A Bayesian Analysis of Fatality in Heart Failure Patients

Joseph S. Kwaik

Synopsis

The aim of this project is to perform a Bayesian probability analysis of fatality in heart failure patients. We will analyze clinical data collected from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad. The data contains the medical records of 299 heart failure patients, all of which had left ventricular systolic dysfunction and previous heart failures. According to the stages of heart failure set by the New York Heart Association, all patients in the dataset are classified as either Class III or Class IV of heart failure. Our analysis will involve the usage of Bayesian statistical methods to estimate probabilities of fatality based on the medical characteristics of the patients. The results of the analysis will provide valuable insights into the factors that may affect the likelihood of fatality in heart failure patients and can be used to guide treatment decisions and improve patient outcomes.

Data Description

The data for this project comes in the form of a .csv, and can easily be downloaded on Kaggle, here. The dataset was originally released in July of 2017 by researchers in statistics who performed a survival analysis of the patients, which can be found here. As aforementioned, the dataset contains the medical records of 299 heart failure patients classified as Class III or Class IV by the New York Heart Association. They consist of 194 men and 105 men, who are all above the age of 40 years old. The data contains the information on 11 clinical features for every patient, including age, gender, anemia, creatine phosphokinase, diabetes, ejection fraction, blood pressure, platelets, serum creatinine, serum sodium, and smoking status.

To begin, I installed and loaded some packages into the R environment that would be useful in the analysis.

```
library(RColorBrewer)
library(ggplot2)
library(readr)
library(caTools)
library(rmarkdown)
library(xaringan)
library(knitr)
library(kableExtra)
```

The data was then imported into R. A dataframe called Heart was created based on the .csv, which is the main dataframe that will be used throughout our analysis.

```
heart_failure_clinical_records_dataset <- read.csv("heart_failure_clinical_records_dataset.csv")
Heart <- heart_failure_clinical_records_dataset
View(Heart)
```

Below is a preview of our data, before any data manipulation and cleaning. The column we are most interested in computing probabilities for is column "DEATH_EVENT". For the columns for the death event, anaemia, diabetes, high blood pressure, and smoking, the values are in binary. Therefore, a 0 value means that they do not have the specified condition or lifestyle habit, and 1 means that they do. For sex, a 0 is a woman, and a 1 is a male.

age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium
75	0	582	0	20	1	265000	1.9	130
55	0	7861	0	38	0	263358	1.1	136
65	0	146	0	20	0	162000	1.3	129
50	1	111	0	20	0	210000	1.9	137
65	1	160	1	20	0	327000	2.7	116
90	1	47	0	40	1	204000	2.1	132
75	1	246	0	15	0	127000	1.2	137
60	1	315	1	60	0	454000	1.1	131
65	0	157	0	65	0	263358	1.5	138
80	1	123	0	35	1	388000	9.4	133

age	anemia	cpkEnzyme	diabetes	ejection	hypertension	platelets	serum_creatinine	serum_sodium	sex	smoking	time	fata
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	1
90	1	47	0	40	1	204000	2.1	132	1	1	8	1
75	1	246	0	15	0	127000	1.2	137	1	0	10	1
60	1	315	1	60	0	454000	1.1	131	1	1	10	1
65	0	157	0	65	0	263358	1.5	138	0	0	10	1
80	1	123	0	35	1	388000	9.4	133	1	1	10	1

Data Manipulation

In order to make our data easier to use, I renamed the columns.

```
colnames(Heart)[colnames(Heart) == "creatinine_phosphokinase"] = "cpkEnzyme"  # Creatine Phosphokinase
colnames(Heart)[colnames(Heart) == "ejection_fraction"] = "ejection"  # Ejection Fraction - % blood lea
colnames(Heart)[colnames(Heart) == "high_blood_pressure"] = "hypertension"
colnames(Heart)[colnames(Heart) == "DEATH_EVENT"] = "fatal"
colnames(Heart)[colnames(Heart) == "anaemia"] = "anemia"  # 'Anaemia' is British English
```

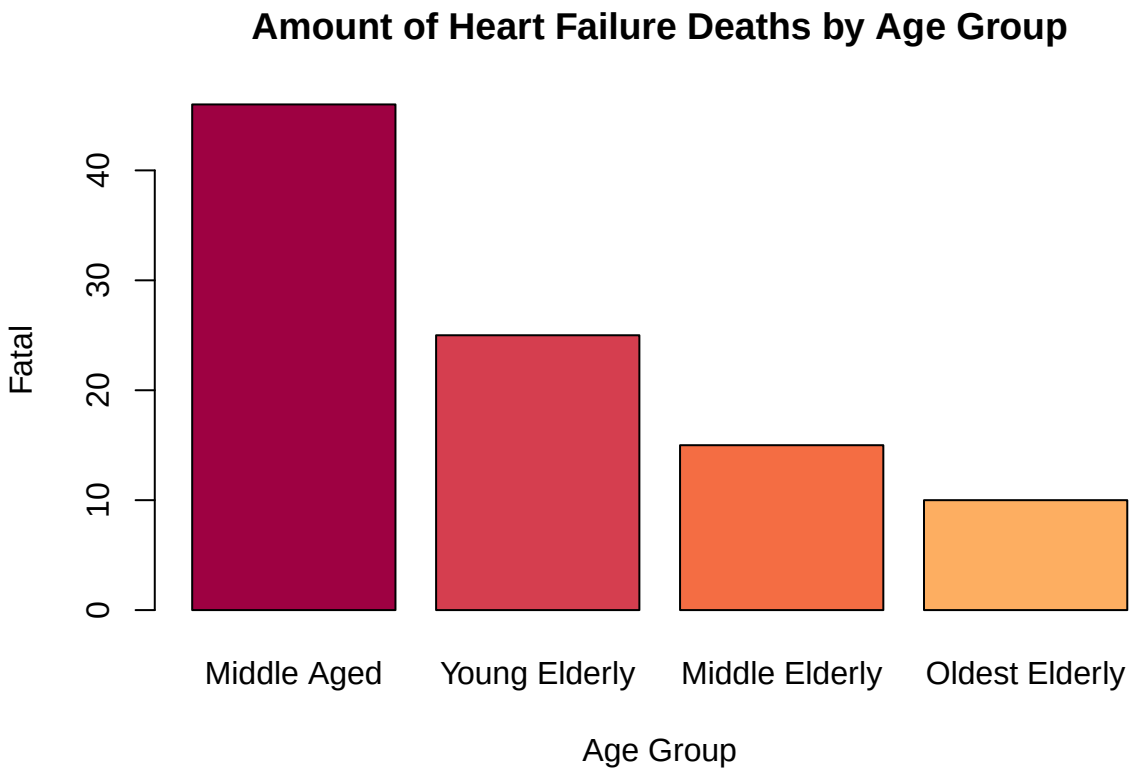
Then, created a column for age classifications. Since illnesses and health conditions increase as age increases, it is not enough to have one classification of the elderly (typically defined as 65+). Instead, we created multiple sub-classifications for the elderly. Our chosen intervals for the classifications is inspired by a study out of Korea published in the National Library of Medicine regarding elderly patients who visit the emergency room. Since our data involves adults above 40 years old but less than 65, there is also a column for the middle aged.

```
AgeGroup <- cut(Heart$age, breaks = c(40,
    64, 74, 84, 100), labels = c("Middle Aged",
    "Young Elderly", "Middle Elderly", "Oldest Elderly"))
Heart$age_group <- AgeGroup
```

Data Visualization

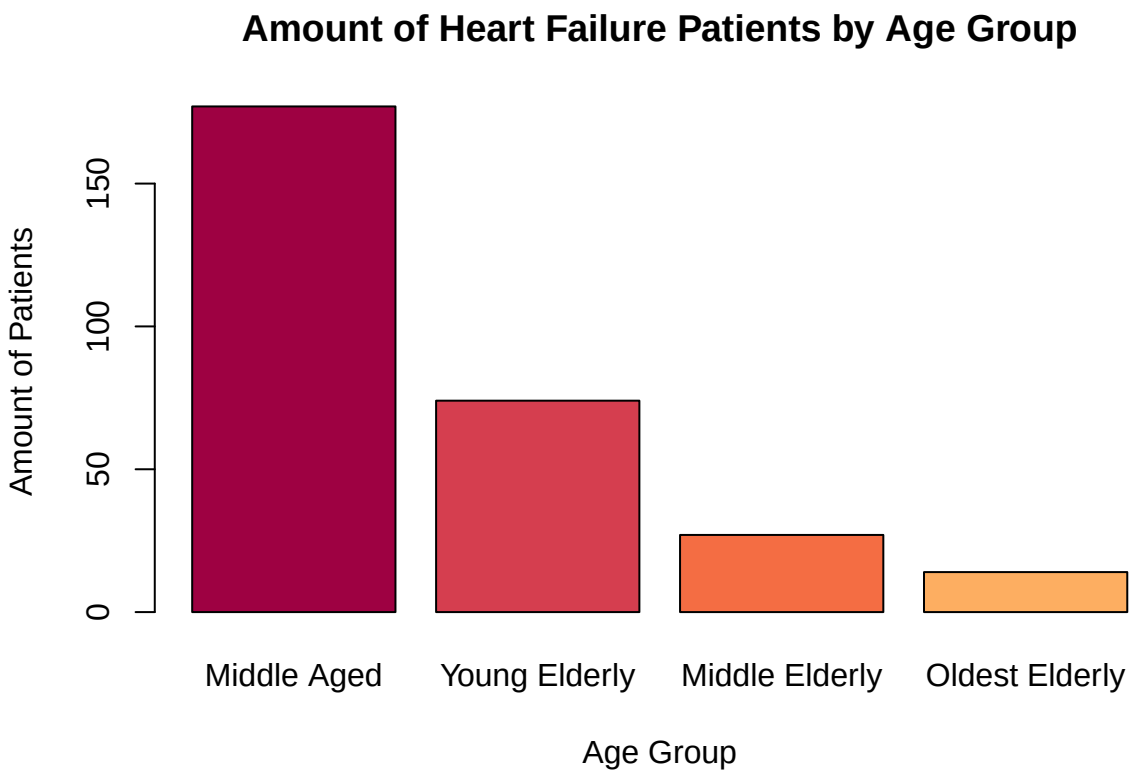
To begin, we want to visualize fatal events based on age group to understand where more of the deaths due to the chronic disease that is heart failure occurs. The following code aims to do so, and the output is below it:

```
FatalHeart <- subset(Heart, fatal == 1)
barplot(table(FatalHeart$age_group), xlab = "Age Group",
    ylab = "Fatal", main = "Amount of Heart Failure Deaths by Age Group",
    col = c(brewer.pal(11, "Spectral")))
```



It doesn't really make sense that most of our heart failure deaths would be in the middle aged because we know that we should be seeing more deaths as age progresses. So, perhaps middle aged patients are overrepresented in this data, and using a barplot that visualizes the **frequency** of failure deaths evidently does not tell the whole story. Let's find out what the distribution of patients looks like in our data, based on age group.

```
barplot(table(Heart$age_group), xlab = "Age Group",
    ylab = "Amount of Patients", main = "Amount of Heart Failure Patients by Age Group",
    col = c(brewer.pal(11, "Spectral")))
```



As we can see, middle aged heart failure patients are indeed over represented in the data. The following code shows us that those patients make up approximately 61% of the data. Additionally, this variable is also equivalent to the probability of selecting a middle aged patient from the data set.

```
pMiddleAged <- nrow(Heart[Heart$age_group ==
    "Middle Aged", ])/nrow(Heart)
pMiddleAged
```

```
## [1] 0.6153846
```

With over half of the dataset being just middle aged patients, it makes much more sense to look at probabilities of fatal events given the age group as opposed to looking at frequencies. An over-representation of a particular group in the dataset can be deceiving, which is one of the many reasons why we must compute probabilities instead.

Bayesian Reasoning

We want to understand the probability of a fatal event based on some of the health features in our data set. To start, we decide to look at how age affects the probability of a fatal event. We recall Bayes Theorem, where A and B are independent events, observations, or evidence, and P(A|B) refers to the probability that A is true or occurs given that B is true or occurs:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

This formula for computing a conditional probability will be used throughout our analysis.

The Probability of a Fatal Event: Age Group In our previous barplot, since middle aged patients are over-represented, it may **appear** that they are the most likely group to experience a fatal event. We perform a Bayesian analysis with Bayes Theorem and compute the probability of a fatal event for two groups; middle aged patients (aged 40 to 64) and oldest elderly patients (aged 85+).

```
# P(Fatal | Middle Aged) = (P(Middle
# Aged | Fatal) * P(Fatal)) / P(Middle
# Aged)
pFatal <- nrow(Heart[Heart$fatal == 1, ])/nrow(Heart)  # P(fatal) = 0.3210702
pMiddleAgedgivenFatal <- nrow(FatalHeart[FatalHeart$age_group ==
    "Middle Aged", ])/nrow(FatalHeart)  # P(Middle Aged | Fatal) = 0.4791667
pFatalgivenMiddleAged <- (pMiddleAgedgivenFatal *
    pFatal)/pMiddleAged
pFatalgivenMiddleAged  # P(Fatal | Middle Aged) = 0.25.
```

```
## [1] 0.25
```

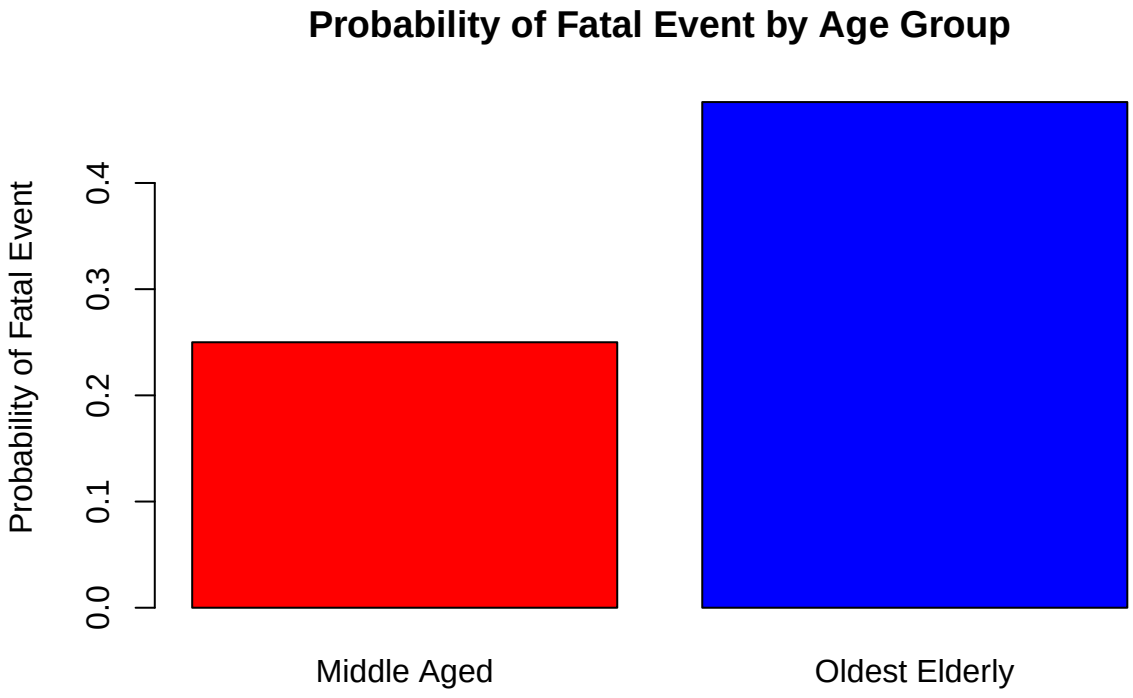
```
# P(Fatal | Oldest Elderly) = (P(Oldest
# Elderly | Fatal) * P(Fatal)) /
# P(Oldest Elderly)
pOldestElderly <- nrow(Heart[Heart$age_group ==
    "Oldest Elderly", ])/nrow(Heart)
pOldestElderlygivenFatal <- nrow(FatalHeart[FatalHeart$age_group ==
    "Oldest Elderly", ])/nrow(FatalHeart)
pFatalgivenOldestElderly <- (pOldestElderlygivenFatal *
    pFatal)/pOldestElderly
pFatalgivenOldestElderly  # P(Fatal | Oldest Elderly) = 0.4761905.
```

```
## [1] 0.4761905
```

```
pFatalgivenOldestElderly - pFatalgivenMiddleAged
```

```
## [1] 0.2261905
```

Our Bayesian analysis confirms that the probability of a fatal event in the oldest elderly group is significantly greater than the probability of a fatal event in our middle aged group. The difference in the probabilities, 0.2261905, is visualized below.



Probability of Fatal Event by Age Group

This means that patients who are part of the oldest elderly group are approximately 22% more likely to

experience a fatal event. The following plot tells much a more accurate story than a plot based on frequencies of fatal deaths does, as previously shown.

The Probability of a Fatal Event based on the Given Health Conditions In the data, we are provided with the patient data on two health conditions; hypertension and anemia.

We continue with our Bayesian analysis to find out how the probability of a fatal event is affected based on those two patient health conditions that are already specified in our dataset.

```
# P(Fatal | Not Anemic) = (P(Not Anemic
# | Fatal) * P(Fatal)) / P(Not Anemic)
pNoAnemia <- nrow(Heart[Heart$anemia == 0,
    ])/nrow(Heart)
pNoAnemiagivenFatal <- nrow(FatalHeart[FatalHeart$anemia ==
    0, ])/nrow(FatalHeart)
pFatalgivenNoAnemia <- (pNoAnemiagivenFatal *
    pFatal)/pNoAnemia
pFatalgivenNoAnemia  # P(Fatal | Not Anemic) = 0.2941176
```

```
## [1] 0.2941176
```

```
# P(Fatal | Anemic) = (P(Anemic |
# Fatal) * P(Fatal)) / P(Anemic)
pAnemic <- nrow(Heart[Heart$anemia == 1,
    ])/nrow(Heart)
pAnemicgivenFatal <- nrow(FatalHeart[FatalHeart$anemia ==
    1, ])/nrow(FatalHeart)
pFatalgivenAnemic <- (pAnemicgivenFatal *
    pFatal)/pAnemic
pFatalgivenAnemic  # P(Fatal | Anemic) = 0.3565891
```

```
## [1] 0.3565891
```

```
pFatalgivenAnemic - pFatalgivenNoAnemia
```

```
## [1] 0.0624715
```

The analysis confirms that the probability of a fatal event given that the patient is anemic (0.3565891) is greater than the probability of a fatal event given that the patient is not anemic (0.2941176). If a patient is anemic, they are approximately 6.25% more likely to experience a fatal event.

```
# P(Fatal | Not Hypertensive) = (P(Not
# Hypertensive | Fatal) * P(Fatal)) /
# P(Not Hypertensive)
pNoHypertension <- nrow(Heart[Heart$hypertension ==
    0, ])/nrow(Heart)
pNoHypertensiongivenFatal <- nrow(FatalHeart[FatalHeart$hypertension ==
    0, ])/nrow(FatalHeart)
pFatalgivenNoHypertension <- (pNoHypertensiongivenFatal *
    pFatal)/pNoHypertension
pFatalgivenNoHypertension  # P(Fatal | Not Hypertensive) = 0.2938144
```

```
## [1] 0.2938144
```

```
# P(Fatal | Hypertensive) =
# (P(Hypertensive | Fatal) * P(Fatal))
# / P(Hypertensive)
pHypertension <- nrow(Heart[Heart$hypertension ==
```

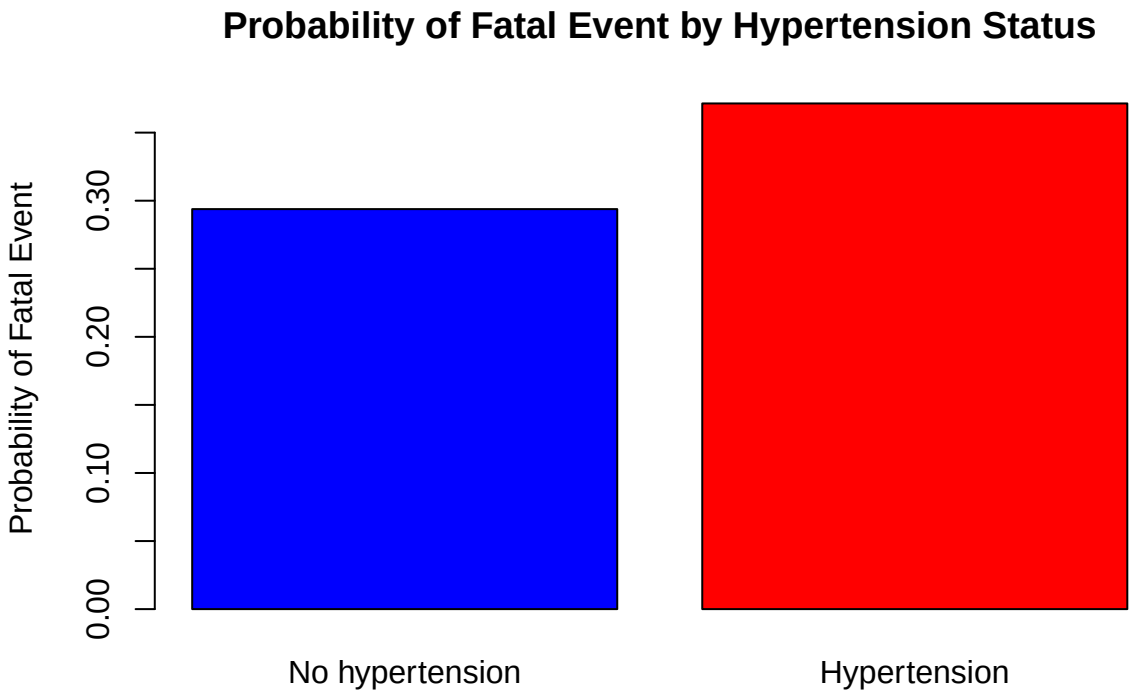
```
    1, ])/nrow(Heart)
pHypertensiongivenFatal <- nrow(FatalHeart[FatalHeart$hypertension ==
    1, ])/nrow(FatalHeart)
pFatalgivenHypertension <- (pHypertensiongivenFatal *
    pFatal)/pHypertension
pFatalgivenHypertension  # P(Fatal | Hypertension) = 0.3714286
```

```
## [1] 0.3714286
```

```
pFatalgivenHypertension - pFatalgivenNoHypertension
```

```
## [1] 0.07761414
```

The analysis confirms that the probability of a fatal event given that the patient is hypertensive is greater than the probability of a fatal event given that the patient is not hypertensive. If a patient is hypertensive, they are approximately 7.76% more likely to experience a fatal event:



Data Manipulation

We performed a Bayesian analysis to uncover the probabilities of a fatal event based on the provided health conditions in the data set, but let's get deeper. Our clinical records contain measurements of creatine phosphokinase (CPK enzyme), platelet levels, serum creatinine, serum sodium, and ejection factor. We want to find out what the relationship between these health features and the probability of a fatal event is. For that, we create new columns for new health classifications based on these features that will help us understand what they mean better.

The Mount Sinai Health System states that normal creatine phosphokinase (CPK enzyme) levels are between 10 to 120 micrograms per liter (mcg/L). Based on that fact, we will create a column that classifies each patient's CPK enzyme level according to whether it is a normal or abnormal level:

```
Heart$cpkLevel <- ifelse(Heart$cpkEnzyme >=
    10 & Heart$cpkEnzyme <= 120, "Normal",
    "Abnormal")
```

age	anemia	cpkEnzyme	diabetes	ejection	hypertension	platelets	serum_creatinine	serum_sodium	sex	smoking	time	fata
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	1
90	1	47	0	40	1	204000	2.1	132	1	1	8	1
75	1	246	0	15	0	127000	1.2	137	1	0	10	1
60	1	315	1	60	0	454000	1.1	131	1	1	10	1
65	0	157	0	65	0	263358	1.5	138	0	0	10	1
80	1	123	0	35	1	388000	9.4	133	1	1	10	1

According to Hopkins Medicine, "a normal count of platelets in the blood is between 150,000 to 450,000 platelets per microliter of blood." Greater than 450,000 platelets is indicative of a condition called thrombocytosis, and less than 150,000 is indicative of thrombocytopenia. We create a column that classifies each patient's platelet level according to whether it is normal or indicative of either health condition:

```
plateletLevel <- cut(Heart$platelets, breaks = c(0,
    149999, 449999, 999999999), labels = c("Thrombocytopenia",
    "Normal", "Thrombocytosis"))
Heart$plateletLevel <- plateletLevel
```

According to the Mayo Clinic, normal serum creatinine levels are between 0.74 to 1.35 mg/Dl for men, and between 0.59 to 1.04 mg/dL for women. Since typical serum creatinine levels are gender dependent, we create a column for gender based on our "sex" column with binary values. Then, an if else statement is used to assign the appropriate classification for each patients creatinine level, considering their gender:

```
Heart$Gender <- ifelse(Heart$sex == 1, "Male", "Female")
Heart$creatinineLevel <- ifelse(Heart$Gender == "Male", # the test
                                ifelse(Heart$serum_creatinine >= 0.74 & Heart$serum_creatinine <= 1.35,
                                ifelse(Heart$serum_creatinine >= 0.59 & Heart$serum_creatinine <= 1.04,
```

Mayo Clinic also states that normal levels of serum sodium are between 135 and 145 (mEq/L). We create a column to classify each patient's serum sodium level:

```
Heart$sodiumLevel <- ifelse(Heart$serum_sodium >=
    135 & Heart$serum_sodium <= 145, "Normal",
    "Abnormal")
```

Normal ejection fraction is also slightly gender dependent, with various classifications for what may be severely, moderately, or mildly abnormal. We create a column to classify each patient's ejection fraction, according to severity of abnormality and gender:

```
MaleEjection <- cut(Heart$ejection, breaks = c(0,
    30, 40, 51, 72, 100), labels = c("Severely Abnormal",
    "Moderately Abnormal", "Mildly Abnormal",
    "Normal", "High"))
FemaleEjection <- cut(Heart$ejection, breaks = c(0,
    30, 40, 53, 74, 100), labels = c("Severely Abnormal",
    "Moderately Abnormal", "Mildly Abnormal",
    "Normal", "High"))
Heart$ejectionTest <- ifelse(Heart$Gender ==
    "Male", as.character(MaleEjection), as.character(FemaleEjection))
```

Here is an update of our data set, with all of our new classifications:

To update our data set, FatalHeart, which is limited to patients who experienced fatal events and is used in our Bayes Theorem calculations, we now redefine it again:

Table 1: A preview of our data set limited to patients who experienced fatal events.

age	anemia	cpkEnzyme	diabetes	ejection	hypertension	platelets	serum_creatinine	serum_sodium	sex	smoking	time	fata
75	0	582	0	20	1	265000	1.9	130	1	0	4	1
55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
65	0	146	0	20	0	162000	1.3	129	1	1	7	1
50	1	111	0	20	0	210000	1.9	137	1	0	7	1
65	1	160	1	20	0	327000	2.7	116	0	0	8	1

```
FatalHeart <- subset(Heart, fatal == 1)
```

Bayesian Reasoning

We now continue with our Bayesian analysis, with our new classifications calculated based on the various lab measurements.

For example, now that we know which patients have normal or abnormal levels of CPK enzyme, we can compute the probability of a fatal event in either group.

```
# P(Fatal | Normal CPK) = (P(Normal CPK
# | Fatal) * P(Fatal)) / P(Normal CPK)
pNormalCPK <- nrow(Heart[Heart$cpkLevel ==
    "Normal", ])/nrow(Heart)
pNormalCPKgivenFatal <- nrow(FatalHeart[FatalHeart$cpkLevel ==
    "Normal", ])/nrow(FatalHeart)
pFatalgivenNormalCPK <- (pNormalCPKgivenFatal *
    pFatal)/pNormalCPK
pFatalgivenNormalCPK  # P(Fatal | Normal CPK) = 0.2467532
```

```
## [1] 0.2467532
```

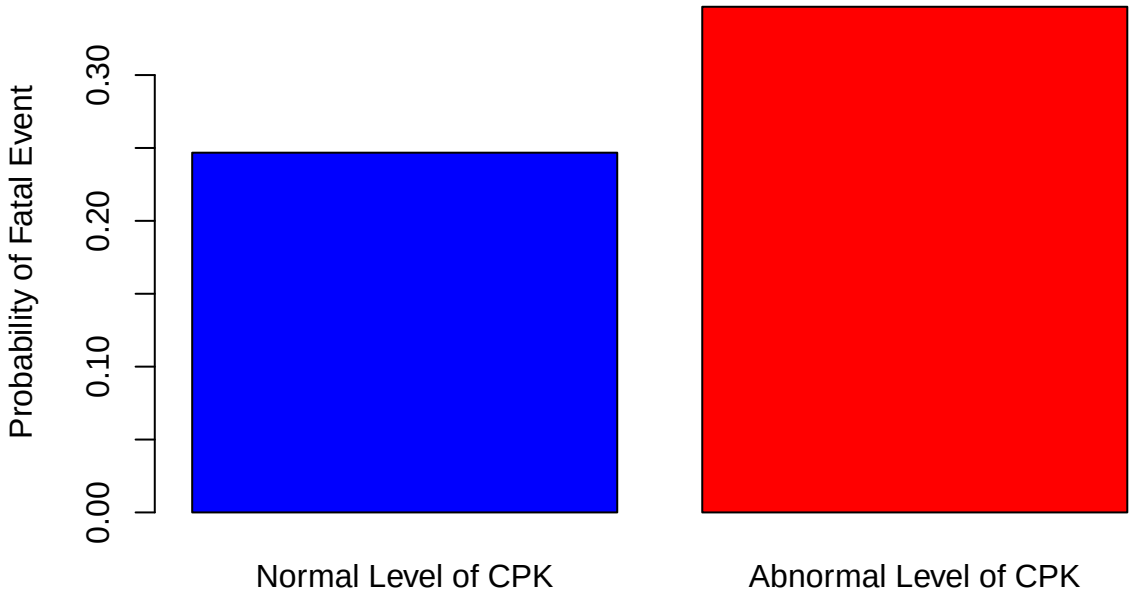
```
# P(Fatal | Abnormal CPK) = (P(Abnormal
# CPK | Fatal) * P(Fatal)) / P(Abnormal
# CPK)
pAbnormalCPK <- nrow(Heart[Heart$cpkLevel ==
    "Abnormal", ])/nrow(Heart)
pAbnormalCPKgivenFatal <- nrow(FatalHeart[FatalHeart$cpkLevel ==
    "Abnormal", ])/nrow(FatalHeart)
pFatalgivenAbnormalCPK <- (pAbnormalCPKgivenFatal *
    pFatal)/pAbnormalCPK
pFatalgivenAbnormalCPK  # P(Fatal | Abnormal CPK) = 0.3468468
```

```
## [1] 0.3468468
```

```
pFatalgivenAbnormalCPK - pFatalgivenNormalCPK
```

```
## [1] 0.1000936
```

The analysis confirms that the probability of a fatal event given that the patient has abnormal levels of CPK enzyme in their blood is greater than the probability of a fatal event given that the patient has normal levels of CPK enzyme in their blood. If a patient's level of CPK enzyme is flagged as abnormal, they are approximately 10% more likely to experience a fatal event:



We continue with our analysis, now looking at status of platelet count in the blood:

```
# P(Fatal | Normal Platelet) =
# (P(Normal Platelet | Fatal) *
# P(Fatal)) / P(Normal Platelet)
pNormalPlatelet <- nrow(Heart[Heart$plateletLevel ==
    "Normal", ])/nrow(Heart)
pNormalPlateletgivenFatal <- nrow(FatalHeart[FatalHeart$plateletLevel ==
    "Normal", ])/nrow(FatalHeart)
pFatalgivenNormalPlatelet <- (pNormalPlateletgivenFatal *
    pFatal)/pNormalPlatelet
pFatalgivenNormalPlatelet  # P(Fatal | Normal Platelet) = 0.3088803
```

```
## [1] 0.3088803
```

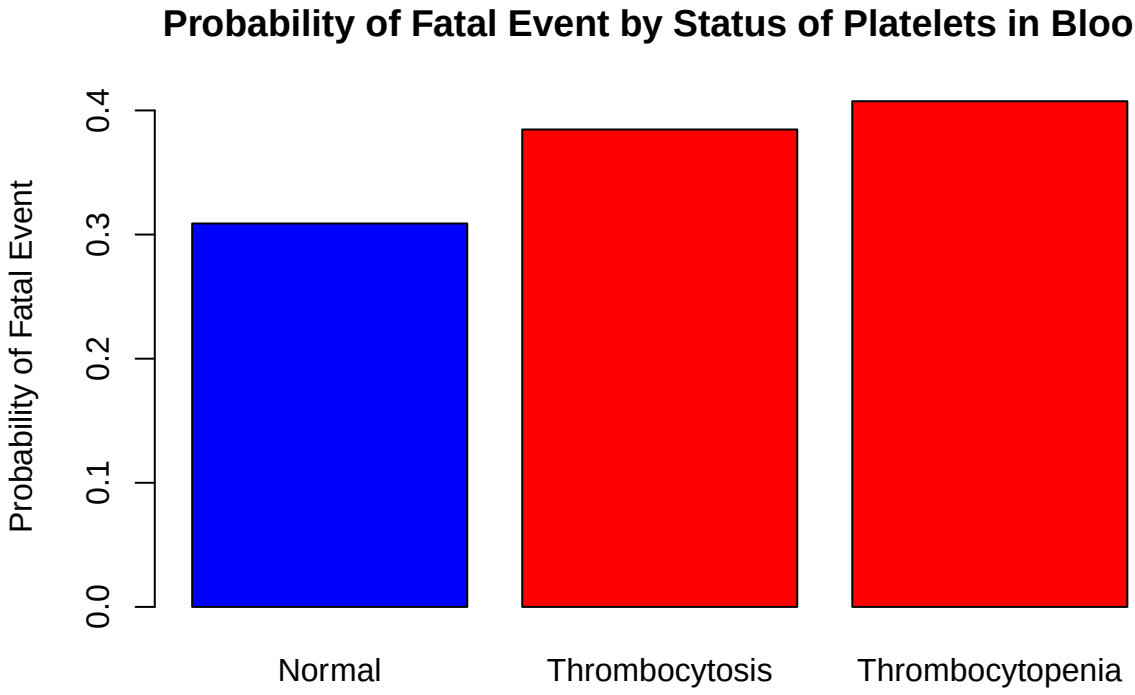
```
# Thrombocytopenia
pThrombocytopenia <- nrow(Heart[Heart$plateletLevel ==
    "Thrombocytopenia", ])/nrow(Heart)
pThrombocytopeniagivenFatal <- nrow(FatalHeart[FatalHeart$plateletLevel ==
    "Thrombocytopenia", ])/nrow(FatalHeart)
pFatalgivenThrombocytopenia <- (pThrombocytopeniagivenFatal *
    pFatal)/pThrombocytopenia
pFatalgivenThrombocytopenia  #P(Fatal | Thrombocytopenia) = 0.4074074
```

```
## [1] 0.4074074
```

```
# Thrombocytosis
pThrombocytosis <- nrow(Heart[Heart$plateletLevel ==
    "Thrombocytosis", ])/nrow(Heart)
pThrombocytosisgivenFatal <- nrow(FatalHeart[FatalHeart$plateletLevel ==
    "Thrombocytosis", ])/nrow(FatalHeart)
pFatalgivenThrombocytosis <- (pThrombocytosisgivenFatal *
    pFatal)/pThrombocytosis
pFatalgivenThrombocytosis  # P(Fatal | Thrombocytosis) = 0.3846154
```

```
## [1] 0.3846154
```

The analysis confirms that the probability of a fatal event given that the patient has an abnormal amount of platelets in their blood is greater than the probability of a fatal event given that the patient has a normal count of platelets in their blood. Below visualizes the difference in the probability of a fatal event based on the condition which their platelet count indicates for, versus a normal platelet count:



We continue with our analysis, now looking at levels of serum creatinine in the blood:

```
# P(Fatal | Normal Creatinine) =
# (P(Normal Creatinine | Fatal) *
# P(Fatal)) / P(Normal Creatinine)
pNormalCreatinine <- nrow(Heart[Heart$creatinineLevel ==
    "Normal", ])/nrow(Heart)
pNormalCreatininegivenFatal <- nrow(FatalHeart[FatalHeart$creatinineLevel ==
    "Normal", ])/nrow(FatalHeart)
pFatalgivenNormalCreatinine <- (pNormalCreatininegivenFatal *
    pFatal)/pNormalCreatinine
pFatalgivenNormalCreatinine  # P(Fatal | Normal Creatinine) = 0.2342857
```

```
## [1] 0.2342857
```

```
# P(Fatal | Abnormal Creatinine) =
# (P(Abnormal Creatinine | Fatal) *
# P(Fatal)) / P(Abnormal Creatinine)
pAbnormalCreatinine <- nrow(Heart[Heart$creatinineLevel ==
    "Abnormal", ])/nrow(Heart)
pAbnormalCreatininegivenFatal <- nrow(FatalHeart[FatalHeart$creatinineLevel ==
    "Abnormal", ])/nrow(FatalHeart)
pFatalgivenAbnormalCreatinine <- (pAbnormalCreatininegivenFatal *
    pFatal)/pAbnormalCreatinine
pFatalgivenAbnormalCreatinine  # P(Fatal | Abnormal Creatinine) = 0.4435484
```

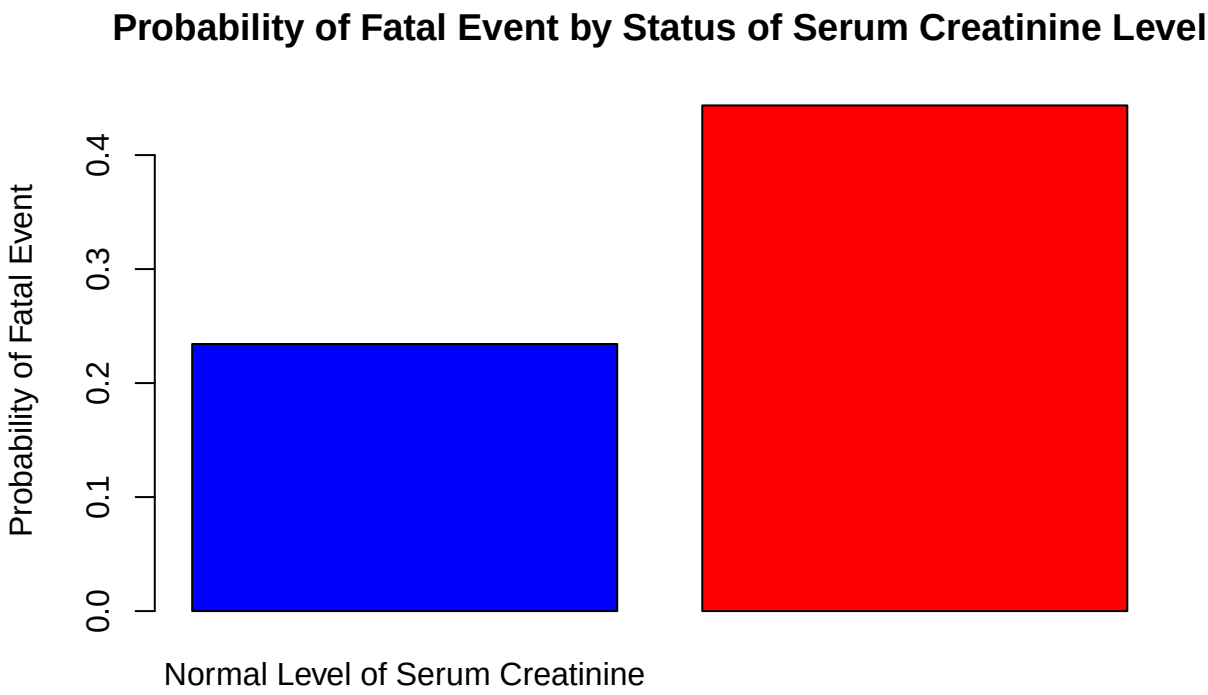
```
## [1] 0.4435484
```

```
pFatalgivenAbnormalCreatinine - pFatalgivenNormalCreatinine
```

```
## [1] 0.2092627
```

```
# Probability of a fatal event is
# higher given the patient has abnormal
# creatinine levels.
```

The analysis confirms that the probability of a fatal event given that the patient has an abnormal level of serum creatinine in their blood is greater than the probability of a fatal event given that the patient has a normal level of serum creatinine in their blood. If a patient has an abnormal level of serum creatinine in their blood, they are a whopping approximately 21% more likely to experience a fatal event:



We continue with our analysis, now looking at levels of serum sodium:

```
# P(Fatal | Normal Sodium) = (P(Normal
# Sodium | Fatal) * P(Fatal)) /
# P(Normal Sodium)
pNormalSodium <- nrow(Heart[Heart$sodiumLevel ==
    "Normal", ])/nrow(Heart)
pNormalSodiumgivenFatal <- nrow(FatalHeart[FatalHeart$sodiumLevel ==
    "Normal", ])/nrow(FatalHeart)
pFatalgivenNormalSodium <- (pNormalSodiumgivenFatal *
    pFatal)/pNormalSodium
pFatalgivenNormalSodium  # P(Fatal | Normal Sodium) = 0.2476636
```

```
## [1] 0.2476636
```

```
# P(Fatal | Abnormal Sodium) =
# (P(Abnormal Sodium | Fatal) *
# P(Fatal)) / P(Abnormal Sodium)
pAbnormalSodium <- nrow(Heart[Heart$sodiumLevel ==
    "Abnormal", ])/nrow(Heart)
pAbnormalSodiumgivenFatal <- nrow(FatalHeart[FatalHeart$sodiumLevel ==
```

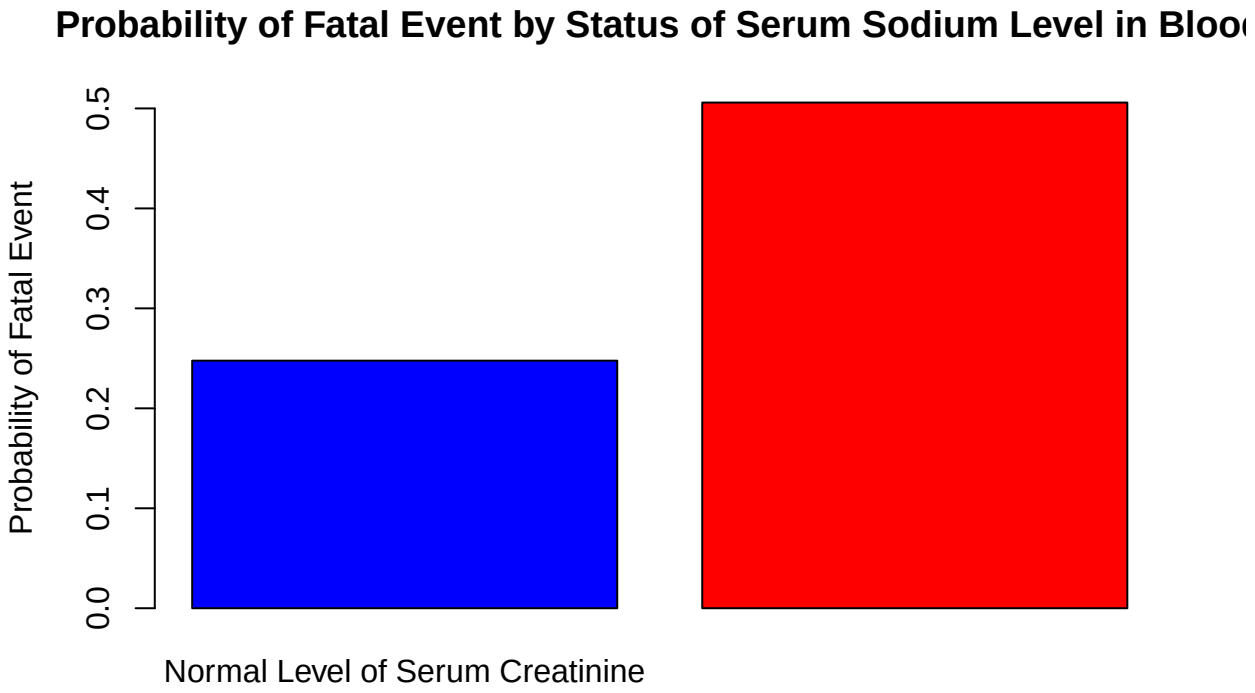
```
    "Abnormal", ])/nrow(FatalHeart)
pFatalgivenAbnormalSodium <- (pAbnormalSodiumgivenFatal *
    pFatal)/pAbnormalSodium
pFatalgivenAbnormalSodium  # P(Fatal | Abnormal Sodium) = 0.5058824
```

```
## [1] 0.5058824
```

```
pFatalgivenAbnormalSodium - pFatalgivenNormalSodium
```

```
## [1] 0.2582188
```

The analysis confirms that the probability of a fatal event given that the patient has an abnormal level of serum sodium in their blood is greater than the probability of a fatal event given that the patient has a normal level of serum sodium in their blood. If a patient has an abnormal level of serum sodium in their blood, they are a whopping approximately 25% more likely to experience a fatal event:



The last health marker that we will analyze is ejection fraction, which has multiple subclassifications for what is abnormal depending on severity.

```
# P(Fatal | Normal Ejection) =
# (P(Normal Ejection | Fatal) *
# P(Fatal)) / P(Normal Ejection)
pNormalEjection <- nrow(Heart[Heart$ejectionTest ==
    "Normal", ])/nrow(Heart)
pNormalEjectiongivenFatal <- nrow(FatalHeart[FatalHeart$ejectionTest ==
    "Normal", ])/nrow(FatalHeart)
pFatalgivenNormalEjection <- (pNormalEjectiongivenFatal *
    pFatal)/pNormalEjection
pFatalgivenNormalEjection  # P(Fatal | Normal Ejection) = 0.2105263
```

```
## [1] 0.2105263
```

```
# P(Fatal | Mildly Abnormal Ejection) =
# (P(Mildly Abnormal Ejection | Fatal)
# * P(Fatal)) / P(Mildly Abnormal
```

```
# Ejection)
pMildAbEjection <- nrow(Heart[Heart$ejectionTest ==
    "Mildly Abnormal", ])/nrow(Heart)
pMildAbEjectiongivenFatal <- nrow(FatalHeart[FatalHeart$ejectionTest ==
    "Mildly Abnormal", ])/nrow(FatalHeart)
pFatalgivenMildAbEjection <- (pMildAbEjectiongivenFatal *
    pFatal)/pMildAbEjection
pFatalgivenMildAbEjection  # P(Fatal | Mildly Abnormal Ejection) = 0.2682927
```

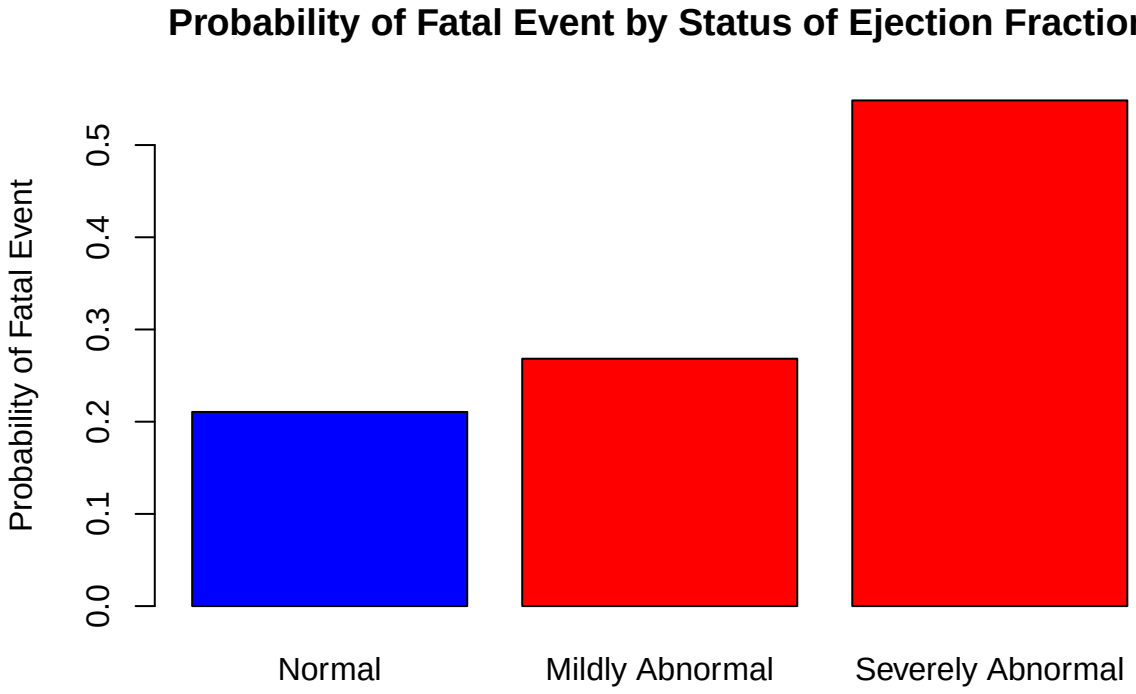
```
## [1] 0.2682927
```

```
# P(Fatal | Severely Abnormal Ejection)
# = (P(Severely Abnormal Ejection |
# Fatal) * P(Fatal)) / P(Severely
# Abnormal Ejection)
pSevereAbEjection <- nrow(Heart[Heart$ejectionTest ==
    "Severely Abnormal", ])/nrow(Heart)
pSevereAbEjectiongivenFatal <- nrow(FatalHeart[FatalHeart$ejectionTest ==
    "Severely Abnormal", ])/nrow(FatalHeart)
pFatalgivenSevereAbEjection <- (pSevereAbEjectiongivenFatal *
    pFatal)/pSevereAbEjection
pFatalgivenSevereAbEjection  # P(Fatal | Mildly Abnormal Ejection) = 0.5483871
```

```
## [1] 0.5483871
```

```
# Probability of a fatal event is
# higher given the patient has abnormal
# ejection fractions. The more severe
# the abnormality, the higher the
# probability.
```

Our analysis confirms that the probability of a fatal event if the patient has abnormal ejection fraction is greater than the probability of a fatal event if the patient has a normal ejection fraction. Additionally, the more severe their abnormality is, the greater the probability of a fatal event is. Ejection fraction is an especially important health marker that health practitioners should pay attention to in heart failure patients. A patient who has a severely abnormal ejection fraction is 34% more likely to experience a fatal event compared to a patient who has normal ejection fraction. Below is a barplot which visualizes the differences in probabilities of a fatal event based on severity of abnormality in ejection fraction:



Conclusions and Insights

Our Bayesian analysis reveals that by analyzing health markers such as age, gender, anemia, creatine phosphokinase enzyme levels, diabetes, ejection fractions, blood pressure, platelet counts, serum creatinine levels, serum sodium levels, and smoking status, we may gain valuable insights into the probability of death in heart failure patients. Generally speaking, we see increased probabilities of fatality if the patient's clinical records are abnormal in any of the health markers. The health marker which most dramatically affected the probability of a fatal event was age, levels of serum creatinine, levels of sodium creatinine, and status of ejection fraction. Health practitioners may utilize these insights to make better treatment decisions to prevent death and promote longevity in their patients. In the future, these markers may be used as variables in a linear or machine learning model which predicts fatality in heart failure patients.

Our analysis may be limited in that we have a relatively small sample of patients from one country, and we may reach more accurate conclusions with expanded data. In addition, our data comprised mostly of middle aged adults, and more data on older populations (85+) may be helpful.