

Lecture 1: Introduction to Introduction to Data Science

Vincent Toups

2020-06-10 Tue

Outline

Lecture 01 - Introduction to the Course

- About Me

- About This Course - Introduction to Data Science

- Goals

- Today: Our Basic Environment and Tools

- ▶ Dr Vincent Touns
- ▶ Neuroscientist by way of the UNC Physics Department
- ▶ Software Engineer and Data Scientist
- ▶ Programming Language Enthusiast / Weirdo
- ▶ Data Operations Guy

- ▶ What I did at my last job
- ▶ What I do at my current job
- ▶ Data Science vs Statistics Perfectly illustrated by the difference between Neural Networks and Regression. Data Scientist is to Neural Networks as the Statistician is to Linear Regression that is
 - ▶ the former wants to extract actionable predictions in a probably non-mission critical setting and isn't particularly concerned about putting a fine point on the interpretation
 - ▶ the latter is more precise, interested in quantifying power, interpretation of models, laying out the exact assumptions that go into an analysis

- ▶ there is overlap
- ▶ I am a data scientist - you will be shocked, I'm sure, about what I don't know about statistics.
- ▶ Data Scientists are also pretty heterogeneous. I have worked with physicists, mathematicians, statisticians, and software engineers who are all called data scientists.
- ▶ Opportunistic skill accumulators: a good data scientist will know some statistics, machine learning, software engineering, dev ops, and be comfortable (like a statistician) becoming familiar with new domains.

This course is about giving you a taste of that opportunistic skill set.

Its also about affirmation: you can do this stuff. The most crufty linux shell wizardry is no more complicated that bonafide biostatistics and probably at least half as useful.

1. Comfort with Developer Tools Docker, Make, Git, Linux
2. Survey of Data Scientific Techniques and How to Understand them

3. Deeper Appreciation of The Art of Programming

1. Demo - Compute
2. Demo - Storage
3. Demo - Docker
4. Demo - Git
5. Demo - Rstudio
6. Demo - Human and Dog Names