

Data Visualization

April Sang

15/05/2019

```
sensors <- read.csv('sensor.csv', header = T, sep = ",")
#head(sensors, 5)

# Separate day and time and add as columns in data frame
days <- as.numeric(substr(sensors$timestamp,8,9))
times <- substr(sensors$timestamp,45,52)
sensors$day <- days
sensors$time <- times
#tail(sensors, 10)

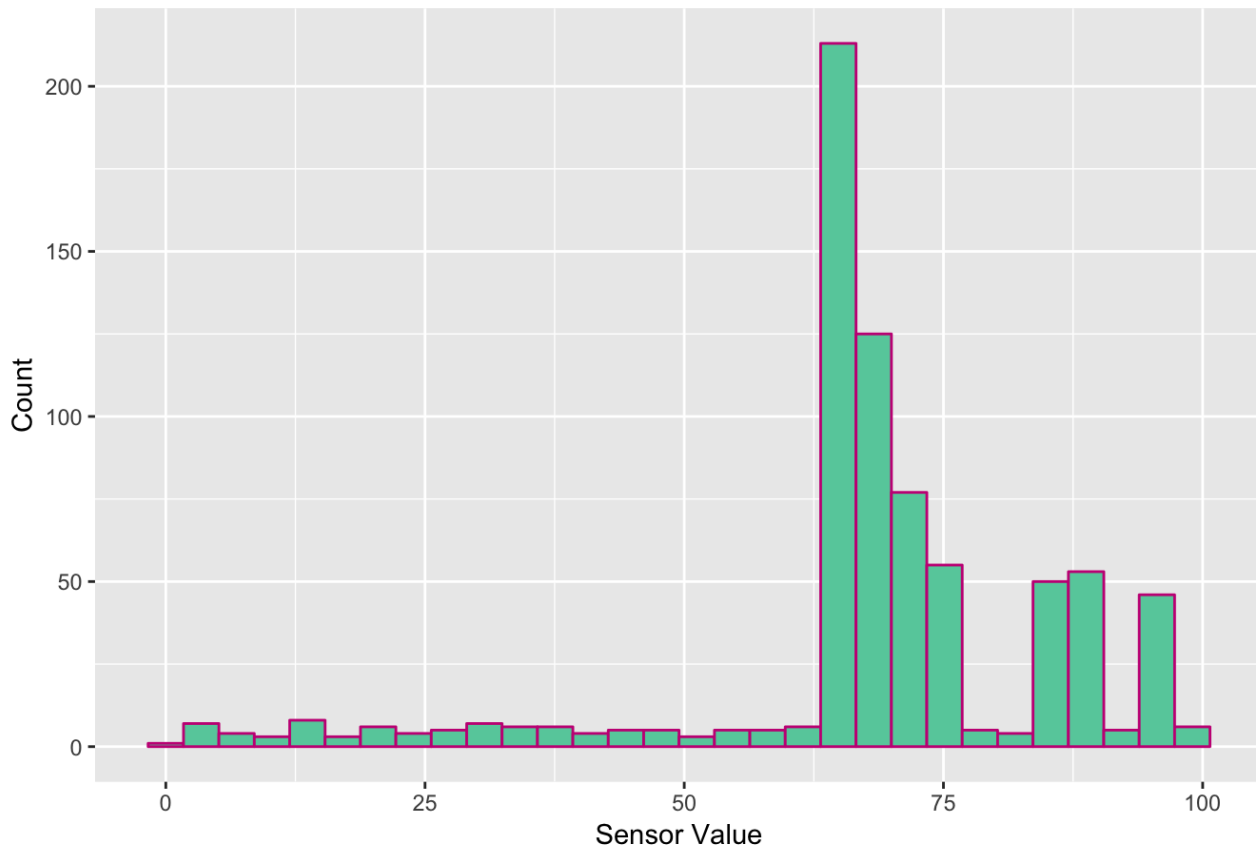
# Create a dataset filters only those observations where value is between 0 and 100 inclusive
sensors_clean <- subset(sensors, as.numeric(value) >=0 & as.numeric(value) <=100)

# Convert column to numeric
sensors_clean$value <- as.numeric(sensors_clean$value)
```

```
# ggplot example
# ggplot histogram
library(ggplot2)
ggplot(sensors_clean, aes(x = value)) + geom_histogram(color = 'mediumvioletred', fill= 'mediumaquamarine') +
  xlab("Sensor Value") + ylab("Count") + ggtitle("Histogram of Sensor Values")
```

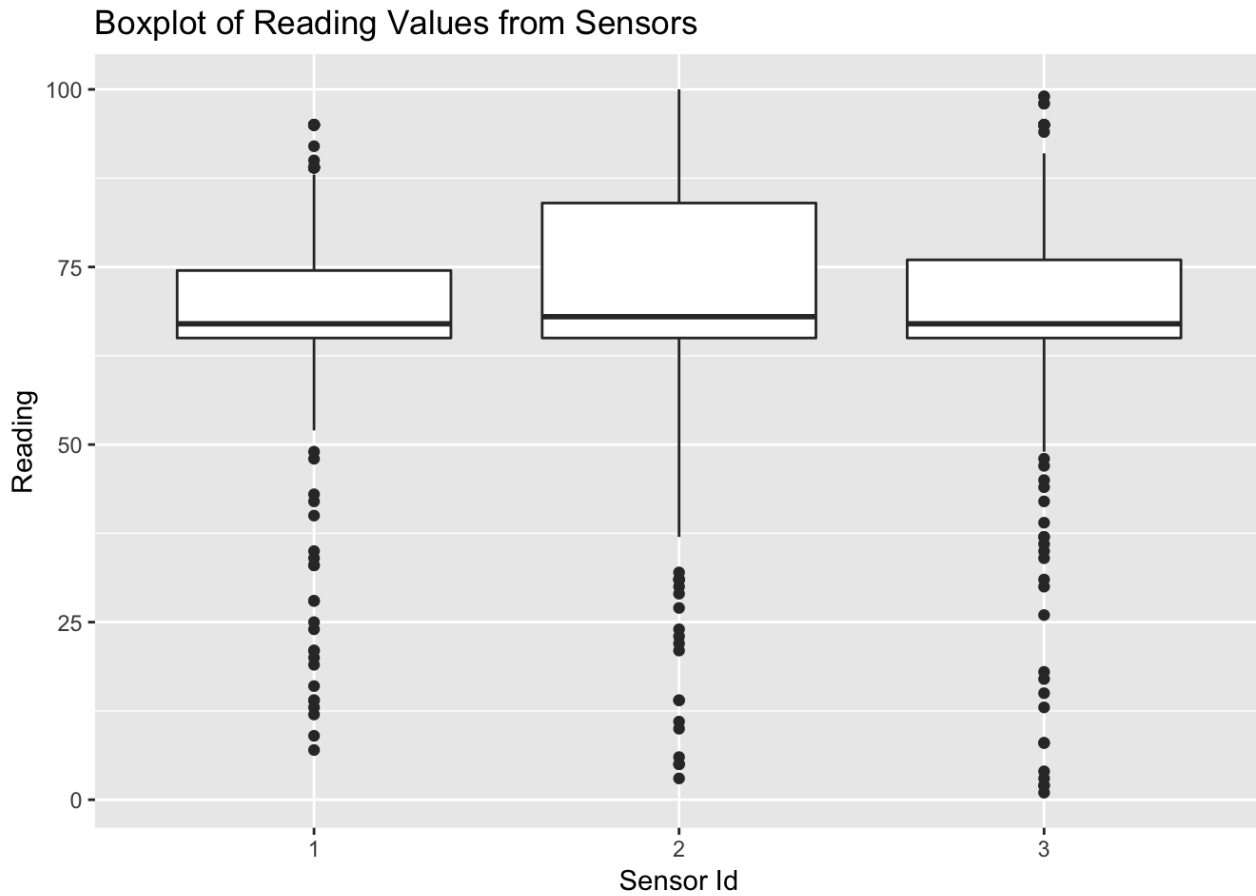
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Histogram of Sensor Values



```
# ggplot boxplot
```

```
ggplot(sensors_clean, aes(x = as.factor(sensorid), y = value)) + geom_boxplot() +  
  xlab("Sensor Id") + ylab("Reading")+ ggtitle("Boxplot of Reading Values from Sensors")
```

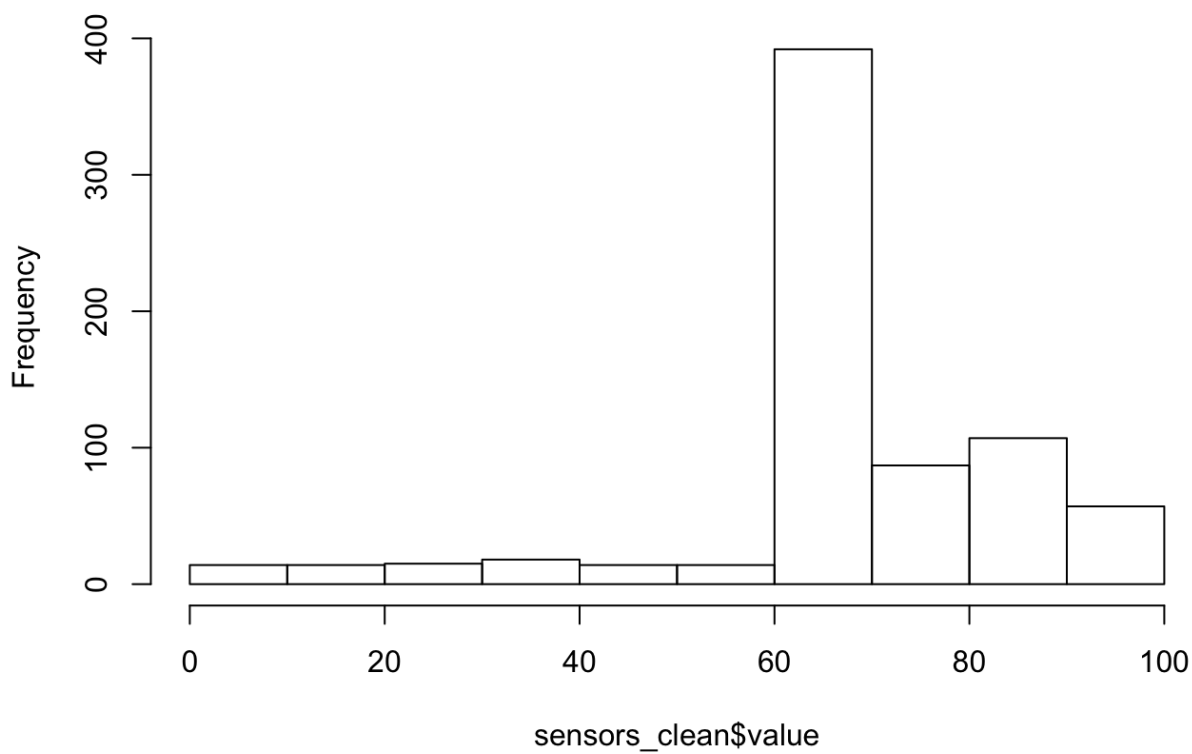


```
# Alternative using built-in hist and boxplot
```

```
# draw a histogram and a boxplot
```

```
hist(sensors_clean$value)
```

Histogram of sensors_clean\$value



```
# boxplot of just sensor ids
```

```
boxplot(value~sensorid, data = sensors_clean, main="Sensor Values by Sensor", xlab='Sensor ID', ylab='Reading')
```

