

Assignment 2 - Adrian og Sindre

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(forcats)
library(modelr)
library(flextable)
library(vtable)
library(labeling)
library(DescTools)
```

1 Innledning

I dette minipaperet skal vi se nærmere på forholdet mellom høyde og inntekt, og prøve å finne svar på om det er høyden som bestemmer inntekten vår, ved hjelp av datasettet Heights. Besvarelsen vil starte med en generell literaturgjennomgang hvor vi ser nærmere på forholdet mellom inntekt og en rekke andre faktorer, blant annet bmi, sivilstatus, utdanning og kjønn.

Deretter vil det gjennomføres en EDA-analyse for å se nærmere på det aktuelle datasettet, og bruke deskriptiv statistikk for å analysere sammenhenger mellom ulike faktorer, og knytte dette opp mot literaturgjennomgangen. Videre brukes grafiske modeller for å studere andre relevante sammenhenger, før det til slutt konkluderes hvorvidt det er høyde som bestemmer inntekt.

2 Literaturgjennomgang

Denne delen av oppgaven vil benytte eksterne kilder, og gjøre rede for hva teori og forskning mener er sammenhengen mellom lønn og henholdsvis, høyde, bmi, sivilstatus, utdanning og kjønn.

2.1 Lønn og høyde

Flere studier, inkludert Census (n.d.) og Britt (2009), viser en sammenheng mellom høyde og lønn. Høyde kan være en indikator på helse og ernæring i barndommen, som videre kan påvirke utdanning og arbeidsmuligheter. De peker på at høyere individer i gjennomsnitt har høyere inntekter, noe som kan være relatert til oppfattelse av høyde som et signal om styrke og lederskap i arbeidslivet. Det kan også være at høyere individer opplever mindre diskriminering i arbeidsmarkedet og blir sett på som mer selvsikre, noe som kan være attraktivt for arbeidsgivere. Denne sammenhengen tyder på at det finnes flere underliggende faktorer som knytter høyde til økonomisk suksess, noe som er viktig å ta hensyn til når man studerer arbeidsmarkedet. I tillegg viser en studie at høyere individer har bedre selvopplevd helse, noe som også kan bidra til høyere inntektsnivå (Census n.d.). Case og Paxson (2006) finner det samme, og hevder at hver tomme med ekstra høyde er assosiert med omtrent 2-2,5% økning i inntekt. De peker blant annet på at høye personer gjør det bedre på kognitive tester som barn, og at dette legger et bedre utgangspunkt for karriere. De samme forskerne mener også at i tillegg til bedre kognitive ferdigheter som barn, vil også høyde ofte gi høyere selvtillit og andre personlige egenskaper som kan være avgjørende for inntekt.

2.2 Lønn og bmi

Böckerman mfl. (2019) finner at høy BMI generelt har en negativ innvirkning på lønn, spesielt for kvinner. I tillegg kan personer med høyere BMI møte utfordringer som lavere selvtillit og færre karrieremuligheter på grunn av samfunnsmessige stigma knyttet til vekt. Studier viser også at sammenhengen mellom BMI og inntekt kan variere mellom kjønn og regioner, noe som antyder at kulturelle normer og holdninger spiller en viktig rolle i hvordan vekt påvirker inntekten. Videre finner Edwards, Bjørngaard, og Kinge (2021) m.fl. at individer med høyere BMI har redusert lønnsutvikling sammenlignet med individer med normal BMI, spesielt over en lengre tidsperiode. For menn er sammenhengen ofte svakere eller ikke-eksisterende, mens kvinner ofte opplever en mer betydelig negativ effekt.

2.3 Lønn og sivilstatus

Når det gjelder sivilstatus, finner Vandenbroucke (n.d.) at gifte individer ofte har høyere lønn enn ugifte. Dette kan skyldes flere faktorer, som at gifte personer kan ha en stabiliserende effekt på økonomien og ofte nyter fordeler som bedre helse og økt nettverk som kan føre til høyere inntekt. I tillegg kan gifte personer ha tilgang til ekstra ressurser og støtte som bidrar til karriereutvikling og jobbstabilitet (On-The-Economy-Blog 2020). Vi ser ofte at inntekt øker med alder, og at det derfor ofte vil være en yngre gruppe vi studerer når vi ser nærmere på ugifte personer (Howard 2018). Sivilstatus kan dermed ha en betydelig påvirkning på en persons inntekt, og det å være gift ser ut til å være en fordel i mange økonomiske sammenhenger.

2.4 Lønn og utdanning

Flere studier, inkludert Census (n.d.) og Britt (2009), viser en sterk positiv sammenheng mellom utdanning og lønn. Høyere utdanning fører ofte til bedre kvalifikasjoner og muligheter for høyere betalte jobber. Utdanning øker ikke bare inntekten direkte, men påvirker også individets evne til å tilpasse seg et arbeidsmarked i endring. Videre gir utdanning tilgang til sosiale nettverk og anerkjennelse, som begge kan være verdifulle for å sikre høyere lønn (Britt 2009). De med høyere utdanning har også større mulighet til å få stillinger som gir økonomisk sikkerhet og karriereutvikling, noe som igjen fører til bedre langsiktige inntektsmuligheter. Dette gjør at utdanning kan sees som en av de viktigste investeringene for å sikre høyere levestandard og sosial mobilitet. Jacob Mincers inntektsfunksjon illustrerer at hvert år med utdanning øker inntekten med omtrent 10 % globalt (Mincer 1974). Dette gjør at utdanning kan sees som en av de viktigste investeringen for å sikre høyere levestandard og inntekt.

2.5 Lønn og kjønn

Studier fremhever kjønnsforskjeller i lønn som et vedvarende problem. Generelt tjener kvinner mindre enn menn, selv når de har tilsvarende kvalifikasjoner og erfaring. Gould, Schieder, og Geier (2016) peker på at årsakene til denne lønnsforskjellen inkluderer diskriminering, forskjellige karrierevalg og at kvinner i større grad velger sektorer med lavere lønnsnivå. I tillegg finner On-The-Economy-Blog (2020) at kvinner ofte har større sannsynlighet for å ta pauser fra arbeidslivet for å ta vare på familie, noe som kan føre til lavere samlet arbeidsopplevelse og dermed lavere lønn. I følge Bureau (2019), er forskjellen så stor at blant de med høyere utdanning tjener kvinner kun 74 cents for hver dollar menn tjener. Selv om det er gjort fremskritt når det gjelder å redusere lønnsgapet mellom kjønnene, viser forskningen at strukturelle og kulturelle barrierer fortsatt hindrer kvinner i å oppnå likelønn. Forskere understreker behovet for politiske tiltak og strukturelle endringer for å rette opp i disse forskjellene, slik at kvinner får likere muligheter til økonomisk fremgang.

3 Utforskende dataanalyse («EDA»)

Vi benytter R, R Core Team (2024), og datasettet *heights* fra pakken *modelr* (Wickham 2023).

Tabell 1: Beskrivende statistikk for datasettet `modelr::heights` før bearbeiding.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	7006	41204	55892	0	166	55000	343830
height	7006	67	4.1	52	64	70	84
weight	6911	188	44	76	157	212	524
age	7006	51	2.2	47	49	53	56

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
marital	7006						
... single	1124	16%					
... married	3806	54%					
... separated	366	5%					
... divorced	1549	22%					
... widowed	161	2%					
sex	7006						
... male	3402	49%					
... female	3604	51%					
education	6996	13	2.6	1	12	15	20
afqt	6744	41	29	0	15	65	100

Liste 1 Kode for å lese inn data og definere noen nye variabler.

```

heights <- heights %>%
  mutate(
    bmi = weight * 0.4536/(height * 2.54/100)^2,
    married = fct_collapse(
      .f = marital,
      married = "married",
      other_level = "not married"
    ),
    edu_fac = cut(
      x = education,
      breaks = c(0, 12, 14, 16, 21),
      labels = c("not_hs", "not_cc", "not_col", "col_plus"),
      right = FALSE
    )
  ) |>
# reorganiserer data s.a. de fire faktor-variablerne kommer
# lengst til høyre
select(income:age, education:bmi, everything()) |>
# Dropper marital og education siden disse ikke skal brukes
select(-education, -marital)

```

Tabell 2: Beskrivende statistikk for datasettet `heights` etter at det er bearbeidet.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	7006	41204	55892	0	166	55000	343830
height	7006	67	4.1	52	64	70	84
weight	6911	188	44	76	157	212	524
age	7006	51	2.2	47	49	53	56
afqt	6744	41	29	0	15	65	100
bmi	6911	29	6.2	13	25	32	75
sex	7006						
... male	3402	49%					
... female	3604	51%					
married	7006						
... married	3806	54%					
... not married	3200	46%					
edu_fac	6996						
... not_hs	1058	15%					
... not_cc	3195	46%					
... not_col	1110	16%					
... col_plus	1633	23%					

Liste 2 Lager nye datasett bestående av deler av opprinnelig datasett. `heightsZeroInc` består av individ med inntekt lik null, mens `heightsNormInc` består av individ med med normal inntekt, dvs. $0 < \text{inntekt} < \$ 343830$.

```
# Inntekt lik 0
heightsZeroInc <- heights |>
  filter(income == 0)
# «Normal» inntekt
heightsNormInc <- heights |>
  filter(income > 0 & income < 343830)
heightsHighInc <- heights |>
  filter(income == 343830)
```

Tabell 3: Beskrivende statistikk for del-datasettet `heightsZeroInc` som består av individ med inntekt lik \$0.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	1740	0	0	0	0	0	0
height	1740	66	4.1	55	63	69	84
weight	1714	187	49	76	153	215	524
age	1740	51	2.2	47	50	53	56
afqt	1662	29	26	0	7.5	44	100
bmi	1714	30	7.4	14	25	33	75
sex	1740						
... male	745	43%					
... female	995	57%					
married	1740						
... married	705	41%					
... not married	1035	59%					
edu_fac	1732						
... not_hs	497	29%					
... not_cc	835	48%					
... not_col	211	12%					
... col_plus	189	11%					

```
heightsNormInc |>
  st(
    out = "return"
  ) |>
  as_flextable(max_row = 30, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  # ag: endrer vidde første kolonne
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
```

Tabell 4: Beskrivende statistikk for del-datasettet `heightsNormInc` som består av individ med normal inntekt, dvs. $\$0 < \text{inntekt} \leq \343830 .

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	5123	46751	33286	45	23000	62000	178000
height	5123	67	4	52	64	70	80

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
weight	5054	188	43	78	159	212	480
age	5123	51	2.2	47	49	53	56
afqt	4939	44	28	0	20	68	100
bmi	5054	29	5.8	13	25	32	67
sex	5123						
... male	2526	49%					
... female	2597	51%					
married	5123						
... married	2983	58%					
... not married	2140	42%					
edu_fac	5121						
... not_hs	559	11%					
... not_cc	2349	46%					
... not_col	886	17%					
... col_plus	1327	26%					

ag: Bare sjekker at kryssreferanser virker. Tabell 4 angir beskrivende statistikk for individ med normal inntekt, mens Tabell 3 angir beskrivende statistikk for individer med 0 i inntekt.

Liste 3 Kode for å lage tabell med deskriptiv statistikk.

```
heights |>
  #fjerner afqt da vi ikke har behov for det her
  select(-afqt) |>
# ag: ser at dere også har funnet st(out = "return")
st(out = "return") |>
as_flextable(max_row = 20, show_coltype = FALSE) |>
line_spacing(space = 0.3, part = "body") |>
fontsize(size = 9, part = "body") |>
fontsize(size = 10, part = "header") |>
# ag: Øker vidden på første kolonne s.a. vi får plass til not-married
# uten at linjen brytes. Når vi kjører så kompakte tabeller med
# line_spacing 0.3 kan vi ikke ha brutte linjer. Gjør at tekst
# overskrives.
width(j = 2:8, width = 1.2, unit = "cm") |>
fit_to_width(max_width = 15, unit = "cm") |>
delete_part("footer")
```

Tabell 5: Deskriptiv statistikk for datasettet `modelr::heights`.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	7006	41204	55892	0	166	55000	343830
height	7006	67	4.1	52	64	70	84
weight	6911	188	44	76	157	212	524
age	7006	51	2.2	47	49	53	56
bmi	6911	29	6.2	13	25	32	75
sex	7006						
... male	3402	49%					
... female	3604	51%					
married	7006						
... married	3806	54%					
... not married	3200	46%					
edu_fac	6996						
... not_hs	1058	15%					
... not_cc	3195	46%					
... not_col	1110	16%					
... col_plus	1633	23%					

ag: Sjekker kryssreferanser. Både Tabell 1 og Tabell 5 angir beskrivende statistikk for hele datasettet.

```
# Totalt antall manglende verdier i hele datasettet
total_missing <- heights %>%
  summarize_all(~ sum(is.na(.))) %>%
  summarise(total_missing = sum(across(everything())) %>%
    pull(total_missing))
# print(total_missing)
# Se nedenfor for hvordan man kan få det samme fint innarbeidet i teksten.
```

Totalt antall manglende verdier (NA) er 462.

```
heights %>%
  summarise(
    Mean = mean(height, na.rm = TRUE),
    SD = sd(height, na.rm = TRUE),
    Min = min(height, na.rm = TRUE),
    Median = median(height, na.rm = TRUE),
    Max = max(height, na.rm = TRUE)
```



```

) |>
round(digits = 3) %>%
# print(height_stats)
# ag: as_flextable har den irriterende egenskapen at tibbles med en rekke
# blir automatisk snudd til en tabell med en kolonne. Gjør at en tabell
# av en named vektor, slik som her ,tar unødvendig stor plass.
# Et triks som bøter på dette er følgende:
rbind(rep("a", 5)) |>
as_flextable(show_coltype = FALSE) |>
delete_rows(i = 2, part = "body") |>
delete_part("footer") |>
theme_booktabs()

```

Tabell 6: Deskriptiv statistikk for høyde.

Mean	SD	Min	Median	Max
67.104	4.082	52	67	84

```

heights |>
filter(income == 0) |>
st(out = "return") |>
as_flextable(max_row = 20, show_coltype = FALSE) |>
line_spacing(space = 0.3, part = "body") |>
fontsize(size = 9, part = "body") |>
fontsize(size = 10, part = "header") |>
# ag: endrer vidde første kolonne
width(j = 2:8, width = 1.2, unit = "cm") |>
fit_to_width(max_width = 15, unit = "cm") |>
delete_part("footer")

```

Tabell 7: Deskriptiv statistikk for delsettet med ingen inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	1740	0	0	0	0	0	0
height	1740	66	4.1	55	63	69	84
weight	1714	187	49	76	153	215	524
age	1740	51	2.2	47	50	53	56
afqt	1662	29	26	0	7.5	44	100
bmi	1714	30	7.4	14	25	33	75
sex	1740						
... male	745	43%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
... female	995	57%					
married	1740						
... married	705	41%					
... not married	1035	59%					
edu_fac	1732						
... not_hs	497	29%					
... not_cc	835	48%					
... not_col	211	12%					
... col_plus	189	11%					

```

heights %>%
  filter(income > 0 & income < 343830) %>%
  st(out = "return") %>%
  as_flextable(max_row = 20, show_coltype = FALSE) %>%
  line_spacing(space = 0.3, part = "body") %>%
  fontsize(size = 9, part = "body") %>%
  fontsize(size = 10, part = "header") %>%
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")

```

Tabell 8: Deskriptiv statistikk for delsettet med normal inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	5123	46751	33286	45	23000	62000	178000
height	5123	67	4	52	64	70	80
weight	5054	188	43	78	159	212	480
age	5123	51	2.2	47	49	53	56
afqt	4939	44	28	0	20	68	100
bmi	5054	29	5.8	13	25	32	67
sex	5123						
... male	2526	49%					
... female	2597	51%					
married	5123						
... married	2983	58%					
... not married	2140	42%					
edu_fac	5121						

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
... not_hs	559	11%					
... not_cc	2349	46%					
... not_col	886	17%					
... col_plus	1327	26%					

```

heights %>%
  filter(income >= 343830) %>%
  st(out = "return") %>%
  as_flextable(max_row = 20, show_coltype = FALSE) %>%
  line_spacing(space = 0.3, part = "body") %>%
  fontsize(size = 9, part = "body") %>%
  fontsize(size = 10, part = "header") %>%
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")

```

Tabell 9: Deskriptiv statistikk for delsettet med høy inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	143	343830	0	343830	343830	343830	343830
height	143	71	3.1	61	69	72	81
weight	143	195	37	123	170	210	335
age	143	51	2.3	48	49	53	55
afqt	143	78	22	3.3	70	94	100
bmi	143	28	4.8	15	25	30	45
sex	143						
... male	131	92%					
... female	12	8%					
married	143						
... married	118	83%					
... not married	25	17%					
edu_fac	143						
... not_hs	2	1%					
... not_cc	11	8%					
... not_col	13	9%					
... col_plus	117	82%					

3.1 Forskjeller mellom de tre inntektsgruppene.

Når vi sammenligner de tre inntektsgruppene — null inntekt, normal inntekt, og høy inntekt — ser vi tydelige forskjeller på flere områder som høyde, vekt, sivilstatus, og utdanningsnivå.

For gruppen uten inntekt, ser vi at gjennomsnittshøyden er 66 tommer og vekten er 187 lbs. BMI ligger på et gjennomsnitt på 30, noe som indikerer en tendens til overvekt. Det er også en høyere andel kvinner (57 %) i denne gruppen enn menn (43 %). Når det gjelder sivilstatus, er 41 % av deltakerne gift, mens de resterende 59 % er ugifte. Utdanningsnivået viser at kun 11 % har høyere utdanning, mens en stor del av gruppen (48 %) har ikke collegeutdanning.

I gruppen med normal inntekt, Tabell 8 er gjennomsnittlig inntekt 46,751. Vi merker oss også en noe høyere gjennomsnittshøyde på 67 tommer, og en vekt på 188 lbs, som gir en BMI på 29. Andelen personer som er gift, er 58 %, noe som indikerer at stabil sivilstatus kan være mer vanlig blant personer med en viss inntekt, sammenlignet med gruppen uten inntekt. Når det gjelder utdanning, har 26 % av deltakerne høyere utdanning, mens 46 % ikke har det. Dette tyder på at utdanning fortsatt spiller en viktig rolle i å nå dette inntektsnivået, men at det ikke nødvendigvis krever en fullført høyere grad.

For gruppen med høy inntekt, ser vi flere markante forskjeller sammenlignet med de to øvrige undergruppene. Gjennomsnittlig høyde her er 71 tommer og vekten ligger på 195 lbs, noe som gir en BMI på 28. Andelen som er gift er også svært høy (83 %), noe som kan tyde på at stabil sivilstatus og høy inntekt er nært knyttet sammen. Utdanningsnivået er også betydelig høyere i denne gruppen, med hele 82 % som har fullført høyere utdanning. Dette trekker i retning av en tydelig sammenheng mellom høyere utdanning og høy inntekt.

Totalt sett er det store forskjeller mellom de tre undergruppene, spesielt basert på kjønn, sivilstatus og utdanningsnivå. Vi merker oss at menn er sterkt overrepresentert blant de med høy inntekt, og kvinner er noe overrepresentert i gruppen uten inntekt. Andelen gifte og andelen med høyere utdanning er betydelig høyere i gruppen med høy inntekt, noe som tyder på at både stabil sivilstatus og utdanning er viktige faktorer for å oppnå høy inntekt. Disse tendensene ser vi også fra gruppen med ingen inntekt opp til gruppen med normale inntekter. Gruppen uten inntekt skiller seg ut med høyere andel kvinner og lavere utdanning, noe som kan tyde på økonomiske utfordringer knyttet til manglende utdanning.

3.2 Splittet på kjønn

```
heights |>
  filter(sex == 'male' & income == 0) |>
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
```

```
# width(width = 16, unit = "mm") |>
width(j = 2:8, width = 1.2, unit = "cm") |>
fit_to_width(max_width = 15, unit = "cm") |>
delete_part("footer")
```

Tabell 10: Deskriptiv statistikk for menn med null inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	745	0	0	0	0	0	0
height	745	70	3.3	55	67	72	84
weight	741	202	46	94	170	225	524
age	745	51	2.2	47	49	53	56
afqt	704	26	24	0	7.6	40	99
bmi	741	29	5.9	14	25	32	69
sex	745						
... male	745	100%					
... female	0	0%					
married	745						
... married	222	30%					
... not married	523	70%					
edu_fac	741						
... not_hs	252	34%					
... not_cc	359	48%					
... not_col	74	10%					
... col_plus	56	8%					

```
heights_zero_inc_female <- heights %>%
  filter(sex == 'female' & income == 0) |>
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
heights_zero_inc_female
```

Tabell 11: Deskriptiv statistikk for kvinner med null inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	995	0	0	0	0	0	0
height	995	64	2.9	55	62	66	79
weight	973	177	49	76	140	200	430
age	995	52	2.2	47	50	53	56
afqt	958	30	27	0	7.2	49	100
bmi	973	30	8.3	15	24	34	75
sex	995						
... male	0	0%					
... female	995	100%					
married	995						
... married	483	49%					
... not married	512	51%					
edu_fac	991						
... not_hs	245	25%					
... not_cc	476	48%					
... not_col	137	14%					
... col_plus	133	13%					

```
heightsNormIncMale <- heightsNormInc %>%
  filter(sex == 'male') |>
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
  # width(width = 16, unit = "mm") |>
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
heightsNormIncMale
```

Tabell 12: Deskriptiv statistikk for menn med normal inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	2526	54236	35622	72	28800	72000	178000
height	2526	70	2.9	57	68	72	80

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
weight	2520	205	39	103	180	225	480
age	2526	51	2.2	47	49	53	56
afqt	2413	44	29	0	18	69	100
bmi	2520	29	5	13	26	32	67
sex	2526						
... male	2526	100%					
... female	0	0%					
married	2526						
... married	1575	62%					
... not married	951	38%					
edu_fac	2524						
... not_hs	332	13%					
... not_cc	1238	49%					
... not_col	372	15%					
... col_plus	582	23%					

```
heightsNormIncFemale <- heightsNormInc %>%
  filter(sex == 'female') %>%
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
  #width(width = 16, unit = "mm") |>
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
heightsNormIncFemale
```

Tabell 13: Deskriptiv statistikk for kvinner med normal inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	2597	39472	29057	45	19000	53000	175000
height	2597	64	2.7	52	62	66	75
weight	2534	172	40	78	144	193	365
age	2597	51	2.2	47	49	53	56
afqt	2526	44	28	0	21	67	100

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
bmi	2534	29	6.5	14	25	33	60
sex	2597						
... male	0	0%					
... female	2597	100%					
married	2597						
... married	1408	54%					
... not married	1189	46%					
edu_fac	2597						
... not_hs	227	9%					
... not_cc	1111	43%					
... not_col	514	20%					
... col_plus	745	29%					

```
heightsHighIncMale <- heightsHighInc %>%
  filter(sex == 'male') |>
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
  #width(width = 16, unit = "mm") |>
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
heightsHighIncMale
```

Tabell 14: Deskriptiv statistikk for menn med høy inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	131	343830	0	343830	343830	343830	343830
height	131	71	2.7	64	69	72	81
weight	131	199	36	130	175	215	335
age	131	51	2.3	48	49	53	55
afqt	131	78	22	3.3	70	94	100
bmi	131	28	4.8	15	25	30	45
sex	131						
... male	131	100%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
... female	0	0%					
married	131						
... married	108	82%					
... not married	23	18%					
edu_fac	131						
... not_hs	2	2%					
... not_cc	10	8%					
... not_col	13	10%					
... col_plus	106	81%					

```
heightsHighIncFemale <- heightsHighInc %>%
  filter(sex == 'female') |>
  st(out = "return") |>
  as_flextable(max_row = 20, show_coltype = FALSE) |>
  line_spacing(space = 0.3, part = "body") |>
  fontsize(size = 9, part = "body") |>
  fontsize(size = 10, part = "header") |>
  #width(width = 16, unit = "mm") |>
  width(j = 2:8, width = 1.2, unit = "cm") |>
  fit_to_width(max_width = 15, unit = "cm") |>
  delete_part("footer")
heightsHighIncFemale
```

Tabell 15: Deskriptiv statistikk for kvinner med høy inntekt.

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
income	12	343830	0	343830	343830	343830	343830
height	12	65	2.5	61	63	67	68
weight	12	151	21	123	136	166	185
age	12	51	2.4	48	50	53	55
afqt	12	78	23	29	68	95	99
bmi	12	25	4.5	19	22	29	34
sex	12						
... male	0	0%					
... female	12	100%					
married	12						
... married	10	83%					

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
... not married	2	17%					
edu_fac	12						
... not_hs	0	0%					
... not_cc	1	8%					
... not_col	0	0%					
... col_plus	11	92%					

Når vi sammenligner de seks gruppene, ser vi flere klare ulikheter mellom dem. Sammenligner vi kvinner uten inntekt og menn uten inntekt finner vi naturligvis forskjeller i fysikk, men vi merker oss spesielt at BMI er høyere for kvinner (30) enn menn (29). Andelen kvinner som er gift er 51 %, mot kun 30 % for menn. Kvinner har også høyere andel med utdanning, sammenlignet med menn.

For normal inntekt, som vi ser i Tabell 12 er menns gjennomsnittlige inntekt 54,236, mens kvinners ligger på 39,472. Menn er selvsagt høyere og tyngre enn kvinner, men vi merker oss at her har kjønnene lik BMI (29). 62 % av mennene med normal inntekt er gift, sammenlignet med 54 % av kvinnene. Utdanningsnivået er også her noe høyere blant kvinner (29 % mot 23 % for menn).

For de med høy inntekt er menn igjen høyere (71 tommer) og tyngre (199 lbs) enn kvinner (65 tommer, 150 lbs). Disse undergruppene skiller seg kraftig ut med at hele 83 % av kvinnene og 82 % av mennene er gift. Høy utdanning er vanlig, med hele 81 % av mennene og 92 % av kvinnene som har fullført høyere utdanning.

Totalt ser vi tydelige mønstre når det gjelder fysisk karakteristikk, sivilstatus og utdanningsnivå, som påvirkes av kjønn. Menn har høyere inntekter, større høyde og vekt, mens kvinner har høyere andel som er gift og høyere utdanning, spesielt i lav- og normalinntektsgruppene. Høy inntekt korrelerer med høyere utdanning for begge kjønn.

3.3 Interessante sammenhenger mellom variabler i datasettet normal inntekt

Basert på datasettet med normal inntekt, kan vi identifisere noen interessante sammenhenger mellom variablene. Vi ser at den gjennomsnittlige høyden i denne gruppen er 67 tommer, og at gjennomsnittlig vekt er 188 lbs, med en BMI på 29. Dette indikerer at både menn og kvinner i denne gruppen i gjennomsnitt ligger nær grensen for overvekt, noe som kan være relevant dersom vi diskuterer helseaspekter i relasjon til inntekt.

Når vi ser på sivilstatus, finner vi at 58 % av personene med normal inntekt er gift, med omtrent like stor andel menn og kvinner. Dette kan tyde på at inntektsnivået har en viss sammenheng med stabil sivilstatus, hvor gifte par er relativt likt fordelt på tvers av kjønn. Andelen personer med høyere utdanning er 26 %, som indikerer at en fjerdedel av gruppen har fullført høyere utdanning. Dette er en faktor som også kan være med på å forklare

lønnsnivået, siden høy utdanning ofte er korrelert med høyere inntekt, selv innenfor en relativt homogen inntektsgruppe.

Et annet interessant funn er at det er en relativt stor forskjell i utdanningsnivå innen normalinntektsgruppen. Mens 46 % av deltakerne har noe collegeutdanning, har kun 26 % høyere utdanning. Dette kan tyde på at høyere utdanning gir en fordel for å nå høyere inntektsnivåer, men at mange også klarer seg på dette nivået med mindre enn fullført collegeutdanning.

Oppsummert viser datasettet med normal inntekt at det er klare sammenhenger mellom variabler som høyde, vekt, sivilstatus og utdanningsnivå. Personer i denne gruppen har gjennomgående lik høyde og vekt, og en stor andel er gift. Utdanningsnivå varierer derimot betydelig, og det kan være interessante diskusjoner rundt hvordan dette påvirker økonomiske og sosiale utfall innen denne gruppen.

3.4 Samsvarer det som fremkommer fra den deskriptive statistikken med funnene fra litteraturen?

Når det kommer til høyde kan det ut fra den deskriptive statistikken vår virke som om litteraturen har rett når det kommer til høyde. I modellene hvor henholdsvis nullinntekt, normal inntekt og høy inntekt er filtrert ut ser vi at høyden øker noe for hver kategori. Likevel må det nevnes at andelen menn øker for hver gruppe, og flere menn vil også dra opp gjennomsnittshøyden. Splittet på kjønn virker det som høyden er mindre avgjørende, på tross av at gjennomsnittet er noe høyere for begge kjønn blant de som tjener best.

I litteraturen kom det frem at det var noe vage sammenhenger mellom BMI og inntekt. Men utfra den deskriptive statistikken vår kan det se ut som om litteraturen har rett i at en økning i BMI kan ha negativ påvirkning på inntekt, Men akkurat slik litteraturen peker på, ser det negative utfallet ut til å være størst blant kvinner.

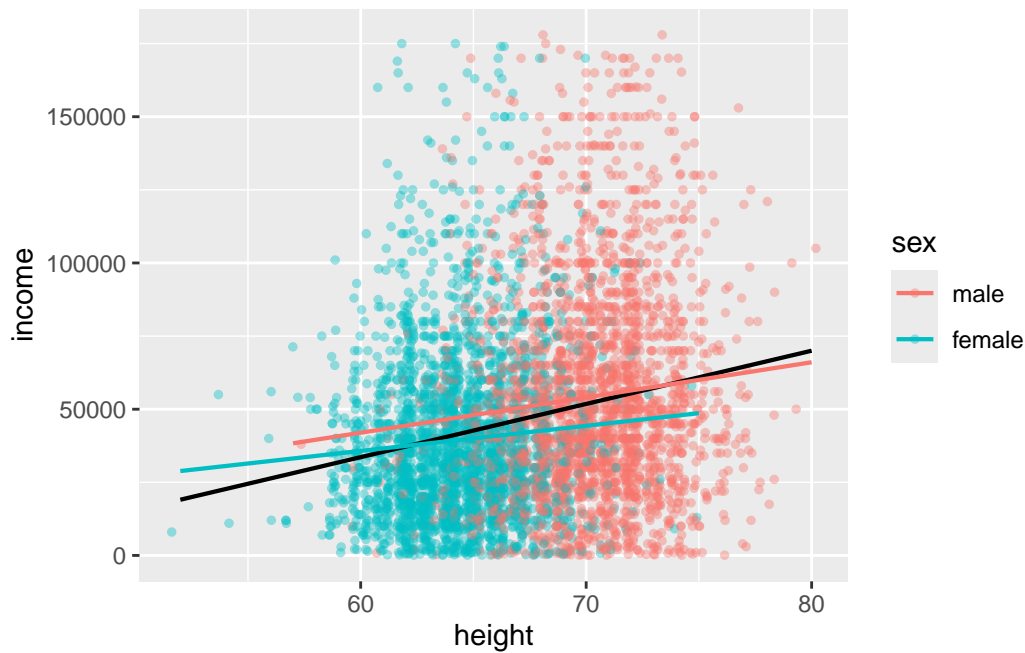
Utfra statistikkene våre er det også åpenbart at utdanning spiller en viktig rolle for inntekt. Dette viser igjen i fordelingen av utdanning for de med henholdsvis nullinntekt, normal inntekt og høy inntekt. Blant de som tjener best er andelen med høyere utdanning svært høy. Dette samsvarer også bra med informasjonen fra literaturdelen. Utfra litteraturedelen virker det samlet sett som om utdanning er blant de mest avgjørende variablene for utfallet av inntekt, og statistikkene våre trekker også i retning av dette.

Når det kommer til sammenhengen mellom kjønn og inntekt, ser vi at tabellene våre indikerer akkurat det samme som litteraturen, at menn generelt har høyere inntekt enn kvinner. Dette gjenspeiles både i at kvinner er overrepresentert blant de med null-inntekter, samt at kun 12 kvinner er blant de 147 med høyest inntekt.

Litteraturen peker på at sivilstatus kan ha stor betydning for inntekten, og det kan virke som om gifte personer generelt har høyest inntekt. Dette samsvarer bra med funnene våre fra den deskriptive statistikken. Det er særlig merkbart hvordan andelen gifte personer øker i takt med de tre inntektsgruppene våre.

4 Visualisering

```
heightsNormInc |>
  ggplot(
    mapping = aes(
      x = height,
      y = income,
      color = sex
    )
  ) +
  geom_jitter(
    size = 1,
    alpha = 0.40
  ) +
  geom_smooth(
    formula = y ~ x,
    method = "lm",
    colour = "black",
    lwd = 0.75,
    se = FALSE
  ) +
  geom_smooth(
    aes(colour = sex),
    formula = y ~ x,
    method = "lm",
    lwd = 0.75,
    se = FALSE)
```



Figur 1: Plot av inntekt mot høyde. Fra den sorte modell-linjen som er lagt inn ser vi at det ser ut til å være en positiv sammenheng mellom høyde og inntekt. Det er også lagt inn separate modell-linjer for kvinner og menn. Vi merker oss at sammenhengen mellom høyde og inntekt virker mindre når kjønn hensyntas i modell-linjene.

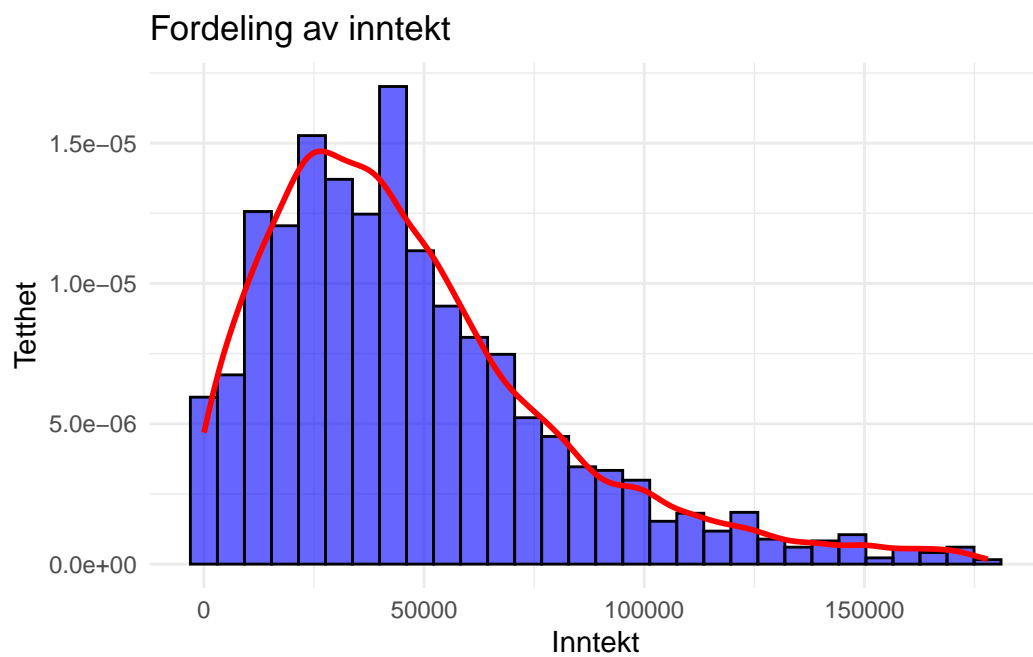
Plottet Figur 1 over forteller oss en rekke ting, og bekrefter en del av observasjonene som er diskutert tidligere i paperet. Fargene på observasjonene bekrefter det vi allerede vet, at menn generelt er høyere enn kvinner, og at flere menn enn kvinner har høy inntekt. Den sorte modell-linjen forteller oss i utgangspunktet at det virker å være en positiv sammenheng mellom høyde og inntekt. De røde og blå linjene viser også sammenheng mellom høyde og inntekt, men for henholdsvis menn og kvinner. Vi ser da at sammenhengen ikke er like tydelig, akkurat som diskutert tidligere i oppgaven, og merker oss at regresjonslinjen er brattere for menn enn for kvinner.

```
heightsNormInc %>%
  ggplot(aes(
    x = income
  )) +
  geom_histogram(aes(
    #y = ..density..,
    # ..density.. is outdated
    y = after_stat(density),
    bins = 30,
    fill = "blue",
    alpha = 0.6,
    color = "black"
  ))
```

```

) +
geom_density(
  color = "red",
  linewidth = 1
) +
labs(title = "Fordeling av inntekt",
  x = "Inntekt",
  y = "Tetthet") +
theme_minimal()

```



Figur 2: Histogram som viser fordeling av inntekt i populasjonen til hele datasettet

```

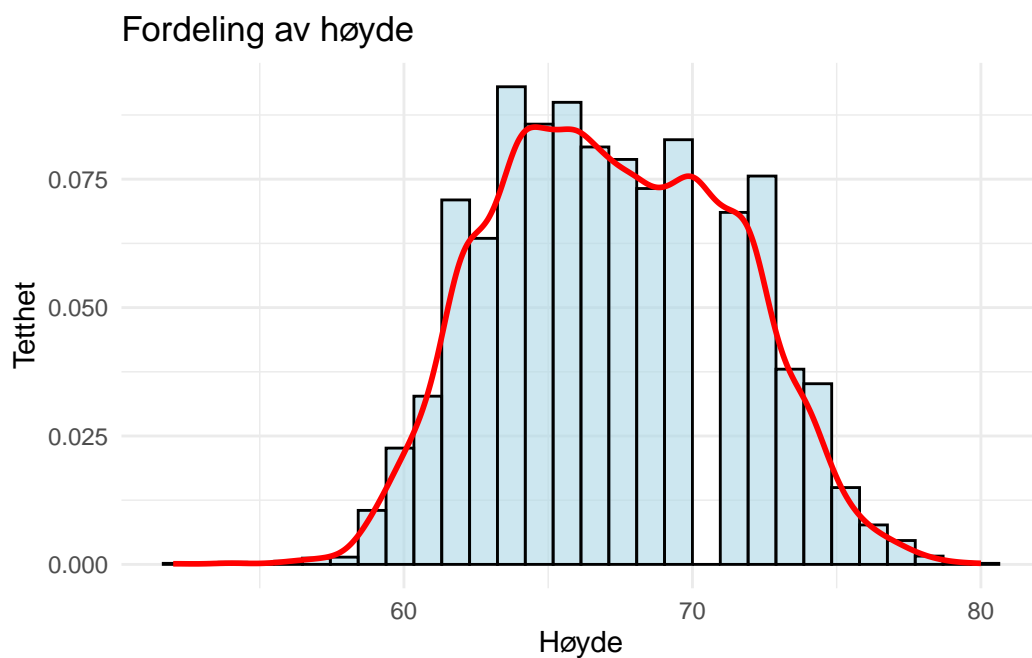
heightsNormInc %>%
  ggplot(
    mapping = aes(
      x = height
    )
  ) +
  geom_histogram(
    bins = 30,
    aes(y = ..density..),
    fill = "lightblue",
    alpha = 0.6,
    color = "black"
  ) +

```

```
geom_density(
  color = "red",
  size = 1
) +
labs(title = "Fordeling av høyde",
  x = "Høyde",
  y = "Tetthet") +
theme_minimal()
```

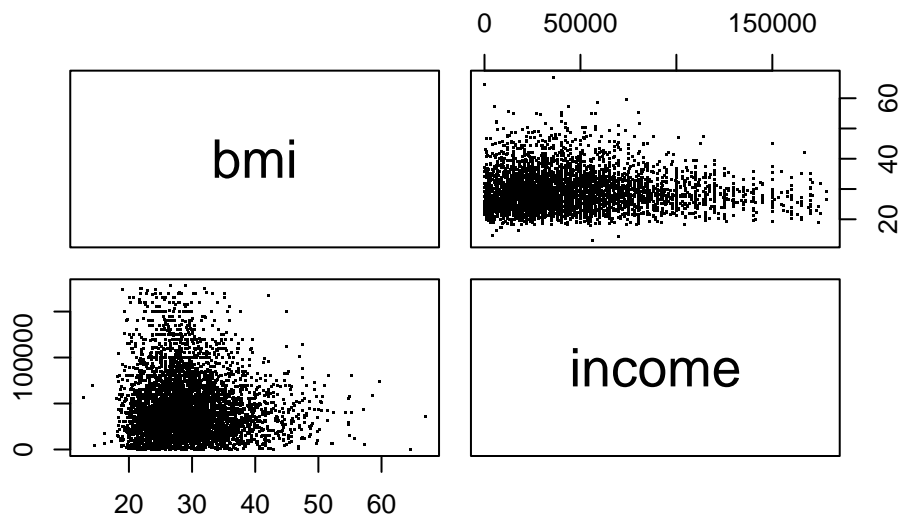
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.



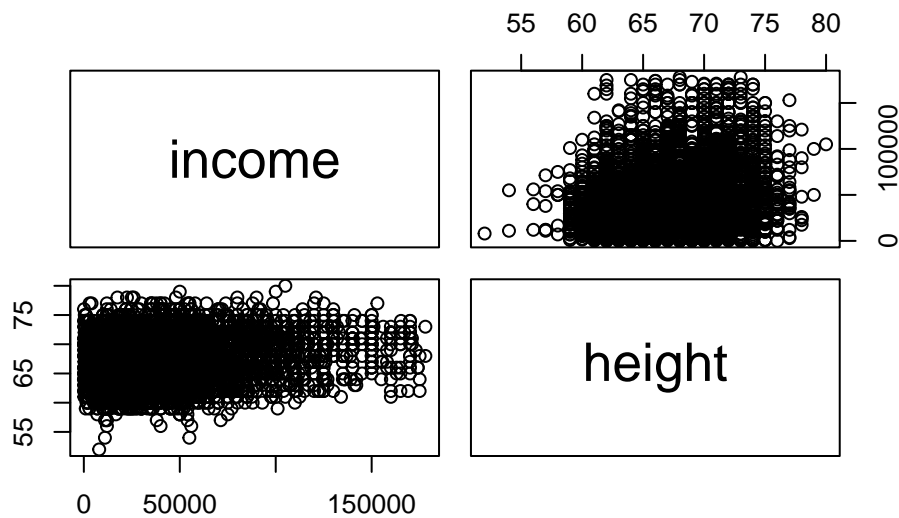
Figur 3: Histogram som viser fordeling av høyde i hele datasettet (populasjonen).

```
# ønsker å bruke pairs for å sammenligne inntekt og høyde
pairs(select(heightsNormInc, bmi, income), pch = '.', cex = 0.1)
```



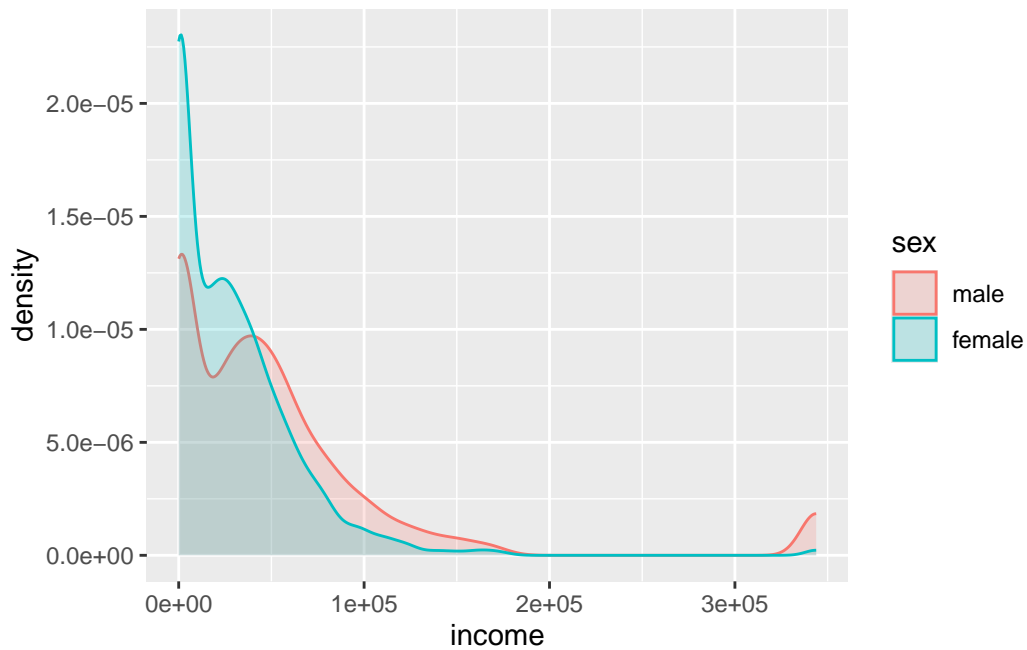
Figur 4: Korrelasjon mellom BMI og inntekt. Fuguren viser at det kan være tendenser til at økt BMI kan medføre redusert inntekt.

```
# ønsker å bruke pairs for å sammenligne inntekt og høyde  
pairs(select(heightsNormInc, income, height))
```

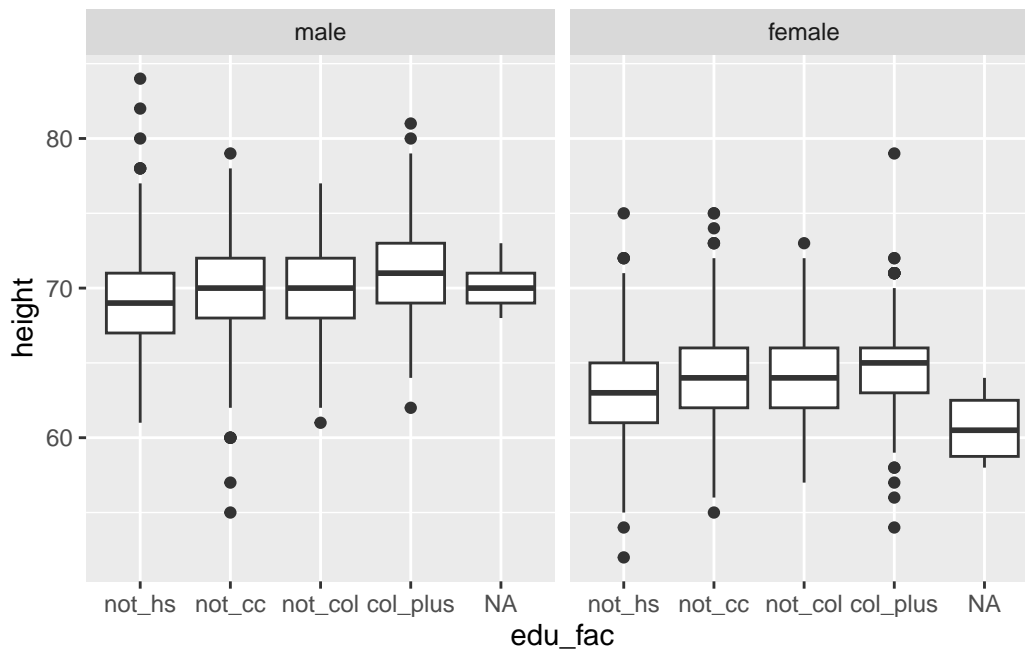
Figur 5: Korrelasjon mellom høyde og inntekt. Figuren viser at det ser ut til å være en korrelasjon mellom høyde og inntekt.

```
heights %>%
  ggplot(mapping = aes(x = income, fill = sex, colour = sex)) +
  geom_density(alpha = 0.2, na.rm = TRUE)
# geom_density(alpha = 0.2, na.rm = TRUE) +
# facet_wrap(~sex)
```



Figur 6: Figuren viser fordeling av inntekt mellom menn og kvinner i populasjonsutvalget. Vi merker oss at langt flere kvinner enn menn har ingen til svært lav inntekt. I tillegg ser vi klart flest menn blant de som tjener aller best.

```
heights %>%
  ggplot(mapping = aes(x = edu_fac, y = height)) +
  facet_wrap(~sex) +
  geom_boxplot()
```

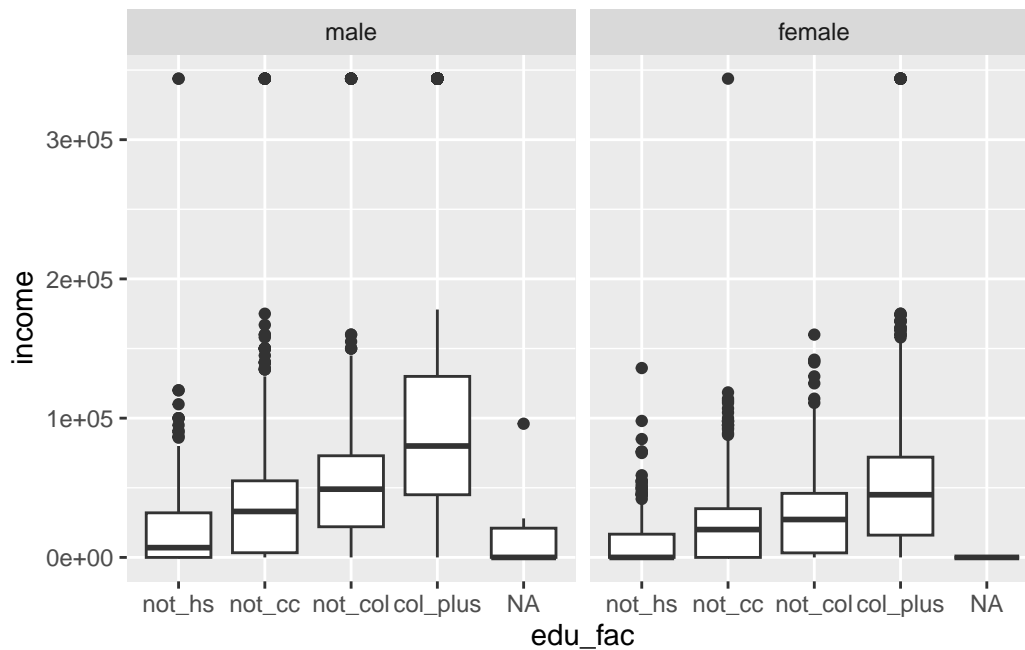


Figur 7: Boksplot av utdanning og høyde. Modellen er splittet på kjønn og viser menn til venstre, og kvinner til høyre. Det kan virke som det er en positiv sammenheng mellom høyde og utdanningsnivå både for kvinner og menn.

Utfra det vi har sett tidligere, både i teoridelen og i den deskriptive statistikken, kan det etter modellen Figur 7 tyde på at utdanning er en svært viktig faktor for inntekt. Den er derfor interessant å sjekke nærmere og høyde kan ha betydning for utdanningsnivå. Vi lager derfor en boxplot som viser utdanning i forhold til høyde, og facer denne mht kjønn.

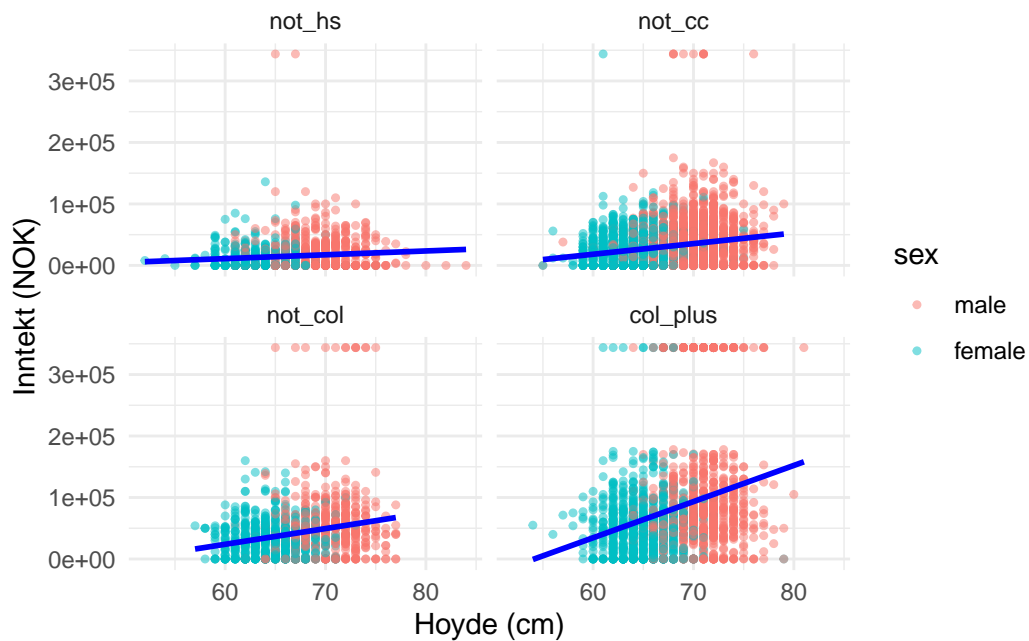
Vi merker oss at høyde ser ut til å påvirke utdanning en del, og videre skal vi se hvordan utdanning preger inntekt, via lignende boxplot.

```
heights %>%
  ggplot(mapping = aes(x = edu_fac, y = income)) +
  facet_wrap(~sex) +
  geom_boxplot()
```



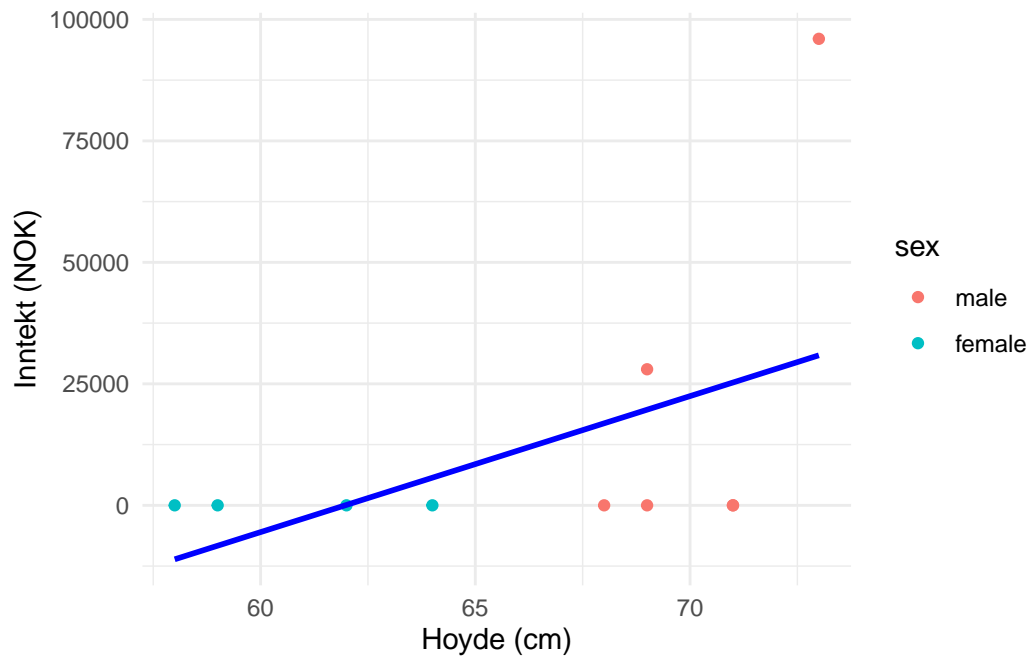
Figur 8: Boksplot av utdanning og inntekt. Modellen er splittet på kjønn og viser menn til venstre, og kvinner til høyre. Modellen viser tydelig at utdanning er en svært viktig variabel for økning av inntekt, både for menn og kvinner.

```
heights |>
  # Foreslår å ta ut observasjonene med NA for edu_fac slik at de resterende fire
  # kan plottes i større versjon
  filter(!is.na(edu_fac)) |>
  ggplot(
    mapping = aes(
      x = height,
      y = income,
      color = sex)) +
  geom_point(mapping = aes(), alpha = 0.5, size = 0.9) +
  geom_smooth(
    # bare for å bli kvitt den irriterende meldingen fra geom_smooth()
    formula = 'y ~ x',
    method = "lm",
    se = FALSE,
    color = "blue") + # Legger til en lineær trendlinje
  labs(
    x = "Høyde (cm)",
    y = "Inntekt (NOK)"
  ) +
  theme_minimal() +
  facet_wrap(~ edu_fac)
```



Figur 9: Figuren viser sammenheng mellom inntekt og høyde, skiller på utdanning og kjønn.

```
heights |>
  # Foreslår å ta ut observasjonene med NA for edu_fac slik at de resterende fire
  # kan plottes i større versjon
  filter(is.na(edu_fac)) |>
  ggplot(
    mapping = aes(
      x = height,
      y = income,
      color = sex)) +
    geom_point() +
    geom_smooth(
      # bare for å bli kvitt den irriterende meldingen fra geom_smooth()
      formula = 'y ~ x',
      method = "lm",
      se = FALSE,
      color = "blue") + # Legger til en lineær trendlinje
    labs(
      x = "Hoyde (cm)",
      y = "Inntekt (NOK)"
    ) +
    theme_minimal()
```

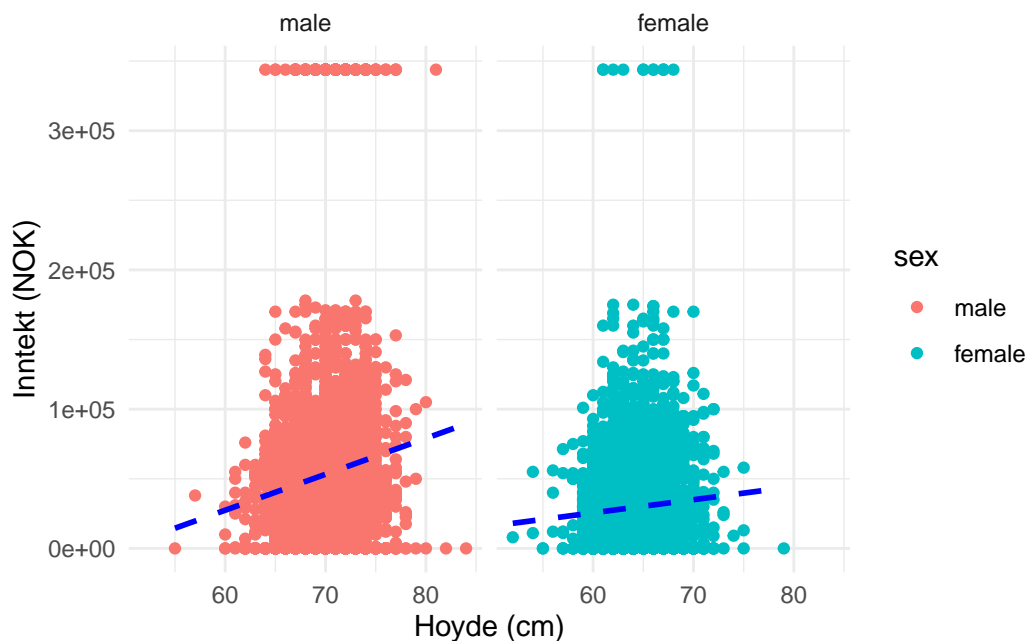


Figur 10: Figuren viser sammenheng mellom inntekt og høyde hvor opplysninger om «education» mangler.

Det kan virke som om høyde spiller en større rolle desto høyere utdanning man har.

```
gg_hic <- ggplot(data = heights,
  aes(
    x = height,
    y = income,
    color = sex)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue", linetype = "dashed") + # Legger
  labs(
    x = "Hoyde (cm)",
    y = "Inntekt (NOK)"
  ) +
  theme_minimal() +
  facet_wrap(~ sex)

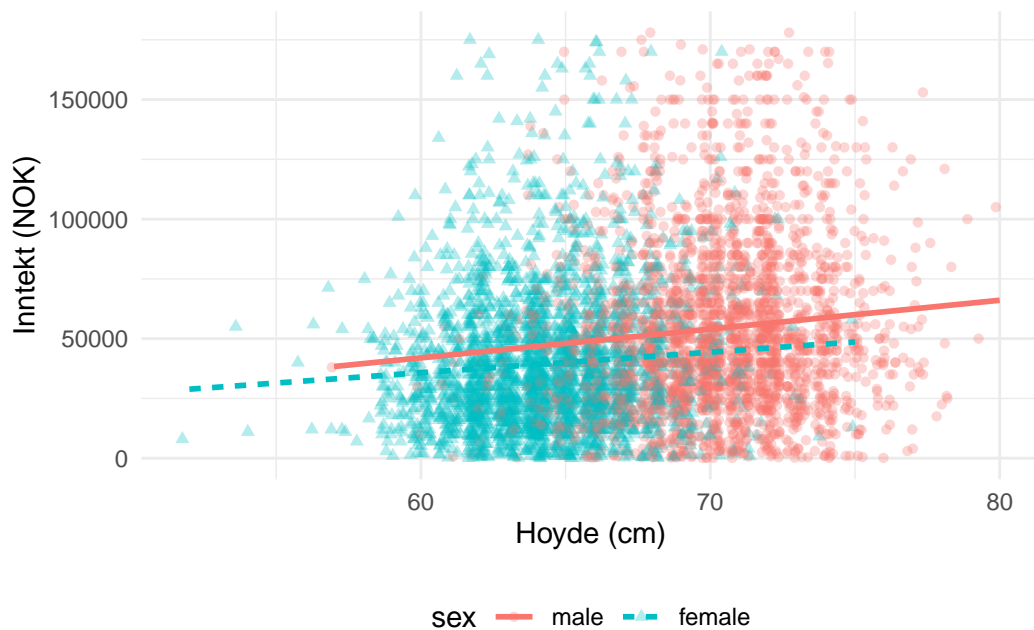
suppressMessages(print(gg_hic))
```



Figur 11: Modellen viser sammenhengden mellom høyde og inntekt for henholdsvis menn og kvinner. Det virker å være en klar sammenheng mellom høyde og inntekt for begge kjønn, men det er interessant å merke seg hvor mye brattere linjen er for menn enn kvinner. Dette kan trekke i retning av at høyde er viktigere for menn sin inntekt enn for kvinner.

```
gg_hic_normInc <- heightsNormInc |>
ggplot(
  aes(
    x = height,
    y = income,
    color = sex)) +
  geom_jitter(mapping = aes(shape = sex), alpha = 0.3) +
  #geom_point(mapping = aes(shape = sex), alpha = 0.3) +
  geom_smooth(
    mapping = aes(group = sex, color = sex, linetype = sex),
    method = "lm", se = FALSE) + # Legger til en lineær trendlinje
  labs(
    x = "Hoyde (cm)",
    y = "Inntekt (NOK)"
  ) +
  theme_minimal() +
  #facet_wrap(~ sex) +
  # for å få litt mer bredde til x aksene
  theme(legend.position = "bottom")
```

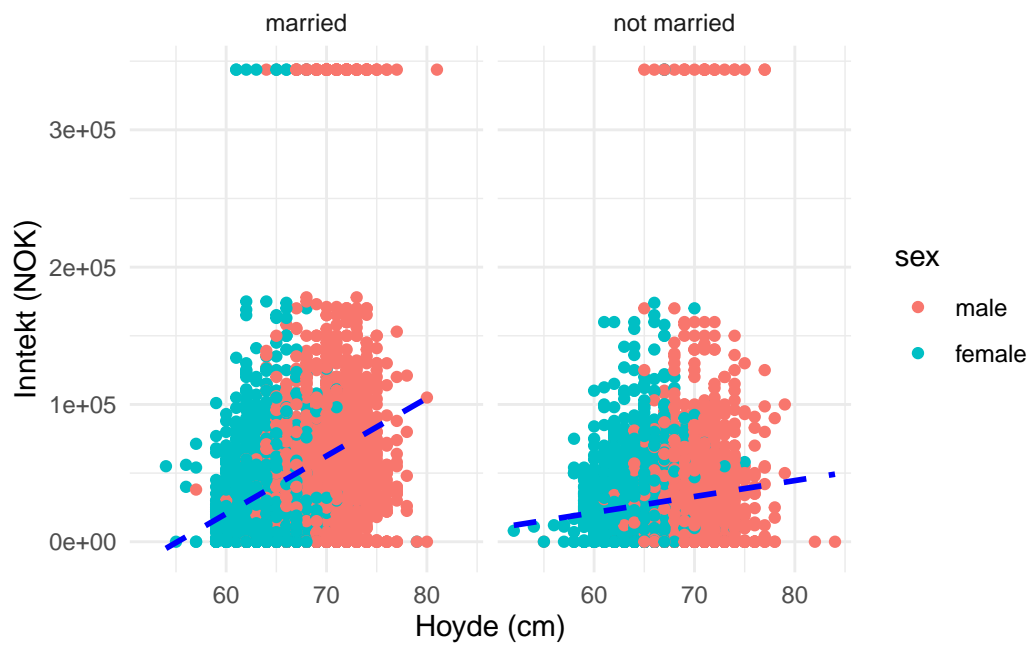
```
suppressMessages(print(gg_hic_normInc))
```



Figur 12: Modellen viser sammenhengden mellom høyde og inntekt for henholdsvis menn og kvinner for gruppen med normal inntekt (dvs. de svært høye inntektene er utelatt). Det virker å være en klar sammenheng mellom høyde og inntekt for begge kjønn, men det er interessant å merke seg hvor mye brattere linjen er for menn enn kvinner. Dette kan trekke i retning av at høyde er viktigere for menn sin inntekt enn for kvinner.

```
ggplot(data = heights,
  aes(
    x = height,
    y = income,
    color = sex)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "blue", linetype = "dashed") + # Legger
  labs(
    x = "Hoyde (cm)",
    y = "Inntekt (NOK)"
  ) +
  theme_minimal() +
  facet_wrap(~ married)
```

```
`geom_smooth()` using formula = 'y ~ x'
```

Figur 13: Sammenheng mellom inntekt og høyde, skiller på utdanning og kjønn.

```
heights %>%
  ggplot(mapping = aes(x = married, y = income)) +
  facet_wrap(~sex) +
  geom_boxplot()
```



Figur 14: Boksplotet viser inntekt mot ekteskap. Modellen er splittet på kjønn og viser menn til venstre, og kvinner til høyre. Det er interessant hvordan gifte menn har betydelig høyere inntekt enn ugifte menn. Det er en liten tendens til dette også for kvinner, men i langt mindre grad.

5 Er det virkelig høyde som bestemmer inntekt?

I dette minipaperet har vi innhentet litteratur fra eksterne kilder, i tillegg til å utføre vår egen EDA-analyse for å se om det virkelig er slik at høyde bestemmer inntekt. Vi har benyttet datasettet “heights”, og laget en rekke deskriptive statistikker for å prøve å avsløre at andre variabler spiller en viktig rolle.

Utfra litteraturdelen kan kunne det virke som om høyde spiller en svært viktig rolle for inntekt. Det ble pekt på flere årsaker til dette, blant annet bedre kognitive ferdigheter, og klare sosiale oppfatninger i samfunnet. Også den deskriptive statistikken vår gav inntrykk for dette, i form av at gjennomsnittshøyden var høyere for gruppen med normal inntekt enn for gruppen med lav inntekt, og vesentlig høyere blant de med høyest inntekt. Det ble likevel også lagt vekt på at menn generelt har høyere inntekt enn kvinner, og at menn av naturlige årsaker er høyere enn kvinner. Det var økende andel menn i de ulike undergruppene, noe som derfor også drar opp gjennomsnittet i gruppene. Påvirkningskraften til høyde som variabel syntest å bli noe mindre etterhvert som undergruppene ble filtrert på kjønn, selv om den fortsatt var merkbar.

Det kommer også klart frem av analysen at det er andre faktorer som også er svært avgjørende for en persons inntekt. Utdanning trekkes kanskje frem som den klareste faktoren, og dette gjenspeiler seg både i statistikken og i grafene våre. Det ble også gjort en interessant

observasjon på dette området som viser at det virker å være korrelasjon mellom høyde og utdanning. Dette er særlig interessant for oppgavens overordnede problemstilling, ettersom utdanning synes å være den potensielt mest forklarende variabelen for inntekt.

Det ble også gjort interessante funn i forholdet mellom sivilstatus og inntekt som vi ser i Figur 14. Gifte personer tjener i gjennomsnitt mer, og andelen som er gift er økende i takt med de tre undergruppene som er benyttet i analysen.

Samlet sett trekker analysene i retning av at vi langt i fra med sikkerhet kan si at det isolert sett er høyde som bestemmer inntekt, og at det er mange flere faktorer som spiller inn. Det er likevel klare indikasjoner på at høyde kan være en svært viktig faktor for inntekt, og at inntekten reelt sett er økende i takt med høyde. Dette virker å henge sammen med at det kan se ut som om høyde påvirker flere av de andre viktige variablene som kan forklare inntekt, deriblant utdanning og sivilstatus. Det virker derfor som om høyde indirekte påvirker inntekt i merkbart grad, men at den alene ikke er ensbetydende for en persons inntekt.

- Britt, Robert Roy. 2009. «Taller People Earn More Money». *LiveScience.com*. <https://www.livescience.com/5552-taller-people-earn-money.html>.
- Bureau, United States Census. 2019. «College Degree Widens Gender Earnings Gap». 2019. <https://www.census.gov/library/stories/2019/05/college-degree-widens-gender-earnings-gap.html>.
- Böckerman, Petri, John Cawley, Jutta Viinikainen, Terho Lehtimäki, Suvi Rovio, Ilkka Seppälä, Jaakko Pehkonen, og Olli Raitakari. 2019. «The Effect of Weight on Labor Market Outcomes: An Application of Genetic Instrumental Variables». *Health Economics* 28 (1): 65–77.
- Case, Anne, og Christina Paxson. 2006. «Stature and Status: Height, Ability, and Labor Market Outcomes». *Princeton University*. https://www.princeton.edu/~accase/downloads/Stature_and_Status.pdf.
- Census, US. n.d. «Among the Educated, Women Earn 74 Cents for Every Dollar Men Make». *Census.gov*. <https://www.census.gov/library/stories/2019/05/college-degree-widens-gender-earnings-gap.html>.
- Edwards, Christina Hansen, Johan Håkon Bjørngaard, og Jonas Minet Kinge. 2021. «The relationship between body mass index and income: Using genetic variants from HUNT as instrumental variables». *Health Economics*. <https://doi.org/10.1002/hec.4285>.
- Gould, Elise, Jessica Schieder, og Kathleen Geier. 2016. «What Is the Gender Pay Gap and Is It Real?: The Complete Guide to How Women Are Paid Less Than Men and Why It Can't Be Explained Away». *Economic Policy Institute*. <https://www.epi.org/publication/what-is-the-gender-pay-gap-and-is-it-real/>.
- Howard, David H. 2018. «Married Men Sit Atop the Wage Ladder». 2018. <https://research.stlouisfed.org/publications/economic-synopses/2018/09/14/married-men-sit-atop-the-wage-ladder>.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. Columbia University Press.
- On-The-Economy-Blog. 2020. «Taking a Closer Look at Marital Status and the Earnings Gap». *St. Louis Fed*. <https://www.stlouisfed.org/on-the-economy/2020/september/taking-closer-look-marital-status-earnings-gap>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Vandenbroucke, Guillaume. n.d. «Married Men Sit Atop the Wage Ladder». <https://research.stlouisfed.org/publications/economic-synopses/2018/09/14/married-men-sit-atop-the-wage-ladder>.
- Wickham, Hadley. 2023. *modelr: Modelling Functions that Work with the Pipe*. <https://CRAN.R-project.org/package=modelr>.