Spotify_Analysis_Project

2023-03-28

```r
file<-read.csv("Top_Hits_2000_2019.csv") # Read the original dataset
summary(file['duration..min.'])
```
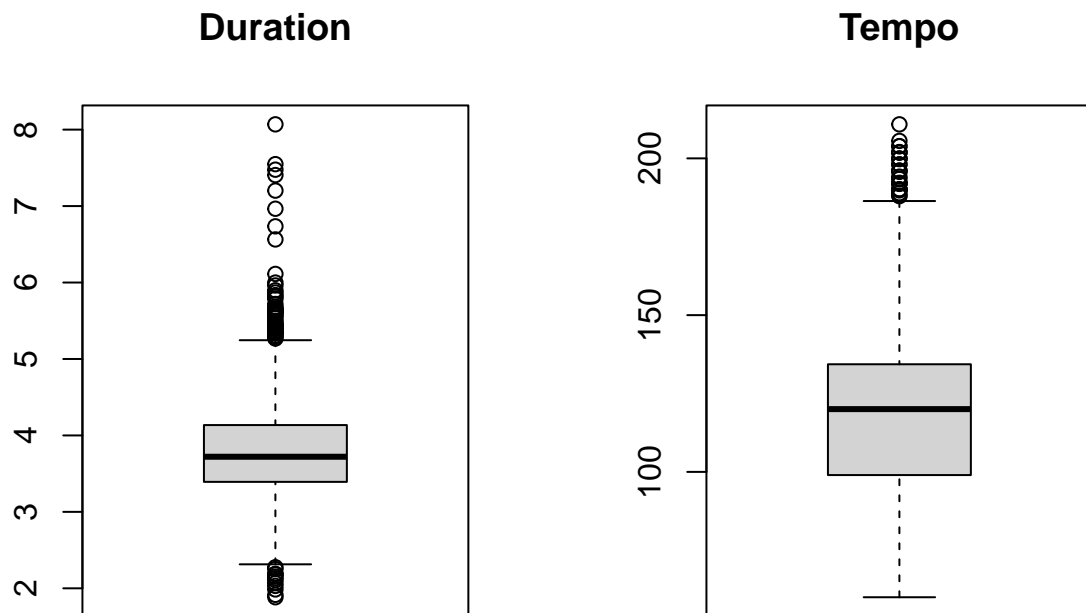
```
##  duration..min.
##  Min.   :1.883
##  1st Qu.:3.393
##  Median :3.721
##  Mean   :3.812
##  3rd Qu.:4.136
##  Max.   :8.069
```

```r
summary(file['tempo..bpm.'])
```

```
##   tempo..bpm.
##  Min.   : 60.02
##  1st Qu.: 98.99
##  Median :120.02
##  Mean   :120.12
##  3rd Qu.:134.27
##  Max.   :210.85
```

Create box plots of duration & tempo here

```r
par(mfrow=c(1,2))    # Plot 2 graphs side by side
boxplot(file['duration..min.'], main=c('Duration'))
boxplot(file['tempo..bpm.'], main=c('Tempo'))
```
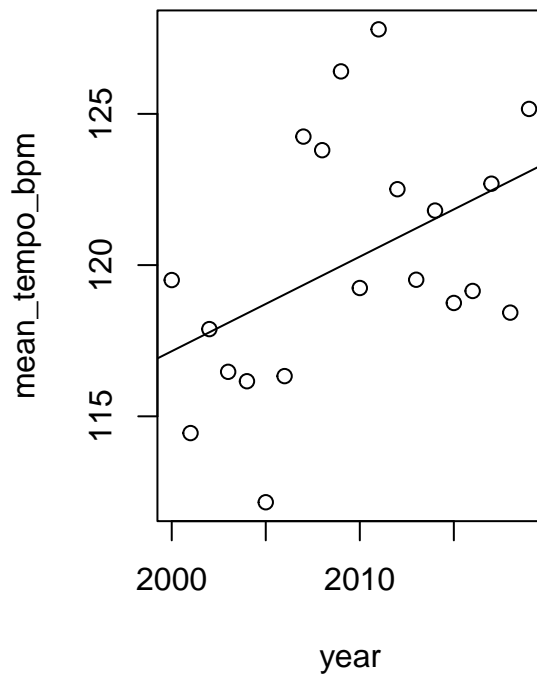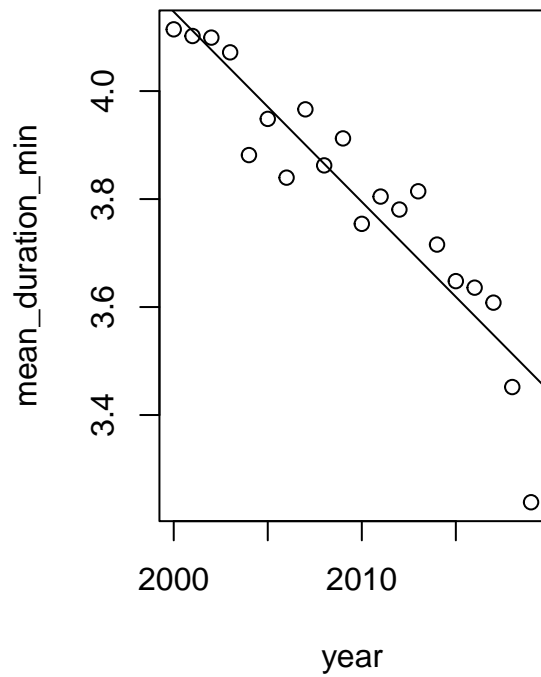
Duration

Tempo

Run linear regression on the parameters

```r
file_mean <- read.csv('Top_Hits_Yearly.csv')

par(mfrow=c(1,2))     # Plot 2 graphs side by side
lm_fit <- lm(mean_duration_min ~ year, data = file_mean)
lm_summary <- summary(lm_fit)
lm_coef1 <- lm_summary$coefficients
plot(mean_duration_min ~ year, data = file_mean)
abline(lm_fit)

lm_fit <- lm(mean_tempo_bpm ~ year, data = file_mean)
lm_summary <- summary(lm_fit)
lm_coef2 <- lm_summary$coefficients
plot(mean_tempo_bpm ~ year, data = file_mean)
abline(lm_fit)
```

```
lm_coef1 # coefficients for duration
```

```
##              Estimate  Std. Error  t value      Pr(>|t|)
## (Intercept) 74.55661733 6.711348888  11.10904 1.729160e-09
## year        -0.03520485 0.003339797 -10.54102 3.945874e-09
```

```
lm_coef2 # coefficients for tempo
```

```
##               Estimate  Std. Error  t value   Pr(>|t|)
## (Intercept) -508.0140084 293.6909343 -1.729757 0.10078366
## year          0.3125835   0.1461506  2.138776 0.04641414
```