

Question 1

Peter Meihuizen^a

^a*Stellenbosch University, South Africa*

Keywords: Multivariate GARCH, Kalman Filter, Copula

JEL classification L250, L100

1. Introduction

For this question I am looking at how different continents were affected by Covid-19.

2. Analysis

Ok so firstly I need to import the data.

I want to compare how Covid affected African countries relative to other continents. In order to do this I'm going to compare various COVID-19 statistics over 6 month intervals for every continent. Therefore I will divide the times period into 5 periods: 2020-01, 2020-07, 2021-01, 2021-07 and 2022-01. I will give the mean value for each of these continents over each 6 month period.

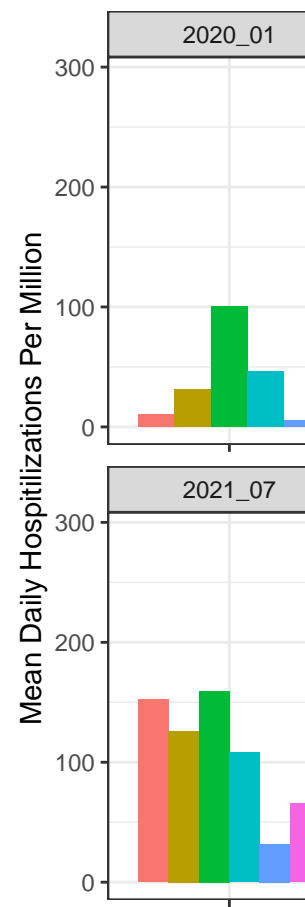
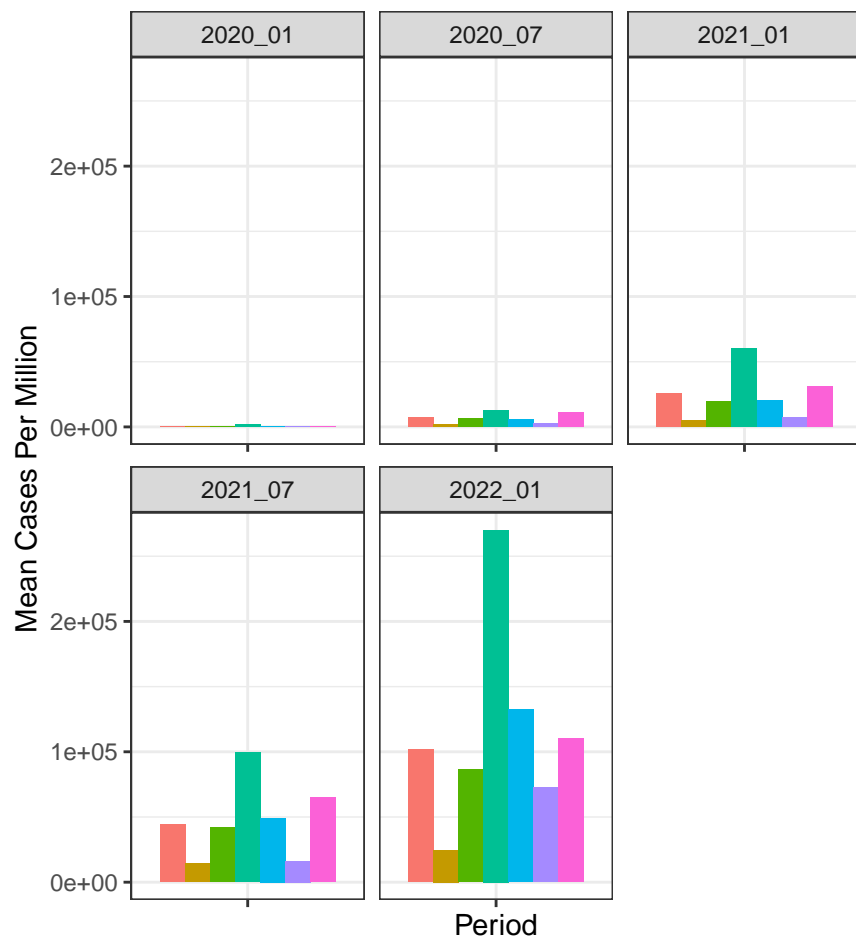
In order to do this I need to build a function which calculates the mean value for a particular country over this a particular 6 month period.

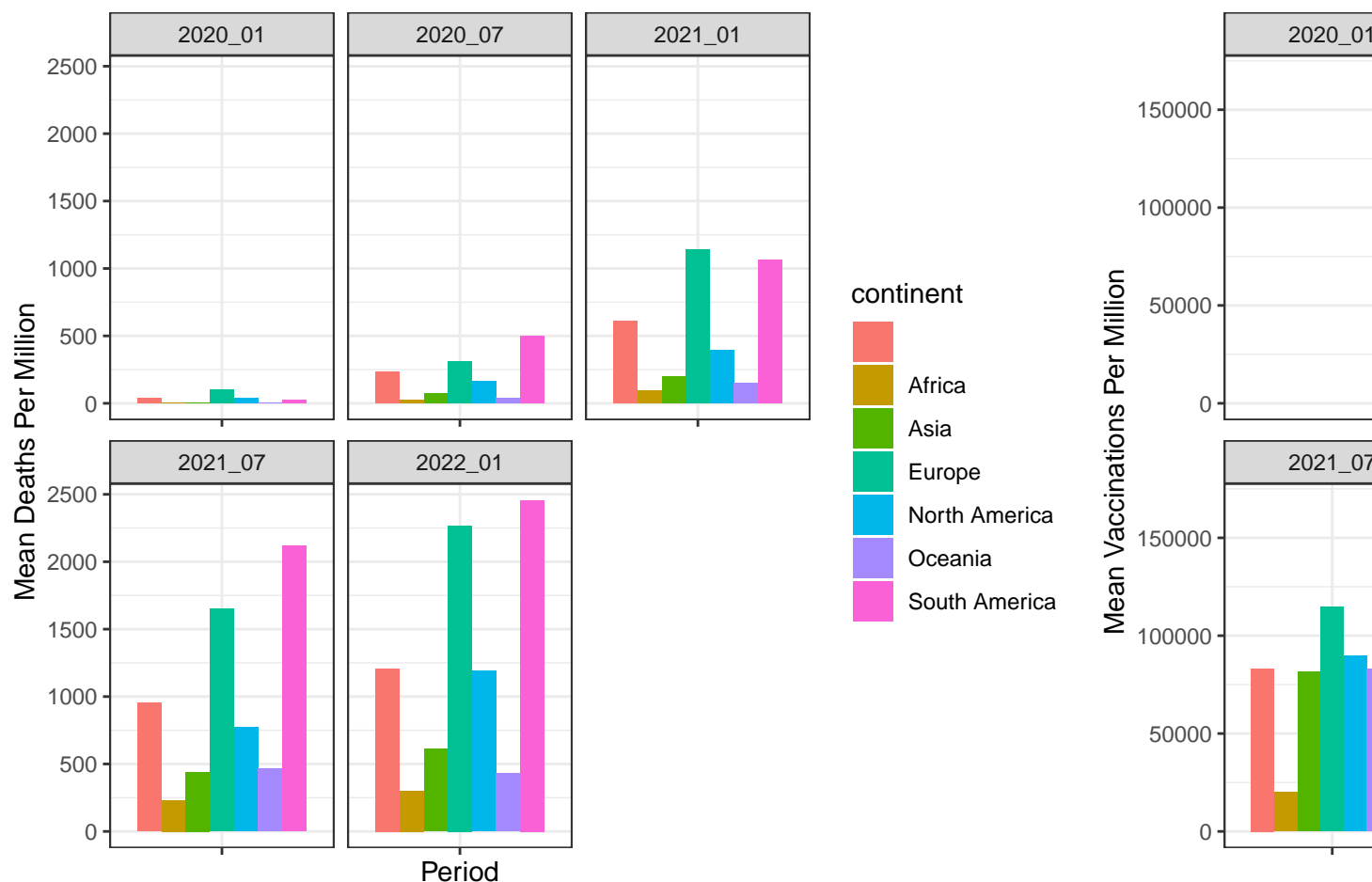
I want to look at the mean value for `total_cases_by_million` for each country so I'll use the function `calculate the mean` for each country based on each period.

Ok so this function is giving me errors and is not working so I'm going to try simplify it to calculate the `mean_total cases` for each country in that particular period, first I'm going to add a period column to the main data frame indicating the 5 different periods.

Now I generate the bargraphs showings the growing rates of the average cases, hospitalisations, deaths and vaccinations.

Email address: 21831041@sun.az.za (Peter Meihuizen)





Ok so as can be seen in the results Africa is the least affected by Covid on all accounts. Of all the continents it has the lowest number of cases, hospitalisations, deaths and vaccinations per million of every continent. It appears the more developed continents, Europe and North America are the most significantly affected.

For the next part of my analysis I want to look at the correlation between smoking and hospitalisations to see if countries with an higher rate of smoking are more likely to be hospitalised. I need to import the death_by_cause data to look at smoking related deaths.

Because Europe was most significantly affected by Covid I want to only look at European countries to do my comparison, I therefore want to look at those countries which had the greatest level of lower respiratory infections as these can cause higher death rates amongst smokers as they have weaker lungs. Therefore I need to look at the European countries with the 3 greatest number of average daily hospitalisations and the 3 lowest. I will find the hospitalisations and populations of all measured European countries.

```
## # A tibble: 30 x 3
##   location mean_hosp mean_pop
##   <chr>      <dbl>    <dbl>
## 1 Bulgaria    469.  6896655
## 2 Serbia      419.  6871547
## 3 Lithuania   362.  2689862
## 4 Romania     326. 19127772
## 5 Hungary     290.  9634162
## 6 France      265. 67422000
## 7 Poland      255. 37797000
## 8 Latvia      231.  1866934
## 9 Croatia     228.  4081657
## 10 Czechia    220. 10724553
## # i 20 more rows
```

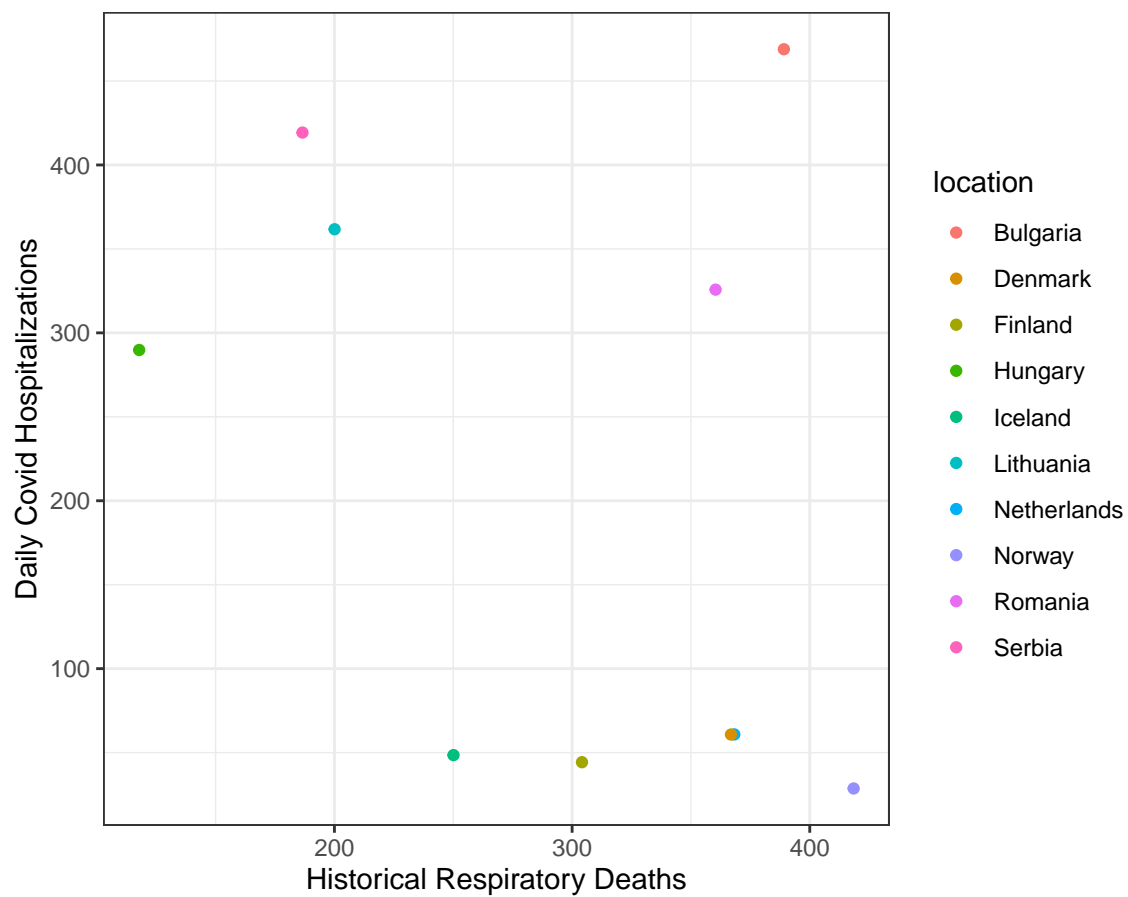
Ok so based on this it shows that the highest daily hospitalisations for Covid in Europe over the five periods were in Bulgaria, Serbia, Lithuania, Romania and Hungary whilst the lowest recorded daily hospitalisations were in Norway, Finland, Iceland, Denmark and the Netherlands.

Now I need to look at each countries prevalence for smoking by averaging their respiratory infection deaths from the Deaths_by_cause data set.

Now I need to combine these two sets to compare respiratory deaths with covid hospitalisation and create a respiratory deaths per population variable.

Now that I have the average number of Covid hospitalisations and the number of lower respiratory infection deaths per million of the population I can create a scatter plot to see if there was any correlation.

Scatterplot of European Countries Covid Hospitalizations and Historic



```
## # A tibble: 10 x 3
##   location    mean_hosp percent_resp_deaths
##   <chr>      <dbl>          <dbl>
## 1 Bulgaria    469.            389.
## 2 Serbia      419.            187.
## 3 Lithuania   362.            200.
## 4 Romania     326.            360.
## 5 Hungary     290.            118.
## 6 Netherlands  60.8            368.
## 7 Denmark     60.8            367.
## 8 Iceland     48.5            250.
## 9 Finland     44.3            304.
## 10 Norway      28.6            418.
```

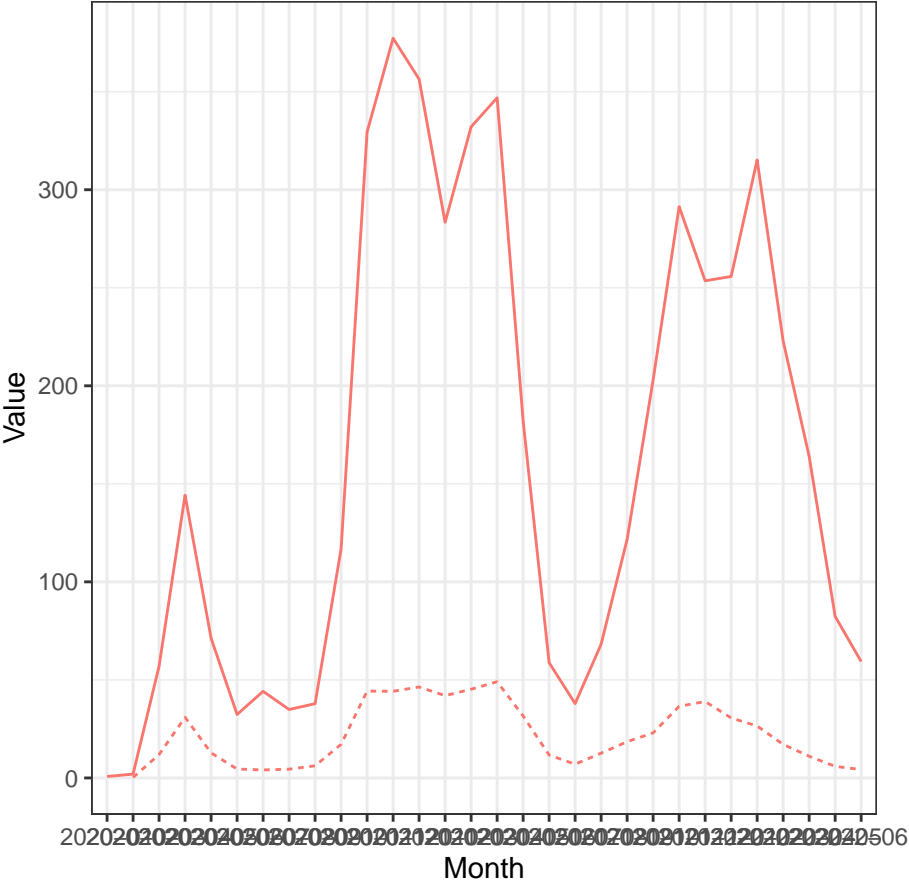
So as can be seen there is almost no correlation between the historical respiratory deaths and daily Covid hospitalisations for countries in Europe. This would suggest that countries with historically high levels of smoking are unlikely to have more or less covid-related hospitalisations.

For the final part of the question I'm going to look at hospitalisation numbers and ICU numbers for each continent and see if the one led or lagged the other. Ok so first I'm going to group the data by continent and by month.

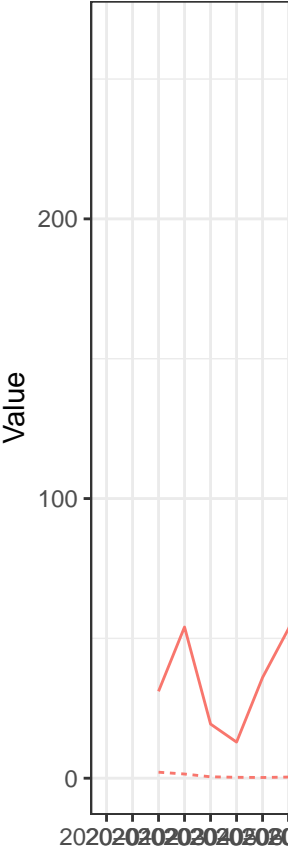
```
## # A tibble: 209 x 4
##   continent year_month cont_hosp cont_ICU
##   <chr>      <chr>      <dbl>    <dbl>
## 1 ""        2020-01         NaN     NaN
## 2 ""        2020-02         NaN     NaN
## 3 ""        2020-03         NaN     NaN
## 4 ""        2020-04         NaN     NaN
## 5 ""        2020-05         NaN     NaN
## 6 ""        2020-06         NaN     NaN
## 7 ""        2020-07         NaN     NaN
## 8 ""        2020-08         NaN     NaN
## 9 ""        2020-09         NaN     NaN
## 10 ""       2020-10         NaN     NaN
## # i 199 more rows
```

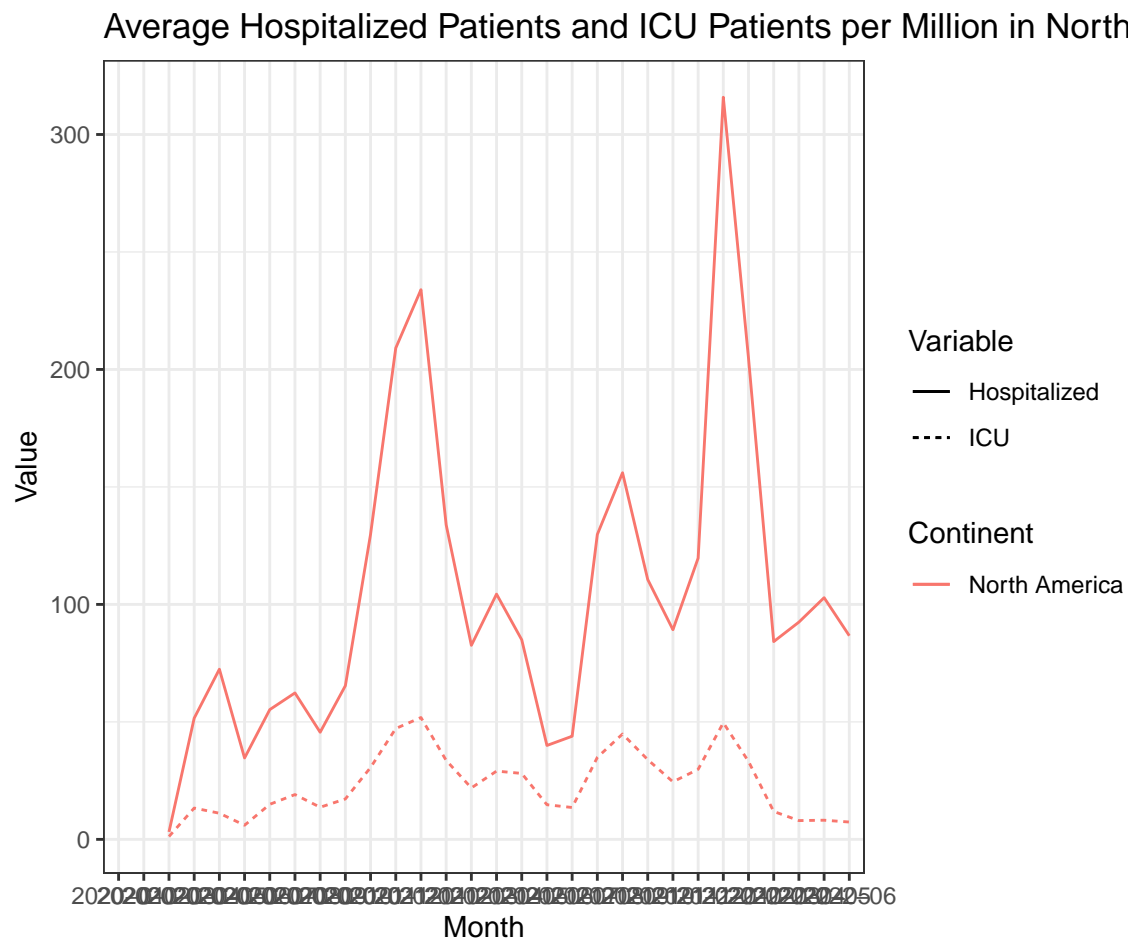
Now I want to plot the data to see if hospitalisations number lead or lag ICU numbers.

Average Hospitalized Patients and ICU Patients per Million in Europe



Average Hos





Therefore it appears that the level of hospitalizations, which is being used as a proxy for hospital capacity, coincides with an increase in ICU admissions for Europe, Asia and North America when looking at the monthly admissions.