

# hent-ssb-data

Kevin Ha - 571821

Ola Andre Olofsson - 170745

•

```
# Vector med relevante kommunenummer  
load("knr.Rdata")
```

## Gjennomsnittlig kvadratmeterpris

I denne oppgaven skal vi studere prisen per kvadratmeter i et bestemt utvalg av kommuner. Oppgaven baserer seg på data som stammer fra, og er produsert av SSB. Følgelig henter vi ut dataene derfra.

### Dataauthenting, SSB

```
pm2_raw <- ApiData(  
  urlToData = "06035",  
  Region = knr,  
  ContentsCode = "KvPris",  
  Boligtype = "01",  
  Tid = c(  
    as.character(2002:2017)))
```

•

```
pm2 <- pm2_raw$dataset %>%  
  tibble() %>%  
  select(-Boligtype, -ContentsCode) %>%  
  rename(  
    knr = Region,  
    aar = Tid,  
    pm2 = value)
```

•

```
names(pm2_raw)[[1]] <- "desc"
```

•

```
pm2 <- pm2 %>%
  mutate(knavn = pm2_raw$desc$region) %>%
  # vent med gruppering
  # group_by(knr) %>%
  select(knr, aar, pm2, knavn)
```

•

```
pm2 <- pm2 %>%
  select(knr, knavn, aar, pm2)
```

•

```
head(pm2)
```

```
## # A tibble: 6 x 4
##   knr   knavn      aar   pm2
##   <chr> <chr>    <chr> <int>
## 1 0101 Halden (-2019) 2002   9070
## 2 0101 Halden (-2019) 2003   9301
## 3 0101 Halden (-2019) 2004   9436
## 4 0101 Halden (-2019) 2005  10846
## 5 0101 Halden (-2019) 2006  12052
## 6 0101 Halden (-2019) 2007  12363
```

•

Boligtypen vi er interessert i er eneboliger. Ifra APIen til SSB så ser vi at denne er gitt ved “01”. Vi har dermed satt våres boligtype i *pm2\_raw* til å være “01” slik at vi kun får eneboliger i datasettet. Ettersom vi vet at vi skal finne pris per kvadratmeter til eneboliger, så kan vi bruke denne forutsetningen til å fjerne *boligtype* og *ContentsCode* fra *pm2\_raw*.

## Fjerning av kommune årstall.

I kommunenavnene er det oppgitt når de ulike kommunene ble oppløst. Vi ønsker å fjerne dette slik at vi kun får navnene på selve kommunene.

## Mønster

```
load("test_string_tib.Rdata")
```

•

```
# Legger inn regex mønster
moenster <- '\\s*\\([\\d\\s-]*\\d*\\)\\s*$'
```

•

```
pm2 <- pm2 %>%
  mutate(
    knavn = str_replace(knavn, moenster, "")
```

## NA-Verdier

Vi sjekker videre hvor mange NA-verdier vi har i datasettet *pm2*.

```
pm2 %>%
  map_df(is.na) %>%
  map_df(sum) %>%
  as.tibble()
```

```
## Warning: 'as.tibble()' was deprecated in tibble 2.0.0.
## Please use 'as_tibble()' instead.
## The signature and semantics have changed, see '?as_tibble'.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was generated.

## # A tibble: 1 x 4
##   knr knavn  aar  pm2
##   <int> <int> <int> <int>
## 1     0     0     0 2903
```

Vi ser ut fra chunken *NA\_Verdier* at vi har 2 903 NA-observasjoner på pm2-rekken.

## Complete Cases

Complete Cases, antall rekker uten NA verdier, fom. 2006 tom. 2017:

```
pm2_2006 <- pm2 %>%
  filter(aar >= 2006) %>%
  pivot_wider(
    names_from = aar,
    values_from = pm2)
```

•

```
pm2_2006 %>%
  complete.cases() %>%
  sum()
```

```
## [1] 197
```

Complete Cases, antall rekker uten NA-verdier, f.o.m. 2008 t.o.m. 2017:

```
pm2_2008 <- pm2 %>%
  filter(aar >= 2008) %>%
```

```
pivot_wider(
  names_from = aar,
  values_from = pm2)
```

•

```
pm2_2008 %>%
  complete.cases() %>%
  sum()
```

```
## [1] 214
```

Left\_joiner

```
pm2 <- pm2 %>%
  left_join(pm2_2008) %>%
  na.omit()
```

```
## Joining, by = c("knr", "knavn")
```

**Konkluderer datainnsamling for gjennomsnittspris.**

Vi trenger ikke lengre pm2\_raw og fjerner denne.

```
rm(pm2_raw, test_string_tib)
```

## Andel i arbeidsfør alder

### Dataauthenting, SSB

```
pop_08_17_ya_raw <- ApiData(
  urlToData = "07459",
  Region = knr,
  Kjonn = c(1, 2),
  Alder = list("agg:TredeltGrupperingB2",
               c("F20-64")),
  Tid = c(as.character(2008:2017))
)$dataset %>%
  select(-ContentsCode, -Alder)
```

•

```
pop_08_17_ya <- pop_08_17_ya_raw %>%
  pivot_wider(
    id_cols = c(Region, Tid),
    names_from = Kjonn,
```

```
names_prefix = "sex",
values_from = value)
```

•

```
names(pop_08_17_ya)[[1]] <- "knr"
names(pop_08_17_ya)[[2]] <- "aar"
names(pop_08_17_ya)[[3]] <- "ya_menn"
names(pop_08_17_ya)[[4]] <- "ya_kvinner"
```

•

```
pop_08_17_ya <- pop_08_17_ya %>%
  mutate(ya_Total = ya_menn+ya_kvinner)
```

•

```
dim(pop_08_17_ya)
```

```
## [1] 4230    5
```

•

```
names(pop_08_17_ya)
```

```
## [1] "knr"          "aar"          "ya_menn"      "ya_kvinner" "ya_Total"
```

## Dataauthenting 2, SSB

```
pop_08_17_raw <- ApiData(
  urlToData = "07459",
  Region = knr,
  Kjonn = c(1, 2),
  Alder = list("agg:TodeltGrupperingB",
               c("H17", "H18")),
  Tid = c(as.character(2008:2017))
)$dataset %>%
  select(-ContentsCode)
```

•

```
pop_08_17 <- pop_08_17_raw %>%
  pivot_wider(
    names_from = Kjonn,
    values_from = value)
```

•

```
names(pop_08_17)[[1]] <- "knr"
names(pop_08_17)[[2]] <- "Alder"
names(pop_08_17)[[3]] <- "aar"
names(pop_08_17)[[4]] <- "Menn"
names(pop_08_17)[[5]] <- "Kvinner"
```

•

```
pop_08_17 <- pop_08_17 %>%
  pivot_wider(
    names_from = Alder,
    values_from = c(Menn, Kvinner))
```

•

```
pop_08_17 <- pop_08_17 %>%
  mutate(Menn_t = Menn_H17 + Menn_H18) %>%
  mutate(Kvinner_t = Kvinner_H17 + Kvinner_H18) %>%
  mutate(Totalt_t = Menn_t + Kvinner_t)
```

•

```
pop_08_17 <- pop_08_17 %>%
  select(knr, aar, Menn_t, Kvinner_t, Totalt_t)
```

•

```
dim(pop_08_17)
```

```
## [1] 4230    5
```

•

```
names(pop_08_17)
```

```
## [1] "knr"      "aar"      "Menn_t"   "Kvinner_t" "Totalt_t"
```

•

## Sammenslåing

```
pop_08_17_ya_p <- merge(pop_08_17, pop_08_17_ya)
```

Muterer nye variabler prosentvis yrkesaktiv alder

```
pop_08_17_ya_p <- pop_08_17_ya_p %>%
  mutate(Menn_ya_p = ya_menn/Menn_t*100) %>%
```

```
mutate(Kvinner_ya_p = ya_kvinner/Kvinner_t*100) %>%
mutate(Totalt_ya_p = ya_Total/Totalt_t*100)
```

•

```
pop_08_17_ya_p <- pop_08_17_ya_p %>%
  select(knr, aar, Menn_ya_p, Kvinner_ya_p, Totalt_ya_p)
```

Sjekker om vi har gjort rett

```
head(pop_08_17_ya_p, n=5)
```

```
##      knr  aar Menn_ya_p Kvinner_ya_p Totalt_ya_p
## 1 0101 2008  59.74892    56.79763    58.26214
## 2 0101 2009  59.77860    57.04693    58.40290
## 3 0101 2010  59.64298    57.06300    58.34376
## 4 0101 2011  59.84630    57.22382    58.53183
## 5 0101 2012  59.45122    57.00467    58.22699
```

Legger til *ya* variablene frå datasettet **pop\_08\_17\_ya\_p** inn i datasettet **pm2**.

```
pm2 <- merge(pm2, pop_08_17_ya_p)
```

•

## Rydder opp rot

```
rm(pop_08_17_raw, pop_08_17_ya_raw, pm2_2006, pm2_2008, pop_08_17, pop_08_17_ya)
```

## Øvre og Nedre Desil/Kvintil

### Dataauthenting, SSB

```
kvintiler_08_17_raw <- ApiData(
  urlToData = "12558",
  Region = knr,
  #Krever tekststrenger
  # Desiler = c(1, 2, 9, 10),
  Desiler = c("01", "02", "09", "10"),
  # ContentsCode = "VerdiDesil",
  ContentsCode = "AndelHush",
  InntektSkatt = "00",
  Tid = c(
    as.character(2008:2017)
  )
)
```

```
)$dataset %>%
select(Region, Desiler, Tid, value)
```

•

```
#Nu går alt så meget bedre ;-)
kvintiler_08_17 <- kvintiler_08_17_raw %>%
  pivot_wider(
    names_from = Desiler,
    values_from = value)
```

•

```
names(kvintiler_08_17)[[1]] <- "knr"
names(kvintiler_08_17)[[2]] <- "aar"
names(kvintiler_08_17)[[3]] <- "Desil_1"
names(kvintiler_08_17)[[4]] <- "Desil_2"
names(kvintiler_08_17)[[5]] <- "Desil_9"
names(kvintiler_08_17)[[6]] <- "Desil_10"
```

•

```
kvintiler_08_17 <- kvintiler_08_17 %>%
  mutate(inc_k1 = Desil_1 + Desil_2) %>%
  mutate(inc_k5 = Desil_9 + Desil_10)
```

•

```
kvintiler_08_17 <- kvintiler_08_17 %>%
  select(knr, aar, inc_k1, inc_k5)
```

•

```
pm2 <- merge(pm2, kvintiler_08_17)
```

•

Rydder opp rot, atter en gang

```
rm(kvintiler_08_17, kvintiler_08_17_raw, pop_08_17_ya_p)
```

## Utdanning

Dataauthenting, SSB. Både menn og Kvinner, Kort og Lang høy-skoleutdanning.

```
utdanning_08_17_raw <- ApiData(
  urlToData = "09429",
```



```

Region = knr,
Nivaa = c("03a", "04a"),
Kjonn = TRUE,
ContentsCode = "PersonerProsent",
Tid = c(as.character(2008:2017)))

```

•

```
uni_p_raw <- utdanning_08_17_raw
```

•

```

uni_p_raw <- tibble(
  knr = uni_p_raw$dataset$Region,
  aar = uni_p_raw$dataset$Tid,
  Kjonn = uni_p_raw$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikk`,
  nivaa = uni_p_raw$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikk`,
  uni_p = uni_p_raw$dataset$value
)

```

•

```

uni_p_raw %>%
  head(uni_p_raw, n=5)

```

```

## # A tibble: 5 x 5
##   knr   aar   Kjonn      nivaa      uni_p
##   <chr> <chr> <chr>      <chr>      <dbl>
## 1 0101  2008  Begge kjønn Universitets- og høgskolenivå, kort 17.8
## 2 0101  2009  Begge kjønn Universitets- og høgskolenivå, kort 18.2
## 3 0101  2010  Begge kjønn Universitets- og høgskolenivå, kort 18.6
## 4 0101  2011  Begge kjønn Universitets- og høgskolenivå, kort 19
## 5 0101  2012  Begge kjønn Universitets- og høgskolenivå, kort 19.6

```

•

```
uni_p <- uni_p_raw
```

•

```
head(uni_p, n=8)
```

```

## # A tibble: 8 x 5
##   knr   aar   Kjonn      nivaa      uni_p
##   <chr> <chr> <chr>      <chr>      <dbl>
## 1 0101  2008  Begge kjønn Universitets- og høgskolenivå, kort 17.8
## 2 0101  2009  Begge kjønn Universitets- og høgskolenivå, kort 18.2
## 3 0101  2010  Begge kjønn Universitets- og høgskolenivå, kort 18.6
## 4 0101  2011  Begge kjønn Universitets- og høgskolenivå, kort 19

```

```
## 5 0101 2012 Begge kjønn Universitets- og høghskolenivå, kort 19.6
## 6 0101 2013 Begge kjønn Universitets- og høghskolenivå, kort 19.9
## 7 0101 2014 Begge kjønn Universitets- og høghskolenivå, kort 20.6
## 8 0101 2015 Begge kjønn Universitets- og høghskolenivå, kort 21
```

## Rekoding

```
uni_p <- uni_p %>%
  mutate(
    nivaa = fct_recode(nivaa,
      "uni_k" = "Universitets- og høghskolenivå, kort",
      "uni_l" = "Universitets- og høghskolenivå, lang")
  )
```

•

```
uni_p <- uni_p %>%
  mutate(
    Kjonn = fct_recode(Kjonn,
      "mf" = "Begge kjønn",
      "f" = "Kvinner",
      "m" = "Menn")
  )
```

•

```
uni_p <- uni_p %>%
  pivot_wider(
    id_cols = c("knr", "aar"),
    names_from = c(Kjonn, nivaa),
    values_from = uni_p)
  )
```

•

## Sammenslåing

```
pm2 <- merge(pm2, uni_p)
```

```
rm(uni_p, uni_p_raw, utdanning_08_17_raw)
```

Rydder rot enda ein atter så gang

# Handelsomsetning per innbygger

## Dataauthenting, SSB.

Hvordan lager vi variabelen “Trade\_pc”

```
trade_08_17 <- ApiData(  
  urlToData = "04776",  
  Region = knr,  
  Tid = c(as.character(2008:2017)))
```

•

```
trade_pc <- tibble(  
  knr = trade_08_17$dataset$Region,  
  aar = trade_08_17$dataset$Tid,  
  value = trade_08_17$dataset$value)
```

•

## Sammenslåing

```
pm2 <- merge(pm2, trade_pc)
```

```
rm(trade_08_17, trade_pc)
```

Fjerner rot

## Avslutter med å kontrollsjekke

```
names(pm2)
```

```
## [1] "knr"          "aar"          "knavn"        "pm2"          "2008"  
## [6] "2009"        "2010"        "2011"        "2012"        "2013"  
## [11] "2014"        "2015"        "2016"        "2017"        "Menn_ya_p"  
## [16] "Kvinner_ya_p" "Totalt_ya_p" "inc_k1"      "inc_k5"      "mf_uni_k"  
## [21] "m_uni_k"     "f_uni_k"     "mf_uni_l"    "m_uni_l"     "f_uni_l"  
## [26] "value"
```

•

```
pm2 <- pm2 %>%
```

```
# Kommentar fra Kevin: lagt til "pm2" midt inni her og ser at vi da får 16 variables
```

```
select(knr, aar, knavn, pm2, Menn_ya_p, Kvinner_ya_p, Totalt_ya_p, inc_k1, inc_k5, mf_
```

Endrer navn på *value* då denne skal no heite **Trade\_pc**.

*# Kommentar fra Kevin: Endret kolonnennummer fra 15 til 16 da "value" ble flyttet lenger*

```
names(pm2)[[16]] <- "Trade_pc"
```

•

```
pm2 %>%  
  select(knr:inc_k5) %>%  
  head(n=8)
```

##	knr	aar	knavn	pm2	Menn_ya_p	Kvinner_ya_p	Totalt_ya_p	inc_k1	inc_k5
## 1	0101	2008	Halden	13427	59.74892	56.79763	58.26214	24.5	13.6
## 2	0101	2009	Halden	13095	59.77860	57.04693	58.40290	24.4	14.1
## 3	0101	2010	Halden	13832	59.64298	57.06300	58.34376	23.9	13.7
## 4	0101	2011	Halden	14915	59.84630	57.22382	58.53183	24.0	14.0
## 5	0101	2012	Halden	15473	59.45122	57.00467	58.22699	23.9	14.0
## 6	0101	2013	Halden	15461	58.97797	56.73872	57.85475	24.1	13.4
## 7	0101	2014	Halden	17164	58.76014	56.72937	57.74260	23.9	13.5
## 8	0101	2015	Halden	17427	58.71457	56.84787	57.78159	24.0	13.7

•

```
pm2 %>%  
  select(mf_uni_k:Trade_pc) %>%  
  head(n=8)
```

##	mf_uni_k	m_uni_k	f_uni_k	mf_uni_l	m_uni_l	f_uni_l	Trade_pc
## 1	17.8	15.1	20.4	3.9	5.4	2.4	56266
## 2	18.2	15.4	20.9	3.9	5.4	2.5	56366
## 3	18.6	15.6	21.6	4.1	5.5	2.7	57210
## 4	19.0	15.8	22.2	4.4	5.8	3.0	58010
## 5	19.6	16.2	22.9	4.6	5.9	3.3	58787
## 6	19.9	16.4	23.3	4.6	5.8	3.4	59453
## 7	20.6	17.0	24.0	4.9	6.1	3.8	63033
## 8	21.0	17.2	24.8	5.2	6.4	4.1	63747

## Skriver .csv fil

```
write_csv(pm2, file="pm2.csv")
```