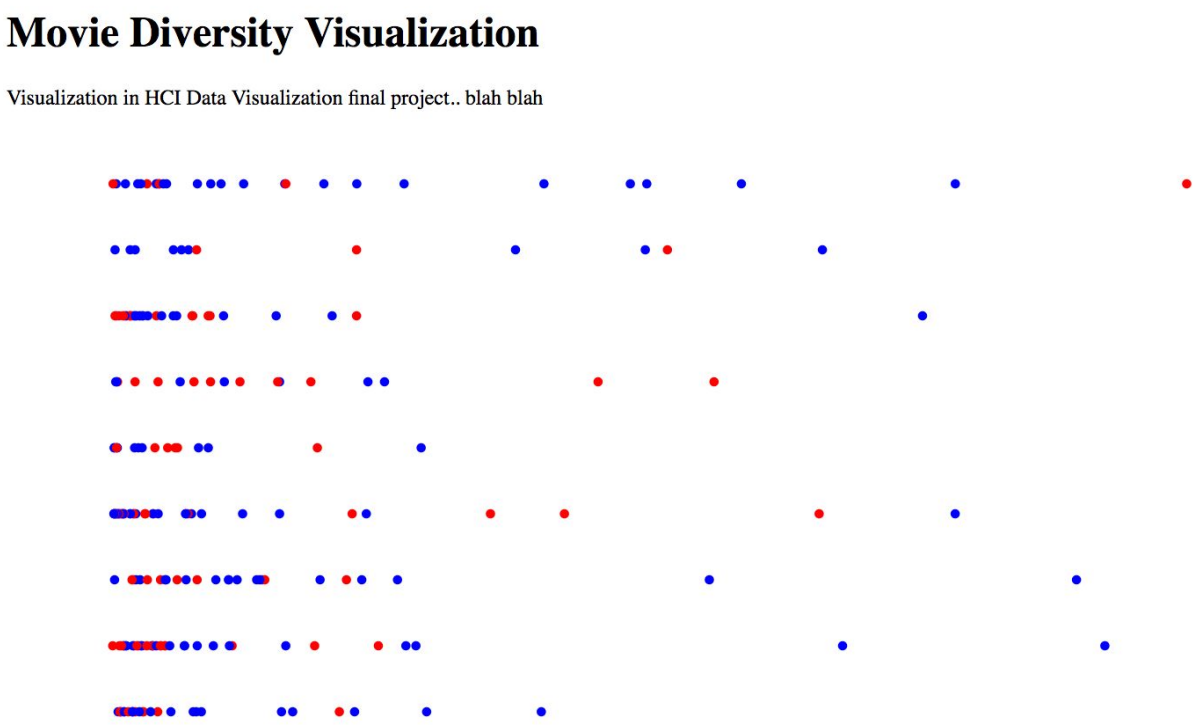
Filmmakers
Each square represents a producer, writer, or director of a film. As you can see, filmmakers are overwhelmingly white and male. However, of the top 5 highest ROI films, 4 out of the 5 have at least one woman behind the camera, and 3 have at least one person of color among the film creators.

Sharing the Code

Once we narrowed in on the structure of our final visualization, Courtney built the scaffolding for the visualization in HTML and d3, and split up the work. We each focused on different components of the visualization, committing our changes to the GitHub repository we shared.

This allowed us to increase our efficiency as a team and keep from weighing down any one team member.

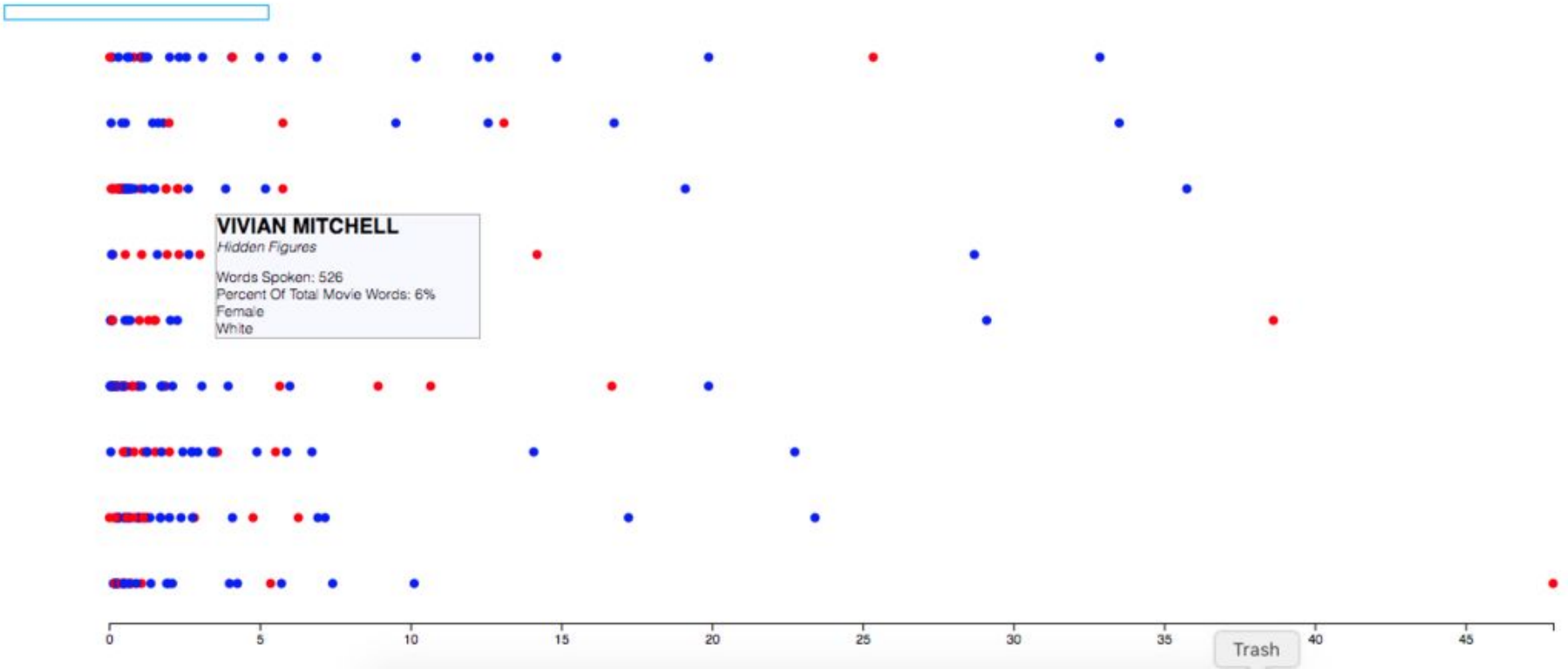
Iterations


Movie Diversity Visualization
Visualization in HCI Data Visualization final project.. blah blah

CODING

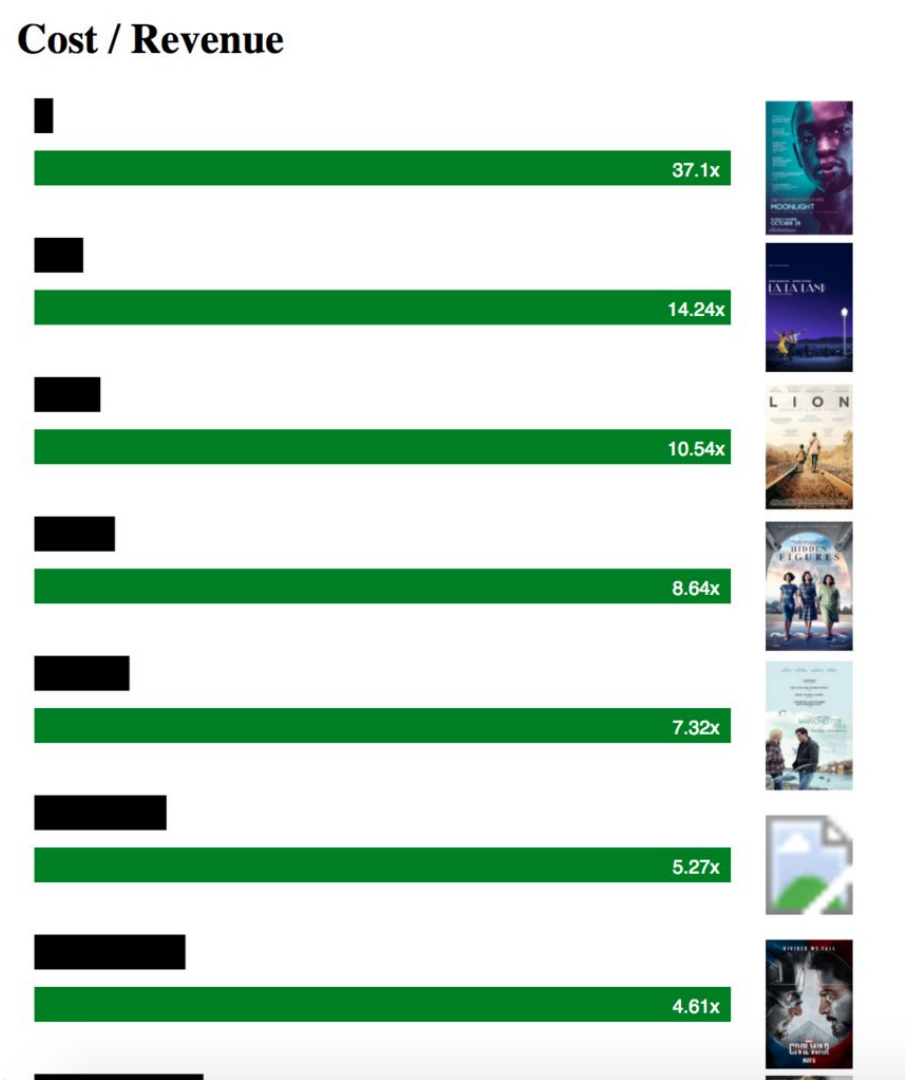
Movie Diversity Visualization

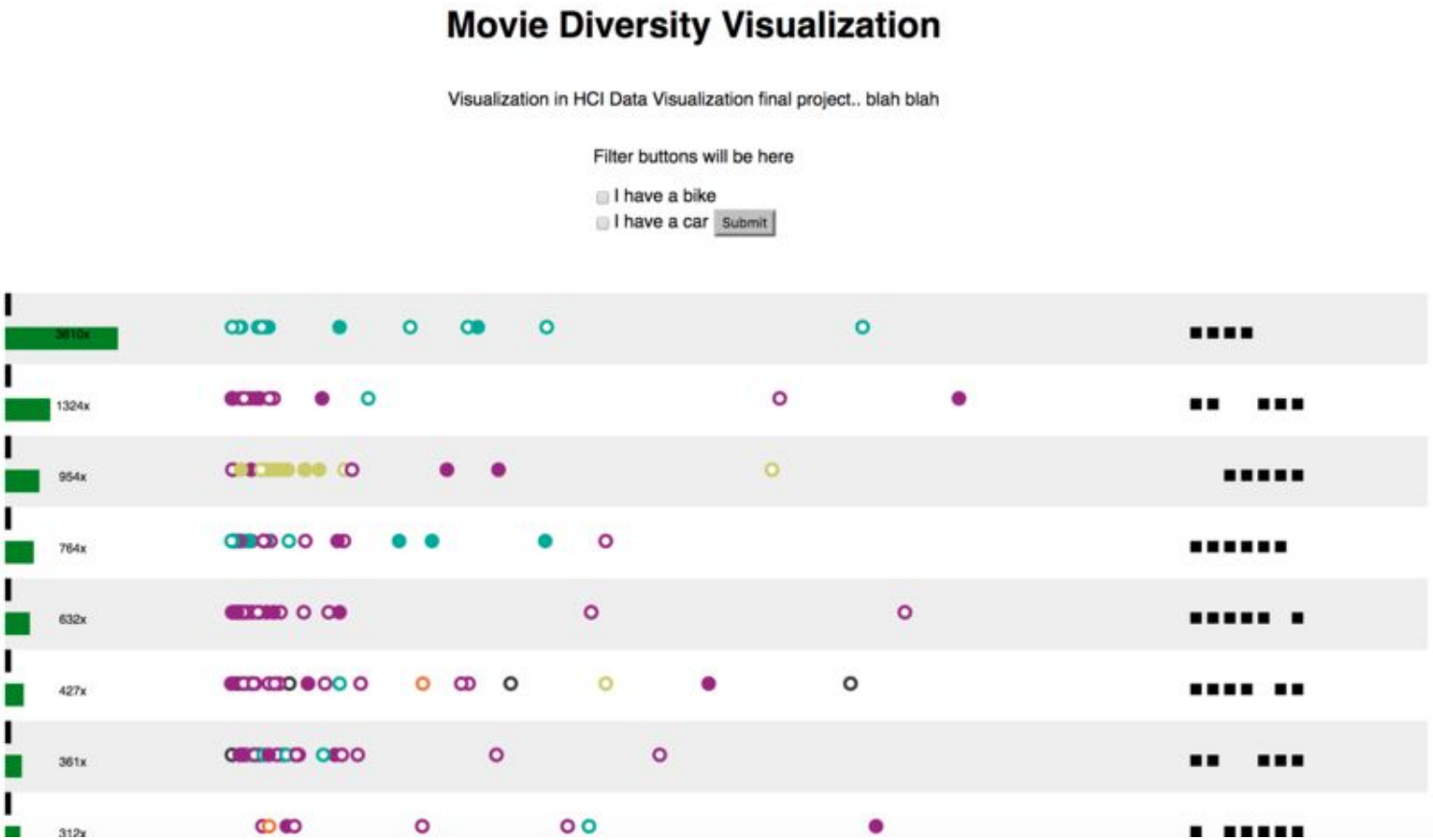
Visualization in HCI Data Visualization final project.. blah blah



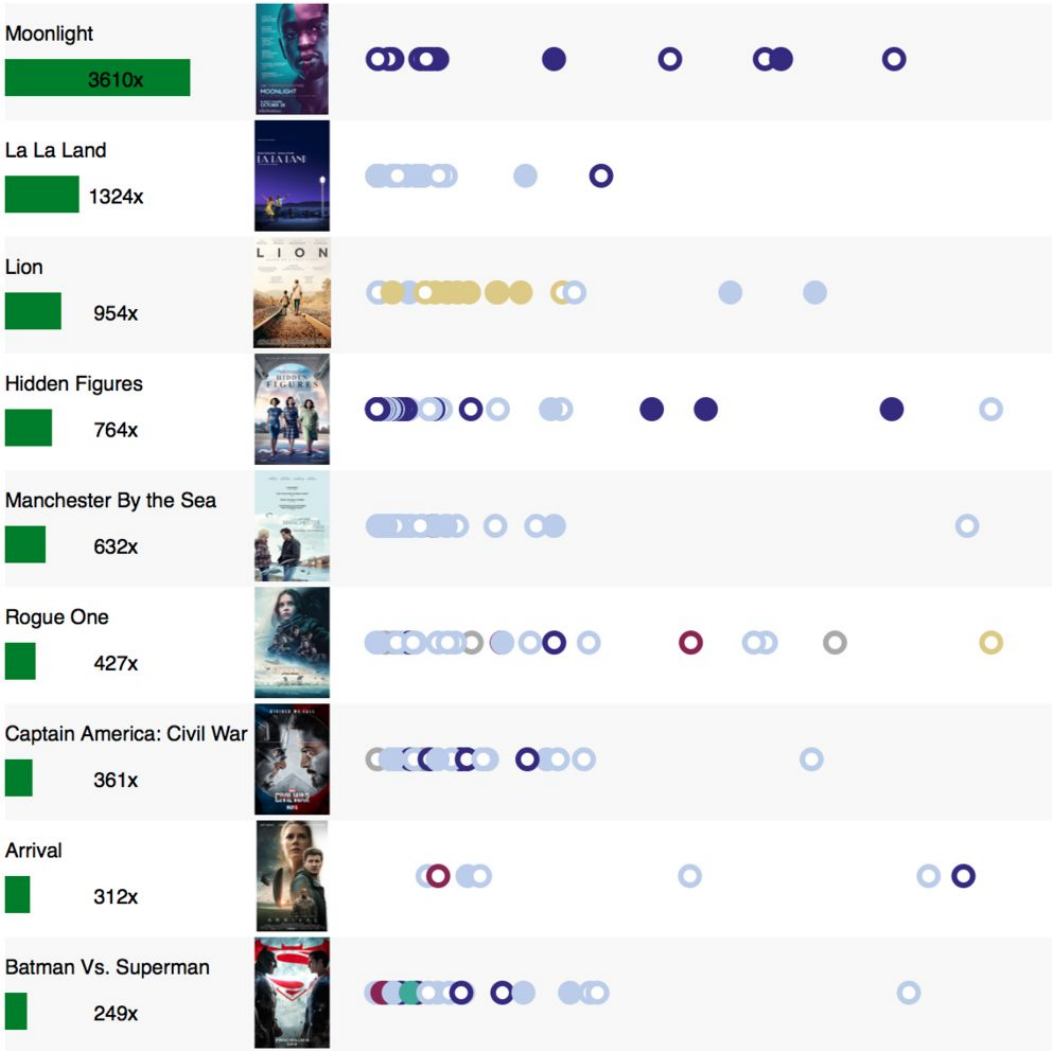
CODING

CODING


Cost / Revenue
37.1x
14.24x
10.54x
8.64x
7.32x
5.27x
4.61x
MOONLIGHT
LA LA LAND
LION
HIDDEN FIGURES

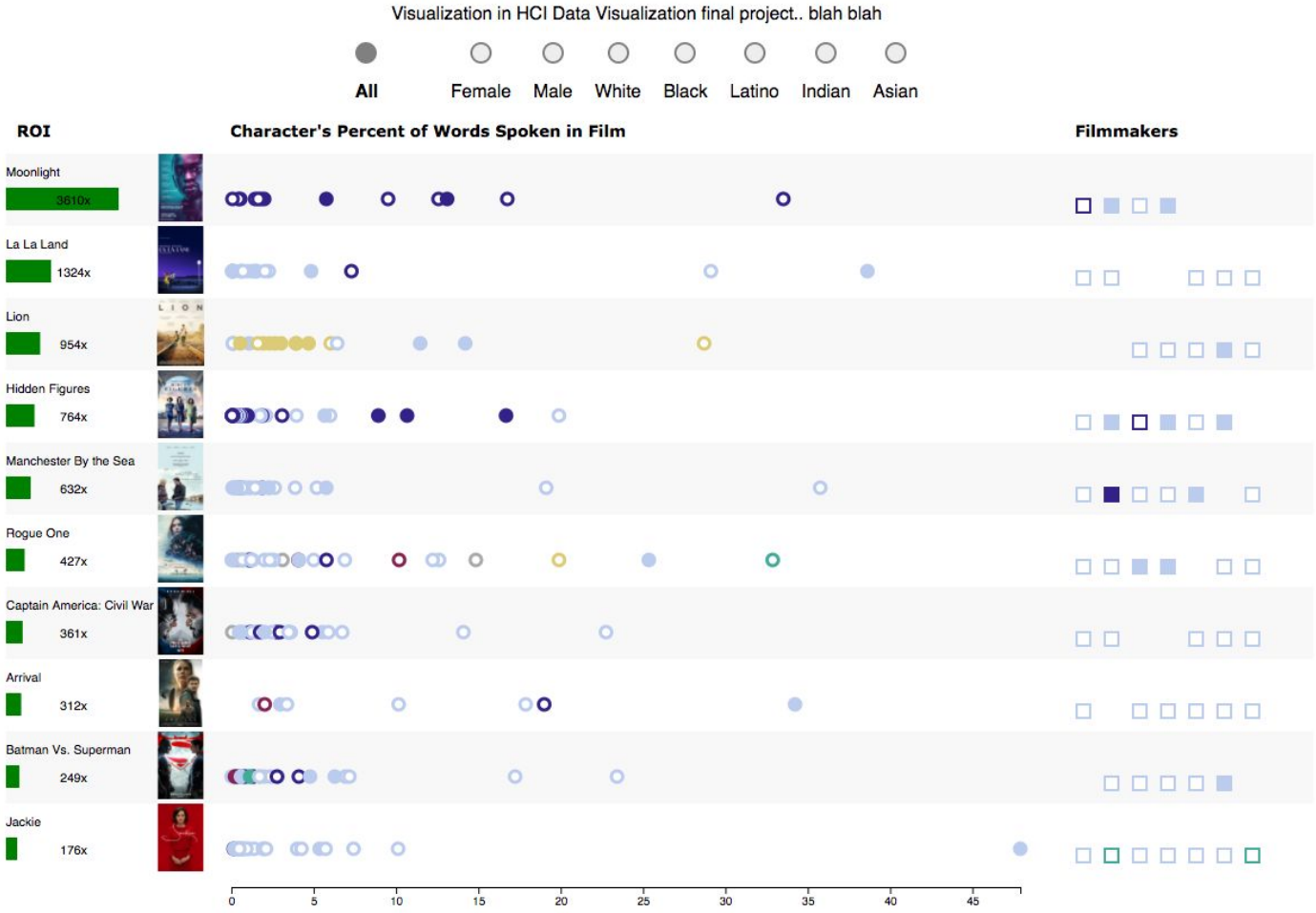

Movie Diversity Visualization
Visualization in HCI Data Visualization final project.. blah blah
Filter buttons will be here
I have a bike
I have a car
Submit
1324x
954x
764x
632x
427x
361x

CODING

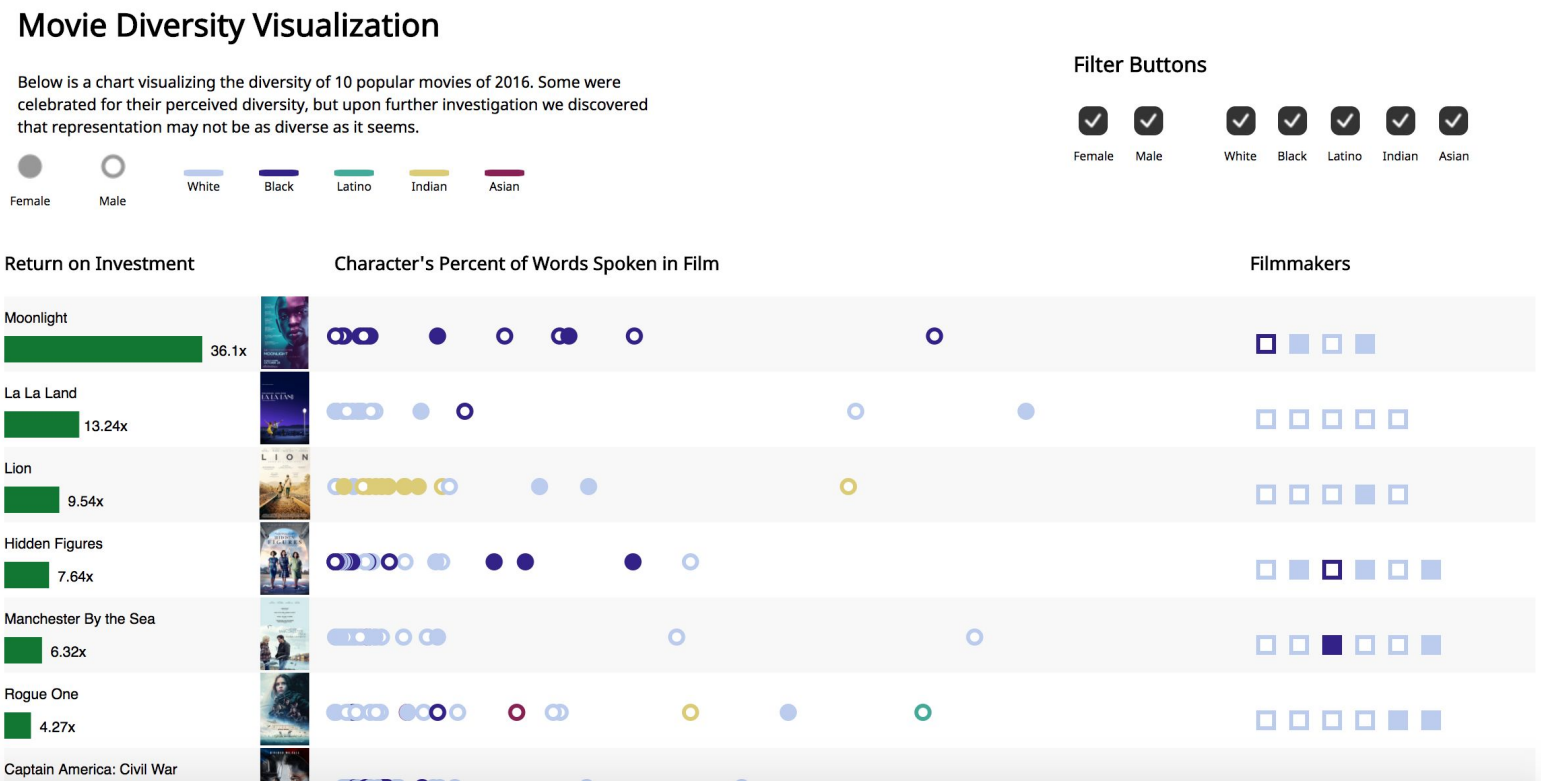

Moonlight
3610x
La La Land
1324x
Lion
954x
Hidden Figures
764x
Manchester By the Sea
632x
Rogue One
427x
Captain America: Civil War
361x
Arrival
312x
Batman Vs. Superman
249x

CODING

Movie Diversity Visualization



Final Product



Thank you!