

# Survey of Machine Learning Methods and Resources

John D. Lee and Linda Ng Boyle

10/28/2019

# Machine Learning Next Steps

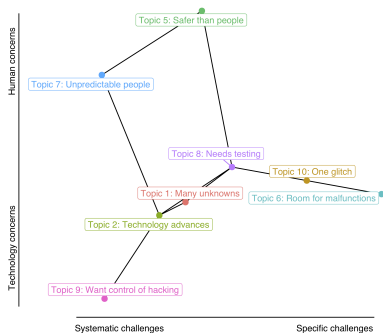
- Machine learning challenges and opportunities
- Other machine learning techniques
- Machine learning resources

Friday 8:30 Grand Ballroom C, Human in Focus: Feature research and application of ubiquitous user monitoring

# Machine Learning Challenges and Opportunities|Issues that behavioral scientists can help address

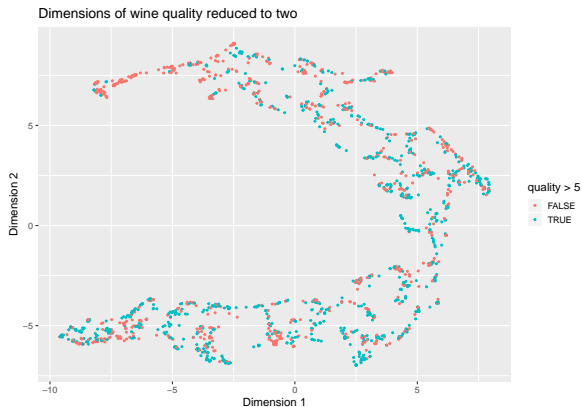
- Representative sampling
- Differential costs of prediction errors
- Tradeoff of model understandability and performance
- Explaining model output (LIME)
- Ethical considerations
- Design thinking and systems thinking

# Machine Learning Techniques|Text analysis: Quantitative analysis of qualitative data



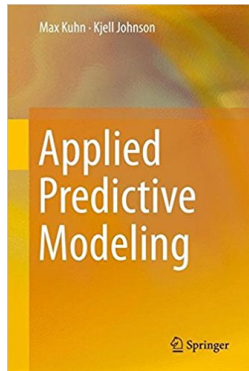
Lee, J. D., & Kolodge, K. (2019). Exploring trust in self-driving vehicles with text analysis. *Human Factors*.

# Machine Learning Techniques|Visualizing multidimensional data



## Machine learning resources: caret

- Comprehensive online description of caret features (<http://topepo.github.io/caret/index.html>)
- Early draft of book on feature engineering by the author of caret (<http://www.feet.engineering>)



# Machine Learning Resources: Keras

- Simple interface to powerful deep learning algorithms
- Support for prototyping through deployment
- Implement and deploy in an afternoon



<https://keras.rstudio.com>

## INSTALLATION

The keras R package uses the Python keras library. You can install all the prerequisites directly from R.

[https://keras.rstudio.com/reference/install\\_keras.html](https://keras.rstudio.com/reference/install_keras.html)

```
library(keras)
install_keras()
```

See ?install\_keras  
for GPU instructions

This installs the required libraries in an Anaconda environment or virtual environment 'r-tensorflow'.

## TRAINING AN IMAGE RECOGNIZER ON MNIST DATA

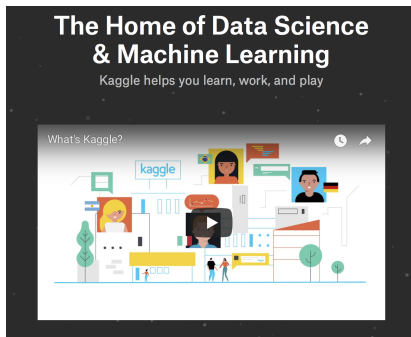
# input layer: use MNIST images

```
mnist <- dataset_mnist()
x_train <- mnist$train$x; y_train <- mnist$train$y
x_test <- mnist$test$x; y_test <- mnist$test$y
```



# Machine Learning Resources: Kaggle

- Data—over 10,000 datasets
- Competitions
- Kernels—code developed for analysis of datasets hosted on kaggle



<https://www.kaggle.com>



# Machine Learning Resources: DataCamp

- Online courses with integrated exercises
- Including basic R and advanced courses
- Machine learning toolbox by the author of caret
- Writing functions by author of the tidyverse



<https://www.datacamp.com/courses/machine-learning-toolbox>

# Workshop Summary

Human Factors contributions to ML and ML contributions to Human Factors

- Data visualization: Don't venture into uncharted territory
- Data reduction and cleaning: 80% of any data analysis
- Multi-level general linear models: Model individual variability
- Machine learning overview and implementation: Adopt different statistical mindset
- Survey of machine learning behavioral issues and applications: Design for people

Repeatable data analysis with R and RStudio

Friday 8:30 Grand Ballroom C, Human in Focus: Feature research and application of ubiquitous user monitoring