**THE UNITED TWEETS OF AMERICA – Presentation on Wednesday, 4/18/2018 – DRAFT / Dow**

I. INTRO

Tweets & Means Committee:

The House of Tweetresentatives:

Eric Blank, Jessica Wilbert, Justin \_\_\_\_, Steve Dow

II. EXAMPLES OF DATA USE

III. EVOLUTION OF PROJECT

* INITIAL GOAL: Provide a general sentiment for all tweets and grouped by location (U.S. state),

BUT:

* CHALLENGES / LIMITATIONS:
  + (1) General vs Specified Sentiment Analysis:
    - We learned that we would need to have SOME search term to analyze <<how did we learn this again?)
  + (2) Honing in on Location:
    - We chose lat/long to hone in on location, BUT since so few people enable geolocation in twitter, we gathered this data in two different ways:
      * By using tweets that referenced “places” (i.e. check-in)
      * Using twitter’s “magic” ‘grouping into cities’ capabilities to gather many more tweets (i.e. @facebook)
  + (3) Dealing with States.
    - (A) Changed state-grouping plans. The plan was to loop through each state and return a similar number of tweet returns for each state, BUT we would likely get stuck waiting on several states with limited populations. Also, from a data-integrity perspective, it did make sense to essentially force a smaller sample size for a state with a large volume of tweets just to match the same quantity of a lesser populated state.
    - (B) Twitter does return a location that often lists “City, State” BUT we found that this was not consistent for each tweet. These fields in other tweets, for example, were populated with “State, USA”. There were also some instance with other data returned as well. In order to avoid this, the plan was to pull the returned lat/long from the tweet and run those through the google geocode tool to match these more precisely to a state. Analysis by state could then more reliably be returned.
    - (C) We ultimately discovered the plotly tool, which effectively plots our sentiment on a US map. (The grouping by state plan was abandoned, but can be added back in if specific state-level analysis is desired).