Amendments summary

Yiru (Earo) Wang   22 November 2019

I thank my examiners for their thorough and constructive comments. The point by point description of changes are below: the examiners' comments are in red and my response is in black.

Professor Antony R. Unwin

I appreciate all the comments and suggestions, and will incorporate them into future work. Specific comments are addressed as follows:

p12 'Then the final coordinates for making the calendar plots of the pedestrian sensor data is' are

Done.

p36 'The "tidy data' principles are extended as tidy temporal data principles in tsibble as follows:" tsibble should be in bold?

'tsibble' here refers to the new data abstraction in normal text throughout the thesis, while 'tsibble' in bold refers to the package **tsibble**.

p36 'It contains 12 observations and 5 variables landing in a tidy data form.' landing?

No changes made. Supervisor agreed on this.

p38 '(such as in the Stata)' in the Stata package?

'Stata' is used throughout Chapter 3. It's statistical software as R, not an R package.

p39 'When constructing a tsibble, any duplicates of key-index (i.e. key-value) pairs will fail' Values are measurements and not the same as the index. Duplicates do not fail.

I removed '(i.e. key-value)'.

p39 'This neatly decouples the expensive data copy from downstream summaries, significantly reducing the storage space.' An example would be useful here, the sentence is unclear.

I rewrote 'This neatly decouples the expensive data copying from downstream summarization, which can significantly reduce the required storage space.'

p40 'Each type of observational units should form a table' unit?

Done.

p40 (and References) McIlroy not Mcilroy

Done.

p41 tidyverts needs a reference of some kind

Done.

p41 'It takes advantage of the wrangling verbs implemented in the dplyr package, and develops a suite of new tools for facilitating temporal manipulation for primarily easing two aspects: implicit missingness handlers and time-aware aggregations.' Confusing.

No changes made. Supervisor agreed on this.

The **tsibble** package leverages the dplyr verbs and contributes new verbs for tackling some common temporal problems with ease.

p41 'Implicit missingness are values' needs to be rephrased

I've rephrased 'missingness' to 'missings'.

p41 'A family of verbs are provided' is

Done.

p41 'It will be a problem for temporal models and operations like lag/lead are applied.' needs rewriting

I rewrote 'It will be a problem for temporal models and operations like lag/lead, that expect consecutive time.'

p42 'The former will reduce to a single summary, whereas the latter will obtain the index and their corresponding summaries.' needs rephrasing

I rewrote 'The former gives a single summary for a data table, and the latter provides the corresponding summary for each index value.'

p42 'the interested measurement' the measurement of interest

Done.

p42 ', hence can' should be 'can hence'

Done.

p47 'A complete and consistent set of tools are available' is available

Done.

p48 'the table verbs extensively use tidy evaluation to evaluate computation in the context of tsibble data and spotlights the "tidy' interface" spotlight

Done.

p50 'interested questions' questions of interest

Done.

p51 The language in the caption to Figure 3.5 could be improved

I rewrote 'Line plots showing departure delay against time of day, faceted by day of week and 50%, 80% and 95% quantiles. The blue horizontal line indicates the 15-minute on-time standard to help grasp the delay severity. Passengers are more likely to experience delays around 18 during a day, and are recommended to travel early. The variations increase substantially as the upper tails.'

p51 'the operations undertaking' undertaken

Done.

p53 'with higher delay percentage in both ends of the year.' at both ends

Done.

p53 'Half-hourly general supply in kwH have been recorded' has been

Done.

p70 'The right panel is the swap between them, showing that the overlapping missings of the Runs type with the Occasions occur to the longer runs.' ?

I've rephrased to 'The right panel interchanges two variables, and indicates that there is little overlap on the long runs.'

p71 'which is referred to "missing data polish"' referred to as?

Done.

p72 'models would likely complain the existence of any missings and some would even decline the job' Rewrite!

models would likely complain *about* the existence of any missings and some would even decline the job.

p74 'can spark overwhelmingness at the first try' Rephrase

No changes made. Supervisor agreed on this.

p76 'is used too illustrate' to

Done.

the bigram 'and etc.', which is used a few times, should just be 'etc.'

Done.

The classical text of Chambers et al from 1983 is dated by its reprinting in 2017, which makes it look like a modern text. Shouldn't the original date be used, or at least mentioned?

Done.

Tierney & Cook (2018a) and (2018b) are the same reference. Presumably one of them should refer to the naniar package.

Done.

Professor Jeffrey T. Leek

I appreciate all the comments and suggestions, and will incorporate them into future work.