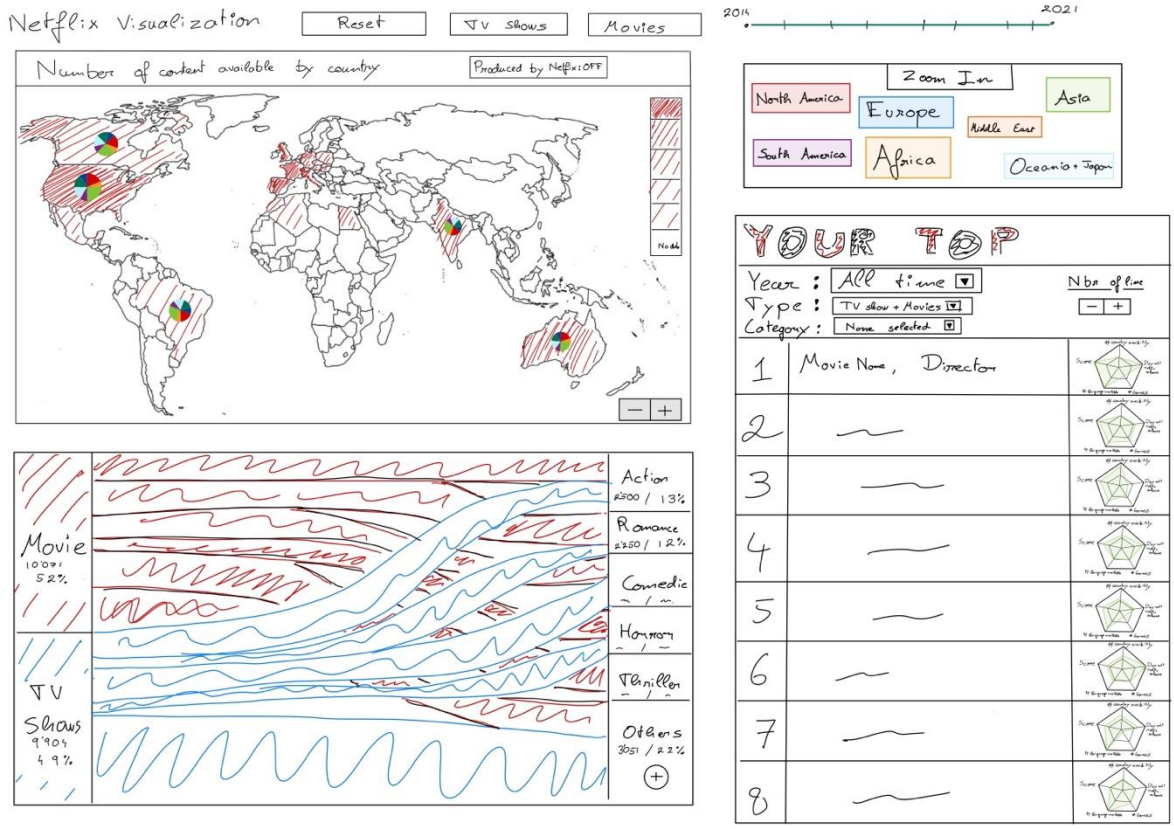
Checkpoint II: Visualization Sketch

Group: 30
Date: <2024/09/24>

Modifications of CPI after improved version : we decided to drop 2 other columns called 'Director' and 'Writer' since they had a significant impact on our data that we realized upon the making of our Sankey Diagram since, almost 90% of the TV Shows did not have a director hence our data had clearly a huge flaw. We fixed this issue.

Dashboard Overview



Our dashboard draws inspiration from the "Hall of Fame" and consists of three interconnected visualizations. Each serves a specific purpose:

1. **Map Visualization**: This map illustrates geographic data, showcasing which countries have the most availability of movies and TV shows. It allows users to investigate regional trends, such as whether northern countries have a greater variety of content or if other regions are more prominent. It is mandatory because we want to be able to examine the difference between Northern and Southern Countries at a glance. We also added some pie charts to all the countries that contained data to instantly have some insights regarding the genres of the TV/Shows, Movies. We added shortcuts to directly zoom into Europe for example where displaying a bar chart with 0 zoom would lead to an uncomprehensible map.

2. **Top Ranking Bar Chart:** This bar chart provides a customizable top rankings feature, enabling users to see the top x up to 20 movies/tv shows. Users can apply filters such as production by Netflix, year of release, time range, type (movies or TV shows), and genre. This allows for a flexible

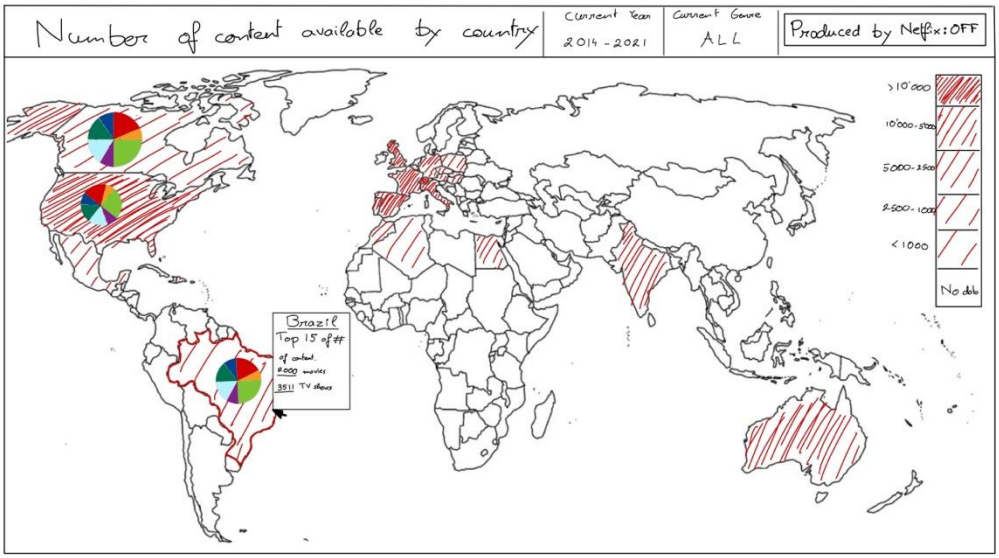
exploration of top content based on personalized criteria. We added the usage of small multiples such that without interactions we can clearly with what kind of movies we are interacting. Let me explain myself, you will see the first movie for example, you will see an information on the availability in countries, you will see it's score of course, the number of days after it released and was added to Netflix.

3. **Sankey Diagram:** This chart visualizes the distribution of genres across TV shows and movies. It clearly displays how different genres are divided between these two formats. The Sankey diagram allows users to see the flow and proportion of each genre, such as action, comedy, romance, etc. Without any interactions, we can directly see the #number of TV shows and Movies in different genres and vice versa.

All these visualizations are interconnected. For instance, if a user selects a genre, like "Action," the map updates to show the countries where action content is most available. Simultaneously, the top list adjusts to reflect the highest-ranking action movies or TV shows based on the filters applied. This synergy between the different visual elements enables users to gain deeper insights and make more informed comparisons across different data dimensions.

Charts

Choropleth Map



This image implements the new pie chart of the map

Marks and Channels

Marks: The Choropleth map uses areas (country borders) as marks. Each country's area is colored to represent the number of available TV shows and movies. It also uses pie charts to display the % repartition of the genres.

Channels:

- Color: The primary visual channel used is color intensity. Darker shades represent a higher count of available content in that region, while lighter shades (or no color) indicate fewer or no data.
- Tooltip: When hovering over a country, detailed data (like the number of TV shows and movies available in Brazil) will appear as a text box. This is an example of using textual encoding through interaction.
- Position: The countries are positioned based on their actual geographic locations, which allows for an intuitive understanding of global availability trends.

Rationale

The Choropleth map is an ideal choice for visualizing geographical data, as it provides an easy way to understand distribution patterns across the globe. The use of color intensity is effective in highlighting differences in the amount of content available per country. Here's why we chose this idiom:

- **Geospatial Clarity**: The map directly connects the dataset (availability of Netflix content) to geographical areas, making it easy to interpret region-specific availability, are there any difference between the northern countries and the southern ?
- **Scalability**: This approach works well even with a large number of data points (countries), avoiding visual clutter.
- **Data Type Fit**: The map effectively handles categorical data (countries) combined with ratio data (amount of content).

Alternatives Considered:

- **Bubble Map**: While bubble maps could show availability with different-sized circles, this approach could clutter regions with smaller landmasses. Additionally, color gradients in a Choropleth map offer a clearer visual comparison across countries. Finally circles are not a great way to compare sizes.

Interaction

The Choropleth map supports the following interaction techniques:

- **Hover**: Users can hover over a country to see a tooltip providing detailed data for that region (e.g., number of TV shows and movies). This dynamic interaction adds clarity without overwhelming the visualization.
- **Click**: Clicking on a country allows users to filter other visualizations. For example, selecting a country will update the Top List and Sankey Diagram to display only the content available in that specific country.
- **Range Filter**: Users can filter by time range (e.g., content available between 2014 and 2021), which dynamically updates the color scale and availability statistics in each country.
- **Genre Filter**: A filter for genres is available so that users can view the availability of specific genres per country. The map adjusts accordingly based on the genre selected.

Chart Integration

The Choropleth map is integrated with other components of the dashboard to enable seamless interactions:

- **Selection Synchronization**: When a user selects a genre or time range from the Sankey diagram or Top List, the map updates to display content availability specific to that genre or period.

- **Highlighting**: If a user clicks on a country in the map, this action highlights the top content available in that country in the Top List and breaks down genre availability in the Sankey diagram.
- **Dynamic Filtering**: The map works in tandem with the filters applied to the Top List (e.g., "Netflix Originals" or "Produced by Netflix: On/Off"). Changes in these filters will be reflected on the map by altering the availability statistics and visual colors of the countries.

Answering the Questions

Question 1: What are the top 9 action movies or TV shows between 2014-2016 Netflix, and in which years were they released?

1. Set the year slider to 2014-2016.
2. Select "Action" on the Sankey diagram to filter the results.
3. Check the "Your Top" list to see the top 9 action titles, along with their release years.

Question 2: How many Netflix movies and Netflix TV shows were released in 2018?

1. Set the year slider to 2018.
2. Hover over the "Movies" and "TV Shows" sections in the Sankey diagram to see the counts.
3. Toggle the “Produced by Netflix” filter if you only want Netflix Originals.

Question 3: How many romantic movies or romantic TV show were added between 2017-2019?

1. Set the year slider to 2017-2019.
2. Select "Romance" in the Sankey diagram.
3. Hover over the "Romance" section for the total count of movies and TV shows.

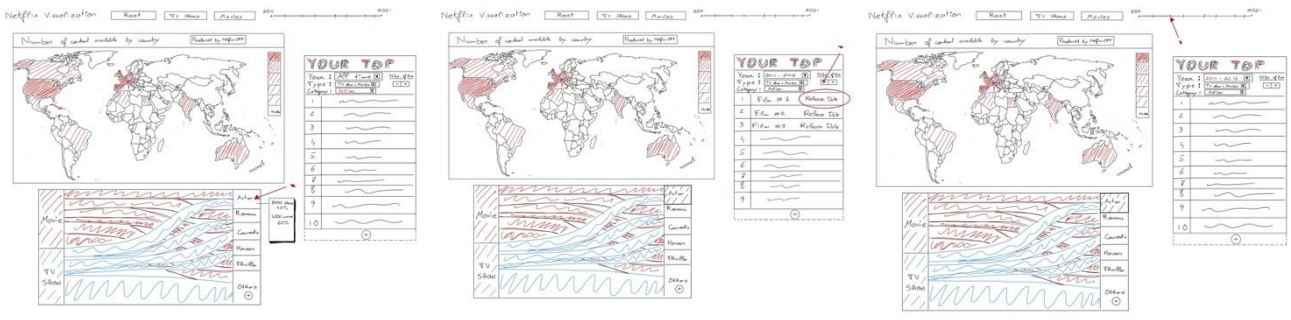
Question 4: Does the USA have the most action movies and TV shows available on Netflix?

1. Select "Action" on the Sankey diagram.
2. Hover over the USA on the map to see the count and its rank compared to other countries.

Question 5: Do northern European countries have more comedy Netflix produced content on Netflix than southern countries?

1. Toggle the "Produced by Netflix" button.
2. Select "Comedy" in the Sankey diagram.
3. Compare northern and southern European countries on the map by their color intensity or hover over them to see exact numbers.

Storyboards



Missed Opportunities

We could have done something completely different like using a Sankey diagram to link a group of common actors per countries and per movies.

If we had kept some budget per movies (unfortunately not enough data was present) we could have compared the budgets of movies with respect to some actors present, or other criteria.

We could have tried to cross match 2 datasets to see the number of views per movies across the years.