Predicting the occurrence of seizures in a population of COVID-19 patients during prolonged hospital stays with Azithromycin and Hydroxychloroquine treatment options.

Savannah Gilkes

Data Science Practicum I

Regis University

Overview

- Purpose and Description
- R and RStudio
- Data Description and Preparation
- Exploratory Data Analysis (EDA) and Data Visualizations
- KNN Classification Model
- Logistic Regression Model
- Conclusions

Purpose

Determine the factors that can be used to predict the occurrence of seizures

- R will be used to perform descriptive statistics, EDA, and produce visualizations
- R will be used to build a K-nearest-neighbor (KNN) classifier and logistic regression model to determine factors in the prediction of seizure occurrence in COVID-19 patients.

R and RStudio

Created by Ross Ihaka and Robert Gentlemen through the R core team in 1993

Statistical computing, data analysis, and graphics

RStudio is IDE created for R

Data Description

Published in 2021 provided by the Chief Resident of the LSU Health Department of Neurology

Originally contained 33 variables with 250 observations

Data Preparation

Age

Sex
0 = Male
1 = Female

Race
0 = American Indian/Native Alaskan
1 = Asian
2 = Native Hawaiian or Pacific Islander
3 = African American
4 = Caucasian

Ethnicity
0 = Non-Hispanic
1 = Hispanic

LOS (hospital length of stay)

BMI (body mass index)

Smoker
0 = Never
1 = Former
2 = Current

Asthma
0 = No
1 = Yes

Obesity
0 = No
1 = Yes

hx_migraine (history of migraine)
0 = No
1 = Yes

hx_epilepsy (history of epilepsy)
0 = No
1 = Yes

hx_cva (history of stroke)
0 = No
1 = Yes

AMS_comp (altered mental status during hospital stay)
0 = No
1 = Yes

Seizure_comp (seizures during hospital stay)
0 = No
1 = Yes

Headache_comp (headache during hospital stay)
0 = No
1 = Yes

Encephalitis_comp (encephalitis during hospital stay)
0 = No
1 = Yes

Ageusia_anosmia_comp (ageusia or anosmia during hospital stay)
0 = No
1 = Yes

Azithromycin
0 = No
1 = Yes

Hydroxychloroquine
0 = No
1 = Yes

Data Cleaning

```
> sum(is.na(df))
[1] 50
> colSums(is.na(df))
                  age                   sex                  race             ethnicity
                    0                     0                    33                     0
                  los                   bmi                smoker                asthma
                    0                     4                     6                     1
              obesity           hx_migraine          hx_epilepsy                hx_cva
                    2                     0                     0                     0
            AMS_comp          Seizure_comp          Status_comp         Headache_comp
                    2                     1                     1                     0
   Encephalitis_comp ageusia_anosmia_comp          Azithromycin   Hydroxychloroquine
                    0                     0                     0                     0
```

```
> sum(is.na(df1))
[1] 0
> colSums(is.na(df1))
                  age                   sex                  race             ethnicity
                    0                     0                     0                     0
                  los                   bmi                smoker                asthma
                    0                     0                     0                     0
              obesity           hx_migraine          hx_epilepsy                hx_cva
                    0                     0                     0                     0
            AMS_comp          Seizure_comp         Headache_comp    Encephalitis_comp
                    0                     0                     0                     0
ageusia_anosmia_comp          Azithromycin   Hydroxychloroquine
                    0                     0                     0
```

EDA

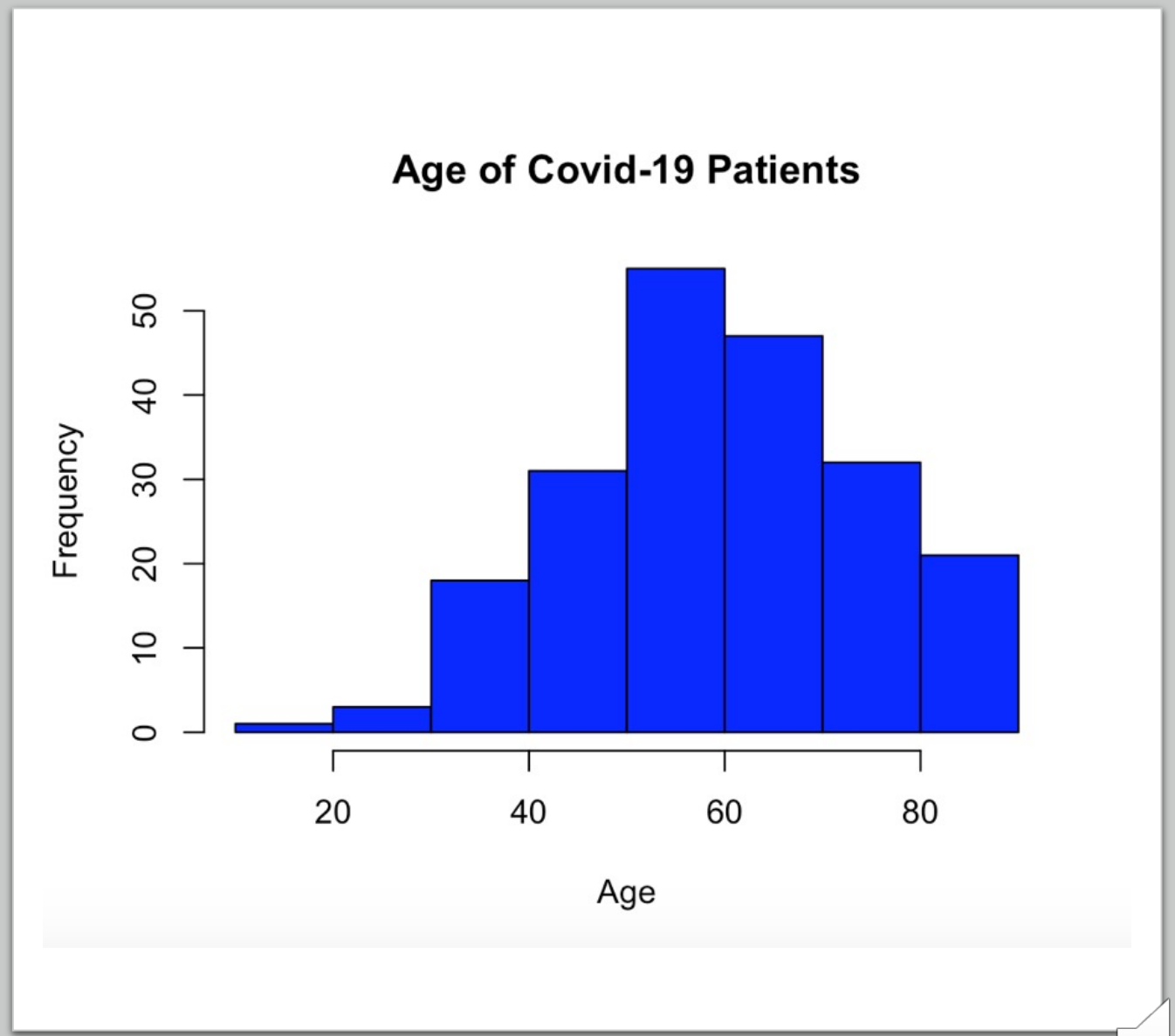
Summary

```
      age              sex              race          ethnicity            los        
 Min.   :18.00   Min.   :0.0000   Min.   :0.000   Min.   :0.000000   Min.   :-2.00  
 1st Qu.:50.00   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:0.000000   1st Qu.: 4.00  
 Median :59.50   Median :1.0000   Median :3.000   Median :0.000000   Median : 8.00  
 Mean   :59.97   Mean   :0.5529   Mean   :3.043   Mean   :0.009615   Mean   :10.41  
 3rd Qu.:71.00   3rd Qu.:1.0000   3rd Qu.:3.000   3rd Qu.:0.000000   3rd Qu.:14.00  
 Max.   :89.00   Max.   :1.0000   Max.   :4.000   Max.   :1.000000   Max.   :49.00  
      bmi             smoker           asthma          obesity        hx_migraine      
 Min.   :16.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000  
 1st Qu.:27.00   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000  
 Median :32.84   Median :0.0000   Median :0.0000   Median :1.0000   Median :0.00000  
 Mean   :34.71   Mean   :0.4615   Mean   :0.1635   Mean   :0.6298   Mean   :0.01442  
 3rd Qu.:40.97   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000  
 Max.   :69.71   Max.   :2.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000  
  hx_epilepsy         hx_cva          AMS_comp      Seizure_comp Headache_comp    
 Min.   :0.00000   Min.   :0.000   Min.   :0.0000   0:200        Min.   :0.00000  
 1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.0000   1:  8        1st Qu.:0.00000  
 Median :0.00000   Median :0.000   Median :0.0000                Median :0.00000  
 Mean   :0.03365   Mean   :0.149   Mean   :0.2644                Mean   :0.07692  
 3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:1.0000                3rd Qu.:0.00000  
 Max.   :1.00000   Max.   :1.000   Max.   :1.0000                Max.   :1.00000  
Encephalitis_comp ageusia_anosmia_comp  Azithromycin    Hydroxychloroquine
 Min.   :0.00000   Min.   :0.00000      Min.   :0.0000   Min.   :0.000     
 1st Qu.:0.00000   1st Qu.:0.00000      1st Qu.:0.0000   1st Qu.:0.000     
 Median :0.00000   Median :0.00000      Median :0.0000   Median :1.000     
 Mean   :0.01442   Mean   :0.01442      Mean   :0.4904   Mean   :0.625     
 3rd Qu.:0.00000   3rd Qu.:0.00000      3rd Qu.:1.0000   3rd Qu.:1.000     
 Max.   :1.00000   Max.   :1.00000      Max.   :1.0000   Max.   :1.000     
```

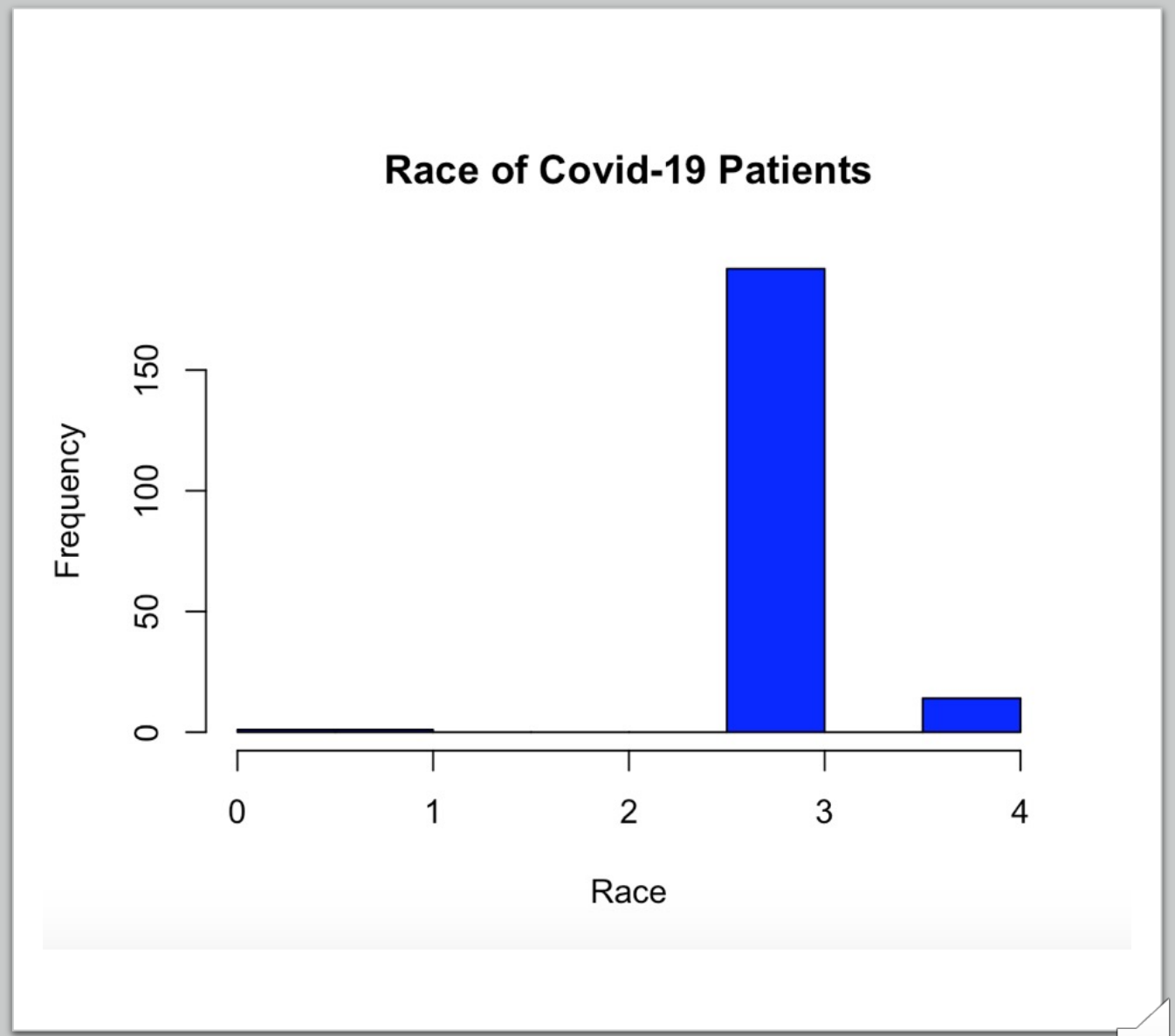
Structure

```
tibble [208 × 19] (S3: tbl_df/tbl/data.frame)
 $ age                 : num [1:208] 89 74 18 44 57 44 67 35 61 72 ...
 $ sex                 : num [1:208] 1 0 1 1 1 1 1 0 1 1 ...
 $ race                : num [1:208] 3 3 4 3 3 3 3 3 3 0 ...
 $ ethnicity           : num [1:208] 0 0 0 0 0 0 0 0 0 0 ...
 $ los                 : num [1:208] 3 25 9 1 14 18 12 6 26 6 ...
 $ bmi                 : num [1:208] 20.2 24.7 44.5 32.6 42.9 ...
 $ smoker              : num [1:208] 0 1 0 1 1 0 1 2 0 0 ...
 $ asthma              : num [1:208] 0 0 0 0 0 0 0 0 0 1 ...
 $ obesity             : num [1:208] 0 0 1 1 1 1 1 1 1 0 ...
 $ hx_migraine         : num [1:208] 0 0 0 0 0 0 0 0 0 0 ...
 $ hx_epilepsy         : num [1:208] 0 0 0 0 0 0 0 0 0 0 ...
 $ hx_cva              : num [1:208] 0 0 0 0 1 0 0 0 1 1 ...
 $ AMS_comp            : num [1:208] 1 1 0 0 0 0 0 0 1 1 ...
 $ Seizure_comp        : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 2 ...
 $ Headache_comp       : num [1:208] 0 0 0 0 0 0 0 0 0 0 ...
 $ Encephalitis_comp   : num [1:208] 0 0 0 0 0 0 0 0 0 0 ...
 $ ageusia_anosmia_comp: num [1:208] 0 0 0 0 0 0 0 1 0 1 ...
 $ Azithromycin        : num [1:208] 1 1 0 0 0 1 1 1 1 1 ...
 $ Hydroxychloroquine  : num [1:208] 0 0 1 1 0 1 1 0 1 1 ...
```

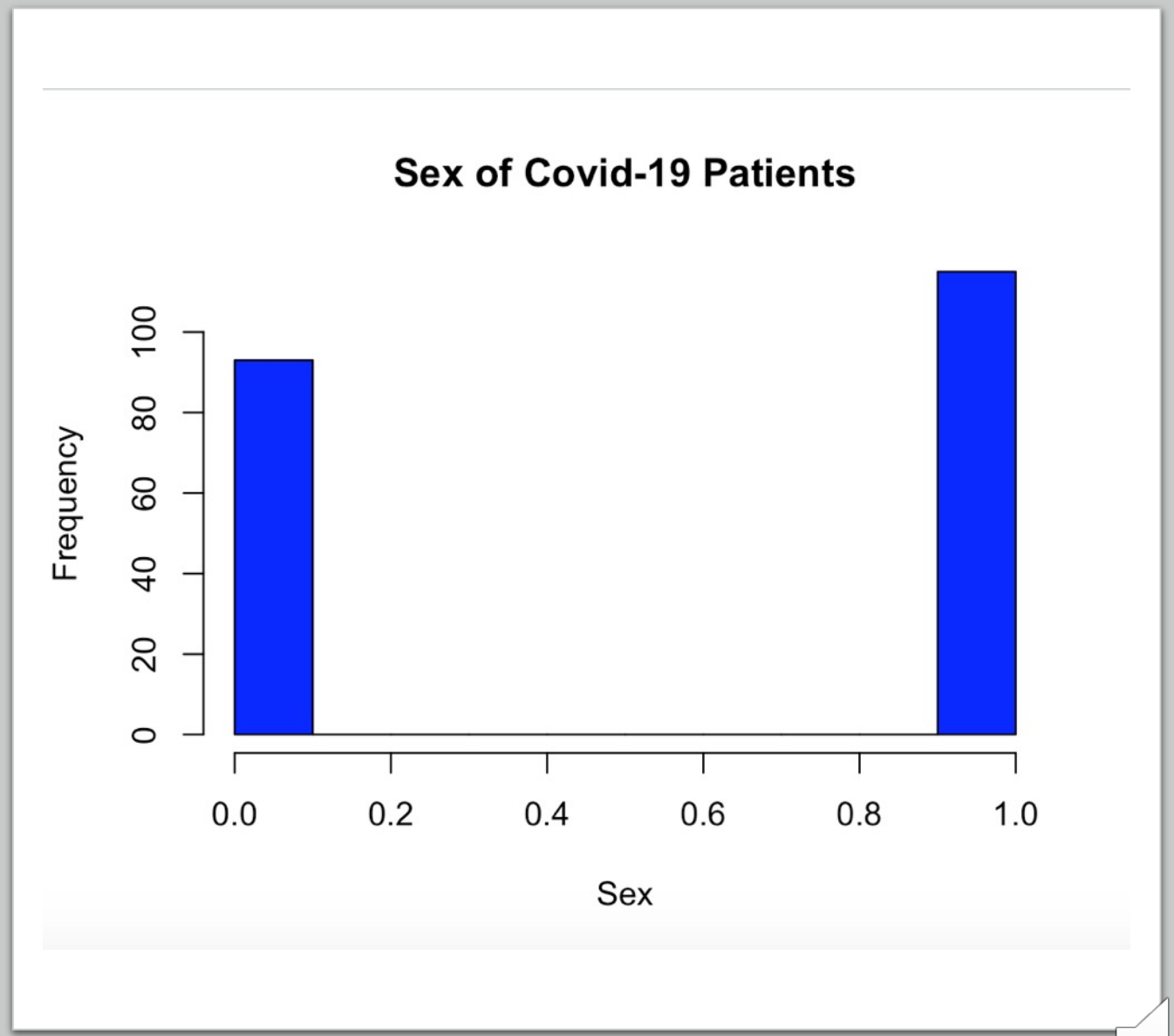
Data Visualizations



Data Visualizations



Data Visualizations



Building Models

KNN Classification Model

k = 1

```
Confusion Matrix and Statistics

   k1
     0  1
  0 60  0
  1  1  1

               Accuracy : 0.9839
                 95% CI : (0.9134, 0.9996)
    No Information Rate : 0.9839
    P-Value [Acc > NIR] : 0.7358

                  Kappa : 0.6593

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 0.9836
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 0.5000
             Prevalence : 0.9839
         Detection Rate : 0.9677
   Detection Prevalence : 0.9677
      Balanced Accuracy : 0.9918

       'Positive' Class : 0
```

k = 3

```
Confusion Matrix and Statistics

   k3
     0  1
  0 60  0
  1  2  0

               Accuracy : 0.9677
                 95% CI : (0.8883, 0.9961)
    No Information Rate : 1
    P-Value [Acc > NIR] : 1.0000

                  Kappa : 0

 Mcnemar's Test P-Value : 0.4795

            Sensitivity : 0.9677
            Specificity :     NA
         Pos Pred Value :     NA
         Neg Pred Value :     NA
             Prevalence : 1.0000
         Detection Rate : 0.9677
   Detection Prevalence : 0.9677
      Balanced Accuracy :     NA

       'Positive' Class : 0
```

Logistic Regression Model

```
Call:
glm(formula = Seizure_comp ~ ., family = binomial(), data = df2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7450  -0.0880  -0.0204  -0.0013   3.5648

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           9.8766     8.5390   1.157   0.2474
age                  -5.2521     6.0726  -0.865   0.3871
sex                   1.3213     1.7843   0.741   0.4590
race                -15.4881     7.7931  -1.987   0.0469 *
ethnicity             6.1146     2.4849   2.461   0.0139 *
los                   9.8299     4.3716   2.249   0.0245 *
bmi                 -21.0388    11.1237  -1.891   0.0586 .
smoker               -1.8724     2.4195  -0.774   0.4390
asthma               -1.2425     2.7011  -0.460   0.6455
obesity               4.0178     2.2476   1.788   0.0738 .
hx_migraine         -22.6440  8239.0503  -0.003   0.9978
hx_epilepsy           8.4558     4.4501   1.900   0.0574 .
hx_cva               -5.9217     3.9037  -1.517   0.1293
AMS_comp              1.8960     1.9531   0.971   0.3317
Headache_comp       -19.2321  3066.9069  -0.006   0.9950
Encephalitis_comp   -11.7325  9044.5369  -0.001   0.9990
ageusia_anosmia_comp  0.3458     3.4976   0.099   0.9212
Azithromycin          1.8995     1.5055   1.262   0.2070
Hydroxychloroquine   -2.2946     1.9176  -1.197   0.2314
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Logistic Regression Model

```r
fit_model <- train(Seizure_comp ~., data = train, method = 'glm', family = binomial())
pred <- predict(fit_model, test)
table(pred, test$Seizure_comp)
confusionMatrix(table(pred, test$Seizure_comp))
```

```
Confusion Matrix and Statistics


pred  0  1
   0 59  1
   1  1  1

               Accuracy : 0.9677
                 95% CI : (0.8883, 0.9961)
    No Information Rate : 0.9677
    P-Value [Acc > NIR] : 0.6767

                  Kappa : 0.4833

 Mcnemar's Test P-Value : 1.0000

            Sensitivity : 0.9833
            Specificity : 0.5000
         Pos Pred Value : 0.9833
         Neg Pred Value : 0.5000
             Prevalence : 0.9677
         Detection Rate : 0.9516
   Detection Prevalence : 0.9677
      Balanced Accuracy : 0.7417

       'Positive' Class : 0
```

Conclusions and Next Steps

- Race, ethnicity, and length of stay at the hospital may have an impact on seizure occurrence
- Can help healthcare workers assess what resources are needed and where they should be allocated
- More data needs to be collected
- Decision tree

References

Chachkhiani, D. (2021). Data for: Neurological Complications of COVID-19 Predict Worse Outcomes During Hospitalization. *Mendeley Data*. doi: 10.17632/njhc957dfj.2

Lantz, B. (2015) *Machine Learning with R: Discover How to Build Machine Learning Algorithms, Prepare Data, and Dig Deep into Data Prediction Techniques with R*. Packt Publishing.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276–282.

Nwanganga, F., Chapple, M. (2020) *Practical Machine Learning in R*. Wiley.

Yu-Wei, C. (2015). *Machine Learning with R Cookbook Explore over 110 Recipes to Analyze Data and Build Predictive Models with the Simple and Easy-to-Use R Code*. Packt Publishing.