Apply og aggregate

sepal.length	sepal.width	variety
5.1	3.5	Setosa
4.9	3	Setosa
4.7	3.2	Setosa
4.6	3.1	Setosa
5	3.6	Setosa
7	3.2	Versicolor
6.4	3.2	Versicolor
6.9	3.1	Versicolor
5.5	2.3	Versicolor
6.5	2.8	Versicolor
6.3	3.3	Virginica
5.8	2.7	Virginica
7.1	3	Virginica
6.3	2.9	Virginica
6.5	3	Virginica
7.6	3	Virginica
4.9	2.5	Virginica

Aggregate sum

Aggregate sum

variety	sepal.length	sepal.width
Setosa	24.3	16.4
Versicolor	32.3	14.6
Virginica	44.5	20.4

Function	Input data type	Output data type
apply	dataframe or matrix or array (with margins)	vector, matrix, array, list
lapply	vector, list, variables in dataframe or matrix	list
sapply	vector, list, variables in dataframe or matrix	matrix, vector, list
mapply (multivariate sapply)	vector, list, variables in dataframe or matrix	matrix, vector, list
tapply	ragged array	array
rapply	vector, list, variables in	list

Aggregate

Syntax for Aggregate() Function in R:

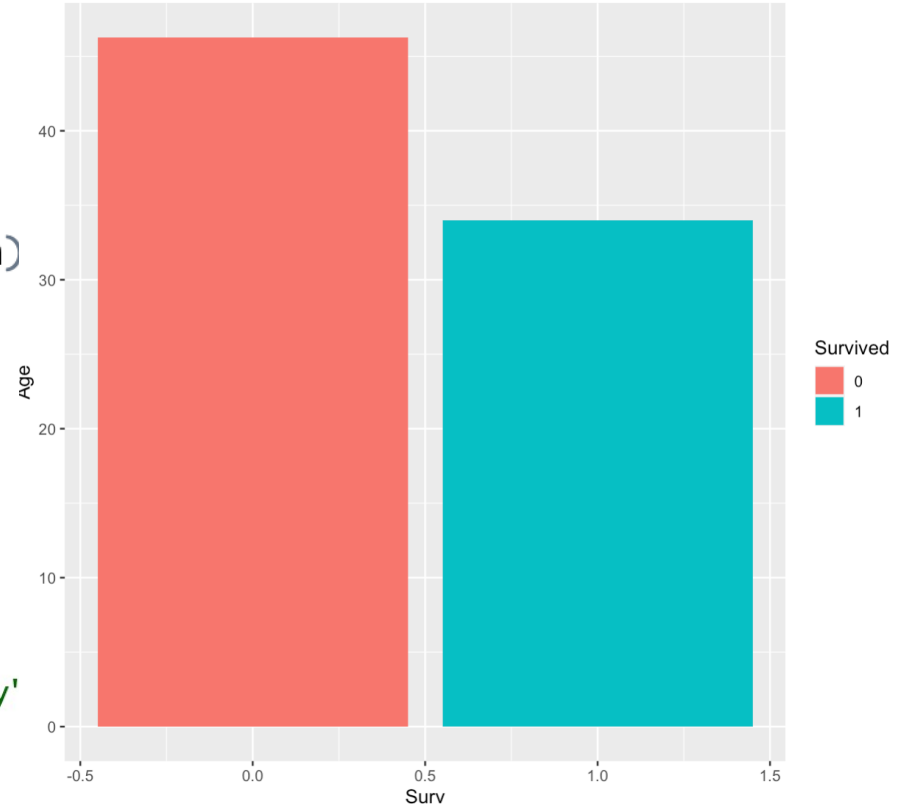
aggregate(x, by, FUN, ..., simplify = TRUE, drop = TRUE)	
X	an R object, Mostly a dataframe
by	a list of grouping elements, by which the subsets are grouped by
FUN	a function to compute the summary statistics
simplify	a logical indicating whether results should be simplified to a vector or matrix if possible
drop	a logical indicating whether to drop unused combinations of grouping values.

```
4  # gennemsnitsalder for de overlevende og døde
5  dd=aggregate(estonia['Age'], list(Surv=estonia$Survived), mean)
```

Surv	Age
0	46.27347
1	34.01460

og plot ..

```
9  ggplot(dd, aes(x=Surv,y=Age, fill=as.factor(Surv)))+geom_bar(stat="identity"
10   labs(fill="Survived")
```



Aggregate

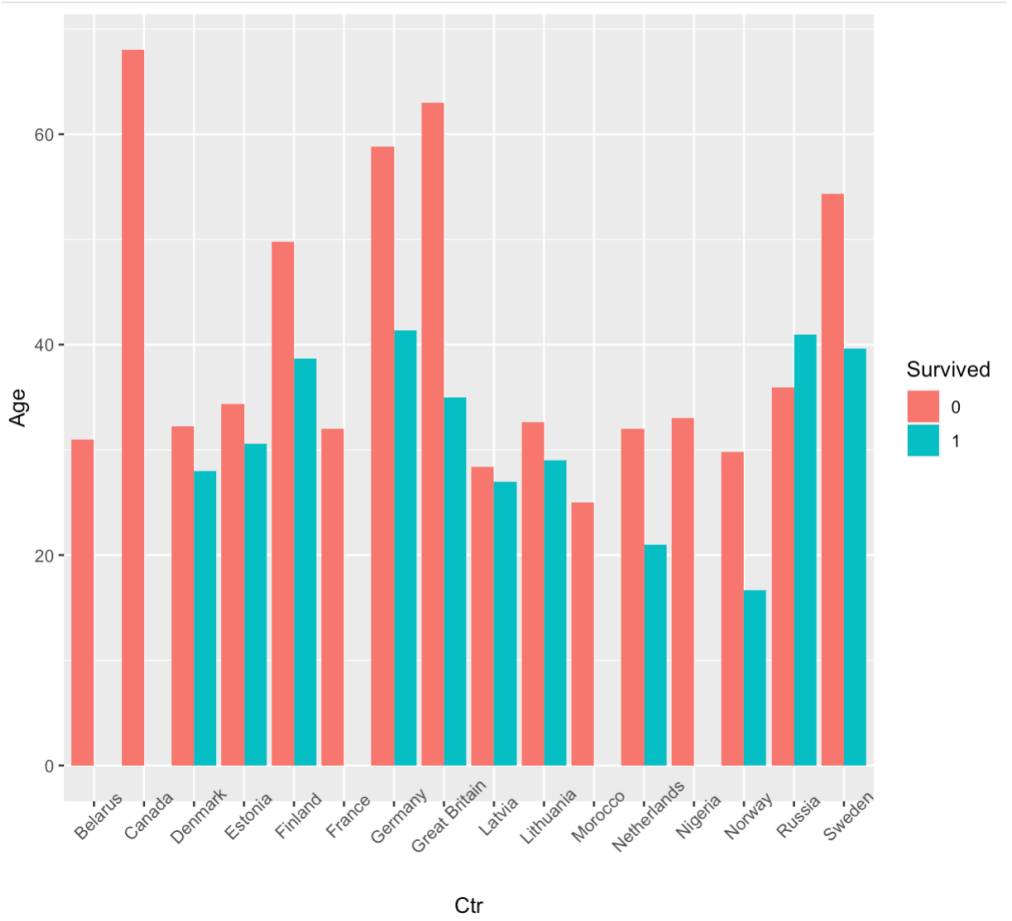
```r
ggplot(sctr, aes(x=Ctr,y=Age, fill=as.factor(Surv)))+
  geom_bar(stat="identity", position="dodge")+
  labs(fill="Survived")+
  theme(axis.text.x = element_text(angle=45))
```

```r
# gennemsnitsalder for overlevelse fordelt på land
sctr <- aggregate(estonia['Age'],
                  list(Surv=estonia$Survived, Ctr=estonia$Country),
                  mean,
                  drop = F
                  )
```

Surv	Ctr	Age
0	Belarus	31
1	Belarus	*NA*
0	Canada	68
1	Canada	*NA*
0	Denmark	32
1	Denmark	28
0	Estonia	34
1	Estonia	30
0	Finland	49
1	Finland	38

Øvelse:
tæl antal overlevevende fordelt på land og køn

og plot …



Apply

Function	Input data type	Output data type
apply	dataframe or matrix or array (with margins)	vector, matrix, array, list
lapply	vector, list, variables in dataframe or matrix	list
sapply	vector, list, variables in dataframe or matrix	matrix, vector, list
mapply (multivariate sapply)	vector, list, variables in dataframe or matrix	matrix, vector, list

lapply

Datatransformation anvendt på hvert element i listen eller vektoren **uden** loops

```
# lav en funktion der deler personer op i barn,ung,voksen,ældre,gammel
obesity$AgeCat <- lapply(obesity$Age, FUN = get_age_cat)
```

10	Male	22	1.72	68.0	adult
11	Male	26	1.85	105.0	adult
12	Female	21	1.72	80.0	adult
13	Male	22	1.65	56.0	adult
14	Male	41	1.80	99.0	senior
15	Male	23	1.77	60.0	adult
16	Female	22	1.70	66.0	adult
17	Male	27	1.93	102.0	adult
18	Female	29	1.53	78.0	adult

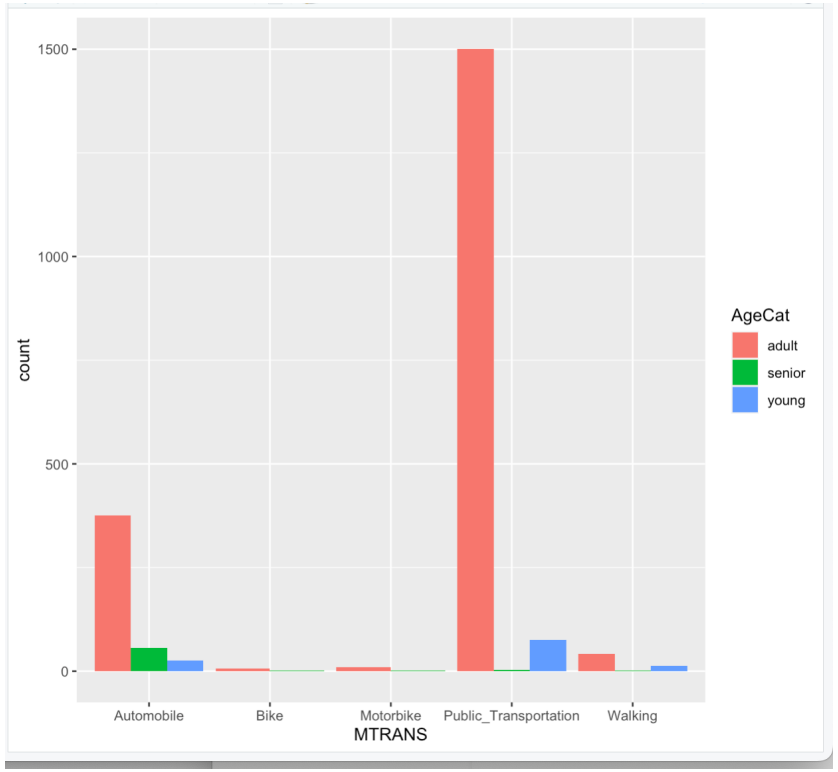
Et summary:

```
adult senior  young
 1935     63    113
```

Øvelse:
Skriv funktionen get_age_cat

lapply

```
34 # plot nu antal observationer per alderskategori aggregeret på transport-form
35 obesity$count=1
36 dftrans <- aggregate(count ~ MTRANS + AgeCat, data=obesity, FUN = sum)
```



Øvelse:
Gør det samme for vægt-kategorien, altså aggregér og plot

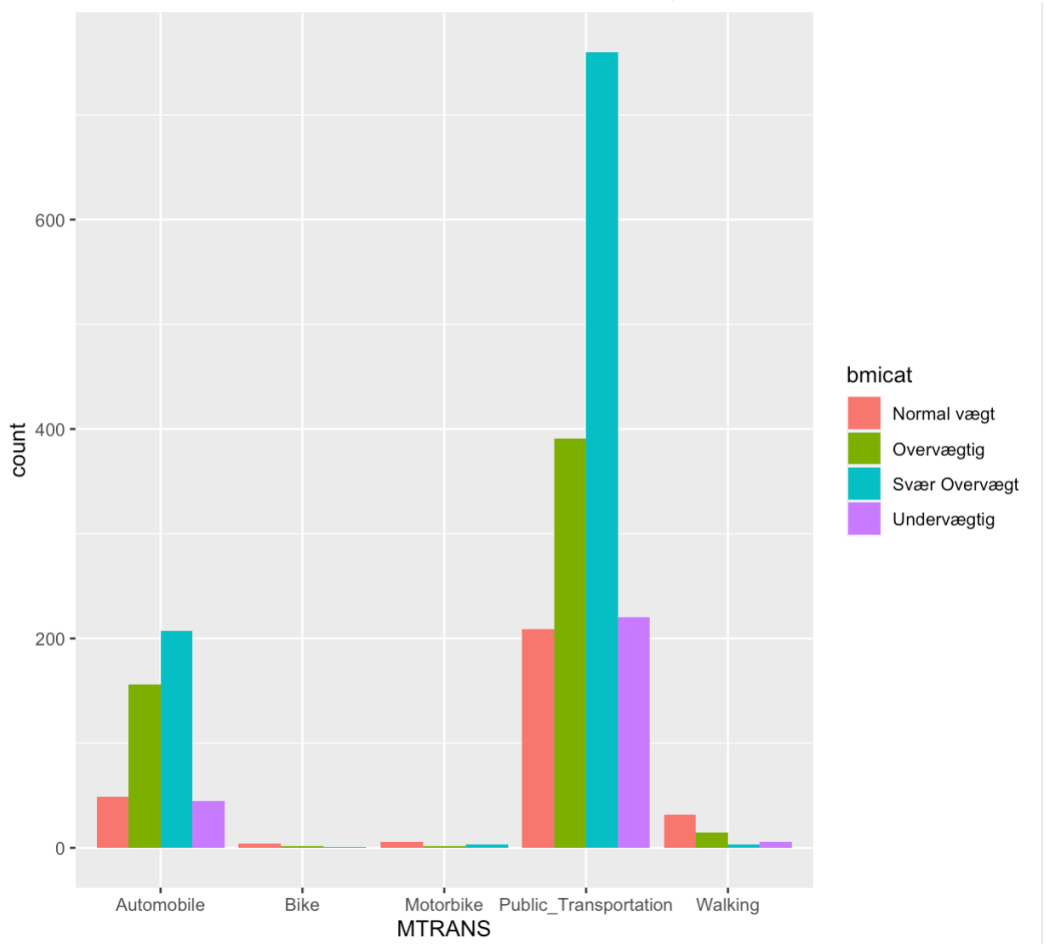
mapply

Datatransformation anvendt på hvert element i listen eller vektoren **uden** loops

```
73 #lav en funktion der deler vægt op i undervægtig, normal, overvægt og svær overvægt
74 bmicat=mapply(FUN=get_weight_cat, w=obesity$Weight, h=obesity$Height)
75 obesity2 = cbind(obesity,bmicat)
76
```

Øvelse:

- Skriv funktionen get_weight_cat
- Aggregér og plot



mapply

Datatransformation anvendt på hvert element i listen eller vektoren **uden** loops

```
73 #lav en funktion der deler vægt op i undervægtig, normal, overvægt og svær overvægt
74 bmicat=mapply(FUN=get_weight_cat, w=obesity$Weight, h=obesity$Height)
75 obesity2 = cbind(obesity,bmicat)
76
```

Øvelse:

- Skriv funktionen get_weight_cat
- Aggregér og plot