

# *COM5507 Social Media Data Acquisition and Processing*

## Week 10. Data exploration episode 1: Processing numerical data (outlines only)

Lectured by: Dr. Xinzhi ZHANG

Research Assistant Professor, Department of Journalism

Hong Kong Baptist University

7 Nov 2018 @ CityU M4003

# Agenda

- Exploring numerical data
  - Examining the data features
  - Descriptive statistics
  - Trend
  - Correlations
  - Outliers

# Data exploration

- Data exploration starts with inspecting the dimensionality, structure and data objects, followed by basic descriptive statistics and various charts like pie charts and histograms.
- Exploration of multiple variables are then demonstrated, including grouped distribution, grouped boxplots, scattered plot and pairs plot.
- Multiple-dimensions analysis is also introduced (as a challenge).
- - Zhao Yanchang, 2012

# In-class demonstration

- Exploring a data set from Gapminder with Python Pandas