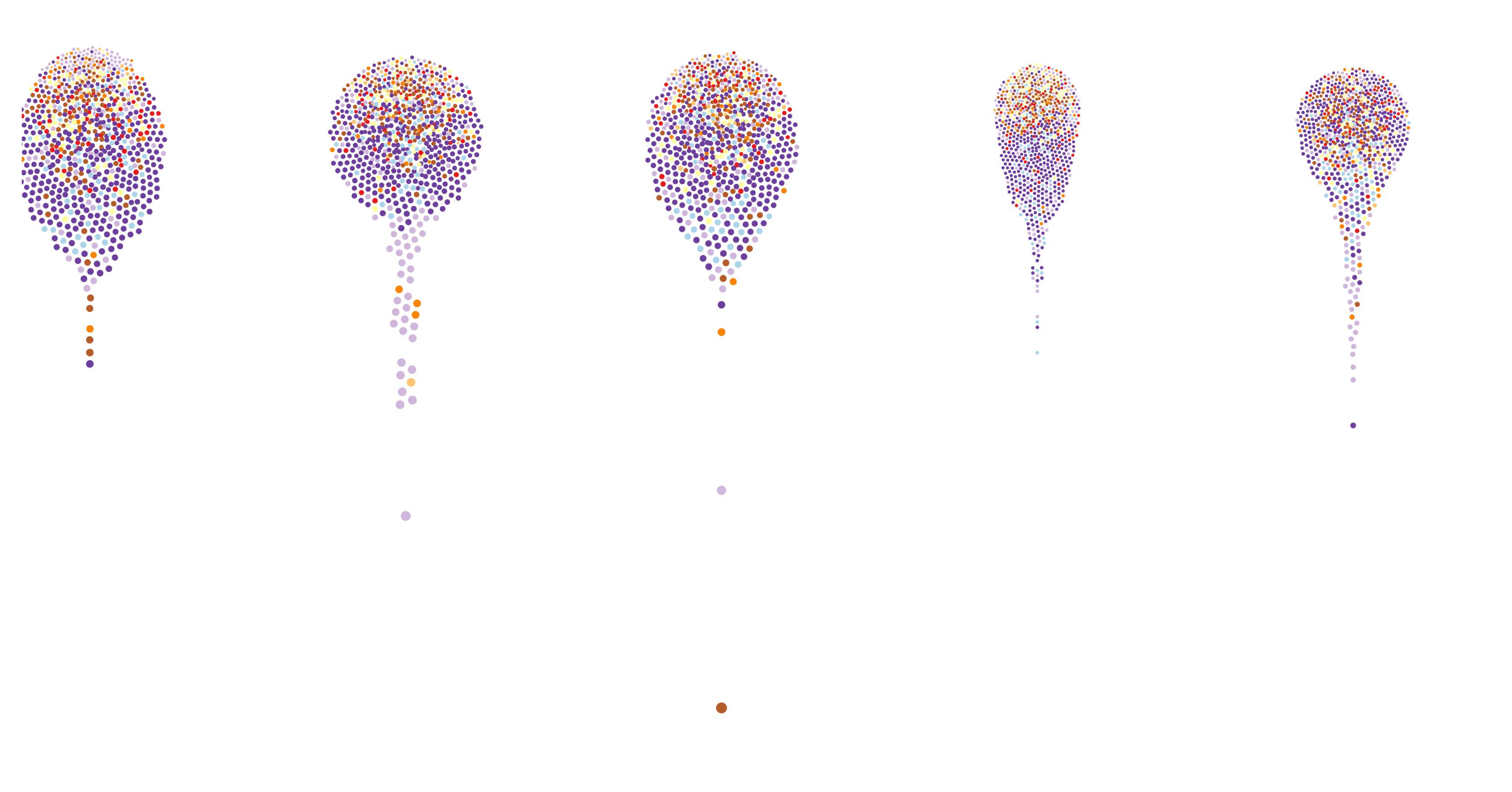
Design choices for the beeswarm visualization

Once we figured out our data(the dish dataset) and our target(visualize Japanese favorite dishes), nutrition is regarded as a critical dimension for the dishes, and beeswarm plot would be a fancy way to visualize the nutritional information of those dishes, since it enables us to quickly identify the distribution of values along several categories, corresponding to the categories of nutrients we are to visualize.

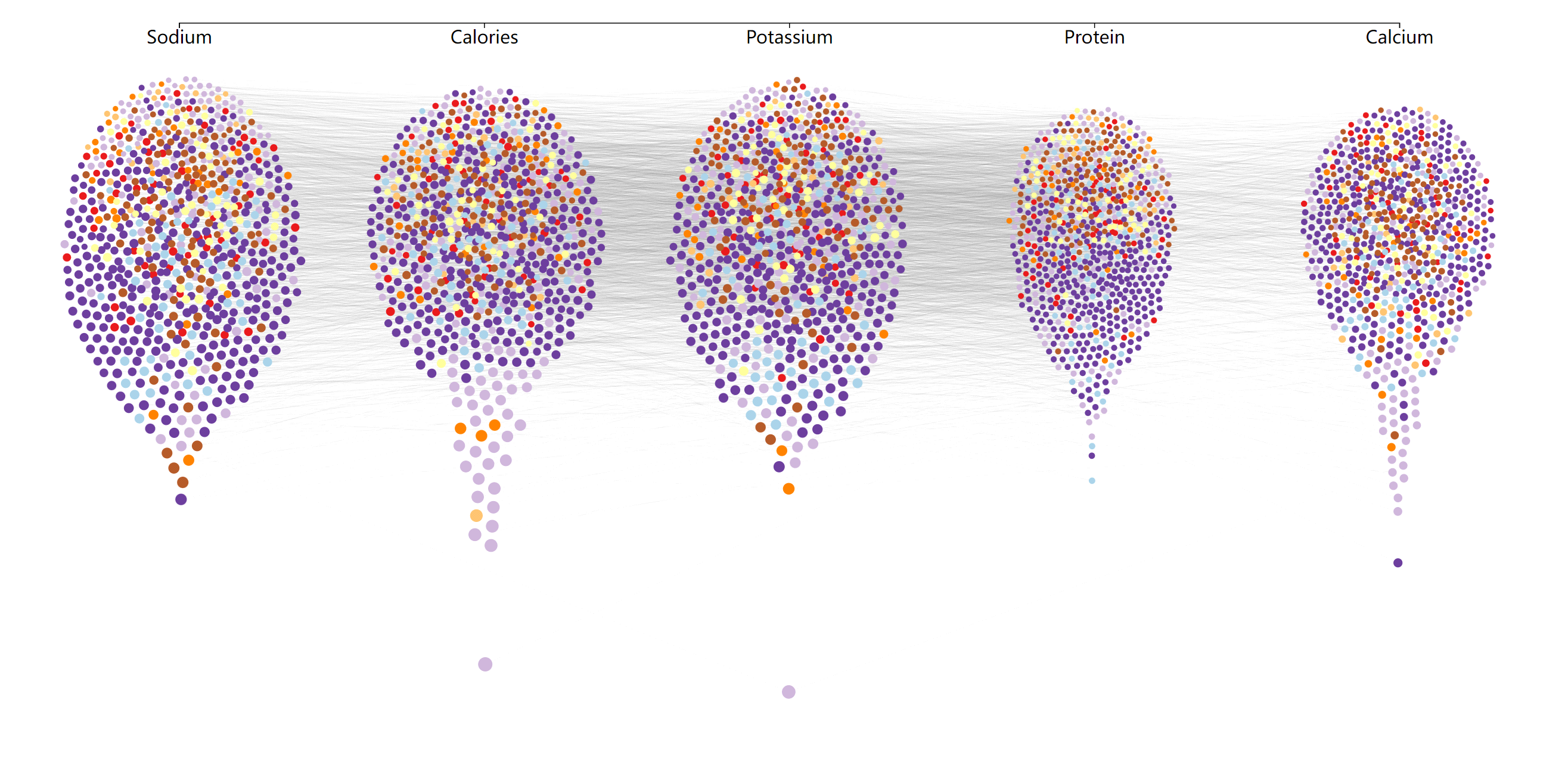
Regarding this beeswarm plot, the very first idea was to clean the JSON data into a dataframe whose observational unit is the value of a nutrient for a single dish, remove outliers, then plot the observations as points in the beeswarm plots; however, it turns out that the number of different nutrients was too large for a single beeswarm plot, whose optimal clusters were 3-5 in practice, so the desicion was to count the number of dishes with each category of nutrient, find the 5 most frequent nutrients, then take the subset of these nutrients.

Till this point, the plot looks like:



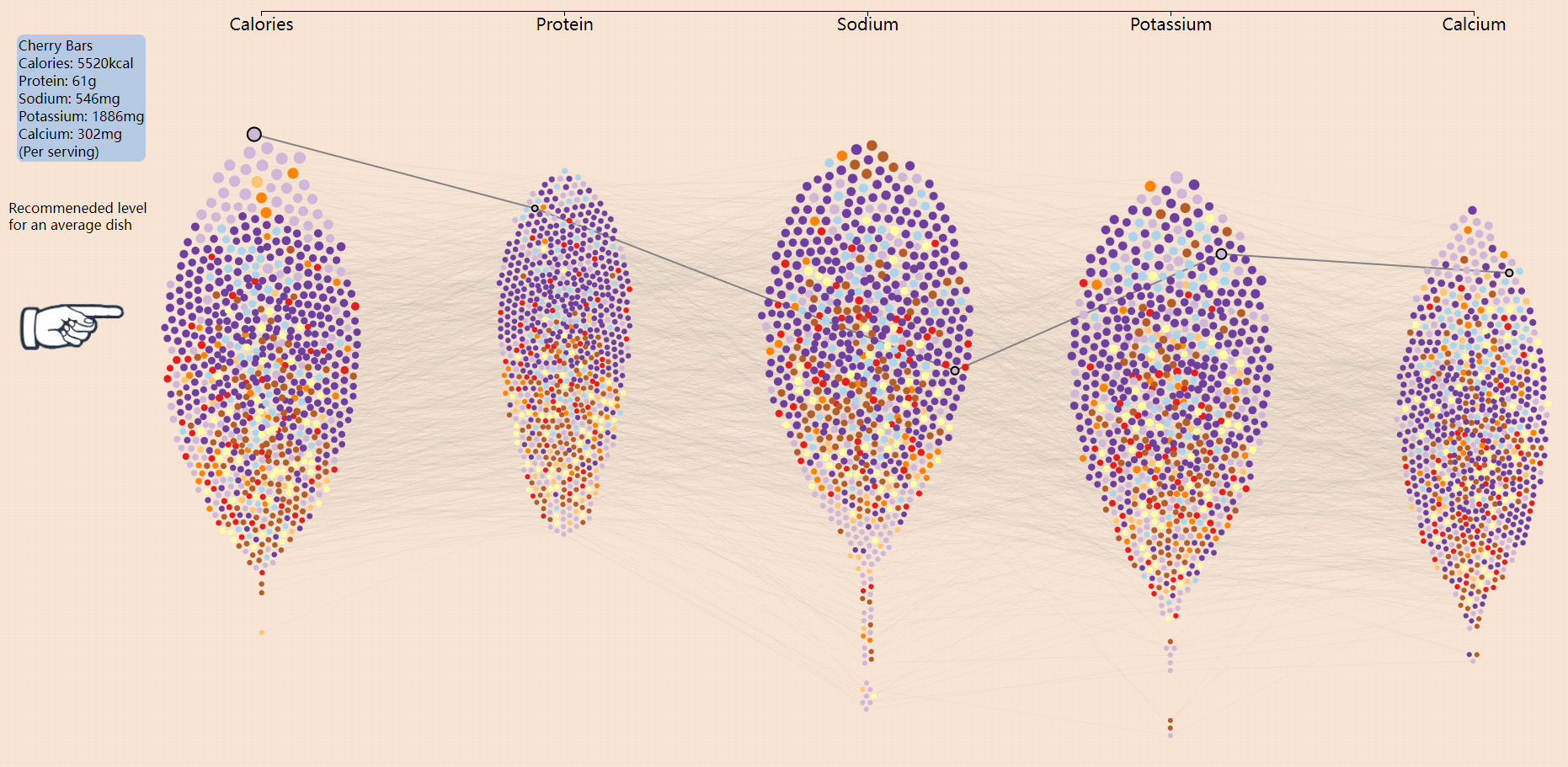
Soon another design problem was found: nutritional value of a dish is dissected into potentially 5 points! In order to show the integrity of a single dish, it would be better to draw edges between nodes representing the same dish, but not all dishes there in the dataset have possessed all the 5 nutrients to make all the 4 intra-cluster edges! Here, the decision was to filter the dishes again to include only those with the 5 nutrients, and connect nodes representing the same dish.

Till this point, the plot looks like:



Now, before adding tooltips and annotations, it turns out that by the current setting of Y-axis transformation the distributions of 5 nutrients were all somewhat similar, in a balloon shape and with similar range. In the original data, these categories all have numeric values in different range, respect to certain nutrients. Here, since the overall goal is to analyze the nutritional pattern of these foods, we decided to scale the value of nutrients by their recommended daily amount. Here we decided to set the normal level of a dish as 1/5 of daily recommended amount, log-transform the data to make clusters more connected, and add the annotation of the recommended level at the left side of the plot. We also provided a multi-functional tooltip reacting to both mouseover and mouse click; when users hover their mouse over a node, all the nodes and edges referring to the same dish would be highlighted, the name of dish would be displayed by the node; if users get interested on a certain dish, they could click on any node referring to that dish to apply an even more highlighted effect on the connected nodes/edges, plus the exact nutritional information of the dish displayed at the top left corner.

Till this point, the plot looks like:



This is the current stage of our beeswarm plot. During the development of the whole project, this plot could help us identify several critical insights about the nutritional pattern of these dishes. First of all, the nutritions display a strong pattern between different types of dishes(appetizers, dessert, etc.); also, even though the beeswarm plot has somewhat distorted the exact location of any given point, the recommended level represented by the finger image still tells us that comparing to regular nutrients like calories and protein, these foods may have possesses less micro-nutrients like potassium and calcium. Although this plot does not help much in identifying the Japanese dish, it is still a powerful tool to visualize and browse the dataset, and more functions could be added into it later, like dish-searching, plot-legends, hyperlinks to actual recipe webpages, etc.