An analysis of music hits across decades: 1950-2009

Naresh N. Vempala

January 24, 2016

This is a dataset consisting of audio attributes for 600 songs. These are the top 100 hit songs from six decades: 1950s, 1960s, 1970s, 1980s, 1990s, and 2000-2009. The data represents derived work collected from two sources - tsort.info/music/ and the Echo Nest database (http://the.echonest.com/). For more information about the dataset and its attributes, how the dataset was created, and how to use it (licensing and citation info), please refer to the README.md file.

More information about song attributes is available here: http://runningwithdata.com/post/1321504427/danceability-and-energy http://developer.echonest.com/acoustic-attributes.html

To replicate the results below in R, first download the dataset with the decades data titled "NVDecades.csv". Then make sure to set your working directory in R to the folder containing this dataset, or mention the full/absolute path of the file in the read.csv function.

Import the dataset into your working directory:

```
datNew <- read.csv(file = "NVDecades.csv", header = TRUE, sep = ",")
```

Load the required packages in R after installing them first:

```
library(plyr)
library(ggplot2)
library(corrplot)
```

Box plots are a great way to summarize and compare distributions of data from each decade for any attribute. Here are some box plots for each decade.

Box plots of danceability:

```
ggplot(datNew, aes(x=decade, y=danceability, fill=decade)) + geom_boxplot() +
        stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```
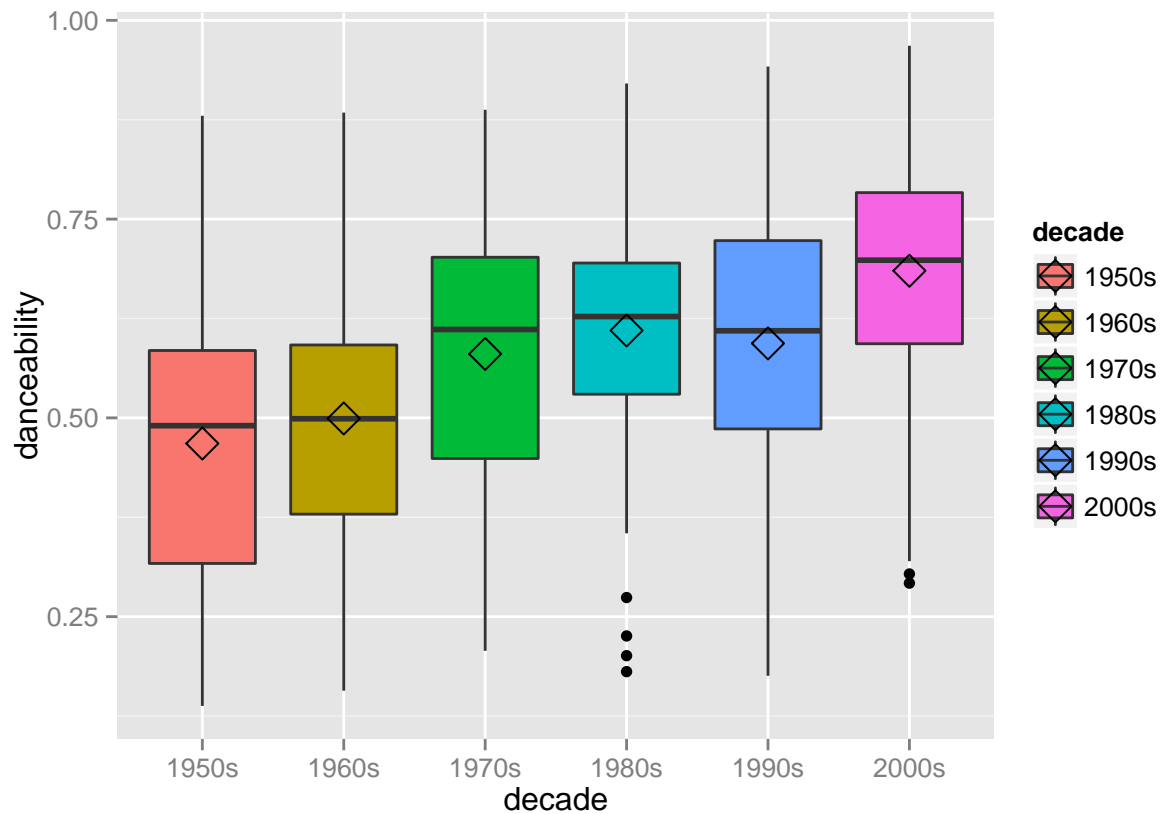
The Echo Nest describes danceability as how suitable a track is for dancing. The Echo Nest computes this attribute using a combination of tempo, rhythm stability, beat strength, and regularity. When considering the median danceability for each decade, we notice two obvious jumps in danceability - one for the 1970s and one for the 2000s. The sudden jump in the 1970s clearly represents how the 70s ushered in the disco and disco-funk era - a departure from the rock & roll and R&B hits of the 50s and 60s.
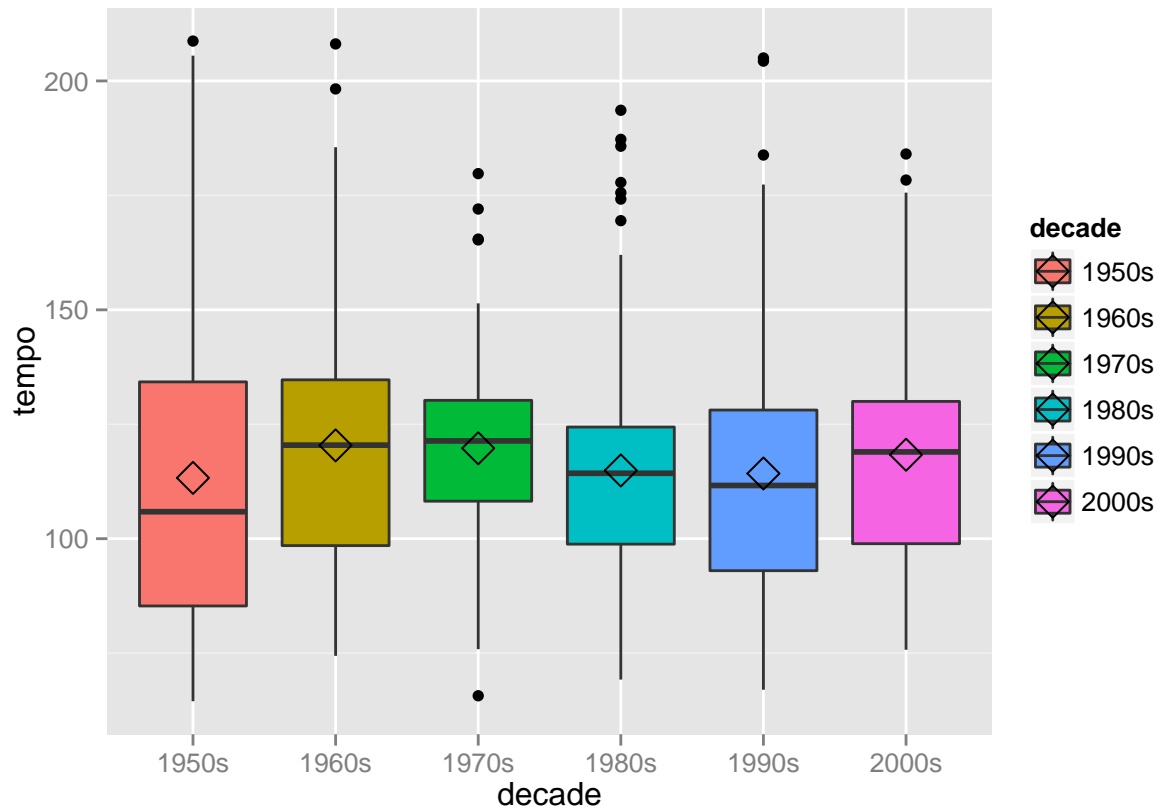
Box plots of energy:

```r
ggplot(datNew, aes(x=decade, y=energy, fill=decade)) + geom_boxplot() +
        stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```
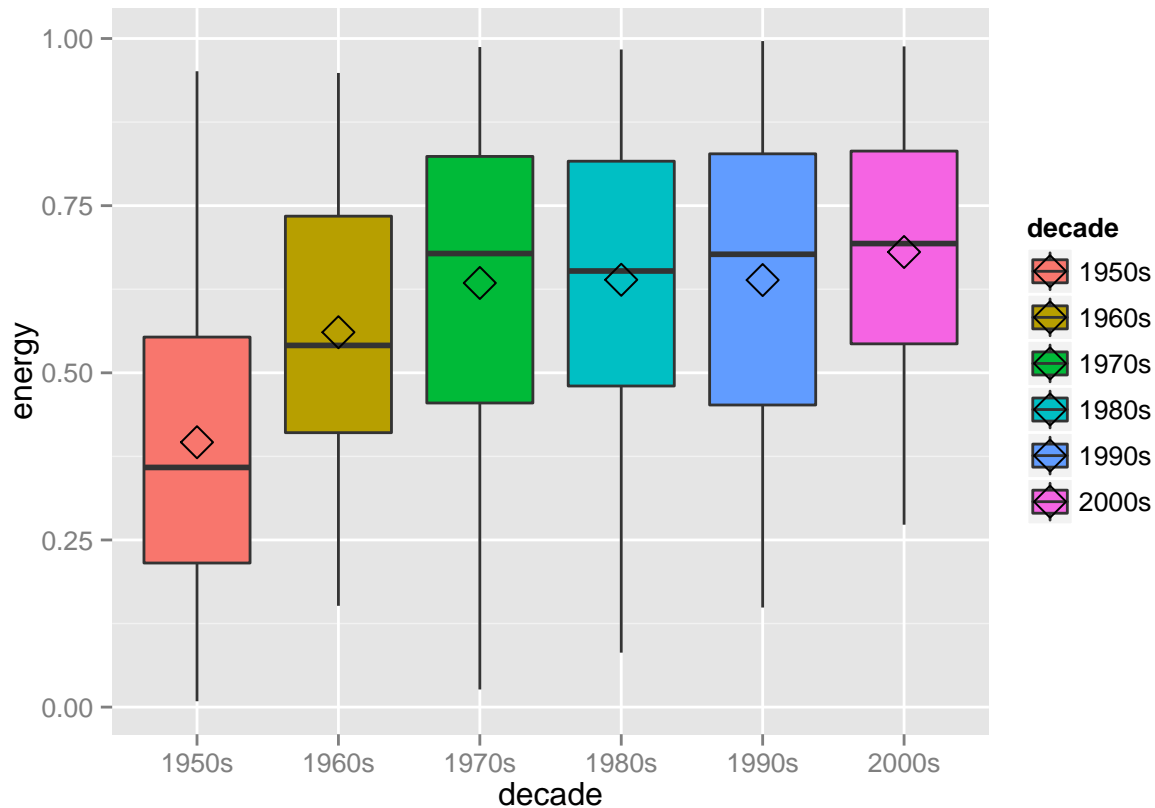
Energy represents how energetic the song is. While danceability is more subjective to the listener, energy is more directly dependent on the audio characteristics of the song. The Echo Nest uses a combination of loudness and segment durations to compute energy. We notice a spike in the energy once in the 60s and again in the 70s. One possible explanation for the increase in energy levels of songs in the 60s could be the popularity of amplified distortion (especially in the latter part of the 60s). As for the 70s, this was when big mixing consoles (still analog, but allowing 32 channels) started gaining popularity in recording studios, representing a big revolution in music recording. More audio content could be accommodated and EQs could be adjusted. Could this perhaps have had something to do with the increase in energy starting with the 70s?
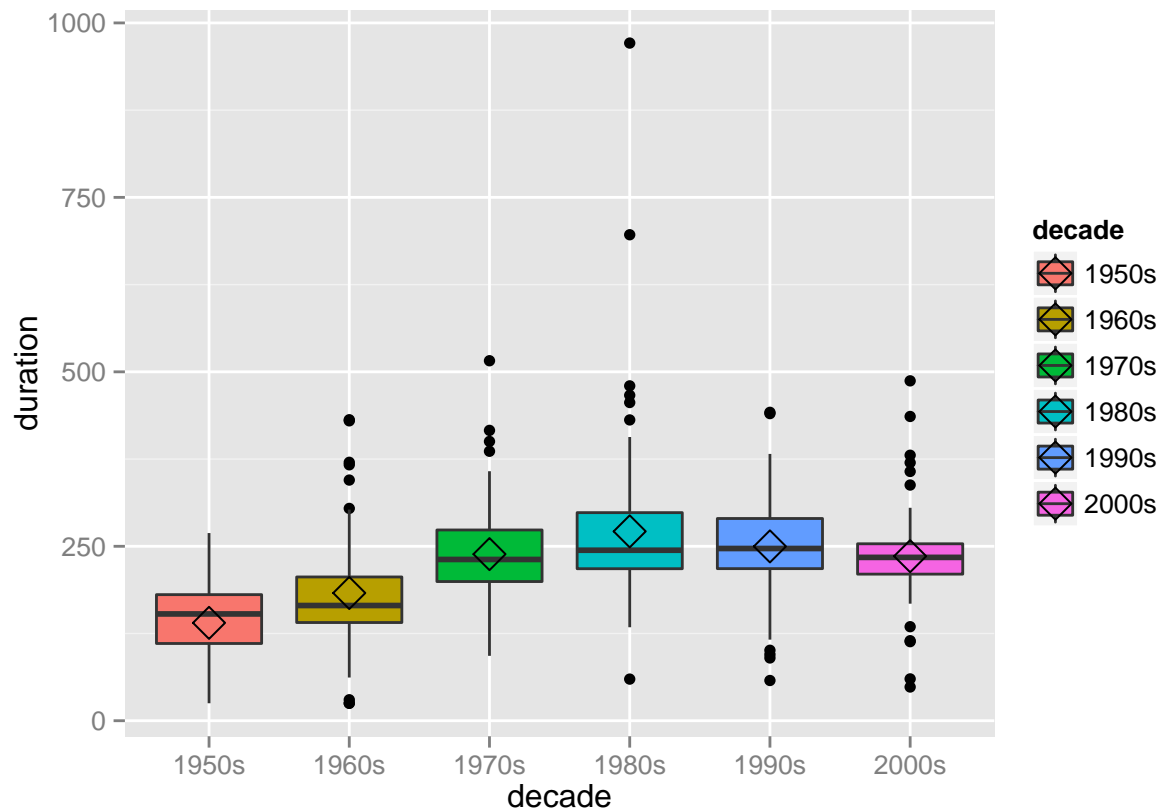
Box plots of tempo:

```
ggplot(datNew, aes(x=decade, y=tempo, fill=decade)) + geom_boxplot() +
        stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```

We don't notice much difference in median tempi between decades. As one might likely expect, songs in the 50s have the lowest tempi. However, an interesting thing to note is that hit songs from the 50s varied more in tempo than in the other decades, as perceived by the size of the rectangle in the box plot. In other words, the 50s exhibited a wider range of tempi.

Box plots of duration:

```
ggplot(datNew, aes(x=decade, y=duration, fill=decade)) + geom_boxplot() +
       stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```

We don't see anything of note here, except that songs in the 50s and 60s were somewhat shorter in duration.
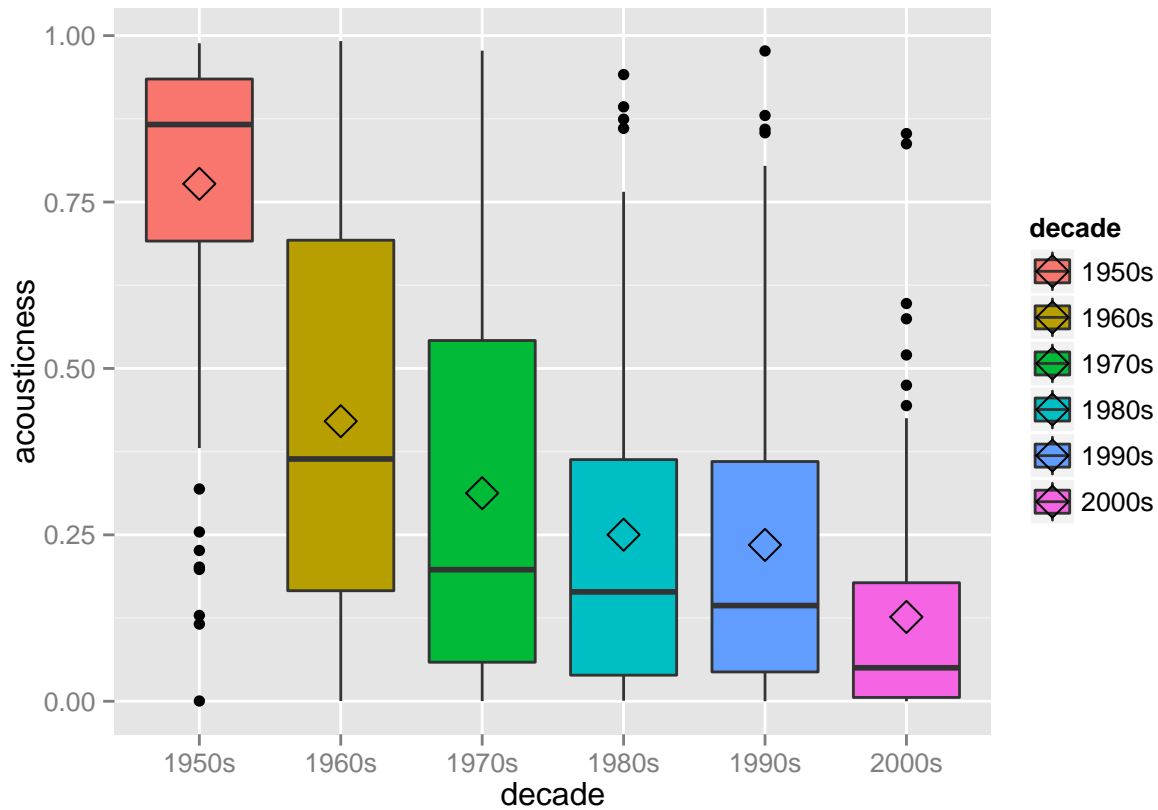
Box plots of acousticness:

```
ggplot(datNew, aes(x=decade, y=acousticness, fill=decade)) + geom_boxplot() +
        stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```

According to the Echo Nest, acousticness is a measure of the likelihood that a song was created by acoustic means such as voice and acoustic instruments that are not electronically synthesized or amplified. So, the inclusion of electric guitars, distortion, synthesizers, auto-tuned vocals, and drum machines will considerably lower the acousticness of a song. Not surprisingly, we notice a big drop in acousticness from the 50s to the 60s when distorted amplifiers and electric guitars started becoming popular. We also notice another drop, although not as huge, between the 60s and the 70s corresponding to the advances in recording consoles as well as the use of moog synthesizers.

Box plots of valence:

```
ggplot(datNew, aes(x=decade, y=valence, fill=decade)) + geom_boxplot() +
        stat_summary(fun.y=mean, geom="point", shape=5, size=4)
```
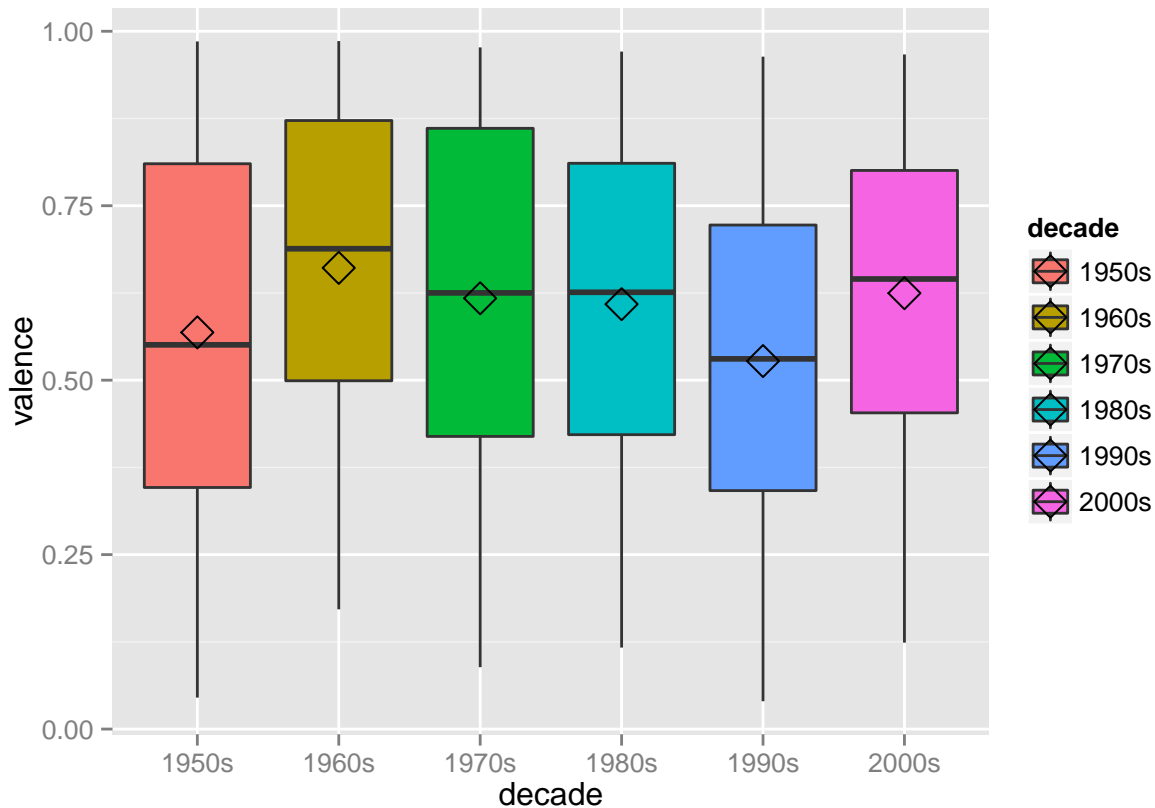
Valence is a subjective measure of pleasantness and is very listener-dependent. It needs to be interpreted with caution. The Echo Nest associates valence with positivity. The slightly lower median valence for the top songs of the 50s and the 90s when compared with other decades is hard to explain.

Now that we have a summarized overview of how hit songs are distributed with respect to these attributes across each decade, we can ask some interesting questions:

(1) The box plots for danceability and energy show a similar pattern. Is a more energetic song generally more danceable or vice versa?
(2) Do faster songs tend to be more energetic? (or more danceable?)
(3) How do danceability, energy, duration, and tempo affect how pleasurable a song is? In other words, how are these attributes correlated with valence?

A good basic method to help address these questions is to compute cross-correlations (i.e., Pearson's correlation coefficients).

```
correlationMatrix <- cor(datNew[,c(6,8,10,14,17,16)])
print(correlationMatrix)
```

```
##                   energy       tempo acousticness    duration danceability
## energy         1.0000000  0.20897553   -0.6036450  0.15925288    0.23183629
## tempo          0.2089755  1.00000000   -0.1022250  0.04413172   -0.05609813
## acousticness  -0.6036450 -0.10222501    1.0000000 -0.31943776   -0.35774373
## duration       0.1592529  0.04413172   -0.3194378  1.00000000   -0.01033423
## danceability   0.2318363 -0.05609813   -0.3577437 -0.01033423    1.00000000
## valence        0.3549200  0.13759455   -0.2140745 -0.24896454    0.51479867
##                  valence
## energy         0.3549200
```

7

```
## tempo         0.1375945
## acousticness -0.2140745
## duration     -0.2489645
## danceability  0.5147987
## valence       1.0000000
```

Perhaps a better way to visualize these correlations would be either as a table of correlations or as a series of ellipses. In the second plot, the closer the ellipse is to a straight line, the more correlated it is. This is also indicated by its color. Additionally, the direction of correlation between any two attributes is also indicated.

```
corrplot(correlationMatrix, method = "number")
```
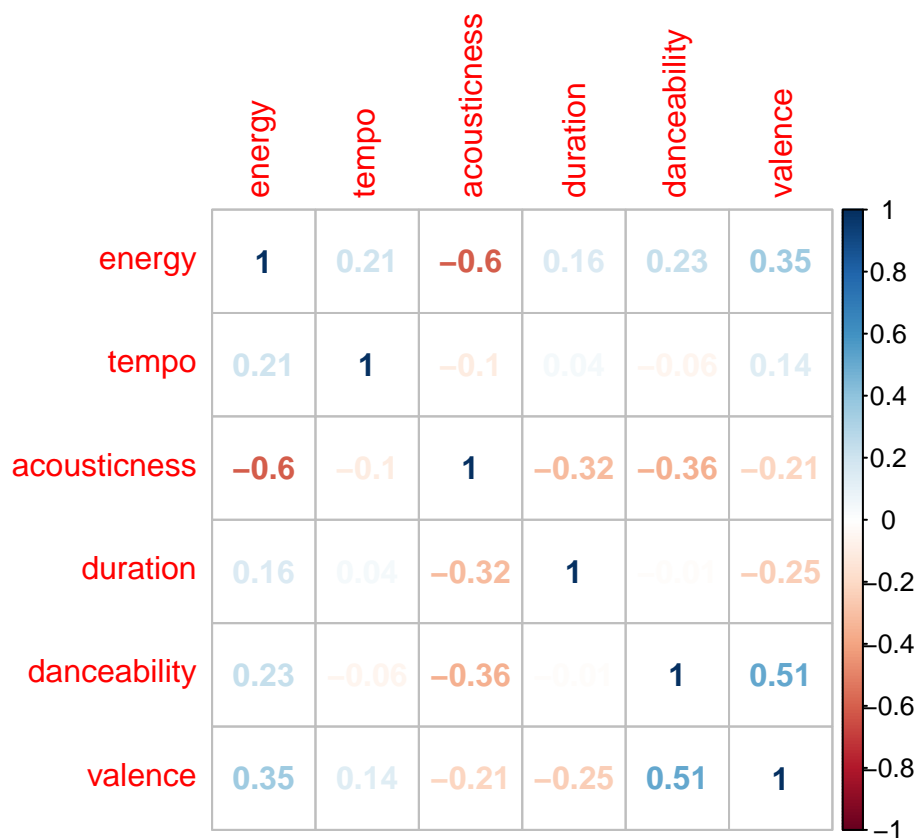


```
corrplot(correlationMatrix, method = "ellipse")
```
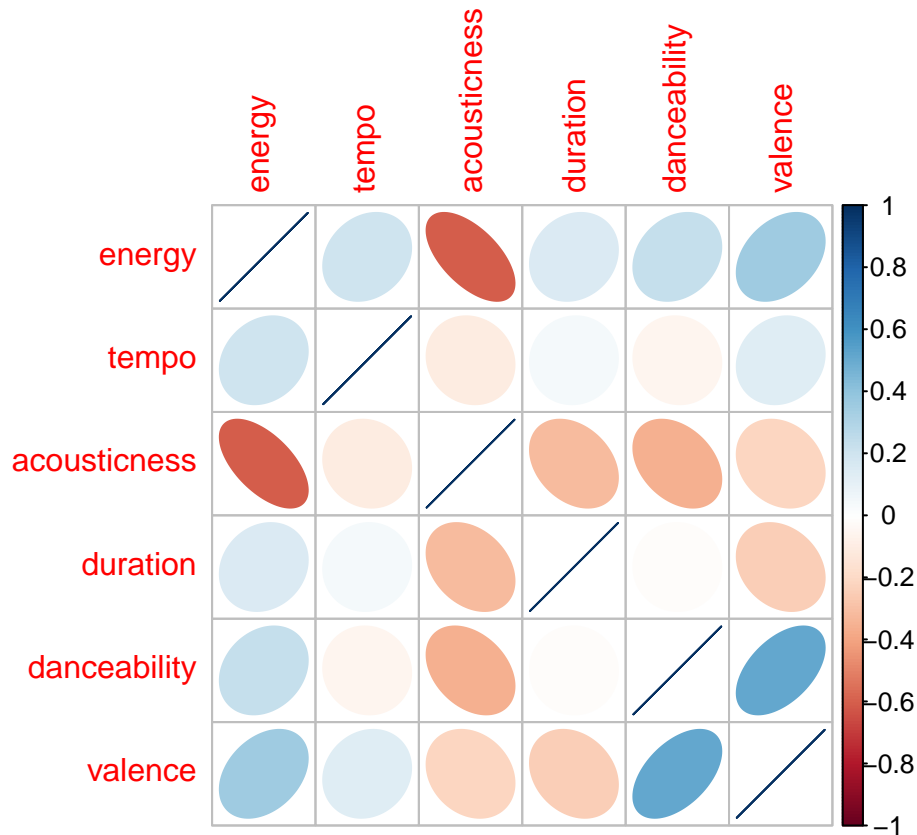
Coming back to our questions, we notice that energy and danceability are positively correlated, although the correlation is weak. Again there is a weak positive correlation between tempo and energy, and tempo and danceability. This suggests that faster songs might sound more energetic or danceable but not always. The positive correlation between danceability and valence is much clearer, as might be expected. The more danceable a song is, the more positive or pleasant the listening experience. Another clear insight is the strong negative correlation between acousticness and energy. So, the less acoustic or more synthesized (i.e., artificially amplified, compressed etc.) a song is, the more energetic it sounds.

We can also dig deeper and examine each decade. Each decade could be thought of as a genre. This becomes clearer in retrospect. For instance, songs of the 70s when thought of as one unified schema, have a distinct sound, a strong association with disco/disco-funk etc. However, when living in the 70s and listening to music, this might not have seemed obvious, given that the 70s had a variety of music - hard rock, early progressive rock and heavy metal, disco, synth rock, funk and so on.

But first, let's find the top 5 songs with maximum danceability, across all decades.

```
datTopdance <- head(arrange(datNew, -danceability), n = 5)
datTopdance[, c(3,18,19,17)]
```

```
##                    artist_name                            title decade
## 1        Justin Timberlake                              SexyBack  2000s
## 2                    Eminem                          Just Lose It  2000s
## 3            Gwen Stefani                        Hollaback Girl  2000s
## 4 Puff Daddy & Faith Evans I'll Be Missing You (Instrumental)  1990s
## 5         Michael Jackson                            Billie Jean  1980s
##   danceability
## 1     0.968045
## 2     0.960278
## 3     0.959466
```

```
## 4      0.941884
## 5      0.920595
```

Three of the top 5 are from the 2000s, two from the 90s, and one from the 80s. Justin Timberlake's "SexyBack" comes out at #1 (note that this dataset contains songs only up to 2009).

What are the top 5 songs with maximum energy, across all decades?

```
datTopenergy <- head(arrange(datNew, -energy), n = 5)
datTopenergy[, c(3,18,19,6)]
```

```
##      artist_name                                          title decade
## 1  Culture Beat                                        Mr. Vain  1990s
## 2 Elvis Presley A Little Less Conversation (JXL Radio Edit Remix)  2000s
## 3  Scatman John                     Scatman (ski-ba-bop-ba-dop-bop)  1990s
## 4 Gloria Gaynor                                    I Will Survive  1970s
## 5         Queen                     Another One Bites The Dust  1980s
##      energy
## 1 0.996188
## 2 0.988198
## 3 0.987580
## 4 0.987276
## 5 0.983594
```

We have a better contribution from the decades here, with at least one song each from the 70s, 80s, 90s, and 2000s. Since we have an officially streamable version available on YouTube, here is Scatman John's Scatman. Gloria Gaynor's "I Will Survive" is an interesting entry, suggesting that the song remains a mood uplifter perhaps not just because of the lyrics but also because of the acoustic characteristics that make the song more energetic.

What are the top 5 most pleasant songs (i.e., songs with maximum valence), across all decades?

```
datTopvalence <- head(arrange(datNew, -valence), n = 5)
datTopvalence[, c(3,18,19,16)]
```

```
##      artist_name               title decade  valence
## 1   Elvis Presley      Stuck on You  1960s 0.985961
## 2 The Kalin Twins              When  1950s 0.985415
## 3       Kingsmen      Louie Louie  1960s 0.985127
## 4   Elvis Presley  Good Luck Charm  1960s 0.977265
## 5 Three Dog Night Joy To The World  1970s 0.976910
```

Elvis scores high with two entries in this list.

What are the top 5 songs with maximum acousticness?

```
datTopacousticness <- head(arrange(datNew, -acousticness), n = 5)
datTopacousticness[, c(3,18,19,10)]
```

```
##                         artist_name           title decade acousticness
## 1                         Kingsmen      Louie Louie  1960s      0.991799
## 2                         Doris Day        Bewitched  1950s      0.988521
## 3 Chris Barber's Jazz & Blues Band     Petite Fleur  1950s      0.988425
## 4                     Elvis Presley  Return To Sender  1960s      0.981618
## 5                   Barbra Streisand  The Way We Were  1970s      0.977376
```

As might be expected, 4 of the songs are from the 50s and the 60s.

Just for fun, what are the 5 songs with minimum acousticness?

```
datLeastacousticness <- head(arrange(datNew, acousticness), n = 5)
datLeastacousticness[, c(3,18,19,10)]
```

```
##           artist_name              title decade acousticness
## 1     Ricky Martin Livin' La Vida Loca  1990s       1.9e-05
## 2 Right Said Fred        I'm Too Sexy  1990s       2.2e-05
## 3  Britney Spears               Toxic  2000s       2.7e-05
## 4      Katy Perry         Hot N Cold  2000s       6.9e-05
## 5      Ace of Base  All That She Wants  1990s       8.1e-05
```

This is a fun list. Let's recall what lesser acousticness means. Lesser acousticness indicates that a song is more electronically synthesized or amplified. The inclusion of electric guitars, distortion, synthesizers, auto-tuned vocals, and drum machines would have considerably lowered the acousticness of a song. Ricky Martin's "Livin' La Vida Loca" requires special mention here. This is a great song, in no small part due to Ricky Martin's delivery and showmanship, but also because of being a watershed moment in music recording history. This song was noted for its extreme use of dynamic range compression to increase the perceived loudness of the song. More information about the recording process behind La Vida Loca is provided in this article.

We can dig deeper and examine the top songs for each decade.

What are the 5 most danceable songs from the 50's?

```
dat50s <- datNew[which(datNew$decade=='1950s'),]
dat50sTopdance <- head(arrange(dat50s, desc(danceability)), n = 5)
dat50sTopdance[, c(3,18,19,17)]
```

```
##             artist_name                                title decade
## 1         Phil Harris                            The Thing  1950s
## 2         Dinah Shore      Dear Hearts and Gentle People  1950s
## 3         Perry Como Don't Let The Stars Get In Your Eyes  1950s
## 4 Les Paul & Mary Ford                   How High the Moon  1950s
## 5     Rosemary Clooney                      This Ole House  1950s
##   danceability
## 1     0.879959
## 2     0.854778
## 3     0.826361
## 4     0.795677
## 5     0.769732
```

What are the 5 most danceable songs from the 60's?

```
dat60s <- datNew[which(datNew$decade=='1960s'),]
dat60sTopdance <- head(arrange(dat60s, desc(danceability)), n = 5)
dat60sTopdance[, c(3,18,19,17)]
```

```
##                  artist_name              title decade danceability
## 1                   Kingsmen        Louie Louie  1960s     0.883955
## 2              The Supremes Where Did Our Love Go  1960s     0.874447
```

11

```
## 3                Elvis Presley      Return To Sender  1960s       0.826889
## 4 Sam the Sham & The Pharaohs         Wooly Bully  1960s       0.791505
## 5                Elvis Presley       Wooden Heart  1960s       0.780824
```

Elvis again shows up twice, another clear reflection of why he was (and possibly still is?) such a popular artist.

What are the 5 most danceable songs from the 70's?

```
dat70s <- datNew[which(datNew$decade=='1970s'),]
dat70sTopdance <- head(arrange(dat70s, desc(danceability)), n = 5)
dat70sTopdance[, c(3,18,19,17)]
```

```
##          artist_name                           title decade danceability
## 1               Chic                         Le Freak  1970s       0.887522
## 2           Boney M.                         Ma Baker  1970s       0.883808
## 3 The Rolling Stones                         Miss You  1970s       0.855037
## 4    Michael Jackson Don't Stop 'Til You Get Enough  1970s       0.837228
## 5       Frankie Valli                          Grease  1970s       0.808027
```

No surprises here with Chic coming on top. The Nile Rodgers/Bernard Edwards combo was legendary in bringing a unique sound that shaped and defined disco funk.

What are the 5 most danceable songs from the 80's?

```
dat80s <- datNew[which(datNew$decade=='1980s'),]
dat80sTopdance <- head(arrange(dat80s, desc(danceability)), n = 5)
dat80sTopdance[, c(3,18,19,17)]
```

```
##          artist_name              title decade danceability
## 1 Michael Jackson        Billie Jean  1980s       0.920595
## 2      Lipps, Inc.          Funkytown  1980s       0.894554
## 3       Diana Ross        Upside Down  1980s       0.853097
## 4       The Police Every Breath You Take  1980s       0.820528
## 5   George Michael              Faith  1980s       0.820484
```

The fact that Michael Jackson shows up on the 80s and 70s lists indicates two things: (a) his exceptional talent as an artist and a performer, and (b) the fantastic production teams he worked with for his "Off the Wall" and "Thriller" albums.

What are the 5 most danceable songs from the 90's?

```
dat90s <- datNew[which(datNew$decade=='1990s'),]
dat90sTopdance <- head(arrange(dat90s, desc(danceability)), n = 5)
dat90sTopdance[, c(3,18,19,17)]
```

```
##                    artist_name                              title decade
## 1 Puff Daddy & Faith Evans I'll Be Missing You (Instrumental)  1990s
## 2               Ini Kamoze         Here Comes the Hotstepper  1990s
## 3                   Shaggy                         Boombastic  1990s
## 4               Vanilla Ice                     Ice Ice Baby  1990s
## 5            Janet Jackson                   Together Again  1990s
##   danceability
```

```
## 1      0.941884
## 2      0.883807
## 3      0.866098
## 4      0.860772
## 5      0.850136
```

Finally, what are the 5 most danceable songs from the 2000's?

```
dat00s <- datNew[which(datNew$decade=='2000s'),]
dat00sTopdance <- head(arrange(dat00s, desc(danceability)), n = 5)
dat00sTopdance[, c(3,18,19,17)]
```

```
##             artist_name                title decade danceability
## 1 Justin Timberlake             SexyBack  2000s     0.968045
## 2            Eminem         Just Lose It  2000s     0.960278
## 3     Gwen Stefani        Hollaback Girl  2000s     0.959466
## 4           50 Cent           In Da Club  2000s     0.913415
## 5            Eminem The Real Slim Shady  2000s     0.907038
```

I have just about scratched the surface here and would love to see other analyses with this dataset.