**GENERATING BIG DATA OF CARDIOVASCULAR RISK FACTORS THROUGH GEOTAGGED TWITTER MINING**

**J. Lee**1, J. Sung2
1Jackson State University, Madison, MS, USA

2Associate Professor, Jackson State University, Madison, MS, USA.

**Objective:** Emotional well-being and neighborhood characteristics are significantly linked to cardiovascular disease. Daily routine may usually disrupt emotional health. These can lead to strong feelings of sadness, stress, or anxiety. To cope with stressful situations, people may tend to share their emotion with others through the social media. This study is to determine the relationship between neighborhood characteristics and emotion expressed on Tweets and discuss the applicability of the big data technology to cardiovascular research.

**Method:** Using Twitter's streaming application programming interface, a total of 155,360 tweets were collected between August 15 and 21, 2017 across 6 metropolitan cities: NY, Denver, Atlanta, New Orleans, Miami and Jackson. Geo-coordinates of where tweets were sent were spatially linked to 2010 census data. To implement sentiment analysis, we utilized the National Research Council lexicon which categorizes words into positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. A series of multilevel analyses were conducted to determine the association between each of sentiments and neighborhood factors.

**Results:** Anger, disgust, sadness, surprise and negativity were significantly higher during weekdays than weekend; no difference was found in anticipation, joy, trust and positivity. A highest score of anticipation, trust, joy and positivity were found in Denver. Positive tweets were associated with median house value, median household income and % owner occupied; negative tweets associated with % labor force population, median household income, % owner occupied and population density.

**Conclusion:**Our study revealed that sentiments expressed in the tweets were significantly associated with neighborhood characteristics, which are persistent with the results from the others. Therefore, social media may be valuable and inexpensive data source for understanding the relation between risk factors of cardiovascular disease.