

R Notebook

```
suppressPackageStartupMessages({
  library(PxWebApiData)
  library(tidyverse)
  library(lubridate)
  library(dplyr)
  library(stringr)
})
#knitr::opts_chunk$set(echo=FALSE, include = FALSE)
```

```
#Gjennomsnittlig kvadratmeterpris
```

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

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

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

```
pm2 <- pm2_raw$dataset %>%
  #NAstatus ikke definert ovenfor
  select(-Boligtype, -ContentsCode) %>%
  rename(
    knr = Region,
    aar = Tid,
    pm2 = value) %>%
  mutate(
    knavn = pm2_raw$desc$region
  )
```

```
load("test_string_tib.Rdata")
# Legg inn regex mønster
moenster <- '\\s*\\([\\d\\s-]*\\d*\\)\\s*$'
test_string_tib %>%
  mutate(
    knavn = str_replace(knavn, moenster, "")
  )
```

```
## # A tibble: 5 x 1
##   knavn
##   <chr>
## 1 Halden
## 2 Sarpsborg
## 3 Våler (Østfold)
```

```
## 4 Haugesund
## 5 Porsanger - Porsángu - Porsanki
```

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

```
pm2%>%
  summarise(
    across(
      everything(),
      ~ sum(is.na(.))
    )
  )
```

```
##   knr aar  pm2 knavn
## 1   0   0 2903     0
```

```
# Alternativ måte å gjøre det samme på, men summarise across med bruk av anonym funksjon er
# jo temmelig fancy i forhold til min ;-)
sum(is.na(pm2))
```

```
## [1] 2903
```

```
pm2fra2006 <- pm2 %>%
  filter(aar >= 2006) %>%
  pivot_wider(
    names_from = aar,
    values_from = pm2
  )
```

Complete cases i perioden 2006-2017:

```
pm2fra2006 %>%
  complete.cases() %>%
  sum()
```

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

```
pm2fra2008 <- pm2 %>%
  filter(aar >= 2008) %>%
  pivot_wider(
    names_from = aar,
    values_from = pm2
  )
```

Complete cases i perioden 2008-2017:

```
pm2fra2008 %>%
  complete.cases() %>%
  sum()
```

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

```
# Virker, men kanskje unødvendig komplisert
# pm2 <- pm2 %>%
# filter(aar >= 2008) %>%
# left_join(pm2fra2008,
#           by = c("knr", "knavn")) %>%
```

```

# na.omit(pm2)
#
#Mitt forslag. Husk at & betyr og
pm2 <- pm2 %>%
  filter (
    aar >= 2008 & complete.cases(.) == TRUE
  )

# Time to clean up
rm(test_string_tib, pm2_raw, pm2fra2006, pm2fra2008)

#Befolkning, hente data fra SSB
# Her vil jeg foreslå å ikke hente ut dataset-biten direkte siden kjønnskoding er litt usikker
# Er 1, mann eller kvinne?
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))
)

# Bearbeiding av befolkning
tmp <- pop_08_17_ya_raw$dataset %>%
  mutate(
    #henter kjønn fra første listen (beskrivelse listen). Bruker [[1]] for å unngå det lange navnet
    Kjonn = pop_08_17_ya_raw[[1]]$kjønn
  ) %>%
  select(
    -Alder,
    -ContentsCode
  )

pop_08_17_ya_raw <- tmp
rm(tmp)

pop_08_17_ya <- pop_08_17_ya_raw %>%
  pivot_wider(
    id_cols = c(Region, Tid),
    names_from = Kjonn,
    values_from = value
  ) %>%
  rename(
    knr = Region,
    aar = Tid,
    ya_menn = Menn,
    ya_kvinner = Kvinner
  )

# Meget lettere vha tidyverse, se ovenfor
# 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
  )

# unødvendig, navnet er satt i mutate
# names(pop_08_17_ya)[[5]] <- "ya_total"

dim(pop_08_17_ya)

## [1] 4230    5

names(pop_08_17_ya)

## [1] "knr"          "aar"          "ya_menn"      "ya_kvinner"  "ya_total"

# Igjen ville jeg her ikke tatt dataset direkte slik at jeg fikk tilgang til kjønn i første liste element
pop_08_17_raw <- ApiData (
  urlToData = "07459",
  Region = knr,
  Kjønn = c(1,2),
  Alder = list("agg:TodeltGrupperingB",
               c("H17", "H18")),
  Tid = c(as.character(2008:2017))
)

tmp <- pop_08_17_raw$dataset %>%
  mutate(
    Kjønn = pop_08_17_raw[[1]]$kjønn
  ) %>%
  rename(
    knr = Region,
    alder = Alder,
    aar = Tid,
  ) %>%
  select(-ContentsCode)

pop_08_17_raw <- tmp
rm(tmp)

# Vil anbefale å først rename variablene før en gjør pivot_wider
pop_08_17 <- pop_08_17_raw %>%
  pivot_wider(
    names_from = Kjønn,
    values_from = value
  ) %>%
  pivot_wider(
    names_from = alder,
    values_from = c(Menn, Kvinner)
  ) %>%
  mutate(
    menn = Menn_H17 + Menn_H18,
    kvinner = Kvinner_H17 + Kvinner_H18,
    totalt = menn + kvinner
  )

```

```

) %>%
select(knr, aar, menn, kvinner, totalt)

# igjen rename i pipe som ovenfor mye lettere og også klarere kode
# 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"

ya_prosent <- cbind(pop_08_17[,1:2], (pop_08_17_ya[,3:5]/pop_08_17[,3:5])*100)

# Mye fin kode i de neste chunk-ene, men helt unødvendig. Alt dere
# trenger ligger i ya_prosent ovenfor
# Dette er gjort ovenfor i chunk-en: utvidelse av hele befolkning
# pop_08_17 <- pop_08_17 %>%
#   mutate(kvinner_t = kvinner_H17 + kvinner_H18,
#          menn_t = menn_H17 + menn_H18) %>%
#   mutate(totalt_t = kvinner_t + menn_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"     "kvinner" "totalt"

# pop_08_17_ya_p <- pop_08_17_ya %>%
#   left_join(pop_08_17,
#             by = c("knr", "aar"))

# pop_08_17_ya_p <- pop_08_17_ya_p %>%
#   mutate(kvinner_ya_p = (ya_kvinner/kvinner_t)*100,
#          menn_ya_p = (ya_menn/menn_t)*100,
#          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)

# rm(pop_08_17_raw, pop_08_17_ya_raw, pop_08_17, pop_08_17_ya)

# dere har det dere trenger i ya_prosent
# gjenstår bare å joine pm2 med ya_prosent
pm2 <- pm2 %>%
  inner_join(ya_prosent,
            by = c("knr", "aar"))

#Inntektsdesiler
inc_k_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
```

```
inc_k_08_17 <- inc_k_08_17_raw %>%
  pivot_wider(
    names_from = Desiler,
    # R liker ikke variabelnavn som starter med tall, derfor "d"
    names_prefix = "d",
    values_from = value
  )
```

```
# Her legger dere bare inn faste verdier. Det går galt
inc_k_08_17 <- inc_k_08_17 %>%
  mutate(
    #Nå bør det virke
    inc_k1 = d01 + d02,
    inc_k5 = d09 + d10
  )
```

```
inc_k_08_17 <- inc_k_08_17 %>%
  select(Region, Tid, inc_k1, inc_k5)
```

```
# Langt enklere å bruke rename() i en pipe
names(inc_k_08_17)[[1]] <- "knr"
names(inc_k_08_17)[[2]] <- "aar"
```

```
names(inc_k_08_17)
```

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

```
dim(inc_k_08_17)
```

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

```
pm2 <- pm2 %>%
  inner_join(inc_k_08_17,
    by = c("knr", "aar"))
```

```
rm(inc_k_08_17_raw)
```

```
#Befolkning med utdanning
```

```
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
```

```
uni_p <- tibble(
  knr = uni_p$dataset$Region,
  aar = uni_p$dataset$Tid,
  Kjonn = uni_p$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikkvariabel og år`$kj
```

```
nivaa = uni_p$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikkvariabel og år`$nivaa
uni_p = uni_p$dataset$value
)
```

```
uni_p <- uni_p[, c("knr", "Kjonn", "nivaa", "uni_p", "aar")]
```

```
names(uni_p)
```

```
## [1] "knr" "Kjonn" "nivaa" "uni_p" "aar"
```

```
dim(uni_p)
```

```
## [1] 25380 5
```

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

```
## # A tibble: 5 x 5
```

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

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

```
uni_p <- uni_p %>%
  mutate(
    Kjonn = fct_recode (Kjonn, mf = "Begge kjønn",
                       f = "Kvinner",
                       m = "Menn")
  ) #Får feilmelding på denne hele tiden, inkludert når jeg kjører alle chunkene på nytt, men ser ut so
```

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

```
pm2 <- pm2 %>%
  inner_join(uni_p,
    by = c("knr", "aar"))
```

```
rm(uni_p_raw)
```

```
#Handelsomsetning
```

```
trade_08_17_raw <- ApiData (
  urlToData = "04776",
  Region = knr,
  ContentsCode = "OmsPerInnb",
  Tid = c(as.character(2008:2017))
)
```

```

)

trade_08_17 <- trade_08_17_raw

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

names(trade_08_17)

## [1] "knr"      "aar"      "trade_pc"
dim(trade_08_17)

## [1] 4230      3
head(trade_08_17, n = 5)

## # A tibble: 5 x 3
##   knr    aar  trade_pc
##   <chr> <chr>    <int>
## 1 0101  2008     56266
## 2 0101  2009     56366
## 3 0101  2010     57210
## 4 0101  2011     58010
## 5 0101  2012     58787

pm2 <- pm2 %>%
  inner_join(trade_08_17,
    by = c("knr", "aar"))

rm(trade_08_17_raw)

#Ferdig
# Neste er kommentert ut. Gir ca. 100 sider med dataautskrift
#na.omit(pm2)
dim(pm2)

## [1] 2665      16

#Ferdig
dim(pm2)

## [1] 2665      16

pm2 <- pm2 %>% select (knr,knavn,aar,pm2,ya_menn,ya_kvinner,ya_total,inc_k1,inc_k5,uni_k_mf,uni_k_m,uni.

dim(pm2)

## [1] 2665      16

# Bruk relative stier til filer. Ellers virker det ikke på andre maskiner
# Bruk heller write_csv() fra tidyverse. Det er en bedre funksjon.
write_csv(pm2, "pm2.csv")

```