

Oppgave 1)

```
require(ggplot2)
```

```
require(stringr)
```

```
theFile = "C:/.../dishes.csv"
```

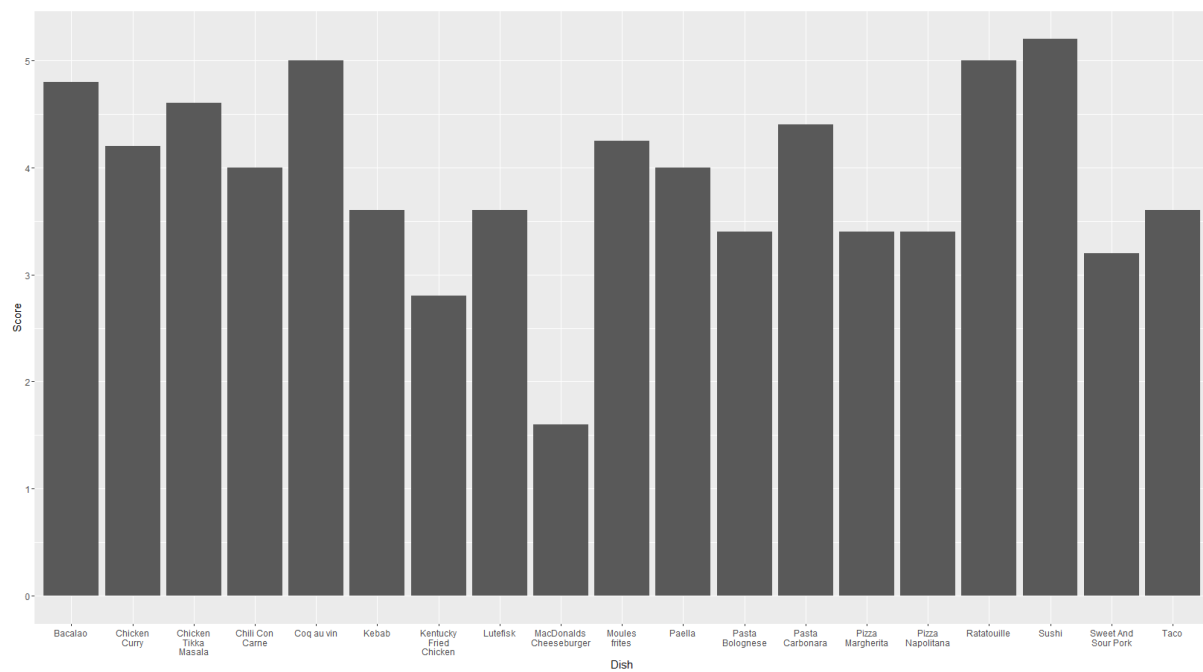
```
table <- read.table(file=theFile,header=TRUE,sep = ";")
```

```
avg = aggregate(Score ~ Dish, data = dishes, FUN = mean)
```

```
avg = subset(avg,select = c(Dish,Score))
```

```
avg$Dish = str_wrap(avg$Dish, width = 10)
```

```
ggplot(avg,aes(x=Dish, y=Score)) + geom_bar(stat = "identity")
```



Oppgave 2)

```
theFile = "C:/.../dishes.csv"
```

```
dishes <- read.table(file=theFile,header=TRUE,sep = ";")
```

```
require(reshape2)
```

```
dish_re = dcast(dishes, UserName ~ Dish)
```

```
users <- dish_re[,1]
```

```
dish_re <- dish_re[,-1]
```

```
d <- dist(dish_re, method = "euclidian")
```

```
d <- as.matrix(d)
```

```
fit <- cmdscale(d,eig=TRUE,k=2)
```

```
x <- fit$points[,1]
```

```
y <- fit$points[,2]
```

```
plot(x,y,xlab = "x",ylab = "y",main="UserName",type="n")
```

```
text(x,y,labels=users,cex=1.5)
```

```
dinnerparty <- function(me,d,k,users){
```

```
  ordered.neighbors <- order(d[me, ])
```

```
  ordered.neighbors <- ordered.neighbors[ordered.neighbors != me]
```

```
  nearest.neighbors <- ordered.neighbors[0:k]
```

```
  result <- vector(mode="character",length=k)
```

```
  count <- 1
```

```
  for(i in nearest.neighbors){
```

```
    result[count] <- as.character(users[i])
```

```
    count <- count + 1
```

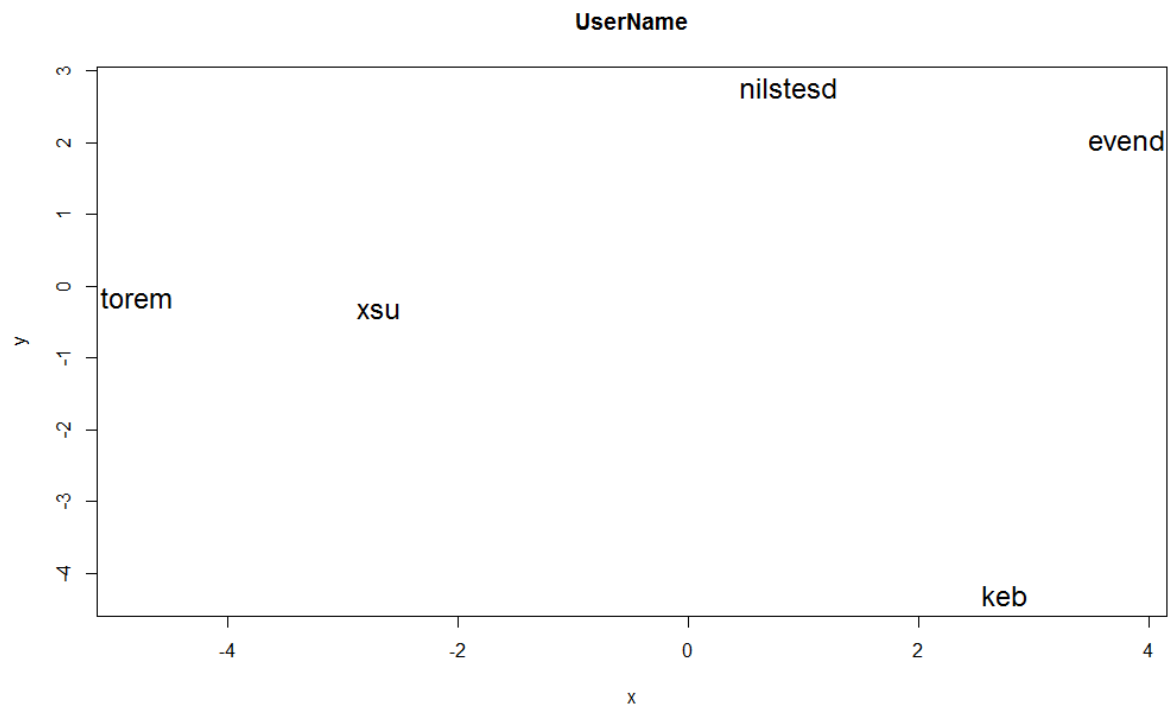
```
  }
```

```
  return(result)
```

```
}
```

```
dinnerparty(1,d,2,users = users)
```

```
# output: "nilsted" "keb"
```



Oppgave 3)

```
theFile = "C/.../winequality-red.csv"
```

```
wines = read.table(file=theFile,header=TRUE,sep=";")
```

```
require(rpart.plot)
```

```
fit <- rpart(quality ~ fixed.acidity + volatile.acidity + residual.sugar + citric.acid + chlorides +  
free.sulfur.dioxide + total.sulfur.dioxide + density + pH + sulphates + alcohol,  
method="anova", data = wines)
```

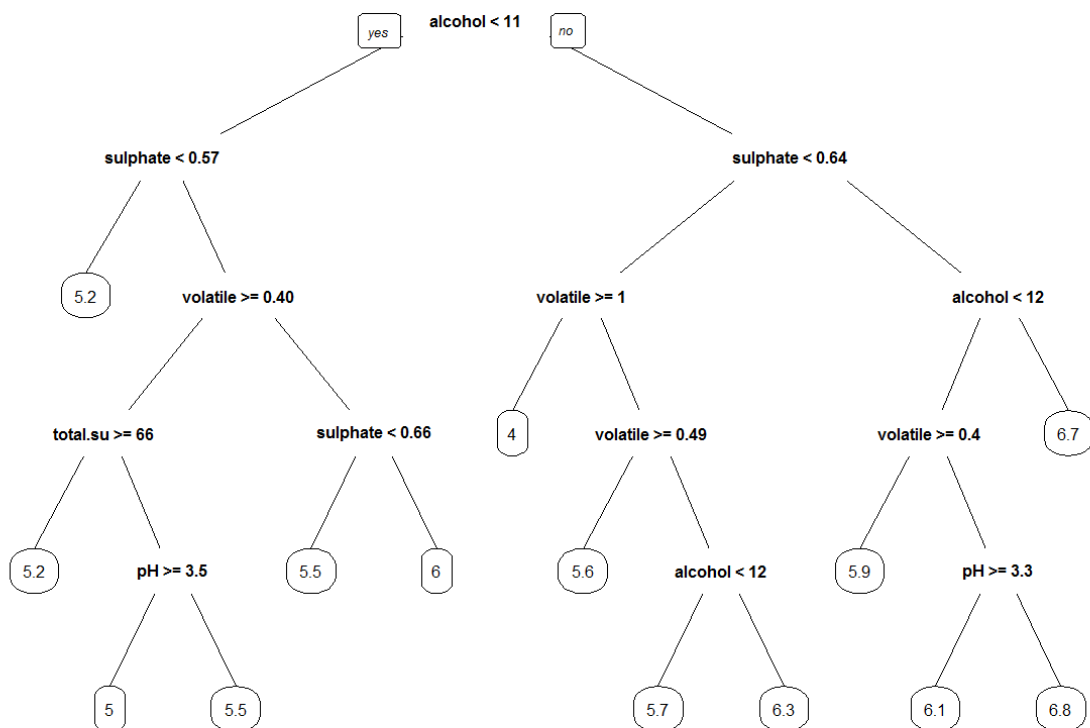
```
printcp(fit)
```

```
#den beste xerror fant jeg ved cp=0.0063626 (fra utskriften til printcp), den hadde
```

```
#xerror=0.677
```

```
fit <- rpart(quality ~ fixed.acidity + volatile.acidity + residual.sugar + citric.acid + chlorides +  
free.sulfur.dioxide + total.sulfur.dioxide + density + pH + sulphates + alcohol, method="anov  
a", data = wines, control = rpart.control(cp=0.0063626))
```

```
prp(fit)
```



```
#lagde også et tre med veldig stor dybde, slik kunne jeg se hvordan xerror utviklet seg
```

Variables actually used in tree construction:

alcohol chlorides citric.acid density fixed.acidity
 free.sulfur.dioxide pH residual.sugar sulphates total.sulfur.dioxide
 volatile.acidity

Root node error: $1042.2/1599 = 0.65176$

n= 1599

	CP	nspl	rel error	xerror	xstd
1	0.1782206	0	1.00000	1.00137	0.037880
2	0.0535887	1	0.82178	0.84275	0.036575
3	0.0297433	2	0.76819	0.80323	0.033488
4	0.0288858	3	0.73845	0.78978	0.032638
5	0.0223428	4	0.70956	0.76253	0.031769
6	0.0192724	5	0.68722	0.74079	0.029448
7	0.0151135	6	0.66795	0.72655	0.029359
8	0.0101591	7	0.65283	0.71896	0.029249
9	0.0091604	9	0.63251	0.69848	0.028765
10	0.0086504	10	0.6233	0.69452	0.028678
11	0.0075264	11	0.6147	0.69544	0.028554
12	0.0074710	12	0.6071	0.68238	0.028418
13	0.0063626	13	0.5997	0.67681	0.027892
14	0.0058783	14	0.5933	0.67990	0.028084
15	0.0055681	17	0.5757	0.68362	0.028235
16	0.0055511	18	0.5701	0.68632	0.028368
17	0.0055502	20	0.5590	0.68632	0.028368
18	0.0049056	22	0.5479	0.68501	0.028251
19	0.0045537	23	0.5430	0.68954	0.029025
20	0.0044724	25	0.5339	0.69436	0.029281
21	0.0040936	26	0.5294	0.68885	0.029310
22	0.0040564	29	0.5171	0.69126	0.029124
23	0.0036905	30	0.5131	0.68942	0.029255
24	0.0034070	31	0.5094	0.69825	0.029886
25	0.0033007	33	0.5026	0.69867	0.029894
26	0.0032899	35	0.4960	0.69695	0.029900
27	0.0032664	36	0.4927	0.69670	0.029903
28	0.0032617	37	0.4894	0.69646	0.029899
29	0.0032333	39	0.4829	0.69761	0.029872
30	0.0032263	40	0.4797	0.69823	0.029775
31	0.0032072	41	0.4764	0.69821	0.029785
32	0.0031249	42	0.4732	0.69737	0.029783
33	0.0030868	45	0.4638	0.70058	0.029923
34	0.0029591	46	0.4608	0.70368	0.029921
35	0.0029243	48	0.4548	0.70207	0.029843
36	0.0028975	50	0.4490	0.70282	0.029815
37	0.0027231	51	0.4461	0.70650	0.029918
38	0.0026604	52	0.4434	0.70859	0.029956
39	0.0026105	53	0.4407	0.70711	0.029883
40	0.0024915	55	0.4355	0.70946	0.030318

Klassifisering av vinene kan gjøres ved å f.eks si at en vin av god kvalitet har en score på mer enn 7, en vin av middels kvalitet mellom 3-7 og en av dårlig kvalitet under 7. Deretter kategoriserer man vinen etter disse standardene. Regresjon gjøres slik jeg har vist i oppgaven, der man gir vinen en eksakt kvalitetsverdi.