

hent-ssb-data

Ingrid-Liv Morkken

Thibiga Kuddyar

```
knitr::opts_chunk$set(eval=TRUE)

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

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

```
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
  )
```

```
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
```

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

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

```
select(knr, knavn, aar, pm2)
```

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

```
moenster <- '\\s*\\([\\d\\s-]*\\d*\\)\\s*$'
```

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

```
pm2 %>%  
  map_df(is.na) %>%  
  map_df(sum) %>%  
  # as.tibble() har gått ut på dato  
  # nå heter den as_tibble()  
  # as.tibble()  
  as_tibble()
```

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

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

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

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

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

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

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

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

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

```
# time to clean up  
rm(pm2_raw)
```

Befolkning

```
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"
```

```
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"
```

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

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/Totalt_t*100)

pop_08_17_ya_p <- pop_08_17_ya_p %>%
  select(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)

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

```

Intektsdesiler

```

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
pm2 <- merge(pm2, inc_08_17)

rm(inc_08_17, inc_08_17_raw, pop_08_17_ya_p)

```

Prosent av befolkning med universitet/høgskolen 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

# Jeg vil foreslå å endre navn på første komponent i uni_p så slipper en
# tekst-strenger som går langt ut i margin. Disse navnene til SSB er ikke særlig
# R vennlige

```

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

```
# bruker desc for å referere til første element i uni_p listen
# slipper uendelig lange navn som går utenfor margen
```

```
uni_p <- tibble(
  knr = uni_p$dataset$Region,
  aar = uni_p$dataset$Tid,
  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 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 %>%
  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"
    )
  )
```

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

rm(uni_p, uni_p_raw)

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)

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 Kvinne_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 write_csv fra tidyverse  
write_csv(pm2, "pm2.csv")
```