The article « nCov2019: An analytics tool for studying the COVID-19 coronavirus pandemic » is an article to present the R package nCov2019. The package intends to provide a way of accessing historical COVID19 data directly from R and to provide analysis tools in the form of a shiny application and plot functions. While I appreciate the effort and the work done, I have several strong critics, that I would summarize in the following points:

1. The main point is that I do not see the novelty of what is proposed here. Several historical datasets are easily available in one command line in R, and plenty of advanced visualization tools are now available, which are not cited in the manuscript. The authors only cite the John Hopkins dataset, but do not compare their tools and dataset with this one, nor do they clarify the novelty of the proposed tool

2. The grammar of the R package is not clear. The package is not published on the CRAN, and the documentation to succinct to help

3. There are multiple problems with the vizualisation tools, especially with the shiny app (wrong scales, missing interactivity on maps, missing axis legend, errors on data selection). The overall impression is that the app is not finished nor properly checked.

4. Some indicators are strongly problematic, namely the growth rates, which are calculated on a daily basis. Some other are undefined and unknown, like the health rate. The prediction given by forecast are rather uninformative (see example below), and not based on any epidemiological background.

5. The data number are different from other institutional COVID19 dataset, and I do not see any comments about it, nor validation or comparison of the present dataset with other institutional dataset.

The first four points are critical for me. Points 2, 3 and 4 would require as subsequent work, as the overall impression is that the package is not finished and rather in a "work in progress" stage. The first point is essential and from what is presented here, I do not think that the package bring something new concerning visualization or data about COVID19. I would therefore not recommend publication in peerJ, and advise the authors to target a R specific publication when the package is finished and made available on CRAN

Below find the detail of each point:

1) The authors, to motivate their work, claim line 69: : "Thus, more specific historical data is needed. However these data are not well obtained at present, as many data sources only provide real-time aggregated data. There is still a lack of clean, easily accessible historical data source and analysis tool". And line 83 that their package provide: "statistics data collection, statistics data query, which include both real-time and historical data query, geographic maps visualization and interactive dashboard.". There are many available aggregated data, including historical data, with advanced interactive dashboards. The query from R to these data is rather easy because it consists in downloading csv data from github or from permalinks. Among others:
   a. https://ourworldindata.org/coronavirus-data-explorer which provide complete exploration of cases, deaths, tests. These variable can be normalized per population or not, cumulative, averaged, over the time, with maps etc. The complete dataset with data from beginning of the Crisis are available from permalinks: https://ourworldindata.org/coronavirus-source-data , and can thus be directly downloaded in R in one line.

b. Each national and transnational CDC institution provide visualization tools, as for the European CDC: https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html

c. The John Hopkins University provide an advanced interactive app: https://coronavirus.jhu.edu/map.html but also plenty of representation of the trends, with accessible data source https://github.com/CSSEGISandData/COVID-19 including historical data at the regional level for China, Japan, Brazil, Germany, Italy, Spain, Denmark, Peru, Columbia, US etc. The authors do cite them at the end of their article, but do not comment nor compare their tools with this one.

d. The authors mention the blog post https://www.statsandr.com/blog/top-r-resources-on-covid-19-coronavirus/ which refers to 39 shiny app about COVID and 8 other R packages about COVID19 data. The authors should at least refer to them and comment on the difference.

2) The package is not available in CRAN, which is a drawback, as CRAN ensures the quality of the package and guarantee that the package is maintained. The grammar detailed in "statistic data query" is not consistent. For example, while x['hubei'] works for China dataset, x['Campania'] does not work for Italy.
The github package provide the code lines:

library(nCov2019)
nCov2019_set_country(country = 'Italy')
*x['province'] # this will return Italy data only.*
Which do not work.
   nCov2019_set_country('Italy')
   y <- load_nCov2019(lang = 'en')
   y['province']
will not work either, and the user has to find that actually, as indicated in the vignette, he should do:
   nCov2019_set_country('Italy')
   y <- load_nCov2019(lang = 'en', source = 'github')
   y['province']
Then trying to plot only Itay, plot(y) does plot the whole world.

3) The article presents the package as allowing to easily explore the data. First, the local and the online shiny app are not the same. Secondly both shiny apps have many flaws.
Here are the principals I noticed after playing few minutes with it:
For the local one:
   a. Growth rate curve does not have any x axis
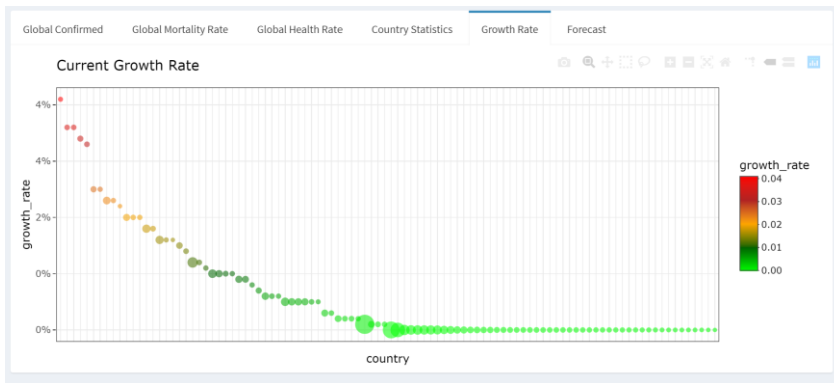
b. Forecast curve either:
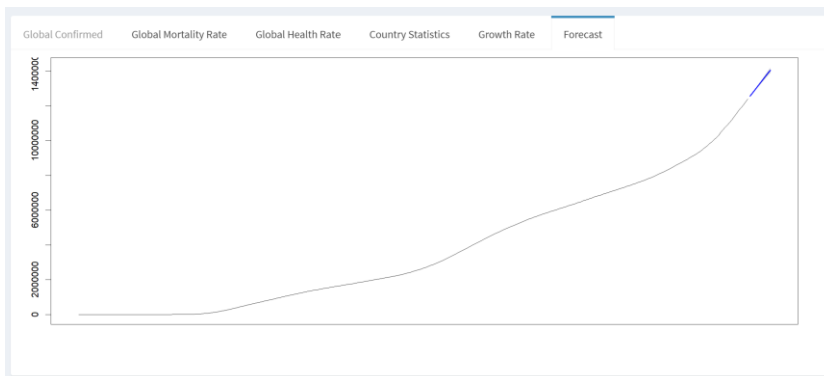
c. It is not possible, once a province is selected, to go back to the entire country

d. There is the possibility to select unpublished sources, that generate errors

e. The ggplotly main plot does not offer the possibility to display the daily curves. Furthermore, the three variable displayed are on very different scales (example below where the red curve appears as flat), and the title "Plot" is rather uninformative
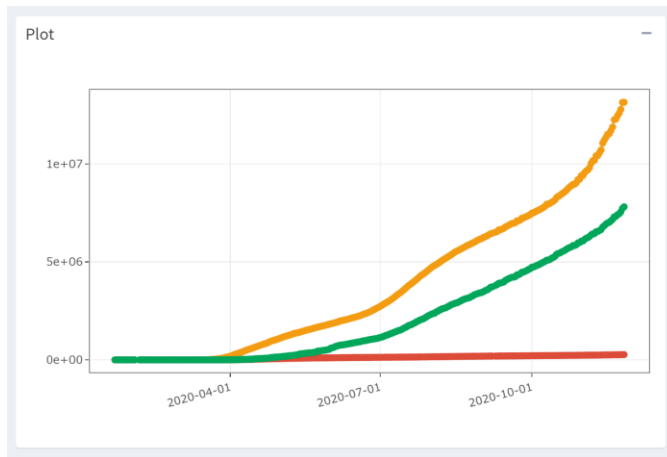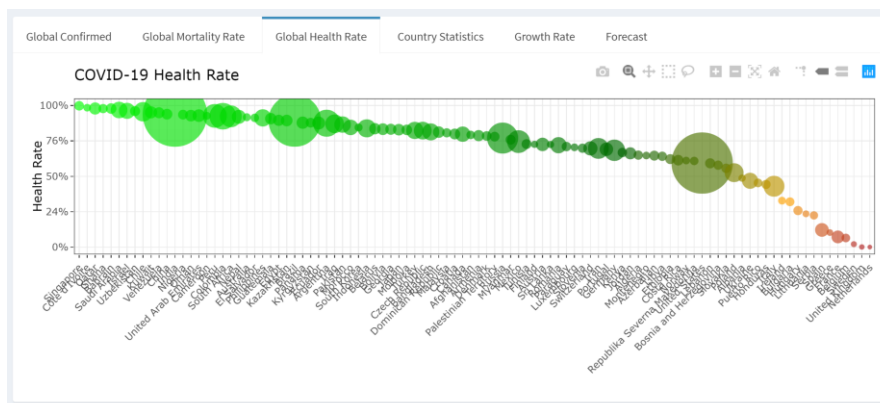
f. There is a mistake when downloading the data file, which appears as .tsv:

chier : coronavirus_histrical_2020-11-26 (1).tsv

g. The map is not interactive, thus giving very limited possibility (no zoom, no indication when hovering with the mouse). Consider leaflet at least, to have the possibility to display and zoom:
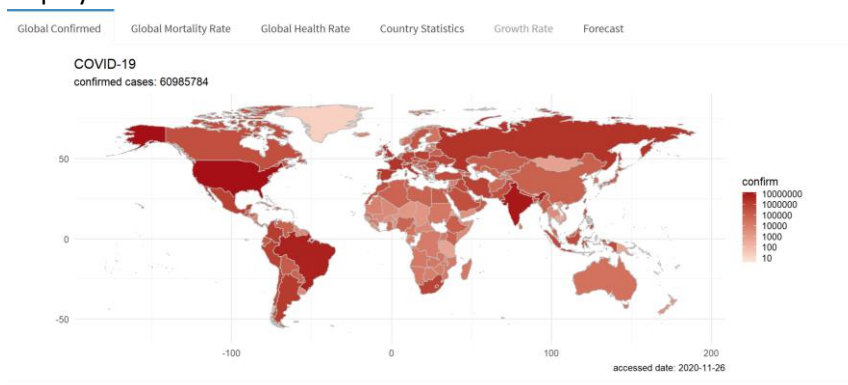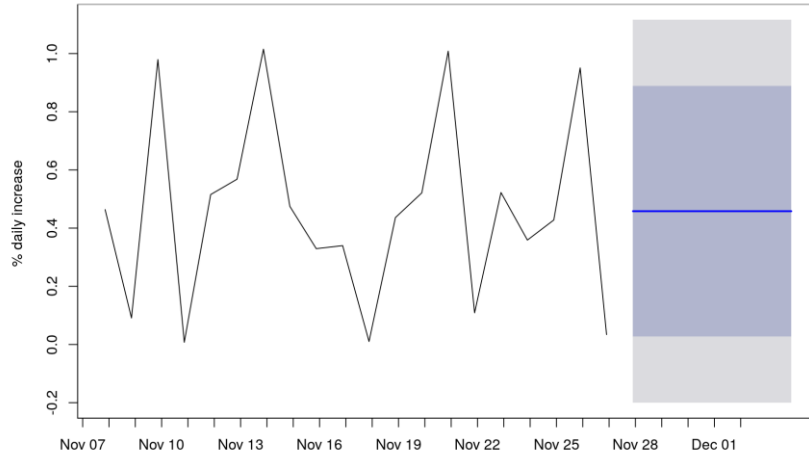

h. The measure health rate is undefined

For the online one:

a) The oneline website has similar problematic and is extremely slow. I tried on two webbrowser without any addblockers, and the forecast tab takes about 2 minutes to load.
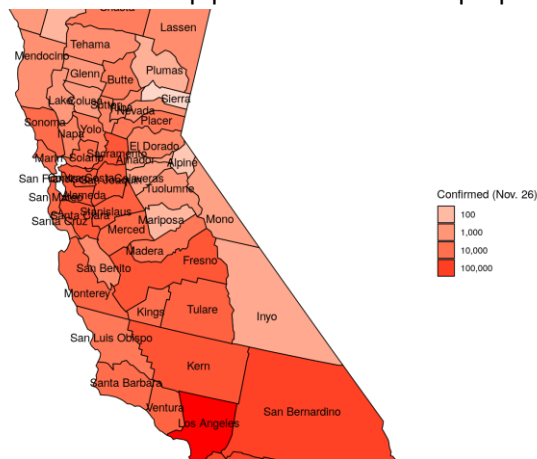b) Forecast produce uninformative curve like this one:

Total cases in India is expected to increase by 0.5% on average daily, reaching 9571304 by Dec 03.

i. Scales in the graphs are totally inappropriate: here the death cannot be seen
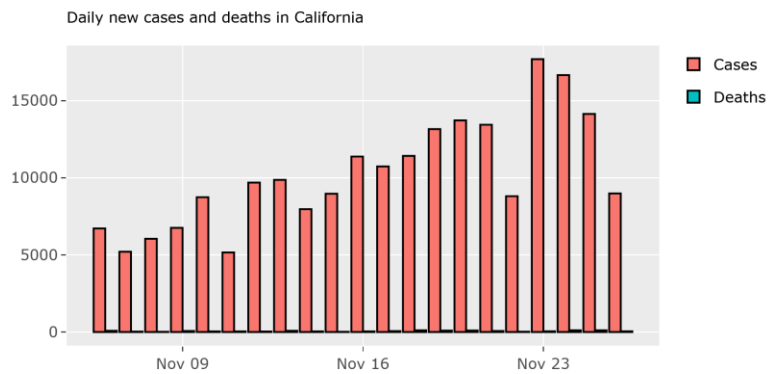
Here in the static map plot all the names superpose:

4) Some indicators do not make sense. The authors claim that their package: "produce convenient and rich visualization images, facilitating mathematical modeling and forecasting of the COVID-19 outbreak.". But the use of ARIMA based model without any epidemiological consideration lead to prediction like the one below:

Total cases in India is expected to increase by 0.5% on average daily, reaching 9571304 by Dec 03.
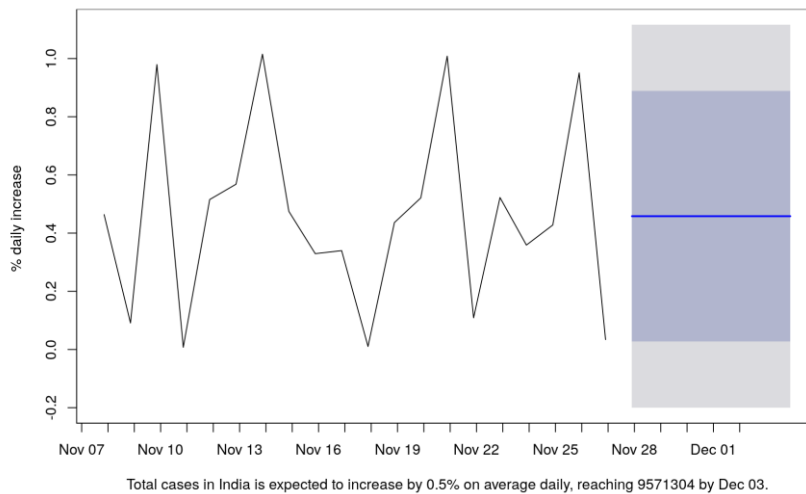
And can not be considered as a proper forecast. Secondly, the growth rate calculate as the daily relative change does not have any valid epidemiological sense, as the COVID data are known to fluctuate strongly on a day to day basis due to case report changes. Growth rate should be calculated on a week basis, or based on a model.

5) On the 27th of novembre, the package gives 13248676 cases for the US, 93219 for China, 2235537 for France, whereas European CDC gives 12.88 billion for US (360 000 cases of difference), 91784 for China, 2.18 million for France (50000 case difference). These difference are not criticals, but as the authors use non institutional sources, they should at least comment, inspect and validate the difference with reference datasets.