

# Prosjektoppgave

André Langvik, Martin Smedstad, Mathias Hetland

December 15, 2021

## Contents

<b>DEL A:</b>	<b>3</b>
Del 1: . . . . .	3
Oppgave A) . . . . .	3
Oppgave B) . . . . .	4
Oppgave C) . . . . .	5
Oppgave D) . . . . .	6
Oppgave E) . . . . .	7
Oppgave F) . . . . .	8
Oppgave G) . . . . .	9
Oppgave H) . . . . .	10
Analyse: . . . . .	11
DEL 2: . . . . .	14
Oppgave A) . . . . .	14
Oppgave B) . . . . .	14
Opgave C) . . . . .	15
<b>DEL B</b>	<b>20</b>
Oppgave A) . . . . .	20
Oppgave B) . . . . .	21
Oppgave C) . . . . .	22
Oppgave D) . . . . .	24
Oppgave E) . . . . .	25
Oppgave F) . . . . .	27
Oppgave G) . . . . .	28
Oppgave H) . . . . .	30
Sammendrag: . . . . .	32
<b>Kilder</b>	<b>32</b>

---

Laster ned nødvendige pakker og henter URL fil

```
skipsdata <- readRDS(file = url("https://bit.ly/3FfJbKa"))

library(tidyverse)

## Registered S3 methods overwritten by 'tibble':
##   method      from
##   format.tbl pillar
##   print.tbl  pillar

## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.5     v purrr    0.3.3
## v tibble   2.1.3     v dplyr    1.0.7
## v tidyr    1.0.2     v stringr  1.4.0
## v readr    1.3.1     vforcats  0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()

library(plyr)

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## -----
## 
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
## 
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize

## The following object is masked from 'package:purrr':
## 
##   compact
```

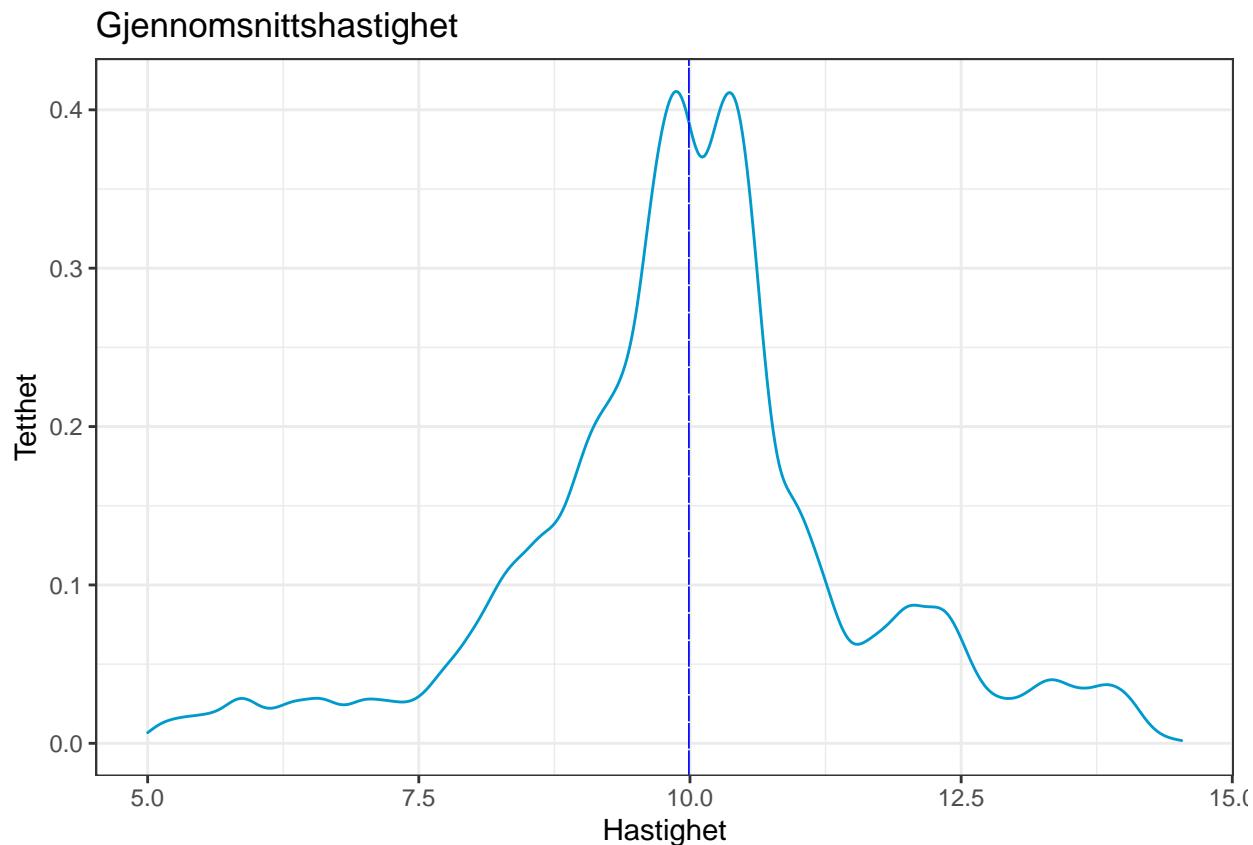
## DEL A:

Del 1:

Oppgave A)

Finner gjennomsnittshastighet:

```
gjennomsnitthastighet <-  
  print(mean(skipsdata$hastighet))  
  
## [1] 9.990076  
  
plot1 <-  
  ggplot(skipsdata, aes(x=hastighet)) +  
  geom_density(color="deepskyblue3") +  
  geom_vline(aes(xintercept=mean(hastighet)),  
             color="blue", linetype="F1", size=0.3) + theme_bw() +  
  labs(title="Gjennomsnittshastighet",  
       y="Tetthet",  
       x="Hastighet")  
plot1
```



Fordelingen viser at det er to topper av makshastighet, som befinner seg på to forskjellige knop-verdier og at gjennomsnittshastigheten ligger rett under 10 knop. Farten før og etter 10 knop har vært stigende og synkede.

## Oppgave B)

Finner gjennomsnittshastighet for mannskap A og B

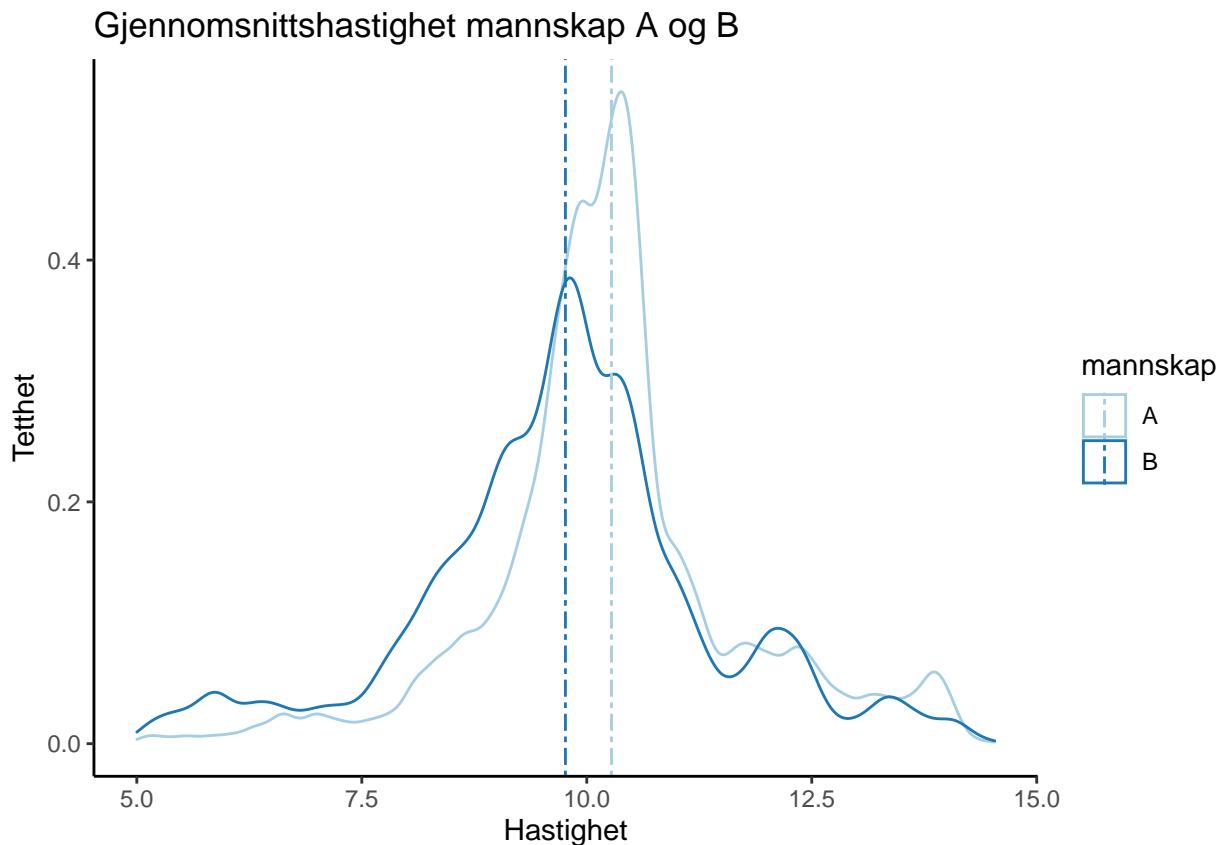
```
mannskapgjennomsnitt <-
  print(ddply(skipsdata, "mannskap", summarise, grp.mean=mean(hastighet)))
```

```
##   mannskap grp.mean
## 1          A 10.27443
## 2          B  9.76241
```

Lager et plot med gj.snittfart for mannskap A og B:

```
plot2 <-
  ggplot(skipsdata, aes(x=hastighet, color=mannskap)) +
  theme_bw() +
  geom_density() +
  geom_vline(data = mannskapgjennomsnitt, aes(xintercept=grp.mean, color=mannskap),
             linetype = "twodash", size= 0.5) +
  scale_color_brewer(palette = "Paired") +
  theme_classic() +
  labs(title="Gjennomsnittshastighet mannskap A og B",
       y="Tetthet",
       x="Hastighet")
```

plot2



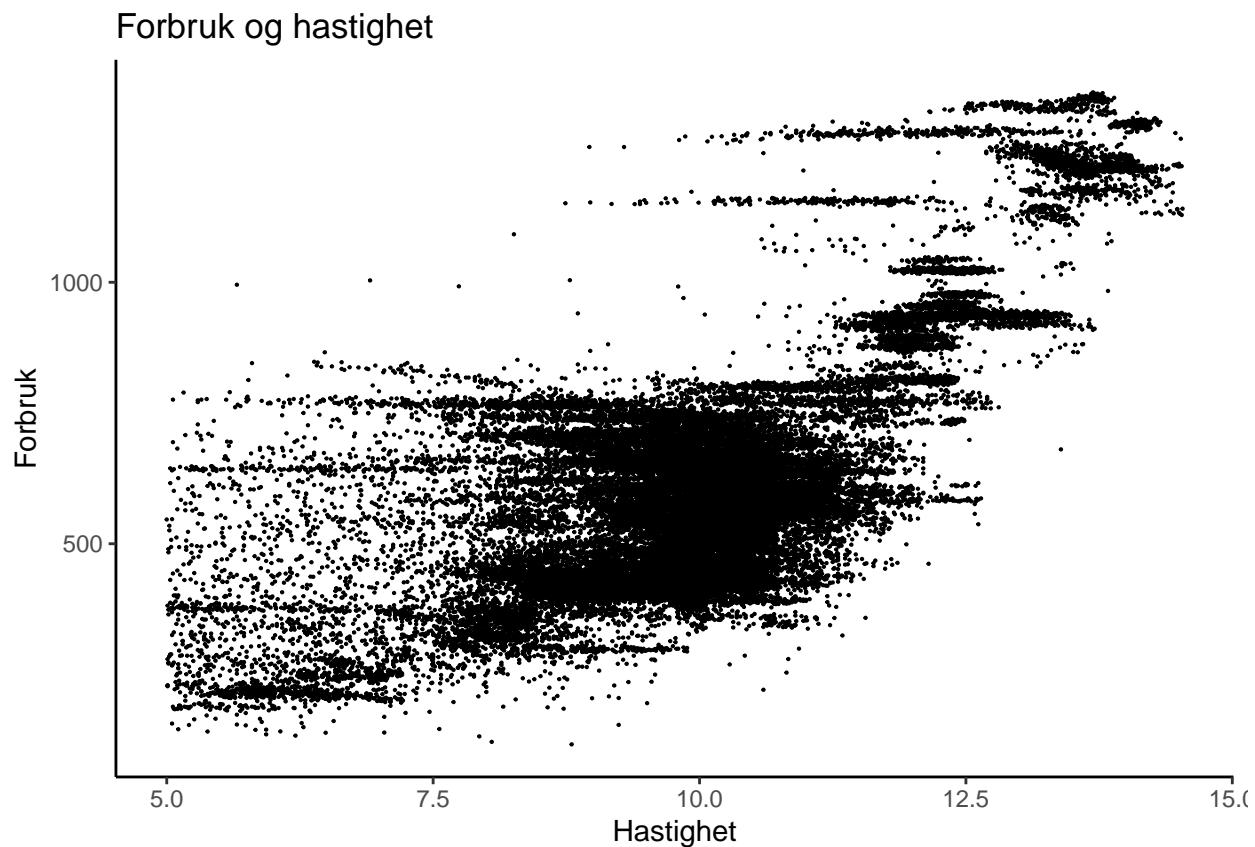
I dette plottet vises gjennomsnittshastighetene til de to mannskapene, der mannskap B har lavere gjennomsnittshastighet i forhold til mannskap A. Sammenlignet med plot i oppgave.A, så ser man at de to høydene i plottet er toppfarten til mannskapene. Gjennomsnittshastigheten for hele mannskapet var nesten 10 knop, man ser at hastighetene til mannskap A og B er på hver sin side av samlet gjennomsnittshastighet.

---

### Oppgave C)

Lager et spredningsplot av drivstoffforbruk:

```
skipsdata %>%
  ggplot(aes(x=hastighet, y=forbruk)) +
  geom_point(lwd=0.1) +
  theme_classic() +
  labs(title="Forbruk og hastighet",
       y = "Forbruk",
       x = "Hastighet")
```



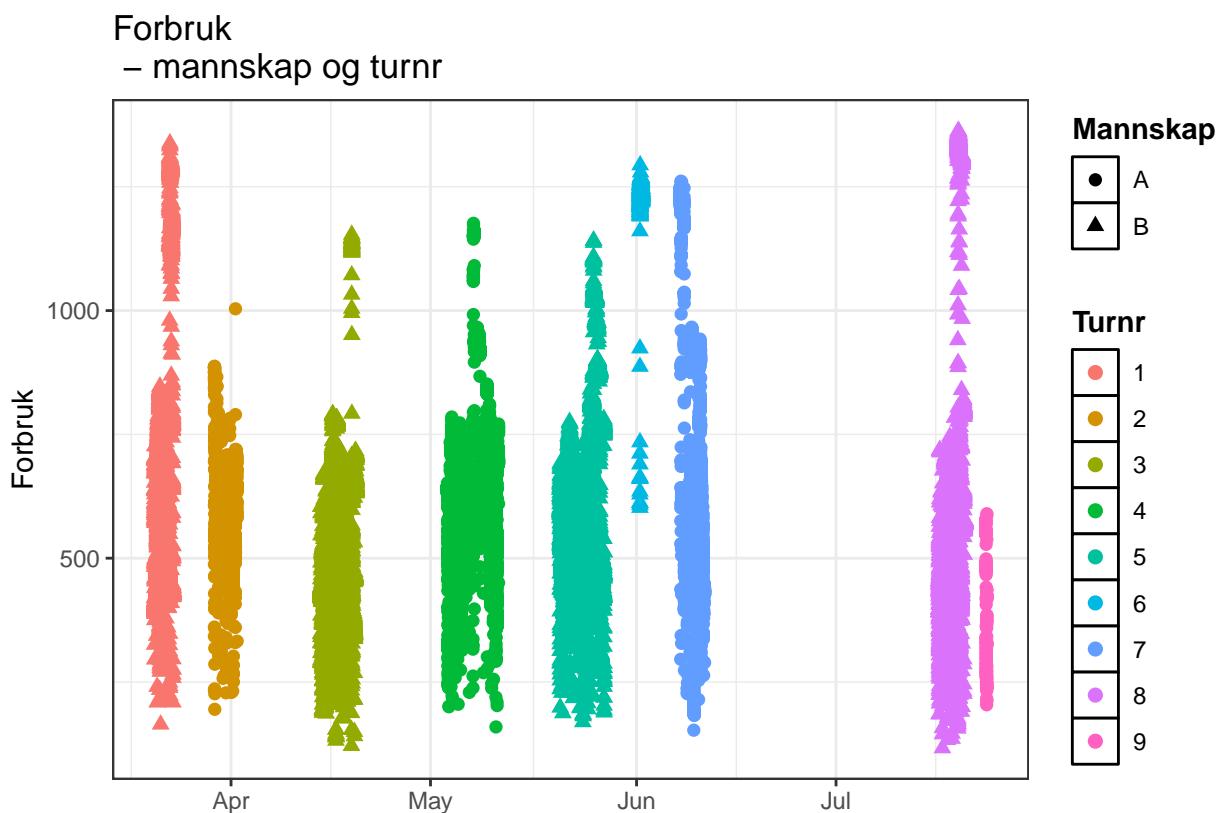
I dette 2D-scatterplotet ser man at jo høyere farten er, jo mer drivstoff det blir brukt. Mønsteret visser at mesteparten av drivstoffet blir brukt rundt gjennomsnittshastigheten til mannskap A og B. Har også blitt observert store mengder forbruk av drivstoff i lav hastighet. Drivstoffforbruks går veldig opp når man passerer rundt 11 knop.

---

## Oppgave D)

Lager en plot som viser sammenheng mellom forbruk og mannskap:

```
skipsdata %>%
  ggplot(aes(dato, forbruk, colour= factor(turnr),
             shape=factor(mannskap))) +
  geom_point(lwd=2) +
  theme_bw() +
  labs(colour="Turnr",
       shape = "Mannskap") +
  theme(legend.key = element_rect(fill="white", color="black")) +
  theme(legend.title = element_text(face="bold")) +
  labs(title="Forbruk \n - mannskap og turnr",
       y = "Forbruk",
       x="")
```



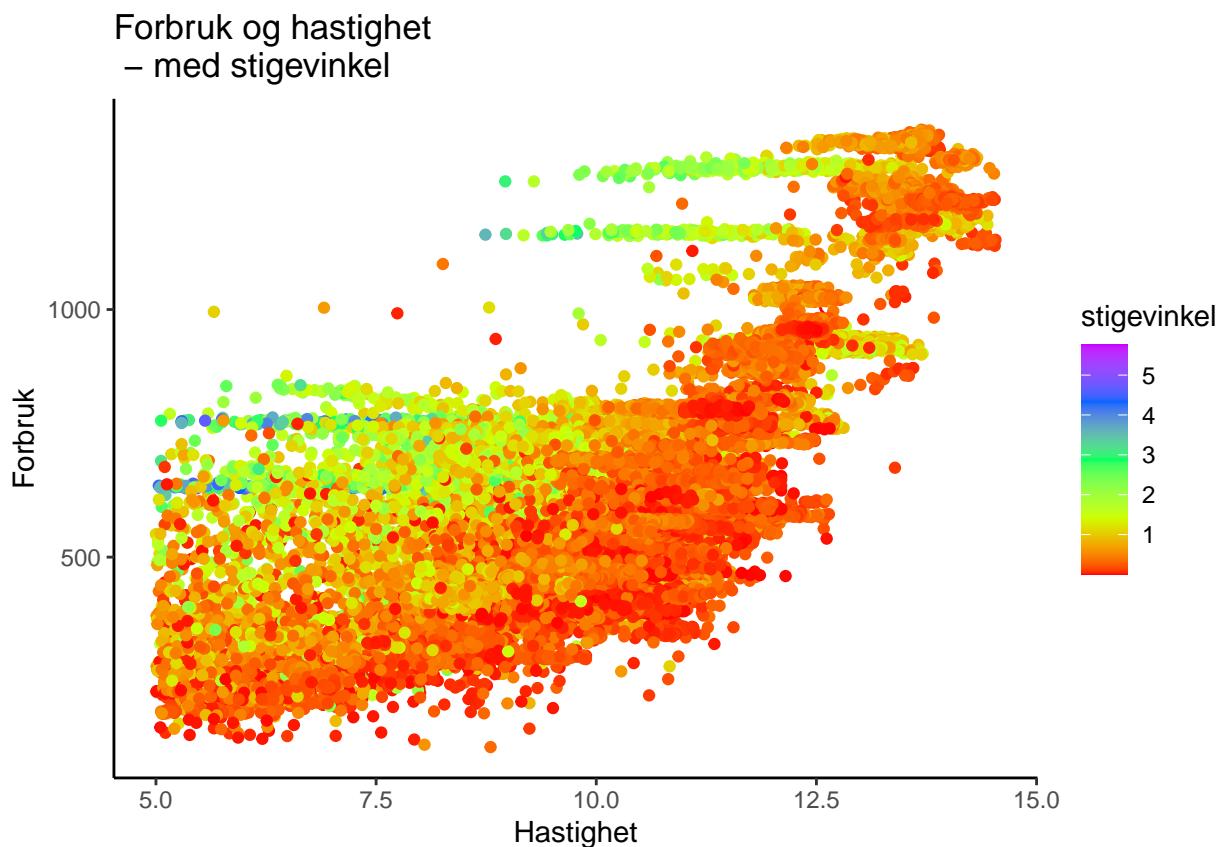
Nytten med denne informasjonen er at man kan se hvor mye forbruk av drivstoff det er per mannskap som er ombord. Man kan se at mannskap B har hatt veldig høyt forbruk hver tur. Mannskap A har brukt mindre drivstoff på sine turer enn B, men også hatt turer med høyt forbruk, selv om de har høyere gjennomsnittshastighet enn mannskap B.

Mannskap A har kortere perioder på havet ved sine turer enn det mannskap B har. Mannskap A har også litt jevnere forbruk på sine turer enn det mannskap B.

## Oppgave E)

Lager et spredningsplot som viser sammenheng mellom hastighet og forbruk med stigevinkel:

```
skipsdata %>%
  ggplot(aes(x=hastighet, y=forbruk)) +
  geom_point(lwd=1) +
  geom_density_2d(color="red") +
  theme_classic() +
  geom_point(aes(colour = stigevinkel)) +
  scale_color_gradientn(colours = rainbow(5)) +
  labs(title = "Forbruk og hastighet \n - med stigevinkel",
       y = "Forbruk",
       x = "Hastighet")
```



Her kan man se at høyere stigevinkel gir høyere forbruk og lavere hastighet. Samtidig så gir lav stigevinkel mulighet for høyere hastighet noe som også gir et høyt forbruk. Der hastighetene har vært veldig lav, har det også vært høy stigevinkel, altså mye bølger. På en stigevinkel på ca 3, vil det altså i snitt være lavere fart enn der gjennomsnittshastighetene og forbruk ligger. Gjennomsnittlig har stigevinekelen ligget på rundt 1-3, og man kan se at det er litt variasjon på hvor mye dette påvirker hastighet. Man kan f.eks. se at en stigevinkel på 3-4 og ca. 10 knop i hastighet gir et veldig høyt forbruk.

## Oppgave F)

Dele datasettet i to:

```
sd1 <-  
  skipsdata %>%  
  filter(stigevinkel >= 0.5)  
  
sd2 <-  
  skipsdata %>%  
  filter(stigevinkel < 0.5)
```

Gjennomsnittsfart for hele mannskap A

```
totA <-  
  skipsdata %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="A") %>%  
  summarise(Averagespeed = mean(hastighet))
```

Gjennomsnittsfart for hele mannskap B

```
totB <-  
  skipsdata %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="B") %>%  
  summarise(Averagespeed = mean(hastighet))
```

Gjennomsnittsfart for mannskap A, stigevinkel større eller lik 0.5

```
sd1A <-  
  sd1 %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="A") %>%  
  summarise(Averagespeed = mean(hastighet))
```

Gjennomsnittsfart for mannskap B, stigevinkel større eller lik 0.5

```
sd1B <-  
  sd1 %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="B") %>%  
  summarise(Averagespeed = mean(hastighet))
```

Gjennomsnittsfart for mannskap A, stigevinkel under 0.5

```
sd2A <-  
  sd2 %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="A") %>%  
  summarise(Averagespeed = mean(hastighet))
```

Gjennomsnittsfart for mannskap B, stigevinkel under 0.5

```
sd2B <-  
  sd2 %>%  
  select(mannskap, hastighet) %>%  
  filter(mannskap=="B") %>%  
  summarise(Averagespeed = mean(hastighet))
```

## Gjennomsnittshastighet:

Data	Mannskap A	Mannskap B	A-B
Datasettet	10.2744320702308	9.76240965220409	0.512022418026701
Stigevinkel < 0.5	10.5008008404285	10.0006530427472	0.50014779768124
Stigevinkel $\geq$ 0.5	10.0839945463535	9.55496085405503	0.529033692298453

I tabellen kan man se at gjennomsnittshastigheten er lavere med en stigevinkel over 0,5 samtidig som at farten er høyere med en stigevinkel på under 0,5. Her kan man også se at gjennomsnittshastigheten til mannskap A også er større enn B.

---

## Oppgave G)

*Gjennomsnittsforbruk for hele mannskap A*

```
totA_consum <-
  skipsdata %>%
  select("mannskap", "forbruk") %>%
  filter(mannskap=="A") %>%
  summarise(Averagesconsum = mean(forbruk))
```

*Gjennomsnittsforbruk for hele mannskap B*

```
totB_consum <-
  skipsdata %>%
  select(mannskap, forbruk) %>%
  filter(mannskap=="B") %>%
  summarise(Averageconsum = mean(forbruk))
```

*Gjennomsnittsforbruk for mannskap A, stigevinkel over eller lik 0.5*

```
sd1_Aconsum <-
  sd1 %>%
  select(mannskap, forbruk) %>%
  filter(mannskap=="A") %>%
  summarise(Averageconsum = mean(forbruk))
```

*Gjennomsnittsforbruk for mannskap B, stigevinkel over eller lik 0.5*

```
sd1_Bconsum <-
  sd1 %>%
  select(mannskap, forbruk) %>%
  filter(mannskap=="B") %>%
  summarise(Averageconsum = mean(forbruk))
```

*Gjennomsnittsforbruk for mannskap A, stigevinkel under 0.5*

```
sd2_Aconsum <-
  sd2 %>%
  select(mannskap, forbruk) %>%
  filter(mannskap=="A") %>%
  group_by(mannskap) %>%
  summarise(Averageconsum = mean(forbruk))
```

Gjennomsnittsforbruk for mannskap B, stigevinkel under 0.5

```
sd2_Bconsum <-
  sd2 %>%
  select(mannskap, forbruk) %>%
  filter(mannskap=="B") %>%
  summarise(Avargeconsum = mean(forbruk))
```

## Gjennomsnittsforbruk

Data	Mannskap A	Mannskap B	A-B
Datasetssettet	642.985227534237	566.875206183514	76.1100213507228
Stigevinkel < 0.5	610.926387629697	531.447878955816	79.478508673881
Stigevinkel $\geq$ 0.5	669.955405816641	597.723308362823	72.232097453818

Det man ser når man ser på gjennomsnittforbruk av drivstoff og gjennomsnittshastighet mellom mannskap A og B, er at mannskap A har en høyere gjennomsnittsforbruk med stigevinkel under og over 0.5. Som visst i tidligere oppgaver så har hastighet og forbruk en sammenheng, det gjør også at mannskap A har et høyere forbruk enn mannskap B. Stigevinkel har en påvirkning på forbruk, men også hastigheten til mannskap A er en avgjørende faktor for at de har et større forbruk enn mannskap B.

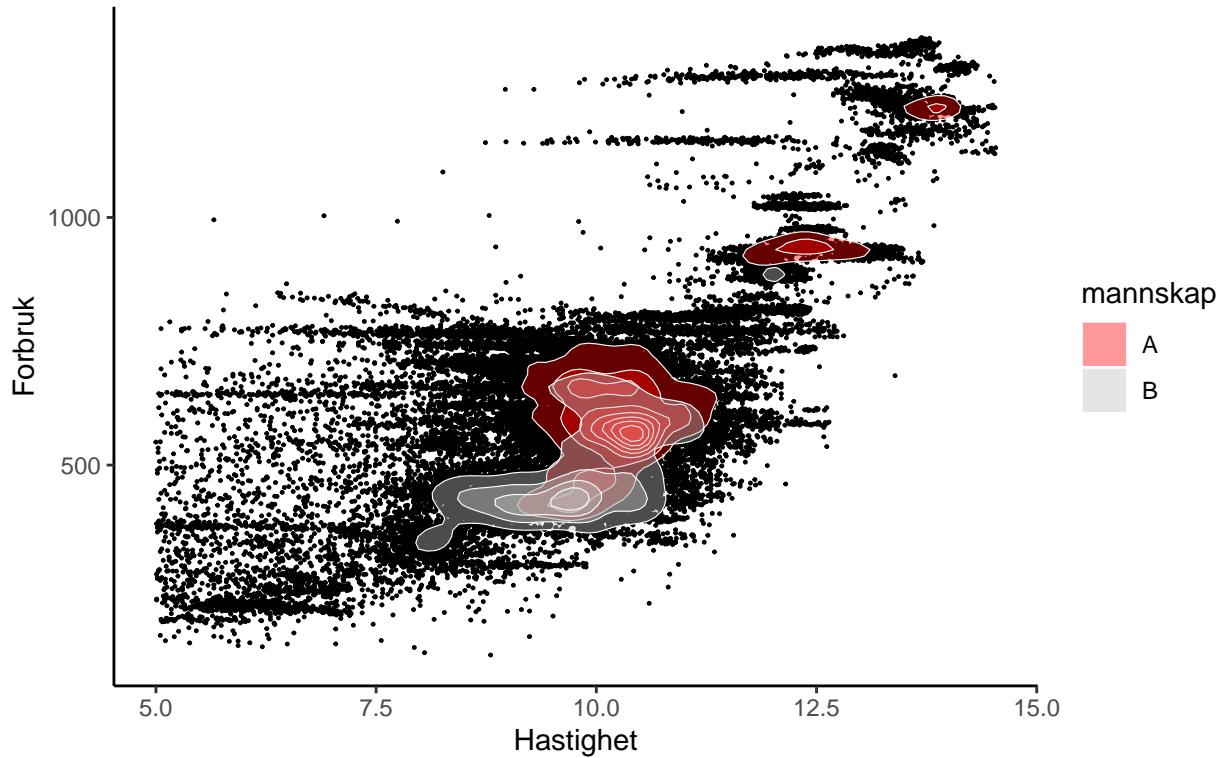
---

## Oppgave H)

Lager et tetthetsplot av forbruk og hastighet mellom mannskap A og B

```
skipssdata %>%
  ggplot(aes(x=hastighet, y=forbruk)) +
  geom_point(lwd=0.2) +
  stat_density_2d(aes(fill = mannskap),
  geom = "polygon", colour = "white", size=0.2, alpha=.4) +
  scale_fill_manual(values=c("red", "grey")) +
  theme_classic() +
  labs(title = "Forbruk og hastighet \n med mannskap A og B",
  y = "Forbruk",
  x = "Hastighet")
```

## Forbruk og hastighet med mannskap A og B



### Analyse:

På bakgrunn av hva som er kommet fram i oppgavene kan man se at at det er et klart skille mellom mannskapene. Der Mannskap A er det raskeste basert på at de har høyere gjennomsnittshastighet enn mannskap B, men har også et høyere forbruk per tur som er høyere enn mannskap B. Mannskap A har også på sine turer brukt mer drivstopp i gjennomsnitt med tanke på stigevinkel, som har gjort at de har et større gjennomsnittligforbruk av drivstoff enn mannskap B. Det er en forskjell i arbeidsmetodene til mannskap A og B, siden det er forskjell på både hastighet og forbruk mellom mannskapene.

Når det gjelder effektivitet til mannskapene er det vanskelig å si hvem av dem som er mest effektiv. Mannskap A kan ha en klar fordel med å kjøre fortare, men det er ingen data som sier hvor mye de potensielt tjener inn på å kjøre fortare enn mannskap B, med tanke på at dem også bruker mer drivstoff.

```

forbruk_kr_pertime_A <-
  totA_consum*4.90

forbruk_kr_pertime_B <-
  totB_consum * 4.90

d <- ((totA_consum*4.90) *1500)
b <- ((totB_consum*4.90) * 1500)
  
```

Vis man regner på hvor mye mannskapene bruker av drivstoff i snitt per time vil mannskap B ha et gjennomsnittsforbruk av drivstoff på 566.875206183514, mens Mannskap A vil ha på

642.985227534237.

Dette vil si at mannskap B har en drivstoffkostnad per time 2777.68851029922 og mannskap A har 3150.62761491776 noe som vil tilsvare en kostnad for B og A på henholdsvis 4166532.76544883 og 4725941.42237664 på 1500 timer.

Dette vil da gi en kostnadsdifferanse mellom mannskapene A og B på 559408.656927813

```
forbruk_perknop_A <-
  ((totA_consum/totA))
```

```
forbruk_perknop_B <-
  (totB_consum/totB)
```

```
effektivitetA <-
  ((forbruk_perknop_A * 1500) * 4.9)
effektivitetB <-
  ((forbruk_perknop_B * 1500)* 4.9)
```

Hvis man skal ta utgangspunkt i hvilket mannskap som er mest effektivt, så vil man se på hvor mange liter som blir brukt per. knop de kjører, som er for Mannskap B 58.067139812713 og mannskap A 62.5810967593262.

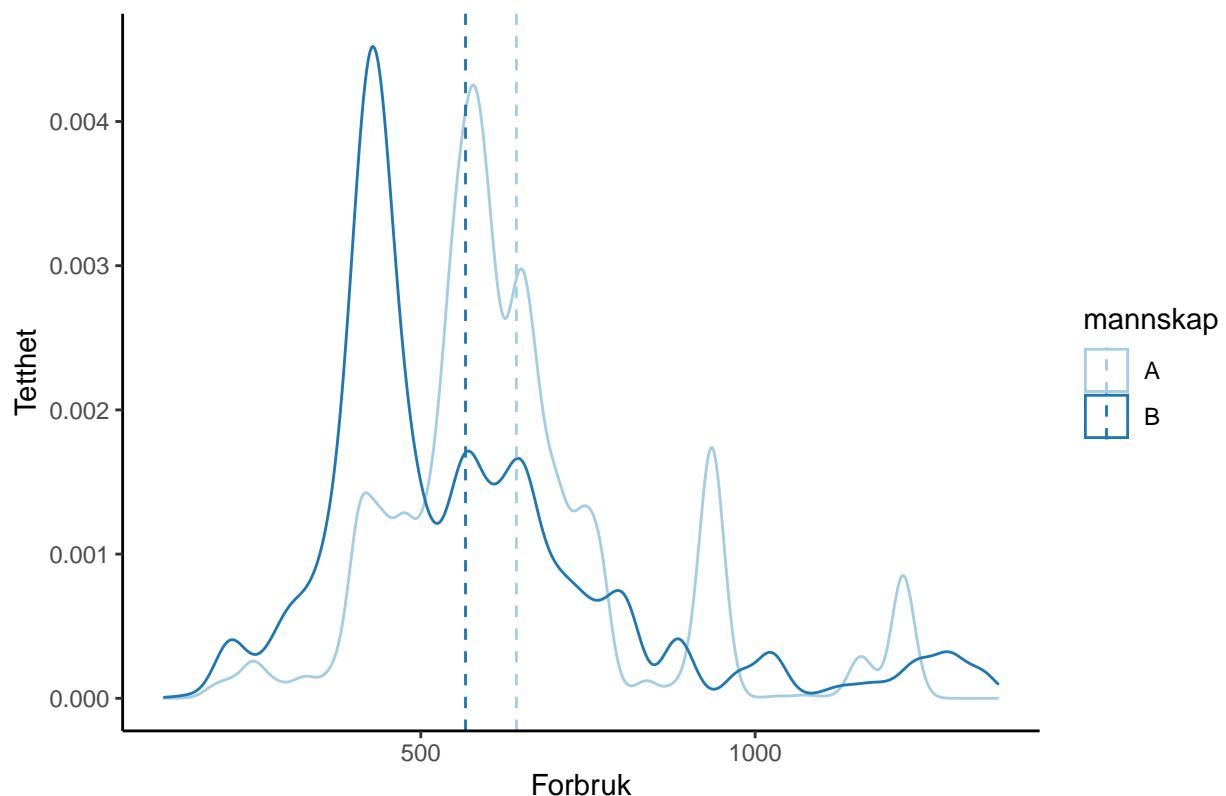
Hvis man da legger til kjøretimer og pris for per liter drivstoff vil mannskap B ha en kostnad på 426793.477623441 og mannskap A 459971.061181047, man kan se at mannskap B er effektive på kostandene med å holde en litt lavere hastighet enn mannskap A, noe som gjør mannskap B til det mest effektive og det mannskapet som gjør at rederiet kan spare 33177.5835576066 kr i året.

Nettoeffekt av en endring i balasten ved gangtid så vil det for mannskap B betyr en redusering i gjennomsnittforbruket med 538.531445874338

```
mannskapgjennomsnittforbruk <-
  ddply(skipsdata, "mannskap", summarise, grp.mean=mean(forbruk))

plot92 <-
  ggplot(skipsdata, aes(x=forbruk, color=mannskap)) +
  theme_bw() +
  geom_density() +
  geom_vline(data = mannskapgjennomsnittforbruk, aes(xintercept=grp.mean, color=mannskap),
             linetype = "dashed", size= 0.5) +
  scale_color_brewer(palette = "Paired") +
  theme_classic() +
  labs(title="Gjennomsnittsforbruk mannskap A og B",
       y="Tetthet",
       x="Forbruk")
plot92
```

### Gjennomsnittsforbruk mannskap A og B



## DEL 2:

### Oppgave A)

Laster ned pakke, laget et langt datasett:

```
load(url("https://bit.ly/3BcnSXA"))
library(tidyverse)

liten <-
  liten %>%
  mutate(motorkombinasjon = "liten")

stor <-
  stor %>%
  mutate(motorkombinasjon = "stor")

liten_stor <-
  liten_stor %>%
  mutate(motorkombinasjon = "liten_stor")

motorkombinasjon <- rbind(liten, stor, liten_stor)
```

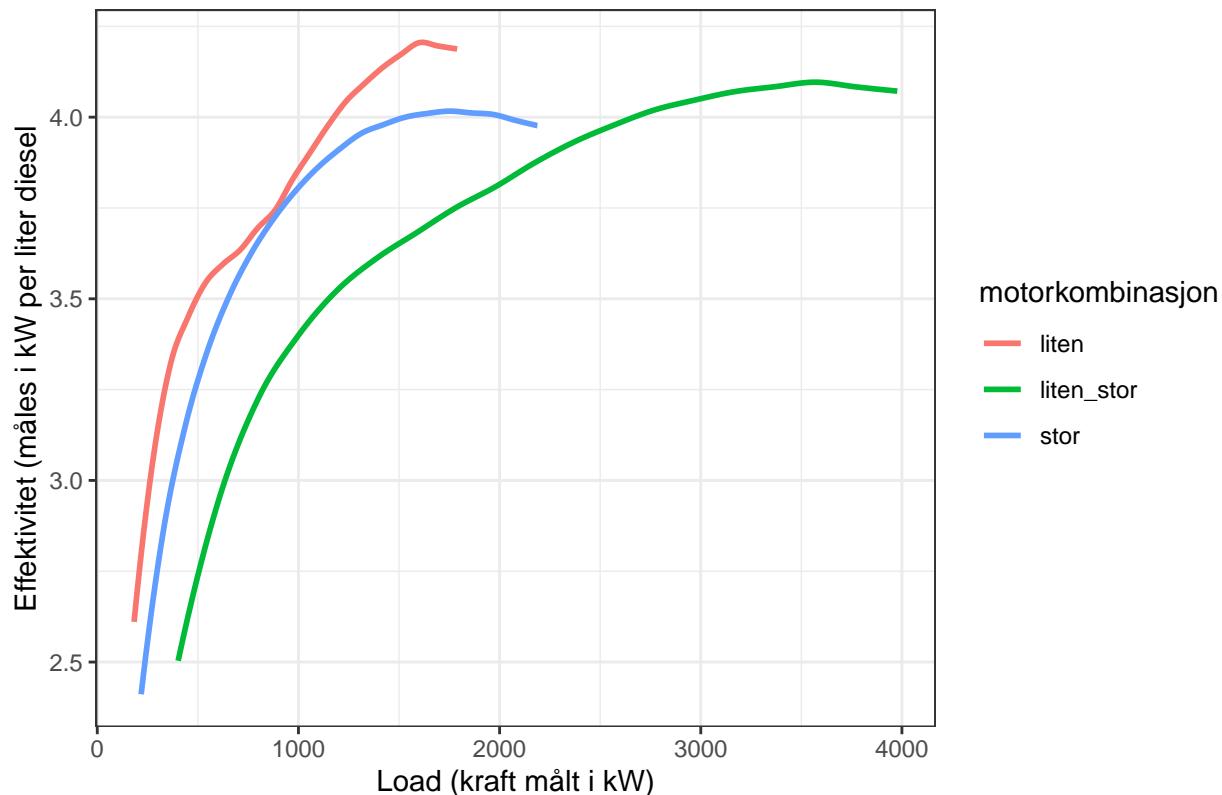
---

### Oppgave B)

Laget et plot av samspill av motorene:

```
motorkombinasjon %>%
  ggplot(aes(load, effektivitet, colour=motorkombinasjon,
             shape=(motorkombinasjon))) +
  geom_line(lwd=1) +
  theme_bw() +
  labs(title = "Samspill mellom load og effektivitet",
       y= "Effektivitet (måles i kW per liter diesel",
       x="Load (kraft målt i kW)")
```

## Samspill mellom load og effektivitet



I dette plottet kan man se at at liten\_stor motor er den mest effektive kombinasjonen, siden den kan belastet med en høyere kraft og samtidig har et relativt lavt forbruk av diesel. Dette kan man se med at effektiviteten (som måles i kW per liter drivstoff) går opp jo mer load (belasting målt i kW) det er på motoren. Hvis man skal kjøre på tomgang så vil den storemotoren være best egent, siden den har et høyere effektivistetsnivå fordi den utnytter dieselen bedre på et lavt belastningsnivå. Den liten\_store er den mest økonomiske kombinasjøen siden den kan tåle et belastningsnivå over lengre tid, siden den bruker mye mindre diesel enn de andre motorene for å produsere kraft, noe som gjør at helningen på effektivitskurven blir slakkere enn på de andre motorene ved høyere load.

---

### Opgave C)

a)

```
library(gridExtra)

##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##   combine
stor1 <-
  max(stor$effektivitet)
```

```

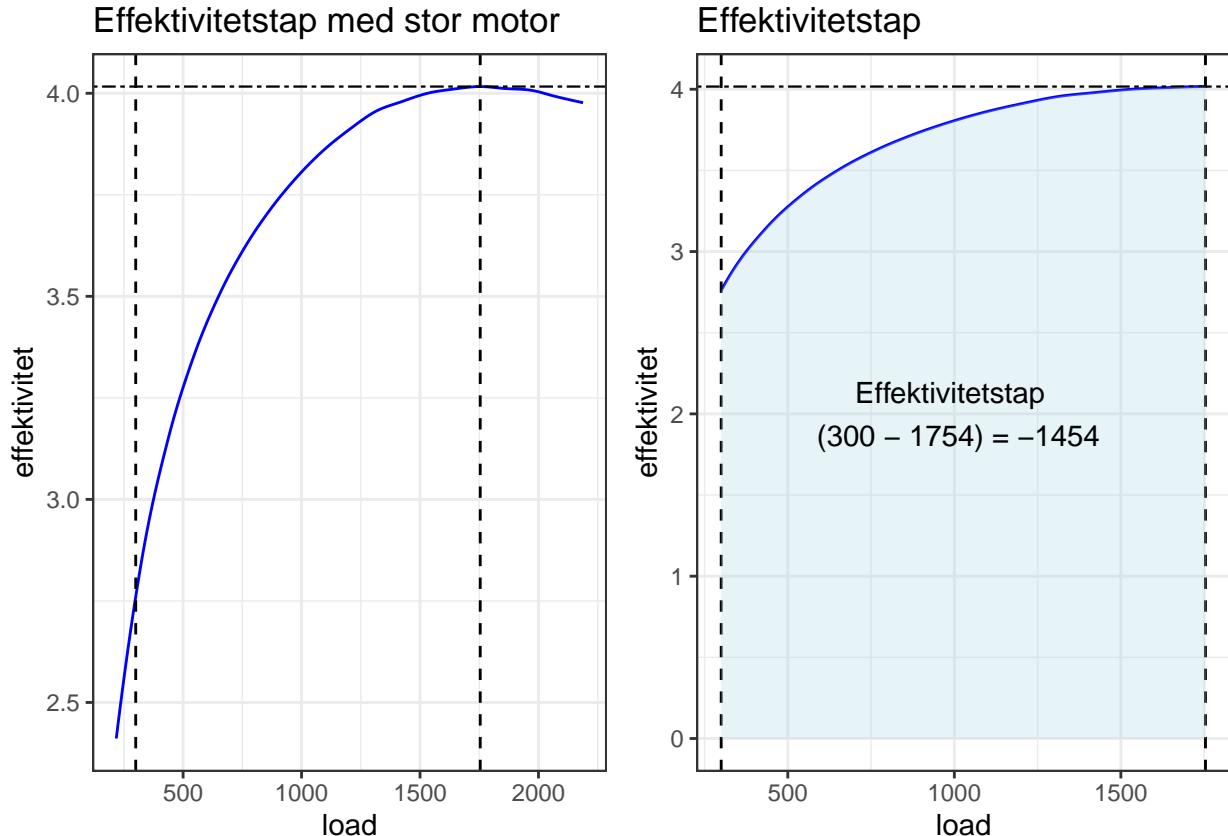
plotstor <-
  stor %>%
  ggplot(aes(x = load, y = effektivitet), lwd=2) + geom_line(color = "blue") +
  geom_vline(xintercept = 300, linetype = "dashed") +
  geom_hline(yintercept = stor1, linetype = "twodash", lwd = 0.4) +
  geom_vline(xintercept = 1754, linetype = "dashed", lwd = 0.5) +
  theme_bw() + ggtitle("Effektivitetstap a \n med stor motor") +
  ggtitle("Effektivitetstap med stor motor")

plot2stor <-
  stor %>%
  ggplot(aes(x= load, y = effektivitet)) + geom_line(color="blue") +
  theme_bw() + geom_vline(xintercept = 300, linetype = "dashed") +
  geom_hline(yintercept = stor1, linetype = "twodash", lwd = 0.4) +
  geom_vline(xintercept = 1754, linetype = "dashed", lwd = 0.5) +
  geom_area(aes(x=load), fill = "lightblue", alpha = 0.3) + xlim(300, 1756) +
  annotate("text", x = 1000, y = 2, label = "Effektivitetstap \n (300 - 1754) = -1454") +
  ggtitle("Effektivitetstap")

grid.arrange(plotstor, plot2stor, ncol = 2)

## Warning: Removed 514 rows containing missing values (position_stack).
## Warning: Removed 514 row(s) containing missing values (geom_path).

```



I disse plotene ser man effektivitetstapet på 1454. Hvis man skulle gått på 300kW så ville det være langt under den effektiviteten som hadde vært optimalt på 1754kW, før effektiviteten

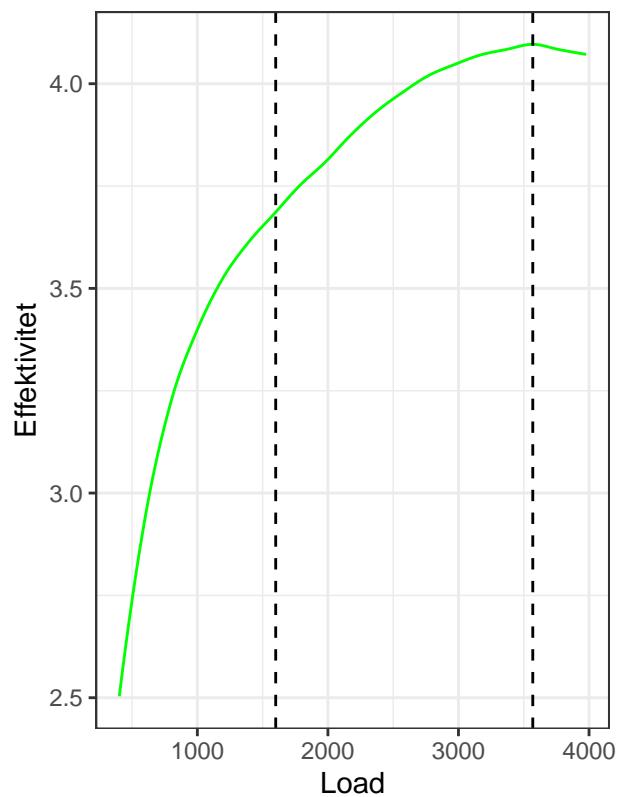
faller litt igjen. Med stor motor ville det da vært mest effektivt å ligge på 1753kW, og vis man da skulle ha 5000 timer på 300kW ville det da være langt under nivå for mest effektivitet

---

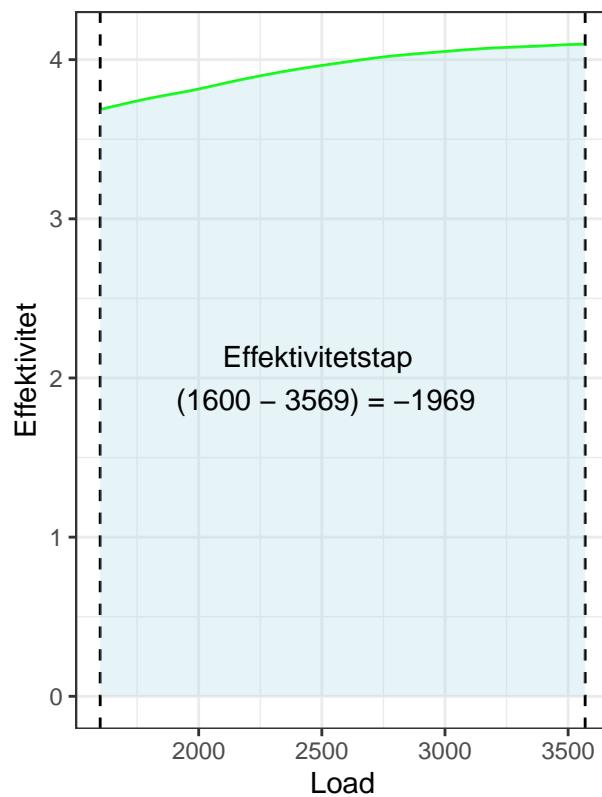
b)

```
stor_liten1 <-  
  max(liten_stor$effektivitet)  
  
plotlitenstor <-  
  liten_stor %>%  
  ggplot(aes(x= load, y = effektivitet)) + geom_line(color = "green") +  
  theme_bw() + geom_vline(xintercept = 1600, linetype = "dashed") +  
  geom_vline(xintercept = 3569, linetype = "dashed", lwd = 0.5) +  
  labs(title = "Effektivitetstap med liten_stor motor",  
       y = "Effektivitet",  
       x = "Load")  
  
plot2litenstor <-  
  liten_stor %>%  
  ggplot(aes(x= load, y = effektivitet)) + geom_line(color = "green") +  
  theme_bw() + geom_vline(xintercept = 1600, linetype = "dashed") +  
  geom_vline(xintercept = 3569, linetype = "dashed", lwd = 0.5) +  
  geom_area(aes(x=load), fill = "lightblue", alpha = 0.3) + xlim(1600, 3569) +  
  annotate("text", x = 2500, y = 2, label = "Effektivitetstap \n (1600 - 3569) = -1969") +  
  labs(title = "Effektivitetstap",  
       y="Effektivitet",  
       x="Load")  
  
grid.arrange(plotlitenstor, plot2litenstor, ncol = 2)  
  
## Warning: Removed 1606 rows containing missing values (position_stack).  
## Warning: Removed 1606 row(s) containing missing values (geom_path).
```

Effektivitetstap med liten\_stor motor



Effektivitetstap



Her ser man i plottene at en motorkraft på 1600 kW på kombinasjonen liten\_stor så ville det vært et stykke under optimalt punkt som er 3569kW, så det betyr at på 1600kW så er man et godt stykke under effektivitetsoptimalen, og effektivitetstapet ved å gå 7000 timer på 1600kW vil være kW differansen på 1969kW.

c)

```
stor_liten1 <-  
  max(liten_stor$effektivitet)  
  
plot3litenor <-  
  liten_stor %>%  
  ggplot(aes(x= load, y = effektivitet)) + geom_line(color = "green") +  
  theme_bw() + geom_vline(xintercept = 2400, linetype = "dashed") +  
  geom_vline(xintercept = 3569, linetype = "dashed", lwd = 0.5) +  
  labs(title="Effektivitetstap med liten_stor motor",  
       y = "Effektivitet",  
       x = "Load")  
  
plot4litenor <-  
  liten_stor %>%  
  ggplot(aes(x= load, y = effektivitet)) + geom_line(color = "green") +  
  theme_bw() + geom_vline(xintercept = 2400, linetype = "dashed") +  
  geom_vline(xintercept = 3569, linetype = "dashed", lwd = 0.5) +
```

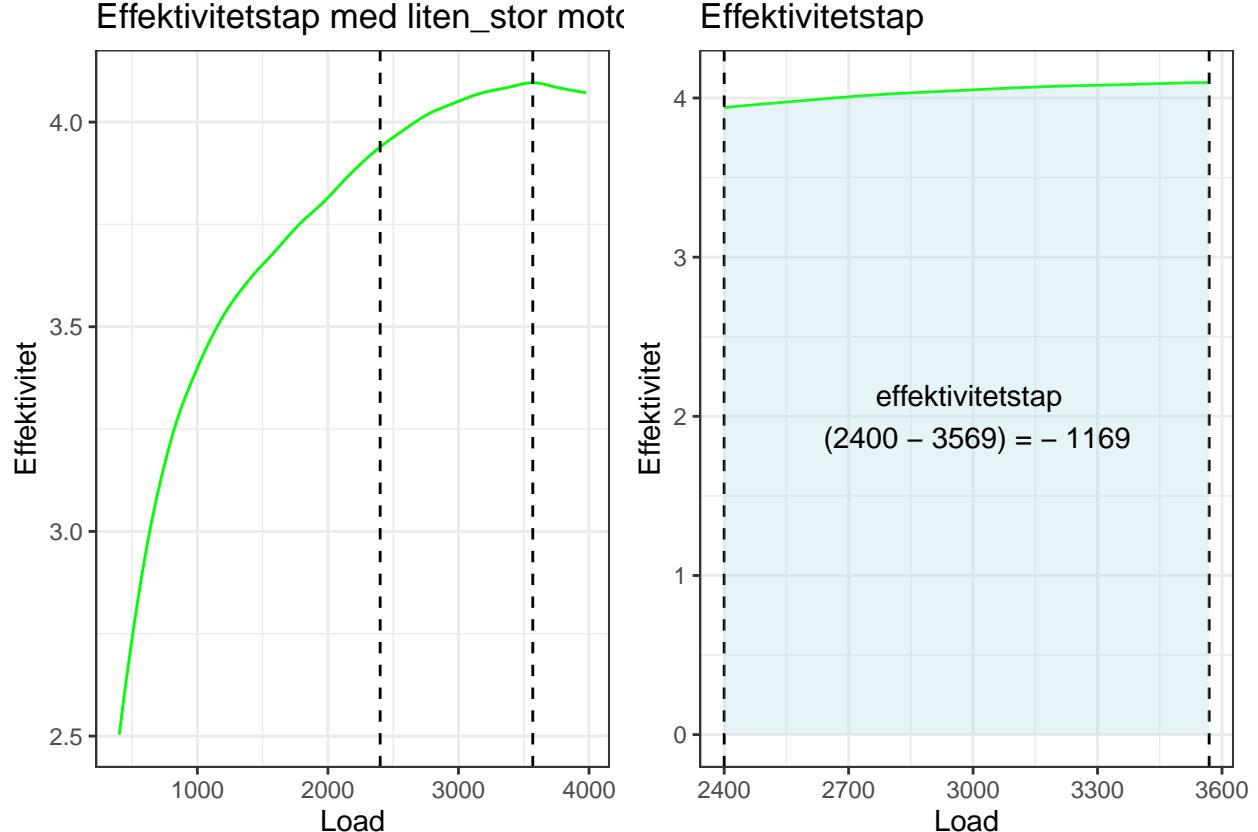
```

geom_area(aes(x=load), fill = "lightblue", alpha = 0.3) + xlim(2400, 3569) +
annotate("text", x = 3000, y = 2, label = "effektivitetstap \n (2400 - 3569) = - 1169") +
labs(title = "Effektivitetstap",
y = "Effektivitet",
x = "Load")

grid.arrange(plot3litenstor, plot4litenstor, ncol = 2)

## Warning: Removed 2406 rows containing missing values (position_stack).
## Warning: Removed 2406 row(s) containing missing values (geom_path).

```



Her ser man i plottene at en motorkraft på 2400 kW på kombinasjonen liten\_stor så ville det vært et stykke under optimalt punkt som er 3560kW, så det betyr at på 2400kW så er man et godt stykke under effektivitetsoptimalen, og effektivitetstapet ved å gå 10000 timer på 2400kW vil være kW differansen på 1160kW, som er nær optimal kW

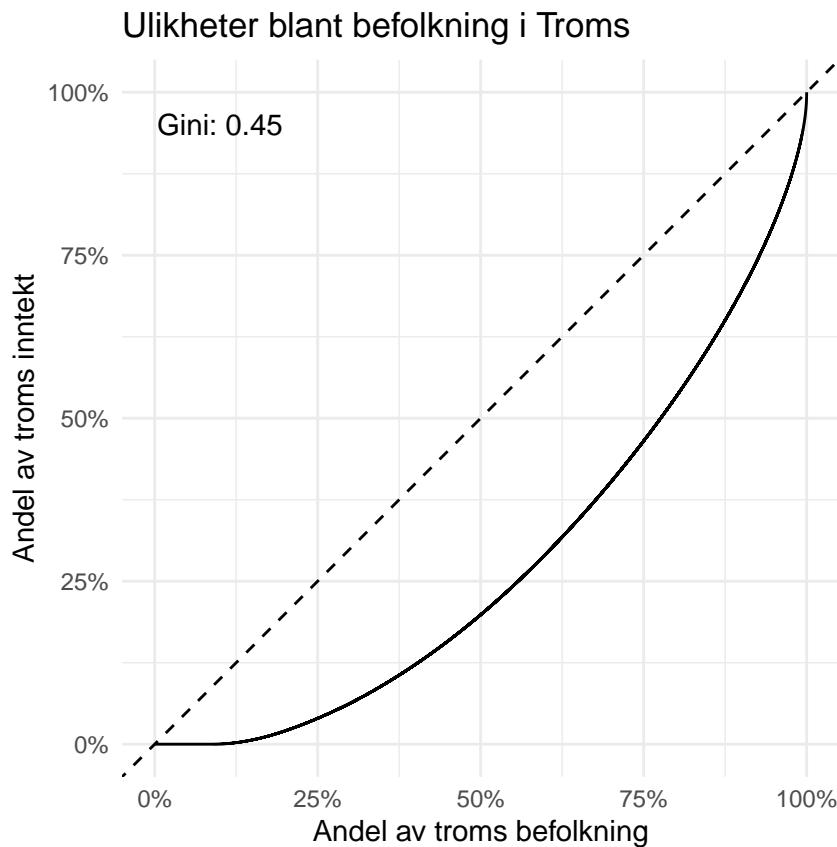
## DEL B

### Oppgave A)

```
library(tidyverse)
library(gglorenz)
library(ineq)

load(url("https://bit.ly/2YBntjg"))

skattetall %>%
  ggplot(aes(innnett)) +
  stat_lorenz(desc = FALSE) +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs (x = "Andel av troms befolkning",
        y = "Andel av troms innnett",
        title = "Ulikheter blant befolkning i Troms") +
  annotate_ineq(skattetall$innnett) +
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(labels = scales::percent)
```



```
gjennomsnittsinnnett <-
  skattetall %>%
```

```

  summarise(gjennomsnittsinntekt = mean(inntekt))

Gini_koffedensienten <-
  ineq(skattetall$inntekt) * 100

```

Fylke	Gini(%)	Inntekt	Personer
Troms	44.5220612	286641.956295435	138 521

En Lorenz-kurve viser grafisk hvor stor andel av samlet inntekt som tilfaller en andel av befolkningen i Troms, når befolkningen er ordnet etter stigende inntekt. På horisontalaksen viser det andeler av befolkningen rangert etter inntekt, den vertikale asken så måles det prosent av samlet inntekt.

Vis man f.eks. velger 0,1 på den horisontalaksen, vil da punktet på Lorenzkurven vise hvor stor andel av samlet inntekt som tilhører de fattigste ti prosenten av Troms befolkningen. I et samfunn der alle hadde hatt lik inntekt, ville da Lorenz-kurven vært i 45-graderslinje som betyr at de fattigste 10 prosentene ville hatt 10 prosent av inntekten osv. Jo lengre vekke Lorenz-kurven er fra 45-graderslinjen, desto mer ulik er inntektsfordelingen

Gini-konfidenten er et tall mellom 0-1, der 0 gir full fordeling av samfunnets inntekt, og 1 er at det er størst ulikhet mellom fordeling av samfunnets inntekt. Troms sin gini er på 0,45, som betyr at det er middels fordelinger av fylkets inntekter.

---

## Oppgave B)

Data fra Tromsø:

```

skattetalltromsø <-
  skattetall %>%
  filter(kommnr == "1902")

gjennomsnittsinntekt.T <-
  skattetalltromsø %>%
  summarise(gjennomsnittsinntekt.T = mean(inntekt))

Gini_tromsø <-
  ineq(skattetalltromsø$inntekt) * 100

befolknings_tromsø <-
  nrow(skattetalltromsø)

andelstromsø <-
  (befolknings_tromsø / 138521) * 100

```

Data fra Harstad

```

skattetallHarstad <-
  skattetall %>%
  filter(kommnr == "1903")

gjennomsnittsinntekt.H <-

```

```

skattetallHarstad %>%
  summarise(gjennomsnittsinntekt.H = mean(inntekt))

Gini_harstad <-
  ineq(skattetallHarstad$inntekt) * 100

befolkning_harstad <-
  nrow(skattetallHarstad)

andelharstad <-
  (befolkning_harstad / 138521) * 100

```

Data fra Omegn:

```

skattetall0megn <-
  skattetall %>%
  filter(kommnr > "1903")

gjennomsnittsinntekt.0 <-
  skattetall0megn %>%
  summarise(gjennomsnittsinntekt.0 = mean(inntekt))

Gini_omegn <-
  ineq(skattetall0megn$inntekt) * 100

befolkning_omegn <-
  nrow(skattetall0megn)

andelomegn <-
  (befolkning_omegn/138521) * 100

```

Kommune	Gini(%)	Inntekt	Personer	Andel (%)
Tromsø	46.6381364	297756.0530297	62795	45.3324767
Harstad	43.2945417	299649.962712367	20248	14.6172782
Omegn	42.012701	269314.442355528	55478	40.0502451

Her kan man se at det er større skjevføring i Tromsø kommune, enn det er i kommunene Omegn og Harstad kommune. På grunn av at Tromsø og Harstad har mange millionærer, så gir det en større grad av skjevføring av inntekt i forhold til Omegn. Dette fører også til at gjennomsnittsinntekten blir større i Tromsø og Harstad.

Sammenlignet med oppgave A, så ligger Harstad og Omegn sin gini under fylkets gini, mens Tromsø sin gini ligger over. Dvs. at det er større skjevføring i Tromsø enn fylke generelt. Mens i Harstad og Omegn er det mindre skjevføring enn i Tromsø fylke.

## Oppgave C)

```
library(cowplot)
```

```

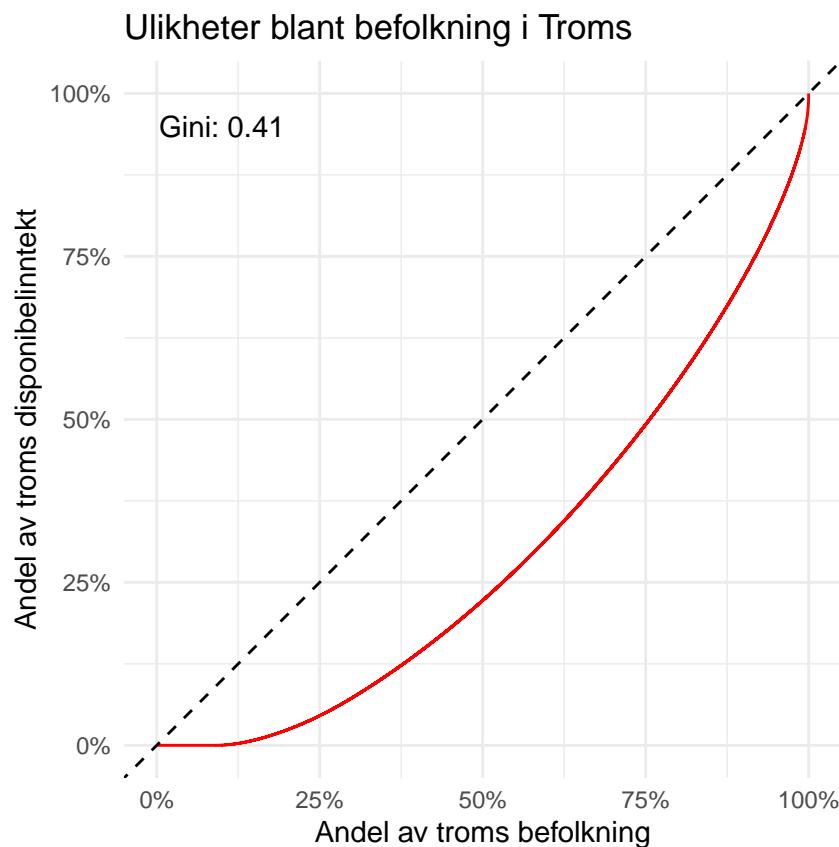
skattetall1 <-
  skattetall %>%
  mutate(disponibelinntekt = (inntekt - skatt))

skattetall1[skattetall1 < 0] <- 0

LC2 <-
  skattetall1 %>%
  ggplot(aes(disponibelinntekt)) +
  stat_lorenz(desc = FALSE, colour = "red") +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "Andel av Troms befolkning",
       y = "Andel av Troms disponibelinntekt",
       title = "Ulikheter blandt befolkning i Troms") +
  annotate_ineq(skattetall1$disponibelinntekt) +
  scale_y_continuous(labels = scales::percent) +
  scale_x_continuous(labels = scales::percent)

```

LC2



```

library(gridExtra)

LC1 <-
  skattetall %>%
  ggplot(aes(inntekt)) +
  stat_lorenz(desc = FALSE, colour = "blue") +

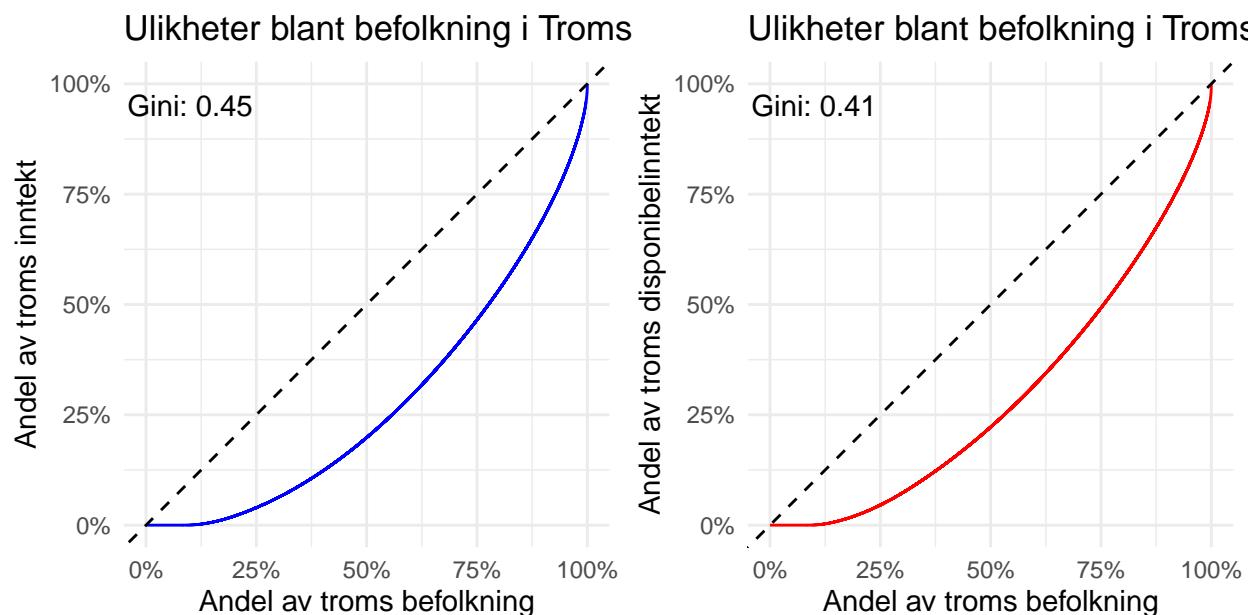
```

```

coord_fixed() +
geom_abline(linetype = "dashed") +
theme_minimal() +
labs(x = "Andel av troms befolkning",
y = "Andel av troms inntekt",
title = "Ulikheter blant befolkning i Troms") +
annotate_ineq(skattetall$inntekt) +
scale_y_continuous(labels = scales::percent) +
scale_x_continuous(labels = scales::percent)

grid.arrange(LC1, LC2, ncol = 2)

```



Disponibilinntekt er “inntekt - skatt”, og er sett på av mange økonominer som en bedre indikator på å beregne Gini, fordi det viser hvor mye penger man faktisk har å konsumere for. Vis man ser på disponibilinntekt, så er Gini-en lavere enn ved inntekt, noe som indikerer at disponibilinntekt er bedre fordelt i samfunnet enn inntekt. Siden disponibilinntekt sin gini er lavere så vil det si at man skatter mer ved høyere inntekter.

#### Oppgave D)

```

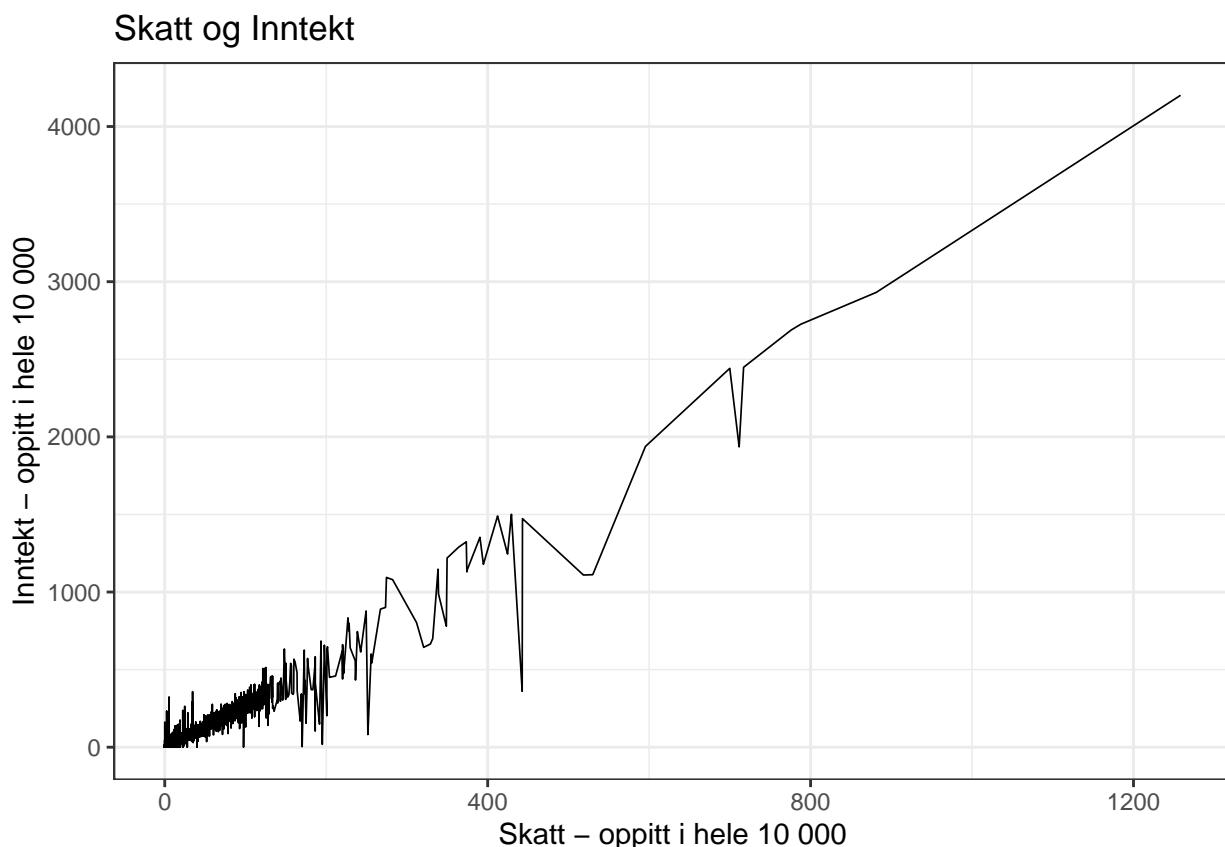
skattetall %>%
  mutate(skatt=skatt/1e4) %>%
  mutate(inntekt=inntekt/1e4) %>%

```

```

ggplot(aes(x = skatt, y = inntekt)) +
  geom_line(lwd=0.3) +
  theme_bw() +
  labs(title="Skatt og Inntekt",
       y="Inntekt - oppitt i hele 10 000",
       x="Skatt - oppitt i hele 10 000")

```



Her kan man se at mye av skatteinntektene kommer fra dem som tjener under 1 million kr, man kan også se at det er noen som ikke tjener så meget men skatter veldig mye, dette kan tyde på at det er skatt fra fomue som gjør at skatteinntekten blir stor. Ellers så ser man at de som tjener mest, skatter mest. Med unntak av unntakene så ser man at kurven er nesten linær, så det betyr at det er et relativt forhold mellom inntekt og skatt.

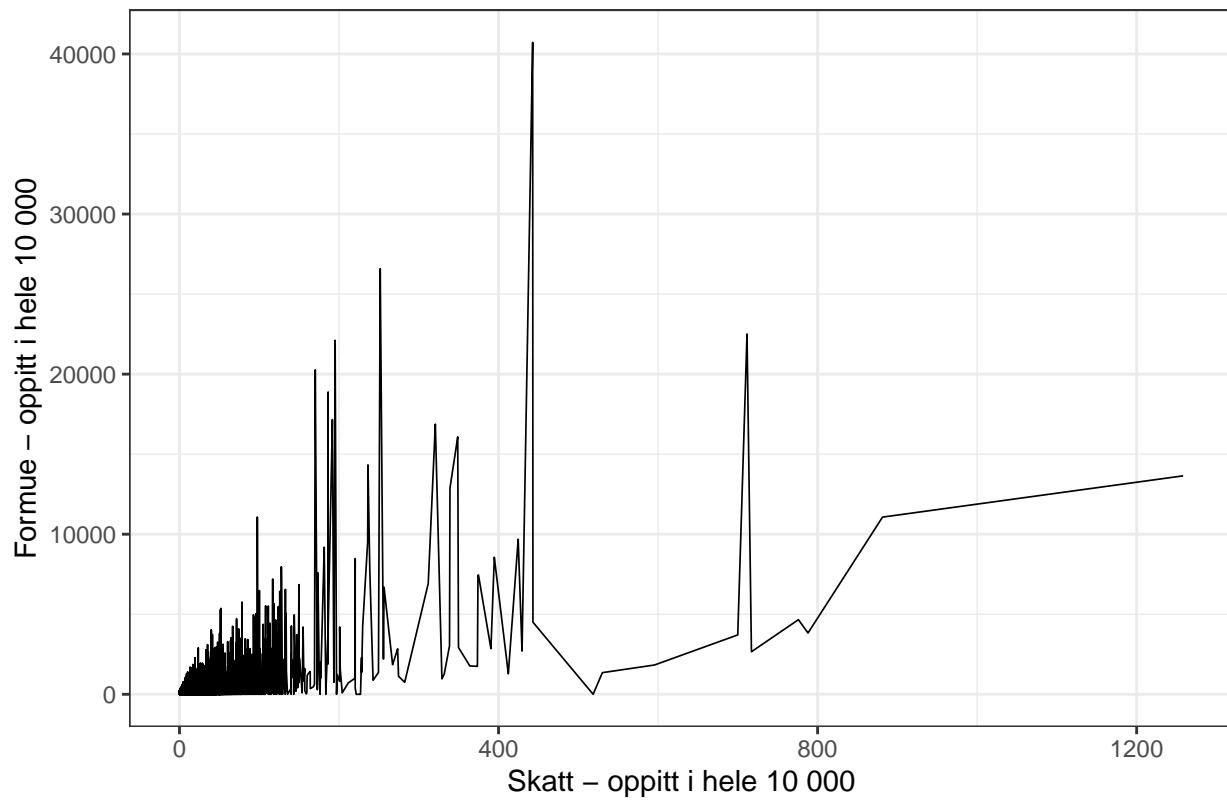
### Oppgave E)

```

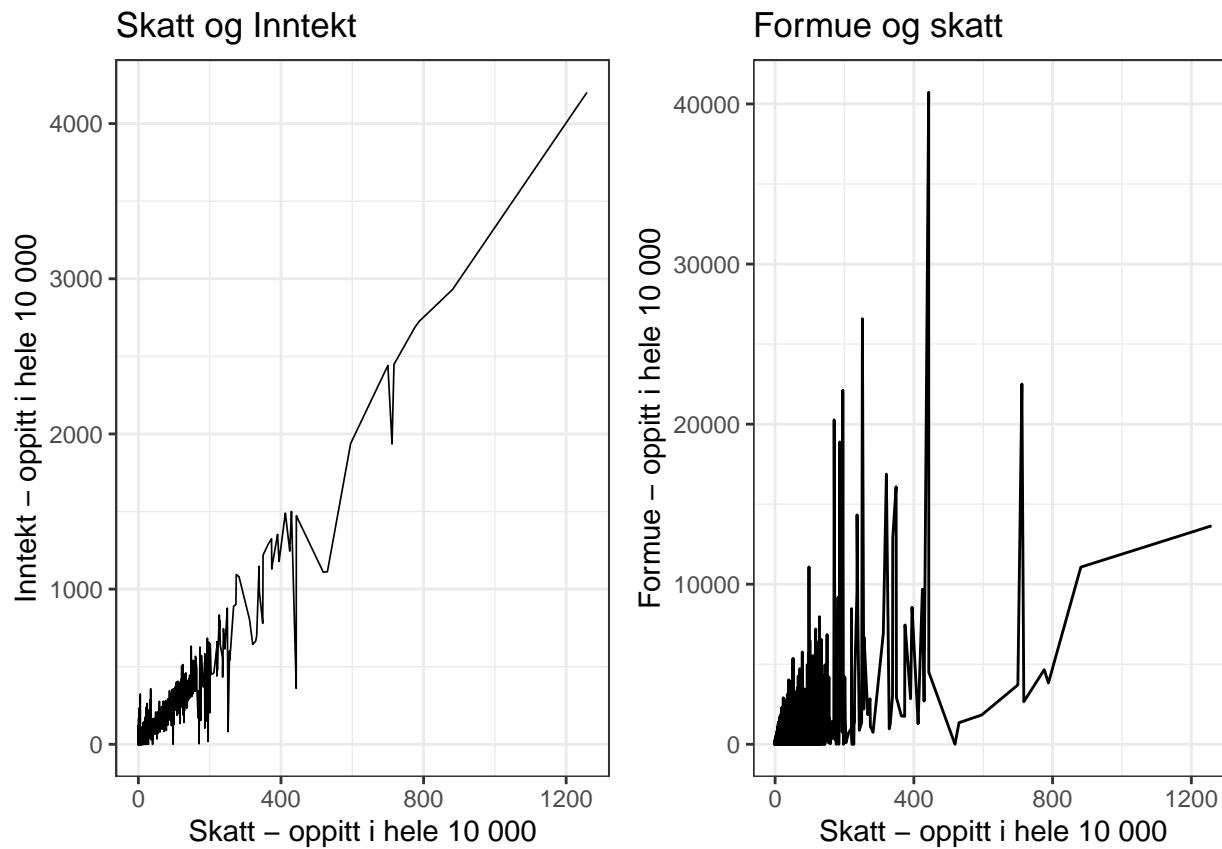
skattetall %>%
  mutate(formue=formue/1e4) %>%
  mutate(skatt=skatt/1e4) %>%
  ggplot(aes(x = skatt, y = formue)) + geom_line(lwd=0.3) +
  theme_bw() +
  labs(title="Formue og skatt",
       y="Formue - oppitt i hele 10 000",
       x="Skatt - oppitt i hele 10 000")

```

## Formue og skatt



```
skattoginntekt <-
  skattetall %>%
  mutate(skatt=skatt/1e4) %>%
  mutate(inntekt=inntekt/1e4) %>%
  ggplot(aes(x = skatt, y = inntekt)) +
  geom_line(lwd=0.3) + theme_bw() +
  labs(title="Skatt og Inntekt",
       y="Inntekt – opprett i hele 10 000",
       x="Skatt – opprett i hele 10 000")
formueogskatt <-
  skattetall %>%
  mutate(formue=formue/1e4) %>%
  mutate(skatt=skatt/1e4) %>%
  ggplot(aes(x = skatt, y = formue)) + geom_line() +
  theme_bw() +
  labs(title="Formue og skatt",
       y="Formue – opprett i hele 10 000",
       x="Skatt – opprett i hele 10 000")
grid.arrange(skattoginntekt, formueogskatt, ncol = 2)
```



Her kan man klart se samfunnets skjevfordelinger av skatt. Ploten formue/skatt viser at dem som har høyest formue, betaler relativt lite skatt, mens dem som har mindre formue, betaler relativt mye skatt på formuen. Dette indikerer at forholdet mellom skatt og formue er veldig skjevfordelt og sammenlignet med skatt/inntekt så er det klare forskjeller på skattenivået, sånn som i skatt/inntekt der forholdet er nokså likt, mens i formue/skatt er det veldig store variasjoner og relativt ujevne forhold.

### Oppgave F)

Gjennomsnittlig formue:

```
Formuetromsø <-
  skattetalltromsø %>%
  summarise(Formuetromsø = mean(formue))

Formueharstad <-
  skattetallHarstad %>%
  summarise(Formueharstad = mean(formue))

Formueomegn <-
  skattetallOmegn %>%
  summarise(Formueomegn = mean(formue))

Formuetroms <-
```

```
skattetall %>%
  summarise(Formuetroms = mean(formue))
```

Gini på formue:

```
Giniformuetromsø <-
  ineq(skattetalltromsø$formue) * 100

Giniformueharstad <-
  ineq(skattetallHarstad$formue) * 100

Giniformueomegn <-
  ineq(skattetall0megn$formue) * 100

Giniformuetroms <-
  ineq(skattetall$formue) * 100
```

Kommune	Gini(%)	Formue	Personer	Andel (%)
Tromsø	86.1965547	491493.342161	62795	45.3324767
Harstad	83.9812301	461465.760815883	20248	14.6172782
Omegn	82.4565137	441224.211921843	55478	40.0502451
Troms	84.5530948	466971.217180067	138521	100

I denne tabellen så ser man at det generelt er en høy skjevfordeling av formue i kommune og fylke. I likhet med gini på inntekten er også Tromsø kommune i denne omgang værst på gini med formue, og ligger over fylkesnittet. Harstad og Omegn har igjen kommet best ut av det, og viser at de har knapp lavere skejvfordeling enn Tromsø kommune og snittet for fylke.

Sammenlignet med a) og b) så ser man at formuefordelingen er ekstremt skjevfordelt i forhold til inntektsfordelingen i kommunene og i fylket. Det vil si at det er et utvalg mennesker som har en høy formue, mens resten av samfunnet har en lavere formue. Man kan se at den største formuefordelingen ligger i Tromsø kommune, som ligger over fylkessnittet mens Harstad og Omegn ligger under fylkesnittet.

## Oppgave G)

```
library(reldist)

## reldist: Relative Distribution Methods
## Version 1.6-6 created on 2016-10-07.
## copyright (c) 2003, Mark S. Handcock, University of California-Los Angeles
## For citation information, type citation("reldist").
## Type help(package="reldist") to get started.

##
## Attaching package: 'reldist'

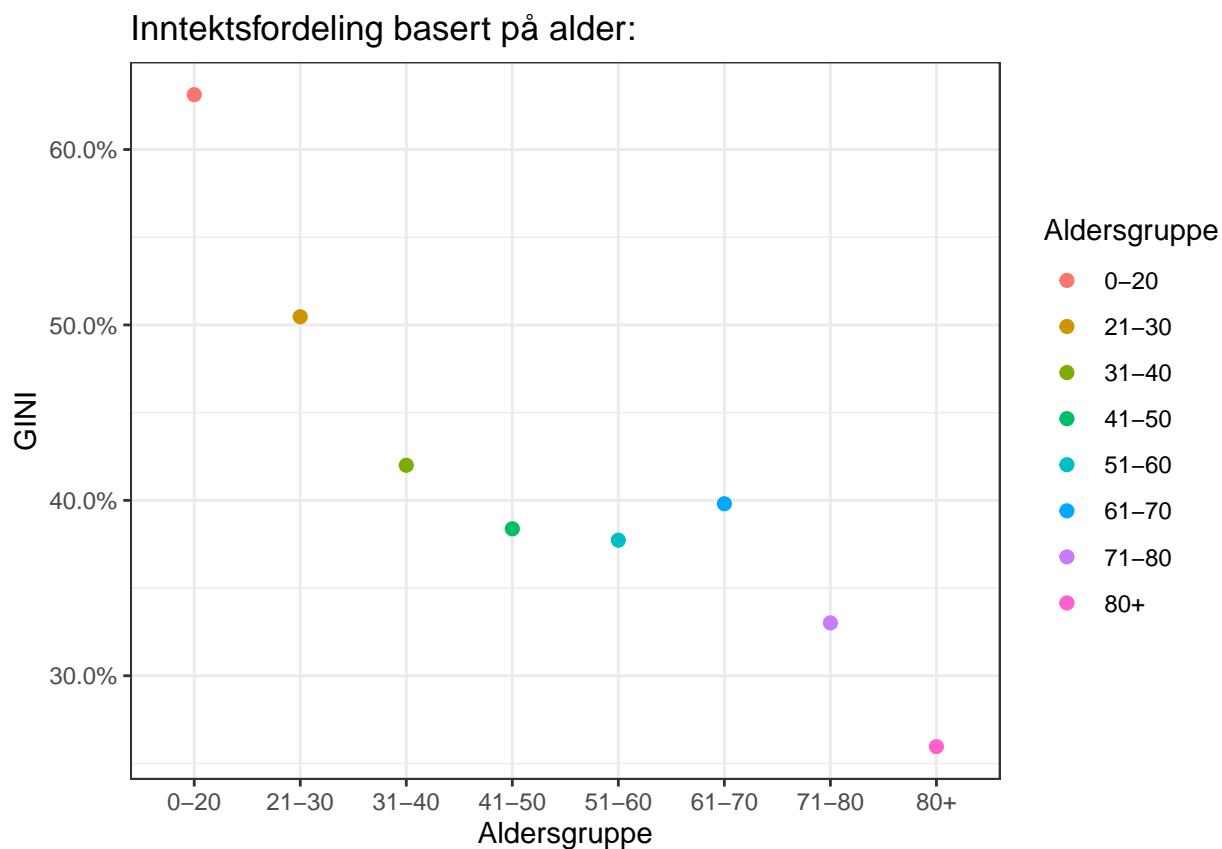
## The following object is masked from 'package:ineq':
## 
##      entropy
```

```

gini_alder <- aggregate(inntekt ~ aldersgruppe,
                        data = skattetall,
                        FUN="gini")
names(gini_alder) <- c("Aldersgruppe", "GINI")

gini_alder %>%
  ggplot(aes(x=Aldersgruppe, y=GINI, colour = Aldersgruppe)) +
  geom_point(lwd = 2) +
  theme_bw() +
  scale_y_continuous(labels = scales::percent) +
  ggtitle("Inntektsfordeling basert på alder:")

```



Her kan man se gini på inntekt fordelt over aldersgruppene. Det er en høy grad av skjevfordeling i de to yngste aldersgruppene, noe som kan ha en sammenheng med at mangen i denne alderen holder på med studier og grunnskole. Gini-inntekt ligger langt over gini-inntektsnittet i Troms.

Skjevfordelingen ser ut til å flattne mer ut når aldersgruppene som kan være yrkesaktive inntreffer og man kan se at aldersgruppen 31-40 ligger nesten på gini-inntektsnittet for Troms. De tre siste aldersgruppene er det relativt lav skjevfordeling og ligger under snittet for gini-inntektsnittet i Troms, noe som kan ha en sammenheng med at dette er de aldersgruppene der pensjonstilværelsen inntreffer.

## Oppgave H)

```
skattetall1 <-
  subset(skattetall, skattetall$kjonn != "N")

dfMale <- skattetall1[which(skattetall$kjonn == "M"),]
dfFemale <- skattetall1[which(skattetall$kjonn == "F"),]

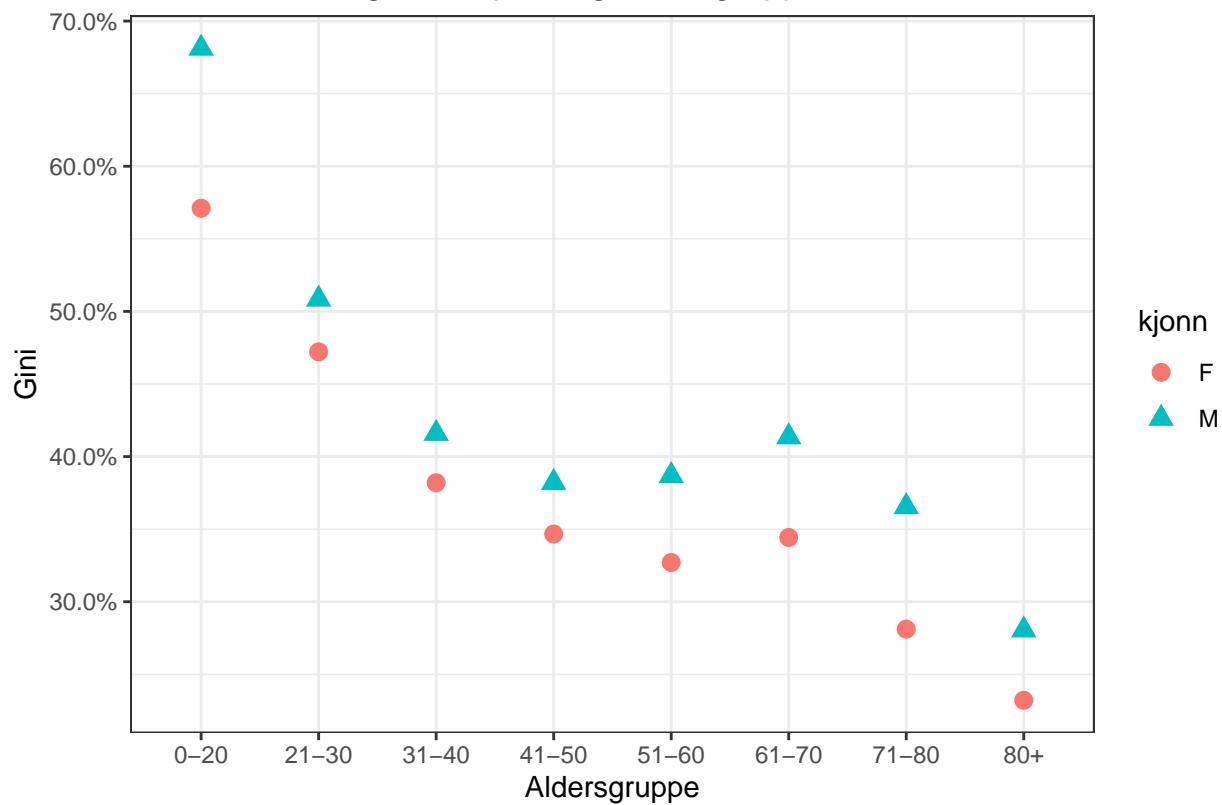
gini_mann <- aggregate(inntekt ~ aldersgruppe,
                        data = dfMale,
                        FUN="gini")
names(gini_mann) <- c("Aldersgruppe", "GINI")
gini_mann <- gini_mann %>%
  mutate(kjonn = "M")

gini_kvinner <- aggregate(inntekt ~ aldersgruppe,
                           data = dfFemale,
                           FUN="gini")
names(gini_kvinner) <- c("Aldersgruppe", "GINI")
gini_kvinner <- gini_kvinner %>%
  mutate(kjonn = "F")

gini_kjønn <- rbind(gini_kvinner, gini_mann)

gini_kjønn %>%
  ggplot(aes(x=Aldersgruppe, y=GINI,
             colour = kjonn, shape = kjonn)) + geom_point(lwd = 3) +
  theme_bw() +
  scale_y_continuous(labels = scales::percent) +
  labs(title = "Inntektsfordeling blant kjønn og aldersgruppe",
       y = "Gini",
       x = "Aldersgruppe")
```

### Inntektsfordeling blant kjønn og aldersgruppe



I dette plottet kan man se gini på inntekt mellom kjønn og aldersgruppene.

Man kan se at kvinner generelt har bedre fordeling av inntekt enn det menn har i nesten alle aldersgruppene med unntak av 21-30, der det nesten er likt mellom kvinner og menn. Noe som drar opp skjevfordelingen i samfunnet er aldersgruppen 0-20 der veldig stor gini på inntekt, og drar opp snittet fra de 6 siste aldersgruppene.

---

## Sammendrag:

Troms fylke har en gini på ca 44%, som viser til at det er en skjevfordeling av inntekt i samfunnet. Noen kommuner skiller seg ut i fylket, som Tromsø kommune med en høyere gini på ca.46% som viser til en større skjevfordeling enn i Harstad - og Omegn kommunene. På fordelingen av formue i fylket, så ser man at Tromsø kommune igjen har en større skjevfordeling en fylkessnitett, mens Harstad - og Omegn kommunene har lavere enn fylkets snittet. Formuefordelingen i samfunnet generelt er veldig høy i Troms, med ca. 84%.

Når det kommer til aldersgruppene er det også store forskjeller mellom inntekts- og formuefordeling, der de yngste aldersgruppene kommer værst ut med en høy gini-prosent, mens de eldste aldersgruppene kommer best ut med lavest gini-prosent. Dette viser til at det er en stor skjevfordeling i de yngste aldersgruppene, men en mye mindre skjevfordeling av både formue og inntekt i de eldre aldersgruppene. Grunnen til at det f.eks. kan være sånt, er fordi de eldste i samfunnet mest sannsynligvis er pensjonsiter der f.eks. staten betaler et maks og fast beløp for pensjonsutbetaling i x antall år, og resten av levetiden er det egne oppsparte midler som er inntekten, noe som kan gjøre at skjevfordelingen er mye mindre på både inntekt og formue.

Ser man på skjevfordeling av inntekt basert på kjonn og aldersgruppe, der kvinner har mindre skjevfordeling av enn det menn har. Det er stor variasjon i mellom kjønn i den yngste aldersgruppen, og deretter er forholdet av skjevfordeling mellom menn og kvinner relativt like til dem når de to eldste aldersgruppene der skjevfordelingen minker.

## Kilder

*Density 2d (2021, December), Hentet fra; <https://www.r-graph-gallery.com/2d-density-chart.html>*

*STHA (2021, December), ggplot2 density plot : Quick start guide - R software and data visualization, Hentet fra: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software-and-data-visualization>*

*GGplot 2 (2021, December), Contours of a 2D density estimate, hentet fra: [https://ggplot2.tidyverse.org/reference/geom\\_density\\_2d.html](https://ggplot2.tidyverse.org/reference/geom_density_2d.html)*