

# **Termpaper**

Sindre og Morten

## **Oppgave 1**

### **A**

Hedoniske modeller blir tatt i bruk for å måle marginal «willingness to pay» (MWTP), denne blir tatt i bruk for å måle boligmarkedet sine miljømessige kvaliteter. Den hedoniske modellen ble fremstilt ved å ta i bruk et stort og avansert datasett fra boligmarkeder i store byer som har en avansert økonomi. Jo mer tilgang vi får på data til disse byområdene, desto bedre analyser får vi med bruk av en hedonisk modell.

Rosens første steg handler om å kunne definere et marked som vil tilfredsstille «law of one function». «Law of one function» er at hus som er identiske blir solgt for den samme prisen i et gitt marked. Steg to handler om data innsamling. Det forteller oss at det er met oppnålig med å ha et tilfeldig utvalg. Slike datainnsamlinger til en hedonisk modell som omhandler eiendomsverdi har som regel fokus på eneboliger.

### **C - Why could it be important to define a housing market as a single metropolitan (or travel to work) area and studying a relatively short period of time, when estimating**

#### **i. hedonic price function?**

Bishop forklarer at i hedonisk modell innebærer det at markedet bør defineres slik at «loven om en prisfunksjon» oppfylles (Bishop mfl. 2020). Med en prisfunksjon menes det at identiske boliger vil selges til samme pris gjennom hele markedet. Likevel, de nøyaktige romlige og tidsmessige grensene som tilfredsstiller denne betingelsen kan variere på tvers av rom og tid ettersom informasjon, institusjoner og flyttekostnader endres (Bishop mfl. 2020). I praksis er det vanlig å definere markedet som et enkelt storbyområde (single metropolitan) over noen år. Flyttekostnader vil egentlig bryte med loven om en prisfunksjon, men for husholdninger som flytter innenfor dette storbyområdet er det lite sannsynlig at disse kostnadene vil variere noe særlig.

Årsaken til det er at de fysiske kostnadene som vel som de økonomiske (f.eks lastebil-leie) ikke endrer seg på tvers av destinasjonssteder i hovedstadsområdet. De psykologiske kostnadene er også mer begrenset ved flytting fordi det tillattes i lettere grad å opprettholde relasjonene til familie, venner og nabolag. Dette gjør at loven om én prisfunksjon opprettholdes mellom lokasjoner i et storbyområde gjennom arbitrasje. Altå, hvis to hus som er tilsvarende like i samme storbyområdet selges, så velger kjøperne naturligvis det rimeligste.

**ii. Explain intuitively, by using an example, why it is important to avoid omitted variable bias when estimating a hedonic price model.**

Gjennom empirien og teorien er det grunn til å tro av miljøfaciliteter er romlig korrelert på grunn av de naturlige trekken ved geografien som for eksempel fjell og hav, miljøtilbakemeldingseffekter (f.eks urbane varme øyer) og stemming på lokale felles godter. Dette potensialet for romlig korrelasjon har ført til utbredt bekymring for utelatt-variabel skjevhets (Bishop mfl. 2020). Det er først og fremst fordi det virker usannsynlig at forskere vil være i stand til å inkludere alle bekvemmeligheter som betyr noe for kjøpere. I tillegg vil uobserverte faciliteter sannsynligvis være korrelert med tilbuddet av interesse, og dermed forårsake skjevhets. Dette kan forklares ved et eksempel: Hvis velstående og velutdannede boligkjøpere flytter til områder med bedre luftkvalitet og deretter stemmer for å øke offentlig skolefinansiering, vil estimater av MWTP for luftkvalitet være skjev oppover hvis skolekvalitet uteslates fra modellen. Potensialet for denne typen oppførsel fra huseiere betyr at for at de resulterende estimatene skal være troverdige, må forskningsdesignet isolere eksogen variasjon i tilbuddet av interesse (Bishop mfl. 2020).

## Oppgave 2

i.

Så over variablene og definisjonene på dem hos Kaggle.

ii.

Laster inn data

```
kc_house_data <- read_csv("kc_house_data.csv")
```

```
Rows: 21613 Columns: 21
-- Column specification -----
Delimiter: ","
chr   (1): id
```

```
dbl  (19): price, bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterf...
dttm  (1): date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

iii.

```
kc_house_data <- arrange(kc_house_data, desc(date))
```

iv.

```
kc_house_data <- kc_house_data %>%
  distinct(id, .keep_all = TRUE)
```

v & vi.

```
kc_house_data_sf <- st_as_sf(kc_house_data,
  coords = c(x = "long",
             y = "lat"),
  crs = 4326) %>%
  st_transform(2926)
```

vii.

koordinatene til Seattle:

Norske (*wikipedia*):

- 47.60621, -122.33207

Engelske (*wikipedia*):

- 47.609722, -122.333056

```
cbd <- st_sfc(st_point(c(-122.33207, 47.60621)), crs = 4326) %>%
  st_transform(2926)
```

viii.

```
kc_house_data_sf <- kc_house_data_sf %>%
  mutate(dist_cbd = st_distance(cbd, .,
                                by_element = TRUE),
        dist_cbd_km = set_units(dist_cbd, km)
  )
```

## Oppgave 3

```
kc_wadoh_map <- here("maps/WADOH_Environmental_Health_Disparities_Index_Calculated_for_King_County___wadoh_map.shp") %>%
  st_read() %>%
  st_transform(2926)

Reading layer `WADOH_Environmental_Health_Disparities_Index_Calculated_for_King_County___wadoh_map` from file
using driver `ESRI Shapefile'
Simple feature collection with 398 features and 192 fields
Geometry type: MULTIPOLYGON
Dimension:     XY
Bounding box:  xmin: -122.528 ymin: 47.08446 xmax: -121.0657 ymax: 47.78058
Geodetic CRS:  WGS 84

kc_wadoh_map <- kc_wadoh_map %>%
  select(
    GEO_ID_TRT,
    EHD_percen,#Environmental Health Index, weighted score many vars
    linguist_2,#Pop. age 5+ speaking English less than "very well"
    poverty_pe,#Percentage people living in poverty
    POC_percen,#People of Color in percentage of pop. in tract
    transporta,#% of income spent on transportation median family in tract
    unemploy_2,#percentage unemployed
    housing_pe,#% of households in group "Unaffordable Housing" (>30% inc.)
    traffic_pe,#% of pop. near heavy traffic roadways
    diesel,# nox concentration
    ozone,# ozone concentration
    PM25, # concentration of Particulate Matter in air
    toxic_rele, # Toxic release from factories
    hazardous_, # Hazardous Waste Treatment Storage and disposal Facilities
    lead_perce, # measure of Lead paint in houses
```

```

superfund, # Proximity to contaminated sites on national list
facilities, # Proximity to Risk Management Plan Facilities
wastewater, # Proximity to wastewater facilities
sen_pop_pe, # % pop. over 65
socio_perc # score social economic determinants, low best
)

acs_b19101_fam_inc <- read.dbf("../maps/censusSHP/acs_b19101_familyincome.dbf")
attach(acs_b19101_fam_inc)

acs_b19101_fam_inc <- acs_b19101_fam_inc %>%
  mutate(low = (E19101138 + E19101139 + E19101140 + E19101141 +
             E19101142 + E19101143)/E19101137) %>%
  mutate(mid = (E19101144 + E19101145 + E19101146 + E19101147 +
             E19101148 + E19101149)/E19101137) %>%
  mutate(high = (E19101150 + E19101151 + E19101152 + E19101153)/E19101137)

acs_b19101_fam_inc <- acs_b19101_fam_inc %>%
  select(GEOIDTRT, low, mid, high) %>%
  rename(GEO_ID_TRT = GEOIDTRT)

kc_wadoh_map_2 <- left_join(
  acs_b19101_fam_inc,
  st_drop_geometry(kc_wadoh_map),
  by = "GEO_ID_TRT")

kc_tracts10 <- here("../maps/censusSHP/tracts10.shp") %>%
  st_read() %>%
  st_transform(2926)

Reading layer `tracts10' from data source
`/Users/sindreespedal/Documents/HVL /Høst 2022/MSB 204 - Bolig - R/maps/censusSHP/tracts10'
using driver `ESRI Shapefile'
Simple feature collection with 398 features and 22 fields
Geometry type: POLYGON
Dimension:      XY
Bounding box:  xmin: 1217085 ymin: 31406.52 xmax: 1583210 ymax: 287947.2
Projected CRS: NAD83(HARN) / Washington North (ftUS)

```

```

kc_tracts10_shore <- here("../maps/censusSHP/tracts10_shore.shp") %>%
  st_read() %>%
  st_transform(2926)

Reading layer `tracts10_shore' from data source
  `/Users/sindreespedal/Documents/HVL /Høst 2022/MSB 204 - Bolig - R/maps/censusSHP/tracts10_
  using driver `ESRI Shapefile'
Simple feature collection with 398 features and 22 fields
Geometry type: MULTIPOLYGON
Dimension:      XY
Bounding box:  xmin: 1220306 ymin: 31406.52 xmax: 1583210 ymax: 287675.5
Projected CRS: NAD83(HARN) / Washington North (ftUS)

kc_tracts10_env_data <- left_join(
  kc_tracts10, kc_wadoh_map_2,
  by = "GEO_ID_TRT"
)
kc_tracts10_shore_env_data <- left_join(
  kc_tracts10_shore, kc_wadoh_map_2,
  by = "GEO_ID_TRT"
)

kc_houses_env_var <- st_join(kc_house_data_sf, kc_tracts10_env_data)
kc_tracts10_shore_env_var <- st_join(kc_house_data_sf, kc_tracts10_shore_env_data)

st_write(kc_house_data, "../maps/kc_house_data.gpkg", append = FALSE)

Deleting layer `kc_house_data' using driver `GPKG'
Writing layer `kc_house_data' to data source
`../maps/kc_house_data.gpkg' using driver `GPKG'
Writing 21436 features with 21 fields without geometries.

st_write(kc_tracts10, "../maps/kc_tracts10.gpkg", append = FALSE)

Deleting layer `kc_tracts10' using driver `GPKG'
Writing layer `kc_tracts10' to data source
`../maps/kc_tracts10.gpkg' using driver `GPKG'
Writing 398 features with 22 fields and geometry type Polygon.

```

```
st_write(kc_tracts10_shore, ".../maps/kc_tracts10_shore.gpkg", append = FALSE)
```

```
Deleting layer `kc_tracts10_shore' using driver `GPKG'  
Writing layer `kc_tracts10_shore' to data source  
`.../maps/kc_tracts10_shore.gpkg' using driver `GPKG'  
Writing 398 features with 22 fields and geometry type Multi Polygon.
```

```
st_write(kc_houses_env_var, ".../maps/kc_houses_env_var.gpkg", append = FALSE)
```

```
Deleting layer `kc_houses_env_var' using driver `GPKG'  
Writing layer `kc_houses_env_var' to data source  
`.../maps/kc_houses_env_var.gpkg' using driver `GPKG'  
Writing 21436 features with 65 fields and geometry type Point.
```

```
st_write(kc_tracts10_shore_env_var, ".../maps/kc_tracts10_shore_env_var.gpkg", append = FALSE)
```

```
Deleting layer `kc_tracts10_shore_env_var' using driver `GPKG'  
Writing layer `kc_tracts10_shore_env_var' to data source  
`.../maps/kc_tracts10_shore_env_var.gpkg' using driver `GPKG'  
Writing 21436 features with 65 fields and geometry type Point.
```

## Oppgave 4

i.

```
#summary(kc_tracts10_env_data) <- La til hashtag for å unngå render-problemer
```

```
#summary(kc_tracts10_shore_env_var) <- La til hashtag for å unngå render-problemer
```

ii.

I QGIS fant vi følgende observasjoner ved å se på *tracts10*, *tracts10\_shore* & *kc\_houses\_env\_var*:

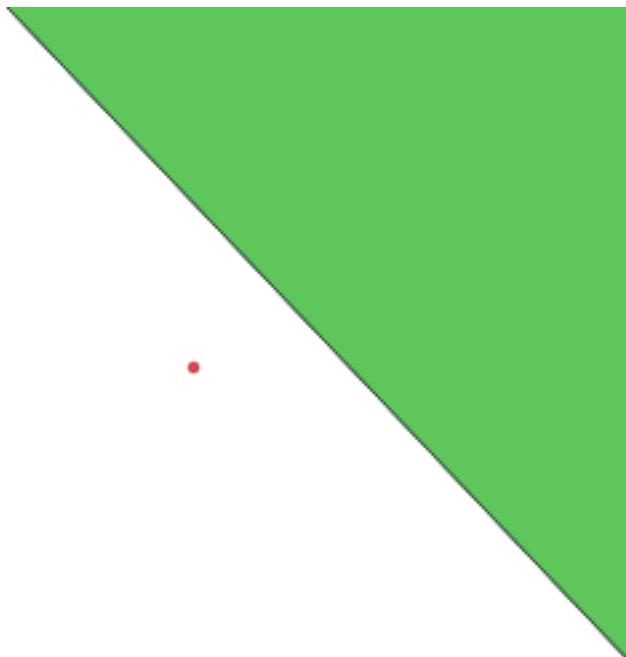


Figure 1: observasjon utenfor WA state

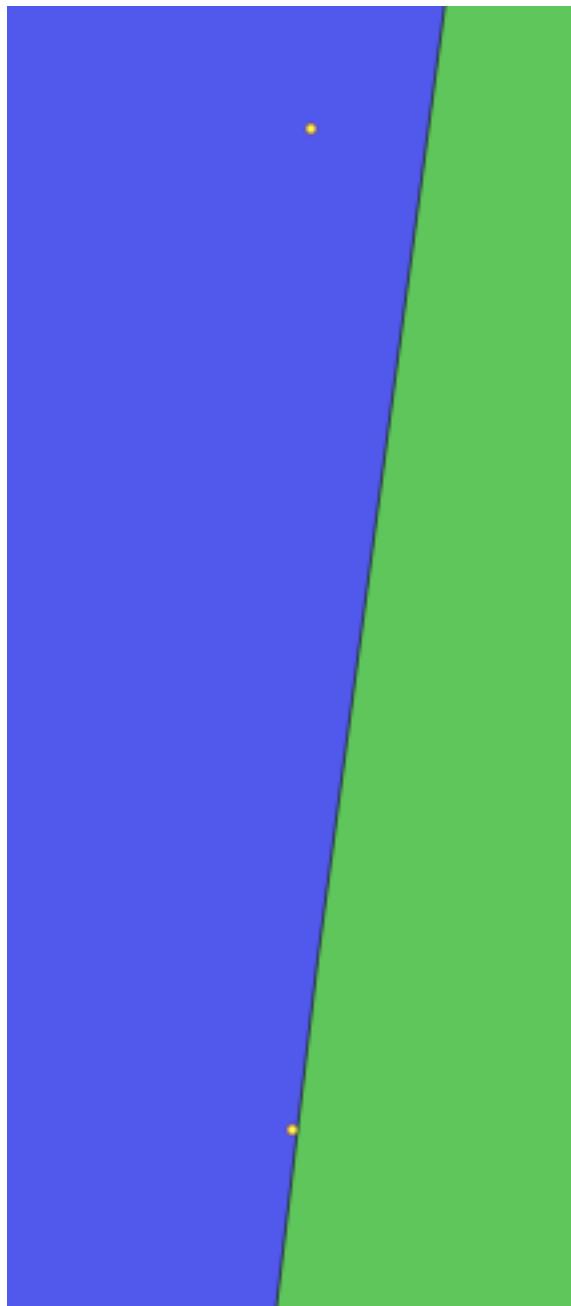


Figure 2: Observasjon utenfor kystlinjen.a



Figure 3: Observasjon utenfor kystlinjen.b

iii.

Dropper Observasjonen 3518000180 ved å:

```
kc_houses_env_var <- arrange(kc_houses_env_var, desc(id))
kc_houses_env_var.omit <- kc_houses_env_var[-c(11997),]

st_write(kc_houses_env_var.omit, ".../maps/kc_houses_env_var.omit.gpkg", append = FALSE)

Deleting layer `kc_houses_env_var.omit' using driver `GPKG'
Writing layer `kc_houses_env_var.omit' to data source
`.../maps/kc_houses_env_var.omit.gpkg' using driver `GPKG'
Writing 21435 features with 65 fields and geometry type Point.

kc_houses_env_var.omit <- kc_houses_env_var.omit %>%
  mutate(
    year_month = substr(date, start = 1, stop = 7))
```

## Oppgave 5

### References

Bishop, Kelly C., Nicolai V. Kuminoff, H. Spencer Banzhaf, Kevin J. Boyle, Kathrine von Gravenitz, Jaren C. Pope, V. Kerry Smith, og Christopher D. Timmins. 2020. «Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality». *Review of Environmental Economics and Policy* 14 (2): 260–81. <https://doi.org/10.1093/reep/reaa001>.

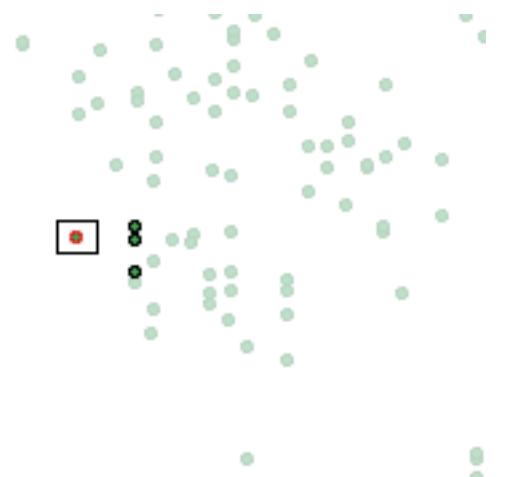


Figure 4: K-nearest 3

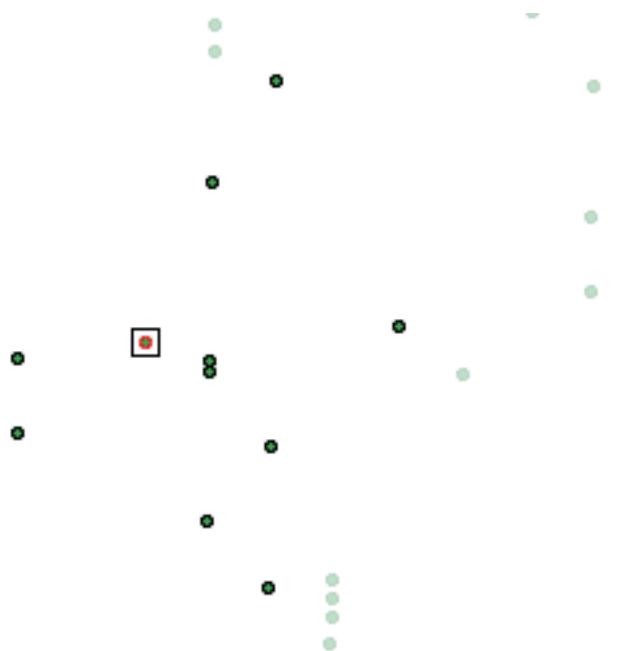


Figure 5: K-nearest 10

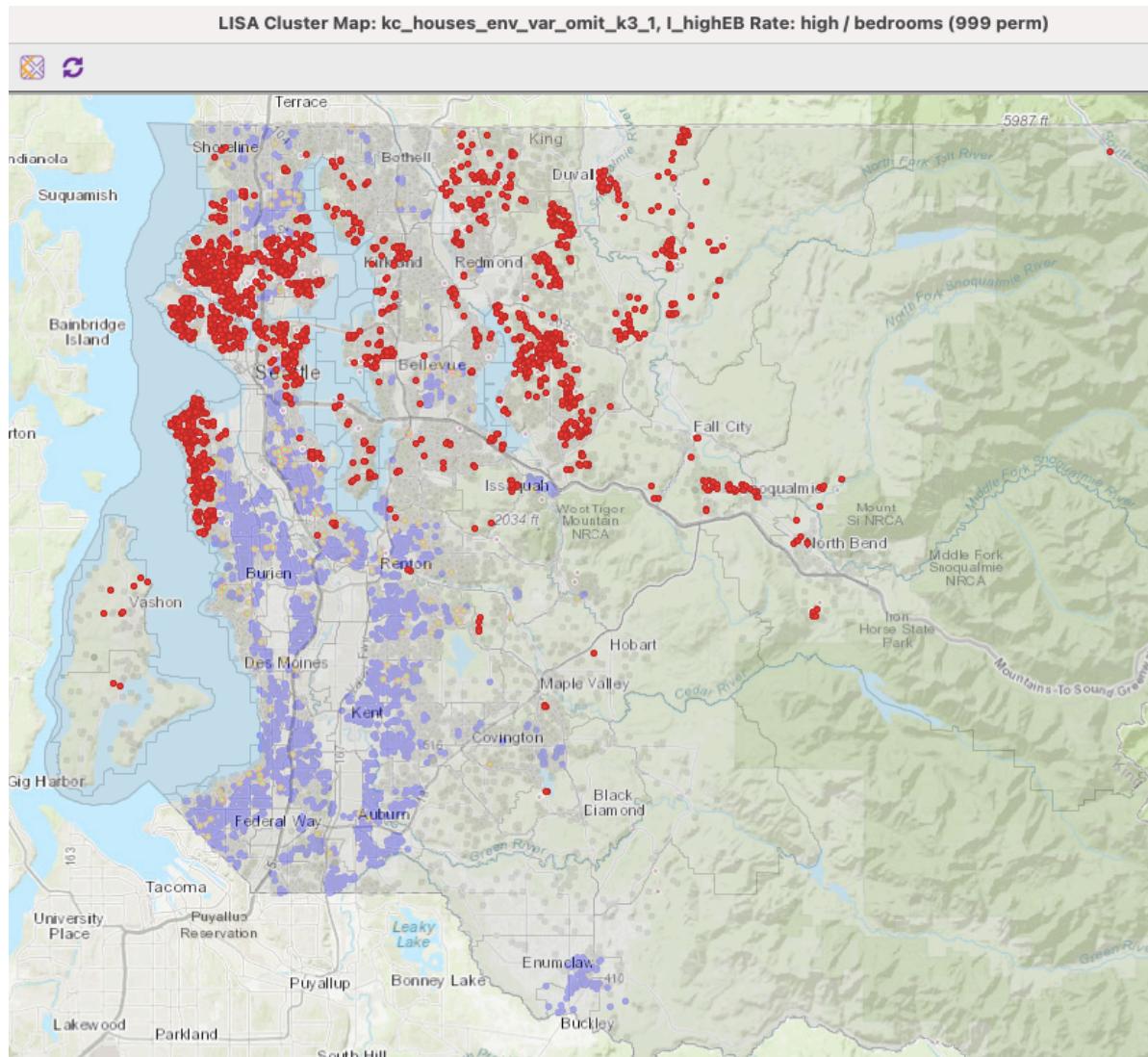


Figure 6: K3 - Store og dyre boliger

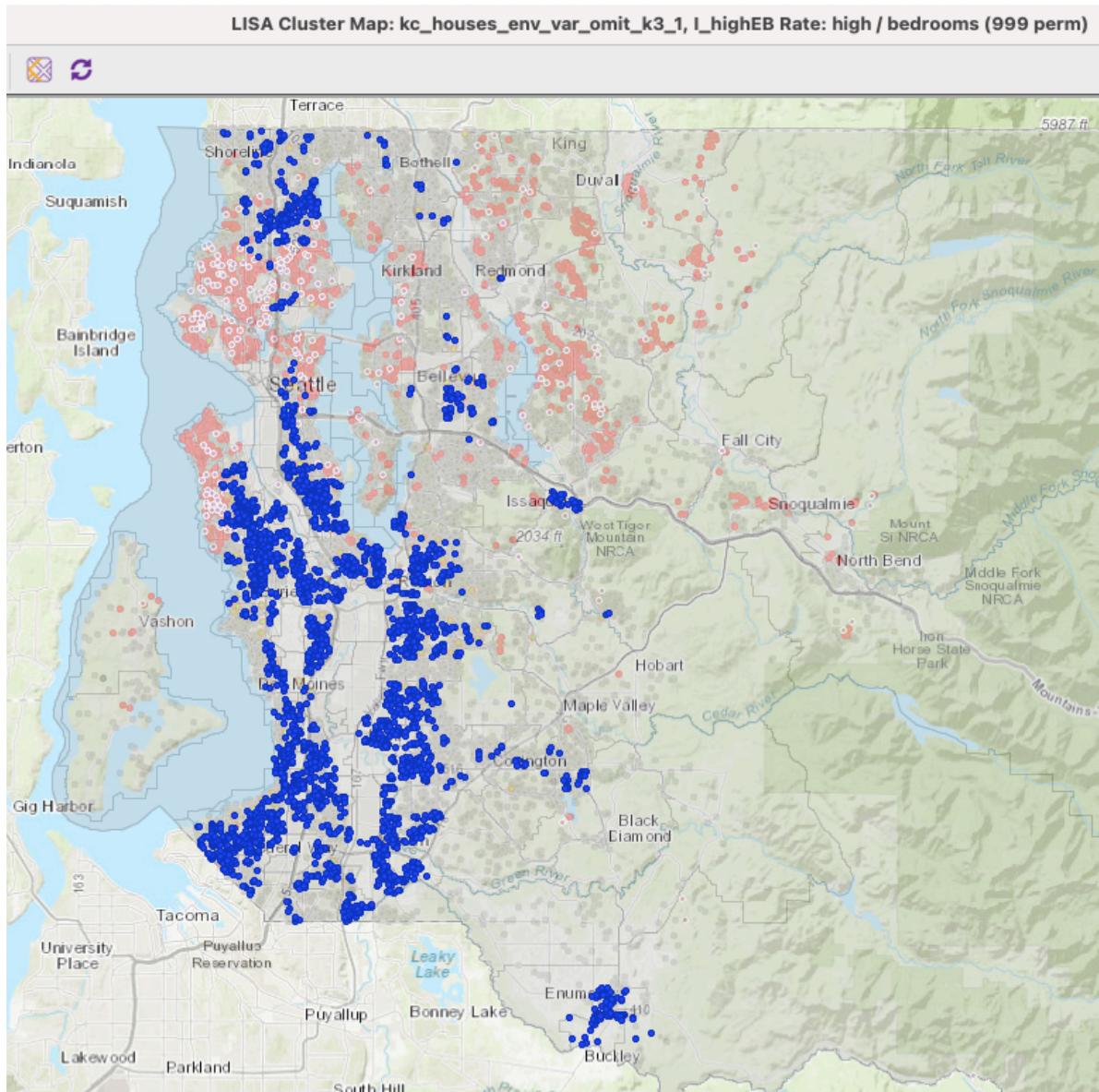


Figure 7: K3 - Små og billige boliger

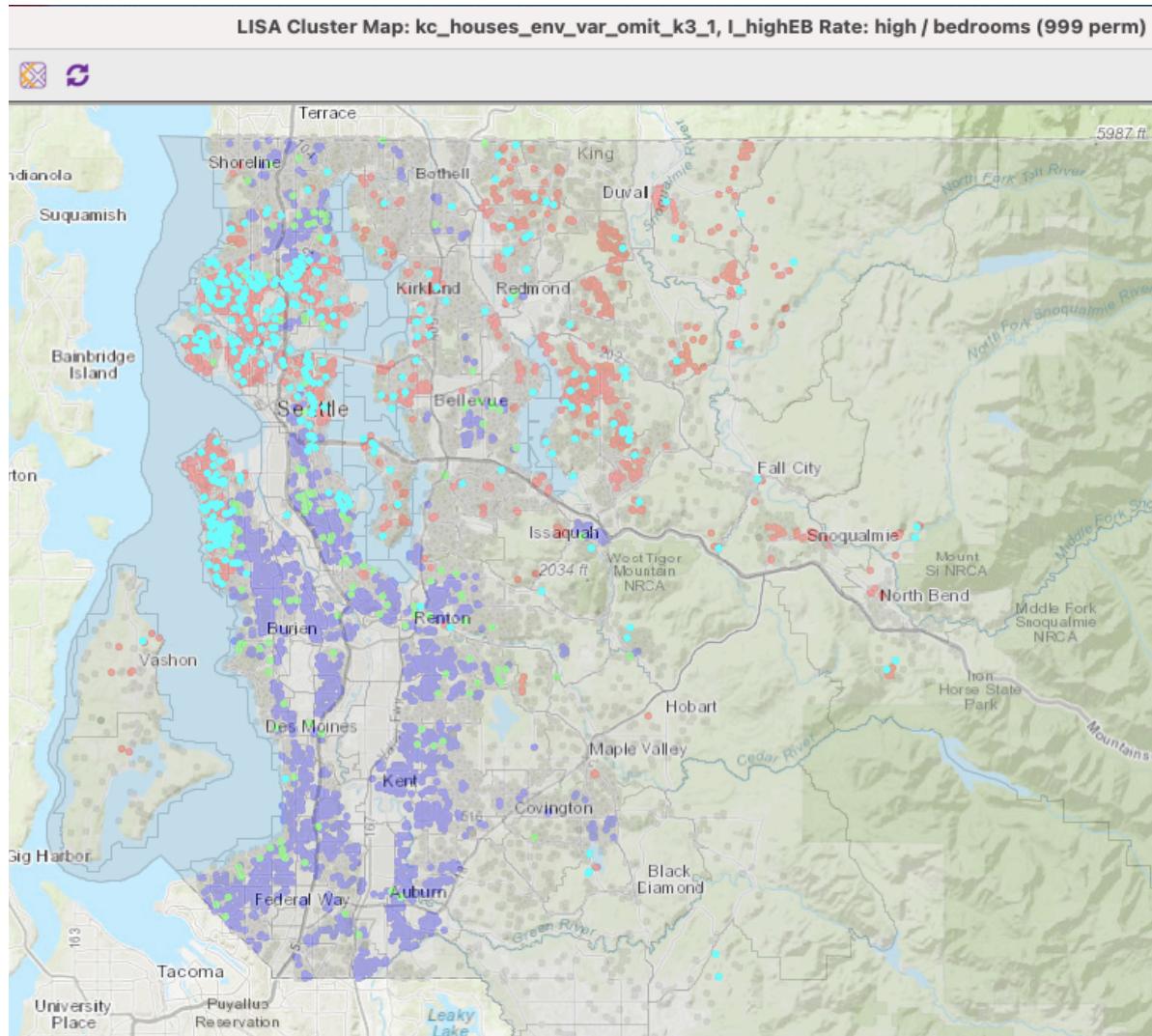


Figure 8: K3 - Små og dyre boliger

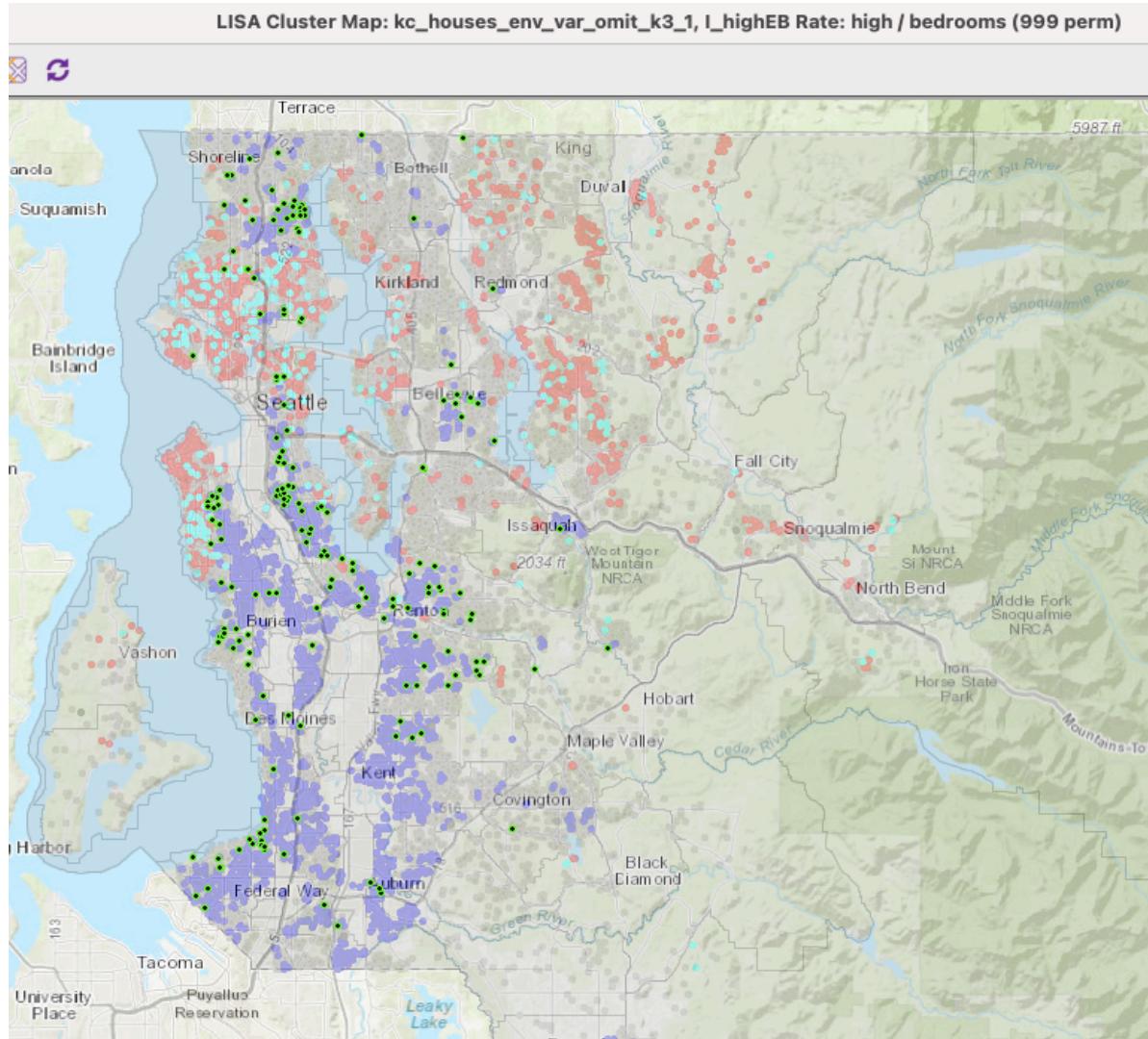


Figure 9: K3 - Store og billige boliger

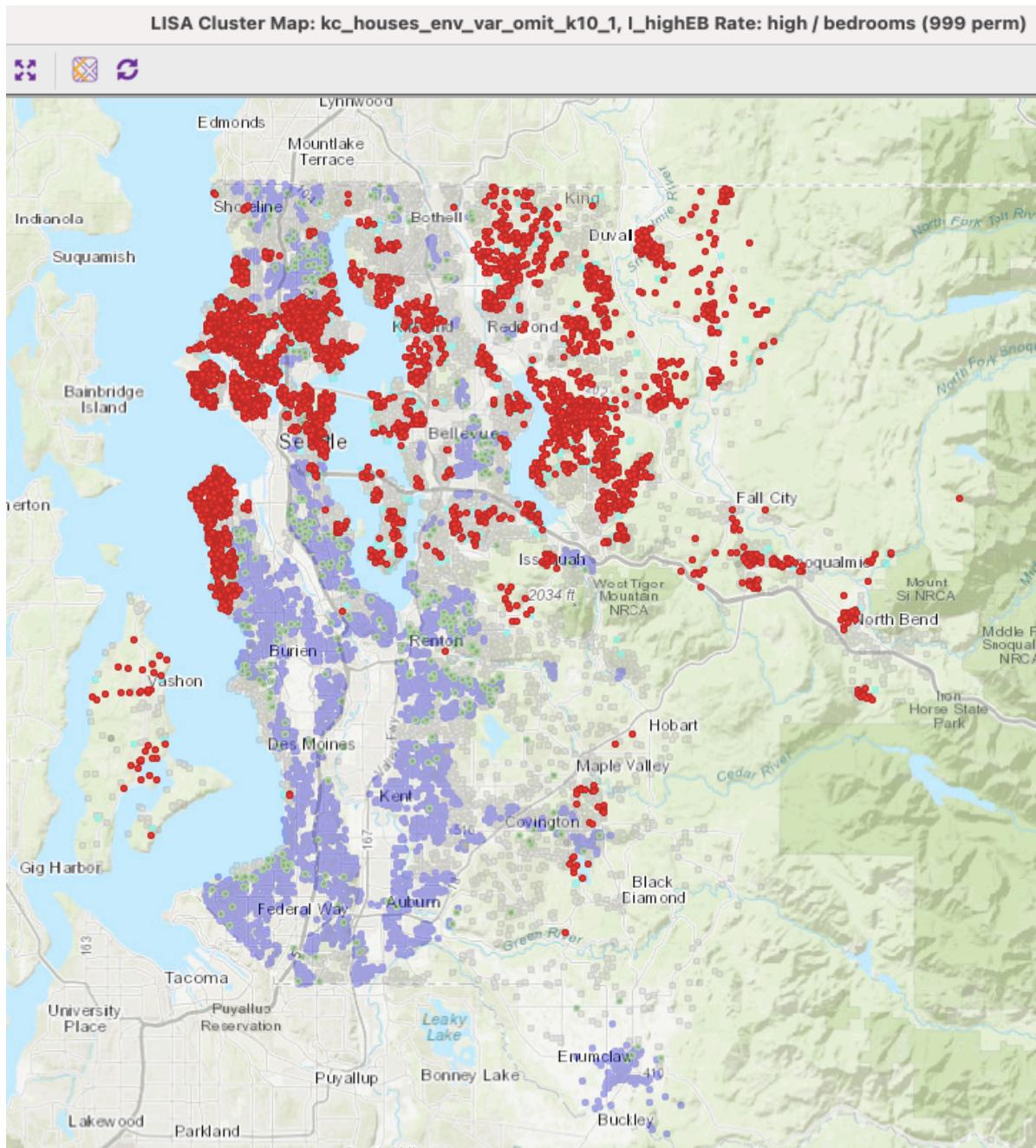


Figure 10: K10 - Store og dyre boliger

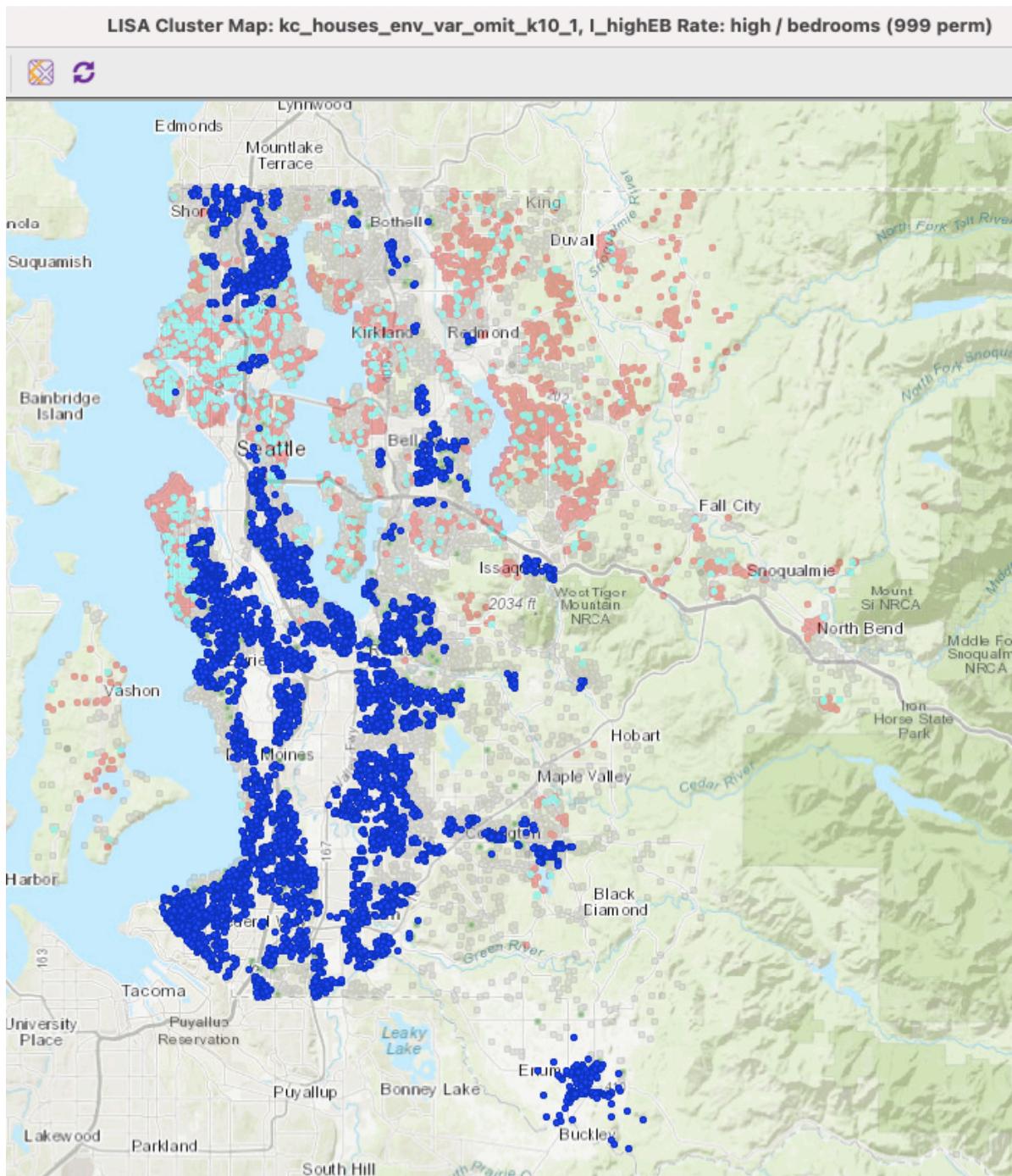


Figure 11: K10 - Små og billige boliger

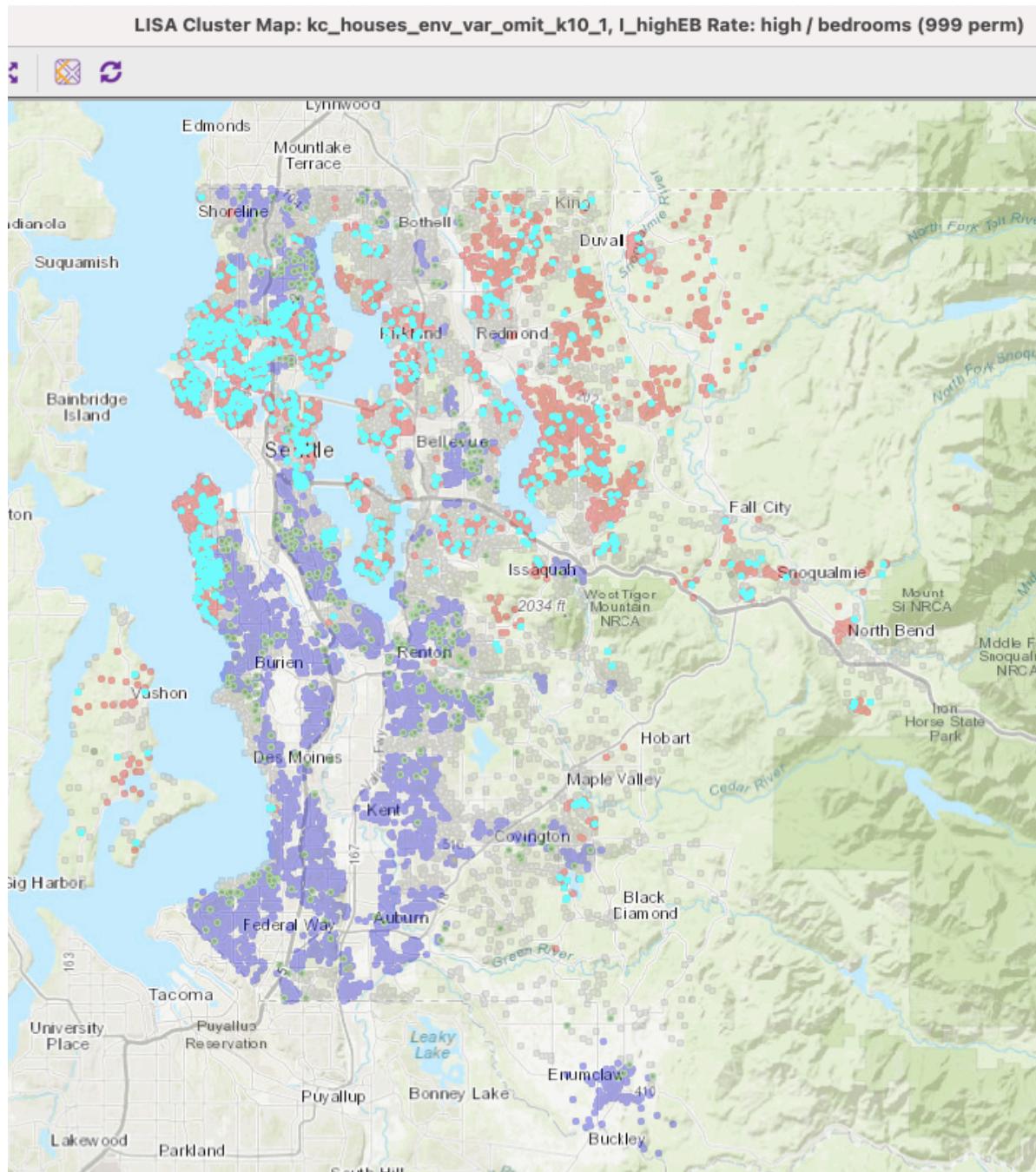


Figure 12: K10 - Små og dyre boliger

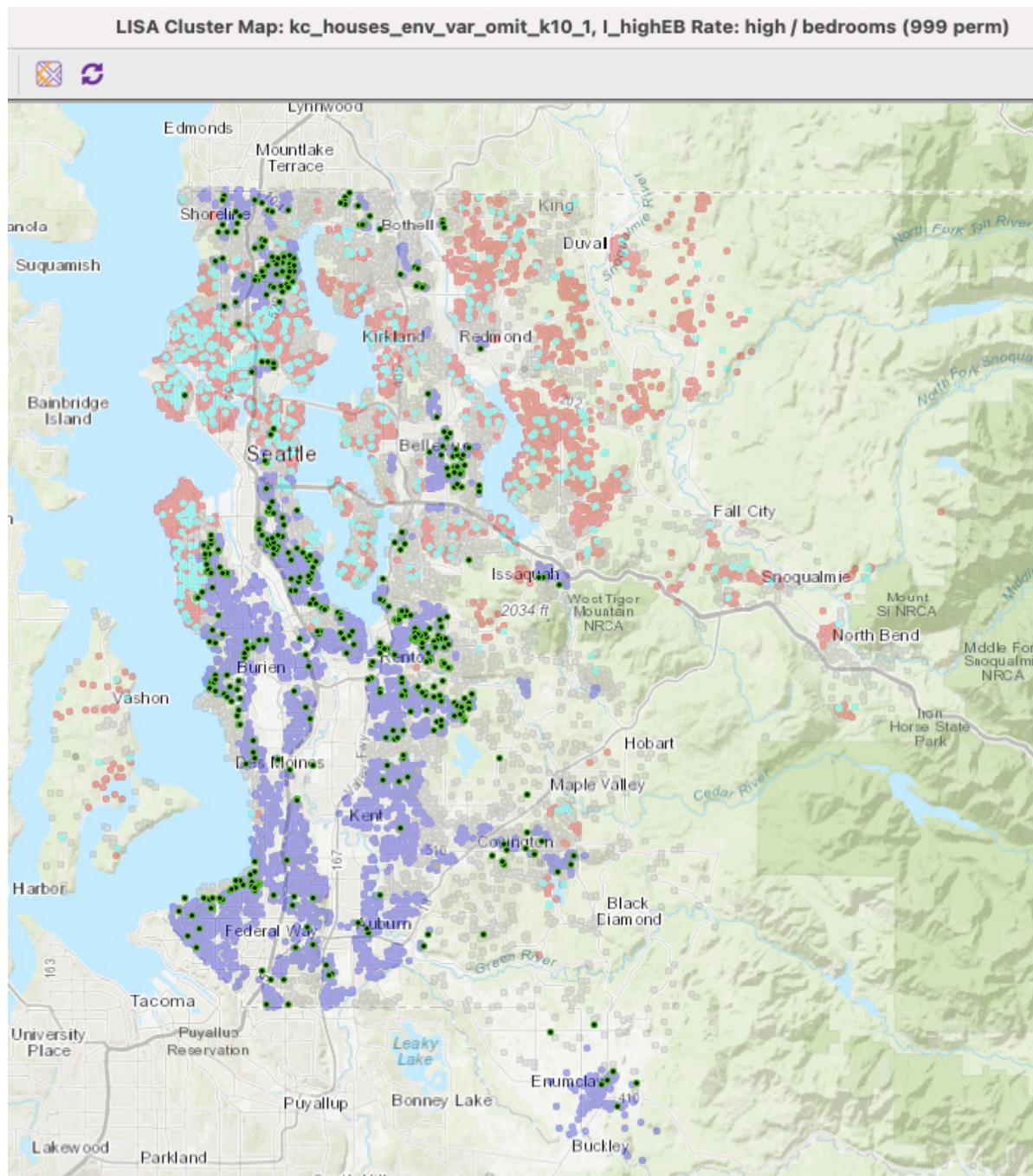


Figure 13: K10 - Store og billige boliger