

# Proposal

## High-level goal

Our overall goal is to build an interactive Shiny dashboard that follows the molecular DNA of 20 world cuisines. We want to create an interactive dashboard to let users explore a global leaflet map to flavors, phylogenetic family trees, ingredient networks, and a live flavor explorer to show relationships geographically between food cultures.

## Description of your goals

For this project we were interested in exploring food across cuisines because of our backgrounds. For example some questvisualziations to let users navigate a wolrd map of cuisines through phlogenetific if they don;t utilize the same ingredients? Are cuisines across different geographies like French and Japanese more similar compared to close geographies like French and Italian? We think that the chemical compounds in food that make up flavors can help us answer these questions.

For background, a paper by Ahn et al in 2011 showed that every cuisine has a molecular fingerprint and that culinary traditions creates clusters that are different from what is expected with geographical distances in cuisines. This project builds on that idea to create the an interactive visual of culinary genomics. Our goal is to make this science explorable and intuitive. We want to create visualziations to let users navigate a wolrd map of cuisines through phlogenetific family trees, compare different flavors, and look at ingredient networks that intersect, and a live molecule explorer to show matches across different culture foods.

Our project draws on three datasets. The **Yummly “What’s Cooking” dataset** (Kaggle) contains data on recipies across different cuisine categories with ingredient lists which we will use as our baseline. The **Ahn et al. 2011 Flavor Network supplementary data** dataset maps ingredients to their flavor compounds aka how many flavor moelcules are in the food. We will use **CulinaryDB** database which has a database of ingredients with flavor moelcule profiles to fill in the gaps from the second dataset.

These three datasets allow us to create a cuisine & flavor matrix. We will use ggtree for the phylogenetic tree, leaflet for the map geographic visualization, and visNetwork to create the interactive ingredients network visualization.

## Yummly What’s Cooking

```
glimpse(recipes)
```

```
Rows: 39,774  
Columns: 3
```

```
$ id      <int> 10259, 25693, 20130, 22213, 13162, 6602, 42779, 3735,
1690...
$ cuisine <chr> "greek", "southern_us", "filipino", "indian", "indian",
"j...
$ ingredients <list> <"romaine lettuce", "black olives", "grape tomatoes",
"ga...
```

```
recipes %>%
  count(cuisine, sort = TRUE) %>%
  kable(col.names = c("Cuisine", "Number of Recipes"))
```

Cuisine	Number of Recipes
italian	7838
mexican	6438
southern_us	4320
indian	3003
chinese	2673
french	2646
cajun_creole	1546
thai	1539
japanese	1423
greek	1175
spanish	989
korean	830
vietnamese	825
moroccan	821
british	804
filipino	755
irish	667
jamaican	526
russian	489
brazilian	467

---

## Ahn et al. Ingredients

```
glimpse(ingredients)
```

```
Rows: 1,530
Columns: 3
$ `# id`          <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,
15...
$ `ingredient name` <chr> "magnolia_tripetala", "calyptranthes_parriculata",
"..."
$ category        <chr> "flower", "plant", "plant derivative", "fish/
seafood..."
```

```
ingredients %>%
  count(category, sort = TRUE) %>%
  kable(col.names = c("Category", "Count"))
```

Category	Count
plant derivative	424
plant	313
fruit	186
vegetable	104
herb	90
flower	66
meat	57
fish/seafood	56
spice	55
alcoholic beverage	50
cereal/crop	39
dairy	39
nut/seed/pulse	33
animal product	18

### Ahn et al. Flavor Compounds

```
glimpse(flavor_compounds)
```

```
Rows: 1,107
Columns: 3
```

```
$ `# id`          <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,
...
$ `Compound name` <chr> "jasmane", "5-methylhexanoic_acid", "l-glutamine",
"1-...
$ `CAS number`   <chr> "488-10-8", "628-46-6", "56-85-9", "1076-56-8",
"103-2...
```

## Ahn et al. Backbone Network

```
glimpse(backbone)
```

```
Rows: 1,470
Columns: 5
$ `0`          <chr> "orange_peel", "orange", "orange", "orange", "orange",
"ora...
$ `1`          <chr> "orange", "orange_peel", "orange_juice", "citrus",
"citrus_...
$ `2`          <dbl> 14, 14, 54, 51, 40, 19, 9, 69, 30, 57, 8, 67, 56, 23, 10,
6...
$ category     <chr> "plant", "fruit", "fruit", "fruit", "fruit", "fruit",
"frui...
$ prevalence   <dbl> 0.02730180, 0.08157129, 0.08157129, 0.08157129,
0.08157129,...
```

```
backbone %>%
  count(category, sort = TRUE) %>%
  kable(col.names = c("Category", "Count"))
```

Category	Count
dairy	335
fruit	271
alcoholic beverage	199
fish/seafood	133
vegetable	126
meat	123
plant derivative	97
spice	63
flower	44
nut/seed/pulse	24

Category	Count
herb	21
plant	16
cereal/crop	15
animal product	3

## CulinaryDB Recipes

```
glimpse(culinary_recipes)
```

```
Rows: 45,772
Columns: 4
$ `Recipe ID` <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17,...
$ Title      <chr> "5 spice vegetable fried rice", "aachar aaloo", "aadu
lass...
$ Source     <chr> "TARLA_DALAL", "TARLA_DALAL", "TARLA_DALAL",
"TARLA_DALAL"...
$ Cuisine    <chr> "Indian Subcontinent", "Indian Subcontinent", "Indian
Subc...
```

```
culinary_recipes %>%
  count(Cuisine, sort = TRUE) %>%
  kable(col.names = c("Cuisine", "Number of Recipes"))
```

Cuisine	Number of Recipes
USA	16118
Italy	7504
Indian Subcontinent	4058
Mexico	3138
France	2703
Canada	1112
Caribbean	1103
British Isles	1075
Middle East	993
China	941

Cuisine	Number of Recipes
Greece	934
Spain	816
Thailand	667
Africa	651
South East Asia	611
Japan	580
Eastern Europe	565
Australia & NZ	494
DACH Countries	487
Scandinavia	404
South America	310
Korea	301
Misc.: Portugal	138
Misc.: Dutch	40
Misc.: Belgian	15
Misc.: Central America	14

## CulinaryDB Ingredients

```
glimpse(culinary_ingredients)
```

```
Rows: 930
Columns: 4
$ `Aliased Ingredient Name` <chr> "Egg", "Bread", "Rye Bread", "Wheaten
Bread"...
$ `Ingredient Synonyms` <chr> "egg", "bread; bun", "bread-rye", "bread-
whe...
$ `Entity ID` <dbl> 0, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
1...
$ Category <chr> "Meat", "Bakery", "Bakery", "Bakery",
"Baker..."
```

```
culinary_ingredients %>%
  count(Category, sort = TRUE) %>%
  kable(col.names = c("Category", "Count"))
```

Category	Count
Fruit	136
Fish	119
Vegetable	73
Dish	68
Plant	64
Herb	51
Meat	49
Beverage Alcoholic	45
Dairy	45
Essential Oil	42
Seafood	35
Bakery	30
Additive	28
Spice	26
Nuts & Seed	25
Cereal	23
Legume	23
Beverage	21
Fungus	12
Flower	9
Maize	6

### CulinaryDB Compound Ingredients

```
glimpse(culinary_compounds)
```

```
Rows: 103
Columns: 5
$ `Compound Ingredient Name` <chr> "Garam Masala", "Ginger Garlic Paste",
...
$ `Compound Ingredient Synonyms` <chr> "garam masala", "ginger garlic paste",
...
$ entity_id <dbl> 2000, 2001, 2002, 2003, 2004, 2005,
200...
```

```
$ `Contituent Ingredients`      <chr> "black pepper, mace, cinnamon, clove,
C...
$ Category                      <chr> "Spice", "Spice", "Spice", "Spice",
"Sp..."
```

```
culinary_compounds %>%
  count(Category, sort = TRUE) %>%
  kable(col.names = c("Category", "Count"))
```

Category	Count
Dish	30
Spice	25
Fish	19
Bakery	6
Meat	5
Cereal	4
Vegetable	4
Fruit	3
Seafood	3
Additive	2
Beverage Alcoholic	1
Dairy	1

### CulinaryDB Recipe-Ingredient Aliases

```
glimpse(culinary_recipe_ing)
```

```
Rows: 456,279
Columns: 4
$ `Recipe ID`                  <dbl> 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3,
3...
$ `Original Ingredient Name`  <chr> "capsicum", "green bell pepper", "soy
sauce...
$ `Aliased Ingredient Name`  <chr> "capsicum", "pepper bell", "soy sauce",
"su...
$ `Entity ID`                <dbl> 362, 362, 291, 426, 61, 332, 258, 2001,
317...
```

```
culinary_recipe_ingredient %>%
  slice_head(n = 10) %>%
  kable()
```

Recipe ID	Original Ingredient Name	Aliased Ingredient Name	Entity ID
1	capsicum	capsicum	362
1	green bell pepper	pepper bell	362
1	soy sauce	soy sauce	291
1	sunflower oil	sunflower	426
2	buttermilk	buttermilk	61
2	cumin	cumin	332
2	fenugreek	fenugreek	258
2	ginger garlic paste	ginger garlic paste	2001
2	black mustard seed oil	mustard oil	317
2	nigella seed	nigella seed	392

## Weekly Plan of Attack

Week	Tasks	By
1	Load Yummly data, build cuisine × ingredient frequency matrix	Cindy
1	Join CulinaryDB to Ahn et al. data, produce cuisine × flavor compound matrix	Reinesse
1	Set up GitHub repo, folder structure, and Shiny skeleton	Hannah
2	Build Tab 2: phylogenetic tree	Cindy
2	Build Tab 3: radar chart and Tab 5: flavor explorer	Reinesse
2	Build Tab 4: ingredient network	Hannah
3	Build Tab 1: Leaflet world map with reactive popups	Cindy
3	Wire Tabs 2 and 3 into Shiny	Reinesse
3	Wire Tabs 4 and 5 into Shiny, test cross-tab reactivity	Hannah
4	Polish UI, theming, and tab narrative text	Cindy
4	Add animations and reactivity	Reinesse & Hannah
5	Rehearse, finalize repo, submit	All

## Final repository organization

We will create the following folders/files:

Folders:

- app
  - – modules folder
  - – app.R
  - – server.R
  - ui.R
- presentation
- R
  - data\_prep.R
- data
  - dataset csv and jsons

Descriptions for each of these folders and files is in each respective README.md