

AGH

Symulacja rozprzestrzeniania patogenu

*Auguścik Szymon,
Błasiok Jan,
Maliszewski Piotr*

2020
Czerwiec

Spis treści

1	Opis problemu	2
2	Modele zjawiska	2
2.1	Model SIR	2
2.2	Model SEIR	3
3	Propozycja symulacji	4
3.1	Model SEIRD	4
3.2	Ogólny zarys rozwiązania	5
4	Implementacja	5
4.1	C++	5
4.1.1	Klasa Region	5
4.1.2	Klasa Simulation	5
4.2	Python	6
4.2.1	DataFrame	6
4.2.2	Window	6
4.3	Szczegółowa specyfikacja rozwiązania	6
5	Wyniki symulacji dla wybranych par "choroba-państwo początkowe"	7
5.1	COVID-19 w Chinach	7
5.2	Odra w Polsce	10
5.3	Grypa hiszpanka w USA	12
5.4	AIDS na Sri Lance	13
6	Wnioski	14
6.1	Różnice między poszczególnymi chorobami	14
6.2	Różnice między wykresami dla państw i dla świata	15
6.3	Różnice pomiędzy symulacją a rzeczywistością	15
6.4	Pozostałe	16
7	Propozycje rozwijania w przyszłości	16
7.1	Uwzględnienie demografii	16
7.2	Gotowe scenariusze	16
7.3	Reakcje państw na epidemię	16
7.4	Optymalizacja generowania map	17

1 Opis problemu

Nasz projekt zajmuje się symulacją rozprzestrzeniania patogenu w skali światowej. Staramy się w jak najlepszy sposób odwzorować rozprzestrzenianie się różnych chorób, które zawarliśmy w naszej bazie (m. in. COVID-19, AIDS, malaria).

2 Modele zjawiska

2.1 Model SIR

Model SIR został po raz pierwszy użyty przez *Kermacka* i *McKendricka* w 1927 roku, a następnie został zastosowany do różnych chorób, zwłaszcza do chorób wieku dziecięcego przenoszonych drogą powietrzną z dożywotnią odpornością po powrocie do zdrowia. takich jak odra, świnka, różyczka i krztusiec. S, I i R reprezentują liczbę osób podatnych (*Susceptible*), zarażonych (*Infectious*) i usuniętych z symulacji (*Removed*), a $N = S + I + R$ to całkowita populacja.

W modelu pojawiają się współczynniki charakteryzujące epidemię. Od nich zależy jej przebieg. Dla istniejących chorób współczynniki te zostały wyznaczone empirycznie. Oznaczają one:

- β – współczynnik zarażenia się chorobą
- γ – współczynnik zakończenia choroby dla danej jednostki (zgon lub przechorowanie)

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

2.2 Model SEIR

Wiele chorób ma fazę utajnią, podczas której jednostka jest zakażona, ale jeszcze nie zaraża. To opóźnienie między nabyciem zakażenia a stanem zakaźnym można włączyć do modelu SIR poprzez dodanie utajnionej populacji, E (*Exposed*), i umożliwienie zakażonym (ale jeszcze niezakażającym) osobnikom przejścia z S do E i z E do I.

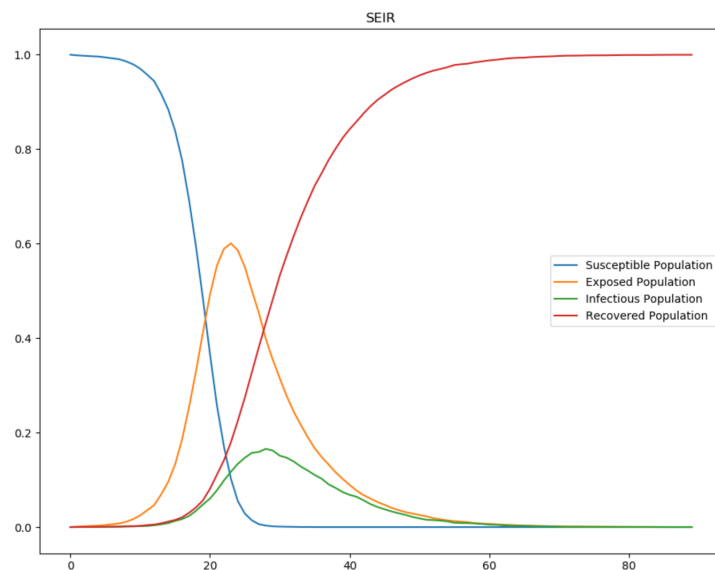
$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

Przez wyróżnienie kolejnej grupy wprowadzony musi być nowy współczynnik

- σ – Współczynnik oznaczający przechodzenie ze stanu zarażonego do stanu zarażającego

Na poniższym wykresie, przedstawione są zależności poszczególnych grup (SEIR) od czasu.

- Niebieski – podatni na zachorowanie
- Pomarańczowy – narażeni na zachorowanie, czyli osoby, które jeszcze nie zarażają, ale są zarażone
- Zielony – osoby zarażone i rozprzestrzeniające chorobę
- Czerwony – osoby usunięte z symulacji, czyli albo wyzdrowiały ze stałą odpornością, albo martwe w wyniku choroby



3 Propozycja symulacji

3.1 Model SEIRD

Z modeli, które braliśmy pod uwagę najciekawszym wydawał się model SEIR, ale postanowiliśmy go zmodyfikować w taki sposób, że z grupy R (*Removed*) wyodrębniliśmy grupy R (*Recovered*) oraz D (*Dead*).

Wyróżnienie tych grup niesie za sobą rozdzielenie współczynnika γ na dwie wartości, oznaczone przez nas γ_1 oraz γ_2

- γ_1 – współczynnik wyzdrowienia
- γ_2 – współczynnik umierania

$$\frac{\partial S}{\partial t} = (\Lambda - \mu)S - \beta \frac{I}{N}S$$

$$\frac{\partial E}{\partial t} = \beta \frac{I}{N} \cdot S - (\mu - \alpha)E$$

$$\frac{\partial I}{\partial t} = \alpha E - (\gamma_1 + \gamma_2)I$$

$$\frac{\partial R}{\partial t} = \gamma_1 I - \mu R$$

$$\frac{\partial D}{\partial t} = \gamma_2 I$$

3.2 Ogólny zarys rozwiązania

Po dyskusji na temat sposobu symulacji stwierdziliśmy, że najlepszym sposobem będzie symulacja każdego kraju osobno każdego dnia. Musieliśmy najpierw uzyskać informacje o krajach na świecie i chorobach, aby móc wykorzystać te dane w naszym modelu matematycznym. Zdecydowaliśmy się, aby samą symulację przeprowadzić korzystając z języka C++, a część wizualną do interakcji z użytkownikiem zostawiliśmy językowi Python.

4 Implementacja

4.1 C++

4.1.1 Klasa Region

Klasa Region obsługuje pojedynczy krok symulacji, tj. oblicza na podstawie równań modelu zmiany zachodzące pomiędzy grupami SEIRD. Ze względu na to, że współczynniki i przyrosty SEIRD są liczbami niecałkowitymi, liczbę osób poszczególnych grup też traktujemy jako taką. Zaokrąglanie powodowałoby problemy w działaniu. Każdy region przechowuje swoją historię, która następnie jest zapisywana w folderze bieżącej epidemii do pliku o nazwie takiej samej jak region.

4.1.2 Klasa Simulation

W tej klasie przeprowadzana obsługiwane są relacje między regionami, jak i uruchamianie jednego kroku symulacyjnego dla zarażonych państw. Wprowadziliśmy zabezpieczenie przed zbyt długim działaniem przez ustawienie górnego limitu dni. Dane ze-

brane z poszczególnych państw są przechowywane w buforze daną liczbę dni(kroków), po wypełnieniu tej tablicy symulacja jest przerywana na moment zapisywania danych poszczególnych państw do osobnych plików jest to wykonywane metodą wielowątkową, aby proces ten był jak najkrótszy. Przed rozpoczęciem głównej pętli symulacji ustawiany jest *pacjent zero* w kraju początkowym. Jeśli kraj nie został wybrany, losowany jest z pośród istniejących krajów.

4.2 Python

4.2.1 DataFrame

Klasa DataFrame jest odpowiedzialna za przechowywanie w części programu napisanej w Pythonie danych załadowanych z plików wygenerowanych podczas symulacji w C++. Jej istotnymi elementami są listy dotyczące poszczególnych grup, a także dni startu i zakończenia pandemii w danym kraju. Istotą tej klasy jest jednokrotne wczytanie danych z plików, by później swobodnie operować nimi w programie np. podczas tworzenia map lub generowania wykresu dla całego świata.

4.2.2 Window

Klasa Window jest odpowiedzialna za bezpośrednią interakcję z użytkownikiem. Pozwala mu na wybór choroby oraz kraju początkowego, na podstawie których symulacja jest przeprowadzana. Jeżeli jednak takie dane zostały już wprowadzone, to program wykorzystuje dane, które uprzednio zostały wyliczone. Klasa również pozwala na wprowadzenie własnych współczynników charakteryzujących epidemię. W obu przypadkach odpowiednie argumenty przekazywane są do programu.

4.3 Szczegółowa specyfikacja rozwiązania

Wykorzystujemy łatwość dzięki językowi Python możliwość projektowania w dwóch językach jednocześnie - Python w bardzo prosty sposób potrafi wykorzystywać programy napisane w innych językach, dzięki temu część odpowiedzialną za symulację mogliśmy napisać w C++, a resztę w Pythonie.

Zadaniem programu napisanego w języku C++ jest wygenerowanie plików opisujących każdy kraj pod względem przebytej pandemii. W wynikowym pliku zapisywane są dni epidemii jako wiersze zawierające dzień pandemii (globalny), a także liczbę osób z każdej grupy z modelu SEIRD. Przykładowy fragment pliku wygląda jak poniżej:

```

471;3.71724e+07;0.4341;0.9871;0.007;0.0059
472;3.71724e+07;0.741852;0.98965;0.165918;0.0112519
473;3.71724e+07;0.987721;0.973354;0.245857;0.0166188
474;3.71724e+07;1.21271;1.88847;0.329739;0.0223616
475;3.71724e+07;1.4392;1.22264;0.42374;0.0287364
476;3.71724e+07;1.68211;1.3969;0.53811;0.03595
477;3.71724e+07;1.95288;1.60355;0.65164;0.0441917
478;3.71724e+07;2.25776;1.84499;0.791149;0.0536526
479;3.71724e+07;2.60712;2.12515;0.951663;0.0645381
480;3.71724e+07;3.00822;2.44914;1.13655;0.0778764
481;3.71724e+07;3.46975;2.82326;1.34963;0.0915264
482;3.71724e+07;4.00138;3.25493;1.59525;0.108184
483;3.71724e+07;4.61407;3.75283;1.87843;0.127388
484;3.71724e+07;5.32836;4.327;2.20493;0.149529
485;3.71724e+07;6.13464;4.9891;2.58137;0.175059
486;3.71724e+07;7.07348;5.75254;3.01543;0.204494
487;3.71724e+07;8.15596;6.63282;3.5159;0.238434
488;3.71724e+07;9.40407;7.64782;4.09295;0.277568
489;3.71724e+07;10.8432;8.81815;4.75831;0.32269
490;3.71724e+07;12.5825;10.1676;5.52549;0.374717
491;3.71724e+07;14.4157;11.7235;6.41807;0.434786
492;3.71723e+07;16.6218;13.5175;7.43882;0.503875
493;3.71723e+07;19.1654;15.5861;8.68684;0.583628
494;3.71723e+07;22.0982;17.9712;9.96284;0.675586
495;3.71723e+07;25.4799;20.7214;11.5255;0.781617
496;3.71723e+07;29.379;23.8923;13.3283;0.903873
497;3.71723e+07;33.8749;27.5485;15.4869;1.04484
498;3.71723e+07;39.0687;31.7643;17.8836;1.20737

```

Jak widać, model wymaga od nas przyjęcia do obliczeń liczb niecałkowitych - jest to warunek konieczny, by symulacja dobiegła końca, tj. liczba osób zarażonych bezobjawowo oraz chorych była zbliżona do 0 (pomijalnie mała).

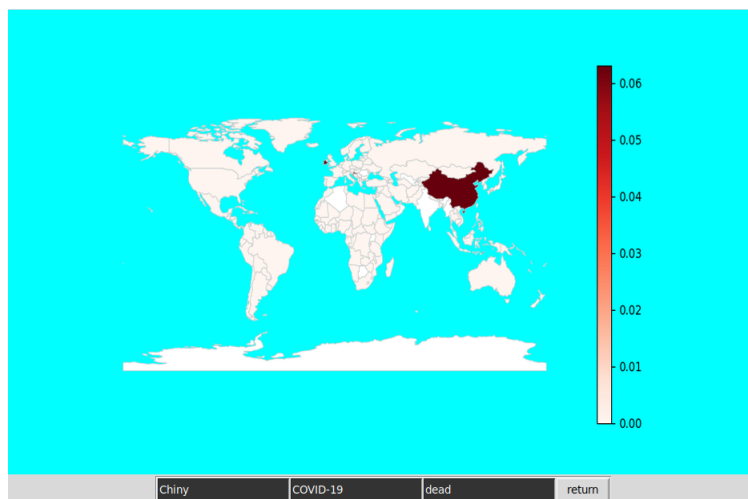
Dane z symulacji zapisywane są w plikach, każdy plik odpowiada konkretnemu państwu w którym panowała epidemia. Program symulując wpisuje dane do bufora (dynamiczna tablica danych). Po zapełnieniu tablicy symulacja jest przerywana aby dane z bufora zostały zapisane do plików a później wyzerowane (aby móc od nowa zapisywać dane) - jest to prowadzone metodą wielowątkową aby przyspieszyć ten proces.

5 Wyniki symulacji dla wybranych par "choroba-państwo początkowe"

5.1 COVID-19 w Chinach

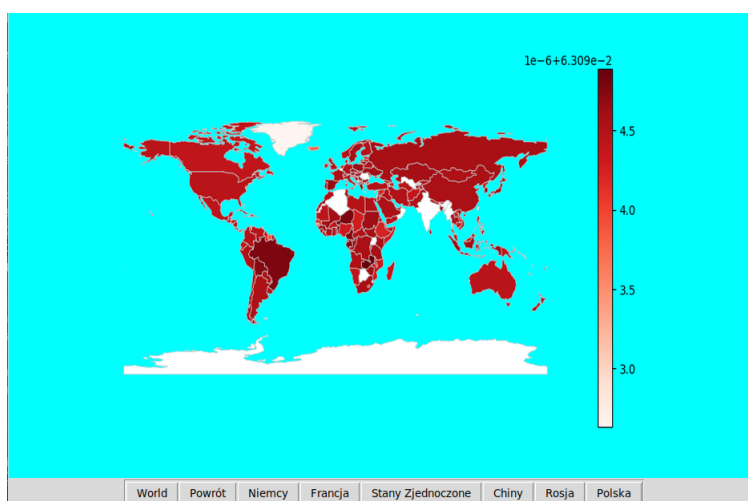
COVID-19 jest chorobą stosunkowo nieznaną - w trakcie pisania dokumentacji świat nadal zmagają się z pandemią koronawirusa wywołującego w/w chorobę. Niemniej jednak, bazując na dostępnych danych, można stwierdzić, że COVID-19 jest chorobą o stosunkowo niskiej śmiertelności, lecz wysokiej zaraźliwości. Chiny natomiast to państwo o największej populacji na świecie, ponadto są one ważnym punktem na szlakach handlowych, transportowych i turystycznych, co dodatkowo generuje spory ruch osobowy między Państwem Środka a pozostałymi państwami globu.

Dla pierwszych kilkunastu dni mapa odsetka osób zmarłych na COVID-19 w poszczególnych krajach wygląda następująco:



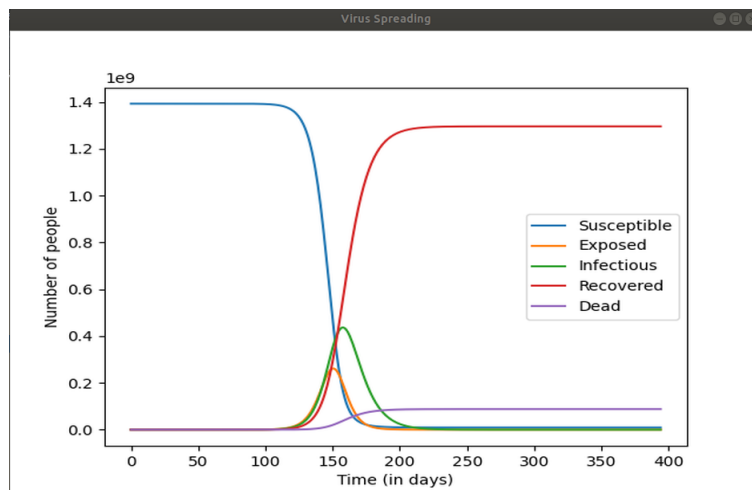
Wirus przedostał się do Irlandii.

Natomiast na koniec mapa prezentuje się następująco:

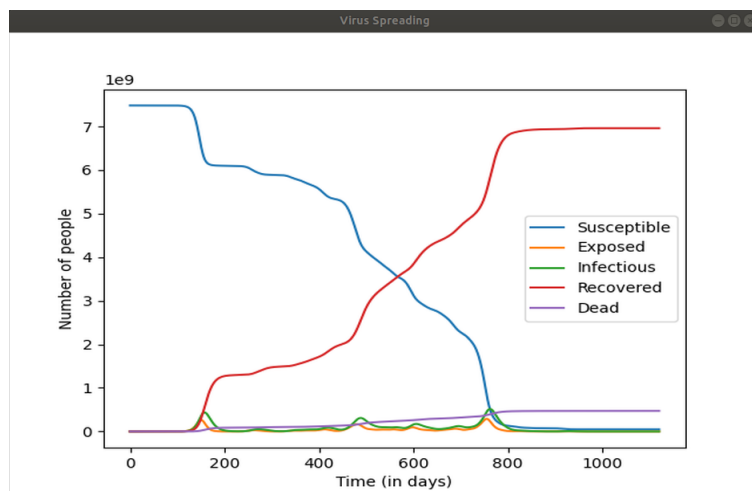


Jak widać, pandemia objęła swoim zasięgiem cały świat, wykazując podobną śmiertelność w różnych regionach globu.

Wykres przebiegu epidemii dla Chin(rok trwania epidemii) prezentuje się w następujący sposób:



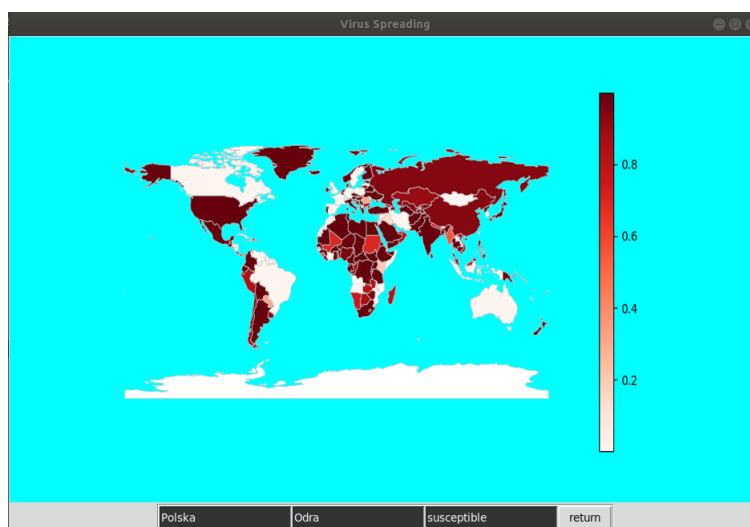
Natomiast wykres przedstawiający rozwój pandemii dla całego świata wygląda w taki sposób:



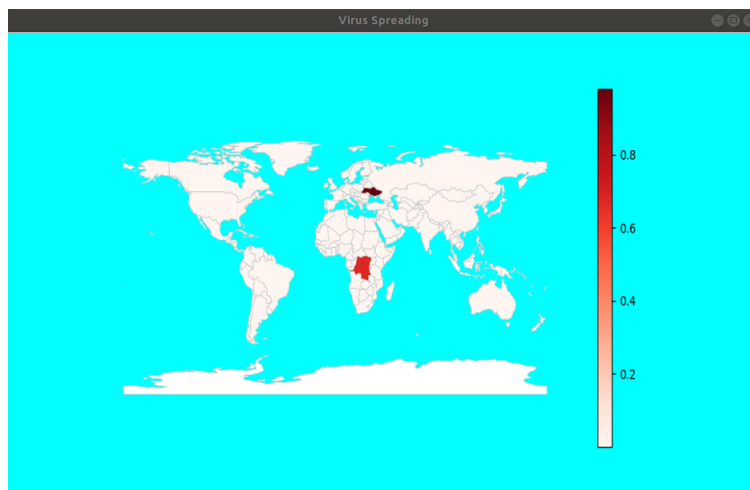
5.2 Odra w Polsce

Odra jest chorobą o stosunkowo małej śmiertelności. Mimo to wg danych z 2018 roku ta choroba zabrała życie ponad 140 tys. ludzi, a zapadło na nią ok. 10 mln osób na całym świecie. Odra jest chorobą wysoce zakaźną, która łatwo rozprzestrzenia się pomiędzy ludźmi. Polska, jako średniej wielkości kraj położony w Europie Środkowo-Wschodniej, stanowi istotny węzeł komunikacyjny pomiędzy Starym Kontynentem a Azją, co może za sobą pociągać spory ruch międzypaństwowy w regionie.

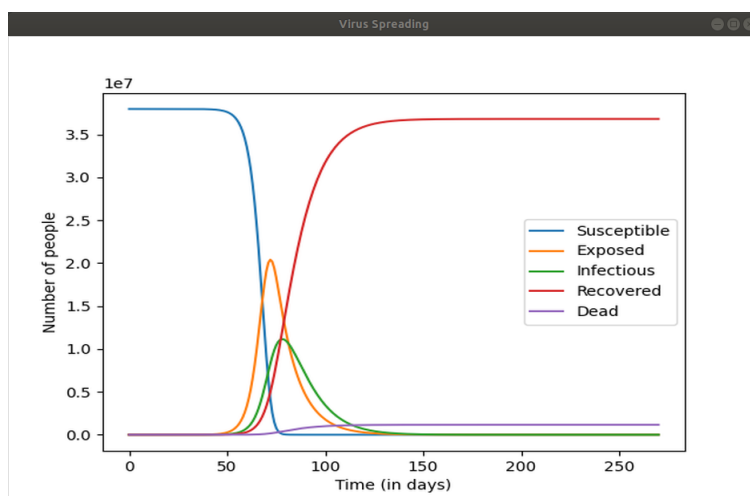
Po kilkudziesięciu pierwszych dniach pandemii wirus odry zdążył rozprzestrzenić się na niemal wszystkich kontynentach, zarażając niektóre z państw praktycznie w całości:



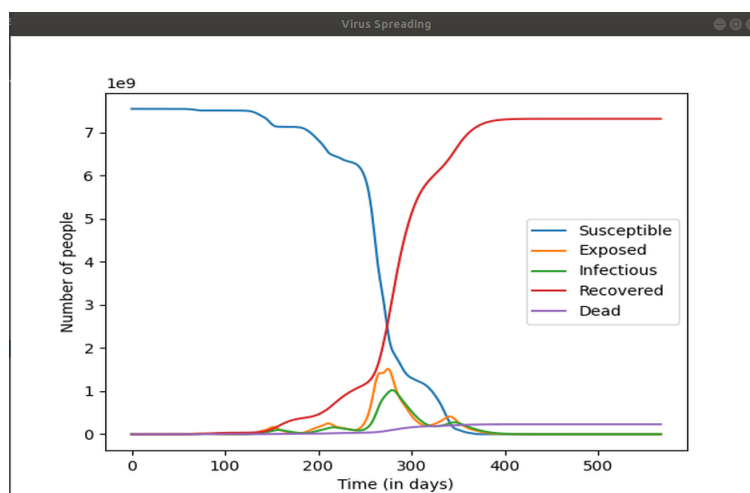
Pod koniec epidemii okazało się, że nieliczne regiony nie zostały zarażone patogenem:



Epidemia przebiegła w Polsce jak na wykresie:



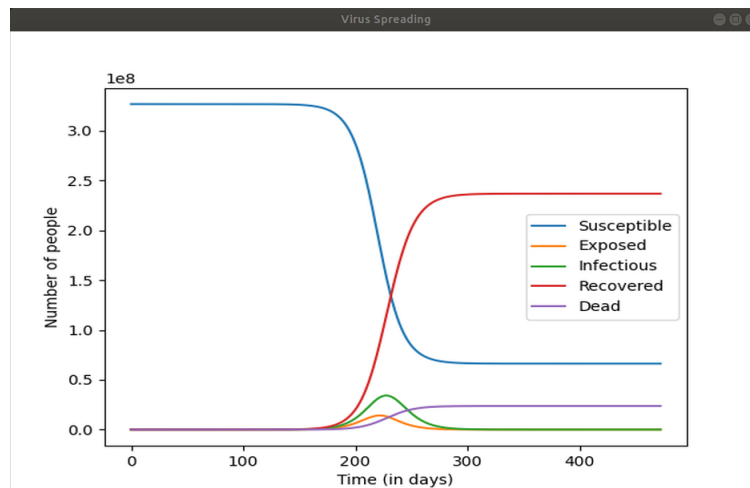
Natomiast dla całego świata wykres przebiegu pandemii wygląda następująco:



5.3 Grypa hiszpanka w USA

Grypa hiszpanka jest doskonałym przykładem na to, jak nieprzewidywalna potrafi być choroba. Epidemia, która wybuchła w latach 1918-1919 charakteryzowała się 3 falami o zmiennej zakaźliwości i równie nierówną śmiertelnością. Dla potrzeb symulacji przyjęliśmy parametry choroby charakterystyczne dla drugiej fali epidemii (charakteryzujące się wysoką śmiertelnością). USA jest największą potęgą gospodarczą, stanowiąc centralny punkt komunikacyjny w przypadku większości tras handlowych oraz charakteryzując się wysoce rozwiniętym transportem międzynarodowym.

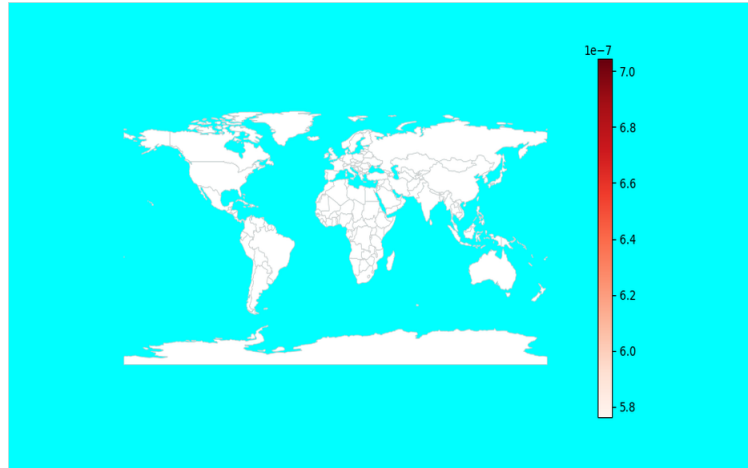
W tym przypadku okazało się, że jedyny kraj, jaki ucierpiał z powodu epidemii, to kraj startowy, czyli USA. Wykres przebiegu pandemii grypy hiszpanki to jednocześnie wykres jej przebiegu w USA. Można to zaobserwować na zdjęciu poniżej:



5.4 AIDS na Sri Lance

Wirus HIV jest odpowiedzialny za jedną z największych pandemii w historii ludzkości - pandemię AIDS. Jest to nieuleczalna choroba, która powoduje trwałe osłabienie odporności, które może być śmiertelne w przypadku przebycia innej choroby (na potrzeby symulacji przyjęliśmy jednak, że AIDS posiada pewną śmiertelność). Sri Lanka to mała wyspa położona u wybrzeży Indii, która ze względu na położenie nie posiada granic lądowych.

W przypadku przebiegu epidemii na Sri Lance, mapa po ok. 8 latach nie zmieniła się praktycznie wcale w stosunku do początku pandemii:



Dodatkowo, AIDS jest jedyną chorobą w naszej symulacji, która przekracza dozwo-
lony limit dni epidemii (domyślnie 5000).

6 Wnioski

6.1 Różnice między poszczególnymi chorobami

Na podstawie naszej symulacji można wywnioskować, że choroby o wysokiej śmiertelności wbrew pozorom są mniej niebezpieczne od chorób o wysokiej zaraźliwości. Dobrym przykładem może być grypa hiszpanka, która mimo wysokiej śmiertelności nie była w stanie zabić więcej osób niż np. odra. Jest to spowodowane tym, że przy wysokiej śmiertelności osoby chore nie są w stanie przenieść patogenu na kolejne osoby, przez co epidemia dosyć szybko wygasa. Natomiast dla chorób o niskiej śmiertelności i wysokiej zaraźliwości obserwuje się odwrotne zjawisko - epidemia trwa dłużej, ale jednocześnie jest w stanie odebrać życie większej liczbie ludzi. Mimo oszacowanej przez nas śmiertelności na poziomie ok. 6% COVID-19 był w stanie rozprzestrzenić się po całym świecie i zabić o jeden rząd wielkości więcej ludzi niż wspomniana grypa hiszpanka.

6.2 Różnice między wykresami dla państw i dla świata

Porównując ze sobą wykresy pandemii dla świata i dla pojedynczego państwa można dojść do wniosku, iż prezentują one dwa kompletnie różne zjawiska. Krzywe dla państw są S-kształtne i "gładkie", podczas gdy wykres dla całego świata sprawia wrażenie chaotycznego i mniej przewidywalnego. Jest to jednak mylny wniosek. Z uwagi na charakter modelu (opisanego za pomocą kilku równań różniczkowych), wygląd krzywych państwowych nie powinien dziwić. Natomiast za charakterystyczny kształt wykresu dla całego świata odpowiada niejednoczesny czas trwania pandemii dla poszczególnych krajów. Państwa przechodzą epidemię w przecinających się, ale nie całkowicie pokrywających się przedziałach czasowych. Oznacza to, iż niektóre państwa mogą zostać zainfekowane nawet kilka lat po wybuchu epidemii. Charakterystyczne grzbiety na krzywych *exposed* oraz *infectious* to efekt zakażenia państw o dużej populacji, jak Chiny czy Indie.

6.3 Różnice pomiędzy symulacją a rzeczywistością

Zaproponowane przez nas rozwiązanie, oparte o model SEIR, daje dość dobre przybliżenie rzeczywistości, aczkolwiek w wielu punktach nie jest z nią zgodne. Najważniejsza rzecz, która wykazuje niezgodność z obserwowanym zjawiskiem w świecie rzeczywistym, to nachylenie krzywych na wykresie przebiegu epidemii. Osoby zainfekowane oraz narażone na patogen są gwałtownie zarażane, a zarazem bardzo szybko przechodzą do grupy *recovered*. Również ta grupa bardzo szybko zwiększa swoją liczebność. Kolejny aspekt niezgodny z rzeczywistością to faktyczna liczba osób, które przejdą chorobę. Według różnych szacunków, aby naród uzyskał tzw. "odporność stadną", patogenem musi zarazić się (a następnie wyzdrowieć) od 60 do 80 procent społeczeństwa w danym kraju. Dla naszego modelu praktycznie każda choroba prędzej czy później jest w stanie zarazić niemalże wszystkich obywateli danego państwa. Może to być spowodowane tym, że współczynniki charakteryzujące model są niezmiennie, co nie występuje w przyrodzie. W obliczu pandemii wiele państw stosuje obostrzenia, które mają na celu zminimalizowanie transfer patogenu, co wpływa na owe współczynniki. Przez to w rzeczywistości możemy zaobserwować efekt wypłaszczenia krzywej zachorowań, czyli jej wydłużenie w czasie z jednoczesnym spadkiem maksymalnej liczby aktywnych zachorowań. Nie zgadzają się również wyniki śmiertelności w różnych państwach. W naszej symulacji występuje ona na przybliżonym poziomie - jest to związane z biernością wszystkich państw wobec patogenu. W rzeczywistości każde państwo może zastosować odmienną strategię walki z patogenem. Poza tym istotny wpływ na rezultaty tych zmagania ma sama charakterystyka państwa - klimat, demografia czy stopień rozwoju transportu międzynarodowego.

6.4 Pozostałe

Mimo wykazanych wyżej różnic, w naszej symulacji dostrzegamy też pewne cechy wspólne do rzeczywistości. Symulacja epidemii AIDS na Sri Lance może prowadzić do dwóch ciekawych wniosków - pierwszy z nich to fakt, że rozpoczęcie epidemii w państwie o słabo rozwiniętym transporcie może mieć wpływ na jej spowolnienie. Drugi z nich to sam czas trwania epidemii AIDS - dla przyjętych przez nas parametrów okazuje się, że epidemia AIDS może trwać nawet 30-40 lat. Jest to niepokojąca zgodność z rzeczywistością, gdyż świat od lat zmaga się z epidemią wirusa HIV, zwłaszcza w Afryce, a wobec braku leku przeciwko temu patogenowi, sytuacja może się znacznie pogorszyć. Ponadto, użyty model może być dobrym przykładem, jak ważne są nie tylko działania prewencyjne państw, ale i skoordynowanie ich przez dane państwa. W przypadku braku reakcji na epidemię, model sugeruje, iż większość patogenów może zarażać nawet ponad 95% ludzkości. Istotnym wnioskiem jest natomiast fakt, że matematyczne modele próbują przewidywać rozwój epidemii dla zadanego zestawu współczynników, natomiast faktycznie współczynniki ustalane są na podstawie zakończonej epidemii.

7 Propozycje rozwijania w przyszłości

7.1 Uwzględnienie demografii

W przyszłości chcielibyśmy uwzględnić śmiertelność spowodowaną innymi chorobami oraz przyrost populacji. Dzięki temu nasza symulacja byłaby bliższa faktycznej sytuacji rozprzestrzeniania się patogenu. Ponadto, istotne może się okazać uwzględnienie takich informacji, jak chociażby średni wiek obywateli danego państwa.

7.2 Gotowe scenariusze

Pomysł, który pragniemy zrealizować to danie użytkownikowi możliwości wyboru przygotowanych przez nas wcześniej gotowych scenariuszy, w których ustalone byłyby współczynniki takie jak: częstotliwość mycia rąk czy zwiększone podstawowe środki przeznaczone na służbę zdrowia. Użytkownik mógłby też sam dostosować te współczynniki wedle swojego uznania. Innym pomysłem na gotowe scenariusze jest przygotowanie słynnych epidemii jak wspomniania grypa hiszpanka czy epidemia dżumy w XIV-wiecznej Europie.

7.3 Reakcje państw na epidemię

W planach mamy również dodanie odpowiednich reakcji państw na epidemię. Są to m. in. zamknięcie granic czy nakaz pozostawania w swoich domach. W ten sposób

moglibyśmy symulować pewne posunięcia państw, jakie mogą one poczynić w obliczu kryzysu związanego z pandemią.

7.4 Optymalizacja generowania map

Niestety w aktualnym stanie projektu generowanie map trwa istotnie dłużej niż sama symulacja. Chcielibyśmy zoptymalizować to, aby użytkownik nie musiał czekać zbyt długo na wyniki.