```r
1   ## Clean Genre Assignments ####
2   suppressMessages(library(tidyverse))
3
4   ## Load data ####
5   DATA_PATH <- '../../data/skate_data_uncleaned.csv'
6   df <- suppressMessages(read_csv(DATA_PATH))
7   df$index <- NULL
8   count_df <- df
9   count_df$genre <- NULL
10
11  ## Reassign genres to broader categories ####
12  count_df$genre[count_df$genre2 %in% c("50's Rock", "60's Rock", "70's Rock", "European
    Pop", "Folk Rock", "New Romantic",
13                                        "Retro Rock Revival", "Western Pop"
14  )] = 'Classic Rock'
15  Oldies <- c("Classic Country", "Classic Pop Vocals")
16  count_df$genre[count_df$genre2 %in% c("Dance & Club", "Dubstep", "Downtempo, Lounge &
    Ambient", "Electronica Fusion",
17                                        "Electronica Mainstream", "Game Music", "House",
    "Indian Subcontinent Pop", "Techno",
18                                        "Trance")] = 'Electronic'
19
20  count_df$genre[count_df$genre2 %in% c("Asian Hip-Hop/Rap", "Animation", "Caribbean
    Pop", "Latin Hip-Hop/Rap", "Reggae",
21                                        "Western Hip-Hop/Rap")] = 'Hip Hop'
22  count_df$genre[count_df$genre2 %in% c("Adult Alternative Rock", "Alternative",
    "Alternative Folk", "Alternative Roots",
23                                        "Brit Pop", "Brit Rock", "Contemporary Era",
    "Folk", "Indie Rock", "New Wave Pop",
24                                        "New Wave Rock", "Synth Pop"
25  )] = 'Indie/Alternative'
26  count_df$genre[count_df$genre2 %in% c("Acoustic Blues", "Bebop & Modern Jazz", "African
    Pop", "African Traditional",
27                                        "Asian Traditional", "Big Band & Swing", "Classic
    R&B/Soul", "Contemporary Jazz & Fusion",
28                                        "Contemporary R&B/Soul", "Funkadelic", "Mateo
    Messina", "Disco", "Early Jazz",
29                                        "Electric Blues", "Jazz Vocals", "Jungle/Drum 'n'
    Bass", "Latin & World Jazz",
30                                        "Smooth Jazz", "South American"
31  )] = "Jazz/Soul"
32  count_df$genre[count_df$genre2 %in% c("Goth", "Hard Rock", "Industrial", "Metal")] =
    'Metal'
33
34  count_df$genre[count_df$genre2 %in% c("Emo & Hardcore", "Punk", "Rockabilly Revival",
    "Ska Revival", "Surf Revival"
35  )] = 'Punk'
36  count_df$genre[count_df$genre2 %in% c("Garage Rock Revival", "Mainstream Rock")] =
    'Rock'
37
38  count_df$genre[count_df$genre2 %in% c("Baroque Era",  "Brazilian", "Celtic",
    "Children's", "Chinese Pop", "Classic Pop Vocals",
39                                        "Classical Crossover", "Comedy", "Country","Data
    & Other",  "Easy Listening", "European Rock",
40                                        "European Traditional", "Holiday", "Indian
    Traditional", "Jam Bands", "Japanese Pop",
```

```
41                                              "Japanese Rock", "Latin Pop", "Latin Rock",
     "Latin Traditional", "Mexican",
42                                              "Mid-East Traditional", "Middle East Pop",
     "Modern Era", "New Age", "Original Film/TV Music",
43                                              "Other Classical", "Other Classical", "Power
     Pop", "Reggaetón", "Religious", "Romantic",
44                                              "Sound Effects & Ringtones", "Spoken Word",
     "Sports Themes", "Stage Musicals"
45  )] = 'Other'
46
47  count_df <- count_df %>%
48    filter( !(genre2 %in% c("Asian Pop", "Classical Era", "Swing Revival")))
49
50
51  count_df$artist[count_df$artist == 'Eternal Summer'] = 'Eternal Summers'
52  count_df$artist[count_df$artist == 'Arthur Russel'] = 'Arthur Russell'
53  count_df$artist[count_df$artist == 'Shit Kickers'] = 'Shitkickers'
54  count_df$artist[count_df$artist == 'Patty Smith'] = 'Patti Smith'
55  count_df$artist[count_df$artist == 'Reverend Horton Heat'] = 'The Reverend Horton Heat'
56  count_df$artist[count_df$artist == 'Native Baron'] = 'Baron'
57  count_df$artist[count_df$artist == 'Matti Costa'] = 'Matt Costa'
58  count_df$artist[count_df$artist == 'Streets'] = 'The Streets'
59  count_df$artist[count_df$artist == 'Mamas and the Papas'] = 'The Mamas & The Papas'
60  count_df$artist[count_df$artist == ''] = ''
61  count_df$artist[count_df$artist == ''] = ''
62  count_df$artist[count_df$artist == ''] = ''
63  count_df$artist[count_df$artist == ''] = ''
64
65  ## Manually assign poorly assigned Artists ####
66  count_df$genre[count_df$artist %in% c('Enoch', 'Mr. Dibbs', 'Odd Nosdam', 'Baron',
     'Adeodat Warfield',
67                                              'DJ Egadz', 'DJ Mutt', 'Air', '((sounder))',
     'Jel', 'DJ Frane',
68                                              'Le Tigre', 'deadxbeat', 'DJ Greyboy', 'DJ
     Kechup', 'DJ Sureshot',
69                                              'Eighty Mile Beach', 'Mophono', 'Shed', 'Busy P',
70                                              'Nobody', 'Aim', 'Dutch Dub', 'Andy Caldwell',
     'Broken Spindles',
71                                              'Vangelis', 'Vito Y Coco', 'Sixtoo', 'Baron',
     'Sol', "Salem",
72                                              "Blvck Ceiling", "Baron", "Emperor Penguin",
     "Digital Cutup Lounge",
73                                              "Ellie Goulding", "Death In Vegas", "Balam Acab",
     "Justin Bates",
74                                              "Martin Solveig & Dragonette", "Martin Solveig",
75  )] = 'Electronic'
76
77  count_df$genre[count_df$artist %in% c('Bad Shit', 'Hightower', 'Pegboy', 'Sweet
     Apple','Graveyard',
78                                              'Black Mountain', 'Nihilist', 'The LSDemons',
     'ASG',
79                                              'Fretblanket', 'Grindline The Band', 'Hum',
     'Tool',
80                                              'Dead Meadow', 'The Sadies', 'Valient Thorr',
     'Viking Skull',
81                                              'Motörhead', 'J Mascis and the Fog', 'Lard',
     'Roky Erickson',
```

```r
 82                                               "Black Rose", "The Invisibles","Hightower",
    "Cky", "CKY",
 83                                               "Black", "Mercyful Fate", "Guns N' Roses",
    "Motörhead",
 84                                               "Slayer", "Earthless", "Green Leaf", "Flower
    Travlin' Band",
 85                                               "Bleed The Sky", "Boris", "Sweet Apple", "Red
    Planet"
 86
 87  )] = 'Metal'
 88
 89  count_df$genre[count_df$artist %in% c('Euphone', 'Caural', 'Duane Pitre', 'Rodrigo &
    Gabriela',
 90                                               'Jet Black Crayon', 'Bob James', 'Mother Earth',
    'Bob Marley',
 91                                               'Sizzla', "Tenor Saw", "Pete Drake", "Carlinhos
    Zodi" )] = 'Jazz/Soul'
 92
 93  count_df$genre[count_df$artist %in% c('Atiba', 'Guru', 'Money Mark', 'Bracket', 'Chad
    Muska',
 94                                               'Hurakan', 'Kanye West', 'Sacred Hoop', 'Swollen
    Members',
 95                                               'The Grouch', 'Extra Prolific', 'Hip Hop For
    Respect',
 96                                               'Slippers', 'Budgie', 'Gorillaz', 'Styles of
    Beyond',
 97                                               "Jigz Crillz", "Rok One", "P.H.A.T", "456
    Productions",
 98                                               "Islah Guzman", "Alf Diggi", "Chris Gentry",
    "Eazy E",
 99                                               "Spark Master Tape", "12 Jewelz", "Sacred Hoop",
    "Jeep Beat Collective",
100                                               "Notorious B.I.G.", "Galliano", "Neptunes",
    "Deadxbeat", "Chad Muska",
101                                               "The Streets"
102  )] = 'Hip Hop'
103
104  count_df$genre[count_df$artist %in% c('Curtis Mayfield', 'Nina Simone', 'Marvin Gaye',
105                                               'Stevie Wonder', 'Al Green', 'Ben Harper',
    'Michael Jackson',
106                                               'Bobby Womack', 'Prince', 'Rick James', 'The
    Commodores',
107                                               'The Temptations', 'Frank Sinatra', 'Ananda
    Shankar', 'Saxman',
108                                               "Ray Barbee", "Homenagon A Mongo", "Lionel Richie
    ",
109                                               "Philip Glass", "Phillip Glass", "Luis Bacalov",
    "Big Band Katowice",
110                                               "Hermon Hitson", "Harry Warren, Jack Brooks",
    "Pat Williams Orchestra",
111                                               "Cee Knowledge", "Slim And The Soulful Saints",
    "Tommy Guerrero",
112                                               "Tommy Guerrero & Monte Vallier", "Cinematic
    Orchestra",
113                                               "The James Taylor Quartet", "Soul Train Soul",
114                                               "Tommy Guerrero & Mark Gonzales", "Tommy Guerrero
    & DJ Pause",
```

```r
115                                     "Brooklyn Funk Essentials", "Solsonics",
       "Jestofunk",
116                                     "Ralph Myerz and the Jack Herren Band", "Wizards
       Of Ooze",
117                                     "Palace of Pleasure", "The Brand New Heavies",
       "Jim Sound Entley",
118                                     "Pink Martini", "Tommy Guerrero & Gadget
       Enhanced", "Paco De Lucia",
119                                     "Peru Negro", "Los Incas", "Ray Barbee"
120    )] = 'Jazz/Soul'
121
122    count_df$genre[count_df$artist %in% c('Belle & Sebastian', 'Eternal Summers', 'Elliot
       Smith',
123                                     'Matt Costa', 'Weird War', "Ariel Pink's Haunted
       Graffiti",
124                                     'Kelley Stoltz', 'The Greenhornes', 'The Organ',
       'The Sounds',
125                                     'Whitey', 'A Band of Bees', 'Birds of Avalon',
       'Doves',
126                                     'Luscious Jackson', 'Madonna', 'Old Canes',
       'Pepper Rabbit',
127                                     'The Go! Team', 'Depeche Mode', 'The Black Lips',
       'The Cry',
128                                     'The Teenagers', 'The Zoo', 'Wonderful Broken
       Thing',
129                                     'Liam The Younger', 'Seafood', 'Spindrift',
       'Amadou & Mariam',
130                                     "Who Made Who", "Afterimage", "The Len Brown
       Society",
131                                     "Gifthorse", "The People's Temple", "Barry Louis
       Polisar",
132                                     "The Downtown Struts", "Die Jungen", "The
       Crocodiles",
133                                     "The Devil Makes Three", "Henries Holborn", "Rock
       Plaza Central",
134                                     "Leeroy Stagger", "White Flight", "Main Fear
       Love", "The Wiseguys",
135                                     "Kid loco", "Kid Loco", "Lana del Rey", "Danger
       Beach", "Chemical Brothers",
136                                     "The Chemical Brothers", "The 2 Bears", "The
       Mynabirds", "Dracula Lewis",
137                                     "Puro Instinct", "Teddybears", "The Similou",
       "Audionom", "Inina Gap",
138                                     "Teddybears Stockholm", "Pedro The Lion", "Matt
       Costa", "Manu Chao",
139                                     "Arthur Russell", "The Minders", "The Nerves",
       "New Order",
140                                     "Depeche Mode", "Yael Naim", "Gwen Stefani",
       "Katy Perry",
141                                     "Edward Sharpe And The Magnetic Zeroes"
142    )] = 'Indie/Alternative'
143
144    count_df$genre[count_df$artist %in% c('The Adolescents', 'The Odd Numbers', 'Bad
       Strip',
145                                     'Division of Laura Lee', 'Sub Society', 'Dance
       Hall Crashers',
```

```
146    'Drunk Injuns', 'Marginal Man', 'Rudimentary
    Peni', 'The Heartaches',
147    'The Saints', 'The Buzzcocks', 'The Needles',
    'The Veebees',
148    'Los Olvidados', 'Smut Peddlers', "The Spits",
    "Shed", "Blur",
149    "Elastica", "Sands", "Plaid Rentina", "Plaid
    Retina", "Embrace",
150    "The Milkshakes", "Pissboner", "Loren Humphrey",
    "Rat Fist",
151    "Shitkickers", "Gangbusters", "Delaware White
    Thrash",
152    "Tilt Wheel", "Devo", "Willie Grundle And The
    Taints", "Yuck",
153    "Vermillion Sands", "The Triggers", "The Plague",
    "Monarch Box",
154    "Golden Shower", "Livingdead", "Johnny & The
    Dudes", "Badshit",
155    "The Scematics", "The Authorities", "GG Allin and
    The Murder Junkies",
156    "The Richmond Sluts", "The Clash", "Devo", "Patti
    Smith",
157    "Elvis Costello", "The Jam", "The Modern Lovers",
    "The Only Ones",
158    "The Cry", "The Cramps", "Reverend Horton Heat",
    "The Legendary Shack Shakers",
159    "The Peacocks"
160 )] = 'Punk'
161
162 count_df$genre[count_df$artist %in% c('Johnny Cash', 'Love', 'Red Octopus', 'Dick
    Dale', 'The Small Faces',
163    'Norman Greenbaum', "Focus", "The Monkeys", "The
    Monkees",
164    "Max Bygraves", "Bobby Pickett and The Crypt",
    "Spencer Davis Group",
165    "The Beatles", "Van Morrison", "The Grass Roots",
    "Dire Straits",
166    "Shocking Blue", "Syd Barrett", "Syd Barret",
    "The Band",
167    "David Bowie & Bing Crosby", "Grateful Dead",
    "Dylan & The Dead",
168    "Little Feet", "Loose Gravel", "Tom Petty & The
    Heartbreakers",
169    "John Mellencamp", "The Rolling Stones", "Tom
    Petty", "The Velvet Underground",
170    "Billy Idol", "Lou Reed", "The Police", "Cheap
    Trick", "The Vapors",
171    "The Romantics", "The Solution", "The Beatles",
    "Mamas and the Papas",
172    "Paul McCartney", "John Lennon", "Cat Stevens",
    "Paul Simon",
173    "Neil Diamond", "Hall & Oates"
174 )] = 'Classic Rock'
175
176 ## Save data to csv ####
177 #write_csv(count_df, '~/Desktop/cleaned_skatedata.csv3')
178
```