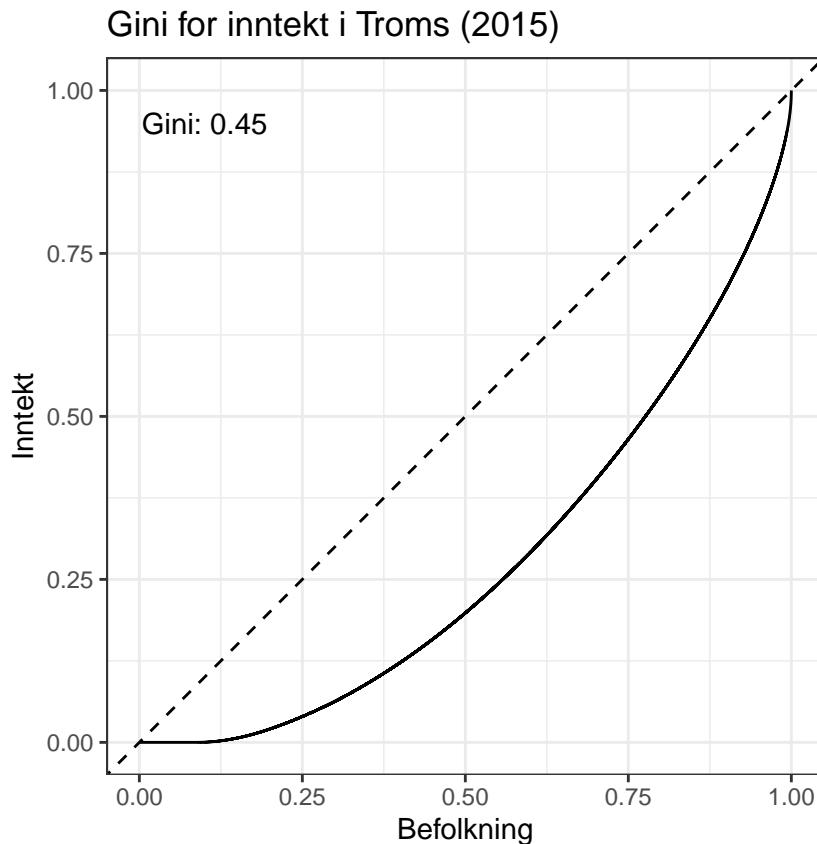


# Semesteroppgave SOK-1004

I denne oppgaven tar vi for oss skattetallene for Troms fylke i 2015. Datasettet er delt inn i følgende variabler: kommunenummer, aldersgruppe, fødselsår, kjønn, formue, inntekt, skatt.

```
inntekt <- skattetall %>%
  arrange(inntekt)

inntekt %>%
  ggplot(aes(inntekt)) +
  stat_lorenz(desc = FALSE) +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "Befolknning",
       y = "Inntekt",
       title= "Gini for inntekt i Troms (2015)") +
  theme_bw() +
  annotate_ineq(inntekt$inntekt)
```



```

befolkning <- inntekt %>%
  summarise(n=n())

tibble(
  "Fylke" = c("Troms"),
  "Gini" = ineq(inntekt$inntekt)*100,
  "inntekt" = mean(skattetall$inntekt),
  "Befolknings" = c(befolkning$n)
)

```

```

## # A tibble: 1 x 4
##   Fylke   Gini inntekt Befolknings
##   <chr>   <dbl>    <dbl>      <int>
## 1 Troms   44.5  286642.    138521

```

Oppgave A: Vi ser på gini indeksen for inntekt i Troms fylke fra 2015, at den var på 0.45. Vi ser at kurven begynner helt flatt. Dette må sees i sammenheng med at det er flere i fylket som ikke er registrert med inntekt, altså 0. Dette skyldes nok i hovedsak aldersgruppen 0-20 år, der de fleste i denne alderen ikke er i jobb, og da er registrert med 0kr i inntekt.

Det vi ser i tabellen er gini indeksen for Troms, gjennomsnittlig inntekt, og hvor mange som var registrert med bosted i Troms i 2015. Vi bruker funksjonen ineq til å regne ut gini indeksen. Vi får da en gini på 44.5, og en gjennomsnittlig inntekt på 286 642kr, fordelt på 138 521 personer i 2015.

```

skattetall1 <- skattetall %>%
  mutate(kommune=kommnr)
skattetall1 <- skattetall1 %>%
  mutate(kommune=recode(kommune,
                        "1902" = "Tromsø",
                        "1903" = "Harstad",
                        ))
skattetall1[is.na(skattetall1)] <- "Omegn"
gini.kommune <- skattetall1 %>%
  group_by(kommune) %>%
  summarise(gini=ineq(inntekt))
inntekt <- skattetall1 %>%
  group_by(kommune) %>%
  summarise(mean(inntekt))
befolkning <- skattetall1 %>%
  group_by(kommune) %>%
  summarise(n=n())
pros_bef <- skattetall1 %>%
  group_by(kommune) %>%
  summarise(percent = 100*n()/nrow(skattetall1))

```

```

tibble(
  "kommune" = c("Harstad", "Omegn", "Tromsø"),
  "Gini (%)" = c(gini.kommune$gini)*100,
  "Inntekt" = c(inntekt$`mean(inntekt)`),
  "Personer" = c(befolknings$n),
  "Andel (%)" = c(pros_bef$percent))

## # A tibble: 3 x 5
##   kommune 'Gini (%)'  Inntekt Personer 'Andel (%)'
##   <chr>      <dbl>    <dbl>    <int>      <dbl>
## 1 Harstad     43.3  299650.    20248     14.6
## 2 Omegn       42.0  269314.    55478     40.1
## 3 Tromsø      46.6  297756.    62795     45.3

```

Oppgave B: I denne tabellen har vi lagt til egne kolonner for de to største byene i Troms, altså Tromsø og Harstad. Mens resten av fylket er samlet i Omegn. Vi har også lagt til antall personer og hvor stor andel Tromsø, Harstad og Omegn utgjør i prosentvis befolkning av Troms fylke. Vi ser at Tromsø har flest innbyggere, mens Harstad har den høyeste gjennomsnittlige inntekten. Mens Omegn har den lavest gini indeks (43.29), og Tromsø har høyest gini med (46.63).

Når vi sammenligner med tabellen for hele Troms, ser vi at gjennomsnittsinntekten i Harstad 299 650kr og Tromsø 297 756kr ligger litt over snittet for hele fylket 286 642kr, mens Omegn 269 314kr ligger noe under snittet i Troms fylke. I storbyene er gjennomsnittsinntekten høyere enn på landet. Dette har nok sammenheng med at det er bedre betalte jobber i byene, og bedre jobbmarked.

Når det kommer til gini indeksen så er Tromsø litt over snittet i Troms, mens Harstad og Omegn ligger under. På grunn av Tromsø sitt høye innbyggertall, så er de med på å dra opp snittet for hele Troms fylke som er (44.52).

```

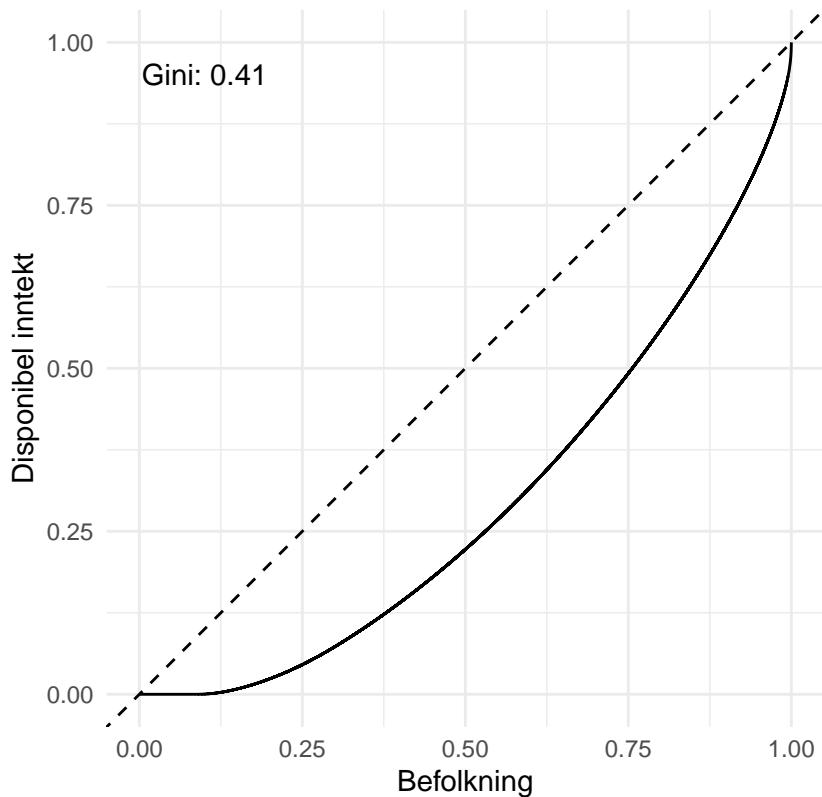
skattetall1 <- skattetall1 %>%
  mutate(disponibel_inntekt = inntekt-skatt)
skattetall1[skattetall1 < 0] <- 0

disponibel_inntekt <- skattetall1 %>%
  arrange(disponibel_inntekt)

disponibel_inntekt %>%
  ggplot(aes(disponibel_inntekt)) +
  stat_lorenz(desc = FALSE) +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "Befolknings",
       y = "Disponibel inntekt",
       title= "Gini for disponibel inntekt i Troms (2015)") +
  annotate_ineq(disponibel_inntekt$disponibel_inntekt)

```

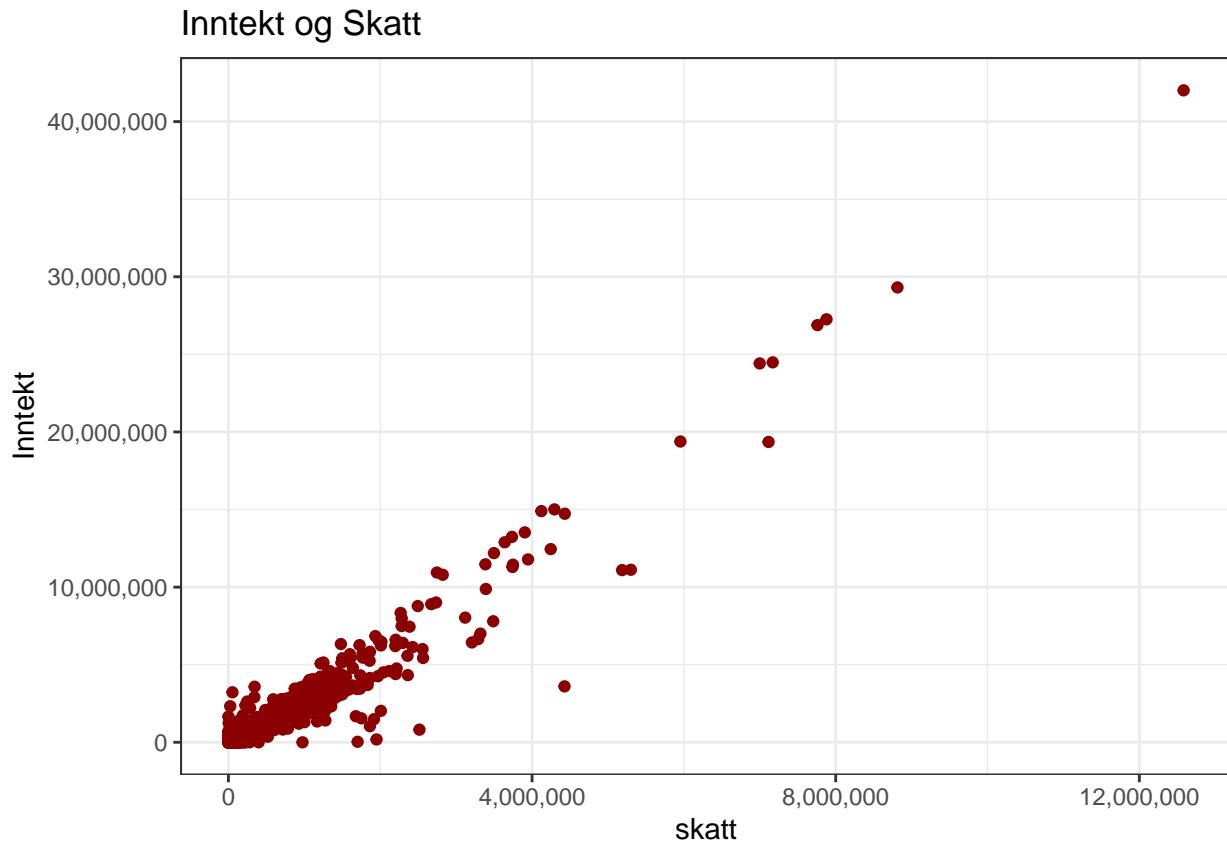
### Gini for disponibel inntekt i Troms (2015)



Oppgave C: På figuren kan vi lese at gini indeksen for disponibel inntekt ligger på 0.41. Det vil si at ulikhettene er mindre etter skatt er betalt.

```
inntekt <- skattetall %>%
  arrange(inntekt)
skatt <- skattetall %>%
  arrange(skatt)

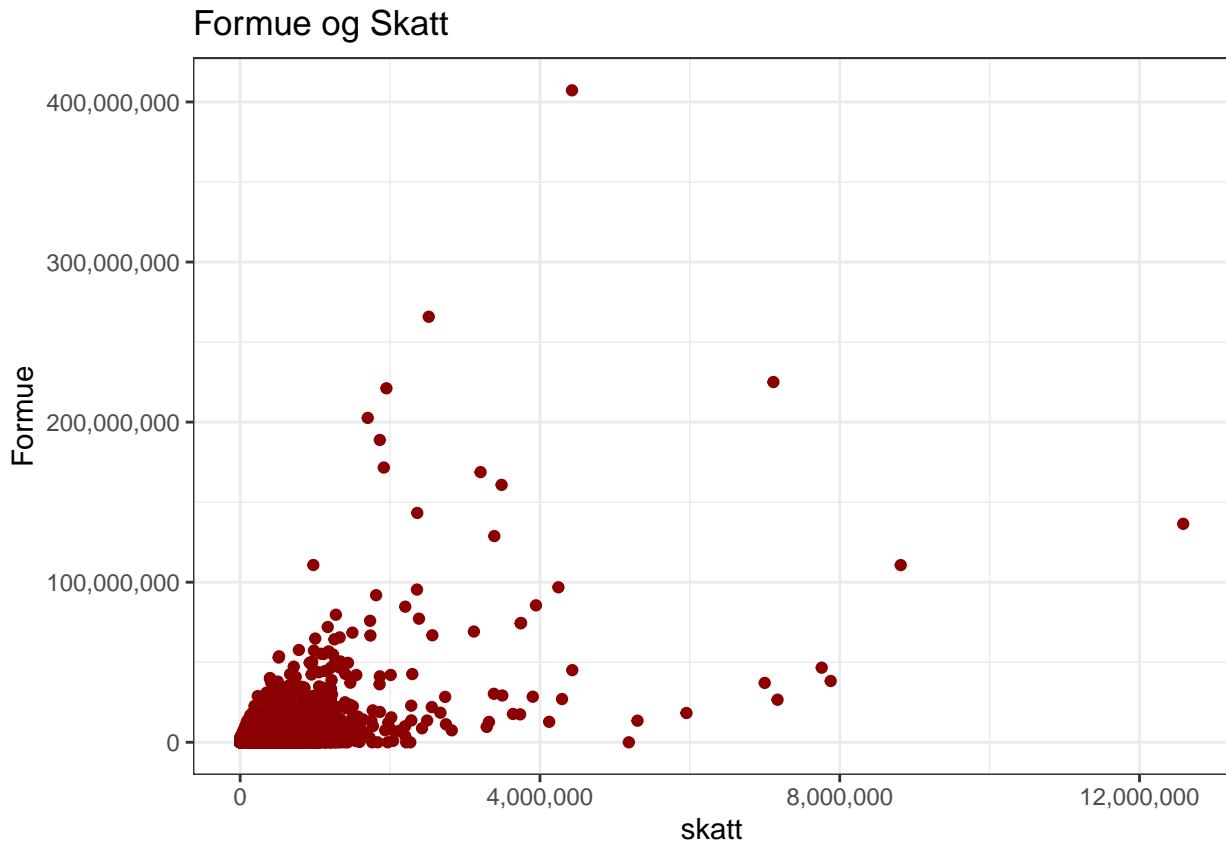
inntekt %>%
  ggplot(aes(x = skatt, y = inntekt)) +
  geom_point(aes(y = inntekt), color="dark red") +
  scale_y_continuous(labels = scales::comma) +
  scale_x_continuous(labels = scales::comma) +
  labs(title = "Inntekt og Skatt",
       x = "skatt",
       y = "Inntekt",) +
  theme_bw()
```



Oppgave D: På figuren har vi skatt på x-aksen og inntekt på y-aksen i millioner kr. Utviklingen er relativt lineær med unntak av noen få avvik. Noen individer har høy skatt i forhold til inntekt, dette er personer med høy formue og dermed høy formueskatt.

```
Formue <- skattetall %>%
  arrange(formue)

Formue %>%
  ggplot(aes(x = skatt, y = formue)) +
  geom_point(aes(y = formue), color="dark red") +
  scale_y_continuous(labels = scales::comma) +
  scale_x_continuous(labels = scales::comma) +
  labs(title = "Formue og Skatt",
       x = "skatt",
       y = "Formue",) +
  theme_bw()
```



Oppgave E: I denne figuren har vi byttet ut inntekt med formue på y-aksen og har fortsatt skatt på x-aksen. I forhold til figuren med skatt og inntekt ser vi at det ikke er like sterk korrelasjon mellom skatt og formue. Dette skyldes en blanding av at flere med høy formue har liten til ingen inntekt og dermed kun formueskatt. De punktene som går skrått oppover ser vi har høy formue og betaler mye skatt også på grunn av høy inntekt.

```

tromso <- skattetall %>%
  filter(kommnr==1902)
harstad <- skattetall %>%
  filter(kommnr==1903)
omegn <- skattetall %>%
  filter(kommnr>=1904)

for_omegn <- mean(omegn$formue)
for_tromso <- mean(tromso$formue)
for_harstad <- mean(harstad$formue)
for_troms <- mean(skattetall$formue)
gini_for_omegn <- ineq(omegn$formue)
gini_for_tromso <- ineq(tromso$formue)
gini_for_harstad <- ineq(harstad$formue)
gini_for_troms <- ineq(skattetall$formue)
per_omegn <- omegn %>%
  summarise(n=n())
per_tromso <- tromso %>%
  summarise(n=n())
per_harstad <- harstad %>%
  summarise(n=n())
pros_bef <- skattetall1 %>%

```

```

group_by(kommune) %>%
  summarise(percent = 100*n()/nrow(skattetall1))
personer <- skattetall1 %>%
  summarise(n=n())
pros_total <- personer %>%
  group_by(n) %>%
  summarise(percent = 100*n()/nrow(personer))
tibble(
  "kommune" = c("Harstad", "Omegn", "Tromsø", "Troms"),
  "Gini (%)" = c(gini_for_harstad*100, gini_for_omegn*100, gini_for_tromso*100, gini_for_troms*100),
  "Formue" = c(for_harstad, for_omegn, for_tromso, for_troms),
  "Personer" = c(per_harstad$n, per_omegn$n, per_tromso$n, personer$n),
  "Andel (%)" = c(pros_bef$percent, pros_total$percent))

## # A tibble: 4 x 5
##   kommune 'Gini (%)'  Formue Personer 'Andel (%)'
##   <chr>      <dbl>    <dbl>    <int>      <dbl>
## 1 Harstad    84.0    461466.    20248     14.6
## 2 Omegn      82.5    441224.    55478     40.1
## 3 Tromsø      86.2    491493.    62795     45.3
## 4 Troms       84.6    466971.   138521     100

```

Oppgave F: Her har vi byttet ut inntekt med formue, samt lagt til Troms i tabellen. Når det kommer til gini indeksen for formue er den mye høyere. Det vil si at den totale formuen er fordelt relativt få personer, altså større ulikhet blant befolkningens formue. Vi så tidligere at Tromsø var byen med høyest gjennomsnittlig inntekt, i denne figuren ser vi at Tromsø også er byen med størst gjennomsnittlig formue.

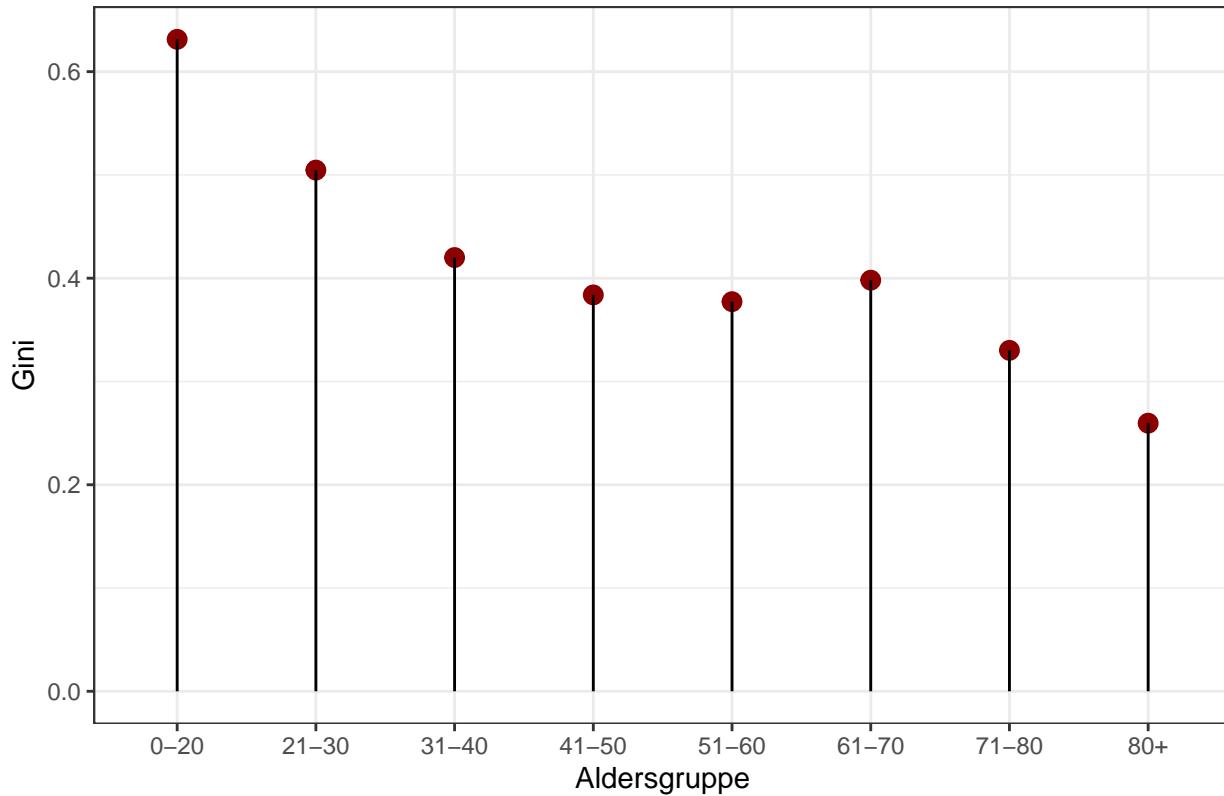
```

aldersgruppe <- skattetall %>%
  group_by(aldersgruppe) %>%
  summarise(gini_aldersgruppe=ineq(inntekt))

aldersgruppe %>%
  ggplot(aes(x=aldersgruppe, y=gini_aldersgruppe)) +
  geom_point(size=3, color="dark red") +
  geom_segment(aes(x=aldersgruppe, xend=aldersgruppe, y=0, yend=gini_aldersgruppe)) +
  labs(title="Gini og aldersgruppe",
       x="Aldersgruppe",
       y="Gini") +
  theme_bw()

```

## Gini og aldersgruppe



Oppgave G: I figuren med inntekt og gini indeks per aldersgruppen. Ser vi at ulikheten blant aldersgruppen 0-20 har en høy gini indeks. Dette er fordi de fleste i denne gruppen er registrert med 0kr i inntekt. Det er en naturlig forklaring da de fleste i denne gruppen er barn, studenter og inne til førstegangstjeneste. Så de få som har inntekt drar gini indeksen kraftig opp. Vi ser så en nedadgående gini indeks for de andre aldersgruppene, med unntak av aldersgruppen 61-70år der den går litt opp igjen. Den nedadgående trenden er nok fordi flere og flere blir ferdig utdannet og kommer i jobb, noe som er med på å utjevne ulikhetene. I aldersgruppen 61-70år, så er det mange som pensjonerer seg i slutten av 60-årene, noe som fører til en økning i gini indeksen for denne gruppen.

```

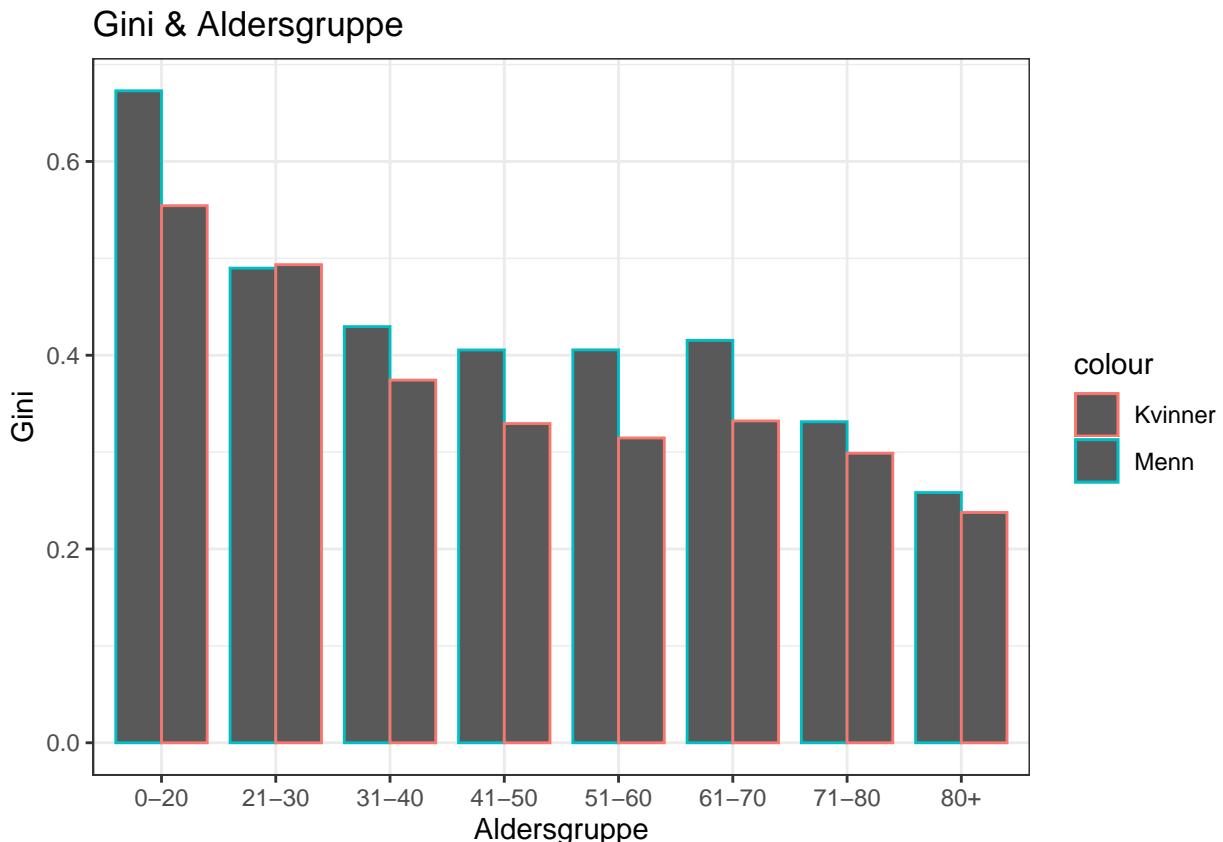
Menn <- skattetall %>%
  filter(kjonn=="M")
Kvinner <- skattetall %>%
  filter(kjonn=="F")

aldersgruppe_menn <- Menn %>%
  group_by(aldersgruppe) %>%
  summarise(gini_aldersgruppe=ineq(inntekt))
aldersgruppe_kvinner <- Kvinner %>%
  group_by(aldersgruppe) %>%
  summarise(gini_aldersgruppe=ineq(inntekt))

ggplot(NULL, aes(x= aldersgruppe, y=gini_aldersgruppe)) +
  geom_bar(data=aldersgruppe_menn, aes(color='Menn'), stat = "identity",
           width=.4, position=position_nudge(x = -0.2)) +
  geom_bar(data=aldersgruppe_kvinner, aes(color='Kvinner'), stat = "identity",
           width=.4, position=position_nudge(x = 0.2)) +

```

```
labs(title="Gini & Aldersgruppe",
  x="Aldersgruppe",
  y="Gini") +
theme_bw()
```



Oppgave H: Vi ser at gini for menn i alle aldersgrupper er høyere enn hos kvinner, bortsett fra 21-30år. Vi tror det er større forskjeller blant menn for de med høyest inntekt i Troms er menn, og det er flest menn som er arbeidsledige, altså det blir større forskjell blant menn enn blant kvinner. Forskjellen i aldersgruppen 0-20år er ganske betydelig, særlig for menn. Vi tror det er fordi flere menn enn kvinner velger yrkesfag. De kommer tidligere i jobb i motsetning til de i denne aldersgruppen som venter noen år med utdanning.

Skriv et kort sammendrag om inntekt, formue og ulikheter i Troms fylke, basert på dine analyser.

Inntekten, formuen og ulikheten i Troms fra 2015 kan fremstilles og analyseres ved bruk av flere figurer. Ved bruk av gini indeksen har vi fremstilt inntektsulikheten i Troms, både før og etter skatt. Ulikheten synker med 0.04 etter skatt noe som kan tilsvare at vi har en skattepoltikk som forebygger ulikhet. Vi har også sett på ulikhetene for formue, og sammenhengen mellom formue og skatt. Det vi ser for formue i Troms er at den har en høy gini indeks, som vil si at den totale formuen er ulikt fordelt i befolkningen. Det er altså en liten andel av befolkningen som har høy formue.

Vi har funnet ut at gini indeksen er synkende per aldersgruppe, fordi det er flere som begynner å jobbe når man blir eldre og dermed blir inntektsforskjellene jevnere over tid. Når vi skiller kvinner og menn fra hverandre og setter det opp i et søylediagram kan vi se at ulikheten er generelt større for menn enn for kvinner. En observasjon vi har gjort er at når det kommer til gini er at Tromsø har den høyeste gini indeksen. Tromsø er den største byen i Troms med 45% av befolkningen og har dermed svært stor påvirkningen på statistikken.