Code

ICU Admissions Term Project, eH705

14 April 2020

1 Statement of Authorship

This report is created by Ekene Olatunji, Catherine Nassralla, Victoria Chin, Basma Chamas, and Sajiya Somji for the eHealth 705 Statistics in eHealth final project.

2 Executive Summary

The ICU Admissions dataset, containing information about 200 individuals admitted to an intensive care unit, was explored and analysed using principles taught during the eHealth 705 course. The response variable was Status, a categorical variable indicating whether individuals lived or died on ICU admission. The analysis of this dataset was focused on understanding factors that affected Status. The demographics of this dataset, such as race and age, were not balanced, limiting some of the generalizability of conclusions drawn from the analysis. Attributes that were explored in more detail include those that might be known about patients prior to admission and lab testing, which might help assess the risk of poor outcomes during ICU admission. This might guide allocation of resources and prioritizing patients, in cases where ICU resources might be strained. It might also help inform healthcare for higher risk individuals. These attributes included Service, Cancer, Previous, Type, Age, Systolic, HeartRate, Consciousness, CPR, Infection, Sex and Renal. Using a logistic regression model, the attributes Age, Cancer, Systolic, Type, Service, Previous, and Consciousness were found to be predictors of Status.

3 About the ICU Admissions Dataset

3.1 Description

The ICU Admission dataset contains observations of patients that were admitted to an adult intensive care unit (ICU). The response variable of this dataset is Status of patients after admission, that is, whether they lived or died after admission to ICU. A description of the variables is available below.

Number of cases: 200
Variable Names:

1. ID: ID number of the patient
2. STA: Vital status (0 = Lived, 1 = Died)
3. AGE: Patient's age in years
4. SEX: Patient's sex (0 = Male, 1 = Female)
5. RACE: Patient's race (1 = White, 2 = Black, 3 = Other)
6. SER: Service at ICU admission (0 = Medical, 1 = Surgical)
7. CAN: Is cancer part of the present problem? (0 = No, 1 = Yes)
8. CRN: History of chronic renal failure (0 = No, 1 = Yes)
9. INF: Infection probable at ICU admission (0 = No, 1 = Yes)

10. CPR: CPR prior to ICU admission (0 = No, 1 = Yes)
11. SYS: Systolic blood pressure at ICU admission (in mm Hg)
12. HRA: Heart rate at ICU admission (beats/min)
13. PRE: Previous admission to an ICU within 6 months (0 = No, 1 = Yes)
14. TYP: Type of admission (0 = Elective, 1 = Emergency)
15. FRA: Long bone, multiple, neck, single area, or hip fracture (0 = No, 1 = Yes)
16. PO2: PO2 from initial blood gases (0 = >60, 1 = ^{2}60)
17. PH: PH from initial blood gases (0 = ^{3}7.25, 1 <7.25)
18. PCO: PCO2 from initial blood gases (0 = ^{2}45, 1 = >45)
19. BIC: Bicarbonate from initial blood gases (0 = ^{3}18, 1 = <18)
20. CRE: Creatinine from initial blood gases (0 = ^{2}2.0, 1 = >2.0)
21. LOC: Level of consciousness at admission (0 = no coma or stupor, 1= deep stupor, 2 = coma)

3.2 Reading in the ICU Admissions Dataset

Hide

```
> ICU <- read.table("./ICUAdmissions.csv", header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

3.3 Structure of the ICU dataset

Hide

```
> str(ICU)
```

```
'data.frame':   200 obs. of  21 variables:
 $ ID           : int  8 12 14 28 32 38 40 41 42 50 ...
 $ Status       : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Age          : int  27 59 77 54 87 69 63 30 35 70 ...
 $ Sex          : int  1 0 0 0 1 0 0 1 0 1 ...
 $ Race         : int  1 1 1 1 1 1 1 1 2 1 ...
 $ Service      : int  0 0 1 0 1 0 1 0 0 1 ...
 $ Cancer       : int  0 0 0 0 0 0 0 0 0 1 ...
 $ Renal        : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Infection    : int  1 0 0 1 1 1 0 0 0 0 ...
 $ CPR          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Systolic     : int  142 112 100 142 110 110 104 144 108 138 ...
 $ HeartRate    : int  88 80 70 103 154 132 66 110 60 103 ...
 $ Previous     : int  0 1 0 0 1 0 0 0 0 0 ...
 $ Type         : int  1 1 0 1 1 1 0 1 1 0 ...
 $ Fracture     : int  0 0 0 1 0 0 0 0 0 0 ...
 $ PO2          : int  0 0 0 0 0 1 0 0 0 0 ...
 $ PH           : int  0 0 0 0 0 0 0 0 0 0 ...
 $ PCO2         : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Bicarbonate  : int  0 0 0 0 0 1 0 0 0 0 ...
 $ Creatinine   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ Consciousness: int  1 1 1 1 1 1 1 1 1 1 ...
```

The ICU Admissions dataset consists of 200 observations with 21 variables. From these observations we found;

1. The dependent variable is the binary variable Vital Status (Status).
2. Nineteen possible predictor variables, both discrete and continuous, were also observed.
3. Most of the variables are integer but from information about the data, most of the variables can be recoded to factor variables. Several of the variables have two to three levels categories and so while they are numerical, their means do not offer any meaningful information.
4. There are only four variables that can be left as numerical variables others can be recoded to categorical/factor variables.
5. There are no missing data.

4 Recoding Variables

4.1 Converting Numerical Variables to Factor Variables

Attributes that could be recoded as factors included Status, Sex, Race, Service, Cancer, Renal, Infection, CPR, Previous, Type, Fracture, PCO2, PH, PO2, Bicarbonate, Creatinine and Consciousness. Labelling the factor levels with level names helps with comparative analysis and visualization.

Hide

```
> ICU <- within(ICU, {
+   Status <- factor(Status, labels=c('Lived','Died'))
+   Sex <- factor(Sex, labels=c('Male','Female'))
+   Race <- factor(Race, labels=c('White','Black','Other'))
+   Service <- factor(Service, labels=c('Medical','Surgical'))
+   Cancer <- factor(Cancer, labels=c('No','Yes'))
+   Renal <- factor(Renal, labels=c('No','Yes'))
+   Infection <- factor(Infection, labels=c('No','Yes'))
+   CPR <- factor(CPR, labels=c('No','Yes'))
+   Previous <- factor(Previous, labels=c('No','Yes'))
+   Type <- factor(Type, labels=c('Elective','Emergency'))
+   Fracture <- factor(Fracture, labels=c('No','Yes'))
+   PCO2 <- factor(PCO2, labels=c('No','Yes'))
+   PH <- factor(PH, labels=c('No','Yes'))
+   PO2 <- factor(PO2, labels=c('No','Yes'))
+   Bicarbonate <- factor(Bicarbonate, labels=c('No','Yes'))
+   Creatinine <- factor(Creatinine, labels=c('No','Yes'))
+   Consciousness <- factor(Consciousness, labels=c('Conscious','Deep Stupor','Coma'))
+ })
```

4.2 Recoding Numeric variables into bins

Several tests required recoding some of our numerical variables into bins. For Age, bins were created using equal count bins. Systolic was binned by blood pressure categories. HeartRate was also binned based on categories of bradycardia and elevated heart rate. the

Binning of the Age variable with equal-count bins:

Hide

```
> ICU$Age.binned <-
+   with(ICU, binVariable(Age,
+    bins=5, method='proportions',
+   labels=c('Group 1','Group 2','Group 3',
+   'Group 4','Group 5')))
```

Binning of the Systolic variable by hypotension and hypertension categories:

Hide

```
> ICU <-
+   within(ICU, {
+   Systolic.grouped <- Recode(Systolic,
+   '0:80="hypotension"; 80:120="normal"; 120:129="elevated"; 130:139="stage 1 hypertension"; 140:180="stage 2 hypertension"; 18
0:260="stage 3 hypertension"; ;',
+    as.factor=TRUE)
+ })
```

Binning of the HeartRate variable by groups of bradycardia, normal and elevated heart rate:

Hide

```
> ICU <-
+   within(ICU, {
+   HeartRate.grouped <- Recode(HeartRate,
+   '0:60="bradycardia"; 60:100="normal"; 100:200="elevated"',
+    as.factor=TRUE)
+ })
```

4.3 Saving Recoded Dataset

Hide

```
> # write.csv(ICU, file="ICUAdmissions_recoded.csv", row.names=FALSE)
```

5 Data Exploration - Statistical Summaries and Graphing of Variables

5.1 Preview of Recoded Dataset

Below is a table containing results from the first and last four variables in this dataset.

Hide

```
> headTail(ICU) %>% datatable(rownames = TRUE, filter="top", options = list(pageLenght = 10, scrollX=T))%>% formatRound(columns=
c(1:17), digits=0)
```

Show 10 entries Search:

	ID	Status	Age	Sex	Race	Service	Cancer	Renal	Infection	CPR
1	8	Lived	27	Female	White	Medical	No	No	Yes	No
2	12	Lived	59	Male	White	Medical	No	No	No	No
3	14	Lived	77	Male	White	Surgical	No	No	No	No
4	28	Lived	54	Male	White	Medical	No	No	Yes	No
...	...		...							
197	752	Died	64	Male	White	Medical	Yes	No	Yes	No
198	789	Died	60	Male	White	Medical	No	No	Yes	No
199	871	Died	60	Male	Other	Surgical	No	Yes	Yes	No
200	921	Died	50	Female	Black	Medical	No	No	No	No

Showing 1 to 9 of 9 entries Previous 1 Next

Hide

```
> str(ICU)
```

```
'data.frame':   200 obs. of  24 variables:
 $ ID               : int  8 12 14 28 32 38 40 41 42 50 ...
 $ Status           : Factor w/ 2 levels "Lived","Died": 1 1 1 1 1 1 1 1 1 1 ...
 $ Age              : int  27 59 77 54 87 69 63 30 35 70 ...
 $ Sex              : Factor w/ 2 levels "Male","Female": 2 1 1 1 2 1 1 2 1 2 ...
 $ Race             : Factor w/ 3 levels "White","Black",..: 1 1 1 1 1 1 1 1 2 1 ...
 $ Service          : Factor w/ 2 levels "Medical","Surgical": 1 1 2 1 2 1 2 1 1 2 ...
 $ Cancer           : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 2 ...
 $ Renal            : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Infection        : Factor w/ 2 levels "No","Yes": 2 1 1 2 2 2 1 1 1 1 ...
 $ CPR              : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Systolic         : int  142 112 100 142 110 110 104 144 108 138 ...
 $ HeartRate        : int  88 80 70 103 154 132 66 110 60 103 ...
 $ Previous         : Factor w/ 2 levels "No","Yes": 1 2 1 1 2 1 1 1 1 1 ...
 $ Type             : Factor w/ 2 levels "Elective","Emergency": 2 2 1 2 2 2 1 2 2 1 ...
 $ Fracture         : Factor w/ 2 levels "No","Yes": 1 1 1 2 1 1 1 1 1 1 ...
 $ PO2              : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...
 $ PH               : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ PCO2             : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Bicarbonate      : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 1 1 1 1 ...
 $ Creatinine       : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ Consciousness    : Factor w/ 3 levels "Conscious","Deep Stupor",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Age.binned       : Factor w/ 5 levels "Group 1","Group 2",..: 1 3 5 2 5 4 3 1 1 4 ...
 $ Systolic.grouped : Factor w/ 6 levels "elevated","hypotension",..: 5 3 3 5 3 3 3 5 3 4 ...
 $ HeartRate.grouped: Factor w/ 3 levels "bradycardia",..: 3 3 3 2 2 2 3 2 1 2 ...
```

Findings

- The following variables were successfully recoded; Status, Sex, Race, Service, Cancer, Renal, Infection, CPR, Previous, Type,Fracture, PCO2,PH, PO2, Bicarbonate, Creatinine, and Consciousness.

5.2 Statistical Summary of the Recoded Dataset

Hide

```
> summary(ICU)
```

```
       ID          Status         Age            Sex         Race
 Min.   :  4.0   Lived:160   Min.   :16.00   Male  :124   White:175
 1st Qu.:210.2   Died : 40   1st Qu.:46.75   Female: 76   Black: 15
 Median :412.5               Median :63.00                Other: 10
 Mean   :444.8               Mean   :57.55
 3rd Qu.:671.8               3rd Qu.:72.00
 Max.   :929.0               Max.   :92.00
     Service    Cancer    Renal     Infection  CPR         Systolic
 Medical : 93   No :180   No :181   No :116   No :187   Min.   : 36.0
 Surgical:107   Yes: 20   Yes: 19   Yes: 84   Yes: 13   1st Qu.:110.0
                                                       Median :130.0
                                                       Mean   :132.3
                                                       3rd Qu.:150.0
                                                       Max.   :256.0
   HeartRate      Previous         Type     Fracture   PO2        PH
 Min.   : 39.00   No :170   Elective : 53   No :185   No :184   No :187
 1st Qu.: 80.00   Yes: 30   Emergency:147   Yes: 15   Yes: 16   Yes: 13
 Median : 96.00
 Mean   : 98.92
 3rd Qu.:118.25
 Max.   :192.00
  PCO2     Bicarbonate Creatinine     Consciousness   Age.binned
 No :180   No :185     No :190    Conscious  :185   Group 1:41
 Yes: 20   Yes: 15     Yes: 10    Deep Stupor:  5   Group 2:39
                                  Coma       : 10   Group 3:43
                                                    Group 4:44
                                                    Group 5:33

             Systolic.grouped   HeartRate.grouped
 elevated            :17      bradycardia: 15
 hypotension         :12      elevated   : 80
 normal              :61      normal     :105
 stage 1 hypertension:30
 stage 2 hypertension:65
 stage 3 hypertension:15
```

5.3 Numerical Summary of Integer Variables

Hide

```
> numSummary(ICU[,c("Age", "HeartRate", "Systolic"), drop=FALSE], statistics=c("mean", "sd", "IQR",
+   "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

```
             mean       sd   IQR 0%    25% 50%    75% 100%   n
Age        57.545 20.05465 25.25 16  46.75  63  72.00   92 200
HeartRate  98.925 26.82962 38.25 39  80.00  96 118.25  192 200
Systolic  132.280 32.95210 40.00 36 110.00 130 150.00  256 200
```

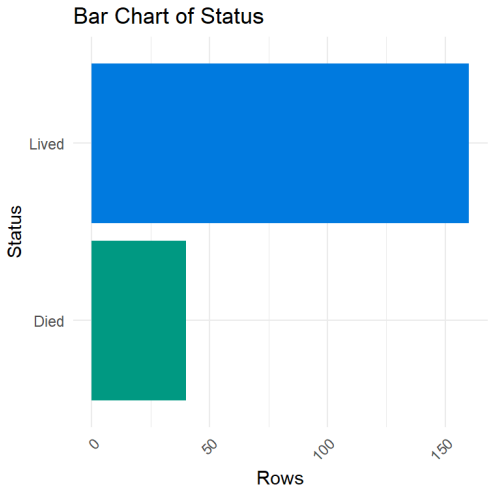
Findings

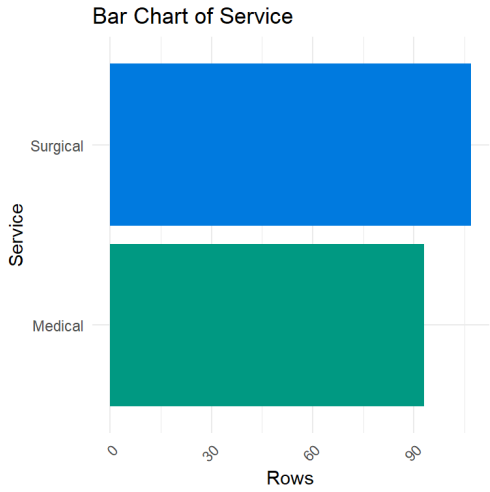
- Age ranges from 16-92 years old with a mean of 57.55 and median of 63
- Systolic blood pressure ranges from 36-256 mmHg with a mean of 132.3 and median of 130
- Heart rate ranges from 39-192 beats per minute with a mean of 98.92 and median of 96

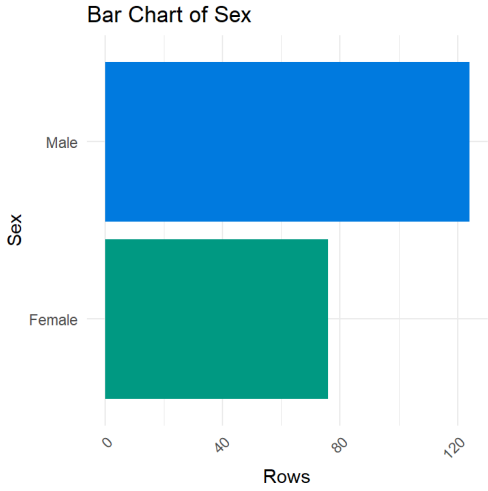
6 Data Exploration - Graphs

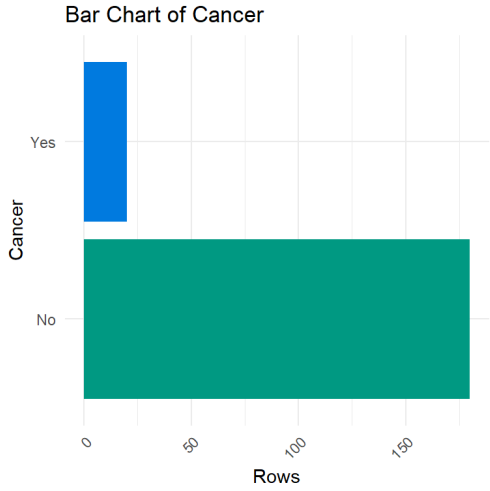
6.1 Examining the Variables Distribution

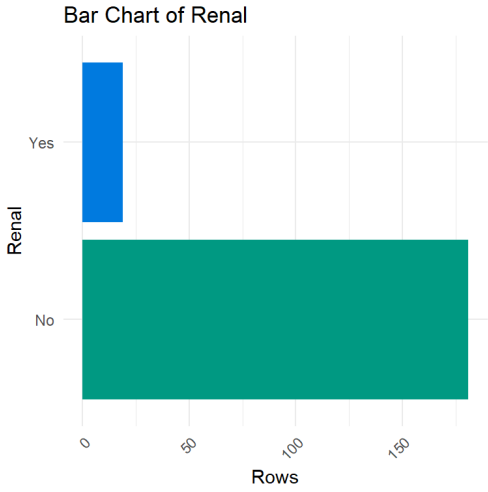
```
================================================================================
```

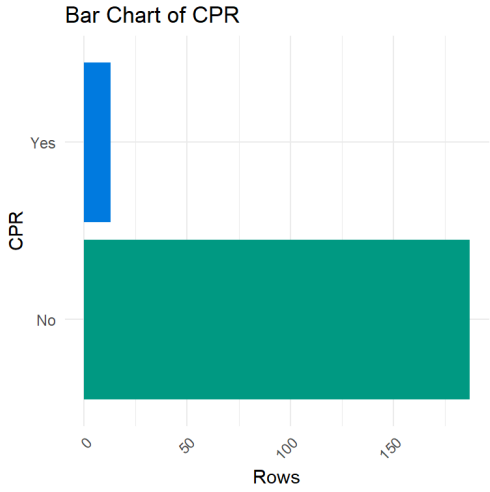

Bar Chart of Status
Lived
Died
Status
0
50
100
150
Rows

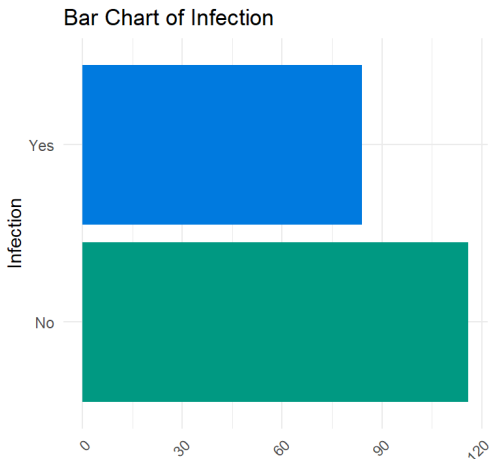

Bar Chart of Service
Surgical
Medical
Service
0
30
60
90
Rows

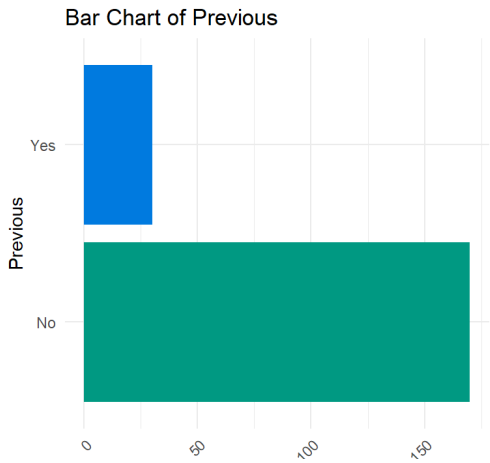

Bar Chart of Sex
Male
Female
Sex
0
40
80
120
Rows

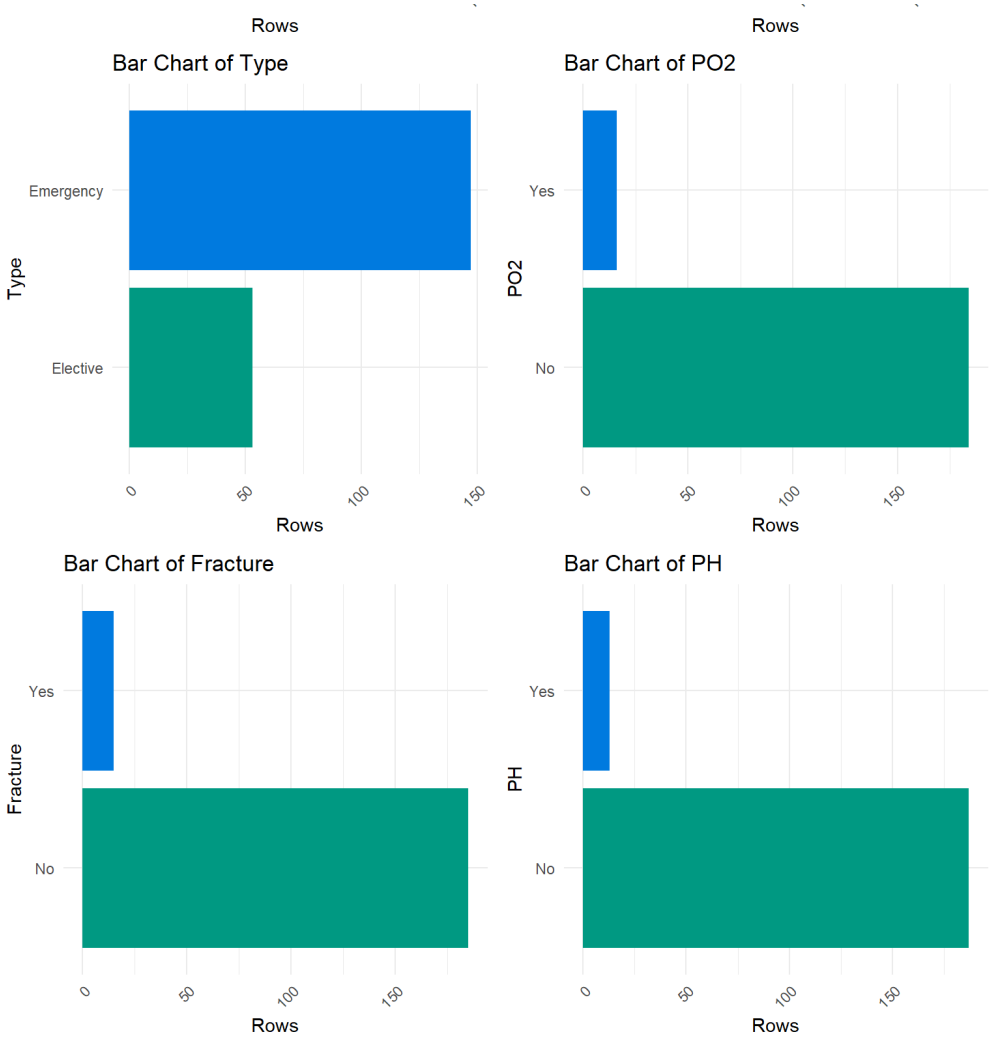

Bar Chart of Cancer
Yes
No
Cancer
0
50
100
150
Rows

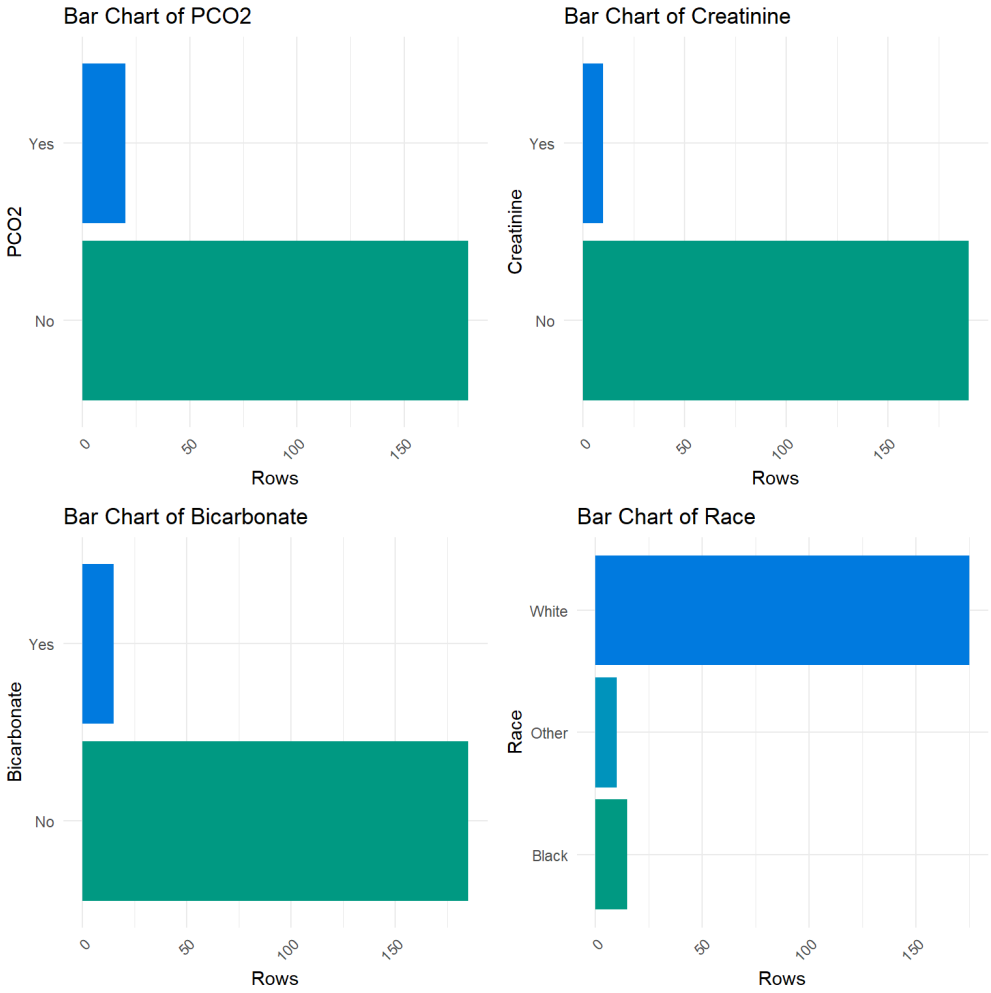

Bar Chart of Renal
Yes
No
Renal
0
50
100
150
Rows

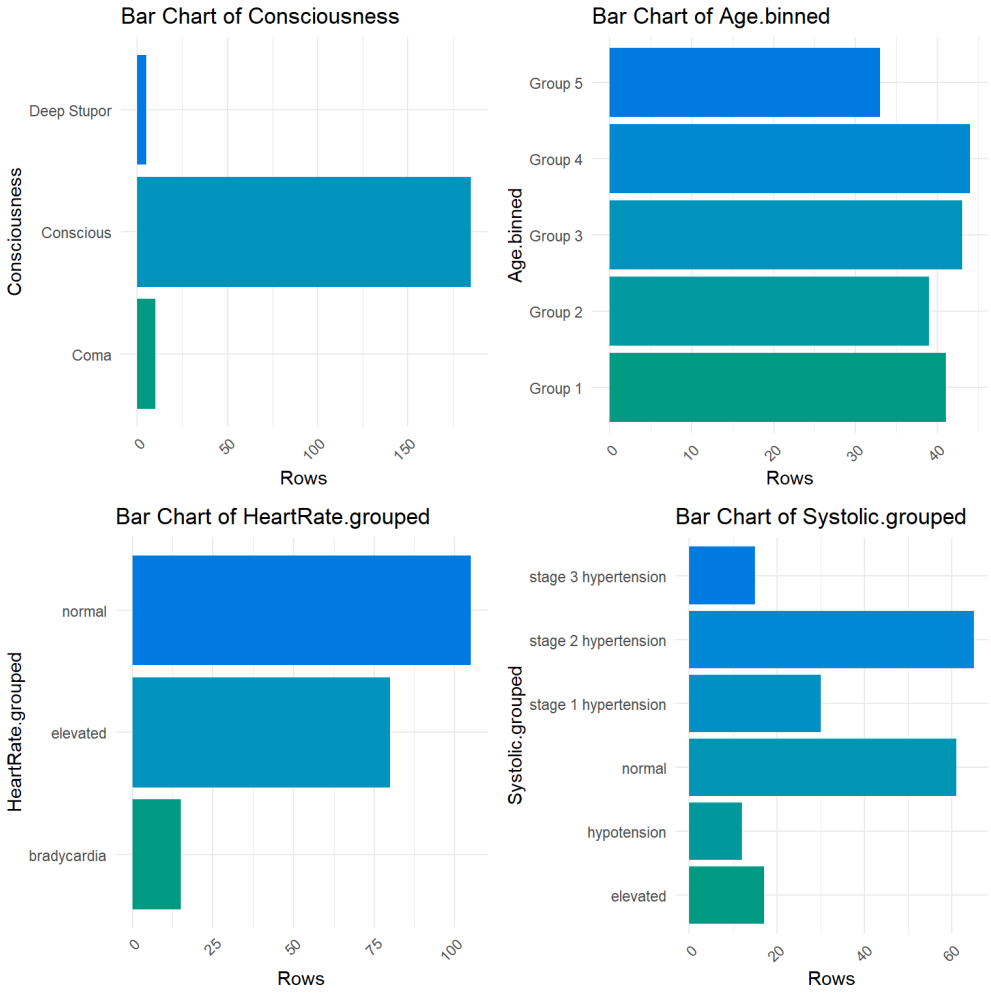

Bar Chart of CPR
Yes
No
CPR
0
50
100
150
Rows

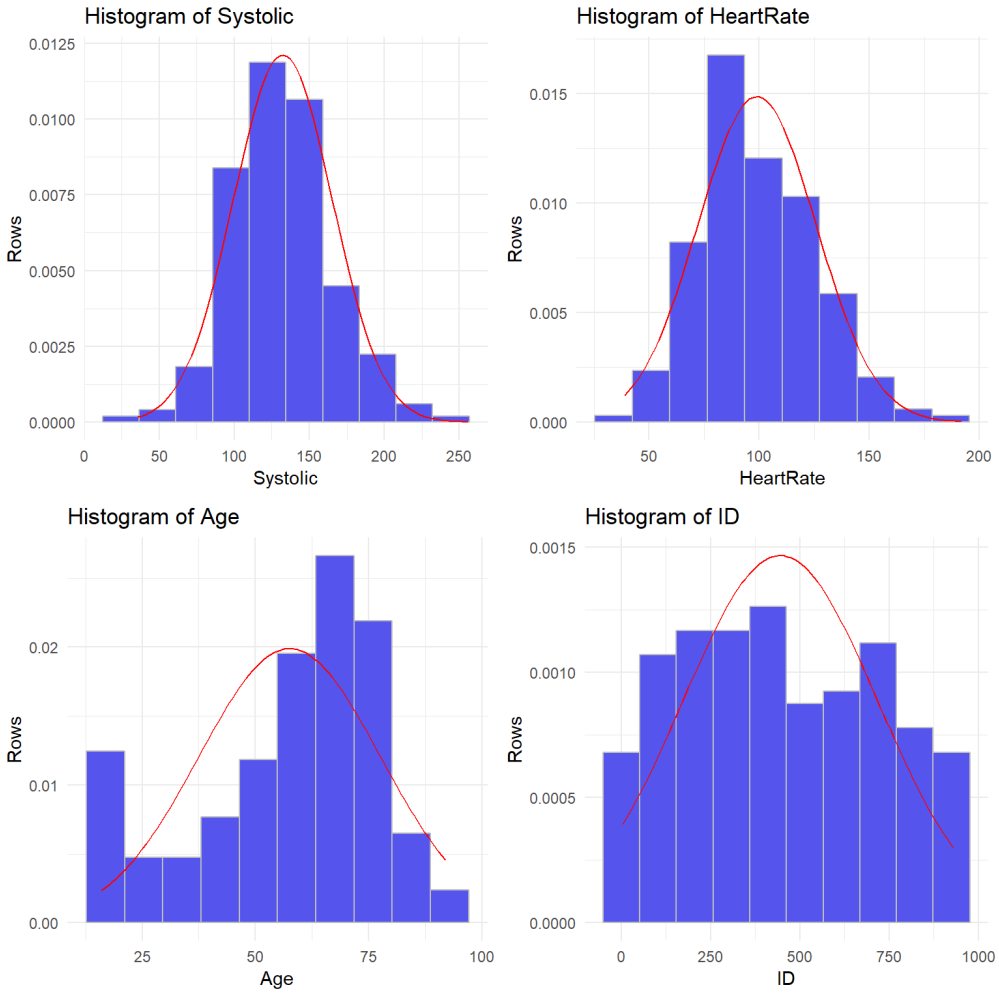

Bar Chart of Infection
Yes
No
Infection
0
30
60
90
120


Bar Chart of Previous
Yes
No
Previous
0
50
100
150


Rows
Bar Chart of Type
Emergency
Elective
Type
0
50
100
150
Rows
Rows
Bar Chart of PO2
Yes
No
PO2
0
50
100
150
Rows
Bar Chart of Fracture
Yes
No
Fracture
0
50
100
150
Rows
Bar Chart of PH
Yes
No
PH
0
50
100
150
Rows


Bar Chart of PCO2
PCO2
Yes
No
0
50
100
150
Rows
Bar Chart of Creatinine
Creatinine
Yes
No
0
50
100
150
Rows
Bar Chart of Bicarbonate
Bicarbonate
Yes
No
0
50
100
150
Rows
Bar Chart of Race
Race
White
Other
Black
0
50
100
150
Rows


Bar Chart of Consciousness
Consciousness
Deep Stupor
Conscious
Coma
0
50
100
150
Rows
Bar Chart of Age.binned
Age.binned
Group 5
Group 4
Group 3
Group 2
Group 1
0
10
20
30
40
Rows
Bar Chart of HeartRate.grouped
HeartRate.grouped
normal
elevated
bradycardia
0
25
50
75
100
Rows
Bar Chart of Systolic.grouped
Systolic.grouped
stage 3 hypertension
stage 2 hypertension
stage 1 hypertension
normal
hypotension
elevated
0
20
40
60
Rows




```
   Variable   p_1 p_10   p_25  p_50   p_75  p_90   p_99
1  Systolic 55.92   92    110   130    150   170 212.12
2       Age 16.99   21  46.75    63     72    78     91
3 HeartRate 45.98   65     80    96 118.25 136.1 162.08
4        ID 11.96 81.3 210.25 412.5 671.75 829.8 924.01
```

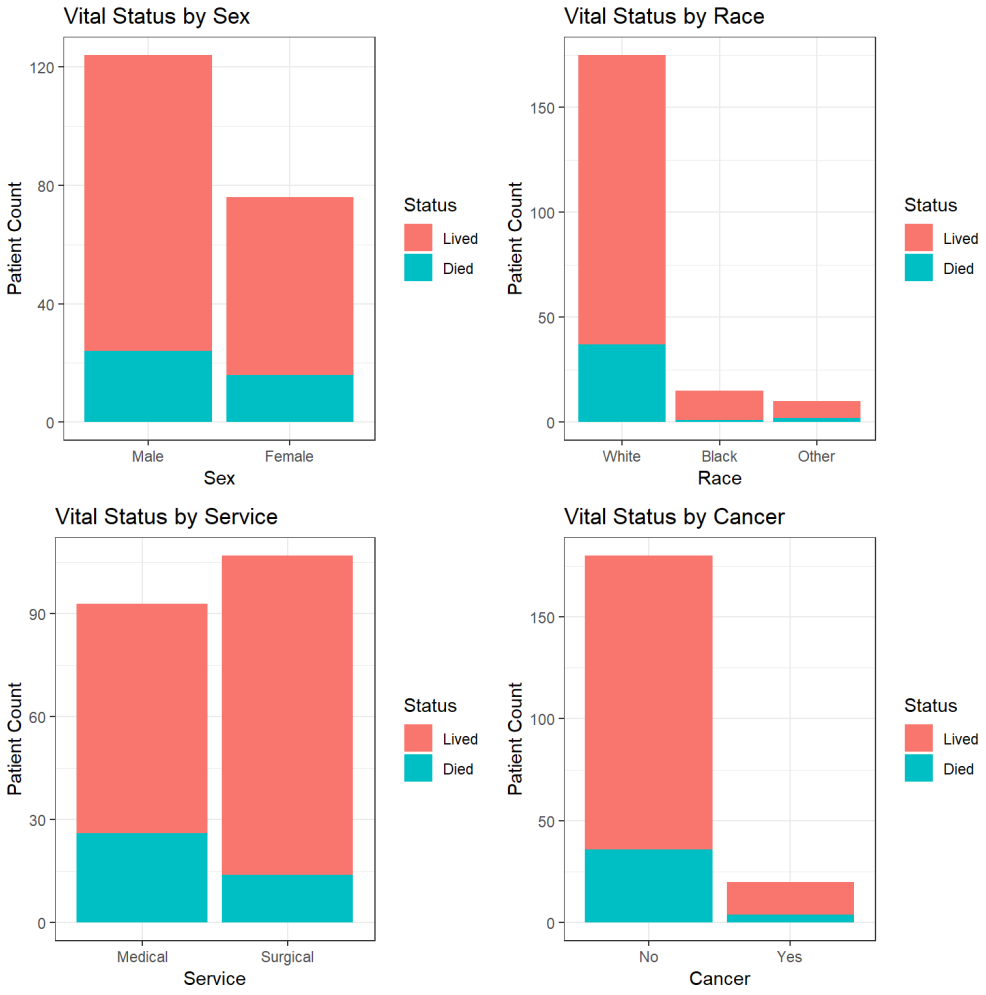
Using the xray package, general trends in the dataset were identified.

1. There are more male observations in the dataset than females.
2. Many of the admissions to the ICU were emergencies, with a about a quarter of admissions being elective. This could relate to elective surgical procedures where morbidity could have been high or complications occurred. It is unclear whether people would be preemptively admitted to the ICU for high-risk procedures or if these elective admissions could be thought of as unforeseen or emergencies in themselves.
3. Looking at Status, more than 75% of observations lived after their ICU admission.
4. While the numbers are roughly even, slightly more procedures were surgical.
5. More admitted patients had no chronic renal failure, no previous admissions to the ICU, fracture, CPR or cancer when admitted. Most admitted patients had PO2 above or equal to 60, blood pH above or equal to 7.25, PCO2 below or equal to 45, and creatinine below or equal to 2. Much more of the observations were also white.
6. Most patients admitted were conscious, with slighly more patients being comatose than in a deep stupor if unconscious.
7. Looking at histograms of the three numerical variables in the dataset, which were Age, Systolic and HeartRate, it could be guess that if any of the distributions were to be normal, they would be Systolic and HeartRate. Age is very obviously not normally distributed. Systolic looks like it is centered around 140, and the mean confirms this as mentioned above. The mean of HeartRate seems to be centered around 100 and its mean confirms this visual estimate.

6.2 Distribution of Vital Status by Factor Variables

Hide

```
> library(ggplot2)
> f01<-ggplot(ICU, aes(x=Sex, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Sex")
>
> f02<-ggplot(ICU, aes(x=Race, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Race")
> f03<-ggplot(ICU, aes(x=Service, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Service")
>
> f04<-ggplot(ICU, aes(x=Cancer, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Cancer")
>
> library(Rmisc)
> multiplot(f01, f02, f03, f04,  layout=matrix(c(1:4), nrow=2, byrow=TRUE))
```



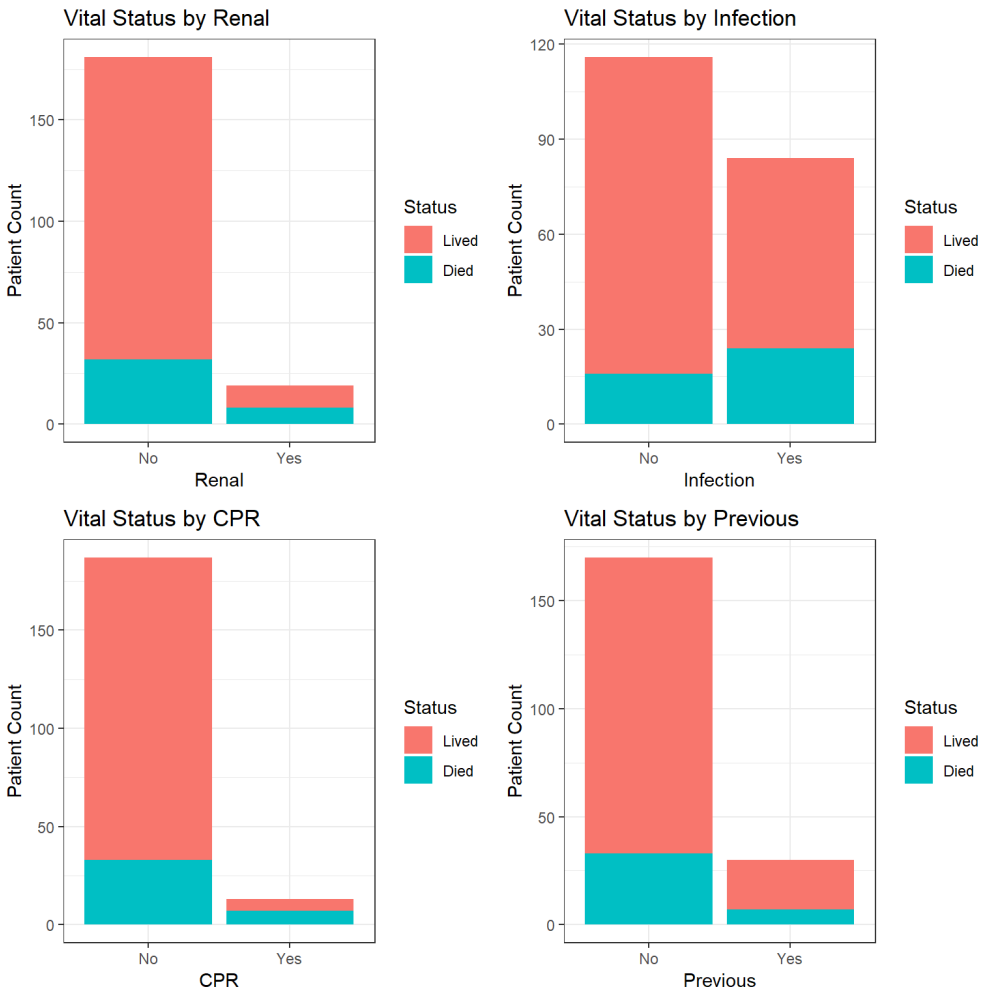
Observation

- Due to the increased ratio in the population ofmale to female, the ratio of the number of patients who lived is found almost directly proportional between male and female.
- From the graph regarding vital Status by race, it can be seen that the patient population of the White race is higher than the patient count for Black and other races. Overall, in the different races there are more cases of patients that lived than died.
- For the Vital status by Service, we can also observe that more patient died receiving medical service than surgical service.

- In the case of the Vital status by Cancer graph, there are more patients without a present case of Cancer, however there are als more patients who died and do not have a present case of Cancer.

Hide

```
> f05<-ggplot(ICU, aes(x=Renal, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Renal")
>
> f06<-ggplot(ICU, aes(x=Infection, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Infection")
>
> f07<-ggplot(ICU, aes(x=CPR, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by CPR")
>
> f08<-ggplot(ICU, aes(x=Previous, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Previous")
>
> library(Rmisc)
> multiplot(f05, f06, f07, f08,  layout=matrix(c(1:4), nrow=2, byrow=TRUE))
```



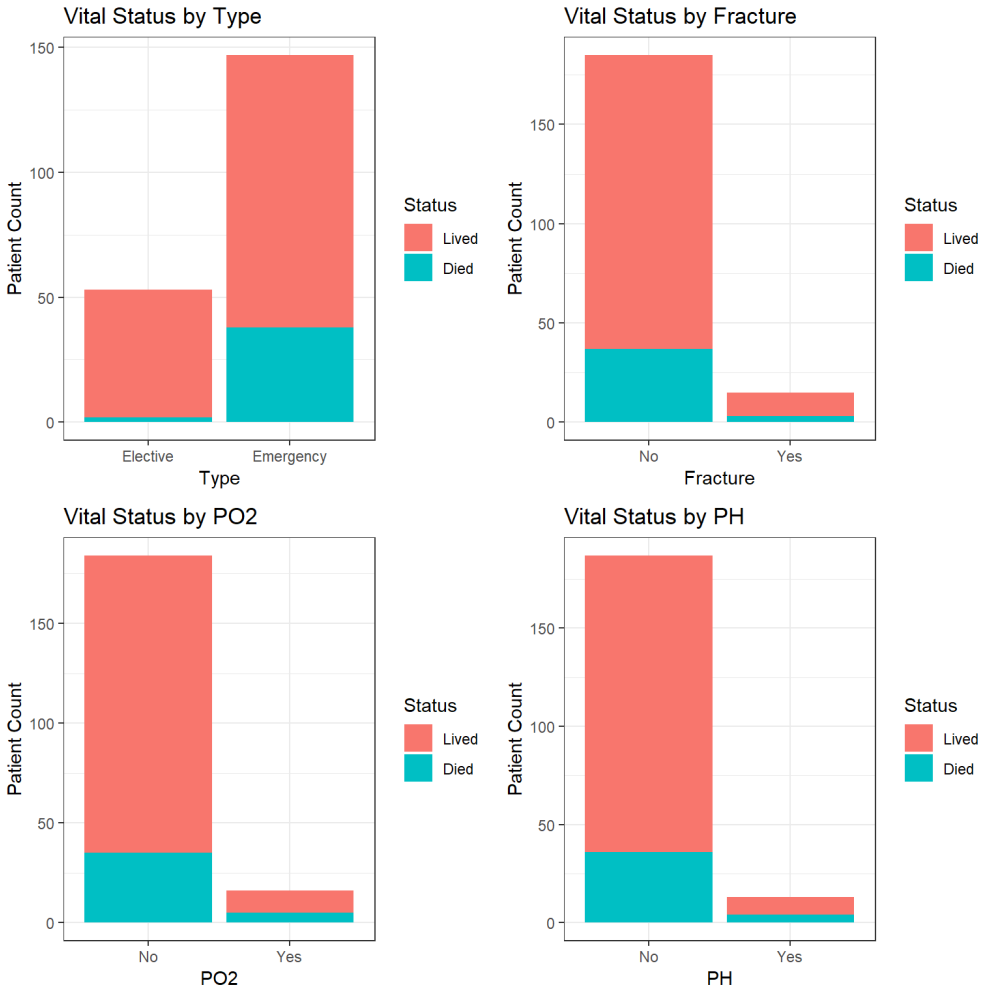
Observation

- For the Vital Status by Renal, there are more patients without history of chronic Renal failure and most of then lived however, it is also observed that there seems to be an equal proportion in Status of patients who had an history with chronic Renal failure.

- In the case of Vital status by Infection, it can be observed that there are more cases of patients that lived than died regardless of if they have infection or not. However, there are more cases of patients who probably had infections and died.
- For the vital Status by CPR, it can be observed that there are more patients who were admitted in the ICU without prior CPR but there is the likelihood of patients dying with prior CPR.
- The visualization of Vital Status by Previous shows that there are more patients without previous ICU admission and the cases of patients who died was low.

Hide

```
> f09<-ggplot(ICU, aes(x=Type, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Type")
>
> f10<-ggplot(ICU, aes(x=Fracture, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Fracture")
>
> f11<-ggplot(ICU, aes(x=PO2, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by PO2")
>
> f12<-ggplot(ICU, aes(x=PH, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by PH")
>
> library(Rmisc)
> multiplot(f09, f10, f11, f12,  layout=matrix(c(1:4), nrow=2, byrow=TRUE))
```

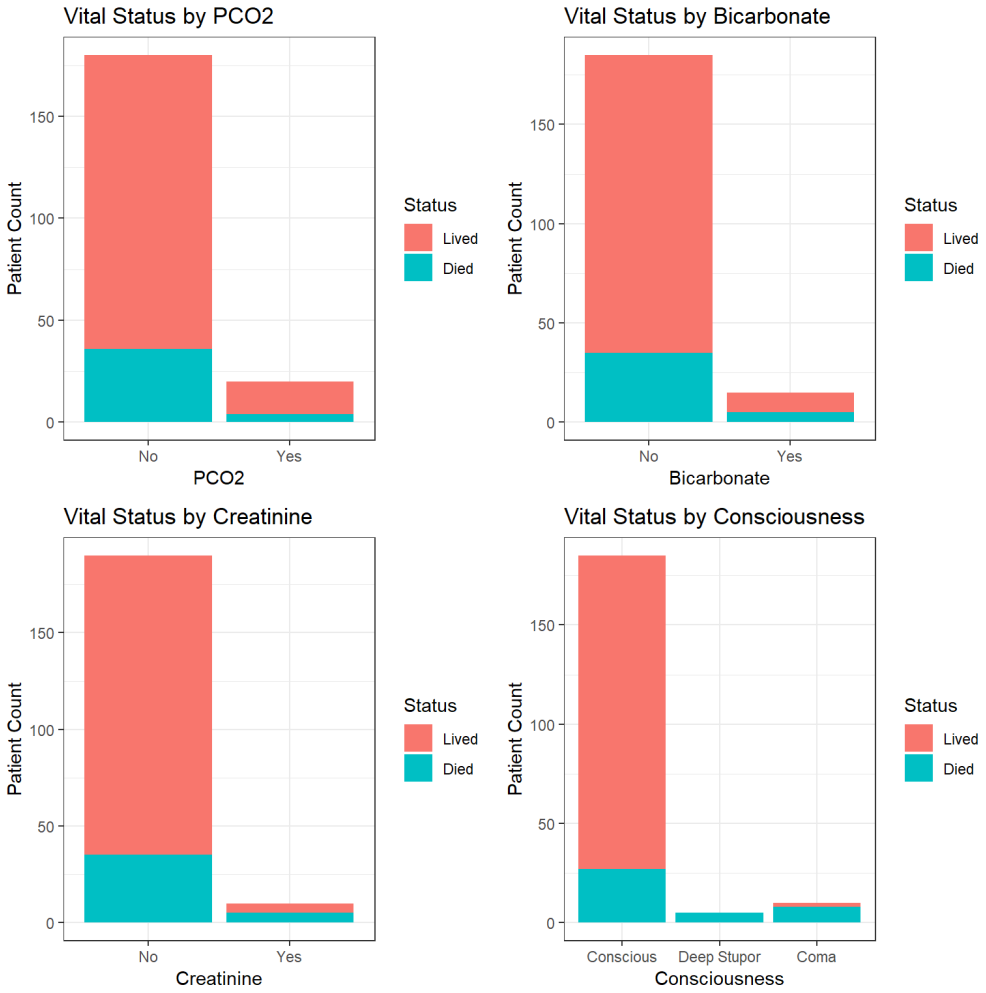


Observation

- For the Vital Status by Type, there are more patients with emergency ICU admission and those also expereince more cases of death than the elective ICU admissions.
- In the case of Vital status by Fracture, it can be observed that there are more cases of patients that lived than died regardless of if they have Fracture or not.
- For the vital Status by PO2, it can be observed that there are more patients who were admitted in the ICU without PO2 but there is the likelihood of patients dying with having PO2.
- The visualization of Vital Status by PH shows that there are more patients without a dangerous level of PH and the cases of patients who died was low.

Hide

```
> f13<-ggplot(ICU, aes(x=PCO2, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by PCO2")
>
> f14<-ggplot(ICU, aes(x=Bicarbonate, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Bicarbonate")
>
> f15<-ggplot(ICU, aes(x=Creatinine, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Creatinine")
>
> f16<-ggplot(ICU, aes(x=Consciousness, fill = Status)) +
+  theme_bw() +
+  geom_bar() +
+  labs(y = "Patient Count",
+       title = "Vital Status by Consciousness")
>
>
> library(Rmisc)
> multiplot(f13, f14, f15, f16, layout=matrix(c(1:4), nrow=2, byrow=TRUE))
```

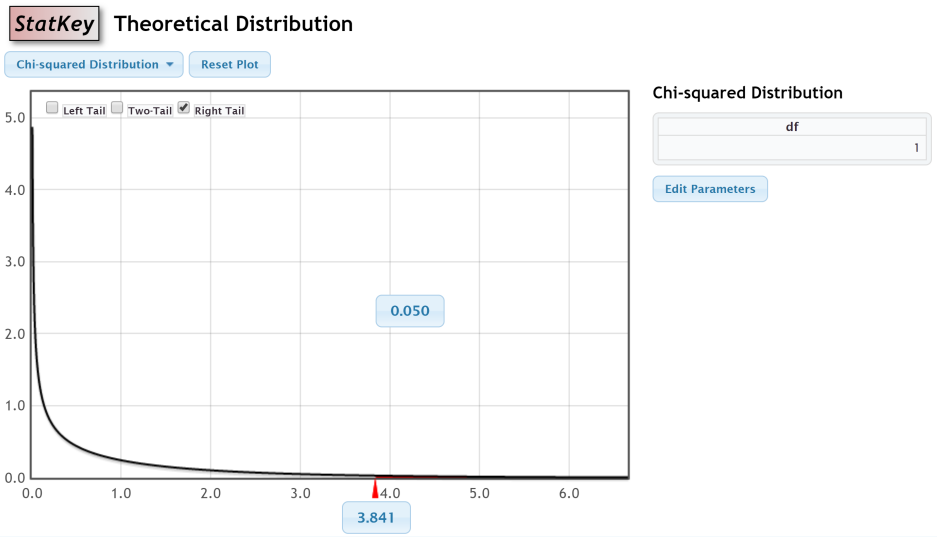


Observations

The most visually prominent chart from the last graph collection is the relationship between vital status and Consciousness which shows that although there is a high count of patients who were conscious at admission, however most patients that were received at a stage of Deep Stupor or Coma lost consciousness and died.

7 Uniform Distribution

The distribution of the ICU attributes will be explored below. Many of the attributes chosen to be tested here relate to the history of the individuals in the data set or the circumstances of the admission, rather than lab tests on admission. Most of the categorical attributes in this dataset have two categories, making the degrees of freedom 1 for chi squared distribution.



Theoretical chi square distribution with 1 degrees of freedom

For Categorical variables with two categories:
Ho: the attribute is distributed uniformally, the chi squared value does not exceed the hypothesized value of 3.841 (p = 0.05)
Ha: the attribute is not distiributed uniformally, The chi squared value exceeds the hypothesized value of 3.841

Chi squared test - Status:

The chi squared test for Status returns a chi squared value above the null hypothesized value with a very small p value (<0.05), indicating that there is a very low risk of Type I error if the null hypothesis is rejected. It can be concluded that Status is not uniformally distributed and that statistically more individuals lived than died on admission to the ICU.

Hide

```
> local({
+   .Table <- with(ICU, table(Status))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Status
Lived  Died
  160    40

percentages:
Status
Lived  Died
   80    20
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 72, df = 1, p-value < 2.2e-16
```

Chi squared test - CPR:

Similarly, the chi squared test for CPR returns a very high chi squared value at 151, and a low p value, allowing us to reject the null hypothesis that this attribute is uniformally distributed. More individuals did not have CPR than did before ICU admission. In fact, as seen from the table below, few individuals received CPR at all, which might make it a poor predictor of ICU outcomes.

Hide

```
> local({
+   .Table <- with(ICU, table(CPR))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
CPR
 No Yes
187  13

percentages:
CPR
  No  Yes
93.5  6.5
```

```
    Chi-squared test for given probabilities

data:  .Table
X-squared = 151.38, df = 1, p-value < 2.2e-16
```

Hide

```
> table(ICU$CPR, ICU$Status)
```

```
     Lived Died
  No    154   33
  Yes     6    7
```

Chi squared test - Infection:

The chi squared value produced from this test is only slightly higher than the null hypothesized value, with a p value of 0.024, indicating a 2.4% chance of a Type I error if the null hypothesis is rejected. While the distribution can be concluded to not be uniform based on the threshold set with the null hypothesis, the uniformity of Infection among patients admitted to the ICU could be explored in further research. It should be noted that in the group of individuals who died, more had infections than not.

Hide

```
> local({
+   .Table <- with(ICU, table(Infection))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```
counts:
Infection
 No Yes
116  84

percentages:
Infection
 No Yes
 58  42
```

```
    Chi-squared test for given probabilities

data:  .Table
X-squared = 5.12, df = 1, p-value = 0.02365
```

Hide

```
> table(ICU$Infection, ICU$Status)
```

```
     Lived Died
  No    100   16
  Yes    60   24
```

Chi squared test - Previous:

The chi squared value here exceeds the null hypothesized value with a p value much smaller than 0.05, allowing us to reject the null hypothesis that this attribute is uniformally distributed. More individuals in both status categories did not have a previous ICU admission.

Hide

```
> local({
+   .Table <- with(ICU, table(Previous))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Previous
 No Yes
170  30

percentages:
Previous
 No Yes
 85  15
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 98, df = 1, p-value < 2.2e-16
```

Chi squared test - Sex:

The chi squared value returned from this test is 11.52, which is not as far from the null hypothesized value as some of the other values generated from other tests. The value and p value of 0.00069 still indicated that the null hypothesis should be reject and that Sex is not uniformally distributed. There are more men in this dataset than females.

Hide

```
> local({
+   .Table <- with(ICU, table(Sex))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Sex
  Male Female
   124     76

percentages:
Sex
  Male Female
    62     38
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 11.52, df = 1, p-value = 0.0006885
```

In both sex categories, despite the inequality in the number of observations in each group, there are still more individuals who lived than died.

Chi squared test - Type:

The distribution of Type can be concluded to be non-uniform as the null hypothesis should be rejected based on the chi squared value produced and low p value. As seen in exploratory data analysis, the there are more individuals who are had emergency admissions than elective admissions. This difference in distribution is statistically significant. Fewer individuals died when they were admitted to the ICU on an elective basis.

Hide

```
> local({
+   .Table <- with(ICU, table(Type))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Type
 Elective Emergency
       53       147

percentages:
Type
 Elective Emergency
     26.5      73.5
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 44.18, df = 1, p-value = 2.995e-11
```

Chi squared test - Service:

The distribution of Service is uniform based on the null hypothesis of a chi squared variable of 3.841. The chi squared value here indicates that there is not a significant difference between a uniform distribution and the distributio not Service. This means that there was no significant difference between the number of individuals admitted who had medical or surgical services performed, as the p value indicates a 32% risk of incorrectly rejecting the null hypothesis.

Hide

```
> local({
+   .Table <- with(ICU, table(Service))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Service
 Medical Surgical
      93      107

percentages:
Service
 Medical Surgical
    46.5     53.5
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 0.98, df = 1, p-value = 0.3222
```

Chi squared test- Renal:

The distribution of renal categories was not uniform, with significantly larger difference from the expected as shown by the chi squared value of 131 which is much larger than the null hypothesized chi squared value. The low p value indicated that it is safe to reject the null hypothesis without concern for Type I error. It can thus be concluded that significantly more people did not have chronic renal failure prior to ICU admission.

Hide

```
> local({
+   .Table <- with(ICU, table(Renal))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.5,0.5)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Renal
 No Yes
181  19

percentages:
Renal
  No  Yes
90.5  9.5
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 131.22, df = 1, p-value < 2.2e-16
```

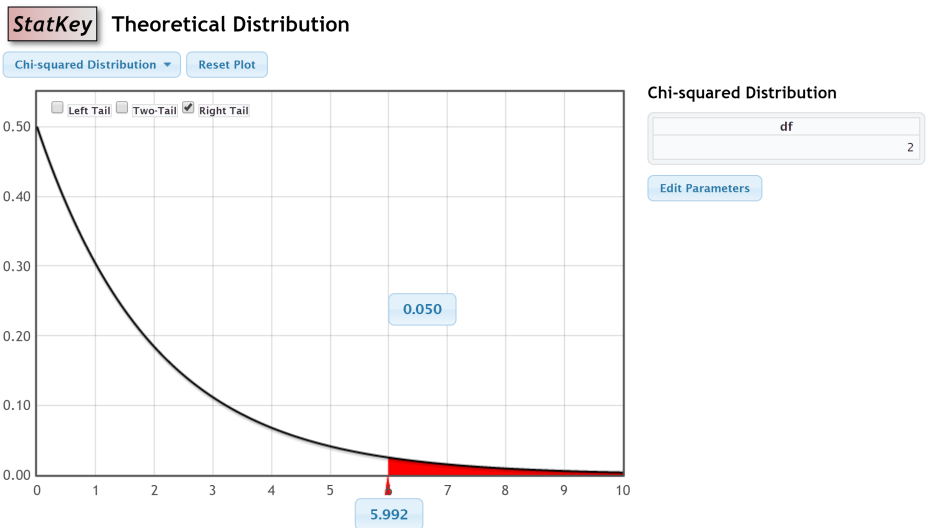
Additionally, more individuals in both status groups did not have chronic renal failure, indicating that chronic renal failure might not be a factor in worse outcomes during ICU admissions.

Hide

```
> table(ICU$Renal, ICU$Status)
```

```
     Lived Died
  No    149   32
  Yes    11    8
```

For Categorical variables with three categories:
Ho: the attribute is distributed uniformally, the chi squared value does not exceed the hypothesized value of 5.992 (p = 0.05)
Ha: the attribute is not distiributed uniformally, The chi squared value exceeds the hypothesized value of 5.992 (p=0.05)



Theoretical chi square distribution with 2 degrees of freedom

Chi squared test - Consciousness:

The null hypothesis for this attribute's distribution is that an equal number of observations will be found in each category of Consciousness. The chi squared value produced by this test exceeds the null hypothesized value with a low p value, indicating that the observations in this category are statistically not uniform. This could have been guessed from counts of observations in each category, where most individuals were in the conscious category.

Hide

```
> local({
+   .Table <- with(ICU,
+   table(Consciousness))
+   cat("\ncounts:\n")
+   print(.Table)
+   cat("\npercentages:\n")
+   print(round(100*.Table/sum(.Table),
+   2))
+   .Probs <- c(0.333333333333333,
+   0.333333333333333,0.333333333333333)
+   chisq.test(.Table, p=.Probs)
+ })
```

```

counts:
Consciousness
  Conscious Deep Stupor        Coma
        185           5          10

percentages:
Consciousness
  Conscious Deep Stupor        Coma
       92.5         2.5         5.0
```

```

    Chi-squared test for given probabilities

data:  .Table
X-squared = 315.25, df = 2, p-value < 2.2e-16
```

8 Normal Distribution

Distribution of the three numerical variables in the ICU dataset was evaluated using quantile-quantile plots and the Shapiro-Wilk's test for normality. Confidence intervals will be produced for the means of each attribute.

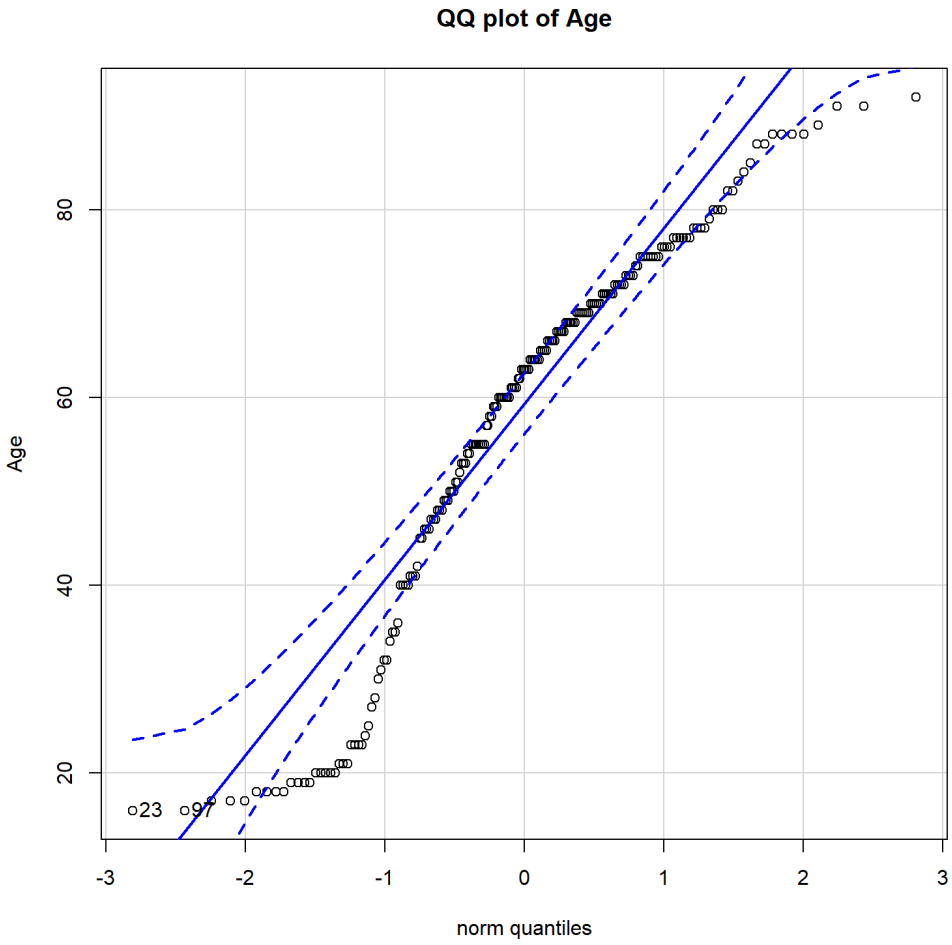
Ho: Age, HeartRate and Systolic are normally distributed
Ha: Age, HeartRate and Systolic are not normally distributed

Tests of normality - Age:

Recall that a quantile-quantile plot should produce a nearly linear plot using dataset values, with the intercept going through zero, if the null hypothesis is satisfied. Below, the plotted points of Age do not conform well to the line of best fit, and are frequently outside of the confidence bands. This supports what might have been hypothesized when the histogram of Age was produced: this distribution is not clearly centered around a mean. This plot suggests that the null hypothesis that Age is normally distributed should be rejected.

Hide

```
> with(ICU, qqPlot(Age, dist="norm", id=list(method="y", n=2, labels=rownames(ICU)), main="QQ plot of Age"))
```




```
[1] 23 97
```

Using the Shapiro-Wilk normality test, the conclusions drawn from the QQ plot can be supported, with a very low risk of Type I error, base on this p value. The Shapiro Wilk test is sensitive when used on larger datasets, so this result should be taken into account with the other tests used.

Hide

```
> normalityTest(~Age, test="shapiro.test", data=ICU)
```

```
    Shapiro-Wilk normality test

data:  Age
W = 0.92836, p-value = 2.507e-08
```

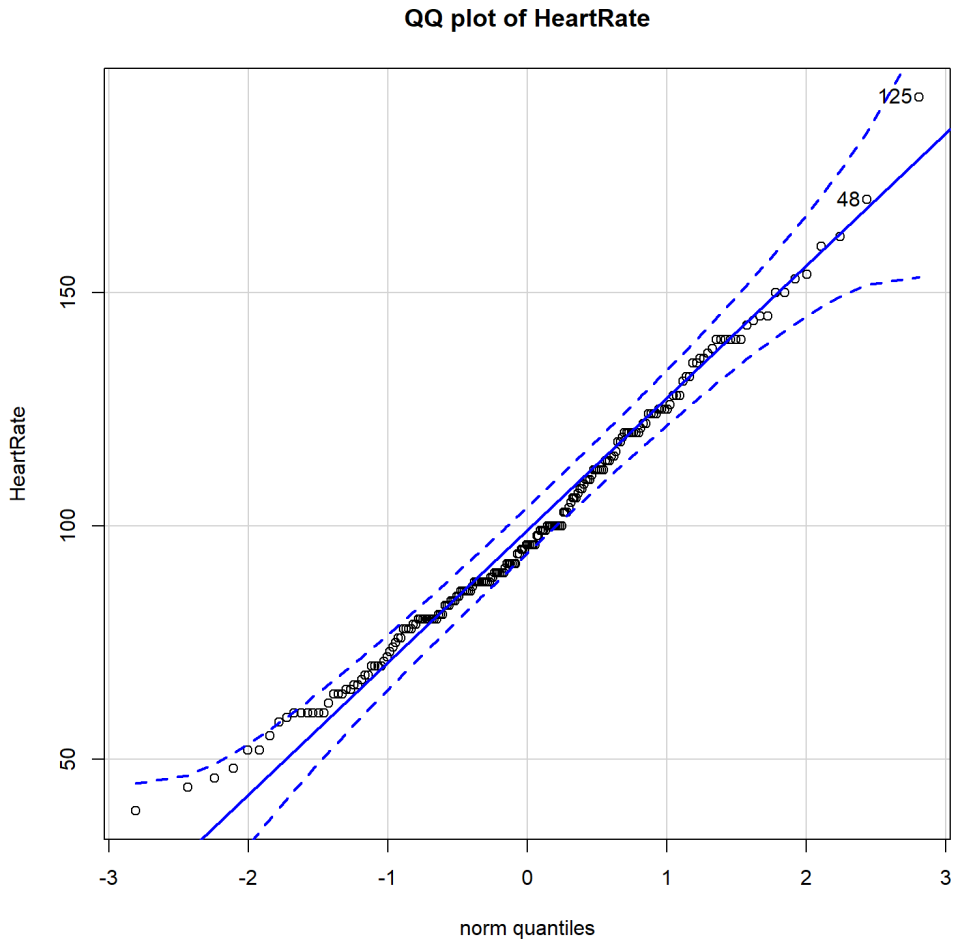
Using the t-test, given the confidence interval for this attribute is between 54.75 and 60.34, an interval that does not contain 0. (choosing not to add this as I do not have a test mean - need to refresh on what a true mean is)

Tests of normality - HeartRate:

The QQ plot of HeartRate appears to adhere well to the line of best fit, despite several points in the middle of the graph lying along one side of the confidence band. Most of the points here fall within the confidence band. There is a positive skew of this data towards the lower values of HeartRate. Visually, one might conclude that this data is normally distributed.

Hide

```
> with(ICU, qqPlot(HeartRate, dist="norm", id=list(method="y", n=2, labels=rownames(ICU)), main="QQ plot of HeartRate"))
```




```
[1] 125  48
```

The results of the Shapiro-Wilk test very narrowly allow one to reject the null hypothesis of normal distribution. However the p value is very close to the threshold p value of 0.05, which decreases the confidence that might be had to reject normality based on this test and the sample size used.

Hide

```
> normalityTest(~HeartRate, test="shapiro.test", data=ICU)
```

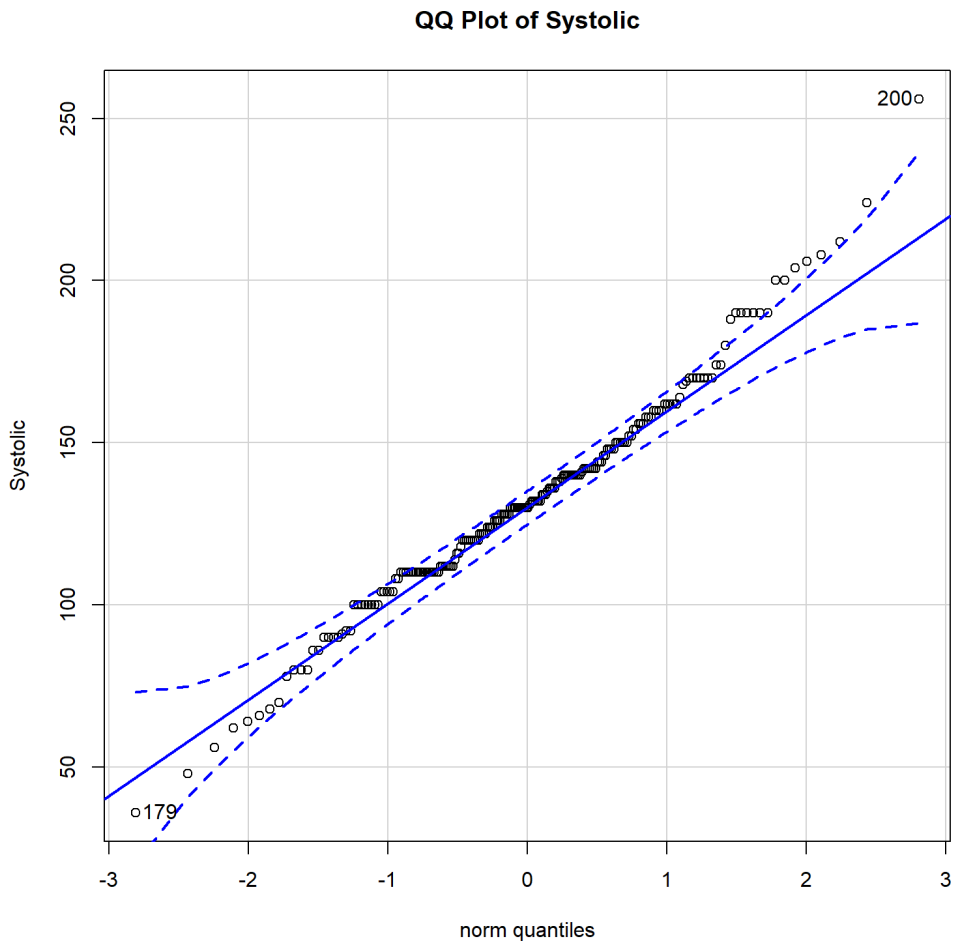
```

    Shapiro-Wilk normality test

data:  HeartRate
W = 0.98598, p-value = 0.04478
```

Test of normality - Systolic:

Similar to the plot of HeartRate, the QQ plot of Systolic displays points that line closely on the line of best fit. The confidence bands here are quite narrow, and it is only along the ends oof the plot the points very clearly exit the confidence bands. One might conclude that the null hypothesis could be rejected for the normal distribution of Systolic based on this plot. There is a very slight negative skew here towards higher values of Systolic.

Hide

```
> with(ICU, qqPlot(Systolic, dist="norm", id=list(method="y", n=2, labels=rownames(ICU)), main="QQ Plot of Systolic"))
```




```
[1] 200 179
```

While more statistically significant compared to the Shapiro Wilk test for normality of the HeartRate distribution, the p value here is not far from 0.05, indicating a 2% risk of Type I error. Based on this test, the null hypothesis could be rejected. The distribution of Systolic appears to be non-normal.

Hide

```
> normalityTest(~Systolic, test="shapiro.test", data=ICU)
```

```

    Shapiro-Wilk normality test

data:  Systolic
W = 0.98369, p-value = 0.0204
```

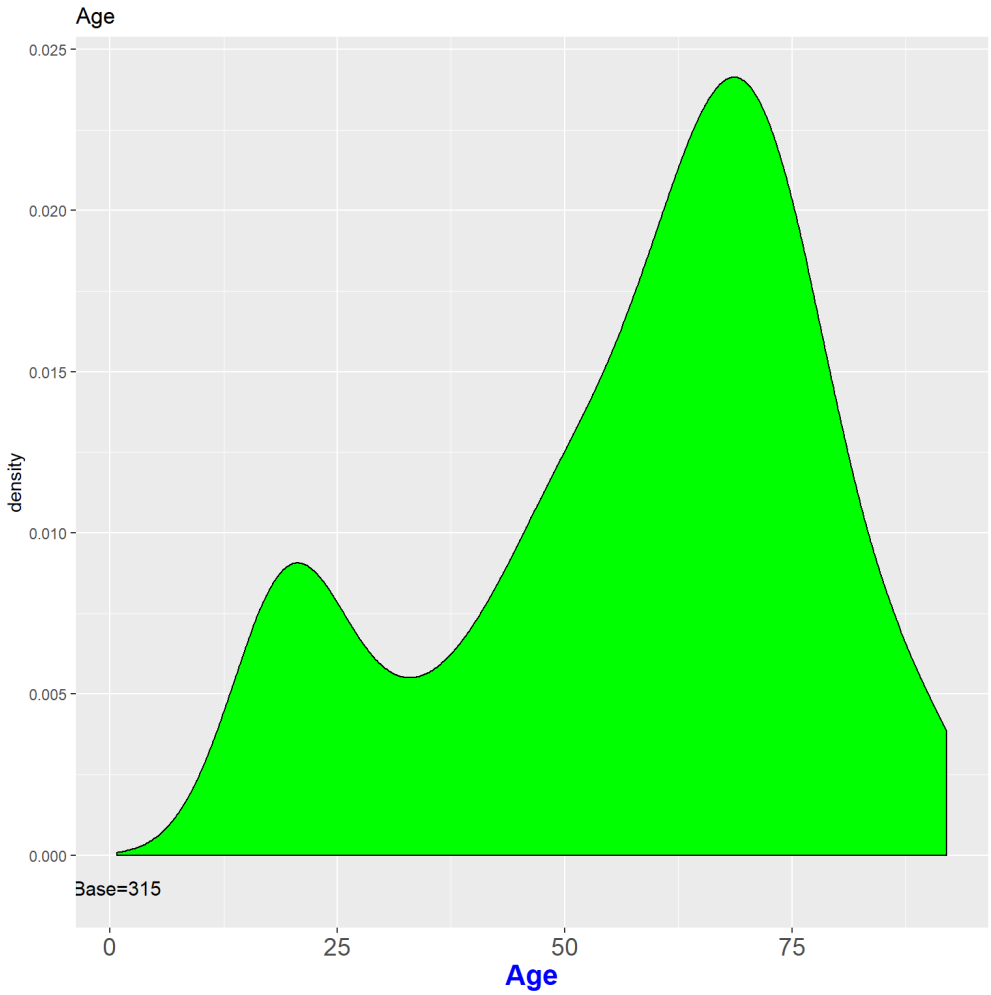
In reference to Ian Fellows' comments about tests for normality, these tests for normality do not add much more to understanding of the ICU dataset and the contributions of these three numerical variables to the determination of Status, beyond what is provided by the xray package. In the same vein, as the central limit theorem and bootstrapping have shown, most distributions can become normal with repeated sampling.

9 Relationships with Age

The mean of Age, referring to the descriptive statistics for this attributes, is about 58 years. Visually, it can be seen that the distribution of Age has two peaks, one around 20 years of age and one around 70 years of age. The majority of subjects admitted to the ICU seem to be between 40-80 years old.

Hide

```
> ggplot(ICU, aes(x=Age)) +
+  geom_density(fill="green") +
+  ggtitle("Age") +
+  theme(axis.title.x=element_text(size=16, face="bold", colour="blue")) +
+  theme(axis.text.x=element_text(size=14 )) +
+  annotate("text", x=0.8, y=-0.001, label="Base=315",  size=4)
```



Critical Values for Difference in Means

Critical values to assess differences in means are based on a degrees of freedom of 199.

Hide

```
> qt(c(0.025), df=199, lower.tail=TRUE)
```

```
[1] -1.971957
```

Hide

```
> qt(c(0.025), df=199, lower.tail=FALSE)
```

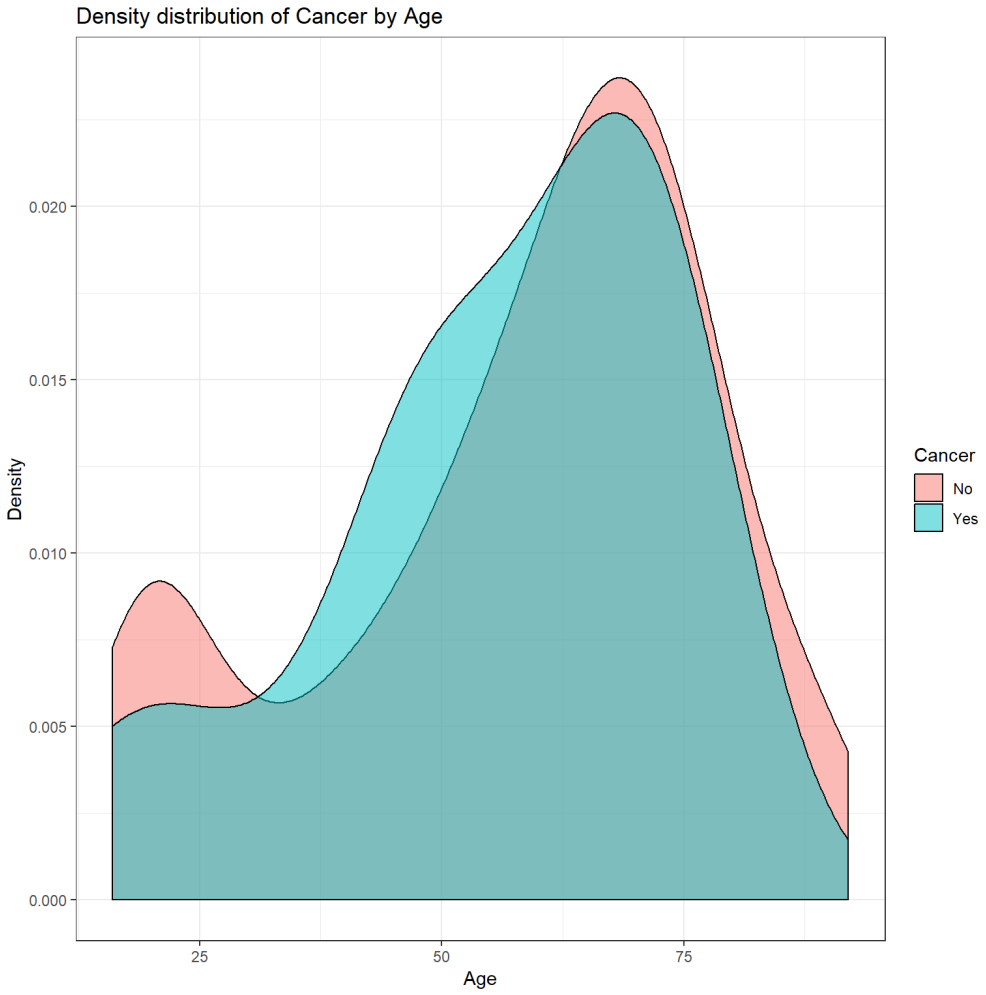
```
[1] 1.971957
```

9.1 Cancer by Age

There are three peaks for Ages where ICU admissions involved cancer. These are around 20, 50 and 70 years of age. Given the relatively small number of individuals who had cancer involvement with their admission, and the variability in ages associated, cancer might not be the best predictor of ICU admission.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Cancer)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Cancer by Age")
```



Difference in Means

Ho: the variances for admissions involving cancer and admission not involving cancer are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Cancer.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Cancer, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
        mean       sd se(mean) Age:n
No  57.70556 20.30953 1.513783   180
Yes 56.10000 17.99971 4.024857    20
```

Since the p-value = 0.5684 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for admissions involving cancer vs those that do not are equal. As such, the Student t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Cancer, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1  0.3265 0.5684
      198
```

As the p-value = 0.735, 0 is within the confidence intervals of -7.736825 to 10.947937 and t = 0.33891 is less than 1.971957, we retain the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for age are the same between the two groups.

Hide

```
> t.test(Age~Cancer, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

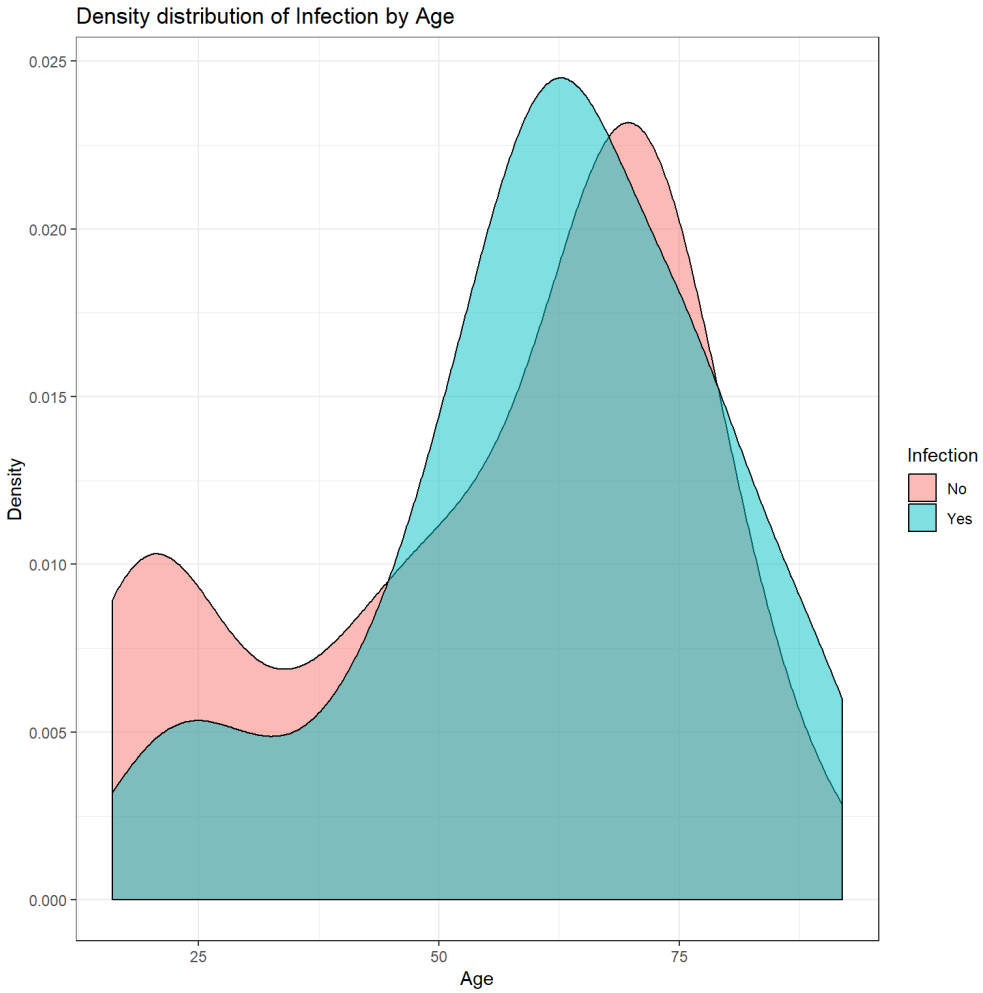
```
    Two Sample t-test

data:  Age by Cancer
t = 0.33891, df = 198, p-value = 0.735
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.736825 10.947937
sample estimates:
 mean in group No mean in group Yes
         57.70556          56.10000
```

9.2 Infection by Age

ICU admissions that involved infection centered around 60-65 years of age. This association is relatively clear visually looking at the plot below.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Infection )) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Infection by Age")
```



Difference in Means

Ho: the variances for admissions involving infection and those that do not are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Infection.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Infection, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1
))
```

```
        mean       sd se(mean) Age:n
No  54.93103 21.21842 1.970081   116
Yes 61.15476 17.82545 1.944917    84
```

Since the p-value = 0.04162 for testing the homogeneity of variances is less than 0.05, we reject the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for admissions involving infection and admissions not involving infection are not equal. As such, the Welch two sample t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Infection, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value  Pr(>F)
group   1  4.2053 0.04162 *
      198
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value = 0.02569, 0 is not within the confidence intervals of -11.6837808 to -0.7636741 and t = -2.2481 is less than -1.971957, we reject the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for age are not the same between the two groups. this suggests that there may be a relationship between Age and Infection.

Hide

```
> t.test(Age~Infection, alternative="two.sided", conf.level=.95, var.equal=FALSE, data=ICU)
```

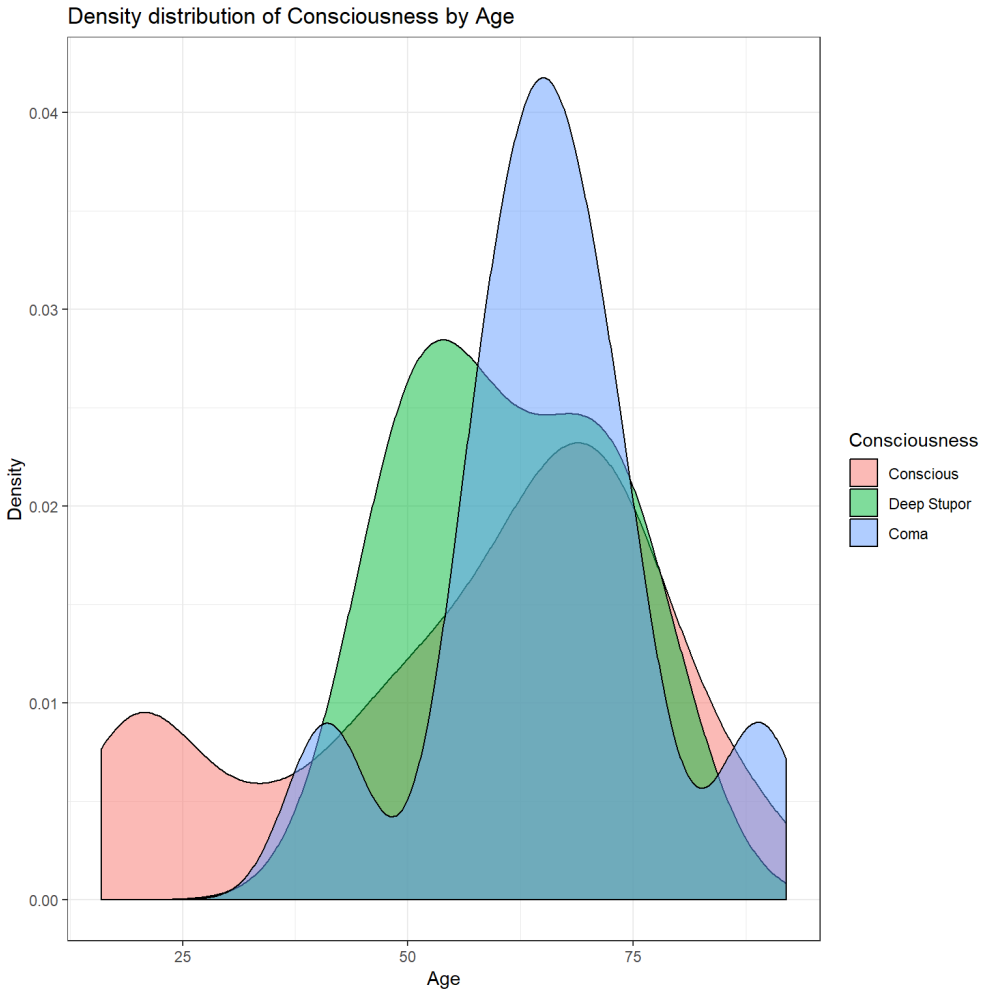
```

    Welch Two Sample t-test

data:  Age by Infection
t = -2.2481, df = 193.6, p-value = 0.02569
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.6837808  -0.7636741
sample estimates:
 mean in group No mean in group Yes
         54.93103          61.15476
```

9.3 Consciousness by Age

Another association that was explored was between Age and Consiousness. All three categories for Consciousness overlap with peaks between 50 and 75 years of age. There is a distinct peak at around 50 for deep stupor, indicating that this age might be associated with deep stupor when related to ICU admissions. However, deep stupor also has a secondary peak at around 75, which make it less clearly how this consciousness category relates to age in this dataset.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Consciousness )) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Consciousness by Age")
```



Difference in Means (ANOVA)

Ho: the variances for Age is the same across all levels of consciousness
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Consciousness.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Consciousness, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

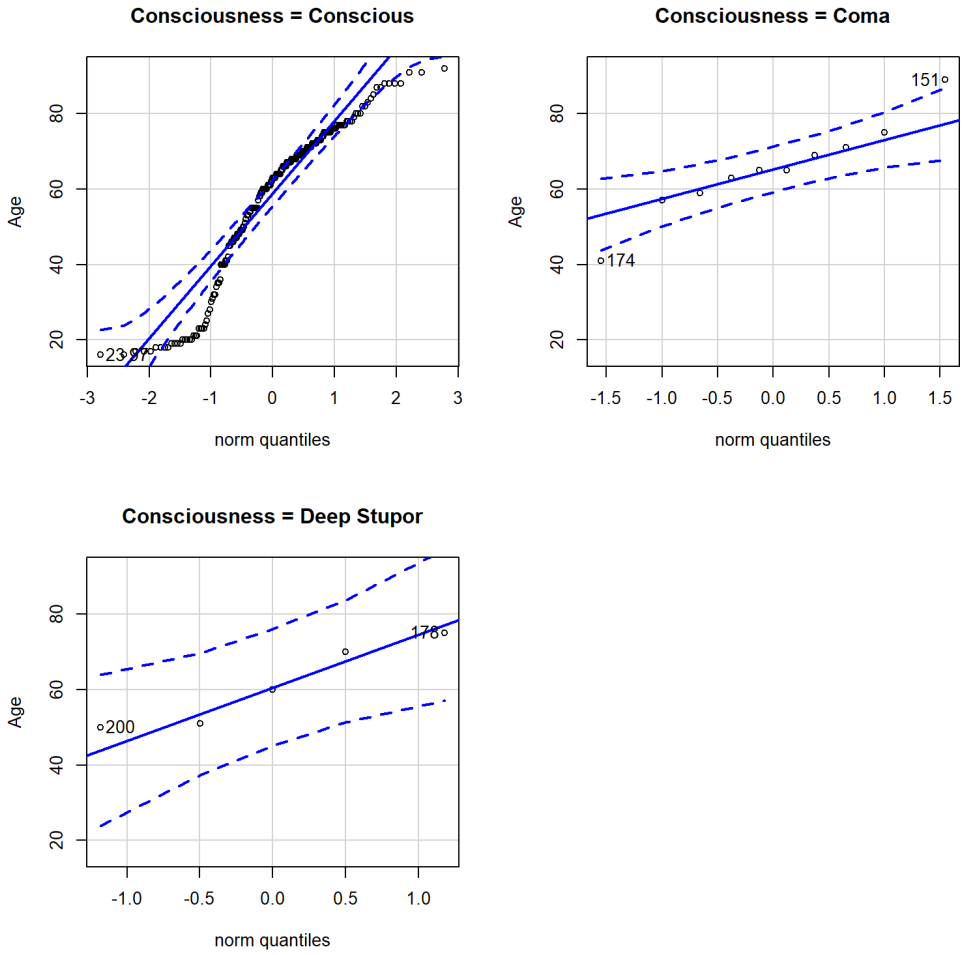
```
               mean       sd se(mean) Age:n
Conscious   57.02162 20.50927 1.507871   185
Deep Stupor 61.20000 11.16692 4.993996     5
Coma        65.40000 12.50067 3.953058    10
```

Testing normality
Ho: the several distributions are normally distributed
Ha: the distributions are not normally distributed

According the q-q plot and Shapiro-Wilk normality test for each of the three levels of consciousness, Age does not appear to be normally distributed for the level of consciousness = conscious. This was found as the q-q plot for conscious does not follow the diagonal line and the p-value = 5.319e-08 which is less than 0.05. Additionally, based on the normality test previously performed on Age, the p-value = 2.507e-08 which is less than 0.05. As such, we would reject the null hypothesis at a 5% risk and conclude that the distributions are not normally distributed. That being said, since normality was rejected, the Kruskal-Wallis test should be used to determine whether the four distributions are the same.

Hide

```
> with(ICU, qqPlot(Age, dist="norm", id=list(method="y", n=2,
+   labels=rownames(ICU)), groups=Consciousness))
```



Hide

```
> normalityTest(Age ~ Consciousness, test="shapiro.test", data=ICU)
```

```
 --------
 Consciousness = Conscious

    Shapiro-Wilk normality test

data:  Age
W = 0.92706, p-value = 5.319e-08

 --------
 Consciousness = Deep Stupor

    Shapiro-Wilk normality test

data:  Age
W = 0.9012, p-value = 0.4165

 --------
 Consciousness = Coma

    Shapiro-Wilk normality test

data:  Age
W = 0.96128, p-value = 0.8003

 --------

 p-values adjusted by the Holm method:
            unadjusted adjusted
Conscious   5.3186e-08 1.5956e-07
Deep Stupor 0.41654    0.83308
Coma        0.80031    0.83308
```

Testing for Homogeneity of Variances
Ho: variances are homogenous
Ha: at least one of the variances is different from the others

As the Levene Test generated a p-value = 0.0942, we would retain the null hypothesis with a 5% risk of a Type 1 error. This means that the difference in variances of Age for the different levels of consciousness is zero, and that the variances are homogeneous.

Hide

```
> leveneTest(Age ~ Consciousness, data=ICU , center=median)
```

```
Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   2  2.3909 0.0942 .
      197
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Non-parametric Kruskal-Wallis test
As 'normality' was rejected, the Kruskal-Wallis test will be used in order to determine whether the three distributions for Age are the same.

The Kruskal-Wallis test produced a p-value of p-value = 0.6468, which is greater than 0.05. As such, we would retain the null hypothesis and conclude that the distributions are the same and that the means are likely the same.

Hide

```
> kruskal.test(Age ~ Consciousness, data=ICU)
```

```
    Kruskal-Wallis rank sum test

data:  Age by Consciousness
Kruskal-Wallis chi-squared = 0.87135, df = 2, p-value = 0.6468
```

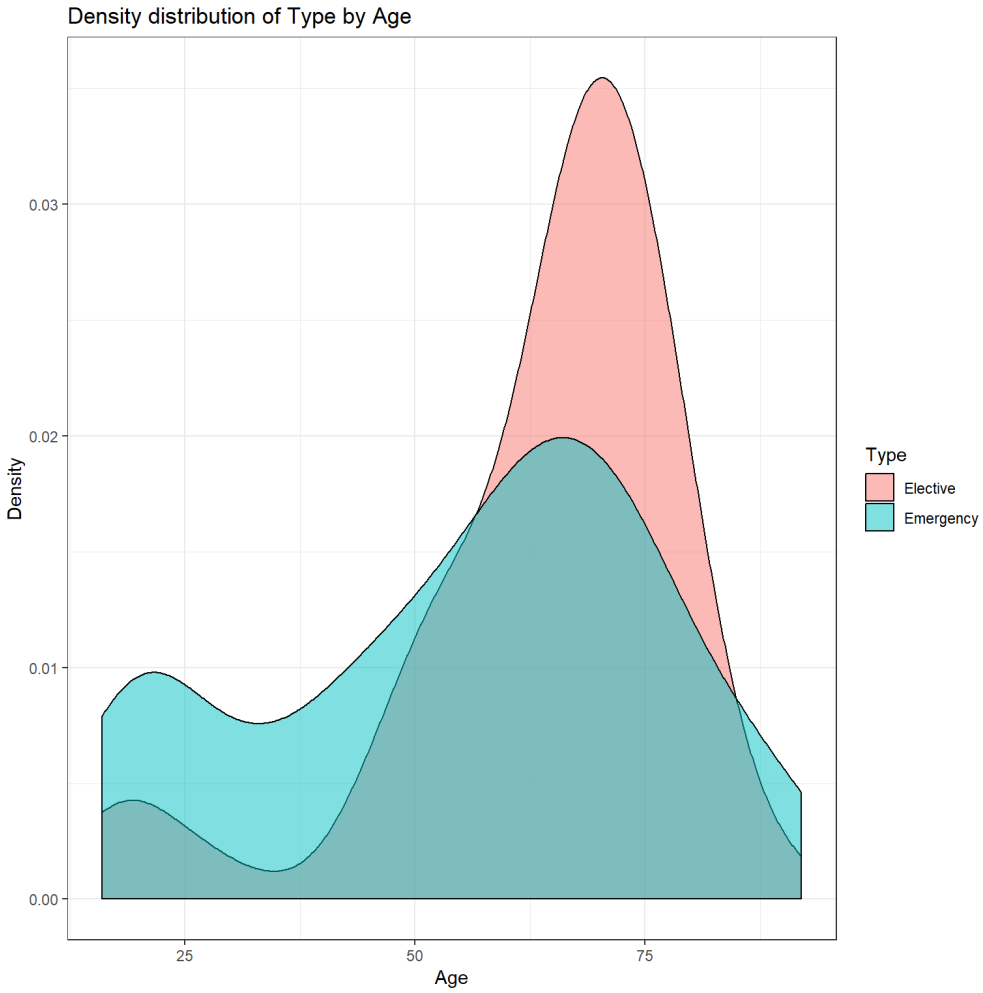
9.4 Type by Age

Both admission Types, elective and emergency, happen more frequently with advancing age. However, elective admissions happen more frequently with older age. This might suggest that older individuals are more likely to have complications during surgical or other procedures that would require admissions to an intensive care unit. Depending on organizational policy or practice, this might also mean that individuals at the study site were

premptively admitted to ICU, perhaps more frequently with advancing age, for fear of complications.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Type )) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Type by Age")
```



Difference in Means

Ho: the variances for elective and emergency are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Type.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Type, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
             mean       sd se(mean) Age:n
Elective  63.77358 15.49769 2.128772    53
Emergency 55.29932 21.05909 1.736924   147
```

Since the p-value = 0.001188 for testing the homogeneity of variances is less than 0.05, we reject the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for elective and emergency admission types are not equal. As such, the Welch two sample t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Type, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value   Pr(>F)
group   1  10.821 0.001188 **
      198
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value = 0.002513, 0 is not within the confidence intervals of 3.036522 to 13.912009 and t = 3.0844 is greater than 1.971957, we reject the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for age are not the same between the two groups. This further confirms our observations from the graph that there is a difference between the two. Additionally, a relationship may exist between Age and admission type.

Hide

```
> t.test(Age~Type, alternative="two.sided", conf.level=.95, var.equal=FALSE, data=ICU)
```

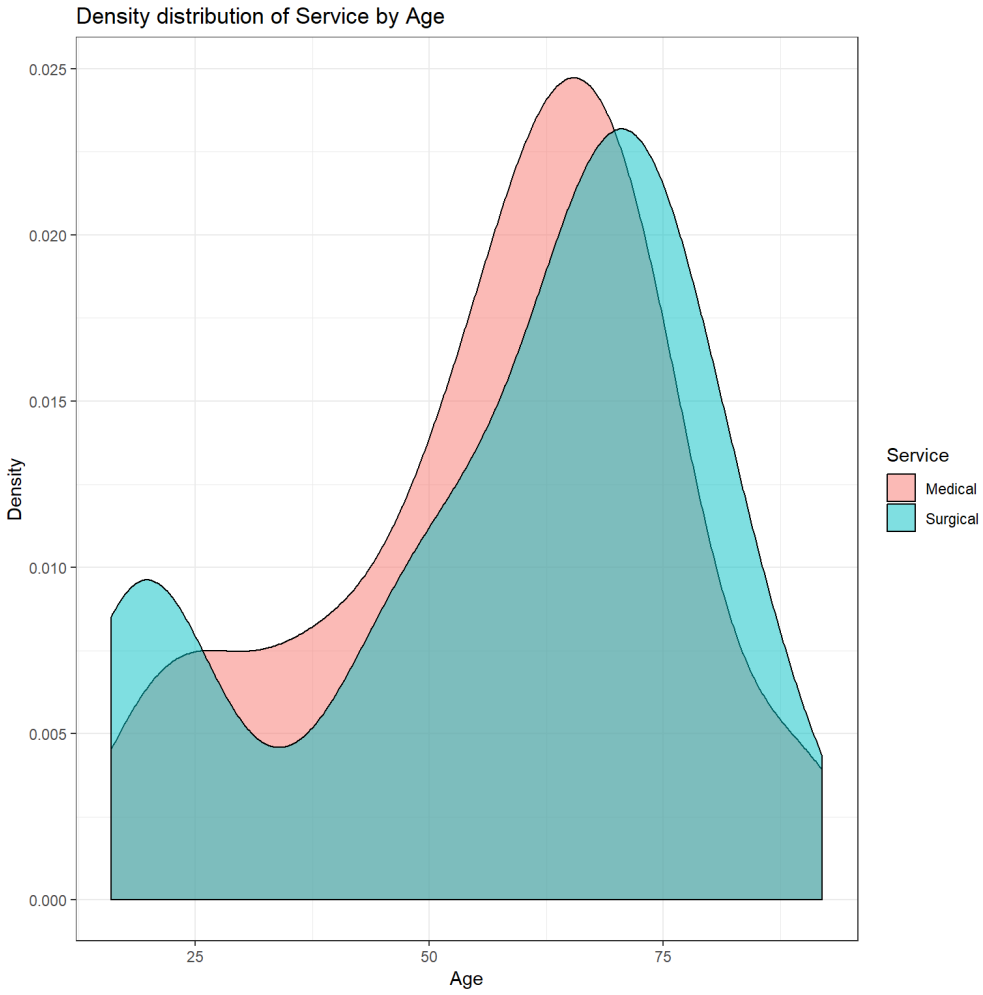
```

    Welch Two Sample t-test

data:  Age by Type
t = 3.0844, df = 124.61, p-value = 0.002513
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  3.036522 13.912009
sample estimates:
 mean in group Elective mean in group Emergency
               63.77358                55.29932
```

9.5 Service by Age

While the peaks for surgical and medical Service by Age overlap considerably, there is interestingly a peak of surgical service type in younger age groups, around 20 years of age. While it is not known what conditions led to these individuals being admitted, it might be concluded that older individuals would have more events that were related to their medication or needing medication, for example falls or stroke. This is a very weak conclusion to draw from the graph below, but if specific reasons for admission were also collected, this theory could be explored.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Service)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Service by Age")
```



Difference in Means

Ho: the variances for medical and surgical are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Service.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Service, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
            mean       sd se(mean) Age:n
Medical  57.45161 18.56511 1.925112    93
Surgical 57.62617 21.35174 2.064150   107
```

Since the p-value = 0.2151 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for medical and surgical are equal. As such, the Student two t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Service, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1  1.5464 0.2151
      198
```

As the p-value = 0.9512, 0 is within the confidence intervals of -5.795344 to 5.446234 and t = -0.061242 is greater than -1.971957, we retain the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for age are the same between the two groups.

Hide

```
> t.test(Age~Service, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

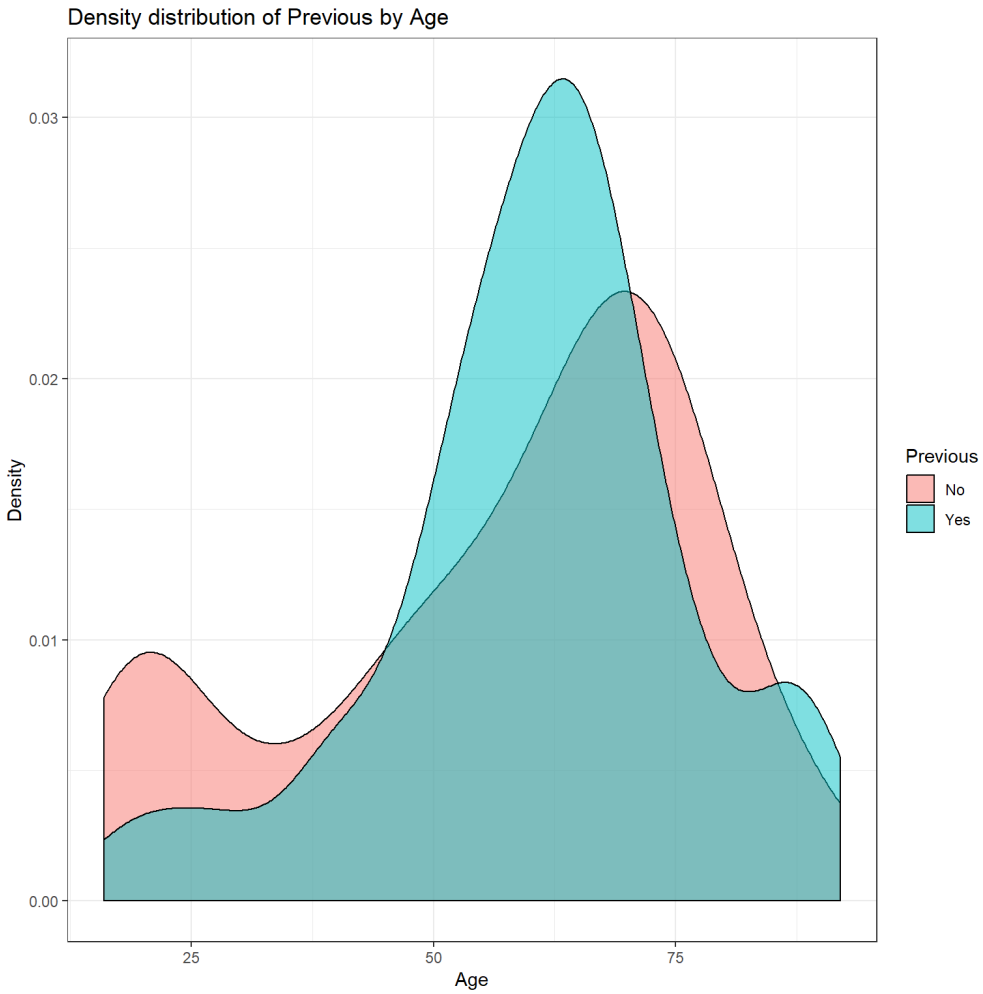
```
    Two Sample t-test

data:  Age by Service
t = -0.061242, df = 198, p-value = 0.9512
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.795344  5.446234
sample estimates:
 mean in group Medical mean in group Surgical
              57.45161               57.62617
```

9.6 Previous by Age

There is a definite an almost independent peak of history of previous ICU admission with older age. It can be noted that there is a smaller peak of relatively older individuals who were admitted who had not had a previous ICU admission within the past 6 months. This indicates that old age does not relate clearly to previous ICU admission and future ICU admission. On the extremes of age, younger individuals are shown to have less frequency of previous ICU admission within the past 6 months, while there is a small peak of older admission of the individuals in this dataset having had a previous ICU admission in the past 6 months. There might be a smaller increased likelihood of ICU admission with previous ICU admission in the past 6 months with age.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Previous )) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Previous by Age")
```



Difference in Means

Ho: the variances for previous admission within the past 6 months and no previous admission within the past 6 months are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances of Age by Previous.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Previous, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1
))
```

```
        mean       sd se(mean) Age:n
No  56.96471 20.69528 1.587256   170
Yes 60.83333 15.83554 2.891161    30
```

Since the p-value = 0.05185 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for previous admission within the past 6 months and no previous admission within the past 6 months are equal. As such, the Student two t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Previous, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value  Pr(>F)
group   1  3.8268 0.05185 .
      198
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value = 0.3312, 0 is within the confidence intervals of -11.701332 to 3.964077 and t = -0.97399 is greater than -1.971957, we retain the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for age are the same between the two groups.

Hide

```
> t.test(Age~Previous, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

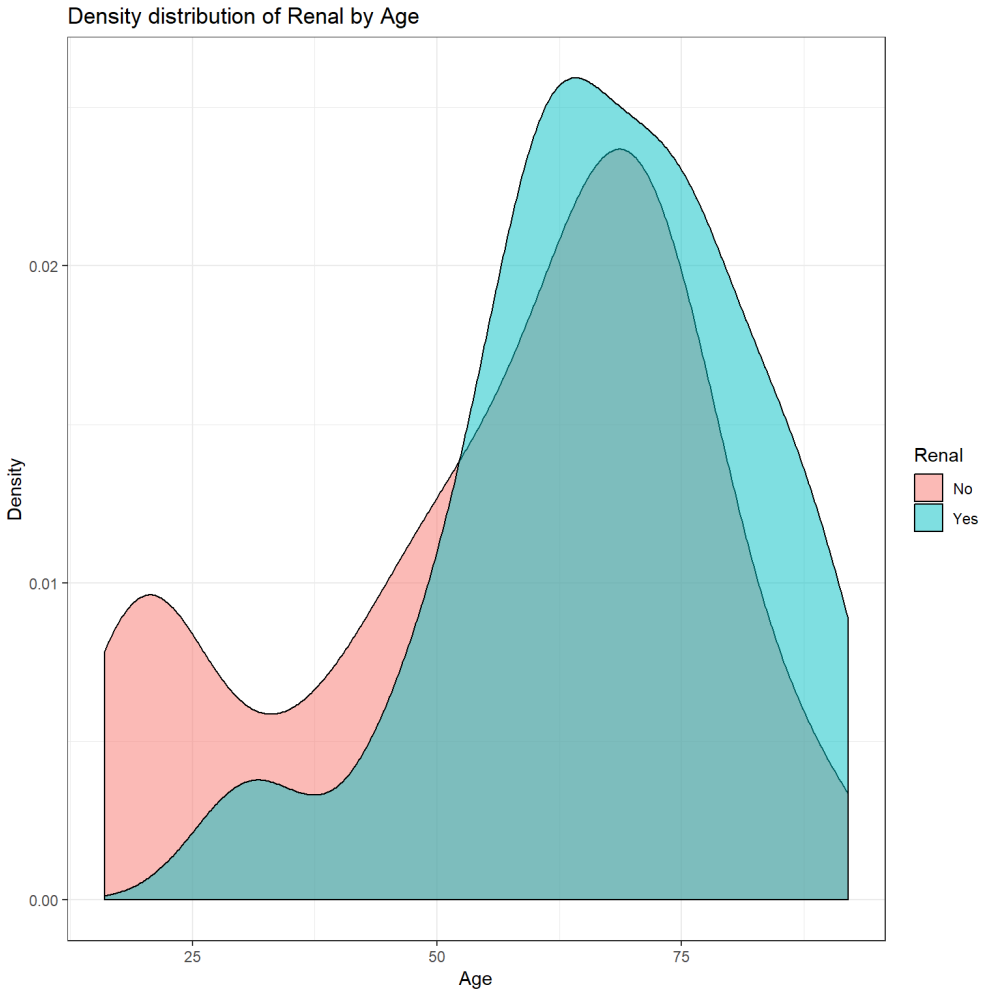
```

    Two Sample t-test

data:  Age by Previous
t = -0.97399, df = 198, p-value = 0.3312
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.701332   3.964077
sample estimates:
 mean in group No mean in group Yes
         56.96471          60.83333
```

9.7 Renal by Age

The plot of density distribution of the two renal categories by Age indicates that both Status groups had a higher frequency of chronic renal failure as age progressed, and peaked at around 70-75 years of age. There is a small peaks at 20 and a smaller peak at about 30 years of age for individuals who did not have chronic renal failure and did have chronic renal failure on admission, respectively. The amount of overlap in the peaks confirms the results of the previous table above indicating that the Renal attribute might not give more insight into ICU outcomes.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Renal )) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Renal by Age")
```



Difference in Means
Ho: the variances for history of chronic renal failure and no history **Ha**: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances between the history of chronic renal failue and no history for Age.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Renal, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
        mean       sd se(mean) Age:n
No  56.53039 20.28315 1.507634   181
Yes 67.21053 14.94649 3.428961    19
```

Since the p-value = 0.1385 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for history and no history are equal. As such, the Student t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Renal, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1  2.2128 0.1385
      198
```

As the p-value = 0.02685, 0 is not within the confidence intervals of -20.123596 to -1.236684 and t = -2.2303 is less than -1.971957, we reject the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for Renal are not the same between the two groups.

Hide

```
> t.test(Age~Renal, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

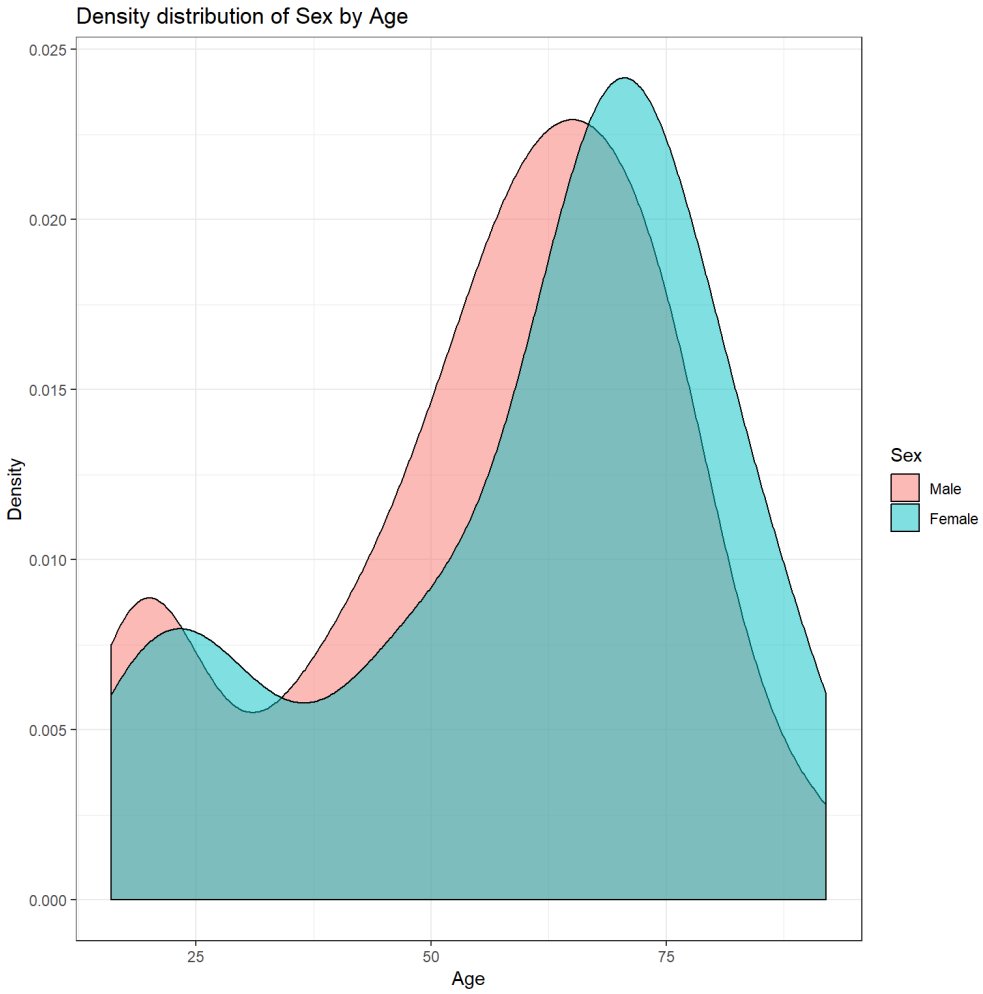
```
    Two Sample t-test

data:  Age by Renal
t = -2.2303, df = 198, p-value = 0.02685
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -20.123596  -1.236684
sample estimates:
 mean in group No mean in group Yes
         56.53039          67.21053
```

9.8 Sex by Age

From the plot below, it seems that females who were admitted to the ICU were older than their male counterparts. There are similar peaks and patterns for younger individuals: females were older if admitted to ICU than their male counterparts.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Sex)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Sex by Age")
```



Difference in Means
Ho: the variances for Males and Females are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances between the two Males and Females for Age.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Sex, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
          mean       sd se(mean) Age:n
Male   56.04032 19.45451 1.747067   124
Female 60.00000 20.89466 2.396781    76
```

Since the p-value = 0.7344 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for Males and Females are equal. As such, the Student t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Sex, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1  0.1154 0.7344
      198
```

As the p-value = 0.1759, 0 is within the confidence intervals of -9.708824 to 1.789469 and t = -1.3582 is greater than -1.971957, we retain the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for Age are the same between the two groups. This further confirms our observations from the graph that the peaks of the two sexes are similar.

Hide

```
> t.test(Age~Sex, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

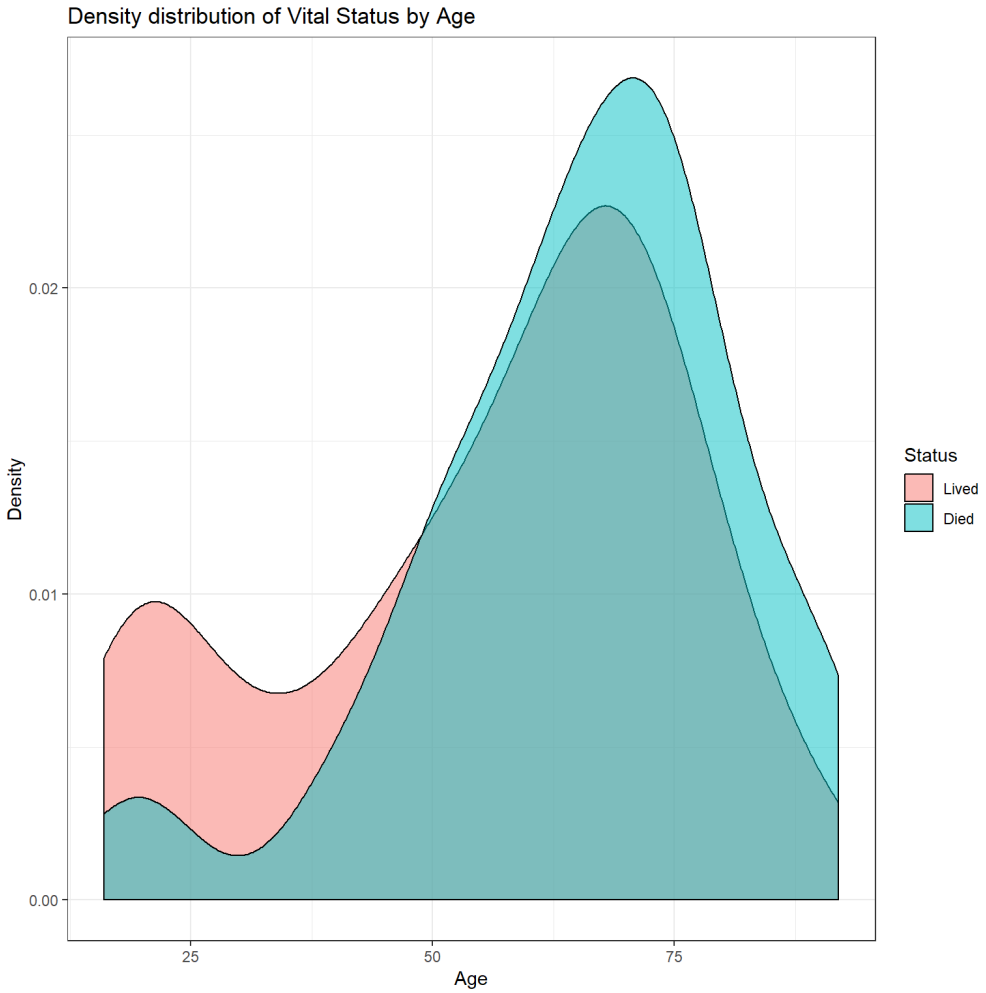
```

    Two Sample t-test

data:  Age by Sex
t = -1.3582, df = 198, p-value = 0.1759
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.708824  1.789469
sample estimates:
  mean in group Male mean in group Female
           56.04032             60.00000
```

9.9 Vital Status by Age

Overlaying Status and Age, density plots show that deaths after admission follow the larger Age peak that contains middle aged to elderly individuals.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Status)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Vital Status by Age")
```

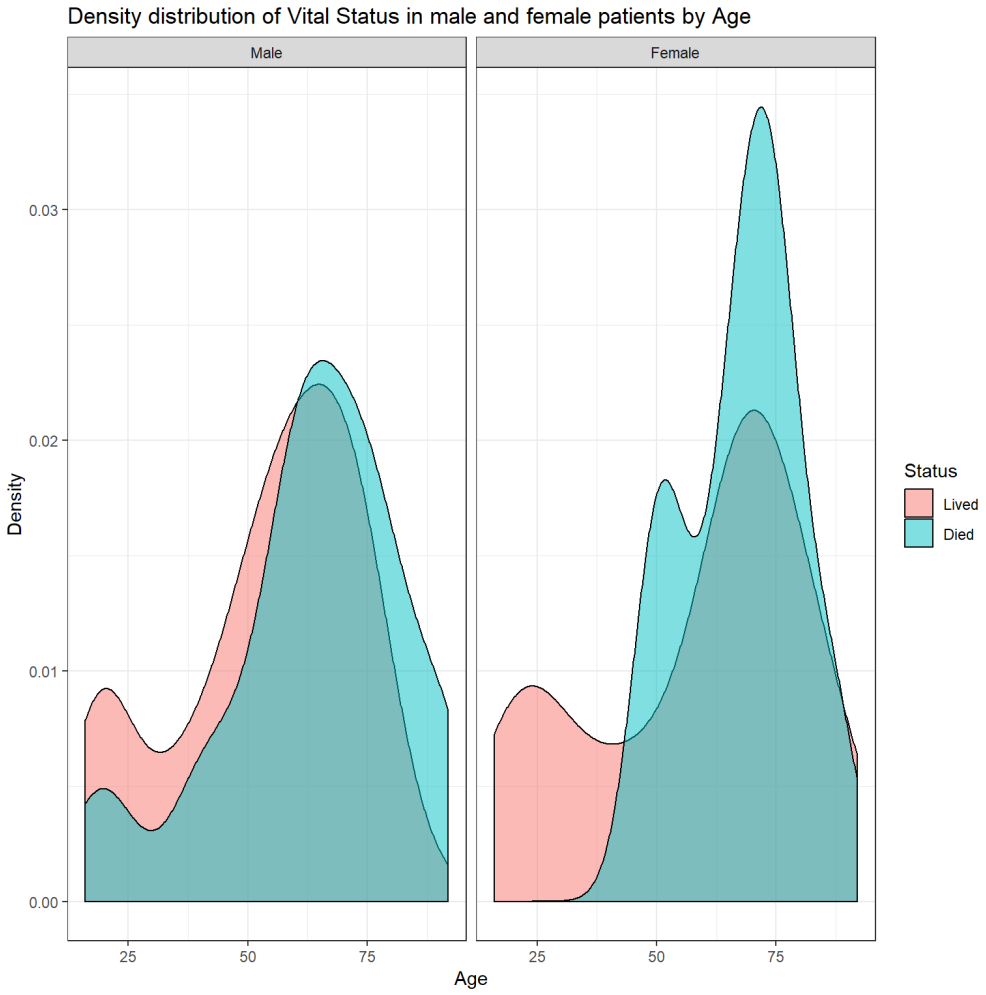


When the additional layer of Status, more information can be gathered from the relationship between Age and Sex. When looking at males, older males more often died than younger males, with similar frequency in advanced age, around the age of 65-70. In younger males, those admitted to the ICU were more often to have lived through their admission, where the opposite is seen in older males.

In females, there is a clearer distribution of older individuals who had died. No females in this dataset under the age of about 40 died on ICU admission. There is a clear peak at around 75 years and a smaller peak at around 50 of females who died on ICU admission. There is a higher frequency of younger females who lived through ICU admission, but there remains a higher frequency of women who were older in this group.

Hide

```
> ggplot(ICU, aes(x=Age, fill = Status)) +
+    theme_bw() +
+    facet_wrap(~ Sex) +
+        geom_density(alpha=0.5) +
+    labs(y = "Density",
+         title = "Density distribution of Vital Status in male and female patients by Age")
```



Difference in Means
Ho: the variances for Lived and Died are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances between the two Status groups for Age.

Hide

```
> numSummary(ICU[,c("Age"), drop=FALSE], groups=ICU$Status, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
       mean       sd se(mean) Age:n
Lived 55.650 20.42818 1.614990   160
Died  65.125 16.64900 2.632438    40
```

Since the p value = 0.07855 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for Lived and Died are equal. As such, the Student t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Age ~ Status, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value  Pr(>F)
group   1   3.127 0.07855 .
      198
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value = 0.007211, 0 is not within the confidence intervals of -16.35688 to -2.59312 and t = -2.7151 is less than -1.971957, we reject the null hypothesis and conclude that the means for Age for the groups Lived and Died are not the same. This suggests that there might be a relationship between Age and Vital Status.

Hide

```
> t.test(Age~Status, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

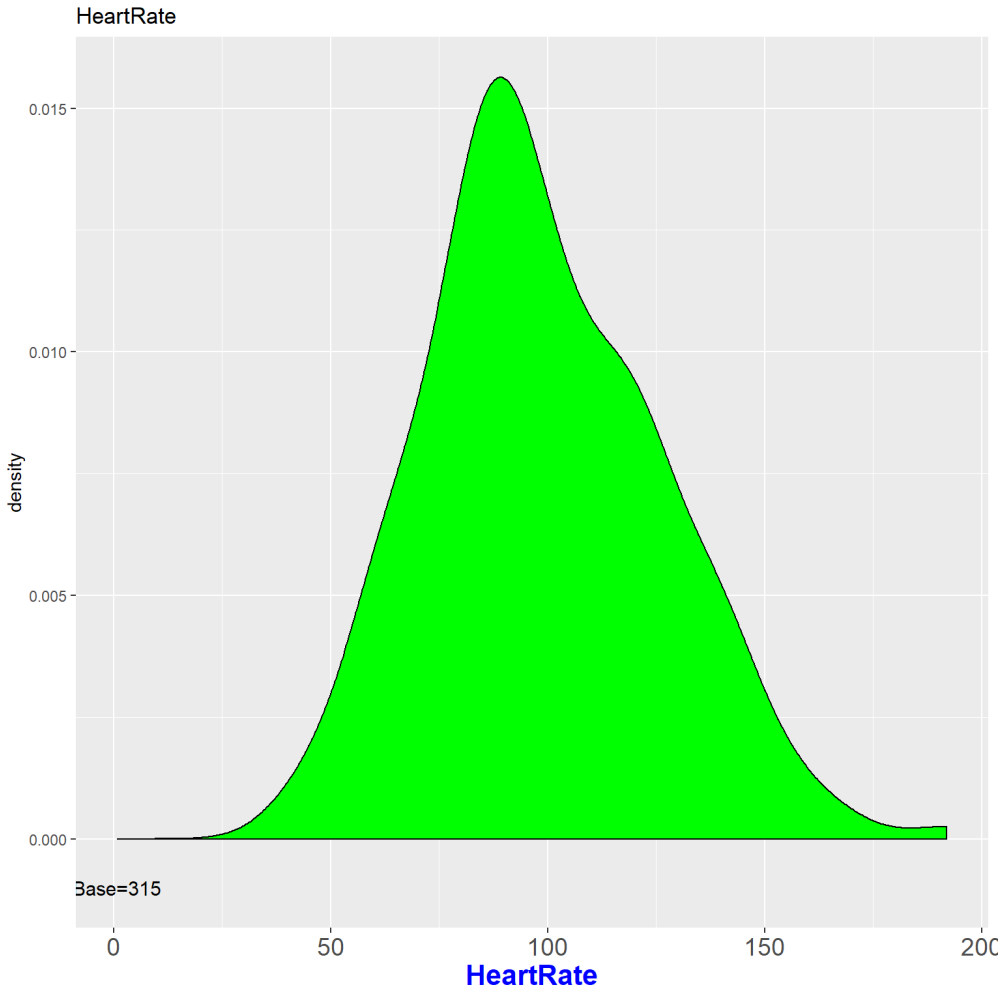
```
    Two Sample t-test

data:  Age by Status
t = -2.7151, df = 198, p-value = 0.007211
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -16.35688  -2.59312
sample estimates:
mean in group Lived  mean in group Died
             55.650              65.125
```

10 Relationships with HeartRate

Using a density estimate plot, the distribution of HeartRate was produced. The main peak here is at about 90 beats per minute. There appears to be a secondary peak around 130 beats per minute. The mean for HeartRate calculated by Rcmdr is 99. Looking at the data, this mean might be said to not accurately portray the frequency of values for HeartRate.

Hide

```
>  ggplot(ICU, aes(x=HeartRate)) +
+  geom_density(fill="green") +
+  ggtitle("HeartRate") +
+  theme(axis.title.x=element_text(size=16, face="bold", colour="blue")) +
+  theme(axis.text.x=element_text(size=14 )) +
+  annotate("text", x=0.8, y=-0.001, label="Base=315",  size=4)
```

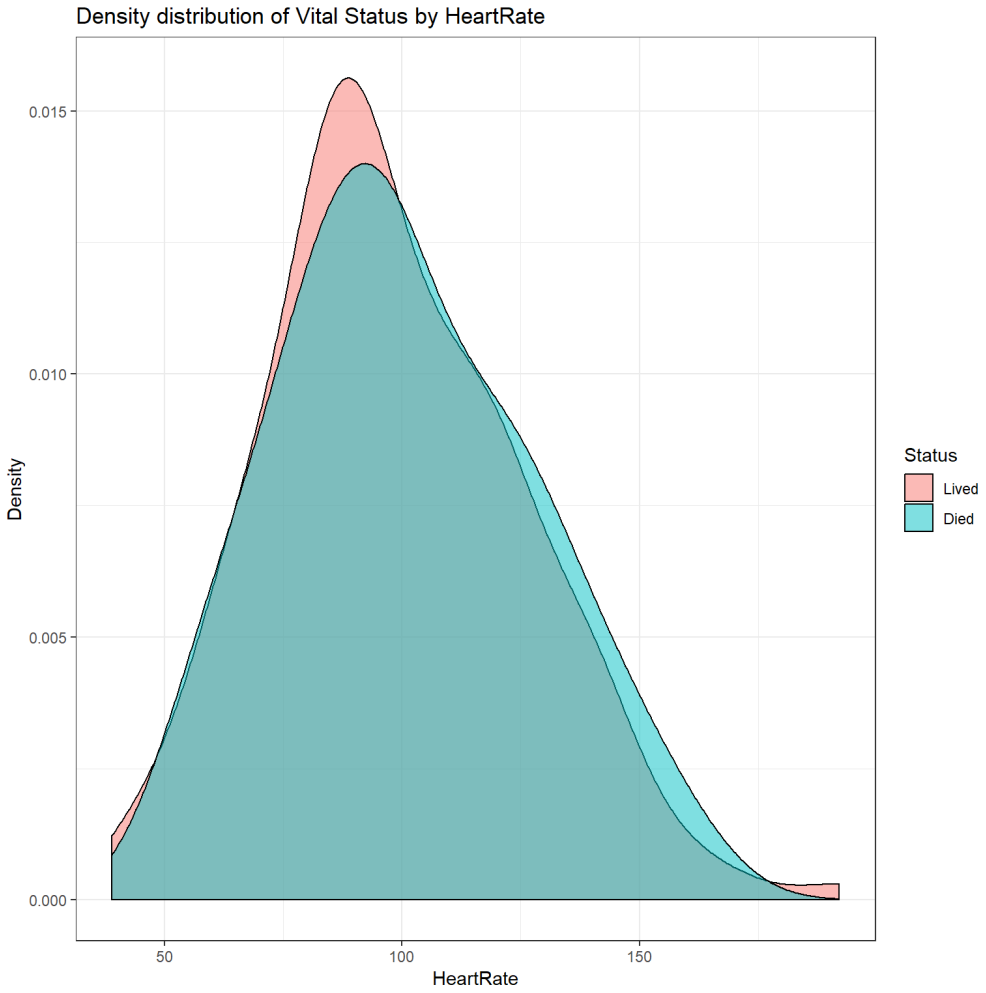


10.1 Vital Status by Heart Rate

Both status categories have similar peaks when plotted on a density estimate plot of HeartRate. This indicates that there might not be a difference in the Status response category based on HeartRate. People who lived did seem to have a higher density of HeartRate values around 90 than those who died. A normal adult heart rate ranges between 60 and 100 beats per minute. This graph might suggest that people who lived more often had heart rates within this range, whereas those who died seemed more likely to have higher heartrates. This could inform further analysis or research.

Hide

```
> ggplot(ICU, aes(x=HeartRate, fill = Status)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Vital Status by HeartRate")
```



Difference in Means
Ho: the variances for Lived and Died are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of means and variances between the two Status groups for HeartRate.

Hide

```
> numSummary(ICU[,c("HeartRate"), drop=FALSE], groups=ICU$Status, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75,1))
```

```
        mean       sd se(mean) HeartRate:n
Lived  98.500 26.97868 2.132852         160
Died  100.625 26.49304 4.188918          40
```

Since the p-value = 0.929 for testing the homogeneity of variances is greater than 0.05, we retain the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for Lived and Died are equal. As such, the Student t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(HeartRate ~ Status, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value Pr(>F)
group   1   0.008  0.929
      198
```

As the p-value = 0.6553, 0 is within the confidence intervals of -11.496845 to 7.246845 and t = -0.44714 is greater than -1.971957, we retain the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for HeartRate are the same among those that lived and those that died. While observations from the graph suggest that people who lived more often had heart rates that were lower, based on the t-test, the means of the two groups are not statistically different.

Hide

```
> t.test(HeartRate~Status, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=ICU)
```

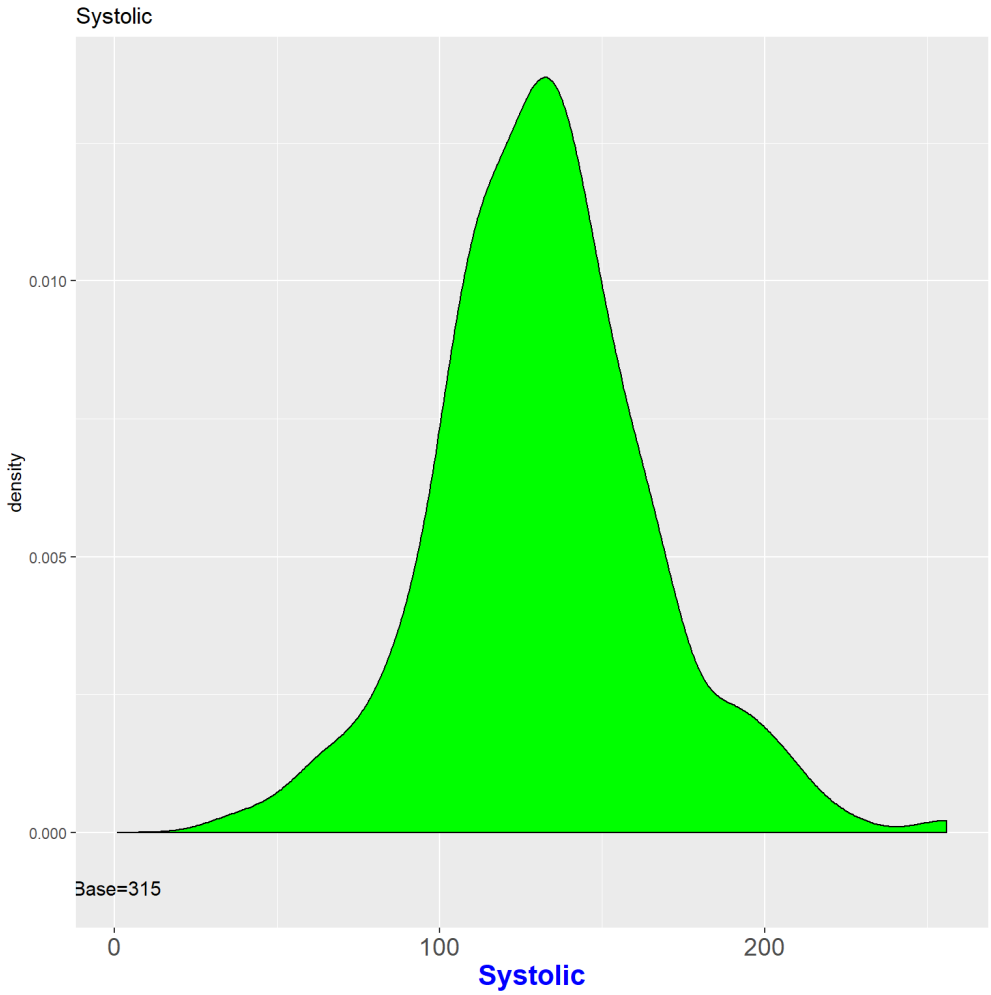
```

    Two Sample t-test

data:  HeartRate by Status
t = -0.44714, df = 198, p-value = 0.6553
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.496845   7.246845
sample estimates:
mean in group Lived  mean in group Died
             98.500             100.625
```

11 Relationships with Systolic

Below is a density plot of Systolic blood pressure in mmHg. A peak appears around 130-140mmHg. The peak is quite narrow and distinct compared to the density plots of Age and HeartRate, suggesting that the majority of individuals had blood pressure readings that are reflective of this plot. The mean for this attribute was 132mmHg, which seems to be more valid than the means of the other numeric variables, based on the distribution of the frequency of values.

Hide

```
> ggplot(ICU, aes(x=Systolic)) +
+  geom_density(fill="green") +
+  ggtitle("Systolic") +
+  theme(axis.title.x=element_text(size=16, face="bold", colour="blue")) +
+  theme(axis.text.x=element_text(size=14 )) +
+  annotate("text", x=0.8, y=-0.001, label="Base=315",  size=4)
```

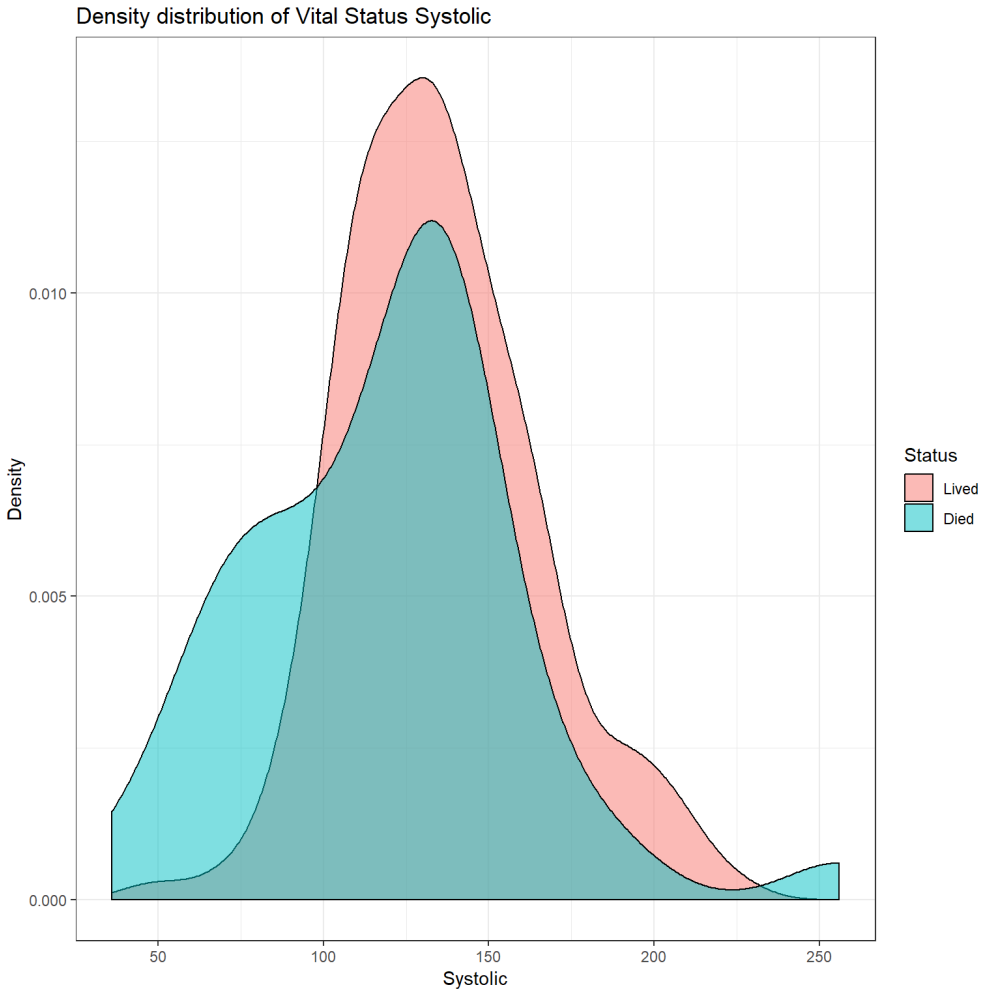


11.1 Vital Status by Systolic

The distribution of Systolic values for those who died seem to be concentrated around 75 mmHg and 140mmHg. A normal Systolic blood pressure can vary widely but the American Heart Association states that blood pressure below 120mmHg is normal. However, excessively low blood pressure could be a result of bleeding, for example, and can result in insufficient blood flow to critical organs. Medications used to restore blood pressure are used when blood pressure becomes too low. This might explain the peak around 75mmHg in the death curve in the graph below.

Hide

```
> ggplot(ICU, aes(x=Systolic, fill = Status)) +
+   theme_bw() +
+       geom_density(alpha=0.5) +
+   labs(y = "Density",
+        title = "Density distribution of Vital Status Systolic")
```



Difference in Means
Ho: the variances for Lived and Died are equal
Ha: the variances are different

Ho: the means are equal
Ha: the means are different

The following is the comparison of variances between the two Status groups for Systolic.

Hide

```
> numSummary(ICU[,c("Systolic"), drop=FALSE], groups=ICU$Status, statistics=c("mean","sd", "se(mean)"), quantiles=c(0,.25,.5,.75
,1))
```

```
         mean       sd se(mean) Systolic:n
Lived 135.6438 29.80151 2.356016        160
Died  118.8250 41.08084 6.495451         40
```

Since the p-value = 0.04205 for testing the homogeneity of variances is less than 0.05, we reject the null hypothesis with a 5% risk of a type 1 error and conclude that the variances for Lived and Died are not equal. As such, the Welch two Sample t-test is used to analyze whether there was a significant difference in means.

Hide

```
> leveneTest(Systolic ~ Status, data=ICU, center="median")
```

```
Levene's Test for Homogeneity of Variance (center = "median")
       Df F value  Pr(>F)
group   1  4.1872 0.04205 *
      198
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value = 0.01856, 0 is not within the confidence intervals of 2.938642 to 30.698858 and t = 2.4341 is greater than 1.971957, we reject the null hypothesis at a 5% risk level of a type 1 error and conclude that the means for systolic blood pressure are not the same among those that lived and those that died. As such, this might suggest that patients who lived had different systolic blood pressures compared to those that died. Therefore a relationship might exist between systolic blood pressure and vital status.

Hide

```
> t.test(Systolic~Status, alternative="two.sided", conf.level=.95, var.equal=FALSE, data=ICU)
```

```
    Welch Two Sample t-test

data:  Systolic by Status
t = 2.4341, df = 49.726, p-value = 0.01856
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  2.938642 30.698858
sample estimates:
mean in group Lived  mean in group Died
           135.6438            118.8250
```

12 Difference in Proportions

We will conduct z-tests between two categorical variables where the dependent variable is Vital status and the independent variable is binary.

12.1 Vital status by Service

Ho: there is no difference between the proportion of patients that died in the medical versus surgical group, risk = 0.05
Ha: there is a difference between the proportion of patients that died in the medical versus surgical group

The following code produces a crosstab of the proportions of Service (Medical and Surgical) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Service, show.col.prc = TRUE)
```

Status	*Service*		***Total***
	Medical	Surgical	
Lived	67 72 %	93 86.9 %	160 80 %
Died	26 28 %	14 13.1 %	40 20 %
Total	93 100 %	107 100 %	200 100 %

χ^2=5.981 · df=1 · φ=0.185 · p=0.014

The following code conducts a z-test and produces a z-value and a p-value for Service.

Hide

```
> x1<-26; x2<-14; n1<-93; n2<-107
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] 2.622729
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.004361436
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of 2.622729 does not fall within this confidence interval, thus we reject the null hypothesis that there is no difference between the proportion of patients dying in the ICU from medical procedures versus surgical procedures.

Further, since the p-value of 0.004361436 is less than 0.05, we, again, reject the null hypothesis that the proportions of the two groups are equal.

Lastly, we can see from the table that more of those that died were there for a medical service and more of those that lived were there for a surgical service.

12.2 Vital status by Sex

Ho: there is no difference between the proportion of male and female patients that died, risk = 0.05 **Ha**: there is a difference between the proportion of male and female patients that died

The following code produces a crosstab of the proportions of Sex (Male and Female) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Sex, show.col.prc = TRUE)
```

Status	*Sex*		***Total***
	Male	Female	
Lived	100 80.6 %	60 78.9 %	160 80 %
Died	24 19.4 %	16 21.1 %	40 20 %
Total	124 100 %	76 100 %	200 100 %

χ^2=0.012 · df=1 · φ=0.021 · p=0.913

The following code conducts a z-test and produces a z-value and a p-value for Sex.

Hide

```
> x1<-24; x2<-16; n1<-124; n2<-76
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -0.2913583
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.6146113
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -0.2913583 does fall within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportion of male and female patients that died in the ICU.

Further, since the p-value of 0.6146113 is greater than 0.05, we, again, fail to reject the null hypothesis that the proportions of the two groups are equal.

Thus, the proportion of men and women that died in the ICU was equal.

We would, however, like to note that there are significantly more males in this dataset than females, which reduces the integrity of the analysis.

12.3 Vital status by Infection

Ho: there is no difference between the proportion of infection probable versus non-infection probable carrying patients that died, risk = 0.05 **Ha**: there is a difference between the proportion of infection probable versus non-infection probable carrying patients that died

The following code produces a crosstab of the proportions of Infection (No and Yes) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Infection, show.col.prc = TRUE)
```

Status	*Infection*		***Total***
	No	Yes	
Lived	100 86.2 %	60 71.4 %	160 80 %
Died	16 13.8 %	24 28.6 %	40 20 %
Total	116 100 %	84 100 %	200 100 %

χ^2=5.759 · df=1 · φ=0.182 · p=0.016

The following code conducts a z-test and produces a z-value and a p-value for Infection.

Hide

```
> x1<-16; x2<-24; n1<-116; n2<-84
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -2.578807
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.9950429
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -2.578807 does not fall within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportion of infection probable versus non-infection probable patients that died.

Further, since the p-value of 0.9950429 is greater than 0.05, we, again, fail to reject the null hypothesis that the proportions of the two groups are equal.

Thus, the proportion of deaths in the infection probable patients is equal to the proportion of deaths in the non-infection probable patients.

12.4 Vital status by Renal

Ho: there is no difference between the proportion of patients that died with a history of chronic renal failure versus without, risk = 0.05 **Ha**: there is a difference between the proportion of patients that died with a history of chronic renal failure versus without

The following code produces a crosstab of the proportions of Renal (No and Yes) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Renal, show.col.prc = TRUE)
```

Status	*Renal*		***Total***
	No	Yes	
Lived	149 82.3 %	11 57.9 %	160 80 %
Died	32 17.7 %	8 42.1 %	40 20 %
Total	181 100 %	19 100 %	200 100 %

χ^2=4.976 · df=1 · φ=0.179 · Fisher's p=0.029

The following code conducts a z-test and produces a z-value and a p-value for Renal.

Hide

```
> x1<-32; x2<-8; n1<-181; n2<-19
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -2.532143
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.9943316
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -2.532143 does not fall within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportion of patients that died with a history of chronic renal failure versus without.

Further, since the p-value of 0.9943316 is greater than 0.05, we, again, fail to reject the null hypothesis that the proportions of the two groups are equal.

Thus, the proportion of deaths in patients with a history of chronic renal failure is equal to the proportion of deaths in patients with no history of chronic renal failure.

12.5 Vital Status by CPR

Ho: there is no difference between the proportion of deaths in patients who recieved CPR upon admission versus patients who did not, risk = 0.05 **Ha**: there is a difference between the proportion of deaths in patients who recieved CPR upon admission versus patients who did not

The following code produces a crosstab of the proportions of CPR (No and Yes) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$CPR, show.col.prc = TRUE)
```

Status	*CPR*		***Total***
	No	Yes	
Lived	154 82.4 %	6 46.2 %	160 80 %
Died	33 17.6 %	7 53.8 %	40 20 %
Total	187 100 %	13 100 %	200 100 %

χ^2=7.821 · df=1 · φ=0.223 · Fisher's p=0.005

The following code conducts a z-test and produces a z-value and a p-value for CPR.

Hide

```
> x1<-33; x2<-7; n1<-187; n2<-13
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -3.155115
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.9991978
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -3.155115 does not fall within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportion of deaths in patients who recieved CPR upon admission versus patients who did not.

Further, since the p-value of 0.9991978 is greater than 0.05, we, again, fail to reject the null hypothesis that the proportions of the two groups are equal.

Thus, the proportion of deaths in patients who recieved CPR upon admission is equal to the proportion of deaths in patients who did not recieve CPR upon admission.

12.6 Vital status by Cancer

Ho: there is no difference between the proportions of patients that died in the cancer versus non-cancer group, risk = 0.05
Ha: there is a difference between the proportions of patients that died in the cancer versus non-cancer group

The following code produces a crosstab of the proportions of Cancer (No and Yes) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Cancer, show.col.prc = TRUE)
```

Status	*Cancer*		***Total***
	No	Yes	
Lived	144 80 %	16 80 %	160 80 %
Died	36 20 %	4 20 %	40 20 %
Total	180 100 %	20 100 %	200 100 %

χ^2=0.000 · df=1 · φ=0.000 · Fisher's p=1.000

The following code conducts a z-test and produces a z-value and a p-value for Cancer.

Hide

```
> x1<-36; x2<-4; n1<-180; n2<-20
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] 0
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.5
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of 0 falls within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportions of patients that died with versus without cancer as part of the present problem.

Further, since the p-value of 0.5 is greater than 0.05, we, again, fail to reject the null hypothesis that the proportions of the Cancer versus non-cancer patients that died are equal.

Thus, there are equal proportions of those who died with cancer as part of the present problem and without, as well as equal proportions of those who lived with cancer as part of the present problem and without.

12.7 Vital status by Previous

Ho: there is no difference between the proportion of patients that died with versus without previous admission to an ICU within 6 months, risk = 0.05
Ha: there is a difference between the proportion of patients that died with versus without previous admission to an ICU within 6 months

The following code produces a crosstab of the proportions of Previous (No and Yes) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Previous, show.col.prc = TRUE)
```

Status	*Previous*		***Total***
	No	Yes	
Lived	137 80.6 %	23 76.7 %	160 80 %
Died	33 19.4 %	7 23.3 %	40 20 %
Total	170 100 %	30 100 %	200 100 %

χ^2=0.061 · df=1 · φ=0.035 · Fisher's p=0.624

The following code conducts a z-test and produces a z-value and a p-value for Previous.

Hide

```
> x1<-33; x2<-7; n1<-170; n2<-30
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -0.4950738
```

Hide

```
> 1-pnorm(zp)
```

```
[1] 0.689726
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -0.4950738 falls within this confidence interval, thus we fail to reject the null hypothesis that there is no difference between the proportion of patients that died with versus without previous admission to an ICU within 6 months of the current admission.

Further, since the p-value of 0.689726 is greater than 0.05, we, again, fail to reject the null hypothesis that there is no difference between the proportion of patients that died whether or not they were previously admitted to the ICU within the past 6 months.

Lastly, we see from the table that there was a lower percentage of patients that died with no prior ICU admission than patients that died with prior ICU admission.

12.8 Vital status by Type

Ho: there is no difference between the proportion of patients that died in the elective admission group versus the emergency admission group, risk = 0.05 **Ha**: there is a difference between the proportion of patients that died in the elective admission group versus the emergency admission group

The following code produces a crosstab of the proportions of Type (Elective and Emergency) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Type, show.col.prc = TRUE)
```

Status	*Type*		***Total***
	Elective	Emergency	
Lived	51 96.2 %	109 74.1 %	160 80 %
Died	2 3.8 %	38 25.9 %	40 20 %
Total	53 100 %	147 100 %	200 100 %

χ^2=10.527 · df=1 · φ=0.244 · p=0.001

The following code conducts a z-test and produces a z-value and a p-value for Type.

Hide

```
> x1<-2; x2<-38; n1<-53; n2<-147
> p1<-x1/n1; p2<-x2/n2
> p<-(x1+x2)/(n1+n2)
> varp<-p*(1-p)*(1/n1 + 1/n2)
> stdp<-sqrt(varp)
> zp<-(p1 - p2)/stdp
> zp
```

```
[1] -3.444743
```

Hide

```
> pnorm(zp)
```

```
[1] 0.000285801
```

When conducting a two-tail z-test with a 5% level of risk of a type 1 error, the critical values are -1.96 and +1.96. The z-value of -3.444743 does not fall within this confidence interval, thus we reject the null hypothesis that there is no difference between the proportion of patients that died in the elective admission group versus the emergency admission group.

Further, since the p-value of 0.000285801 is less than 0.05, we, again, reject the null hypothesis that the proportions of the two groups are equal.

We can see from the table that the emergency admission type had a higher proportion of deaths than the elective admission type did.

We will now conduct a prop.test for categorical variables where the independent variable has more than 2 subgroups.

12.9 Vital status by Age

Ho: there is no difference between the proportion of patients that died across the 5 age groups, risk = 0.05
Ha: the proportion of patients that died is different in at least one of the 5 age groups

We will now bin the Age variable by equal-width bins:

Hide

```
> ICU$Age.Binned <- with(ICU, binVariable(Age, bins=5, method='intervals', labels=c('Age group 1','Age group 2','Age group 3','A
ge group 4',
+   'Age group 5')))
```

The following code produces a crosstab of the proportions of Age (Groups 1-5) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Age.Binned, show.col.prc = TRUE)
```

Status	*Age.Binned*					***Total***
	Age group 1	Age group 2	Age group 3	Age group 4	Age group 5	
Lived	29 93.5 %	17 89.5 %	36 78.3 %	57 76 %	21 72.4 %	160 80 %
Died	2 6.5 %	2 10.5 %	10 21.7 %	18 24 %	8 27.6 %	40 20 %
Total	31 100 %	19 100 %	46 100 %	75 100 %	29 100 %	200 100 %

χ^2=6.502 · df=4 · Cramer's V=0.180 · Fisher's p=0.144

The following code conducts a z-test and produces a p-value, a chi-squared value and sample proportion estimates for Age.

Hide

```
> Died  <- c( 2, 2, 10, 18, 8 )
> Total <- c( 31, 19, 46, 75, 29 )
> prop.test(Died, Total)
```

```
    5-sample test for equality of proportions without continuity
    correction

data:  Died out of Total
X-squared = 6.5023, df = 4, p-value = 0.1646
alternative hypothesis: two.sided
sample estimates:
    prop 1     prop 2     prop 3     prop 4     prop 5
0.06451613 0.10526316 0.21739130 0.24000000 0.27586207
```

Since the p-value of 0.1646 is greater than 0.05, we fail to reject the null hypothesis that there is no difference between the proportion of patients that died across the 5 age groups.

Further, when conducting a chi-square test with a 5% level of risk of a type 1 error and 4 degrees of freedom, the critical value is 9.49. Since the chi-squared value of 6.5023 is less than this critical value, we, again, fail to reject the null hypothesis that there is no difference between the proportion of patients that died across the 5 age groups.

However, we can see from the sample estimates produced by the proportions test that Age group 1 and Age group 2 have slightly lower proportions than age groups 3, 4 and 5. Thus, the number of patients that died in age group 1 (the youngest age group) and age group 2 is much lower than the rest. The greatest proportion of patients that died is in group 5 (the oldest age group). Overall, proportions of death increased slightly in each age group, with closer proportions in the 3 oldest age groups and very low proportions in the 2 youngest age groups.

12.10 Vital status by Systolic Blood Pressure

Ho: there is no difference between the proportion of patients that died across the 6 systolic blood pressure level groups, risk = 0.05
Ha: the proportion of patients that died is different in at least one of the 6 systolic blood pressure level groups

We grouped Systolic into medically characterized categories of hypotension, normal blood pressure, elevated blood pressure and stage 1, 2 and 3 hypertension. Hypotension is characterized by a blood pressure level of less than 80mmHg, normal blood pressure is characterized by a blood pressure level between 80 to 120mmHg, elevated blood pressure is characterized by a blood pressure level between 120 and 129mmHg, stage 1 hypertension is characterized by a blood pressure level 130-139mmHg, stage 2 hypertension is characterized by a blood pressure level 140-180mmHg, and stage 3 hypertension is characterized by a blood pressure level greater than 180mmHg.

Hide

```
> ICU <-
+   within(ICU, {
+   Systolic.grouped <- Recode(Systolic,
+   '0:80="hypotension"; 80:120="normal"; 120:129="elevated"; 130:139="stage 1 hypertension"; 140:180="stage 2 hypertension"; 18
0:260="stage 3 hypertension"; ;',
+    as.factor=TRUE)
+ })
```

The following code produces a crosstab of the proportions of Systolic (Hypotension, normal, elevated, 3 categories of hypertension) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Systolic.grouped, show.col.prc = TRUE)
```

Status	*Systolic.grouped*						***Total***
	elevated	hypotension	normal	stage 1 hypertension	stage 2 hypertension	stage 3 hypertension	
Lived	14 82.4 %	3 25 %	52 85.2 %	24 80 %	54 83.1 %	13 86.7 %	160 80 %
Died	3 17.6 %	9 75 %	9 14.8 %	6 20 %	11 16.9 %	2 13.3 %	40 20 %
Total	17 100 %	12 100 %	61 100 %	30 100 %	65 100 %	15 100 %	200 100 %

χ^2=24.597 · df=5 · Cramer's V=0.351 · Fisher's p=0.002

The following code conducts a z-test and produces a p-value, a chi-squared value and sample proportion estimates for Systolic.

Hide

```
> Died  <- c( 3, 9, 9, 6, 11, 2 )
> Total <- c( 17, 12, 61, 30, 65, 15 )
> prop.test(Died, Total)
```

```
    6-sample test for equality of proportions without continuity
    correction

data:  Died out of Total
X-squared = 24.597, df = 5, p-value = 0.0001667
alternative hypothesis: two.sided
sample estimates:
   prop 1    prop 2    prop 3    prop 4    prop 5    prop 6
0.1764706 0.7500000 0.1475410 0.2000000 0.1692308 0.1333333
```

Since the p-value of 0.0001667 is less than 0.05, we will reject the null hypothesis that there is no difference between the proportion of patients that died in the 6 systolic blood pressure level groups. We can see from the sample estimates that the proportions are indeed different between the 6 groups, with the highest proportion in the hypotenstion range and in the stage 1 hypertension range. The proportions in the elevated, normal, stage 2 hypertension and stage 3 hypertension groups are much lower and are quite close to each other. Thus, the hypotension and stage 1 hypertension groups had higher proportions of patients that died.

Further, when conducting a chi-square test with a 5% level of risk of a type 1 error and 5 degrees of freedom, the critical value is 11.07. Since the chi-squared value of 24.597 is greater than this critical value, we will, again, reject the null hypothesis that there is no difference between the proportion of patients that died in the 6 systolic blood pressure level groups.

12.11 Vital status by Heart Rate

Ho: there is no difference between the proportion of patients that died across the bradychardia, normal heart rate and elevated heart rate groups, risk = 0.05
Ha: the proportion of patients that died is different in at least one of the 3 heart rate categories

We grouped heart rate into medically characterized categories of bradycardia, normal heart rate and tachycardia. Bradycardia is characterized by a heart rate less than 60 beats per minute, normal heart rate is characterized by a heart rate between 60-100 beats per minute, and tachycardia is characterized by a heart rate greater than 100 beats per minute.

Hide

```
> ICU <-
+   within(ICU, {
+   HeartRate.grouped <- Recode(HeartRate,
+   '0:60="bradycardia"; 60:100="normal"; 100:200="tachycardia"',
+    as.factor=TRUE)
+ })
```

The following code produces a crosstab of the proportions of Status (Bradycardia, Normal and Tachycardia) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$HeartRate.grouped, show.col.prc = TRUE)
```

Status	*HeartRate.grouped*			***Total***
	bradycardia	normal	tachycardia	
Lived	12 80 %	84 80 %	64 80 %	160 80 %
Died	3 20 %	21 20 %	16 20 %	40 20 %
Total	15 100 %	105 100 %	80 100 %	200 100 %

χ^2=0.000 · df=2 · Cramer's V=0.000 · Fisher's p=1.000

The following code conducts a z-test and produces a p-value, a chi-squared value and sample proportion estimates for Heart Rate.

Hide

```
> Died  <- c( 3, 16, 21 )
> Total <- c( 15, 80, 105 )
> prop.test(Died, Total)
```

```
    3-sample test for equality of proportions without continuity
    correction

data:  Died out of Total
X-squared = 0, df = 2, p-value = 1
alternative hypothesis: two.sided
sample estimates:
prop 1 prop 2 prop 3
   0.2    0.2    0.2
```

Since the p-value of 1 is greater than 0.05, we fail to reject the null hypothesis that there is no difference between proportion of patients that died across the bradychardia, normal heart rate and elevated heart rate groups. An exact p-value of 1 means that the difference in proportions of patients that died between the 3 heart rate groups is exactly 0. Thus, there were an equal number of patients that died with bradycardia, normal heart rate and elevated heart rate.

Further, when conducting a chi-square test with a 5% level of risk of a type 1 error and 2 degrees of freedom, the critical value is 5.99. Since the chi-squared value of 0 is lower than this critical value, we, again, fail to reject the null hypothesis that there is no difference between the proportion of patients that died having bradychardia, normal heart rate and elevated heart rate.

12.12 Vital status by Consciousness

Ho: there is no difference between the proportion of patients that died across the Conscious, Deep Stupor and Coma groups, risk = 0.05
Ha: the proportion of patients that died is different in at least one of the Conscious, Deep Stupor and Coma groups

The following code produces a crosstab of the proportions of Consciousness (Conscious, Deep Stupor and Coma) by Status (Lived or Died).

Hide

```
> pacman::p_load(sjPlot)
> sjt.xtab(ICU$Status, ICU$Consciousness, show.col.prc = TRUE)
```

Status	*Consciousness*			***Total***
	Conscious	Deep Stupor	Coma	
Lived	158 85.4 %	0 0 %	2 20 %	160 80 %
Died	27 14.6 %	5 100 %	8 80 %	40 20 %
Total	185 100 %	5 100 %	10 100 %	200 100 %

χ^2=45.878 · df=2 · Cramer's V=0.479 · Fisher's p=0.000

The following code conducts a z-test and produces a p-value, a chi-squared value and sample proportion estimates for Consciousness.

Hide

```
> Died  <- c( 27, 5, 8 )
> Total <- c( 185, 5, 10 )
> prop.test(Died, Total)
```

```
    3-sample test for equality of proportions without continuity
    correction

data:  Died out of Total
X-squared = 45.878, df = 2, p-value = 1.091e-10
alternative hypothesis: two.sided
sample estimates:
   prop 1    prop 2    prop 3
0.1459459 1.0000000 0.8000000
```

Since the p-value of 1.091e-10 is less than 0.05, we will reject the null hypothesis that there is no difference between the proportion of patients that died across the Conscious, Deep Stupor and Coma groups.

Further, when conducting a chi-square test with a 5% level of risk of a type 1 error and 2 degrees of freedom, the critical value is 5.99. Since the chi-squared value of 45.878 is greater than this critical value, we will, again, reject the null hypothesis that there is no difference between the proportion of patients that died across the 3 Consciousness groups.

With a proportion of 1, the Deep Stupor group had the highest proportion of patients that died as all the patients with a deep stupor died. The Coma group has the second highest proportion of patients that died with a sample estimate of 80% dying. The lowest number of deaths was in the group of patients that had no coma or stupor, which is expected. Those with a deep stupor had a higher proportion of deaths than those in a coma.

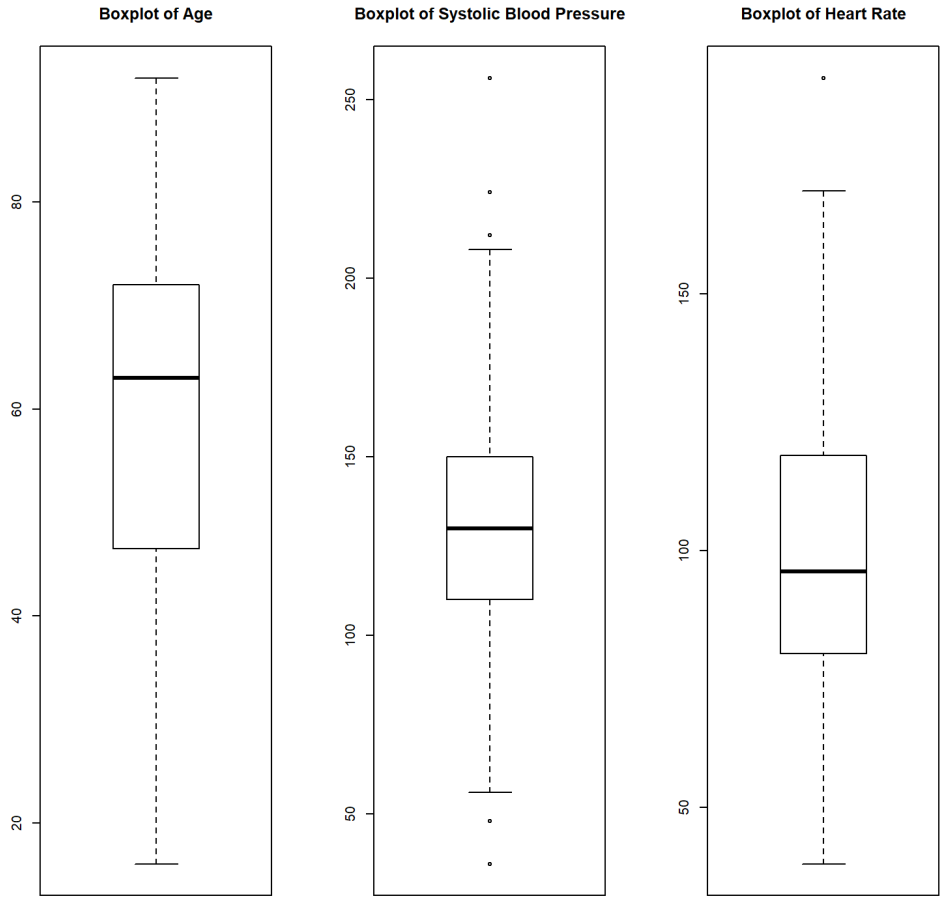
From the tests of equal proportions, we note that Service (type of procedure), Type (of admission) and Conciousness had a statistically significant impact on the Vital status of patients. On the other hand, Sex, Infection, Renal, CPR, Cancer, Previous (ICU admission within the last 6 months), Age, Systolic Blood Pressure and Heart rate were not significant predictors of Vital status of patients.

13 Detection of outliers of Int Variables

The following variables were not significant predictors of Vital status of the patient: Cancer, Previous (ICU admission within the last 6 months), Age, Systolic Blood Pressure, Heart Rate and Race.

Hide

```
> par(mfrow=c(1,3))
> boxplot(ICU$Age, main="Boxplot of Age")
> boxplot(ICU$Systolic, main="Boxplot of Systolic Blood Pressure")
> boxplot(ICU$HeartRate, main="Boxplot of Heart Rate")
```



As we can see above, the outliers are present in systolic blood pressure and heart rate. To see the values of the outliers, please see below, where the first row includes the outliers for systolic blood pressure, and the second includes outliers for heart rate.

Hide

```
> systolic_outlier<-boxplot.stats(ICU$Systolic)
> heartrate_outlier<-boxplot.stats(ICU$HeartRate)
> systolic_outlier$out
```

```
[1]  48 212 224  36 256
```

Hide

```
> heartrate_outlier$out
```

```
[1] 192
```

14 Crosstabs (relations between categorical variables)

Hide

```
> sjt.xtab(ICU$Status, ICU$Sex, show.col.prc = TRUE)
```

Status	*Sex*		***Total***
	Male	Female	
Lived	100 80.6 %	60 78.9 %	160 80 %
Died	24 19.4 %	16 21.1 %	40 20 %
Total	124 100 %	76 100 %	200 100 %

χ^2=0.012 · df=1 · φ=0.021 · p=0.913

Hide

```
> sjt.xtab(ICU$Status, ICU$Race, show.col.prc = TRUE)
```

Status	*Race*			***Total***
	White	Black	Other	
Lived	138 78.9 %	14 93.3 %	8 80 %	160 80 %
Died	37 21.1 %	1 6.7 %	2 20 %	40 20 %
Total	175 100 %	15 100 %	10 100 %	200 100 %

χ^2=1.810 · df=2 · Cramer's V=0.095 · Fisher's p=0.502

Hide

```
> sjt.xtab(ICU$Status, ICU$Service, show.col.prc = TRUE)
```

Status	*Service*		***Total***
	Medical	Surgical	
Lived	67 72 %	93 86.9 %	160 80 %
Died	26 28 %	14 13.1 %	40 20 %
Total	93 100 %	107 100 %	200 100 %

χ^2=5.981 · df=1 · φ=0.185 · p=0.014

Hide

```
> sjt.xtab(ICU$Status, ICU$Cancer, show.col.prc = TRUE)
```

Status	*Cancer*		***Total***
	No	Yes	
Lived	144 80 %	16 80 %	160 80 %
Died	36 20 %	4 20 %	40 20 %
Total	180 100 %	20 100 %	200 100 %

χ^2=0.000 · df=1 · φ=0.000 · Fisher's p=1.000

Hide

```
> sjt.xtab(ICU$Status, ICU$Renal, show.col.prc = TRUE)
```

Status	*Renal*		***Total***
	No	Yes	
Lived	149 82.3 %	11 57.9 %	160 80 %
Died	32 17.7 %	8 42.1 %	40 20 %
Total	181 100 %	19 100 %	200 100 %

χ^2=4.976 · df=1 · φ=0.179 · Fisher's p=0.029

Hide

```
> sjt.xtab(ICU$Status, ICU$Infection, show.col.prc = TRUE)
```

Status	*Infection*		***Total***
	No	Yes	
Lived	100 86.2 %	60 71.4 %	160 80 %
Died	16 13.8 %	24 28.6 %	40 20 %
Total	116 100 %	84 100 %	200 100 %

χ^2=5.759 · df=1 · φ=0.182 · p=0.016

Hide

```
> sjt.xtab(ICU$Status, ICU$CPR, show.col.prc = TRUE)
```

Status	*CPR*		***Total***
	No	Yes	
Lived	154 82.4 %	6 46.2 %	160 80 %
Died	33 17.6 %	7 53.8 %	40 20 %
Total	187 100 %	13 100 %	200 100 %

χ^2=7.821 · df=1 · φ=0.223 · Fisher's p=0.005

Hide

```
> sjt.xtab(ICU$Status, ICU$Previous, show.col.prc = TRUE)
```

Status	*Previous*		***Total***
	No	Yes	
Lived	137 80.6 %	23 76.7 %	160 80 %
Died	33 19.4 %	7 23.3 %	40 20 %
Total	170 100 %	30 100 %	200 100 %

χ^2=0.061 · df=1 · φ=0.035 · Fisher's p=0.624

Hide

```
> sjt.xtab(ICU$Status, ICU$Type, show.col.prc = TRUE)
```

Status	*Type*		***Total***
	Elective	Emergency	
Lived	51 96.2 %	109 74.1 %	160 80 %
Died	2 3.8 %	38 25.9 %	40 20 %
Total	53 100 %	147 100 %	200 100 %

χ^2=10.527 · df=1 · φ=0.244 · p=0.001

Hide

```
> sjt.xtab(ICU$Status, ICU$Fracture, show.col.prc = TRUE)
```

Status	*Fracture*		***Total***
	No	Yes	
Lived	148 80 %	12 80 %	160 80 %
Died	37 20 %	3 20 %	40 20 %
Total	185 100 %	15 100 %	200 100 %

χ^2=0.000 · df=1 · φ=0.000 · Fisher's p=1.000

Hide

```
> sjt.xtab(ICU$Status, ICU$PO2, show.col.prc = TRUE)
```

Status	*PO2*		***Total***
	No	Yes	
Lived	149 81 %	11 68.8 %	160 80 %
Died	35 19 %	5 31.2 %	40 20 %
Total	184 100 %	16 100 %	200 100 %

χ^2=0.718 · df=1 · φ=0.083 · Fisher's p=0.324

Hide

```
> sjt.xtab(ICU$Status, ICU$PH, show.col.prc = TRUE)
```

Status	*PH*		***Total***
	No	Yes	
Lived	151 80.7 %	9 69.2 %	160 80 %
Died	36 19.3 %	4 30.8 %	40 20 %
Total	187 100 %	13 100 %	200 100 %

χ^2=0.416 · df=1 · φ=0.071 · Fisher's p=0.297

Hide

```
> sjt.xtab(ICU$Status, ICU$PCO2, show.col.prc = TRUE)
```

Status	*PCO2*	***Total***

	No	Yes	
Lived	144 80 %	16 80 %	160 80 %
Died	36 20 %	4 20 %	40 20 %
Total	180 100 %	20 100 %	200 100 %

χ^2=0.000 · df=1 · φ=0.000 · Fisher's p=1.000

Hide

```
> sjt.xtab(ICU$Status, ICU$Bicarbonate, show.col.prc = TRUE)
```

Status	*Bicarbonate*		***Total***
	No	Yes	
Lived	150 81.1 %	10 66.7 %	160 80 %
Died	35 18.9 %	5 33.3 %	40 20 %
Total	185 100 %	15 100 %	200 100 %

χ^2=1.014 · df=1 · φ=0.095 · Fisher's p=0.187

Hide

```
> sjt.xtab(ICU$Status, ICU$Creatinine, show.col.prc = TRUE)
```

Status	*Creatinine*		***Total***
	No	Yes	
Lived	155 81.6 %	5 50 %	160 80 %
Died	35 18.4 %	5 50 %	40 20 %
Total	190 100 %	10 100 %	200 100 %

χ^2=4.112 · df=1 · φ=0.172 · Fisher's p=0.029

Hide

```
> sjt.xtab(ICU$Status, ICU$Consciousness, show.col.prc = TRUE)
```

Status	*Consciousness*			***Total***
	Conscious	Deep Stupor	Coma	
Lived	158 85.4 %	0 0 %	2 20 %	160 80 %
Died	27 14.6 %	5 100 %	8 80 %	40 20 %
Total	185 100 %	5 100 %	10 100 %	200 100 %

χ^2=45.878 · df=2 · Cramer's V=0.479 · Fisher's p=0.000

Statistically significant crosstabs include: Creatinine, Type, CPR, Infection,Renal, Service.

15 Goodman Kruskall's Lambda

We will also conduct the Goodman Kruskall's Lambda test to measure the proportional reduction in error in the crosstabs conducted above. As we can see in the lambda calculation below, there is a weak association between the dependent and independent variable, as the lambda is higher than 0 but lower than 1, in which the latter would denote a strong relationship.

Hide

```
> ICUtable<-as.table(cbind(c(ICU$Service, ICU$Cancer, ICU$Previous, ICU$Type, ICU$Age.binned, ICU$Systolic.grouped, ICU$HeartRat
e.grouped, ICU$Consciousness, ICU$CPR, ICU$Infection, ICU$Sex, ICU$Renal), c(ICU$Status)))
> Lambda(ICUtable, direction=c("symmetric", "row", "column"), conf.level = 0.95)
```

```
    lambda     lwr.ci     upr.ci
0.02407445 0.01881298 0.02933592
```

16 Logistic Regression

In order to deduce the model that will predict the outcome of the status based on our chosen variables, we had decided to conduct a logistic regression, due to the fact that our respondent variable is binomial.

According to our logistic regression model below, statistically significant indicators include age, cancer, systolic blood pressure, type, service, previous and consciousness. The large enough difference between the null and residual deviance provides us with some confidence in our model, although our confidence intervals do not as they have 0 in them. However, before we state our model based on the calculations below, we will reduce our model to its significant predictors, and further investigate whether our model is sturdy.

Hide

```
> ICU.m<-glm(formula=Status~Age.binned+Sex+Service+Cancer+Renal+Infection+CPR+Systolic.grouped+HeartRate.grouped+Previous+Type+P
O2+Consciousness, family=binomial(logit), data=ICU)
> summary(ICU.m)
```

```
Call:
glm(formula = Status ~ Age.binned + Sex + Service + Cancer + 
    Renal + Infection + CPR + Systolic.grouped + HeartRate.grouped + 
    Previous + Type + PO2 + Consciousness, family = binomial(logit), 
    data = ICU)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-1.8829  -0.4709  -0.2601  -0.0640   2.4897  

Coefficients:
                                    Estimate Std. Error z value
(Intercept)                        -6.63439    2.10826  -3.147
Age.binnedGroup 2                   0.79631    0.91905   0.866
Age.binnedGroup 3                  -0.02831    0.98049  -0.029
Age.binnedGroup 4                   2.02247    0.93499   2.163
Age.binnedGroup 5                   2.59815    0.95435   2.722
SexFemale                          -0.66070    0.54423  -1.214
ServiceSurgical                    -1.07449    0.63645  -1.688
CancerYes                           3.29458    1.11696   2.950
RenalYes                            0.12988    0.75184   0.173
InfectionYes                       -0.03277    0.59850  -0.055
CPRYes                              0.83400    0.99999   0.834
Systolic.groupedhypotension         2.48531    1.31637   1.888
Systolic.groupednormal             -0.24664    0.90602  -0.272
Systolic.groupedstage 1 hypertension    0.17345    1.04589   0.166
Systolic.groupedstage 2 hypertension    0.06994    0.89875   0.078
Systolic.groupedstage 3 hypertension   -2.70641    1.64322  -1.647
HeartRate.groupednormal             1.11165    1.25553   0.885
HeartRate.groupedtachycardia        0.46650    1.25586   0.371
PreviousYes                         1.54889    0.71295   2.173
TypeEmergency                       3.47421    1.29731   2.678
PO2Yes                             -1.12282    1.00804  -1.114
ConsciousnessDeep Stupor           22.15464 1493.02278   0.015
ConsciousnessComa                   3.40765    1.34549   2.533
                                  Pr(>|z|)
(Intercept)                        0.00165 **
Age.binnedGroup 2                  0.38625
Age.binnedGroup 3                  0.97696
Age.binnedGroup 4                  0.03053 *
Age.binnedGroup 5                  0.00648 **
SexFemale                          0.22475
ServiceSurgical                    0.09136 .
CancerYes                          0.00318 **
RenalYes                           0.86285
InfectionYes                       0.95634
CPRYes                             0.40428
Systolic.groupedhypotension        0.05903 .
Systolic.groupednormal             0.78545
Systolic.groupedstage 1 hypertension  0.86828
Systolic.groupedstage 2 hypertension  0.93797
Systolic.groupedstage 3 hypertension  0.09956 .
HeartRate.groupednormal            0.37594
HeartRate.groupedtachycardia       0.71030
PreviousYes                        0.02982 *
TypeEmergency                      0.00741 **
PO2Yes                             0.26534
ConsciousnessDeep Stupor           0.98816
ConsciousnessComa                  0.01132 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 200.16  on 199  degrees of freedom
Residual deviance: 113.90  on 177  degrees of freedom
AIC: 159.9

Number of Fisher Scoring iterations: 16
```

Hide

```
> confint(ICU.m)
```

```
                                       2.5 %     97.5 %
(Intercept)                       -11.2943293 -2.9183994
Age.binnedGroup 2                  -0.9845178  2.7306204
Age.binnedGroup 3                  -1.9841072  1.9925381
Age.binnedGroup 4                   0.2829682  4.0461747
Age.binnedGroup 5                   0.8427448  4.6641819
SexFemale                          -1.7796928  0.3753063
ServiceSurgical                    -2.3924683  0.1322101
CancerYes                           1.2162772  5.7378792
RenalYes                           -1.4328515  1.5686034
InfectionYes                       -1.2264859  1.1514648
CPRYes                             -1.2229901  2.7790263
Systolic.groupedhypotension         0.0492908  5.2865460
Systolic.groupednormal             -1.9894751  1.6416260
Systolic.groupedstage 1 hypertension -1.8759776  2.3029733
Systolic.groupedstage 2 hypertension -1.6435452  1.9585803
Systolic.groupedstage 3 hypertension -6.4558896  0.2342551
HeartRate.groupednormal            -1.0472842  4.0750579
HeartRate.groupedtachycardia       -1.7630972  3.3596867
PreviousYes                         0.1451916  2.9812219
TypeEmergency                       1.3260542  6.7287588
PO2Yes                             -3.3410665  0.7062389
ConsciousnessDeep Stupor         -131.5510230         NA
ConsciousnessComa                   0.9466277  6.3280206
```

Hide

```
> exp(coef(ICU.m))
```

```
                       (Intercept)                  Age.binnedGroup 2
                      1.314382e-03                       2.217333e+00
                 Age.binnedGroup 3                  Age.binnedGroup 4
                      9.720847e-01                       7.557003e+00
                 Age.binnedGroup 5                         SexFemale
                      1.343889e+01                       5.164922e-01
                   ServiceSurgical                          CancerYes
                      3.414735e-01                       2.696616e+01
                       RenalYes                          InfectionYes
                      1.138693e+00                       9.677654e-01
                           CPRYes          Systolic.groupedhypotension
                      2.302504e+00                       1.200485e+01
           Systolic.groupednormal Systolic.groupedstage 1 hypertension
                      7.814207e-01                       1.189401e+00
Systolic.groupedstage 2 hypertension Systolic.groupedstage 3 hypertension
                      1.072445e+00                       6.677639e-02
          HeartRate.groupednormal         HeartRate.groupedtachycardia
                      3.039363e+00                       1.594400e+00
                       PreviousYes                        TypeEmergency
                      4.706237e+00                       3.227245e+01
                             PO2Yes             ConsciousnessDeep Stupor
                      3.253609e-01                       4.184463e+09
                 ConsciousnessComa
                      3.019420e+01
```

16.1 Final model

As stated above, we wanted to reduce our model to a final model, based on its significant predictors.

As we can see below, the predictors are still significant, and the difference between the null and residual deviance are still large enough to provide us with some confidence in our model, while our confidence intervals still do not. However, this does not mean that we won't test the strength of our final model, which we will do below.

Hide

```
> ICU.final.m<-glm(formula=Status~Age.binned+Service+Cancer+Systolic.grouped+Consciousness+Previous+Type, family=binomial(logi
t),data=ICU)
> summary(ICU.final.m)
```

```
Call:
glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Consciousness + Previous + Type, family = binomial(logit),
    data = ICU)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.30437  -0.51110  -0.28167  -0.07613   2.54830

Coefficients:
                                   Estimate Std. Error z value
(Intercept)                        -5.99744    1.71089  -3.505
Age.binnedGroup 2                   0.74928    0.86787   0.863
Age.binnedGroup 3                   0.02170    0.94389   0.023
Age.binnedGroup 4                   1.42785    0.84632   1.687
Age.binnedGroup 5                   2.28294    0.86739   2.632
ServiceSurgical                    -0.91971    0.55234  -1.665
CancerYes                           3.22412    1.06372   3.031
Systolic.groupedhypotension         1.88610    1.16403   1.620
Systolic.groupednormal             -0.21794    0.85482  -0.255
Systolic.groupedstage 1 hypertension    0.09025    0.96050   0.094
Systolic.groupedstage 2 hypertension    0.23025    0.84390   0.273
Systolic.groupedstage 3 hypertension   -2.60303    1.51526  -1.718
ConsciousnessDeep Stupor           21.44683 1572.96411   0.014
ConsciousnessComa                   3.19219    1.03515   3.084
PreviousYes                         1.32364    0.66415   1.993
TypeEmergency                       3.47964    1.29137   2.695
                                  Pr(>|z|)
(Intercept)                       0.000456 ***
Age.binnedGroup 2                 0.387942
Age.binnedGroup 3                 0.981661
Age.binnedGroup 4                 0.091579 .
Age.binnedGroup 5                 0.008489 **
ServiceSurgical                   0.095888 .
CancerYes                         0.002438 **
Systolic.groupedhypotension       0.105164
Systolic.groupednormal            0.798761
Systolic.groupedstage 1 hypertension 0.925137
Systolic.groupedstage 2 hypertension 0.784972
Systolic.groupedstage 3 hypertension 0.085819 .
ConsciousnessDeep Stupor          0.989121
ConsciousnessComa                 0.002044 **
PreviousYes                       0.046265 *
TypeEmergency                     0.007049 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 200.16  on 199  degrees of freedom
Residual deviance: 119.76  on 184  degrees of freedom
AIC: 151.76

Number of Fisher Scoring iterations: 16
```

Hide

```
> confint(ICU.final.m)
```

```
                                        2.5 %     97.5 %
(Intercept)                          -9.872787e+00 -2.9758853
Age.binnedGroup 2                    -9.269851e-01  2.5835764
Age.binnedGroup 3                    -1.852435e+00  1.9673144
Age.binnedGroup 4                    -1.597920e-01  3.2532500
Age.binnedGroup 5                     6.943378e-01  4.1819402
ServiceSurgical                      -2.062763e+00  0.1239095
CancerYes                             1.241663e+00  5.5829700
Systolic.groupedhypotension          -2.975703e-01  4.3473638
Systolic.groupednormal               -1.850946e+00  1.5794947
Systolic.groupedstage 1 hypertension -1.803025e+00  2.0529066
Systolic.groupedstage 2 hypertension -1.364411e+00  2.0213164
Systolic.groupedstage 3 hypertension -6.109662e+00  0.1221235
ConsciousnessDeep Stupor             -1.409386e+02         NA
ConsciousnessComa                     1.316358e+00  5.5297057
PreviousYes                           1.278671e-04  2.6418700
TypeEmergency                         1.362320e+00  6.7326806
```

Hide

```
> exp(coef(ICU.final.m))
```

```
                          (Intercept)                   Age.binnedGroup 2
                         2.485115e-03                        2.115474e+00
                  Age.binnedGroup 3                    Age.binnedGroup 4
                         1.021934e+00                        4.169717e+00
                  Age.binnedGroup 5                      ServiceSurgical
                         9.805506e+00                        3.986338e-01
                             CancerYes          Systolic.groupedhypotension
                         2.513138e+01                        6.593579e+00
              Systolic.groupednormal Systolic.groupedstage 1 hypertension
                         8.041750e-01                        1.094451e+00
Systolic.groupedstage 2 hypertension Systolic.groupedstage 3 hypertension
                         1.258920e+00                        7.404889e-02
            ConsciousnessDeep Stupor                    ConsciousnessComa
                         2.061761e+09                        2.434157e+01
                          PreviousYes                         TypeEmergency
                         3.757058e+00                        3.244790e+01
```

16.2 How strong is our model?

We will assess our model by dividing it into training and testing samples. This way, we can assess how good our model is at predicting results in the testing sample, while also further scrutinizing our training and testing model.

16.3 Testing & training the model

Below we have split our ICU dataset into training and testing datasets, with 150 observations in the former, and 50 in the latter.

Hide

```
> pacman::p_load(rsample,DT)
> set.seed(78)
> train_test_split <- initial_split(ICU)
> train <- training(train_test_split)
> test <- testing(train_test_split)
> (samp <- dim(train_test_split))
```

```
  analysis assessment          n          p
       150         50        200         25
```

Hide

```
> datatable(train)
```

Show 10 ▼ **entries** **Search:**

	ID	Status	Age	Sex	Race	Service	Cancer	Renal	Infection	CPR	Systolic	HeartRate	F
1	8	Lived	27	Female	White	Medical	No	No	Yes	No	142	88	No

	ID	Status	Age	Sex	Race	Service	Cancer	Renal	Infection	CPR	Systolic	HeartRate	F
3	14	Lived	77	Male	White	Surgical	No	No	No	No	100	70	No
4	28	Lived	54	Male	White	Medical	No	No	Yes	No	142	103	No
5	32	Lived	87	Female	White	Surgical	No	No	Yes	No	110	154	Ye
6	38	Lived	69	Male	White	Medical	No	No	Yes	No	110	132	No
9	42	Lived	35	Male	Black	Medical	No	No	No	No	108	60	No
10	50	Lived	70	Female	White	Surgical	Yes	No	No	No	138	103	No
12	53	Lived	48	Male	Black	Surgical	Yes	No	No	No	162	100	No
13	58	Lived	66	Female	White	Surgical	No	No	No	No	160	80	Ye
14	61	Lived	61	Female	White	Medical	No	Yes	No	No	174	99	No

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 ... 15 Next

Below we have conducted a logistic regression on our training dataset, using the finalized coefficients that we have deduced above. We can observe below that intercept, age, cancer, consciousness, service, previous and type are statistically significant. However, systolic is no longer statistically significant. Yet, the large difference between the null and residual deviance leads us with some confidence in our model.

Hide

```
> ICU.tr<-glm(Status~Age.binned+Service+Cancer+Systolic.grouped+Previous+Consciousness+Type, family=binomial(logit),data=train)
> summary(ICU.tr)
```

```
Call:
glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.23081  -0.45015  -0.07818  -0.00002   2.43434

Coefficients:
                                   Estimate Std. Error z value Pr(>|z|)
(Intercept)                        -40.1228  4921.3736  -0.008   0.9935
Age.binnedGroup 2                    0.2123     1.0653   0.199   0.8420
Age.binnedGroup 3                    0.3680     1.0733   0.343   0.7317
Age.binnedGroup 4                    0.3774     1.0534   0.358   0.7201
Age.binnedGroup 5                    1.5838     1.0429   1.519   0.1288
ServiceSurgical                     -1.2139     0.7177  -1.691   0.0908
CancerYes                            4.0976     1.6469   2.488   0.0128
Systolic.groupedhypotension         20.3716  4350.2391   0.005   0.9963
Systolic.groupednormal              17.8232  4350.2390   0.004   0.9967
Systolic.groupedstage 1 hypertension   18.5208  4350.2390   0.004   0.9966
Systolic.groupedstage 2 hypertension   18.6956  4350.2390   0.004   0.9966
Systolic.groupedstage 3 hypertension   15.7793  4350.2393   0.004   0.9971
PreviousYes                          1.0038     0.8410   1.194   0.2327
ConsciousnessDeep Stupor            40.7402  6647.7825   0.006   0.9951
ConsciousnessComa                    3.9932     1.5764   2.533   0.0113
TypeEmergency                       19.7312  2301.1604   0.009   0.9932

(Intercept)
Age.binnedGroup 2
Age.binnedGroup 3
Age.binnedGroup 4
Age.binnedGroup 5
ServiceSurgical                     .
CancerYes                           *
Systolic.groupedhypotension
Systolic.groupednormal
Systolic.groupedstage 1 hypertension
Systolic.groupedstage 2 hypertension
Systolic.groupedstage 3 hypertension
PreviousYes
ConsciousnessDeep Stupor
ConsciousnessComa                   *
TypeEmergency
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144.406  on 149  degrees of freedom
Residual deviance:  70.376  on 134  degrees of freedom
AIC: 102.38

Number of Fisher Scoring iterations: 19
```

Hide

```
> exp(coef(ICU.tr))
```

```
                       (Intercept)                  Age.binnedGroup 2
                      3.757317e-18                       1.236558e+00
                 Age.binnedGroup 3                  Age.binnedGroup 4
                      1.444786e+00                       1.458544e+00
                 Age.binnedGroup 5                    ServiceSurgical
                      4.873390e+00                       2.970322e-01
                         CancerYes         Systolic.groupedhypotension
                      6.019612e+01                       7.035111e+08
            Systolic.groupednormal Systolic.groupedstage 1 hypertension
                      5.501992e+07                       1.105283e+08
Systolic.groupedstage 2 hypertension Systolic.groupedstage 3 hypertension
                      1.316382e+08                       7.125962e+06
                       PreviousYes            ConsciousnessDeep Stupor
                      2.728531e+00                       4.934409e+17
                 ConsciousnessComa                      TypeEmergency
                      5.422580e+01                       3.708115e+08
```

16.4 Testing & training the model

As we can see below, where we have used the training model to predict the models in the testing dataset, we do get mixed results regarding the statistical significance of the predictors based on the CIs. For example, all predictors are positive as we can see above, yet only age binned group 5, cancer,consciousness coma, previous, and type have their lower CIs above 1. Intercept does not have a lower coefficient above 1, nor does systolic thus further cementing the fact that it is not statistically significant.

Hide

```
> predicted.val<-predict(ICU.tr, newdata=test)
> head(predicted.val)
```

```
        2          7          8         11         15         17
 -1.196692 -23.145574  -1.696054 -24.341391  -4.244406 -22.428841
```

Hide

```
> predicted.prob <- predict(ICU.tr, test, type = "response")
> head( predict( ICU.tr, test, type="response") )
```

```
          2           7           8          11          15
2.320642e-01 8.871659e-11 1.549813e-01 2.683295e-11 1.414141e-02
          17
1.816679e-10
```

Hide

```
> predicted.classes <- ifelse( predicted.prob > 0.5, "Lived", "Dead" )
> head(predicted.classes)
```

```
     2      7      8     11     15     17
"Dead" "Dead" "Dead" "Dead" "Dead" "Dead"
```

Hide

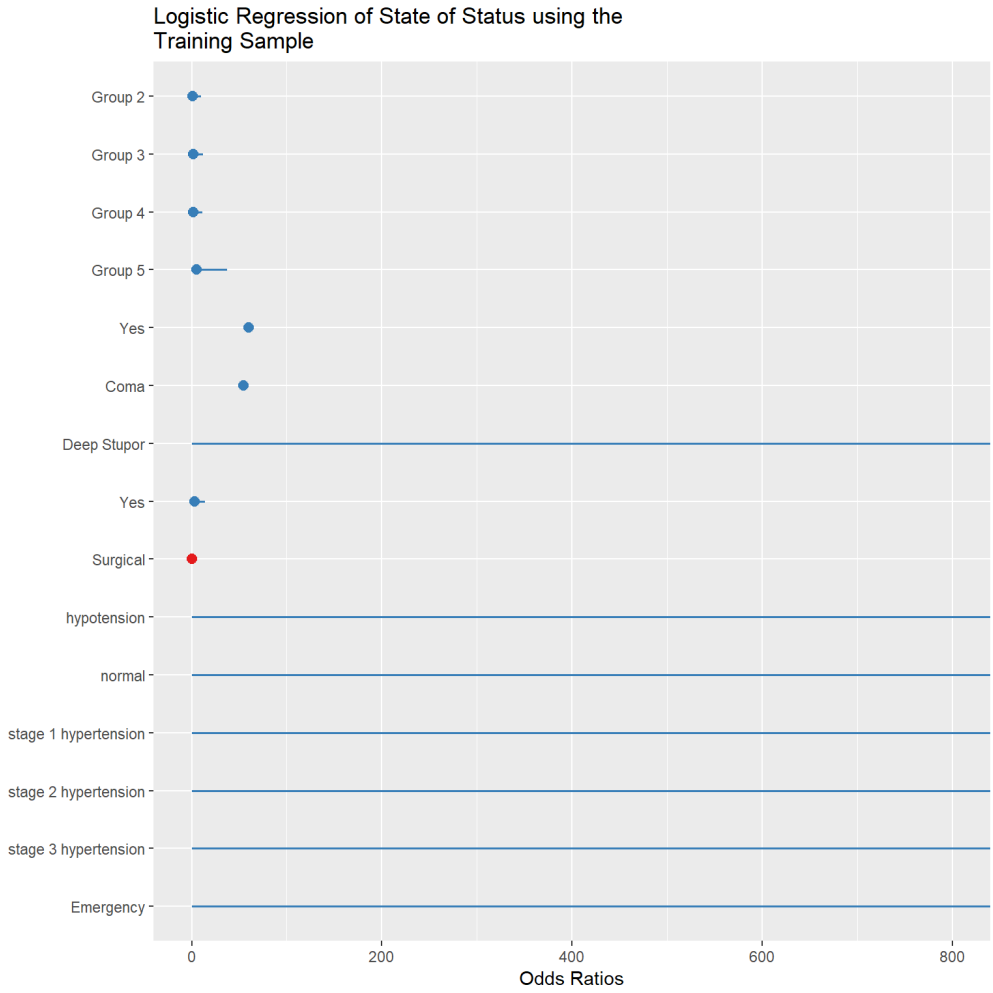
```
> tab_model(
+   ICU.tr,
+   title = "Logistic Regression Odds Ratios for Status using the Training Sample",
+   show.stat = TRUE,
+   digits = 3,
+   string.stat = "z-value",
+   string.p = "p (sig)",
+   show.fstat = TRUE,
+   show.dev = TRUE,
+   show.aic = TRUE,
+   CSS = list(
+     css.depvarhead = 'color: red;',
+     css.centeralign = 'text-align: left;',
+     css.firsttablecol = 'font-weight: bold;',
+     css.summary = 'color: blue;'
+   )
+ )
```

Logistic Regression Odds Ratios for Status using the Training Sample

		Status		
Predictors	Odds Ratios	CI	z-value	p (sig)
(Intercept)	0.000	0.000 – Inf	-0.008	0.993
Group 2	1.237	0.153 – 9.977	0.199	0.842
Group 3	1.445	0.176 – 11.842	0.343	0.732
Group 4	1.459	0.185 – 11.498	0.358	0.720
Group 5	4.873	0.631 – 37.628	1.519	0.129
Surgical	0.297	0.073 – 1.213	-1.691	0.091
Yes	60.196	2.386 – 1518.541	2.488	**0.013**
hypotension	703511090.034	0.000 – Inf	0.005	0.996
normal	55019918.417	0.000 – Inf	0.004	0.997
stage 1 hypertension	110528316.737	0.000 – Inf	0.004	0.997
stage 2 hypertension	131638249.462	0.000 – Inf	0.004	0.997
stage 3 hypertension	7125961.558	0.000 – Inf	0.004	0.997
Yes	2.729	0.525 – 14.184	1.194	0.233
Deep Stupor	4934408520868531 84.000	0.000 – Inf	0.006	0.995
Coma	54.226	2.468 – 1191.428	2.533	**0.011**
Emergency	370811484.789	0.000 – Inf	0.009	0.993
Observations		150		
Cox & Snell's R^2 / Nagelkerke's R^2		0.390 / 0.630		
Deviance		70.376		
AIC		102.376		

Hide

```
> plot_model(
+   ICU.tr,
+   title = "Logistic Regression of State of Status using the Training Sample",
+   show.p = TRUE
+ )  + ylim(0, 800)
```



With regards to the multicollinearity present within the model, according to the VIFs below, none are equal to 4, thus demonstrating the lack of multicollinarity in this model.

Hide

```
> pacman::p_load(car,corrplot)
> vf<-vif(ICU.tr)
> vf
```

```
                   GVIF Df GVIF^(1/(2*Df))
Age.binned       1.928971  4        1.085590
Service          1.182827  1        1.087579
Cancer           1.463451  1        1.209732
Systolic.grouped 2.173679  5        1.080736
Previous         1.310050  1        1.144574
Consciousness    1.637686  2        1.131248
Type             1.136135  1        1.065897
```

Although we have deduced that there is no multicollinearity present, we will now test the model against the null model, based on the training data. Below, we see that it's intercept is still significant.

Hide

```
> fit0<-glm(Status~1, family=binomial(logit), data=train)
> summary(fit0)
```

```
Call:
glm(formula = Status ~ 1, family = binomial(logit), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6428  -0.6428  -0.6428  -0.6428   1.8322

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.4718     0.2095  -7.024 2.16e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144.41  on 149  degrees of freedom
Residual deviance: 144.41  on 149  degrees of freedom
AIC: 146.41

Number of Fisher Scoring iterations: 4
```

Below, however, we attempt to fit all of the predictors, using the training data. We still see that the all of the predictors are statistically significant, thus demonstrating that this model is worth pursuing. This is further confirmed by the ANOVA chi-square conducted below, which is also statistically significant.

Hide

```
> fitall<-glm(Status~Age.binned+Service+Cancer+Systolic.grouped+Previous+Consciousness+Type, family=binomial(logit), data=train)
> summary(fitall)
```

```
Call:
glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.23081  -0.45015  -0.07818  -0.00002   2.43434

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -40.1228  4921.3736  -0.008   0.9935
Age.binnedGroup 2                  0.2123     1.0653   0.199   0.8420
Age.binnedGroup 3                  0.3680     1.0733   0.343   0.7317
Age.binnedGroup 4                  0.3774     1.0534   0.358   0.7201
Age.binnedGroup 5                  1.5838     1.0429   1.519   0.1288
ServiceSurgical                   -1.2139     0.7177  -1.691   0.0908
CancerYes                          4.0976     1.6469   2.488   0.0128
Systolic.groupedhypotension       20.3716  4350.2391   0.005   0.9963
Systolic.groupednormal            17.8232  4350.2390   0.004   0.9967
Systolic.groupedstage 1 hypertension   18.5208  4350.2390   0.004   0.9966
Systolic.groupedstage 2 hypertension   18.6956  4350.2390   0.004   0.9966
Systolic.groupedstage 3 hypertension   15.7793  4350.2393   0.004   0.9971
PreviousYes                        1.0038     0.8410   1.194   0.2327
ConsciousnessDeep Stupor          40.7402  6647.7825   0.006   0.9951
ConsciousnessComa                  3.9932     1.5764   2.533   0.0113
TypeEmergency                     19.7312  2301.1604   0.009   0.9932

(Intercept)
Age.binnedGroup 2
Age.binnedGroup 3
Age.binnedGroup 4
Age.binnedGroup 5
ServiceSurgical                   .
CancerYes                         *
Systolic.groupedhypotension
Systolic.groupednormal
Systolic.groupedstage 1 hypertension
Systolic.groupedstage 2 hypertension
Systolic.groupedstage 3 hypertension
PreviousYes
ConsciousnessDeep Stupor
ConsciousnessComa                 *
TypeEmergency
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 144.406  on 149  degrees of freedom
Residual deviance:  70.376  on 134  degrees of freedom
AIC: 102.38

Number of Fisher Scoring iterations: 19
```

Hide

```
> anova(fitall, fit0, test="Chisq")
```

```
Analysis of Deviance Table

Model 1: Status ~ Age.binned + Service + Cancer + Systolic.grouped + Previous +
    Consciousness + Type
Model 2: Status ~ 1
  Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
1       134     70.376
2       149    144.406 -15   -74.03 8.47e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have also conducted a Wald test to test the significance of the predictors in the training model as well. According to the results, cancer, type,consciousness, and previous are statistically significant.

Hide

```
> pacman::p_load(survey)
> regTermTest(ICU.tr, "Age.binned")
```

```
Wald test for Age.binned
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  0.714926  on  4  and  134  df: p= 0.58313
```

Hide

```
> regTermTest(ICU.tr, "Cancer")
```

```
Wald test for Cancer
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  6.190383  on  1  and  134  df: p= 0.014071
```

Hide

```
> regTermTest(ICU.tr, "Systolic.grouped")
```

```
Wald test for Systolic.grouped
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  1.456501  on  5  and  134  df: p= 0.20829
```

Hide

```
> regTermTest(ICU.tr, "Type")
```

```
Wald test for Type
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  7.352134e-05  on  1  and  134  df: p= 0.99317
```

Hide

```
> regTermTest(ICU.tr, "Service")
```

```
Wald test for Service
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  2.860549  on  1  and  134  df: p= 0.093101
```

Hide

```
> regTermTest(ICU.tr, "Consciousness")
```

```
Wald test for Consciousness
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  3.208147  on  2  and  134  df: p= 0.043556
```

Hide

```
> regTermTest(ICU.tr, "Previous")
```

```
Wald test for Previous
 in glm(formula = Status ~ Age.binned + Service + Cancer + Systolic.grouped +
    Previous + Consciousness + Type, family = binomial(logit),
    data = train)
F =  1.424478  on  1  and  134  df: p= 0.23478
```

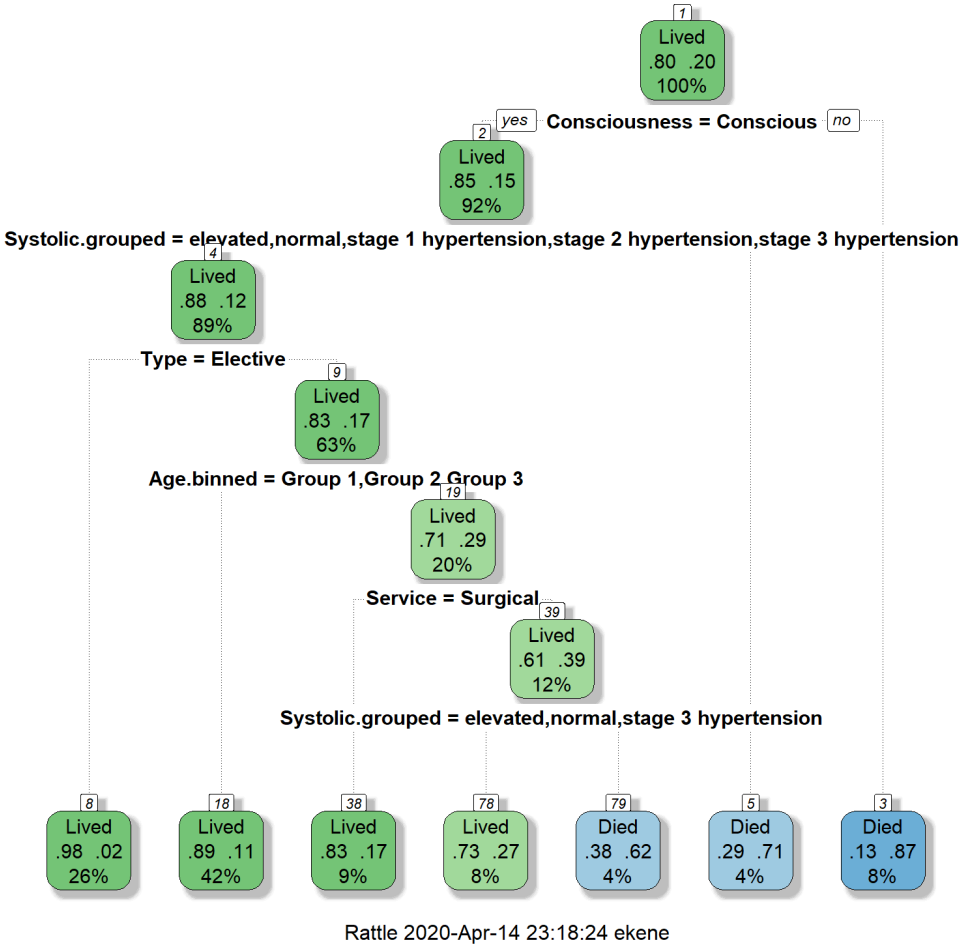
Now that we have deduced that our model is indeed robust, we will now move on to deduce our classificantion and regression trees, based on our variables of interest.

16.5 Classification and Regression Trees

As we can see below, the variables of interest are age, systolic, service and type. Pruning the tree doesn't seem plausible, as the xerror is steady from lines 2 to 4 and our cp graph denotes a stable cp from size 3 to 7, while the mean acuracy for our regression tree is 0.88.

Hide

```
> pacman::p_load(rpart, rpart.plot, rattle, dplyr)
> set.seed(2715)
>
> icutree=rpart(Status~Age.binned+Cancer+Systolic.grouped+Type+Consciousness+Type+Service+Renal+Infection+CPR+HeartRate.grouped+
Previous+Consciousness, method="class", data=ICU)
> fancyRpartPlot(icutree)
```



Hide

```
> printcp(icutree)
```

```
Classification tree:
rpart(formula = Status ~ Age.binned + Cancer + Systolic.grouped +
    Type + Consciousness + Type + Service + Renal + Infection +
    CPR + HeartRate.grouped + Previous + Consciousness, data = ICU,
    method = "class")

Variables actually used in tree construction:
[1] Age.binned       Consciousness    Service          Systolic.grouped
[5] Type

Root node error: 40/200 = 0.2

n= 200

      CP nsplit rel error xerror    xstd
1 0.2750      0     1.000  1.000 0.14142
2 0.0750      1     0.725  0.725 0.12449
3 0.0125      2     0.650  0.775 0.12795
4 0.0100      6     0.600  0.825 0.13123
```

Hide

```
> predicted.classes <- icutree %>% predict(, data=icu, type = "class")
> head(predicted.classes, 12)
```

```
    1     2     3     4     5     6     7     8     9    10    11    12
Lived Lived Lived Lived Lived Lived Lived Lived Lived Lived Lived Lived
Levels: Lived Died
```

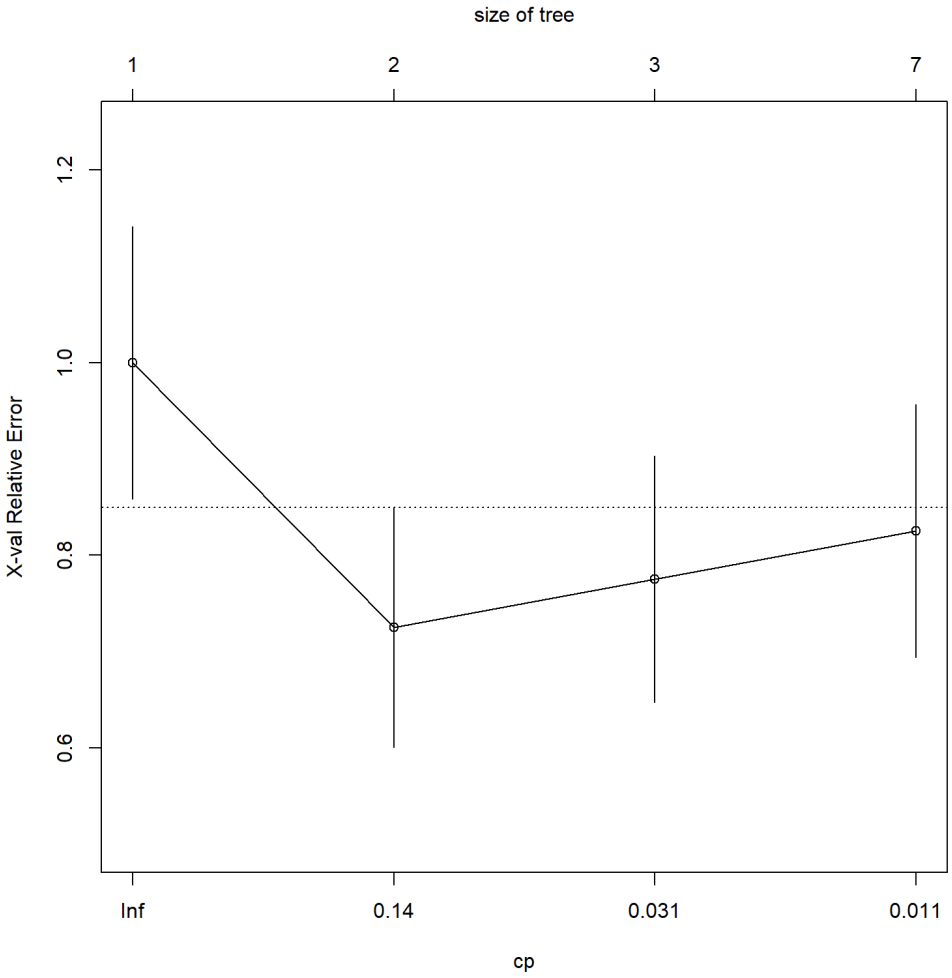
Hide

```
> mean(predicted.classes == ICU$Status)
```

```
[1] 0.88
```

Hide

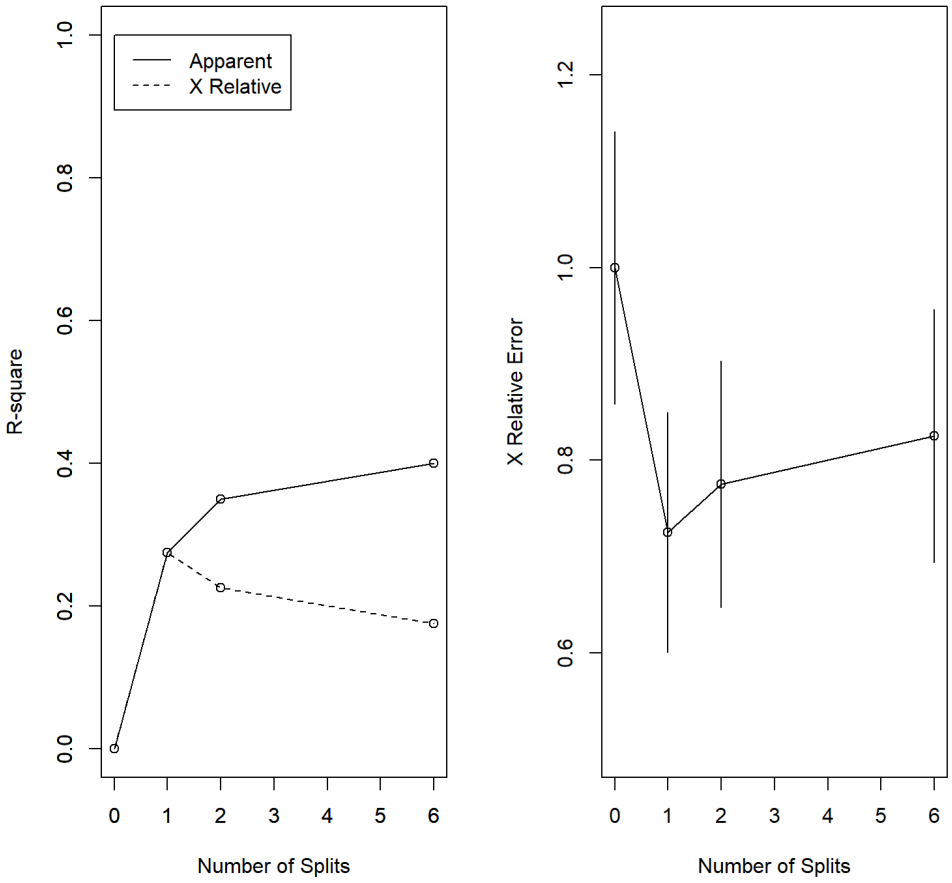
```
> plotcp(icutree)
```



Hide

```
> par(mfrow=c(1,2))
> rsq.rpart(icutree)
```

```
Classification tree:
rpart(formula = Status ~ Age.binned + Cancer + Systolic.grouped +
    Type + Consciousness + Type + Service + Renal + Infection +
    CPR + HeartRate.grouped + Previous + Consciousness, data = ICU,
    method = "class")

Variables actually used in tree construction:
[1] Age.binned       Consciousness    Service          Systolic.grouped
[5] Type

Root node error: 40/200 = 0.2

n= 200

      CP nsplit rel error xerror    xstd
1 0.2750      0     1.000  1.000 0.14142
2 0.0750      1     0.725  0.725 0.12449
3 0.0125      2     0.650  0.775 0.12795
4 0.0100      6     0.600  0.825 0.13123
```

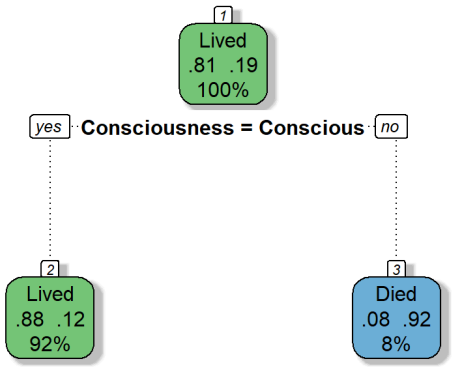


16.6 Classification tree built on training data

Below is our classification tree built on our training data. While our tree has 3 branches, the X error rises from line 2 to line 3, indicating that our training regression tree needs to be pruned, which according to our cp plot is around 2 splits.

Hide

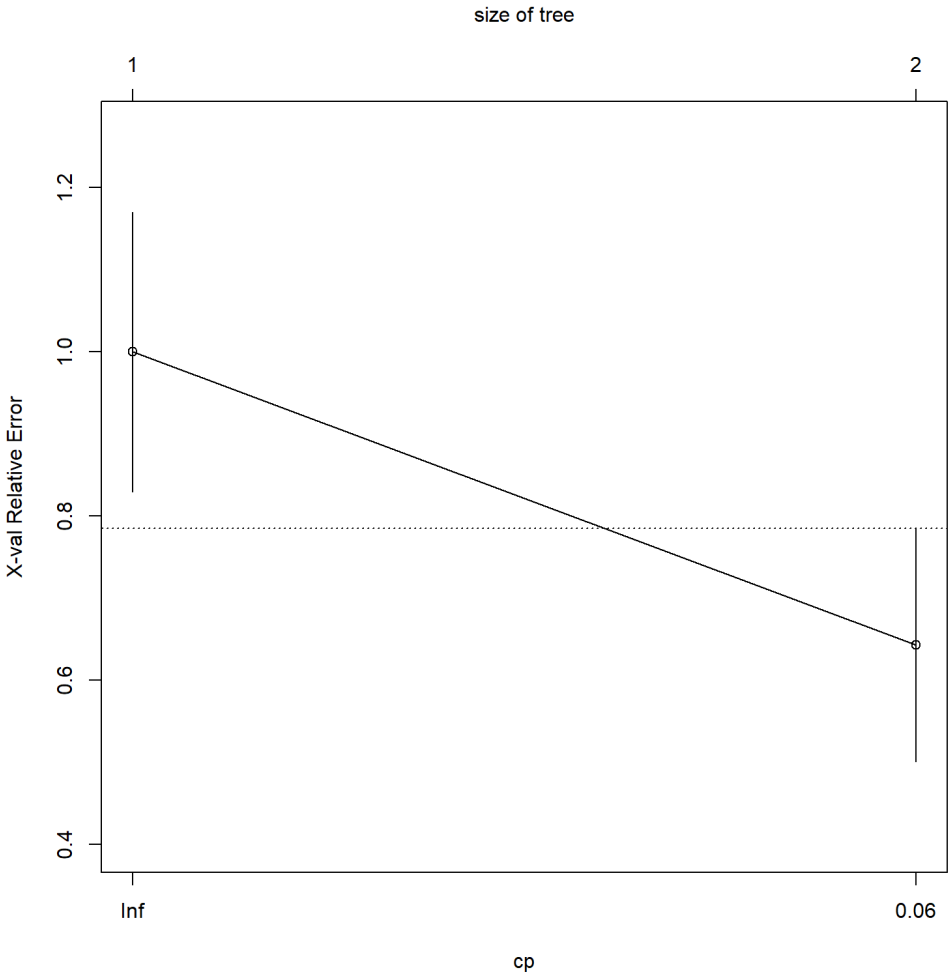
```
> icutree.train=rpart(Status~Service+Cancer+Previous+Type+Age.binned+Systolic.grouped+HeartRate.grouped+Consciousness+CPR+Infect
ion+Sex+Renal, method="class", data=train, minsplit=20)
> fancyRpartPlot(icutree.train)
```



Rattle 2020-Apr-14 23:18:28 ekene

Hide

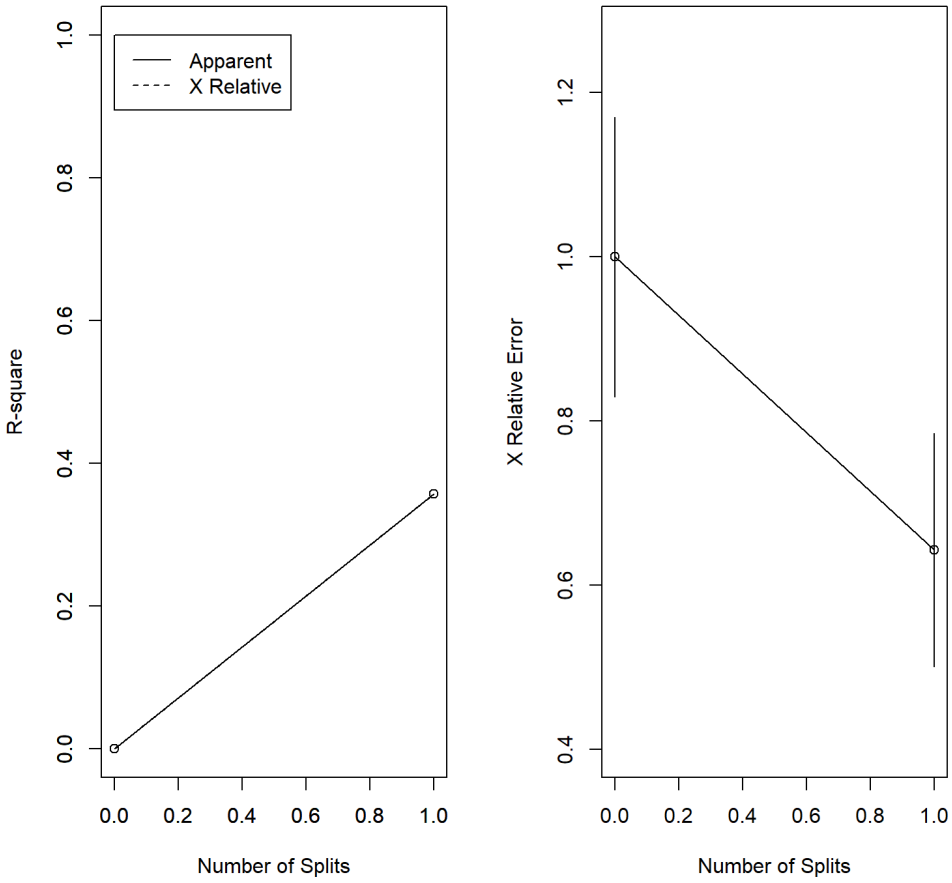
```
> plotcp(icutree.train)
```



Hide

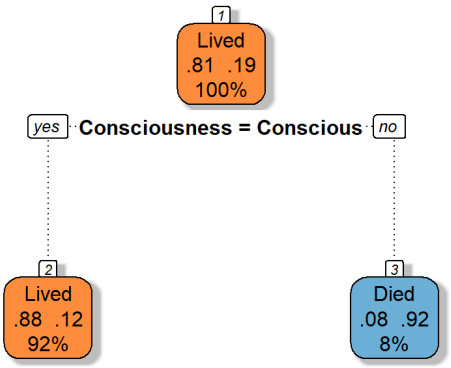
```
> par(mfrow=c(1,2))
> rsq.rpart(icutree.train)
```

```
Classification tree:
rpart(formula = Status ~ Service + Cancer + Previous + Type +
    Age.binned + Systolic.grouped + HeartRate.grouped + Consciousness +
    CPR + Infection + Sex + Renal, data = train, method = "class",
    minsplit = 20)

Variables actually used in tree construction:
[1] Consciousness

Root node error: 28/150 = 0.18667

n= 150

       CP nsplit rel error  xerror    xstd
1 0.35714      0   1.00000 1.00000 0.17043
2 0.01000      1   0.64286 0.64286 0.14214
```



Hide

```
> trainfit<-prune(icutree.train,cp=icutree.train$cptable[which.min(icutree.train$cptable[,"xerror"]),"CP"])
> fancyRpartPlot(trainfit, palettes = "YlOrRd" , caption=NULL  )
```



16.7 Prediction accuracy of our full model on training data

According to our results below, the prediction accuracy based on training data is statistically significant, with an 87 percent accuracy for predicting lived and dead patients in the ICU.

Hide

```
> tree.pred.train.full=predict(icutree.train, newdata=train, type="class")
>
> table_mat_full<-table(train$Status, tree.pred.train.full)
>
> sjt.xtab(train$Status, tree.pred.train.full, var.labels=c("Actual Status", "Predicted Status"), show.row.prc = TRUE)
```

Actual Status	*Predicted Status*		***Total***
	Lived	Died	
Lived	121 99.2 %	1 0.8 %	122 100 %
Died	17 60.7 %	11 39.3 %	28 100 %
Total	138 92 %	12 8 %	150 100 %

χ^2=40.706 · df=1 · φ=0.552 · Fisher's p=0.000

Hide

```
> (accuracy_Train_Test <- sum(diag(table_mat_full)) / sum(table_mat_full))
```

```
[1] 0.88
```

However, when we build our model on the testing data, we get 72% accuracy, and a non-significant result.

Hide

```
> tree.pred.test.full = predict(icutree.train, newdata=test, type = "class")
> table_mat_tr_full <- table(test$Status, tree.pred.test.full)
> sjt.xtab(test$Status, tree.pred.test.full, var.labels = c("Actual Status", "Predicted Status"), show.row.prc = TRUE)
```

Actual Status	*Predicted Status*		***Total***
	Lived	Died	
Lived	37 97.4 %	1 2.6 %	38 100 %
Died	10 83.3 %	2 16.7 %	12 100 %
Total	47 94 %	3 6 %	50 100 %

χ^2=1.183 · df=1 · φ=0.252 · Fisher's p=0.139

Hide

```
> (accuracy_Train_Test_full <- sum(diag(table_mat_tr_full)) / sum(table_mat_tr_full))
```

```
[1] 0.78
```

Finally, the classification rules based on the training sample are below:

Hide

```
> rpart.rules(trainfit)
```

```
 Status
   0.12 when Consciousness is           Conscious
   0.92 when Consciousness is Deep Stupor or Coma
```

Specifically, we should expect 13 percent to die based if consciousness is conscious, and 83 percent to die if consciousness is deep stupor or coma.

17 Discussion

17.1 Exploratory Data Analysis

From initially examining the data set, a number of findings about the attributes(variables) and relationships were observed. We can deduce from initial exploration of the study that race of most of the observations were White, the number of patients who lived and survived were higher than the population who died. More patients from the population who died were from age 50 and above for both male and female. The female population experienced more deaths than men of the same age bracket. The density plot peaked at age 75 for both male and female meaning more patients died at age 75. The relationship between vital status and consciousness shows that although there is a high count of patients who were conscious at admission, most patients that were received at a stage of deep stupor or coma lost consciousness and died.

Furthermore, most admitted patients had no chronic renal failure, no previous admissions to the ICU, fracture, CPR or cancer when admitted. Also most admitted patients had PO2 above or equal to 60, blood pH above or equal to 7.25, PCO2 below or equal to 45, and creatinine below or equal to 2. Due to the large difference between groups some of these variables were not further explored, and the focus of exploratory data analysis and understanding of distribution was on variables where values might have been prior to admission. These might aid clinicians who work in the ICU in understanding the likelihood of survival for individuals who are admitted before lab work has been ordered.

Additionally, large differences in gender and race, may limit how applicable this data set is to the general population. The frequency of type of admissions was unclear as the circumstances of admission to ICU in an elective setting can either be proactive or as a result of an emergency during surgery. More individuals who were elderly were admitted on an elective basis. It is unclear whether this meant they were proactively admitted due to fear of complication or complications occurred during treatment on an elective basis. Clarification of these two levels might provide more insight into how this attribute contributes to status. The distribution of age has two peaks which limits the value of the mean when looking at the whole dataset. This distribution might have affected other attribute counts, where there was often a peak at about 20 and a larger peak around 70 years of age. We investigate these findings further by taking a closer look at the distribution and associations.

17.2 Relationships

17.2.1 Distribution of Attributes

Most of the attributes explored further had non-uniform distribution, indicating larger/ significant differences in categories. Infection was narrowly significantly non-uniform - future research may show that infection might not disproportionately affect individuals admitted to ICU. More men than women were included in this dataset but it is unclear if that is an accurate distribution for people who are admitted to an ICU.

Statistically, all three numeric variables were non-normal. HeartRate and Systolic seemed to be normal based on quantile-quantile plots, but this conclusion was rejected based on the Shapiro-Wilk (SW) test. Given that the SW test can detect small effects in larger datasets, it is unclear whether this is an accurate conclusion to draw. Additionally, as per the central limit theorem, essentially all three of these distributions could be normal with repeated sampling. Therefore, normality testing did not add to the insights that could be drawn from this dataset.

17.2.2 Associations with Numeric Attributes

17.2.2.1 Age

Many of the attributes explored were more frequent with advancing age, which might be expected as the risk of cancer and comorbidity become higher as people age. This was further supported when comparing differences in means for age with the various variables. Statistically significant differences in means for age were found for infection, vital status, ICU admission, type of admission and history of chronic renal failure. While these statistically significant findings may suggest that there is a link to age, further analysis is required to assess whether the relationships are statistically significant.

Surgical service peaks at a younger age, indicating the conditions that result in admissions from this group might have been acute and accident related rather than related to a medical condition. Based on plots of sex by age, it might be concluded that more females who were older were admitted to the ICU as compared to males. Based on plots of status by sex, a higher frequency of deaths occurred in advancing age, a conclusion that seems reasonable as individuals become frail, require surgery or are diagnosed with other medical conditions, including cancer.

Additionally, through the difference of proportions test, although age was not statistically significant by accordance of the p-value, the proportion numbers show that the number of patients that died in the two youngest age groups is much lower than the rest and the greatest proportion of deaths was in the oldest age group. It is logical that age would be a predictor of vital status, especially in an ICU.

When status was examined with status and sex, it was found that more older women were admitted to the ICU and older women also died more often during ICU admission. There was a similar frequency of older men who lived and died, with more men dying with older age.

17.2.3 Associations with Numeric Attributes

17.2.3.1 HeartRate

The distribution of this attribute shows a second peak at 130 bpm in addition to a primary peak at about 90 bpm, which would be considered a normal heart rate. When examined by status, there seemed to be more deaths as heart rate increased. Whether this is related to the reason for admission or to the anxiety of having a medical emergency and whether elevated heart rate contributed to death is unclear. Despite these observations, when assessing for differences in means, there was no statistical difference between the two status groups nor was there a statistical significance in terms of proportions between vital status and heart rate.

17.2.4 Associations with Numeric Attributes

17.2.4.1 Systolic

This attribute seems to be visually normally distributed, a conclusion that was rejected when testing for normality, albeit narrowly. When examined by status, there are two peaks at 75 and 140mmHg among those who died. This suggests that hypotension and hypertension may have contributed to death during admission. This could be related to a bleed, sepsis or even dehydration in hypotension and stress, hypertension or hypertensive crisis with higher readings. Critical care often involves the use of vasopressors to increase blood pressure and perfusion to manage hypotension. When means were examined, there was a statistically significant difference in means for status by systolic blood pressure. As this dataset did not include any history about these patients, conclusions cannot be made about what conditions could have resulted in these readings on admission.

Systolic blood pressure was also found to be statistically significant through the difference in proportions test. It was found that the hypotension group had a much higher proportion of patient deaths in comparison to the other systolic blood pressure categories. Again, this is an understandable predictor due to the life threatening nature of some of the potential causes of hypotension.

17.3 Differences in Proportions for Categorical Attributes

From the tests of equal proportions, we note that whether patients were receiving medical or surgical service at ICU admission was a statistically significant predictor of the vital status of the patient. The proportions of patients that died was higher for those that were there for a medical service was higher than those there for a surgical service. This makes sense since more patients in the ICU for a medical service could be admitted due to a life threatening emergency than those in the ICU for surgery, who are often scheduled to be admitted in advance. Thus, this could explain the higher death rate in medical service receivers in comparison to surgery receivers.

We also found that type of admission was a significant predictor of vital status in ICU patients, when conducting equal proportions tests. Patients that were admitted due to an emergency had a higher rate of death than those admitted electively. Again, this is understandable since patients admitted in an emergency would likely be in life-threatening condition and, thus, would be more likely to die than those who are scheduled to be admitted to the ICU.

The third variable we found to be a statistically significant predictor of vital status upon conducting the equal proportions tests was the patient's level of consciousness at admission. Those who had a deep stupor upon admission were the highest proportion of patients to die, with all patients with a deep stupor dying in the ICU. Those who had a coma had a 80% death rate and those who had no stupor or coma had the lowest rate of death, which makes sense. It is interesting that the deep stupor group had a higher death rate since they had less patients in that group than the coma group and a coma is considered to be more medically severe than a deep stupor, thus you would expect the coma group to have a higher proportion of deaths.

On the other hand, sex, infection at admission, a history of chronic renal failure, the reception of CPR upon admittance, a cancer diagnosis and prior ICU admission within the last 6 months were not deemed to be significant predictors of the vital status of patients. It is important to note that the dataset includes more males than females, which reduces the integrity of the proportions test on sex and vital status.

Additionally, we thought it was interesting that infection at admission, a history of chronic renal failure, cancer and prior ICU admission were not significant predictors. But we concluded that it would depend on the severity of the patient's condition.

Lastly, we also followed up our results with a Goodman Kruskall's Lambda analysis, and found that while the relationship between the predictors we focused on and the response variable was not 0 as it was approximately 0.25, it also was not 1 therefore denoting a somewhat weak relationship.

17.4 Logistic Regression

When we conducted a logistic regression, the statistically significant indicators that were found included age, cancer, systolic blood pressure, type, service, previous, and consciousness. However, when the logistic regression was conducted on our training and testing sample, we found systolic was no longer significant. With regards to determining how good our model was, we used the null model; due to the fact that it was found to be significant, it therefore denoted that our model is worth pursuing. With regards to determining the importance of the predictors within our model, the Wald Test found that consciousness, systolic blood pressure, type, age and service to be significant predictors within the model.
Our final model deduced was Status=(1.118927e-03)+(6.130498e+00) Age.binnedGroup4+(1.025275e+01)Age.binnedGroup5+(3.130311e-01)ServiceSurgical+(2.183970e+01)CancerYes+(6.624874e+00)PreviousYes+(3.463936e+01)ConsciousnessComa+(2.575717e+01)TypeEmergency

17.5 Possibility of survival based on classification and regression trees

In both our original classification and regression tree and our finalized classification and regression tree based on our training data, consciousness was a key player in a patient's status. Whereas systolic, type, age and service were key categories in the classification and regression tree conducted on the original dataset, our training dataset required pruning, which eliminated all branches but consciousness and maintained an accuracy of approximately 87 percent.

17.6 Comments regarding our dataset & analysis

Due to the difference in results between the classification and regression trees with both the training data and the original dataset, along with the difference in models in both the training data and original dataset, there is doubt that both our models and dataset are representative of true "cases" observed within a typical ICU unit, not to mention other anomalies present within our data as previously mentioned. Furthermore, we attempted to deduce a logistic regression based on historical variables, which also contributed to our unique results.

18 Conclusion

Our analysis on the relationship between variables related to patient history and status within the ICU. Common predictors that were found to be statistically significant and thus pertinent, among all tests include systolic, service, and consciousness. Recommendations include conducting the study on a more diverse population with a large number of observations, along with more research to be conducted on the systolic blood pressure, type of service and consciousness on status within an ICU stay.

A work by Ekene Olatunji, Catherine Nassralla, Victoria Chin, Basma Chamas, and Sajiya Somji

Msc. eHealth

DeGroote School of Business

eHealth 705 Statistics for eHealth - Final Project