Epidemic simulations - Final Report
Project Embodied AI: Group Hotel

Iman Firouzifard
2670146

Othmane Aifoute
2676740

Danila Rusinkiewicz
2675515

Mariya Velychko
2664879

Sofia Fraile
2659410

Vrije Universiteit Amsterdam
June 29, 2021

Abstract

This paper addresses the process of simulating individual behaviour in an epidemic environment. Various real-life situations were modelled and examined using agent-based simulations. These situations consisted out of a base-case scenario where agents could move freely, both partial and lockdown scenarios, a scenario with a central hub, a scenario with a quarantine zone and a scenario with social distancing restrictions. In all scenarios, infected agents were able to spread the virus to other susceptible agents, and eventually recover after a certain time. The recovery period of individual agents was modelled based on their age. With the use of the simulations, the spread of the virus among a population of agents was measured using two metrics: the susceptible ratio, and the reproduction number. For each scenario, a total of 30 simulations have been performed, and for statistical analysis, the ANOVA-test has been performed for the susceptible ratio. The results showed that the scenarios including the lockdown, quarantine-zone, and social-distancing had significantly different results compared to the base scenario. It was concluded that these measures are effective to control the infection numbers of COVID-19.

1 Introduction

Traditional AI involves mostly algorithmic approaches that have several limitations. The agents perform marvellously in controlled conditions or environments, however, the main limitation is when the agent requires to interact with the real world, with real-world problems that are not entirely controlled or set up. Therefore, algorithmic approaches have trouble in containing interactions that require "common sense" or knowledge of human experiences.

Embodied AI, on the other hand, tackles this limitation by working with real-world physical systems. For instance, during the pandemic, the ability to work with physical agents was impacted. Hence, deploying an embodied AI system into a virtual world offers a unique opportunity to evaluate and develop embodied AIs within an environment where human and computer are nearly united (1).

Swarm robotics is a unique method of coordinating a large number of relatively basic droids that are inspired by social insects (2). The collective behaviour of these droids stems from the interaction between themselves and the environment which is perceived through sensory inputs or internal memory, and guided by swarm intelligence that has the characteristics of being composed by multiple agents. The agents are all identical or identify with different typologies, and their interactions are based on behavioural rules that use local information or information that is either exchanged between the agents directly or through their environment.

For this paper we concentrate on the collective behaviour of the population in a pandemic situation such as COVID-19 and our proposed approach stems from the idea of swarm intelligence where the

agents collect their knowledge through the interactions with other agents and their environment, which both affect their overall behaviour. In this case, the SIR model has been used as a basis for the modelling of the spread of COVID-19. Specifically, it has been modelled whether individuals are susceptible, infected or recovered from COVID-19, and based on which factors do the infected agents spread the virus to others.

1.1 Problem statement

The COVID-19 outbreak has shown that taking specific measures is vital to the well-being of a country as not taking them will lead to multiple repercussions such as the rise of the infection numbers. Therefore, strict measures need to be taken to minimize the impact (3). Accordingly, as a problem statement, we are focusing on understanding how the COVID-19 virus spreads amongst a population in a specific environment and which type of measures allow to keep the number of susceptible agents as high as possible. A useful tool to measure the impact of a virus is by examining the number of new infections. Hence, the problem at hand is how to keep the infections as low as possible, which consequently means keeping the number of susceptible agents as high as possible and additionally having the smallest reproduction number R0 which estimates the speed at which a disease is capable of spreading in a population.

In this paper, the specific research questions are: What type of scenarios enhance the spread of the COVID-19, and consequently, what type of scenarios can rather help limit the number of infections in a population? Therefore, to answer these questions, different scenarios will be implemented and simulated, since different environments may lead to different collective dynamics regarding the spread of the virus. The scenarios will be compared against each other and evaluated based on the described metrics.

1.2 Related work

Due to the large social and economic impact of the pandemic on a global scale, much research is being conducted into the understanding of the spread of COVID-19 and possible preventive measures. As seen in the research paper by (4), the mathematical SIR model was utilized to simulate COVID-19 waves in different regions with real-life data (4).

The SIR model is a classic model that uses three equations for the changes in the number of susceptible persons (S); infected persons spreading the infection (I) and removed persons (R), which is the sum of isolated, immunized, and deceased individuals. This model is especially useful for the prediction of the duration of epidemics within different environments, such as different countries, and for the detection of new waves due to changes in pandemic conditions.

Furthermore, as it is a desirable outcome to limit the consequences of the pandemic as much as possible, much research has been done into types of measures that can limit the spread of the virus effectively. For instance, in the research conducted by Murphy (3), the effectiveness of public measures such as social distancing has been interrogated by looking at real-life data from different countries. This modelling approach was also taken by Chen *et al.* (5), where the model was extended to be time-dependent and include detected and undetected cases of the virus. This was done to give a more realistic approach to find optimal solutions and restrictions, which show that avoiding contact is the leading approach to reducing the number of infections.

Lastly, Manotosh *et al.* (6) takes a mathematical approach toward the spread of the virus, which provides more insight into the calculation of the reproduction number R to minimise the impact of the different variants of COVID-19 in overly populated areas.

2 Methodology

2.1 SIR Model

To find an answer to the problem statement, an agent-based simulation has been used to model the population dynamics during an epidemic. To simulate the agent-based disease spread, the SIR model has been used (see figure 1). This model allocates each member of the agent population to one of the following groups:

- The *Susceptible* group contains agents who are vulnerable to exposure to infectious agents.
- The *Infectious* group represents the infected people that can pass the disease to susceptible people and can be recovered after some specific period.
- The group of *Recovered or Removed* agents that have acquired immunity to the disease. Thus, once an agent has recovered, it cannot be infectious or susceptible to the same disease anymore. It should

be noted that for the purpose of this simulation, it has been decided that deceased agents are included within the group of Recovered agents.
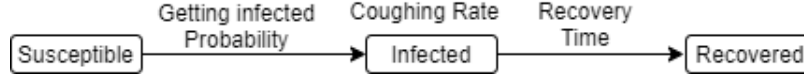

Figure 1: The SIR model

2.2 Recovery Time

Furthermore, a decision has been made to extend the SIR model with the concept of age to be able to model individual differences in recovery time and therefore to allow more realistic simulations. For this, the agent population has been divided into five age-groups where its distribution was based on the population distribution of the Dutch population (see figure 3) (7). Using this distribution, each agent was randomly given a specific age within the age range. Additionally, it can be noted from the figure 3 that individuals younger than 12 years old have not been included considering that prior research has shown that young children play only a small role in the spread of the disease so far (8). This approach has been chosen due to its ability to provide insightful information on real-life situations.

The general formula used to calculate the recovery time of a specific agent takes into account the minimum recovery time together with the age of the agent which is divided by the maximum age of the population and multiplied with the difference between maximum recovery time and minimum recovery time. Ultimately, some Gaussian noise has been added to represent the individual differences. This Recovery Time formula is the following:

$$RecoveryTime = MinRT + \frac{age}{maxAge} * (MaxRT - MinRT) + Noise \tag{1}$$

The minimum and the maximum recovery times (MinRT and MaxRT) have been based on the research of Voinsky *et al.* (9), where the Recovery periods range from 8 to 20 days. Based on this research, the minimum and the maximum recovery times in the simulations have been set to be between 8 and 20 seconds respectively, where each second represents a day.

2.3 Probability of getting infected

A final addition to the SIR model is the probability of getting infected. This probability takes into account both the number of infected agents who cough and are within an unsafe distance of the agent, and the average distance of all these agents from the susceptible agent. Both the coughing rate and the safe distance can be adjusted, however, based on the research by Agrawal *et al.* (10), the safe distance to maintain to reduce the risk of infection is set to be 2 meters and the coughing rate is 0.68 as Li *et al.* found in their research that this is the average rate (11).

The formula which calculates the probability of getting infected is:

$$P_{Infected} = 1 - \frac{averageDistance^{numAgents}}{(safeDistance - averageDistance)^{numAgents} + averageDistance^{numAgents}} \tag{2}$$

This formula was based on the Sigmoid function and has been defined in such a way that having multiple infected agents within the safe distance of the susceptible agent will exponentially increase the probability of getting infected.

2.4 Different Scenarios

In order to be able to use the above-mentioned SIR model to understand how the virus spreads in a specific environment and which type of measures allows for limiting the number of infections and thus have the highest susceptible rations, different scenarios have to be simulated. For this research, six different scenarios have been implemented:

1. The first implemented scenario is the base case scenario, which consisted of an environment where the agents could move freely without any measures or restrictions. Here, all infected agents could spread the virus to all susceptible individuals of the population. Due to this reason, it was expected that this scenario would result in a faster and more significant spread of the virus.

2. The next implemented scenario was a partial lockdown, which includes 8 lockdown areas in total. This partial lockdown allowed agents to leave their initial lockdown sites sporadically through small openings. As the agents are party-limited in their mobility, it is expected that more agents will remain susceptible in this scenario.

3. The previous scenario was subsequently expanded to a full lockdown scenario, which also consisted of 8 lockdown areas, however, without any openings such that the agents were not able to leave their lockdown sites. The design choice of the infected agents being inside was created to test the spread of the virus in a controlled area and display the impact of the virus containment. As the agents were especially limited in their mobility in this scenario, it was expected that more agents would be infected with the virus when compared to the partial lockdown.

4. As an additional scenario, a central hub was implemented, that can be visited by some proportion of agents depending on a adjustable probability. The probability of entering the hub can be set to a certain number, such that not all the agents would enter. Whenever an agent enters the hub, which in our simulation is a supermarket, it will wander in the area for a specific time period. It was expected that the presence of a supermarket would enhance the spread of the virus as there is a chance of a faster infection process within the site.

5. A quarantine scenario has also been simulated to show the effects of containing the virus. In this scenario, all infected agents are transferred to the enclosed and secluded quarantined zone after a specific wandering time and will only be released once they are recovered and are therefore not infectious anymore. This scenario was chosen to demonstrate the consequences of having a quarantine place as well as the necessary transportation to such a place and the risk of infecting others on the way. It was expected that this scenario would have a smaller number of infections when compared to the base scenario and therefore a higher number of susceptible agents.

6. The final implemented scenario included social distancing, where agents will keep a certain distance from others if that is possible. Simulating the social distancing restriction was chosen to demonstrate the social impact of this measure in the spread of the virus. Here, a drop in the number of infected agents was expected when compared to the base case thus a decrease of the COVID-19 spread.

2.5 Metrics

As previously mentioned, to evaluate the spread of the virus in different scenarios, two metrics have been used. The first metric is the ratio of the susceptible agents. The formula of the Susceptible ratio is:

$$Susceptible ratio = \frac{nr.SusceptibleAgents}{totalnr.AgentsInPopulation} \qquad (3)$$

This metric was used to evaluate what particular scenario keeps the susceptible rate high, to find the scenario in which fewer infections occur.

Furthermore, another metric that has been used for the evaluation of different scenarios is the reproduction factor R0, which expresses the number of agents every infected agent infects. After each iteration, the reproduction factor will be determined by the ratio between the infected and susceptible individuals at that point in time. This results in having a different reproduction rate for each iteration which is used to show how this metric progresses throughout the entire simulation. The corresponding formula is:

$$R_0 = \frac{nr.InfectedAgents}{nr.SusceptibleAgents} \qquad (4)$$

The metric of using the reproduction factor was selected after similar research conducted by Bagal *et al.* that used an R-value as one of the metrics to evaluate the spread of COVID in their models (12).

3 Experiments

3.1 Main Parameters

The spread of the coronavirus within a specific environment is influenced by several parameters. To have realistic as possible simulations, the parameter values of the mentioned scenarios are mainly based on literature and prior research. For the experiment, to perform a proper analysis and a comparative study, the parameter values have been kept constant such that the results are comparable.

One of the main initialized parameters is the number of agents in the simulation environment, which has been set to 100 such that agents are likely to get in close proximity to each other while wandering. Moreover, the initial infection rate which is the percentage of infected agents at the initialization of each simulation has been set to 5%, as all epidemics initially start with a small number of infected individuals.

3.2 Set-up

In order to be able to draw reliable conclusions from the comparison of the six different scenarios, an experimental setup is necessary. For the setup, it has been decided that each scenario will be run 30 times in total, such that normality can be assumed (13).

Furthermore, it has been observed that in the base case scenario, all infected agents recover around 3000 iterations. Therefore, it has been decided that each of the individual simulations shall consist of 3000 iterations.

The results of the reproduction rates of all simulations have been generalized into one graph which will be addressed in the section below. Additionally, the average susceptible final ratios for each scenario after 30 simulations have been inspected, together with their respective highest infected ratio.

3.3 Results

3.3.1 Visual Analysis

By running the simulations for each of the mentioned scenarios, we have obtained various results regarding the two metrics which were mentioned in the Set-up section.

In Figure 4, the evolution of the reproduction rate throughout the simulations of each scenario is shown, where each color represents a different scenario. Additionally, in Figure 5, the bar-chart exhibits the highest reproduction rates, thus the peaks of the previous graph. It can be noted from the graph and the bar chart that the virus spreads faster when there are no measures since those scenarios have the highest peak of reproduction rates. The graph in Figure 4 also reveals that having any measures, such as incorporating the social distancing, the quarantining, or any lockdown, significantly reduces the value of the reproduction rate.

Furthermore, from figure 5, it can be seen that the maximum R-value in the partial lockdown seems to be lower than the maximum R-value of the full lockdown scenario. A possible explanation for this observation is that in the full lockdown scenario, the agents were enclosed in the sites leading to a faster spread in a short amount of time, whereas in the partial lockdown the agents could leave the areas and avoid an outbreak. However, it is important to highlight that the overall reproduction rate of the partial lockdown is generally higher as it has a longer tail (seen in figure 4). Finally, the lowest peak of the reproduction rate belongs to the social distancing scenario, meaning that from this visual analysis, this scenario has the lowest spread of the virus.

In figure 2, a comparison is shown between the final susceptible ratio versus the highest ratio of the infected agents for all six scenarios. Here, it can be observed that in the supermarket scenario the final susceptible ratio is 0.14, which is higher than the scenario with no measures, which has the lowest susceptible ratio of 0.12. Moreover, the scenario with no restrictions have in fact the highest number of infections. Therefore, these two scenarios seem to be leading to the highest infection ratio and the lowest susceptible ratio. A possible explanation for this may be that in these scenarios, agents have more chance to be close to each other and therefore infect more agents.
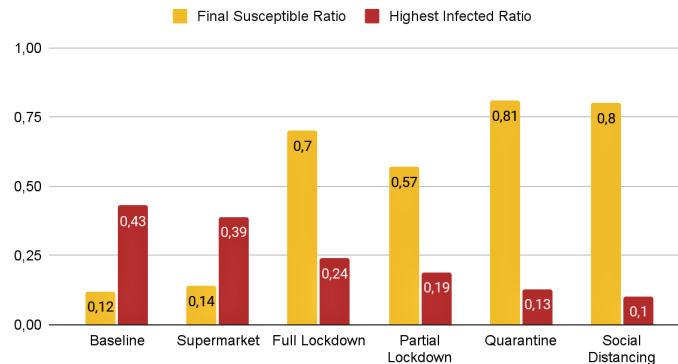

Figure 2: Final Susceptible Ratio vs. Highest Infected Ratio

Additionally, it can be observed from figure 2 that the partial lockdown with the final susceptible ratio of 0.57 was quite effective compared to the scenario with no restrictions. Nonetheless, the partial lockdown is undoubtedly less effective compared to the full lockdown where the final susceptibility ratio is 0.70. Finally, it can be noted that the quarantine and social distancing scenarios have the highest susceptible ratios, which means they are the best-considered policies that decrease the spread of the virus. However, the measured infected ratio is lower when social distancing is applied rather than when the quarantine measures are implemented which suggests that social distancing may be a better policy.

Still, to confirm whether there is any significant difference between the employment of the two restrictions, a statistical analysis is necessary.

3.3.2 Statistical Analysis: Susceptible Ratio

To interrogate whether the differences in the susceptible ratios between different scenarios are significant, statistical analysis has to be performed. Since multiple scenarios will have to be compared against one another, a two-sided ANOVA statistical test has been chosen to compare whether the differences in the mean values of the susceptible ratios are significant between the six scenarios. Here, it has been decided to keep the focus on the comparison between each of the five additional scenes against the base scenario - to answer the research question. Furthermore, the pairwise test between the quarantine and the social distancing will be reviewed as well to determine whether their difference in the mean is statistically significant.

The chosen significance level is set to be 95%, thus alpha is 0.05. Using the ANOVA test, we have obtained different p-values for different scenarios. It can be concluded that if the p-value of a pairwise comparison of these scenarios is higher than alpha, then the difference is not statistically significant.

Table 1 shows all the p-values of the six pairwise comparisons. In this table, it can be seen from the supermarket-baseline comparison, that its respective p-value of 0.945 is much greater than alpha. Therefore, the difference in the susceptible ratios is not statistically significant, which means there is not enough evidence to state that the supermarket scenario is different from the baseline with no restrictions. Additionally, when comparing the quarantine and the social distancing scenarios, it can be seen that the p-value of 0.997 is also higher than alpha, thus the difference of the result between these two is also not statistically significant.

Nonetheless, the first four pairwise comparison tests have a p-value that is nearly 0. As these values are all lower than the alpha of 0.05, the differences in all these comparisons are statistically significant. Therefore, it can be concluded that these first four scenarios have significantly different susceptible ratios when compared to the base case that consists of no restrictions.

4 Conclusion

This paper has addressed how the spread of the COVID-19 virus could be simulated as realistic as possible using a SIR model. Additionally, it has been examined what policies could limit the number of infections in a population. Throughout several simulations of the six mentioned scenarios, we observed that certain policies such as full and partial lockdowns were effective to limit the number of infected people. Nonetheless, these were less convincing and less successful in reducing the number of infections when compared to the quarantine or social distancing scenarios. These two latter scenarios have been undeniably the most effective with the highest final susceptible ratios.

Although the findings of the simulations were matching our expectations, the most challenging part of this research was the need for a realistic model. This is crucial when epidemiologists desire to understand the process of a pandemic and apply the findings of their simulation-based studies to a real-world environment. Therefore, further research could be conducted to understand the spread of the virus and possibly even prevent future epidemics. For instance, the SIR model could be expanded to a SIRD model, that would include deceased individuals and therefore be resembling authentic situations. Another potential future development could be to enforce a model with possible variants of the virus, the spread of which would not be contained by the different policies. Lastly, an ambitious idea could be the expansion of the model with more demographic characteristics such as gender, ethnicity, education or social-economic status, to obtain a more complex and representative model.

Authors' contributions

All authors have equally contributed to the presentations and to the report. The simulations were mainly performed by Iman Firouzifard.

References

[1] D. J. H. Burden, "Deploying embodied ai into virtual worlds," *Applications and Innovations in Intelligent Systems XVI*, pp. 103–115, 2009.

[2] E. Şahin, "Swarm robotics: From sources of inspiration to domains of application," *Swarm Robotics*, pp. 10–20, 2005.

[3] P. Murphy, *COVID-19 : proportionality, public policy and social distancing ISBN: 9789811575136*. Palgrave Pivot, 2020.

[4] I. Nesteruk, *COVID-19 pandemic dynamics : mathematical simulations ISBN: 9789813364158*. Singapore, Singapore Springer, 2021.

[5] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A time-dependent sir model for covid-19 with undetectable infected persons," *IEEE Transactions on Network Science and Engineering*, pp. 1–1, 2020.

[6] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, and T. Kar, "A model based study on the dynamics of covid-19: Prediction and control," *Chaos, Solitons, and Fractals*, vol. 136, p. 109889, 07 2020.

[7] C. B. v. d. Statistiek, "bevolking leeftijd https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/bevolking."

[8] "Kinderen, school en covid-19 https://www.rivm.nl/coronavirus-covid-19/kinderen."

[9] I. Voinsky, G. Baristaite, and D. Gurwitz, "Effects of age and sex on recovery from covid-19: Analysis of 5769 israeli patients," *Journal of Infection*, vol. 81, pp. e102–e103, 08 2020.

[10] A. Agrawal and R. Bhardwaj, "Probability of covid-19 infection by cough of a normal person and a super-spreader," *Physics of Fluids*, vol. 33, p. 031704, 03 2021.

[11] L. Li, T. Huang, Y. Wang, Z. Wang, Y. Liang, T. Huang, H. Zhang, W. Sun, and Y. Wang, "Covid-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis," *Journal of Medical Virology*, vol. 92, pp. 577–583, 03 2020.

[12] D. K. Bagal, A. Rath, A. Barua, and D. Patnaik, "Estimating the parameters of susceptible-infected-recovered model of covid-19 cases in india during lockdown periods," *Chaos, Solitons Fractals*, vol. 140, p. 110154, 11 2020.

[13] M. F. Triola, *Elementary statistics ISBN: 9781323578179*. Pearson Education, Inc, 2018.
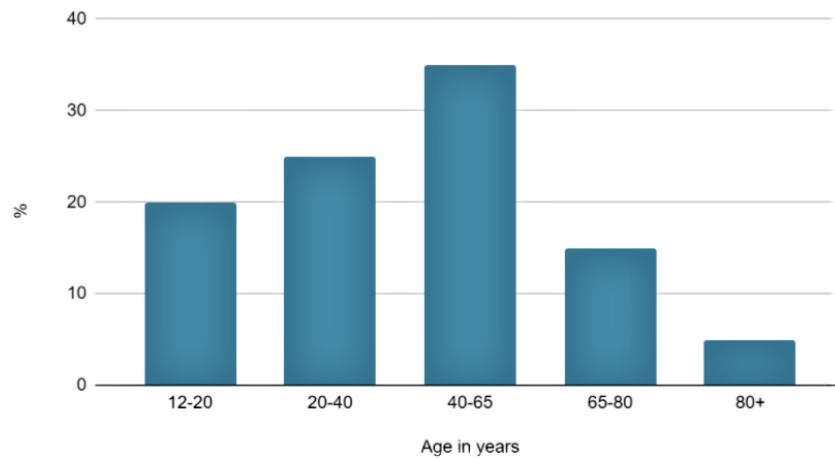
APPENDIX
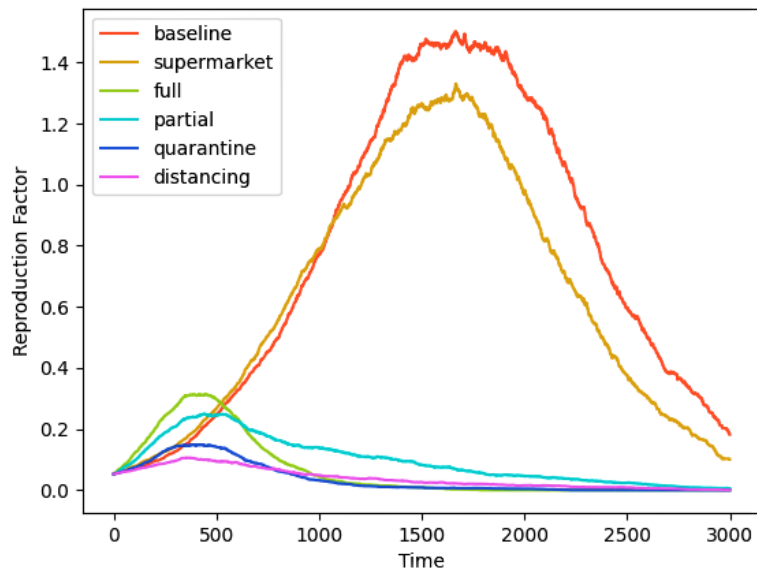

Figure 3: Age distribution of the agent population

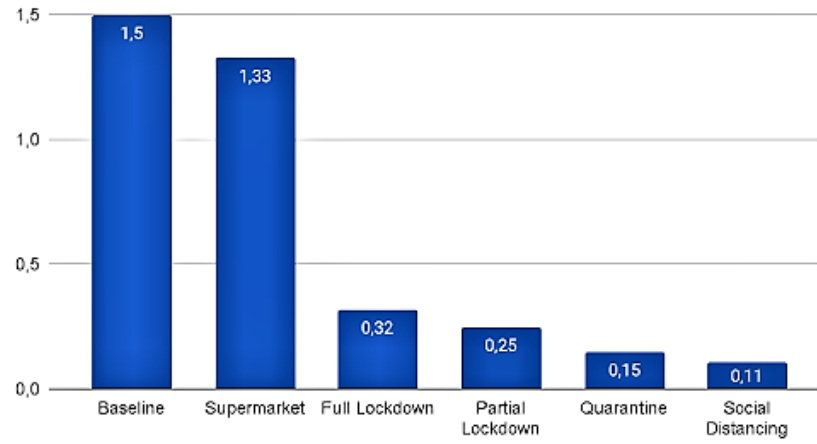Figure 4: Reproduction rate for all six scenarios

Figure 5: Highest reproduction rate

Scenario1 - Scenario2	P-value
baseline - distancing	1.331633e-74
baseline - full	9.276678e-65
baseline - partial	3.702299e-51
baseline - quarantine	3.223087e-75
baseline - supermarket	9.452085e-01
quarantine - distancing	9.973971e-01

Table 1: P-values for each pairwise comparison test