# Homework (HW11) Sentiment Analysis

### General Instructions

For this homework you will upload 1 R file into blackboard.

Reminder:

* All HW must start with an Identification Block like this sample…

################################################

# IST 387/687, Standard Homework Heading

#

# Student name:

# Homework number:

# Date due:

#

# Attribution statement: (choose the statements that are true)

# 1. I did this work by myself, with help from the book and the professor

# 2. I did this work with help from the book and the professor and these Internet sources: <provide the urls>

# 3. I did this work with coaching from <Name of another student> but did not cut and paste any code

# Run these three functions to get a clean test of homework code

dev.off() # Clear the graph window

cat('\014') # Clear the console

rm(list=ls()) # Clear all user objects from the environment!!!

# Set working directory

# Change to the folder containing your homework data files

setwd("~/MyDesktop/ISTX87/Homework")

This homework builds on our efforts from the Prep Exercise and depends on a careful read of Chapter 14 and 15 of *An Introduction to Data Science*. As usual we’ll use the Prep Ex R file that you created as a starting point. In this homework, we build upon the activities described in Chapters 14 and 15 to count and plot the negative and positive words in a political speech. We have cleaned up the speech file and the list of positive and negative words to streamline your work.

### HW11

**Step 1: Create a list of word counts from the speech.**

1. Create a named list of word counts by frequency, then sort as shown below:

**wordCounts <- sort(wordCounts, decreasing=TRUE)**

1. Run the *head(wordCounts)* command and explain what you see in a block comment.
2. In a block comment, state how many *unique* words there are in the speech and what R code did you use to determine this answer?
3. In a block comment, state how many *total* words there are in the speech and what R code did you use to determine this answer?

**Step 2: Match the speech words with positive and negative words.**

1. Using the code provided below match the words from the speech to the list of positive words.

**matchedP <- match(names(wordCounts), posWords, nomatch = 0)**

1. Create a similar line of code to match the speech to the negative words. In a block comment explain the code that you wrote.
2. Write a block comment that described the information that “matchedP” contains. How long is that vector? What do the numeric values in it represent?

**Step 3: Making bar charts of the words that matched.**

1. Using ggplot, make a bar chart of the positive matches.
2. Using ggplot, create a bar chart of the top 20 negative matches.
3. Create two additional bar charts that only shows the negative, and positive words, that occurred at least twice.
4. Calculate the ratio that shows the proportion of positive words relative to the total number of words in the speech. Do the same for negative words. In a comment, explain if the speech was a positive or negative speech.

***You must submit all Homework to blackboard prior to the deadline specified for each assignment.*** Late HW assignments will not be accepted for credit.

*Materials provided by the instructors of this course are copyrighted 2018 by Jeffrey Saltz and Jeffrey Stanton, or by the respective instructor that produced them. These materials are provided to enrolled students for their exclusive use during the semester of instruction. Uploading instructor-produced materials to Internet sites (e.g., Coursehero) without permission is considered a breach of professional ethics and will be reported as a possible academic integrity violation.*