

# گزارش: بررسی فیلم و ژانر

وحید رمضانی

## سوال پژوهشی:

1. تولید فیلم در سال های مختلف  
a. چه توزیعی دارد؟  
b. همچنین بر اساس ژانر
2. کدام(ها) ژانرهای فیلم  
a. بیشتر از بقیه تولید میشوند؟  
b. کدام(ها) پرفدارترین است؟
3. آیا توزیع امتیاز فیلمها در سالهای مختلف نظم خاصی دارد؟
4. برای یک فیلم پرامتیاز، نظر تماشاگران چگونه تغییر کرده است؟
5. بیشترین امتیاز - از نظر تماشاگران - از میان فیلمهای ساخته شده را کدام فیلمها دارند؟

## متغیرها

متغیر پیوسته: امتیاز برای فیلم یا ژانر

متغیر پیوسته: زمان برای امتیاز(رای) تماشاگران

متغیر گسسته چند رده ای: سالها، از 1997 تا 2015

## منبع آمار

MovieLens.org

## نرم افزار مورد استفاده

MS Office (2016) Excel

## تفسیر

برای دست یافتن به پاسخ‌ها، به تناسب سوال، از نمودارهای مختلفی استفاده شد: میله‌ای (bar)، جعبه‌ای (box)، پراکنش (scatter)، حبابی (bubble)، سطح (area)، دایره‌ای (pie) و خطی (line)

**توضیح 1:** یک فیلم می‌تواند چند ژانر داشته باشد. به عبارت دیگر فیلم‌ها بر اساس ژانر افراز نشده‌اند، بلکه برچسب خورده‌اند.

**توضیح 2:** آمار سال 2015 کامل نیست، بنابراین روی آن بحث نمی‌کنیم؛ آمار مربوط به 1997 تا 2014 است.

به ترتیب به تفسیر و پاسخ دادن به سوالات مطرح شده می‌پردازیم:

### تفسیر سالانه تولید فیلم‌ها (سوال 1a)

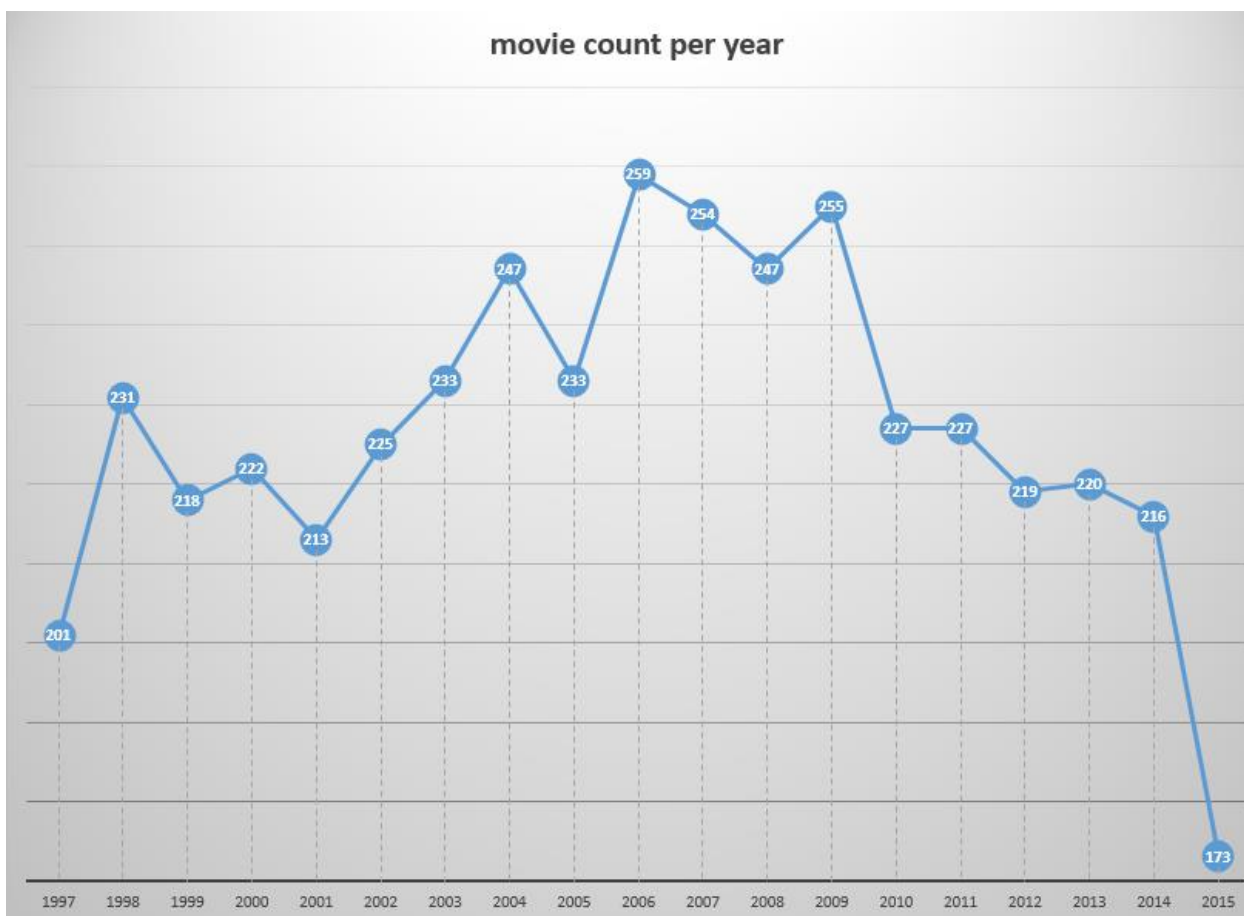


Figure 1: توزیع سالانه تولید فیلم

**پاسخ 1a:** طبق نمودار 1، مشاهده می‌شود که طی سال‌های متوالی از 1997 تا 2015، تولید فیلم نوسان‌هایی داشته است. بیشترین تعداد فیلم در سال 2006 (259) با یک صعود از سال قبلش و با رتبه دوم در سال 2010 (255) بوده است؛ لذا در سال‌های 2006 تا 2010 اوج ساخت فیلم را شاهدیم.

با این همه، توزیع منظم نیست؛ قدری شبیه توزیع heavy-tailed است.

پس از سال 2010، سقوط ناگهانی‌ای در تعداد فیلم وجود دارد!

## تفسیر سالانه تولید فیلم به تفکیک ژانر (سوال 1b)

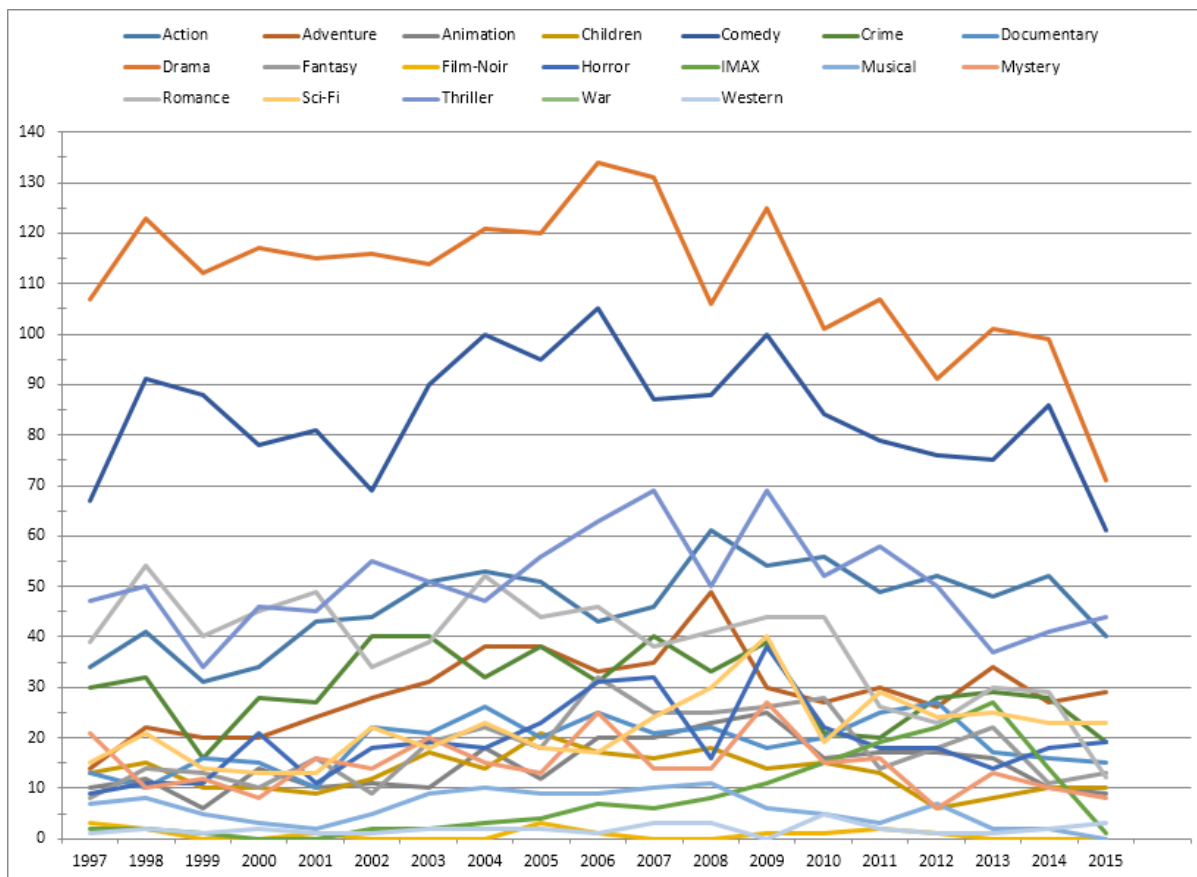


Figure 2: توزیع سالانه فیلم به تفکیک ژانر

پاسخ (اجمالی) 1b: بیشترین تولید مربوط به Drama و Comedy و کمترین تولید مربوط به Film-Noir و Western است:

Genre	Total	Mean	SD	cv	Median	min	Max
Drama	2111	111.10	14.79	0.13	114	71	134
Comedy	1600	84.21	11.69	0.14	86	61	105
Film-Noir	15	0.79	1.03	1.3	0	0	3
Western	35	1.84	1.12	0.61	2	0	5

پیوند: جدول کامل نمودار 2 در فایل tools.xlsx: "Genre Movie Count Table"

توضیح: بررسی بیشتر برای ژانرهای پرتیراژ در پاسخ به سوال 2a صورت میگیرد

بر اساس تیراژ

رشد	ثبات	نزول کم	نزول بیشتر	نزول چشمگیر	افت و خیز
IMAX	Animation Children Documentary Film-Noir Horror	Sci-Fi Mystery War	Romance Action Crime	Drama Comedy Thriller البته با این وجود هر	Adventure Fantasy

	3 ژانر پرطرفدارترین‌ها هستند			Musical Western	
--	------------------------------	--	--	-----------------	--

برای وضوح بیشتر، همچنین برای بررسی موشکافانه‌تر مسئله، 4 ژانر با بیشترین تولید را انتخاب میکنیم: Drama، Comedy، Thriller و Action.

چولگی این 4 ژانر بر اساس جدول:

Comedy: -0.16	Drama: -0.97
Action: -0.35	Thriller: 0.42

دو ژانر Comedy و Action دارای چولگی کمی هستند

توزیع این 4 ژانر مطابق نمودار 3 است:

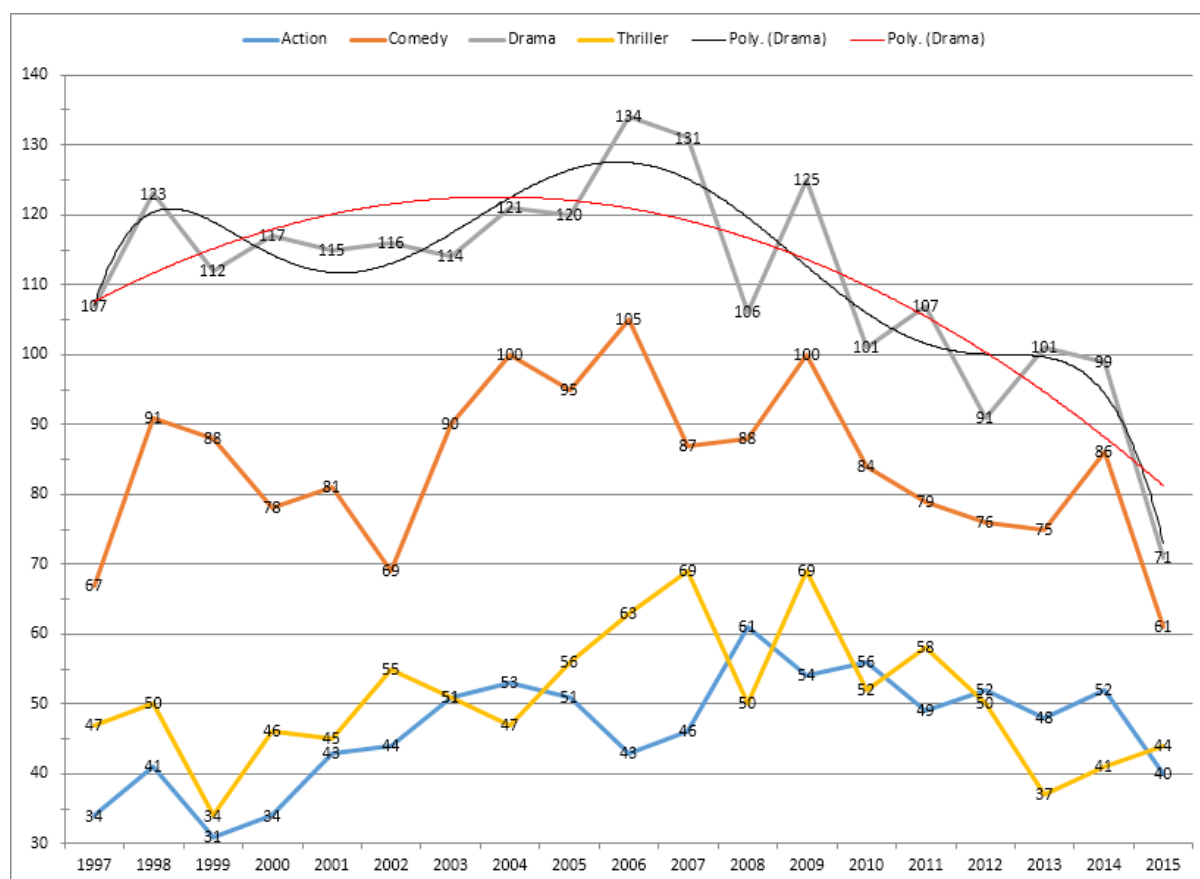


Figure 3

با توجه به نمودار 3 و خطوط چندجمله‌ای (polynomial) درجه دوم و ششم، میتوان نتیجه گرفت که در سال‌های اخیر، تولید ژانرهای پرتیراژ هم روند نزولی دارد! البته برای ژانر سوم و چهارم (به ترتیب Action و Thriller)، این شیب ملایم‌تر است، و حتی شاید در آینده رشد داشته باشد.

## بیشترین تولید به تفکیک ژانر (سوال 2a)

**پاسخ 2a:** نتیجه دیگری که با دقت در نمودار 3 عاید میشود این است که تولید ژانر اول (Drama) همیشه بیشتر از ژانر دوم (Comedy) بوده و تولید این دو با فاصله قابل ملاحظه‌ای از دو ژانر سوم و چهارم بیشتر بوده است.

Drama > Comedy >> Thriller ~ Action

## پرطرفدارترین ژانرها (سوال 2b)

برای این منظور، رای هر تماشاگر را برای یک یک فیلم‌هایی که دیده بود را جمع کردیم، و آن فیلم‌ها را ژانر بندی نمودیم. از روی داده‌ها جدول "Genre Vote Count Table" را ساختیم که نشانگر تعداد رای‌های برای هر ژانر، به تفکیک سال، میانه، میانگین، انحراف معیار و ضریب تغییرات آنها است.

بررسی همگی بر اساس مخاطب(رای)

رشد	ثبات	نزول کم	نزول شدید	افت و خیز
IMAX Action Adventure Fantasy		Musical	Romance War	Film-Noir Comedy Crime Fantasy Horror Drama Thriller Western

مورد قابل توجه! ژانر Film-Noir تنها در 3 سال طرفدار داشته: 1997، 1998 و 2005 و دست آخر در سال 2011 علاقه کمی به این ژانر وجود داشته.

**پاسخ 2b:** با دقت در میانه و میانگین (و انحراف معیار) تعداد رای‌ها برای هر ژانر، به ترتیب پرطرفدارترین ژانرها: Drama ، Comedy ، Action و Thriller اند!

این 4 ژانر همان دقیقا همان 4 ژانر پرتیراژند! و دقیقا با همان رتبه؛ دوباره

Drama >> Comedy > Action ~ Thriller

نتیجه نمودار 4 است:

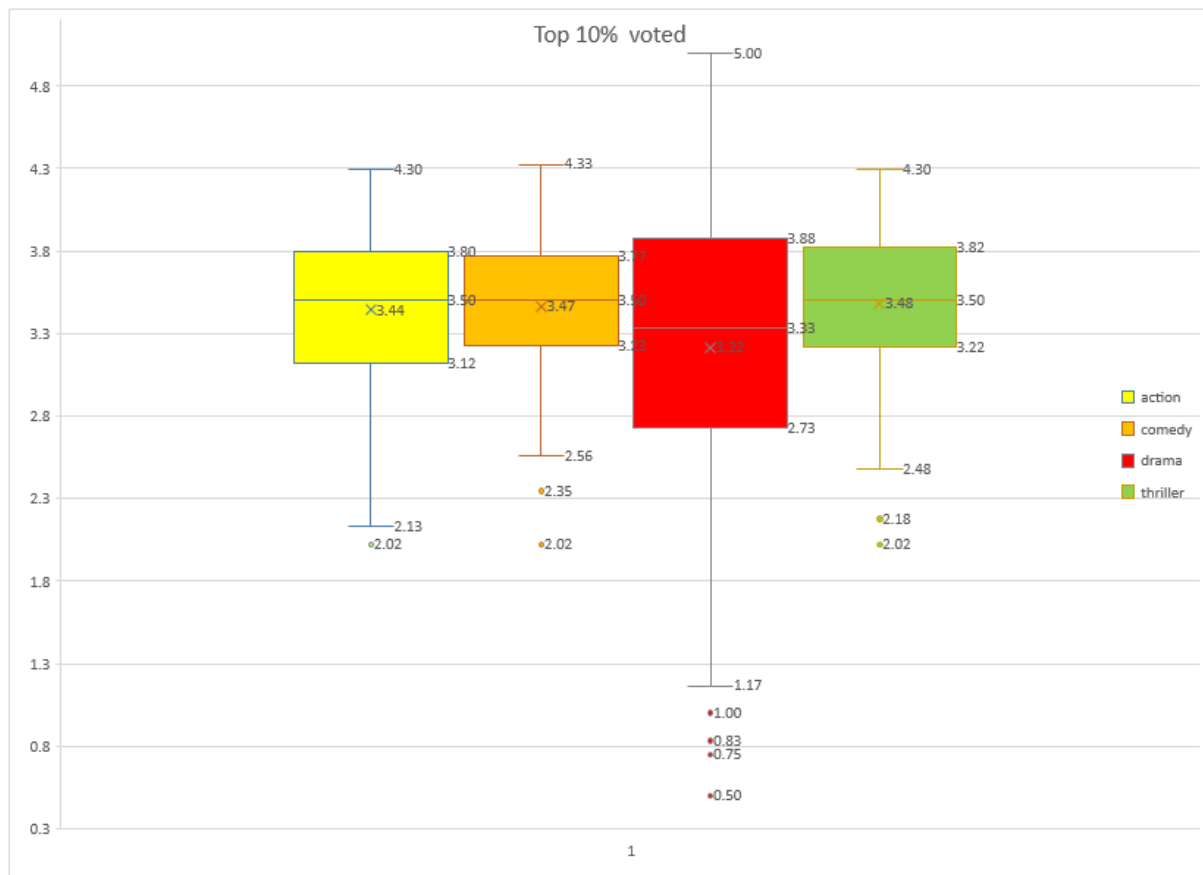


Figure 4: پرترفدارترین ژانرها در پرترفدارترین 10 درصد کل فیلمها؛ X نمایانگر میانگین است.

نتیجه‌های نمودار 4 (امتیاز 4 ژانر پرترفدار)

- با اینکه Drama بیشترین تعداد طرفدار را دارد، در امتیاز هم میانگین و هم میانه، از 3 ژانر دیگر کمتر است!
- امتیاز Comedy از 3 مورد دیگر قابل اعتمادتر است (Range و IQR کمتر). در مقابل Drama
- 3 ژانر Action، Comedy و Thriller دارای میانه یکسان (3.50) و میانگین (3.44 تا 3.48) بسیار نزدیک هستند

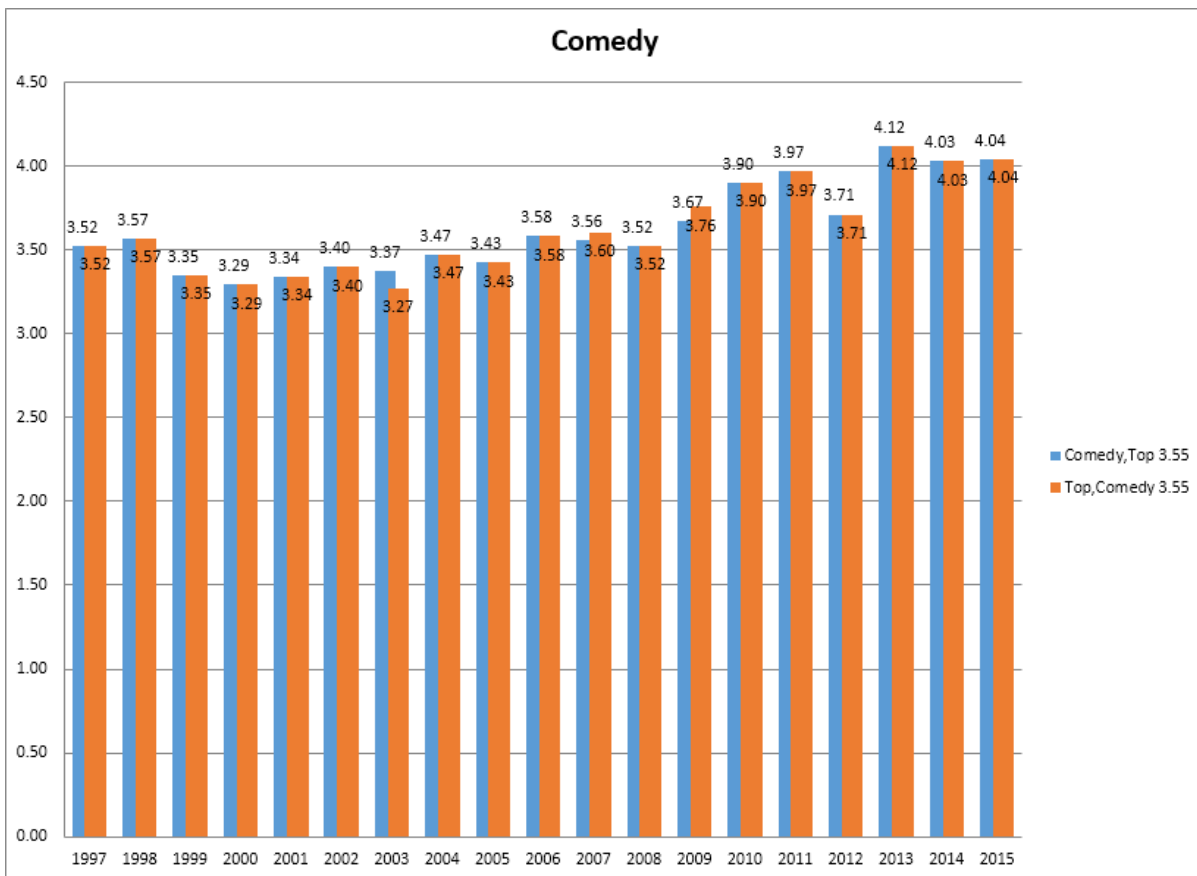
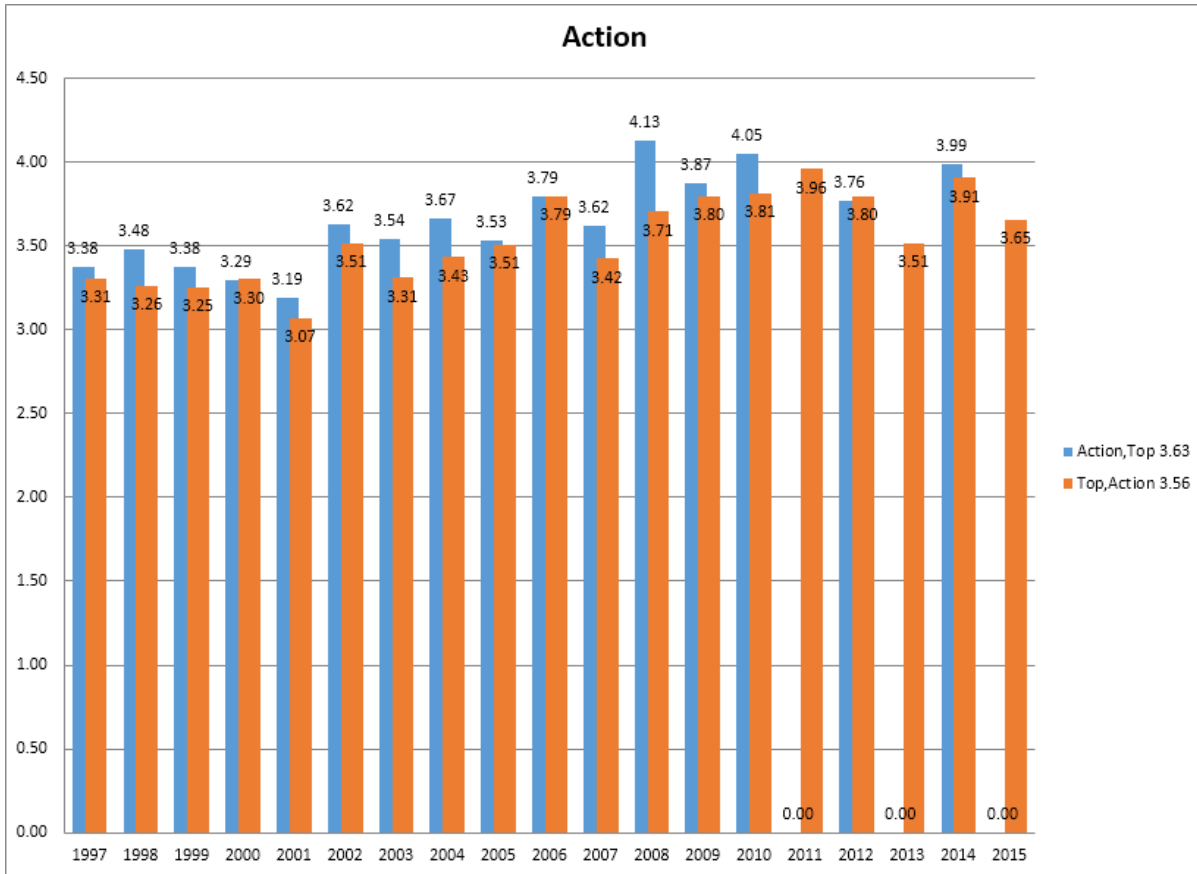
○ هر 3 دارای امتیازهای نسبتاً یکسان‌اند

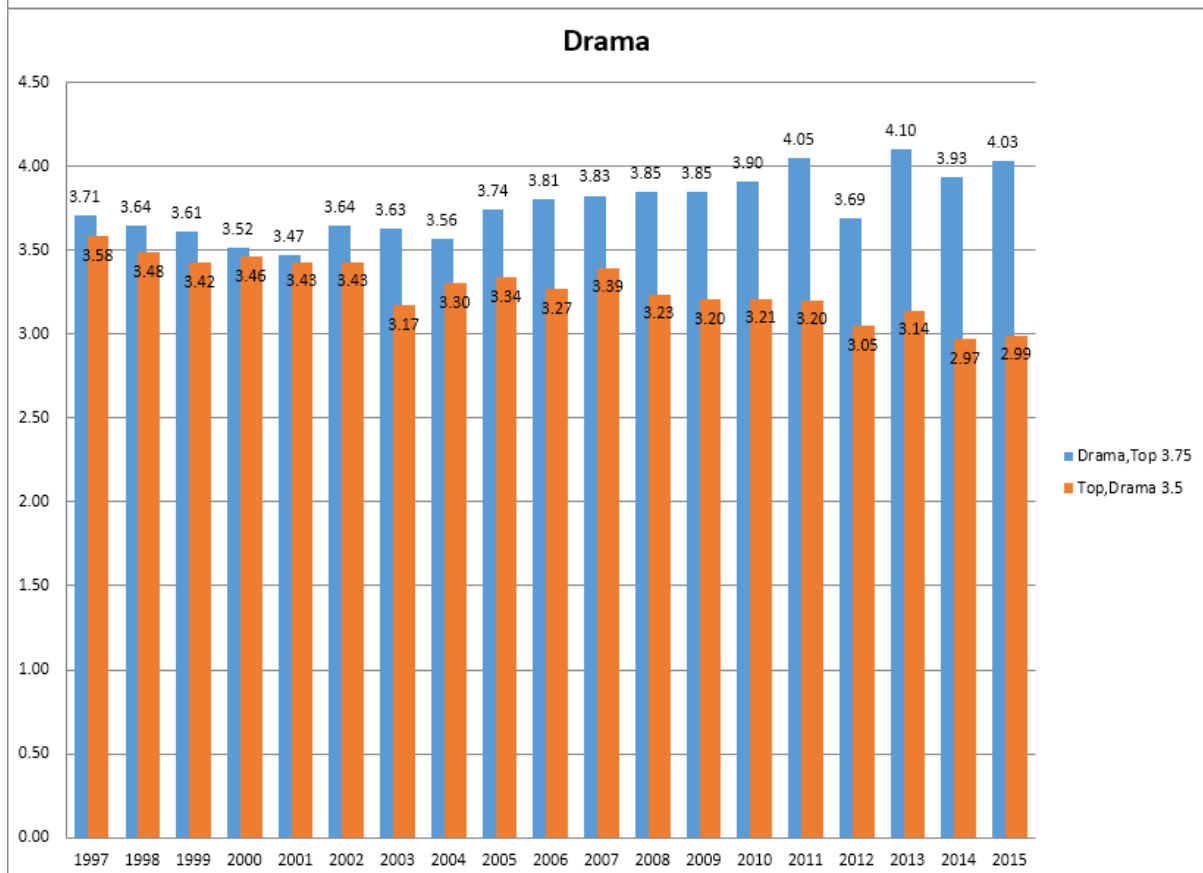
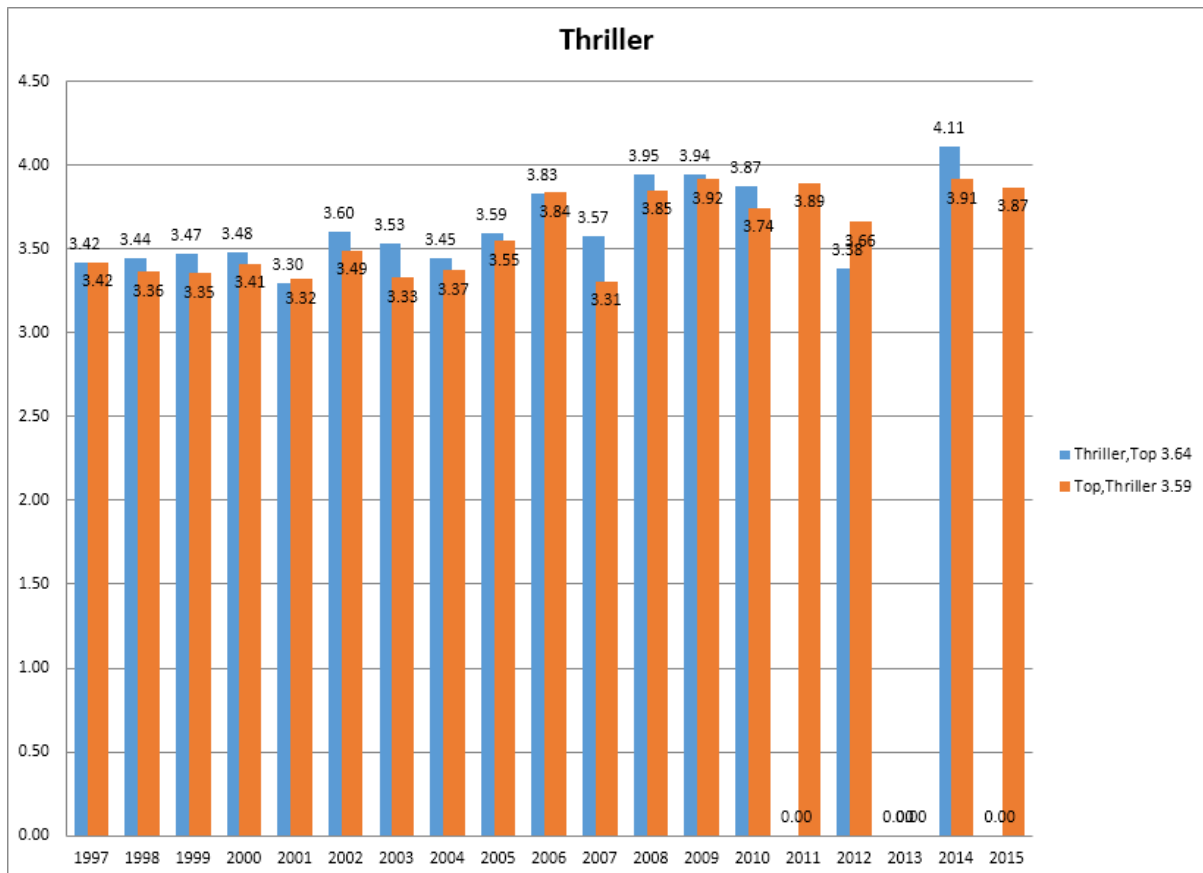
○ Action و Comedy دارای min بسیار نزدیک به هم‌اند.

برای خلاصه، امتیازها بدین صورت‌اند:

Action ~ = Comedy ~ > = Thriller > Drama

برای بررسی بیشتر این 4 ژانر امتیاز آنها را در سال‌های مختلف در نمودارهای زیر مشاهده کنید:







مشاهده میکنیم که نتایج حاصل از نمودار جعبه‌ای با رنگ نارنجی هر 4 نمودار میله‌ای فوق مطابقت دارد؛ به عنوان نمونه:

- مشاهده میکنیم که در Drama، بین فیلم‌های برتر، امتیاز ژانر به صورت نزولی است و بازه آن بیشتر از 3 ژانر دیگر است.

- دو ژانر Action و Comedy شباهت زیادی به هم دارند

### توزیع امتیاز فیلم (کلی) در سال های مختلف (سوال 3)

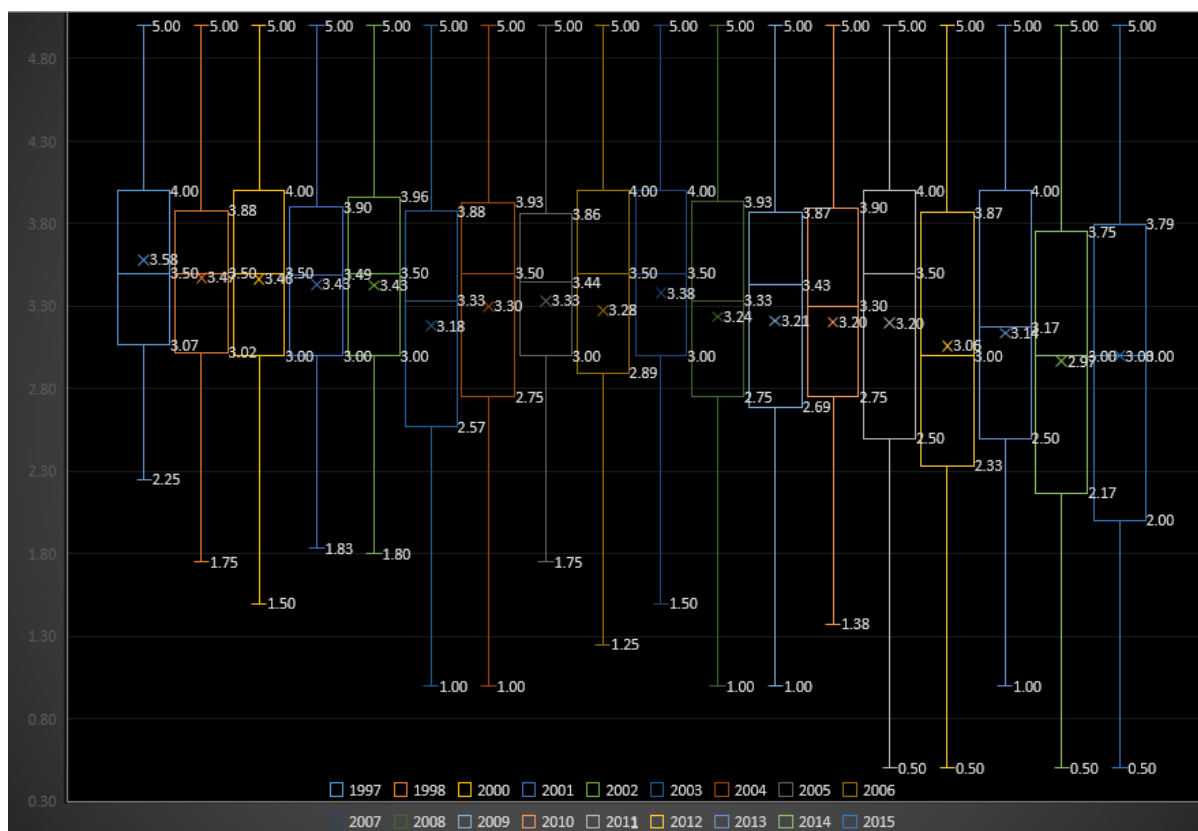
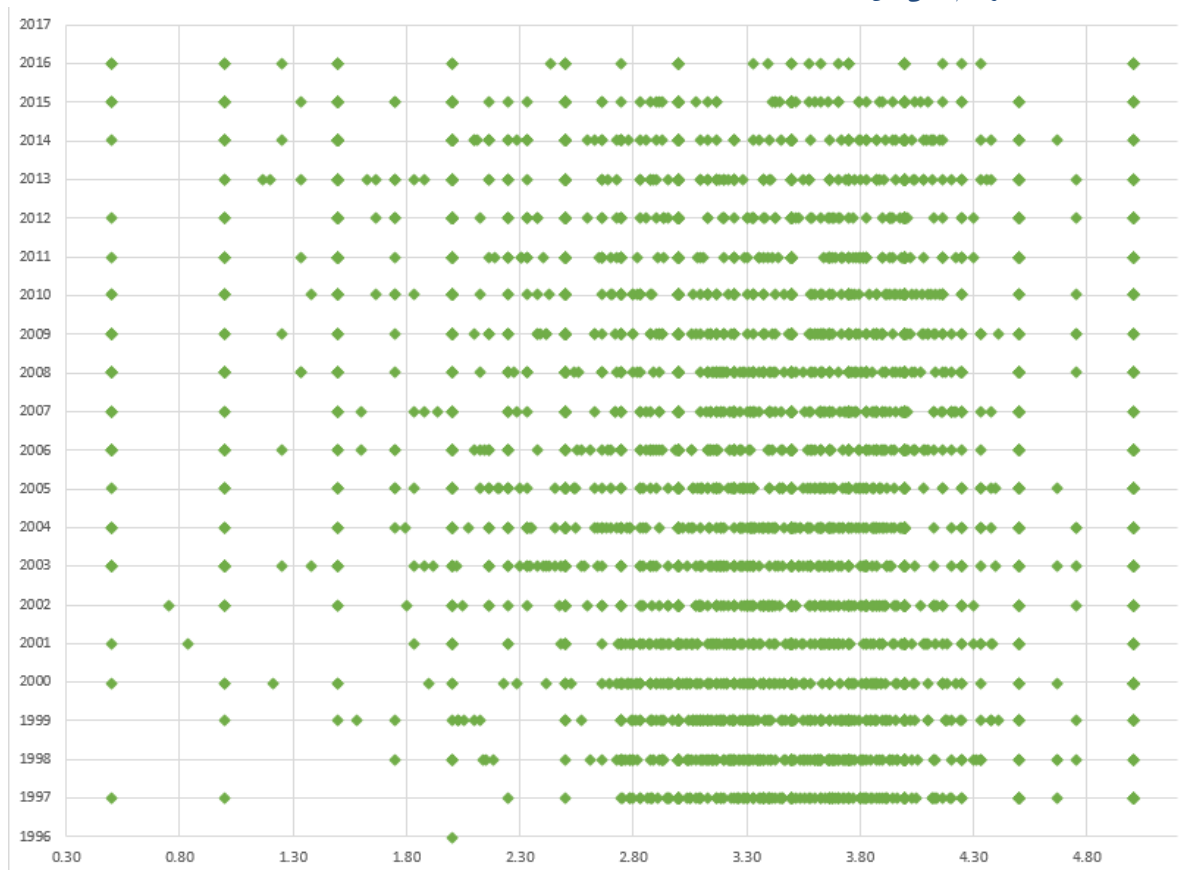


Figure 5 امتیاز فیلم ها طی سال ها



با دقت در 2 نمودار 5 و 6 مشاهده میکنیم که در 5 سال اخیر، پراکندگی امتیازها زیاد است ولی به طور کلی، امتیاز حول 3.50 است.

در سال‌های اولیه، توزیع امتیازها تجمع بسیار دارد و میتوان به جرات اظهار کرد که طی سالهای 1997 تا 2001، قریب به اتفاق امتیازها از 2.78 تا 4.2 هستند؛ مطابق نمودار جعبه‌ای (نمودار 5)، IQR طی این 5 سال، حدوداً از 3.0 تا 3.95 است. حتی با توجه به وسط بودن در IQR و تطابق میانه و میانگین، میتوان گفت در این 5 سال، توزیع امتیازها نرمال بوده است.

از سال 2002 تا 2010، میانگین کمتر از میانه است (چوله به چپ) و اکثراً داده‌ها در بازه وسیعی هستند (IQR نسبتاً بزرگ). میانه حدود 3.4 و میانگین حدود 3.25 است؛ بنابراین اغلب فیلم‌ها دارای امتیازی بیشتر از 3.4 اند ولی شدت کم بودن امتیاز فیلم‌های زیر 3.4 بیشتر است. به عبارت دیگر، برای یک فیلم که به صورت اتفاقی انتخاب شده، ه تقریب خوبی، اگر فیلم خوبی باشد، بالای 3.4 است، ولی اگر فیلم بدی باشد، امتیاز آن حتی شاید کمتر از 2 باشد!

البته بین این سال‌ها، سال 2005، (به دلیل کوچک بودن IQR) دارای فیلم‌های یکدست‌تری است؛ این مشاهده در نمودار 6 (پراکنش) هم دیده میشود. یعنی نقاط به هم نزدیک اند و حول بازه خاصی هستند. (3.00 تا 3.85)

طبق نمودار جعبه‌ای در 4 سال اخیر امتیازها افت بارزی دارند؛ میانه و میانگین در طی 2012 تا 2015 حدود 3.05 اند، که نشان دهنده افت کیفیت فیلم‌هاست. با این حال در سال 2011، نصف فیلم‌ها (میانه) امتیازی بیشتر از 3.5 داشته اند، که قابل قبول است. البته در این سال، امتیازهای گوناگون (کم و زیاد) بسیاری حتی بیشتر از هر سالی وجود داشته (IQR وسیع)

## بررسی نظرات برای فیلم the Matrix(1999) (سوال 4)

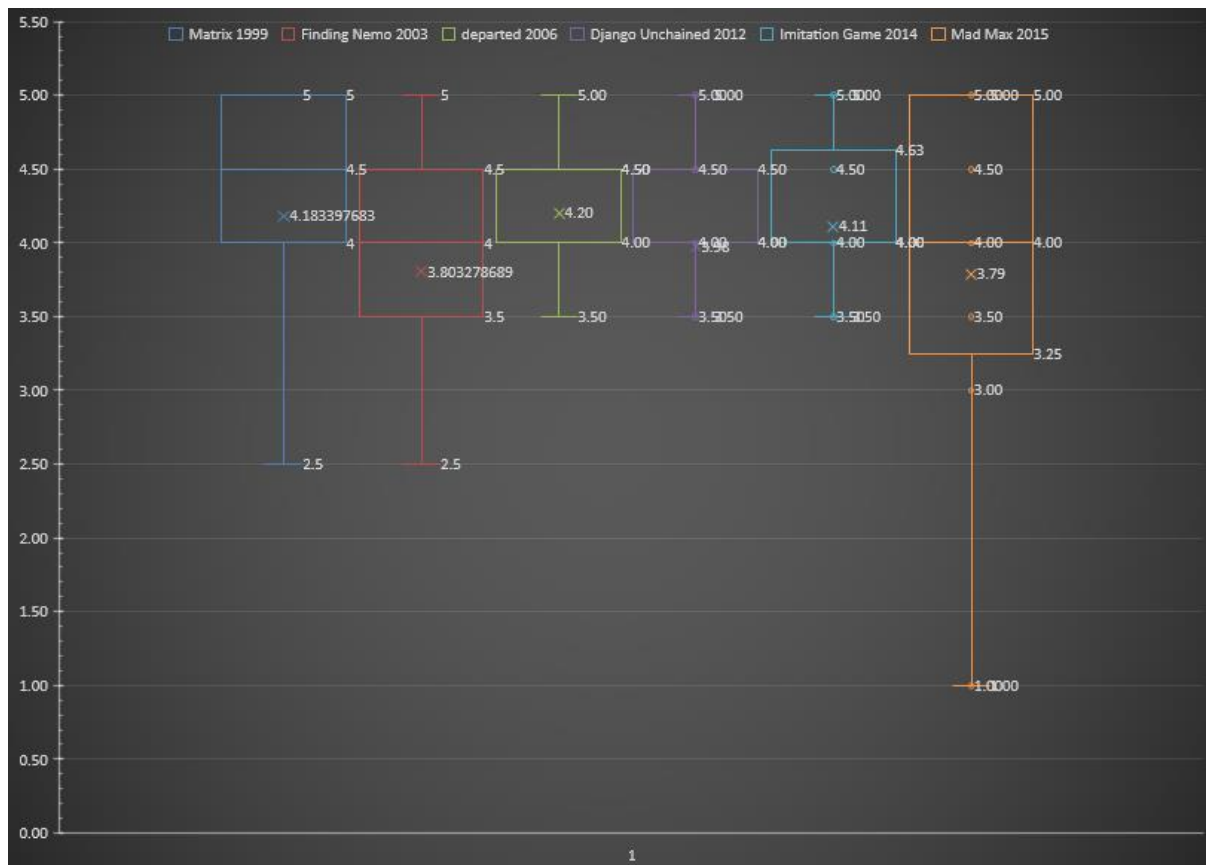


Figure 7 نمودار جعبه‌ای برای آرای چند فیلم

آرا تماشاگران را برای the Matrix (1999) بررسی میکنیم

count	259
sum	1083.50
min	1
max	5
mean	4.18
median	4.50
SD	0.90
cv	0.22
skew	-1.28
mean - median	-0.32

Figure 8 آمارهای Matrix1999

طبق جدول 8، آرا چوله به چپ(کمتر)، دارای میانگین 4.18 (انحراف معیار 0.90)، میانه 4.50 و حداقل 1 هستند.

با توجه به نمودار 7 (جعبه‌ای)، نصف امتیازها در بازه IQR: 4.00 to 5.00 هستند، و نصف کل امتیازها از 4.5 بیشترند. با این حال چون شدت کم بودن امتیازهای زیر 4.00 زیاد است(به هر دلیلی!) میانگین 4.18 است. این مشاهدات توسط نمودار 12(دایره‌ای) تایید میشوند.

طبق نمودار 11 (حبابی مطلق)، بیشترین تعداد رای در به ترتیب در 2 سال 2015 و 2000 بوده است.

همچنین این نمودار (11) نشانگر افت و خیز نظر تماشاگران درباره این فیلم است.

برای the Matrix (1999) این آمار خشنود کننده است؛ این فیلم یکی از پرفرودارترین فیلم‌هاست (لیست شده

در پاسخ سوال 5)

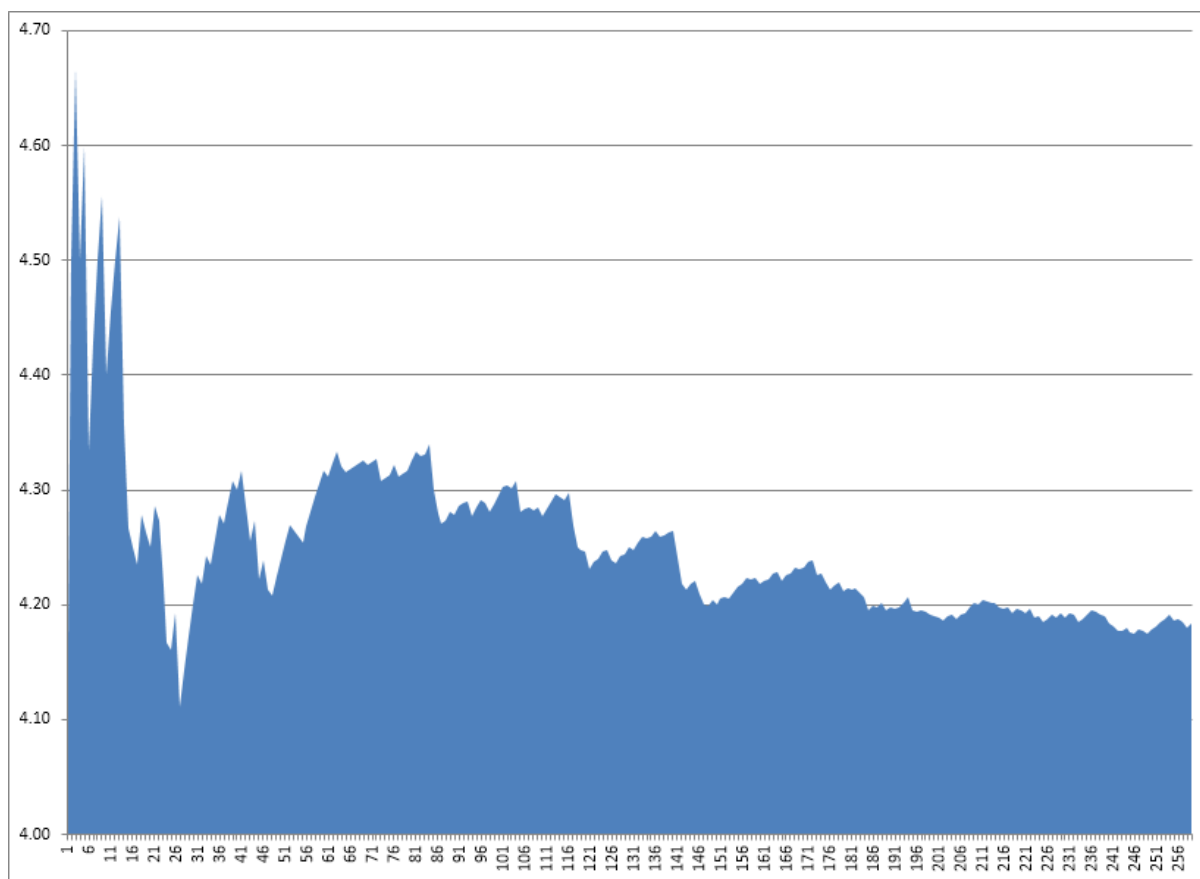


Figure 9 میانگین آرا به صورت تجمعی، بعد از هر رای برای Matrix1999

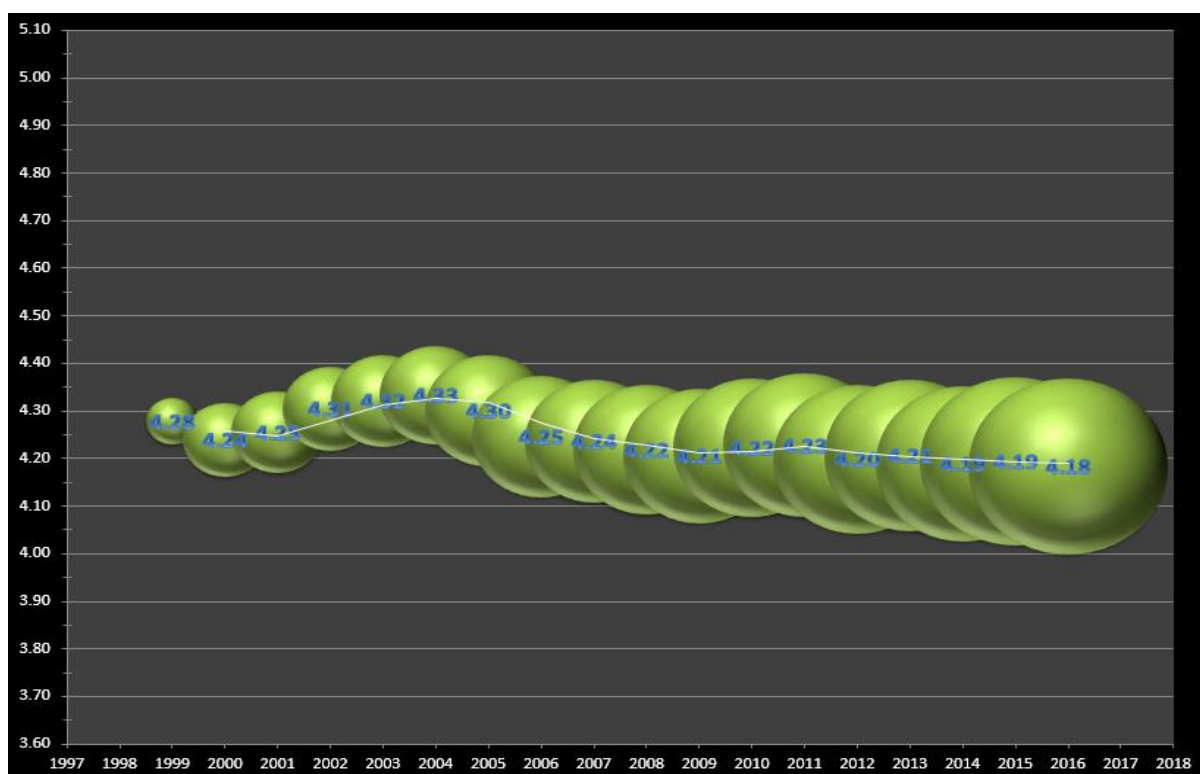


Figure 10 میانگین آرا به صورت تجمعی برای Matrix1999؛ قطر دایره‌ها متناسب با تعداد آرا است

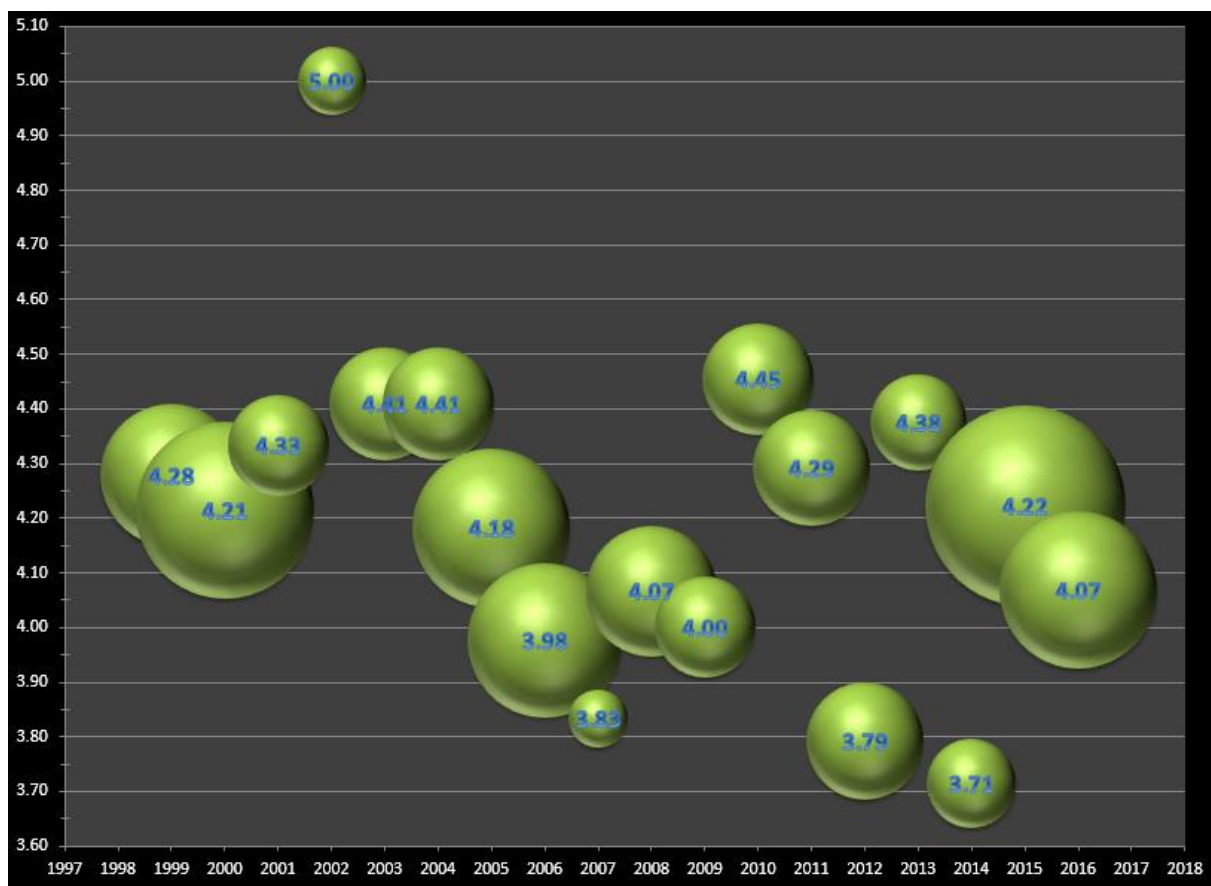


Figure 11 رای (مطلق) سالانه برای Matrix1999 قطر دایره‌ها متناسب با تعداد آرا است

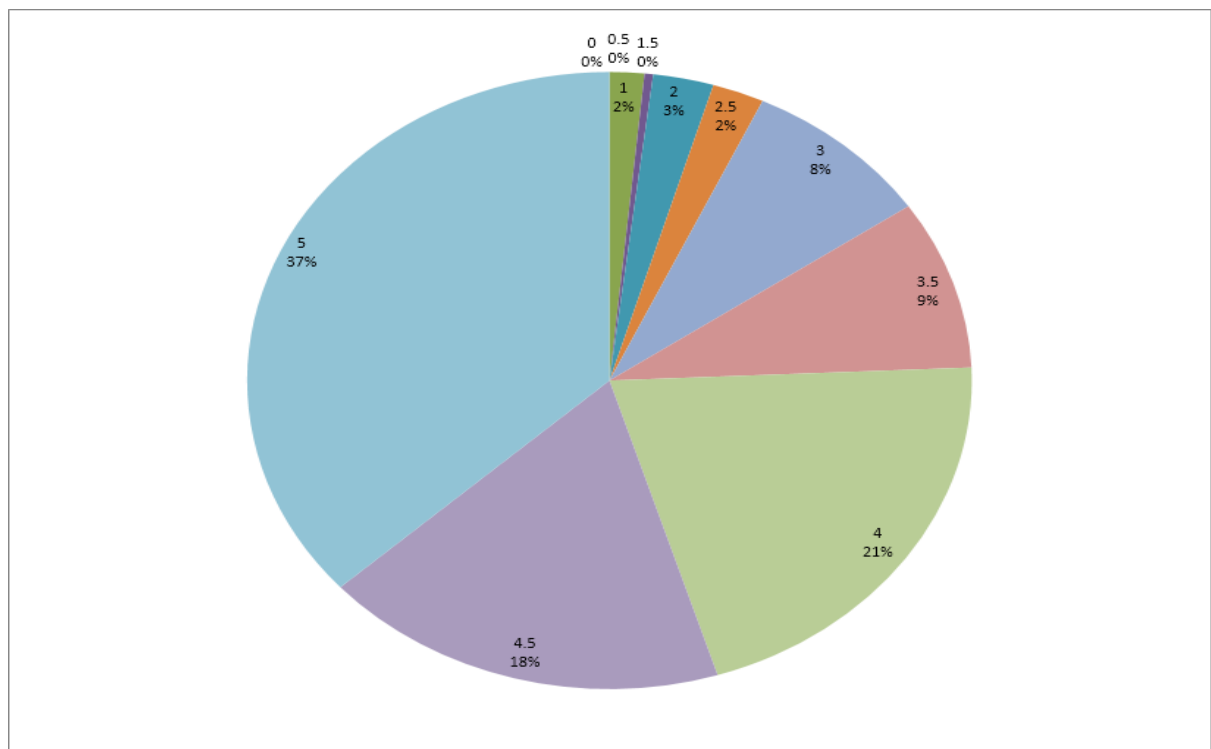


Figure 12 دسته بندی برای Matrix1999

## فیلم‌هایی با بیشترین امتیاز (سوال 5)

Name	Genre	Score
City of God (Cidade de Deus) (2002)	Action Adventure Crime Drama Thriller	4.30
American Beauty (1999)	Drama Romance	4.24
Dark Knight, The (2008)	Action Crime Drama IMAX	4.24
Memento (2000)	Mystery Thriller	4.23
Departed, The (2006)	Crime Drama Thriller	4.20

Figure 13 پراستیمایزترین‌ها

این 5 فیلم (جدول 13) دارای بیشترین امتیاز طی این 19 سال (از 1997 تا 2015) هستند.

**توضیح:** برای دقیق شدن نظرات، فیلم‌هایی که در 5٪ بیشترین تعداد آراء (دارای بیشتر از 23 رای) بوده‌اند، بررسی شده‌اند.

Name	Genre	Votes	Score
Matrix, The (1999)	Action Sci-Fi Thriller	259	4.18
<b>American Beauty (1999)</b>	Drama Romance	220	4.24
Fight Club (1999)	Action Crime Drama Thriller	202	4.18
Lord of the Rings: The Fellowship of the Ring, The (2001)	Adventure Fantasy	200	4.18

Figure 14 بیشترین تعداد رای

با این وجود، فیلم‌هایی که دارای بیشترین تعداد رای (بیشترین تماشاگر/بیننده/طرفدار) هستند، طبق جدول 14، به

ترتیب Lord of the Rings: The Fellowship of the Ring, The (2001)، American Beauty (1999)، Fight Club (1999)، Matrix, The (1999) هستند.

The End.