Next Steps

Introduction to Data Science
Nina Zumel
John Mount

Next steps

- Find a free data source

- Ask an interesting question

- Try to make your own predictive model

A *few* example sources

UCI Machine learning repository

- One of the first machine learning data set repositories

- https://archive.ics.uci.edu/ml/

Journal of Statistics Education

- Well chosen data sets for teach statistical concepts

- http://www.amstat.org/publications/jse/jse_data_archive.htm

KD Nuggets list

- A big list of lists!

- http://www.kdnuggets.com/
  datasets/index.html

Try to find more on your own

- Governments (countries, cities)

- NGOs

- Industry groups

- Machine learning interest groups

- Academic sources

- and more

One way to form your own interesting questions

- Find a fact that is valuable and hard to measure in a given data set

- Using the methods outlined in this course: see if other facts in the data set that are easy to measure can be used to make a good predictive model

Thank you