

General Assembly - DAT-19

Predicting NaNoWriMo

Winners

NICOLE FRONDA

What is NaNoWriMo?



National Novel Writing Month

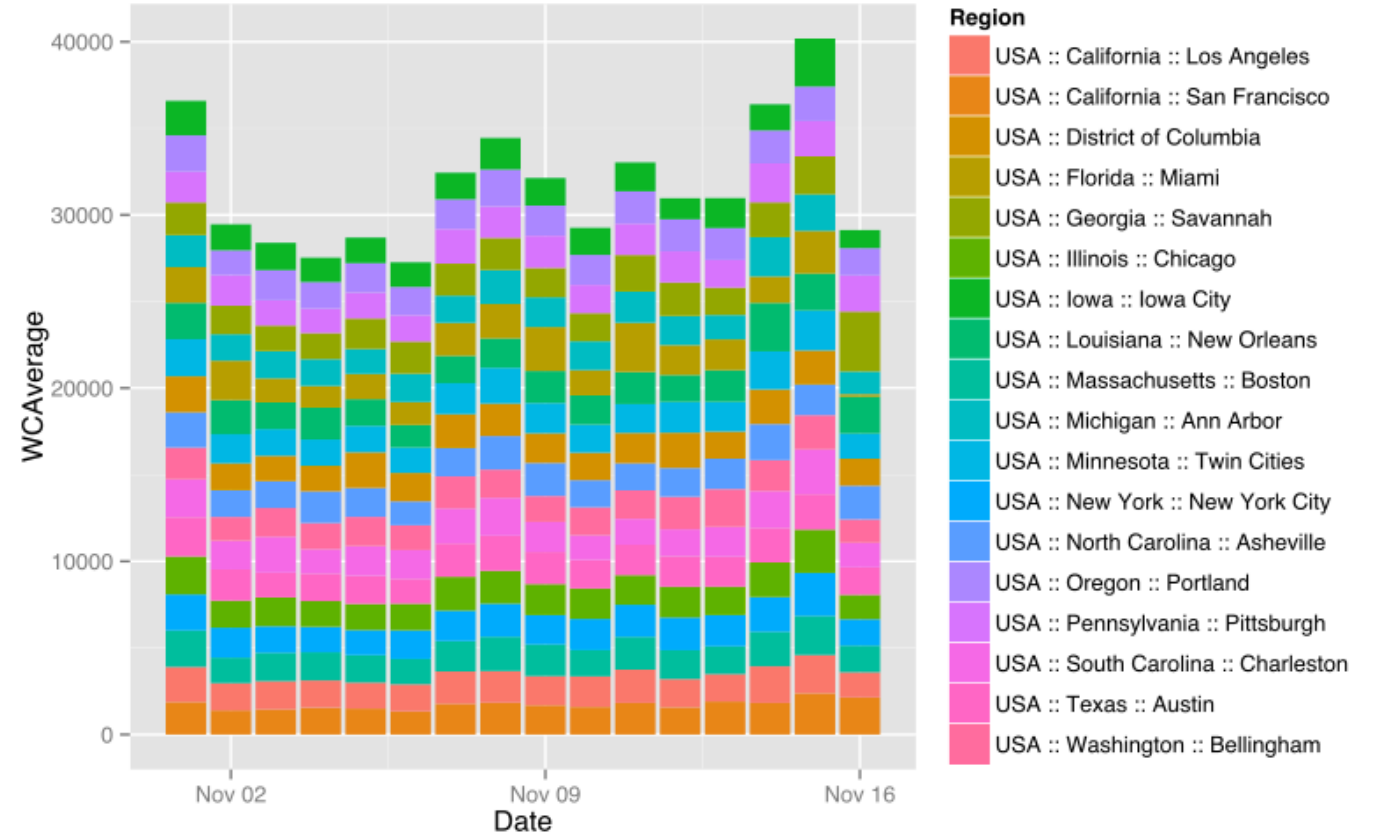
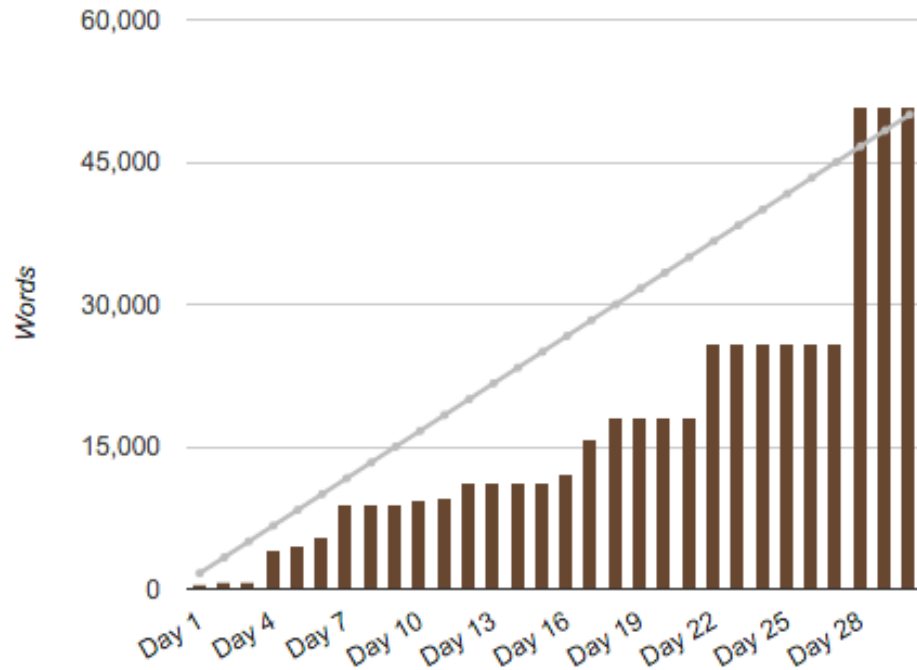
“On November 1, participants begin working towards the goal of writing a **50,000-word** novel by 11:59 PM on November 30.

Valuing enthusiasm, determination, and a deadline, NaNoWriMo is for *anyone* who has ever thought about writing a novel.”

- <http://nanowrimo.org/about>

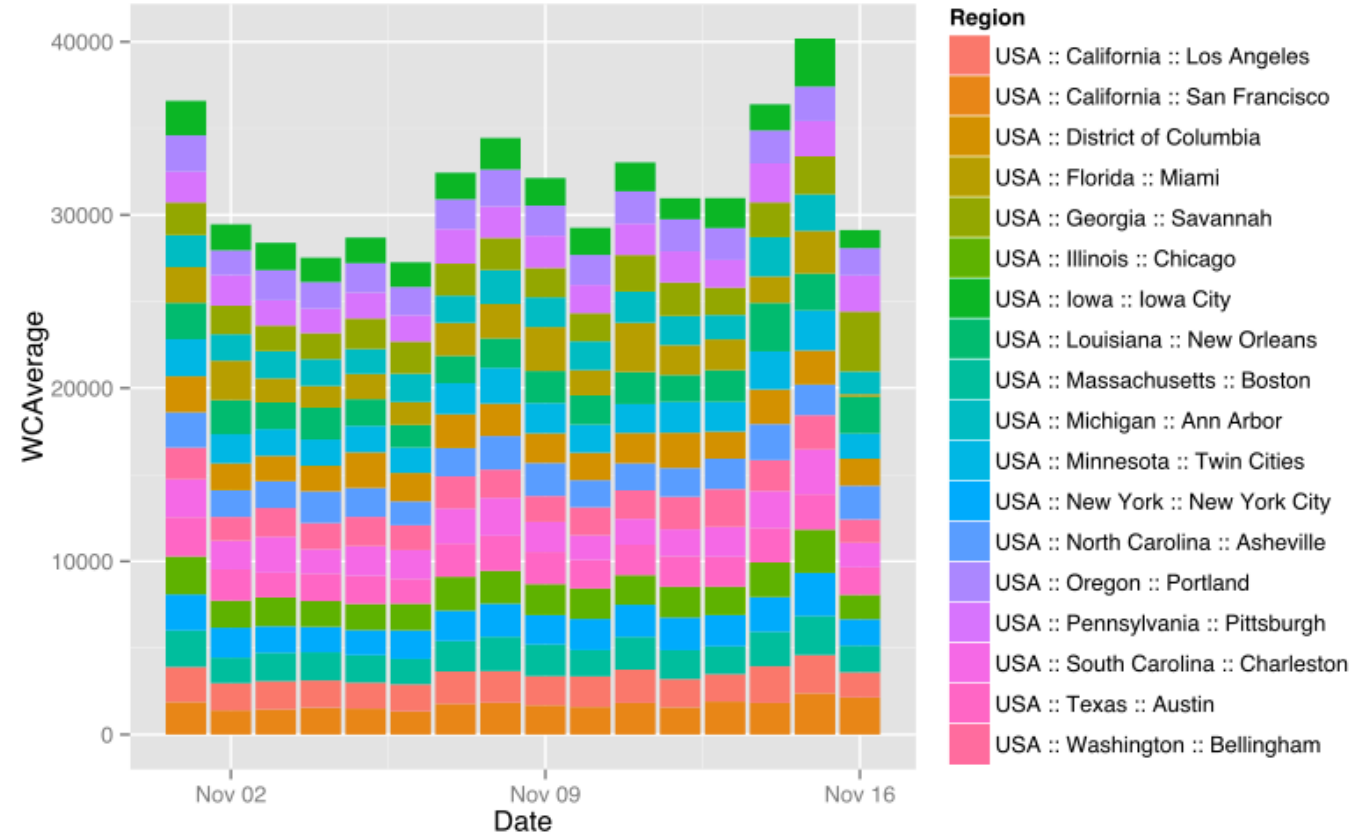
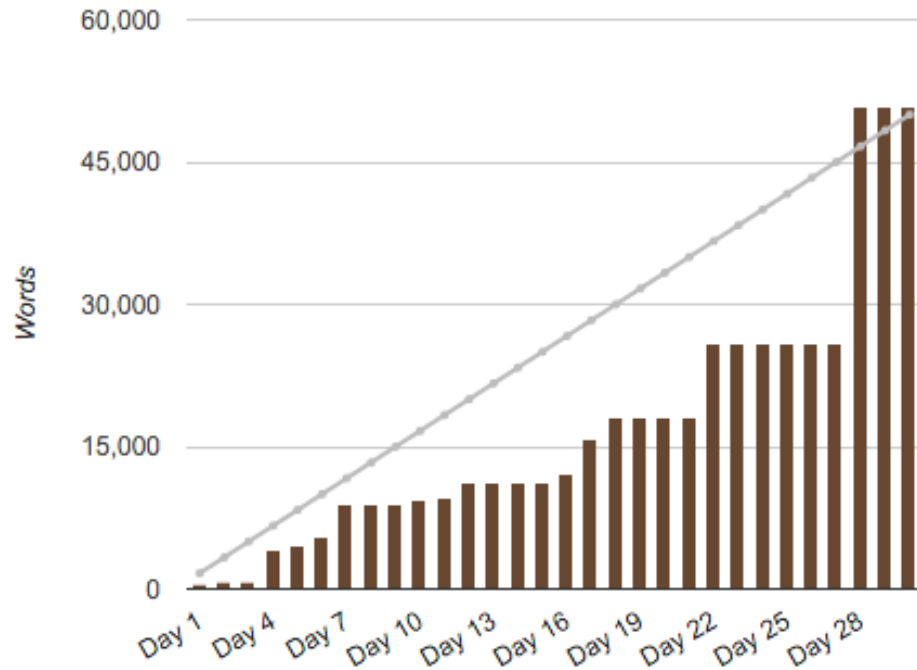
Motivation:

Tracking my Writing Progress



Motivation:

Tracking my Writing Progress



Goal: Predict if a writer or novel will win the next contest

Writer Data

Past NaNoWriMos		Ongoing NaNoWriMo	
<u>Numerical</u>	<u>Binary</u>	<u>Numerical</u>	<u>Binary</u>
Member Length	Municipal Liaison	First Week Word Count	Donor
Lifetime Word Count	Sponsorship	First Week Num Submissions	
Number of Novels		Second Week Word Count	
Count Wins		Second Week Num Submissions	
Count Donations		<div>Novel Data</div>	
Average Submission			
Daily Average			
Num Consecutive Years Participated			
Num Consecutive Winning Years			

Getting the Data

Stats from Most Recent Contest – Word Count API

http://nanowrimo.org/modules/wordcount_api/wchistory/ *nicaless*

```
<wchistory>
  <uid>30837</uid>
  <uname>NewMexicoKid</uname>
  <user_wordcount>25001</user_wordcount>
  <winner>false</winner>
  <wordcounts>
    <wcentry>
      <wc>2699</wc>
      <wcdte>2005-11-01</wcdte>
    </wcentry>
    <wcentry>
      <wc>2315</wc>
      <wcdte>2005-11-02</wcdte>
    </wcentry>
  </wordcounts>
</wchistory>
```

	Writer	Word Count	Winner	1	2	3	4	5	6	7	...
0	Nicaless	24229	false	3497	1132	677	296	456	631	0	...
1	Rachel B. Moore	50603	true	5885	1760	1771	592	1050	706	4958	...
2	abookishbabe	29299	false	28632	0	0	0	0	0	0	...
3	alexabexis	78800	true	3052	4242	3296	4554	3065	2011	5140	...
4	AllYellowFlowers	11000	false	0	520	455	170	204	451	0	...

Getting the Data

Web pages – Kimono Labs

The screenshot shows the NaNoWriMo website interface. At the top, there's a browser address bar with "kimono" and a "Member Length" dropdown showing a row of colored circles with numbers 1, 1, 1, 1, 1, 1, 1, 3, and a "+" sign. The website header includes a logo with a blue shield containing icons of a coffee cup, a laptop, a pen, and a stack of papers, topped with a blue bull head. To the right of the logo, it says "Hello, Nicaless!" with a mail icon and the number 2, followed by links for "FAQ", "Search", and "Sign Out". A "Save" button is in the top right corner. Below the header, a large blue banner reads "National Novel Writing Month" and "November 1-30 • The world needs your novel." A dark brown navigation bar contains links: "My NaNoWriMo", "Regions", "Inspiration", "Conversation", "Shop", and "Donate". The main content area features a profile for "Nicaless", a member for about 2 years, with a profile picture of a dragon. It lists their current novel "Lauren's Birthday" in the "Young Adult" genre, with a progress bar showing 24,229 words. Below the profile, there are tabs for "Author Info", "Novels", "Badges", "Writing Buddies", "Buddy Of", and "Stats". The "Author Info" tab is active, showing an "Author Bio" section with an "Edit" button. The bio text reads: "I've been writing stories since I've been able to hold a pencil in my hand, yet I've always been too scared to share, or worse, too scared to finish. I hope to change that this time." and "It's my first NaNoWriMo! Good luck to all of us." To the right of the bio, there's a "Years Done/Won" section showing two circular progress indicators for years 14 and 15. At the bottom right, a "Lifetime Achievement: Total NaNo Word Count" section shows a total of 75,148 words.

kimono Member Length 1 1 1 1 1 1 1 3 +

Hello, Nicaless! 2 FAQ Search Sign Out Save

National Novel Writing Month

November 1-30 • The world needs your novel.

My NaNoWriMo Regions Inspiration Conversation Shop Donate

Nicaless
Member for about 2 years

Novel: Lauren's Birthday 24229 words
Genre: Young Adult

Author Info Novels Badges Writing Buddies Buddy Of Stats

Author Bio Edit

I've been writing stories since I've been able to hold a pencil in my hand, yet I've always been too scared to share, or worse, too scared to finish. I hope to change that this time.

It's my first NaNoWriMo! Good luck to all of us.

Years Done/Won

14 15

Lifetime Achievement: Total NaNo Word Count

75,148

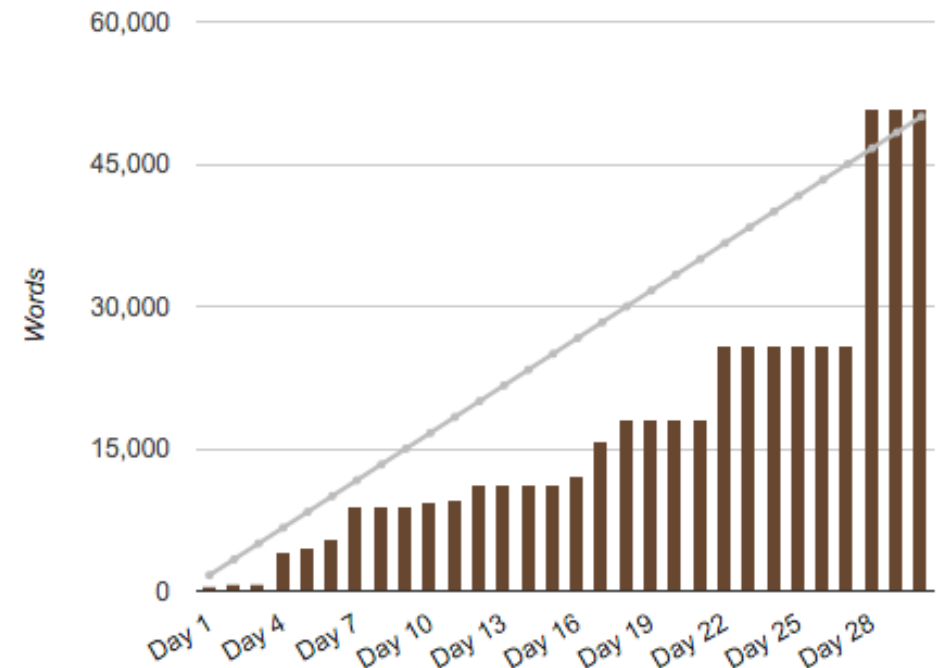
Getting the Data

Web pages – BeautifulSoup

Author Info	Novels	Badges	Writing Buddies	Buddy Of
A Mystery in the Kingdom of Aermon				
Author: Nicaless				
Genre: Fantasy				
Synopsis				
Hitoshi is appointed the youngest Judge and master of the Veris Saber in the history of Aermon. His new responsibilities may prove to be a bit more challenging than he first anticipated. It certainly does not help that he's standing trial for murder...				
Excerpt				
This story, funnily enough, started out as a history textbook. An academic at heart, I always knew I wanted to leave something behind for future students of the world. However, while trying to engross myself in the process of writing history, I was swept away in the process of making it.				
The events described in the following collection of newspaper article, letters, diary entries, etc., that I have procured over the years and arranged like so for this book/website(?), tell a story of magic and mystery, of science and disbelief, of treachery and thievery, of love and war.				

```
# open the url
page = urllib.urlopen(i).read()
# parse the page
soup = BeautifulSoup(page, 'html.parser')
# find the novel_synopsis tag
synopsis = soup.find(id='novel_synopsis').encode_contents()
```

```
page = urllib.urlopen(u).readlines()
for line in page:
    # find the line that starts with rawCamperData
    if line[10:23] == "rawCamperData":
        # split the string to get the wc submission values
        mystr = line[27:len(line)-3]
        mystr = mystr.split(",")
        mystr = list(mystr)
        mystr = map(int, mystr)
```

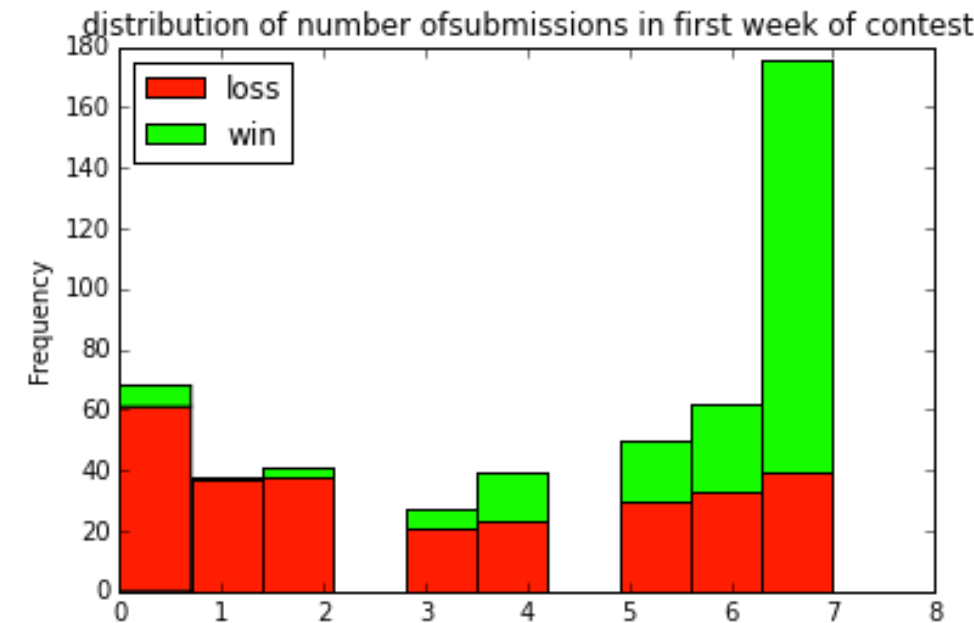
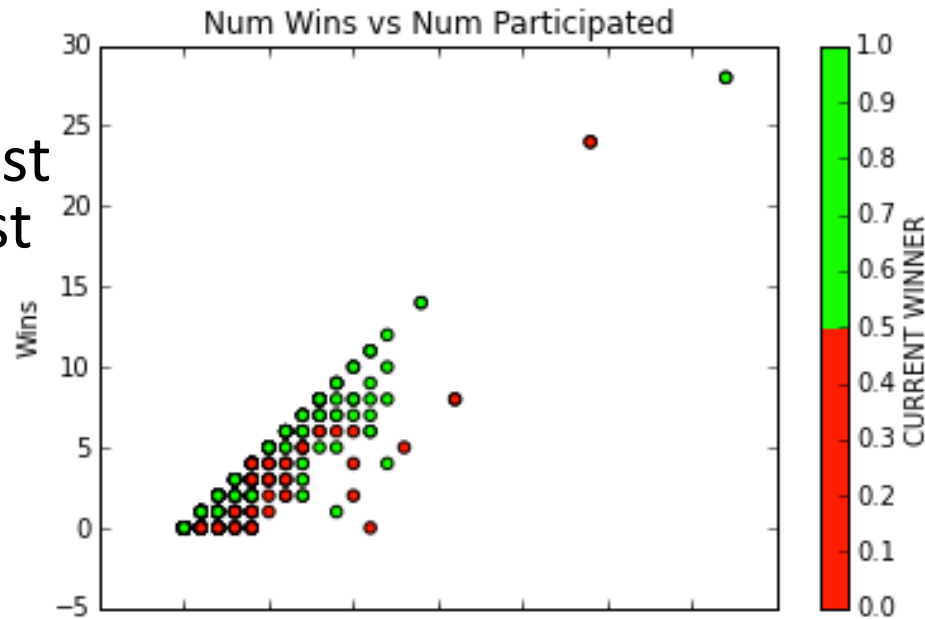


Exploring the Data: Writers

501 Writers
219 won recent contest
282 lost recent contest

Writers with
Sponsors
2x likely to Win

Municipal Liaisons
6x likely to Win



Writer Data

Past NaNoWriMos

Numerical

Binary

Member Length

Municipal Liaison

Lifetime Word Count

Sponsorship

Number of Novels

Count Wins

Count Donations

Average Submission

Daily Average

Num Consecutive Years

Participated

Num Consecutive Winning
Years

Ongoing NaNoWriMo

Numerical

Binary

First Week Word Count

Donor

First Week Num Submissions

Second Week Word Count

Second Week Num Submissions

Writer Data

Past NaNoWriMos	
<u>Numerical</u>	<u>Binary</u>
Member Length	Municipal Liaison
Lifetime Word Count	Sponsorship
Number of Novels	
Count Wins	
Count Donations	
Average Submission	
Daily Average	
Num Consecutive Years Participated	
Num Consecutive Winning Years	

Ongoing NaNoWriMo	
<u>Numerical</u>	<u>Binary</u>
First Week Word Count	Donor
First Week Num Submissions	
Second Week Word Count	
Second Week Num Submissions	

Logistic Regression – Predicting Winning Writers

	Actual 0	Actual 1
Predicted 0	46	7
Predicted 1	22	34

	Precision	Recall	F1-Score	Support
0	0.69	0.87	0.77	55
1	0.77	0.52	0.62	46
avg/total	0.73	0.71	0.70	101

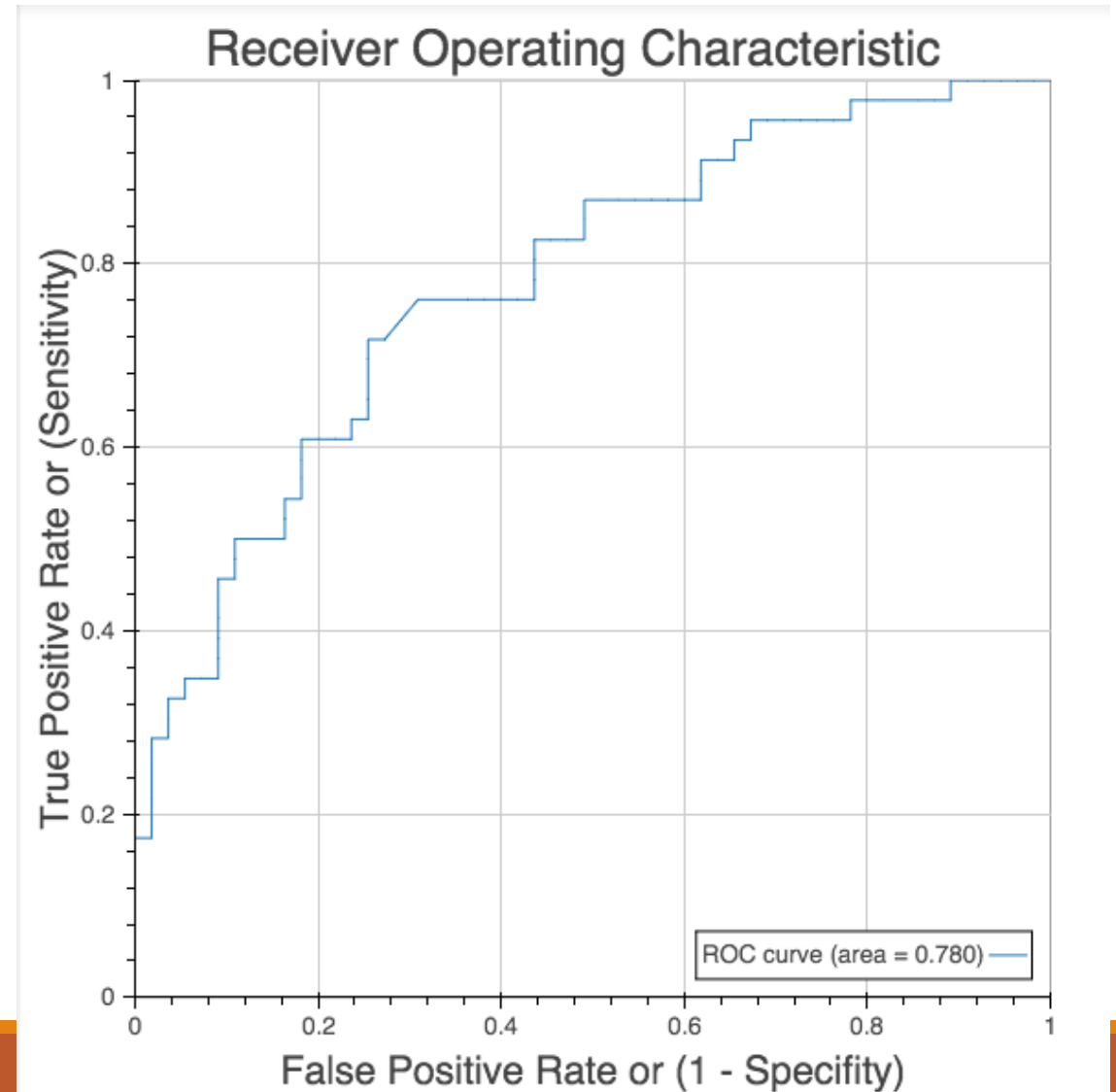
Other Models:

Naïve Bayes – 67%

Support Vector Machine – 72%

Decision Tree – 65%

Random Forest – 74%



Writer Data

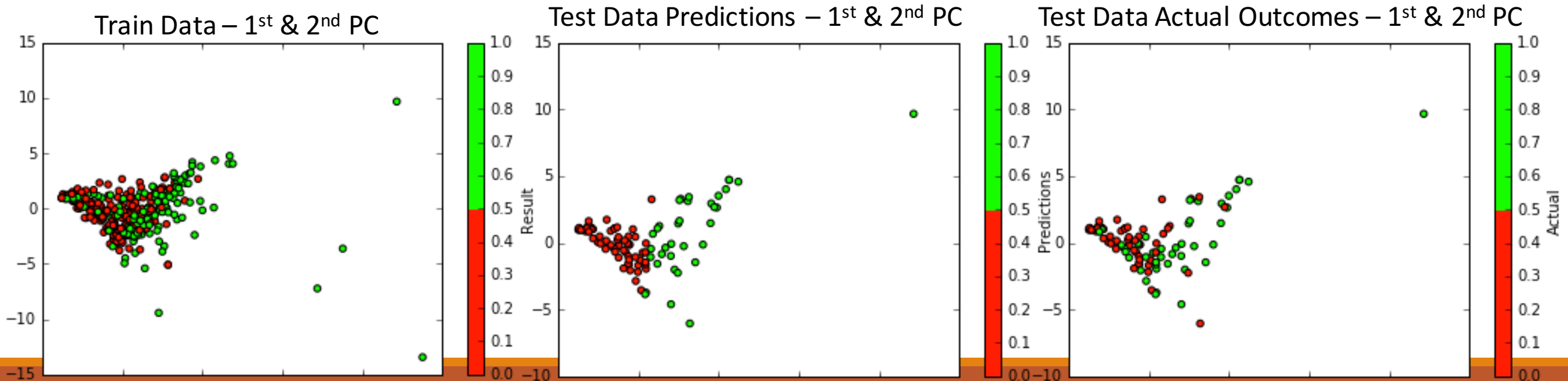
Past NaNoWriMos		Ongoing NaNoWriMo	
<u>Numerical</u>	<u>Binary</u>	<u>Numerical</u>	<u>Binary</u>
Member Length	Municipal Liaison	First Week Word Count	Donor
Lifetime Word Count	Sponsorship	First Week Num Submissions	
Number of Novels		Second Week Word Count	
Count Wins		Second Week Num Submissions	
Count Donations			
Average Submission			
Daily Average			
Num Consecutive Years Participated			
Num Consecutive Winning Years			

Logistic Regression – Predicting Winning Writers

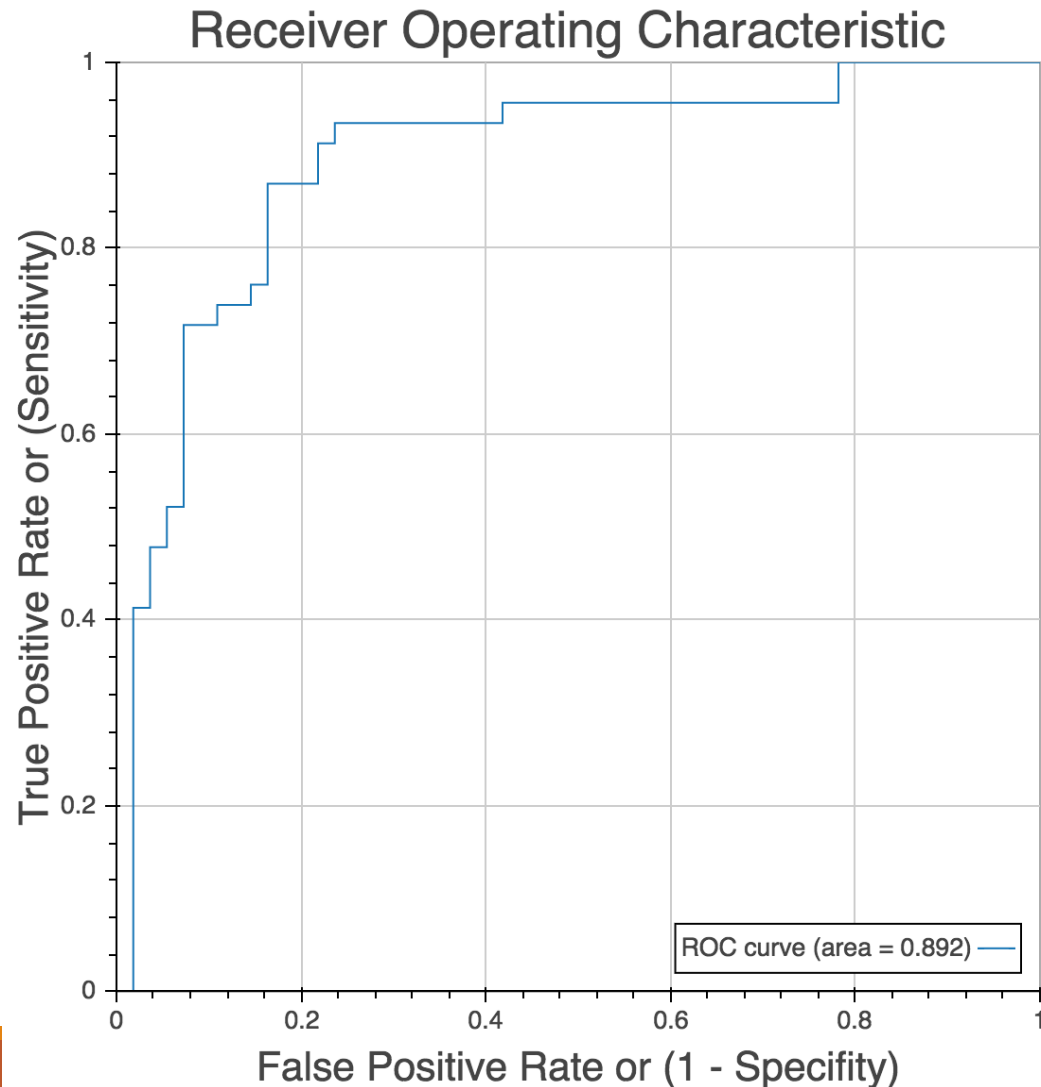
Include Word Count from current contest

	Actual 0	Actual 1
Predicted 0	46	9
Predicted 1	8	38

	Precision	Recall	F1-Score	Support
0	0.85	0.84	0.84	55
1	0.81	0.83	0.82	46
avg/total	0.83	0.83	0.83	101



Logistic Regression – Predicting Winning Writers



Feature	Measured Importance
Second Week Word Count	0.712847
LifetimeWordCount	0.057194
Second Week Num Submissions	0.045414
Expected Avg Submission	0.039117
Consecutive Part	0.036241

Data of progress in first few weeks helps improve predictions

Next Steps

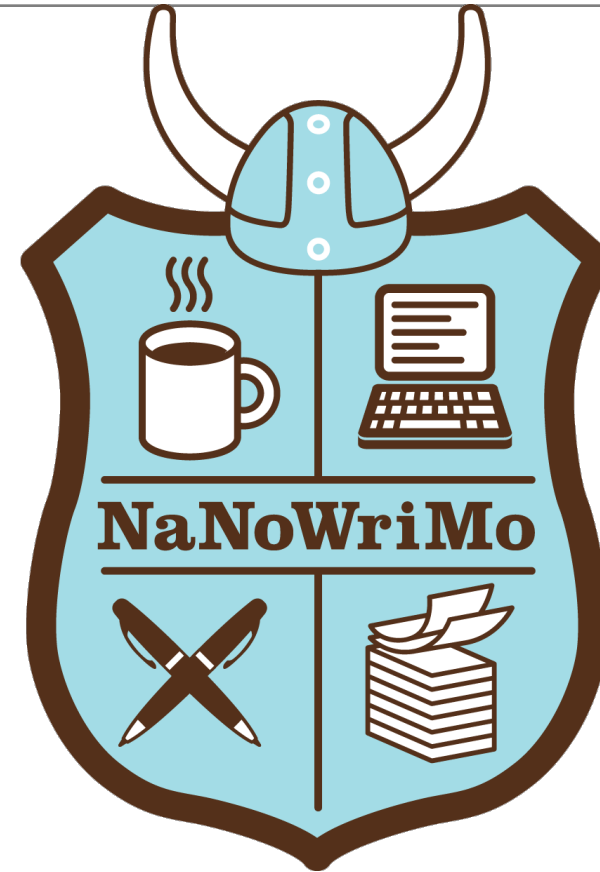
Improving the model

- Collect more data
- Feature Engineering

Predicting Final Word Count

Discovering clusters of writers

- (Not just winners and losers)



Next Steps

Improving the model

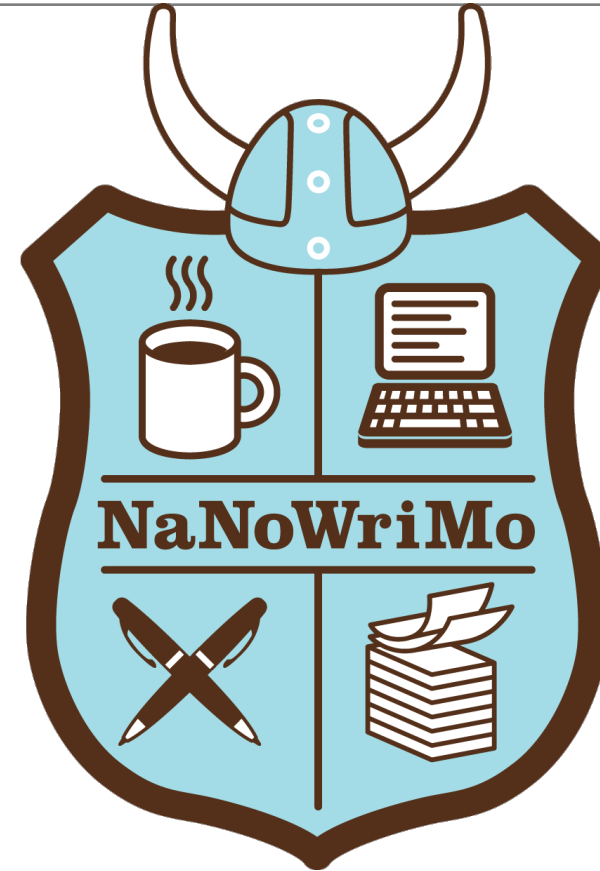
- Collect more data
- Feature Engineering

Predicting Final Word Count

Discovering clusters of writers

- (Not just winners and losers)

Win the Next NaNoWriMo 😊



Questions?
