
A TEMPLATE FOR THE *ARXIV* STYLE

A PREPRINT

H.Sherry Zhang

Department of Econometrics and Business Statistics
Monash University
Melbourne, Australia
`huize.zhang@monash.edu`

Dianne Cook

Department of Econometrics and Business Statistics
Monash University
Melbourne, Australia
`dicook@monash.edu`

Ursula Laa

Institute of Statistics
University of Natural Resources and Life Sciences
Vienna, Austria
`ursula.laa@boku.ac.at`

Nicolas Langrené

34 Village Street, Docklands VIC 3008 Australia
CSIRO Data61
Melbourne, Australia
`nicolas.langrene@csiro.au`

Patricia Menéndez

Department of Econometrics and Business Statistics
Monash University
Melbourne, Australia
`patricia.menendez@monash.edu`

August 4, 2021

Abstract

Enter the text of your abstract here.

Keywords blah · blee · bloo · these are optional and can be removed

1 Introduction

Spatio-temporal data

Motivation: Many focus on spatio/ temporal solely -> spatio-temporal vector data structure for data analysis sections

2 Literature Review

Existing R pkgs for spatio/ temporal data

- `sp` + `spacetime` (time series data from package `xts`)
- `sf`

- simple features geometry
- only spatio
- good for geographical but these geometry types i.e. multipolygon, can be hard to manipulate. Would be nice to have a vector data structure
- **stars**
 - a newer structure allow both spatio + temporal
 - array

3 A new data structure for spatio-temporal data

Intro to cubble:

- list-column: rowwise_df with temporal variables, including the time index, nested.
 - Focus on spatio: those output per station
- long form: grouped_df
 - Focus on temporal

Compatible with tidyverse manipulation and tsibble

4 Examples

Daily climate data (prcp, tmax, and tmin) from RNOAA - lots of stations across Australia

An exploratory data analysis questions: What's the climate profile look like in Australia

- General features: Any general trend/ fluctuation in prcp, tmax, and tmin?
- Local features: Any station stands out from the crowd?

4.1 Manipulation

- data quality check: filter out stations have variables not properly recorded
- data summary:
 - daily -> monthly/ weekly,
 - summarise by mean for tmax/ tmin, sum for prcp
-

4.2 Graphics

Static + interactive -> tooltip to show additional information upon hovering

- Where are those stations on the map?
 - Mention mostly aero, airport, and lighthouse

5 Summary