

# Pstat 131 Final Project

Yu Tian

Spring 2022-04-10

## Data Demo

```
# Read the full data set into R using read_csv()
fat_supply_quantity <- read.csv(file = 'Fat_Supply_Quantity_Data.csv')
fat_supply_quantity %>%
  head()
```

	Country	Alcoholic.Beverages	Animal.Products	Animal.fats		
## 1	Afghanistan	0	21.6397	6.2224		
## 2	Albania	0	32.0002	3.4172		
## 3	Algeria	0	14.4175	0.8972		
## 4	Angola	0	15.3041	1.3130		
## 5	Antigua and Barbuda	0	27.7033	4.6686		
## 6	Argentina	0	30.3572	3.3076		
	Aquatic.Products..Other	Cereals...Excluding.Beer	Eggs	Fish..Seafood		
## 1	0	8.0353	0.6859	0.0327		
## 2	0	2.6734	1.6448	0.1445		
## 3	0	4.2035	1.2171	0.2008		
## 4	0	6.5545	0.1539	1.4155		
## 5	0	3.2153	0.3872	1.5263		
## 6	0	1.3316	1.5706	0.1664		
	Fruits...Excluding.Wine	Meat	Miscellaneous Milk...Excluding.Butter	Offals		
## 1	0.4246	6.1244	0.0163	8.2803 0.3103		
## 2	0.6418	8.7428	0.0170	17.7576 0.2933		
## 3	0.5772	3.8961	0.0439	8.0934 0.1067		
## 4	0.3488	11.0268	0.0308	1.2309 0.1539		
## 5	1.2177	14.3202	0.0898	6.6607 0.1347		
## 6	0.2091	19.2693	0.0000	5.8512 0.1878		
	Oilcrops	Pulses	Spices	Starchy.Roots	Stimulants	Sugar.Crops
## 1	1.0452	0.1960	0.2776	0.0490	0.0980	0
## 2	3.1622	0.1148	0.0000	0.0510	0.5270	0
## 3	1.1983	0.2698	0.1568	0.1129	0.2886	0
## 4	3.9902	0.3282	0.0103	0.7078	0.1128	0
## 5	1.3579	0.0673	0.3591	0.0449	1.0549	0
## 6	0.0640	0.0213	0.0213	0.1110	0.2475	0
	Sugar...Sweeteners	Treenuts	Vegetal.Products	Vegetable.Oils	Vegetables	
## 1	0	0.7513	28.3684	17.0831	0.3593	
## 2	0	0.9181	17.9998	9.2443	0.6503	
## 3	0	0.8595	35.5857	27.3606	0.5145	
## 4	0	0.0308	34.7010	22.4638	0.1231	
## 5	0	0.2020	22.2995	14.4436	0.2469	
## 6	0	0.1366	19.6449	17.3147	0.1878	

```
## Obesity Undernourished Confirmed Deaths Recovered Active
## 1 4.5 29.8 0.14213420 0.006185779 0.1233739 0.012574497
## 2 22.3 6.2 2.96730092 0.050951374 1.7926357 1.123713883
## 3 26.6 3.9 0.24489709 0.006558153 0.1675722 0.070766734
## 4 6.8 25 0.06168747 0.001460550 0.0568077 0.003419224
## 5 19.1 <NA> 0.29387755 0.007142857 0.1908163 0.095918367
## 6 28.5 4.6 4.35614739 0.108226635 3.9051921 0.342728695
## Population Unit..all.except.Population.
## 1 38928000 %
## 2 2838000 %
## 3 44357000 %
## 4 32522000 %
## 5 98000 %
## 6 45377000 %
```

```
protein_supply_quantity <- read.csv(file = 'Protein_Supply_Quantity_Data.csv')
protein_supply_quantity %>%
  head()
```

```
## Country Alcoholic.Beverages Animal.Products Animal.fats
## 1 Afghanistan 0.0000 9.7523 0.0277
## 2 Albania 0.1840 27.7469 0.0711
## 3 Algeria 0.0323 13.8360 0.0054
## 4 Angola 0.6285 15.2311 0.0277
## 5 Antigua and Barbuda 0.1535 33.1901 0.1289
## 6 Argentina 0.1704 31.9799 0.0097
## Aquatic.Products..Other Cereals...Excluding.Beer Eggs Fish..Seafood
## 1 0 35.9771 0.4067 0.0647
## 2 0 14.2331 1.8069 0.6274
## 3 0 26.5633 1.2916 0.6350
## 4 0 20.3882 0.1756 5.4436
## 5 0 10.5108 0.4850 8.2146
## 6 0 13.6702 2.0593 1.0223
## Fruits...Excluding.Wine Meat Milk...Excluding.Butter Offals Oilcrops
## 1 0.5824 3.1337 5.5278 0.5916 0.2034
## 2 1.2757 7.6582 16.4750 1.1084 0.3722
## 3 1.1624 3.5088 8.0616 0.3283 0.1830
## 4 1.2754 7.6248 1.1460 0.8133 2.1534
## 5 1.2586 16.0670 7.4349 0.8534 0.7674
## 6 0.5209 21.6250 5.8322 1.4313 0.0097
## Pulses Spices Starchy.Roots Stimulants Sugar.Crops Sugar...Sweeteners
## 1 1.2479 0.1664 0.1941 0.5546 0 0.0000
## 2 1.4555 0.0000 0.8867 0.2635 0 0.0042
## 3 2.5509 0.1776 1.4638 0.4628 0 0.0000
## 4 4.0850 0.0000 5.1941 0.1017 0 0.0092
## 5 0.8841 0.3438 0.4666 0.4113 0 0.0000
## 6 0.2434 0.0292 1.3096 0.9688 0 0.0049
## Treenuts Vegetal.Products Vegetable.Oils Vegetables Miscellaneous Obesity
## 1 0.1387 40.2477 0.0000 1.1370 0.0462 4.5
## 2 0.2677 22.2552 0.0084 3.2456 0.0544 22.3
## 3 0.2745 36.1694 0.0269 3.1267 0.1399 26.6
## 4 0.0092 34.7782 0.0092 0.8133 0.0924 6.8
## 5 0.0737 16.8161 0.0430 1.6024 0.2947 19.1
## 6 0.0438 18.0176 0.0000 1.0516 0.0000 28.5
## Undernourished Confirmed Deaths Recovered Active Population
```

```
## 1      29.8 0.14213420 0.006185779 0.1233739 0.012574497 38928000
## 2      6.2 2.96730092 0.050951374 1.7926357 1.123713883 2838000
## 3      3.9 0.24489709 0.006558153 0.1675722 0.070766734 44357000
## 4      25 0.06168747 0.001460550 0.0568077 0.003419224 32522000
## 5      <NA> 0.29387755 0.007142857 0.1908163 0.095918367 98000
## 6      4.6 4.35614739 0.108226635 3.9051921 0.342728695 45377000
## Unit..all.except.Population.
## 1      %
## 2      %
## 3      %
## 4      %
## 5      %
## 6      %
```

```
food_supply_quantity_kg <- read.csv(file = 'Food_Supply_Quantity_kg_Data.csv')
food_supply_quantity_kg %>%
  head()
```

```
##          Country Alcoholic.Beverages Animal.fats Animal.Products
## 1      Afghanistan      0.0014      0.1973      9.4341
## 2      Albania      1.6719      0.1357      18.7684
## 3      Algeria      0.2711      0.0282      9.6334
## 4      Angola      5.8087      0.0560      4.9278
## 5 Antigua and Barbuda      3.5764      0.0087      16.6613
## 6      Argentina      4.2672      0.2234      19.3454
## Aquatic.Products..Other Cereals...Excluding.Beer Eggs Fish..Seafood
## 1      0      24.8097 0.2099      0.0350
## 2      0      5.7817 0.5815      0.2126
## 3      0      13.6816 0.5277      0.2416
## 4      0      9.1085 0.0587      1.7707
## 5      0      5.9960 0.2274      4.1489
## 6      0      8.4102 0.9979      0.4693
## Fruits...Excluding.Wine Meat Milk...Excluding.Butter Miscellaneous Offals
## 1      5.3495 1.2020      7.5828      0.0728 0.2057
## 2      6.7861 1.8845      15.7213      0.1123 0.2324
## 3      6.3801 1.1305      7.6189      0.1671 0.0870
## 4      6.0005 2.0571      0.8311      0.1165 0.1550
## 5      10.7451 5.6888      6.3663      0.7139 0.2219
## 6      6.0435 7.0421      10.2328      0.0000 0.3779
## Oilcrops Pulses Spices Starchy.Roots Stimulants Sugar...Sweeteners
## 1      0.0700 0.2953 0.0574      0.8802      0.3078      1.3489
## 2      0.9377 0.2380 0.0008      1.8096      0.1055      1.5367
## 3      0.3493 0.4783 0.0557      4.1340      0.2216      1.8342
## 4      0.4186 0.6507 0.0009      18.1102      0.0508      1.8495
## 5      0.2172 0.1840 0.1524      1.4522      0.1564      3.8749
## 6      0.0116 0.0528 0.0122      3.0420      0.4378      3.0536
## Sugar.Crops Treenuts Vegetable.Oils Vegetables Vegetal.Products Obesity
## 1      0      0.0770      0.5345      6.7642      40.5645      4.5
## 2      0      0.1515      0.3261      11.7753      31.2304      22.3
## 3      0      0.1152      1.0310      11.6484      40.3651      26.6
## 4      0      0.0061      0.6463      2.3041      45.0722      6.8
## 5      0      0.0253      0.8102      5.4495      33.3233      19.1
## 6      0      0.0200      0.9541      4.3503      30.6559      28.5
## Undernourished Confirmed Deaths Recovered Active Population
## 1      29.8 0.14213420 0.006185779 0.1233739 0.012574497 38928000
```

```
## 2      6.2 2.96730092 0.050951374 1.7926357 1.123713883 2838000
## 3      3.9 0.24489709 0.006558153 0.1675722 0.070766734 44357000
## 4      25 0.06168747 0.001460550 0.0568077 0.003419224 32522000
## 5      <NA> 0.29387755 0.007142857 0.1908163 0.095918367 98000
## 6      4.6 4.35614739 0.108226635 3.9051921 0.342728695 45377000
## Unit..all.except.Population.
## 1      %
## 2      %
## 3      %
## 4      %
## 5      %
## 6      %
```

```
food_supply_kcal <- read.csv(file = 'Food_Supply_kcal_Data.csv')
food_supply_kcal %>%
  head()
```

```
##      Country Alcoholic.Beverages Animal.Products Animal.fats
## 1      Afghanistan      0.0000      4.7774      0.8504
## 2      Albania      0.9120      16.0930      1.0591
## 3      Algeria      0.0896      6.0326      0.1941
## 4      Angola      1.9388      4.6927      0.2644
## 5      Antigua and Barbuda      2.3041      15.3672      1.5429
## 6      Argentina      1.4354      14.9869      1.0650
## Aquatic.Products..Other Cereals...Excluding.Beer Eggs Fish..Seafood
## 1      0      37.1186 0.1501      0.0000
## 2      0      16.2107 0.8091      0.1471
## 3      0      25.0112 0.4181      0.1195
## 4      0      18.3521 0.0441      0.8372
## 5      0      13.7215 0.2057      1.7280
## 6      0      16.7927 0.8643      0.2006
## Fruits...Excluding.Wine Meat Milk...Excluding.Butter Miscellaneous Offals
## 1      1.4757 1.2006      2.4512      0.0250 0.1251
## 2      3.8982 3.8688      9.9441      0.0588 0.2648
## 3      3.1805 1.2543      3.9869      0.1045 0.0597
## 4      2.3133 2.9302      0.5067      0.0661 0.1102
## 5      3.6824 7.0356      4.6904      0.3086 0.1646
## 6      1.4663 9.4459      3.1641      0.0000 0.2624
## Oilcrops Pulses Spices Starchy.Roots Stimulants Sugar.Crops
## 1      0.1751 0.5003 0.1001      0.3252      0.0750      0
## 2      1.0886 0.8091 0.0000      1.2651      0.2501      0
## 3      0.2688 1.0900 0.1195      1.9262      0.1493      0
## 4      1.0795 1.4981 0.0000      12.6239      0.0441      0
## 5      0.5966 0.4526 0.3497      0.8434      0.4937      0
## 6      0.0309 0.1235 0.0309      1.4045      0.2315      0
## Sugar...Sweeteners Treenuts Vegetal.Products Vegetable.Oils Vegetables
## 1      2.2261      0.1251      45.2476      2.3012      0.7504
## 2      3.4422      0.3972      33.9070      2.8244      2.7508
## 3      3.9869      0.2240      43.9749      5.7638      2.0457
## 4      2.7539      0.0000      45.3184      4.2741      0.3525
## 5      5.8218      0.0823      34.6225      4.6904      1.2960
## 6      7.0536      0.0463      34.9900      5.5410      0.8643
## Obesity Undernourished Confirmed Deaths Recovered Active
## 1      4.5      29.8 0.14213420 0.006185779 0.1233739 0.012574497
## 2      22.3      6.2 2.96730092 0.050951374 1.7926357 1.123713883
```

```
## 3      26.6          3.9 0.24489709 0.006558153 0.1675722 0.070766734
## 4       6.8          25 0.06168747 0.001460550 0.0568077 0.003419224
## 5      19.1        <NA> 0.29387755 0.007142857 0.1908163 0.095918367
## 6      28.5          4.6 4.35614739 0.108226635 3.9051921 0.342728695
##      Population Unit..all.except.Population.
## 1    38928000                                     %
## 2     2838000                                     %
## 3    44357000                                     %
## 4    32522000                                     %
## 5       98000                                     %
## 6    45377000                                     %
```

```
descriptions <- read.csv(file = 'Supply_Food_Data_Descriptions.csv')
descriptions %>%
  head()
```

```
##              Categories
## 1    Alcoholic Beverages
## 2              Animal fats
## 3      Animal Products
## 4 Aquatic Products, Other
## 5 Cereals - Excluding Beer
## 6                Eggs
##
## 1
## 2
## 3 Aquatic Animals, Others; Aquatic Plants; Bovine Meat; Butter, Ghee; Cephalopods; Cream; Crustacean
## 4
## 5
## 6
```

I may only use the first four files.

## An overview of your dataset

**What does it include ?**

COVID-19 Healthy Diet Dataset in different counties

**Where and how will you be obtaining it? Include the link and source .**

[https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?select=Food\\_Supply\\_kcal\\_Data.csv](https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?select=Food_Supply_kcal_Data.csv)

**About how many observations? How many predictors ?**

170 observations

about 101 predictors

**What types of variables will you be working with ?**

I will mainly work with the data related with covid-19.

**Is there any missing data? About how much? Do you have an idea for how to handle it ?**

no (temporary)

## **An overview of your research question(s)**

**What variable(s) are you interested in predicting? What question(s) are you interested in answering ?**

What kind of food is best for avoid infection of covid-19 and helpful for recovery? What kind of food combinations are good for avoid infection of covid-19 and helpful for recovery?

**Name your response/outcome variable(s) and briefly describe it/them .**

food\_type

**Will these questions be best answered with a classification or regression approach ?**

maybe classification

**Which predictors do you think will be especially useful ?**

percentage of COVID-19 active/death/confirmed/recover cases

**Is the goal of your model descriptive, predictive, inferential, or a combination? Explain.**

combination

## **Your proposed project timeline**

**When do you plan on having your data set loaded, beginning your exploratory data analysis, etc? Provide a general timeline for the rest of the quarter.**

I have loaded the data set, but they still need to be resort to research

## **Any questions or concerns**

####Are there any problems or difficult aspects of the project you anticipate? ####Any specific questions you have for me/the instructional team?

the predictors is too many, so choosing a specific direction to research is difficult for me now