

R-seminar 2 Oppgaver

STV1020 Vår 2021

Uke 11

1. Sett ditt working directory og last inn datasettet *gapminder*.

```
## Setter working directory
setwd("filstiendin")

## Error in setwd("filstiendin"): cannot change working directory

## Laster inn datasettet gapminder (fra working directory)
load("gapminder.Rdata")

View(gapminder)
head(gapminder)

##      country continent year lifeExp      pop gdpPercap
## 1 Afghanistan      Asia 1952  28.801  8425333  779.4453
## 2 Afghanistan      Asia 1957  30.332  9240934  820.8530
## 3 Afghanistan      Asia 1962  31.997 10267083  853.1007
## 4 Afghanistan      Asia 1967  34.020 11537966  836.1971
## 5 Afghanistan      Asia 1972  36.088 13079460  739.9811
## 6 Afghanistan      Asia 1977  38.438 14880372  786.1134
```

2. Opprett en ny variabel hvor landene i Asia får verdien 1, Amerika verdien 2, Oseania verdien 3, Europa verdien 4 og Afrika verdien 5. Undersøk om omkodning er gjort riktig.

```
## Lager ny variabel over kontinenter med 5 kategorier
gapminder$continent_5 <- ifelse(gapminder$continent == "Asia", 1,
                                ifelse(gapminder$continent == "Americas", 2,
                                        ifelse(gapminder$continent == "Oceania", 3,
                                              ifelse(gapminder$continent == "Europe", 4,
                                                    ifelse(gapminder$continent == "Africa", 5, NA))))))

## Sjekker at det ble riktig
table(gapminder$continent, gapminder$continent_5)

##
##      1  2  3  4  5
## Africa 0  0  0  0 624
```

```
## Americas 0 300 0 0 0
## Asia 396 0 0 0 0
## Europe 0 0 0 360 0
## Oceania 0 0 24 0 0

# 625 observasjoner for Afrika for eksempel
```

3. Opprett en ny variabel hvor verdien på variablene lifeExp er ganget med 1000. Undersøk om du har omkodet riktig.

```
## Lager ny variabel der lifeExp ganges med 1000
gapminder$lifeExp_1000 <- gapminder$lifeExp * 1000

## Sjekker om det ble riktig
summary(gapminder$lifeExp)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 23.60  48.20   60.71   59.47  70.85   82.60

summary(gapminder$lifeExp_1000)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 23599  48198   60713   59474  70846   82603
```

4. Velg ut tre variabler fra datasettet og opprett et nytt subset av gapminder-datasett bestående av disse tre variablene. Inkluder kun enheter med verdien 1977 på variabelen year. (Tips: Husk at R jobber sekvensielt, så pass på hvilket argument som kommer først).

```
## Henter relevante pakker opp fra biblioteket
library(tidyverse)
library(dplyr)

## Lager nytt datasett med kun relevante variabler og filtrert årstall
gapminder_2 <- gapminder %>%
  filter(year == 1977) %>%
  select(pop, country, lifeExp)

View(gapminder_2)
```

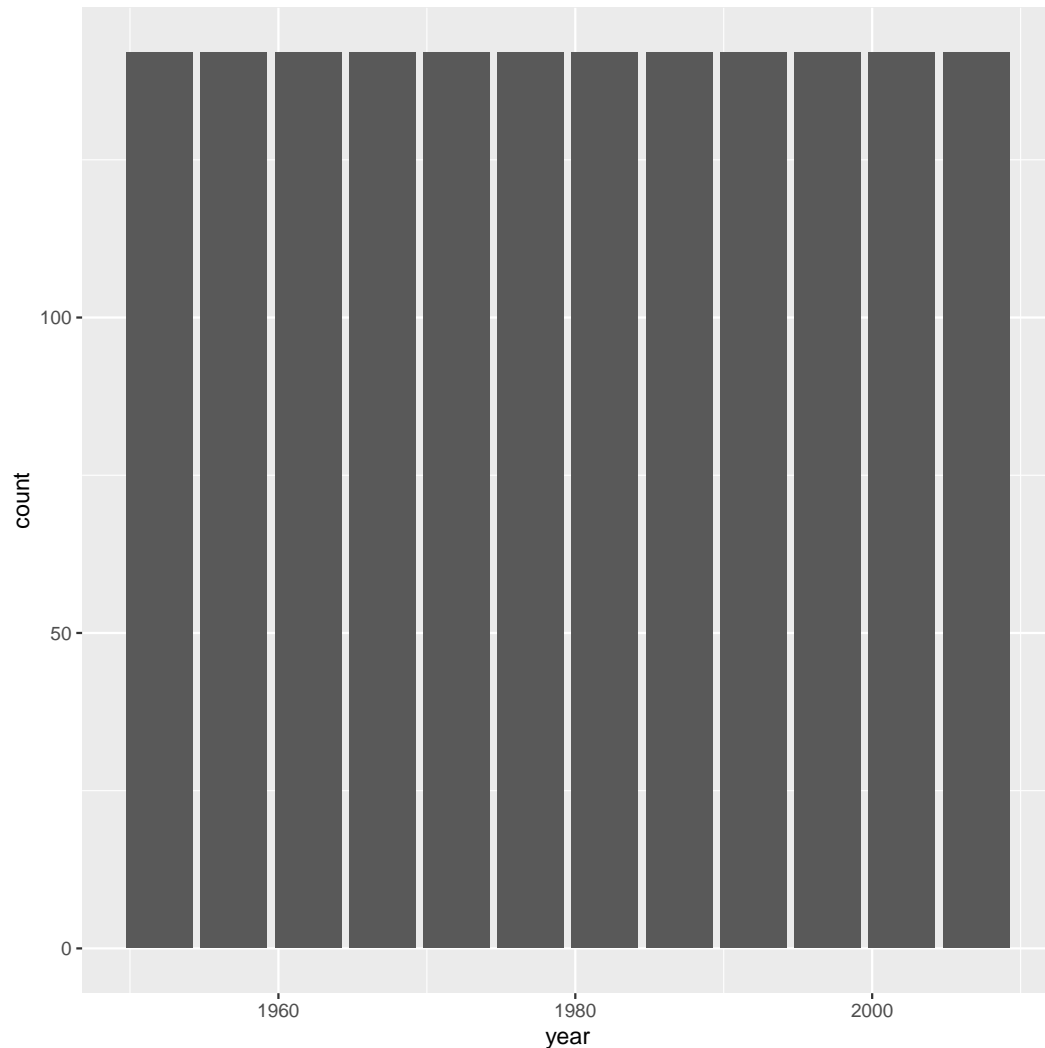
5. Opprett et nytt subset av gapminder-datasettet bestående av europeiske land som har en forventet levealder på 75 år eller høyere.

```
## Lager nytt datasett med kun europeiske land som har en forventet levealder
# på 75+
gapminder_3 <- gapminder %>%
  filter(continent == "Europe",
         lifeExp >= 75) # Verdien skal være større > eller den samme = som 75 år

View(gapminder_3)
```

6. Dataene i Gapminder er hentet inn hvert femte år. Vi er interessert i å vite om det er like mange observasjoner for hvert av årene. Lag et søylediagram for å undersøke dette.

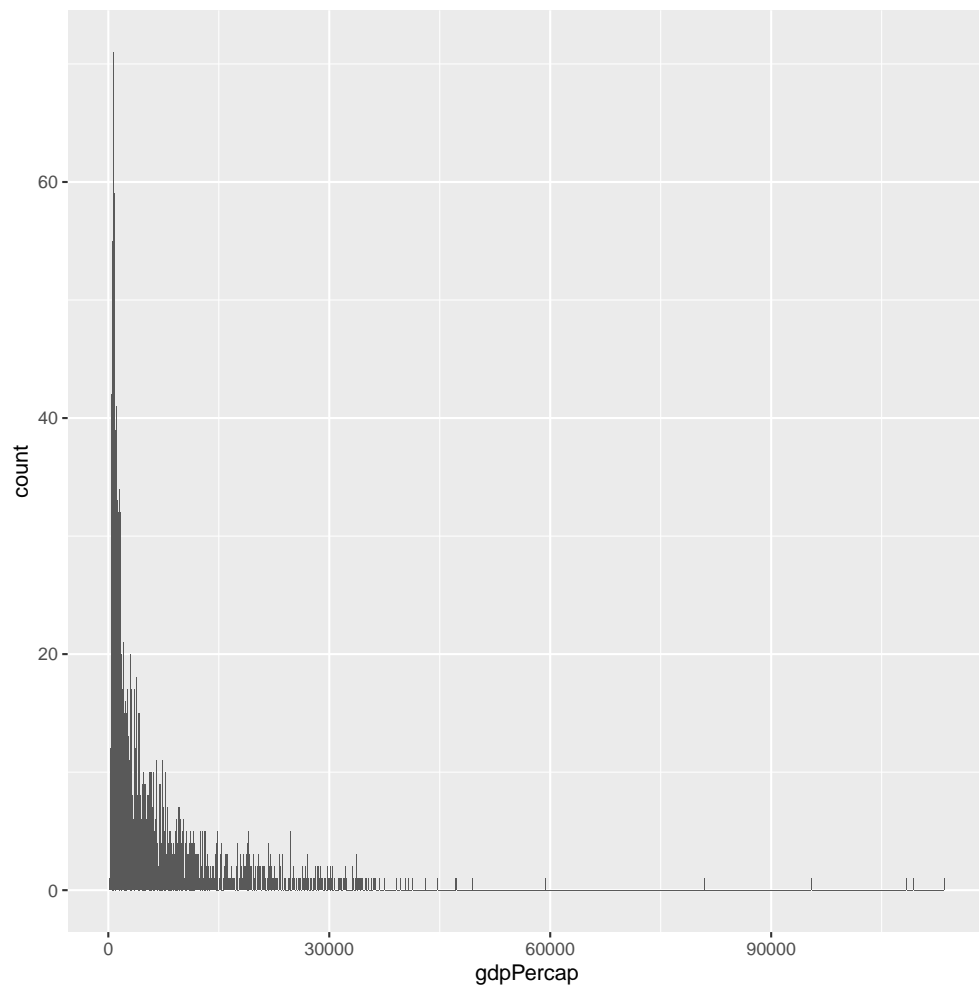
```
## Lager søylediagram over år
ggplot(gapminder, aes(x = year)) +
  geom_bar()
```



```
# Det er like mange observasjoner hvert år
```

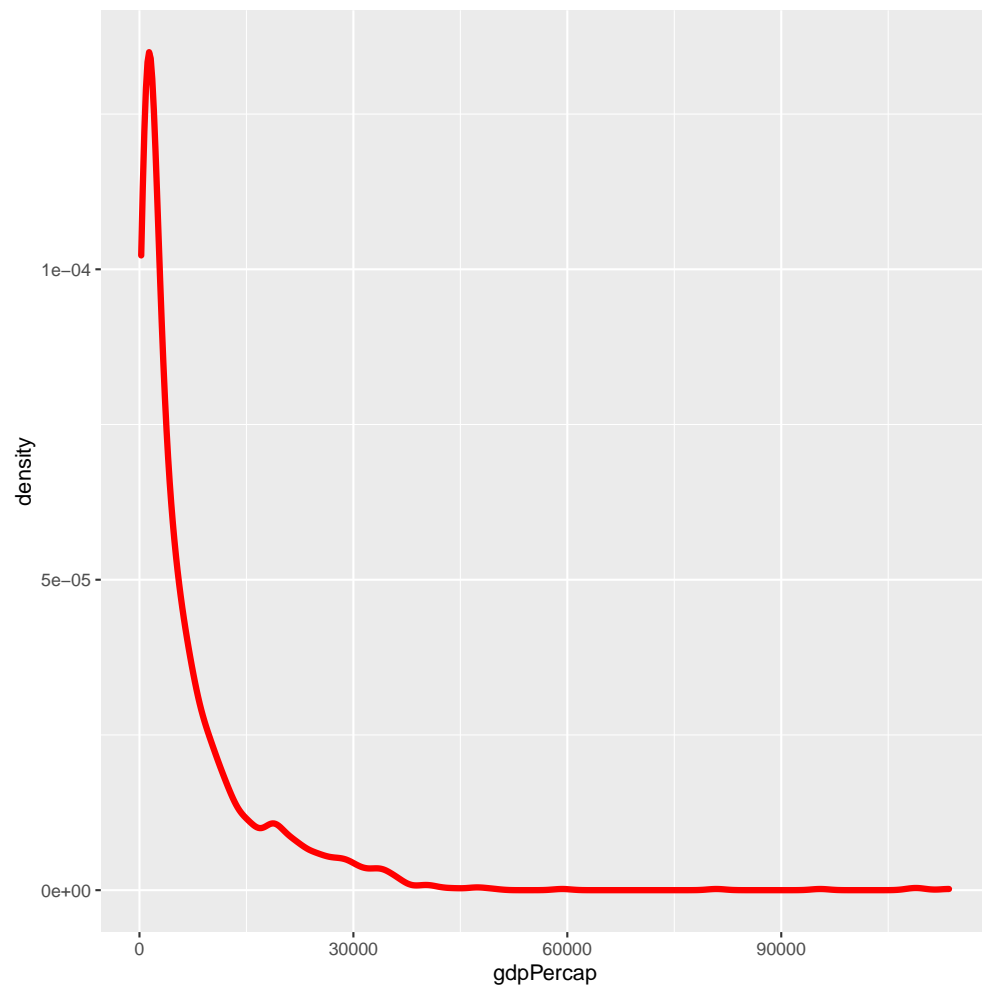
7. • Lag et histogram for variabelen GDP per capita (gdpPercap).

```
## Lager histogram over GDP per capita
ggplot(gapminder, aes(x = gdpPercap)) +
  geom_histogram(binwidth = 100)
```



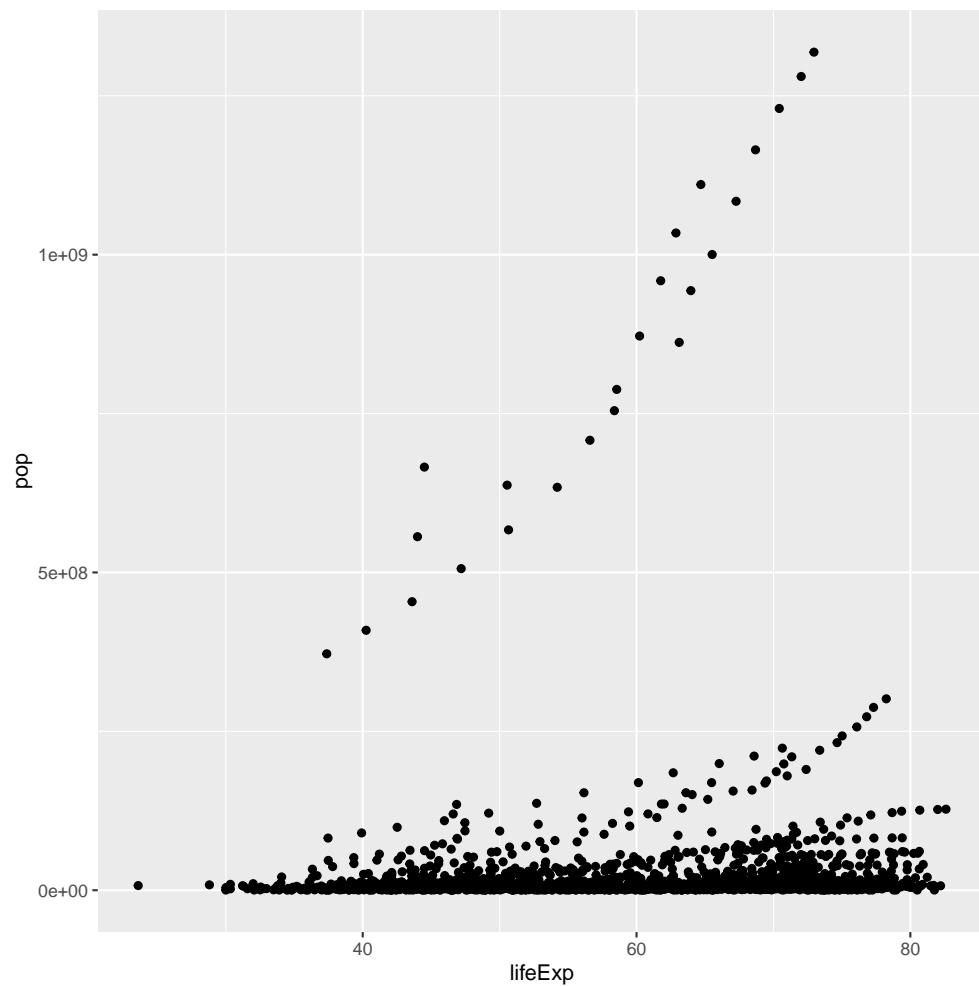
- Lag et density plot for den samme variabelen. Gjør linja rød, og sett tykkelsen til 1.5

```
## Lager density plot for GDP per capita
ggplot(gapminder, aes(x = gdpPercap)) +
  geom_density(size = 1.5, col = "red") # Størrelse 1.5 og rød farge
```



8. • Lag et scatterplot over forventet levealder (lifeExp) og befolkningstall (pop).

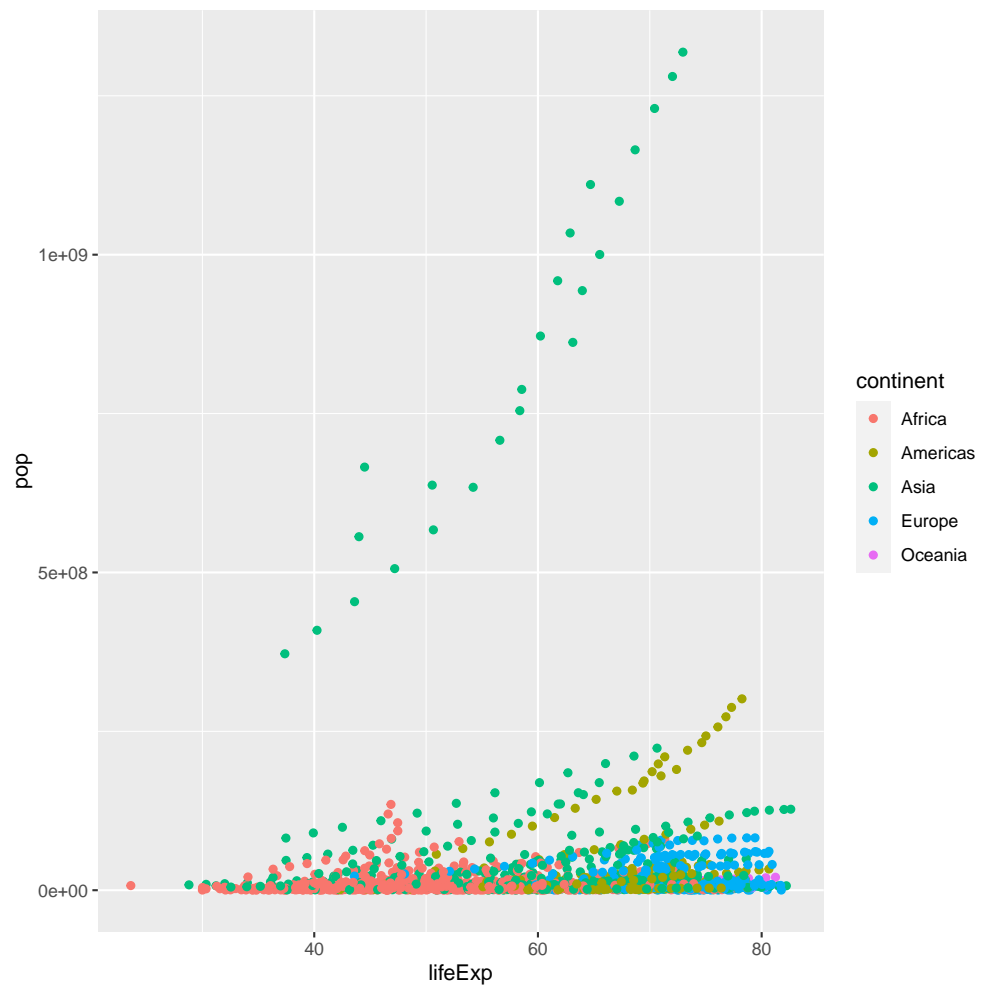
```
## Lager scatterplot som viser sammenhengen mellom levealder og  
# befolkningstall  
ggplot(gapminder, aes(x = lifeExp, y = pop)) +  
  geom_point()
```



```
# Noen observasjoner som skiller seg ut, øverste punkter
# Generelt sett har forventet levealder økt i takt med økt/høyere
# befolkningstall
```

- Del opp observasjonene i ulike farger for hvert kontinent.

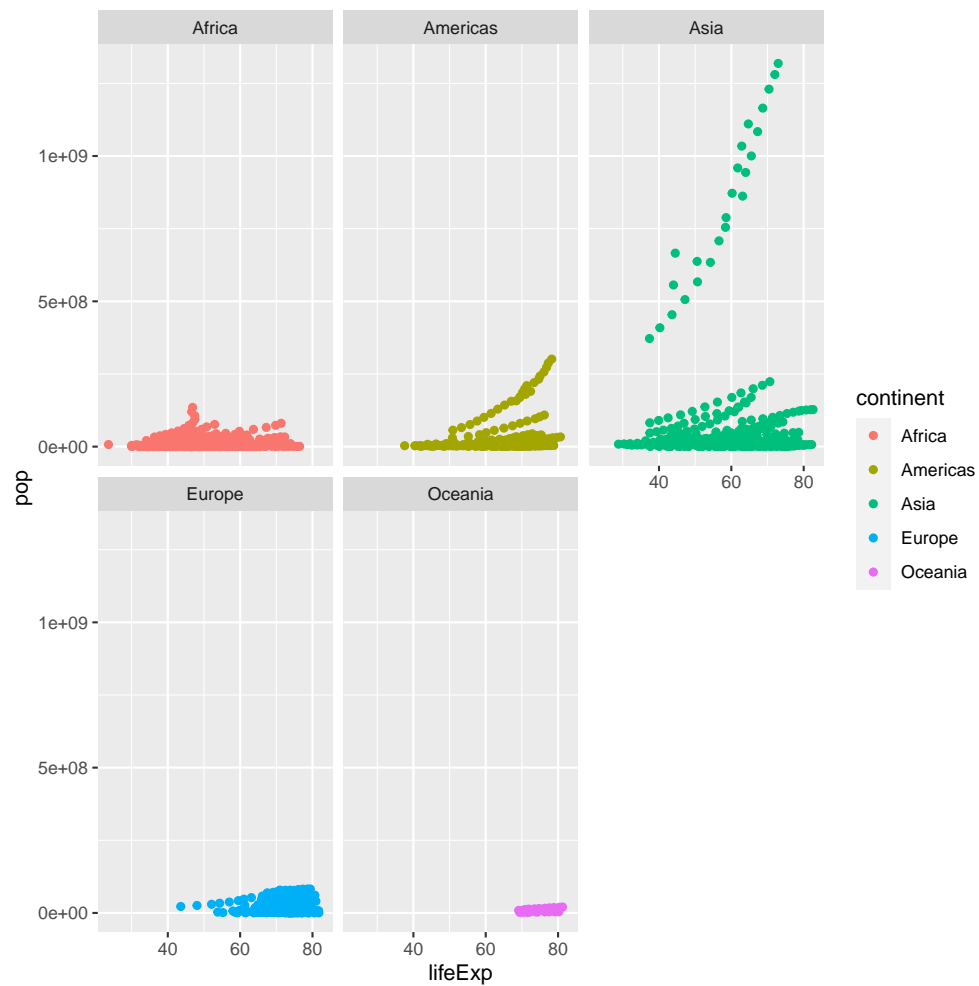
```
## Scatterplot med hensyn til kontinent, der hvert kontinent får egen farge
ggplot(gapminder, aes(x = lifeExp, y = pop, color = continent)) +
  geom_point()
```



Ser nå at observasjoner for Asia skiller seg ut

- Kan du bruke `facet_wrap` for å gjøre plottet mer oversiktlig?

```
## Bruker facet_wrap for å få fem figurer istedenfor alle i én
ggplot(gapminder, aes(x = lifeExp, y = pop, col = continent)) +
  geom_point() + facet_wrap(vars(continent))
```

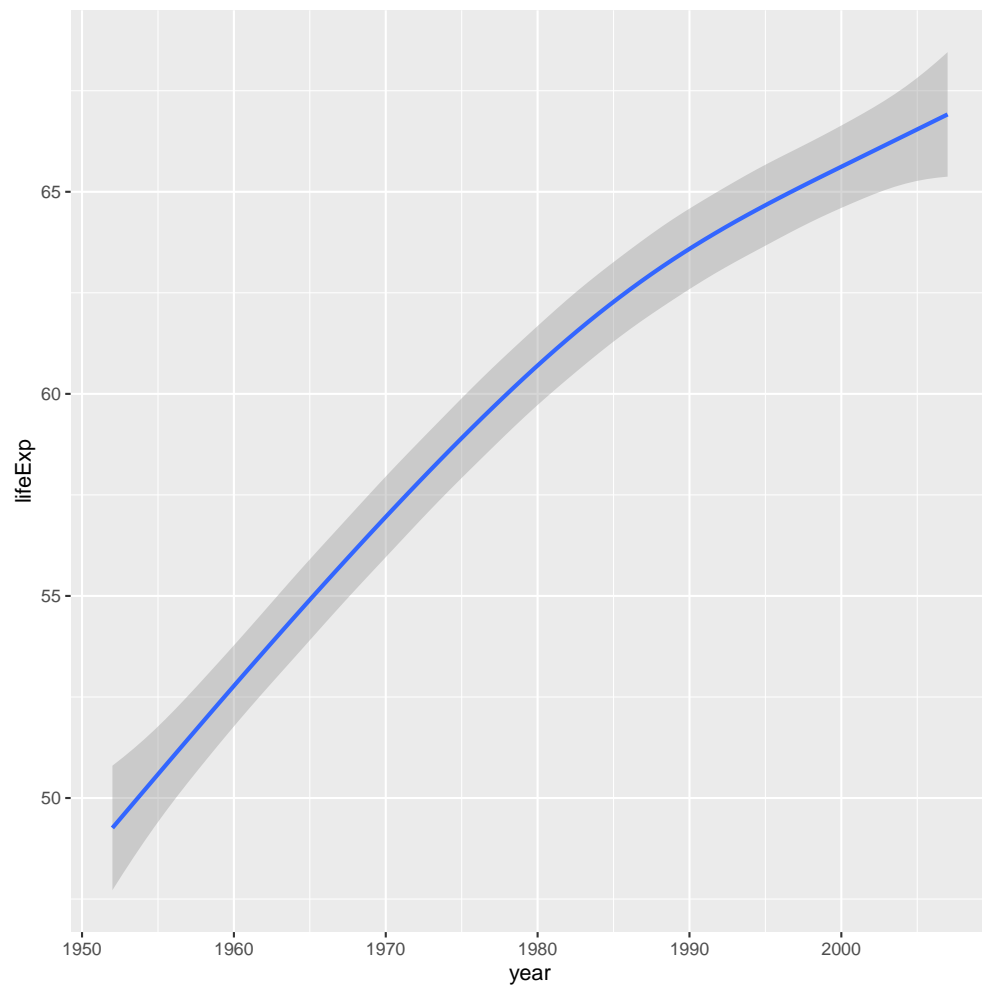


```
# Ser nå enda tydeligere at det kun er noen observasjoner i Asia som
# skiller seg ut
```

9. • Plott en graf (ei linje) over hvordan forventet levealder (lifeExp) har endret seg over tid.

```
## Lager graf over endringen i forventet levealder over tid
ggplot(gapminder, aes(x = year, y = lifeExp)) +
  geom_smooth()

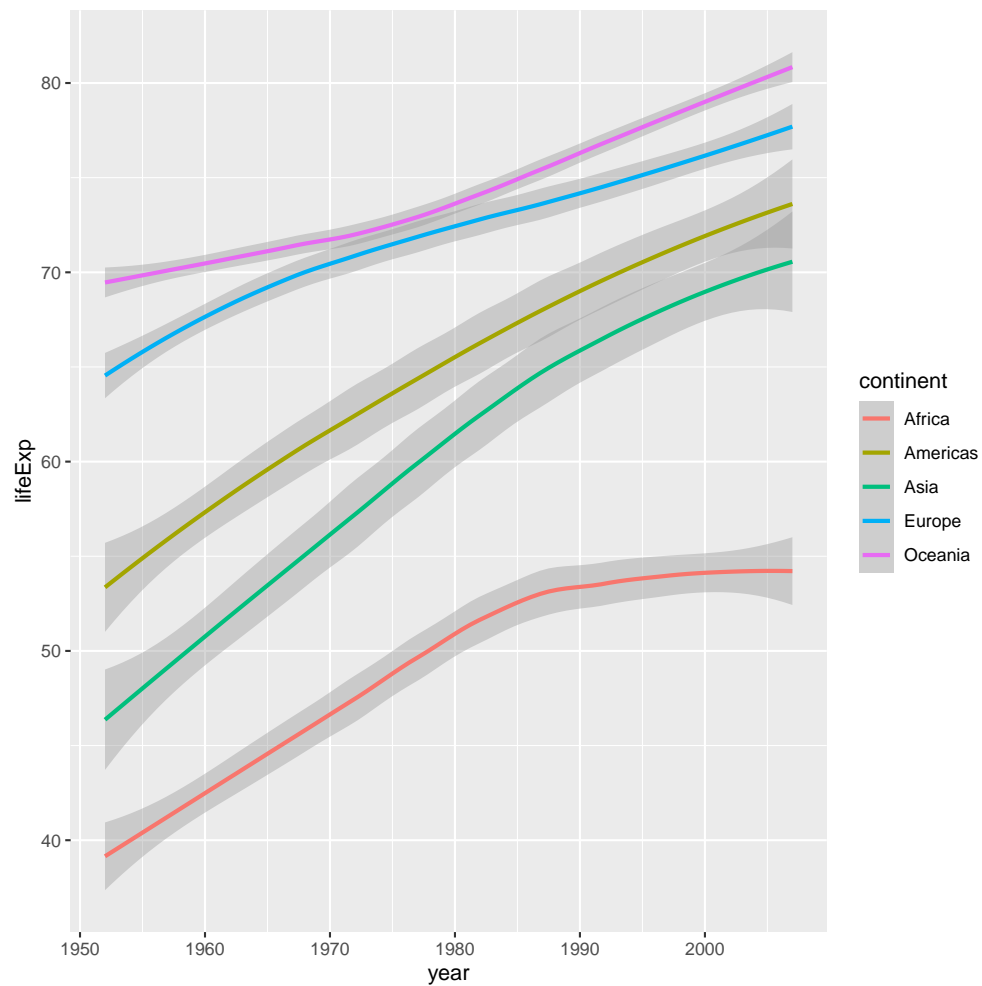
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
# Forventet levealder har økt over tid
```

- Hvordan ser dette trenden ut på ulike kontinenter? Lag ei linje for hvert kontinent.

```
## Lager graf som viser endringen i forventet levealder for ulike  
# kontinenter  
ggplot(data=gapminder, aes(x = year, y = lifeExp, col = continent)) +  
  geom_smooth()  
  
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



*# Forventet levealder har flatet ut noe for Afrika siden rundt 1985
 # Forventet levealder i Amerika og Asia har økt mest*

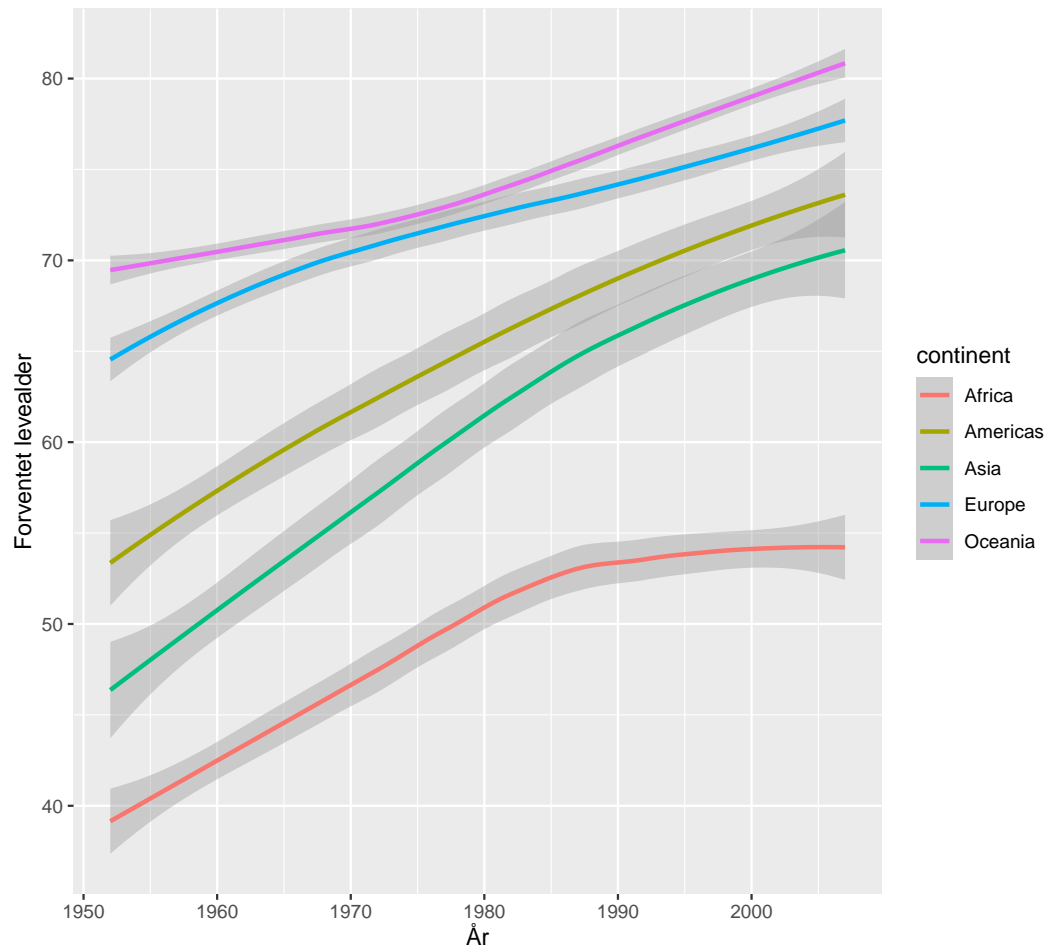
10. Legg til en tittel på plottet fra oppgave 9, samt nye labels til x- og y-aksen. Lagre plottet ditt som et objekt i environment og lagre det deretter på pc-en din, både som .png- og .pdf-fil.

```
## Gir plottet en tittel og labels til x- og y-aksen
ggplot(gapminder, aes(x = year, y = lifeExp, col = continent)) +
  geom_smooth() +
  labs(x = "År",
       y = "Forventet levealder",
       title = "Forventet levealder fra 1952 til 2007",
       subtitle = "For fem kontinenter", # Sper på med undertittel
       caption = "Kilde: Gapminder") # og en caption til figuren

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

Forventet levealder fra 1952 til 2007

For fem kontinenter



Kilde: Gapminder

```
## Oppretter plottet som et objekt
levealder <- ggplot(gapminder, aes(x = year, y = lifeExp, col = continent)) +
  geom_smooth() +
  labs(x = "År",
       y = "Forventet levealder",
       title = "Forventet levealder fra 1952 til 2007",
       subtitle = "For fem kontinenter", # Sper på med undertittel
       caption = "Kilde: Gapminder") # og en caption til figuren

## Lagrer grafen som .png-fil
ggsave("levealderplott.png", plot = levealder)

## Saving 7 x 7 in image
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'

## Lagrer som .pdf-fil
ggsave("levealderplott.pdf", plot = levealder)

## Saving 7 x 7 in image
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Ekstraoppgave:

1. Opprett et datasett (basert på *gapminder*) der du kun inkluderer afrikanske land fra året 1952 som har en høyere forventet levealder enn gjennomsnittet for alle afrikanske landene i 1952. (Denne oppgaven er litt vanskelig, men prøv før du ser på løsningsforslaget)

```
## Lager nytt datasett for kun Afrika i 1952
gapminder_africa <- gapminder %>%
  filter(year == 1952,
         continent == "Africa")

## Lager datasett som har høyere forventet levealder enn gjennomsnittet
# for de afrikanske landene
gapminder_africa_2 <- gapminder_africa %>%
  filter(lifeExp > mean(lifeExp))
```