Data Analysis Project

Mohamed Mohsen
Abdelrahman Omar
Abdelrahman Rabie
Alyaa Hassan
Menna Hassan

Points:

- Data Cleaning
- Data visualizations
- Methods and roles

Data Cleaning

FR

```
import pandas as pd
df = pd.read_csv('/content/covid19 (1).csv')
df.head()
```

	Unnamed: 0	Country,Other	TotalCases	NewCases	TotalDeaths
0	0	USA	29383324.0	12619.0	529515
1	1	India	11156250.0	16927.0	157471
2	2	Brazil	10647845.0	NaN	257562
3	3	Russia	4278750.0	10535.0	87348
4	4	UK	4188400.0	NaN	123296

```
df.tail(10)
```

	Unnamed: 0	Country,Other	TotalCases	NewCases	TotalDeaths
219	219	Vanuatu	1.0	NaN	
220	220	China	89933.0	10.0	4636
221	221	33720845	23073.0	765344.0	1378
222	222	25223160	72549.0	400595.0	742
223	223	18134879	6577.0	470591.0	128
224	224	34442286	119818.0	822446.0	2724
225	225	3949024	5653.0	104626.0	64
226	226	51331	16.0	1091.0	NaN
227	227	721	NaN	15.0	NaN
228	228	115522246	227686.0	2564708.0	5036

FR

```
[ ]  #removing incorrect data (data with no countries)
     df2=df.drop([221,222,223,224,225,226,227,228],axis=0)

     #replacing null values with 0
     df2=df2.fillna(0)
```

```
[ ]  #set 0 in all TotalDeaths that cannot be replaced with fillna
     for i in range (220):
       if df2['TotalDeaths'][i]==' ':
         df2['TotalDeaths'][i]=0
     df2
```

FR

```
#transforming TotalDeath into float
df2['TotalDeaths']=df2['TotalDeaths'].astype('f')
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 221 entries, 0 to 220
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      221 non-null    int64
 1   Country,Other   221 non-null    object
 2   TotalCases      221 non-null    float64
 3   NewCases        221 non-null    float64
 4   TotalDeaths     221 non-null    float32
dtypes: float32(1), float64(2), int64(1), object(1)
memory usage: 17.6+ KB
```

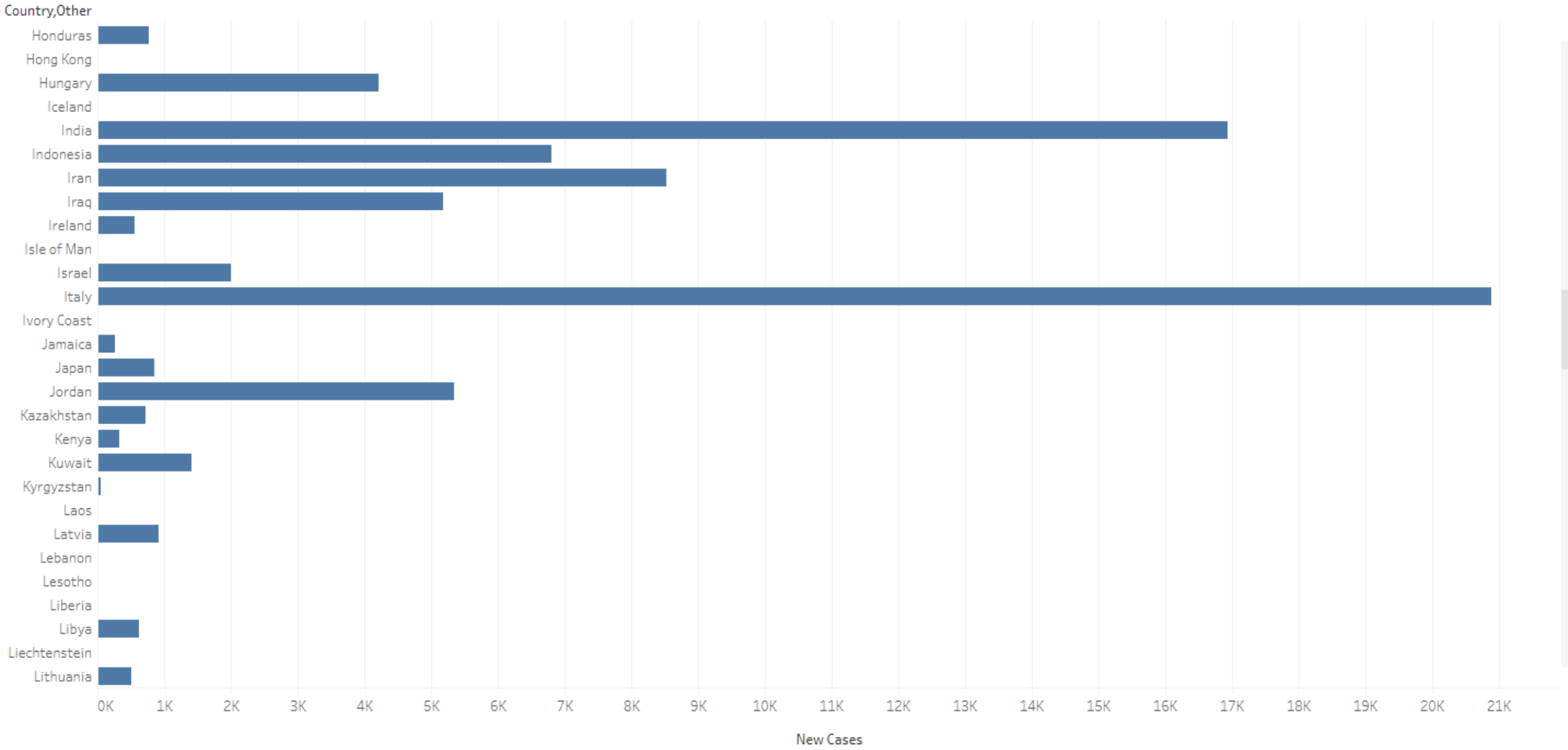
```
#saving the new spreadsheet
df2.to_csv('/content/covid19_new.csv')
```

Data visualization

Columns SUM(New Cases)

Rows Country,Other

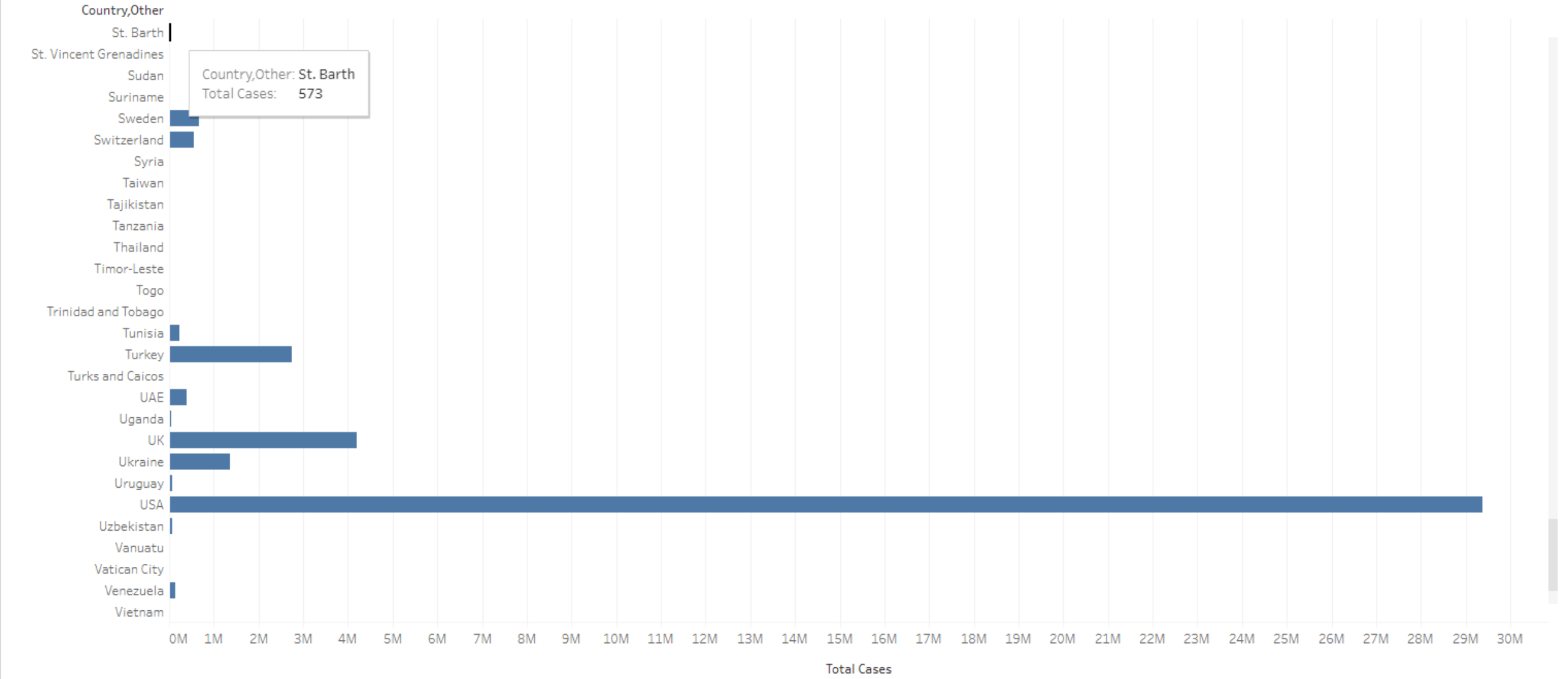
Relation between Countries and No.new cases



Columns SUM(Total Cases)

Rows Country,Other

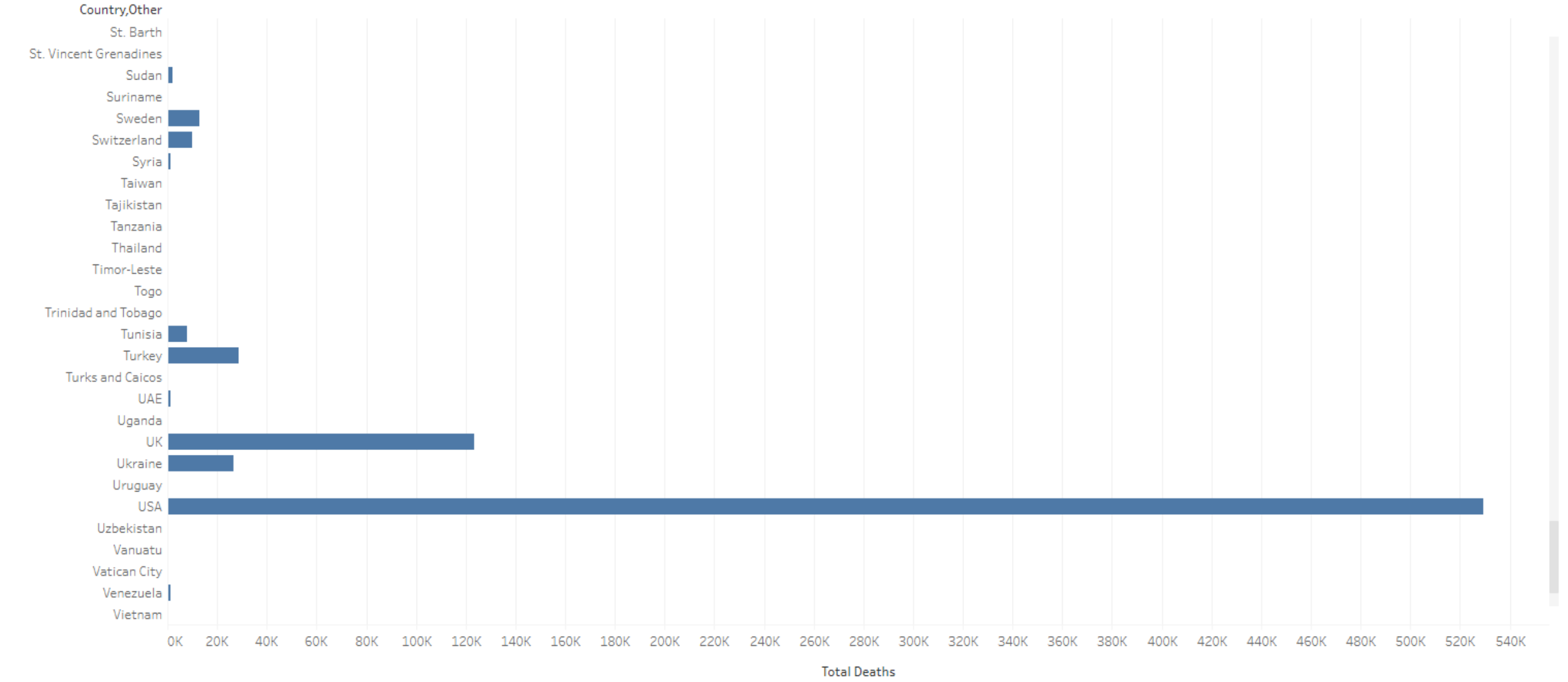
Relation between Countries and No.total cases



Columns SUM(Total Deaths)

Rows Country,Other

Relation between Countries and No.deaths



FR

Heat map

to locate where most cases lie

File Data Worksheet Dashboard Analysis Map Format Help

New Workbook (Tableau Public)

Publish As... Mohamed Moh...

Publish to save your changes

Show Me

Data Analytics

covid19_new

Search

Tables

- Country.Other
- *Measure Names*
- New Cases
- Total Cases
- Total Deaths
- *covid19_new.csv (Count)*
- *Latitude (generated)*
- *Longitude (generated)*
- *Measure Values*

Pages

Filters

Marks

Automatic

Color Size Label

Detail Tooltip

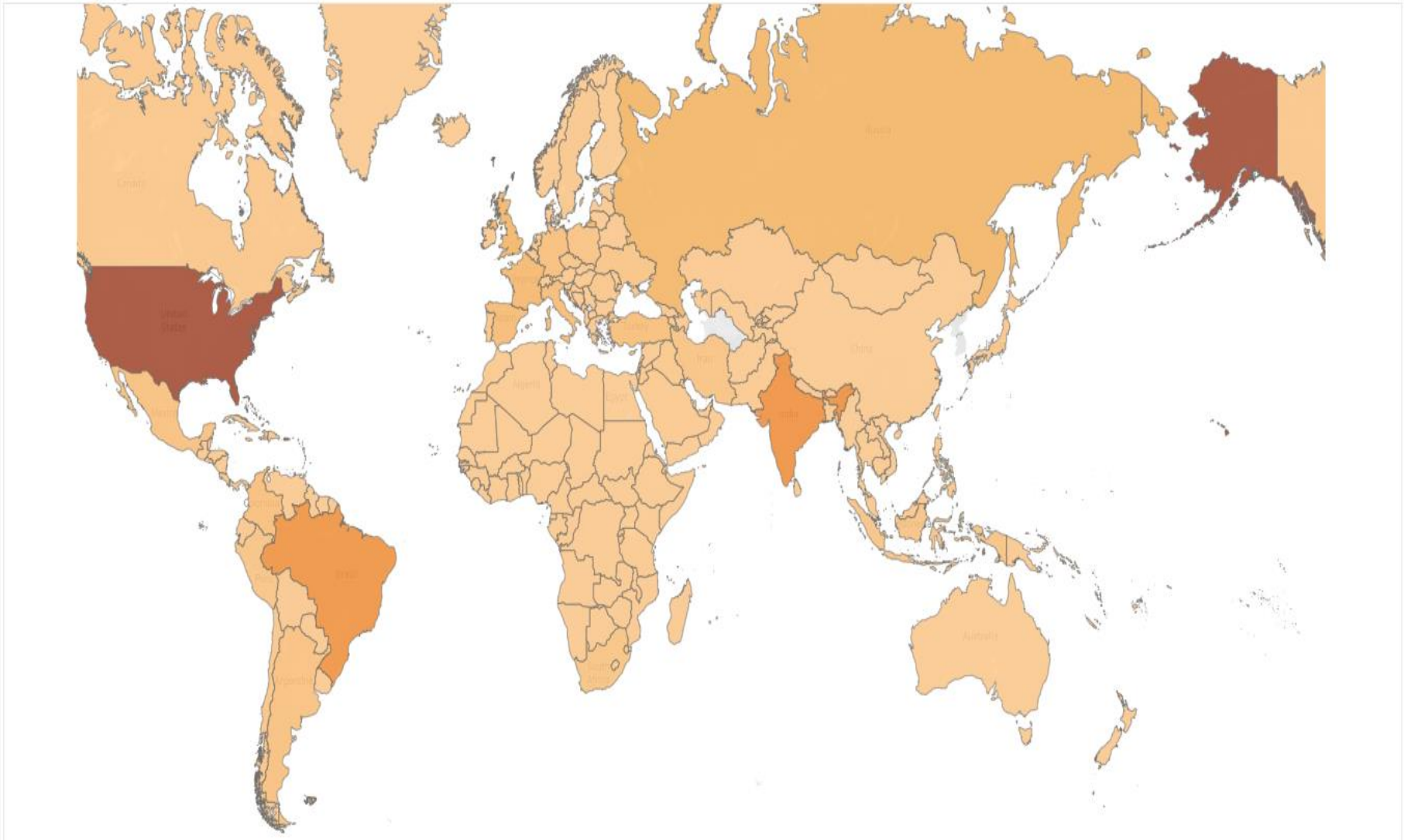
SUM(Total Cases)

Country.Other

Columns Longitude (generated)

Rows Latitude (generated)

Sheet 2



Total Cases

1 29,383,324

Publish to save your changes

Show Me

Pages

Filters

Marks

Automatic

Color | Size | Label | Detail | Tooltip

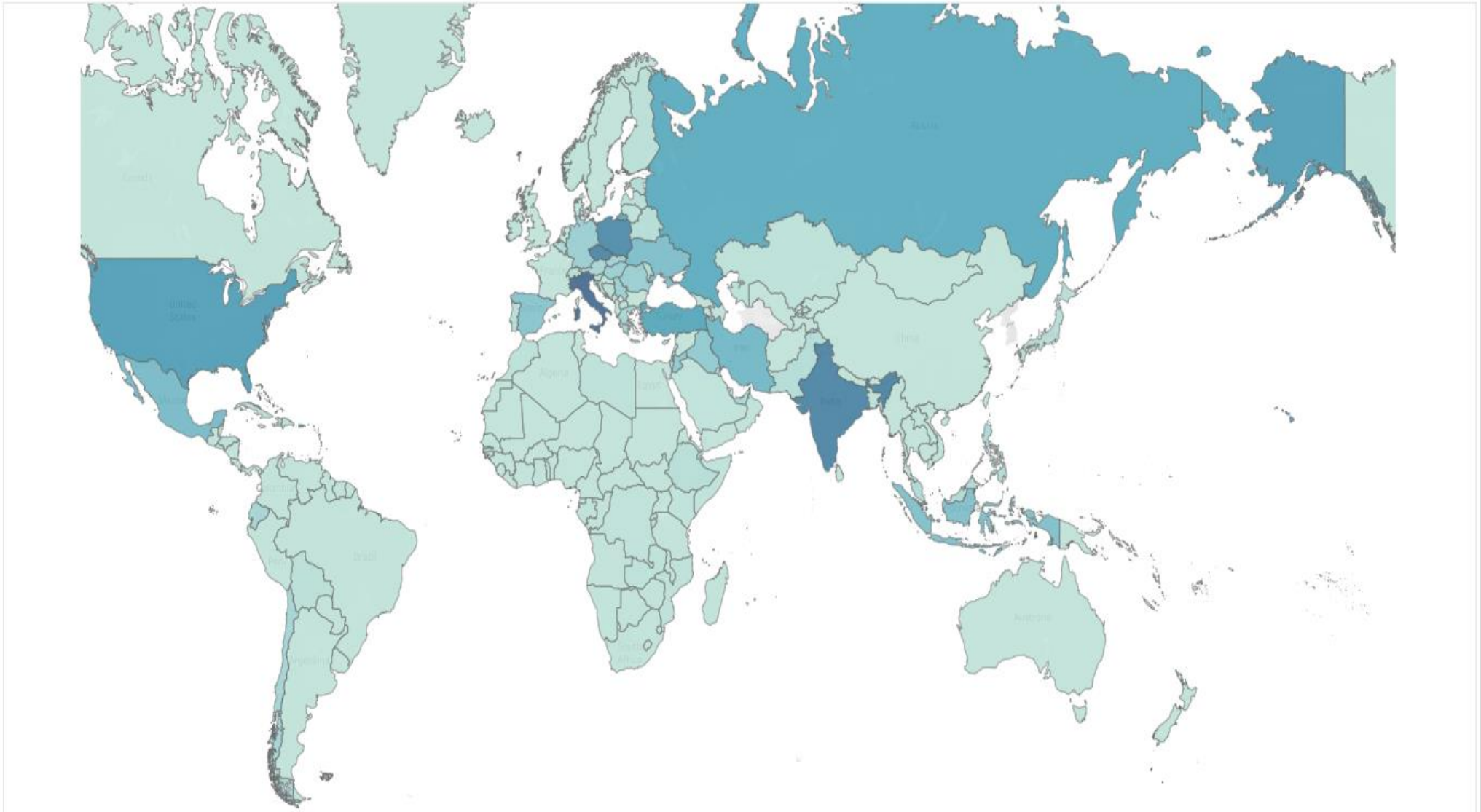
SUM(New Cases)

Country,Other

Columns: Longitude (generated)

Rows: Latitude (generated)

Sheet 2



New Cases

0 20,884

Pages

Filters

Marks

Automatic

Color | Size | Label | Detail | Tooltip

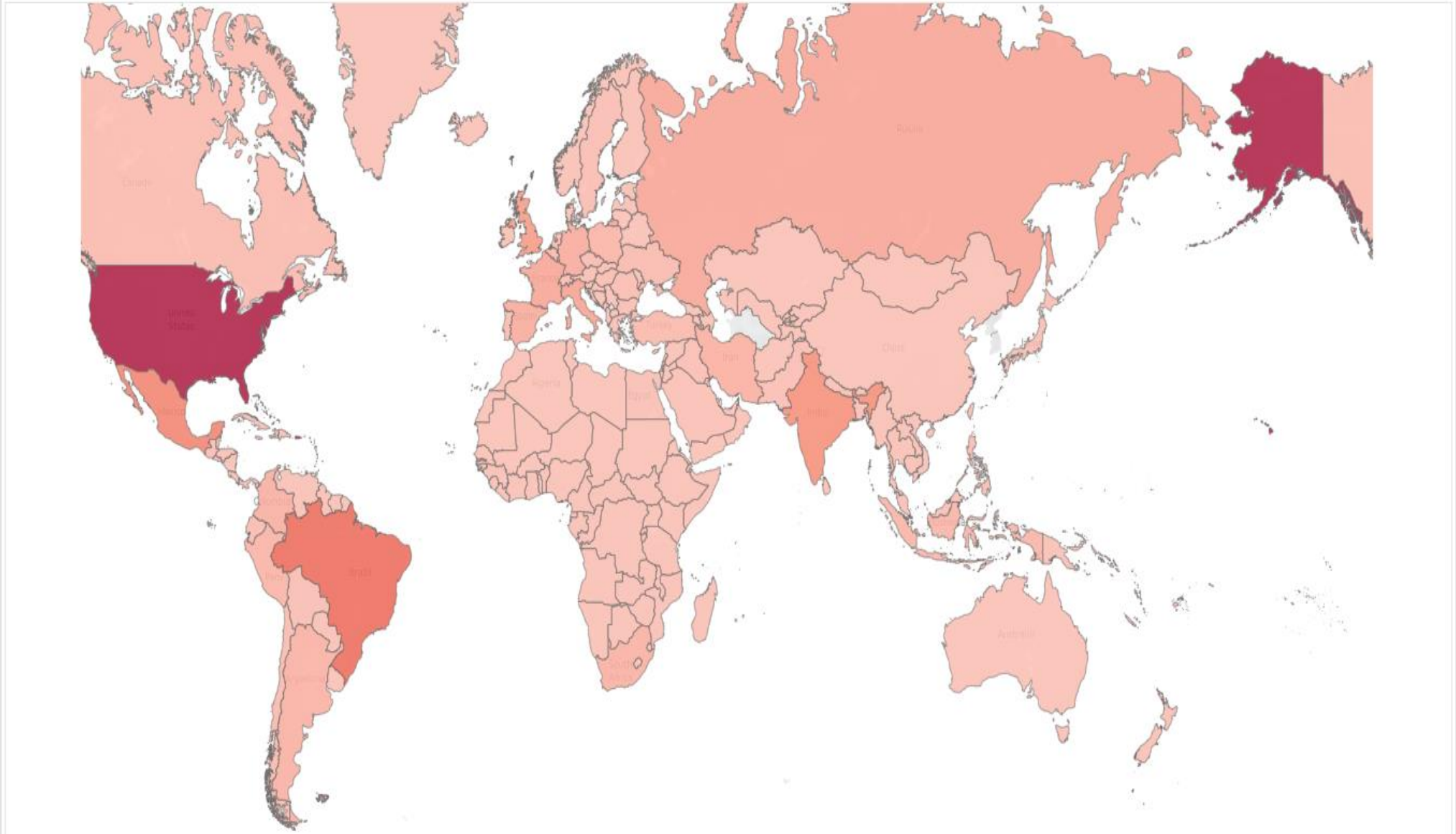
SUM(Total Deaths)

Country,Other

Columns: Longitude (generated)

Rows: Latitude (generated)

Sheet 2



Total Deaths

0 529,515

FR

Finally , a scatter plotting for the cases with respect to no. deaths

(1) WhatsApp | Edit | project

public.tableau.com/authoring/project_16631759341710/Sheet3#1

File Data Worksheet Dashboard Analysis Map Format Help

project (Tableau Public)

Publish | Mohamed Moh...

Show Me

Data | Analytics

covid19_new

Search

Tables

- Country,Other
- *Measure Names*
- New Cases
- Total Cases
- Total Deaths
- *covid19_new.csv (Count)*
- *Latitude (generated)*
- *Longitude (generated)*
- *Measure Values*

Pages

Filters

Marks

Circle

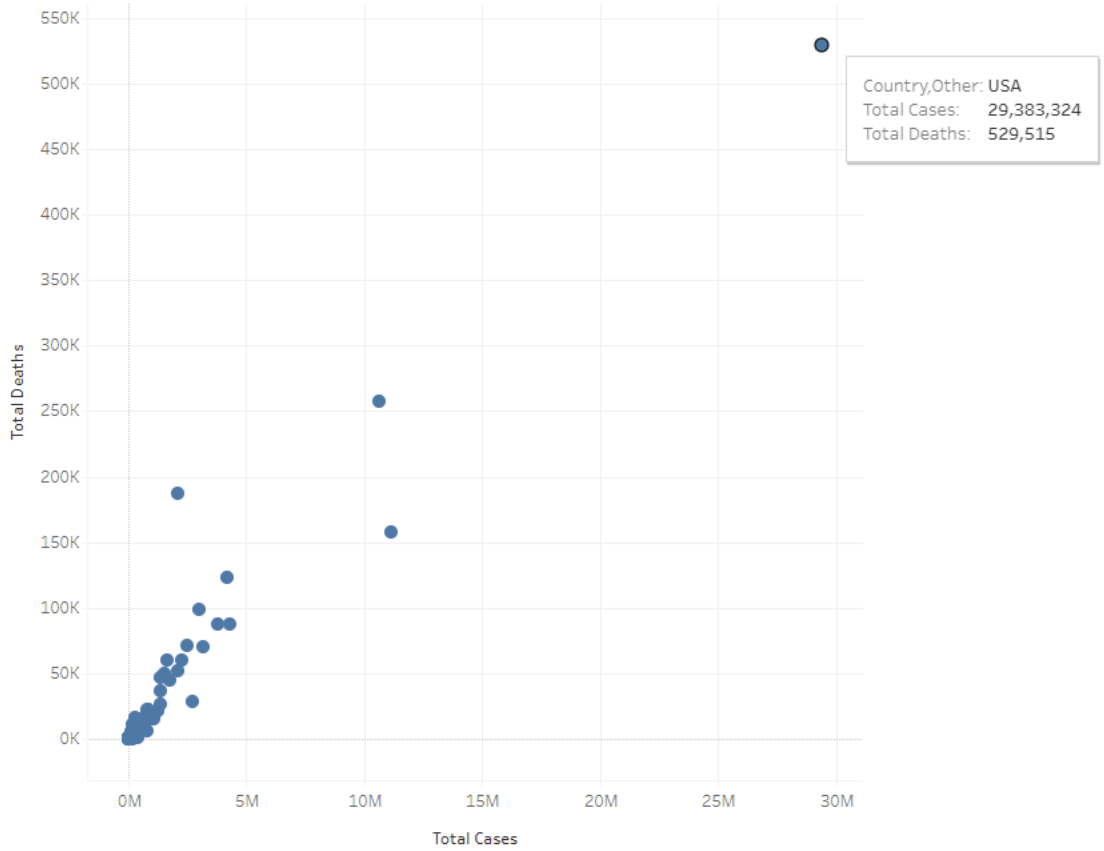
Color | Size | Label | Detail | Tooltip

Country,Other

Columns: SUM(Total Cases)

Rows: SUM(Total Deaths)

Sheet 3



Data Source | Sheet 1 | Sheet 2 | Sheet 3

221 marks | 1 row by 1 column | SUM(Total Cases): 115,522,246

covid19_new.csv | covid19 (1).csv | xah6tBWNQxeoer....xlsx | ifKytSWMRtaysrUlj....xlsx | RNcnGH5XS1yXJx....xlsx | 7b-USYCkQvy_lEm....xlsx | Show all

34°C صافي غالبًا | ENG | 8:11 PM | 9/14/2022

FR

Methods and Roles

METHODS:

1- SPREADSHEETS (EXCEL)
2- GOOGLE COLAB (TO RUN PYTHON CODES TO CLEAN THE DATA)
3- TABLEAU TOOL (FOR VISUALIZATIONS)

ROLES:

CLEANING PROCESS:

1- ABDELRAHMAN RABIE
2- ABDELRAHMAN OMAR

VIZUALIZATIONS USING TABLEAU
1-MOHAMED MOHSEN
2- MENNA HASSAN

PRESENTATION AND RESEARCH (FOR SUITABLE TOOLS AND FUNCTIONS)

ALYAA HASSAN