

COMP 333 — Week 3 Welcome

Welcome to COMP 333 Data Analytics

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

It is very important to understand your data.

We saw many aspects of DDA in the examples:

- ▶ Example 2 is an official dataset from OECD on educational performance of 15-year olds.
- ▶ Example 3 is an investigation of survival on the Titanic when it sank.

In Week 3 (this lecture) we will cover

- ▶ Using the Python `scipy` library for *descriptive statistics* to describe the data.
- ▶ Re-iterate the types of data: *categorical* vs *continuous*, and *nominal*, *ordinal*, *interval*, and *ratio*, as these types allow different methods of description using statistics and plots.
- ▶ An introduction to plotting with the Python libraries `matplotlib` and `seaborn` to create bar charts, histograms, boxplots, and scatter plots.
- ▶ How to put these together as Descriptive Data Analysis.

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 lectures.

All the best, Greg.