

```

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
13   Pop", "Folk Rock", "New Romantic",
14   "Retro Rock Revival", "Western Pop"
15   )] = 'Classic Rock'
16 Oldies <- c("Classic Country", "Classic Pop Vocals")
17 count_df$genre[count_df$genre2 %in% c("Dance & Club", "Dubstep", "Downtempo, Lounge &
18   Ambient", "Electronica Fusion",
19   "Electronica Mainstream", "Game Music", "House",
20   "Indian Subcontinent Pop", "Techno",
21   "Trance")] = 'Electronic'
22 count_df$genre[count_df$genre2 %in% c("Asian Hip-Hop/Rap", "Animation", "Caribbean
23   Pop", "Latin Hip-Hop/Rap", "Reggae",
24   "Western Hip-Hop/Rap")] = 'Hip Hop'
25 count_df$genre[count_df$genre2 %in% c("Adult Alternative Rock", "Alternative",
26   "Alternative Folk", "Alternative Roots",
27   "Brit Pop", "Brit Rock", "Contemporary Era",
28   "Folk", "Indie Rock", "New Wave Pop",
29   "New Wave Rock", "Synth Pop"
30   )] = 'Indie/Alternative'
31 count_df$genre[count_df$genre2 %in% c("Acoustic Blues", "Bebop & Modern Jazz", "African
32   Pop", "African Traditional",
33   "Asian Traditional", "Big Band & Swing", "Classic
34   R&B/Soul", "Contemporary Jazz & Fusion",
35   "Contemporary R&B/Soul", "Funkadelic", "Mateo
36   Messina", "Disco", "Early Jazz",
37   "Electric Blues", "Jazz Vocals", "Jungle/Drum 'n'
38   Bass", "Latin & World Jazz",
39   "Smooth Jazz", "South American"
40   )] = "Jazz/Soul"
41 count_df$genre[count_df$genre2 %in% c("Goth", "Hard Rock", "Industrial", "Metal")] =
42   'Metal'
43
44 count_df$genre[count_df$genre2 %in% c("Emo & Hardcore", "Punk", "Rockabilly Revival",
45   "Ska Revival", "Surf Revival"
46   )] = 'Punk'
47 count_df$genre[count_df$genre2 %in% c("Garage Rock Revival", "Mainstream Rock")] =
48   'Rock'
49
50 count_df$genre[count_df$genre2 %in% c("Baroque Era", "Brazilian", "Celtic",
51   "Children's", "Chinese Pop", "Classic Pop Vocals",
52   "Classical Crossover", "Comedy", "Country", "Data
53   & Other", "Easy Listening", "European Rock",
54   "European Traditional", "Holiday", "Indian
55   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', '((sunder))',
'Jel', 'DJ Frane',
68         'Le Tigre', 'deadxbeat', 'DJ Greyboy', 'DJ
Ketchup', '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',

```

```

82     "Black Rose", "The Invisibles", "Hightower",
83     "Cky", "CKY",
84     "Black", "Mercyful Fate", "Guns N' Roses",
85     "Motörhead",
86     "Slayer", "Earthless", "Green Leaf", "Flower
87     Travlin' Band",
88     "Bleed The Sky", "Boris", "Sweet Apple", "Red
89     Planet"
90 ]] = 'Metal'
91
92 count_df$genre[count_df$artist %in% c('Euphone', 'Caural', 'Duane Pitre', 'Rodrigo &
93 Gabriela',
94     'Jet Black Crayon', 'Bob James', 'Mother Earth',
95     'Bob Marley',
96     'Sizzla', "Tenor Saw", "Pete Drake", "Carlinhos
97 Zodi" )] = 'Jazz/Soul'
98
99 count_df$genre[count_df$artist %in% c('Atiba', 'Guru', 'Money Mark', 'Bracket', 'Chad
100 Muska',
101     'Hurakan', 'Kanye West', 'Sacred Hoop', 'Swollen
102 Members',
103     'The Grouch', 'Extra Prolific', 'Hip Hop For
104 Respect',
105     'Slippers', 'Budgie', 'Gorillaz', 'Styles of
106 Beyond',
107     'Jigz Crillz', "Rok One", "P.H.A.T", "456
108 Productions",
109     "Islah Guzman", "Alf Diggi", "Chris Gentry",
110     "Eazy E",
111     "Spark Master Tape", "12 Jewelz", "Sacred Hoop",
112     "Jeep Beat Collective",
113     "Notorious B.I.G.", "Galliano", "Neptunes",
114     "Deadxbeat", "Chad Muska",
115     "The Streets"
116 )] = 'Hip Hop'
117
118 count_df$genre[count_df$artist %in% c('Curtis Mayfield', 'Nina Simone', 'Marvin Gaye',
119     'Stevie Wonder', 'Al Green', 'Ben Harper',
120     'Michael Jackson',
121     'Bobby Womack', 'Prince', 'Rick James', 'The
122 Commodores',
123     'The Temptations', 'Frank Sinatra', 'Ananda
124 Shankar', 'Saxman',
125     "Ray Barbee", "Homenagon A Mongo", "Lionel Richie
126 ",
127     "Philip Glass", "Phillip Glass", "Luis Bacalov",
128     "Big Band Katowice",
129     "Hermon Hitson", "Harry Warren, Jack Brooks",
130     "Pat Williams Orchestra",
131     "Cee Knowledge", "Slim And The Soulful Saints",
132     "Tommy Guerrero",
133     "Tommy Guerrero & Monte Vallier", "Cinematic
134 Orchestra",
135     "The James Taylor Quartet", "Soul Train Soul",
136     "Tommy Guerrero & Mark Gonzales", "Tommy Guerrero
137 & DJ Pause",

```

```

115 "Jestofunk",
116 "Of Ooze",
117 "Jim Sound Entley",
118 Enhanced", "Paco De Lucia",
119 )) = 'Jazz/Soul'
120
121 count_df$genre[count_df$artist %in% c('Belle & Sebastian', 'Eternal Summers', 'Elliot
122 Smith',
123 'Matt Costa', 'Weird War', 'Ariel Pink's Haunted
124 Graffiti',
125 'Kelley Stoltz', 'The Greenhornes', 'The Organ',
126 'The Sounds',
127 'Whitey', 'A Band of Bees', 'Birds of Avalon',
128 'Doves',
129 'Luscious Jackson', 'Madonna', 'Old Canes',
130 'Pepper Rabbit',
131 'The Go! Team', 'Depeche Mode', 'The Black Lips',
132 'The Cry',
133 'The Teenagers', 'The Zoo', 'Wonderful Broken
134 Thing',
135 'Liam The Younger', 'Seafood', 'Spindrift',
136 'Amadou & Mariam',
137 'Who Made Who', 'Afterimage', 'The Len Brown
138 Society',
139 'Gifthorse', 'The People's Temple', 'Barry Louis
140 Polisar',
141 'The Downtown Struts', 'Die Jungen', 'The
142 Crocodiles',
143 'The Devil Makes Three', 'Henries Holborn', 'Rock
144 Plaza Central',
145 'Leeroy Stagger', 'White Flight', 'Main Fear
146 'Kid loco', 'Kid Loco', 'Lana del Rey', 'Danger
147 'The Chemical Brothers', 'The 2 Bears', 'The
148 'Puro Instinct', 'Teddybears', 'The Similou',
149 'Teddybears Stockholm', 'Pedro The Lion', 'Matt
150 'Arthur Russell', 'The Minders', 'The Nerves',
151 'Depeche Mode', 'Yael Naim', 'Gwen Stefani',
152 'Edward Sharpe And The Magnetic Zeroes'
153 )) = 'Indie/Alternative'
154
155 count_df$genre[count_df$artist %in% c('The Adolescents', 'The Odd Numbers', 'Bad
156 Strip',
157 'Division of Laura Lee', 'Sub Society', 'Dance
158 Hall Crashers',

```

```

146 Peni', 'The Heartaches',
147 'The Veebees',
148 "Shed", "Blur",
149 Retina", "Embrace",
150 "Rat Fist",
151 Thrash",
152 Taints", "Yuck",
153 "Monarch Box",
154 Dudes", "Badshit",
155 The Murder Junkies",
156 Smith",
157 "The Only Ones",
158 "The Legendary Shack Shakers",
159
160 )] = 'Punk'
161
162 count_df$genre[count_df$artist %in% c('Johnny Cash', 'Love', 'Red Octopus', 'Dick
163 Dale', 'The Small Faces',
164 'Norman Greenbaum', "Focus", "The Monkeys", "The
165 Monkees",
166 "Max Bygraves", "Bobby Pickett and The Crypt",
167 "The Beatles", "Van Morrison", "The Grass Roots",
168 "Shocking Blue", "Syd Barrett", "Syd Barret",
169 "David Bowie & Bing Crosby", "Grateful Dead",
170 "Little Feet", "Loose Gravel", "Tom Petty & The
171 Heartbreakers",
172 "John Mellencamp", "The Rolling Stones", "Tom
173 Petty", "The Velvet Underground",
174 "Billy Idol", "Lou Reed", "The Police", "Cheap
175 Trick", "The Vapors",
176 "The Romantics", "The Solution", "The Beatles",
177 "Paul McCartney", "John Lennon", "Cat Stevens",
178 "Paul Simon",
179 "Neil Diamond", "Hall & Oates"
180
181 )] = 'Classic Rock'
182
183 ## Save data to csv ####
184 #write_csv(count_df, '~/Desktop/cleaned_skatedata.csv3')

```

