

# McNulty and Associates

Principal Predictors of Heart Disease

# The Data

Data from four hospital databases was combined to form the complete data set for this analysis. The data comprises 899 records and 76 fields, some of which are missing data.

# Feature Selection

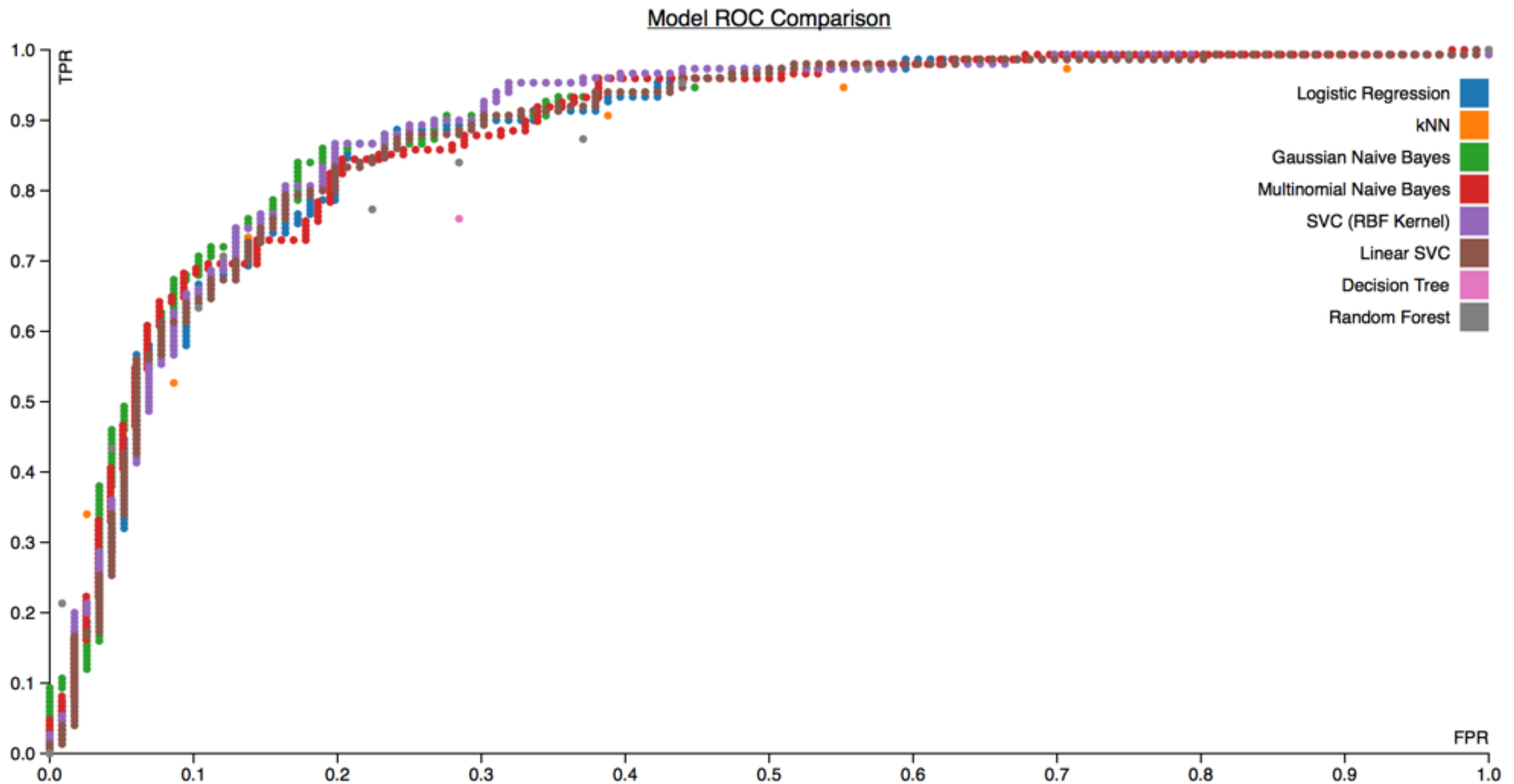
- Began with over 75 features
- Removed features with more than 100 null values
- Performed recursive feature selection on remaining features
  - Used a Logistic Regression Model
- Final feature count reduced to 14

# Models Explored

Using the 14 final features, several models were built to evaluate which is optimal

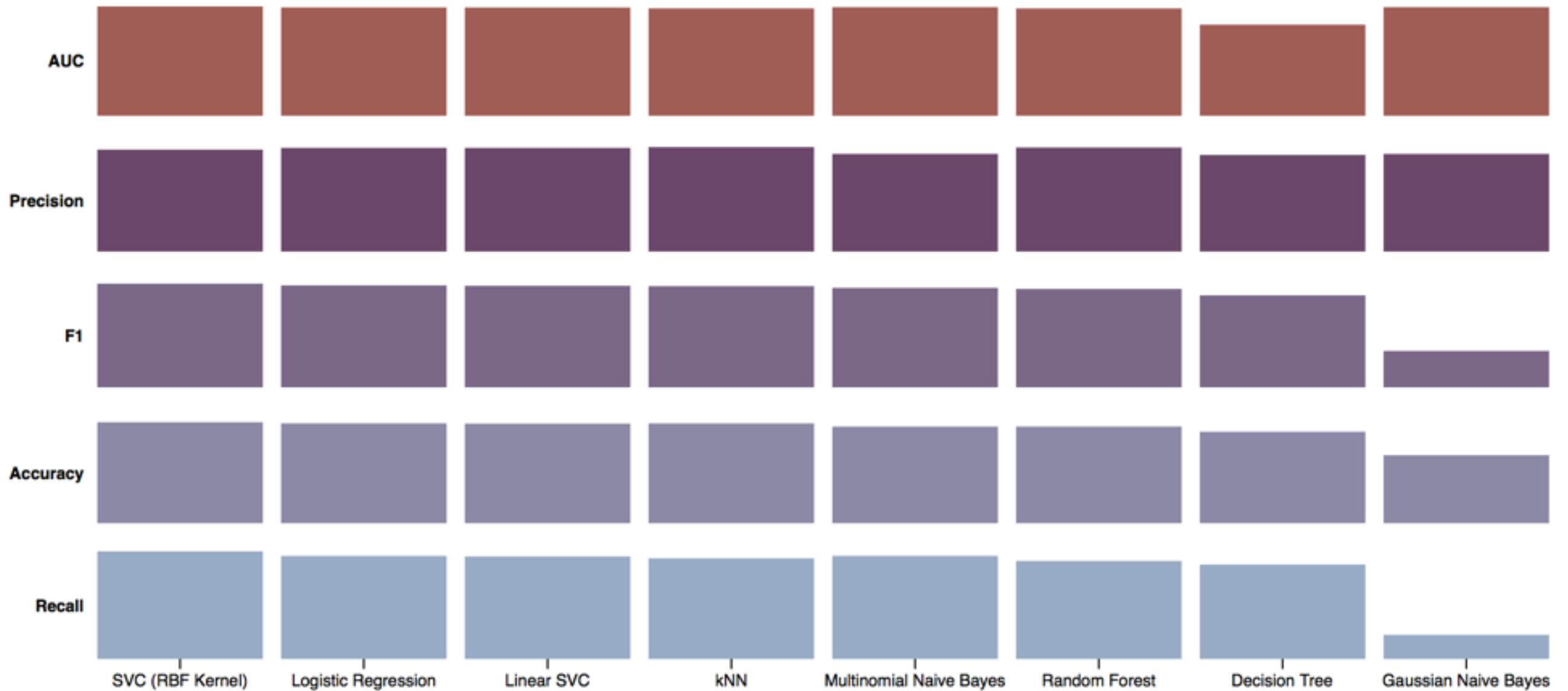
- Logistic Regression
- kNN
- SVC (RBF Kernel)
- SVC (Linear Kernel)
- Gaussian Naive Bayes
- Multinomial Naive Bayes
- Decision Tree
- Random Forest

# Models Evaluated



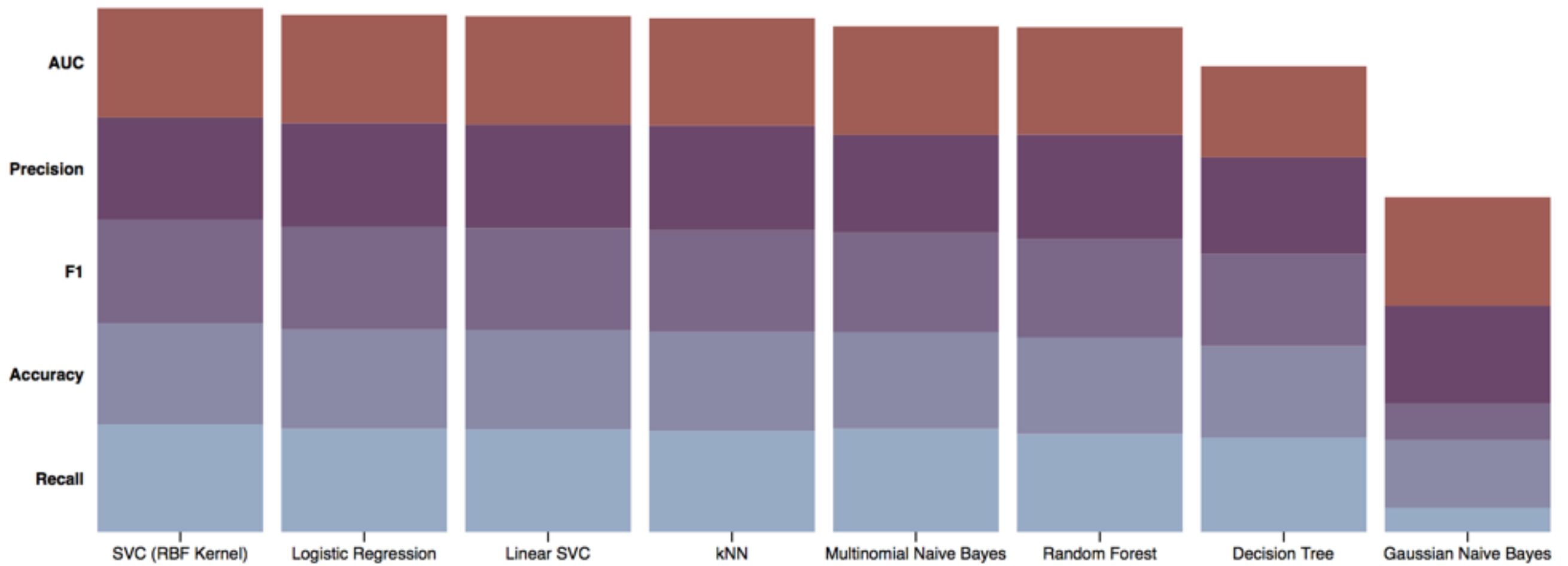
# Models Evaluated

### Model Score Comparison

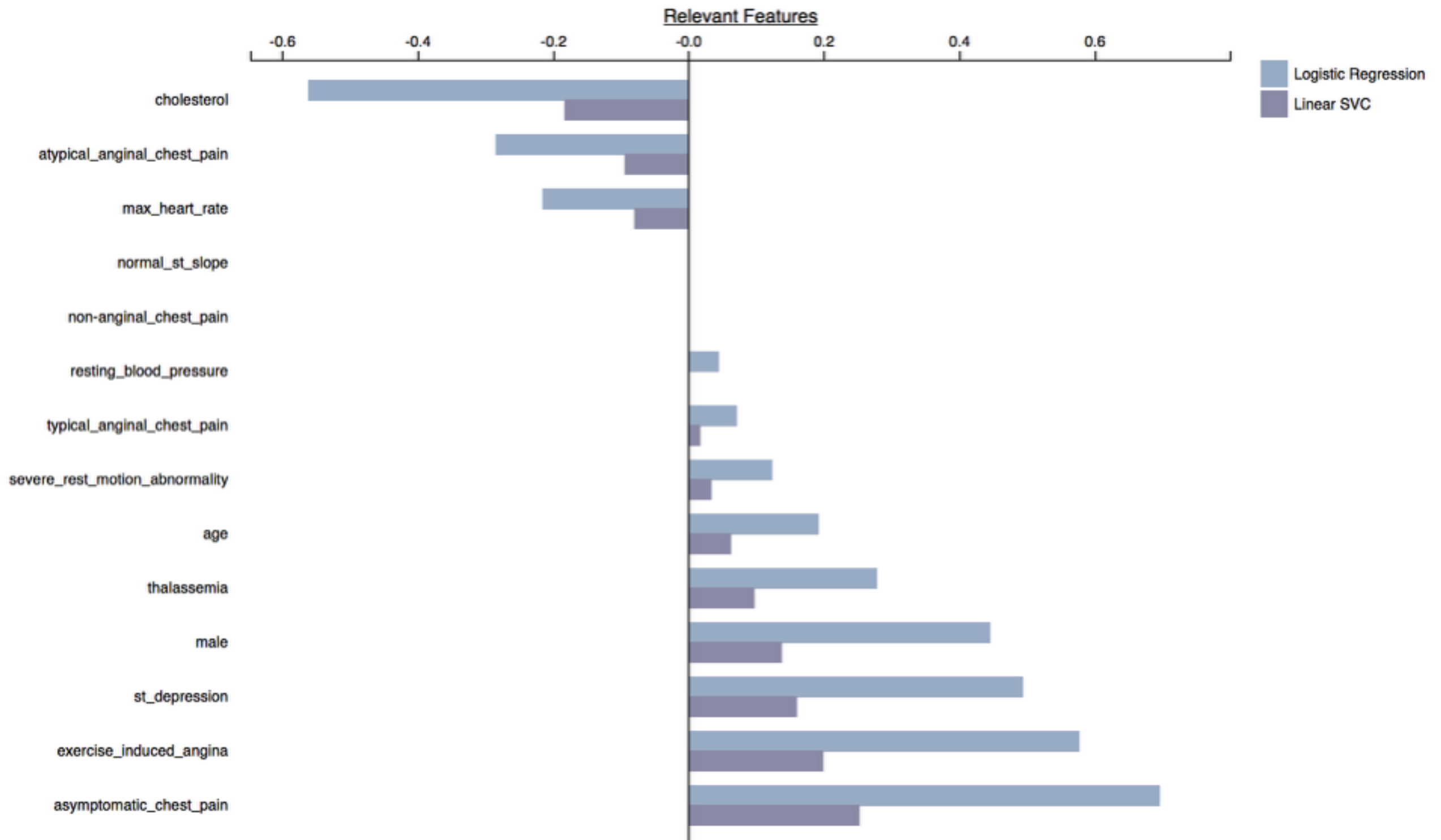


# Models Evaluated

## Model Score Comparison



# Feature Importance





# Takeaways

- It is possible to determine, with reasonable accuracy, whether or not someone has heart disease
- This can be done with only 14 features results, most of which are readily available or easy to obtain