

Project Walkthrough

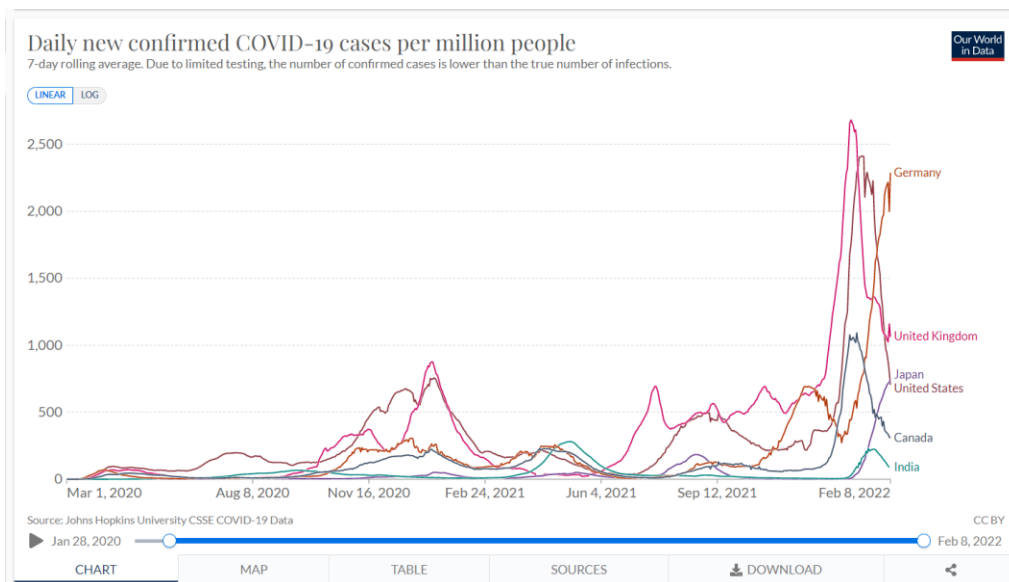
Over the past couple of years, we've been living in this pandemic that has affected the entire world. In this project we'll gather the data and dive into the statistics to find out the rate of infection for each country and number of deaths.

Tools used: Excel, SQL, & Tableau

[Full SQL Script](#) & [Tableau Dashboard](#)

Gathering and cleaning the data

- 1) Data is gathered from <https://ourworldindata.org/coronavirus>

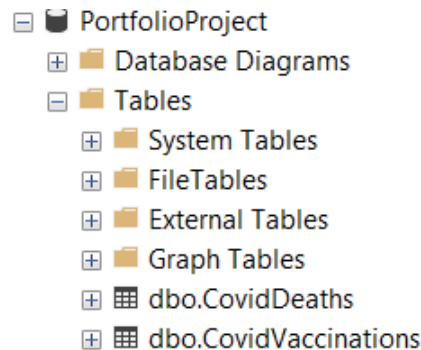


- 2) The downloaded datasets contain the total number of deaths and vaccinations. They are then cleaned by removing columns that are not relevant to the project.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
iso_code	continent	location	date	population	total_case	new_case	new_case	total_deat	new_deat	new_deat	total_case	new_case	new_case
AFG	Asia	Afghanistan	2/24/2020	39835428	5	5					0.126	0.126	
AFG	Asia	Afghanistan	2/25/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/26/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/27/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/28/2020	39835428	5	0					0.126	0	
AFG	Asia	Afghanistan	2/29/2020	39835428	5	0	0.714				0.126	0	0.018
AFG	Asia	Afghanistan	3/1/2020	39835428	5	0	0.714				0.126	0	0.018
AFG	Asia	Afghanistan	3/2/2020	39835428	5	0	0				0.126	0	0
AFG	Asia	Afghanistan	3/3/2020	39835428	5	0	0				0.126	0	0
AFG	Asia	Afghanistan	3/4/2020	39835428	5	0	0				0.126	0	0

A	B	C	D	E	F	G	H	I	J	K
iso_code	continent	location	date	new_tests	total_tests	total_tests	new_tests	new_tests	new_tests	positive_r
AFG	Asia	Afghanistan	2/24/2020							
AFG	Asia	Afghanistan	2/25/2020							
AFG	Asia	Afghanistan	2/26/2020							
AFG	Asia	Afghanistan	2/27/2020							
AFG	Asia	Afghanistan	2/28/2020							
AFG	Asia	Afghanistan	2/29/2020							
AFG	Asia	Afghanistan	3/1/2020							
AFG	Asia	Afghanistan	3/2/2020							
AFG	Asia	Afghanistan	3/3/2020							
AFG	Asia	Afghanistan	3/4/2020							

- 3) Data is then imported into Microsoft SQL server where it will be queried. Two tables are created in the database.



SQL Queries

- 1) SQL scripts are used to get different looks of the data. These scripts contain calculations to determine the percentage of the population infected and the highest death count per country.

```

Full SQL Script.sql...SUS-G11\mrodr (62)*
19 -- Shows percentage of population infected
20 Select Location, date, total_cases, population, (total_cases/population)*100 as InfectionRate
21 From PortfolioProject..CovidDeaths
22 --Where location like '%states%'
23 order by 1,2
24
25 -- Looking at Countries with Highest Infection Rate Compared to Population
26 Select Location, population, MAX(total_cases) as HighestInfectionCount, MAX((total_cases/population))*100 as InfectionRate
27 From PortfolioProject..CovidDeaths
28 group by Location, population
29 order by population desc
30
31 -- Countries With Highest Death Count per population
32 Select Location, MAX(cast(Total_deaths as int)) as TotalDeaths
33 From PortfolioProject..CovidDeaths
34 Where continent is not null
35 group by Location
36 order by TotalDeaths desc
37
100 %
Results Messages
Location TotalDeaths
1 United States 869026
2 Brazil 623636
3 India 490462
4 Russia 320178
5 Mexico 303301
6 Peru 204404
7 United Kingdom 154042
8 Indonesia 144227
9 Italy 143875
10 Colombia 132477
11 Iran 132251
12 France 129088
13 Argentina 119444
14 Germany 116967
15 Ukraine 105871
16 Poland 103846
17 South Africa 94625
Query executed successfully.
ASUS-G11\SQLEXPRESS (15.0 RTM)

```

- 2) The death rate for the world is calculated by creating a query that takes the total number of cases and the total number of deaths divided by the total number of cases then multiplied by 100.

```
Full SQL Script.sql...SUS-G11\mrodr (62)* X
46 -- Global Numbers
47
48 Select date, SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths,
49 SUM(cast(new_deaths as int))/SUM(new_cases)*100 as DeathPercentage
50 From PortfolioProject..CovidDeaths
51 --Where location like '%states%'
52 Where continent is not null
53 Group By date
54 order by 1,2
55
56
57 -- Death Rate For World
58 Select SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths,
59 SUM(cast(new_deaths as int))/SUM(new_cases)*100 as DeathPercentage
60 From PortfolioProject..CovidDeaths
61 --Where location like '%states%'
62 Where continent is not null
63 --Group By date
64 order by 1,2
65
```

100 %

Results Messages

	total_cases	total_deaths	DeathPercentage
1	354385307	5579947	1.57454242311462

- 3) A view is created to store data for the visualizations that will be created in Tableau

```
121
122 --Creating View to store data for visualizations
123 Create View DeathRateForWorldView as
124 Select SUM(new_cases) as total_cases, SUM(cast(new_deaths as int)) as total_deaths,
125 SUM(cast(new_deaths as int))/SUM(new_cases)*100 as DeathPercentage
126 From PortfolioProject..CovidDeaths
127 --Where location like '%states%'
128 Where continent is not null
129 Group By date
130 --order by 1,2
131
132 Select *
133 from DeathRateForWorld
```

Visualizations

- 1) The data is imported into Tableau. We will be looking at the global numbers and the rate of infection delineated by country.

Sheet1

2 fields 6 rows

Name

Sheet1

Fields

Type	Field Name	Physical Table	Remote Fi...
Abc	Continent	Sheet1	location
#	Total Death Count	Sheet1	TotalDeathC...

Abc	#
Sheet1	Sheet1
Continent	Total Death Count
Europe	1,598,454
Asia	1,278,334
North America	1,268,796
South America	1,192,927
Africa	235,852
Oceania	5,584

- 2) The first sheet contains a simple table describing the global numbers for the total number of cases, total, deaths, and the death percentage.

Pages

Filters

Measure Names

Marks

Automatic

Color

Size

Text

Detail

Tooltip

Measure Values

Measure Values

SUM(Total Cases)

SUM(Total Deaths)

SUM(Death Percent..)

Columns

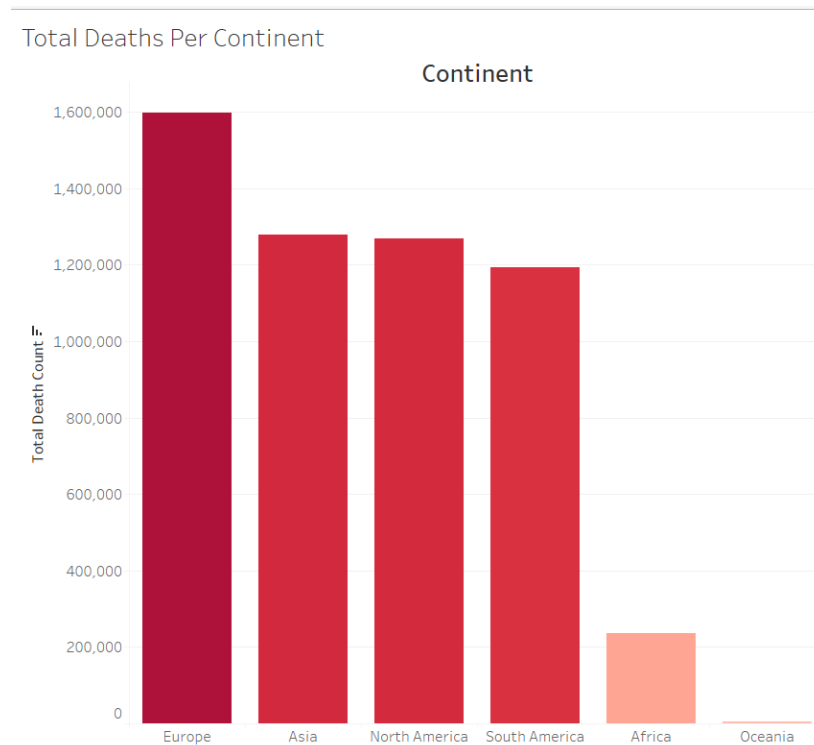
Measure Names

Rows

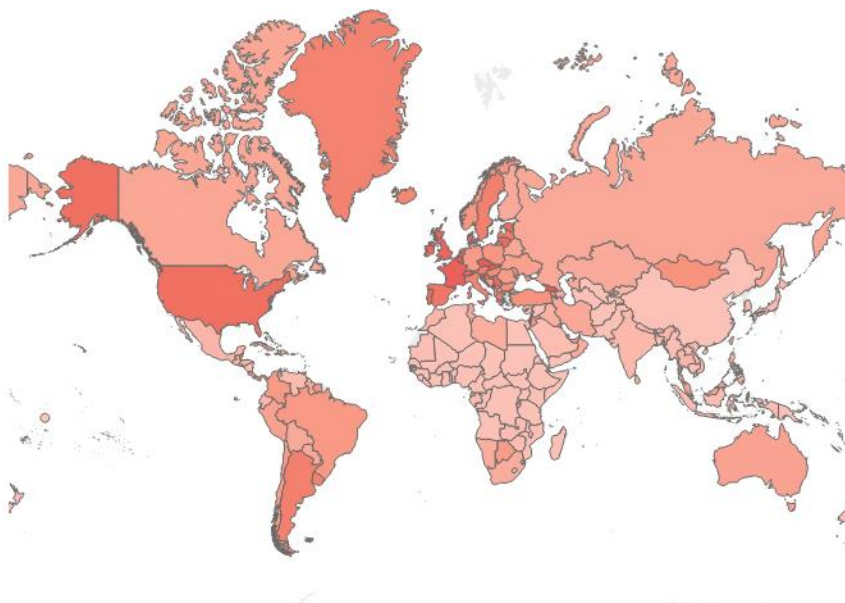
Global Numbers

Total Cases	Total Deaths	Death Percentage
354,385,307	5,579,947	1.57

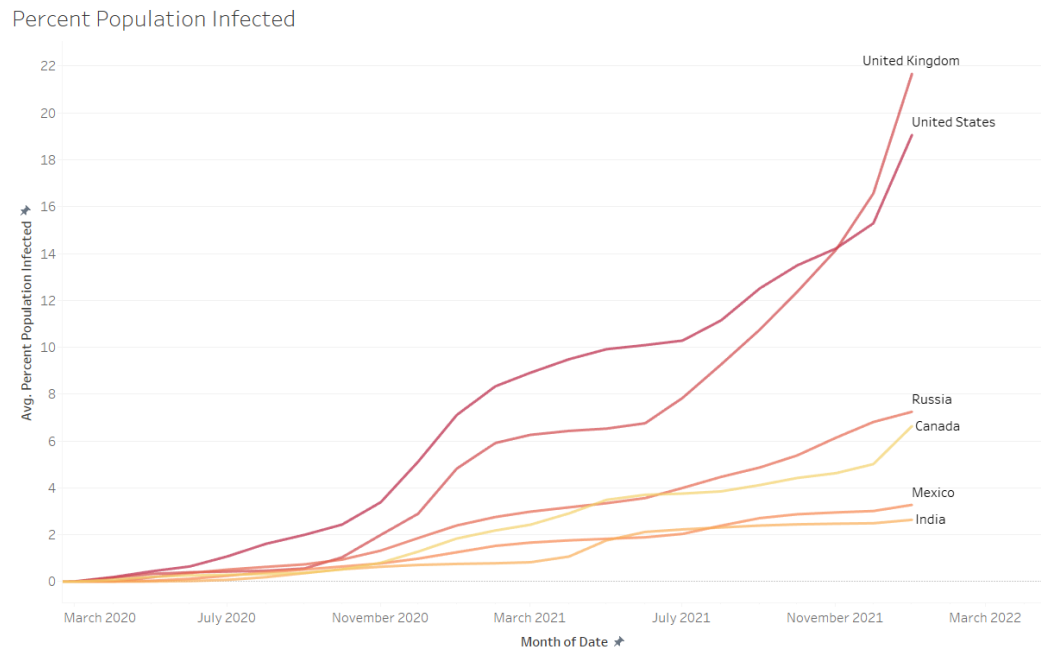
- 3) The 2nd sheet is a bar chart that is easy to digest. It shows the number of deaths per continent with Europe being the highest at 1,598,454 deaths.



- 4) The third sheet is an interactive map of the world where you can hover over each country and see the rate of infection. The darker countries have the highest infection rate for its population.



- 5) The last sheet contains the average percent of the population infected per month since the beginning of the pandemic. This is also a comparison between 6 countries. These countries are Canada, India, Mexico, Russia, the United Kingdom, and the United States. We can see that the US and the UK have the highest numbers.



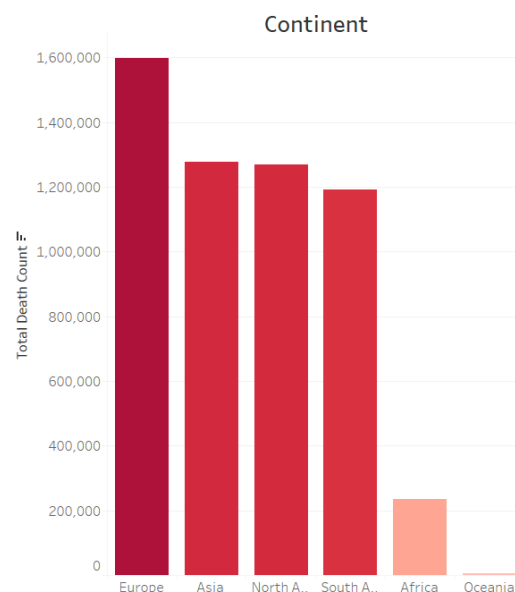
Dashboard

The final piece of the project is an interactive dashboard containing all the previous sheets. You can see the full visualization [Here](#)

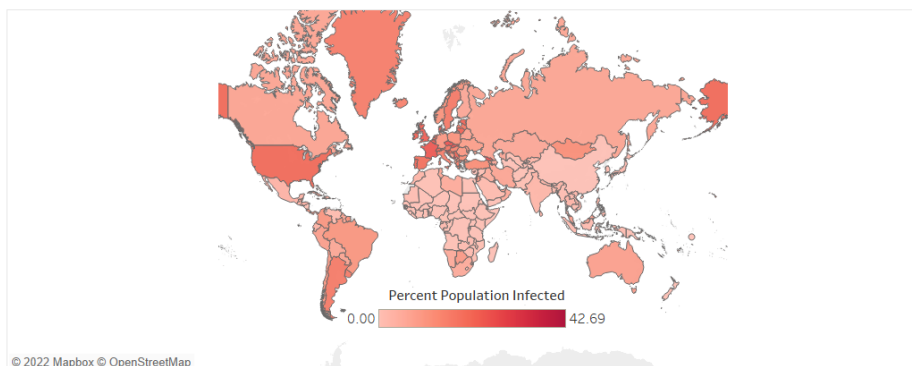
Global Numbers

Total Cases	Total Deaths	Death Percentage
354,385,307	5,579,947	1.57

Total Deaths Per Continent



Percent Population Infected Per Country



Percent Population Infected

