Stat 222 Project 3: Twitter

1 Data Description

For this project, your primary data source will be Twitter. And you will be expected to work with the Twitter API using Python to access this data. On Monday (2/23), we will be briefly review Python, introduce JSON (Javascript Object Notation), and demonstrate how to interact with the Twitter API using the Python Twitter Tools. However, you will need to do outside reading to get up to speed with these tools.

Python Twitter Tools is one of several Python packages for interacting with the Twitter API. It is fairly minimal and is the package used in chapter 1 and 9 of *Mining the Social Web*.

- Chapter 1: Mining Twitter: Exploring Trending Topics, Discovering What People Are Talking About, and More

- Chapter 9: Twitter Cookbook

2 Your Assignment

Each group is responsible for creating a presentation that visually answers a set of questions about this data. *Which* questions you answer is up to you, but think about telling a story. The story will be more interesting if the questions you address are related to each other in some way. Here are a few example topics:

- investigate the relation of breaking news on Twitter versus traditional news sources

- compare stop words usage on Twitter versus NY Times

- chart how the ratio of positive versus negative words used in tweets involving some event (or issue) change over time

- relate tweets about a TV show/movie/book to their viewers/ticket sales/sales over time

Timeline

- Monday (2/23) Begin Twitter project

- Wednesday (2/25) Poster presentations for airline data

- Monday (3/2) Introduction to text mining (e.g., NLTK)

- Wednesday (3/4) Pecha Kucha

- Monday (3/9)

- Wednesday (3/11) Practice presentations

- Monday (3/16) Final presentations

3   Initial Guidance (to do for first data debrief)

First,

4   Next Steps

On Wednesday

5   Presentation Details

All presentations will be given in the **Pecha Kucha**[1] style. Pecha Kucha presentations have a very rigid format and have grown in popularity. A Pecha Kucha presentation consists of 20 slides that are automatically advanced 20 seconds (20x20). The complete presentation lasts exactly 6 minutes and 40 seconds.

This is a very constrained format. So you will need to carefully plan and prepare your talk. Each group will have 4 members and each member will be responsible for presenting 5 slides. Since the slides will automatically advance, you will need to practice your talks before you present in class.

We will also be required to have a practice presentation

How to make slides

There are many ways to create slides, but make sure that you are able to save your slides as a PDF. Here are some possibilities for you to explore:

- Beamer
  http://web.mit.edu/rsi/www/pdfs/beamer-tutorial.pdf

- Pandoc
  http://johnmacfarlane.net/pandoc/demo/example9/producing-slide-shows-with-pandoc

- Powerpoint

- Keynote

How NOT to make slides

- Tufte's *PowerPoint Is Evil*
  http://archive.wired.com/wired/archive/11.09/ppt2.html

- Norvig's *Gettysburg Cemetery Dedication*
  http://norvig.com/Gettysburg/sld001.htm

- Efron's *Thirteen rules*
  http://statweb.stanford.edu/~ckirby/brad/other/2013ThirteenRules.pdf

---

[1]http://en.wikipedia.org/wiki/PechaKucha