

COMP 333 — Week 4 Welcome

Welcome to COMP 333 Data Analytics

In Week 3 and Week 4 (this lecture) we will cover Descriptive Data Analysis which is a basic technique of Data Analytics.

It is very important to understand your data.

This lecture looks at aspect of data that you will meet during data analysis even though they may not seem central to the course they are key things to keep in mind.

In Week 4 (this lecture) we will cover Data, namely

- ▶ Numerical Data and issues of accuracy, precision, significant digits, and normalization
- ▶ Structured and Unstructured Data, leading to the Big Data tsunami
- ▶ Metadata, Self-Descriptive Data, Data Formats relevant to data analysis
- ▶ `pandas` DataFrames
- ▶ Tidy Data, the basic format for easy manipulation of dataframes in `pandas`

READ the files marked READ.

For these topics, it is very worthwhile to also read/watch the supplementary material.

The recap will re-iterate the points covered in the lecture.

All the best, Greg.