Pence Kaine EDA

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

```
checkdata = read.csv("data/checkdata.csv")
pencekaine = read.csv("data/pence_kaine_2016.csv")
```

```
checked_claims = checkdata %>%
  filter(Checked. == "1") %>%
  arrange(desc(Ratings))

checked_claims %>%
  group_by(Claimant) %>%
  summarize(avgrating = mean(Ratings))
```

```
## # A tibble: 2 x 2
##   Claimant   avgrating
##   <chr>          <dbl>
## 1 Mike Pence     0.423
## 2 Tim Kaine      0.383
```

Lowest claimbuster score of a claim that was checked: 0.12

The highest claimbuster score of a claim that was checked: 0.71

Total claims checked: 112

Average claimbuster score Mike Pence: *~.42*

Average claimbuster score Tim Kaine: *~.38*

On average Tim Kaine's claims are slightly less checkable (0.04 / 4%).

```
checkdata %>%
  group_by(Claimant) %>%
  count()
```

```
## # A tibble: 3 x 2
## # Groups:   Claimant [3]
##   Claimant         n
##   <chr>        <int>
## 1 ""               2
## 2 "Mike Pence"   542
## 3 "Tim Kaine"    582
```

```
checked_claims %>%
  group_by(Claimant) %>%
  count()
```

```
## # A tibble: 2 x 2
## # Groups:   Claimant [2]
##   Claimant       n
##   <chr>      <int>
## 1 Mike Pence    53
## 2 Tim Kaine     59
```

Mike Pence had a total of 542 factual claims. 53 of his factual claims were checked *~9.8%*

Tim Kaine had a total of 582 factual claims. 59 of his factual claims were checked *~10%*

No major bias by fact checkers according to candidate Note: may be worth running a t-test on this to confirm no statistical significance.

```
checked_claims %>%
  group_by(times.fact.checked) %>%
  count(Claimant)
```

```
## # A tibble: 10 x 3
## # Groups:   times.fact.checked [5]
##    times.fact.checked Claimant       n
##                 <int> <chr>      <int>
##  1                  1 Mike Pence    34
##  2                  1 Tim Kaine     35
##  3                  2 Mike Pence    11
##  4                  2 Tim Kaine     16
##  5                  3 Mike Pence     6
##  6                  3 Tim Kaine      6
##  7                  4 Mike Pence     1
##  8                  4 Tim Kaine      1
##  9                  6 Mike Pence     1
## 10                  6 Tim Kaine      1
```

As expected (knowing that each claimant was fact-checked a roughly equal amount of times) both claimants have similar amounts of claims checked repeatedly.

The only *times.fact.checked* category that has different numbers of claims checked for each Claimant is *2* where Mike Pence had *11* checked claims and Tim Kaine has *16* checked claims.

```
PF = checked_claims %>%
  filter(PF == "1") %>%
  group_by(Claimant) %>%
  summarize(PFcheck = n())

WP = checked_claims %>%
  filter(WP == "1") %>%
  group_by(Claimant) %>%
  summarize(WPcheck = n())
```

```
FC = checked_claims %>%
  filter(FC == "1") %>%
  group_by(Claimant) %>%
  summarize(FCcheck = n())

NYT = checked_claims %>%
  filter(NYT == "1") %>%
  group_by(Claimant) %>%
  summarize(NYTcheck = n())

CNN = checked_claims %>%
  filter(CNN == "1") %>%
  group_by(Claimant) %>%
  summarize(CNNcheck = n())

AP = checked_claims %>%
  filter(AP == "1") %>%
  group_by(Claimant) %>%
  summarize(APcheck = n())

TG = checked_claims %>%
  filter(TG == "1") %>%
  group_by(Claimant) %>%
  summarize(TGcheck = n())

checkers = left_join(PF, left_join(WP
                                   , left_join(FC
                                               , left_join(NYT
                                                           , left_join(CNN
                                                                       , left_join(AP, TG,
                                                                                   by = "Claimant"))))))
```

```
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
```

```
checkers
```

```
## # A tibble: 2 x 8
##   Claimant   PFcheck WPcheck FCcheck NYTcheck CNNcheck APcheck TGcheck
##   <chr>        <int>   <int>   <int>    <int>    <int>   <int>   <int>
## 1 Mike Pence      13      13       4       12       16      14      12
## 2 Tim Kaine       23      26       7        7       14       6      12
```

1/7 checker checked both candidates equally (TG)

3/7 checkers checked Pence more than Kaine (NYT, CNN, AP)

3/7 checkers checked Kaine more than Pence (PF, WP, FC)

No apparent partisan bias by news sources for one candidate over the other

```
`0.2checked` = checked_claims %>%
  filter(Ratings >= 0.2)

`0.2data` = checkdata %>%
  filter(Ratings >= 0.2)

`0.2checked` %>%
  group_by(Claimant) %>%
  summarize(avgrating = mean(Ratings))
```

```
## # A tibble: 2 x 2
##   Claimant   avgrating
##   <chr>          <dbl>
## 1 Mike Pence     0.440
## 2 Tim Kaine      0.39
```

```
`0.2data` %>%
  group_by(Claimant) %>%
  count()
```

```
## # A tibble: 2 x 2
## # Groups:   Claimant [2]
##   Claimant       n
##   <chr>      <int>
## 1 Mike Pence   358
## 2 Tim Kaine    402
```

```
`0.2checked` %>%
  group_by(Claimant) %>%
  count()
```

```
## # A tibble: 2 x 2
## # Groups:   Claimant [2]
##   Claimant       n
##   <chr>      <int>
## 1 Mike Pence    50
## 2 Tim Kaine     57
```

Average claimbuster score Mike Pence: *~.44*

Average claimbuster score Tim Kaine: *~.39*

On average Tim Kaine's claims are slightly less checkable (0.05 / 5%).

Mike Pence had a total of 358 factual claims. 50 of his factual claims were checked *~14%*

Tim Kaine had a total of 402 factual claims. 57 of his factual claims were checked *~14%*

No major bias by fact checkers according to candidate Note: may be worth running a t-test on this to confirm no statistical significance.

```
PF.2 = `0.2checked` %>%
  filter(PF == "1") %>%
  group_by(Claimant) %>%
```

```
  summarize(PFcheck = n())

WP.2 = `0.2checked` %>%
  filter(WP == "1") %>%
  group_by(Claimant) %>%
  summarize(WPcheck = n())

FC.2 = `0.2checked` %>%
  filter(FC == "1") %>%
  group_by(Claimant) %>%
  summarize(FCcheck = n())

NYT.2 = `0.2checked` %>%
  filter(NYT == "1") %>%
  group_by(Claimant) %>%
  summarize(NYTcheck = n())

CNN.2 = `0.2checked` %>%
  filter(CNN == "1") %>%
  group_by(Claimant) %>%
  summarize(CNNcheck = n())

AP.2 = `0.2checked` %>%
  filter(AP == "1") %>%
  group_by(Claimant) %>%
  summarize(APcheck = n())

TG.2 = `0.2checked` %>%
  filter(TG == "1") %>%
  group_by(Claimant) %>%
  summarize(TGcheck = n())

checkers.2 = left_join(PF.2
                       , left_join(WP.2
                                   , left_join(FC.2
                                               , left_join(NYT.2
                                                           , left_join(CNN.2
                                                                       , left_join(AP.2, TG.2,
                                                                                   by = "Claimant"))))))
```

```
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
## Joining, by = "Claimant"
```

```
checkers.2
```

```
## # A tibble: 2 x 8
##   Claimant   PFcheck WPcheck FCcheck NYTcheck CNNcheck APcheck TGcheck
##   <chr>        <int>   <int>   <int>    <int>    <int>   <int>   <int>
## 1 Mike Pence      13      12       4       12       15      13      11
## 2 Tim Kaine       21      25       7        7       13       6      12
```

3/7 checkers checked Pence more than Kaine (NYT, CNN, AP)

4/7 checkers checked Kaine more than Pence (PF, WP, FC, TG)

No apparent partisan bias by news sources for one candidate over the other