

M & M Histogram and Exercises

Ulrich Matter

2/19/2017

Preparations: Install/load packages

```
# Install required packages (if not yet installed)  
# install.packages(c("gsheet", "ggplot2", "data.table"))  
  
# Load required packages  
library(gsheet)  
library(ggplot2)  
library(data.table)
```

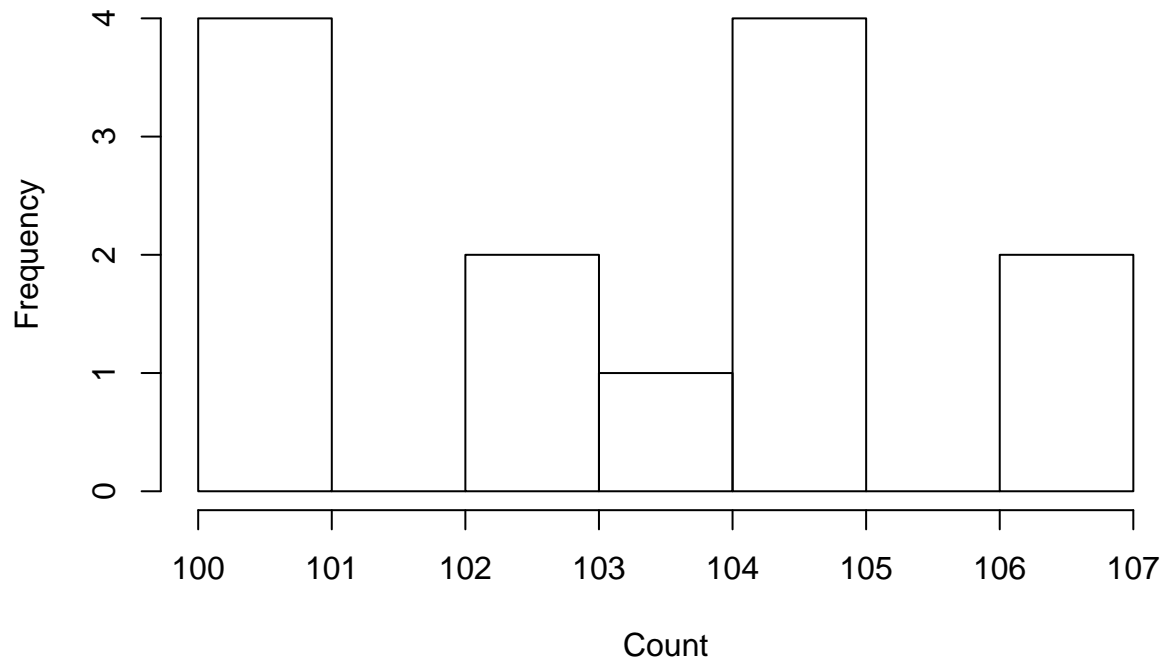
Fetch data from our Google-spreadsheet

```
# Fetch data from google docs  
URL <- 'https://docs.google.com/spreadsheets/d/1rQZasD4o2aLTf3-sfcFt0TMiRJudWOR_qW7ZXgw8Jt8/edit?usp=sh  
mnm_data <- gsheet2tbl(URL)
```

Example I: Plot a simple histogram for total counts

```
# Plot histogram of total occurrences  
hist(mnm_data$Total, main = "M&M Bag Histogram", xlab = "Count")
```

M&M Bag Histogram

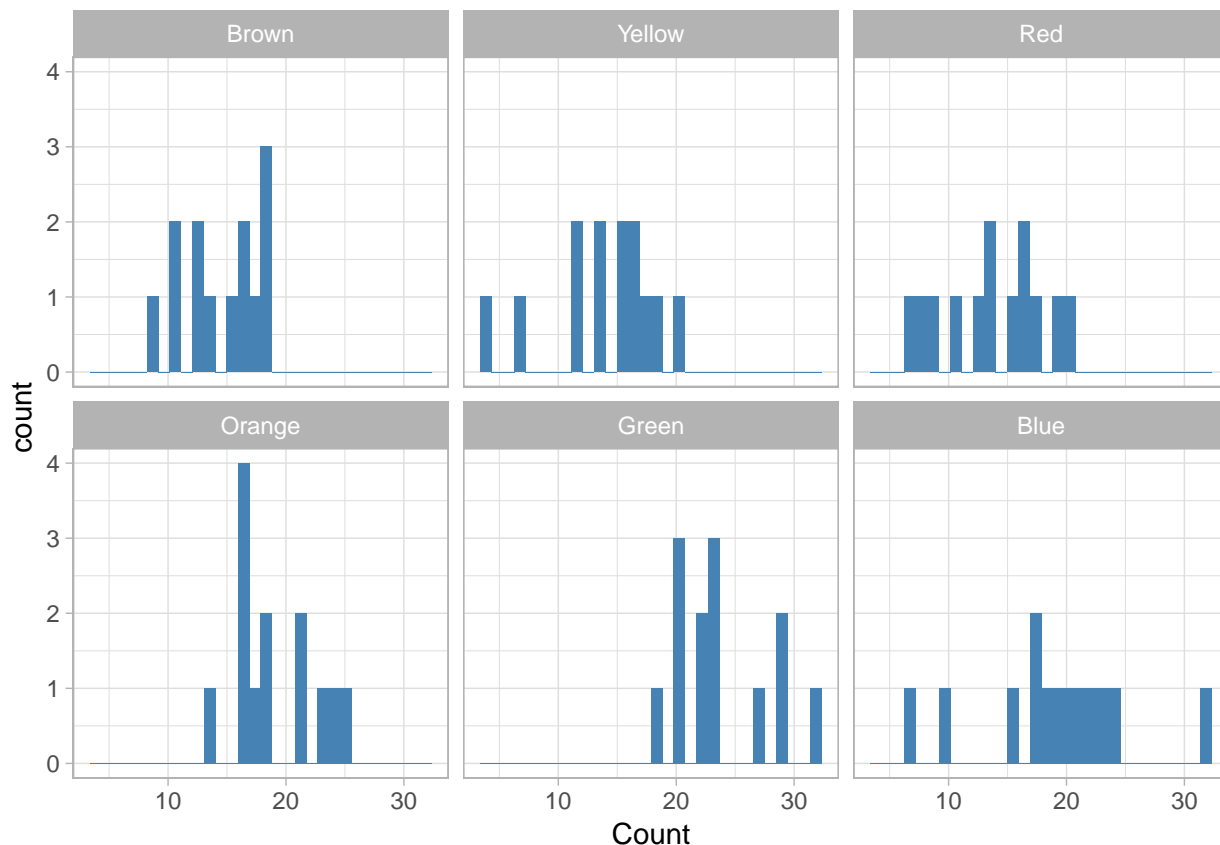


Example II: Plot a histogram for each color

Note: here we use **ggplot2**, a powerful R-package to visualize data.

```
# II ) FREQUENCY DISTRIBUTION OF M & Ms PER COLOR -----  
  
# get data for each color in 'long' format  
mnm_long <- as.data.table(mnm_data[, -ncol(mnm_data)])  
mnm_long <- melt(mnm_long, id.vars = "Name")  
names(mnm_long) <- c("Name", "Color", "Count")  
  
# Plot counts per color  
ggplot(data = mnm_long, aes(x = Count, group=1)) +  
  geom_histogram(fill = "steelblue") +  
  facet_wrap(~Color) +  
  theme_light()
```

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



Exercises

1. What is the average number of M&Ms per bag? For each color?

```
# compute the average number (i.e., the mean) of M&Ms per bag in our sample
```

```
mean(mnm_data$Total)
```

```
## [1] 103.5385
```

```
# compute the average number for each color
```

```
# a) select only those columns containing data on the count of M&Ms of a specific color  
# (i.e., columns 2 to 7)
```

```
mnm_colors <- mnm_data[, 2:7]
```

```
# b) use the already implemented R-function 'colMeans' to compute the mean for each column (color)
```

```
colMeans(mnm_colors)
```

```
##      Brown      Yellow      Red      Orange      Green      Blue
```

```
## 14.53846 13.84615 13.76923 18.84615 23.69231 18.84615
```

```
# or alternatively use sapply (see ?sapply for what this function does!)
```

```
sapply(mnm_colors, mean)
```

```
##      Brown      Yellow      Red      Orange      Green      Blue
```

```
## 14.53846 13.84615 13.76923 18.84615 23.69231 18.84615
```

2. What is the variance and standard deviation?

```
# for the total number of M&Ms
```

```
var(mnm_data$Total)
```

```
## [1] 5.935897
```

```
sd(mnm_data$Total)
```

```
## [1] 2.43637
```

```
# per color
```

```
sapply(mnm_colors, var)
```

```
##      Brown      Yellow      Red      Orange      Green      Blue
## 8.935897 18.974359 16.525641 12.641026 18.064103 39.474359
```

```
sapply(mnm_colors, sd)
```

```
##      Brown      Yellow      Red      Orange      Green      Blue
## 2.989297 4.355957 4.065174 3.555422 4.250189 6.282862
```

3. What is the range, median, and mode?

```
# for the total number of M&Ms
```

```
range(mnm_data$Total) # the range
```

```
## [1] 100 107
```

```
median(mnm_data$Total) # the median
```

```
## [1] 104
```

```
mymode(mnm_data$Total) # the mode (NOTE: we have implemented this function ourselves, see above!)
```

```
## [1] 105
```

```
# for each color
```

```
sapply(mnm_colors, range)
```

```
##      Brown Yellow Red Orange Green Blue
## [1,]      9      4  7      14      18      7
## [2,]     18     20 20     25     32     32
```

```
sapply(mnm_colors, median)
```

```
##      Brown Yellow      Red Orange      Green      Blue
##      15      15      14      18      23      19
```

```
sapply(mnm_colors, mymode)
```

```
##      Brown Yellow      Red Orange      Green      Blue
##      18      12      14      16      20      17
```