Associations Between Diet and COVID-19 Rates

Alex Shen

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

Introduction

When the novel coronavirus (COVID-19) pandemic spread through the world in late 2019 and 2020, it caused quarantines resulting in worldwide shutdowns of travel and commerce. Despite many limitations on travel, mask mandates, testing protocols, and social distancing requirements, the disease still caused shortages in hospital resources, including PPE, ventilators, and ICU beds. For example, the pandemic has caused up to 300,000 people in the US needing ICU beds simultaneously, while estimates for total ICU beds in the US are around 100,000.

The pervasiveness of the disease begs the question whether or not individual behaviors and life choices can decrease one's risk of either contracting or dying from the disease. Obesity, for example, is associated with higher rates of severe illness from COVID-19 (https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html). While obesity is influenced by genetic and environmental factors, it can be controlled by lifestyle choices such as exercise and healthy eating.

On the topic of healthy eating, Harvard T.H. Chan School of Public Health says that a "balanced diet consisting of a range of vitamins and minerals, combined with healthy lifestyle factors. . . most effectively primes the body to fight infection and disease." As a result, we are interested in whether or not dietary choices can effect rates of contracting COVID-19 or rates of serious illness/death.

The data is called the COVID-19 Healthy Diet Dataset and comes from Kaggle courtesy of user Maria Ren, and was last updated the week of 02/06/2021, giving plenty of data from the COVID-19 pandemic to work with. There are 4 datasets provided, each of which contains information about a part of people's diet. One contains information about people's consumption of fat, another for protein, another for overall food intake (in kg), and another for overall food intake (in kcal). Each dataset has 170 rows (countries) and the same 32 columns. There is a column for country name and many different columns representing food groups that contain information about what percentage of people's diet (in the category defined by the dataset it comes from) that each food group takes up. For example, the animal fats column in the Fats dataset represents the percentage of a countries total fat intake that is comprised of animal fats. Meanwhile, the column animal fats in the Overall Food Intake (kg) dataset will represent the percentage of countries total food intake in kg that is comprised of animal fats. In addition, there are columns (we will refer to these following columns as demographic/outcome information) for obesity rate, undernourished rate, confirmed COVID-19 case rate, COVID-19 death rate, active COVID-19 case rate, and total population. All of the rates are calculated as percentage of total population, for example, confirmed COVID-19 case rate in a country is the percent of people in that country that have COVID-19. The demographic/outcome information is identical for all 4 datasets.

The 4 datasets were combined by joining them together on the country field (all the countries in the 4 sets are the same), and deleting the duplicate demographic/outcome information. My overall research question will be to examine any associations between the different proportions certain food groups take up of people's diet and COVID-19 rates on a country level.
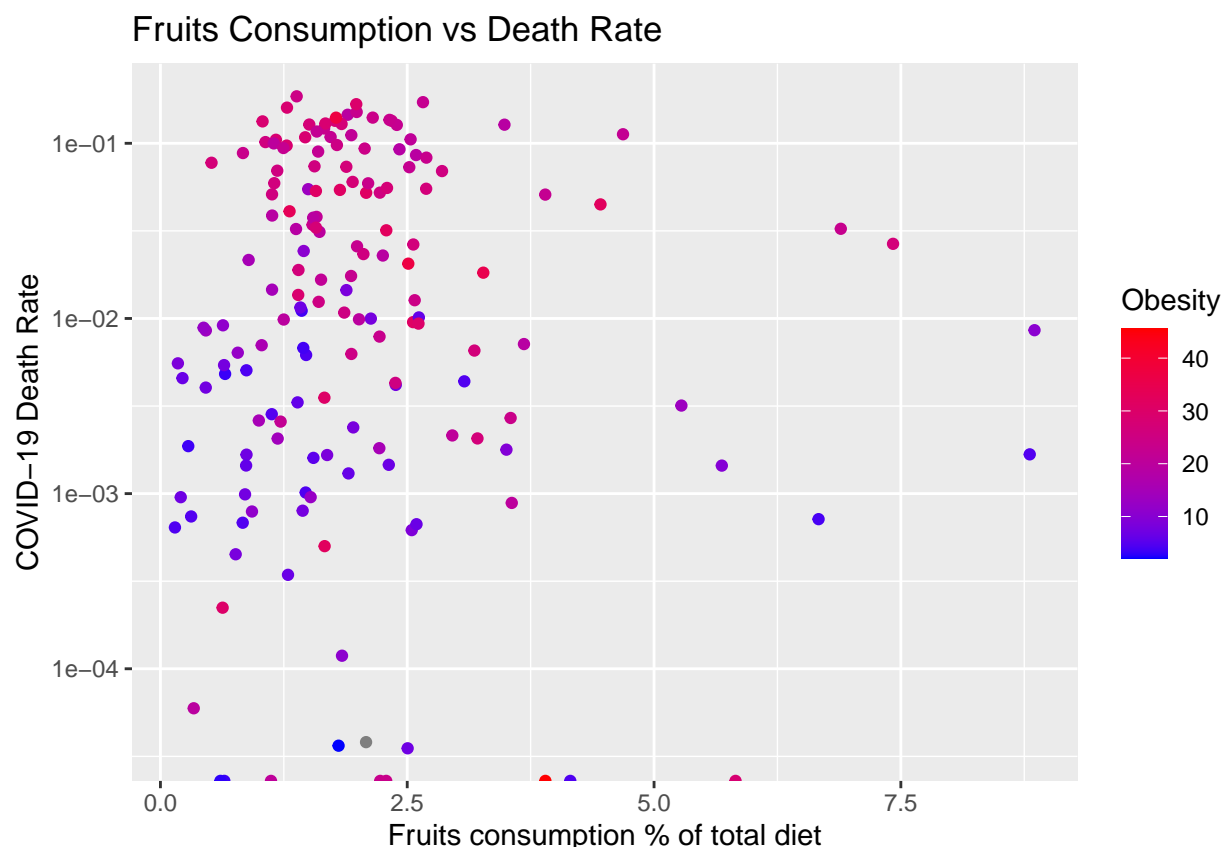
EDA

Will make the plots smaller later.

Vegetable Consumption vs Death Rate
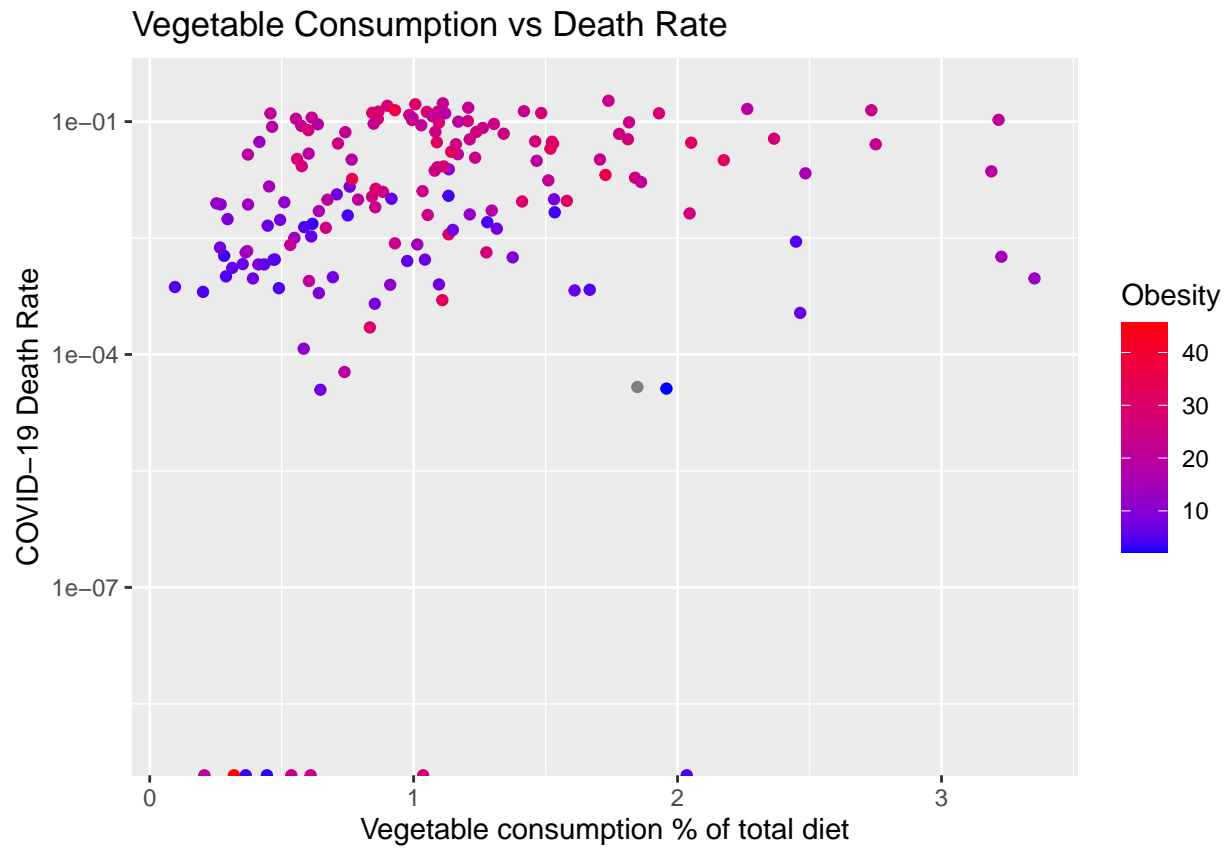
Animal Fats Consumption vs Death Rate

Fruits Consumption vs Death Rate
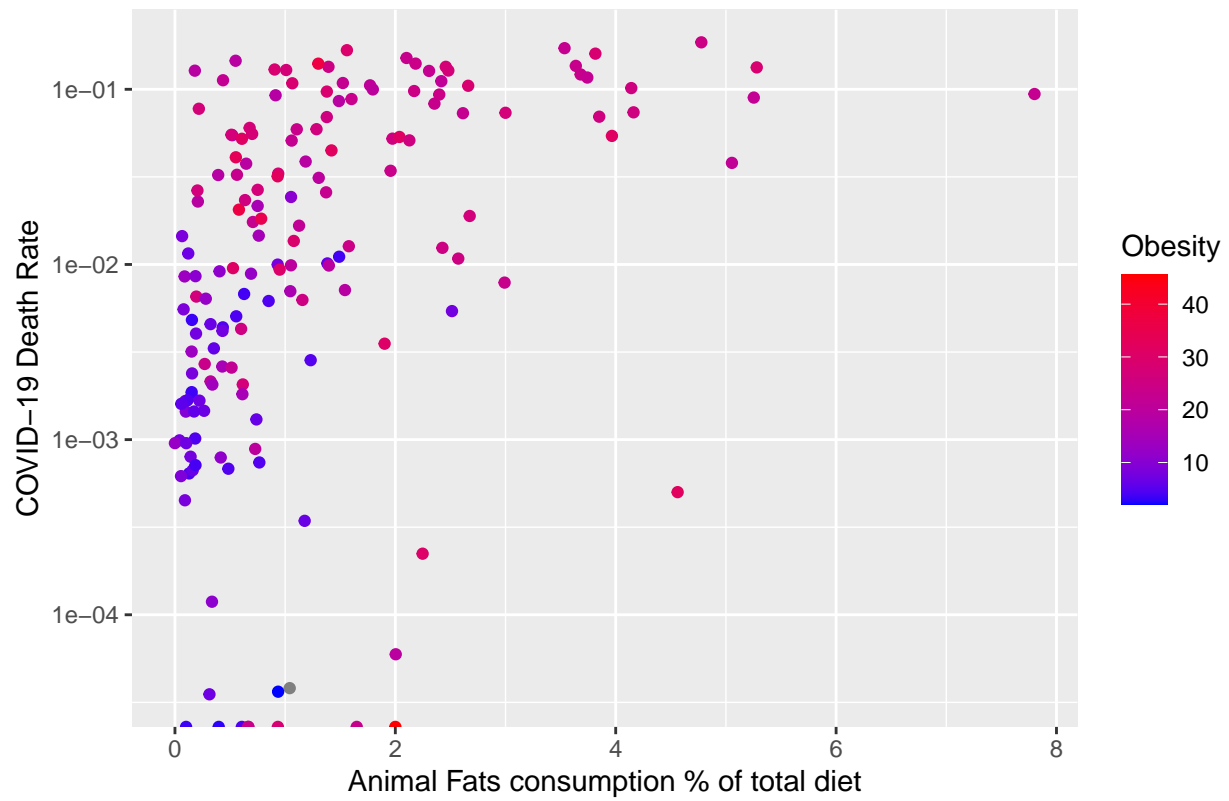
In these 3 plots, we have the percentage of each country's diet taken up by vegetables, animal fats, and fruit respectively on the x axes, and the rate of death from COVID-19 on the y axis. The darker the points, the lower the obesity rate is in the nation. The lighter and more yellow the points, the higher the obesity rate. From all 3 graphs, we can generally see that high death rates are associated with a higher rate of obesity, which makes sense.

For the vegetable graph, it is hard to make any discerning observations without stretching the imagination. From the animal fats graph, we can see that generally, countries with higher consumption of animal fats tend to be be ones with higher death rates. For the fruits graph, it seems that most countries with high death rates are clustered around 2.25% of their diet being fruits.

Methodology

Model formulation/selection

We decided to concentrate on two response variables in particular, rates of COVID-19 related deaths and confirmed cases. We will fit two models, one for each response variable. We are mostly interested in the relationship between each of the four predictor variables vegetable, fruit, meat, and animal fat consumption as a percentage of total food consumed (in kcal) and the response variables. These variables will go into my model. We also decided to control for the consumption of other dietary food groups as a percentage of total food consumed (in kcal). As a result, we also added variables for alcoholic beverages, cereals (wheat, rice, and other grain products), eggs, fish and seafood, milk, offals, oilcrops (including coconuts, peanuts, rapeseed, mustardseed, sunflower seeds, soybeans, etc.), pulses (legumes), spices, starchy root vegetables, spices, stimulants (including coffee, tea, and cacao products), tree nuts, vegetable oils, sugar and sweeteners, and finally miscellaneous, all representing percentage of total food consumed in kcal in order to control for these effects.

In addition, given that there was data available for the percentage of total protein and fats each of the food

groups took up, we decided to also control for these. However, in the interest of being parsimonious, we used variables representing larger food groups to control for them. We added in variables for percentage of total fat intake from animal products (not including seafood), seafood, vegetal products (any plant related products not including oils), and vegetable oils. Similar variables were added to the model but instead of fat, they measured percentage of total protein intake.

We also added in variables for obesity rate and population to control for these factors as well. We decided to not to include undernourished rate. This was because the data for undernourished rate was recorded in a strange way where anything under 2.5% was recorded in the same way. We did not see a principled way to create numerical values for this data, and using another source of data was not feasible because we were not sure exactly where and how the undernourished rate was defined in the original data set. As a result, we decided to omit the undernourished rate from the model. Lastly, from a quick initial survey, it seems that how affluent a country is may effect COVID-19 case rates, deaths, etc. As a result, we downloaded the 2020 GDP per capita for all countries, and merged it into my dataset using the country name. We went through and cleaned it up and hand inputted a few data points due to certain countries being named slightly differently.

let $PropF\ X$ = proportion of a country's fat consumption coming from food X, and $PropP\ X$ = proportion of a country's protein consumption coming from food X, and $PropT\ X$ = proportion of a country's total food consumption in kcal coming from food X.

let $p_i = Death\ Rate_i | PropF\ Animals_i, PropF\ Seafood_i, PropF\ Vegetal\ Products_i, PropF\ Vegetable\ OilsF,$
$PropP\ Animals_i, PropP\ Seafood_i, PropP\ Vegetal\ Products_i,$
$PropP\ Vegetable\ Oils_i, PropT\ Alcohol_i, PropT\ Animal\ Fats_i,$
$PropT\ Cereals_i, PropT\ Eggs_i, PropT\ Seafood_i,$
$PropT\ Meat_i, PropT\ Fruits_i, PropT\ Milk_i, PropT\ Miscellaneous_i,$
$PropT\ Offals_i, PropT\ Oilcrops_i, PropT\ Pulses_i, PropT\ Spices_i,$
$PropT\ Starchy\ Roots_i, PropT\ Stimulants_i, PropT\ Sugar\ and\ Sweeteners_i,$
$PropT\ Treenuts_i, PropT\ Vegetable\ Oils_i, PropT\ Vegetables_i,$
$Obesity_i, GDP\ Per\ Capita_i, Population_i$

$log(p_i/1 - p_i) = \beta_0 + \beta_1 PropF\ Animals_i + \beta_2 PropF\ Seafood_i + \beta_3 PropF\ Vegetal\ Products_i + \beta_4 PropF\ Vegetable\ OilsF + \beta_5 PropP\ Animals_i + \beta_6 PropP\ Seafood_i + \beta_7 PropP\ Vegetal\ Products_i + \beta_8 PropP\ Vegetable\ Oils_i + \beta_9 PropT\ Alcohol_i + \beta_{10} PropT\ Animal\ Fats_i + \beta_{11} PropT\ Cereals_i + \beta_{12} PropT\ Eggs_i + \beta_{13} PropT\ Seafood_i + \beta_{14} PropT\ Meat_i + \beta_{15} PropT\ Fruits_i + \beta_{16} PropT\ Milk_i + \beta_{17} PropT\ Miscellaneous_i + \beta_{18} PropT\ Offals_i + \beta_{19} PropT\ Oilcrops_i + \beta_{20} PropT\ Pulses_i + \beta_{21} PropT\ Spices_i + \beta_{22} PropT\ Starchy\ Roots_i + \beta_{23} PropT\ Stimulants_i + \beta_{24} PropT\ Sugar\ and\ Sweeteners_i + \beta_{25} PropT\ Treenuts_i + \beta_{26} PropT\ Vegetable\ Oils_i + \beta_{27} PropT\ Vegetables_i + \beta_{28} Obesity_i + \beta_{29} GDP\ Per\ Capita_i + \beta_{30} Population_i$

The second model with confirmed COVID_19 case rate as the response has the same formulation, except replacing $Death\ Rate_i$ with $Confirmed\ Case\ Rate_i$

Missing Data

After subsetting the data for only predictors we are interested in, 7 of the 170 countries in the dataset contain missing values. All missing values occur in either the variables for rates of obesity, confirmed cases and deaths. Missing data for a country in the confirmed cases variable was always accompanied with a missing deaths variable. Generally speaking, most countries with missing values are either countries with very small populations or are special cases such as Taiwan or North Korea. We assume this data is MAR since it seems like the missing values are mostly explained by something observed in the data (population). Most of the missing data comes in the response variables of confirmed cases and deaths, and imputing response variables does not make much sense. As a result, we will drop these 7 observations with missing values and perform a complete case analysis.

In addition, we will drop 8 additional observations that have values of 0 for deaths. This is because we think that it is unfeasible to have 0 deaths from COVID-19 by February 2021, and also because beta regression

models must have response variables between 0 and 1 non-inclusive. We will keep in mind that dropping these 15 total observations will bias the results.

Model Diagnostics and Assumptions

One assumption necessary is the independence of observations. This assumption may not be completely met. For example, bordering countries may often share similar diets due to similar cultures. Having close geographical proximity may imply similar climate and habitablity for plant and species, leading to similar diets as well. Countries sharing borders may have more similar rates of COVID-19 as well due to more people (adjusted for population) moving between 2 bordering countries compared to 2 countries across the world. As a result, observations are not independent in this case, but we will proceed.

Diagnostic plots not working due to singular values occurring. Using the plot function, first two plots are outputted, after that it breaks due to singular values. Is something like the below plot good for checking linearity?
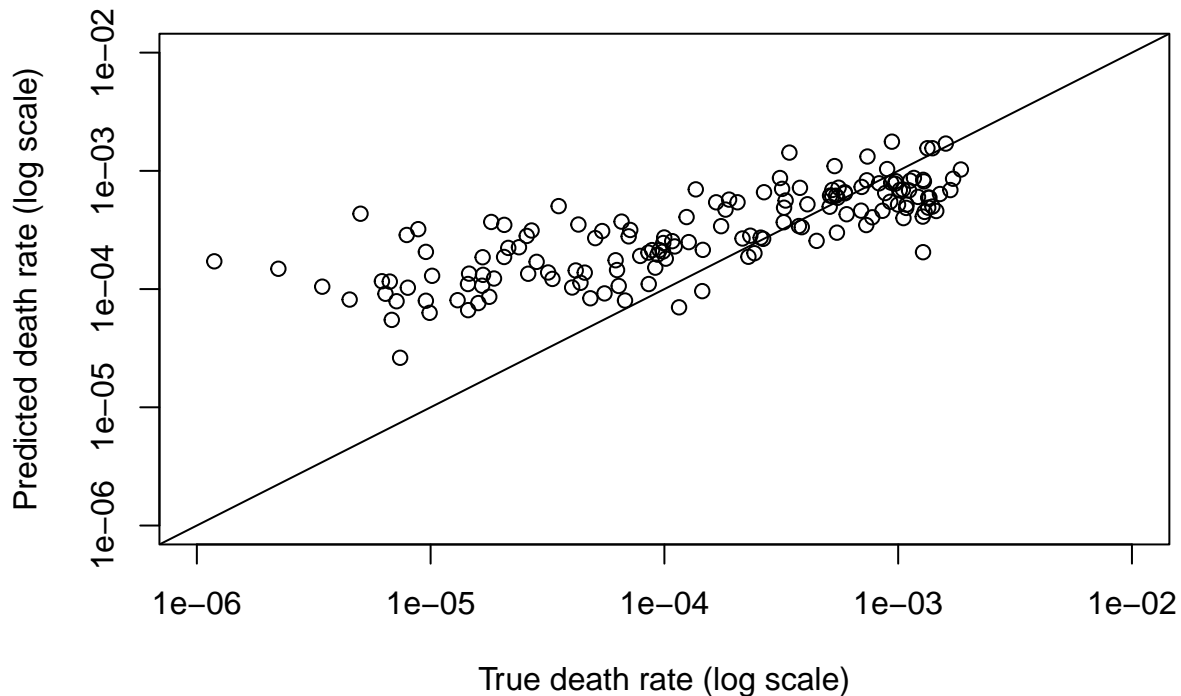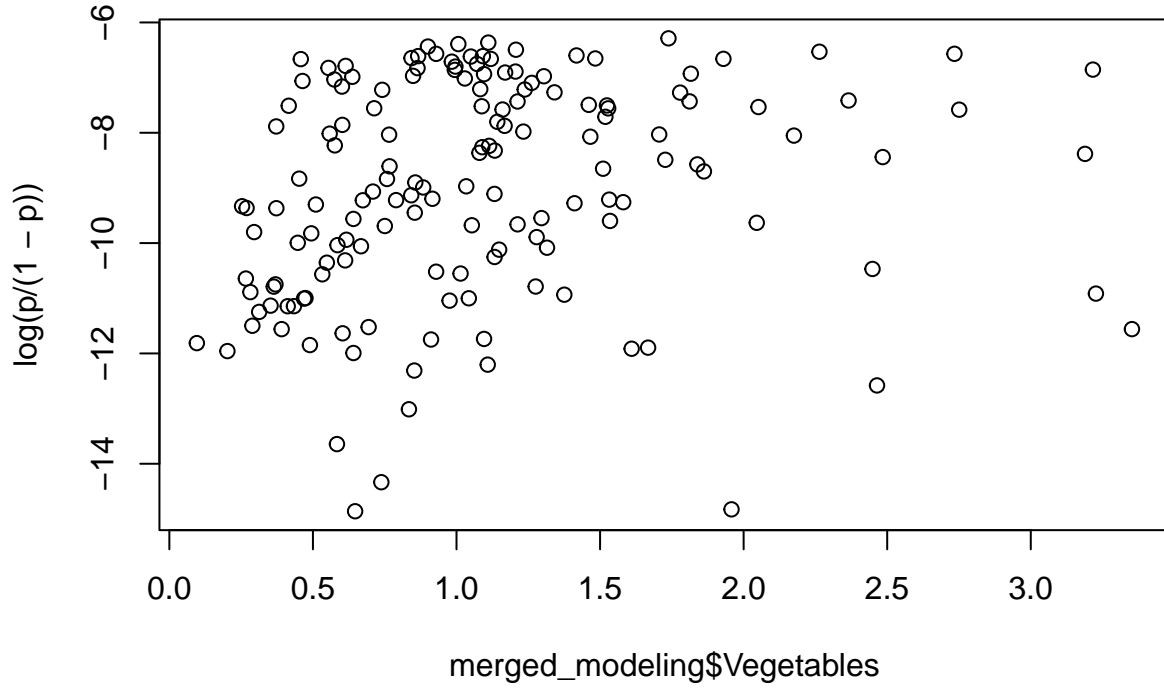
Results

The first model has death rates as the response variable, and the second model has confirmed case rates as the response. Otherwise, the formulations are the same. Some of the main parts of diet we were focused on was the percentage of total diet (in kcal) taken up by animal fats, fruits, and vegetables. For percentage of total diet taken up by animal fats, the coefficient from the model was 0.179. This means that holding other predictors constant, for every one unit increase in percentage animal fats take up of a country's diet, we expect to see a 0.179 increase in the logit of the deaths due to COVID-19 in that country as a proportion of total population with a 95% confidence interval of [-1.9574, 2.3154]. Holding all other predictors constant, for every one unit increase in percentage of a country's total diet taken up by fruits, we expect to see a 0.295 increase in the logit of death rate with a 95% confidence interval of [-1.84924, 2.43924]. Holding all other predictors constant, for every one unit increase in the percentage of a country's total diet taken up by vegetables, we expect to see a 0.190 increase in the logit of death rate with a 95% confidence interval from [-1.98168, 2.36168].

For the second model modeling confirmed cases as a proportion of population, holding all other predictors constant, for every one unit increase in the percentage of a country's diet taken up by animal fats (in kcal), we expect to see a 0.307 increase in the logit of confirmed cases as a proportion of total population, with a 95% confidence interval of [-1.7864, 2.3884]. For every one unit increase in the percentage of a country's diet taken up by fruit, we expect to see a 0.461 increase in the logit of rate of confirmed cases holding all other predictors constant, with a 95% confidence interval of [-1.63228, 2.55428]. Lastly, for every one unit increase in percentage of a country's diet taken up by vegetables, we expect to see a 0.410 increase in the logit of rate of confirmed cases holding all other predictors constant, with a 95% confidence interval of [-1.7068, 2.5268].

Curiously, for both models, at a significance level of 0.05, no predictors in either model are significant given the current model formulations.

Discussion and Conclusions

The fact that no predictors in either model were significant was really surprising because we expected to see at least a few significant predictors, even if they did not come in the variables of interest. For example, in both models we control for obesity. As stated in the introduction, obesity is a known risk factor for serious complications or death in COVID-19 cases. However, in the first model, we found that obesity was not a significant predictor for death. The reason for this may be due to the data we have on diet sharing some of the same explanatory power that obesity has, since information about a country's diet can presumably be associated with obesity rates. Removing the diet related information, we can see that obesity is a significant predictor for death. However, this does not extend in the other direction; removing obesity in both models does not change the fact that all predictors in both are not significant.

One limitation is the missing data. Dropping missing values and unfeasible values to conduct a complete case analysis causes bias in the resulting estimates and results. Imputation would have to be performed with some point imputation method, perhaps using something like KNN due to MICE not working. This would fail to capture variation in the data that MICE can do, but might be better than just dropping these values.

Another limitation is a the issue of countries not being independent observations. It might be possible to add random effects to the model to cope with the independence violations. With further research with groupings of countries can be formed based on a combination of geography, cultural boundaries, and other factors to be introduced as random effects.

Overall, no significant associations between diet and COVID-19 death rates were found given the current model formulation. Given that dietary choices have been shown to be associated with rates of influenza, it might be worth further investigating this topic, especially given the large impact that COVID-19 has on people's lives and the disruptive influence on public health systems worldwide. Further research may be conducted with different model formulations or acquiring more information to inform the model. However, digging deeper into this topic, a pivot might be in order to first examine the relationship between the breakdown of each country's diet and rates of obesity, in order to determine whether or not it is appropriate to control for obesity in the model.

Appendix

Table 1: Death Rate Model

Term	Estimate	Standard Error	Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	-479.25	1144.52	-0.42	0.68	-2722.48	1763.98
Proportion of fat consumption from Animals	-16.01	16.51	-0.97	0.33	-48.37	16.36
Proportion of fat consumption from Seafood	9.32	22.89	0.41	0.68	-35.54	54.18
Proportion of fat consumption from Vegetal Products	-16.01	16.52	-0.97	0.33	-48.38	16.36
Proportion of fat consumption from Vegetable Oils	-16.07	16.51	-0.97	0.33	-48.44	16.30
Proportion of protein consumption from Animals	25.27	15.03	1.68	0.09	-4.19	54.73
Proportion of protein consumption from Seafood	-0.17	0.14	-1.20	0.23	-0.45	0.11
Proportion of protein consumption from Vegetal Products	25.21	15.02	1.68	0.09	-4.23	54.65
Proportion of protein consumption from Vegetable Oils	26.03	15.19	1.71	0.09	-3.73	55.80

Term	Estimate	Standard Error	Statistic	P-Value	95% CI Lower	95% CI Upper
Proportion of total consumption from Alcohol	0.37	1.09	0.34	0.73	-1.76	2.50
Proportion of total consumption from Animal Fats	0.18	1.09	0.16	0.87	-1.96	2.32
Proportion of total consumption from Cereals	0.22	1.09	0.21	0.84	-1.91	2.36
Proportion of total consumption from Eggs	0.41	1.10	0.37	0.71	-1.75	2.58
Proportion of total consumption from Seafood	0.46	1.39	0.33	0.74	-2.26	3.18
Proportion of total consumption from Meat	0.06	1.10	0.05	0.96	-2.09	2.20
Proportion of total consumption from Fruits	0.29	1.09	0.27	0.79	-1.85	2.44
Proportion of total consumption from Milk	0.15	1.09	0.13	0.89	-1.99	2.28
Proportion of total consumption from Miscellaneous	-0.49	1.12	-0.43	0.66	-2.69	1.72
Proportion of total consumption from Offals	-1.63	1.35	-1.21	0.23	-4.28	1.02
Proportion of total consumption from Oilcrops	0.04	1.09	0.04	0.97	-2.10	2.18
Proportion of total consumption from Pulses	0.12	1.09	0.11	0.91	-2.02	2.26
Proportion of total consumption from Spices	-0.08	1.17	-0.07	0.95	-2.37	2.21
Proportion of total consumption from Starchy Roots	0.21	1.09	0.19	0.85	-1.92	2.34
Proportion of total consumption from Stimulants	0.58	1.12	0.52	0.6	-1.62	2.78
Proportion of total consumption from Sugar/Sweeteners	0.24	1.09	0.22	0.82	-1.89	2.38
Proportion of total consumption from Treenuts	0.40	1.10	0.37	0.71	-1.74	2.55
Proportion of total consumption from Vegetable Oils	0.39	1.10	0.35	0.72	-1.77	2.55
Proportion of total consumption from Vegetables	0.19	1.11	0.17	0.86	-1.98	2.36
Obesity	0.02	0.01	1.31	0.19	-0.01	0.04
GDP per capita	0.00	0.00	-0.95	0.34	0.00	0.00
Population	0.00	0.00	0.78	0.44	0.00	0.00

Table 2: Confirmed Case Rate Model

Term	Estimate	Standard Error	Statistic	P-Value	95% CI Lower	95% CI Upper
Intercept	137.84	1045.61	0.13	0.9	-1911.53	2187.20
Proportion of fat consumption from Animals	-23.59	15.18	-1.55	0.12	-53.35	6.16

Term	Estimate	Standard Error	Statistic	P-Value	95% CI Lower	95% CI Upper
Proportion of fat consumption from Seafood	-2.89	20.90	-0.14	0.89	-43.86	38.07
Proportion of fat consumption from Vegetal Products	-23.61	15.18	-1.55	0.12	-53.37	6.15
Proportion of fat consumption from Vegetable Oils	-23.65	15.18	-1.56	0.12	-53.41	6.10
Proportion of protein consumption from Animals	20.47	13.96	1.47	0.14	-6.90	47.83
Proportion of protein consumption from Seafood	-0.11	0.13	-0.81	0.42	-0.36	0.15
Proportion of protein consumption from Vegetal Products	20.36	13.95	1.46	0.14	-6.99	47.71
Proportion of protein consumption from Vegetable Oils	20.10	14.08	1.43	0.15	-7.49	47.69
Proportion of total consumption from Alcohol	0.50	1.06	0.47	0.64	-1.58	2.58
Proportion of total consumption from Animal Fats	0.31	1.07	0.29	0.77	-1.78	2.40
Proportion of total consumption from Cereals	0.39	1.07	0.37	0.71	-1.70	2.48
Proportion of total consumption from Eggs	0.59	1.08	0.55	0.59	-1.53	2.70
Proportion of total consumption from Seafood	-0.01	1.33	-0.01	0.99	-2.61	2.59
Proportion of total consumption from Meat	0.14	1.07	0.13	0.89	-1.95	2.24
Proportion of total consumption from Fruits	0.46	1.07	0.43	0.67	-1.63	2.55
Proportion of total consumption from Milk	0.29	1.06	0.27	0.79	-1.80	2.37
Proportion of total consumption from Miscellaneous	0.14	1.09	0.13	0.9	-2.00	2.28
Proportion of total consumption from Offals	-1.60	1.29	-1.24	0.22	-4.14	0.93
Proportion of total consumption from Oilcrops	0.21	1.07	0.20	0.84	-1.88	2.31
Proportion of total consumption from Pulses	0.34	1.07	0.32	0.75	-1.75	2.43
Proportion of total consumption from Spices	0.04	1.13	0.03	0.97	-2.18	2.26
Proportion of total consumption from Starchy Roots	0.38	1.06	0.35	0.72	-1.71	2.46
Proportion of total consumption from Stimulants	1.09	1.09	1.00	0.32	-1.05	3.22
Proportion of total consumption from Sugar/Sweeteners	0.38	1.06	0.35	0.72	-1.71	2.46
Proportion of total consumption from Treenuts	0.63	1.07	0.59	0.55	-1.47	2.73
Proportion of total consumption from Vegetable Oils	0.51	1.07	0.47	0.64	-1.60	2.61

Term	Estimate	Standard Error	Statistic	P-Value	95% CI Lower	95% CI Upper
Proportion of total consumption from Vegetables	0.41	1.08	0.38	0.7	-1.71	2.53
Obesity	0.02	0.01	1.69	0.09	0.00	0.04
GDP per capita	0.00	0.00	-0.53	0.59	0.00	0.00
Population	0.00	0.00	0.43	0.67	0.00	0.00

```
summary(model1)
```

```
##
## Call:
## betareg(formula = Deaths ~ Animal_no_seafood_fat + Fish..Seafood_fat +
##     Vegetal_no_oil_fat + Vegetable.Oils_fat + Animal_no_seafood_protein +
##     Fish..Seafood_protein + Vegetal_no_oil_protein + Vegetable.Oils_protein +
##     Alcoholic.Beverages + Animal.fats + Cereals...Excluding.Beer + Eggs +
##     Fish..Seafood + Meat + Fruits...Excluding.Wine + Milk...Excluding.Butter +
##     Miscellaneous + Offals + Oilcrops + Pulses + Spices + Starchy.Roots +
##     Stimulants + Sugar...Sweeteners + Treenuts + Vegetable.Oils + Vegetables +
##     Obesity + GDP_per_cap + Population, data = merged_modeling)
##
## Standardized weighted residuals 2:
##     Min      1Q  Median      3Q     Max
## -4.7227 -0.5944  0.1107  0.6995  1.8463
##
## Coefficients (mean model with logit link):
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -4.792e+02  1.145e+03  -0.419   0.6754
## Animal_no_seafood_fat     -1.601e+01  1.651e+01  -0.969   0.3324
## Fish..Seafood_fat          9.316e+00  2.289e+01   0.407   0.6840
## Vegetal_no_oil_fat        -1.601e+01  1.652e+01  -0.969   0.3324
## Vegetable.Oils_fat        -1.607e+01  1.651e+01  -0.973   0.3305
## Animal_no_seafood_protein  2.527e+01  1.503e+01   1.681   0.0927 .
## Fish..Seafood_protein     -1.691e-01  1.413e-01  -1.197   0.2314
## Vegetal_no_oil_protein     2.521e+01  1.502e+01   1.678   0.0933 .
## Vegetable.Oils_protein     2.603e+01  1.519e+01   1.714   0.0865 .
## Alcoholic.Beverages        3.698e-01  1.086e+00   0.340   0.7335
## Animal.fats                1.792e-01  1.090e+00   0.164   0.8694
## Cereals...Excluding.Beer   2.250e-01  1.090e+00   0.206   0.8365
## Eggs                       4.125e-01  1.104e+00   0.374   0.7086
## Fish..Seafood              4.603e-01  1.386e+00   0.332   0.7398
## Meat                       5.701e-02  1.095e+00   0.052   0.9585
## Fruits...Excluding.Wine    2.946e-01  1.094e+00   0.269   0.7877
## Milk...Excluding.Butter    1.465e-01  1.088e+00   0.135   0.8929
## Miscellaneous             -4.871e-01  1.124e+00  -0.433   0.6647
## Offals                    -1.630e+00  1.351e+00  -1.207   0.2276
## Oilcrops                   4.102e-02  1.092e+00   0.038   0.9700
## Pulses                     1.179e-01  1.091e+00   0.108   0.9140
## Spices                    -7.619e-02  1.169e+00  -0.065   0.9480
## Starchy.Roots              2.099e-01  1.088e+00   0.193   0.8469
## Stimulants                 5.825e-01  1.123e+00   0.519   0.6039
## Sugar...Sweeteners         2.437e-01  1.090e+00   0.224   0.8231
## Treenuts                   4.020e-01  1.095e+00   0.367   0.7136
```

```
## Vegetable.Oils                3.890e-01  1.102e+00    0.353    0.7240
## Vegetables                    1.899e-01  1.108e+00    0.171    0.8639
## Obesity                       1.646e-02  1.260e-02    1.307    0.1912
## GDP_per_cap                  -3.895e-06  4.093e-06   -0.951    0.3414
## Population                    3.493e-10  4.506e-10    0.775    0.4382
##
## Phi coefficients (precision model with identity link):
##        Estimate Std. Error z value Pr(>|z|)
## (phi)    3356.6      442.6   7.584 3.35e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood:  1144 on 32 Df
## Pseudo R-squared: 0.6054
## Number of iterations: 226 (BFGS) + 35 (Fisher scoring)
```

```
print(model1$coefficients)
```

```
## $mean
##              (Intercept)      Animal_no_seafood_fat         Fish..Seafood_fat
##             -4.792494e+02            -1.600508e+01              9.315819e+00
##          Vegetal_no_oil_fat       Vegetable.Oils_fat Animal_no_seafood_protein
##             -1.601039e+01            -1.607075e+01              2.526866e+01
##      Fish..Seafood_protein   Vegetal_no_oil_protein    Vegetable.Oils_protein
##             -1.690972e-01             2.521012e+01              2.603365e+01
##        Alcoholic.Beverages              Animal.fats   Cereals...Excluding.Beer
##              3.698036e-01             1.792165e-01              2.249923e-01
##                     Eggs             Fish..Seafood                      Meat
##              4.125272e-01             4.603114e-01              5.700554e-02
##    Fruits...Excluding.Wine   Milk...Excluding.Butter             Miscellaneous
##              2.946357e-01             1.464839e-01             -4.870693e-01
##                   Offals                  Oilcrops                    Pulses
##             -1.630102e+00             4.101743e-02              1.178536e-01
##                   Spices              Starchy.Roots                Stimulants
##             -7.618806e-02             2.099496e-01              5.824697e-01
##         Sugar...Sweeteners                  Treenuts             Vegetable.Oils
##              2.436501e-01             4.019693e-01              3.890247e-01
##               Vegetables                   Obesity                GDP_per_cap
##              1.899189e-01             1.646498e-02             -3.894629e-06
##               Population
##              3.493389e-10
##
## $precision
##    (phi)
## 3356.643
##
## Call:
## betareg(formula = Confirmed ~ Animal_no_seafood_fat + Fish..Seafood_fat +
##     Vegetal_no_oil_fat + Vegetable.Oils_fat + Animal_no_seafood_protein +
##     Fish..Seafood_protein + Vegetal_no_oil_protein + Vegetable.Oils_protein +
##     Alcoholic.Beverages + Animal.fats + Cereals...Excluding.Beer + Eggs +
##     Fish..Seafood + Meat + Fruits...Excluding.Wine + Milk...Excluding.Butter +
##     Miscellaneous + Offals + Oilcrops + Pulses + Spices + Starchy.Roots +
```

```
##       Stimulants + Sugar...Sweeteners + Treenuts + Vegetable.Oils + Vegetables +
##       Obesity + GDP_per_cap + Population, data = merged_modeling)
##
## Standardized weighted residuals 2:
##     Min      1Q  Median      3Q     Max
## -4.6308 -0.4410  0.1838  0.7501  2.2494
##
## Coefficients (mean model with logit link):
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                1.378e+02  1.046e+03   0.132   0.8951
## Animal_no_seafood_fat     -2.359e+01  1.518e+01  -1.554   0.1202
## Fish..Seafood_fat         -2.892e+00  2.090e+01  -0.138   0.8899
## Vegetal_no_oil_fat        -2.361e+01  1.518e+01  -1.555   0.1200
## Vegetable.Oils_fat        -2.365e+01  1.518e+01  -1.558   0.1193
## Animal_no_seafood_protein  2.047e+01  1.396e+01   1.466   0.1426
## Fish..Seafood_protein     -1.054e-01  1.303e-01  -0.809   0.4186
## Vegetal_no_oil_protein     2.036e+01  1.395e+01   1.459   0.1445
## Vegetable.Oils_protein     2.010e+01  1.408e+01   1.428   0.1533
## Alcoholic.Beverages        5.011e-01  1.063e+00   0.471   0.6373
## Animal.fats                3.072e-01  1.065e+00   0.288   0.7730
## Cereals...Excluding.Beer   3.911e-01  1.065e+00   0.367   0.7134
## Eggs                       5.887e-01  1.080e+00   0.545   0.5856
## Fish..Seafood             -1.194e-02  1.325e+00  -0.009   0.9928
## Meat                       1.439e-01  1.070e+00   0.135   0.8930
## Fruits...Excluding.Wine    4.606e-01  1.068e+00   0.431   0.6663
## Milk...Excluding.Butter    2.892e-01  1.064e+00   0.272   0.7857
## Miscellaneous              1.395e-01  1.094e+00   0.128   0.8985
## Offals                    -1.603e+00  1.295e+00  -1.238   0.2156
## Oilcrops                   2.147e-01  1.068e+00   0.201   0.8406
## Pulses                     3.406e-01  1.066e+00   0.320   0.7493
## Spices                     3.949e-02  1.131e+00   0.035   0.9721
## Starchy.Roots              3.768e-01  1.063e+00   0.354   0.7230
## Stimulants                 1.088e+00  1.089e+00   0.998   0.3181
## Sugar...Sweeteners         3.754e-01  1.065e+00   0.352   0.7245
## Treenuts                   6.332e-01  1.071e+00   0.591   0.5542
## Vegetable.Oils             5.078e-01  1.074e+00   0.473   0.6365
## Vegetables                 4.100e-01  1.080e+00   0.380   0.7042
## Obesity                    1.982e-02  1.173e-02   1.689   0.0911 .
## GDP_per_cap               -1.858e-06  3.491e-06  -0.532   0.5946
## Population                 1.927e-10  4.469e-10   0.431   0.6664
##
## Phi coefficients (precision model with identity link):
##       Estimate Std. Error z value Pr(>|z|)
## (phi)    90.60      11.51   7.874 3.43e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 540.4 on 32 Df
## Pseudo R-squared: 0.6441
## Number of iterations: 388 (BFGS) + 21 (Fisher scoring)
```

References

https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?resource=download&select=Food_Supply_Quantity_kg_Data.csv

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7306972/

https://www.hsph.harvard.edu/nutritionsource/nutrition-and-immunity/

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7919160/

https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html

https://www.cdc.gov/healthyweight/index.html

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8954090/