

Modeling Strategy

INFO 370

Learning Objectives

Discuss the insights from Notebook 4

Address questions **about data science** raised by the article

Discuss **hierarchical** (nested) models

Discuss assignment 3 (to be done **in pairs**)

Start working on Notebook 5, if there's time

Notebook 4

Using the data/insights from your analysis, tell a detailed and insightful story about gender discrimination in the department.

count

sl

sx

female 14

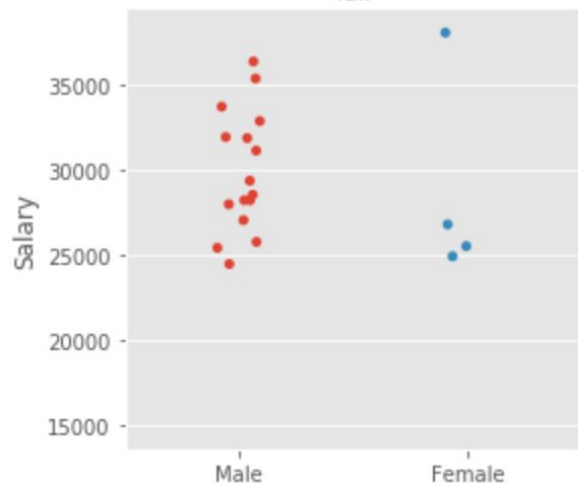
male 38

sx

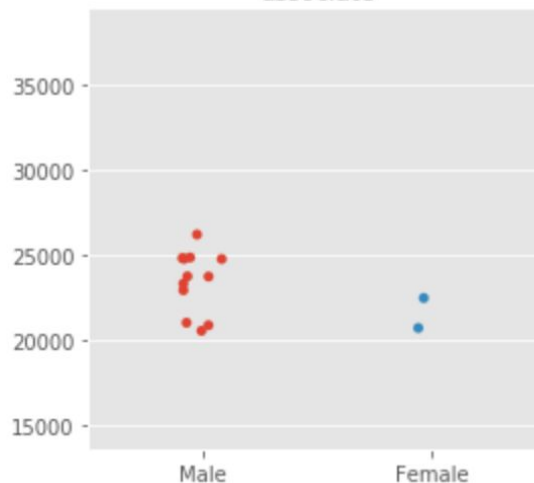
female 21357.142857

male 24696.789474

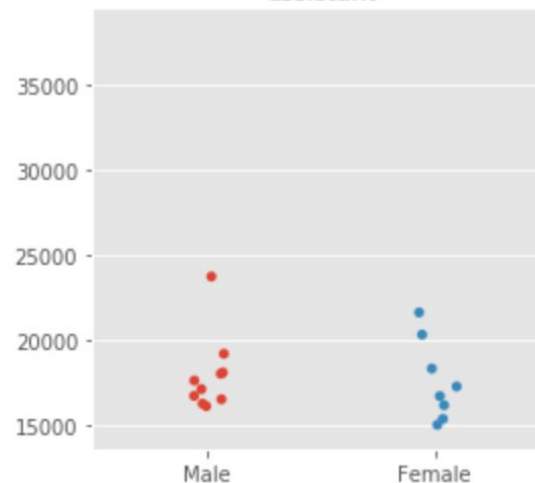
full



associate



assistant



```
# Test for difference using `ttest_ind`
x = ttest_ind(males[['sl']], females[['sl']]) # not significant at p = .05
x.pvalue

array([ 0.07060394])
```

What did this exercise reveal about gender discrimination in the department?

	coef	std err	t	P> t 	[0.025	0.975]
Intercept	1.637e+04	757.573	21.615	0.000	1.49e+04	1.79e+04
sx[T.male]	-505.6541	843.489	-0.599	0.552	-2203.508	1192.200
rk[T.associate]	4253.7176	961.967	4.422	0.000	2317.379	6190.056
rk[T.full]	9508.1038	923.022	10.301	0.000	7650.157	1.14e+04
dg[T.masters]	314.4186	782.166	0.402	0.690	-1260.000	1888.837
yr	385.1491	77.415	4.975	0.000	229.322	540.977

What did this exercise reveal about gender discrimination in the department?

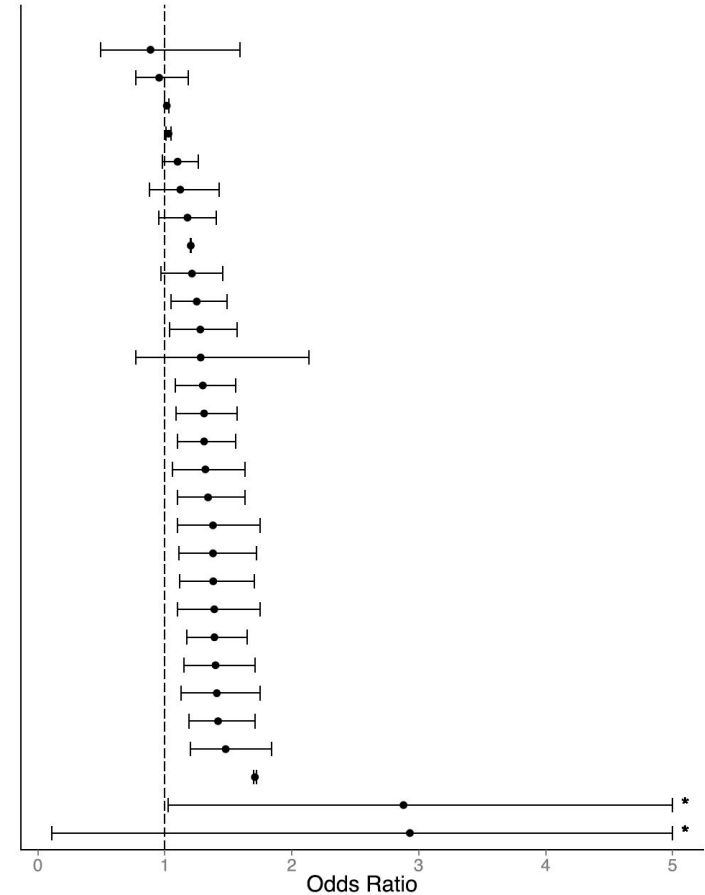
Reading Discussion

Covariate	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	20	21	23	24	25	26	27	28	30	31	32	% used
Position																														62%
Height																														38%
Weight																														38%
Age																														24%
League Country																														17%
Goals																														17%
Referee Country																														17%
Victories																														10%
Club																														7%
Referee																														7%
Player Cards																														7%
Player																														3%
Referee Cards																														3%
Draws																														3%
N Covariates	7	6	2	3	0	3	0	0	2	3	3	2	1	6	1	2	2	2	1	3	2	3	4	6	1	2	3	4	1	

Table 2. This overview shows the covariates used by each team. Team numbers are listed on the top and covariates on the left. A shaded box indicates that the corresponding team used the covariate in their final model. The table is ordered by the frequency by which each covariate was used.

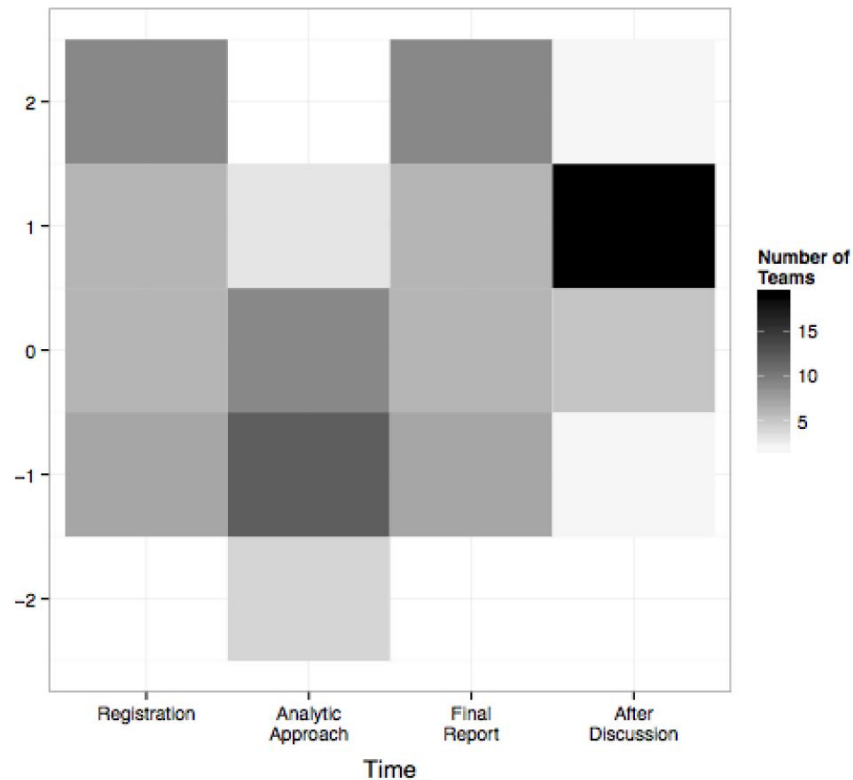
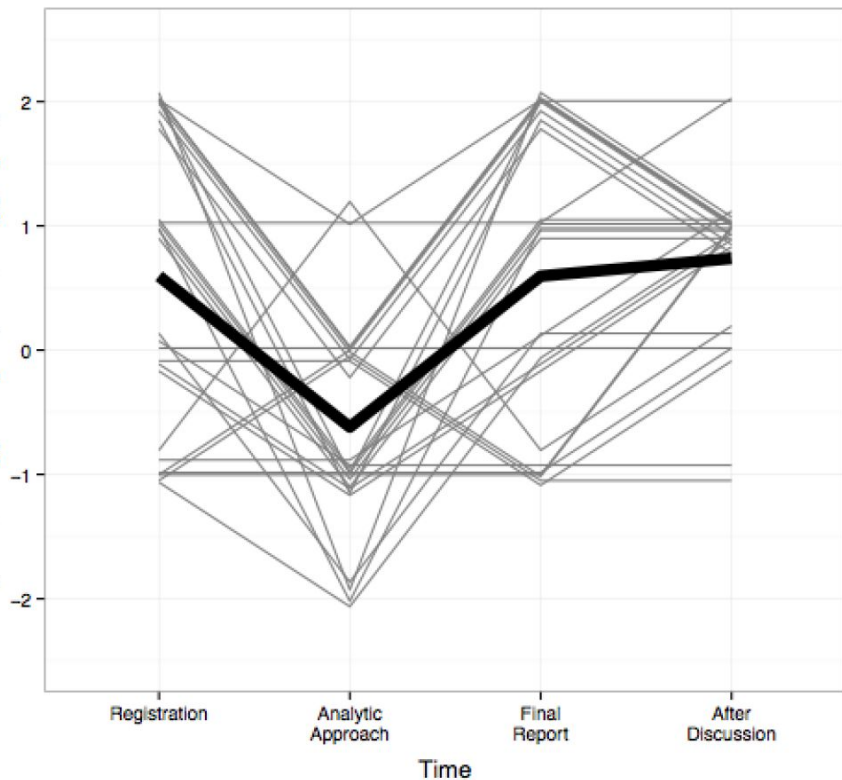
What does this mean? What does it mean we should do?

Team	Analytic Approach	OR
12	Zero-inflated Poisson regression	0.89
17	Bayesian logistic regression	0.96
15	Hierarchical log-linear modeling	1.02
10	Multilevel regression and logistic regression	1.03
18	Hierarchical Bayes model	1.10
31	Logistic regression	1.12
1	Ordinary least squares with robust standard errors, logistic regression	1.18
4	Spearman correlation	1.21
14	Weighted least squares regression with clustered standard errors	1.21
11	Multiple linear regression	1.25
30	Clustered robust binomial logistic regression	1.28
6	Linear Probability Model	1.28
26	Three-level hierarchical generalized linear modeling with Poisson sampling	1.30
3	Multilevel Binomial Logistic Regression using bayesian inference	1.31
23	Mixed model logistic regression	1.31
16	Hierarchical Poisson Regression	1.32
2	Linear probability model, logistic regression	1.34
5	Generalized linear mixed models	1.38
24	Multilevel logistic regression	1.38
28	Mixed effects logistic regression	1.38
32	Generalized linear models for binary data	1.39
8	Negative binomial regression with a log link analysis	1.39
20	Cross-classified multilevel negative binomial model	1.40
13	Poisson Multi-level modeling	1.41
25	Multilevel logistic binomial regression	1.42
9	Generalized linear mixed effects models with a logit link function	1.48
7	Dirichlet process Bayesian clustering	1.71
21	Tobit regression	2.88
27	Poisson regression	2.93



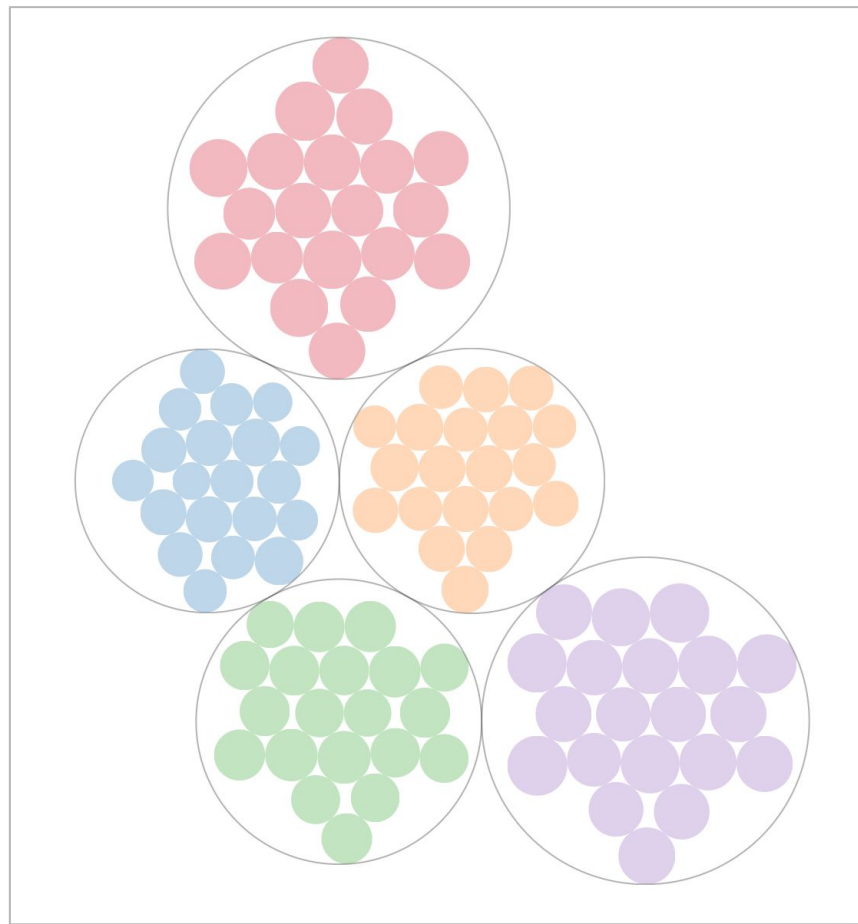
What does this mean? What does it mean we should do?

Researcher Subjective Belief in the Effect for Research Question 1
(-2 = Strongly Disagree, 2 = Strongly Agree)



What does this mean? What does it mean we should do?

Hierarchical Models



Hierarchical models ([link](#))

Assignment 3

Notebook Set 5

nb-set-5

Upcoming...

a3-modeling due ***next Tuesday***

Notebook 5 due **Friday night**

Later this week

- Logistic regression