Project report

Max Narvaez & Ryan Harvey

9/24/2020

Article Review

This article examines the recognition of hit 90's songs today in order to see how 90's music will be characterized by future generations. The article gathered data through a music quiz where participants were asked if they recognized certain Top 100 Billboard songs from the 90's. Different charts were produced to determine how songs stand the test of time and which songs are most recognized today.

Description of the Dataset

Using a markdown table describe the variables from your dataset that you used for your analysis and whether they were in the original table or were obtained by transforming the original dataset. For variables that were not in the original dataset, briefly describe how you obtain them.

Variable Name	Description
artist_song	Artist name and song title
generation	Number of years until birth of subject when the song was released
recognition	Proportion of subjects that recognized the song
latest.recognition	Latest recognition data point (songs were debuted in different years, this is the last data point for a song)
diff.from.trend	The difference between the song's popularity and the average popularity for that age

The first three variables were from the original datasets.

Helper functions used later

```
flip_sign <- function(l) {
     l <- as.numeric(l)*(-1)
     parse(text=l)
}

last_non_na_value <- function(r) {
    cols <- length(r)
    col <- cols - 1
    while (is.na(r[col])) {
        col <- col - 1
    }
    r[col]
}
```

Prepare Dataset

```
x <- nrow(rec)
while (x > 0) {
  y <- 25
  while (y > 1) {
```

```
    if (rec[x, y] == 0) {
      rec[x, y] <- NA
      y <- y - 1
    } else {
      break
    }
  }
  x <- x - 1
}
```

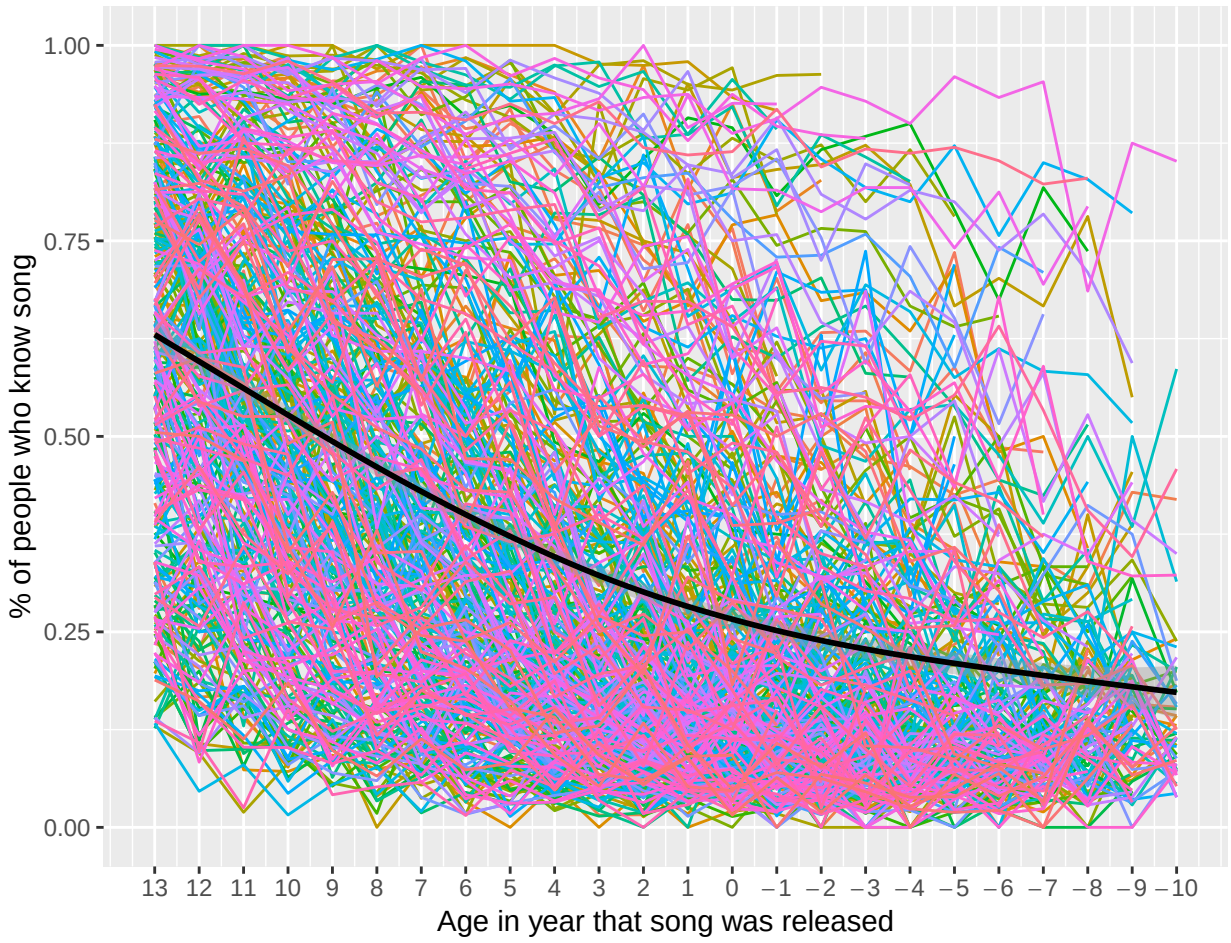
Main Conclusions

Use well-written paragraphs and professional-looking tables and/or graphs to reproduce the main results of your article of interest.

The percent of songs from the 90's recognized currently decays with time

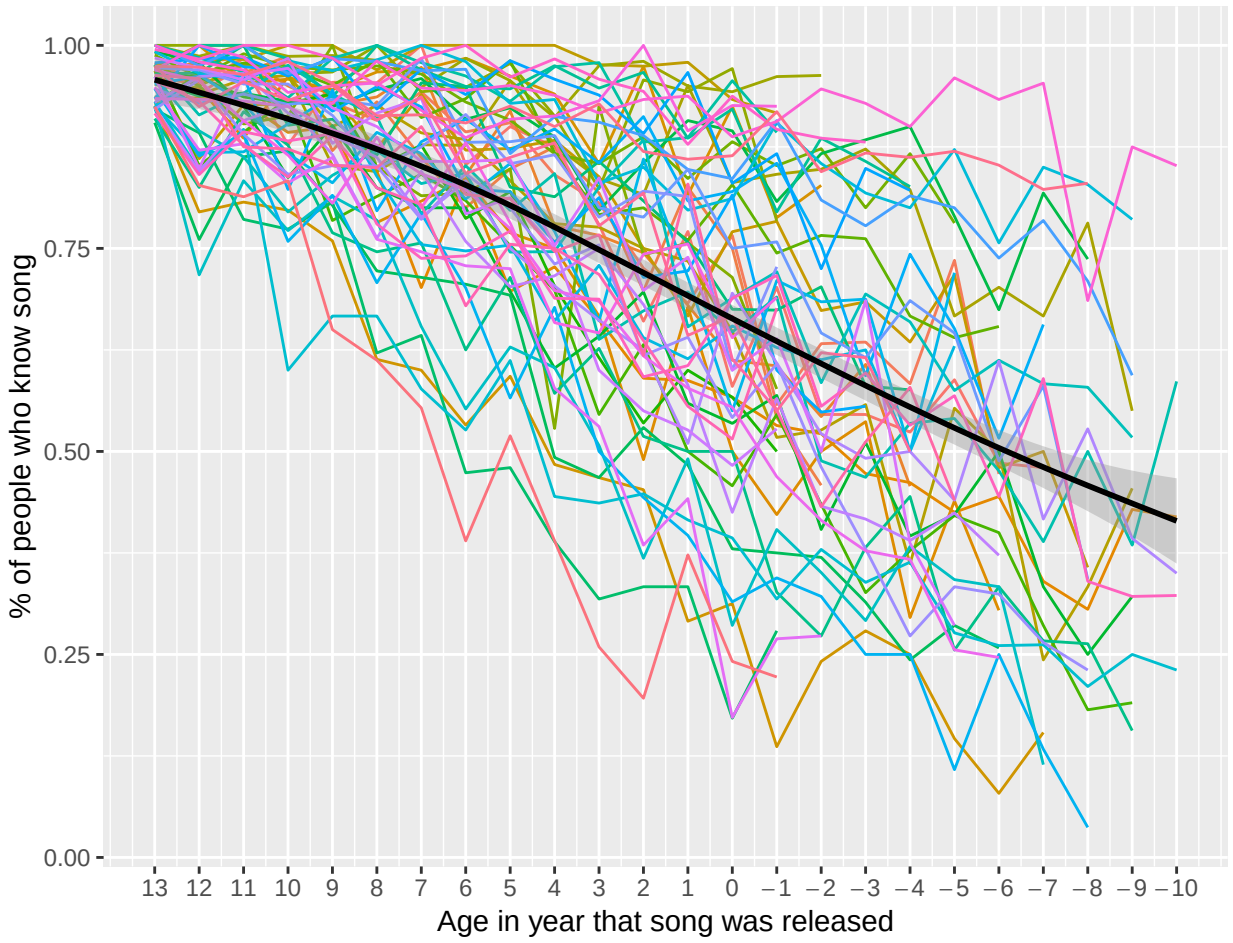
```
rec %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition)) %>%
    ggplot() +
    geom_line(aes(x = as.numeric(generation), recognition, color = song)) +
    geom_smooth(aes(as.numeric(generation), recognition), color = "black") +
    scale_x_continuous(breaks = seq(-13, 10, by = 1), labels = flip_sign) +
    theme(legend.position = "none") +
    labs(x = "Age in year that song was released",
         y = "% of people who know song")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
rec %>%
    filter(`-13` >= .9) %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition)) %>%
    ggplot() +
    geom_line(aes(x = as.numeric(generation), recognition, color = song)) +
    geom_smooth(aes(as.numeric(generation), recognition), color = "black") +
    scale_x_continuous(breaks = seq(-13, 10, by = 1), labels = flip_sign) +
    theme(legend.position = "none") +
    labs(x = "Age in year that song was released",
         y = "% of people who know song")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



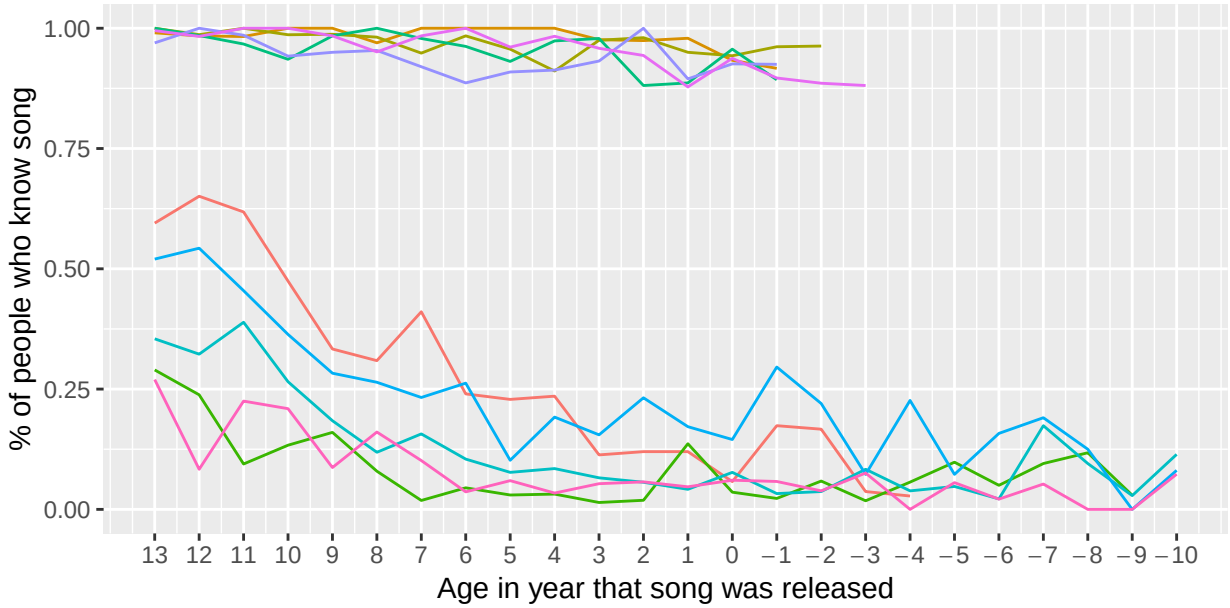
A graph of the top-5 / bottom-5 songs as they are currently remembered

```
current.rec <- rec %>%
    rowwise() %>%
    transmute(artist_song = artist_song,
              latest.recognition = last_non_na_value(c(`-13`, `-12`, `-11`, `-10`, `-9`,
                                                       `-8`, `-7`, `-6`, `-5`, `-4`,
                                                       `-3`, `-2`, `-1`, `0`, `1`,
                                                       `2`, `3`, `4`, `5`, `6`,
                                                       `7`, `8`, `9`, `10`))) %>%
    arrange(desc(latest.recognition))

top.bottom.5.songs <- union(head(current.rec, 5), tail(current.rec, 5)) %>%
    pull(artist_song)

rec %>%
    filter(artist_song %in% top.bottom.5.songs) %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition)) %>%
    ggplot() +
    geom_line(aes(x = as.numeric(generation), recognition, color = song)) +
```

```
    scale_x_continuous(breaks = seq(-13, 10, by = 1), labels = flip_sign) +
    theme(legend.position = "bottom", legend.direction = "vertical") +
    guides(color = guide_legend(ncol=2)) +
    labs(x = "Age in year that song was released",
         y = "% of people who know song",
         color = "Artist/Song")
```



A graph of the top-5 songs that were popular in the 90's but are rarely heard after that

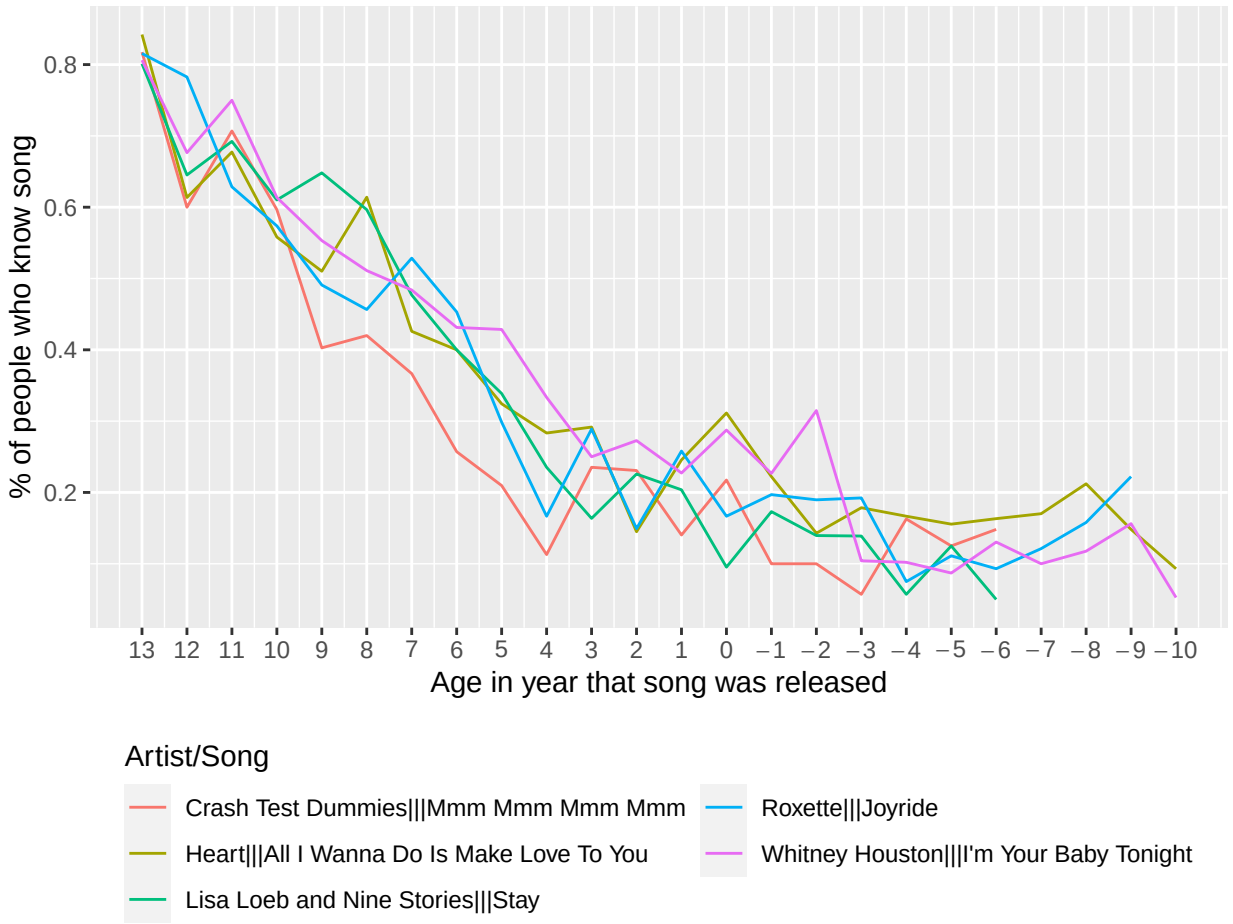
```
avg.table <- summarize_at(rec,
                          vars(`-13`:`10`),
                          ~ mean(., na.rm = TRUE))

avgs <- c()
avgs["-13"] = avg.table$`-13`
avgs["-12"] = avg.table$`-12`
avgs["-11"] = avg.table$`-11`
avgs["-10"] = avg.table$`-10`
avgs["-9"] = avg.table$`-9`
avgs["-8"] = avg.table$`-8`
avgs["-7"] = avg.table$`-7`
avgs["-6"] = avg.table$`-6`
avgs["-5"] = avg.table$`-5`
```

```
avgs["-4"] = avg.table$`-4`
avgs["-3"] = avg.table$`-3`
avgs["-2"] = avg.table$`-2`
avgs["-1"] = avg.table$`-1`
avgs["0"] = avg.table$`0`
avgs["1"] = avg.table$`1`
avgs["2"] = avg.table$`2`
avgs["3"] = avg.table$`3`
avgs["4"] = avg.table$`4`
avgs["5"] = avg.table$`5`
avgs["6"] = avg.table$`6`
avgs["7"] = avg.table$`7`
avgs["8"] = avg.table$`8`
avgs["9"] = avg.table$`9`
avgs["10"] = avg.table$`10`

low_songs <- rec %>%
    filter(`-13` >= .8) %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition)) %>%
    mutate(diff = recognition - avgs[as.character(generation)]) %>%
    group_by(song) %>%
    summarize(diff.from.trend = sum(diff)) %>%
    arrange(desc(diff.from.trend)) %>%
    tail(5) %>%
    pull(song)

rec %>%
    filter(artist_song %in% low_songs) %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition)) %>%
    ggplot() +
    geom_line(aes(x = as.numeric(generation), recognition, color = song)) +
    scale_x_continuous(breaks = seq(-13, 10, by = 1), labels = flip_sign) +
    theme(legend.position = "bottom", legend.direction = "vertical") +
    guides(color = guide_legend(ncol=2)) +
    labs(x = "Age in year that song was released",
         y = "% of people who know song",
         color = "Artist/Song")
```



A graph/table describing the change in recognition across different generations

```
knitr::kable(head(arrange(generation.rec.data,
                          desc(generation.rec.data$mean_gen_z_recognition)), 50),
             digits = 2,
             col.names = c("Artist/Song", "Millenial", "Gen Z"))
```

Artist/Song	Millenial	Gen Z
Celine Dion|||My Heart Will Go On	0.97	0.96
Britney Spears|||..Baby One More Time	0.99	0.96
Spice Girls, The|||Wannabe	0.98	0.93
Smash Mouth|||All Star	0.95	0.92
Lou Bega|||Mambo No. 5	0.98	0.92
Cher|||Believe	0.98	0.90
Snap!|||The Power	0.92	0.88
Los Del Rio|||Macarena	0.97	0.87
Backstreet Boys, The|||Everybody	0.97	0.86
Whitney Houston|||I Will Always Love You	0.91	0.85
R. Kelly|||I Believe I Can Fly	0.96	0.82
Coolio|||Gangsta's Paradise	0.92	0.82
Michael Jackson|||Black Or White	0.89	0.81
House Of Pain|||Jump Around	0.89	0.80
Ricky Martin|||Livin' La Vida Loca	0.92	0.79

Artist/Song	Millenial	Gen Z
Sir Mix-A-Lot|||Baby Got Back	0.84	0.77
R.E.M.|||Losing My Religion	0.89	0.74
Elton John|||Can You Feel The Love Tonight	0.91	0.72
Peabo Bryson and Regina Bell|||A Whole New World	0.86	0.70
Christina Aguilera|||Genie In A Bottle	0.92	0.69
C and C Music Factory|||Gonna Make You Sweat	0.87	0.69
TLC|||No Scrubs	0.89	0.67
Aerosmith|||I Don't Want To Miss A Thing	0.91	0.67
Joan Osborne|||One Of Us	0.91	0.66
Sixpence None The Richer|||Kiss Me	0.89	0.66
Puff Daddy and Faith Evans|||I'll Be Missing You	0.89	0.65
Proclaimers, The|||I'm Gonna Be	0.87	0.65
Michael Jackson|||You Are Not Alone	0.77	0.63
Montell Jordan|||This Is How We Do It	0.85	0.62
Marky Mark and The Funky Bunch|||Good Vibrations	0.79	0.62
Enrique Iglesias|||Bailamos	0.91	0.60
Ace Of Base|||The Sign	0.88	0.60
Alanis Morissette|||Ironic	0.89	0.59
Toni Braxton|||Un-Break My Heart	0.90	0.58
Santana|||Smooth	0.93	0.57
Kris Kross|||Jump	0.72	0.55
Red Hot Chili Peppers|||Under The Bridge	0.84	0.54
Ace Of Base|||All That She Wants	0.83	0.54
Billy Ray Cyrus|||Achy Breaky Heart	0.80	0.53
Celine Dion|||All By Myself	0.75	0.53
BLACKstreet|||No Diggity	0.79	0.53
Seal|||Kiss From A Rose	0.88	0.51
Tag Team|||Whoomp! There It Is	0.82	0.51
TLC|||Waterfalls	0.85	0.50
Nicki French|||Total Eclipse Of The Heart	0.68	0.47
Madonna|||Vogue	0.69	0.47
Vanilla Ice|||Ice Ice Baby	0.68	0.46
Third Eye Blind|||Semi-Charmed Life	0.73	0.44
Real McCoy|||Another Night	0.75	0.44
Notorious B.I.G., The|||Hypnotize	0.63	0.44

Original question

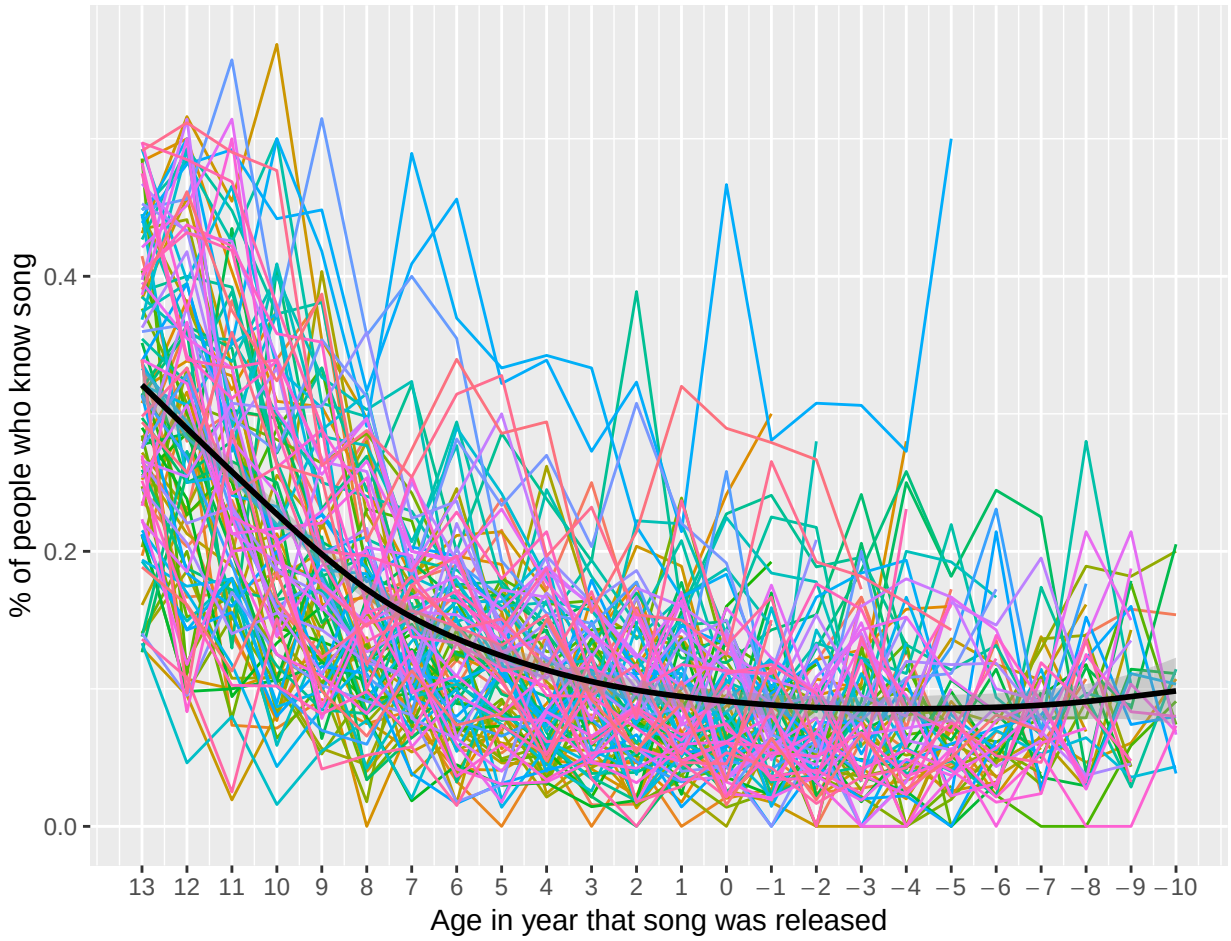
We want to know how recognized songs are today that were only recognized by less than 50% of people that were 13-15 years old at the time of release. Do some of these songs gain popularity or do most die out?

```
low_rec_at_release <- rec %>%
    filter(`-13` <= .5) %>%
    pivot_longer(`-13`:`10`,
                 names_to = "generation", values_to = "recognition") %>%
    transmute(song = artist_song, generation = generation, recognition = recognition) %>%
    filter(!is.na(recognition))

ggplot(low_rec_at_release) +
geom_line(aes(x = as.numeric(generation), recognition, color = song)) +
geom_smooth(aes(as.numeric(generation), recognition), color = "black") +
scale_x_continuous(breaks = seq(-13, 10, by = 1), labels = flip_sign) +
```

```
theme(legend.position = "none") +
labs(x = "Age in year that song was released",
     y = "% of people who know song")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
low_rec_at_release %>%
    mutate(diff = recognition - avgs[as.character(generation)]) %>%
    group_by(song) %>%
    summarize(diff.from.trend = sum(diff)) %>%
    arrange(desc(diff.from.trend)) %>%
    head(5) %>%
    pull(song)
```

```
## [1] "Notorious B.I.G., The|||The What"
## [2] "Notorious B.I.G., The|||One More Chance-Stay With Me"
## [3] "Whitney Houston|||I Believe In You And Me"
## [4] "Whitney Houston|||Exhale"
## [5] "Brandy|||Have You Ever?"
```

Most of these songs appear to die out and not be recognized today. We can tell this by the decreasing slope of the trend line for these songs that we recognized by less than 50% of 13-15 year olds when the song was released. This isn't too surprising of a result. It is interesting to see which of these songs are most recognized today though. The top 2 songs are by Notorious B.I.G. and the next 2 most recognized are by Whitney Houston. It makes sense that these artists are at the top since they are still popular today and have many

fans that could recognize some of their least popular songs.