Analyzing the Coronavirus Curve with SAS

Ching-Ching Wang 40038448

Alexander Guenov 40028815

Chavind Mungun 40062544

Yue Peng 40053397

Prof:
Dr. KRZYSZTOF DZIECIOLOWSKI
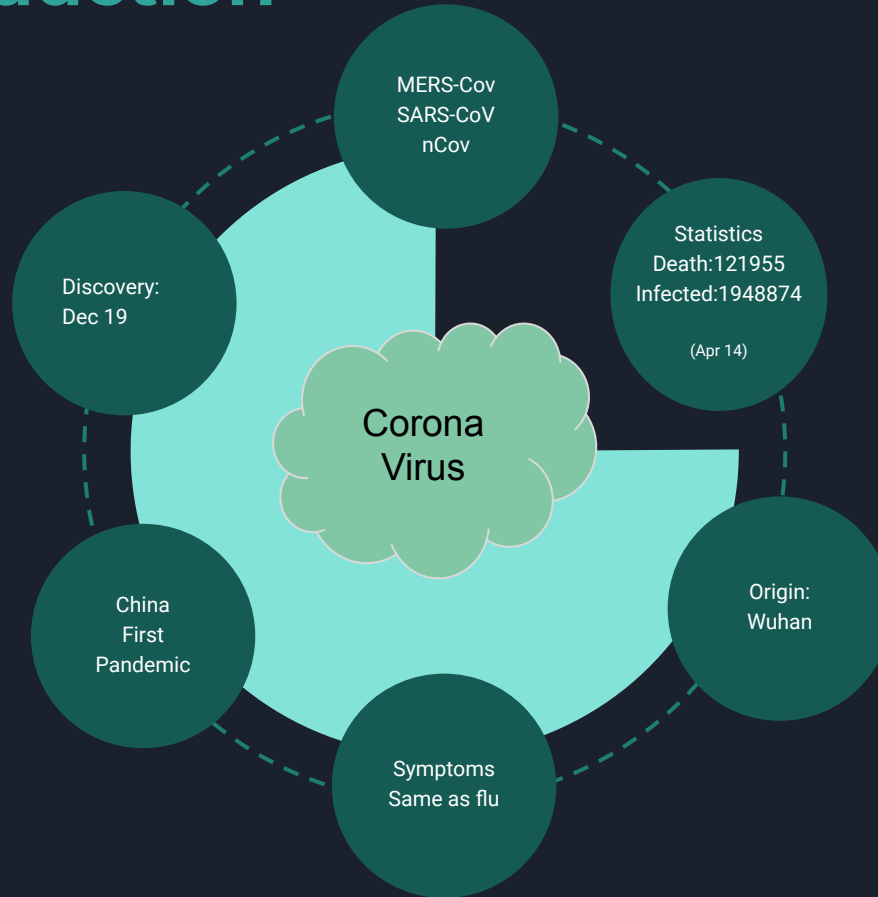
CONTENTS

- ❖ Introduction
- ❖ Raw Dataset Inspection
- ❖ Normalyzing Data
- ❖ Analytics
- ❖ Summary
- ❖ Disclaimer
- ❖ Bibliography

Introduction

MERS-Cov
SARS-CoV
nCov

Statistics
Death:121955
Infected:1948874

(Apr 14)

Discovery:
Dec 19

Corona
Virus

Origin:
Wuhan

China
First
Pandemic

Symptoms
Same as flu

Project Goals and Contents

★ Statistical analysis of the effects of the COVID-19

★ Present the data in such a way where countries' effective implementation of preventatives measures (such as social distancing, emergency health services, and so) on can be assessed.
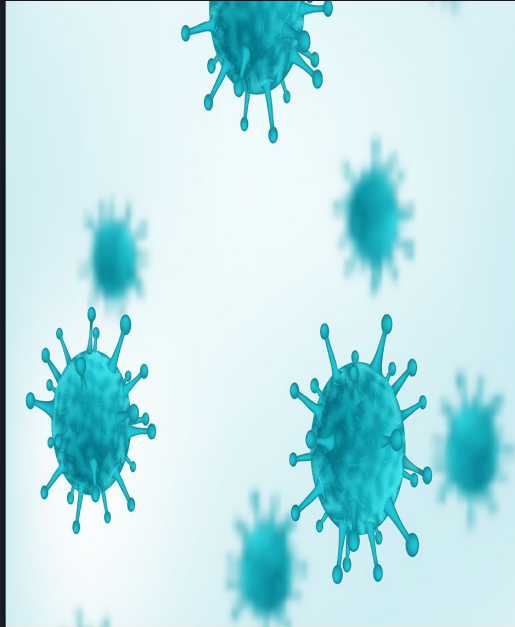
★ Compare the data of these countries

★ Normalize the data with basic SAS inputs and use PROC FREQ to compare countries.

★ We will attempt to use PROC GMAP, sgplot and proc means and use a SAS-supplied map data set of the world to make a world map of the key countries ranked by color for their response rating with an Http get request.

The Dataset

❏ Up to date
❏ 22 jan-bcurrentDate

Global Statements Used

1. Libname to specifystorage of permanent dataset;
2. dataset temp where data is temporarily stored
3. Tell SAS to allow "nonstandard" names
4. Graphics specifications

```
1.   libname libraryy "F:\school\BSTA445";
2.   filename probly temp;
3.   options validvarname=any;
4.   ods graphics on /
     width=30 in   height=30 in
     outputfmt=gif
     imagemap=on    imagename="Corona"
     border=off;
```

Loading the Dataset Into SAS

proc http makes a get request to get an uptodate version of the corona virus Dataset.

It afterwards stores in a temp file called probly.

```
/*Actual corona dataset*/
proc http
url="https://raw.githubusercontent.com/datasets/covid-19/master/data/time-series-19-covid-combined.csv"
method="GET"
 out=probly;
run;

/*Population */
proc import datafile="F:\school\BSTA445\population\data\population.csv"
out=mydata dbms=csv replace;
run;
```

Cleaning & Normalizing the Data

```
proc import  file=probly
  out=probly replace
  dbms=csv;
run;
```

Based on the above statement, we are using proc import to get the data and we are normalizing it into table-like structure.

	Date	Country/Regi...	Province/Sta...		Lat		Long		Confirmed		Recover...
1	2020-01-22	Canada	A		53.9333		-116.5765		0		
2	2020-01-23	Canada	A		53.9333		-116.5765		0		
3	2020-01-24	Canada	A		53.9333		-116.5765		0		
4	2020-01-25	Canada	A		53.9333		-116.5765		0		
5	2020-01-26	Canada	A		53.9333		-116.5765		0		
6	2020-01-27	Canada	A		53.9333		-116.5765		0		
7	2020-01-28	Canada	A		53.9333		-116.5765		0		

Focusing on Areas

We will be focusing primarily on a specific country, and province and its confirmed cases.

Hence we create a permanent dataset, and let's filter cases in canada Quebec for now. So we will be using the Data statement with some conditions.

```
Data libraryy.dataset;
Set probly
(where=
        (
                'Country/Region'n = 'Canada'
        and
                ('Province/State'n ='Q'
                )
        )
);
Run;
```

	Country/Regi..	Province/Sta..	Date	Lat	Long	Confirmed	Recovered	Deaths
1	Canada	Q	2020-01-22	52.9399	-73.5491	0	.	0
2	Canada	Q	2020-01-23	52.9399	-73.5491	0	.	0
3	Canada	Q	2020-01-24	52.9399	-73.5491	0	.	0
4	Canada	Q	2020-01-25	52.9399	-73.5491	0	.	0
5	Canada	Q	2020-01-26	52.9399	-73.5491	0	.	0
6	Canada	Q	2020-01-27	52.9399	-73.5491	0	.	0
7	Canada	Q	2020-01-28	52.9399	-73.5491	0	.	0
8	Canada	Q	2020-01-29	52.9399	-73.5491	0	.	0
9	Canada	Q	2020-01-30	52.9399	-73.5491	0	.	0
10	Canada	Q	2020-01-31	52.9399	-73.5491	0	.	0
11	Canada	Q	2020-02-01	52.9399	-73.5491	0	.	0
12	Canada	Q	2020-02-02	52.9399	-73.5491	0	.	0
13	Canada	Q	2020-02-03	52.9399	-73.5491	0	.	0
14	Canada	Q	2020-02-04	52.9399	-73.5491	0	.	0
15	Canada	Q	2020-02-05	52.9399	-73.5491	0	.	0
16	Canada	Q	2020-02-06	52.9399	-73.5491	0	.	0
17	Canada	Q	2020-02-07	52.9399	-73.5491	0	.	0
18	Canada	Q	2020-02-08	52.9399	-73.5491	0	.	0
19	Canada	Q	2020-02-09	52.9399	-73.5491	0	.	0
20	Canada	Q	2020-02-10	52.9399	-73.5491	0	.	0
21	Canada	Q	2020-02-11	52.9399	-73.5491	0	.	0
22	Canada	Q	2020-02-12	52.9399	-73.5491	0	.	0
23	Canada	Q	2020-02-13	52.9399	-73.5491	0	.	0
24	Canada	Q	2020-02-14	52.9399	-73.5491	0	.	0
25	Canada	Q	2020-02-15	52.9399	-73.5491	0	.	0
26	Canada	Q	2020-02-16	52.9399	-73.5491	0	.	0
27	Canada	Q	2020-02-17	52.9399	-73.5491	0	.	0
28	Canada	Q	2020-02-18	52.9399	-73.5491	0	.	0

```
proc print data=libraryy.dataset label
;

var Date 'Province/State'n
Confirmed;
run;
```

Displaying the Data

Using Proc print to display only Date provinces and Confirmed cases gives the following

Corona Cases in Quebec

Obs	Date	Province/State	Confirmed
1	2020-01-22	Q	0
2	2020-01-23	Q	0
3	2020-01-24	Q	0
4	2020-01-25	Q	0
5	2020-01-26	Q	0
6	2020-01-27	Q	0
7	2020-01-28	Q	0
8	2020-01-29	Q	0
9	2020-01-30	Q	0
10	2020-01-31	Q	0

```
Data libraryy.dataset;
Set libraryy.dataset;
 Zero = 0;
 run;

title 'Corona Cases in Quebec';
footnote j=l 'Bar Chart on Discrete Axis';

ods graphics on /  /*Chart characteristics*/
    width=30 in
            height=30 in
    outputfmt=gif
    imagemap=on
    imagename="MyBoxplot"
    border=off;

proc sgplot data=libraryy.dataset;
  refline 1 1.5 2 / lineattrs=graphgridlines;
  highlow x=Date high=Confirmed low=Zero / type=bar
  group='Province/State'n
  groupdisplay=cluster lineattrs=(color=black);
  xaxis discreteorder=data display=(nolabel);
  yaxis label='Value (/ULN)' offsetmin=0;;
  run;
```

Plotting the Data Using SGPLOT

Explaining SGPLOT Code

```
Data libraryy.dataset;
Set libraryy.dataset;
 Zero = 0;
 run;

title 'Corona Cases in Quebec';
footnote j=l 'Bar Chart on Discrete Axis';

ods graphics on /  /*Chart characteristics*/
    width=30 in
            height=30 in
    outputfmt=gif
    imagemap=on
    imagename="MyBoxplot"
    border=off;

proc sgplot data=libraryy.dataset;
  refline 1 1.5 2 / lineattrs=graphgridlines;
  highlow x=Date high=Confirmed low=Zero / type=bar
  group='Province/State'n
  groupdisplay=cluster lineattrs=(color=black);
  xaxis discreteorder=data display=(nolabel);
  yaxis label='Value (/ULN)' offsetmin=0;;
  run;
```

Add a zero value for reference

Title the dataset's chart.

ODS Graphics for:
- Customizing graph
- Width / Height
- Removing border

- High/Low Chart
- Group Display: Cluster

To begin with, we know that the virus originated from china. So we tracked each country of our data set and observed when they each observed their first case of Corona.

1. First we sort the data by date and take all the cases where there is only 1 case of corona confirmed

2. Using proc Sql we are using a join to eliminate the redundant case of confirmed case 1 and take only the minimum date per country and per state

3. Plotting a scatter plot to show the result with refline for easier reading

```
/*1*/
proc sort data=probly out=sorted;
where Confirmed=1;
by  Date;
by 'Province/State'n;
run;


/*2*/
proc sql;
create table want as
select a.*
from sorted a, (select 'Country/Region'n,'Province/State'n,Confirmed as
_regimen  from sorted group by 'Province/State'n having date=max(date)) b
where a.'Province/State'n = b.'Province/State'n
group by a.'Country/Region'n ,a.'Province/State'n
having date=min(date)
and _regimen=Confirmed;
quit;

/*3*/
PROC SGPLOT Data = want;
    scatter x = Date   y = 'Country/Region'n;
        refline 'Country/Region'n /axis= y;
        refline Date /axis= x;
    title "First Case in various Countries";
run;
```

|	Date	Country/Regi...	Province/Sta...	Lat	Long	Confirmed	Recovered	Deaths
1	2020-02-24	Afghanistan		33	65	1	0	0
2	2020-02-25	Algeria		28.0339	1.6596	1	0	0
3	2020-03-02	Andorra		42.5063	1.5218	1	0	0
4	2020-03-20	Angola		-11.2027	17.8739	1	0	0
5	2020-03-13	Antigua and		17.0608	-61.7964	1	0	0
6	2020-03-03	Argentina		-38.4161	-63.6167	1	0	0
7	2020-03-01	Armenia		40.0691	45.0382	1	0	0
8	2020-03-13	Australia	A	-35.4735	149.0124	1	0	0
9	2020-03-04	Australia	N	-12.4634	130.8456	1	0	0
10	2020-01-29	Australia	Q	-28.0167	153.4	1	0	0
11	2020-02-01	Australia	S	-34.9285	138.6007		0	0
12	2020-03-02	Australia	T	-41.4545	145.9707		0	0
13	2020-01-26	Australia	V	37.8136	144.9631			0
14	2020-03-16	Bahamas			-77.39			0
15	2020-02-24	Bahrain						0
16	2020-02-28	Belarus		53.709				0
17	2020-02-04	Belgium		50.8333				0
18	2020-03-23	Belize		13.1939	-5			0
19	2020-03-16	Benin		9.3077				0
20	2020-03-06	Bhutan		27.5142	9			0
21	2020-02-26	Brazil		-14.235	-5			0
22	2020-03-09	Brunei		4.5353	114.			0
23	2020-03-10	Burkina Fas		12.2383	-1.56			0
24	2020-03-20	Cabo Verde		16.5388	-23.041			0
25	2020-01-27	Cambodia		11.55	104.9167			
26	2020-03-06	Cameroon		3.848	11.5021		0	0
27	2020-03-06	Canada	A	53.9333	-116.5765	1		0
28	2020-01-28	Canada	B	49.2827	-123.1207	1		0
29	2020-03-11	Canada	N	46.5653	-66.4619	1		0
30	2020-01-26	Canada	O	51.2538	-85.3232	1		0
31	2020-03-15	Canada	P	46.5107	-63.4168	1		0
32	2020-02-28	Canada	Q	52.9399	-73.5491	1		
33	2020-03-15	Central Afr		6.6111	20.9394	1	0	0
34	2020-03-19	Chad		15.4542	18.7322	1	0	0
35	2020-03-03	Chile		-35.6751	-71.543	1	0	0
36	2020-01-22	China	A	31.8257	117.2264	1	0	0
37	2020-01-22	China	F	26.0789	117.9874	1	0	0
38	2020-01-22	China	G	26.8154	106.8748	1	0	0
39	2020-01-22	China	H	39.549	116.1306	1	0	0
40	2020-01-24	China	I	44.0935	113.9448	1	0	0
41	2020-01-22	China	J	32.9711	119.455	1	0	0
42	2020-01-22	China	M	22.1667	113.55	1	0	0
43	2020-01-22	China	N	37.2692	106.1655	1	0	0
44	2020-01-25	China	Q	35.7452	95.9956	1	0	0
45	2020-01-22	China	S	37.5777	112.2922	1	0	0
46	2020-01-30	China	T	31.6927	88.0924	1	0	0
47	2020-01-22	China	Y	24.974	101.487	1	0	0
48	2020-03-06	Colombia		4.5709	-74.2973	1	0	0

As we can observe, we get the countries and the various states in various countries

First Case in various Countrues
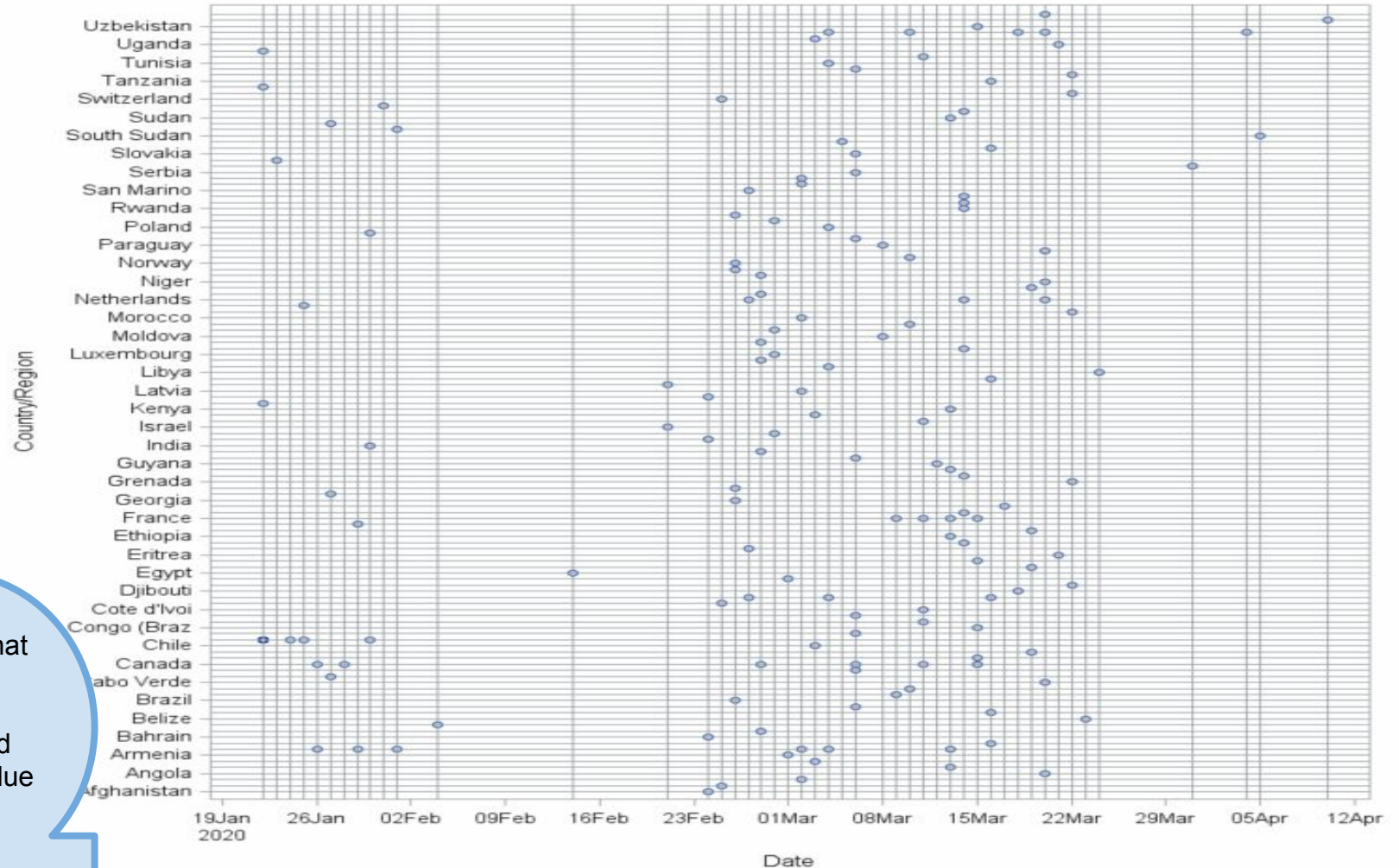
Not all countries that were described above could be shown due to lack of space

If we limit the above result to focus only on the first cases in Canada

1. To achieve this, we simply filter the country to Canada

2. Stays the same

3. Our Scatter plot will be based on Provinces

```
/*1*/
proc sort data=probly out=sorted;
where 'Country/Region'n ='Canada' and Confirmed=1;
by  Date;
by 'Province/State'n;
run;


/*2*/
proc sql;
create table want as
select a.*
from sorted a, (select 'Country/Region'n,'Province/State'n,Confirmed as
_regimen  from sorted group by 'Province/State'n having date=max(date)) b
where a.'Province/State'n = b.'Province/State'n
group by a.'Country/Region'n ,a.'Province/State'n
having date=min(date)
and _regimen=Confirmed;
quit;

/*3*/
PROC SGPLOT Data = want;
    scatter x = Date   y = 'Province/State'n;
          refline 'Province/State'n /axis= y;
          refline Date /axis= x;
/*        group= 'Country/Region'n, 'Province/State'n */
/*        groupdisplay=cluster lineattrs=(color=black);*/
    title "First Case in Canada Provinces";
run;
```

	Date	Country/Regi...	Province/Sta...		Lat		Long		Confirmed		Recovered		Deaths
1	2020-01-26	Canada	O		51.2538		-85.3232		1		.		
2	2020-01-27	Canada	O		51.2538		-85.3232		1		.		
3	2020-01-28	Canada	B		49.2827		-123.1207		1		.		
4	2020-01-28	Canada	O		51.2538		-85.3232		1		.		
5	2020-01-29	Canada	B		49.2827		-123.1207				.		
6	2020-01-29	Canada	O		51.2538								
7	2020-01-30	Canada	B		49.2827								
8	2020-01-30	Canada	O		51.2538		-85						
9	2020-01-31	Canada	B		49.2827		-123.1						
10	2020-02-01	Canada	B		49.2827		-123.1207						
11	2020-02-02	Canada	B		49.2827		-123.1207				.		
12	2020-02-03	Canada	B		49.2827		-123.1207		1		.		
13	2020-02-04	Canada	B		49.2827		-123.1207		1		.		
14	2020-02-28	Canada	Q		52.9399		-73.5491		1		.		
15	2020-02-29	Canada	Q		52.9399		-73.5491		1		.		
16	2020-03-01	Canada	Q		52.9399		-73.5491		1		.		
17	2020-03-02	Canada	Q		52.9399		-73.5491		1		.		
18	2020-03-03	Canada	Q		52.9399		-73.5491		1		.		
19	2020-03-04	Canada	Q		52.9399		-73.5491		1		.		
20	2020-03-06	Canada	A		53.9333		-116.5765		1		.		
21	2020-03-11	Canada	N		46.5653		-66.4619		1		.		

As we can observe, we get the provinces

Date of First Case in Different Canada Provinces
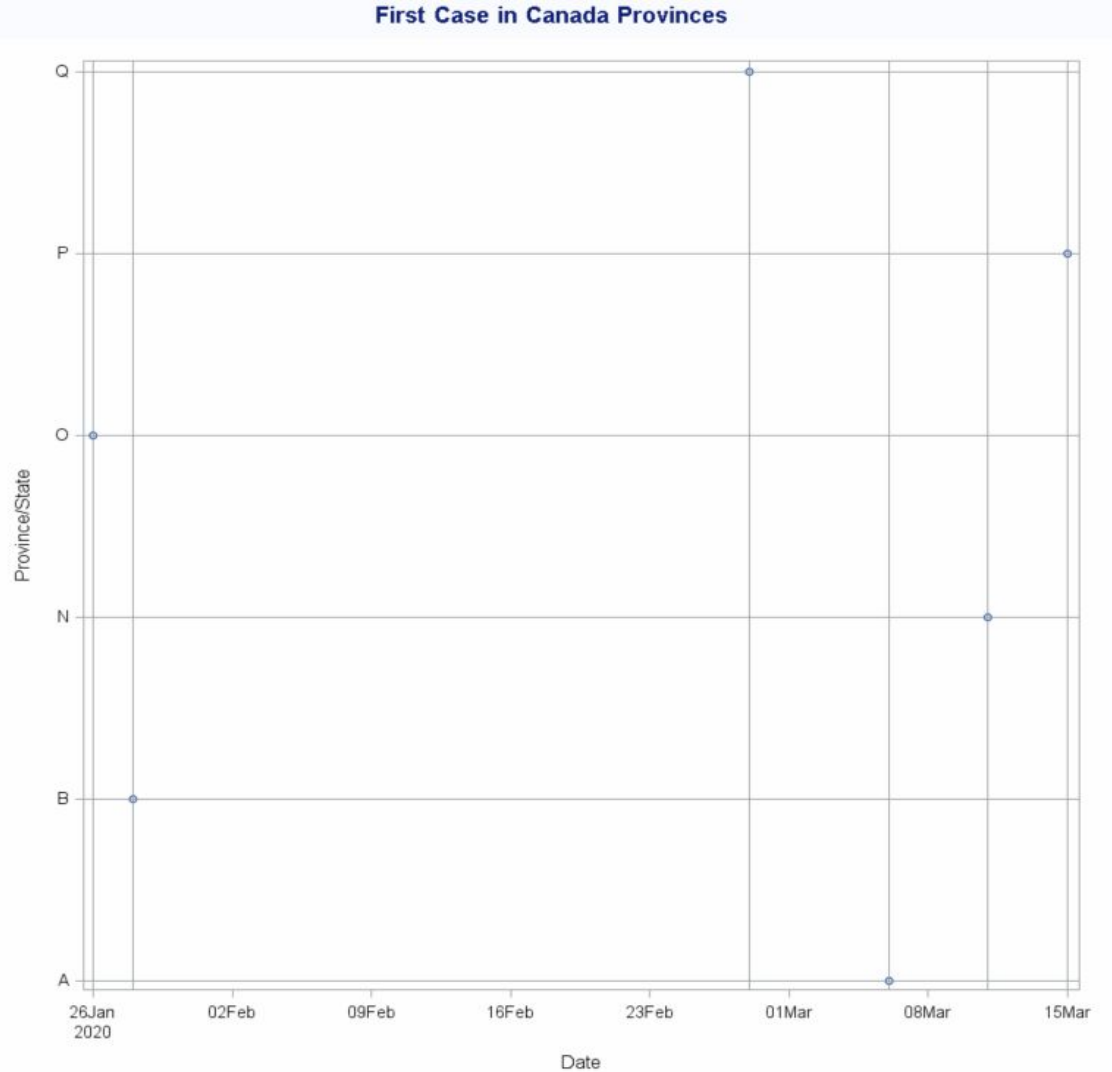
The confirmed cases in Canada

1. We filter Canada as a country

2. Specified our graph width

3. plotting with a high low sgplot grouping each provinces

```
Data libraryy.dataset;
Set probly
(where=            (                     'Country/Region'n = 'Canada'

          ));


/*All cases in Canada confirmed cases*/

title 'Corona Cases in Canadas provinces';
footnote j=l 'Bar Chart on Discrete Axis';

ods graphics on /  /*Chart characteristics*/
    width=10 in
          height=10 in
    outputfmt=gif
    imagemap=on
    imagename="MyBoxplot"
 ;


proc sgplot data=libraryy.dataset;
  refline 1 1.5 2 / lineattrs=graphgridlines;
  highlow x=Date high=Confirmed low=Zero / type=bar
  group='Province/State'n
  groupdisplay=cluster lineattrs=(color=black);
  xaxis discreteorder=data display=(nolabel);
  yaxis label='Value (/ULN)' offsetmin=0;;
  run;
```

Coronavirus Cases in Canada Provinces

Corona Cases in Canadas provinces
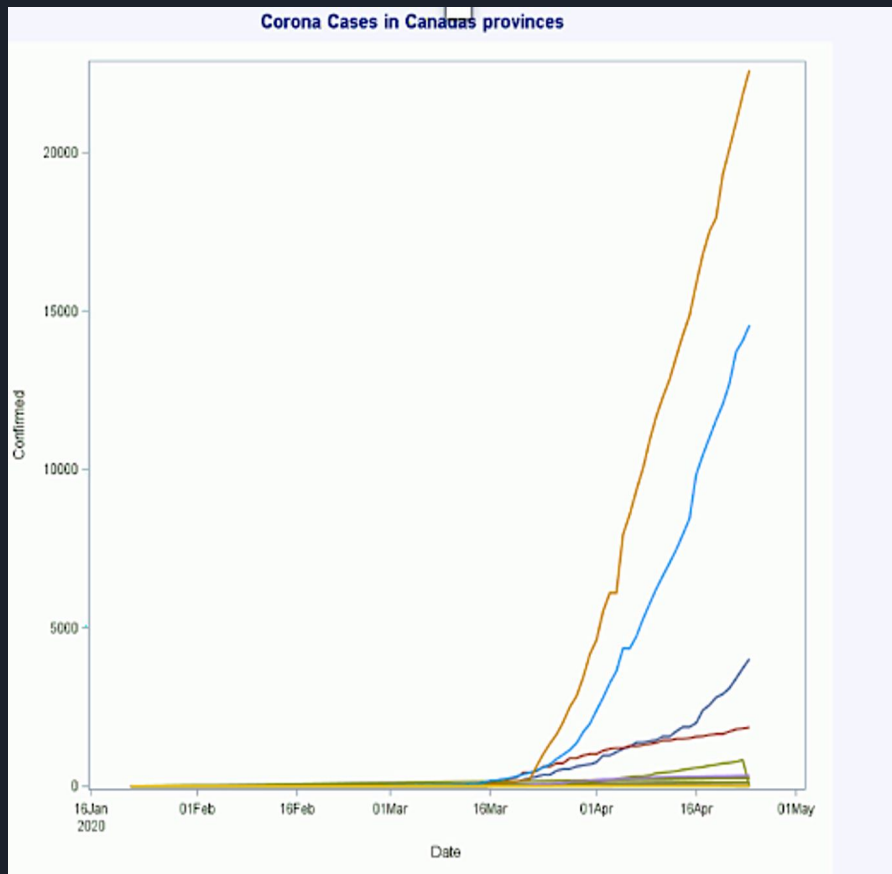
Quebec(Q) has the highest Number of Confirmed cases

We can observe the exponential increase in the number of cases.
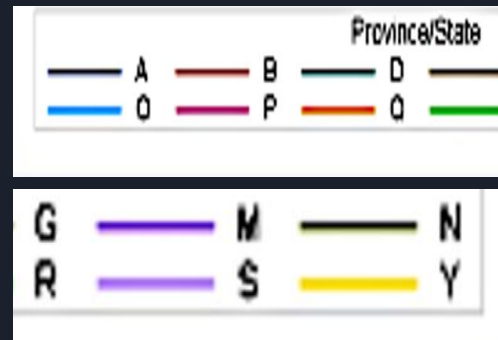
Logarithmized confirmed number of cases

1. Everything stays the same as the confirmed cases in canada

2. We add a new column and logConfirmed

```
Data libraryy.dataset;
Set libraryy.dataset;
 Zero = 0;
  logConfirmed = log(Confirmed);
 run;
```
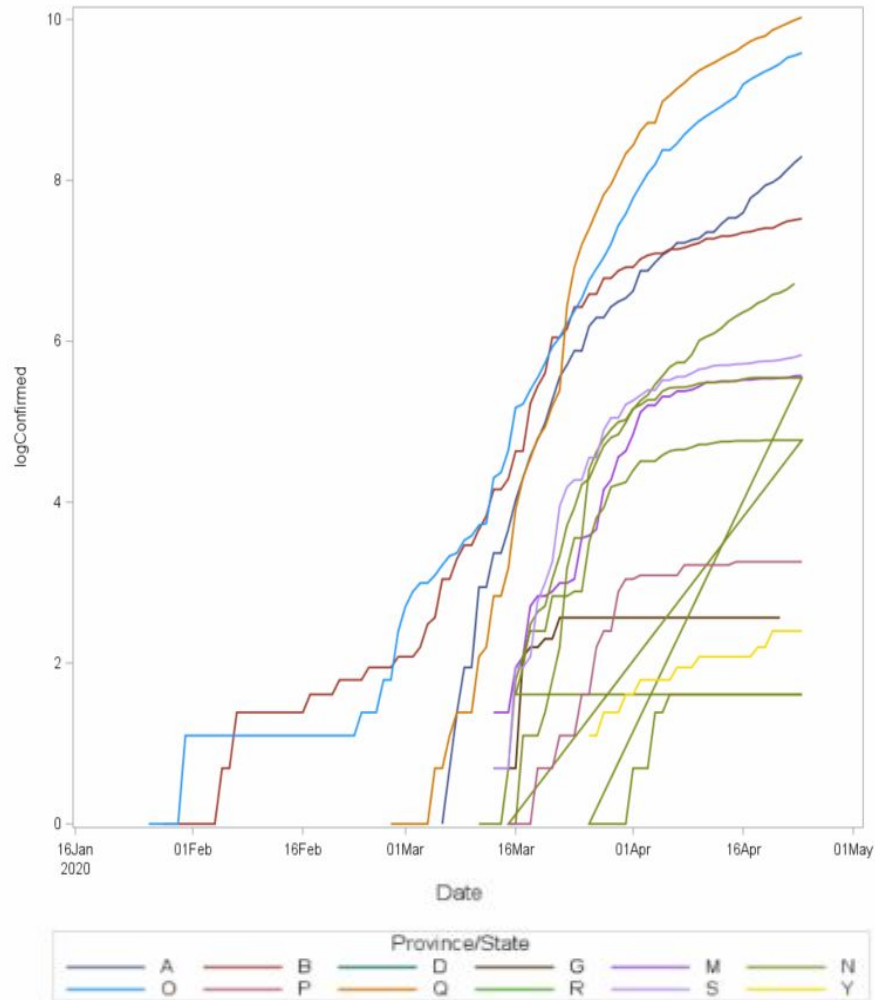
Coronavirus Cases (Log)

Creating a variable log of the deaths for the vertical scale versus the days in the horizontal scale, the slope of the curve shows the growth rate of deaths.

With this information, we can analyze which provinces are slowing down in terms of deaths.

We do this because analyzing an exponential graph to figure out when the growth slows down is tricky.

Finding the Number of Cases Solely in Quebec

Filter Canada as a country and Quebec as a province

```
 Data QuebecConfirmed;
Set libraryy.dataset(where=
         (                    'Country/Region'n = 'Canada'
               and
         ('Province/State'n ='Q'            )        )         );
 Zero = 0;

 run;


title 'Corona Cases in Quebec';
footnote j=l 'Bar Chart on Discrete Axis';

ods graphics on /  /*Chart characteristics*/
     width=30 in
          height=30 in
     outputfmt=gif
     imagemap=on
     imagename="MyBoxplot"
     border=off;


PROC SGPLOT data=libraryy.dataset;
 SERIES X = Date Y = logConfirmed / LEGENDLABEL = 'Quebec'
 group='Province/State'n
  LINEATTRS = (THICKNESS = 2);
XAXIS TYPE = TIME;

run;
```
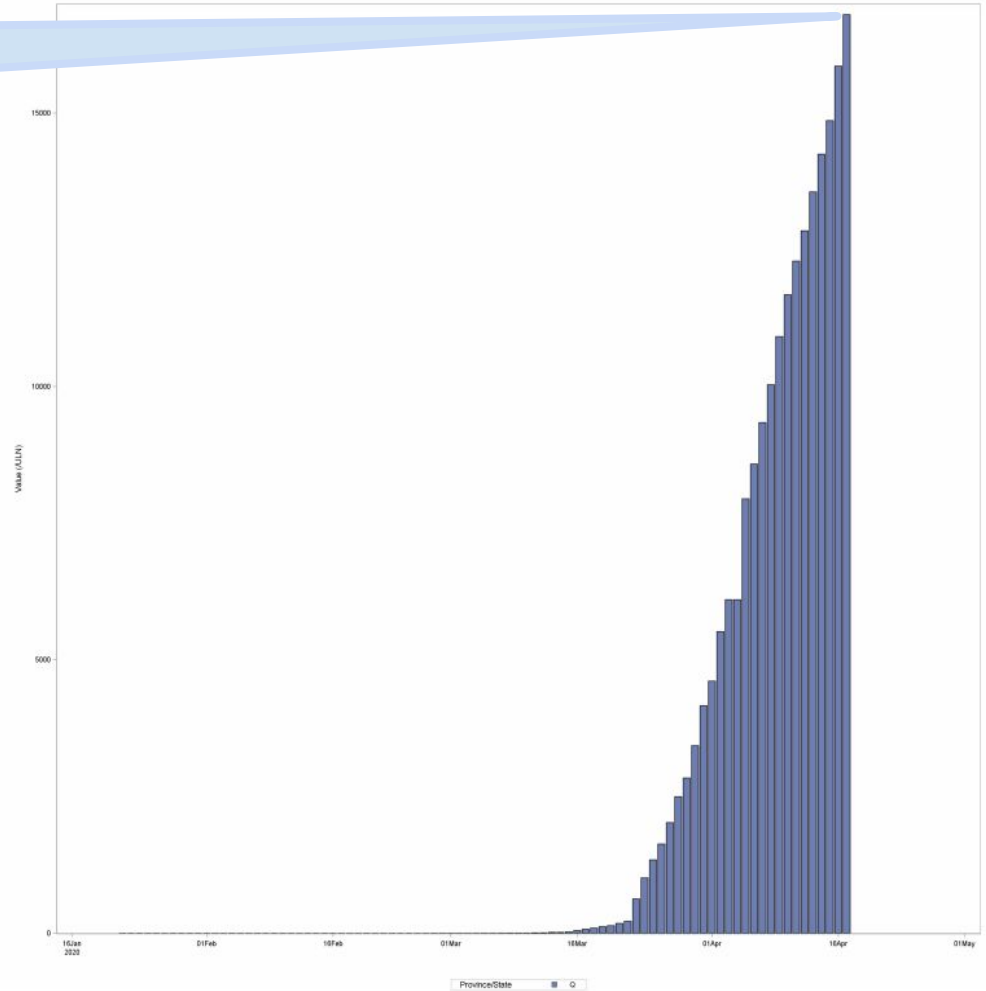
Countries with the Highest Rate of Death

```
proc sql;
create table want as
select a.*


from sorted a,
        (select 'Country/Region'n,'Province/State'n,Deaths
                as _regimen  from sorted group by 'Province/State'n having
date=max(date)) b
where a.'Country/Region'n  = b.'Country/Region'n
group by  a.'Country/Region'n
/*,Deaths,_regimen */
having date=max(date)
and _regimen=Deaths
order by _regimen desc
;


quit;
```

Countries with the Highest Rate of Death

|	Date	Country/Regi...	Province/Sta...	Lat	Long	Confirmed	Recovered	Deaths
1	2020-04-24	US		37.0902	-95.7129	905358	99079	51949
2	2020-04-24	Italy		43	12	192994	60498	25969
3	2020-04-24	Spain		40	-4	219764	92355	22524
4	2020-04-24	France		46.2276	2.2137	158636	43493	22245
5	2020-04-24	United King		55.3781	-3.436	143464	0	19506
6	2020-04-24	Belgium		50.8333	4	44293	10122	6679
7	2020-04-24	Germany		51	9	154999	109800	5760
8	2020-04-24	Iran		32	53	88194	66599	5574
9	2020-04-24	China	H	30.9756	112.2707	68128	63593	4512
10	2020-04-24	Netherlands		52.1326	5.2913	36535	0	4289
11	2020-04-24	Brazil		-14.235	-51.9253	54043	27655	3704
12	2020-04-24	Turkey		38.9637	35.2433	104912	21737	2600
13	2020-04-24	Sweden		63	16	17567	1005	2152
14	2020-04-24	Switzerland		46.8182	8.2275	28677	21000	1589
15	2020-04-24	Canada	Q	52.9399	-73.5491	22616	.	1340
16	2020-04-24	Mexico		23.6345	-102.5528	11633	2627	1069
17	2020-04-24	Ireland		53.1424	-7.6921	18184	9233	1014
18	2020-04-24	Canada	O	51.2538	-85.3232	14550	.	862

Mortality Rate Among Top 20 Countries

➢ We know that just the mortality rate by itself does not mean too much.

➢ So, we took the mortality rate to get an estimate of the number of death in proportion to the number of confirmed cases

➢ We make three dataset for easiness of reuse later. But the ultimate aim is to get the top 20 countries having the highest percentage Deaths/ConfirmedCases

```
Data wantDeaths10;
Set want;
if (Deaths > 100) then do;
  MortalityRate = Deaths/Confirmed*100;
  end;
  label MortalityRate = "Rate of Mortality";
/* by Deaths;*/
 run;


 proc sort data=wantDeaths10 out=wantDeathsTop10;
/*where */
by descending MortalityRate ;



run;

Data wantDeaths20;
zero =0;
Set wantDeathsTop10(obs=20);

 run;
```

	zero	Date	Country/Regi...	Province/Sta...	Lat	Long	Confirmed	Recovered	Deaths	MortalityRate
1	0	2020-04-24	Belgium		50.8333	4	44293	10122	6679	15.079132143
2	0	2020-04-24	France		46.2276	2.2137	158636	43493	22245	14.022668247
3	0	2020-04-24	United King		55.3781	-3.436	143464	0	19506	13.596442313
4	0	2020-04-24	Italy		43	12	192994	60498	25969	13.455858731
5	0	2020-04-24	Algeria		28.0339	1.6596	3127	1408	415	13.271506236
6	0	2020-04-24	Sweden		63	16	17567	1005	2152	12.250241931
7	0	2020-04-24	Netherlands		52.1326	5.2913	36535	0	4289	11.739427946
8	0	2020-04-24	Hungary		47.1625	19.5033	2443	458	262	10.724519034
9	0	2020-04-24	Spain		40	-4	219764	92355	22524	10.249176389
10	0	2020-04-24	Mexico		23.6345	-102.5528	11633	2627	1069	9.1893750537
11	0	2020-04-24	Indonesia		-0.7893	113.9213	8211	1002	689	8.39118256
12	0	2020-04-24	Egypt		26	30	4092	1075	294	7.1847507331
13	0	2020-04-24	Brazil		-14.235	-51.9253	54043	27655	3704	6.8538016024
14	0	2020-04-24	Philippines		13	122	7192	762	477	6.6323692992
15	0	2020-04-24	China	H	30.9756	112.2707	68128	63593	4512	6.6228276186
16	0	2020-04-24	Iran		32	53	88194	66599	5574	6.3201578339
17	0	2020-04-24	Canada	Q	52.9399	-73.5491	22616	.	1340	5.9250088433
18	0	2020-04-24	Canada	O	51.2538	-85.3232	14550	.	862	5.9243986254
19	0	2020-04-24	US		37.0902	-95.7129	905358	99079	51949	5.7379511751
20	0	2020-04-24	Ireland		53.1424	-7.6921	18184	9233	1014	5.5763308403

Top Countries by Mortality Rate
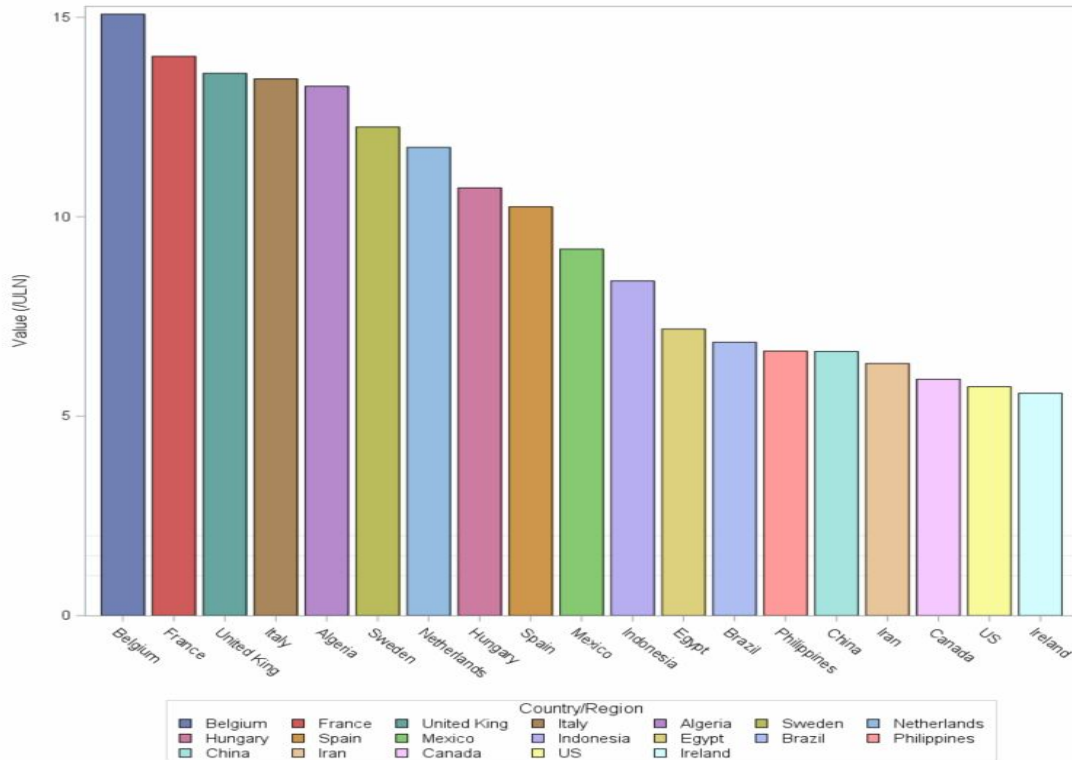
Recovery Rate Among Top 20 Countries

➢ So, we took the recovered Patients/confirmed cases to find the percentage (rate) of recoveries.

➢ We plotted the Top 10 countries with the best recovery rates.

```
Data wantDeaths10;
Set want;
if (Recovered > 100) then do;
  RecoveredRate = Recovered/Confirmed*100;
  end;
  label RecoveredRate = "Rate of Recovery";
/* by Deaths;*/
 run;

 proc sort data=wantDeaths10 out=wantDeathsTop10;
/*where */
by descending RecoveredRate ;


run;

Data wantDeaths20;
zero =0;
Set wantDeathsTop10(obs=100);

 run;

PROC SGPLOT Data = wantDeaths20;
 refline 1 1.5 2 / lineattrs=graphgridlines;
  highlow x='Country/Region'n   high=RecoveredRate low=Zero / type=bar
  group='Country/Region'n
  groupdisplay=cluster lineattrs=(color=black);
  xaxis discreteorder=data display=(nolabel);
  yaxis label='Value (/ULN)' offsetmin=0;;
   title "Top 20 countries mortality rate";
run;
```
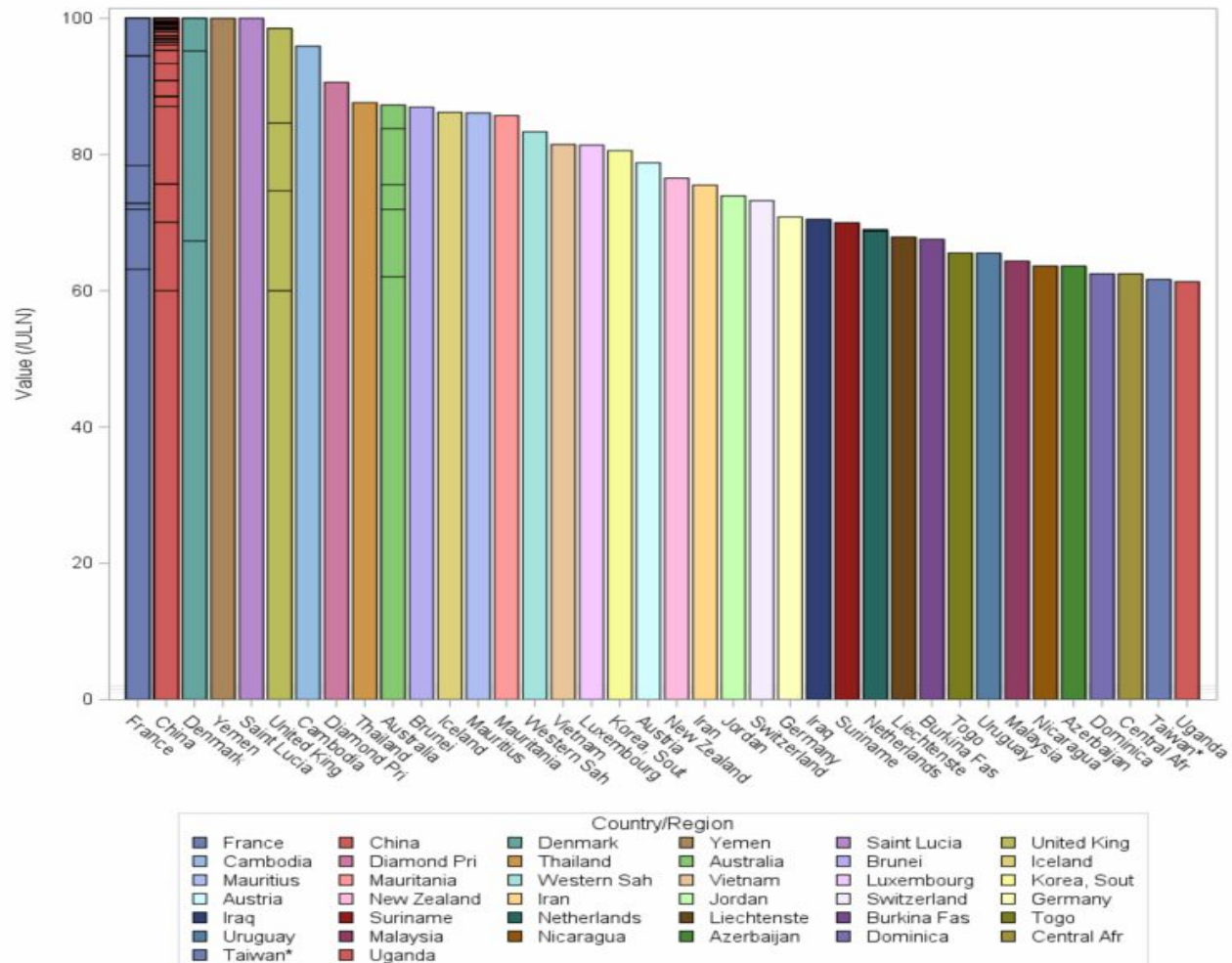
Most Confirmed COVID-19 Cases Per Capita

➢ A better estimate would be the number of confirmed cases per capita

➢ totalConfirmed cases / Population Size *100

➢ First we prepare the data for the population and sort both data in order of countries

```
proc import
datafile="F:\school\BSTA445\population\data\population.csv"
out=mydata dbms=csv replace;
run;


data ret;
set pop.sasSort;
run;

proc sort data=mydata out=sortedPop2;
/*  by country;*/
/* by province;*/
/*   by descending date_confirmation ;*/

by 'Country Name'n;

/*/*by 'Country/Region'n;*/*/;


run;
```

Most Confirmed COVID-19 Cases Per Capita

➢ Then we do a left to keep all our data and merge it with the population csv

➢ We take the latest dataset for population which is 2018
➢ Then we apply our formula to get the confirmed cases per capita

```
PROC SQL;
CREATE TABLE C AS
select *

FROM  totalby as A left JOIN  sortedPop2 as B

ON A.'Country/Region'n =B.'Country Name'n
where
B.Year=2018
/*and */
/*B.'Country Name'n is null*/
;
QUIT;


Data wantDeaths10;
Set C;
Zero=0;
if (Value > 0) then do;
  MortalityRate = (TConfirmed/Value)*100;
 end;
 label MortalityRate = "Contamination percentage";
/* by Deaths;*/
 run;
proc sort data=wantDeaths10(DROP=Deaths Recovered
Confirmed ) out=wantDeaths10sorted;
  by descending MortalityRate;
run;


data we;
 set wantDeaths10sorted(obs=30);
 run;
```
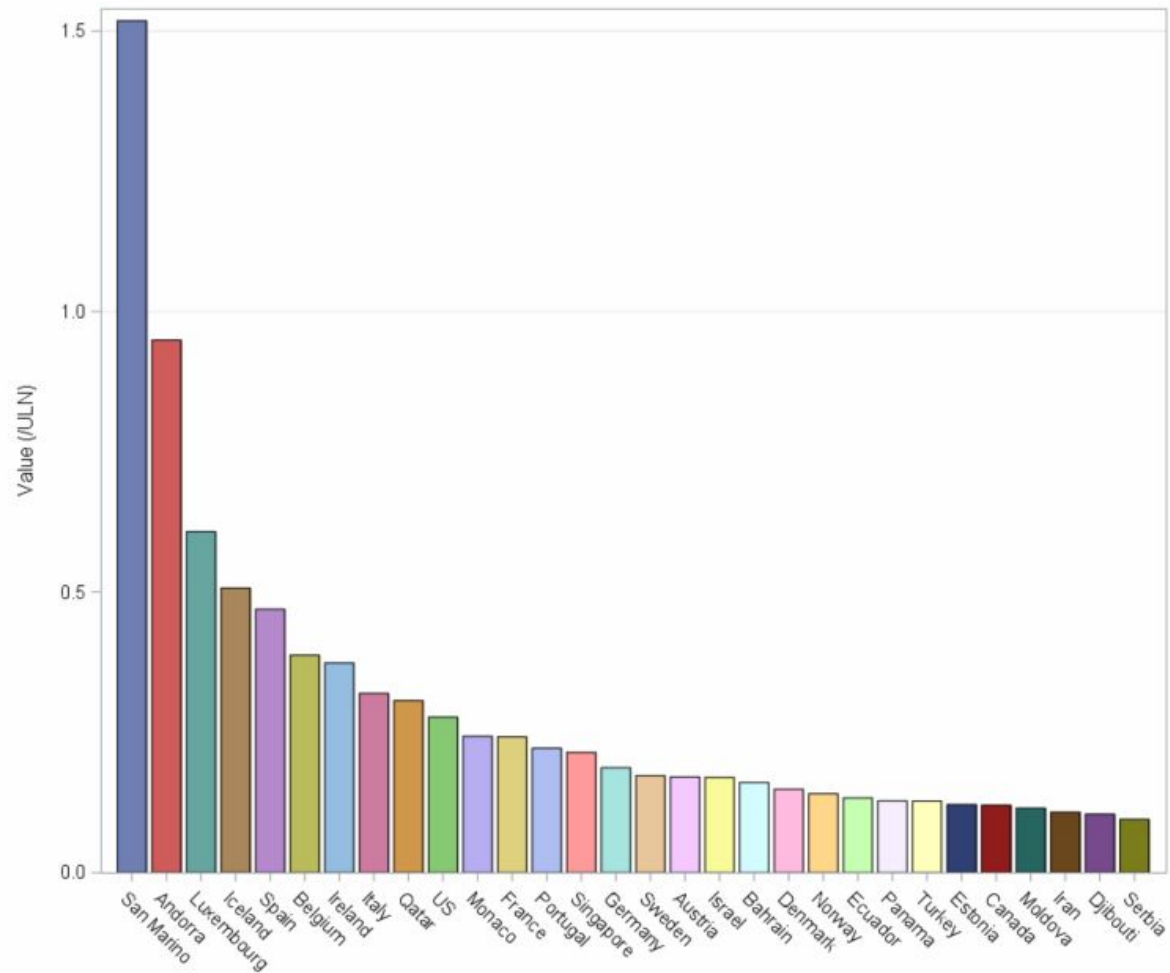
COVID-19 Most Deaths per Capita

➢ A better estimate would be the number of confirmed cases per capita

➢ total Deathcases / Population Size *100

➢ First we prepare the data for the population and sort both data in order of countries

```
proc import datafile="F:\school\BSTA445\population\data\population.csv"
out=mydata dbms=csv replace;
run;



data ret;
set pop.sasSort;
run;

proc sort data=mydata out=sortedPop2;
/*  by country;*/
/* by province;*/
/*   by descending date_confirmation ;*/

by 'Country Name'n;

/*/*by 'Country/Region'n;*/*/;



run;
```
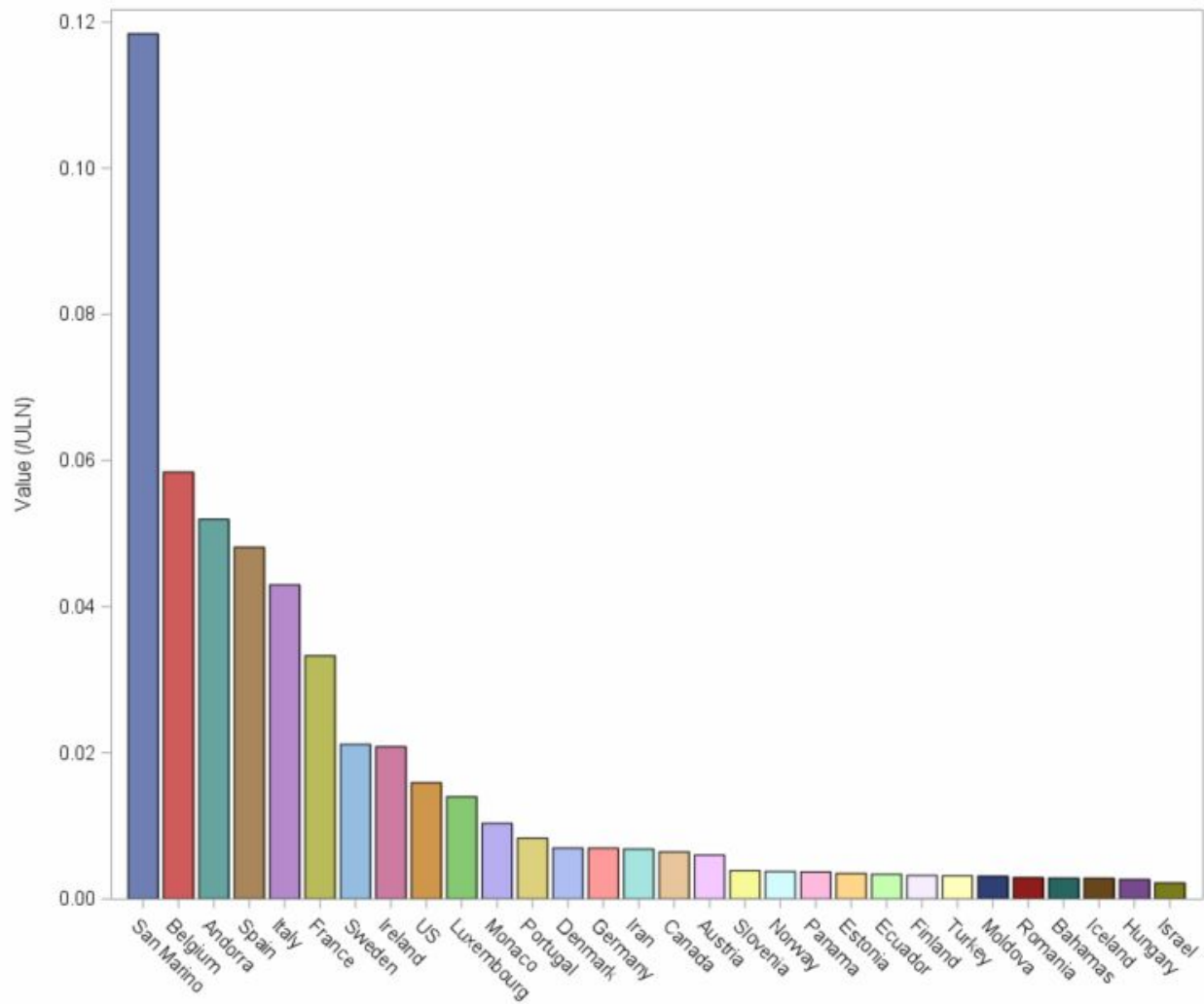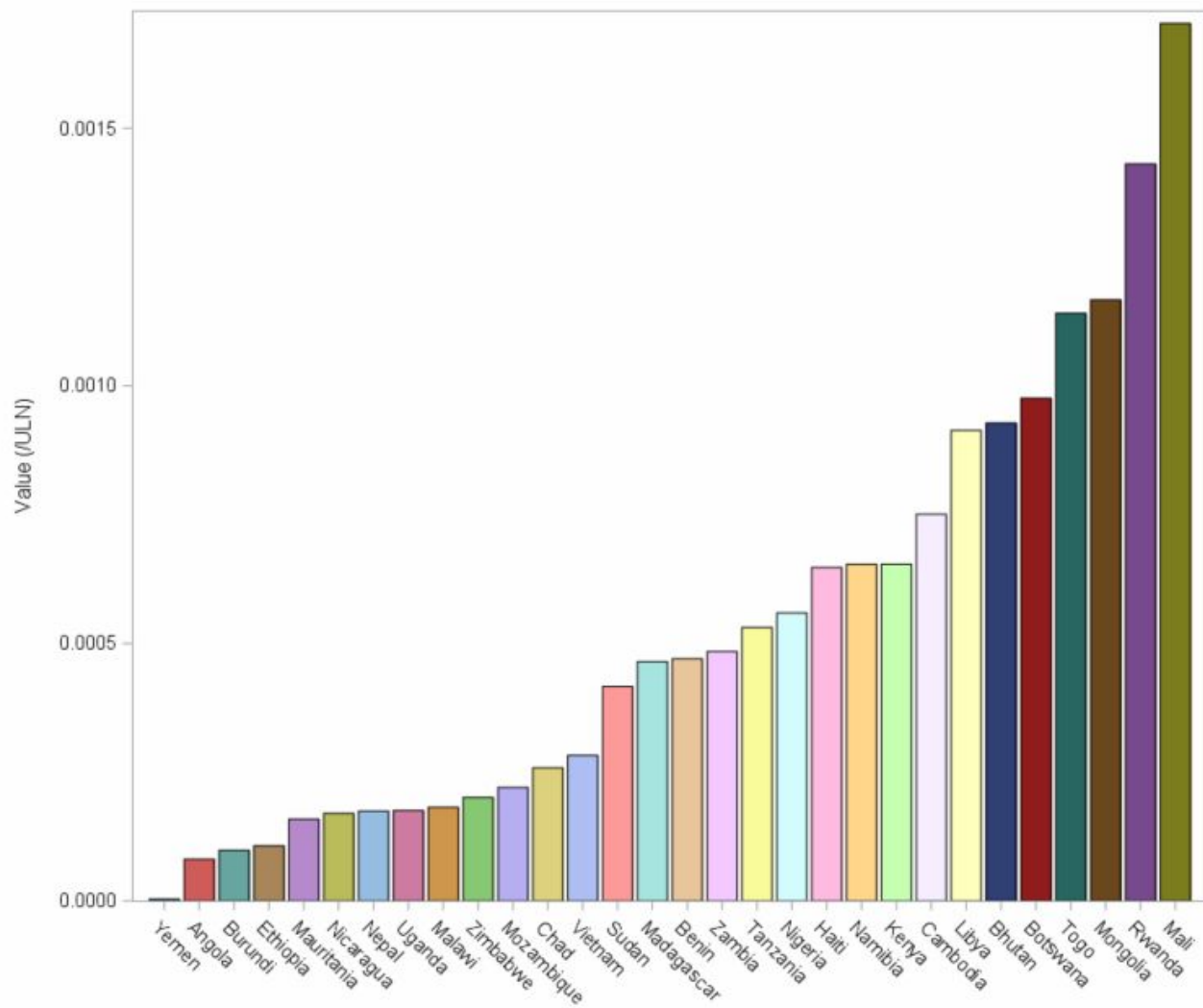
Least Confirmed COVID-19 Cases Per Capita

➢ We keep the same code as in the previous example but we do not in descending order.

```
Data wantDeaths10;
Set C;
Zero=0;
if (Value > 0) then do;
   TConfirmedN= (TConfirmed/Value)*100;
  end;
  label MortalityRate = "Contamination percentage";
/* by Deaths;*/
 run;


proc sort data=wantDeaths10(DROP=Deaths Recovered Confirmed )
out=wantDeaths10sorted;
  by  TConfirmedN;
run;
```
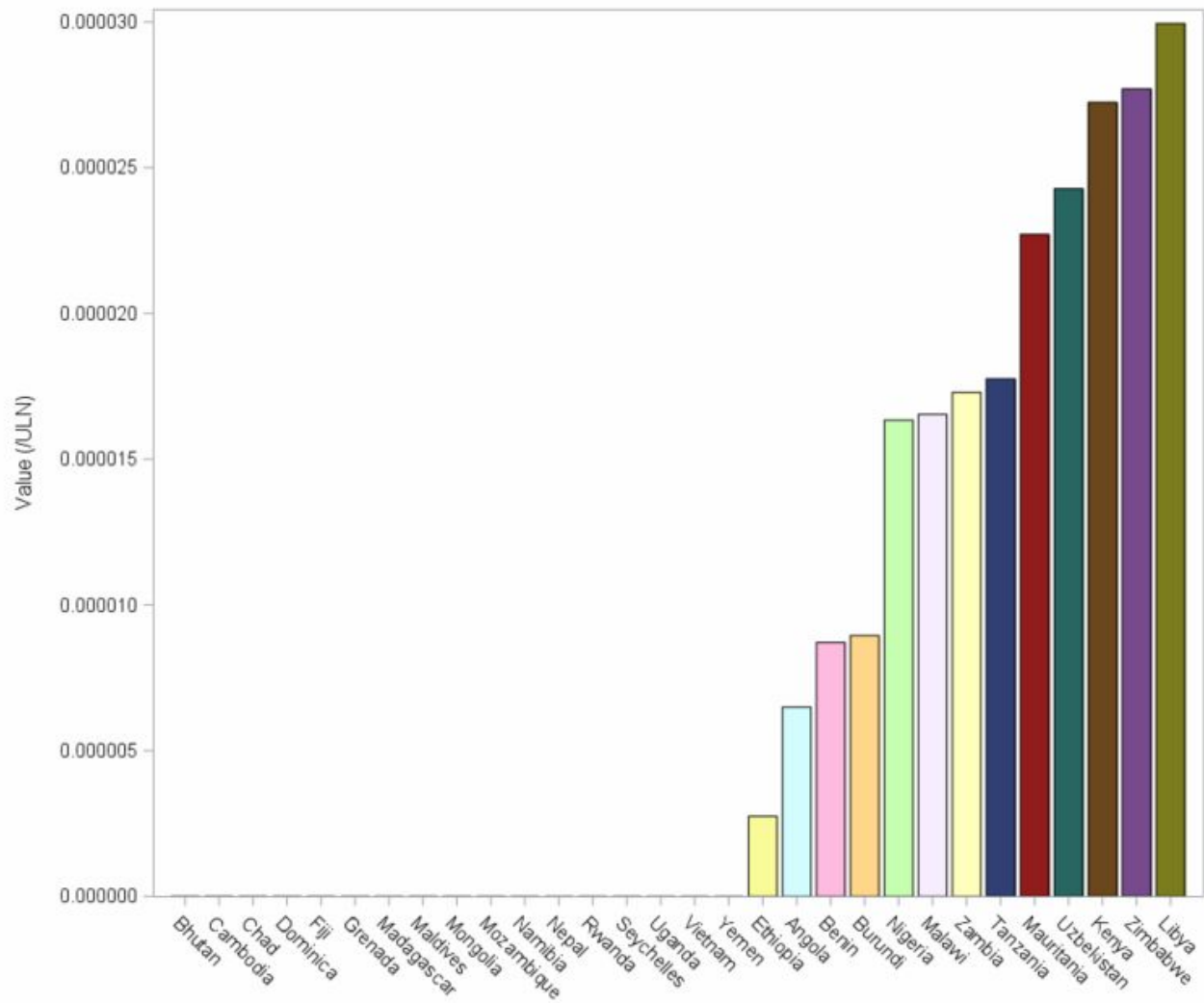
Least Deaths COVID-19 Cases Per Capita

➢ We keep the same code as in the previous example but we do not in descending order.

```
Data wantDeaths10;
Set C;
Zero=0;
if (Value > 0) then do;
  MortalityRate = (TDeaths/Value)*100;
  end;
  label MortalityRate = "Contamination percentage";
/* by Deaths;*/
 run;

proc sort data=wantDeaths10(DROP=Deaths Recovered Confirmed )
out=wantDeaths10sorted;
  by  MortalityRate;
run;
```

Conclusion

We can observe there is a link between the number of confirmed and the number of deaths with covid-19 as our most and least death per capita and confirmed cases per capita showed us.

We have seen that many countries are on the right track for recovery. But drastic measures taken by countries to stay at home and take precautions while having a lockdown seems to be more effective.

Countries like Monaco, Vietnam, Madagascar, Ethopia, Nigeriam Andorra, San Marino, Mauritius, Australia, UK, Yemen, Jordan have taken the extra step to go on lockdown earlier and / or have taken drastic measures. For example Mauritius have segmentalized it's consumer's purchasing habits by imposing regulations that forced consumers to shop only 2 times a week . These countries' strategy lead to the least confirming/ deaths rates while Belgium, Italy, US, France , Spain, Canada, Quebec, Israel, and others don't seem to have implemented the best measures or have adopted them late.

Limitations

However, we have to be considerate, the countries with the lowest rates of deaths and confirmed cases are not targeted by massive incoming flights(Example Canada Italy) or high political barriers(Canada couldn't close entry from us but later did) or high bureaucratic measures that has to go through many levels to be approved.

Each and every country has suffered through this pandemic, and my team and i seriously hope that so each economy can get on their feet with as less damage as possible and if not much, but sufficient experience has been gathered to break through the next pandemic.

What we should take from this huge economic collapse is even the least spark detected from far in any remote country should be taken seriously and that the biggest thing to fear is not wars or anything but small microbes (as bill gates predicated in a ted talk 4 years ago).

Disclaimer

We guarantee in no way that the information presented above is 100% accurate since :

1)Dataset change and vary every second and we use only the source from one dataset.

2) Our population dataset dates back to 2018 and might not reflect up to date values

3) Our merging of population dataset with our corona dataset went through some data loss but it wasn't of a huge impact and ascertained that the values were negligible

4)We thank the professor Dr. Krzysztof Dzieciolowski , the T.A Dariia Dziuba , github repositories, sas documentation and other 3rd parties for their direct and indirect help.

5)Our code is accessible from github.

 We apologize beforehand for any wrongly stated fact or any missing recognition, feel free to consult/add/commit/review on the github repository

Bibliography

- ★ https://blogs.sas.com/content/graphicallyspeaking/2012/04/16/bar-chart-on-interval-axis/#prettyPhoto

- ★ https://github.com/datasets/covid-19

- ★ https://gist.github.com/cjdinger/2950c1a62acd7f50c33473c311b0d9fe

- ★ https://github.com/NajiElKotob/COVID-19

- ★ https://documentation.sas.com/?docsetId=grstatproc&docsetTarget=n0yadqm4wsfapgn1hzekfvu1ex0g.htm&docsetVersion=9.4&locale=en

- ★ file:///C:/Users/Joker/Downloads/Coronavirus-data-analysis-china.pdf

- ★ Yanchang Zhao yanchang@RDataMining.com http:// RDataMining.com

- ★ https://www.businessinsider.com/countries-on-lockdown-coronavirus-italy-2020-3

- ★ https://documentation.sas.com/

- ★ https://github.com/cmungun/population/blob/master/data/population.csv

THANK YOU

find our full code at :
https://github.com/cmungun/SasProject/