

Assignment 4

Heidi and Ann Elisabeth

```
suppressPackageStartupMessages({
  library(PxWebApiData)
  library(tidyverse)
  library(lubridate)
})
knitr::opts_chunk$set(echo = TRUE, include = TRUE)
```

```
load("knr.Rdata")
```

Vi begynner oppgaven med å se på data som viser gjennomsnittlig kvadratmeterpris for eneboliger i perioden 2002 til 2017. Vi henter ned tabell 06035. Gjennomsnittet er regnet ut for hver kommune.

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

```
#ApiData("http://data.ssb.no/api/v0/no/table/06035")
```

```
# Et alternativ er å sette argumenter direkte på str() funksjonen
# Her har jeg spesifisert at bare de 3 første verdiene skal vises som
# eksempel, at linjene skal brytes og at maks vidde er 80 tegn
str(pm2_raw, vec.len = 3, strict.width = "wrap", width = 80)
```

```
## List of 2
## $ 06035: Selveierboliger, etter region, boligtype, statistikkvariabel og
##   år:'data.frame': 6768 obs. of 5 variables:
##   ..$ region : chr [1:6768] "Halden (-2019)" "Halden (-2019)" "Halden (-2019)"
##   ...
##   ..$ boligtype : chr [1:6768] "Eneboliger" "Eneboliger" "Eneboliger" ...
##   ..$ statistikkvariabel: chr [1:6768] "Gjennomsnittlig kvadratmeterpris (kr)"
##   "Gjennomsnittlig kvadratmeterpris (kr)" "Gjennomsnittlig kvadratmeterpris
##   (kr)" ...
##   ..$ år : chr [1:6768] "2002" "2003" "2004" ...
##   ..$ value : int [1:6768] 9070 9301 9436 10846 12052 12363 13427 13095 ...
## $ dataset :'data.frame': 6768 obs. of 5 variables:
##   ..$ Region : chr [1:6768] "0101" "0101" "0101" ...
##   ..$ Boligtype : chr [1:6768] "01" "01" "01" ...
##   ..$ ContentsCode: chr [1:6768] "KvPris" "KvPris" "KvPris" ...
##   ..$ Tid : chr [1:6768] "2002" "2003" "2004" ...
##   ..$ value : int [1:6768] 9070 9301 9436 10846 12052 12363 13427 13095 ...
```

Ser at resultatet fra ssb er en liste med 2 elementer. Element 2 er kalt dataset og er stort sett der vi henter

verdier. For pm2_raw er vi imidlertid også interessert i det første elementet der vi vil hente ut kommunenavn fra variabelen region. Se malens beskrivelse av hvordan en kan endre navn på dette elementet fra '06035: Selveierboliger, etter region, boligtype, statistikkvariabel og år:' til 'desc' for å gjøre ting enklere.

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

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

```
n=5
```

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

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

```
load("test_string_tib.Rdata")
moenster <- '\\s*\\([\\d\\s-]*\\d*\\)\\s*$'
```

```
dim(test_string_tib)
```

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

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

```
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    aar   pm2 knavn
##   <int> <int> <int> <int>
## 1      0      0  2903      0
```

Det er 2903 NA verdier i pm2.

Nå ønsker vi å se på hvor mange complete cases vi har fra 2006 til 2017

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

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

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

Det er 197 completed cases i perioden 2006 til 2017.

```
pm2_2008 <- pm2 %>%
  filter(aar >= 2008) %>%
  pivot_wider(
    names_from = aar,
    values_from = pm2
  )
```

Complete cases i 2008

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

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

Vi har 214 complete cases i 2008.

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

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

da har vi fått variablene knr inn i datasettet pm2

```
names(pm2_2008)
```

```
## [1] "knr"    "knavn"  "2008"   "2009"   "2010"   "2011"   "2012"   "2013"   "2014"
## [10] "2015"   "2016"   "2017"
```

```
rm(pm2_raw)
```

#Befolkning

Vi bruker en ny tabell som heter 07459. Her ønsker vi å finne prosentandelen av befolkningen i hver kommune som er i yrkesaktiv alder. I tabellen kan vi finne personer i alderen 20 til 64 år. Vi kan også i denne tabellen finne total befolkning i hver kommune. ut i fra det finner vi prosent i yrkesaktiv alder. det gjør vi ved å dividere de to størrelsene og multiplisere ved 100.

```
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_prefix = "sex",
    names_from = Kjonn,
    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"

```

her skal vi finne totalen for menn og kvinner

```

pop_08_17_ya <- pop_08_17_ya %>%
  mutate(ya_Total = ya_Menn+ya_Kvinner)

```

```
dim(pop_08_17_ya)
```

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

Vi får tallene 4230 og 5.

```
names(pop_08_17_ya)
```

```
## [1] "knr"          "aar"          "ya_Menn"      "ya_Kvinner"  "ya_Total"
```

ii. befolkningen totalt henting av nye data fra tabell 07459 for å finne totalen for hele befolkningen

```

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)

```

rydde i dataene

```

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

```

ordner opp i navn

```

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"
```

variablene for menn og kvinner med hensyn til alder

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

mutate for å definere nye variabler

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

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

```
dim(pop_08_17)
```

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

```
names(pop_08_17)
```

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

i. slå sammen pop_08_17ya og pop_08_17

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

nye variabler

```
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(Total_ya_p = ya_Total/Total_t*100)
```

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

```
names(pop_08_17_ya_p)
```

```
## [1] "knr"      "aar"      "Menn_ya_p" "Kvinner_ya_p" "Total_ya_p"
```

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

```
##   knr  aar Menn_ya_p Kvinner_ya_p Total_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
```

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

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

rydde opp

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

```
rm(test_string_tib)
```

#Inntektsdesiler henter inne nye data fra tabell 12558

```
inc_08_17_raw <- ApiData(  
  urlToData = "12558",  
  Region = knr,  
  Desiler = c("01", "02", "09", "10"),  
  ContentsCode = "AndelHush",  
  InntektSkatt = "00",  
  Tid = c(  
    as.character(2008:2017))  
)$dataset %>%  
  select(Region, Desiler, Tid, value)
```

```
inc_08_17 <- inc_08_17_raw %>%  
  pivot_wider(  
    names_from = Desiler,  
    values_from = value  
  )
```

```
names(inc_08_17)[[1]] <- "knr"  
names(inc_08_17)[[2]] <- "aar"  
names(inc_08_17)[[3]] <- "Desil_1"  
names(inc_08_17)[[4]] <- "Desil_2"  
names(inc_08_17)[[5]] <- "Desil_9"  
names(inc_08_17)[[6]] <- "Desil_10"
```

```
inc_08_17 <- inc_08_17 %>%  
  mutate(inc_k1 = Desil_1 + Desil_2) %>%  
  mutate(inc_k5 = Desil_9 + Desil_10)
```

```
inc_08_17 <- inc_08_17 %>%  
  select(knr, aar, inc_k1, inc_k5)
```

```
names(inc_08_17)
```

```
## [1] "knr"      "aar"      "inc_k1"   "inc_k5"
```

```
dim(inc_08_17)
```

```
## [1] 4230      4
```

Vi ender opp med 4230 og 4.

```
pm2 <- merge(pm2, inc_08_17)
```

```
rm(inc_08_17, inc_08_17_raw, pop_08_17_ya_p)
```

#Prosent av befolkning med universitets/høgskole utdanning

her henter vi først inn en ny tabell, med nye data

```
uni_p_raw <- ApiData(  
  urlToData = "09429",  
  Region = knr,  
  Nivaa = c("03a", "04a"),
```

```

Kjonn = TRUE,
ContentsCode = "PersonerProsent",
Tid = c(
  as.character(2008:2017)
)
)

uni_p <- uni_p_raw

names(uni_p)[[1]] <- "desc"

uni_p <- tibble(
  knr = uni_p$dataset$Region,
  aar = uni_p$dataset$Tid,
  # Virker dårlig med pdf output, går langt utenfor margen
  # har derfor skiftet navn til desc ovenfor
  # Kjonn = uni_p$`09429: Personer 16 år og over,
  # etter region, nivå, kjønn, statistikkvariabel og år`$kjønn,
  # nivaa = uni_p$`09429: Personer 16 år og over, etter region,
  # nivå, kjønn, statistikkvariabel og år`$nivå,
  Kjonn = uni_p$desc$kjønn,
  nivaa = uni_p$desc$nivå,
  uni_p = uni_p$dataset$value
)

head(uni_p, 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 høghskolenivå, kort 17.8
## 2 0101  2009  Begge kjønn Universitets- og høghskolenivå, kort 18.2
## 3 0101  2010  Begge kjønn Universitets- og høghskolenivå, kort 18.6
## 4 0101  2011  Begge kjønn Universitets- og høghskolenivå, kort 19
## 5 0101  2012  Begge kjønn Universitets- og høghskolenivå, kort 19.6

i. rekode nivåene for variabelen nivaa til uni_k og uni_l

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(nivaa, Kjonn),
    values_from = uni_p
  )

```

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

```
## # A tibble: 8 x 8
##   knr   aar   uni_k_mf uni_k_m uni_k_f uni_l_mf uni_l_m uni_l_f
##   <chr> <chr>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 0101  2008     17.8    15.1    20.4     3.9     5.4     2.4
## 2 0101  2009     18.2    15.4    20.9     3.9     5.4     2.5
## 3 0101  2010     18.6    15.6    21.6     4.1     5.5     2.7
## 4 0101  2011     19      15.8    22.2     4.4     5.8     3
## 5 0101  2012     19.6    16.2    22.9     4.6     5.9     3.3
## 6 0101  2013     19.9    16.4    23.3     4.6     5.8     3.4
## 7 0101  2014     20.6    17      24       4.9     6.1     3.8
## 8 0101  2015     21      17.2    24.8     5.2     6.4     4.1

```

```
dim(uni_p)
```

```
## [1] 4230    8
```

Flettes sammen:

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

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

#Handelsomsetning pr innbygger

henter inn nye data fra tabell 04776

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

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

```
Trade_pc <- trade_08_17
```

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

Slette de objektene som ikke lenger trengs

```
rm(trade_08_17, Trade_pc)
```

```
dim(pm2)
```

```
## [1] 2140    16
```

```
names(pm2)
```



```
## [1] "knr"          "aar"          "knavn"        "pm2"          "Menn_ya_p"
## [6] "Kvinner_ya_p" "Total_ya_p"   "inc_k1"       "inc_k5"       "uni_k_mf"
## [11] "uni_k_m"      "uni_k_f"      "uni_l_mf"     "uni_l_m"      "uni_l_f"
## [16] "Trade_p"
```

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

```
##   knr  aar  knavn  pm2 Menn_ya_p Kvinner_ya_p Total_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(knr:inc_k5) %>%
  head(n=8)
```

```
##   knr  aar  knavn  pm2 Menn_ya_p Kvinner_ya_p Total_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(uni_k_mf:Trade_p) %>%
  head(n=8)
```

```
##   uni_k_mf uni_k_m uni_k_f uni_l_mf uni_l_m uni_l_f Trade_p
## 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
```

```
# Bruk tidyverse versjonen( "_" istedenfor "."). Den er bedre
write_csv(pm2, "pm2.csv")
```