

Washington Post

Sofia Bliss-Carrascosa

10/05/2022

After the initial data cleaning described in our Data Cleaning document, there were 554 fact-checks published by the Washington Post between January 1st 2016 and June 30th, 2021. The dataset only includes fact-checks in which the claimant was a political figure, not Donald Trump, and assigned a Democratic or Republican affiliation. This dataset is called Version 8. We then did some light data cleaning prior to our analysis to clarify claimant names and streamline aberrant textual ratings. That process is described below.

```
v8_wapo <- read.csv("v8wapo.csv")
```

Claimant Name Cleaning + Initial Rating Clean Up V9

In a program called OpenRefine, we created a new column called ‘claimant_clean’ and manually sorted the names of all claimants in our dataset to determine when there were different spellings of the same name. This was done via the cluster function embedded in the application. When varying iterations were found of the same name, we selected a single name to refer to each person by. For example Charles Schumer was changed to Chuck Schumer. For the claimant column, this cleaning process resulted in a reduction of claimants from 188 individual names to 169. This did not reduce the number of claims, just streamlined claimant names so the same political figure was not listed under several different names. The original claimant names are stored under the variable “claimant”

Next we had to address the issue in textualRating. 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.

ADD FORMAL DEF

```
v8_wapo %>%  
  group_by(textualRating) %>%  
  count() %>%  
  arrange(desc(n)) %>%  
  print(n=65)
```

```
## # A tibble: 60 x 2  
## # Groups:   textualRating [60]  
##   textualRating      n  
##   <chr>          <int>  
## 1 Four Pinocchios    136  
## 2 Three Pinocchios   100  
## 3 Two Pinocchios     89  
## 4 FALSE              46  
## 5 Needs context     21  
## 6 Not the whole story 19
```

## 7 Not the Whole Story	19
## 8 Lacks context	12
## 9 In dispute	7
## 10 Wrong	7
## 11 One Pinocchio	6
## 12 Cherry-picked number	5
## 13 Exaggerated	5
## 14 Mostly False	5
## 15 Spins the facts	5
## 16 Geppetto Checkmark	4
## 17 Lacks Context	4
## 18 Misleading	4
## 19 No evidence	4
## 20 Correct	3
## 21 Half true	3
## 22 Mis- leading	3
## 23 Needs Context	3
## 24 TRUE	3
## 25 Doubtful	2
## 26 Fuzzy math	2
## 27 Mostly false	2
## 28 Not the whole story.	2
## 29 Twists the facts	2
## 30 Accurate	1
## 31 Data is fishy	1
## 32 Depends on math	1
## 33 Depends on stats used	1
## 34 Disputed	1
## 35 Evidence gets stronger	1
## 36 Evidence is mixed	1
## 37 Facile comparison	1
## 38 Flip flop	1
## 39 Flip- Flop	1
## 40 Flubbed Talking Point	1
## 41 For the Record	1
## 42 Half Flop	1
## 43 Half True	1
## 44 Incorrect	1
## 45 Issue is more complex	1
## 46 It's a complex story	1
## 47 Lacks evidence	1
## 48 Mis- Spoke	1
## 49 Needs more context	1
## 50 Not based on science	1
## 51 Not clear yet	1
## 52 Not quite right	1
## 53 Not the same	1
## 54 Quoted out of context	1
## 55 Spin	1
## 56 Spinning the facts	1
## 57 Staff error	1
## 58 Study in Dispute	1
## 59 Techni- cally right	1
## 60 Verdict Pending	1

However, in version 8 of our data, there were 335 normal claims (those rated on the WaPo scale of 5 standardized ratings) and 219 anomaly claims with a non-standardized text based rating.

Of those 219, 188 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: 32 x 2
##   textualRating      count
##   <chr>            <int>
## 1 Data is fishy          1
## 2 Disputed              8
## 3 Doubtful              2
## 4 Evidence gets stronger 1
## 5 Evidence is mixed     1
## 6 Facile comparison     1
## 7 Flip flop             2
## 8 Flubbed Talking Point 1
## 9 For the Record        1
## 10 Four Pinocchios      190
## 11 Fuzzy math           2
## 12 Geppetto Checkmark   11
## 13 Half Flop            1
## 14 Issue is more complex 1
## 15 It's a complex story 1
## 16 Lacks evidence       1
## 17 Misleading           7
## 18 Misspoke             1
## 19 Needs context        41
## 20 No evidence          4
## 21 Not based on science 1
## 22 Not clear yet        1
## 23 Not quite right      1
## 24 Not the same         1
## 25 One Pinocchio        6
## 26 Quoted out of context 1
## 27 Staff error          1
## 28 Study in Dispute     1
## 29 Technically right    1
## 30 Three Pinocchios     121
## 31 Two Pinocchios       140
## 32 Verdict Pending      1

##           claimant_clean
## 1           Hillary Clinton
## 2           Reince Priebus
```

## 3	Cory Booker
## 4	Steve Scalise
## 5	Chuck Schumer
## 6	Paul Ryan
## 7	Maxine Waters
## 8	Stephen Miller
## 9	Bill Akins
## 10	Sean Spicer
## 11	Tammy Baldwin
## 12	Marsha Blackburn
## 13	Claire McCaskill
## 14	Nancy Pelosi
## 15	Tom Perez
## 16	Jeff Sessions
## 17	Mick Mulvaney
## 18	Kevin Brady
## 19	Chris Murphy
## 20	Harry Reid
## 21	Kirsten Gillibrand
## 22	John Conyers; Jr.
## 23	Mike Pence
## 24	Jon Ossoff
## 25	Scott Pruitt
## 26	Susan Rice
## 27	Barack Obama
## 28	Valerie Jarrett
## 29	Markwayne Mullin
## 30	Ben Sasse
## 31	Ted Yoho
## 32	Loretta Lynch
## 33	Amy Klobuchar
## 34	Joyce Beatty
## 35	Bob Goodlatte
## 36	Kamala Harris
## 37	Ivanka Trump
## 38	Bernie Sanders
## 39	Tom Price
## 40	Charmaine Yoest
## 41	Kris Kobach
## 42	Mitch McConnell
## 43	Sarah Huckabee Sanders
## 44	Robert E. Lighthizer
## 45	Ted Cruz
## 46	Joe Arpaio
## 47	Charles E. Grassley
## 48	Jeff Merkley
## 49	Bill Cassidy
## 50	Lindsey Graham
## 51	Rick Santorum
## 52	Gary Cohn
## 53	Steve Mnuchin
## 54	Tim Kaine
## 55	James Lankford
## 56	Marco Rubio

## 57	Ryan Zinke
## 58	H.R. McMaster
## 59	Josh Earnest
## 60	Kellyanne Conway
## 61	John Kerry
## 62	Dianne Feinstein
## 63	Ed Gillespie
## 64	Devin Nunes
## 65	Tucker Carlson
## 66	Sebastian Gorka
## 67	Bob Casey
## 68	John Cornyn
## 69	Tom MacArthur
## 70	Jeb Hensarling
## 71	Chris Christie
## 72	Joseph Kennedy III
## 73	Kevin McCarthy
## 74	Raul Labrador
## 75	Tina Smith
## 76	Jim Mattis
## 77	Peter Navarro
## 78	Elaine Chao
## 79	Liz Cheney
## 80	John Boehner
## 81	Larry Kudlow
## 82	Rand Paul
## 83	Steve Cortes
## 84	Kirstjen Nielsen
## 85	Rudy Giuliani
## 86	Bill Clinton
## 87	Andrew Cuomo
## 88	Kevin Cramer
## 89	Elizabeth Warren
## 90	Carolyn B. Maloney
## 91	Neera Tanden
## 92	Phil Bredeesen
## 93	Hope Hicks
## 94	Wilbur Ross
## 95	Alexandria Ocasio-Cortez
## 96	Heidi Heitkamp
## 97	Richard J. Durbin
## 98	Bill Nelson
## 99	Ann Wagner
## 100	Duncan D. Hunter
## 101	Beto O'Rourke
## 102	Patrick Leahy
## 103	Susan Collins
## 104	Tom Cotton
## 105	Dave Brat
## 106	Matt Gaetz
## 107	Sheldon Whitehouse
## 108	Peter Roskam
## 109	Rodney Davis
## 110	Erik Paulsen

## 111	Mike Kelly
## 112	Robert Casey Jr.
## 113	Martha McSally
## 114	David Cicilline
## 115	Ilhan Omar
## 116	Mike Pompeo
## 117	Ari Fleischer
## 118	Betsy DeVos
## 119	John Hickenlooper
## 120	William Barr
## 121	Joe Biden
## 122	Jared Kushner
## 123	Don Beyer
## 124	Tim Ryan
## 125	Julian Castro
## 126	Steny Hoyer
## 127	Katie Hill
## 128	David Perdue
## 129	John Delaney
## 130	Doug Collins
## 131	Bill de Blasio
## 132	Pete Buttigieg
## 133	Mike Bloomberg
## 134	Cory Gardner
## 135	Steve Daines
## 136	Ronna McDaniel
## 137	Jason Smith
## 138	John Barrasso
## 139	Gabriel Sterling
## 140	Roger Wicker
## 141	Roy Blunt
## 142	Jim Banks
## 143	Andrew Wheeler
## 144	Adam Schiff
## 145	Ron Johnson
## 146	Pat Cipollone
## 147	Tulsi Gabbard
## 148	Ron Klain
## 149	Corky Messner
## 150	Amy McGrath
## 151	Richard Grenell
## 152	Tim Scott
## 153	Elise Stefanik
## 154	Rick Scott
## 155	Kristi L. Noem
## 156	Marjorie Taylor Greene
## 157	Jim Jordan
## 158	Greg Abbott
## 159	Jen Psaki
## 160	Vanita Gupta
## 161	Lauren Boebert
## 162	Ed Markey
## 163	Joaquin Castro
## 164	Cindy Axne

## 165	Shelley Moore Capito
## 166	Mo Brooks
## 167	George P. Bush
## 168	Trent Franks
## 169	Kelly Loeffler

Table 1: Claim Counts by Party

claimant_party	n
Democratic	279
Republican	247

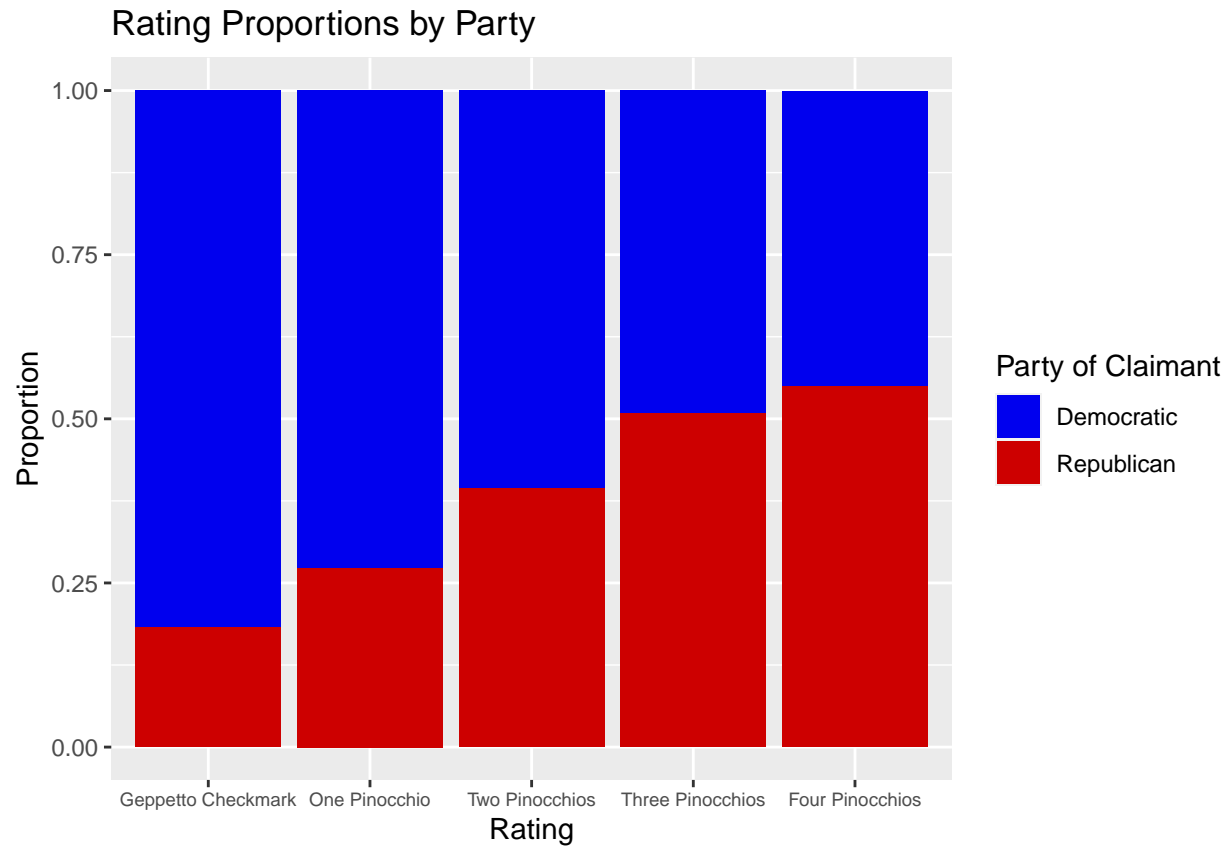
Table 2: Claim Counts by Rating

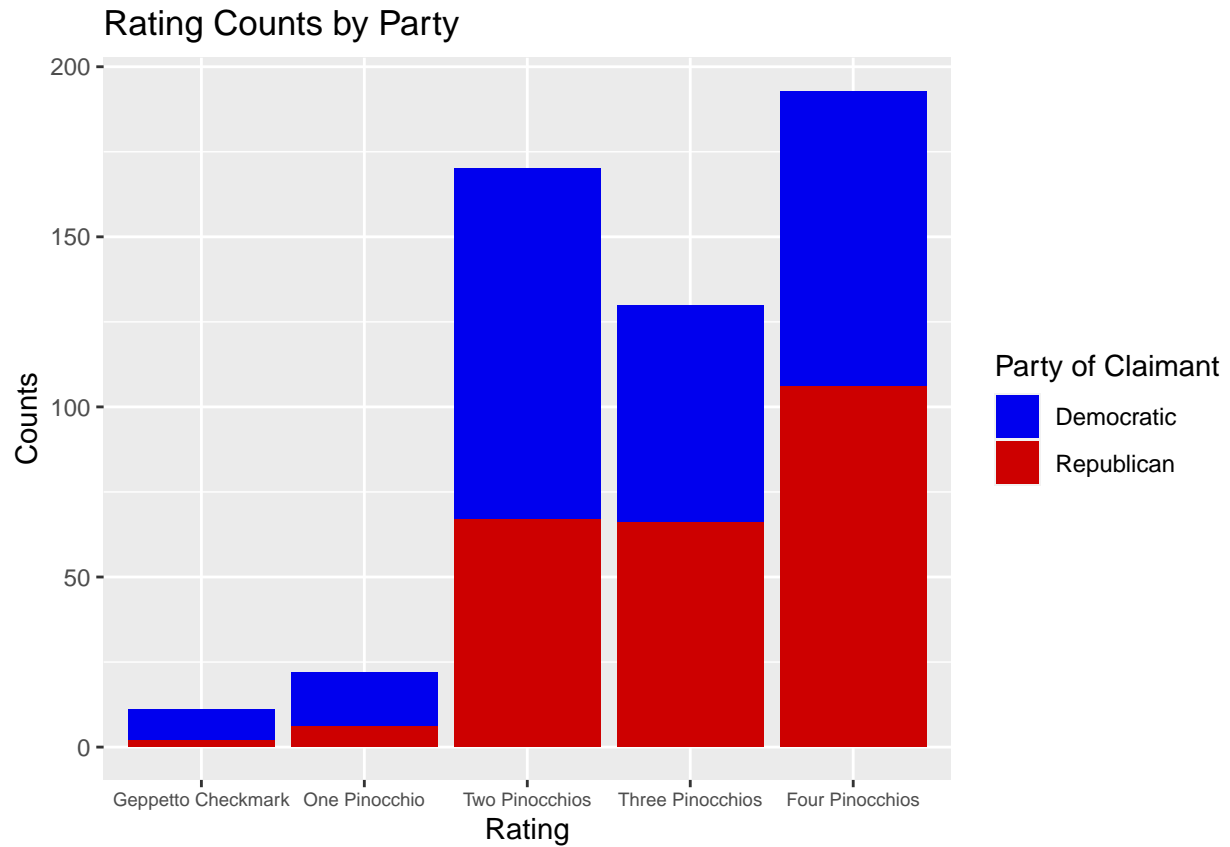
textualRating	n
Geppetto Checkmark	11
One Pinocchio	22
Two Pinocchios	170
Three Pinocchios	130
Four Pinocchios	193

Table 3: Claim Counts by Party and Rating

claimant_party	Geppetto Checkmark	One Pinocchio	Two Pinocchios	Three Pinocchios	Four Pinocchios
Democratic	9	16	103	64	87
Republican	2	6	67	66	106

Likely due to selection bias towards correcting falsehoods and/or an extremely high threshold of proof for truth, there are very few claims rated Geppetto Checkmark or One Pinocchio. As such, its hard to draw any significant conclusions from the limited data. The little information we have does appear to superficially indicate a higher ratio of Democratic truths to their Republican counterparts. There are dozens more datapoints under the Two, Three, and Four Pinocchios rating, both for Democratic and Republican politicians. By count, Democrat falsehoods exceed Republican's in the Two Pinocchios category, but they account for a lower proportion of claims rated Three and Four Pinocchios.





By proportion of total claims alone, and keeping in mind the limitations of any conclusion due to sample size, it appears that Democrats are fact-checked on more truthful claims, whereas Republicans account for a higher share of the egregious falsehoods checked by the Washington Post.

