Projekt semestralny.

Programowanie i analiza danych w R.

Co robimy?

- Pobieramy od Spotify dane w formacie json.
- Przetwarzamy je w programie rStudio.
- Tworzymy podsumowanie i wizualizację.

W jakim celu?

- Uzyskane w ten sposób wyniki podsumowują pewne aspekty działalności użytkowników platformy Spotify.
- Podsumowanie to adresowane jest do osób zajmujących się muzyką i poszukujących podstawowych statystyk tej dziedzinie.

Import Search Queries

```
### creating dataframe with date, devices and country
Search_Queries_df <- function(folder_path){
  file_path <-list.files(folder_path,"SearchQueries")
  select(jsonlite::fromJSON(file_path),1:3) %>%
  mutate(date = ymd(date))
}
```

	date	platform	country
1	2019-09-30	ANDROID	PL
2	2019-09-30	DESKTOP	PL
3	2019-09-30	DESKTOP	PL
4	2019-10-01	ANDROID	PL
5	2019-10-01	ANDROID	PL
6	2019-10-01	ANDROID	PL
7	2019-10-01	ANDROID	PL
8	2019-10-01	ANDROID	PL
9	2019-10-01	ANDROID	PL

Import Streaming History

```r
##returns dataframe with data from spotify
Streaming_History_df <- function(folder_path){
  file_paths<- list.files(folder_path,"StreamingHistory")
  read_files <- lapply(file_paths, jsonlite::fromJSON)
  bind_rows(read_files)
}
```

	endTime	artistName	trackName	msPlayed
1	2018-10-29 17:49	King Crimson	Starless	480009
2	2018-10-30 05:03	Guns N' Roses	14 Years	8170
3	2018-10-30 05:03	Van Halen	Take Your Whiskey Home - 2015 Remaster	1186
4	2018-10-30 05:11	Genesis	Dancing With The Moonlit Knight - Remastered 2008	483053
5	2018-10-30 05:15	Genesis	I Know What I Like (In Your Wardrobe) - Remastered 2008	250400
6	2018-10-30 06:19	Genesis	Firth Of Fifth - Remastered 2008	526259
7	2018-10-30 06:23	Lao Che	Kapitan Polska	245377
8	2018-10-30 06:28	Ozzy Osbourne	Crazy Train	296200
9	2018-10-30 06:29	Lao Che	United Colours Of Armagedon	6888
10	2018-10-30 06:33	Ozzy Osbourne	Mama, I'm Coming Home	251866
11	2018-10-30 06:37	Ozzy Osbourne	Bark at the Moon	257120
12	2018-10-30 06:39	Ozzy Osbourne	No More Tears	85205

Czyszczenie Streaming History

```
#### creating and preparing dataframe
Streaming_History_Complete <- function(folder_path){
  Streaming_History_df(folder_path) %>%
  names_change() %>%
  mutate(end_time = ymd_hm(end_time)) %>%
  mutate(s_played = dmilliseconds((s_played))) %>%
  add_start_time() %>%
  add_skipped() %>%
  add_weekday()
}
```

	end_time	artist_name	track_name	s_played	start_time	skipped	weekday
1	2018-10-29 17:49:00	King Crimson	Starless	480.009s (~8 minutes)	2018-10-29 17:40:59	FALSE	pon\.
2	2018-10-30 05:03:00	Guns N' Roses	14 Years	8.17s	2018-10-30 05:02:51	TRUE	wt\.
3	2018-10-30 05:03:00	Van Halen	Take Your Whiskey Home - 2015 Remaster	1.186s	2018-10-30 05:02:58	TRUE	wt\.
4	2018-10-30 05:11:00	Genesis	Dancing With The Moonlit Knight - Remastered 2008	483.053s (~8.05 minutes)	2018-10-30 05:02:56	FALSE	wt\.
5	2018-10-30 05:15:00	Genesis	I Know What I Like (In Your Wardrobe) - Remastered 2008	250.4s (~4.17 minutes)	2018-10-30 05:10:49	FALSE	wt\.
6	2018-10-30 06:19:00	Genesis	Firth Of Fifth - Remastered 2008	526.259s (~8.77 minutes)	2018-10-30 06:10:13	FALSE	wt\.
7	2018-10-30 06:23:00	Lao Che	Kapitan Polska	245.377s (~4.09 minutes)	2018-10-30 06:18:54	FALSE	wt\.

Czyszczenie Streaming History cd.

```r
names_change <- function(streaming_history, column_names = c("end_time", "artist_name", "track_name", "s_played"))
  names(streaming_history) <- column_names
  streaming_history
}
##returns streaming_history with added start_time column using end_time and ms_played
add_start_time <- function(streaming_history){

  start_time <- streaming_history[["end_time"]] - streaming_history[["s_played"]]
  streaming_history <- cbind(streaming_history, start_time)
  streaming_history

}

##returns streaming history with added "skipped" column [true or false]
add_skipped <- function(streaming_history){
  skipped <- (streaming_history[["s_played"]] < duration(10, "seconds"))
  streaming_history <- cbind(streaming_history, skipped)
}

##returns streaming history with added weekdays column
add_weekday <- function(streaming_history){
  weekday <- wday(streaming_history[["start_time"]], label = TRUE)
  streaming_history <- cbind(streaming_history, weekday)
}
```

Import Playlist

```r
### creating dataframe with names of playlists, string containing song names separated
#by ";;;" and artist names separated by ";;;"
Playlist_df_function <- function(folder_path){
  file_path <-list.files(folder_path,"Playlist")
  a <- jsonlite::fromJSON(file_path)
  b <- select(a[[1]], name, items)
  num_of_rows <- dim(b)[1]
  vec_num_of_rows <- 1:num_of_rows
  df <- data.frame(b$name,NA, NA)
  colnames(df) <- c("Playlist names","Song names","Artist names")
  for (i in vec_num_of_rows){
    df[i,2] <- paste(b[i,2][[1]]$track$trackName, collapse = ";;;")
    df[i,3] <- paste(b[i,2][[1]]$track$artistName, collapse = ";;;")
  df[,1] <- as.character(df[,1])
  }

  return(df)
}
```

	Playlist names	Song names	Artist names
1	Spotify.Me	Breaking the Law;;;So Sentimental;;;Space Groove II (Edit);;;Br...	Judas Priest;;;Violent Soho;;;King Crimson;;;The Rolling Stone...
2	My Shazam Tracks	Wonderful Life;;;SICKO MODE;;;Mo Bamba;;;Freedom at 21	Black;;;Travis Scott;;;Sheck Wes;;;Jack White
3	Pobrane	Owner of a Lonely Heart;;;Hold On;;;It Can Happen;;;Change...	Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Yes;;;Ye...
4	Łowełek	Breaking the Law;;;Kickstart My Heart;;;Paranoid;;;Panama;;;(D...	Judas Priest;;;Mötley Crüe;;;Black Sabbath;;;Van Halen;;;Blue ...
5	Szit którego sie się wstydzę	Mope;;;I Hope You Die;;;Venus In Furs;;;99 Luftballons;;;Wavin...	Bloodhound Gang;;;Bloodhound Gang;;;The Velvet Undergro...
6	Urodzinowa playlista	Gimme! Gimme! Gimme! (A Man After Midnight);;;Girls Just ...	ABBA;;;Cyndi Lauper;;;Bajm;;;MC Hammer;;;Queen;;;Stevie W...
7	KMINA_V2	Tetris;;;Tetris;;;Tetris;;;Tetris;;;Tetris;;;Tetris;;;Tetris	The Game Music Committee;;;The Game Music Committee;;;...

Streaming History i Playlist

```
# creating dataframe similar to Streaming_History_Complete, but this one has additional
#column with lists of playlists that including that song
Playlist_df_Str_his <- function(folder_path){
  Str_his_df <- Streaming_History_Complete(folder_path)
  Playlist_df <- Playlist_df_function(folder_path)
  In_which_playlist <- select(data.frame(1:dim(Str_his_df)[1],NA),2)
  colnames(In_which_playlist) <- c("In_playlists")
  for (i in 1:dim(Playlist_df)[1]){
    splited_song_names <- strsplit(Playlist_df[i,2],";;;")

    splited_artist_names <- strsplit(Playlist_df[i,3],";;;")

    for (j in 1:dim(Str_his_df)[1]){
      for(k in 1:length(splited_artist_names[[1]])){
        if(Str_his_df[j,2]==splited_artist_names[[1]][k] & Str_his_df[j,3]==splited_song_names[[1]][k] )
        {
         if(is.na(In_which_playlist[j,1]==TRUE)){
           In_which_playlist[j,1] <- Playlist_df[i,1]
           break
         }
          else {In_which_playlist[j,1]<- paste(In_which_playlist[j,1], ";",Playlist_df[i,1])
            break
          }
        }
     }
    }
  }
  strsplit_playlist <- function(dataframe){
    dataframe$In_playlists <- strsplit(dataframe$In_playlists,";")
    dataframe
  }
  cbind(Str_his_df,In_which_playlist) %>%
  strsplit_playlist
}
```

Streaming History i Playlist

17547	2019-06-29 18:37:00	Iron Maiden	2 Minutes to Midnight - 2015 Remaster	365.580s (~6.06 minutes)	2019-06-29 18:30:...	FALSE	sob\.	*NA*
17548	2019-06-29 18:40:00	Steppenwolf	Born To Be Wild	210.373s (~3.51 minutes)	2019-06-29 18:36:...	FALSE	sob\.	Łowełek
17549	2019-06-29 18:45:00	Metallica	For Whom The Bell Tolls - Remastered	309.973s (~5.17 minutes)	2019-06-29 18:39:...	FALSE	sob\.	Łowełek
17550	2019-06-29 18:49:00	Alice Cooper	School's Out	210.106s (~3.5 minutes)	2019-06-29 18:45:...	FALSE	sob\.	c("Łowełek ", " Szit którego sie się wstydzę")
17551	2019-06-29 18:54:00	Bon Jovi	Bad Medicine	316.706s (~5.28 minutes)	2019-06-29 18:48:...	FALSE	sob\.	Łowełek
17552	2019-06-29 18:58:00	AC/DC	T.N.T.	214.666s (~3.58 minutes)	2019-06-29 18:54:...	FALSE	sob\.	Łowełek
17553	2019-06-29 19:02:00	Van Halen	Hot for Teacher - 2015 Remaster	282.746s (~4.71 minutes)	2019-06-29 18:57:...	FALSE	sob\.	c("Łowełek ", " Szit którego sie się wstydzę")
17554	2019-06-29 19:08:00	Blue Öyster Cult	(Don't Fear) The Reaper	308.12s (~5.14 minutes)	2019-06-29 19:02:...	FALSE	sob\.	Łowełek
17555	2019-06-29 19:12:00	ZZ Top	Sharp Dressed Man - 2008 Remaster	258.026s (~4.3 minutes)	2019-06-29 19:07:...	FALSE	sob\.	Łowełek
17556	2019-06-29 19:15:00	Judas Priest	Breaking the Law	153.84s (~2.56 minutes)	2019-06-29 19:12:...	FALSE	sob\.	c("Spotify.Me ", " Łowełek")
17557	2019-06-29 19:20:00	Iron Maiden	Speed of Light	301.745s (~5.03 minutes)	2019-06-29 19:14:...	FALSE	sob\.	Łowełek
17558	2019-06-29 19:25:00	Iron Maiden	The Number of the Beast - 2015 Remaster	290.586s (~4.84 minutes)	2019-06-29 19:20:...	FALSE	sob\.	*NA*
17559	2019-06-29 19:28:00	Black Sabbath	Paranoid	168.44s (~2.81 minutes)	2019-06-29 19:25:...	FALSE	sob\.	c("Łowełek ", " Szit którego sie się wstydzę")
17560	2019-06-29 19:34:00	Led Zeppelin	Whole Lotta Love - 1990 Remaster	333.893s (~5.56 minutes)	2019-06-29 19:28:...	FALSE	sob\.	Łowełek
17561	2019-06-29 19:34:00	Motörhead	Ace of Spades	8.438s	2019-06-29 19:33:...	TRUE	sob\.	Łowełek
17562	2019-06-29 19:37:00	Iron Maiden	Aces High - 2015 Remaster	200.138s (~3.34 minutes)	2019-06-29 19:33:...	FALSE	sob\.	*NA*
17563	2019-06-29 19:42:00	Black Sabbath	Children of the Grave	314.72s (~5.25 minutes)	2019-06-29 19:36:...	FALSE	sob\.	Łowełek
17564	2019-06-29 19:47:00	Led Zeppelin	Black Dog - Remaster	295.386s (~4.92 minutes)	2019-06-29 19:42:...	FALSE	sob\.	Łowełek
17565	2019-06-29 19:50:00	Judas Priest	Hell Bent for Leather	160.866s (~2.68 minutes)	2019-06-29 19:47:...	FALSE	sob\.	Łowełek
17566	2019-06-29 19:54:00	Van Halen	Panama - 2015 Remaster	210.226s (~3.5 minutes)	2019-06-29 19:50:...	FALSE	sob\.	*NA*
17567	2019-06-29 20:00:00	Judas Priest	Painkiller	365.826s (~6.1 minutes)	2019-06-29 19:53:...	FALSE	sob\.	Łowełek
17568	2019-06-29 20:04:00	Scorpions	Rock You Like a Hurricane	255.573s (~4.26 minutes)	2019-06-29 19:59:...	FALSE	sob\.	Łowełek
17569	2019-06-29 20:08:00	KISS	I Was Made For Lovin' You	271.24s (~4.52 minutes)	2019-06-29 20:03:...	FALSE	sob\.	c("Łowełek ", " Urodzinowa playlista")
17570	2019-06-29 20:12:00	Judas Priest	Electric Eye	222.44s (~3.71 minutes)	2019-06-29 20:08:...	FALSE	sob\.	Łowełek
17571	2019-06-29 20:17:00	Mötley Crüe	Kickstart My Heart	282.92s (~4.72 minutes)	2019-06-29 20:12:...	FALSE	sob\.	Łowełek
17572	2019-06-29 20:22:00	Judas Priest	You've Got Another Thing Coming	310.08s (~5.17 minutes)	2019-06-29 20:16:...	FALSE	sob\.	Łowełek
17573	2019-06-29 20:24:00	Judas Priest	Turbo Lover - Single Version	111.483s (~1.86 minutes)	2019-06-29 20:22:...	FALSE	sob\.	Łowełek

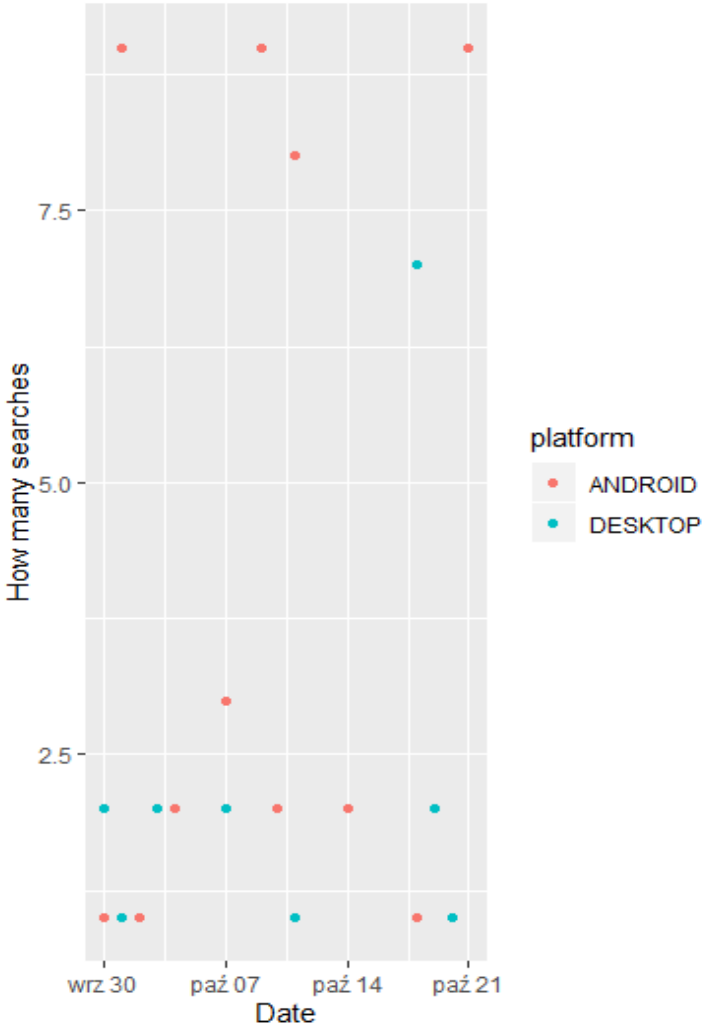
Przykładowe wywołanie

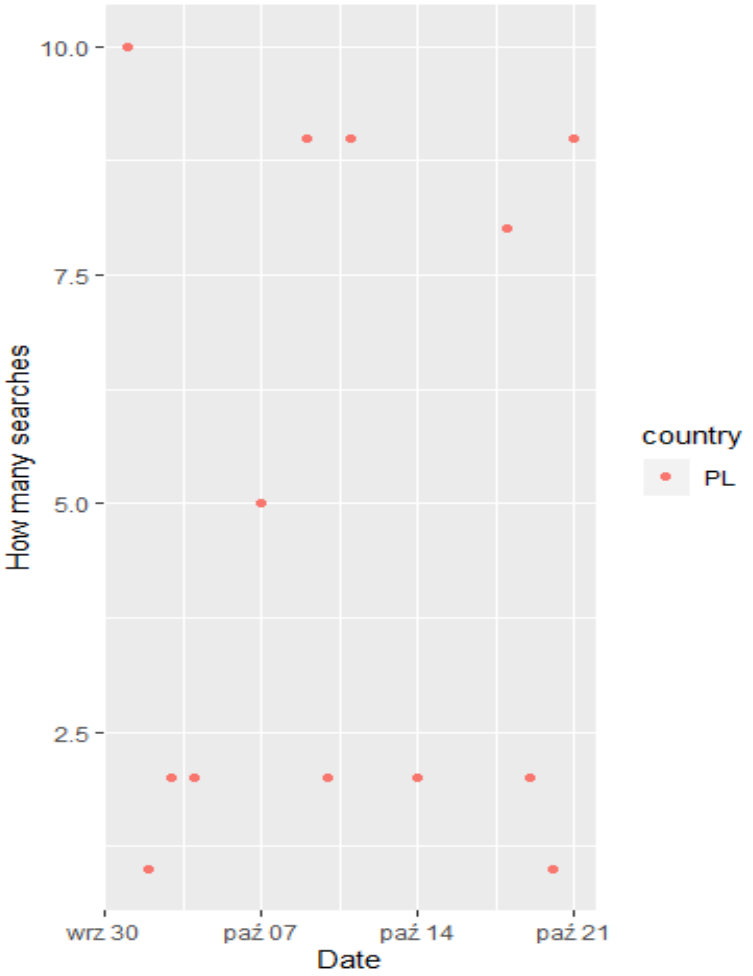
```
> how_long_listened(streaming_history, "2019-06-21", "2019-07-24")
[1] "289098.45s (~3.35 days)"
> how_long_listened(streaming_history, "2019-06-21", "2019-07-24", as_percentage = TRUE)
[1] "10.14%"
>
> most_played_track(streaming_history, "2019-06-21", "2019-07-24", how_many = 10)
# A tibble: 10 x 2
   track_name                                          number
   <chr>                                                <int>
 1 Where Is My Mind?                                       16
 2 When You Die                                            13
 3 Igra Rok En Rol Cela Jugoslavija                         9
 4 Thela Hun Ginjeet                                        6
 5 There Is a Light That Never Goes Out - 2011 Remaster     6
 6 Blackjack                                                5
 7 Cyboogie                                                 5
 8 Disorder - 2007 Remaster                                 5
 9 Elephant Talk                                            5
10 Fishing For Fishies                                      5
> most_skipped_track(streaming_history, "2019-06-21", "2019-09-24")
# A tibble: 10 x 2
   track_name                                          number
   <chr>                                                <int>
 1 Disorder - 2007 Remaster                                12
 2 Igra Rok En Rol Cela Jugoslavija                         7
 3 Behind Blue Eyes                                         6
 4 Dancing With The Moonlit Knight - Remastered 2008        6
 5 Forever Young                                            6
 6 Hate That I Love You                                     6
 7 There Is a Light That Never Goes Out - 2011 Remaster     6
 8 99 Red Balloons                                          5
 9 Hyper Hyper                                              5
10 Jest taki samotny dom                                    5
```

Przykładowe wywołanie

```
> most_played_artist(streaming_history, "2019-06-21", "2019-07-24")
# A tibble: 10 x 2
   artist_name                    number
   <chr>                           <int>
 1 King Gizzard & The Lizard Wizard     93
 2 King Crimson                         60
 3 Pink Floyd                           48
 4 MGMT                                 46
 5 Kraftwerk                            41
 6 Death Grips                          37
 7 Tame Impala                          32
 8 Scooter                              31
 9 The Kinks                            27
10 Radiohead                            24
> most_skipped_artist(streaming_history, "2019-06-21", "2019-07-24", how_many = 5)
# A tibble: 5 x 2
  artist_name  number
  <chr>         <int>
1 Joy Division     10
2 Deep Purple       7
3 Alphaville        6
4 King Crimson      6
5 Led Zeppelin      6
>
> how_many_skipped(streaming_history, "2019-06-21", "2019-07-01")
[1] 202
> how_many_skipped(streaming_history, "2019-06-01", "2019-07-29", as_percentage = TRUE)
[1] "3.283%"
```

Przykładowe wywołanie - Wizualizacja



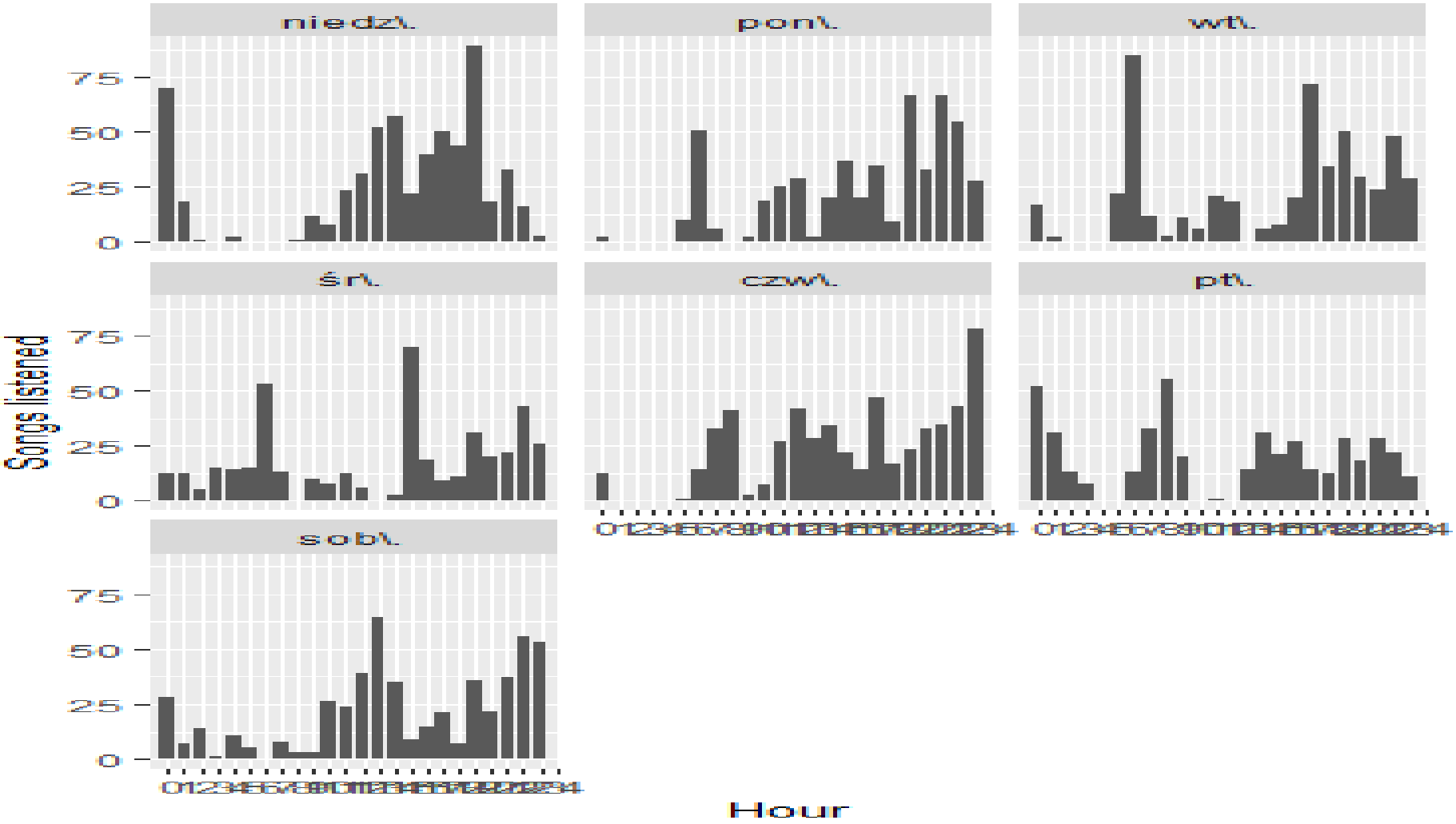


Przykładowe wywołanie - Wizualizacja

```
##
platform_used_by_date<- function(search_queries, start_date, end_date){
  filter(search_queries, date >= ymd(start_date), date <= ymd(end_date)) %>%
    ggplot(aes(x = date, color = platform))+
    geom_point(stat = "count")+
    xlab("Date")+
    ylab("How many searches")
}

##
country_by_date <- function(search_queries, start_date, end_date){
  filter(search_queries, date >= ymd(start_date), date <= ymd(end_date)) %>%
    ggplot(aes(x = date, color = country))+
    geom_point(stat = "count" )+
    xlab("Date")+
    ylab("How many searches")
}
```

Przykładowe wywołanie - Wizualizacja



Przykładowe wywołanie - Wizualizacja

```
##visualizes number of songs played in given time period at different hours
number_of_songs_listened_by_hour <- function(streaming_history, start_date, end_date,
                                             by_weekday = FALSE, dont_show_skipped = TRUE){
  filtered <- filter(streaming_history, start_time >= ymd(start_date), start_time <= ymd(end_date))
  if (dont_show_skipped) filtered <- filter(filtered, skipped == FALSE)
  vis <- ggplot(filtered, aes(x = hour(start_time)))+
    geom_bar()+
    scale_x_discrete(limits = 0:24)+
    xlab("Hour")+
    ylab("Songs listened")

  if (by_weekday){
    vis <- vis+
      facet_wrap(~weekday)
  }
  vis
}
```

Do zrobienia

- Przy pomocy zewnętrznego API, dla danego utworu i wykonawcy, uzyskać informację o gatunku danego utworu.
- Analiza danych wykorzystująca, gatunek danego utworu oraz dane playlisty
- Aplikacja w Shiny