Third Obama-Romney Presidential Debate
Claimbuster Data, EDA

Sofia Bliss-Carrascosa

FACT CHECKED DATA SET

README Info from Claimbuster Datasets

The ClaimBuster dataset consists of six files, the three used are the following: groundtruth.csv, (1032 sentences) crowdsourced.csv, (22,501 sentences) all_sentences.csv, (32,072 sentences)

Both groundtruth.csv and crowdsourced.csv files contain the following attributes. - Sentence_id: A unique numerical identifier to identify sentences in the dataset. - Text: A sentence spoken by a debate participant. - Speaker: Name of the person who verbalized the Text. - Speaker_title: Speaker's job at the time of the debate. - Speaker_party: Political affiliation of the Speaker. - File_id: Debate transcript name. - Length: Number of words in the Text.
- Line_number: A numerical identifier to to indicate the order of the Text in the debate transcript. - Sentiment: Sentiment score of the Text. The score ranges from -1 (most negative sentiment) to 1 (most positive sentiment). - Verdict: Assigned class label (1 when the sentence is CFS,0 when the sentence is UFS, and -1 when sentence is NFS).

all_sentences.csv file contains all presidential debate sentences. It has all the features shown above except for "Verdict''. It also includes the following attribute: - Speaker_role: It depicts the role of the Speaker in the debate as a participant.

Joined relevant claims in `groundtruth` and `crowdsourced` dataset the gather the most complete look at the data.

FACT CHECKED DATA SET

```
## Joining, by = c("Verdict", "Speaker", "Text", "Speaker_title", "Speaker_party", "File_id", "Sentence_
```

Table 1: # of Checked Claims

Speaker	count
Obama	48
Romney	59

Total claims checked: 107

```
ob_rom_factcheck %>%
  group_by(Speaker) %>%
  summarize(count = n()) %>%
  kable(caption = "# of Total Claims")
```

Table 2: # of Total Claims

Speaker	count
Obama	367
Romney	489

Table 3: Unchecked versus Checked Claims

Checked	Obama	Romney
0	319	430
1	48	59

13.1% of Obama's statements were checked, compared to 12.1% of Romney's.

No apparent partisan bias

Table 4: Checkability Score for All Claims

Checkability	Obama	Romney
Checkworthy Factual Statement	107	115
Not Factual Statement	205	327
Uncheckworthy Factual Statement	55	47

Table 5: Checkability Score for Checked Claims

Checkability	Obama	Romney
Checkworthy Factual Statement	25	27
Not Factual Statement	14	16
Uncheckworthy Factual Statement	9	16

Of all claims checked: 52.1% of Obama's claims were Checkworthy Factual Statements, compared to 45.8% of Romney's. 29.2% of Obama's claims were Not Factual Statements, compared to 27.1% of Romney's. 18.8% of Obama's claims were Uncheckworthy Factual Statements, compared to 27.1% of Romney's.

Of claims checked, the most notable difference was in Uncheckworthy Factual Statements, with about 8.3% more of Romney's UFS checked.

Of all total claims: 29.2% of Obama's claims were Checkworthy Factual Statements, compared to 23.5% of Romney's. 55.9% of Obama's claims were Not Factual Statements, compared to 66.9% of Romney's. 15.0% of Obama's claims were Uncheckworthy Factual Statements, compared to 9.6% of Romney's.

The distribution of total claims is much different than that of checked claims. Checkworthy Factual Statments make up a significantly smaller percentage of total claims than checked claims, which makes logical sense. The greatest partisan difference among total claims is that Romney made about 11% more Not Factual Statements than Obama, and Obama made about 1.6x as many Uncheckworthy Factual Statements as Romney's.

Comparing checked claims to total claims: Obama 23.4% of Checkworthy Factual Statements were checked. 6.8% of Not Factual Statements were checked. 16.4% of Uncheckworthy Factual Statements were checked.

Romney 23.5% of Checkworthy Factual Statements were checked. 4.9% of Not Factual Statements were checked. 34.0% of Uncheckworthy Factual Statements were checked.

Obama and Romney had an extremely similar percentage of CFS checked. Obama had a higher percentage (by about 1.4x) of his Non Factual Statements checked. Romney had a much higher percentage (about 2.1x) of his Uncheckworthy Factual Statements Checked. It may be worth further investigating if this shows a partisan bias.

Table 6: More or Less Factual Statements?

Speaker	avg
Obama	-0.2670300
Romney	-0.4335378

With -1 meaning Not Factual Statement, 0 meaning Uncheckworthy Factual Statement, and 1 meaning Checkworthy Factual Statement, I can conclude that in this debate, both candidates made more Nonfactual statements. This data indicates that Obama may make more Factual Statements than Romney.

Table 7: Instances of Repeated Checks

times.fact.checked	Obama	Romney
0	316	426
1	31	34
2	10	15
3	6	6
4	4	5
6	0	1
7	0	1

The distribution of number of times a claim was checked by speaker is relatively equal, Romney had significantly more unchecked statements as he originally made more statements than Obama.

Table 8: Fact Checks by Publisher

Speaker	CNNcheck	Rcheck	APcheck	NYTcheck	Atlcheck	PFcheck	WPcheck	FCcheck
Obama	6	6	13	22	7	8	13	11
Romney	12	5	16	21	6	8	25	21

1/8 fact checkers checked both speakers equally. (PolitiFact) 3/8 fact checkers checked Obama more than Romney. (Reuters, the Atlantic, the NYT) 4/8 fact checkers checked Romney more than Obama. (CNN, AP News, the Washington Post, Fact Check)