Hand picking similar behaving group of customers to check clustering results

A clean dataset is obtained by choosing minimum sum of JS distances from each typical customer based on $quantiles = seq(0.1, 0.9, 0.1)$. The objective is to see if the clustering algorithm then picks the least distant ones as the group.
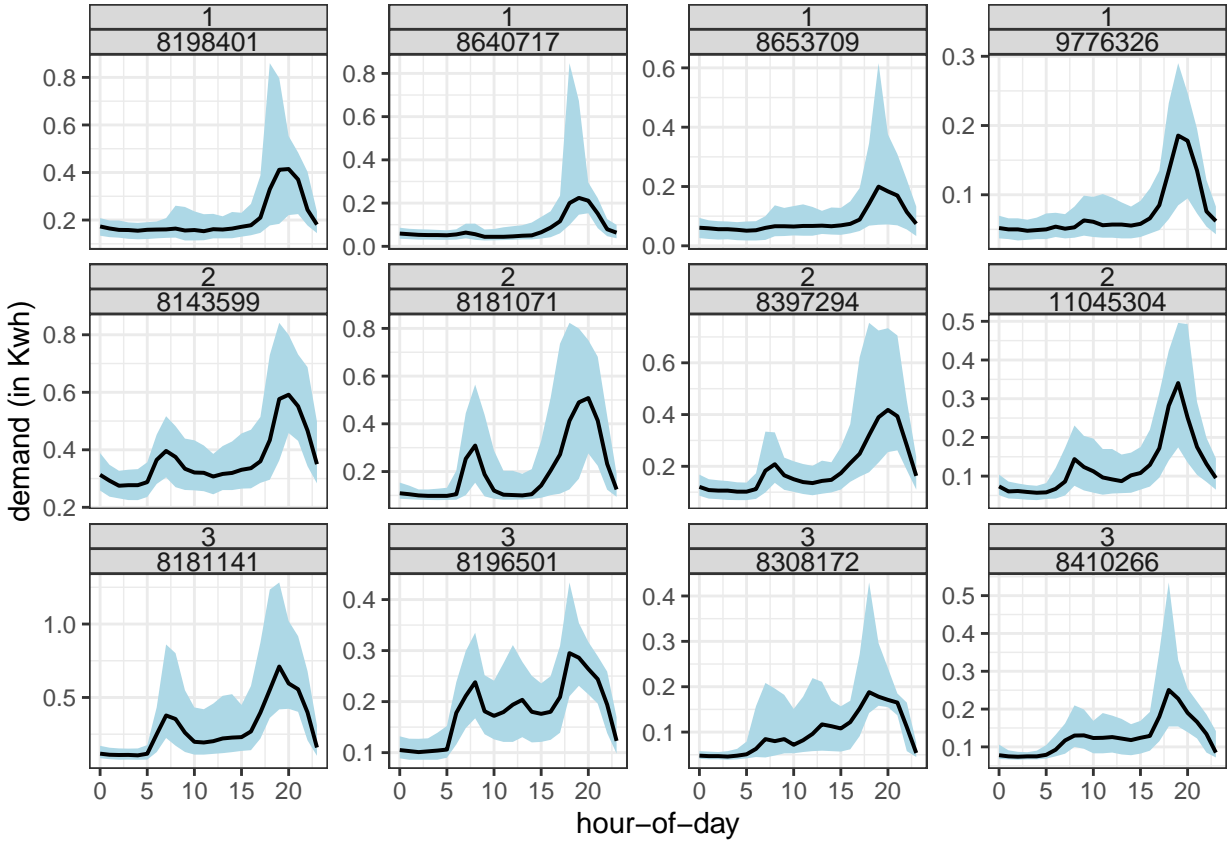
only hod

```
## [1] 11013154
```


Figure 1: Median (black) and quartile deviation (blue region) of hourly demand drawn for few customers showing similar behaviors. Roughly speaking, Design 1 has one evening peak, Design 2 has two peaks and Design 3 has three peaks in a day. Each of design 1, 2 and 3 have six similar behaving customers resulting to 18 time series. We want our clustering results to group each of the designs together.

Do they look similar on the transformed scale?
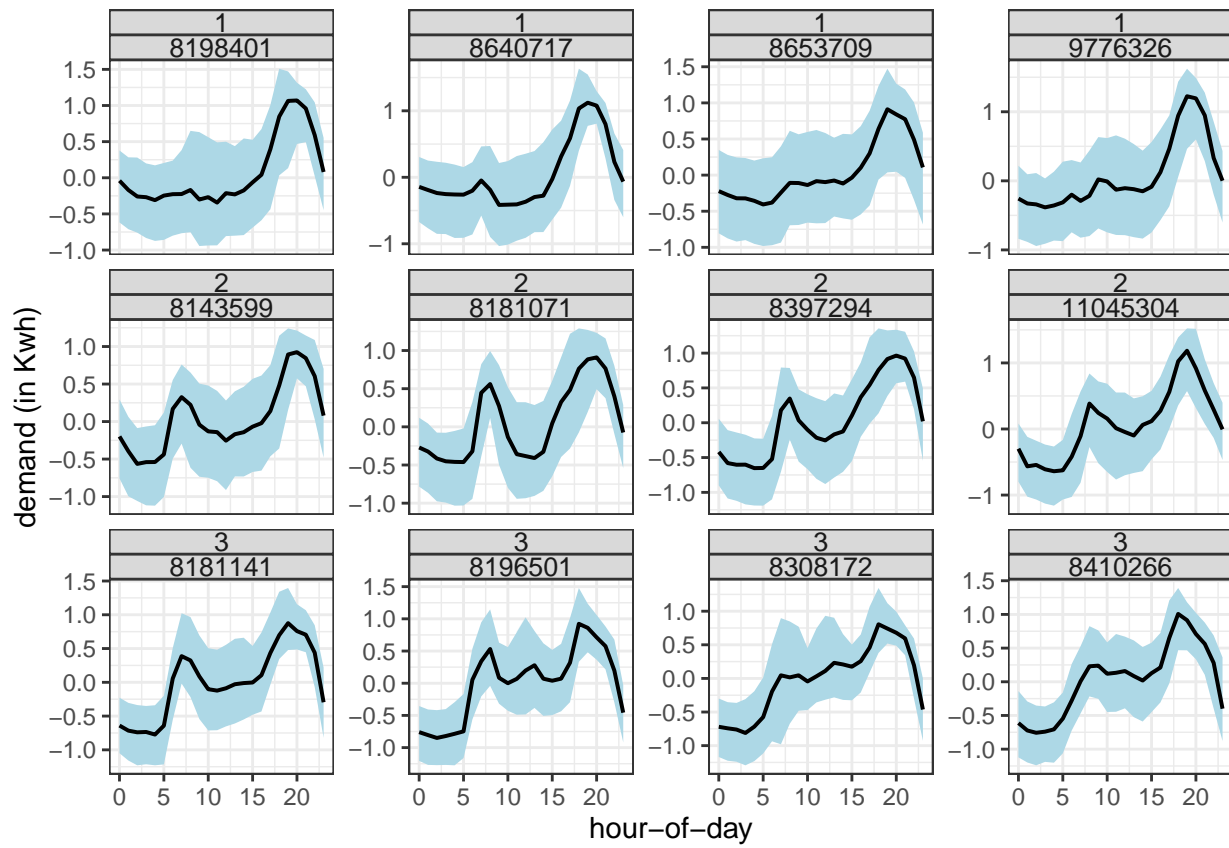
Do they look similar on the robust transformed scale?
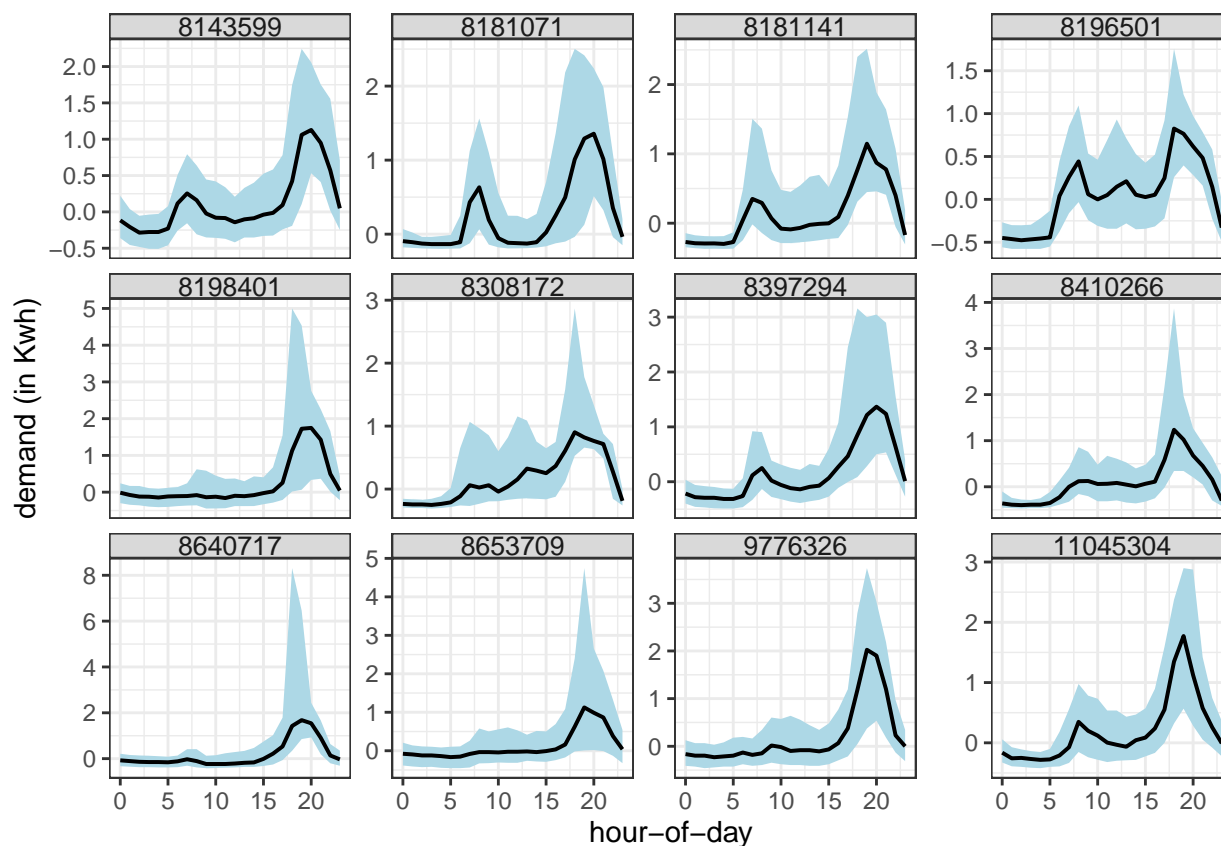
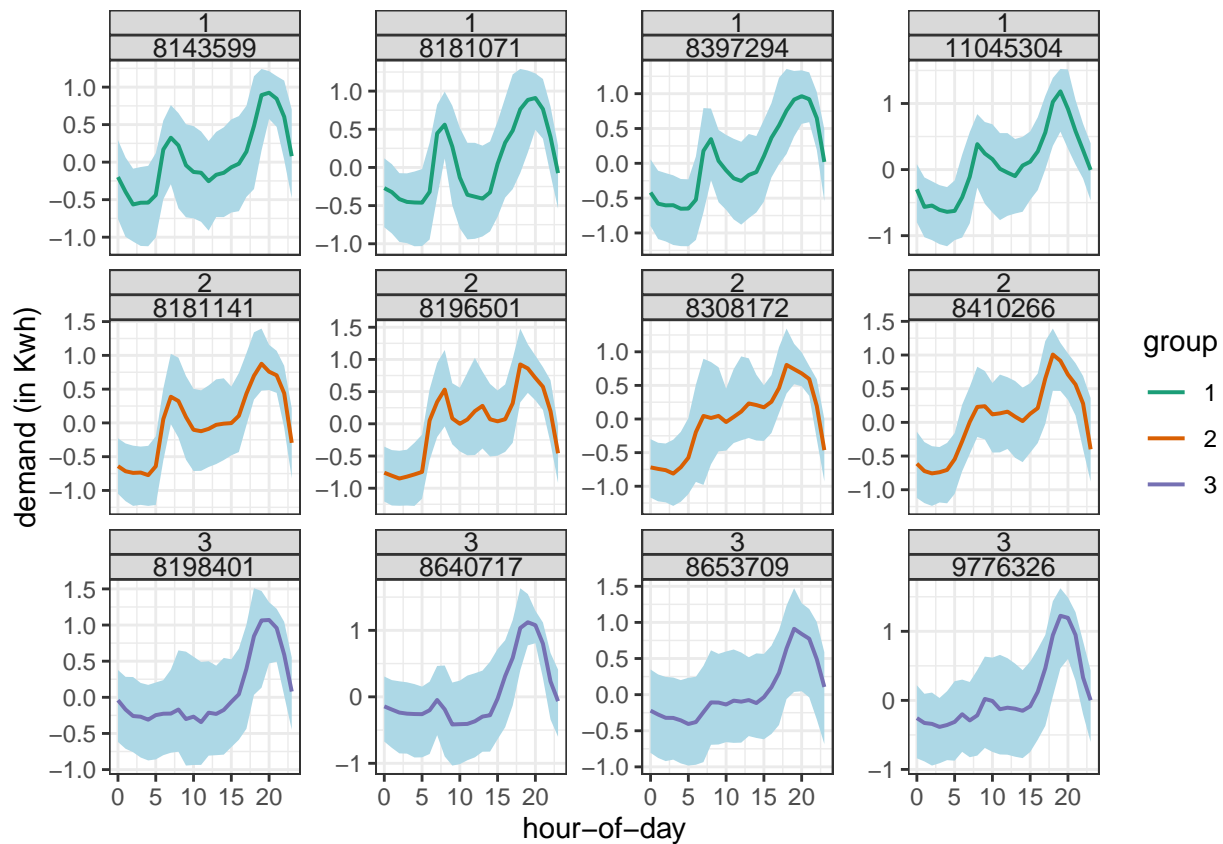## Does the method with NQT scaling identifies the groups?

Yes it does.

```
#quantile_prob_val = c(0.5, 0.75)
#data_pick <- data_pick %>% filter(!(customer_id %in% c(8485375, 8952846)))
library(gracsr)
v2 <- suppressWarnings(
  scaled_dist_gran(data_pick, "hour_day",
                   response = "general_supply_kwh",
                   quantile_prob_val = quantile_prob_clust)) %>% rename("dist_hod" = "dist")
v3 <- suppressWarnings(
  scaled_dist_gran(data_pick, "day_month",
                   response = "general_supply_kwh",
                   quantile_prob_val = quantile_prob_clust)) %>% rename("dist_dom" = "dist")

data_dist <- v3 %>%
  left_join(v2) %>%
  mutate(dist = dist_hod + dist_dom) %>%
    pivot_wider(-c(3, 4),
                names_from = customer_to,
              values_from = dist) %>%
  rename("customer_id" = "customer_from")

## # A tibble: 3 x 2
##   group       n
```

```
##    <int> <int>
## 1     1     4
## 2     2     4
## 3     3     4
```


Does the method with robust scalings identifies the groups?

Yes it does.

```r
library(gracsr)
v2 <- suppressWarnings(
  scaled_dist_gran_iqr(data_pick, "hour_day",
                response = "general_supply_kwh",
                quantile_prob_val = quantile_prob_clust)) %>% rename("dist_hod" = "dist")
v3 <- suppressWarnings(
  scaled_dist_gran_iqr(data_pick, "day_month",
                response = "general_supply_kwh",
                quantile_prob_val = quantile_prob_clust)) %>% rename("dist_dom" = "dist")

data_dist <- v3 %>%
  left_join(v2) %>%
  mutate(dist = dist_hod) %>%
    pivot_wider(-c(3, 4),
                names_from = customer_to,
              values_from = dist) %>%
  rename("customer_id" = "customer_from")
```

```
## # A tibble: 3 x 2
```

```
##   group      n
##   <int>  <int>
## 1     1      8
## 2     2      3
## 3     3      1
```