UNIVERSITY OF LONDON

Reading List - CM3005 Data Science

Topics 1–5

Berberian, S.K. *Linear algebra*. (Mineola, NY: Dover Publications, 2014) Chapter 4 Matrices and Chapter 6 Determinants.

Blaikie, N. *Analyzing quantitative data: from description to explanation*. (London: Sage, 2003). 'What is data analysis?' section, pp.46–51.

EMC Education Services and EMC Education Services *Data science and big data analytics: discovering, analyzing, visualizing and presenting data*. (Indianapolis, IN: John Wiley & Sons, Inc., 2015) Chapter 1 Introduction to Big Data Analytics and Chapter 2 Data Analytics Lifecycle.

Grus, J. *Data science from scratch: first principles with Python*. (Sebastopol, CA: O'Reilly Media Inc., 2019) 2nd edition, Chapter 11 Machine Learning.

Kirk, A. *Data visualization: A successful design process*. (Birmingham, UK: Packt Publishing, 2012) Chapter 1, pp.22–47.

Myatt, G.J. and W.P. Johnson. *Making sense of data I: a practical guide to exploratory data analysis and data mining*. (Hoboken, NJ: John Wiley & Sons, Inc., 2014) 2nd edition.

- pp.22–24: 2.4 Central Tendency
- pp.24-36
- Chapter 4
    - 4.2.1 Scatterplots
    - 4.2.2 Summary Tables and Charts
    - 4.2.3 Cross-classification Tables
    - 4.3 Calculating Metrics about Relationships

Myatt, G.J. and W.P. Johnson *Making sense of data II: a practical guide to data visualization, advanced data mining methods, and applications*. (Hoboken, NJ: John Wiley & Sons, Inc., 2009).

- pp.49–59
- pp.32–49
    - 2.3 Tables
    - 2.4 Univariate data visualisation

Provost, F. and T. Fawcett. *Data science for business: what you need to know about data mining and data-analytic thinking*. (Sebastopol, CA: O'Reilly Media, Inc., 2013).

- Chapter 1
- pp.111–12
- Chapter 5, pp.126–29

VanderPlas, J. *Python data science handbook: essential tools for working with data*. (Sebastopol, CA: O'Reilly Media, Inc., 2016).

- Chapter 2 Introduction to NumPy
- pp.359–63
- Chapter 3 Data Manipulation with Pandas
- Chapter 5 Machine Learning, pp.331–54
- pp.363–75
- pp.375–82

Ware, C. *Information visualization: perception for design*. (Waltham, MA: Morgan Kaufmann, 2012), 3rd edition, pp.1–20. (3.202).

Yau, N. *Data points: visualization that means something*. (Indianapolis, IN: John Wiley & Sons, Inc., 2013) Chapter 4 Exploring Data Visually, pp.143–53 and pp.189–99.

- 4.2 Visualizing Categorical Data
- 4.6 Distributions

Web resources

Topics 1–5

Python tutorial. Python Software Foundation. https://docs.python.org/3/tutorial/

JupyterLab Getting started: Overview. https://jupyterlab.readthedocs.io/en/stable/getting_started/overview.html

Jupyter user guide:

- JupyterLab interface
- Text editor
- Notebooks
- Code consoles.

NumPy:

- Data Types
- Array Creation
- Array Objects: Indexing
- Routines: Statistics
- Linear algebra (numpy.linalg)

Pandas User Guide:

- Introduction
- Overview
- Timestamps vs. time spans
- Converting to timestamps
- Generating ranges of timestamps
- Timestamp limitations
- Indexing is here but cover this later

- Time/date components
- DateOffset objects
- Time Series-Related Instance Methods
- Resampling

Time series / date functionality. Pandas development team

Click the links below to read a brief overview of the fundamental data structures in pandas:

- pandas getting started: Intro to data structures. pandas development team.
- pandas API reference: Series. pandas development team.
- pandas API reference: DataFrame. pandas development team.