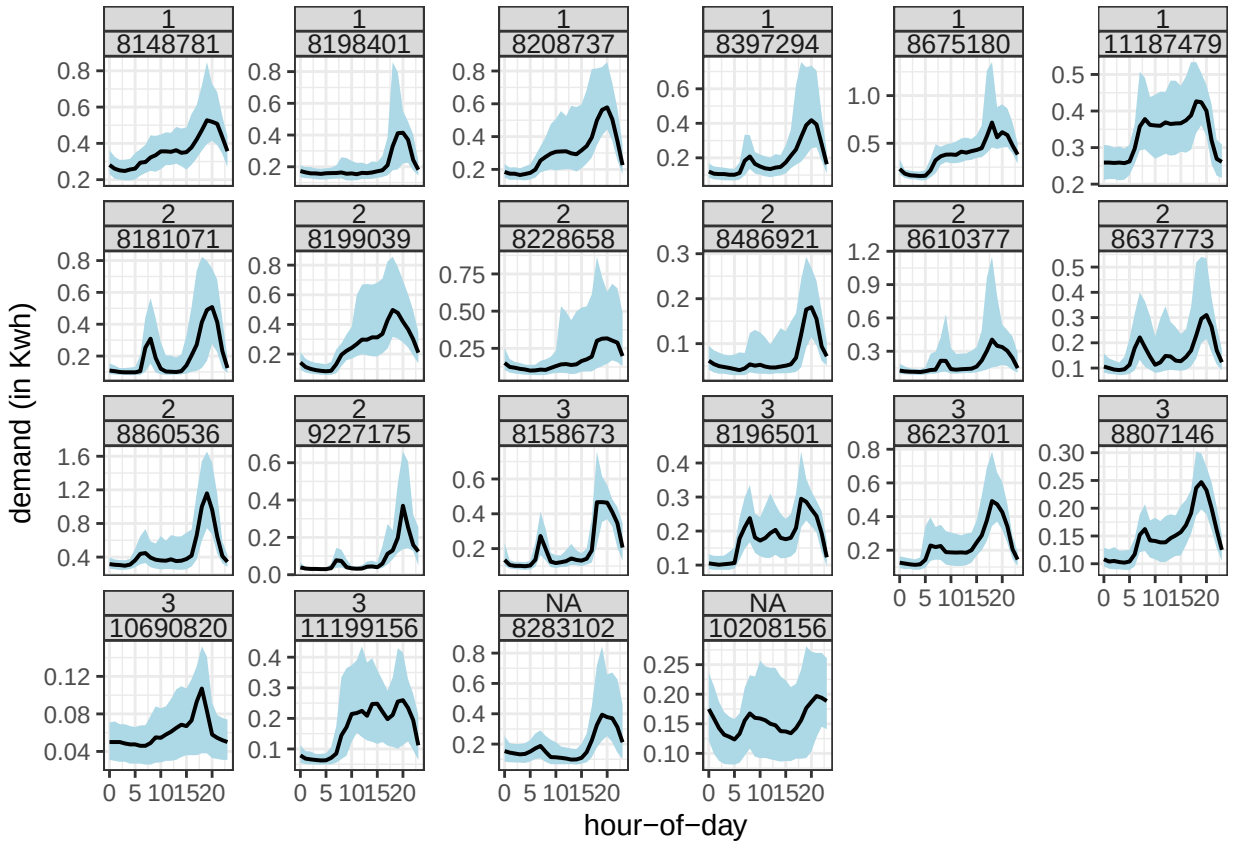
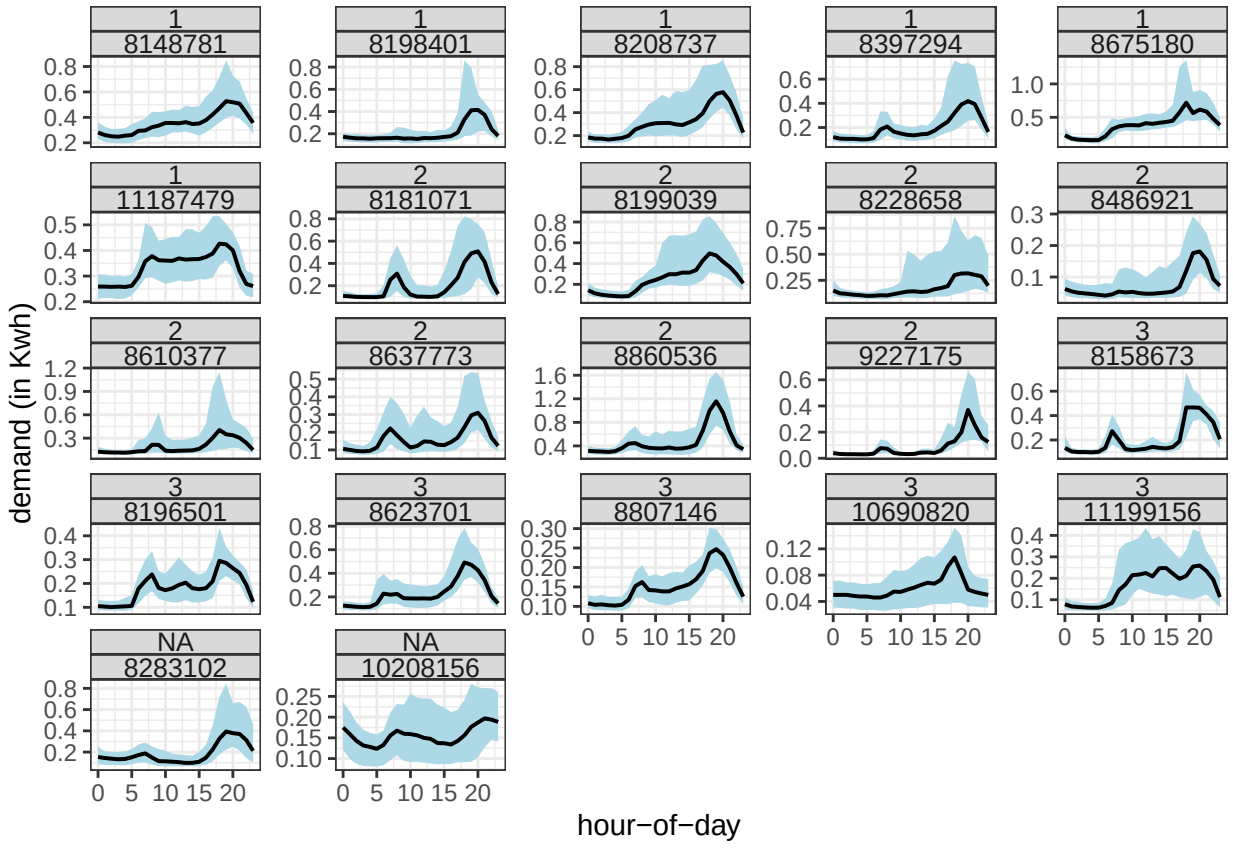
Hand picking similar behaving group of customers to check clustering results

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



Do they look similar on the transformed scale?



only hod

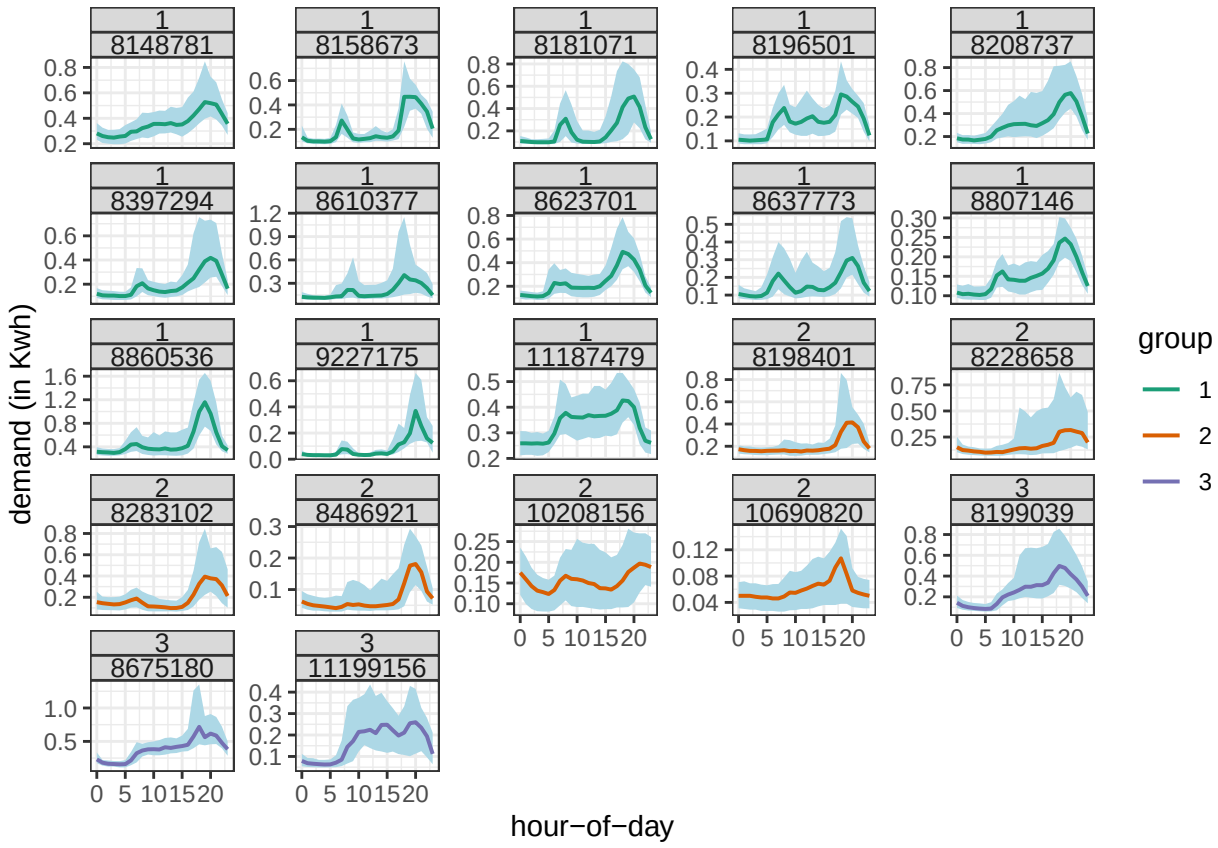
Does hod as the only variable correctly identifies the groups?

```
#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     13
## 2      2      6
## 3      3      3
```


1 8148781
1 8158673
1 8181071
1 8196501
1 8208737
1 8397294
1 8610377
1 8623701
1 8637773
1 8807146
1 8860536
1 9227175
1 11187479
2 8198401
2 8228658
2 8283102
2 8486921
2 10208156
2 10690820
3 8199039
3 8675180
3 11199156
demand (in Kwh)
hour−of−day
group
1
2
3