Reading List - CM3005 Data Science

Berberian, Sterling 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? (pp. 46–51)

EMC Education Services and EMC Education Services. Data Science and Big Data Analytics : Discovering, Analyzing, Visualizing and Presenting Data : Discovering, Analyzing, Visualizing and Presenting Data (Chapter 1 and 2)

Grus, Joel. *Data science from scratch: first principles with Python*. O'Reilly Media, 2019. Chapter 11, Machine learning

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

G. J. Myatt and W. P. Johnson (2014). *Making sense of data I: a practical guide to exploratory data analysis and data mining*. 2nd ed. Hoboken, NJ, US: John Wiley & Sons, Incorporated. ISBN: 9781118422014.

- pp.22–24: 2.4 Central tendency
- pp.24-36
- Chapter 4.
  - 4.2.1 Scatterplots
  - 4.2.2 Summary charts and tables
  - 4.2.3 Cross-classification tables
  - 4.3 Calculating metrics about relationships

G. J. Myatt and W. P. Johnson (2009). *Making sense of data II: a practical guide to data visualization, advanced data mining methods, and applications*. Hoboken, NJ, US: John Wiley & Sons, Incorporated. ISBN: 9780470417393,

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

Provost, Foster and Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. (O'Reilly Media, Inc., 2013)

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

VanderPlas, Jake *Python data science handbook: Essential tools for working with data*. (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. (2012). *Information visualization: perception for design*. 3rd edition. Waltham, MA US: Morgan Kaufmann, pp. 1–20. (3.202)

N. Yau (2013). *Data points: visualization that means something*. Indianapolis, IN, US: John Wiley & Sons, Incorporated. ISBN: 9781118462195, pp.143–153 and 189–199.

- 4.2 Categorical data
- 4.6 Distributions

Web resources

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.