Project 8, Group 2

Dylan, Meera, Molly

November 13, 2017

# Question 1a

tp100song <- function(y) {  
 mydoc <- htmlParse(paste("~/290/gp08g2/scratch/billboardhot100/",y,sep=""))  
 mysongs <- xpathSApply(mydoc, "//\*/div[@class='chart-row\_\_title']/h2[@class='chart-row\_\_song']", xmlValue)  
 return(mysongs)  
}  
head(tp100song("2017-11-11"))

## [1] "Rockstar" "Bodak Yellow (Money Moves)"  
## [3] "1-800-273-8255" "Feel It Still"   
## [5] "Thunder" "Sorry Not Sorry"

# Question 1b

num1b <- sapply(head(dates,n=2),function(y) head(tp100song(y)), simplify = FALSE)  
unlist(num1b)

## [1] "Poor Little Fool" "Patricia"   
## [3] "Splish Splash" "Hard Headed Woman"   
## [5] "When" "Rebel-'rouser"   
## [7] "Poor Little Fool" "Nel Blu Dipinto Di Blu (VolarÃ©)"  
## [9] "Patricia" "Splish Splash"   
## [11] "When" "My True Love"

# Question 1c

tp100art <- function(y) {  
 mydoc1 <- htmlParse(paste("~/290/gp08g2/scratch/billboardhot100/",y,sep=""))  
 myartists <- xpathSApply(mydoc1, "//\*/div[@class='chart-row\_\_title']/span[@class='chart-row\_\_artist']|//\*/div[@class='chart-row\_\_title']/a[@class='chart-row\_\_artist']", xmlValue)   
 myartists <- sub("^\\s+", "", myartists)  
 myartists <- sub("\\s+$", "", myartists)  
 return(myartists)  
}  
head(tp100art("2017-11-11"))

## [1] "Post Malone Featuring 21 Savage"   
## [2] "Cardi B"   
## [3] "Logic Featuring Alessia Cara & Khalid"  
## [4] "Portugal. The Man"   
## [5] "Imagine Dragons"   
## [6] "Demi Lovato"

# Question 1d

num1d <- sapply(head(dates,n=2),function(y) head(tp100art(y)),simplify = FALSE)  
unlist(num1d)

## [1] "Ricky Nelson"   
## [2] "Perez Prado And His Orchestra"   
## [3] "Bobby Darin"   
## [4] "Elvis Presley With The Jordanaires"   
## [5] "Kalin Twins"   
## [6] "Duane Eddy His Twangy Guitar And The Rebels"  
## [7] "Ricky Nelson"   
## [8] "Domenico Modugno"   
## [9] "Perez Prado And His Orchestra"   
## [10] "Bobby Darin"   
## [11] "Kalin Twins"   
## [12] "Jack Scott"

# Question 2

fnFrame <- function(y) {  
   
 theDF <- data.frame( tp100song(y) , tp100art(y) , rep(y,100) , c(1:100) )  
 names(theDF) <- c("Song","Artist","Week","Rank")  
 return(theDF)  
}  
  
q2ans <- sapply(c(head(dates),tail(dates)),fnFrame, simplify = FALSE)  
dfbind <- do.call(rbind,q2ans)  
head(dfbind)

## Song Artist Week  
## 1 Poor Little Fool Ricky Nelson 1958-08-04  
## 2 Patricia Perez Prado And His Orchestra 1958-08-04  
## 3 Splish Splash Bobby Darin 1958-08-04  
## 4 Hard Headed Woman Elvis Presley With The Jordanaires 1958-08-04  
## 5 When Kalin Twins 1958-08-04  
## 6 Rebel-'rouser Duane Eddy His Twangy Guitar And The Rebels 1958-08-04  
## Rank  
## 1 1  
## 2 2  
## 3 3  
## 4 4  
## 5 5  
## 6 6

tail(dfbind)

## Song  
## 1195 Heartache On The Dance Floor  
## 1196 Too Much To Ask  
## 1197 B.E.D.  
## 1198 Bedroom Floor  
## 1199 Too Hotty  
## 1200 Three  
## Artist Week Rank  
## 1195 Jon Pardi 2017-11-11 95  
## 1196 Niall Horan 2017-11-11 96  
## 1197 Jacquees 2017-11-11 97  
## 1198 Liam Payne 2017-11-11 98  
## 1199 Quality Control Featuring Quavo, Takeoff & Offset 2017-11-11 99  
## 1200 Future & Young Thug 2017-11-11 100

# Question 3

tp200alb <- function(y) {  
 mydoc3 <- htmlParse(paste("~/290/gp08g2/scratch/billboardtop200albums/",y,sep=""))  
 myalbums <- xpathSApply(mydoc3, "//\*/div[@class='chart-row\_\_title']/h2[@class='chart-row\_\_song']", xmlValue)  
 return(myalbums)  
}  
tp200art <- function(y) {  
 mydoc4 <- htmlParse(paste("~/290/gp08g2/scratch/billboardtop200albums/",y,sep=""))  
 myartists <- xpathSApply(mydoc4, "//\*/div[@class='chart-row\_\_title']/span[@class='chart-row\_\_artist']|//\*/div[@class='chart-row\_\_title']/a[@class='chart-row\_\_artist']", xmlValue)   
 myartists <- sub("^\\s+", "", myartists)  
 myartists <- sub("\\s+$", "", myartists)  
 return(myartists)  
}  
fn3Frame <- function(y) {  
 theDF3 <- data.frame( tp200alb(y) , tp200art(y) , rep(y,200) , c(1:200) )  
 names(theDF3) <- c("Album","Artist","Week","Rank")  
 return(theDF3)  
}  
n3dates <- seq( as.Date("1968-01-06"), as.Date("2017-11-11"), "week")  
num3b <- sapply(c(head(n3dates),tail(n3dates)),fn3Frame,simplify=FALSE)  
df3bnd <- do.call(rbind,num3b)  
head(df3bnd)

## Album Artist  
## 1 Magical Mystery Tour (Soundtrack) The Beatles  
## 2 Their Satanic Majesties Request The Rolling Stones  
## 3 Pisces, Aquarius, Capricorn, And Jones Ltd. The Monkees  
## 4 Diana Ross And The Supremes Greatest Hits Diana Ross & The Supremes  
## 5 Sgt. Pepper's Lonely Hearts Club Band The Beatles  
## 6 Doctor Zhivago Soundtrack  
## Week Rank  
## 1 1968-01-06 1  
## 2 1968-01-06 2  
## 3 1968-01-06 3  
## 4 1968-01-06 4  
## 5 1968-01-06 5  
## 6 1968-01-06 6

tail(df3bnd)

## Album Artist  
## 2395 The Perfect LUV Tape Lil Uzi Vert  
## 2396 Recovery Eminem  
## 2397 + Ed Sheeran  
## 2398 Under Pressure Logic  
## 2399 Greatest Hits Foo Fighters  
## 2400 I Will Always Love You: The Best Of Whitney Houston Whitney Houston  
## Week Rank  
## 2395 2017-11-11 195  
## 2396 2017-11-11 196  
## 2397 2017-11-11 197  
## 2398 2017-11-11 198  
## 2399 2017-11-11 199  
## 2400 2017-11-11 200

# Question 4

## Dylan

### Are songs often on the top 100 for more than one week?

isRockstar <- function(y) {  
 tp100song(y) == "Rockstar"  
}  
  
n4dates <- seq( as.Date("2017-09-16"), as.Date("2017-11-11"), "week")  
num4 <- sapply(n4dates,isRockstar,simplify=FALSE)  
num4 <- unlist(num4)  
num4ans <- fractions(sum(num4, TRUE) /8 )  
num4ans

## [1] 7/8

### We see that the song "Rockstar" was in the top 100 for 7 of 8 weeks since its release.

# Question 5

## Dylan

### What artist, from 1958 to 2017, holds the record for most weeks in the top 100?

num1d <- sapply(dates,function(y) tp100art(y),simplify = FALSE) tail( sort( table( unlist( num1d) ) ), n=2 ) This code takes a LONG time to run! The code works, and returns Elton John with 889 weeks in the top100!

# Question 6

## Meera

### Did the song "Panda" make it into the top 10 in 2016?

n6dates <- seq( as.Date("2016-01-02"), as.Date("2016-12-24"), "week")  
num6 <- sapply(n6dates,function(y) head(tp100song(y),n=10),simplify=FALSE)  
df6bnd <- do.call(rbind,num6)  
table(df6bnd == "Panda")

##   
## FALSE TRUE   
## 503 17

True represents the number of times the song 'Panda' was in the top ten, which is 17 times.

# Question 7

## Meera

### Who were the top 10 artists in 2014?

n7dates <- seq( as.Date("2014-01-04"), as.Date("2014-12-27"), "week")  
num7 <- sapply(n7dates,function(y) tp100art(y),simplify = FALSE)  
df7bnd <- do.call(rbind,num7)  
tail(sort(table(df7bnd)),n=10)

## df7bnd  
## John Legend Sam Smith One Direction Luke Bryan   
## 51 53 59 60   
## Ed Sheeran Pharrell Williams Katy Perry Lorde   
## 61 62 64 65   
## OneRepublic Imagine Dragons   
## 65 72

The number associated with each artist represents the number of weeks they were in the top 100.

# Question 8

## Meera

### What was the most popular song in 2005?

n8dates <- seq( as.Date("2005-01-01"), as.Date("2005-12-31"),"week")  
num8 <- sapply(n8dates,function(y) tp100song(y),simplify=FALSE)  
df8bnd <- do.call(rbind,num8)  
tail(sort(table(df8bnd)),n=1)

## You And Me   
## 46

The song 'You and Me' is the longest lasting song in the top 100 for 2005. #Question 9 ##Meera ###How many songs in 2006 were in the top 100 for only one week?

n9dates <- seq( as.Date("2006-01-07"), as.Date("2006-12-30"),"week")  
num9 <- sapply(n9dates,function(y) tp100song(y),simplify=FALSE)  
df9bnd <- do.call(rbind,num9)  
sum( sort(table(df9bnd)) == "1" )

## [1] 53

# Question 10

## Molly

### What was the number one track for 2016-08-06? What number was that track for 2016-09-03, if it was on the chart at all?

tp100song <- function(y) {  
 mydoc <- htmlParse(paste("~/290/gp08g2/scratch/billboardhot100/",y,sep=""))  
 mysongs <- xpathSApply(mydoc, "//\*/div[@class='chart-row\_\_title']/h2[@class='chart-row\_\_song']", xmlValue)  
 return(mysongs)  
}  
head(tp100song("2016-08-06"),n=1)

## [1] "Cheap Thrills"

head(tp100song("2016-09-03"),n=2)

## [1] "Closer" "Cheap Thrills"

The song 'Cheap Thrills' was number one for 2016-06-08, and dropped only to second place by 2016-09-03.

# Question 11

## Molly

### Which artist had the most hits in the top 100 for 2016-08-06?

tp100art <- function(y) {  
 mydoc1 <- htmlParse(paste("~/290/gp08g2/scratch/billboardhot100/",y,sep=""))  
 myartists <- xpathSApply(mydoc1, "//\*/div[@class='chart-row\_\_title']/span[@class='chart-row\_\_artist']|//\*/div[@class='chart-row\_\_title']/a[@class='chart-row\_\_artist']", xmlValue)   
 myartists <- sub("^\\s+", "", myartists)  
 myartists <- sub("\\s+$", "", myartists)  
 return(myartists)  
}  
head(tp100art("2016-08-06"),n=1)

## [1] "Sia Featuring Sean Paul"

# Question 12

## Molly

### Was the song that was the number 5 hit on 2016-08-06 higher or lower on the top 100 on 2016-08-27?

head(tp100song("2016-08-06"))

## [1] "Cheap Thrills" "One Dance"   
## [3] "This Is What You Came For" "Can't Stop The Feeling!"   
## [5] "Don't Let Me Down" "Ride"

head(tp100song("2016-08-27"))

## [1] "Cheap Thrills" "Cold Water"   
## [3] "This Is What You Came For" "Heathens"   
## [5] "One Dance" "Closer"

If we look at the entire tp100song for the week rather than the head, we can tell the artist Drake is featured 10 times.