Gunpla data viz:

1. First I had scraped the data from reddit using the PRAW following the method as indicated from this tutorial (<https://medium.com/swlh/scraping-reddit-using-python-57e61e322486>).
2. Then I had launched it as csv file and obtained the top 1000 posts from the gunpla subreddit from 2021 to 2022 to visualize **top 10 most popular gunpla within the past year**
3. Using the data scraped from the subreddit, I primarily used the titles of the posts to determine the type of gunpla being shown.
   1. The approach was to use select keywords (e.g., \_\_\_\_ gundam, zaku) from the posts to determine which model kit was being shown
      1. For this to work, I used regex to split the title into a list of individual words as strings (using space as separators)
   2. Therefore, I had created a list of keywords distinguishing each model kits. For example, the original gundam used the keyword “RX78-2”. Some models had to be group together as they were variations of one another OR they had duplicate words within them making it difficult to create their own individual keywords. These examples include the zakus and unicorns.
   3. After the list was created, nested for loops were used to look for the number of occurrences of the keywords in the post titles for all titles in the data (and count them)
4. Once a list of count data (list) was created, a pictogram using pyWaffle was created following this tutorial (<https://towardsdatascience.com/2-efficient-ways-of-creating-fancy-pictogram-charts-in-python-8b77d361d500>) to showcase the top 10 most popular gunpla within the past year.

A picture containing diagram

Description automatically generated