Nutrition

Hesham Al Kayed

8/13/2019

1 Introduction

This project aims to build a model that can predict which food is `Raw` and which is `Processed` by examining the food nutrition values and use them as predictors. We will use Nutrition data from `U.S. DEPARTMENT OF AGRICULTURE` website.[1].

In `Section 2.1` we Build our data set and format the result as needed. In `Section 2.2` we explore our data to find the difference between `Raw` and `Processed` food, this exploration will be based on `food groups`, which will help us make sense of the data at hand. In `Section 3` we will build a tree based model and compare the performance of different models.

2 Exploratory data analysis

2.1 Data Structure

Downloaded data will contain a documentation file and data files. as mentioned in the documentation, data files are carrot '^' delimited. We will use and connect `NUT_DATA`, `NUTR_DEF`, `FOOD_DESC` and `FD_GROUP` tables as described in the documentation. then we will select needed fields and rename them as follow **ProdID = NDB_No , NutrVal = Nutr_Val, ShrtDesc = Shrt_Desc , FoodGroup = FdGrp_Desc**. (left of the equal sign is the new name)

Table 1: Data Set

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
01001	717	kcal	Energy	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	0	mg	Caffeine	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	0	mg	Theobromine	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	24	mg	Calcium, Ca	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	2	mg	Magnesium, Mg	BUTTER,WITH SALT	NA	Dairy and Egg Products
01001	24	mg	Phosphorus, P	BUTTER,WITH SALT	NA	Dairy and Egg Products

Table 2: Types and Missing Data

	ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
Type	character	double	character	character	character	character	character
NA	0	0	0	0	0	426413	0
Empty	0	0	0	0	0	0	0

Table 3: Unique Values

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
7793	17446	5	97	7790	139	25

As we can see in *Table.2* all variables are complete except for `ManufacName`, which as per downloaded documentation is only available if possible. Our data set have 7 variables and 468669 records. we can see from *Table.3* that we have 7793 products and 7790 descriptions which indicates a duplicated product ID.

[1] https://www.ars.usda.gov/ARSUserFiles/80400525/Data/SR-Legacy/SR-Leg_ASC.zip

Table 4: Duplicated IDs

ProdID	ShrtDesc
04657	OIL,INDUSTRIAL,PALM KERNEL (HYDROGENATED),CONFECTION FAT
04658	OIL,INDUSTRIAL,PALM KERNEL (HYDROGENATED),CONFECTION FAT
13351	BEEF,CHUCK,UNDER BLADE CNTR STEAK,BNLESS,DENVER CUT,LN,0" FA
13352	BEEF,CHUCK,UNDER BLADE CNTR STEAK,BNLESS,DENVER CUT,LN,0" FA
25000	POPCORN,OIL-POPPED,LOFAT
25001	POPCORN,OIL-POPPED,LOFAT

by removing duplicated IDs we get the below table.

Table 5: Unique Values

ProdID	NutrVal	Units	NutrDesc	ShrtDesc	ManufacName	FoodGroup
7790	17443	5	97	7790	139	25

2.2 Data Description

As we want to build a model to classify *raw foods* from *processed food* we will start by exploring the difference between raw and processed food. we will separate raw foods from processed foods by searching for **'RAW'**, in the short description field using the following Reg. expression `'(?<![:alpha:])RAW(?![:alpha:])'`, and we will consider every product in `Sausages and Luncheon Meats` as processed product and every product in `Spices and Herbs` as raw food, Available data have 6374 Processed products and 1416 raw foods.[2].

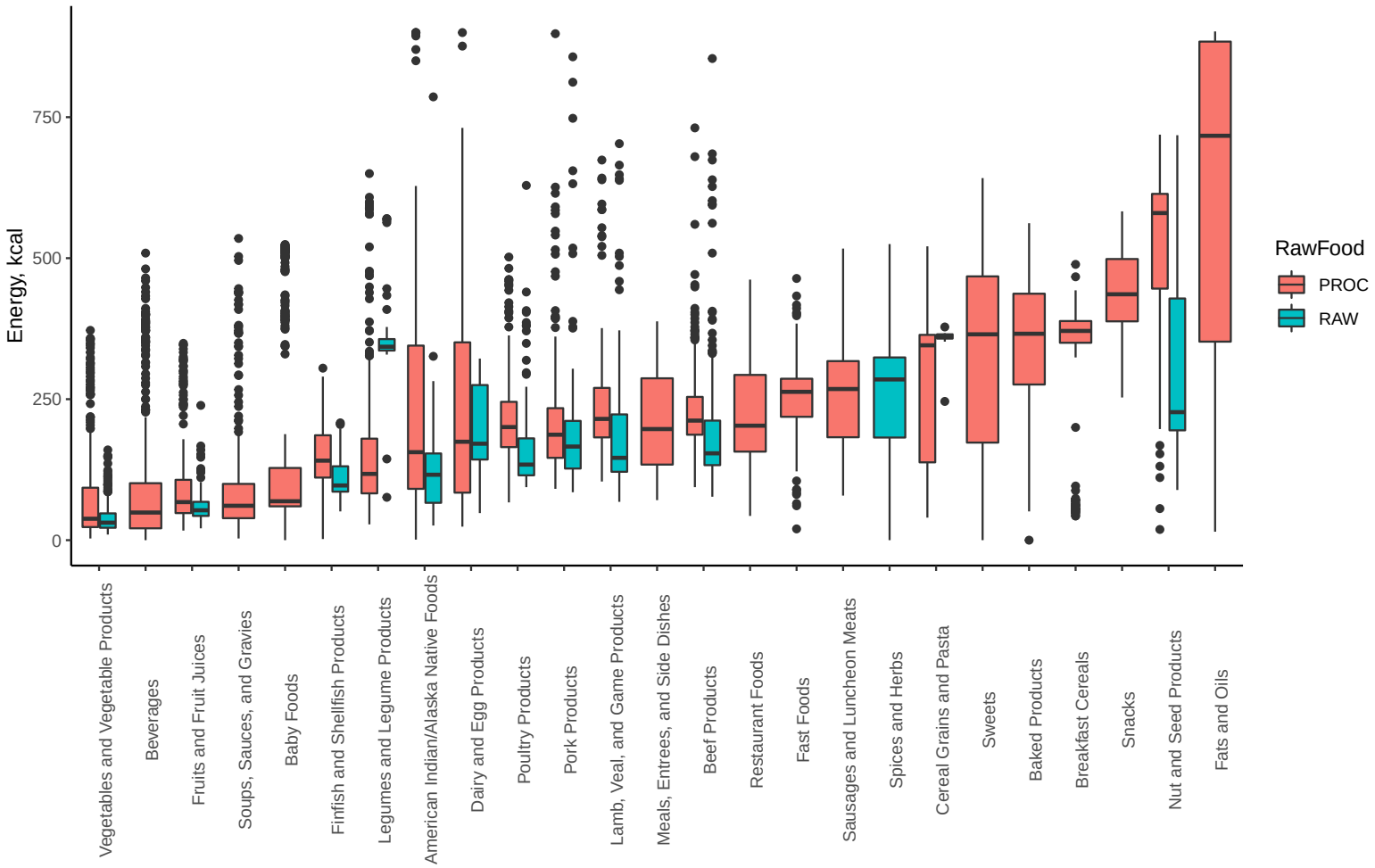

Figure 1: Comparing Raw and Processed Foods Energy Values

[2]We have to note that this method of differentiating between raw and processed food may not include all cases. but because we don't know how the data is constructed and their is no mention to this information in the documentation, we will assume that this method is correct

Fig.1 shows that Processed food on average have more kaloris per 100g. we can see that this does not apply to `Legumes and Legume Product`. *Fig.2* Shows the median value for each of the energy sources in food, we can see that most raw foods do not have sugar in them and processed food does, raw `Legumes and Legume Product` have more protein and more carbohydrate than processed `Legumes and Legume Product`, which can explain why raw `Legumes and Legume Product` has more energy than processed `Legumes and Legume Product` and other food groups don't.

Fiber content per 100g does not deffer much between processed and raw foods, except for `Legumes and Legume Products` as shown in *Fig.3*. we can see in *Fig.4* that all raw foods except `Legumes and Legume Product` have more water content.

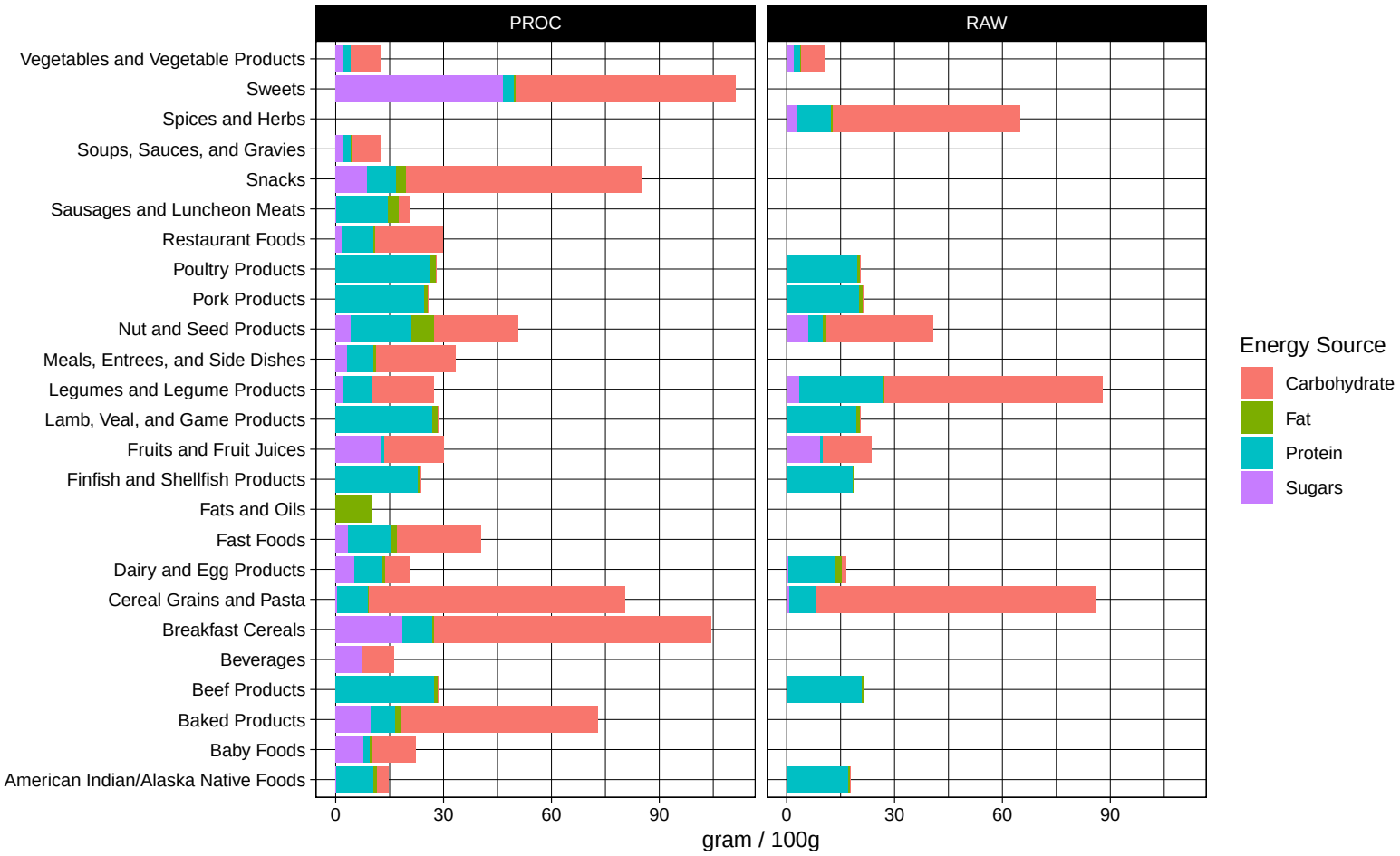

Figure 2: Energy Sources In Food

Fig.5 and *Fig.6* show the content of `Minerals` and `Vitamins` in each food group and enable us to compare raw and processed foods in that group. scale used is based on the **median value** for each group. Numerical values shows **q75, median and q25** (Upper, meddle and lower).

At first glance we can see that raw foods rarely contains any `Fluoride, F` and most processed foods contain `Fluoride, F`. `Sodium, Na` will range higher in processed products compared to raw foods, but again `Legumes and Legume Products` values are higher in its raw form compared to its processed form.

`Calcium, Ca` to my surprise scours the highest in `Spices and Herbs` in its raw form, with more than double the value compared to `Dairy and Egg Products` in its processed form. `Spices and Herbs` have the highest `Iron, Fe`, `Manganese, Mn` in all food groups, and high content value for other Minerals, although you cant eat much `Spices and Herbs` in one day, mixing them with food seems like a nutritious idea.

we can find another difference in `Vintamin C` as we can see that it rangers higher in raw foods, and its main source are plants sourced food, `Vegetables, Fruits, Nuts/Seeds and Spices/Herbs`.

above are a small part of our data. we can further explore the ratio of fat and carbohydrates different types in each food group and much more. but we cant explore every detail in this project because this is a much larger subject, and i don't have domain knowledge to do so.

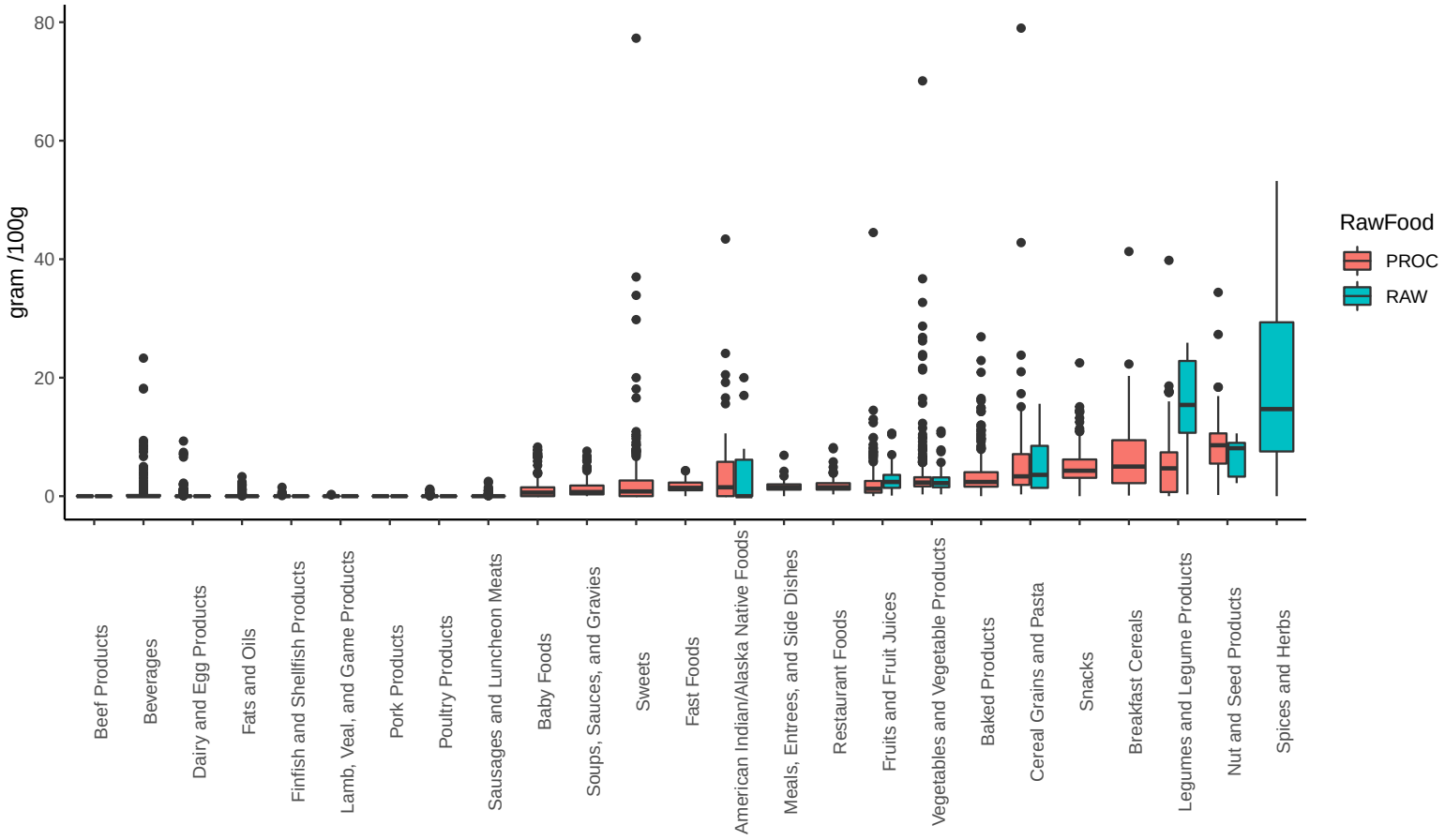

Figure 3: Fiber In Food

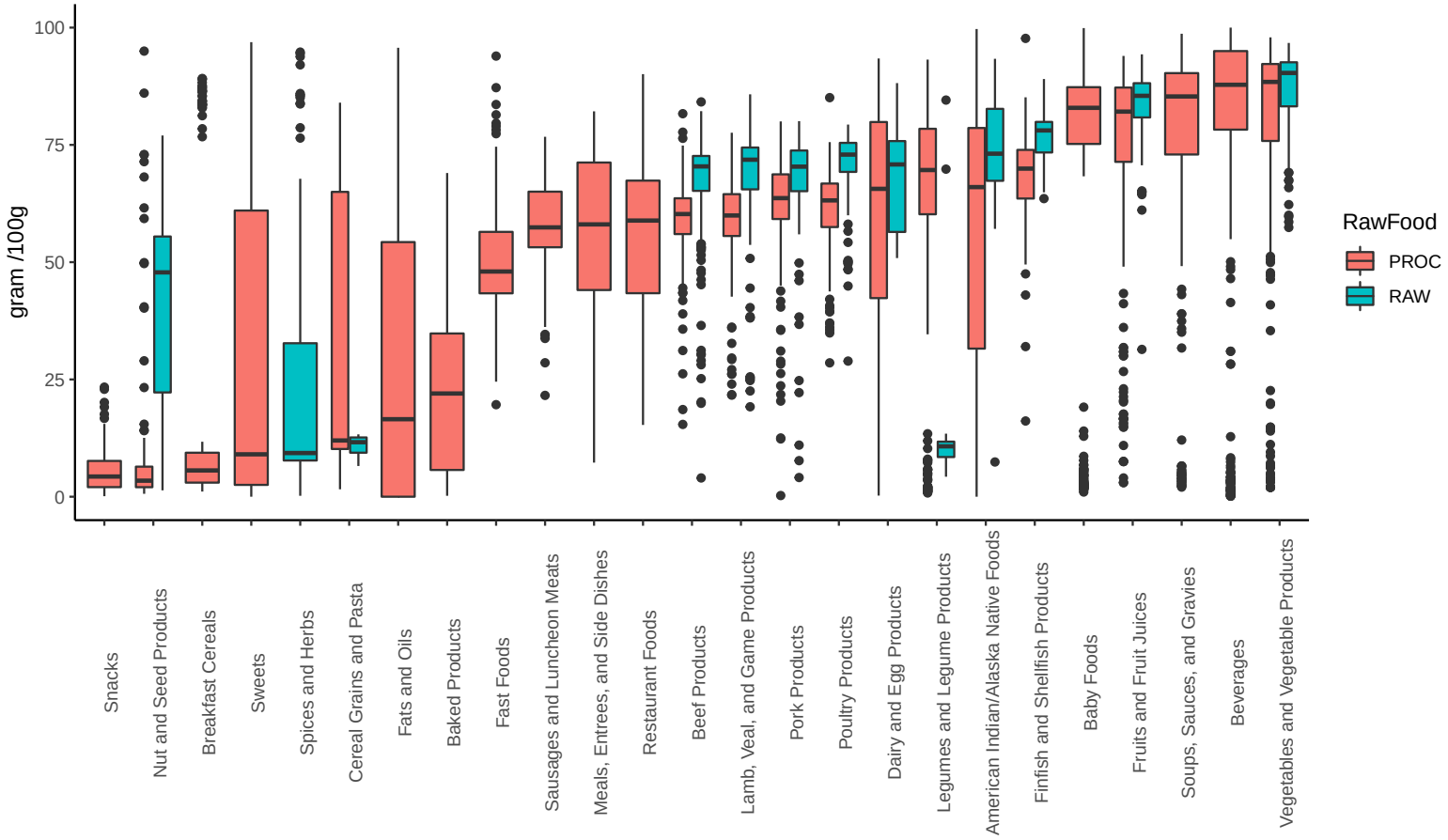

Figure 4: Water In Food

Figure 5: Minerals in Food

Figure 6: Vitamins in Food

3 Building Our Model

We want to build a model that can classify if a food product is `Raw` or `Processed` using its nutrition values. We will use tree based models, because they are easy to use and they will reveal more differences between raw and processed foods by finding variable importance in our model. we will split our data set based on `RawFood` field into training and test sets. any nutrients defined as `added` will be deleted. Below shows the structure of our training set.

we will compare *ctree* from *party* package, *gbm* from *gbm* package and *ranger* from *ranger* package. we will not use `FoodGroup` field to build our model, but we will keep it to compare the performance of our models through different food groups.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    3895 obs. of  97 variables:
##  $ FoodGroup                          : Factor w/ 25 levels "American Indian/Alaska Native Foods",..: 8 8 8 8 8
##  $ RawFood                            : Factor w/ 2 levels "PROC","RAW": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Alanine                            : num  0.029 0.67 0.659 0.409 0.639 0.823 0.958 0.691 0.741 0.757 ...
##  $ Alcohol, ethyl                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Arginine                           : num  0.031 0.874 0.883 0.467 0.47 ...
##  $ Ash                                : num  2.11 3.18 3.6 1.27 5.2 3.79 4.3 3.55 3.27 3.83 ...
##  $ Aspartic acid                      : num  0.064 1.588 1.502 0.963 0.779 ...
##  $ Beta-sitosterol                    : num  4 0 0 0 0 0 0 0 0 0 ...
##  $ Betaine                            : num  0.3 0 0 0.6 0 0 0 0 0 0.7 ...
##  $ Caffeine                           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Calcium, Ca                        : num  24 674 643 111 493 ...
##  $ Campesterol                        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Carbohydrate, by difference        : num  0.06 2.79 4.78 4.76 3.88 1.55 0.36 0.68 2.77 5.58 ...
##  $ Carotene, alpha                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Carotene, beta                     : num  158 76 0 6 3 32 33 78 41 51 ...
##  $ Cholesterol                        : num  215 94 103 12 89 116 110 89 64 64 ...
##  $ Choline, total                     : num  18.8 15.4 0 16.3 15.4 15.4 15.4 15.4 15.4 14.2 ...
##  $ Copper, Cu                         : num  0 0.024 0.042 0.033 0.032 0.025 0.032 0.032 0.025 0.033 ...
##  $ Cryptoxanthin, beta                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Cystine                            : num  0.008 0.131 0.117 0.062 0.083 0.261 0.304 0.123 0.144 0.124 ...
##  $ Dihydrophylloquinone               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Energy                             : num  717 371 387 81 265 389 413 373 254 295 ...
##  $ Fatty acids, total monounsaturated: num  21.021 8.598 8.671 0.516 4.623 ...
##  $ Fatty acids, total polyunsaturated: num  3.043 0.784 0.87 0.083 0.591 ...
##  $ Fatty acids, total saturated       : num  51.37 18.76 19.48 1.24 13.3 ...
##  $ Fatty acids, total trans           : num  3.278 0 0 0.067 0 ...
##  $ Fatty acids, total trans-monoenoic: num  2.982 0 0 0.054 0 ...
##  $ Fatty acids, total trans-polyenoic: num  0.296 0 0 0.013 0 0 0 0 0 0.138 ...
##  $ Fiber, total dietary               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Fluoride, F                        : num  2.8 0 0 0 0 0 0 0 0 0 ...
##  $ Folate, DFE                        : num  3 20 18 8 32 6 10 18 9 27 ...
##  $ Folate, food                       : num  3 20 18 8 32 6 10 18 9 27 ...
##  $ Folate, total                      : num  3 20 18 8 32 6 10 18 9 27 ...
##  $ Folic acid                         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Fructose                           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Galactose                          : num  0 0 0 0.12 0 0 0 0 0 0.78 ...
##  $ Glucose (dextrose)                 : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Glutamic acid                      : num  0.178 5.515 5.718 2.446 2.421 ...
##  $ Glycine                            : num  0.018 0.437 0.403 0.209 0.097 0.457 0.533 0.422 0.464 0.551 ...
##  $ Histidine                          : num  0.023 0.823 0.821 0.306 0.397 ...
##  $ Hydroxyproline                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Iron, Fe                           : num  0.02 0.43 0.21 0.13 0.65 0.23 0.17 0.72 0.22 0.22 ...
##  $ Isoleucine                         : num  0.051 1.137 1.451 0.556 0.803 ...
##  $ Lactose                            : num  0 0 0 3.87 0 0 0 0 0 1.12 ...
##  $ Leucine                            : num  0.083 2.244 2.238 1.049 1.395 ...
##  $ Lutein + zeaxanthin                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Lycopene                           : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Lysine                             : num  0.067 2.124 1.945 0.878 1.219 ...
##  $ Magnesium, Mg                      : num  2 24 21 9 19 14 36 27 23 27 ...
##  $ Maltose                            : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Manganese, Mn                      : num  0 0.012 0.012 0.015 0.028 0.014 0.017 0.011 0.01 0.039 ...
##  $ Menaquinone-4                      : num  0 0 0 0 0 0 0 0 0 4.1 ...
##  $ Methionine                         : num  0.021 0.565 0.612 0.253 0.368 0.706 0.822 0.641 0.677 0.551 ...
```

```
##  $ Niacin                          : num  0.042 0.118 0.08 0.103 0.991 0.15 0.106 0.093 0.105 0.111 ...
##  $ Pantothenic acid                : num  0.11 0.288 0.413 0.524 0.967 0.429 0.562 0.21 0.079 0.429 ...
##  $ Phenylalanine                   : num  0.041 1.231 1.231 0.543 0.675 ...
##  $ Phosphorus, P                   : num  24 451 464 150 337 346 605 444 463 548 ...
##  $ Phytosterols                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Potassium, K                    : num  24 136 95 125 62 64 81 81 84 188 ...
##  $ Proline                         : num  0.082 2.575 2.634 1.155 1.378 ...
##  $ Protein                         : num  0.85 23.24 23.37 10.45 14.21 ...
##  $ Retinol                         : num  671 286 220 68 125 258 268 192 124 223 ...
##  $ Riboflavin                      : num  0.034 0.351 0.293 0.251 0.844 0.204 0.279 0.39 0.303 0.353 ...
##  $ Selenium, Se                    : num  1 14.5 14.5 11.9 15 14.5 14.5 14.5 14.4 27.6 ...
##  $ Serine                          : num  0.046 1.289 1.366 0.601 1.169 ...
##  $ Sodium, Na                      : num  643 560 700 308 1139 ...
##  $ Starch                          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Stigmasterol                    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Sucrose                         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Sugars, total                   : num  0.06 0.51 0 4 0 1.55 0.36 0.5 1.13 1.9 ...
##  $ Theobromine                     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Thiamin                         : num  0.005 0.014 0.046 0.02 0.154 0.021 0.06 0.015 0.018 0.024 ...
##  $ Threonine                       : num  0.038 0.882 0.832 0.47 0.637 ...
##  $ Tocopherol, beta                : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tocopherol, delta               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tocopherol, gamma               : num  0 0 0 0 0 0 0 0 0 0.03 ...
##  $ Tocotrienol, alpha              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tocotrienol, beta               : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tocotrienol, delta              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Tocotrienol, gamma              : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Total lipid (fat)               : num  81.11 29.68 30.6 2.27 21.49 ...
##  $ Tryptophan                      : num  0.012 0.324 0.3 0.138 0.2 0.361 0.421 0.315 0.339 0.551 ...
##  $ Tyrosine                        : num  0.041 1.115 1.128 0.568 0.668 ...
##  $ Valine                          : num  0.057 1.472 1.56 0.703 1.065 ...
##  $ Vitamin A, IU                   : num  2499 1080 985 236 422 ...
##  $ Vitamin A, RAE                  : num  684 292 233 68 125 261 271 198 127 227 ...
##  $ Vitamin B-12                    : num  0.17 1.26 0.83 0.47 1.69 1.68 1.6 0.83 0.82 1.68 ...
##  $ Vitamin B-6                     : num  0.003 0.065 0.074 0.057 0.424 0.083 0.081 0.079 0.07 0.1 ...
##  $ Vitamin C, total ascorbic acid  : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Vitamin D                       : num  0 22 0 0 16 23 24 22 12 15 ...
##  $ Vitamin D (D2 + D3)             : num  0 0.5 0 0 0.4 0.6 0.6 0.6 0.3 0.4 ...
##  $ Vitamin D2 (ergocalciferol)     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ Vitamin D3 (cholecalciferol)    : num  0 0.5 0 0 0.4 0.6 0.6 0.6 0.3 0.4 ...
##  $ Vitamin E (alpha-tocopherol)    : num  2.32 0.26 0 0.08 0.18 0.27 0.28 0.26 0.14 0.5 ...
##  $ Vitamin K (phylloquinone)       : num  7 2.5 0 0 1.8 2.6 2.7 2.5 1.6 1.3 ...
##  $ Water                           : num  16.2 41.1 37.6 81.2 55.2 ...
##  $ Zinc, Zn                        : num  0.09 2.6 2.79 0.51 2.88 3.5 3.9 3 2.76 3.62 ...
```

Table 6: Model Performance, Training Set

	ctree_Training	gbm_Training	ranger_Training	ctree_Test	gbm_Test	ranger_Test
Sensitivity	0.811	0.842	1.000	0.742	0.758	0.770
Specificity	0.968	0.978	1.000	0.947	0.970	0.975
Pos Pred Value	0.848	0.896	1.000	0.758	0.850	0.872
Neg Pred Value	0.958	0.965	1.000	0.943	0.948	0.950
Precision	0.848	0.896	1.000	0.758	0.850	0.872
Recall	0.811	0.842	1.000	0.742	0.758	0.770
F1	0.829	0.868	1.000	0.749	0.801	0.818
Prevalence	0.182	0.182	0.182	0.182	0.182	0.182
Detection Rate	0.147	0.153	0.182	0.135	0.138	0.140
Detection Prevalence	0.174	0.171	0.182	0.178	0.162	0.160
Balanced Accuracy	0.889	0.910	1.000	0.844	0.864	0.872

Table.6 shows the performance of each model, we can see that `ranger` has the highest accuracy of **1**, followed by `gbm`. a perfect accuracy may be a result of over fitting. applying our models to the test set we can see that accuracy of `ranger` falls to **.87** and `gbm` falls to **.86**. which indicates that `gbm` model is more consistent and not over fitted.

One more thing to notice that if we set `RAW` as our positive test result we get a `Prevalence` of **0.182**, thus the high `Specificity` relative to `Sensitivity`. trying to balance our data set by sampling up and down we get the results in *table.7*.

Table 7: Model Performance, Balanced Data Set

	up	gbm	down
Sensitivity	0.968	0.842	0.980
Specificity	0.930	0.978	0.901
Pos Pred Value	0.754	0.896	0.687
Neg Pred Value	0.992	0.965	0.995
Precision	0.754	0.896	0.687
Recall	0.968	0.842	0.980
F1	0.847	0.868	0.808
Prevalence	0.182	0.182	0.182
Detection Rate	0.176	0.153	0.178
Detection Prevalence	0.233	0.171	0.259
Balanced Accuracy	0.949	0.910	0.941

as we can see in *Table.8*, sampling up/down increased `Balanced Accuracy` and increased `Sensitivity` on the expense of `Specificity`, `F1` and `Precision`. this means that our model started to predict more `raw foods`, enough to correctly cover *0.97-0.98* of the `RAW` population but on the expense of predicting more `processed foods` as `raw foods`.

Below shows a breakdown for the performance of `gbm`, and indeed we see that sampling enhanced `Sensitivity`, but decreased `Specificity`. so in conclusion `gbm` is the more stable model and to chose to sample up, down or not to sample at all depends on the importance of predicting positive(raw) test results more than negative results. but in our case i would prefer a balanced model and will chose `gbm` without sampling.

Table 8: GBM, Model Performance Per Group

FoodGroup	Sensitivity	Specificity	Precision	Recall	F1	Prevalence	Balanced Accuracy
Dairy and Egg Products	0.71	1.00	1.00	0.71	0.83	0.05	0.86
Spices and Herbs	0.77	NA	1.00	0.77	0.87	1.00	NA
Baby Foods	NA	0.96	0.00	NA	NA	0.00	NA
Fats and Oils	NA	1.00	NA	NA	NA	0.00	NA
Poultry Products	0.92	0.98	0.96	0.92	0.94	0.32	0.95
Soups, Sauces, and Gravies	NA	1.00	NA	NA	NA	0.00	NA
Sausages and Luncheon Meats	NA	0.99	0.00	NA	NA	0.00	NA
Breakfast Cereals	NA	1.00	NA	NA	NA	0.00	NA
Snacks	NA	1.00	NA	NA	NA	0.00	NA
Fruits and Fruit Juices	0.67	0.99	0.98	0.67	0.80	0.34	0.83
Pork Products	0.81	0.99	0.97	0.81	0.89	0.26	0.90
Vegetables and Vegetable Products	0.67	0.93	0.77	0.67	0.72	0.25	0.80
Nut and Seed Products	0.33	1.00	1.00	0.33	0.50	0.14	0.67
Beef Products	0.99	0.94	0.92	0.99	0.95	0.41	0.96
Beverages	NA	1.00	NA	NA	NA	0.00	NA
Finfish and Shellfish Products	0.88	0.88	0.86	0.88	0.87	0.45	0.88
Legumes and Legume Products	0.92	0.99	0.96	0.92	0.94	0.16	0.95
Lamb, Veal, and Game Products	0.98	0.94	0.93	0.98	0.95	0.43	0.96
Baked Products	NA	1.00	NA	NA	NA	0.00	NA
Sweets	NA	1.00	NA	NA	NA	0.00	NA
Cereal Grains and Pasta	0.00	1.00	NA	0.00	NA	0.04	0.50
Fast Foods	NA	1.00	NA	NA	NA	0.00	NA
Meals, Entrees, and Side Dishes	NA	1.00	NA	NA	NA	0.00	NA
American Indian/Alaska Native Foods	0.38	0.98	0.89	0.38	0.53	0.24	0.68
Restaurant Foods	NA	1.00	NA	NA	NA	0.00	NA

Table 9: GBM, Model Performance Per Group, Sampling UP

FoodGroup	Sensitivity	Specificity	Precision	Recall	F1	Prevalence	Balanced Accuracy
Dairy and Egg Products	0.86	1.00	1.00	0.86	0.92	0.05	0.93
Spices and Herbs	0.91	NA	1.00	0.91	0.95	1.00	NA
Baby Foods	NA	0.94	0.00	NA	NA	0.00	NA
Fats and Oils	NA	0.99	0.00	NA	NA	0.00	NA
Poultry Products	1.00	0.91	0.85	1.00	0.92	0.32	0.96
Soups, Sauces, and Gravies	NA	0.99	0.00	NA	NA	0.00	NA
Sausages and Luncheon Meats	NA	0.99	0.00	NA	NA	0.00	NA
Breakfast Cereals	NA	0.99	0.00	NA	NA	0.00	NA
Snacks	NA	1.00	NA	NA	NA	0.00	NA
Fruits and Fruit Juices	0.95	0.91	0.85	0.95	0.90	0.34	0.93
Pork Products	1.00	0.94	0.84	1.00	0.91	0.26	0.97
Vegetables and Vegetable Products	0.97	0.76	0.58	0.97	0.72	0.25	0.87
Nut and Seed Products	1.00	0.91	0.64	1.00	0.78	0.14	0.95
Beef Products	0.99	0.90	0.87	0.99	0.93	0.41	0.94
Beverages	NA	0.99	0.00	NA	NA	0.00	NA
Finfish and Shellfish Products	0.94	0.75	0.76	0.94	0.84	0.45	0.85
Legumes and Legume Products	1.00	0.90	0.67	1.00	0.80	0.16	0.95
Lamb, Veal, and Game Products	0.99	0.85	0.83	0.99	0.90	0.43	0.92
Baked Products	NA	1.00	0.00	NA	NA	0.00	NA
Sweets	NA	1.00	NA	NA	NA	0.00	NA
Cereal Grains and Pasta	0.00	0.96	0.00	0.00	NaN	0.04	0.48
Fast Foods	NA	1.00	NA	NA	NA	0.00	NA
Meals, Entrees, and Side Dishes	NA	1.00	NA	NA	NA	0.00	NA
American Indian/Alaska Native Foods	0.81	0.80	0.57	0.81	0.67	0.24	0.80
Restaurant Foods	NA	1.00	NA	NA	NA	0.00	NA

4 Conclusion

Data exploration shows that there is a clear difference between raw and processed food if we consider food as groups, such as the differences in `Fluoride, F`, `Sodium, Na` and `Vitamin C` values between raw and processed food. Our model shows that we can predict raw/processed food with a `Balanced Accuracy` of **0.91**, *Fig.7* shows the most important differences between raw and processed foods, again we can see that `Sodium, Na` and `Vitamin C` are important differences between raw and processed foods.

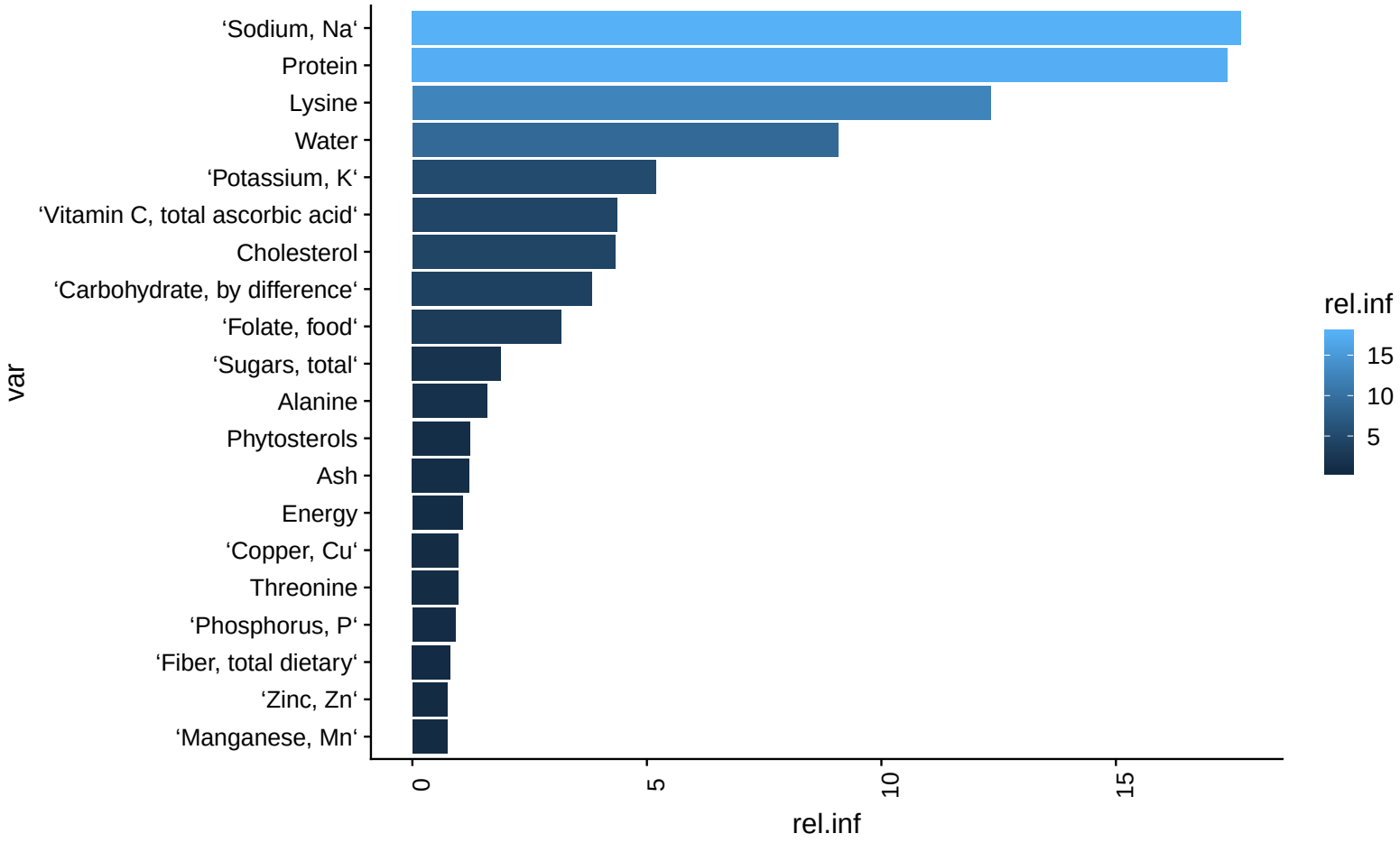

Figure 7: Var Importance