Project 1

Sanskriti Purohit & Caleb Woo

2022-10-13

```r
knitr::opts_chunk$set(message = F, warning = F)
```

```r
library(tidyverse)
data_net <- read_csv("data/sonoma-data-net.csv")
data_log <- read_csv("data/sonoma-data-log.csv")
```

```r
dates <- read.table("data/sonoma-dates", sep = " ")
dates <- data.frame(t(dates)[-2,], row.names = NULL)
names(dates) <- dates[1,]
dates <- dates[-1,]
dates <- dates[-13001,]
dates[1,] <- c("1", "Tue Apr 27 17:10:00 2004", "12536.0069444444")
row.names(dates) <- NULL

dates <- dates %>%
  mutate(epochNums = as.integer(epochNums),
         epochDates = as.POSIXct(epochDates, format = "%a %b %d %H:%M:%S %Y"),
         epochDays = as.numeric(epochDays))
names(dates)[1] <- "epoch"
```
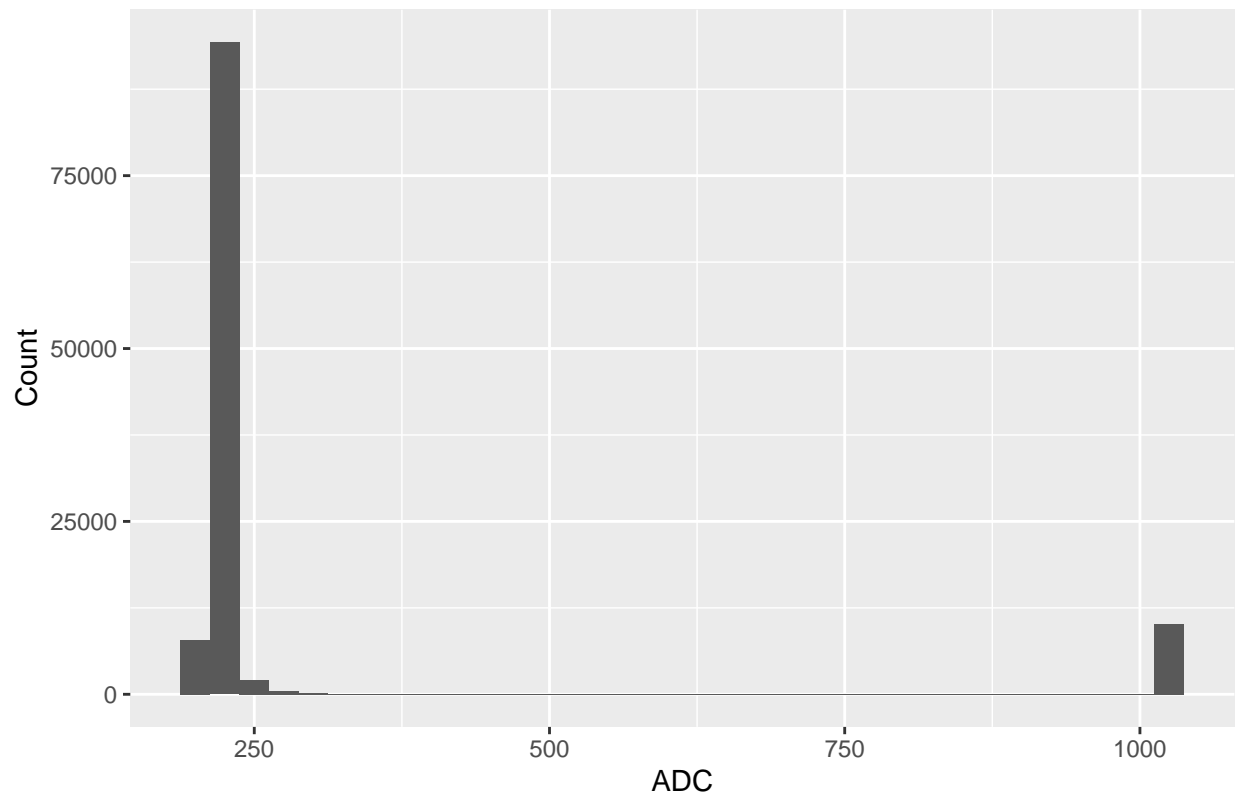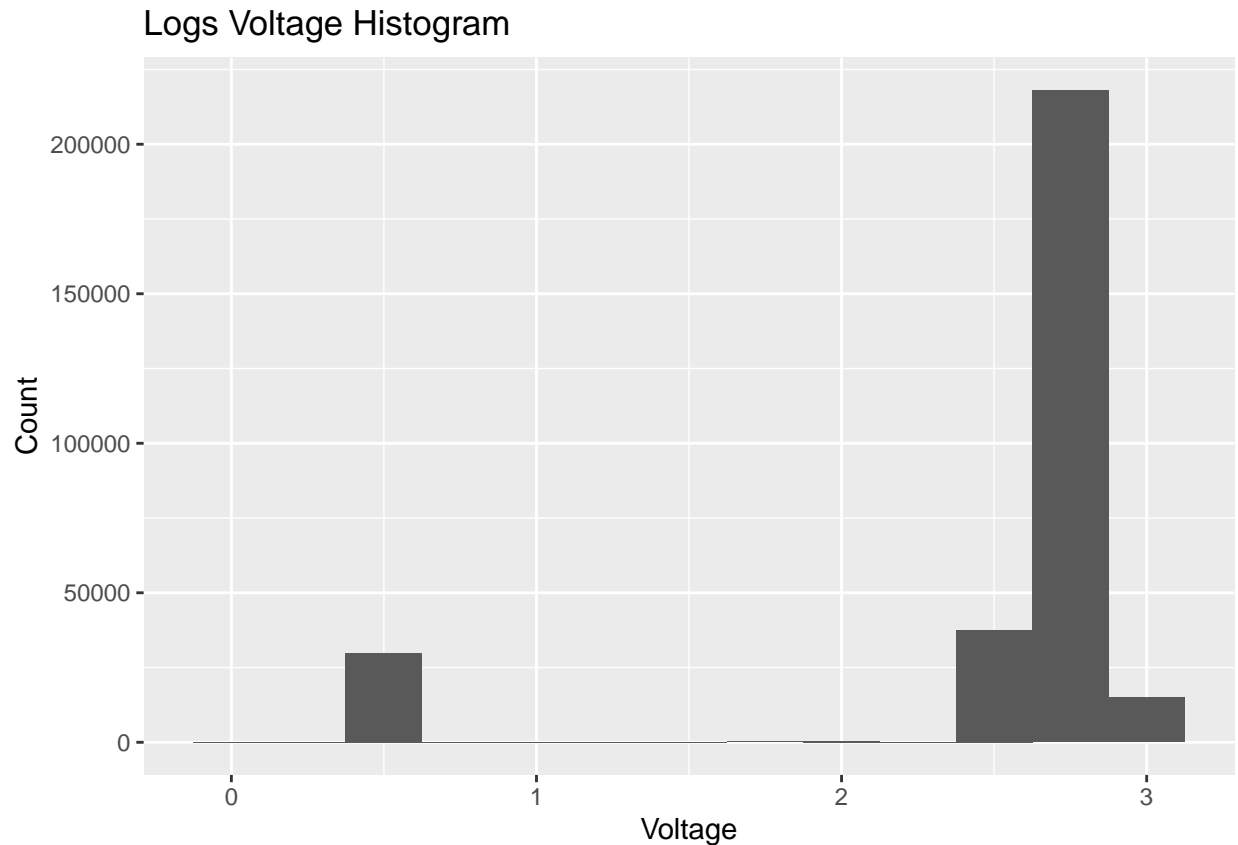
2.

a)

Voltage

```r
data_net %>%
  ggplot(data=., aes(x=voltage)) +
  geom_histogram(binwidth=25) +
  labs(title="Network ADC Histogram", x="ADC", y="Count")
```
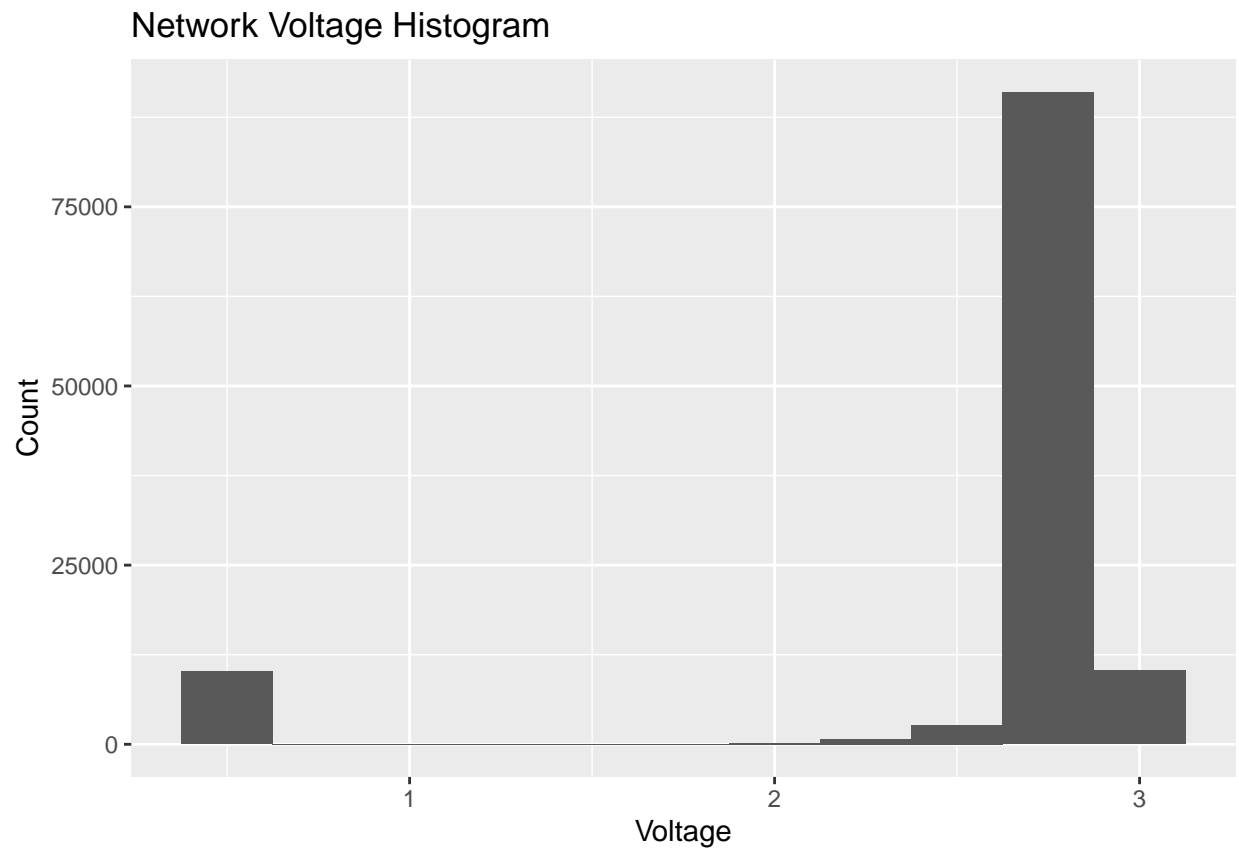
Network ADC Histogram


```
data_log %>%
  ggplot(data=., aes(x=voltage)) +
  geom_histogram(binwidth=0.25) +
  labs(title="Logs Voltage Histogram", x="Voltage", y="Count")
```
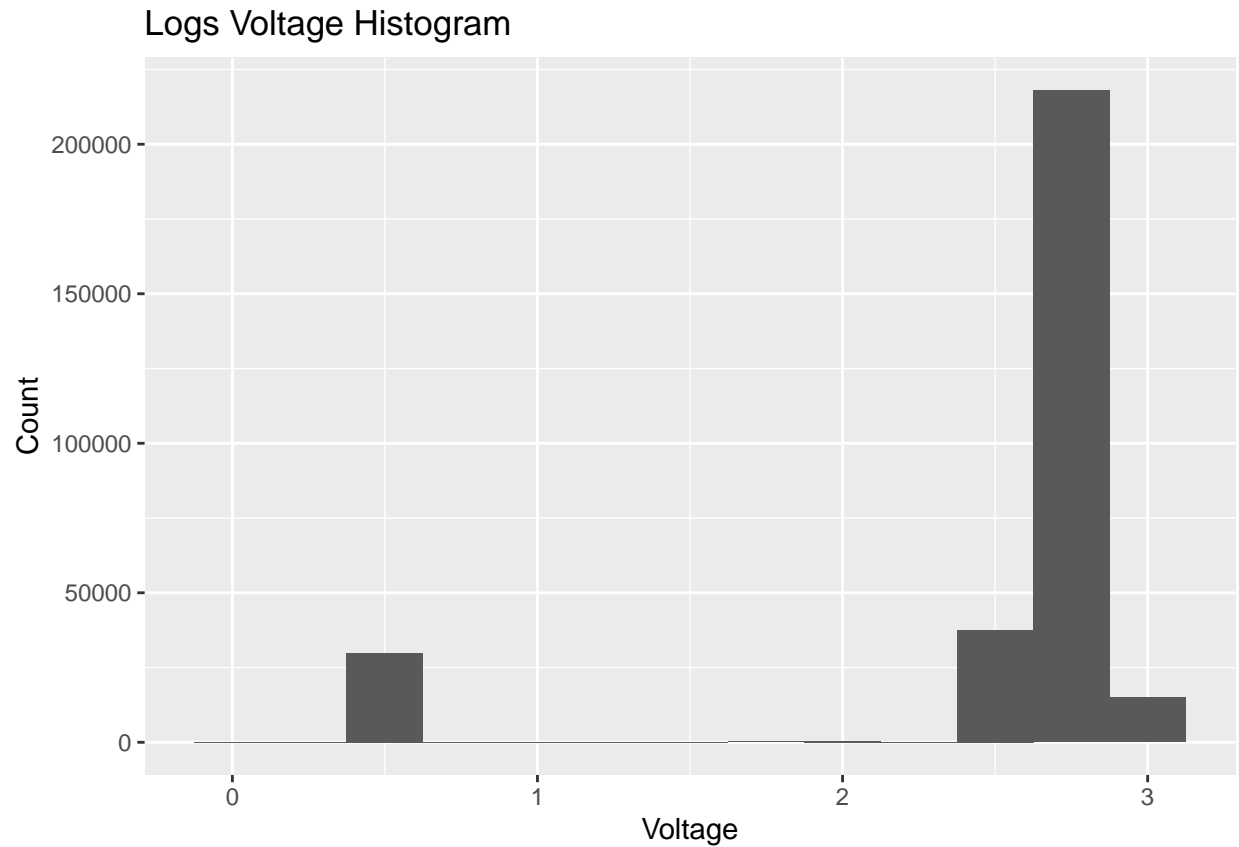
Logs Voltage Histogram

Voltage is inconsistent, link below for ADC to voltage conversion

- http://www-db.ics.uci.edu/pages/research/quasar/MPR-MIB%20Series%20User%20Manual%207430-0021-06__A.pdf

- page 23 (25 of pdf) has MICA2DOT conversion

```
data_net <- data_net %>%
  mutate(voltage = 0.6*(1024/voltage))

data_net %>%
  ggplot(data=., aes(x=voltage)) +
  geom_histogram(binwidth=0.25) +
  labs(title="Network Voltage Histogram", x="Voltage", y="Count")
```

Network Voltage Histogram


```
data_log %>%
  ggplot(data=., aes(x=voltage)) +
  geom_histogram(binwidth=0.25) +
  labs(title="Logs Voltage Histogram", x="Voltage", y="Count")
```

Logs Voltage Histogram

Count vs Voltage histogram showing most observations between 2.4 and 3 volts, with a smaller group near 0.5 volts.

Humidity

```
data_net %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4) %>%
  select(humidity) %>%
  pull() %>%
  min()
```

```
## [1] 19.5147
```

```
data_log %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4) %>%
  select(humidity) %>%
  pull() %>%
  min()
```

```
## [1] 16.2653
```
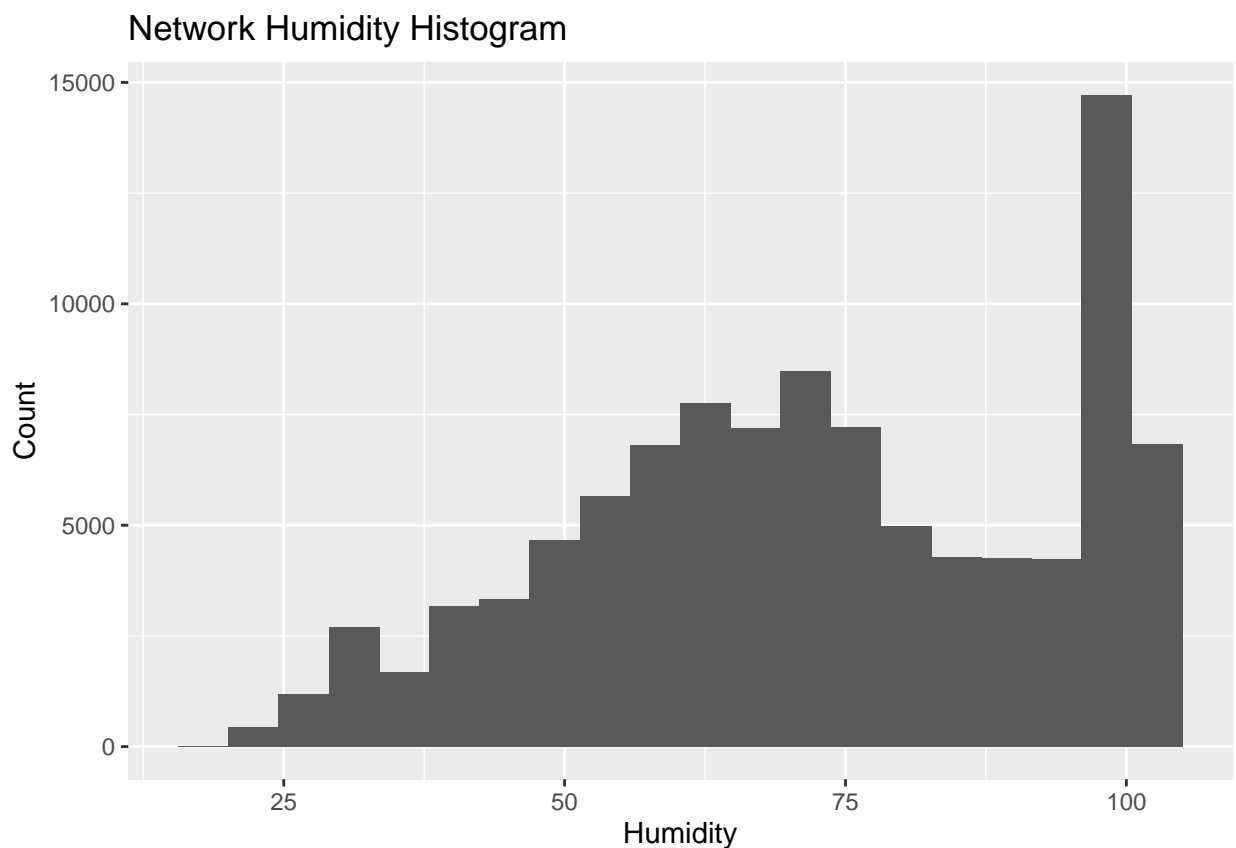
```
data_net %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4 & humidity < 105) %>%
  select(humidity) %>%
  pull() %>%
  max()
```

```
## [1] 104.385
```

```
data_log %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4) %>%
  select(humidity) %>%
  pull() %>%
  max()
```
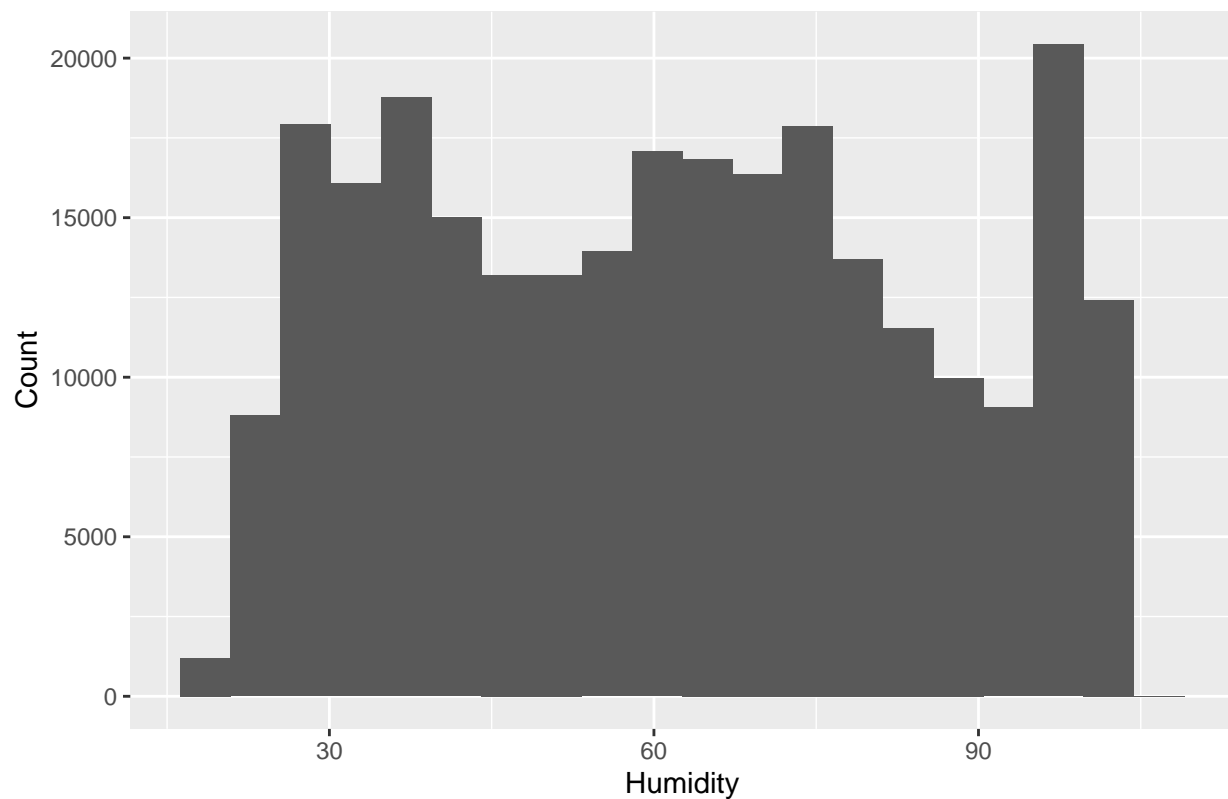
```
## [1] 104.405
```

```
data_net %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4 & humidity < 105) %>%
  ggplot(data=., aes(x=humidity)) +
  geom_histogram(bins=20) +
  labs(title="Network Humidity Histogram", x="Humidity", y="Count")
```


Network Humidity Histogram

```
data_log %>%
  filter(!is.na(humidity) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=humidity)) +
  geom_histogram(bins=20) +
  labs(title="Logs Humidity Histogram", x="Humidity", y="Count")
```

Logs Humidity Histogram

Adjusted Humidity

```r
data_net %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_adj) %>%
  pull() %>%
  min()
```

```
## [1] 19.3107
```

```r
data_log %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_adj) %>%
  pull() %>%
  min()
```

```
## [1] 16.2282
```
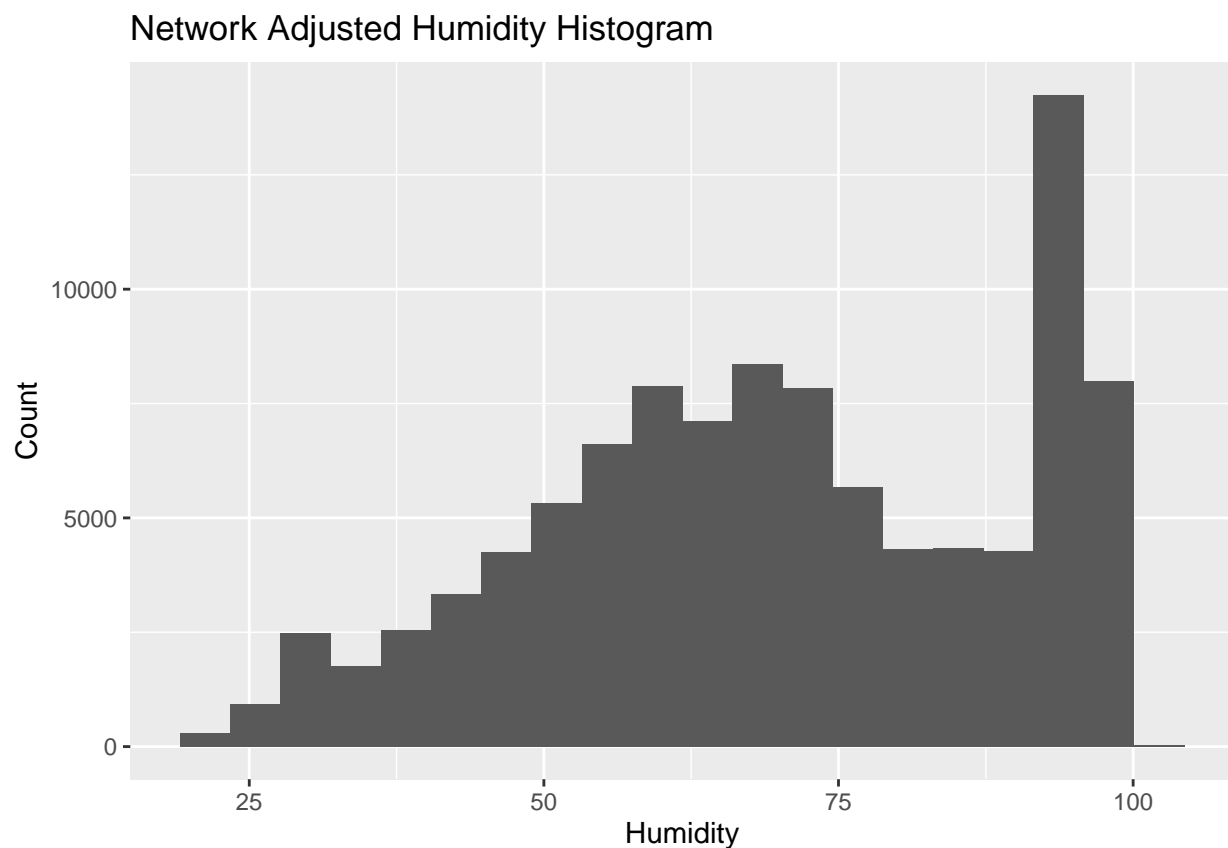
```r
data_net %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4 & humid_adj < 105) %>%
  select(humid_adj) %>%
  pull() %>%
  max()
```

```
## [1] 100.223
```
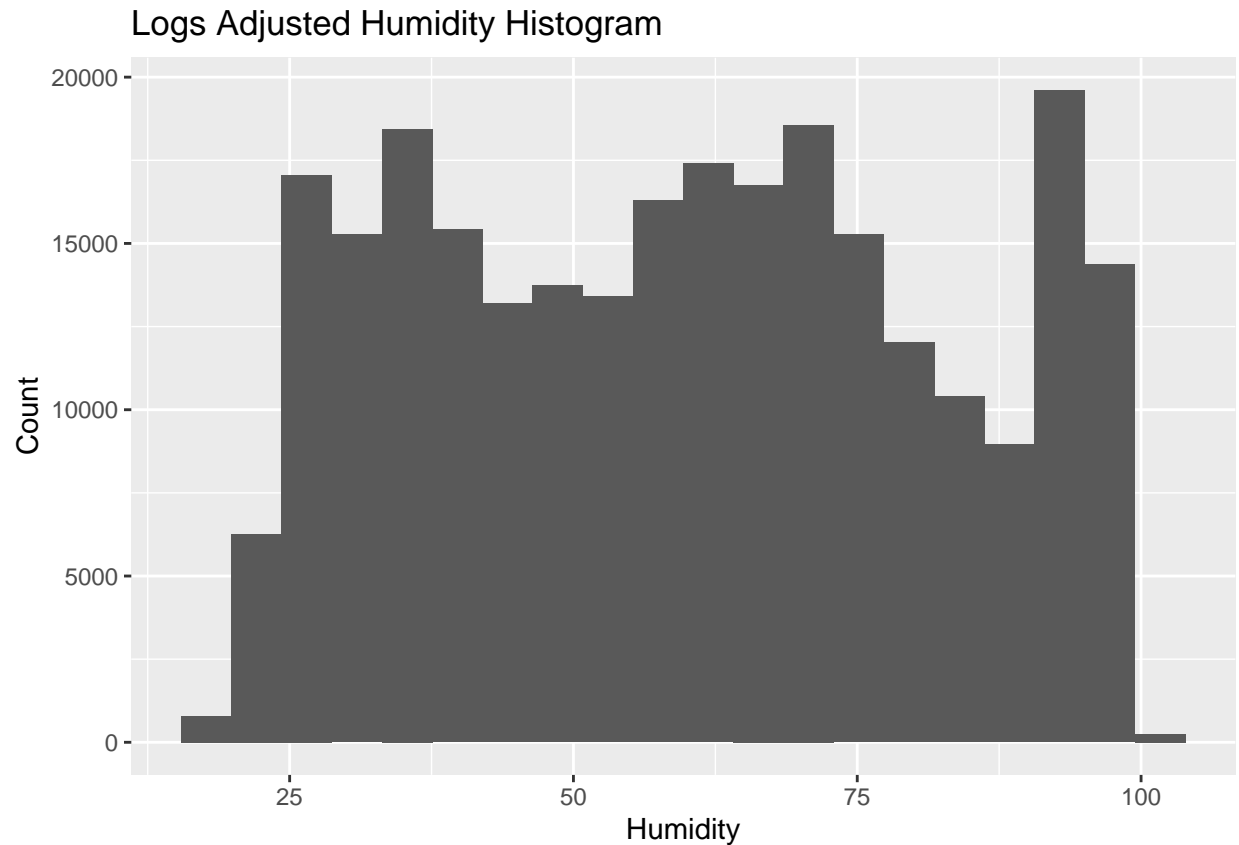
```
data_log %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_adj) %>%
  pull() %>%
  max()
```

```
## [1] 100.223
```

```
data_net %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4 & humid_adj < 105) %>%
  ggplot(data=., aes(x=humid_adj)) +
  geom_histogram(bins=20) +
  labs(title="Network Adjusted Humidity Histogram", x="Humidity", y="Count")
```

Network Adjusted Humidity Histogram


```
data_log %>%
  filter(!is.na(humid_adj) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=humid_adj)) +
  geom_histogram(bins=20) +
  labs(title="Logs Adjusted Humidity Histogram", x="Humidity", y="Count")
```

Logs Adjusted Humidity Histogram

Temperature

```
data_net %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_temp) %>%
  pull() %>%
  min()
```

```
## [1] 6.582
```

```
data_log %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_temp) %>%
  pull() %>%
  min()
```

```
## [1] 6.582
```
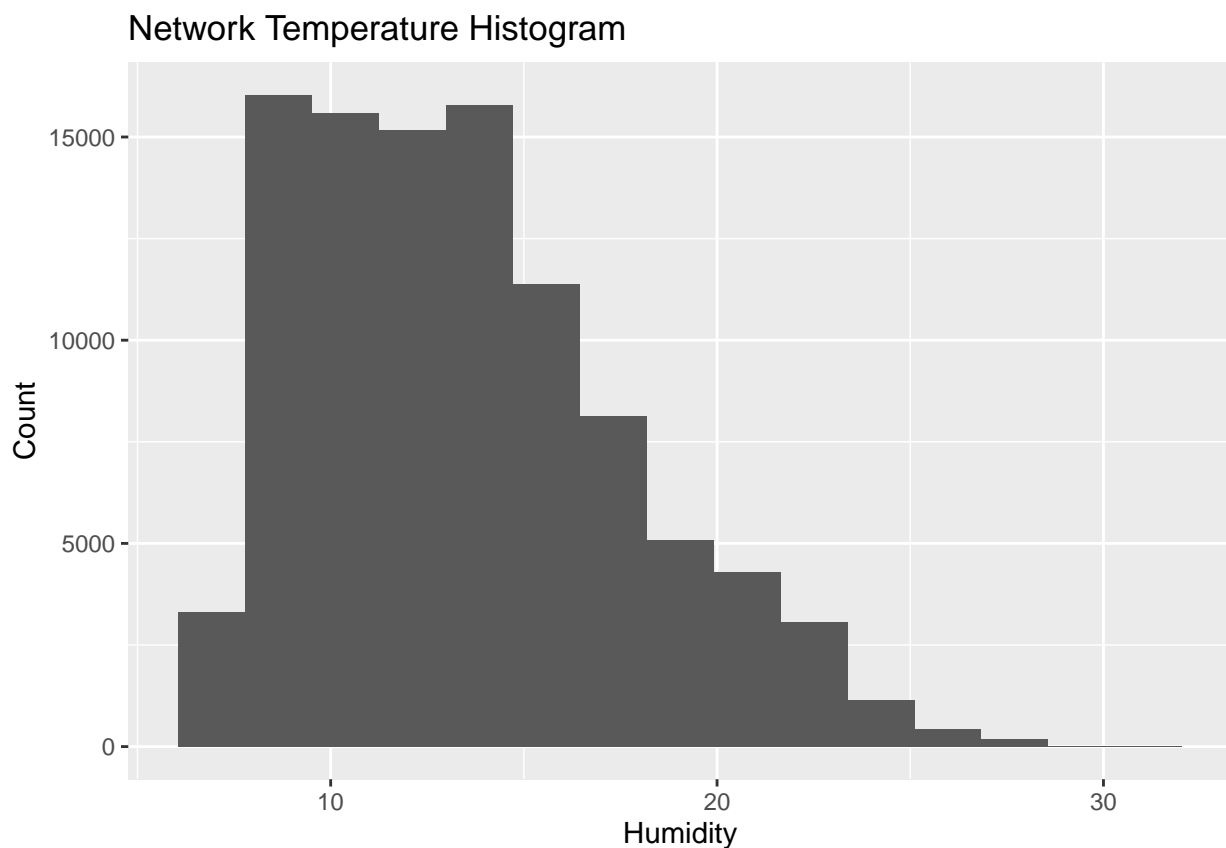
```
data_net %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4 & humid_temp < 35) %>%
  select(humid_temp) %>%
  pull() %>%
  max()
```

```
## [1] 30.8272
```

```
data_log %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4) %>%
  select(humid_temp) %>%
  pull() %>%
  max()
```

```
## [1] 32.5814
```

```
data_net %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4 & humid_temp < 35) %>%
  ggplot(data=., aes(x=humid_temp)) +
  geom_histogram(bins=15) +
  labs(title="Network Temperature Histogram", x="Humidity", y="Count")
```



```
data_log %>%
  filter(!is.na(humid_temp) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=humid_temp)) +
  geom_histogram(bins=15) +
  labs(title="Logs Temperature Histogram", x="Humidity", y="Count")
```

Logs Temperature Histogram
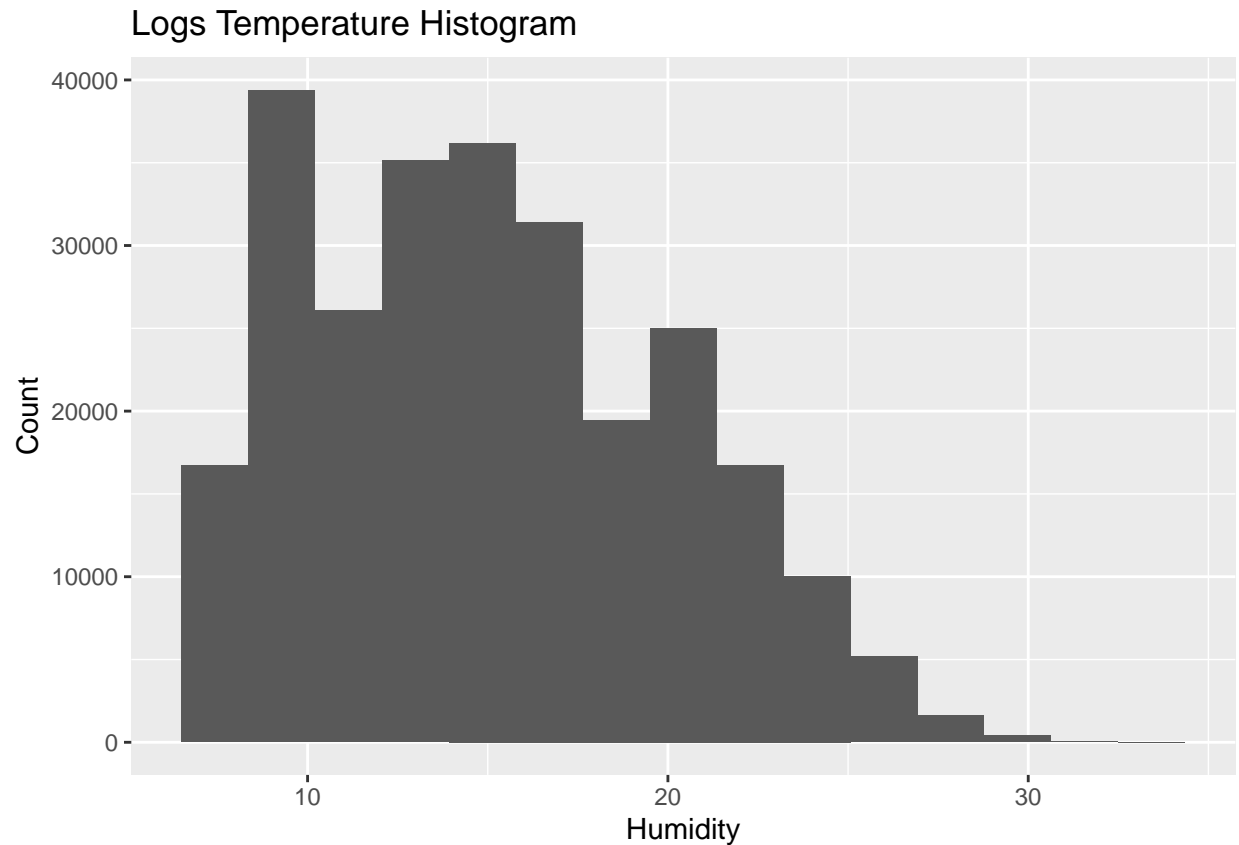
Incident PAR

```
data_net %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamatop) %>%
  pull() %>%
  min()
```

```
## [1] 0
```

```
data_log %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamatop) %>%
  pull() %>%
  min()
```

```
## [1] 0
```

```
data_net %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamatop) %>%
  pull() %>%
  max()
```

```
## [1] 113376
```

```r
data_log %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamatop) %>%
  pull() %>%
  max()
```

```
## [1] 180255
```

Hamatop is inconsistent, link below for Lux to PPFD conversion

- https://www.apogeeinstruments.com/conversion-ppfd-to-lux/

```r
data_net <- data_net %>%
  mutate(hamatop=0.0185*hamatop)

data_log <- data_log %>%
  mutate(hamatop=0.0185*hamatop)

data_net %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamatop) %>%
  pull() %>%
  max()
```
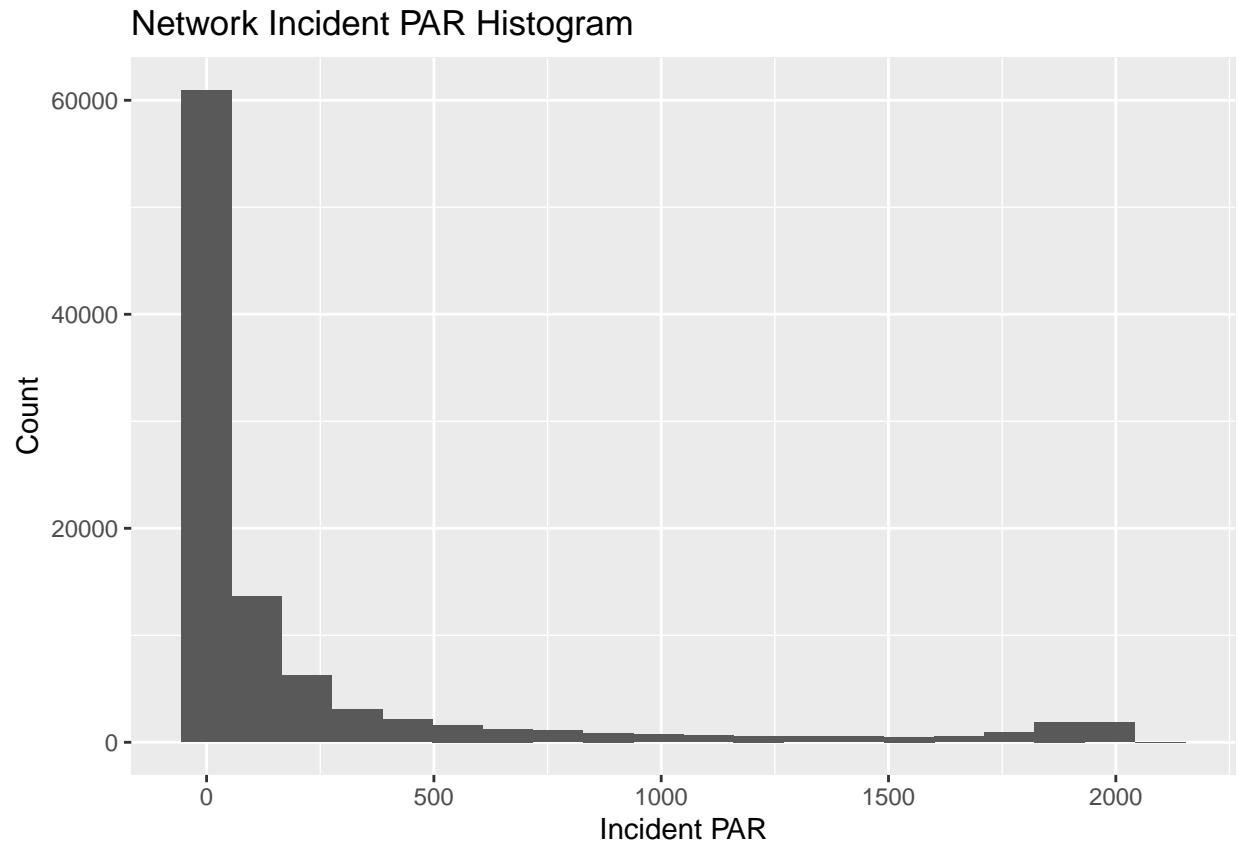
```
## [1] 2097.456
```

```r
data_log %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4 & hamatop < 2200) %>%
  select(hamatop) %>%
  pull() %>%
  max()
```

```
## [1] 2146
```

```r
data_net %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=hamatop)) +
  geom_histogram(bins=20) +
  labs(title="Network Incident PAR Histogram", x="Incident PAR", y="Count")
```

Network Incident PAR Histogram


```
data_log %>%
  filter(!is.na(hamatop) & voltage <= 3 & voltage >= 2.4  & hamatop < 2200) %>%
  ggplot(data=., aes(x=hamatop)) +
  geom_histogram(bins=20) +
  labs(title="Logs Incident PAR Histogram", x="Incident PAR", y="Count")
```

Logs Incident PAR Histogram
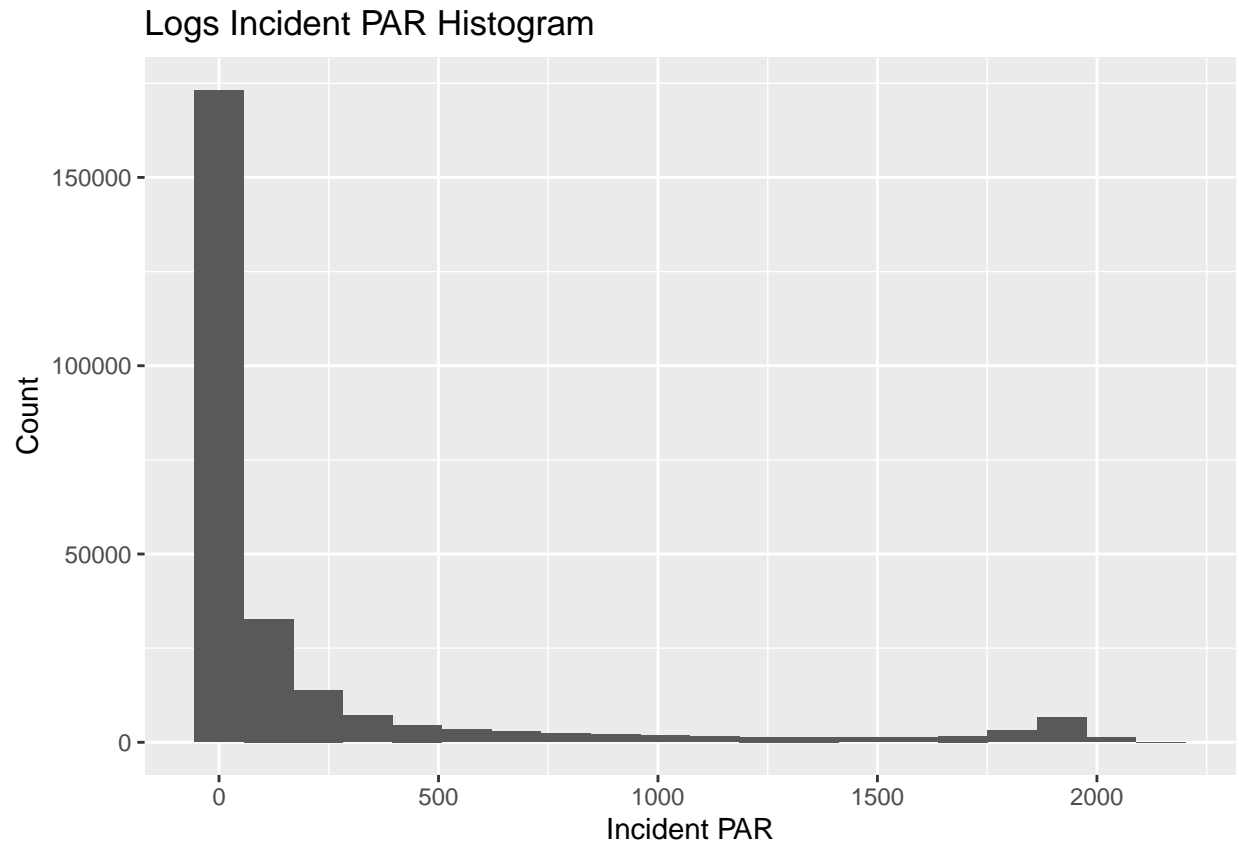
Reflected PAR

```
data_net %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  min()
```

```
## [1] 0
```

```
data_log %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  min()
```

```
## [1] 0
```

```
data_net %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  max()
```

```
## [1] 9480.77
```

```
data_log %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  max()
```

```
## [1] 9142.86
```

Fix hamabot like hamatop

```
data_net <- data_net %>%
  mutate(hamabot=0.0185*hamabot)

data_log <- data_log %>%
  mutate(hamabot=0.0185*hamabot)

data_net %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  max()
```
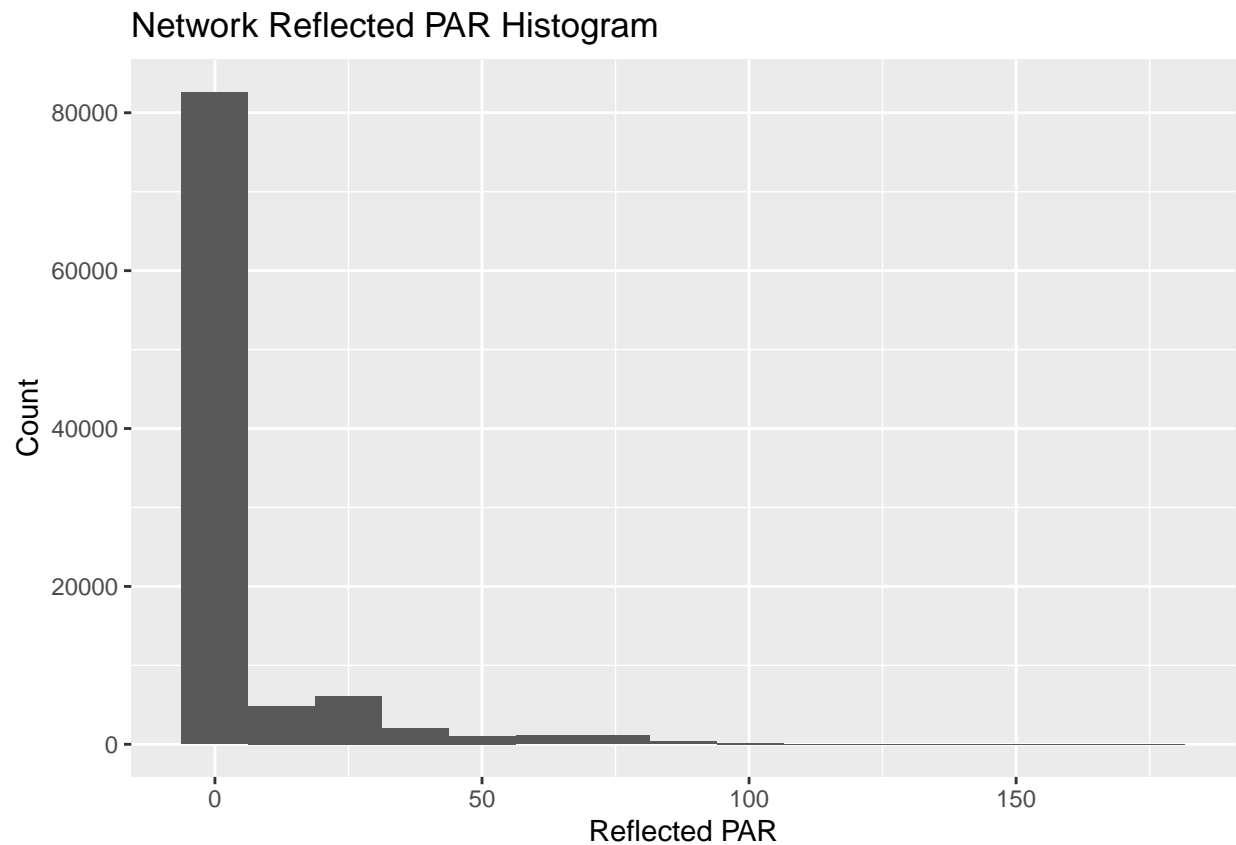
```
## [1] 175.3942
```

```
data_log %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  select(hamabot) %>%
  pull() %>%
  max()
```

```
## [1] 169.1429
```

```
data_net %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=hamabot)) +
  geom_histogram(bins=15) +
  labs(title="Network Reflected PAR Histogram", x="Reflected PAR", y="Count")
```

Network Reflected PAR Histogram


```
data_log %>%
  filter(!is.na(hamabot) & voltage <= 3 & voltage >= 2.4) %>%
  ggplot(data=., aes(x=hamabot)) +
  geom_histogram(bins=15) +
  labs(title="Logs Reflected PAR Histogram", x="Reflected PAR", y="Count")
```

Logs Reflected PAR Histogram
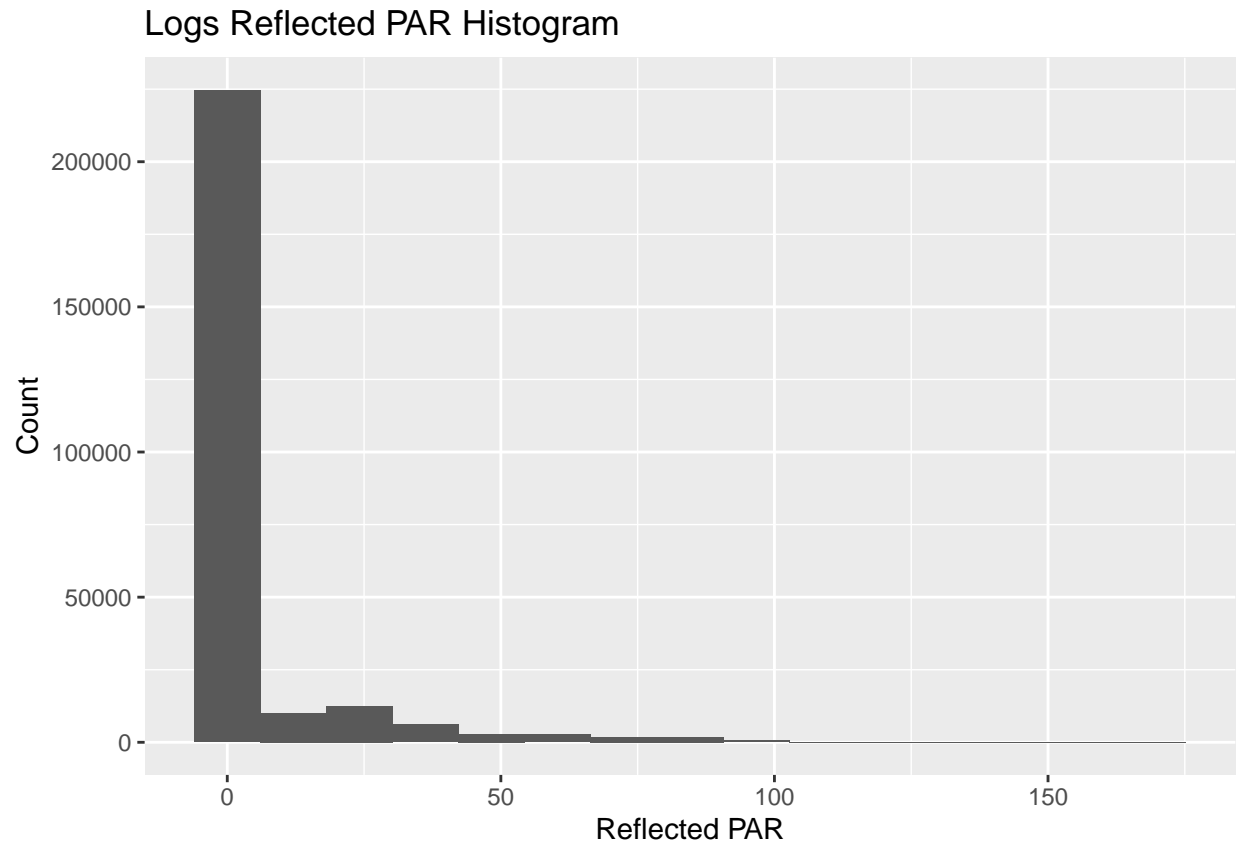
b)

```
data_both <- merge(data_net, data_log, by = c("epoch", "nodeid"), all.y = T, suffixes = c(".net", ".log
data_all <- merge(data_both, dates, by = "epoch", all.x = T)
```

```
missing_data <-  data_all %>%
  select(epoch, nodeid, ends_with(".log"), epochDates, epochDays) %>%
  filter_all(any_vars(is.na(.)))
```

```
paste0("# of missing measurements: ", nrow(missing_data))
```

```
## [1] "# of missing measurements: 8760"
```

```
missing_data %>%
  select(epochDates) %>%
  summarise(min = min(epochDates),
            max = max(epochDates))
```

```
##                   min                 max
## 1 2004-04-30 08:05:00 2004-05-25 21:15:00
```

```
data_all <- data_all[!row.names(data_all) %in% row.names(missing_data),]
row.names(data_all) <- NULL
```