

Project 1 Redwood Data Report

Yicheng Shen (Student ID: 2806571) & Steve Bao (Student ID: XXXXXX)

October 13, 2022

Data Cleaning

```
sonoma_data_all <- read.csv("data/sonoma-data-all.csv")
sonoma_data_log <- read.csv("data/sonoma-data-log.csv")
sonoma_data_net <- read.csv("data/sonoma-data-all.csv")
```

```
summary(sonoma_data_all)
```

```
## result_time      epoch      nodeid      parent
## Length:416036    Min.   :    2    Min.   :   2.00    Min.   :    0
## Class :character  1st Qu.: 1625    1st Qu.:   55.00    1st Qu.:   42
## Mode  :character  Median : 3817    Median :  110.00    Median :  118
##              Mean  : 4220    Mean  :   93.61    Mean  : 1727
##              3rd Qu.: 6589    3rd Qu.:  127.00    3rd Qu.:  140
##              Max.   :12635    Max.   :65535.00    Max.   :65535
##
## voltage          depth      humidity      humid_temp
## Min.   :   0.0091    Min.   :   0.0    Min.   : -9375.37    Min.   : -38.40
## 1st Qu.:   2.6514    1st Qu.:   2.0    1st Qu.:   44.48    1st Qu.:  10.56
## Median :   2.7496    Median :   3.0    Median :   64.80    Median :  14.10
## Mean   :   82.7331    Mean   :  29.5    Mean   :   64.35    Mean   :  14.82
## 3rd Qu.:  214.0000    3rd Qu.:   7.0    3rd Qu.:   83.64    3rd Qu.:  17.94
## Max.   :1023.0000    Max.   :255.0    Max.   :  114.89    Max.   :603.84
##              NA's   :12532    NA's   :12532
## humid_adj        hamatop      hamabot
## Min.   : -6334.83    Min.   :    0    Min.   :    0.0
## 1st Qu.:   43.88    1st Qu.:    0    1st Qu.:    0.0
## Median :   63.02    Median :    0    Median :    0.0
## Mean   :   62.36    Mean   : 11049    Mean   :   252.8
## 3rd Qu.:   80.43    3rd Qu.:  7250    3rd Qu.:    0.0
## Max.   :  147.69    Max.   :22592200    Max.   :465820.0
## NA's   :12532      NA's   :12532      NA's   :12532
```

```
summary(sonoma_data_log)
```

```
## result_time      epoch      nodeid      parent
## Length:301056    Min.   :    2    Min.   :   2.00    Min.   :    0
## Class :character  1st Qu.: 1166    1st Qu.:   49.00    1st Qu.:   46
```

```
## Mode :character      Median : 2420      Median : 105.00      Median : 118
##                      Mean   : 3323      Mean   : 93.08      Mean   : 2338
##                      3rd Qu.: 5249      3rd Qu.: 127.00     3rd Qu.: 141
##                      Max.   :12635      Max.   :65535.00    Max.   :65535
##
##      voltage          depth          humidity          humid_temp
## Min.   :0.009063      Min.   : 0.00      Min.   : -9375.37    Min.   : -38.40
## 1st Qu.:2.627960      1st Qu.: 2.00      1st Qu.: 40.03      1st Qu.: 10.87
## Median :2.699640      Median : 3.00      Median : 61.58      Median : 14.72
## Mean   :2.506449      Mean   : 39.83      Mean   : 61.41      Mean   : 15.02
## 3rd Qu.:2.775330      3rd Qu.: 19.00     3rd Qu.: 80.20      3rd Qu.: 18.81
## Max.   :3.030200      Max.   :255.00      Max.   : 104.41      Max.   :603.84
##                      NA's   :8270      NA's   :8270
##      humid_adj          hamatop          hamabot
## Min.   : -6334.83      Min.   : 0      Min.   : 0.0
## 1st Qu.: 39.53      1st Qu.: 0      1st Qu.: 0.0
## Median : 60.00      Median : 0      Median : 0.0
## Mean   : 59.55      Mean   : 10870      Mean   : 245.5
## 3rd Qu.: 77.22      3rd Qu.: 6762      3rd Qu.: 0.0
## Max.   : 100.22      Max.   :22592200      Max.   :465820.0
## NA's   :8270      NA's   :8270      NA's   :8270
```

```
summary(sonoma_data_net)
```

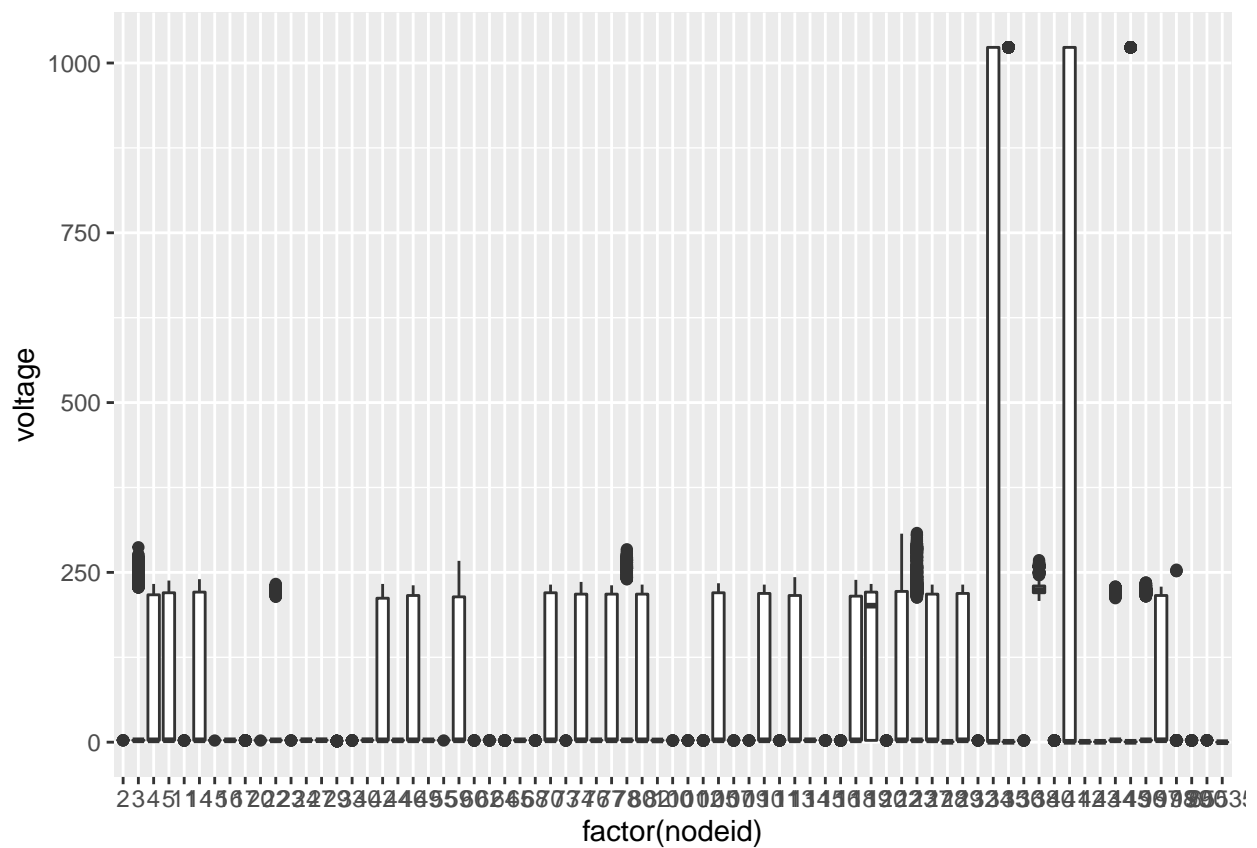
```
## result_time          epoch          nodeid          parent
## Length:416036      Min.   : 2      Min.   : 2.00      Min.   : 0
## Class :character    1st Qu.: 1625      1st Qu.: 55.00      1st Qu.: 42
## Mode :character     Median : 3817      Median : 110.00     Median : 118
##                      Mean   : 4220      Mean   : 93.61      Mean   : 1727
##                      3rd Qu.: 6589      3rd Qu.: 127.00     3rd Qu.: 140
##                      Max.   :12635      Max.   :65535.00    Max.   :65535
##
##      voltage          depth          humidity          humid_temp
## Min.   : 0.0091      Min.   : 0.0      Min.   : -9375.37    Min.   : -38.40
## 1st Qu.: 2.6514      1st Qu.: 2.0      1st Qu.: 44.48      1st Qu.: 10.56
## Median : 2.7496      Median : 3.0      Median : 64.80      Median : 14.10
## Mean   : 82.7331      Mean   : 29.5      Mean   : 64.35      Mean   : 14.82
## 3rd Qu.: 214.0000      3rd Qu.: 7.0      3rd Qu.: 83.64      3rd Qu.: 17.94
## Max.   :1023.0000      Max.   :255.0      Max.   : 114.89      Max.   :603.84
##                      NA's   :12532      NA's   :12532
##      humid_adj          hamatop          hamabot
## Min.   : -6334.83      Min.   : 0      Min.   : 0.0
## 1st Qu.: 43.88      1st Qu.: 0      1st Qu.: 0.0
## Median : 63.02      Median : 0      Median : 0.0
## Mean   : 62.36      Mean   : 11049      Mean   : 252.8
## 3rd Qu.: 80.43      3rd Qu.: 7250      3rd Qu.: 0.0
## Max.   : 147.69      Max.   :22592200      Max.   :465820.0
## NA's   :12532      NA's   :12532      NA's   :12532
```

```
table(sonoma_data_net$nodeid)
```

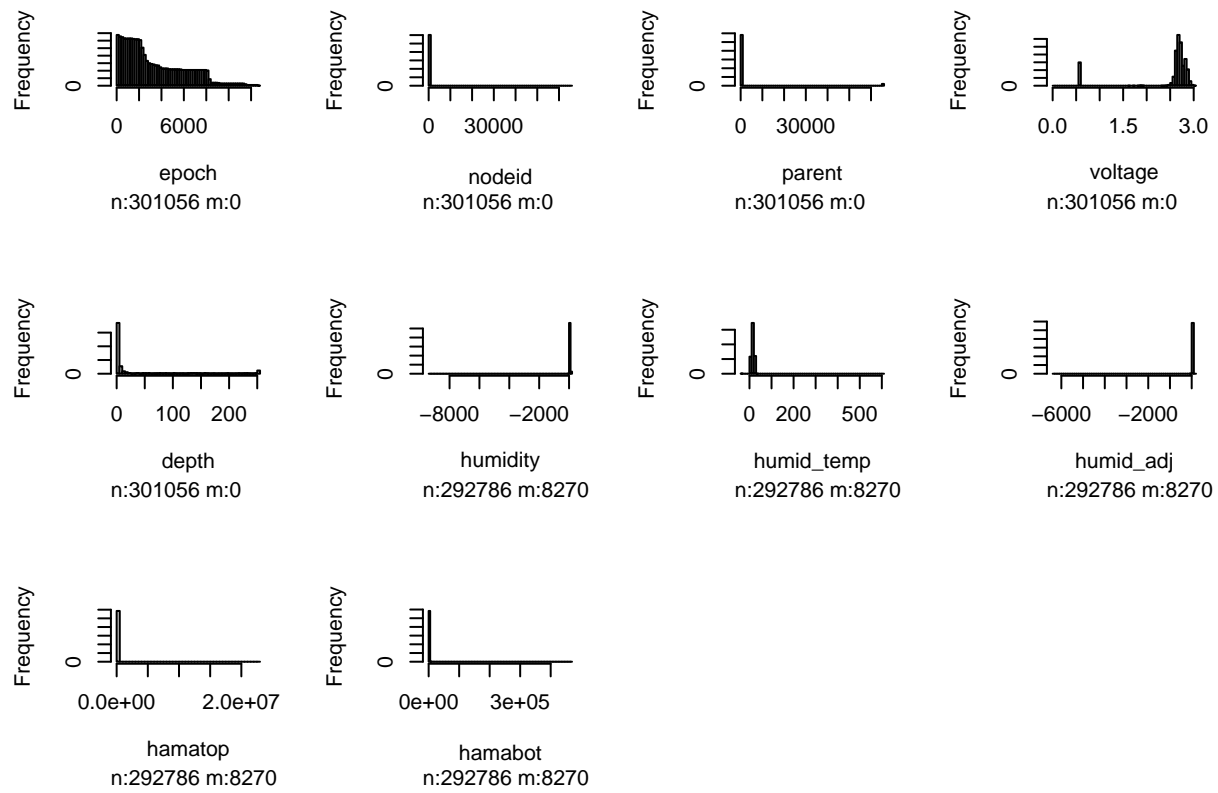
```
##
##      2      3      4      5      11      14      15      16      17      20      22      23      24
```

```
##      65  3929 12233 16261  2187 12982  2344  2171  2724  2517  4227  2171   57
##      27    29   33    40   42   44    46   49   55   59   60   62   64
##      81   660  2584   98 17202  2732 13534  2064  2667  7852  2237   73 2509
##      65    68   70   73   74   76   77   78   80   82   100  101  102
## 2427 2435 13470 2625 13053 2427 13047 3139 11834 1468 2204 2221 2212
##   105   107   109   110   111   113   114   115   116   118   119   120   122
## 13663 2328   345 13640 2427 12103   536 2311  2559 23453 16609 2220 12371
##   123   127   128   129   132   134   135   136   138   140   141   142   143
## 5881 16834 2204 12233 2319 13639 4166 2469 4104 2963 12417 2410 2261
##   144   145   196   197   198   199   200 65535
## 4171 2963 4356 19158 3468 2666    65    1
```

```
ggplot(sonoma_data_net) + geom_boxplot(aes(x = factor(nodeid), y = voltage))
```



```
hist.data.frame(sonoma_data_log%>%dplyr::select(-result_time))
```



```
summary(sonoma_data_log$humidity)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.    NA's
## -9375.37   40.03     61.58    61.41    80.20   104.41   8270
```