

Hospital Readmissions

Audrey Arocha | MBAN FY | April 8, 2023
A2: EDA and Modeling | DAT-5323

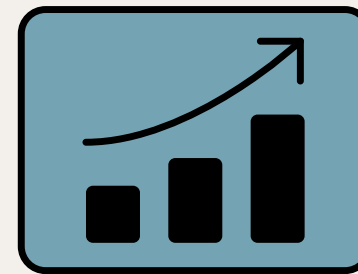


Problems and Objectives



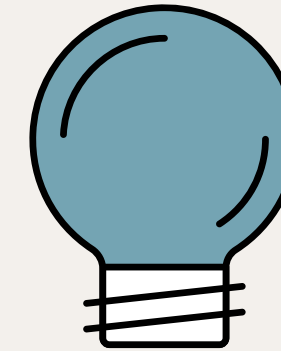
Prediction

How do we predict the probability of patient readmissions?



Analysis

What elements have the greatest impact on this probability?



Action

What can we do better for our patients health and well-being?

Process & Methodology



Sampling

Datasets were already divided between train and test. It was decided for cross-validation to be performed in modeling.

Modification

Included herein was null imputation, dummification, and feature engineering. Some features were also removed.

Analysis

Feature importance analysis and further EDA was done to derive actionable insights and recommendations.

Exploration

Features were explored to identify steps for cleaning and preparation. Relationships and distributions were also noted.

Modeling

The models tested were **Logistic Regression**, **Decision Tree** and **Random Forest** for this classification problem.

Feature Engineering



Simplified Categories

Categorical features were simplified to only top categories, with others lumped together. Upon dummification, uncommon, irrelevant, and undiscerning categories were also then removed (e.g., Expired, Others or None).

Text Analysis for Diagnoses Descriptions

Three descriptions of diagnoses were included in the datasets. These were run through text analyses to capture the most common words. The top 10 words stems were then tallied and added as additional features.

Dropping Obscure, Personalized, or Quasi-constant Features

Features were dropped if they were difficult to decode (e.g., payer_code), too specific to the patient (e.g., identifier), or quasi-constant (e.g., medications that were not present in around 98% of the patients).

Assumptions



Representative Data

The databases provided are representative of the usual diabetic patients that Great River Medical Center get. Particularly in regards to the distribution of ethnicity, gender, and other demographic attributes in the community they serve.

Null Values are MCAR

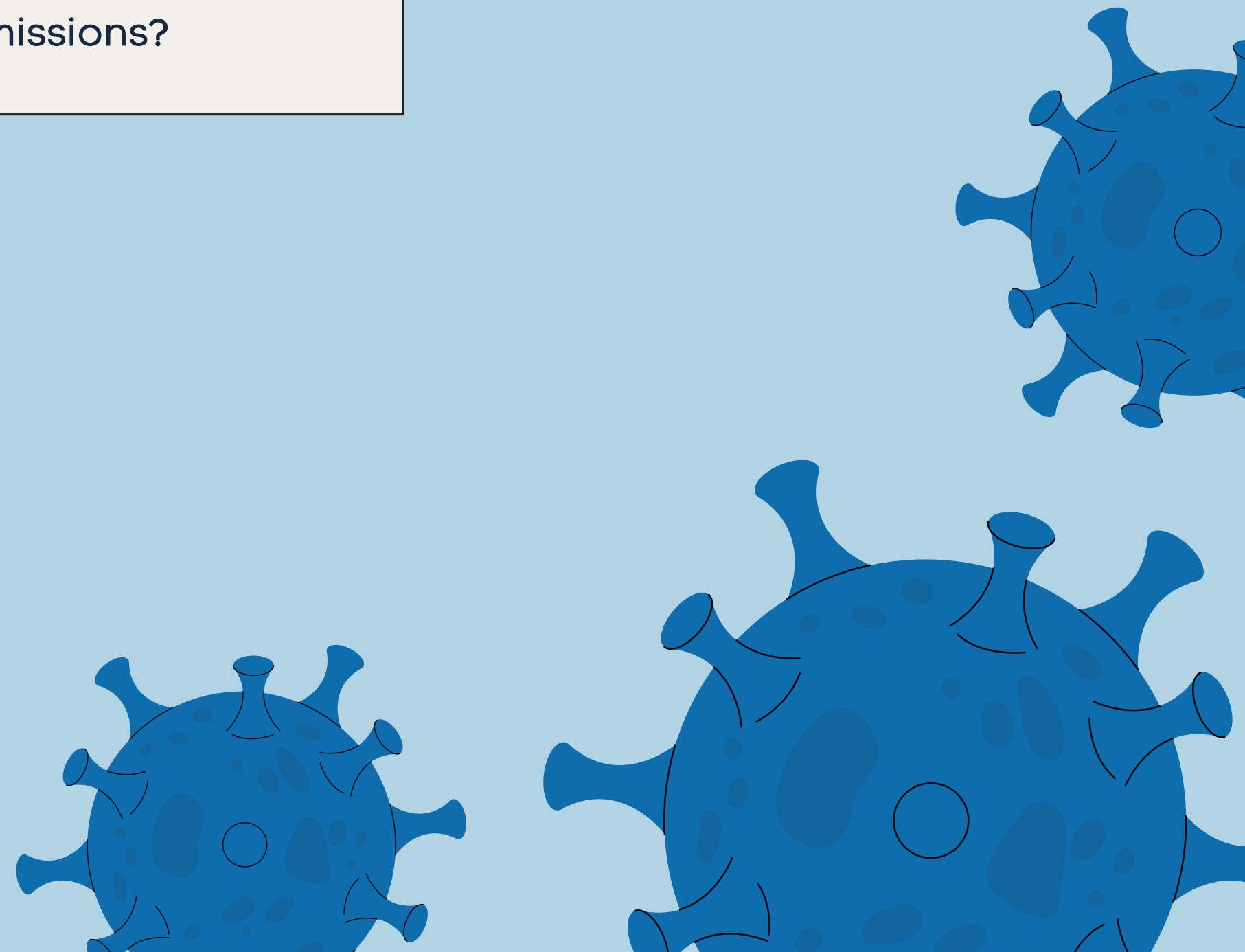
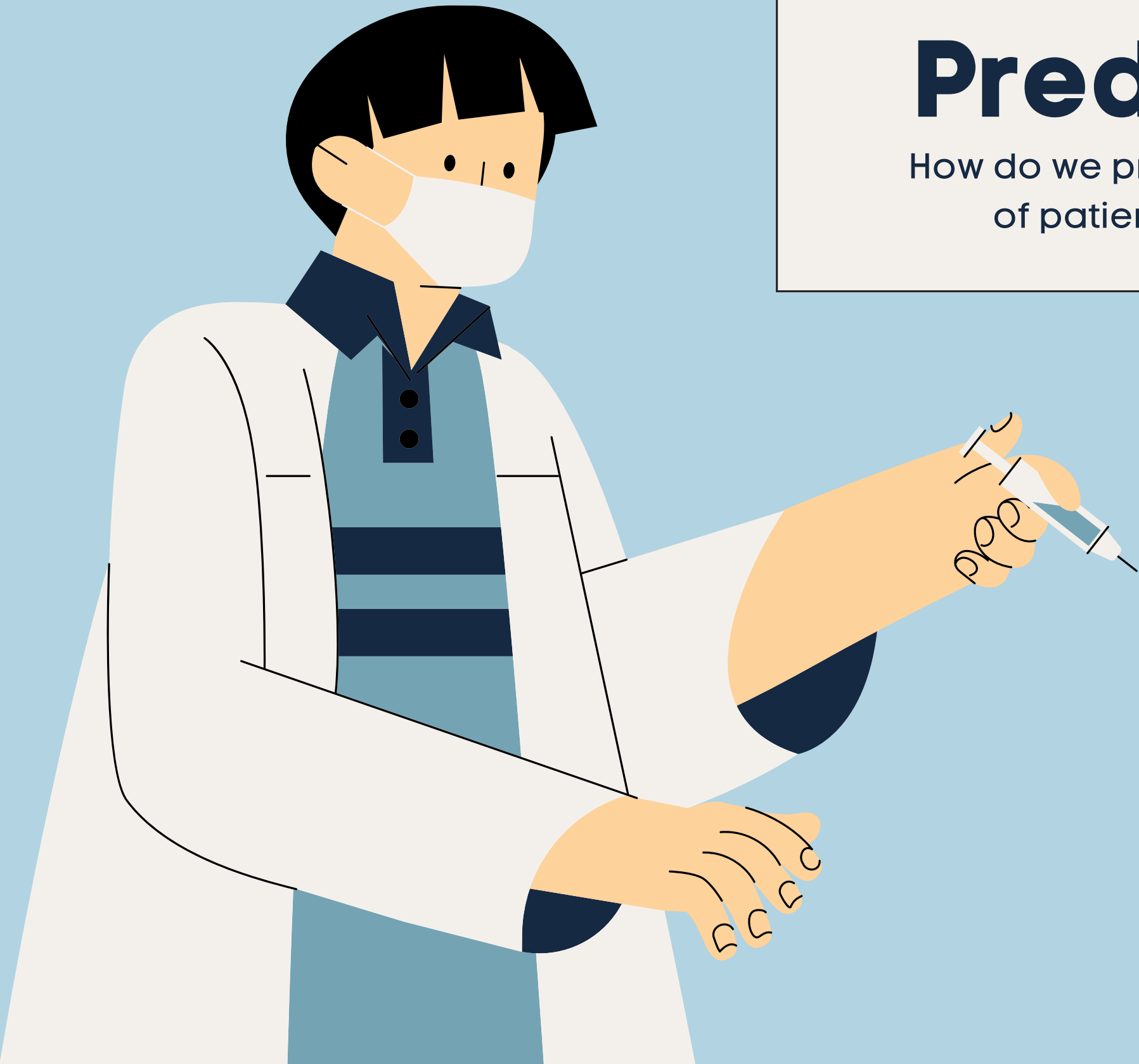
There were minimal null values present. Nonetheless, MICE Predictive Mean Matching (PMM) was used to impute the same. It was assumed that these null values were missing completely at random and were not biased towards other observations or other features.

Non-medical Perspective

The recommendations herein provided are coming from a non-medical perspective. It is solely based on the data provided and other easily obtainable public information. It does not consider intimate knowledge of diabetes, its related symptoms, and its medications.

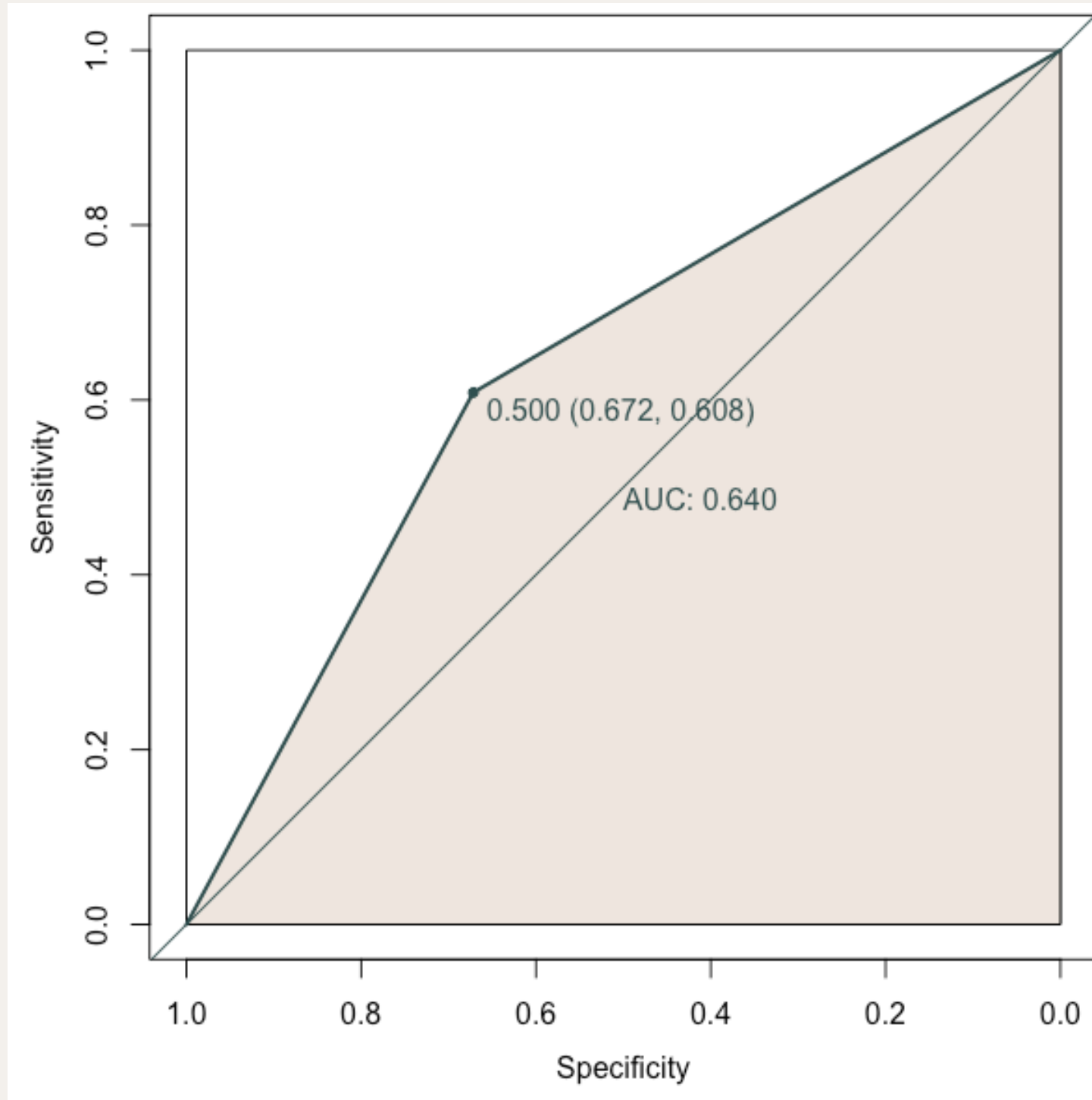
Predictions

How do we predict the probability
of patient readmissions?

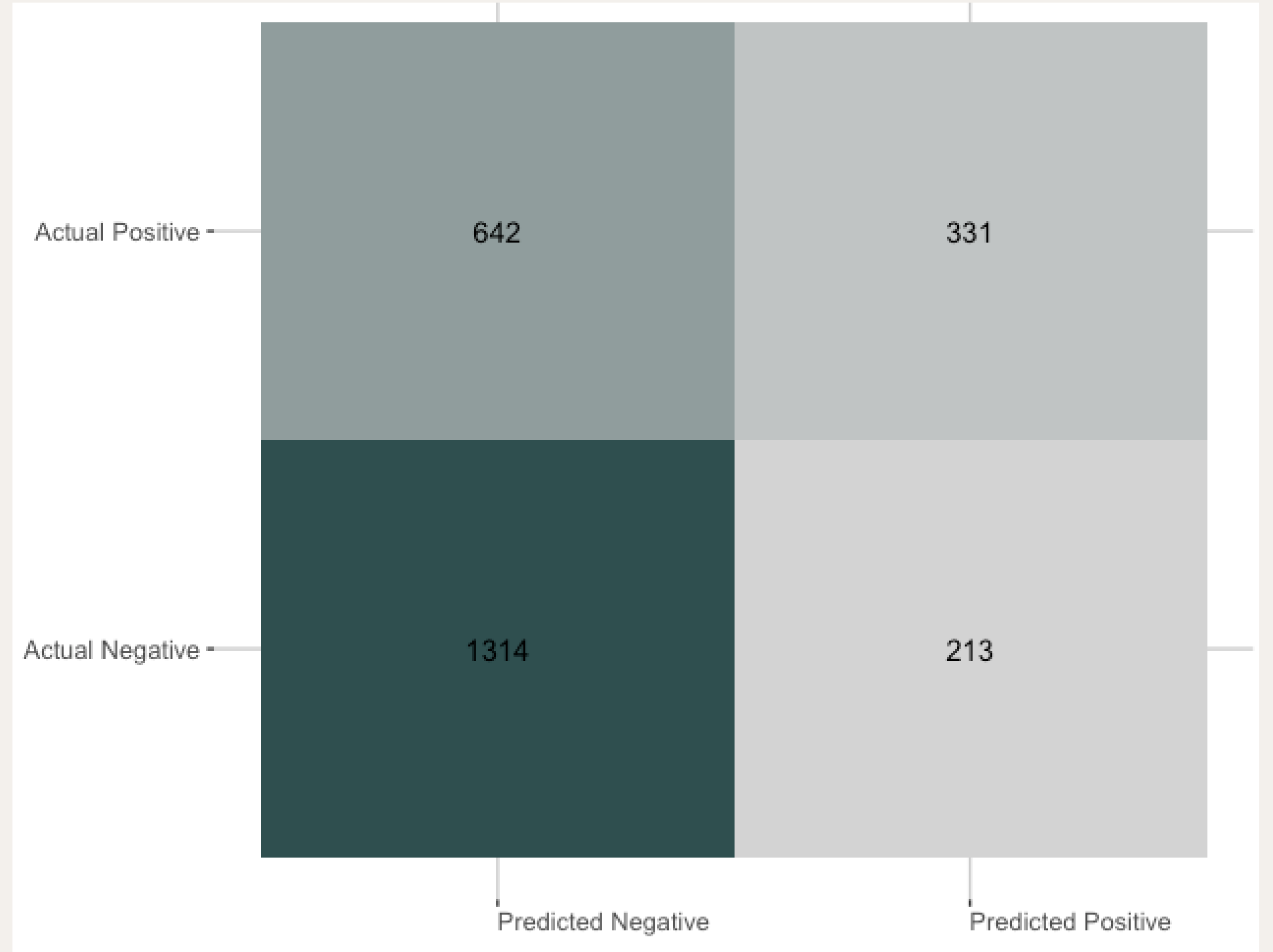


Logistic Regression

ROC CURVE

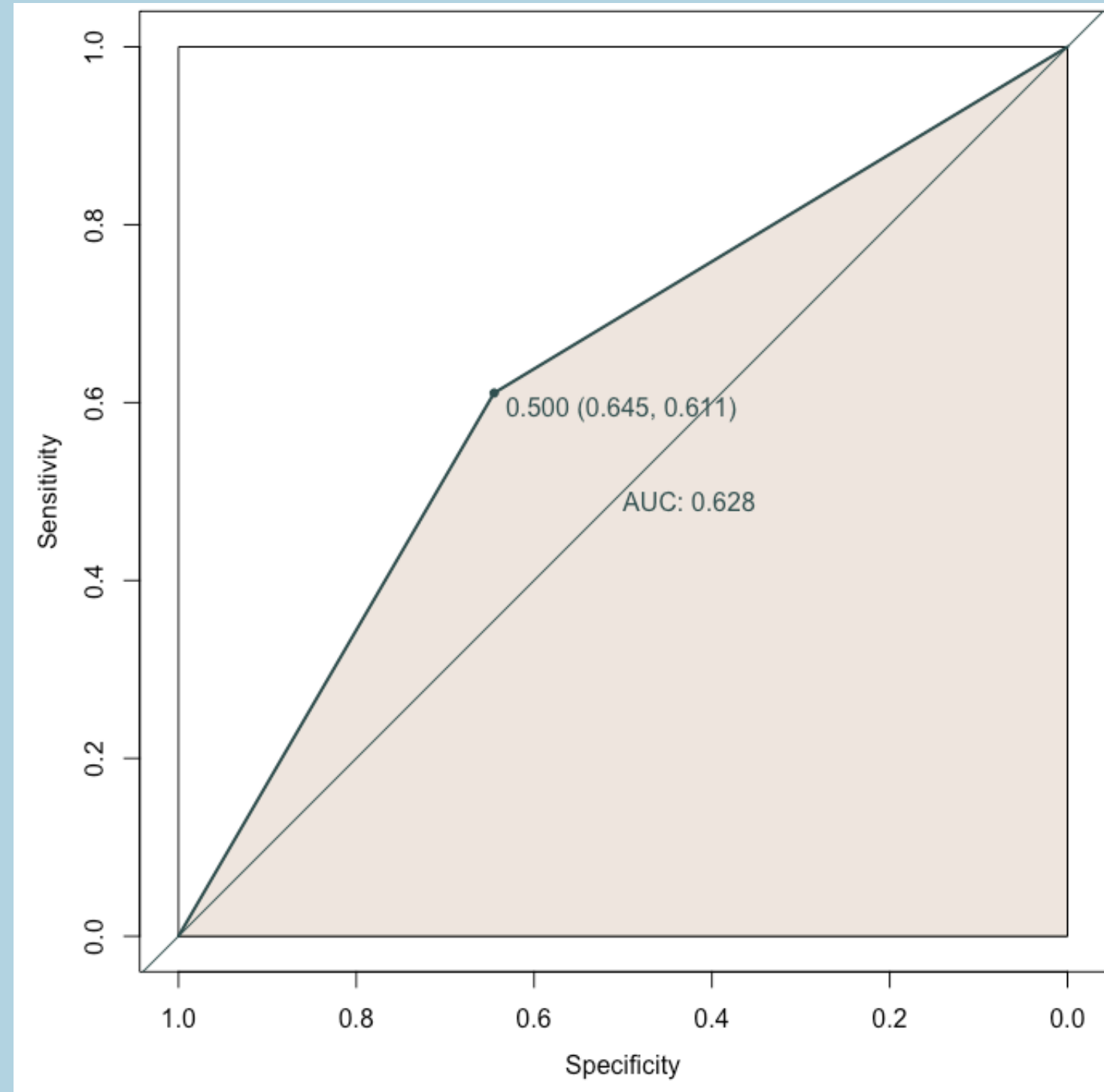


CONFUSION MATRIX

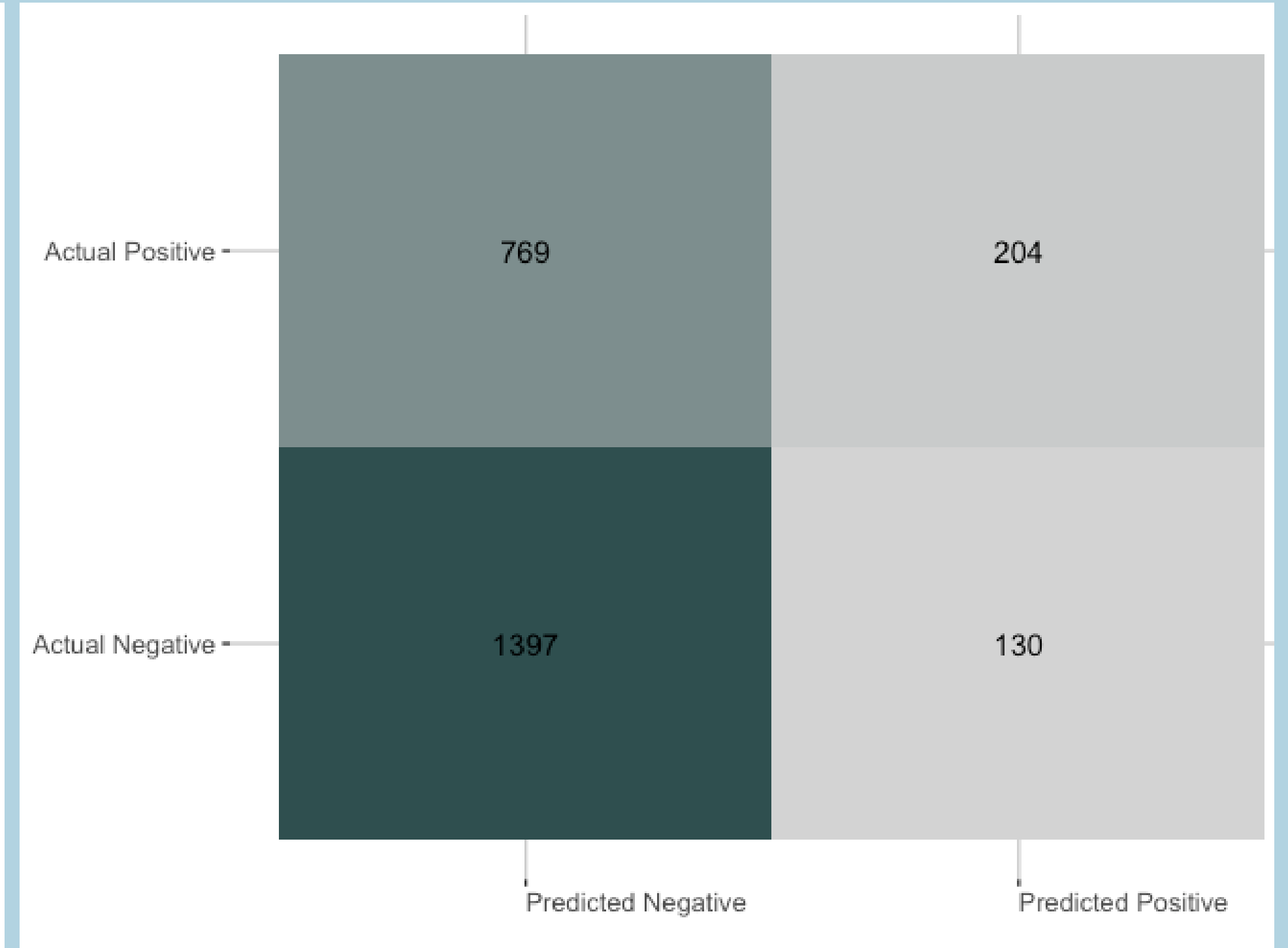


Decision Tree

ROC CURVE

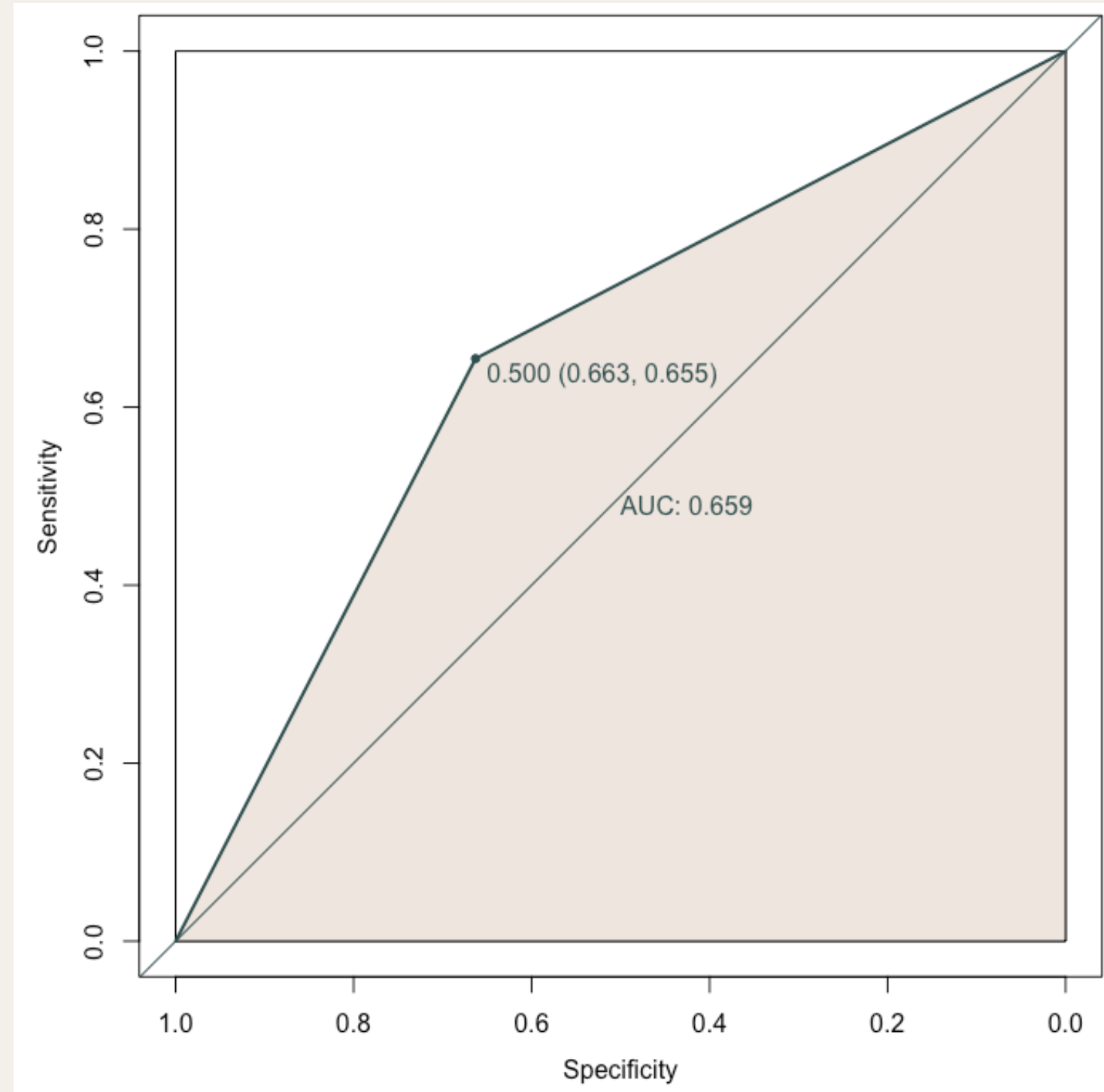


CONFUSION MATRIX

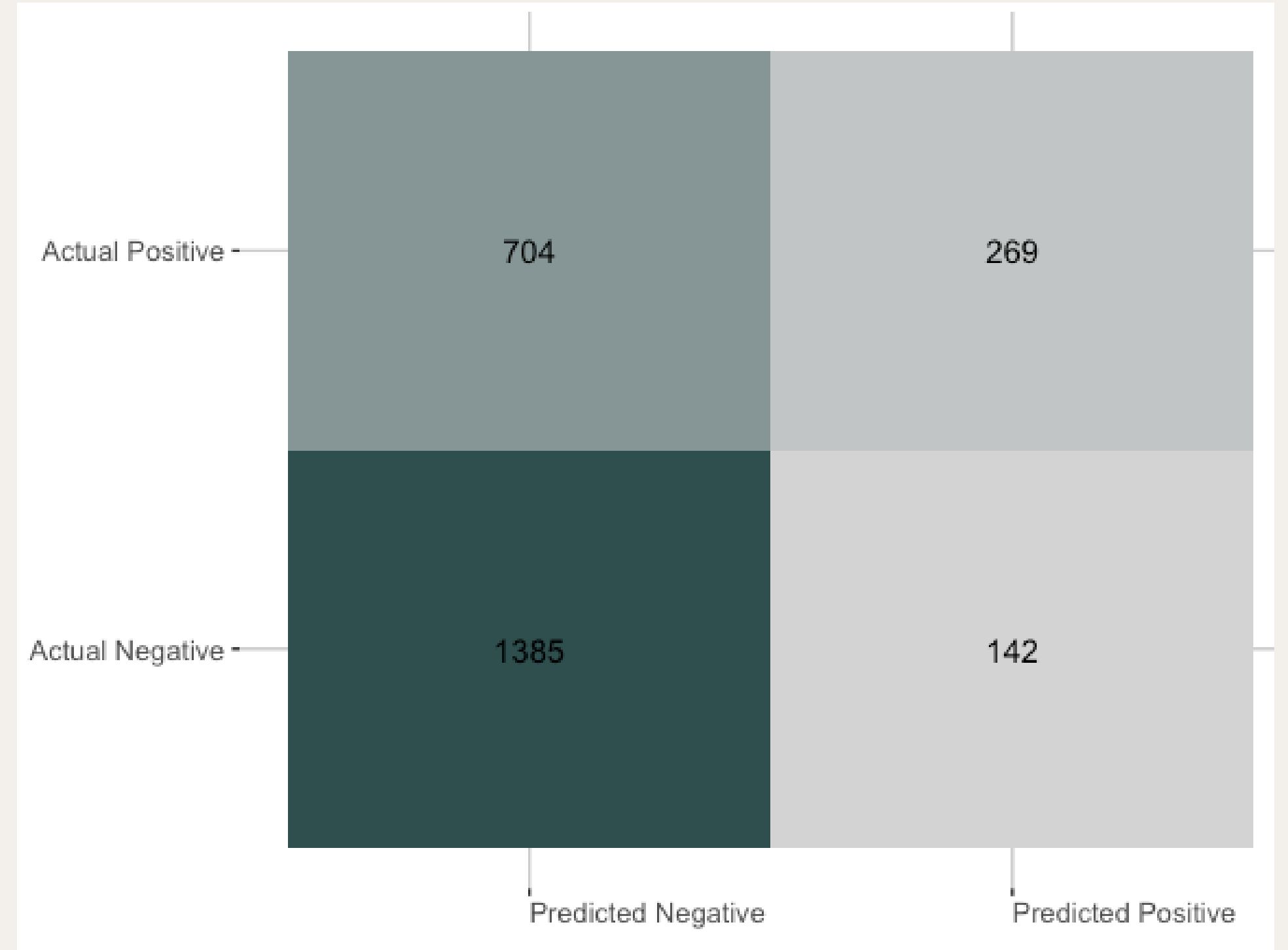


Random Forest

ROC CURVE



CONFUSION MATRIX



Logistic Regression

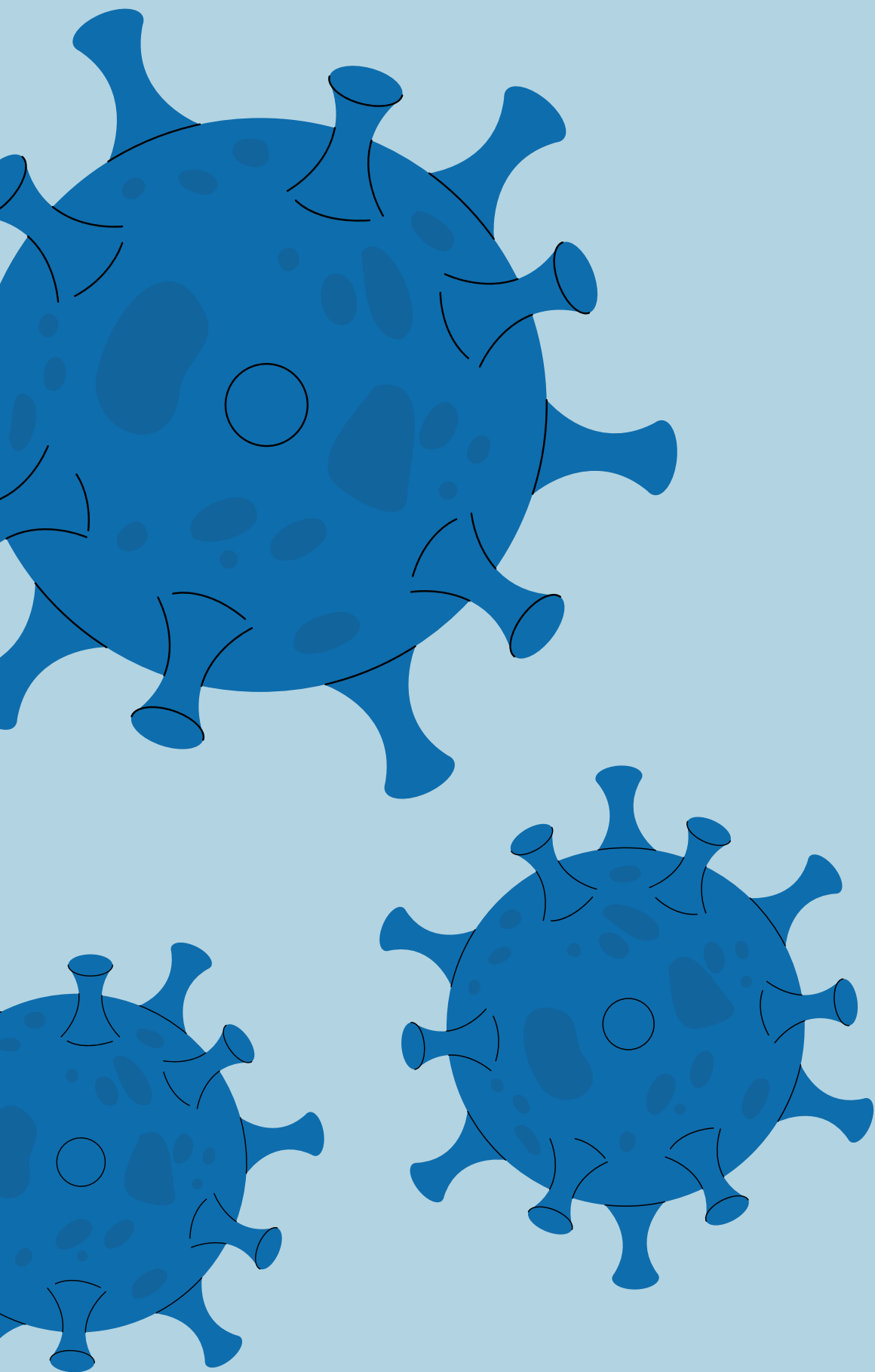
Accuracy	0.6580
AUC	0.6400
F1 Score	0.7545
MAE	0.3420
RMSE	0.5848

Decision Tree

Accuracy	0.6404
AUC	0.6280
F1 Score	0.7566
MAE	0.3596
RMSE	0.5997

Random Forest

Accuracy	0.6616
AUC	0.6590
F1 Score	0.7660
MAE	0.3384
RMSE	0.5817

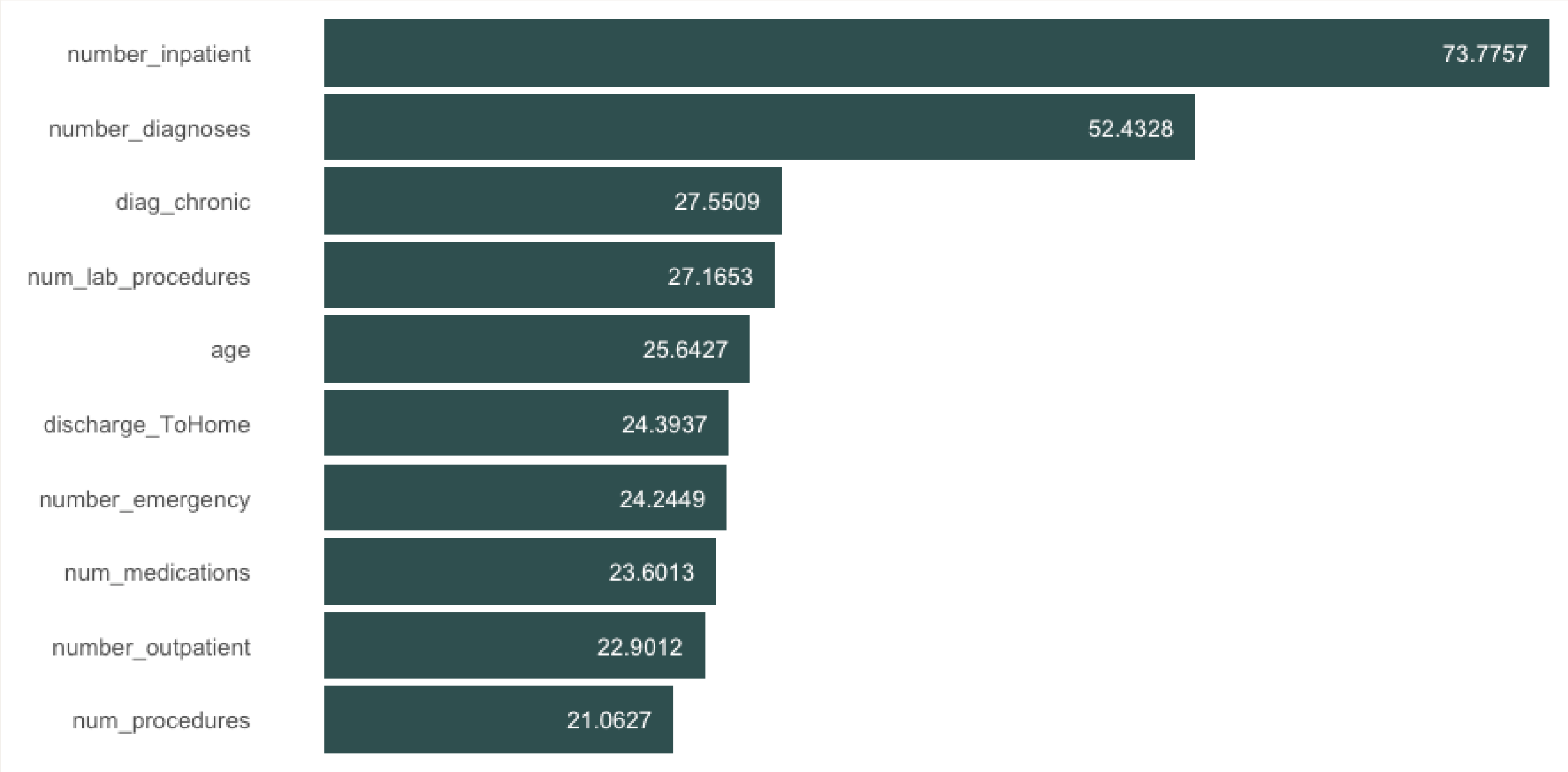


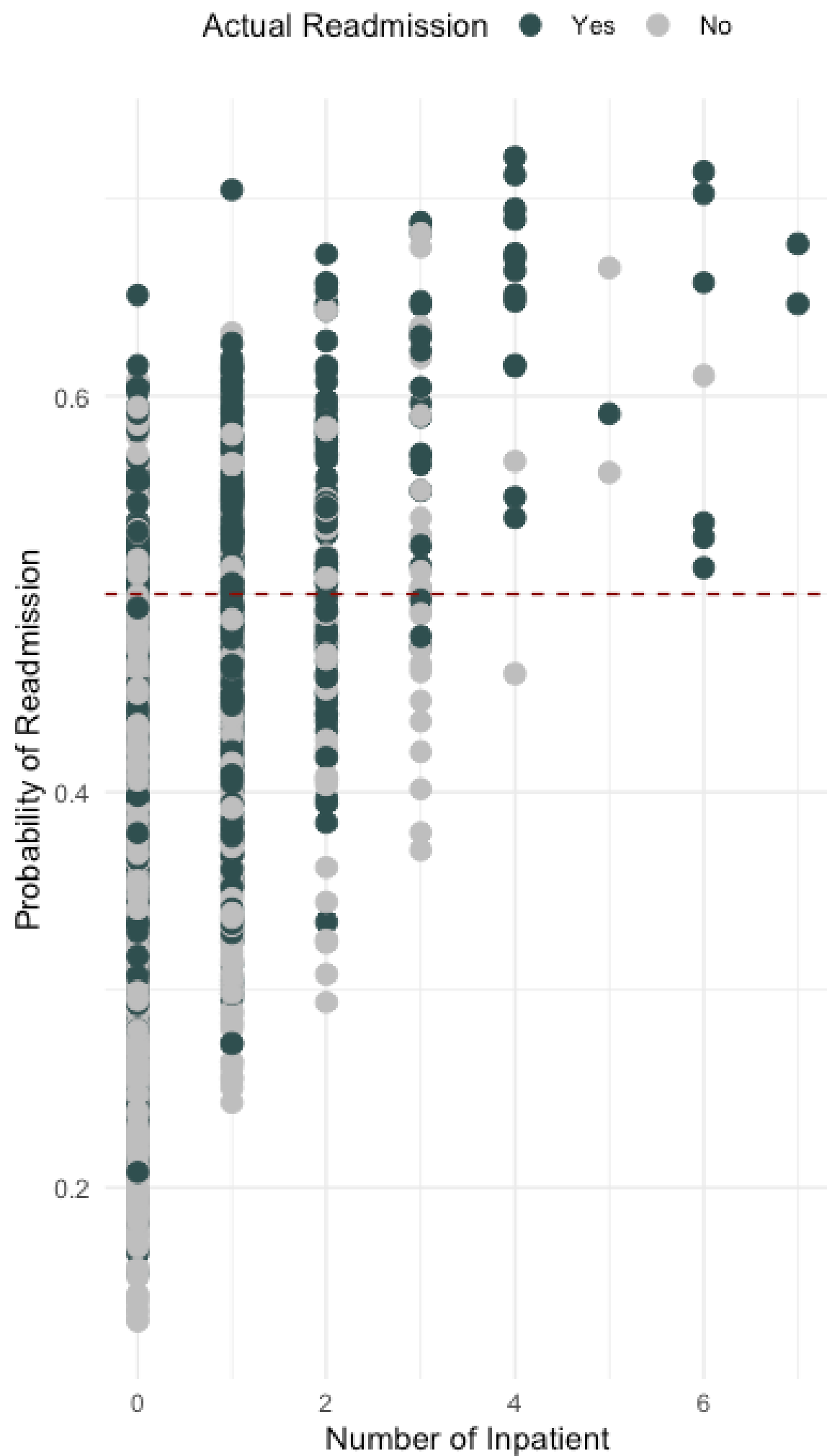
Analysis

What elements have the greatest impact on this probability?



Top 10 Feature Importances



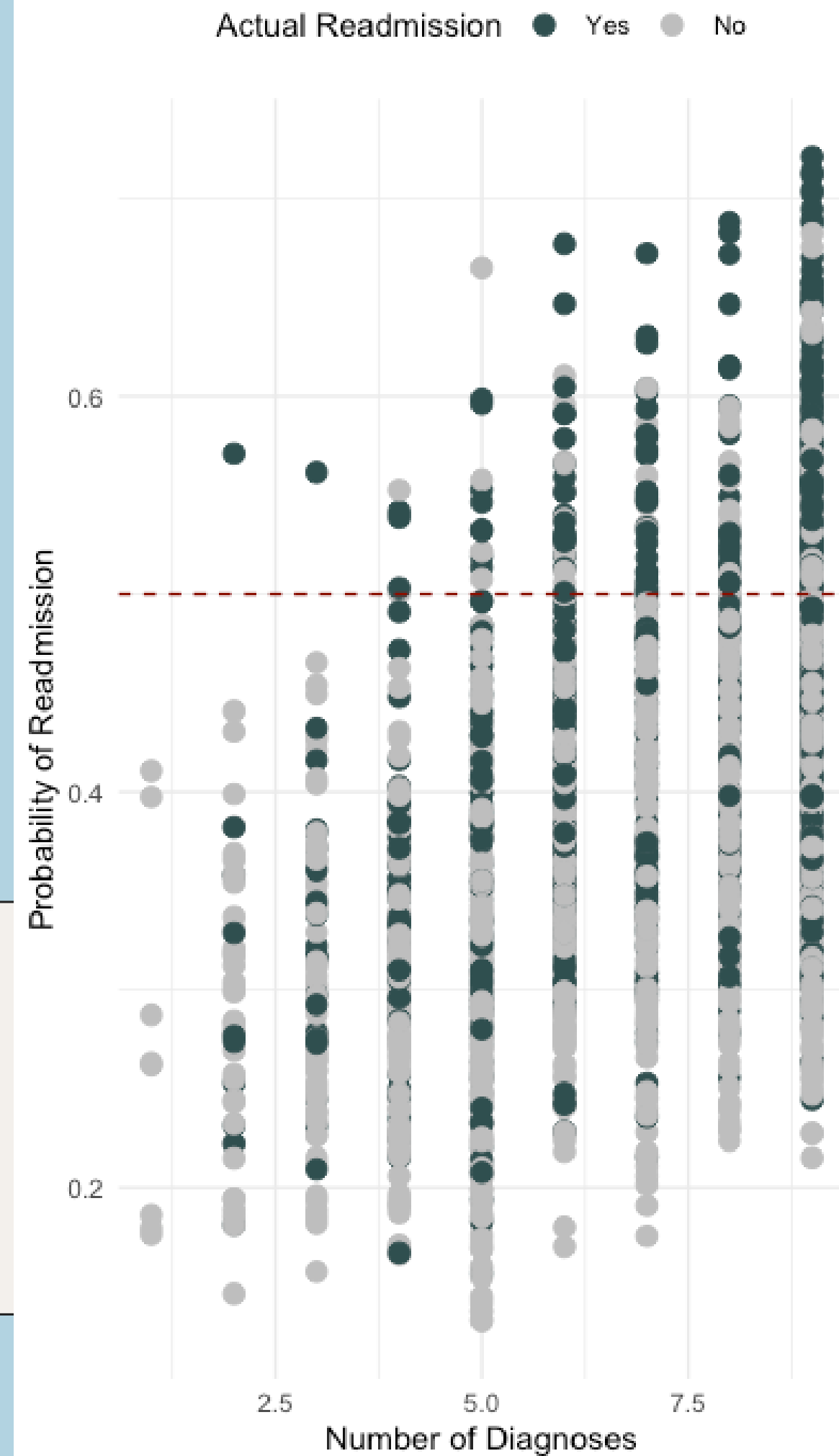


Inpatient

The higher the number of inpatient care given, the higher the probability of readmission. This may be due to the recurring nature of their condition.

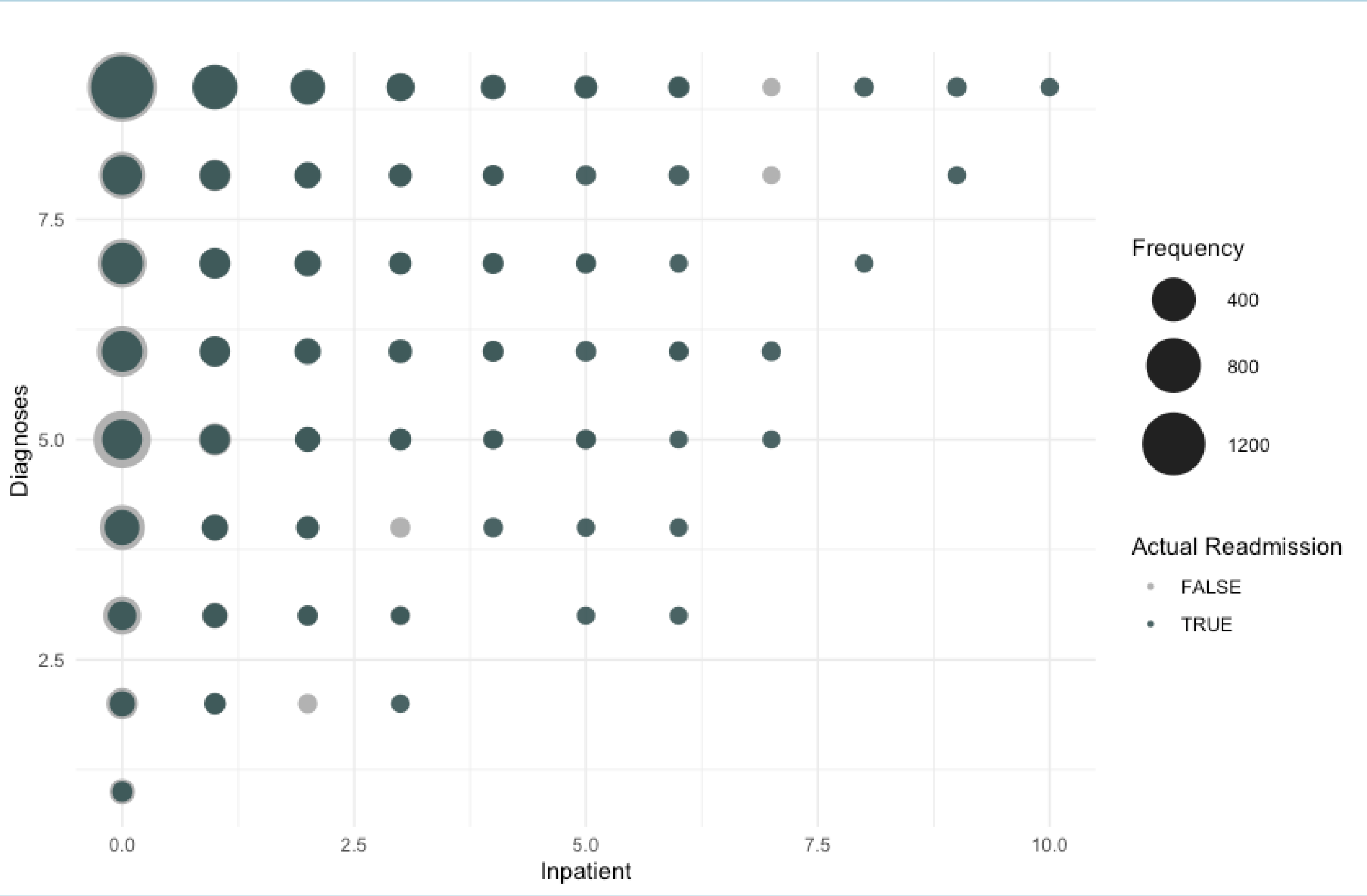
Diagnoses

The higher the number of diagnoses, the higher the probability of readmission. This could be an indicator of recurring conditions.



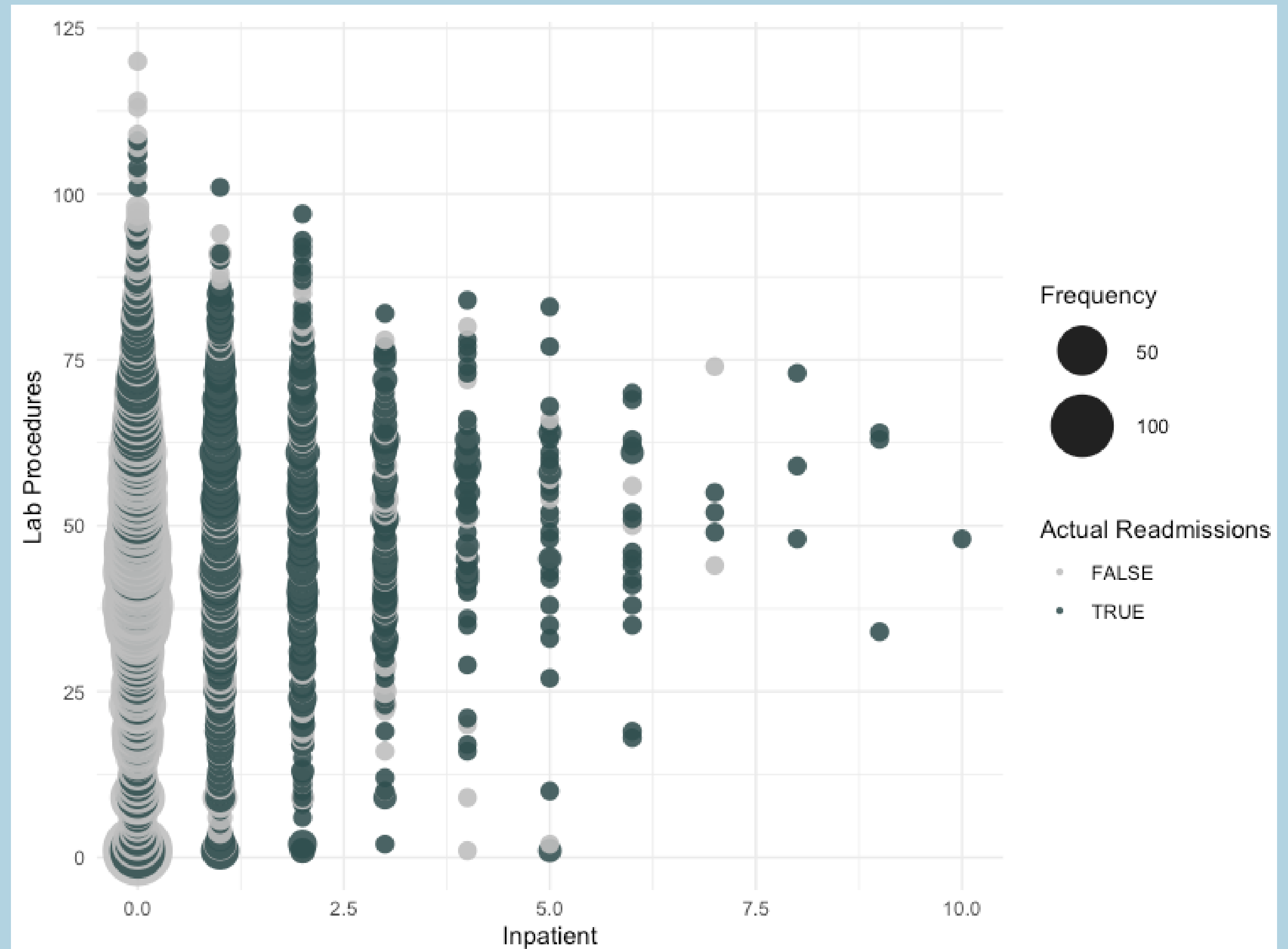
Inpatient vs Diagnoses

There may be patients who need to be admitted or be admitted for a longer time but are not getting the care necessary.



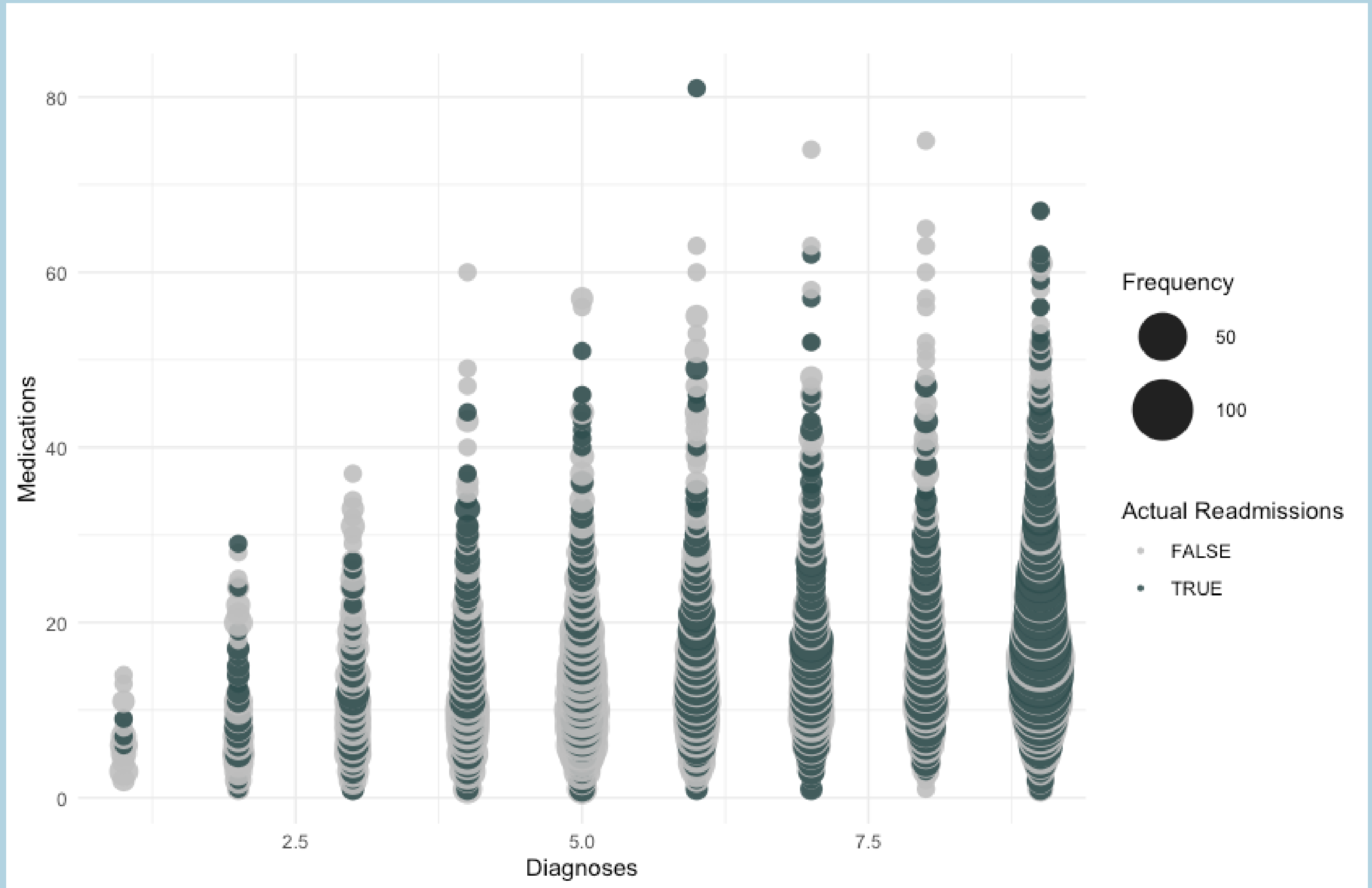
Lab Procedures vs Inpatient

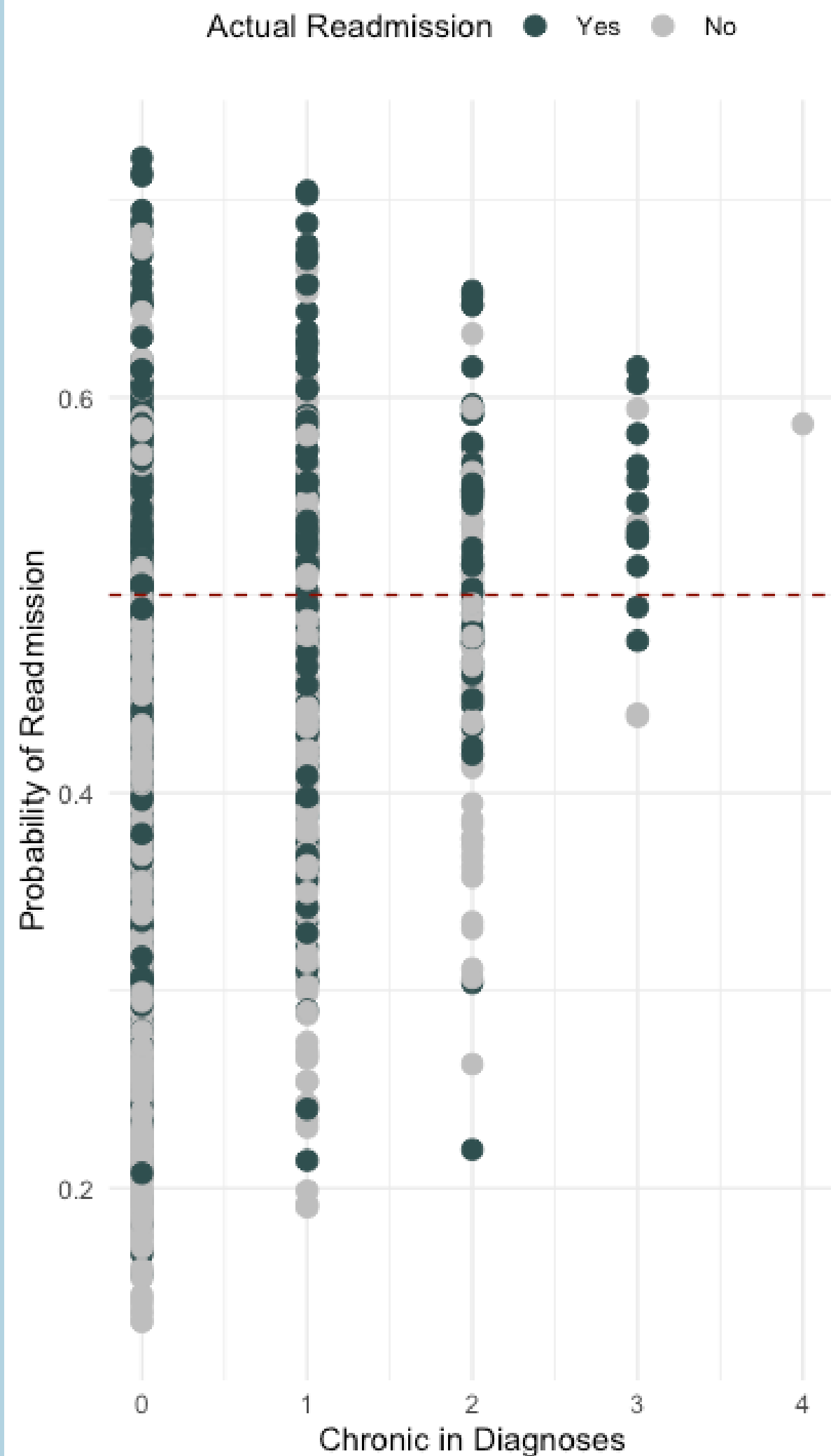
The range of lab procedures given to those with less inpatient care are wider than those with more inpatient care. Those who have a lot of lab procedures also generally do not need inpatient care.



Medication vs Diagnoses

There are a number of patients who have a lot of conditions but are taking very few medications. This could be indicative of untreated conditions.

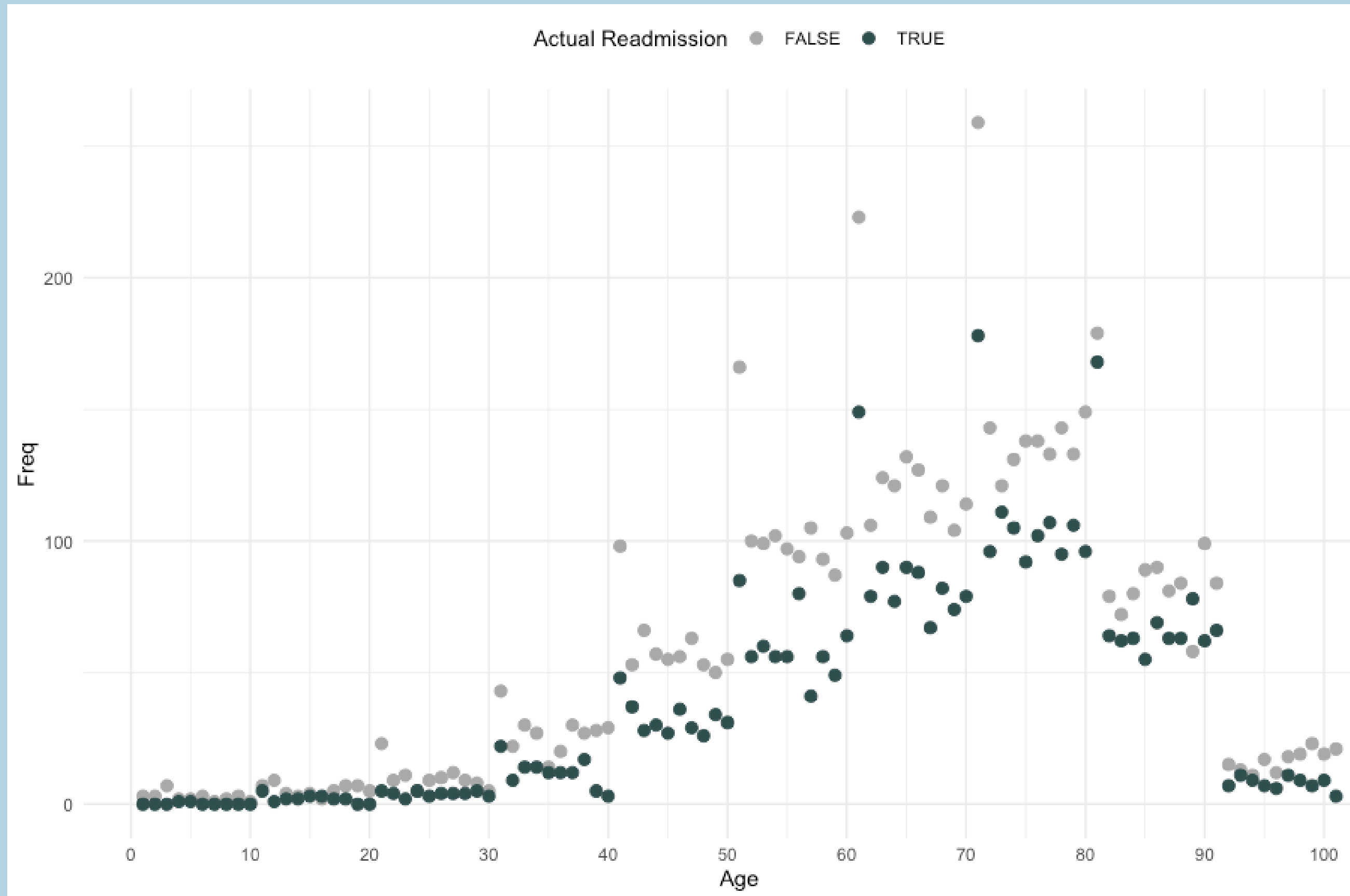




Chronic in Diagnoses

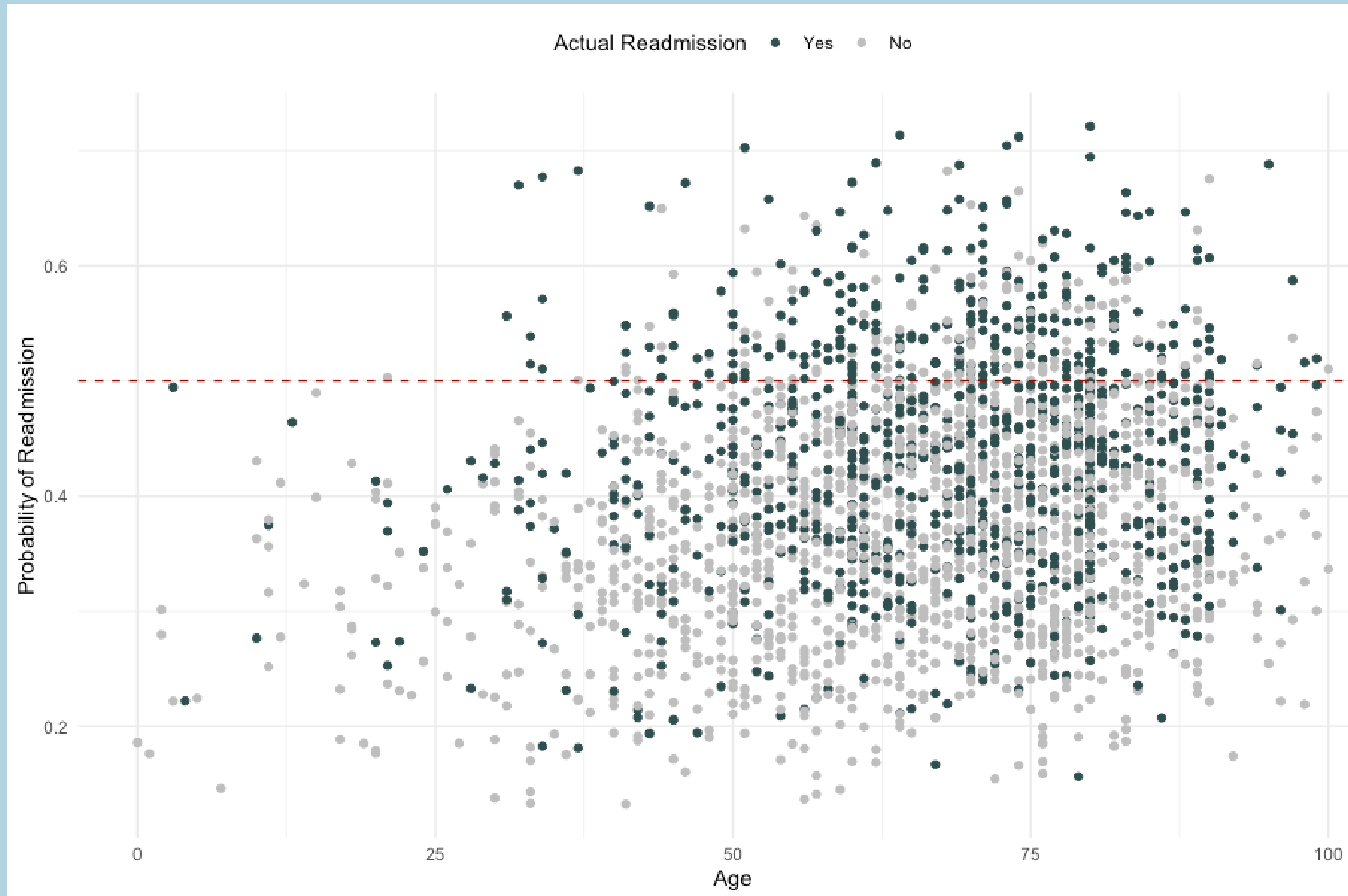
A chronic condition is a health condition or disease that is persistent or otherwise long-lasting in its effects or a disease that comes with time.

There is a higher probability of readmission if they were diagnosed with a chronic condition. This aligns with recurring conditions needing repeated inpatient care.



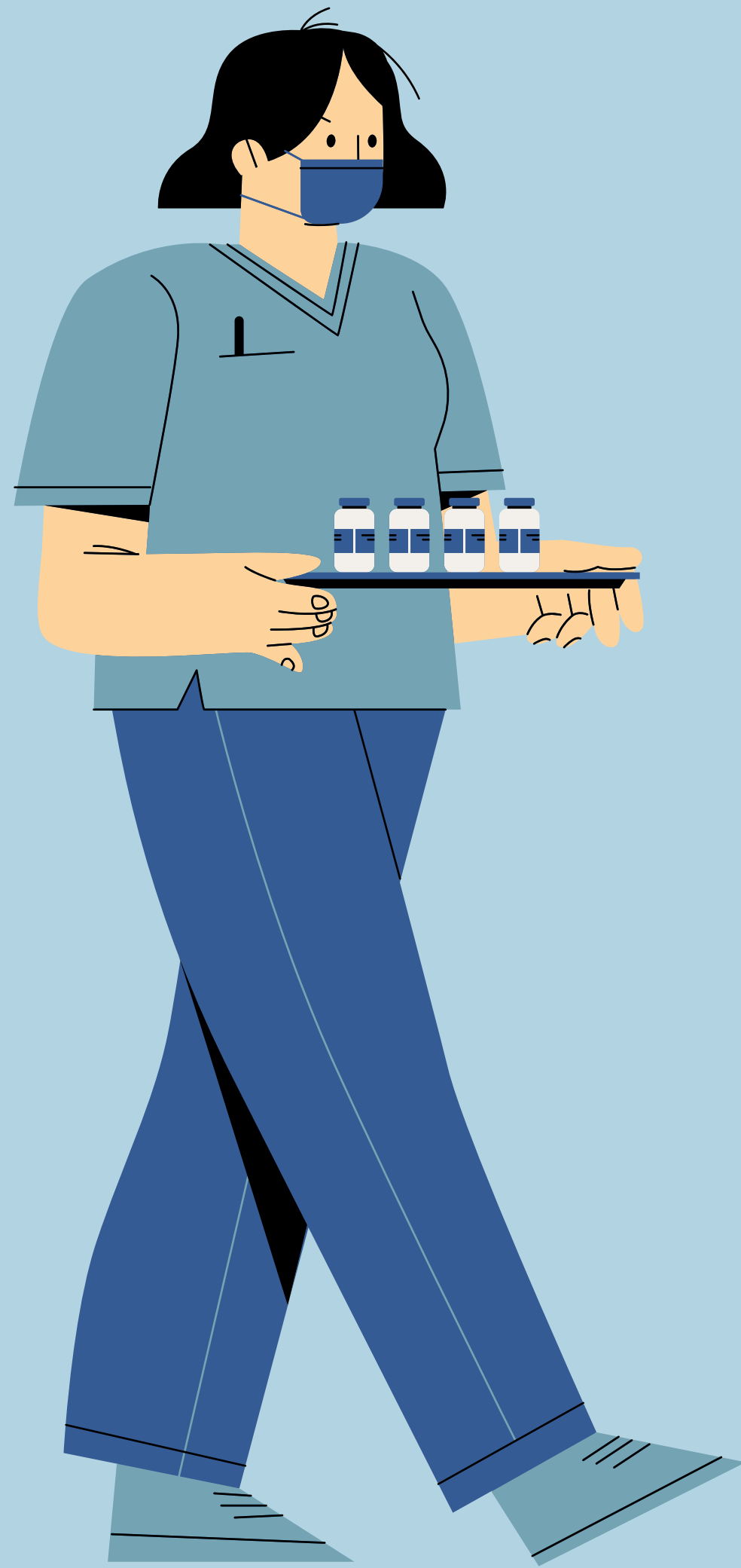
Age Distribution

There is a consistent interval of ages that are noticeably more frequent. The interval is around every 10 years, starting around 40 years old.



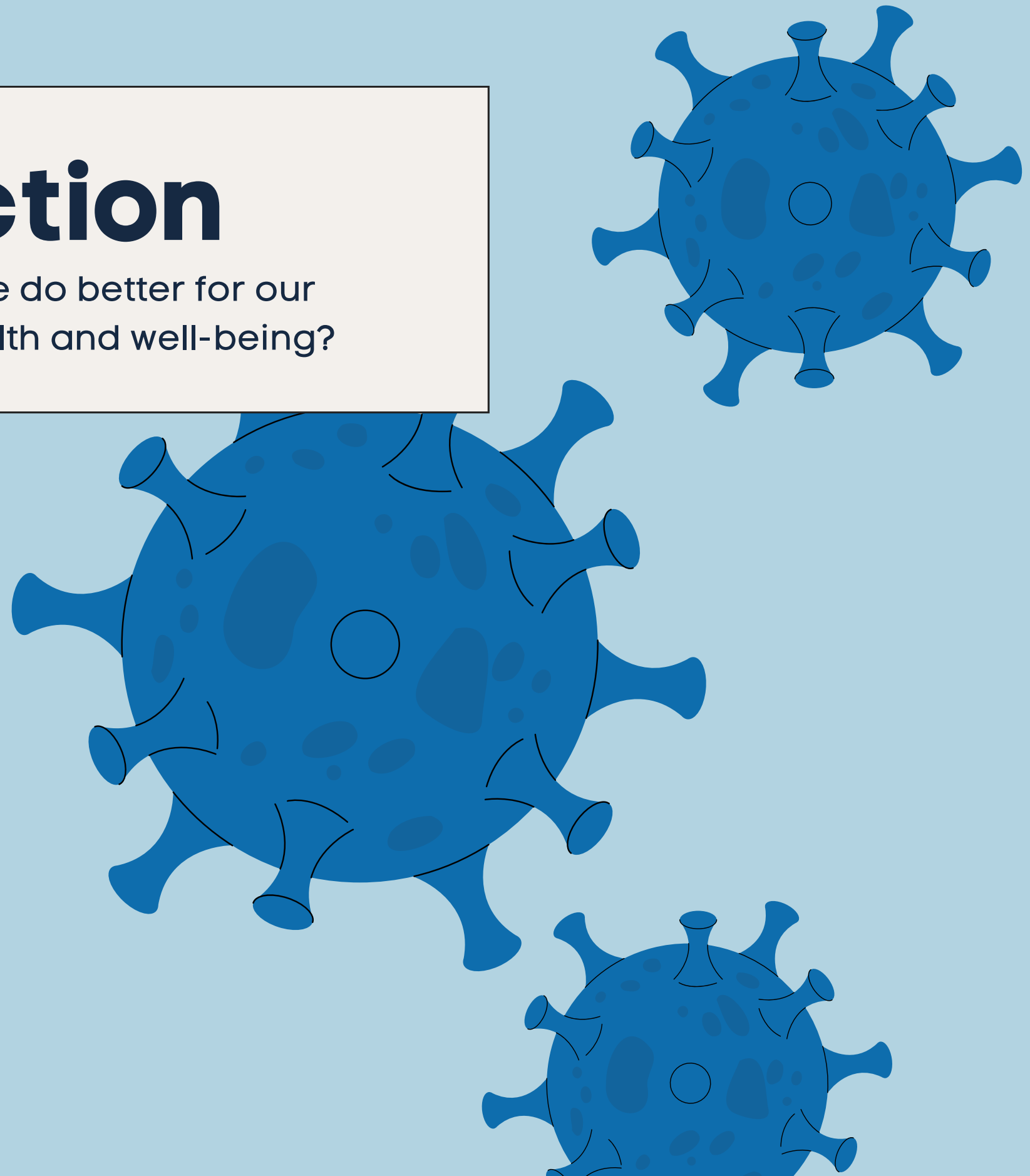
Age Probability

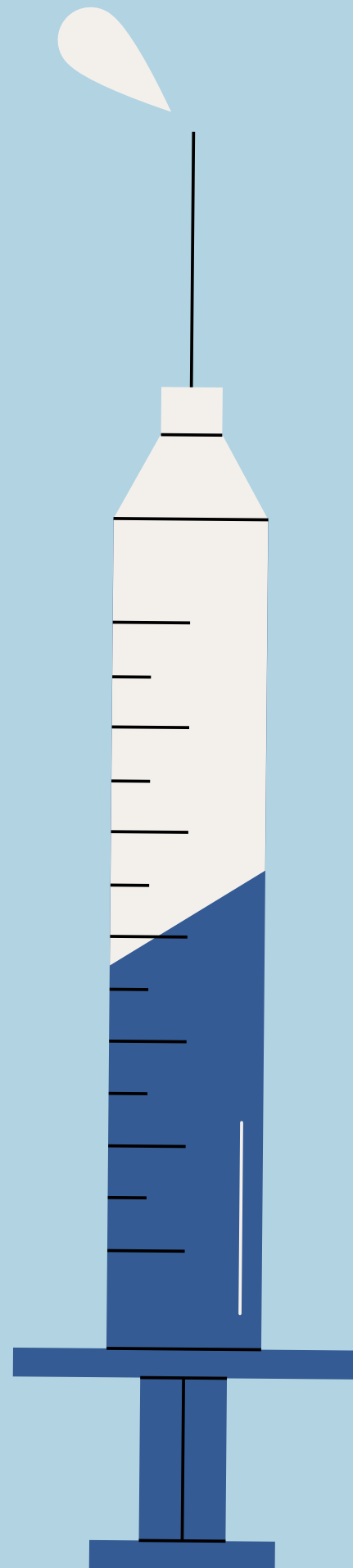
Bulk of patients are 50 to 90 years old. There is no clear relationship in terms of probability, however there seem to be more readmission around this age range as well.



Action

What can we do better for our patients health and well-being?





Recurring conditions

A combination of high inpatient treatment in the past, a high number of diagnoses, and a chronic diagnosis look to be indicative of a high chance of readmission.

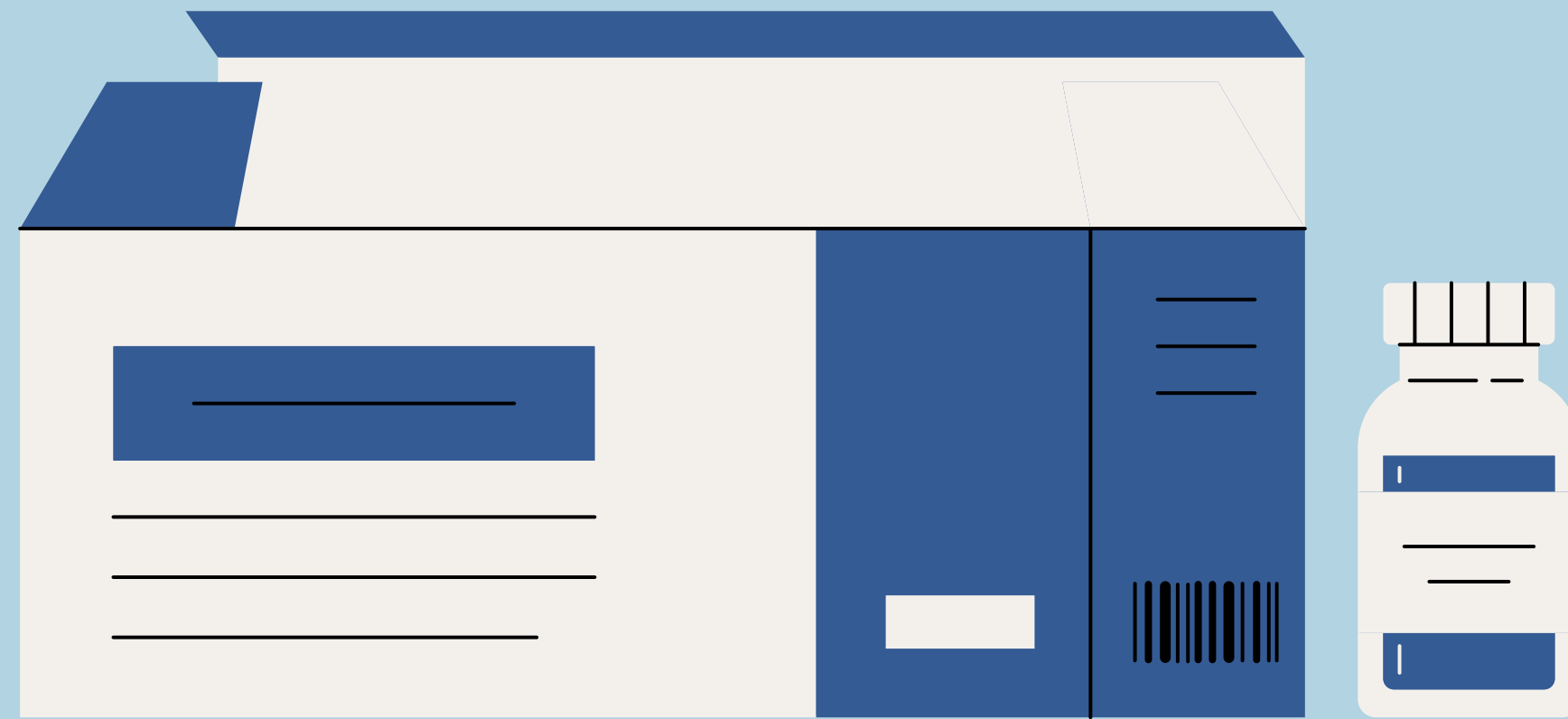
This is expected but maybe indicative that the initial admission of the patient did not provide for the level of quality of care actually needed. The patient may have needed more lab procedures or more medication prescribed, to reduce the chance of readmission.

Next Steps

Increase quality and quantity of care provided to patients fitting the profile of having recurring conditions or symptoms.

Monitor chronic diabetes symptoms more closely and improve service delivery systems for long term problems and self-management guidance.

Increase screening and monitoring as patients get older. Particularly upon they turn 30 and more frequently as they go above 50.



Thank you.

Audrey Arocha | MBAN FY | April 8, 2023
A2: EDA and Modeling | DAT-5323

