Analyzing Vaccine Supply in Texas

2021-03-29

Table of Contents

# Analysis Setup

Before we start building out our reproducible analysis, let’s go ahead and make sure any R packages are loaded and installed properly. The code to install necessary packages and load them can be viewed by clicking on the “Show Code” arrow.

knitr::opts\_chunk$set(warning = FALSE, message = FALSE)  
# In case these aren't installed, uncomment this and run it.  
# install.packages("janitor", "tidyverse", "gt")  
# devtools::install\_github("utexas-lbjp-data/lbjdata")  
  
library(janitor) # Package with useful + convenient data cleaning functions  
library(tidyverse) # Core Set of R Data Science Tools (dplyr, ggplot2, tidyr, readr, etc.)

# Analysis

## Import Our Vaccine Provider and Supply Data

This data comes from the Texas Department of State Health Services and contains the list of vaccine providers across the state of Texas, which can be found on [this page](https://dshs.texas.gov/coronavirus/additionaldata/). They use it for their own interactive mapping application of vaccine provider sites.[[1]](#footnote-22) Each provider is assigned a type and has a report of how much vaccine supply they have for each of the three approved vaccines. We’ll use the read\_csv() function to read in the data straight from the DSHS website. This will help make sure our analysis is “living”, meaning any chart we make will update whenever the feed from DSHS gets updated, and “reproducible”, meaning anyone who takes this R Markdown document can run it in their RStudio IDE and get the exact same thing you did.

The read\_csv() comes from the [readr](https://readr.tidyverse.org) package that was loaded when we ran library(tidyverse) in the setup chunk above (lines 18:30 in the RMarkdown document).

provider\_data\_raw <- readr::read\_csv("https://genesis.soc.texas.gov/files/accessibility/vaccineprovideraccessibilitydata.csv") %>%   
 janitor::clean\_names() # This function makes column headers machine readable  
  
dplyr::glimpse(provider\_data\_raw) # glimpse() lets you preview a data object

## Rows: 3,380  
## Columns: 17  
## $ name <chr> "Hamarsyl Pharmacy LLc", "Baylor Scott & White Cl…  
## $ type <chr> "Pharmacy", "Other", "Pharmacy", "Pharmacy", "Com…  
## $ tsa <chr> NA, NA, NA, NA, NA, "C", NA, NA, NA, NA, NA, NA, …  
## $ street <chr> NA, "2608 BROCKTON DR", "Bill Mack Rd", "1305 N M…  
## $ city <chr> "McAllen", "AUSTIN", "Shamrock", "Vidor", "Pearla…  
## $ county <chr> NA, NA, "Wheeler", "Orange", "Brazoria", "Wilbarg…  
## $ address <chr> "2528 Buddy Owens Blvd, McAllen, TX 78504", "2608…  
## $ zip <chr> "78504", "78758", "79079", "77662", "77581", "763…  
## $ last\_update\_vac <chr> NA, "03/16/2021", "03/29/2021", "03/29/2021", "03…  
## $ last\_update\_time\_vac <time> NA, 07:17:37, 11:13:52, 13:21:14, 08:39:14…  
## $ pfizer\_available <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ moderna\_available <dbl> 0, 0, 150, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ jj\_available <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0…  
## $ vaccines\_available <dbl> 0, 0, 150, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,…  
## $ total\_shipped <dbl> 100, 100, 500, 300, 500, 2300, 100, 1400, 1000, 1…  
## $ public\_phone <chr> "(956) 627-4358", "512-654-4050", NA, "(409) 769-…  
## $ website <chr> "https://www.hamarsylpharmacy.com/", "bswhealth.c…

## Transform our Vaccine Data

Now that we’ve imported it and created a data object called provider\_data\_raw, we can call on that object and use a handful of functions from the [dplyr](https://dplyr.tidyverse.org) package to transform our data into the shape we want for visualizing.

**The question we’ll trying to answer is simple**: “Among all providers, how much of each vaccine exists in Texas?”

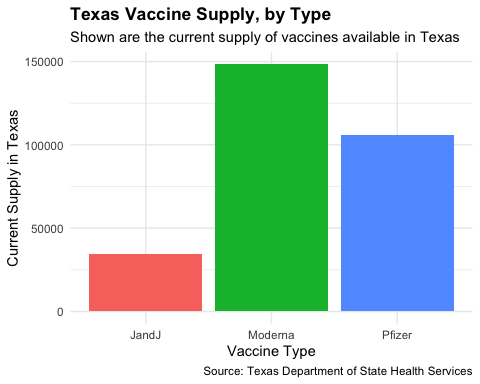
supply\_data <- provider\_data\_raw %>%   
 dplyr::mutate(state = "Texas") %>% # This adds a column where every entry is the word "Texas"  
 dplyr::group\_by(state) %>% # This groups any future functions I write by the state column I created  
 dplyr::summarise( # This begins the summarise() function  
 Pfizer = sum(pfizer\_available), # Creates a column with all pfizer supply  
 Moderna = sum(moderna\_available), # Creates a column with all pfizer supply  
 JandJ = sum(jj\_available) # Creates a column with all pfizer supply  
 ) %>% # This ends the summarise() function  
 tidyr::pivot\_longer(cols = c(Pfizer, Moderna, JandJ), # reshapes our data from wide to long  
 names\_to = "vaccine\_type",  
 values\_to = "supply")  
  
dplyr::glimpse(supply\_data) # glimpse() lets you preview a data object

## Rows: 3  
## Columns: 3  
## $ state <chr> "Texas", "Texas", "Texas"  
## $ vaccine\_type <chr> "Pfizer", "Moderna", "JandJ"  
## $ supply <dbl> 106189, 148417, 34755

## Visualize our Vaccine Data

Now that our data’s in shape, we’ll make a simple bar chart to show the distribution of vaccine supply in Texas.

supply\_chart <- supply\_data %>% # Call on the data  
 ggplot2::ggplot() + # Draw A Chart Canvas  
 ggplot2::aes(x = vaccine\_type, y = supply, fill = vaccine\_type) + # Define How Data Gets Mapped  
 ggplot2::geom\_col() + # Translate into a bar chart format  
 ggplot2::theme\_minimal() + # Add a basic ggplot2 theme  
 ggplot2::theme(legend.position = "none", # Hide the legend  
 plot.title = element\_text(face = "bold")) + # Make the title bold  
 ggplot2::labs(title = "Texas Vaccine Supply, by Type", # Add a title  
 subtitle = "Shown are the current supply of vaccines available in Texas", # Add a subtitle  
 caption = "Source: Texas Department of State Health Services", # Add a caption  
 x = "Vaccine Type", # Add an X axis title  
 y = "Current Supply in Texas") # Add a Y axis title  
   
supply\_chart



## Export our Transformed Dataset and Visualization

Now that we’ve done all of this, we want to share our data and our chart, so we’ll use a couple of functions to save this each time we run it.

## Export Our Data to a CSV File For Sharing  
readr::write\_csv(supply\_data, "clean\_supply\_data.csv")  
  
## Export Our Chart to a PNG File For Sharing  
ggplot2::ggsave("vaccine\_supply\_chart.png", supply\_chart, device = "png", dpi=300, width = 10, height = 6)

# Bonus

## Regression Example

### Regresssion Table

# install.packages("modelsummary") # Uncomment this if you have not installed modelsummary  
library(modelsummary) # Load the {modelsummary package}  
  
model\_1 <- lm(formula=total\_shipped ~ type, # Run a regression using base R  
 data=provider\_data\_raw)  
  
modelsummary::modelsummary(model\_1, stars = TRUE) # Show regression results in a table

|  | Model 1 |
| --- | --- |
| (Intercept) | 1712.051\*\* |
|  | (795.302) |
| typeHospital | 2570.499\*\* |
|  | (1062.232) |
| typeLocal Health Department | 1439.045 |
|  | (1826.421) |
| typeMedical Practice | -903.311 |
|  | (998.975) |
| typeOther | -680.865 |
|  | (934.748) |
| typePharmacy | -615.130 |
|  | (897.054) |
| typeVaccine Hub | 65322.357\*\*\* |
|  | (1659.655) |
| Num.Obs. | 3380 |
| R2 | 0.371 |
| R2 Adj. | 0.370 |
| AIC | 74160.5 |
| BIC | 74209.5 |
| Log.Lik. | -37072.263 |
| F | 331.125 |
| \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01 | |

### Regression Chart

modelsummary::modelplot(model\_1) + # Draw a chart using modelsummary package   
 ggplot2::theme\_dark() + # Add ggplot2 dark theme  
 ggplot2::theme(legend.position = "none", # Hide legend  
 plot.title = element\_text(face = "bold")) + # Make title bold  
 ggplot2::labs(title = "Regression Chart: total\_shipped ~ type", # Add a title  
 subtitle = "How do vaccine shipments and provider type relate?", # Add a subtitle  
 caption = "Source: Texas Department of State Health Services", # Add a caption note  
 x = "", # Add a title for the X Axis  
 y = "") # Add a title for the Y Axis

### Regression Equation

# install.packages("equatiomatic")  
  
equatiomatic::extract\_eq(model\_1) # Extract LaTeX equation with equatiomatic package

$$
\operatorname{total\\_shipped} = \alpha + \beta\_{1}(\operatorname{type}\_{\operatorname{Hospital}}) + \beta\_{2}(\operatorname{type}\_{\operatorname{Local\ Health\ Department}}) + \beta\_{3}(\operatorname{type}\_{\operatorname{Medical\ Practice}}) + \beta\_{4}(\operatorname{type}\_{\operatorname{Other}}) + \beta\_{5}(\operatorname{type}\_{\operatorname{Pharmacy}}) + \beta\_{6}(\operatorname{type}\_{\operatorname{Vaccine\ Hub}}) + \epsilon
$$

1. The link for this map is google.com [↑](#footnote-ref-22)