A. A description of the work done by each team member. (Consider this your final status update.)

Nicole (ntc27):

1. CSS styling
2. Pot animations
3. Interactive elements where the recipe's information is displayed when recipe name is clicked on
4. Macronutrient calculator
5. Graphics

Melody Li (ml867):

1. Created data to visual mapping elements (graphs)
2. Filtered data in javascript to remove problematic recipes
3. Implemented search function (implemented with the help of fuse.js and underscore.js)
4. Implemented on click for ingredients
5. Implemented back-end of updating HTML and graph elements

Janna Yu (qy28)

- Data parsing and filtering two datasets (formatted into a JSON file)
- Unit parsing.
- Performing the matching using python and adjusted the results
- Created new relational data files to perform user search.

B. Data:

- **Source:** The two raw datasets we used are: full_format_recipe.json and nutrition.csv.
    - The first dataset is taken from kaggle.com by the user HugoDarwood. The dataset contains around 20,000 recipes from Epicurious by rating, nutritional content, and categories. Link Here
    - The second dataset is taken from the USDA website. It contains the nutritional values and common household units of over 9,000 ingredients commonly found in the United States. The documentation and description could be found here.
    - For both datasets, we wanted the data to be as comprehensive as possible. The Epicurious dataset is chosen also because of our interest in applicability, since it's highly prestigious and is one of the biggest recipe datasets out there that

American home cooks actually use. The usda nutrition dataset was chosen because it was authoritative and has been updated over the years.

- **Variables:**
    - For the recipe dataset, each data point represents a recipe. Each recipe contains:
        - A. Essential information such as title, rating, categories and directions.
        - B. ingredients together with quantities, expressed in common kitchen units rather than metric units. E.g. “1 1/2 cups whipping cream,” “1 tablespoons canned chicken broth.”
    - For the nutrition dataset, each data point represents a common ingredient. The fields could be separated into two categories:
        - A. Nutritional values (per 100 grams). E.g. Fat, calories, iron, Vitamin A.
        - B. Gram amounts delineated by common kitchen units rather than metric units. E.g. For butter with salt, 1 cup = 227 grams.
- **Modifications to the datasets:**
    - For the recipe dataset, we first reduced the number of recipes. Many recipes contain excessively long lists of ingredients and/or irregular formatting, such as “13-16 oz / 5-6 teaspoon vegetable oil.” We ended up using only 5,000 recipes that are straightforward and short (fewer than 10 ingredients).

      We also parsed out the quantity, unit and name of each ingredient from each 1-sentence description, for the purpose of aggregating the gram amounts.

      Using a python unit conversion library, we were able to convert all units to grams. To keep things simple without distorting the data, we use the common measure of 15 g = 1tbsp, though the exact amount may vary, depending on the nature of each ingredient.

    - For the nutrition dataset, we only used nutritional values that appear on FDA nutrition labels for commercial foods. This amounts to 11 nutritional values that an ordinary health-conscious person would care about.

      An additional complication arises when the recipe dataset does not contain a quantifiable unit at all, e.g. “2 large carrots.” We solved this by taking the unit data from the household unit field in the nutritional data and use it as one serving. After browsing the data, we found that this is often accurate.
- **Integration of the two datasets:**
    - The overall idea is to show how the aggregate nutritional values contained in each recipe, and how each ingredient in the recipe contributes to the overall calorie/fat/sugar level of the recipe.

- One challenge we had to overcome is the difference between ingredients described in natural language in instructions and the standardized ingredient names in the nutrition database. The ingredient names are not coded or standardized, so to achieve as close a matching to the nutrition database as possible, we used a Fuzzywuzzy search to perform string matches and manually overrode some complications.

- The final step is to do a simple aggregation of nutritional values, weighted by the amount of each ingredient that appear in the recipe.

C. Data to Visual:
    a. The scales used in this project were rather straightforward, For each type of nutrition (calories, fat, etc.) in the nutritional breakdown, the scale (and corresponding axis) was calculated depending on the recipe that was selected. Sometimes, the sum of all of the ingredients (multiplied by a ratio to get per serving results) would exceed the original axis domain (about 1.2 times the recommended nutrition facts per day for a 2000 calorie diet), so we implemented axises in a way that would allow the domain in these axises to increase in such cases. This meant that sometimes the recommended daily value per serving would move around in a graph. As a result, we felt that it was necessary to demarcate the recommended daily value with purple lines so that users could gauge how much of their daily value a single serving of a recipe would take. All scales were linear.

    b. Since many recipes from epicurious also contained the calories per serving size, we totaled up the aggregate calories (from the ingredients in the recipe) and calculated a ratio that would normalize the aggregate nutrients for all other nutrient values to per serving segments.

    c. In order to facilitate mapping ingredients to the bars that their nutritional value corresponds to, each large blue bar that initially appears is made up of many smaller bars that correspond to each ingredient's contribution to each nutrient. That way, when clicked, the bars are already in the svg and are easily selectable/made visible.
    d. We also didn't think it was incredibly relevant for viewers to view exactly how much of each nutrient each ingredient contained. The overall point was to be able to compare contribution across multiple ingredients (hence being able to highlight

multiple ingredient sections) and to the aggregate nutritional value and recommended nutritional intake..

D. Story:

a. The visualization is unusual because as meal-planning, special diets and vegetarianism are on the rise, comprehensive nutritional data for the recipes everyone follows are non-existent for many big recipe-sharing websites. As a result, even if you religiously follow each recipe, it is very unclear how much calorie intake you would get, let alone all the other macro- and micro-nutrients people might care about. USDA does have a website where you can create your own recipe by searching for and saving a list of ingredients. However, to use the website, you'll have to know the portions, the sizes and the exact name of the ingredient, which is beyond what an ordinary person would be willing to do. We made this visualization so that people can browse the nutrition data as they contemplate their next dinner — no need to go out of their way.

b. The visualization is designed to be interactive. The user often check out recipe websites when they have some food in their fridge/pantry and they feel like making a dish out of it. This is how they would start exploring the visualization — by searching for recipes by the ingredient at hand. Then, as they click through each available recipe, they would have the nutritional data to contemplate, "If I cook this dish, would it satisfy my day's calorie intake?" By further exploring each ingredient, they can also see the breakdown by ingredients and: 1. find out if there are "star ingredients" that will enrich their diet without incurring a lot of calorie costs; 2. conversely, they might also choose to omit one ingredient in their cooking if they find out that it alone contributes to all of the sugar intake.

c. One thing we found surprising is, a lot of the desert and cocktail recipes actually contain a lot of calories and sugar. Indeed this shouldn't be surprising, since each recipe could contain more than 4 servings, but this is not immediately obvious from the recipe descriptions until you see the visualization.