EDA

Chen Liang

2024-11-16

```
cirrhosis <- read_csv("data/cirrhosis.csv")|>
  janitor::clean_names()
```

```
## 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.
```

```
summary(cirrhosis)
```

```
##       id            n_days        status            drug
## Min.   : 1.0   Min.   :  41   Length:418         Length:418
## 1st Qu.:105.2  1st Qu.:1093   Class :character   Class :character
## Median :209.5  Median :1730   Mode  :character   Mode  :character
## Mean   :209.5  Mean   :1918
## 3rd Qu.:313.8  3rd Qu.:2614
## Max.   :418.0  Max.   :4795
##
##      age            sex              ascites          hepatomegaly
## Min.   : 9598   Length:418         Length:418         Length:418
## 1st Qu.:15644   Class :character   Class :character   Class :character
## Median :18628   Mode  :character   Mode  :character   Mode  :character
## Mean   :18533
## 3rd Qu.:21272
## Max.   :28650
##
##     spiders           edema             bilirubin        cholesterol
## Length:418         Length:418         Min.   : 0.300   Min.   : 120.0
## Class :character   Class :character   1st Qu.: 0.800   1st Qu.: 249.5
## Mode  :character   Mode  :character   Median : 1.400   Median : 309.5
##                                       Mean   : 3.221   Mean   : 369.5
##                                       3rd Qu.: 3.400   3rd Qu.: 400.0
##                                       Max.   :28.000   Max.   :1775.0
##                                                        NA's   :134
##     albumin          copper          alk_phos           sgot
## Min.   :1.960   Min.   :  4.00   Min.   : 289.0   Min.   : 26.35
## 1st Qu.:3.243   1st Qu.: 41.25   1st Qu.: 871.5   1st Qu.: 80.60
```

1

```
##   Median :3.530   Median : 73.00   Median : 1259.0   Median :114.70
##   Mean   :3.497   Mean   : 97.65   Mean   : 1982.7   Mean   :122.56
##   3rd Qu.:3.770   3rd Qu.:123.00   3rd Qu.: 1980.0   3rd Qu.:151.90
##   Max.   :4.640   Max.   :588.00   Max.   :13862.4   Max.   :457.25
##                   NA's   :108      NA's   :106       NA's   :106
##   tryglicerides     platelets      prothrombin       stage
##   Min.   : 33.00   Min.   : 62.0   Min.   : 9.00   Min.   :1.000
##   1st Qu.: 84.25   1st Qu.:188.5   1st Qu.:10.00   1st Qu.:2.000
##   Median :108.00   Median :251.0   Median :10.60   Median :3.000
##   Mean   :124.70   Mean   :257.0   Mean   :10.73   Mean   :3.024
##   3rd Qu.:151.00   3rd Qu.:318.0   3rd Qu.:11.10   3rd Qu.:4.000
##   Max.   :598.00   Max.   :721.0   Max.   :18.00   Max.   :4.000
##   NA's   :136      NA's   :11      NA's   :2       NA's   :6
```