

EDA_1.R

PAILLA SHIVA CHARAN

2023-12-12

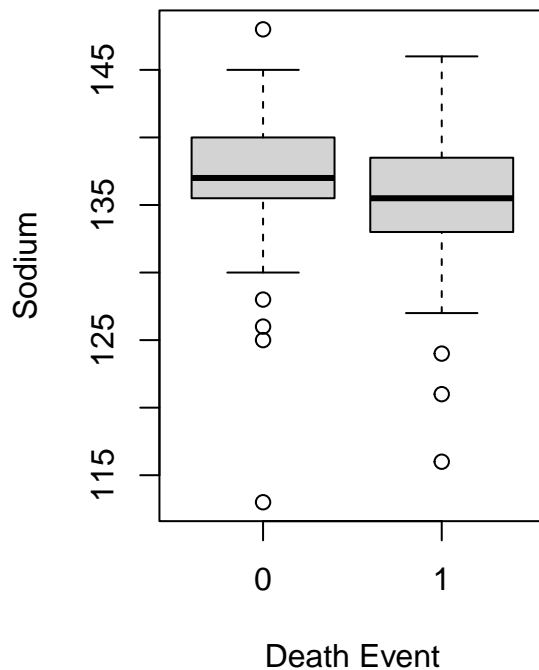
```
library(ggplot2)
```

```
data <- read.csv("S1data.csv")  
str(data)
```

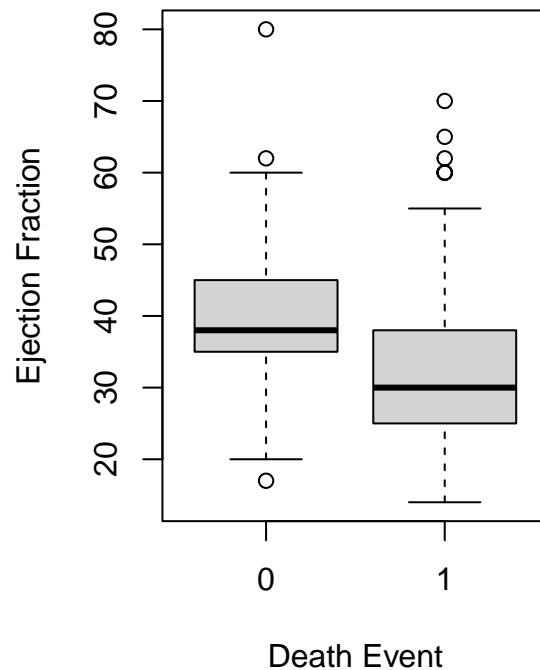
```
## 'data.frame':    299 obs. of  13 variables:  
##  $ TIME          : int   97 180 31 87 113 10 250 27 87 87 ...  
##  $ Event          : int    0 0 1 0 0 1 0 1 0 0 ...  
##  $ Gender         : int    0 1 1 1 1 1 1 1 1 1 ...  
##  $ Smoking        : int    0 1 1 0 0 0 1 0 0 1 ...  
##  $ Diabetes       : int    0 1 0 0 0 0 0 1 0 0 ...  
##  $ BP             : int    0 0 1 0 0 0 0 1 1 0 ...  
##  $ Anaemia        : int    1 1 0 1 0 1 0 0 0 0 ...  
##  $ Age            : num   43 73 70 65 64 75 70 94 75 80 ...  
##  $ Ejection.Fraction: int   50 30 20 25 60 15 40 38 45 25 ...  
##  $ Sodium         : int  135 142 134 141 137 137 136 134 137 144 ...  
##  $ Creatinine     : num   1.3 1.18 1.83 1.1 1 1.2 2.7 1.83 1.18 1.1 ...  
##  $ Pletelets      : num  237000 160000 263358 298000 242000 ...  
##  $ CPK            : int   358 231 582 305 1610 246 582 582 582 898 ...
```

```
# Boxplots  
par(mfrow = c(1,2))  
boxplot(Sodium ~ Event, data = data,  
        main = "Sodium Levels by Event",  
        xlab = "Death Event", ylab = "Sodium")  
boxplot(Ejection.Fraction ~ Event, data = data,  
        main = "Ejection Fraction by Event",  
        xlab = "Death Event", ylab = "Ejection Fraction")
```

Sodium Levels by Event



Ejection Fraction by Event



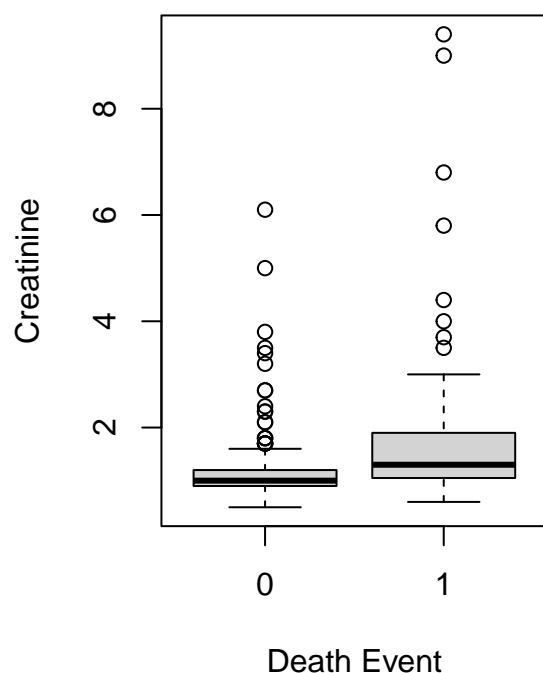
```
boxplot(Creatinine ~ Event, data = data,
        main = "Creatinine by Event",
        xlab = "Death Event", ylab = "Creatinine")
```

Shows the distribution of sodium levels and ejection fraction is lower when a death event occurs.

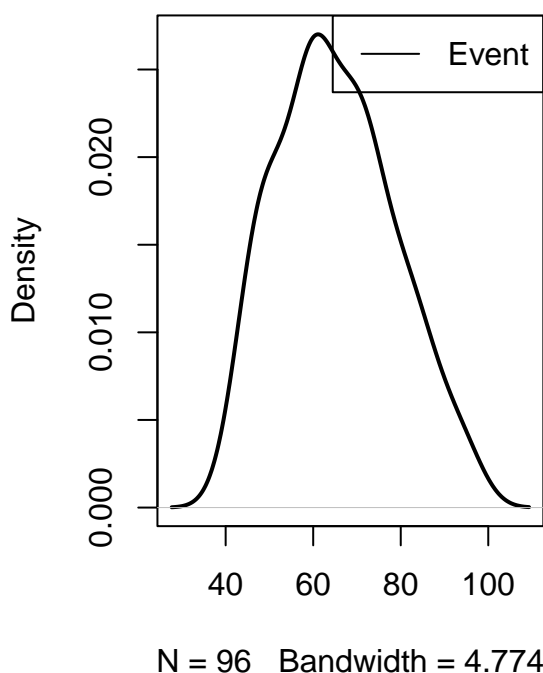
Density plots

```
dens_event = density(data$Age[data$Event==1])
plot(dens_event, lwd = 2, main = "Density Plot of Age by Event")
legend("topright", c("Event"), lwd = 1, col = c("black"))
```

Creatinine by Event



Density Plot of Age by Event



The density plot suggests age may be higher on average when death events occur.

Convert the Gender column to a factor variable

```
data$Gender <- factor(data$Gender, levels = c(0, 1), labels = c("Female", "Male"))
```

Convert the Event column to a factor variable

```
data$Event <- factor(data$Event, levels = c(0, 1), labels = c("Alive", "Dead"))
```

Check distribution of death events by gender

```
with(data, table(Gender, Event))
```

```
##           Event
## Gender  Alive Dead
##  Female    71   34
##   Male   132   62
```

Visualize death events by gender

```
ggplot(data, aes(x=Gender, fill=Event)) +
  geom_bar(position="dodge") +
  labs(x="Sex", y="Proportion",
       title="Gender by Event")
```

