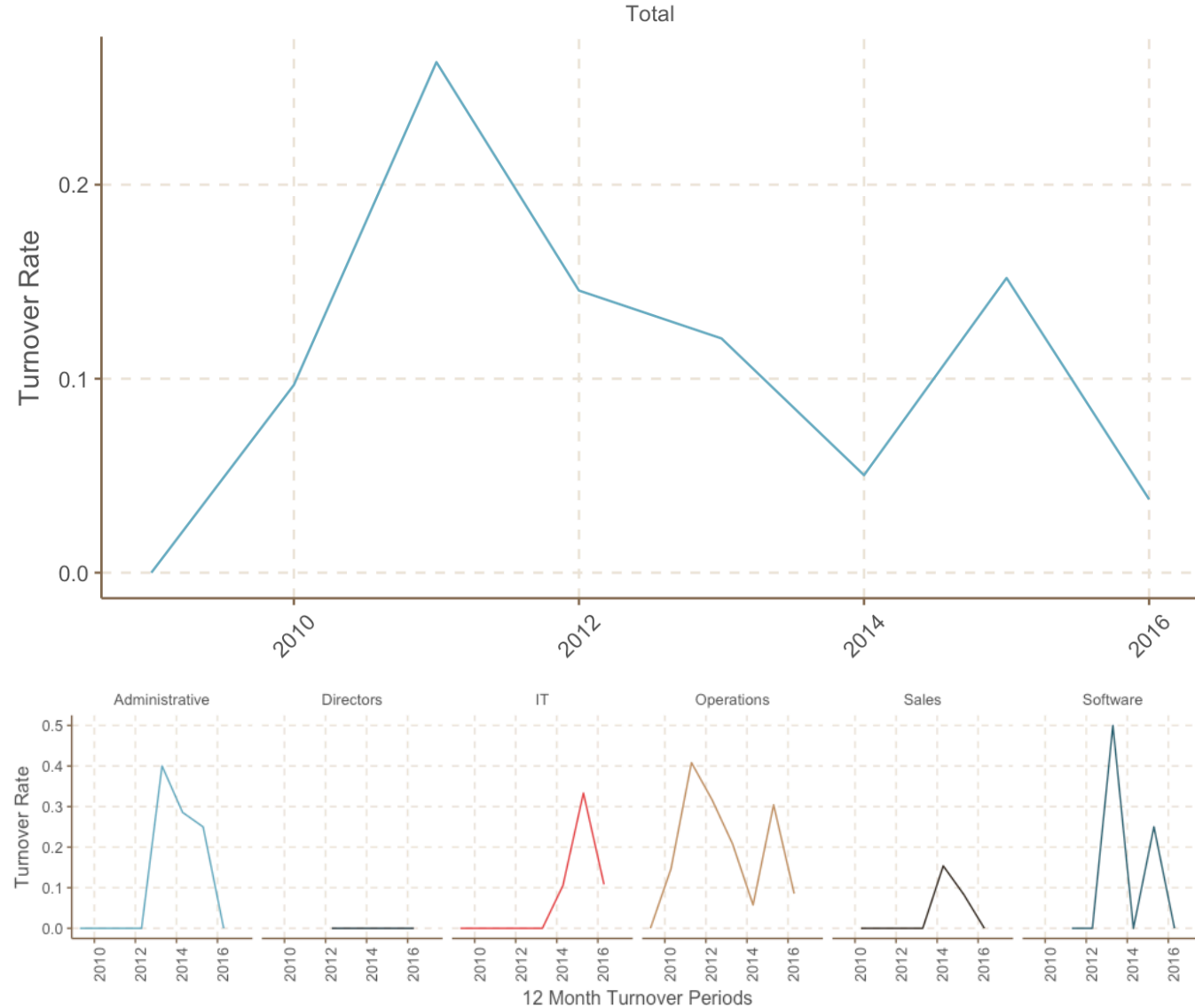


NOSYD: TURNOVER BROKEN DOWN

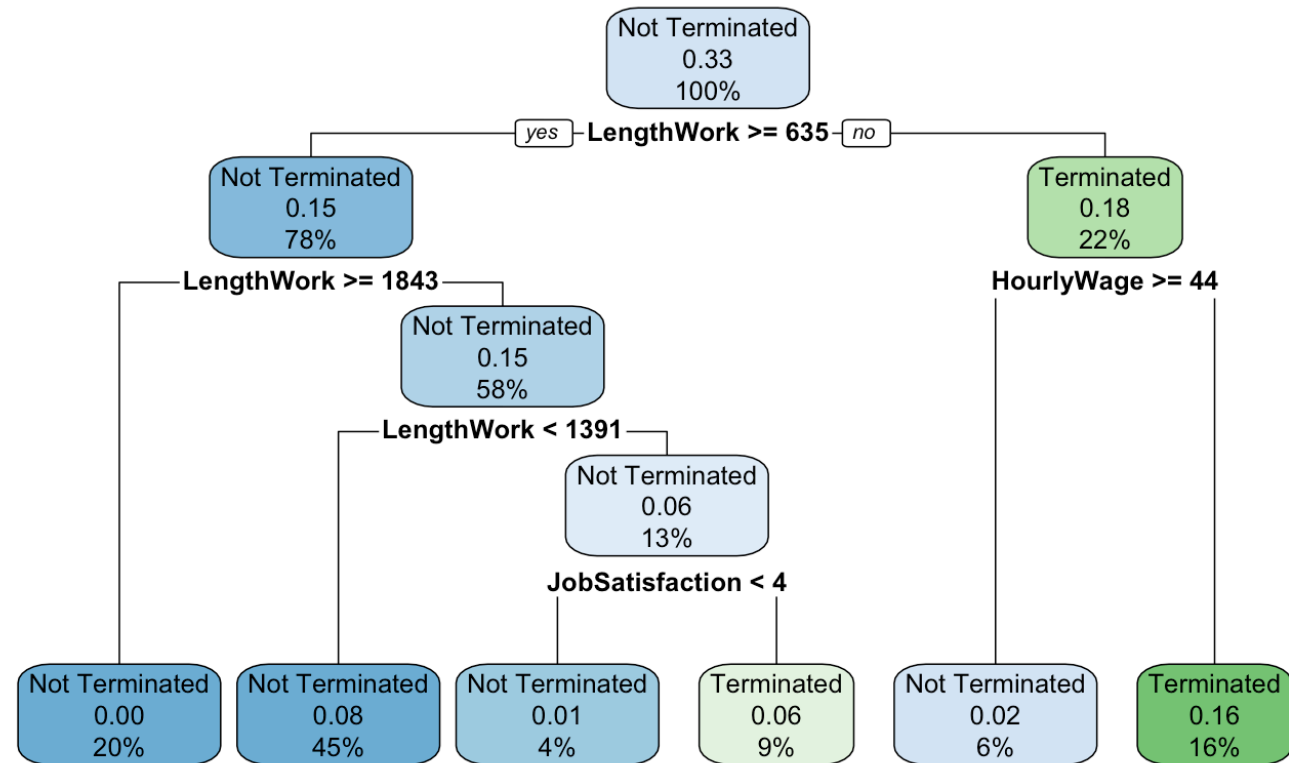
- Today's Date: April 2017
- Date Periods: April, last year, to April, this year
- Turnover rates are down
- Forecast Predictions

Turnover Rates from 2009 - 2016



PERSON-LEVEL TERMINATION PREDICTIONS

- Decision Tree
- Factors of importance
- Inferring causality?
 - Logistic regression
 - Two stage least squares



COMPANY STRUCTURE

- Department frequencies
- Total: 185 people
- Length of stay
- Average job satisfaction by department

Frequency* Distribution of Employees by Department



*Frequency distribution does not include new hires who have not yet begun working

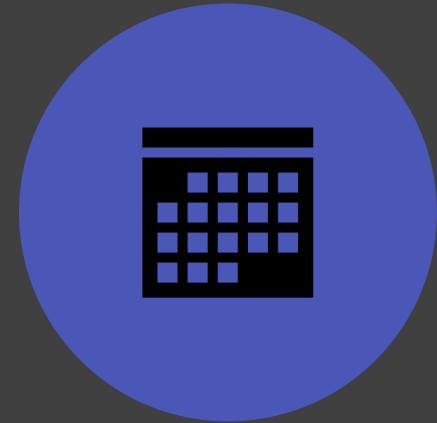
NEXT STEPS



SURVEYS



PANEL DATA



RERUNNING MODELS
WITH NEW TIME PERIODS

APPENDIX: RANDOM FORESTS

- The random forest classifier (RFC) was able to properly predict about 71% of the test data set correctly; while 71% is not laughable, RFC is known for over-fitting, so the results should be interpreted with a grain of salt. Additionally, while I initially thought that it would be smart to run a principle components regression or exploratory factor analysis for dimensionality reduction, it would honestly not be that useful because much of the data, based on my assumptions of relevance, is necessary. A future possible step could be to run a hierarchical logistic regression and see if the main variables (i.e., job satisfaction, performance, hourly wage, length of work, and engagement) will actually explain a statistically significant amount of our variance after controlling for all other factors.
- Additionally, our RFC was not as useful as I had hoped. Ultimately, we have a basic idea, assuming that the RFC is true, that length of work and hourly wage are the strongest classifiers for whether someone will be terminated or not. RFC results suggest that there is a 16% likelihood that people who have worked for over about 2 years and make at least \$44/hr will be terminated. Again, the problem with our RFC is that our sample is still very small, and we have no way of controlling for time effects. That being said, our RFC also suggests that people who have worked at Nosyd for less than 635 days have a 46% chance of leaving - all based on length of stay at Nosyd and no other nodes.

APPENDIX: 2SLS LOGISTIC REGRESSION

- While our RFC had a better prediction rate than our logistic regression (71% vs 41%), the logistic regression's independent variables are honestly not as useful as I had originally hoped. That is why I ran a two stage least squares logistic regression - to see if we could draw out causality.
- Assumptions
 - While I was not specifically provided information on what type of job fair people attended, I am assuming that this job fair is a general job fair at which Nosyd attended and not a Nosyd-specific fair. With this assumption in mind, that means that people who attended the job fair have a better mental framework and model around which the company works, allowing them to choose to work at this company. With more information to build their theoretical framework, people now have a better idea of the company beforehand and will choose to work there if their values and culture align, which is why attending the job fair and job satisfaction covary with each other.
 - Excludability basically means that termination and attending the job fair do not interact directly. I believe this to be the case because attending the job fair has no inherent impact on work flow. While some may say attending job fairs interact with work ethic, performance can be a proxy for work ethic - which is captured in our model. From our 2sls, we discover that a decrease in the job satisfaction rating of 2 points will increase the probability of being terminated 100%. This makes sense: if someone is dissatisfied with his or her job, then he or she may leave voluntarily for other opportunities or may be fired because of a decrease in "good work."