**Potential Project Ideas for**

**Methods for Data Analysis (Course 141030/141033)**

**Data Science (Program 4693/4694)**

1. USA Today keeps a record of all NFL arrests/crimes in this web page: <http://www.usatoday.com/sports/nfl/arrests/> . Are these arrests seasonally correlated? Are they correlated whether the teams wins or loses? Are the different types of crimes related?
2. Twitter Projects:
   1. Sentiment analysis comparison between different search terms.
   2. Finding related words.
   3. Topic modeling on tweets.
   4. Comparing tweet subject frequencies.
3. Spam/Fraud modeling of emails (<https://archive.ics.uci.edu/ml/datasets/Spambase> ), texts (<https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection> ), or other text sources.
4. Kaggle’s public data sets: <https://www.kaggle.com> .
5. Also think about taking any of the above data sets and merging/correlating them with:
   1. Census Data: <http://www.census.gov/data/data-tools.html#>
   2. Weather Data: See weather\_retrieval.R
6. Many major websites have API’s that you can access for free.
   1. Twitter
   2. Linkedin
   3. Foursquare
   4. Open Street Map
7. Amazon released an anonymized purchase dataset here: <http://snap.stanford.edu/data/amazon-meta.html> , which is part of a larger collection of network type data here: <http://snap.stanford.edu/data/index.html>
8. Game Databases:
   1. Chess game database: <http://www.ficgs.com/directory_databases.html>
   2. Poker hand database: <https://archive.ics.uci.edu/ml/datasets/Poker+Hand>
9. Wikipedia traffic statistics:
   1. <https://aws.amazon.com/datasets/6025882142118545>