Raw Data

More information on this data can be found here:
https://github.com/Derek-Jones/renzo-pomodoro/tree/master

The data are a product of Renzo Borgatti tracking planned work tasks for each day for ten years, along with his estimate of how long they would take, and how long they actually took.

Estimates were in units of 'pomodoro' sessions - a 25 minute block of work.

Column Name	Explanation
X.words	One-word tags used to describe the type of task
word_cnt	An anonymised version of the task description - each word in the description is simply replaced with a '1'.
description	A number given to each unique description.
DONE	1 if the task was successfully completed on the day it was planned, 0 if not.
date	The date the task was planned
estimate	The estimated number of pomodoro sessions required to complete the task
actual	The actual number of pomodoro sessions required to complete the task.

Processed Data

In the course of analysing the data, many more columns were added to this dataset. These additional columns are explained below.

difference	The difference between the estimated and actual number of pomodoro sessions required. Calculated data$estimate - data$actual
Year	The year the task took place in. Character class. Format YYYY
yearMonth	The year and month the task took place in. Character class. Format YYYY-MM
original_order	The original chronological order of the tasks in the dataset.

previous_date	The previous date on which any task was planned. Date class.
next_date	The next date on which any task was planned. Date class.
done_previous_day	1 if the same task was planned and completed on the previous working day, 0 if not.
not_done_previous_day	1 if the same task was planned but not completed on the previous working day, 0 if not.
task_next_day	1 if the same task was planned the next day (regardless of whether it was completed or not)
rolling_estimation_mean	The rolling mean of the difference column. n = 50, align = "center"
tags	The tags in the original X.words column, now split so there is a maximum of one per row.
tag_craft … tag_other	There are 9 columns, comprising the 9 most common descriptive tags. These columns simply indicate whether a task was given one of these descriptions, with a 1 in the corresponding column if so, and a 0 if not.
Status	There are five options here, depending on certain characteristics of the task: Done_Only: a task which was planned and marked as done on the same day
	Done_Eventually: a task which was not completed on the day originally planned, but was marked as done a day or more later.
	Done_Maybe_Only: a task which was not marked as done, but was given a corresponding 'actual' number on the same day as it was planned. Thus the data are conflicting about whether the task was completed or not.
	Done_Maybe_Eventually: a task which was given an 'actual' value a day or more later than when it was planned, but not marked as Done.
	Done_Never: a task which was planned but

	not completed on that day, and either not planned or not completed on the subsequent working day.
proportional_error	Since a difference of 1 between estimate and actual columns is much more when the estimate was 1 than 9, for example, this column calculates the degree of error proportional to the estimate. Calculated as the difference divided by estimate.
running_total_tasks	The running total of times that unique task description has occurred before.
running_total_tags	The running total of times that tag has occurred before.
date_minus_x	The date the task was planned minus the value of x (x varies by column). These columns were used to calculate the next set of columns.
task_frequency_xweeks	The number of occurrences of that unique task description over the last x weeks (x varies by column)
maybe_status	FALSE if the task_status is Done_Only, Done_Eventually or Done_Never. TRUE if Done_Maybe or Done_Maybe_Eventually
rolling_error	The rolling mean of the absolute value of the proportional error, across 500 rows. (align = 'center')
rolling_estimate	The rolling mean of the task length estimate across 500 rows. (align = 'center')