EDA

Chen Liang

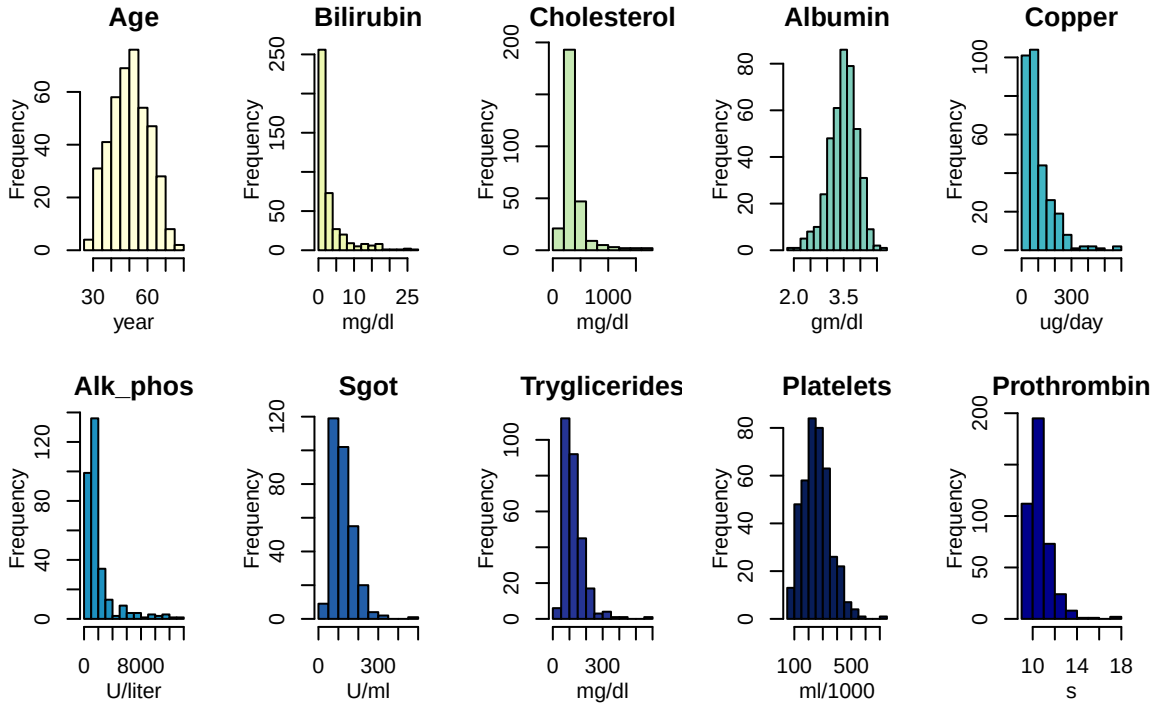
2024-11-16

Load data

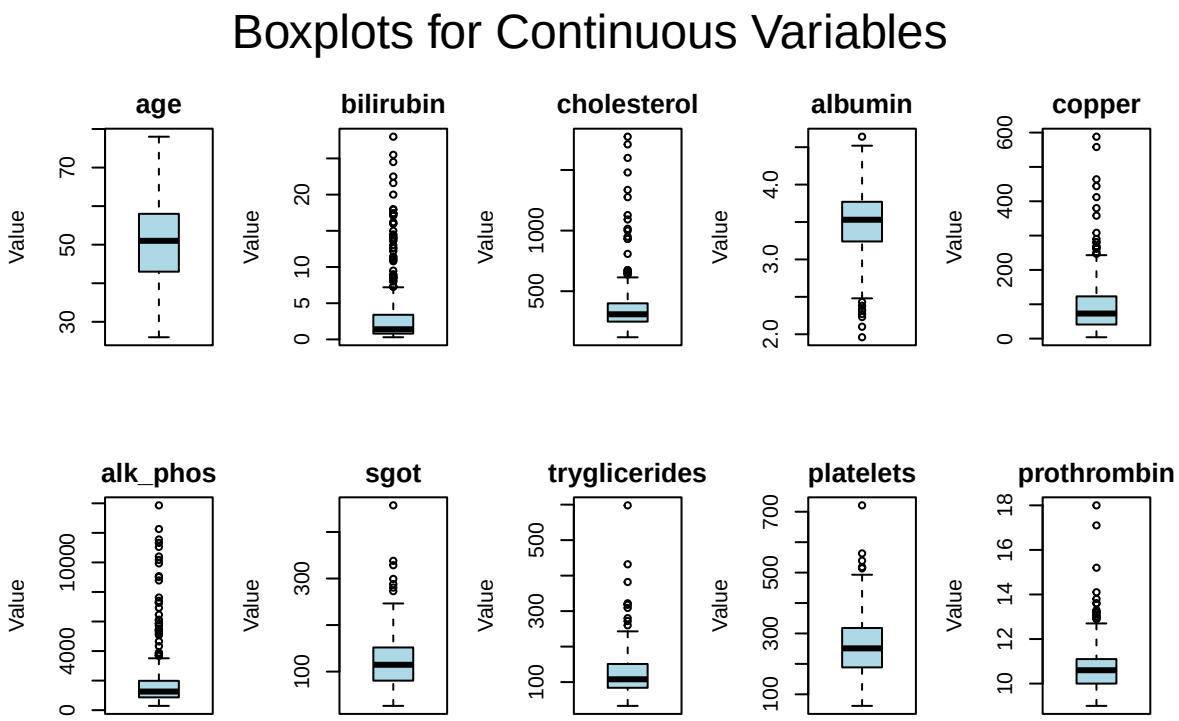
```
## Rows: 418 Columns: 20
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): Status, Drug, Sex, Ascites, Hepatomegaly, Spiders, Edema
## dbl (13): ID, N_Days, Age, Bilirubin, Cholesterol, Albumin, Copper, Alk_Phos...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
##            id        n_days        status          drug           age
##             0             0             0           106             0
##           sex       ascites  hepatomegaly       spiders         edema
##             0           106           106           106             0
##     bilirubin   cholesterol       albumin        copper      alk_phos
##             0           134             0           108           106
##          sgot tryglicerides     platelets   prothrombin         stage
##           106           136            11             2             6
```

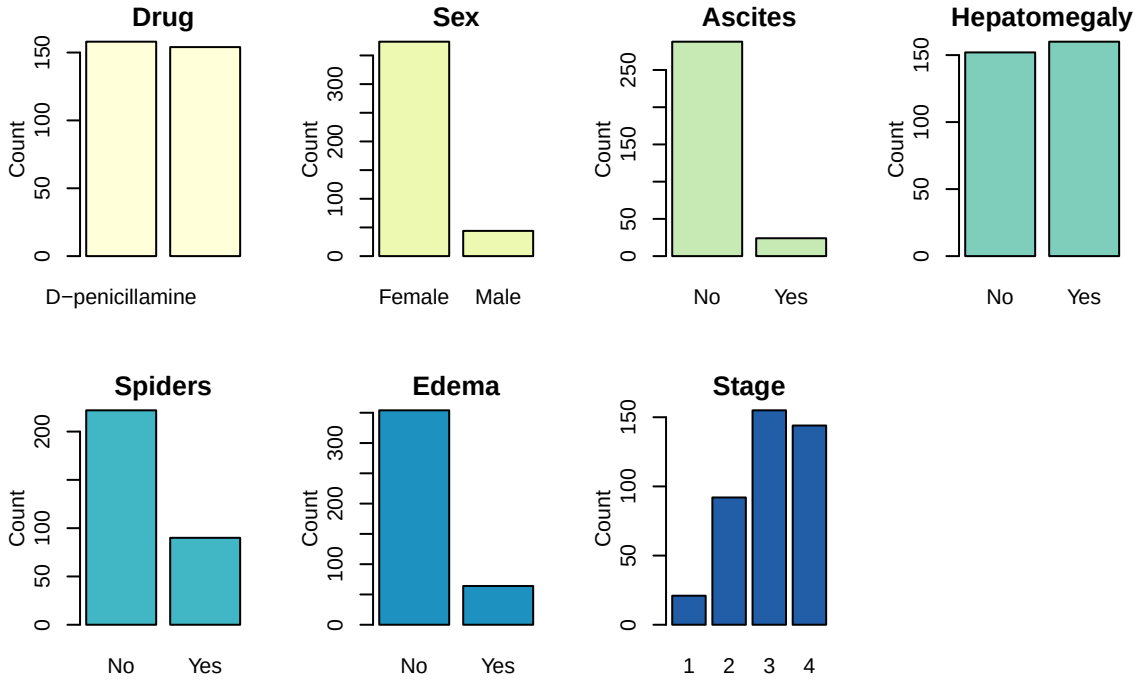
Historgram Plots for continuouse variables



Boxplot for continuous variables



Bar Plots for categorical vairables



Correlation Plot

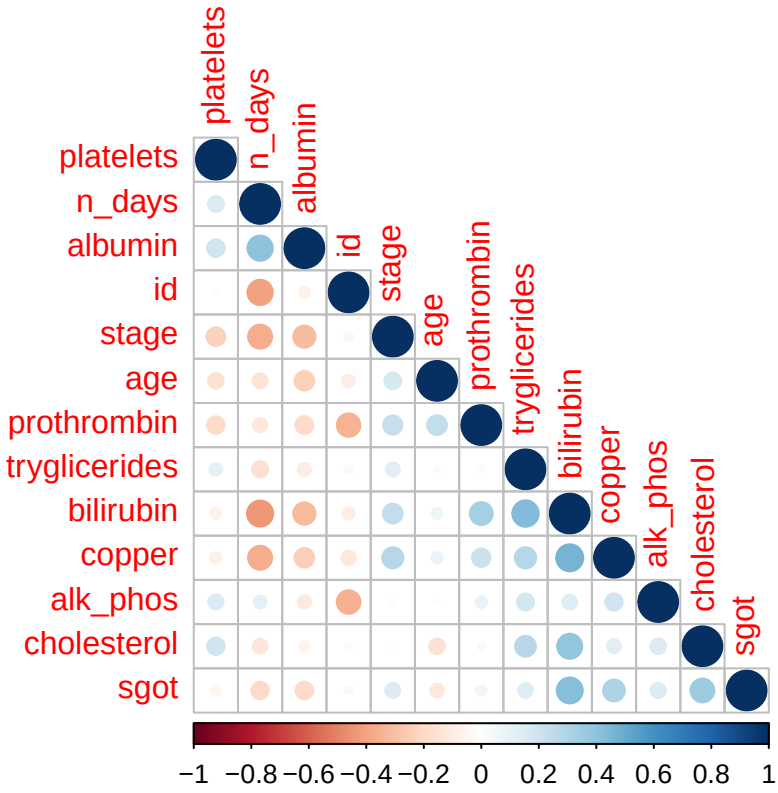


Table 1: Baseline Characteristics

```
## Setting theme "New England Journal of Medicine"
```

```
## Warning: fonts used in `flextable` are ignored because the `pdflatex` engine is
## used and not `xelatex` or `lualatex`. You can avoid this warning by using the
## `set_flextable_defaults(fonts_ignore=TRUE)` command or use a compatible engine
## by defining `latex_engine: xelatex` in the YAML header of the R Markdown
## document.
```

Table 1: Baseline Characteristics

Characteristic	**Censored** N = 232[1]	**Censored due to liver tx** N = 25[1]	**Death** N = 161[1]
N_days	2,333.2 / 2,186.5 (994.7)	1,546.2 / 1,435.0 (753.1)	1,376.9 / 1,083.0 (1,049.2)
Drug			
D-penicillamine	83 (49%)	10 (53%)	65 (52%)
Placebo	85 (51%)	9 (47%)	60 (48%)
Age	49.6 / 50.0 (10.4)	41.6 / 41.0 (6.3)	54.0 / 54.0 (9.8)
Sex			
Female	215 (93%)	22 (88%)	137 (85%)
Male	17 (7.3%)	3 (12%)	24 (15%)
Ascites	1 (0.6%)	0 (0%)	23 (18%)
Hepatomegaly	60 (36%)	12 (63%)	88 (70%)
Spiders	33 (20%)	5 (26%)	52 (42%)
Edema	16 (6.9%)	3 (12%)	45 (28%)
Bilirubin	1.6 / 0.9 (1.9)	3.6 / 3.1 (3.6)	5.5 / 3.2 (5.8)
Cholesterol	326.5 / 292.0 (165.8)	439.5 / 343.5 (335.5)	415.8 / 339.0 (275.0)
Albumin	3.6 / 3.6 (0.4)	3.5 / 3.5 (0.5)	3.4 / 3.4 (0.5)
Copper	66.6 / 52.0 (57.1)	124.0 / 102.0 (100.1)	135.4 / 111.0 (98.5)
Alk_phos	1,578.1 / 1,107.5 (1,633.1)	1,535.2 / 1,345.0 (837.7)	2,594.4 / 1,664.0 (2,677.1)
SGOT	107.3 / 94.6 (52.8)	130.1 / 127.0 (36.9)	141.9 / 134.9 (58.4)
Tryglicerides	111.8 / 104.0 (48.3)	133.9 / 124.0 (70.5)	140.5 / 122.0 (79.3)
Platelets	261.2 / 256.0 (88.6)	309.6 / 304.0 (102.7)	242.5 / 224.0 (107.9)
Prothrombin	10.5 / 10.4 (0.9)	10.4 / 10.3 (0.5)	11.2 / 11.0 (1.0)
Stage			
1	19 (8.3%)	0 (0%)	2 (1.3%)
2	64 (28%)	5 (20%)	23 (15%)
3	97 (42%)	10 (40%)	48 (31%)
4	50 (22%)	10 (40%)	84 (54%)

[1]Mean / Median (SD); n (%)

Table 2: Multivariate Cox Proportional Hazards Analysis

Characteristic	**HR**[1]	**95% CI**[1]	**p-value**
age	1.02	1.00 to 1.04	0.019
sex			
Female	—	—	
Male	1.30	0.75 to 2.26	0.35
bilirubin	1.12	1.08 to 1.16	<0.001
albumin	0.35	0.22 to 0.56	<0.001
copper	1.00	1.00 to 1.01	0.002
prothrombin	1.32	1.12 to 1.57	0.001
stage	1.46	1.13 to 1.88	0.003

[1]HR = Hazard Ratio, CI = Confidence Interval

Multivariate analysis

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `status = case_when(...)`.
## Caused by warning:
## ! NAs introduced by coercion
```