Do many Physicians keep you healthy?

Benno Weber

Dataset used

This analysis and report is based on the following dataset:

- World Development Indicators Dataset (WDI-Dataset)

Motivation

The WDI-Dataset offers a great opportunity to analyse developments regarding health issues across countries and to see how they changed over the last couple of years.

This report presents findings from exploratory data analysis on the WDI-Dataset and analyses how the number of Physicians per 1000 people influences different indicators regarding a population’s health such as life expectancy at birth, survival rate to the last grade of primary education and fertility rate.

Insights could be used to identify countries suffering from lack of highly educated medical personnel and thereby providing help in forming strategies to enhance health indicators.

Do many Physicians keep you healthy?

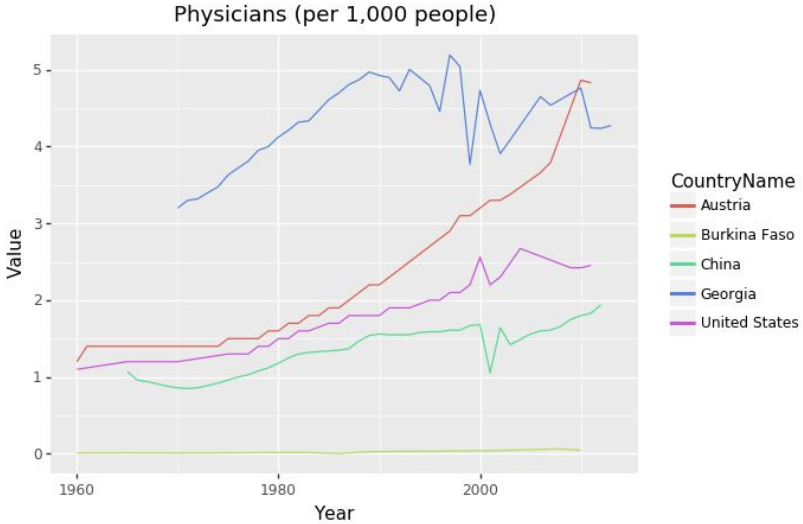
This report explores how the number of physicians per 1000 people in a population influences the following three indicators regarding a population's health:

- life expectancy at birth
- survival rate to the last grade of primary education
- fertility rate, total (births per woman)

The WDI-dataset was split so that data of 5 countries including Austria, USA, China, Georgia and Burkina Faso where analysed.

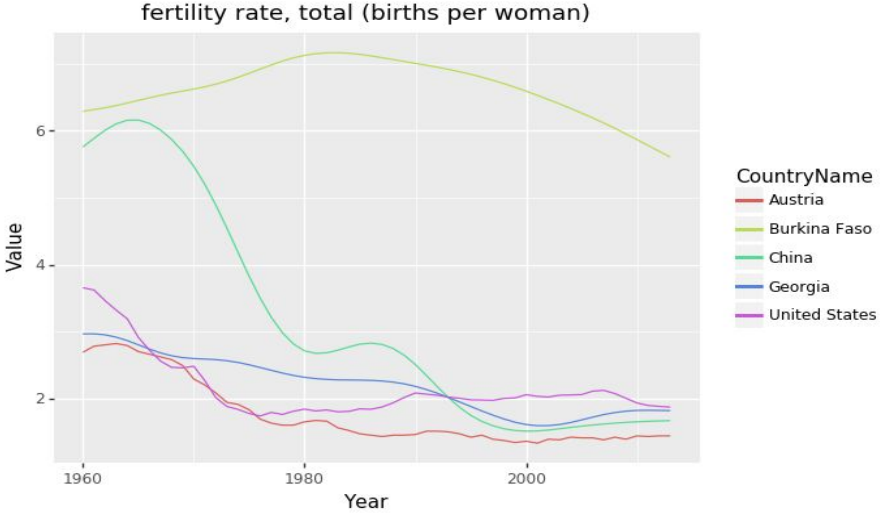
Finding 1 - The number of Physicians differs greatly

Where as Burkina Faso barely reaches a number of 0.12 physicians/1000 people in recent years, Austria and Georgia show numbers much higher than the worldwide average (~1.58)



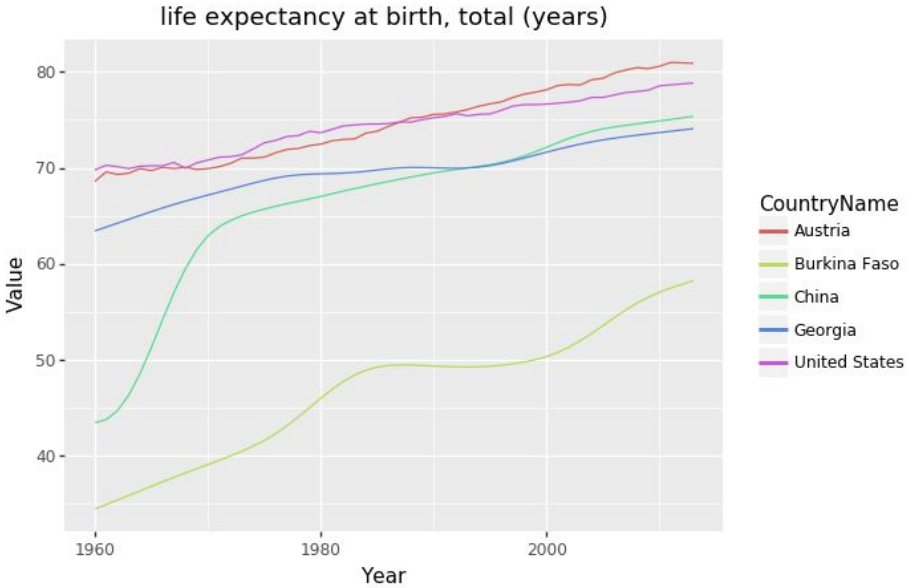
Finding 2 - Fertility rates decrease world wide but at different speeds

Fertility rates in Austria, China, Georgia and the US remain relatively stable in the past 30 years after rapidly decreasing in the 1960's. Rates in Burkina Faso also decrease but are still very high in comparison.



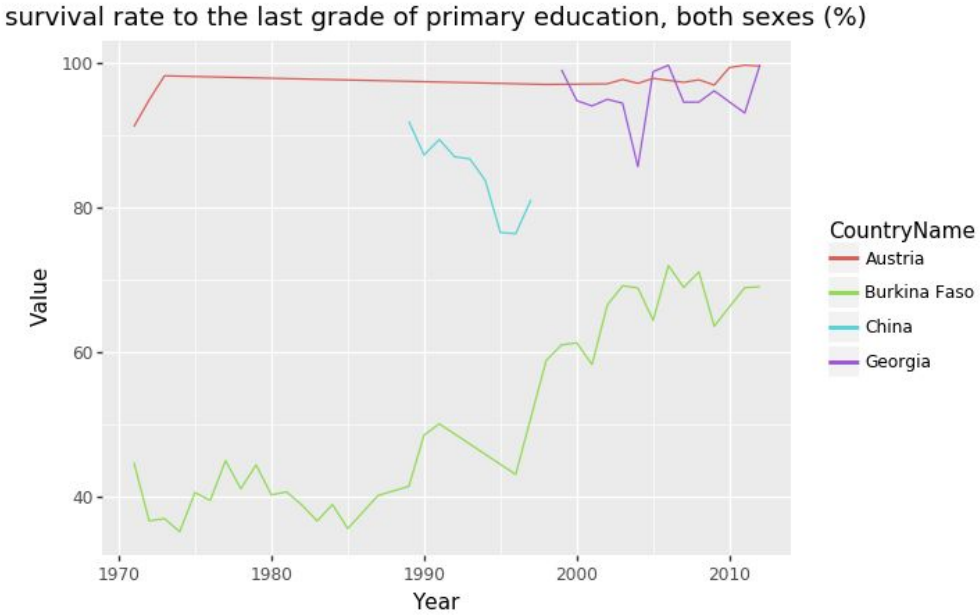
Finding 3 - People tend to live longer

Life expectancy is on the rise in all countries.



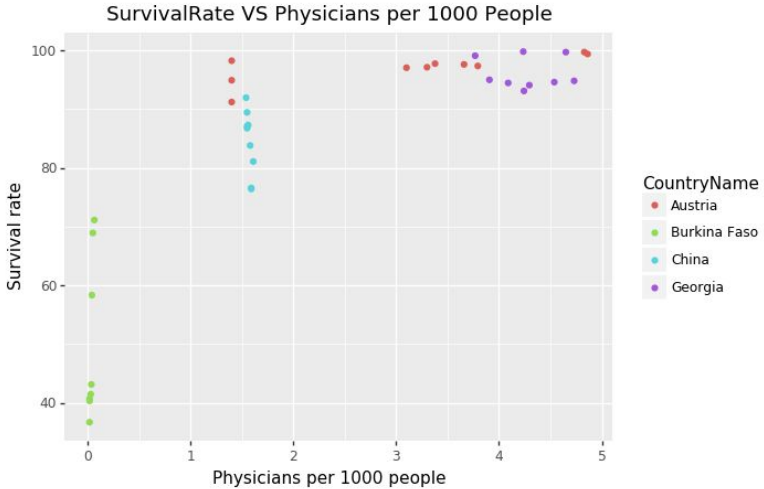
Finding 4 - Survival rates seem to rise

Unfortunately the dataset includes some missing values but it seems as if Survival rates are rising in all countries observed.



Finding 5 - Correlation of Survival and number of Physicians

The correlation coefficient between number of physicians and survival rate is approx. 0.793 which suggest that there is a correlation between those variables. Plotting survival rate against number of physicians and grouping by country, reveals this connection further.



Acknowledgements

I didn’t receive feedback from anyone regarding my work for this report.

References

The work conducted for this report was conducted by myself alone.

In [4]:

```
import pandas as pd
import numpy as np
from pandas.api.types import CategoricalDtype
from plotnine import *
%matplotlib inline
```

In [6]:

```
data = pd.read_csv('./Indicators.csv')
```

In [7]:

```
countries = data["CountryName"].unique().tolist()
indicators = data["IndicatorName"].unique().tolist()
years = data["Year"].unique().tolist()
```

In [8]:

```
phys= [s for s  in indicators if "Physi" in s]
#only keep countries where Physicians(per 1000 people) data is available
maskPhysician = data["IndicatorName"]==phys[0]
#How many observations do we have?
maskPhysician.sum()
#How many countries reported data?
physicianDF=data[maskPhysician]
len(physicianDF["CountryName"].unique())
#whats the average value of Physicians per 1000 people
physicianDF["Value"].mean()
```

Out[8]:

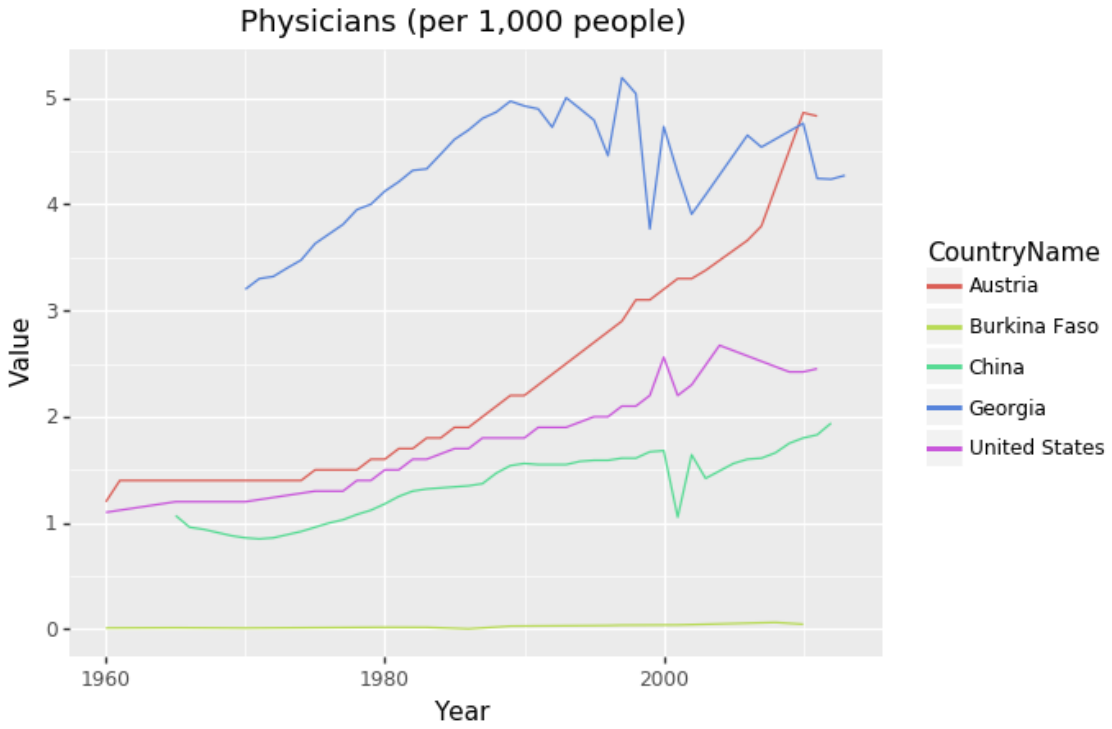
```
1.5858357056558288
```

In [9]:

```
#select 5 Countries for comparison
countryComparison= ("Austria","United States","Burkina Faso","China","Georgia")
maskFiveCOuntries = physicianDF["CountryName"].isin(countryComparison)
fiveCountriesPhysicianDF = physicianDF[maskFiveCOuntries]
```

In [10]:

```
(ggplot(fiveCountriesPhysicianDF)         # defining what data to use
 + aes(x='Year',y="Value", colour="CountryName")    # defining what variable to use
 + geom_line() # defining the type of plot to use
 + ggtitle(phys[0])
)
```



Out[10]:

<ggplot: (-9223371935169535756)>

In [11]:

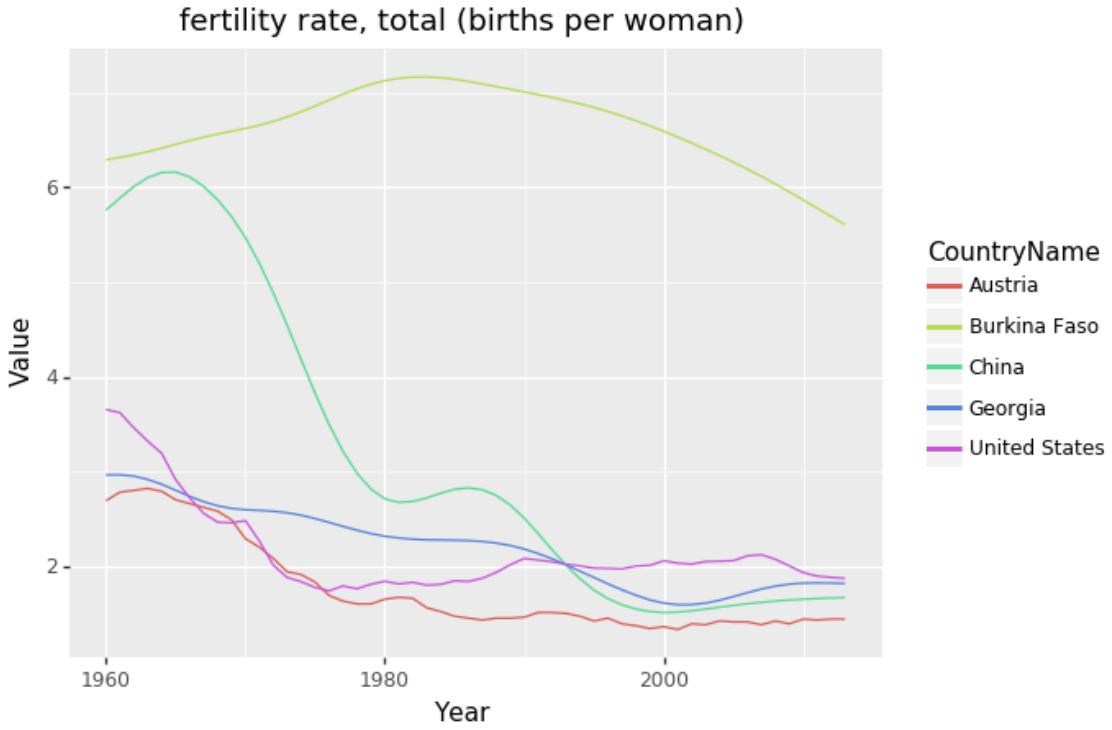
```
#Look for interesting comparison indicators
indicators=[x.lower() for x in indicators]
healthMatch = ("life expectancy at birth, total (years)","survival rate to the last gra
de of primary education, both sexes (%)","fertility rate, total (births per woman)")
healthIndicators = [s for s in indicators if any(xs in s for xs in healthMatch)]
data["IndicatorName"]= [x.lower() for x in data["IndicatorName"]]
maskHealthIndicators = data["IndicatorName"].isin(healthMatch)
#How many Datapoints do we have?
maskHealthIndicators.sum()
maskFiveCOuntriesData = data["CountryName"].isin(countryComparison)
healthDF = data[maskHealthIndicators & maskFiveCOuntriesData]
```

In [12]:

```
#Splitt HealthDF in 3 Dataframes for different Indicators
FertilityDF = healthDF[healthDF["IndicatorName"]==healthIndicators[0]]
LifeExpectancyDF = healthDF[healthDF["IndicatorName"]==healthIndicators[1]]
SurvivalDF = healthDF[healthDF["IndicatorName"]==healthIndicators[2]]
```

In [13]:

```
#Visualize fertility rate
(ggplot(FertilityDF)         # defining what data to use
 + aes(x='Year',y="Value", colour="CountryName")    # defining what variable to use
 + geom_line() # defining the type of plot to use
 + ggtitle(healthIndicators[0])
)
```

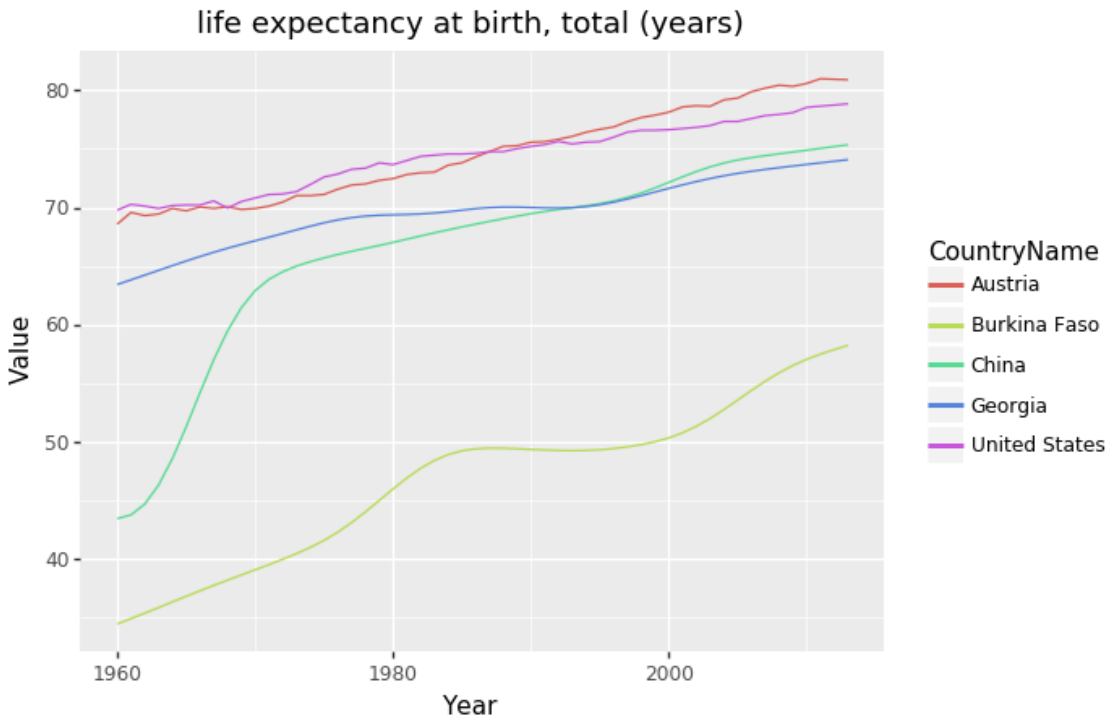


Out[13]:

<ggplot: (-9223371935133173252)>

In [16]:

```
#Visualize Life Expectancy
(ggplot(LifeExpectancyDF)         # defining what data to use
 + aes(x='Year',y="Value", colour="CountryName")    # defining what variable to use
 + geom_line() # defining the type of plot to use
 + ggtitle(healthIndicators[1])
)
```

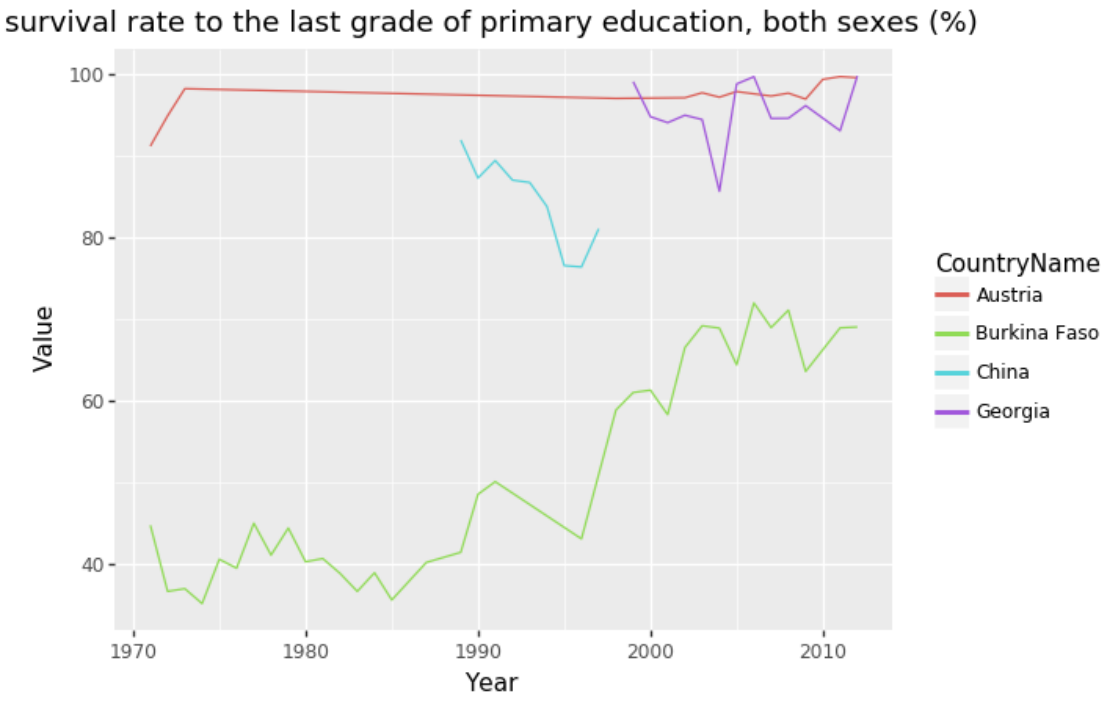


Out[16]:

<ggplot: (-9223371935176205652)>

In [15]:

```
#Visualize Survival rate
(ggplot(SurvivalDF)         # defining what data to use
 + aes(x='Year',y="Value", colour="CountryName")    # defining what variable to use
 + geom_line() # defining the type of plot to use
 + ggtitle(healthIndicators[2])
)
```



Out[15]:

```
<ggplot: (-9223371935181839372)>
```

In [17]:

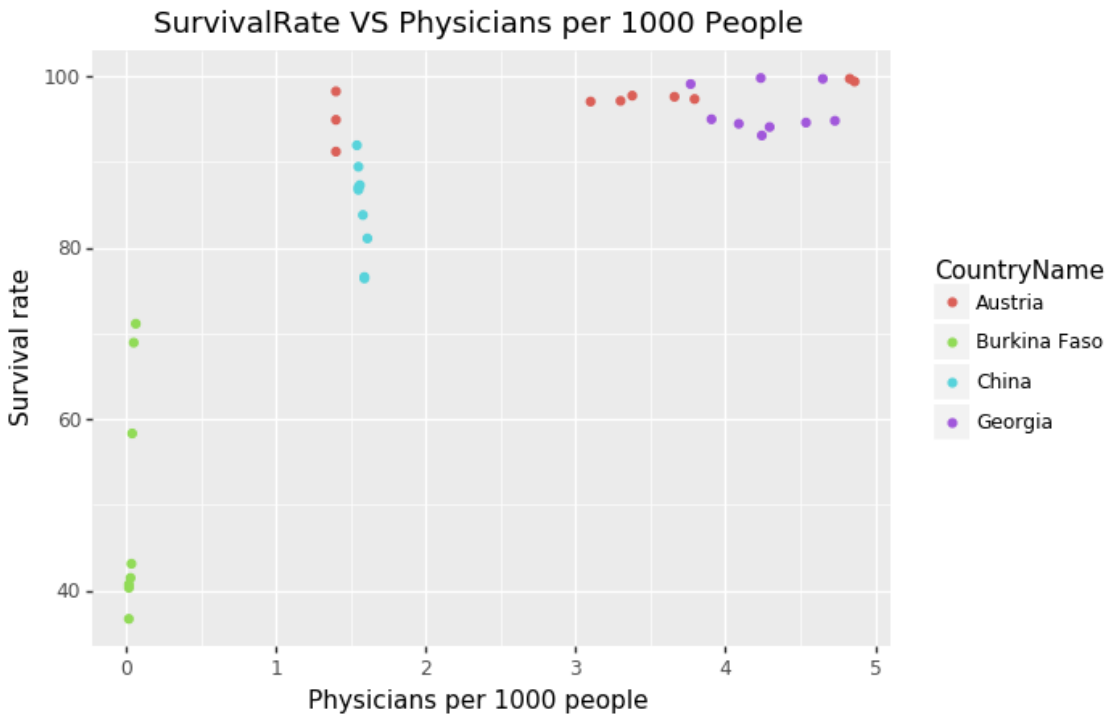
```
#Is there a correlation between Number of Physicians and SurvivalRate?
mergedSurvival = pd.merge(SurvivalDF,fiveCountriesPhysicianDF,on=["CountryName","Year"
])
mergedSurvival["Value_x"].corr(mergedSurvival["Value_y"])
```

Out[17]:

```
0.7934505357762084
```

In [18]:

```
#Visualize the correlation
(ggplot(mergedSurvival)         # defining what data to use
 + aes(x='Value_y',y="Value_x", colour="CountryName")    # defining what variable to us
e
 + geom_point() # defining the type of plot to use
 + ggtitle("SurvivalRate VS Physicians per 1000 People")
 + labs(x = "Physicians per 1000 people", y = "Survival rate")
)
```



Out[18]:

```
<ggplot: (-9223371935180189140)>
```

In [ ]: