

A Covid-19 Data Analysis During the First Five Months in Senegal

Hamidou Dathe[1,2] and Python Ndekou Tandong Paul[2(✉)]

[1] Sorbonne Université, UMMISCO, UMI 209, 93143 Bondy, France
hamidou.dathe@ucad.edu.sn
[2] Department of Mathematics and Computer Science,Cheikh Anta Diop University, Dakar, Senegal
python.paul@ucad.edu.sn

Abstract. In this paper, we present an analysis of Covid-19 data during the first 138 days in Senegal. The data come from the website of the Ministry of Health and Social Action. We build a database with R and Python scripts helping to carry out data analysis for answering questions that support decision-makers. We compute the basic reproduction rate with the collected data using the SIRD model approach (susceptible, infected, recovered, death). Results show that there are two phases in the evolution of infected cases, 4.75% are between the first day of the Covid-19 outbreak and the 50th day, 95.25% are between the 51st day and the 138th day corresponding to the period of the reopening of airports. The average value of the basic reproduction rate during the first 138 days is 2.77. We observed a strong correlation between infected cases and community infected cases. The obtained results could help decision-makers to evaluate the statistical evolution of infected persons.

Keywords: Data analysis · Covid-19 disease · Statistical analysis · Basic reproduction rate · UML language · Mathematical modeling

1 Introduction

The Covid-19 pandemic has officially appeared in Senegal since March 2, 2020, [2]. On April 26, 736 cases of infection and nine deaths were confirmed [2]. For fighting against the spread of covid-19, Senegal Government had canceled international flights since Friday, March 20, 2020 [2]. On March 14, 2020, the public authority closed schools, prohibited public events, canceled events involving more than 50 people [2]. On March 23, 2020, Senegal Government declared a state of emergency throughout the territory and extended it for 30 days on April 4. On April 19, the Government imposed the mandatory wearing of masks in public and private services, shops and transports [2]. Each day, Senegal Government presents collected data as follows: the number of infected cases, the number of tested cases, the number of the imported cases, the number of community cases, the number of infected cases under treatment, and the number of

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12950, pp. 580–588, 2021.
https://doi.org/10.1007/978-3-030-86960-1_41

deaths cases. Data mining models have been used to evaluate the features of the covid-19 outbreak in several countries in the world [4]. For fighting against the transmission of infectious diseases, one can use data mining technics. Jahanbin et al. [5] have used social networks mining [10] in the management of the infectious disease systems. Huang and et al. have worked on the clinical features of patients infected with the covid-19 in China [6]. The study of the epidemiological characteristics of the novel coronavirus diseases has helped public authorities all over the world to develop new strategies for fighting the covid-19 disease [7–9]. The objectives of this work are: collect the Covid-19 disease data, develop a database containing these data, carry out statistical analysis of Covid-19 data, find correlations between statistical variables, examine statistical variations in the different critical periods during the evolution of Covid-19 infection cases.

2 Basic Mathematical Model

For computing the average basic reproduction rate with the real data, we first formulate the covid-19 mathematical model and carry out mathematical analysis that helps to compute the value of R_0. We use the compartments S,I,R,D (Susceptible, Infectious, Recovered, Death) [1] during the covid-19 infection period. We present the ordinary differential equations describing the covid-19 transmission as follows:

$$\begin{cases} \dfrac{dS}{dt} = \lambda - \mu S - \dfrac{\beta IS}{N} \\ \dfrac{dI}{dt} = \dfrac{\beta IS}{N} - \gamma I - \mu I - \alpha I \\ \dfrac{dR}{dt} = \gamma I - \mu R \\ \dfrac{dD}{dt} = \alpha I \end{cases} \tag{1}$$

where the susceptible population is characterized by the death rate μ, the susceptible recruitment rate λ, the infectious contact rate β, the death rate due to the disease α, the recoved rate γ, the total population N. The Mathematical analysis shows that the disease-free equilibrium (DFE) point is: $(S, I, R, D) = \left(\dfrac{\lambda}{\mu}, 0, 0, 0\right)$. We used the next generation matrix approach described by Diekmann et al. [3] for the computation of R_0. We have found that:

$$R_0 = \frac{\beta}{(\mu + \alpha) + \gamma}$$

when posing $R_0(t) = \dfrac{\beta \dfrac{\beta I(t) S(t)}{N}}{(\mu + \alpha)\dfrac{\beta I(t) S(t)}{N} + \gamma \dfrac{\beta I(t) S(t)}{N(t)}}$

assume that NIP(t) = number of infected persons at time t, NDP(t) = number of death persons at time t, NRP(t) = number of recovered persons at time t and let $R_1(t)$ be $R_1(t) = \dfrac{NIP(t)}{NDP(t) + NRP(t)}$.
If $R_1(t) < 1$ there are insufficient new infection, and the covid-19 cannot invade the population. When $R_1(t) > 1$, the covid-19 disease may persist in the environment. We have used $R_1(t)$ formula for computing the basic reproduction rate using collected data.

3 Data Modeling

3.1 Development of the Database

Like several African countries, Senegal had been preparing for the Covid-19 outbreak and developed a plan to fight against this disease. Hospitals and hotels had been recommended for the treatment of infected individuals. Covid-19 test centers have also been identified and placed at the disposal of the Ministry of Public Health. From the appearance of the first cases of the infection of this disease, the following data were published by the Minister of Public Health each day: the number of tested cases, the number of infected cases, the number of infectious contact cases, the number of deaths, the number of recovered cases, the number of infectious community cases. We used a UML (unified modeling language) class diagram (Fig. 1) to show the relationships between the interacting entities in Senegal. The UML model helps to build the Covid-19 database. The dataset used in this study comes from the different test centers and is updated every day. It's possible to see all the data on the Ministry of Public Health (http://www.sante.gouv.sn) website. In the first step of the data treatment, data were collected in each test center with the detection of errors. We carried out different scripts for plotting the data. The study of the correlation between two or several variables helps to know the intensity of the relationship between variables. We have computed for each couple of variables (x, y) the coefficient of linear correlation and have obtained the matrix of correlation. Mathematical models used to compute correlations between two statistical variables x and y are presented as follows:

$$\begin{cases} rcol = & \dfrac{cov(x,y)}{\sigma_x \sigma_y} \\ cov(x,y) = & \dfrac{1}{p}\displaystyle\sum_{i=1}^{p} x_i y_i - (E(x)E(y)) \\ \sigma_x^2 = & \dfrac{1}{p}\displaystyle\sum_{i=1}^{p} n_i x_i^2 - (E(x))^2, \\ \sigma_y^2 = & \dfrac{1}{p}\displaystyle\sum_{i=1}^{p} n_i y_i^2 - (E(y))^2. \end{cases} \tag{2}$$

p is the size of data, n_i is the size of x_i. rcol is the linear correlation between x and y, σ_x is a standard deviation of x, the standard deviation helps to measure the dispersion of the data.

3.2 Description of Algorithms

In this part, we present the following algorithms (Table 1) used to compute the statistical results: mean(X, n) computes the average of data belonging to X. var(X, n) computes the variance of data belonging to X. cov(X, Y, n) computes the covariance between X and Y.

Table 1. Algorithms used for data analysis

Algorithm 1	Algorithm 2	Algorithm 3
function mean(X,n):real	function var(X,n):real	function cov(X,Y,n):real
1. s1 = 0	1.s2 = 0	1. s3 = 0
2. for i = 1 to n do	2. for j = 1 to n do	2. for k = 1 to n do
3 s1 = s1 + X[i]	3. s2 = s2 + X[j]**2	3. s3 = s3 + X[k] * Y[k]
4. endfor	4. endfor	4 endfor
5. xb1 = s1/n	5. xb2 = s2/n	5. xb3 = s3/n
6. return(xb1)	6. V = xb2-(mean(X,n)**2)	6. CV = xb3-(mean(X,n) * mean(Y,n))
7. endfunction	7. return(V)	7. return(CV)
	8. endfunction	8. endfunction

3.3 Matrix of Linear Correlation Coefficients

We suppose that x_i represents the data value at the intersection of row i and column x, y_i represents the data value at the intersection of row i and column y. The database file of Covid-19 data is called Covid-19.csv, using the R software language, the following command has helped to record data into the dataframe. dataframeData<-read.csv(Covid-19.csv, sp = ";", decimal = "."). The database has p fields, the dataframe also has p columns, Data[, i] allows to get all data from column i. A correlation matrix is a set of values allowing to measure the intensity of the relationship between variables. The correlation matrix contains the linear correlation coefficients between each variable and others. We used the function rcorr () coming from the R language to compute the Pearson and Spearman correlations (Table 2).

3.4 Results and Discussion

We collected data during 138 days. We observed that there were 20 days without death cases and recovered cases. There were two phases in the evolution of

Table 2. correlations between variables during the first 138 days of the presence of Covid-19

	InfectedC	ImportedC	ContactC	CommunityC	RecoveredC	DeathC
InfectedC	1.00	−0.004	0.98	0.71	0.67	0.60
ImportedC	−0.004	1.00	−0.087	0.028	0.037	0.12
ContactC	0.98	−0.08	1.00	0.59	0.64	0.53
CommunityuC	0.71	0.028	0.59	1.00	0.58	0.62
RecoveredC	0.67	0.037	0.64	0.58	1.00	0.50
DeathC	0.603	0.121	0.53	0.62	0.50	1.00

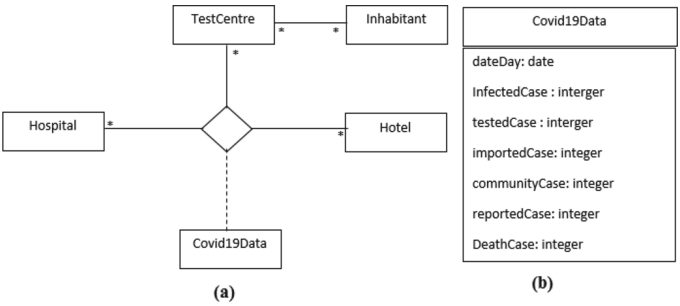


Fig. 1. UML class diagram showing different sources of data represented by a graphic formalism(a) and the Covid19 data class used for creating the database containing all data concerning the spread of the coronavirus 2019 in Senegal(b)

infected cases: 4.75 % of infected cases were between the first day and the 50th day, 95.25% of infected cases were between the 51st day and the 138th day (Fig. 2). There were three phases in the evolution of infectious imported cases: 49.72 % of infectious imported cases were between the first day and the 33rd day, 6.43 % of the infectious imported cases were between the 34th day and the 100th day, 43.85% of infectious imported cases were between the 101st day and the 138th day (Fig. 3). There were two phases in the evolution of infectious contact cases: 3.98% of infectious contact cases were between the first day and the 50th day, 96.02 % of infectious contact cases were between the 51st day and the 138th day (Fig. 4).

Concerning infectious community cases, we observed two scenarios: the first scenario in two phases showed that there were 3.71 % of infectious community cases between the first day and the 50th day, 96.29% of infectious community cases between the 51st day and the 138th day (Fig. 5). A second scenario showed that there were 3.71% of infectious community cases between the first day and the 50th day, 22.88% of infectious community cases were between the 51st day and the 90th day, 73.41% of infectious community cases were between the 91st day and the 138th day (Fig. 5). For the recovered cases, we observed two scenarios: the first scenario in two phases showed that there were 4.14% of recovered

cases between the first day and the 50th day, 95.86% of recovered cases were between the 51st day and the 138th day (Fig. 5).

The second scenario in three phases showed that there were 6.08% of recovered cases between the first day and the 60th day, 43.17% of recovered cases between the 61st day and the 100th day, 50.75% of recovered cases between the 101st day and the 138th day (Fig. 6). For death cases, we observed two scenarios: a first scenario showed that there were 3.06% of death cases between the first day and the 50th day, 95.94% of death cases between the 51st day and the 138th day (Fig. 7). A second scenario showed that there were 11.66% of death cases between the first day and the 70th day, 20.24% of death cases between the 71st day and the 100th day, 68.1% of death cases between the 101st day and the 138th day (Fig. 7). The Computation of daily infection rates showed that the average infection rate (infected cases/number of tests) between the first day and the 31st day was 29.54%, the average infection rate between the 32nd day and the 138th day was 8.68% (Fig. 8). For the computation of the basic reproduction number, we observed four phases during 119 days: between the first and the 11th day, the average basic reproduction rate was 4.77, between day 12 and day 31 the average basic reproduction rate was 1.48, between the day 32 and day 46 the average basic reproduction rate was 7.87, between the day 47 and day 119 the average basic reproduction rate was 1.7 (Fig. 9).

The average basic reproduction rate between the first day and the 138th day was 2.72, explaining the persistence of Covid19 disease. The matrix of correlations between variables showed that there was a very strong correlation between infected cases and infectious contact cases, a strong correlation between infected cases and infectious community cases.

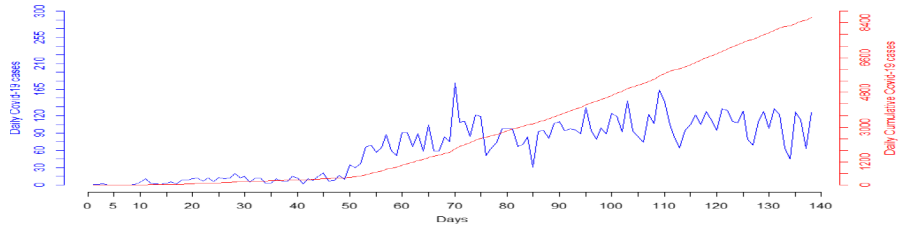


Fig. 2. Evolution of infected cases

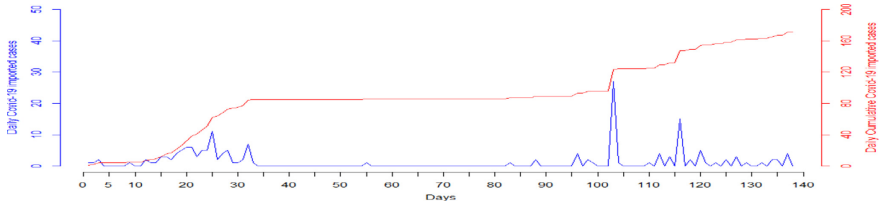


Fig. 3. Evolution of imported cases

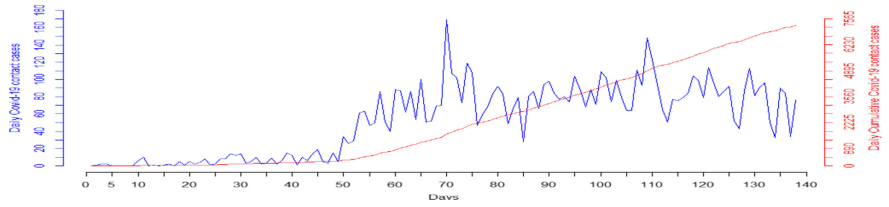


Fig. 4. Evolution of contact cases

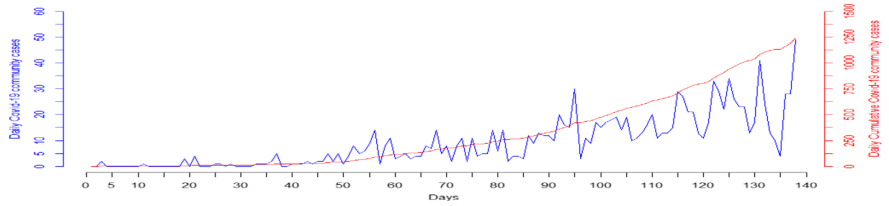


Fig. 5. Evolution of community cases

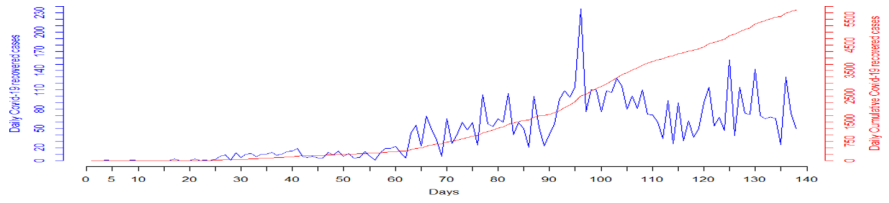


Fig. 6. Evolution of recovered cases

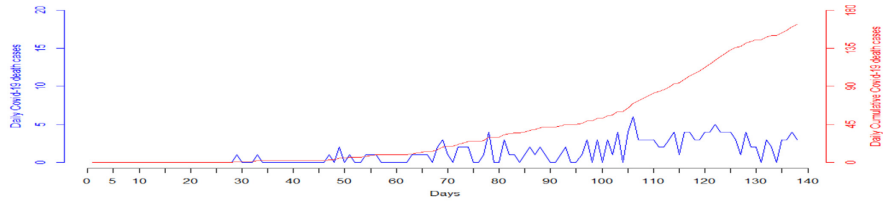


Fig. 7. Evolution of the death cases

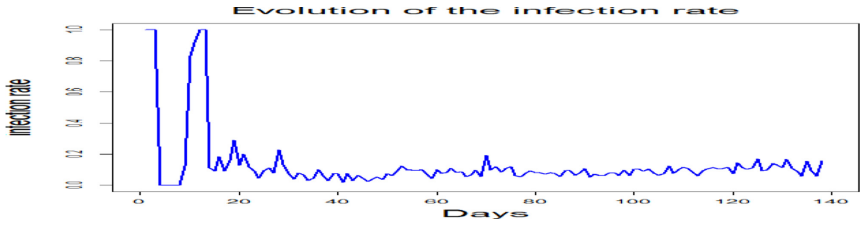


Fig. 8. Evolution of infection cases

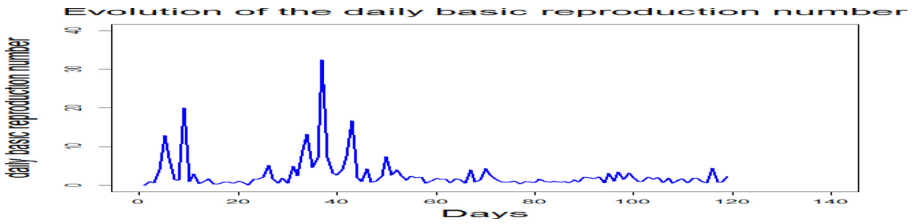


Fig. 9. Evolution of daily basic reproduction rate. The average basic reproduction rate is 2.72

We computed the rate q of the population that can be infected at the end of the covid-19 in Senegal and observed that this rate could be 90.54% of the total population.

4 Conclusion

In this study, we collected data from the different test centers of the covid-19 disease in Senegal. With those data, we have built a database that allowed us to carry out the statistical analysis. Results coming from the statistical analysis helped the decision-makers to evaluate the dynamics of covid-19 infection. During 138 days, corresponding to the 138 first days of the covid-19 presence disease in Senegal, we observed that imported cases of covid-19 had a high impact on the growth of the infected cases. We also observed that when the community infected cases and contact infected cases increase at the same time, we had exponential growth of infected cases. Due to the difficulty of the Ministry of Public Health and Government to evaluate the real impact of each barrier measure, the graphics that we have drawn could help to estimate the effects of each type of measure barrier (mask, social distancing, hand washing). With the statistical results, the Ministry of public health could evaluate the efficiency of barrier measures during the first 138 days of the presence of the covid-19 in Senegal.

References

1. Ross, R.: An application of the theory of probabilities to the study of a priori pathometry. Part I. Proc. R. Soc. London **92** (638) (1916). https://doi.org/10.1098/rspa.1916.0007
2. Covid19 Cases number. http://www.sante.gouv.sn/Actualites/
3. Diekmann, O., Heesterbeek, J.A., Metz, J.A.: On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**(4), 36582 (1990). ISSN 0303-6812. https://doi.org/10.1007/BF00178324
4. Li, X.Q., Cuomo, R., Purushothaman, V.: Data mining and content analysis of the Chinese social media platform weibo during the early COVID-19 outbreak: retrospective observational infoveillance study. JMIR Public Health Surveill. **6**(2), e18700 (2020). PMID: 32293582 PMCID: PMC7175787 https://doi.org/10.2196/18700

5. Jahanbin, K., Rahmanian, F., Rahmanian, V., Sotoodeh, J.A.: Application of Twitter and web news mining in infectious disease surveillance systems and prospects for public health. GMS Hyg. Infect. Control **14**, 1–12 (2019)
6. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al.: Clinical features of patients infected with novel coronavirus in Wuhan. China. Lancet (2019). https://doi.org/10.1016/S0140-6736(20)30183-5
7. Nishiura, H., Jung, S.M., Linton, N.M., Kinoshita, R., Yang, Y., Hayashi, K., et al.: The extent of transmission of novel coronavirus in Wuhan, China. J Clin Med. (2020). https://doi.org/10.3390/jcm9020330
8. Majumder, M., Mandl, K.D.: Early transmissibility assessment of a novel coronavirus in Wuhan, China. SSRN (2020). https://doi.org/10.2139/ssrn.3524675
9. Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F.: A novel coronavirus outbreak of global health concern. Lancet. **395**(10223), 470–473 (2020). [Europe PMC free article] [Abstract] [CrossRef] [Google Scholar]. https://doi.org/10.1016/S0140-6736(20)30185-9
10. Guellil, I., Boukhalfa, K.: Social big data mining: a survey focused on opinion mining and sentiments analysis. In: Conference of ISPS 2015: 12th International Symposium on Programming and Systems, Algiers. IEEE (2015). https://doi.org/10.1109/ISPS.2015.7244976