The purpose of this project is to take data from a game called Pokemon Sleep, and analyze it to show the nuances of the spawns present in the data. Pokemon Sleep is a game that tracks your sleep, and then in the morning between 3 and 8 Pokemon spawn and you have a chance to catch them. The varieties of Pokemon that spawns seems random, but through this project I was hoping to reveal if there are certain conditions in which specific Pokemon are able to spawn. The one set of conditions that the game makes obvious is which of the 3 islands you select: Greengrass Isle (gg), Cyan Beach (cb), or Taupe Hollow (th). This is why the labels of each possible Pokemon to find are split for each of the 3 islands, I want to find the groupings treating a Pokemon spawning on each different island as a separate entity. Because of these 3 islands and there being 450 varieties you can find (less species, but each has 3 or 4 varieties), there are 1,350 total vertices in the graph. It gets constructed with data separated by spaces that is set up as follows:

Total number of possible varieties (450 currently in the game)
List of all styles possible
Island (gg, cb, or th) followed by the index number of each of the styles found in the sleep
Rest of data same format as line 3

This data was collected using an OCR bot in discord (a chat platform) that my friends developed and I helped out on that takes in-game screenshots and automatically records the data to this format, but it was not written in rust so I'm not going to go into more detail here. We currently have over 1000 data points of sleeps, but less than that was used for this project because special cases such as events that change the spawns were removed.

To explain how the code processes a line from the data into information for the graph, I'll just use an example line:

gg 0 15 89 214 20

First, it takes the island and sets a specific variable to an index of that island to make sure labels are applied properly. Each variety is present in the vector of labels once for each island, so to do a proper indexing you take the variety's original index (in the data) multiplied by 3, and add the variable that changes based on what island the line is (either 0, 1 or 2). The code then takes this line and iterating through each Pokemon that spawned creates an edge between them in the graph. So this line would result in the vertices 0, 45, 267, 642, and 60 all getting edges created between them, because Greengrass Isle has an island index of 0.

For the actual analysis, to try and find spawn groups or conditions I decided it would be best to run code that splits the varieties into component groups, being groups that either showed up in a sleep together, or have a mutual (or chain or mutual) varieties that they were both seen with. For each component I also calculated a "Representative Style" which is the variety with the lowest average distance to all other varieties in its component group. If 2 varieties were seen in a sleep together their distance would be 1, if they have a mutual that they were both seen with their distance would be 2, etc. This was done to see if there was a common Pokemon from each component that we could more easily use to draw conclusions about what conditions could cause the more rare ones from that same component to spawn. If 2 or more varieties had the same average distance, they all got printed as the representative style. The program only prints "non-singular" components, being ones that have more than varieties present within it. This was done because there are many varieties that weren't seen in our data set (either aren't possible or are very rare) so they have no edges, and printing all of those singular components massively clutters the output without adding much useful information.

From our dataset the program groups the vertices into 9 components total, which also aligned to 3 components for each island. Another interesting note was that different varieties of the same species were always together in the same component, so it's likely whatever condition determines which of these components you pull spawns from is on a species level, and potentially not on a per variety basis.