ANALIZA DANYCH EPIDEMIOLOGICZNYCH

Aleksandra Gomółka, Szymon Szrajer ● WBBIB UJ

CELE PRACY

- Analiza danych dotyczących wirusa COVID-19.
- Pozyskiwanie i wizualizacja danych za pomocą programów takich jak: LaTeX, Bash, Gnuplot, PyCharm.
- Porównanie uzyskanych, rzeczywistych wyników z danymi modelowymi.

POZYSKIWANIE DANYCH

W celu pozyskania danych skorzystaliśmy ze strony: https://www.worldometers.info.

		7,806,655	+79,562	430,111	+2,422	4,006,003	3,370,541	53,886	1,002	55.2			
	World												
1	USA	2,128,118	+11,196	117,088	+263	843,548	1,167,482	16,578	6,431	354	23,978,670	72,463	330,906,997
2	Brazil	831,064	+1,162	41,952	+51	427,610	361,502	8,318	3,911	197	1,476,057	6,947	212,483,982
3	Russia	520,129	+8,706	6,829	+114	274,641	238,659	2,300	3,564	47	14,574,117	99,870	145,931,552
4	India	321,076	+11,473	9,205	+315	162,026	149,845	8,944	233	7	5,507,182	3,993	1,379,307,780
5	UK	294,375	+1,425	41,662	+181	N/A	N/A	492	4,337	614	6,624,676	97,610	67,868,549
6	Spain	290,289		27,136		N/A	N/A	617	6,209	580	4,465,338	95,507	46,753,936
7	Italy	236,651	+346	34,301	+78	174,865	27,485	220	3,914	567	4,564,191	75,484	60,465,875
8	Peru	220,749		6,308		107,133	107,308	1,089	6,700	191	1,315,417	39,925	32,947,280
9	Germany	187,352	+101	8,864	+1	171,900	6,588	447	2,236	106	4,694,147	56,035	83,771,107
10	Iran	184,955	+2,410	8,730	+71	146,748	29,477	2,755	2,204	104	1,219,400	14,528	83,936,074
11	Turkey	175,218		4,778		149,102	21,338	664	2,079	57	2,541,903	30,156	84,292,030
12	Chile	167,355	+6,509	3,101	+231	137,296	26,958	1,656	8,758	162	819,999	42,914	19,107,883
13	France	156,287		29,374		72,572	54,341	879	2,395	450	1,384,633	21,215	65,266,664
14	Mexico	139,196	+5,222	16,448	+504	101,767	20,981	378	1,080	128	393,714	3,055	128,862,724
15	Pakistan	132,405	+6,472	2,551	+88	50,056	79,798	111	600	12	839,019	3,802	220,650,245
16	Saudi Arabia	123,308	+3,366	932	+39	82,548	39,828	1,843	3,545	27	1,087,021	31,250	34,784,389
17	Canada	98,368	+425	8,105	+56	59,333	30,930	1,929	2,607	215	2,072,096	54,926	37,725,348
18	Bangladesh	84,379	+2,856	1,139	+44	17,827	65,413	1	513	7	489,960	2,977	164,605,029
19	China	83,075	+11	4,634		78,367	74		58	3			1,439,323,776
20	Qatar	78,416	+1,828	70		55,252	23,094	232	27,928	25	286,830	102,155	2,807,805
21	South Africa	61,927		1,354		35,006	25,567	208	1,045	23	1,060,425	17,892	59,269,206
22	Belgium	59,918	+99	9,650	+4	16,547	33,721	88	5,171	833	1,010,059	87,170	11,587,177
23	Belarus	53,241	+721	303	+5	29,111	23,827	92	5,634	32	714,324	75,594	9,449,466

- Potrzebne dane pobieraliśmy do programu Excel a następnie modyfikowaliśmy.

- Korzystaliśmy głównie z komend Ctrl+H i usuń komórki.

- Gotowe pliki zapisywaliśmy w rozszerzeniu.csv.

#	Country, Other	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious, Critical	Tot Cases/ 1M pop	Deaths/ 1M pop	Total Tests	T
	World	7,810,449	83,356	430,133	2,444	4,007,563	3,372,953	53,906	1,002	55.2		
	North America	2,441,085	19,987	143,695	873	1,040,980	1,256,410	19,252				
	Europe	2,189,451	14,806	182,553	486	1,158,084	848,814	6,715				
1	USA	2,129,684	12,762	117,096	271	843,548	1,169,040	16,578	6,436	354	23,978,670	
	Asia	1,577,204	38,519	39,457	720	965,878	571,869	15,432				
	South America	1,364,004	8.56	58,169	325	727,498	578,337	11,977				
2	Brazil	831,064	1,162	41,952	51	427,610	361,502	8,318	3,911	197	1,476,057	
3	Russia	520,129	8,706	6,829	114	274,641	238,659	2,3	3,564	47	14,574,117	
4	India	321,406	11,803	9,205	315	162,320	149,881	8,944	233	7	5,507,182	
5	UK	294,375	1,425	41,662	181	N/A	N/A	492	4,337	614	6,624,676	
6	Spain	290,685	396	27,136		N/A	N/A	617	6,217	580	4,465,338	
7	Italy	236,651	346	34,301	78	174,865	27,485	220	3,914	567	4,564,191	
	Africa	229,071	1,472	6,12	40	105,872	117,079	524				
8	Peru	220,749		6,308		107,133	107,308	1,089	6,7	191	1,315,417	
9	Germany	187,356	105	8,864	1	171,900	6,592	447	2,237	106	4,694,147	
10	Iran	184,955	2,41	8,73	71	146,748	29,477	2,755	2,204	104	1,219,400	
11	Turkey	176,677	1,459	4,792	14	150,087	21,798	684	2,096	57	2,586,995	
12	Chile	167,355	6,509	3,101	231	137,296	26,958	1,656	8,758	162	819,999	
13	France	156,287		29,374		72,572	54,341	879	2,395	450	1,384,633	
14	Mexico	139,196	5,222	16,448	504	101,767	20,981	378	1,08	128	393,714	
15	Pakistan	132,405	6,472	2,551	88	50,056	79,798	111	600	12	839,019	
16	Saudi Arabia	123,308	3,366	932	39	82,548	39,828	1,843	3,545	27	1,087,021	
17	Canada	98,368	425	8,105	56	59,333	30,930	1,929	2,607	215	2,072,096	
18	Bangladesh	84,379	2,856	1,139	44	17,828	65,412	1	513	7	489,960	
19	China	83,075	11	4,634		78,367	74		58	3		
20	Qatar	78,416	1,828	70		55,252	23,094	232	27,928	25	286,830	
21	South Africa	61,927		1,354		35,006	25,567	208	1,045	23	1,060,425	
22	Belgium	59,918	99	9,65	4	16,547	33,721	88	5,171	833	1,010,059	
23	Belarus	53,241	721	303	5	29,111	23,827	92	5,634	32	714,324	
24	Sweden	50,931	180	4,874	20	N/A	N/A	272	5,045	483	325,000	
25	Netherlands	48,640	179	6,057	4	N/A	N/A	77	2,839	354	455,067	
26	Colombia	46,858		1,545		18,715	26,598	335	921	30	470,351	
27	Ecuador	45,778		3,828		22,679	19,271	219	2,597	217	128,175	
28	UAE	41,990	491	288	1	26,761	14,941	1	4,248	29	2,626,000	
29	Egypt	41,303		1,422		11,108	28,773	41	404	14	135,000	
30	Singapore	40,197	347	25		28,808	11,364	1	6,874	4	488,695	
31	Indonesia	37,420	1,014	2,091	43	13,776	21,553		137	8	495,527	
32	Portugal	36,463	283	1,512	7	22,438	12,513	77	3,575	148	975,737	
33	Kuwait	35,466	514	289	4	25,882	9,295	176	8,311	68	332,288	
34	Switzerland	31,094	31	1,938		28,800	356	15	3,594	224	453,745	

#	Country	Total Cases	New Cases	Total Deaths	New Deaths	Total Recovered	Active Cases	Serious Critical	Tot Cases/ 1M	Deaths/ 1M	Total Tests	Tests/ 1M pop	Population
1	USA	1901783	20578	109142	1083	68867	1103971	16939	5748	330	19096671	57717	3,31E+08
2	Brazil	58398	27312	32547	1269	266132	285301	8318	2749	153	930013	4378	2,12E+08
3	Russia	432277	8536	5215	178	195957	231105	23	2962	36	11426045	78298	1,46E+08
4	Spain	287406	394	27128	1	N/A	N/A	617	6147	580	4063843	86921	46753542
5	UK	279856	1871	39728	359	N/A	N/A	1559	4124	585	4786219	7053	67860771
6	Italy	233836	321	33601	71	160938	39297	353	3867	556	3999591	66144	60467811
7	India	216824	9633	6088	259	104071	106665	8944	157	4	4103233	2975	1,38E+09
8	Germany	184425	334	8699	25	1673	8426	632	2202	104	4348880	51917	83765266
9	Peru	178914	403	4894	127	72319	101701	964	5432	149	1115615	33871	32937238
10	Turkey	166422	867	4609	24	130852	30961	612	1975	55	2155349	25576	84272204
11	Iran	160696	3134	8012	70	125206	27478	2557	1915	95	997009	11882	83912575
12	France	151677	352	29021	81	69455	53201	121	2324	445	1384633	21216	65263516
13	Chile	113628	4942	1275	87	86173	2618	1446	5948	67	628318	32889	19104300
14	Mexico	97326	3891	10637	470	70308	16381	378	755	83	293078	2275	1,29E+08
15	Canada	93085	675	7498	103	51048	34539	1721	2468	199	1755926	46554	37718122
16	Saudi_Arab	91182	2171	579	30	68159	22444	1321	2622	17	870963	25047	34772530
17	China	83021		4634		78314	73	3	58	3			1,44E+09
18	Pakistan	80463	4065	1688	67	28923	49852	111	365	8	595344	2699	2,21E+08
19	Qatar	6216	1901	45	2	37542	24573	237	22138	16	236437	84207	2807805
20	Belgium	58685	70	9522	17	15959	33204	172	5065	822	884386	76332	11586077
21	Bangladesh	5514	2695	746	37	1159	42804	1	335	5	345583	21	1,65E+08
22	Netherland	46733	86	5977	10	N/A	N/A	116	2728	349	368529	21511	17132248
23	Belarus	45116	861	248	5	20171	24697	92	4774	26	573699	60712	9449534
24	Ecuador	40966	552	3486	48	20019	17461	252	2325	198	119375	6774	17622682
25	Sweden	40803	2214	4542	74	N/A	N/A	308	4042	450	2755	27291	10094775
26	South_Afri	37525	1713	792	37	19682	17051	208	633	13	785979	13265	59252862
27	Singapore	36405	569	24		23582	12799	5	6226	4	408495	69864	5847022

```
#;Country, Other;Total Cases;New Cases;Total Deaths;New Deaths;Total Recovered;Active Cases;Serious, Critical;Tot Cases/ 1M pop;Deaths/ 1M pop;Total Tests 1M pop;Population
1;USA;2,007,449;18,905;112,469;373;761,708;1,133,272;16,923;6,067;340;21,291,677;64,348;330,885,824
2;Brazil;691,962;18,375;36,499;542;302,084;353,379;8,318;3,257;172;999,836;4,706;212,467,494
3;Russia;467,673;8,984;5,859;134;226,731;235,083;2,3;3,205;40;12,721,549;87,175;145,930,871
4;Spain;288,630;240;27,136;1;N/A;N/A;617;6,173;580;4,063,843;86,92;46,753,739
5;UK;286,194;1,326;40,542;77;N/A;N/A;604;4,217;597;5,581,073;82,238;67,864,628
6;India;257,486;10,864;7,207;261;123,848;126,431;8,944;187;5;4,666,386;3,383;1,379,159,619
7;Italy;234,998;197;33,899;53;165,837;35,262;287;3,886;561;4,236,535;70,064;60,466,843
8;Peru;196,515;4,757;5,465;164;86,219;104,831;1,062;5,965;166;1,191,956;36,183;32,942,259
9;Germany;185,869;173;8,776;7;169,100;7,993;568;2,219;105;4,348,880;51,916;83,768,186
10;Iran;171,789;2,364;8,281;72;134,349;29,159;2,596;2,047;99;1,084,857;12,927;83,924,324
11;Turkey;170,132;914;4,692;23;137,969;27,471;613;2,019;56;2,338,593;27,747;84,282,117
12;France;153,977;343;29,155;13;70,842;53,980;1,053;2,359;447;1,384,633;21,216;65,265,090
13;Chile;134,150;6,405;2,19;649;95,631;36,329;1,558;7,021;115;708,773;37,097;19,106,092
14;Mexico;113,619;3,593;13,511;341;81,544;18,564;378;882;105;332,326;2,579;128,847,928
15;Saudi Arabia;101,914;3,045;712;36;72,817;28,385;1,564;2,93;20;958,237;27,553;34,778,459
16;Pakistan;98,943;4,96;2,002;67;33,465;63,476;111;449;9;683,608;3,099;220,603,292
17;Canada;95,699;642;7,8;27;54,233;33,666;1,816;2,537;207;1,896,822;50,285;37,721,735
18;China;83,036;6;4,634;78,332;70;1;58;3;;;1,439,323,776
19;Qatar;68,790;1,595;54;3;44,338;24,398;245;24,5;19;255,533;91,008;2,807,805
20;Bangladesh;65,769;2,743;888;42;13,903;50,978;1;400;5;397,987;2,418;164,587,111
21;Belgium;59,226;154;9,595;15;16,291;33,340;111;5,112;828;935,670;80,754;11,586,627
22;Belarus;48,630;879;269;6;23,647;24,714;92;5,146;28;622,313;65,857;9,449,500
23;South Africa;48,285;2,312;998;46;24,364;22,923;208;815;17;920,064;15,526;59,261,034
24;Netherlands;47,574;239;6,013;2;N/A;N/A;97;2,777;351;388,113;22,653;17,132,661
25;Sweden;45,065;344;4,659;3;N/A;N/A;1998;4,464;461;275,500;27,289;10,095,462
26;Ecuador;43,120;392;3,621;13;21,020;18,479;222;2,446;205;127,414;7,229;17,625,612
27;Colombia;39,236;1,209;1,259;54;15,322;22,655;335;772;25;410,719;8,077;50,848,887
28;UAE;38,808;540;276;1;21,806;16,726;1;3,927;28;2,500,000;252,964;9,882,820
29;Singapore;37,910;383;25;;24,886;12,999;3;6,483;4;408,495;69,858;5,847,525
30;Portugal;34,693;342;1,479;5;20,995;12,219;58;3,402;145;873,998;85,7;10,198,365
31;Egypt;34,079;1,467;1,237;39;8,961;23,881;41;333;12;135,000;1,321;102,204,915
32;Kuwait;31,848;717;264;10;20,205;11,379;196;7,465;62;315,285;73,898;4,266,472
33;Indonesia;31,186;672;1,851;50;10,498;18,837;;114;7;405,992;1,485;273,342,397
34;Switzerland;30,965;9;1,921;28,700;344;28;3,579;222;423,468;48,952;8,650,765
35;Ukraine;26,999;485;788;11;12,054;14,157;304;617;18;424,046;9,693;43,747,990
36;Poland;26,561;575;1,157;4;12,855;12,549;160;702;31;1,056,396;27,911;37,848,958
37;Ireland;25,201;18;1,679;1;22,698;824;36;5,107;340;348,416;70,611;4,934,313
38;Argentina;22,794;774;664;16;6,909;15,221;274;505;15;193,923;4,293;45,170,095
39;Philippines;21,895;555;1,003;9;4,530;16,362;82;200;9;429,332;3,921;109,487,582
40;Romania;20,479;189;1,333;11;14,638;4,508;144;1,064;69;503,200;26,148;19,244,598
```

- Dane sortowaliśmy wykorzystując program Bash.
- W tym celu stworzyliśmy skrypt w pliku tekstowym, który sortował dane zebrane na przestrzeni ośmiu dni.
- Wynikiem sortowania 8 plików było 20 plików docelowych.

```bash
#!/bin/bash
#top 10 testow/1M populacji - ostatni dzien

cut -d';' -f2,13 10.csv > 10testy1.csv
(head -n 2 10testy1.csv && tail -n +3 10testy1.csv |sort -t';' -k2 -r -n) > 10testy2.csv
head -12 10testy2.csv > 10testy_Wynik.dat
rm 10testy1.csv 10testy2.csv


#top 10 testow - ostatni minus pierwszy

cut -d';' -f2,12 3.csv > 3tottesty1.csv
(head -n 2 3tottesty1.csv && tail -n +3 3tottesty1.csv |sort -t';' -k2 -r -n) > 3tottesty2.csv
head -12 3tottesty2.csv > 3tottesty_Wynik.dat
rm 3tottesty1.csv 3tottesty2.csv

cut -d';' -f2,12 10.csv > 10tottesty1.csv
(head -n 2 10tottesty1.csv && tail -n +3 10tottesty1.csv |sort -t';' -k2 -r -n) > 10tottesty2.csv
head -12 10tottesty2.csv > 10tottesty_Wynik.dat
rm 10tottesty1.csv 10tottesty2.csv

#top 10 zgonow/1M populacji - ostatni dzien

cut -d';' -f2,11 10.csv > 10zgony1.csv
(head -n 2 10zgony1.csv && tail -n +3 10zgony1.csv |sort -t';' -k2 -r -n) > 10zgony2.csv
head -12 10zgony2.csv > 10zgony_Wynik.dat
rm 10zgony1.csv 10zgony2.csv

#top 10 zgonow ogolem - ostatni dzien

cut -d';' -f2,5 10.csv > 10totzgony1.csv
(head -n 2 10totzgony1.csv && tail -n +3 10totzgony1.csv |sort -t';' -k2 -r -n) > 10totzgony2.csv
head -12 10totzgony2.csv > 10totzgony_Wynik.dat
rm 10totzgony1.csv 10totzgony2.csv

#top 10 zgonow ogolem w osiem dni - ostatni minus pierwszy

cut -d';' -f2,5 3.csv > 3totzgony1.csv
(head -n 2 3totzgony1.csv && tail -n +3 3totzgony1.csv |sort -t';' -k2 -r -n) > 3totzgony2.csv
head -12 3totzgony2.csv > 3totzgony_Wynik.dat
rm 3totzgony1.csv 3totzgony2.csv
```

KREŚLENIE WYKRESÓW

- Jeżeli było to konieczne dane reorganizowaliśmy w programie Excel.
- Na podstawie uzyskanych plików w formacie .dat wykreśliliśmy wykresy.
- Wykresy stworzone w programie Gnuplot przedstawiają dane dla państw, które stoją na czele w danych kategoriach.

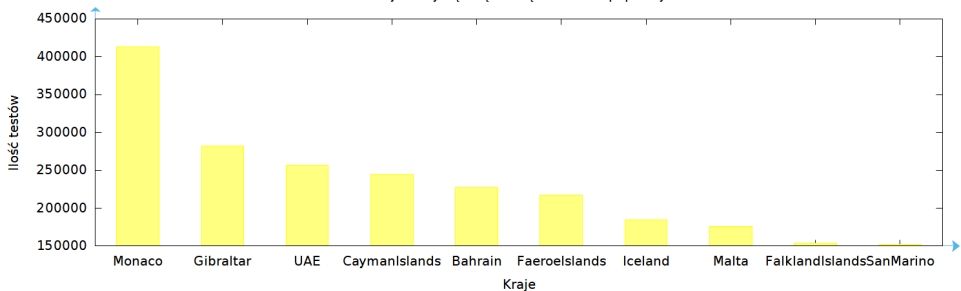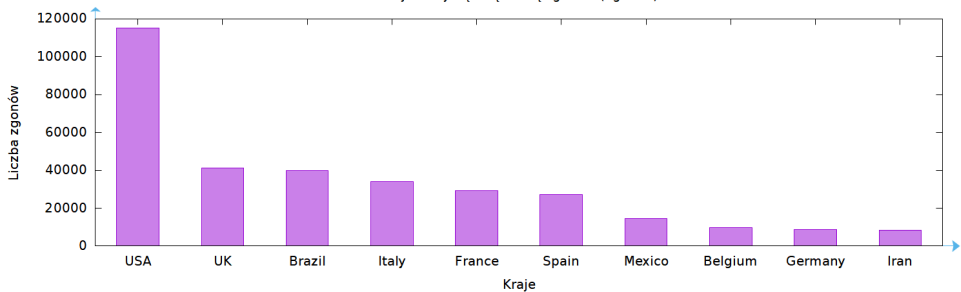Kraje z największą liczbą przypadków zakażeń (ogółem)
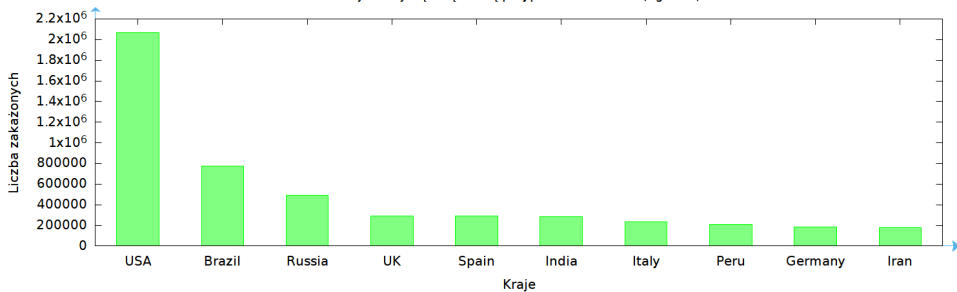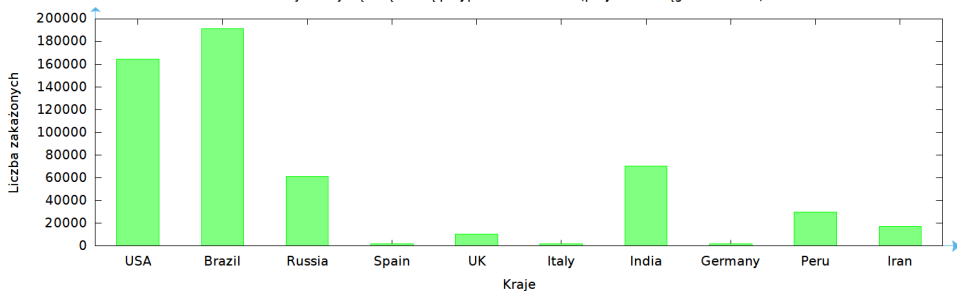
Kraje z największą liczbą zgonów/1M populacji
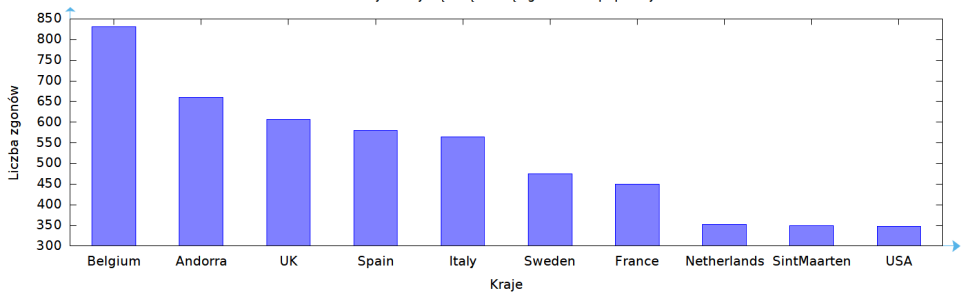
Kraje z największą ilością testów/1M populacji
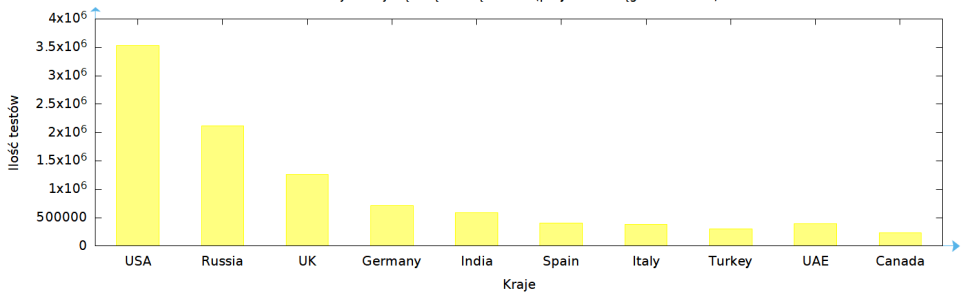
Kraje z największą liczbą zgonów (ogółem)

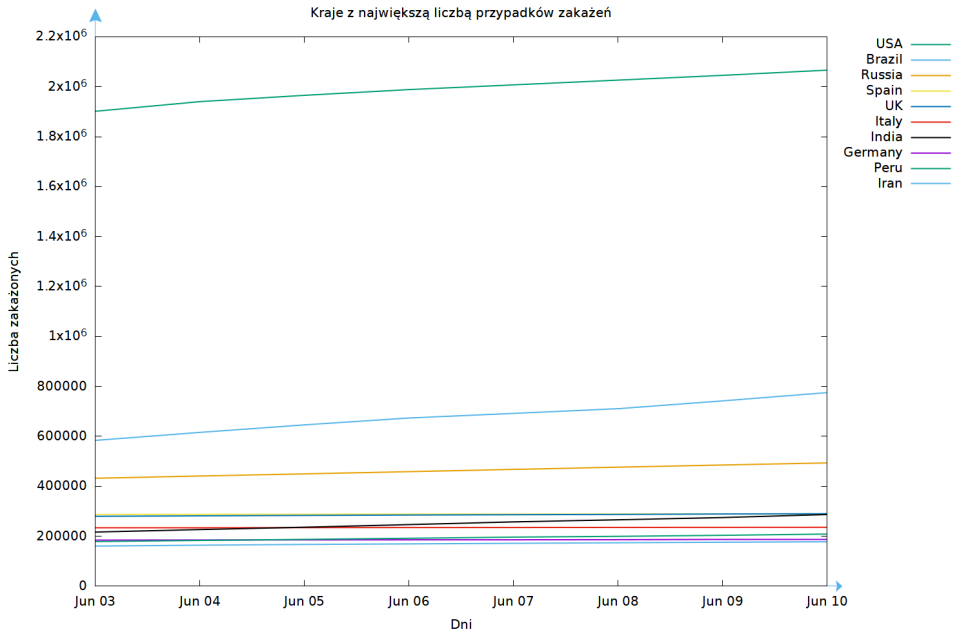Kraje z największą liczbą przypadków zakażeń (przyrost w ciągu ośmiu dni)
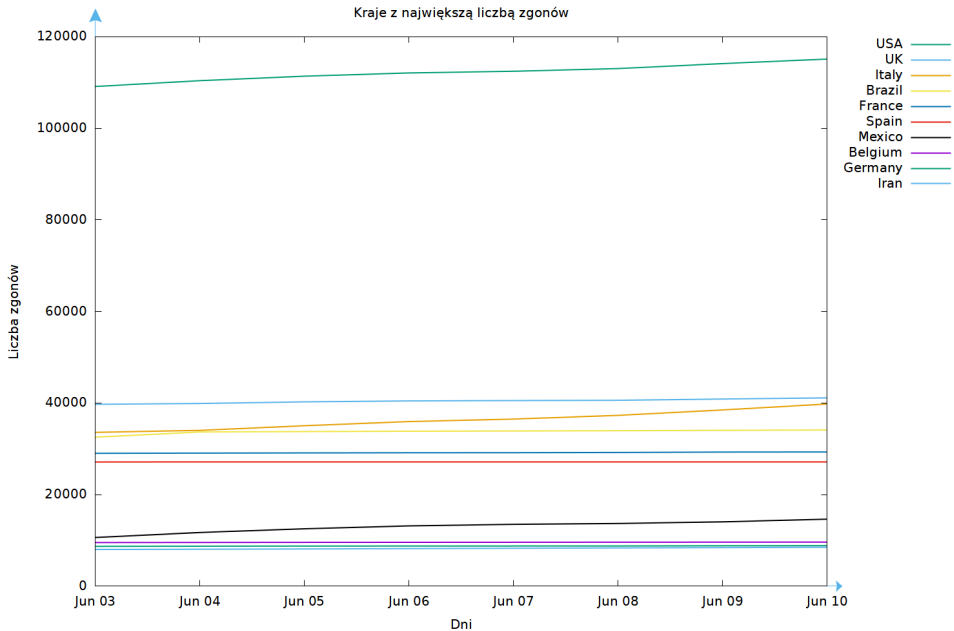
Kraje z największą liczbą zgonów (przyrost w ciągu ośmiu dni)
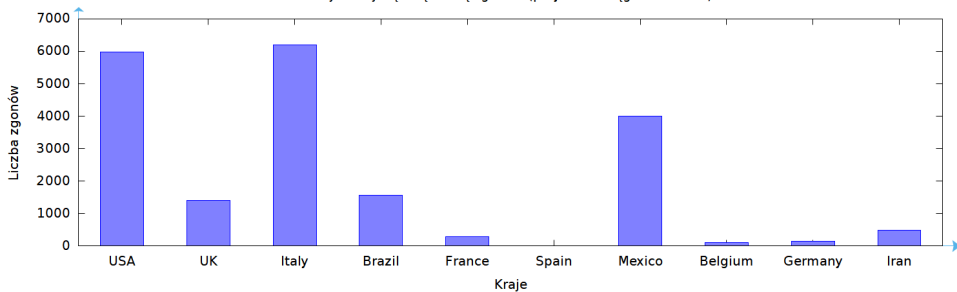
Kraje z największą ilością testów (przyrost w ciągu ośmiu dni)

Kraje z największą liczbą przypadków zakażeń

Kraje z największą liczbą zgonów

Legend:
- USA
- UK
- Italy
- Brazil
- France
- Spain
- Mexico
- Belgium
- Germany
- Iran

Y-axis: Liczba zgonów

X-axis: Dni (Jun 03, Jun 04, Jun 05, Jun 06, Jun 07, Jun 08, Jun 09, Jun 10)

MODEL SIR

Korzystając z języka programowania Python stworzyliśmy interpretację modelu SIR(Suspectible, Infected, Recovered) aby sprawdzić jak dane rzeczywiste mają się do tych, które przewidują modele.

```python
import numpy as np
from scipy.integrate import odeint
import matplotlib.pyplot as plt


N = 37848924 #populacja
I = 24687 # osoby zarażone na w pierwszym dniu badania
beta = 1.4 # liczba kontaktów dziennie
K = 14 # średni czas trwania infekcji wyrażony w dniach
time = 8 # liczba dni oznacza czas brany pod uwagę podczas badań


R = 0 # początkowa liczba osób, które wyzdrowiały
S = N - I - R # liczba osób podatnych na zakażenie
k = 1/K # współczynnik wyzdrowień
t = np.linspace(0,time,time) # siatka punktów czasowych, oś x

# funkcja generująca równania różniczkowe
def derivatives(y, t, N, beta, k):
    S, I, R = y
    dSdt = -beta * S * I / N
    dIdt = beta * S * I / N - k * I
    dRdt = k * I
    return dSdt, dIdt, dRdt
```

```python
# wykreślenie danych na trzech osobnych krzywych dla S(t), I(t) i R(t)
# kolor tła
backg = plt.figure(facecolor='w')
# opis osi
axis = backg.add_subplot(111, facecolor='#E5E8f9', axisbelow=True)
axis.set_xlabel('Time [days]')
axis.set_ylabel('Population (in 1 000 000s)')
# wykresy SIR, czas, wartości, kolor, szerokość, nazwa
axis.plot(t, S/1000000, '#46F0FF', lw=2, label='Susceptible')
axis.plot(t, I/1000000, '#FF7996', lw=2, label='Infected')
axis.plot(t, R/1000000, '#8EFF3B', lw=2, label='Recovered with immunity')
# ograniczenie osi y
axis.set_ylim(0, 38)
# "linie" - widoczne, które, kolor, szerokość, styl
axis.grid(b=True, which='both', c='w', lw=1, ls='-')
#legenda
legend = axis.legend()
plt.show()
print(I)
```
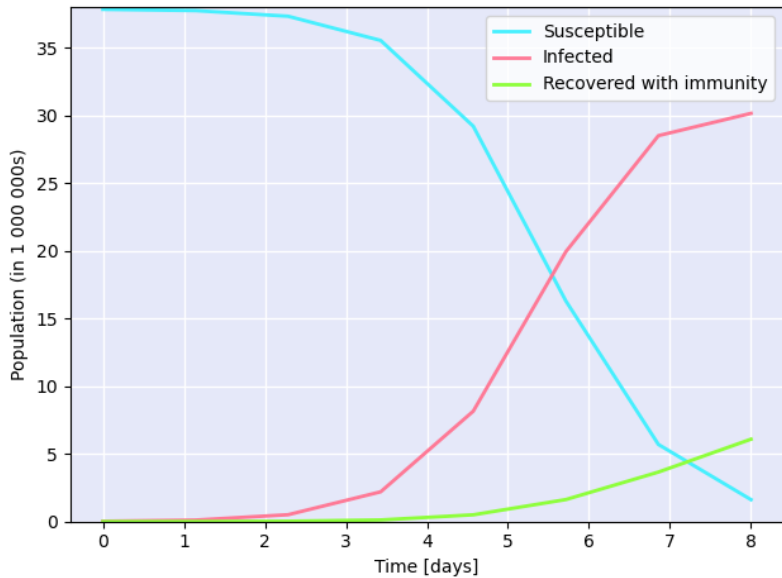
PREZENTACJA

Prezentację wykonaliśmy w programie LaTeX na podstawie bazowej klasy "fancyslides", którą później modyfikowaliśmy dla potrzeb określonego slajdu.

BIBLIOGRAFIA:
https://www.worldometers.info/coronavirus/
Paweł Lupkowski-fancyslides
https://www.pinterest.co.uk/pin/489836896963829968/?

DZIĘKUJEMY ZA UWAGĘ