

# Third Obama-Romney Presidential Debate

## Claimbuster Data, EDA

```
library(rvest)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyr)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v stringr 1.4.0
## v readr   2.1.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter()      masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()         masks stats::lag()
```

```
library(knitr)
```

```
all_sentences = read_csv("data/all_sentences.csv")
```

```
## Rows: 32072 Columns: 10
```

```
## -- Column specification -----
## Delimiter: ","
## chr (6): Text, Speaker, Speaker_title, Speaker_party, Speaker_role, File_id
## dbl (4): Sentence_id, Length, Line_number, Sentiment
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
crowdsourced = read_csv("data/crowdsourced.csv")
```

```
## Rows: 22501 Columns: 10
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (5): Text, Speaker, Speaker_title, Speaker_party, File_id  
## dbl (5): Sentence_id, Length, Line_number, Sentiment, Verdict  
  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
groundtruth = read_csv("data/groundtruth.csv")
```

```
## Rows: 1032 Columns: 10
```

```
## -- Column specification -----  
## Delimiter: ","  
## chr (5): Text, Speaker, Speaker_title, Speaker_party, File_id  
## dbl (5): Sentence_id, Length, Line_number, Sentiment, Verdict  
  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
rom_ob_all = all_sentences %>%  
  filter(File_id == "2012-10-22.txt")
```

```
rom_ob_crowd = crowdsourced %>%  
  filter(File_id == "2012-10-22.txt")
```

```
rom_ob_truth = crowdsourced %>%  
  filter(File_id == "2012-10-22.txt")
```

```
rom_ob_entire <- rom_ob_all %>%  
  filter(Speaker == "Mitt Romney" |  
         Speaker == "Barack Obama")
```

```
full_join(rom_ob_crowd, rom_ob_truth, Sentence_id = Sentence_id)
```

```
## Joining, by = c("Sentence_id", "Text", "Speaker", "Speaker_title", "Speaker_party", "File_id", "Leng
```

```
## # A tibble: 835 x 10
```

```
##   Sentence_id Text          Speaker Speaker_title Speaker_party File_id Length  
##   <dbl> <chr>          <chr>    <chr>          <chr>      <chr>  <dbl>  
## 1     26902 And thank yo~ Mitt Ro~ Governor    REPUBLICAN  2012-1~    11  
## 2     26903 Thank you to~ Mitt Ro~ Governor    REPUBLICAN  2012-1~     9  
## 3     26904 And Mr. Pres~ Mitt Ro~ Governor    REPUBLICAN  2012-1~    10
```

```
## 4      26905 We were toge~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      21
## 5      26907 This is obvi~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      40
## 6      26908 With the Ara~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      40
## 7      26909 But instead,~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      13
## 8      26910 Of course we~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      15
## 9      26911 We see in --~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      29
## 10     26912 Our hearts a~ Mitt Ro~ Governor      REPUBLICAN      2012-1~      10
## # ... with 825 more rows, and 3 more variables: Line_number <dbl>,
## #      Sentiment <dbl>, Verdict <dbl>
```

```
#essentially, these datasets are identical
write.csv(rom_ob_truth, 'rom_ob.csv', row.names=FALSE)
```

## README Info from Claimbuster Datasets

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

## FACT CHECKED DATA SET

```
ob_rom_factcheck <- read.csv("data/obrom_checkeddata.csv")

ob_rom_factcheck <- ob_rom_factcheck %>%
  mutate(Checked = Checked.)
```

```
ob_rom_factcheck %>%
  filter(Checked == 1) %>%
  group_by(Speaker) %>%
  summarize(count = n()) %>%
  kable()
```

Speaker	count
Obama	48
Romney	57

Total claims checked: 105

```
ob_rom_factcheck %>%
  count(Checked, Speaker) %>%
  pivot_wider(id_cols = c(Checked, Speaker),
               names_from = Speaker,
               values_from = n,
               values_fill = 0) %>%
  kable()
```

Checked	Obama	Romney
0	311	417
1	48	57

13% of Obama's statements were checked, compared to 12% of Romney's.

*No apparent partisan bias*

```
ob_rom_factcheck <- ob_rom_factcheck %>%
  mutate(Checkability =
    case_when(Verdict == "-1" ~ "Not Factual Statement",
              Verdict == "0" ~ "Uncheckworthy Factual Statement",
              Verdict == "1" ~ "Checkworthy Factual Statement"))

ob_rom_factcheck %>%
  count(Checkability, Speaker) %>%
  pivot_wider(id_cols = c(Checkability, Speaker),
               names_from = Speaker,
               values_from = n,
               values_fill = 0) %>%
  kable()
```

Checkability	Obama	Romney
Checkworthy Factual Statement	106	112
Not Factual Statement	198	318
Uncheckworthy Factual Statement	55	44

```
ob_rom_factcheck %>%
  filter(Checked == 1) %>%
  count(Checkability, Speaker) %>%
  pivot_wider(id_cols = c(Checkability, Speaker),
               names_from = Speaker,
               values_from = n,
               values_fill = 0) %>%
  kable()
```

Checkability	Obama	Romney
Checkworthy Factual Statement	25	26
Not Factual Statement	14	15

Checkability	Obama	Romney
Uncheckworthy Factual Statement	9	16

*Of all claims checked:* 52.1% of Obama's claims were Checkworthy Factual Statements, compared to 45.6% of Romney's. 29.2% of Obama's claims were Not Factual Statements, compared to 26.3% of Romney's. 18.8% of Obama's claims were Uncheckworthy Factual Statements, compared to 28.1% of Romney's.

**Of claims checked, the most notable difference was in Uncheckworthy Factual Statements, with about 10% more of Romney's UFS checked.**

*Of all total claims:* 29.5% of Obama's claims were Checkworthy Factual Statements, compared to 23.6% of Romney's. 55.2% of Obama's claims were Not Factual Statements, compared to 67.1% of Romney's. 15.3% of Obama's claims were Uncheckworthy Factual Statements, compared to 9.3% 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 12% more Not Factual Statements than Obama, and Obama made about 1.5x as many Uncheckworthy Factual Statements as Romney's.**

*Comparing checked claims to total claims:* Obama 23.6% of Checkworthy Factual Statements were checked. 7.1% of Not Factual Statements were checked. 16.3% of Uncheckworthy Factual Statements were checked.

Romney 23.2% of Checkworthy Factual Statements were checked. 4.7% of Not Factual Statements were checked. 36.4 % of Uncheckworthy Factual Statements were checked.

**Obama and Romney had an extremely similar percentage of CFS checked. Obama had a higher percentage (by about 1.5x) of his Non Factual Statements checked. Romney had a much higher percentage (about 2x) of his Uncheckworthy Factual Statements Checked. It may be worth further investigating if this shows a partisan bias.**

```
ob_rom_factcheck %>%
  group_by(Speaker) %>%
  summarize(avg = mean(Verdict)) %>%
  kable()
```

Speaker	avg
Obama	-0.2562674
Romney	-0.4345992

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.

```
ob_rom_factcheck %>%
  filter(times.fact.checked != "#REF!") %>%
  count(times.fact.checked, Speaker) %>%
  pivot_wider(id_cols = c(times.fact.checked, Speaker),
    names_from = Speaker,
    values_from = n,
    values_fill = 0) %>%
  kable()
```

times.fact.checked	Obama	Romney
0	308	413
1	31	33
2	10	15
3	6	5
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.

```
CNN <- ob_rom_factcheck %>%
  filter(CNN == "1") %>%
  group_by(Speaker) %>%
  summarize(CNN_check = n())

Reuters <- ob_rom_factcheck %>%
  filter(Reuters == "1") %>%
  group_by(Speaker) %>%
  summarize(Reuters_check = n())

APNews <- ob_rom_factcheck %>%
  filter(AP.News == "1") %>%
  group_by(Speaker) %>%
  summarize(AP_check = n())

NYT <- ob_rom_factcheck %>%
  filter(NYT == "1") %>%
  group_by(Speaker) %>%
  summarize(NYT_check = n())

Atl <- ob_rom_factcheck %>%
  filter(Atlantic == "1") %>%
  group_by(Speaker) %>%
  summarize(Atl_check = n())

PF <- ob_rom_factcheck %>%
  filter(Politifact == "1") %>%
  group_by(Speaker) %>%
  summarize(PF_check = n())

WP <- ob_rom_factcheck %>%
  filter(WashPost == "1") %>%
  group_by(Speaker) %>%
  summarize(WP_check = n())

FC <- ob_rom_factcheck %>%
  filter(Fact.Check == "1") %>%
  group_by(Speaker) %>%
  summarize(FC_check = n())
```

```

checkers = left_join(CNN,
                     left_join(Reuters,
                               left_join(APNews,
                                         left_join(NYT,
                                                  left_join(Atl,
                                                            left_join(PF, left_join(WP, FC, by = "Speaker",

```

```

## Joining, by = "Speaker"
## Joining, by = "Speaker"
## Joining, by = "Speaker"
## Joining, by = "Speaker"
## Joining, by = "Speaker"
## Joining, by = "Speaker"

```

```

checkers %>%
  kable()

```

Speaker	CNN_check	Reuters_check	AP_check	NYT_check	Atl_check	PF_check	WP_check	FC_check
Obama	6	6	13	22	7	8	13	11
Romney	12	5	15	20	6	8	24	20

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)