2)Import the dataset into R

```
spotify_data<-read.csv("C:\\Users\\pakan\\Desktop\\CS\\spotify_songs.csv")
head(spotify_data, n=3)
```

```
##                 track.id                                                track_name
## 1 6f807x0ima9a1j3VPbc7VN I Don't Care (with Justin Bieber) - Loud Luxury Remix
## 2 0r7CVbZTWZgbTCYdfa2P31                        Memories - Dillon Francis Remix
## 3 1z1Hg7Vb0AhHDiEmnDE79l                        All the Time - Don Diablo Remix
##   track_artist track_popularity         track_album_id
## 1   Ed Sheeran               66 2oCs0DGTsRO98Gh5ZSl2Cx
## 2     Maroon 5               67 63rPSO264uRjW1X5E6cWv6
## 3 Zara Larsson               70 1HoSmj2eLcsrR0vE9gThr4
##                                    track_album_name
## 1 I Don't Care (with Justin Bieber) [Loud Luxury Remix]
## 2                       Memories (Dillon Francis Remix)
## 3                       All the Time (Don Diablo Remix)
##   track_album_release_date playlist_name            playlist_id playlist_genre
## 1                6/14/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 2               12/13/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 3                7/5/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
##   playlist_subgenre danceability energy key loudness mode speechiness
## 1         dance pop        0.748  0.916   6   -2.634    1      0.0583
## 2         dance pop        0.726  0.815  11   -4.969    1      0.0373
## 3         dance pop        0.675  0.931   1   -3.432    0      0.0742
##   acousticness instrumentalness liveness valence   tempo duration_ms
## 1       0.1020         0.00e+00   0.0653   0.518 122.036      194754
## 2       0.0724         4.21e-03   0.3570   0.693  99.972      162600
## 3       0.0794         2.33e-05   0.1100   0.613 124.008      176616
```

3)Print out descriptive statistics for a selection of quantitative and categorical variables.

```
#quantitative variable
summary(spotify_data$track_popularity)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   24.00   45.00   42.48   62.00  100.00
```

```
#categorical variable
table(head(spotify_data$track_artist))
```

```
##
##       Ed Sheeran    Lewis Capaldi         Maroon 5 The Chainsmokers
##                2                1                1                1
##     Zara Larsson
##                1
```

4)Transform at least one variable. It doesn't matter what the transformation is

```
# Transform a quantitative variable
spotify_data$log_transformed_var <- log(spotify_data$track_popularity)
head(spotify_data)
```

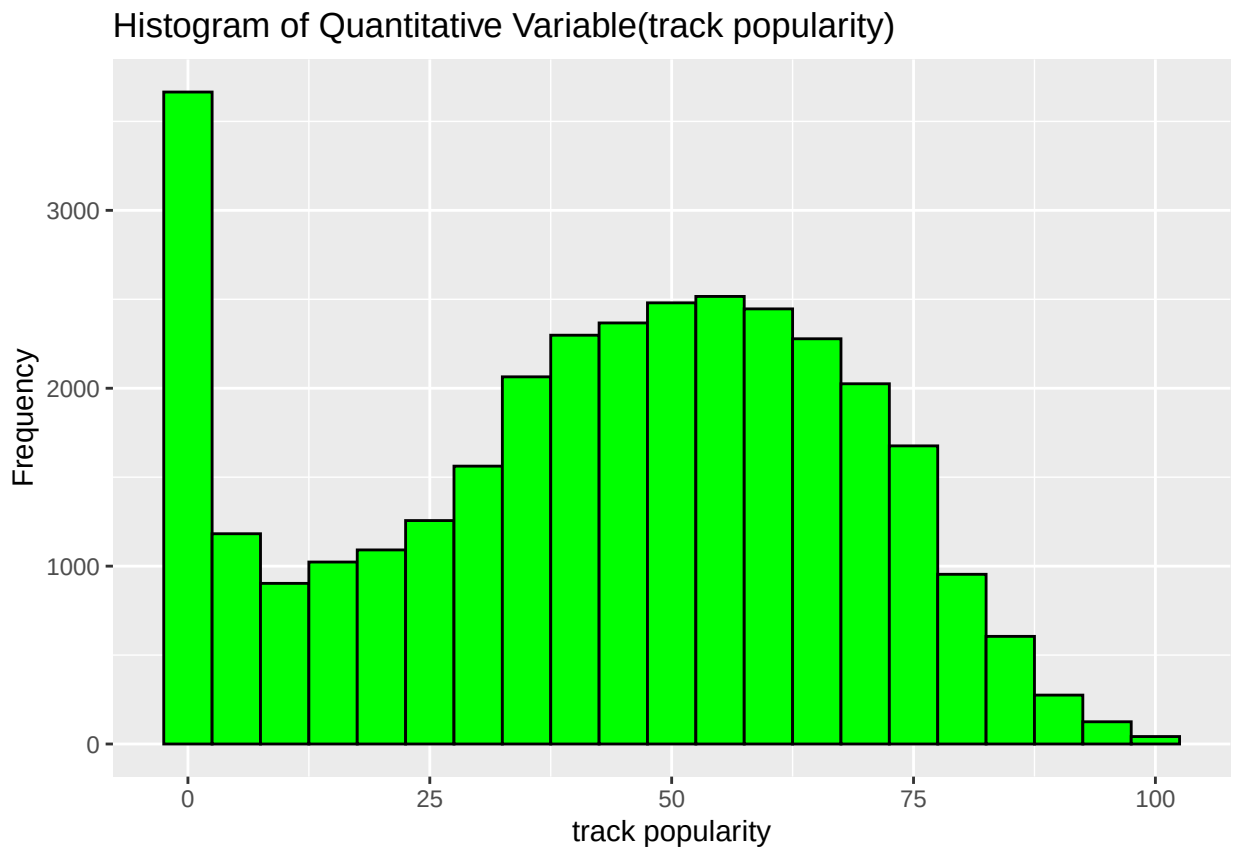
```
##                 track.id                                              track_name
## 1 6f807x0ima9a1j3VPbc7VN I Don't Care (with Justin Bieber) - Loud Luxury Remix
## 2 0r7CVbZTWZgbTCYdfa2P31                        Memories - Dillon Francis Remix
## 3 1z1Hg7Vb0AhHDiEmnDE79l                        All the Time - Don Diablo Remix
## 4 75FpbthrwQmzHlBJLuGdC7                      Call You Mine - Keanu Silva Remix
## 5 1e8PAfcKUYoKkxPhrHqw4x                 Someone You Loved - Future Humans Remix
## 6 7fvUMiyapMsRRxr07cU8Ef     Beautiful People (feat. Khalid) - Jack Wins Remix
##       track_artist track_popularity         track_album_id
## 1       Ed Sheeran               66 2oCs0DGTsRO98Gh5ZSl2Cx
## 2         Maroon 5               67 63rPSO264uRjW1X5E6cWv6
## 3     Zara Larsson               70 1HoSmj2eLcsrR0vE9gThr4
## 4 The Chainsmokers               60 1nqYsOef1yKKuGOVchbsk6
## 5    Lewis Capaldi               69 7m7vv9wlQ4i0LFuJiE2zsQ
## 6       Ed Sheeran               67 2yiy9cd2QktrNvWC2EUi0k
##                                     track_album_name
## 1 I Don't Care (with Justin Bieber) [Loud Luxury Remix]
## 2                       Memories (Dillon Francis Remix)
## 3                       All the Time (Don Diablo Remix)
## 4                           Call You Mine - The Remixes
## 5                Someone You Loved (Future Humans Remix)
## 6     Beautiful People (feat. Khalid) [Jack Wins Remix]
##   track_album_release_date playlist_name            playlist_id playlist_genre
## 1                6/14/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 2               12/13/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 3                7/5/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 4               7/19/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 5                3/5/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
## 6               7/11/2019     Pop Remix 37i9dQZF1DXcZDD7cfEKhW            pop
##   playlist_subgenre danceability energy key loudness mode speechiness
## 1         dance pop        0.748  0.916   6   -2.634    1      0.0583
## 2         dance pop        0.726  0.815  11   -4.969    1      0.0373
## 3         dance pop        0.675  0.931   1   -3.432    0      0.0742
## 4         dance pop        0.718  0.930   7   -3.778    1      0.1020
## 5         dance pop        0.650  0.833   1   -4.672    1      0.0359
## 6         dance pop        0.675  0.919   8   -5.385    1      0.1270
##   acousticness instrumentalness liveness valence   tempo duration_ms
## 1       0.1020         0.00e+00   0.0653   0.518 122.036      194754
## 2       0.0724         4.21e-03   0.3570   0.693  99.972      162600
## 3       0.0794         2.33e-05   0.1100   0.613 124.008      176616
## 4       0.0287         9.43e-06   0.2040   0.277 121.956      169093
## 5       0.0803         0.00e+00   0.0833   0.725 123.976      189052
## 6       0.0799         0.00e+00   0.1430   0.585 124.982      163049
##   log_transformed_var
## 1            4.189655
## 2            4.204693
## 3            4.248495
## 4            4.094345
## 5            4.234107
## 6            4.204693
```

5)Plot at least one quantitative variable, and one scatter plot

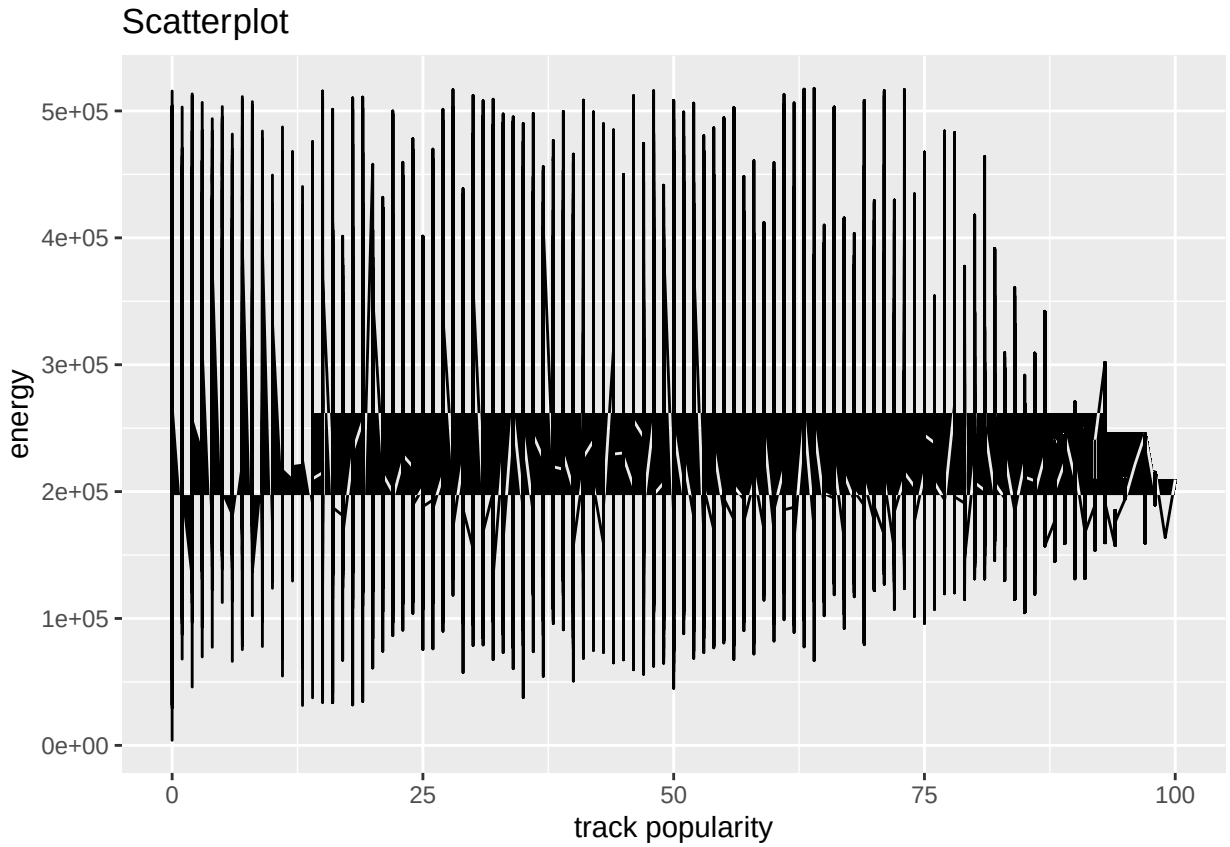
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.2
```

```
ggplot(spotify_data, aes(x = track_popularity)) +
  geom_histogram(binwidth = 5, fill = "green", color = "black") +
  labs(title = "Histogram of Quantitative Variable(track popularity)", x = "track popularity", y = "Fre
```




```
ggplot(spotify_data, aes(x = track_popularity, y =duration_ms)) +
  geom_line() +
  labs(title = "Scatterplot", x = "track popularity", y = "energy")
```


Scatterplot
energy
5e+05
4e+05
3e+05
2e+05
1e+05
0e+00
0
25
50
75
100
track popularity