

Title: ac1

Date: September 17, 2017

In AC1, I explored the possible relationship between “total_rain” and “music_genre”. I found that people living in low “total_rain” area like to listen New Age, Classical and Blues. People living in high “total_rain” area like to listen Instrumental or Electronica.

I first visualized the distribution of “total_rain”. Most of the data have a value less than 200. In order to better visualize the major population, I eliminated 4 points with a total rain value above 200. To do so, I used for loop to check each element in column “total_rain”. If the value >200, the value of this element was converted to “NA”. Also, I use another for loop to convert all “Electronic Dance Music” to “EDM”, and convert “Electronic” to “Electronica”. Then I used ggplot () + geom_boxplot() to generate the figure below.

Part of the code for generating this box plot are shown below.

```
ggplot(AC1, aes(reorder(music_genre_EDM, -total_rain_less_than_200, median),  
total_rain_less_than_200)) + geom_boxplot() + coord_flip() + scale_x_discrete("music_genre")
```

