NZSSN Courses: Introduction to R

Session 6 – Advanced Graphics

Statistical Consulting Centre

consulting@stat.auckland.ac.nz
The Department of Statistics
The University of Auckland

15 November, 2016

SCIENCE
DEPARTMENT OF STATISTICS

Plot means in context

```
with(issp.df, tapply(total.lik, age.group,
                     mean, na.rm = TRUE))

Under 35 36 to 60  Over 61
13.38871 12.45516 10.78836
```

- Means are all but meaningless unless they are presented in context.
- Always present with standard deviations (SDs) or standard errors of means (SEs) or confidence intervals.
- Plot means with 95% confidence intervals ($\pm$ 1.96 $\times$ SE).
  - $\pm$ 1 $\times$ SE yields (approx.) a 68% confidence interval. Equivalent to using a 16% level of significance!!!!
  - $\pm$ 1 $\times$ SD tells us **ABSOLUTELY NOTHING** about whether two means are statistically different from one another.

Calculating 95% CIs

- 95% CI = Mean $\pm$ 1.96 $\times$ SE
- Standard Errors = $\frac{\text{Standard Deviation}}{\sqrt{\text{Sample Size}}}$

```
my.m <- with(issp.df, tapply(total.lik, age.group, mean,
                             na.rm = TRUE))
my.m

Under 35 36 to 60  Over 61
13.38871 12.45516 10.78836

my.sd <- with(issp.df, tapply(total.lik, age.group, sd,
                              na.rm = TRUE))
my.sd

Under 35 36 to 60  Over 61
2.139623 2.156049 1.964491
```

Calculating 95% CIs

```
my.n <- with(issp.df, tapply(total.lik, age.group,
            function(x)length(which(!is.na(x)))))
my.n

Under 35 36 to 60  Over 61
     319       446       189

my.stder <- my.sd/sqrt(my.n)
ci.upper <- my.m + 1.96*my.stder
ci.lower <- my.m - 1.96*my.stder
```

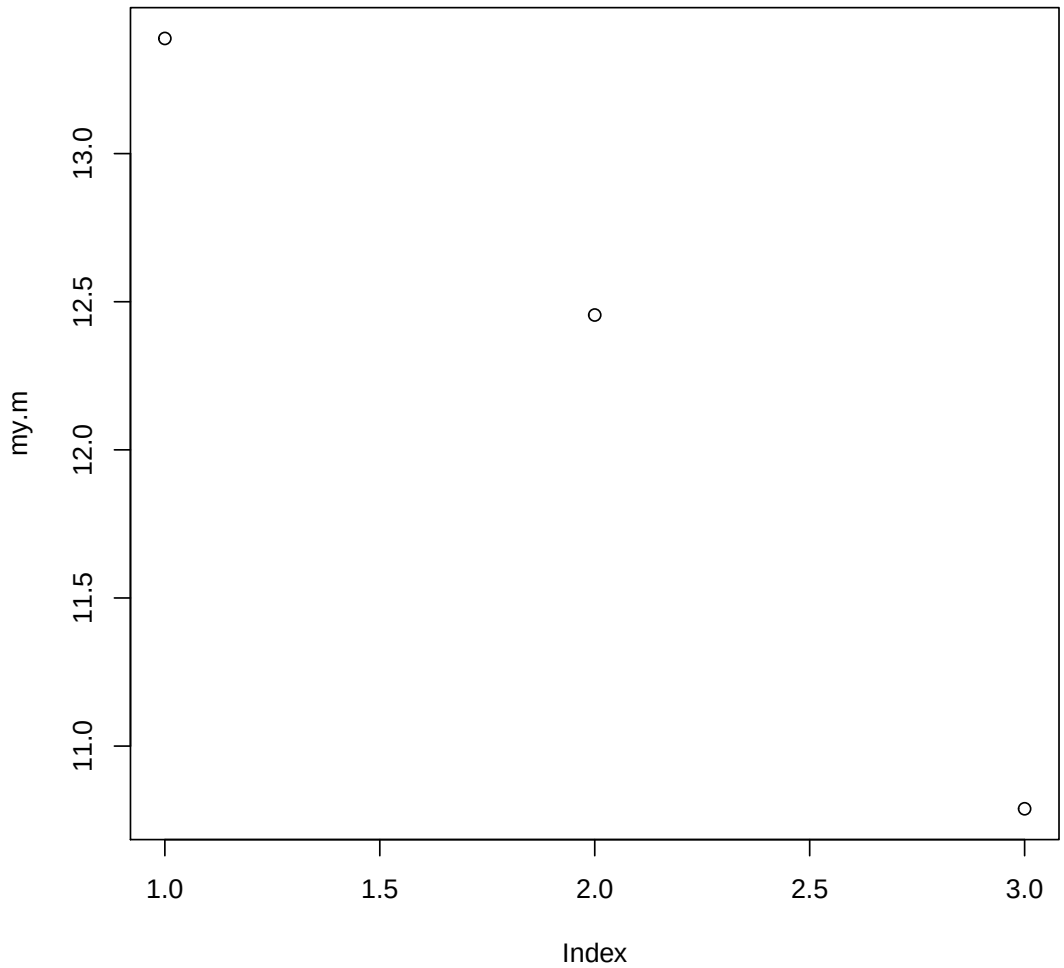
Calculating 95% CIs

```
cbind(my.m, ci.lower, ci.upper)
```

```
             my.m ci.lower ci.upper
Under 35 13.38871 13.15391 13.62351
36 to 60 12.45516 12.25506 12.65526
Over 61  10.78836 10.50828 11.06844
```

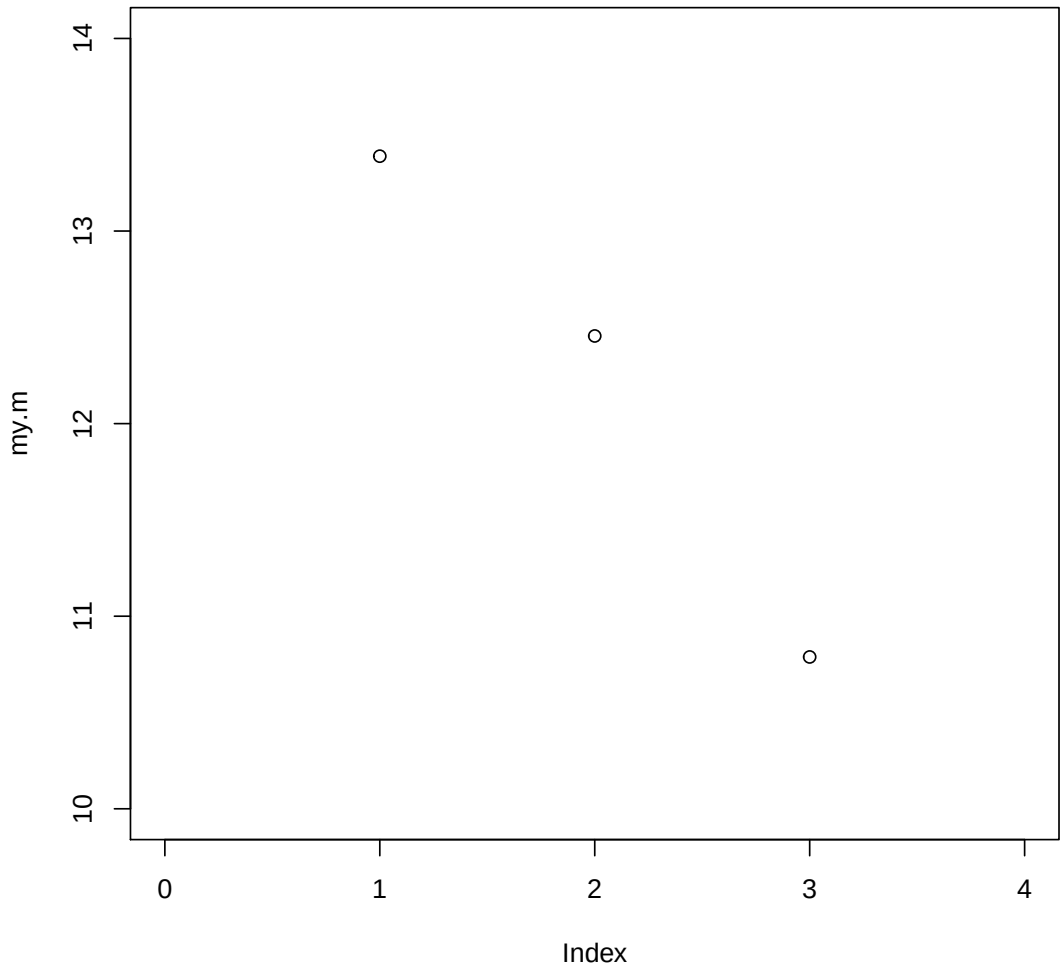
Plot the means

```
plot(my.m)
```



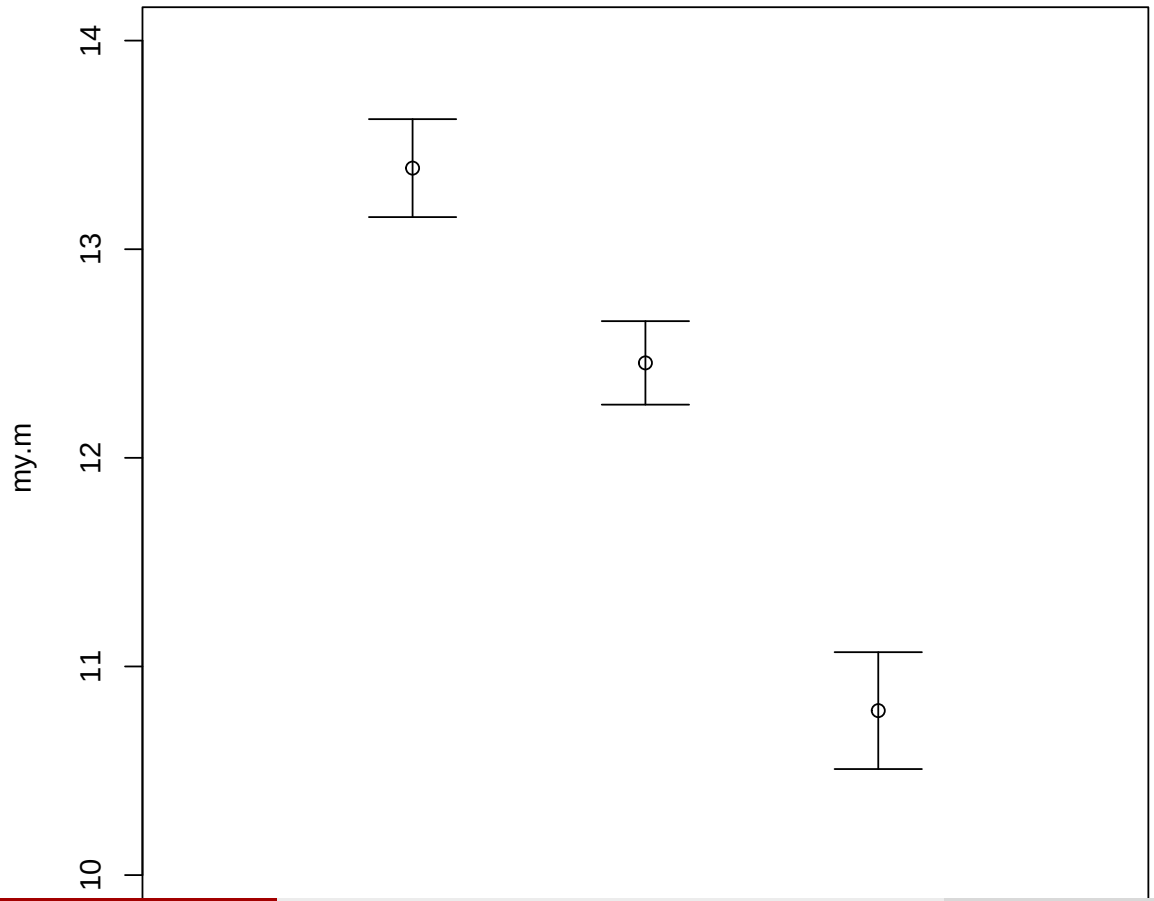
Make room for the SE bars

```
plot(my.m, ylim = c(10, 14), xlim = c(0, 4))
```



Adding error bars

```
plot(my.m, ylim = c(10, 14), xlim = c(0, 4))
arrows(1:3, ci.lower, 1:3, ci.upper, code = 3,
       angle = 90)
```



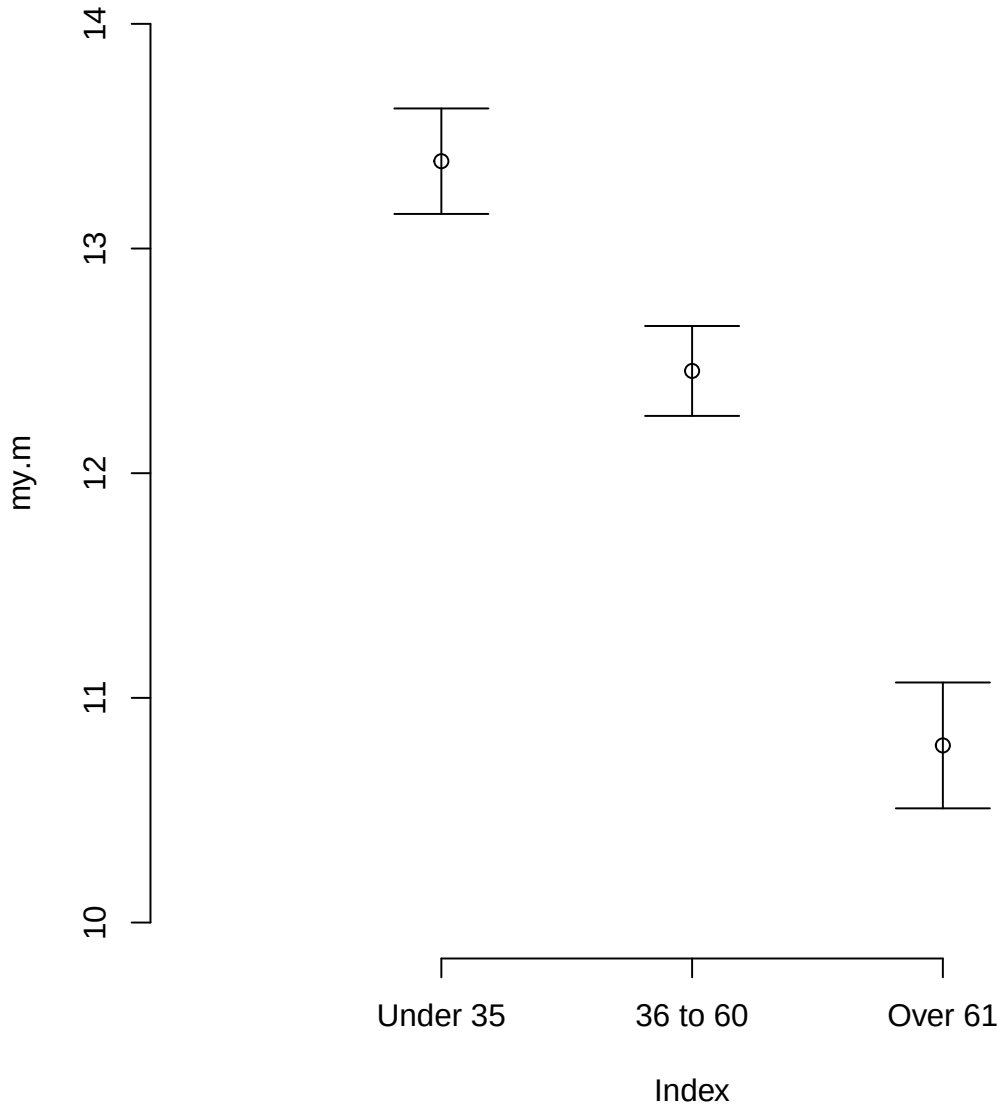
Essential embellishments

The labels on x-axis are wrong:

- Use `axes = FALSE` in `plot()` to prevent the drawing of axes.
- Use `axis()` function to draw the axes manually, with correct labels

```
plot(my.m, ylim = c(10, 14), xlim = c(0, 4), axes = FALSE)
arrows(1:3, ci.lower, 1:3, ci.upper, code = 3,
       angle = 90)
axis(2)
axis(1, at = 1:3,
     label = c("Under 35", "36 to 60", "Over 61"))
```

Essential embellishments

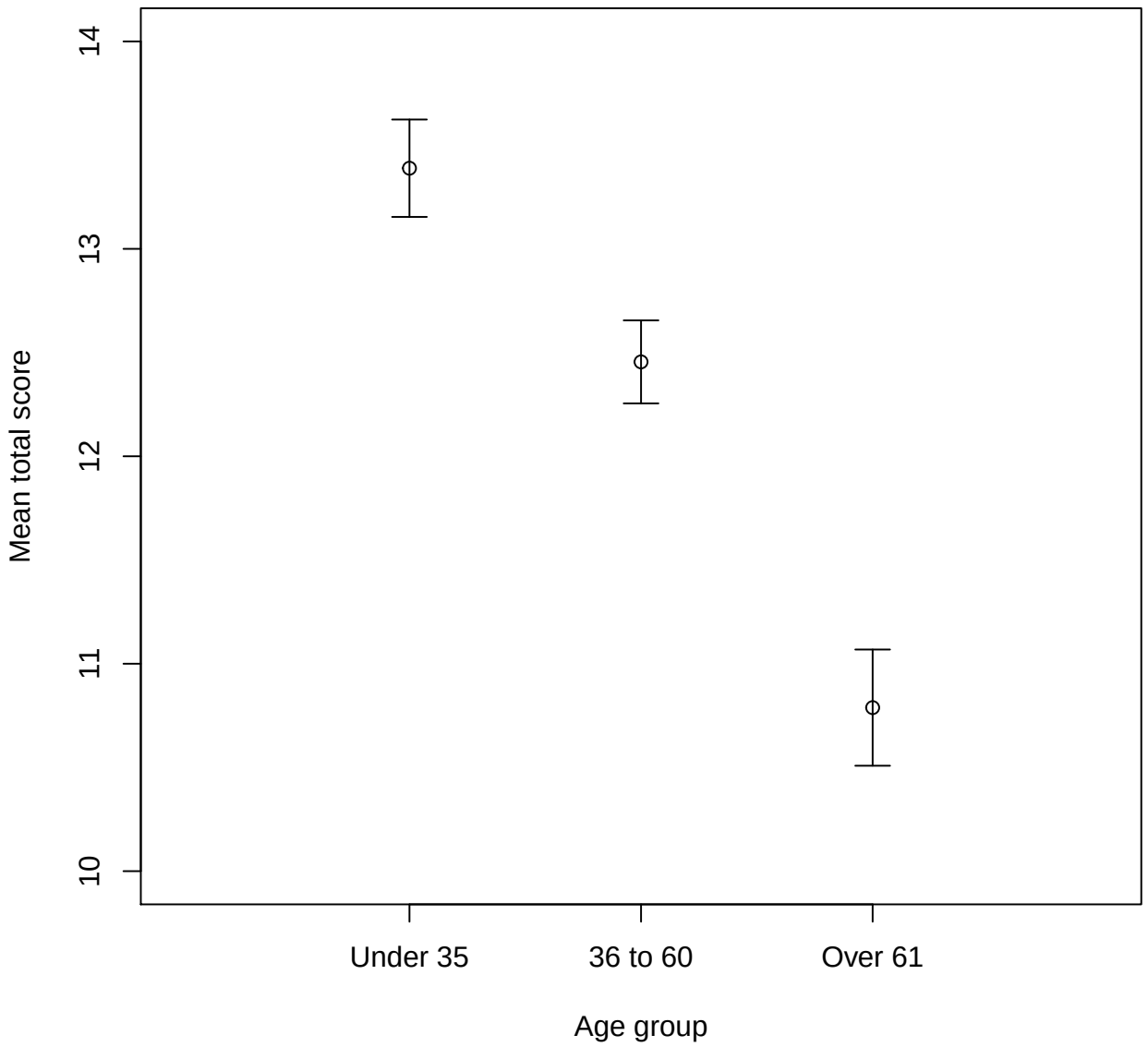


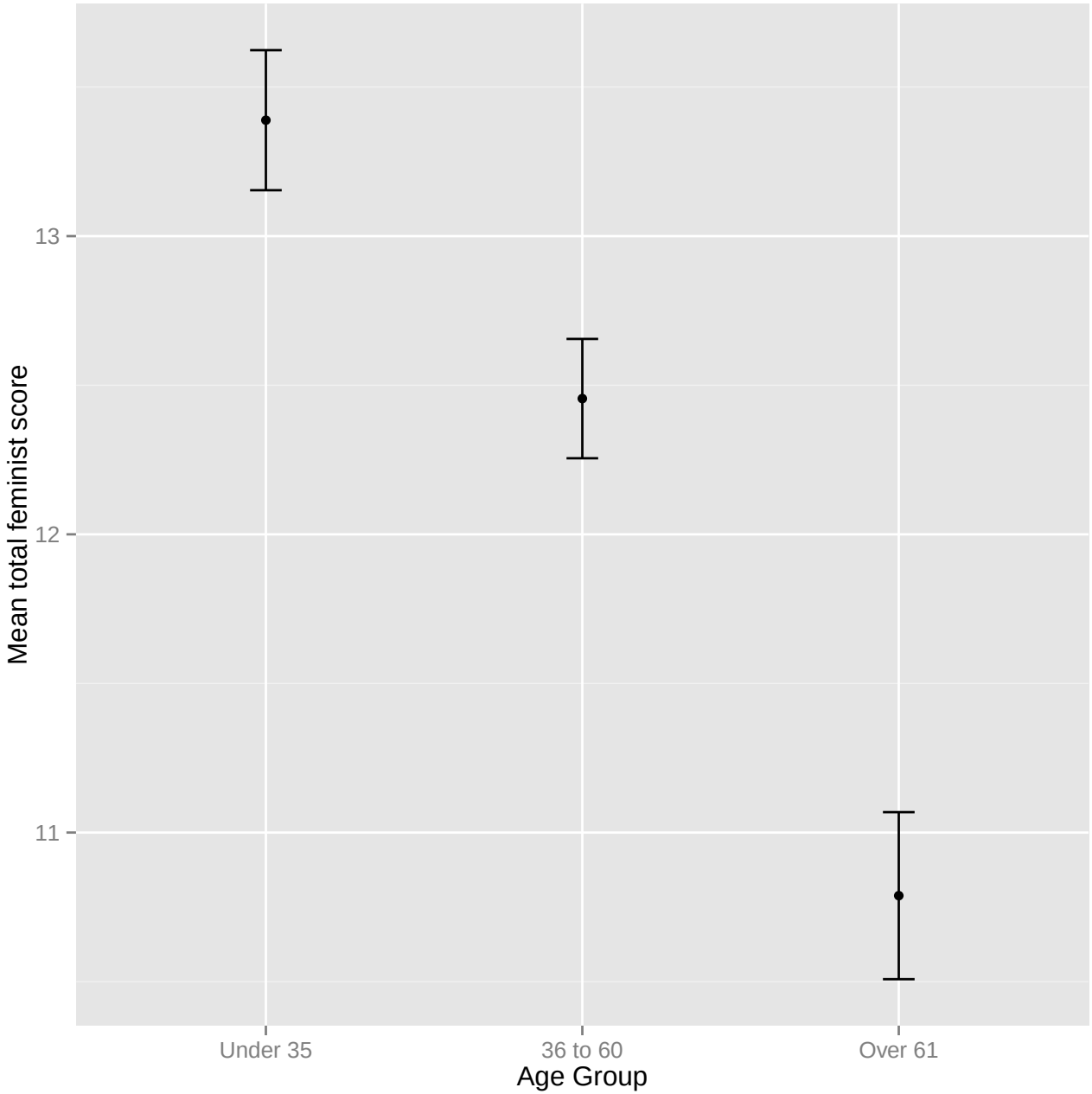
Essential embellishments

- Need to draw an outer box that connects x-axis and y-axis.
- The length of the heads of the error bars are too long.
- Give correct labels for the axes

```
plot(my.m, ylim = c(10, 14), xlim = c(0, 4), axes = FALSE,
     xlab = "Age group",
     ylab = "Mean total score")
arrows(1:3, ci.lower, 1:3, ci.upper, code = 3,
       length = .1, angle = 90)
axis(2)
axis(1, at = 1:3,
     label = c("Under 35", "36 to 60", "Over 61"))
box()
```

Essential embellishments




13
12
11
Mean total feminist score
Under 35
36 to 60
Over 61
Age Group

Any interactions between Gender and Age group?

```
GA.m <- with(issp.df, tapply(total.lik,
           list(Gender, age.group), mean, na.rm = TRUE))
GA.m
```

```
       Under 35 36 to 60  Over 61
Female  13.7500 12.64576 10.87234
Male    12.8963 12.16092 10.70526
```

We must plot these means in context. Therefore, we must calculate their corresponding (95%) CIs.

Calculating 95% CIs

```
GA.sd <- with(issp.df, tapply(total.lik,
                list(Gender, age.group), sd, na.rm = TRUE))
GA.n <- with(issp.df, tapply(total.lik,
              list(Gender,age.group),
              function(x)length(which(!is.na(x)))))
GA.stder <- GA.sd/sqrt(GA.n)
GA.upper <- GA.m + 1.96*GA.stder
GA.lower <- GA.m - 1.96*GA.stder
GA.lower
```

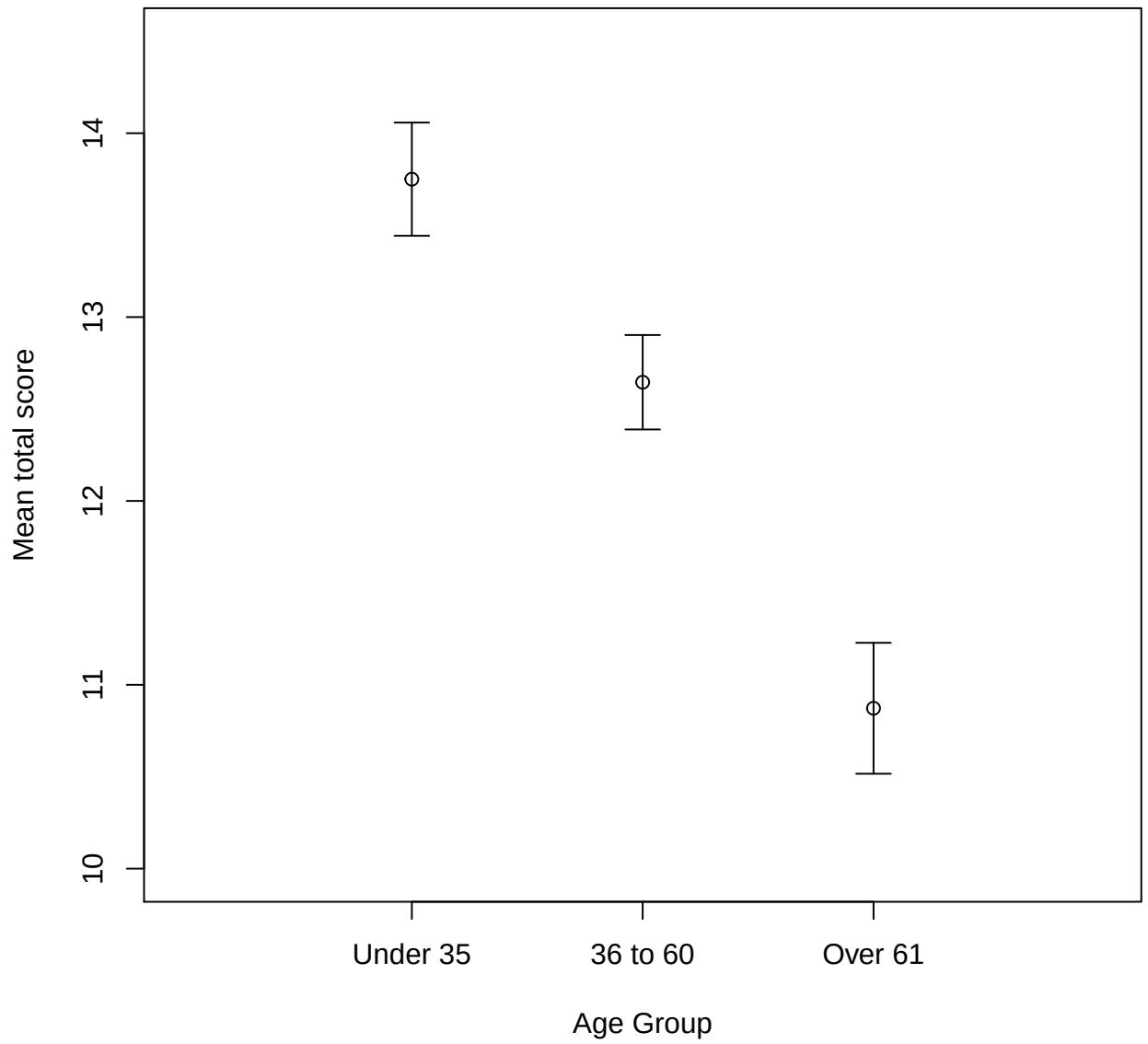
```
       Under 35 36 to 60 Over 61
Female 13.44218 12.38896 10.5163
Male   12.54882 11.84390 10.2723
```

Step-by-step

- The first (second) row of each of `GA.m`, `GA.lower`, `GA.upper` contains the mean, lower 95% CI and upper 95% CI, respectively, for female (male) respondents.
- Plot the points and error bars (corresponding to lower/upper 95% confidence limits) for females.
- Repeat for males.

```
plot(GA.m[1, ], xlab = "Age Group",
     ylab = "Mean feminist score",
     axes = FALSE, xlim = c(0, 4), ylim = c(10, 14.5))
arrows(1:3, GA.upper[1, ], 1:3, GA.lower[1, ],
       angle = 90, code = 3, length = .1)
axis(1, at = 1:3, label = colnames(GA.m))
axis(2)
box()
```

Plotting mean ± 95% CI: Females

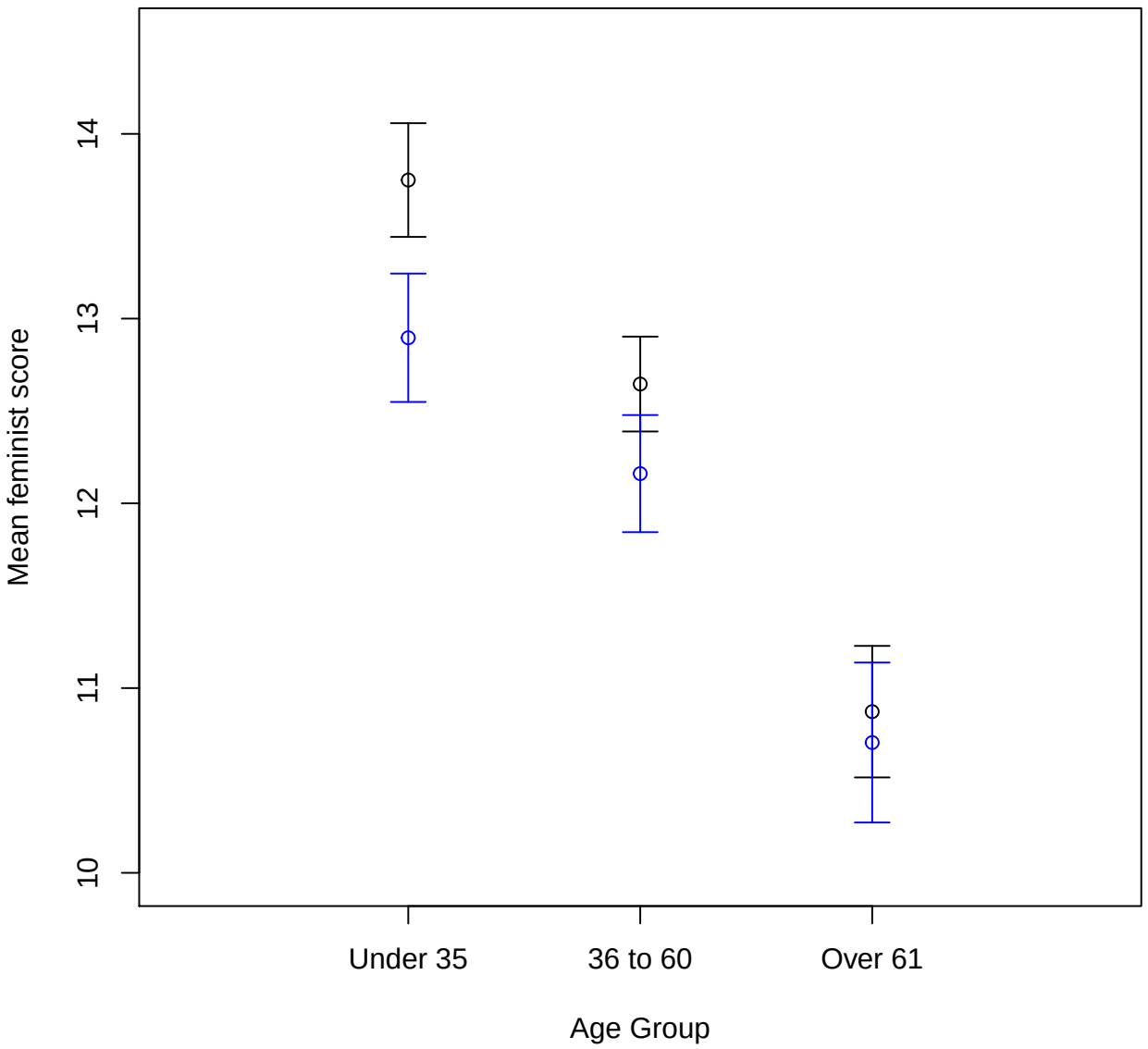


Plotting mean ± 95% CI: Males

Using blue symbols/lines for males.

```
plot(GA.m[1, ], xlab = "Age Group",
     ylab = "Mean total score",
     axes = FALSE, xlim = c(0, 4), ylim = c(10, 14.5))
arrows(1:3, GA.upper[1, ], 1:3, GA.lower[1, ],
       angle = 90, code = 3, length = .1)
axis(1, at = 1:3, label = colnames(GA.m))
axis(2)
box()
points(GA.m[2, ], col = "blue")
arrows(1:3, GA.upper[2, ], 1:3, GA.lower[2, ],
       angle = 90, code = 3, length = .1, col = "blue")
```

Add the means for males



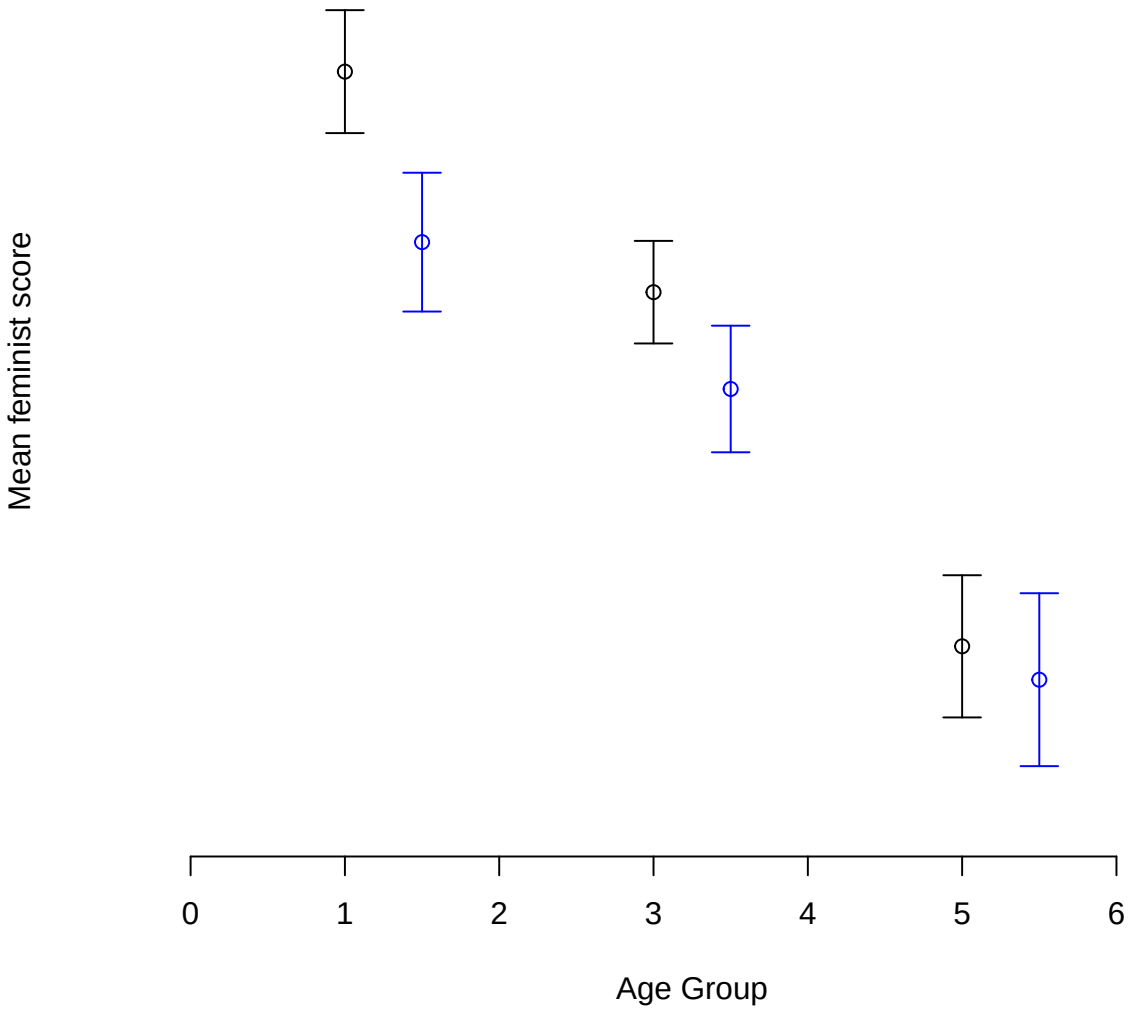
Improvements

- Points/error bars overlay one another. Need to separate them?
- So far, we have plotted means/error bars at x = 1, 2 and 3. We can use different x coordinates for males and females.
- Probably the space between age groups should be larger than the space between males and females.
- How about:
  - Female: x = 1, 3, 5.
  - Male: x = 1.5, 3.5, 5.5

Side-by-side

```
plot(c(1, 3, 5), GA.m[1, ], xlab = "Age Group",
     ylab = "Mean feminist score",col = 1,
     axes = FALSE, xlim = c(0, 6.5), ylim = c(10, 14.5))
arrows(c(1, 3, 5), GA.upper[1, ], c(1, 3, 5),
       GA.lower[1, ],
       angle = 90, code = 3, length = .1, col = 1)
points(c(1.5, 3.5, 5.5),GA.m[2, ], col = "blue")
arrows(c(1.5, 3.5, 5.5), GA.upper[2, ], c(1.5, 3.5, 5.5),
       GA.lower[2, ],
       angle = 90, code = 3, length = .1, col = "blue")
axis(1)
```

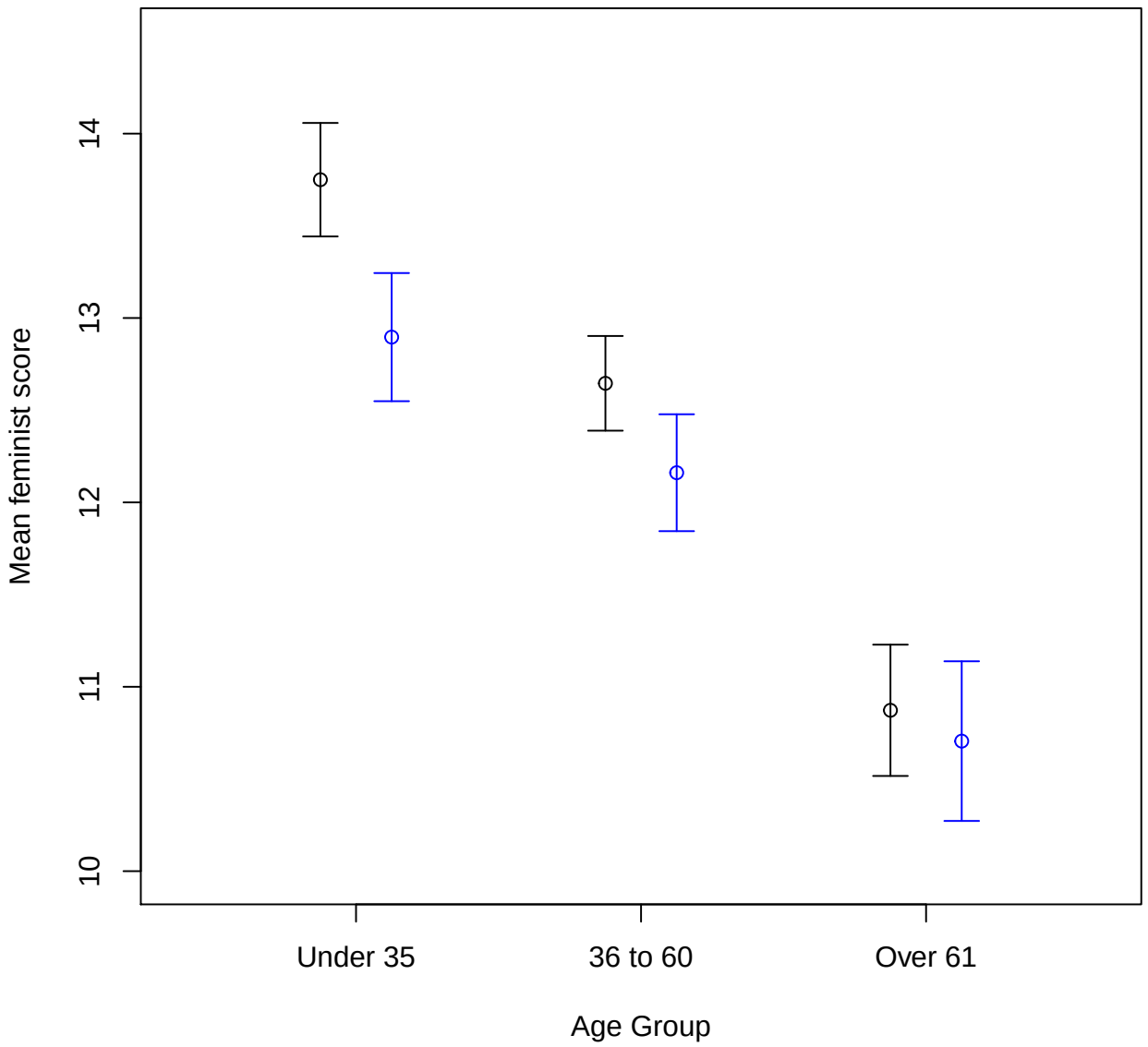
Side-by-side



Tidy up tick marks and labels on x-axis

```
plot(c(1, 3, 5), GA.m[1, ], xlab = "Age Group",
     ylab = "Mean feminist score",col = 1,
     axes = FALSE, xlim = c(0, 6.5), ylim = c(10, 14.5))
arrows(c(1, 3, 5), GA.upper[1, ], c(1, 3, 5),
       GA.lower[1, ],
       angle = 90, code = 3, length = .1, col = 1)
points(c(1.5, 3.5, 5.5),GA.m[2, ], col = "blue")
arrows(c(1.5, 3.5, 5.5), GA.upper[2, ], c(1.5, 3.5, 5.5),
       GA.lower[2, ],
       angle = 90, code = 3, length = .1, col = "blue")
axis(1, at = c(1.25, 3.25, 5.25), label = colnames(GA.m))
axis(2)
box()
```

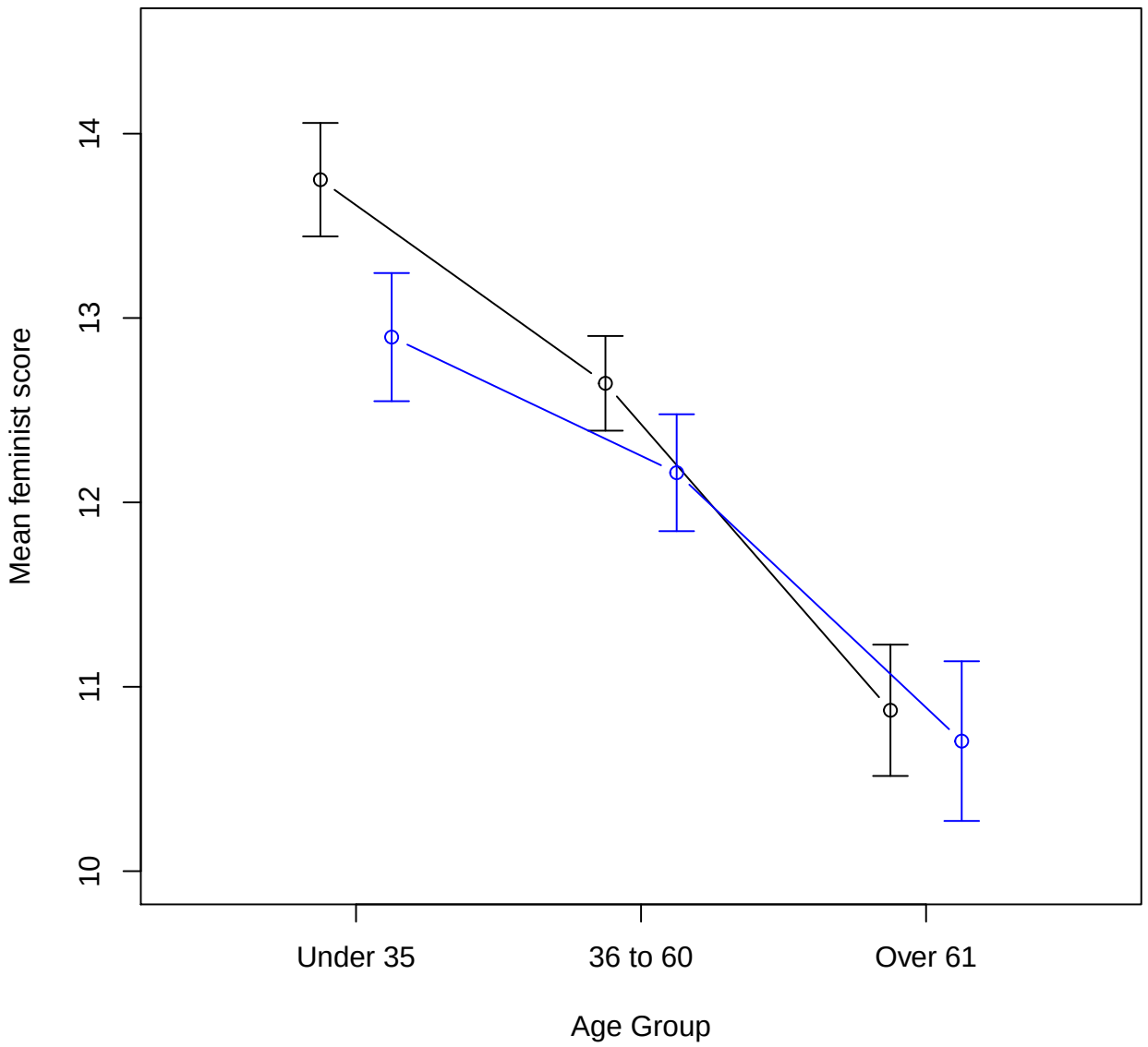
Add a box around the figure



Connect the points?

```
plot(c(1, 3, 5), GA.m[1, ], xlab = "Age Group",
     ylab = "Mean feminist score",col = 1, type = "b",
     axes = FALSE, xlim = c(0, 6.5), ylim = c(10, 14.5))
arrows(c(1, 3, 5), GA.upper[1, ], c(1, 3, 5),
       GA.lower[1, ],
       angle = 90, code = 3, length = .1, col = 1)
points(c(1.5, 3.5, 5.5),GA.m[2, ], col = "blue", type = "b")
arrows(c(1.5, 3.5, 5.5), GA.upper[2, ], c(1.5, 3.5, 5.5),
       GA.lower[2, ],
       angle = 90, code = 3, length = .1, col = "blue")
axis(1, at = c(1.25, 3.25, 5.25), label = colnames(GA.m))
axis(2)
box()
```

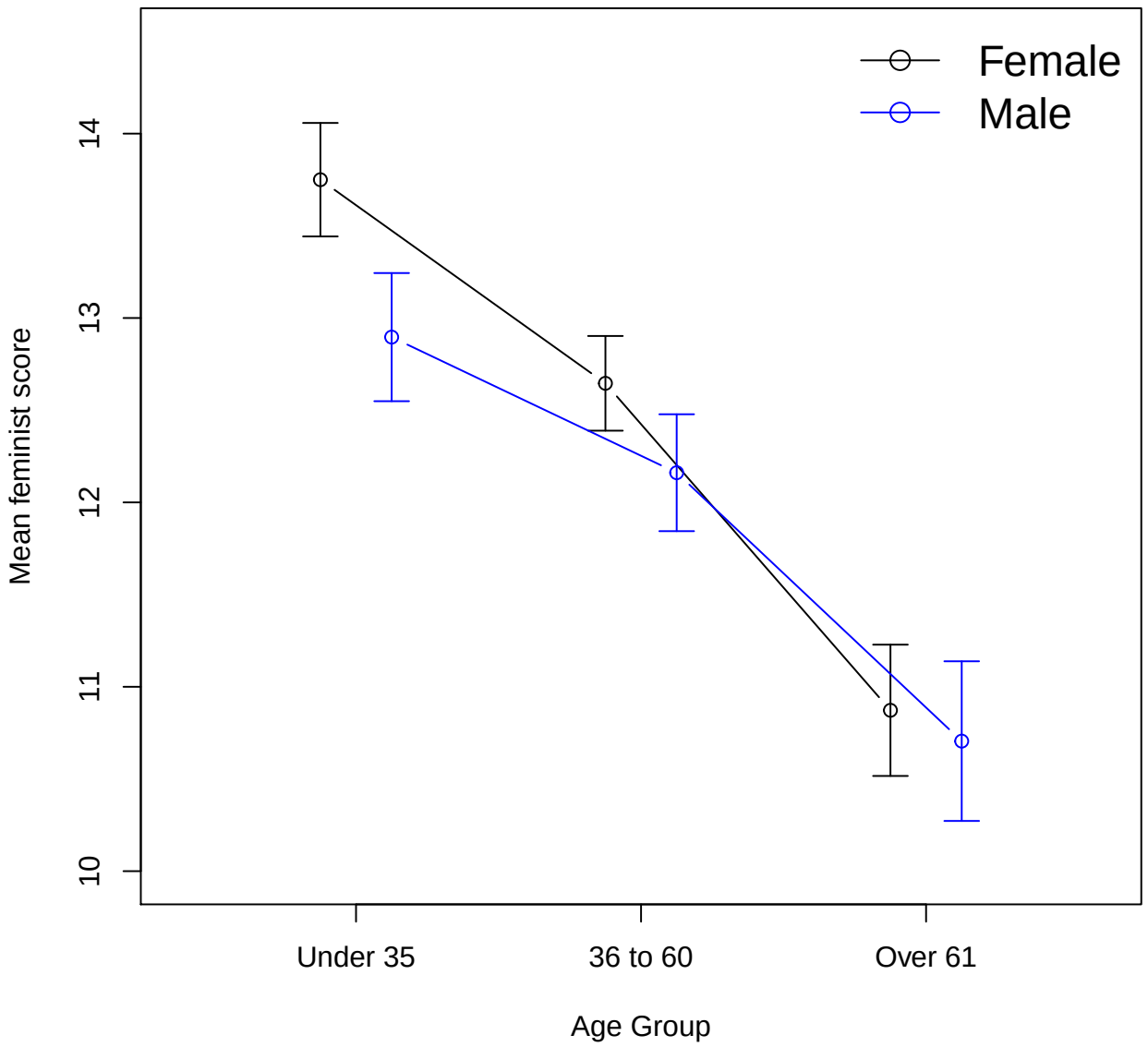
Connect the points

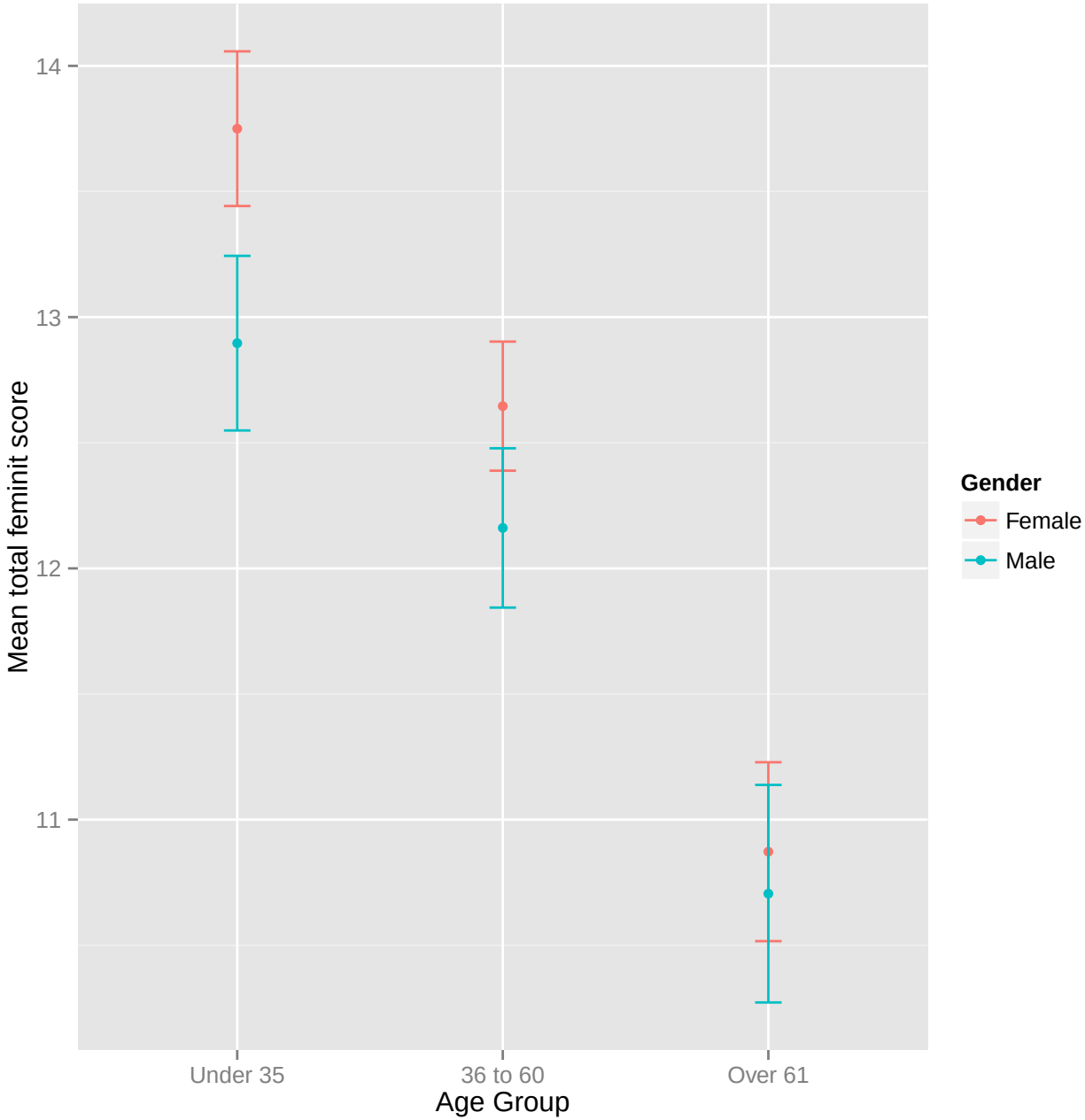


Add legend

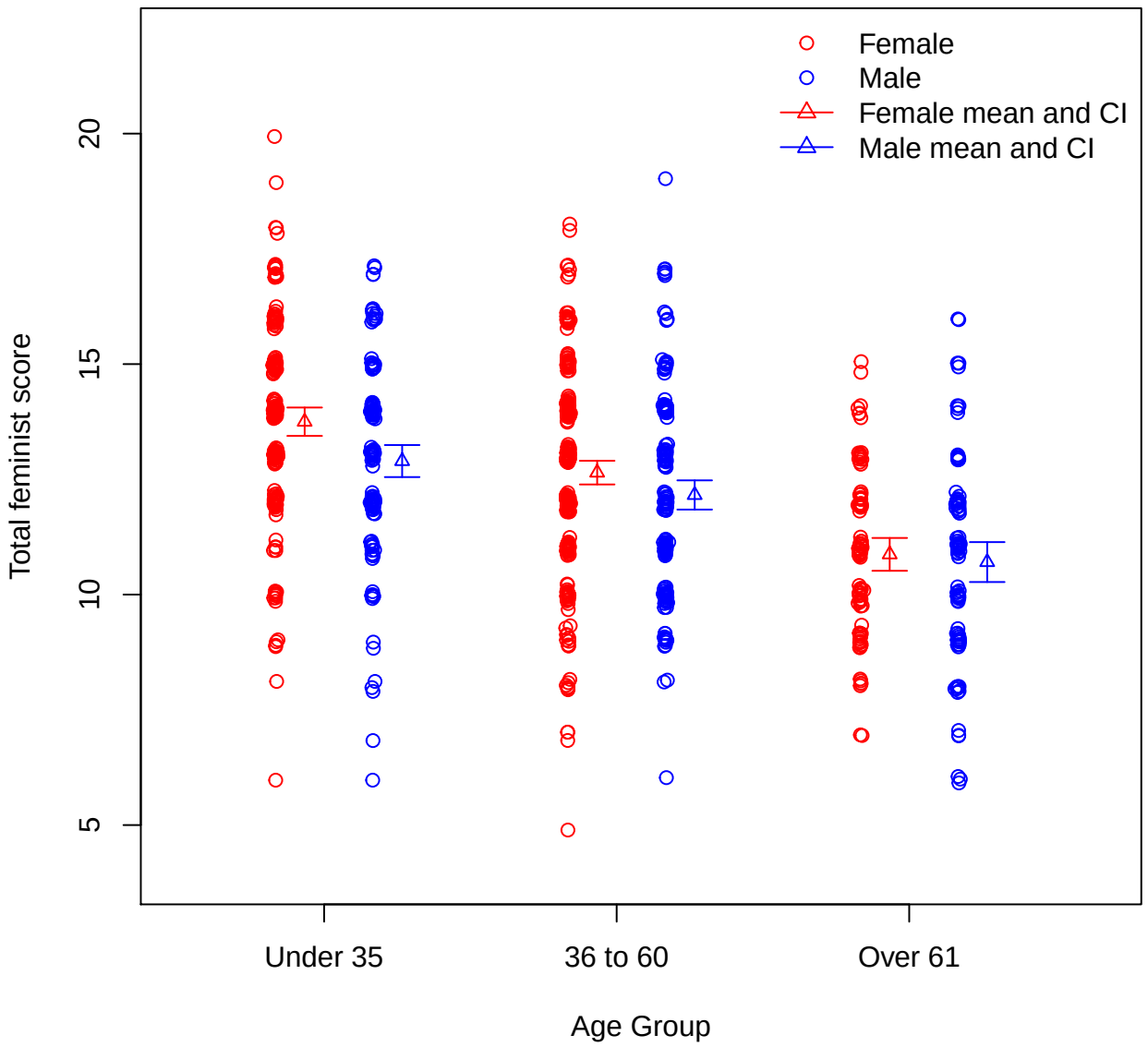
```
plot(c(1, 3, 5), GA.m[1, ], xlab = "Age Group",
     ylab = "Mean feminist score",col = 1, type = "b",
     axes = FALSE, xlim = c(0, 6.5), ylim = c(10, 14.5))
arrows(c(1, 3, 5), GA.upper[1, ], c(1, 3, 5),
       GA.lower[1, ],
       angle = 90, code = 3, length = .1, col = 1)
points(c(1.5, 3.5, 5.5),GA.m[2, ], col = "blue", type = "b")
arrows(c(1.5, 3.5, 5.5), GA.upper[2, ], c(1.5, 3.5, 5.5),
       GA.lower[2, ],
       angle = 90, code = 3, length = .1, col = "blue")
axis(1, at = c(1.25, 3.25, 5.25), label = colnames(GA.m))
axis(2)
box()
legend("topright", pch = 21, lty = 1, bty = "n",
       col = c("black", "blue"), cex = 1.5,
       legend = c("Female", "Male"))
```

Add legend


Female
Male
Mean feminist score
10
11
12
13
14
Under 35
36 to 60
Over 61
Age Group


14
13
12
11
Mean total feminit score
Under 35
36 to 60
Over 61
Age Group
Gender
Female
Male

How far can we go?



Lattice plots