Next Steps

BSDS 100, Spring 2021
Michael Ruddy

What can I do after this course?

- Import, wrangle, and visualize tabular data and create basic models using R programming language

What can I do after this course?

- Import, wrangle, and visualize tabular data and create basic models using R programming language

- What about other programming languages for data science?!

What can I do after this course?

- Import, wrangle, and visualize tabular data and create basic models using R programming language

- What about other programming languages for data science?!
    - Python in Jupyter Notebook
    - tibbles -> pandas dataframes
    - ggplot -> matplotlib
    - relational data -> SQL
    - basic modelling -> scikit-learn

What can I do after this course?

- Different routes to go with data science

    - Statistics

    - Machine Learning

    - Deep Learning

    - Software Engineering w/ Data Science twist

    - Social Sciences (Data Ethicist?)

What can I do after this course?

Different options:

- Formal Programs

    - Courses at USF

    - MSDS programs

    - Bootcamps

- Self Study

    - Textbooks available through USF

    - Free(-ish) online courses

    - YouTube (Stanford lectures)

- Important component: Portfolio with Project(s)

What can I do after this course?

- USF Courses

    - BSDS 200: Python and SQL (prereq CS 110 & conc. MATH 230)

    - [Data Science Major](Math, Stats, and CS tasting menu)

- MSDS Courses

    - Usually 1-2 year programs, background can be varied

    - [USF MSDS Program] is 1 year

- Bootcamps

    - Variety of types: do your research before paying them any $$$

What can I do after this course?

Books

- ggplot2 by Hadley Wickman

- 24.4: Suggestions for modelling in R

- Data Science from Scratch

- Variety of O'Reilly books (deep

  learning for life sciences, NLP,

  PyTorch, etc.)
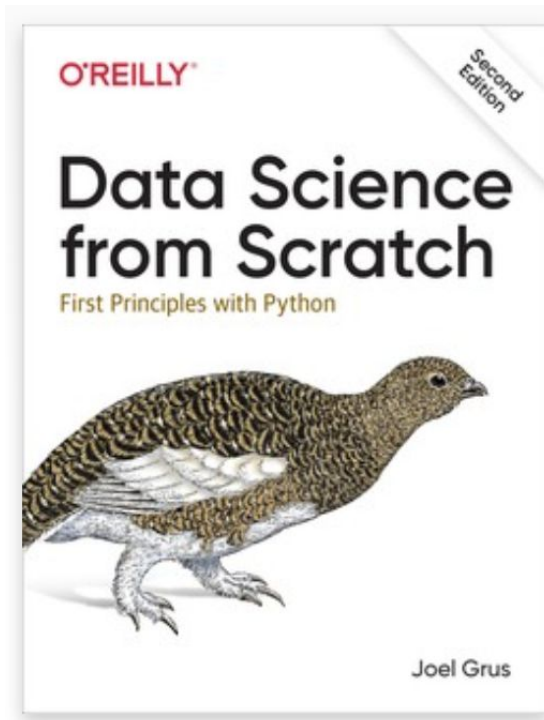
What can I do after this course?

Online Free(-ish) Self-Study Materials (from my own experience)

- Kaggle [mini-courses](#) (Good for quick intro to a lot of topics)

- Coursera ([Andrew Ng's Deep Learning Specialization](#))

  - No Prereqs, very detailed and clear explanations

- [fast.ai](#) courses

  - USF's own Jeremy Howard/Rachel Thomas

  - Great lecturers

- Stanford online [YouTube lectures](#)

What can I do after this course?

Good Statistics resources (from what I've heard)

- [Online Penn State](#)

- [An Introduction to Statistical Learning](#)

Other Stuff

- The following online Notebooks will let you use so many GPU hours (essential for Deep Learning) per week:
    - Kaggle Notebooks
    - Google Colab
- Cocalc is an online R notebook
- Learning how to use git/GitHub is **ESSENTIAL** for any communication involving code
    - version control and open source projects

Some Caution

- Data Science is a hot field right now (though not everyone understands it well)

- Big data is here to stay, but the work required is always changing

- Be careful of those offering you "quick" and "easy" ways to learn data science or get a data science job, especially if they are asking you for payment in exchange

Questions?

- Please feel free to ask me any questions, even after this course about how to study Data Science as a field

- Please fill out my Evals

- Thanks for a great class (still more for you to do though!)

    - Case Study 3: Tomorrow

    - Final Project Description due TODAY

    - Final Project (next slide)

Info about your Final Project

- Please get your Final Project Description to me TODAY!

- Every group will have 10 minutes to present

    - Presentation must include R code (doesn't have to be live)

- You must also submit a Jupyter Notebook that serves as the "essay" for your project

    - Must include Honor Statement (Same submission style as Case Study)

    - Must be 3rd-party readable!