Hospital Readmissions EDA & Modeling

Pablo Diaz
M.Sc Business Analytics

GRMC

Great River Medical Center

AGENDA

HULT INTERNATIONAL BUSINESS SCHOOL

1. PROBLEM STATEMENT
2. DATASETS
3. DATA EXPLORATION
4. FEATURE SELECTION
5. MODEL DEPLOYMENT AND EVALUATION
6. CONCLUSION

GRMC

PROBLEM STATEMENT

HULT INTERNATIONAL BUSINESS SCHOOL

COST OF ADMISSIONS

The cost of readmissions within 30 days is large for patient and hospital, ranging from $10,900 to $15,200

TOP CAUSES OF HOSPITAL READMISSION

1. Septicemia
2. Heart Failure
3. Diabetes
4. COPD

GRMC

*Didn’t encode factor features due to the large amount of existing columns

DATASETS

HULT INTERNATIONAL BUSINESS SCHOOL

FILE NAME	ROWS	COLUMNS	DESCRIPTION
diabetesHospitalInfoTrain.csv	7500	16	information gathered at the hospital i.e.. number of lab procedures and diagnosis descriptions
diabetesHospitalInfoTest.csv	2500	16	
diabetesMedsTrain.csv	7500	23	patients medicine information i.e.. patient is taking a specific drug like Metformin
diabetesMedsTest.csv	2500	23	
diabetesPatientTrain.csv	7500	7	patients' demographic and insurance information i.e.. race, gender, age, weight
diabetesPatientTest.csv	2500	7	

GRMC

DATA EXPLORATION

FILE NAME	ROWS	COLUMNS	DESCRIPTION
patientsComined.csv	10,000	45	All data from the previous datasets

* Creating train/test column to distinguish observations

DATA PREPARATION

* Removed columns with only one value or +95% of observations concentrated in 1 value (Low cardinality)
* Discarded features diagnosis 2 and 3 since we don't know if they were readmitted before 30 days after the first diagnosis
* Discarded Discharge Disp. values related to patients' death and Hospice
* Divided columns into factor and numeric
* Checked for outliers

GRMC

DATA EXPLORATION

MISSING VALUES IMPUTATION

NUMERIC

No N.A. values found in the dataset

FACTOR

Vtreat as an algorithm for feature manipulation due to the low quantity of features with missing values

FEATURE SELECTION

* Splited training data into treatment plan data (10%), training data (72%) and validation data (18%)

* Applied treatment plan to perform feature engineering

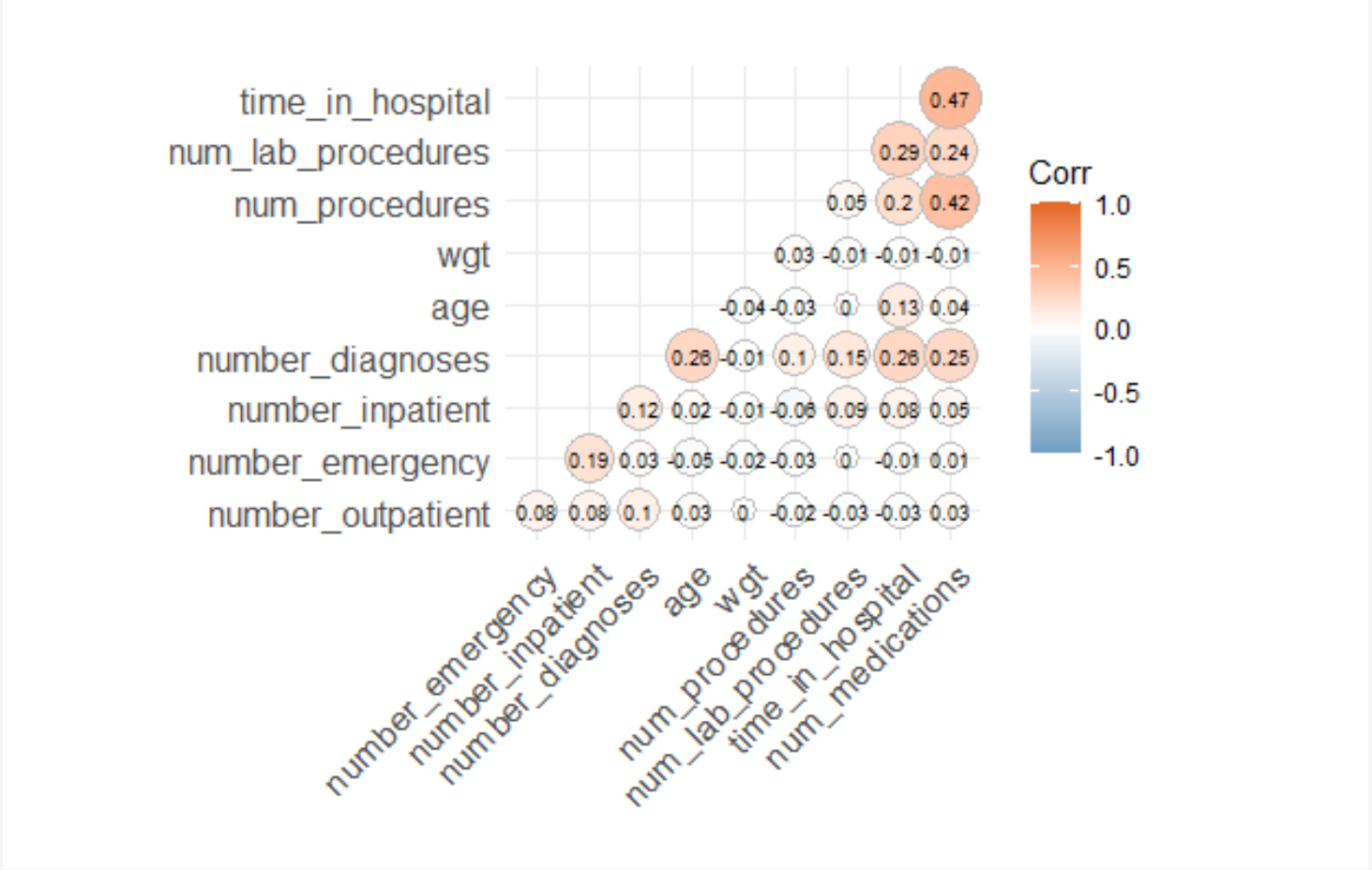
* Applied Boruta feature selection algorithm (random forest model & shadow feature set) to identify important features.

*Discarded redundant and duplicated features

DATA EXPLORATION

HULT INTERNATIONAL BUSINESS SCHOOL

CORRELATION MATRIX

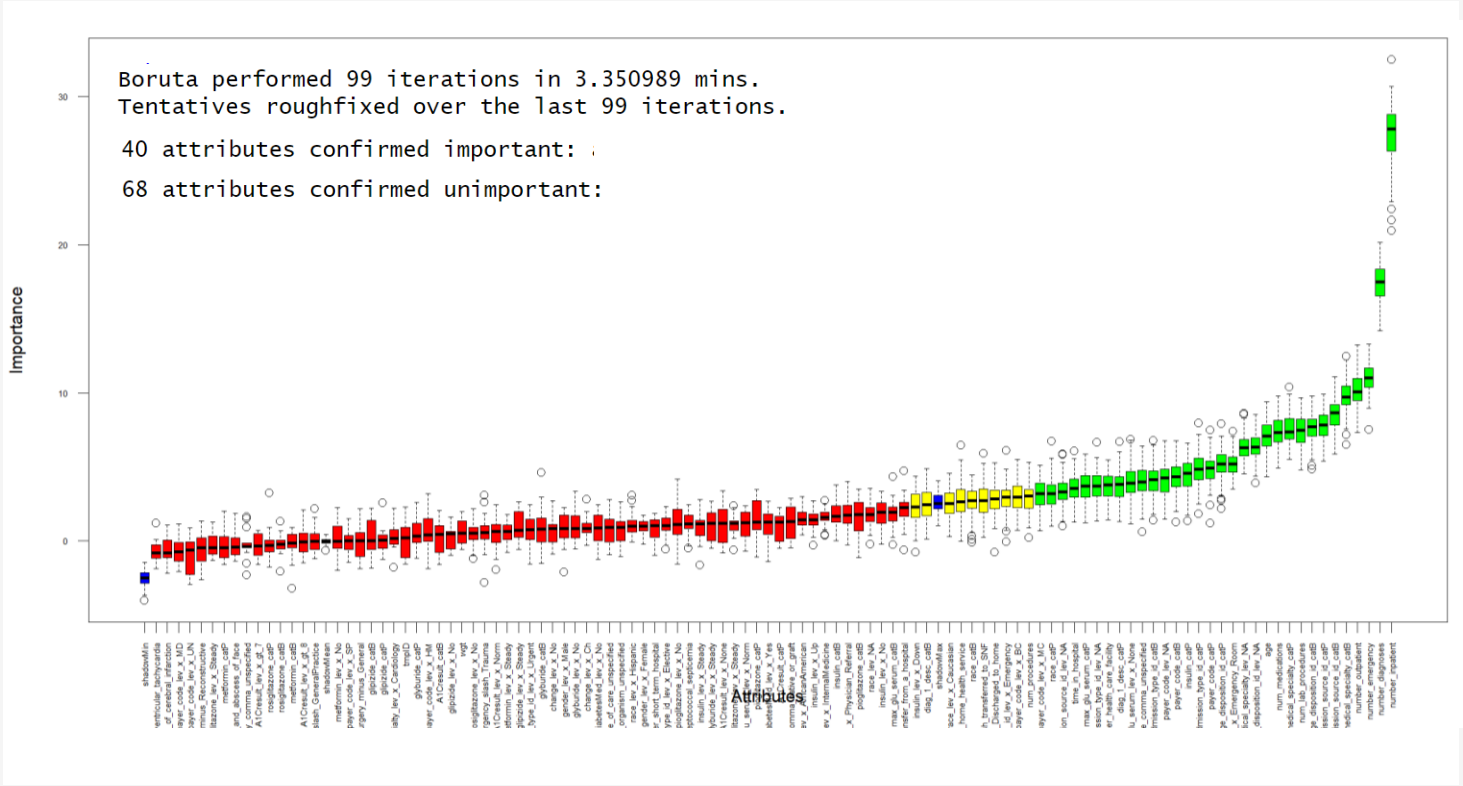


GRMC

FEATURE SELECTION

HULT INTERNATIONAL BUSINESS SCHOOL

BOROUTA FEATURE IMPORTANCE



GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

MODEL LIST

- LOGISTIC REGRESION
- DECISION TREE
- NAIBE BAYES
- RANDOM FOREST
- NEURAL NETWORKS
- GRADIENT BOOSTING

* Models trained and evaluated with training and validation data sets.
* Performed backward stepwise regression to eliminate noise in LR model.

GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

ACCURACY COMPARISON

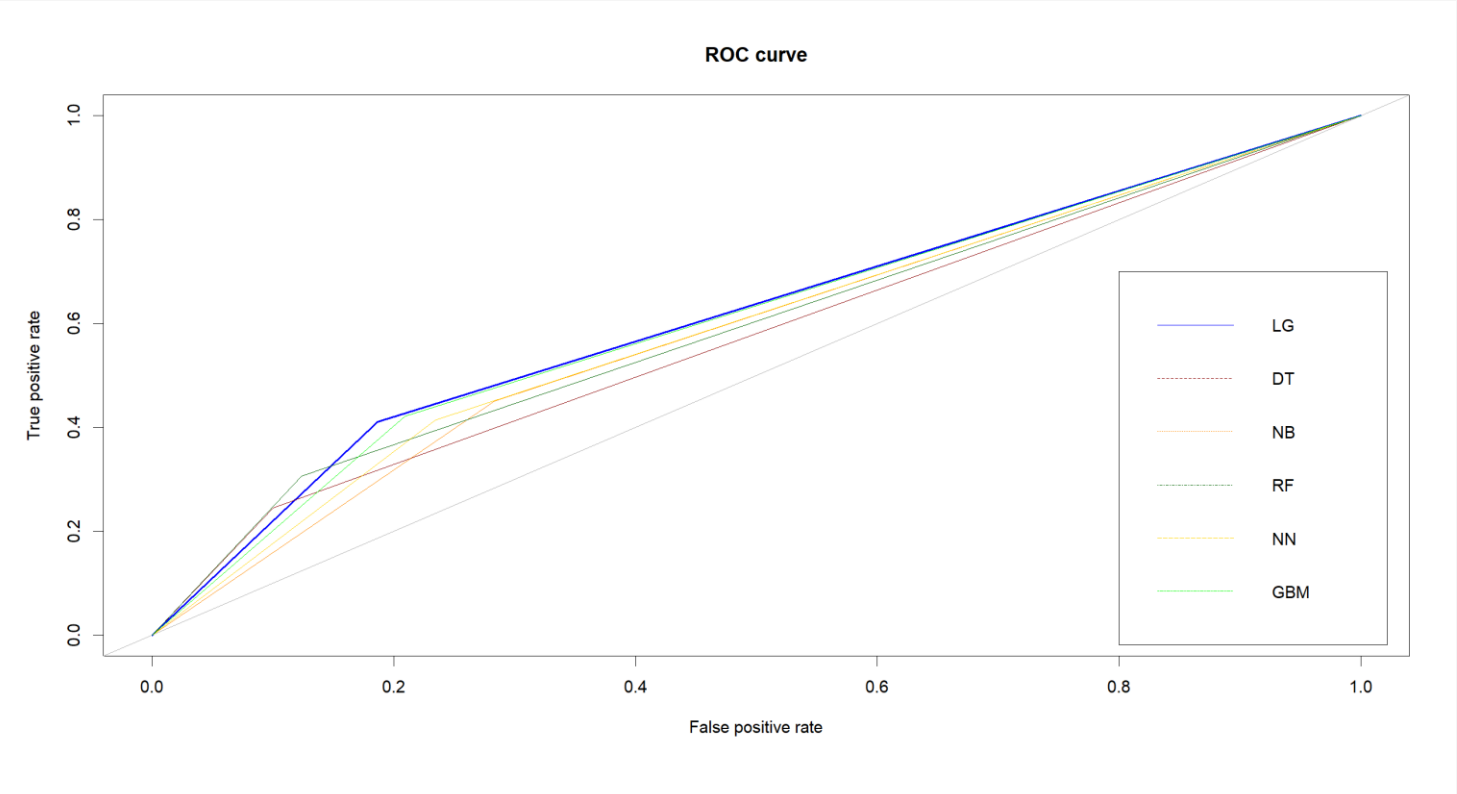
```
Models: LR, DT, RF, NB, NN, GB
Number of resamples: 10

Accuracy
        Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
LR 0.6337761 0.6359758 0.6495696 0.6482410 0.6559080 0.6736243    0
DT 0.5958254 0.6212663 0.6324774 0.6298212 0.6365800 0.6565465    0
RF 0.6349810 0.6399431 0.6454373 0.6467222 0.6527514 0.6653992    0
NB 0.5920304 0.6200190 0.6283270 0.6296322 0.6440421 0.6546490    0
NN 0.6242884 0.6364510 0.6444867 0.6482475 0.6589112 0.6793169    0
GB 0.6280835 0.6389546 0.6425856 0.6503333 0.6630237 0.6787072    0

Kappa
        Min.   1st Qu.    Median      Mean   3rd Qu.      Max. NA's
LR 0.1968504 0.2109614 0.2386247 0.2343025 0.2523625 0.2847358    0
DT 0.0797967 0.1563945 0.1845117 0.1711788 0.2017906 0.2251322    0
RF 0.1793672 0.1975395 0.2083110 0.2077823 0.2174913 0.2488763    0
NB 0.1307967 0.1907341 0.2157001 0.2132613 0.2391249 0.2740497    0
NN 0.1849879 0.2158363 0.2376605 0.2441968 0.2731321 0.3148104    0
GB 0.1873171 0.2206987 0.2270323 0.2420725 0.2702509 0.3042001    0
```

GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

ROC CURVE

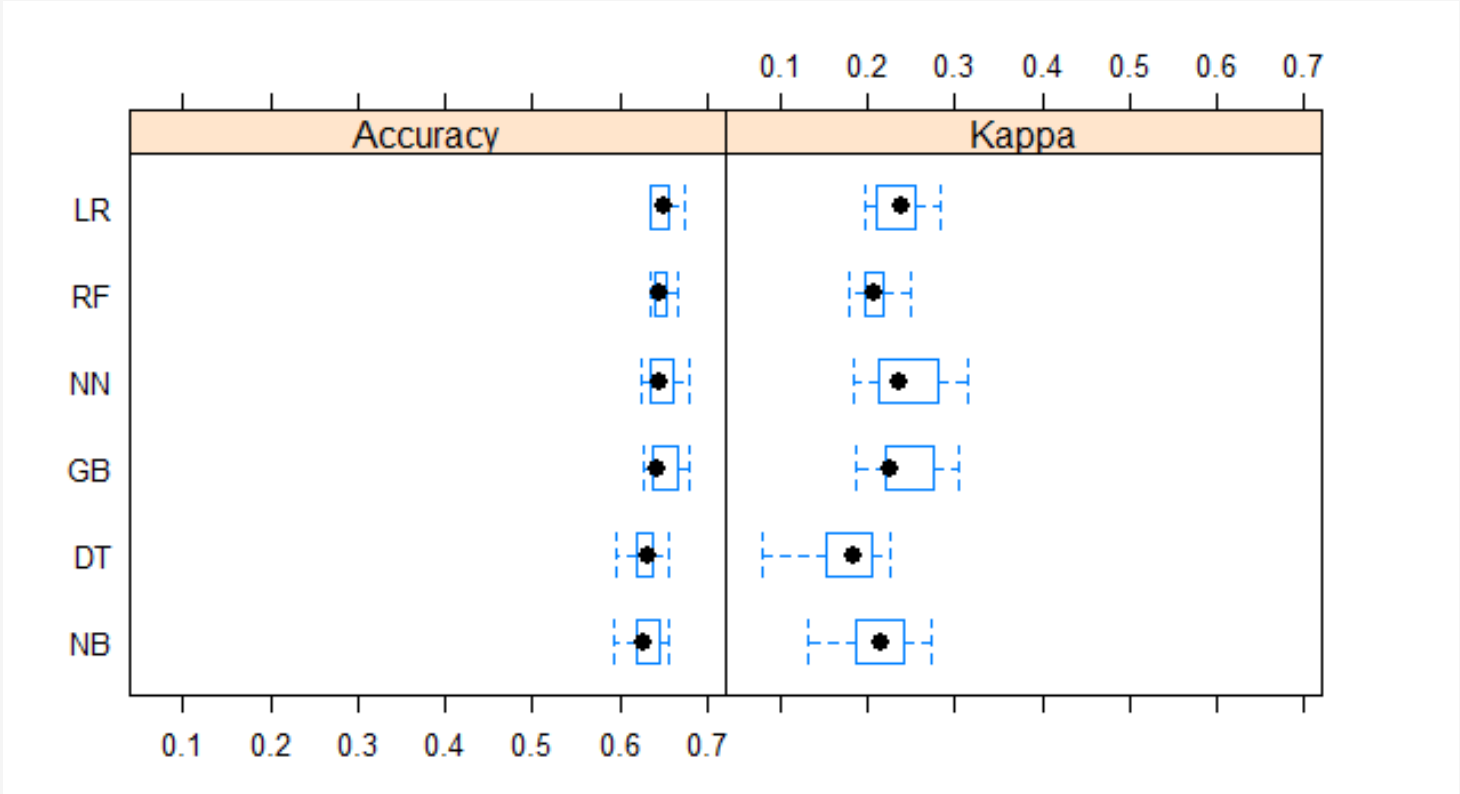


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

ROC CURVE



GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

CONFUSION MATRIX

GRADIENT BOOSTING

Prediction	Reference FALSE	TRUE
FALSE	631	300
TRUE	167	219

Sensitivity	0.7907
Specificity	0.4220
Pos Pred Value	0.6778
Neg Pred Value	0.5674
Prevalence	0.6059
Detection Rate	0.4791
Detection Prevalence	0.7069
Balanced Accuracy	0.6063

RANDOM FOREST

Prediction	Reference FALSE	TRUE
FALSE	699	360
TRUE	99	159

Sensitivity	0.8759
Specificity	0.3064
Pos Pred Value	0.6601
Neg Pred Value	0.6163
Prevalence	0.6059
Detection Rate	0.5308
Detection Prevalence	0.8041
Balanced Accuracy	0.5911

GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

CONFUSION MATRIX

NEURAL NETWORKS

```
          Reference
Prediction FALSE TRUE
     FALSE   611  304
     TRUE    187  215

          Sensitivity : 0.7657
          Specificity : 0.4143
       Pos Pred Value : 0.6678
       Neg Pred Value : 0.5348
           Prevalence : 0.6059
       Detection Rate : 0.4639
 Detection Prevalence : 0.6948
    Balanced Accuracy : 0.5900
```

LOGISTIC REGRESSION

```
          Reference
Prediction  no yes
       no  745 416
       yes  53 103

          Sensitivity : 0.9336
          Specificity : 0.1985
       Pos Pred Value : 0.6417
       Neg Pred Value : 0.6603
           Prevalence : 0.6059
       Detection Rate : 0.5657
 Detection Prevalence : 0.8815
    Balanced Accuracy : 0.5660
```

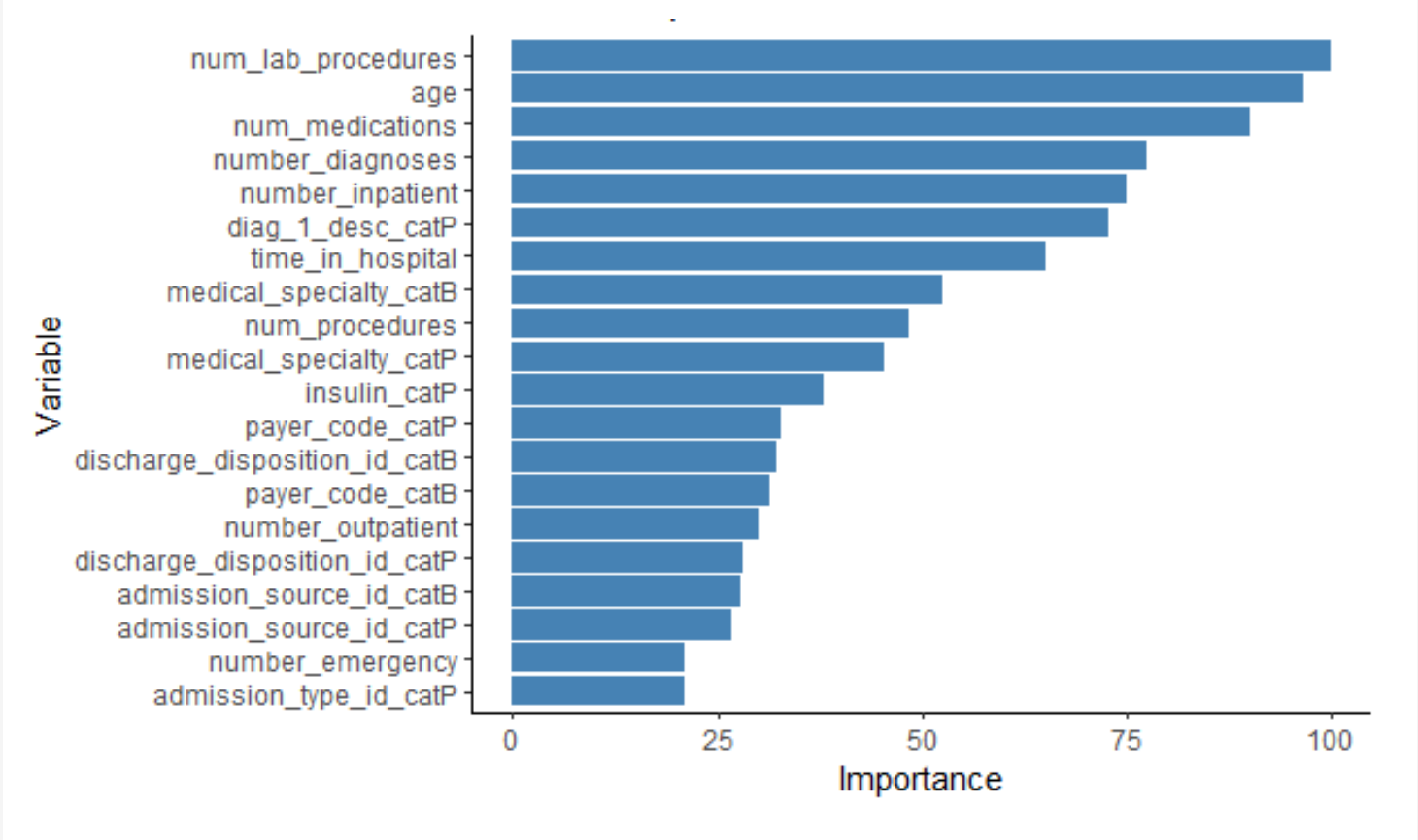
GRMC

MODEL DEPLOYMENT AND EVALUATION

GRMC

HULT INTERNATIONAL BUSINESS SCHOOL

RANDOM FOREST FEATURE IMPORTANCE

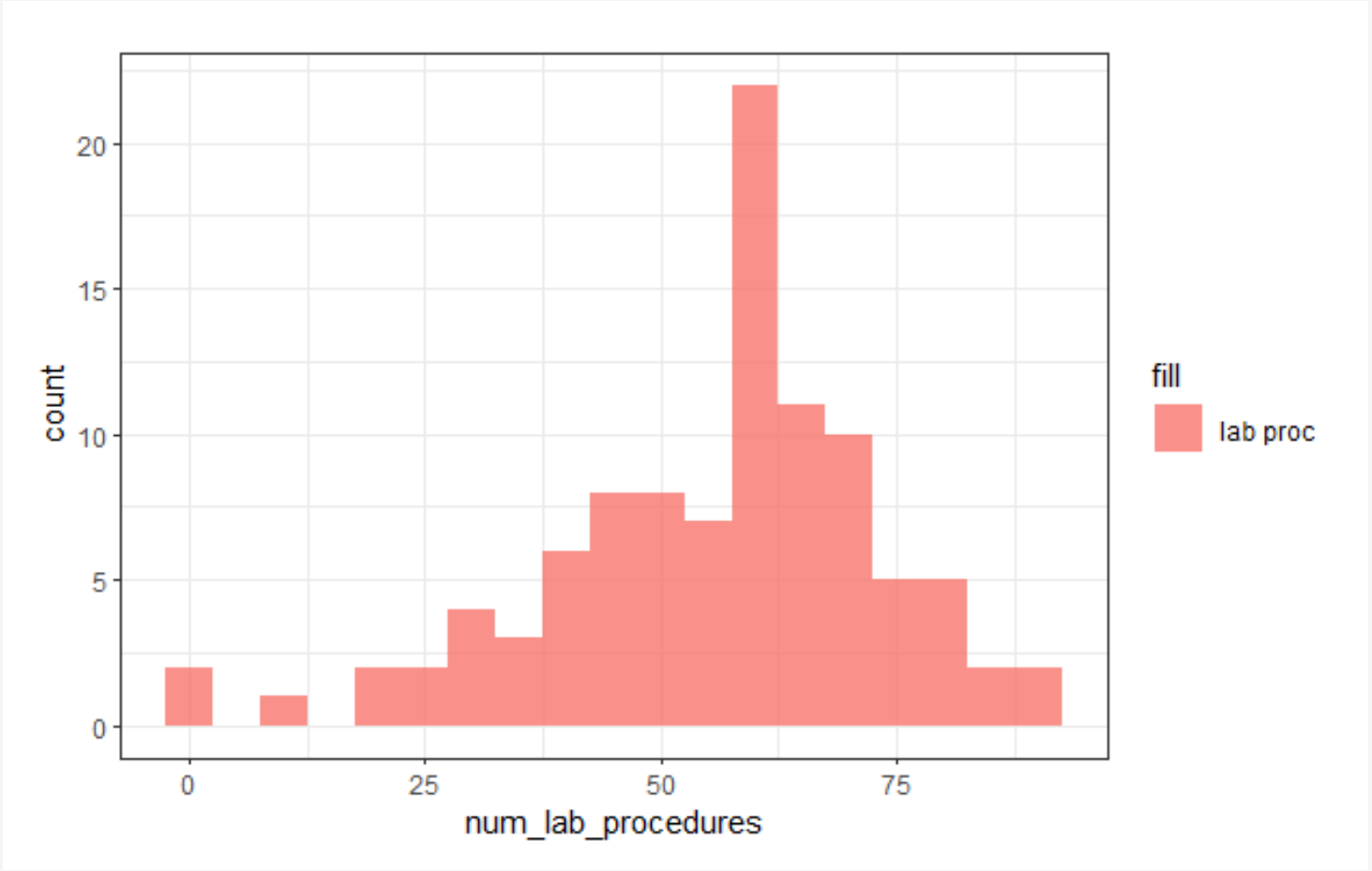


* Top 100 patients with highest probability of readmission with probabilities ranging 66.4% – 87.2%

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF LAB PROCEDURES

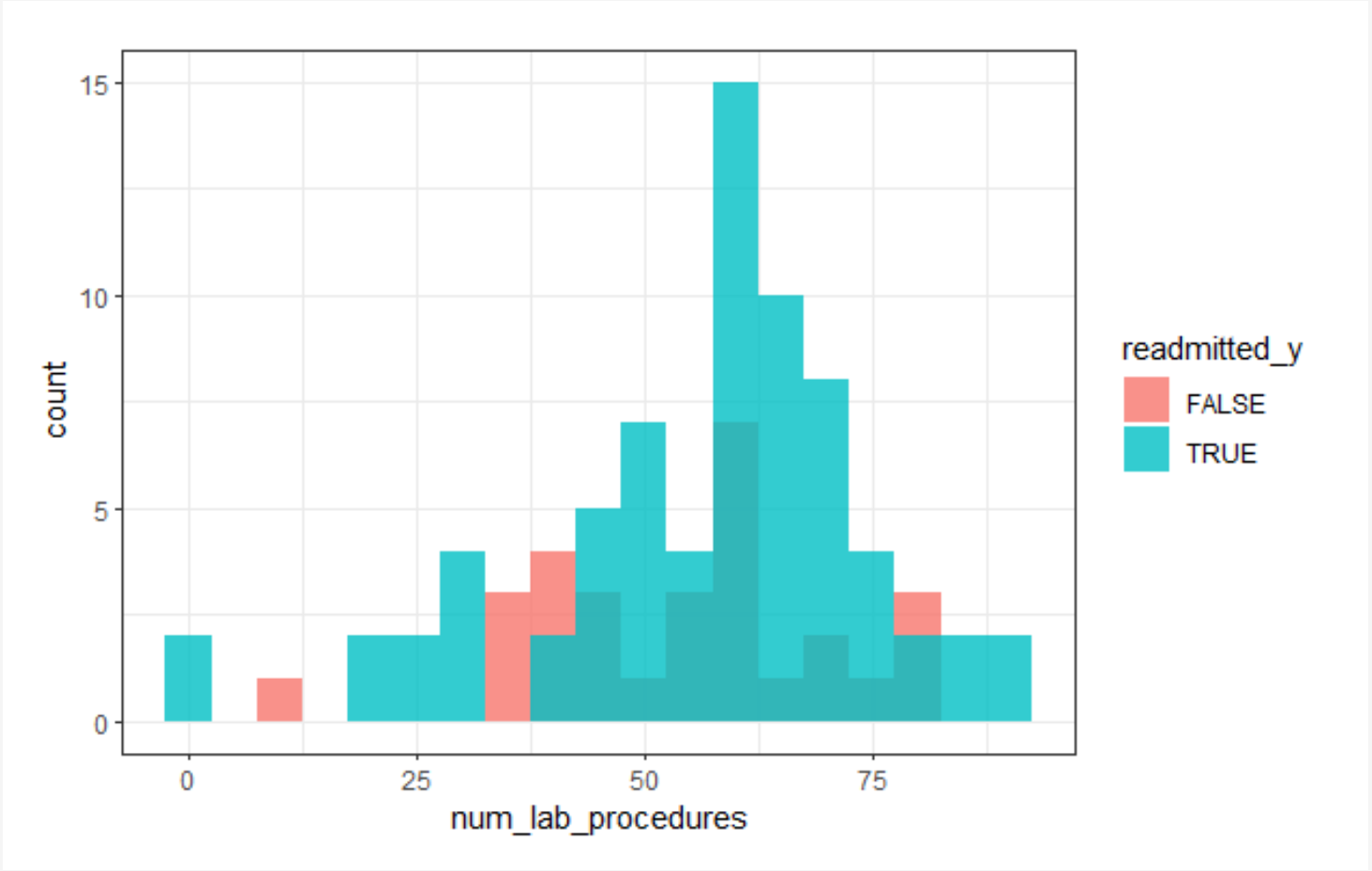


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF LAB PROCEDURES

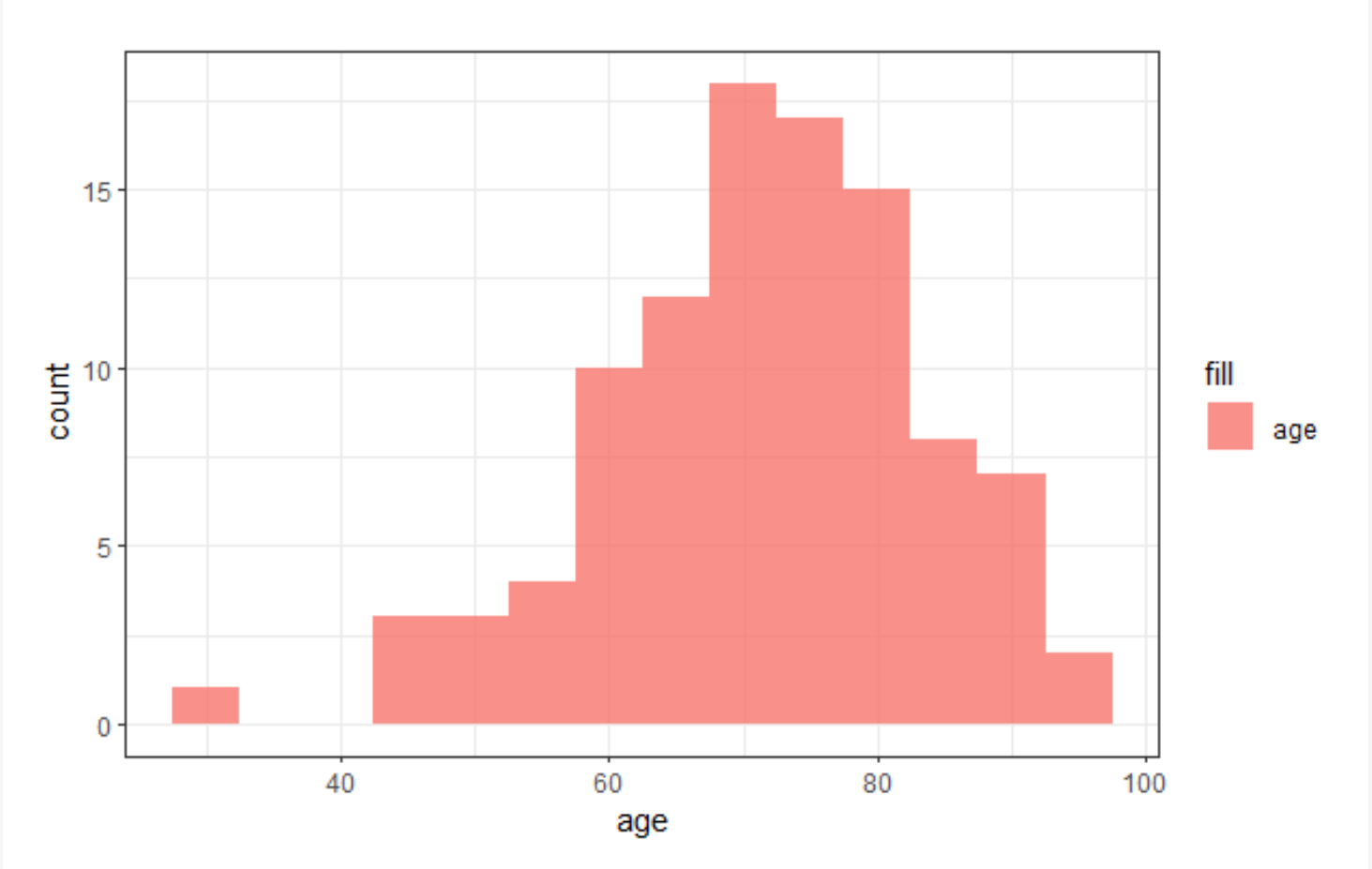


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

AGE OF PATIENTS

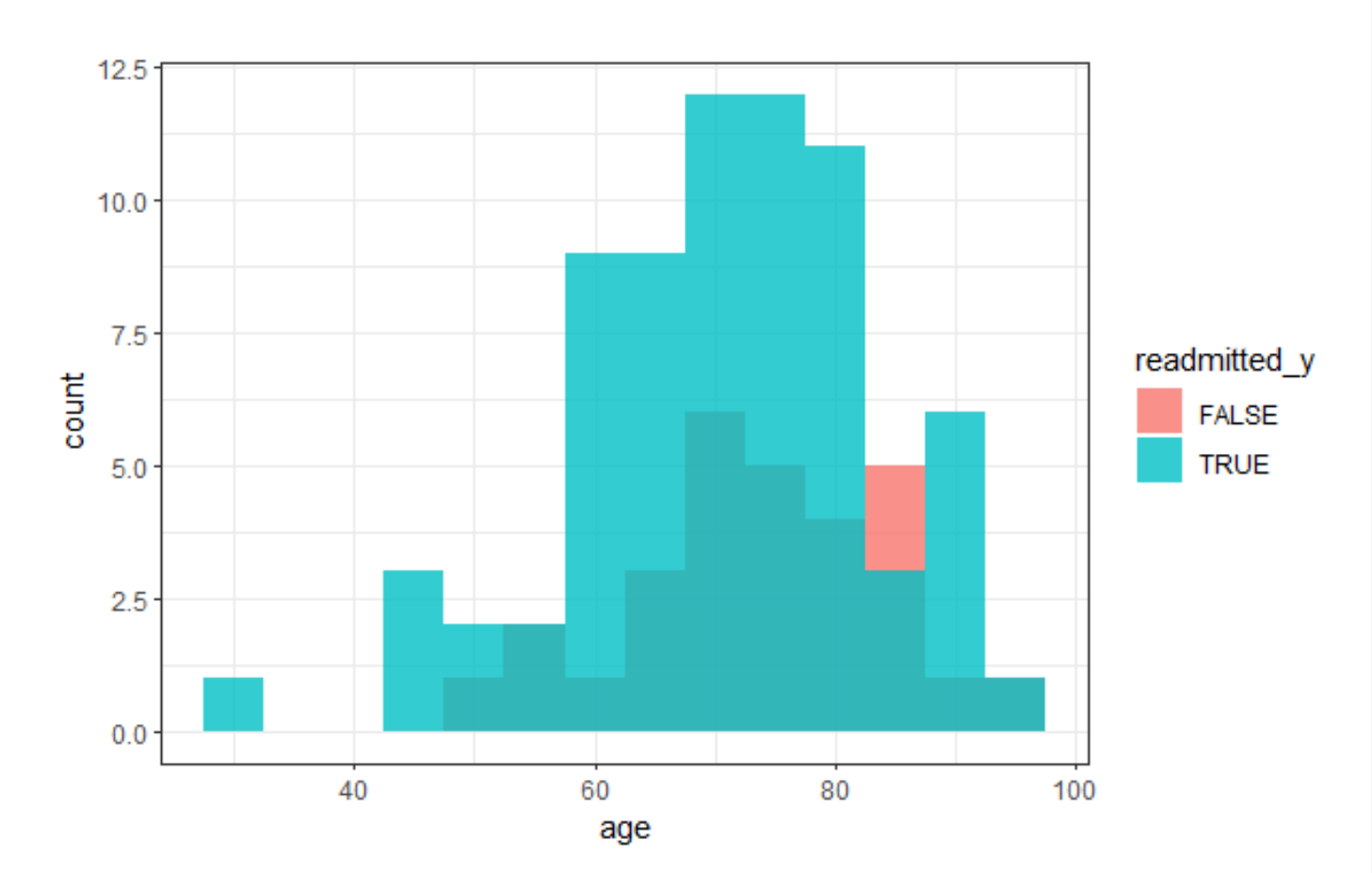


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

AGE OF PATIENTS

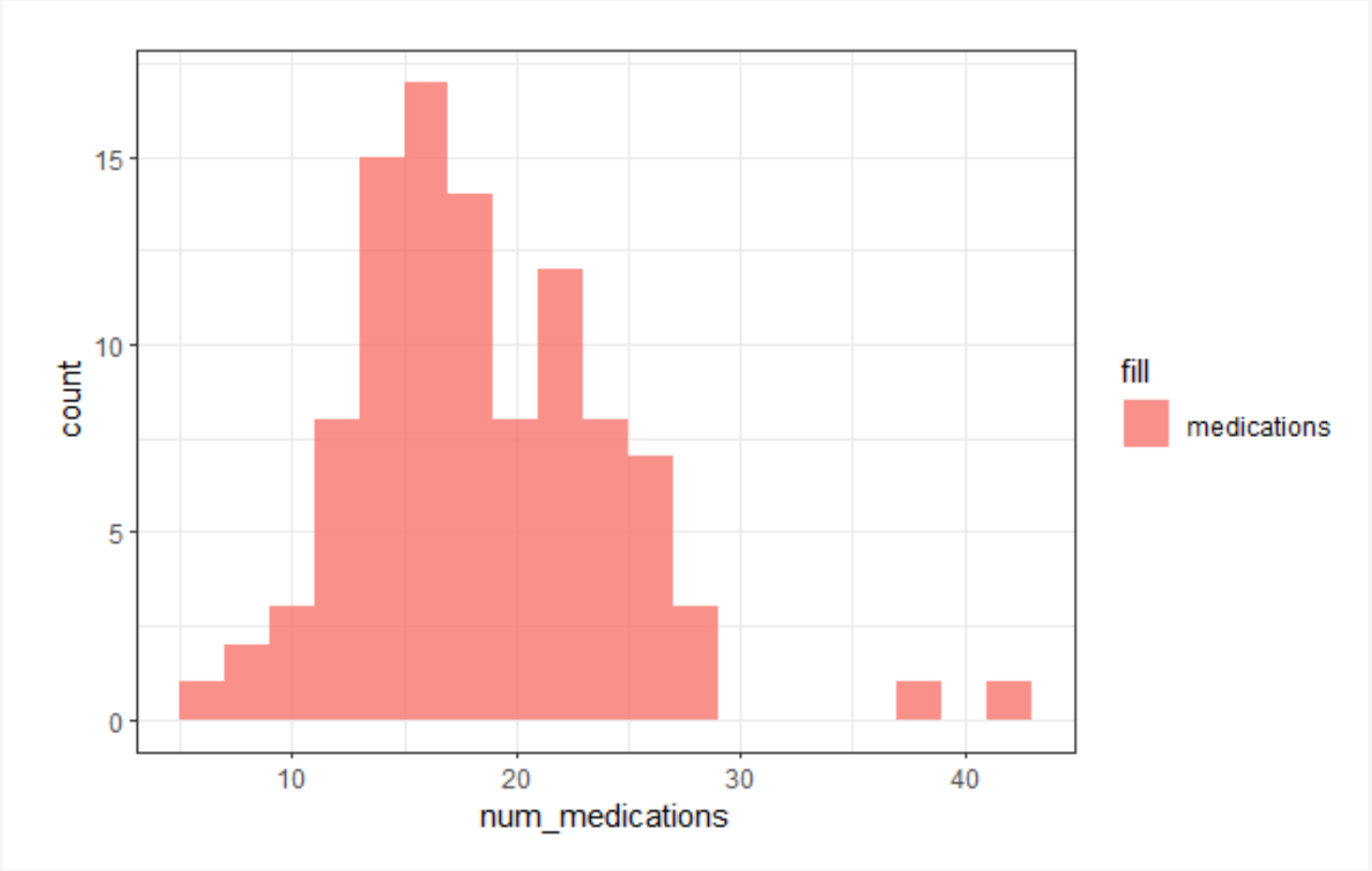


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF MEDICATIONS

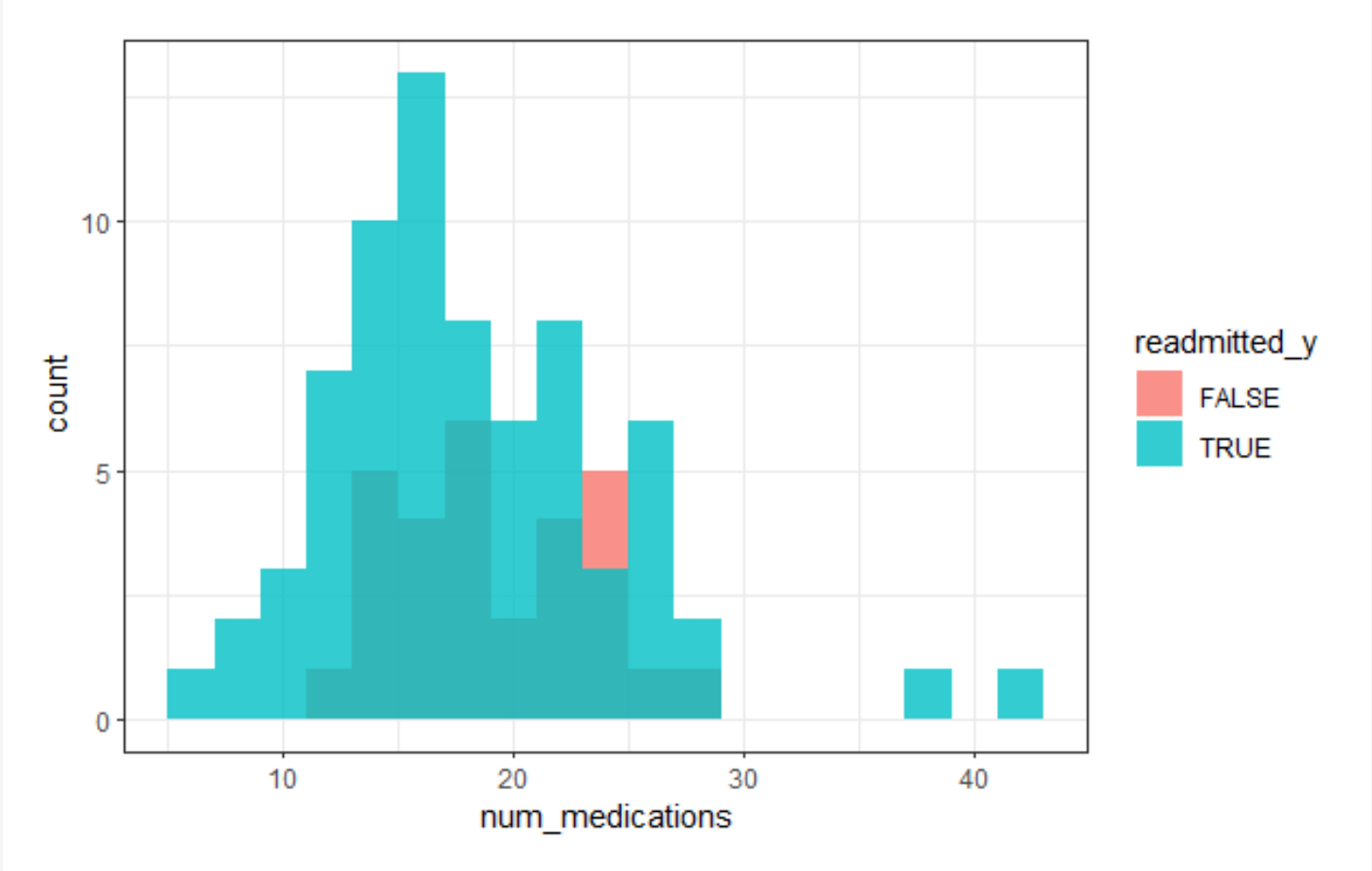


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

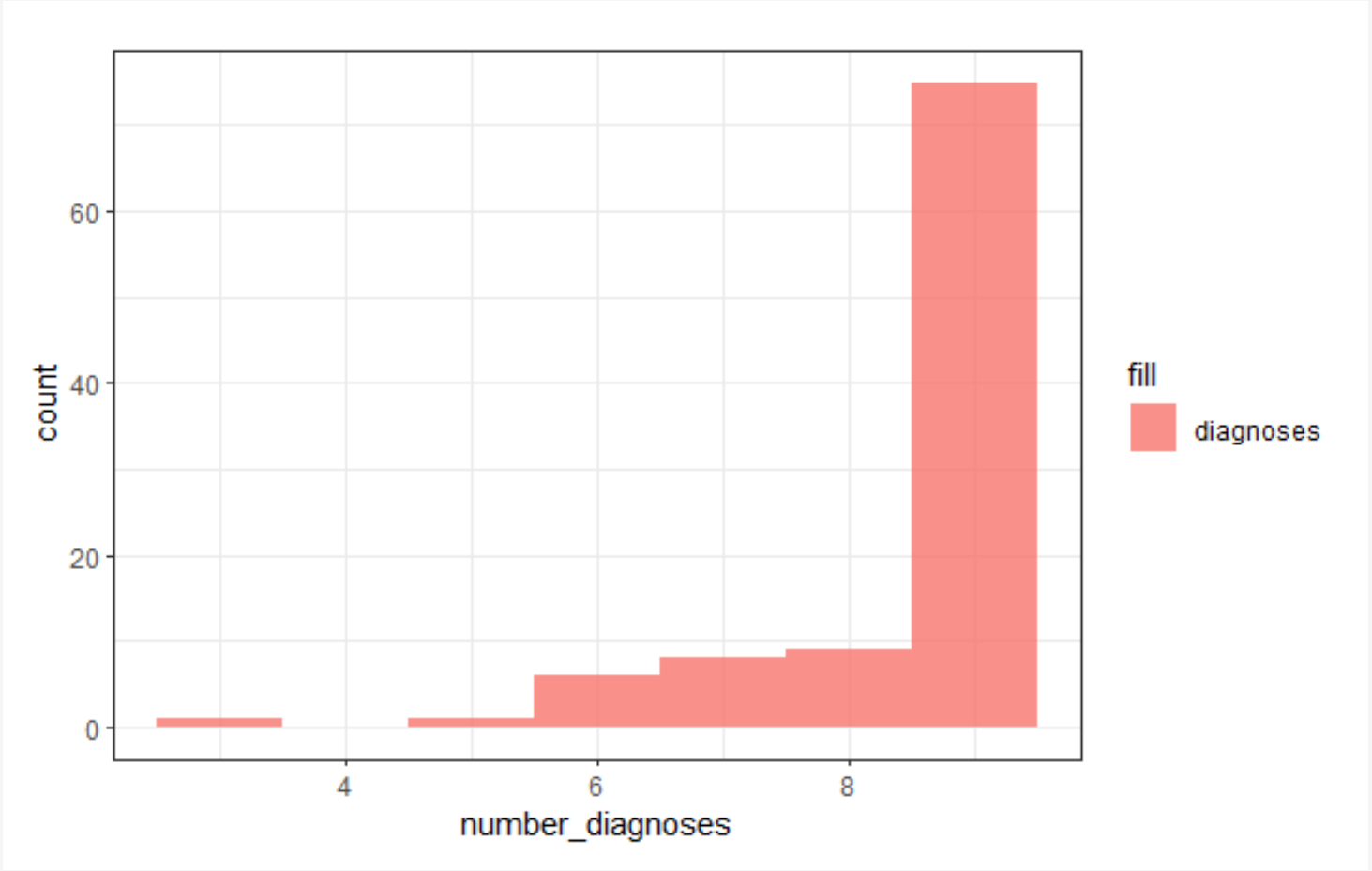
NUMBER OF MEDICATIONS



GRMC

MODEL DEPLOYMENT AND EVALUATION

NUMBER OF DIAGNOSES

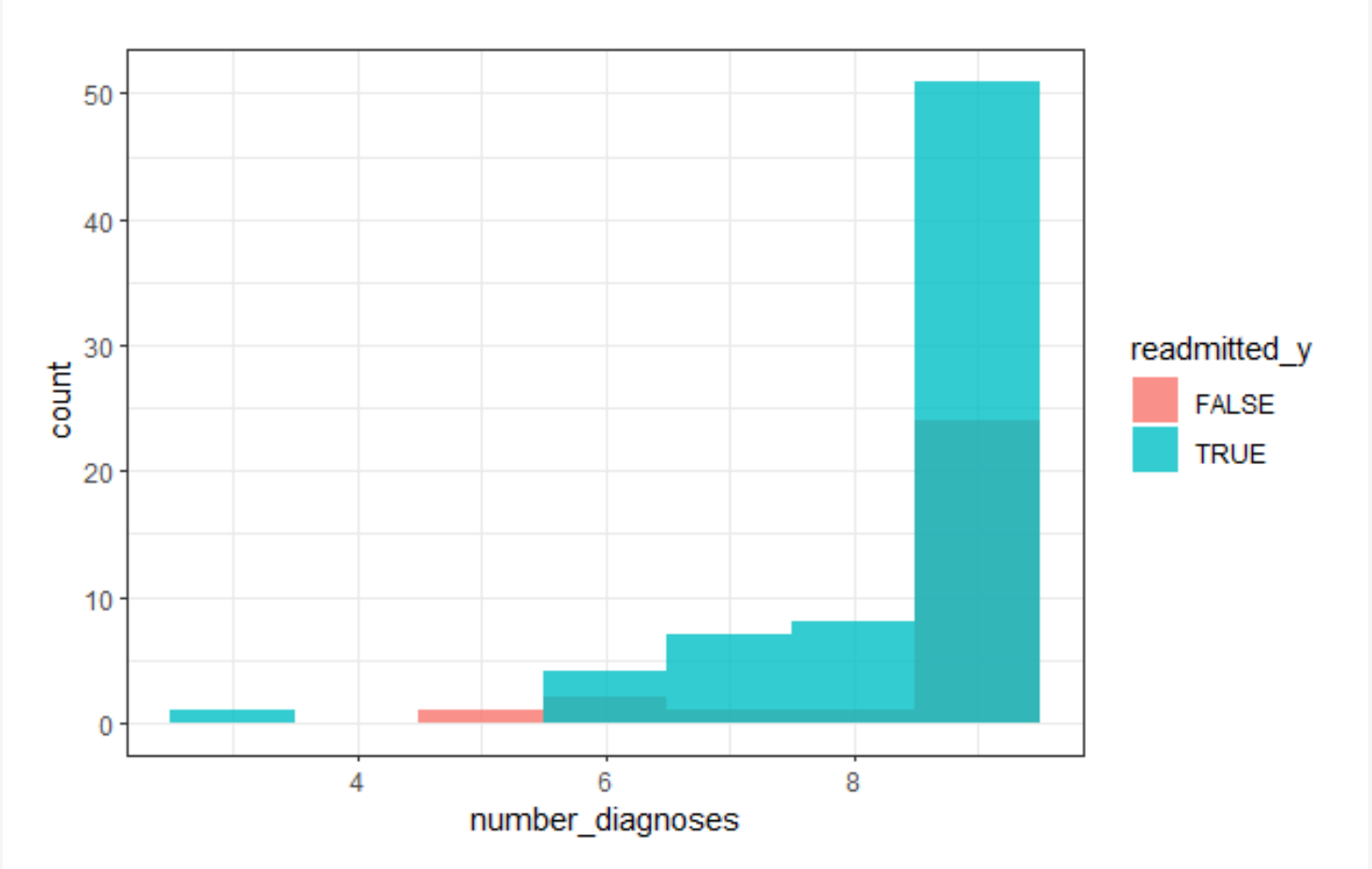


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF DIAGNOSES

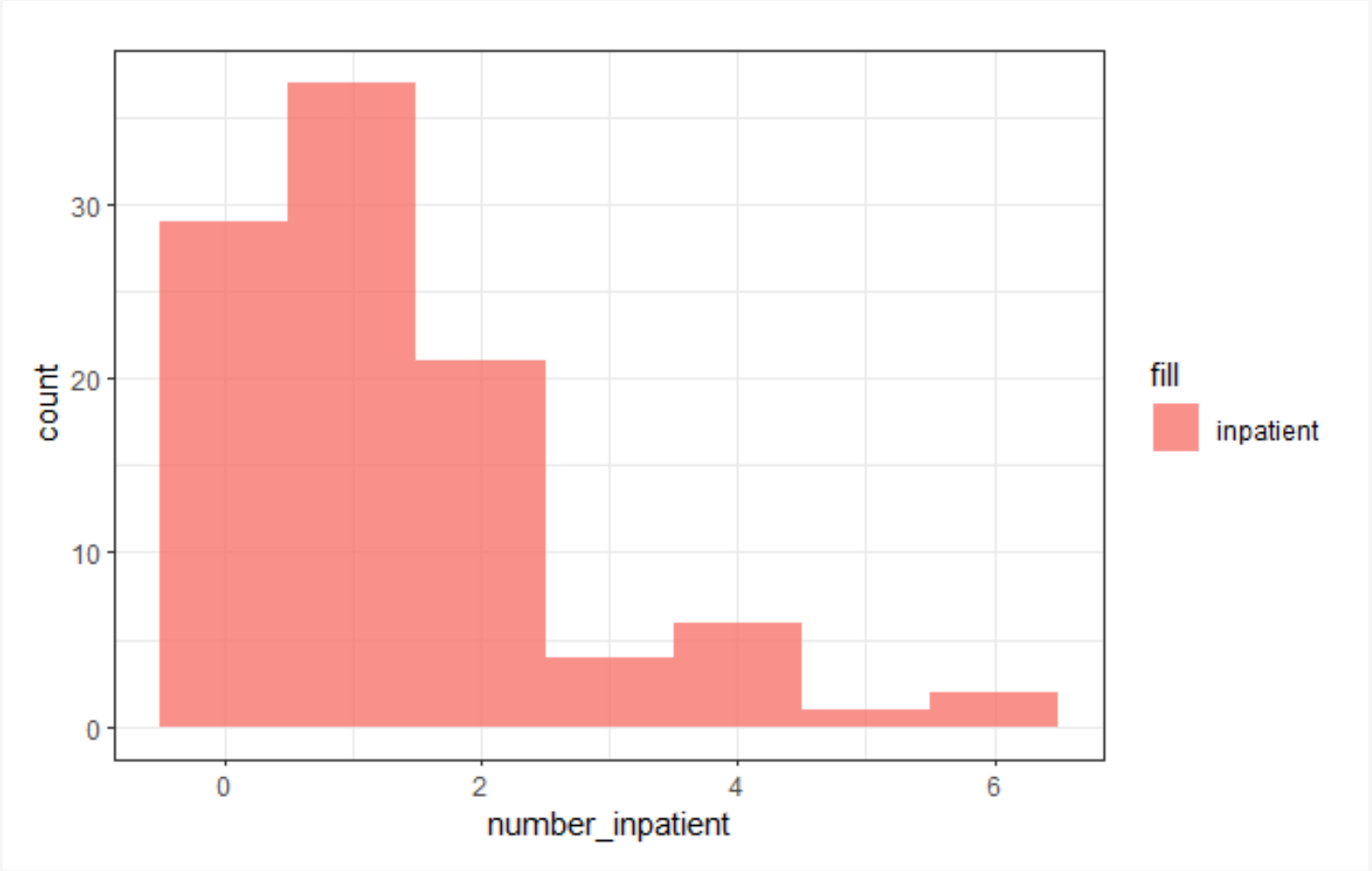


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF INPATIENT

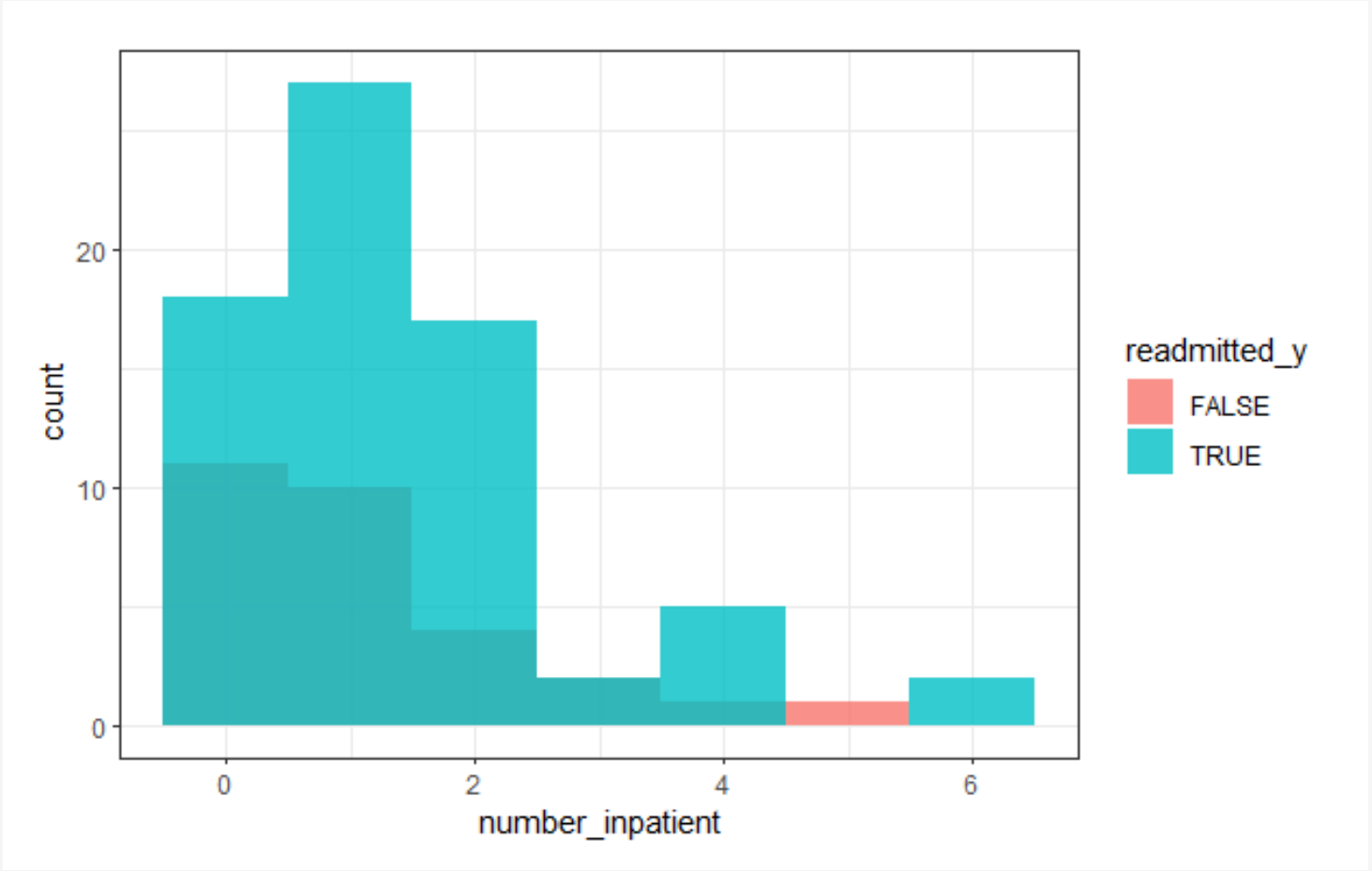


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

NUMBER OF INPATIENT

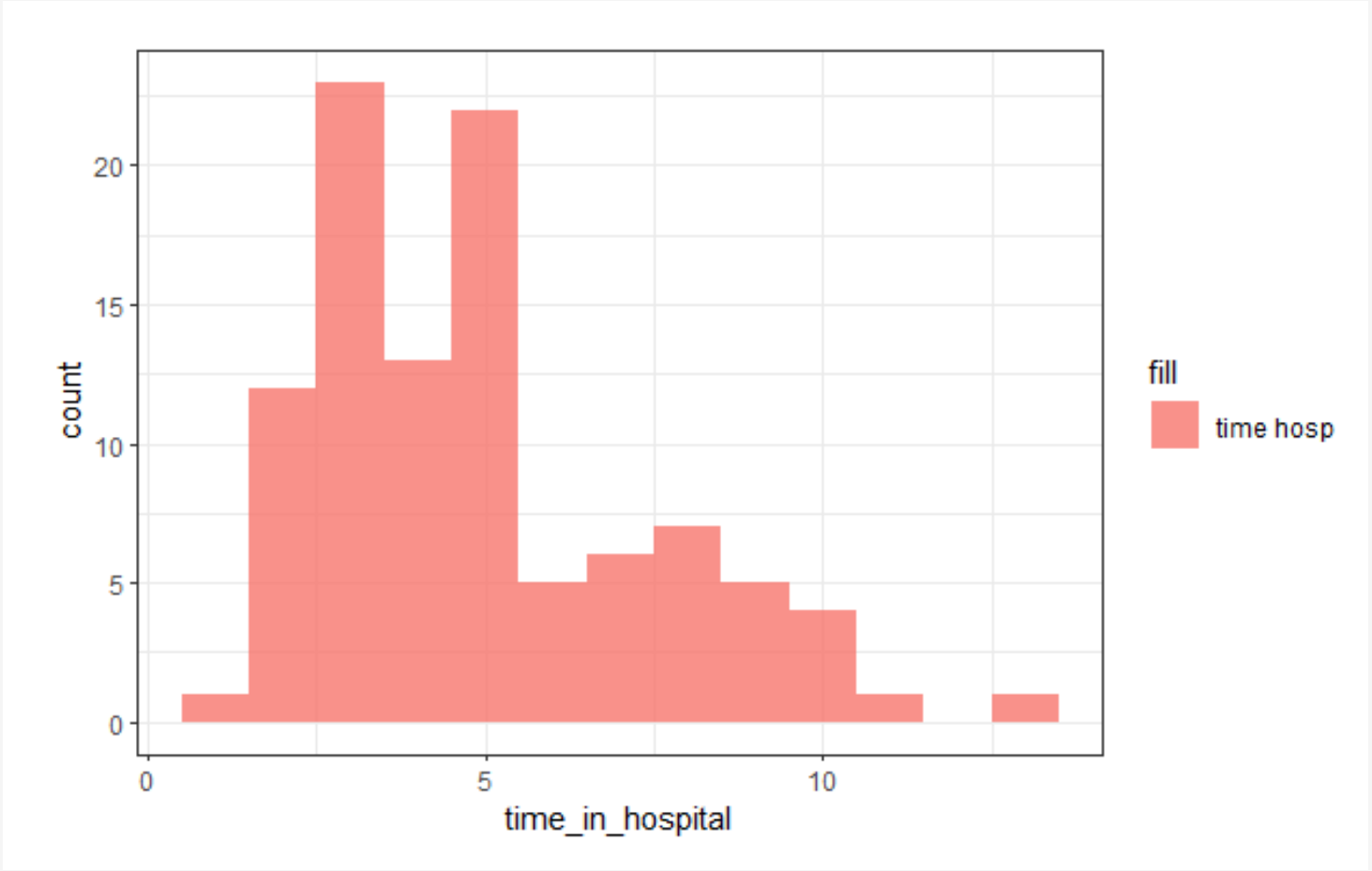


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

TIME IN THE HOSPITAL

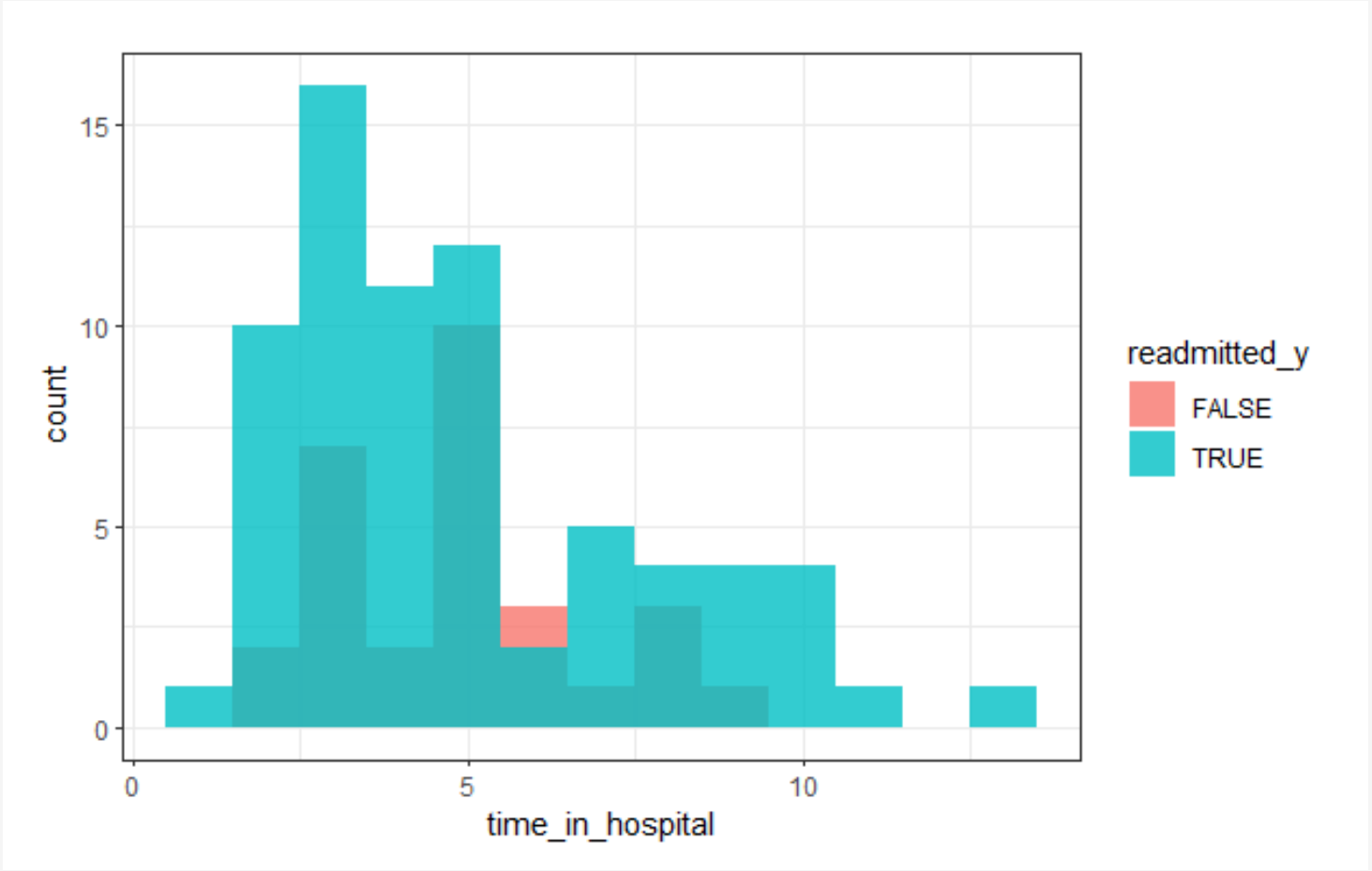


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

TIME IN THE HOSPITAL

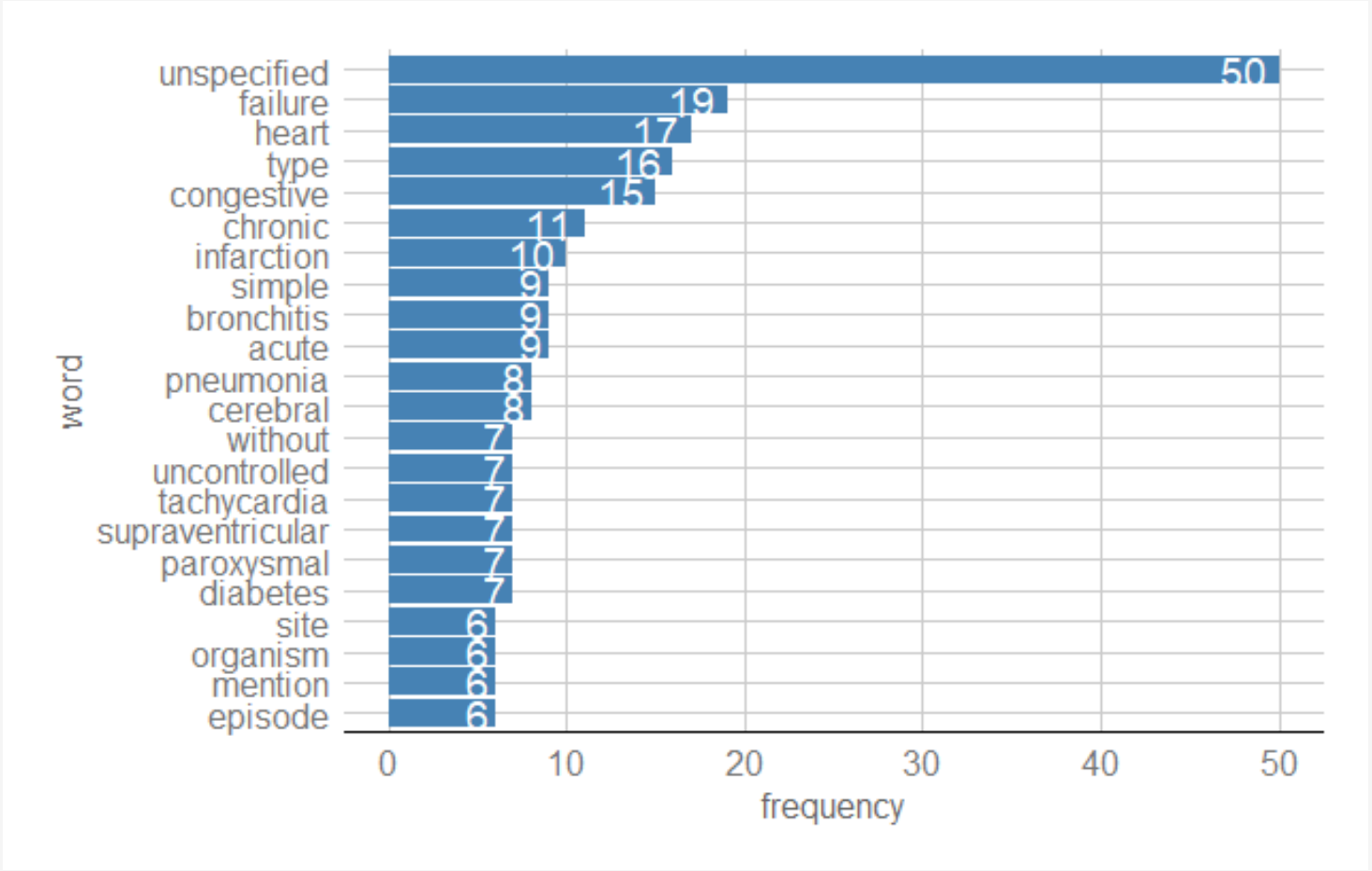


GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

TOP WORDS IN DIAGNOSIS



GRMC

MODEL DEPLOYMENT AND EVALUATION

HULT INTERNATIONAL BUSINESS SCHOOL

TOP WORDS IN DIAGNOSIS

GRMC

CONCLUSION

HULT INTERNATIONAL BUSINESS SCHOOL

In order to decrease readmission rates for diabetic patients, hospitals should focus on several factors

* Monitoring the number of laboratory procedures performed on a patient, as a higher number of procedures can indicate underlying health concerns and increase the risk of readmission

* Pay attention to the age of the patient, as older patients are more susceptible to developing additional health conditions that can lead to complications and readmission.

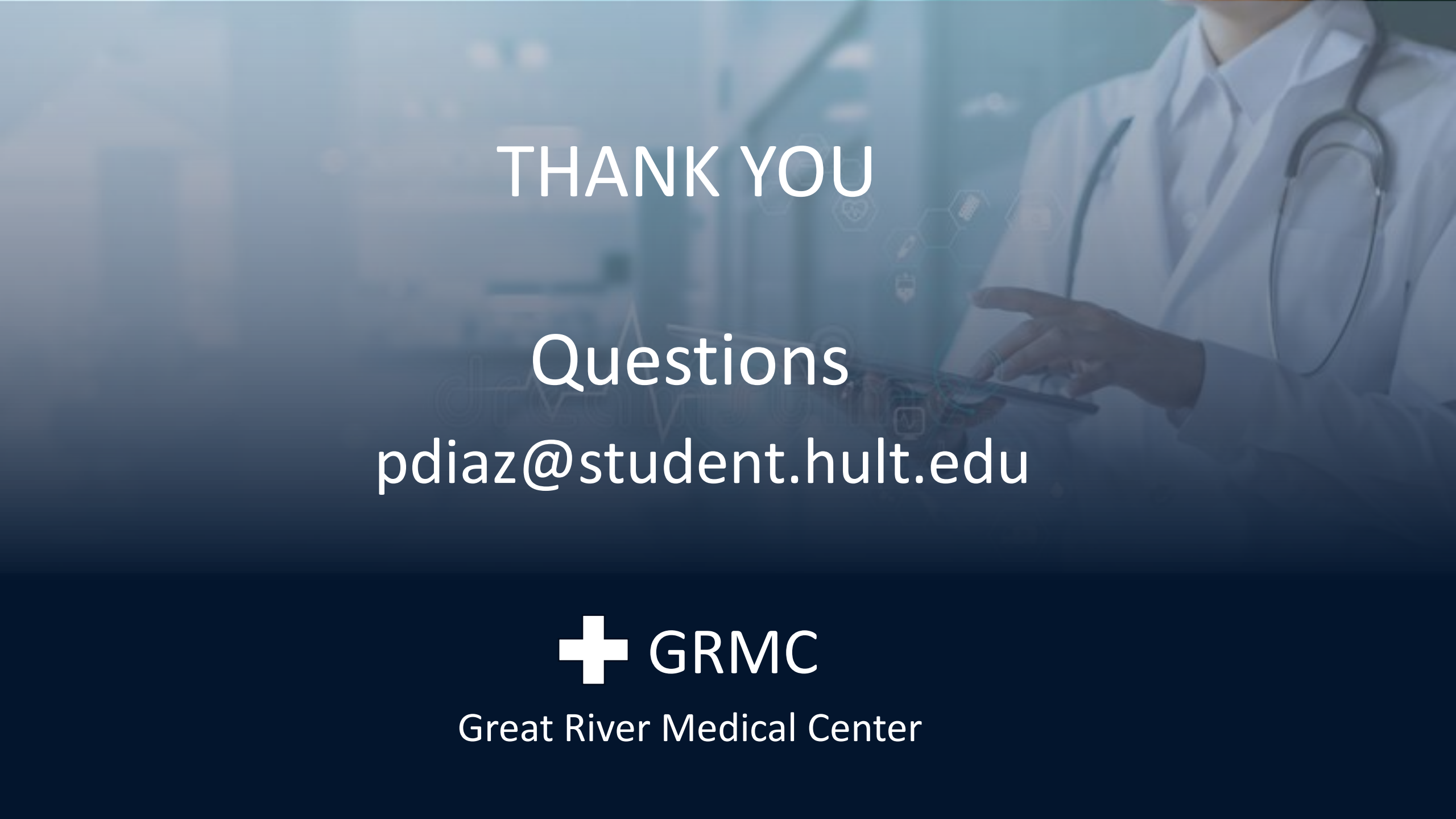
* Number of medications a patient is taking should be monitored, as patients taking a higher number of medications are more likely to be readmitted.

* Patients with multiple diagnoses and inpatient admissions are at a higher risk of readmission

* Hospitals should monitor comorbidities such as heart disease, COPD, bronchitis, pneumonia, and tachycardia in diabetic patients, as these conditions can lead to complications and readmission

By focusing on these factors, hospitals can help decrease readmission rates for diabetic patients and improve overall health outcomes.

GRMC



THANK YOU

Questions

pdiaz@student.hult.edu

GRMC

Great River Medical Center