



REDDIT SHOWDOWN MAGA vs. Resistance

Joseph Darby

1

What's the problem?

Define the problem

Given the text data of a random post from a binary selection of subreddits, can we predict on which subreddit it was posted?

Subreddit Selection Process

- Research topics where classes are hard to differentiate:
 - Socialism vs. Social Democracy? Trickle-Down Economics vs. Capitalism?
- Decided on Political Extremism in the United States:

The_Donald | r/esist



What's the problem?

2

Gather the Data

Data Gathering

Process

- Used requests library to import JSON data
- Indexed into child data to target post titles
- Created for loop to hit the Reddit API for target titles
- Used list comprehensions to isolate and extract post titles and subreddit names
- Saved to CSV



Reddi-Trump

3

Data Exploration

Post Distribution

The_Donald

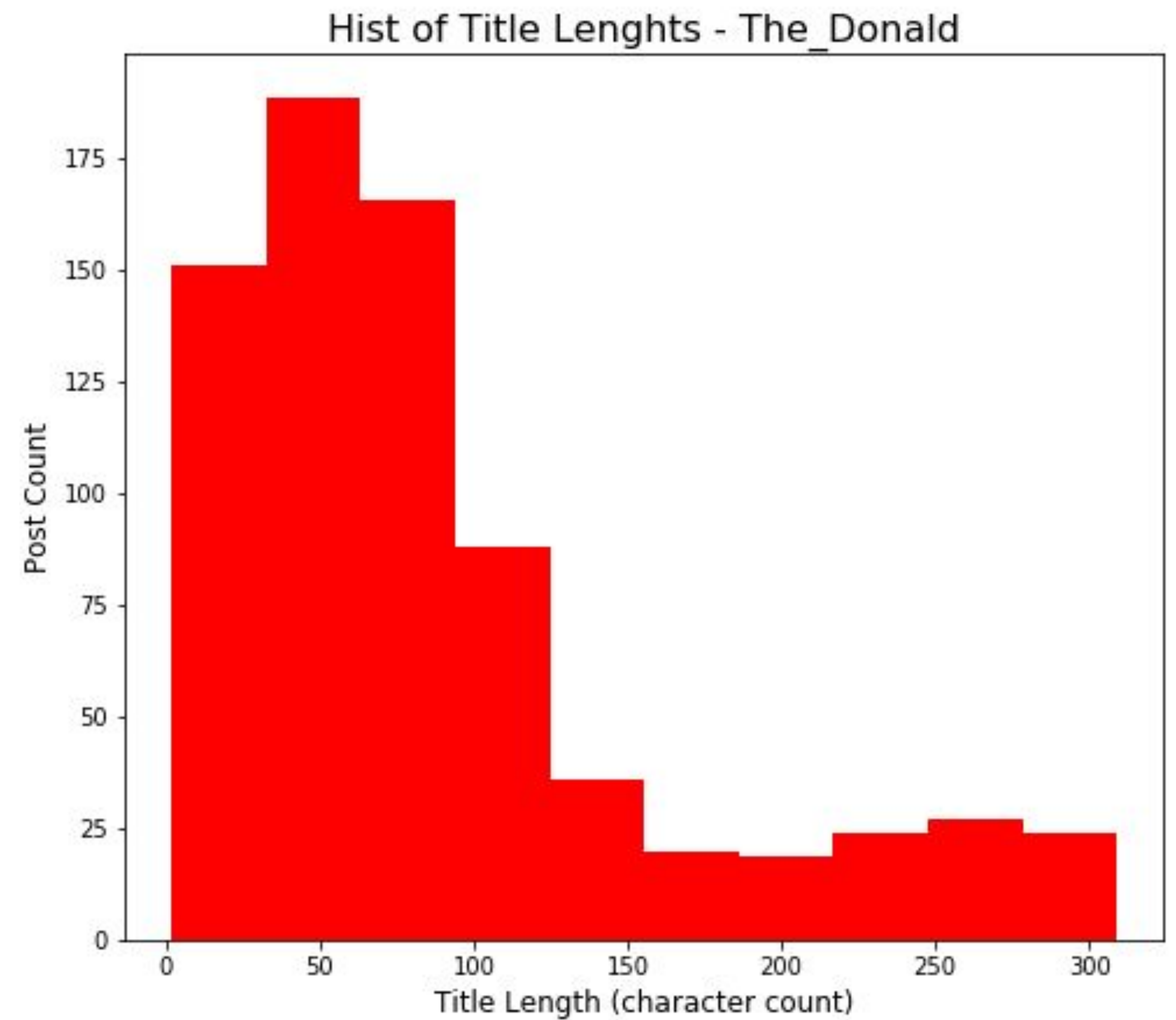
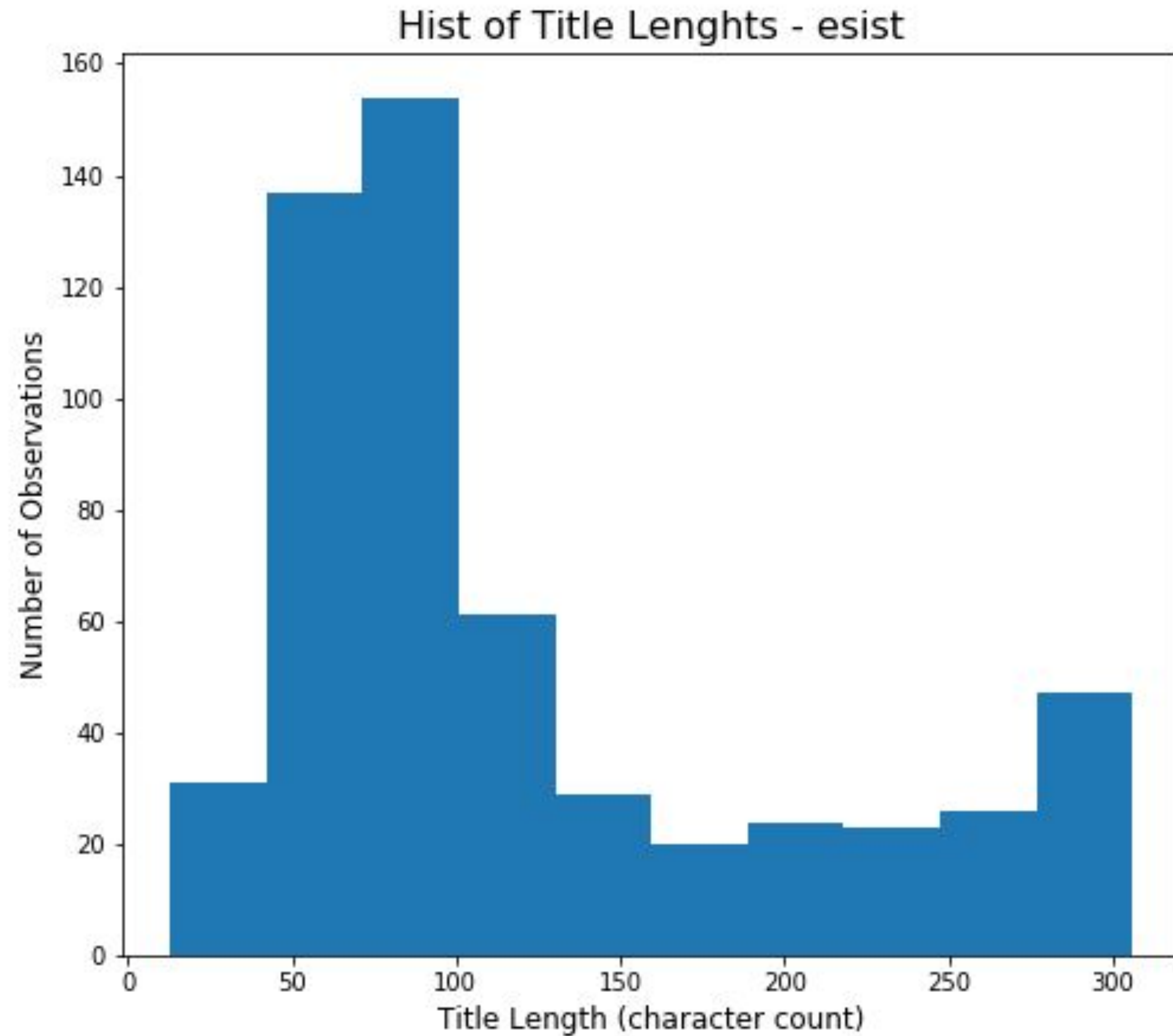
r/esist

997 | 752

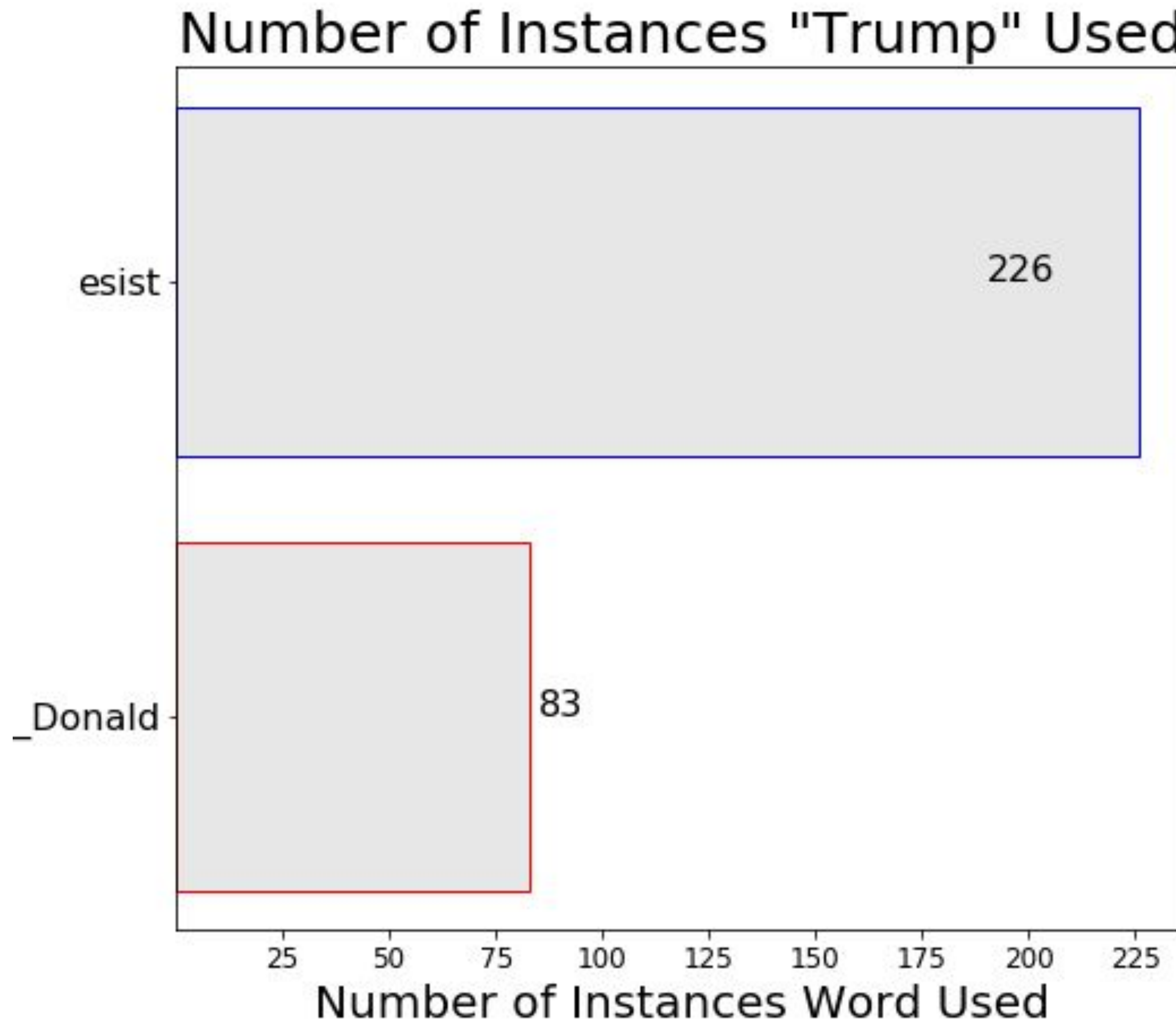
1729 total

56.51% baseline

Title Length Distribution



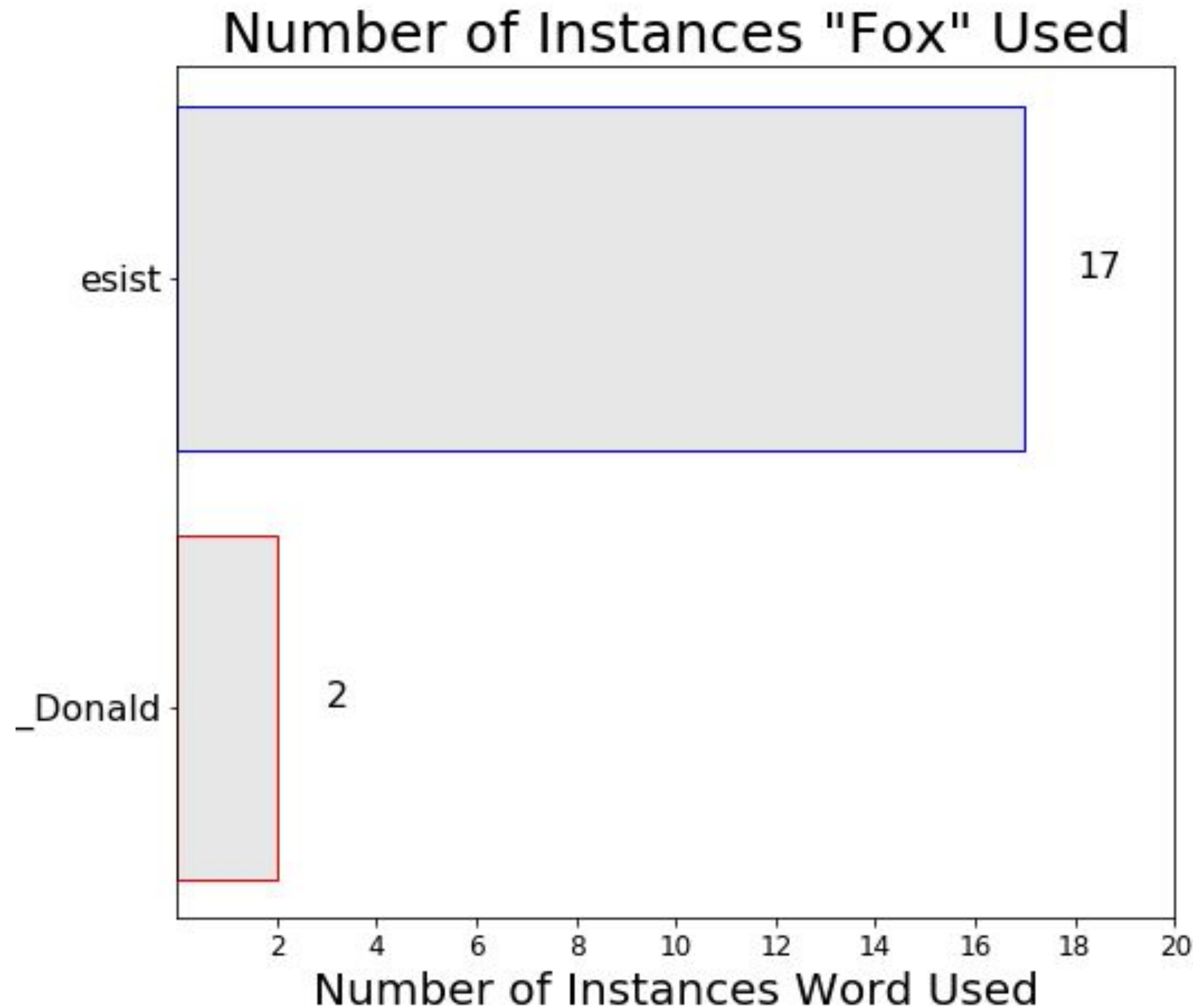
Shared Words



Notes:

- Most used word by both subreddits
- We can infer that it's all about Trump
- The left complains, while the right defends or complements Trump

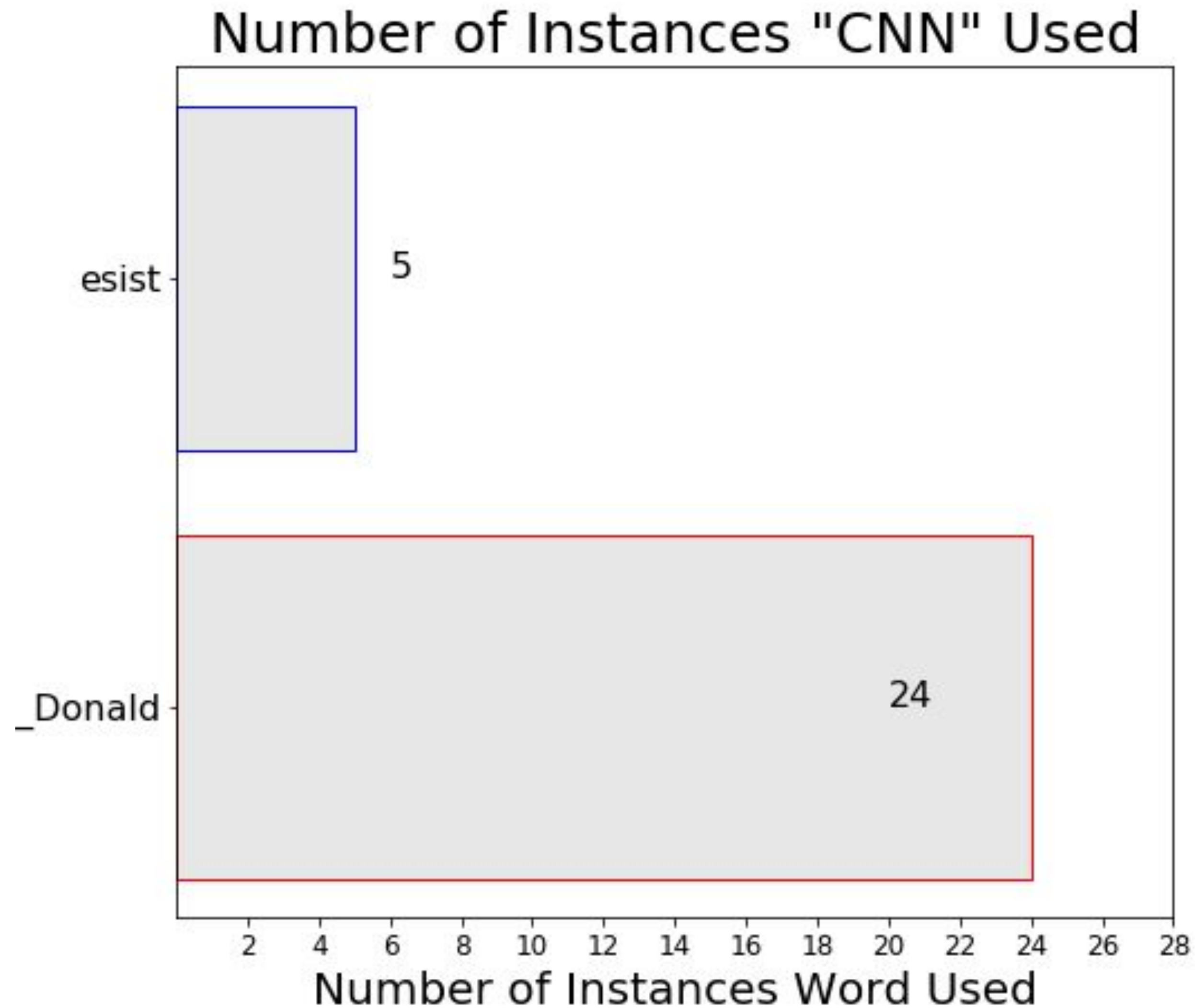
Shared Words



Notes:

- News networks that are associated with one side are generally vilified and attacked by the opposing political party
- Where one gets their news is an indicator of his or her political leaning

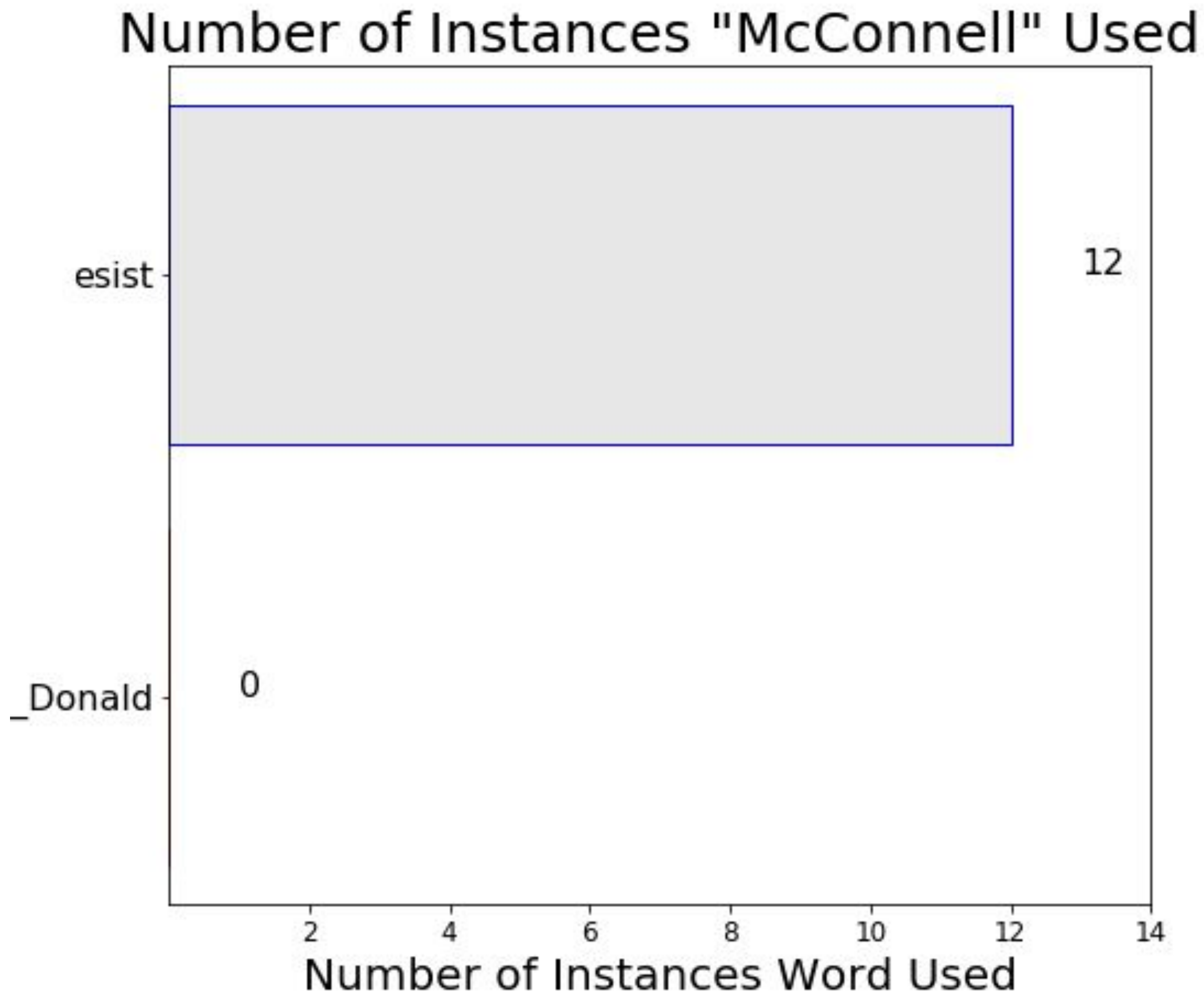
Shared Words



Notes:

- News networks that are associated with one side are generally vilified and attacked by the opposing political party
- Where one gets their news is an indicator of his or her political leaning

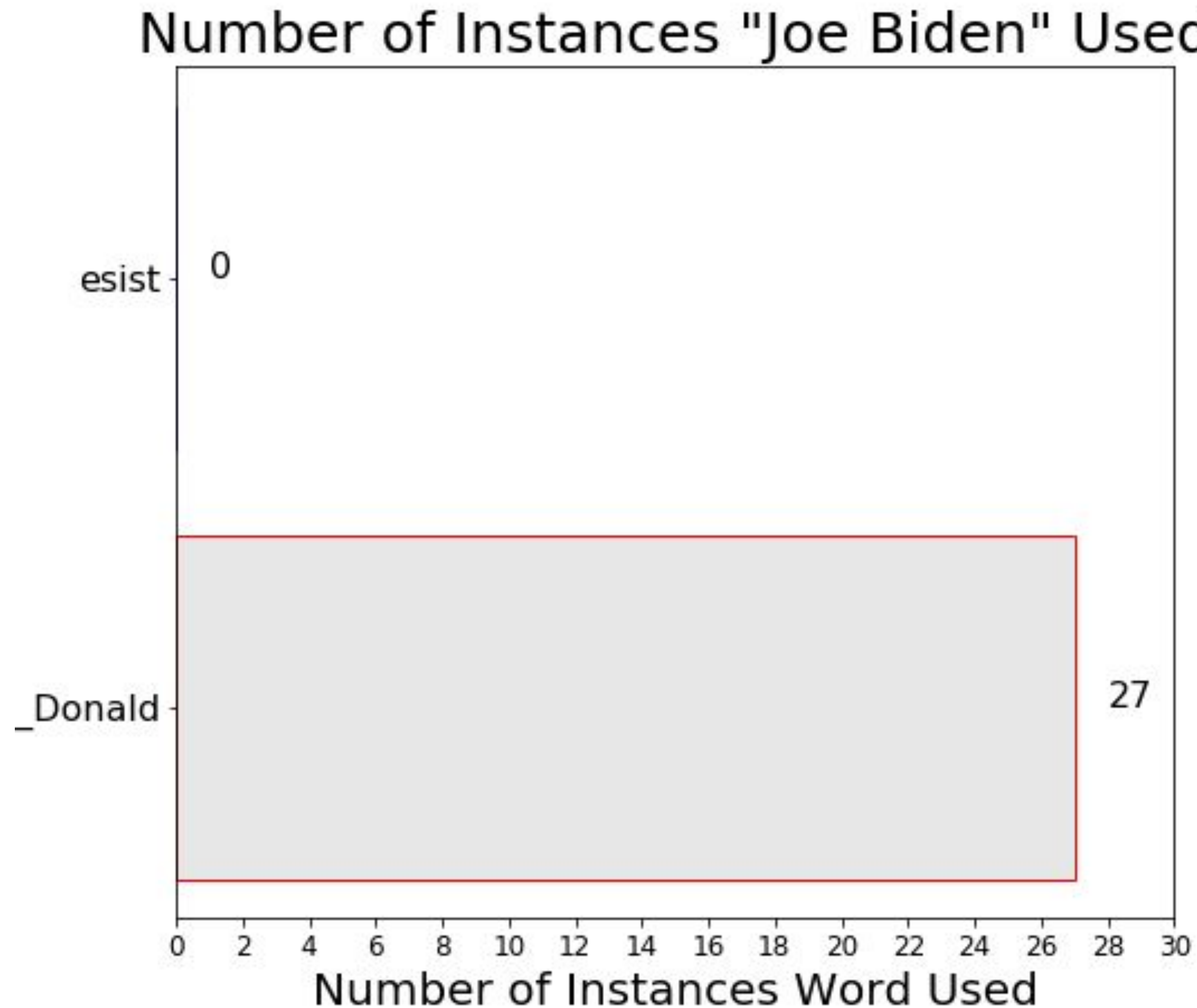
Attack, Attack, Attack!



Notes:

- Each side tends to focus on certain representatives of the opposite view point to consistently vilify

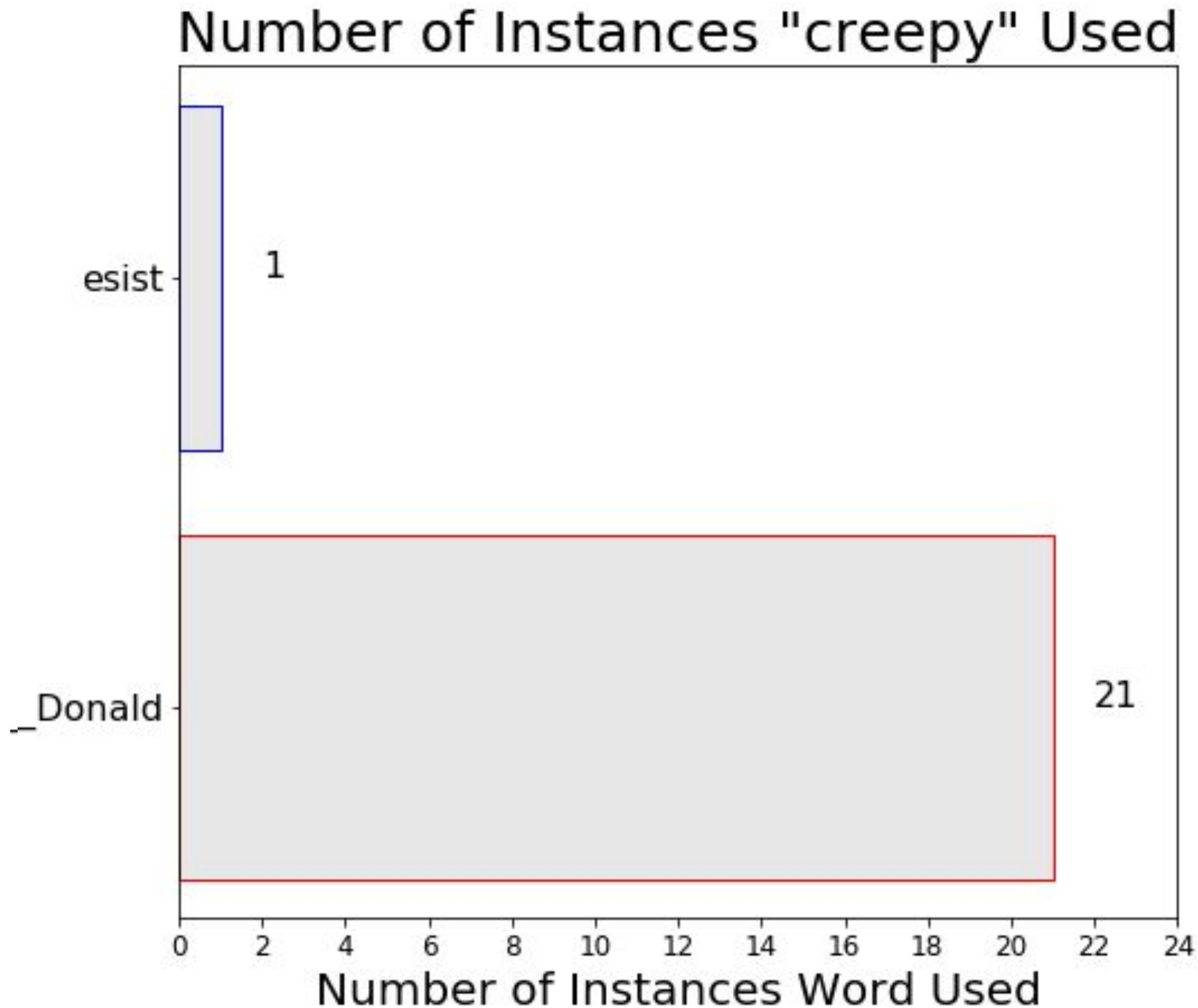
Attack, Attack, Attack!



Notes:

- Each side tends to focus on certain representatives of the opposite view point to consistently vilify

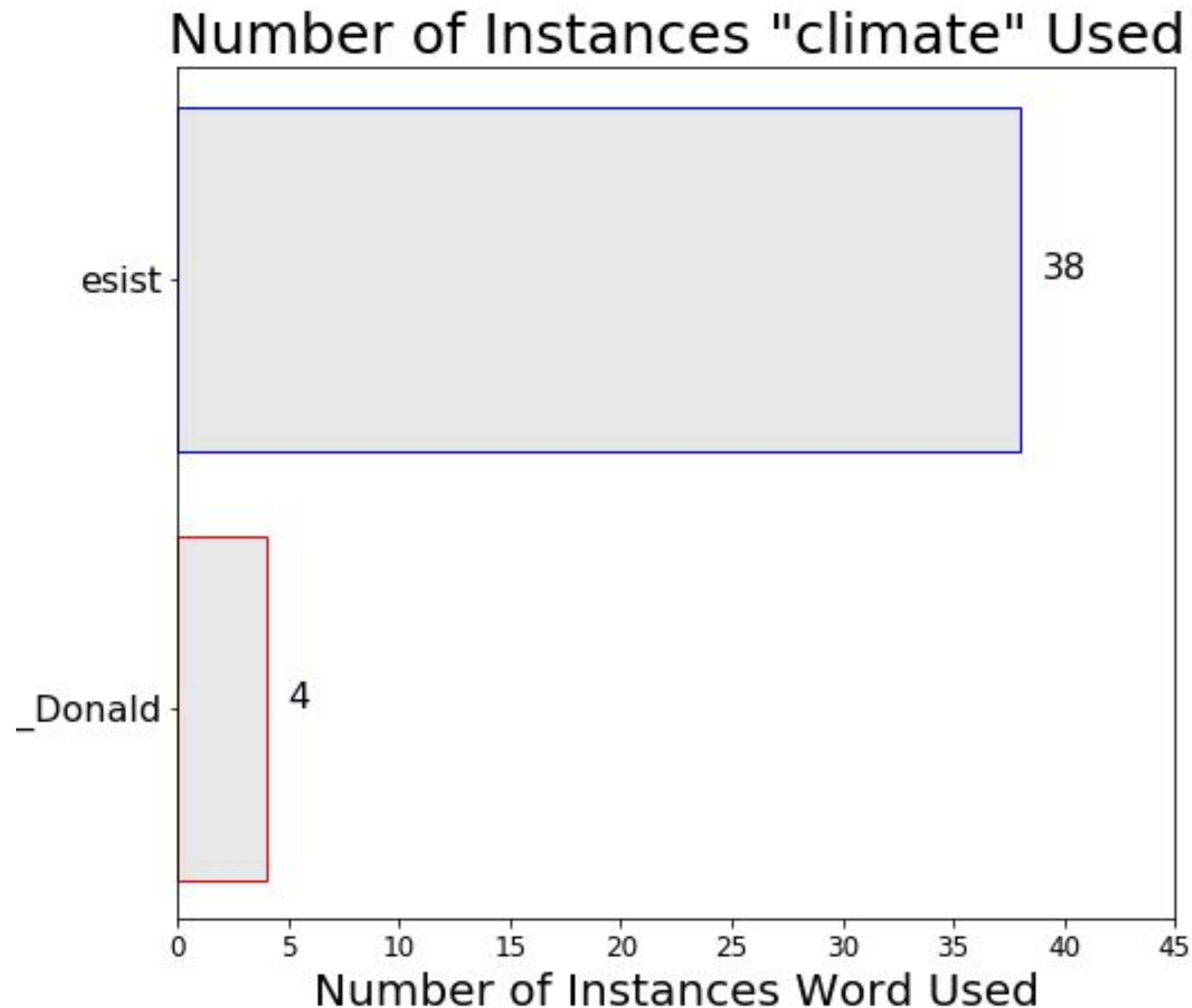
Attack, Attack, Attack!



Notes:

- The right has come up with a new nickname for Joe Biden
- Creepy Uncle Joe

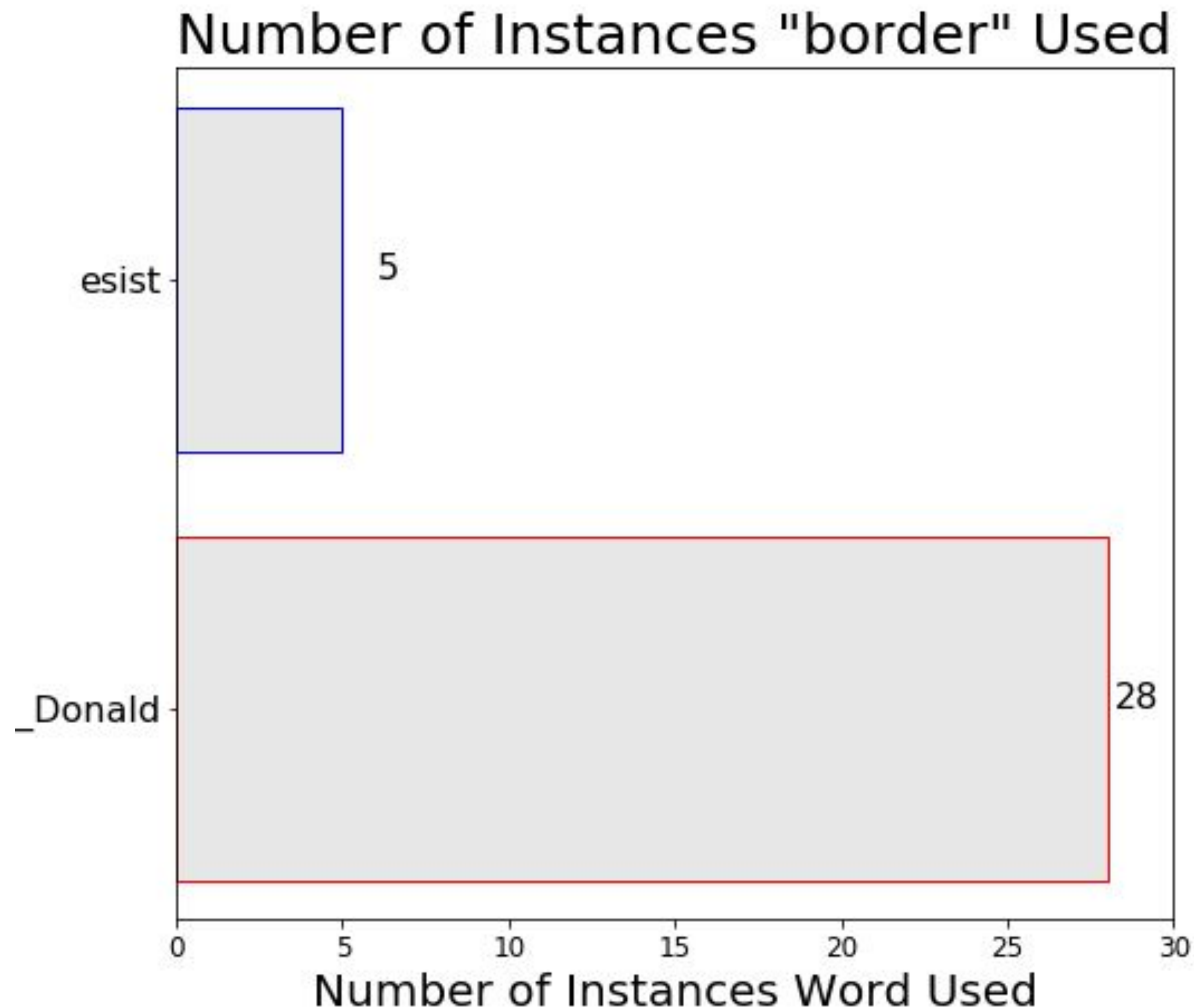
Wedge Issues & Differences



Notes:

- There are particular issues that each side defends dearly and simply cannot agree on
- They are frequently topics of conversation within the subreddits
- “climate change” used by esist: 19 | The_Donald: 3

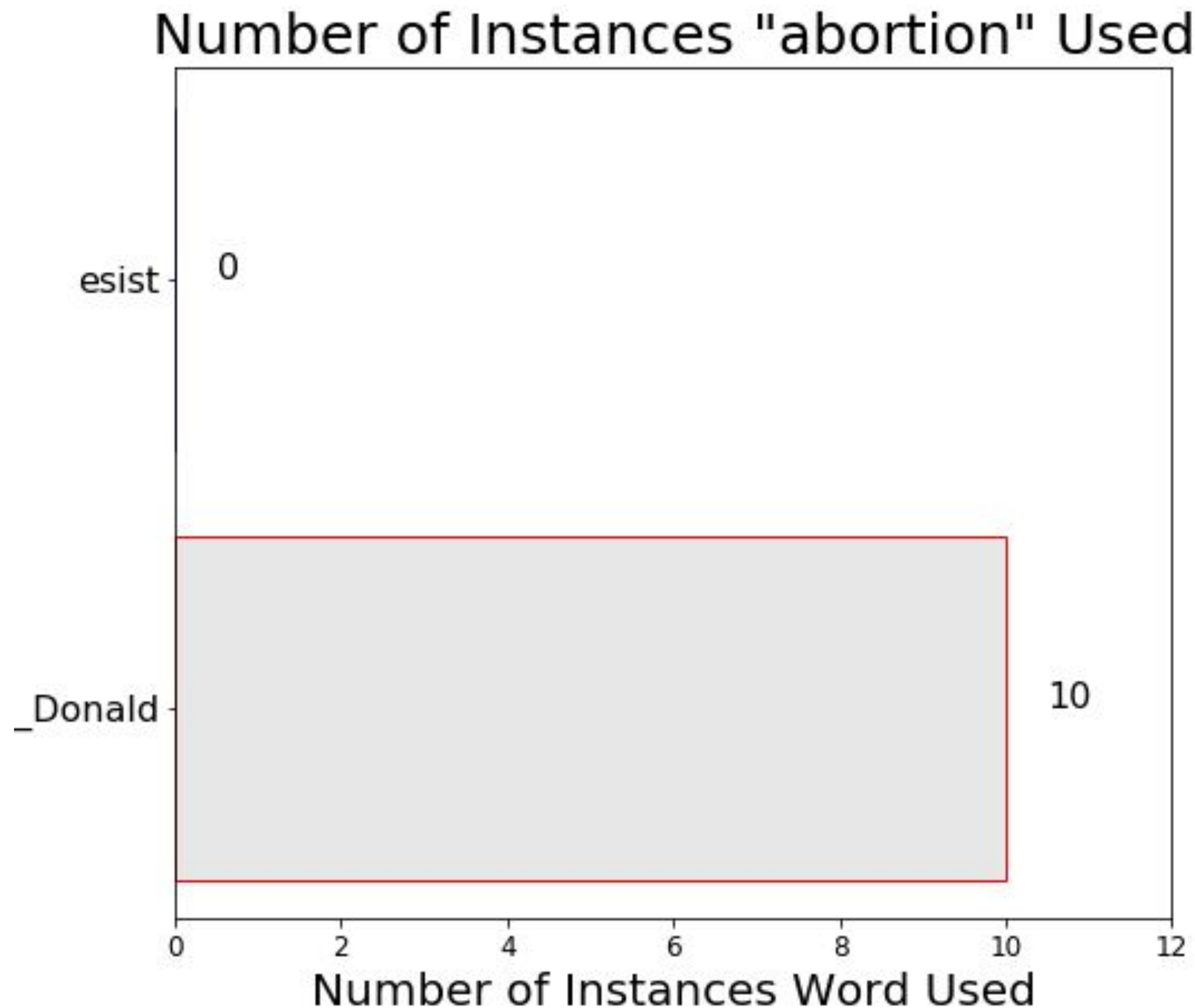
Wedge Issues & Differences



Notes:

- Border Security
- There are particular issues that each side defends dearly and simply cannot agree on
- They are frequently topics of conversation within the subreddits

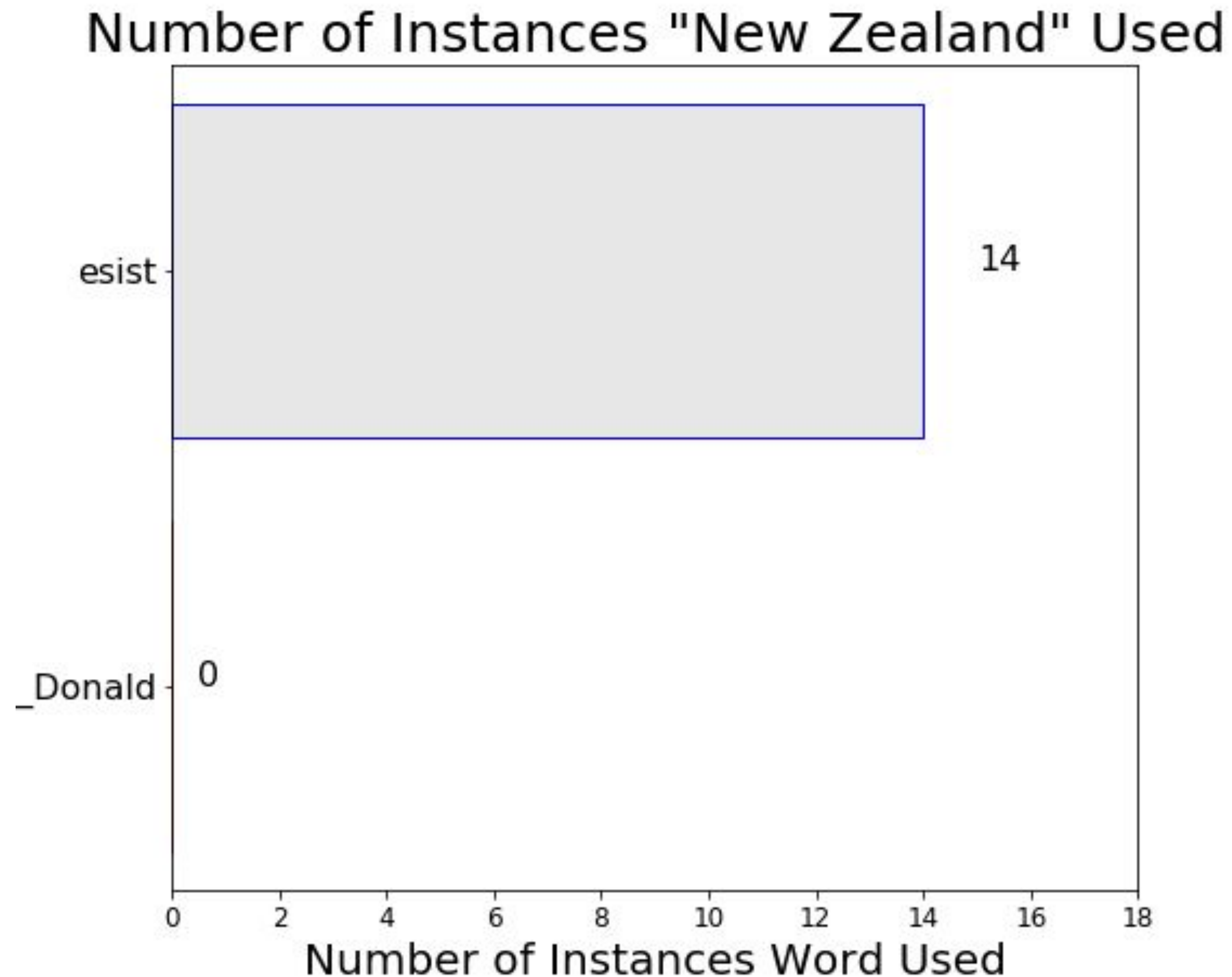
Wedge Issues & Differences



Notes:

- Abortion
- There are particular issues that each side defends dearly and simply cannot agree on
- They are frequently topics of conversation within the subreddits

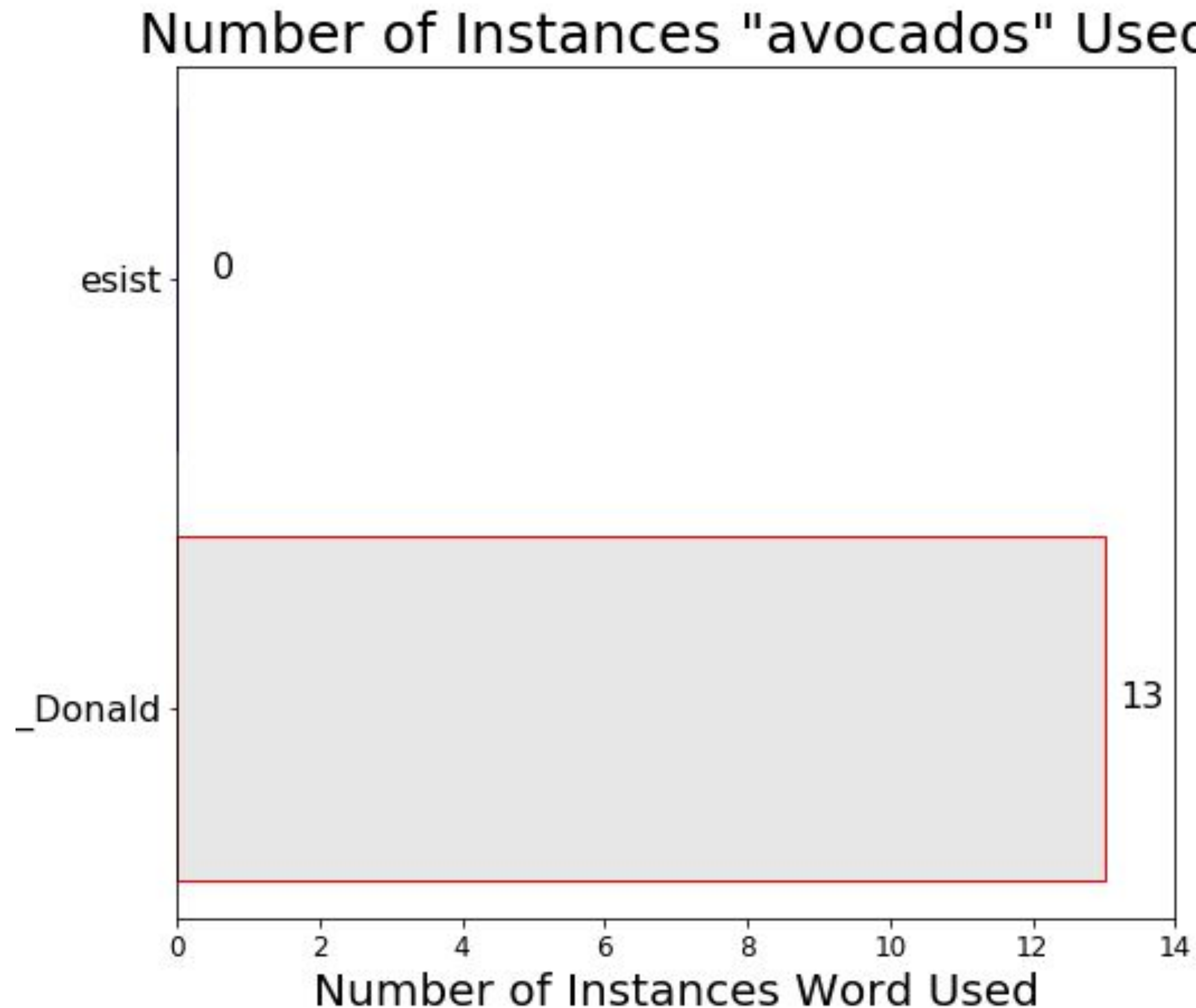
Wedge Issues & Differences



Notes:

- Gun Control
- The left has capitalized on the recent mass shooting in New Zealand to push the topic

Holy Guacamole!



Notes:

- A recent report in USA Today indicated that prices of avocados could rise and supply could dwindle if Trump decides to close the border with Mexico as he has recently tweeted about.

4

Data Modelling

Model Selection & Process

Logistic Regression

X & y set up & TTS

Create Pipeline

CountVectorizer &
Logistic Regression

Set up pipeline
parameters

GridSearchCV & Score

Make Predictions

Evaluate Confusion
Matrix

Multinomial Naive Bayes

X & y set up & TTS

Create Pipeline

CountVectorizer &
MultinomialNB

Set up pipeline
parameters

GridSearchCV & Score

Make Predictions

Evaluate Confusion
Matrix

— 5 | 6 —

**Model Evaluation
&
Answer the Problem**

Model Evaluation

Logistic Regression

GS Best Score: 0.7839

Training Score: 0.9707

Test Score: 0.8176

Overfit: 0.1531

Baseline: 0.5651

Multinomial Naive Bayes

GS Best Score: 0.7879

Training Score: 0.9244

Test Score: 0.7925

Overfit: 0.1322

Baseline: 0.5651

Logistic Regression

	Predicted Negative	Predicted Positive
Actual Negative	148 TN	52 FP
Actual Positive	27 FN	206 TP

Confusion Metrics

- **Accuracy = 0.8175519630484989**
- **Misclassification = 0.18244803695150116**
- **Sensitivity = 0.8841201716738197**
- **Specificity = 0.74**
- **Precision = 0.7984496124031008**

Multinomial Naive Bayes

	Predicted Negative	Predicted Positive
Actual Negative	148 TN	40 FP
Actual Positive	50 FN	195 TP

Confusion Metrics

- Accuracy = 0.792147806004619
- Misclassification = 0.2078521939
- Sensitivity = 0.7959183673469388
- Specificity = 0.7872340425531915
- Precision = 0.8297872340425532



In Summary

- Trump is largely the topic of conversation
- Vilify opposing representatives & news networks
- Wedge issues are ingrained and are not up for debate
- Yes, we can build an accurate classification model
 - Can improve through gathering more text (“selftext” field in JSON), Adjust min_df, max_features, new model?

References

- Reddit Logo(head)
- Reddit Logo (pez)
- The Federalist: Creep Uncle Joe
- USA Today: Avocados & Trump's Border Fight