ProjectThema9

Jurrien

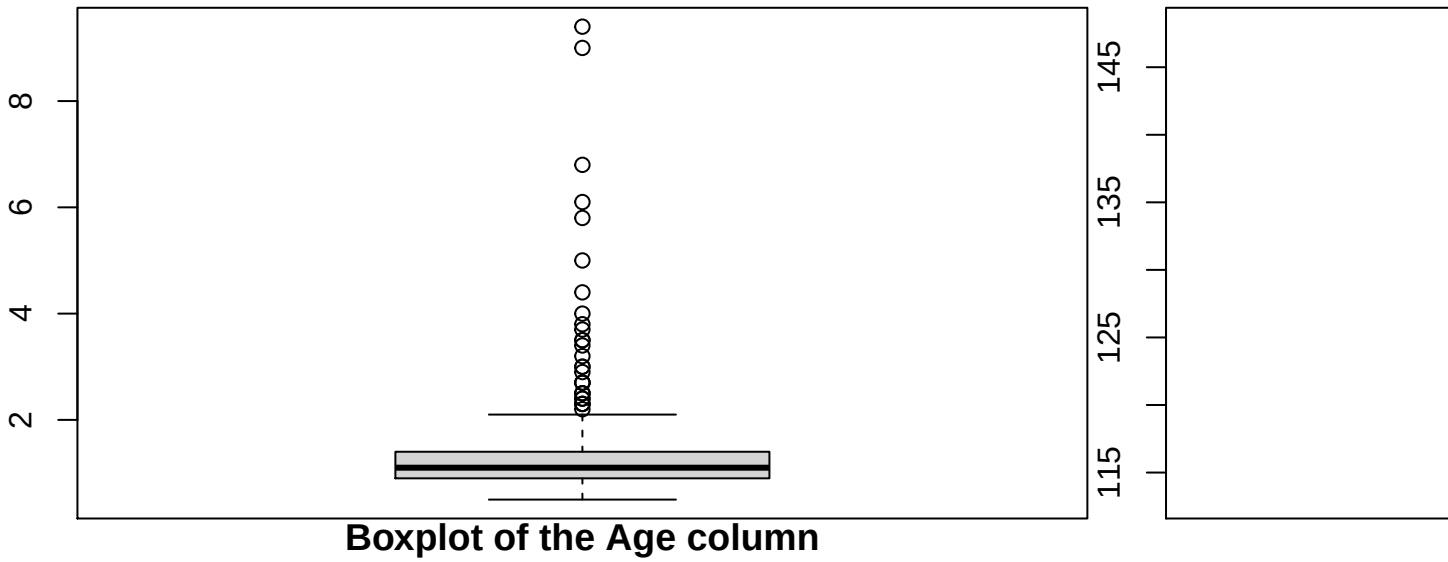
9/15/2021

The dataset analyzed for this research contains the medical records of 299 heart failure patients collected at the Faisalabad Institute of Cardiology and at the Allied Hospital in Faisalabad (Punjab, Pakistan), during April–December 2015 [52, 66]. The patients consisted of 105 women and 194 men, and their ages range between 40 and 95 years old.

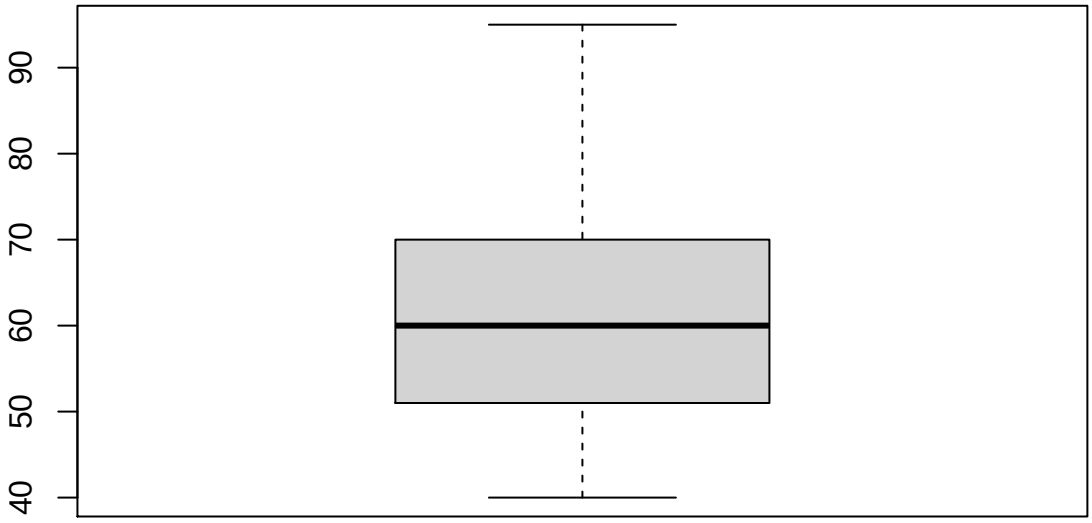
Name	Full.Name	Data.Type	Unit	Description.
age	age	int	years	age in years;
Anaemia	aneamia	boolean	0 (no) or 1 (yes)	decrease of red blood cells or hemoglobin;
HBP	high blood pressure	boolean	0 (no) or 1 (yes)	if the patient has hypertension;
CPK	creatinine phosphokinase	mcg/l (int)	level of the CPK enzyme in the blood;	
diabetes	diabetes	boolean	0 (no) or 1 (yes)	if the patient has diabetes;
ejection fraction	ejection fraction	int	percentage	percentage of blood leaving the heart at each contraction;
platelets	platelets	double	kiloplatelets/mL	platelets in the blood;
sex	sex	boolean	N/A	woman or man;
serum creatinine	serum creatinine	double	mg/dL	level of serum creatinine in the blood;
serum sodium	serum sodium	int	mEq/L	level of serum sodium in the blood;
smoking	smoking	boolean	0 (no) or 1 (yes)	if the patient smokes or not;
time	time	int	time in days	follow-up period;
death event	death event	boolean	0 (no) or 1 (yes)	if the patient deceased during the follow-up period;

The question this research is aiming to give an answer to is: “Can a death event be predicted when blood ejection fraction and age data is given using machine learning techniques?”

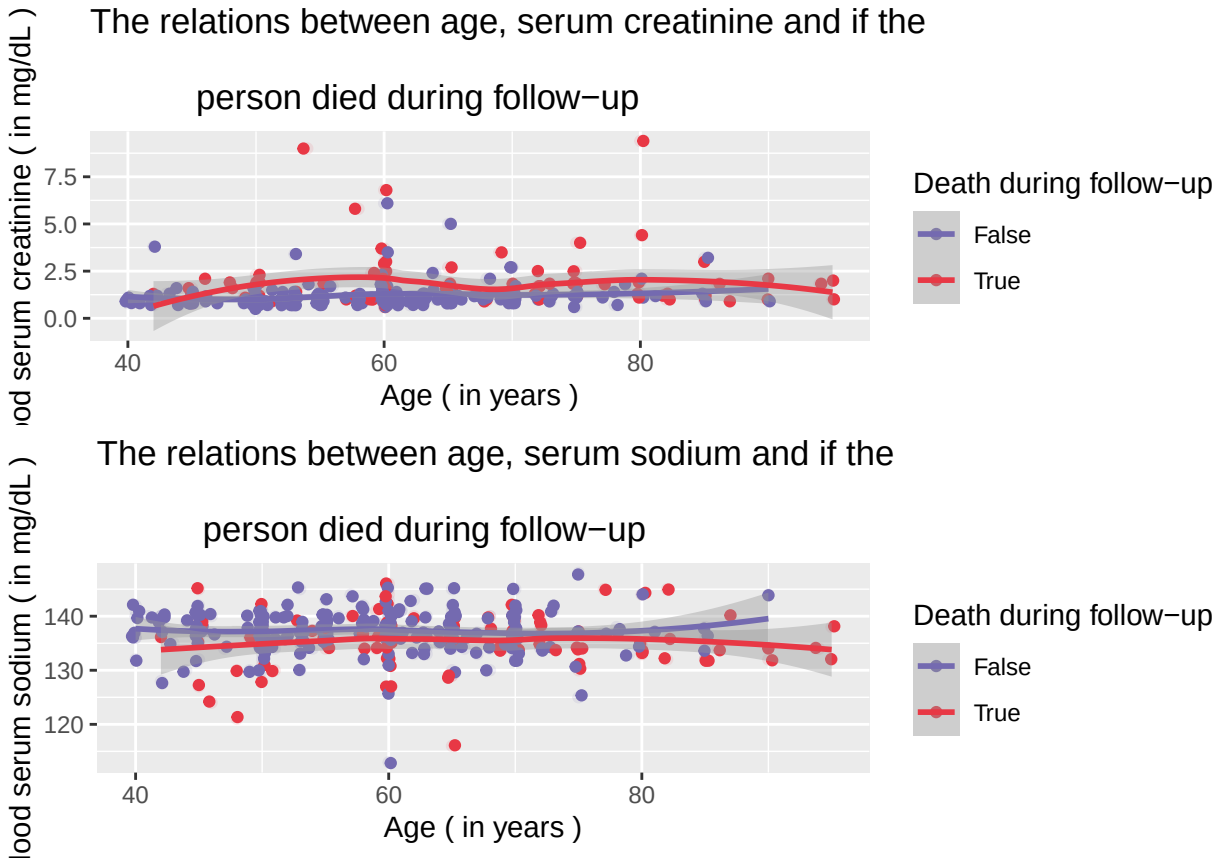
At first we are gonna check if the data contains missing values or major outliers,



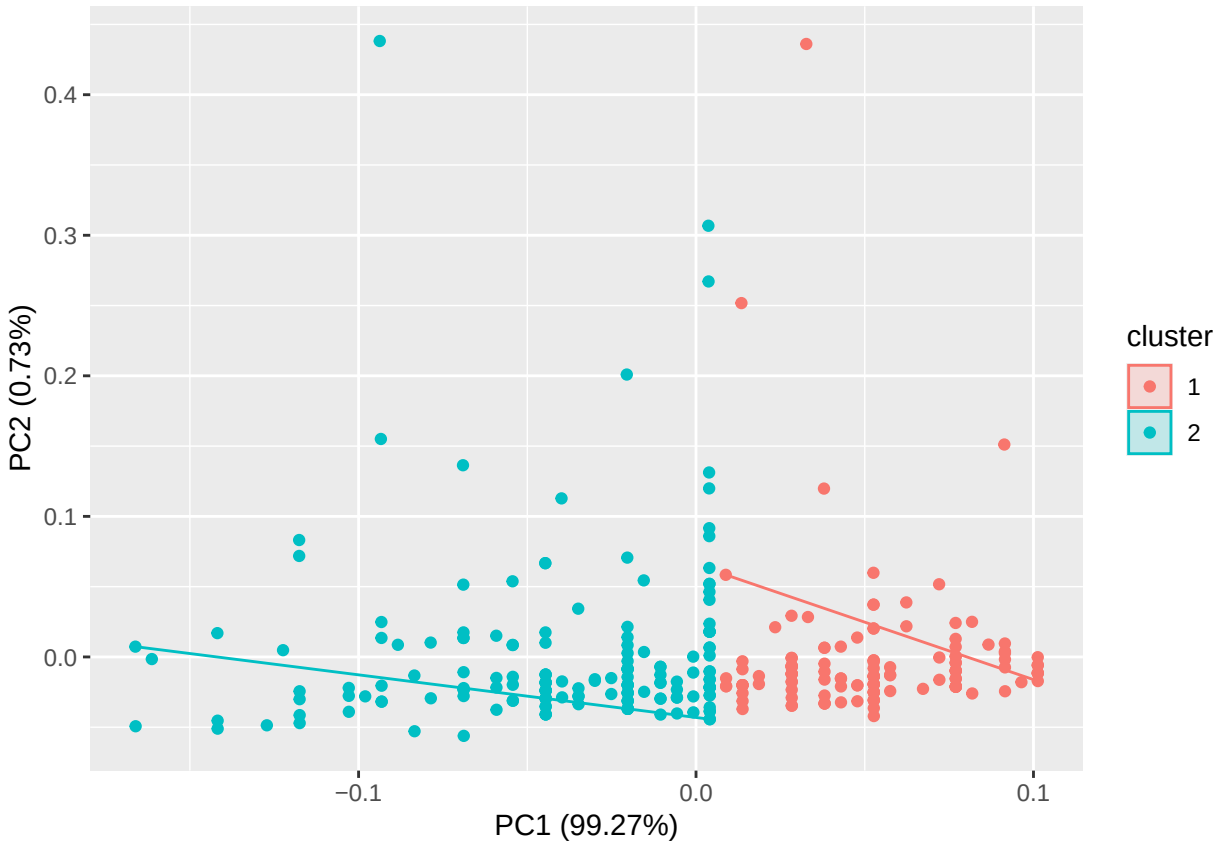
Boxplot of the Age column



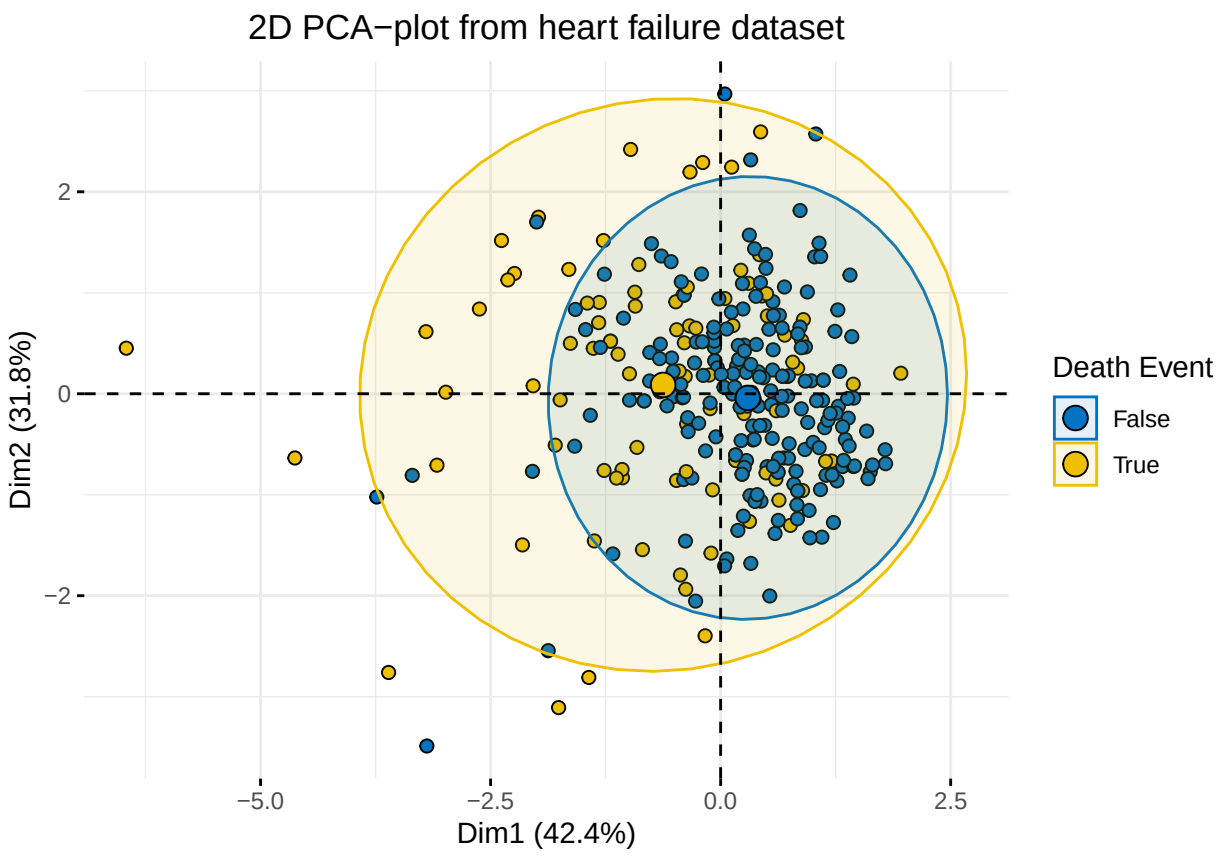

```
## # A tibble: 299 x 13
##      age anaemia creatinine_phosphokinase diabetes ejection_fracti~ high_blood_pres~
##    <int> <fct>                      <int> <fct>                <int> <fct>
##  1    75 False                        582 False                   20 True
##  2    55 False                       7861 False                   38 False
##  3    65 False                        146 False                   20 False
##  4    50 True                         111 False                   20 False
##  5    65 True                         160 True                    20 False
##  6    90 True                          47 False                   40 True
##  7    75 True                         246 False                   15 False
##  8    60 True                         315 True                    60 False
##  9    65 False                        157 False                   65 False
## 10    80 True                         123 False                   35 True
## # ... with 289 more rows, and 7 more variables: platelets <dbl>,
## #   serum_creatinine <dbl>, serum_sodium <int>, sex <fct>, smoking <fct>,
## #   time <int>, DEATH_EVENT <fct>
```




```
##        age serum_creatinine
## 1 50.00775         1.217132
## 2 69.04118         1.528000
```




```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```


2D PCA−plot from heart failure dataset
Dim2 (31.8%)
2
0
−2
−5.0
−2.5
0.0
2.5
Dim1 (42.4%)
Death Event
False
True