

Data visualization for functional time series

Rob J Hyndman

11 December 2018

Outline

- 1 Using ggplot2 for functional time series
- 2 Time-indexed probability distributions

Outline

- 1 Using ggplot2 for functional time series
- 2 Time-indexed probability distributions

Outline

- 1 Using ggplot2 for functional time series
- 2 Time-indexed probability distributions

Smart metre data

vec.ausnetservices.com.au

Request your Smart Meter Data

NOTE: You will need your electricity bill handy to complete this form

Once you submit this form you will be sent an email with instructions on how to have your smart meter data redirected to the Victorian Energy Compare website where you can compare energy retail offers and claim your \$50 Power Saving Bonus

First Name

Enter the First Name provided to your energy retailer 

Last Name

Enter the Last Name provided to your energy retailer 

NMI (National Meter Identifier)



630XXXXXXXX

Meter Number



Phone Number

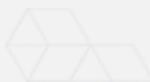


Postcode

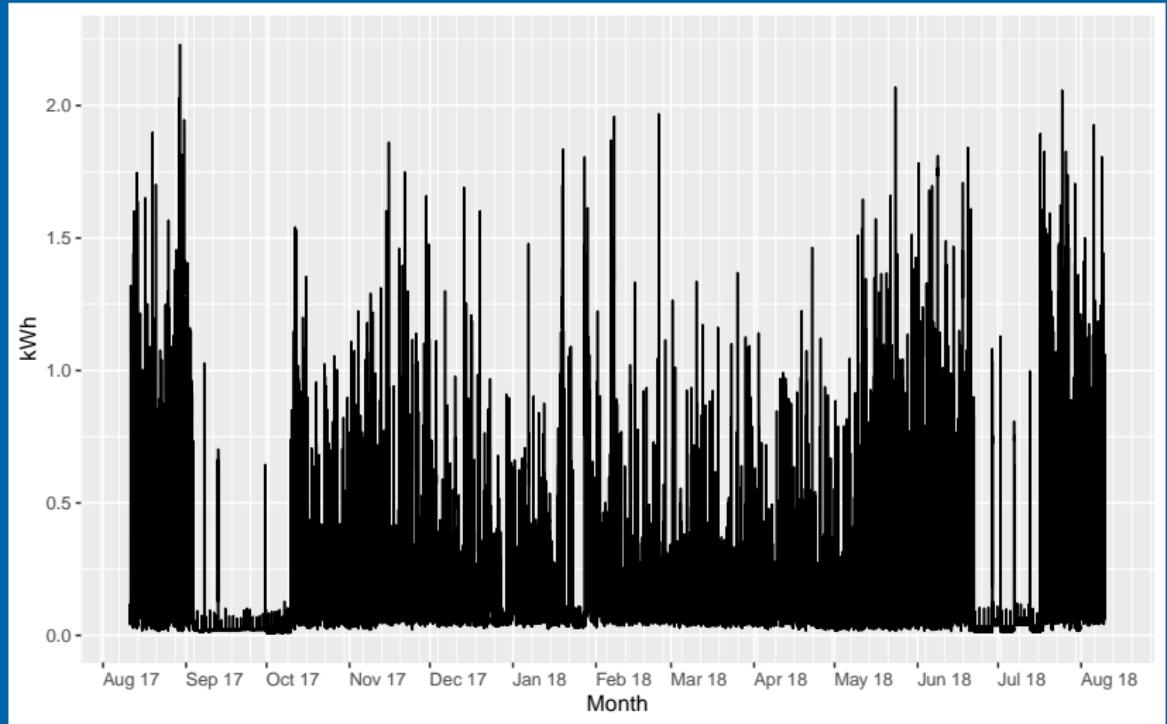
Email Address



Confirm Email



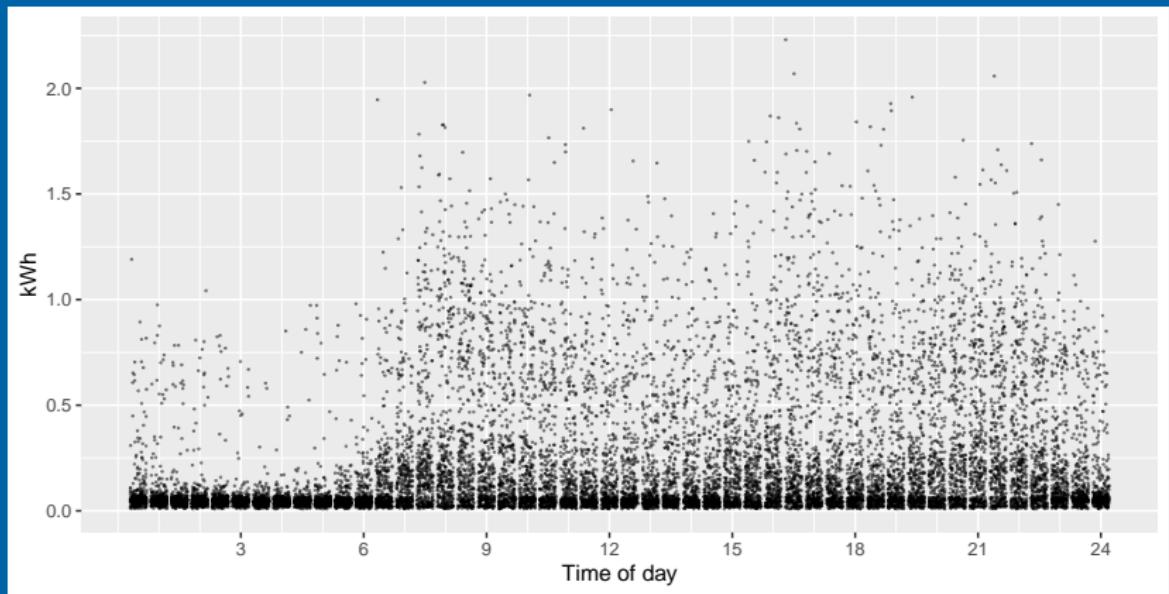
George's data



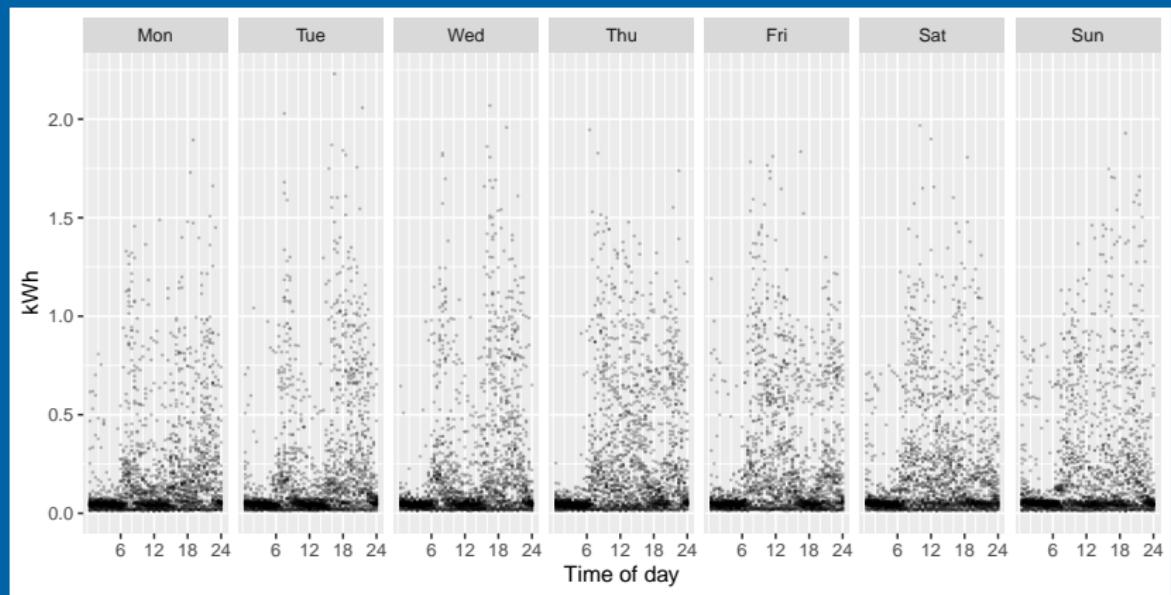
George's data



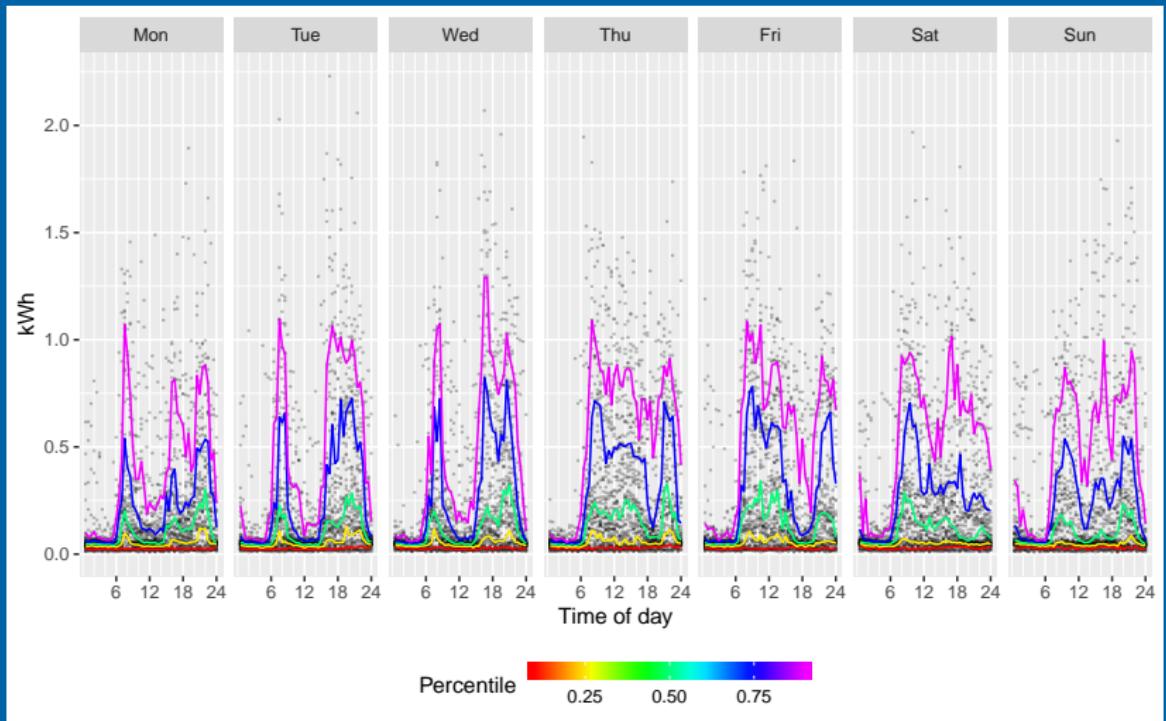
George's data



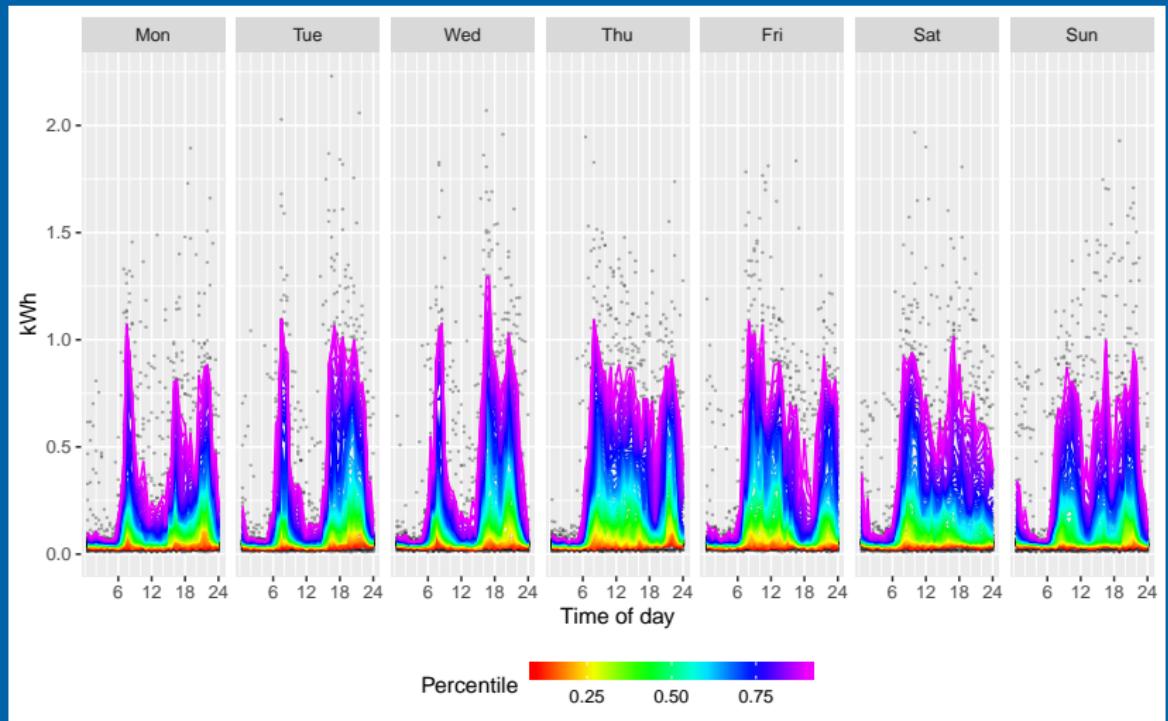
George's data



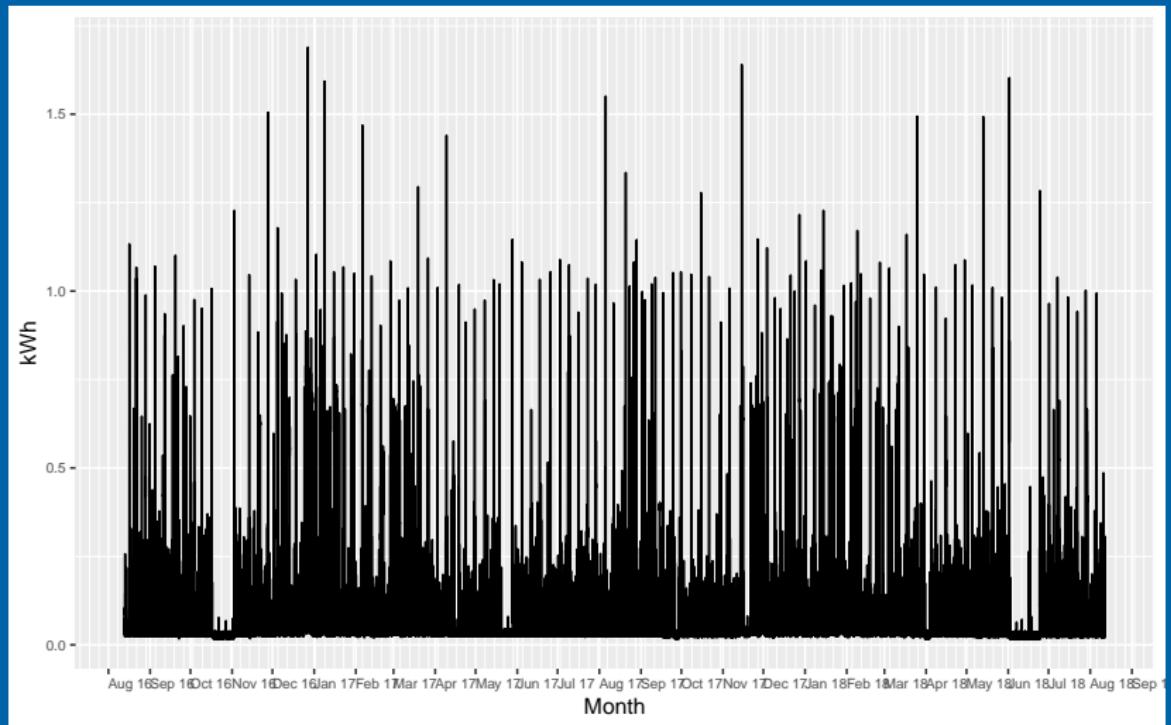
George's data



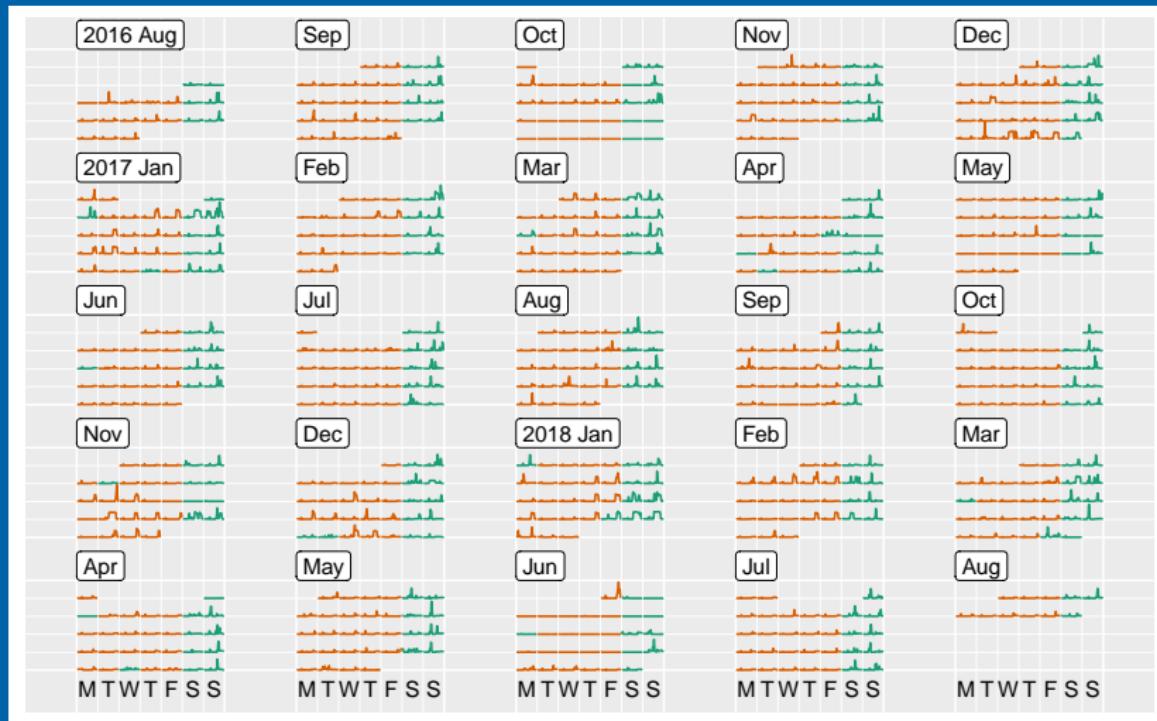
George's data



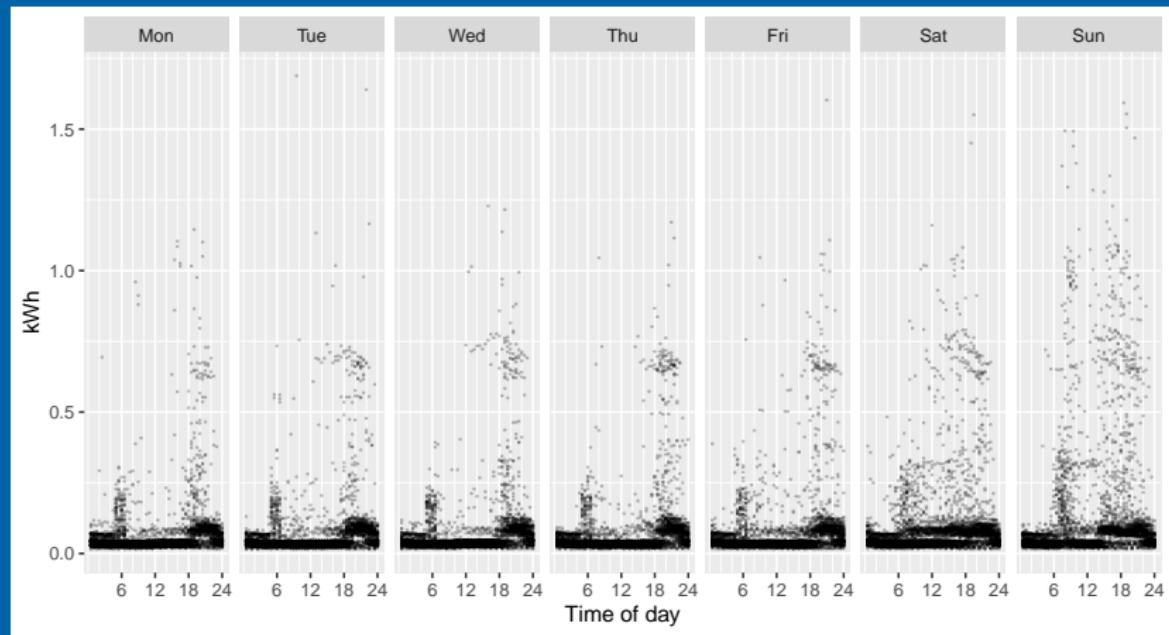
Clare's data



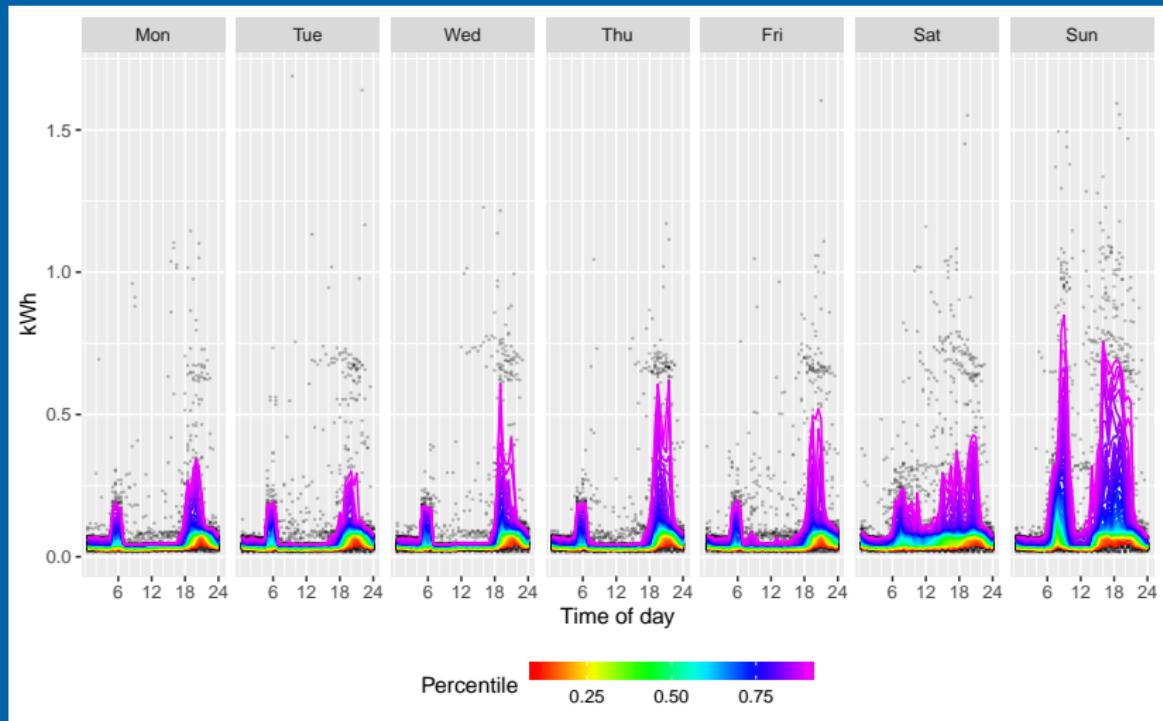
Clare's data



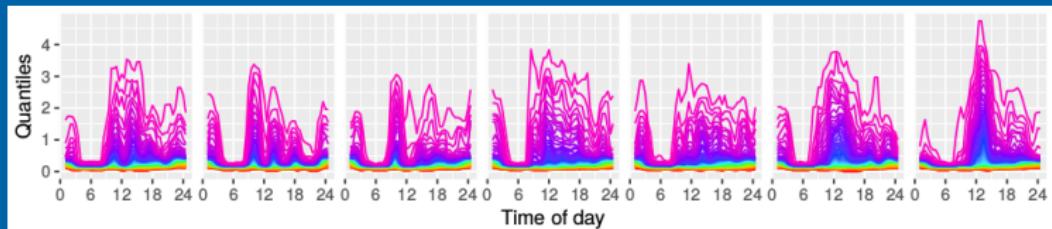
Clare's data



Clare's data



Percentiles conditional on time of week

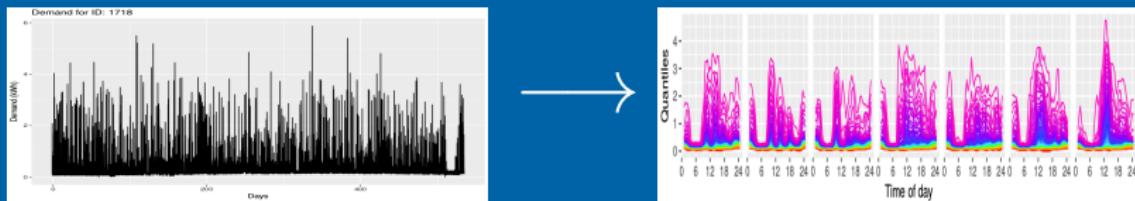


- Percentiles for each household and each half-hour of the week.
- Provides a unique fingerprint of typical usage for a given household.
- 336 probability distributions per household.
- Avoids missing data issues and variation in series length
- Avoids timing of household events, holidays, etc.
- Allows clustering of households based on probabilistic behaviour rather than coincident behaviour.
- A more complicated version also allows it to change across the year.

Finding anomalous smart metres

Irish smart metre data

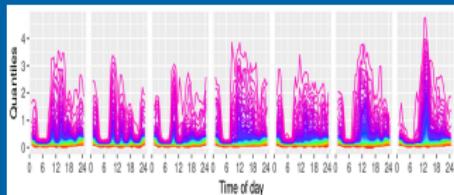
- 500 households from smart metering trial:
14 July 2009 – 31 December 2010.
- Electricity consumption at 30-minute intervals.
- Heating/cooling energy usage excluded.



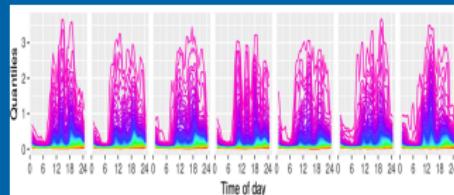
The time series of 535×48 observations per household is mapped to a set of $7 \times 48 \times 99$ percentiles giving a bivariate surface for each household.

Finding anomalous smart metres

Can we compute pairwise distances between all households?

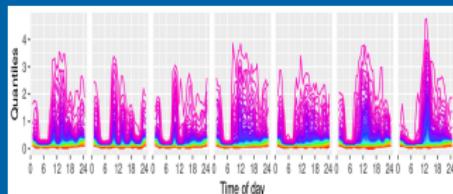


← ? →
Distance

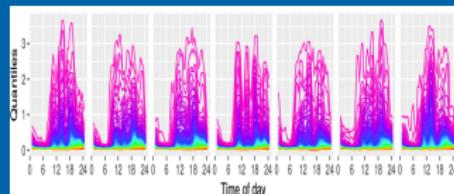


Finding anomalous smart metres

Can we compute pairwise distances between all households?



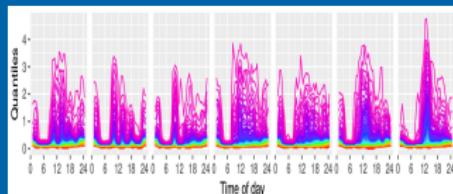
← ? →
Distance



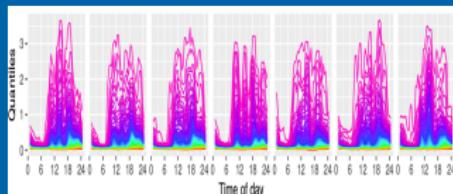
- Jensen-Shannon measure gives distance between two densities

Finding anomalous smart metres

Can we compute pairwise distances between all households?



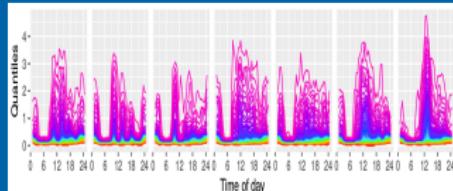
← ? →
Distance



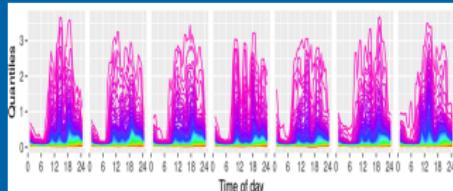
- Jensen-Shannon measure gives distance between two densities
- Distance between household i and household j :
 $\Delta_{ij} = \text{sum of } 7 \times 48 \text{ JS distances.}$

Finding anomalous smart metres

Can we compute pairwise distances between all households?



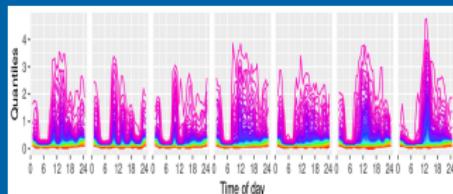
← ? →
Distance



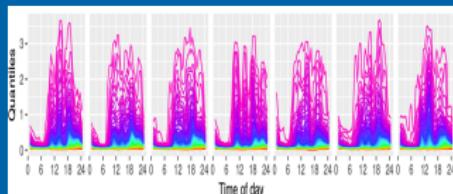
- Jensen-Shannon measure gives distance between two densities
- Distance between household i and household j :
 $\Delta_{ij} = \text{sum of } 7 \times 48 \text{ JS distances.}$
- Similarity between two households:
 $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$

Finding anomalous smart metres

Can we compute pairwise distances between all households?

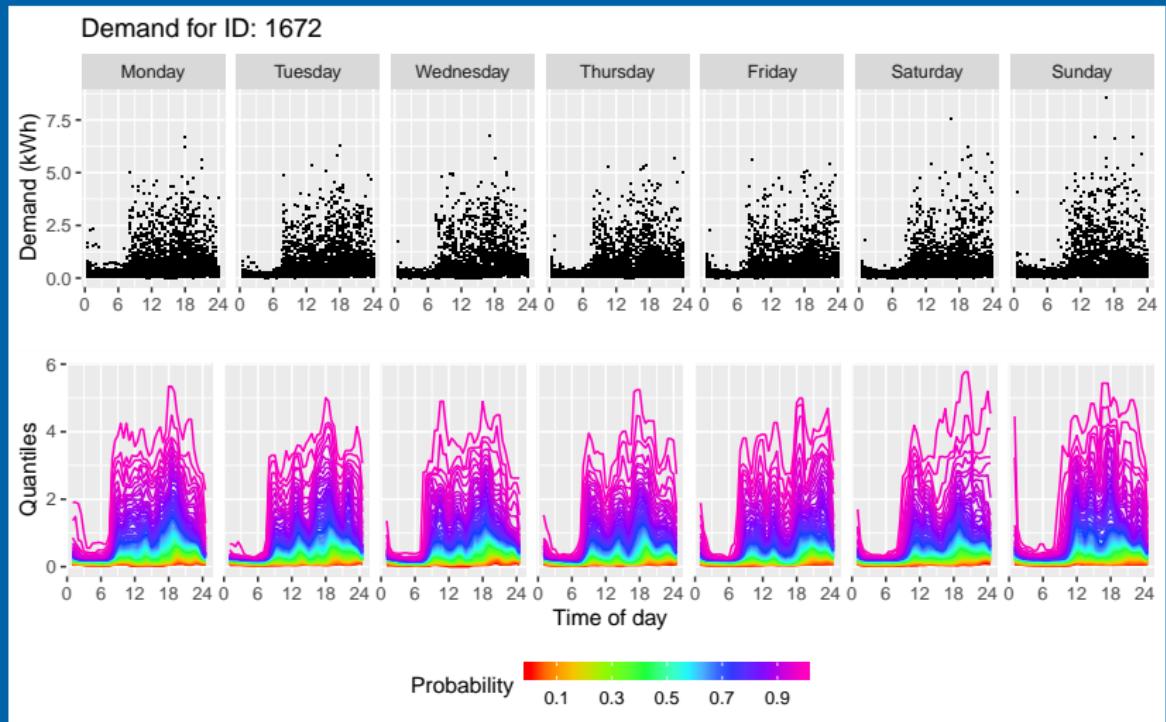


← ? →
Distance

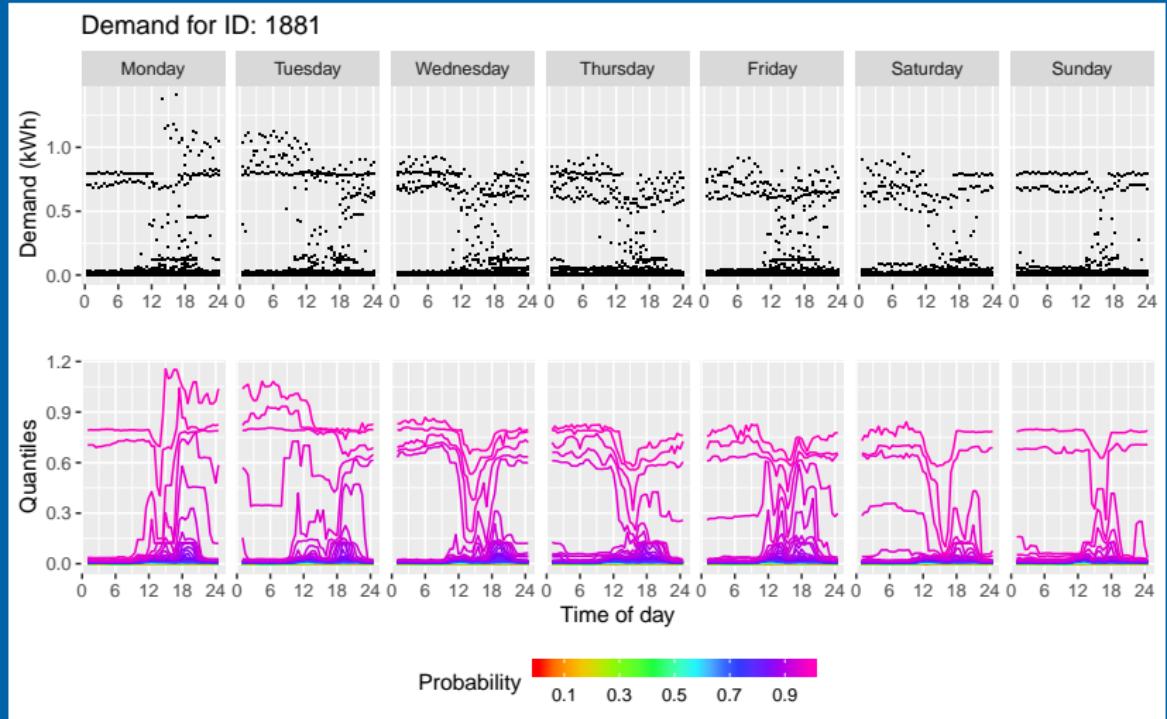


- Jensen-Shannon measure gives distance between two densities
- Distance between household i and household j :
 $\Delta_{ij} = \text{sum of } 7 \times 48 \text{ JS distances.}$
- Similarity between two households:
 $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$
- Household typicality: $f_i = \sum_j w_{ij}.$

Most typical household



Most anomalous household



Laplacian eigenmaps

- **Idea:** Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.

Laplacian eigenmaps

- **Idea:** Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.
- Let $\mathbf{W} = [w_{ij}]$ where $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$.
$$\mathbf{D} = \text{diag}(\hat{f}_i) \quad \text{where } \hat{f}_i = \sum_{j=1}^n w_{ij}$$
$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (\text{the Laplacian matrix}).$$

Laplacian eigenmaps

- **Idea:** Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.
- Let $\mathbf{W} = [w_{ij}]$ where $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$.
$$\mathbf{D} = \text{diag}(\hat{f}_i) \quad \text{where } \hat{f}_i = \sum_{j=1}^n w_{ij}$$
$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (\text{the Laplacian matrix}).$$
- Solve generalized eigenvector problem: $\mathbf{L}\mathbf{e} = \lambda\mathbf{D}\mathbf{e}$.

Laplacian eigenmaps

- **Idea:** Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.
- Let $\mathbf{W} = [w_{ij}]$ where $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$.
$$\mathbf{D} = \text{diag}(\hat{f}_i) \quad \text{where } \hat{f}_i = \sum_{j=1}^n w_{ij}$$
$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (\text{the Laplacian matrix}).$$
- Solve generalized eigenvector problem: $\mathbf{L}\mathbf{e} = \lambda\mathbf{D}\mathbf{e}$.
- Let \mathbf{e}_k be eigenvector corresponding to *kth smallest eigenvalue*.

Laplacian eigenmaps

- **Idea:** Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.
- Let $\mathbf{W} = [w_{ij}]$ where $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$.
$$\mathbf{D} = \text{diag}(\hat{f}_i) \quad \text{where } \hat{f}_i = \sum_{j=1}^n w_{ij}$$
$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (\text{the Laplacian matrix}).$$
- Solve generalized eigenvector problem: $\mathbf{L}\mathbf{e} = \lambda\mathbf{D}\mathbf{e}$.
- Let \mathbf{e}_k be eigenvector corresponding to *kth smallest eigenvalue*.
- Then \mathbf{e}_2 and \mathbf{e}_3 create an embedding of households in 2d space.

Key property of Laplacian embedding

Let $y_i = (e_{2,i}, e_{3,i})$ be the embedded point corresponding to household i .

Then the Laplacian eigenmap minimizes

$$\sum_{ij} w_{ij}(y_i - y_j)^2 = \mathbf{y}' \mathbf{L} \mathbf{y} \quad \text{such that} \quad \mathbf{y}' \mathbf{D} \mathbf{y} = 1.$$

Key property of Laplacian embedding

Let $y_i = (e_{2,i}, e_{3,i})$ be the embedded point corresponding to household i .

Then the Laplacian eigenmap minimizes

$$\sum_{ij} w_{ij}(y_i - y_j)^2 = \mathbf{y}' \mathbf{L} \mathbf{y} \quad \text{such that} \quad \mathbf{y}' \mathbf{D} \mathbf{y} = 1.$$

- the most similar points are as close as possible.

Key property of Laplacian embedding

Let $y_i = (e_{2,i}, e_{3,i})$ be the embedded point corresponding to household i .

Then the Laplacian eigenmap minimizes

$$\sum_{ij} w_{ij}(y_i - y_j)^2 = \mathbf{y}' \mathbf{L} \mathbf{y} \quad \text{such that} \quad \mathbf{y}' \mathbf{D} \mathbf{y} = 1.$$

- the most similar points are as close as possible.
- First eigenvalue is 0 due to translation invariance.

Key property of Laplacian embedding

Let $y_i = (e_{2,i}, e_{3,i})$ be the embedded point corresponding to household i .

Then the Laplacian eigenmap minimizes

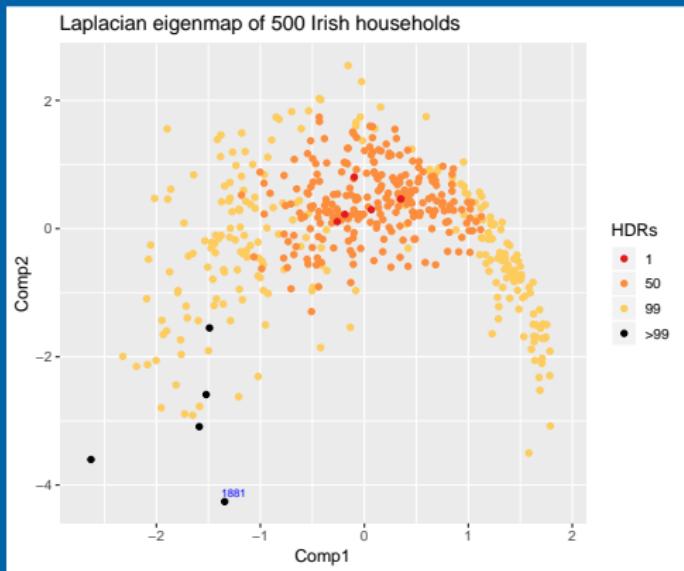
$$\sum_{ij} w_{ij}(y_i - y_j)^2 = \mathbf{y}' \mathbf{L} \mathbf{y} \quad \text{such that} \quad \mathbf{y}' \mathbf{D} \mathbf{y} = 1.$$

- the most similar points are as close as possible.
- First eigenvalue is 0 due to translation invariance.
- Equivalent to optimal embedding using Laplace-Beltrami operator on manifolds.

Visualization via embedding

Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.

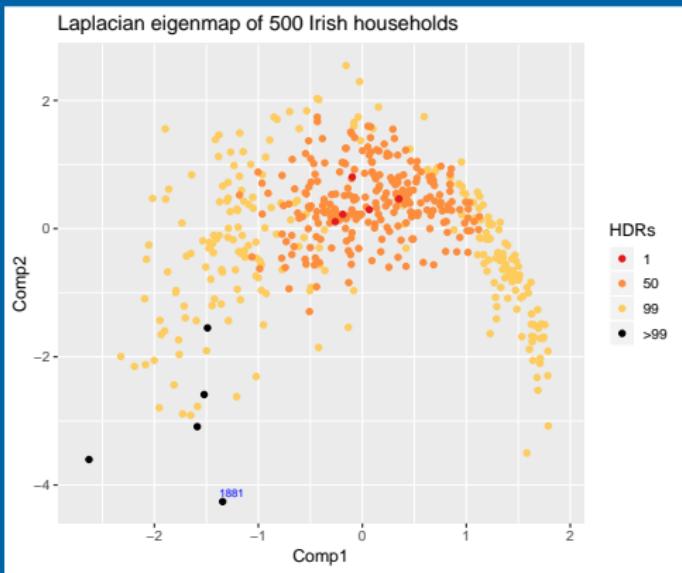
- With more data, we might be able to find clusters of similar households.
- Anomalous households may be due to:
 - malfuctioning equipment
 - unusual schedules or behaviours
 - nefarious activity



Visualization via embedding

Embed conditional densities in a 2d space where the distances are preserved “as far as possible”.

- With more data, we might be able to find clusters of similar households.
- Anomalous households may be due to:
 - malfuctioning equipment
 - unusual schedules or behaviours
 - nefarious activity



Further information:
robjhyndman.com