Visual Analysis of COVID-19 Report

Group: 13
Sai Vinay Nandigam
Tarun Upputuri

Overview & Motivation:

In Wuhan, China, an unknown infectious disease that appears to be pneumonia was discovered in December 2019. Now called Corona Virus, which is a contagious disease. The persons infected with this disease had previously suffered from respiratory illnesses. From the beginning of the sickness until the present day, many individuals have died due to it, and people are constantly fighting it. There are numerous options for halting the disease's spread. Sharing daily cases and trends from the beginning of the disease's transmission is one of the most effective methods.

Many APIs, repositories, and companies are collecting data about COVID-19 to identify helpful information for controlling the spread of COVID-19. There are several methods for extracting useful information from data. Visual analysis is one of them. A user can better understand and acquire essential insights from the data by looking at the visuals. Also, how quickly a user may obtain insights from the data is dependent on the visualizations.

The knowledge gathered through visualization of COVID-19 data can be helpful in various ways. Here are a handful of them: If a person wants to go to a particular region, they can check the COVID-19 status in that country to see if they should go or not. A researcher can learn about the severity of COVID-19 in different nations to learn more about the condition, and the information gained may help determine preventive strategies. The governments of the respective countries could prohibit travel from the seriously afflicted countries.

Different visualizations and their merits and demerits were evaluated and addressed in Big Data Visualization and Visual Analytics of Covid-19 Data. We assessed both advantages and disadvantages in this project, attempted to add interactive features to the visualizations presented in the paper, and gave alternative visualizations.

Related work:

Many dashboards are available online. The renowned dashboards are:

- The WHO covid-19 dashboard.
- Johns Hopkins University dashboard.
- The Government of Canada's website.
- The covid-19 dashboard by the European centre for disease prevention and control.

In addition to these dashboards and websites, COVID-19 visuals were described in detail in the paper[1]. The author went over the advantages and disadvantages of each visualization. According to the author, the most common methods for showing spatial data are bubble maps and choropleth maps. The bubble's radius indicated the confirmed cases on the bubble map. The wider the bubble, the more serious the cases. It was difficult to find cases in countries that were closer and smaller in size.
On the other hand, the choropleth map uses different color hues to show the intensity of covid-19. While the choropleth map solved the problem of the bubble overlapping, it also created a new difficulty: countries with smaller areas were difficult to identify. The author discussed a column chart as an alternative to using a map to visualize spatial data. He also mentioned that temporal data was necessary for displaying daily trends in addition to spatial data. Line graphs, column charts, and stacked areas under curves are visualizations that can represent both spatial and temporal information.

We were inspired to show COVID-19 data by a geographical map described by Dr. Federico Iuricich in CPSC 6030(Data Visualization). One of the critical motivations for including a few of the interactions in our project was the interactions mentioned by the professor in class.

Data: The data for this study was obtained from the COVID-19 Data Repository, which is maintained by Johns Hopkins University's Center for Systems Science and Engineering (CSSE). COVID-19 data from throughout the world is included in the dataset. Countries and their daily confirmed cases, recovered, and daily death cases were the most important attributes in the dataset. There were three datasets for three attributes: one for confirmed cases, one for recovered cases, and one for fatalities count. All of the attributes have been combined into a single data set. The geo.json file, which contains the country name, latitudes, longitudes, and country code, has also been used. To allocate the values to the individual countries, we matched the country codes in both datasets. There were no nation codes in the Covid dataset. We've given them manually in the dataset. The COVID-19 cases were collected regionally for countries like China, Australia, and The United Kingdom in the dataset. We have united all the cases of individual regions and linked the data with their respective countries.

The source for the dataset: https://github.com/CSSEGISandData/COVID-19

Another dataset was collected from WHO website, which helped us to visualize the daily trends of COVID-19. It has the daily cases data and also the cumulative data of the daily cases.

The source for the WHO dataset: https://covid19.who.int/WHO-COVID-19-global-data.csv

Questions:

The following questions will be answered from our visualizations:-

How many confirmed cases does each country possess?
How many confirmed, recovered, and death counts do each country possess?
What are the daily trends of confirmed cases in each country?
What are the daily trends of confirmed cases and deaths across the world?
What are the top-10 most affected countries and their evolution of confirmed cases?

At first, we came up with a question, “What are the confirmed, recovered, and death counts in top-5 affected countries?". Later working on the project, we have implemented an interactive bar chart comparing the confirmed, recovered, and deaths count for a selected country on the map. So we came up with the question, "what are the top-10 most affected countries and their evolution of confirmed cases?"

Exploratory Data Analysis:

The visualizations discussed in the paper[1] were visualized by using Johns Hopkins COVID-19 repository and World Health Organization(WHO). At first, we have considered the dataset available in the Johns Hopkins repository, and we have visualized a color map using the data set. This color map displays the severity of covid-19 by using the cumulative cases available for each country in the Johns Hopkins dataset. Later, we went through the WHO dataset, which has both daily and cumulative cases in the dataset, and found it useful for showing daily trends, whereas the Johns Hopkins dataset has only the cumulative cases, and it is not a good option to show the daily trends. So we have utilized both datasets by using the d3.js library, which was our first proposal, and we have not used any other JavaScript libraries.

Design Evolution:

At first, we have considered a **color map** for displaying the severity of covid-19 as presented in the paper[1] by using d3.js. A **line graph** showing the daily trends of confirmed cases is also provided in the paper [1], so we decided to show the daily trends of both confirmed cases and deaths. In the paper [1], a column chart representing the no of people tested in the provinces of Canada was presented, but our dataset didn't have the data related to testing, so we came up with a **grouped bar chart** for comparing confirmed cases, recovered cases, and deaths for the top five affected countries. These three visualizations were considered at first in the proposal.

Later, we have successfully displayed the confirmed cases data in the color map as mentioned in the proposal. We have tried comparing the daily confirmed cases and deaths of a particular country in the line graph. As we can see in the figure, the confirmed values were more and not even close with the deaths on any date. So we have used an interactive bar chart for displaying the daily trends of confirmed and death counts across the globe.

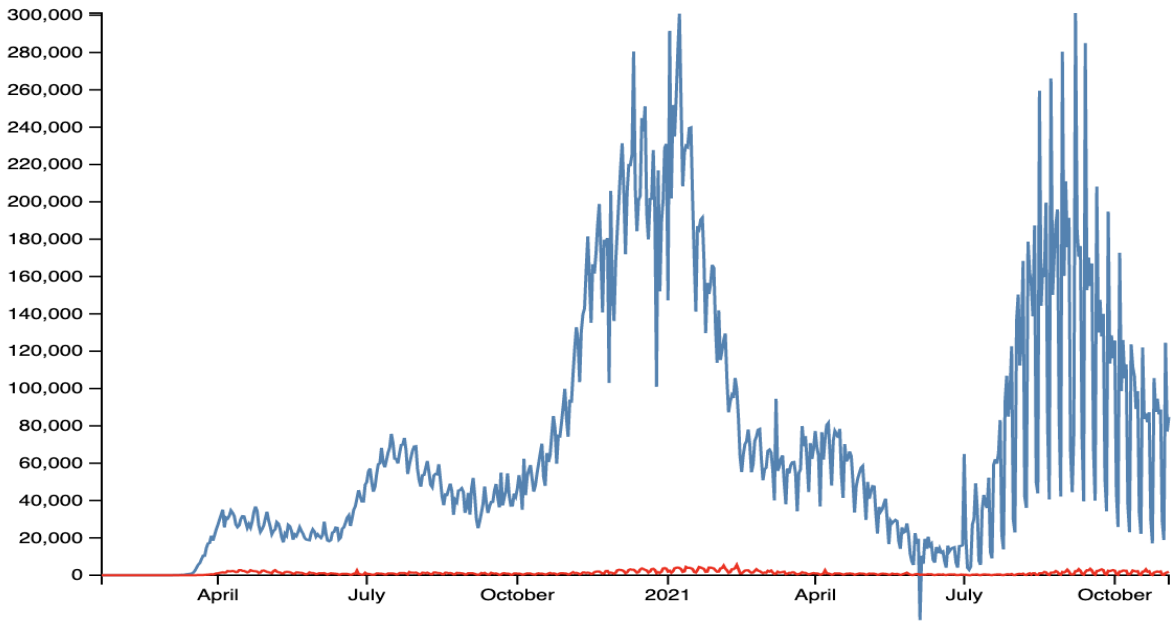

Fig: Comparing the daily trends of both confirmed and deaths counts

We have also implemented a grouped bar chart that displays the top 5 COVID-19 affected countries and compares their confirmed, recovered, and deaths count.

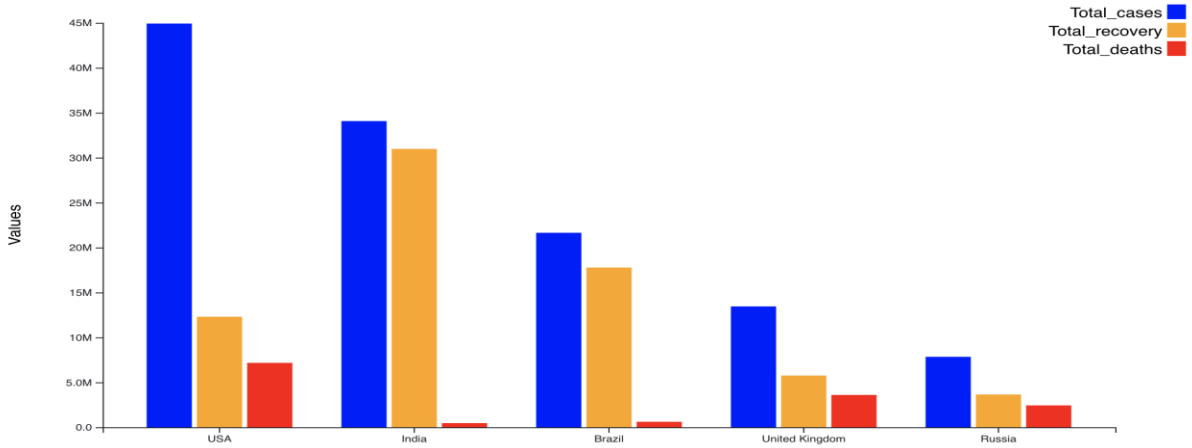

Fig 1.1 A grouped bar chart comparing the top-5 affected countries with their confirmed, recovered, and deaths count.

In the feedback and presentation sessions, professor Federico has suggested us to link the visualization with the main color map visualization. Instead of comparing only the top five affected countries, we come with the idea to compare every country's confirmed, recovered, and deaths count when we click on a particular country. So we have implemented this interaction to our main visualization instead of using a grouped bar chart.

We also felt that showing every country's daily trends is also an important visualization to understand the severity of covid-19 in that country. So we have made an interaction when a user selects a particular country, It also displays the daily trends of the chosen country. So, we have provided two interactions for the color map.

In paper [1], a representation of the no of people tested in the provinces of Canada was presented; we have deviated from the proposal. A bar chart is required to show the top 10 most affected countries to show the spatial and temporal information from the data. We have implemented a bar chart for the top 10 affected countries, but it was ineffective. So, we have added a raced bar chart, which shows the evolution of covid-19 confirmed cases in the top 10 affected countries over a while.

Implementation:

Visualization 1: A Color map displaying the severity of Covid-19.

This visualization shows the severity of covid-19 by using the shades of the color red. Countries possessing confirmed cases represent any one shade of the color red. The darker the shade of the color red represents the level of severity of covid-19 in that particular country. There is also a color scale in the visualization representing the confirmed cases associated with the values.
Interactions were visualized. We have implemented a hover feature for the color map, which helps users easily identify the countries with smaller areas. We also have implemented an on-click function. Whenever the user clicks on a particular country, on the top, the name of the country and its confirmed, recovered, and death count will be displayed. At the same time, there were two interactive graphs. On selecting a particular country on the color map, a bar graph visualizing the Confirmed, Recovered, and Death rates will be shown in the first interaction, followed by the line graph that lists the daily cases of the COVID particular country from its inception to till date.

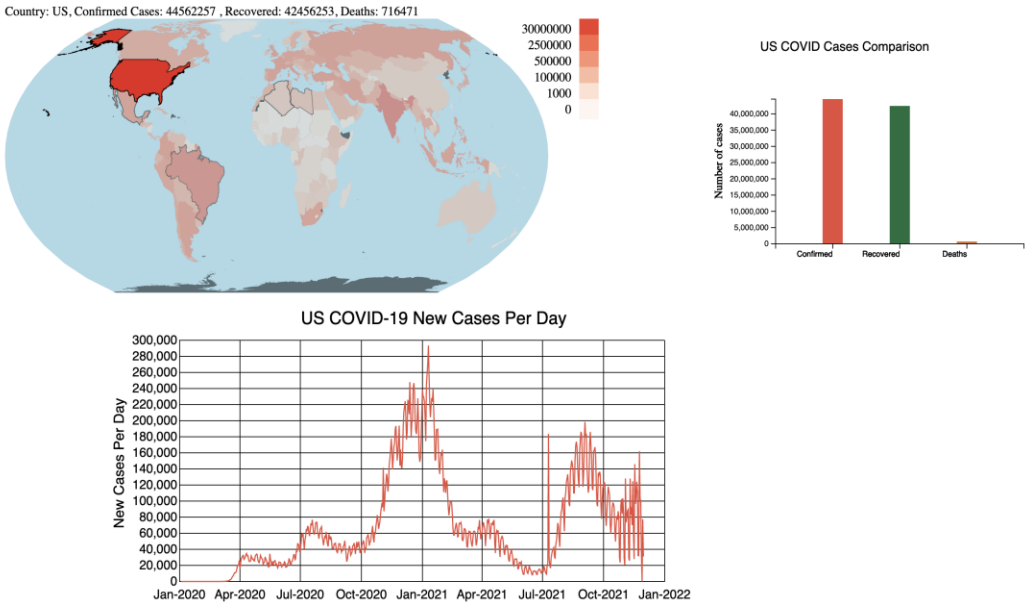

Fig 1.2 Overview of the color map and it's interactions when the user selected the country US

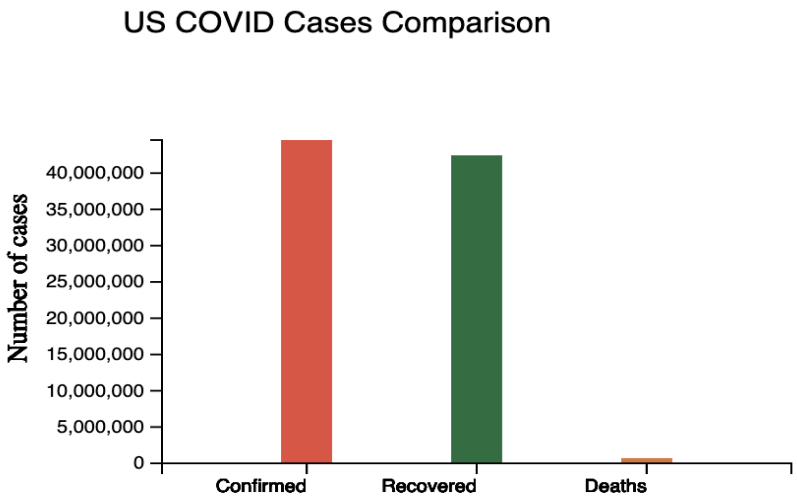

Fig 1.3 A bar chart comparing the confirmed, recovered, and death counts of US

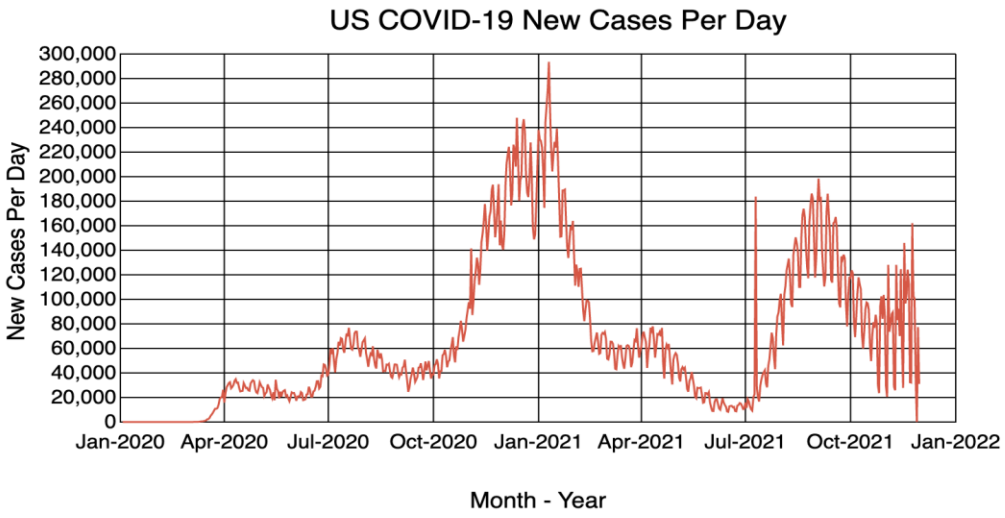

Fig 1.4 A line graph displaying the Daily Confirmed cases in the US

Visualization-2: World Wide Daily Counts

In this visualization, a user can observe the daily trends and overall trend of both confirmed and deaths count of covid-19 worldwide. The user can get the exact confirmed and deaths count of the date selected on the bar.

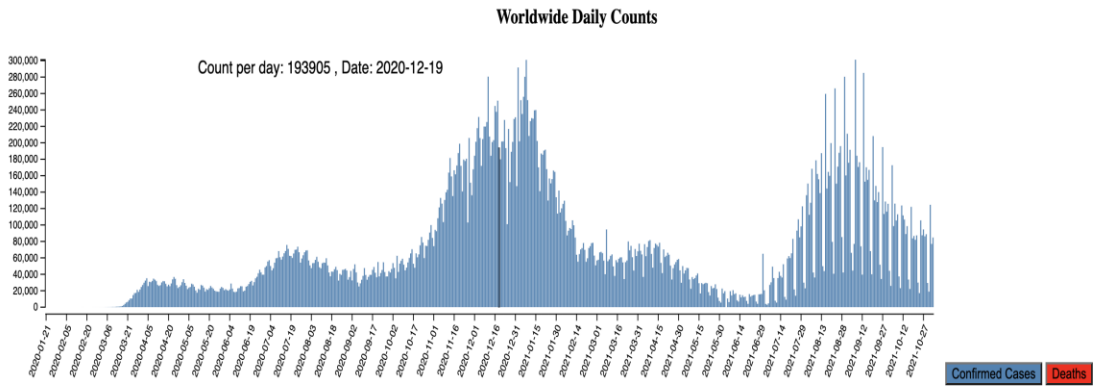

Fig 1.5 A bar graph displaying the daily trends of covid-19 confirmed cases across the globe

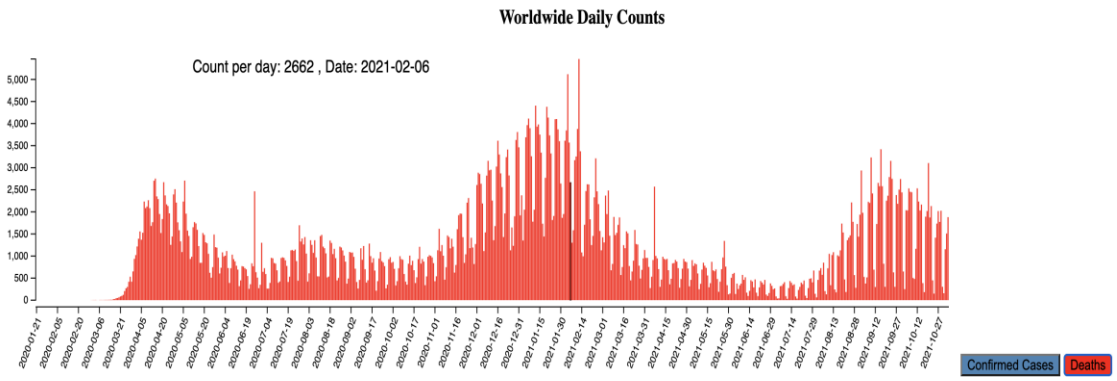

Fig 1.6 A bar graph displaying the daily trends of covid-19 Deaths count across the globe

Visualization-3: A Raced Bar Chart displaying the top-10 Covid-19 affected countries.

In this visualization, a user can watch the change of covid-19 in different countries over some time. A user can select the pace of the data loading in the visualization as we have provided a dropdown indicating slow, medium, and fast. From this visualization, we can see after the data was loaded fully, The USA, India, Brazil, UK, Russia, Turkey, France, Iran, Germany, Argentina, Spain, Colombia, Italy, Indonesia, and Mexico are the top 10 covid-19 affected countries.

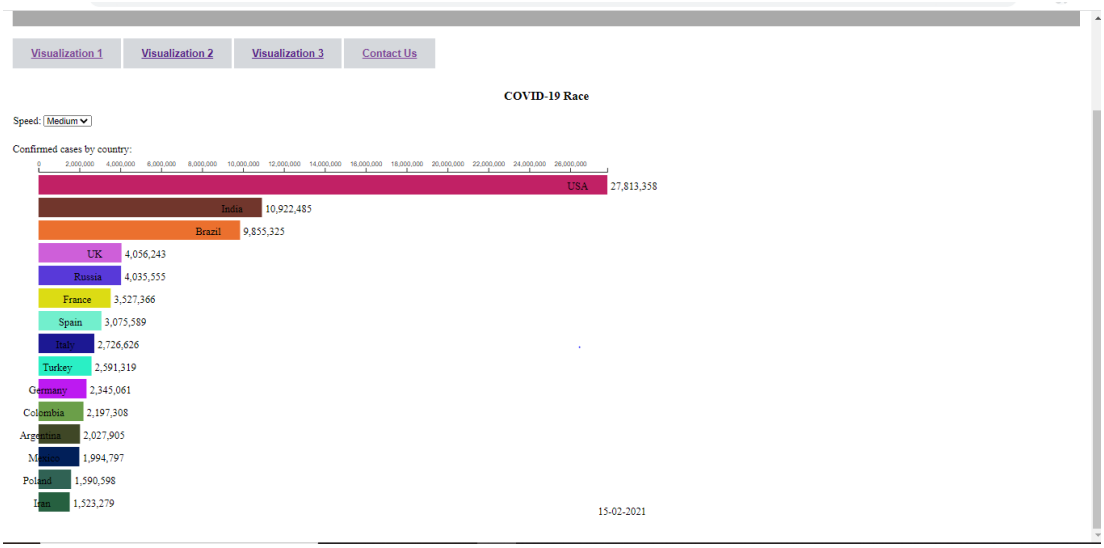

Fig 1.7 A Raced Bar chart visualizing the data of top-10 affected countries.

Evaluation:

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you improve it?

The visualizations answered all our questions. From the first visualization colour map, we can address the highly affected covid-19 countries by observing the countries represented by the darker shade of the colour red. We can consider India, the USA, and Brazil the most affected countries on the map as they possess the darker shade of red. From the scale, we can infer that the darker red colour was associated with the countries with more than 30 million confirmed cases. This visualization helped us answer the question, **"How many confirmed cases does each country possess?".** When we click on a particular country, the country's name, confirmed cases, recovered cases, and deaths count are available at the top of the map. It helps us know the exact confirmed, recovered, and death counts of a selected country. Also, a bar graph comparing confirmed, recovered, and death counts of a selected country helped us to know them visually. These features enabled us to answer the question, **"How many confirmed, recovered, and death counts do each country possess?"**

Also, a line graph displaying the daily confirmed cases of a selected country helped us identify the daily trends of confirmed cases from the beginning of the covid-19 in that country. This visualization helped us answer the question,**" What are the daily trends of confirmed cases in each country?"**

An interactive bar graph displaying the daily trends of confirmed cases and deaths count helped us identify the count on each day and the overall trend. This visualization helped us answer the question, **"What are the daily trends of confirmed cases and deaths across the world?"** The raced bar chart helped us answer the question, **"What are the top-10 most affected countries and their evolution of confirmed cases?"**

Our visualizations helped us to gain information quickly. These visualizations are user-friendly. We hope that anyone can easily understand the visualizations and gain insights quickly. We have presented better visualizations than we had proposed earlier. We also hope that a user can easily understand the severity of covid-19 in different countries.

Our visualization could be improved by implementing a filter on the page for displaying the data related to a country selected in the filter. For example, suppose a user wants to know the severity of covid-19 in the USA. If the user had input USA in the filter, the visualizations would display the data related to the USA in all the visualizations associated with the colormap.

References:

[1] Big Data Visualization and Visual Analytics of COVID-19 Data, Carson K. Leung; Yubo Chen; Calvin S.H. Hoi; Siyuan Shang; Yan Wen; Alfredo Cuzzocrea, IEEE, https://ieeexplore.ieee.org/document/9373130