

Final Project Step 2

Lincoln Brown

8/8/2021

Quick Navigation:

How to import and clean my data

Cleaning the data

What does the final data set look like?

Final Data Set

What information is not self-evident?

Information

What are different ways you could look at this data?

Different Views

How do you plan to slice and dice the data?

- State
- Death
- Age
- Vaccine Type
- Vaccine Manufacturer

How could you summarize your data to answer key questions?

Summarized Data

What types of plots and tables will help you to illustrate the findings to your questions?

Plots and Tables

Do you plan on incorporating any machine learning techniques to answer your research questions? Explain.

Machine Learning

Questions for future steps.

Questions

Libraries

```
library(readxl)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.2      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(stringr)
library(dplyr)
library(broom)
library(scales)

##
## Attaching package: 'scales'
## The following object is masked from 'package:purrr':
##
##   discard
## The following object is masked from 'package:readr':
##
##   col_factor

library(coefplot)
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

library(QuantPsyc)

## Loading required package: boot
## Loading required package: MASS

##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##   select
##
## Attaching package: 'QuantPsyc'
## The following object is masked from 'package:base':
##
##   norm
```

```
library(knitr)
library(dplyr)
library(ggplot2)
```

Datasets

```
data19 <- read.csv("2019VAERSDATA.csv")
symp19 <- read.csv("2019VAERSSYMPTOMS.csv")
vax19 <- read.csv("2019VAERSVAX.csv")
data20 <- read.csv("2020VAERSDATA.csv")
symp20 <- read.csv("2020VAERSSYMPTOMS.csv")
vax20 <- read.csv("2020VAERSVAX.csv")
data21 <- read.csv("2021VAERSDATA.csv")
symp21 <- read.csv("2021VAERSSYMPTOMS.csv")
vax21 <- read.csv("2021VAERSVAX.csv")
```

Merge the datasets by year

We will merge each of the three years separate datasets into one and then find the number of unique VAERS_IDs to determine the number of people affected by adverse vaccine events.

```
merged19 <- merge(data19, symp19)
dim(merged19)
```

```
## [1] 60214    45
```

```
colnames(merged19)
```

```
## [1] "VAERS_ID"      "RECVDATE"      "STATE"         "AGE_YRS"
## [5] "CAGE_YR"       "CAGE_MO"       "SEX"           "RPT_DATE"
## [9] "SYMPTOM_TEXT"  "DIED"          "DATEDIED"      "L_THREAT"
## [13] "ER_VISIT"      "HOSPITAL"      "HOSPDAYS"      "X_STAY"
## [17] "DISABLE"       "RECOVD"        "VAX_DATE"      "ONSET_DATE"
## [21] "NUMDAYS"       "LAB_DATA"      "V_ADMINBY"     "V_FUNDBY"
## [25] "OTHER_MEDS"    "CUR_ILL"       "HISTORY"       "PRIOR_VAX"
## [29] "SPLTTYPE"      "FORM_VERS"     "TODAYS_DATE"   "BIRTH_DEFECT"
## [33] "OFC_VISIT"     "ER_ED_VISIT"   "ALLERGIES"     "SYMPTOM1"
## [37] "SYMPTOMVERSION1" "SYMPTOM2"      "SYMPTOMVERSION2" "SYMPTOM3"
## [41] "SYMPTOMVERSION3" "SYMPTOM4"      "SYMPTOMVERSION4" "SYMPTOM5"
## [45] "SYMPTOMVERSION5"
```

```
merged19 <- merge(merged19, vax19)
ids19 <- unique(merged19$VAERS_ID)
```

```
no_dup19 <- merged19 |> dplyr::distinct(VAERS_ID, .keep_all = TRUE)
```

```
merged20 <- merge(data20, symp20)
dim(merged20)
```

```
## [1] 60082    45
```

```
colnames(merged20)
```

```
## [1] "VAERS_ID"      "RECVDATE"      "STATE"         "AGE_YRS"
## [5] "CAGE_YR"       "CAGE_MO"       "SEX"           "RPT_DATE"
## [9] "SYMPTOM_TEXT"  "DIED"          "DATEDIED"      "L_THREAT"
```

```
## [13] "ER_VISIT"      "HOSPITAL"      "HOSPDAYS"      "X_STAY"
## [17] "DISABLE"       "RECOVD"        "VAX_DATE"      "ONSET_DATE"
## [21] "NUMDAYS"       "LAB_DATA"      "V_ADMINBY"     "V_FUNDBY"
## [25] "OTHER_MEDS"    "CUR_ILL"       "HISTORY"       "PRIOR_VAX"
## [29] "SPLTTYPE"     "FORM_VERS"     "TODAYS_DATE"   "BIRTH_DEFECT"
## [33] "OFC_VISIT"     "ER_ED_VISIT"   "ALLERGIES"     "SYMPTOM1"
## [37] "SYMPTOMVERSION1" "SYMPTOM2"     "SYMPTOMVERSION2" "SYMPTOM3"
## [41] "SYMPTOMVERSION3" "SYMPTOM4"     "SYMPTOMVERSION4" "SYMPTOM5"
## [45] "SYMPTOMVERSION5"
```

```
merged20 <- merge(merged20, vax20)
ids20 <- unique(merged20$VAERS_ID)
```

```
no_dup20 <- merged20 |> dplyr::distinct(VAERS_ID, .keep_all = TRUE)
```

```
merged21 <- merge(data21, symp21)
dim(merged21)
```

```
## [1] 542931      45
```

```
colnames(merged21)
```

```
## [1] "VAERS_ID"      "RECVDATE"      "STATE"         "AGE_YRS"
## [5] "CAGE_YR"       "CAGE_MO"       "SEX"           "RPT_DATE"
## [9] "SYMPTOM_TEXT"  "DIED"          "DATEDIED"      "L_THREAT"
## [13] "ER_VISIT"     "HOSPITAL"      "HOSPDAYS"      "X_STAY"
## [17] "DISABLE"      "RECOVD"        "VAX_DATE"      "ONSET_DATE"
## [21] "NUMDAYS"      "LAB_DATA"      "V_ADMINBY"     "V_FUNDBY"
## [25] "OTHER_MEDS"   "CUR_ILL"       "HISTORY"       "PRIOR_VAX"
## [29] "SPLTTYPE"     "FORM_VERS"     "TODAYS_DATE"   "BIRTH_DEFECT"
## [33] "OFC_VISIT"    "ER_ED_VISIT"   "ALLERGIES"     "SYMPTOM1"
## [37] "SYMPTOMVERSION1" "SYMPTOM2"     "SYMPTOMVERSION2" "SYMPTOM3"
## [41] "SYMPTOMVERSION3" "SYMPTOM4"     "SYMPTOMVERSION4" "SYMPTOM5"
## [45] "SYMPTOMVERSION5"
```

```
merged21 <- merge(merged21, vax21)
ids21 <- unique(merged21$VAERS_ID)
```

```
no_dup21 <- merged21 |> dplyr::distinct(VAERS_ID, .keep_all = TRUE)
```

Now that the three datasets are merged, we will look at the total amount of unique VAERS_IDs to get a total number of patients that experienced adverse reactions to vaccines in the years 2019, 2020, and 2021.

```
total19 <- length(ids19)
total20 <- length(ids20)
total21 <- length(ids21)
```

```
cat("Totals for 2019:", total19)
```

```
## Totals for 2019: 48443
```

```
cat("\n")
```

```
cat("Totals for 2020:", total20)
```

```
## Totals for 2020: 48901
```

```
cat("\n")
```

```
cat("Totals for 2021:", total21)
```

```
## Totals for 2021: 405998
```

Cleaning the data

[Return to Top](#)

Now we will filter the years 2020 and 2021 for COVID vaccines. Analysis conducted in COVID19 vaccines will not include the year 2019 because there is no data to support it.

```
filter20 <- filter(merged20, grepl("COVID19", merged20$VAX_TYPE))
filter21 <- filter(merged21, grepl("COVID19", merged21$VAX_TYPE))
```

```
total_covid_20 <- length(unique(filter20$VAERS_ID))
total_covid_21 <- length(unique(filter21$VAERS_ID))
```

```
cat("Total COVID Vaccine Adverse Events for 2020:", total_covid_20, "\n")
```

```
## Total COVID Vaccine Adverse Events for 2020: 10859
```

```
cat("Total COVID Vaccine Adverse Events for 2021:", total_covid_21, "\n")
```

```
## Total COVID Vaccine Adverse Events for 2021: 402528
```

At this stage, the data sets are loaded, but are highly convoluted with information that will not be necessary for final analysis. It is necessary at this point to remove unwanted columns so that the final data set is concise and easier to work with.

```
keep_cols <- c(1,3,4,7,9,10,12,21,23,28,35,36,38,40,42,44,46,47,48,49,52)
```

```
refined19 <- merged19[,keep_cols]
refined20 <- filter20[,keep_cols]
refined21 <- filter21[,keep_cols]
```

Final Data Sets

[Return to Top](#)

This is what the datasets will look like before I begin slicing and dicing to subset the data by variables.

```
print(str(refined20))
```

```
## 'data.frame':   14080 obs. of  21 variables:
##  $ VAERS_ID      : int   896636 896636 902418 902440 902446 902464 902465 902465 902468 902468 ...
##  $ STATE         : chr    "SC" "SC" "NJ" "AZ" ...
##  $ AGE_YRS       : num   47 47 56 35 55 42 60 60 59 59 ...
##  $ SEX           : chr    "F" "F" "F" "F" ...
##  $ SYMPTOM_TEXT   : chr    "ARM SWELLING WITH PAIN, FEVER, FATIGUE, THEN CONFUSION, FORGETFULNESS, BR
##  $ DIED          : chr    "" "" "" "" ...
##  $ L_THREAT      : chr    "" "" "" "" ...
##  $ NUMDAYS       : int     4 4 0 0 0 0 0 0 0 0 ...
##  $ V_ADMINBY     : chr    "OTH" "OTH" "PVT" "PVT" ...
##  $ PRIOR_VAX     : chr    "" "" "" "" ...
##  $ ALLERGIES      : chr    "tetracycline" "tetracycline" "none" "" ...
##  $ SYMPTOM1      : chr    "Arthralgia" "Memory impairment" "Hypoaesthesia" "Headache" ...
##  $ SYMPTOM2      : chr    "Confusional state" "Pain in extremity" "Injection site hypoaesthesia" "" .
##  $ SYMPTOM3      : chr    "Fatigue" "Peripheral swelling" "" "" ...
```

```
## $ SYMPTOM4      : chr "Feeling abnormal" "Physiotherapy" "" "" ...
## $ SYMPTOM5      : chr "Head discomfort" "Pyrexia" "" "" ...
## $ VAX_TYPE      : chr "COVID19" "COVID19" "COVID19" "COVID19" ...
## $ VAX_MANU      : chr "MODERNA" "MODERNA" "PFIZER\\BIONTECH" "PFIZER\\BIONTECH" ...
## $ VAX_LOT       : chr "" "" "EH9899" "EH 9899" ...
## $ VAX_DOSE_SERIES: chr "UNK" "UNK" "1" "1" ...
## $ VAX_NAME      : chr "COVID19 (COVID19 (MODERNA))" "COVID19 (COVID19 (MODERNA))" "COVID19 (COVID19 (MODERNA))" ...
## NULL
print(str(refined21))
```

```
## 'data.frame':    557702 obs. of  21 variables:
## $ VAERS_ID      : int  916600 916601 916602 916603 916604 916606 916607 916608 916609 916610 ...
## $ STATE         : chr   "TX" "CA" "WA" "WA" ...
## $ AGE_YRS       : num   33 73 23 58 47 44 50 33 71 18 ...
## $ SEX           : chr   "F" "F" "F" "F" ...
## $ SYMPTOM_TEXT  : chr   "Right side of epiglottis swelled up and hinder swallowing pictures taken B
## $ DIED          : chr   "" "" "" "" ...
## $ L_THREAT      : chr   "" "" "" "" ...
## $ NUMDAYS       : int    2 0 0 0 7 0 1 2 8 1 ...
## $ V_ADMINBY     : chr   "PVT" "SEN" "SEN" "WRK" ...
## $ PRIOR_VAX     : chr   "" "" "" "got measles from measel shot, mums from mumps shot, headaches and
## $ ALLERGIES     : chr   "Pcn and bee venom" "\"Dairy\"" "Shellfish" "Diclofenac, novacaine, lidocain
## $ SYMPTOM1      : chr   "Dysphagia" "Anxiety" "Chest discomfort" "Dizziness" ...
## $ SYMPTOM2      : chr   "Epiglottitis" "Dyspnoea" "Dysphagia" "Fatigue" ...
## $ SYMPTOM3      : chr   "" "" "Pain in extremity" "Mobility decreased" ...
## $ SYMPTOM4      : chr   "" "" "Visual impairment" "" ...
## $ SYMPTOM5      : chr   "" "" "" "" ...
## $ VAX_TYPE      : chr   "COVID19" "COVID19" "COVID19" "COVID19" ...
## $ VAX_MANU      : chr   "MODERNA" "MODERNA" "PFIZER\\BIONTECH" "MODERNA" ...
## $ VAX_LOT       : chr   "037K20A" "025L20A" "EL1284" "unknown" ...
## $ VAX_DOSE_SERIES: chr   "1" "1" "1" "UNK" ...
## $ VAX_NAME      : chr   "COVID19 (COVID19 (MODERNA))" "COVID19 (COVID19 (MODERNA))" "COVID19 (COVID19 (MODERNA))" ...
## NULL
```

State Datasets

Return to Top

I will be breaking the data down state by state, so I want to verify that the States are clean. I will be removing instances of blank states when creating the states20 and states21 variables.

```
uniq_state20 <- unique(toupper(filter20$STATE))
uniq_state20[7]
```

```
## [1] ""
```

```
states20 <- refined20[refined20$STATE!=uniq_state20[7],]
dim(states20)
```

```
## [1] 12947    21
```

```
uniq_state21 <- unique(toupper(filter21$STATE))
uniq_state21[14]
```

```
## [1] ""
```

```
states21 <- refined21[refined21$STATE!=uniq_state21[14],]
dim(states21)
```

```
## [1] 504834    21
```

Death datasets

Return to Top

```
died20 <- dplyr::filter(refined20, grepl("Y", refined20$DIED))
died21 <- dplyr::filter(refined21, grepl("Y", refined21$DIED))
```

```
died20_nodup <- died20 |> dplyr::distinct(VAERS_ID, .keep_all = TRUE)
dim(died20_nodup)
```

```
## [1] 16 21
```

```
died21_nodup <- died21 |> dplyr::distinct(VAERS_ID, .keep_all = TRUE)
dim(died21_nodup)
```

```
## [1] 5231    21
```

```
cat("Total deaths related to COVID vaccinations in 2020:", length(unique(died20$VAERS_ID)), "\n")
```

```
## Total deaths related to COVID vaccinations in 2020: 16
```

```
cat("Total deaths related to COVID vaccinations in 2021:", length(unique(died21$VAERS_ID)), "\n")
```

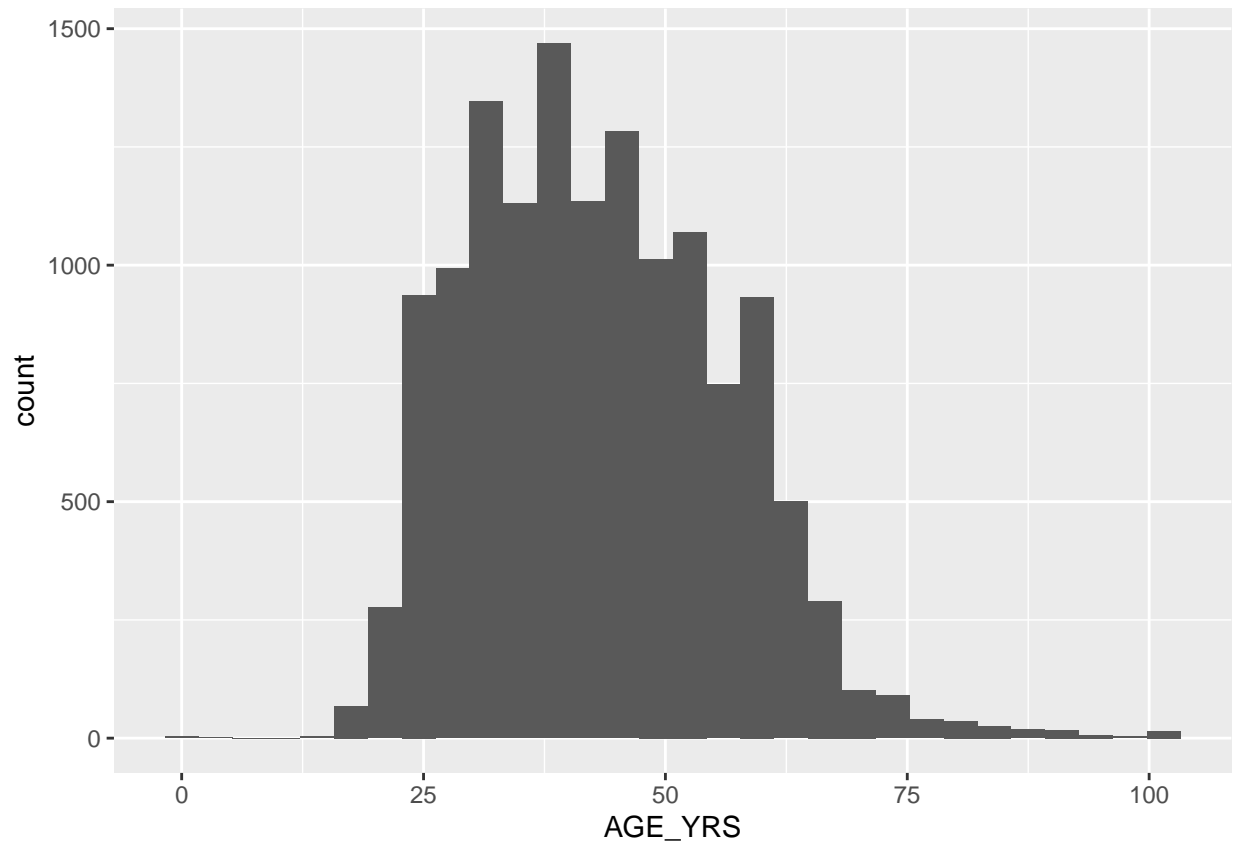
```
## Total deaths related to COVID vaccinations in 2021: 5231
```

Age

Return to Top

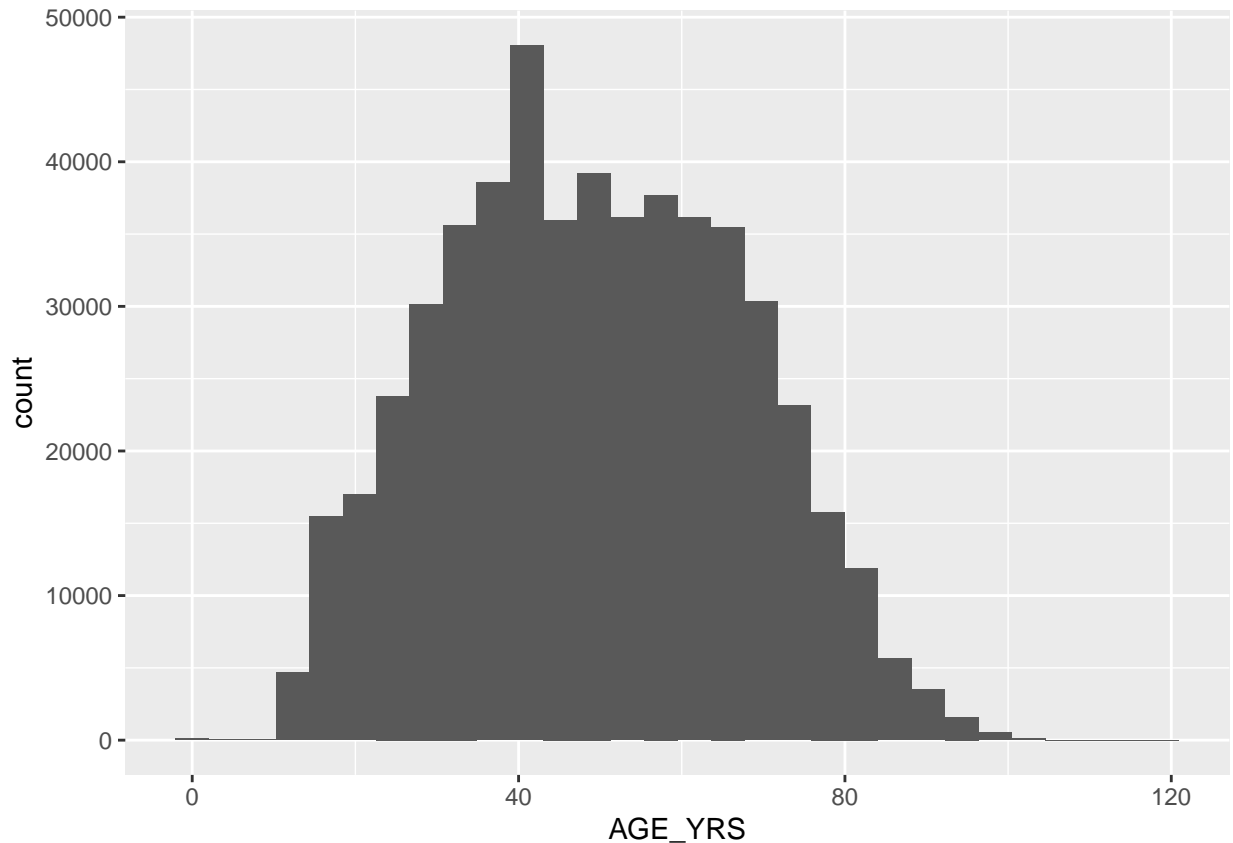
```
age20 <- refined20[!is.na(refined20$AGE_YRS),]
age20_hist <- ggplot(refined20, aes(AGE_YRS)) + geom_histogram(bins=30)
age20_hist
```

```
## Warning: Removed 527 rows containing non-finite values (stat_bin).
```



```
age21 <- refined20[!is.na(refined21$AGE_YRS),]  
age21_hist <- ggplot(refined21, aes(AGE_YRS)) + geom_histogram(bins=30)  
age21_hist
```

```
## Warning: Removed 30918 rows containing non-finite values (stat_bin).
```

```
## Vaccine Type {#vax_type} Return to Top
```

```
vax_type19 <- unique(merged19$VAX_TYPE)
vax_type20 <- unique(merged20$VAX_TYPE)
vax_type21 <- unique(merged21$VAX_TYPE)
```

```
print("Vaccine types 2019: ")
```

```
## [1] "Vaccine types 2019: "
```

```
print(vax_type19)
```

```
## [1] "FLU4"      "VARZOS"    "DTAPIPV"   "MMRV"      "HEPA"
## [6] "FLU3"      "PNC13"     "FLUC4"     "FLUX"      "MMR"
## [11] "VARCEL"    "PPV"       "JEV1"      "SMALL"     "HEP"
## [16] "CHOL"      "HPV9"      "TDAP"      "MNQ"       "DTAPHEPBIP"
## [21] "HIBV"      "RV5"       "TYP"       "DTAPIPVHIB" "HPVX"
## [26] "YF"        "FLUA3"     "DTAP"      "IPV"       "PNC"
## [31] "DTPIPV"    "TD"        "RVX"       "HEPAB"     "RV1"
## [36] "RAB"       "FLUR4"     "HPV4"      "MEN"       "FLUR3"
## [41] "UNK"       "ANTH"      "MENB"      "FLUN3"     "FLUN4"
## [46] "ADEN_4_7"  "DT"        "TTOX"      "DTP"       "HBHEPB"
## [51] "JEV"       "DTOX"      "6VAX-F"    "FLUC3"     "FLUX(H1N1)"
## [56] "TBE"       "BCG"       "HPV2"      "PER"       "MEA"
## [61] "MNQHIB"    "ADEN"     "OPV"       "RV"
```

```
print("Vaccine types 2020: ")
```

```
## [1] "Vaccine types 2020: "
```

```
print(vax_type20)
```

```
## [1] "VARZOS"      "UNK"         "FLU3"        "FLUR4"       "FLU4"
## [6] "PPV"         "PNC13"       "DTAPIPV"     "HEP"         "FLUC4"
## [11] "HEPA"        "HPV9"       "FLUX"        "MNQ"         "DTAP"
## [16] "MMR"         "MMRV"       "MENB"        "VARCEL"      "RV1"
## [21] "RV5"         "TDAP"       "FLUN4"       "FLUA3"       "MEN"
## [26] "IPV"         "DTAPHEPBIP" "HEPAB"       "DTAPIPVHIB" "HPV4"
## [31] "HIBV"        "TYP"        "DTP"         "RAB"         "TD"
## [36] "PNC10"       "FLUN3"      "ANTH"        "YF"          "JEV1"
## [41] "ADEN_4_7"    "SMALL"      "HPVX"        "PNC"         "DT"
## [46] "TTOX"        "CHOL"       "FLUR3"       "RVX"         "HPV2"
## [51] "BCG"         "JEVX"       "DTOX"        "HBHEPB"      "FLUA4"
## [56] "RUB"         "FLUC3"      "MENHIB"      "COVID19"     "DF"
## [61] "FLUX(H1N1)" "EBZR"
```

```
print("Vaccine types 2021: ")
```

```
## [1] "Vaccine types 2021: "
```

```
print(vax_type21)
```

```
## [1] "COVID19"     "FLUC4"       "UNK"         "FLU4"       "FLUA3"
## [6] "VARZOS"      "DTAPHEPBIP" "RV1"         "HIBV"       "MMRV"
## [11] "DTAP"        "FLUR4"       "PNC13"       "TDAP"       "FLUX"
## [16] "HEP"         "MMR"         "PPV"         "FLUA4"      "FLUN4"
## [21] "MENB"        "HEPA"       "DTAPIPV"     "HPV9"       "DT"
## [26] "TTOX"        "CHOL"       "FLU3"        "FLUC3"      "VARCEL"
## [31] "HPV4"        "MNQ"        "RV5"         "ANTH"       "TYP"
## [36] "IPV"         "EBZR"       "TD"          "HPVX"       "FLUX(H1N1)"
## [41] "RVX"         "DTAPIPVHIB" "HEPAB"       "DTP"        "ADEN_4_7"
## [46] "RAB"         "DF"         "PNC"         "OPV"        "MEN"
## [51] "YF"         "JEV1"       "MENHIB"      "FLU(H1N1)" "SMALL"
## [56] "FLUN3"      "MNQHIB"
```

Vaccine Manufacturers

Return to Top

```
vax_manu19 <- unique(merged19$VAX_MANU)
vax_manu20 <- unique(merged20$VAX_MANU)
vax_manu21 <- unique(merged21$VAX_MANU)
```

```
print("Vaccine Manufacturers 2019: ")
```

```
## [1] "Vaccine Manufacturers 2019: "
```

```
print(vax_manu19)
```

```
## [1] "GLAXOSMITHKLINE BIOLOGICALS"      "MERCK & CO. INC."
## [3] "SANOFI PASTEUR"                   "PFIZER\\WYETH"
## [5] "SEQIRUS, INC."                    "UNKNOWN MANUFACTURER"
## [7] "INTERCELL AG"                     "PAXVAX"
## [9] "NOVARTIS VACCINES AND DIAGNOSTICS" "BERNA BIOTECH, LTD."
## [11] "PROTEIN SCIENCES CORPORATION"     "EMERGENT BIOSOLUTIONS"
## [13] "CSL LIMITED"                       "MEDIMUNE VACCINES, INC."
## [15] "TEVA PHARMACEUTICALS"             "MASS. PUB HLTH BIOL LAB"
```

```
## [17] "DYNAVAX TECHNOLOGIES CORPORATION" "SMITHKLINE BEECHAM"
## [19] "MICHIGAN DEPT PUB HLTH" "CONNAUGHT LABORATORIES"
## [21] "AVENTIS PASTEUR"

cat("\n")

print("Vaccine Manufacturers 2020: ")

## [1] "Vaccine Manufacturers 2020: "
print(vax_manu20)

## [1] "GLAXOSMITHKLINE BIOLOGICALS" "UNKNOWN MANUFACTURER"
## [3] "SANOFI PASTEUR" "PROTEIN SCIENCES CORPORATION"
## [5] "MERCK & CO. INC." "PFIZER\\WYETH"
## [7] "SEQIRUS, INC." "NOVARTIS VACCINES AND DIAGNOSTICS"
## [9] "DYNAVAX TECHNOLOGIES CORPORATION" "MEDIMMUNE VACCINES, INC."
## [11] "CSL LIMITED" "BERNA BIOTECH, LTD."
## [13] "MASS. PUB HLTH BIOL LAB" "EMERGENT BIOSOLUTIONS"
## [15] "INTERCELL AG" "TEVA PHARMACEUTICALS"
## [17] "MEDEVA PHARMA, LTD." "PAXVAX"
## [19] "MODERNA" "PFIZER\\BIONTECH"

cat("\n")

print("Vaccine Manufacturers 2021: ")

## [1] "Vaccine Manufacturers 2021: "
print(vax_manu21)

## [1] "MODERNA" "PFIZER\\BIONTECH"
## [3] "SEQIRUS, INC." "UNKNOWN MANUFACTURER"
## [5] "SANOFI PASTEUR" "NOVARTIS VACCINES AND DIAGNOSTICS"
## [7] "GLAXOSMITHKLINE BIOLOGICALS" "MERCK & CO. INC."
## [9] "PROTEIN SCIENCES CORPORATION" "PFIZER\\WYETH"
## [11] "MEDIMMUNE VACCINES, INC." "PAXVAX"
## [13] "EMERGENT BIOSOLUTIONS" "BERNA BIOTECH, LTD."
## [15] "CSL LIMITED" "TEVA PHARMACEUTICALS"
## [17] "DYNAVAX TECHNOLOGIES CORPORATION" "JANSSEN"
## [19] "MASS. PUB HLTH BIOL LAB" "INTERCELL AG"
```

Information about the data

[Return to Top](#)

It is important to remember that the number of rows in the data set are not necessarily the number of people affected. Looking at the number of unique VAERS_IDs is the correct way to see the number of people affected by adverse events from vaccinations.

Different ways to view the data

[Return to Top](#)

- Age
- Sex
- State

- Vaccine Manufacturer
- Died

Summarize data to answer key questions

Return to Top

A lot of the key questions related to this data set can be answered with simple functions or plots.

Here are a few questions that are pertinent to the analysis.

What is the average age per year that a person experiences an event?

In 2019 the average age was: 40.0555507.

In 2020 the average age was: 41.1683647.

In 2021 the average age was: 49.0881903.

Which vaccine type was associated with the highest number of adverse reactions?

```
num_type19 <- table(no_dup19$VAX_TYPE)
cat("Table of adverse events by vaccine type for 2019:\n")
```

```
## Table of adverse events by vaccine type for 2019:
```

```
print(num_type19)
```

```
##
##      6VAX-F      ADEN  ADEN_4_7      ANTH      BCG      CHOL      DT
##      11          1          8      145          2          19      26
##      DTAP DTAPHEPBIP  DTAPIPV DTAPIPVHIB  DTOX      DTP      FLU3
##      1039        512        847      750          1          14      2366
##      FLU4      FLUA3      FLUC3      FLUC4      FLUN3      FLUN4      FLUR3
##      4422        595          1      1326        10          44          8
##      FLUR4      FLUX FLUX(H1N1)      HEP      HEP      HEPAB      HIBV
##      520        583          1      837      1550        117          507
##      HPV2      HPV4      HPV9      HPVX      IPV      JEV      JEV1
##      6          241      1815        49      197          1          32
##      MEA      MEN      MENB      MMR      MMRV      MNQ      PER
##      2          33          913      1685      1050        662          1
##      PNC      PNC13      PPV      RAB      RV1      RV5      RVX
##      11        1170      2452      177          52          568          14
##      SMALL      TD      TDAP      TTOX      TYP      UNK      VARCEL
##      79         137      1281        35      131      1111      1097
##      VARZOS      YF
##      17134        45
```

```
num_type20 <- table(no_dup20$VAX_TYPE)
cat("\nTable of adverse events by vaccine type for 2020:\n")
```

```
##
## Table of adverse events by vaccine type for 2020:
```

```
print(num_type20)
```

```
##
##      ADEN_4_7      ANTH      BCG      CHOL      COVID19      DF      DT
##      3          90          5      18      10859          1      14
##      DTAP DTAPHEPBIP  DTAPIPV DTAPIPVHIB      DTP      EBZR      FLU3
```

```
##      587      361      593      578      7      1      415
##      FLU4      FLUA3      FLUA4      FLUC3      FLUC4      FLUN3      FLUN4
##      6276      378      688      8      1083      20      50
##      FLUR3      FLUR4      FLUX FLUX(H1N1)      HBHEPB      HEP      HEPA
##      7      813      2306      1      1      607      813
##      HEPAB      HIBV      HPV2      HPV4      HPV9      HPVX      IPV
##      60      497      2      180      1229      8      131
##      JEV1      MEN      MENB      MMR      MMRV      MNQ      PNC
##      19      49      603      667      704      390      19
##      PNC10      PNC13      PPV      RAB      RV1      RV5      RVX
##      1      610      1868      120      43      560      1
##      SMALL      TD      TDAP      TTOX      TYP      UNK      VARCEL
##      17      104      730      19      750      1008      895
##      VARZOS      YF
##      11020      14
```

```
num_type21 <- table(no_dup21$VAX_TYPE)
cat("\nTable of adverse events by vaccine type for 2021:\n")
```

```
##
## Table of adverse events by vaccine type for 2021:
print(num_type21)
```

```
##
##      ADEN_4_7      ANTH      CHOL      COVID19      DF      DT      DTAP
##      8      9      7      402313      3      3      33
##      DTAPHEPBIP      DTAPIPV      DTAPIPVHIB      DTP      EBZR      FLU(H1N1)      FLU3
##      21      25      27      1      2      2      13
##      FLU4      FLUA3      FLUA4      FLUC3      FLUC4      FLUN4      FLUR4
##      422      29      64      12      155      6      61
##      FLUX FLUX(H1N1)      HEP      HEPA      HEPAB      HIBV      HPV4
##      198      4      35      31      4      25      13
##      HPV9      HPVX      IPV      JEV1      MEN      MENB      MENHIB
##      65      3      9      1      1      29      1
##      MMR      MMRV      MNQ      OPV      PNC      PNC13      PPV
##      47      10      34      1      1      58      85
##      RAB      RV1      RV5      RVX      SMALL      TD      TDAP
##      9      4      8      1      1      2      55
##      TTOX      TYP      UNK      VARCEL      VARZOS      YF
##      4      7      1172      26      837      1
```

Tables for Vaccine Manufacturers associated with COVID-19 vaccine adverse events that resulted in death

```
death_type20 <- table(died20_nodup$VAX_MANU)
cat("Number of adverse events that resulted in death associated with each COVID19-vaccine manufacturer :
## Number of adverse events that resulted in death associated with each COVID19-vaccine manufacturer in
print(death_type20)
```

```
##
##      MODERNA PFIZER\BIONTECH
##      5      11
```

```

death_type21 <- table(died21_nodup$VAX_MANU)
cat("Number of adverse events that resulted in death associated with each COVID19-vaccine manufacturer :

## Number of adverse events that resulted in death associated with each COVID19-vaccine manufacturer in
print(death_type21)

##
##                JANSSEN                MODERNA                PFIZER\BIONTECH
##                490                2485                2232
## UNKNOWN MANUFACTURER
##                24

```

Adverse reactions by sex:

```

cat("Adverse reactions by sex for 2019: \n")

```

```

## Adverse reactions by sex for 2019:

```

```

print(table(no_dup19$SEX))

```

```

##
##      F      M      U
## 28070 13730  6643

```

```

cat("\nAdverse reactions by sex for 2020:\n")

```

```

##
## Adverse reactions by sex for 2020:

```

```

print(table(no_dup20$SEX))

```

```

##
##      F      M      U
## 29331 12595  6975

```

```

cat("\nAdverse reactions by sex for 2021:\n")

```

```

##
## Adverse reactions by sex for 2021:

```

```

print(table(no_dup21$SEX))

```

```

##
##      F      M      U
## 283744 112011 10243

```

Plots and Tables

Return to Top

- Histograms - Look at the distribution of data for specific variables.
- Scatterplots - Identify relationships between variables
- Residual plots - Look for outliers in the distributions
- Density plots - Observe smoothed distributions to check assumptions.
- Box plots - Look for outliers in the distributions

- Tables:
 - Vaccine types
 - Variables used
 - Common Symptoms
 - Reactions by state
 - Prior Allergies
 - Administered by variable explanation

Machine Learning

[Return to Top](#)

I do not plan to use any machine learning techniques at this time.

Questions

[Return to Top](#)

I want to do research into the symptoms, but I am unsure of how far I will get due to my limitations in text data mining. Improve use of categorical variables in regression. Best way to handle missing data.