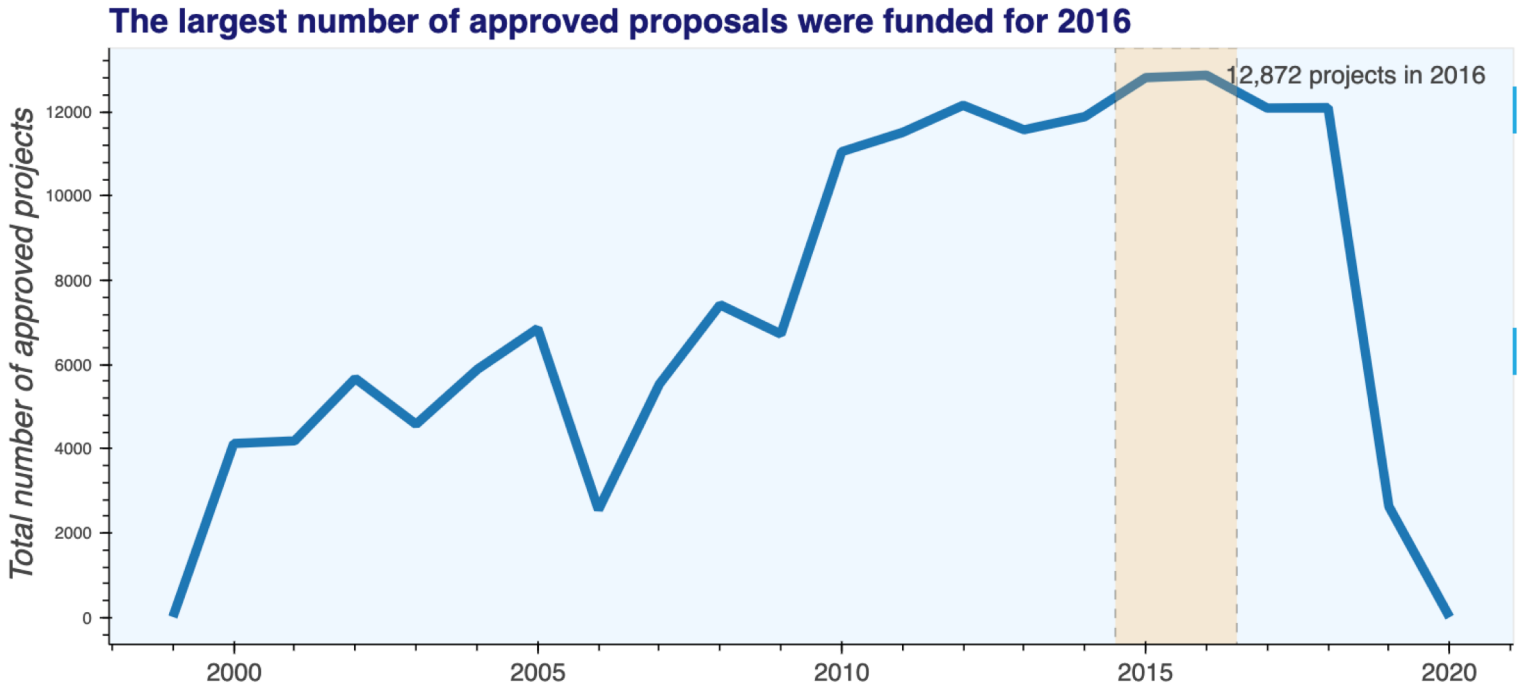
Text Analysis NSF Abstract Awards

Using text of funded project proposals to predict funding

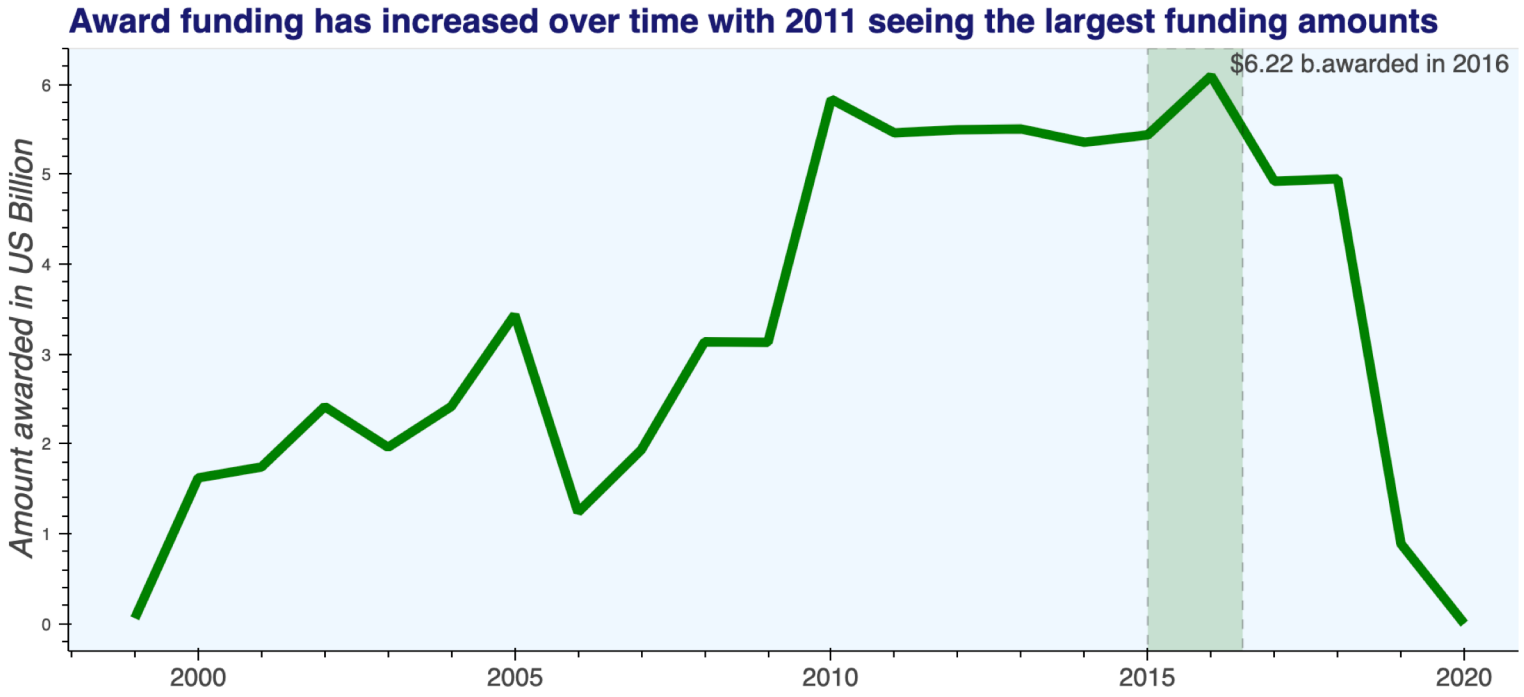
Prepared by
Geovanna Meier

Government funding is one of the most common ways to get funding



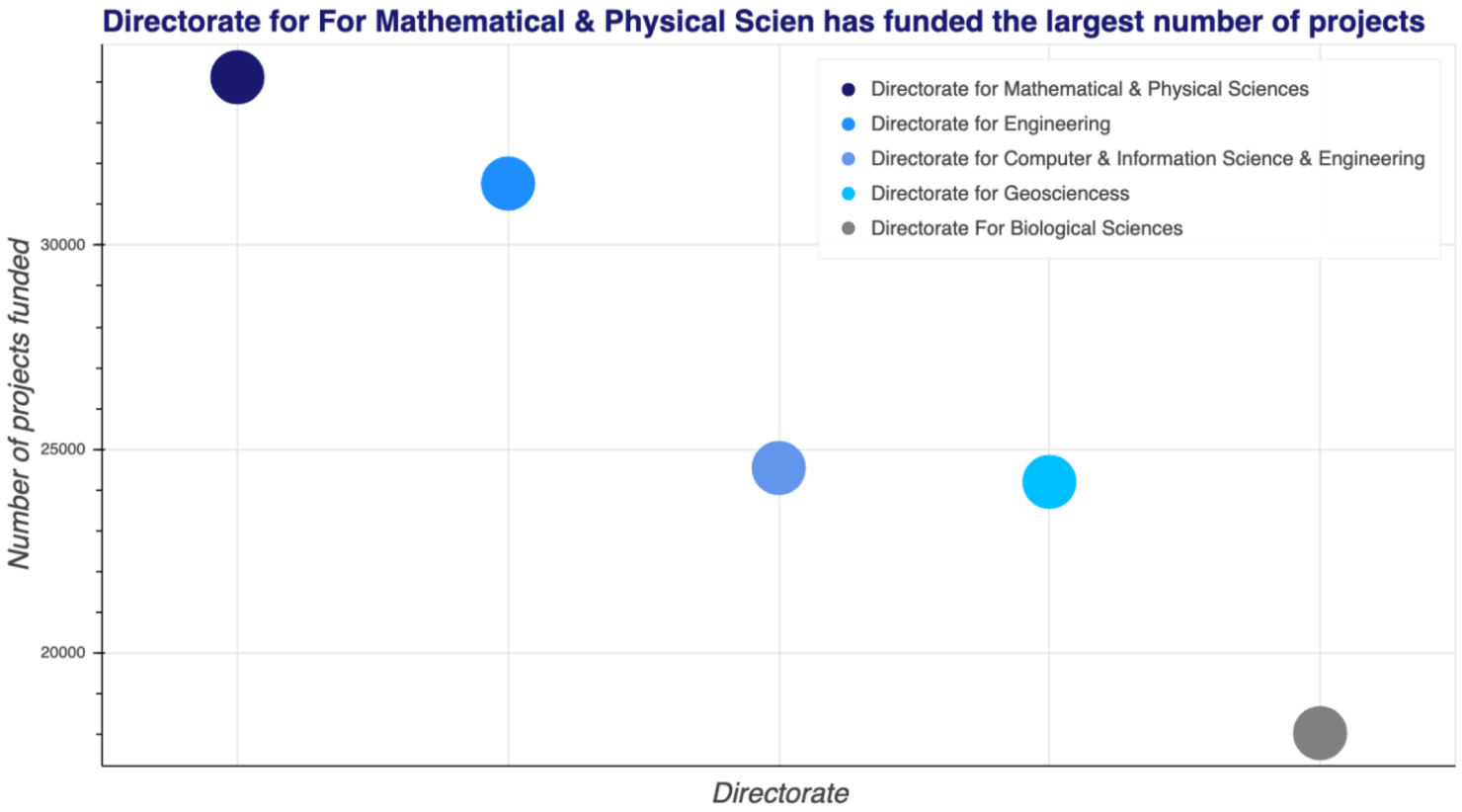
Since 2000, thousands of projects have been funded by the National Science Foundation through its directorates. The largest number of funded projects was in 2016 with over 12,000

Funding has increased steadily over the years



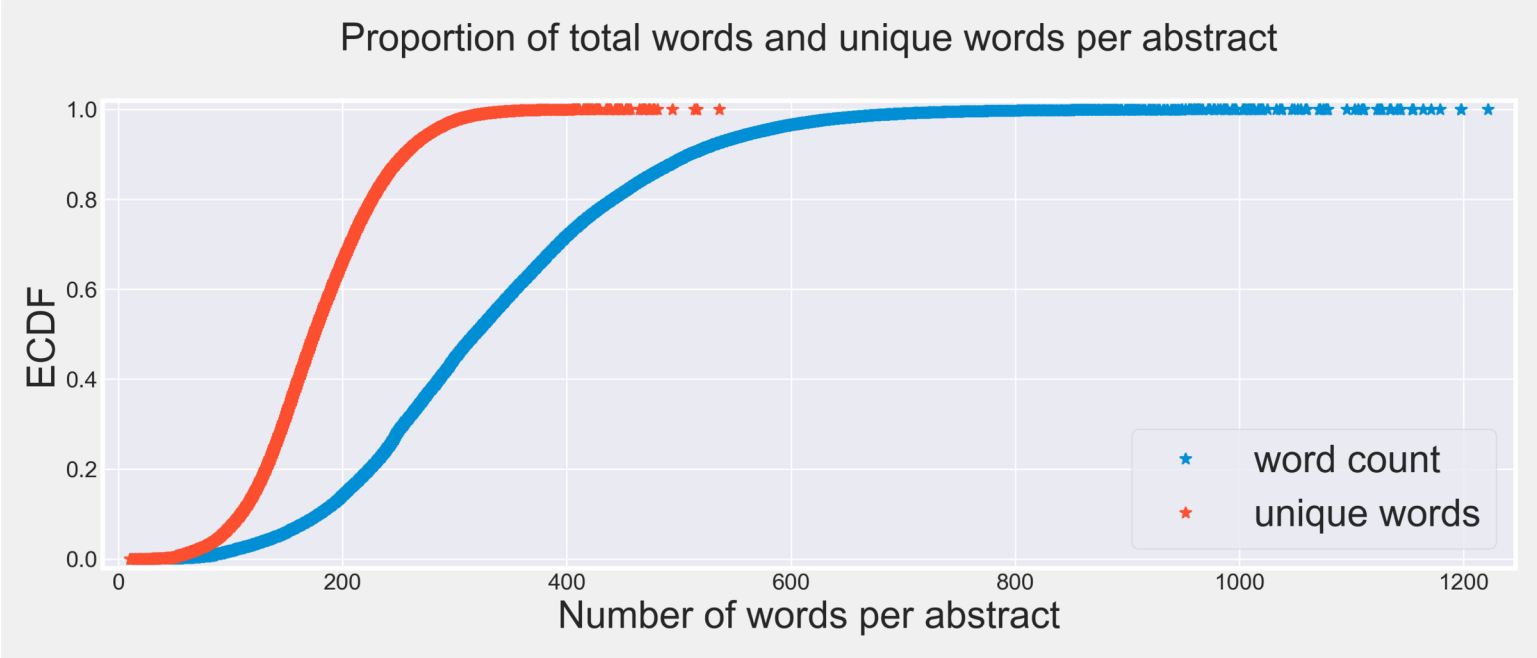
Only year with decline in funding was 2006 due to congressional budget cuts. However, in general funding has increased with 2016 having the largest funding amounts

What does it take to get funded? More Math and physical science research.



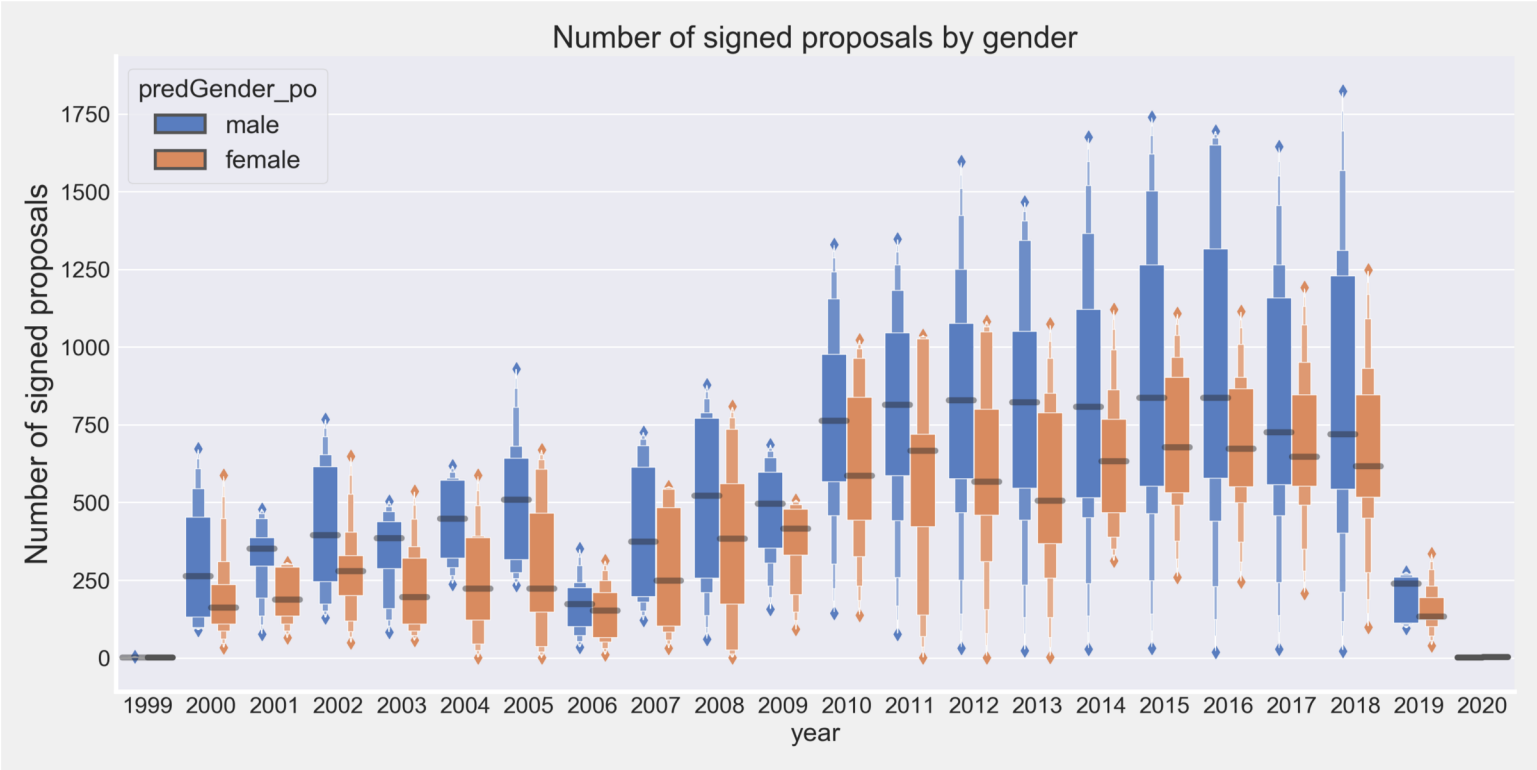
The NSF division that provides the largest amount of funding is the directorate for Mathematical and Physical Science. Hence, projects around physical science tend to receive largest amounts of aid.

Words count



Although the number of words per abstract did not have a direct impact on the amount awarded, the number of unique words did. More unique words has a positive impact on funding obtained.

Gender of approver.



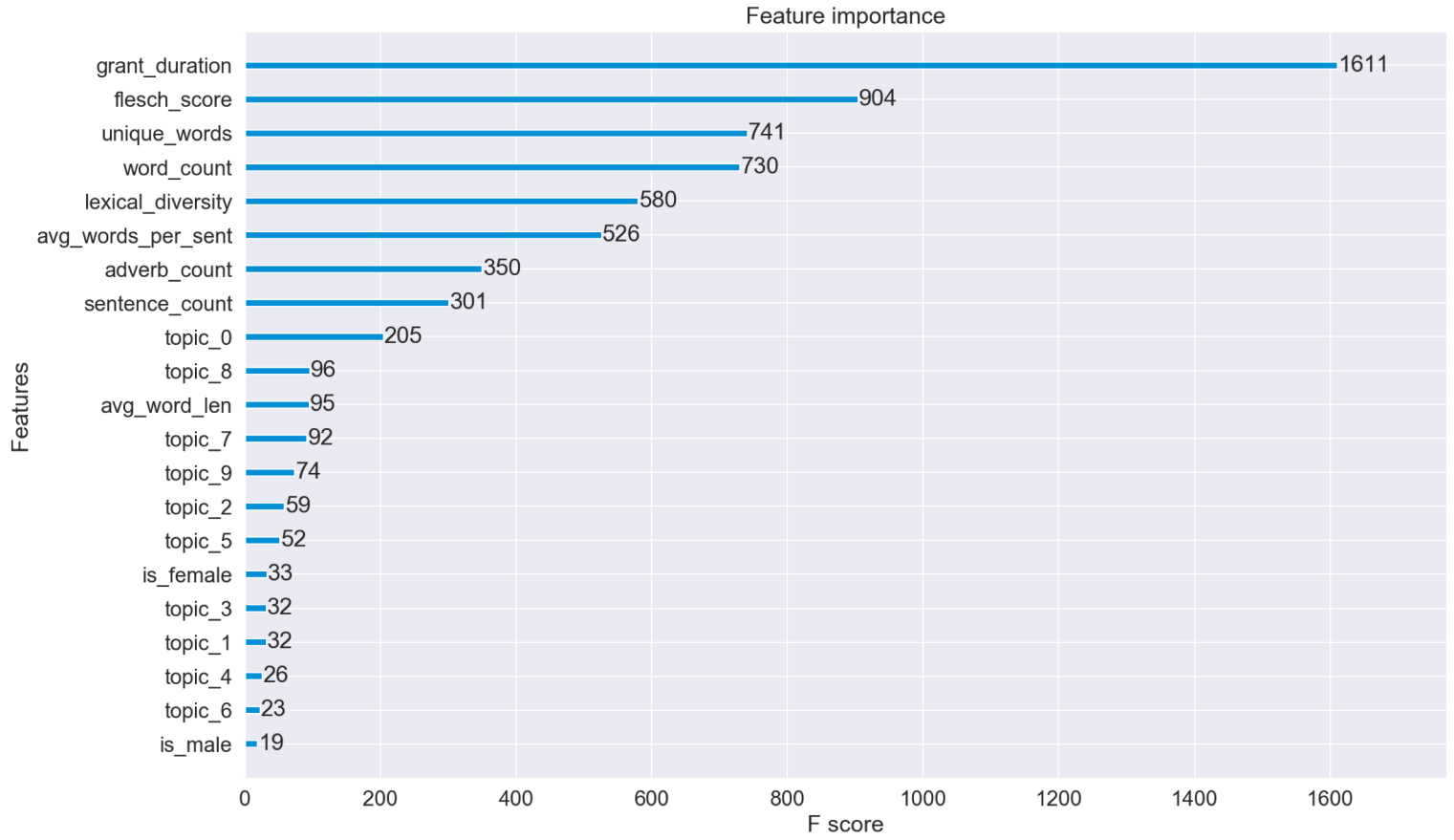
In general, more projects were signed and approved by males. 2006 was an even year. This doesn't necessarily indicate that males approve more awards, but it could indicate that there are less women in the program officer role.

Predicting Award Amounts

XGBoost

- The goal of XGBoosting is to train multiple models and sequentially combine them to improve predictive accuracy until the there is no more 'relevant' improvement and 'errors' has been minimized to the possible extent.
- Target variable: Two models were created using two variances of award amounts:
  - Total award amount
  - Per day award amount.

Project duration, ease of read and unique words are worth more

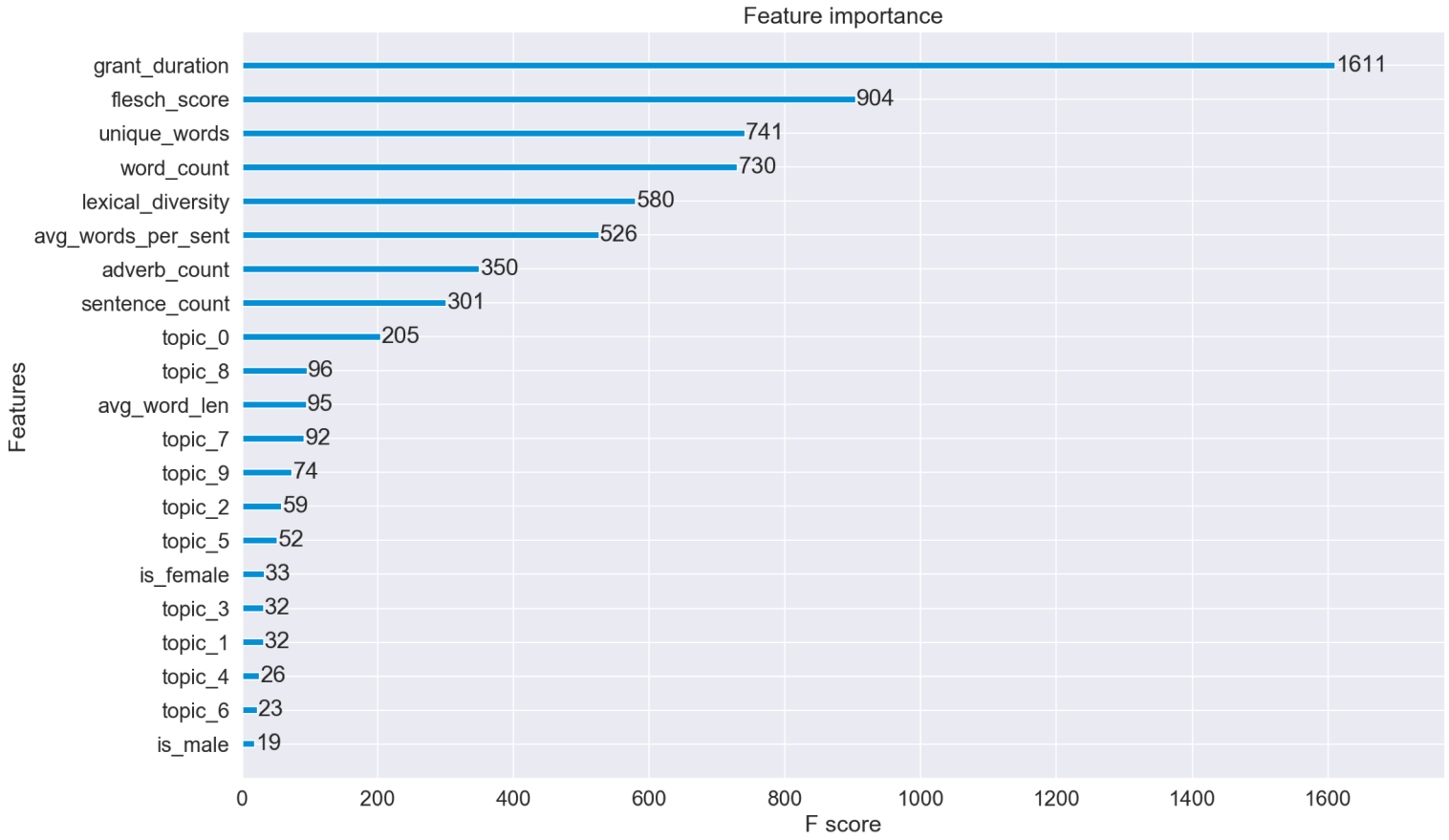


The XGBoost predictive model found that project duration, Flesch reading ease score and the number of unique words are the most important predictors of funding received.

Project duration: Projects that last more days get more funding.

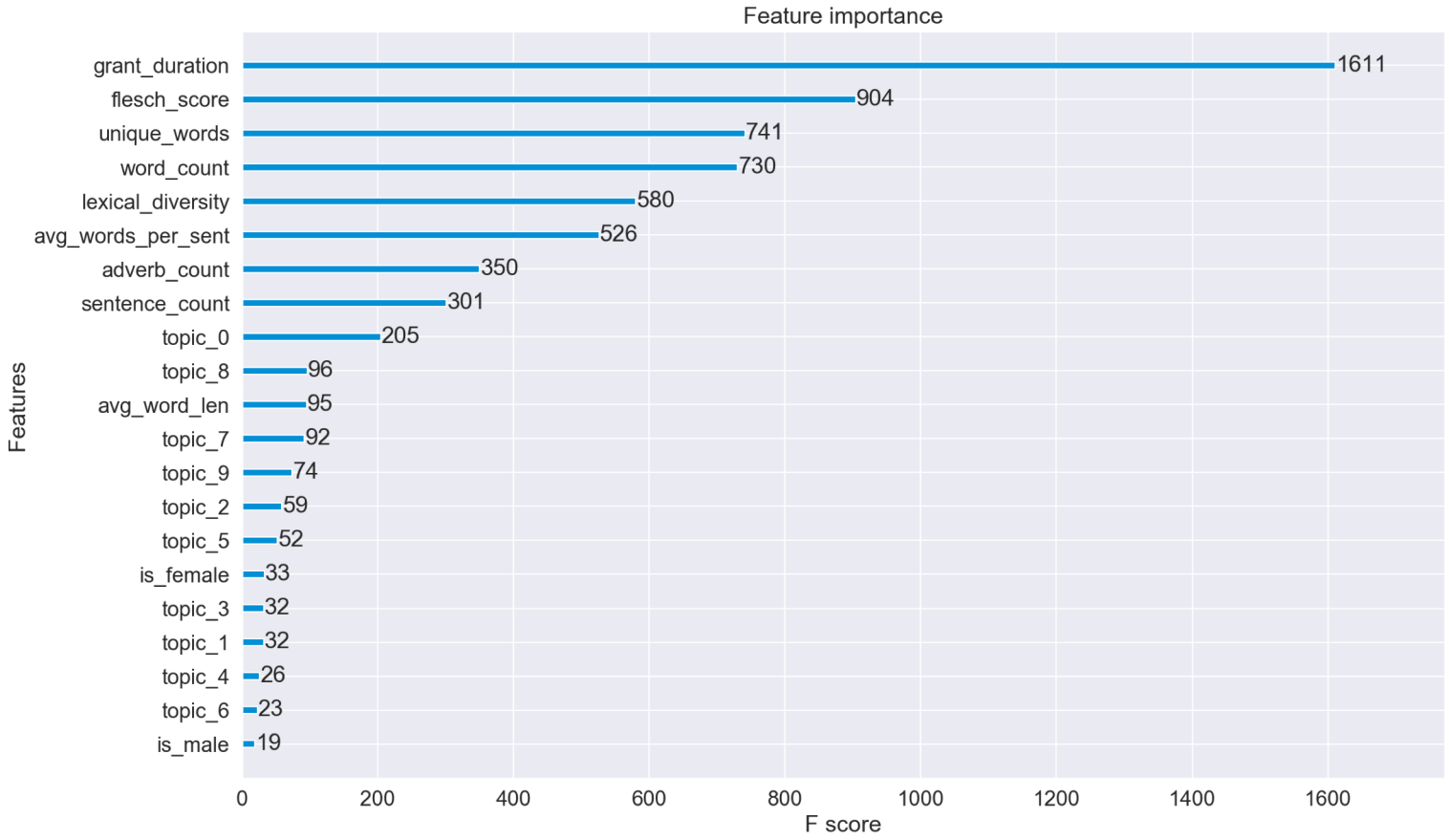
Grant duration: longer projects tend to have more depth and involve greater collaboration because they are generally working on breakthrough technology. Hence more funded is needed.

Ease of read: Clarity helps funding.



Flesch reading ease score: The flesch reading ease score measures the clarity and ease of read of a text. A lower the score, means that the text is more difficult to read. Easier to read proposals tend to get more funding.

Unique words are worth more.



Unique words: Abstracts with a greater number of unique words show less ambiguity. Redundancy can hurt funding as it makes the text more difficult to follow and understand.

Text Analysis provides insights on funding

Conclusions

- Mindful and clear writing (more unique words) as well as readability are two of the most important factors when it comes to ensuring funding.
- Proposals need to be persuasive, clear and understandable.
- Less redundancies = More funding
- Longer projects tend to obtain greater funding