Quest SRI High School Research Project 2023

Investigation of the Impact of COVID-19 Vaccine in the United States

Team Members

Faith Aylesworth, Parsh Gandhi , Jai Kashyap, Rithek Shankar, Burhan Yasakci

Mentor : Prof.Necmi Mutlu

Background

- The COVID-19 pandemic has had a devastating impact on the United States, with over 800,000 deaths as of March 8, 2023. One of the most effective ways to prevent COVID-19 is to get vaccinated.
- Vaccinations have been shown to be highly effective in preventing serious illness, hospitalization, and death from COVID-19.

Jai Kashyap

Research Question

What is the impact of the COVID-19 vaccine on number of confirmed deaths in the United States?

Jai Kashyap

Hypothesis:

We hypothesized that the number of COVID-19 vaccine doses administered will be inversely correlated with the number of confirmed deaths from COVID-19.

In other words, we hypothesized that as the number of vaccine doses administered increases, the number of confirmed deaths from COVID-19 will decrease.

By Faith Aylesworth

Our Method:

We collected data on the number of COVID-19 vaccines administered and the number of confirmed deaths from COVID-19 in the United States from the website https://ourworldindata.org/.

We then developed C++ program to analyze the data and look for a correlation between the two time frames

- November 1 2020- March 1st in 2021 before vaccination
- November 1 2021- March 1st in 2022 after vaccination

Finally calculate the Correlation between New Cases vs New deaths within each time frame and compare the two Correlations.

By Faith Aylesworth

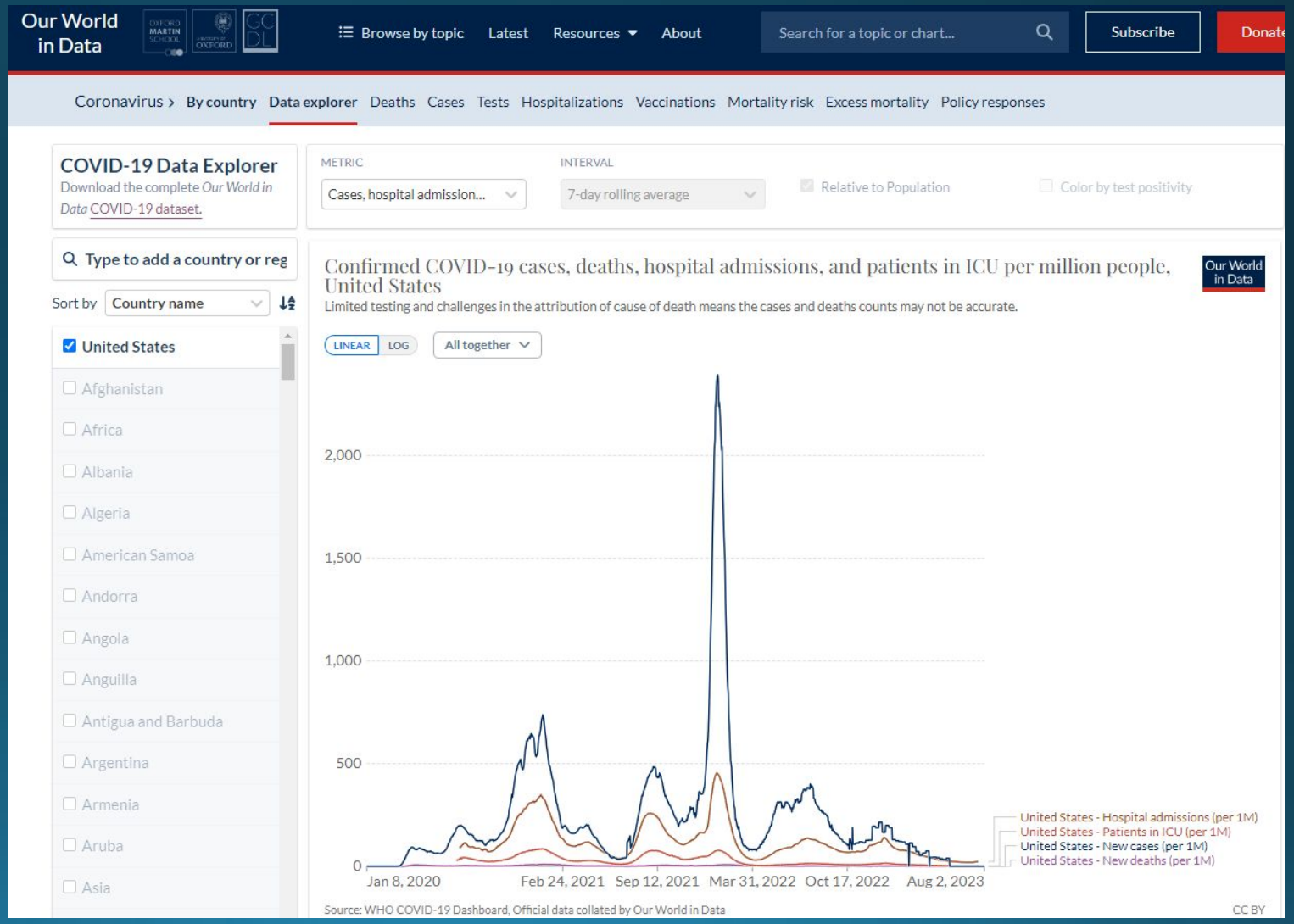
Data Source | Our World in Data

Our data came from the Our World in Data Covid-19 Data Explorer.

The URL is https://ourworldindata.org/

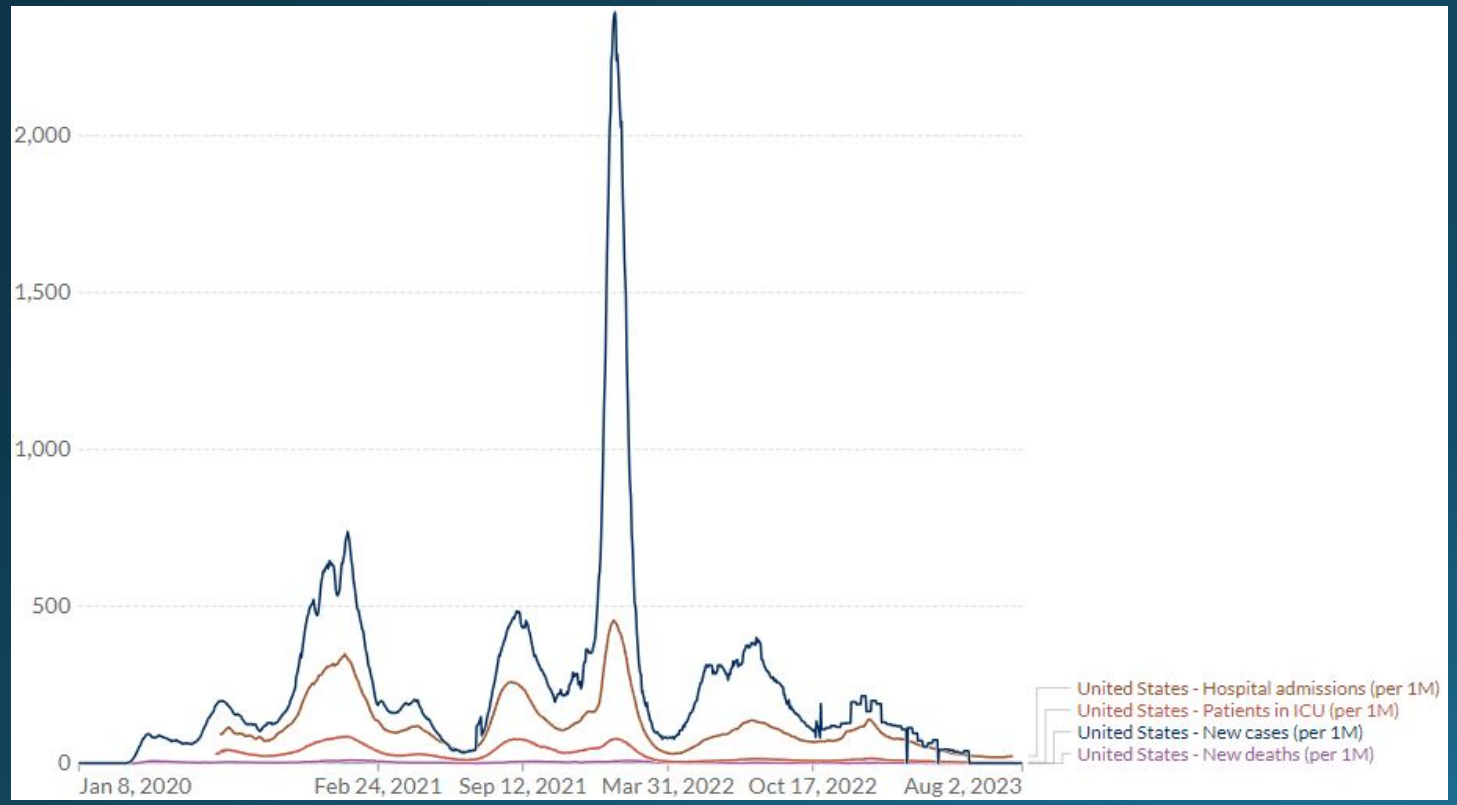
Our World in Data is a scientific online publication that focuses on large global problems such as poverty, disease, hunger, climate change, war, existential risks, and inequality

Rithek Shankar



Collected data time frames

- November 1 2020- March 1st in 2021
- November 1 2021- March 1st in 2022



1	date	new_case	new_deaths
2	11/1/2020	97518	908
3	11/2/2020	92783	942
4	11/3/2020	141582	708
5	11/4/2020	87188	738
6	11/5/2020	98499	1031
7	11/6/2020	113875	1298
8	11/7/2020	122830	1227
9	11/8/2020	133337	1260
10	11/9/2020	129810	1095
11	11/10/2020	111648	882
12	11/11/2020	124487	930
13	11/12/2020	145951	1674
14	11/13/2020	153660	1518
15	11/14/2020	162300	1347
16	11/15/2020	176989	1271
17	11/16/2020	164239	1472
18	11/17/2020	137564	1027
19	11/18/2020	153374	1018
20	11/19/2020	167352	1591
21	11/20/2020	170904	1833
22	11/21/2020	188681	1920
23	11/22/2020	193369	1959
24	11/23/2020	173861	1724
25	11/24/2020	149783	1303
26	11/25/2020	160411	1413
27	11/26/2020	175183	1889
28	11/27/2020	194100	2289
29	11/28/2020	145029	1674

123	date	new_case	new_deaths
124	11/1/2021	34950	453
125	11/2/2021	25347	302
126	11/3/2021	117847	1196
127	11/4/2021	74593	1450
128	11/5/2021	78766	1188
129	11/6/2021	88850	1204
130	11/7/2021	89858	1614
131	11/8/2021	35987	476
132	11/9/2021	31524	256
133	11/10/2021	116391	1161
134	11/11/2021	82883	1587
135	11/12/2021	93749	1345
136	11/13/2021	73647	970
137	11/14/2021	122999	1734
138	11/15/2021	69276	544
139	11/16/2021	33593	217
140	11/17/2021	133590	1080
141	11/18/2021	96915	1464
142	11/19/2021	105810	1442
143	11/20/2021	123870	3867
144	11/21/2021	113822	1662
145	11/22/2021	46194	374
146	11/23/2021	38751	177
147	11/24/2021	145378	1177
148	11/25/2021	101066	1560
149	11/26/2021	106535	1374
150	11/27/2021	72723	465
151	11/28/2021	81747	563

Rithek Shankar

C++ Program Structure & Results

- The size of the data is found & stored
- Stores each data point into two separate 1-D arrays, one for the # of cases and one for the number of deaths
- Uses the data, now stored in the 1-D arrays, to calculate the Correlation Coefficient
- Uses the Correlation Coefficient to reach a conclusion
- The same process is repeated for both the November 1, 2020 - March 1st, 2021 and November 1, 2021 - March 1st, 2022 time frames.
- The correlation coefficient and conclusion are displayed

Parsh Gandhi

C++ Program Structure

```cpp
int sizeOfArray(string filename) // function definition with the parameter of the name of the file to be opened

{
    //Definition of function specific variables
    string line;

    // lines will keep track of the number of lines read in from the file (both
    // so far as each line is read in, and in total).  We initialize it to 0
    // because we have not read any lines from the file yet.
    int lines = 0;

    // ifstream file object is used to access the file
    ifstream file;

    // Use the open member function of the ifstream file object to open the file
    // with the filename provided by the user.
    file.open(filename);

    // If the file fails to open for some reason (e.g. the file doesn't exist),
    // the fail member function of the ifstream file object will return true.
    // If it does, we exit the program with an error message and status.
    if (file.fail())
    {
        // Output an error message for the user
        cout << "File failed to open." << endl;
    }

    // The .eof() member function will return true when we reach the end of the
    // file, so we continue the loop so long as this is NOT true using !file.eof()
    while (!file.eof())
    {
        // getline() will read the next line from the file and store it into the
        // string in the array at the index lines.
        getline(file, line);

        // Increment lines so that the next line in the file is stored in the next
        // index of the array.
        lines++;
    }

    // Close the file as we are now done working with it
    file.close();

    return lines;
}
```

Size of the data is found & stored

```cpp
void data(string filename, double arr1[], double arr2[], int size) // Function definition where input paremeters are the 2 arrays, filename, and size of the array
{
    // ifstream file object is used to access the file
    ifstream inputData;

    // Use the open member function of the ifstream file object to open the file
    // with the filename provided by the user.
    inputData.open(filename);

    // For loop will scan through each array index
    for (int i = 0; i < size; i++)
    {
        inputData >> arr1[i]; // Will put the number in index of the array of the first column of the file in the first array
        inputData >> arr2[i]; // Will put the number in index of the array of the first column of the file in the second array
    }

    // Close the file as we are now done working with it
    inputData.close();
}
```

Stores each data point into two separate 1-D arrays, one for the # of cases and one for the # of ICU patients

```cpp
double myCoorelationCoefficient(double arr1[], double arr2[], int size) // Coorelation coefficient function that will take 3 inputs: the two arrays and the size of them
{
    //Definition of function specific variables
    int i; // This integer variable represents the integer which will be used in the for loop
    double sum1 = 0; // Represents the sum of the numbers of the first array
    double sum2 = 0; // Represents the sum of the numbers of the second array
    double mean1; // Represents the average of the numbers of the first array
    double mean2; // Represents the average of the numbers of the second array
    double radicand1; // Represents the sum of powers from the first array
    double radicand2; // Represents the sum of powers from the second array
    double numerator; // Represents the numerator of the coorelation coefficient
    double denominator; // Represents the numerator of the coorelation coefficient
    double coorelationCoefficient; // Represents the coorelation coefficient

    for (i = 0; i < size; i++) // This for loop finds the sum of the numbers in each array
    {
        sum1 += arr1[i];
        sum2 += arr2[i];
    }

    // Calculate the mean of each array by using the sum of the numbers and dividing it by the given size of the array
    mean1 = sum1 / size;
    mean2 = sum2 / size;

//numerator calculations
    for (i = 0; i < size; i++) // To find the numerator we use this for loop to get the sum of the product of the subtraction between two arrays and their resepctive averages
    {
        numerator += ((arr1[i] - mean1) * (arr2[i] - mean2));

    }

//denominator calculations
    for (i = 0; i < size; i++) // To find the denominator we first use a for loop to find the sum of the subtraction between two arrays and their resepctive averages raised to the second power
    {
        radicand1 += pow((arr1[i] - mean1), 2);
        radicand2 += pow((arr2[i] - mean2), 2);

    }

    denominator = sqrt(radicand1 * radicand2); // The second step to finding thee denominator is to get the square root of the product of the two arrays sum

//coorelation coefficient calculations
    coorelationCoefficient = numerator / denominator; // Finally to get the coorelation coefficient we dividie the numerator and denominator
    return coorelationCoefficient; // We return the result of the coorelation coefficient
}
```

Uses the data, now stored in the 1-D arrays, to calculate the Correlation Coefficient

The same process is repeated for both the November 1, 2020 - March 1st, 2021 and November 1, 2021 - March 1st, 2022 time frames. The correlation coefficient and conclusion are displayed

```cpp
void conclusion(float CC) // Function definition with the parameter of the coorelation coefficient
{
    // If loop to go through and display the findings according to the table below
    // Value of CC  | Conclusion
    // 0 =          | There is no correlation
    // 0<=CC<0.2    | There is a very weak correlation
    // 0.2<=CC<0.4  | There is a weak correlation
    // 0.4<=CC<0.6  | There is a moderate correlation
    // 0.6<=CC<0.8  | There is a strong correlation
    // 0.8<=CC<1    | There is a very strong correlation
    // = 1          | There is a perfect coorelation

    if (CC == 0.0)
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is equal to 0, there is no correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else if (0.0 < CC && CC < 0.2)
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is between 0 and 0.2, there is a very weak correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else if (0.2 <= CC && CC < 0.4)
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is between 0.2 and 0.4, there is a weak correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else if (0.4 <= CC && CC < 0.6)
    {
        cout << "The Coorelation Coefficient is: " << to_string(CC) << ". As it is between 0.4 and 0.6, there is a moderate correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else if (0.6 <= CC && CC < 0.8)
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is between 0.6 and 0.8, there is a strong correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else if (0.8 <= CC && CC < 1.0)
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is between 0.8 and 1, there is a very strong correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
    else
    {
        cout << "the Coorelation Coefficient is: " << to_string(CC) << ". As it is equal to 1, there is a perfect correlation between COVID-19 vaccination and the number of ICU cases." << endl;
    }
}
```

Uses the Correlation Coefficient to reach a conclusion

Results

Within November 1, 2020 - March 1st, 2021 time frame the **Correlation Coefficient is: 0.49**.

Within November 1, 2021 - March 1st, 2022 time frame the **Correlation Coefficient is: 0.44.**

Parsh Gandhi

Conclusion

As our hypothesis states, as the administered number of vaccine doses increases, the number of confirmed deaths in correlation to cases from COVID-19 does decrease.

From our data, we were able to conclude that our hypothesis was true, and this can be seen in our research, with the smaller number of correlation coefficient of .44(lower correlation) in the second timeframe compared to the first timeframe of .49(higher correlation).

The correlation was decreased after the vaccination

Burhan Yasakci

Any Questions?

Thank you