**PS 239T Final Project Proposal**

**Chuyue Frances Huang**

**Title:**

Text Analysis of Chinese Social Media Relating to the 2016 Presidential Election

**Data Collection:**

1. Sina Weibo(the Chinese microblogging website, equivalent of Quora): I will be accessing their API and getting the Chinese tweets there. Many Chinese scholars have done this before and performed comparative studies on Sina Weibo (the Chinese microblogging website) and Twitter.
2. Zhihu (the Chinese equivalence of Quora): I will be scrapping their website, and in particular, I will be scrapping questions under the topics relating to to politics and the 2016 American election. Some of the megadata I will collect are requester, respondents, question title, date, and number of likes. I have checked out the structure of their webpages, and I think I will be able to identify the XML structure and parse the html accordingly.

**Research Question:**

Anecdotal evidence (and my own experience) indicates that the Chinese, in general, lack profound understanding of American politics. Yet amidst the 2016 US presidential election, the surprising outcome leads to a surge of interests among the Chinese, who are increasingly reliant on social media platforms for news.

In what ways has the Chinese respond to the result of the 2016 presidential election? Has there been a general increase of interests in learning about American politics? Is there a difference between how the Chinese netizens view the two candidates and the promise of democracy before and after the election?

**Analysis:**

1. First, I will visualize the number of tweets/posts before and after the election and see if there is a significant jump at the day of the election.
2. Second, I will use sentiment analysis in Chinese and try to assign a sentiment score to all the posts/tweets relating to the same topic and compare them across different subjects. I will have a separate plot showing what kind of tweets/posts (positive or negative) tend to get t the most likes.
3. Third, I want to use text analysis to see what are the top terms/words that people associate with the election.

**Skills:**

1. Webscrapping (for the Chinese quora website)
2. Accessing the API
3. Text Analysis and creating a corpus

*Note: I found some softwares that would allow me to perform text analysis in Chinese. The one tool I will use to parse the Chinese text is called “txtorg.”*

1. Data visualization using ggplot in R

**Additional Comment:**

I might also try to see if I can get any data from WeChat, the largest social media platform (equivalence of Facebook), although as of now it looks very encrypt and inaccessible.

**For sentiment analysis:**

**Semantria:** https://online.semantria.com/

**Plan update: (11-28)**

* Kicking up Sina Weibo, only doing **Zhihu**

Part 1:using the知乎日报API to get the info before and after the election

Part 2: another 知乎API to get the latest number of questions on the topic

Part 3: scarpe the search pages to get the answers to questions on the following topic

* Search terms:
  + 美国大选
  + 川普
  + 克林顿
  + 希拉里
  + 特兰普
  + 大选
  + 美国政治