

Final Presentation

```
library(tidyverse)
library(tidymodels)
```

```
food <- readr::read_csv("data/Food_Supply_kcal_Data.csv")
```

#Introduction and Data:

We are interested in looking at the relationship between COVID-19 outcomes and nutrition worldwide. The USDA Center for Nutrition Policy and Promotion suggests a dietary intake which consists of 30% grains, 40% vegetable, 10% fruits, and 20% proteins (dietaryguidelines.gov). Previous studies demonstrate an increased mortality in patients infected with COVID-19 which have chronic inflammatory diseases such as obesity, diabetes, and hypertension.

The prevalence of these chronic inflammatory diseases are known to be correlated with an individual's diet (Onishi 2020). Furthermore, previous studies show that maintaining a healthy diet can decrease risk of severe infection by promoting the immune system (Messina et al. 2020, Iddir et. al 2020). Adequate protein consumption is essential for antibody production and poor nutrient consumption has been shown to increase inflammation and oxidative stress (Iddir et. al 2020). We are ultimately interested in seeing if countries that tend to consume similar diets to those suggested by the USDA show decreased rates of mortality from COVID-19 and how it/if it is related to income levels.

This dataset, "COVID-19 Healthy Diet Dataset" comes from Kaggle. The dataset provides energy intake (kcal) as percentages of total diet by food group. In addition, it provides percentages of obesity and undernourished individuals. Finally it provides data for total confirmed COVID-19 cases, recovered COVID cases, COVID deaths, and active COVID cases for 170 countries. The food supply quantities in addition to the prevalence of obesity and undernourishment in the populations were obtained from the Food and Agricultural Organization of the United Nations, the population count was taken from the Population Reference Bureau, and the Johns Hopkins Center for Systems Science and Engineering was used for COVID-19 data and was last updated in February of 2021.

Works Cited:

Iddir M., et al. Strengthening the Immune System and Reducing Inflammation and Oxidative Stress through Diet and Nutrition: Considerations during the COVID-19 Crisis. *Nutrients*. 2020;12(6):1562.

Messina G., et al. Functional Role of Dietary Intervention to Improve the Outcome of COVID-19: A Hypothesis of Work. *International Journal of Molecular Sciences*. 2020; 21(9):3104

Onishi J., et al. Can Dietary Fatty Acids Affect the COVID-19 Infection Outcome in Vulnerable Populations? *mBio*. 11(4).

INCOME DATA FROM:

```
food <- food %>%
  mutate(income = ifelse(Country %in% c("Afghanistan","Burkina Faso", "Central African Republic", "Chad",
                                         "Cote d'Ivoire", "Democratic Republic of Congo", "Ecuador", "Egypt",
                                         "Guinea", "Honduras", "Kenya", "Liberia", "Madagascar", "Mali",
                                         "Mauritius", "Mozambique", "Niger", "Nigeria", "Papua New Guinea",
                                         "Rwanda", "Senegal", "Sierra Leone", "Somalia", "South Africa", "Tajikistan",
                                         "Tanzania", "Togo", "Tunisia", "Uganda", "Yemen", "Zambia", "Zimbabwe",
                                         "Algeria", "Angola", "Bangladesh", "Belize", "Benin", "Bolivia", "Bosnia and Herzegovina",
                                         "Brazil", "Cambodia", "Cameroon", "Canada", "Chile", "China", "Colombia", "Costa Rica",
                                         "Croatia", "Cuba", "Cyprus", "Czechia", "Denmark", "Dominican Republic", "Ecuador",
                                         "Egypt", "El Salvador", "Estonia", "Finland", "France", "Germany", "Ghana", "Greece",
                                         "Guatemala", "Guinea", "Guyana", "Honduras", "Hungary", "Iceland", "India", "Indonesia",
                                         "Israel", "Italy", "Japan", "Jordan", "Kazakhstan", "Kenya", "Korea", "Kuwait", "Kyrgyzstan",
                                         "Laos", "Latvia", "Lebanon", "Lithuania", "Luxembourg", "Madagascar", "Malawi", "Malaysia",
                                         "Maldives", "Mali", "Mauritius", "Mexico", "Moldova", "Mongolia", "Montenegro", "Morocco",
                                         "Mozambique", "Myanmar", "Namibia", "Netherlands", "New Zealand", "Nicaragua", "Nigeria",
                                         "Norway", "Oman", "Pakistan", "Panama", "Paraguay", "Peru", "Philippines", "Poland", "Portugal",
                                         "Romania", "Russian Federation", "Saudi Arabia", "Senegal", "Serbia", "Sierra Leone", "Singapore",
                                         "Slovakia", "Slovenia", "South Africa", "South Korea", "Spain", "Sri Lanka", "Sweden", "Switzerland",
                                         "Tajikistan", "Tanzania", "Thailand", "Timor-Leste", "Togo", "Tunisia", "Turkey", "Ukraine",
                                         "United Kingdom", "United States", "Uruguay", "Uzbekistan", "Venezuela", "Vietnam", "Yemen", "Zambia",
                                         "Zimbabwe", "UMIC", "HIC"))))
```

```

food <- food %>%
  mutate(Grains = `Cereals - Excluding Beer` + `Starchy Roots`) %>%
  mutate(Vegetables2 = Pulses + `Vegetal Products` + Vegetables) %>%
  mutate(Fruits = `Fruits - Excluding Wine`) %>%
  mutate(Fats = `Animal fats` + Oilcrops + Treenuts + `Vegetable Oils`) %>%
  mutate(Protein = `Animal Products` + Eggs + `Fish, Seafood` + Meat + `Milk - Excluding Butter` + Offa
  mutate('Alcohol/Stimulants' = `Alcoholic Beverages` + Stimulants) %>%
  mutate(Other = `Aquatic Products, Other` + Miscellaneous + Spices + `Sugar Crops` + `Sugar & Sweetene

```

Our dataset included more specific dietary composition than the USDA values. Therefore, to be able to compare them, we grouped the datasets into broader dietary categories including grains, vegetables, fruits, fats, proteins, and a general miscellaneous category. We did so by reading the descriptions of what composes each variable in the dataset.

```

meanfood <- food %>%
  summarize(meangrains = mean(Grains), meanveg = mean(Vegetables2), meanfruit = mean(Fruits), meanfat =

meanfoodincome <- food %>%
  group_by(income) %>%
  summarize(meangrains = mean(Grains), meanveg = mean(Vegetables2), meanfruit = mean(Fruits), meanfat =
  add_row(income="USDA", meangrains=30, meanveg=40, meanfruit=10, meanprotein=20) %>%
  pivot_longer(cols=meangrains:meanother, names_to = "Means", values_to = "PercentDiet")

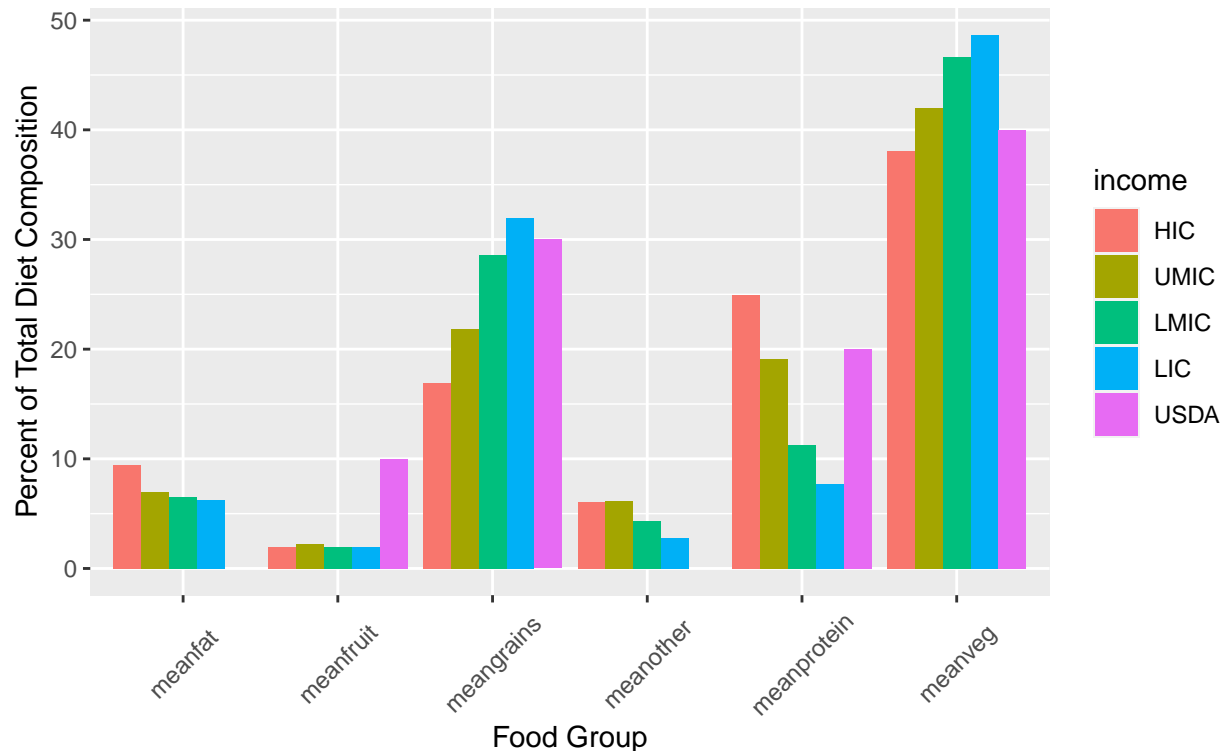
meanfoodincome %>%
  mutate(income = fct_relevel(income,
    "HIC", "UMIC", "LMIC",
    "LIC", "USDA")) %>%
  ggplot(aes(x=Means, y=PercentDiet, fill=income))+geom_col(position="dodge") + labs(x="Food Group", y=

## Warning: Removed 2 rows containing missing values (geom_col).

```

Diet Make up by Food Group

broken down by income category and compared to USDA suggested diet



In this graph, we summarized the break down of diet by income country and also introduced a new column with the recommended values from USDA. This visualization shows that low income countries seem to be consuming less than the recommended amount of protein and more grains than recommended. We can also see that higher income countries consume more protein and less grains than lower income countries. Furthermore, while USDA does not provide a value for suggested fat, it is a nutritional group represented in our data set.

This graph suggests that in our analysis we should include income level as a possible predictor of covid outcomes since countries vary in their nutritions by income level.

```
food <- food %>%
  mutate(rec_rate=Recovered/(Confirmed)) %>%
  mutate(adherence_grains = (30 - Grains)^2) %>%
  mutate(adherence_veg = (40 - Vegetables2)^2) %>%
  mutate(adherence_fruit = (10 - Fruits)^2) %>%
  mutate(adherence_protein = (20 - Protein)^2) %>%
  mutate(adherence_tot = adherence_grains + adherence_protein + adherence_fruit + adherence_veg) %>%
  filter(rec_rate>0)

tfood<-food %>%
  group_by(income) %>%
  summarize(meangrains = mean(Grains), meanveg = mean(Vegetables2), meanfruit = mean(Fruits), meanfat =
  add_row(income="USDA", meangrains=30, meanveg=40, meanfruit=10, meanprotein=20)

summary(aov(rec_rate~income,data=food))
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## income      3  0.324  0.10785    2.546  0.058 .
```

```
## Residuals    156    6.607 0.04236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, we created a variable for how much the average diet in each country differs from the suggested USDA value. We did so by subtract the mean value from each food group from the USDA value and then squaring that value. We then summed up this value from each food group to create adherence total variable. We squared them because eating too much or too little of a food group should not balance each other out, rather it should increase the deviation from the suggested intake.

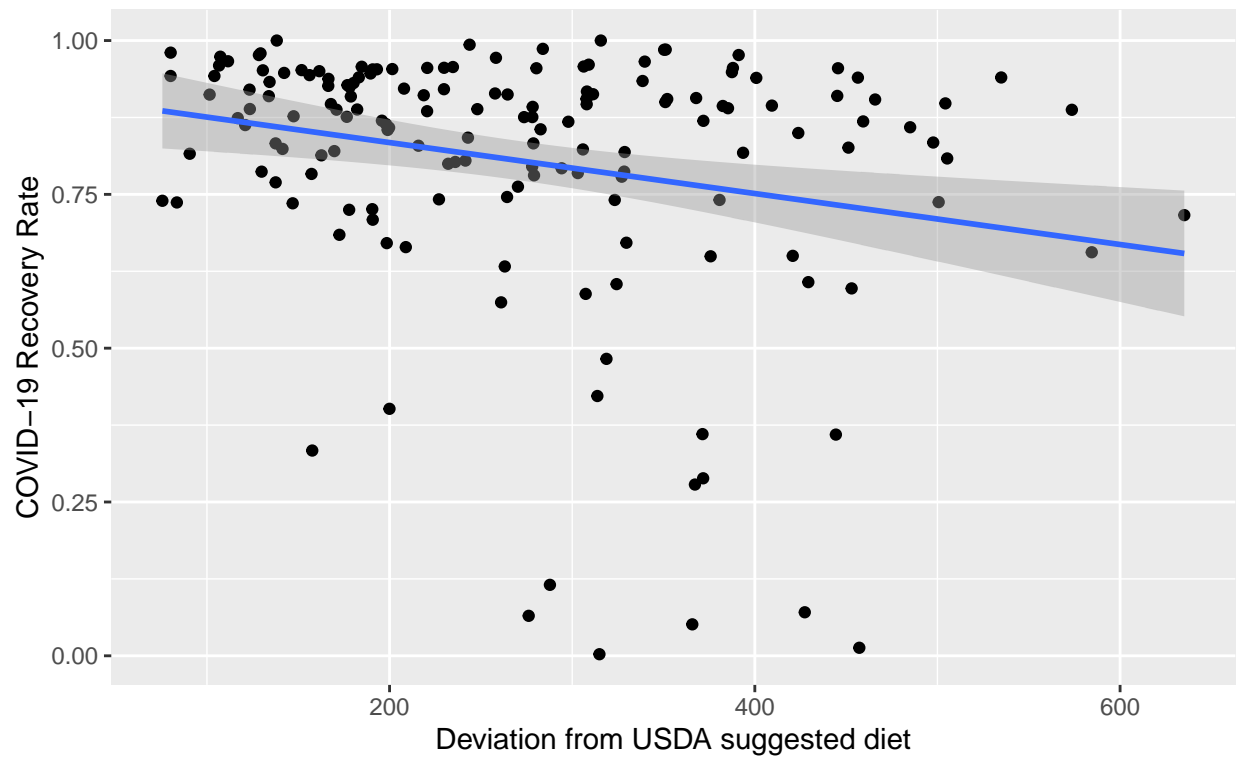
To see if income level has an effect on recovery rate, we performed an ANOVA test. The results of our test indicate that

```
food <- food %>%
  mutate(rec_rate=Recovered/(Confirmed)) %>%
  mutate(adherence_grains = (30 - Grains)^2) %>%
  mutate(adherence_veg = (40 - Vegetables2)^2) %>%
  mutate(adherence_fruit = (10 - Fruits)^2) %>%
  mutate(adherence_protein = (20 - Protein)^2) %>%
  mutate(adherence_tot = adherence_grains + adherence_protein + adherence_fruit + adherence_veg) %>%
  filter(rec_rate>0)

food %>%
  filter(adherence_tot<750) %>%
ggplot(aes(x = adherence_tot, y = rec_rate)) +
  geom_point()+geom_smooth(method="lm") + labs(
  title = "COVID-19 Recovery Rate in relationship to deviation from USDA suggested Dietary Intake",
  subtitle="linear model by income category",
  x = "Deviation from USDA suggested diet",
  y = "COVID-19 Recovery Rate")

## `geom_smooth()` using formula 'y ~ x'
```

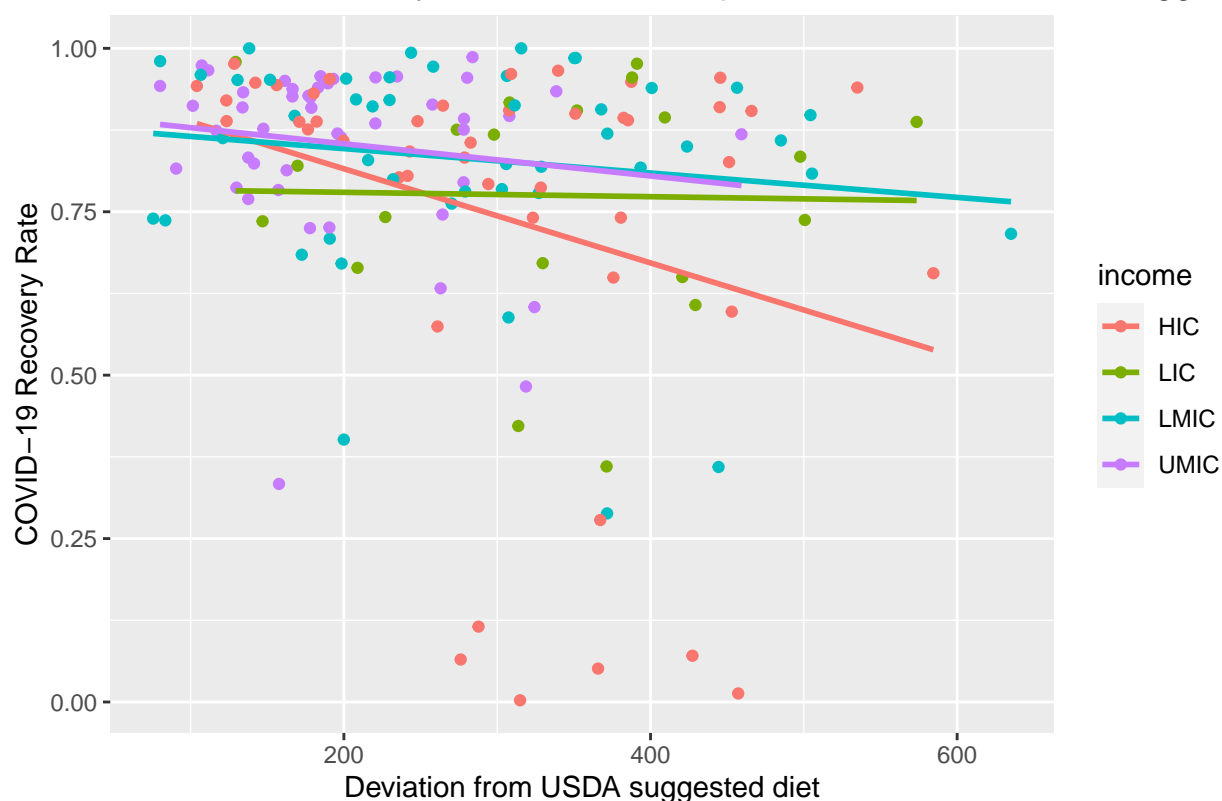
COVID-19 Recovery Rate in relationship to deviation from USDA suggested
linear model by income category



```
food %>%
  filter(adherence_tot<750) %>%
  ggplot(aes(x = adherence_tot, y = rec_rate, color=income)) +
  geom_point()+geom_smooth(method="lm", se = FALSE) + labs(
    title = "COVID-19 Recovery Rate in relationship to deviation from USDA suggested Dietary Intake",
    x = "Deviation from USDA suggested diet",
    y = "COVID-19 Recovery Rate")

## `geom_smooth()` using formula 'y ~ x'
```

COVID-19 Recovery Rate in relationship to deviation from USDA suggested



Here we compared the relationship between deviation from USDA suggested intake with recovery rates from COVID-19. The recovery rates were calculated by calculated the percent of the population that recovered from COVID-19 with the percent of the population that was diagnosed with COVID-19.

In the first graph, we see that there is a negative correlation between recovery rates and deviation from USDA suggested dietary intake. This suggests that as ones dietary intake begins to differ from the USDA suggested intake, they on average seem to have a lower change of recovering from COVID-19.

To further look at this relationship, we compared this relationship by income level. We can see that while lower middle income and upper middle income countries have a slightly negative relationship, lower income countries almost show no relationship between deviation rate and recovery rate. In contrast, high income countries show the strongest correlation between deviation rate and recovery rate.

WHY?

```
res <- linear_reg() %>%
  set_engine("lm") %>%
  fit(rec_rate ~ adherence_tot, data = food) %>%
  tidy() %>%
  print()
```

```
## # A tibble: 2 x 5
##   term          estimate std.error statistic  p.value
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    0.886    0.0377     23.5 1.83e-53
## 2 adherence_tot -0.000292 0.000123     -2.36 1.94e- 2
```

This linear model shows that there is a significant relationship between recovery rate and deviation rate since our p value is less than the alpha level of 0.05. For every integer increase in adherence rate, the chances of

recovering are on average 0.0002 lower.