

Assignment 4

Assignment 4 i kurset Data Science 2021

Karoline Midtbø

Morten Knutsen

```
suppressPackageStartupMessages({  
library(tidyverse)  
library(lubridate)  
library(modelr)  
library(broom)  
library(lmtest)  
library(sandwich)  
library(viridis)  
})
```

```
knitr::opts_chunk$set(echo = FALSE, include = FALSE)
```

Legger til pm2

i

i.

Vi ser at `aar_f2009` ikke er signifikant. Alle de andre årene er signifikante på et 0.1% nivå..
Vi ser at koeffisientene fra `aar_f2009` og frem til `aar_f2017` øker fra år til år.

ii.

godt spørsmål

heteroskedastisitet

i.

ii.

Ja, vi har problemer med heteroskedas

iii.

iv.

Dette er gjort i tidligere oppgave.

v.

iv.

vii til x.

Finner gjennomsnittet til hvert fylke hvert år.

Dummy fylke og år

i og ii.

iii.

iv.

DEL-PLOT

i + ii

I modell 2 er ikke kvaliteten helt optimal, fordi den mangler noen variabler. Det kan skyldes heteroskedastisitet at det er en stor variasjon. Rogaland er et stort residual.

iii

Mangler noen farger for kommunen.

i

ii

De kommunene som ligger nærmest Stavanger overvurderes.

Modell for hvert år

i)

ii)

Siste delen mangler

får ikke til siste del.

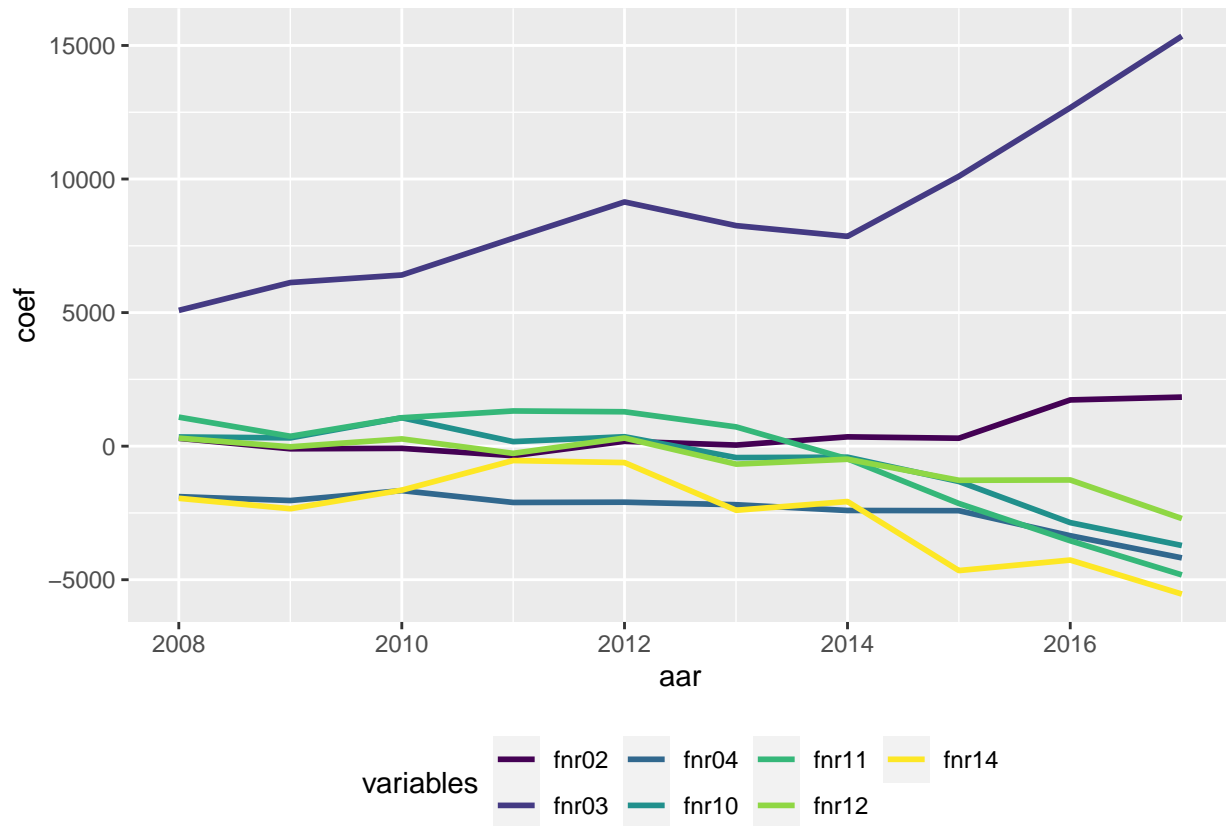
Her er de nødvendige besvergelsene

```
coef_df <- mod_sum$model %>%  
  # .f = 1 gir oss første element i liste model i mod_sum  
  # .f = 2 gir oss andre elementet som er residualene  
  # ikke så intuitivt derfor hint ;-)  
  map_df(.f = 1) %>%  
  tibble()
```

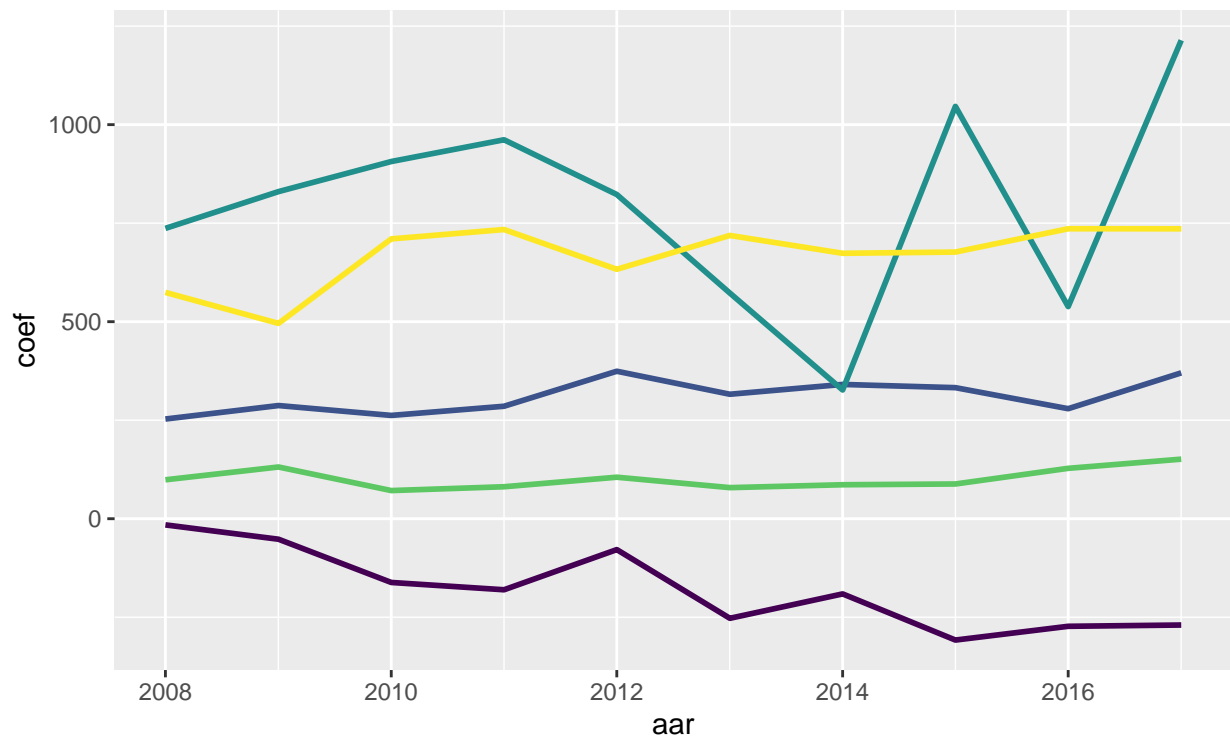
- i. Lag en ny variabel av type date i `coef_df` som angir år.
- ii. Pivot_longer `coef_df` til `coef_df_long`.
- iii. Lag så et plot av fylke-faktorvariablenes koeffisienter for fylkene “fnr02”, “fnr03”, “fnr04”, “fnr10”, “fnr11”, “fnr12”, “fnr14” fra år 2008 til 2017.
- iv. Hva sier plot-et oss om prisutviklingen i disse fylkene?
- v. Hva skjedde i 2014?

Her er kode for å lage plottet:

```
coef_df_long %>%  
  select(aar, variables, coef) %>%  
  filter(  
    variables %in% c("fnr02", "fnr03", "fnr04", "fnr10", "fnr11", "fnr12", "fnr14")  
  ) %>%  
  ggplot(mapping = aes(x = aar, y = coef, colour = variables)) +  
  scale_color_viridis(discrete = TRUE, option = "D") +  
  geom_line(aes(group = variables), lwd = 1) +  
  theme(legend.position = 'bottom')
```



- Lag et plot tilsvarende det ovenfor for fnr, men nå for variablene Total_ya_p, inc_k1, inc_k5, uni_k_mf, uni_l_mf og Trade_pc_100K. (Plottet er gjengitt nedenfor, dere skal gjenskape det vha ggplot)
- Diskuter om koeffisientene ser ut til å være stabile over tid.



variables inc_k1 inc_k5 Trade_pc_100K uni_k_mf uni_l_mf