

빌보드 차트 뽀개기 : 1960년부터 2009년까지 G.G.L.분석



디지털인문예술 융합전공
인문데이터 마이닝과 파이썬

Input(A+)
구윤제, 정민, 주영훈

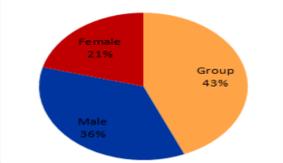
지도교수: 한수미

Objective

1960~2009년까지 Billboard Top 100에 오른 음악의 가수의 성별(Gender), 장르(Genre) 그리고 가사(Lyrics)를 파이썬의 데이터 마이닝과 다양한 시각화를 통한 시대별 변천사에 대한 분석

Gender

전체 시대의 가수 비율 1960~2009년

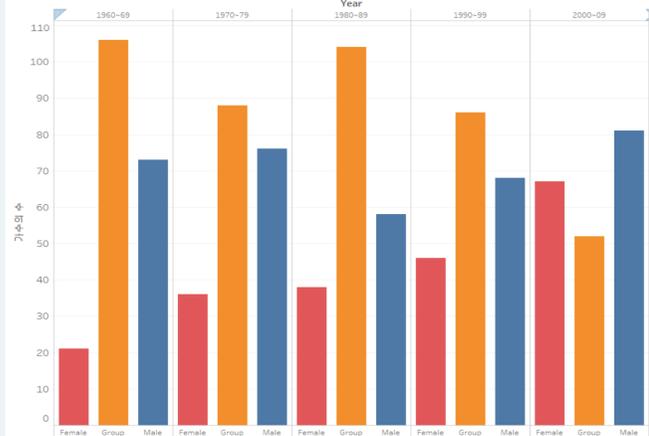


전체 시대에서 가장 높은 비율을 차지하는 것은 그룹이다. 현재는 많은 여성 가수들이 상위권에 들어섰지만 50년의 기간을 합산해 보았을 때, 그룹 비율이 전체 빈도의 절반 정도에 해당한다.

데이터 설명

1960년대부터 1990년대까지는 그룹의 비중이 가장 높았고, 여성의 비율은 낮았다. 하지만, 시대가 흘러갈수록 여성의 비율이 점차 증가하기 시작했다. 그리고 2000년대에 들어와서는 그룹의 비중이 솔로 가수보다 매우 줄어들었고, 여성의 비율이 남성을 거의 따라잡은 것을 볼 수 있다. 앞으로도 여성 가수의 비율이 더 증가할 것으로 전망된다.

시대별 성별 및 그룹 가수 수의 변천사



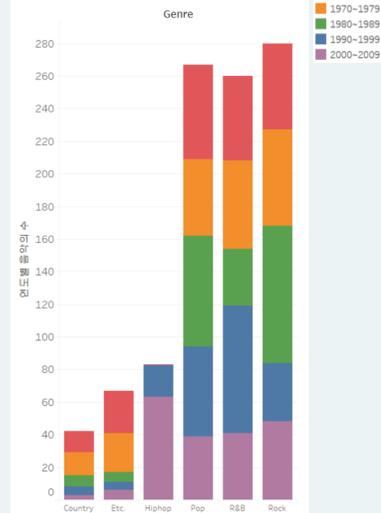
분석방법

1960~2009년까지의 데이터 세트에서 먼저 10년 단위로 5번을 나누었다. 그리고 1년을 단위로 Top20 안에서 장르, 성별, 가사를 세분화해서 분석을 했다. 빌보드차트 웹사이트에 올라와 있는 정보를 바탕으로 총 1000개의 데이터를 얻을 수 있었다. 누락된 정보는 검색을 통해 재조사했다. 최종적으로 python의 textblob, split, token, freq, sort 함수를 사용해 단어를 나누어 빈도수를 분류했다. 이후 가사에서 최다 빈출 단어는 wordle를 이용하고, 성별과 장르는 tableau를 이용하여 그래프로서 데이터 시각화하였다.

Genre

Genre	1960~1969	1970~1979	1980~1989	1990~1999	2000~2009
Rock	53	59	84	36	48
Pop	58	47	68	55	39
Hiphop	0	0	0	20	63
R&B	52	54	35	78	41
Country	13	14	7	5	3
Etc.	26	24	6	5	6

장르의 연도별 음악의 수



다양한 장르가 있었으나 비슷한 장르, 파생된 장르끼리 묶어, 6개의 범주로 나눔

- R&B는 Soul의 상위 장르이기 때문에 둘을 같이 묶었다.
- 그 외에 적은 비중을 차지하는 Jazz, Christian, Blues, Disco, Doo-Wop, Ska, Folk, Instrumental 등은 Etc.로 묶었다.

데이터 설명

1960년대부터 록 음악의 인기가 점차 증가하여, 항상 높은 비중을 차지하던 팝과 함께 1980년대 차트의 대부분을 차지하는 것을 볼 수 있다. 그러다 1980년대까지 차트에는 들지 못하던 힙합이 1990년대부터 강세를 보여 알앤비와 차트를 점령하고, 2000년대에는 최고의 인기 장르가 된 것을 볼 수 있다. 1960~80년대가 록/팝이 강세를 보였다면, 1990년대부터 지금까지의 차트는 힙합과 알앤비가 대세가 되었다고 볼 수 있다. 그밖에 컨트리에는 어느 시대에도 가장 유행하는 장르는 아니었으나, 지속적으로 인기가 줄고 있다. 포크는 1960년대 이후로 차트에 모습을 보이지 않았고, 60년대 초의 두왑, 70년대 말의 디스코 등(Etc.) 한때에만 유행한 장르도 있다. 차트를 통해 10년 단위로 장르의 빈도를 보면서, 음악 장르의 세대 교체를 한눈에 알아 볼 수 있다.

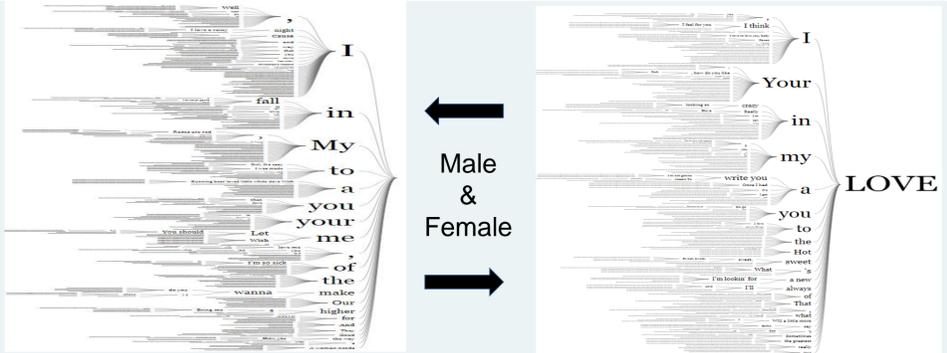
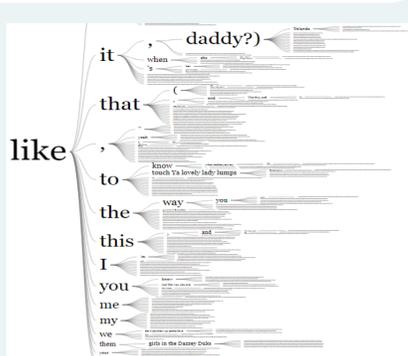
Lyrics



차트 속 노래들에서 가장 많이 나온 단어들을 wordle를 통해 시각화했다. 장르와 성별의 변화는 두드러진 반면, 가사 자체의 변화는 크지 않은 것을 확인할 수 있다. 시대와 장르, 성별을 떠나서 역대 가장 히트한 노래들의 가사에서 가장 많이 등장한 단어는 love와 baby 등 비슷한 단어이다. 그 밖에 눈에 띄는 단어들도 like, girl, sweet 등 사랑에 관련된 단어임을 알 수 있다. 시대가 변하고 유행이 변해도, 사람들은 가사 속에서 사랑에 관한 긍정적인 메시지를 원한다는 것을 알 수 있다.



← Rock & Hiphop



← Male & Female →

가사는 비슷해도 표현 방식에서 장르와 성별로 차이가 있음을 wordtree를 활용하여 특정 키워드 전후에 어떤 문장이 오는지를 통해 분석해왔다. 록과 힙합 장르에선, 일반적으로 많이 쓰이는 like를 통해 표현 방식의 차이에 대해 알아봤다. 록 가사에선, like는 '그것과 같다'라는 의미로, 주로, 대상의 아름다움이나 특성을 비유하는 형식으로 쓰인다. 반면, 힙합에선, '나는 그것이 좋다'라는 의미로 더 많이 쓰인 것을 알 수 있다. 이를 통해 주로 화자가 말하고자 하는 것을 비유를 통해 표현하는 록 가사와 직설적으로 자신이 원하는 것을 표현하는 힙합 가사의 차이점을 알 수 있다. 남자와 여자의 표현방식의 차이는 love라는 단어 전에 어떤 단어가 왔는지의 차이를 통해 알 수 있다. Love 전엔, 남자에 비해 여자가 your이라는 단어의 빈도수가 훨씬 높다는 것을 알 수 있다. 이를 통해 여자 노래의 가사에서 여자는 보통 사랑을 받는, 수동적인 역할을 하는 경우가 더 많다는 것을 알 수 있다.

Conclusion

차트 분석을 통하여 시대의 변화가 히트곡들에 어떤 영향을 끼쳤는지 알아보았다. 시대가 지남에 따라 유행하는 장르가 바뀌고, 성비도 바뀌었지만, 오히려 유행하는 노래들의 메시지는 크게 변하지 않았음을 확인하였다. 하지만, 겉보기에 비슷해보이는 가사들도 성별과 장르에 따라, 시대에 따라 표현방식도 세밀하게 바뀌었음을 확인할 수 있었다. 대부분의 히트곡의 가사는 사랑에 관한 것이지만, 사랑에 대한 표현방식은 조금씩 달랐다. 90년대부터 차트를 점령한 힙합음악의 가사는 비유를 주로 사용하는 락과 팝에 비해 훨씬 직설적이고, 여자가 부른 사랑노래의 화자는 대개 남자에 비해 훨씬 수동적이다. 수천곡을 데이터 마이닝해서 얻어낸 결과로 유행가들에는 시대가 변해도 바뀌지 않는 공통점이 있으면서도, 장르와 성별별로 다른 차이점도 있다는 것을 알 수 있었다.