

What Makes an Award Winning Film Script?

Allison Ragan
Allison Shafer

Agenda

- I. Cleaning
- II. Sentiment Analysis
- III. Topic Modeling
- IV. K-Means Clustering
- V. Areas of Opportunity



Cleaning the Data

- ◆ Remove extraneous text
- ◆ Parse dialogue from full script
- ◆ Normalize for vectorization
- ◆ Introduce supplementary metadata on the film

INT. THROMBEY ESTATE - 2ND FLOOR - DAWN

Follow one housekeeper named FRAN carrying a tray of coffee up a flight of stairs.

Second floor: a hallway, doors all closed. The house has not woken up, and Fran steps lightly. Up a much narrower creaky flight of steep stairs.

INT. THROMBEY ESTATE - 3RD FLOOR MASTER BEDROOM - DAWN

Third floor: the master bedroom suite.

FRAN
Morning Mr Thrombey

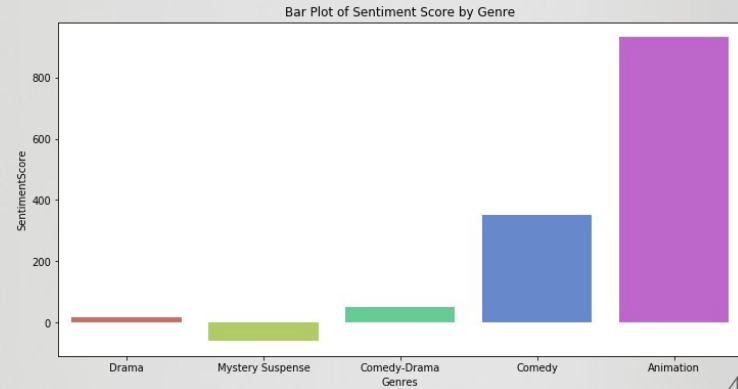
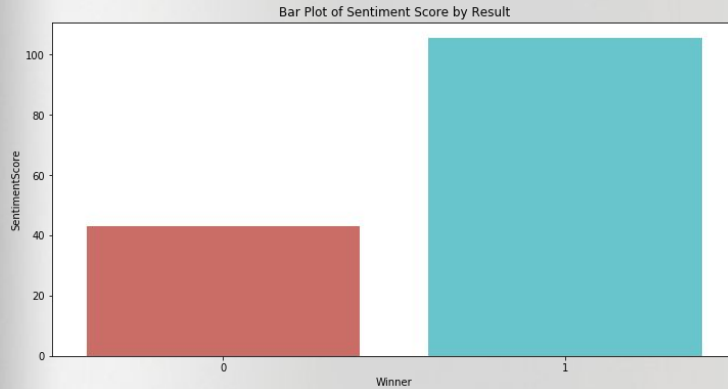
But the bed is empty, unslept in. A robe thrown across it.

Fran heads out onto the landing and UP an EVEN NARROWER half flight of stairs, which leads to a single door.

FRAN (cont'd)
Mr Thrombey you up there? Mr Thrombey
I'm coming in

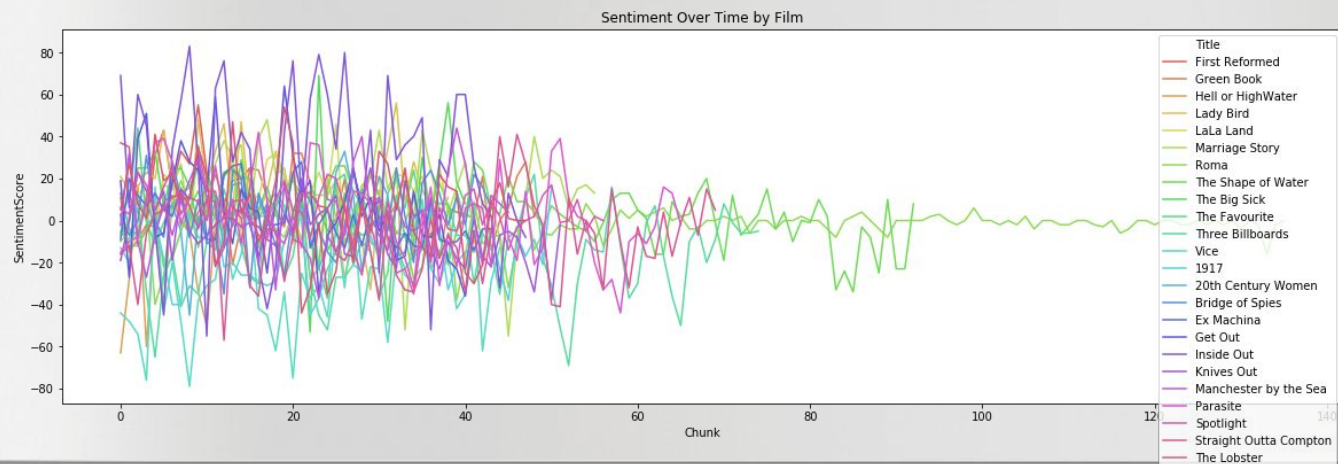
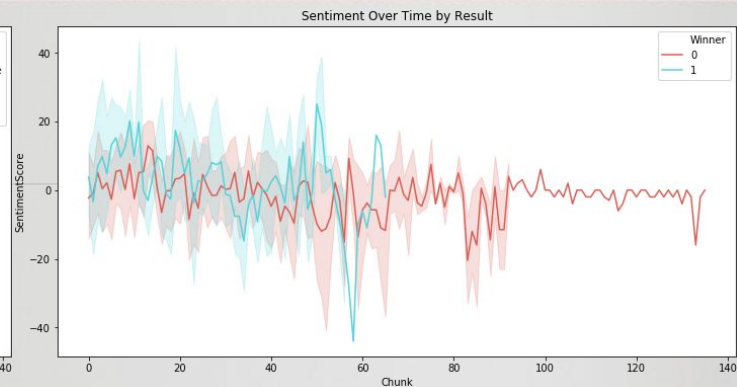
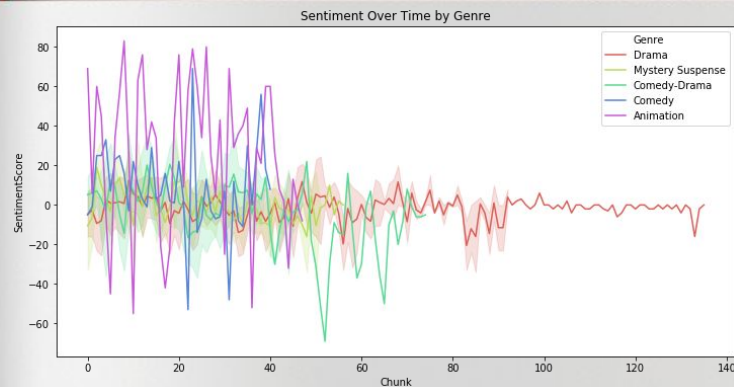
Sentiment Analysis

- ◆ AFINN Lexicon (the “new ANEW”)
- ◆ Drama had lower sentiment scores, comedy had higher
- ◆ Winners were overall positive



Top Words by Sentiment

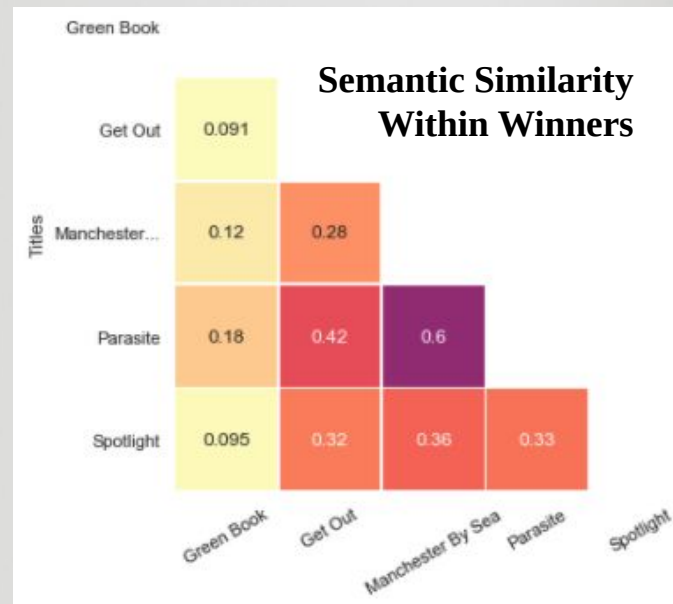
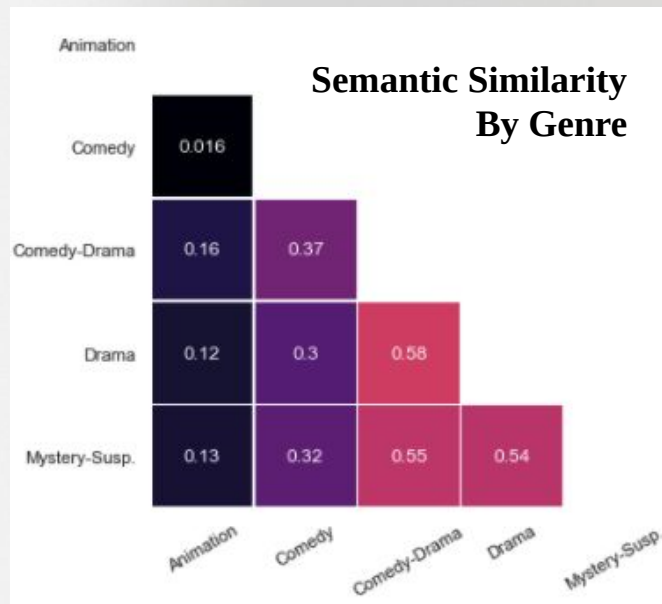




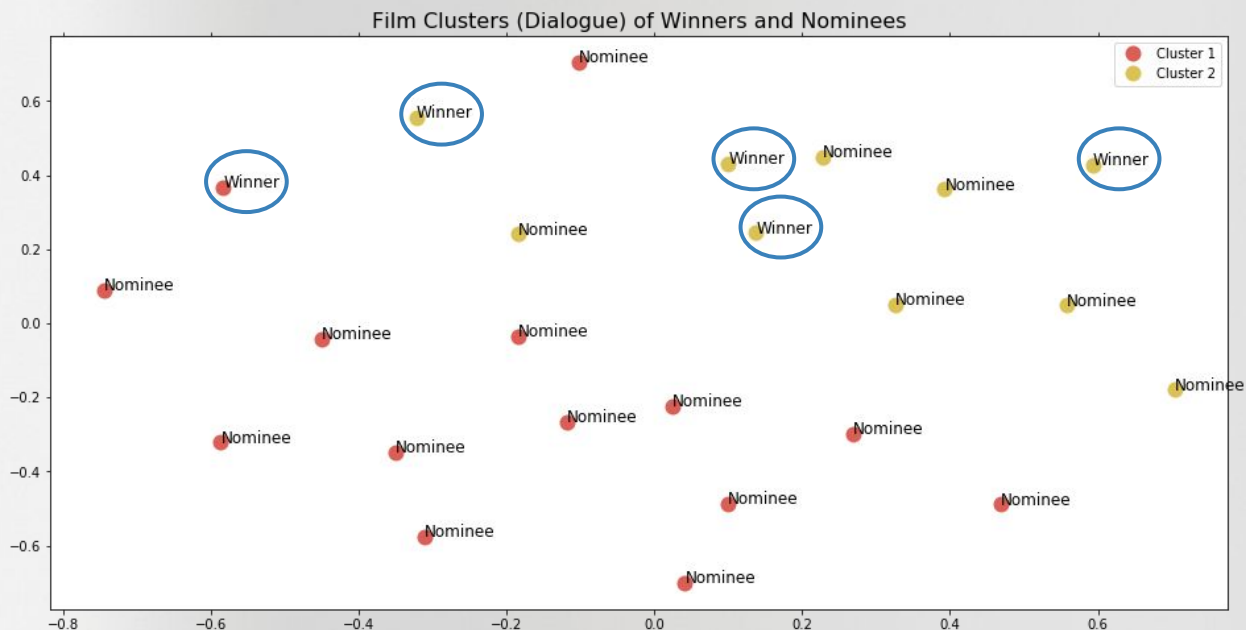
Topic Modeling - For All Scripts - By Dialogue

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10
0	beat	dont	gon	dick	lady	charlie	los	joy	rose	lip
1	sir	youre	fucking	power	bird	new york	sofa	sadness	rod	piano
2	nod	thats	bank	president	mom	york	patio	memory	picture	christmas
3	sign	shes	fuck	mary	dad	theater	doctor	island	brooklyn	glance
4	pie	david	beth	mike	suddenly	mom	hall	dad	driver	record
5	egg	didnt	god	agent	nod	fucking	pie	thats	dont	gon
6	blood	whats	wan	war	shall	lawyer	van	dont	lawn	george
7	grab	doesnt	cube	american	horse	hesitates	dice	fear	dining	hell
8	german	dont know	pause	office	rabbit	pause	child	mom	richard	stage
9	corridor	ill	cop	state	dress	envelope	nurse	train	jim	letter
10	glass	william	bitch	beat	college	divorce	street	youre	ing room	chief
11	suddenly	theyre	kinda	judge	happy	court	old man	anger	begin	road
12	record	gon	church	secretary	war	costume	stair	shes	chair	eat
13	men	joe	police	government	hurt	fuck	mike	mom dad	detective	beat
14	push	mom	leaning	soldier	mail	money	baby	happy	thats	stone

Semantic Similarity using Cosine



K-Means Clustering



Areas of Opportunity

- ◆ Parts of Speech Tagging
- ◆ Larger sample size
- ◆ Classification model

THANKS!

Any questions?

ar2579a@american.edu

&

as8273a@american.edu

