Why the Rate of Active Wikipedia Users Has Been Rapidly Decreasing

EJ Seong, Indiana University, Bloomington

Thomas Johnson, Elizabeth City State University

INTRODUCTION

The goal of the project was to extract data from Wikipedia regarding a specific cohort to monitor the effects that Wikipedia's current system of maintenance and management has on the users that are active during said time. We observed the factors that may cause members of said cohort to remain editing within Wikipedia or to abandon Wikipedia altogether.
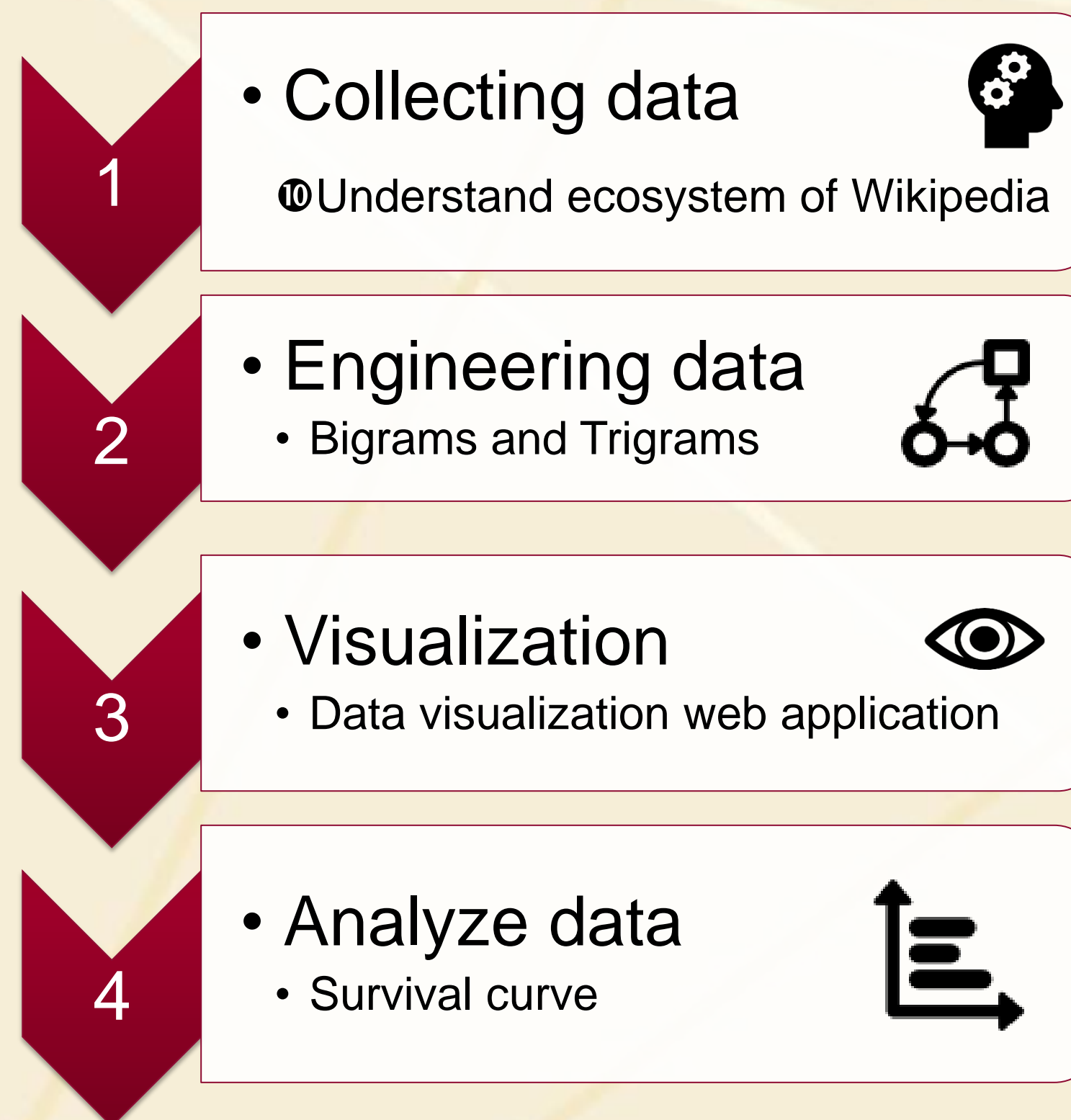
- Total Number of Users: 257,560
- Total Number of Users Who Edit: 54,165
- Total Number of Edits: 616,134
- Main Article Edits: 408,308

METHODS

- To collect related data:
  - Wikipedia API.
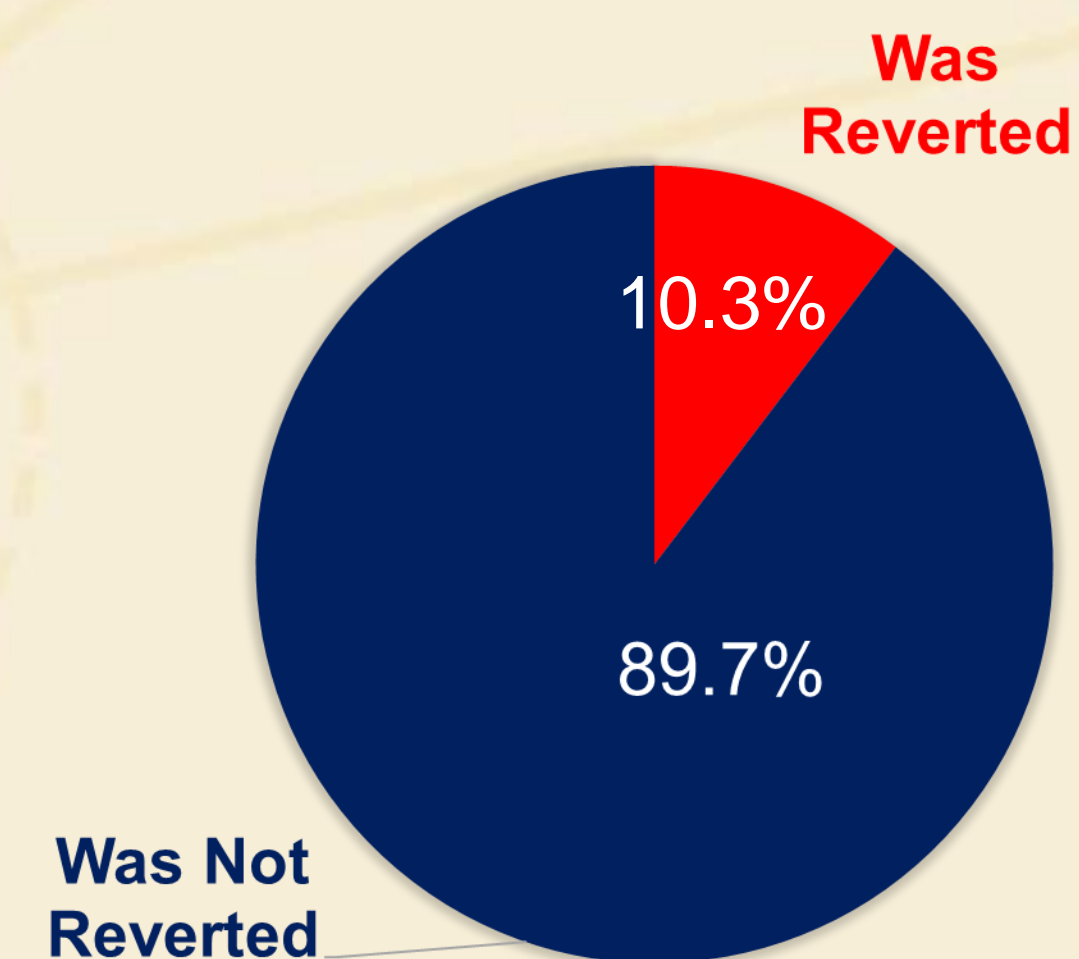 - Parse comments by using regular expressions.
- Make new features

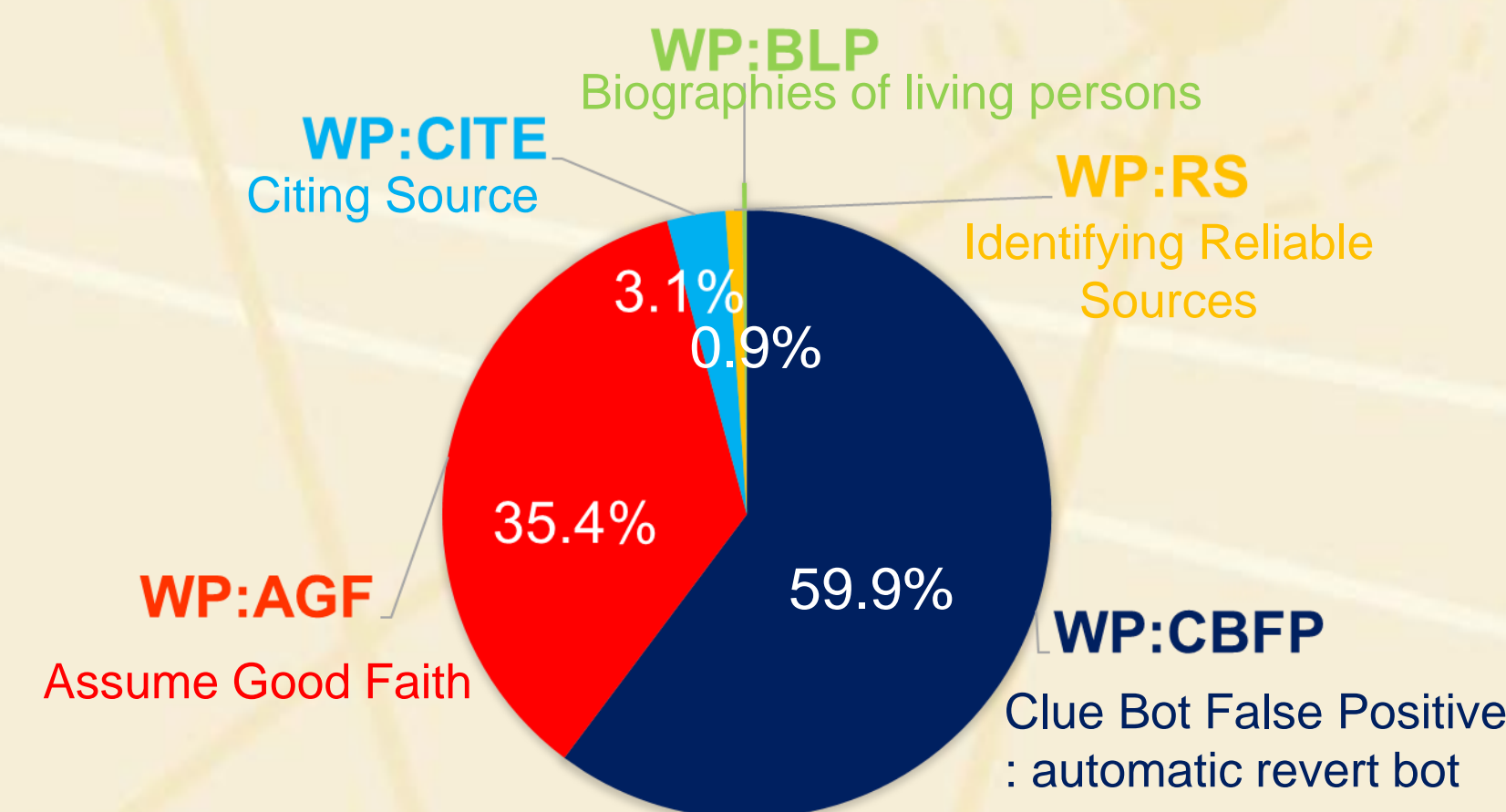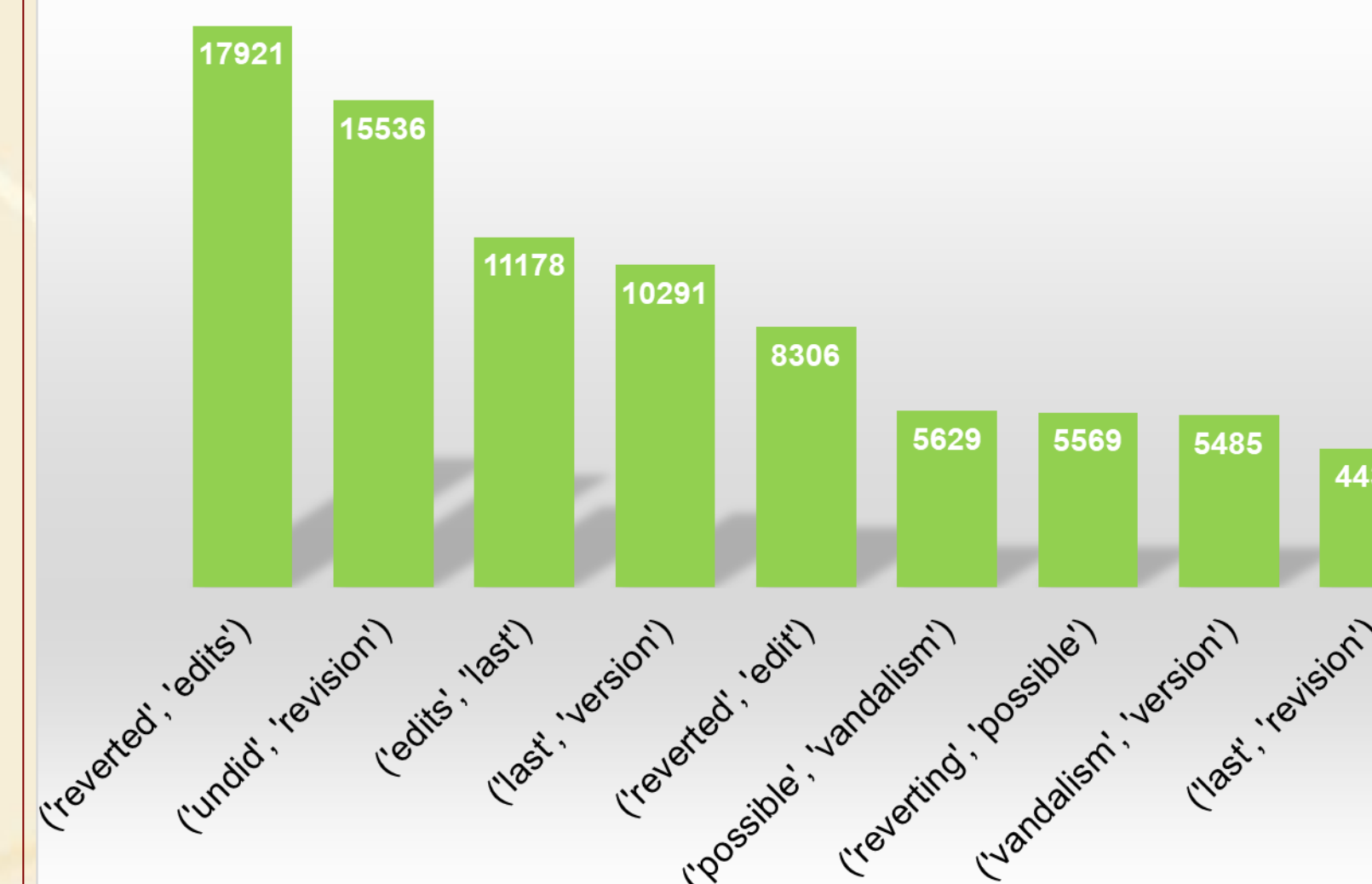| Was Reverted	Whether Wikipedia edits were reverted by users
Used Policies	What policies users cite when they revert edits
Used Tools	What tools users implemented when they revert edits

PROCESS

1. • Collecting data
   • Understand ecosystem of Wikipedia
2. • Engineering data
   • Bigrams and Trigrams
3. • Visualization
   • Data visualization web application
4. • Analyze data
   • Survival curve

RESULT

WAS REVERTED VS WAS NOT REVERTED

Was Reverted 10.3%
Was Not Reverted 89.7%

WAS REVERTED : MOST USED POLICY

WP:BLP Biographies of living persons
WP:CITE Citing Source
WP:RS Identifying Reliable Sources
WP:AGF Assume Good Faith
WP:CBFP Clue Bot False Positive : automatic revert bot

3.1%
0.9%
35.4%
59.9%

Frequency of Bigrams Chart

17921 ('reverted', 'edits')
15536 ('undid', 'revision')
11178 ('edits', 'last')
10291 ('last', 'version')
8306 ('reverted', 'edit')
5629 ('possible', 'vandalism')
5569 ('reverting', 'possible')
5485 ('vandalism', 'version')
4433 ('last', 'revision')

User Survival Graph by Receiving Revert

- 25% Reverted
- 25%-50% Reverted
- 50%-75% Reverted
- 75% Reverted
- Not Reverted

At risk						
<25% Reverted	2622	885	479	350	259	212
25% - 50% Reverted	2149	68	21	7	4	4
50% - 75% Reverted	3137	17	3	2	1	1
> 75% Reverted	9457	3	0	0	0	0
Not Reverted	25198	378	102	49	28	15
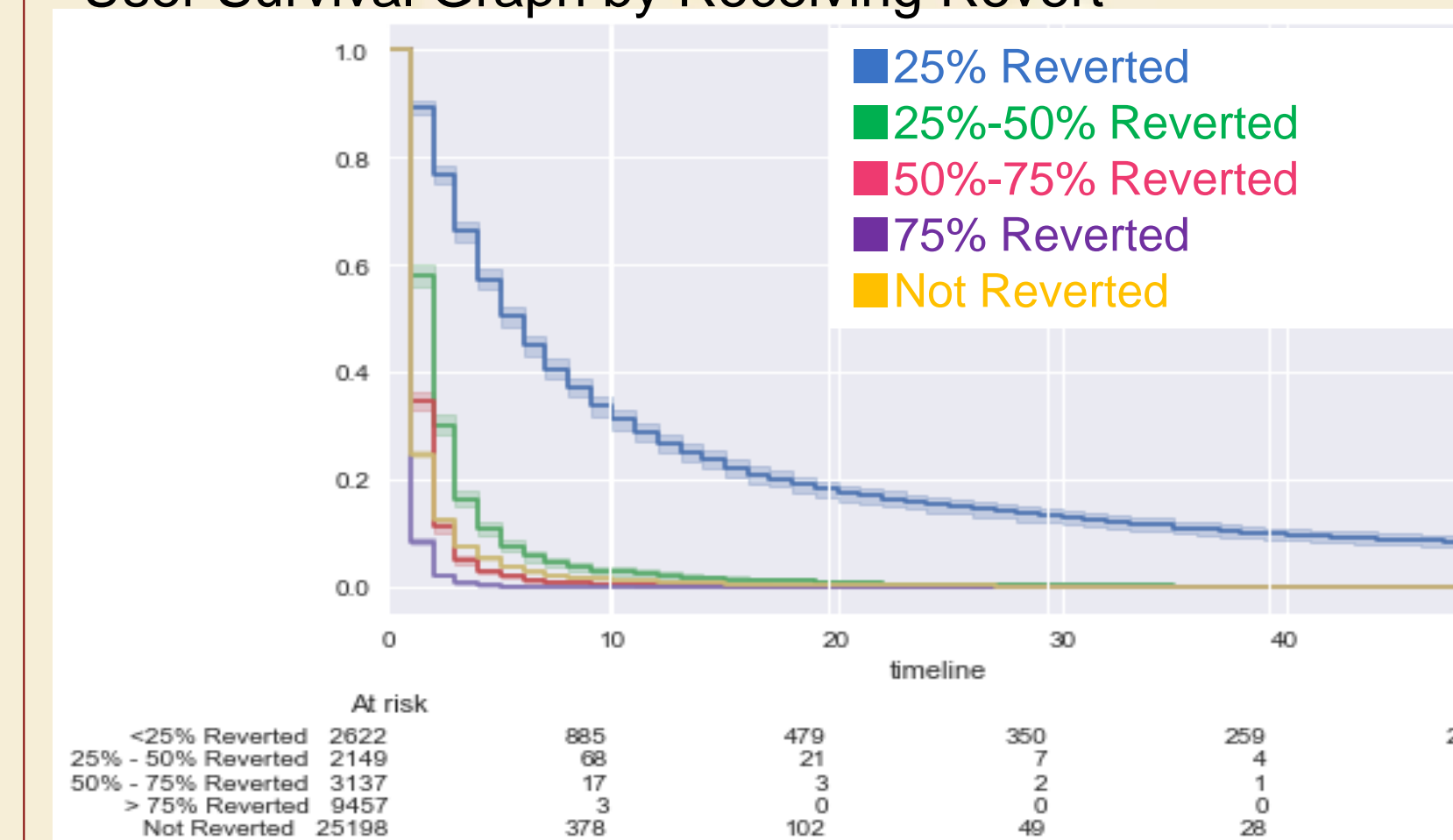
CONCLUSION

- Actual revert rate is 10% but as displayed in the bigrams chart, revert related words are used most frequently.
- Considering the usage of WP:CBFP as the most cited policy and the prevalence of "vandalism" in the bigram statistics, it can be inferred that Wikipedia still holds potential vandalism at a high priority.
- Proper amount of reverts (25%-50%) help new Wikipedia users to survive by providing interaction with other users rather than inattention.

FUTURE WORK

- Test data in different conditions
- Survival curve
  - Cox hazard model
- Bayesian model
- Upgrade the data visualization web application

ACKNOWLEDGEMENTS

INDIANA UNIVERSITY BLOOMINGTON
SCHOOL OF INFORMATICS AND COMPUTING

National Science Foundation
WHERE DISCOVERIES BEGIN