

Covid-19 Report and Analysis

Introduction

This report analyzes data collected on 204 countries, from approximately February 24, 2020 through approximately October 03, 2021.

Body

Data

This report was written using MyJupyterNotebooks. First I imported Pandas, Matplotlib, Seaborn, and Geopandas to analyze and write reports on the data. Next I imported the Covid-19 data from a .csv file. Then I requested the number of rows and columns (121034 and 65 respectively), the data types of each column (all were float or string/"Object"), and the column names.

Method

First, I created the a dataframe for each variable that I wanted to visualize, for clarity and readability:

Vaccinations: location, total_cases, total_vaccinations, population (grouped by location)

Poverty/GDP: location, total_cases_per_million, gdp_per_capita, extreme_poverty (grouped by location)

Age: location, total_cases, median_age, aged_65_older, aged_70_older (grouped by location)

Smokers: location, total_cases, population, female_smokers, male_smokers (grouped by location)

Population: location, population_density (grouped by location)

Preexisting conditions: location, cardiovasc_death_rate, diabetes_prevalence (grouped by location)

For the **Vaccinations** and **Smokers** dataframes, I used the Population column to calculate a percentage of the population for the total_cases, female_smokers, and male_smokers fields. I did this by dividing each column by population, then multiplying by 100. Due to the nature of the data, I believed that this percentage would more accurately represent the population than number of cases and number of smokers alone.

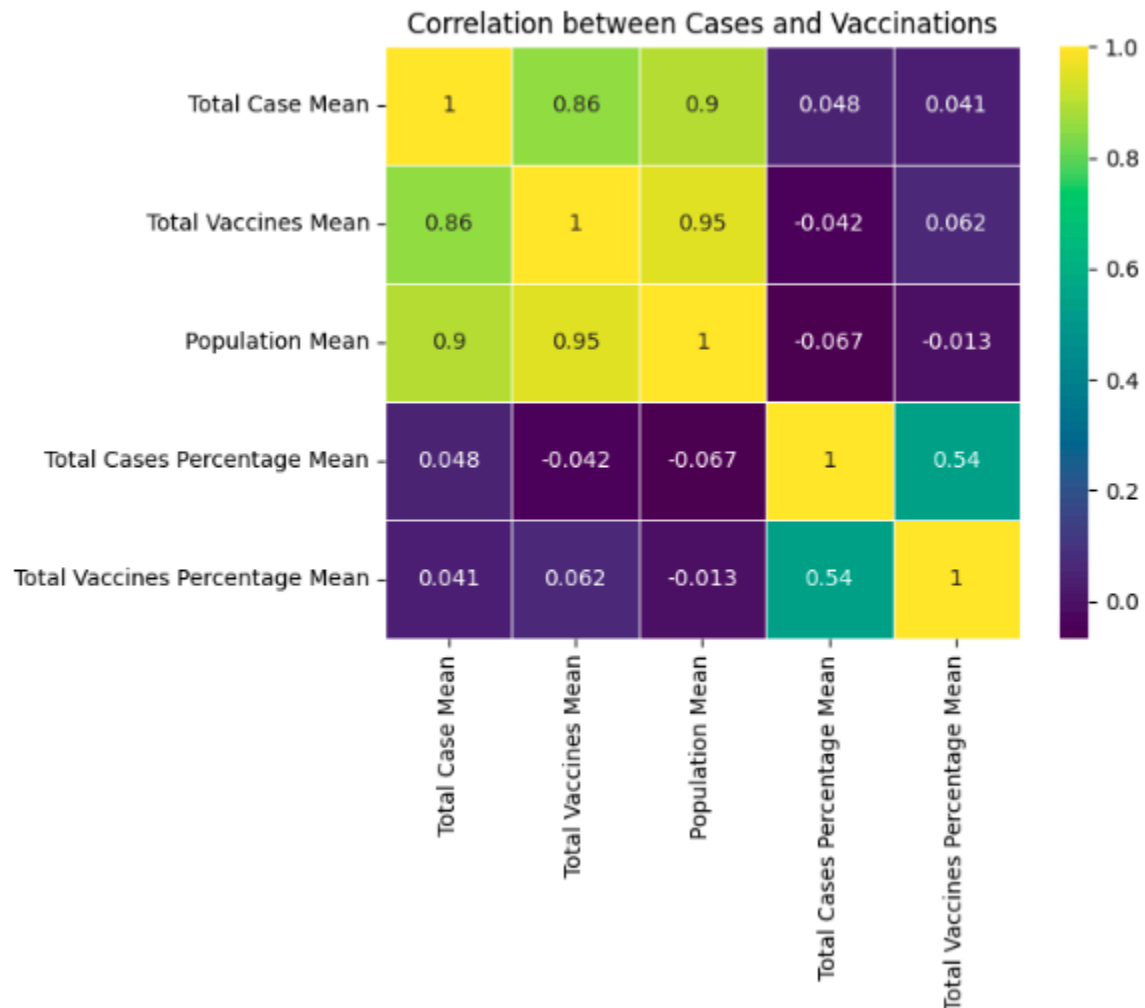
For all dataframes, I calculated the correlation coefficient between the columns in the dataframe, then created a heatmap to show the strength of the correlation. After doing this, I generated relational plots to visualize the data further, and then a longform plot to show the linear regression line for two variables at a time.

Results

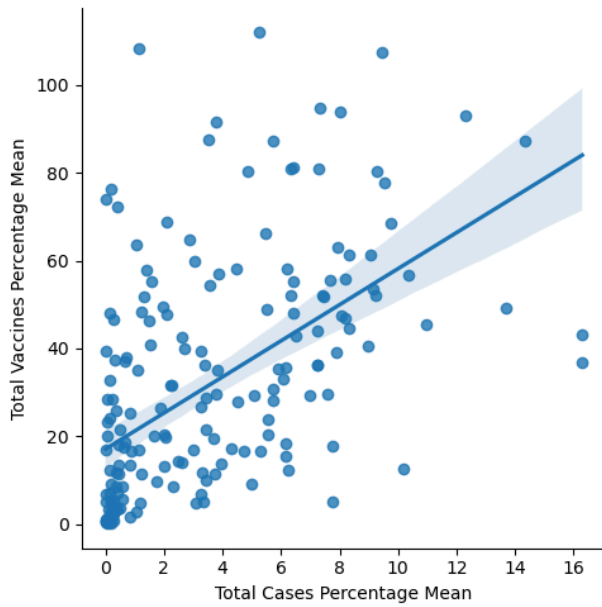
When using number of total cases as a comparison, I found that number of vaccinations, GDP, and number of smokers showed a correlation. Extreme poverty, median age, and preexisting conditions did not show a correlation.

Analysis

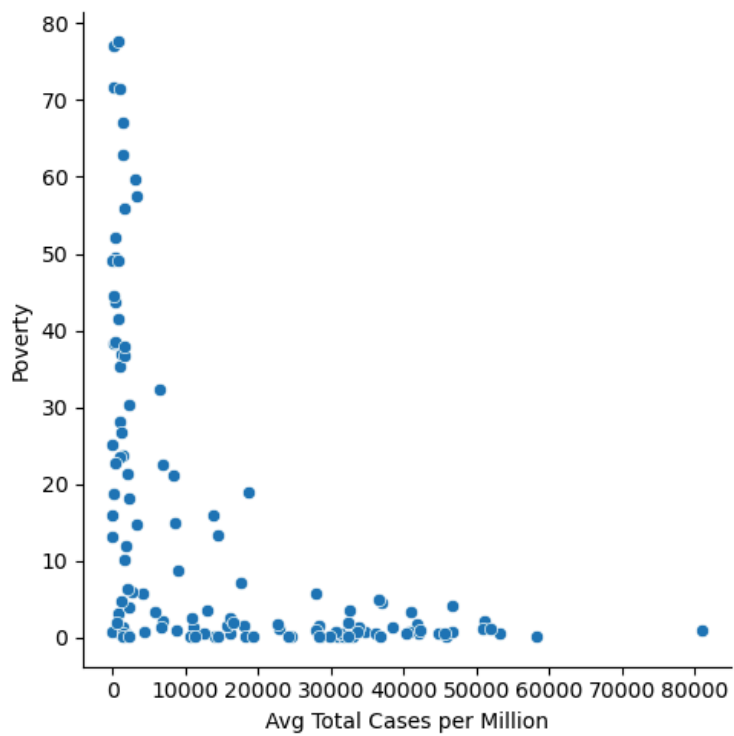
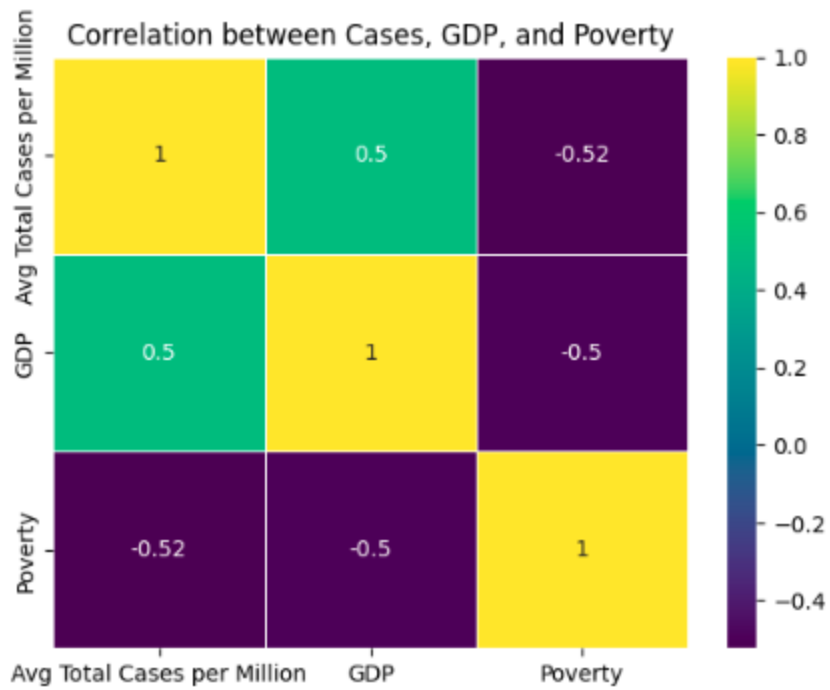
Heatmap of the correlation between percentage of vaccinations (by population) and percentage of total cases (by population):

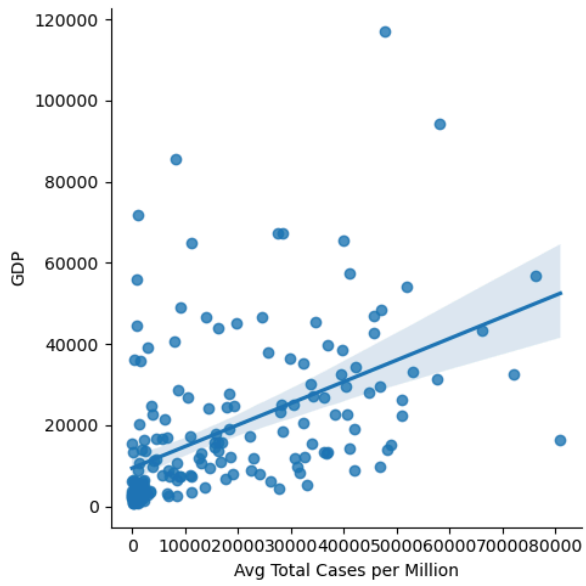


Longform plot of vaccinations vs. total cases:

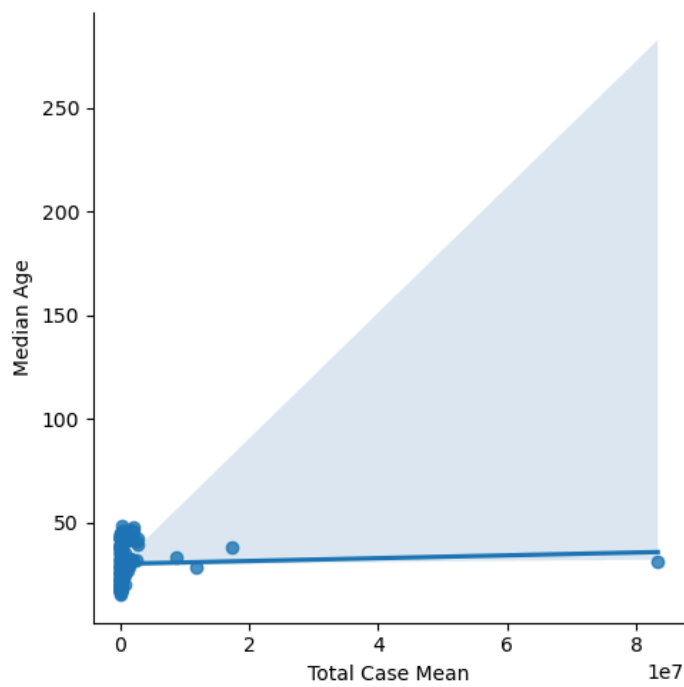


Heatmap showing correlation between total cases, country GDP, and extreme poverty; relational plot showing correlation between total cases and extreme poverty; and longform plot showing correlation between total cases and GDP:

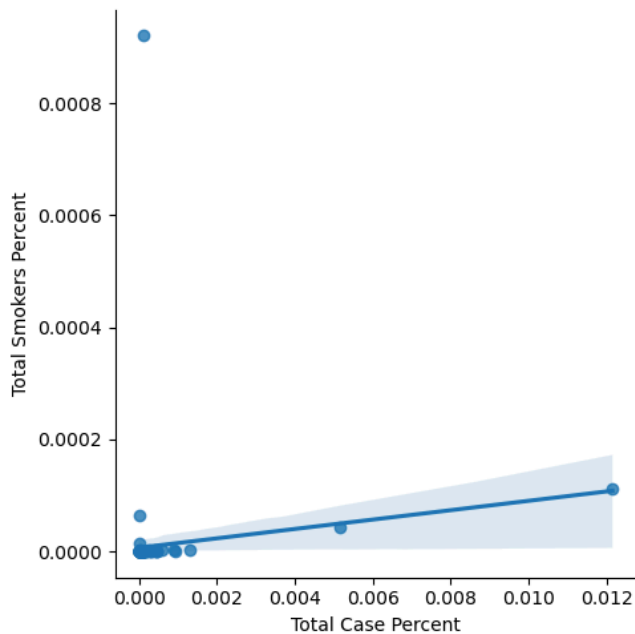




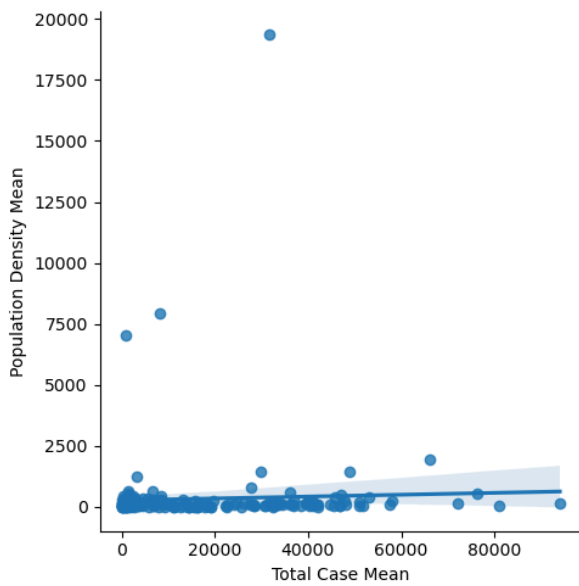
Longform plot showing correlation between total cases and median age:



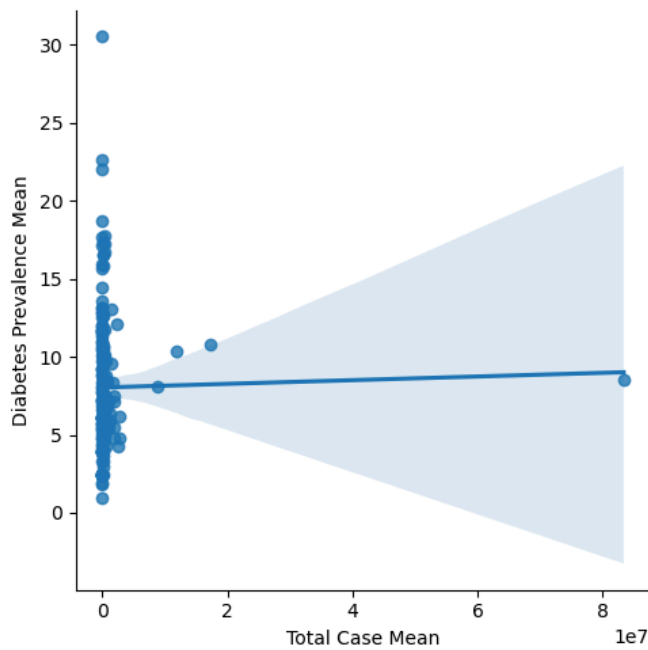
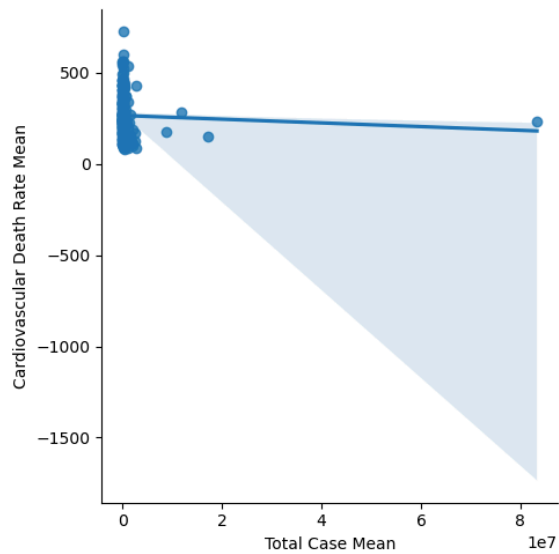
Longform plot showing correlation between percentage of total cases and smokers:



Longform plot showing comparison of cases per million and population density:



Longform plots showing comparison of total cases, cardiovascular deaths, and diabetes prevalence:



Conclusion

There is a notable correlation between percentage of vaccinations given and percentage of total cases. There are also slight correlations between country GDP per capita and total cases, and percentage of total smokers vs. percentage of total cases.