

Software Design for Data Scientists

Course Introduction, Command Line & Data Essentials

Bernease Herman^{1,3}, David Beck^{1,2,3}, Joseph Hellerstein^{1,3},
Daniel Jones³

¹eScience Institute

²Chemical Engineering

³Computer Science

University of Washington

September 25, 2019



What's the class about? Who are we?

- Objectives
 - Teach how to create and collaborate on data- and computation-intensive research projects
 - Provide practical software skills for data analysis in research & industry.
 - Elevate coding in science to the level of technical writing.
- Instructor introductions



Agenda

- Why data science?
- Course overview
- Pronto data
- Getting data with shell scripts



Why Data Science?



MAY 28, 2013 @ 09:09 AM 103,702 VIEWS

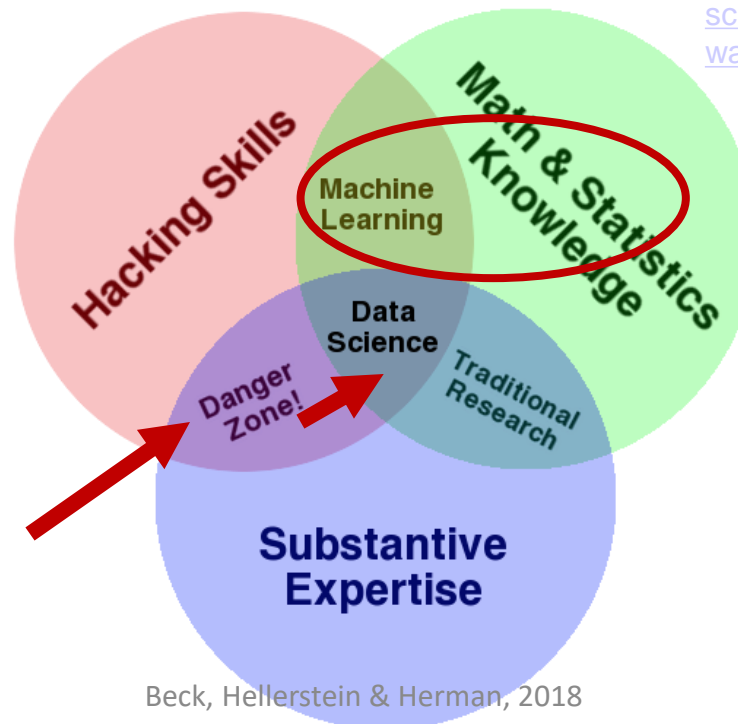
A Very Short History Of Data Science



Gil Press
CONTRIBUTOR

The story of how data scientists became sexy is mostly the story of the coupling of the mature discipline of statistics with a very young one—computer science. The term “Data Science” has emerged only recently to specifically designate a new profession that is expected to make sense of the vast stores of big data. But making sense of data has a long history and has been discussed by scientists, statisticians, librarians, computer scientists and others for years. The following timeline traces the evolution of the term “Data Science” and its use, attempts to define it, and related terms.

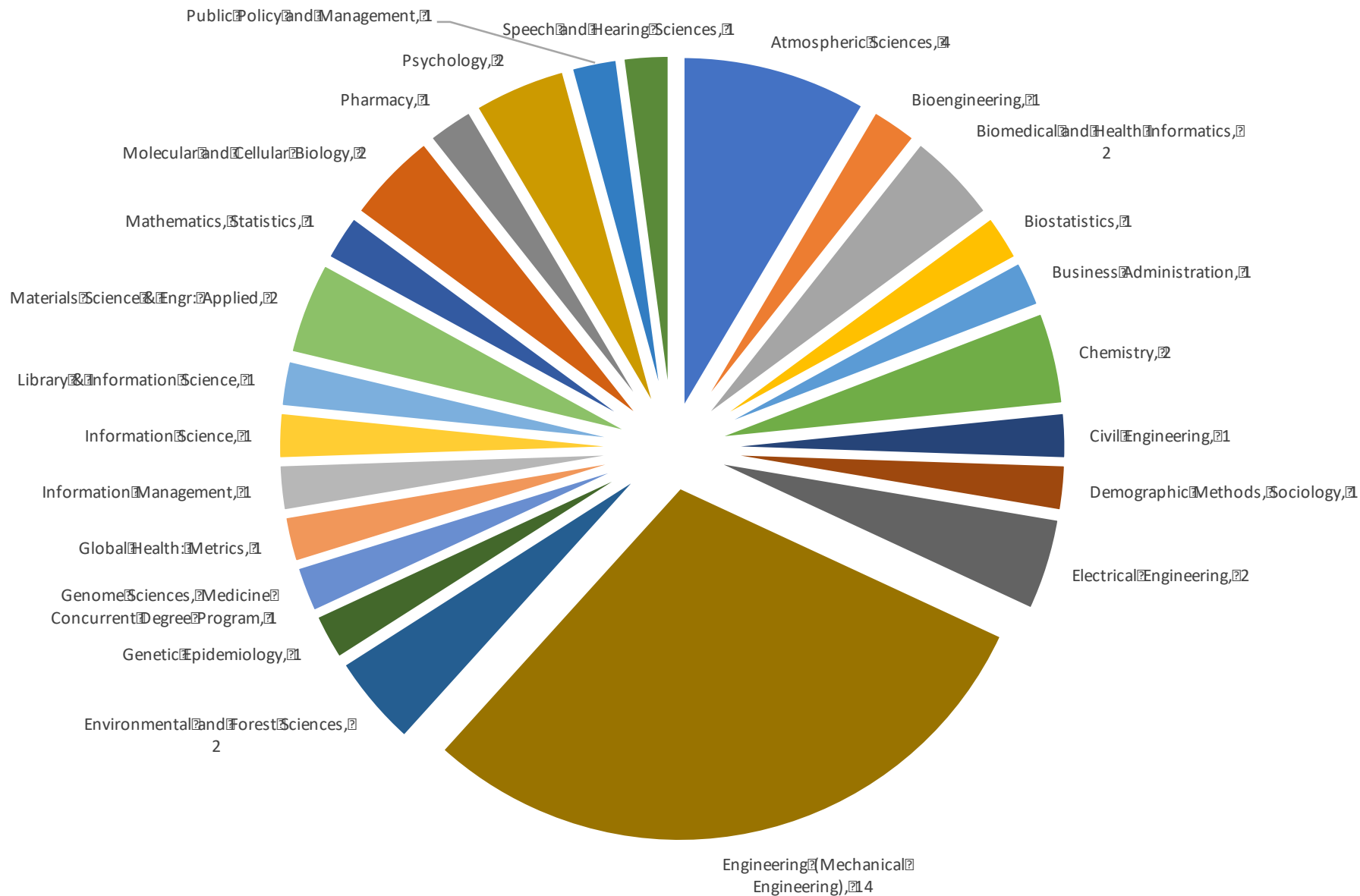
<https://escience.washington.edu/data-science-courses-at-the-university-of-washington/>



Beck, Hellerstein & Herman, 2018



Fall 2018, Degree Program Enrollment



Course Overview



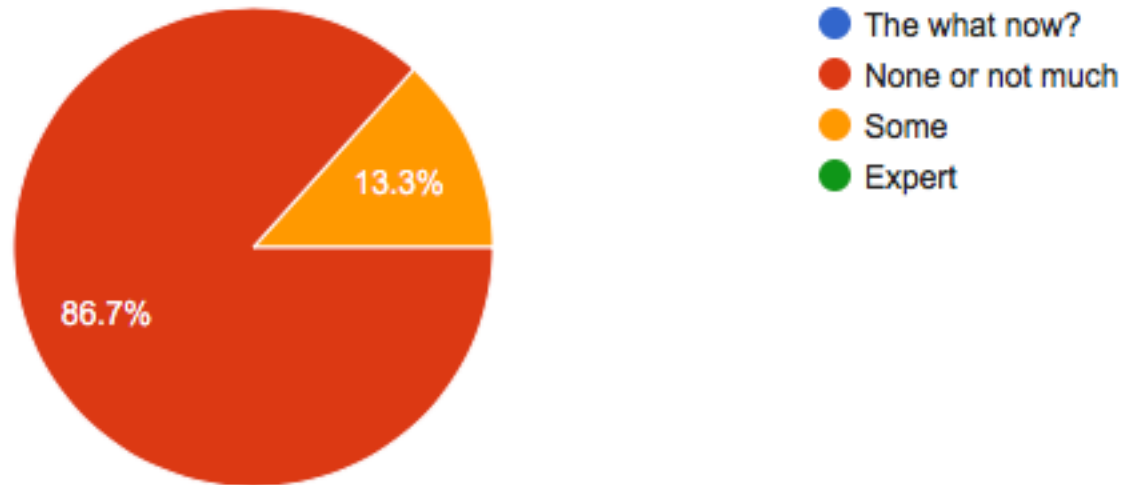
Survey Results

Question	Response
Years of programming?	Mostly > 1 yr
Years of python?	< 1 yr
Experience with a text editor	Mostly "Yes"
Comfort with if-statements?	Yes
Comfort with for-statements?	Yes
Comfort with functions?	Mostly "Yes"
Python packages (scipy, pandas, ...)	Mostly "No"
Experience with iPython?	Very little
Experience with github?	Very little



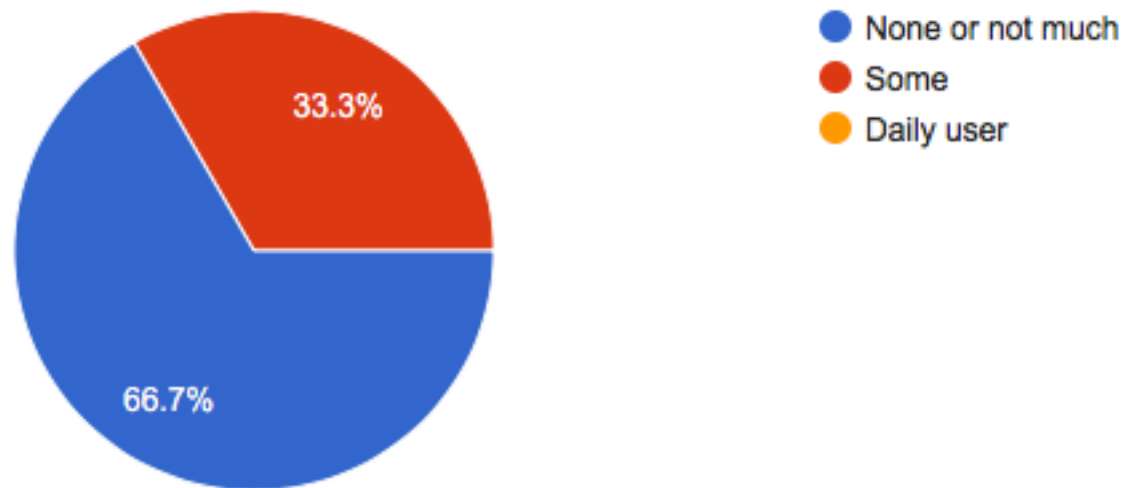
Survey Results

Do you have any previous experience with the bash shell? (15 responses)



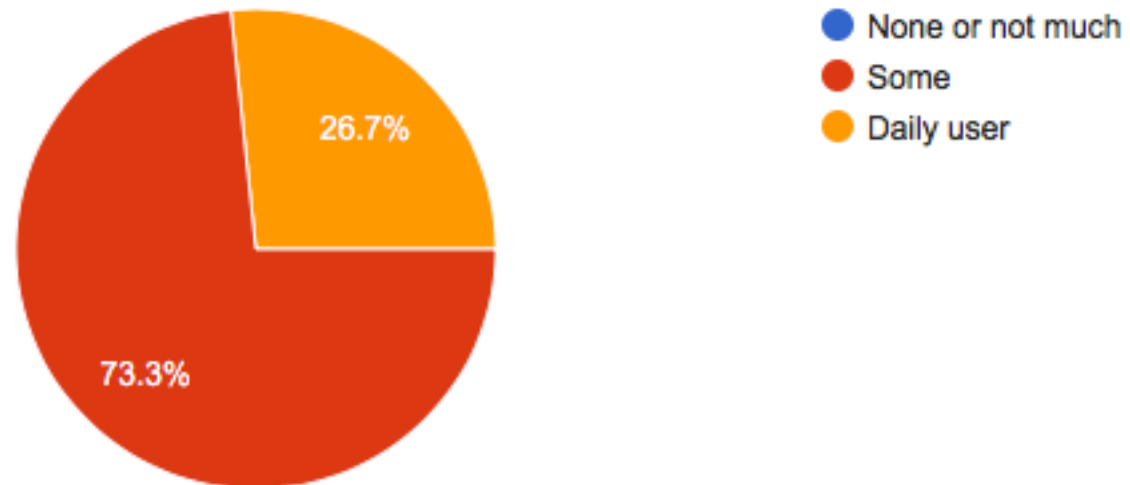
Survey Results

Do you have any previous experience with version control? (15 responses)



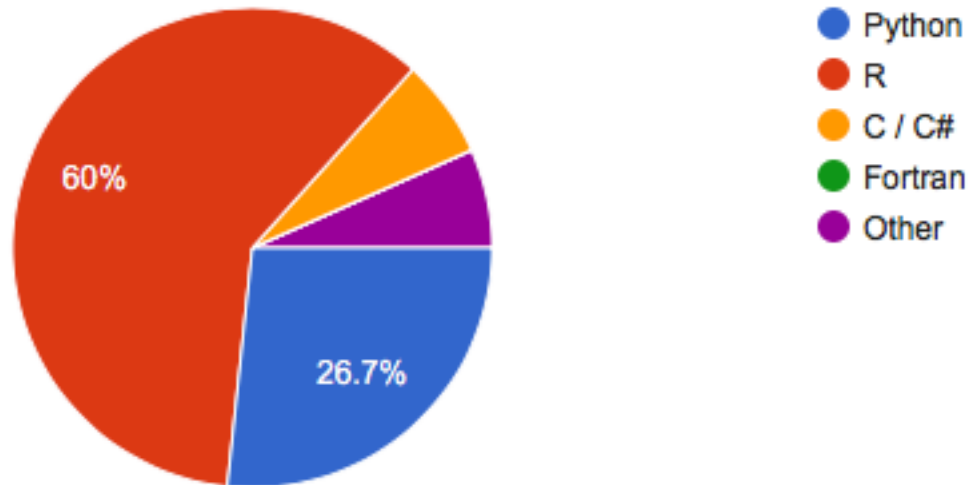
Survey Results

Do you have any previous programming experience? (15 responses)



Survey Results

If yes to the above, what languages have you used? (15 responses)



Skills Taught

- Program in python using the Python scientific stack, including numpy, pandas, and matplotlib.
- Search, evaluate, and integrate into a project externally developed Python packages; create your own Python packages.
- Develop unit tests that validate important aspects of the project implementation.
- Develop software that it can be used by others including: shared code on github, documentation, installing packages, setup, and running computational studies.
- Create technical specifications for what a program should do and how this is accomplished.



Course Structure

<http://uwseds.github.io/syllabus.html>

- Programming basics (3 weeks)
 - Version control, Python, data manipulation
- Software development (3 weeks)
 - Debugging, documentation, design, collaboration
- Data manipulations in detail (2 weeks)
 - Visualization, machine learning
- Software engineering practicum (2 weeks)



Programming vs. Software Engineering

Analogy: What is the difference between the following kinds of writing:

1. Note to yourself
2. An article in the NY Times



Relating Writing to Software

How Learn Skills

Reporter

Writing
quality

Content

Structure

Review

Freshman
English

Composition & literature
classes, professional writer

SW Eng

Code
quality

Features

Design

Testing

Programming
class

Computer Science degree,
Collaborate on a big project



Grading

Item	%
Homework (~7 homeworks)	35
Class participation	15
Class project	50



Grading

- How we grade homework...
 - You submit homework using GitHub (version control tool) by the due date
 - We take up to a week to review the homework and give feedback with a GitHub Issue
 - You have an additional week to revise the homework to address any weaknesses
 - Your final grade for a homework given
 - **There is no reason you can't get 5/5 on all HW**



Course Web Page

<http://uwseds.github.io>

