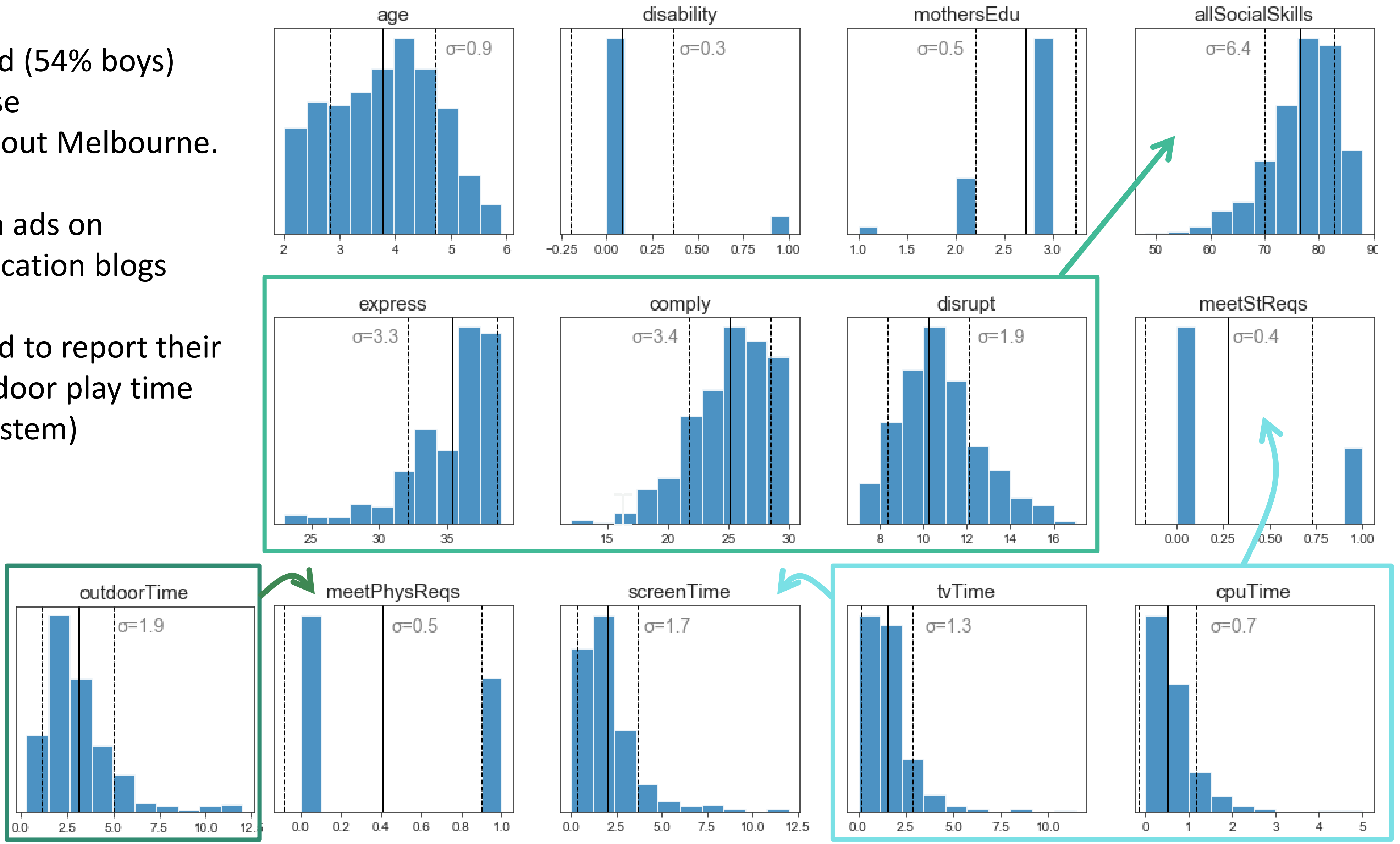
Children’s Screen Time & Social Skills

Dataset:

- 575 mothers with a child (54% boys) aged 2–5 years in diverse neighborhoods throughout Melbourne.
- Mothers were found via ads on parenting and child education blogs
- Mothers were instructed to report their child's screen time, outdoor play time and social skills (ASBI system)



https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193700

Topics

1. Can we predict a child’s “Social Skill” score be predicated based on demographics and activities (screen time, outdoor time)?
2. Can we predict whether a child will show disruptive behavior?
   - Sidebar: What does PCA do (visually) when we create components out of binary variables?

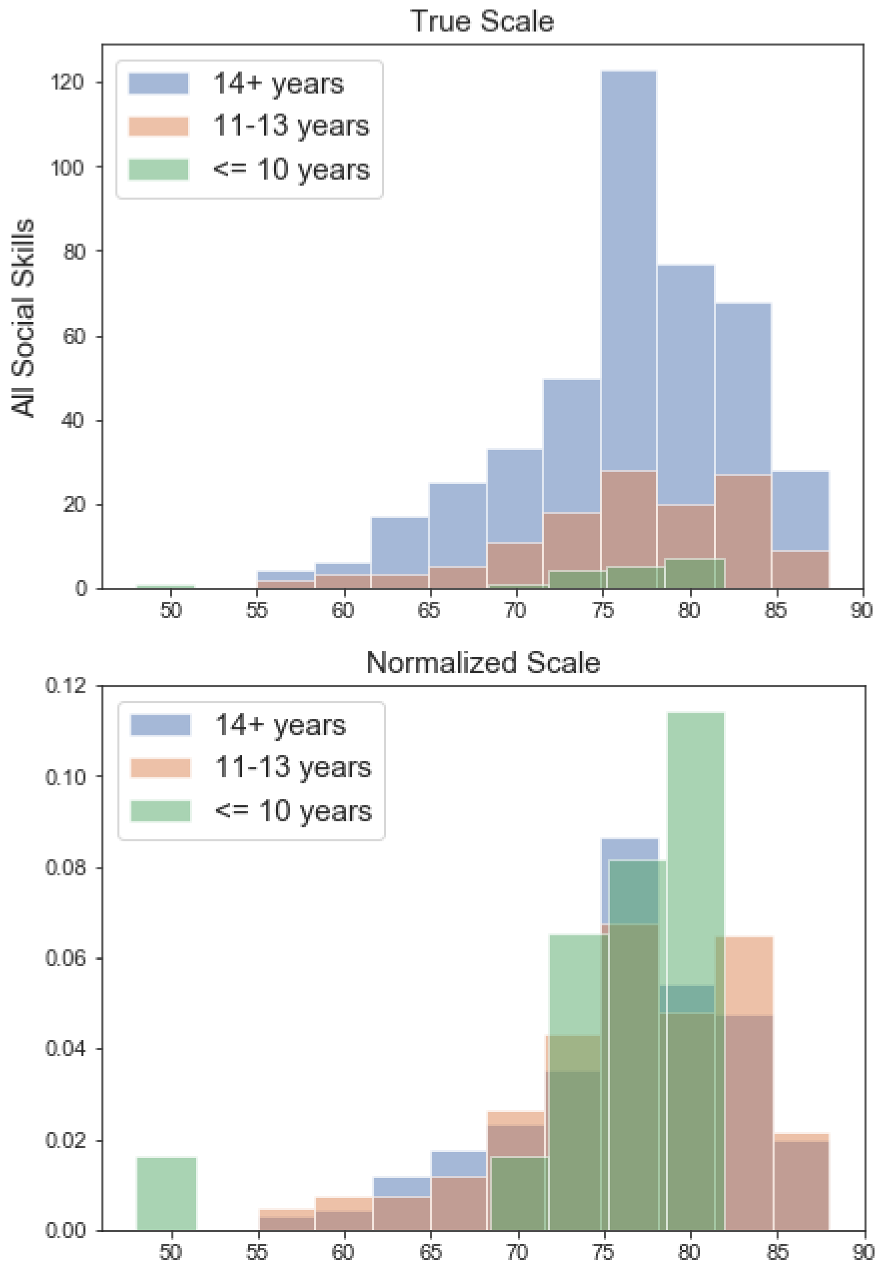
Predicting Social Skill

1) One-hot encode 'gender' and 'mother's education'

2) Use Lasso Regression to find the useful features:

```
[('age', '1.66'),
 ('disability', '-2.80'),
 ('meetStReqs', '0.00'),
 ('meetPhysReqs', '-0.00'),
 ('tvTime', '-0.54'),
 ('cpuTime', '-0.28'),
 ('outdoorTime', '0.43'),
 ('gender_Female', '1.23'),
 ('gender_Male', '-0.00'),
 ('mothersEdu_1', '-0.00'),
 ('mothersEdu_2', '0.00'),
 ('mothersEdu_3', '-0.00')]
```

?

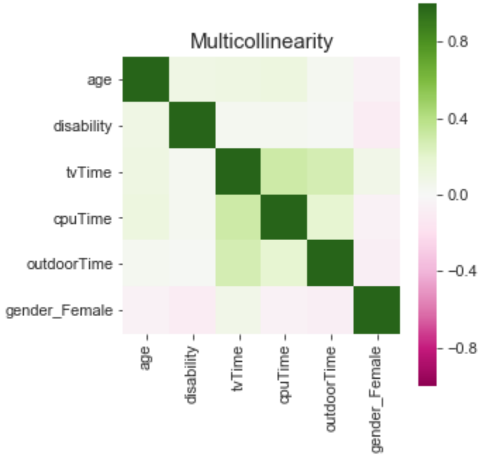


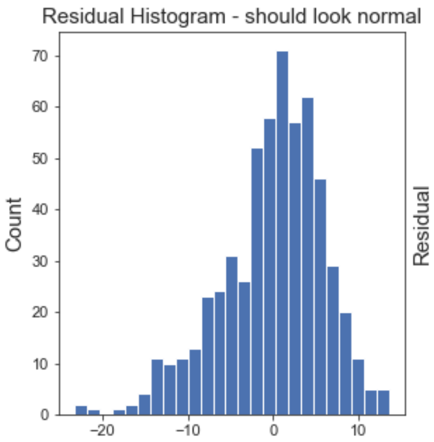
Predicting Social Skill

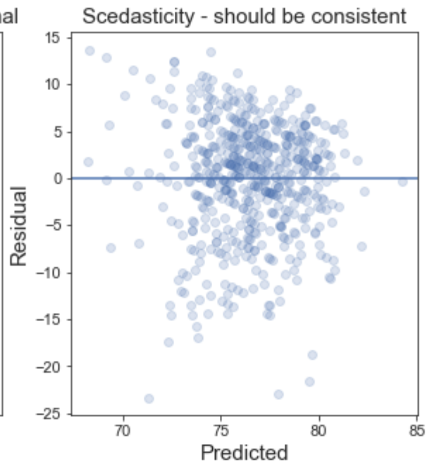
3) Simple Linear Regression – examine output

Dep. Variable:	allSocialSkills	R-squared:	0.132
Model:	OLS	Adj. R-squared:	0.123
Method:	Least Squares	F-statistic:	14.45
Date:	Sat, 05 Jan 2019	Prob (F-statistic):	2.28e-15
Time:	10:53:25	Log-Likelihood:	-1843.6
No. Observations:	575	AIC:	3701.
Df Residuals:	568	BIC:	3732.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	68.8026	1.137	60.524	0.000	66.570	71.035
age	1.8390	0.269	6.834	0.000	1.310	2.368
disability	-4.0338	0.899	-4.489	0.000	-5.799	-2.269
tvTime	-0.6026	0.205	-2.945	0.003	-1.005	-0.201
cpuTime	-0.5011	0.401	-1.250	0.212	-1.288	0.286
outdoorTime	0.4907	0.136	3.596	0.000	0.223	0.759
gender_Female	1.5896	0.511	3.113	0.002	0.587	2.593



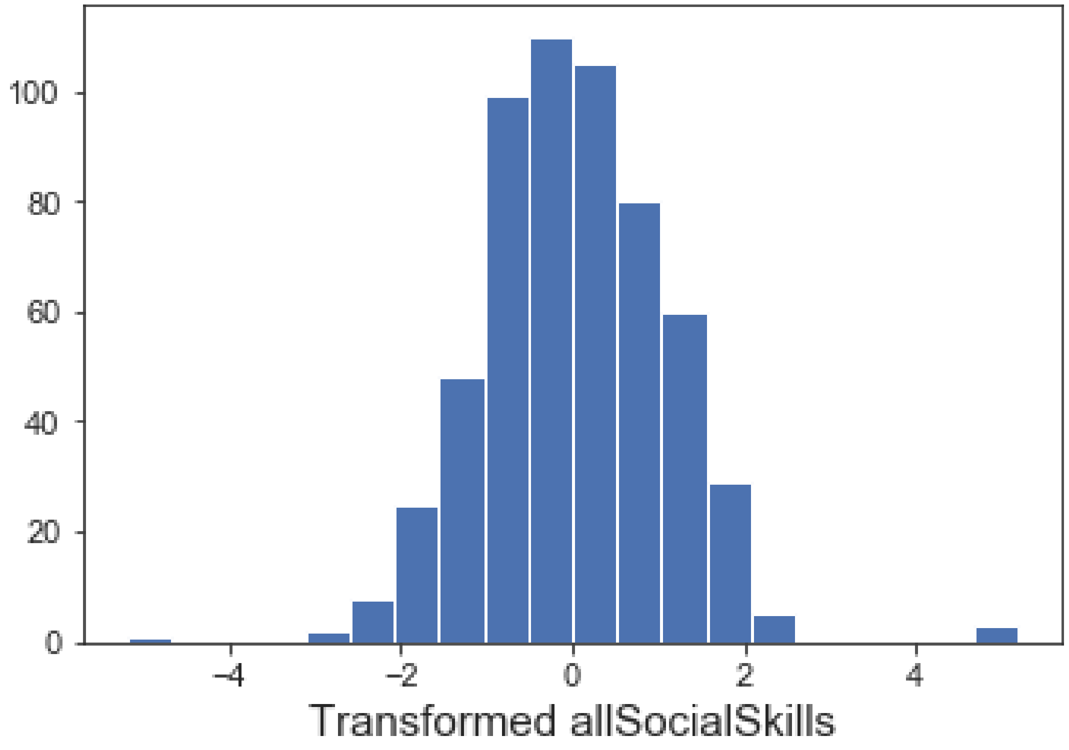




Predicting Social Skill

4) Our Target variable is a skewed distribution. Does Quantile-transforming it help?

```
y_trans = quantile_transform(pd.DataFrame(df.allSocialSkills),output_distribution='normal').squeeze()
```

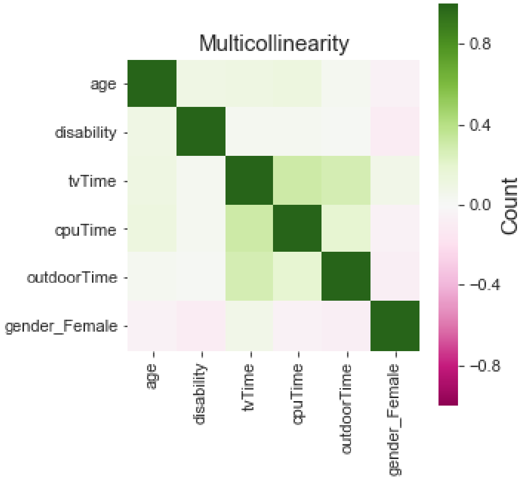


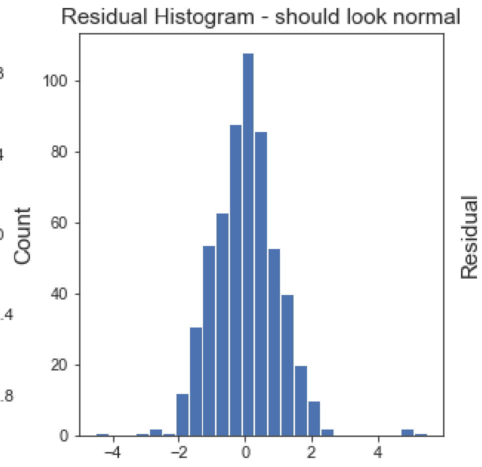
Predicting Social Skill

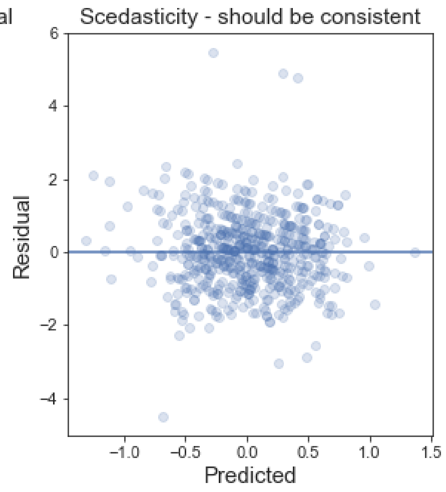
4) Our Target variable is a skewed distribution. Does Quantile-transforming it help?

Dep. Variable:	ytrans	R-squared:	0.128
Model:	OLS	Adj. R-squared:	0.119
Method:	Least Squares	F-statistic:	13.87
Date:	Sat, 05 Jan 2019	Prob (F-statistic):	9.59e-15
Time:	10:53:26	Log-Likelihood:	-811.75
No. Observations:	575	AIC:	1638.
Df Residuals:	568	BIC:	1668.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.3245	0.189	-7.010	0.000	-1.696	-0.953
age	0.3180	0.045	7.110	0.000	0.230	0.406
disability	-0.5579	0.149	-3.736	0.000	-0.851	-0.265
tvTime	-0.0951	0.034	-2.796	0.005	-0.162	-0.028
cpuTime	-0.0968	0.067	-1.453	0.147	-0.228	0.034
outdoorTime	0.0882	0.023	3.888	0.000	0.044	0.133
gender_Female	0.2185	0.085	2.575	0.010	0.052	0.385



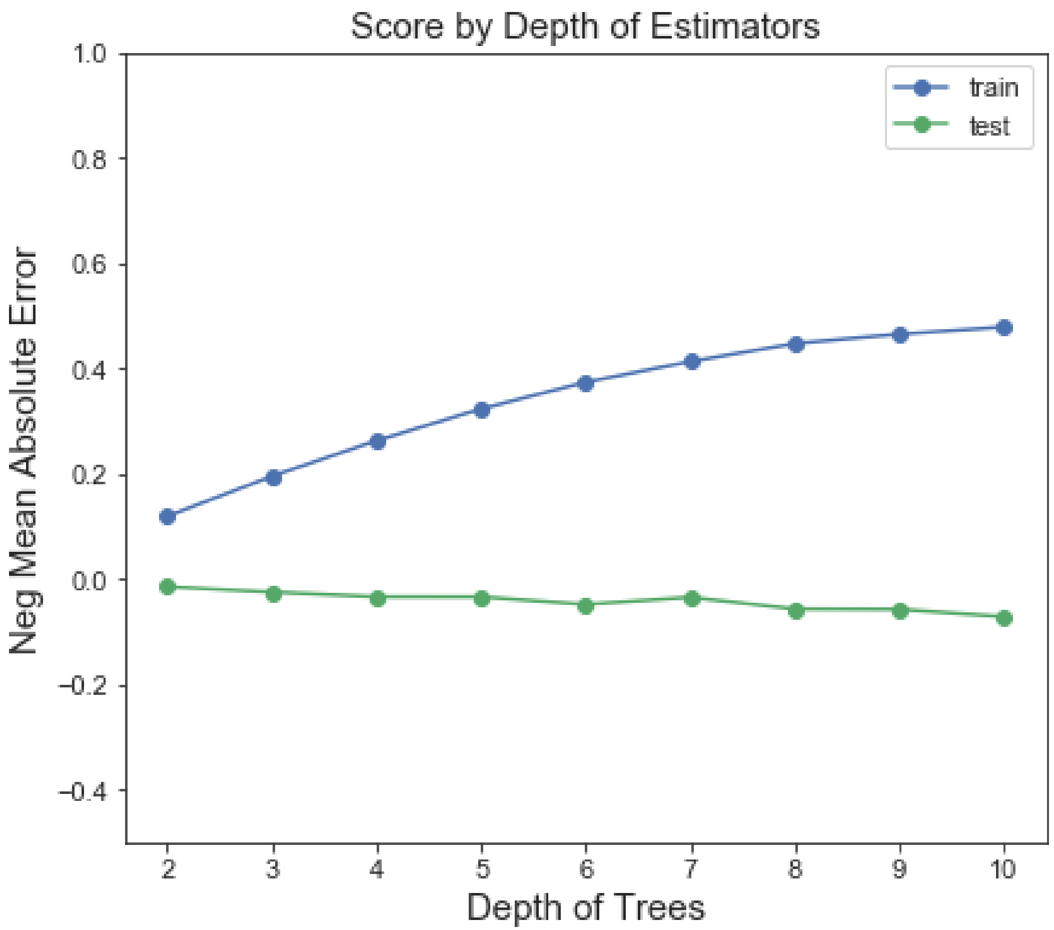




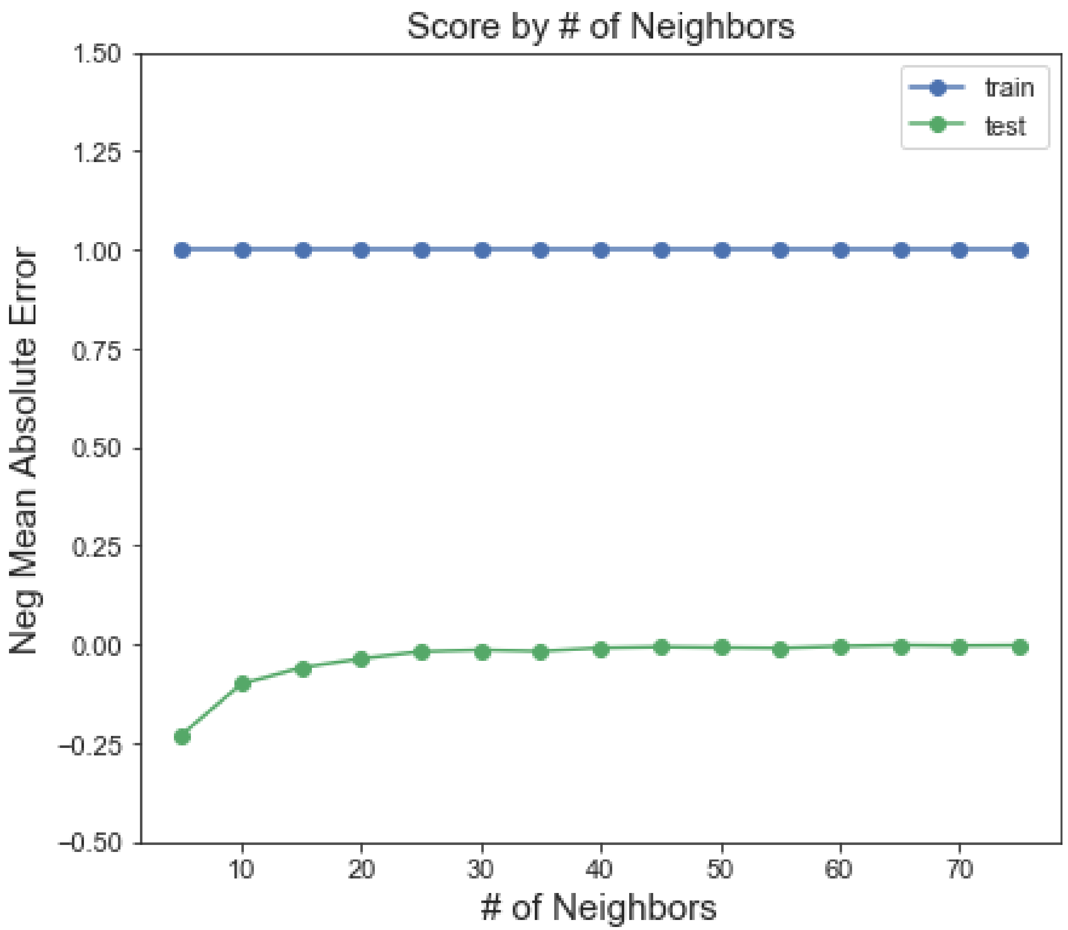
Predicting Social Skill

4) Can other Regressor Types do better? Grid searching Random Forest, KNN Regressors

Random Forest Regressor



KNN Regressor

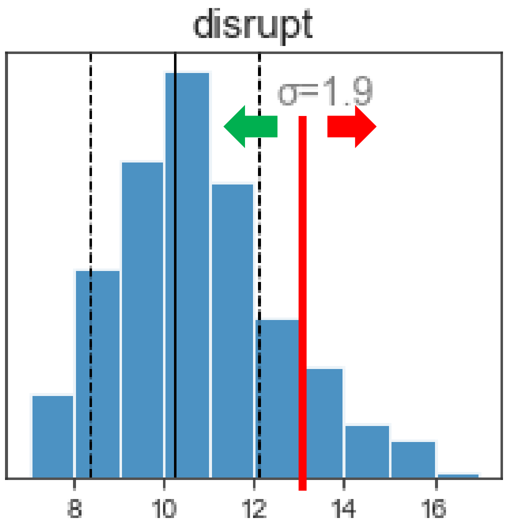


Topics

1. *Can we predict a child's "Social Skill" score be predicated based on demographics and activities (screen time, outdoor time)? – Not with these predictors*

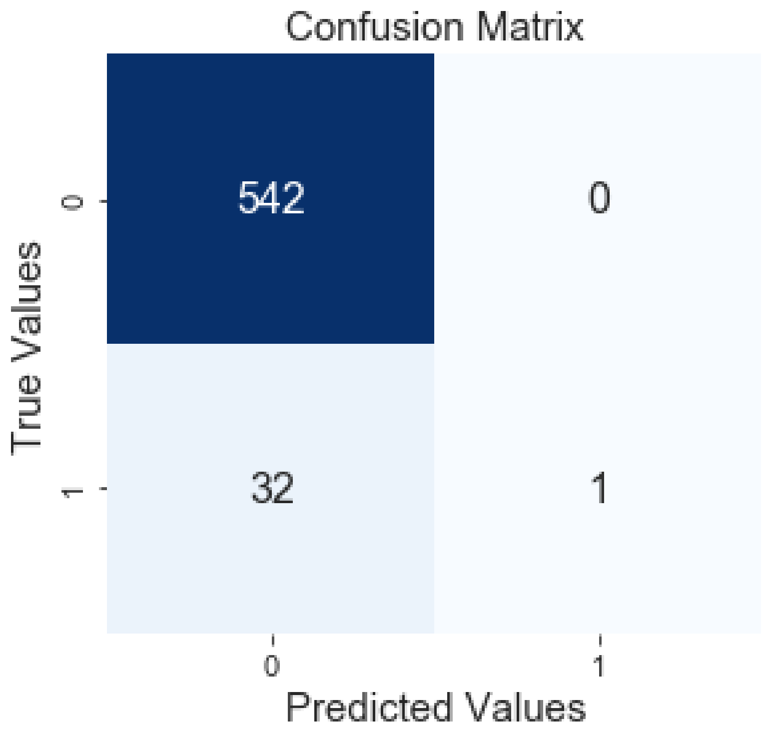
2. Can we predict whether a child will show disruptive behavior?
   - Sidebar: What does PCA do (visually) when we create components out of binary variables?

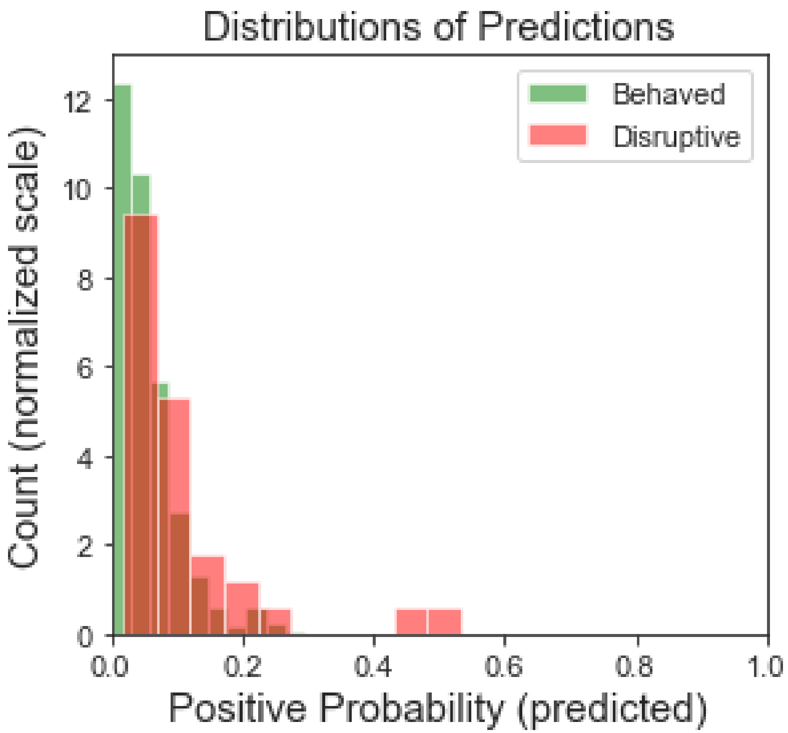
Predicting Disruptive Behavior

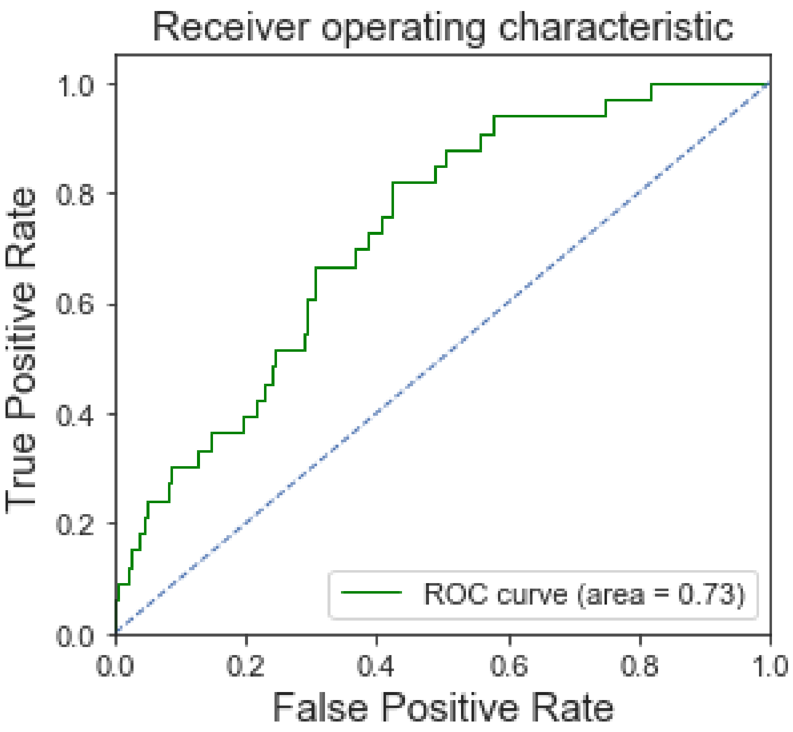


Example: A school administration is trying to prevent classrooms from having too many disruptive students, so they are trying to identify which of their incoming students are likely to be disruptive in advance (disrupt score > x)

<u>Basic Logistic Regression</u>







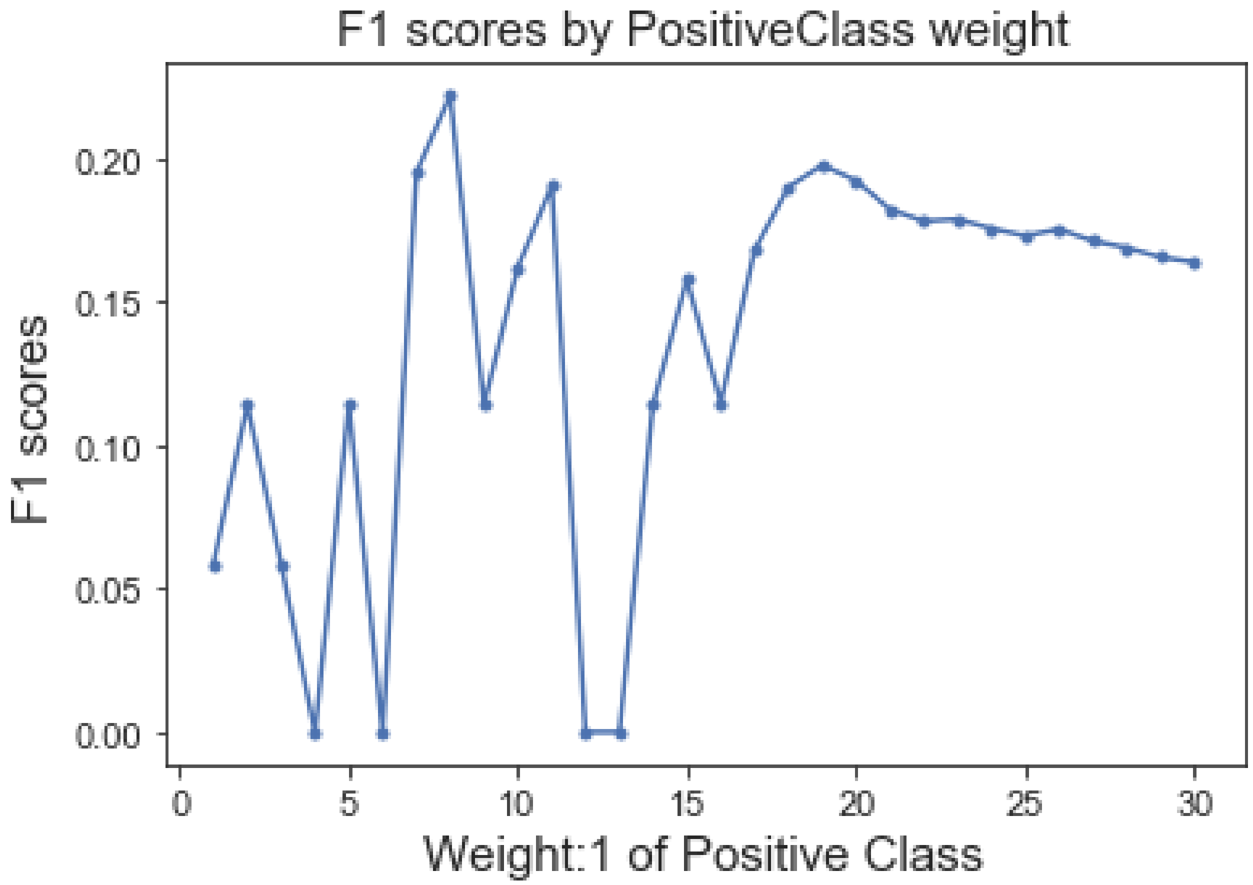
Predicting Disruptive Behavior

Search for the correct class balance, then retry:

$$precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

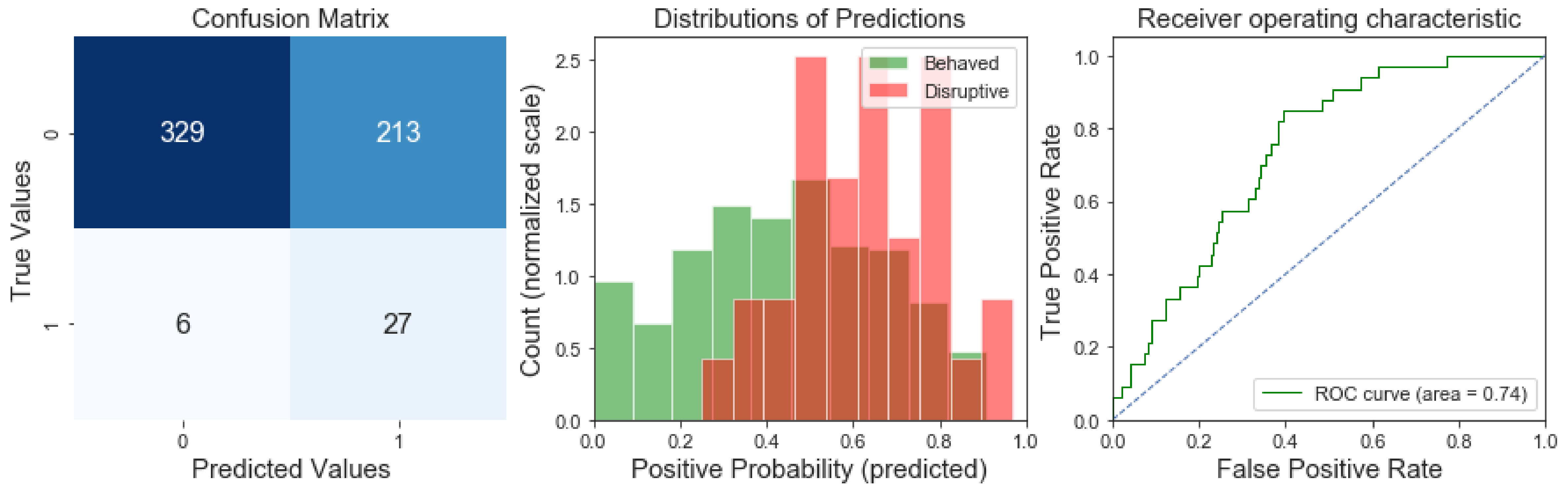
$$F1 = \frac{2 * precision * recall}{precision + recall}$$



Bottom line – F1 (aka F-measure or F-score) gives us a single number to understand the confusion matrix, enabling grid searching that doesn't penalize small classes

Predicting Disruptive Behavior

Logistic Regression with Class weights 1:19



At this point, we could tell the school something like this, based on the ROC curve above and choosing a decision threshold:

*"Our current model could correctly identify ~ **85**% of disruptive students; however, it would also result in almost **40**% of well-behaved students being mistakenly labelled as disruptive"*

Feature Analysis

Name	Description	Type
age	Age, ranges 2-6	Continuous Float
express	Child's Ability to express themselves	"continuous" integer (really a sum of 13 Likert scales)
comply	Child's Ability to comply	"continuous" integer (really a sum of 10 Likert scales)
tvTime	Hours watching TV / day	Continuous Float
cpuTime	Hours on Computer / day	Continuous Float
outdoorTime	Hours outside / day	Continuous Float
disability	Whether Mother considers child disabled	Boolean
mothersEdu	Mother's education level: 1 = <10 years, 2 = 11-13 years, 3 = 14+ years	Discrete
meetStReqs	Whether child meets government recommendations for screen time	Boolean
meetPhysReqs	Whether child meets government recommendations for outdoor time	Boolean
gender	Male/female	Boolean

Creating components from these continuous variables removes any multicollinearity, but doesn't change model performance…

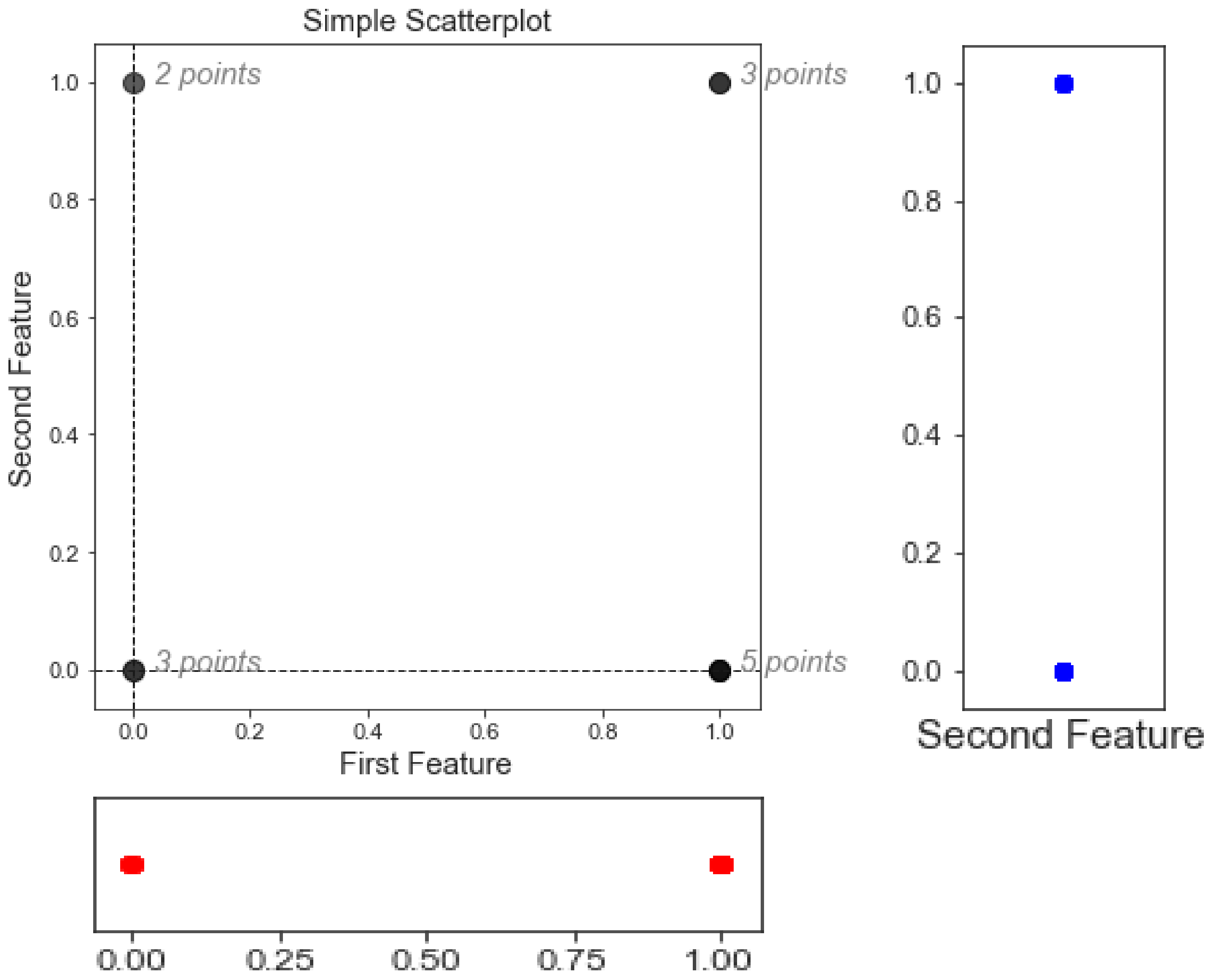
What would PCA do to our Boolean variables?

Sidebar: Visualizing PCA for Boolean Variables

	one	two
0	1	1
1	0	0
2	1	1
3	1	1
4	0	0
5	0	0
6	0	1
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	0	1

How would these look on a scatter plot?

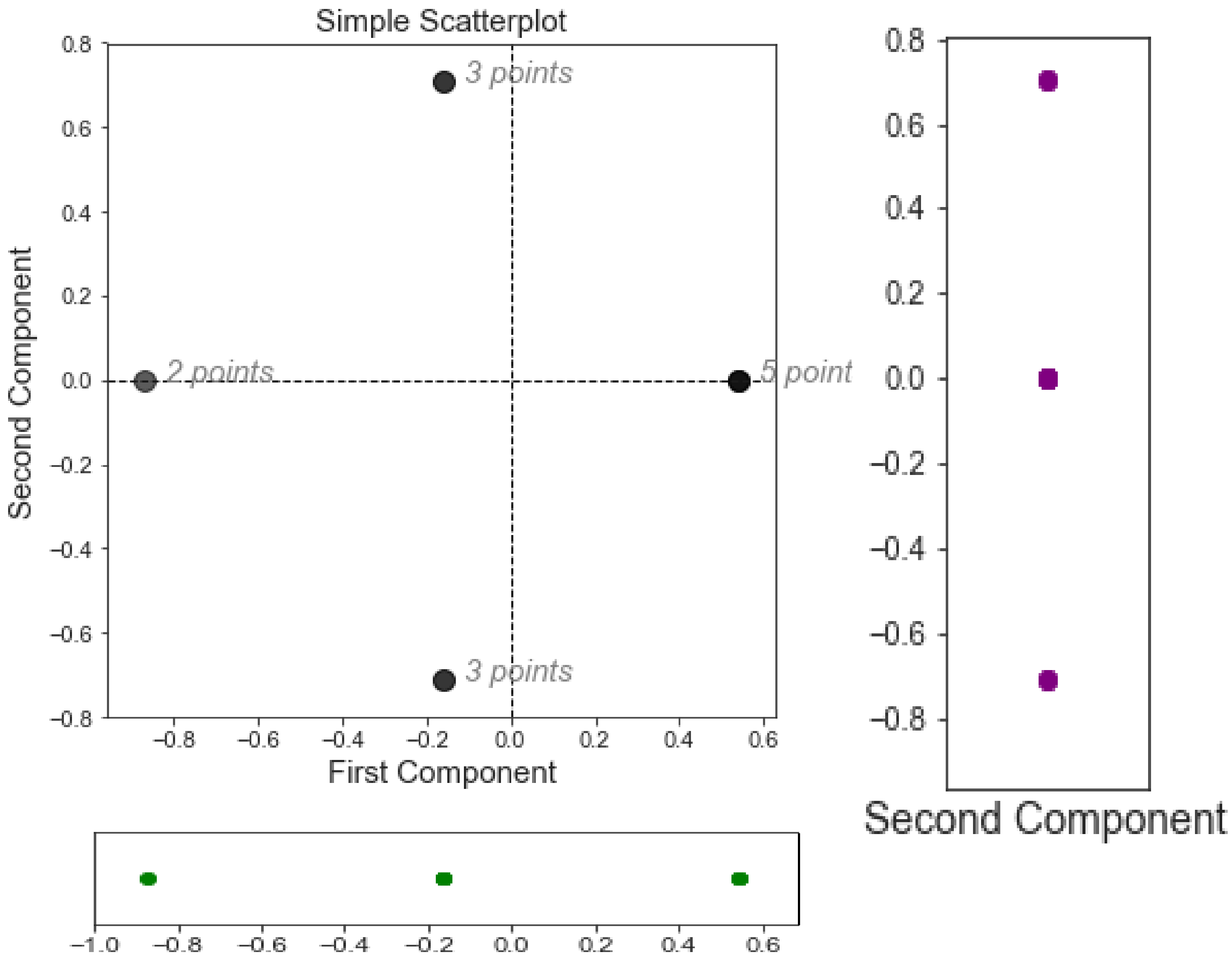
We'd expect just 4 clusters of points:



Sidebar: Visualizing PCA for Boolean Variables

	one	two
0	1	1
1	0	0
2	1	1
3	1	1
4	0	0
5	0	0
6	0	1
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	0	1

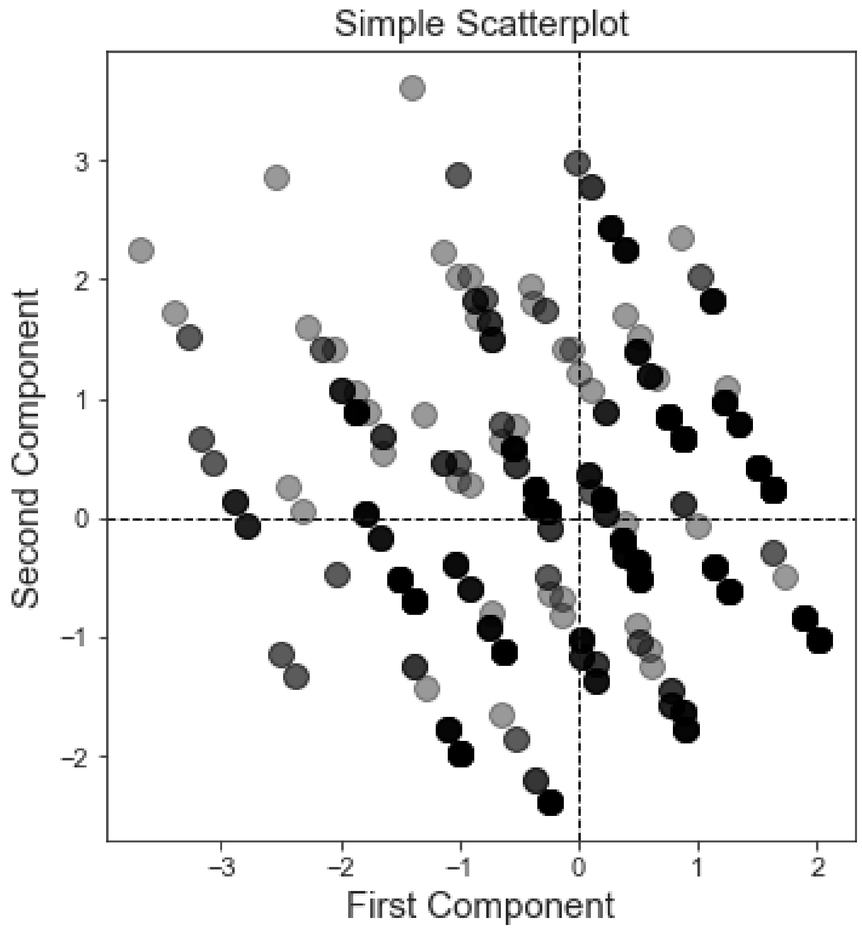
What if we ran PCA on these 2 features with mean-only scaling?



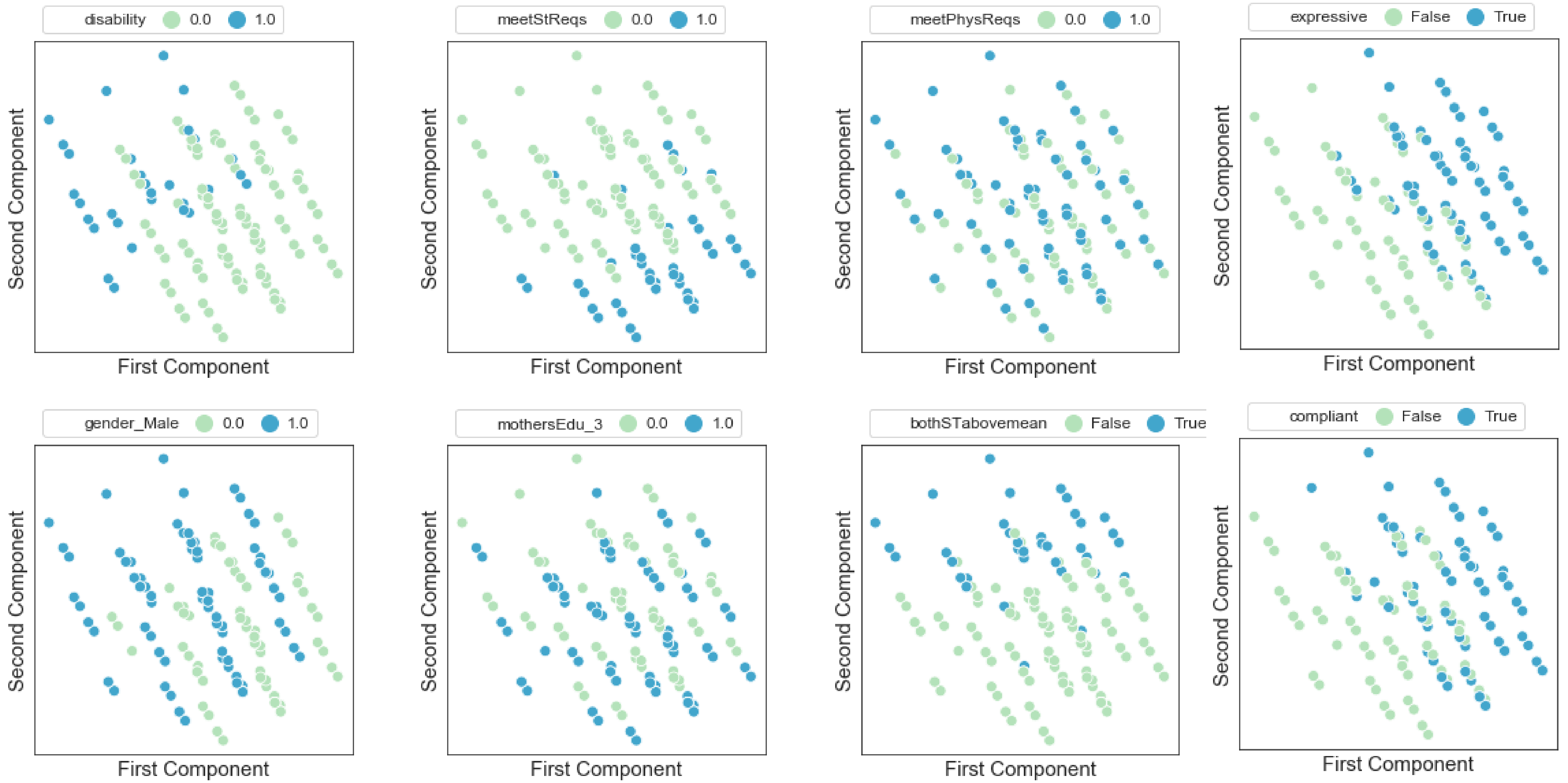
Back to our Data…

Features Used:	True if:
Disability	Mother considers child disabled
meetStReqs	Child meets gov't recommendations for screen time
meetPhysReqs	Child meets gov't recommendations for outdoor time
Gender_Male	Child is male
MothersEdu_3	Mother educated 14+ years
bothSTabovemean	Both tv time and cpu time over the mean
expressive	Above average expressiveness
compliant	Above average compliance

First 2 components created by PCA on 8 Boolean features



Back to our Data…



Predicting Disruptive Behavior

Logistic Regressor Summary	n Features	Class Weights	F1Score	AUC
Basic	15	1:1	.06	.73
Balanced	15	1:16	.12	.65
"Optimized" Weights	15	1:19	.20	.74
Ridge	15	1:19	.20	.74
With continuous components	14	1:19	.20	.74
With continuous & binary components	12	1:7	.24	.73

Appendix

allSocialSkills relationship to its precedents

```
regr = linear_model.LinearRegression()
x = df[['express', 'comply', 'disrupt']]
y = df.allSocialSkills
regr.fit(x,y)
scores = cross_val_score(regr,x,y,cv=10)
print("Fold Scores: ",scores)
print("\nAverage Score: ",np.mean(scores))
```

```
Fold Scores:  [0.98852157 0.99074261 0.98442617
 0.98661701 0.99385063 0.9916524  0.99170743]

Average Score:  0.9910974050546434
```