Analyzing Vaccine Supply in Texas

true

2021-03-29

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. They use it for their own interactive mapping application of vaccine provider sites.[1] 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` 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/vaccineprovidera
  janitor::clean_names() # This function makes column headers machine readable

dplyr::glimpse(provider_data_raw) # glimpse() lets you preview a data object
```

```
## Rows: 3,382
## Columns: 17
## $ name                 <chr> "HILL COUNTRY MEMORIAL HOSP - FREDERICKSBURG", "A~
## $ type                 <chr> "Vaccine Hub", "Other", "Other", "Other", "Medica~
## $ tsa                  <chr> "P", NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ street               <chr> "1020 S State Highway 16", "Main St", "3322 E Wal~
## $ city                 <chr> "Fredericksburg", "Jourdanton", "Pearland", "Lock~
## $ county               <chr> "Gillespie", "Atascosa", NA, NA, NA, NA, NA, "Col~
```

[1] The link for this map is google.com

```
## $ address              <chr> "1020 S State Highway 16", "915 Main St", "3322 E~
## $ zip                  <chr> "78624", "78026", "77581", "78644", "76028", "793~
## $ last_update_vac      <chr> "03/28/2021", "03/24/2021", "03/24/2021", NA, "03~
## $ last_update_time_vac <time> 21:41:18, 08:44:35, 14:58:56,       NA, 19:45:27~
## $ pfizer_available     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ moderna_available    <dbl> 2400, 300, 0, 0, 0, 0, 0, 500, 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, 3~
## $ vaccines_available   <dbl> 2400, 300, 0, 0, 0, 0, 0, 500, 0, 0, 0, 0, 0, 0, ~
## $ total_shipped        <dbl> 22000, 3300, 100, 200, 100, 100, 100, 100, 100, 1~
## $ public_phone         <chr> "830-990-6648", "830-769-2702", NA, NA, "817-618-~
## $ website              <chr> "https://www.hillcountrymemorial.org/covid-19-upd~
```

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` 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> 106151, 146063, 35092
```

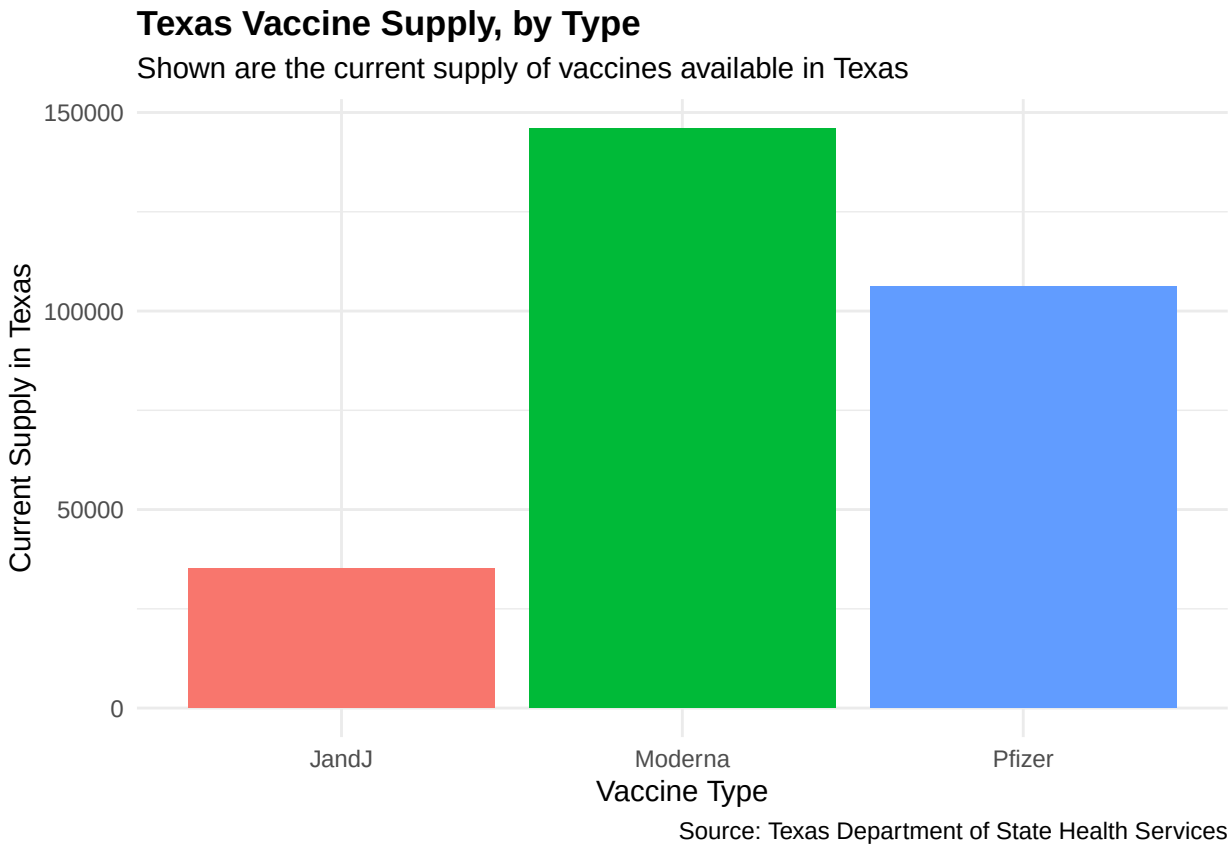
Visualize our Vaccine Data

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 subti
                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 =
```

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**
	(795.067)
typeHospital	2570.499**
	(1061.919)
typeLocal Health Department	1439.045
	(1825.881)
typeMedical Practice	-902.940
	(998.680)
typeOther	-679.726
	(934.630)
typePharmacy	-617.298
	(896.538)
typeVaccine Hub	65322.357***
	(1659.165)
Num.Obs.	3382
R2	0.371
R2 Adj.	0.370
AIC	74202.4
BIC	74251.4
Log.Lik.	-37093.202
F	331.333

* 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 = "Statistic", # Add a title for the X Axis
       y = "Provider Types") # Add a title for the Y Axis
```

Regression Equation

```
# install.packages("equatiomatic")

equatiomatic::extract_eq(model_1) # Extract LaTeX equation with equatiomatic package
```

$$\text{total\_shipped} = \alpha + \beta_1(\text{type}_{\text{Hospital}}) + \beta_2(\text{type}_{\text{Local Health Department}}) + \beta_3(\text{type}_{\text{Medical Practice}}) + \beta_4(\text{type}_{\text{Other}}) + \beta_5(\text{type}_{\text{Pharmacy}})$$