Team Project Multivariate Analysis

Adrian White, Cesar Conejo, Xavier Bryant

11/14/2020

Team members

- Adrian White: 1004391004
- Cesar Conejo: 100443596 (Representative member)
- Xavier Bryant: 100445659

Introduction data set

We have selected the CRASH-2 data set provided by Vanderbilt School of Biostatistics for our project. It describes the outcome of a randomized controlled trial and economic valuation of the effects of tranexamic acid on death, vascular occlusive events and transfusion requirement in bleeding trauma patients. Tranexamic acid reduces bleeding in trauma patients undergoing surgery, but is an expensive treatnment option. The trial's objective was to assess the effects and cost effectiveness of an early administration of this medication.

Participants of the study were adults with, or at risk of, significant bleeding within 8 hours of injury. Sample randomization was determined by the allocation of an eight digit sequence randomly generated by a computer. **Patients and staff were masked to treatment allocation of the tranexamic acid. Health outcome was measured by number of life-years (LYs) gained, while cost data was obtained from the hospitals and measured by international collars ($) per LY. The trial was undertaken by two hundred and seventy-four hospitals world wide.**

Some interesting links related with this data set are:

- https://pubmed.ncbi.nlm.nih.gov/23477634/
- https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60835-5/fulltext

Summary variables in the data set

The variables in this dataset are the following: **add units of measurements**

- entryid: (Numerical) Unique Numbers for Entry Forms
- sex: (Boolean) The sex of the patient
- age : (Numerical) Age of the patient
- injurytime: (Numerical) Hours since injury
- injurytype: (Categorical) Type of injury {Blunt, Penetrating, Blunt and Penetrating}
- sbp: (Numerical) Systolic Blood Pressure
- rr: (Numerical) Respiratory Rate
- cc: (Numerical) Central Capillary Refille Time
- hr: (Numerical) Heart Rate

- ndaysicu: (Numerical) Number of days in ICU
- btransf: (Boolean) Blood Products Transfusion
- ncell: (Numerical) Number of Units of Red Call Products Transfused
- nplasma: (Numerical) Number of Units of Fresh Frozen Plasma Transfused
- nplatelets: (Numerical) Number of Units of Platelets Transfused
- ncryo: (Numerical) Number of Units of Cryoprecipitate Transfused
- bvii: (Boolean) Recombinant Factor VIIa Given
- Death: (Boolean) 0 if the participant is death at the end of the trial, 1 alive.
- bloading (Boolean) Complete Loading Dose of Trial Drug Given.

A summmary of data type is the following

variable	type_variable	sub_type_variable
entryid	Quantitative	Continuous
sex	Qualitative	Nominal
age	Quantitative	Continuous
injurytime	Quantitative	Continuous
injurytype	Qualitative	Nominal
sbp	Quantitative	Continuous
rr	Quantitative	Continuous
cc	Quantitative	Continuous
hr	Quantitative	Continuous
ndaysicu	Quantitative	Discrete
btransf	Quanlitative	Nominal
ncell	Quantitative	Discrete
nplasma	Quantitative	Discrete
nplatelets	Quantitative	Discrete
ncryo	Quantitative	Discrete
bvii	Qualitative	Nominal
death	Qualitative	Nominal
bloading	Qualitative	Nominal

Summary and Graphical display

A review of the structure of the dataset is the following:

```
## 'data.frame':    9497 obs. of  18 variables:
##  $ entryid   : int  1 3 4 6 7 8 9 11 12 14 ...
##  $ sex       : Factor w/ 2 levels "male","female": 1 1 1 1 1 1 1 1 1 2 ...
##  $ age       : int  50 30 40 19 27 16 29 41 56 37 ...
##  $ injurytime: num  1 1 2 3 0.5 1 1 0.5 0.5 8 ...
##  $ injurytype: Factor w/ 3 levels "blunt","penetrating",..: 1 1 2 2 2 2 1 2 1 2 ...
##  $ sbp       : int  75 70 60 90 90 90 116 120 60 104 ...
##  $ rr        : int  28 26 20 30 26 28 15 15 9 23 ...
##  $ cc        : int  5 6 5 5 5 2 3 3 3 5 ...
##  $ hr        : int  120 130 120 90 96 118 118 70 100 92 ...
##  $ ndaysicu  : num  0 6 2 9 7 0 7 7 23 2 ...
##  $ btransf   : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
##  $ ncell     : num  1 2 4 2 1 1 16 8 4 4 ...
##  $ nplasma   : int  0 0 0 0 0 0 9 11 9 0 ...
##  $ nplatelets: int  0 0 0 0 0 0 22 10 0 0 ...
##  $ ncryo     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ bvii      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ death     : Factor w/ 2 levels "0","1": 2 1 2 2 1 1 1 1 1 1 ...
##  $ bloading  : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
```

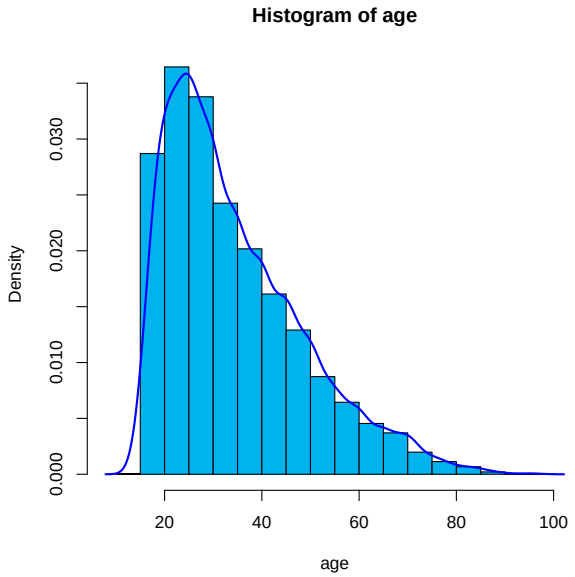
A summary of the values in the data set are:

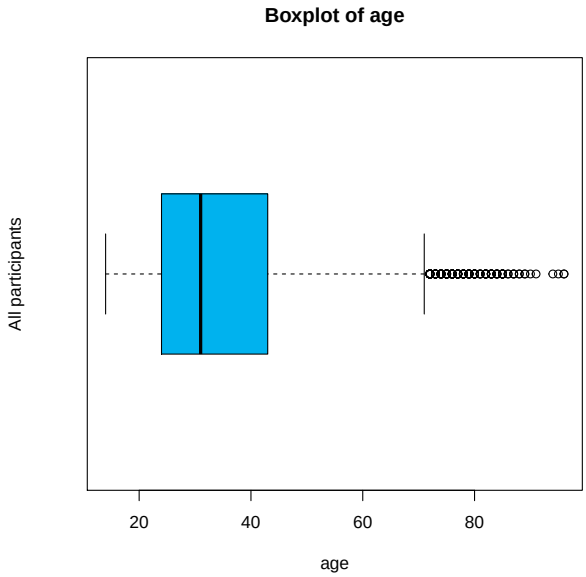
```
##     entryid          sex            age         injurytime
##  Min.   :    1   male  :7906   Min.   :14.0   Min.   : 0.10
##  1st Qu.: 4720   female:1591   1st Qu.:24.0   1st Qu.: 1.00
##  Median : 9333                 Median :31.0   Median : 3.00
##  Mean   : 9657                 Mean   :34.7   Mean   : 3.09
##  3rd Qu.:14598                 3rd Qu.:43.0   3rd Qu.: 4.50
##  Max.   :20270                 Max.   :96.0   Max.   :48.00
##                  injurytype        sbp              rr             cc
##  blunt                :5211   Min.   :  4.0   Min.   : 2.0   Min.   : 1.00
##  penetrating          :2937   1st Qu.: 80.0   1st Qu.:20.0   1st Qu.: 2.00
##  blunt and penetrating:1349   Median : 90.0   Median :22.0   Median : 3.00
##                               Mean   : 93.1   Mean   :23.5   Mean   : 3.44
##                               3rd Qu.:104.0   3rd Qu.:28.0   3rd Qu.: 4.00
##                               Max.   :225.0   Max.   :91.0   Max.   :20.00
##        hr          ndaysicu     btransf      ncell          nplasma
##  Min.   :  3   Min.   : 0.00   0:  12   Min.   : 0.00   Min.   : 0.00
##  1st Qu.: 96   1st Qu.: 0.00   1:9485   1st Qu.: 2.00   1st Qu.: 0.00
##  Median :110   Median : 1.00            Median : 3.00   Median : 0.00
##  Mean   :108   Mean   : 4.14            Mean   : 3.91   Mean   : 1.44
##  3rd Qu.:120   3rd Qu.: 5.00            3rd Qu.: 5.00   3rd Qu.: 1.00
##  Max.   :220   Max.   :58.00            Max.   :60.00   Max.   :60.00
##    nplatelets        ncryo       bvii     death    bloading
##  Min.   : 0.00   Min.   : 0.00   0:9456   0:7672   0:  39
##  1st Qu.: 0.00   1st Qu.: 0.00   1:  41   1:1825   1:9458
##  Median : 0.00   Median : 0.00
##  Mean   : 0.54   Mean   : 0.26
##  3rd Qu.: 0.00   3rd Qu.: 0.00
##  Max.   :87.00   Max.   :61.00
```

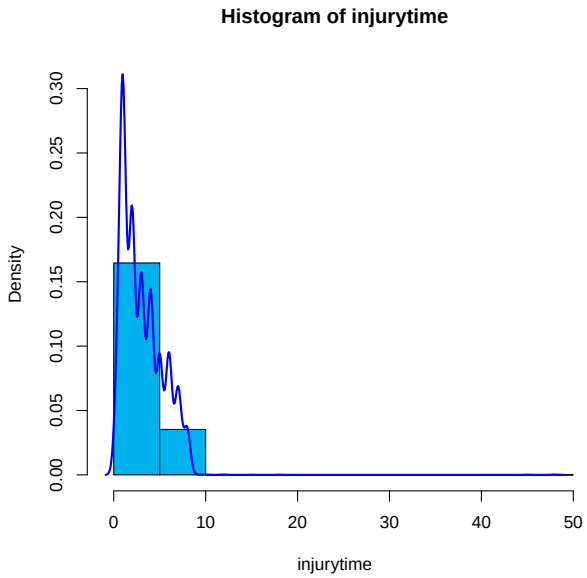
Finally, the list of different values by column is the following:

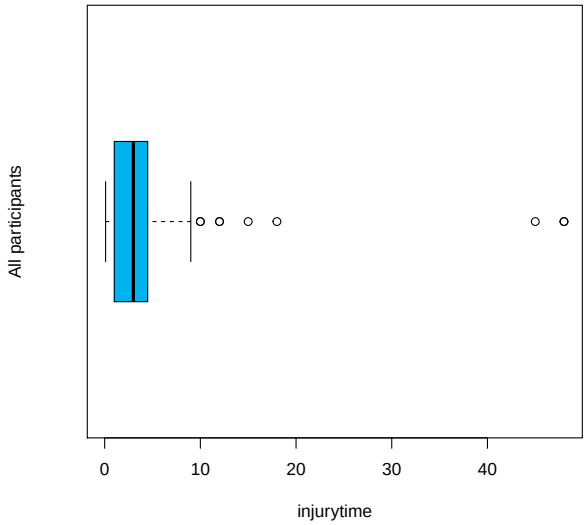
```
##     entryid         sex         age injurytime injurytype         sbp          rr
##        9497           2          81          78          3         153          58
##          cc          hr    ndaysicu    btransf      ncell    nplasma nplatelets
##          16         154          47          2         47          45          39
##       ncryo        bvii       death   bloading
##          28           2           2          2
```

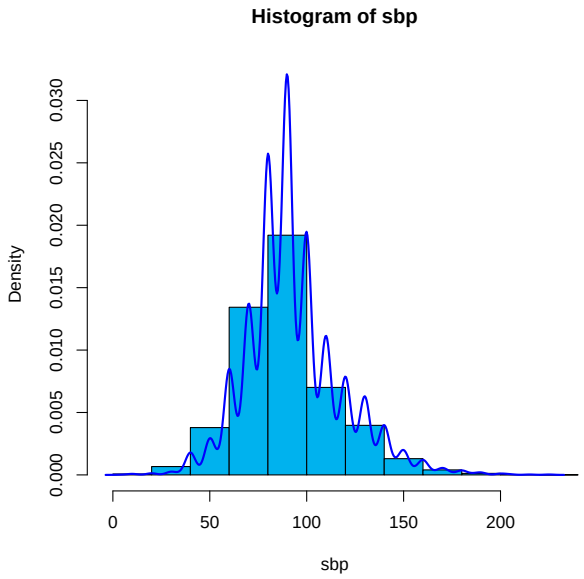
Some visualizations of the quantitative variables are:

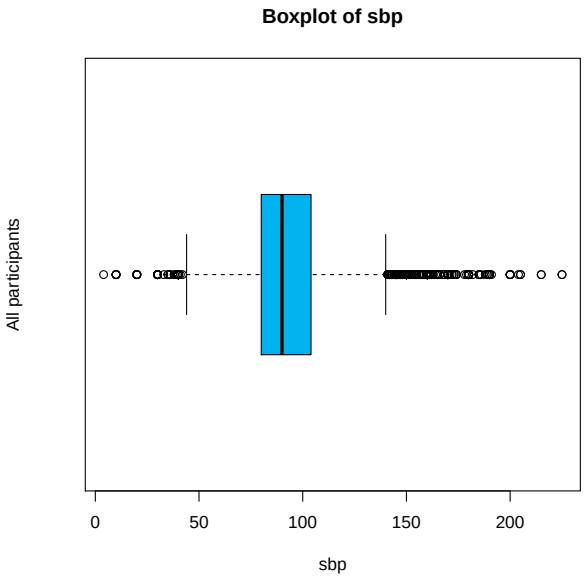


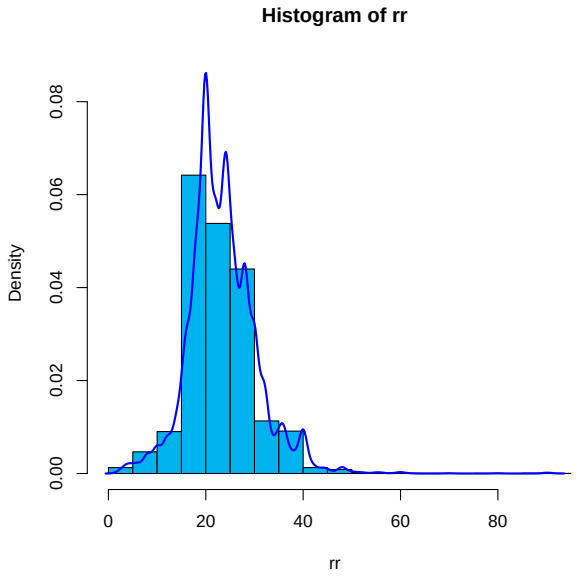


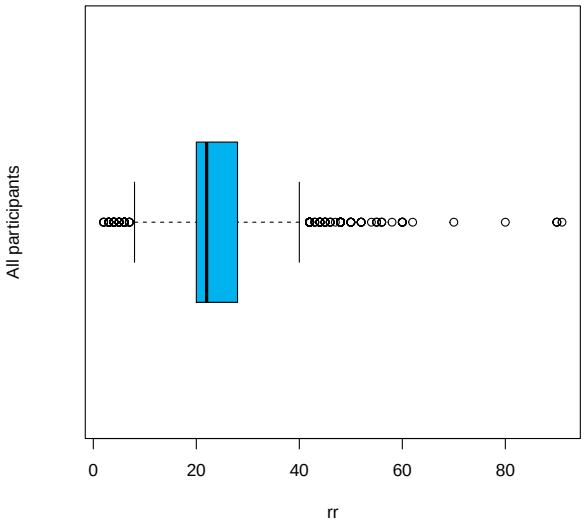


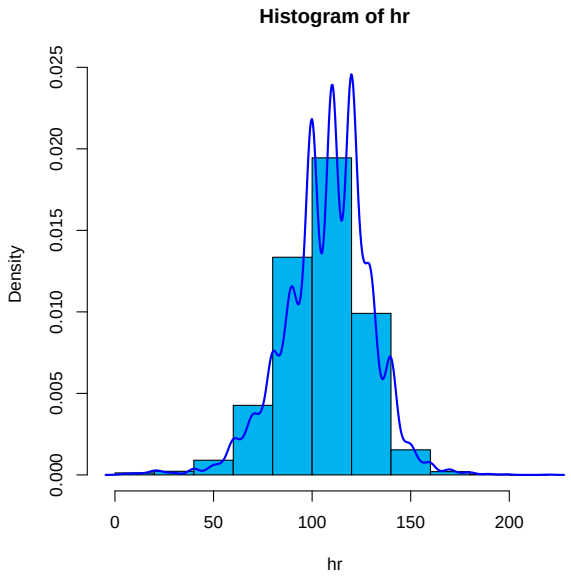


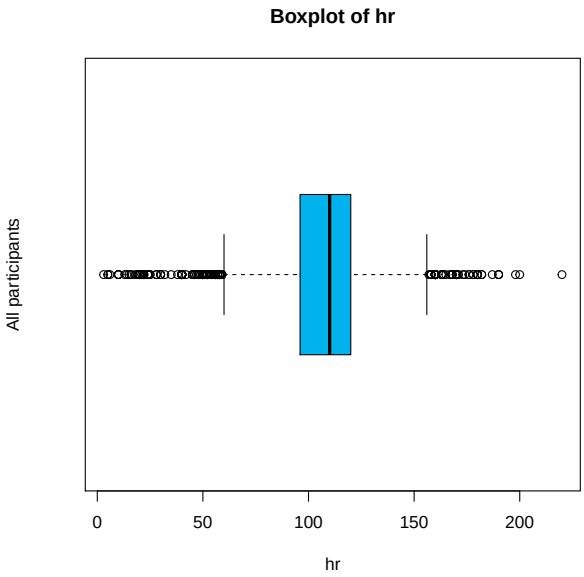

Histogram of sbp
Density
0.000
0.005
0.010
0.015
0.020
0.025
0.030
0
50
100
150
200
sbp

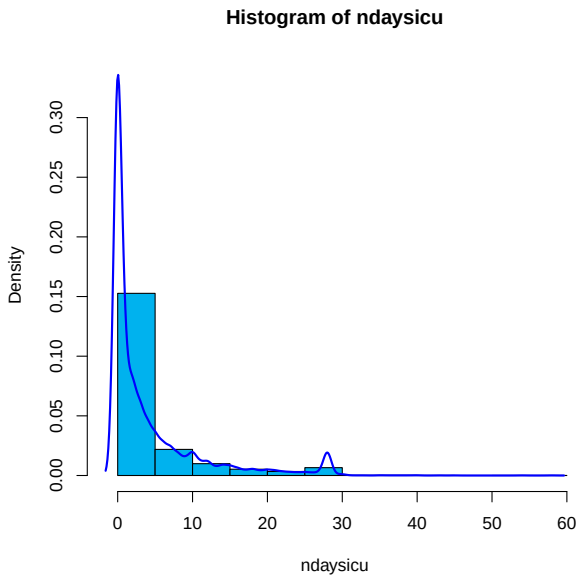

Boxplot of sbp
All participants
0
50
100
150
200
sbp

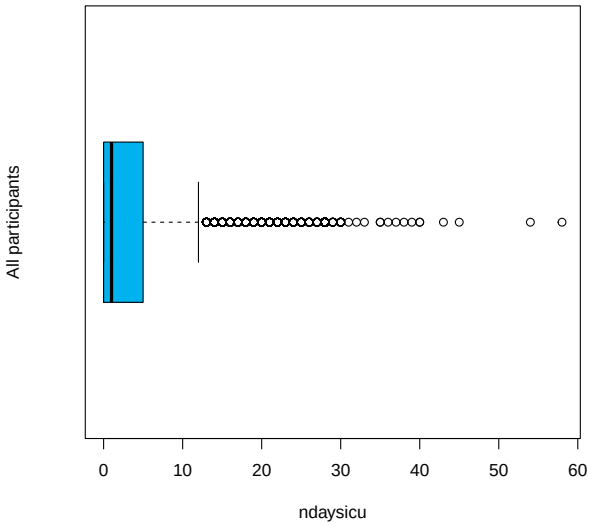

Histogram of rr
Density
0.00
0.02
0.04
0.06
0.08
0
20
40
60
80
rr

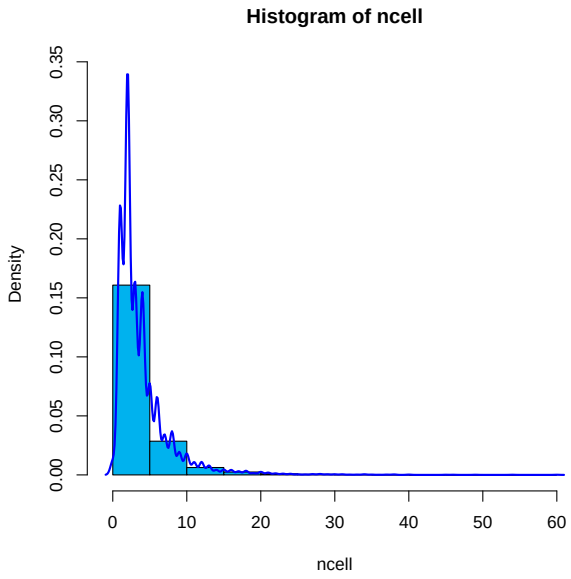

Boxplot of rr
All participants
0
20
40
60
80
rr

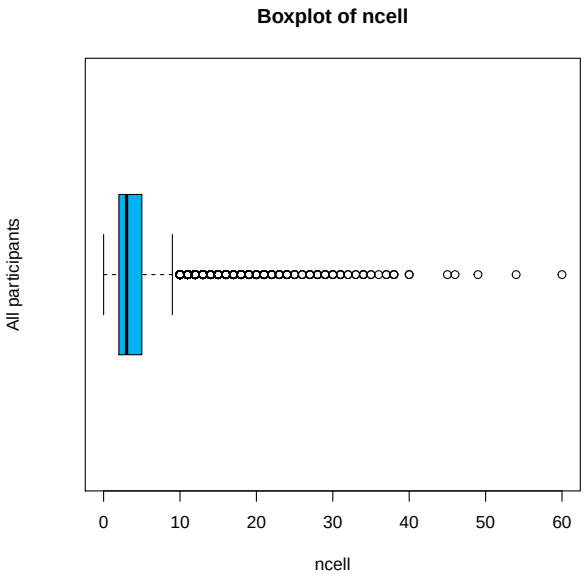

Histogram of hr
Density
0.000
0.005
0.010
0.015
0.020
0.025
0
50
100
150
200
hr

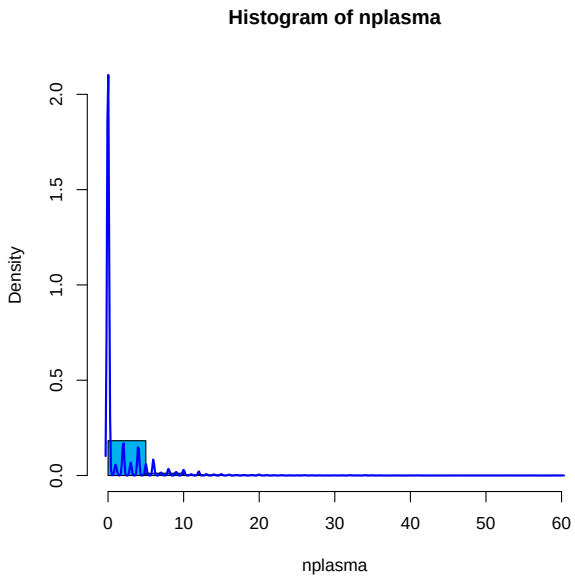

Boxplot of hr
All participants
0
50
100
150
200
hr

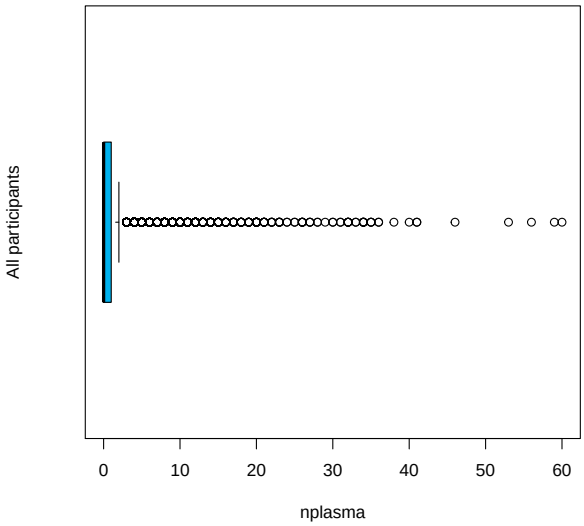

Histogram of ndaysicu
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
10
20
30
40
50
60
ndaysicu

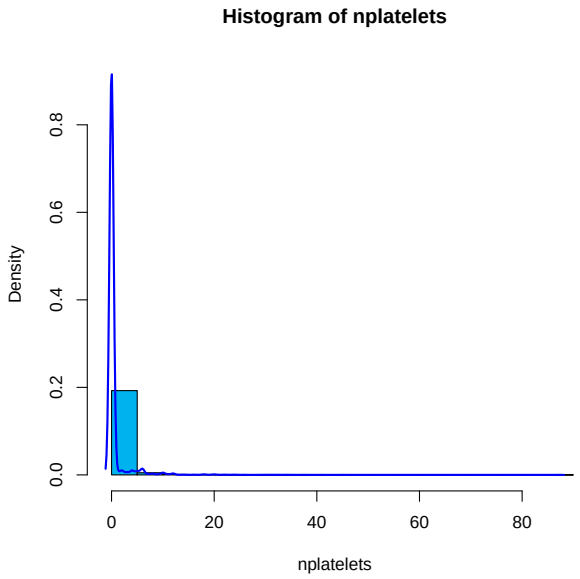

Boxplot of ndaysicu
All participants
0
10
20
30
40
50
60
ndaysicu

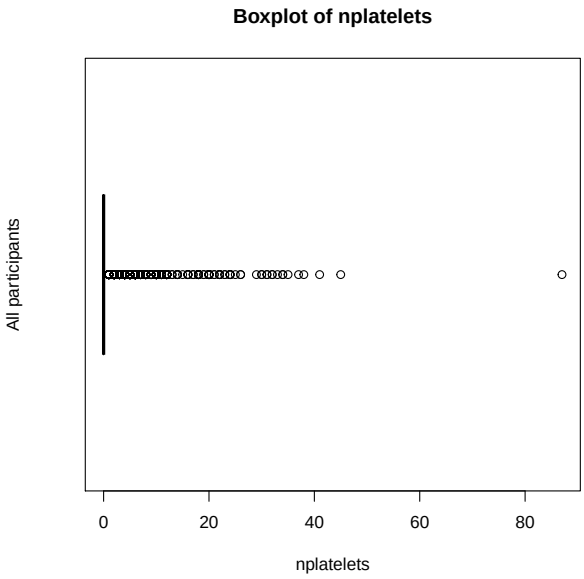

Histogram of ncell
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50
60
ncell

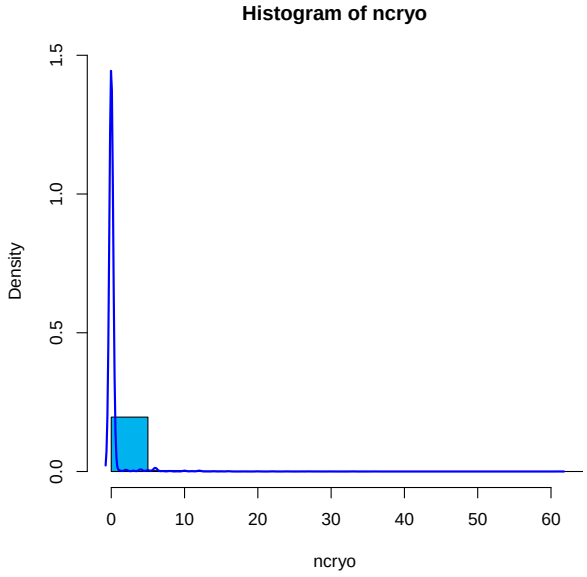

Boxplot of ncell
All participants
0
10
20
30
40
50
60
ncell

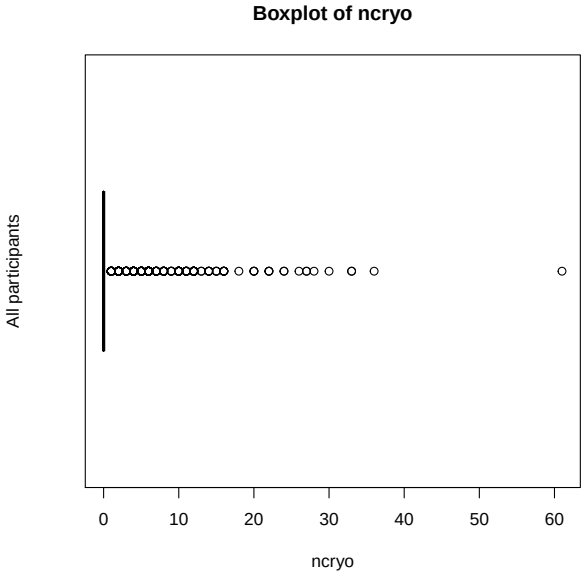

Histogram of nplasma
Density
0.0
0.5
1.0
1.5
2.0
0
10
20
30
40
50
60
nplasma


Boxplot of nplasma
All participants
0
10
20
30
40
50
60
nplasma


Histogram of nplatelets
Density
0.0
0.2
0.4
0.6
0.8
0
20
40
60
80
nplatelets


Boxplot of nplatelets
All participants
0
20
40
60
80
nplatelets


Histogram of ncryo
Density
0.0
0.5
1.0
1.5
0
10
20
30
40
50
60
ncryo


Boxplot of ncryo
All participants
0
10
20
30
40
50
60
ncryo