An in-depth statistical analysis of the COVID-19 pandemic's initial spread in the WHO African region

Ananthu James [1, 9 *] (ananthu89jms@gmail.com), Jyoti Dalal [9 *] (jyotidalal20@gmail.com), Timokleia Kousi [2, 9 *] (Timokleia.Kousi@etu.unige.ch), Daniela Vivacqua [3, 9] (danivivacqua@gmail.com), Daniel Cardoso Portela Câmara [4, 5, 9] (dcpcamara@gmail.com), Izabel Cristina dos Reis [4, 5, 9] (izabio2005@gmail.com), Sara Botero-Mesa [2, 9] (sara.botero@etu.unge.ch), Beat Stoll [2] (Beat.Stoll@unige.ch), Cristina Barroso Hofer [3, 9] (cbhofer@hucff.ufrj.br), Flávio Codeço Coelho [6, 9] (fccoelho@fgv.br), Olivia Keiser [2, 9 **] (olivia.keiser@unige.ch), Jessica L Abbate [7, 8, 9 **] (jessie.abbate@gmail.com)

1. Department of Chemical Engineering, Indian Institute of Science, Bengaluru, India
2. Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland
3. Department of Pediatric Infectious Diseases, UFRJ, Rio de Janeiro, Brazil
4. Laboratório de Mosquitos Transmissores de Hematozoários - LATHEMA, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
5. Núcleo Operacional Sentinela de Mosquitos Vetores - NOSMOVE, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil
6. School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, Brazil
7. UMI TransVIHMI (Institut de Recherche pour le Développement, Institut National de la Santé et de la Recherche Médicale, Université de Montpellier), Montpellier, France
8. Geomatys, Montpellier, France
9. The Global Research and Analysis for Public Health (GRAPH) Network (https://thegraphnetwork.org), Association Actions en Santé, Geneve, Switzerland

Corresponding author: Ananthu James (ananthu89jms@gmail.com)

[*] first authorship

[**] Equal contribution

Abstract:

Objective: To quantify the initial spread of COVID-19 in the WHO African region, and to investigate the possible drivers responsible for variation in the epidemic among member states.
Design: A cross-sectional study.
Setting: COVID-19 daily case and death data from the initial case through 29 November 2020.
Participants: 46 countries comprising the WHO African region.
Main outcome measures: We used five pandemic response indicators for each country: speed at which the pandemic reached the country, speed at which the first 50 cases accumulated, maximum monthly attack rate, cumulative attack rate, and crude case fatality ratio (CFR). We studied the effect of 13 predictor variables on the country-level variation in them using a principal component analysis, followed by regression.
Results: Countries with higher tourism activities, GDP per capita, and proportion of older people had higher monthly ($p < 0.001$) and cumulative attack rates ($p < 0.001$) and lower CFRs ($p = 0.052$). Countries having more stringent early COVID-19 response policies experienced greater delay in arrival of the first case ($p < 0.001$). The speed at which the first 50 cases occurred was slower in countries whose neighbors had higher cumulative attack rates ($p = 0.06$).
Conclusions: While global connectivity and tourism could facilitate the spread of airborne infectious agents, the observed differences in attack rates between African countries might also be due to differences in testing capacities or age distribution. Wealthy countries managed to minimize adverse outcomes. Further, careful and early implementation of strict government policies, such as restricting tourism, could be pivotal to controlling the COVID-19 pandemic. Evidently, good quality data and sufficient testing capacities are essential to unravel the epidemiology of an outbreak. We thus urge decision-makers to reduce these barriers to ensure rapid responses to future threats to public health and economic stability.

Summary Box:

What is already known on this topic

- The diversity in the COVID-19 trajectories across Africa.
- The heterogeneity in the reporting and testing capacities among countries.

What this study adds

- The heterogeneities in the attack rates during the first wave of COVID-19 in Africa could have emerged due to tourism and connectivity, the variations in age distribution, and the testing facilities between nations. Although richer countries, having a higher fraction of older people and more tourism activities, documented higher attack rates, they succeeded in reducing mortalities.
- African countries that had more stringent early COVID-19 response policies managed to delay the onset of the outbreak at the national level. Moreover, the speed at which the first 50 cases occurred was slower in nations whose neighbors had higher cumulative attack rates.
- Good quality data and timely reporting are essential in understanding and controlling a public health crisis.

Keywords: Africa; COVID-19; initial spread; attack rates; CFR; GDP; age distribution; testing capacity; stringency

Word count: ~3,600

Introduction

The first confirmation of a COVID-19 case in the African continent occurred in Egypt on 14 Feb 2020 (Massinga et al., 2021). Following that introduction, along with others elsewhere in the continent, COVID-19 cases and related fatalities rose exponentially, eventually reaching all African countries. Whether due to its younger populations, lower rates of obesity and other comorbidities, higher rates of immune-modulating parasitic diseases, relatively low population densities and urbanization, a disproportionate lack of data on confirmed cases and deaths, or some other feature unique to the African diaspora, countries on the continent appear to have fared better during the initial wave of the pandemic than elsewhere in the world, with lower attack rates and many orders of magnitude fewer deaths (Kousi et al., 2021; Wolday et al., 2021; Chitungo, 2020; Zahid et al., 2021; Rice et al., 2021). While the as-yet-undetermined cause of this difference in confirmed COVID-19 cases and deaths setting African countries apart could be common across the continent, there was nevertheless heterogeneity in the evolution of the epidemic between countries (Twahirwa Rwema et al., 2020). It is important to understand the factors that may contribute to the differences among these countries, as these insights can inform policy and priorities for future public health emergencies, when comparison to the rest of the world is marred by problematic biases (Salyer et al., 2021).

Motivated by this objective to examine drivers of heterogeneous COVID-19 spread across African countries, we quantified how the first wave of the COVID-19 pandemic unfolded in the 47 member states comprising the World Health Organization (WHO) African region using data scraped from official country health ministry announcements and published daily on the WHO COVID-19 dashboard (World Health Organization Dashboard, 2020). Following this, we performed a statistical analysis using policy related characteristics during the pandemic as well as largely pre-pandemic socioeconomic, and demographic aspects specific to each country.

Methods

Study design and settings

This is a cross-sectional analysis of the COVID-19 data reported as of 29 Nov 2020, among the 47 member states comprising the WHO African region. We analysed the data gathered by the WHO on the daily number of new cases and deaths published by country health ministries through official channels. These data eventually came to be systematically published on the WHO COVID-19 dashboard (World Health Organization Dashboard, 2020), and are thus freely accessible. However, groups of confirmed cases marked as ‘probable’ due to sole availability of

rapid diagnostic (viral antigen) tests - representing only 70 cases from Comoros and one case from the Democratic Republic of Congo (DRC) - are now also considered part of the official counts and were included as such in the present study. We included the cases with information about patient outcome status (alive, recovered, or dead), excluding those with missing values. Because of the lack of data reported from Tanzania, our analysis focused on the remaining 46 countries (Figure 1).

Variables

The response indicators: For each of the 46 countries, we included the five response indicators (Fig. 2A). They described not only the severity/burden of the outbreak but the evolution of the pandemic in the context of each country as well as the whole WHO African region, as explained below. 1) The per-capita cumulative number of cases (cases per million inhabitants), 2) the maximum monthly attack rate (new cases per million inhabitants counted over consecutive 4-week intervals), 3) the crude case fatality ratio (CFR, cumulative number of deaths / cumulative number of cases), 4) the speed of the epidemic within each country, measured as the inverse of the time interval between the first case and the 50th case,

$$speed_first50cases = \frac{1}{Day_50th_case - Day_first_case}$$

and 5) the speed at which the epidemic reaches each country, following its first confirmation in the region, measured as the inverse of the time delay between the first reported case in the region (on 25 February 2020 in Algeria) and the first case in each country, denoted by *start_delay*:

$$start_speed = \frac{1}{1 + start_delay}$$

In the above expression, the addition of 1 in the denominator ensures that we do not incur in a division by zero. The within-country speed of the outbreak was *a priori* expected to correlate with population density. Our final response indicator, concerning the regional correlation, was motivated by Li et al., 2021, representing a delay in either the exposure, testing capacity, data reporting, or some combination thereof.

Predictor variables: To explain the variation in the response variables between countries, we collected 13 predictor variables (Fig. 2B) from public data repositories (see Table S3 and the Supplementary Information (SI)): Gross Domestic Product (GDP) per capita, fishing volume per capita, proportion of children, fraction of males, population density, urban population density (urbanization), latitude, attack rate in neighbouring countries, per capita money generated from tourism, number of tourists arriving per capita, stringency index at the time of reporting of the first case, difference between the mean value of stringency index (Hale et al., 2021), corresponding to the study period, and its minimum, and the level of preparedness in a country

(World Health Organisation, 2005). For every country, we used the most recent values available for every parameter we measured (indexmundi, 2020; worldbank, 2020). (For a few of our studied countries, there was no recent available data for some of our indicators. In that case, we used the latest available value as we hypothesized it would better depict the situation of the country.)

Statistical Analysis

Missing data and imputation: We identified 11 missing values in our dataset corresponding to variables: Tourism_arrivals_percap, Stringency_mean, Stringency_day1, across six countries. To preserve these in our analysis, we used data imputation, substituting the missing values corresponding to the explanatory variables: Tourism_arrivals_percap and Stringency_mean with the corresponding median values. For the Stringency_day1, we substituted the missing values with the value computed using

$$Stringency_day1 \ = \ [median(\sqrt{Stringency_day1})]^2.$$

(In the analysis involving countries having at least 10 deaths, even the mean stringency was imputed like this.) Moreover, we used the OWID (Our World in Data, 2021) data to compute the neighbors_attack_rate for the countries we did not have data regarding the total number of positive cases.

Transformations: Following data imputation, we log-transformed skewed variables: Tourism_arrivals_percap, Tourism_dollars_percap, neighbors_attack_rate, GDP_percap, Fishing_volume_percap, and Pop_density, so that they could be approximated by normal distributions. We used square root transformation for the Stringency_day1. (In the analysis with only the countries having at least 10 deaths, the mean stringency (Stringency_mean) also was transformed like this.) We also transformed the response variables so that they would follow a normal distribution. We log-transformed the cases_per_million, max_monthly_attackrate, and start_speed, whereas speed_first50cases was square root transformed and CFR was arcsine transformed.

The Principal Component Analysis (PCA): We performed the PCA on our explanatory variable dataset with two objectives in mind: 1) reducing the number of original variables (from 13) and 2) to discover if the 46 analyzed African countries can be classified according to a composite profile of socioeconomic, geographic, and epidemiological dimensions (Borcard et al., 2011) (see Fig. 2C; correlation matrix plot), scaling variables to unit variance (using function "PCA" from the *FactoMineR* package; Le et al., 2008). The resulting informative Principal Component (PC) dimensions were selected following the Kaiser criterion (eigenvalue > 1; Kaiser, H. E, 1960). PCA loadings were used to verify the direction and contribution of each variable when interpreting the resulting dimensions. The scores of the resulting informative dimensions were used as explanatory variables in regression analyses.

Regression analysis: We performed a linear regression to understand the dependence of response indicators on the five composite explanatory variables (PCA dimensions 1-5 for the main analysis and the analysis to study the robustness). For each response indicator, we used the glmulti function from the glmulti package (Calcagno and de Mazancourt, 2010) to select the best fitting model among those including 0-5 of the explanatory PCA dimensions. We have provided the full summary tables corresponding to the best fitting regression model in the Supplementary Information.

Data were processed using both Python (version 3.7.10) and R (version 4.0.2) (https://www.r-project.org/). All statistical analyses were performed in R (version 4.0.2).

Robustness to variation in data quality: In order to check the robustness of our results to data quality issues, we assumed countries with fewer than 10 reported deaths as of 29 Nov 2020 to be suspected of under-detection or under-reporting of cases. Thus, we repeated the above analyses after removing the four countries that had reported fewer than 10 deaths (i.e., Burundi (deaths = 1), Comoros (deaths = 7), Eritrea (deaths = 0), and Seychelles (deaths = 0)) to check for qualitative differences in the results.

Patient and public involvement

Neither patients nor the public were involved in the design, conduct, reporting, or dissemination plans of our research.

Results

An overview of the response indicators included

The COVID-19 response variables (before transformation) showed substantial variation across the 46 countries (see Figures 2A and 4A). The cumulative attack rates (cases_per_million) ranged from 58.97 in Burundi to 19140.62 in Cabo Verde (median: 886.67). Maximum monthly attack rate (max_monthly_attackrate) varied from 14.83 in Burundi to 4843.33 in South Africa (median: 247.48). CFR ranged from 0.00% in Seychelles and Eritrea to 6.00% in Chad (median: 1.76%). Speed of the occurrence of first 50 cases at the country-level ($speed_first50cases$) varied from an average of 0.01 (equivalent of 0.49 cases per day) in Gambia to 0.125 (or 0.061 cases per day) in Mauritius (median: 0.19), and the speed at which the pandemic reached each country, following its first confirmation in the region ($start_speed$), ranged from 0.013 (delay of 75 days) in Lesotho to 1.00 in Algeria (median: 0.048).

The main analysis

The reduced list of predictors (principal components)

With the help of a PCA on the 13 predictors (Figure 2B-C), we reduced the number of dimensions to five, with all of them having eigenvalue > 1 and accounted for 75.7% of the variance of the dataset. Figure 3 shows the five PC dimensions, the percentage of variance explained by them, their descriptions, and correlations with the predictors.

The first dimension, PC1, accounted for 29.4% of the explained variance. PC1 was higher in countries with higher income and visitors from tourism, higher GDP, and generally older populations ($p < 0.001$ for all). The second dimension (PC2) accounted for 14.9% of the total variance and was significantly positively correlated with urbanization and the proportion of males, and negatively with overall population density ($p < 0.001$ for all). PC3 accounted for 14% and was negatively correlated with the cumulative attack rate in neighboring countries ($p < 0.001$). PC4 accounted for 9.6% of the explained variance. It was positively correlated with epidemic preparedness and negatively correlated with initial government response stringency ($p < 0.001$ for both). PC5 accounted for 7.9% of the total variance. It varied positively with per-capita fishing volume and negatively with the proportion of males and population density, though these correlations were not significant ($p > 0.05$).

The main predictors associated with the response indicators

Our regression suggested a negative relationship ($p = 0.052$) between CFR and PC1 (composed of high GDP, high tourism, and also higher proportion of older individuals). Both cases per million (cases_per_million) ($p < 0.001$) and monthly attack rate (max_monthly_attackrate) ($p < 0.001$) were positively correlated with PC1 (composed of high tourism, high GDP, older age). On the other hand, monthly attack rate ($p = 0.086$) showed a non-significant negative relationship with PC3, implying that it was positively associated with the cumulative attack rate of neighbouring countries. The relative indicator $start_speed$ showed a positive correlation with PC2 (high urbanization) ($p = 0.016$) and PC4 (higher preparedness and lower initial stringency) ($p < 0.001$). The $speed_first50cases$showed a positive trend with PC3 (lower neighbour attack rate) ($p = 0.06$; Tables 1 and S1).

The geographic distributions of the response variables as well as the PC dimensions can be seen in Figure 4.

The analysis following the removal of countries with <10 deaths

The principal components

In this case, PC1 remained the same as the main analysis, whereas PC2 was positively correlated with urbanization ($p < 0.001$) and the proportion of males ($p < 0.001$). Further, PC3 was negatively correlated with cases per million from neighbours and positively correlated with population density ($p < 0.001$). PC4 was negatively associated with the stringency index on the first day ($p < 0.001$). PC5 showed a negative correlation with fishing volume and preparedness and a positive association with proportion of males, with none of the correlations being significant ($p > 0.05$) (Figure S1).

The main predictors associated with the response indicators

As before, CFR remained negatively associated with PC1 (tourism, GDP, and older age) ($p = 0.02$). The monthly and cumulative attack rates were positively related to PC1 ($p < 0.001$). The speed of first case arrival($start_speed$)was associated positively with PC2 ($p = 0.042$) and PC4 ($p < 0.001$) and showed a non-significant negative trend with PC5 ($p = 0.085$). Therefore, countries in which the first infection happened quickly had less stringent restrictions that time, and also had high urbanization and male proportion. The within-country speed of the pandemic, speed to first 50 cases ($speed_first50cases$), was positively related to PC3 ($p < 0.001$), high population density and low infections in neighbors (see Table S2).

Discussion

Major findings and trends

This study considered globalisation and geographic, demographic, and socioeconomic factors as potential determinants of the spread and severity of COVID-19 in 46 WHO AFRO countries. Our two primary findings were the following: 1) the strong positive associations of tourism, GDP, and a high fraction of older people with the attack rates; 2) the negative relationship between tourism/GDP/older age and CFR. The influence of air travel, connectivity, and tourism in the spread (Rice et al., 2021; Kousi et al., 2021) and mortality (Pana et al., 2020) associated with COVID-19 infection in sub-Saharan Africa (SSA) is well documented in the literature. This finding may underline the importance of the design, and early implementation of surveillance and restriction measures at the points of entry, including appropriate screening and isolation policies (Pana et al., 2020), since these could result in lower infection rates (Pana et al., 2020; Sun et al., 2020; Farzanegan et al., 2020). Our findings could also be explained by the heterogeneity in the testing capacity between countries (Kousi et al., 2021), which could lead to underreporting of cases. Besides, policies to test only at the onset of the symptoms in quarantined individuals could further misinterpret the actual community spread (Skrip et al., 2020). It is unclear whether the low reported cases represent an actual decreased epidemic risk or

more asymptomatic infections (Abraha et al., 2021). In fact, COVID-19 severity being higher in older populations (Abraha et al., 2021) could be another reason behind detecting more cases from countries having a higher fraction of older people. African countries were also reported to have a lower number of COVID-19 infections among children than adults (Velásquez et al., 2021), in accordance with our finding. Moreover, previous research showed that countries with higher GDP per capita have higher life expectancy at birth, which leads to a higher percentage of older people, leading to the correlation between GDP and age observed by us (Miladinov, 2020). The lower CFR could be attributed to the better healthcare system, along with more patients with milder symptoms possibly being detected owing to more testing capacities, in richer countries (Skrip et al., 2020).

We found that urbanisation is positively associated with a faster occurrence of the first infection. It has been previously recorded that large, global cities reported positive COVID-19 cases at an earlier stage, due to their strong connectedness to other areas (OECD, 2020). Additionally, we found that the speed at which the first 50 infections happen in a country is positively related to its population density. Urban, dense areas usually produce the prerequisites for crowding, greater number of social interactions and close contact, which increases the risk for the spread of an infectious agent (IZA, 2020; Neiderud et al., 2015; Sigler et al., 2021).

We observed that countries having less stringent measures were the ones in which the COVID-19 outbreak happened quickly. As expected, early implementation of restriction and control measures probably resulted in delaying the onset of the outbreak. This result could help the authorities to implement the necessary measures to more effectively control the upcoming surge of cases and protect public health. Previous research showed that a common characteristic among countries with delayed onset was the implementation of effective border measures and various preparedness activities at an early stage (Kousi et al., 2020; Li et al., 2021; Ayouni et al., 2021). Conversely, countries that had a higher value of preparedness index were among the first ones in which the outbreak initiated. Perhaps, countries better prepared to control an epidemic had also enhanced testing capabilities, and as a result were able to detect and identify cases earlier. Indeed, a study focusing on the 24 SSA countries that were COVID-19 free as of March 30, 2020 documented that only 38% of them had COVID-19 testing capacities (Li et al., 2021). Although we found that stringent policies are helpful in controlling the pandemic, many of these measures were associated with other negative consequences (Katana et al., 2021; Bourgault et al., 2021). Therefore, strict government actions should be implemented carefully. Future studies may address the overall impact of such measures by considering more parameters like quality-adjusted life years (QALYs) and effects on mental health (Briggs et al., 2021; Mohler-Kuo et al., 2021; Singh et al., 2020; Brigdland et al., 2021).

An interesting finding was the negative association between the speed at which the first 50 cases happen in a country and the attack rate in neighbouring countries. It is possible that an outbreak

in one country would alert its neighbours to take early precautions to avoid the spread of the virus in their territory. These precautions could include stricter border control and decreased mobilisation of people, resulting in lower early cases. However, given the geographic complexity of the continent and the knowledge gap on the drivers of geographical diffusion between areas, further research is needed to examine this hypothesis in detail (Sigler et al., 2021).

Strengths and limitations

Our study is one of the most comprehensive studies describing the first wave of COVID-19 pandemic in Africa, where it is less studied. Using the 5 response variables and 13 predictors we analyzed, we believe that we managed to get a thorough understanding of the burden, severity, spatial trends, and the evolution of the pandemic in the WHO African region. The other strength of our study is the robustness of our estimates. The analysis conducted with countries having at least 10 deaths confirmed the roles of tourism, age distribution, and GDP on the attack rates and CFR. It also established that nations with strict governmental policies had a late onset of the COVID-19 outbreak and those having a high attack rate in neighbouring countries managed to reduce the speed at which the first 50 infections occur in them.

However, there exists heterogeneity across the countries in this continent, which limits our ability to extract general conclusions. One important limitation of this study is the lack of standardization in testing policies (Velasquez et al., 2021) and notification of cases between countries. Additionally, the availability of individual patient data corresponding to each predictor variable, where possible, is useful in obtaining a clearer insight of the pandemic situation. However, such studies are limited in the context of Africa and have addressed the impacts of a few individual-level factors only (Dalal et al., 2021). Similar to other regions, Africa was also subject to under-reporting of confirmed cases and deaths (Chitungo et al., 2020; Dyer, 2021).

We transformed many of the predictors as well as the response indicators so that they are well approximated by normal distributions. Hence, our results need to be interpreted carefully. Moreover, the number of tourists arriving as well as the stringency variables were missing for a few countries. We relied on imputation on such occasions. We also recognize that data was not reported consistently for all countries for some of the indicators. For instance, we noticed that the latest reported data on some of our variables was not documented the same year for all countries and was collected before the pandemic. For a few of them, the absolute value could change in the pandemic time, and the proportion by which the change happens could vary across countries. There are other potential variables not accounted for in our study such as air pollution, climatic variation, economic inequality (GINI index), diet, etc., that could affect the attack rates, severity, and speed corresponding to COVID-19 (Rice et al., 2021; Kim et al., 2021).

Conclusions

Important evidence can be extracted from the presented research, and inform decision makers on the factors to be considered when designing their plan to effectively and rapidly control a future outbreak in the African context. Rich countries, with high tourism activities, experienced a higher number of cases, indicating links between air travel, connectivity, human mobility, and viral transmission, along with suggesting the roles of better testing facilities and age structure. However, since most of these nations also had better healthcare, they managed to minimize the number of deaths, resulting in their low CFRs. Urbanised areas with high population density underwent a faster increase in the number of cases in the beginning of the outbreak. Also, countries with weaker control measures faced a COVID-19 outbreak quicker. These findings stress the need for appropriate non-pharmaceutical measures at an early stage, with emphasis on densely populated areas, and famous tourist destinations. Lastly, it is very important for the countries to ensure robust surveillance, and testing capacities, so that they can base their control strategy on data that depicts the epidemiological situation of the country correctly. Good quality of data will be helpful to better evaluate and adjust the various implemented measures during the course of the outbreak.

Contributions:
AJ conceived the study, with the support of JD, TK, JLA, and DV. AJ, JD, TK, DV designed the study, with the support of JLA, DCPC, and SBM. AJ, JD, TK, DV participated in weekly meetings and exchanged ideas that were helpful for the progress of the draft. JD acquired and cleaned the response variables data with the support of JLA and AJ. TK collected the predictor variable data with the support of JD, AJ, DV, and JLA. AJ performed the statistical analysis with the support of JD, JLA, and DCPC. JLA, DCPC, FCC provided suggestions on statistical analysis. JD, AJ made the plots with the support of JLA, DCPC, and ICR. TK, AJ conducted the literature review, with the support of JD and JLA. AJ, JD, TK interpreted the data, connected it to evidence from literature, and wrote the initial draft with the support of JLA, DCPC, and DV. AJ, JD, TK, DV wrote the initial draft with the support of JLA, DCPC. AJ, JD improved and finalized the draft with the support of JLA and TK. OK, FCC, CBH, BS critically revised the manuscript for important intellectual content and supervised the study. All authors contributed to final approval of the version to be submitted. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. AJ, JD, TK, JLA, and OK are the guarantors.

Funding: This study was funded by the Swiss National Science Foundation (grants no. 196270 and 163878).

Conflicts of Interest: The authors declare no competing interests.

Data Availability Statement: COVID-19 data used in this manuscript is available at https://github.com/JyotiDalal93/COVID-19_SSA/blob/main/Data_sub_saharan_Africa.zip. Moreover, data dictionaries, statistical code etc., that support the findings of this study are available upon reasonable request from the corresponding author.

Ethical Approval: Data was collected for surveillance purposes under IHR, and no ethical approval was needed.

Transparency statement: AJ, JD, TK, JLA, and OK affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Patient consent for publication: Not required.

Figures and tables (4 Figures and 1 Table):

Figure 1. The 46 countries from the WHO African region included in this study. We excluded Tanzania because of underreporting as the last case reported publicly was on 7 May 2020.

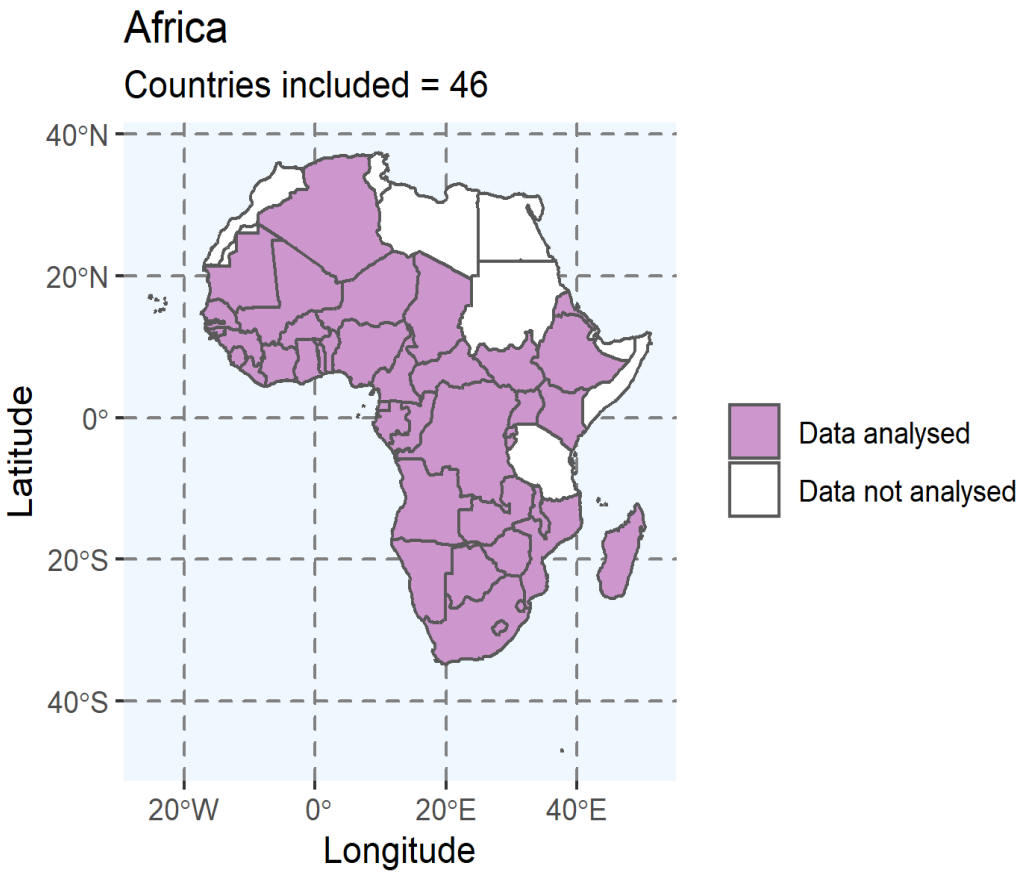
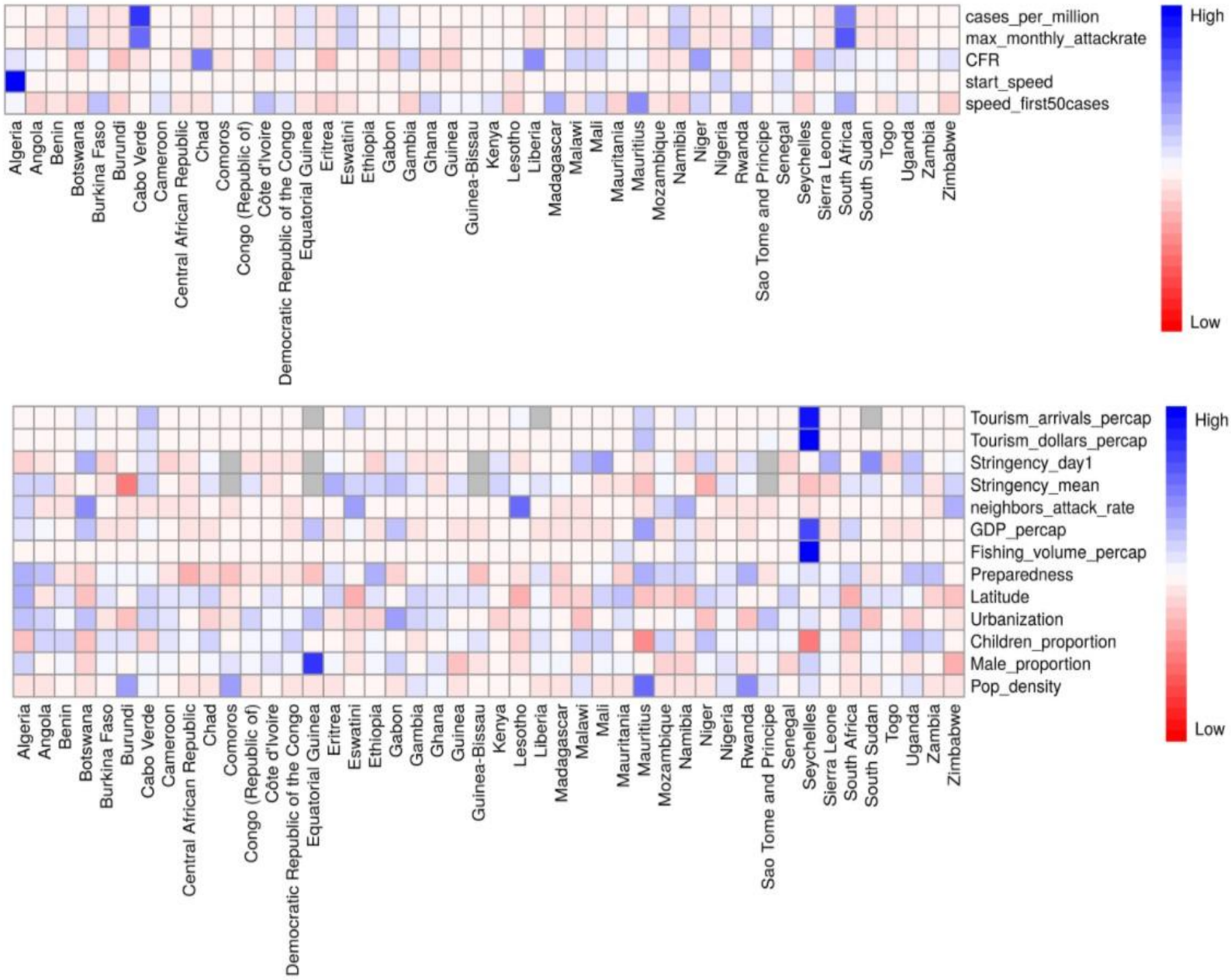


Figure 2. A: Heatmap showing all the 5 response indicators from the 46 WHO AFRO countries reporting cases. **B:** Heatmap showing the variation of 13 explanatory variables across the countries. The grey color corresponds to the missing entries. For both the heatmaps, blue color represents high values whereas red color denotes low values. Each indicator (in the sets of response and predictor variables) here is scaled by the standard deviation and centered by subtracting the mean before plotting. **C:** Correlation matrix depicting the predictor variables included in the PCA. So as to preserve the epidemiological significance, we did not exclude any predictor based on its strong correlation with other ones.


cases_per_million
max_monthly_attackrate
CFR
start_speed
speed_first50cases
High
Low
Tourism_arrivals_percap
Tourism_dollars_percap
Stringency_day1
Stringency_mean
neighbors_attack_rate
GDP_percap
Fishing_volume_percap
Preparedness
Latitude
Urbanization
Children_proportion
Male_proportion
Pop_density
High
Low
Algeria
Angola
Benin
Botswana
Burkina Faso
Burundi
Cabo Verde
Cameroon
Central African Republic
Chad
Comoros
Congo (Republic of)
Côte d'Ivoire
Democratic Republic of the Congo
Equatorial Guinea
Eritrea
Eswatini
Ethiopia
Gabon
Gambia
Ghana
Guinea
Guinea-Bissau
Kenya
Lesotho
Liberia
Madagascar
Malawi
Mali
Mauritania
Mauritius
Mozambique
Namibia
Niger
Nigeria
Rwanda
Sao Tome and Principe
Senegal
Seychelles
Sierra Leone
South Africa
South Sudan
Togo
Uganda
Zambia
Zimbabwe

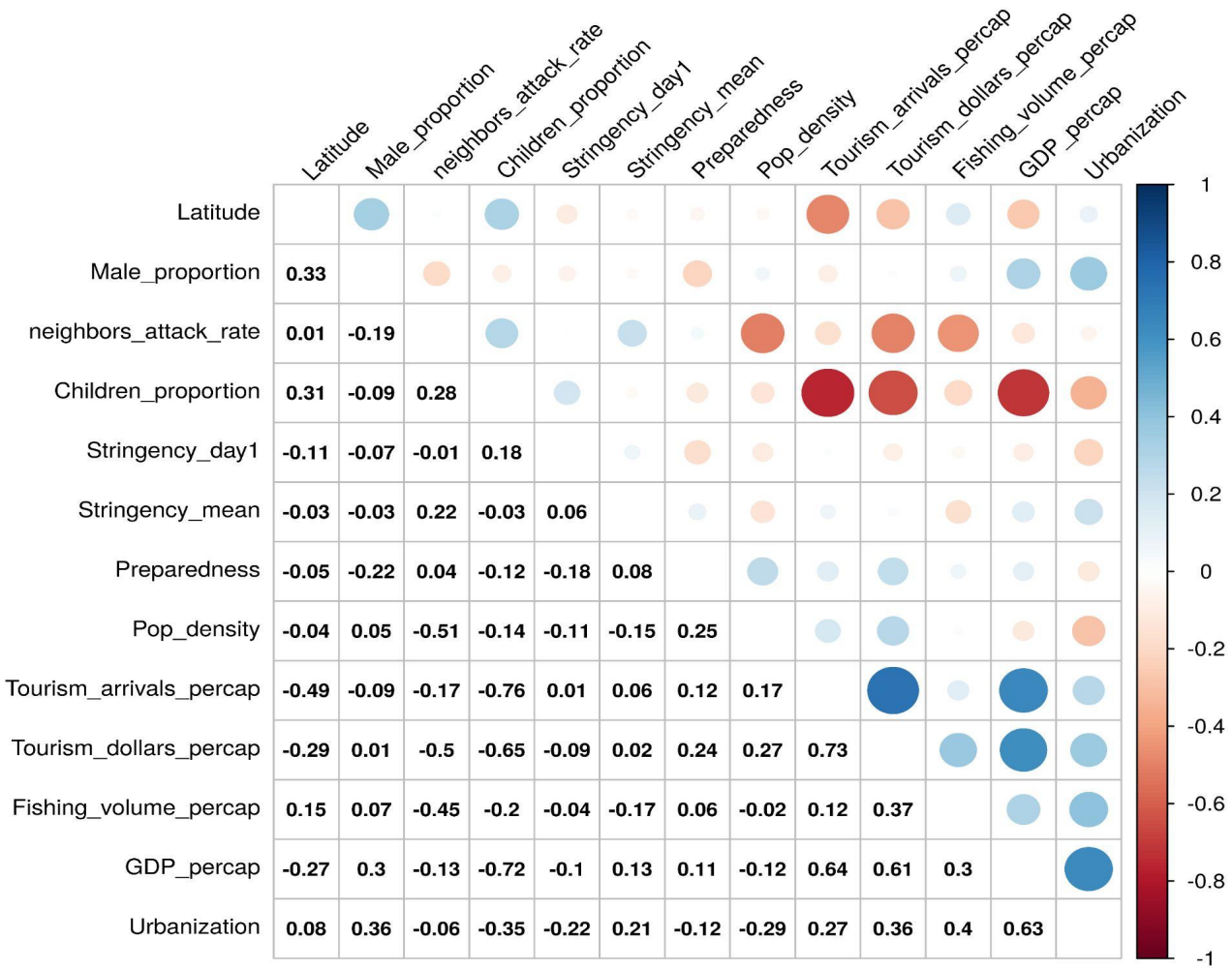
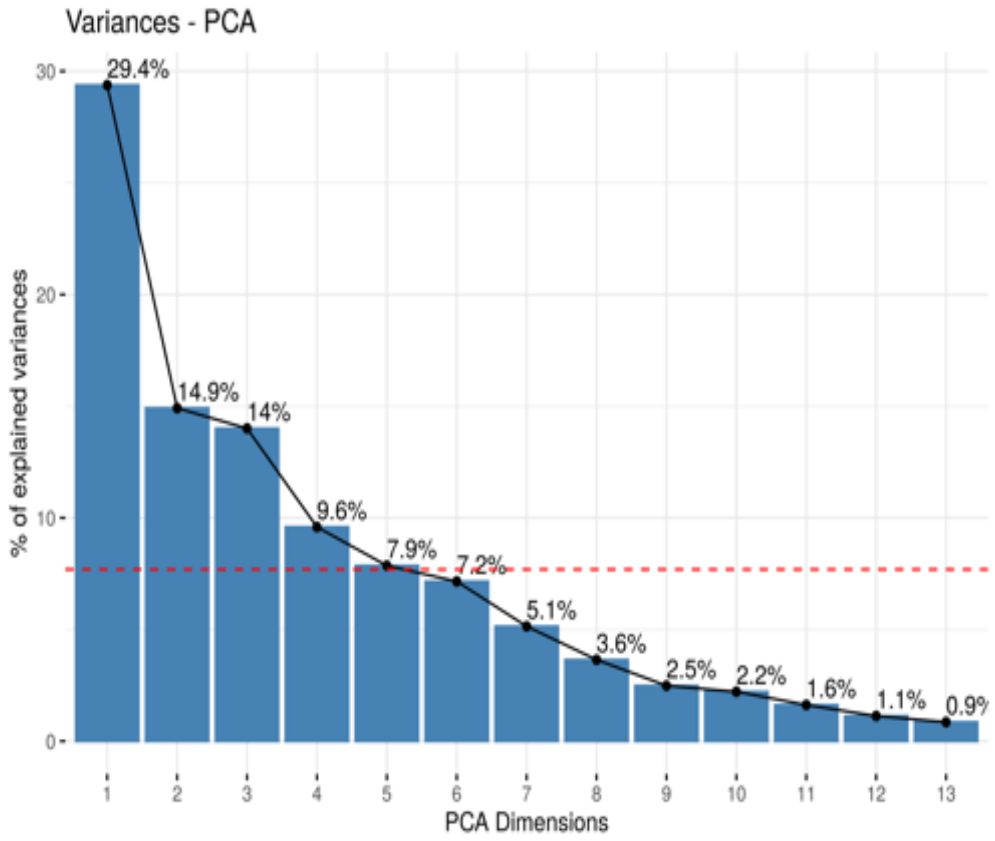
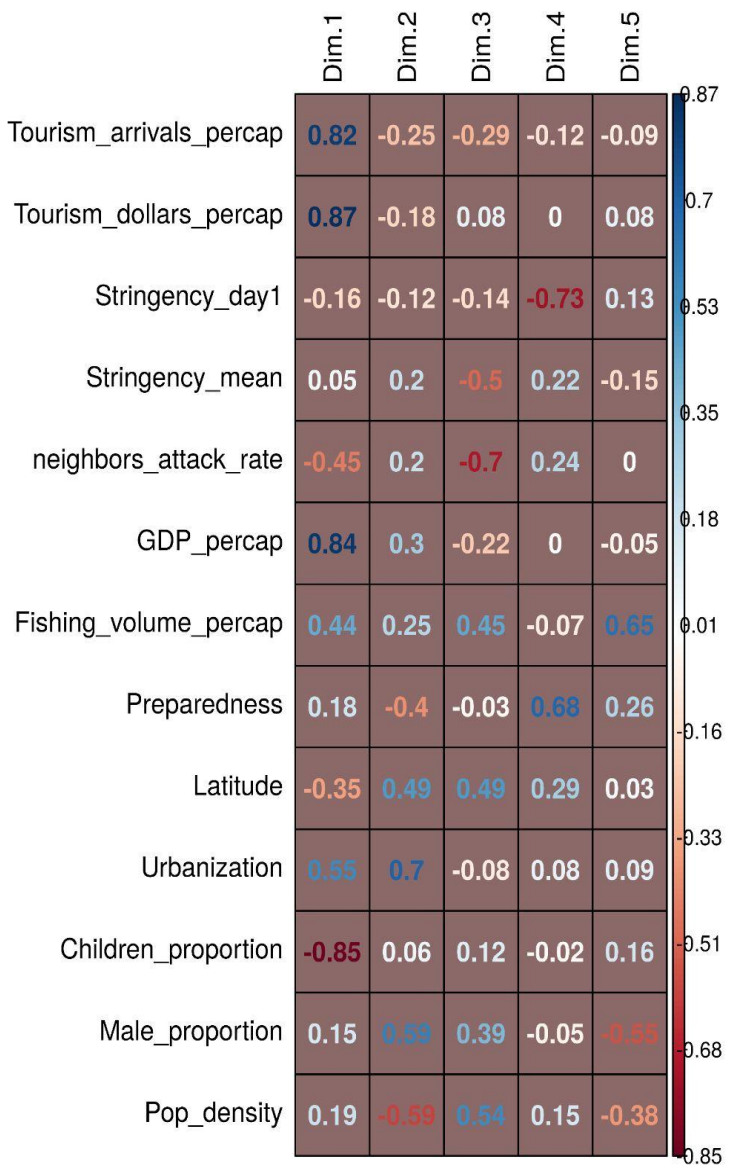


Figure 3. **A:** Scree plot depicting the percentage of variance explained by each PCA dimension. The red line differentiates the 5 principal components with eigenvalue > 1. **B:** Correlation matrix showing the predictor variables and their correlations with each PC dimension. **C:** Correlation matrix showing the contribution of the predictors to each PC.


Variances - PCA
% of explained variances
PCA Dimensions
29.4%
14.9%
14%
9.6%
7.9%
7.2%
5.1%
3.6%
2.5%
2.2%
1.6%
1.1%
0.9%
0
10
20
30
1
2
3
4
5
6
7
8
9
10
11
12
13


Dim.1
Dim.2
Dim.3
Dim.4
Dim.5
Tourism_arrivals_percap 0.82 -0.25 -0.29 -0.12 -0.09
Tourism_dollars_percap 0.87 -0.18 0.08 0 0.08
Stringency_day1 -0.16 -0.12 -0.14 -0.73 0.13
Stringency_mean 0.05 0.2 -0.5 0.22 -0.15
neighbors_attack_rate -0.45 0.2 -0.7 0.24 0
GDP_percap 0.84 0.3 -0.22 0 -0.05
Fishing_volume_percap 0.44 0.25 0.45 -0.07 0.65
Preparedness 0.18 -0.4 -0.03 0.68 0.26
Latitude -0.35 0.49 0.49 0.29 0.03
Urbanization 0.55 0.7 -0.08 0.08 0.09
Children_proportion -0.85 0.06 0.12 -0.02 0.16
Male_proportion 0.15 0.59 0.39 -0.05 -0.55
Pop_density 0.19 -0.59 0.54 0.15 -0.38
0.87
0.7
0.53
0.35
0.18
0.01
-0.16
-0.33
-0.51
-0.68
-0.85

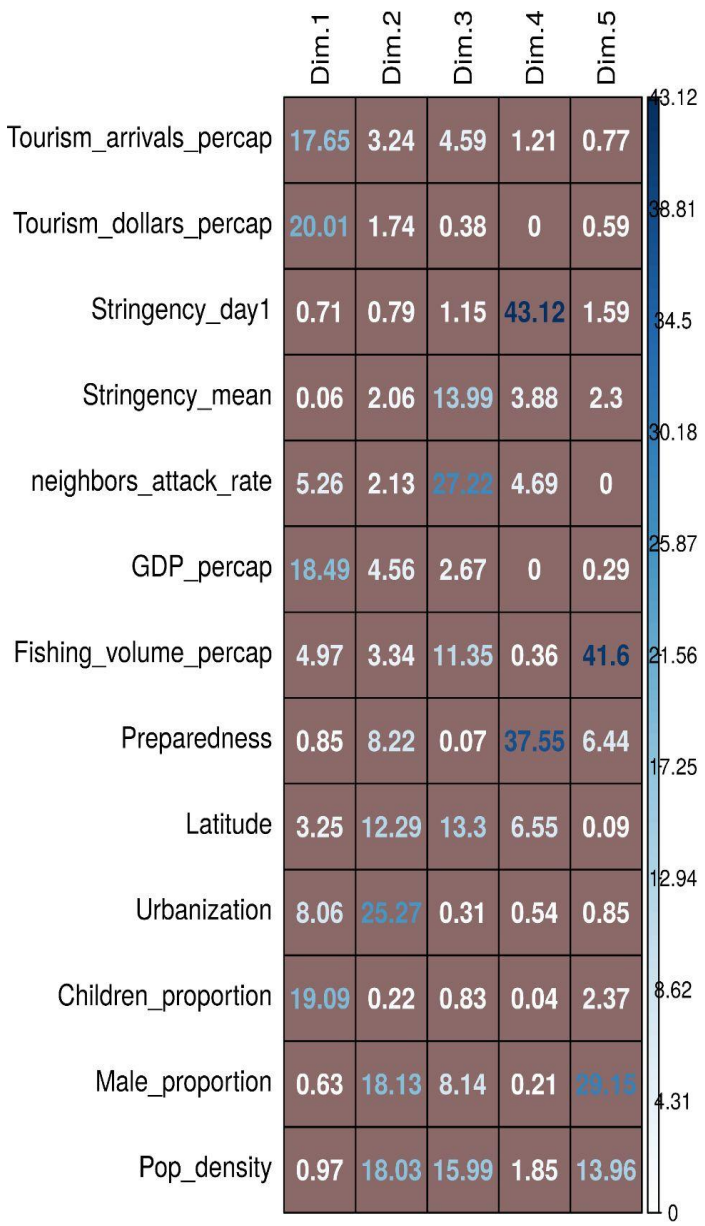
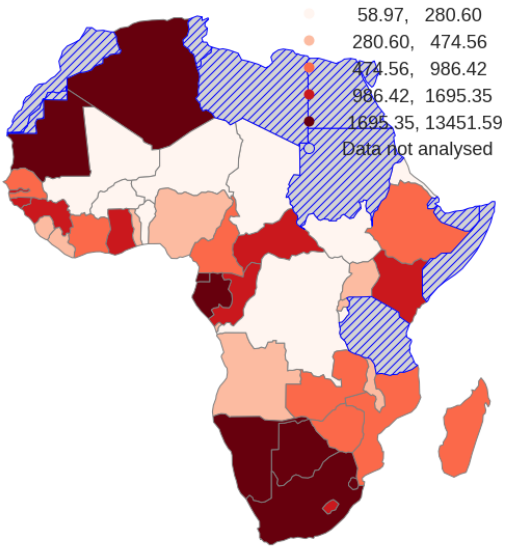
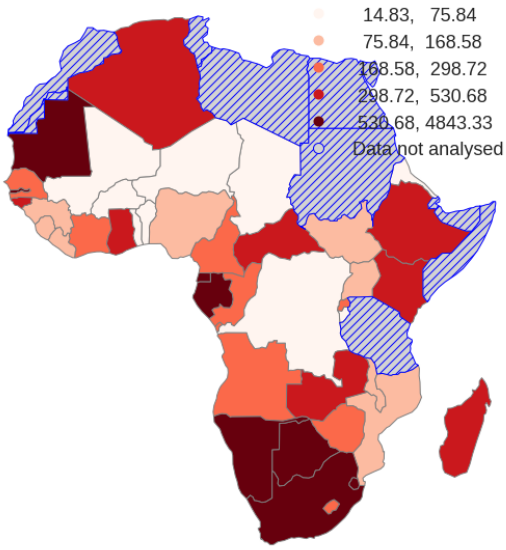


Figure 4. The geographic distributions of all the **A:** response variables and **B:** PC dimensions. The strong positive association between the attack rates (cases per million and maximum monthly attack rate) with PC1 is evident from the maps.

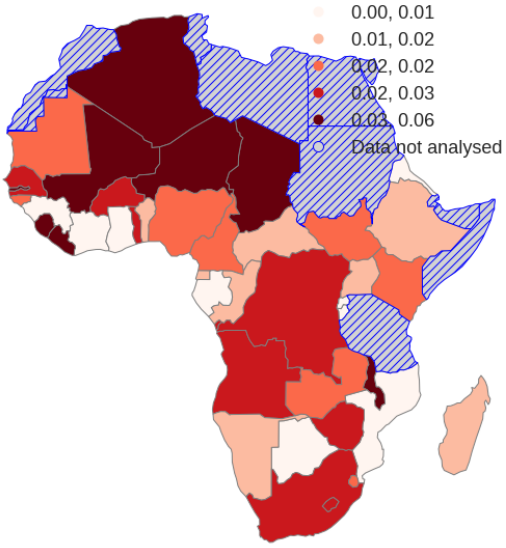
Cases per million



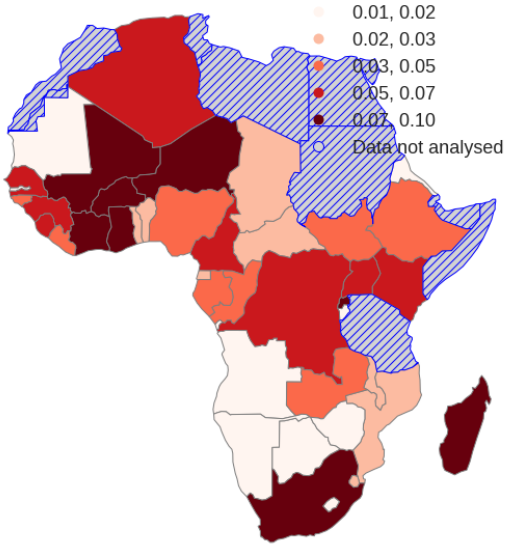
Max. Monthly Attack Rate



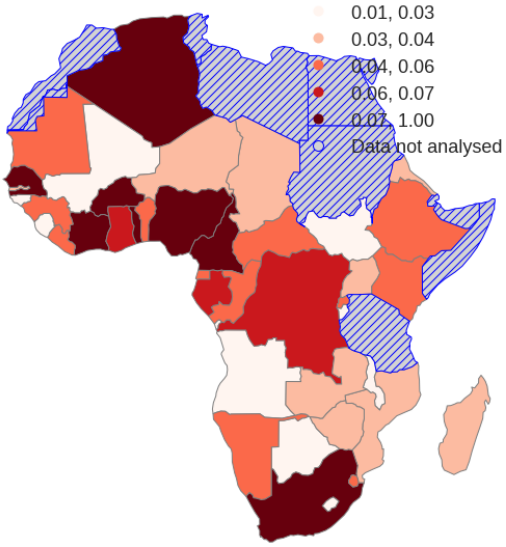
Case fatality rate



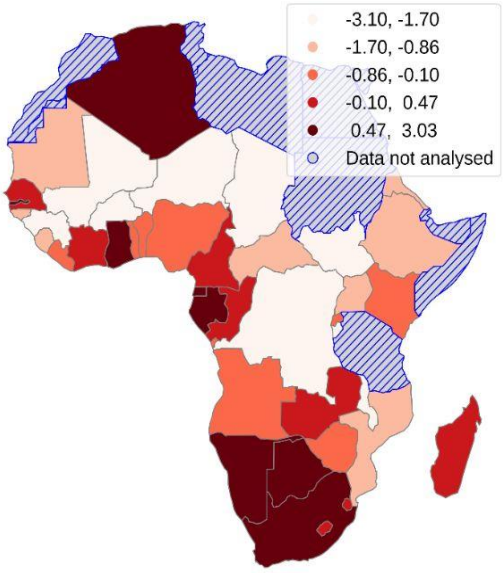
Speed of first case arrival



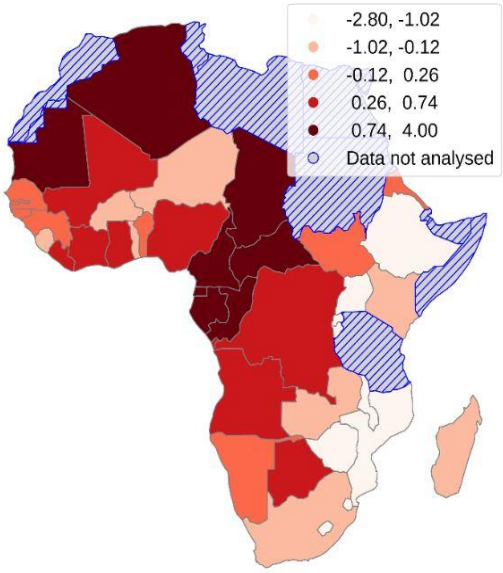
Speed of accumulation of first 50 cases



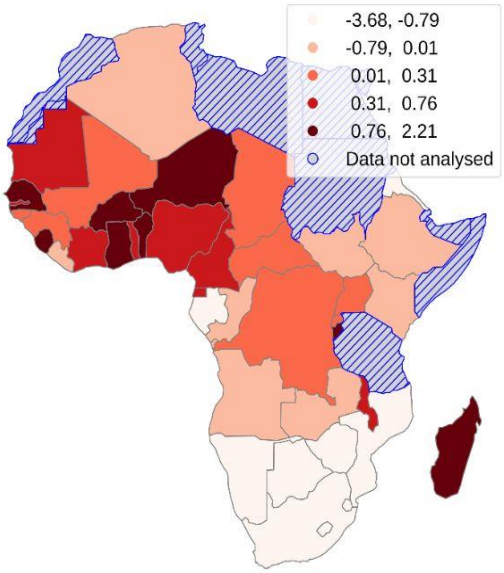
PC 1



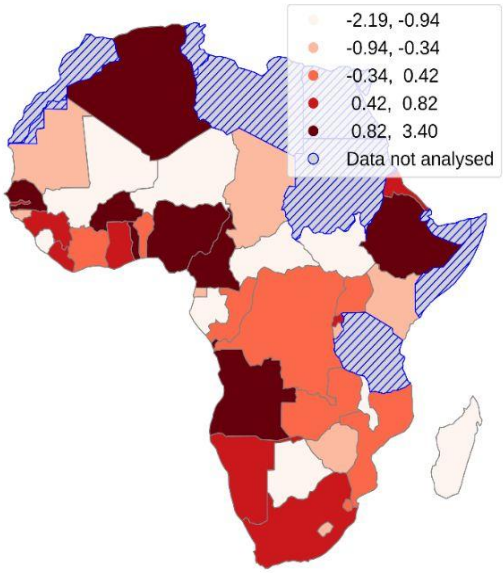
PC 2



PC 3



PC 4



PC 5

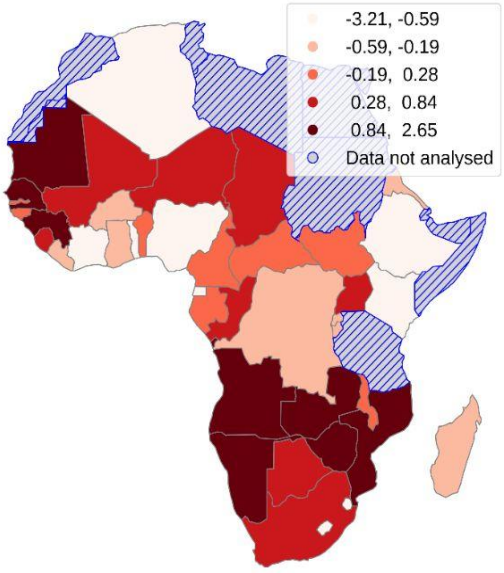


Table 1. *Main regression analysis:* Best fitting regression model along with their coefficients for each of the five response indicators. The significant codes: '***' for < 0.001, '**' for < 0.01, '*' for < 0.05, and '.' for <0.1.

	Response Indicators	Best Fitting Model	Coefficients		
1.	Cumulative Attack Rate (cases_per_million)	~ (PC1 + PC3)	Intercept 1931.38 ***	PC1 927.36 ***	PC3 -450.90
2.	Maximum Monthly Attack Rate (max_monthly_attackrate)	~ (PC1 + PC3)	Intercept 618.91 ***	PC1 302.55 ***	PC3 -158.41 .
3.	Case Fatality Ratio (CFR)	~ PC1	Intercept 0.13 ***	PC1 -0.0065 .	
4.	Speed of first case arrival ("start_speed")	~ (PC2 + PC4)	Intercept 0.0753 ***	PC2 0.031 *	PC4 0.070 ***
5.	Speed of accumulation of first 50 cases ("speed_first50cases")	~ (PC2 + PC3)	Intercept 0.0447 ***	PC2 -0.0041	PC3 0.0056 .

References:

1. Massinga Loembé M., Tshangela, A., Salyer, S.J. et al. COVID-19 in Africa: the spread and response. Nat Med 26, 999–1003 (2020). https://doi.org/10.1038/s41591-020-0961-x
2. Kousi, T., Vivacqua, D., Dalal, J., James, A., Câmara, D. C. P., Mesa, S. B., … Hofer, C. B. (2021, July 31). The COVID-19 pandemic in Africa's Island Nations: A descriptive study of the course of the COVID-19 outbreaks and the national response plans. https://doi.org/10.31219/osf.io/xmqac
3. Wolday, D., Gebrecherkos, T., & Arefaine, Z. G. (2021). Effect of co-infection with intestinal parasites on COVID-19 severity: A prospective observational cohort study. EClinicalMedicine, 39, 101054. https://doi.org/10.1016/j.eclinm.2021.101054
4. Chitungo I, Dzobo M, Hlongwa M, Dzinamarira T. COVID-19: Unpacking the low number of cases in Africa. Public Health Pract (Oxf). 2020;1:100038. doi:10.1016/j.puhip.2020.100038
5. Zahid MN, Perna S. Continent-Wide Analysis of COVID 19: Total Cases, Deaths, Tests, Socio-Economic, and Morbidity Factors Associated with the Mortality Rate, and Forecasting Analysis in 2020–2021. International Journal of Environmental Research and Public Health. 2021; 18(10):5350. https://doi.org/10.3390/ijerph18105350
6. Rice, B.L., Annapragada, A., Baker, R.E. et al. Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa. Nat Med 27, 2021, 447–453. https://doi.org/10.1038/s41591-021-01234-8
7. Twahirwa Rwema, O. J. et al. COVID-19 Across Africa: Epidemiologic Heterogeneity and Necessity of Contextually Relevant Transmission Models and Intervention Strategies. Ann Intern Med.2020;173:752-753. [Epub ahead of print 18 June 2020]. doi:10.7326/M20-2628
8. Salyer, S. J. et al. The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. The Lancet, 2021. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(21)00632-2/fulltext
9. World Health Organisation, WHO Coronavirus (COVID-19) Dashboard, Accessed on 30 November 2020. https://covid19.who.int/
10. Pana TA, Bhattacharya S, Gamble DT, Country-level determinants of the severity of the first global wave of the COVID-19 pandemic: an ecological study, BMJ Open 2021;11:e042034. doi: 10.1136/bmjopen-2020-042034
11. World Health Organisation. Guidance document for the state party self-assessment annual reporting tool. (2005) https://apps.who.int/iris/bitstream/handle/10665/272438/WHO-WHE-CPI-2018.17-eng.pdf?sequence=1
12. The World Bank. Datasets. Retrieved March 18, 2021, from https://data.worldbank.org/
13. IndexMundi. Country Facts. Retrieved March 18, 2021, from https://www.indexmundi.com/
14. Our World in Data. Datasets. Retrieved March 18, 2021, from https://ourworldindata.org/

15. Hale T., Angrist N., Goldszmidt R. et al. “A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker).” Nature Human Behaviour. 2021;5, 529–538. https://doi.org/10.1038/s41562-021-01079-8
16. Cao Y, Hiyoshi A, Montgomery S. COVID-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data BMJ Open 2020;10:e043560. doi: 10.1136/bmjopen-2020-043560
17. Borcard, D., Gillet, F., Legendre, P., 2011. Numerical ecology with R. Springer-Verlag, New York City, NY.
18. Li, Z., Jones, C., Ejigu, G.S. et al. Countries with delayed COVID-19 introduction – characteristics, drivers, gaps, and opportunities. Global Health 17, 28 (2021). https://doi.org/10.1186/s12992-021-00678-4
19. Le et al., FactoMineR: An R Package for Multivariate Analysis, J. Stat. Software, 2008
20. Calcagno and de Mazancourt, glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models, J. Stat. Software, 2010
21. Kaiser, H. E (1960). The application of electronic computers to factor analysis. Education & Psychological Measurement, 20, 14I-151
22. Gilbert M., Pullano G., Pinotti, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. Lancet (London, England), 2020, 395(10227): 871–877. https://doi.org/10.1016/S0140-6736(20)30411-6
23. Sun M.C., Lan C.B., Wah C.. Lessons to be learnt from the COVID-19 public health response in Mauritius, Public Health in Practice, 2020;1 https://doi.org/10.1016/j.puhip.2020.100023
24. Farzanegan, M. R., Gholipour, H. F., Feizi, M., Nunkoo, R., & Andargoli, A. E. International Tourism and Outbreak of Coronavirus (COVID-19): A Cross-Country Analysis. Journal of Travel Research, 2021:60(3): 687–692. https://doi.org/10.1177/0047287520931593
25. Skrip L., Selvaraj P., et al., Seeding COVID-19 across sub-Saharan Africa: an analysis of reported importation events across 48 countries, The American Journal of Tropical Medicine and Hygiene, 2021;104:5:1694–1702. https://doi.org/10.1101/2020.04.01.20050203
26. Abraha, H. E. et al., (2021). Clinical features and risk factors associated with morbidity and mortality among patients with COVID-19 in northern Ethiopia; https://www.sciencedirect.com/science/article/pii/S1201971221002551
27. Velásquez, S. R., Jacques, L., & Dalal, J. (2021). The toll of COVID-19 on African children: A descriptive analysis on the COVID-19-related morbidity and mortality among the pediatric population in Sub-Saharan Africa. International Journal of Infectious Diseases. Published. https://doi.org/10.1016/j.ijid.2021.07.060
28. Miladinov G., Socioeconomic development and life expectancy relationship: evidence from the EU accession candidate countries. Genus, 2020; 76: 2 . https://doi.org/10.1186/s41118-019-0071-0

29. OECD. (2021). OECD Regional Outlook 2021 Addressing COVID-19 and Moving to Net Zero Greenhouse Gas Emissions. OECD.
30. Sigler, T., Mahmuda, S., Kimpton, A. et al. The socio-spatial determinants of COVID-19 diffusion: the impact of globalisation, settlement characteristics and population. Global Health 2021;17,56. https://doi.org/10.1186/s12992-021-00707-2
31. IZA Institute of Labor Economics. (2020, July). Urban Density and COVID-19. http://ftp.iza.org/dp13440.pdf
32. Neiderud C. J. How urbanization affects the epidemiology of emerging infectious diseases. Infection ecology & epidemiology, 2015;5: 27060. https://doi.org/10.3402/iee.v5.27060
33. Ayouni, I., Maatoug, J., Dhouib, W. et al. Effective public health measures to mitigate the spread of COVID-19: a systematic review. BMC Public Health, 2021; 21: 1015. https://doi.org/10.1186/s12889-021-11111-1
34. Kousi T, Mitsi L-C, Simos J. The Early Stage of COVID-19 Outbreak in Greece: A Review of the National Response and the Socioeconomic Impact. International Journal of Environmental Research and Public Health. 2021; 18(1):322. https://doi.org/10.3390/ijerph18010322
35. Katana, E., Amodan, B.O., Bulage, L. et al. Violence and discrimination among Ugandan residents during the COVID-19 lockdown. BMC Public Health, 2021; 21: 467. https://doi.org/10.1186/s12889-021-10532-2
36. Center for Global Development. Violence Against Women and Children During COVID-19— One Year On and 100 Papers In A Fourth Research Round Up. 2021 https://www.cgdev.org/sites/default/files/vawc-fourth-roundup.pdf
37. Briggs, A., Vassall, A. Count the cost of disability caused by COVID-19. Nature, 2021; 593. https://media.nature.com/original/magazine-assets/d41586-021-01392-2/d41586-021-01392-2.pdf
38. Bridgland VME, Moeck EK, Green DM, et al. Why the COVID-19 pandemic is a traumatic stressor. PLoS ONE. 2021;16(1): e0240146. https://doi.org/10.1371/journal.pone.0240146
39. Singh, S., Roy, D., Sinha, K., Parveen, S., Sharma, G., & Joshi, G. Impact of COVID-19 and lockdown on mental health of children and adolescents: A narrative review with recommendations. Psychiatry research, 2020; 293: 113429. https://doi.org/10.1016/j.psychres.2020.113429
40. Mohler-Kuo M, Dzemaili S, Foster S, Werlen L, Walitza S. Stress and Mental Health among Children/Adolescents, Their Parents, and Young Adults during the First COVID-19 Lockdown in Switzerland. International Journal of Environmental Research and Public Health. 2021; 18(9):4668. https://doi.org/10.3390/ijerph18094668
41. Yamagishi, T., Kamiya, H., Kakimoto, K., Suzuki, M., & Wakita, T. Descriptive study of COVID-19 outbreak among passengers and crew on Diamond Princess cruise ship,

Yokohama Port, Japan, 20 January to 9 February 2020. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 2020; 25(23): 2000272. https://doi.org/10.2807/1560-7917.ES.2020.25.23.2000272

42. Kawther Aabed, Maha M.A. Lashin. An analytical study of the factors that influence COVID-19 spread, Saudi Journal of Biological Sciences, 2021; 28: 2: 1177-1195. https://doi.org/10.1016/j.sjbs.2020.11.067.
43. Quigley, A., Nguyen, P., Stone, H., Lim, S., MacIntyre, C. Cruise Ship Travel and the Spread of COVID-19 – Australia as a Case Study. International Journal of Travel Medicine and Global Health, 2020;9(1): 10-18. doi: 10.34172/ijtmgh.2021.03
44. Dyer O. Covid-19: Study claims real global deaths are twice official figures BMJ 2021; 373 :n1188. doi:10.1136/bmj.n1188
45. Dalal, J. et al. COVID-19 Mortality in Women and Men in Sub-Saharan Africa: A Cross Sectional Study; https://www.medrxiv.org/content/10.1101/2021.07.31.21261422v1