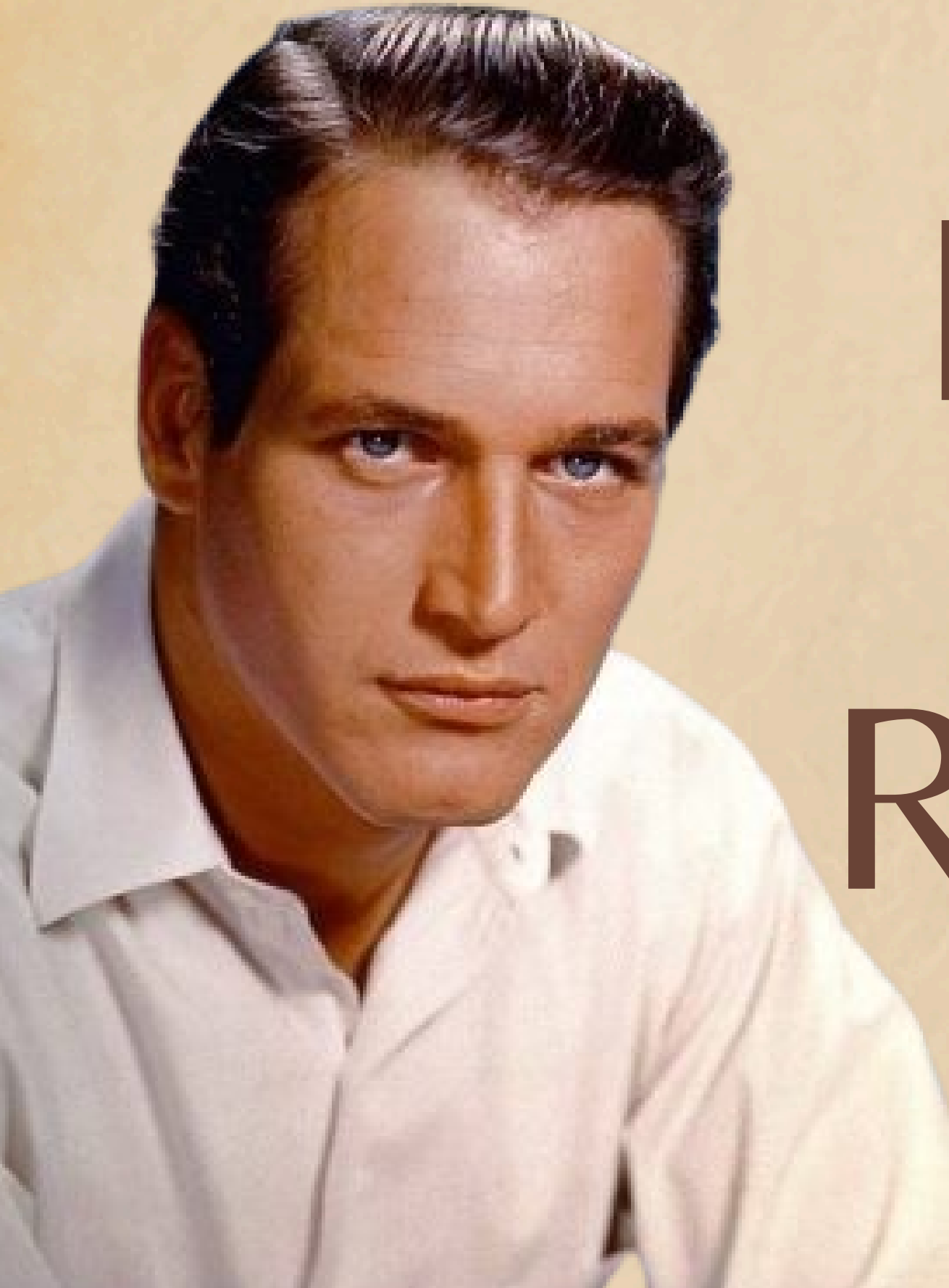

Matteo and Barbara

BIAS IN FILM RATINGS

Is the screen neutral?



DATASET PIPELINE



INITIAL DATASETS



UNIVERSITY OF MINNESOTA

MovieLens 1M movie ratings (2003)

- vote by users
- information of voter

MovieLens 32M movie ratings (2024)

- links with IMDb



IMDb Non-Commercial Datasets

- information about actors

REFINED DATASET

Actors

	actor_id	actor_name	actor_gender	actor_birth_year	actor_death_year
1	158	Tom Hanks	M	1956	NA
2	741	Tim Allen	M	1953	NA
3	725543	Don Rickles	M	1926	2017
4	1815	Jim Varney	M	1949	2000
5	1728	Wallace Shawn	M	1943	NA

Cast

	movie_id	actor_id	ordering
1	1	158	1
2	1	741	2
3	1	725543	3
4	1	1815	4
5	1	1728	5

Movies

	movie_id	movie_title	movie_year	movie_genres
1	1	Toy Story	1995	Animation Children's Comedy
2	2	Jumanji	1995	Adventure Children's Fantasy
3	3	Grumpier Old Men	1995	Comedy Romance
4	4	Waiting to Exhale	1995	Comedy Drama
5	5	Father of the Bride Part II	1995	Comedy

Voters

	voter_id	voter_gender	voter_age	voter_occupation	voter_zip_code
1	1	F	1	10	48067
2	6	F	50	9	55117
3	8	M	25	12	11413
4	9	M	25	17	61614
5	10	F	35	1	95370

Ratings

	voter_id	movie_id	rating
1	1	1	5
2	6	1	4
3	8	1	4
4	9	1	5
5	10	1	5



QUESTIONS

→ OCCUPATION BIAS

→ GENRES DIFFERENCES

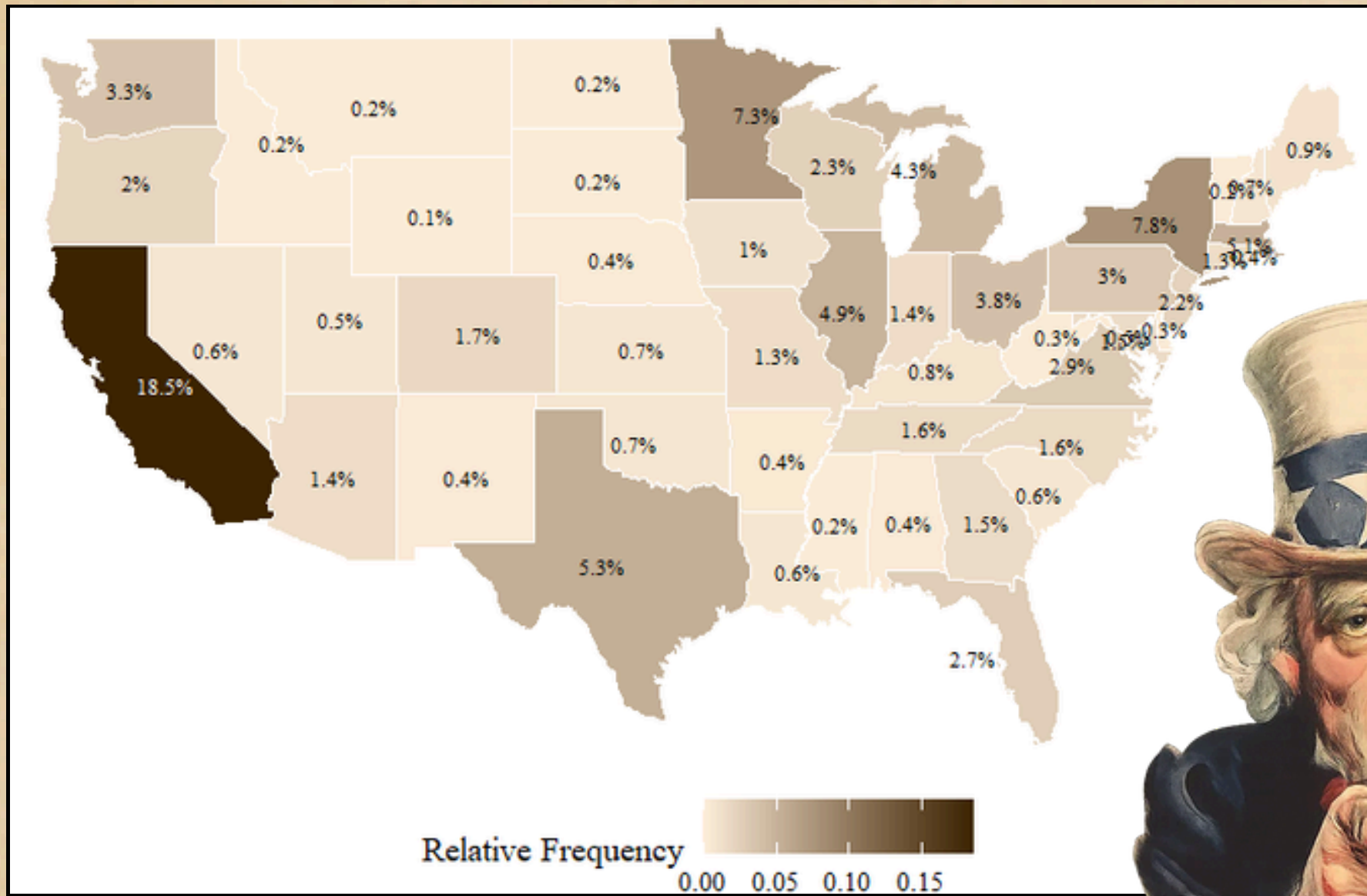
→ GENDER BIAS

→ GENDER STAR BIAS

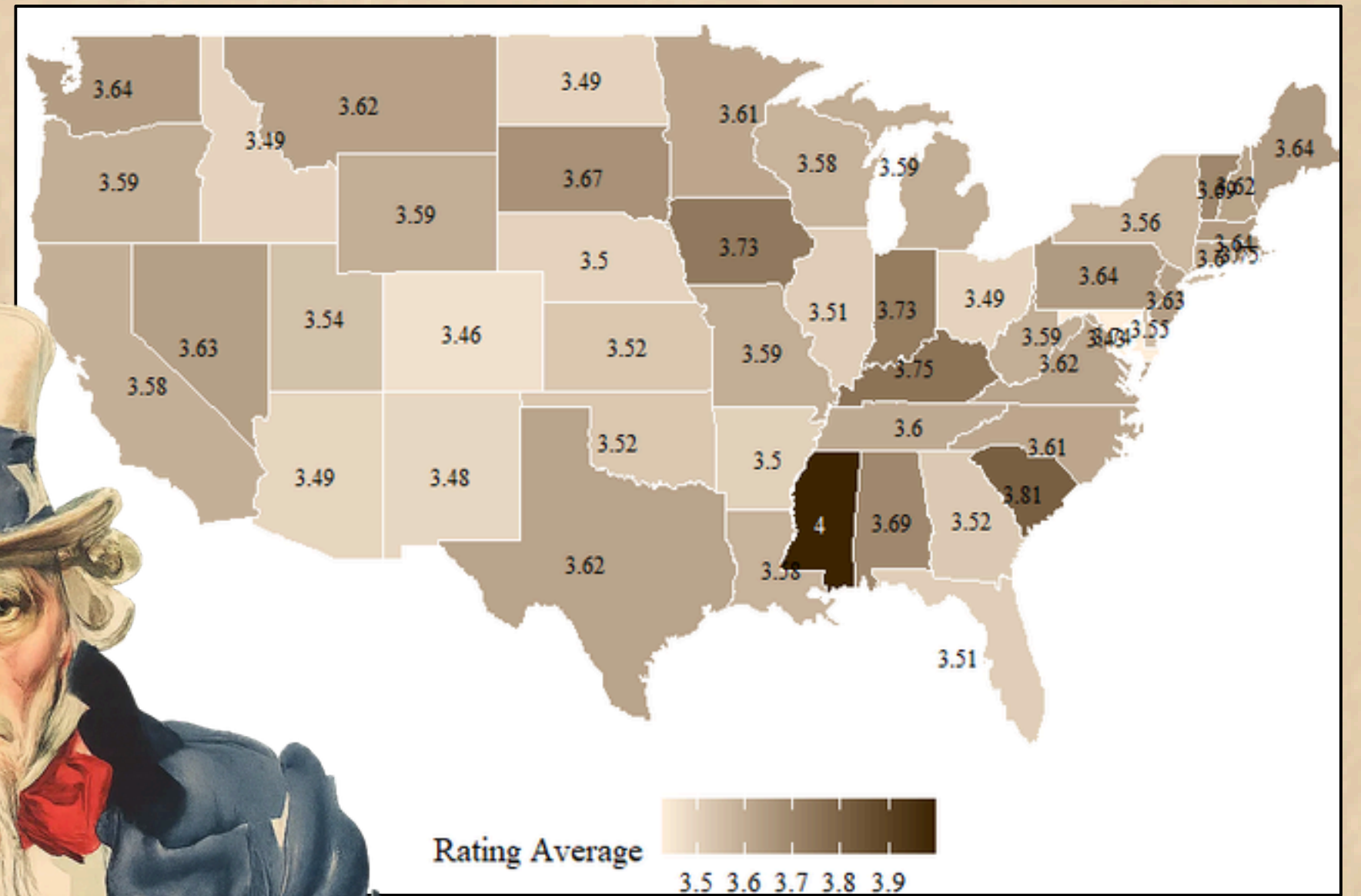


SOME EXPLANATORY MAPS

RELATIVE FREQUENCY PER STATE

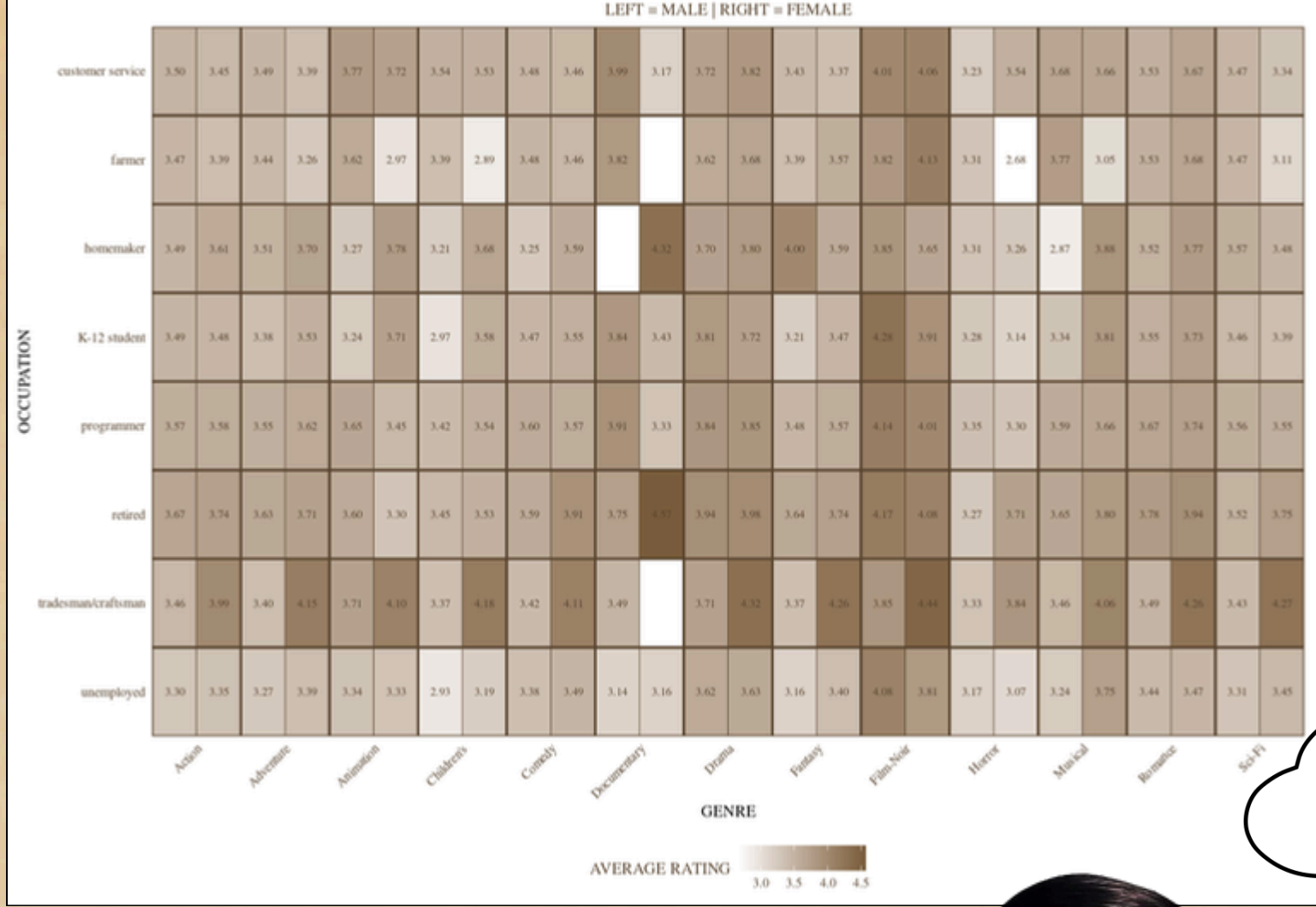


AVERAGE RATING PER STATE

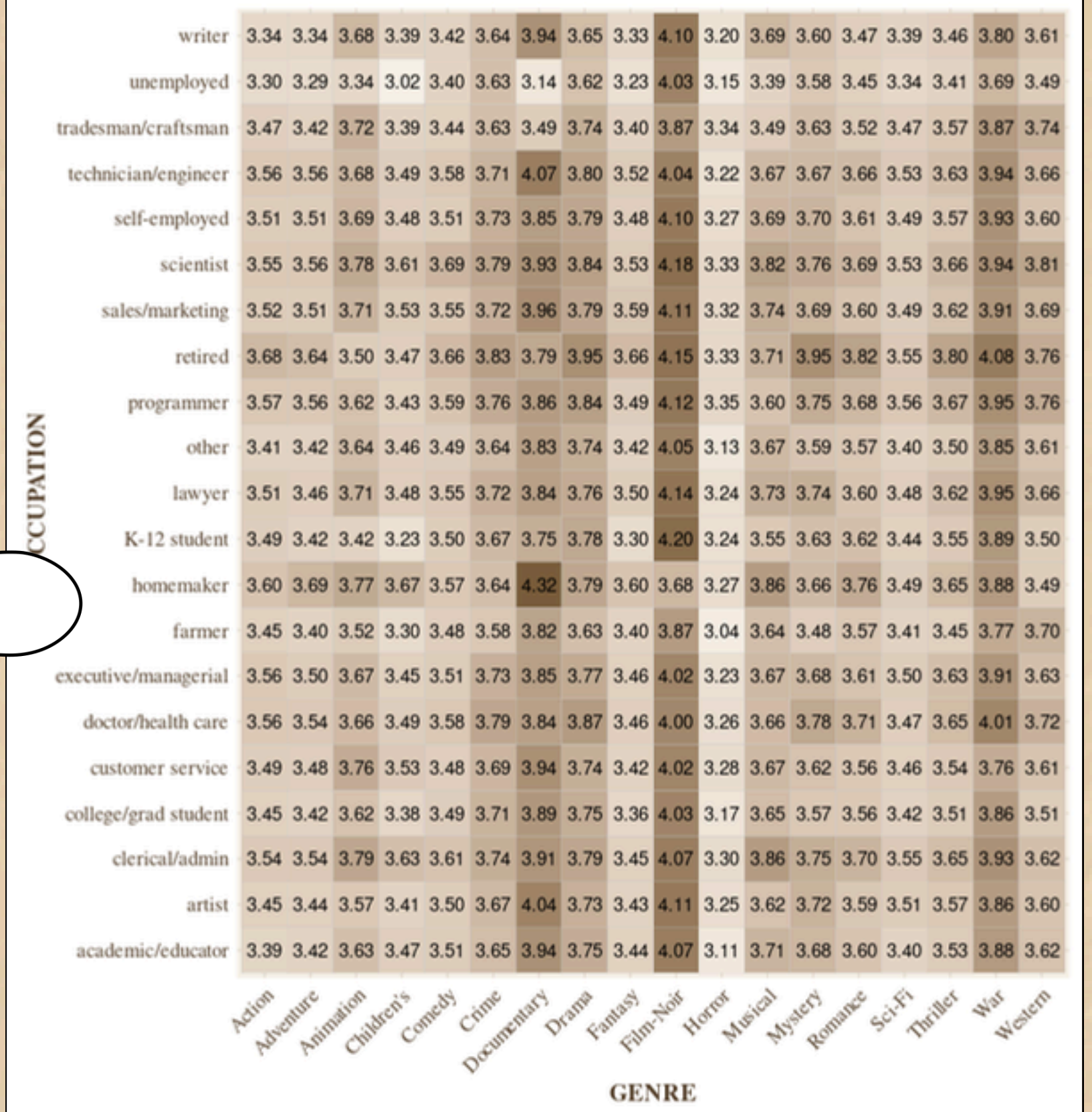


AVERAGE RATING BY GENRE, OCCUPATION AND GENDER

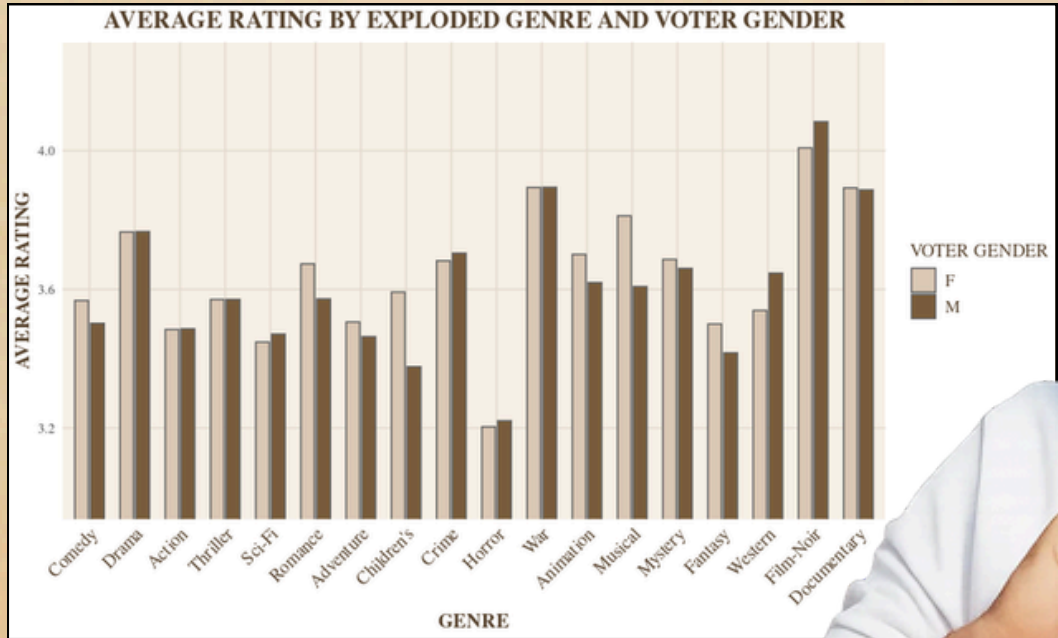
AVERAGE RATING BY GENDER AND OCCUPATION (DIFFERENCE ≥ 0.5)



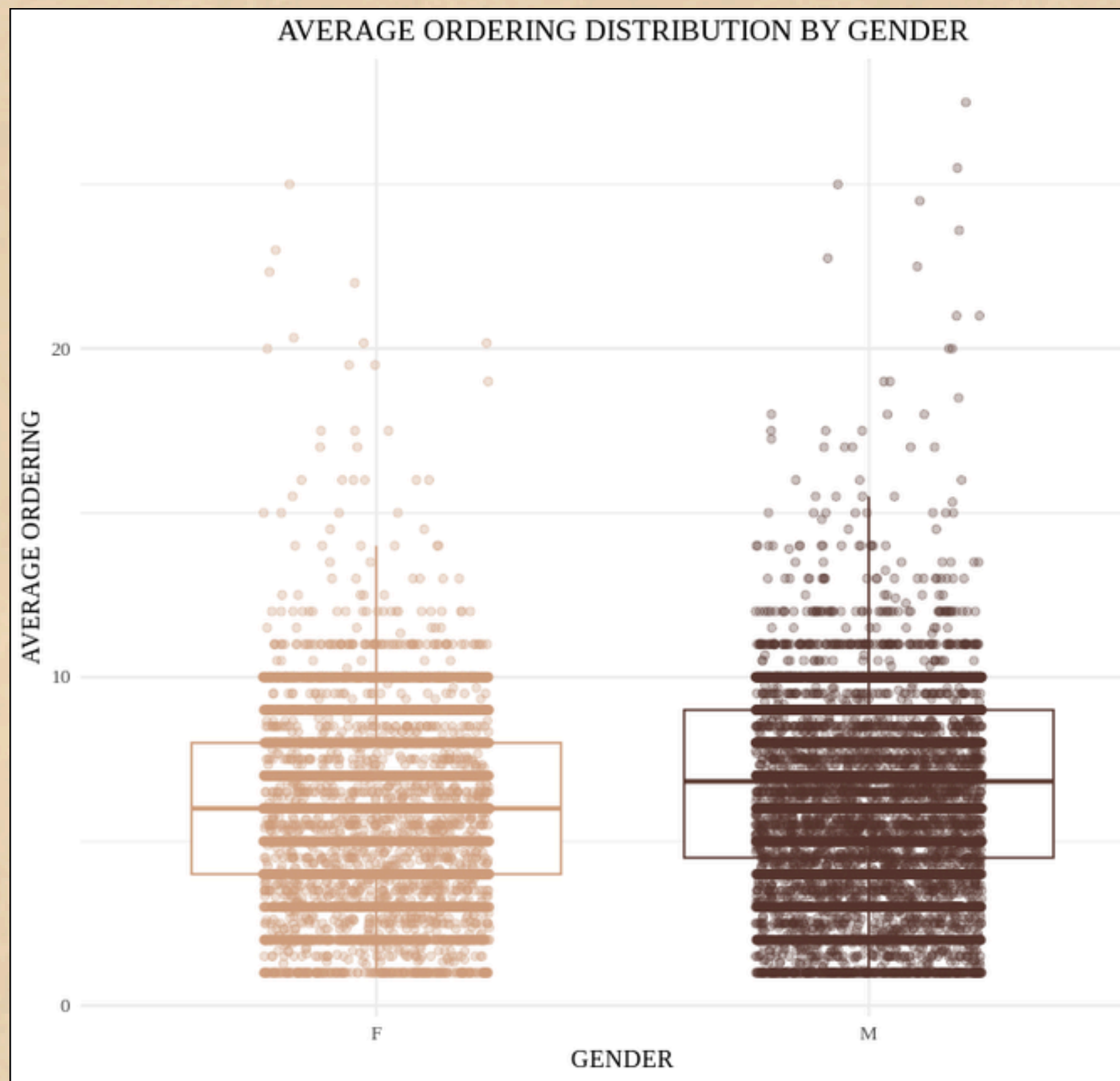
AVERAGE RATING BY GENRE AND OCCUPATION



WERE YOU ABLE TO SEGMENT THE MARKET?



HOW ACTOR AND ACTRESS DISTRIBUTED IN CREDITS

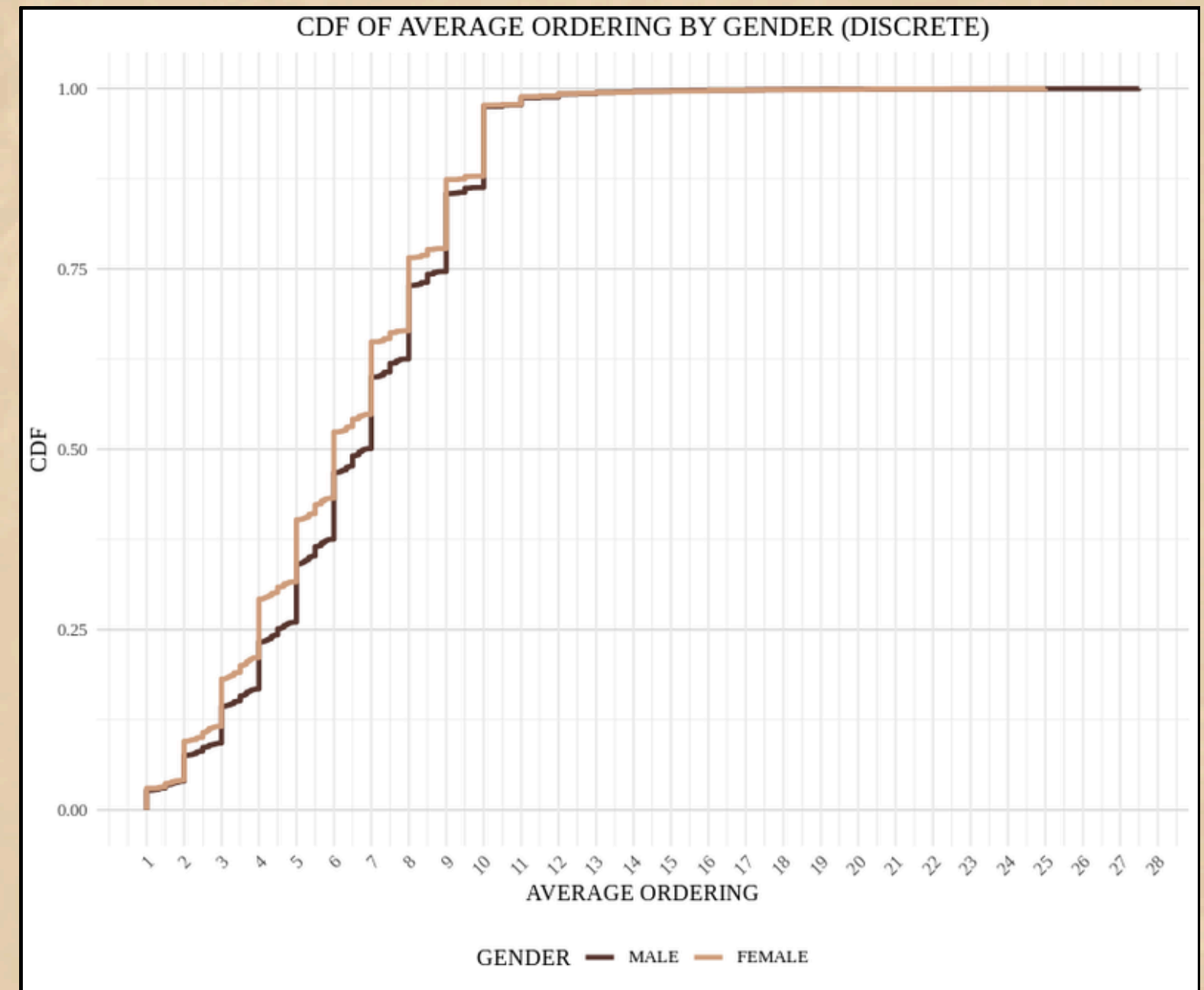
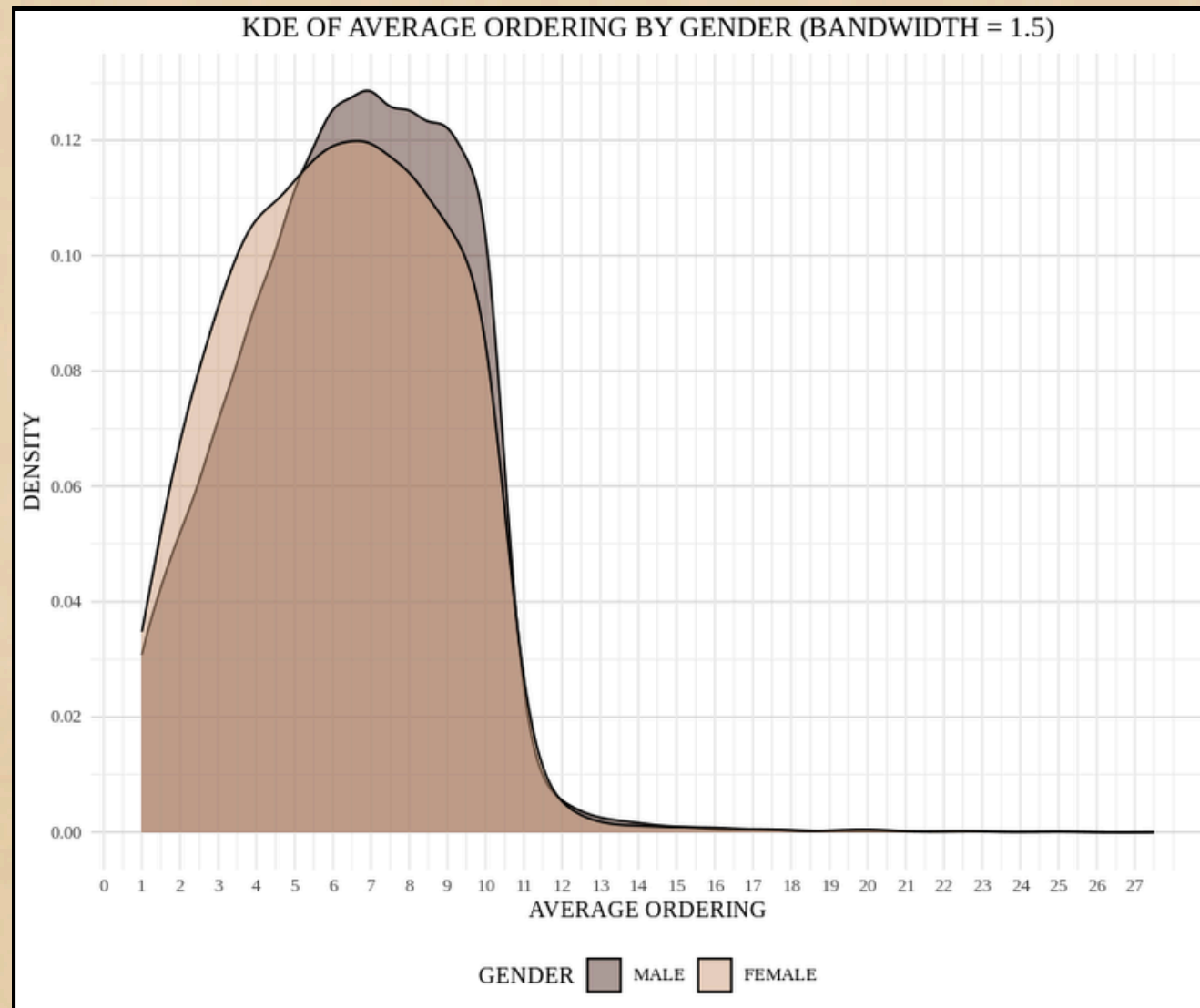




FIND STARS

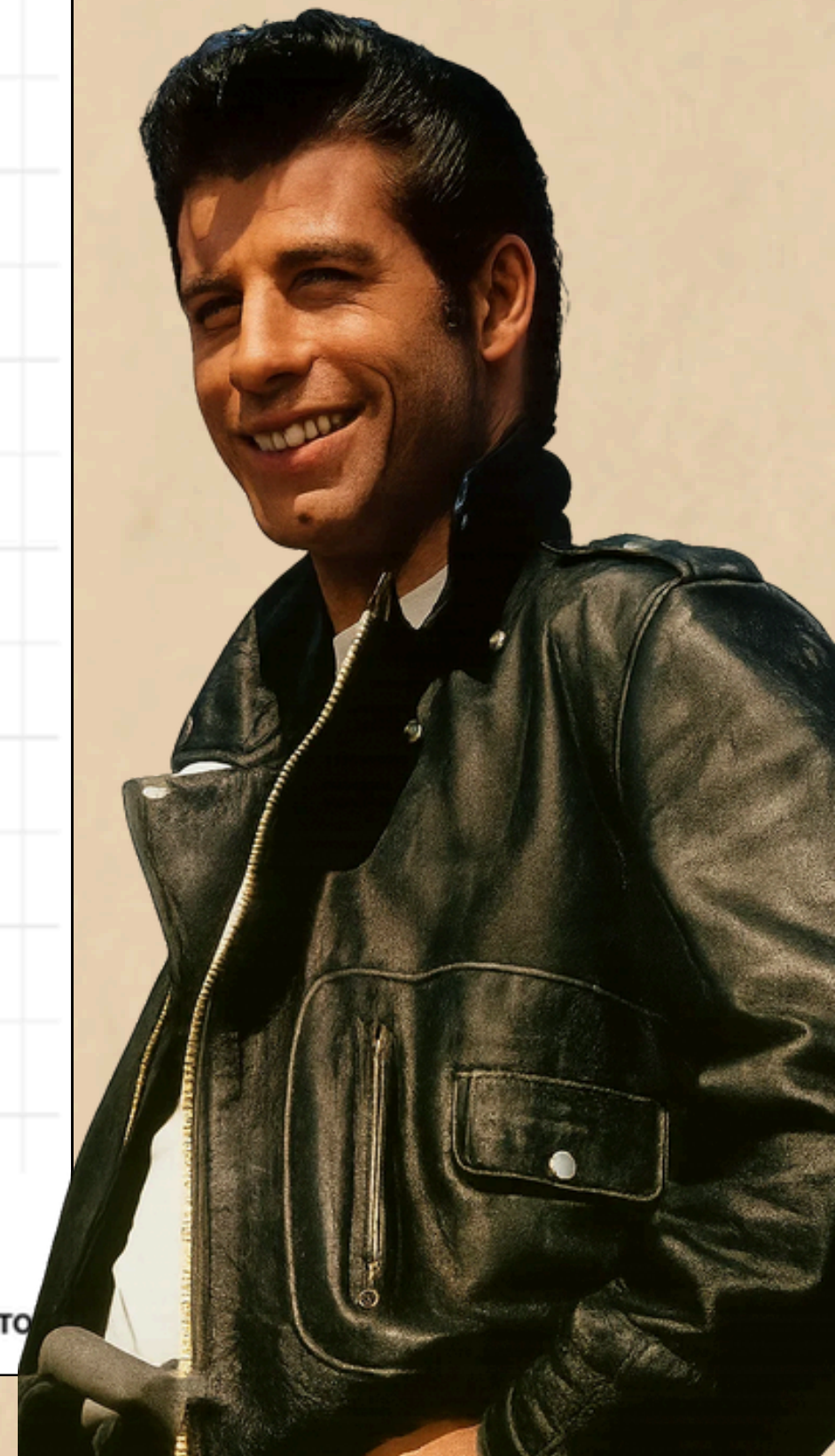
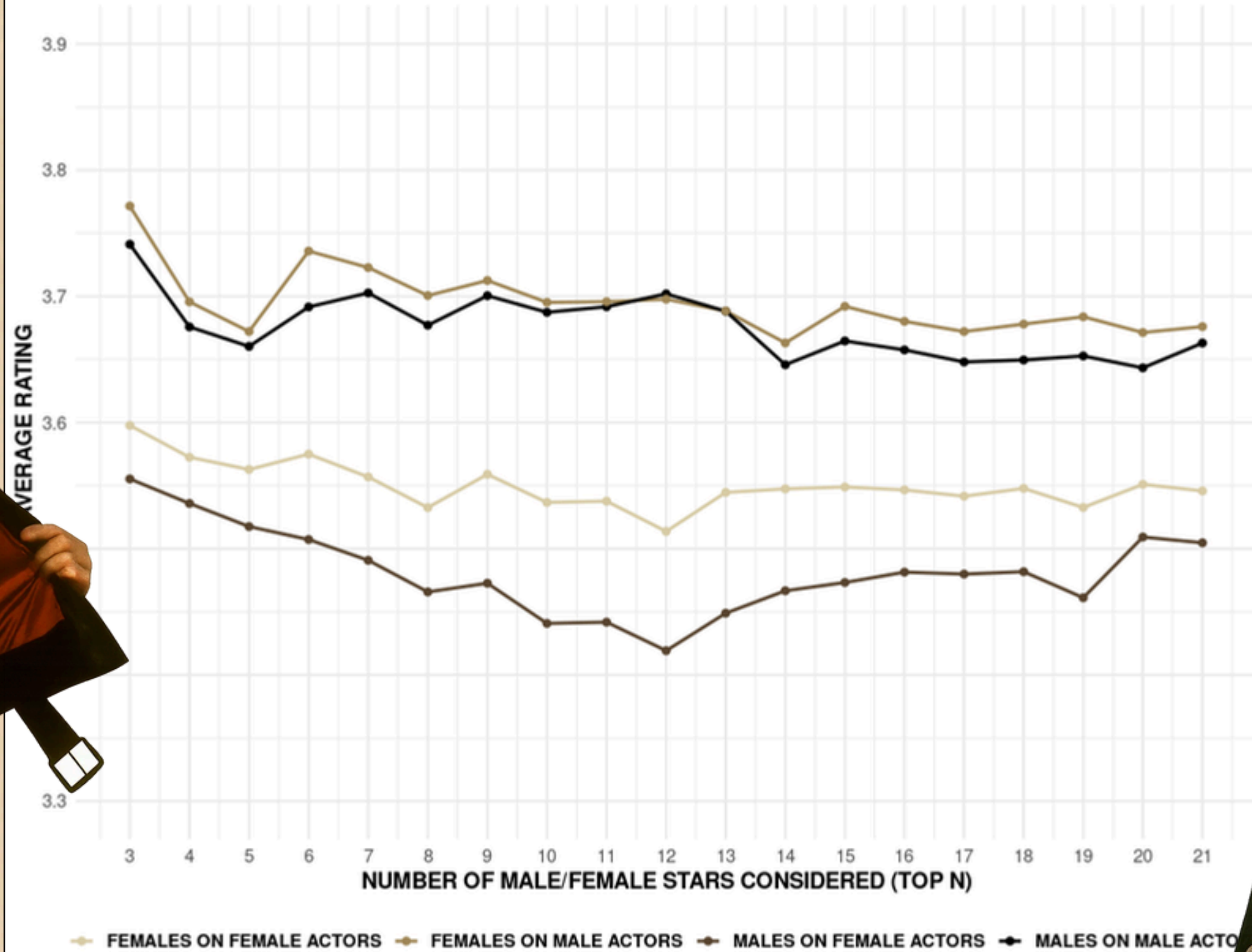
(WONDERUL) HEURISTIC
to determine whether an actor/actress is a star

$$\text{score}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \text{ordering}_j^{(i)} - \frac{\log(n_i)}{n_i}$$



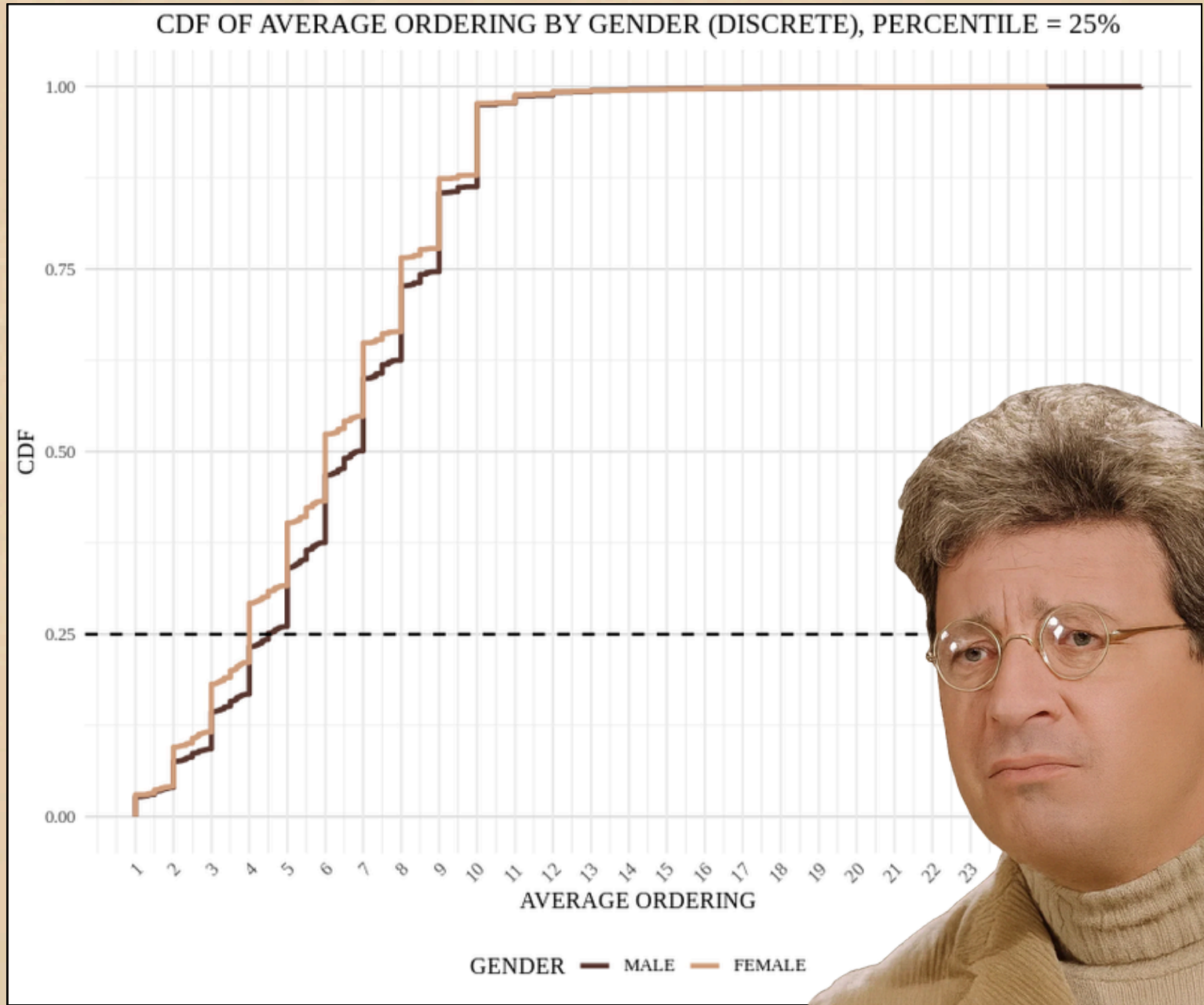
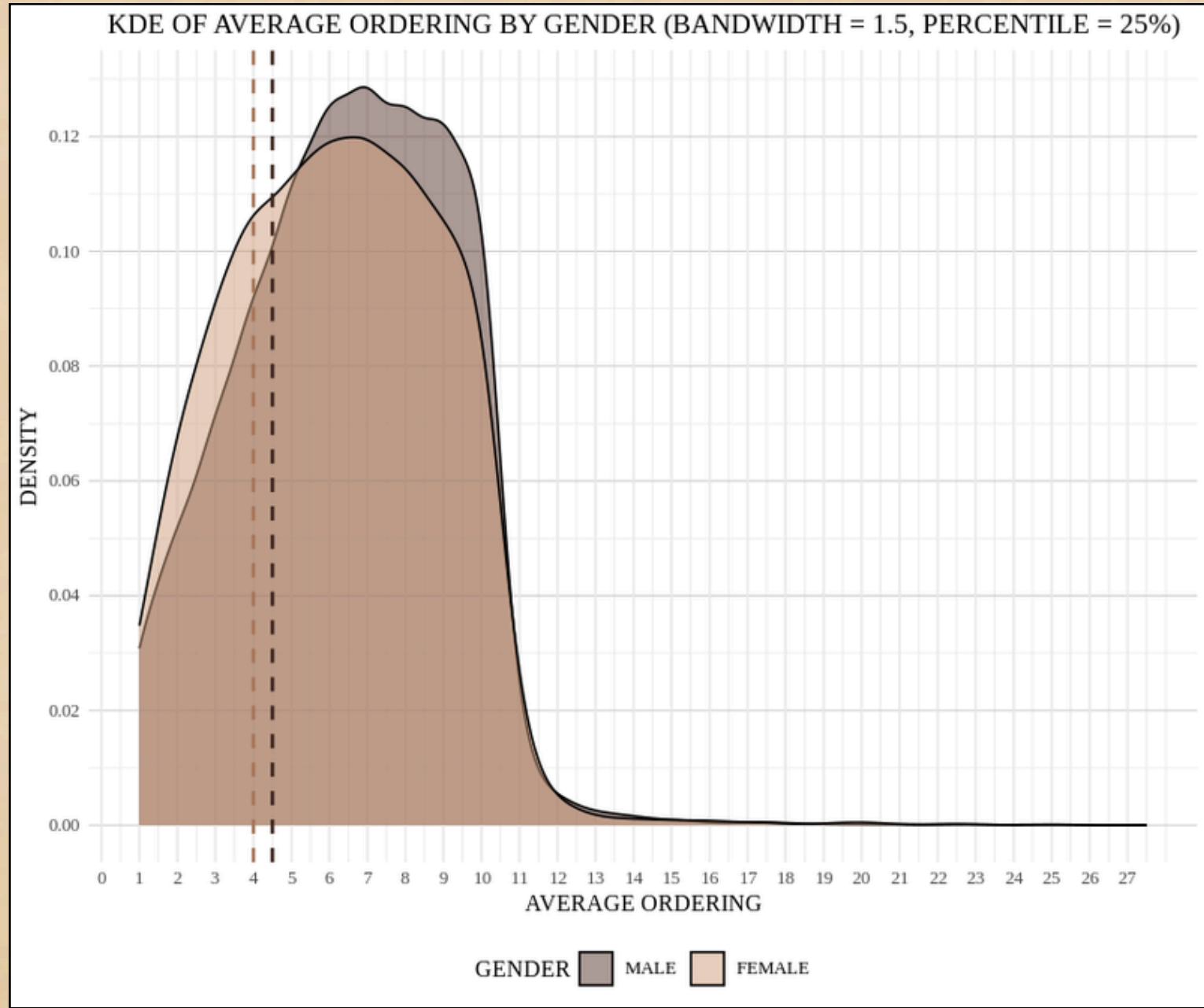
A TREND

AVERAGE RATINGS BY VOTER GENDER AND STAR GENDER (MIN 12 FILMS)



A TRADE - OFF

	FALSE	TRUE
M	9557	3209
F	4835	1999

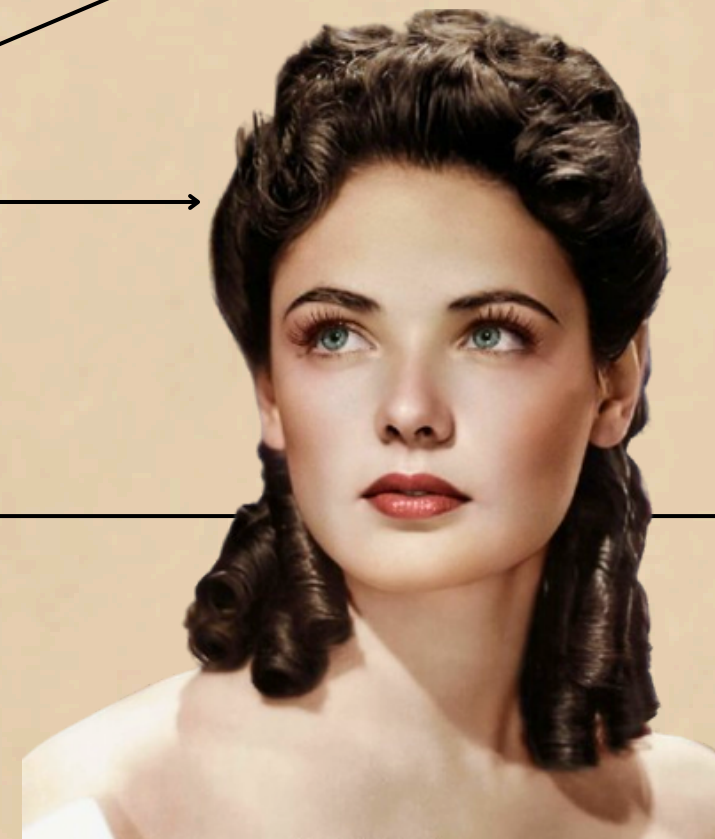


EXAMPLES OF STARS

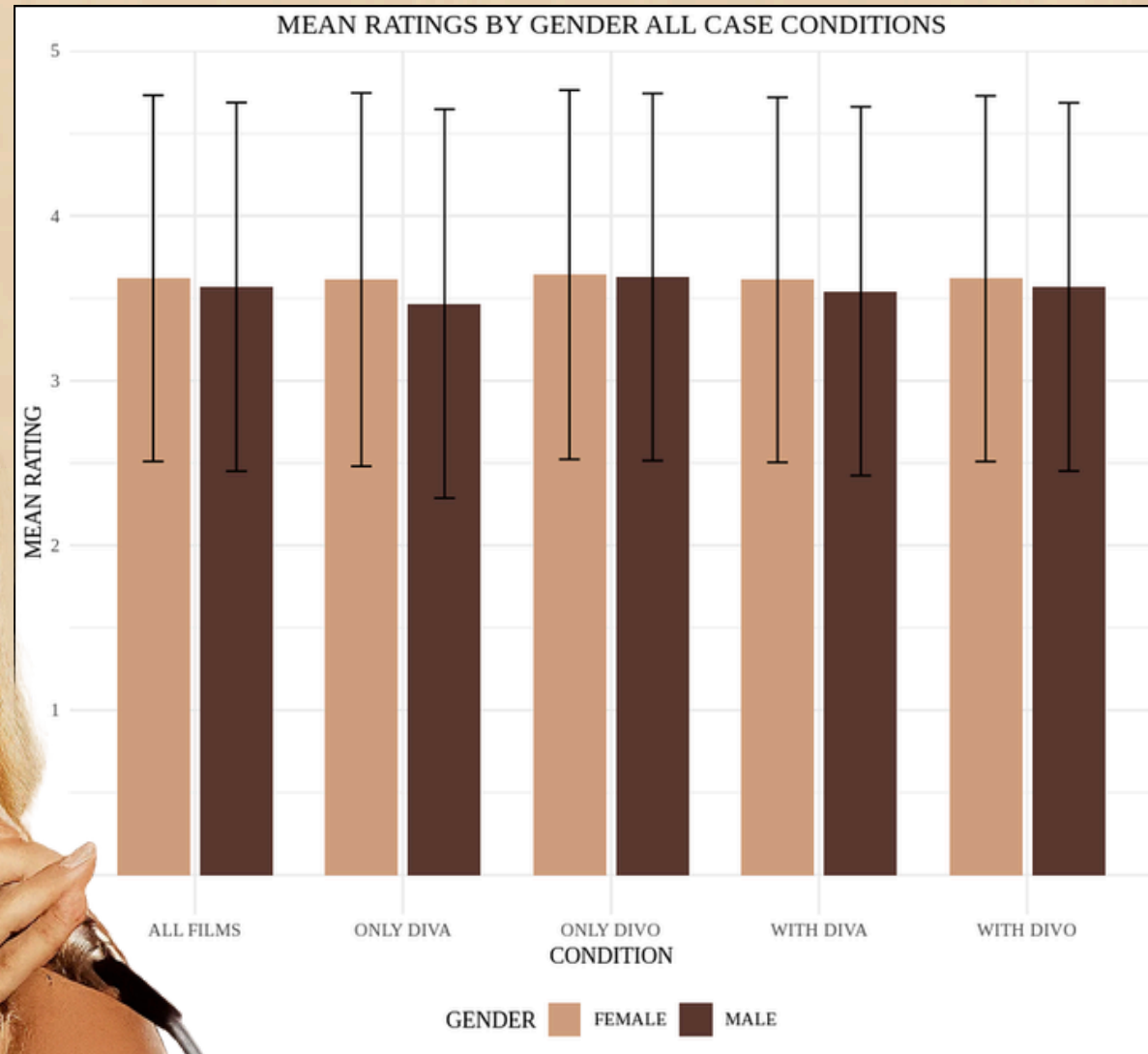
	actor_id	actor_name	avg_ordering
1	4	John Belushi	1
2	7	Humphrey Bogart	1
3	42	Alan Ladd	1
4	50	Groucho Marx	1
5	61	Tyrone Power	1
6	62	Elvis Presley	1
7	140	Michael Douglas	1
8	154	Mel Gibson	1



	actor_id	actor_name	avg_ordering
1	23	Judy Garland	1
2	73	Shirley Temple	1
3	74	Gene Tierney	1
4	97	Pamela Anderson	1
5	486	Heather Langenkamp	1
6	659	Barbra Streisand	1
7	742	June Allyson	1
8	845	Tallulah Bankhead	1



IS THERE ANY BIAS BY GENDER?



COHEN'S D - EFFECT SIZE

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

FILM WITH AT LEAST A STAR OF GENDER

MALE

FEMALE

Welch Two Sample t-test

data: rating by voter_gender
 t = 19.015, df = 401766, p-value < 2.2e-16
 alternative hypothesis: true difference in means between group F and group M is not equal to 0
 95 percent confidence interval:
 0.04492606 0.05525196
 sample estimates:
 mean in group F mean in group M
 3.618616 3.568527

Welch Two Sample t-test

data: rating by voter_gender
 t = 23.118, df = 333341, p-value < 2.2e-16
 alternative hypothesis: true difference in means between group F and group M is not equal to 0
 95 percent confidence interval:
 0.06280669 0.07444314
 sample estimates:
 mean in group F mean in group M
 3.611539 3.542914

Cohen's d

d estimate: 0.04491305 (negligible)
 95 percent confidence interval:
 lower upper
 0.04026839 0.04955772

Cohen's d

d estimate: 0.06148657 (negligible)
 95 percent confidence interval:
 lower upper
 0.05624729 0.06672584

FILM WITH ONLY STARS OF GENDER

MALE

FEMALE

Welch Two Sample t-test

data: rating by voter_gender
 t = 2.5843, df = 84884, p-value = 0.00976
 alternative hypothesis: true difference in means between group F and group M is not equal to 0
 95 percent confidence interval:
 0.003376331 0.024576399
 sample estimates:
 mean in group F mean in group M
 3.642497 3.628521

Welch Two Sample t-test

data: rating by voter_gender
 t = 9.0097, df = 15193, p-value < 2.2e-16
 alternative hypothesis: true difference in means between group F and group M is not equal to 0
 95 percent confidence interval:
 0.1139474 0.1773127
 sample estimates:
 mean in group F mean in group M
 3.612812 3.467182

Cohen's d

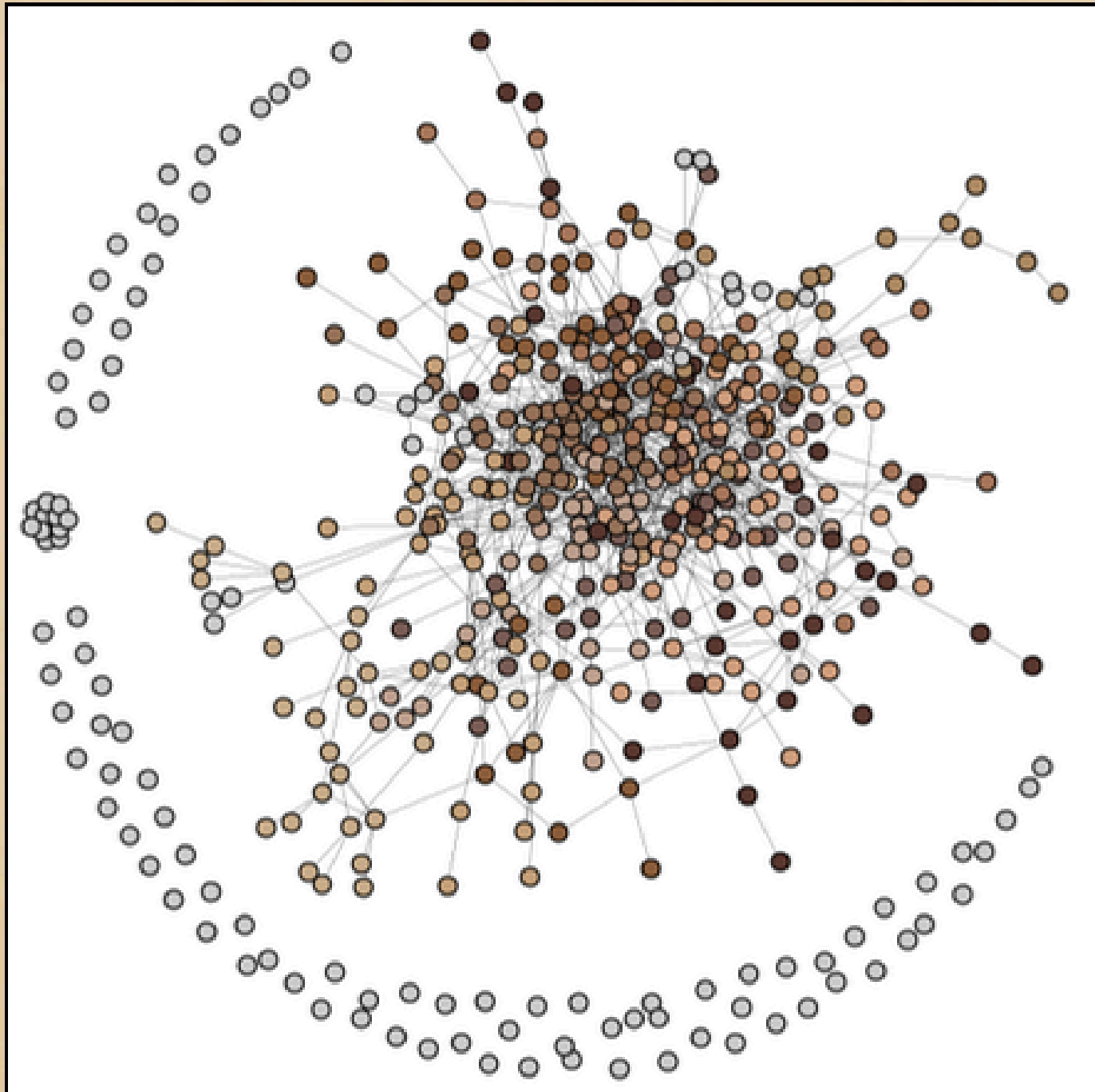
d estimate: 0.01252704 (negligible)
 95 percent confidence interval:
 lower upper
 0.003055613 0.021998467

Cohen's d

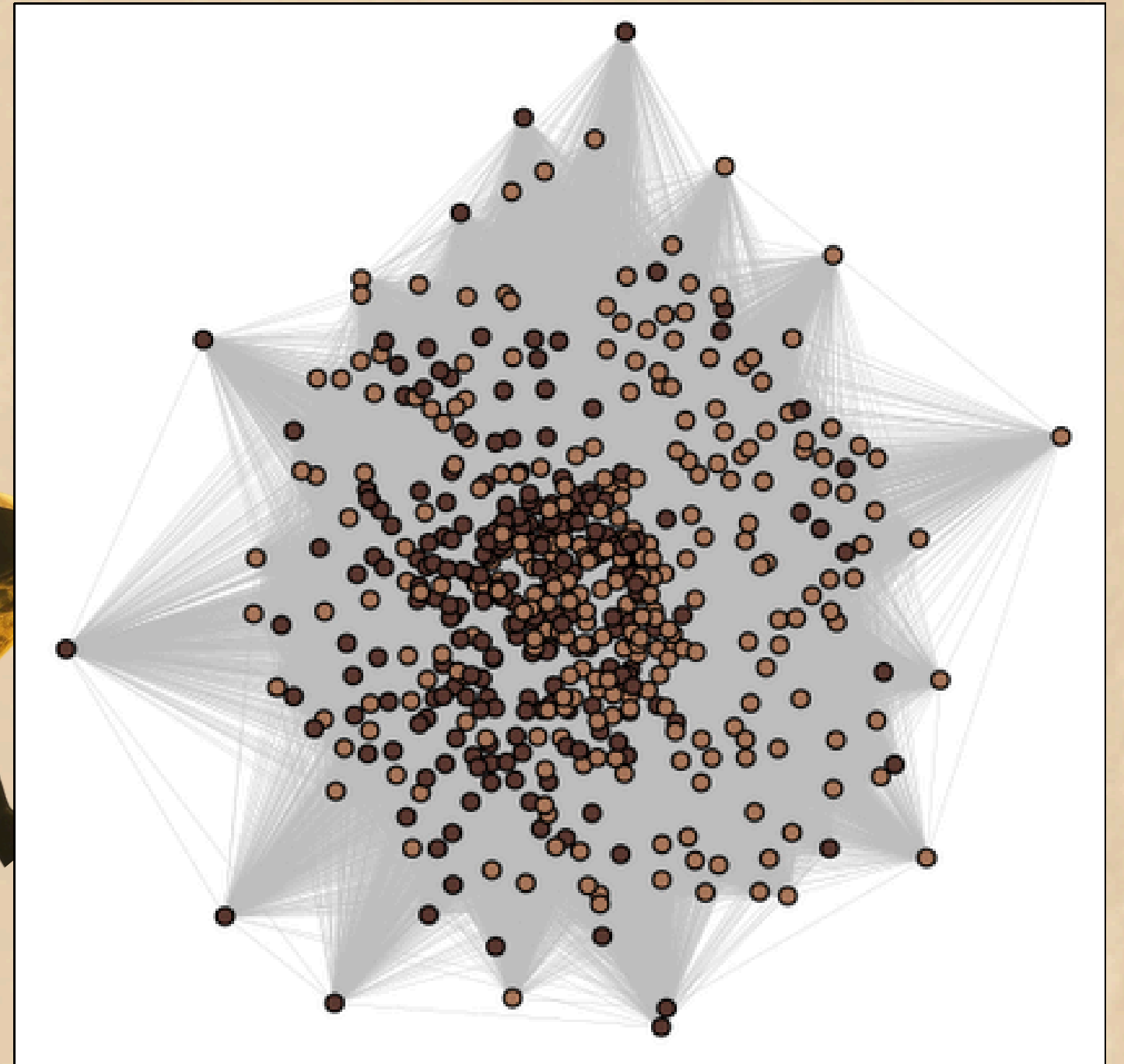
d estimate: 0.1250164 (negligible)
 95 percent confidence interval:
 lower upper
 0.09739761 0.15263516

SOME NETWORKS ON SAMPLES

FAST GREEDY CLUSTERING
ON A RANDOM SAMPLE OF 500 ACTORS



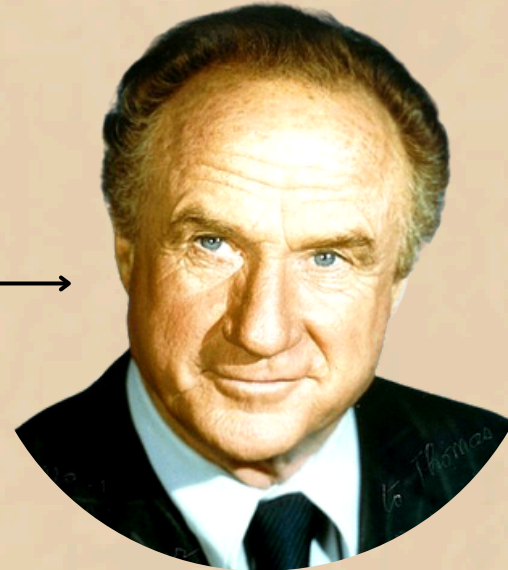
FAST GREEDY CLUSTERING
ON A RANDOM SAMPLE OF 500 VOTERS



NETWORK METRICS AND STARS

TOP 10 BETWEENNESS ACTORS

	actor_name	bet	star
1	Jack Nicholson	3876082	TRUE
2	Robert De Niro	3493204	TRUE
3	Sean Connery	2995169	TRUE
4	Jack Lemmon	2560896	TRUE
5	Gregory Peck	2356768	TRUE
6	Jack Warden	2339106	FALSE
7	John Malkovich	2216569	TRUE
8	Gene Hackman	2057926	TRUE
9	Demi Moore	2042339	TRUE
10	Harvey Keitel	2041292	TRUE

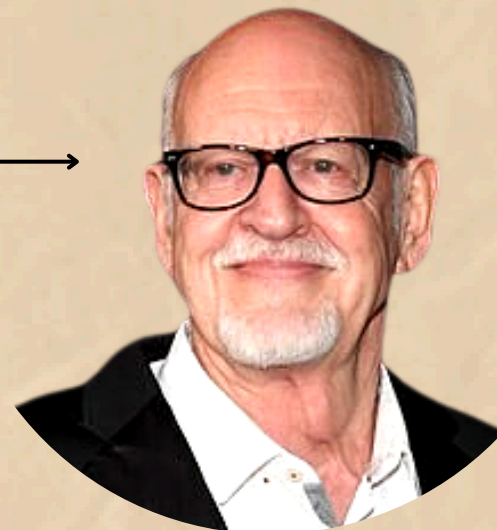


TOP 10 CLOSENESS ACTORS

	actor_name	clos	star
1	Louise Smith	1	TRUE
2	Judith Fyfe	1	TRUE
3	David Riva	1	TRUE
4	Leopold Kozlowski	1	TRUE
5	Timmy Prairie	1	FALSE
6	Juan Munoz	1	FALSE
7	Jerry Archbold	1	TRUE
8	Desiree Densiti	1	FALSE
9	John McGarr	1	TRUE
10	Alan Rosenberg	1	FALSE

TOP 10 DEGREE ACTORS

	actor_name	deg	star
2	Robert De Niro	35	TRUE
3	Frank Oz	27	FALSE
4	Samuel L. Jackson	27	TRUE
5	Robin Williams	26	TRUE
6	Demi Moore	26	TRUE
7	Jack Nicholson	26	TRUE
8	Sean Connery	26	TRUE
9	Whoopi Goldberg	26	TRUE
10	Gene Hackman	25	TRUE
11	Bruce Willis	25	TRUE



TOP 10 EIGN. ACTORS

	actor_name	eig	star
1	Frank Oz	1.0000000	FALSE
2	Jerry Nelson	0.9953334	FALSE
3	Dave Goelz	0.8951082	FALSE
4	Steve Whitmire	0.5818869	FALSE
5	Jim Henson	0.5441778	TRUE
6	Richard Hunt	0.5359189	FALSE
7	Bill Barretta	0.2686210	FALSE
8	Kevin Clash	0.2377687	FALSE
9	Robert Tygner	0.1359383	FALSE
10	David Rudman	0.1353919	FALSE



THANK YOU FOR
YOU ATTENTION

