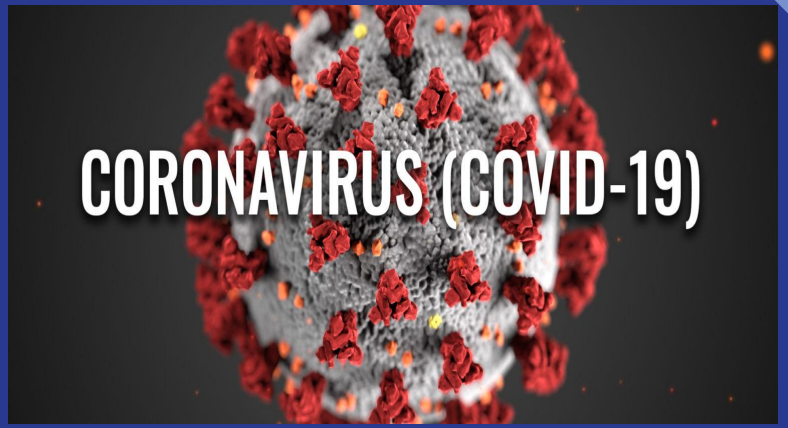
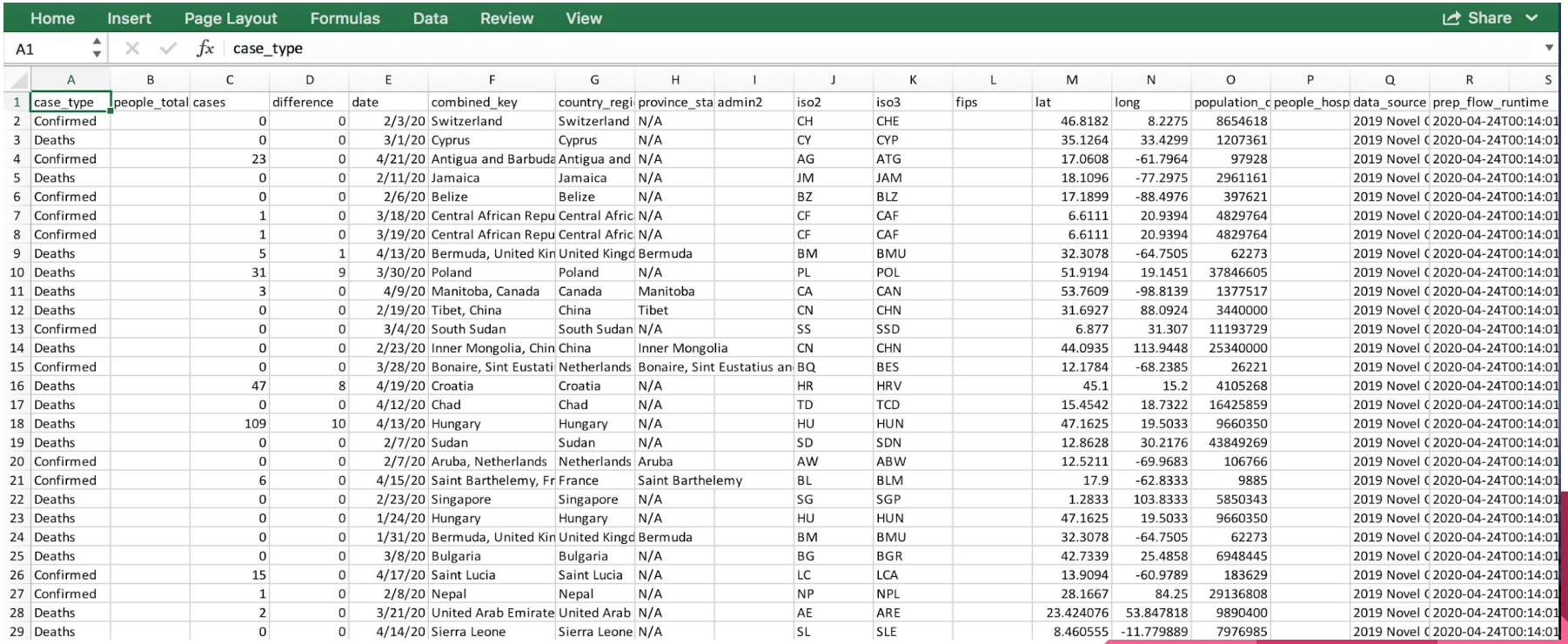
Data Processing and Analytics COVID-19 Analysis



By: Hector Pedro, Israel Alegria, Julio Montiel

Dataset

https://www.dropbox.com/s/smw5aein7i8zr1d/covid-19-data-resource-hub-covid-19-case-counts%20%281%29.zip?dl=0
https://github.com/juliom12/covid_19_cases



Home Insert Page Layout Formulas Data Review View Share

A1 fx case_type

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	case_type	people_total	cases	difference	date	combined_key	country_regi	province_sta	admin2	iso2	iso3	fips	lat	long	population_c	people_hosp	data_source	prep_flow_runtime	
2	Confirmed		0	0	2/3/20	Switzerland	Switzerland	N/A		CH	CHE		46.8182	8.2275	8654618		2019 Novel C	2020-04-24T00:14:01	
3	Deaths		0	0	3/1/20	Cyprus	Cyprus	N/A		CY	CYP		35.1264	33.4299	1207361		2019 Novel C	2020-04-24T00:14:01	
4	Confirmed		23	0	4/21/20	Antigua and Barbuda	Antigua and	N/A		AG	ATG		17.0608	-61.7964	97928		2019 Novel C	2020-04-24T00:14:01	
5	Deaths		0	0	2/11/20	Jamaica	Jamaica	N/A		JM	JAM		18.1096	-77.2975	2961161		2019 Novel C	2020-04-24T00:14:01	
6	Confirmed		0	0	2/6/20	Belize	Belize	N/A		BZ	BLZ		17.1899	-88.4976	397621		2019 Novel C	2020-04-24T00:14:01	
7	Confirmed		1	0	3/18/20	Central African Repu	Central Afric	N/A		CF	CAF		6.6111	20.9394	4829764		2019 Novel C	2020-04-24T00:14:01	
8	Confirmed		1	0	3/19/20	Central African Repu	Central Afric	N/A		CF	CAF		6.6111	20.9394	4829764		2019 Novel C	2020-04-24T00:14:01	
9	Deaths		5	1	4/13/20	Bermuda, United Kin	United Kingd	Bermuda		BM	BMU		32.3078	-64.7505	62273		2019 Novel C	2020-04-24T00:14:01	
10	Deaths		31	9	3/30/20	Poland	Poland	N/A		PL	POL		51.9194	19.1451	37846605		2019 Novel C	2020-04-24T00:14:01	
11	Deaths		3	0	4/9/20	Manitoba, Canada	Canada	Manitoba		CA	CAN		53.7609	-98.8139	1377517		2019 Novel C	2020-04-24T00:14:01	
12	Deaths		0	0	2/19/20	Tibet, China	China	Tibet		CN	CHN		31.6927	88.0924	3440000		2019 Novel C	2020-04-24T00:14:01	
13	Confirmed		0	0	3/4/20	South Sudan	South Sudan	N/A		SS	SSD		6.877	31.307	11193729		2019 Novel C	2020-04-24T00:14:01	
14	Deaths		0	0	2/23/20	Inner Mongolia, Chin	China	Inner Mongolia		CN	CHN		44.0935	113.9448	25340000		2019 Novel C	2020-04-24T00:14:01	
15	Confirmed		0	0	3/28/20	Bonaire, Sint Eustati	Netherlands	Bonaire, Sint Eustatius an		BQ	BES		12.1784	-68.2385	26221		2019 Novel C	2020-04-24T00:14:01	
16	Deaths		47	8	4/19/20	Croatia	Croatia	N/A		HR	HRV		45.1	15.2	4105268		2019 Novel C	2020-04-24T00:14:01	
17	Deaths		0	0	4/12/20	Chad	Chad	N/A		TD	TCD		15.4542	18.7322	16425859		2019 Novel C	2020-04-24T00:14:01	
18	Deaths		109	10	4/13/20	Hungary	Hungary	N/A		HU	HUN		47.1625	19.5033	9660350		2019 Novel C	2020-04-24T00:14:01	
19	Deaths		0	0	2/7/20	Sudan	Sudan	N/A		SD	SDN		12.8628	30.2176	43849269		2019 Novel C	2020-04-24T00:14:01	
20	Confirmed		0	0	2/7/20	Aruba, Netherlands	Netherlands	Aruba		AW	ABW		12.5211	-69.9683	106766		2019 Novel C	2020-04-24T00:14:01	
21	Confirmed		6	0	4/15/20	Saint Barthelemy, Fr	France	Saint Barthelemy		BL	BLM		17.9	-62.8333	9885		2019 Novel C	2020-04-24T00:14:01	
22	Deaths		0	0	2/23/20	Singapore	Singapore	N/A		SG	SGP		1.2833	103.8333	5850343		2019 Novel C	2020-04-24T00:14:01	
23	Deaths		0	0	1/24/20	Hungary	Hungary	N/A		HU	HUN		47.1625	19.5033	9660350		2019 Novel C	2020-04-24T00:14:01	
24	Deaths		0	0	1/31/20	Bermuda, United Kin	United Kingd	Bermuda		BM	BMU		32.3078	-64.7505	62273		2019 Novel C	2020-04-24T00:14:01	
25	Deaths		0	0	3/8/20	Bulgaria	Bulgaria	N/A		BG	BGR		42.7339	25.4858	6948445		2019 Novel C	2020-04-24T00:14:01	
26	Confirmed		15	0	4/17/20	Saint Lucia	Saint Lucia	N/A		LC	LCA		13.9094	-60.9789	183629		2019 Novel C	2020-04-24T00:14:01	
27	Confirmed		1	0	2/8/20	Nepal	Nepal	N/A		NP	NPL		28.1667	84.25	29136808		2019 Novel C	2020-04-24T00:14:01	
28	Deaths		2	0	3/21/20	United Arab Emirate	United Arab	N/A		AE	ARE		23.424076	53.847818	9890400		2019 Novel C	2020-04-24T00:14:01	
29	Deaths		0	0	4/14/20	Sierra Leone	Sierra Leone	N/A		SL	SLE		8.460555	-11.779889	7976985		2019 Novel C	2020-04-24T00:14:01	

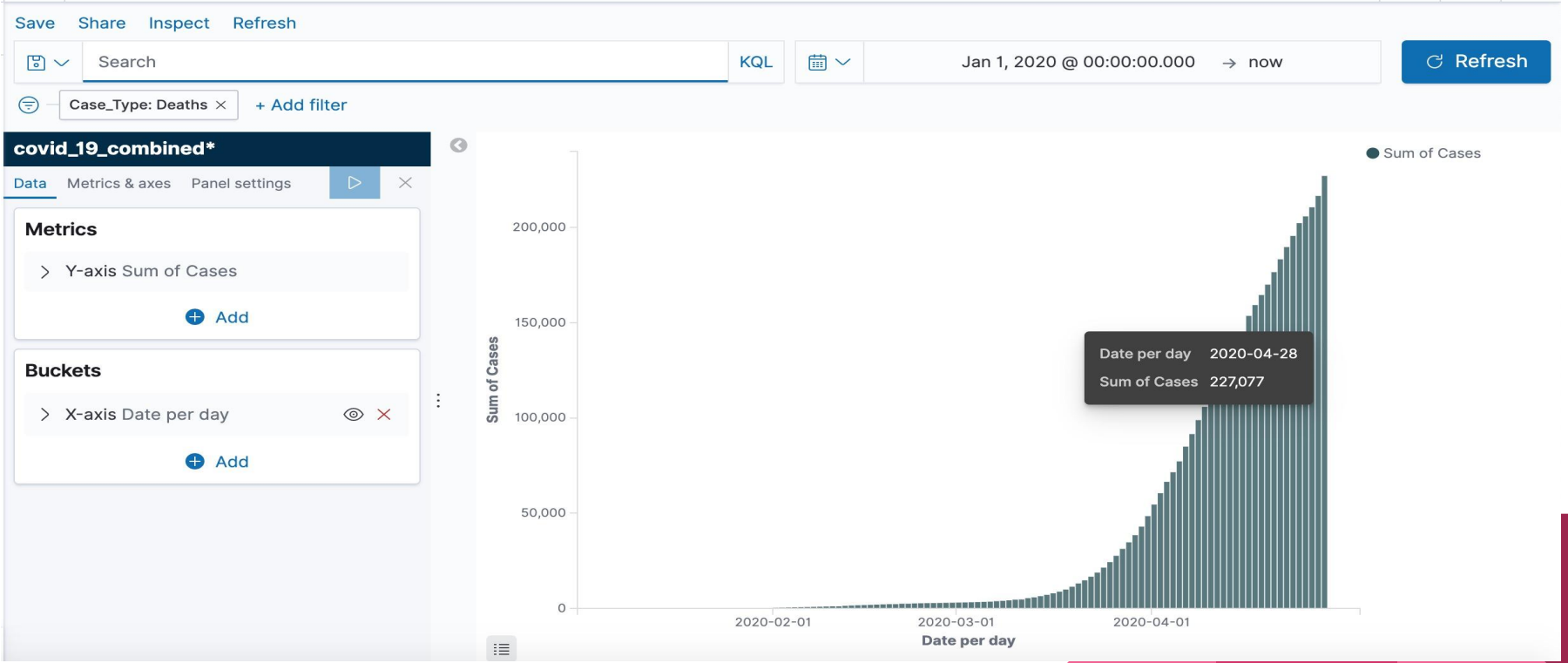
Project goals

The purpose of data processing and analyzing this Covid-19 dataset is to receive more insight on the current pandemic i.e.

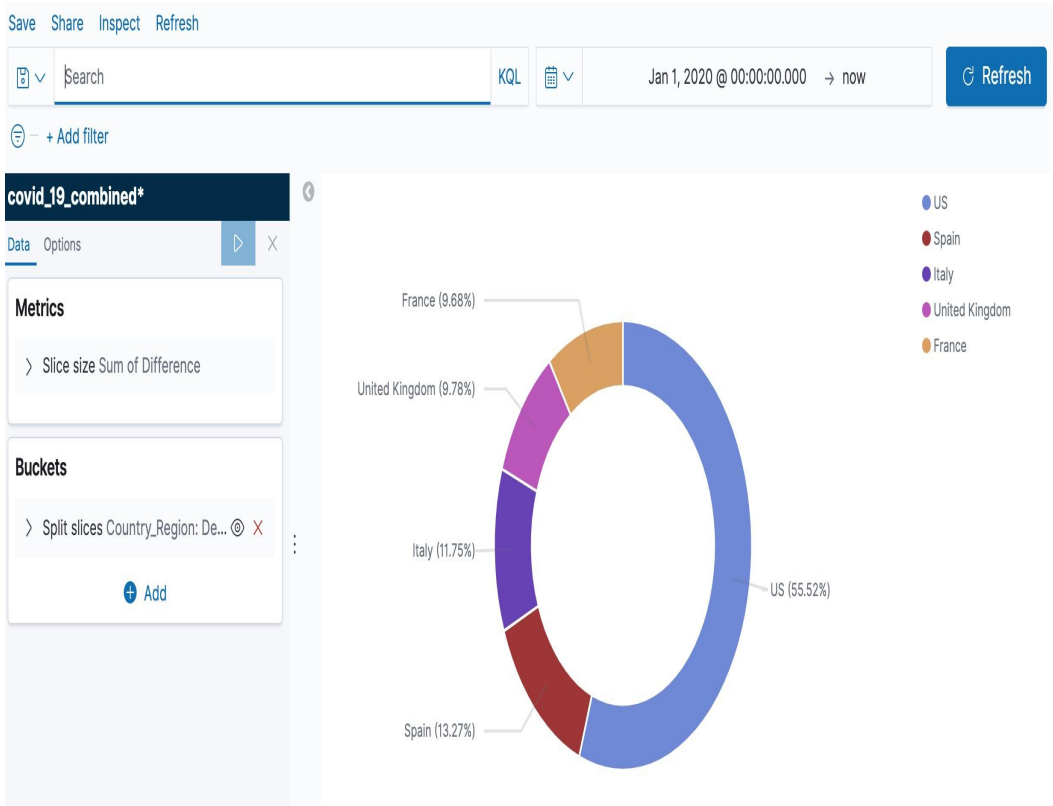
- The comparison of confirmed cases to deaths
- Countries with the highest death cases and confirmed cases
- Map of death cases around the world
- The count of cases per day
- Each countries percentage of cases

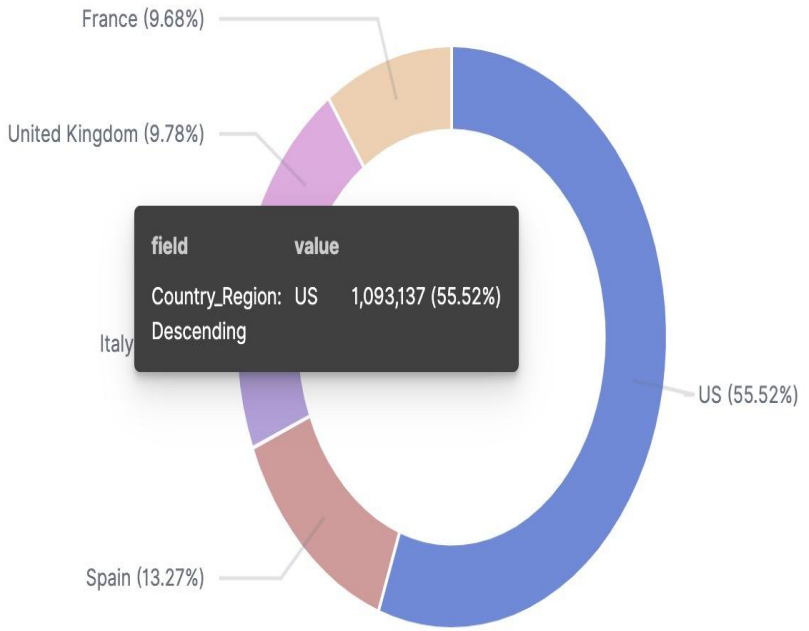
Elastic Visualization 1: Geospatial Map of Death Cases

Elastic Visualization 2: Sum of cases per day

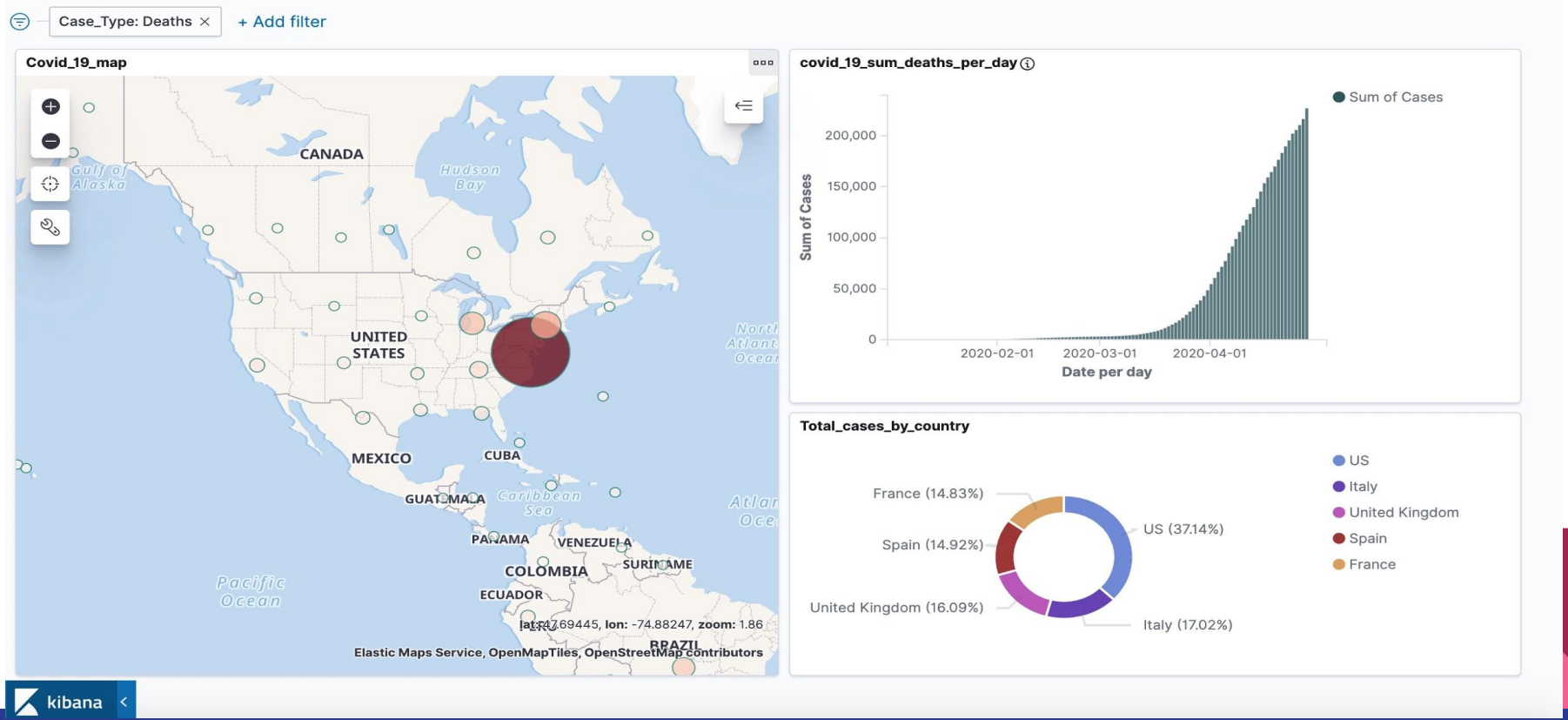


Elastic Visualization 3: Total Case % by country





Elastic Covid-19 Dashboard



Azure ML https://gallery.cortanaintelligence.com/Experiment/Covid-19-2

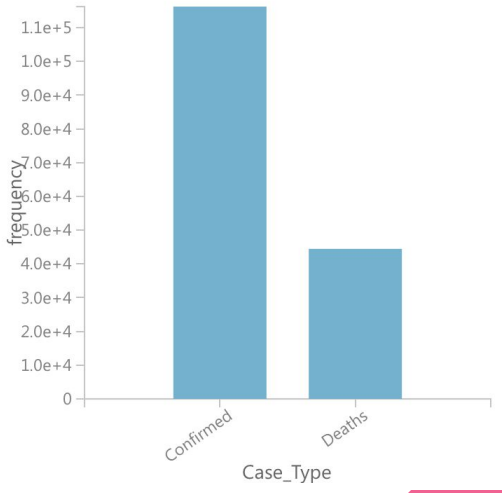
Azure ML

Covid-19 › Clean Missing Data › Cleaned dataset

rows 160693 columns 8

view as

Case_Type	Cases	Difference	Date	Country_Region
Confirmed	4815	344	2020-04-06T00:00:00Z	Chile
Confirmed	2	0	2020-04-14T00:00:00Z	US
Confirmed	145	2	2020-04-04T00:00:00Z	France
Confirmed	4	0	2020-04-02T00:00:00Z	US
Confirmed	22	2	2020-04-28T00:00:00Z	US
Confirmed	40	2	2020-04-15T00:00:00Z	US
Confirmed	562	51	2020-03-23T00:00:00Z	Saudi Arabia

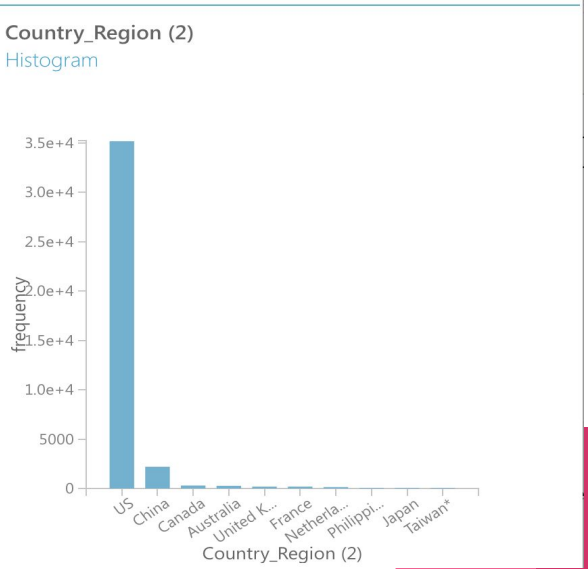


Azure ML

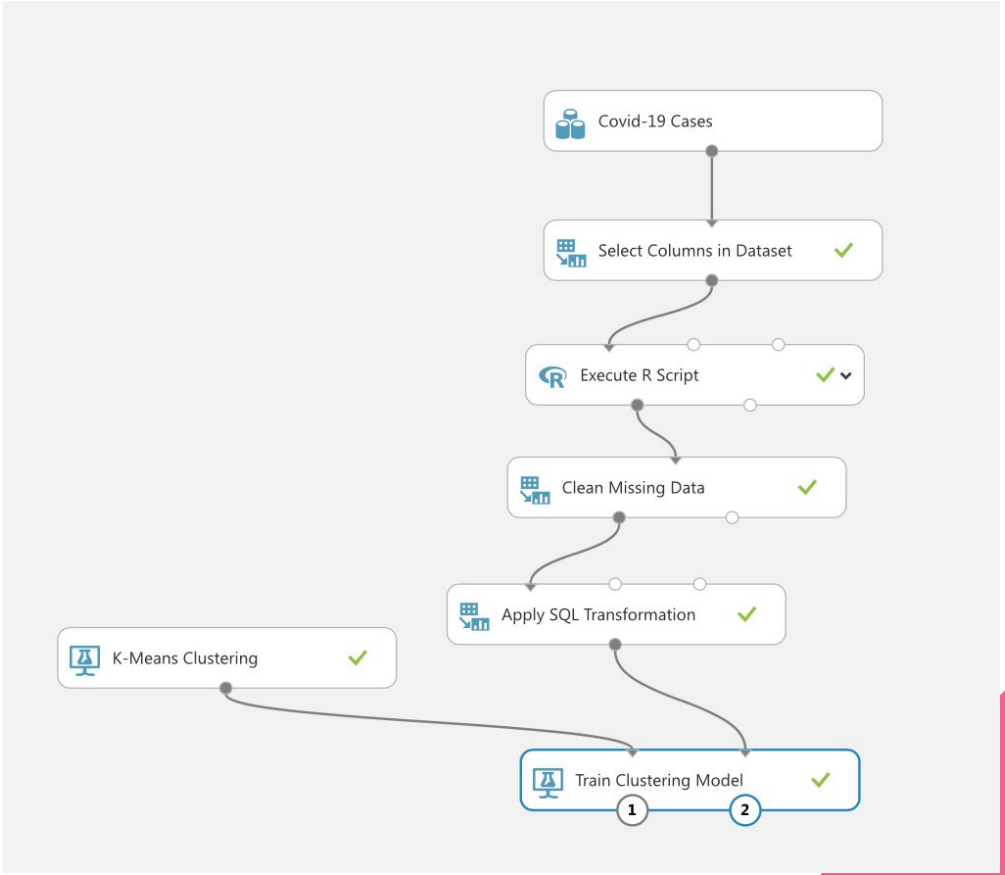
Covid-19 › Apply SQL Transformation › Results dataset

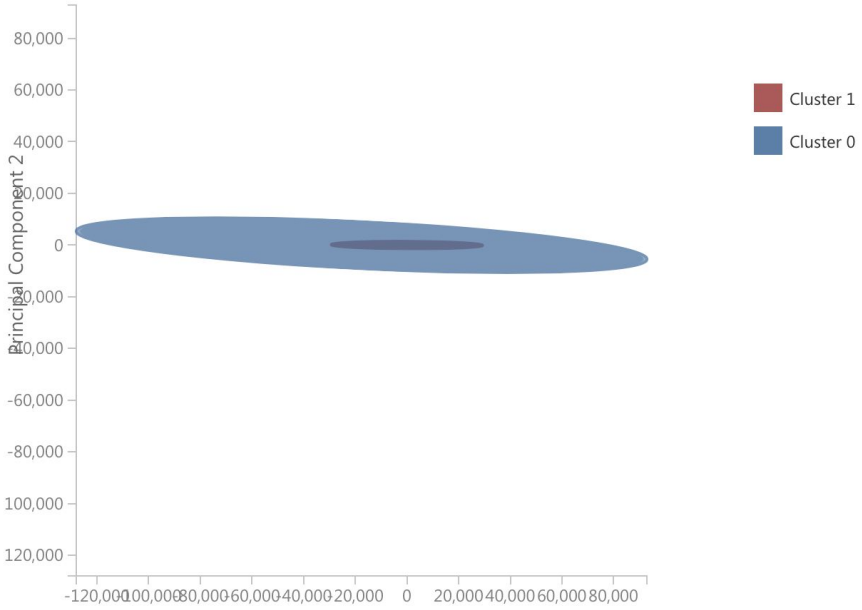
rows	columns
44463	11

ion	Province_State	Lat	Long	Country_Region (2)	Cases (2)	Case_Type (2)
	Guam	13.4443	144.7937	US	4	Deaths
	Oklahoma	36.119396	-95.94013939	US	5	Deaths
	Indiana	38.089227	-87.2716085	US	4	Deaths
	Montana	48.655776	-111.6942414	US	6	Deaths
	Texas	33.205219	-97.11691153	US	5	Deaths
	N/A	21.521757	-77.781167	Cuba	9	Deaths
	Louisiana	32.136752	-91.67589857	US	2	Deaths
	Texas	34.070449	-101.8268114	US	2	Deaths



Azure ML


Covid-19 Cases
Select Columns in Dataset
Execute R Script
Clean Missing Data
Apply SQL Transformation
K-Means Clustering
Train Clustering Model
1
2

Azure ML

Covid-19 › Train Clustering Model › Results dataset