DSC 105 Grades Analysis

1. Download data into data frame `ds1`

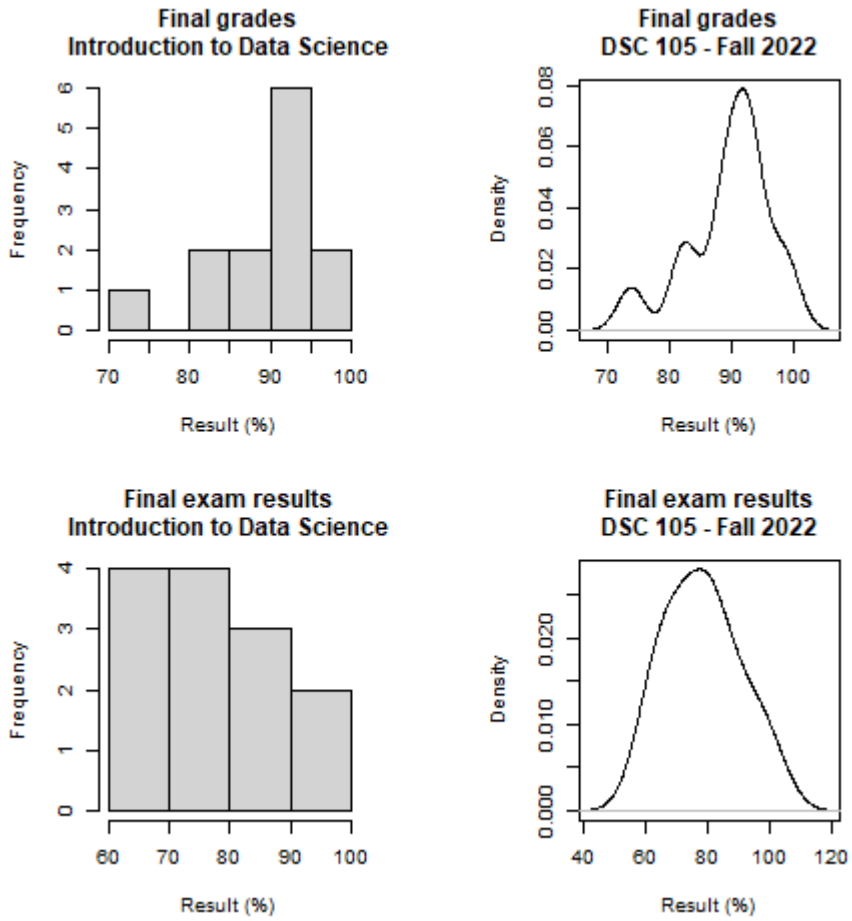
```R
ds1 <- read.csv(
  file="../data/ds1_grades.csv",
  header=TRUE)
str(ds1)
```

2. Final grades `final_ds1` and final exam results `exam_ds1`

```
final_ds1 <- (as.numeric(ds1$Final.Score[3:nrow(ds1)-1]))
final_ds1
exam_ds1 <- as.numeric(ds1$Final.exam.Final.Score[3:nrow(ds1)-1])
exam_ds1
```

3. Plot final grades and final exam results:

```
par(mfrow=c(2,2), pty='s')
hist(final_ds1,
     xlab="Result (%)",
     main="Final grades\nIntroduction to Data Science")
plot(density(final_ds1),
     xlab="Result (%)",
     main="Final grades\nDSC 105 - Fall 2022")
hist(exam_ds1,
     xlab="Result (%)",
     main="Final exam results\nIntroduction to Data Science")
plot(density(exam_ds1),
     xlab="Result (%)",
     main="Final exam results\nDSC 105 - Fall 2022")
```



4. Final projects - grade table and stats summary

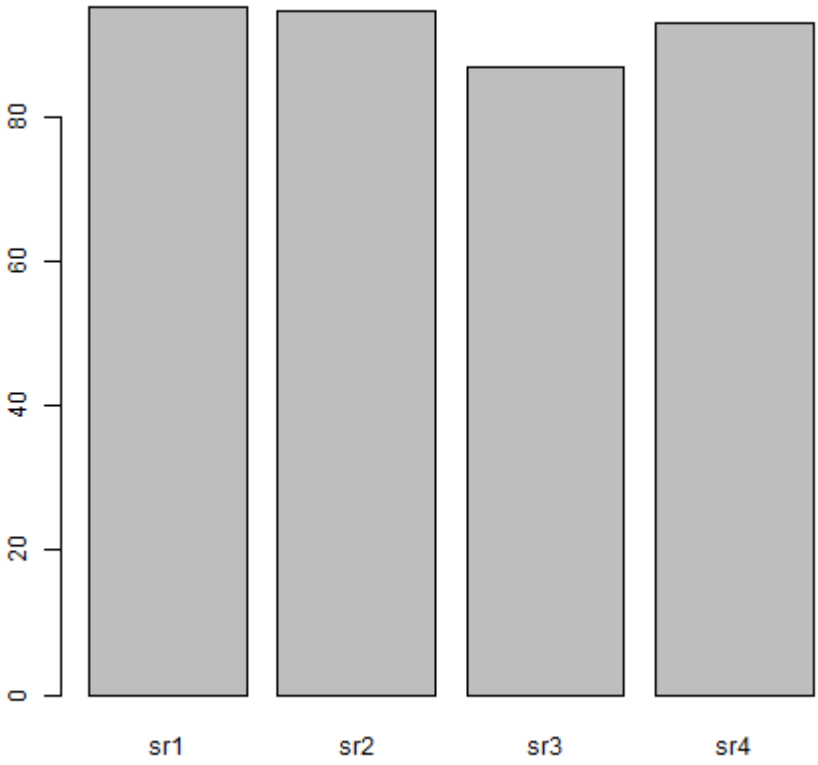
```
proj <- data.frame(
  "sr1" = 100*ds1[3:nrow(ds1)-1,grep("X1st", colnames(ds1))]/20,
  "sr2" = 100*ds1[3:nrow(ds1)-1,grep("X2nd", colnames(ds1))]/20,
  "sr3" = 100*ds1[3:nrow(ds1)-1,grep("X3rd", colnames(ds1))]/20,
  "sr4" = 100*ds1[3:nrow(ds1)-1,grep("Fourth", colnames(ds1))]/40)
proj
summary(proj)
```

5. Barplots - sprint review averages

```
avg_proj_ds1 <- sapply(X=proj,FUN=mean)
str(avg_proj_ds1)
```

Barplot of project averages by sprint review:

```
barplot(avg_proj_ds1)
```

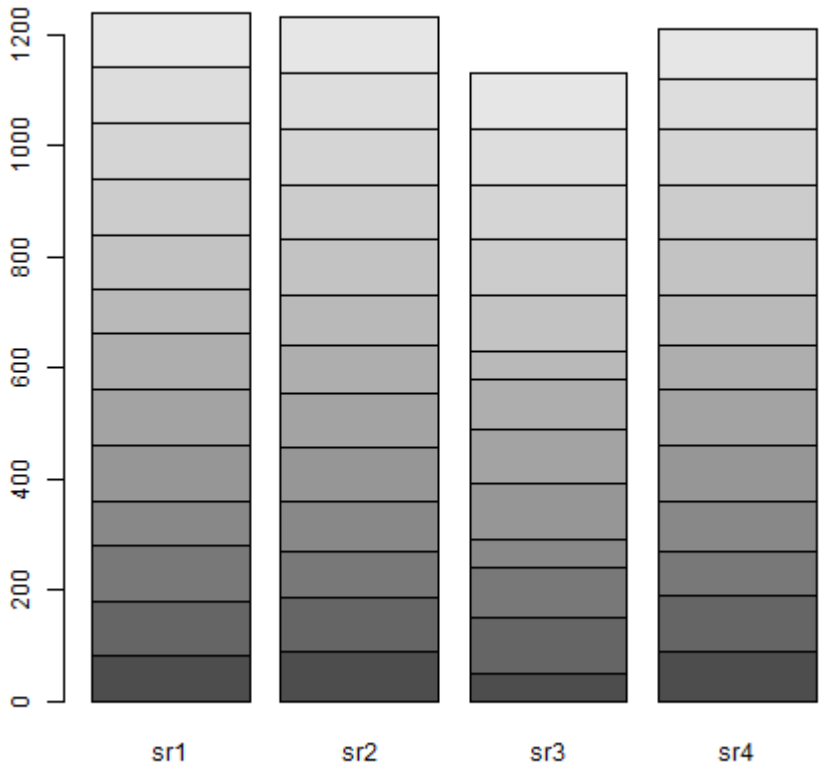


With the individual member results (stacked barplot):

```
as.matrix(proj)
```

Barplot of project averages by team member and sprint review:

```
barplot(as.matrix(proj))
```



But what we want is grouping the results by team. We have:

```
##ds1$Student[3:nrow(ds1)-1]  # all students
soccer_idx <- sort(c(grep(pattern="Pedro",x=ds1$Student,fixed=TRUE)-1,
                     grep(pattern="Andrei",x=ds1$Student,fixed=TRUE)-1,
                     grep(pattern="Fortu",x=ds1$Student,fixed=TRUE)-1))
soccer_idx
rocket_idx <- sort(c(grep("Brittany",ds1$Student)-1,
                     grep("Jordan",ds1$Student)-1))
rocket_idx
baseball_idx <- sort(c(grep("Nick",ds1$Student)-1,
                       grep("Hayden",ds1$Student)-1))
baseball_idx
sticky_idx <- c(grep("Harrod",ds1$Student)-1)
sticky_idx
ceramics_idx <- sort(c(grep("Kim",ds1$Student)-1,
                       grep("Balah",ds1$Student)-1,
                       grep("David",ds1$Student)-1))
ceramics_idx
titanic_idx <- sort(c(grep("Jacob",ds1$Student)-1,
                      grep("Brandon",ds1$Student)-1))
titanic_idx
```

Unname project data:

```
proj_unnamed <- unname(proj)
proj_unnamed
```

Extract team data (by index):

```
ds1_proj_team <- proj_unnamed[soccer_idx[1],]
ds1_proj_team[2,] <- proj_unnamed[rocket_idx[1],]
ds1_proj_team[3,] <- proj_unnamed[baseball_idx[1],]
ds1_proj_team[4,] <- proj_unnamed[sticky_idx[1],]
ds1_proj_team[5,] <- proj_unnamed[ceramics_idx[1],]
ds1_proj_team[6,] <- proj_unnamed[titanic_idx[1],]
ds1_proj_team
```

Turns out unnaming is not necessary:

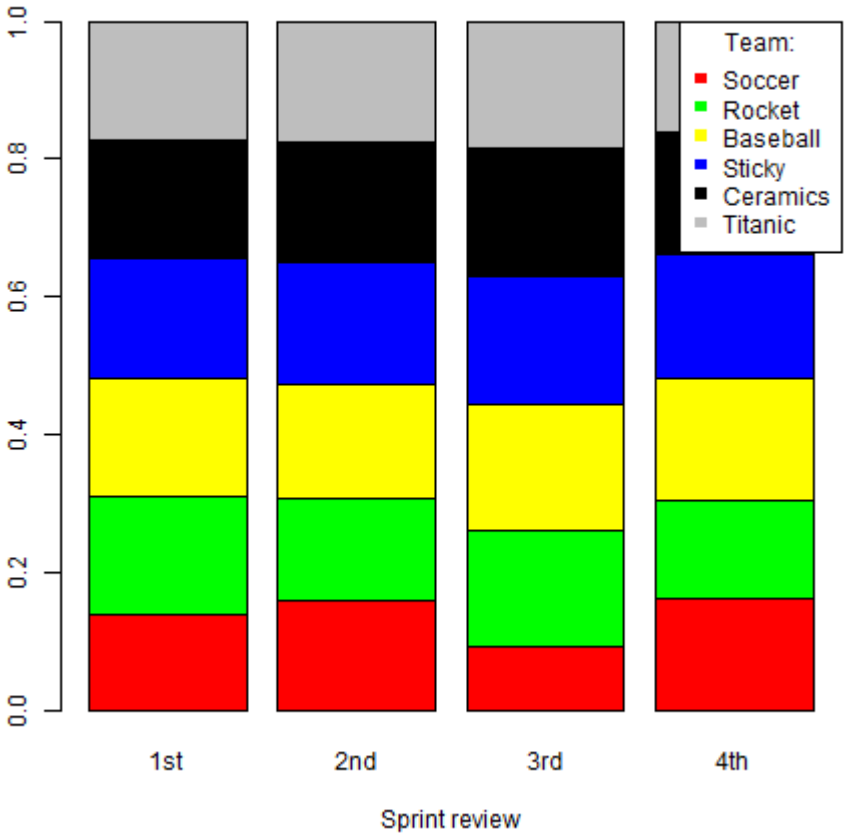
```
ds1_proj_team <- proj[soccer_idx[1],]
ds1_proj_team[2,] <- proj[rocket_idx[1],]
ds1_proj_team[3,] <- proj[baseball_idx[1],]
ds1_proj_team[4,] <- proj[sticky_idx[1],]
ds1_proj_team[5,] <- proj[ceramics_idx[1],]
ds1_proj_team[6,] <- proj[titanic_idx[1],]
rownames(ds1_proj_team) <- c("Soccer",
                             "Rocket",
                             "Baseball",
                             "Sticky",
                             "Ceramics",
                             "Titanic")
ds1_proj_team
```

```
p <- prop.table(ds1_proj_team)
p
```

```
ds1_teams <- apply(X=ds1_proj_team,
                   MARGIN=2,
                   FUN=function(x){
                     x/sum(x,na.rm=TRUE)})
sum(ds1_teams[,1])
ds1_teams
class(ds1_teams)
```

6. Barplot of project averages by team and sprint review:

```
colors <- c("red","green","yellow","blue","black","gray")
barplot(ds1_teams,
        col=colors,
        xlab="Sprint review",
        names.arg=c("1st","2nd","3rd","4th"))
legend("topright",
       legend=rownames(ds1_proj_team),
       pch=15,
       title="Team:",
       col=colors)
```



7. Final grades for report

```
data.frame(
  "Student"=ds1$Student[3:nrow(ds1)-1],
  "Grade"=final_ds1)
mean(final_ds1)
```

Created: 2022-12-13 Tue 11:14