

Assignment 4, 2021

I denne siste oppgaven skal vi studere nærmere pris per kvadratmeter for eneboliger i et utvalg norske kommuner. Vi vil studere tidsperioden 2008 til 2017. Utvalget av kommuner vil være i overkant av 200 og betsår av de kommunene som har registrerte data for hele perioden. Dette betyr i praksis at de mindre kommune er utelatt.

For å forklare prisforskjellene mellom kommunene vil vi benytte et sett variabler hentet direkte fra SSB via en api. Denne api-en er implementert i R-pakken `PxWebApiData`. Denne pakken, `tidyverse` og `lubridate` må lastes i en chunk kalt `setup` i starten av dokumentet `hent-ssb-data.Rmd`. Vi vil hente dataene og gjøre datasettet *tidy* i samme dokument da det synes å være mest hensiktsmessig i dette tilfellet.

Arbeidet skal utføres i et RStudio-prosjekt knyttet til github på vanlig måte. Kall prosjektet for `ass4-h21-initialerStud1-initialerStud2`. I prosjektet skal det være en mappe kalt `data`. I denne mappen skal det være en R notebook kalt `hent-ssb-data.Rmd` der koden for å hente data fra SSB og å bygge et *tidy* datasett skal skrives. Som hjelp har jeg lagt ut en fil `mal-hent-ssb-data.pdf` med beskrivelse av ting som må gjøres og sjekk-punkter undervegs så dere kan sjekke at dere er på rett spor. Vi skal til slutt ende opp med en *tidy* tibble med 16 kolonner og 2140 rekker. Denne skal vi lagre som datafilen `pm2.csv` i mappen `data`.

Dokumentet `hent-ssb-data.Rmd` skal skrives slik at prosessen med å hente data-ene er reproducerbar. I teorien kan dataene hos kilden endre seg. Før vi skriver ut det endelige datasettet vi vil bruke legger vi inn en kommentar om datoen filen ble generert og legger inn opsjonen

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

i den første kode-chunken kalt `setup`. Dette sikrer at vi ikke vil kjøre dokumentet på nytt, ved et uhell, og generere ny datafil.

Når datasettet er på plass flytter vi opp fra datamappen. Her skal vi ha et dokument kalt `model.Rmd` der vi skal foreta selve analysen. Vi starter med å lese inn datafilen `pm2.csv` vha. kommandoen `read_csv()`.

```
pm2 <- read_csv("data/pm2.csv", show_col_types = FALSE)
```

igjen har jeg skrevet et dokument `mal-model.pdf` som prøver å lede dere igjennom prosessen. Se hvilke `libraries` som trengs i starten av dette dokumentet. Dere vil særlig få bruk for funksjoner dekket i kapitlene 5 (`select`, `filter` etc.), 12 (`pivot_wider` etc.), 19 (`map_*` etc), og 25 (`nest`, `unnest`, `broom::glance` etc.) og også 3 (`ggplot` etc.). Det siste for å lage flotte plot av resultatene.

Oppgaven favner vidt så skulle dere ha behov for litt oppfriskning foreslår jeg å ta en titt i slides først for så eventuelt å sjekke læreboken (`r4ds`).

Skulle dere kjøre dere fast svarer jeg som vanlig på spørsmål. Gir dere meg tilgang til github repository-en er det enklere for meg å gjøre små *fixes*. Prøv å skrive klar og fin kode og rydd opp undervegs (dvs slett objekter som ikke lenger trengs). Det gjør det lettere for meg å hjelpe.

Lykke til!

Siste frist for levering kommer som kunngjøring på Canvas.