TEAM RESEARCH AND DEVELOPMENT PROJECT REPORT

Group: 7COM1079-new_group 44

Student ID: 19025662, 18021677, 18044139, 19025802, 19063511

Table of Contents

Introduction

Almost everyone likes entertainment because entertainment is mind relaxing and some of the entertainment forms include football, music (Bates and Ferri, 2010, pp. 1-20). The entertainment industry specifically the music industry is growing more competitive as artists become more and more creative in their artistry skills (Power and Hallencreutz, 2007, pp. 377-389). As a consequence, this has led to some songs being more popular than others (Berns et al., 2010, pp. 2687-2696). For music artists to remain relevant and continue producing songs that are hit and hit, then they have to determine factors that will lead to the popularity of their songs. Therefore, to help the artists in the music industry determine factors that affect or lead to the popularity of a song, this report will show the following factors that affect music popularity: valence, danceability, acoustic ness, instrumentals, liveness, loudness, speechiness as well as energy.

Data set Research Question and Hypothesis

The research collected data from Spotify for a significant number of songs for the year 2020. The data set was collected from the link, https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks. Of the various variables available, the most significant variables to be analyzed were as follows: Popularity of the song was the dependent variable while independent variables included the valence, danceability, acoustiscness, instrumentalness, liveness, loudness, speechiness as well as energy (Ashrith, 2018). For convenient transfer of the data into R programming for analysis, the data was converted and stored in Comma-Separated Values. Before data analysis, the data was cleaned whereby any missing values were checked, non-numerical values also were checked, the data was then sorted and arranged in required terms to facilitate effective analysis (Chu et al., 2016, pp. 2201-2206).

The research also formulated a hypothesis statement that will guide the research analysis. A hypothetical statement is a statement that can be true or false until it is proved (Ehmke, Lusk and List, 2008, pp. 498-500). Therefore, the Null Hypothesis for the research was as follows:

Ho: No characteristics of a song influence its popularity.

On the other hand, the alternative hypothesis for the research was as follows:

Ha: Some characteristics can influence the popularity of a song.

The above hypothesis will be determined to be true or false after the research has analyzed various data factors collected.

Another significant guidance to the research is the research question which will keep and guide the research. The specific research question in this case was:

RQ: Is there a pattern in song instrumental data that influences its popularity?

After data has been collected, processed into CSV and cleaned, a hypothesis identified and then a research question formulated, the data will then be ready to be analyzed using R-Statistical Programming software for correlation or association analysis.

Correlation/Association Analysis

R-Programming software was used for analysis in which correlation function, as well as the plot function, was used. In the plot function, the options used were xlim, ylim, and pch whereby xlim=limit of x which was from 0-1, ylim = limit of y which was from 0-100 and pch was used to specify the simples of the points in which specifically pch=3 simple was used to represent a plus (+) simple or points of spread (Alboukadel, 2020). In this case analysis, black color was set as the default color in the R programming. The output results were scattered plus signs and decision rule that the research used to make was where there was more concentration.

Influence of Valence on Song to become a Hit

Song popularity was the dependent variable while valence was the independent variable. Valence or the hedonic tone of a song can be defined as the affective quality of the song which refers to the intrinsic attractiveness or goodness or the averseness of a song- positive valence (Ashrith, 2018). Valence can also be negative in which the song will not be attractive.

From the analysis of the results, positive valence of the song was considered- x axis of 0-1. The result table showed that the positive valence had a significant influence on the song popularity for the results indicated song popularity of approximately 70%-80%. The result indicated that the valence be it small (0.2) or large (1.0), it had the same effect on the popularity of the song. However, zero valence had no impact on the popularity of the song. The results clearly indicate that the more positive the valence or attractiveness of a song is, the more the song will become popular and hit.

Influence of acousticsness on the popularity of the song

Again, acousticsness was the independent variable while the popularity of the song was the dependent variable. Music acousticsness can be defined as a multidisciplinary field that deals with researching as well as describing the physics of music, for instance, the way sound is employed to make music (Ashrith, 2018). Acoustics can entail the study of how musical instruments function, human voice as well as computer analysis of the melody (Ashrith, 2018).

From the results analysis output, the less acousticsness in the music the more and increased chances of the song becoming popular- popularity of approximately 70%-80%. In fact, the zero acousticsness in the music the more popular the song becomes. The higher the acousticsness in the song the less the chances of the song becoming popular. This result indicates that acousticsness in the song does not guarantee that a song will become a hit. This is true because even if an artist has studied music acoustics, there are still a significant number of factors that affect song popularity especially in the 21st century where there is a shift in the demands or music consumption content among people.

Influence of danceability on the popularity of a song

In this case, again, danceability is the independent variable while popularity of the song is the dependent variable. The danceability of a song can be determined by the use of the song mixture features such as the stability of the tempo, strength of the beat, and the overall tempo (Ashrith, 2018). In the end, the value that is returned determines the easiness with which an individual can dance to the entire song (Ashrith, 2018).

From the results analysis, the less the danceability to the song, the less the popularity of the song. As the song records high features of danceability, the more popular the song becomes. The result shows that danceability scale of approximately 0.9 have positive popularity to the song by approximately 70%-80%. Therefore, this is a clear indication to the musicians that they should focus more on the tempo, beat the strengths of the song for it to become a hit. This is the reason why music producers are behind successful hit songs because they are the ones concerned with creating music beats and tempos.

Influence of Instrumentalness on a song to become a hit

Again, instrumentalness is the independent variable while popularity of the song is the dependent variable. Instrumentalness can be defined as the vocal amount in the song and that a given specific song is more instrumental when its instrumentalness approaches 1 (Ashrith, 2018). Instrumentalness is related to the acousticsness of the song and this is because a song with instrumentalness score of 1 implies the song is acoustic (Ashrith, 2018).

Now, let us consider the results. From the presented results, it indicates that instrumentalness does not affect the popularity of the song. It does not influence the ability of a song to become a hit. From the instrumentalness scale from 0-1, there is no influence on the song's popularity which implies that low or high instrumentalness in a song has no impact on its popularity. Therefore, the results simply imply that a song can become a hit in absence of instrumentalness. This effectively agrees with its relation to acousticsness that they have a positive relationship. This is true because even from the earlier analysis, acousticsness of a song has no significant impact on the popularity of the song and this coincidence between the two shows that the data, as well as results analysis, is reliable and accurate.

Influence of liveliness of a song to become a hit

Again, the popularity of the song is the dependent variable while liveliness of the song is the independent variable. Liveliness means that the audience was a presence during the recording session of the song (Ashrith, 2018). If a given song was performed live, then it will have high values of liveliness (Ashrith, 2018).

From the results output, the lower the liveliness of a song the more popular a song will become. As the song becomes more and more live filled values, the less popular the song will become. Therefore, these findings implied that liveliness of a song has no significant impact on the popularity of the song. As a consequence, musicians should not pay more attention to the liveliness values of a song in their song recordings because this does not guarantee the popularity of the song.

Influence of Loudness of a song to become a hit

Loudness of the song was the independent variable while popularity of the song was the dependent variable (Ashrith, 2018). The values of loudness are averaged across the whole song. In this case, the loudness variable had a scale of both negative and positive values (Ashrith, 2018).

The results indicated that loudness of a song has no significant impact on the popularity of a song. In fact, the result showed a negative value of loudness to the popularity of a song. This implies that the more-loud a song is, the less the song will become popular. This is relatable because playing loud music will be noisy and not enjoyable. Therefore, artists and in this case music deejays and also the general public should not play loud music for it does not guarantee that the particular song being played will become popular.

Influence of speechiness on the popularity of the song

Speechiness was the independent variable while song popularity was the dependent variable. Speechiness can be defined as the spoken word presence in a track (Ashrith, 2018). This includes a speech-like recording, for instance, talk shows or poetry in a given song (Ashrith, 2018).

From the results table output, speechiness had a significant impact on the popularity of the song. The results show that 0.5 speechiness scale have an important impact on the popularity of a song. However, as the speechiness values in the song increases, this does not guarantee that the song will become popular. Therefore, music artists need to consider speechiness factor in their song recordings but they should not overdo it. This is because speechiness in the song has a significant impact on the popularity of the song.

Influence of energy on the popularity of a song

Song popularity is the dependent variable while energy is the independent variable. Energy in a song can be defined as the perceptual measure of intensity as well as activity in a song (Ashrith, 2018). For instance, music or songs that are energetic in nature are loud, fast as well as noisy.

From the results table, the factor of energy in a song significantly affects the popularity of the song. The result indicates that a less energetic song cannot become a hit. On the other hand, a

song that is more energetic have improved and higher chances of becoming popular. Therefore, music artists have to importantly focus and consider the aspect of energy in their music recordings because this feature is significant in influencing the popularity of a song.

Analysis

We used Pearson's correlation analysis on various characteristics of a song. Since it is based on the covariance principle, it is known as the best method of calculating the relationship between variables of interest. It provides data on the extent of the relationship, or correlation, as well as the direction of the relationship. The findings of the analysis are as follows:

Energy vs Popularity

Pearson's product-moment correlation

data: songs_data$energy and songs_data$popularity
t = 229.11, df = 170651, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4813682 0.4886252
sample estimates:
   cor
0.485005

Valence vs Popularity

Pearson's product-moment correlation

data: songs_data$valence and songs_data$popularity
t = 5.8668, df = 170651, p-value = 4.452e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.009456557 0.018943671

sample estimates:

cor

0.01420043

Acousticness vs Popularity

Pearson's product-moment correlation

data: songs_data$acousticness and songs_data$popularity

t = -288.94, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.5763390 -0.5699672

sample estimates:

cor

-0.5731618

Danceability vs Popularity

Pearson's product-moment correlation

data: songs_data$danceability and songs_data$popularity

t = 84.151, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 0.1950464 0.2041573

sample estimates:

cor

0.1996062

Instrumental Value vs Popularity
Pearson's product-moment correlation

data: songs_data$instrumentalness and songs_data$popularity

t = -128.37, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.3010709 -0.2924174

sample estimates:

     cor

-0.2967503

Liveliness vs Popularity
Pearson's product-moment correlation

data: songs_data$liveness and songs_data$popularity

t = -31.68, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.08117914 -0.07174559

sample estimates:

     cor

-0.07646407

Loudness vs Popularity
Pearson's product-moment correlation

data: songs_data$loudness and songs_data$popularity

t = 212.28, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 0.4532891 0.4607959

sample estimates:

   cor

0.4570506

Speechiness vs Popularity

Pearson's product-moment correlation

data: songs_data$speechiness and songs_data$popularity

t = -72.119, df = 170651, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

 -0.1765792 -0.1673708

sample estimates:

    cor

-0.1719787

Conclusion

It is clear that several factors affect the popularity of a song for it to become a hit. This research collected data from Spotify till the year 2020 so as to determine factors influencing the popularity of a song. The collected data was processed and cleaned, converted into CSV format before being analyzed in R Statistical programming software. The Null Hypothesis was: No characteristics of a song influence its popularity. From the analysis of the results, it was found that speechiness, valence, danceability, energetic features have a very significant impact on the popularity of a song. However, factors such as acousticsness, instrumentalness, loudness, and liveliness have a less significant impact on the popularity of a song. Therefore, from the analysis of the results, the null hypothesis is false and the alternative hypothesis is true. Therefore, certain factors significantly affect the ability of a song to become popular. Therefore, music artists have to consider valence, danceability, acoustiscness, instrumentalness, liveness, loudness, speechiness as well as energy factors in their music recordings because these factors affect the popularity of a song.

Through Pearson Correlation we found out the most significant characteristics that could affect the popularity of a song.

Energy vs Popularity: cor 0.485005

Loudness vs Popularity: cor 0.4570506

References

Alboukadel, 2020. DATA NOVIA, PCH IN R BEST TIPS. Retrieved from:

https://www.datanovia.com/en/blog/pch-in-r-best-tips/

Ashrith, 2018. What Makes a Song Likeable? Retrieved from:

https://towardsdatascience.com/what-makes-a-song-likeable-dffdb7abe404.

Bates S., and Ferri A.J., 2010. What's entertainment? Notes toward a definition. *Studies in Popular Culture 33 (1),* pp. 1-20.

Berns G.S., Capra C.M., Moore S., Noussair C., 2010. Neural mechanisms of the influence of popularity on adolescents' ratings of music. Neuroimage 49 (3), pp. 2687-2696.

Chu X., Ilyas I.F., Krishnan S., Wang J.,2016. Data cleaning: Overview and emerging challenges. *Proceedings of the 2016 International Conference on Management of Data*, pp. 2201-2206.

Ehmke M.D., Lusk J.L., List J.A., 2008. Is hypothetical bias a universal phenomenon? A multinational investigation. *Land economics 84 (3),* pp. 489-500.

Power D., and Hallencreutz D., 2007. 'Competitiveness, local production systems and global commodity chains in the music industry: entering the US market'. *Regional Studies 41 (3)*, pp. 377-389.

Appendix

Output.R

2021-01-06

```r
#pdf("visualization.pdf")
library(base)

songs_data<-read.csv("D:/Downloads/data.csv", header = TRUE)
#loading the data

#View(songs_data)
#checking the data

selected_data<-subset(songs_data, year==2020)
#extracting data we want to work on


plot(selected_data$valence, selected_data$popularity, xlab="Valence", ylab="P
opularity",main="Influence of Valence on Song to become HIT(2020)",xlim = c(0
,1), ylim = c(0,100), pch=3)
```
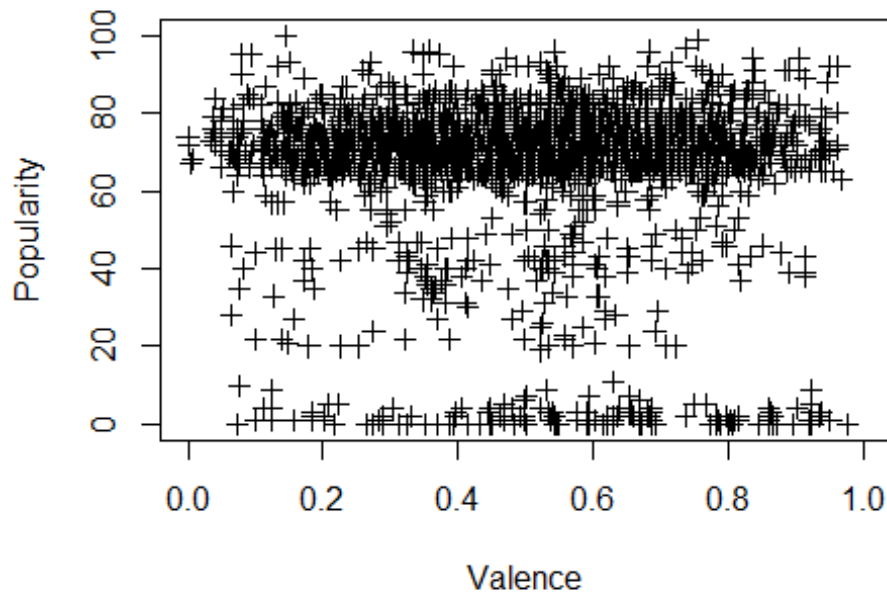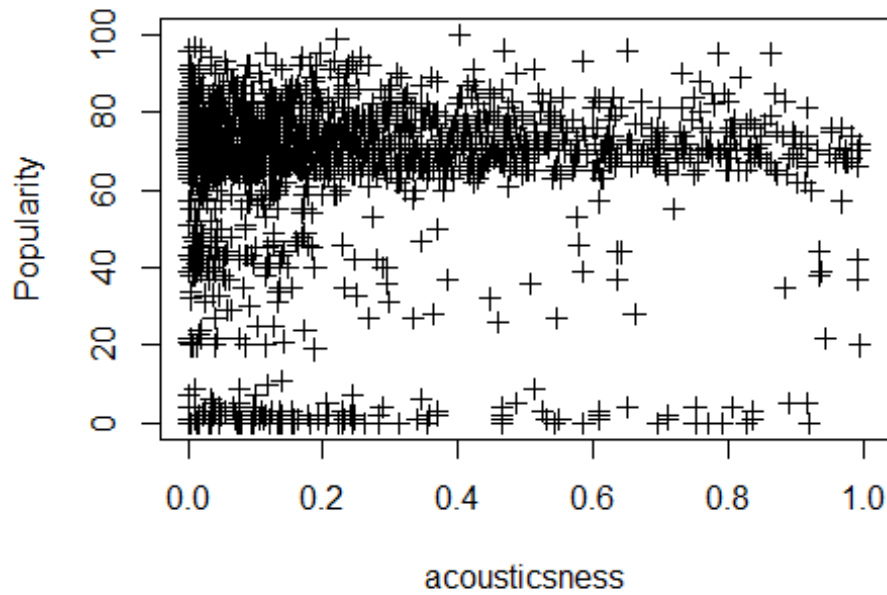

Influence of Valence on Song to become HIT(2020)

```
plot(selected_data$acousticness, selected_data$popularity, xlab="acousticsnes
s", ylab="Popularity",main="Influence of acousticsness on Song to become HIT(
2020)",xlim = c(0,1), ylim = c(0,100), pch=3)
```

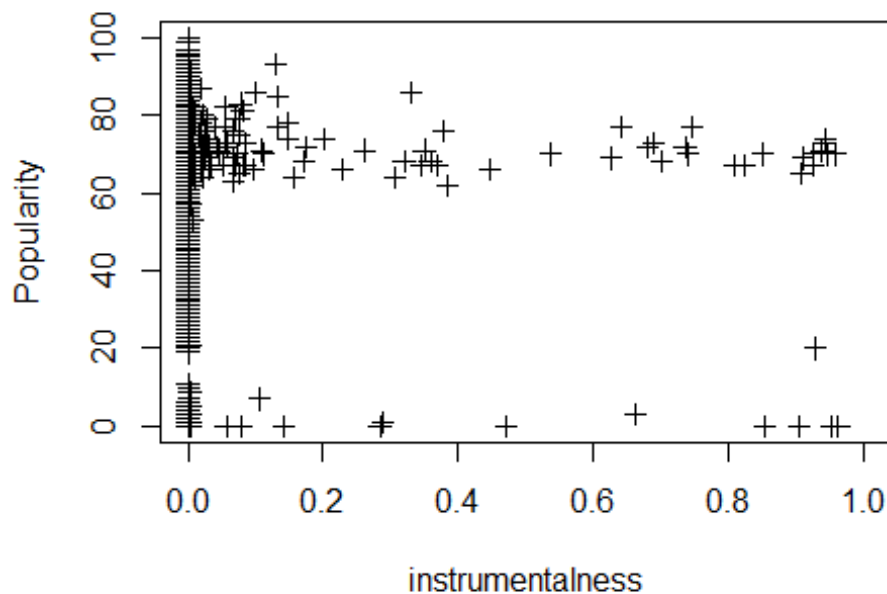## nfluence of acousticsness on Song to become HIT(2


```
plot(selected_data$danceability, selected_data$popularity, xlab="danceability
", ylab="Popularity",main="Influence of danceability on Song to become HIT(20
20)",xlim = c(0,1), ylim = c(0,100), pch=3)
```

Influence of danceability on Song to become HIT(20
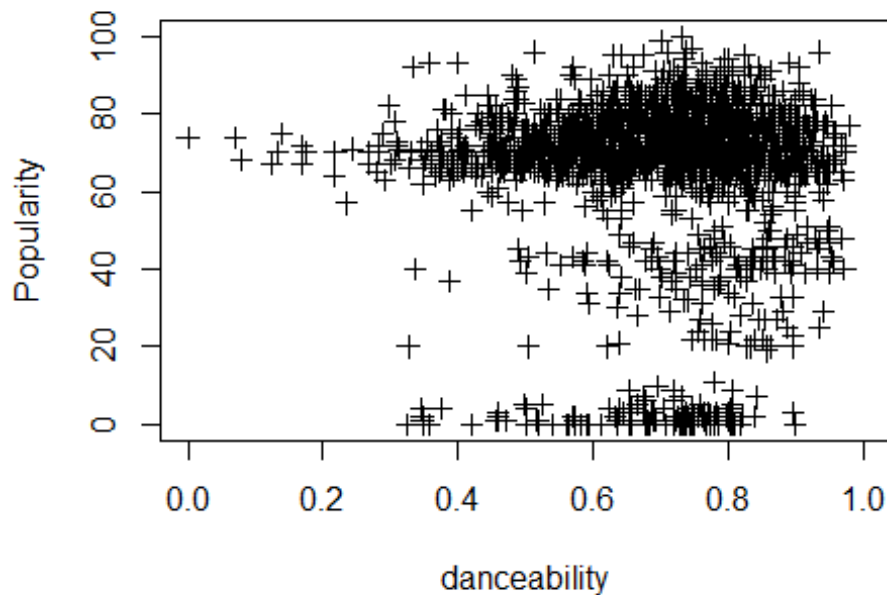

```
plot(selected_data$instrumentalness, selected_data$popularity, xlab="instrume
ntalness", ylab="Popularity",main="Influence of instrumentalness on Song to b
ecome HIT(2020)",xlim = c(0,1), ylim = c(0,100), pch=3)
```
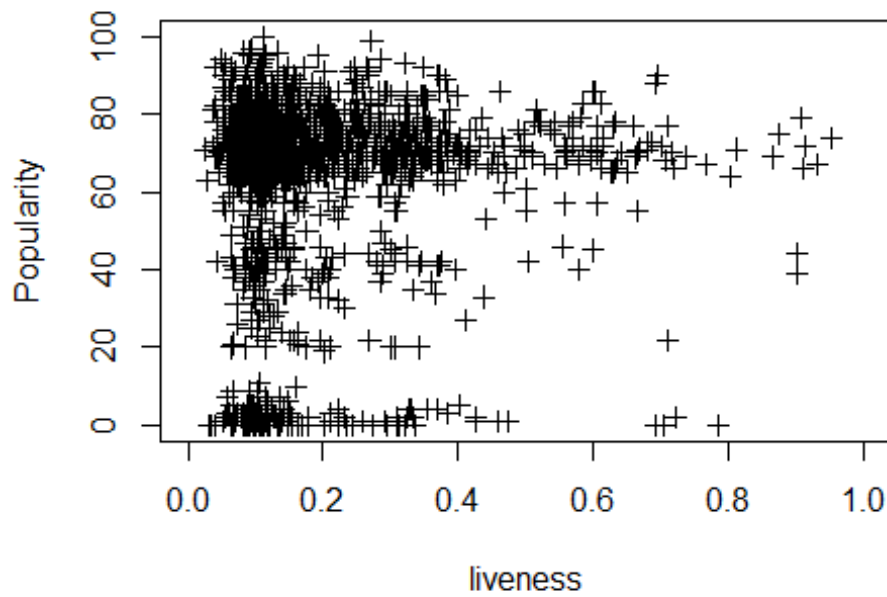
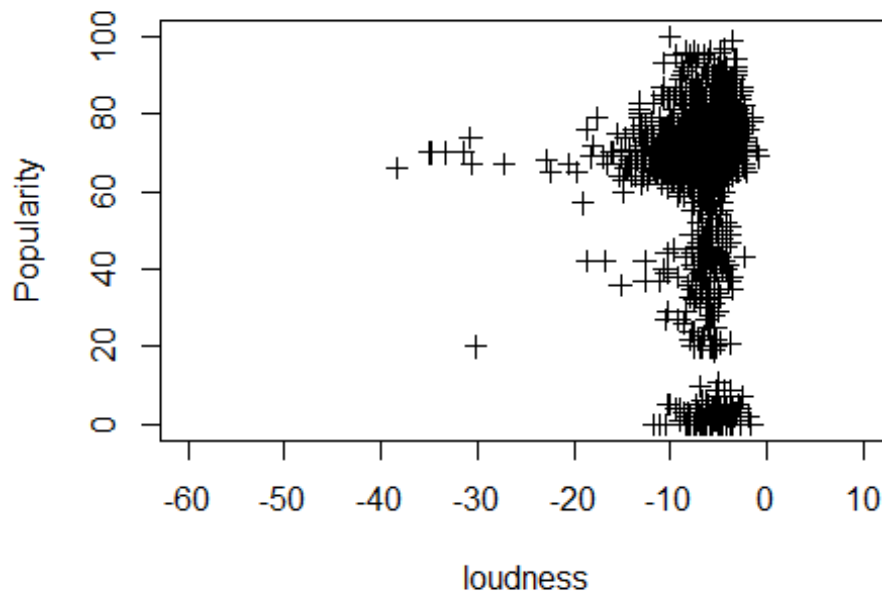## fluence of instrumentalness on Song to become HIT

```
plot(selected_data$liveness, selected_data$popularity, xlab="liveness", ylab=
"Popularity",main="Influence of liveness on Song to become HIT(2020)",xlim =
c(0,1), ylim = c(0,100), pch=3)
```

**Influence of liveness on Song to become HIT(2020**


```
plot(selected_data$loudness, selected_data$popularity, xlab="loudness", ylab=
"Popularity",main="Influence of loudness on Song to become HIT(2020)",xlim =
c(-60,10), ylim = c(0,100), pch=3)
```

Influence of loudness on Song to become HIT(202


```r
plot(selected_data$speechiness, selected_data$popularity, xlab="speechiness",
ylab="Popularity",main="Influence of speechiness on Song to become HIT(2020)"
,xlim = c(0,1), ylim = c(0,100), pch=3)
```
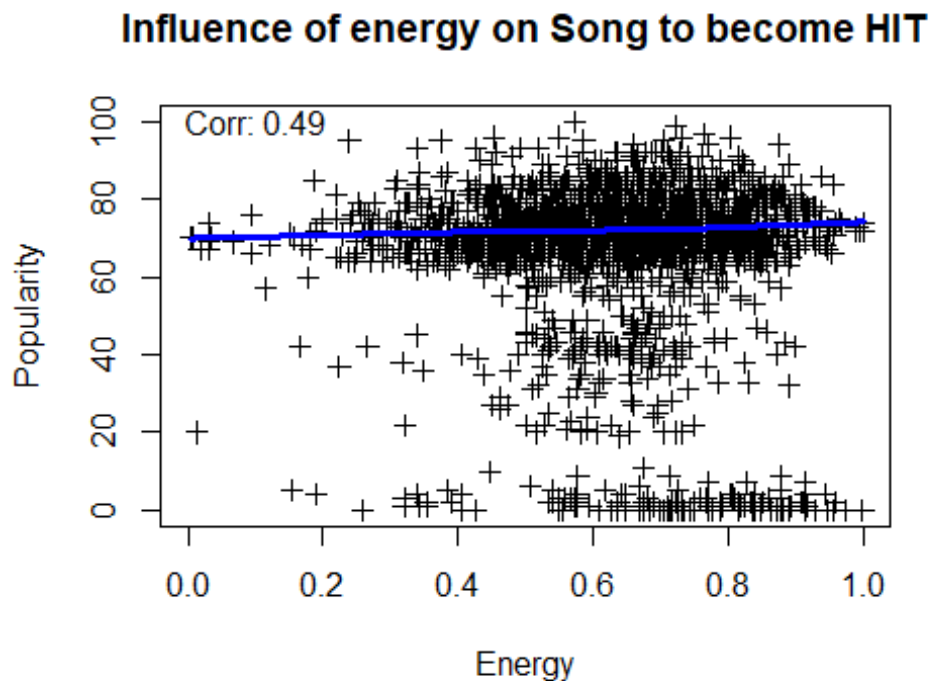
Influence of speechiness on Song to become HIT(2(
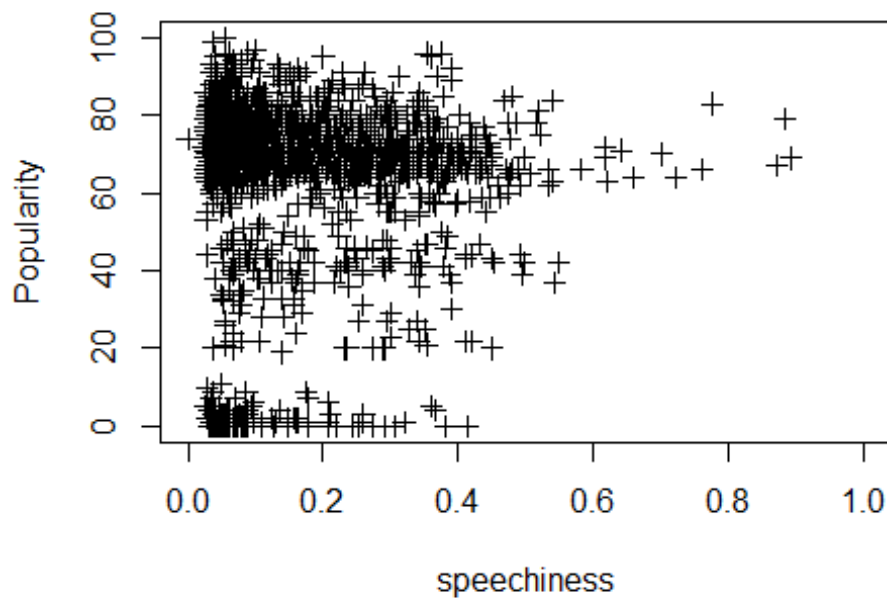
```
#Highest Influence on song popularity according to data.
Corr <- cor(songs_data$energy, songs_data$popularity)

plot(selected_data$energy, selected_data$popularity, xlim = c(0,1), ylim = c(
0,100), main="Influence of energy on Song to become HIT", xlab="Energy", ylab
="Popularity", pch=3)
lines(lowess(selected_data$energy, selected_data$popularity), col="blue",lwd=
3)
text(paste("Corr:", round(Corr,2)), x=0.1, y=100)
```


Influence of energy on Song to become HIT

```
dev.off()

## null device
##           1
```