UNDERNOURISHMENT, OBESITY, AND COVID-19

How are they related?

By: Iris Michelsen, Anthony Abushacra, XueYi Lu, Sowmya Kotipalli, Thanushree Vale

TABLE OF CONTENTS

GUIDING QUESTIONS

Our analysis aims to answer a few guiding questions about the relationship between malnutrition (either obesity or undernourishment) and COVID-19 susceptibility.

- What do the variables tell us on their own? Is the data clean, accurate, and consistent? How is it distributed?
- What are the relationships between the variables? Is there a correlation between obesity/undernourishment and COVID-19 confirmed cases? What about COVID-19 deaths?
- What stories do these trends tell?

Our group project seeks to answer these questions using data analysis and visualizations.

DATA CLEANING

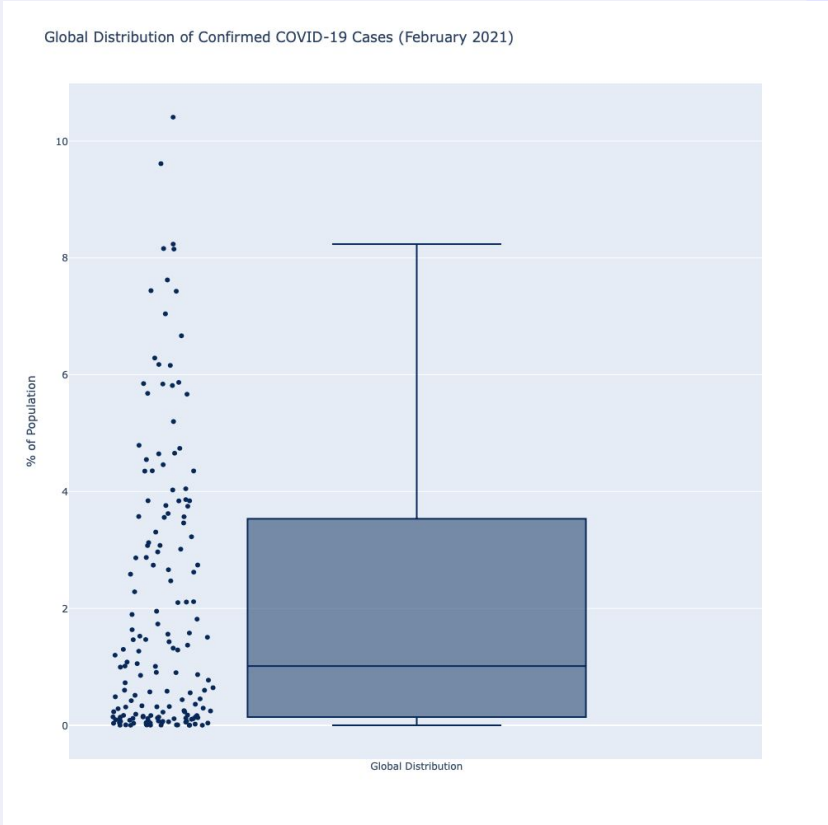
Here are the steps we took to clean our data and prepare it for analysis:

- Import all of our tools - pandas, matplotlib, and more
- Import raw data and format as data frames
- Remove rows with empty cells
- Create variables for columns of interest
- Find global averages

Now we're ready to begin analyzing our variables!

CONFIRMED CASES

- Data for confirmed COVID-19 cases comes from Johns Hopkins University Coronavirus Resource Center.
- Two outliers- Montenegro (10.4%) and Czechia (9.61%)
- Many lower reported rates are from developing countries
- Stats for confirmed COVID-19 diagnosis:
  - p-value: 2.1542800095891712e-12
  - Mean: 2.07
  - Median: 1.07
  - Range: 10.41
  - Variance: 5.65
  - Standard Deviation: 2.38
  - Highest %: Montenegro 10.4%
  - Lowest %: Vanuatu 0.0003%



COVID-19 DEATHS

Top Statistics

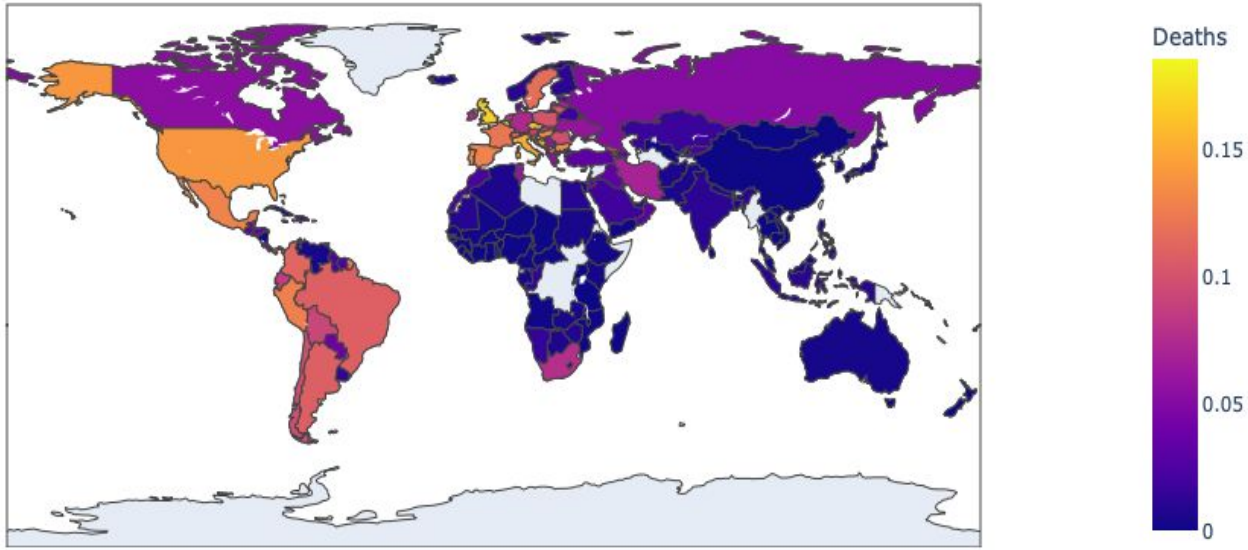
- There are two outliers in the data, both on the higher end: Belgium with a death rate of 0.185% and Slovenia with a death rate of 0.171%
- The three countries with the highest rates of COVID deaths are: Belgium with 0.1855, Slovenia with 0.171%, and The United Kingdom with 0.167%
- The three countries with the lowest rates of COVID deaths are: Cambodia, Dominica, and Lao People's Democratic Republic, all with a rate of 0.0%

Shapiro-Wilk Test p-value: 5.083525831187097e-14
The data is not normally distributed

Average COVID-19 Death Rate 0.040%
Median COVID-19 Death Rate 0.012%

Roughly 68% of the data is between -0.0095 and 0.088%
Roughly 95% of the data is between -0.058% and 0.137%
Roughly 99.7% of the data is between -0.106% and 0.185%

COVID-19 Death Rates by Country



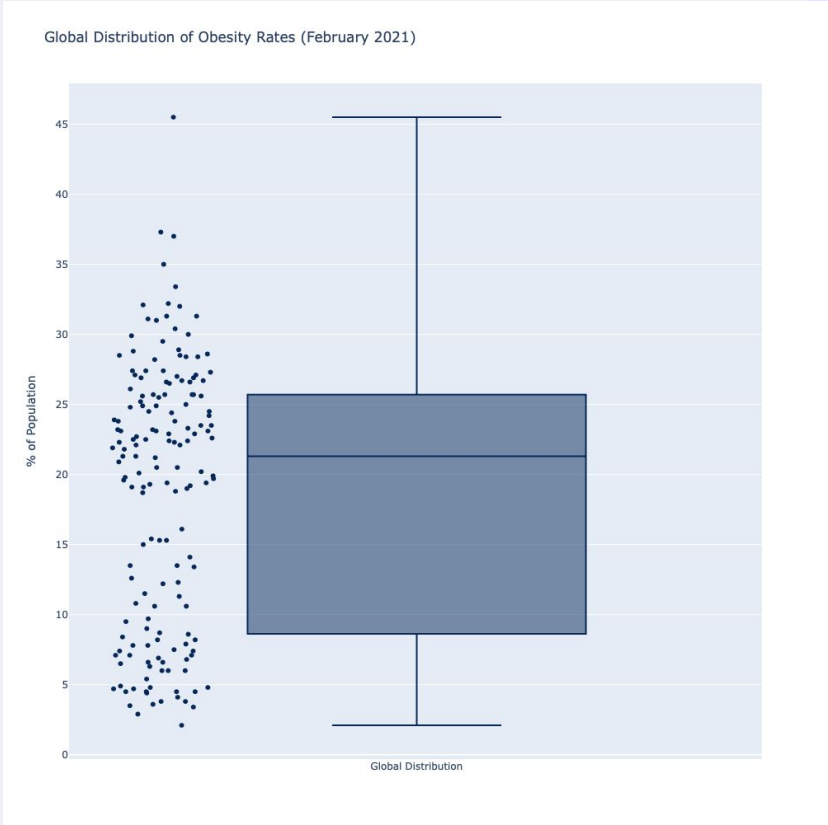
OBESITY RATES

- The Food and Agriculture Organization of the United Nations (where our obesity data was pulled from) defines obesity as:

"Percentage of adults ages 18 and over whose Body Mass Index (BMI) is more than 30 kg/m2."

- Interesting gap in the data between Botswana (16.1%) and Bolivia (18.7%)

- Stats for Obesity:
  - p-value: 1.0834194199560443e-06
  - Mean: 18.60
  - Median: 21.55
  - Range: 43.4
  - Variance: 90.60
  - Standard Deviation: 9.52
  - Highest rate: Samoa 45.5%
  - Lowest rate: Vietnam 2.1%

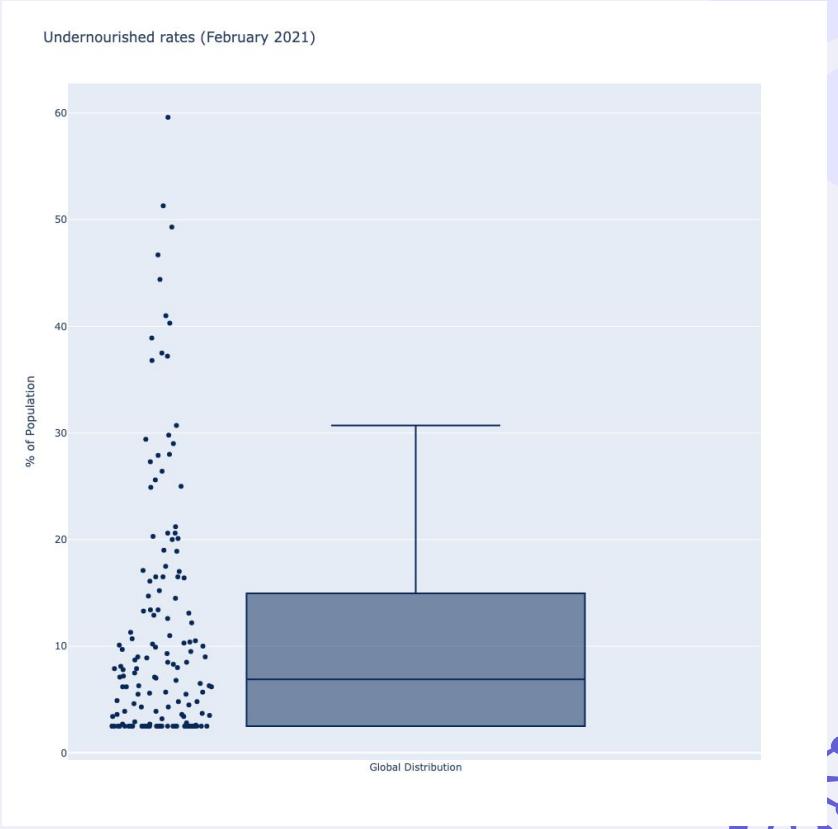


UNDERNOURISHED RATES

- The Food and Agriculture Organization of the United Nations (where the undernourished data was obtained) defines undernourished as:

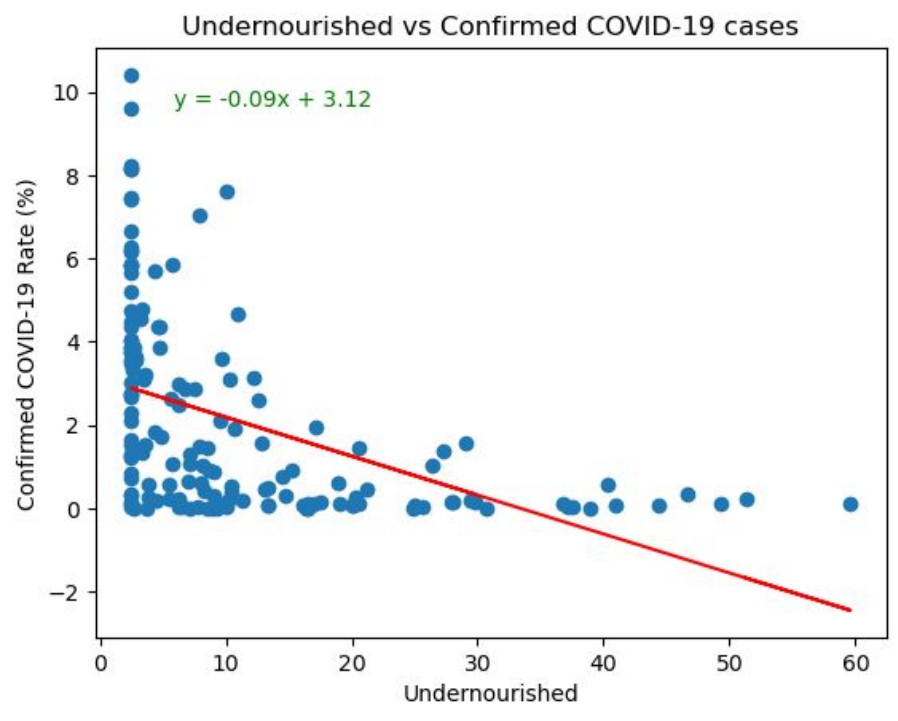
"[the] estimate of the proportion of the population whose habitual food consumption is insufficient to provide the dietary energy levels that are required to maintain a normal active and healthy life."

- Stats for Undernourished Rates:
  - p-value: 6.974955061087159e-15
  - Mean: 11.21
  - Median: 6.9
  - Variance: 136.87
  - Standard Deviation: 11.70
  - Highest rate: Central African Republic 59.6%
  - Lowest rate: Unknown



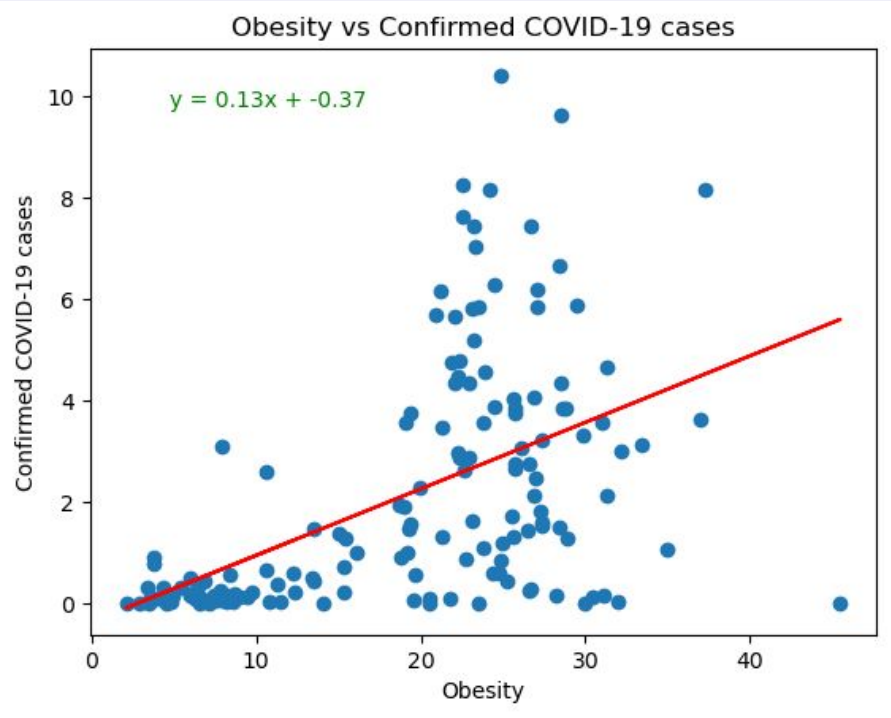
UNDERNOURISHED V. CONFIRMED

- The r-squared is: 0.2117166835649667
  - This means that around 21.17% of the variability in the Confirmed COVID-19 cases can be explained by the variability in the undernourished variable based on the linear regression model.
- The P-value is 1.511194582840137e-09
  - p-value provides strong evidence against the null hypothesis.



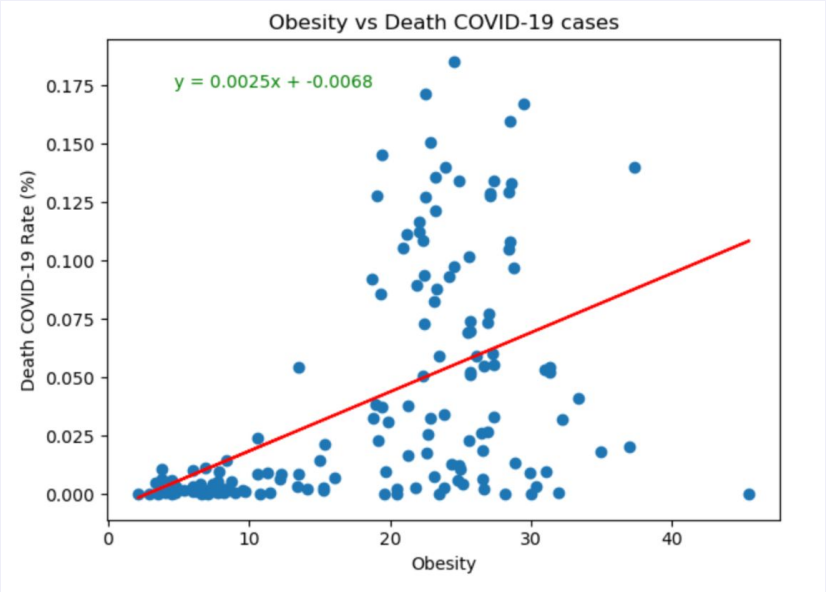
OBESITY V. CONFIRMED

- The r-squared is: 0.27548993574147734
  - Low r-squared, therefore low variability
- The P-value is 2.012809548086103e-12
  - Strong evidence against null hypothesis

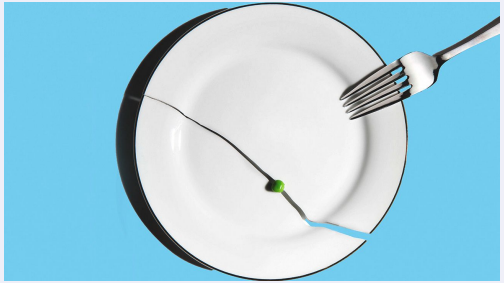


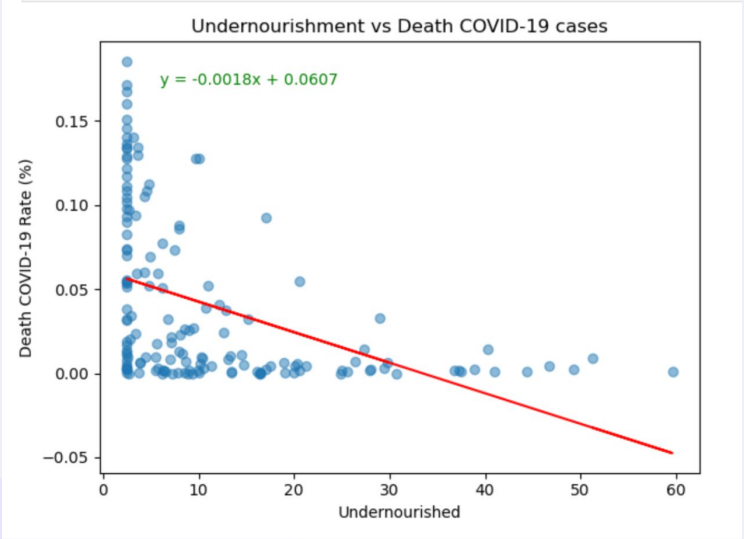
OBESITY V. DEATHS

- The r-squared is: 0.2425095259439095
- P-value: 6.589453144445402e-11
- Moderate positive linear relationship



UNDERNOURISHED V. DEATHS





- The r-squared is: 0.18749991267890193
- P-value: 1.64454424178202 18e-08
- Weak negative linear relationship

FINAL CONCLUSIONS

P-values -

- Confirmed - 2.1542800095891712e-12
- Deaths - 1.4726930883546957e-13
- Obesity - 1.0834194199560443e-06
- Undernourishment - 6.974955061087159e-15

R-squared -

- Obesity vs Confirmed COVID-19 cases – 0.27548993574147734
- Undernourishment vs Confirmed COVID-19 cases - 0.2117166835649667
- Obesity vs COVID-19 Deaths - 0.2425095259439059
- Undernourishment vs COVID-19 Deaths - 0.18749991267890193

THANK YOU!

Dataset:

Ren, Maria. (2021). COVID-19 Healthy Diet Dataset [Description of form]. Retrieved from https://www.kaggle.com/datasets/mariaren/covid19-healthy-diet-dataset?resource=download