First load the file "MR_data.csv" into R:

Input:

```
MR <- read.csv(file.choose(), header = TRUE)
head(MR)
```

Output:

```
> MR <- read.csv(file.choose(), header = TRUE)
> head(MR)
  InvoiceNumber InvoiceDate Division            ProjectType HoursRequired ClientFeedbackScore InvoiceAmount RepeatClient
1        176563   4/21/2018   Mobile             Web_Design         100.7                  100      12856.62            0
2        176835   7/10/2018 Portland        Market_Research         174.4                   95      22822.32            1
3        177176   6/14/2018 Portland        Market_Research          94.9                   98      13489.30            1
4        177565   5/12/2018 Portland        Customer_Survey          80.1                   99      10122.28            1
5        177958   10/5/2018 Portland        Market_Research          72.1                   83      10640.40            0
6        178262   8/28/2018  Atlanta Social_Media_Analytics          67.6                   73       4194.47            1
>
```

1) Which of the firm's divisions offered web design services in 2018 for their clients?

Answer: The firm's divisions '**Mobile'** and '**Denver'** offered web design services in 2018 to their clients.

Input:

```
div <- subset(MR, MR$ProjectType == "Web_Design", c("Division"))
div
unique(div)
```

Output:

```
> div <- subset(MR, MR$ProjectType == "Web_Design", c("Division"))
> div
    Division
1     Mobile
10    Denver
11    Mobile
12    Mobile
16    Mobile
>
> unique(div)
   Division
1    Mobile
10   Denver
>
```

2) What is the average Client Feedback Score in the Atlanta division for projects which required more than 100 hours?

Answer: The average Client Feedback Score in the Atlanta division for projects that required more than 100 hours, is **61**.

Input:

```
AT <- MR[MR$Division =='Atlanta', ]
AT
AT_100 <- AT[AT$HoursRequired > 100, ]
```

AT_100
mean(AT_100$ClientFeedbackScore)

Output:

```
> AT <- MR[MR$Division =='Atlanta', ]
> AT
    InvoiceNumber InvoiceDate Division           ProjectType HoursRequired ClientFeedbackScore InvoiceAmount RepeatClient
6          178262   8/28/2018  Atlanta Social_Media_Analytics          67.6                  73      4194.470            1
23         182249    5/3/2018  Atlanta Social_Media_Analytics          80.1                  68      6417.504            0
25         182784   2/17/2018  Atlanta Social_Media_Analytics          91.8                  70      8750.409            0
41         187241   8/11/2018  Atlanta Social_Media_Analytics          90.5                  60     12625.969            0
42         187428   3/31/2018  Atlanta Social_Media_Analytics          70.8                  57      7385.005            0
57         190434  12/16/2018  Atlanta Social_Media_Analytics          88.7                  70     10170.297            0
63         191761   2/26/2018  Atlanta Social_Media_Analytics          89.2                  52      8929.196            0
66         192353    3/9/2018  Atlanta Social_Media_Analytics          89.2                  68      9891.326            0
71         193635    6/6/2018  Atlanta Social_Media_Analytics          67.2                  56      6140.590            0
99         200460   6/10/2018  Atlanta Social_Media_Analytics          76.5                  61      6947.698            0
105        202102   8/30/2018  Atlanta Social_Media_Analytics          85.8                  48      9916.879            1
121        205698  11/17/2018  Atlanta Social_Media_Analytics          62.6                  54      5257.010            0
134        208329  10/30/2018  Atlanta Social_Media_Analytics          82.7                  40      9631.106            0
139        209699   8/26/2018  Atlanta Social_Media_Analytics          75.2                  53     10404.764            0
149        211866   10/9/2018  Atlanta Social_Media_Analytics          86.0                  63      9304.126            1
182        220214   8/13/2018  Atlanta Social_Media_Analytics          90.4                  55      7500.974            0
197        223772    4/1/2018  Atlanta        Market_Research         133.9                  57     14224.613            0
202        224751  12/15/2018  Atlanta Social_Media_Analytics          79.9                  55      7138.371            0
>
> AT_100 <- AT[AT$HoursRequired > 100,]
> AT_100
    InvoiceNumber InvoiceDate Division           ProjectType HoursRequired ClientFeedbackScore InvoiceAmount RepeatClient
197        223772    4/1/2018  Atlanta        Market_Research         133.9                  57     14224.613            0
330        257284    7/4/2018  Atlanta        Market_Research         125.0                  87     15709.106            1
458        289850   3/16/2018  Atlanta        Market_Research         103.5                  53      7757.702            0
462        290487   5/23/2018  Atlanta        Market_Research         216.7                  58     24085.390            0
463        290814   12/9/2018  Atlanta Social_Media_Analytics         110.6                  50      9987.136            0
>
>
> mean(AT_100$ClientFeedbackScore)
[1] 61
>
```

3) What is the range (max and min) and average hourly rate for market research projects across all divisions?

Answer: For the market research projects across all divisions, the **range** (min and max) of hourly rate is **65.75 to 163.89**, while the **average** hourly rate is **132.73**

Input:

```
MR$HR <- (MR$InvoiceAmount / MR$HoursRequired)
head(MR)
Mini <- min(MR$HR[MR$ProjectType=='Market_Research'])
Mini
Maxi <- max(MR$HR[MR$ProjectType=='Market_Research'])
Maxi
range(MR$HR[MR$ProjectType=='Market_Research'])
AvgHR <- mean(MR$HR[MR$ProjectType=='Market_Research'])
AvgHR
```

Output:

```
> MR$HR <- (MR$InvoiceAmount / MR$HoursRequired)
> head(MR)
  InvoiceNumber InvoiceDate Division            ProjectType HoursRequired ClientFeedbackScore InvoiceAmount RepeatClient        HR
1        176563   4/21/2018   Mobile             Web_Design         100.7                  100      12856.62            0 127.67251
2        176835   7/10/2018 Portland        Market_Research         174.4                   95      22822.32            1 130.86190
3        177176   6/14/2018 Portland        Market_Research          94.9                   98      13489.30            1 142.14226
4        177565   5/12/2018 Portland        Customer_Survey          80.1                   99      10122.28            1 126.37048
5        177958   10/5/2018 Portland        Market_Research          72.1                   83      10640.40            0 147.57835
6        178262   8/28/2018  Atlanta Social_Media_Analytics          67.6                   73       4194.47            1  62.04837
> Mini <- min(MR$HR[MR$ProjectType=='Market_Research'])
> Mini
[1] 65.74583
> Maxi <- max(MR$HR[MR$ProjectType=='Market_Research'])
> Maxi
[1] 163.8851
> range(MR$HR[MR$ProjectType=='Market_Research'])
[1]  65.74583 163.88515
> AvgHR <- mean(MR$HR[MR$ProjectType=='Market_Research'])
> AvgHR
[1] 132.7266
>
```

4) What was the invoice number and the invoice date for the market research project that had the lowest hourly rate?

Answer: For the market research project that had the lowest hourly rate, the **invoice number** is **313125** and the **invoice date** is **11/2/2018**

Input:

MR$InvoiceNumber[MR$ProjectType == 'Market_Research' & MR$HR == Mini]

MR$InvoiceDate[MR$ProjectType == 'Market_Research' & MR$HR == Mini]

Output:

```
>
> MR$InvoiceNumber[MR$ProjectType == 'Market_Research' & MR$HR == Mini]
[1] 313125
> MR$InvoiceDate[MR$ProjectType == 'Market_Research' & MR$HR == Mini]
[1] "11/2/2018"
>
```

5) An analyst at corporate headquarters took the projects from 2018 across all divisions and divided them into quintiles (bottom 20%, ..., top 20%) by Client Feedback Score. Which division had more projects in the bottom 20% Client Feedback Scores than any other division?

Answer: The division which had most projects in the bottom 20% of Client Feedback Scores is **Mobile**, with score of 61.

Input:

Q <- quantile(MR$ClientFeedbackScore, probs = seq(0, 1, 1/5))

QuintTable <- cut(MR$ClientFeedbackScore, quantile(MR$ClientFeedbackScore, probs = seq(0, 1, 1/5)), include.lowest = TRUE)

A <- table(QuintTable, MR$Division)

A

max(A[1,])

Output:

```
>
> Q <- quantile(MR$ClientFeedbackScore, probs = seq(0, 1, 1/5))
> QuintTable <- cut(MR$ClientFeedbackScore, quantile(MR$ClientFeedbackScore, probs = seq(0, 1, 1/5)), include.lowest = TRUE)
> A <- table(QuintTable, MR$Division)
> A

QuintTable  Atlanta Denver Mobile Portland
  [11,72]        44     22     61        0
  (72,84.2]       9     56     36       20
  (84.2,91]       2     24     31       75
  (91,97]         1     20     14       82
  (97,100]        0      9     38       75
>
> max(A[1, ])
[1] 61
>
```

6) The analyst also created the table below, showing the distribution of each division's project Client Feedback Scores. For example, 30% of Portland's projects had scores between 97 and 100, which was the top 20% score group across all divisions. Submit the R code to recreate this table.

	[11,72]	(72,84.2]	(84.2,91]	(91,97]	(97,100]
Atlanta	0.79	0.16	0.04	0.02	0.00
Denver	0.17	0.43	0.18	0.15	0.07
Mobile	0.34	0.20	0.17	0.08	0.21
Portland	0.00	0.08	0.30	0.33	0.30

Answer/ Input:

```
percent_CFS <- function(CFS_data, val_low, val_high) {
 num <- length(CFS_data)
 num_in_range <- CFS_data[CFS_data > val_low & CFS_data <= val_high]
 percent <- length(num_in_range)/num
 return(percent)
}

atlanta_CFS = MR[MR$Division == "Atlanta", c("ClientFeedbackScore")]
a1 <- percent_CFS(atlanta_CFS, 11, 72)
a2 <- percent_CFS(atlanta_CFS, 72, 84.2)
a3 <- percent_CFS(atlanta_CFS, 84.2, 91)
a4 <- percent_CFS(atlanta_CFS, 91, 97)
a5 <- percent_CFS(atlanta_CFS, 97, 100)
A <- round(c(a1, a2, a3, a4, a5), digits = 2)
A

denver_CFS = MR[MR$Division == "Denver", c("ClientFeedbackScore")]
d1 <- percent_CFS(denver_CFS, 11, 72)
d2 <- percent_CFS(denver_CFS, 72, 84.2)
d3 <- percent_CFS(denver_CFS, 84.2, 91)
d4 <- percent_CFS(denver_CFS, 91, 97)
d5 <- percent_CFS(denver_CFS, 97, 100)
D <- round(c(d1, d2, d3, d4, d5), digits = 2)
D
```

```
mobile_CFS = MR[MR$Division == "Mobile", c("ClientFeedbackScore")]
m1 <- percent_CFS(mobile_CFS, 11, 72)
m2 <- percent_CFS(mobile_CFS, 72, 84.2)
m3 <- percent_CFS(mobile_CFS, 84.2, 91)
m4 <- percent_CFS(mobile_CFS, 91, 97)
m5 <- percent_CFS(mobile_CFS, 97, 100)
M <- round(c(m1, m2, m3, m4, m5), digits = 2)
M

portland_CFS <- MR[MR$Division == "Portland", c("ClientFeedbackScore")]
p1 <- percent_CFS(portland_CFS, 11, 72)
p2 <- percent_CFS(portland_CFS, 72, 84.2)
p3 <- percent_CFS(portland_CFS, 84.2, 91)
p4 <- percent_CFS(portland_CFS, 91, 97)
p5 <- percent_CFS(portland_CFS, 97, 100)
P <- round(c(p1, p2, p3, p4, p5), digits = 2)
P

CFStable <- matrix(c(A, D, M, P), nrow = 4, byrow = TRUE)

rname <- c("Atlanta", "Denver", "Mobile", "Portland")
rownames(CFStable) <- rname
cname <- c("[11,72]", "(72,84.2]", "(84.2,91]", "(91-97]", "(97-100]")
colnames(CFStable) <- cname

CFStable
```

Output:

```
>
> percent_CFS <- function(CFS_data, val_low, val_high) {
+   num <- length(CFS_data)
+   num_in_range <- CFS_data[CFS_data > val_low & CFS_data <= val_high]
+   percent <- length(num_in_range)/num
+   return(percent)
+ }
> atlanta_CFS = MR[MR$Division == "Atlanta", c("ClientFeedbackScore")]
> a1 <- percent_CFS(atlanta_CFS, 11, 72)
> a2 <- percent_CFS(atlanta_CFS, 72, 84.2)
> a3 <- percent_CFS(atlanta_CFS, 84.2, 91)
> a4 <- percent_CFS(atlanta_CFS, 91, 97)
> a5 <- percent_CFS(atlanta_CFS, 97, 100)
> A <- round(c(a1, a2, a3, a4, a5), digits = 2)
> A
[1] 0.79 0.16 0.04 0.02 0.00
> denver_CFS = MR[MR$Division == "Denver", c("ClientFeedbackScore")]
```

```
> p2 <- percent_CFS(portland_CFS, 72, 84.2)
> p3 <- percent_CFS(portland_CFS, 84.2, 91)
> p4 <- percent_CFS(portland_CFS, 91, 97)
> p5 <- percent_CFS(portland_CFS, 97, 100)
> P <- round(c(p1, p2, p3, p4, p5), digits = 2)
> P
[1] 0.00 0.08 0.30 0.33 0.30
> CFStable <- matrix(c(A, D, M, P), nrow = 4, byrow = TRUE)
> rname <- c("Atlanta", "Denver", "Mobile", "Portland")
> rownames(CFStable) <- rname
> cname <- c("[11,72]", "(72,84.2]", "(84.2,91]", "(91-97]", "(97-100]")
> colnames(CFStable) <- cname
> CFStable
         [11,72] (72,84.2] (84.2,91] (91-97] (97-100]
Atlanta     0.79      0.16      0.04    0.02     0.00
Denver      0.17      0.43      0.18    0.15     0.07
Mobile      0.33      0.20      0.17    0.08     0.21
Portland    0.00      0.08      0.30    0.33     0.30
>
```

7) Create a conditional box plot of ClientFeedbackScore by RepeatClient. How does the customer satisfaction of new clients compare to repeat clients?

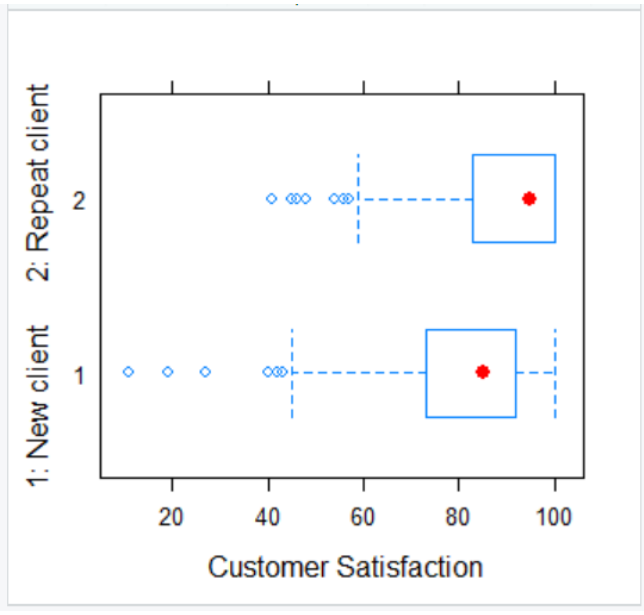
Answer: Client Feedback Score of repeat clients varies much less than that of new clients and is also concentrated towards the higher end, thus can be considered to be more consistent and dependable to understand customer satisfaction.

Input:

```
any(is.na(MR))
library(lattice)
bwplot(MR$RepeatClient ~ MR$ClientFeedbackScore, dataset = MR, xlab = " Customer Satisfaction",
    ylab = "1: New client     2: Repeat client", col = "red")
```

Output:

```
> any(is.na(MR))
[1] FALSE
> library(lattice)
> bwplot(MR$RepeatClient ~ MR$ClientFeedbackScore, dataset = MR, xlab = " Customer Satisfaction", ylab = "1:
 New client      2: Repeat client", col = "blue")
>
> bwplot(MR$RepeatClient ~ MR$ClientFeedbackScore, dataset = MR, xlab = " Customer Satisfaction",
+        ylab = "1: New client      2: Repeat client", col = "red")
>
```



8) Create a visualization of Hours Required by Project Type. You may choose what type of chart to use, pick one that you think makes it easy to discern patterns in the data.

Answer/ Input:

any(is.na(MR))

library(ggplot2)

ggplot(MR, aes(x = HoursRequired, y = ProjectType)) + geom_boxplot(col = “red”)

Output:

```
> any(is.na(MR))
[1] FALSE
> library(ggplot2)
> ggplot(MR, aes(x = HoursRequired, y = ProjectType)) + geom_boxplot(col = "red")
>
```