

# Big Data: Data Analysis Boot Camp

## Introduction and Overview

Chuck Cartledge, PhD

19 January 2018

# Table of contents (1 of 1)

## 1 Introduction

- The global view

## 2 Overview

- The world from 50,000 feet.

## 3 Administrivia

- Miscellaneous and necessary things

## 4 Q & A

## 5 Conclusion

## 6 References

## 7 Vita

# Big Data: Data Analysis 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 data analysis beyond the marketing hype of the 3Vs of volume, variety, and velocity,



Image from [1].

More detailed information at:

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

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

- Friday
  - ① Administrivia
  - ② What is BD?
  - ③ What is R?
  - ④ Looking at the built-in iris and Titanic datasets
- Saturday
  - ① Visualizing data with different packages
  - ② Exploring cluster analysis (of different types)
  - ③ Linear regression and some variants
  - ④ Classification techniques
  - ⑤ Text analysis
  - ⑥ Serial vs. parallel processing
- Sunday
  - ① R limitations
  - ② R and Hadoop
  - ③ R and SQL and No-SQL DBMs
  - ④ Hands-on with real-world crime data
  - ⑤ Wrap-up

# All things related to paper work.

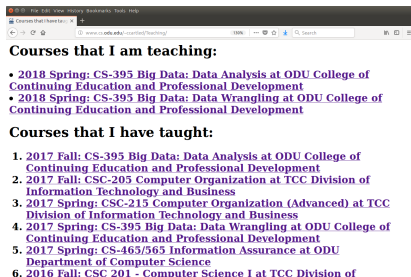
- Parking – front and back without permits
- Breaks – yes we'll have them.
- Lunch – yes places near by: right a main light to “fast food”
- Text books – recommended but not necessary, has good ideas, techniques
- Non-credit option
- Credit option – two additional assignment
- Hours – 9AM to 5PM with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

Other things as well.

# Soft copies available from Internet

- All information (presentations, scripts, and data) is available on your VM desktop (static)
- All information is available via the I'net (dynamic)
- Errata updated nightly

I'm not a web designer, nor do I play one on TV.



The screenshot shows a web browser window with the address bar displaying `www.cs.odu.edu/~ccartled/teaching/`. The page content includes:

**Courses that I am teaching:**

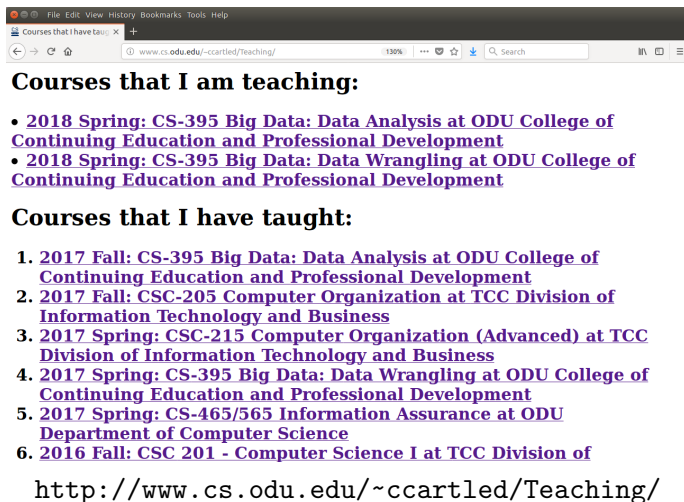
- 2018 Spring: CS-395 Big Data: Data Analysis at ODU College of Continuing Education and Professional Development
- 2018 Spring: CS-395 Big Data: Data Wrangling at ODU College of Continuing Education and Professional Development

**Courses that I have taught:**

1. 2017 Fall: CS-395 Big Data: Data Analysis at ODU College of Continuing Education and Professional Development
2. 2017 Fall: CSC-205 Computer Organization at TCC Division of Information Technology and Business
3. 2017 Spring: CSC-215 Computer Organization (Advanced) at TCC Division of Information Technology and Business
4. 2017 Spring: CS-395 Big Data: Data Wrangling at ODU College of Continuing Education and Professional Development
5. 2017 Spring: CS-463/563 Information Assurance at ODU Department of Computer Science
6. 2016 Fall: CSC 201 - Computer Science I at TCC Division of

`http://www.cs.odu.edu/  
~ccartled/Teaching/`

# Same image.



The screenshot shows a web browser window with the address bar displaying `www.cs.odu.edu/~ccartled/Teaching/`. The page content is as follows:

**Courses that I am teaching:**

- **2018 Spring: CS-395 Big Data: Data Analysis at ODU College of Continuing Education and Professional Development**
- **2018 Spring: CS-395 Big Data: Data Wrangling at ODU College of Continuing Education and Professional Development**

**Courses that I have taught:**

1. **2017 Fall: CS-395 Big Data: Data Analysis at ODU College of Continuing Education and Professional Development**
2. **2017 Fall: CSC-205 Computer Organization at TCC Division of Information Technology and Business**
3. **2017 Spring: CSC-215 Computer Organization (Advanced) at TCC Division of Information Technology and Business**
4. **2017 Spring: CS-395 Big Data: Data Wrangling at ODU College of Continuing Education and Professional Development**
5. **2017 Spring: CS-465/565 Information Assurance at ODU Department of Computer Science**
6. **2016 Fall: CSC 201 - Computer Science I at TCC Division of**

`http://www.cs.odu.edu/~ccartled/Teaching/`

## Q & A time.

Q: What is the square root of  $4b^2$ ?

A: To be or not to be.





# What have we covered?

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



Now!! On to exploring the world of Big Data!

## References (1 of 1)

- [1] Vangie Beal, Big Data,  
[https://www.webopedia.com/TERM/B/big\\_data.html](https://www.webopedia.com/TERM/B/big_data.html),  
2017.

# Who am I?

- Father
- Husband (only 41 years, but it seems longer)
- PhD, Computer Science, 2014
- CAPT, USN retired 2004 (31+ years)
- Professional software developer (38 years)
- A perennial student
- 1<sup>st</sup> 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

