Objects to Data

Group Proposal

Martin Roger Ze Schneider, Lars Mangel, Dominique Brouwer, Cassandra Sijtsma, Dianne Ocampo

Question

How often are Hillary Clinton and Donald Trump mentioned on media outlets such as The New York Times or The Guardian? Does the frequency increase as Election Day nears? Who is mentioned more and how are they discussed?

In hindsight, considering the last week, I think it will be inevitable. Given Trump's inclination to put his foot in his mouth, you may want to nuance this question a bit.

Method

- Scrape data off NYT API and Guardian API
- Filter results according to timeline of August to October
  - August is after both the RNC and DNC, October is just after the first Presidential Debate.
- Filter results according to ranking of keywords in articles
- Filter results according to what section articles appear in on the website
- When there are high spikes in frequency, we can decide whether or not to analyze further and see what context the candidate is discussed in (positive or negative?)
- Use RStudio to plot

I think we need a bit more clarity here. This list is fine, but it is just a list. Explain a bit more why you are doing all this.
This is also a mixture of your method and dataset discussion. These should be separated (more on the DATA bit below.

Visualisation

- How often is Hillary Clinton mentioned from August to October? Linear distribution
- How often is Donald Trump mentioned? Linear distribution
- Who is mentioned more and when? Shaded range

This tells me what you are going to visualise, but very little about how. I could use some more information here about what you mean by a "shaded range" and what kind of linear distribution (i.e. what variables)?

Data

- NYT API
- Guardian API