Project Report - 90's Music Canon

Max Narvaez & Ryan Harvey

9/30/2020

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

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. The `latest.regocnition` variable is found by using the `last_non_na_value()` function to find the trailing non `NA` value for each song. The `diff.from.trend` variable is found by taking the sum of the difference between a song's recognition data point and the average recognition rate for each age.

Helper functions used later

```
# Flip -13 to 13, 10 to -10, 0 to 0, etc.
#
# Used to format the labels for the ages
flip_sign <- function(l) {
  l <- as.numeric(l)*(-1)
  parse(text=l)
}

# Get the last value in a row that isn't NA
#
# Used to find the latest data point for each song
last_non_na_value <- function(r) {
  cols <- length(r)
  col <- cols - 1
  while (is.na(r[col])) {
    col <- col - 1
  }
  r[col]
}
```

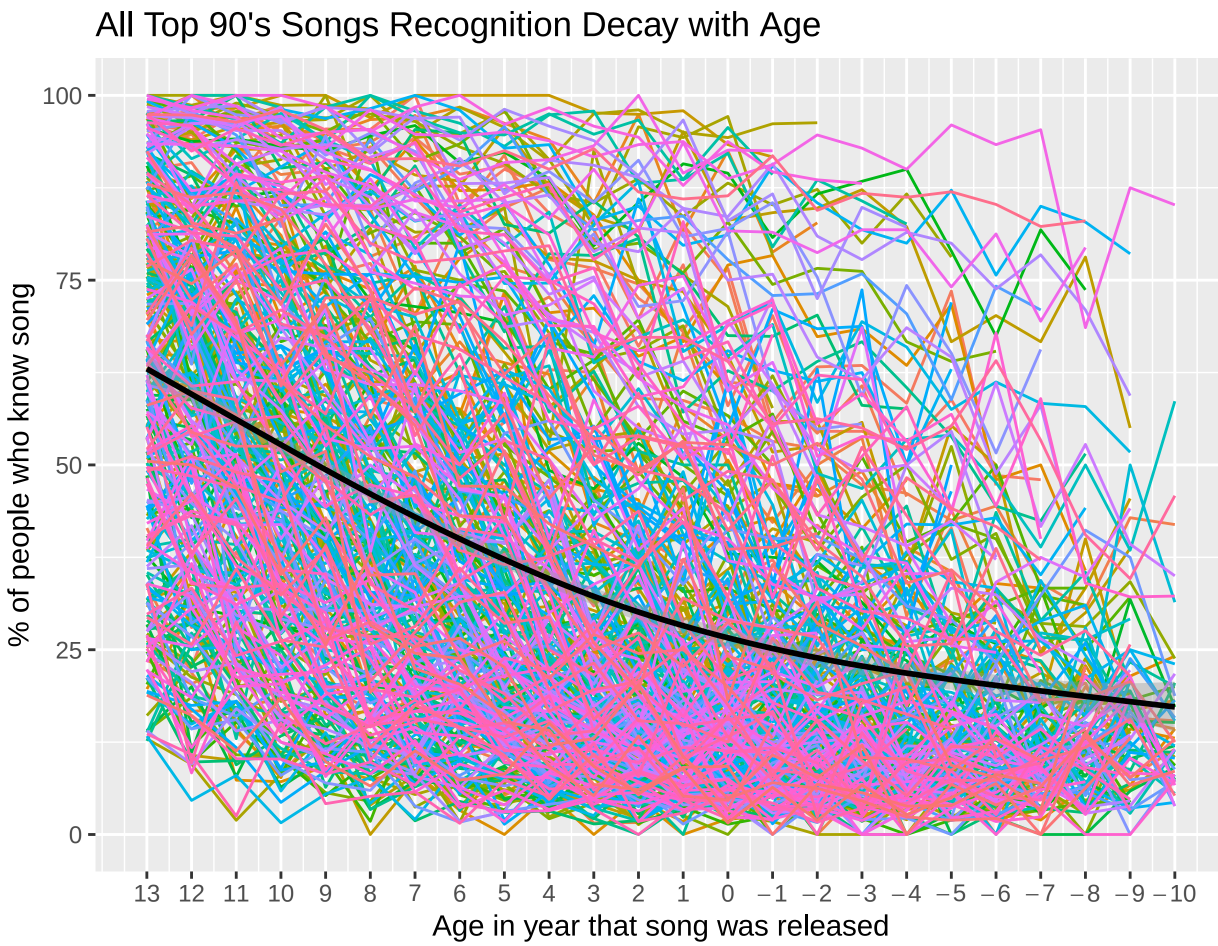
Prepare Dataset

```r
# For each row, start at the last column and
#  work backwards towards the first column,
#  replacing any 0 with NA until a non-0 value is found,
#  then move to the next row
#
# This cleans up the data so non-existent data points
#  can be identified and filtered out easily while
#  preventing us from losing data that should be 0.
# Technically we could be losing some data if songs ended
#  with a 0% recognition rate, but because of how the data
#  is presented, this would be difficult to distinguish
#  from a song that has 0s filling the extra data points.
# Thus it is likely not a large issue if it is at all.
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 from Article

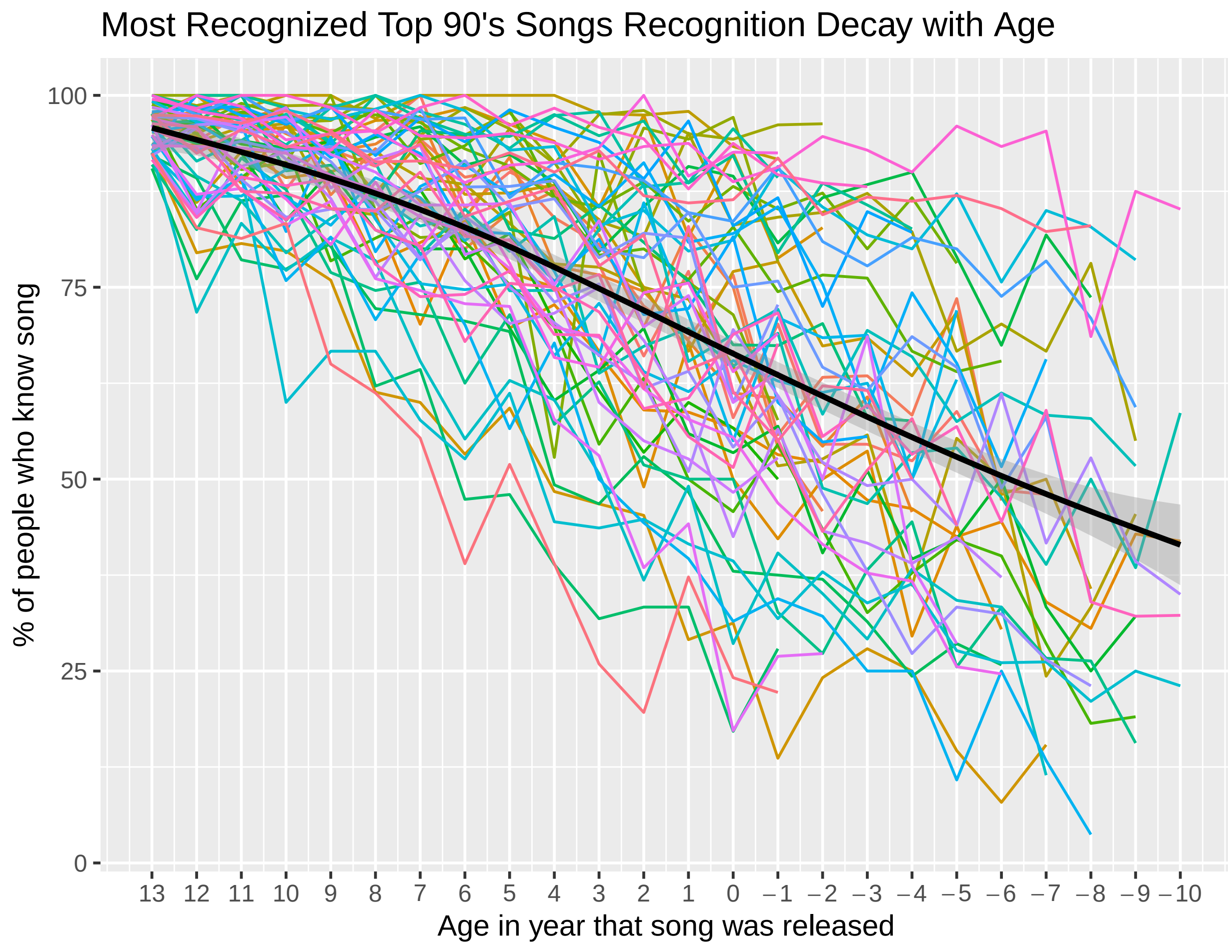
(1) The percent of songs from the 90's recognized currently decays with time

```r
rec %>%
  pivot_longer(`-13`:`10`,
               names_to = "generation",
               values_to = "recognition") %>%
  transmute(song = artist_song,
            generation = generation,
            recognition = recognition * 100) %>%
  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",
       title = "All Top 90's Songs Recognition Decay with Age")
```



This plot includes every top 90's song recognition based on age. There is an overall downward trend, which makes sense since certain songs will be forgotten over time. We will use this trend line as our average recognition rate for some of the following plots in this report. This plot doesn't tell us how some songs hold up over time. The next plot tries to do that.

```
rec %>%
  filter(`-13` >= .9) %>%
  pivot_longer(`-13`:`10`,
               names_to = "generation",
               values_to = "recognition") %>%
  transmute(song = artist_song,
            generation = generation,
            recognition = recognition * 100) %>%
  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",
       title = "Most Recognized Top 90's Songs Recognition Decay with Age")
```



This plot is all the top 90's songs that were recognized by at least 90% of those who were 13-15 years old at the time of the song's release. Once again, there is a downward trend, but the slope is less steep than the previous plot. This reveals that the songs that almost all Millenials recognize are much more likely to be recognized by Gen Z.

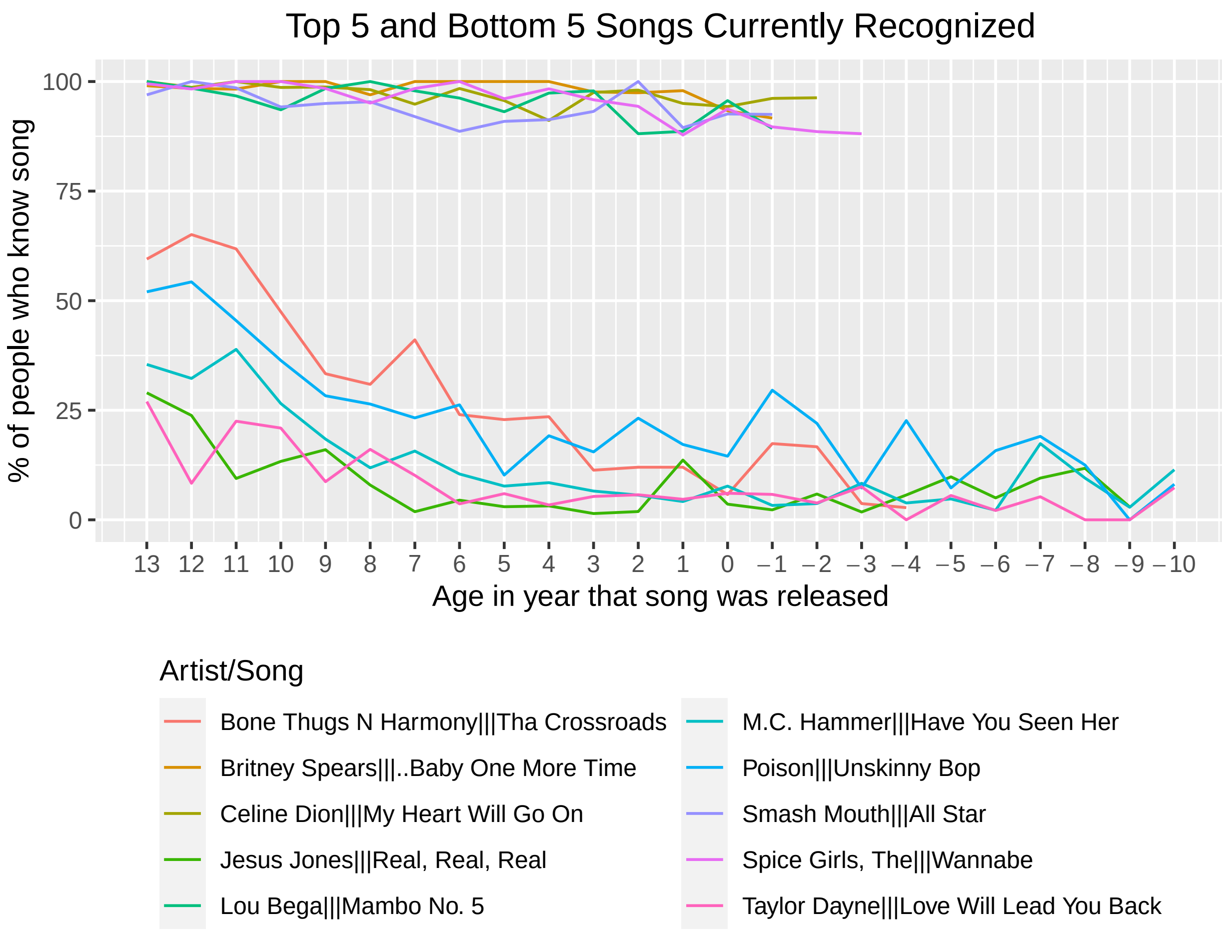
(2) A graph of the top-5 / bottom-5 songs as they are currently remembered

```
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 * 100) %>%
  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",
        plot.title = element_text(hjust = 0.5)) +
  guides(color = guide_legend(ncol=2)) +
  labs(x = "Age in year that song was released",
       y = "% of people who know song",
       title = "Top 5 and Bottom 5 Songs Currently Recognized",
       color = "Artist/Song")
```



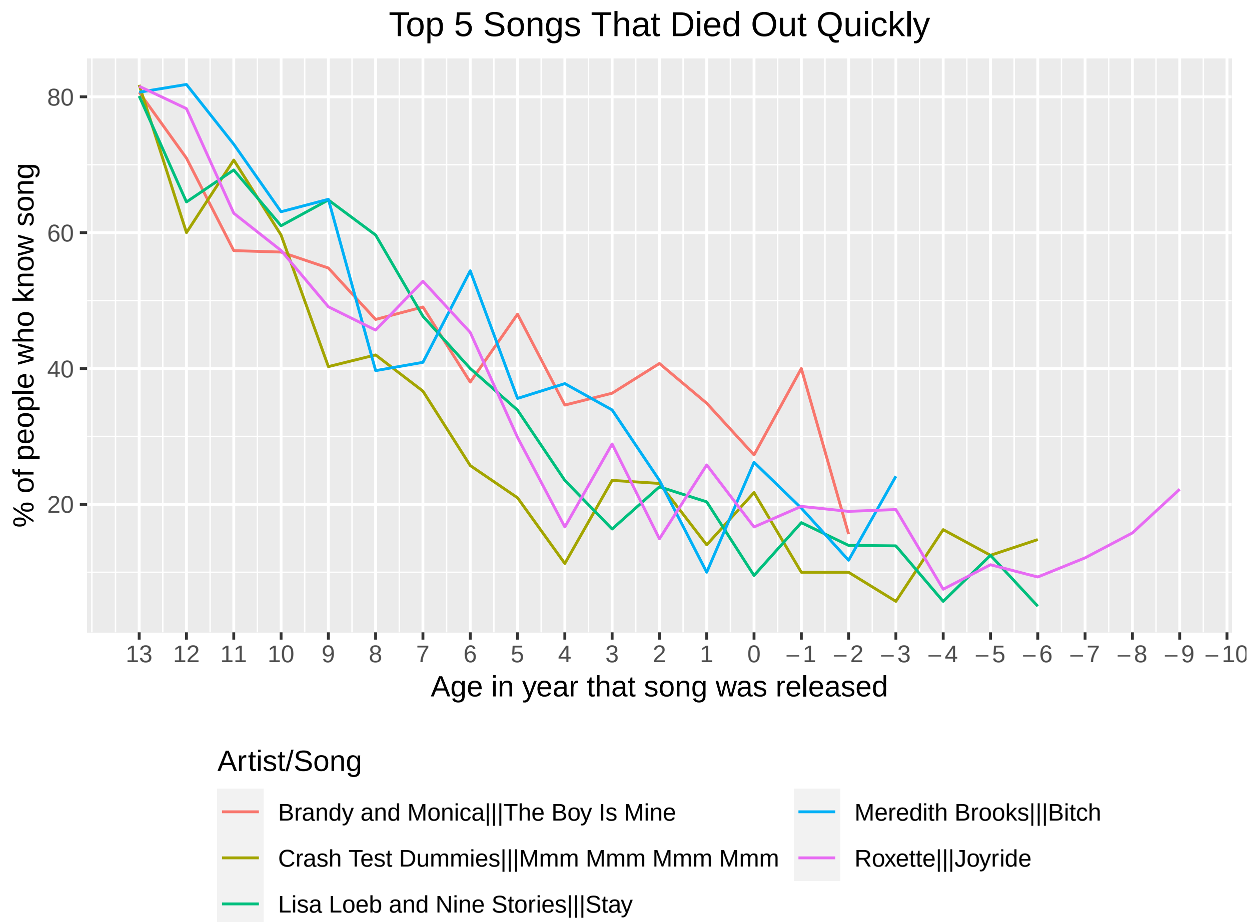
This graph is showing the current top 5 and bottom 5 songs from the 90s. It isn't surprising that the lowest songs started low and the highest songs started high. One interesting observation is that the high songs are more recent (indicated by the lack of data from people that were born more than 2-3 years after the release).

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

```
# Get the average for each year and store to a named vector
avgs <- unlist(summarize_at(rec,
                            vars(`-13`:`10`),
                            ~ mean(., na.rm = TRUE))[1,])

# Determine the top 5 originally but not currently popular songs
#  based on their "distance" from the trendline
#
# "Distance" means the sum of the distance of each data point from
#  the average of all data points for that year.
# A positive distance means it did better than average overall,
#  while a negative distance means it did worse than average overall
low_songs <- rec %>%
  filter(`-13` >= .8) %>%
  pivot_longer(`-13`:`10`,
               names_to = "generation",
               values_to = "recognition") %>%
  transmute(song = artist_song,
            generation = generation,
            recognition = recognition * 100) %>%
  filter(!is.na(recognition)) %>%
  # Find the distance from the trendline for each data point
  mutate(diff = recognition - avgs[as.character(generation)]) %>%
  group_by(song) %>%
  # Sum the distances to get the overall "distance" from the trendline
  summarize(diff.from.trend = sum(diff)) %>%
  arrange(desc(diff.from.trend)) %>%
  tail(5) %>%
  pull(song)

# Create the chart for the songs found above
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 * 100) %>%
  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",
        plot.title = element_text(hjust = 0.5)) +
  guides(color = guide_legend(ncol=2)) +
  labs(x = "Age in year that song was released",
       y = "% of people who know song",
       title = "Top 5 Songs That Died Out Quickly",
       color = "Artist/Song")
```



There is a clear plummet in the songs' popularity over the years. This makes sense because after starting so well they have to do really poorly to make up for their prior good performance. One thing to note is that if the threshold for a "popular" song is shifted, i.e. from 80% to 85%, the makeup of the songs selected will change. This is because a song that starts off worse has a better shot at being far away from the trendline overall.

(4) 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

Artist/Song	Millenial	Gen Z
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
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

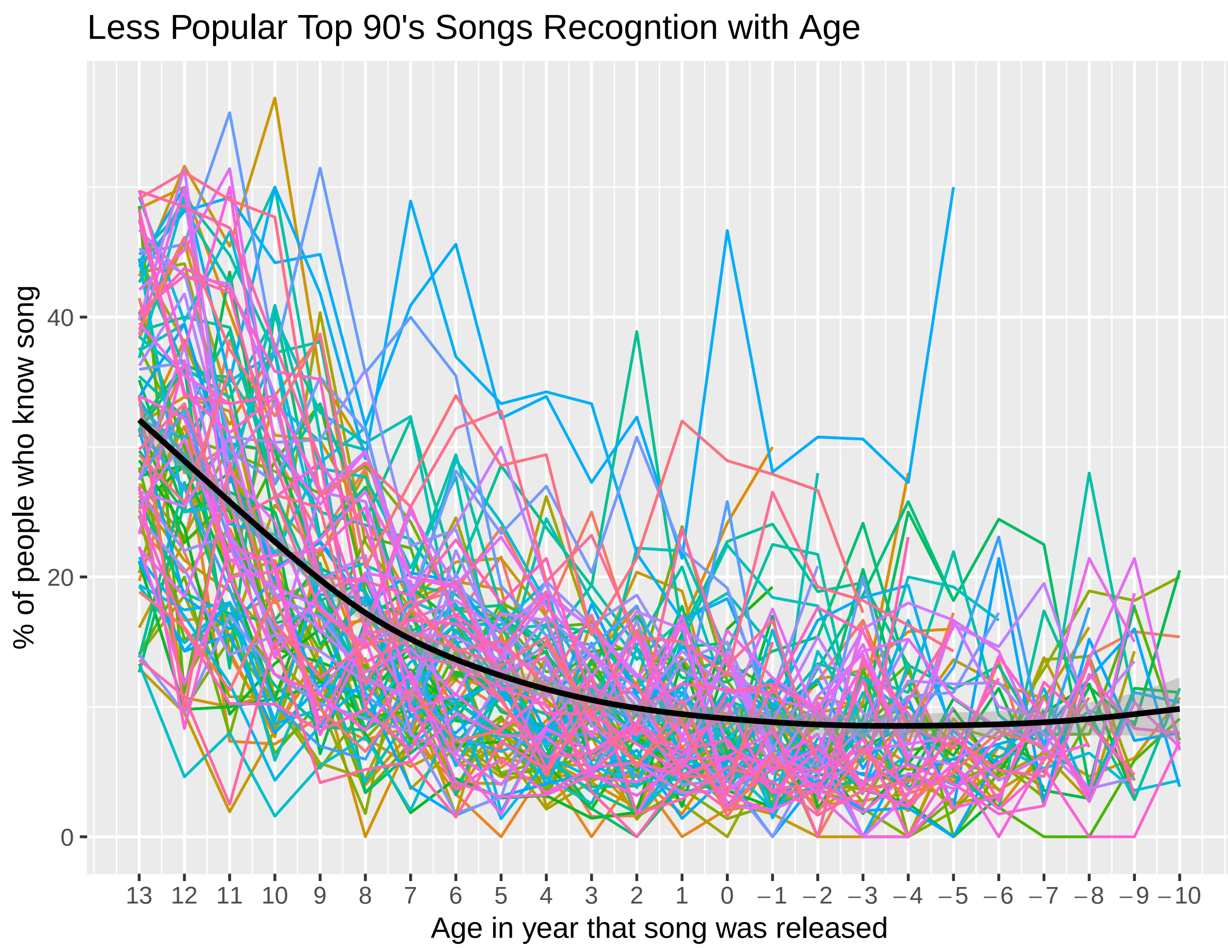
This table is a top 50 ranking of top 90's songs based on Gen Z recognition rank. It is interesting to note which songs have a large gap between Millenial and Gen Z recognition rate. These songs would have been very popular back when they came out, but have lost their popularity. The few songs that are both recognized by almost all Millenials and Gen Z are the songs that will live on. The higher the recognition rate of songs by Gen Z, the more likely they will continue to define the 90's music canon.

Our 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?

```r
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 * 100) %>%
  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",
       title = "Less Popular Top 90's Songs Recogntion with Age")
```




```r
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)) %>%
```

```r
  head(5) %>%
  pull(song) %>%
  knitr::kable(col.names = c("Artist/Song"))
```

Artist/Song
Notorious B.I.G., The|||The What
Notorious B.I.G., The|||One More Chance-Stay With Me
Luther Vandross|||Power Of Love
Whitney Houston|||Exhale
Mariah Carey|||I Don't Wanna Cry

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. Most of these songs are only recognized by 10% of Gen Z. 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 The 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.