GDA of *course.df* (from **s20x**)

Background Data on 146 students taking a Stats summer school

Aims Which students do well? Which factors affect student performance? Are there differences between the genders?

Source Auckland Statistics Department

Structure 146 observations on 15 variables (7 numeric and 8 factors)

```
library(gridExtra)
data(course.df, package="s20x")
a1 <- ggplot(course.df, aes(Degree)) + geom_bar()
a2 <- ggplot(course.df, aes(Gender)) + geom_bar()
a3 <- ggplot(course.df, aes(Attend)) + geom_bar()
grid.arrange(a1,a2,a3, nrow=1, widths=c(3,2,2))
```

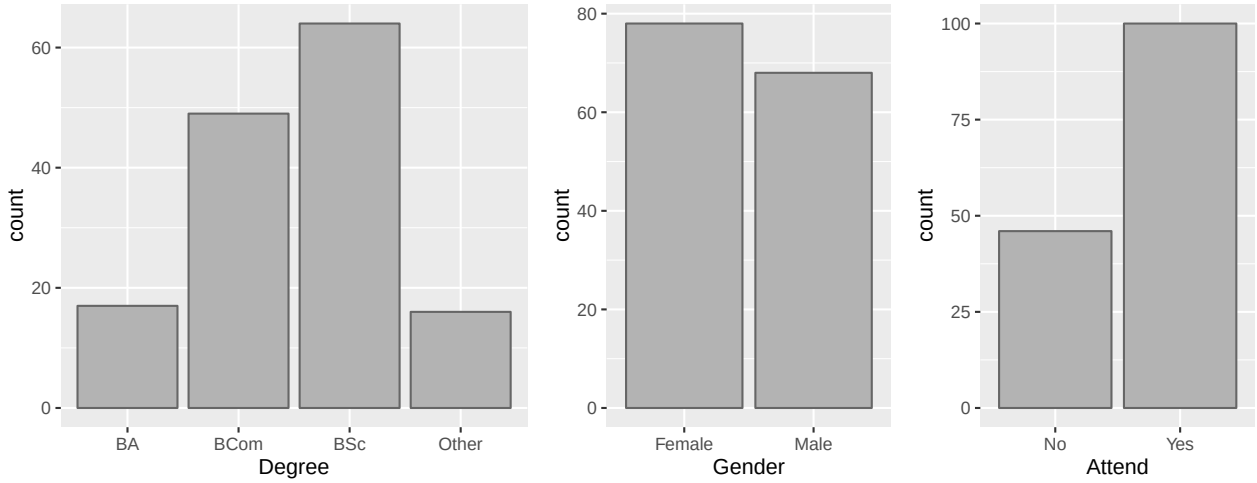

Fig 1: The three barcharts show that students mostly had a BSc or BCom, that there were slightly more females than males, and that just under a third did not regularly attend the course.

```
ggplot(course.df, aes(Degree)) + geom_bar() + facet_grid(.~Gender)
```

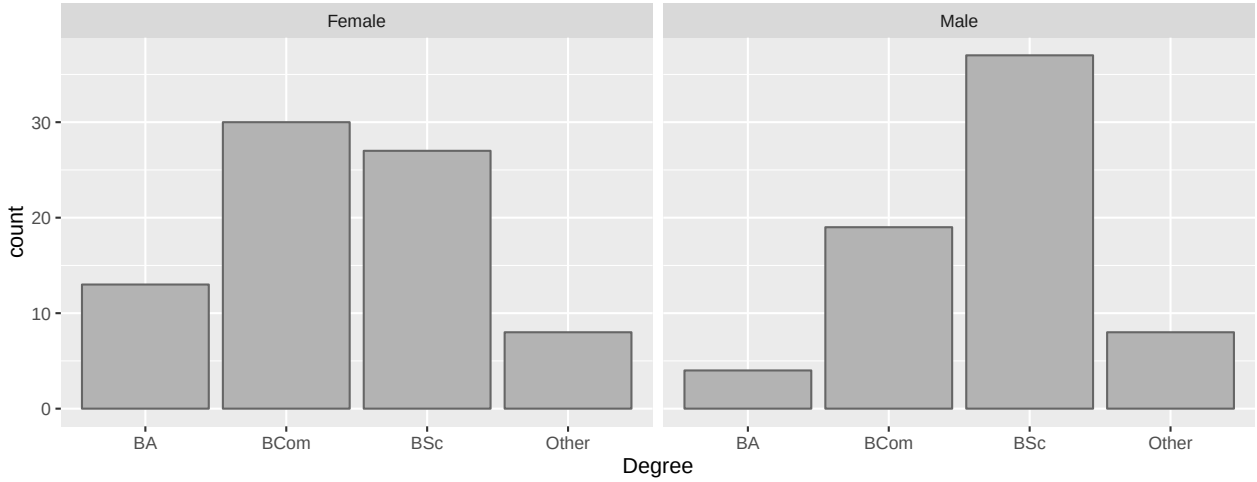

Fig 2: A majority of the males had BSc's, while quite a few females had BCom degrees.

```
ggplot(course.df, aes(Degree, Exam)) + geom_boxplot() + facet_grid(.~Gender)
```

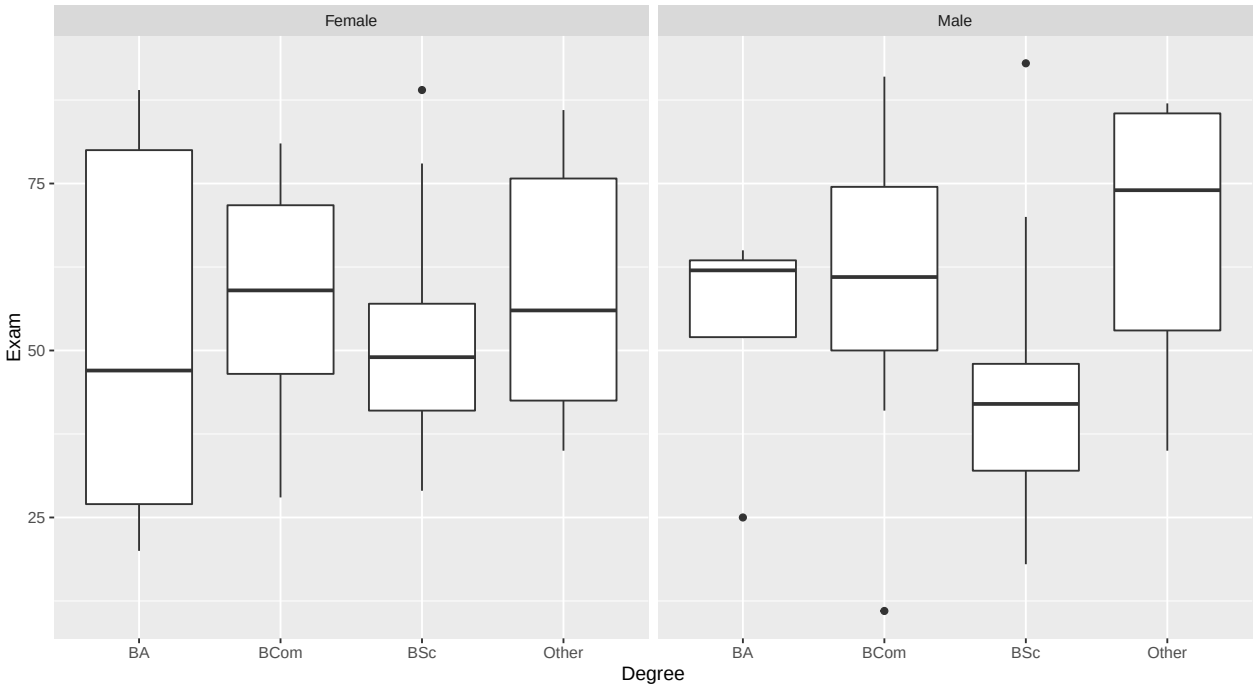

Fig 3: The group of male BSc students did particularly badly in the exam (although the top mark was gained by one of them).

```
ggplot(course.df, aes(Degree)) + geom_bar() + facet_grid(Attend~Gender)
```

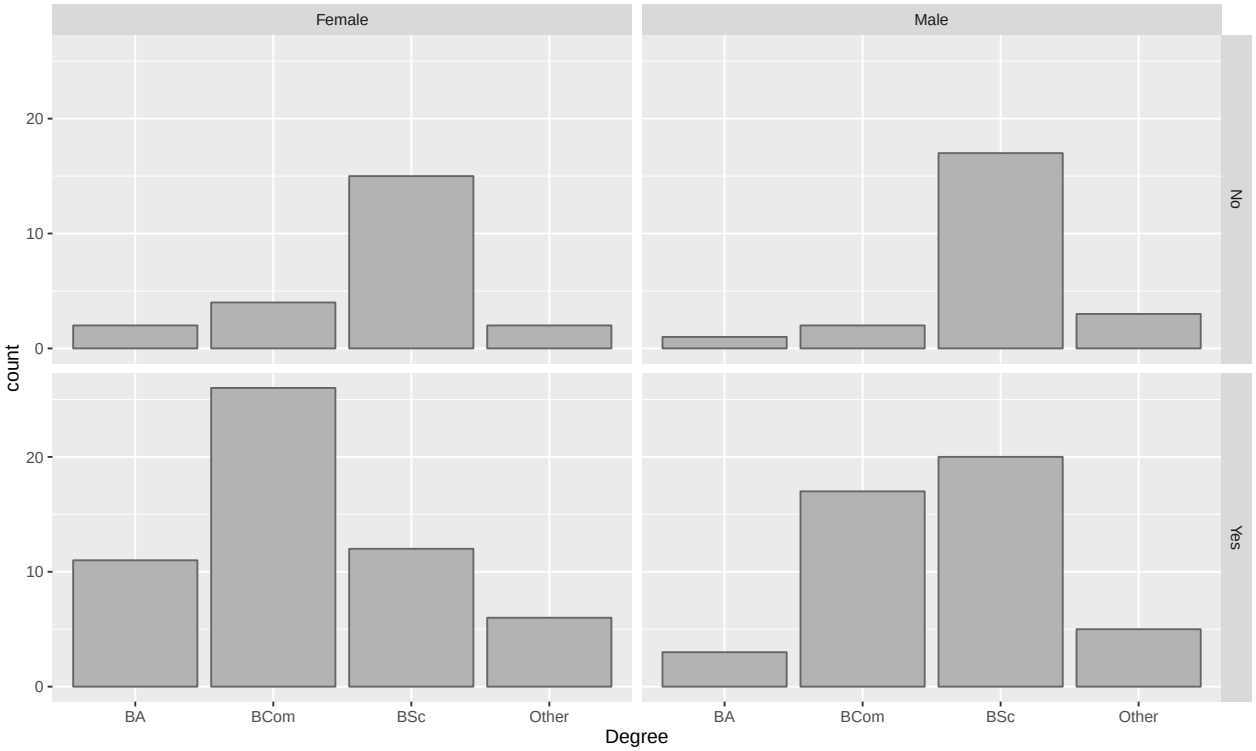

Fig 4: Regular attendance was poor amongst the BSc degree students for both males and females.

```
library(dplyr)
library(vcd)
c2 <- course.df %>% mutate(Rep1=factor(Repeat, levels=c("Yes","No")))
doubledecker(Pass~Attend+Rep1, data=c2, gp = gpar(fill = c("grey90", "red")))
```

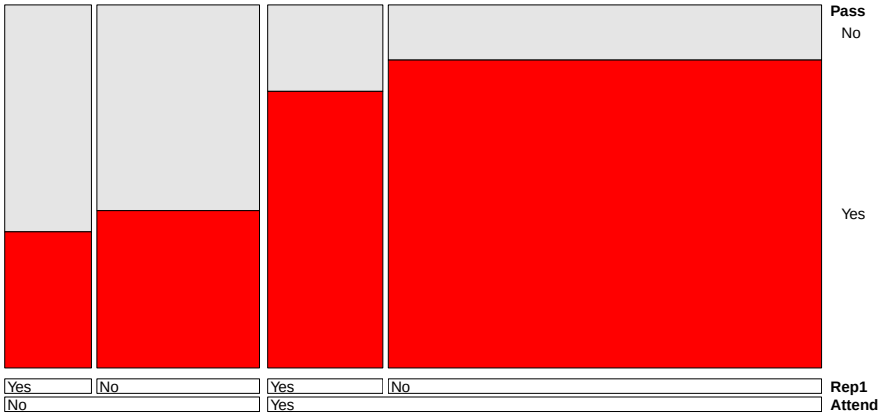

Fig 5: Pass rates were best for students who attended and were not repeating.

```
ggplot(course.df, aes(Assign, Exam)) + geom_point() + ylim(0,100) + geom_smooth()
```

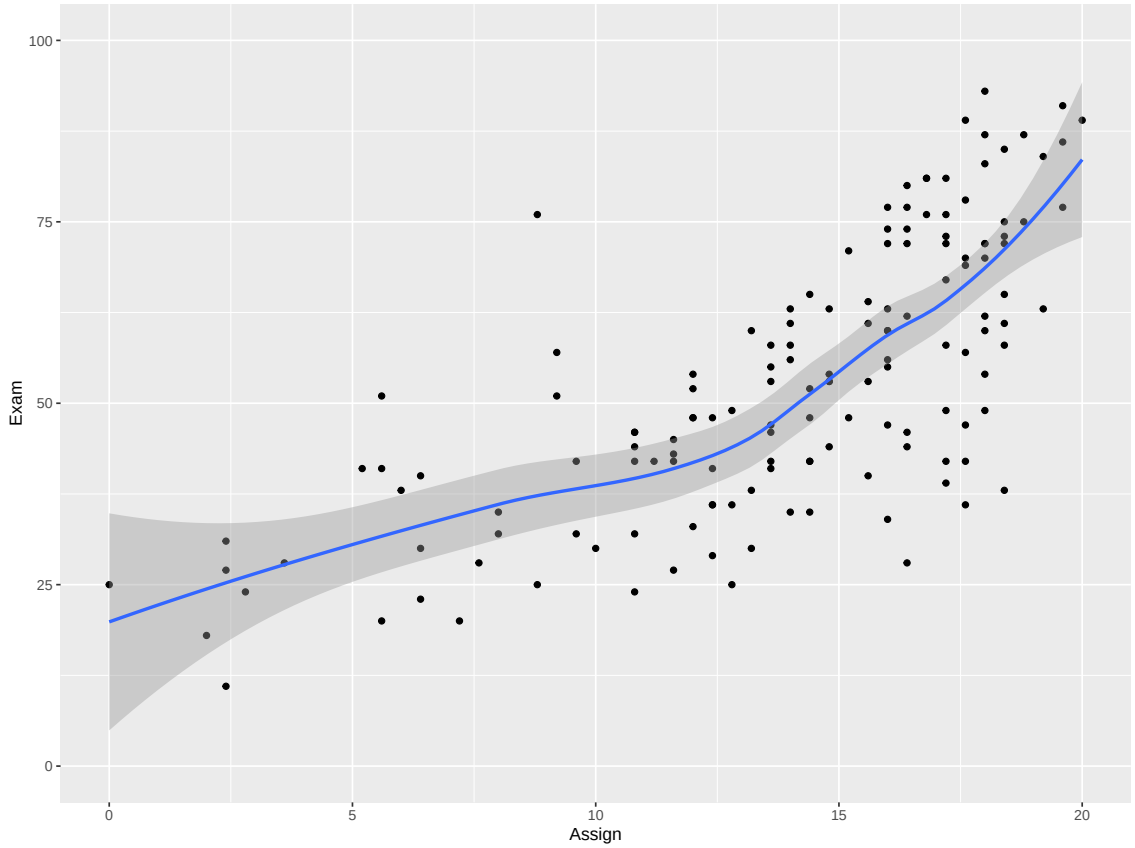

Fig 6: Poor assignment marks imply poor exam marks (apart from one notable exception with an exam mark over 75). For assignment marks over 12 there is a roughly linear relationship between exam marks and assignment marks.

```
library(GGally)
ggparcoord(course.df, columns=c(3,7:11), groupColumn="Grade", scale="globalminmax") +
  xlab("") + ylab("")
```

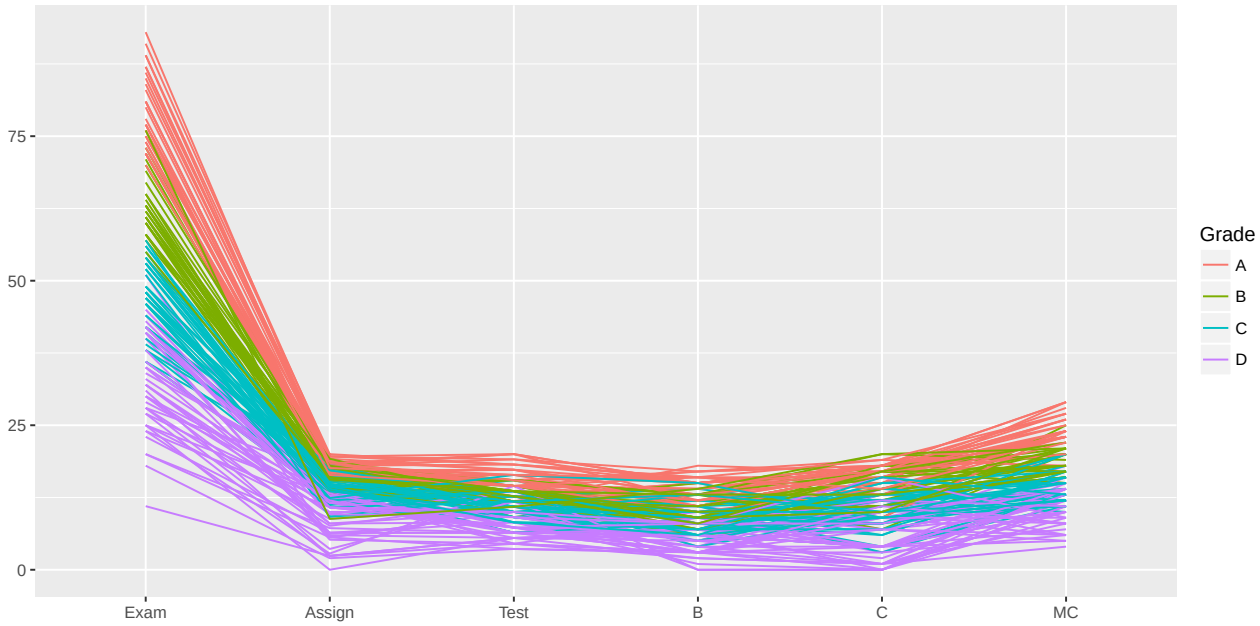

Fig 7: Final grades (A was the best, D was a fail) depended mainly on the exam mark. There was some variation in the exam components (B, C, and MC) and in particular the best mark in section C was achieved by someone with a B grade.

```
ggparcoord(course.df, columns=c(3,7:11), groupColumn="Grade", scale="uniminmax") +
  xlab("") + ylab("")
```

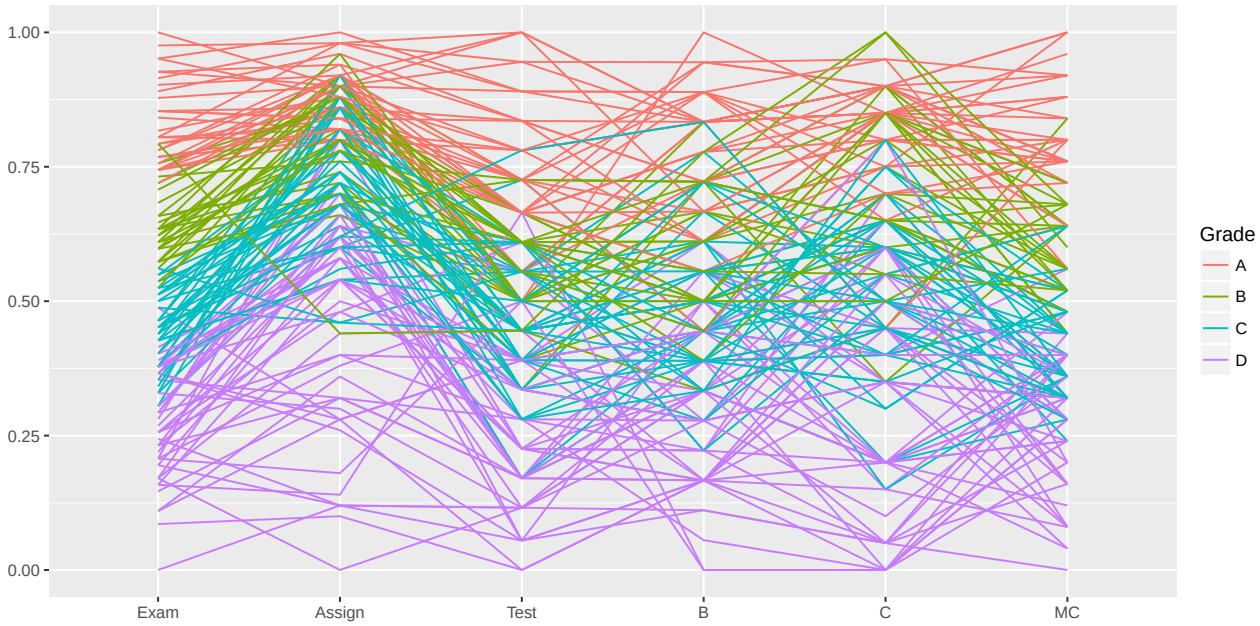

Fig 8: The same display as in Fig 7 but with a unified scaling to show the variablity between the results for the non-exam marks.