

Clustering real data with two customers

Contents

1	What you are doing?	1
1.1	entire data	1
1.2	how do they look	2
1.3	make one granularity	4
1.4	conditional list across categories of one granularity	4
1.5	conditional quantiles across categories of one granularity	4
1.6	how do their deciles look for different categories of cyclic granularity	4
1.7	10 iterations of each customer	6
1.8	conditional list across categories of one granularity	8
1.9	conditional quantiles across categories of one granularity	8
1.10	JS Pairwise distances between households	8
2	without scaling leads to mix of two houses	9
3	so trying scaling	9
3.1	conditional list across categories of one granularity	9
3.2	conditional quantiles across categories of one granularity	9
3.3	JS Pairwise distances between households	9

1 What you are doing?

Taking real customers from the smart meter data set and cluster them and see if their distributions are different

- just one cyclic granularity (hod)
- two cyclic granularities (hod and dow)

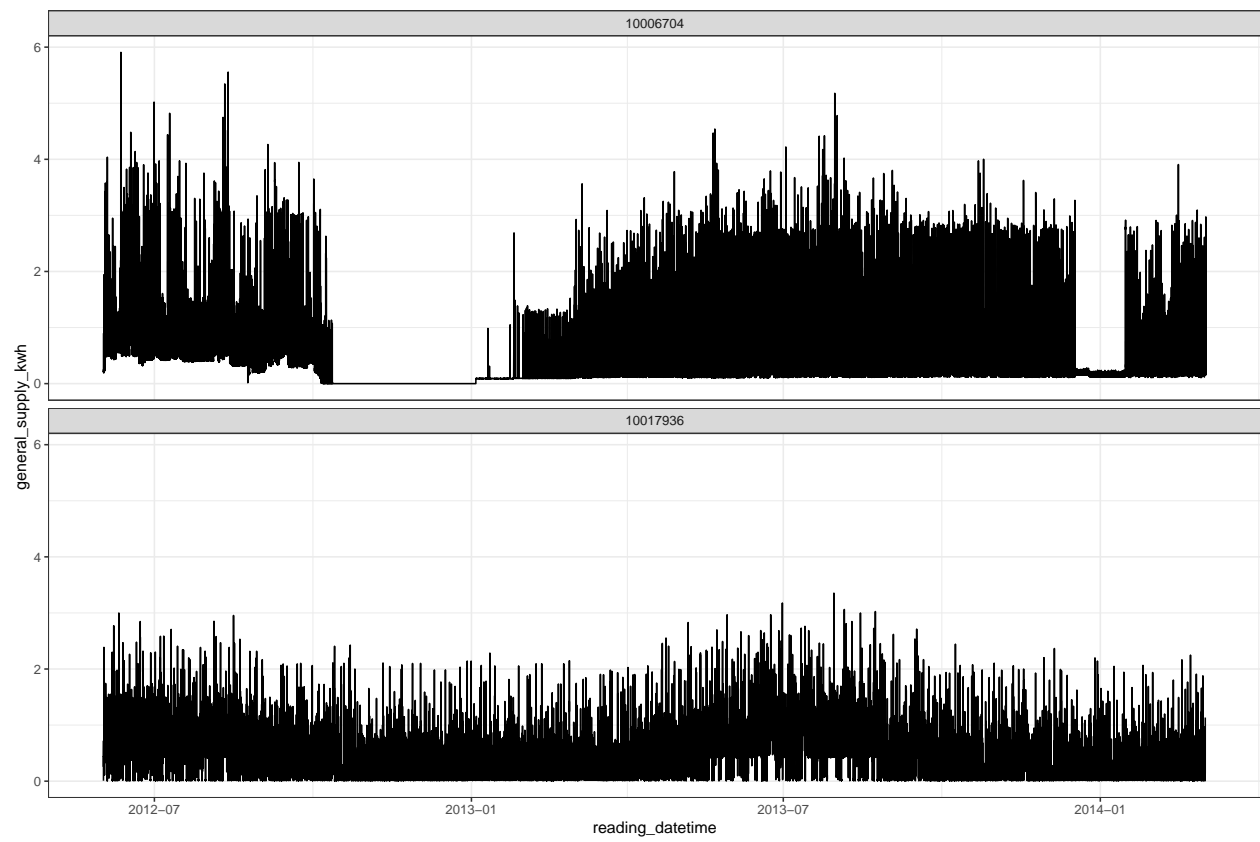
Compute quantiles of conditional distributions Conditional quantiles are obtained for each combination of categories.

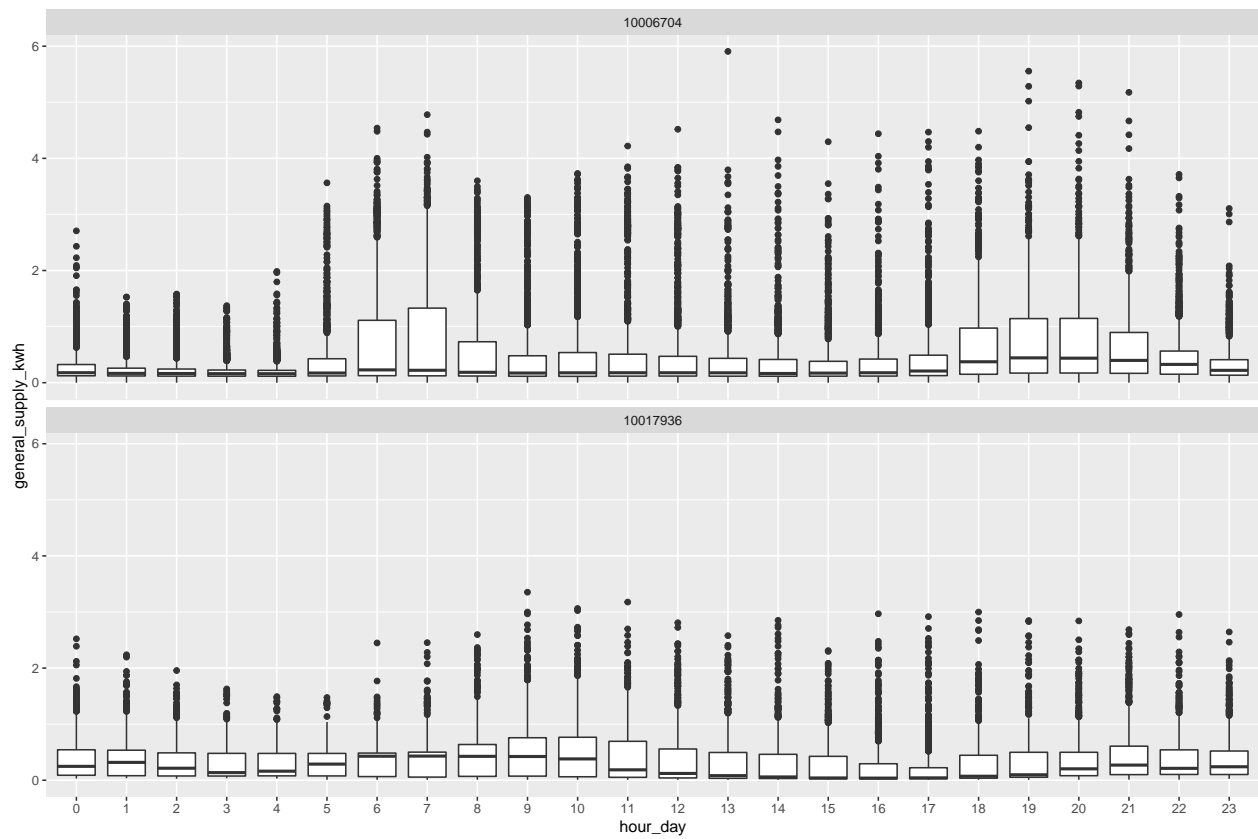
1.1 entire data

```
## # A tibble: 60,925 x 3
##   customer_id reading_datetime   general_supply_kwh
##   <chr>       <dtm>                <dbl>
## 1 10006704    2012-06-01 10:00:00         0.24
## 2 10006704    2012-06-01 10:30:00         0.245
## 3 10006704    2012-06-01 11:00:00         0.206
## 4 10006704    2012-06-01 11:30:00         0.217
## 5 10006704    2012-06-01 12:00:00         0.214
## 6 10006704    2012-06-01 12:30:00         0.23
## 7 10006704    2012-06-01 13:00:00         0.898
## 8 10006704    2012-06-01 13:30:00         0.197
## 9 10006704    2012-06-01 14:00:00         0.255
## 10 10006704    2012-06-01 14:30:00         0.199
```

```
## # ... with 60,915 more rows
```

1.2 how do they look



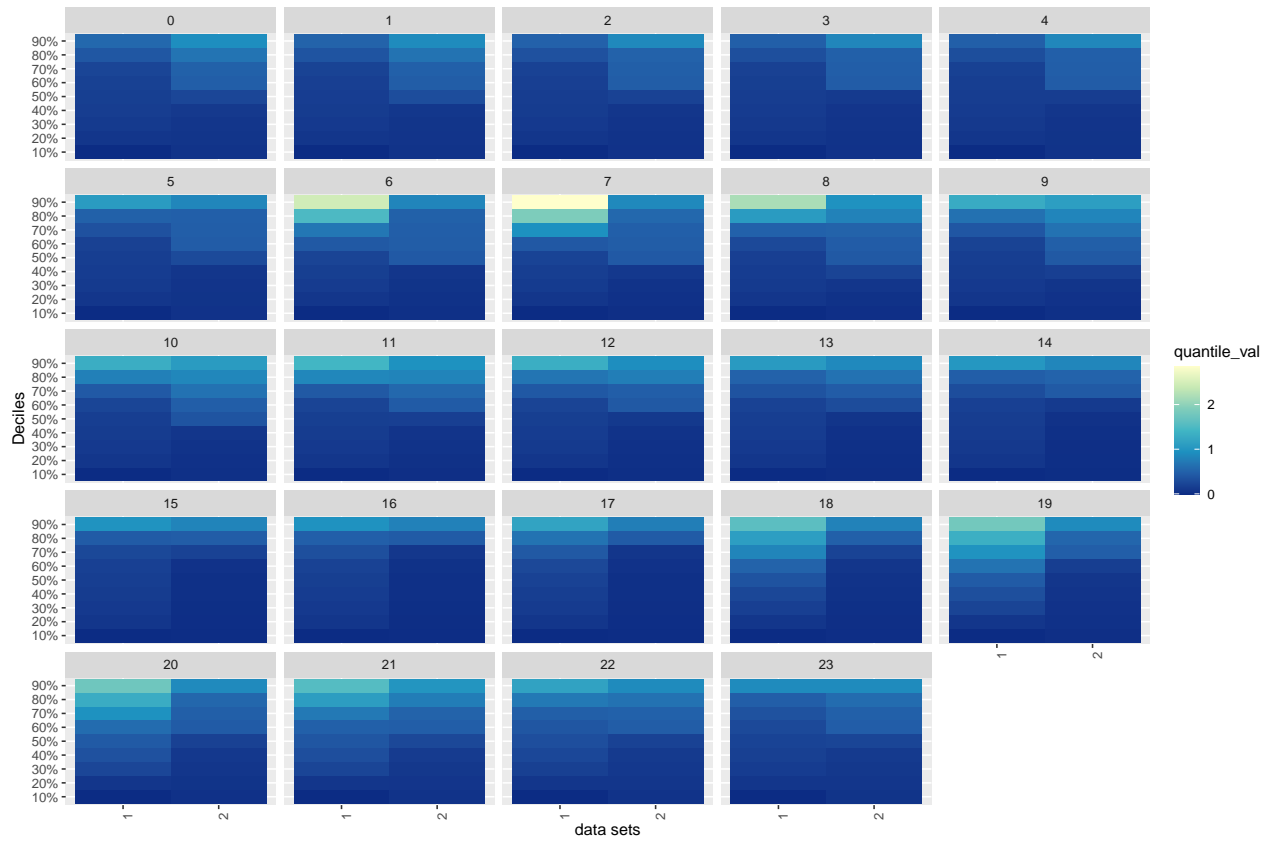


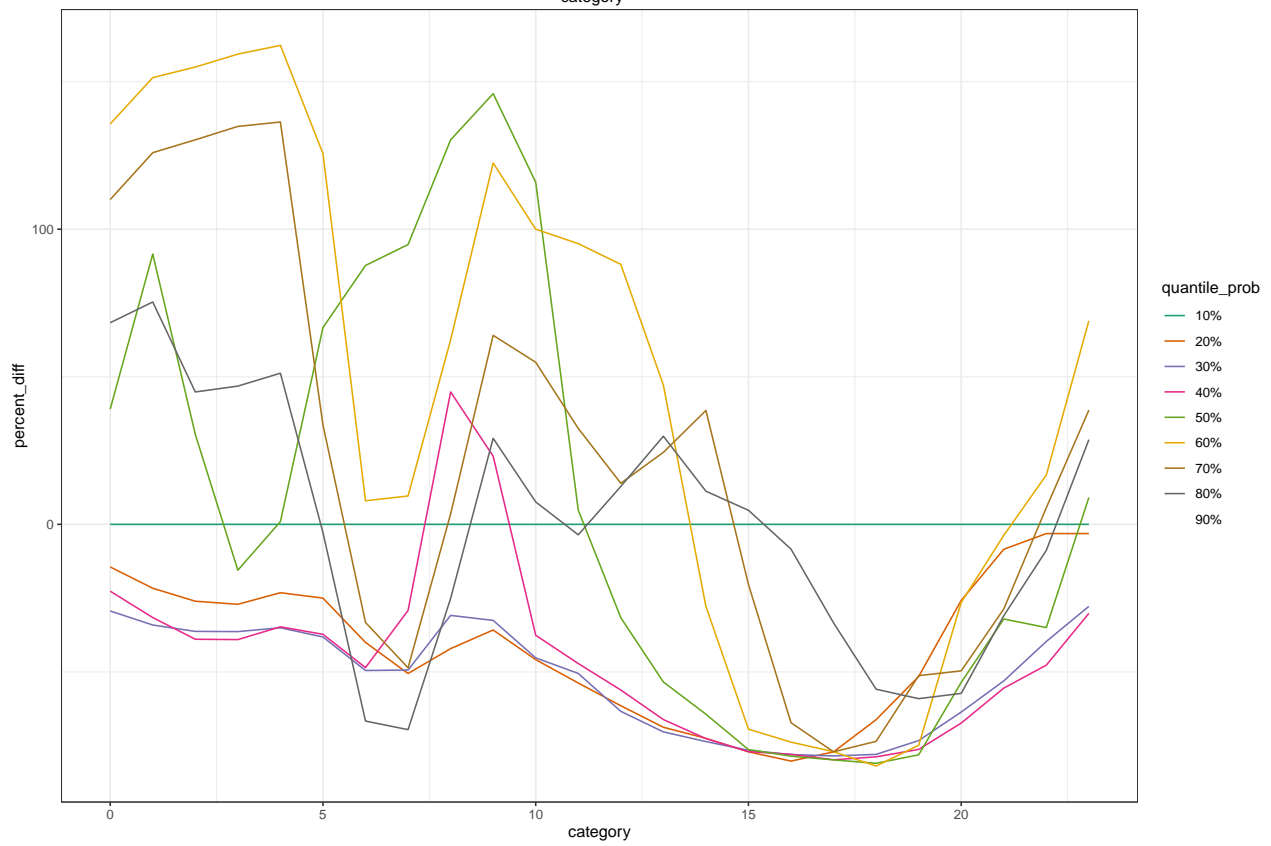
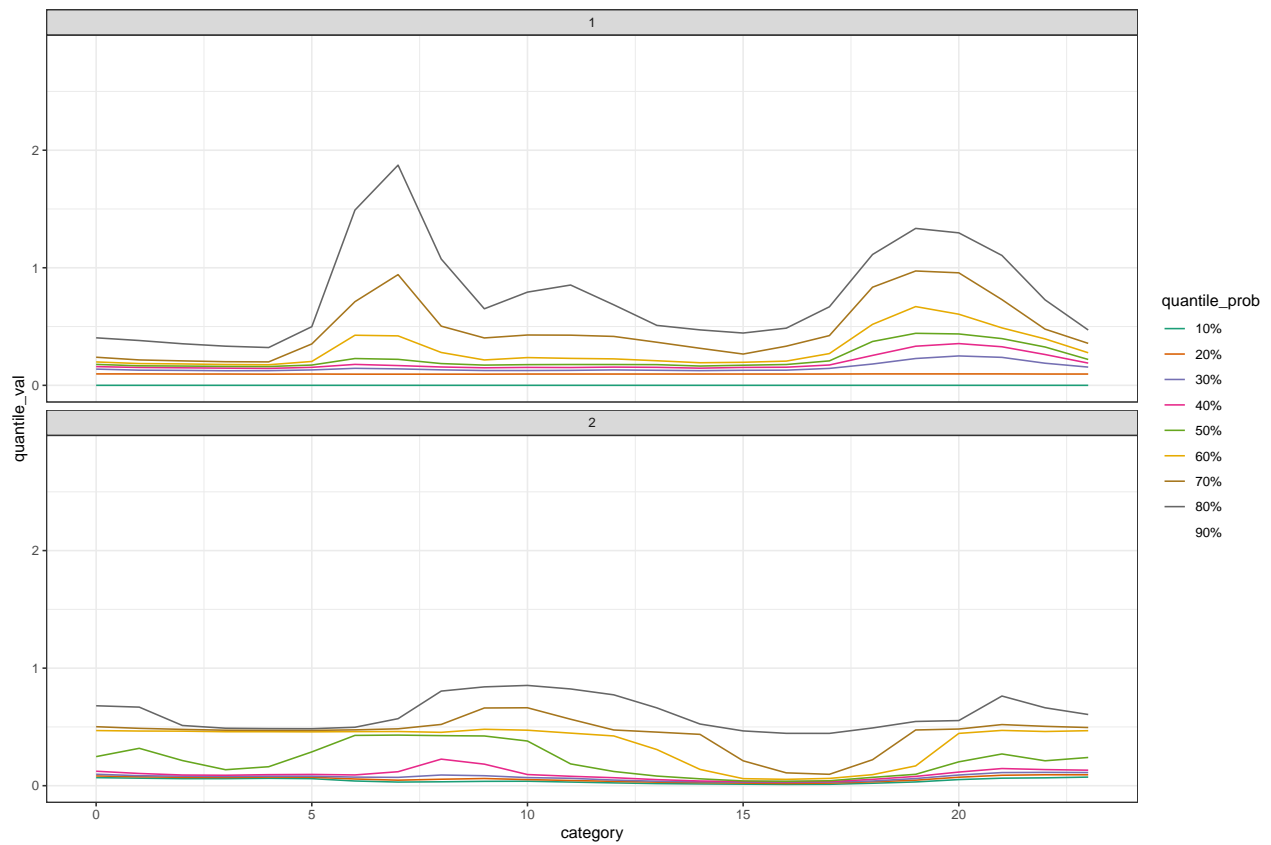
1.3 make one granularity

1.4 conditional list across categories of one granularity

1.5 conditional quantiles across categories of one granularity

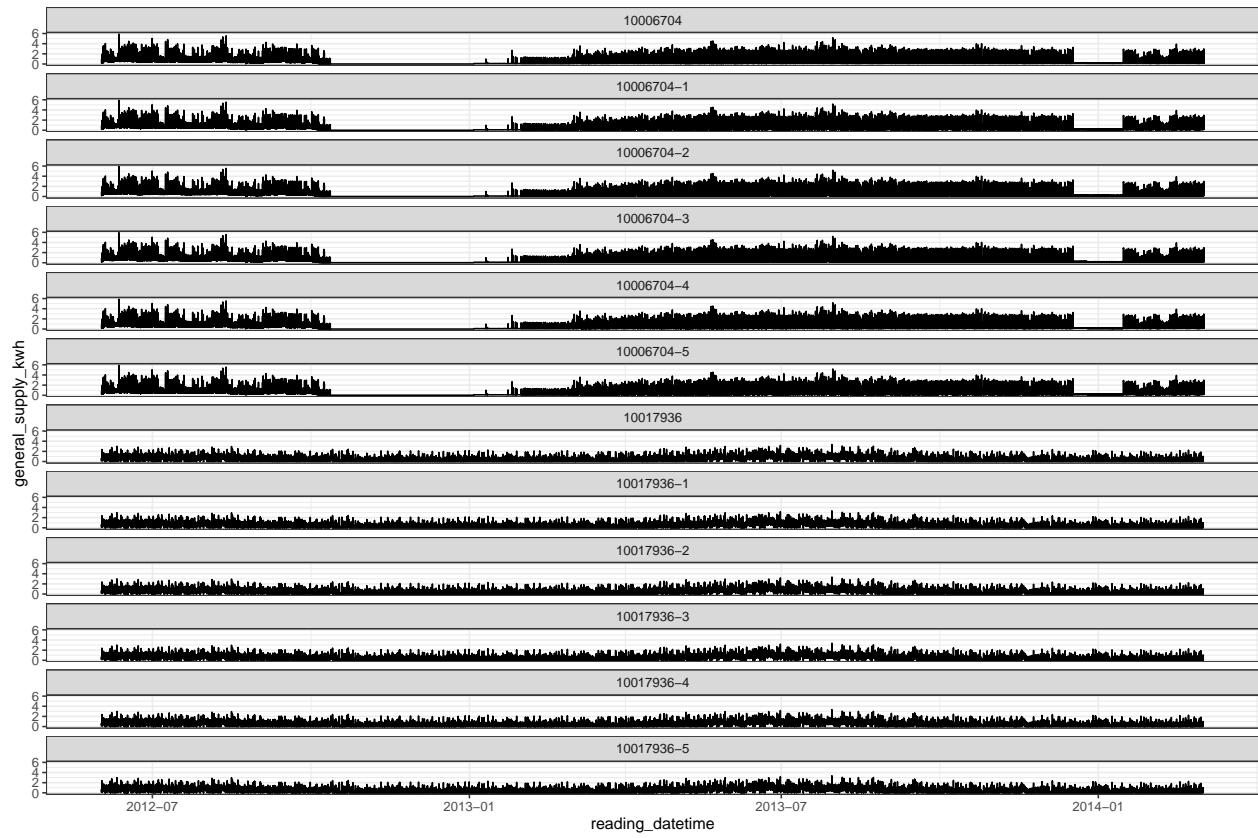
1.6 how do their deciles look for different categories of cyclic granularity



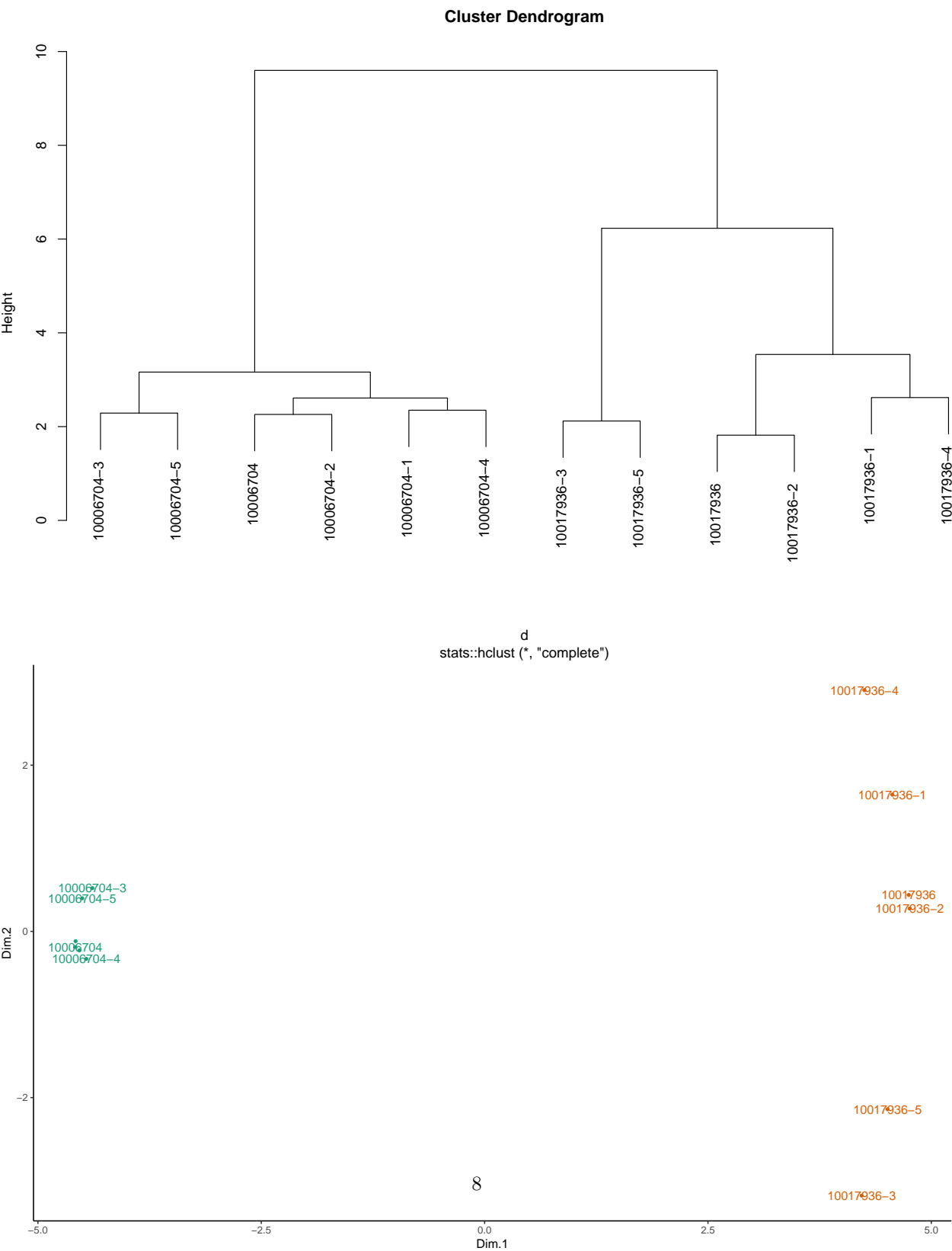


1.7 10 iterations of each customer

1.7.1 Data preparation



- 1.7.2 make one granularity
- 1.8 conditional list across categories of one granularity
- 1.9 conditional quantiles across categories of one granularity
- 1.10 JS Pairwise distances between households
 - 1.10.1 hc



2 without scaling leads to mix of two houses

3 so trying scaling

3.0.1 make one granularity

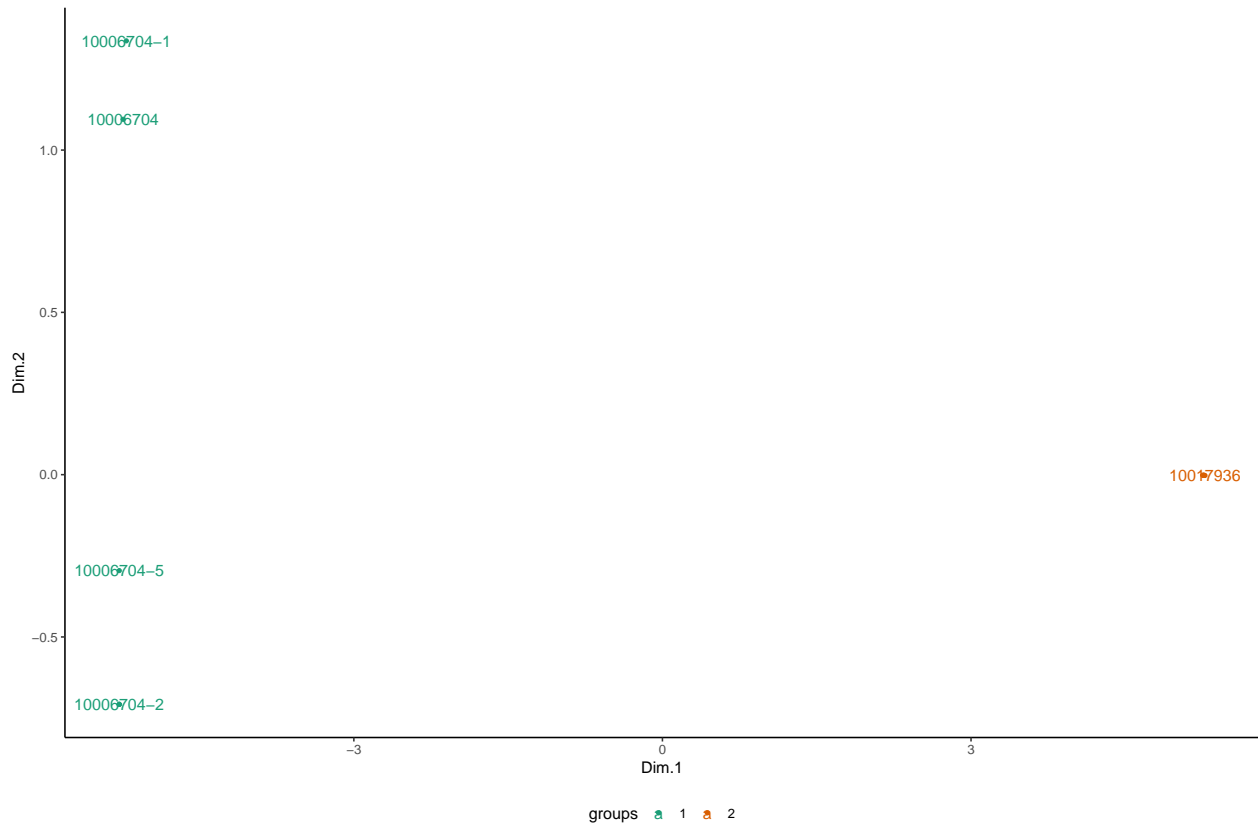
3.1 conditional list across categories of one granularity

3.2 conditional quantiles across categories of one granularity

3.3 JS Pairwise distances between households

3.3.1 hc





JS Pairwise distances between datasets Distance between the data sets is computed as the sum of JS distances across different categories.

Hierarchical clustering with 4 clusters Hierarchical clustering is performed using $k = 4$ and dendrogram observed

Clusters obtained visualized using MDS Each cluster represents data sets from a separate design