ExploreSpotify

Melvis Onoriode

2022-05-18

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
spotify<-read.csv("data.csv",sep="#")
head(spotify)
```

```
##   Position                              Track.Name           Artist Streams
## 1        1                                 Starboy       The Weeknd 3135625
## 2        2                                  Closer The Chainsmokers 3015525
## 3        3                          Let Me Love You         DJ Snake 2545384
## 4        4 Rockabye (feat. Sean Paul & Anne-Marie)     Clean Bandit 2356604
## 5        5                               One Dance            Drake 2259887
## 6        6                               Fake Love            Drake 2137437
##         Date
## 1 2017-01-01
## 2 2017-01-01
## 3 2017-01-01
## 4 2017-01-01
## 5 2017-01-01
## 6 2017-01-01
##                                                                                Genre
## 1                                    ['canadian pop', 'canadian contemporary r&b', 'pop']
```

```
## 2                           ['pop', 'pop dance', 'tropical house', 'edm', 'electropop', 'dance pop']
## 3                             ['pop', 'electronic trap', 'dance pop', 'edm', 'pop dance', 'pop rap']
## 4 ['pop', 'uk dance', 'dance pop', 'uk funky', 'tropical house', 'pop dance', 'post-teen pop', 'edm']
## 5                   ['toronto rap', 'canadian pop', 'canadian hip hop', 'rap', 'pop rap', 'hip hop']
## 6                   ['toronto rap', 'canadian pop', 'canadian hip hop', 'rap', 'pop rap', 'hip hop']
```

```
spotify1<-mutate(spotify,Year=year(spotify$Date))

year17<-spotify[spotify1$Year==2017,]
year18<-spotify[spotify1$Year==2018,]
year19<-spotify[spotify1$Year==2019,]
year20<-spotify[spotify1$Year==2020,]
year21<-spotify[spotify1$Year==2021,]
```

```
paste("Total Songs in the Database: ",length(unique(spotify1$Track.Name)))
```

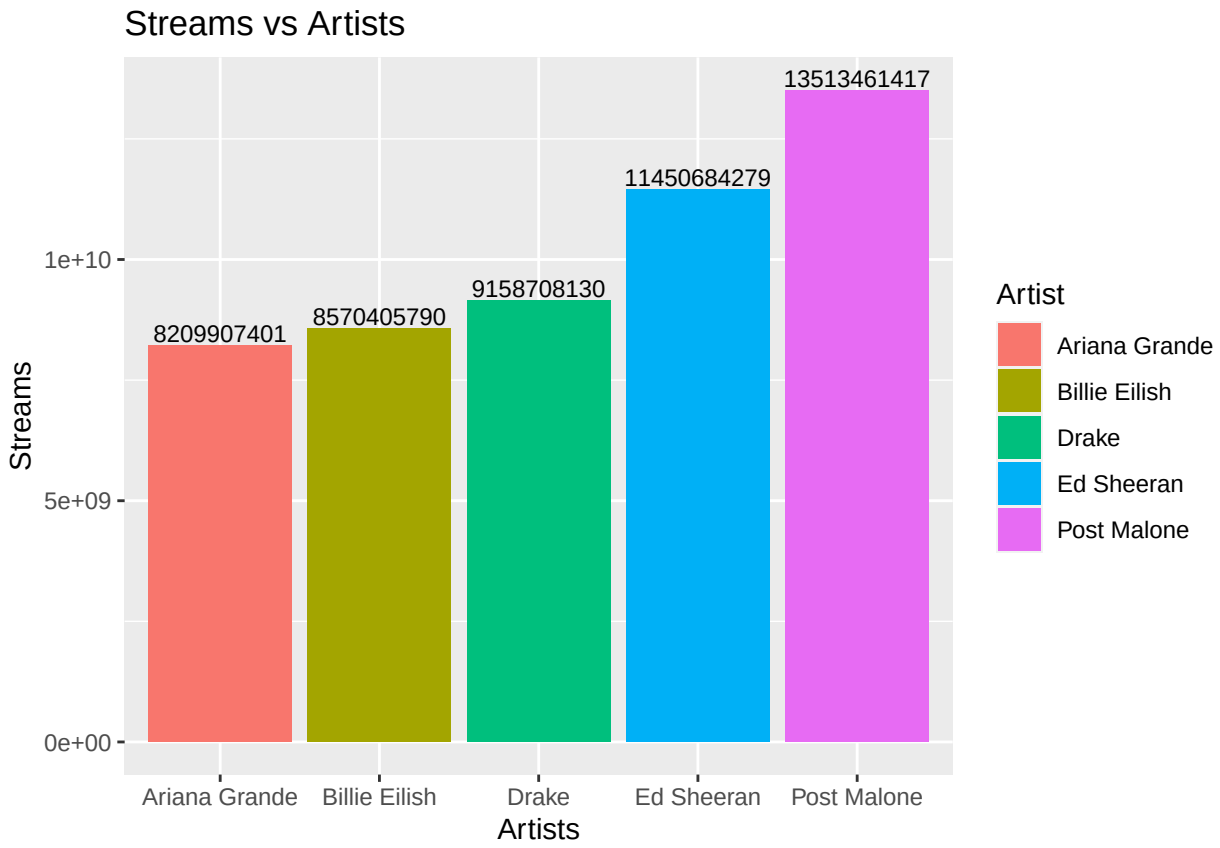
```
## [1] "Total Songs in the Database:  5497"
```

```
paste("Total Artists in the Database: ",length(unique(spotify1$Artist)))
```

```
## [1] "Total Artists in the Database:  1128"
```

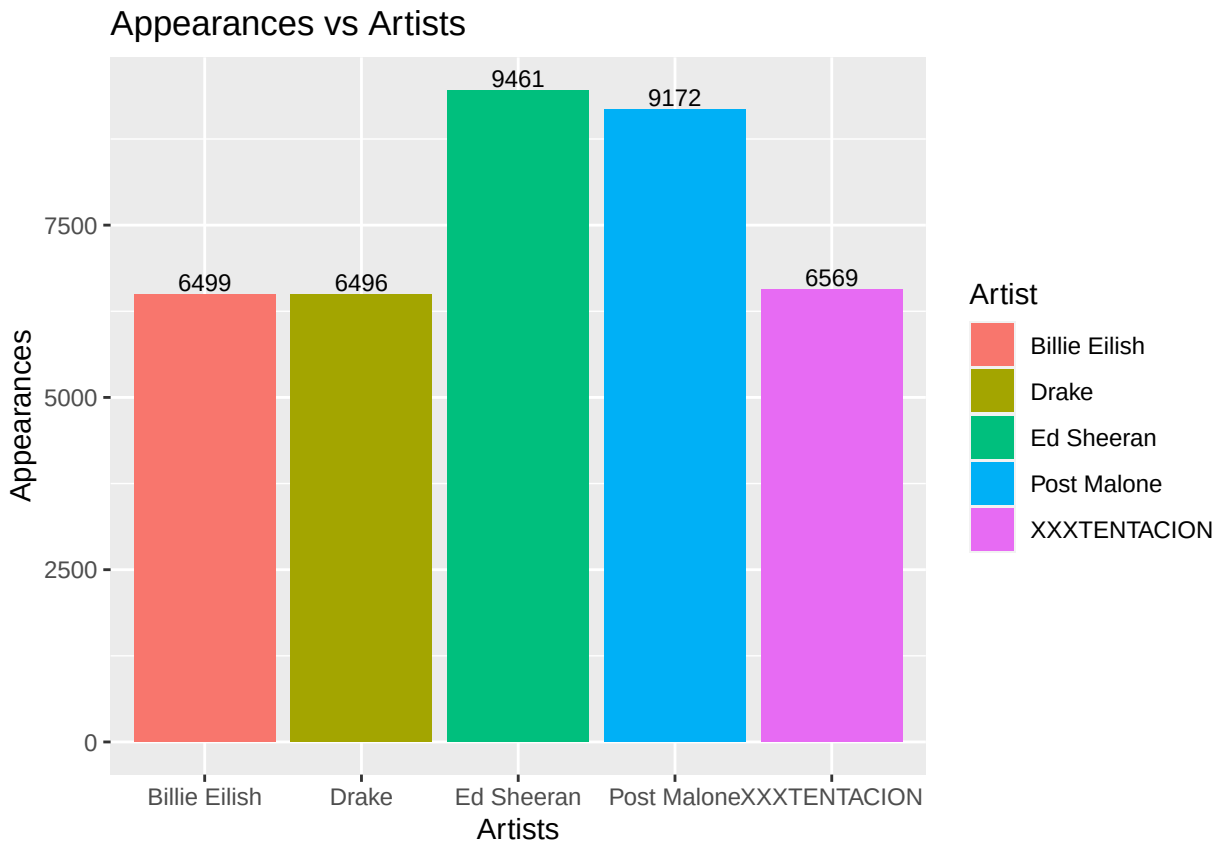
```
# The Total number of Streams of an Artist
streams<-aggregate(x=spotify1$Streams,by=list(spotify1$Artist),FUN=sum)
    colnames(streams)<-c("Artist","Streams")
    streams<-streams[order(streams$Streams,decreasing=T),]

    ggplot(data=streams[1:5,],aes(x=Artist,y=Streams,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=Streams),vjust=-0.2,size=3)+
      xlab("Artists")+
      ylab("Streams")+
      ggtitle("Streams vs Artists")
```



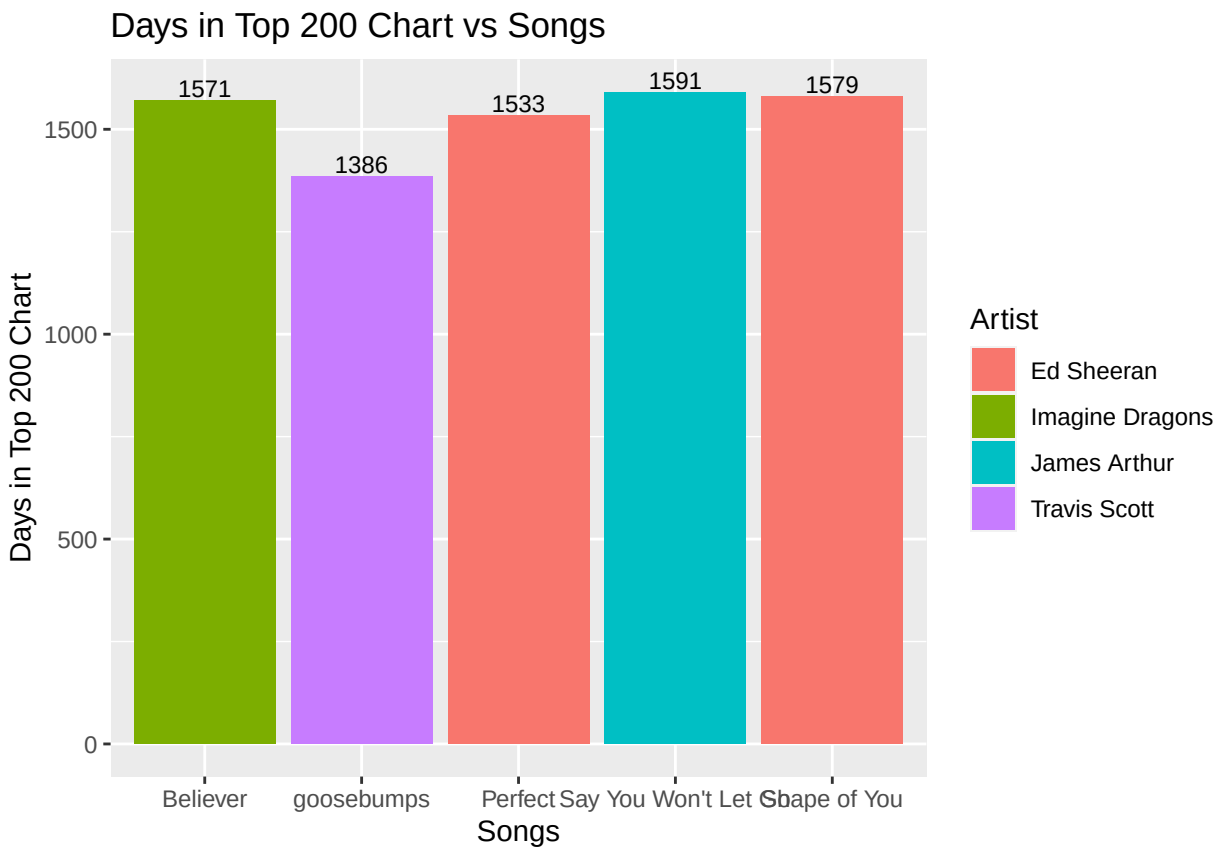

```
# Number of Appearances of Artist in the 'Corresponding Data'
artistapp<-as.data.frame(table(spotify1$Artist))
    colnames(artistapp)<-c("Artist","No._of_Appearance_in_Top_200_Chart")
    artistapp<-artistapp[order(artistapp$`No._of_Appearance_in_Top_200_Chart`,decreasing=T),]

    ggplot(data=artistapp[1:5,],aes(x=Artist,y=No._of_Appearance_in_Top_200_Chart,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=No._of_Appearance_in_Top_200_Chart),vjust=-0.2,size=3)+
      xlab("Artists")+
      ylab("Appearances")+
      ggtitle("Appearances vs Artists")
```

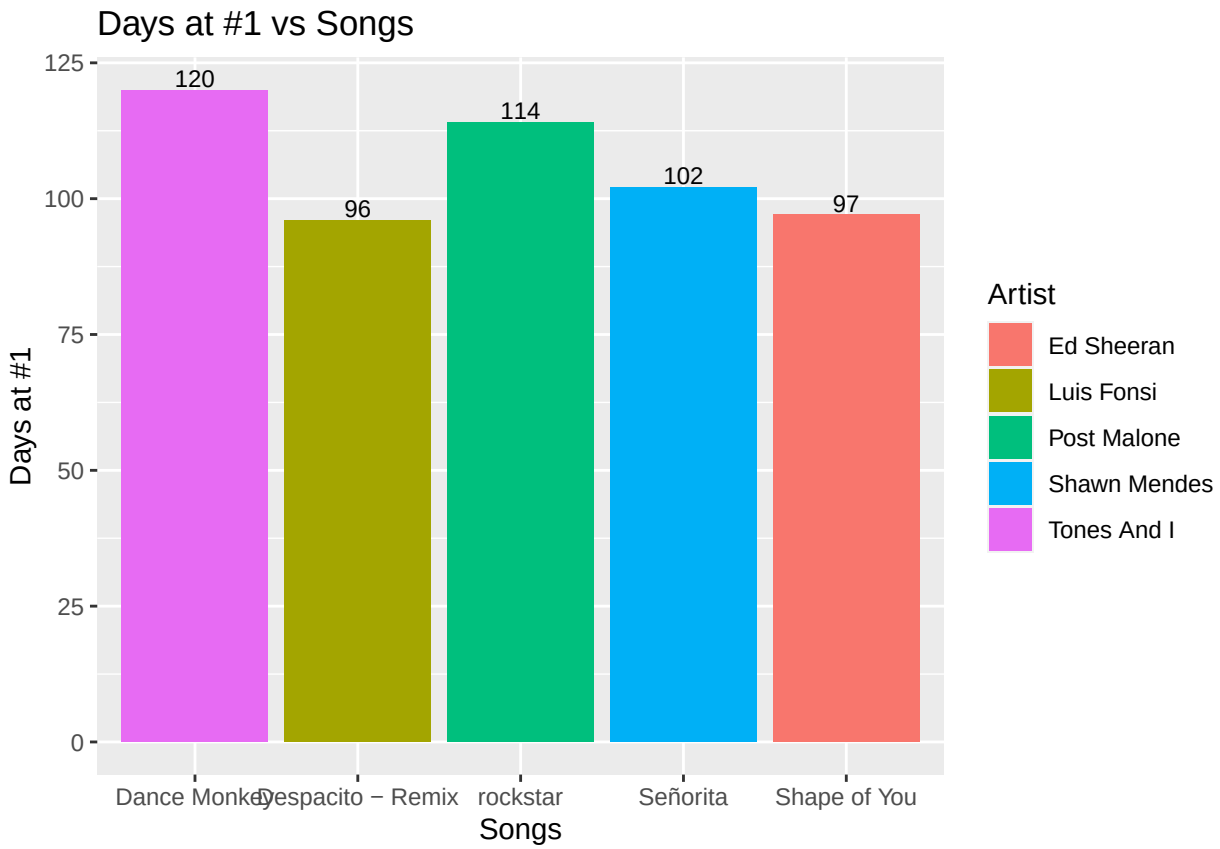



```
#No. of days a song stays in Top 200.
songs<-spotify$Track.Name
    artname<-spotify[spotify1$Track.Name==songs,]$Artist
    days<-data.frame("Artist"=artname,"Song"=songs)
    days<-aggregate(x=days$Song,by=list(days$Artist,days$Song),FUN=length)
    days<-as.data.frame(days)
    colnames(days)<-c("Artist","Song","No._of_Days")
    days<-days[order(days$No._of_Days,decreasing=T),]

    ggplot(data=days[1:5,],aes(x=Song,y=No._of_Days,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=No._of_Days),vjust=-0.2,size=3)+
      xlab("Songs")+
      ylab("Days in Top 200 Chart")+
      ggtitle("Days in Top 200 Chart vs Songs")
```

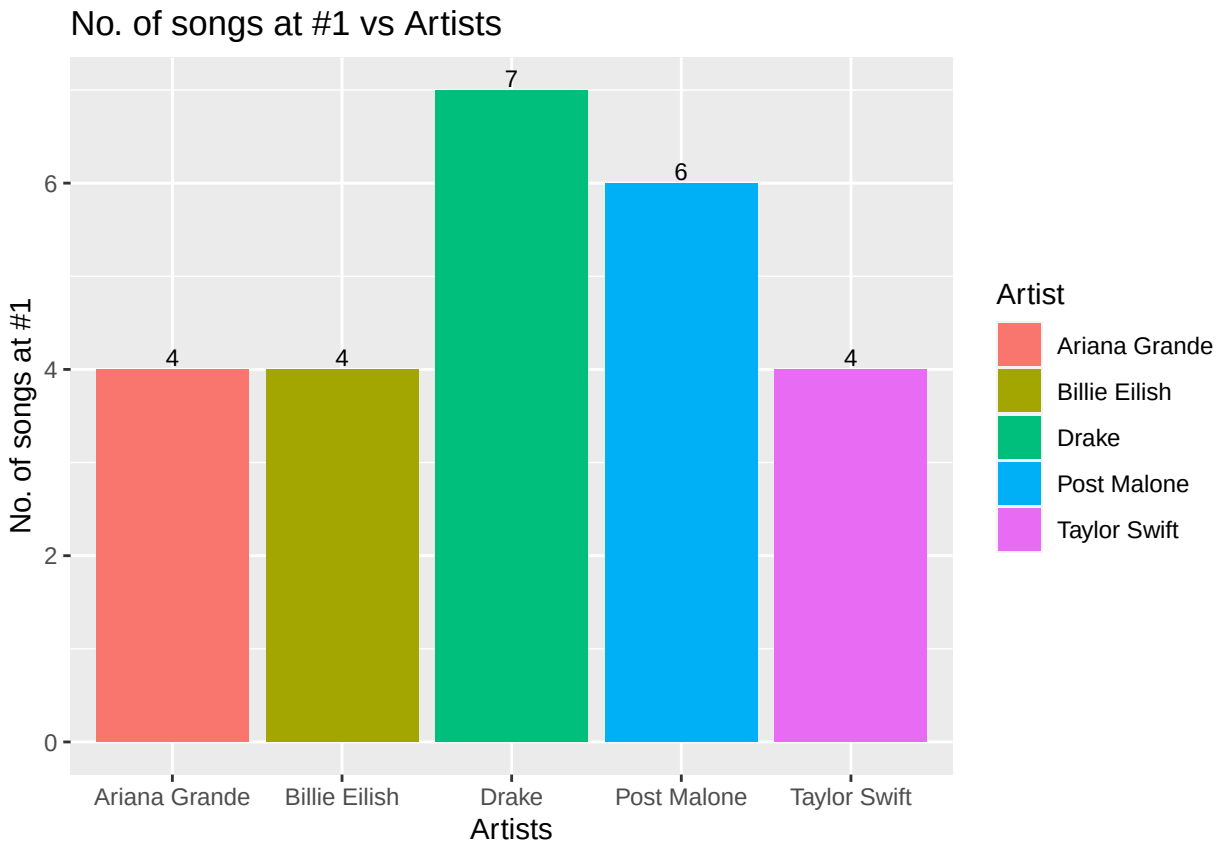



```
#No. of days a song stays at No. 1 in Top 200 Charts.
top<-spotify[spotify1$Position==1,]
    numeruno<-aggregate(x=top$Position,by=list(top$Track.Name,top$Artist),FUN=sum)
    colnames(numeruno)<-c("Song","Artist","Days_at_1")
    numeruno<-numeruno[order(numeruno$`Days_at_1`,decreasing = T),]
      ggplot(data=numeruno[1:5,],aes(x=Song,y=Days_at_1,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=Days_at_1),vjust=-0.2,size=3)+
      xlab("Songs")+
      ylab("Days at #1")+
      ggtitle("Days at #1 vs Songs")
```




```
#  Artists with most No. 1 songs
mostnumerunos<-as.data.frame(table(numeruno$Artist))
    colnames(mostnumerunos)<-c("Artist","No._of_Songs_at_1")
    mostnumerunos<-mostnumerunos[order(mostnumerunos$`No._of_Songs_at_1`,decreasing = T),]

    ggplot(data=mostnumerunos[1:5,],aes(x=Artist,y=No._of_Songs_at_1,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=No._of_Songs_at_1),vjust=-0.2,size=3)+
      xlab("Artists")+
      ylab("No. of songs at #1")+
      ggtitle("No. of songs at #1 vs Artists")
```




```
# Year 2019 in a Nutshell

    paste("Total Songs that featured in Top 200 in 2019: ",length(unique(spotify1$Track.Name)))
```

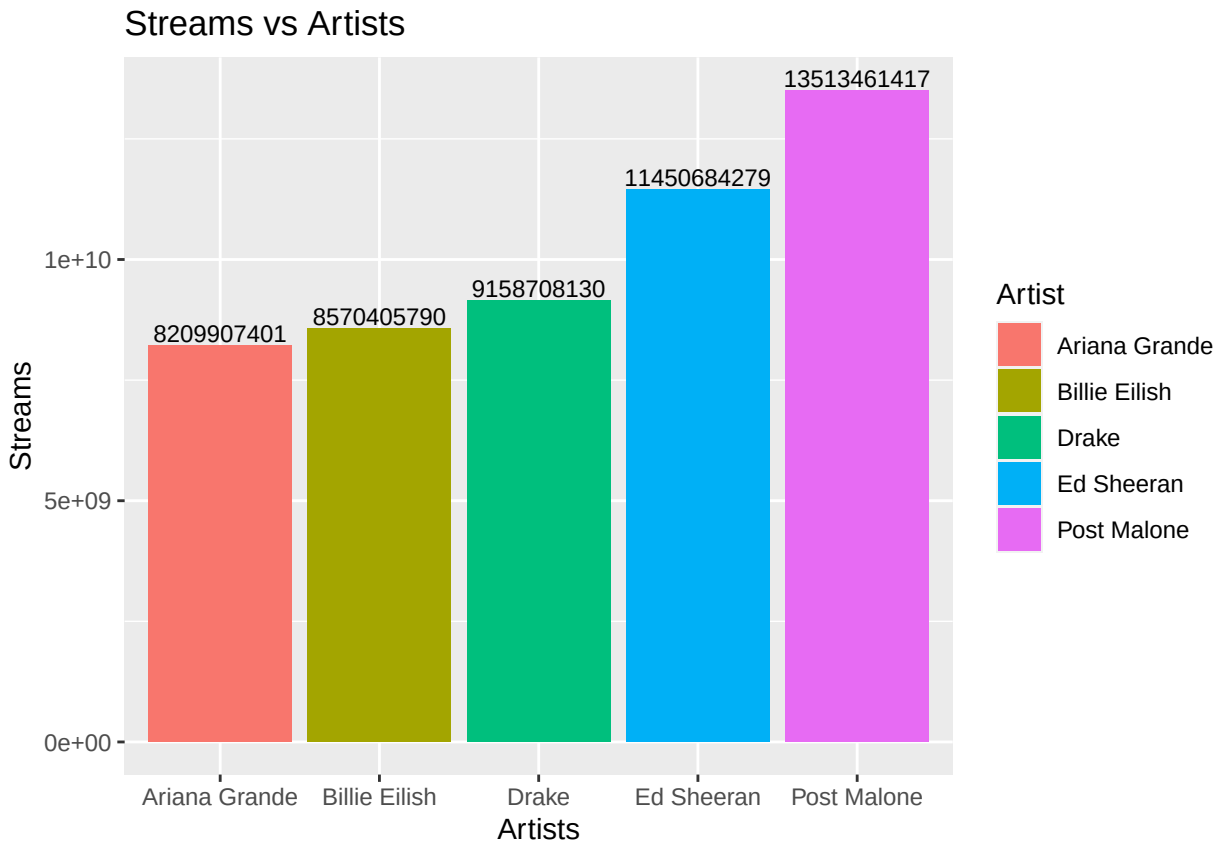
```
## [1] "Total Songs that featured in Top 200 in 2019:  5497"
```

```
    paste("Total Artists that featured in Top 200 in 2019: ",length(unique(spotify1$Artist)))
```

```
## [1] "Total Artists that featured in Top 200 in 2019:  1128"
```

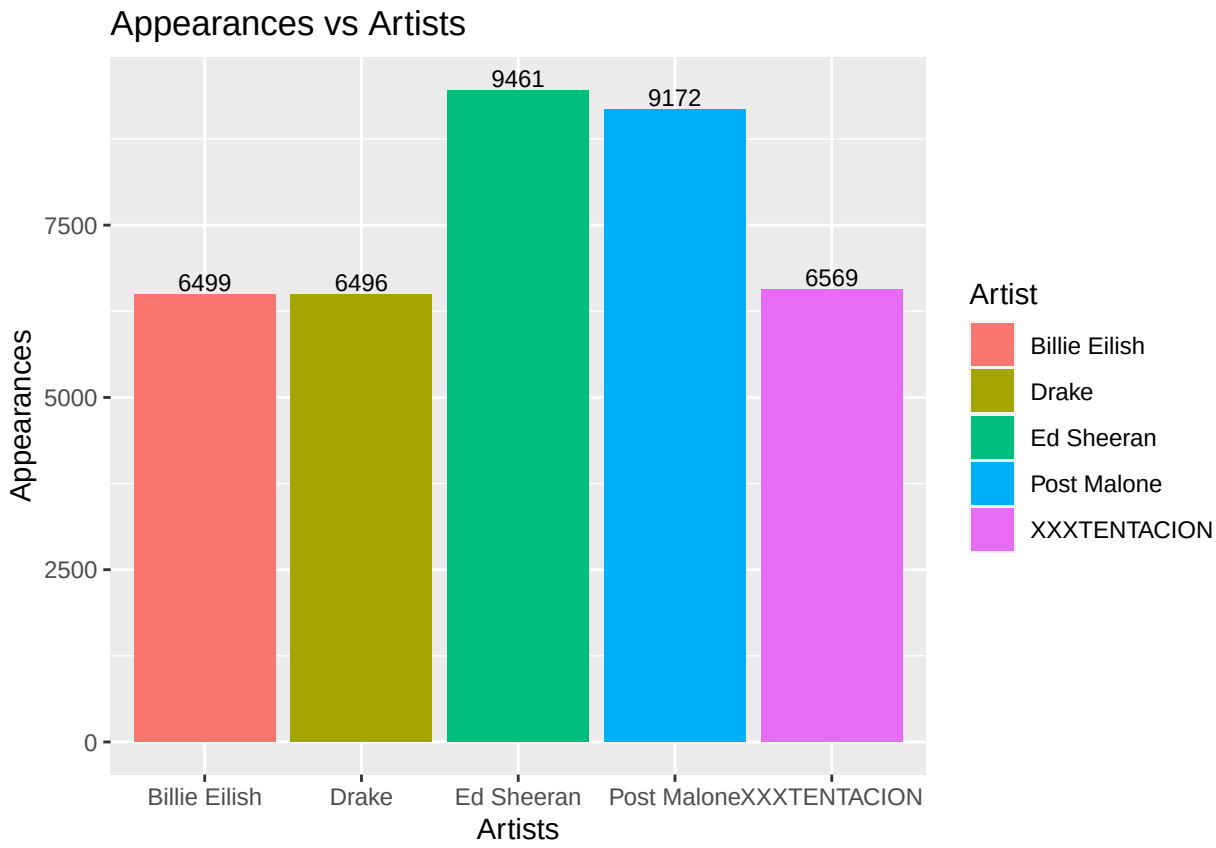
```
    streams<-aggregate(x=spotify1$Streams,by=list(spotify1$Artist),FUN=sum)
    colnames(streams)<-c("Artist","Streams")
    streams<-streams[order(streams$Streams,decreasing=T),]

    ggplot(data=streams[1:5,],aes(x=Artist,y=Streams,fill=Artist))+
      geom_bar(stat="Identity")+
      geom_text(aes(label=Streams),vjust=-0.2,size=3)+
      xlab("Artists")+
      ylab("Streams")+
      ggtitle("Streams vs Artists")
```



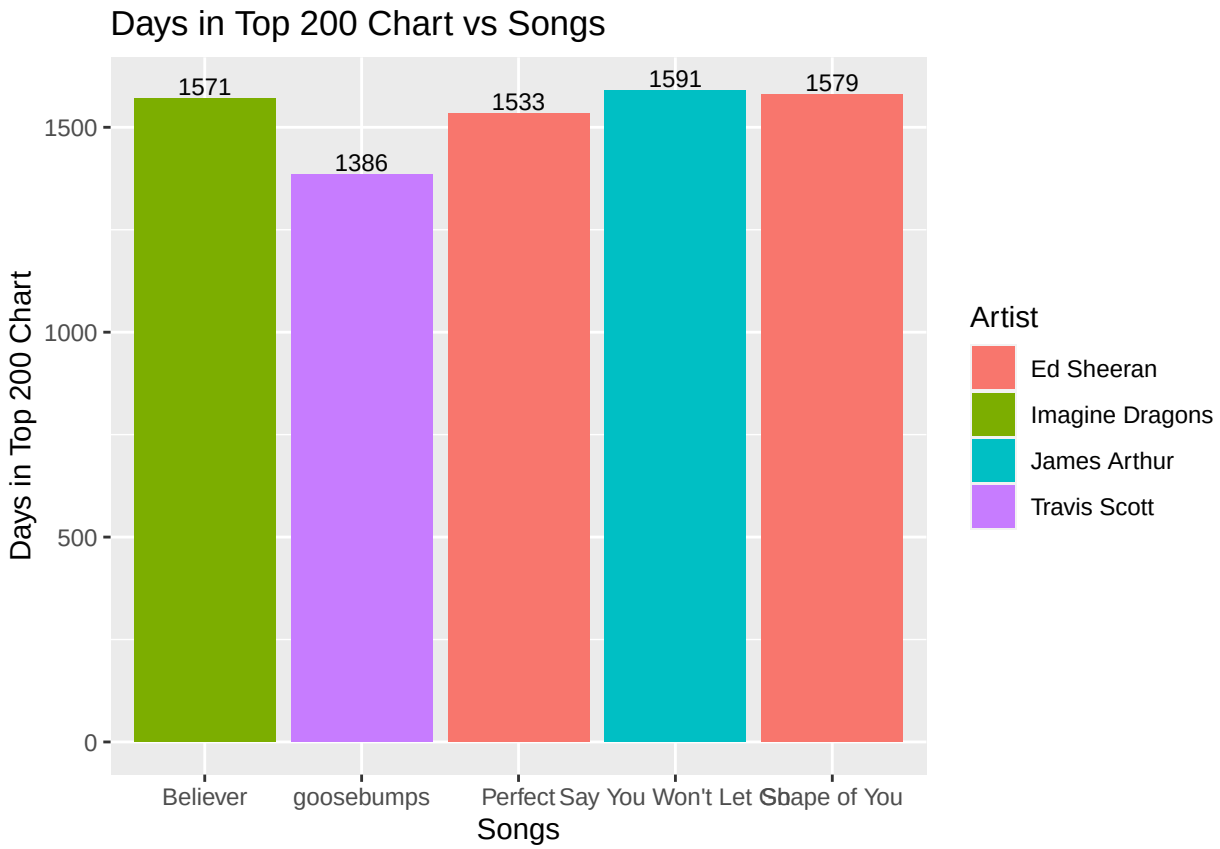

```
artistapp<-as.data.frame(table(spotify1$Artist))
colnames(artistapp)<-c("Artist","No._of_Appearance_in_Top_200_Chart")
artistapp<-artistapp[order(artistapp$No._of_Appearance_in_Top_200_Chart,decreasing=T),]

ggplot(data=artistapp[1:5,],aes(x=Artist,y=No._of_Appearance_in_Top_200_Chart,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=No._of_Appearance_in_Top_200_Chart),vjust=-0.2,size=3)+
  xlab("Artists")+
  ylab("Appearances")+
  ggtitle("Appearances vs Artists")
```



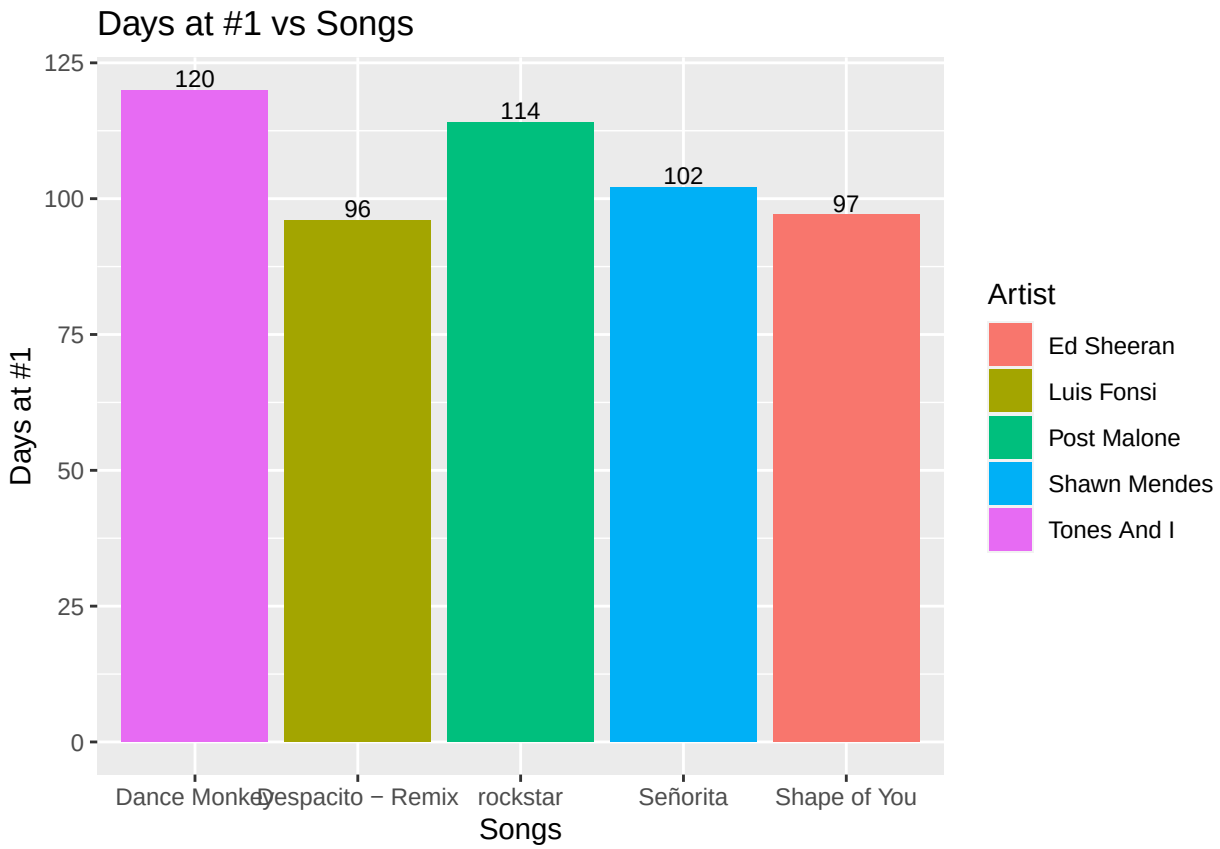

```
songs<-spotify1$Track.Name
artname<-spotify1[spotify1$Track.Name==songs,]$Artist
days<-data.frame("Artist"=artname,"Song"=songs)
days<-aggregate(x=days$Song,by=list(days$Artist,days$Song),FUN=length)
days<-as.data.frame(days)
colnames(days)<-c("Artist","Song","No._of_Days")
days<-days[order(days$No._of_Days,decreasing=T),]

ggplot(data=days[1:5,],aes(x=Song,y=No._of_Days,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=No._of_Days),vjust=-0.2,size=3)+
  xlab("Songs")+
  ylab("Days in Top 200 Chart")+
  ggtitle("Days in Top 200 Chart vs Songs")
```



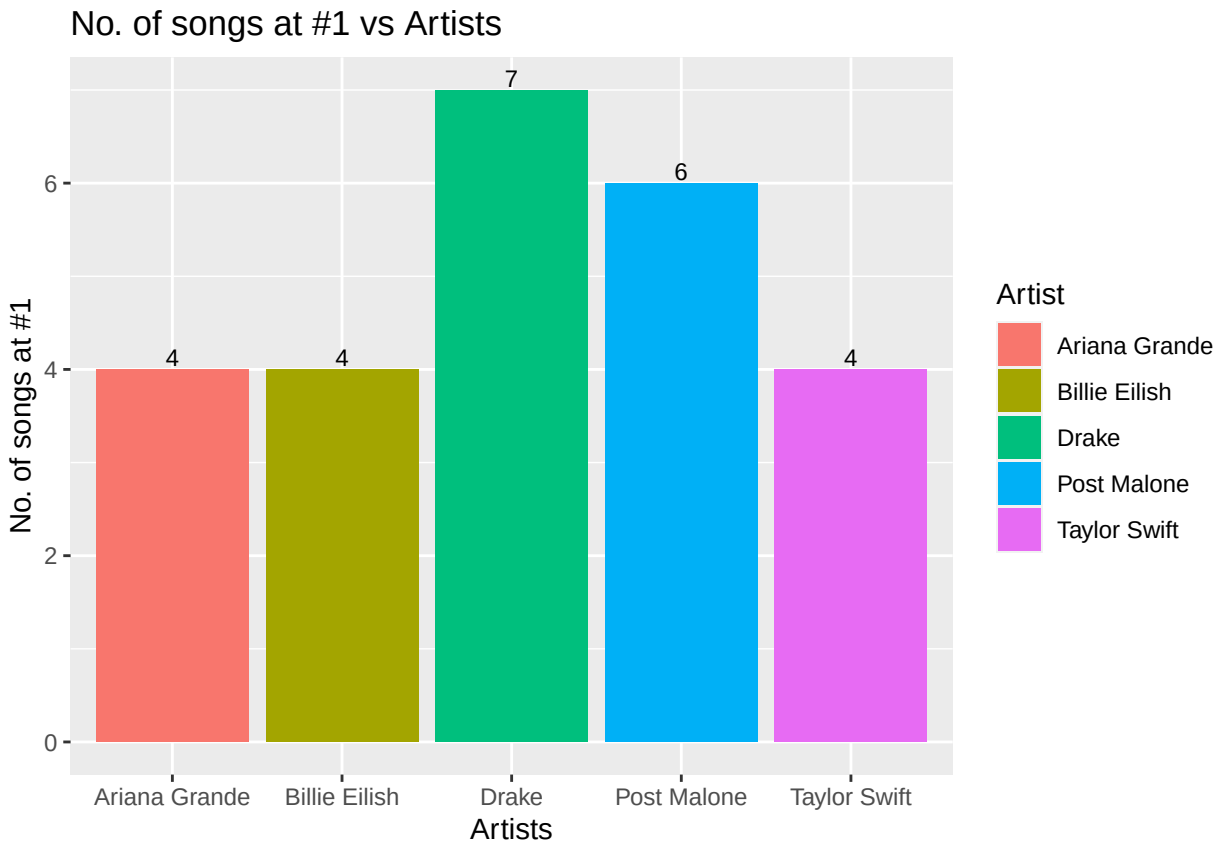

```
top<-spotify1[spotify1$Position==1,]
numeruno<-aggregate(x=top$Position,by=list(top$Track.Name,top$Artist),FUN=sum)
colnames(numeruno)<-c("Song","Artist","Days_at_1")
numeruno<-numeruno[order(numeruno$`Days_at_1`,decreasing = T),]

ggplot(data=numeruno[1:5,],aes(x=Song,y=Days_at_1,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=Days_at_1),vjust=-0.2,size=3)+
  xlab("Songs")+
  ylab("Days at #1")+
  ggtitle("Days at #1 vs Songs")
```



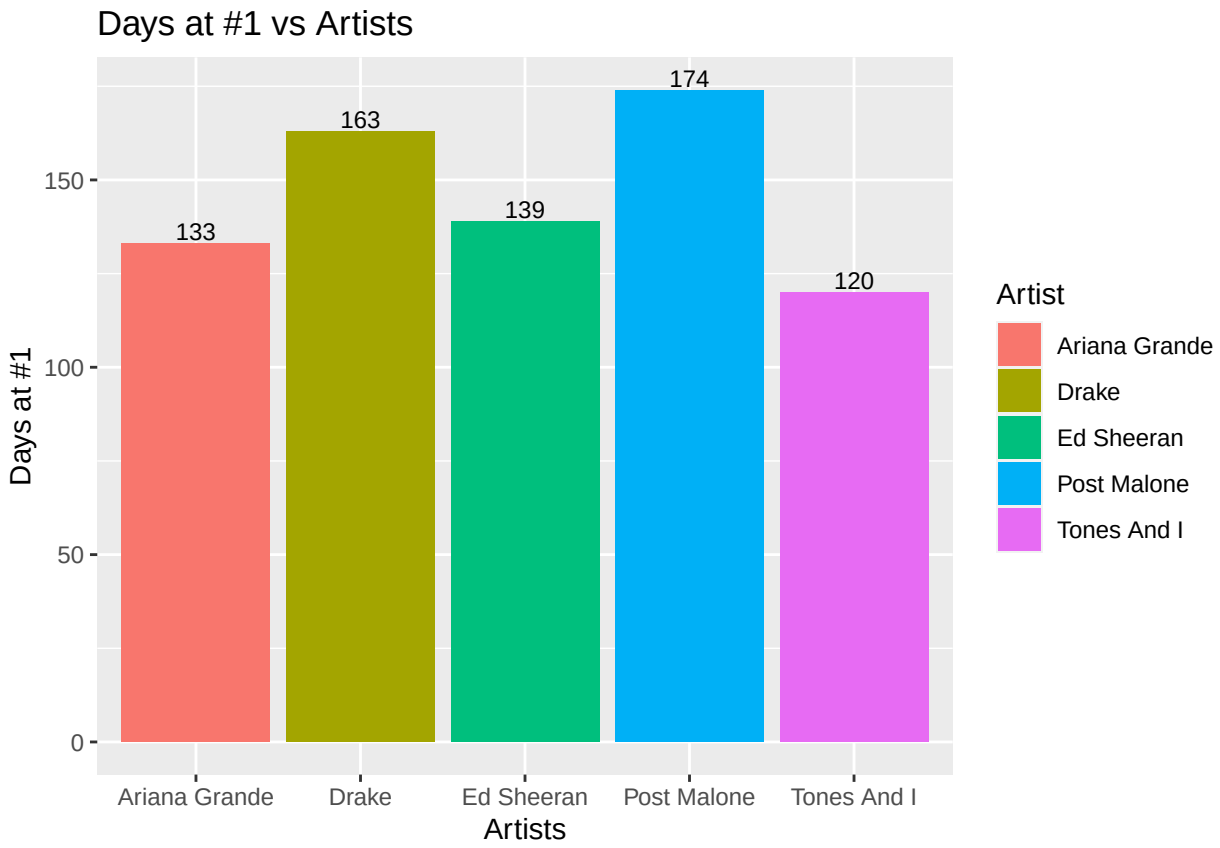

```
mostnumerunos<-as.data.frame(table(numeruno$Artist))
colnames(mostnumerunos)<-c("Artist","No._of_Songs_at_1")
mostnumerunos<-mostnumerunos[order(mostnumerunos$`No._of_Songs_at_1`,decreasing = T),]

ggplot(data=mostnumerunos[1:5,],aes(x=Artist,y=No._of_Songs_at_1,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=No._of_Songs_at_1),vjust=-0.2,size=3)+
  xlab("Artists")+
  ylab("No. of songs at #1")+
  ggtitle("No. of songs at #1 vs Artists")
```



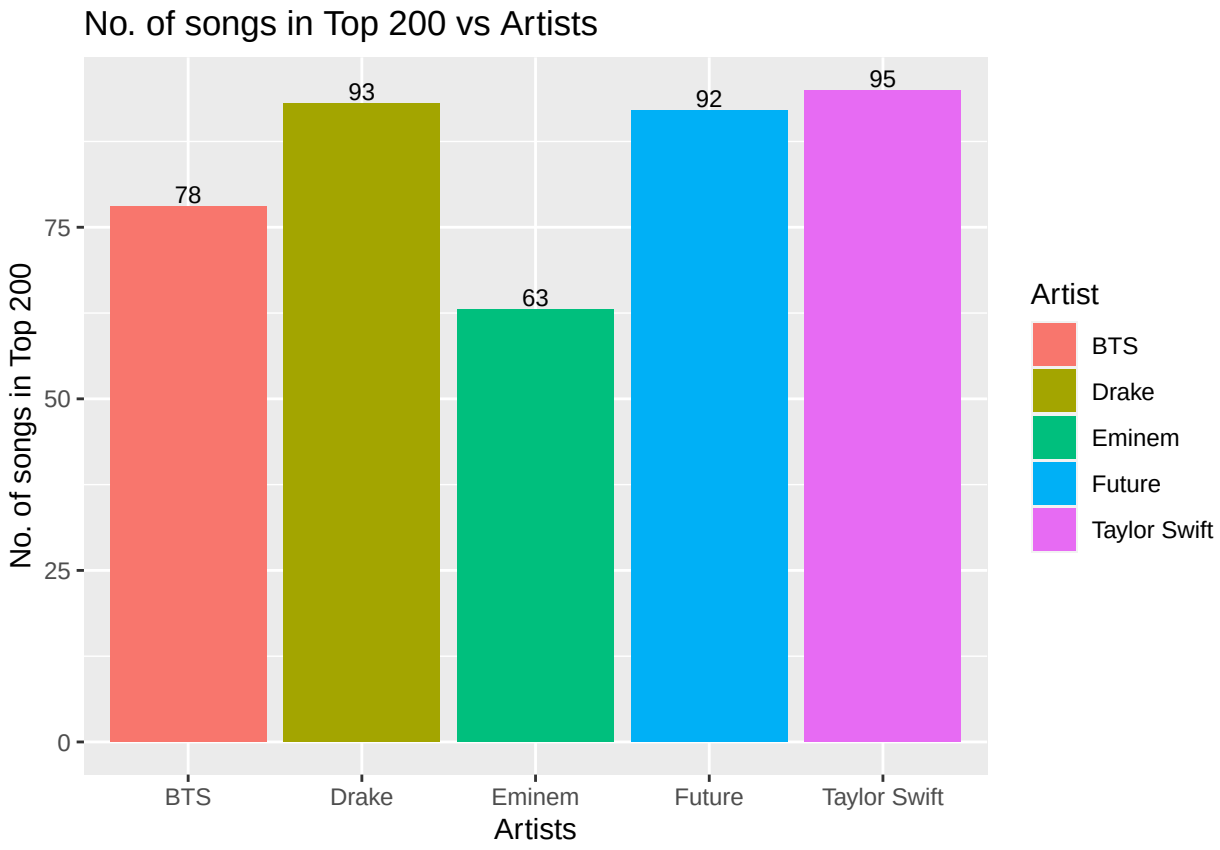

```
numerunodays<-aggregate(x=numeruno$Days_at_1,by=list(numeruno$Artist),FUN=sum)
colnames(numerunodays)<-c("Artist","Days_at_1")
numerunodays<-numerunodays[order(numerunodays$Days_at_1,decreasing=T),]

ggplot(data=numerunodays[1:5,],aes(x=Artist,y=Days_at_1,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=Days_at_1),vjust=-0.2,size=3)+
  xlab("Artists")+
  ylab("Days at #1")+
  ggtitle("Days at #1 vs Artists")
```




```
art<-unique(spotify1$Artist)
count<-numeric(length(art))
for(i in 1:length(art)){
  count[i]<-length(unique(spotify1[spotify1$Artist==art[i],c("Track.Name")]))
}
artist<-data.frame(Artist=art,No._of_Songs_in_Top_200=count)
artist<-artist[order(artist$No._of_Songs_in_Top_200,decreasing=T),]


ggplot(data=artist[1:5,],aes(x=Artist,y=No._of_Songs_in_Top_200,fill=Artist))+
  geom_bar(stat="Identity")+
  geom_text(aes(label=No._of_Songs_in_Top_200),vjust=-0.2,size=3)+
  xlab("Artists")+
  ylab("No. of songs in Top 200")+
  ggtitle("No. of songs in Top 200 vs Artists")
```


No. of songs in Top 200 vs Artists
No. of songs in Top 200
0
25
50
75
78
93
63
92
95
BTS
Drake
Eminem
Future
Taylor Swift
Artists
Artist
BTS
Drake
Eminem
Future
Taylor Swift