

jmaditat_Quant_Assignment1

```
chooseCRANmirror(graphics = getOption("menu.graphics"), ind = 79,  
                 local.only = FALSE)  
install.packages("dplyr")
```

```
## Installing package into 'C:/Users/ibeme/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'dplyr' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'dplyr'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:  
## \Users\ibeme\Documents\R\win-library\4.1\00LOCK\dplyr\libs\x64\dplyr.dll to C:  
## \Users\ibeme\Documents\R\win-library\4.1\dplyr\libs\x64\dplyr.dll: Permission  
## denied
```

```
## Warning: restored 'dplyr'
```

```
##
```

```
## The downloaded binary packages are in  
## C:\Users\ibeme\AppData\Local\Temp\RtmpOwlhSl\downloaded_packages
```

```
install.packages("Hmisc")
```

```
## Installing package into 'C:/Users/ibeme/Documents/R/win-library/4.1'  
## (as 'lib' is unspecified)
```

```
## package 'Hmisc' successfully unpacked and MD5 sums checked
```

```
## Warning: cannot remove prior installation of package 'Hmisc'
```

```
## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:  
## \Users\ibeme\Documents\R\win-library\4.1\00LOCK\Hmisc\libs\x64\Hmisc.dll to C:  
## \Users\ibeme\Documents\R\win-library\4.1\Hmisc\libs\x64\Hmisc.dll: Permission  
## denied
```

```
## Warning: restored 'Hmisc'
```

```
##
```

```
## The downloaded binary packages are in  
## C:\Users\ibeme\AppData\Local\Temp\RtmpOwlhSl\downloaded_packages
```

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(Hmisc)
```

```
## Loading required package: lattice  
  
## Loading required package: survival  
  
## Loading required package: Formula  
  
## Loading required package: ggplot2  
  
##  
## Attaching package: 'Hmisc'  
  
## The following objects are masked from 'package:dplyr':  
##  
##   src, summarize  
  
## The following objects are masked from 'package:base':  
##  
##   format.pval, units
```

```
#Dataset from Kaggle
```

```
Cereal_data<- read.csv("Cereals.csv")  
View(Cereal_data)  
head(Cereal_data)
```

```
##           name mfr type calories protein fat sodium fiber carbo  
## 1      100%_Bran   N    C        70         4   1   130  10.0   5.0  
## 2  100%_Natural_Bran Q    C       120         3   5    15   2.0   8.0  
## 3        All-Bran   K    C        70         4   1   260   9.0   7.0  
## 4 All-Bran_with_Extra_Fiber K    C        50         4   0   140  14.0   8.0  
## 5      Almond_Delight R    C       110         2   2   200   1.0  14.0  
## 6 Apple_Cinnamon_Cheerios G    C       110         2   2   180   1.5  10.5  
##   sugars potass vitamins shelf weight cups   rating  
## 1      6     280       25     3      1 0.33 68.40297  
## 2      8     135        0     3      1 1.00 33.98368  
## 3      5     320       25     3      1 0.33 59.42551  
## 4      0     330       25     3      1 0.50 93.70491  
## 5      8      NA       25     3      1 0.75 34.38484  
## 6     10      70       25     1      1 0.75 29.50954
```

```
#Statistics of the data
describe(Cereal_data)
```

```
## Cereal_data
##
## 16 Variables      77 Observations
## -----
## name
##      n missing distinct
##      77      0      77
##
## lowest : 100%_Bran      100%_Natural_Bran      All-Bran      All-Bran_with
## highest: Triples      Trix      Wheat_Chex      Wheaties
## -----
## mfr
##      n missing distinct
##      77      0      7
##
## lowest : A G K N P, highest: K N P Q R
##
## Value      A      G      K      N      P      Q      R
## Frequency    1    22    23     6     9     8     8
## Proportion 0.013 0.286 0.299 0.078 0.117 0.104 0.104
## -----
## type
##      n missing distinct
##      77      0      2
##
## Value      C      H
## Frequency    74     3
## Proportion 0.961 0.039
## -----
## calories
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      77      0      11    0.933    106.9    19.86      70      90
##      .25      .50      .75      .90      .95
##      100     110     110     124     140
##
## lowest :  50  70  80  90 100, highest: 120 130 140 150 160
##
## Value      50      70      80      90     100     110     120     130     140     150     160
## Frequency    3      2      1      7     17     29     10      2      3      2      1
## Proportion 0.039 0.026 0.013 0.091 0.221 0.377 0.130 0.026 0.039 0.026 0.013
## -----
## protein
##      n missing distinct      Info      Mean      Gmd
##      77      0      6    0.912     2.545     1.166
##
## lowest : 1 2 3 4 5, highest: 2 3 4 5 6
##
## Value      1      2      3      4      5      6
## Frequency   13     25     28     8      1      2
## Proportion 0.169 0.325 0.364 0.104 0.013 0.026
```

```

## -----
## fat
##      n missing distinct      Info      Mean      Gmd
##      77      0      5      0.892      1.013      1.049
##
## lowest : 0 1 2 3 5, highest: 0 1 2 3 5
##
## Value      0      1      2      3      5
## Frequency    27     30     14      5      1
## Proportion 0.351 0.390 0.182 0.065 0.013
## -----
## sodium
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      77      0      27      0.995     159.7     93.51      0      0
##      .25     .50     .75     .90     .95
##     130     180     210     254     282
##
## lowest :  0  15  45  70  75, highest: 250 260 280 290 320
## -----
## fiber
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      77      0      13      0.966     2.152     2.289      0.0      0.0
##      .25     .50     .75     .90     .95
##      1.0     2.0     3.0     4.4     5.2
##
## lowest :  0.0  1.0  1.5  2.0  2.5, highest:  5.0  6.0  9.0 10.0 14.0
##
## Value      0.0  1.0  1.5  2.0  2.5  2.7  3.0  4.0  5.0  6.0  9.0
## Frequency    19   16   3   10   1   1   15   4   4   1   1
## Proportion 0.247 0.208 0.039 0.130 0.013 0.013 0.195 0.052 0.052 0.013 0.013
##
## Value      10.0  14.0
## Frequency     1    1
## Proportion 0.013 0.013
## -----
## carbo
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      76      1      21      0.994     14.8     4.434      8.75     10.50
##      .25     .50     .75     .90     .95
##     12.00    14.50    17.00    21.00    21.00
##
## lowest :  5  7  8  9 10, highest: 19 20 21 22 23
## -----
## sugars
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      76      1      16      0.991     7.026     5.061      0.0      1.5
##      .25     .50     .75     .90     .95
##      3.0     7.0     11.0     13.0     14.0
##
## lowest :  0  1  2  3  4, highest: 11 12 13 14 15
##
## Value      0      1      2      3      4      5      6      7      8      9     10
## Frequency    7      1      3     13      1      5      7      4      5      4      5
## Proportion 0.092 0.013 0.039 0.171 0.013 0.066 0.092 0.053 0.066 0.053 0.066

```

```

##
## Value          11    12    13    14    15
## Frequency       5     7     4     3     2
## Proportion 0.066 0.092 0.053 0.039 0.026
## -----
## potass
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      75      2      35    0.998    98.67    74.3    25.0    30.0
##      .25     .50     .75     .90     .95
##      42.5    90.0    120.0   190.0   246.0
##
## lowest : 15 20 25 30 35, highest: 240 260 280 320 330
## -----
## vitamins
##      n missing distinct      Info      Mean      Gmd
##      77      0      3    0.451    28.25    15.64
##
## Value          0    25   100
## Frequency       8    63     6
## Proportion 0.104 0.818 0.078
## -----
## shelf
##      n missing distinct      Info      Mean      Gmd
##      77      0      3    0.86     2.208    0.8941
##
## Value          1     2     3
## Frequency      20    21    36
## Proportion 0.260 0.273 0.468
## -----
## weight
##      n missing distinct      Info      Mean      Gmd
##      77      0      7    0.426     1.03    0.1102
##
## lowest : 0.50 0.83 1.00 1.25 1.30, highest: 1.00 1.25 1.30 1.33 1.50
##
## Value          0.50 0.83 1.00 1.25 1.30 1.33 1.50
## Frequency       2     1    64     2     1     5     2
## Proportion 0.026 0.013 0.831 0.026 0.013 0.065 0.026
## -----
## cups
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      77      0      12    0.926     0.821    0.2522    0.466    0.500
##      .25     .50     .75     .90     .95
##      0.670    0.750    1.000    1.000    1.026
##
## lowest : 0.25 0.33 0.50 0.67 0.75, highest: 1.00 1.13 1.25 1.33 1.50
##
## Value          0.25 0.33 0.50 0.67 0.75 0.80 0.88 1.00 1.13 1.25 1.33
## Frequency       1     3     7    13    16     1     2    30     1     1     1
## Proportion 0.013 0.039 0.091 0.169 0.208 0.013 0.026 0.390 0.013 0.013 0.013
##
## Value          1.50
## Frequency       1
## Proportion 0.013

```

```
## -----
## rating
##      n missing distinct      Info      Mean      Gmd      .05      .10
##      77      0       77         1    42.67    15.48    22.67    27.92
##      .25     .50     .75     .90     .95
##    33.17    40.40    50.83    60.09    68.27
##
## lowest : 18.04285 19.82357 21.87129 22.39651 22.73645
## highest: 68.23588 68.40297 72.80179 74.47295 93.70491
## -----
```

```
summary(Cereal_data)
```

```
##      name      mfr      type      calories
## Length:77      Length:77      Length:77      Min.   : 50.0
## Class :character Class :character Class :character 1st Qu.:100.0
## Mode  :character Mode  :character Mode  :character Median :110.0
##                                     Mean   :106.9
##                                     3rd Qu.:110.0
##                                     Max.   :160.0
##
##      protein      fat      sodium      fiber
## Min.   :1.000    Min.   :0.000    Min.   : 0.0    Min.   : 0.000
## 1st Qu.:2.000    1st Qu.:0.000    1st Qu.:130.0   1st Qu.: 1.000
## Median :3.000    Median :1.000    Median :180.0   Median : 2.000
## Mean   :2.545    Mean   :1.013    Mean   :159.7   Mean   : 2.152
## 3rd Qu.:3.000    3rd Qu.:2.000    3rd Qu.:210.0   3rd Qu.: 3.000
## Max.   :6.000    Max.   :5.000    Max.   :320.0   Max.   :14.000
##
##      carbo      sugars      potass      vitamins
## Min.   : 5.0    Min.   : 0.000    Min.   : 15.00   Min.   : 0.00
## 1st Qu.:12.0    1st Qu.: 3.000    1st Qu.: 42.50   1st Qu.: 25.00
## Median :14.5    Median : 7.000    Median : 90.00   Median : 25.00
## Mean   :14.8    Mean   : 7.026    Mean   : 98.67   Mean   : 28.25
## 3rd Qu.:17.0    3rd Qu.:11.000    3rd Qu.:120.00   3rd Qu.: 25.00
## Max.   :23.0    Max.   :15.000    Max.   :330.00   Max.   :100.00
## NA's   :1      NA's   :1      NA's   :2
##      shelf      weight      cups      rating
## Min.   :1.000    Min.   :0.50    Min.   :0.250    Min.   :18.04
## 1st Qu.:1.000    1st Qu.:1.00    1st Qu.:0.670    1st Qu.:33.17
## Median :2.000    Median :1.00    Median :0.750    Median :40.40
## Mean   :2.208    Mean   :1.03    Mean   :0.821    Mean   :42.67
## 3rd Qu.:3.000    3rd Qu.:1.00    3rd Qu.:1.000    3rd Qu.:50.83
## Max.   :3.000    Max.   :1.50    Max.   :1.500    Max.   :93.70
##
```

```
#List of manufacturers with their total calories
```

```
data_by_mfr <- Cereal_data %>% group_by(mfr)
Calories_Count_by_mfr<-summarise(data_by_mfr,TotalCalories_of_each_mfr=sum(calories))
Calories_Count_by_mfr
```

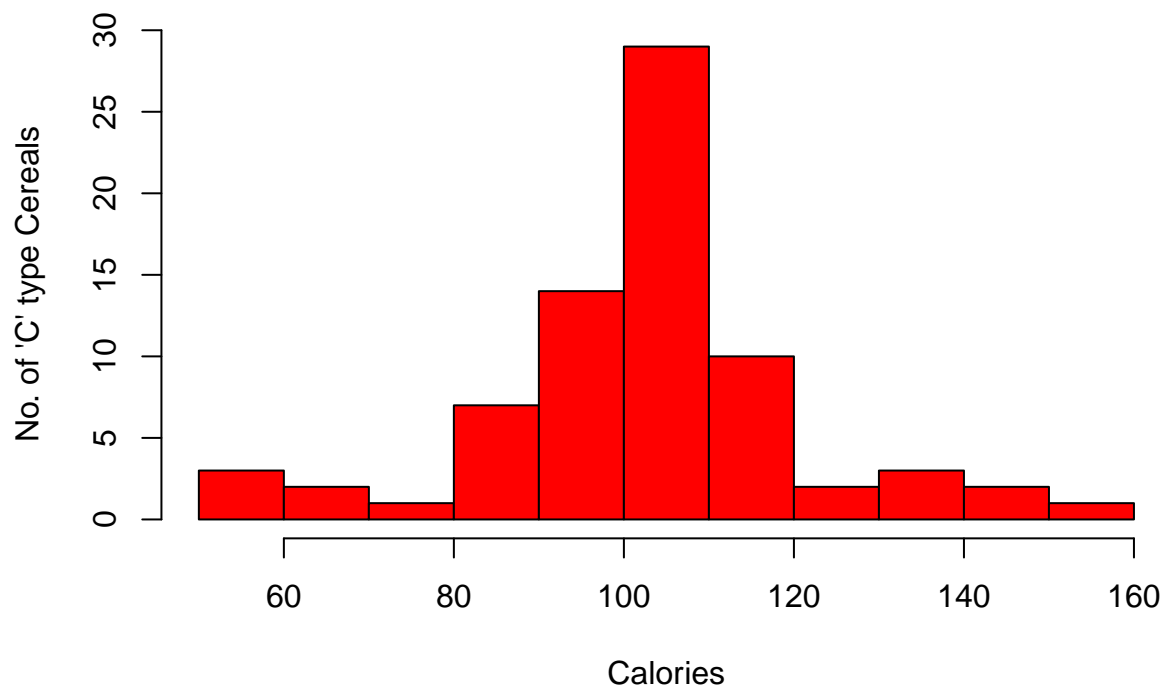
```
## # A tibble: 7 x 2
##   mfr      TotalCalories_of_each_mfr
```

```
##    <chr>                <int>
## 1 A                      100
## 2 G                     2450
## 3 K                     2500
## 4 N                      520
## 5 P                      980
## 6 Q                      760
## 7 R                      920
```

#Plots for calorie distribution of Hot and cold type of cereals

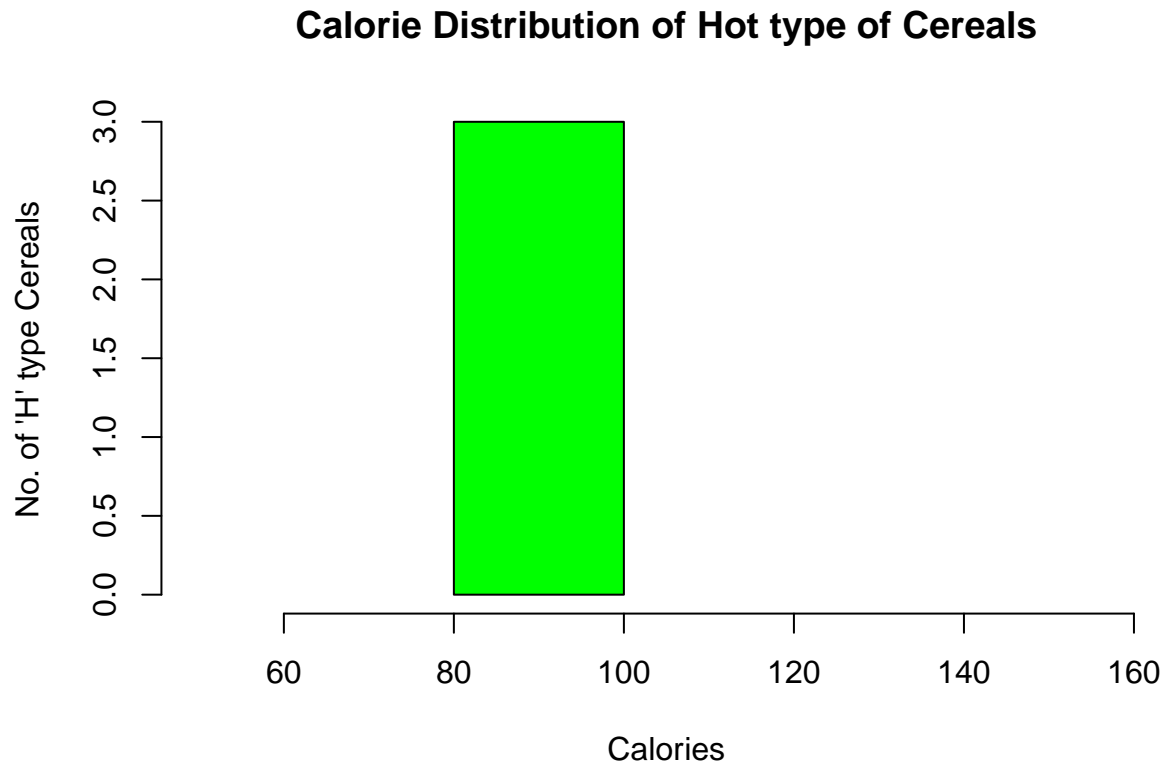
```
hist(Cereal_data$calories [Cereal_data$type == "C"],
     xlim = c(50,160) ,#Limiting the scale on x
     breaks = 10,
     main="Calorie Distribution of Cold Type of Cereals",
     xlab= "Calories",
     ylab="No. of 'C' type Cereals",
     col="red")
```

Calorie Distribution of Cold Type of Cereals



```
hist(Cereal_data$calories [Cereal_data$type == "H"],
     xlim = c(50,160) ,#Limiting the scale on x
     breaks = 10,
     main="Calorie Distribution of Hot type of Cereals",
     xlab="Calories",
     ylab="No. of 'H' type Cereals",
```

```
#add= TRUE,  
col="Green")
```



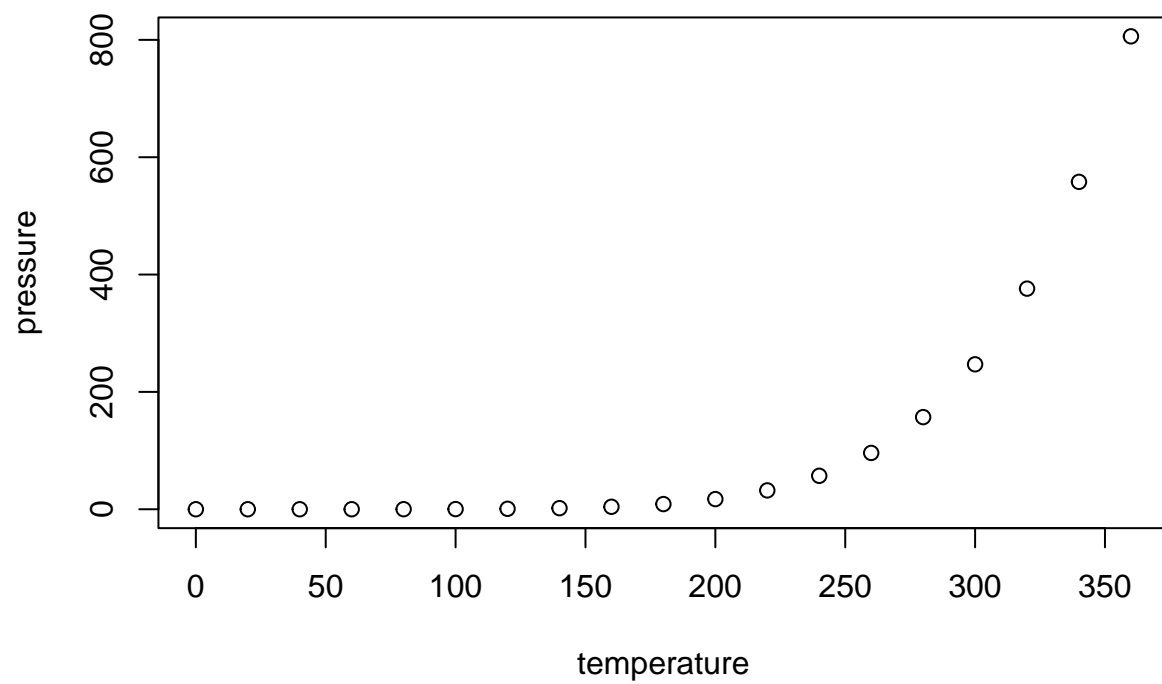
R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.