The Music by Me

Giulio

2016-10-04

Contents

```r
knitr::opts_chunk$set(warning = FALSE, message = FALSE)
```

1 Introduction

This is *the* data analysis of **songs**.

2 Material

The data is online:

```r
link1 <- "http://kris.shaffermusic.com/media/MSDSubset-extracted-data.csv"
link2 <- "http://kris.shaffermusic.com/media/MSDSubset-topics.csv"
```

We use a bunch of packages:

3 Methods

We wrote two functions:

```r
#' This is the function that gets the data from the web
gather_music_data <- function(web_link1,web_link2){
  if( ! is.character(web_link1)) stop("I want an URL!")
  if( ! is.character(web_link2)) stop("I want an URL!")

  if( ! url.exists(web_link1)) stop("Internet no good")
  if( ! url.exists(web_link2)) stop("Internet no good")

  #' Use url.exists() to check if a web link exists or not
  #' and if not, stop.
```

```
  msd_data <- read_csv(web_link1) #' this is the MSD data
  genre_data <- read_csv(web_link2) #' this is the genre and lyrics data

  music_data <- left_join(msd_data,genre_data, by = "tid") %>%
    select(-ends_with(".x")) %>%
    rename(artist_name = artist_name.y,
           title = title.y)

  return(music_data)
}
```

and the other function:

```
music_plotter <- function(music_data,
                           xvariable,
                           yvariable,
                           colorvariable,
                           do_smooth) {
  music_plotted <- music_data %>%
    filter(year > 1000) %>%
    ggplot(aes_string(y = yvariable, x = xvariable, color = colorvariable)) +
    geom_point()

  if(do_smooth){
    music_plotted <- music_plotted +
      geom_smooth(method = "lm")
  } else {
    music_plotted <- music_plotted
  }

  return(music_plotted)
}
```

4   Results

We got the data:

```
all_data <- gather_music_data(link1,link2)
```

we plotted it:

```
music_plotter(all_data,"year","loudness","mode",do_smooth = TRUE)
```