

A2 AEDV João Victor Mendes

Passo 1: A Escolha da Base de Dados

Assistindo as apresentações da A1, me interessei muito pelo tema abordado e pela base abordada nos trabalhos do grupo de Rian e no trabalho individual do Cristiano, a base em questão se trata da base do Spotify disponível no kaggle (<https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks>), a qual consta com diversos dados sobre as músicas presentes no App, como por exemplo o ano em que foram lançadas e a popularidade atual (no caso, da última vez que a base foi atualizada, 2 meses atrás).

Em um destes trabalhos da A1 me recorro de ter sido feito um comentário como “nota-se que a única das músicas no Top 10 que está fora dos padrões das músicas mais populares é Drivers License da Olivia Rodrigo”, e isso ficou por um bom tempo na minha cabeça, pois estudando sobre música sempre ouvimos os grandes artistas falarem sobre que “para inovar, romper barreiras e entrar pra história na música, primeiro você precisa dominar o básico”, então seguindo essa ideia, realmente as músicas que mais fariam sucesso não estariam na média esperada, e este foi o caso de “Drivers License”, sendo até agora disparadamente o maior sucesso do ano.

Pensando nisso, resolvi fazer uma análise levando em conta a atual música mais popular de algumas décadas passadas, resolvi considerar apenas da década de 1950 até a década de 2000 (da década de 40 para trás só haviam músicas natalinas -as quais desconsidereei por serem músicas que sempre são resgatadas em uma época específica, que cabe a outro tipo de análise e não este que estou fazendo sobre músicas que sempre são lembradas- entre as mais populares, e penso que as da década de 2010 para cá ainda são muito recentes para fazermos um julgamento sobre serem “músicas que marcaram a história”), e comparar essas faixas com a média de suas décadas.

Passo 2: A Justificativa dos Elementos Gráficos

Desde o primeiro pensamento envolvendo usar essa base eu já tinha a ideia toda em mente, porém ainda não sabia como fazê-la de maneira programática; foram longas pesquisas e conversas com colegas e monitores até conseguir tirá-la da caixinha de massa cerebral e colocá-la na caixinha preta do console: uma análise de 5 gráficos ao mesmo tempo, levando em conta o quão energética uma música é, o quão acelerado é seu ritmo, o quão “dançável” é a faixa, qual a sua duração e se a música está mais para “feliz” ou “triste”, comparando as faixas escolhidas de cada década com a média dos anos, mas com o desafio de que na base do Spotify essas informações quanto as médias dos anos e as músicas estão contidas em datasets diferentes.

Como o tema eram músicas na plataforma do Spotify, tentei usar um esquema de cores parecido com o deles nas representações gráficas, e também um estilo minimalista como o deles;

Passo 3: Produção de Gráfico com função Explanatória

Aqui irei mostrar o passo-a-passo da construção de meu gráfico;

meu primeiro passo foi limpar as bases, selecionando apenas os dados que eu irei utilizar nas análises, sendo eles na base das músicas o artista e o nome da canção (para poder selecioná-las na análise), a popularidade atual da música, o ano de lançamento, “dançabilidade”, “energia”, “duração” (esta em milissegundos), “tempo” (ritmo) e “valência” (o quão alegre é sua sonoridade).

```
data_clean <- select(data, artists,name,popularity,year,
                     danceability,energy,duration_ms,
                     tempo,valence)

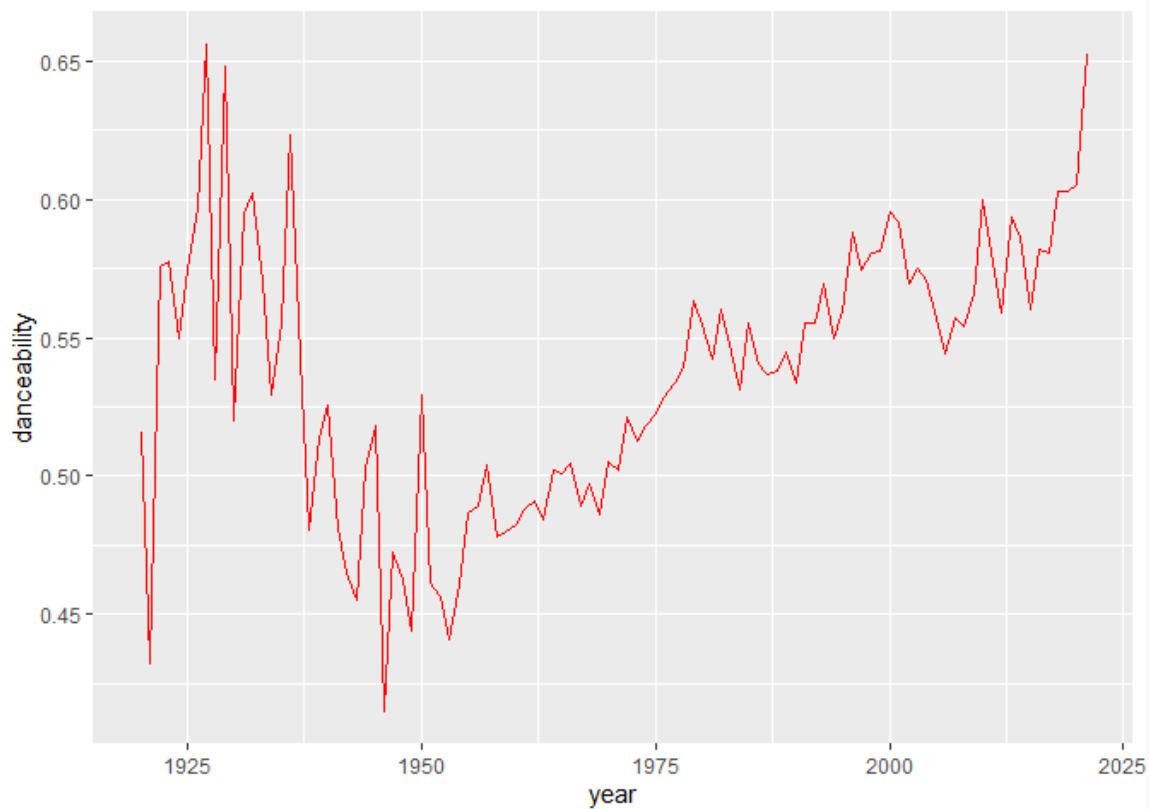
year_clean <- select(by_year, year,danceability,energy,duration_ms,
                    tempo,valence)
```

irei deixar apenas uma delas como exemplo para evitar do trabalho ficar demasiadamente longo

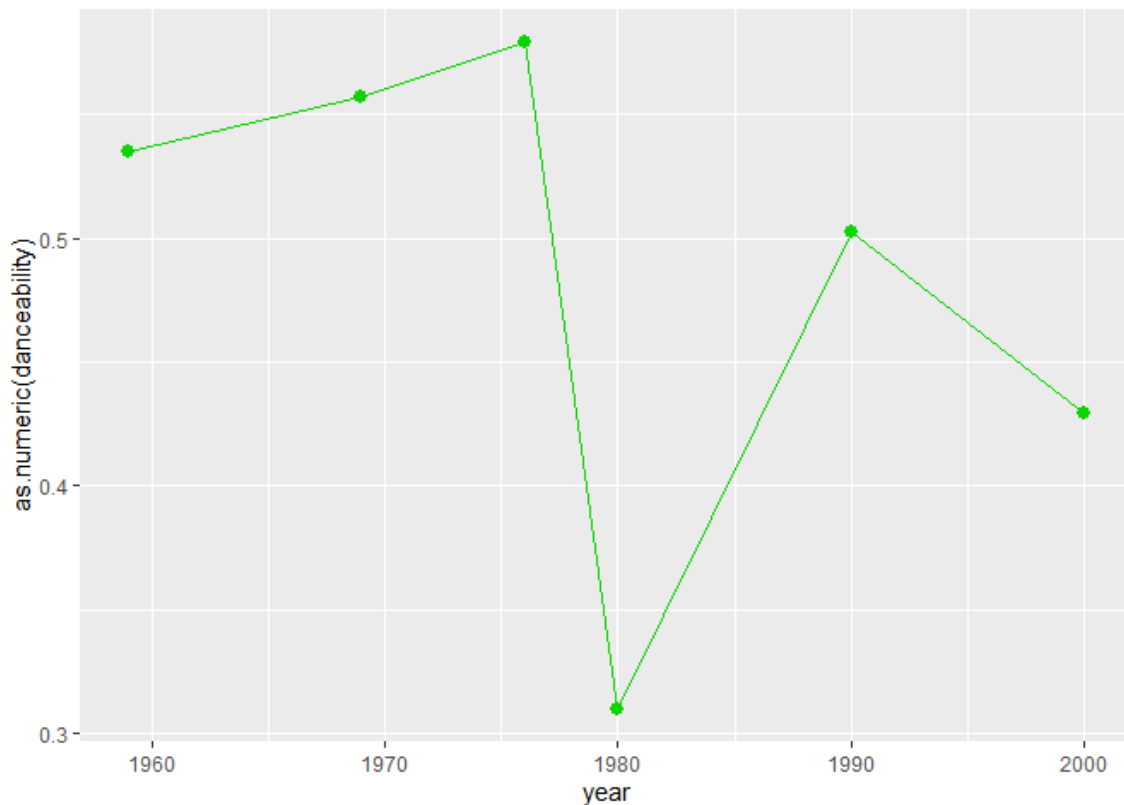
Depois, plotei um gráfico de linha em vermelho com as médias anuais de cada uma das variáveis em que eu iria trabalhar.

Plotei também outro gráfico selecionando apenas as informações referentes as canções identificadas como as mais populares atualmente de cada uma das décadas citadas, respectivamente em cada um dos pontos, da esquerda para a direita, “Johnny B. Goode” da década de 50, “Here Comes The Sun” da década de 60, “Hotel California” da década de 70, “Back In Black” da década de 80, “Thunderstruck” da década de 90, e por último, “Yellow” da década de 2000. Selecionei a cor da linha como sendo uma similar a do Spotify. usei “as.numeric” em todos exceto duração pois a informação não estava sendo lida como um número, então a ordem não estava sendo representada corretamente; em duração usei /60000 para ser representado em minutos e não milissegundos;

```
ggplot()+
geom_line(data=year_clean,aes(x=year, y=danceability, group=1),
          color="#ff0000")
```



```
data_clean %>%  
  filter(name == "Johnny B. Goode" |  
         name == "Here Comes The Sun" |  
         name == "Hotel California" |  
         name == "Back In Black" |  
         name == "Thunderstruck" |  
         name == "Yellow") %>%  
  group_by(year) %>%  
  ggplot(aes(x = year, y = as.numeric(danceability))) +  
  geom_line(color="#0bd400")+  
  geom_point(shape = 21, size = 2, stroke = 1,color="#0bd400",  
            fill = "#0bd400")
```

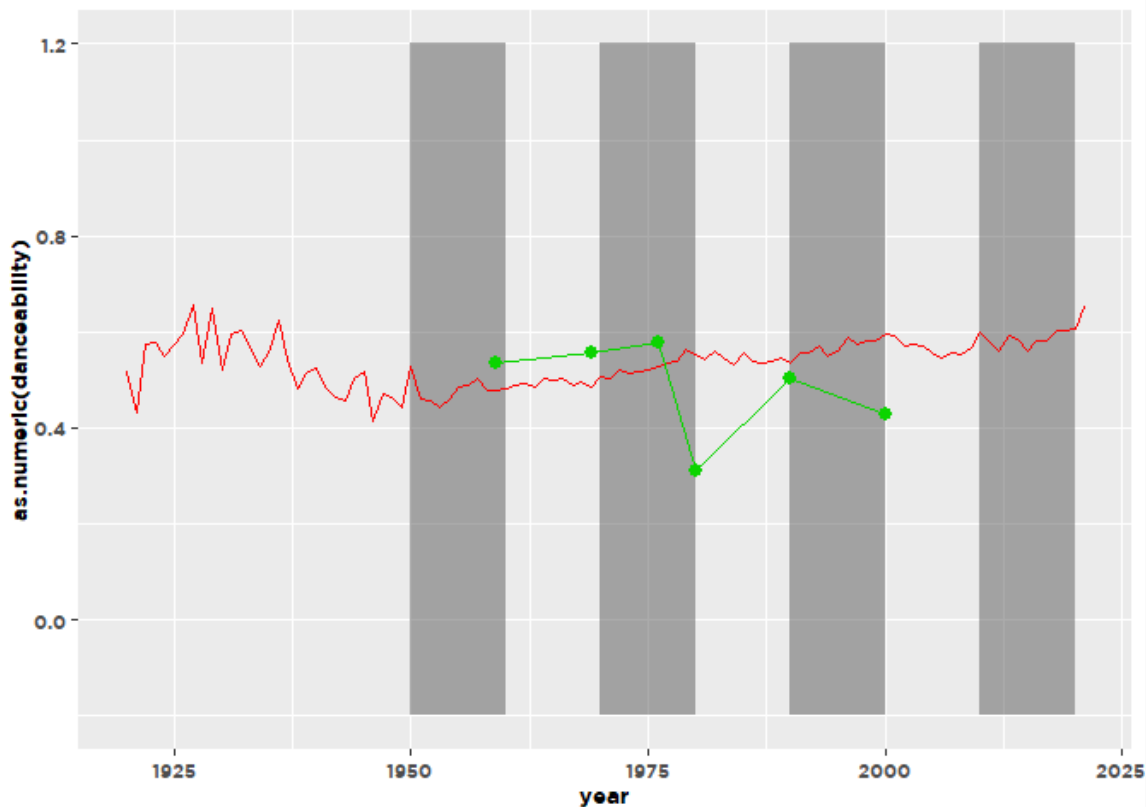


Aqui, depois de mais de 12h seguidas pesquisando sobre como fazer isso, com a ajuda do monitor consegui plotar os dois gráficos de datasets diferentes em um só, usei da função `annotate` para destacar as décadas mudando a cor de fundo do gráfico para tons mais escuros e mais claros de um cinza similar ao presente na plataforma do Spotify. Também adicionei a fonte utilizada no App.

Selecionei apenas da década de 50 pra frente pois as anteriores não aparecerão no gráfico final;

```
name_danc <- data_clean %>%
  filter(name == "Johnny B. Goode" |
         name == "Here Comes The Sun" |
         name == "Hotel California" |
         name == "Back In Black" |
         name == "Thunderstruck" |
         name == "Yellow") %>%
  group_by(year) %>%
  ggplot(aes(x = year, y = as.numeric(danceability)))+
    annotate("rect",xmin = 1950,xmax = 1960,
            ymin = -0.2,ymax = 1.2,alpha=0.5)+
    annotate("rect",xmin = 1970,xmax = 1980,
            ymin = -0.2,ymax = 1.2,alpha=0.5)+
    annotate("rect",xmin = 1990,xmax = 2000,
            ymin = -0.2,ymax = 1.2,alpha=0.5)+
    annotate("rect",xmin = 2010,xmax = 2020,
            ymin = -0.2,ymax = 1.2,alpha=0.5)+
  geom_line(data=year_clean,aes(x=year, y=danceability, group=1),
           color="#ff0000") + geom_line(color="#0bd400")+
  geom_point(shape = 21, size = 2, stroke = 1,color="#0bd400",fill="#0bd400")+
  theme(text=element_text(size=10, family="Gotham"))
```

name_danc



Aqui criei uma caixa com os nomes das canções analisadas para ser adicionado no final na hora de juntar todos os plots;

```
c<-ggplot() + annotate("text", x =0, y = 0, label = "
1950-Johnny B. Goode
1960-Here Comes The Sun
1970-Hotel California
1980-Back In Black
1990-Thunderstruck
2000-Yellow",color="white")+
labs(x =NULL, y=NULL)+guides(color = FALSE)+
theme(panel.grid.major = element_line(color = "#000000", size = .25),
panel.grid.minor = element_line(color = "#000000", size = .15),
panel.background = element_rect(fill = "#000000"),
plot.background = element_rect(fill = "#000000"))+
theme(legend.position = "none")+
theme(axis.text.x = element_text(color="#000000"),
axis.title.x = element_text(color="#000000"),
axis.text.y = element_text(color="#000000"),
axis.title.y = element_text(color="#000000"))
```

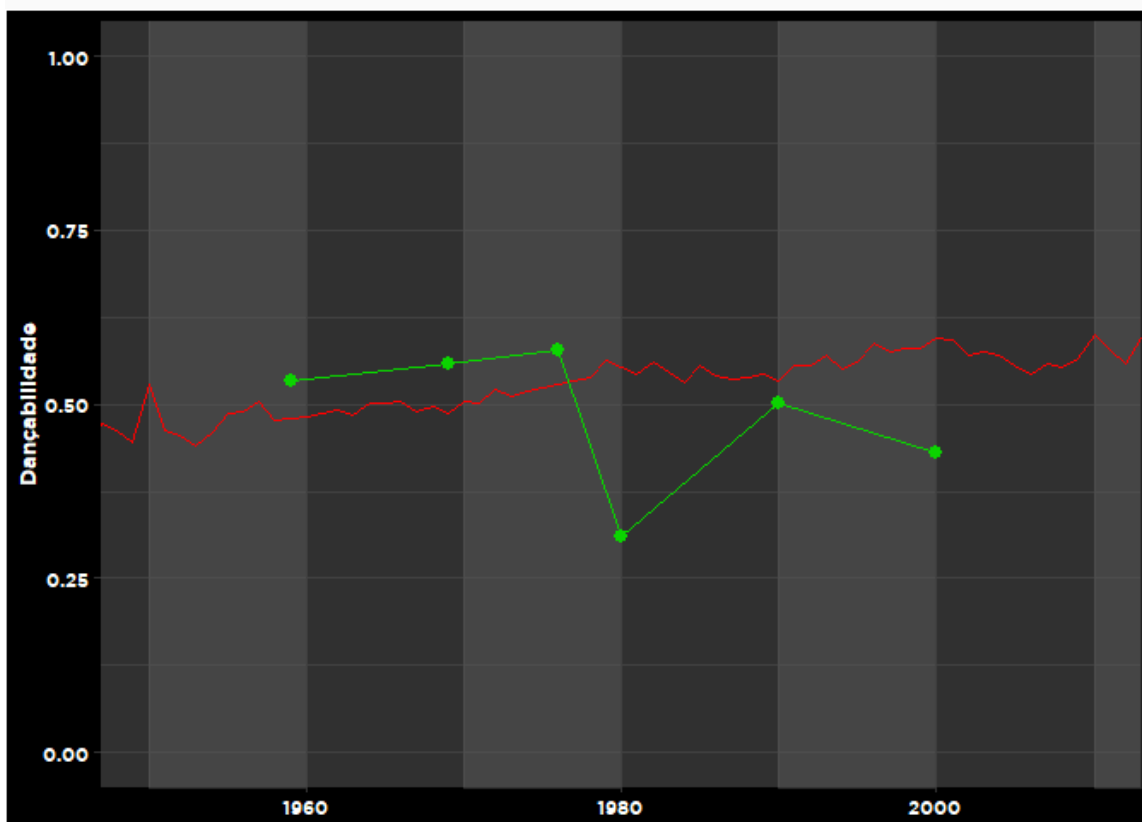
Criei outro objeto com um nome mais simples para juntar todos no final, alternei os nomes dos eixos e escondi eixos que não queria que aparecessem nos plots, defini os limites de 'x' e 'y' para melhor visualização e alternei as cores de fundo para um tema mais parecido com o do Spotify.

```

an <- name_danc + coord_cartesian(xlim = c(1950, 2010), ylim = c(0, 1))+
  labs(x =NULL, y="Dançabilidade")+guides(color = FALSE)+
  theme(panel.grid.major = element_line(color = "#424242", size = .25),
        panel.grid.minor = element_line(color = "#424242", size = .15),
        panel.background = element_rect(fill = "#303030"),
        plot.background = element_rect(fill = "#000000"))+
  theme(legend.position = "none")+
  theme(axis.text.x = element_text(color="#ffffff"),
        axis.title.x = element_text(color="#ffffff"),
        axis.text.y = element_text(color="#ffffff"),
        axis.title.y = element_text(color="#ffffff"))

```

an

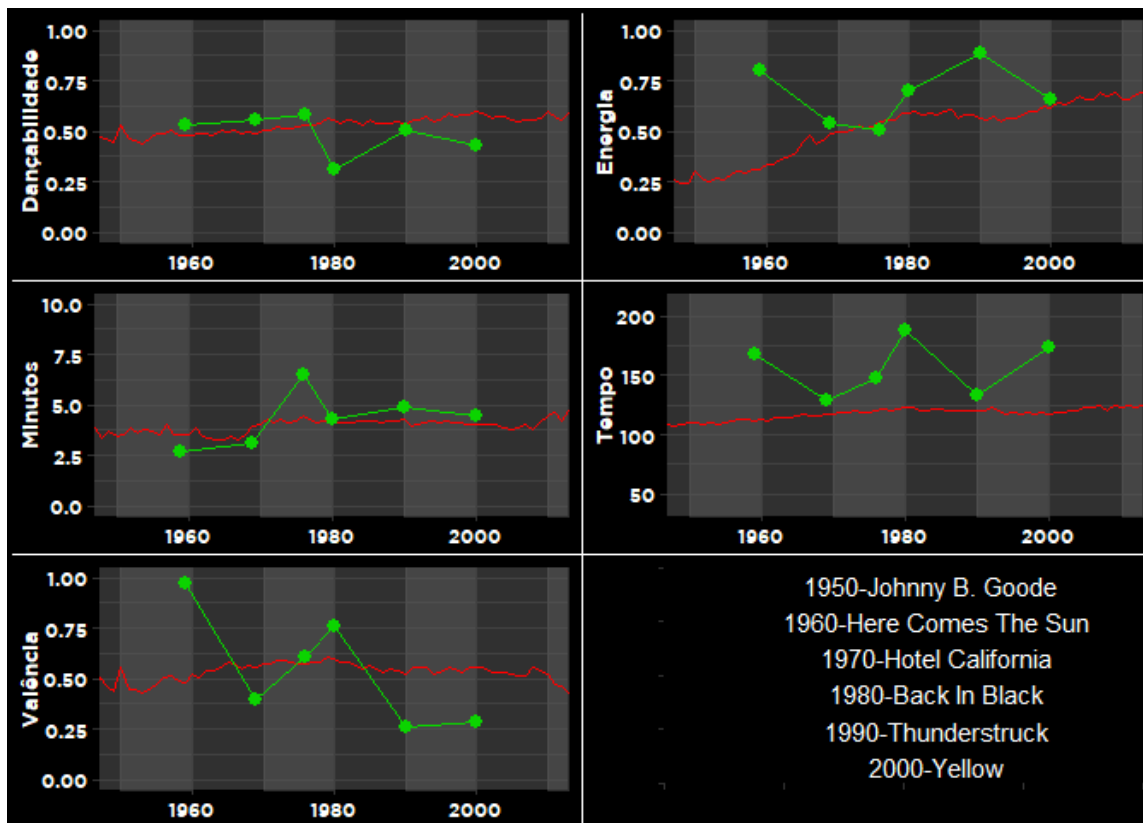


Aqui juntei todos os gráficos anteriormente feitos;

```

plot_grid(an,bn,cn,dn,en,c, ncol = 2, nrow = 3, color="#000000", fill="#000000")
grid.rect(width = 1, height = 1, gp = gpar(lwd = 5, col = "#000000", fill = NA))

```



Passo 4: Comentários e Conclusão

Terminado o gráfico, chega-se a hora de fazer as análises.

Começando pelo primeiro gráfico, o que avalia o quesito “dançabilidade”, vemos que a maior parte das faixas analisadas ficou próxima a média do ano, com exceção da quarta, “*Back in Black*”, que ficou bastante abaixo da média de “dançabilidade”, o que é esperado para uma faixa de metal, então provavelmente nesse quesito as faixas que mais fazem sucesso precisam sim manter uma certa média de “dançabilidade” e não tender para nenhum extremo para sobreviverem as mazelas do tempo, o que também é corroborado pelos dados temporais, que mostram que por mais que as médias anuais tenham subido, foi uma subida bem sutil;

Na segunda categoria analisada, “energia”, vemos que muitas estão bem acima da média dos anos em que foram lançadas e as mais baixas estão exatamente na média do ano, o que indica uma tendência de que as músicas das décadas passadas que romperam o padrão vindo com sonoridades mais energéticas mantêm um nível de sucesso mais alto hoje em dia, o que faz sentido quando olhamos pra linha vermelha que mostra como a média vem crescendo ao longo dos anos;

Na terceira categoria, “minutos”, nota-se que uma duração padrão sempre foi importante para as músicas fazerem sucesso, a média subiu suavemente nas últimas décadas, mas mesmo para o padrão atual a única que não está dentro da média é “*Hotel California*”, estando cerca de 1 minuto e 15 segundos acima da média atual, sendo esta menos de 40 segundos superior a do ano em que a música foi lançada;

Na quarta, “tempo”, temos um padrão ainda mais evidente do que no tópico “energia”, todos os sucessos analisados tem um ritmo muito mais rápido do que a média de seus anos, e nessas médias anuais também notamos uma tendência sutil de crescimento;

No último gráfico, “valência”, não se nota nenhum padrão nos grandes sucessos, alguns estando próximos do extremo de músicas tristes, outros estando próximos do extremo de músicas alegres; nota-se que dos anos 50 para cá tivemos um crescimento na média de “alegria” das músicas por volta de 1953 até 1980, uma leve decaída de 1980 até 1990, altos e baixos mas mantendo a mesma média de 1990 até os anos 2000, um grande caimento de 2000 até por volta de 2007, um súbito aumento no ano seguinte e por fim dali pra frente uma enorme queda na valência das canções, o que indica que as canções de um ano serem mais felizes ou tristes não seguem nenhuma tendência típica e são altamente influenciadas por diversas questões e ocorrências ao longo dos anos;

Em conclusão, músicas mais energética e de ritmo acelerado tem maior a tendência de fazer sucesso ao longo do tempo (pelo menos para os dados atuais) e músicas que se mantém perto da média anual de duração e dançabilidade também entram nesse contexto; quanto a música ser feliz ou triste, provavelmente depende da ocasião e da época em que ela é produzida para ter o sucesso inicial no ano de lançamento, depois disso, se fazendo conhecida não se notam tendências sobre músicas mais tristes ou mais felizes serem mais ouvidas hoje em dia;