COVID-19 Assiciative Plots

untitled

03/04/2021

Contents

Introduction

Our World in Data COVID-19 data set contains up-to-date data on confirmed cases, deaths, hospitalizations, testing, and vaccinations as well as other variables of potential interest. The data set was derived from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University and can be found here: https://github.com/CSSEGISandData/COVID-19. Analyzing this data set will give us some incredible insights into how the covid-19 virus has affected several countries across the globe. We will explore such insights through exploratory data analysis right after we clean and preprocess the data.

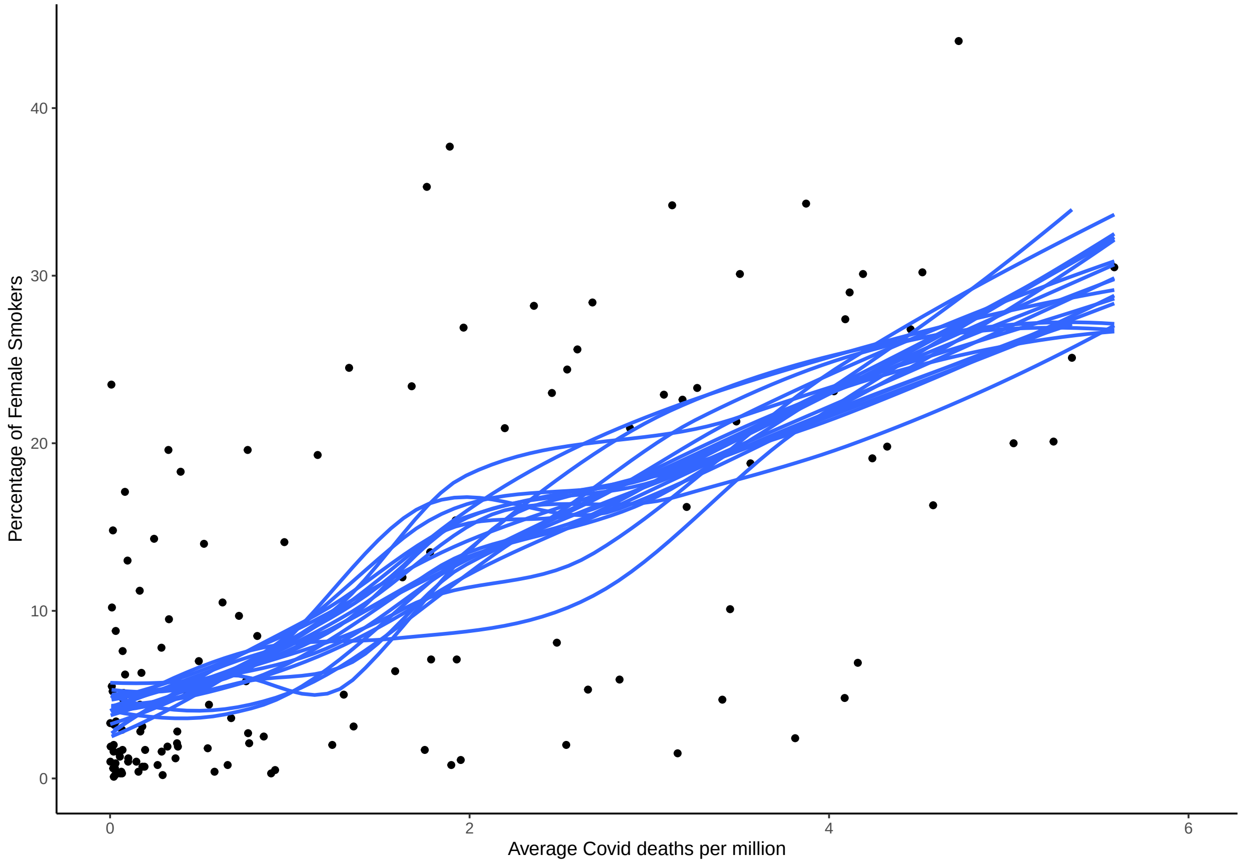
Data Structure

In this data set, we have a mix of character and numeric variables. It has 59 variables and 70,645 observations.

Exploratory Data Analysis

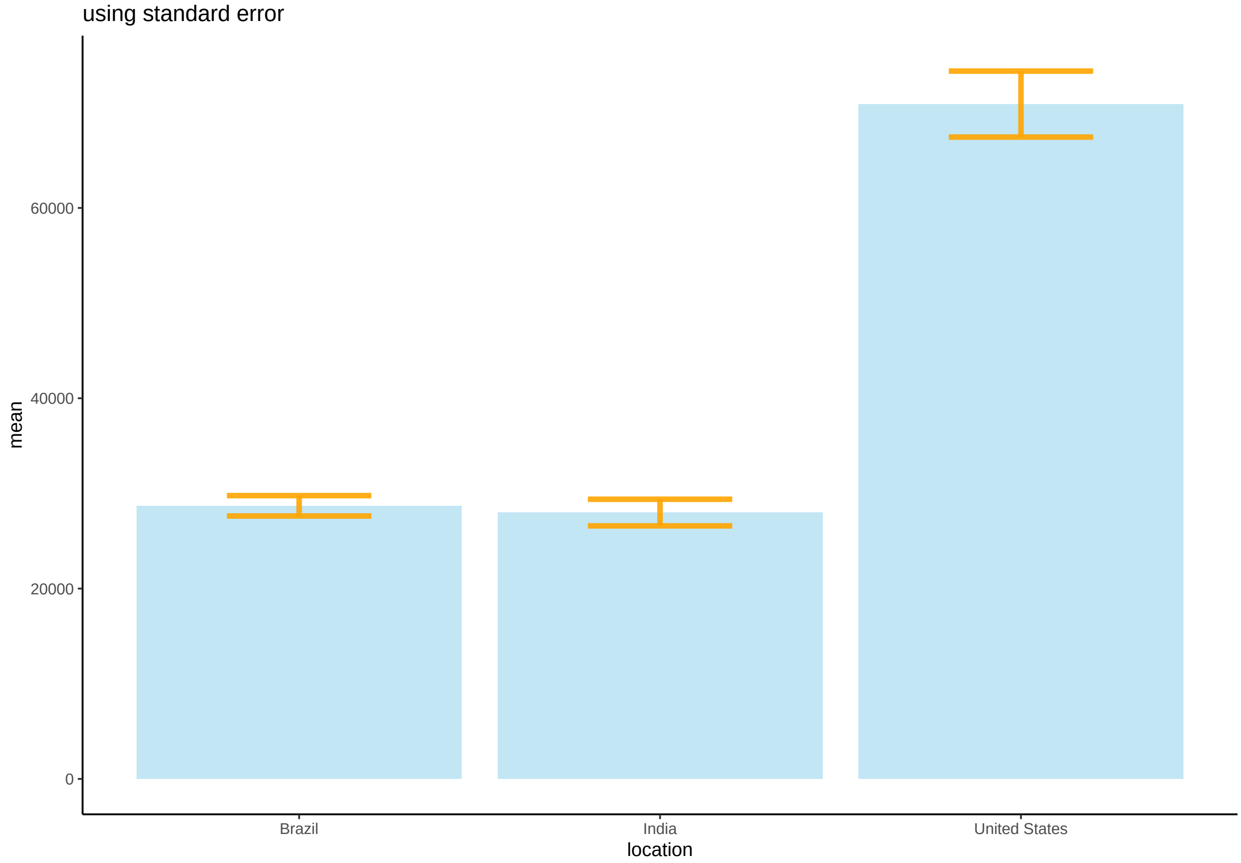
A. Average Covid deaths per million vs Percentage of Female Smokers

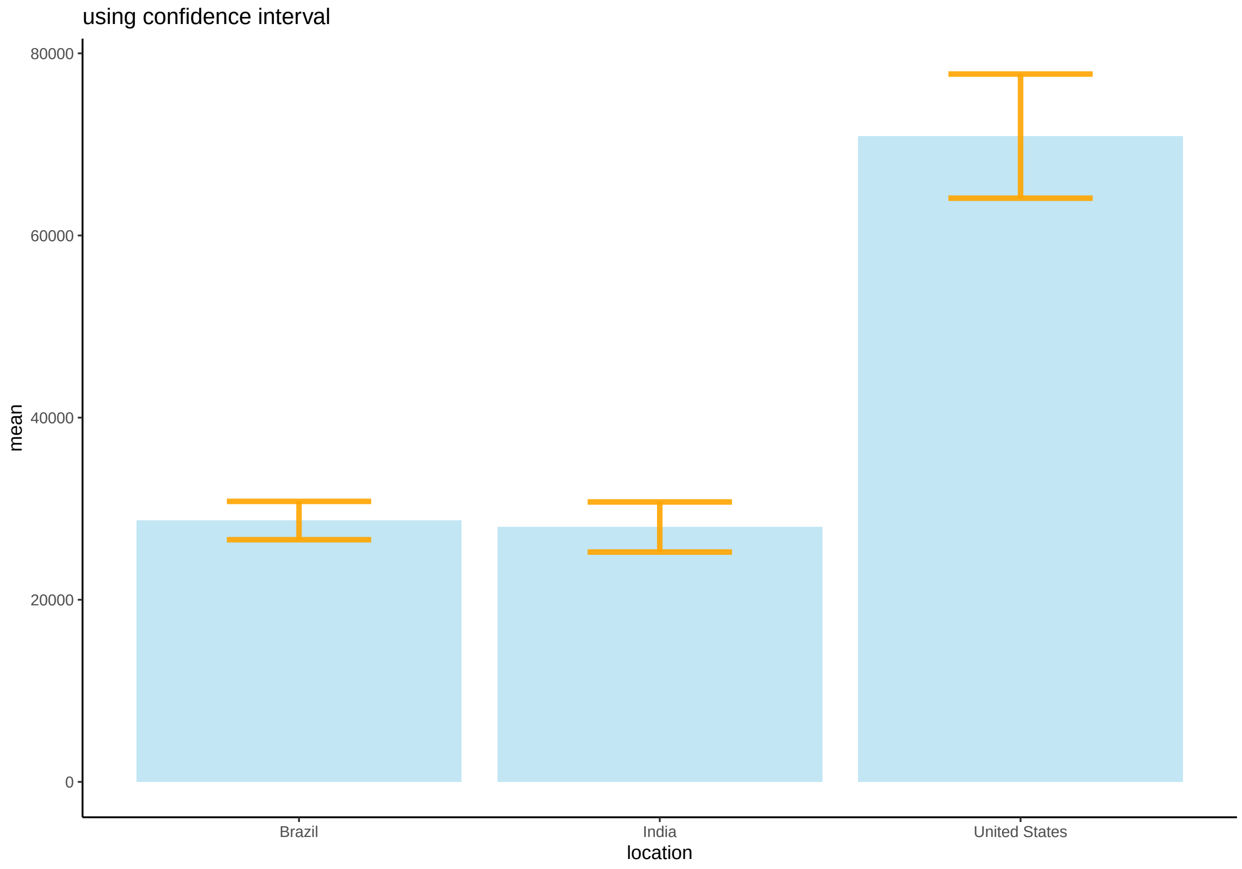
In the diagram below we show the relationship between average covid deaths per million and the percentage of female smokers in a country. There is a positive relationship between these two variables meaning countries with higher female smokers suffer more covid deaths on average whiles countries with a lower proportion of female smokers suffer fewer covid deaths per million on average. The non-linear curve fits demonstrate this relationship and shows equally likely alternative fits drawn from the posterior distribution.



B. Average Cases for the Top Three Infected Countries

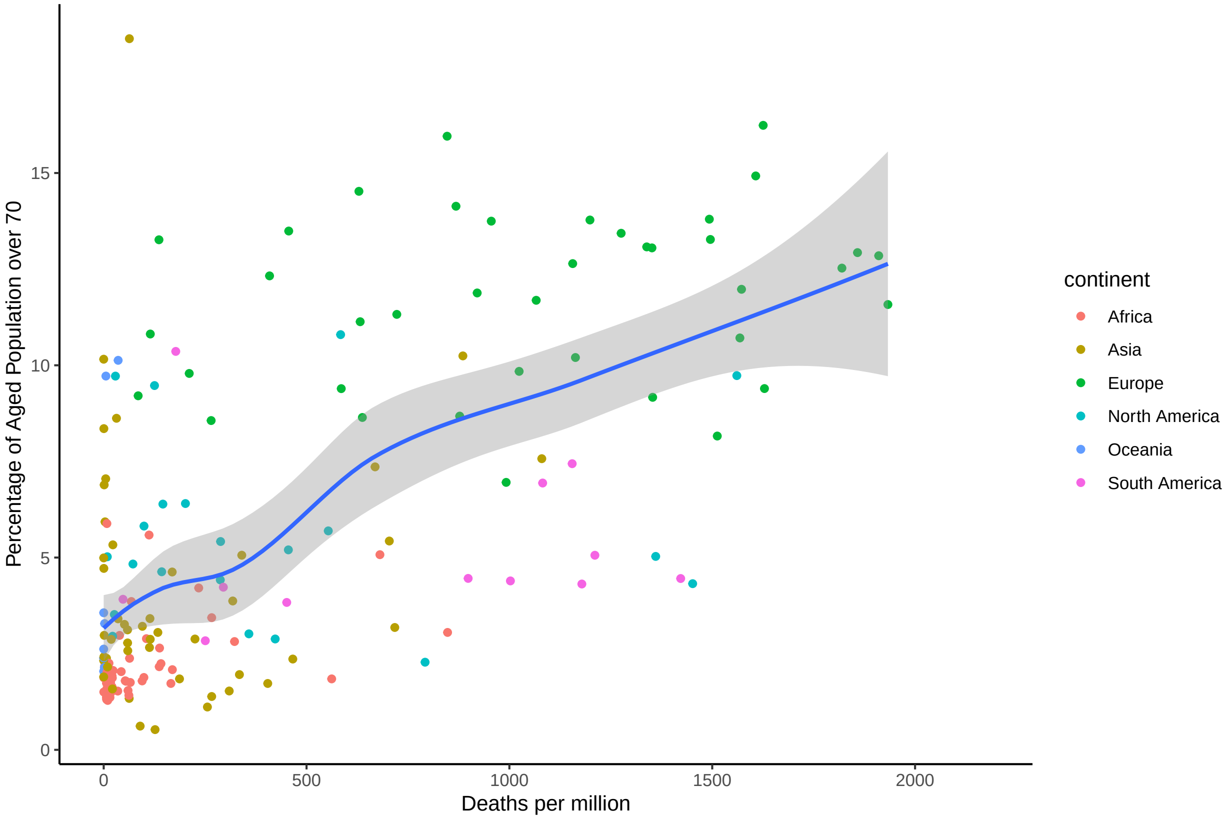
This diagram shows the average cases for the top three covid infected countries. These countries are the USA, India and Brazil. The USA takes the lead with the highest case count followed by Brazil and India. Additionally, error bars indicating the standard error and confidence intervals were included. The standard error is the standard deviation of the vector sampling distribution. The confidence interval on the other hand is defined so that there is a specified probability that the mean positive cases for the country lie within it.




using confidence interval
80000
60000
40000
20000
0
mean
Brazil
India
United States
location

C. Deaths per Million vs Percentage of Aged Population Over 70

In this diagram, we see a direct relationship between covid deaths per million and the percentage of the aged population over 70 years old. This implies that countries with a more aged population suffer more deaths with countries having a less proportion of the aged population over 70 experiencing fewer deaths. European countries that have a high aged population, for instance, suffer the most deaths with the relatively younger population of Africa having fewer deaths. The fit spline illustrates the positive relationship between these two variables and the confidence band above and below the fitted line represent confidence intervals and uncertainty in an estimate of a curve.


Percentage of Aged Population over 70
Deaths per million
continent
Africa
Asia
Europe
North America
Oceania
South America