MONASH University

MONASH BUSINESS SCHOOL

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 meter 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 compar
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

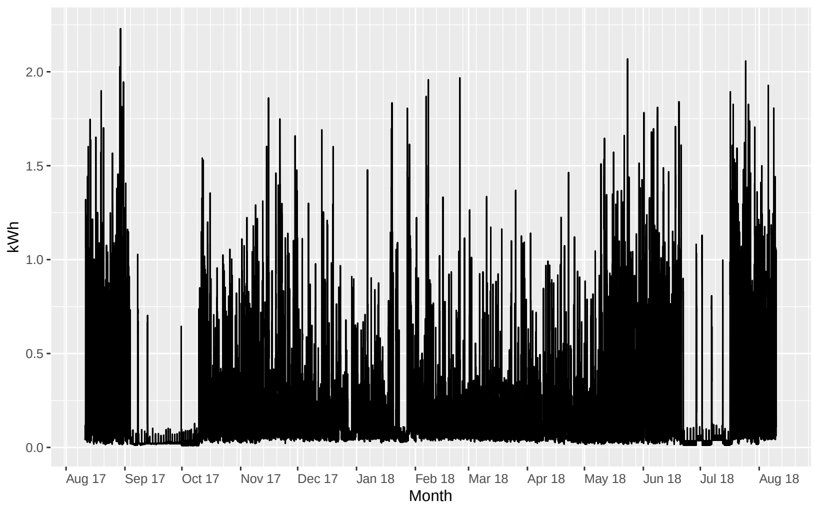
Phone Number

Postcode

Email Address

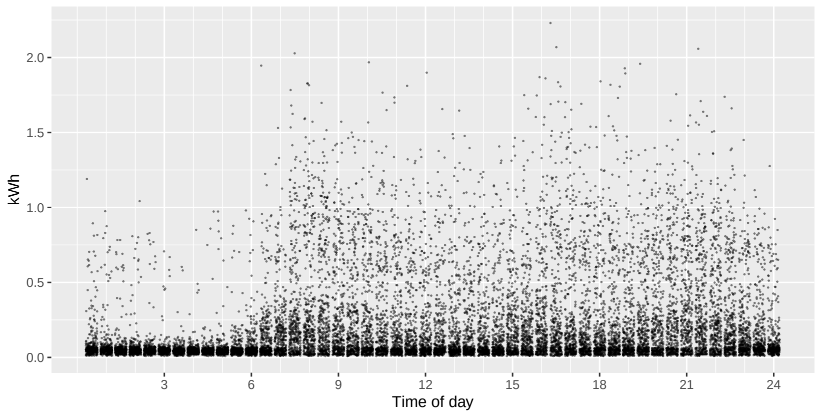
Confirm Email

George’s data

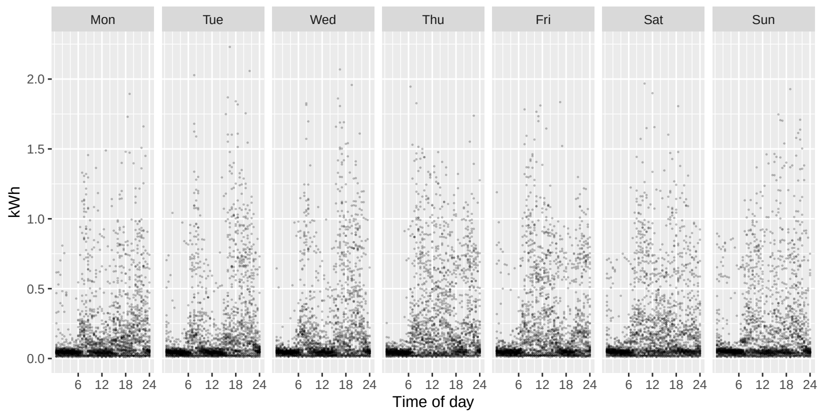


George’s data

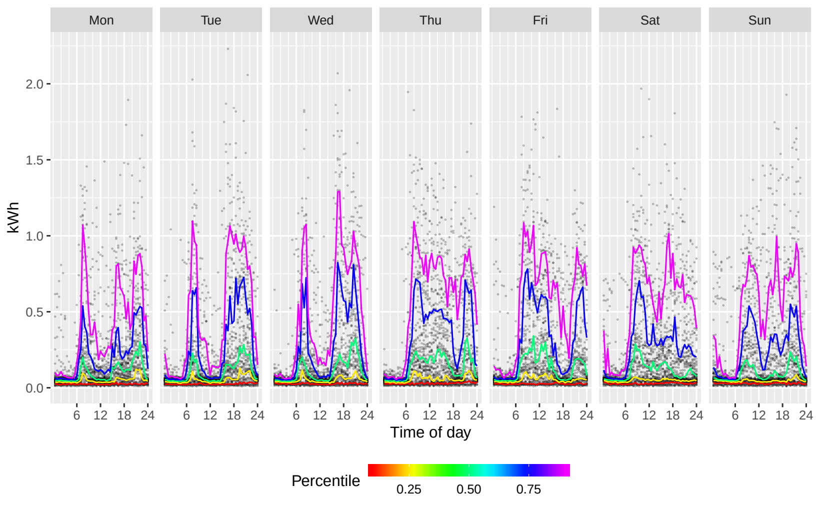
George's data



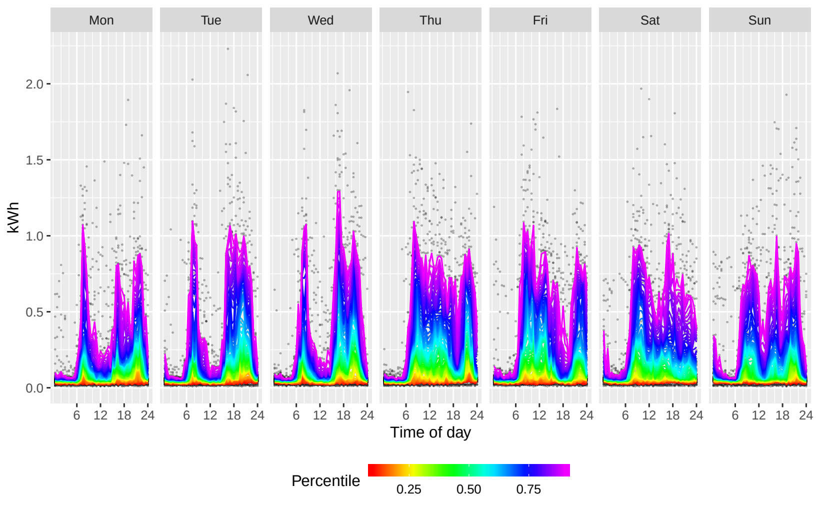
George’s data



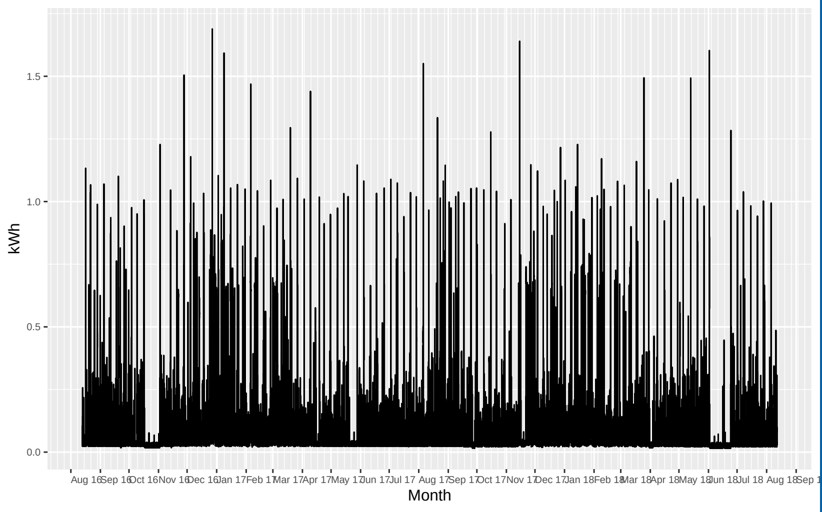
George’s data



George’s data

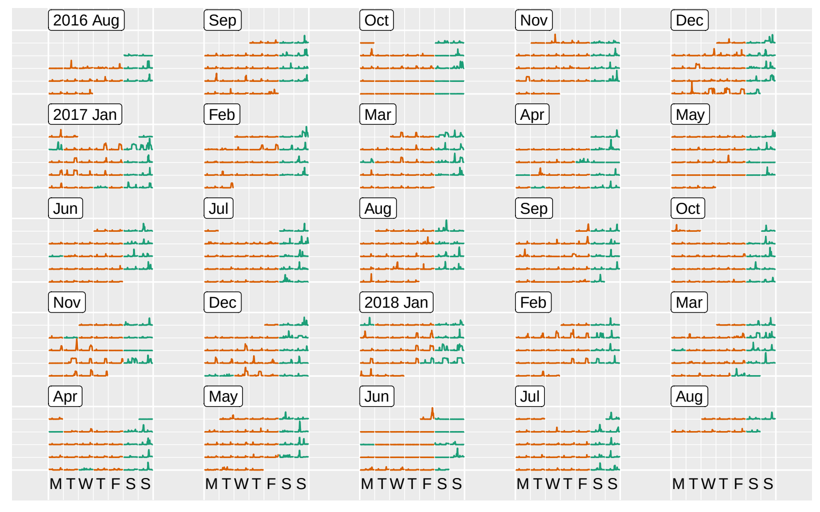


Clare’s data



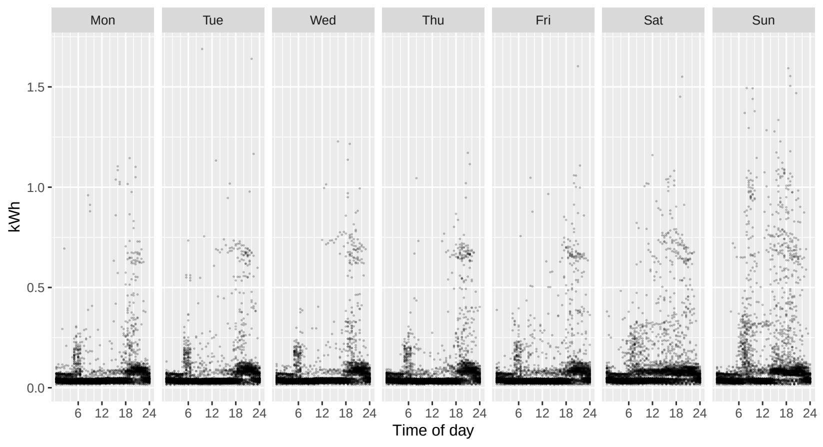
1.5
1.0
0.5
0.0
kWh
Aug 16 Sep 16 Oct 16 Nov 16 Dec 16 Jan 17 Feb 17 Mar 17 Apr 17 May 17 Jun 17 Jul 17 Aug 17 Sep 17 Oct 17 Nov 17 Dec 17 Jan 18 Feb 18 Mar 18 Apr 18 May 18 Jun 18 Jul 18 Aug 18
Month

Clare’s data

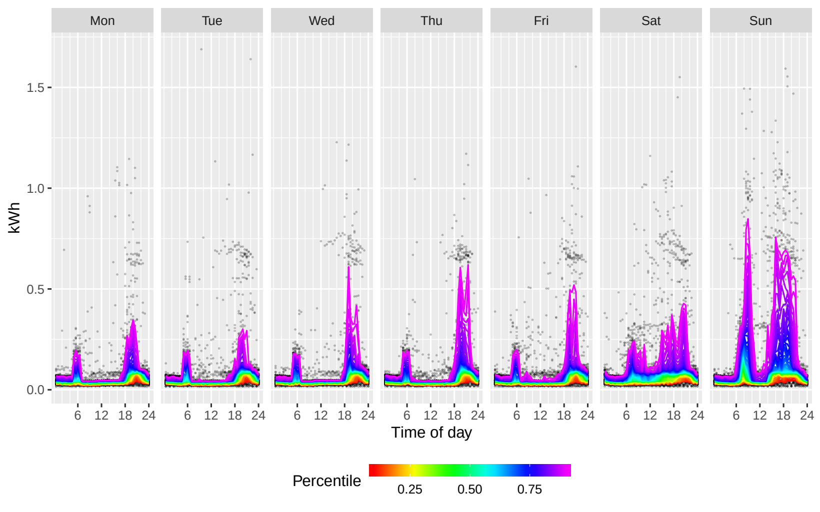


2016 Aug
Sep
Oct
Nov
Dec
2017 Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
2018 Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
M T W T F S S

Clare's data

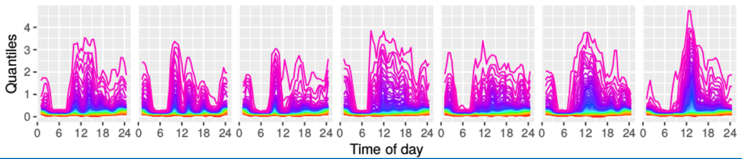


Clare’s data



Mon
Tue
Wed
Thu
Fri
Sat
Sun
kWh
1.5
1.0
0.5
0.0
6 12 18 24
Time of day
Percentile
0.25
0.50
0.75

Percentiles conditional on time of week

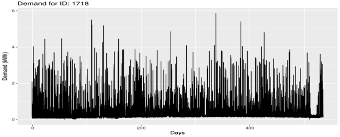


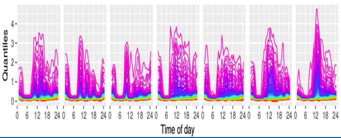
- Percentiles for each household and each half-hour of the week.
- Provides a unique fingerprint of typical usage for a given household.
- 336 probability distributions per household.
- Avoids missing data issues and variation in series length
- Avoids timing of household events, holidays, etc.
- Allows 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 meters

Irish smart meter data

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

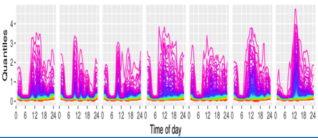




The time series of 535 $\times$ 48 observations per household is mapped to a set of 7 $\times$ 48 $\times$ 99 percentiles giving a bivariate surface for each household.

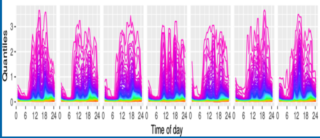
Finding anomalous smart meters

Can we compute pairwise distances between all households?



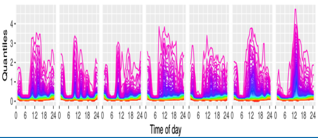
← ? →

Distance



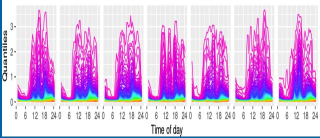
Finding anomalous smart meters

Can we compute pairwise distances between all households?



$\leftarrow$? $\rightarrow$

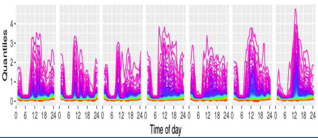
Distance



- Jensen-Shannon measure gives distance between two densities

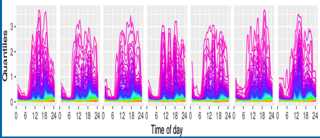
Finding anomalous smart meters

Can we compute pairwise distances between all households?



$\leftarrow$? $\rightarrow$

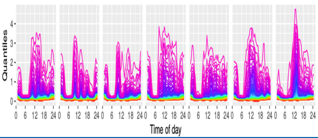
Distance



- Jensen-Shannon measure gives distance between two densities
- Distance between household *i* and household *j*: Δ_{ij} = sum of 7×48 JS distances.

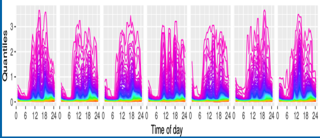
Finding anomalous smart meters

Can we compute pairwise distances between all households?



$\leftarrow$? $\rightarrow$

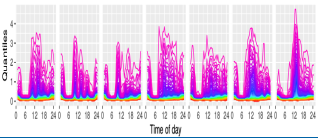
Distance



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

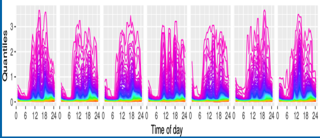
Finding anomalous smart meters

Can we compute pairwise distances between all households?



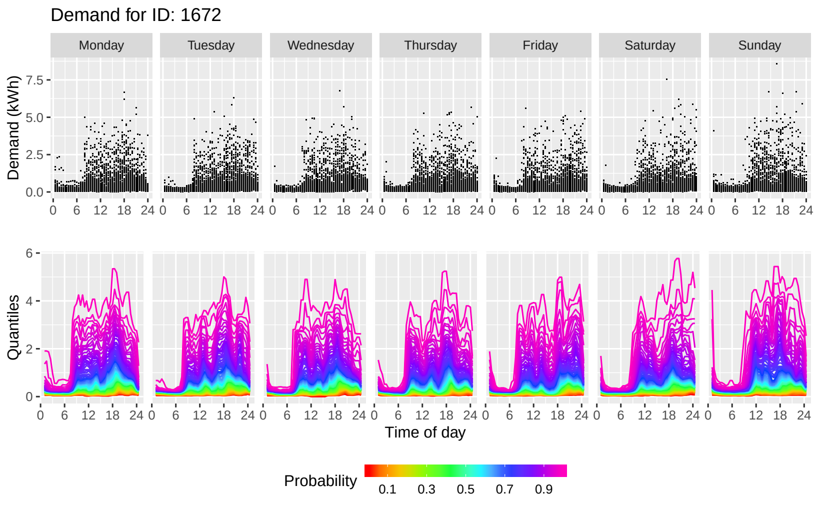
$\leftarrow$? $\rightarrow$

Distance

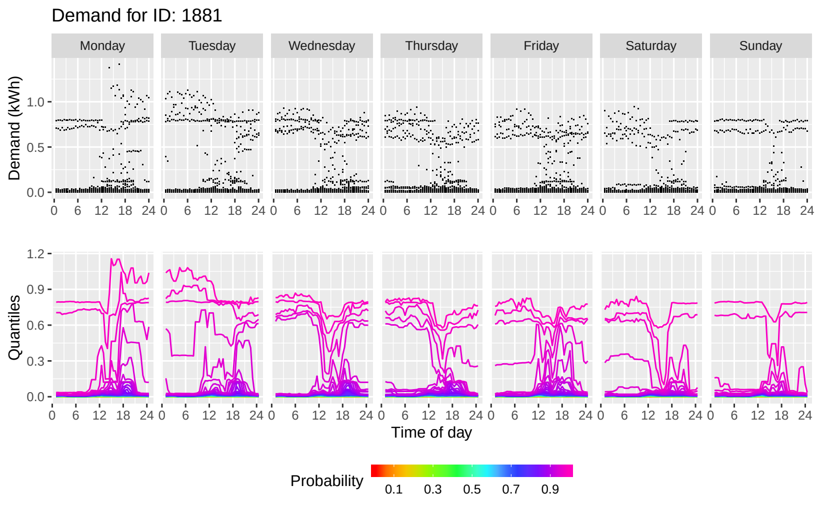


- Jensen-Shannon measure gives distance between two densities
- Distance between household i and household j: Δ_{ij} = sum of 7×48 JS distances.
- Similarity between two households: $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$
- Household typicality: $f_i = \sum_j w_{ij}$.

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 $\boldsymbol{W} = [w_{ij}]$ where $w_{ij} = \exp(-\Delta_{ij}^2/h^2)$.

$$\boldsymbol{D} = \mathrm{diag}(\hat{f}_i) \quad \text{where } \hat{f}_i = \sum_{j=1}^{n} w_{ij}$$

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W} \quad \text{(the Laplacian matrix).}$$

Laplacian eigenmaps

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

Laplacian eigenmaps

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

$$\begin{aligned} \boldsymbol{W} &= [w_{ij}] && \text{where } w_{ij} = \exp(-\Delta_{ij}^2/h^2). \\ \boldsymbol{D} &= \operatorname{diag}(\hat{f}_i) && \text{where } \hat{f}_i = \sum_{j=1}^{n} w_{ij} \\ \boldsymbol{L} &= \boldsymbol{D} - \boldsymbol{W} && \text{(the Laplacian matrix).} \end{aligned}$$

- Solve generalized eigenvector problem: $\boldsymbol{L}\boldsymbol{e} = \lambda \boldsymbol{D}\boldsymbol{e}$.
- Let $\boldsymbol{e}_k$ 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

$$
\begin{aligned}
\boldsymbol{W} &= [w_{ij}] && \text{where } w_{ij} = \exp(-\Delta_{ij}^2/h^2).\\
\boldsymbol{D} &= \operatorname{diag}(\hat{f}_i) && \text{where } \hat{f}_i = \sum_{j=1}^{n} w_{ij}\\
\boldsymbol{L} &= \boldsymbol{D} - \boldsymbol{W} && \text{(the Laplacian matrix).}
\end{aligned}
$$

- Solve generalized eigenvector problem: $\boldsymbol{L}\boldsymbol{e} = \lambda \boldsymbol{D}\boldsymbol{e}$.
- Let $\boldsymbol{e}_k$ be eigenvector corresponding to kth *smallest* eigenvalue.
- Then $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ 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

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

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

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

- 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

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

- 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

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

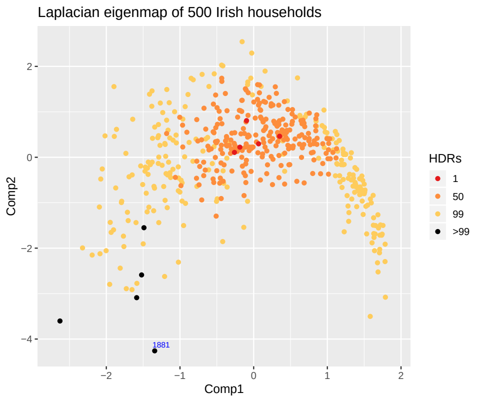
- the most similar points are as close as possible.
- First eigenvalue is 0 due to translation invariance.
- Equivalent to optimal embedding using Laplace-Beltrami operator on manifolds.

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:
  - malfunctioning equipment
  - unusual schedules or behaviours
  - nefarious activity

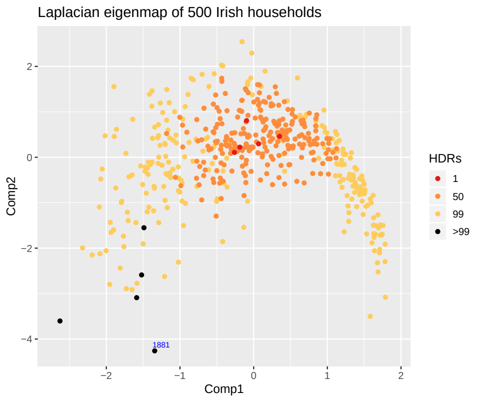


Visualization via embedding

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:
  - malfunctioning equipment
  - unusual schedules or behaviours
  - nefarious activity



Further information:
robjhyndman.com