Report: week of Dec. 28:

- This week, I tried to analyze the periods I found in Adams' positivity / negativity graph in which positivity and negativity were decreasing most rapidly.
  - For example, I found that both positivity and negativity decreased almost linearly between 1775 and 1777. Furthermore, I found that they increased almost linearly between 1777 and 1778.
  - To understand what caused this change, I performed a PMI analysis on the set of all letters written by Adams during both of these time periods.
  - Some of the results that I found are displayed below:
    - This first set of words is the sorted set of words for Adams' letters between 1775 and 1777. The words are sorted in reverse order by: s(w) = PMI(w, “good”)

good 2.81628804683
may 1.41503749928
shall 1.0
upon 0.736965594166
great 0.65207669658
would 0.0
every -0.13750352375
one -0.192645077942
must -0.700439718141
make -0.765534746363
much -1.0
well -1.37851162325
think -1.41503749928
give -1.68805599369
know -1.83007499856
new -2.0
general -2.0
congress -2.0
wish -2.15527822548
letter -2.22239242134
hope -2.28540221886
us -2.45943161864
could -2.58496250072
without -2.64385618977
made -2.66296501272
many -2.80735492206
america -3.08746284125
adams -3.16992500144
men -3.24792751344
people -3.39231742278
take -3.58496250072
first -3.62449086491
public -3.7725895039
time -3.90689059561
let -4.12928301694

ever -4.1963972128
account -4.21723071622
american -4.24792751344
john -4.32192809489
opinion -4.33390073655
never -4.35755200462
states -4.39231742278
whole -4.40939093614
send -4.48112668974
put -4.48112668974
present -4.49185309633
part -4.52356195606
see -4.55458885168
might -4.58496250072
officers -4.60880924268
cannot -4.70043971814
gentlemen -4.7414669864
since -4.80735492206
army -4.80735492206
long -4.88264304936
man -4.89077093025
state -4.90689059561
said -4.95419631039
taken -4.9772799235
ought -5.0
friend -5.01495034147
little -5.1497471195
hear -5.16992500144
england -5.18982455888
say -5.20945336563
nothing -5.20945336563
necessary -5.26052755022
find -5.28540221886
things -5.32192809489
even -5.33985000288
believe -5.43740531231
yet -5.45943161864
answer -5.47031993478
whose -5.59991284219
service -5.61470984412
humble -5.67242534197
servant -5.72792045456
received -5.72792045456
honour -5.75488750216
best -5.76818432478
making -5.86418614465
soon -5.93073733756
pay -5.94251450534
certainly -5.96578428466

The next set of words is the sorted set of words for Adams' letters between 1777 and 1778. The words are sorted in reverse order by: s(w) = PMI(w, “good”).

good 1.88582898008
may 0.736965594166
shall 0.0
upon -0.263034405834
one -1.41503749928
much -1.73696559417
great -1.80735492206
every -1.87446911792
would -2.0
make -2.08746284125
general -2.58496250072
congress -2.58496250072
must -2.70043971814
us -2.87446911792
time -2.90689059561
hope -2.96347412397
think -3.0
sir -3.16992500144
new -3.22239242134
made -3.24792751344
give -3.27301849441
america -3.50250034053
well -3.70043971814
know -4.0
wish -4.02974734339
cannot -4.11547721742
take -4.16992500144
never -4.35755200462
states -4.39231742278
dear -4.4262647547
whether -4.45943161864
without -4.64385618977
even -4.75488750216
ever -4.78135971352
therefore -4.80735492206
whole -4.82442843542
state -4.90689059561
said -4.95419631039
send -5.06608919046
england -5.18982455888
first -5.20945336563
find -5.28540221886
pay -5.35755200462
people -5.39231742278
court -5.4262647547
french -5.44294349585

go -5.45943161864
man -5.47573343097
present -5.49185309633
account -5.53915881111
see -5.55458885168
might -5.58496250072
whose -5.59991284219
gentlemen -5.7414669864
honour -5.75488750216
best -5.76818432478
among -5.78135971352
opportunity -5.80089989992
army -5.80735492206
men -5.83289001416
least -5.87036471958
ships -5.90689059561
hear -5.90689059561
english -6.0
believe -6.02236781303
war -6.02236781303
boston -6.04439411936
put -6.06608919046
british -6.08746284125
power -6.16992500144
dont -6.2288186905
france -6.26678654069
country -6.32192809489
gentleman -6.33539035469
public -6.35755200462
also -6.37503943135
greatest -6.38370429247
troops -6.44846050082
news -6.46760555008
commissioners -6.533978572
three -6.56985560833
order -6.61470984412
enough -6.70043971814
come -6.70043971814
let -6.71424551767
however -6.75488750216
obliged -6.82017896242
money -6.83289001416
making -6.86418614465
opinion -6.91886323727
officers -6.93073733756
indeed -6.94836723158
perhaps -6.95419631039

The next step will be to find more time periods automatically and figure out what other kinds of analysis to do on them. A positive or negative “plateau” in positivity or negativity might be useful to look at.

– can also try to do this calculation for smaller periods of time (2 to 6 months).

For all years in which I have a lot of data for John Adams (1775-1785), I performed a separate pmi analysis (i.e. which words were most commonly associated with the word “good” in these time periods?). Some of my observations and findings are presented below:

- letters' average sentiment don't change too much over time
- There is a very noticeable and definite negative correlation between number of letters and sentiment of letters. This should be accounted for & adjusted in the final data collection.
  - I believe this is due to the fact that when more letters are written, a larger percentage of them happen to be low-sentiment letters.
- One particularly good period:
  - 1776 has the same percentage of low-sentiment letters as 1777, however, 1776 has a much higher average negativity.
- To do next:
  - factor out the more businesslike letters?
  - Compare the pmi for each year with the normal pmi for adams; then try some analysis.

Next, I performed an analysis to find, for each year, the biggest differences between this pmi score, and the overall pmi score for all of Adams' letters. This should tell me which words John Adams is using more than usual in these time periods, and if mined properly, should tell me something about his personality.

- Question: how to proceed after this? I have found some interesting & heterogeneous word lists, but I'm not completely sure how to mine this properly
- Perhaps find groups of words that go together using LDA, or some other clustering metric? Will this be a useful way of generally defining his responses to events that happened from year to year?