

Instruksjoner

Denne oppgaven skal IC8ses interaktivt i RStudio ved C% legge inn egen kode og kommentarer. Det ferdige dokumentet lagres med kandidatnummeret som navn [kandidatnummer]_SOK1004_C1_H22.qmd og lastes opp pC% deres GitHub-side. Hvis du har kandidatnummer 43, sC% vil filen hete 43_SOK1004_C1_H22.qmd. PC%se at koden kjC8rer og at dere kan eksportere besvarelsen til pdf. Dere leverer lenken til GitHub-repositoriet i Canvas.

Bakgrunn

Vi skal analysere utviklingen i bruttonasjonalprodukt (BNP) per person i Norge. Vi bruker data Statistisk SentralbyrC% (SSB), tabell "09842: BNP og andre hovedstC8rrelser (kr per innbygger), etter statistikkvariabel og C%r". Tabellen inneholder C%rlige data pC% BNP per innbygger, fra 1970 til 2021.

I. API, visualisering

SSB gir oss tilgang til sine data via en [API](#) (*Application Programming Interface*), programvare som lar to applikasjoner kommunisere med hverandre. SSB tilbyr en API medB [ferdige datasett](#). Her er det om lag 250 kontinuerlig oppdaterte datasett med en fast URL over de mest brukte tabellene i Statistikkbanken.

For C% fC% tilgang til tabellen med bruttonasjonalprodukt mC% vi benytte tjenesten [PxWebApi](#). Her finner du en [API konsoll](#) med en sC8kefunksjon. PrC8v C% sC8k pC% "bnp" og merk forslaget: tabell 09842. SC8k pC% denne, og noter URL-en. Den vil vi bruke etterpC%.

Til C% laste ned dataene skal vi bruke en R-pakke, [PxWebApiData](#), som SSB har laget. I fC8rste omgang skal vi bruke funksjonen ApiData(). Syntaksen er ikke den samme som i t idyverse, og har noen litt uvante egenskaper, herunder lagring i regnformat og en kombinasjon av norsk og engelsk.

Tips: Det er typisk instruktivt C% se pC% [eksempel pC% bruk](#). Da har man et intuitivt utgangspunkt for hvordan koden kan brukes.

Jeg vil nC% vise dere trinnviss hvordan C% laste ned dataene. FormC%let er C% gi dere en idC) pC% hvordan man kan IC&re seg C% bruke en ny pakke eller funksjon. Vi begynner med C% laste inn nC8dvendige pakker:

```
rm(list=ls())
library(tidyverse)

# Attaching packages
# ggplot2 3.3.6      # purrr   0.3.4
# tidbbl  3.1.8      # dplyr   1.0.10
# tidyr   1.2.1      # stringr 1.4.1
# readr   2.1.2      # forcats 0.5.2

# Conflicts:
# dplyr::filter() masks stats::filter()
# dplyr::lag()     masks stats::lag()

library(PxWebApiData)

NB! Du mC% installere PxWebApiData fC8rst. KjC8r kommandoen install.packages("PxWebApiData") i konsollen. Det mC% kun gjC8res C)n gang.

Vi bruker funksjonen ApiData() til C% hente tabell 09842. Som notert ovenfor fant vi URL-en ved hjelp av sC8kefunksjonen til SSB. FC8rst prC8ver vi C% laste ned dataene direkte, uten ytterligere tilvalg, og tar en titt pC% hva vi fC%r.

lenke <- "http://data.ssb.no/api/v0/no/table/09842"

df <- lenke %>%
  ApiData()

df %>%
  print()

$`09842: BNP og andre hovedstørrelser (kr per innbygger), etter statistikkvariabel og år`
  statistikkvariabel  år value
1 Bruttonasjonalprodukt 1970 23616
2 Bruttonasjonalprodukt 2020 633965
3 Bruttonasjonalprodukt 2021 765836
4 Konsum i husholdninger og ideelle organisasjoner 1970 12283
5 Konsum i husholdninger og ideelle organisasjoner 2020 278844
6 Konsum i husholdninger og ideelle organisasjoner 2021 298804
7 MEMO: Bruttonasjonalprodukt. Faste 2015-priser 1970 214756
8 MEMO: Bruttonasjonalprodukt. Faste 2015-priser 2020 604951
9 MEMO: Bruttonasjonalprodukt. Faste 2015-priser 2021 625077

$dataset
  ContentsCode Tid value
1 BNP 1970 23616
2 BNP 2020 633965
3 BNP 2021 765836
4 KonsumHIO 1970 12283
5 KonsumHIO 2020 278844
6 KonsumHIO 2021 298804
7 MEMOBNP 1970 214756
8 MEMOBNP 2020 604951
9 MEMOBNP 2021 625077
```

Merk fC8lgende: df inneholder to datasett i formatet data.frame. Datasettene heter "09842: BNP og andre hovedstC8rrelser (kr per innbygger), etter statistikkvariabel og C%r" og dataset. Datasettene inneholder 9 verdier av 3 variabler. Variabelen value er identisk. Variablene C%r og Tid inneholder de identiske verdiene "1970", "2020" og "2020". Merk at disse er i tegnformat <chr> (derav anfC8rselstegnene) og ikke en numerisk verdi, for eksempel <dbl>. Variabelen statistikkvariabel og ContentsCode inneholder henholdsvis verdiene BNP, KonsumHIO MEMOBNP og Bruttonasjonalprodukt, Konsum i husholdninger og ideelle organisasjoner og MEMO: Bruttonasjonalprodukt. Faste 2015-priser.

Vi har altsC% ikke fC%tt hele tabell 09842, men verdiene for tre statistikkvariabler over tre tidsperioder, lagret med forskjellige variabelnavn og verdier.

Det vi trenger er **metadata**: Informasjon som beskriver innholdet i dataene, slik at vi kan filtrere API-spC8rringen. KjC8r fC8lgende kode.

```
metadata <- lenke %>%
  ApiData(returnMetadata = TRUE)

C pner vi listen metadata fra minnet sC% kan vi se nC&rmere pC% den i C8vre venstre vindu i Rstudio. Her ser vi to lister kalt [[1]] og [[2]]. Listene beskriver variablene vi kan filtrere pC%. Liste [[1]] har fire variable: code, text, values, og valueTexts. Alle variablene er <chr>. Liste [[2]] har de samme foregC%ende fire variablene samt en variabel time.
```

- code viser navnene pC% variablene vi bruker i funksjonen ApiData() for C% filtrere. Den tar verdiene ContentsCode og Tid. Legg merke til at utviklerne i SSB her blander norsk og engelsk.
- text er en unik tekstverdi tilknyttet verdien pC% code som forklarer hva vi ser pC%. Den tar verdien statistikkvariabel og C%r. Vi kan altsC% filtrere pC% statistikkvariabel og C%r.
- values viser hvilke verdier av statistikkvariabel og C%r vi kan velge, med henholdsvis 6 og 52 forskjellige verdier. Du vil kjenne igjen tre av hver fra den fC8rste spC8rringen ovenfor.
- valueTexts gir en unik tekstverdi tilknyttet verdien pC% values som forklarer oss hva vi ser pC%. For Tid og C%r er de identiske, men for ContentsCode og statistikkvariabel fC%r vi en mer fullstendig forklaring.
- time er en logisk variabel, og tar derfor to verdier: TRUE og FALSE. I dette tilfellet indikerer den at variabelen Tid mC%ler tid, hvilket gjC8r at funksjonene i pakken vil behandle Tid pC% en annen mC%te enn en statistikkvariabel.

Vi har nC% informasjonen vi trenger til C% laste ned BNP-tall mellom 1970 og 2021. Jeg velger C% ta BNP med bC%de IC8pende og faste priser.

```
df <- lenke %>%
  ApiData(Tid = paste(1970:2021), ContentsCode = c("BNP", "MEMOBNP"))
```

PC% venstre side av likhetstegnet bruker vi code fra metadata. PC% hC8yre side velger vi verdier fra values. Merk at jeg bruker funksjonen paste() for C% konvertere numeriske verdier, for eksempel <dbl> til tegn <chr>.

La oss rydde i data. Det er tre ting C% ta tak i:

- df lagrer informasjonen i to tabeller med samme informasjon, som vist over. Det er unC8dvendig.
- C rstallene er lagret som tegn, <chr>. Disse skulle heller vC&re heltall, <int>.
- Formatet data.frame er underlegent tidbbl.

Oppgave Ia: Rydd i data

Skriv kode som lagrer dataene som C)n tidbbl med anstendige variabelnavn og C%rstall som heltall. Fremover bruker jeg "var", "tid", og "verdi" for "statistikkvariabel", "Tid", og "value".

```
# Oppgave Ia IC8ses her
df <- df[2] $dataset
df$Tid <- as.integer(df$Tid)
df$value <- as.numeric(df$value)
df <- tidbbl(df)
df <- rename(df, var = ContentsCode)
df <- rename(df, tid = Tid)
df <- rename(df, verdi = value)
```

Oppgave Ib: Lag en figur

FC8lgende kode skaper en enkel figur.

```
df %>%
  filter(var == "BNP") %>%
  ggplot(aes(x=tid,y=verdi)) +
  geom_line()
```



Lag en pen figur som viser BNP i tusener av kroner per person, i bC%de IC8pende og faste priser, mellom 2000 og 2021. Skriv en tydelig forklaring og tolkning av figuren. Hvordan har inntektene utviklet seg? Forklar forskjellen mellom BNP i IC8pende og faste priser. Til hvilke formC%l er de mest relevante?

```
# Oppgave Ib IC8ses her
df %>%
  filter(var == "BNP") %>%
  filter(tid %in% c(2000:2021)) %>%
  ggplot(aes(x=tid,y=verdi)) +
  scale_y_continuous(labels=scales::comma) +
  geom_line()
```

[Besvar oppgave Ib]

I denne grafen kan vi se at i begynnelsen av 2000 begynte BNP'en i Norge og vokse. Vi kan se at den fortsetter ?? vokse fram til den ??konomiske krisen som traff verden i 2008. Denne gjorde ikke et s?? altfor stort inntrykk p?? Norge men gjorde litt likevel. Etter dette fortsetter den ?? vokse med noe nedgang her og der. Alt i alt kan vi si at den norske BNPen har vokst ganske mye de siste 20 ??rene.

Fant bare et dataset for denne grafen s?? det var ikke mulig ?? presentere l??pende og faste.

II. Transformasjon, visualisering

VC%re data er en tidsserie, hvilket betyr at rekkefC8lgen i observasjonene er ordnet etter tid. Vi skal nC% regne prosentvis, C%rlig endring. La x_t vC&re BNP i C%r t . For eksempel vil x_{1970} vC&re 23616.

Den C%rlige endringen i BNP fra C%r $t - 1$ til t er gitt ved $x_t - x_{t-1}$. I samfunnsC8konomi er det vanlig C% betegne dette som $\Delta x_t := x_t - x_{t-1}$. Tegnet Δ er den greske bokstaven delta og betegner differanse. For eksempel vil $\Delta x_{1971} = 26363 - 23616 = 2747$ kroner.

I mange tilfeller er vi interesserte i relativ vekst: Hvor mye C8kte BNP, relativt til hva den var i utgangspunkt? Den mest brukte enheten er hundredeler eller prosentvis endring, gitt ved $100 \times \Delta x_t / x_{t-1}$. For eksempel var den prosentvise endringen i BNP i 1971 $100 \times \Delta x_{1971} / x_{1970} = 100 \times (2747 / 23616) \approx 11.6$, hvor \approx betegner "omtrent lik" da jeg viser svaret med kun C)n desimal. Tilsvarende kan man skrive at $\Delta x_{1971} / x_{1970} = 2747 / 23616 \approx 0.116 = 11.6\%$, hvor tegnet % betegner at belC8pet oppgis i hundredeler eller prosent.

Oppgave IIa: Omorganisere datasett med pivot_wider()

Vi skal lage to variable dBNP og dMEMOBNP som viser relativ endring i BNP og MEMOBNP. Til dette formC%let skal vi bruke kommandoene pivot_wide() og pivot_long() til C% omorganisere dataene. Jeg anbefaler dere fC8rst C% lese [kapittel 12.3](#) i pensum. Betrakt fC8lgende kode.

```
df_wide <- df %>%
  pivot_wider(names_from = var, values_from = verdi) 
```

Beskriv konkret hva koden gjorde. Sammenlign df og df_wide.

[Besvar oppgave IIa her]

df_wide har omrøkkert p?? radene og utvidet tabellen, som gj??r tabellen litt mer oversiktlig og enkel ?? lese.

Oppgave IIb: Beregn vekst

Til C% beregne endring er funksjonen lag() meget nyttig. I denne konteksten er begrepet *lag* et engelsk verb som beskriver foregC%ende observasjon. Bruker vi funksjoenen lag() pC% en variabel (kolonne) sC% returnerer den en ny kolonne hvor verdien er lik foregC%ende observasjon. Betrakt fC8lgende kode:

```
df_wide <- df_wide %>%
  mutate(LBNP = lag(BNP,n=1L)) %>%
  mutate(LMEMOBNP = lag(MEMOBNP,n=1L))
```

legger variablene i rekkefC8lge

```
df_wide <- df_wide %>%
  relocate("LBNP", .before = "MEMOBNP")
```

```
df_wide
```

```
# A tibble: 52 × 5
  tid   BNP   LBNP MEMOBNP LMEMOBNP
<int> <dbl> <dbl>   <dbl>   <dbl>
1 1970 23616 NA      214756   NA
2 1971 26363 23616  225352  214756
3 1972 29078 26363  235557  225352
4 1973 32805 29078  244518  235557
5 1974 37734 32805  252539  244518
6 1975 42884 37734  263586  252539
7 1976 48711 42884  277636  263586
8 1977 54652 48711  287968  277636
9 1978 60091 54652  297971  287968
10 1979 66069 60091  309942  297971
# ... with 42 more rows
```

Hvis vi bruker den matematiske notasjonen diskutert tidligere sC% har vi nC% kolonner med x_t (BNP, MEMOBNP) og x_{t-1} (LBNP, LMEMOBNP).

Bruk funksjonen mutate() til C% lage en ny variabel med relativ endring i BNP og MEMOBNP i df_wide og lagre de som dBNP og dMEMOBNP.

```
# Besvar oppgave IIb her
df_wide <- df_wide %>%
  mutate(dBNP = BNP - LBNP) %>%
  mutate(dMEMOBNP = MEMOBNP - LMEMOBNP)
```

```
df_wide <- df_wide %>%
  relocate("dBNP", .before = "MEMOBNP")
```

Oppgave IIc: Omorganisere datasett med pivot_longer()

Bruk nC% funksjonen pivot_longer() til C% transformere df_wide til det opprinnelige formatet, altsC% med variablene var og verdi. Kall den transformerte tabellen for df_long.

NB! Husk C% bruk anfC8rselstegn ("variabelnavn") nC%r du definerer nye variable i pivot_longer().

```
# Besvar oppgave IIc
df_long <- df_wide %>%
  pivot_longer(cols = everything(), names_to = "var", values_to = "verdi")
```

Oppgave IId: Figur med vekst

Lag en pen figur med prosentvis vekst i nominelt og reell BNP per person fra 1970 til 2021. Finnes det observasjoner med negativ vekst i reell BNP? Hva skyldes dette?

Merknad: Det er en del stC8yi i data. PrC8v C% kombinere geom_point() og geom_smooth() for C% fC% et bedre inntrykk av den langsiktige utviklingen.

```
# Besvar oppgave IId her
df_wide %>% mutate(prosBNP = 100*(BNP - lag(BNP))/lag(BNP))
```

```
# A tibble: 52 × 8
  tid   BNP   LBNP   dBNP MEMOBNP LMEMOBNP dMEMOBNP prosBNP
<int> <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
1 1970 23616 NA      NA      214756 NA      NA      NA
2 1971 26363 23616  2747  225352 214756 10596  11.6
3 1972 29078 26363  2715  235557 225352 10205  10.3
4 1973 32805 29078  3727  244518 235557  8961  12.8
5 1974 37734 32805  4929  252539 244518  8021  15.0
6 1975 42884 37734  5150  263586 252539 11047  13.6
7 1976 48711 42884  5827  277636 263586 14050  13.6
8 1977 54652 48711  5941  287968 277636 10332  12.2
9 1978 60091 54652  5439  297971 287968 10803  9.95
10 1979 66069 60091  5978  309942 297971 11971  9.95
# ... with 42 more rows
```

```
df_wide %>%
  mutate(prosBNP = 100*(BNP - lag(BNP))/lag(BNP)) %>%
  filter(tid >=1971) %>%
  ggplot(aes(x=tid, y=prosBNP)) +
  geom_line() +
  geom_smooth() +
  geom_point()
```

```
`geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Vi ser p?? denne grafen at det var negativ vekst i ca 2002/2003, 2008 og i 2020.

Vi kan forklare disse periodene med negativ vekst ganske enkelt.

2002, Nedgang for den Amerikanske ??konomien n??dde Norsk ??konomi og dro oss ned etter.

2007/2008 Var det nedgang pga. den globale finanskrisen som hendte.

2020 Nedgang pga. covid-19 viruset som skapte en global pandemi.