

Oppgave 3, Seigmannoppgaven

Vi kjører for begge datasett samtidig. Altså begge seigmann typer. Vi begynner med å vise hvordan du kan importere data i stedet for å skrive rett inn i R. Her importerer vi rådata, altså enkeltdata med farge og lengde.

```
# Om importen går feil, sjekk innstillinger for desimal-tegn fra excel, og karakter-koding (helst UTF-8)
# read_csv2 har ";" som default seperator
# read_csv har "," som default seperator

dfLabanRAW <- read.csv2(file = "seigmennRAW1.csv", header = TRUE, sep=";", dec=".")
dfBrynildRAW <- read.csv(file = "seigmennRAW2.csv", header = TRUE, sep=",", dec=".")

# Dette er rådata. Vi gjør det om til 2-veis tabell med table:

dfLabanTable<-table(dfLabanRAW)

# ... og så til vanlig 1-dim. frekvenstabell med to kriterier Farge og Lengde
# med as.data.frame

dfLaban <- as.data.frame(dfLabanTable)

# Har du hatt en tabell av dette slaget fra Excel, er du allerede her ved kommandoene
# dfLaban <- read.csv2(file = "seigmenn1.csv", header = TRUE, sep=";", dec=".")
# dfLaban <- read.csv(file = "seigmenn2.csv", header = TRUE, sep=",", dec=".")
# så vi tar dette som startpunkt.

dfBrynild <- as.data.frame(table(dfBrynildRAW))

# Om det trengs, så kan vi tvinge variablene til å være av riktig type:
# Og det trengs, for table gjør om Lengde og Farge tilen "factor", og den får vi
# tilbake til tallform ved TO konverteringer. (Direkte konvertering funker ikke)
```

```
dfLaban$Farge = as.character(dfLaban$Farge)
dfBrynild$Farge = as.character(dfBrynild$Farge)
dfLaban$Lengde = as.double(as.character(dfLaban$Lengde))
dfBrynild$Lengde = as.double(as.character(dfBrynild$Lengde))
```

3a

Frekvenstabell for farger. Lager du den manuelt, skriver du den bare opp, slik:

Farge	Antall
Grønn	8
Gul	4
Rød	13
Oransj	5

Men vil du skrive det inn i R, så skriv input slik:

```
# Alt. 1: Skriv inn fargene en om gangen
# fargene <- c("Grønn", "Rød", "Grønn", etc )

# Alt. 2: Bruk rep for å skrive inn farge+antall.
# farger<-c(rep("Grønn",8),rep("Gul",4),rep("Rød",13),rep("Oransj",5))

# Siden vi allerede har dataene importert, så får vi ut rene rådata med farger
# ved metode nummer 2:

farger1 <- c(rep(dfLaban$Farge,dfLaban$Freq))
farger2 <- c(rep(dfBrynild$Farge,dfBrynild$Freq))
```

Så til det vi egentlig skulle gjøre, output:

```
# Frekvenstabell i R blir da ganske enkel, ved å bruke kommandoen table:
```

```
table(farger1)
```

```
## farger1
##   Grønn   Gul Oransj   Rød
##     17    14     11     8
```

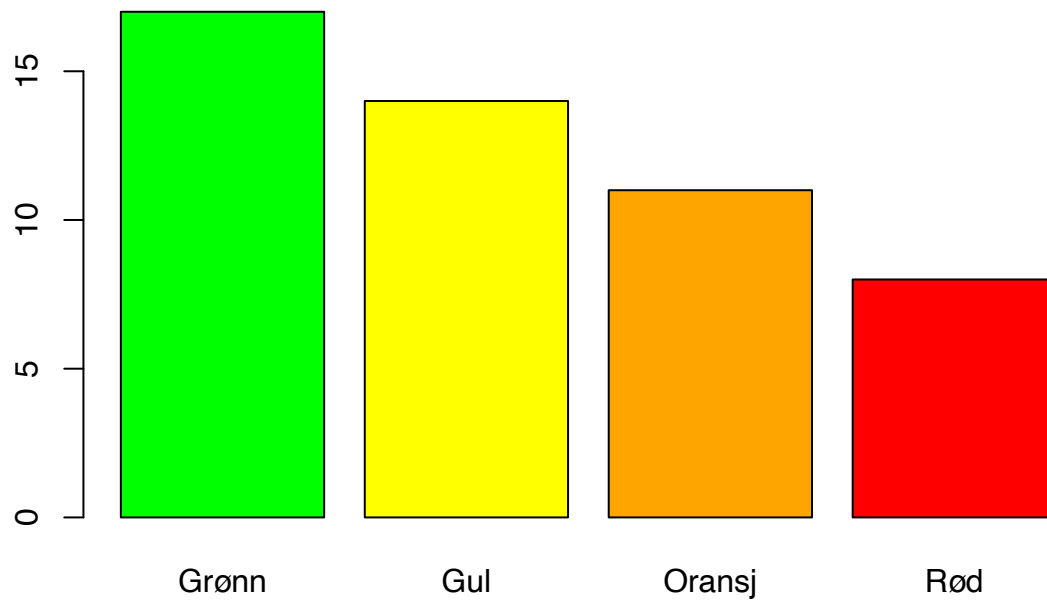
```
table(farger2)
```

```
## farger2
##   Grønn   Gul Oransj   Rød
##     10    10     11    19
```

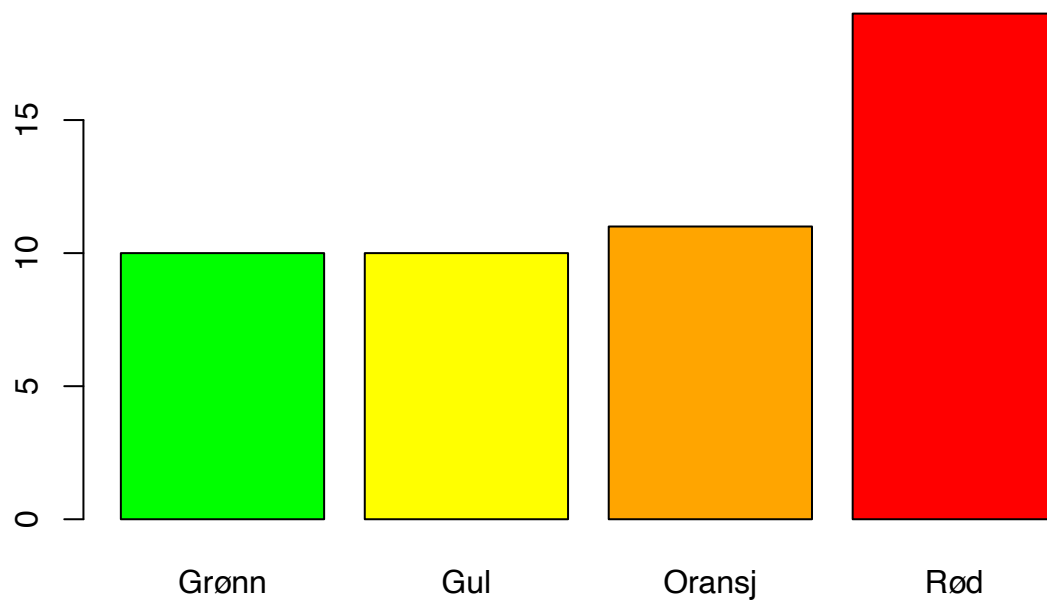
3b

Frekvensdiagram for farger.

```
barplot(table(farger1),col=c("green","yellow","orange","red"))
```



```
barplot(table(farger2),col=c("green","yellow","orange","red"))
```



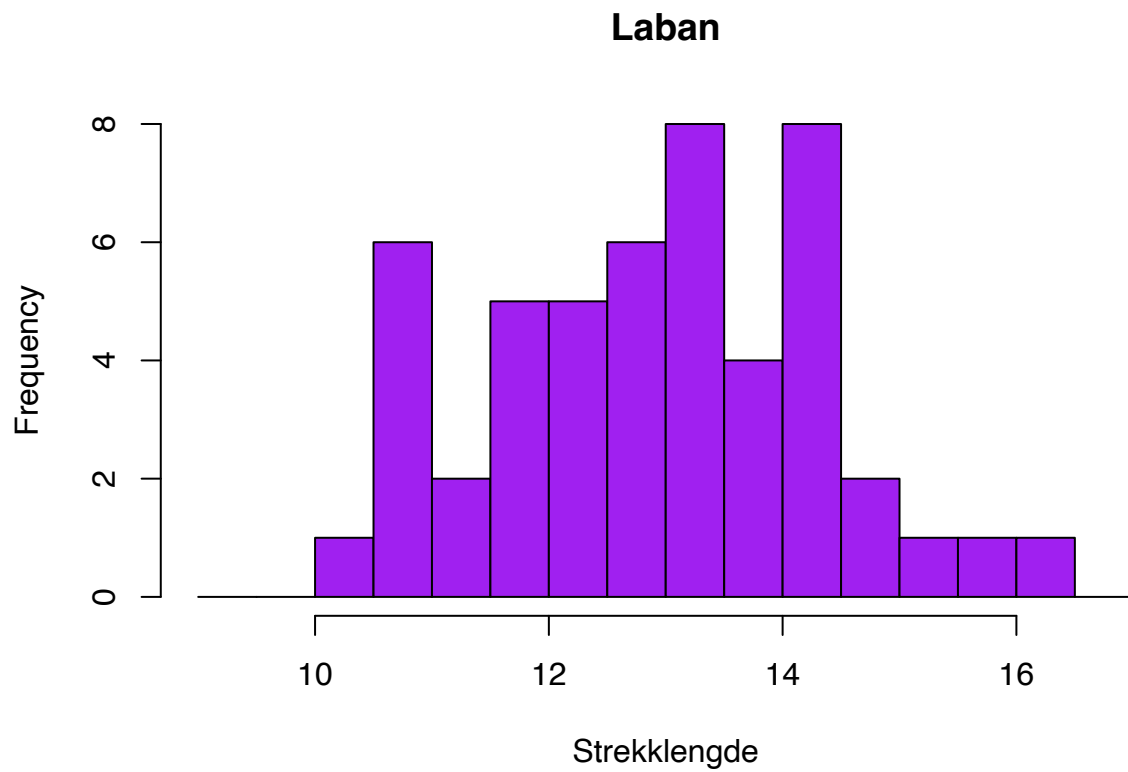
3cde

```
# Vi tar utgangspunkt i frekvensdata. Har du rådata, hopper du over dette:

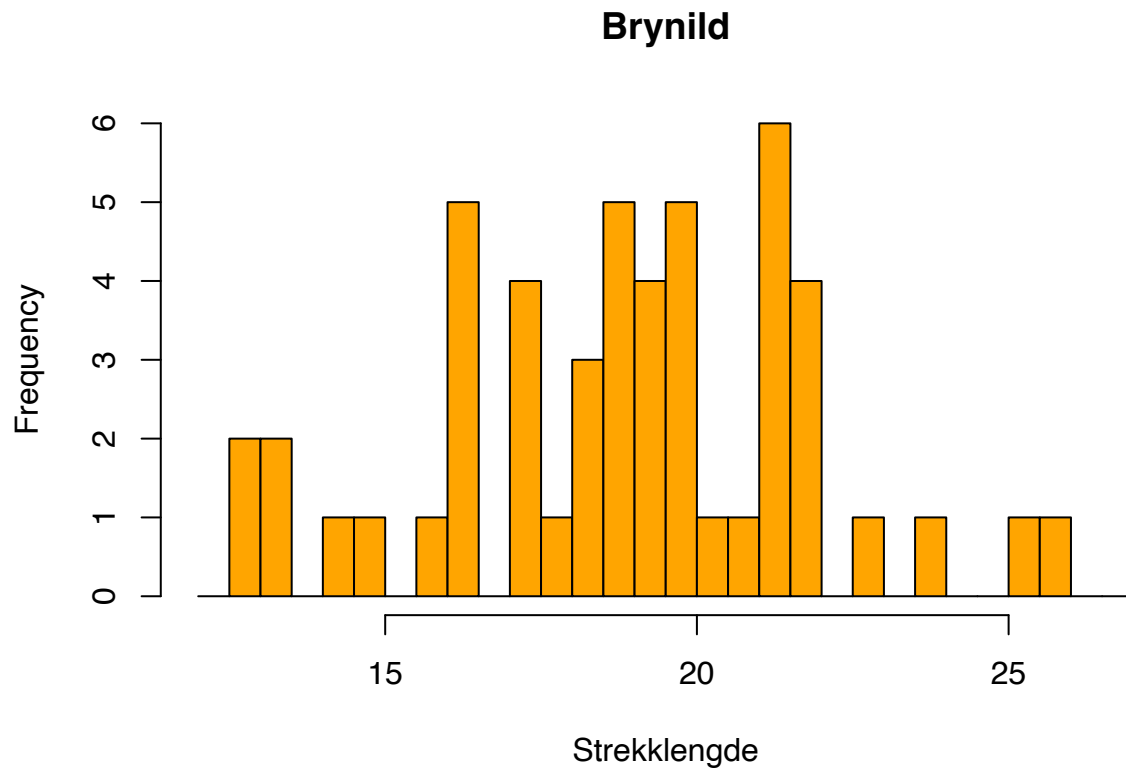
lengthsLaban <- rep(dfLaban$Lengde,dfLaban$Freq)
lengthsBrynild <- rep(dfBrynild$Lengde,dfBrynild$Freq)
```

3f

```
hist(lengthsLaban, col="purple", breaks=seq(9,17,0.5), main="Laban", xlab="Strekklengde")
```



```
hist(lengthsBrynild, col="orange", breaks=seq(12,27,0.5), main="Brynild", xlab="Strekklengde")
```



3g

Frekvenstabell basert på datasettet.

Hvis du ikke setter opp manuelt

```
Laban <- setNames(data.frame(table(lengthsLaban)), c("Lengde", "Antall"))
Brynild <- setNames(data.frame(table(lengthsBrynild)), c("Lengde", "Antall"))
```

Vi må passe på å gjøre tall til tall og tekst til tekst:

```
Laban$Lengde = as.double(as.character(Laban$Lengde))
Brynild$Lengde = as.double(as.character(Brynild$Lengde))
Laban$Antall = as.integer(as.character(Laban$Antall))
Brynild$Antall = as.integer(as.character(Brynild$Antall))
```

Da er tabellene:

Laban

```
##   Lengde Antall
## 1   10.5      1
## 2   11.0      6
## 3   11.5      2
## 4   12.0      5
## 5   12.5      5
## 6   13.0      6
## 7   13.5      8
```

```
## 8      14.0      4
## 9      14.5      8
## 10     15.0      2
## 11     15.5      1
## 12     16.0      1
## 13     16.5      1
```

Brynild

```
##      Lengde Antall
## 1      13.0      2
## 2      13.5      2
## 3      14.5      1
## 4      15.0      1
## 5      16.0      1
## 6      16.5      5
## 7      17.5      4
## 8      18.0      1
## 9      18.5      3
## 10     19.0      5
## 11     19.5      4
## 12     20.0      5
## 13     20.5      1
## 14     21.0      1
## 15     21.5      6
## 16     22.0      4
## 17     23.0      1
## 18     24.0      1
## 19     25.5      1
## 20     26.0      1
```

3h

Vi tar utgangspunkt i intervaller på 0.5cm, og at tallet vi har operert med er midtpunktet. Mer avanserte ting går an, men det krever mer avansert koding. Modifiser koden her dersom du har kommet til det stadiet.

Vi endrer navnet Lengde til Midtpunkt. Dette er 1. kolonne, så ...

```
names(Laban)[1]="Midtpunkt"
```

Å lage nye kolonne i dataframes er ENKELT:

```
Laban$Bredde=0.5
Laban$Kumulativ=cumsum(Laban$Antall)
Laban$Venstre=Laban$Midtpunkt-0.5*Laban$Bredde
Laban$Høyre=Laban$Midtpunkt+0.5*Laban$Bredde
```

Hvis vi vil liste kolonnene i ønsket rekkefølge skriver vi ...

```
Laban[c("Venstre", "Høyre", "Midtpunkt", "Bredde", "Antall", "Kumulativ")]
```

```
##      Venstre Høyre Midtpunkt Bredde Antall Kumulativ
## 1      10.25 10.75      10.5   0.5      1          1
## 2      10.75 11.25      11.0   0.5      6          7
```

```
## 3    11.25 11.75      11.5    0.5    2      9
## 4    11.75 12.25      12.0    0.5    5     14
## 5    12.25 12.75      12.5    0.5    5     19
## 6    12.75 13.25      13.0    0.5    6     25
## 7    13.25 13.75      13.5    0.5    8     33
## 8    13.75 14.25      14.0    0.5    4     37
## 9    14.25 14.75      14.5    0.5    8     45
## 10   14.75 15.25      15.0    0.5    2     47
## 11   15.25 15.75      15.5    0.5    1     48
## 12   15.75 16.25      16.0    0.5    1     49
## 13   16.25 16.75      16.5    0.5    1     50
```

Og så for Brynild

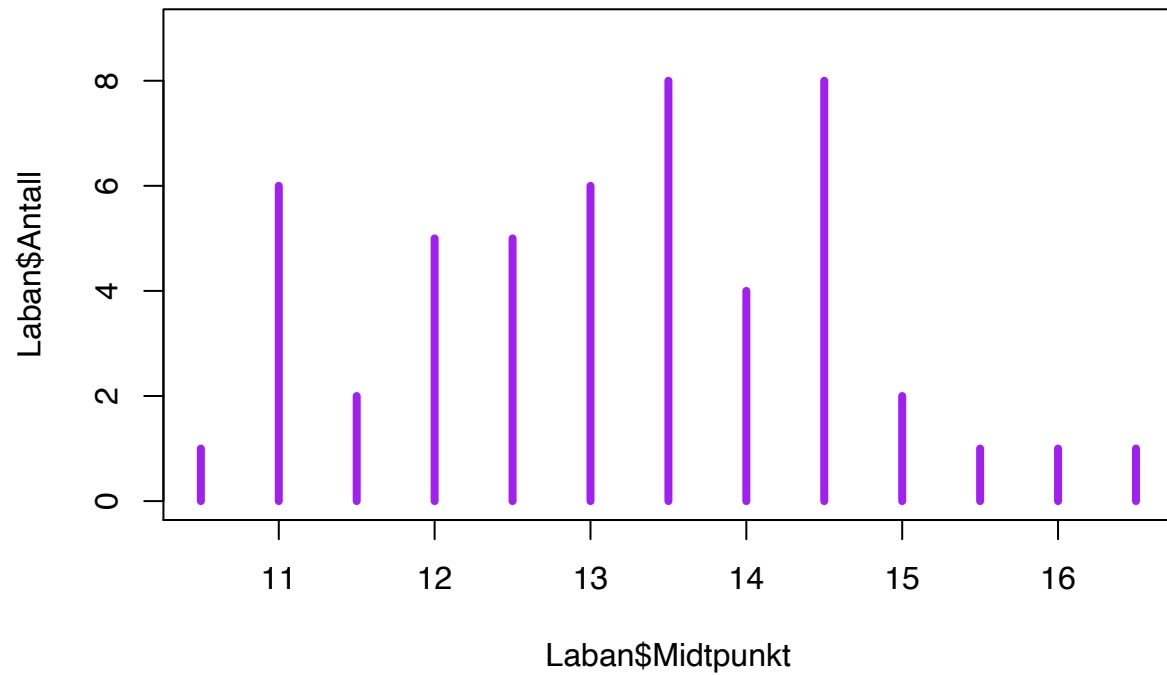
```
names(Brynild)[1]="Midtpunkt"
Brynild$Bredde=0.5
Brynild$Kumulativ=cumsum(Brynild$Antall)
Brynild$Venstre=Brynild$Midtpunkt-0.5*Brynild$Bredde
Brynild$Høyre=Brynild$Midtpunkt+0.5*Brynild$Bredde
Brynild[c("Venstre", "Høyre", "Midtpunkt", "Bredde", "Antall", "Kumulativ")]
```

```
##      Venstre Høyre Midtpunkt Bredde Antall Kumulativ
## 1    12.75 13.25      13.0    0.5    2      2
## 2    13.25 13.75      13.5    0.5    2      4
## 3    14.25 14.75      14.5    0.5    1      5
## 4    14.75 15.25      15.0    0.5    1      6
## 5    15.75 16.25      16.0    0.5    1      7
## 6    16.25 16.75      16.5    0.5    5     12
## 7    17.25 17.75      17.5    0.5    4     16
## 8    17.75 18.25      18.0    0.5    1     17
## 9    18.25 18.75      18.5    0.5    3     20
## 10   18.75 19.25      19.0    0.5    5     25
## 11   19.25 19.75      19.5    0.5    4     29
## 12   19.75 20.25      20.0    0.5    5     34
## 13   20.25 20.75      20.5    0.5    1     35
## 14   20.75 21.25      21.0    0.5    1     36
## 15   21.25 21.75      21.5    0.5    6     42
## 16   21.75 22.25      22.0    0.5    4     46
## 17   22.75 23.25      23.0    0.5    1     47
## 18   23.75 24.25      24.0    0.5    1     48
## 19   25.25 25.75      25.5    0.5    1     49
## 20   25.75 26.25      26.0    0.5    1     50
```

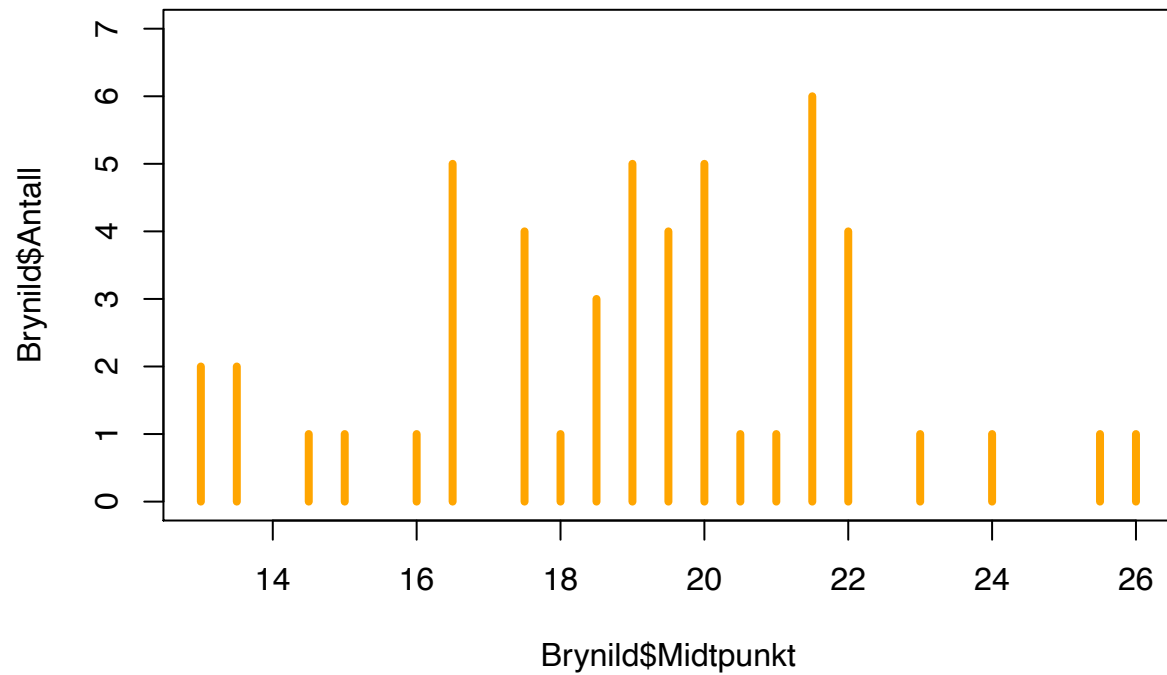
3i

Dette har du essensielt sett gjort om du tegnet histogrammet med riktige bredder, men med dataframe har du også muligheten til å bruke plot.

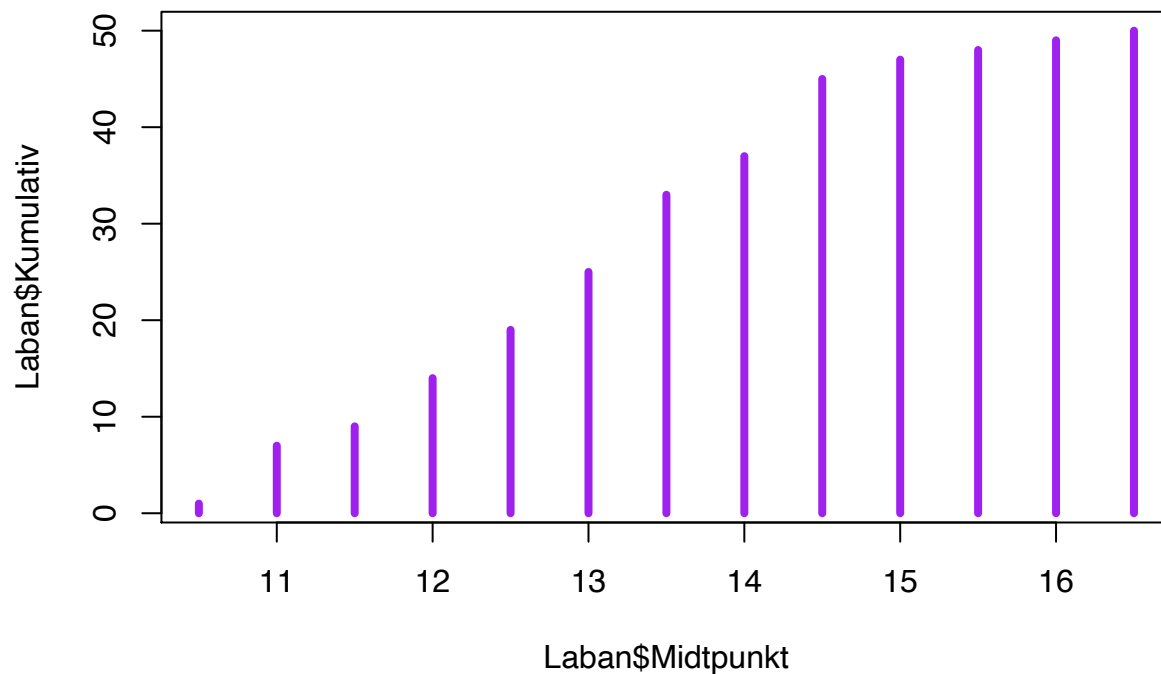
```
plot(Laban$Midtpunkt,Laban$Antall,ylim=c(0,max(Laban$Antall+1)),type="h",lwd=4,col="purple")
```



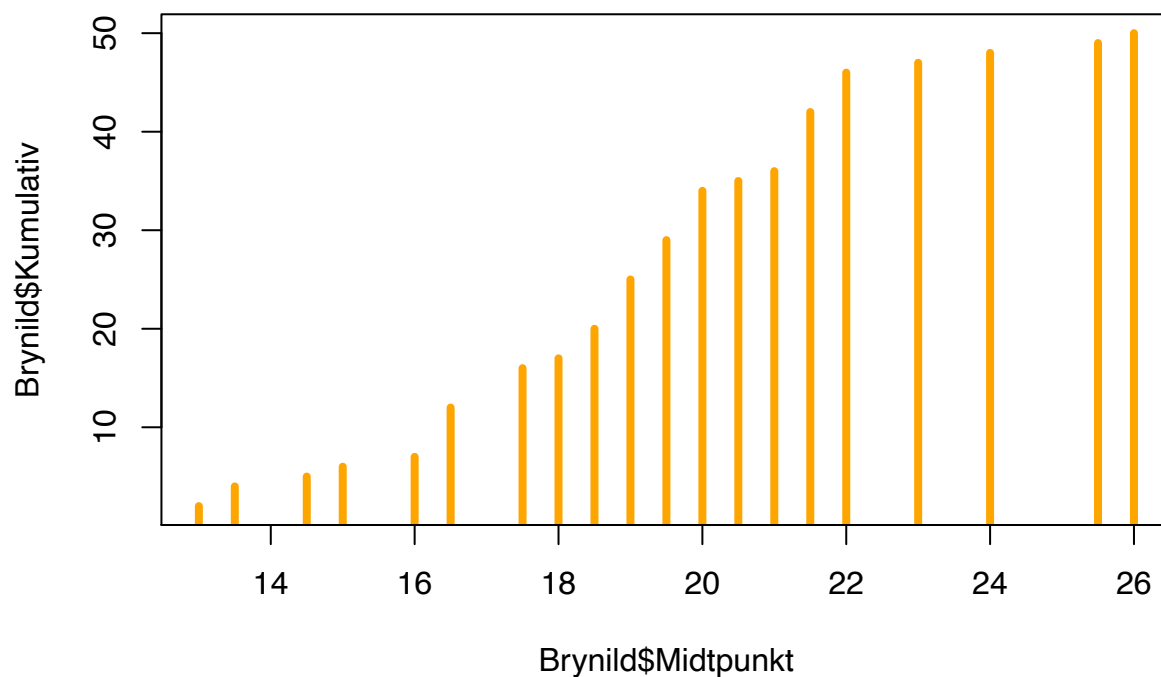
```
plot(Brynild$Midtpunkt,Brynild$Antall,ylim=c(0,max(Brynild$Antall+1)),type="h",lwd=4,col="orange")
```



```
plot(Laban$Midtpunkt,Laban$Kumulativ,type="h",lwd=4,col="purple")
```

```
plot(Brynild$Midtpunkt,Brynild$Kumulativ,type="h",lwd=4,col="orange")
```



```
meanLaban = mean(lengthsLaban)
medLaban = median(lengthsLaban)
maxLaban = max(Laban$Antall)
modeLaban = Laban[Laban$Antall==maxLaban,]$Midtpunkt

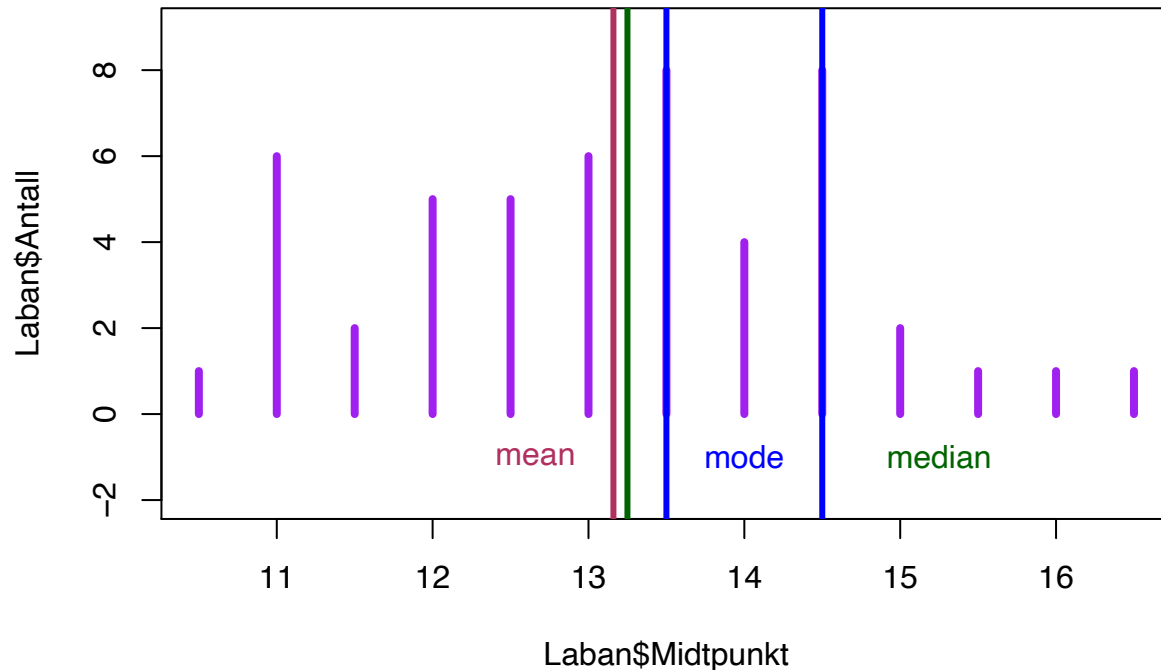
meanBrynild = mean(lengthsBrynild)
medBrynild = median(lengthsBrynild)
maxBrynild = max(Brynild$Antall)
modeBrynild = Brynild[Brynild$Antall==maxBrynild,]$Midtpunkt
```

Det forventes IKKE at dere kan finne mode ved koding. det gjør dere
manuelt ved å se gjennom tabellen

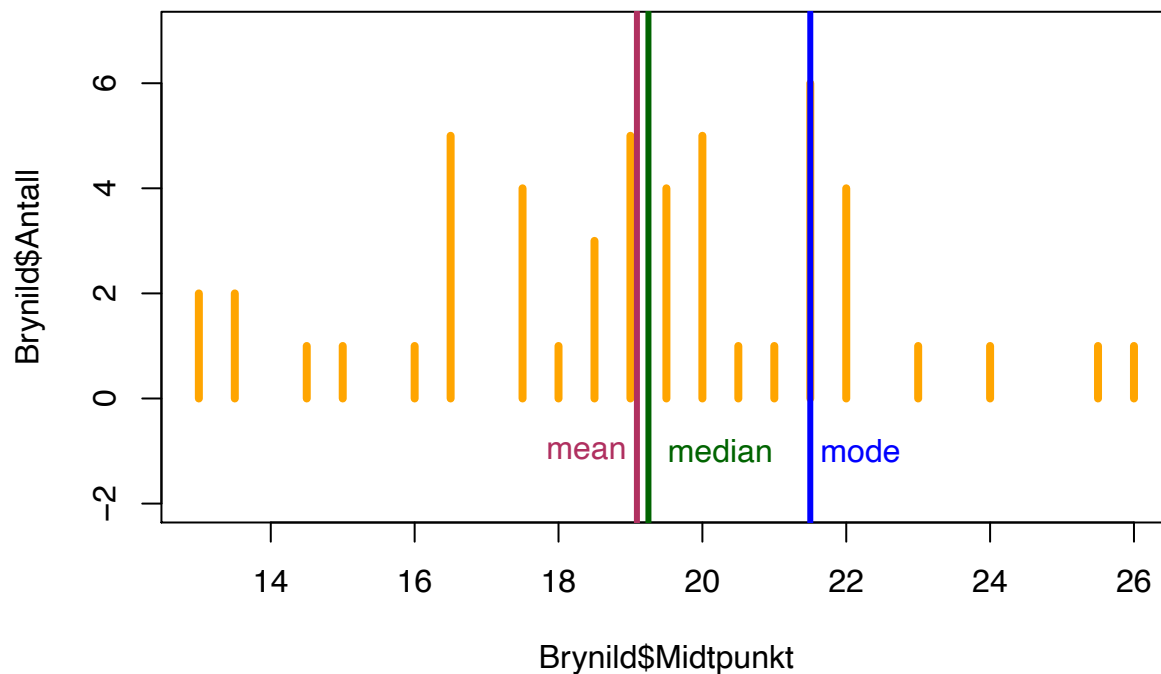
	Laban	Brynild
Gjennomsnitt	13.16	19.09
Median	13.25	19.25
Modus (Typetall)	13.5, 14.5	21.5

31

```
plot(Laban$Midtpunkt, Laban$Antall, ylim=c(-2, max(Laban$Antall+1)), type="h", lwd=4, col="purple")
abline(v = meanLaban, col = "maroon", lwd = 3)
text(meanLaban-0.5, -1, "mean", col="maroon")
abline(v = medLaban, col = "darkgreen", lwd = 3)
text(medLaban+2, -1, "median", col="darkgreen")
abline(v = modeLaban, col = "blue", lwd = 3)
text(mean(modeLaban), -1, "mode", col="blue")
```



```
plot(Brynild$Midtpunkt, Brynild$Antall, ylim=c(-2, max(Brynild$Antall+1)), type="h", lwd=4, col="orange")
abline(v = meanBrynild, col = "maroon", lwd = 3)
text(meanBrynild-0.7, -1, "mean", col="maroon")
abline(v = medBrynild, col = "darkgreen", lwd = 3)
text(medBrynild+1, -1, "median", col="darkgreen")
abline(v = modeBrynild, col = "blue", lwd = 3)
text(modeBrynild+0.7, -1, "mode", col="blue")
```



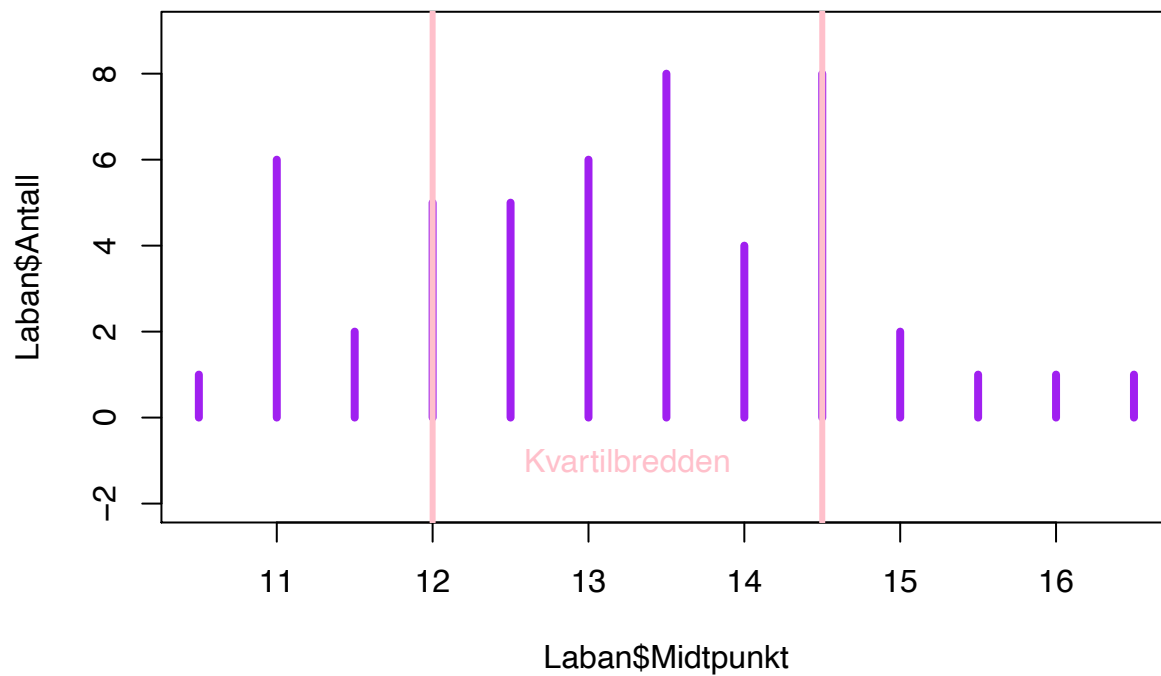
```
q1Laban = quantile(lengthsLaban,0.25,type=6)
q3Laban = quantile(lengthsLaban,0.75,type=6)
kvartilbreddeLaban = q3Laban-q1Laban
n=length(lengthsLaban)
sampSdLaban = sd(lengthsLaban)
popSdLaban = sqrt((n-1)/n)*sampSdLaban

q1Brynild = quantile(lengthsBrynild,0.25,type=6)
q3Brynild = quantile(lengthsBrynild,0.75,type=6)
kvartilbreddeBrynild = q3Brynild-q1Brynild
n=length(lengthsBrynild)
sampSdBrynild = sd(lengthsBrynild)
popSdBrynild = sqrt((n-1)/n)*sampSdBrynild
```

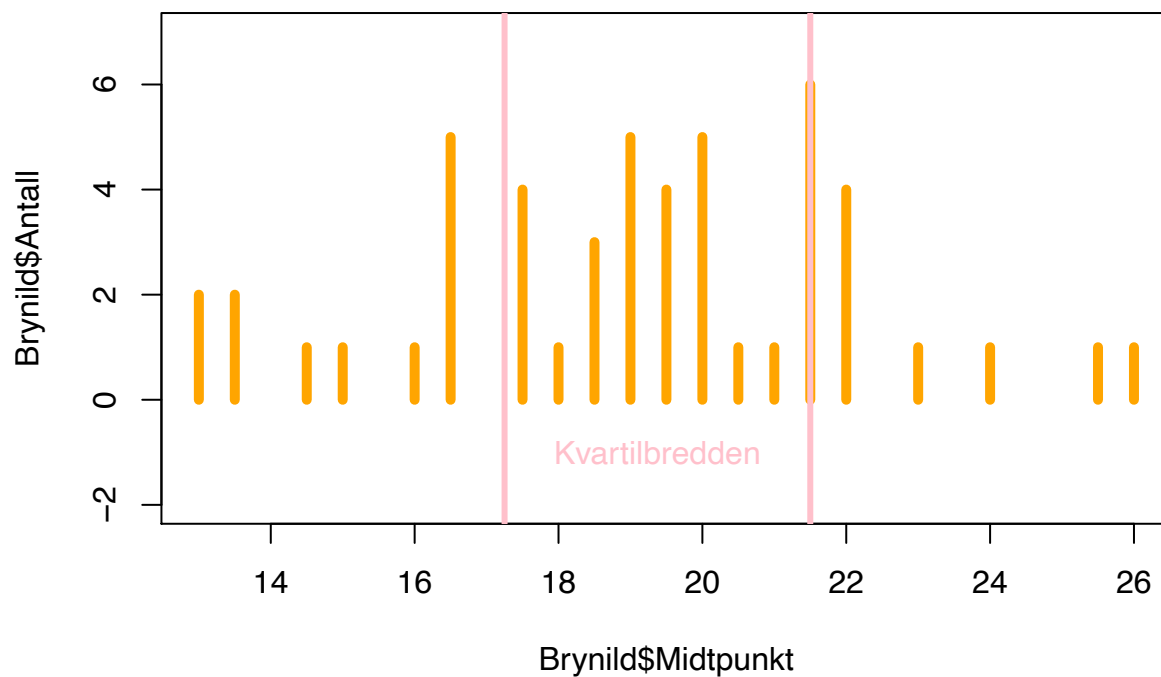
	Laban	Brynild
1. kvartil	12	17.25
3. kvartil	14.5	21.5
Kvartilbredde	2.5	4.25
Utvalgsstandardavvik	1.4229891	2.9875422
Populasjonsstandardavvik	1.4086873	2.9575158

3n

```
plot(Laban$Midtpunkt,Laban$Antall,ylim=c(-2,max(Laban$Antall+1)),type="h",lwd=4,col="purple")
abline(v = c(q1Laban,q3Laban), col = "pink", lwd = 3)
text((q1Laban+q3Laban)/2,-1,"Kvartilbredden",col="pink")
```



```
plot(Brynild$Midtpunkt,Brynild$Antall,ylim=c(-2,max(Brynild$Antall+1)),type="h",lwd=5,col="orange")
abline(v = c(q1Brynild,q3Brynild), col = "pink", lwd = 3)
text((q1Brynild+q3Brynild)/2,-1,"Kvartilbredden",col="pink")
```



30

Vi har gjennomgående gjort alt for begge posere underveis