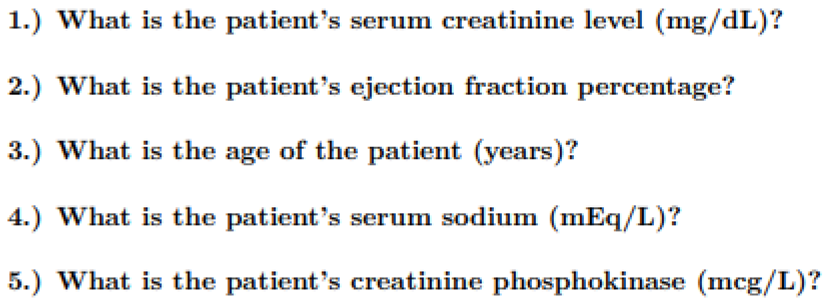
CART to Predict and Prevent Death Events

Brian Wisniewski

Content/Exposition:

The primary purposes of this report are to fit a classification tree model with the goal of predicting a death event based on other variables and to create a questionnaire, based on those findings, that doctors could use to assess a patient's risk. This was accomplished by first selecting variables that appeared to correlate with death events and then fitting the tree model using those variables. To determine the variables through correlation assessment, various plots (Figure 1 & Figure 2) were created to gain a preliminary understanding of their correlation with death events. Based on my interpretation of those plots and the lack of correlation associated with them, I chose to exclude the platelets and time variables. Once the variables were chosen, I fitted and plotted the classification tree model (Figure 3). Additionally, I created plots for models without some additional variables (Figure 4 & 5) to assess if they performed better than Figure 3. The variables that were removed are indicated in the titles for Figure 4 & 5. After careful consideration, I determined that my original model (Figure 3) provided the most informative yet easily interpretable results, so I used it as the basis for my questionnaire. It is worth noting that when I removed age (Figure 4), more variables were included in the tree model. However, this did not contribute any meaningful information in my opinion, as most of the logic led to a 0-death event. By observing the tree model plot (Figure 3), I deduced survey questions based on the conditional logic and overall structure of the branches. Consequently, I structured the questionnaire as follows, guided by the logic presented in Figure 3, below:

Questionnaire:



1.) What is the patient's serum creatinine level (mg/dL)?

2.) What is the patient's ejection fraction percentage?

3.) What is the age of the patient (years)?

4.) What is the patient's serum sodium (mEq/L)?

5.) What is the patient's creatinine phosphokinase (mcg/L)?

Plots to Assess Correlation of Categorical Factors to Death Events

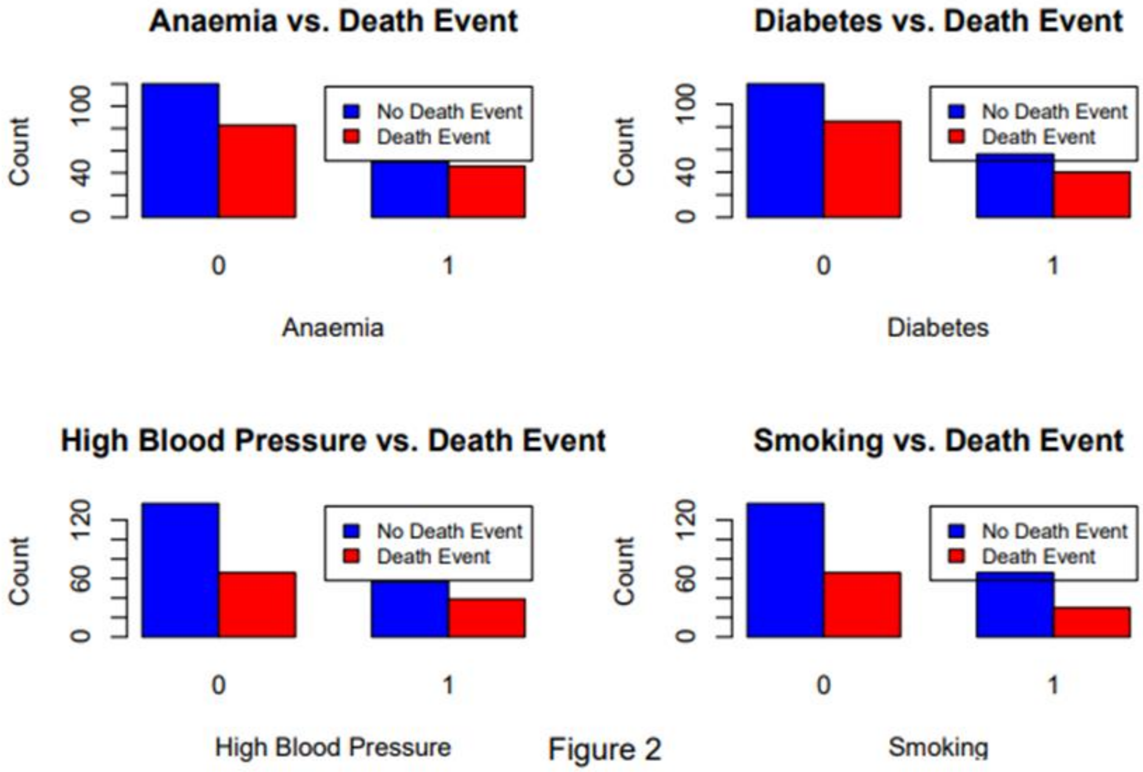


Figure 2

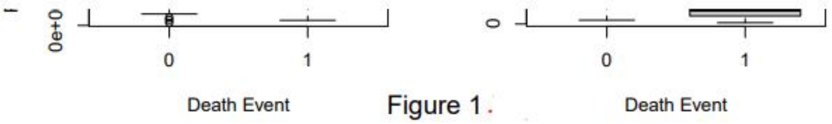


Figure 1.

Plots to Assess Correlation:

Plots to Assess Correlation of Continuous Variables to Death Event

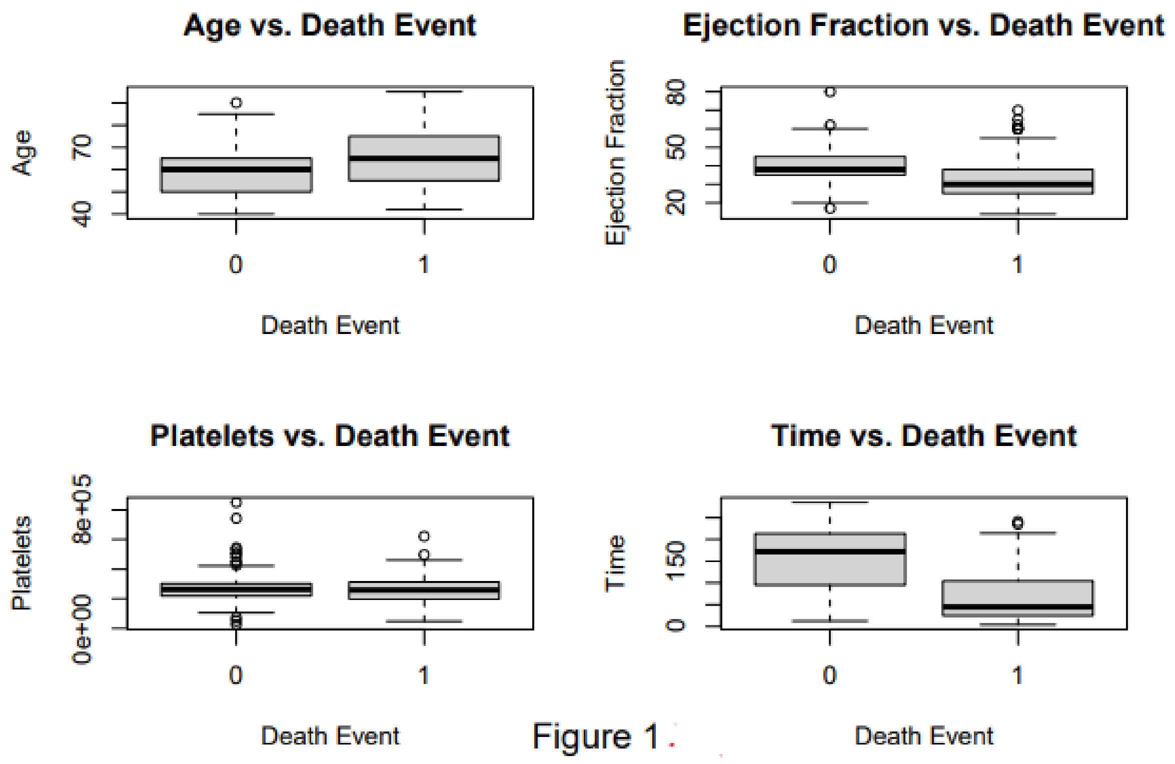
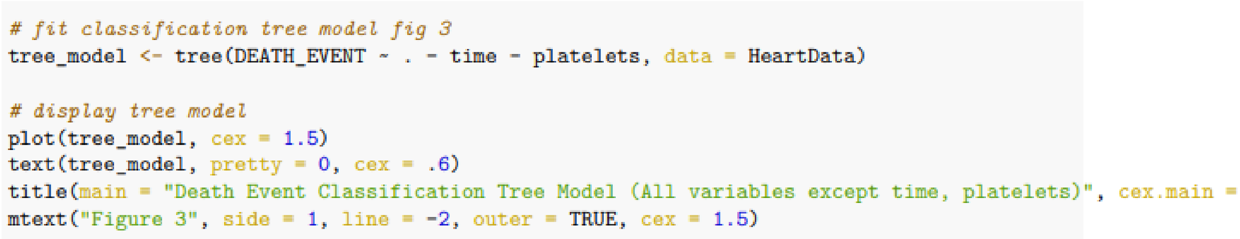


Figure 1 .

Classification Tree Models

Classification Tree Model (Figure 3)



```
# fit classification tree model fig 3
tree_model <- tree(DEATH_EVENT ~ . - time - platelets, data = HeartData)

# display tree model
plot(tree_model, cex = 1.5)
text(tree_model, pretty = 0, cex = .6)
title(main = "Death Event Classification Tree Model (All variables except time, platelets)", cex.main =
mtext("Figure 3", side = 1, line = -2, outer = TRUE, cex = 1.5)
```

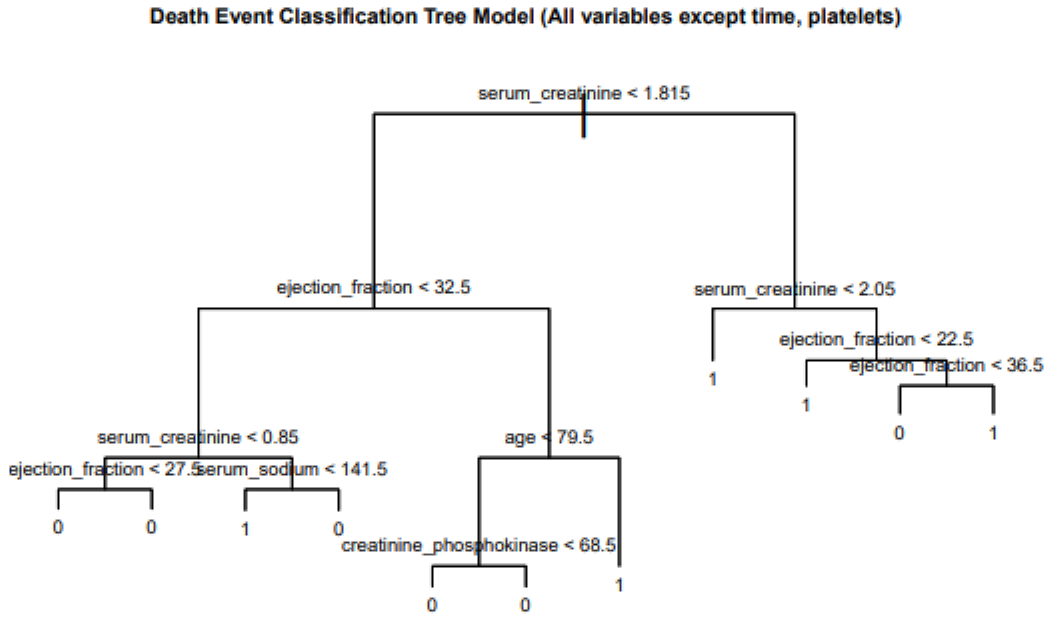


Figure 3

Classification Tree Model (Figure 4)

```
# fit classification tree model fig 4
tree_model2 <- tree(DEATH_EVENT ~ . - time - platelets - age, data = HeartData)
# display tree model
plot(tree_model2, cex=1.5)
text(tree_model2, pretty = 0, cex =.6)
title(main = "Death Event Classification Tree Model (All variables except time, platelets, age)",cex.ma
mtext("Figure 4", side = 1, line = -2, outer = TRUE, cex = 1.5)
```

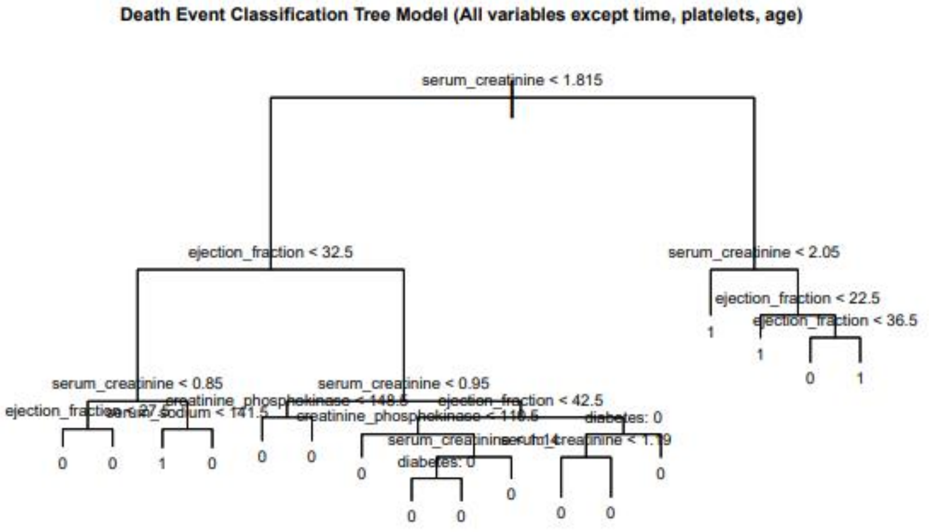
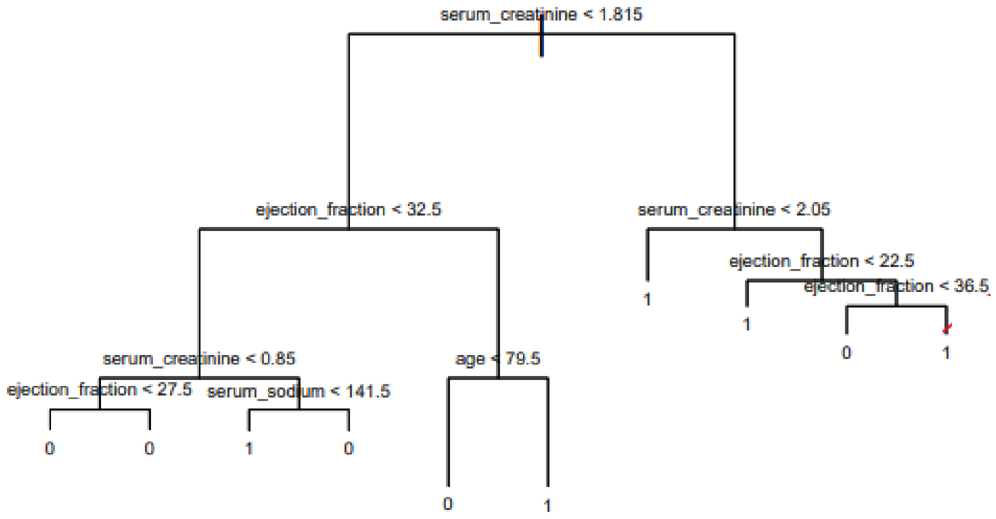

Death Event Classification Tree Model (All variables except time, platelets, age)

Figure 4

Classification Tree Model (Figure 5)

```
#fit classification tree model fig 5
tree_model3 <- tree(DEATH_EVENT ~ . - time - platelets - creatinine_phosphokinase, data = HeartData)
# display the tree model
plot(tree_model3, cex=1.5)
text(tree_model3, pretty = 0,cex=.6)
title(main = "Death Event Classification Tree Model (All variables except time, platelets, creatinine p
mtext("Figure 5", side = 1, line = -2, outer = TRUE, cex = 1.5)
```

Death Event Classification Tree Model (All variables except time, platelets, creatinine phosphokinase)


serum_creatinine < 1.815
ejection_fraction < 32.5
serum_creatinine < 2.05
ejection_fraction < 22.5
ejection_fraction < 36.5
1
1
0
1
serum_creatinine < 0.85
age < 79.5
ejection_fraction < 27.5
serum_sodium < 141.5
0
0
1
0
0
1