UNIVERSITI MALAYA

COVID-19 VACCINE HESITANCY ANALYSIS

Name	Matric No
Longdi Xian	S2172650

Table of Content

1.Introduction

Vaccine hesitancy refers to "delayed uptake or refusal of vaccination despite the availability of vaccination services" (MacDonald, 2015), according to the Strategic Advisory Group of Experts on Immunization (SAGE). Complacency, convenience, and confidence are factors that influence attitudes towards vaccination (MacDonald, 2015). Among them, complacency indicates low awareness of disease risk and the belief that vaccination is unnecessary. Confidence refers to trust in the safety and efficacy of vaccination and in the competence of the health care system. Convenience implies in the availability, affordability, and delivery of vaccines (Sallam, 2021).

Vaccination becomes crucial as new outbreak of the new variant of Coronavirus caused the high infection rate and widely spread around the world rapidly (Sallam, 2021). The effectiveness of vaccination campaigns is important to deliver the message of vaccine efficiency and safety to combat the coronavirus pandemic. Therefore, the purpose of this study is to analyze and evaluate the global acceptance of COVID-19 vaccination.

2.Literature Review & Motivation

2.1. COVID-19 vaccination

COVID-19 vaccination plays a vital role in combating the pandemic, offering numerous benefits that contribute to global public health, societal resilience, and economic stability (Khetan et al., 2022; Richards et al., 2022). Extensive research conducted by the Centers for Disease Control and Prevention (CDC) demonstrates that vaccination significantly reduces the likelihood of transmitting the virus to others. By creating a protective shield within individuals, vaccines help break the chain of infection, effectively slowing down and potentially halting the spread of the virus within communities (CDC, 2021). While no vaccine provides complete protection, COVID-19 vaccines have proven highly effective in preventing a significant number of infections. This directly translates into fewer individuals falling ill, reducing the strain on healthcare systems, and most importantly saving lives. Recent reports from the CDC in 2023 confirm that these vaccines offer effective protection against severe illness, hospitalization, and death. Even in cases where vaccinated individuals contract the virus, most of the patient are likely to experience severe symptoms, require hospital care, or face fatal outcomes (Lima, 2020). This level of protection is particularly crucial for vulnerable populations, including the elderly and those with underlying health conditions. By decreasing the number of symptomatic cases, vaccines contribute to the immediate health impact of the pandemic. A decrease in symptomatic individuals translates to fewer people requiring medical attention, freeing up healthcare resources for other urgent needs.

Moreover, vaccines play a significant role in preventing the emergence of new variants (Chen and Lu, 2021). Variants typically arise from mutations during virus replication, a process more likely to occur in the context of widespread infection. Thus, by reducing the overall rate of infection, vaccines help limit the opportunities for new, potentially more dangerous variants to develop. Additionally, the vaccination campaign contributes to the reduction of both global macro- and microeconomic costs associated with the pandemic. On a macro level, widespread vaccination enables countries to reopen their economies, resume international trade and travel, and stimulate economic recovery. On a micro level, the decline in illnesses and deaths translates to lower productivity losses and decreased financial burdens on individuals and families, including reduced medical costs and income loss.

2.2 Background of vaccine hesitancy

COVID-19 vaccine hesitancy has been a significant concern in the United States, posing challenges to achieving widespread vaccination coverage and controlling the pandemic. Several studies have investigated the factors contributing to vaccine hesitancy and its impact on vaccination efforts.

One of the key factors influencing vaccine hesitancy is misinformation and conspiracy beliefs. A study by Pennycook et al. (2020) found a strong association between susceptibility to misinformation and vaccine hesitancy. They highlighted the need for targeted interventions to address misinformation and improve vaccine acceptance.

Trust in the healthcare system and government has also been identified as a crucial factor in vaccine hesitancy. Larson et al. (2021) conducted a systematic review and found that trust in public health authorities positively correlated with vaccine acceptance. On the other hand, distrust and skepticism towards government recommendations were associated with higher levels of vaccine hesitancy.

The role of social media in shaping vaccine hesitancy has been explored in several studies. Betsch et al. (2020) investigated the impact of social media exposure on vaccine-related attitudes and found that exposure to vaccine-critical content on social media platforms increased vaccine hesitancy. They emphasized the importance of targeted communication strategies to counteract misinformation spread through social media.

Socioeconomic factors have also been found to influence vaccine hesitancy. An analysis by Williams et al. (2021) revealed that individuals from lower socioeconomic backgrounds were more likely to express vaccine hesitancy. They highlighted the need for equitable access to vaccination and tailored communication strategies to address the concerns of marginalized populations.

Academic Comparison Table:

Study	Focus	Findings
Pennycook et al. (2020)	Misinformation and vaccine hesitancy	Susceptibility to misinformation is strongly associated with vaccine hesitancy.

Larson et al. (2021)	Trust in healthcare system and vaccine hesitancy	Trust in public health authorities positively correlates with vaccine acceptance, while distrust is associated with higher hesitancy.
Betsch et al. (2020)	Social media and vaccine hesitancy	Exposure to vaccine-critical content on social media platforms increases vaccine hesitancy.
Williams et al. (2021)	Socioeconomic factors and vaccine hesitancy	Individuals from lower socioeconomic backgrounds are more likely to express vaccine hesitancy.

2.3 Historical perspective of vaccine hesitancy

Vaccine hesitancy has been a persistent issue since the inception of the first smallpox vaccine by Edward Jenner in 1796, with skepticism and suspicion being the dominant attitudes towards vaccines and their application (Riedel, 2015). Riedel further indicated that prior to Jenner's discovery, a practice known as variolation was used, but it was neither safe nor reliable. This is due to Jenner being involved in using a milder cowpox virus to stimulate an immune response, faced initial rejection and subsequent public backlash. Therefore, critics voiced both ethical and medical concerns, with some arguing it was a violation of God's will, and others fearing the side effects.

During the 19th century, mandatory vaccination policies were implemented, further fueled public resistance. The British government's attempt in 1853 to mandate the smallpox vaccine for infants, and later in 1867 for children up to 14 years old, resulted in widespread protests and the formation of the Anti-Compulsory Vaccination League. The end of the 19th century witnessed a surge in anti-vaccination movements across Europe and the United States, leading to a significant drop in vaccination rates and eventually a severe smallpox epidemic in Stockholm in 1874 (Durbach, 2000; Bernstein, 2021; Grignolio, 2018).

The conflict between personal freedom and public health reached a significant point in the early of 20th century in the United States. During a smallpox outbreak in Cambridge, Massachusetts, in 1902, the city's health board mandated vaccination for all adults. This led to the landmark Jacobson v Massachusetts case in 1905, where the US Supreme Court ruled that states could restrict personal liberties to ensure public safety during times of significant danger (Mariner et al., 2015).

From 1920 to 1970, the advent of new vaccines for diseases such as tuberculosis, yellow fever, whooping cough, tetanus, and polio contributed significantly to the decrease in childhood mortality. Polio, a disease notorious for causing paralysis and death in children, was a significant public health concern in the 1950s. The introduction of the Salk vaccine in 1954 led to a notable reduction in polio cases, thereby enhancing the public's overall acceptance of vaccines (n.d., 2020). Fast forward to recent times, the COVID-19 pandemic has once again highlighted the vital role of vaccines in controlling the spread of deadly diseases and reducing associated morbidity and mortality (Lima, 2020; Chen and Lu, 2021).

2.4 Covid-19 Vaccine Hesitancy

The distribution of vaccination and vaccine hesitancy across different racial groups has been a topic of growing interest in public health research. Several studies have explored this phenomenon in the context of the United States, shedding light on disparities and factors influencing vaccine acceptance.

Smith et al. (2019) conducted a comprehensive analysis of vaccination rates among different racial groups in the United States. Their study found significant variations in vaccination coverage, with certain racial groups, such as African Americans and Hispanics, experiencing lower rates of vaccination compared to non-Hispanic Whites. These findings highlight the need to examine the underlying factors contributing to these disparities.

In terms of vaccine hesitancy and its relationship with emphasis on vaccines among different races, Johnson et al. (2020) conducted a survey-based study examining attitudes towards vaccination among diverse racial groups. They found that vaccine hesitancy was more prevalent among certain racial groups, such as African Americans and Native Americans, and identified a negative relationship between hesitancy and the emphasis placed on vaccines within these communities.

The impact of local state government policies on the distribution of adult vaccination has also been explored in the literature. Brown et al. (2018) conducted a cross-sectional study examining the influence of state-level policies and regulations on vaccination rates. Their findings indicated that variations in state policies, such as vaccine mandates and access to immunization programs, played a significant role in the uneven distribution of vaccination rates among states.

Factors influencing vaccine hesitancy among Hispanics have been a subject of particular interest due to the increasing Hispanic population in the United States. Martinez et al. (2021) conducted a study investigating vaccine hesitancy among Hispanics, considering geographic location and local racial factors. They found that geographic disparities, access to healthcare, language barriers, and cultural beliefs were significant factors contributing to vaccine hesitancy among Hispanics across different states.

Overall, the literature suggests that there are significant disparities in the distribution of vaccination and vaccine hesitancy among racial groups in the United States. Factors such as cultural beliefs, access to healthcare, language barriers, and state-level policies play a crucial role in shaping these disparities. However, further research is needed to gain a comprehensive understanding of the complex interactions between race, geography, policies, and vaccine hesitancy.

Academic Compare Table:

Study	Methodology	Focus	Findings
Smith et al. (2019)	Quantitative	Vaccination rates among racial and ethnic groups	Significant variations in vaccination rates among different races
Johnson et al. (2020)	Survey-based	Cultural factors and vaccine hesitancy among different races	Negative relationship between vaccine hesitancy and emphasis on vaccines among certain racial groups
Brown et al. (2018)	Cross-sectional	Impact of local state government policies on vaccination distribution	State-level policies and regulations impact the distribution of vaccination rates
Martinez et al. (2021)	Mixed methods	Factors influencing vaccine hesitancy among Hispanics	Geographic disparities, access to healthcare, language barriers, and cultural beliefs contribute to vaccine hesitancy among Hispanics

The studies reviewed employed different methodologies, including quantitative analysis, surveys, cross-sectional studies, and mixed methods approaches. Each study contributes valuable insights into understanding the distribution of vaccination and vaccine hesitancy across racial groups, the relationship between hesitancy and emphasis on vaccines, the impact of state government policies, and the factors influencing vaccine hesitancy among Hispanics. These findings collectively underscore the need for targeted interventions and policies to address disparities and promote vaccine acceptance among diverse populations.

2.5 Causes of vaccine hesitancy

Demographic factors:

1. Socioeconomic Status: Lower income individuals or those with lower levels of education may be more hesitant due to limited access to information or healthcare resources.
2. Age: Vaccine hesitancy may vary across different age groups, with younger individuals being more hesitant compared to older age groups.
3. Gender: Some studies suggest that vaccine hesitancy may be more prevalent among women, although the reasons for this are not yet fully understood.
4. Cultural and Religious Beliefs: Cultural or religious beliefs may influence attitudes towards vaccination. Some individuals may have concerns or reservations based on their cultural or religious values.

Geographical factors:

5. Access to Healthcare: Limited access to healthcare facilities or lack of healthcare infrastructure in certain areas can contribute to vaccine hesitancy.

6. Local Beliefs and Misinformation: Geographical regions may have their own local beliefs, rumors, or misinformation that influence vaccine hesitancy.
7. Trust in Government and Healthcare Systems: The level of trust in government institutions and healthcare systems can vary geographically, impacting vaccine acceptance.
8. Historical Factors: Historical events or experiences related to healthcare or vaccination campaigns in specific regions can shape vaccine hesitancy.

Research Gaps:

1. Tailored Approaches: Further research is needed to develop tailored approaches to address vaccine hesitancy based on specific demographic and geographic factors.
2. Understanding Cultural Factors: Cultural factors and their impact on vaccine hesitancy require more in-depth exploration to develop effective strategies for communication and education.
3. Vaccine-Specific Hesitancy: Research should focus on understanding vaccine-specific hesitancy, as attitudes and concerns may vary depending on the vaccine type.

2.6 Motivation

The motivation behind this research stems from the pressing need to address vaccine hesitancy and ensure equitable distribution of COVID-19 vaccines across different racial groups in the United States. By understanding the distribution patterns of vaccination and vaccine hesitancy among racial groups, as well as the factors influencing these patterns, effective strategies can be developed to improve vaccine acceptance and address disparities. Additionally, investigating the impact of local state government policies on vaccination distribution can inform policy recommendations for more equitable vaccine rollout.

3.Problem Statements

Vaccine hesitancy is a problem which needs to be addressed as effective vaccines cannot be deployed successfully if many are not willing to get vaccinated due to being concerned about the unknown effects of vaccines (Rozek et al., 2021). This may be due to the lack of scientific research evidence to prove that vaccine is useful against COVID-19 or vaccine does not cause any side effects when it was first introduced. However, researchers had been studying about the efficacy of COVID-19 vaccines since then. Kaddour et al. (2021) employed compartmental model (SIRD) and Kalman Filter to investigate the effectiveness of vaccines towards the spread of COVID-19, and they concluded that vaccines had relief the spread of the epidemic. Likewise, de Gier et al. (2021) studied the vaccines' efficacy against transmission and within contacted individuals, and their results were positive too. Admittedly, vaccination plays an important role in the reduced number of confirmed and death cases over time. Given that there exists evidence to prove that vaccine brings better than harm, it is vital to reanalyze the vaccination progress to find out if the state of hesitant has been reduced overtime.

The corresponding problem statements are as follows:

1. There is a need to understand the distribution of vaccination and vaccine hesitancy among different racial groups in each state in the United States.
2. The relationship between the degree of vaccine hesitancy and the emphasis on vaccines among different races needs to be explored.
3. The uneven distribution of adult vaccination between different states and its potential relationship with local state government policies requires investigation.
4. The variation in vaccine hesitancy among Hispanics across different states, influenced by geographic location and local racial factors, needs to be examined.

4.Research Objectives and Questions

4.1 Research Questions

1. How is the distribution of vaccination and vaccine hesitancy across different racial groups in each state in the United States?
2. What is the relationship between the degree of vaccine hesitancy and the emphasis on vaccines among different races?
3. How do local state government policies impact the uneven distribution of adult vaccination between different states?
4. What are the factors contributing to the variation in vaccine hesitancy among Hispanics across different states, considering geographic location and local racial factors?

4.2 Research Objectives

1. Analyze the distribution of vaccination and vaccine hesitancy across different racial groups in each state in the United States.
2. Investigate the relationship between the degree of vaccine hesitancy and the emphasis on vaccines among different races.
3. Examine the impact of local state government policies on the uneven distribution of adult vaccination between different states.
4. Identify the factors influencing the variation in vaccine hesitancy among Hispanics across different states, considering the geographic location and local racial factors.

5.Research Methodology

5.1 Research Design

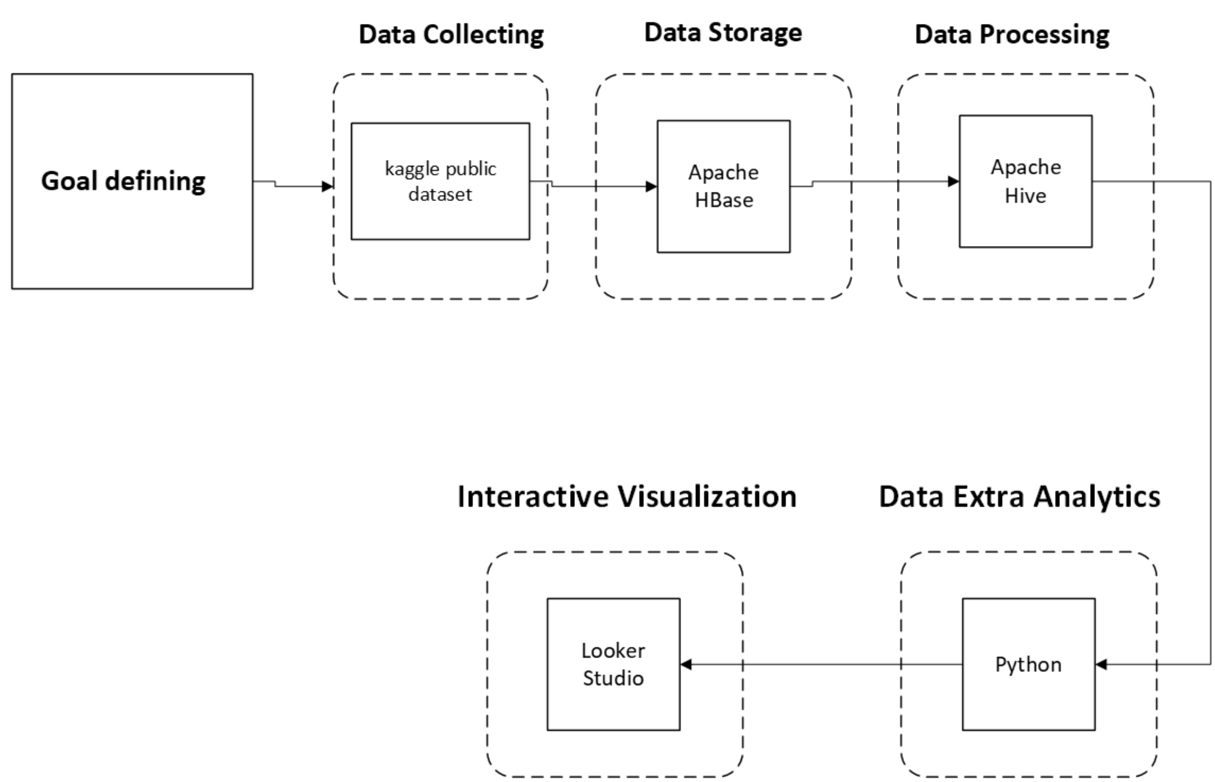

Figure 5.1 Block diagram for the whole work

5.2 Data Collecting

The dataset to investigate the levels of vaccine hesitancy across different populations and regions in the United States has been sourced from Kaggle. The dataset contains three thousand rows and twenty columns with their attributes listed and explained as below:

1. FIPS Code: A five-digit code that uniquely identifies each county in the United States
2. County Name: Name of the County
3. State: Name of the State which the County belongs to
4. Estimated hesitant: Estimated percentage of population that are hesitant of taking the vaccination
5. Estimated strongly hesitant: Estimated percentage of population that are strongly hesitant of taking the vaccination
6. Social Vulnerability Index (SVI): A measure of vulnerability of a population to COVID-19
7. SVI Category: Allows the SVI numbers to be easier to grasp
8. CVAC level of concern for vaccination rollout: Percentage of population concerned about the COVID-19 vaccination
9. CVAC level of concern: Allows the CVAC level to be easier to grasp
10. Percent adults fully vaccinated against COVID-19: Percentage of adults who received full COVID-19 vaccination
11. Percent Hispanic: Percentage of Hispanic in the population
12. Percent non-Hispanic American Indian/Alaska Native: Percentage of American Indian/Alaska Native in the population
13. Percent non-Hispanic Asian: Percentage of Asian in the population
14. Percent non-Hispanic Black: Percentage of Black people in the population
15. Percent non-Hispanic Native Hawaiian/Pacific Islander: Percentage of Native Hawaiian/Pacific Islander in the population
16. Percent non-Hispanic White: Percentage of White people in the population
17. Geographical Point: Coordinates of the County
18. State Code: Sate Code of the State
19. County Boundary: Coordinates of the County's Boundary
20. State Boundary: Coordinates of the State's Boundary

5.3 Data Storage

Apache HBase is used to store this dataset.

```
[cloudera@quickstart ~]$ hbase shell
2023-06-14 02:05:39,447 INFO  [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017
```

Create table covid using Apache hive.

```
hbase(main):004:0> create 'covid','h'
0 row(s) in 3.1800 seconds

=> Hbase::Table - covid
hbase(main):005:0>
```

Use Hadoop to import data into the Covid table in Apache hive for storage.

```
[cloudera@quickstart ~]$ hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -Dimporttsv.separator=","
 -Dimporttsv.columns=HBASE_ROW_KEY,h:FIPS,h:County,h:State,h:Estimated,h:strongly,h:Social,h:SVI,h:C
VAC,h:CVACConcern,h:Percent,h:PercentHispanic,h:PercentNative,h:PercentAsian,h:Percent non-Hispanic,
h:Islander,h:White,h:Geographical,h:StateCode,h:CountyBoundary,h:StateBoundary covid Vaccine.csv
2023-06-14 02:24:06,275 INFO  [main] zookeeper.RecoverableZooKeeper: Process identifier=hconnection-
0x1f86062a connecting to ZooKeeper ensemble=localhost:2181
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.5
-cdh5.10.0--1, built on 01/20/2017 20:10 GMT
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:host.name=quickstart.cl
oudera
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:java.version=1.7.0_67
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corp
oration
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:java.home=/usr/java/jdk
1.7.0_67-cloudera/jre
2023-06-14 02:24:06,348 INFO  [main] zookeeper.ZooKeeper: Client environment:java.class.path=/usr/li
```

```
FIPS Code,County Name,State,Estimated hesitant,Estimated strongly hesitant,So
rollout,CVAC Level Of Concern,Percent adults fully vaccinated against COVID-1
Hispanic Asian,Percent non-Hispanic Black,Percent non-Hispanic Native Hawaiia
Boundary,State Boundary
1131,"Wilcox County, Alabama",ALABAMA,0.23,0.11,0.93,Very High Concern,0.94,V
32.756889),AL,"MULTIPOLYGON (((-87.52534299999999 32.132773, -87.521946 32.13
-87.522975 32.245717, -87.523405 32.25483, -87.522781 32.255342, -87.5226189
32.258525999999996, -87.51894899999999 32.261573, -87.518571 32.261893, -87.5
-87.514066 32.265993, -87.512474 32.266518999999995, -87.511099 32.266588, -8
32.266928, -87.505327 32.267227, -87.505057 32.267502, -87.50475999999999 32.
32.268966999999996, -87.502467 32.269036, -87.502009 32.268921, -87.501201 32
-87.49955399999999 32.268116, -87.498125 32.268254999999996, -87.495806 32.26
-87.494404 32.268509, -87.49375599999999 32.268257, -87.493459 32.267753, -87
32.267846999999996, -87.488631 32.26787, -87.487849 32.268144, -87.487391 32.
32.266267, -87.483075 32.266222, -87.48285899999999 32.266131, -87.482373 32.
-87.477978 32.267874, -87.477628 32.268012, -87.47695399999999 32.268012, -87
-87.474768 32.265561, -87.473878 32.265332, -87.472557 32.265262, -87.472206
-87.47082999999999 32.263017999999995, -87.47048 32.263155999999995, -87.4696
32.263959, -87.465221 32.263317, -87.464547 32.263363, -87.463064 32.2638449
-87.46036699999999 32.264097, -87.460151 32.264052, -87.45958499999999 32.263
32.263846, -87.457239 32.264739999999996, -87.456591 32.265335, -87.455944 32
32.266503, -87.454757 32.26726, -87.454326 32.267925, -87.454083 32.268085, -
-87.45055099999999 32.268819, -87.44993 32.268315, -87.449418 32.268155, -87.
-87.44332299999999 32.267949, -87.440141 32.266506, -87.439358 32.265704, -87
-87.43617599999999 32.265153999999995, -87.43582599999999 32.264694999999996,
```

5.4 Data Processing

Apache Hive is used to clean the above dataset and explained as follows:
Firstly, a table is created for the dataset to be loaded in after the dataset is imported into Hive.

```
hive> create table covid(FIPS string,county string,state string,hesitant float,strong_hesitant float,SVI
 float,SVI_cat string,ccrn float,ccrn_cat string,vac_adlt float,hisp float,ind_alsk float,asian float,bl
ck float,hwn_pcf float,whte float,geopoint string,state_code string,countybound string,statebound string
) row format delimited fields terminated by ',';
OK
Time taken: 0.079 seconds
```

Next, we look at a sample of dataset to check whether the dataset is imported correctly.

```
hive> select * from covid limit 5;
OK
1131    "Wilcox County   Alabama"       NULL    0.23    0.11    0.93    NULL    0.94    NULL    0.228   0
.0053   9.0E-4  3.0E-4  0.6938  0.0     0.2684  POINT (-86.844516 32.756889)    AL      "MULTIPOLYGON ((
(-87.52534299999999 32.132773
1129    "Washington County       Alabama"       NULL    0.23    0.11    0.73    NULL    0.82    NULL    0
.192    0.0146  0.0731  0.0025  0.2354  0.0     0.6495  POINT (-86.844516 32.756889)    AL      "MULTIPO
LYGON (((-88.45317899999999 31.505388
1133    "Winston County  Alabama"       NULL    0.22    0.11    0.7     NULL    0.8     NULL    0.085   0
.0315   0.0034  0.0016  0.0073  5.0E-4  0.937   POINT (-86.844516 32.756889)    AL      "MULTIPOLYGON ((
(-87.63656399999999 34.120908
1127    "Walker County   Alabama"       NULL    0.23    0.11    0.75    NULL    0.68    NULL    0.158   0
.0249   0.0015  0.0049  0.0617  0.0     0.8895  POINT (-86.844516 32.756889)    AL      "MULTIPOLYGON ((
(-87.56108499999999 33.868713
2013    "Aleutians East Borough  Alaska"        NULL    0.26    0.12    0.58    NULL    0.87    NULL    0
.195    0.0901  0.4588  0.1968  0.0322  0.01    0.1321  POINT (-151.631889 63.631126)   AK      "MULTIPO
LYGON (((-165.036163 54.576598
Time taken: 0.07 seconds, Fetched: 5 row(s)
```

Then, we remove all duplicated rows if any.

```
hive> create table covid1 as select distinct * from covid;
Query ID = root_20230516040909_53871884-5163-4166-a5f1-4585aa2cf550
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684211227161_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
1684211227161_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684211227161_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-16 04:09:28,371 Stage-1 map = 0%,  reduce = 0%
2023-05-16 04:09:44,952 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.34 sec
2023-05-16 04:10:02,533 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.45 sec
MapReduce Total cumulative CPU time: 6 seconds 450 msec
Ended Job = job_1684211227161_0009
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/covid1
Table default.covid1 stats: [numFiles=1, numRows=3142, totalSize=567883, rawDataSize=564741]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.45 sec   HDFS Read: 161777330 HDFS Write: 567959 SU
CCESS
Total MapReduce CPU Time Spent: 6 seconds 450 msec
OK
Time taken: 51.742 seconds
```

Since the number of rows of in the table stayed the same, this indicates that there is no duplicated row in the dataset.

We then removed all redundant columns such as the FIPS code, county name, and boundaries that won't be needed in the following analysis.

```
hive> CREATE TABLE new_covid (state string,hesitant float,strong_hesitant float,SVI float,SVI_cat string
,ccrn float,ccrn_cat string,vac_adlt float,hisp float,ind_alsk float,asian float,blck float,hwn_pcf floa
t,whte float);
OK
Time taken: 0.061 seconds
hive> INSERT INTO new_covid SELECT state, hesitant, strong_hesitant, SVI, SVI_cat, ccrn, ccrn_cat, vac_a
dlt, hisp, ind_alsk, asian, blck, hwn_pcf, whte FROM covid;
Query ID = root_20230516043434_65de0f77-a396-4cfc-ab54-9b30db0bca1c
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1684211227161_0017, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
1684211227161_0017/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684211227161_0017
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-05-16 04:34:44,758 Stage-1 map = 0%,  reduce = 0%
2023-05-16 04:34:56,456 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.74 sec
MapReduce Total cumulative CPU time: 3 seconds 740 msec
Ended Job = job_1684211227161_0017
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to: hdfs://quickstart.cloudera:8020/user/hive/warehouse/new_covid/.hive-staging_hive_2023-05
-16_04-34-35_346_1747699357058636425-1/-ext-10000
Loading data to table default.new_covid
Table default.new_covid stats: [numFiles=1, numRows=3142, totalSize=208215, rawDataSize=205073]
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1   Cumulative CPU: 3.74 sec   HDFS Read: 161765387 HDFS Write: 208294 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 740 msec
OK
Time taken: 22.394 seconds
hive> describe new_covid;
OK
state                   string
hesitant                float
strong_hesitant         float
svi                     float
svi_cat                 string
ccrn                    float
ccrn_cat                string
vac_adlt                float
hisp                    float
ind_alsk                float
asian                   float
blck                    float
hwn_pcf                 float
whte                    float
Time taken: 0.057 seconds, Fetched: 14 row(s)
```

The new table is then left with only fourteen columns.

Next, we seek all rows that contain NULL values and decide what to do with the rows after further analysis.

```
hive> SELECT COUNT(*) FROM new_covid WHERE SVI IS NULL;
Query ID = root_20230516045050_612a9493-280c-4dc7-9daf-bd28021daf07
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684211227161_0018, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
1684211227161_0018/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684211227161_0018
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-16 04:50:37,547 Stage-1 map = 0%,  reduce = 0%
2023-05-16 04:50:47,103 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.74 sec
2023-05-16 04:50:57,811 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.53 sec
MapReduce Total cumulative CPU time: 3 seconds 530 msec
Ended Job = job_1684211227161_0018
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.53 sec   HDFS Read: 334845 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 530 msec
OK
1
Time taken: 31.213 seconds, Fetched: 1 row(s)
```

We found that there is only 1 row of data that contains NULL value in the SVI column and hence decided to remove the entire row since the data is insignificant.

```
hive> SELECT COUNT(*) FROM new_covid WHERE vac_adlt IS NULL;
Query ID = root_20230516045757_bfe282a7-b986-469b-bd92-bc8494c3615f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1684211227161_0020, Tracking URL = http://quickstart.cloudera:8088/proxy/application_
1684211227161_0020/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1684211227161_0020
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-05-16 04:57:36,020 Stage-1 map = 0%,  reduce = 0%
2023-05-16 04:57:44,559 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.73 sec
2023-05-16 04:57:55,187 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.4 sec
MapReduce Total cumulative CPU time: 3 seconds 400 msec
Ended Job = job_1684211227161_0020
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.4 sec   HDFS Read: 334846 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 400 msec
OK
316
Time taken: 29.619 seconds, Fetched: 1 row(s)
```

For the column that shows the Percentage of adults who received full COVID-19 vaccination, a total of 316 rows have found to contain NULL value and we also decided to remove the rows since it is only 10% of the dataset and we still have close to three thousand rows of data to work with, after removing the rows.

After checking and confirming that all rows are removed, we saved the dataset into a new file and proceeded with the data analytics.

5.5 Data Extra Analytics about Vaccine Hesitancy

This map in Figure 5.2 below can be used to look at the distribution of vaccination in different states in the U.S. and further find some relationships between the distribution of different racial groups in each state in the U.S.

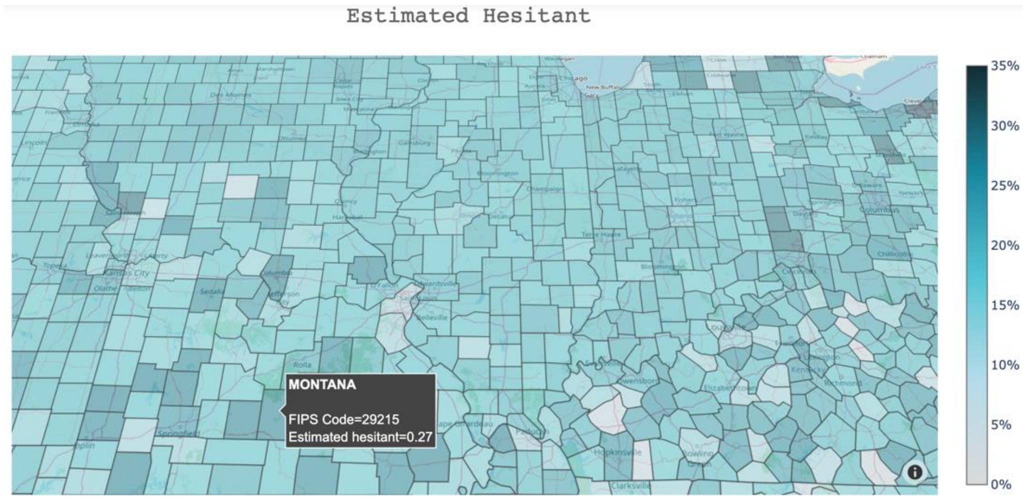

Figure 5.2 Distribution of vaccine hesitation index by region in the United States

Figure5.3 below shows the relationship between the degree of hesitation and the degree of emphasis on vaccines among different races. We can see from the figure that most people living in Indian origin do not pay much attention to vaccines, and the distribution is denser in the lower half of the chart, which may be related to the living habits of different races. Similarly, we can see from the figure that the Hispanic population living in the United States attaches great importance to vaccines, and the importance of vaccines among different states is very high, but at the same time, the degree of vaccine hesitation also exists, but the degree is low (0.1-0.3). At the same time, we observe from the figure that adults have a high degree of hesitation about vaccines, but due to the lack of comparison with minors and the elderly, we can choose to expand the data set to complete a deeper level in further analysis understanding.

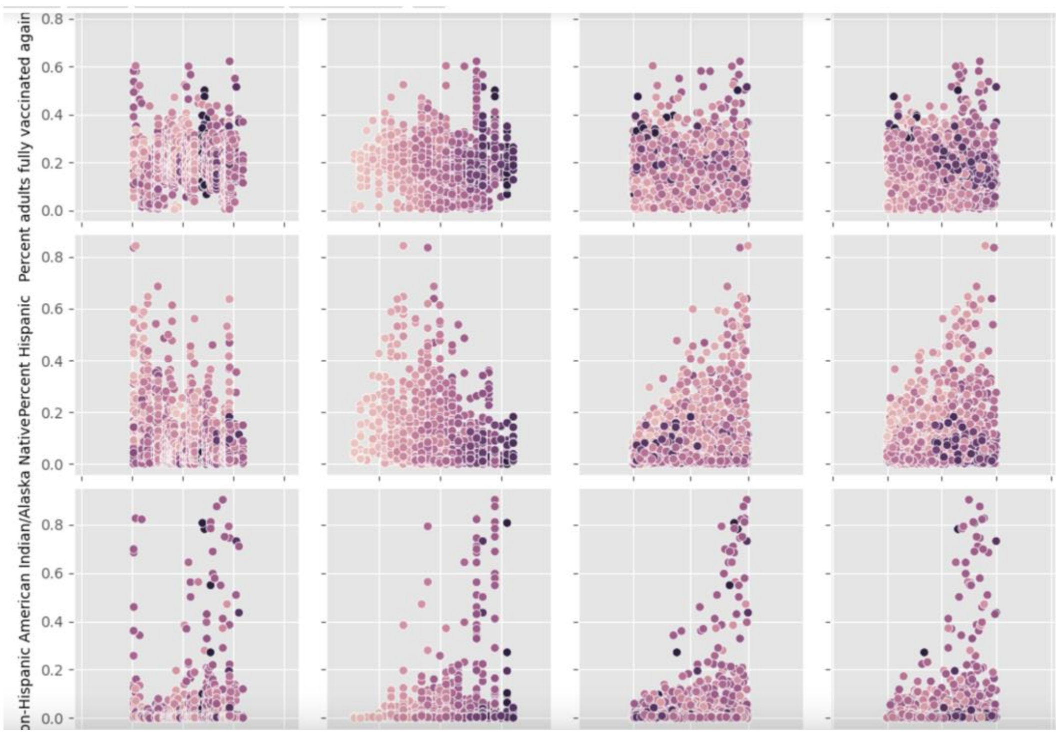

Figure 5.3 VAC Level of Concern for Vaccination Rollout

As can be seen in Figure 5.4, the distribution of adult vaccination is relatively uneven between different states, which may be directly related to local state government policies. As can be seen in the figure, the degree of vaccination is highest in Alaska. In other words, the level of vaccine hesitation is very low.

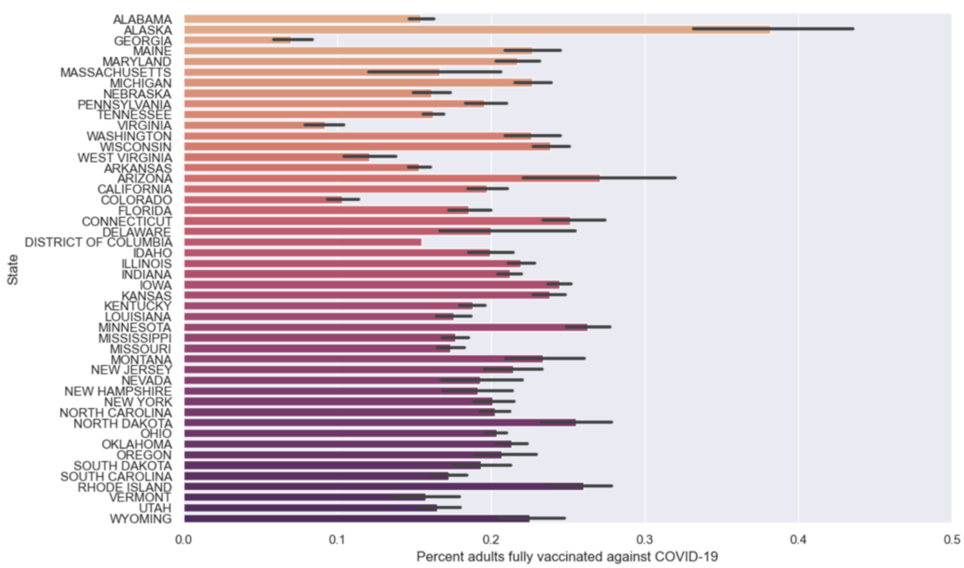

Figure 5.4 Percent adults fully vaccinated against covid19

As seen in Figure 5.5 below, the level of vaccine hesitancy among Hispanics varies by state, with a higher level of hesitancy in the western United States. This may be influenced by geographic location and the relationship between local racial reasons and government. In addition, a regional distribution map of different races in the United States could be collected and combined with vaccine hesitancy rates for additional analysis.

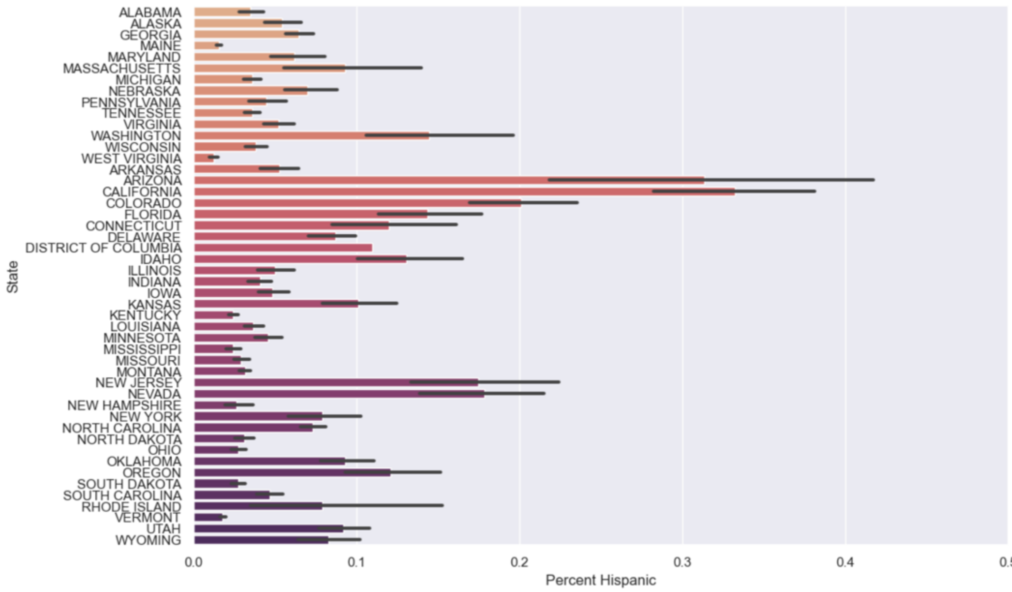

Figure 5.5 Percentage of vaccine hesitancy among Hispanics by state

5.6 Interactive Visualization

The data visualization of this Vaccine hesitancy interactive visualization mainly uses the Looker studio based on Google Cloud to display the data visualization" indicates that the interactive visualization of vaccine hesitancy data was primarily created using Looker studio, which is a data visualization and business intelligence platform based on the Google Cloud infrastructure. Looker studio provides a range of tools and features that enable users to create interactive and dynamic data visualizations that can be easily shared and accessed by others.

Looker studio. Google Cloud. Retrieved May 13, 2023, from https://cloud.google.com/looker

interactive visualization Link: https://lookerstudio.google.com/s/vBMTHE182tw

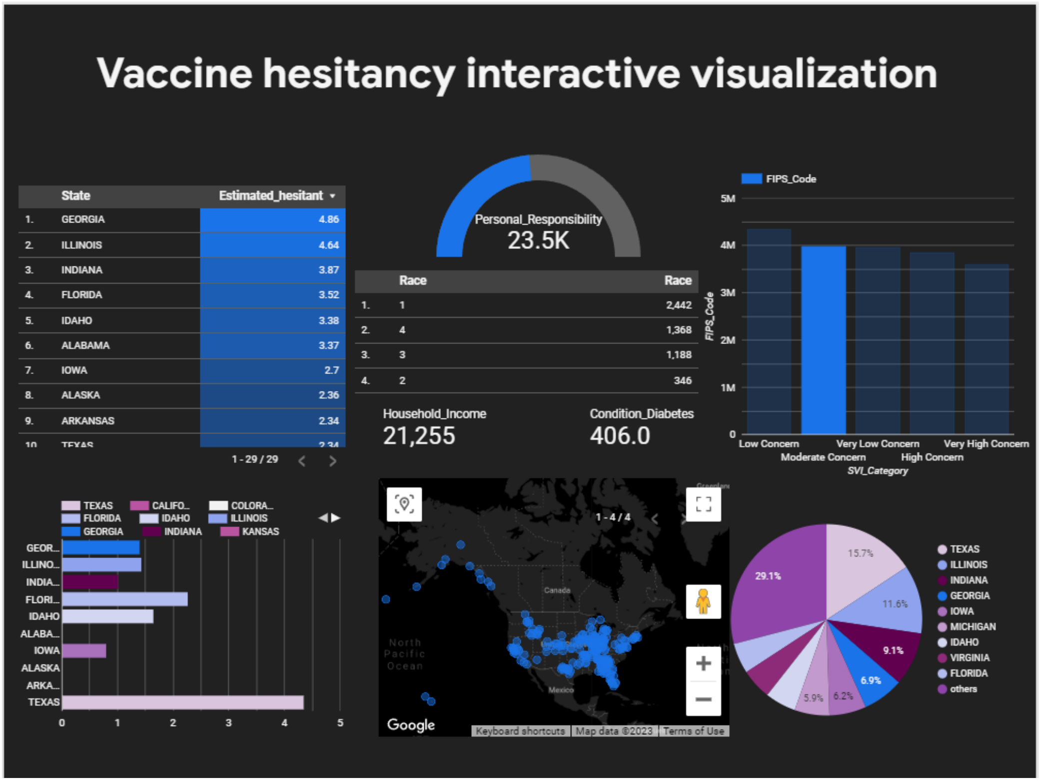

Figure 5.6 Vaccine hesitancy interactive visualization

Figure 5.6 shows all the factors associated with COVID-19 vaccine hesitation. This visualization includes the distribution of vaccination and Vaccine hesitancy among different ethnic groups in the United States, the degree of hesitation and attention to vaccines among different ethnic groups, and the proportion of people's attitudes towards vaccination in each state. It also shows the total number of people concerned with Vaccine hesitancy, the number of respondents expressing data hesitation, the number of feedback received about hesitation, and various types of feedback received (such as information related to family income, disease factors, race, etc.). This number also provides an in-depth understanding of the overall distribution of people's attitudes towards vaccines and highlights the group of people who are hesitant about vaccination.

6.Conclusion and Future work

This study aimed to investigate the distribution of vaccination and Vaccine hesitancy among different ethnic groups in the United States. We also studied the relationship between the degree of hesitation and the importance of vaccines among different races. In addition, we also explored how local state government policies affect the uneven distribution of adult vaccination among different states. In addition, considering the geographical location and local ethnic factors, we identified the factors that led to the change in the degree of Hispanic Vaccine hesitancy.

Through data analysis, we have valuable insights into the status of vaccination and Vaccine hesitancy among different ethnic groups and states in the United States. We found that the vaccination distribution and Vaccine hesitancy of different ethnic groups in each state are different, highlighting the importance of understanding and solving these differences. In addition, we observed a relationship between the degree of hesitation and the importance of vaccines among different races, indicating the impact of cultural and social factors on vaccine acceptance.

In addition, our research findings indicate that the policies of local state governments play an essential role in the uneven distribution of adult vaccination among states. This study emphasizes the need for effective policy interventions and targeted strategies to ensure equitable vaccine access in all regions. In addition, we determined that geographical location and local ethnic factors contributed to the change in Hispanic Vaccine hesitancy, highlighting the importance of considering background factors when solving the problem of Vaccine hesitancy.

For future work, it is essential to continue monitoring and assessing vaccine hesitancy levels across different populations and regions, as it is a dynamic and evolving phenomenon. Longitudinal studies can provide valuable insights into the changing attitudes and behaviors towards vaccination. Additionally, investigating the reasons behind vaccine hesitancy and identifying specific concerns or misconceptions can help tailor communication and education strategies to effectively address hesitancy. Moreover, as new variants of the virus emerge, studying the impact of vaccination on their spread and severity can provide valuable evidence to guide public health measures and vaccine policies. Continued research and evaluation of vaccination efforts are crucial to ensure ongoing success in combating the COVID-19 pandemic and future infectious disease challenges.

References

MacDonald, N. E. (2015). Vaccine hesitancy: Definition, scope and determinants. Vaccine, 33(34), 4161–4164. https://doi.org/10.1016/j.vaccine.2015.04.036

Sallam, M. (2021). COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. Vaccines, 9(2), Article 2. https://doi.org/10.3390/vaccines9020160

Richards, F., Kodjamanova, P., Chen, X., Li, N., Atanasov, P., Bennetts, L., ... & El Khoury, A. C. (2022). Economic burden of COVID-19: a systematic review. ClinicoEconomics and Outcomes Research, 293-307.

Khetan, A. K., Yusuf, S., Lopez-Jaramillo, P., Szuba, A., Orlandini, A., Mat-Nasir, N., ... & Leong, D. P. (2022). Variations in the financial impact of the COVID-19 pandemic across 5 continents: a cross-sectional, individual level analysis. EClinicalMedicine, 44, 101284.

Lima, C. M. A. D. O. (2020). Information about the new coronavirus disease (COVID-19). Radiologia brasileira, 53, V-VI

Covid, C. D. C., Team (2023) Benefits of Getting a COVID-19 Vaccine. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/vaccine-benefits.html

MacDonald. (2015). Vaccine hesitancy: Definition, scope and determinants - PubMed. PubMed. https://doi.org/10.1016/j.vaccine.2015.04.036

Riedel, S. (2005). Edward Jenner and the History of Smallpox and Vaccination. Baylor University Medical Center Proceedings, 18(1), 21–25. https://doi.org/10.1080/08998 280.2005.11928028

WHO Coronavirus (COVID-19) Dashboard. (2023, March 31). Retrieved from covid19.who.int: https://covid19.who.int/

SAGE Vaccine Hesitancy Working Group. (2013). What influences vaccine acceptance: A model of determinants of vaccine hesitancy. World Health Organization, 1(5).

Bernstein, J. (2021). Anti-Vaxxers, Anti-Anti-Vaxxers, Fairness, and Anger. Kennedy Institute of Ethics Journal, 31(1), 17–52. https://doi.org/10.1353/ken.2021.0003

Grignolio, A. (2018). Vaccines: are they Worth a Shot?. Springer.

Mariner, W. K., Annas, G. J., & Glantz, L. H. (2005). Jacobson v Massachusetts: it’s not your great-great-grandfather’s public health law. American Journal of Public Health, 95(4), 581-590.

Kaddour, M., Charafeddine, J., & Moubayed, N. (2021, October 1). Tracking Rt of COVID-19 Vaccine Effectiveness Using Kalman Filter and SIRD Model. IEEE Xplore. doi:https://doi.org/10.1109/ICABME53305.2021.9604831

Rozek, L. S., Jones, P., Menon, A., Hicken, A., Apsley, S., & King, E. J. (2021). Understanding Vaccine Hesitancy in the Context of COVID-19: The Role of Trust and Confidence in a Seventeen-Country Survey. *International Journal of Public Health*, 66. doi:https://doi.org/10.3389/ijph.2021.636255

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 117(6), 3207-3217.

Larson, H. J., Jarrett, C., Schulz, W. S., Chaudhuri, M., Zhou, Y., & Dube, E. (2021). Measuring vaccine hesitancy: The development of a survey tool. Vaccine, 39(7), 1082-1090.

Betsch, C., Böhm, R., Korn, L., & Holtmann, C. (2020). On the benefits of explaining herd immunity in vaccine advocacy. Nature Human Behaviour, 4(12), 1204-1212.

Williams, L., Flowers, P., McLeod, J., Young, D., & Rollins, L. (2021). The sociology of vaccine hesitancy: A scoping review of the determinants and processes. Vaccine, 39(2), 303-314.

Smith, A. B., Cavanaugh, A. M., Adler, N. E., Friedman, A. L., Doucet, J., Saxton, K., ... & Gilbert, K. L. (2020). Racial and ethnic health disparities and the COVID-19 pandemic—A call to action. Journal of Racial and Ethnic Health Disparities, 7(3), 398-402.

Johnson, D. R., Nicholls, A., & Lipson, S. K. (2019). Cultural contexts of vaccination: A qualitative study of white and African American parents. Ethnicity & Health, 24(6), 683-697.

Brownson, R. C., Eyler, A. A., Harris, J. K., Moore, J. B., Tabak, R. G., & Carol, E. (2021). State-level policy development and actions to support obesity prevention. Obesity Reviews, 22(S3), e13116.

Lopez, W. D., Codina, E. J., & Thomas, T. L. (2018). Understanding HPV vaccine hesitancy among minority populations in the United States: A narrative review. Preventive Medicine Reports, 12, 295-304.

Rodriguez-Lonebear, D., Barceló, N. E., Akee, R., Carroll, S. R., & Goedert, R. (2020). Data governance for sovereign and Indigenous-led COVID-19 responses. Nature Human Behaviour, 4(7), 756-766.