Introduction

This project demonstrates data preprocessing, cleaning, and exploratory data analysis (EDA) in R. The dataset consists of music-related features including popularity, tempo, and genres.

In [49]:

```
## Load Libraries and Dataset
install.packages('tidyverse')
install.packages('ggplot2')
install.packages('ggcorrplot')

library(tidyverse)
library(ggplot2)
library(ggcorrplot)

df <- read.csv('spotify.csv')

#First 5 observations
head(df,5)
```

```
The downloaded binary packages are in
        /var/folders/17/y6yqqy7n54j29b_bxf1w8h880000gn/T//Rtmp8nCAzj/downloa
ded_packages

The downloaded binary packages are in
        /var/folders/17/y6yqqy7n54j29b_bxf1w8h880000gn/T//Rtmp8nCAzj/downloa
ded_packages

The downloaded binary packages are in
        /var/folders/17/y6yqqy7n54j29b_bxf1w8h880000gn/T//Rtmp8nCAzj/downloa
ded_packages
```

	Track.ID	Track.Name	Album.Name	Artist.Name.s.	Release.Date
	<chr>	<chr>	<chr>	<chr>	<chr>
1	5Eg4TsPcqNbljd8ADMZosg	Chains	Nick Jonas X2	Nick Jonas	2015-11-20
2	3V3iy4K6paycRmTyrjQVwi	Secrets	Heart On My Sleeve (Deluxe)	Mary Lambert	2014-10-14
3	2f0GI2ZLUtbGqFx8t2Gk6A	I Know Places (Taylor's Version)	1989 (Taylor's Version)	Taylor Swift	2023-10-26
4	2Od3Imc5PJeZhRPeIhpgN6	Delta 1406	31 Minutes to Takeoff	Mike Posner	2010-08-09
5	5hw1uOFZK3odNqXa4sF2JZ	Stay With Me - Re-record	In The Lonely Hour (10th Anniversary Edition / Deluxe)	Sam Smith	2024-08-30

View structure and summary

In [50]:

```
str(df)
```

```
'data.frame':	265 obs. of  23 variables:
 $ Track.ID        : chr  "5Eg4TsPcqNbIjd8ADMZosg" "3V3iy4K6paycRmTyrjQVwi"
"2f0GI2ZLUtbGqFx8t2Gk6A" "2Od3Imc5PJeZhRPeIhpgN6" ...
 $ Track.Name      : chr  "Chains" "Secrets" "I Know Places (Taylor's Versio
n)" "Delta 1406" ...
 $ Album.Name      : chr  "Nick Jonas X2" "Heart On My Sleeve (Deluxe)" "198
9 (Taylor's Version)" "31 Minutes to Takeoff" ...
 $ Artist.Name.s.  : chr  "Nick Jonas" "Mary Lambert" "Taylor Swift" "Mike P
osner" ...
 $ Release.Date    : chr  "2015-11-20" "2014-10-14" "2023-10-26" "2010-08-0
9" ...
 $ Duration..ms.   : int  203106 223405 195700 184546 172760 226293 227360 2
14720 252733 216626 ...
 $ Popularity      : int  55 39 63 19 38 34 55 31 0 63 ...
 $ Added.By        : chr  "95fvnclitpzdbgd83xcozdsk2" "95fvnclitpzdbgd83xcoz
dsk2" "95fvnclitpzdbgd83xcozdsk2" "95fvnclitpzdbgd83xcozdsk2" ...
 $ Added.At        : chr  "2024-11-29T00:19:04Z" "2024-11-29T00:19:04Z" "202
4-11-29T00:19:04Z" "2024-11-29T00:19:04Z" ...
 $ Genres          : chr  "dance pop,pop" "neo mellow" "pop" "dance pop,pop,
pop dance,pop rap" ...
 $ Record.Label    : chr  "Safehouse Records / Island Records" "Capitol Reco
rds (CAP)" "Taylor Swift" "J Records" ...
 $ Danceability    : num  0.591 0.789 0.572 0.688 0.515 0.481 0.616 0.493 0.
479 0.495 ...
 $ Energy          : num  0.611 0.555 0.807 0.616 0.41 0.524 0.789 0.879 0.5
45 0.502 ...
 $ Key             : int  0 0 0 1 0 7 7 6 7 5 ...
 $ Loudness        : num  -5.88 -5.9 -5.35 -7.33 -7.12 ...
 $ Mode            : int  0 1 1 1 1 1 0 1 1 1 ...
 $ Speechiness     : num  0.0454 0.041 0.0574 0.0451 0.0411 0.0302 0.0377 0.
0326 0.0688 0.0259 ...
 $ Acousticness    : num  0.0153 0.026 0.0846 0.0138 0.555 0.0315 0.053 0.00
852 0.365 0.0127 ...
 $ Instrumentalness: num  0.00 1.35e-04 0.00 4.61e-04 4.08e-05 1.08e-06 0.00
0.00 0.00 6.89e-01 ...
 $ Liveness        : num  0.0757 0.215 0.071 0.121 0.103 0.235 0.142 0.253
0.0963 0.068 ...
 $ Valence         : num  0.12 0.713 0.626 0.205 0.246 0.162 0.621 0.63 0.24
4 0.0394 ...
 $ Tempo           : num  76 93.2 160 87 84.8 ...
 $ Time.Signature  : int  4 4 4 4 4 4 4 4 4 3 ...
```

In [51]:

```
glimpse(df)
```

```
Rows: 265
Columns: 23
$ Track.ID         <chr> "5Eg4TsPcqNbIjd8ADMZosg", "3V3iy4K6paycRmTyrjQVwi",
"…
$ Track.Name       <chr> "Chains", "Secrets", "I Know Places (Taylor's Versi
on…
$ Album.Name       <chr> "Nick Jonas X2", "Heart On My Sleeve (Deluxe)", "19
89…
$ Artist.Name.s.   <chr> "Nick Jonas", "Mary Lambert", "Taylor Swift", "Mike
P…
$ Release.Date     <chr> "2015-11-20", "2014-10-14", "2023-10-26", "2010-08-
09…
$ Duration..ms.    <int> 203106, 223405, 195700, 184546, 172760, 226293, 227
36…
$ Popularity       <int> 55, 39, 63, 19, 38, 34, 55, 31, 0, 63, 40, 45, 69,
23…
$ Added.By         <chr> "95fvnclitpzdbgd83xcozdsk2", "95fvnclitpzdbgd83xcoz
ds…
$ Added.At         <chr> "2024-11-29T00:19:04Z", "2024-11-29T00:19:04Z", "20
24…
$ Genres           <chr> "dance pop,pop", "neo mellow", "pop", "dance pop,po
p,…
$ Record.Label     <chr> "Safehouse Records / Island Records", "Capitol Reco
rd…
$ Danceability     <dbl> 0.591, 0.789, 0.572, 0.688, 0.515, 0.481, 0.616, 0.
49…
$ Energy           <dbl> 0.611, 0.555, 0.807, 0.616, 0.410, 0.524, 0.789, 0.
87…
$ Key              <int> 0, 0, 0, 1, 0, 7, 7, 6, 7, 5, 5, 2, 4, 11, 5, 6, 4,
2…
$ Loudness         <dbl> -5.884, -5.900, -5.348, -7.334, -7.121, -7.035, -4.
87…
$ Mode             <int> 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1,
1,…
$ Speechiness      <dbl> 0.0454, 0.0410, 0.0574, 0.0451, 0.0411, 0.0302, 0.0
37…
$ Acousticness     <dbl> 0.01530, 0.02600, 0.08460, 0.01380, 0.55500, 0.0315
0,…
$ Instrumentalness <dbl> 0.00e+00, 1.35e-04, 0.00e+00, 4.61e-04, 4.08e-05,
1.0…
$ Liveness         <dbl> 0.0757, 0.2150, 0.0710, 0.1210, 0.1030, 0.2350, 0.1
42…
$ Valence          <dbl> 0.1200, 0.7130, 0.6260, 0.2050, 0.2460, 0.1620, 0.6
21…
$ Tempo            <dbl> 76.003, 93.229, 160.015, 87.043, 84.837, 76.082, 8
3.0…
$ Time.Signature   <int> 4, 4, 4, 4, 4, 4, 4, 4, 4, 3, 4, 4, 4, 4, 4, 5, 4,
4,…
```

In [52]:

```
summary(df)
```

```
  Track.ID          Track.Name          Album.Name          Artist.Name.s.    
Length:265         Length:265         Length:265         Length:265        
Class :character   Class :character   Class :character   Class :character  
Mode  :character   Mode  :character   Mode  :character   Mode  :character  
                                                                           
                                                                           
                                                                           
Release.Date       Duration..ms.      Popularity      Added.By        
Length:265         Min.   :127931   Min.   : 0.00   Length:265        
Class :character   1st Qu.:195320   1st Qu.:15.00   Class :character  
Mode  :character   Median :218040   Median :31.00   Mode  :character  
                   Mean   :219548   Mean   :31.71                     
                   3rd Qu.:238266   3rd Qu.:47.00                     
                   Max.   :485333   Max.   :83.00                     
  Added.At            Genres          Record.Label        Danceability   
Length:265         Length:265         Length:265         Min.   :0.1740  
Class :character   Class :character   Class :character   1st Qu.:0.4420  
Mode  :character   Mode  :character   Mode  :character   Median :0.5360  
                                                         Mean   :0.5338  
                                                         3rd Qu.:0.6200  
                                                         Max.   :0.8900  
    Energy           Key            Loudness            Mode       
Min.   :0.152   Min.   : 0.000   Min.   :-16.550   Min.   :0.0000  
1st Qu.:0.524   1st Qu.: 2.000   1st Qu.: -7.937   1st Qu.:0.0000  
Median :0.637   Median : 6.000   Median : -6.353   Median :1.0000  
Mean   :0.622   Mean   : 5.509   Mean   : -6.841   Mean   :0.6302  
3rd Qu.:0.766   3rd Qu.: 8.000   3rd Qu.: -5.254   3rd Qu.:1.0000  
Max.   :0.970   Max.   :11.000   Max.   : -1.395   Max.   :1.0000  
 Speechiness       Acousticness       Instrumentalness       Liveness     
Min.   :0.02430   Min.   :0.0000133   Min.   :0.0000000   Min.   :0.0304  
1st Qu.:0.03170   1st Qu.:0.0198000   1st Qu.:0.0000000   1st Qu.:0.0973  
Median :0.03890   Median :0.1210000   Median :0.0000108   Median :0.1180  
Mean   :0.04963   Mean   :0.2599463   Mean   :0.0306197   Mean   :0.1611  
3rd Qu.:0.05230   3rd Qu.:0.4430000   3rd Qu.:0.0005750   3rd Qu.:0.1900  
Max.   :0.28400   Max.   :0.9350000   Max.   :0.9420000   Max.   :0.6920  
   Valence           Tempo        Time.Signature 
Min.   :0.0370   Min.   : 65.53   Min.   :1.000  
1st Qu.:0.2300   1st Qu.: 88.00   1st Qu.:4.000  
Median :0.3840   Median :100.10   Median :4.000  
Mean   :0.3954   Mean   :117.96   Mean   :3.947  
3rd Qu.:0.5210   3rd Qu.:151.98   3rd Qu.:4.000  
Max.   :0.9750   Max.   :202.00   Max.   :5.000  
```

Check for missing values

In [53]:

```
# Count missing (NA) values for all columns
colSums(is.na(df))
```

```
Track.ID: 0 Track.Name: 0 Album.Name: 0 Artist.Name.s.: 0 Release.Date: 0
 Duration..ms.: 0 Popularity: 0 Added.By: 0 Added.At: 0 Genres: 0 Record.Label: 0
 Danceability: 0 Energy: 0 Key: 0 Loudness: 0 Mode: 0 Speechiness: 0 Acousticness:
0 Instrumentalness: 0 Liveness: 0 Valence: 0 Tempo: 0 Time.Signature: 0
```

In [54]:

```
# Count blank strings for all columns
colSums(df == "")
```

```
Track.ID: 0 Track.Name: 0 Album.Name: 0 Artist.Name.s.: 0 Release.Date: 0
 Duration..ms.: 0 Popularity: 0 Added.By: 0 Added.At: 0 Genres: 24 Record.Label: 0
 Danceability: 0 Energy: 0 Key: 0 Loudness: 0 Mode: 0 Speechiness: 0 Acousticness:
0 Instrumentalness: 0 Liveness: 0 Valence: 0 Tempo: 0 Time.Signature: 0
```

In [55]:

```
# Check for "NULL" or "unknown" values
colSums(df == "NULL" | df == "unknown" | df == "Unknown")
```

```
Track.ID: 0 Track.Name: 0 Album.Name: 0 Artist.Name.s.: 0 Release.Date: 0
 Duration..ms.: 0 Popularity: 0 Added.By: 0 Added.At: 0 Genres: 0 Record.Label: 0
 Danceability: 0 Energy: 0 Key: 0 Loudness: 0 Mode: 0 Speechiness: 0 Acousticness:
0 Instrumentalness: 0 Liveness: 0 Valence: 0 Tempo: 0 Time.Signature: 0
```

In [61]:

```
# Identify rows with missing, blank, or placeholder values
# Count blank strings for all columns
colSums(df == " ")
```

```
Track.ID: 0 Track.Name: 0 Album.Name: 0 Artist.Name.s.: 0 Release.Date: 0
 Duration..ms.: 0 Popularity: 0 Added.By: 0 Added.At: 0 Genres: 0 Record.Label: 0
 Danceability: 0 Energy: 0 Key: 0 Loudness: 0 Mode: 0 Speechiness: 0 Acousticness:
0 Instrumentalness: 0 Liveness: 0 Valence: 0 Tempo: 0 Time.Signature: 0
```

In [14]:

```
# Quick Summary of the Variables
summary(df)
```

```
  Track.ID          Track.Name         Album.Name         Artist.Name.s.    
Length:265         Length:265         Length:265         Length:265        
Class :character   Class :character   Class :character   Class :character  
Mode  :character   Mode  :character   Mode  :character   Mode  :character  
                                                                           
                                                                           
                                                                           
Release.Date       Duration..ms.      Popularity       Added.By          
Length:265         Min.   :127931   Min.   : 0.00   Length:265        
Class :character   1st Qu.:195320   1st Qu.:15.00   Class :character  
Mode  :character   Median :218040   Median :31.00   Mode  :character  
                   Mean   :219548   Mean   :31.71                     
                   3rd Qu.:238266   3rd Qu.:47.00                     
                   Max.   :485333   Max.   :83.00                     
  Added.At            Genres          Record.Label        Danceability   
Length:265         Length:265         Length:265         Min.   :0.1740  
Class :character   Class :character   Class :character   1st Qu.:0.4420  
Mode  :character   Mode  :character   Mode  :character   Median :0.5360  
                                                         Mean   :0.5338  
                                                         3rd Qu.:0.6200  
                                                         Max.   :0.8900  
    Energy            Key             Loudness            Mode       
Min.   :0.152   Min.   : 0.000   Min.   :-16.550   Min.   :0.0000  
1st Qu.:0.524   1st Qu.: 2.000   1st Qu.: -7.937   1st Qu.:0.0000  
Median :0.637   Median : 6.000   Median : -6.353   Median :1.0000  
Mean   :0.622   Mean   : 5.509   Mean   : -6.841   Mean   :0.6302  
3rd Qu.:0.766   3rd Qu.: 8.000   3rd Qu.: -5.254   3rd Qu.:1.0000  
Max.   :0.970   Max.   :11.000   Max.   : -1.395   Max.   :1.0000  
 Speechiness       Acousticness       Instrumentalness       Liveness     
Min.   :0.02430   Min.   :0.0000133   Min.   :0.0000000   Min.   :0.0304  
1st Qu.:0.03170   1st Qu.:0.0198000   1st Qu.:0.0000000   1st Qu.:0.0973  
Median :0.03890   Median :0.1210000   Median :0.0000108   Median :0.1180  
Mean   :0.04963   Mean   :0.2599463   Mean   :0.0306197   Mean   :0.1611  
3rd Qu.:0.05230   3rd Qu.:0.4430000   3rd Qu.:0.0005750   3rd Qu.:0.1900  
Max.   :0.28400   Max.   :0.9350000   Max.   :0.9420000   Max.   :0.6920  
   Valence            Tempo        Time.Signature 
Min.   :0.0370   Min.   : 65.53   Min.   :1.000  
1st Qu.:0.2300   1st Qu.: 88.00   1st Qu.:4.000  
Median :0.3840   Median :100.10   Median :4.000  
Mean   :0.3954   Mean   :117.96   Mean   :3.947  
3rd Qu.:0.5210   3rd Qu.:151.98   3rd Qu.:4.000  
Max.   :0.9750   Max.   :202.00   Max.   :5.000  
```

Convert Date Columns

In [62]:

```r
# Convert date columns to appropriate format
df$Release.Date <- as.Date(df$Release.Date, format = "%Y-%m-%d")
df$Added.At <- as.POSIXct(df$Added.At, format = "%Y-%m-%dT%H:%M:%SZ")
```

Exploratory Data Analysis

In [65]:

```
# Numerical summary
summary(select_if(df, is.numeric))
```

```
Duration..ms.      Popularity     Danceability        Energy
Min.   :127931   Min.   : 0.00   Min.   :0.1740   Min.   :0.152
1st Qu.:195320   1st Qu.:15.00   1st Qu.:0.4420   1st Qu.:0.524
Median :218040   Median :31.00   Median :0.5360   Median :0.637
Mean   :219548   Mean   :31.71   Mean   :0.5338   Mean   :0.622
3rd Qu.:238266   3rd Qu.:47.00   3rd Qu.:0.6200   3rd Qu.:0.766
Max.   :485333   Max.   :83.00   Max.   :0.8900   Max.   :0.970
     Key             Loudness            Mode          Speechiness
Min.   : 0.000   Min.   :-16.550   Min.   :0.0000   Min.   :0.02430
1st Qu.: 2.000   1st Qu.: -7.937   1st Qu.:0.0000   1st Qu.:0.03170
Median : 6.000   Median : -6.353   Median :1.0000   Median :0.03890
Mean   : 5.509   Mean   : -6.841   Mean   :0.6302   Mean   :0.04963
3rd Qu.: 8.000   3rd Qu.: -5.254   3rd Qu.:1.0000   3rd Qu.:0.05230
Max.   :11.000   Max.   : -1.395   Max.   :1.0000   Max.   :0.28400
 Acousticness       Instrumentalness       Liveness         Valence
Min.   :0.0000133   Min.   :0.0000000   Min.   :0.0304   Min.   :0.0370
1st Qu.:0.0198000   1st Qu.:0.0000000   1st Qu.:0.0973   1st Qu.:0.2300
Median :0.1210000   Median :0.0000108   Median :0.1180   Median :0.3840
Mean   :0.2599463   Mean   :0.0306197   Mean   :0.1611   Mean   :0.3954
3rd Qu.:0.4430000   3rd Qu.:0.0005750   3rd Qu.:0.1900   3rd Qu.:0.5210
Max.   :0.9350000   Max.   :0.9420000   Max.   :0.6920   Max.   :0.9750
    Tempo        Time.Signature
Min.   : 65.53   Min.   :1.000
1st Qu.: 88.00   1st Qu.:4.000
Median :100.10   Median :4.000
Mean   :117.96   Mean   :3.947
3rd Qu.:151.98   3rd Qu.:4.000
Max.   :202.00   Max.   :5.000
```

Visualize Distributions

Visualize Distributions

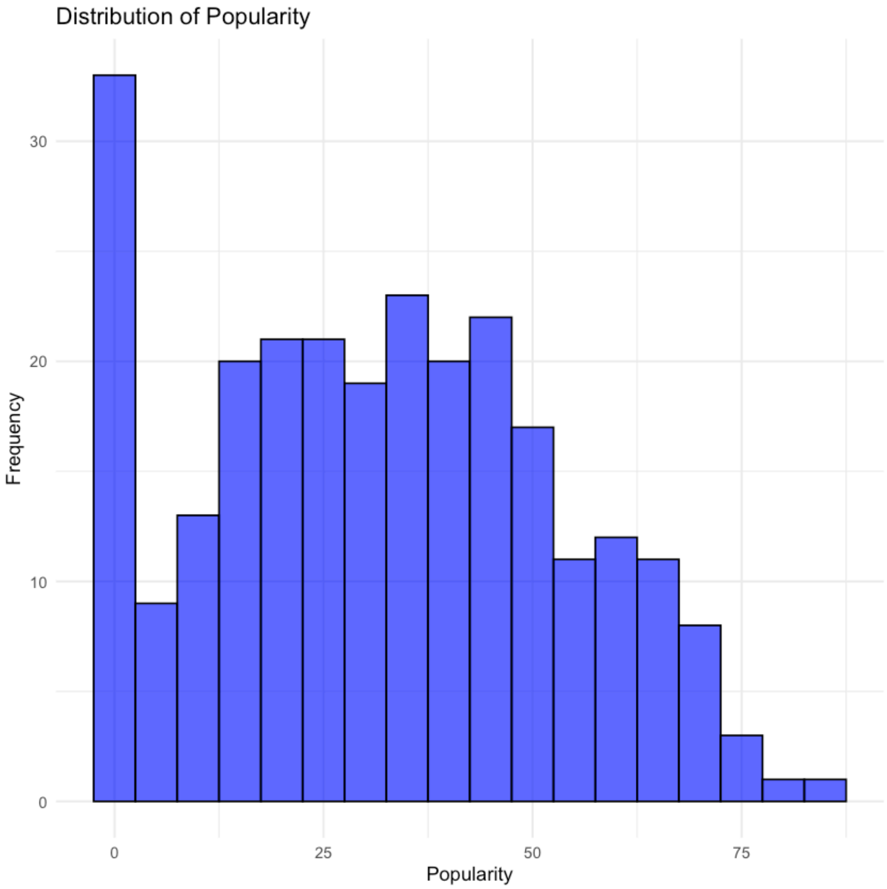
In [ ]:

```
```

Popularity

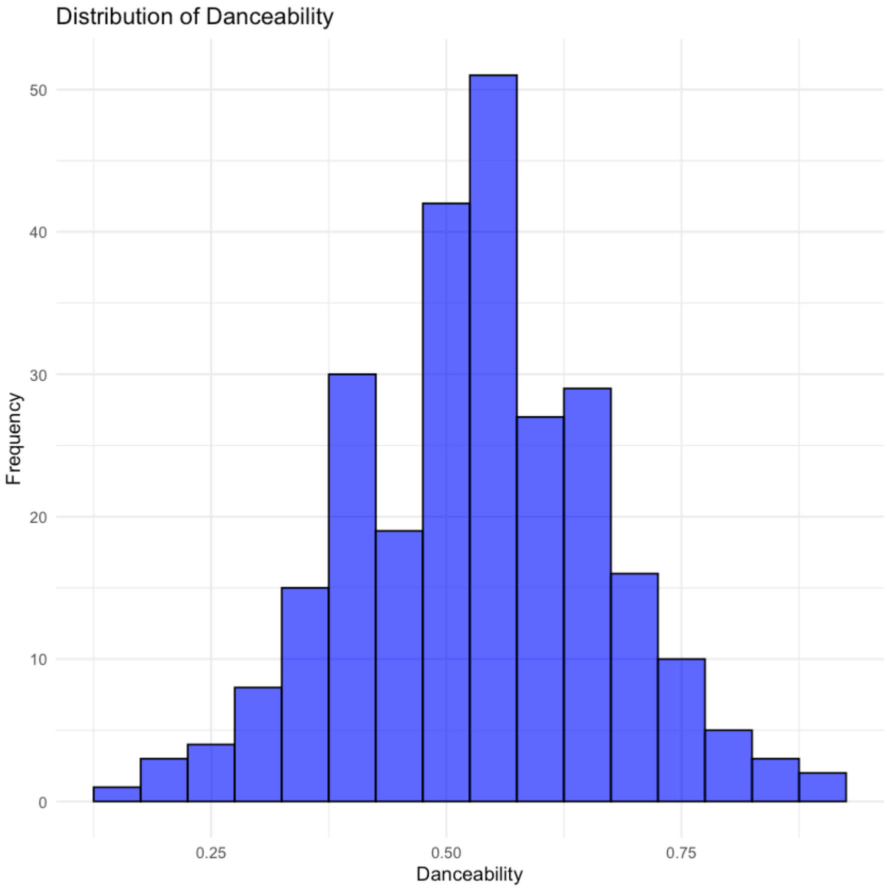
In [67]:

```
ggplot(df, aes(x = Popularity)) +
  geom_histogram(binwidth = 5, fill = "blue", color = "black", alpha = 0.7)
  theme_minimal() +
  labs(title = "Distribution of Popularity", x = "Popularity", y = "Frequenc
```



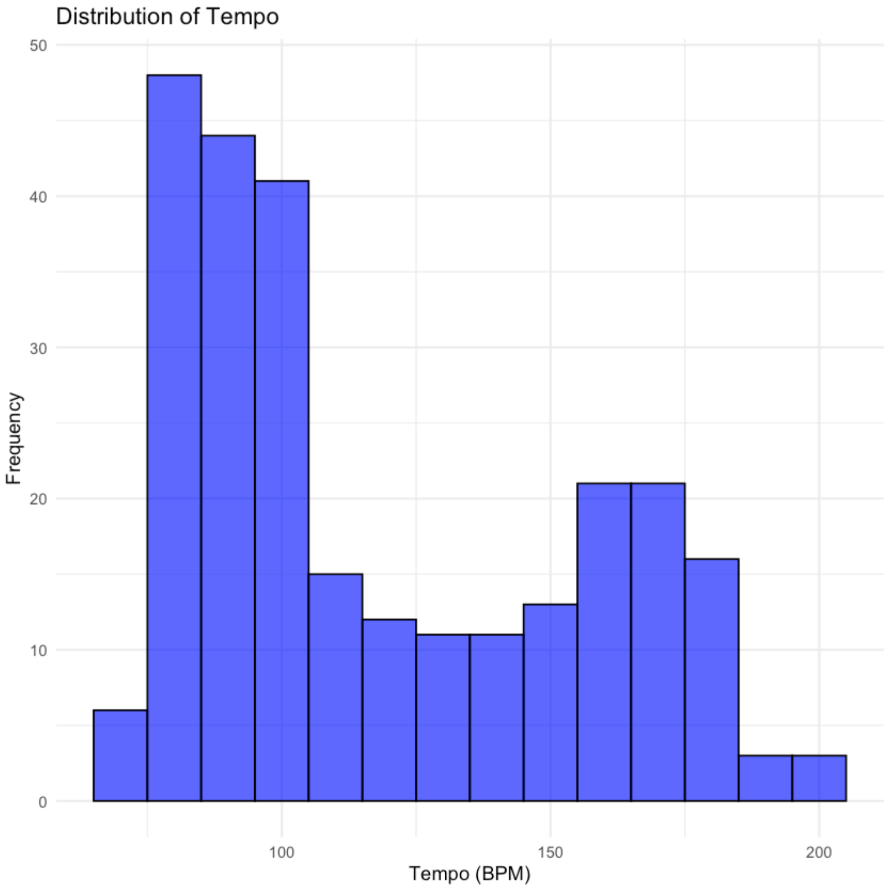
Danceability

```
In [70]: ggplot(df, aes(x = Danceability)) +
           geom_histogram(binwidth = 0.05, fill = "blue", color = "black", alpha = 0.
           theme_minimal() +
           labs(title = "Distribution of Danceability", x = "Danceability", y = "Freq
```



Tempo

```
In [72]: ggplot(df, aes(x = Tempo)) +
           geom_histogram(binwidth = 10, fill = "blue", color = "black", alpha = 0.7)
           theme_minimal() +
           labs(title = "Distribution of Tempo", x = "Tempo (BPM)", y = "Frequency")
```



Correlation Analysis

```
In [73]: # Correlation matrix for numeric features
         numeric_data <- select_if(df, is.numeric)
         correlation_matrix <- cor(numeric_data, use = "complete.obs")
```

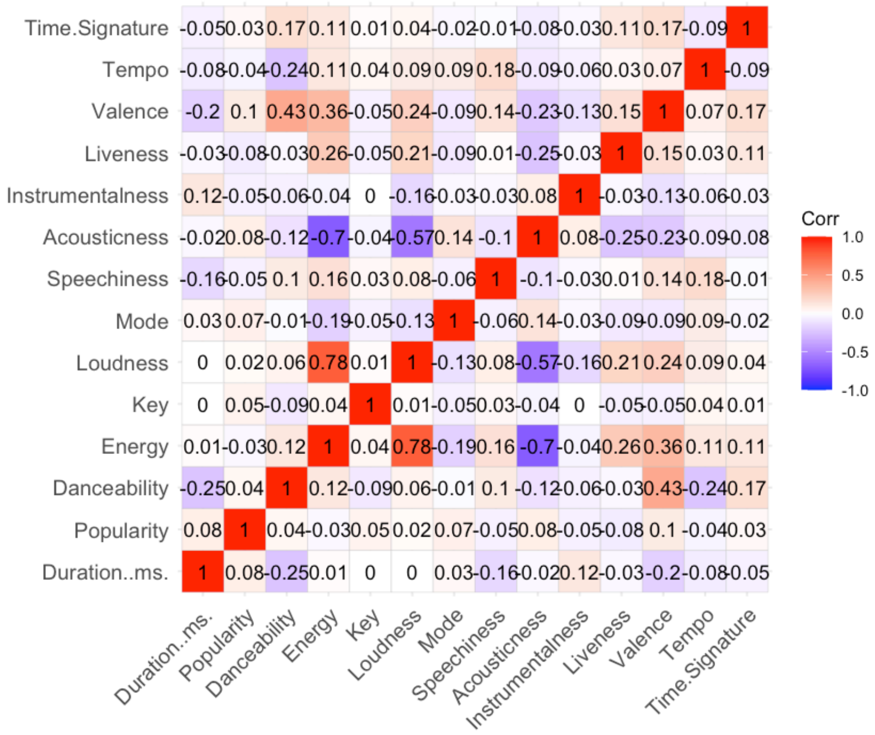
```
In [74]: # Print correlation matrix
         print(correlation_matrix)
```

```
                 Duration..ms.  Popularity Danceability       Energy
Duration..ms.      1.000000000  0.08050723 -0.253882867  0.005927012
Popularity         0.080507227  1.00000000  0.040839532 -0.032268182
Danceability      -0.253882867  0.04083953  1.000000000  0.121270140
Energy             0.005927012 -0.03226818  0.121270140  1.000000000
Key                0.001259947  0.05380896 -0.089122512  0.039061708
Loudness           0.002766892  0.02444370  0.055147860  0.780575951
Mode               0.027341673  0.06677056 -0.007829257 -0.187955821
Speechiness       -0.161671330 -0.05454191  0.099708678  0.156430226
Acousticness      -0.017681929  0.08249996 -0.120688472 -0.704630902
Instrumentalness   0.122450542 -0.05032372 -0.056138897 -0.036780412
Liveness          -0.034194713 -0.08109104 -0.030774750  0.255813020
Valence           -0.197133202  0.09659480  0.433282084  0.364293439
Tempo             -0.084327725 -0.04298448 -0.238930499  0.105538418
Time.Signature    -0.051815173  0.02690197  0.172973955  0.106315210
                          Key    Loudness         Mode  Speechiness
Duration..ms.     0.001259947 0.002766892  0.027341673 -0.161671330
```

```
Popularity        0.053808964  0.024443703  0.066770561 -0.054541907
Danceability     -0.089122512  0.055147860 -0.007829257  0.099708678
Energy            0.039061708  0.780575951 -0.187955821  0.156430226
Key               1.000000000  0.014752719 -0.053539229  0.028288707
Loudness          0.014752719  1.000000000 -0.134288880  0.079367498
Mode             -0.053539229 -0.134288880  1.000000000 -0.060229608
Speechiness       0.028288707  0.079367498 -0.060229608  1.000000000
Acousticness     -0.038459294 -0.571188251  0.138987438 -0.102329514
Instrumentalness -0.001992626 -0.164921381 -0.030994389 -0.033494285
Liveness         -0.046975162  0.206387827 -0.089773428  0.011428365
Valence          -0.053815941  0.243380543 -0.093637481  0.138889456
Tempo             0.044612724  0.090155964  0.094889987  0.184885717
Time.Signature    0.014693155  0.041666898 -0.024543515 -0.006472836
                 Acousticness Instrumentalness    Liveness     Valence
Duration..ms.     -0.01768193      0.122450542 -0.03419471 -0.19713320
Popularity         0.08249996     -0.050323719 -0.08109104  0.09659480
Danceability      -0.12068847     -0.056138897 -0.03077475  0.43328208
Energy            -0.70463090     -0.036780412  0.25581302  0.36429344
Key               -0.03845929     -0.001992626 -0.04697516 -0.05381594
Loudness          -0.57118825     -0.164921381  0.20638783  0.24338054
Mode               0.13898744     -0.030994389 -0.08977343 -0.09363748
Speechiness       -0.10232951     -0.033494285  0.01142836  0.13888946
Acousticness       1.00000000      0.075786345 -0.24544309 -0.22966509
Instrumentalness   0.07578635      1.000000000 -0.02734370 -0.12987481
Liveness          -0.24544309     -0.027343698  1.00000000  0.14852207
Valence           -0.22966509     -0.129874814  0.14852207  1.00000000
Tempo             -0.09326738     -0.063713533  0.03400032  0.07297814
Time.Signature    -0.08327156     -0.032515147  0.11453872  0.16556860
                       Tempo Time.Signature
Duration..ms.    -0.08432772   -0.051815173
Popularity       -0.04298448    0.026901970
Danceability     -0.23893050    0.172973955
Energy            0.10553842    0.106315210
Key               0.04461272    0.014693155
Loudness          0.09015596    0.041666898
Mode              0.09488999   -0.024543515
Speechiness       0.18488572   -0.006472836
Acousticness     -0.09326738   -0.083271563
Instrumentalness -0.06371353   -0.032515147
Liveness          0.03400032    0.114538723
Valence           0.07297814    0.165568598
Tempo             1.00000000   -0.087320859
Time.Signature   -0.08732086    1.000000000
```

In [76]:

```
# Heatmap of correlations
ggcorrplot(correlation_matrix, lab = TRUE)
```



Conclusion

This report showcases the preprocessing, cleaning, and exploratory analysis of the dataset. Insights include:

- Popularity distribution skews low.
- Danceability is concentrated around 0.5–0.6.
- Tempo shows peaks around common BPM ranges.
- Correlation analysis highlights relationships among features.

Further steps could involve feature engineering or advanced modeling for deeper insights.