SURVIVAL ANALYSIS OF THE BRAZILIAN SPOTIFY RANKING:

DIFFERENCES BETWEEN NATIONAL AND INTERNATIONAL ARTISTS

SILVIO NUNES AUGUSTO JR.[1]

VINÍCIUS BASSETO FÉLIX.[2]

Resumo

Esse artigo utilizou dados públicos fornecidos pela empresa de *streaming Spotify* com objetivo de analisar a permanência de artistas no *ranking* brasileiro ao longo do ano de 2017. Um total de 1084 músicas foram acessadas utilizando o pacote *spotifycharts*, que permite a extração das 200 mais ouvidas, diariamente, em todos os 53 países nos quais os serviços de *streaming* da *Spotify* estão presentes. Para a manipulação, o tratamento e a análise dos dados foram utilizados os pacotes *lubridate* (GROLEMUND, 2011), *dplyr* (WICKHAM e FRANCOIS, 2015), *forcats* (WICKHAM, 2017) e *ggplot2* (WICKHAM, 2016). Os resultados mostram maior proporção de artistas nacionais em todos os meses do ano de 2017. Em vista de que um artista pode ter mais de uma música fazendo sucesso ao mesmo tempo, as análises mostram artistas com mais de 20 músicas entre as mais ouvidas. A curva de sobrevivência das músicas internacionais diminui mais rápido do que a curva de sobrevivências das músicas nacionais em ambos os *rankings* analisados, Top 200 e Top 10. Por fim, discute-se a utilização da técnica de análise de sobrevivência para interpretar a permanência de músicas no *ranking* das mais ouvidas.

Palavras-chave: *Spotify*, Música, Análise de Sobrevivência

Abstract

This article used public data provided by the streaming company Spotify with the objective of analyzing the permanence of artists in the Brazilian ranking throughout the year of 2017. A total of 1084 songs were accessed using the spotifycharts package, which allows the extraction of the daily Top 200 songs of any of the 53 countries that Spotify streaming services are present. For manipulation, treatment and analysis of the data, the packages lubridate (GROLEMUND, 2011), dplyr (WICKHAM; FRANCOIS, 2015), forcats (WICKHAM, 2017) and ggplot2 (WICKHAM, 2016) were used. The results show a greater proportion of

[1] Universidade de São Paulo (USP). Emails:silvio.augusto@usp.br ou silvioaugustojr@gmail.com

[2] Universidade Estadual de Maringá (UEM). Email: vinicius1b7f@gmail.com

national artists in all months of the year 2017. Given that an artist can have more than one song at the same time, the analysis shows artists with more than 20 songs among the most listened. On top of that, International songs' survival curve decreases faster than the national's survival curve on both Top 200 and Top 10 rankings. Finally, the use of the survival analysis technique to interpret the permanence of songs in the ranking of the most heard is discussed.

Keywords: Spotify, Music, Survival Analysis.

Introduction

Over the past few years, streaming services became so popular that streaming and downloadable films and songs are now ahead of DVDs and Blu-ray discs at UK market[3]. When it comes specifically about music streaming service, research shows that the consumption of digital music leads to a loss in the perceived sense of ownership (BARTMANSKI; WOODWARD, 2015), suggesting that consumers are experiencing a period of post-ownership economy (SINCLAIR; TINSON, 2017). Compared to other companies, Spotify and Apple Music are leading the competition in this market[4]. However, apart from all concerns about negative effects of streaming services on sales, Aguiar and Waldfogel (2015) were able to show that losses are out weighted by gains.

While research has been published about user behavior and music consumption (ZHANG; KREITZ; ISAKSSON; UBILLOS; URDANETA; POUWELSE; EPEMA, 2013; GREENBERG; RENTFROW, 2017), and packages has been developed to understand sentiment tendencies in artists or even in playlists created by users (see Sentify[5]), no research about the style of the songs or about the effect of an artist nationality among different countries where streaming services is offered was found.

By using the *spotifycharts* package, which allows data extraction of the 200 top ranking songs in 53 countries, we tracked the available streaming information from January to December of 2017. Our focus for this paper was to understand survival rates of songs and artists in Brazil - or by other means, how long a song or an artist can hold this rank status.

[3] **Information about UK streaming market**: https://www.theguardian.com/media/2017/jan/05/film-and-tv-streaming-and-downloads-overtake-dvd-sales-for-first-time-netflix-amazon-uk

[4] **Information about Spotify and Apple music**: https://www.statista.com/chart/5152/music-streaming-subscribers/

[5] **Analyze musical sentiment for your favorite artists and Spotify playlists**: http://www.rcharlie.net/sentify/

Both international and national songs and artists were considered for analysis, and visualization shows a tendency for certain types of music, such as the Brazilian rhythm Sertanejo and the prevalence of artists such as Ed Sheeran with more than 20 songs in the ranking.

Objective

The main objective of this paper is to compare survival rates of national and international artists and their songs on the Brazilian Spotify ranking. The proportion of national and international artists is also considered. On top of that, we describe the application of the survival analysis in order to identify how prevalent those songs are.

Methods and Materials

In this paper, the data source used for the analysis is *Spotify Charts*, a platform which updates a daily-ranking of the top 200 most listened songs on Spotify. The data was downloaded through the *spotifycharts* package[6]. We used a list of packages for data manipulation, exploration and visualization, such as *lubridate* (GROLEMUND, 2011), *dplyr* (WICKHAM; FRANCOIS, 2015), *forcats* (WICKHAM, 2017) and *ggplot2* (WICKHAM, 2016). Since we intended to analyze differences between national and international artists, we manually added a binary variable classification for all artists – national for Brazilian, and International for all foreigners.

To perform survival analysis of songs comparing both national and international origin of artists, we used the Kaplan-Meier estimator (KAPLAN; MEYER, 1958) with the *survival* package (THERNEAU, 2016). This estimator is an adaptation of the empirical survival function which in absence of constraints is defined as:

$$S(t) = \frac{n_t}{n},$$

which n_t is the number of observation that do not have fail until the time t, and n is the total number of observations in the study. The graphic estimator, $\hat{S}(t)$, is a ladder function which the steps in the observed fail times have $\frac{1}{n}$ size. If there is a draw in any t time, the size of the step is multiplied by the total account of draws. The formula of the Kaplan-Meier estimator is:

[6] **More information about *spotifycharts* package is available at github**: https://github.com/mikkelkrogsholm/spotifycharts.

$$\hat{S}(t) = \prod_{j:t_j<t}\left(1 - \frac{d_j}{n_j}\right),$$

which $t_1 < t_2 < \cdots < t_k$, are the k distinct and ordinate fail time, d_j the number of observed fails in t_j, $j = 1, \ldots, k$, and n_j the number of individuals under risk at the t_j moment, which means the number of individuals that do not had failed and were not censored until the precisely moment before t_j.

After the Kaplan-Meier estimator, it is necessary to use a technique to compare both national and international curves, so a log-rank test was applied (MANTEL, 1966). The number of observed and expected events are calculated by each group, and then add to the general calculation. At any time $t_j, j = 1, \ldots, k,$ which occurs $d_{.j}$ fails and $n_{.j}$ individuals under risk until the precisely moment before t_j to the total sample. Analogously, to both d_{ij} and n_{ij} groups, $i = 1{,}2,$ conditioned to fails, we have that d_{2j} follows a hypergeometric distribution (BERKOPEC, 2007).

$$\frac{\binom{n_{1j}}{d_{ij}}\binom{n_{2j}}{d_{2j}}}{\binom{n_{.j}}{d_{.j}}}.$$

The statistics is given by:

$$T = \left[\sum_{j=1}^{k}(d_{2j} - w_{2j})\right]^2 \sum_{j=1}^{k} V_{2j},$$

which w_{2j} and V_{2j} are the respective average and variance of d_{2j}. For further information and detailed equations see Giolo and Colosimo (2006).

To visualize the results from both the estimated curve and the statistical test, the graphical features from *survminer* package were used (KASSAMBARA, 2017).

Results and Discussion

The dataset has 71790 observations, with the following variables: 1) position of the song on the Top 200 ranking; 2) songs' names; 3) artists' names; 4) number of streams; 5) ranking date; 6) artists' gender; 7) artists' nationality.

The proportion of artists and songs was analyzed according to their origin, national or international, throughout the year of 2017. The dataset contains a total of 293 unique artists, among which 131 (45%) are national and 162 (55%) are international. There were a total of 1084 unique songs, among which 456 (42%) are national, while 628 (58%) are international.

Since there are a considerable proportion of artists (92%) and songs (82%) that are repeated on the ranking along the year, these repetitions were not considered in the Figure 1.

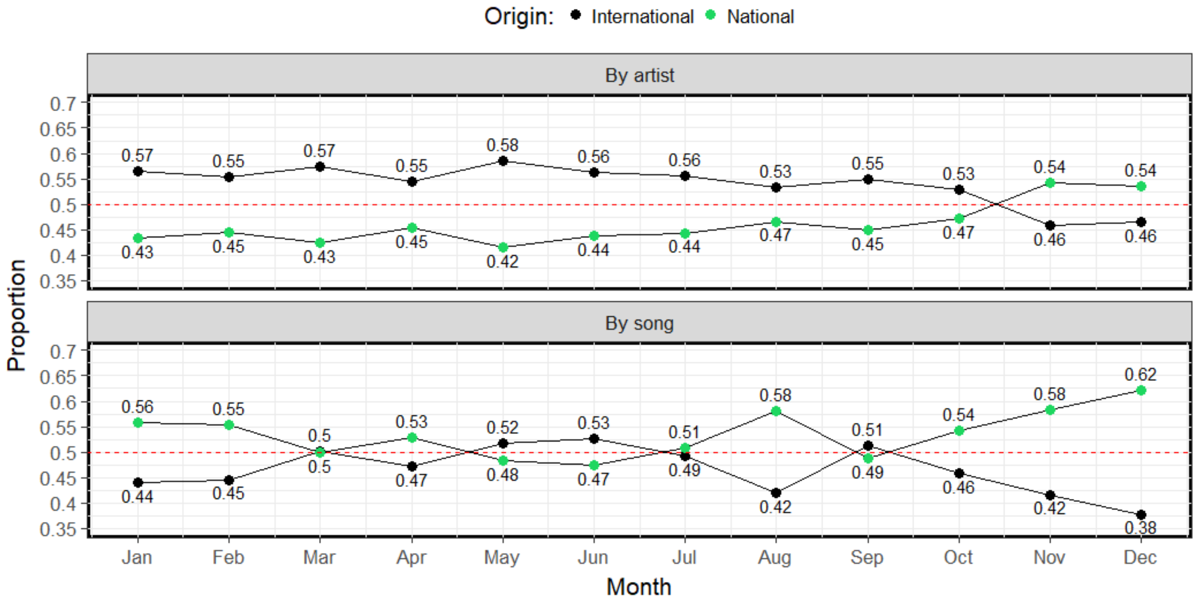

Figure 1: Proportion of national and international artists and songs without considering repeated songs.

The reason why repetition was not considered in Figure 1 is to disclose the real proportion of national and international artists and songs. Even though the proportion of international artists (57%) is greater than national artists (43%) in January, the proportion of Brazilian songs (56%) is greater than international songs (44%).

This pattern suggests that in months where the proportion of artists does not match the proportion of songs, one or more artists have more than one song at same time on the ranking. For instance, while some Brazilian artists have more than 10 songs at the same time, the majority of the international artists has less than 4 songs. The only international artist with 5 songs at the same time on the ranking is the group *The Chainsmokers*.

Considering that repetition seems to play a role on the ranking, the Figure 2 considered repetition in the proportion.

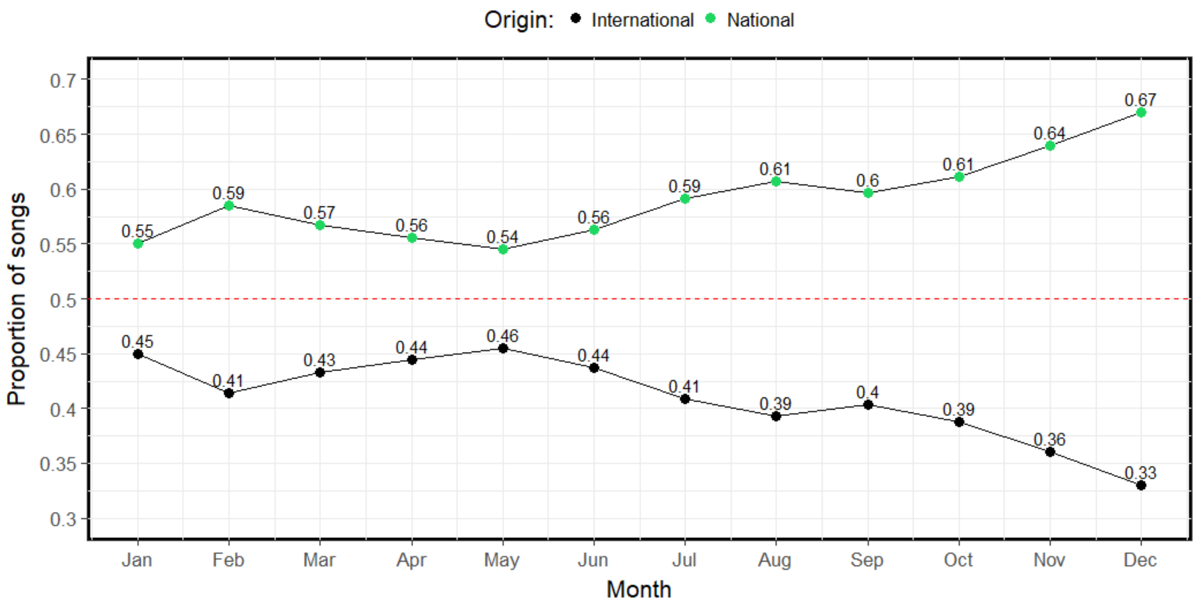

Figure 2: Proportion of artists by origin and month considering repetition.

The proportion of national songs is greater than 50% from January to July, and greater than 60% from August to December. The increasing of this rate in the end of the year might be explained due to Christmas and the New Year’s holiday. The increasing of 0.04 in February might be due to Carnival holiday. The results suggest that even though there are months with more international artists than national artists, Brazilian artists have a greater number of songs along the year.

The Figure 3 shows the total number of streams by the number of tracks in the Spotify Brazilian ranking:

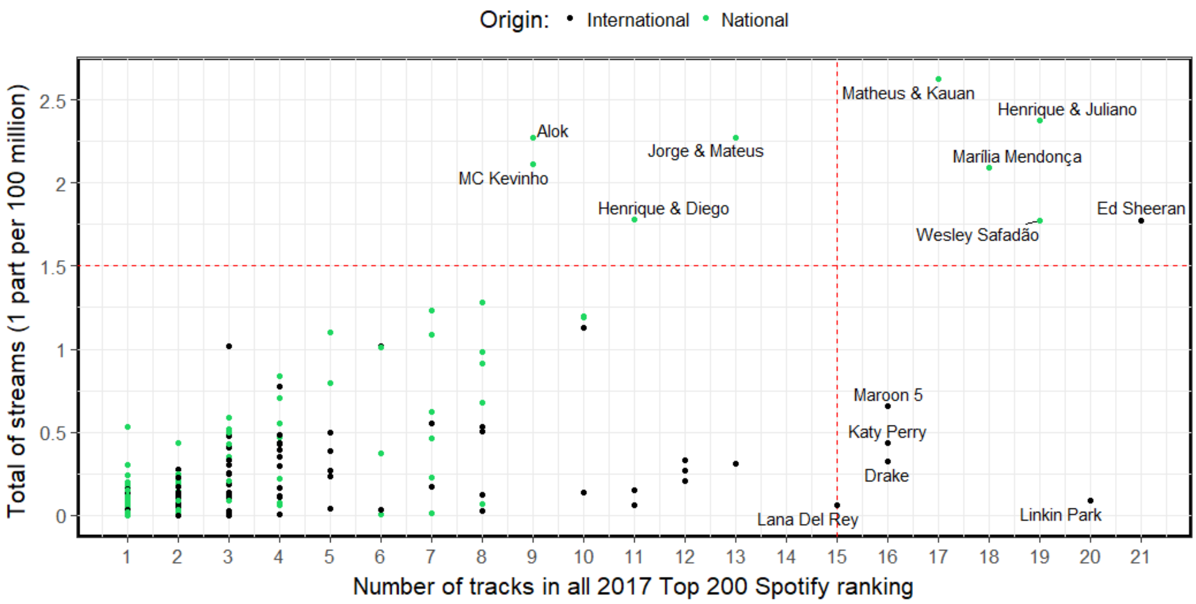

Figure 3: Total of streams (1 part per 100 million) x number of tracks in all 2017 Top 200 Spotify ranking.

The graph discloses that there is a concentration of songs with less than 100 million streamings and less than 8 songs per artist on the ranking. It also identifies few artists that can be considered outliers, since they have the majority of both streamings and songs, such as Ed Sheeran with more than 170 million streamings and 21 songs. Other interesting examples are Linkin Park band group with 20 songs and less than 100 million streamings, maybe due to the death of their singer Chester Bennington in 2017, also that there are many duos from the Sertanejo genre such as Jorge & Mateus, Matheus & Kauan e Henrique e Juliano, all of them with more than 200 million streamings. The total streaming discloses that many artists hold more than one song on the ranking along the year.

In order to explore the survival rate of these songs, data was analyzed by counting how many songs appeared just for one day on the both Top 200 and Top 10 rankings. As summarized in Figure 4, a substantial number of songs cannot hold a top position for more than one day. From the total of 1084 songs on the Top 200 ranking, 199 (64 national, 135 international) of then appeared just for one day, representing 18% of the population. From the total of 86 songs on the filtered Top 10 ranking, 14 (4 national, 10 international) of then appeared just for one day, representing 16% of the sample. Besides the fact that some songs stay on the Top 200 ranking along the year, this number decreases considerably from day two, and 43% of songs do not appear on the Top 200 ranking after day 9.

The comparison between the Top 200 and Top 10 rankings in Figure 4 shows a pattern between the number of appearances.

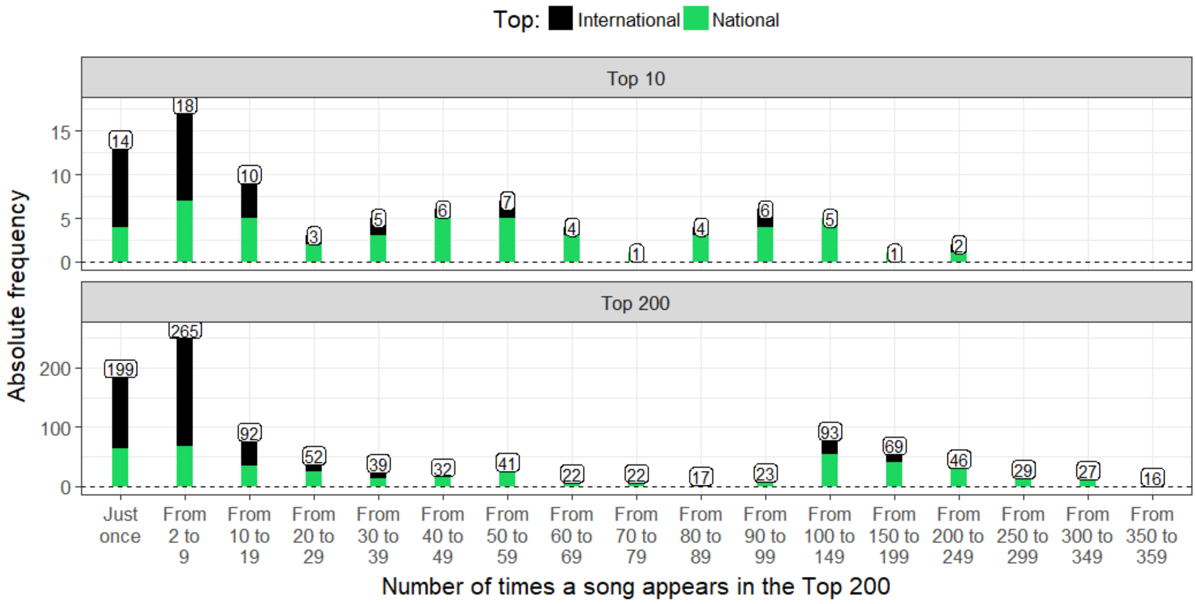

Figure 4: Absolute frequency of the number of times a song appears on the Top 10 and Top 200 rankings.

Summarizing Figures 1 to 4, these findings suggest that while a few artists are consistent and able to hold the first positions of the ranking for 10 days or more (a few of them for all the year round), the majority of artists and songs are inconsistent, sometimes appearing just for one day along the year. That is the main reason why the ranking was filtered to account only the first 10 songs from each day, allowing us to compare both survival curves.

While exploratory findings suggest that the likelihood of a song surviving after the tenth day decreases considerably, survival analysis suggest that there is a statistically significance difference between national and international artists survival curves. These results are represented in the Figure 5:

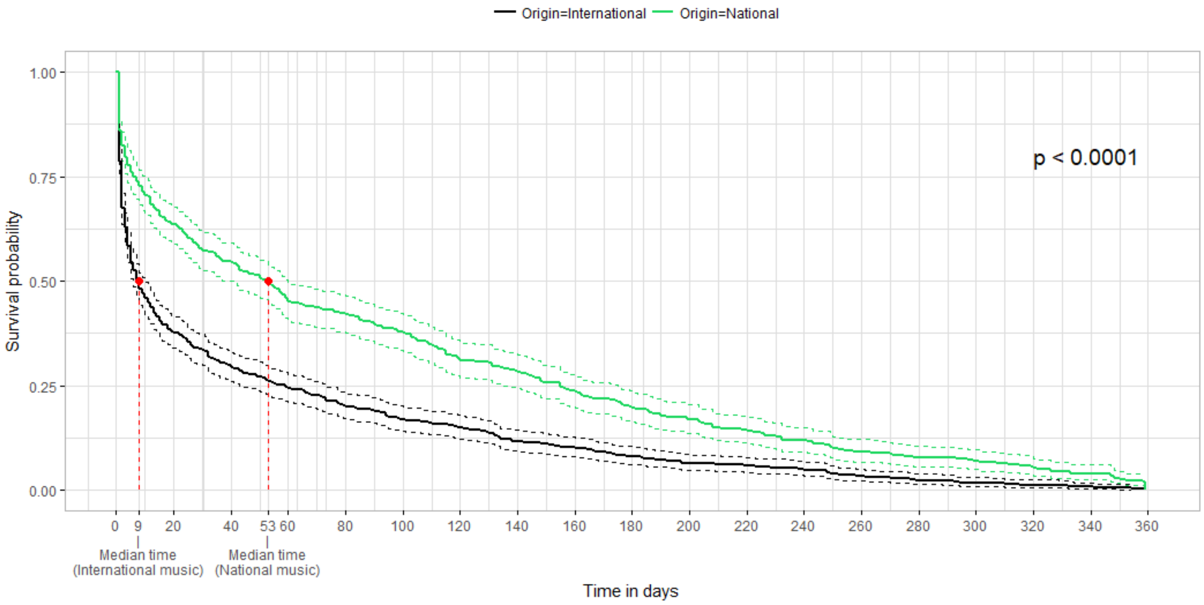

Figure 5: Top 200 survival time curves using Kaplan-Meier estimator to compare origin of artists and songs.

Survival rate in days are represented by two main curves in the graph, one green for national songs and one black for international songs. The dashed lines indicate their respective 95% confidence interval and median points are highlighted in red. The p-value was rejected in the null hypothesis test, suggesting that there is evidence of differences between the time of survival curves. International songs decrease to a 50% likelihood chance of survival in day 9, while national songs reach this point only in day 53.

In an unexpected way, Top 10 ranking shows similar results in Figure 6:

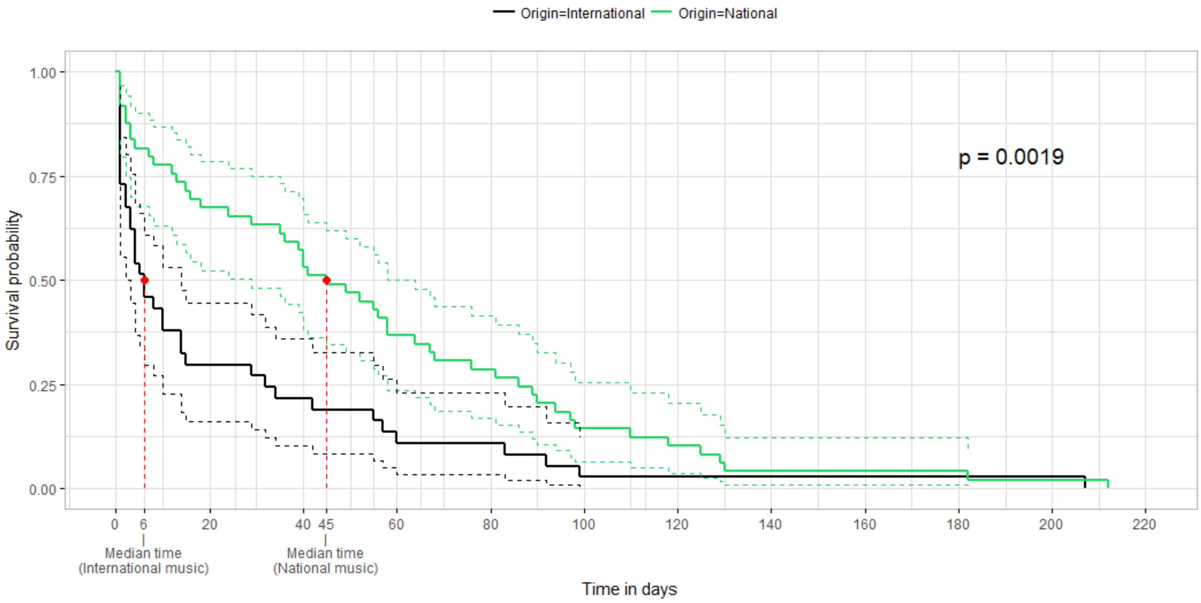

Figure 6: Top 10 survival time curves using Kaplan-Meier estimator to compare origin of artists and songs.

The Figure 6 also shows two main curves, dashed lines, median points and a significant p-value, therefore the description and the interpretation of Figure 6 are the same as described above for Figure 5. International songs decrease to a 50% likelihood chance of survival in day 6 and national songs reach this point only in day 45. These median points are smaller compared to median points in Figure 5. On top of that, the 95% confidence interval are bigger probably due to the reduced number of observations, since we filtered only songs that were able to hold a Top 10 position.

Conclusion

A total of 1084 unique songs (national = 456 songs) were collected from the Spotify Brazilian Top 200 ranking using spotifycharts package. The exploratory analysis suggests that 43% of songs do not stay more than 9 days on the Top 200 ranking (national = 131, international = 333), 37% of songs do not stay more than 9 days on the Top 10 ranking (national = 11, international = 21).

To compare national and international survival rates of artists and their songs, the Kaplan-Meier estimator was used, since it is unbiased for larger samples, converges asymptotically to a gaussian process and is the maximum likelihood estimator for $S(t)$ (GIOLO; COLOSIMO, 2006).

The main finding of this analysis suggests that the international songs survival curve decreases faster than the national's survival curve on both Top 200 and Top 10 rankings.

Even though survival analysis is often applied to health research, this paper showed that it can be used to understand a cultural issue such as music consumption, allowing comparisons between the survival rates of songs according to national and international origin of an artist.

References

AGUIAR, Luis; WALDFOGEL, Joel. **Streaming reaches flood stage: Does spotify stimulate or depress music sales?**. National Bureau of Economic Research, 2015.

BARTMANSKI, Dominik; WOODWARD, Ian. The vinyl: The analogue medium in the age of digital reproduction. **Journal of consumer culture**, v. 15, n. 1, p. 3-27, 2015.

BERKOPEC, Aleš. **HyperQuick algorithm for discrete hypergeometric distribution**. Journal of Discrete Algorithms, v. 5, n. 2, p. 341-347, 2007.

GREENBERG, David M.; RENTFROW, Peter J. Music and big data: a new frontier. **Current Opinion in Behavioral Sciences**, v. 18, p. 50-56, 2017.

GIOLO, Suely Ruiz; COLOSIMO, Enrico Antônio. **Análise de sobrevivência aplicada**. Edgard Blucher, 2006.

GROLEMUND, Garrett et al. **Dates and times made easy with lubridate**. Journal of Statistical Software, v. 40, n. 3, p. 1-25, 2011.

KAPLAN, Edward L.; MEIER, Paul. **Nonparametric estimation from incomplete observations**. Journal of the American statistical association, v. 53, n. 282, p. 457-481, 1958.

KASSAMBARA, Alboukadel et al. **survminer: Drawing Survival Curves using'ggplot2'**. R package version 0.3, v. 1, 2017.

MANTEL, Nathan. **Evaluation of survival data and two new rank order statistics arising in its consideration**. Cancer Chemother. Rep., v. 50, p. 163-170, 1966.

SINCLAIR, Gary; TINSON, Julie. Psychological ownership and music streaming consumption. **Journal of Business Research**, v. 71, p. 1-9, 2017.

THERNEAU, T. **A Package for Survival Analysis in S**. version 2.38. 2015. Reference Source, 2016.

WICKHAM, Hadley; FRANCOIS, Romain. dplyr: A grammar of data manipulation. R package version 0.4, v. 3, 2015.

WICKHAM, Hadley. **ggplot2: elegant graphics for data analysis**. Springer, 2016.

WICKHAM, Hadley. **forcats: Tools for Working with Categorical Variables (Factors).** R package version 0.2. 0. URL: https://CRAN. R-project. org/package= forcats, 2017.

ZHANG, Boxun et al. Understanding user behavior in spotify. In: **INFOCOM, 2013 Proceedings IEEE**. IEEE, 2013. p. 220-224.

Attachment

```
library(spotifycharts);library(lubridate);library(dplyr);library(purrr);library(forcats);
library(survival);library(survminer);library(stringr)
import_spotify<-function(month,year){
  if(is.numeric(month) == F) {
    stop ("O mês deve ser informado pelo seu numeral")}
  if(nchar(year) != 4) {
    stop ("O ano deve ser informado com 4 dígitos")}
    chart_daily() %>%     filter(month(days) == month & year(days) == year) ->days
    dias<-rev(days$days);    df<-data.frame();    n<-length(dias)
    for( i in 1:n){
    cat(100*round(i/n,2),"% ",sep = "")
    aux<-chart_top200_daily(region = "br", days = dias[i])
    aux$date<-rep(dias[i],200);    df<-rbind(df,aux)
  }
  name<-paste0("top200_",month.abb[month],"_",year,".Rds");  saveRDS(df,name)
  }
1:12 %>% map(~ import_spotify(.x,year=2017))
green_spotify<-rgb(maxColorValue = 255,30,215,96)
df<- list.files(pattern = "top200_")%>%
            map(readRDS) %>%    rbind_list() df_artists %>% filter(artist!="" ) %>%
  mutate(genre_artist = fct_recode(genre_artist, Male = "M", Female = "F",
Mixed = "Mi")) %>%
  mutate(tamanho = fct_recode(tamanho, Dupla = "D",Grupo = "G",Individual = "I")) %>%
  mutate(nacionalidade = fct_recode(nacionalidade,National = "N",
International = "I"))->df_artists
df<- df %>%   left_join(df_artists)
df %>%   mutate(date = month(lubridate::ymd(date))) %>%
  group_by(date,nacionalidade) %>%
  summarise(n=n()) %>%   ungroup() %>%
  group_by(date) %>%   mutate(N= sum(n),p=n/N) %>%
  ggplot(aes(x= date, y=p,group = nacionalidade))+
  geom_text(aes(label = round(p,2) ),vjust= -.5)+  geom_line()+
  geom_point(size=3,aes(col=nacionalidade))+
  geom_hline(linetype="dashed",yintercept = .5)+
  scale_x_continuous(breaks = 1:12,labels=month.abb)+
  labs(x = "Month",y="Proportion of artists",col="Origin:")+
  theme_bw(base_size = 16)+  theme(legend.position = "top")+
  scale_y_continuous(breaks = seq(0,1,.1),labels = seq(0,1,.1),limits=c(0.3,.7))+
  scale_color_manual(values=c("black",green_spotify))+
  theme(panel.background =element_rect(colour = "black",size = 2),
    panel.grid = element_line(colour="grey79"))
df %>%  group_by(artist,nacionalidade) %>%
  summarise(n_track=n_distinct(track.name),soma=sum(streams)) ->df_track
streams<- 1.5*10^8; ntracks <- 15
df_track %>%   ggplot(aes(n_track,soma)) +
  geom_vline(xintercept = ntracks,linetype = "dashed",col=green_spotify)+
  geom_hline(yintercept = streams,linetype = "dashed",col=green_spotify)+
  ggrepel::geom_text_repel(data=df_track %>%
                          filter(n_track >= ntracks | soma >= streams),aes(label = artist))+
  geom_point(aes(col=nacionalidade))+
  labs(x = "Number of tracks in all 2017 Top 200 Spotify rank",
       y="Total of streams (1 part per 100 million) ",col="Origin:")+
  theme_bw(base_size = 16)+  theme(legend.position = "top")+
  scale_x_continuous(breaks = 1:21,labels=1:21)+
  scale_y_continuous(breaks = seq(0,3*10^8, by =.5*10^8),
                 labels = seq(0,3*10^8, by =.5*10^8)/10^8 )+
  scale_color_manual(values=c("black",green_spotify))+
  theme(panel.grid = element_line(colour="grey79"),
      panel.background =element_rect(colour = "black",size = 2))
top<-10
df_all %>%  filter(position<=top) %>%
  group_by(track.name,nacionalidade,url) %>%   summarise(Time = n() )%>%
  rename(Origin = nacionalidade) ->df_survival
df_survival$Time<-Surv(df_survival$Time)
fit <- survfit(Time ~ Origin,
              data = df_survival, conf.type = "log-log")
ggsurv <- ggsurvplot(fit,data = df_survival, pval = TRUE, pval.coord = c(180,0.8),
  conf.int = TRUE, xlim = c(0,220), xlab = "Time in days", break.time.by = 100,
  ggtheme = theme_light(), conf.int.style = "step",  surv.median.line = "none"
)
seq<-seq(0,220,by=20); mi <- 6; mn <- 45
ggsurv+
  geom_segment(xend = c(mi),yend = 0.5,y=0,x=c(mi),col="red",linetype="dashed")+
  geom_segment(xend = c(mn),yend = 0.5,y=0,x=c(mn),col="red",linetype="dashed")+
  geom_point(aes(x = mi,y=.5),col="red",size=2)+
  geom_point(aes(x = mn,y=.5),col="red",size=2)+
  scale_x_continuous(breaks = c(seq,mi,mn),
                 labels = c(seq,"6\n|\nMedian time\n(International music)","45\n|\nMedian time\n(National music)"),limits=c(0,220))+
  scale_color_manual(values=c("black",green_spotify))+ labs(col="")
top<-200
df_all %>%  filter(position<=top) %>%
  group_by(track.name,nacionalidade,url) %>%
  summarise(Time = n() )%>%
  rename(Origin = nacionalidade) ->df_survival
df_survival$Time<-Surv(df_survival$Time)
fit <- survfit(Time ~ Origin, data = df_survival, conf.type = "log-log")
ggsurv <- ggsurvplot(fit,data = df_survival,  pval = TRUE,pval.coord = c(320,0.8),
  conf.int = TRUE xlim = c(0,360), xlab = "Time in days",   break.time.by = 100,
  ggtheme = theme_light(), conf.int.style = "step",  surv.median.line = "none"
 )
seq<-seq(0,360,by=20);mi <- 8; mn <- 53
ggsurv+
  geom_segment(xend = c(mi),yend = 0.5,y=0,x=c(mi),col="red",linetype="dashed")+
  geom_segment(xend = c(mn),yend = 0.5,y=0,x=c(mn),col="red",linetype="dashed")+
      geom_point(aes(x = mi,y=.5),col="red",size=2)+
      geom_point(aes(x = mn,y=.5),col="red",size=2)+
      scale_x_continuous(breaks = c(seq,mi,mn),labels = c(seq,"9\n|\nMedian time\n(International
music)","53\n|\nMedian time\n(National music)"),limits=c(0,360))+
      scale_color_manual(values=c("black",green_spotify))+ labs(col="")
    df_all %>%  group_by(artist,track.name,url) %>%  summarise(n=n()) %>%
      mutate(Top = "Top 200")->df_n200
    df_all %>% filter(position<=10) %>%
      group_by(track.name,url,artist) %>%  summarise(n=n()) %>%  select(-url) %>%
      mutate(Top = "Top 10") ->df_n10
    df_n <-df_n200 %>%  full_join(df_n10)
    values<-c(2,seq(10,90,by=10),seq(100,350,by=50),360);n<-length(values)
    lbs<-c("Just\nonce",paste0("From\n",values[1:(n-1)],
                          " to ",values[2:n]-1))
    df_n %>%  mutate(n2 = cut(n,c(1,values),include.lowest=T,right = F)) %>%
      ggplot(aes(n2))+
      theme_bw(16)+
      geom_bar(width=.25,col="black",position = "dodge",
               fill = green_spotify)+
      labs(x="Number of times a song appears in the Top 200",
           y="Absolute frequency",fill = "Top:")+
      geom_label(stat="count",aes(label=..count..,y=..count..),
                 position = position_dodge(.5),vjust=.7)+
      scale_x_discrete(labels = lbs)+  theme(legend.position = "top")+  scale_fill_brewer(palette =
"Set1")+ facet_wrap(~Top,scales = "free_y",ncol=1)
    df_all %>%  mutate(date = month(lubridate::ymd(date))) %>%
      group_by(date,nacionalidade) %>%  summarise(n=n_distinct(artist)) %>%
      ungroup() %>%  group_by(date) %>%  mutate(N= sum(n),p_artist=n/N) %>%
      select(p_artist, nacionalidade) -> df_art
    df_all %>%  mutate(date = month(lubridate::ymd(date))) %>%
      group_by(date,nacionalidade) %>%  summarise(n=n_distinct(url)) %>%
      ungroup() %>%  group_by(date) %>%  mutate(N= sum(n),p_song=n/N) %>%
      select(p_song,nacionalidade)->df_song
    df_art %>%  left_join(df_song) %>%  tidyr::gather(var,value,p_artist, p_song) %>%
      mutate(var = ifelse(var == 'p_artist', "By artist","By song")) ->df_p
    df_p %>%  mutate(pos = case_when(value <  0.5  ~  1.4,value >  0.5  ~  -.75,value == 0.5  ~
0)) ->df_p
    df_p %>%  ggplot(aes(x= date, y=value,group = nacionalidade))+
      geom_text(aes(label = round(value,2), vjust = pos))+  geom_line()+
      geom_point(size=3,aes(col=nacionalidade))+
      geom_hline(linetype="dashed",yintercept = .5,col="red")+
      scale_x_continuous(breaks = 1:12,labels=month.abb)+
      labs(x = "Month",y="Proportion",col="Origin:")+
      theme_bw(base_size = 16)+  theme(legend.position = "top")+
      scale_y_continuous(breaks = seq(0,1,.05),labels = seq(0,1,.05),limits=c(0.35,.7))+
      facet_wrap(~var,ncol = 1)+  scale_color_manual(values=c("black",green_spotify))+
      theme(panel.background =element_rect(colour = "black",size = 2),
      panel.grid = element_line(colour="grey79"))
```