Manuscript for a Data Analysis Project

Katie Wells & Connor Ross

3/16/23

Warning: package 'here' was built under R version 4.2.2

Warning: package 'knitr' was built under R version 4.2.2

Warning: package 'tidyverse' was built under R version 4.2.2

Warning: package 'ggplot2' was built under R version 4.2.2

Warning: package 'tidyr' was built under R version 4.2.2

Warning: package 'readr' was built under R version 4.2.2

Warning: package 'purrr' was built under R version 4.2.2

Warning: package 'dplyr' was built under R version 4.2.2

Warning: package 'stringr' was built under R version 4.2.2

Warning: package 'forcats' was built under R version 4.2.2

Warning: package 'lubridate' was built under R version 4.2.2

Warning: package 'tidymodels' was built under R version 4.2.2

Warning: package 'dials' was built under R version 4.2.2

Warning: package 'scales' was built under R version 4.2.2

Warning: package 'infer' was built under R version 4.2.2

Warning: package 'modeldata' was built under R version 4.2.2

Warning: package 'parsnip' was built under R version 4.2.2

Warning: package 'recipes' was built under R version 4.2.2

Warning: package 'rsample' was built under R version 4.2.2

Warning: package 'tune' was built under R version 4.2.2

Warning: package 'workflows' was built under R version 4.2.2

Warning: package 'workflowsets' was built under R version 4.2.2

Warning: package 'yardstick' was built under R version 4.2.2

# 1. Introduction

## 1.1 General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis*

## 1.2 Description of data and data source

This dataset contains US COVID-19 vaccine deliveries and administration data at national and jurisdiction level from all vaccine partners, including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities. It was obtained from https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc. It has 109 columns and 37,500 rows of administration, distribution, and series data of COVID-19 vaccinations across the US.

## 1.3 Questions/Hypotheses to be addressed

Question: What factors (like region) influenced COVID-19 vaccine distribution across the US? Hypotheses: H0 - There is not a difference in COVID-19 vaccine (across manufacturers) popularity across all the regions in the United States. HA - There is differences in COVID-19 vaccine (manufacturer) popularity across the regions in the United States.

# 2. Summary/Abstract

*Write a summary of your project.*

# 3. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

## 3.1 Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

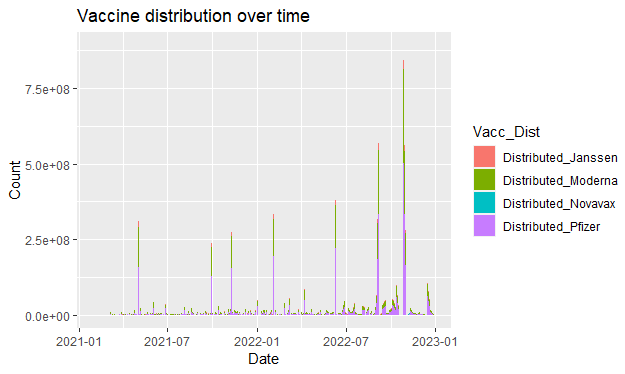
## 3.2 Statistical analysis

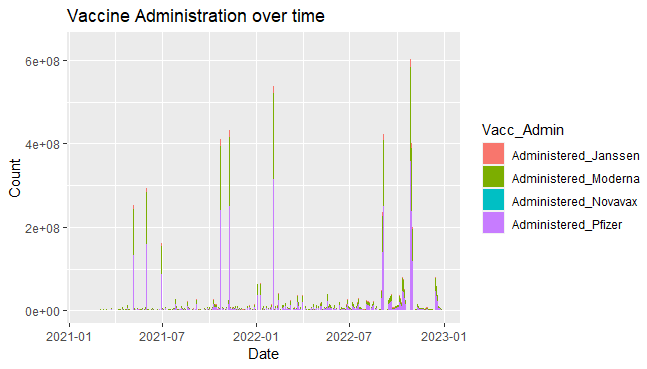
After placing each state in a region of the US, we will conduct a time series analysis to determine the change in popularity over time. Further, we intended to plot overall administration of manufacturer-specific vaccines in each region to determine popularity after adjusting for distribution and time.

# 4. Results

## 4.1 Exploratory/Descriptive analysis

Here are the main results from our exploratory analysis where we plotted the distributed and administered vaccines over the observed period. Individual bar graphs for each region can be found in the ‘exploration.qmd’ file in the ‘processing\_code’ file.





## 4.2 Basic statistical analysis

Here are the main results from our initial fitting analysis. Note: we may not keep all of these fits in our final project but we calculated them all to view the results. Further checking of the models can be found in the ‘fitting.qmd’ in the ‘analysis\_code’ folder.

resulttable1 <- readRDS(here("code", "analysis\_code", "lm7.rds"))  
kable(resulttable1)

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -3.967848e+04 | 11621.815926 | -3.414138 | 0.0006408 |
| RegionNorthEast | 2.105466e+05 | 15658.150692 | 13.446454 | 0.0000000 |
| RegionSouthEast | -3.884536e+05 | 15043.823274 | -25.821468 | 0.0000000 |
| RegionSouthWest | -3.807442e+05 | 21597.652416 | -17.628964 | 0.0000000 |
| RegionWest | 2.474770e+05 | 16052.402876 | 15.416821 | 0.0000000 |
| Distributed | 8.055944e-01 | 0.000401 | 2009.211299 | 0.0000000 |

resulttable2 <- readRDS(here("code", "analysis\_code", "lm8.rds"))  
kable(resulttable2)

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | 7072182.7 | 138731.2 | 50.977583 | 0.0000000 |
| RegionNorthEast | 853523.1 | 196195.6 | 4.350369 | 0.0000136 |
| RegionSouthEast | 225224.6 | 188498.6 | 1.194834 | 0.2321627 |
| RegionSouthWest | 4913232.6 | 268651.9 | 18.288474 | 0.0000000 |
| RegionWest | 1437161.7 | 201040.6 | 7.148613 | 0.0000000 |

resulttable3 <- readRDS(here("code", "analysis\_code", "lm9.rds"))  
kable(resulttable3)

| term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- |
| (Intercept) | -6.876137e+04 | 6922.7485546 | -9.932669 | 0 |
| Distributed | 8.048692e-01 | 0.0004176 | 1927.469243 | 0 |

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# 6. References