MONASH University

MONASH BUSINESS SCHOOL

Analysing sub-daily time series data

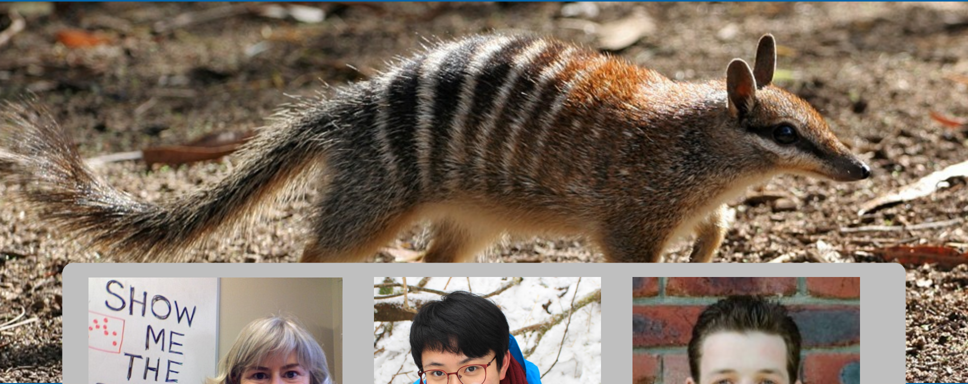
Rob J Hyndman
Earo Wang
Mitchell O'Hara-Wild

NUMBATS

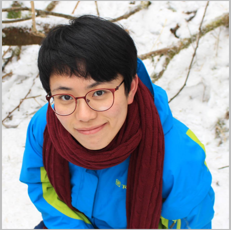
Non-Uniform Monash Business Analytics Team

NUMBATS

Non-Uniform Monash Business Analytics Team



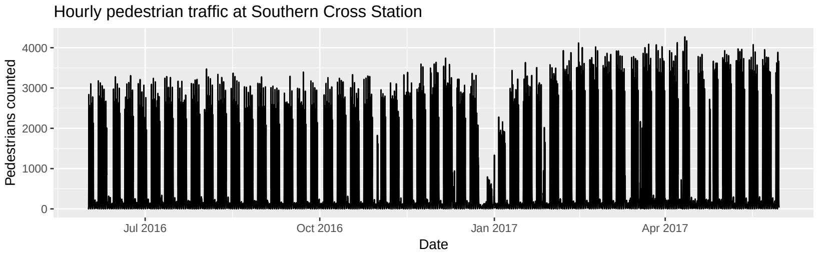
Di Cook

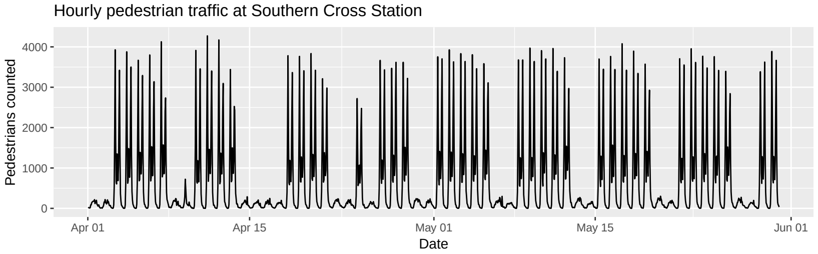


Earo Wang

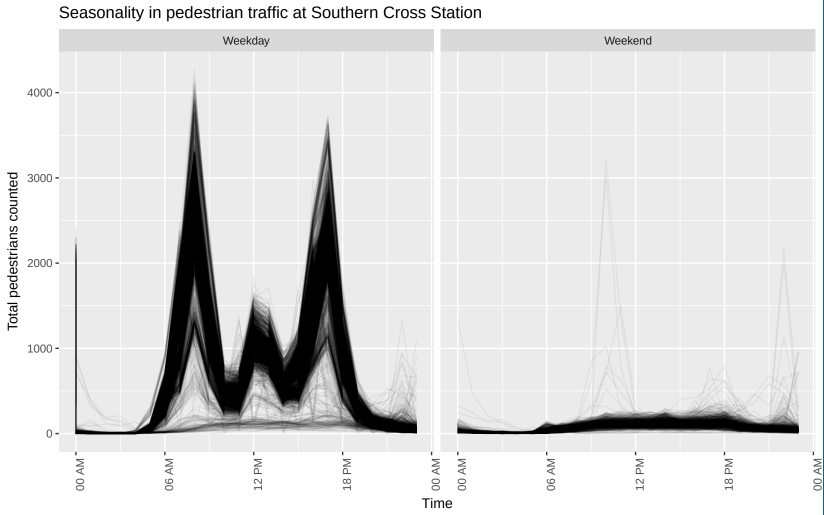
Mitchell O'Hara-Wild

Pedestrian counts

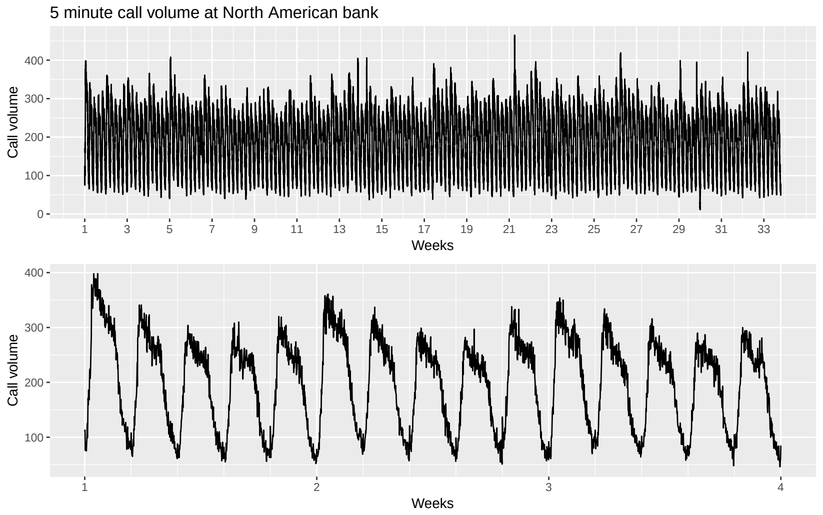




Pedestrian counts

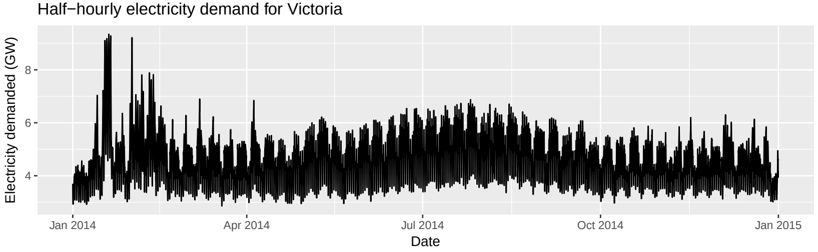


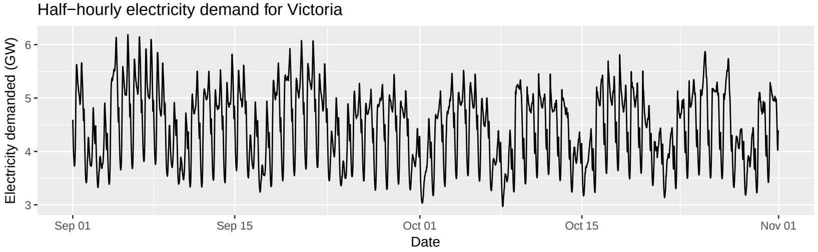
Call volume



5 minute call volume at North American bank
Call volume
400
300
200
100
0
1
3
5
7
9
11
13
15
17
19
21
23
25
27
29
31
33
Weeks
Call volume
400
300
200
100
1
2
3
4
Weeks

Electricity demand





Challenges

Visualization

- Even plotting a single time series comprising one year of data, it is hard to see the interesting features.

Challenges

Visualization

- Even plotting a single time series comprising one year of data, it is hard to see the interesting features.

R classes

- The `ts`, `zoo`, `xts` and other time series classes do not work well with sub-daily data.
- Newer packages (`timetk` and `tibbletime`) do not play nicely with modelling functions.

Earo Wang

Challenges

Visualization

- Even plotting a single time series comprising one year of data, it is hard to see the interesting features.

R classes

- The `ts`, `zoo`, `xts` and other time series classes do not work well with sub-daily data.
- Newer packages (`timetk` and `tibbletime`) do not play nicely with modelling functions.

Challenges

Visualization

- Even plotting a single time series comprising one year of data, it is hard to see the interesting features.

R classes

- The `ts`, `zoo`, `xts` and other time series classes do not work well with sub-daily data.
- Newer packages (`timetk` and `tibbletime`) do not play nicely with modelling functions.

Forecasting

- Most time series modelling frameworks handle sub-daily data poorly.
- Available models include `tbats` and `prophet`, but they have limitations.

TBATS model

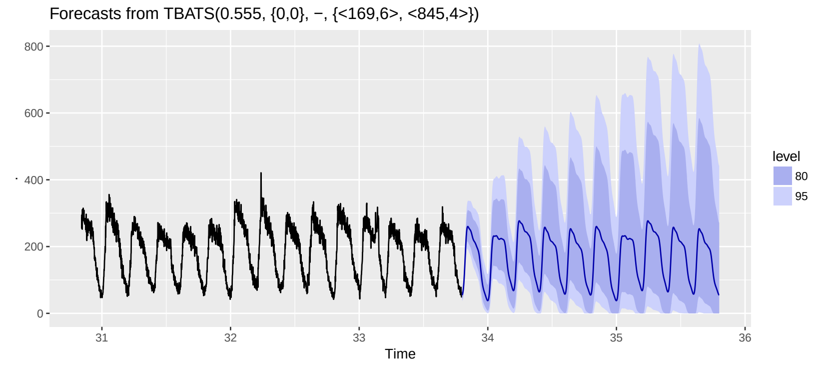
TBATS

Trigonometric terms for seasonality
Box-Cox transformations for heterogeneity
ARMA errors for short-term dynamics
Trend (possibly damped)
Seasonal (including multiple and non-integer periods)

- Handles non-integer seasonality, multiple seasonal periods.
- Entirely automated
- Prediction intervals often too wide
- Very slow on long series
- No exogenous predictors

TBATS model

```r
library(forecast)
calls %>% tbats %>% forecast %>%
  autoplot(include=2500)
```



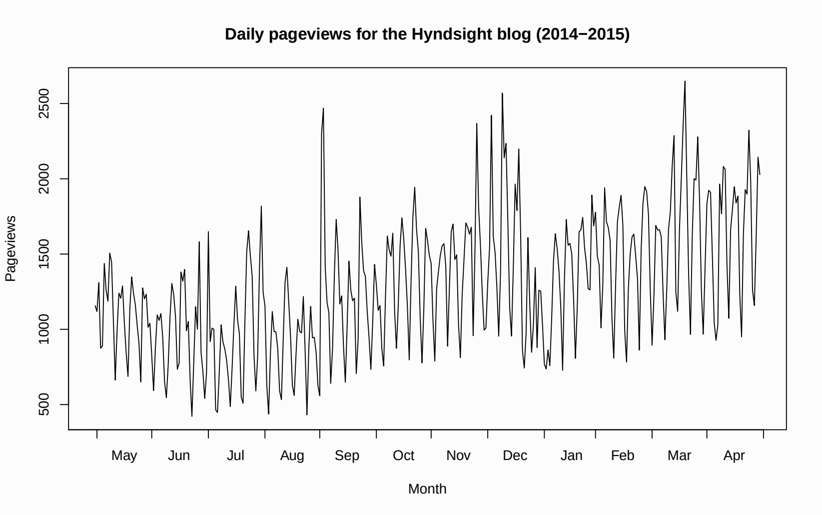
prophet

Additive regression model developed at Facebook

$$y_t = g_t + s_t + h_t + \varepsilon_t$$

- y_t = time series.
- g_t = piecewise linear growth function
- s_t = Fourier seasonal terms: daily, weekly and/or yearly
- h_t = holiday effect.
- ε_t = error (can be ARMA errors).
- Estimated as a Bayesian regression using Stan

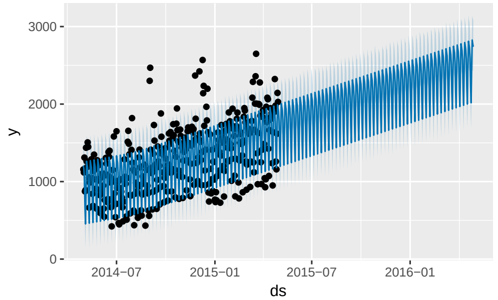
Daily blog traffic



prophet example

```r
library(prophet)
m <- prophet(hyndsight)
future <- make_future_dataframe(m, periods = 365)
forecast <- predict(m, future)
```

```r
plot(m, forecast)
```



prophet pros and cons

Pros

- Completely automatic including changepoints
- Handles multiple seasonality and holiday effects

Cons

- Seems to overfit annual seasonality
- Number of Fourier terms is hard-coded

Mitchell O'Hara-Wild

Watch this space

https://github.com/earowang/tsibble

http://pkg.earo.me/sugrrants

https://github.com/mitchelloharawild/fasster

http://pkg.robjhyndman.com/forecast

http://pkg.earo.me/hts

Slides available at robjhyndman.com