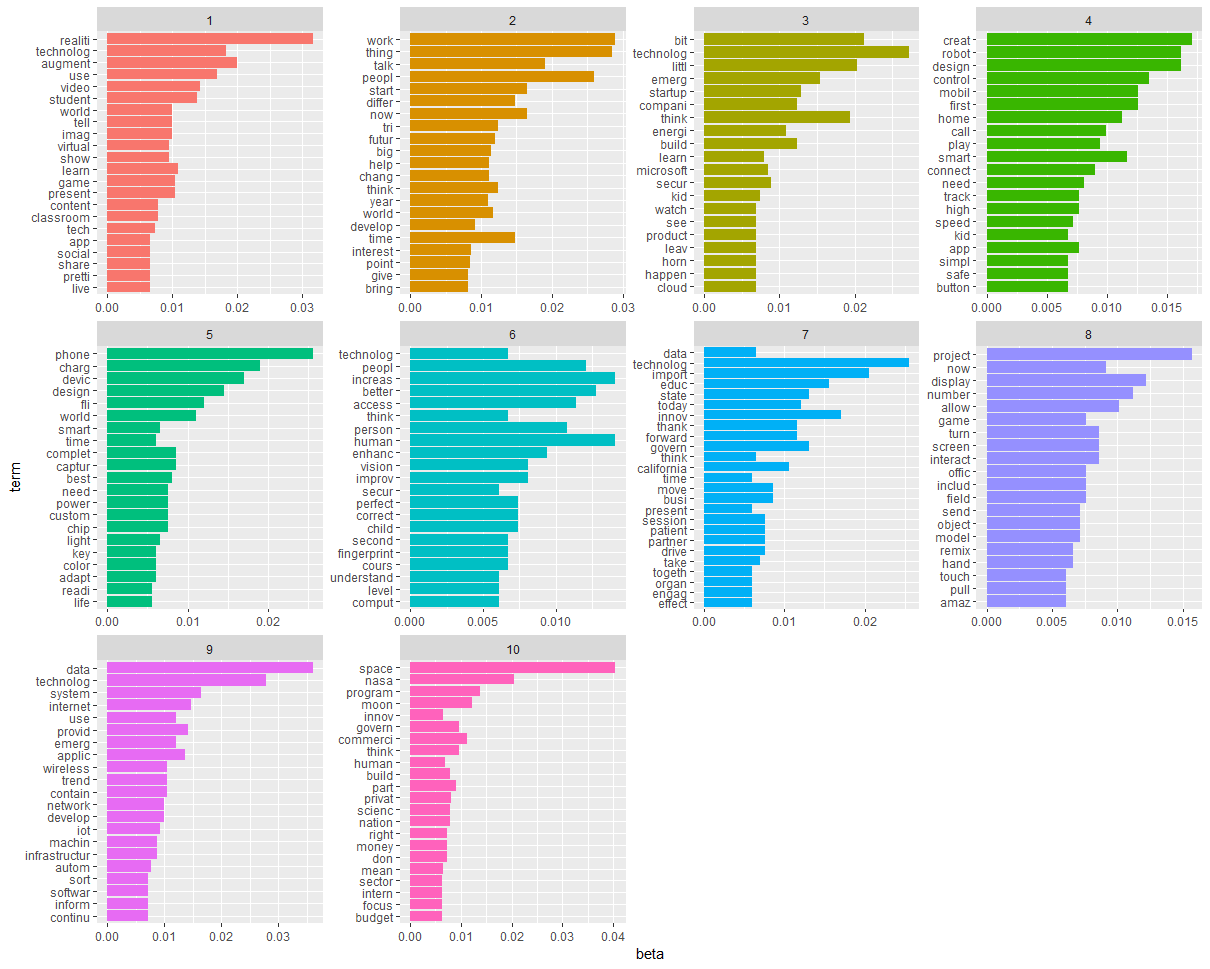
**Class Project Part 2: Topic Model Analysis and Data Mart Reports**

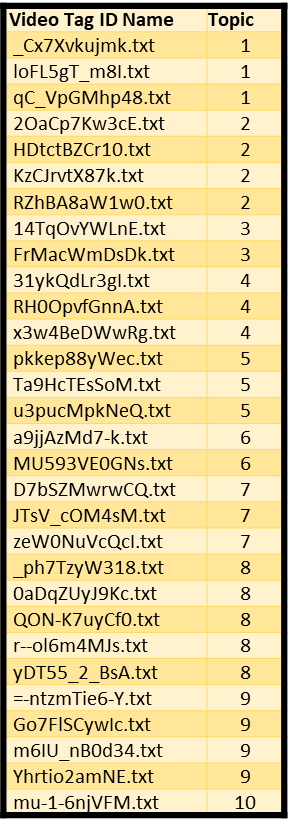
Individuals contributing to this document: Alyson Loschiavo, Shane Varnum, Alexander Ragasa

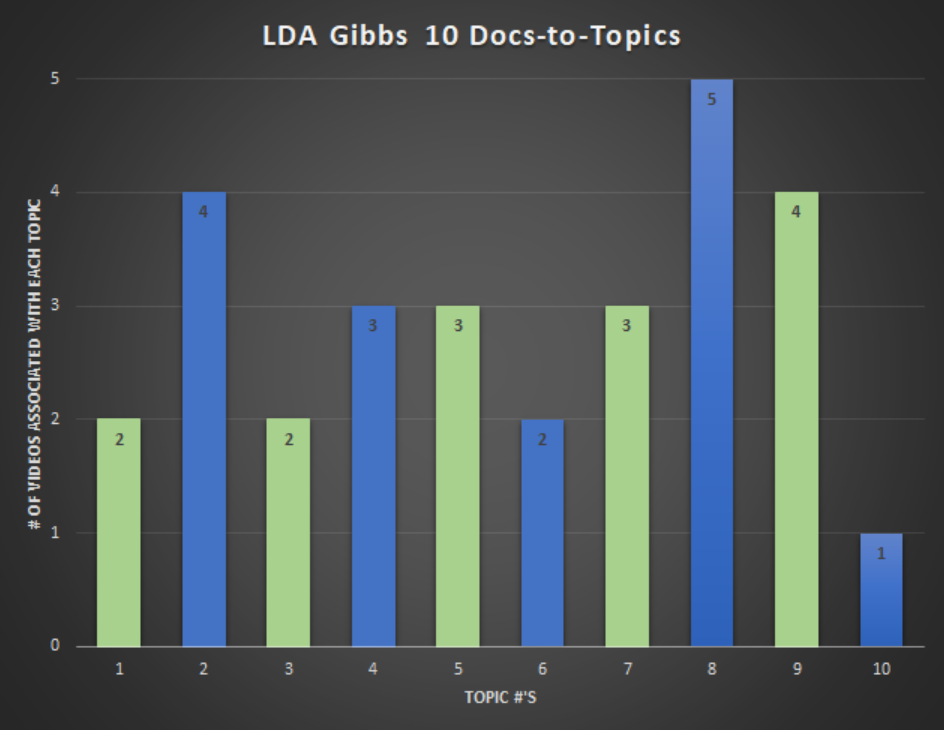
**Topic Model Analysis**



**Topic Plot description:** Topic modeling exists to discover hidden structure patterns of word use and then link them to documents that exhibit similar patterns. The image that was derived from the topic analysis showed the statistical probability/distribution of each word and then clustered the word sets into topics. This is due to the way that several of the natural language processing techniques work – mainly focusing on the Latent Dirichlet Allocation model. This model views each corpus document as a mixture of words and can classify topics based on the probability observed when analyzing the words and documents. Beta is used here and a higher beta value implies that each topic should be likely to contain a specific word mix. It can be implied that a higher beta value leads to topics becoming more similar in regard to the words they contain, or the probability of the term being generated from the topic i.e. “beta spread”. This allows for basically an inference of a topic based on the words grouped within the topic cluster, shown by the plots above. We chose to classify words from their respective documents into a set of k = 10 topics.

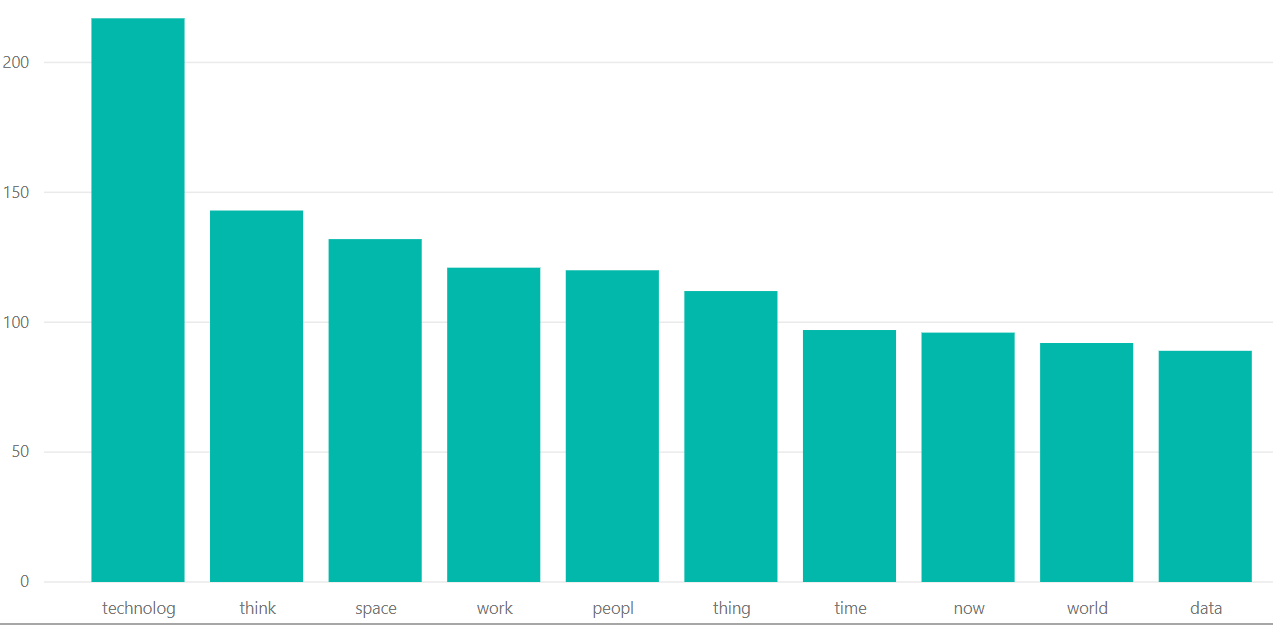
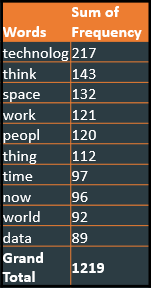
**Docs to Topics Report**





**Explanation:** This Docs-to-Topics reporting places each of the videos into one of the 10 topics created in the topic analysis reporting based on each video’s word use. Essentially, as previously stated, each topic should be likely to contain a specific word mix. The reporting above reflects which of the 30 videos are likely to fall under a specific topic based on the mix of words found.

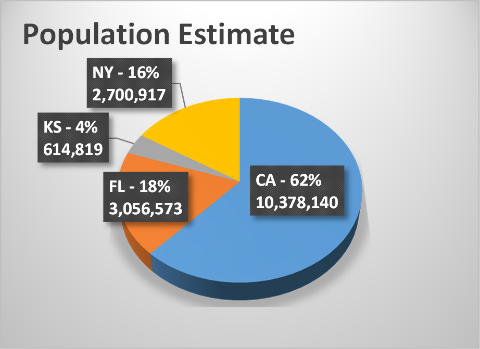
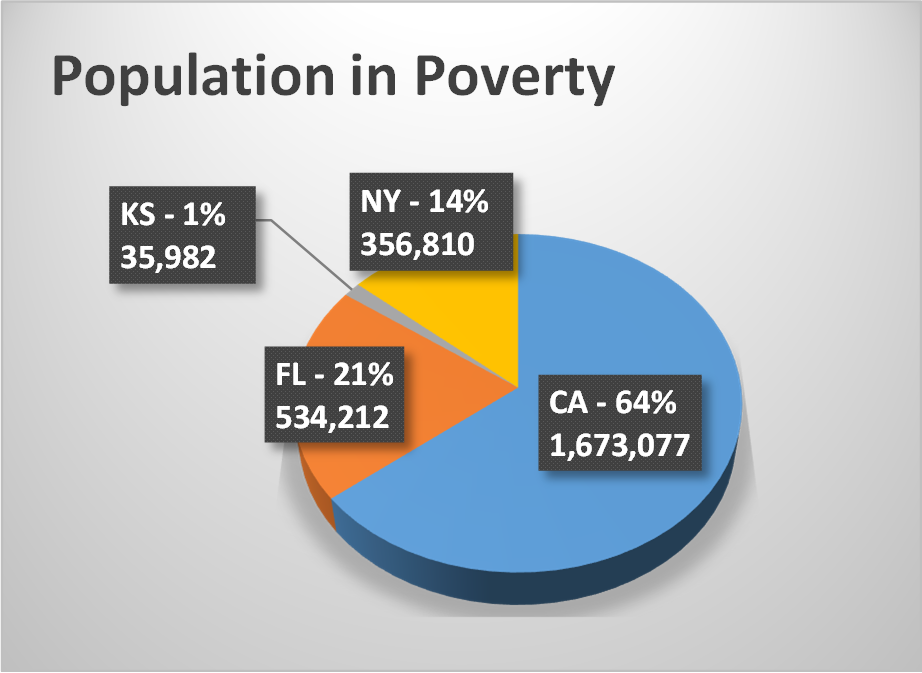
**Word Frequency Report**



**Explanation:** In the Topic Modeling Reporting, the word trend was broken down into 10 different topics. Here in the word frequency reporting we have an overall ranking of the most used words in all 30 videos, and can be reviewed by clicking the excel icon labeled Word Frequency. The visuals above reflect the 10 most used words that came in at an impressive 1219 times and not surprisingly the top spot was claimed by technology being used over 200 times on its own.

**Economic Demographics**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **State** | **Area Name** | **2013 Urban Influence Code** | **2013 Rural urban Continuum Code** | **Population in poverty** | **Population Estimate 2016** |
| CA |  |  |  |  |  |
|  | Alpine County | 4 | 8 | 199 | 1,057 |
|  | Butte County | 2 | 3 | 43,428 | 226,525 |
|  | Los Angeles County | 1 | 1 | 1,629,450 | 10,150,558 |
| FL |  |  |  |  |  |
|  | Liberty County | 7 | 8 | 1,468 | 8,319 |
|  | Escambia County | 2 | 2 | 45,044 | 311,711 |
|  | Miami-Dade County | 1 | 1 | 487,700 | 2,736,543 |
| KS |  |  |  |  |  |
|  | Greeley County | 12 | 9 | 149 | 1280 |
|  | Ellis County | 8 | 5 | 3,453 | 28,823 |
|  | Johnson County | 1 | 1 | 32,380 | 584,716 |
| NY |  |  |  |  |  |
|  | Schuyler County | 7 | 6 | 2,684 | 17,998 |
|  | Rockland County | 1 | 1 | 46,260 | 326,875 |
|  | Queens County | 1 | 1 | 307,866 | 2,356,044 |



**Explanation:** What we were looking to see in the Economic Demographic Data Mart was whether or not there was correlation between the size of a population and the amount of poverty within it. At first glance at the visuals, it would appear they are pretty correlated however if you look a little closer at the state of Kansas, there is a 3% decrease when comparing the population contribution of Kansas vs. their poverty levels and could be worth exploring a little deeper to see if there is meaning behind the change or if it is insignificant. Additionally, we included the Urban Influence Code and Rural Urban Continuum Code reporting to see if living in a metropolitan area or an urban area has any effect on the poverty levels.

**SQL code report 1:**

SELECT Education.DimState.[State], Education.DimAreaName1.[Area Name], Education.DimCode.[2013 Urban Influence Code], Education.DimCode.[2013 Rural-urban Continuum Code],Education.FactTable.[Poverty All] AS 'Population in poverty', Education.FactTable.[Population Estimate 2016]

FROM Education.FactTable

INNER JOIN Education.DimCode

ON Education.FactTable.[Code ID] = Education.DimCode.[Code ID]

INNER JOIN Education.DimAreaName1

ON Education.FactTable.[Area Name ID] = Education.DimAreaName1.[Area Name ID]

INNER JOIN Education.DimState

ON Education.DimAreaName1.StateID = Education.DimState.StateID

WHERE Education.DimState.State IN ('CA', 'KS', 'NY', 'FL')

AND Education.DimAreaName1.[Area Name] IN ('Los Angeles County', 'Butte County','Alpine County','Johnson County', 'Ellis County', 'Greeley County', 'Queens County', 'Rockland County', 'Schuyler County', 'Miami-Dade County', 'Escambia County', 'Liberty County')

ORDER BY Education.DimState.[State], Education.FactTable.[Poverty All];

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | **Economic Demographics** | |  |  |  |
|  | |  | | --- | |  | |  |
|  |  |  |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **State** | **Area Name** | **2013 Urban Influence Code** | **2013 Rural urban Continuum Code** | **Population in poverty** | **Population Estimate 2016** | | CA |  |  |  |  |  | |  | Alpine County | 4 | 8 | 199 | 1,057 | |  | Butte County | 2 | 3 | 43,428 | 226,525 | |  | Los Angeles County | 1 | 1 | 1,629,450 | 10,150,558 | | FL |  |  |  |  |  | |  | Liberty County | 7 | 8 | 1,468 | 8,319 | |  | Escambia County | 2 | 2 | 45,044 | 311,711 | |  | Miami-Dade County | 1 | 1 | 487,700 | 2,736,543 | | KS |  |  |  |  |  | |  | Greeley County | 12 | 9 | 149 | 1280 | |  | Ellis County | 8 | 5 | 3,453 | 28,823 | |  | Johnson County | 1 | 1 | 32,380 | 584,716 | | NY |  |  |  |  |  | |  | Schuyler County | 7 | 6 | 2,684 | 17,998 | |  | Rockland County | 1 | 1 | 46,260 | 326,875 | |  | Queens County | 1 | 1 | 307,866 | 2,356,044 | | | |  |

**Explanation:**

**SQL code report 2:**

SELECT Education.DimState.[State], Education.DimAreaName1.[Area Name], Education.FactTable.[Less than Highschool Diploma 2012-16],

Education.FactTable.[Highschool Diploma Only 2012-16], Education.FactTable.[Some College or Degree 2012-16],

Education.FactTable.[Bachelor's degree or higher 2012-16],Education.FactTable.[Median Household Income 2016]

FROM Education.FactTable

INNER JOIN Education.DimCode ON Education.FactTable.[Code ID] = Education.DimCode.[Code ID]

INNER JOIN Education.DimAreaName1 ON Education.FactTable.[Area Name ID] = Education.DimAreaName1.[Area Name ID]

INNER JOIN Education.DimState ON Education.DimAreaName1.[StateID] = Education.DimState.[StateID]

WHERE Education.DimState.[State] IN( 'CA','KS', 'NY', 'FL')

AND Education.DimAreaName1.[Area Name] IN ('Los Angeles County', 'Butte County','Alpine County','Johnson County', 'Ellis County', 'Greeley County', 'Queens County', 'Rockland County', 'Schuyler County', 'Miami-Dade County', 'Escambia County', 'Liberty County')

ORDER BY Education.FactTable.[Median Household Income 2016];

SELECT Education.DimState.[State], Education.DimAreaName1.[Area Name], Education.FactTable.[Less than Highschool Diploma 2012-16],

Education.FactTable.[Highschool Diploma Only 2012-16], Education.FactTable.[Some College or Degree 2012-16],

Education.FactTable.[Bachelor's degree or higher 2012-16],Education.FactTable.[Population Estimate 2016],Education.FactTable.[Median Household Income 2016]

FROM Education.FactTable

INNER JOIN Education.DimCode ON Education.FactTable.[Code ID] = Education.DimCode.[Code ID]

INNER JOIN Education.DimAreaName1 ON Education.FactTable.[Area Name ID] = Education.DimAreaName1.[Area Name ID]

INNER JOIN Education.DimState ON Education.DimAreaName1.[StateID] = Education.DimState.[StateID]

WHERE Education.DimState.[State] IN( 'CA','KS', 'NY', 'FL')

AND Education.DimAreaName1.[Area Name] IN ('Los Angeles County', 'Butte County','Alpine County','Johnson County', 'Ellis County', 'Greeley County', 'Queens County', 'Rockland County', 'Schuyler County', 'Miami-Dade County', 'Escambia County', 'Liberty County')

ORDER BY Education.FactTable.[Median Household Income 2016];

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |  | | --- | | **Regional Demographics** | |  |  |  |
|  | |  | | --- | |  | |  |
|  |  |  |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **State** | **Area Name** | **Median Household Income 2016** | **Unemployment Rate 2016** | **Population in Group Quarters** | **Population Estimate 2016** | | CA |  |  |  |  |  | |  | Alpine County | $57,481.00 | 6.5% | 24 | 1,057 | |  | Butte County | $45,097.00 | 6.6% | 5,394 | 226,525 | |  | Los Angeles County | $61,308.00 | 5.3% | 176,715 | 10,150,558 | | FL |  |  |  |  |  | |  | Liberty County | $38,900.00 | 5.1% | 2,018 | 8,319 | |  | Escambia County | $45,004.00 | 4.9% | 18,079 | 311,711 | |  | Miami-Dade County | $45,886.00 | 5.3% | 41,757 | 2,736,543 | | KS |  |  |  |  |  | |  | Greeley County | $47,620.00 | 2.1% | 26 | 1,280 | |  | Ellis County | $52,265.00 | 3.1% | 1,033 | 28,823 | |  | Johnson County | $80,872.00 | 3.3% | 5,160 | 584,716 | | NY |  |  |  |  |  | |  | Schuyler County | $48,109.00 | 6.0% | 215 | 17,998 | |  | Rockland County | $84,358.00 | 4.2% | 7,004 | 326,875 | |  | Queens County | $61,844.00 | 4.5% | 27,505 | 2,356,044 | | | | |

**Explanation:**

**SQL code report 3:**

SELECT Education.DimState.[State], Education.DimAreaName1.[Area Name], Education.FactTable.[Median Household Income 2016],

Education.FactTable.[UR 2016] AS 'Unemployment Rate 2016', Education.FactTable.[GQ Estimates 2016] AS 'Population in Group Quarters',

Education.FactTable.[Population Estimate 2016]

FROM Education.FactTable

INNER JOIN Education.DimCode ON Education.FactTable.[Code ID] = Education.DimCode.[Code ID]

INNER JOIN Education.DimAreaName1 ON Education.FactTable.[Area Name ID] = Education.DimAreaName1.[Area Name ID]

INNER JOIN Education.DimState ON Education.DimAreaName1.[StateID] = Education.DimState.[StateID]

WHERE Education.DimState.[State] IN( 'CA','KS', 'NY', 'FL')

AND Education.DimAreaName1.[Area Name] IN ('Los Angeles County', 'Butte County','Alpine County','Johnson County', 'Ellis County', 'Greeley County', 'Queens County', 'Rockland County', 'Schuyler County', 'Miami-Dade County', 'Escambia County', 'Liberty County')

ORDER BY Education.DimState.[State], Education.FactTable.[Population Estimate 2016];