

Rcode2_Spotify_Team3

2022-08-09

```
library("ggplot2")
library("magrittr")
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("readxl")

library(readxl)
Spotify <- read_excel("E:/Canada/College/GBC/T405/Semester 2/4066 Introduction to Analytics Systems/Ass

#1 Print the structure of your dataset
str(Spotify)

## # tibble [53 x 17] (S3: tbl_df/tbl/data.frame)
## $ id : chr [1:53] "6DCZcSspjsKoFjzjrWoCd" "3ee8Jmje8o58CHK66QrVC" "0e7ipj03S05BNilyu5b"
## $ Title : chr [1:53] "God's Plan" "SAD!" "rockstar (feat. 21 Savage)" "Psycho (feat. Ty D
## $ Artist : chr [1:53] "Drake" "XXXTENTACION" "Post Malone" "Post Malone" ...
## $ Genre : chr [1:53] "Hip-Hop/Rap" "Hip-Hop/Rap" "Hip-Hop/Rap" "Hip-Hop/Rap" ...
## $ Danceability : num [1:53] 75 74 59 75 84 68 82 79 84 63 ...
## $ Energy : num [1:53] 45 61 52 56 63 58 73 86 54 88 ...
## $ Key : num [1:53] 7 8 5 8 1 10 5 9 7 9 ...
## $ dB : num [1:53] -9.21 -4.88 -6.09 -8.01 -5.83 ...
## $ Mode : num [1:53] 1 1 0 1 1 1 0 0 1 0 ...
## $ Speechiness : num [1:53] 11 14 7 11 13 4 13 11 9 5 ...
## $ Acousticness : num [1:53] 3 26 12 55 6 33 10 4 4 21 ...
## $ Instrumentalness: num [1:53] 8.29e-05 3.72e-03 6.56e-05 0.00 6.00e-05 0.00 0.00 2.19e-05 0.00 0.00
## $ Liveness : num [1:53] 55 12 13 11 40 14 37 8 8 13 ...
## $ Valence : num [1:53] 36 47 13 46 35 34 65 59 51 53 ...
## $ Tempo : num [1:53] 77 75 160 140 91 145 136 124 97 95 ...
## $ Duration_secs : num [1:53] 199 167 218 221 218 231 253 215 218 203 ...
## $ Popularity : num [1:53] 87 88 89 85 84 88 85 86 84 85 ...
```

```
#2 List the variables in your dataset  
names(Spotify)
```

```
## [1] "id"           "Title"         "Artist"        "Genre"  
## [5] "Danceability" "Energy"        "Key"          "dB"  
## [9] "Mode"          "Speechiness"   "Acousticness" "Instrumentalness"  
## [13] "Liveness"      "Valence"       "Tempo"        "Duration_secs"  
## [17] "Popularity"
```

```
#3 Print the top 15 rows of your dataset  
head(Spotify, n=15)
```

```
## # A tibble: 15 x 17  
##   id     Title Artist Genre Danceability Energy Key dB Mode Speechiness  
##   <chr>  <chr> <chr> <chr>    <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 6DCZcSs~ God'~ Drake Hip--    75     45    7 -9.21  1    11  
## 2 3ee8Jmj~ SAD!~ XXXTE~ Hip--   74     61    8 -4.88  1    14  
## 3 0e7ipj0~ rock~ Post ~ Hip--  59     52    5 -6.09  0     7  
## 4 3swc6WT~ Psyc~ Post ~ Hip--  75     56    8 -8.01  1    11  
## 5 2G7V7zs~ In M~ Drake Hip--  84     63    1 -5.83  1    13  
## 6 7dt6x5M~ Bett~ Post ~ Hip--  68     58   10 -5.84  1     4  
## 7 58q2HKr~ I Li~ Cardi~ Hip--  82     73    5 -4.00  0    13  
## 8 7ef4Dls~ One ~ Calvi~ Danc~  79     86    9 -3.24  0    11  
## 9 76cy1WJ~ IDGAF Dua L~ Pop   84     54    7 -5.98  1     9  
## 10 08bNPGL~ FRIE~ Marsh~ Danc~  63     88    9 -2.38  0     5  
## 11 1rfofaq~ Hava~ Camil~ Pop   77     52    2 -4.33  1     3  
## 12 0s3nnoM~ Luci~ Juice~ Hip--  51     57    6 -7.23  0    20  
## 13 3CA9pLi~ Nice~ Drake Hip--  59     91    8 -6.47  1     7  
## 14 7fa9MBX~ Girl~ Maroo~ Pop   85     54    0 -6.82  1     5  
## 15 09IStsI~ The ~ Zedd Pop   75     66    7 -3.06  1     4  
## # ... with 7 more variables: Acousticness <dbl>, Instrumentalness <dbl>,  
## #   Liveness <dbl>, Valence <dbl>, Tempo <dbl>, Duration_secs <dbl>,  
## #   Popularity <dbl>
```

```
#4 Write a user defined function using any of the variables from the data set  
duration_in_mins <-function(x){  
  x/60  
}  
duration_in_mins(Spotify$Duration_secs)
```

```
## [1] 3.316667 2.783333 3.633333 3.683333 3.633333 3.850000 4.216667 3.583333  
## [9] 3.633333 3.383333 3.616667 4.000000 3.516667 3.933333 3.083333 3.866667  
## [17] 3.433333 2.883333 2.250000 3.016667 3.516667 6.966667 2.183333 3.383333  
## [25] 3.483333 3.900000 3.366667 2.733333 1.983333 4.383333 3.883333 3.716667  
## [33] 3.716667 2.616667 2.900000 2.900000 3.783333 3.300000 2.033333 3.050000  
## [41] 3.683333 3.333333 5.216667 3.116667 3.700000 3.600000 3.616667 3.466667  
## [49] 3.533333 3.616667 3.466667 3.533333 3.600000
```

```
#5 Use data manipulation techniques and filter rows based on any logical criteria that exist in your da  
Spotify %>% filter(Genre == "Pop")
```

```

## # A tibble: 17 x 17
##   id      Title Artist Genre Danceability Energy Key   dB Mode Speechiness
##   <chr>    <chr> <chr> <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 76cy1WJ~ IDGAF Dua L~ Pop        84     54    7 -5.98  1     9
## 2 1rfFAQ~ Hava~ Camil~ Pop        77     52    2 -4.33  1     3
## 3 7fa9MBX~ Girl~ Maroo~ Pop       85     54    0 -6.82  1     5
## 4 09IStsI~ The ~ Zedd Pop        75     66    7 -3.06  1     4
## 5 2qT1uLX~ no t~ Arian~ Pop       70     71    9 -5.51  0     6
## 6 7uzmGii~ Mine Bazzi Pop        71     79    4 -3.87  1     7
## 7 2iUXsYO~ Youn~ 5 Sec~ Pop      60     85    7 -5.11  0    46
## 8 2xJCMIJ~ New ~ Dua L~ Pop      76     70    9 -6.02  0     7
## 9 7qiZfU4~ Shap~ Ed Sh~ Pop      83     65    1 -3.18  0     8
## 10 4e4fqjx~ Mean~ Bebe ~ Pop     64     77   10 -6.61  1     8
## 11 0tgVpDi~ Perf~ Ed Sh~ Pop     60     45    8 -6.31  1     2
## 12 63Sevsz~ Neve~ Denni~ Pop     60     69    3 -8.35  1    20
## 13 0d2iYfp~ East~ benny~ Pop     56     68    6 -7.65  0    32
## 14 4eWQlBR~ Neve~ Camil~ Pop     64     71    0 -4.33  1     7
## 15 1zB4vmk~ Thun~ Imagi~ Pop     61     82    0 -4.83  1     4
## 16 4hQ6UGy~ Back~ Selen~ Pop     60     72    6 -4.86  1     5
## 17 4hQ6UGy~ Back~ Selen~ Pop     60     72    6 -4.86  1     5
## # ... with 7 more variables: Acousticness <dbl>, Instrumentalness <dbl>,
## #   Liveness <dbl>, Valence <dbl>, Tempo <dbl>, Duration_secs <dbl>,
## #   Popularity <dbl>

```

```

#6 Identify the dependent & independent variables and use reshaping techniques and create a new data frame
independentVar = Spotify$Popularity
dependentVar = as.data.frame(Spotify %>% select(4:12))
newdf = as.data.frame(cbind(dependentVar, independentVar))
names(newdf)[10] = "Popularity"
newdf

```

```

## # A tibble: 22 x 8
##   id      Genre Danceability Energy Key   dB Mode Speechiness
##   <dbl>    <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl>
## 1 1      Hip-Hop/Rap  75     45    7 -9.211  1     11
## 2 2      Hip-Hop/Rap  74     61    8 -4.880  1     14
## 3 3      Hip-Hop/Rap  59     52    5 -6.090  0     7
## 4 4      Hip-Hop/Rap  75     56    8 -8.011  1     11
## 5 5      Hip-Hop/Rap  84     63    1 -5.833  1     13
## 6 6      Hip-Hop/Rap  68     58   10 -5.843  1     4
## 7 7      Hip-Hop/Rap  82     73    5 -3.998  0     13
## 8 8      Dance/Electronic 79     86    9 -3.240  0     11
## 9 9      Pop          84     54    7 -5.975  1     9
## 10 10     Dance/Electronic 63     88    9 -2.384  0     5
## 11 11     Pop          77     52    2 -4.333  1     3
## 12 12     Hip-Hop/Rap  51     57    6 -7.230  0     20
## 13 13     Hip-Hop/Rap  59     91    8 -6.474  1     7
## 14 14     Pop          85     54    0 -6.825  1     5
## 15 15     Pop          75     66    7 -3.061  1     4
## 16 16     Hip-Hop/Rap  70     63    8 -4.946  1     6
## 17 17     Pop          70     71    9 -5.507  0     6
## 18 18     Reggaeton    59     75    9 -4.736  0     6
## 19 19     Hip-Hop/Rap  92     54    9 -5.723  0     8
## 20 20     Hip-Hop/Rap  92     58   10 -7.495  1    27
## 21 21     Dance/Electronic 65     81    0 -4.057  1     5
## 22 22     Reggaeton    90     68   11 -3.445  0    21

```

## 23	Pop	71	79	4	-3.874	1	7
## 24	Pop	60	85	7	-5.114	0	46
## 25	Pop	76	70	9	-6.021	0	7
## 26	Pop	83	65	1	-3.183	0	8
## 27	R&B	71	65	6	-5.626	1	4
## 28	Pop	64	77	10	-6.610	1	8
## 29	Hip-Hop/Rap	87	39	0	-9.144	0	24
## 30	Pop	60	45	8	-6.312	1	2
## 31	Hip-Hop/Rap	88	56	0	-7.442	1	12
## 32	Dance/Electronic	74	64	11	-4.546	0	4
## 33	R&B	56	54	8	-5.408	0	4
## 34	Pop	60	69	3	-8.354	1	20
## 35	Reggaeton	73	89	0	-3.655	1	4
## 36	Pop	56	68	6	-7.648	0	32
## 37	Pop	64	71	0	-4.333	1	7
## 38	Dance/Electronic	72	80	11	-4.590	0	4
## 39	R&B	67	31	11	-10.068	1	3
## 40	Dance/Electronic	68	75	6	-5.335	1	16
## 41	Hip-Hop/Rap	75	75	8	-5.916	0	52
## 42	Reggaeton	78	84	1	-3.456	1	5
## 43	Hip-Hop/Rap	83	73	8	-3.714	1	22
## 44	Pop	61	82	0	-4.833	1	4
## 45	Reggaeton	78	78	0	-4.449	0	9
## 46	Dance/Electronic	75	83	3	-2.867	0	5
## 47	R&B	70	86	5	-4.877	0	10
## 48	Pop	60	72	6	-4.856	1	5
## 49	Hip-Hop/Rap	66	71	5	-5.665	0	12
## 50	R&B	70	86	5	-4.877	0	10
## 51	Pop	60	72	6	-4.856	1	5
## 52	Hip-Hop/Rap	66	71	5	-5.665	0	12
## 53	Dance/Electronic	75	83	3	-2.867	0	5
## Acousticness Instrumentalness Popularity							
## 1	3	8.29e-05	87				
## 2	26	3.72e-03	88				
## 3	12	6.56e-05	89				
## 4	55	0.00e+00	85				
## 5	6	6.00e-05	84				
## 6	33	0.00e+00	88				
## 7	10	0.00e+00	85				
## 8	4	2.19e-05	86				
## 9	4	0.00e+00	84				
## 10	21	0.00e+00	85				
## 11	18	3.56e-05	86				
## 12	35	0.00e+00	94				
## 13	9	1.09e-04	82				
## 14	57	0.00e+00	81				
## 15	17	0.00e+00	85				
## 16	6	1.94e-04	80				
## 17	4	3.11e-06	84				
## 18	3	1.08e-03	79				
## 19	56	4.04e-03	87				
## 20	0	5.86e-05	81				
## 21	19	0.00e+00	81				
## 22	54	1.28e-05	82				

```

## 23          2    2.77e-06      83
## 24          2    0.00e+00      85
## 25          0    1.56e-05      84
## 26         58    0.00e+00      87
## 27         10    0.00e+00      81
## 28          5    0.00e+00      72
## 29         47    4.13e-06      88
## 30         16    0.00e+00      88
## 31          2    0.00e+00      84
## 32          4    6.66e-05      81
## 33          7    0.00e+00      86
## 34         16    9.78e-06      81
## 35          4    0.00e+00      71
## 36         56    0.00e+00      86
## 37         18    6.37e-04      78
## 38         12    0.00e+00      84
## 39         88    0.00e+00      86
## 40         26    1.10e-05      64
## 41         14    0.00e+00      80
## 42         17    0.00e+00      77
## 43          1    0.00e+00      89
## 44          1    1.34e-01      87
## 45          5    0.00e+00      79
## 46         37    0.00e+00      81
## 47          2    0.00e+00      78
## 48          9    1.68e-06      82
## 49         31    0.00e+00      84
## 50          2    0.00e+00      78
## 51          9    1.68e-06      82
## 52         31    0.00e+00      84
## 53         37    0.00e+00      81

```

```

#7 Remove missing values in your dataset
# ONE ROW IS REMOVED WHERE ID = NULL
spotify_fil <- Spotify %>% filter(!is.na(id))
spotify_fil

```

```

## # A tibble: 52 x 17
##   id      Title Artist  Genre Danceability Energy Key   dB Mode Speechiness
##   <chr>   <chr> <chr> <chr>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 6DCZcSs~ God'~ Drake Hip--     75    45    7 -9.21  1    11
## 2 3ee8Jmj~ SAD! XXXTE~ Hip--     74    61    8 -4.88  1    14
## 3 0e7ipj0~ rock~ Post ~ Hip--    59    52    5 -6.09  0     7
## 4 3swc6WT~ Psyc~ Post ~ Hip--    75    56    8 -8.01  1    11
## 5 2G7V7zs~ In M~ Drake Hip--    84    63    1 -5.83  1    13
## 6 7dt6x5M~ Bett~ Post ~ Hip--    68    58    10 -5.84 1     4
## 7 58q2HKr~ I Li~ Cardi~ Hip--    82    73    5 -4.00  0    13
## 8 7ef4Dls~ One ~ Calvi~ Danc~    79    86    9 -3.24  0    11
## 9 76cy1WJ~ IDGAF Dua L~ Pop     84    54    7 -5.98  1     9
## 10 08bNPGL~ FRIE~ Marsh~ Danc~   63    88    9 -2.38  0     5
## # ... with 42 more rows, and 7 more variables: Acousticness <dbl>,
## #   Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>, Tempo <dbl>,
## #   Duration_secs <dbl>, Popularity <dbl>

```

```
#8 Identify and remove duplicated data in your dataset
# THREE DUPLICATE ROWS ARE REMOVED
spotify_fil <- spotify_fil[!duplicated(spotify_fil),]
spotify_fil
```

```
## # A tibble: 49 x 17
##   id      Title Artist  Genre Danceability Energy  Key    dB Mode Speechiness
##   <chr>   <chr> <chr> <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 6DCZcSs~ God'~ Drake Hip--      75    45    7 -9.21  1    11
## 2 3ee8Jmj~ SAD! XXXTE~ Hip--      74    61    8 -4.88  1    14
## 3 0e7ipj0~ rock~ Post ~ Hip--     59    52    5 -6.09  0     7
## 4 3swc6WT~ Psyc~ Post ~ Hip--     75    56    8 -8.01  1    11
## 5 2G7V7zs~ In M~ Drake Hip--     84    63    1 -5.83  1    13
## 6 7dt6x5M~ Bett~ Post ~ Hip--     68    58    10 -5.84 1     4
## 7 58q2HKr~ I Li~ Cardi~ Hip--    82    73    5 -4.00  0    13
## 8 7ef4Dls~ One ~ Calvi~ Danc~    79    86    9 -3.24  0    11
## 9 76cy1WJ~ IDGAF Dua L~ Pop     84    54    7 -5.98  1     9
## 10 08bNPGL~ FRIE~ Marsh~ Danc~    63    88    9 -2.38  0     5
## # ... with 39 more rows, and 7 more variables: Acousticness <dbl>,
## #   Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>, Tempo <dbl>,
## #   Duration_secs <dbl>, Popularity <dbl>
```

```
#9 Reorder multiple rows in descending order
spotify_fil %>% arrange(desc_Mode, desc_Key, desc_Danceability)
```

```
## # A tibble: 49 x 17
##   id      Title Artist  Genre Danceability Energy  Key    dB Mode Speechiness
##   <chr>   <chr> <chr> <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 7AFASz~ chan~ XXXTE~ R&B      67    31    11 -10.1  1     3
## 2 4qKcDk~ Look~ BlocB~ Hip--     92    58    10 -7.50  1    27
## 3 7dt6x5~ Bett~ Post ~ Hip--     68    58    10 -5.84  1     4
## 4 4e4fqj~ Mean~ Bebe ~ Pop     64    77    10 -6.61  1     8
## 5 2xLMif~ SICK~ Travi~ Hip--    83    73    8 -3.71  1    22
## 6 3swc6W~ Psyc~ Post ~ Hip--     75    56    8 -8.01  1    11
## 7 3ee8Jm~ SAD! XXXTE~ Hip--     74    61    8 -4.88  1    14
## 8 3GCdLU~ All ~ Kendr~ Hip--     70    63    8 -4.95  1     6
## 9 0tgVpD~ Perf~ Ed Sh~ Pop     60    45    8 -6.31  1     2
## 10 3CA9pL~ Nice~ Drake Hip--    59    91    8 -6.47  1     7
## # ... with 39 more rows, and 7 more variables: Acousticness <dbl>,
## #   Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>, Tempo <dbl>,
## #   Duration_secs <dbl>, Popularity <dbl>
```

```
#10 Rename some of the column names in your dataset
spotify_fil %>% rename(song_id = id, song_title = Title, artist_name = Artist)
```

```
## # A tibble: 49 x 17
##   song_id    song_title artist_name  Genre Danceability Energy  Key    dB Mode
##   <chr>      <chr>     <chr>       <chr>     <dbl>  <dbl> <dbl> <dbl> <dbl>
## 1 6DCZcSspj~ God's Plan Drake Hip--      75    45    7 -9.21  1
## 2 3ee8Jmje8~ SAD! XXXTENTACI~ Hip--      74    61    8 -4.88  1
## 3 0e7ipj03S~ rockstar ~ Post Malone Hip--     59    52    5 -6.09  0
## 4 3swc6WTsr~ Psycho (f~ Post Malone Hip--     75    56    8 -8.01  1
```

```

## 5 2G7V7zsVD~ In My Fee~ Drake Hip-- 84 63 1 -5.83 1
## 6 7dt6x5M1j~ Better Now Post Malone Hip-- 68 58 10 -5.84 1
## 7 58q2HKrzh~ I Like It Cardi B Hip-- 82 73 5 -4.00 0
## 8 7ef4Dlsgr~ One Kiss ~ Calvin Har~ Danc~ 79 86 9 -3.24 0
## 9 76cy1WJvN~ IDGAF Dua Lipa Pop 84 54 7 -5.98 1
## 10 08bNPGLD8~ FRIENDS Marshmello Danc~ 63 88 9 -2.38 0
## # ... with 39 more rows, and 8 more variables: Speechiness <dbl>,
## # Acousticness <dbl>, Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>,
## # Tempo <dbl>, Duration_secs <dbl>, Popularity <dbl>

```

#11 Add new variables in your data frame by using a mathematical function
`spotify_fil %>% mutate(db_by_acous = dB/Acousticness)`

```

## # A tibble: 49 x 18
##   id      Title Artist Genre Danceability Energy Key   dB Mode Speechiness
##   <chr>    <chr> <chr> <chr>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 6DCZcSs~ God'~ Drake Hip-- 75 45 7 -9.21 1 11
## 2 3ee8Jmj~ SAD! XXXTE~ Hip-- 74 61 8 -4.88 1 14
## 3 0e7ipj0~ rock~ Post ~ Hip-- 59 52 5 -6.09 0 7
## 4 3swc6WT~ Psyc~ Post ~ Hip-- 75 56 8 -8.01 1 11
## 5 2G7V7zs~ In M~ Drake Hip-- 84 63 1 -5.83 1 13
## 6 7dt6x5M~ Bett~ Post ~ Hip-- 68 58 10 -5.84 1 4
## 7 58q2HKr~ I Li~ Cardi~ Hip-- 82 73 5 -4.00 0 13
## 8 7ef4Dls~ One ~ Calvi~ Danc~ 79 86 9 -3.24 0 11
## 9 76cy1WJ~ IDGAF Dua L~ Pop 84 54 7 -5.98 1 9
## 10 08bNPGL~ FRIE~ Marsh~ Danc~ 63 88 9 -2.38 0 5
## # ... with 39 more rows, and 8 more variables: Acousticness <dbl>,
## #   Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>, Tempo <dbl>,
## #   Duration_secs <dbl>, Popularity <dbl>, db_by_acous <dbl>

```

#12 Create a training set using random number generator engine.

```

set.seed(1234)
training_set <- spotify_fil %>% sample_frac(0.8, replace = FALSE)
training_set

```

```

## # A tibble: 39 x 17
##   id      Title Artist Genre Danceability Energy Key   dB Mode Speechiness
##   <chr>    <chr> <chr> <chr>     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 4e4fqjx~ Mean~ Bebe ~ Pop 64 77 10 -6.61 1 8
## 2 3GCdLUS~ All ~ Kendr~ Hip-- 70 63 8 -4.95 1 6
## 3 3V8UKqh~ Te B~ Nio G~ Regg~ 90 68 11 -3.44 0 21
## 4 4eWQlBR~ Neve~ Camil~ Pop 64 71 0 -4.33 1 7
## 5 1zB4vmk~ Thun~ Imagi~ Pop 61 82 0 -4.83 1 4
## 6 76cy1WJ~ IDGAF Dua L~ Pop 84 54 7 -5.98 1 9
## 7 2G7V7zs~ In M~ Drake Hip-- 84 63 1 -5.83 1 13
## 8 0tBbt8C~ Wolv~ Selen~ Danc~ 72 80 11 -4.59 0 4
## 9 4hQ6UGy~ Back~ Selen~ Pop 60 72 6 -4.86 1 5
## 10 3swc6WT~ Psyc~ Post ~ Hip-- 75 56 8 -8.01 1 11
## # ... with 29 more rows, and 7 more variables: Acousticness <dbl>,
## #   Instrumentalness <dbl>, Liveness <dbl>, Valence <dbl>, Tempo <dbl>,
## #   Duration_secs <dbl>, Popularity <dbl>

```

```
#13 Print the summary statistics of your dataset  
summary(spotify_fil)
```

```
##      id          Title        Artist       Genre  
##  Length:49      Length:49      Length:49      Length:49  
##  Class :character  Class :character  Class :character  Class :character  
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character  
##  
##  
##  
##  Danceability     Energy        Key         dB  
##  Min.   :51.00    Min.   :31.00    Min.   : 0.000  Min.   :-10.068  
##  1st Qu.:64.00   1st Qu.:57.00   1st Qu.: 3.000  1st Qu.: -6.312  
##  Median :72.00   Median :69.00   Median : 7.000  Median : -5.335  
##  Mean   :71.92   Mean   :67.59   Mean   : 5.816  Mean   : -5.453  
##  3rd Qu.:78.00   3rd Qu.:78.00   3rd Qu.: 9.000  3rd Qu.: -4.333  
##  Max.   :92.00   Max.   :91.00   Max.   :11.000  Max.   : -2.384  
##  
##      Mode        Speechiness  Acousticness Instrumentalness  
##  Min.   :0.0000  Min.   : 2.00  Min.   : 0.00  Min.   :0.0000000  
##  1st Qu.:0.0000 1st Qu.: 5.00  1st Qu.: 4.00  1st Qu.:0.0000000  
##  Median :1.0000  Median : 7.00  Median :12.00  Median :0.0000000  
##  Mean   :0.551   Mean   :11.27  Mean   :19.22  Mean   :0.0029435  
##  3rd Qu.:1.0000  3rd Qu.:13.00 3rd Qu.:26.00  3rd Qu.:0.0000356  
##  Max.   :1.0000  Max.   :52.00  Max.   :88.00  Max.   :0.1340000  
##  
##      Liveness      Valence        Tempo      Duration_secs  
##  Min.   : 2.00  Min.   : 8.00  Min.   : 65.0  Min.   :119.0  
##  1st Qu.:11.00  1st Qu.:34.00  1st Qu.: 95.0  1st Qu.:185.0  
##  Median :13.00  Median :46.00  Median :107.0  Median :211.0  
##  Mean   :18.67  Mean   :48.08  Mean   :115.8  Mean   :208.6  
##  3rd Qu.:20.00  3rd Qu.:61.00  3rd Qu.:136.0  3rd Qu.:223.0  
##  Max.   :64.00  Max.   :93.00  Max.   :180.0  Max.   :418.0  
##  
##      Popularity  
##  Min.   :64.00  
##  1st Qu.:81.00  
##  Median :84.00  
##  Mean   :83.04  
##  3rd Qu.:86.00  
##  Max.   :94.00
```

```
#14 Use any of the numerical variables from the dataset and perform the following statistical functions  
#Mean  
mean(spotify_fil$Popularity)
```

```
## [1] 83.04082
```

```
#Median  
median(spotify_fil$Tempo)
```

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

```
#Mode
find_mode <- function(x) {
  u <- unique(x)
  tab <- tabulate(match(x, u))
  u[tab == max(tab)]
}
find_mode(spotify_fil$key)
```

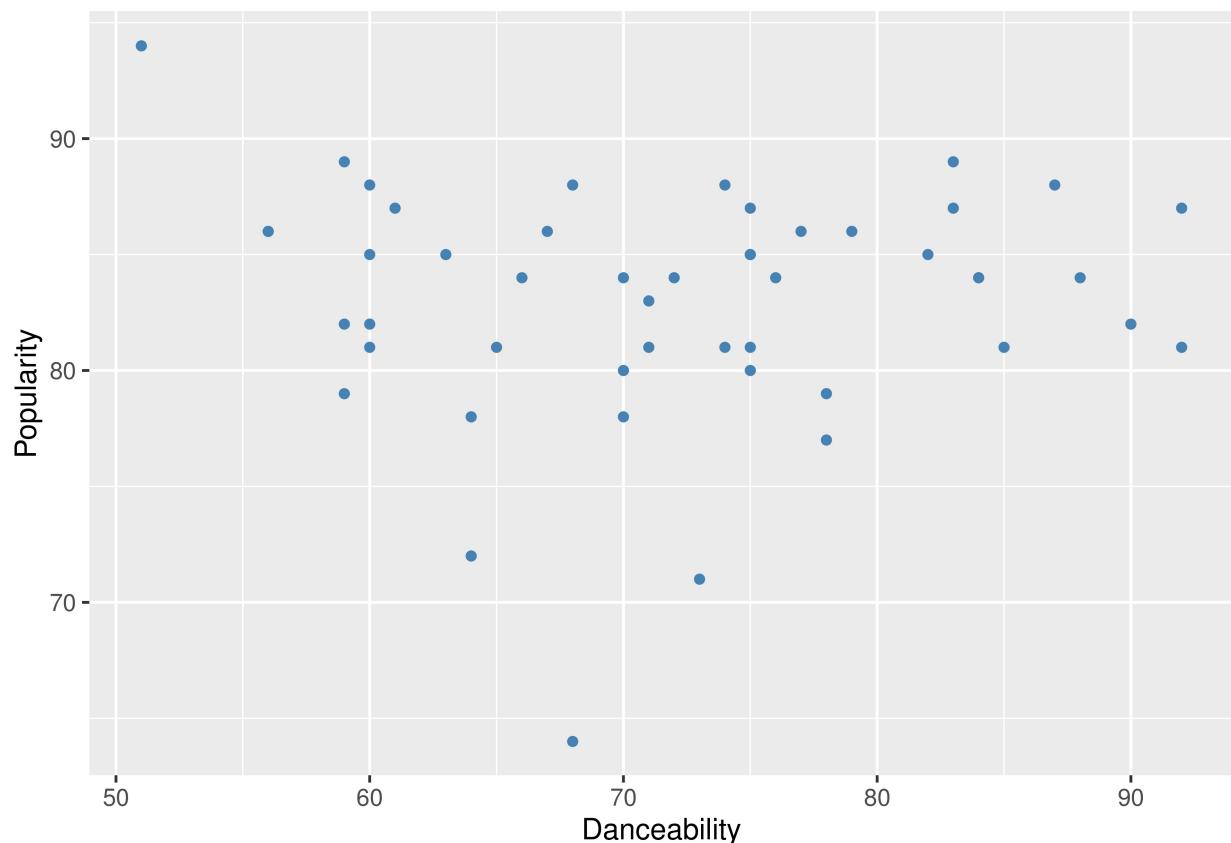
```
## [1] 8 0
```

```
#Range
range(spotify_fil$Energy)
```

```
## [1] 31 91
```

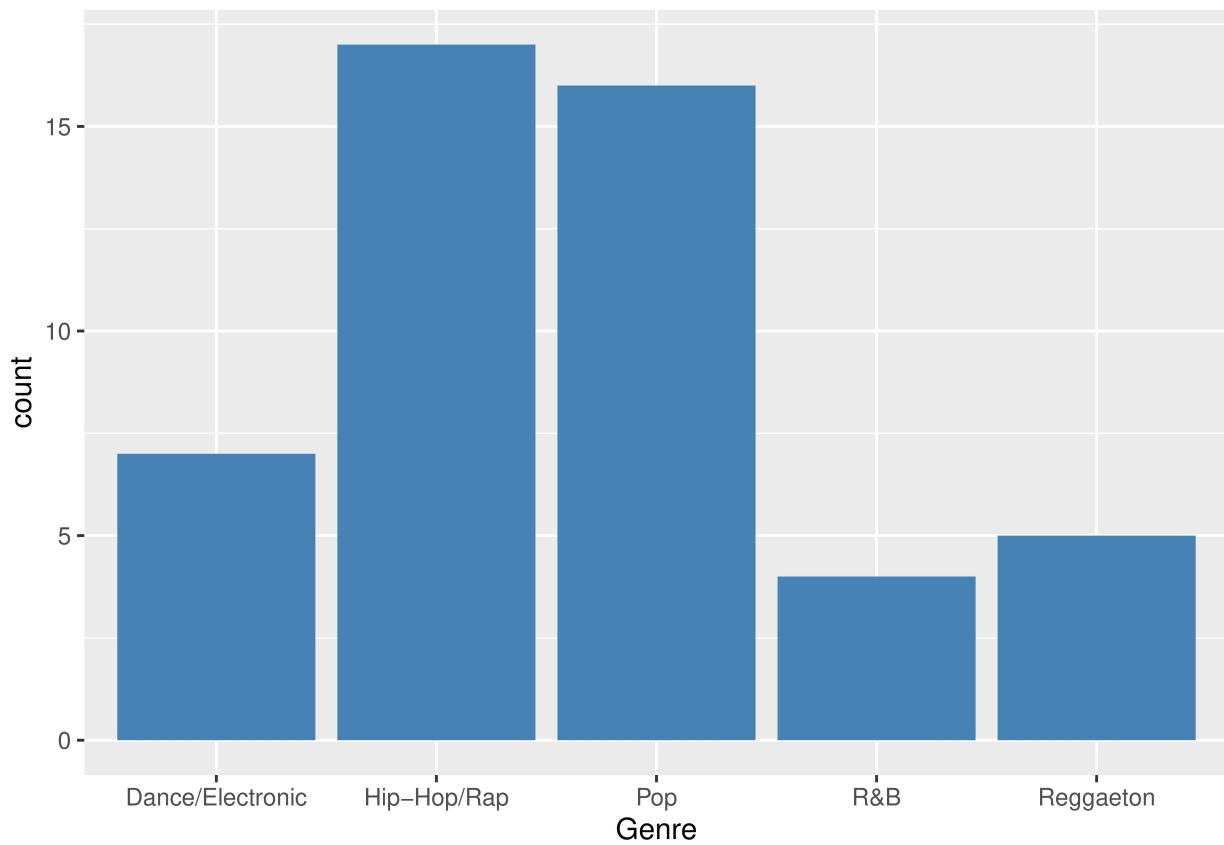
#15 Plot a scatter plot for any 2 variables in your dataset

```
ggplot(spotify_fil , aes(x=Danceability , y=Popularity)) +
  geom_point(size=1.3, color ="steelblue", shape=19)
```

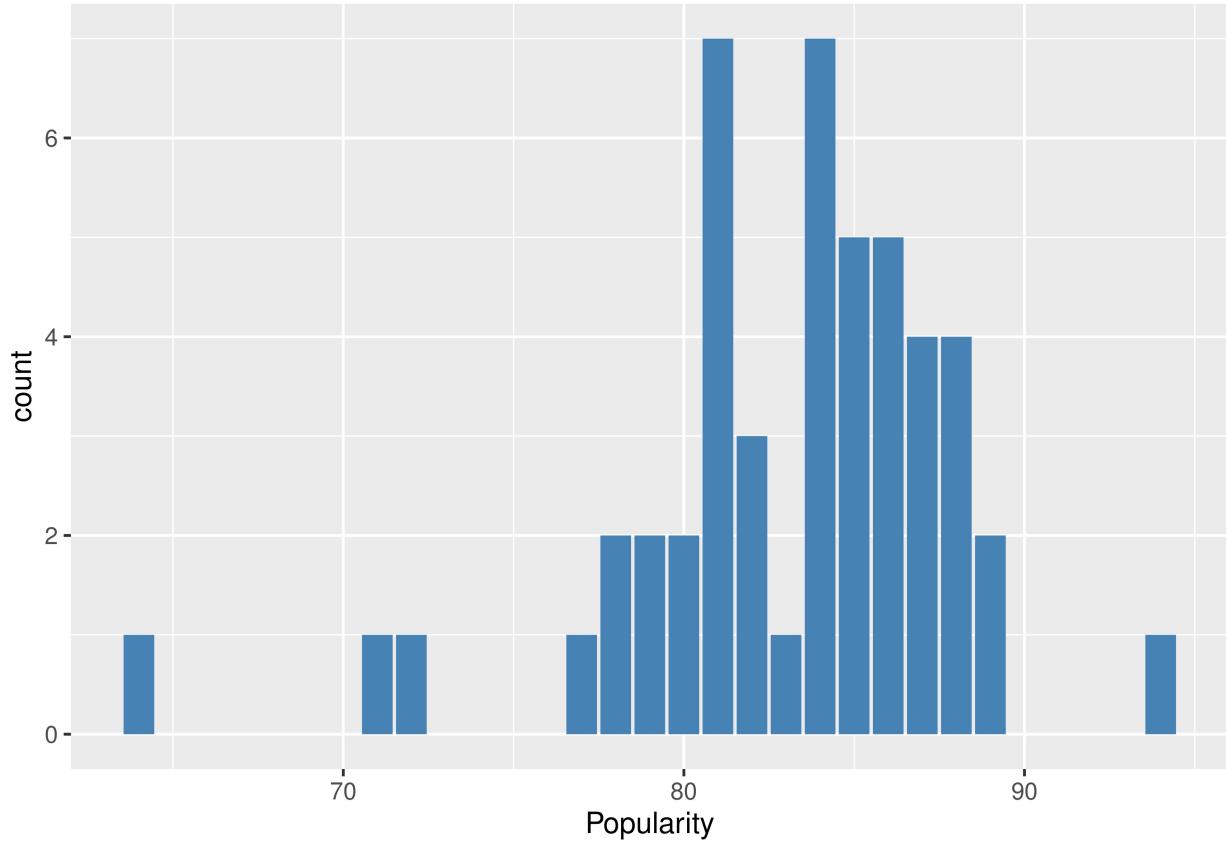


#16 Plot a bar plot for any 2 variables in your dataset

```
ggplot(data = spotify_fil, aes (x = Genre))+geom_bar(fill = "steelblue")
```



```
ggplot(data = spotify_fil, aes (x = Popularity))+geom_bar(fill = "steelblue")
```



```
#17 Find the correlation between any 2 variables by applying least square linear regression model
y = unlist(spotify_fil[,"Popularity"])
x = unlist(spotify_fil[,"Danceability"])
head(x)

## Danceability1 Danceability2 Danceability3 Danceability4 Danceability5
##          75           74           59           75           84
## Danceability6
##          68

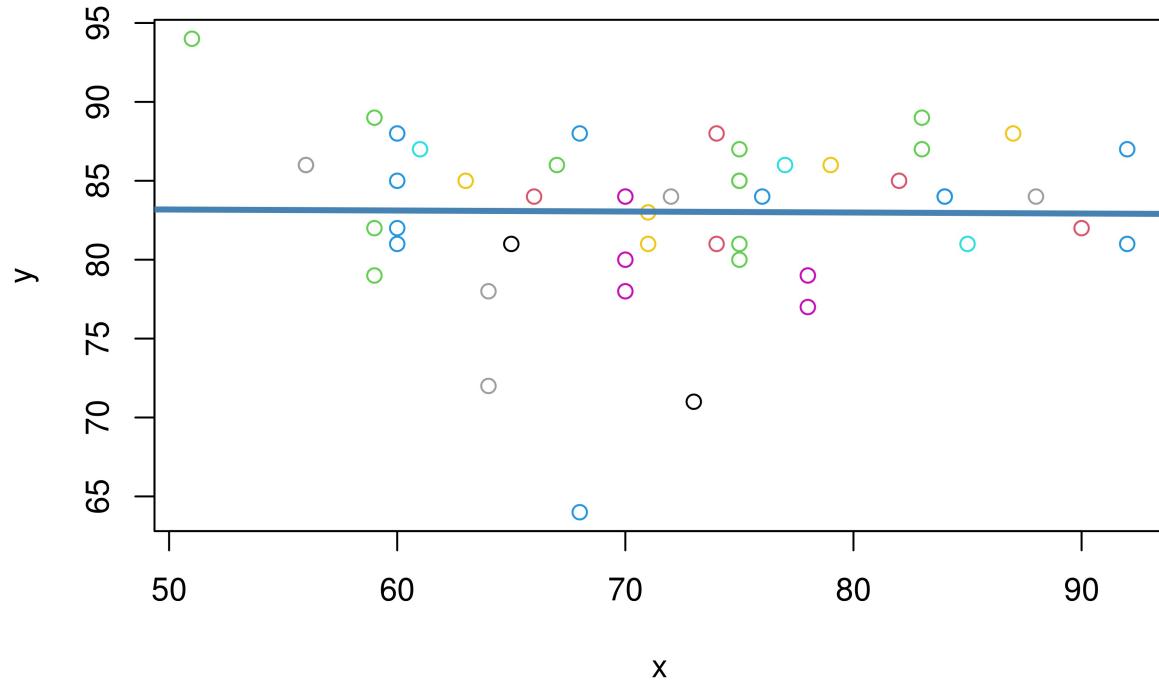
head(y)

## Popularity1 Popularity2 Popularity3 Popularity4 Popularity5 Popularity6
##          87           88           89           85           84           88

xycorr <-cor(x,y, method = "pearson")
xycorr

## [1] -0.01310046

plot(y~x, col = x)
model1 = lm(y~x)
abline(model1,col = "steelblue", lwd = 3)
```



```
# Github link : https://github.com/abhisheksalwan/Rcode2\_Team3
```