**SStats 401 final project proposal**

Team member： Jiyuan Wang, Luyao Wang, Peng Wang

1. Introduction

"Genshin Impact," developed by miHoYo, is a globally popular action role-playing game that integrates an in-game lottery system, commonly known as the "Gacha" mechanism, to acquire characters and weapons. The “Gacha” system, which originates from Japanese mobile games, encourages players to spend real or in-game currency for a chance to get their desired in-game items. There are 3 banners in the game: “up” character,weaponand regular banner. You spend a pull to draw from the banners. We want to analyze the real probability of getting a character or specific weapon and verify whether it corresponds with the probability with that published by miHoYo.

1. Datasets

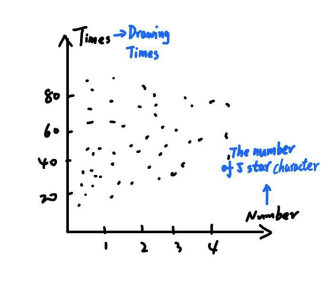
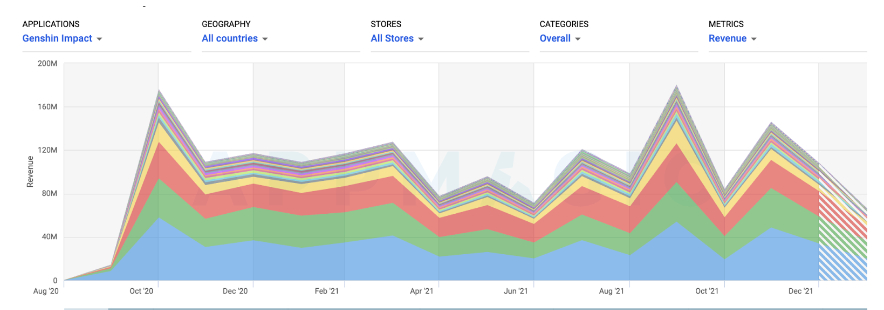
We use datasets from 3 sources: open-source dataset, scraped data, and data from API. The open source “Gacha” dataset is downloaded from github (https://github.com/OneBST/GI\_Gacha\_dataset) which is constructed through data voluntarily disclosed by players, about 15.59 million pulls. We also include scraped team composition statistics (<https://akashadata.com/total/>). For the player-level current "Gacha" record, we can acquire the data via the official web API from miHoYo (https://github.com/jvergerolle/Genshin-Impact-Wish-history-API).

1. Time river of the Recharge turnover in “up” character, weapon, and regular banner.

We employ Python to cleanse and scrutinize the Gacha game dataset, focusing on player spending across different banners—specifically, the 'Up' character, weapon, and standard banners. Through our program, we estimate the approximate recharge turnover for individual players and aggregate this data to identify trends across various banners. Utilizing a time-series visualization, which we refer to as a 'Time River,' we then map the shifting proportions of recharge turnover over time, providing insights into player spending behaviors.

1. Sketches and references

This visualization emphasizes the dynamic nature of player preferences by animating the shifts in the top five most acquired characters over time. Presented in the second and third graphs in the second row below, this feature offers a real-time tracking mechanism that aggregates the total count of each character obtained by all players. Through this animated timeline, we capture the evolving popularity of individual characters, offering valuable insights into changing trends and player behavior.

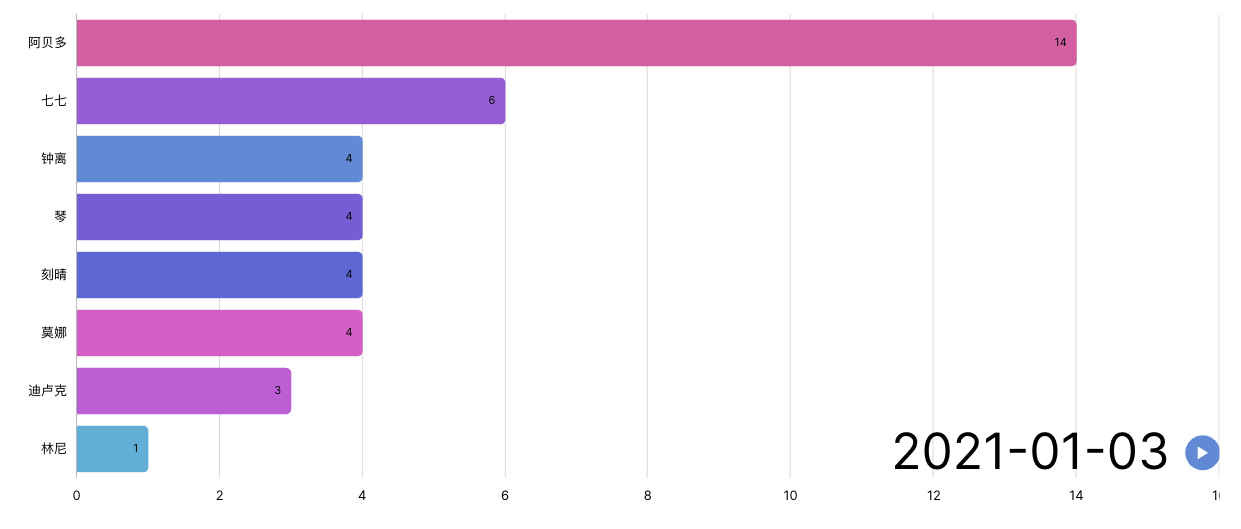
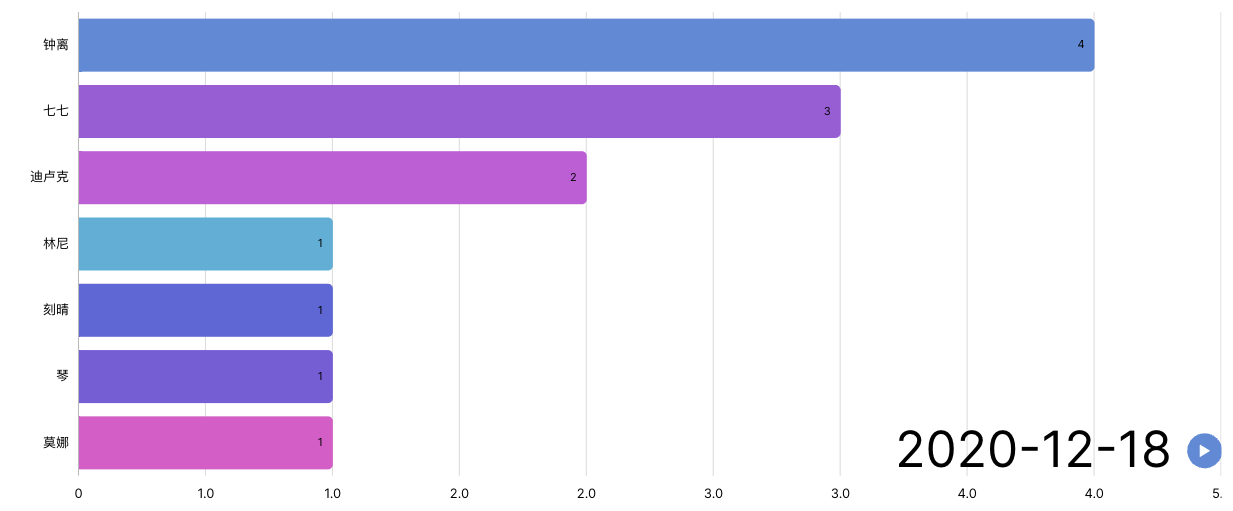
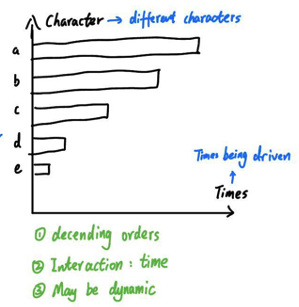


Figure 1. Visualization 1-3

Visualization 1: Average polls taken to draw a 5-star character vs number of polls.

Visualization 2: Time river of the Recharge turnover

Visualization 3: Total character count for all the open source “Gacha” data, allows for animation showing the change of character count over time.

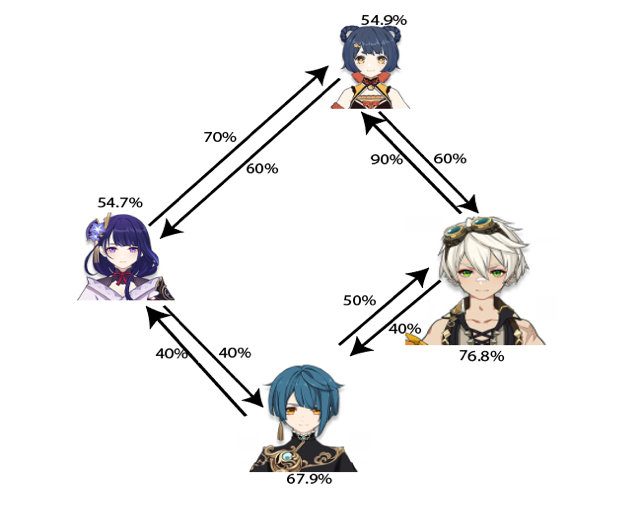
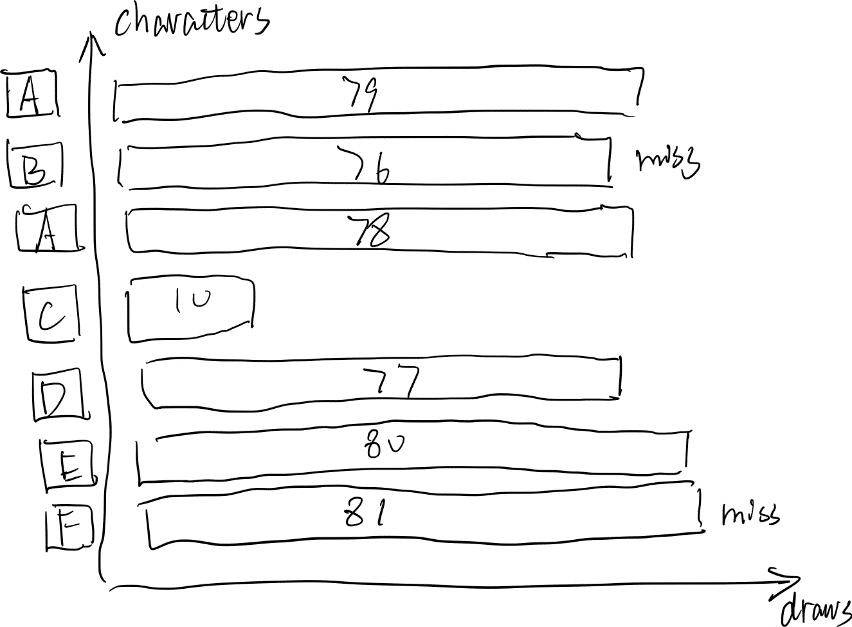


Figure 2 & 3. Visualization 4-5

Visualization 4: This visualization shows the “Gacha” history, average pulls needed to get an “up” character, etc.

Interactions: change the banner: “up” character/weapon/regular banner; Click on each entry to view the detailed information, e.g., the time that pull was made

Visualization 5: Team composition network

In Genshin Impact, a team consists of four characters. The network answers mainly two questions: 1) The character usage rate of each character, reflected by the icon size. 2) How important is one character to another? An arrow from A towards B means out of all the team composition occurrences of B, what is the percentage of A’s occurrences. A higher percentage is marked by a bolder line. Users can choose to only view a selected number of characters or filter out character usage rate below a certain amount.

1. Roles and responsibilities
   1. Jiyuan Wang: data visualization on D3 coding part, basically the first chart.
   2. Luyao Wang: react app development and frontend visualization, data acquisition.the third and the fourth graph.
   3. Peng Wang: data cleaning and preprocessing and data visualization. The second chart and the last chart
2. Deliverables for the Interim Presentation
   1. Use the open source “Gacha” datasets to analyze how the Genshin Impact “Gacha” mechanism corresponds to its official probability.
   2. “Gacha” record as a useful tool to review one’s own pull history.
   3. Team composition as a reference to building popular and strong teams.
3. Timeline and milestones
   1. Week 4: data acquisition, react app set up.
   2. Week 5: data cleaning and data analysis.
   3. Week 6: about 2 visualizations done.
   4. Week 7: all 5 visualizations done, webpage integration.