# Washington Post

Sofia Bliss-Carrascosa

10/05/2022

We began with a dataset of Washington Post fact checks from Claim Review. We began with 554 observations and in the process of cleaning our data, described below, we ended with 530.

In OpenRefine, we created a new column called 'claimant_clean' and manually sorted the names of all claimants to determine when there were different spellings of the same name. This was done via the cluster function embedded in the application. When iterations were found, we selected a single name to refer to each person by. For the claimant column, this cleaning process resulted in a reduction of claimants from 188 individual names to 169.

The Washington Post is not entirely consistent in its determinations of claim ratings. In theory, it has a standard ranking system ordered as follows: Geppetto Checkmark, One Pinocchio, Two Pinocchios, Three Pinocchios, Four Pinocchios. However, in our data, there were 335 are normal claims (those rated on the WaPo scale of 5 standardized ratings) and 173 anomaly claims with a non-standardized text based rating. Of those 219, 189 were labeled with an anomaly claim repeated more than once. To accord these common anomalies with the standard rating system, we took the data to Glenn Kessler, the singular WaPo fact-checker and the individual most familiar with the scale. At this stage we sent the more common ratings (occasionally grouping similar ones together) and asking him where they fall on the ratings scale. Claims marked: Wrong, Incorrect, False –> Four Pinocchios Mostly false, Spins or Twists facts, Exaggerated –> Three Pinocchios Half true, Not the whole story, Depends on the math, Cherry-picked number –> Two Pinocchios True, Correct, Accurate –> Geppetto Checkmark

For the more unclear categories, we sent 61 claims for individual review. 57 of these were manually reassigned a textual rating according to his instructions, 4 were removed. For the purpose of concision, we removed claims that were tagged with uncommon anomaly ratings (1-2 observations). We began this process with 60 textual ratings, having the goal of eventually reducing down to the 5 rating ranking system.

```
## # A tibble: 28 x 2
## # Groups:   textualRating [28]
##    textualRating          n
##    <chr>              <int>
##  1 Four Pinocchios      193
##  2 Two Pinocchios       170
##  3 Three Pinocchios     130
##  4 One Pinocchio         19
##  5 Geppetto Checkmark    11
##  6 Needs context          4
##  7 One Pinocchios         3
##  8 Doubtful               2
##  9 Flip flop              2
## 10 Fuzzy math             2
## # ... with 18 more rows
```

Table 1: Claim Counts by Party

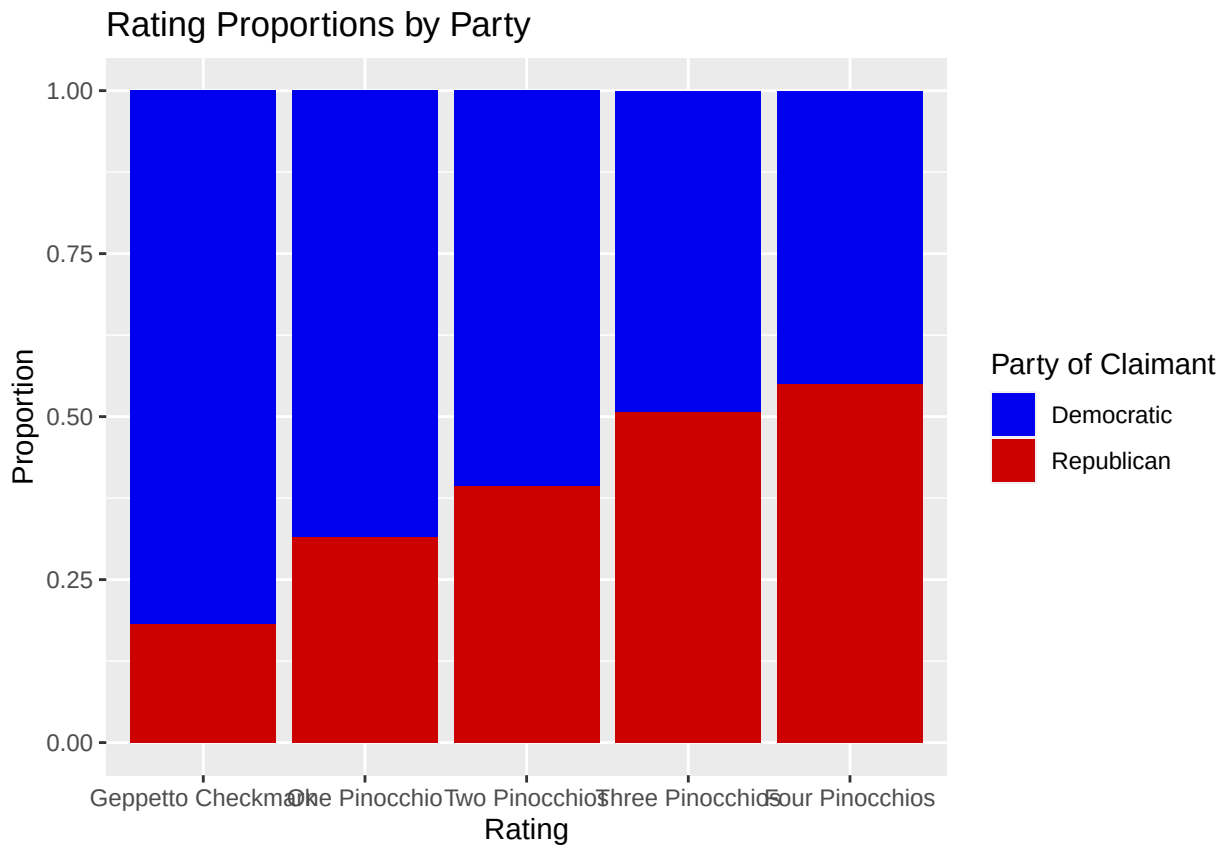
| claimant_party | n |
|---|---|
| Democratic | 276 |
| Republican | 247 |

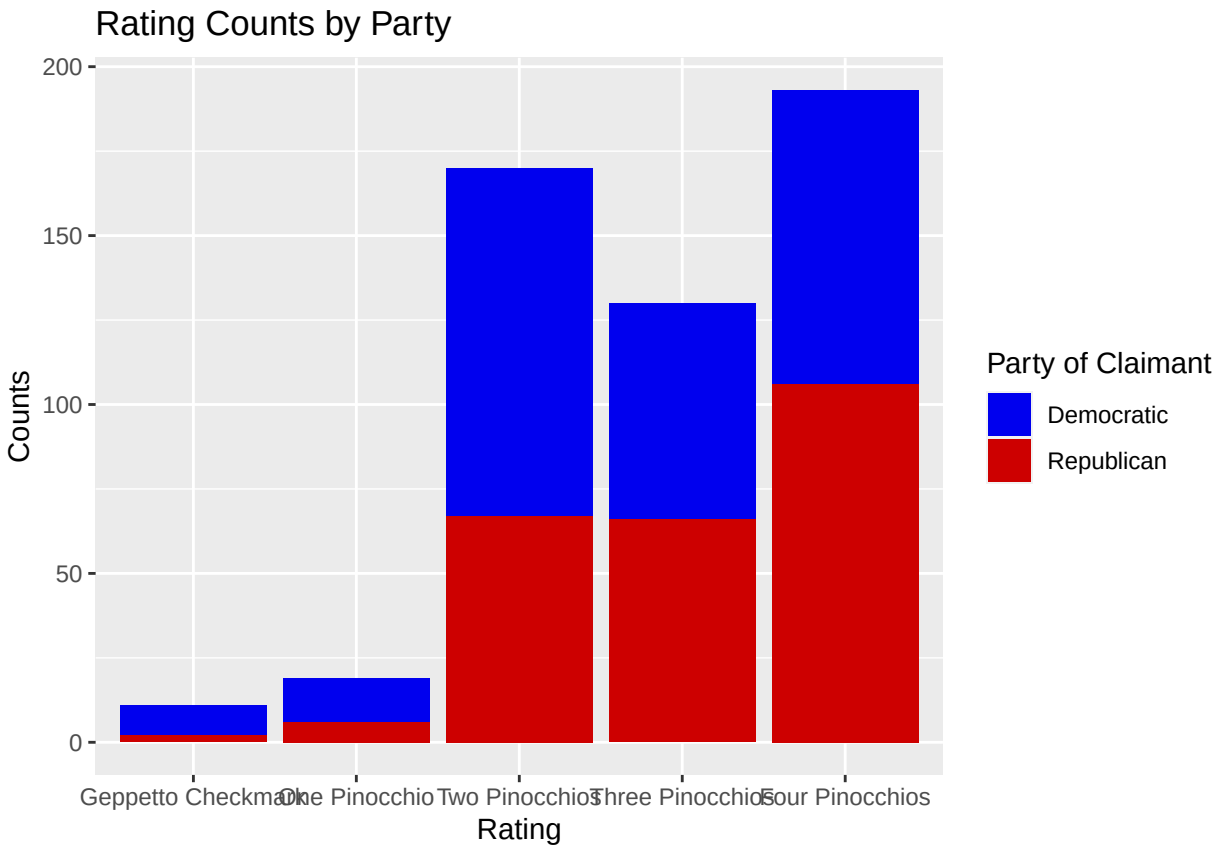
Table 2: Claim Counts by Rating

| textualRating | n |
|---|---|
| Geppetto Checkmark | 11 |
| One Pinocchio | 19 |
| Two Pinocchios | 170 |
| Three Pinocchios | 130 |
| Four Pinocchios | 193 |

Table 3: Claim Counts by Party and Rating

| claimant_party | Geppetto Checkmark | One Pinocchio | Two Pinocchios | Three Pinocchios | Four Pinocchios |
|---|---|---|---|---|---|
| Democratic | 9 | 13 | 103 | 64 | 87 |
| Republican | 2 | 6 | 67 | 66 | 106 |




Rating Counts by Party
200
150
100
50
0
Counts
Geppetto Checkmark
One Pinocchio
Two Pinocchios
Three Pinocchios
Four Pinocchios
Rating
Party of Claimant
Democratic
Republican