Assignment 4

Assignment 4 i kurset Data Science 2021

Karoline Midtbø Morten Knutsen

Oppgvave 4

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

##Loader inn data

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

I denne delen skal vi se på prisen per kvm.

```
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, aar, pm2, knavn)
```

I denne delen har vi valgt ut hva vi skal ha som variabler, der vi har valt vekk *boligtyper* og *contentscode*.

```
load("test_string_tib.Rdata")
# Legg inn regex mønster
moenster <- '\\s*\\([\\d\\s-]*\\d*\\)\\s*$'
```

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

```
## # A tibble: 6,768 x 4
## # Groups:   knr [423]
##    knr   aar     pm2 knavn
##    <chr> <chr> <int> <chr>
##  1 0101  2002   9070 Halden
##  2 0101  2003   9301 Halden
##  3 0101  2004   9436 Halden
##  4 0101  2005  10846 Halden
##  5 0101  2006  12052 Halden
##  6 0101  2007  12363 Halden
##  7 0101  2008  13427 Halden
##  8 0101  2009  13095 Halden
##  9 0101  2010  13832 Halden
## 10 0101  2011  14915 Halden
## # ... with 6,758 more rows
```

Sjekke hvor mange NA verdier det er i 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   aar   pm2 knavn
##   <int> <int> <int> <int>
## 1     0     0  2903     0
```

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

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

Complete.cases fra 2006 til 2017

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

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

Det er 197 complete cases i fra 2016 til 2017, det vil si antall kommuner som har data for alle årene f.o.m. 2006 t.o.m. 2017

Complete.cases fra 2008 til 2017

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

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

Det er 214 complete cases i fra 2008 til 2017, som vil si 214 kommuner om har data for alle årene f.o.m. 2008 t.o.m. 201

Vi ser at ved å velge perioden 2008-2017 istedenfor 2006-2017 får vi 17 ekstra complete.cases. Velger derfor å studere perioden 2008-2017.

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

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

Fjerner data

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

Datasett henting fra 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
```

sjekker navn

```
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(kvinner, menn)
 )
```

```
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)
```

Henter data fra SSB om 0-17 0g 18+

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

```
names(pop_08_17)
```

```
## [1] "knr"       "aar"       "menn_t"    "kvinner_t" "totalt_t"
```

merge data til pop_08_17_ya_p

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

Legger sammen navnene og finner dem i prosent

```
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_kvinner/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

```
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    28.61313
## 2 0101 2009  59.77860     57.04693    28.72944
## 3 0101 2010  59.64298     57.06300    28.73575
## 4 0101 2011  59.84630     57.22382    28.68241
## 5 0101 2012  59.45122     57.00467    28.52452
```

Merger data til pm2

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

```r
pm2 <- pm2 %>%
  select(knr, knavn, aar, pm2, Menn_ya_p, kvinner_ya_p, totalt_ya_p)
```

Rydder opp

```r
rm( pop_08_17_raw, pop_08_17_ya_raw, pm2_2006, pm2_2008)
```

```r
rm(test_string_tib,pop_08_17_ya)
```

Desiler

Vi henter inn data fordelt opp i desiler

```r
inc_08_17_raw <- ApiData(
  urlToData = "12558",
  Region = knr,
  #  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)
```

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

```r
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"
```

```r
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)
```

Rydder opp

```
rm(inc_08_17,inc_08_17_raw, pop_08_17_ya_p, pop_08_17)
```

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$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikkvar
  nivaa = uni_p$`09429: Personer 16 år og over, etter region, nivå, kjønn, statistikkvar
  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
```

```
dim(uni_p)
```

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

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

```
rm(pop_08_17, uni_p, uni_p_raw)
```

```
## Warning in rm(pop_08_17, uni_p, uni_p_raw): object 'pop_08_17' not found
```

Handelsomsetning per innbygger

```
trade_08_17 <- ApiData (
  urlToData = "04776",
  Region = knr,
  Tid = c(as.character(2008:2017))
) $dataset %>%
  select(-ContentsCode)
```

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

```
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" "totalt_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 totalt_ya_p inc_k1
## 1 0101 2008 Halden (-2019) 13427  59.74892     56.79763    28.61313   24.5
## 2 0101 2009 Halden (-2019) 13095  59.77860     57.04693    28.72944   24.4
## 3 0101 2010 Halden (-2019) 13832  59.64298     57.06300    28.73575   23.9
## 4 0101 2011 Halden (-2019) 14915  59.84630     57.22382    28.68241   24.0
## 5 0101 2012 Halden (-2019) 15473  59.45122     57.00467    28.52452   23.9
## 6 0101 2013 Halden (-2019) 15461  58.97797     56.73872    28.46051   24.1
## 7 0101 2014 Halden (-2019) 17164  58.76014     56.72937    28.42493   23.9
## 8 0101 2015 Halden (-2019) 17427  58.71457     56.84787    28.41269   24.0
##   inc_k5
## 1   13.6
## 2   14.1
```

```
## 3   13.7
## 4   14.0
## 5   14.0
## 6   13.4
## 7   13.5
## 8   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
```

```
#Bruke heller write_csv() fra tidyverse
write_csv(pm2, "pm2.csv")
```

```
#siste
```