|  |
| --- |
| Photo displaying partial image of two pie charts on a canvas-textured page |
| Correlation between Physical Height and Contributing Factors  Coursera Capstone Project Report |
| |  |  |  | | --- | --- | --- | | Gilburt | 11/13/20 | Applied Data Science Capstone | |

Contents

[Introduction 2](#_Toc56120547)

[Antithesis 2](#_Toc56120548)

[Data 2](#_Toc56120549)

[Data Acquisition 2](#_Toc56120550)

[Data Setup 3](#_Toc56120551)

[Method 4](#_Toc56120552)

[Calculation of Target Variables 4](#_Toc56120553)

[Height 4](#_Toc56120554)

[Meat Consumption 5](#_Toc56120555)

[Milk Consumption 6](#_Toc56120556)

[Wealth 7](#_Toc56120557)

[Results and Discussion 7](#_Toc56120558)

[Conclusion 10](#_Toc56120559)

[Future Next Steps 10](#_Toc56120560)

# Introduction

The topic or **business problem** for the data science project is determining what factors contribute to a country’s average height? Many might say genetics. However, environment cannot be overlooked. This is potentially an interest of everyone (the general population) who are curious about why some countries are taller than others.

My theory is nutrition/diet leads to a high quality of life and on-average, taller people for a country but the study will also look at other factors.

There have been many studies that say wealth leads to better food and healthcare which directly leads to taller people.

## Antithesis

There is also a thought that I could be looking at the wrong data. How wealthy a country or how much meat and dairy they consume now may have extraordinarily little to do with their current average height and more to do with the future height of the nation.

# Data

This section describes the data used to solve the problem and the source of the data.

To determine what factors, contribute to a country’s average adult height, I will first be identifying the 10 tallest countries by average height for a male and woman, then the 10 shortest countries. There will also be a study on all of the countries (that we have data for) and their correlation between height, meat consumption, milk consumption, and wealth.

## Data Acquisition

Data sources are listed below for each attribute targeted in the study: Average height (in inches) per country, average meat consumption per capita (lbs.), average dairy consumption per capita (gallons),

*Data for height:*

<http://www.averageheight.co/average-male-height-by-country> #website used for a Business insider article on tallest countries (2016)

<http://www.ncdrisc.org/data-downloads-height.html> #dataset on height by country (2017)

* “NCD Risk Factor Collaboration (NCD-RisC) is a network of health scientists around the world that provides rigorous and timely data on major risk factors for non-communicable diseases for all of the world’s countries.”

<https://www.worlddata.info/average-bodyheight.php#by-population> #dataset on height by country

**Step 1 – List of countries and height:** Aggregate and average out the heights provided in the (3) datasets and use the final dataset to determine the 10 tallest countries (data needed – country name and height for male and female) and 10 shortest countries.

Next, we will look to gather data on each country.

**Step 2 – Factors that might contribute to a country’s height**: Determine various environmental characteristics that may lead to a country being tall.

These datasets will be scraped from each website and converted to data frames.

* Diet and Nutrition: Theory – a heavy meat (protein) and dairy-based (calcium) diet leads to taller people.
  + Meat consumption per capita
    - <https://en.wikipedia.org/wiki/List_of_countries_by_meat_consumption>
    - (Did not find many other sources with complete data on world meat consumption per capita broken down by country)
  + Dairy consumption per capita
    - <https://www.statista.com/statistics/272003/global-annual-consumption-of-milk-by-region/> #csv pulled from website with 2020 data on dairy consumption per capita
* Wealth: Theory – wealthier countries and their citizens have a better quality of life which may lead to taller people.
  + GDP per capita (current) and GDP per capita (year 2000) – country was wealthy 20 years ago which contributed to the people today being tall
    - <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD> #csv file pulled from World Bank showing GDP per capita for every year up until 2019. Use the columns from 2016-2019 and averaged the GDP per capita.

Ideas: Future data to analyze and correlate with height (more difficult to find this information)

* Sleep: Theory – the more sleep you get, the taller you get (data might be too difficult to find which will lead to this factor being omitted)
  + Average sleep time
* Gravitational Pull: (more off-the-wall) Theory –gravitational pull ranges around the world and maybe that contributes to where people grow tall. Need to confirm.
  + Gravitational pull (by longitude and latitude coordinates of the center of each country)
    - https://www.sensorsone.com/local-gravity-calculator/

**Step 3 – Wrangle and organize all data then compare and correlate:** Procure, format, and build the datasets then plots graphs and compute correlation between various datasets.

## Data Setup

The first task was to procure and aggregate a final **List of countries and height.** Data was pulled from 3 sources then cleansed and formatted before being averaged out to show the following:

* **Country name**
* **Average height (male) in inches**

Then the list was sorted from tallest to shortest and then 10 tallest countries and 10 shortest countries were taken and put into 2 lists: (1) Top 10 tallest countries, (2) Top 10 shortest countries. The study will also look at all countries.

Next, wealth (Global Domestic Product per Capita) and average meat/dairy consumption per capita for each of these countries will be procured and aggregated into the main height lists:

* **Country name**
* **Average height (male) in inches**
* **Average meat consumption (Pounds)**
* **Average dairy consumption (Gallons)**
* **Average GDP per capita (USD)**

Once the final list is compiled, then further analysis can be performed. FourSquare data will also be analyzed.

# Method

This section is where I discuss and describe any exploratory data analysis performed and highlight inferential statistical testing.

## Calculation of Target Variables

The first target variable: Average height for men (in inches) needed to be calculated first to find the top 10 tallest and shortest countries. Then add in average GDP per capita, average dairy consumption, restaurant pricing, average sleep time, and gravitational pull before analyzing correlations.

### Height

Average height for each country was first scraped from (3) source websites: Average height dot com, NCD Risk Factor Collaboration (NCD-RisC), and World Data. Each web page needed to be scraped and then the average height for each country converted from meters, feet and inches, or centimeters to just inches. Once this was complete, the 3 columns of average height were then averaged once more to create a final dataset. The final dataset was then sorted by ascending heights and the top 10 tallest and top 10 shortest countries were taken. The results are shown below:

#### Top 10 Tallest Countries

|  |  |
| --- | --- |
| **Country** | **Average Height(in)** |
| Bosnia & Herzegovina | 72.4 |
| Netherlands | 72.19 |
| Denmark | 71.73 |
| Montenegro | 71.48 |
| Serbia | 71.46 |
| Norway | 71.46 |
| Iceland | 71.19 |
| Germany | 71.11 |
| Croatia | 71.1 |
| Czech Republic | 70.96 |

#### Top 10 Shortest Countries

|  |  |
| --- | --- |
| **Country** | **Average Height(in)** |
| Indonesia | 62.93 |
| Vietnam | 63.82 |
| Bolivia | 63.89 |
| Philippines | 63.91 |
| Nepal | 64.08 |
| Cambodia | 64.09 |
| Sri Lanka | 64.68 |
| Peru | 64.73 |
| Nigeria | 64.77 |
| Ecuador | 64.84 |

### Meat Consumption

For meat consumption, the study looked to get a data frame that showed each country and the average amount of meat consumed per capita (kg/person). The first data source came from Wikipedia (<https://en.wikipedia.org/wiki/List_of_countries_by_meat_consumption>). However, the latest data (2017) in the chart is incomplete, therefore we are using the data from 2009. The data was scraped from the website, formatted, and sorted by country. Then the top 10 tallest and top 10 shortest countries were pulled out of the list along with their meat consumption per capita.

#### Meat consumption per capita of Top 10 Tallest Countries

|  |  |
| --- | --- |
| **Country** | **Meat per capita (kg/person)** |
| Croatia | 66.3 |
| Czech Republic | 83.4 |
| Denmark | 95.2 |
| Germany | 88.1 |
| Iceland | 86.2 |
| Montenegro | 57.7 |
| Netherlands | 85.5 |
| Norway | 66 |
| Serbia | 45.2 |
| Bosnia & Herzegovina | 27.9 |

#### Meat consumption per capita of Top 10 Shortest Countries

|  |  |
| --- | --- |
| **Country** | **Meat per capita (kg/person)** |
| Bolivia | 59.1 |
| Cambodia | 16.6 |
| Ecuador | 56.4 |
| Indonesia | 11.6 |
| Nepal | 9.9 |
| Nigeria | 8.8 |
| Peru | 20.8 |
| Philippines | 33.6 |
| Sri Lanka | 6.3 |
| Vietnam | 49.9 |

### Milk Consumption

For milk consumption, the study looked to get a data frame that showed each country and the average amount of milk consumed per capita (kg/person). The data source came from <https://www.statista.com/statistics/272003/global-annual-consumption-of-milk-by-region/> as a CSV file which was then converted into a data frame. The procedure to get the list below was similar as the meat consumption data.

#### Milk consumption per capita of Top 10 Tallest Countries

|  |  |
| --- | --- |
| **Country** | **Milk Consumption (kg/person)** |
| Bosnia & Herzegovina | 197.24 |
| Croatia | 242.15 |
| Czech Republic | 190.93 |
| Denmark | 307.03 |
| Germany | 267.52 |
| Iceland | 255.76 |
| Montenegro | 389.91 |
| Netherlands | 340.35 |
| Norway | 252.82 |
| Serbia | 173.12 |

#### Milk consumption per capita of Top 10 Shortest Countries

|  |  |
| --- | --- |
| **Country** | **Milk Consumption (kg/person)** |
| Bolivia | 47.75 |
| Cambodia | 3.31 |
| Ecuador | 107.93 |
| Indonesia | 6.74 |
| Nepal | 54.12 |
| Nigeria | 2.18 |
| Peru | 56.68 |
| Philippines | 1.2 |
| Sri Lanka | 11.19 |
| Vietnam | 8.66 |

### Wealth

The GDP per capita data was found on the World Bank website. The information was imported from the CSV file, having GDP per capita for the years 2016 to 2019. All of the values were averaged, then the top 10 tallest countries and top 10 shortest countries were extracted.

#### Wealth of Top 10 Tallest Countries

|  |  |
| --- | --- |
| **Country** | **GDP/capita** |
| Bosnia & Herzegovina | $ 5,633.68 |
| Croatia | $ 13,886.56 |
| Czech Republic | $ 21,248.00 |
| Denmark | $ 58,254.46 |
| Germany | $ 45,086.60 |
| Iceland | $ 68,172.82 |
| Montenegro | $ 8,129.79 |
| Netherlands | $ 50,044.75 |
| Norway | $ 75,777.51 |
| Serbia | $ 6,672.28 |

#### Wealth of Top 10 Shortest Countries

|  |  |
| --- | --- |
| **Country** | **GDP/capita** |
| Bolivia | $ 3,382.11 |
| Cambodia | $ 1,452.52 |
| Ecuador | $ 6,188.34 |
| Indonesia | $ 3,857.48 |
| Nepal | $ 949.57 |
| Nigeria | $ 2,101.79 |
| Peru | $ 6,708.61 |
| Philippines | $ 3,233.52 |
| Sri Lanka | $ 3,974.25 |
| Vietnam | $ 2,459.93 |

# Results and Discussion

The data analysis involved many different comparisons (via scatterplot and correlation) of data. First, the top 10 tallest countries and top 10 shortest countries data frames were assessed individually. These studies included:

* Height versus Meat consumption per capita (Top 10 tallest countries only)
* Height versus Meat consumption per capita (Top 10 shortest countries only)
* Height versus Milk consumption per capita (Top 10 tallest countries only)
* Height versus Milk consumption per capita (Top 10 shortest countries only)
* Height versus GDP per capita (Top 10 tallest countries only)
* Height versus GDP per capita (Top 10 shortest countries only)

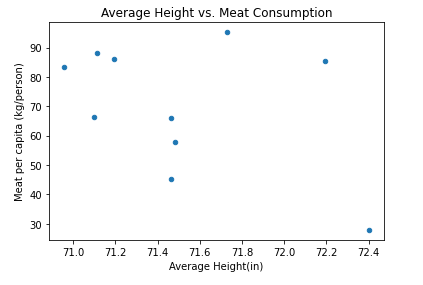
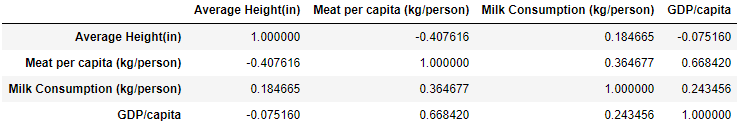


Figure . Average Height versus Meat Consumption per Capita for the top 10 tallest countries showed little to no correlation



As expected, the initial analysis did not show much correlation when only looking at the 10 countries on either extreme of the height chart.

Next, the data frame for the top 10 tallest countries was combined with the top 10 shortest countries and the same analysis was performed (scatterplots and correlation).

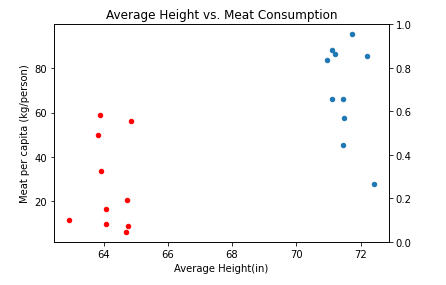
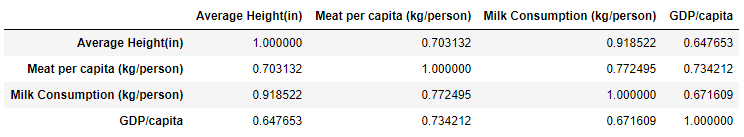
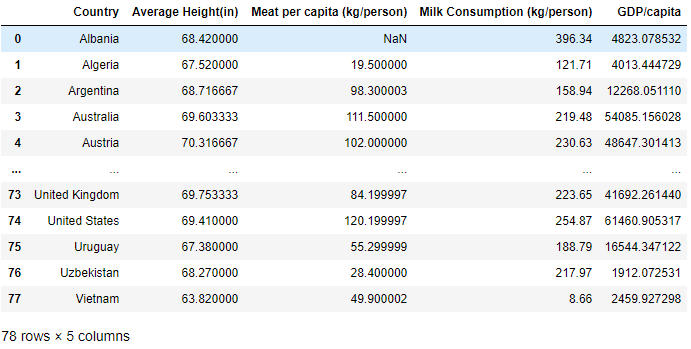


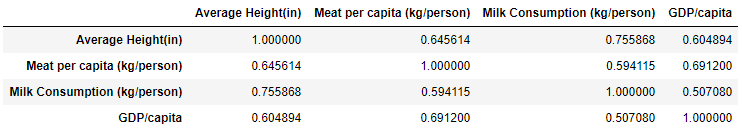
Figure . Once the 2 data frames were combined (Top 10 tallest and Top 10 shortest), they were plotted and positive correlation was more evident in all comparisons with height. This can also be seen in the correlation chart: strong positive correlation for both height versus meat consumption and height versus milk consumption.



Finally, the comparison was performed on all countries (this should have been done first as it would have saved the study from wasting time). Not enough data was available to have a full data frame for all 195 countries in the world. So, the final data frame was composed of only 78 countries.



We can see even when comparing 78 countries, there is a very strong positive correlation between height and milk consumption (>0.7) and a moderate positive correlation between height and meat consumption, wealth (<0.7 and >0.3).



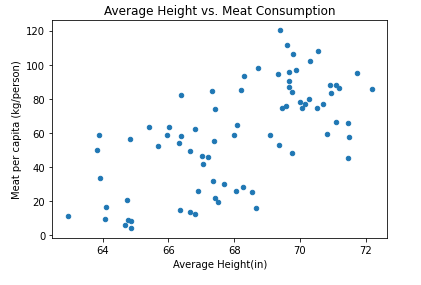


Figure . There is a very strong positive correlation between Average Height and Meat Consumption (via correlation chart and the scatter plot).

Based on the data, it looks like milk consumption has a direct positive correlation to height. Countries with higher milk consumption per person also tend to have taller people. This is also correlated with meat consumption and wealth, but not quite as strong.

# Conclusion

Based on the study results, there is evidence that there is a direct correlation between height and milk consumption. The more milk is consumed, the taller the people as told by the results. There is not enough evidence to suggest a direct correlation between meat consumption and wealth but there is still some positive correlation between the different factors. Also, the study is quite simple as it was a report that needed to be completed for a course within a time frame while I was juggling other work (e.g full-time job, life, etc.). I would like to take more time to perform other types of computational analysis.

# Future Next Steps

The data set was not as complete and the analysis not as strenuous as it could have been. The data was difficult to find for meat consumption per capita and milk consumption per capita. It would make for a stronger case had we found more data and allowed for correlation to be performed on more than 78 countries. This study also could have been biased as I assumed these factors would be correlated to height and potentially setup the study to confirm this. I would need to brainstorm other approaches. Also, more factors that can possibly contribute to height should be brainstormed and then studied (e.g. average amount of sleep – which is another set of data that is hard to find). There is more work to be done in figuring out why certain countries are so much taller on average than others.