

# **Termpaper MSB205**

## **Arbeidskrav (kan finne en annen tittel)**

### **Oppg. 1**

Bishop et al. (2020) skriver at den mest direkte illustrasjonen på hvordan private markeder kan avsløre forbrukernes betalingsvillighet er den hedoniske eiendomsverdi modellen. Modellen går ut på at en ser for seg at kjøpere velger eiendommer basert på egenskapene til boligen, som kan for eksempel være antall soverom og bad. Den er også basert på stedsspesifikke fasiliteter for en bolig som for eksempel er luftkvaliteten, nærhet til sentrum/parker og flomrisiko.

I løpet av de siste årene har forventninger til kvaliteten på dataen og den økonometrisk åpenhet økt. Det har også forbedret forståelse om hvordan en identifisert gjennom kvasi-eksperimentelle forskningsdesign forholder seg til velferds mål, som vil si mål på betalingsvillighet. Det ble brukt rike data fra boligmarkedet i storbyområdene i en avansert økonomi i de fleste studiene, som etablerte den beste praksisen i den hedoniske modellen. Denne dataen beskriver boligtransaksjoner, egenskaper og fasiliteter som stadig blir mer tilgjengelig rundt om i verden, som gir muligheten for å bruke den hedoniske modellen.

Den deriverte prisfunksjonen kan tolkes som å indikere tilbudets implisitte priser, denne kan brukes til å beregne husholdningens marginale betalingsvilligheten for tilbuddet. Den marginale betalingsvilligheten bidrar med til å informere om politikk. Den moderne hedoniske eiendomsverdimodellen er den fremste tilnærmingene til å verdsette endringer i miljøfasiliteter, da innen områder som akademisk forskning, rettssaker og offentlig politikk.

Roses (1974) hadde en banebrytende artikkel om det hedoniske rammeverket. Der han delte det inn i to trinn, første trinn handler om definere markedet og andre trinn om å samle inn data. Modellen blir sett på som en likevektsmodell for å kunne forstå hva differensierte produktpriser kunne avsløre om forbrukernes etterspørsel etter produktattributter. Likevekten er forholdet mellom boligpriser og huskarakteristikk, som igjen avslører hver kjøpers marginale betalingsvillighet. Den marginale betalingsvilligheten til kjøpere kan endres over tid, som kan gjenspeiles som en endring i funksjonen for de implisitte prisene for fasilitetene. Grunnen til at den marginale betalingsvilligheten kan endres over tid kan skyldes at en øker arbeidernes

produktivitet, induserer migrasjon, gir ny informasjon om fasilitetene eller endrer bekjemmeligheters nivåer, dette kan være for eksempel endringer i regler i luftkvalitet.

Dette gir et grunnlag for å kunne bruke den hedoniske eiendomsverdimodellen, og en bruker den hedoniske modellen til å estimere marginale betalingsvilligheter for miljøfasiliteter.

Første trinnet i modellen er å definere et marked. Et marked bør bli definert slik at det tilfredsstiller “loven om én prisfunksjon.” Det betyr at identiske hus bør bli solgt for samme pris i et gitt marked. En kan definere marked som et stort område over kort tid, men kan også være et større område over lengre tid. For å følge opp “loven om én prisfunksjon” gjelder det å defineret markedet som et stort område over kort tid. Da unngår en ta stilling til flytting, fordi de fysiske og økonomiske kostnadene ved å flytte ikke viser seg til å endre på ulike destinasjoner innenfor et stort område. I motsetning til dette er det mindre sannsynlig at loven om én prisfunksjon blir tilfredsstilt hvis markedet er definert til å omfatte flere storbyområder og/eller flere år. Om en skulle definert et marked som et større område over lengre tid er at det vil føre til større flyttekostnader, fordi det omfatter flere storbyområder. Dette ville påvirket den marginale betalingsvilligheten til flere, fordi en kan bli påtvunget flytting på grunn av for eksempel jobb, kan også være endringer i lokale skattekostnader, som gjør at en velger å flytte.

I trinn to handler det om å samle inn data. Et tilfeldig utvalg er det beste når det kommer til datainnsamling i hedoniske eiendomsverdistudier, da av boligtransaksjonspriser og egenskaper for det aktuelle studieområdet og hovedsakelig for enebolig. Ved innsamling av data kan det oppstå utfordringer som kan oppstå i mindre enn ideelle datainnstillingar, inkludert regulering av priser, sparsomme transaksjoner og mangel på transaksjonspriser.

Når en skal samle inn data om boligsalg har en forventning på at informasjonen ligge offentlig tilgjengelig, her kan det oppstå problemer som gjør at en ikke fanger opp dette. Ved å identifisere og ta vekk disse dataene kan en redusere muligheten for å få målefeil. Dette kan være for eksempel at kjøper og selger har likt etternavn i et salg, fordi sannsynligheten for at de er slekt er stor. Er også vanlig å fjerne tvangssalg, og kjøp fra eiendomsinvesteringsforetak og uteliggere som tydelig indikerer datainntastingsfeil.

Det er også viktig å inkludere hvordan kjøpere oppfatter fasilitetsnivåene på hvert boligsted. Der forskerne må inkludere romlig interpolasjon, luftspredningsmodeller eller spådommer fra satellitter for å tildele forurensningsnivåer til hus. En kan også se på nærhet til rekreasjonssteder som strender, innsjøer og parker måles etter geografisk avstand, kjøreavstand, total reisetid eller andelen land som er viet til denne rekreasjonsbruken innenfor et geografisk område rundt et hus, det er for å se på hva som har en betydning for kjøperen.

Når dataen om salgspriser og kjennetegn ved en enemoligtransaskjon ikke er tilgjengelig, kan en bruke data om anslitte priser, leiepriser og salg av barmark, samt romlig aggregerte oppsummeringsmål som middel eller medianer. Dette kan gi utfordringer for å kunne tolke prisfunksjonsparametere som mål på den marginale betalingsvilligheten. Det kan også være ideelt å bruke spørreundersøkelser for å kunne estimere en verdi. Det er også mulig å bruke eiendomsvurderinger eller andre selskaper. Transaksjonpriser er foretrukket fremfor en anslått pris, dette

fordi predikerte priser er at de inkluderer målefeil, som korrelerer med kjøperens demografi boligkarakteristikker og nabolagsfasiliteter, og fører dermed til skjevheter i prisfunksjonens parameterestimater. Under leiepriser kan det oppstå uklarheter fordi det er opp til den som leier ut hva som er viktig, da om leieren betaler for vedlikehold og utstyr. Ofte leier en i kortere periode, og derfor prioriterer ikke fasiliteter eller nabobelagget.

For å velge en økonometrisk spesifikasjon for den hedoniske prisfunksjonen, bør en først se på at prisfunksjonen antas å være ikke-lineær. Dette er fordi dette gir en mer nøyaktige estimer av gjennomsnittlig marginal betalingsvillighet for boligkarakteristikker, enn om det skulle vært lineære og lig-lineær. Det tillater også at markedslikevekt gjenspeiler komplementaritet mellom fasiliteter. Med tanke på kriminalitet, støy og luftkvalitet. Det er også viktig å se på at ingen informasjon er utelatt, som da er vitkig for huskjøperen. Da med tanke på skoler, nabolag og natur. Det viktigste med denne modellen er å se på hva kjøperens betalingsvilligheten er for miljøfasiliteter

## Oppg. 2

I

Lastet først ned datasettet House Sales in King County, USA fra Kaggle. Deretter sjekket vi definisjonene om at de var riktige.

II

Leser inn hus salgene i King County i USA som vi har lastet ned fra Kaggle.

```
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

Nå sorterer vi salgene etter dato.

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

### IV

I denne omgang bruker vi dplyr:: distinct til å velge siste salg der vi multiple salg av samme eiendom.

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

### V og VI

Bruker st\_as\_sf() til å konvertere house data til et sf objekt vha. long lat og setter til geografisk projekson.

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

### VII

Koordinater Seattle som er hentet fra Wikipedia er : 47.3622, -122,1955

```
cbd <- st_sfc(st_point(c(-122.1955, 47.3622)), crs = 4326) %>%
  st_transform(2926)
```

### VIII

Her finner vi avstanden mellom punktet EPSG:2926 og samtlige hus i datasettet i luftlinje. Deretter konverterer vi det til km og legger dem inn i variabelen dest\_CBD.

```
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)
)
```

## Oppg. 3

### I og III

Leser inn filen WADOH King County.

```
kc_wadoh_map <- here("WADOH_Environmental_Health_Disparities_Index_Calculated_for_King_Cou  
st_read() %>%  
st_transform(2926)
```

```
Reading layer `WADOH_Environmental_Health_Disparities_Index_Calculated_for_King_County___wado  
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
```

### II

Plukker ut variablene som er angitt i oppgaveteksten.

```
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
)

```

## IV

Laster inn acs\_b19101\_familyincome.

```
acs_b19101_fam_inc <- read.dbf("../Maps/censusSHP/acs_b19101_familyincome.dbf")
```

Lager inntektsvariabelen.

```

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/kinemakestad/Documents/Master i sivilokonom/3. Semester/Boligmarked og spatial økonomi
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/kinemakestad/Documents/Master i sivilokonom/3. Semester/Boligmarked og spatial økonomi
  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)

```

Nå blir det brukt left join for å legge dataene inn i WADOH King County.

```

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")

summary(kc_tracts10)

  GEO_ID_TRT           FEATURE_ID        TRACT_LBL        TRACT_STR
Length:398       Min.    :10153  Length:398       Length:398
Class :character  1st Qu.:25818  Class :character  Class :character
Mode  :character  Median :44344   Mode  :character  Mode  :character
                           Mean   :36731
                           3rd Qu.:45226
                           Max.   :45837
  TRACT_INT          TRACT_FLT        TRACT_DEL        TRTLABEL_F
Min.    : 100   Min.    : 1.00  Length:398       Length:398
1st Qu.: 9625  1st Qu.: 96.25  Class :character  Class :character
Median : 24150  Median : 241.50  Mode  :character  Mode  :character
Mean   : 23022  Mean   : 230.22
3rd Qu.: 30076  3rd Qu.: 300.76
Max.   :990100  Max.   :9901.00
  TRTLABEL_C          TRTLABEL_T        COUNTY_STR        COUNTY_INT
Length:398       Length:398       Length:398       Min.    :33
Class :character  Class :character  Class :character  1st Qu.:33
Mode  :character  Mode  :character  Mode  :character  Median :33
                                         Mean   :33

```

				3rd Qu.:33
				Max. :33
STATE_STR	STATE_INT	LEVEL_1	LEVEL_2	
Length:398	Min. :53	Length:398	Length:398	
Class :character	1st Qu.:53	Class :character	Class :character	
Mode :character	Median :53	Mode :character	Mode :character	
	Mean :53			
	3rd Qu.:53			
	Max. :53			
LEVEL_3	TRACT_AREA	TRACT_PERI	LOGRECNO	
Length:398	Min. :2.466e+06	Min. : 7060	Length:398	
Class :character	1st Qu.:1.933e+07	1st Qu.: 20586	Class :character	
Mode :character	Median :3.362e+07	Median : 29573	Mode :character	
	Mean :1.616e+08	Mean : 44019		
	3rd Qu.:5.601e+07	3rd Qu.: 43667		
	Max. :1.526e+10	Max. :738820		
Shape_area	Shape_len	geometry		
Min. :2.466e+06	Min. : 7060	POLYGON :398		
1st Qu.:1.933e+07	1st Qu.: 20586	epsg:2926 : 0		
Median :3.362e+07	Median : 29573	+proj=lcc ...: 0		
Mean :1.616e+08	Mean : 44019			
3rd Qu.:5.601e+07	3rd Qu.: 43667			
Max. :1.526e+10	Max. :738820			

```

kc_house_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_house_env_var, ".../Maps/kc_house_env_var.gpkg", append = FALSE)
```

```
Deleting layer `kc_house_env_var' using driver `GPKG'  
Writing layer `kc_house_env_var' to data source  
`.../Maps/kc_house_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.
```

## Oppg. 4

|

Her sjekker vi områdevariablene fra WADOH ved hjelp av summary for både tracts10 og tracts10 shore.

```
summary(kc_tracts10_env_data)
```

GEO_ID_TRT	FEATURE_ID	TRACT_LBL	TRACT_STR
Length:398	Min. :10153	Length:398	Length:398
Class :character	1st Qu.:25818	Class :character	Class :character
Mode :character	Median :44344	Mode :character	Mode :character
	Mean :36731		
	3rd Qu.:45226		
	Max. :45837		

TRACT_INT	TRACT_FLT	TRACT_DEL	TRTLABEL_F
Min. : 100	Min. : 1.00	Length:398	Length:398
1st Qu.: 9625	1st Qu.: 96.25	Class :character	Class :character
Median : 24150	Median : 241.50	Mode :character	Mode :character
Mean : 23022	Mean : 230.22		
3rd Qu.: 30076	3rd Qu.: 300.76		
Max. :990100	Max. :9901.00		
TRTLABEL_C	TRTLABEL_T	COUNTY_STR	COUNTY_INT
Length:398	Length:398	Length:398	Min. :33
Class :character	Class :character	Class :character	1st Qu.:33
Mode :character	Mode :character	Mode :character	Median :33
			Mean :33
			3rd Qu.:33
			Max. :33
STATE_STR	STATE_INT	LEVEL_1	LEVEL_2
Length:398	Min. :53	Length:398	Length:398
Class :character	1st Qu.:53	Class :character	Class :character
Mode :character	Median :53	Mode :character	Mode :character
	Mean :53		
	3rd Qu.:53		
	Max. :53		
LEVEL_3	TRACT_AREA	TRACT_PERI	LOGRECNO
Length:398	Min. :2.466e+06	Min. : 7060	Length:398
Class :character	1st Qu.:1.933e+07	1st Qu.: 20586	Class :character
Mode :character	Median :3.362e+07	Median : 29573	Mode :character
	Mean :1.616e+08	Mean : 44019	
	3rd Qu.:5.601e+07	3rd Qu.: 43667	
	Max. :1.526e+10	Max. :738820	
Shape_area	Shape_len	low	mid
Min. :2.466e+06	Min. : 7060	Min. :0.009298	Min. :0.0000
1st Qu.:1.933e+07	1st Qu.: 20586	1st Qu.:0.053302	1st Qu.:0.2391
Median :3.362e+07	Median : 29573	Median :0.092424	Median :0.3339
Mean :1.616e+08	Mean : 44019	Mean :0.125013	Mean :0.3327
3rd Qu.:5.601e+07	3rd Qu.: 43667	3rd Qu.:0.166534	3rd Qu.:0.4261
Max. :1.526e+10	Max. :738820	Max. :1.000000	Max. :0.6790
		NA's :1	NA's :1
high	EHD_percen	linguist_2	poverty_pe
Min. :0.0000	Min. : 1.00	Min. : 0.45	Min. : 1.97

1st Qu.:0.4006	1st Qu.: 25.00	1st Qu.: 3.88	1st Qu.:10.53
Median :0.5637	Median : 50.00	Median : 8.72	Median :16.75
Mean :0.5423	Mean : 50.38	Mean :10.62	Mean :20.42
3rd Qu.:0.6955	3rd Qu.: 75.00	3rd Qu.:15.38	3rd Qu.:27.48
Max. :0.8816	Max. :100.00	Max. :46.76	Max. :75.48
NA's :1	NA's :1	NA's :5	NA's :1
POC_percen	transporta	unemploy_2	housing_pe
Min. : 7.54	Min. :10.00	Min. : 1.000	Min. :15.14
1st Qu.:23.36	1st Qu.:18.00	1st Qu.: 3.350	1st Qu.:27.34
Median :36.29	Median :19.00	Median : 4.480	Median :32.26
Mean :38.64	Mean :18.97	Mean : 5.099	Mean :33.75
3rd Qu.:51.01	3rd Qu.:21.00	3rd Qu.: 6.460	3rd Qu.:39.13
Max. :92.70	Max. :26.00	Max. :24.400	Max. :81.89
NA's :1	NA's :1	NA's :3	NA's :1
traffic_pe	diesel	ozone	PM25
Min. : 0.00	Min. : 0.14	Min. :46.73	Min. :3.787
1st Qu.: 0.00	1st Qu.: 6.65	1st Qu.:48.91	1st Qu.:5.642
Median : 3.60	Median :12.65	Median :49.78	Median :6.180
Mean :16.07	Mean :17.10	Mean :50.62	Mean :6.186
3rd Qu.:26.17	3rd Qu.:18.99	3rd Qu.:51.28	3rd Qu.:6.872
Max. :97.75	Max. :92.63	Max. :62.89	Max. :7.897
NA's :1	NA's :1	NA's :1	NA's :1
toxic_rele	hazardous_	lead_perce	superfund
Min. : 823.9	Min. :0.02303	Min. : 0.24	Min. :0.03454
1st Qu.: 5180.9	1st Qu.:0.04168	1st Qu.: 6.46	1st Qu.:0.07358
Median : 10186.5	Median :0.05160	Median :13.79	Median :0.13133
Mean : 19398.3	Mean :0.08190	Mean :17.08	Mean :0.24645
3rd Qu.: 20058.1	3rd Qu.:0.09280	3rd Qu.:26.20	3rd Qu.:0.28436
Max. :186434.6	Max. :0.63781	Max. :54.68	Max. :1.46778
NA's :1	NA's :1	NA's :1	NA's :1
facilities	wastewater	sen_pop_pe	socio_perc
Min. :0.0523	Min. :0.00e+00	Min. : 1.00	Min. : 1.00
1st Qu.:0.1612	1st Qu.:5.50e-06	1st Qu.: 25.00	1st Qu.: 25.00
Median :0.3652	Median :5.30e-04	Median : 50.00	Median : 50.00
Mean :0.6046	Mean :2.62e-02	Mean : 50.38	Mean : 50.38
3rd Qu.:0.9119	3rd Qu.:8.70e-03	3rd Qu.: 75.00	3rd Qu.: 75.00
Max. :3.3682	Max. :6.40e-01	Max. :100.00	Max. :100.00
NA's :1	NA's :1	NA's :1	NA's :1
geometry			
POLYGON :398			
epsg:2926 : 0			
+proj=lcc ...: 0			

```
summary(kc_tracts10_shore_env_data)
```

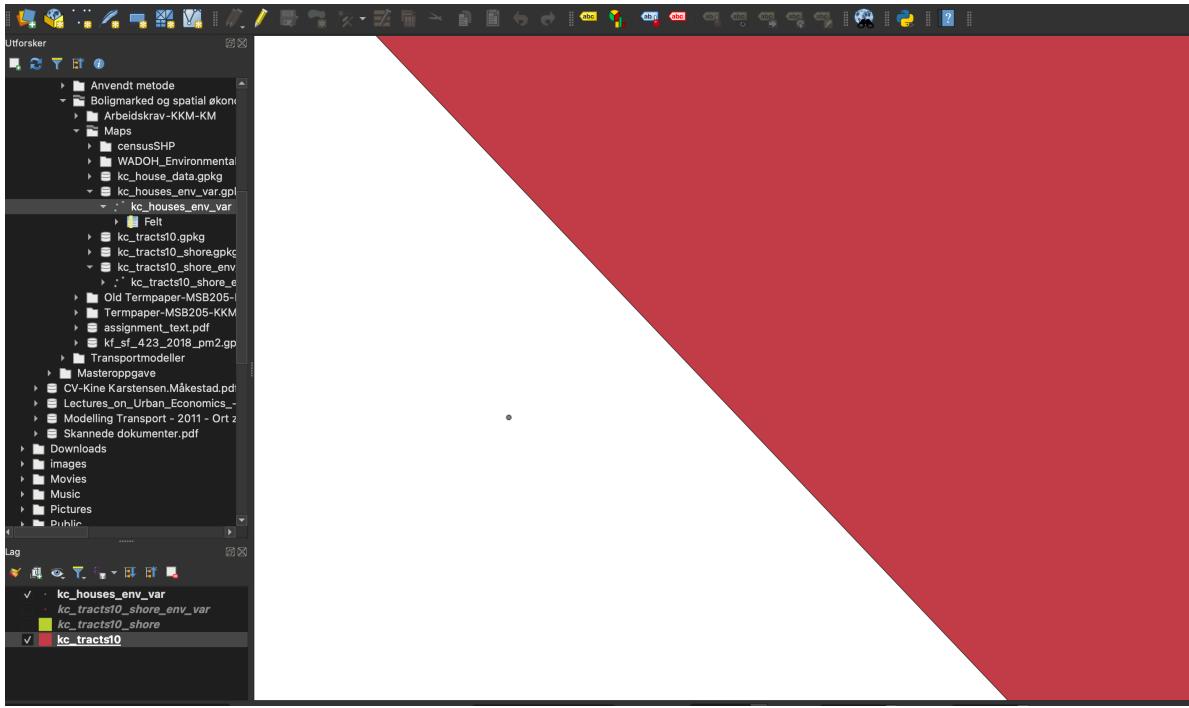
GEO_ID_TRT	FEATURE_ID	TRACT_LBL	TRACT_STR
Length:398	Min. :10153	Length:398	Length:398
Class :character	1st Qu.:27069	Class :character	Class :character
Mode :character	Median :44458	Mode :character	Mode :character
	Mean :36834		
	3rd Qu.:45197		
	Max. :45838		
TRACT_INT	TRACT_FLT	TRACT_DEL	TRTLABEL_F
Min. : 100	Min. : 1.00	Length:398	Length:398
1st Qu.: 9625	1st Qu.: 96.25	Class :character	Class :character
Median : 24150	Median : 241.50	Mode :character	Mode :character
Mean : 23022	Mean : 230.22		
3rd Qu.: 30076	3rd Qu.: 300.76		
Max. :990100	Max. :9901.00		
TRTLABEL_C	TRTLABEL_T	COUNTY_STR	COUNTY_INT
Length:398	Length:398	Length:398	Min. :33
Class :character	Class :character	Class :character	1st Qu.:33
Mode :character	Mode :character	Mode :character	Median :33
			Mean :33
			3rd Qu.:33
			Max. :33
STATE_STR	STATE_INT	LEVEL_1	LEVEL_2
Length:398	Min. :53	Length:398	Length:398
Class :character	1st Qu.:53	Class :character	Class :character
Mode :character	Median :53	Mode :character	Mode :character
	Mean :53		
	3rd Qu.:53		
	Max. :53		
LEVEL_3	TRACT_AREA	TRACT_PERI	LOGRECNO
Length:398	Min. :2.466e+06	Min. : 7060	Length:398
Class :character	1st Qu.:1.933e+07	1st Qu.: 20586	Class :character

Mode :character	Median :3.362e+07	Median : 29573	Mode :character
	Mean :1.616e+08	Mean : 44023	
	3rd Qu.:5.601e+07	3rd Qu.: 43919	
	Max. :1.526e+10	Max. :738820	
 Shape_area	 Shape_len	 low	 mid
Min. :7.819e+05	Min. : 7060	Min. :0.009298	Min. :0.0000
1st Qu.:1.794e+07	1st Qu.: 20297	1st Qu.:0.053302	1st Qu.:0.2391
Median :2.964e+07	Median : 28874	Median :0.092424	Median :0.3339
Mean :1.504e+08	Mean : 41303	Mean :0.125013	Mean :0.3327
3rd Qu.:5.020e+07	3rd Qu.: 40590	3rd Qu.:0.166534	3rd Qu.:0.4261
Max. :1.526e+10	Max. :738820	Max. :1.000000	Max. :0.6790
		NA's :1	NA's :1
 high	 EHD_percen	 linguist_2	 poverty_pe
Min. :0.0000	Min. : 1.00	Min. : 0.45	Min. : 1.97
1st Qu.:0.4006	1st Qu.: 25.00	1st Qu.: 3.88	1st Qu.:10.53
Median :0.5637	Median : 50.00	Median : 8.72	Median :16.75
Mean :0.5423	Mean : 50.38	Mean :10.62	Mean :20.42
3rd Qu.:0.6955	3rd Qu.: 75.00	3rd Qu.:15.38	3rd Qu.:27.48
Max. :0.8816	Max. :100.00	Max. :46.76	Max. :75.48
NA's :1	NA's :1	NA's :5	NA's :1
 POC_percen	 transporta	 unemploy_2	 housing_pe
Min. : 7.54	Min. :10.00	Min. : 1.000	Min. :15.14
1st Qu.:23.36	1st Qu.:18.00	1st Qu.: 3.350	1st Qu.:27.34
Median :36.29	Median :19.00	Median : 4.480	Median :32.26
Mean :38.64	Mean :18.97	Mean : 5.099	Mean :33.75
3rd Qu.:51.01	3rd Qu.:21.00	3rd Qu.: 6.460	3rd Qu.:39.13
Max. :92.70	Max. :26.00	Max. :24.400	Max. :81.89
NA's :1	NA's :1	NA's :3	NA's :1
 traffic_pe	 diesel	 ozone	 PM25
Min. : 0.00	Min. : 0.14	Min. :46.73	Min. :3.787
1st Qu.: 0.00	1st Qu.: 6.65	1st Qu.:48.91	1st Qu.:5.642
Median : 3.60	Median :12.65	Median :49.78	Median :6.180
Mean :16.07	Mean :17.10	Mean :50.62	Mean :6.186
3rd Qu.:26.17	3rd Qu.:18.99	3rd Qu.:51.28	3rd Qu.:6.872
Max. :97.75	Max. :92.63	Max. :62.89	Max. :7.897
NA's :1	NA's :1	NA's :1	NA's :1
 toxic_rele	 hazardous_	 lead_perce	 superfund
Min. : 823.9	Min. :0.02303	Min. : 0.24	Min. :0.03454
1st Qu.: 5180.9	1st Qu.:0.04168	1st Qu.: 6.46	1st Qu.:0.07358
Median : 10186.5	Median :0.05160	Median :13.79	Median :0.13133
Mean : 19398.3	Mean :0.08190	Mean :17.08	Mean :0.24645
3rd Qu.: 20058.1	3rd Qu.:0.09280	3rd Qu.:26.20	3rd Qu.:0.28436

```
Max.    :186434.6   Max.    :0.63781   Max.    :54.68    Max.    :1.46778
NA's     :1          NA's     :1          NA's     :1          NA's     :1
  facilities       wastewater      sen_pop_pe    socio_perc
Min.    :0.0523    Min.    :0.00e+00  Min.    : 1.00    Min.    : 1.00
1st Qu.:0.1612    1st Qu.:5.50e-06  1st Qu.: 25.00  1st Qu.: 25.00
Median  :0.3652    Median :5.30e-04   Median : 50.00  Median : 50.00
Mean    :0.6046    Mean    :2.62e-02  Mean    : 50.38  Mean    : 50.38
3rd Qu.:0.9119    3rd Qu.:8.70e-03  3rd Qu.: 75.00  3rd Qu.: 75.00
Max.    :3.3682    Max.    :6.40e-01  Max.    :100.00  Max.    :100.00
NA's     :1          NA's     :1          NA's     :1          NA's     :1
  geometry
MULTIPOLYGON :398
epsg:2926     : 0
+proj=lcc ....: 0
```

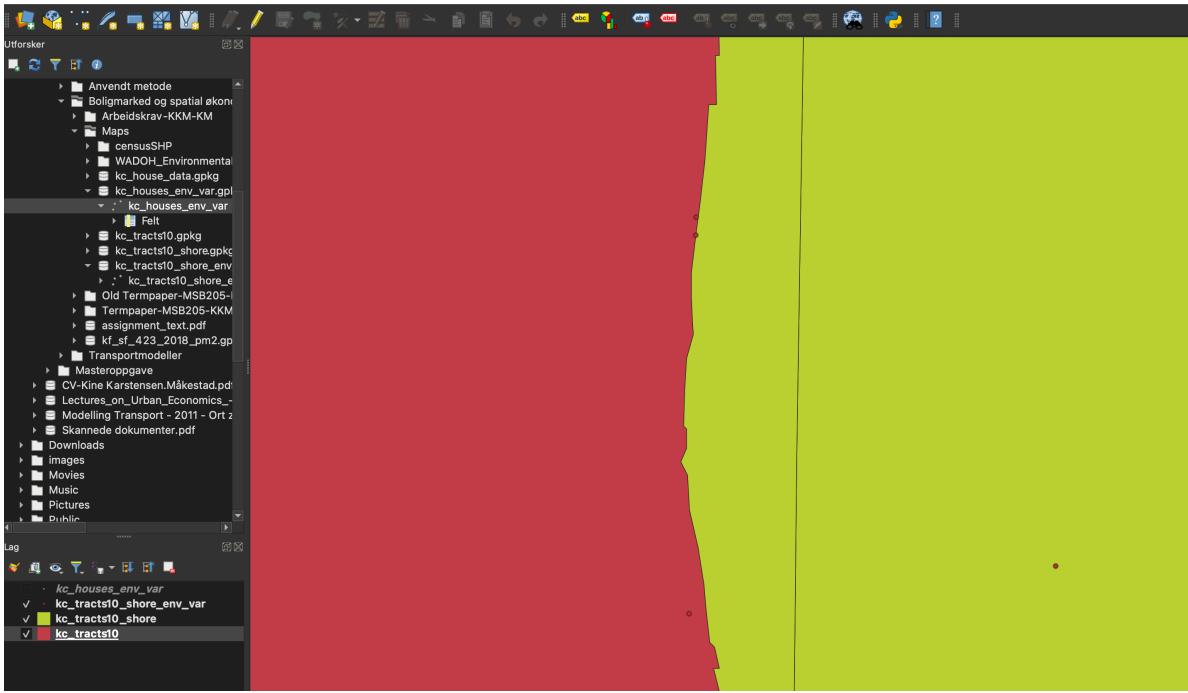
||

Når vi ser på disse dataene fra tracts10 både i R og i QGIS, så ser vi at kc\_tracts10\_env\_data har en observasjon som går utenfor kommunegrensene.



Som en kan se på dette bilde er det den lille prikken som ligger utenfor kommunegrensen.

For den andre kc\_tracts10\_shore\_env\_data kan en se at det er ligger alle observasjonene innenfor kommunegrensene. Men nå vi la til shore-kartet kan en se at ved vannlinjen er det 25 observasjoner som ligger rett utenfor. Det er også grunnen til at det gir 25 NA.



iii

Her har vi droppet områdevariabelen med id 3518000180.

```

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

st_write(kc_house_env_var.omit, "../Maps/kc_house_env_var.omit.gpkg", append = FALSE)

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

```

Nå skal vi lage en faktor-variabel av år og måned ut fra Date.

```

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

```

Deretter skal vi slette gpkg filen fra husdataene og lagre den på ny.

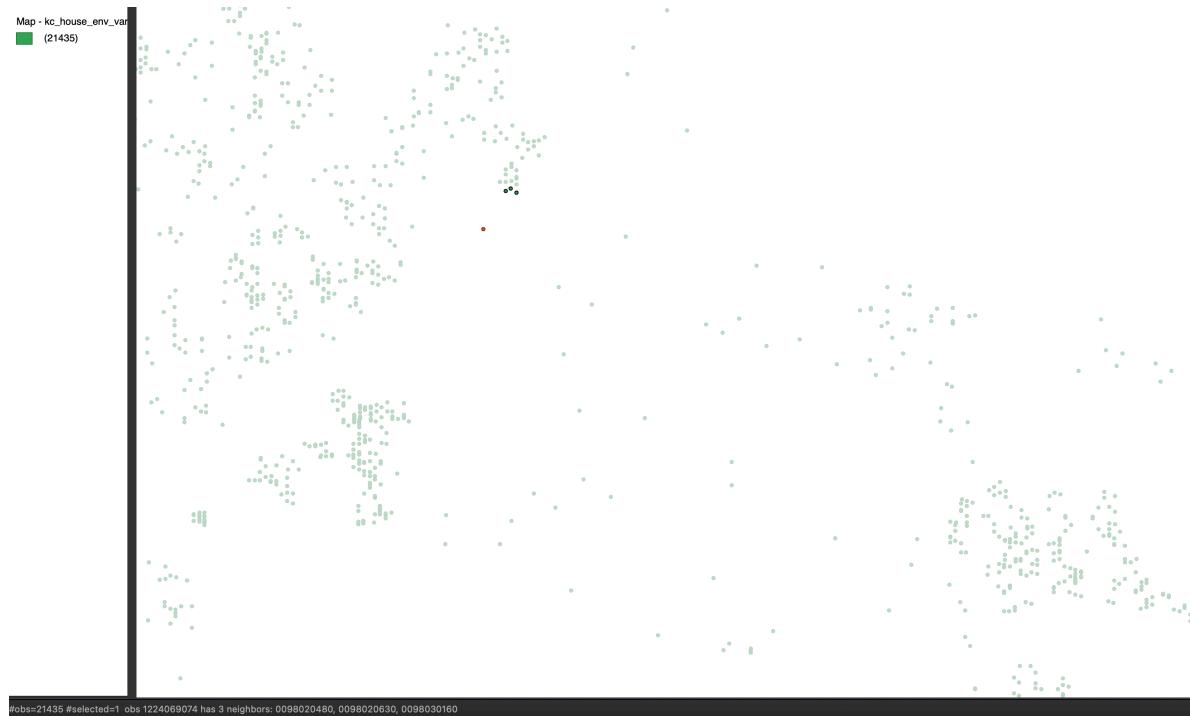
```
st_write(kc_house_env_var OMIT, ".../Maps/kc_house_env_var OMIT.gpkg", append = FALSE)
```

```
Deleting layer `kc_house_env_var OMIT' using driver `GPKG'  
Writing layer `kc_house_env_var OMIT' to data source  
`.../Maps/kc_house_env_var OMIT.gpkg' using driver `GPKG'  
Writing 21435 features with 66 fields and geometry type Point.
```

## Oppg. 5

III

Her genererer vi en vekt fil utfra 3 nærmeste nabøer.



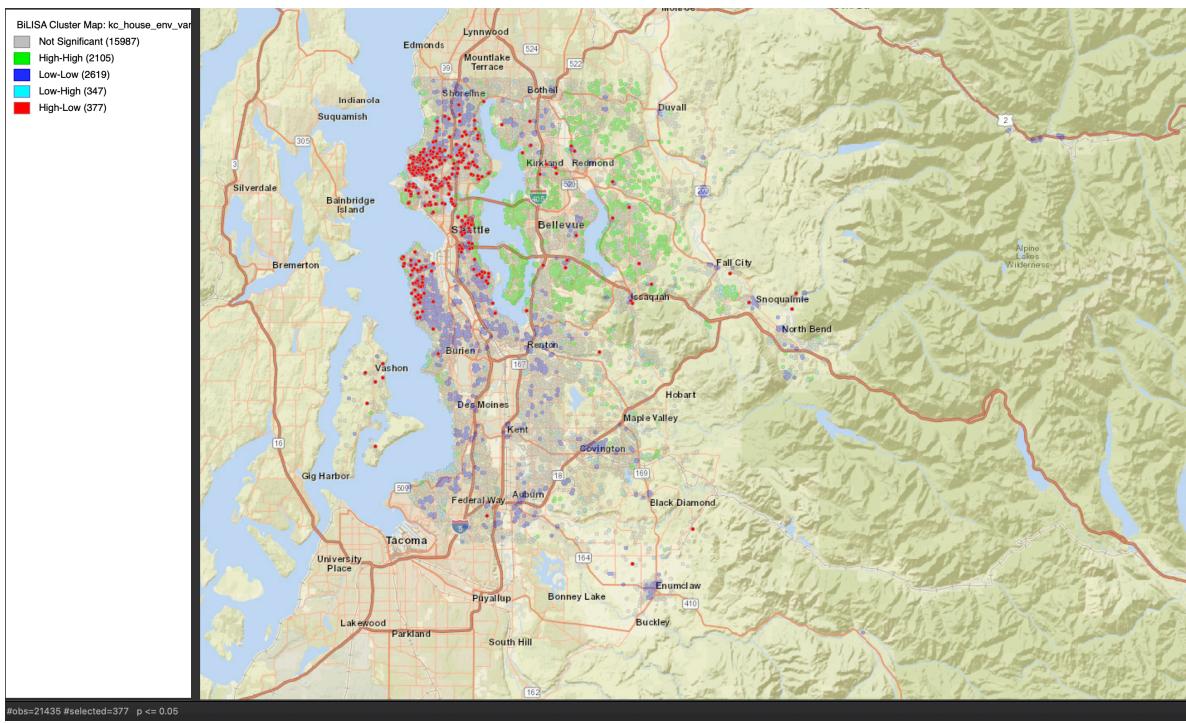
Nå skal vi generere en fil utfra en en plass med 10 nærmeste nabøer.



## IV

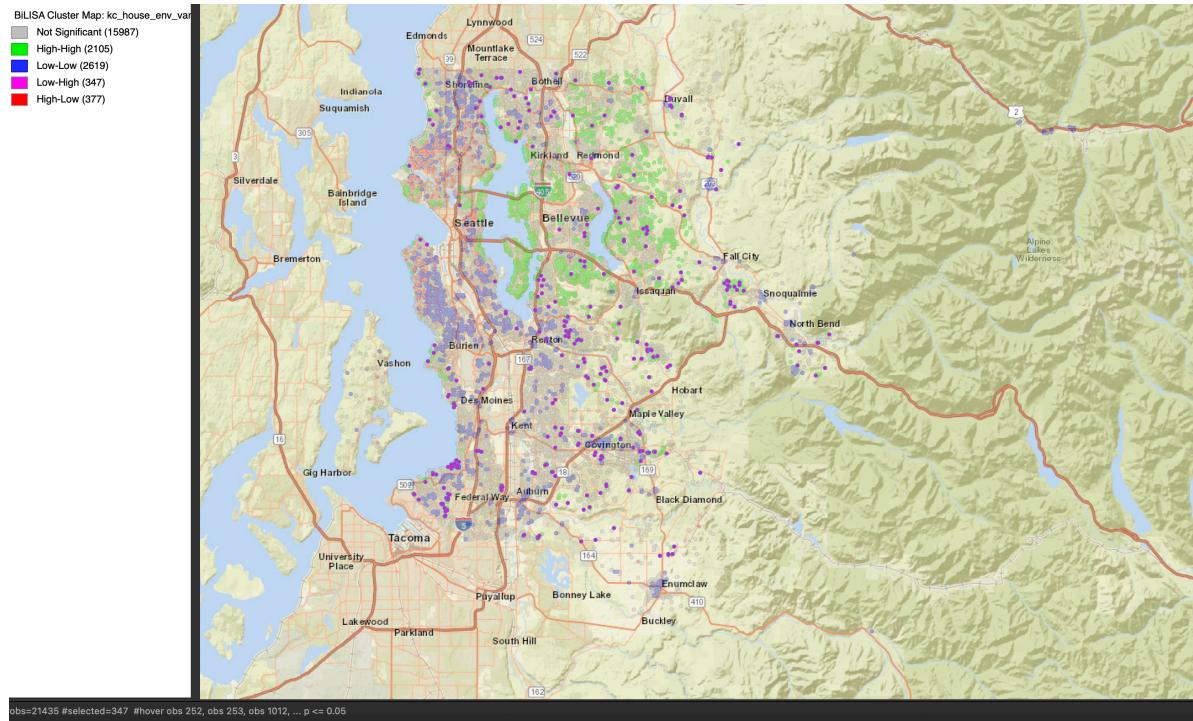
### Nærmeste nabo 3:

Store og billige boliger.

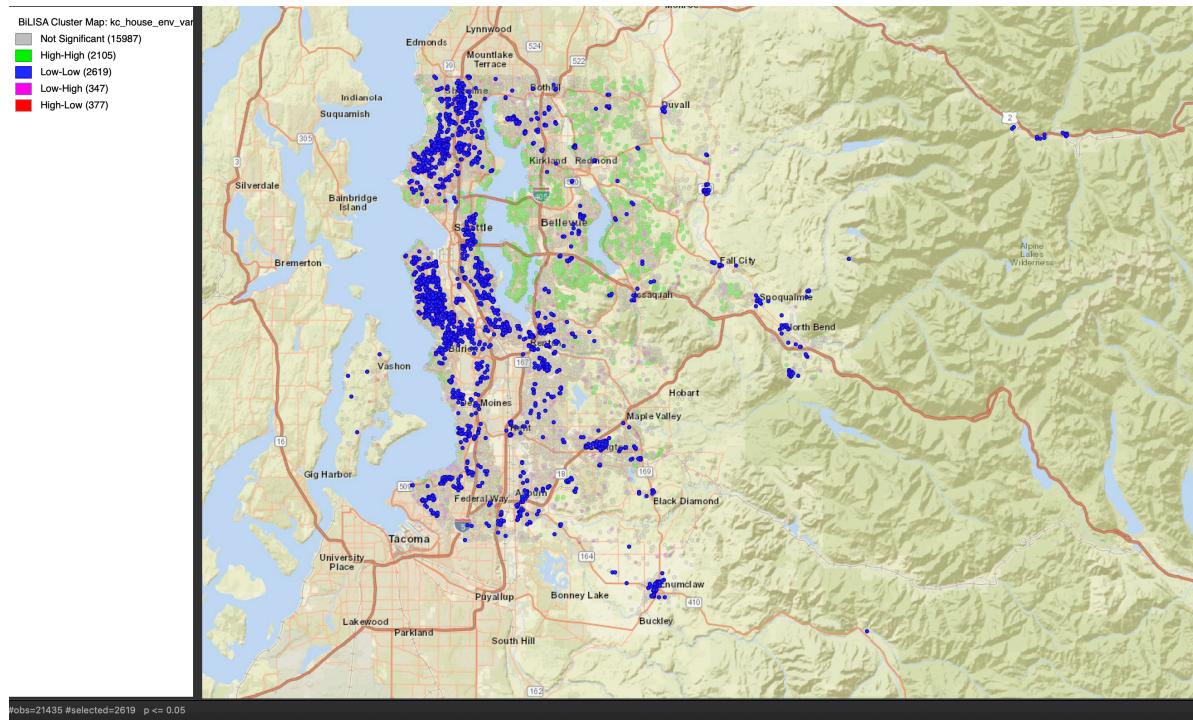


På bilde ovenfor ser man på de røde prikkene, og det er de plassene som er store og små boliger.

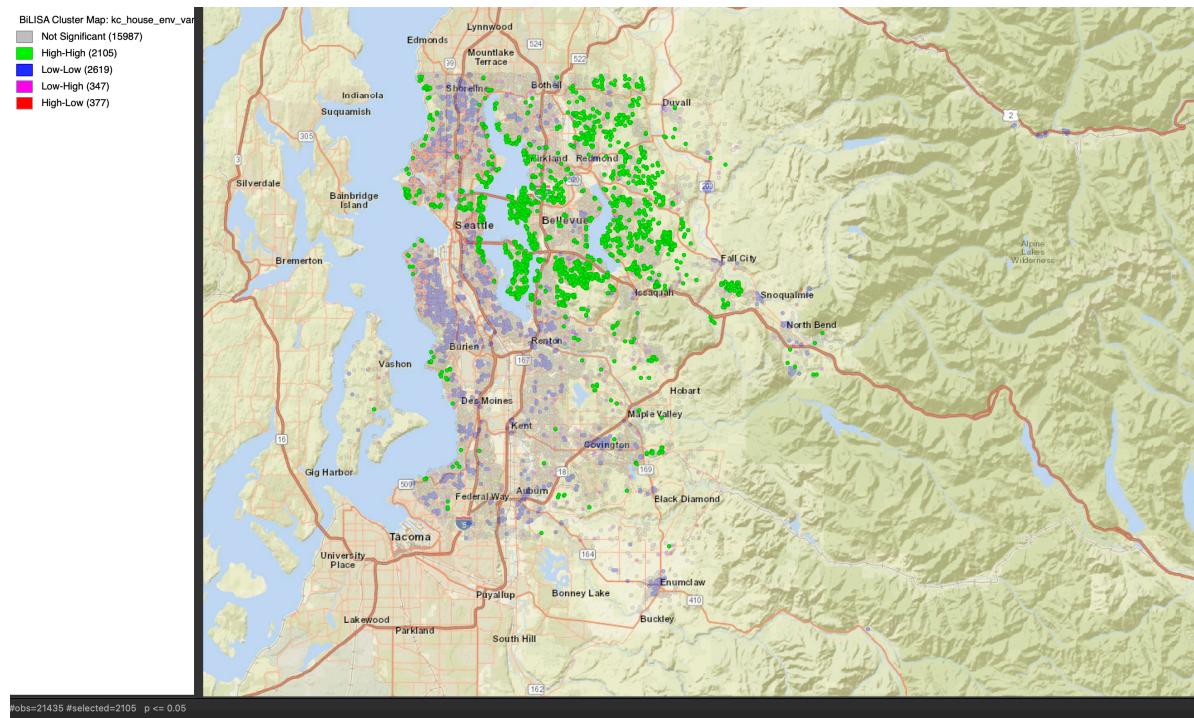
Neste er små og dyre boliger som er markert i rosa.



Deretter kommer de små og billige boligene som er markert i blå.

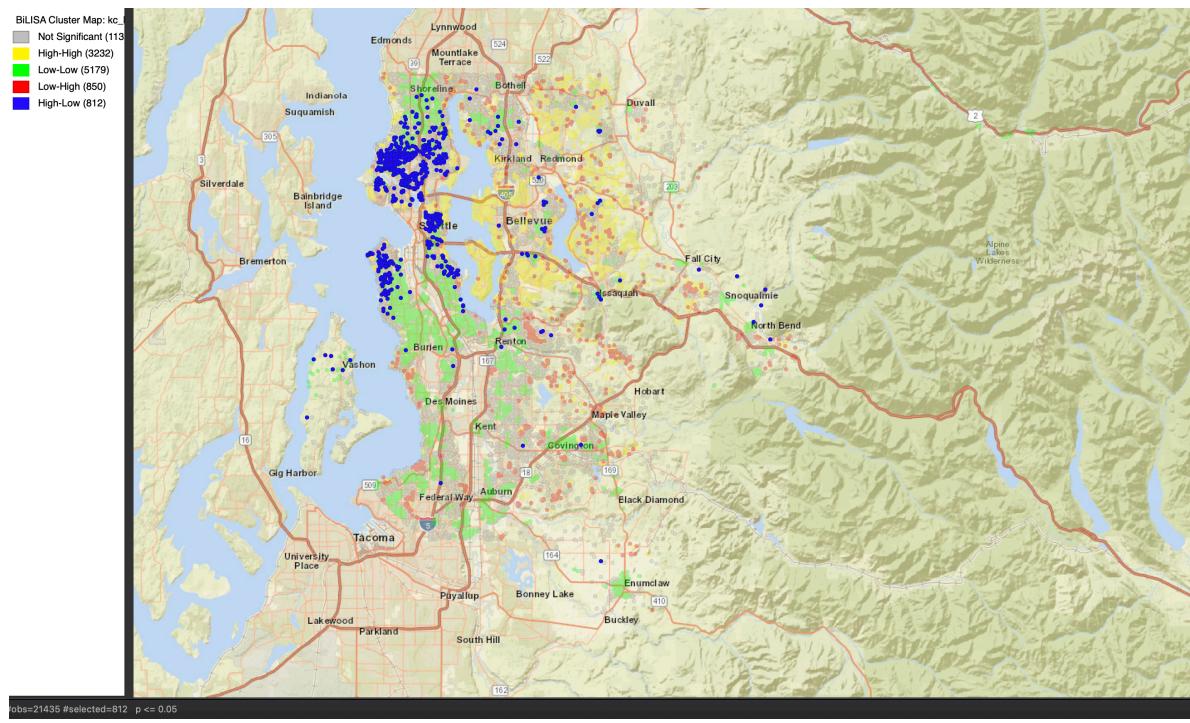


Til slutt er det de store og dyre boligene som er markert i grønn.

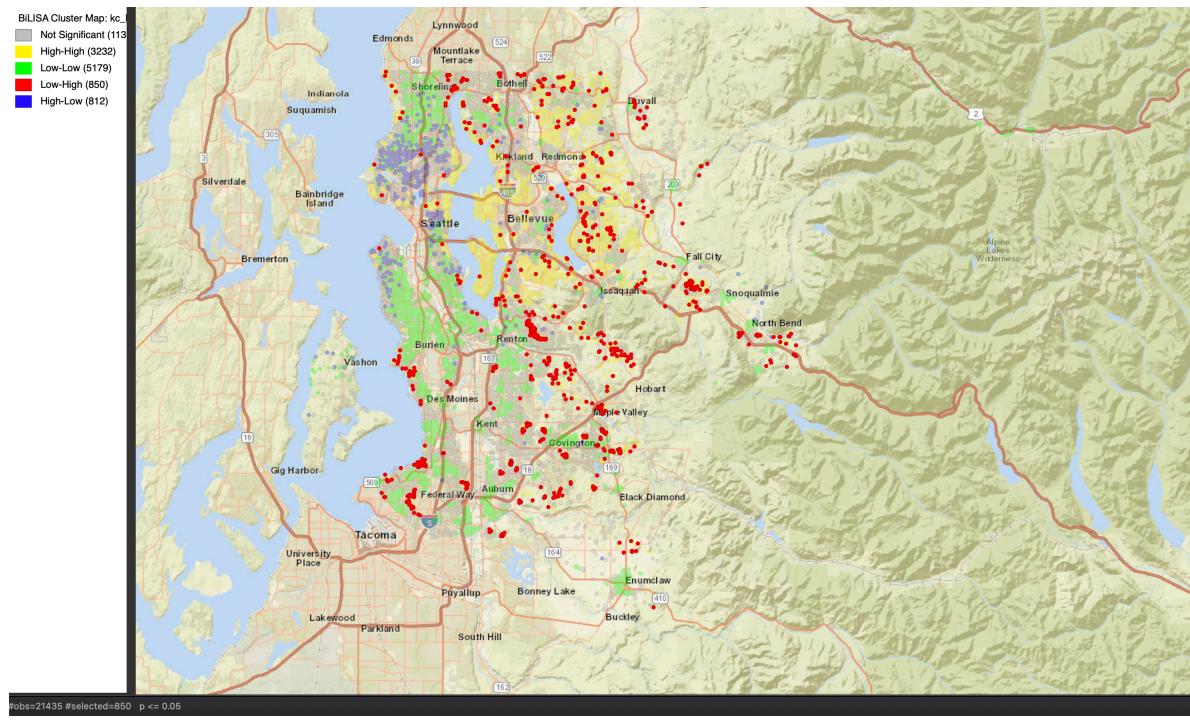


### Nærmeste nabo 10:

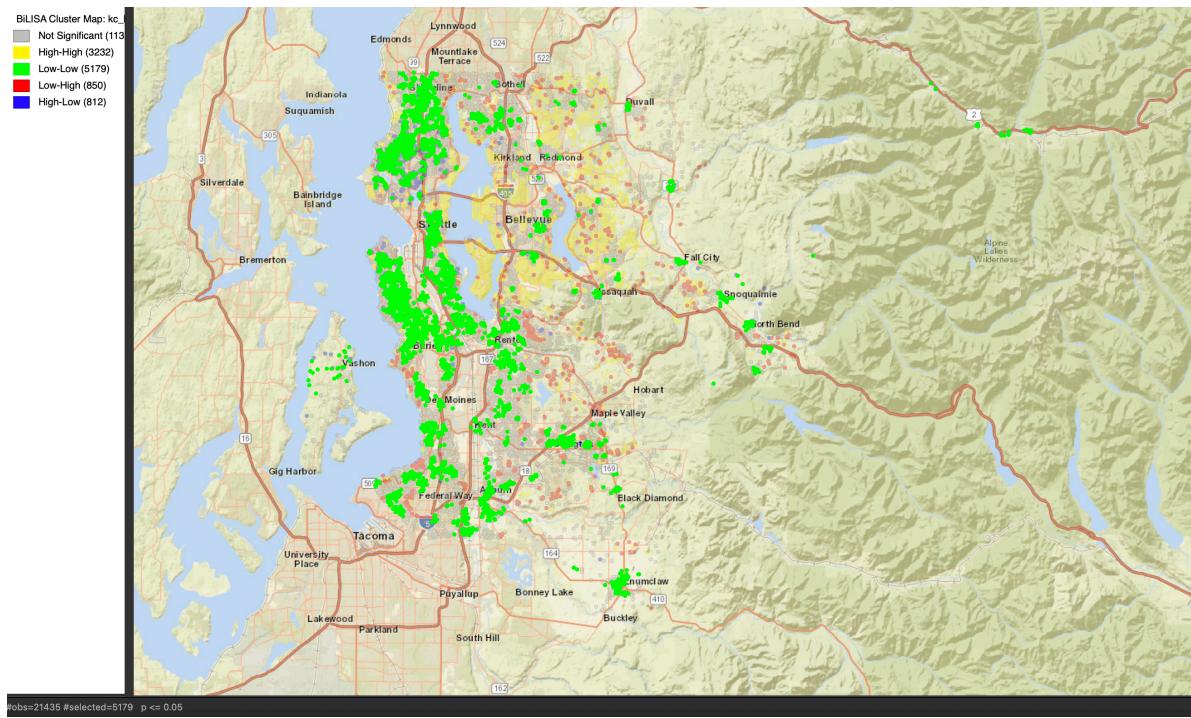
Første er store og billige boligene som en kan se på bilde nedenfor som er markert i blått.



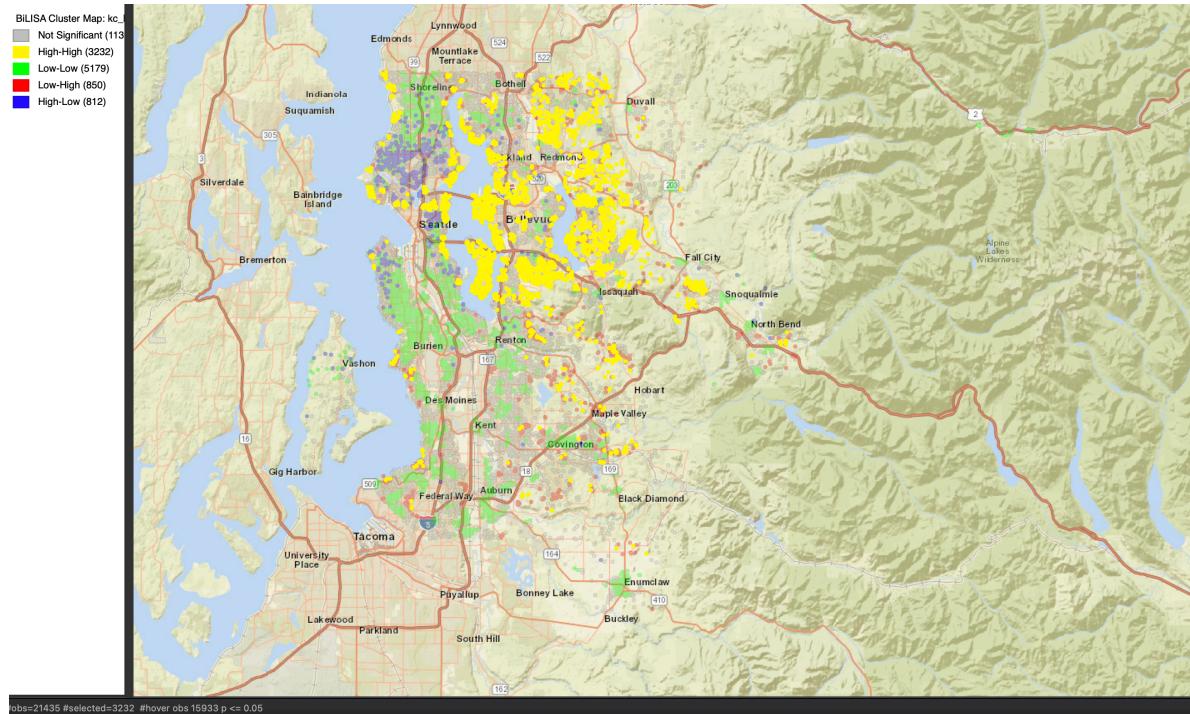
Så kommer små og dyre boliger, og det er markert i rødt.



Deretter kommer små og billige boliger, og dette er markert med grønt på kartet.



Til slutt kommer store og dyre boligene, og de er markert med gult på kartet nedenfor.



## Oppg. 6

### I Sammenfatning av funnene fra EDA

Etter å ha sett på pris i forhold til størrels epå boligen, kan en se at de små og billige boligene sentrer seg rundt Seattle sentrum. Da har vi sett i fordelingen på knear = 3 og knear = 10, der begge er nokså like. Ser en på de store og dyre boligene er de både med tre og ti nabover plasser rundt Bellevue og mer øst for Seattle. Under små og dyr boliger er mye spredt rundt Seattle, mest utenfor sentrum området. Mens store og billige boliger er veldig sentrert i sentrum av Seattle, som de andre er det likt både med tre nabover.

Etter å ha sett på Moran's I får en verdier på tre nabover = 0,399 og med ti nabover = 0.350. Ved positiv verdi vil det si at det klynger, og perfekt klynging er 1, der våre verdier viser til litt klyning, men også spredning. Noen kan se i kartene ovenfor. Dette vil stemme overens med Moran's I verdier som sier at perfekt tilnærming er 0.

II

1

Huskarakteristiska og tids-dummier

```
attach(kc_house_env_var_omit)
```

```
mod1 <- "price ~ bedrooms + bathrooms + sqft_living + sqft_living15 + sqft_lot + sqft_lot1
```

## 2

Huskarateristika, dist\_CBD, relevante tract variabler og tids-dummier.

```
mod2 <- "price ~ bedrooms + bathrooms + sqft_living + sqft_living15 + sqft_lot + sqft_lot1  
hazardous_ + lead_perce + socio_perc"
```

## 3

Huskarateristika, dist\_CBD, EHD indeks og tids-dummier.

```
mod3 <- "price ~ bedrooms + bathrooms + sqft_living + sqft_living15 + sqft_lot + sqft_lot1
```

Hedoniske modeller

```
hedon1 <- lm(mod1, data = kc_house_env_var_omit)  
hedon2 <- lm(mod2, data = kc_house_env_var_omit)  
hedon3 <- lm(mod3, data = kc_house_env_var_omit)
```

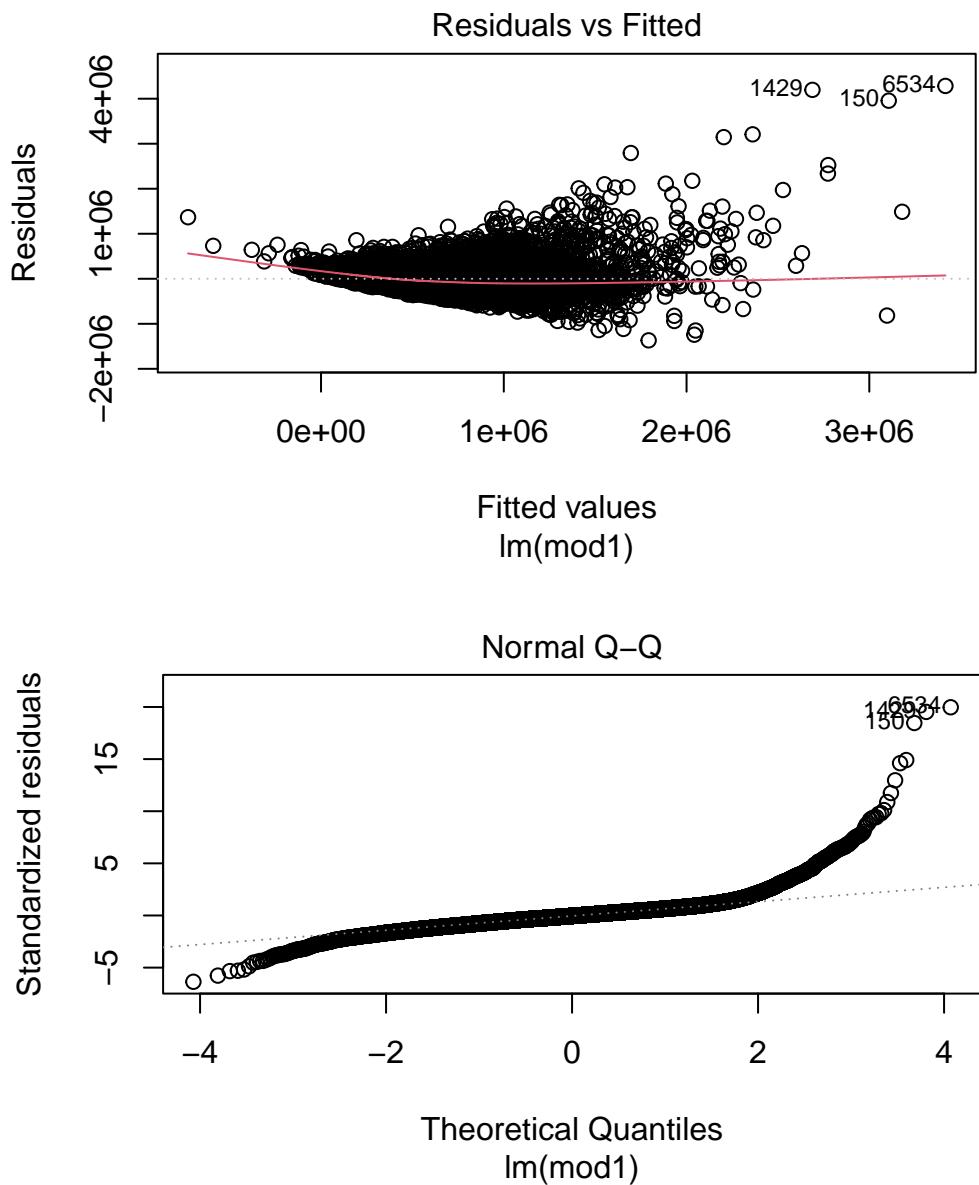
```
huxreg("Hedon1" = hedon1, "Hedon2" = hedon2, "Hedon3" = hedon3, error_format = "[{statisti
```

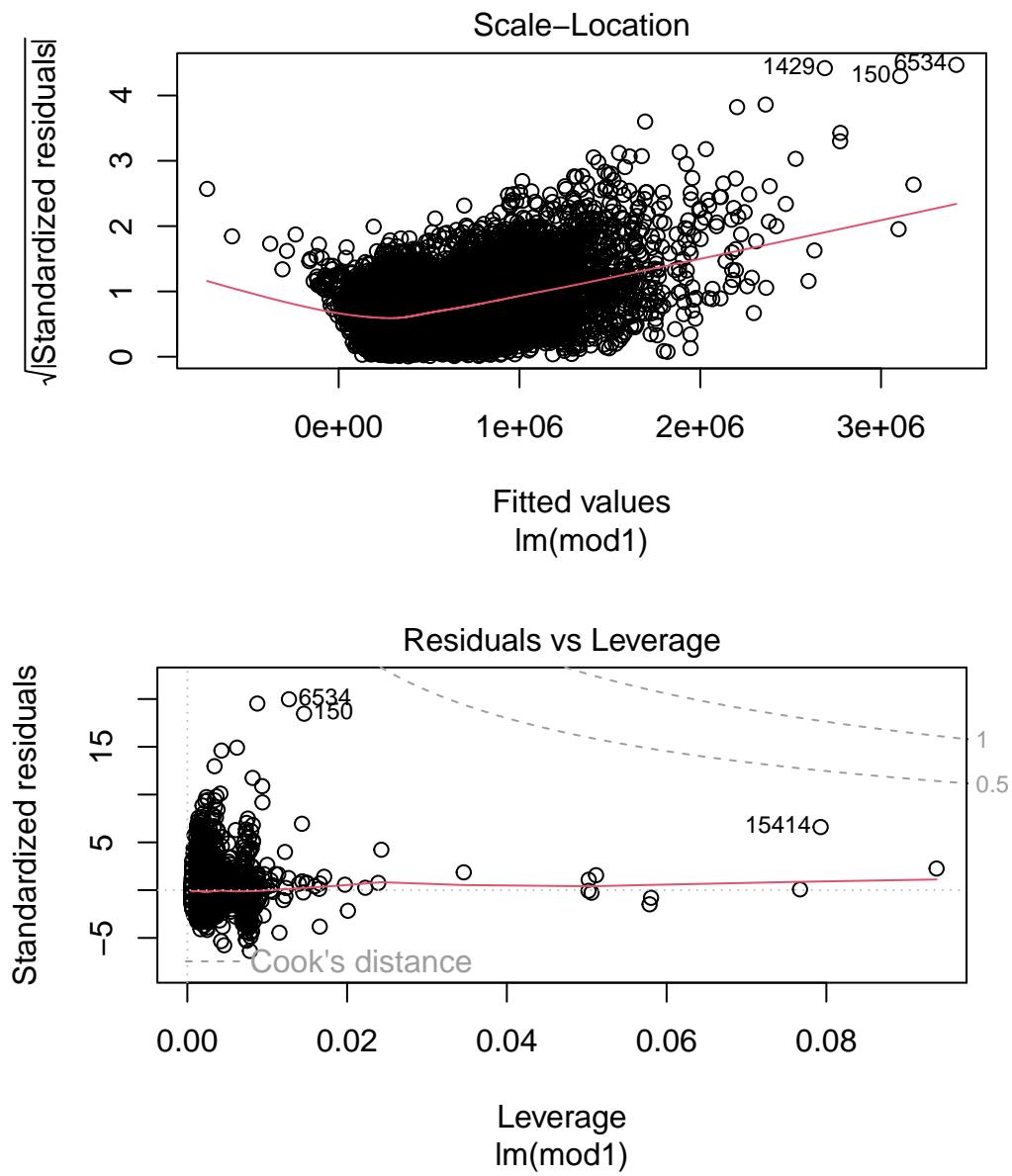
## iv

Etter å ha sett dataen i tre ulike hedoniske modeller, kan en se at mod2 har en bedre marginal forklaringskraft i forhold til mod3. Grunnen til dette er fordi mod3 har mindre variabler og EHD-percen inneholder alle de miljømessige varibler fra mod2. En kan se alle de variablene en trenger i mod2.

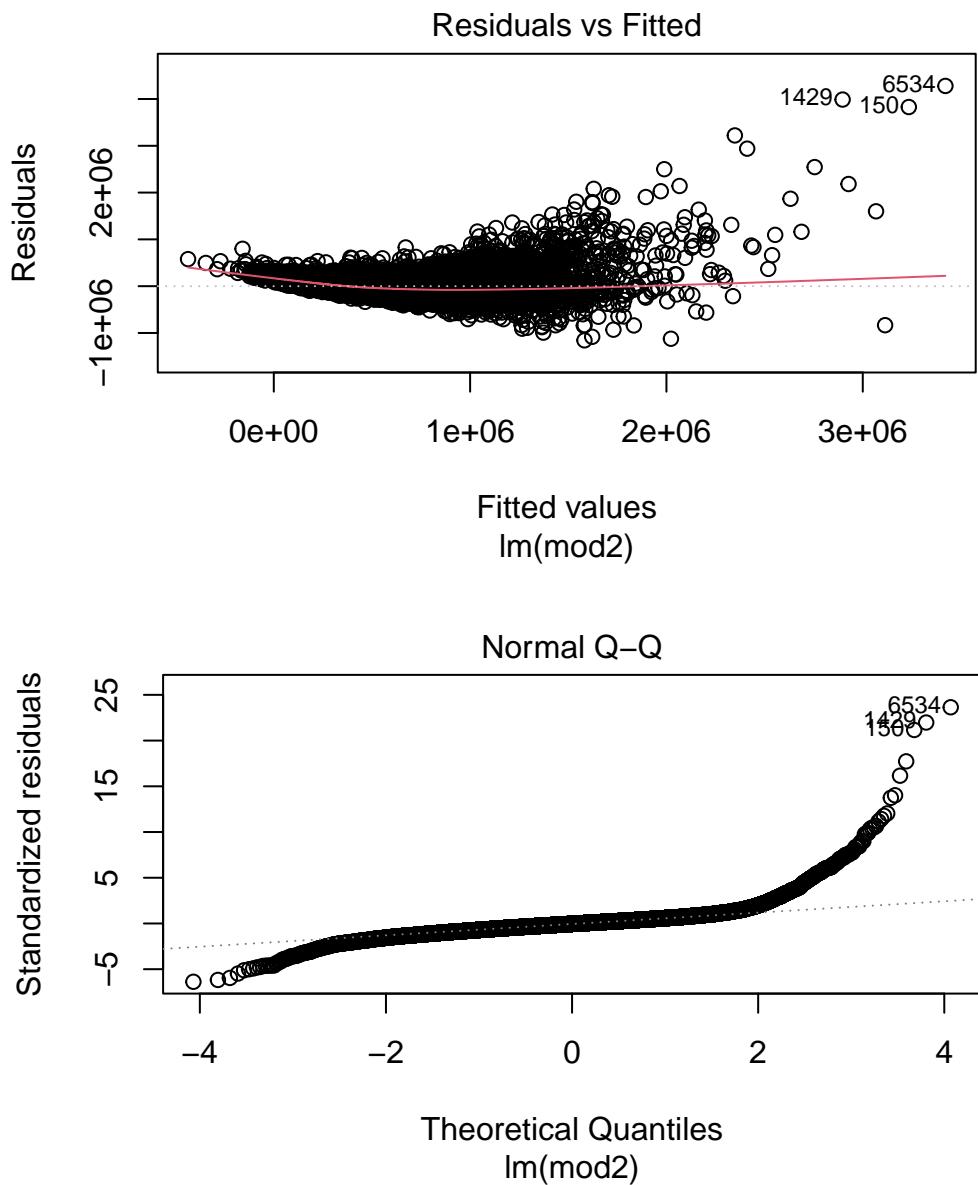
Ser på plot(km\_navnet)

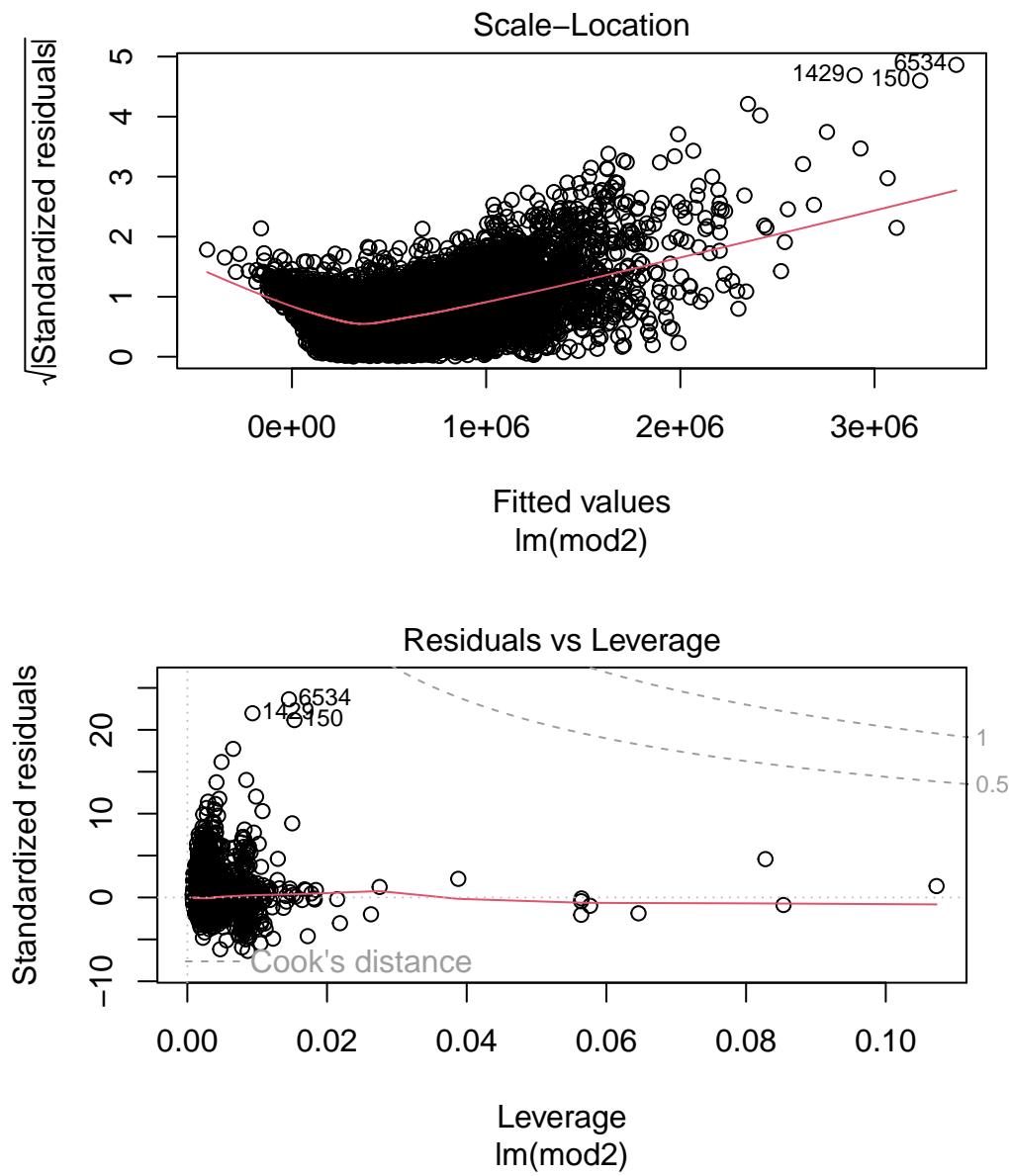
```
hedon1 %>%  
plot()
```



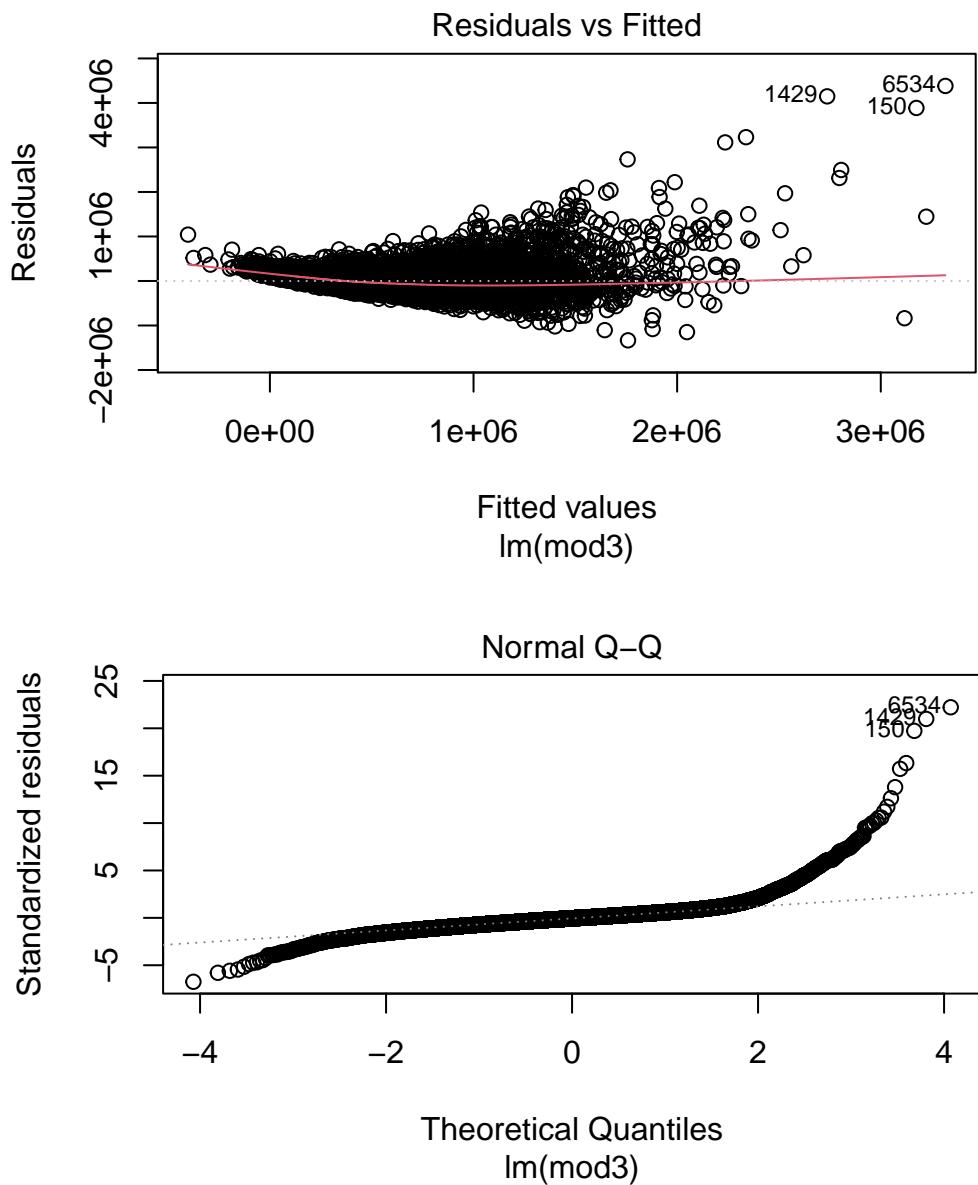


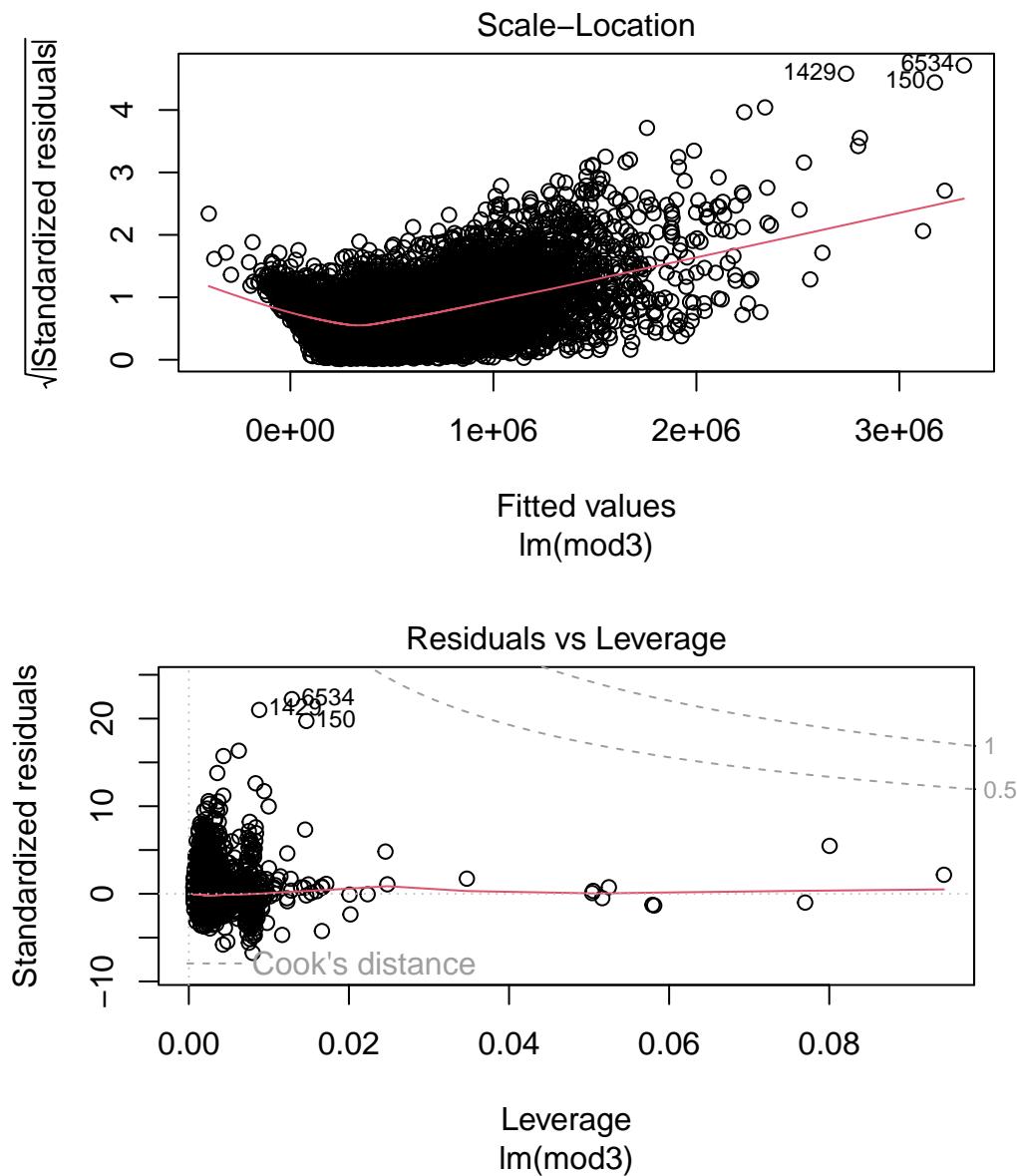
```
hedon2 %>%
  plot()
```





```
hedon3 %>%
  plot()
```





## Oppg. 7

Her bruker vi en simultan test på tids-dummiene for å avgjøre om vi trenger disse i modellen.

```
hedoni %>%
  linearHypothesis(c("year_month2014-06=0", "year_month2014-07=0",
                     "year_month2014-08=0", "year_month2014-09=0",
                     "year_month2014-10=0", "year_month2014-11=0",
```

```

"year_month2014-12=0", "year_month2015-01=0",
"year_month2015-02=0", "year_month2015-03=0",
"year_month2015-04=0", "year_month2015-05=0"),
white_adjust = hc3)

hedon2 %>%
  linearHypothesis(c("year_month2014-06=0", "year_month2014-07=0",
                     "year_month2014-08=0", "year_month2014-09=0",
                     "year_month2014-10=0", "year_month2014-11=0",
                     "year_month2014-12=0", "year_month2015-01=0",
                     "year_month2015-02=0", "year_month2015-03=0",
                     "year_month2015-04=0", "year_month2015-05=0"),
                     white_adjust = hc4)

hedon3 %>%
  linearHypothesis(c("year_month2014-06=0", "year_month2014-07=0",
                     "year_month2014-08=0", "year_month2014-09=0",
                     "year_month2014-10=0", "year_month2014-11=0",
                     "year_month2014-12=0", "year_month2015-01=0",
                     "year_month2015-02=0", "year_month2015-03=0",
                     "year_month2015-04=0", "year_month2015-05=0"),
                     white_adjust = hc1)

```

Når vi ser på disse testene ser vi på F-verdiene og p-verdiene for å se om vi skal forkaste H0. I følge testene vil vi da forkaste H0 da p-verdien er lavere signifikantnivået. Dette betyr da at vi vil kunne trenge disse tids-dummiene i modellen.

## Oppg. 8

|

Her laster vi inn gruppens utvalg for relevant gpkg fil.

	Hedon1	Hedon2	Hedon3
(Intercept)	6210567.524 *** [44.647]	3372625.196 *** [21.435]	3953788.747 *** [29.006]
bedrooms	-39596.759 *** [-19.444]	-26343.443 *** [-15.106]	-32336.935 *** [-17.223]
bathrooms	46467.945 *** [13.256]	27539.039 *** [9.196]	40542.973 *** [12.568]
sqft_living	167.682 *** [35.847]	129.590 *** [32.127]	157.120 *** [36.379]
sqft_living15	24.038 *** [6.662]	33.258 *** [10.233]	-5.904 [-1.721]
sqft_lot	-0.003 [-0.061]	0.211 *** [4.598]	0.068 [1.433]
sqft_lot15	-0.556 *** [-7.100]	-0.178 * [-2.564]	-0.503 *** [-6.938]
sqft_above	-6.495 [-1.426]	82.422 *** [20.251]	25.189 *** [5.938]
floors	26914.959 *** [7.100]	-60197.734 *** [-17.189]	-4573.083 [-1.295]
grade	120002.514 *** [53.176]	66185.300 *** [32.615]	91928.479 *** [43.269]
yr_built	-3584.346 *** [-50.313]	-663.856 *** [-8.839]	-2604.253 *** [-38.234]
yr_renovated	10.616 ** [2.706]	31.350 *** [9.305]	16.397 *** [4.537]
waterfront	579243.823 *** [31.092]	614229.057 *** [38.668]	613889.766 *** [35.803]
condition	20271.885 *** 33 [8.057]	30566.687 *** [14.046]	25868.926 *** [11.091]
view	42889.032 *** [18.823]	47818.802 *** [24.372]	48886.142 *** [23.304]
year_month2014-06	3436.683	8506.371	6587.530

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
2.14e+04	1e+15				
2.14e+04	9.97e+14	12	5.35e+12	9.57	7.55e-19

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
2.12e+04	7.09e+14				
2.12e+04	7.03e+14	12	6e+12	15	4.97e-32

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
2.14e+04	8.5e+14				
2.14e+04	8.43e+14	12	6.15e+12	13	4.17e-27