

Big Data: Data Wrangling Boot Camp

Introduction and Overview

Chuck Cartledge, PhD

16 September 2016

Table of contents I

1 Introduction

2 Overview

3 Administrivia

4 Q & A

5 Conclusion

6 Vita

Big Data: Data Wrangling Boot Camp

We will cover aspects common to all Big Data investigations, including: defining Big Data, surveying tools and techniques for processing Big Data, and visualizing selected aspects of Big Data.

The emphasis of the camp is to understand what is Big Data beyond the marketing hype of the 3Vs of volume, variety, and velocity,



More detailed information at:

<https://www.odu.edu/cepd/bootcamps/data-wrangling>

Things we'll be covering over the next three days:

- Friday

- 1 Administrivia
- 2 What is BD?
- 3 What are the BD Vs?
- 4 Publicly available sources of BD
- 5 BD tools and techniques
- 6 Twitter developer's account

- Saturday

- 1 Software requirements
- 2 Software design

- 3 Compare Python to other languages
- 4 IDE
- 5 Working with the code
- 6 Scalability issues

- Sunday

- 1 Compare R to other languages
- 2 IDE
- 3 Working with the code
- 4 Scalability issues
- 5 Wrap-up

All things related to paper work.

- Parking – front and back without permits
- Breaks – yes we'll have them.
- Lunch – yes places near by: right a main light to “fast food”
- Text book – not necessary, has good ideas, techniques
- Non-credit option
- Credit option – one additional assignment
- Hours – 9AM to 5PM with break for lunch
- Sunday access – yes check in with security
- Soft copies – all presentations, and software are available
- Computer logins and passwords – TBD
- Break room – across hall
- Bathrooms – around elevator

Other things as well.

Q & A time.

“‘The Answer to the Great Question . . . Of Life, the Universe and Everything . . . is . . . forty-two,’ said Deep Thought, with infinite majesty and calm.”

Douglas Adams, The Hitchhiker's Guide to the Galaxy



What have we covered?

- Where we are.
- Where we're going.
- How we'll get there.



Now!! Onto exploring the world of Big Data!

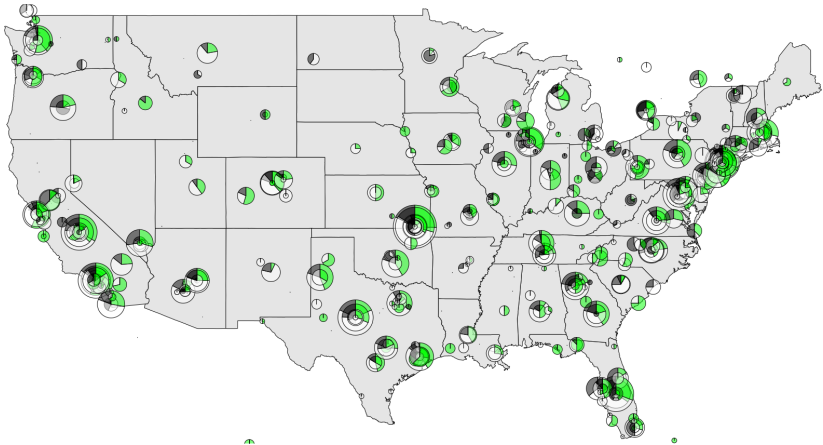
Who am I?

- Father
- Husband
- PhD, Computer Science
- CAPT, USN retired 2004 (31+ years)
- A perennial student
- 1st computer: 1970, donated ICBM guidance computer, machine code, paper/mylar tape, and drum memory



Interests: autonomic systems, real-time applications, distributed processing, long-term preservation of digital data, Big Data

Sentiment by geographic location



Sentiment by sending device

