

Exploring probability distributions for bivariate temporal granularities

Sayani Gupta - Department of Econometrics and Business Statistics, Monash University

Smart meters measuring energy usage at finer temporal scales, are installed on many households in many countries now. Providing tools to explore this type of data is an important activity. We propose to provide methods to operate on time in an automated way, to deconstruct it in many different ways. Deconstructions of time that respect the linear progression of time like days, weeks and months are defined as linear time granularities and those that accommodate for periodicities in time like hour of the day or day of the month are defined as circular granularities or calendar categorizations. Often visualizing data across these circular granularities are a way to go when we want to explore periodicities, pattern or anomalies in the data. Also, because of the large volume of data in recent days, using probability distributions for display is a potentially useful approach. This work provides these techniques into the tidy workflow, so that probability distributions can be examined in the range of graphics available in the ggplot2 package. In particular, this work establishes the following:

- Provides functions to create multiple-order-up time granularities. For example, week of the month, hour of the week or quarter hour of the day. This is an extension to the lubridate package, which also allows creating calendar categorizations, usually one-order-up.
- Provides checks on the feasibility of plotting or drawing inference from two granularities together. It categorizes pairs of granularities as either a harmony or clash, where harmonies are pairs of circular granularities that aid exploratory data analysis. Clashes are pairs that are incompatible with each other for exploratory analysis. Appropriate data structures are provided to visualize with the grammar of graphics for harmonies.