

# Textual Data

⚠ Changes before 22/23

- ☐ Add Learning Outcomes for each week
- ☐ Link to regex/working with strings in Week 2 CASA0005

## Overview

! Important

This week's Learning Outcomes are:

- 1.
- 2.
- 3.

## Preparation

### Lectures

Come to class prepared to present:

Session	Video	Presentation	Notes
Notebooks as Documents	Video	Slides	Notes
Patterns in Text	Video	Slides	Notes
Cleaning Text	Video	Slides	Notes
Analysing Text	Video	Slides	Notes

## Readings

Come to class prepared to present:

- (Ladd 2020) URL
- (Lavin 2019) URL

## Activities

- Padlet: [Collaborative Agenda]
- Complete the short Moodle quiz associated with this week's activities.

## Practical

### Changes before 22/23

- ☐ Add Learning Outcomes for each week
- ☐ Standardise delivery by TAs where practicals are TA-led (e.g. FSDS): clearer guidance on how to take students through each week.
- ☐ Clearer articulation of difficulty levels/targets for components *within* each week's practicals.
- ☐ Make links between lecture and practical content explicit; ideally trace a question through the whole process (*e.g.* "If I wanted to know if and where blue plaques are clustered in London how would I find out?")

The practical can be downloaded from GitHub.

## References

- Ladd, John R. 2020. "Understanding and Using Common Similarity Measures for Text Analysis." *The Programming Historian*, no. 9. <https://doi.org/10.46430/phen0089>.
- Lavin, Matthew J. 2019. "Analyzing Documents with TF-IDF." *The Programming Historian*, no. 8. <https://doi.org/10.46430/phen0082>.