Final project sections analysis part

Tianyang Liu

5/30/2021

Stats MAS405 Spring 2021

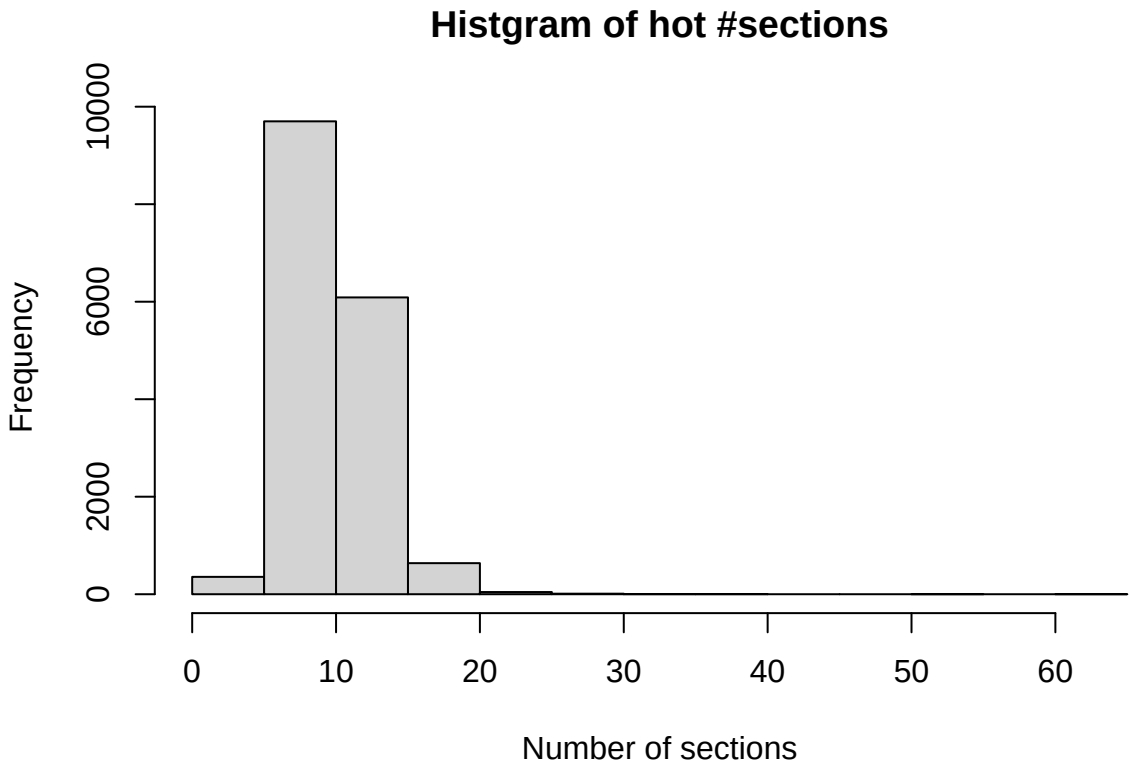
Tianyang Liu

Introduction

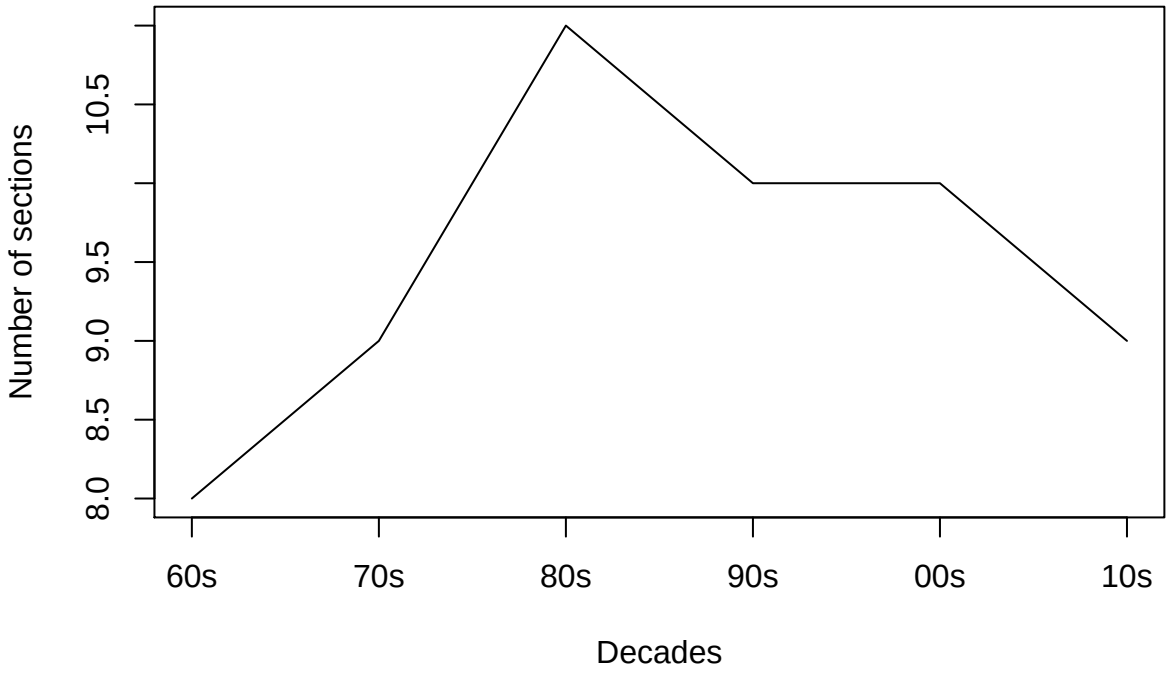
First load all the data from the database

```
# Helper functions to get mode
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
```

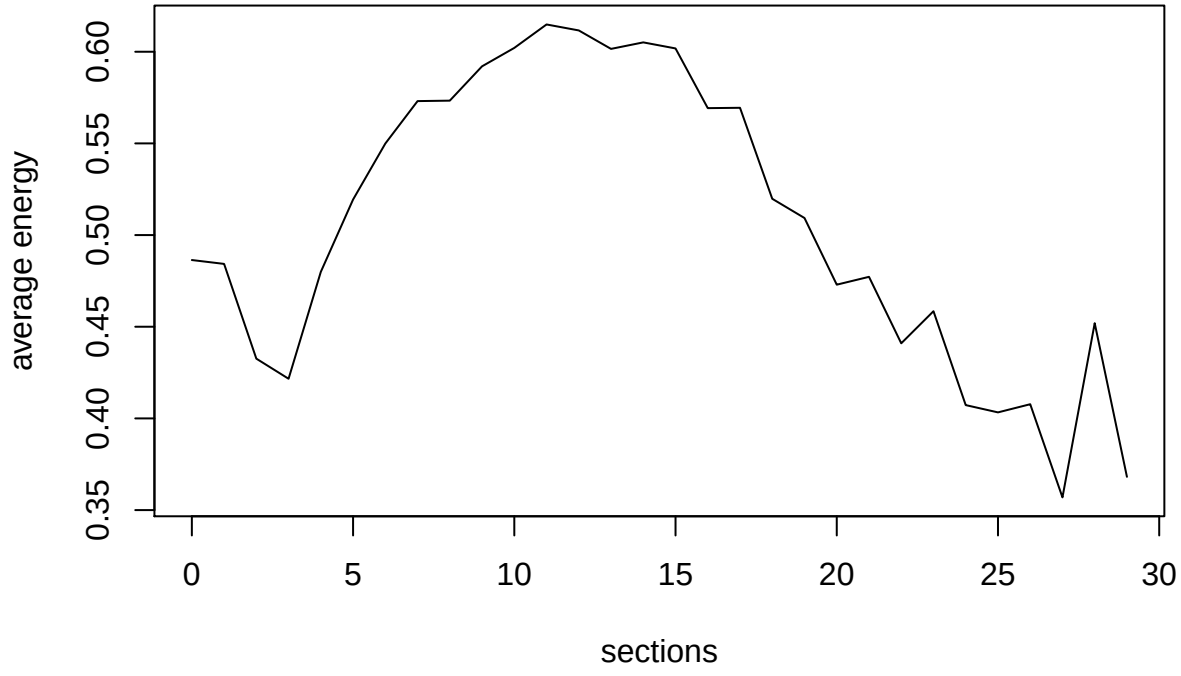
Find the most common number of sections within Hot-100 tracks. It seems like the most common number of sections in Hot-100 tracks is 9.

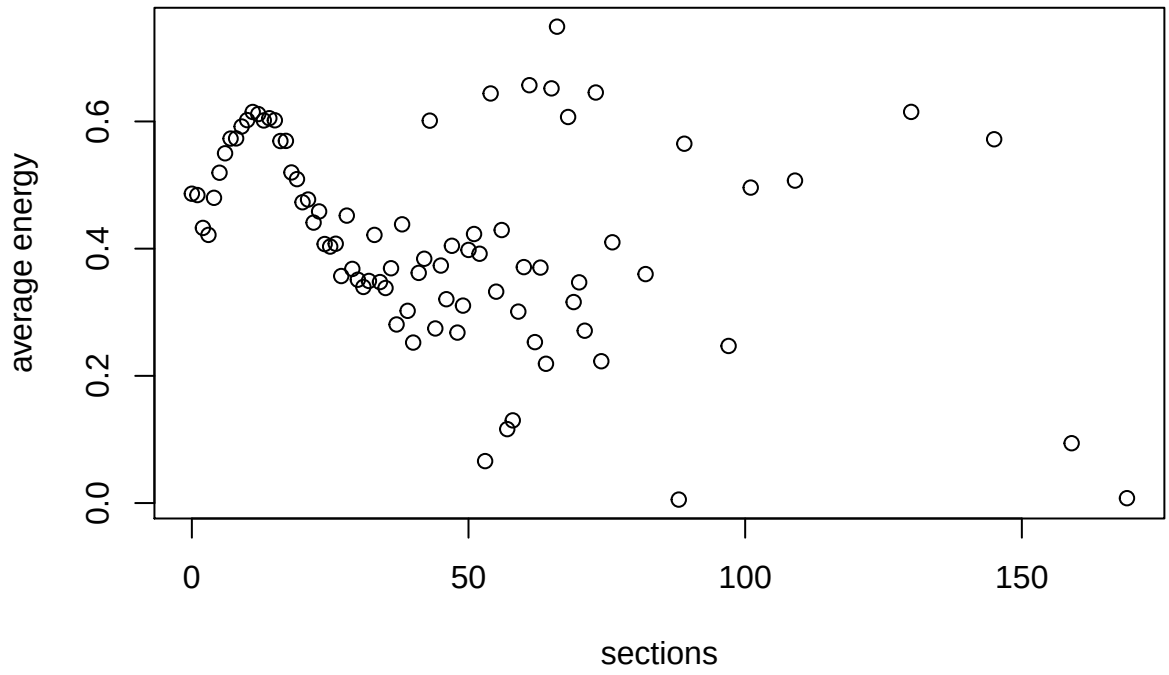


Split the dataset into 6 decades. Then I found out the most common number of sections in each decade. The line plot describes the main trend of number of section along with decades.



Analysis of the relationship between energy and the number of sections. It turns out that they might have a quadratic relationship between average energy and the number of sections from the range of 1-30 sections.



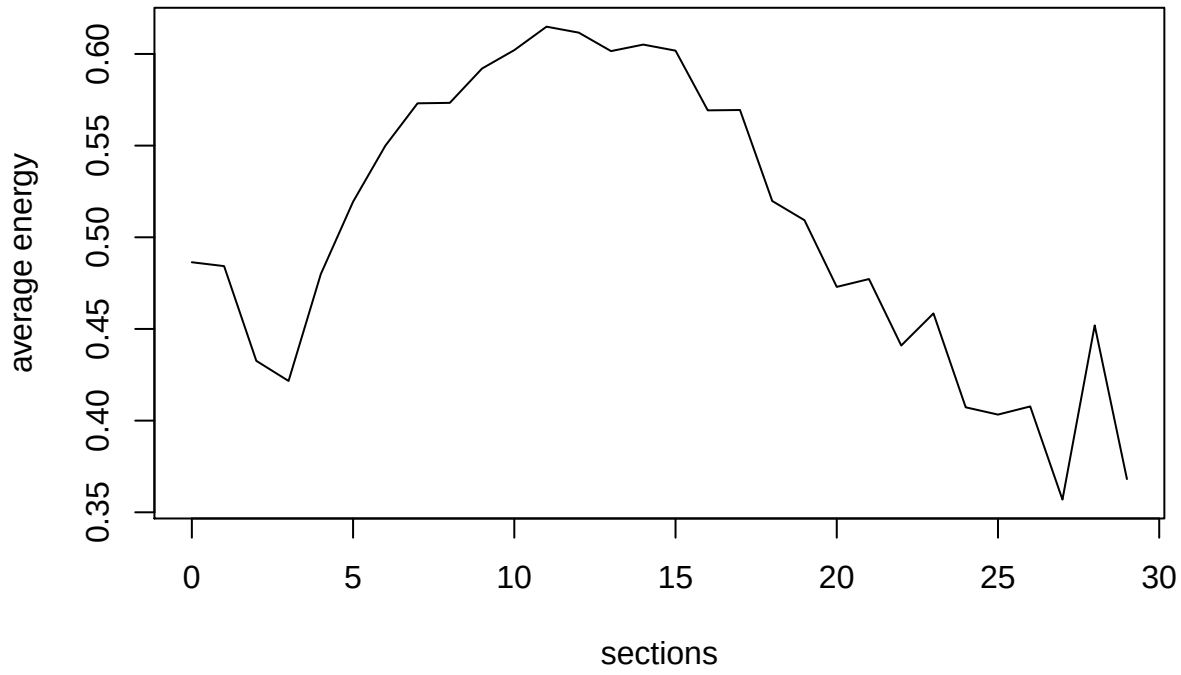


Analysis of the relationship between danceability and the number of sections. It turns out that they might also have a quadratic relationship between average danceability and the number of sections from the range of 1-30 sections.

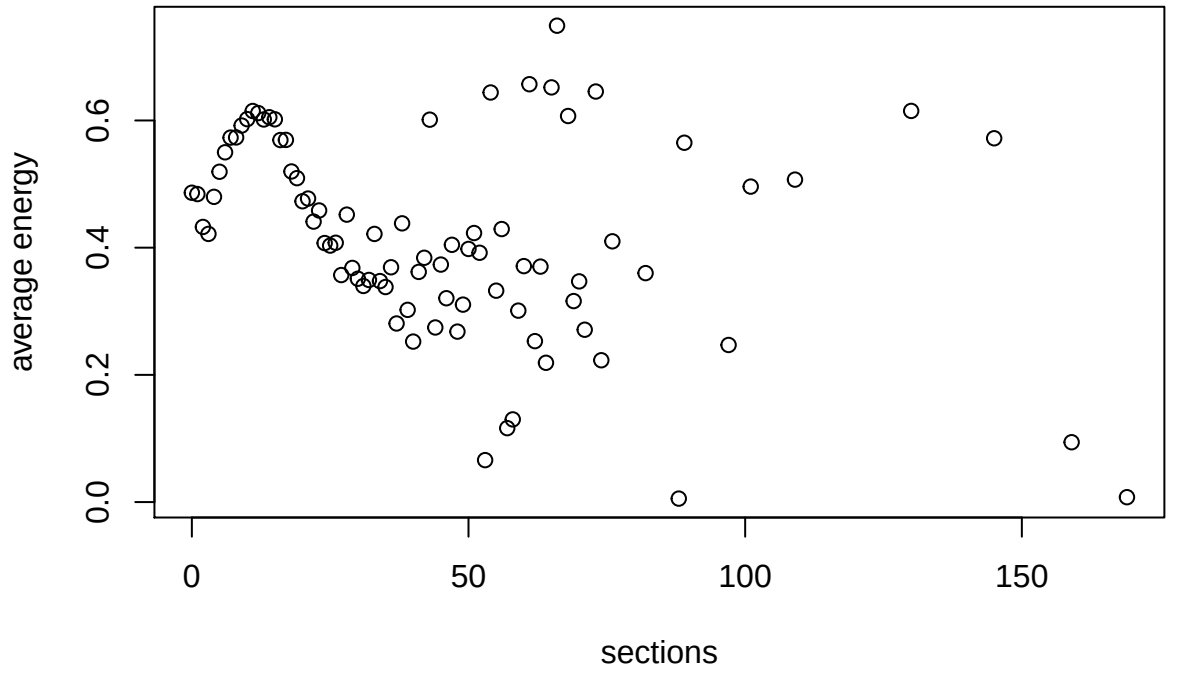
```
# sections and danceability
qstr <- paste0("SELECT sections, AVG(danceability) as average_dance FROM music GROUP BY sections ORDER
sections_danceability <- dbGetQuery(con, qstr)
head(sections_danceability)
```

```
##   sections average_dance
## 1        0     0.4930667
## 2        1     0.4352727
## 3        2     0.4451206
## 4        3     0.4363454
## 5        4     0.4623787
## 6        5     0.4861642
```

```
sd_related = sections_danceability[1:30, ]
plot(se_related$sections, se_related$average_energy, xlab = "sections", ylab = "average energy", type =
```




```
# The relationship between sections and average energy in total range
plot(sections_energy$sections, sections_energy$average_energy, xlab = "sections", ylab = "average energy
```



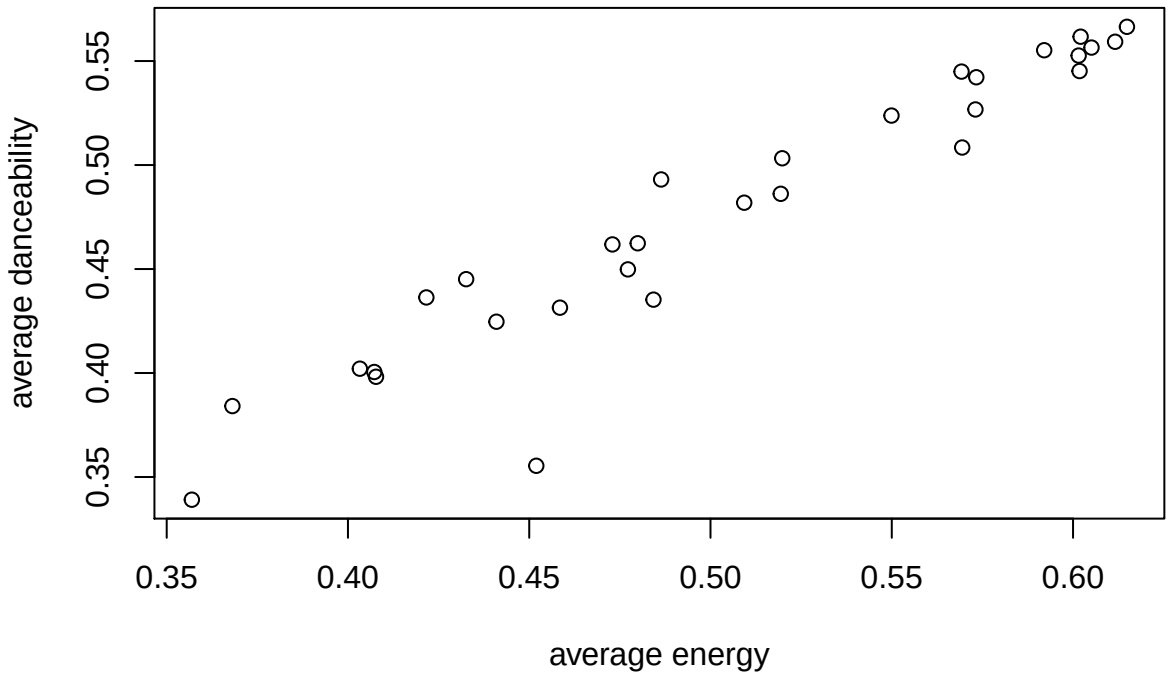
Combine the average energy, average danceability, sections into one dataframe. Surprisingly, it turns out that average energy and average danceability is kind of having some linear relationship within the range of [1, 30] sections

```
sd_related$average_energy <- se_related$average_energy
fig <- plot_ly(x = sd_related$sections, y = sd_related$average_energy, z = sd_related$average_dance, ty
```

```
fig <- fig %>% layout(
    title = "Relation between sections and energy&danceability",
    scene = list(
      xaxis = list(title = "sections"),
      yaxis = list(title = "avg_energy"),
      zaxis = list(title = "avg_danceability")
    ))
fig
```

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, pleas
```

```
plot(se_related$average_energy, sd_related$average_dance, xlab = "average energy", ylab = "average danc
```




```
avgvalue_30section <- data.frame(cbind(se_related$average_energy, sd_related$average_dance))
colnames(avgvalue_30section) <- c("avg_energy", "avg_danceability")
head(avgvalue_30section)
```

```
##   avg_energy avg_danceability
## 1  0.4863800        0.4930667
## 2  0.4842727        0.4352727
## 3  0.4325989        0.4451206
## 4  0.4216268        0.4363454
## 5  0.4799220        0.4623787
## 6  0.5193806        0.4861642
```