

## COMP 333 — Week 8 Exploratory Data Analysis

### Exploratory Data Analysis

In Week 8 (this lecture) and Week 9 we cover Exploratory Data Analysis(EDA).

EDA involves Data Wrangling, including Data Cleaning

EDA involves repeated Descriptive Data Analysis (DDA)

EDA is about *exploring* the data,  
the relationships between variables,  
to gain *insight*  
so you can answer your questions  
and achieve your goals.

EDA builds *models* to capture the insight  
to predict outcomes in new situations  
as aids to decision-making.

You should revisit Example 3 of Week 2 on the Titanic  
as it is a very good introduction to EDA.

*A Gentle Introduction to Exploratory Data Analysis* by Daniel Bourke

<https://towardsdatascience.com/a-gentle-introduction-to-exploratory-data-analysis-f11d8>

This week we cover

- ▶ Exploratory Data Analysis
- ▶ Principal Component Analysis  
which is a common technique for *Dimension Reduction*  
for situations where the data has many variables  
(that is, high-dimensional data).
- ▶ Clustering  
which groups your observations based on the data.  
This is useful for *insight*  
but is also a technique for *feature engineering*.

Next week we cover

- ▶ Feature Engineering
- ▶ Models
  - ▶ Regression
  - ▶ Classification
  - ▶ Prediction
  - ▶ Simulation

READ the files marked READ.

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

Professor Skiena's Lecture 20: Clustering is informative

<http://www3.cs.stonybrook.edu/~skiena/519/>

All the best, Greg.