

Data Visualization Project 2 Report

Unicorn Startup Companies Data Visualization

Name: Josh Hernett Tan

Student ID: 32421982

Lab Class: Tutorial 05 (Friday, 9 am - 12 pm)

Words: 890 (excluding Cover, References, and Appendix)

https://josh8t.github.io/FIT3179_Assignment2/

Domain

The domain chosen is “Unicorn Startup Companies”. Unicorns, referring to privately held startups valued at over \$1 billion, are a critical landscape in the business ecosystem due to these entities' significant influence on technological innovation, economic vitality, and market trends.

These rare successes, known for reshaping industries, are invaluable study subjects for a diverse audience, including investors, entrepreneurs, market analysts, policymakers, and job seekers.

Analyzing the trajectories and attributes of unicorns offers deep insights into high-stakes entrepreneurship and investment, potentially forecasting broader market dynamics and contributing to informed policy-making and strategic business decisions.

Data

Original Source: <https://www.kaggle.com/datasets/tahzeer/unicorn-startup-companies-july-2023>

Cleaned Version:

https://github.com/Josh8T/FIT3179_Assignment2/blob/main/unicorn_startup_companies_cleaned.csv

Data cleaning was done to fix some null values. For this assignment, the cleaned version of the dataset was used. The dataset contained 1222 lines and 9 columns. The dataset is only recorded up to July 2023.

- **Rank:** The standing of the company in terms of its current valuations compared to other companies
- **Company:** Name of the unicorn
- **Valuation (\$B):** Current valuation in billion dollars
- **Date Joined:** Date it reached 1 billion valuations.
- **Country:** The country where the startup was founded.
- **City:** The city where the startup was founded.
- **Industry:** Field of the startup
- **Select Investors:** The main investors in the company.
- **Continent:** The continent where the startup was founded

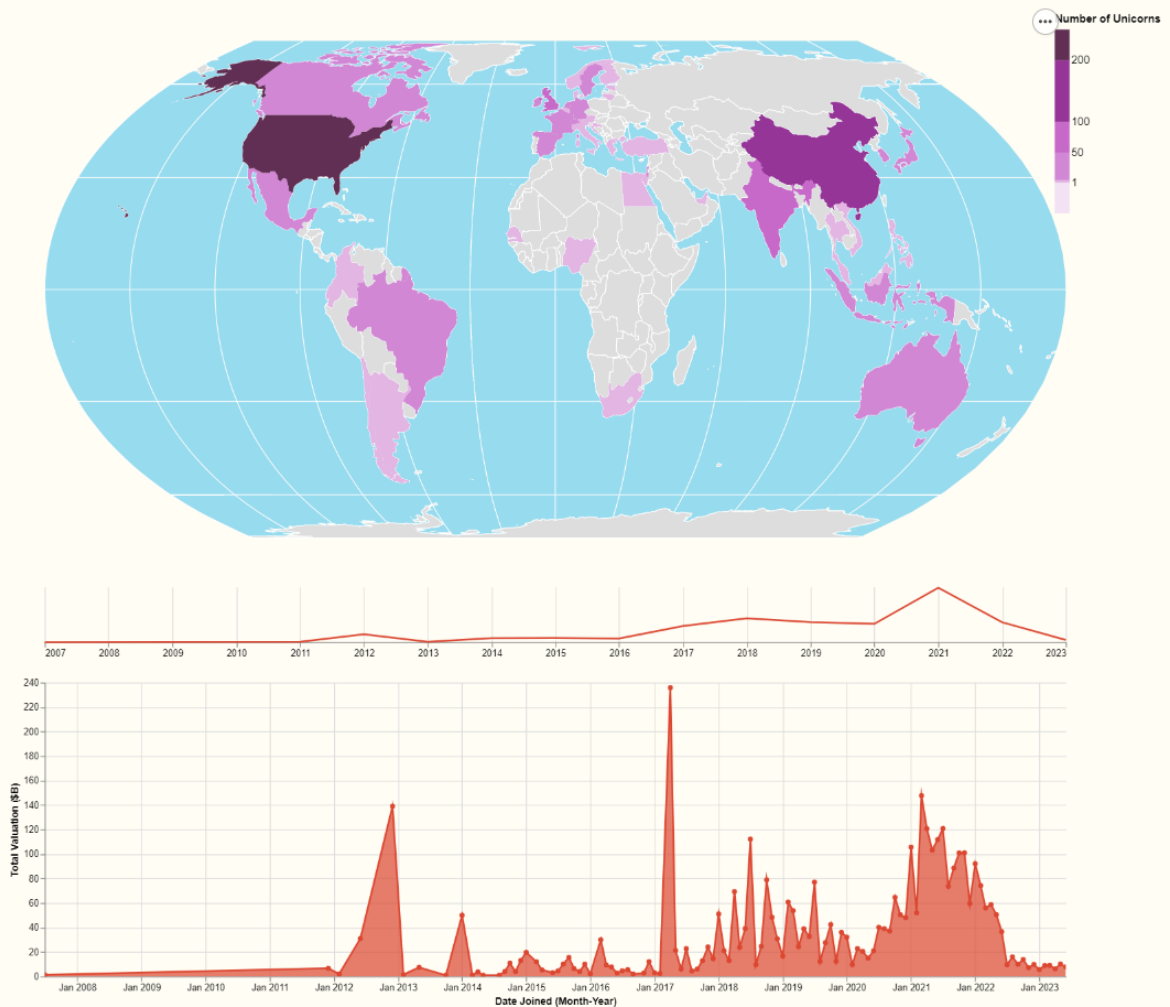
Website Overview

Unicorn Startups Companies

An Analytical View into Unicorn Startups

Defined as privately held startups valued at \$1 billion or more, unicorns are recognized for their substantial impact on the industries they operate within and their notable investment traction. Utilizing Vega-Lite visualizations, this platform offers you a detailed analysis, enabling an examination of patterns, trends, and insights into the world where business, innovation, and significant capital coalesce.

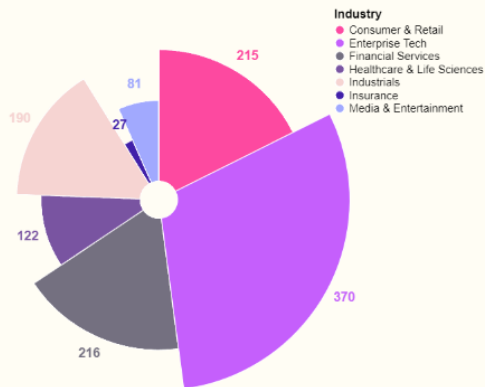
Choropleth Number of Unicorns per Country & Growth of Unicorns Globally



The choropleth map gives an overview of the **number of unicorn startups by country**. The area chart tracks the **growth of unicorn startup companies globally**. Using the line chart provided, you can isolate specific range of time to view for more details. This will provide a visual representation of the **economic impact** of these startups on different countries.

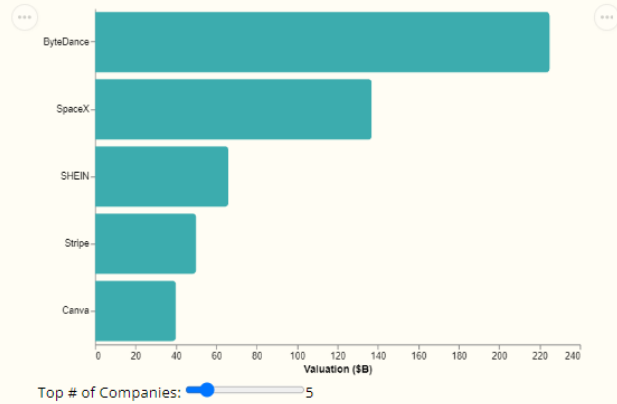
Figure 1 Entire View of Website (Part 1)

Distribution of Unicorn Industries



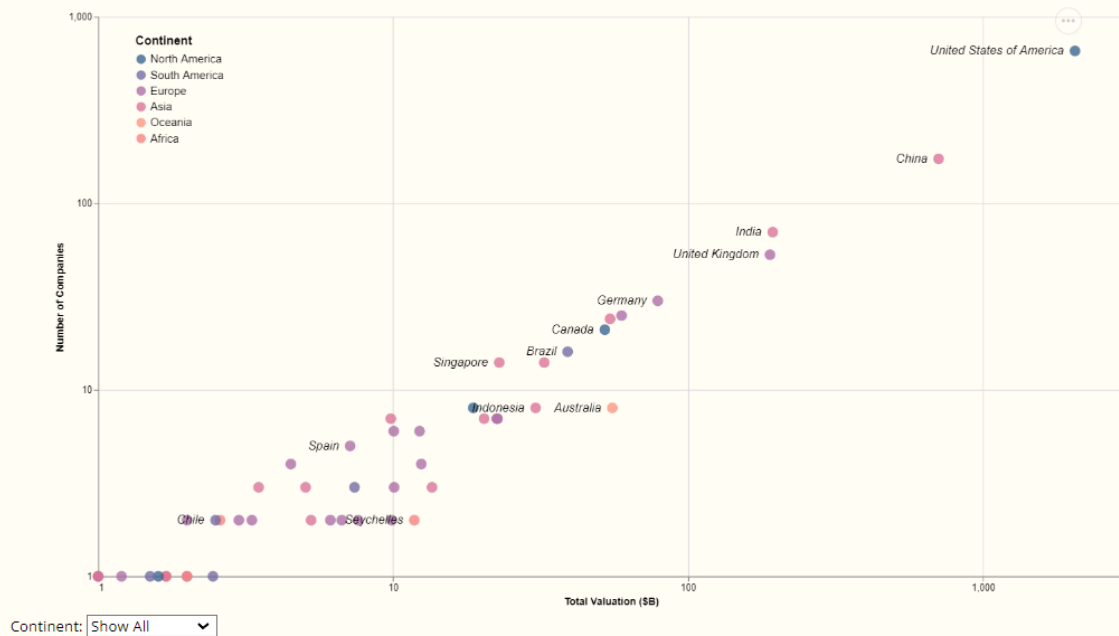
This pie chart depicts the **distribution of unicorn startups across various industries**. Each slice represents a distinct sector, with the accompanying number indicating the total unicorn companies in that domain. Notably, the **Media & Entertainment** industry leads with a staggering **370** startups, showcasing its dominant presence in the unicorn ecosystem.

Highest Valuation Unicorn



This bar chart showcases the **top-tier unicorn startups based on their astounding valuations**. Leading the pack is ByteDance, reaching valuations soaring **above \$200 billion**. Use the interactive feature below to adjust the number of companies displayed, always ranked by their valuation prowess.

Unicorn Distribution by Country: Count and Total Valuation



This scatterplot to explore how countries stack up in terms of unicorn company count and combined valuation. Notably, the **United States** and **China** stand out as the titans of the startup world. Utilize the 'Continent' filter to narrow down your view and uncover regional trends.

Figure 2 Entire View of Website (Part 2)

Idiom

Choropleth Map

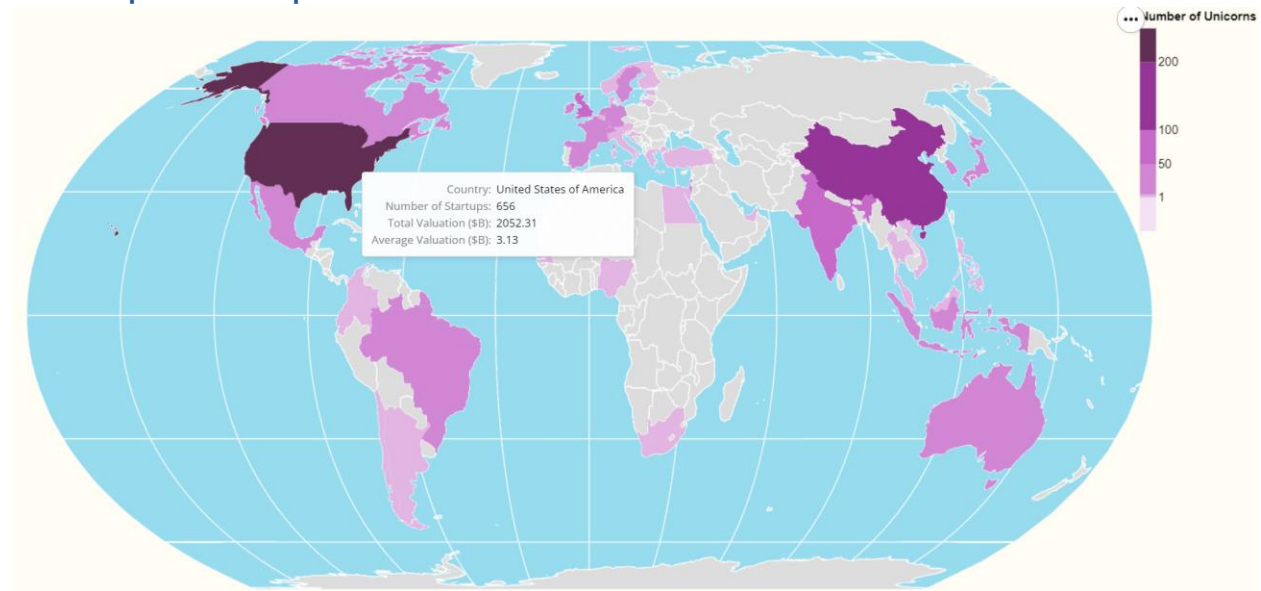


Figure 3 Choropleth Map

- *Figure 3* is used to visualize the number of unicorns per country.
- A choropleth map is used to offer a balanced view of unicorn distribution worldwide.
- Utilizing a sequential color palette, a gradient from light pink to deep purple, to represent the concentration or intensity, with deeper hues indicating more unicorns.
- Tooltips provide in-depth insights into specific countries, complementing the broad visual overview.

Area Chart

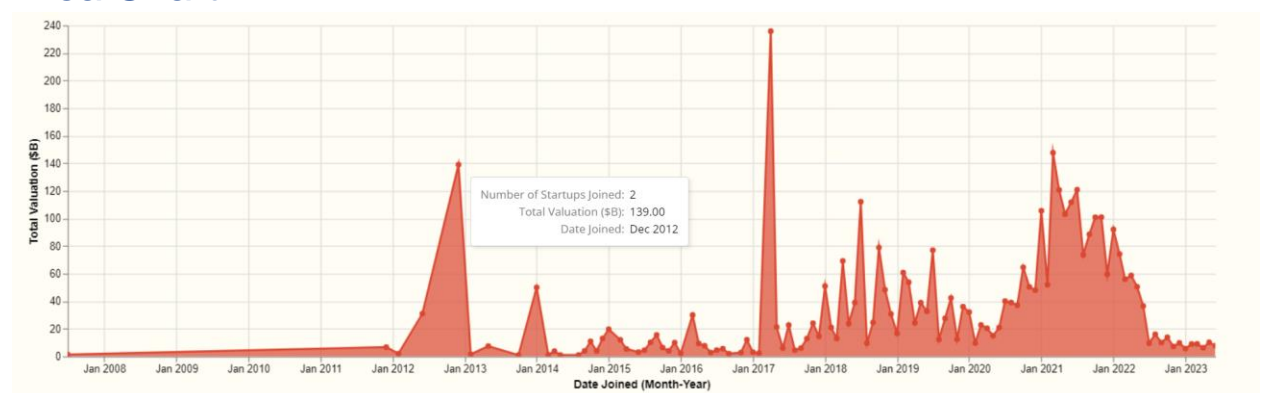


Figure 4 Area Chart

- *Figure 4* is used to visualize the number of startups getting the 'Unicorn' title over time.
- An area chart is used to offer a continuous representation of the unicorn club's growth over the years.

- A consistent shade of orange is used throughout the area under the line. This is to contrast the beige background, emphasizing attention.

Line Chart

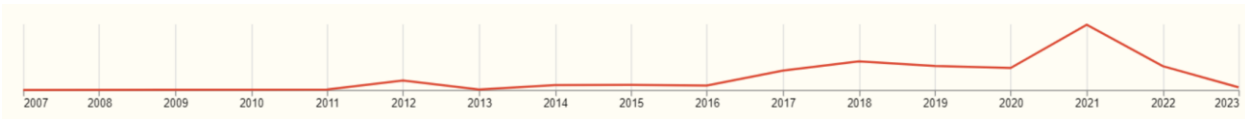


Figure 5 Line Chart (Default)



Figure 6 Line Chart (Using Time Brush)

- *Figure 5* is a line chart used as an interactive feature. By clicking on the chart and dragging it across the x-axis, a rectangular highlight appears and will continuously stretch until the viewer releases the click (as seen in *Figure 6*). The highlight then can be dragged across the x-axis to filter the range of dates for both *Figure 3* and *Figure 4*. This feature allows the viewer to observe much greater detail.

Radial Plot

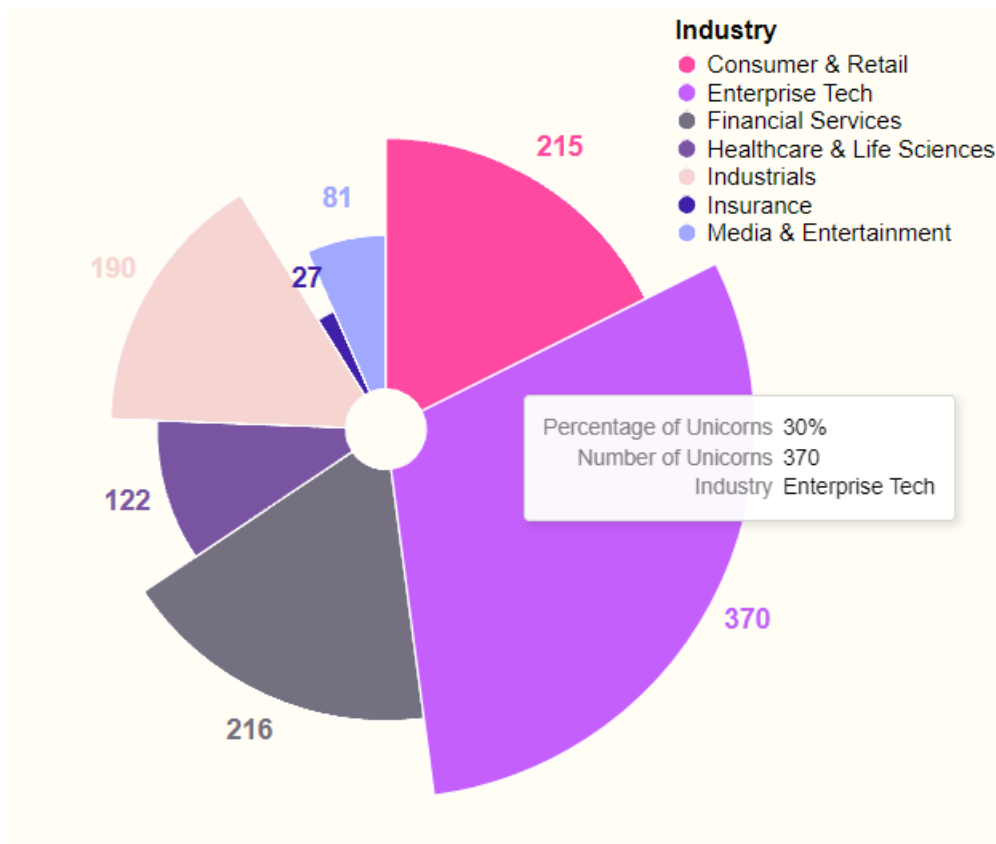


Figure 7 Radial Plot

- *Figure 7* depicts the distribution of unicorn startups across various industries/fields.
- Used a radial plot as it offers a circular arrangement to display the distribution of startups across various industries. Each sector's arc length directly corresponds to the number of startups it represents, thereby visually indicating the prominence of each field.
- Various distinct and vibrant colors are utilized to differentiate each industry, ensuring no two segments are easily confused.
- Numbers indicating the count of startups in each industry are strategically placed close to their respective arcs. This ensures immediate association between the count and the industry.

Bar Chart

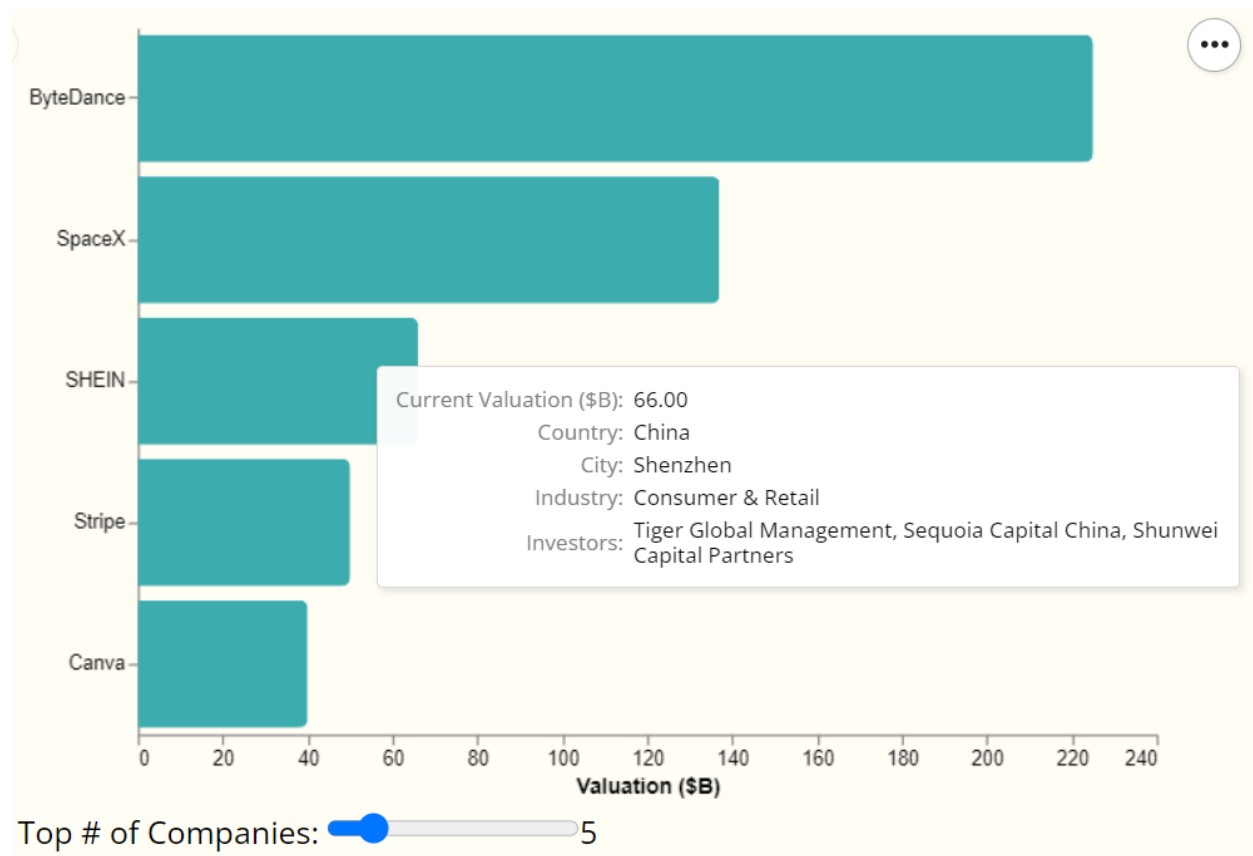


Figure 8 Bar Chart

- *Figure 8* visualizes the top highest valuation unicorn startup.
- Usage of a bar chart employs horizontal bars to represent the valuation of the top startups, facilitating easy comparison of their relative magnitudes. The longer the bar, the higher the valuation of the corresponding company.
- Consistent teal color is used for each bar, maintaining uniformity and focus on the data.
- At the bottom of *Figure 8*, there's an interactive slider labeled "Top # of Companies," currently set to 5. By dragging the slider to the right, and increasing the value, viewers can see more companies. This provides a dynamic aspect to the data presentation.

Scatter Plot

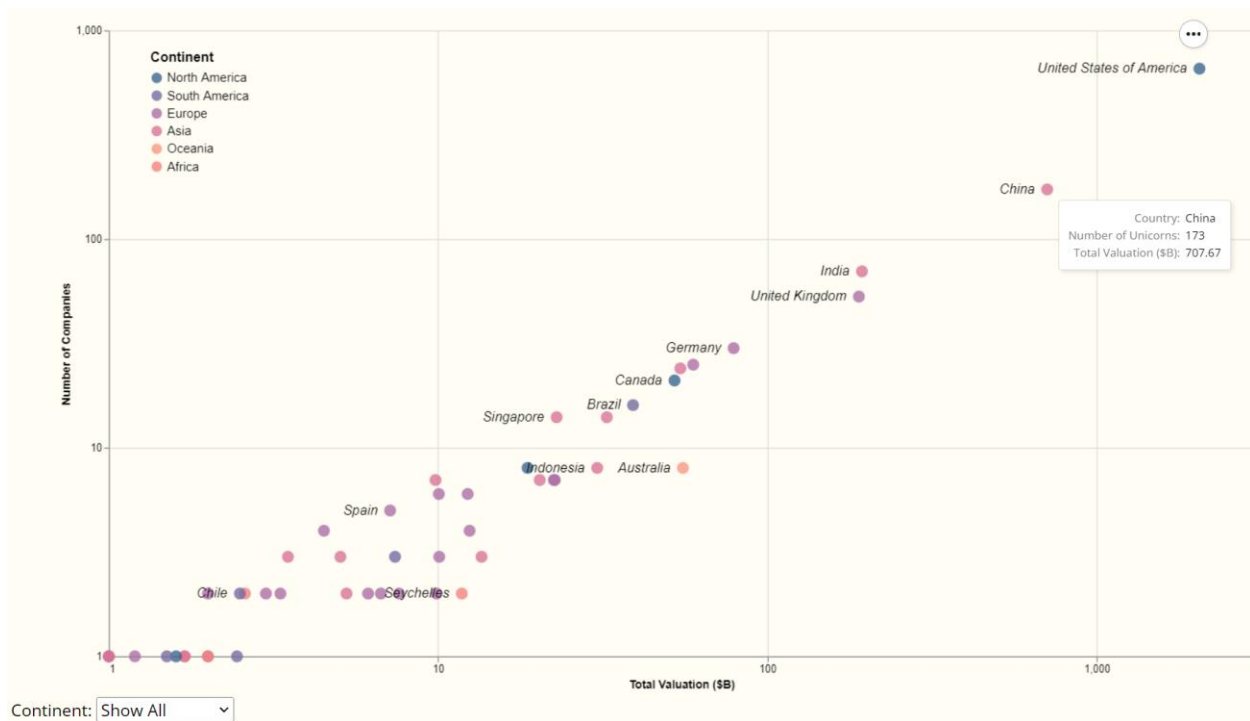


Figure 9 Scatter Plot

- *Figure 9* visualizes how countries stack up in terms of unicorn company count and combined valuation.
- A scatter plot is used here because it allows for the simultaneous visualization of two quantitative variables for each country: the number of unicorn companies (y-axis) and their total valuation (x-axis).
- *Figure 9* employs a logarithmic scale on both axes, allowing for a clear representation of countries' unicorn distribution across wide ranges of company numbers and valuations.
- A distinct color palette is used for *Figure 9* to not confuse viewers with the previous visualizations.
- At the bottom of *Figure 9*, there's a dropdown menu labeled "Continent," this allows viewers to filter the display of countries based on specific continents. This provides an additional layer of granularity and adaptability to the viewer's interest.

Website Design

- Layout
 - The layout of the dashboard has easily distinguishable sightlines and balance, where the left space of the dashboard is used for the idioms while the right is used for text.
- Colour
 - The main color of the dashboard is beige, for simplicity and to contrast the color palettes that the idioms used for better distinguishment.
 - Color palettes used for each idiom are unique, to not confuse viewers between the representation of values on the visualizations.

- Typography
 - The typography used on the entirety of the website is *Open Sans* as it has a classy and neutral appearance, in addition to its clarity and readability.
 - The title is configured to have the highest weight and size, the titles of each idiom also have the same weight as the title but are smaller in size. This is done for hierarchal purposes.
 - The text prompts have the lowest weight and size as they have the lowest standing on the hierarchy.

References

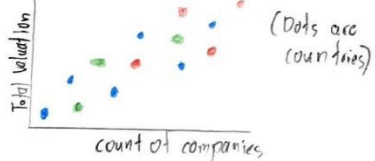
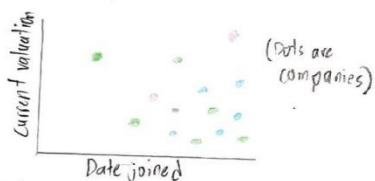
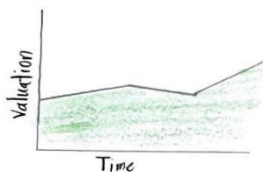
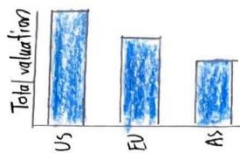
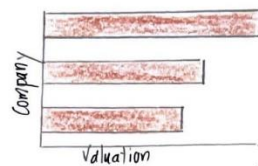
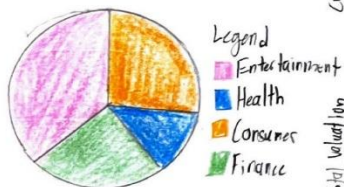
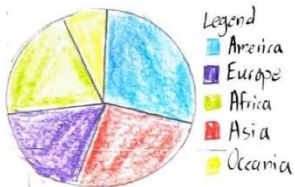
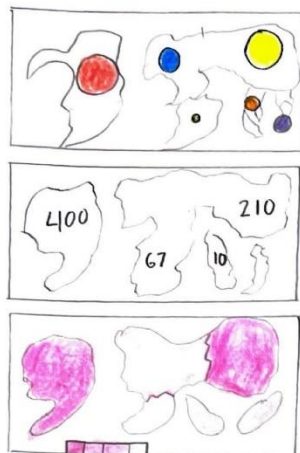
OpenAI. (2023). ChatGPT (September 25 Version) [Large language model].
<https://chat.openai.com>

Unicorn Startup Companies July 2023. (2023, August 3).

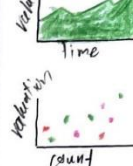
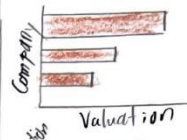
Kaggle.<https://www.kaggle.com/datasets/tahzeer/unicorn-startup-companies-july-2023>

Appendix

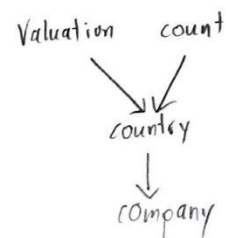
IDEAS



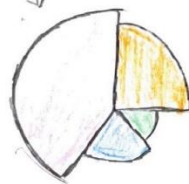
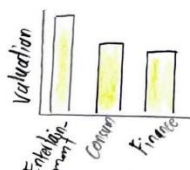
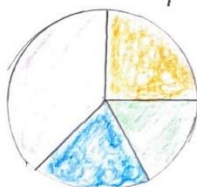
FILTER



CATEGORIZE



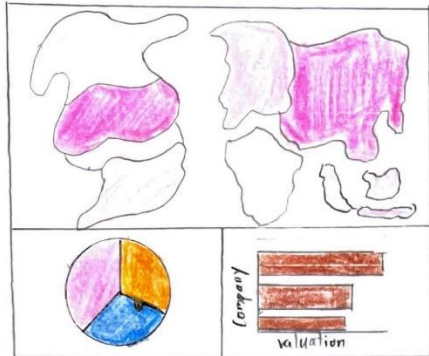
COMBINE & REFINE



Question

- 1) Are implementation feasible?
- 2) Does it show how startups effect country economy?
- 3) Shows top startups?

LAYOUT



Title: Starting design

Author: Josh Hemmett Tan

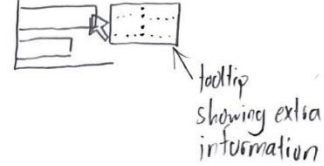
Date: 1/10/2023

Sheet: 2

Task: Simple Implementation (trial & error phase)

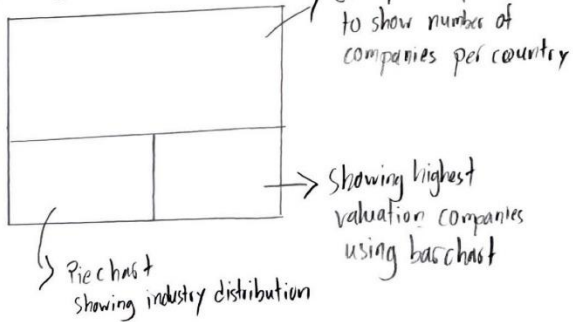
OPERATIONS

Tooltips implemented
ex:



FOCUS

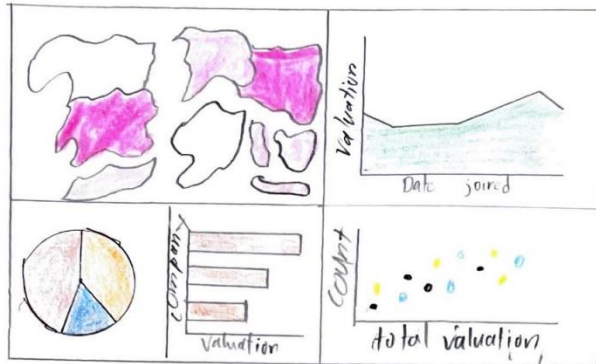
focusing on starting visualizations:



DISCUSSIONS

- Need to add more visualization to answer all questions
 - ↳ No time series chart implemented

LAYOUT

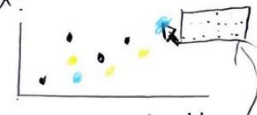


Title : Dashboard design
 Author : Josh Harnett Tan
 Date : 1/10/2023
 Sheet : 3
 Task : Implementation 2 (All visualizations)

OPERATIONS

Tooltip implemented for all visualisation

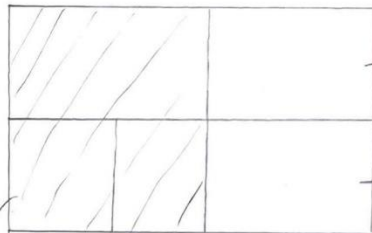
ex:



tooltip popup when cursor come in contact with data

FOCUS

Focusing on implementing all the visualization properly.
 Ensuring no errors and null values for all visualization.



→ Area chart implemented

→ Scatter plot implemented

entire section is the same as previous sheet

DISCUSSIONS

- Each visualization have their own colour palette to ensure no confusion with data representation.
- Implemented all visualization necessary
- No interactions?

LAYOUT



Title: Dashboard
 Author: Josh Bennett Tan
 Date: 2/10/2023
 Sheet: 4
 Task: Implementation 2
 (Adding features)

OPERATIONS

Continent select filter

Continent: Asia

Clicking on this allows viewer to filter the scatterplot based on continent

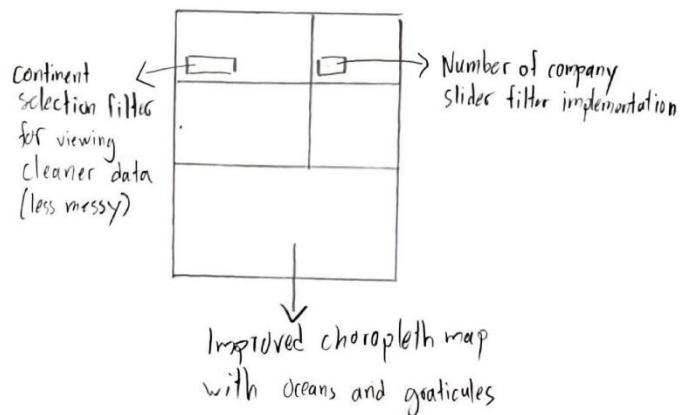
Number of company shown
 Slider filter

Number: [slider]

Going further to the right will show more companies on the bar chart

FOCUS

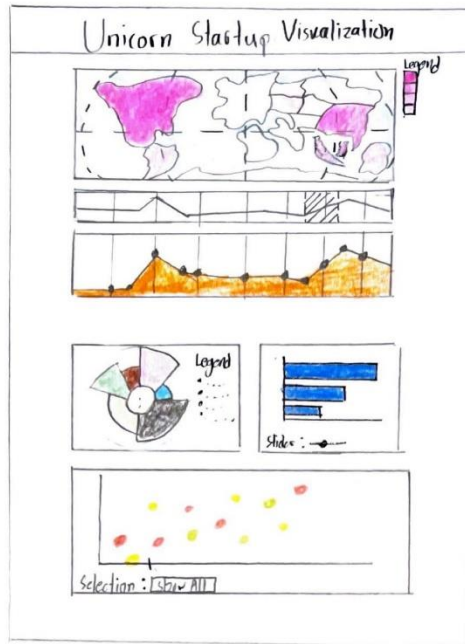
Focusing on visualization improvement such as interactivity.



DISCUSSION

- Visualizations and interactivity are all implemented
- layout improvements needed for final design.
- Any more improvements possible?

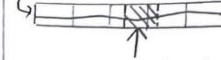
LAYOUT



Title: Final Design
 Author: Josh Kenneth Tan
 Date: 3/10/2023
 Sheet: 5
 Task: Implementation Design (Final)

OPERATIONS

line chart



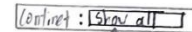
- Time brush interaction
- Changes choropleth map and Area chart based on the time range selected

Slider



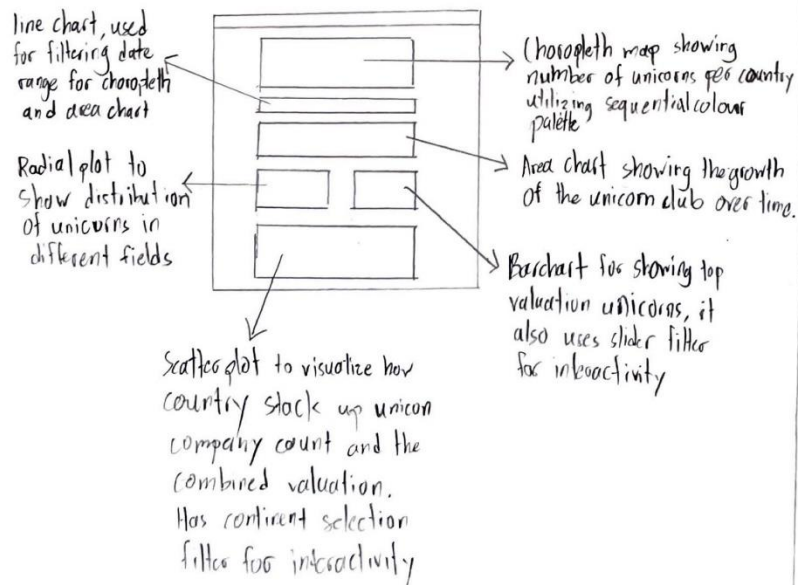
- Slider filter interaction
- Shows more unicorn on the bar chart going further to the right

Selection



- Selection filter interaction
- Filters the content of scatterplot based on which continent selected

FOCUS



DISCUSSION

- Contains necessary information
- Have interactions for viewers
- Time required: 1 week +
- Improved pie chart to radial plot
- Added line chart as interactive filter