Today's Agenda

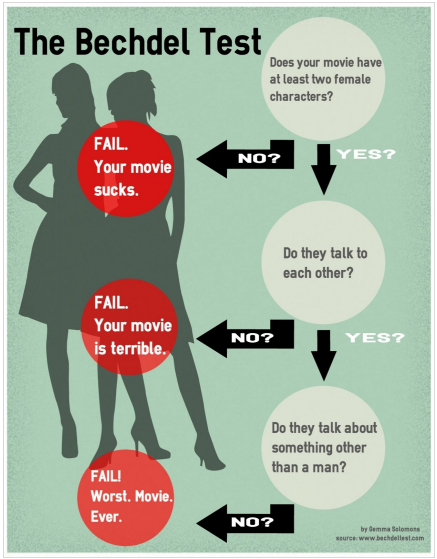
Practice extending the OLS Regression

1. Dummy predictors
2. Categorical predictors
3. Transforming the variables
4. Transforming the model

Hickey, W. (2014, Apr 1). The Dollar-And-Cents Case Against Hollywood's Exclusion of Women. FiveThirtyEight.

	A	B	C	D	E	F	G	H
1	year	title	bechdel	bechdel_why	budget_2013	domgross_2013	intgross_2013	total_gross_2013
2	2013	21 & Over	FAIL	notalk	13000000	25682380	42195766	67878146
3	2012	Dredd 3D	PASS	ok	45658735	13611086	41467257	55078343
4	2013	12 Years a Slave	FAIL	notalk	20000000	53107035	158607035	211714070
5	2013	2 Guns	FAIL	notalk	61000000	75612460	132493015	208105475
6	2013	42	FAIL	men	40000000	95020213	95020213	190040426
7	2013	47 Ronin	FAIL	men	225000000	38362475	145803842	184166317
8	2013	A Good Day to Die Hard	FAIL	notalk	92000000	67349198	304249198	371598396
9	2013	About Time	PASS	ok	12000000	15323921	87324746	102648667
10	2013	Admission	PASS	ok	13000000	18007317	18007317	36014634
11	2013	After Earth	FAIL	notalk	130000000	60522097	244373198	304895295
12	2013	American Hustle	PASS	ok	40000000	148430908	249484909	397915817
13	2013	August: Osage County	PASS	ok	25000000	37304874	50304874	87609748
14	2013	Beautiful Creatures	PASS	ok	50000000	19452138	55940671	75392809
15	2013	Blue Jasmine	PASS	ok	18000000	33345833	68447831	101793664
16	2013	Captain Phillips	FAIL	notalk	55000000	107136417	218743570	325879987
17	2013	Carrie	PASS	ok	30000000	35266619	85001659	120268278

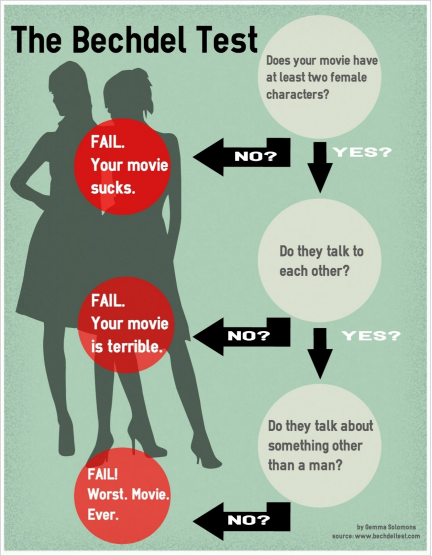
The Bechdel Test: Sexism at the Movies



The Bechdel Test: Sexism at the Movies

Let's explore!

- Top 10 movies pass the test?
- Top 20?
- Any surprises?



Do sexist movies (Bechdel = Fail) make more money at the box office?

- Regress total_gross_2013 on bechdel

	Box Office (millions)
bechdel_pass	−83.22*
	(19.16)
Constant	330.95*
	(12.81)
Observations	1,776
Adjusted R^2	0.01
Residual Std. Error	401.41 (df = 1774)
F Statistic	18.87* (df = 1; 1774)
Note:	*p<0.05

Is the international market more sexist then the domestic market?

Fit and evaluate the following two models:

- Model 1: Regress domgross_2013 on bechdel
- Model 2: Regress intgross_2013 on bechdel

	Box Office (millions)	Domestic	International
	(1)	(2)	(3)
bechdel_pass	−83.22*	−28.18*	−55.17*
	(19.16)	(5.98)	(13.44)
Constant	330.95*	107.77*	222.53*
	(12.81)	(4.00)	(8.99)
Observations	1,776	1,776	1,783
Adjusted R^2	0.01	0.01	0.01
Residual Std. Error	401.41 (df = 1774)	125.22 (df = 1774)	282.26 (df = 1781)
F Statistic	18.87* (df = 1; 1774)	22.24* (df = 1; 1774)	16.84* (df = 1; 1781)

Note: *p<0.05

What kinds of sexism make the most money at the box office?

bechdel_why

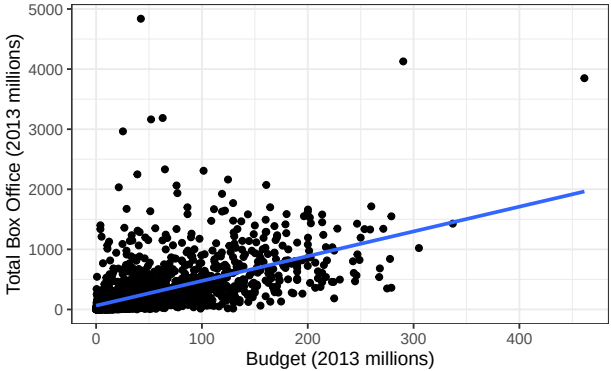
1. nowomen = Fewer than two women
2. notalk = Women don't talk to each other
3. men = Women only talk about men
4. ok = Passes the test

What kinds of sexism make the most money at the box office?

1. Convert bechdel_why to three dummies: nowomen, notalk and men (Baseline = "ok")
2. Regress total_gross_2013 on bechdel_why dummies

	Box Office (millions)
nowomen	59.01
	(36.64)
notalk	107.93*
	(22.54)
men	40.22
	(31.88)
Constant	247.73*
	(14.10)
Observations	1,635
Adjusted R^2	0.01
Residual Std. Error	397.25 (df = 1631)
F Statistic	7.72* (df = 3; 1631)

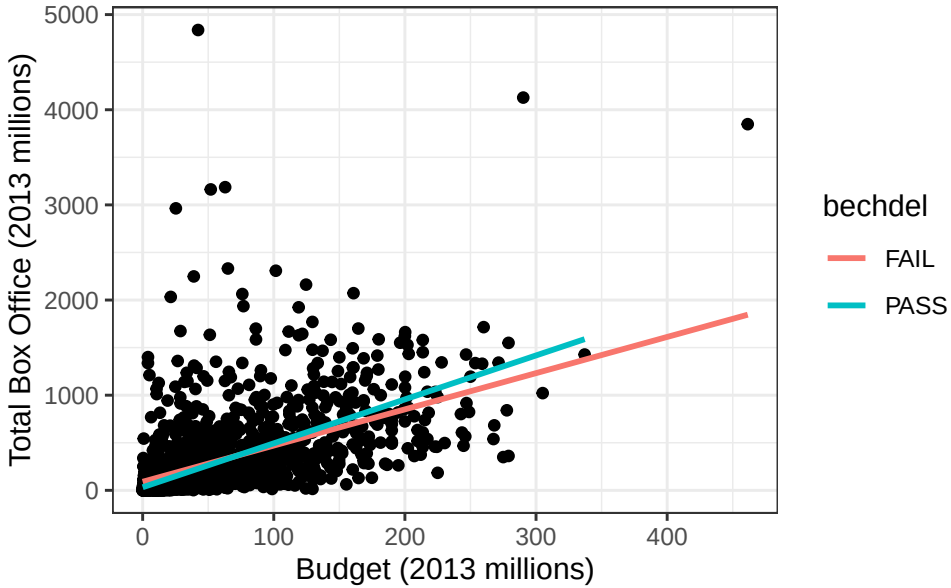
Is box office about sexism or budget size?



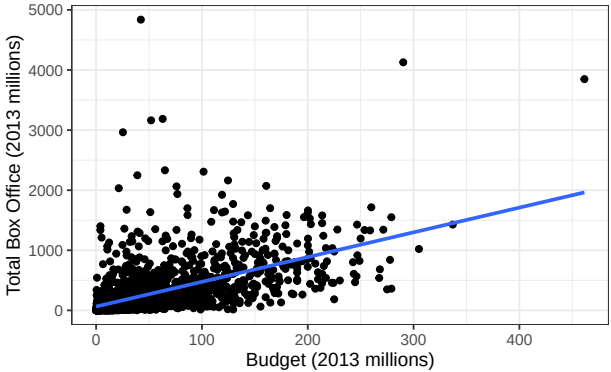
- Regress total_gross_2013 on bechdel and budget_2013_millions

	Box Office (millions)	
	(1)	(2)
bechdel_pass	−83.22*	−14.92
	(19.16)	(16.12)
budget_2013_millions		4.10*
		(0.15)
Constant	330.95*	71.47*
	(12.81)	(14.10)
Observations	1,776	1,776
Adjusted R^2	0.01	0.31
Residual Std. Error	401.41 (df = 1774)	333.96 (df = 1773)
F Statistic	18.87* (df = 1; 1774)	408.66* (df = 2; 1773)

Note: *p<0.05


Total Box Office (2013 millions)
5000
4000
3000
2000
1000
0
0
100
200
300
400
Budget (2013 millions)
bechdel
FAIL
PASS

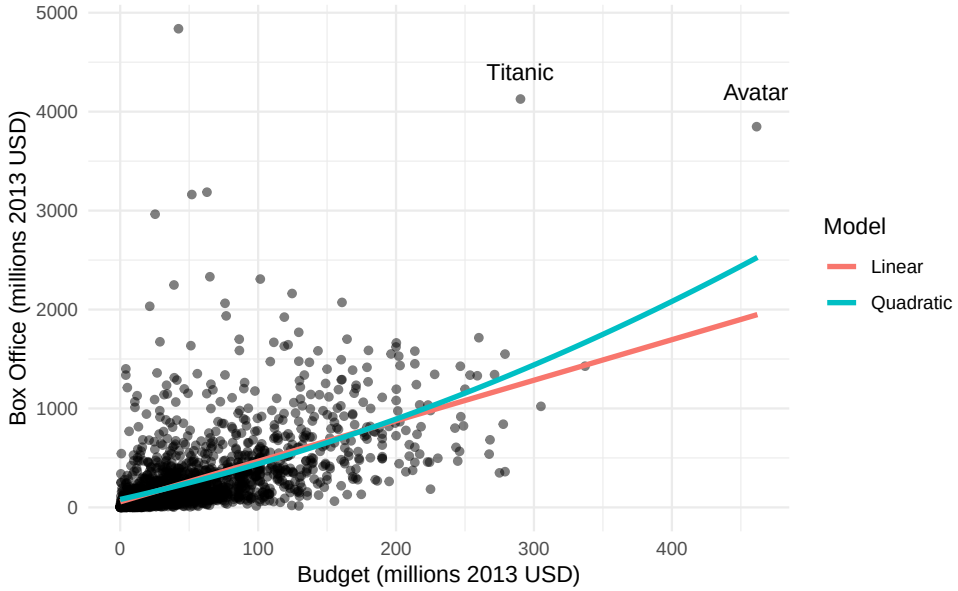
Is box office about sexism or budget size?



- Redo the last model but include a quadratic function of budget_2013_millions

	Box Office (millions)	
	(1)	(2)
bechdel_pass	−14.92	−17.54
	(16.12)	(16.11)
budget_2013_millions	4.10*	3.12*
	(0.15)	(0.35)
budget_2013_millions2		0.005*
		(0.002)
Constant	71.47*	98.29*
	(14.10)	(16.53)
Observations	1,776	1,776
Adjusted R^2	0.31	0.32
Residual Std. Error	333.96 (df = 1773)	333.16 (df = 1772)
F Statistic	408.66* (df = 2; 1773)	276.92* (df = 3; 1772)

Note: *p<0.05


Titanic
Avatar
Box Office (millions 2013 USD)
Budget (millions 2013 USD)
5000
4000
3000
2000
1000
0
0
100
200
300
400
Model
Linear
Quadratic