NZSSN Courses: Introduction to R

Session 5 – Graphics

Statistical Consulting Centre

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

2 March, 2017

SCIENCE
DEPARTMENT OF STATISTICS

Exploring your data

- Examine for patterns, relationships, structures, and other features.
- Do this using graphs, tables and summary statistics.
- Patterns are more easily seen in graphs.
- Remember:

> "R is a free software environment for statistical computing and *graphics*"

Designing graphs

Two important considerations when designing a graph:

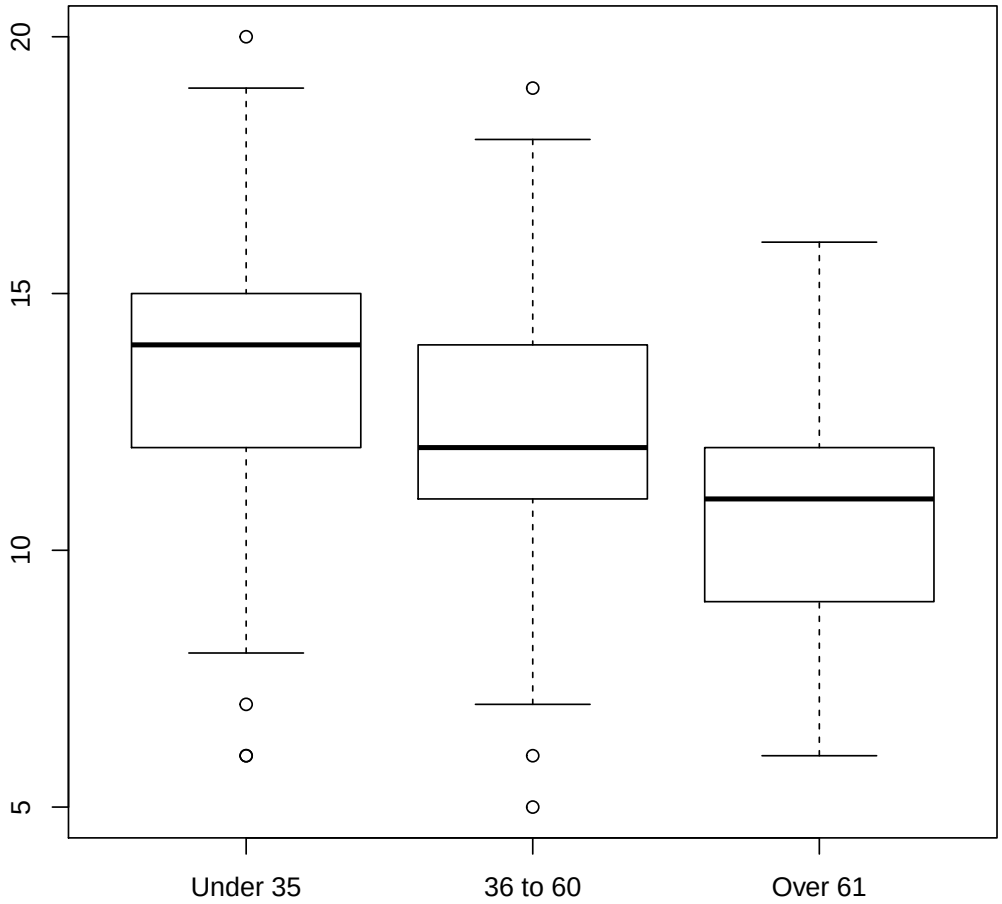
- It should have something to say.
  - Decide what information you want your graph to display.
  - Graphical software is capable of producing any number of different graphs, so generating graphs with meaningless information is quite simple to do!
- It should be easy to interpret.
  - Simplicity is key!
  - Sufficiently large fonts, all axes labelled, clearly defined legends, etc.

Some functions for drawing graphs

- Usually, we use `plot()` to create a graph and then `lines()`, `points()`, `text()`, and other commands to annotate it.
- `plot()` does appropriate things for different types of variables (see later examples).
- Variables of type `character` are not suitable for plotting (and analysis). Convert such variables to type `factor`.
- There are no "erase" or "undo" functions, so store your commands in an `R` script.

Boxplots

```
with(issp.df, plot(age.group, total.lik))
```



Boxplots

```
with(issp.df, plot(Gender, total.lik))

Warning in xy.coords(x, y, xlabel, ylabel, log):  NAs
introduced by coercion
Warning in min(x):  no non-missing arguments to min;
returning Inf
Warning in max(x):  no non-missing arguments to max;
returning -Inf
Error in plot.window(...):  need finite 'xlim' values
```

Boxplots

`age.group` is a factor but `Gender` is not.

```
class(issp.df$Gender)
```

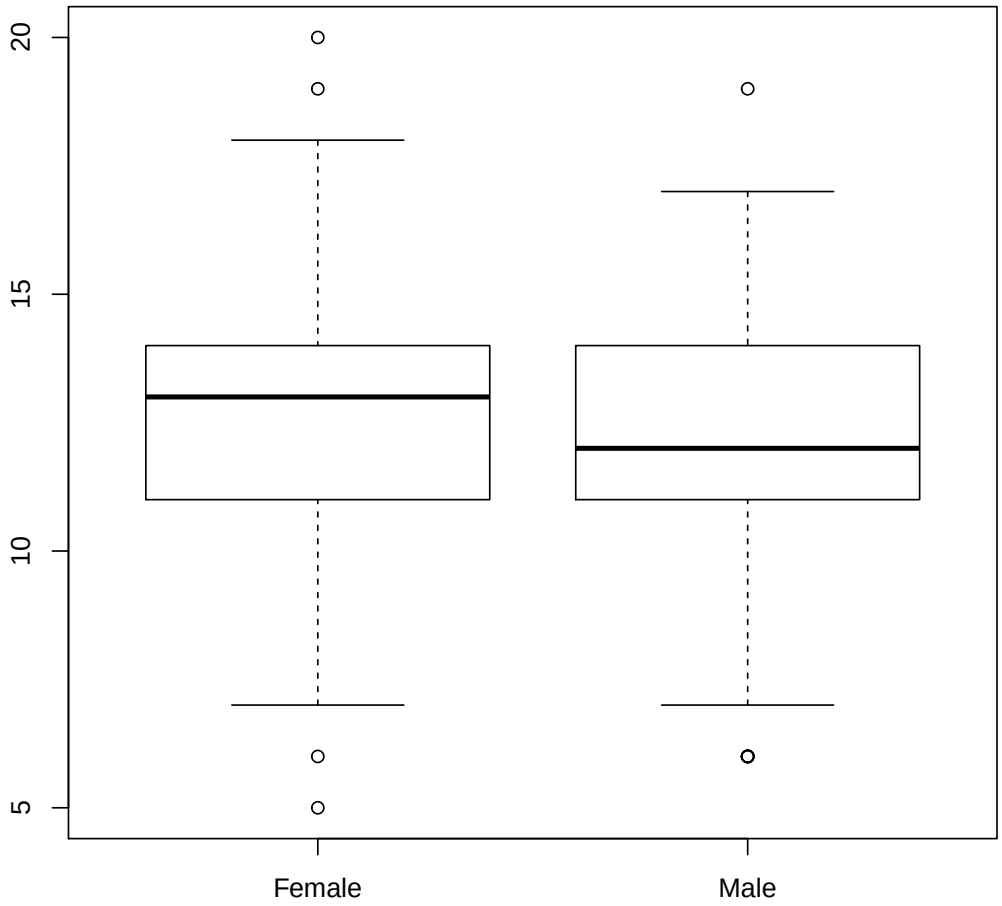
```
[1] "character"
```

```
class(issp.df$age.group)
```

```
[1] "factor"
```

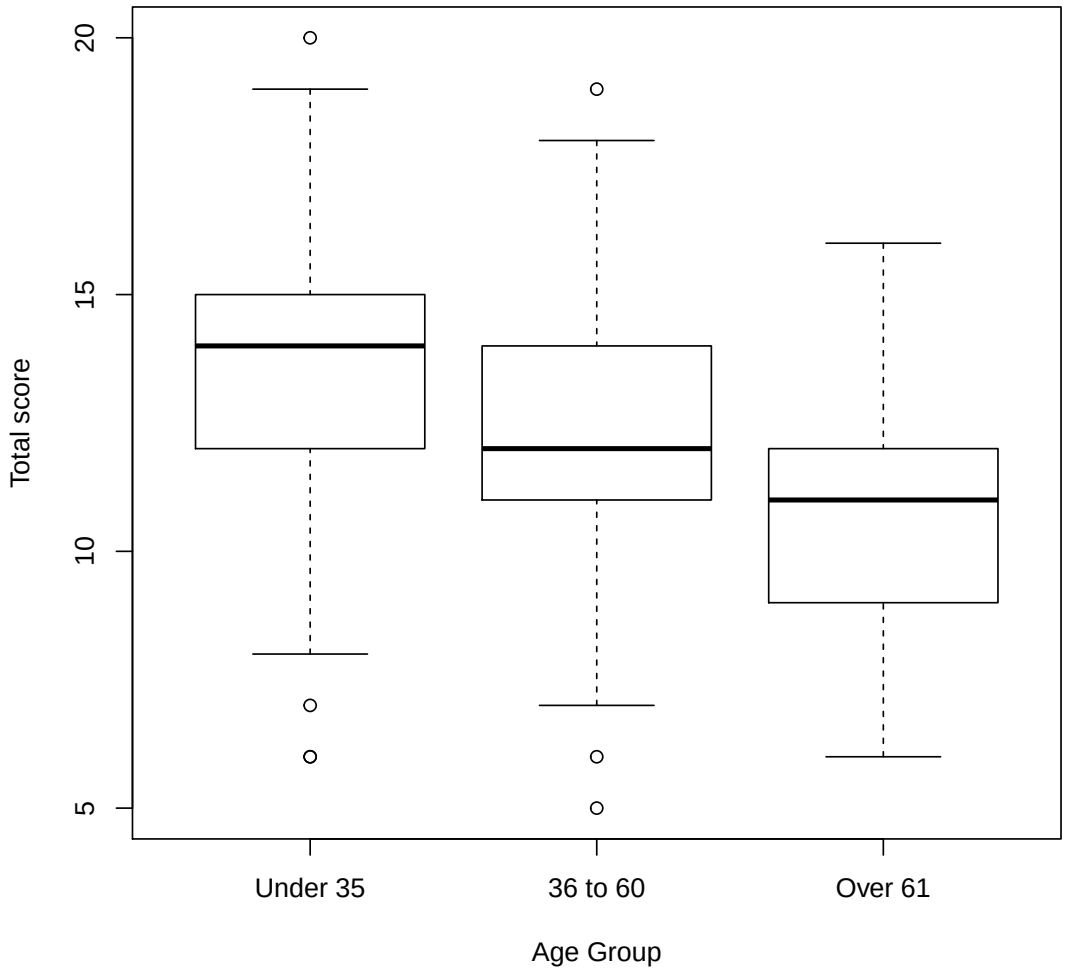
Boxplots

```
with(issp.df, plot(factor(Gender), total.lik))
```



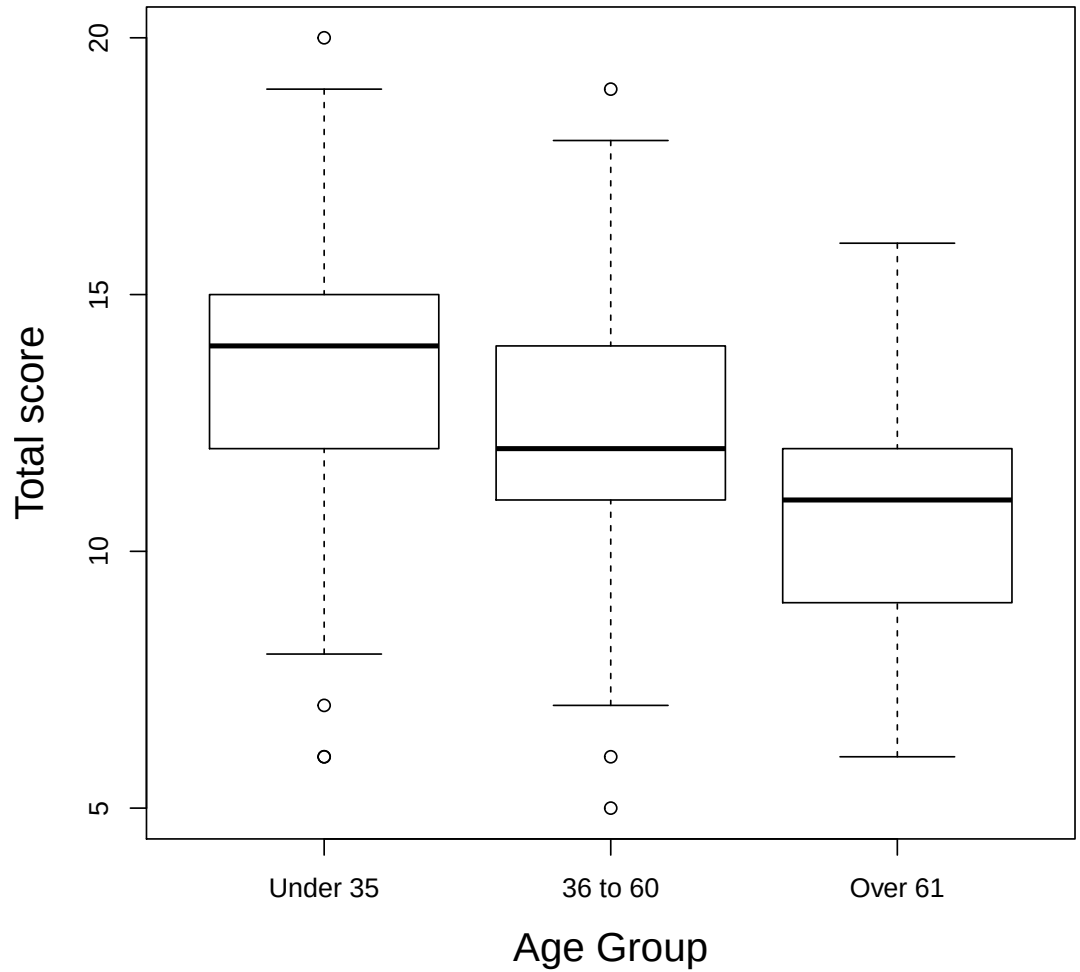
Axes need labels

```
with(issp.df, plot(age.group, total.lik,
     xlab = "Age Group", ylab = "Total score"))
```



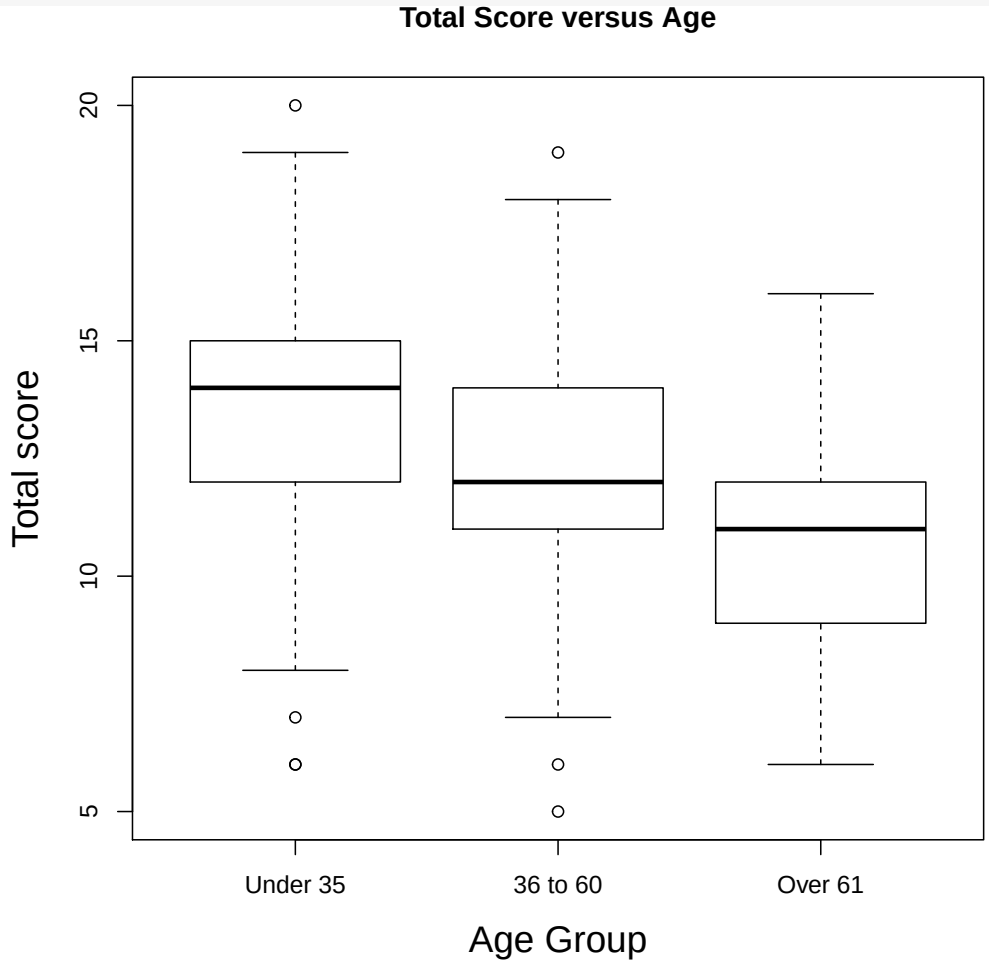
Axis labels are too small!!

```
with(issp.df, plot(age.group, total.lik, cex.lab = 1.5,
    xlab = "Age Group", ylab = "Total score"))
```



Sometimes you need a title

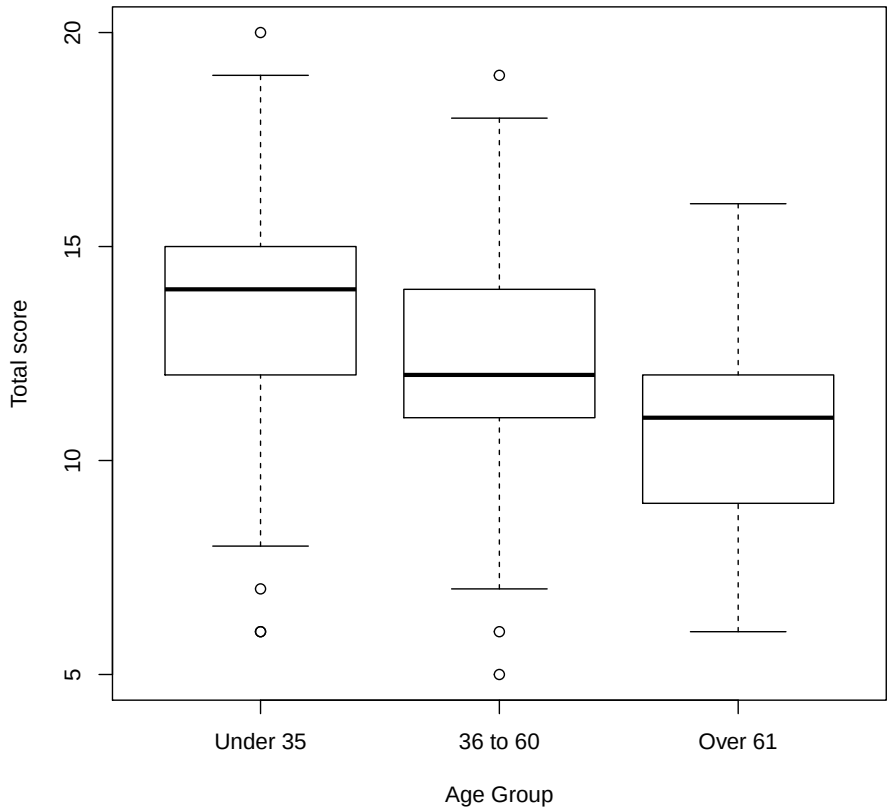
```
with(issp.df, plot(age.group, total.lik, cex.lab = 1.5,
    xlab = "Age Group", ylab = "Total score",
     main = "Total Score versus Age"))
```



boxplot()

More flexible than `plot()`, **if** drawing boxplots. Check the help documentation for more details.

```
with(issp.df, boxplot(total.lik~age.group,
    xlab = "Age Group", ylab = "Total score"))
```



Importance of childhood obedience by age group

Q5: Is obedience important in terms of preparing children for life?

Examine the data via a two-way frequency table.

```
with(issp.df, table(Q5, age.group))
```

```
                    age.group
Q5                   Under 35 36 to 60 Over 61
  be obedient              38       74       75
  can t choose, dk         30       34       12
  NA, refused               6        9        9
  think themselves        259      353      122
```

```
issp.df$Q5 <- with(issp.df,
                   ifelse(Q5 == "can t choose, dk" |
                          Q5 == "NA, refused", NA, Q5))
```

Importance of childhood obedience by age group

First, convert counts to percentages. Why?

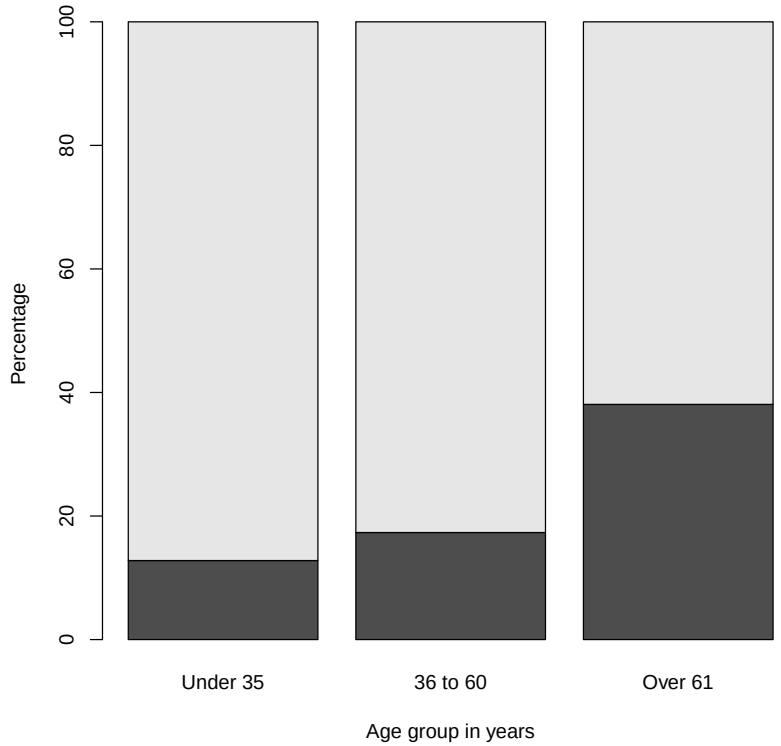
```
q5.age <- with(issp.df,
               100*prop.table(table(Q5, age.group), 2))
round(q5.age, 1)
```

```
                   age.group
Q5                  Under 35 36 to 60 Over 61
  be obedient            12.8     17.3    38.1
  think themselves       87.2     82.7    61.9
```

How do we graphically display the information contained in this table?

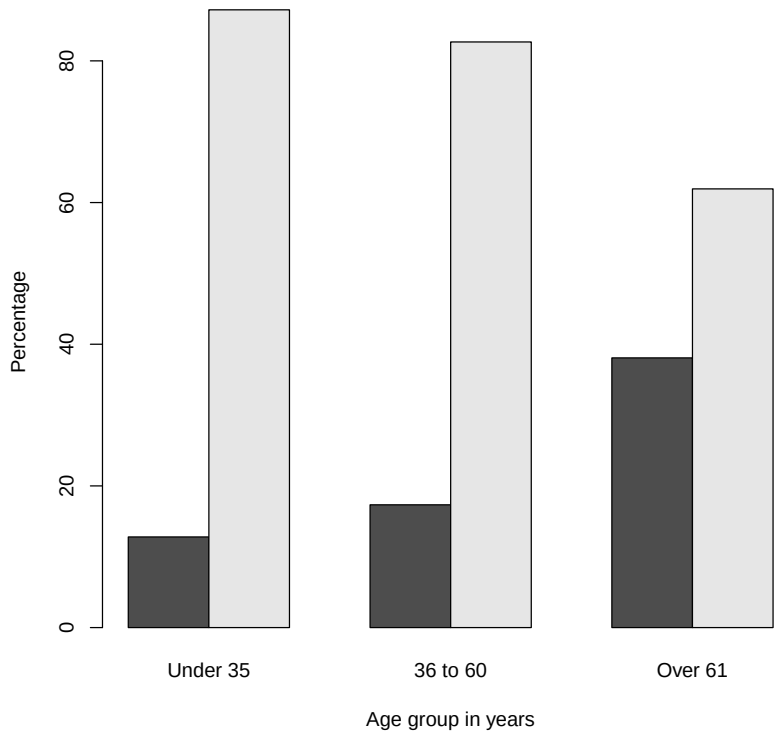
Absolute disaster!!!

```
barplot(q5.age, xlab = "Age group in years",
        ylab = "Percentage")
```



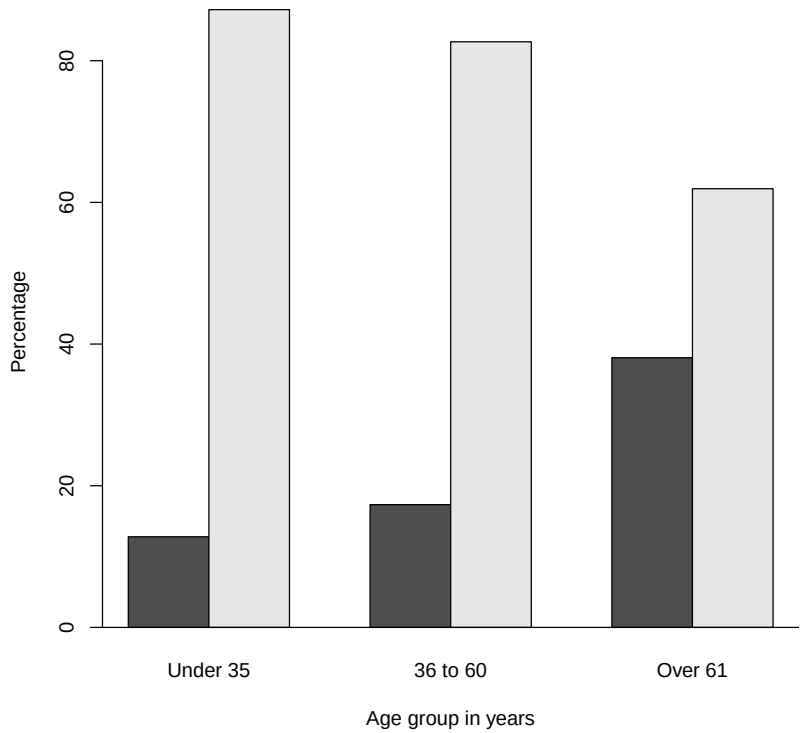
Better

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years")
```



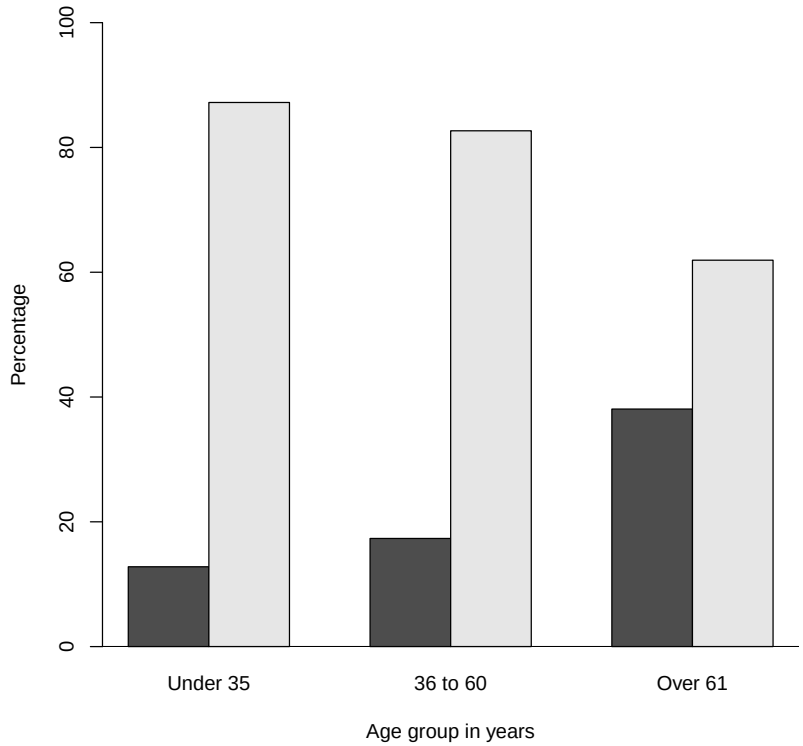
Line along horizontal axis?

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years")
abline(h = 0)
```



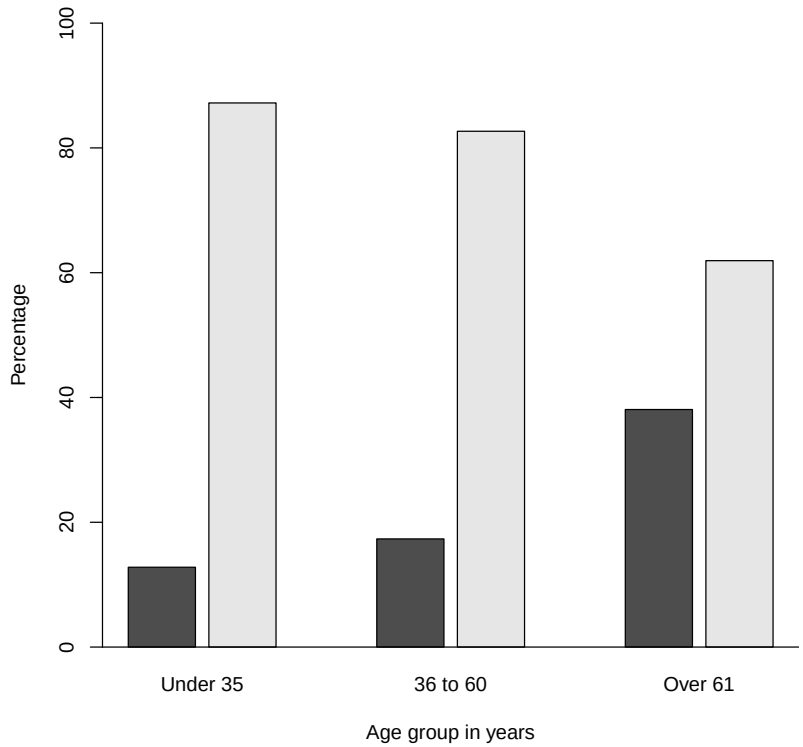
Vertical axis higher than bars?

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100))
abline(h = 0)
```



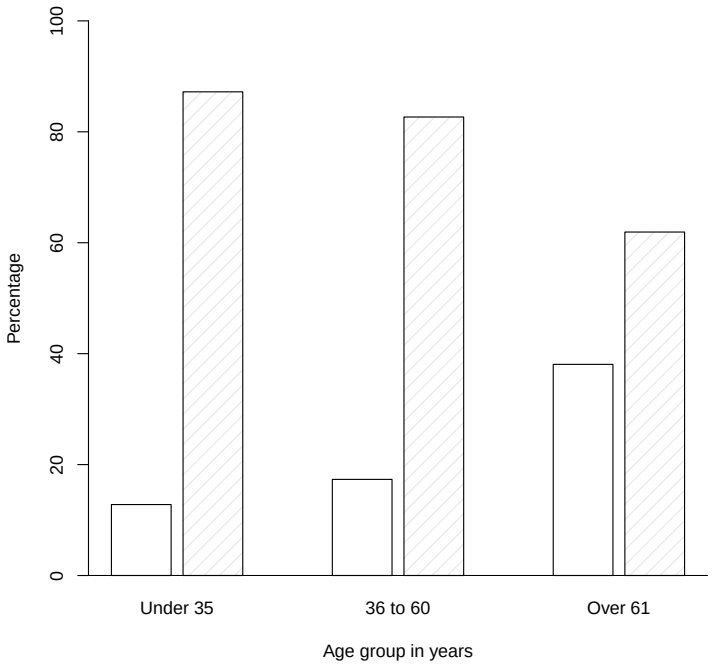
Adding space between bars

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5))
abline(h = 0)
```



Shading with lines?

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5), density = c(0, 10))
abline(h = 0)
```



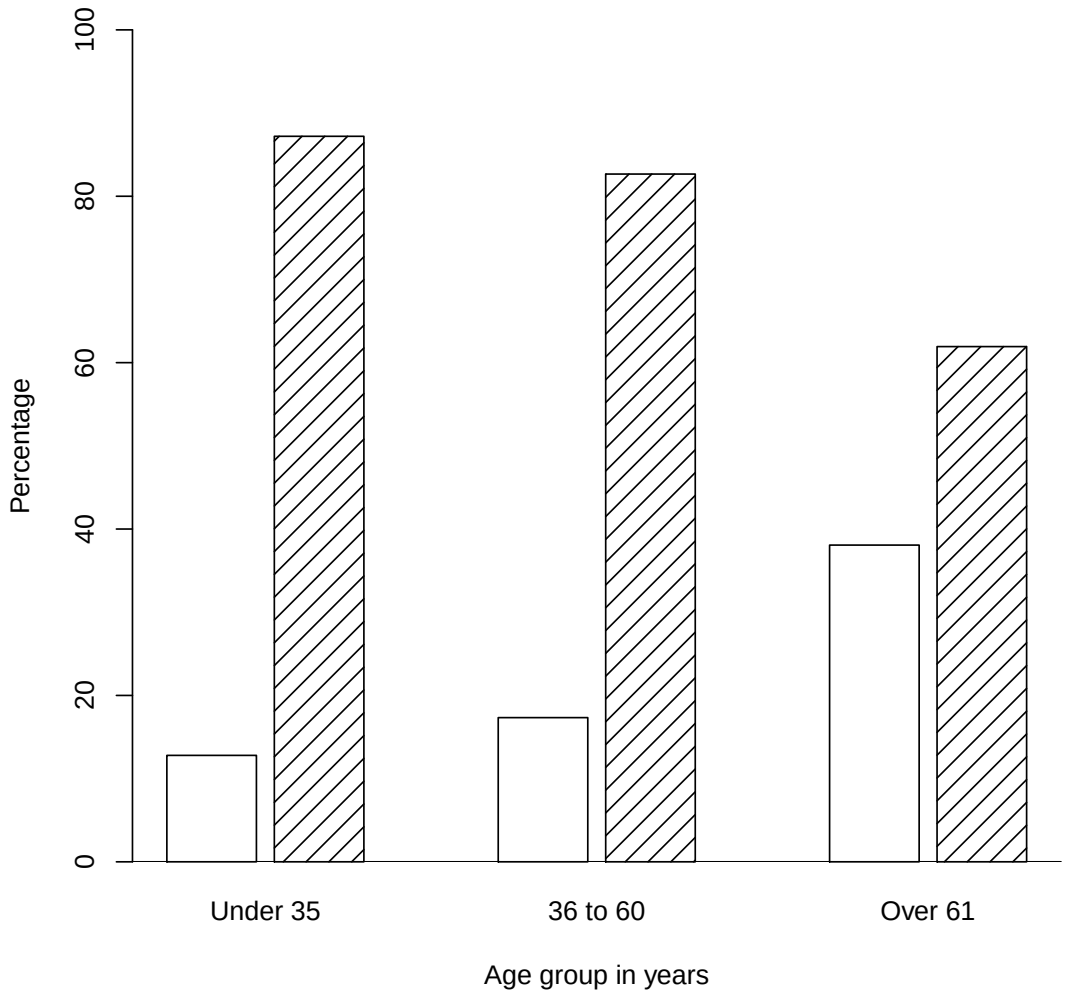
Can't see the shading?

Shading with lines?

The default colour for shading lines is pale white! Change to black using the `col` argument.

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5), density = c(0, 10),
        col = "black")
abline(h = 0)
```

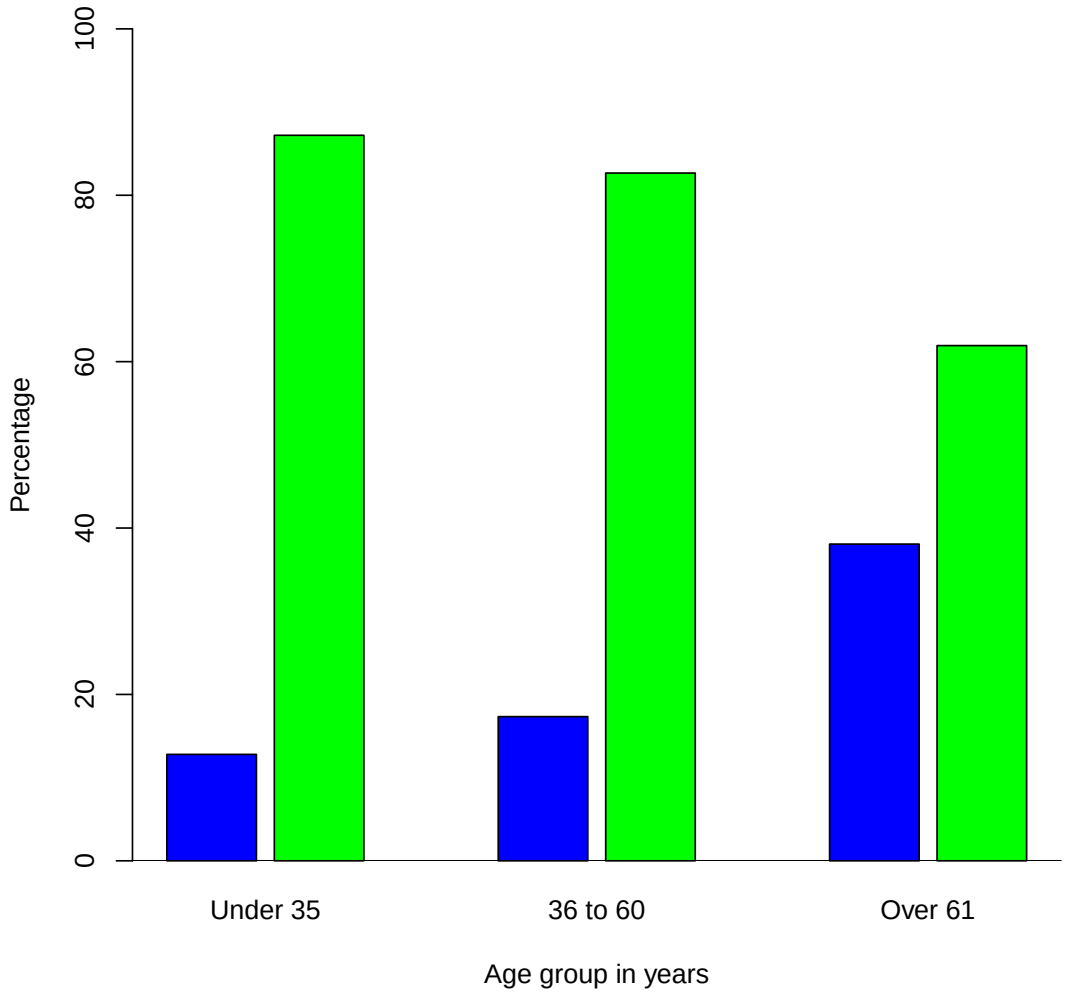
Shading with lines?



Color-filled bars?

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5), col = c("blue", "green"))
abline(h = 0)
```

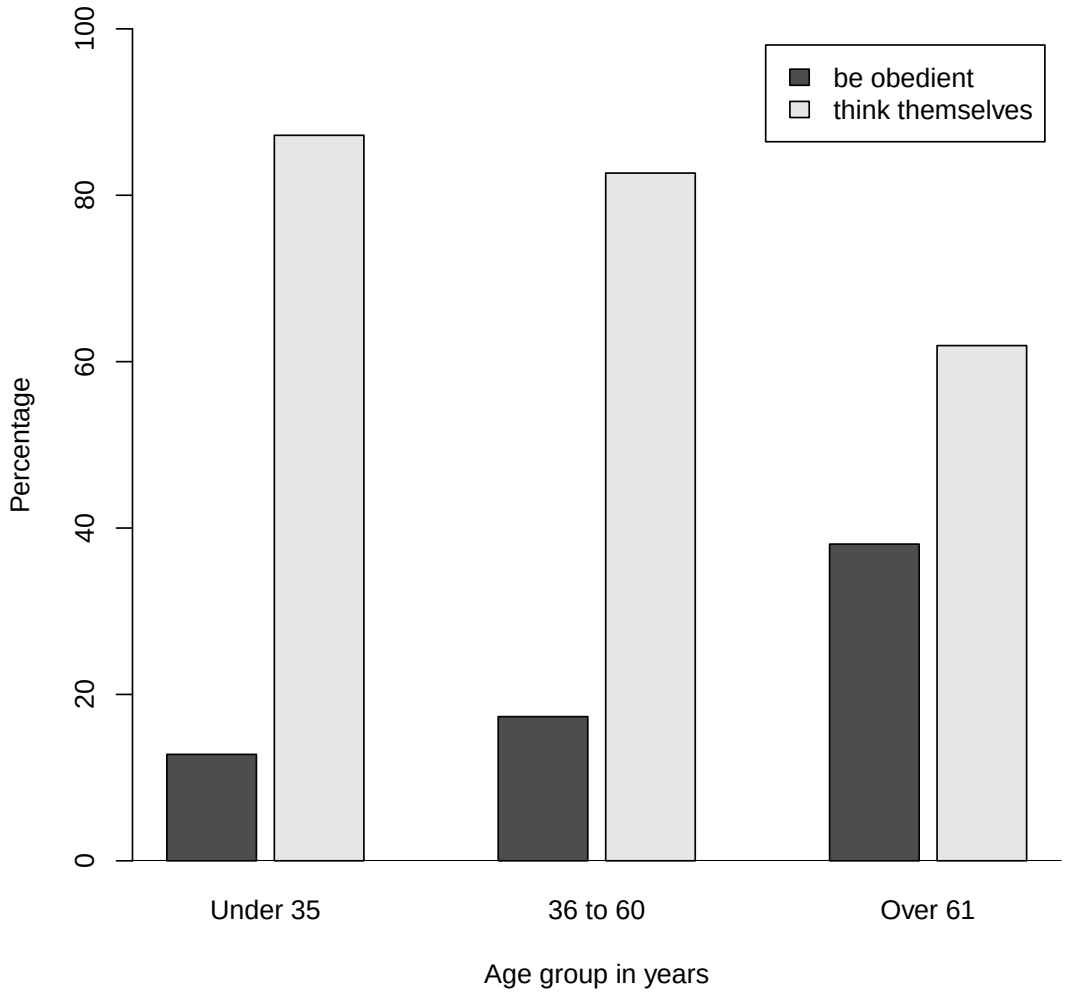
Color-filled bars?



Legend?

```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5), legend.text = TRUE)
abline(h = 0)
```

Legend?

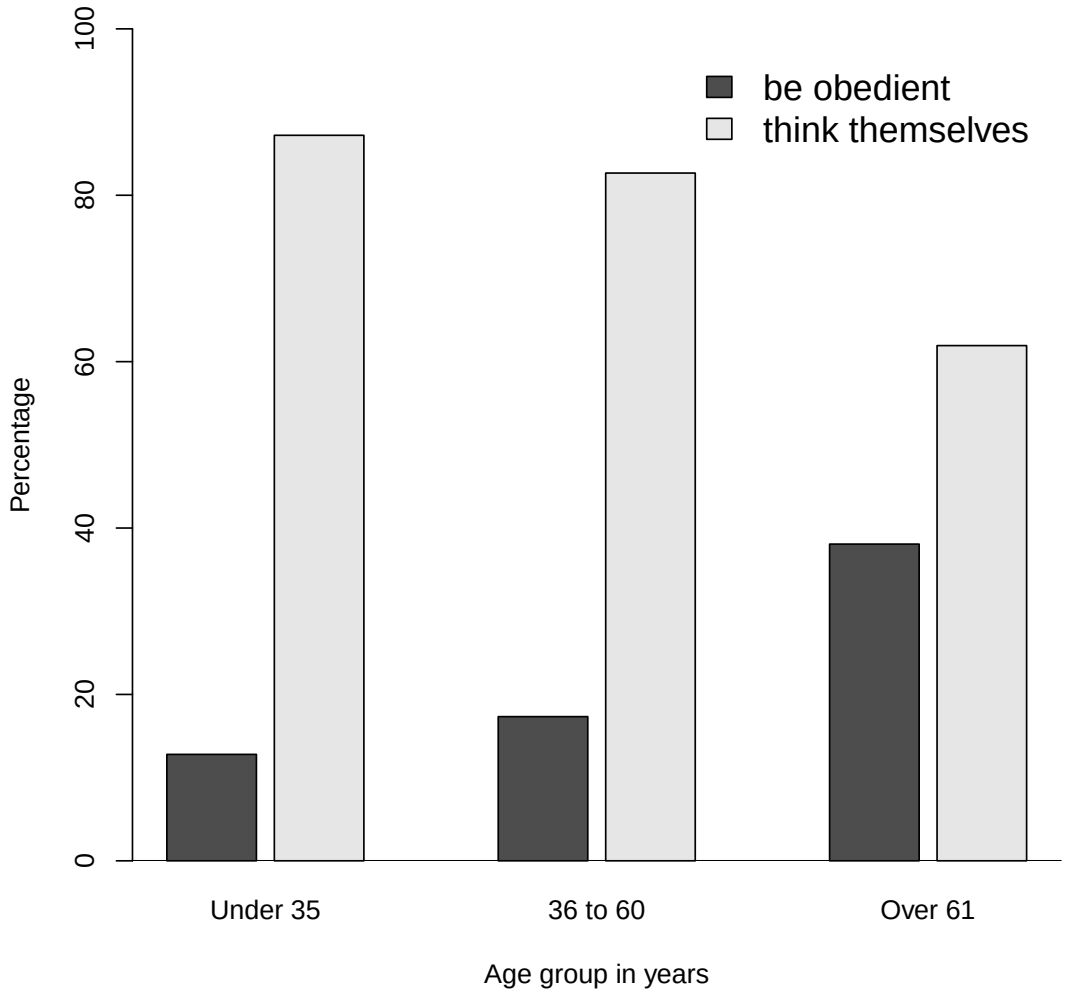

be obedient
think themselves
Percentage
0
20
40
60
80
100
Under 35
36 to 60
Over 61
Age group in years

Further improvements

- Larger title for axes
- Get rid of the box drawn outside the legend
- Larger legend

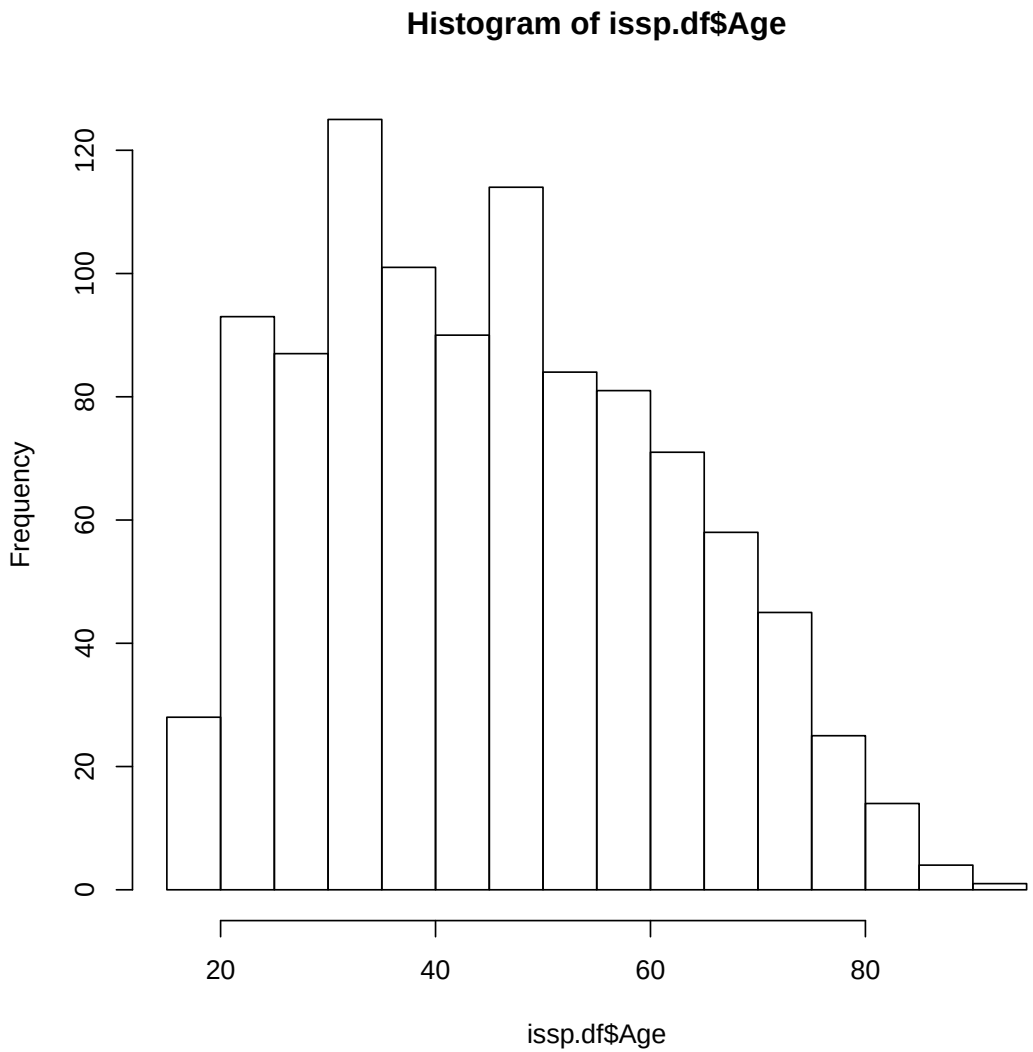
```
barplot(q5.age, beside = TRUE, ylab = "Percentage",
        xlab = "Age group in years", ylim = c(0, 100),
        space = c(0.2, 1.5), legend.text = TRUE,
        args.legend = list(bty = "n", cex = 1.3))
abline(h = 0)
```

Further improvements



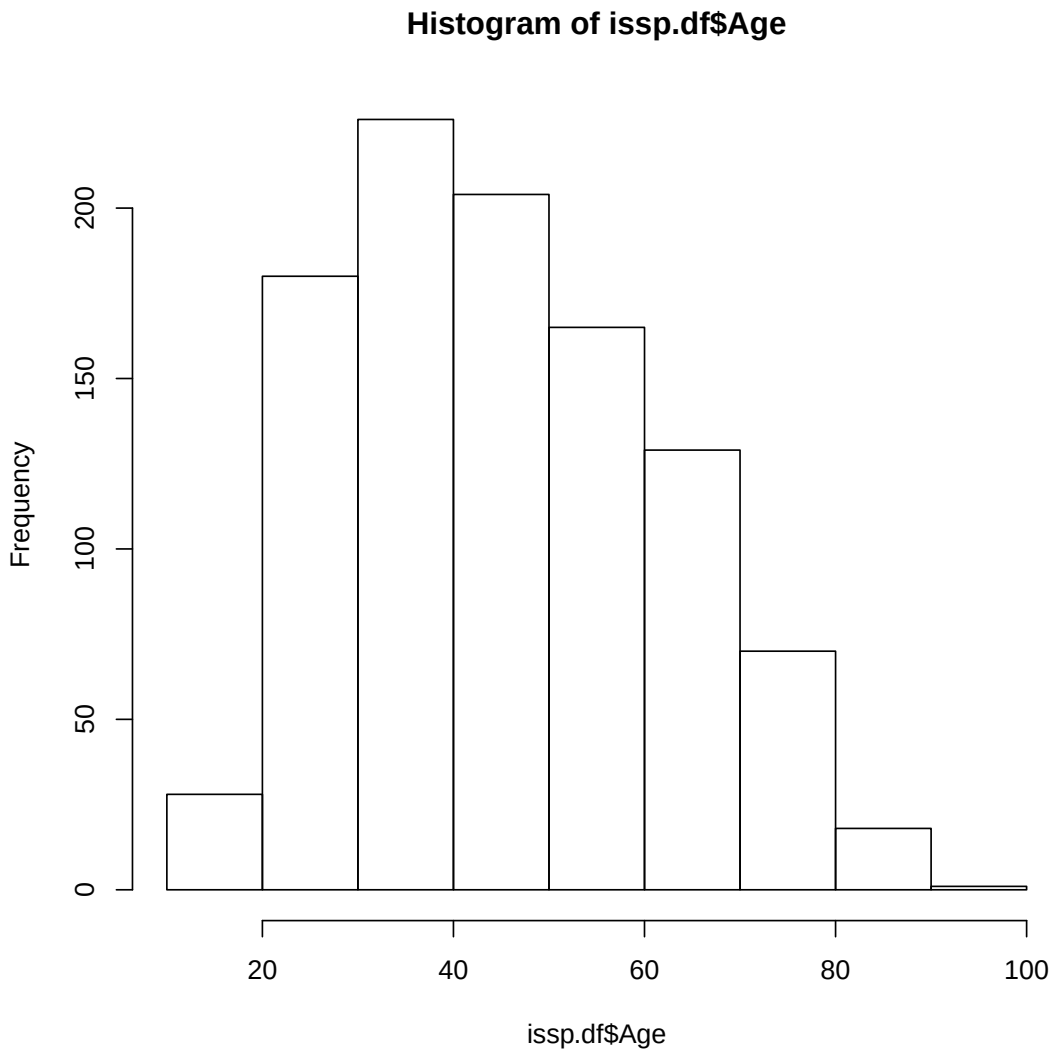
Histograms

```
hist(issp.df$Age)
```



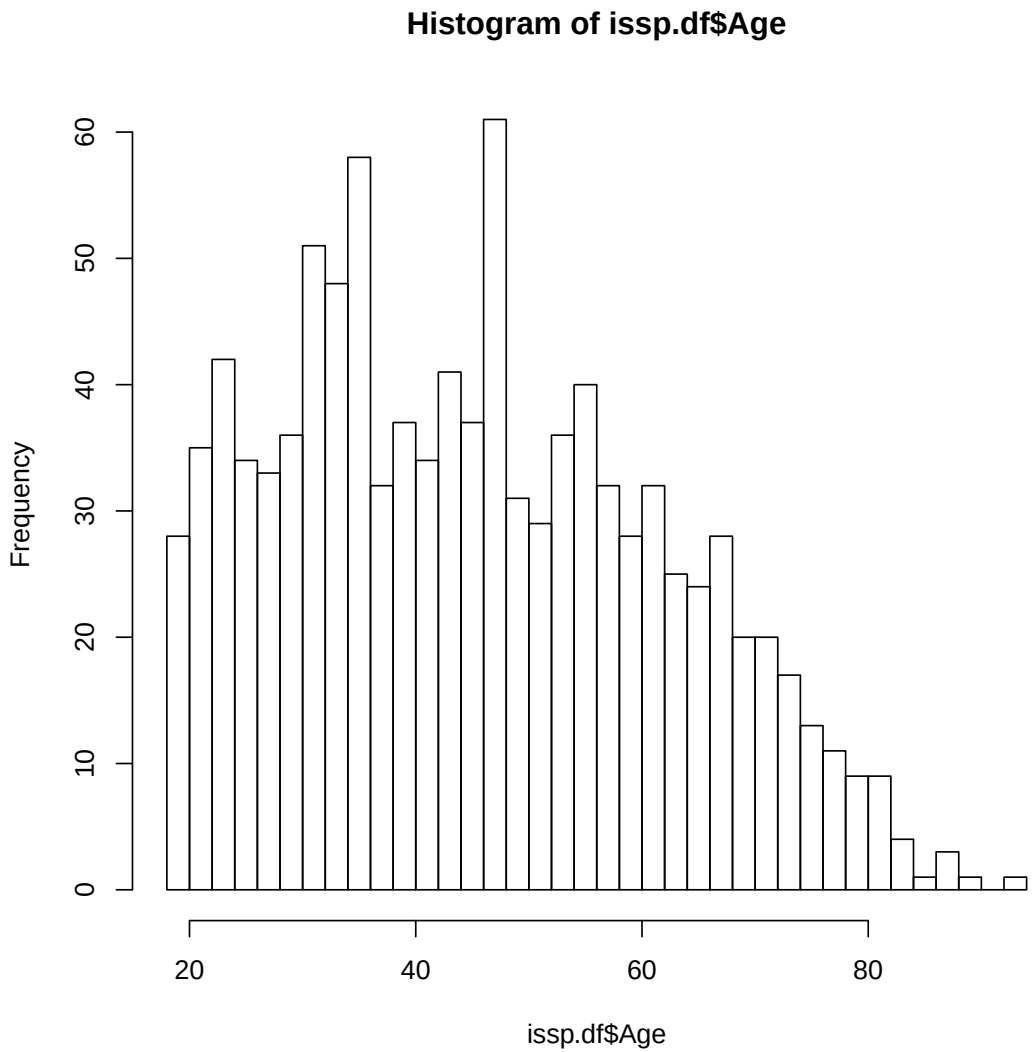
Fewer bins?

```
hist(issp.df$Age, breaks = 10)
```



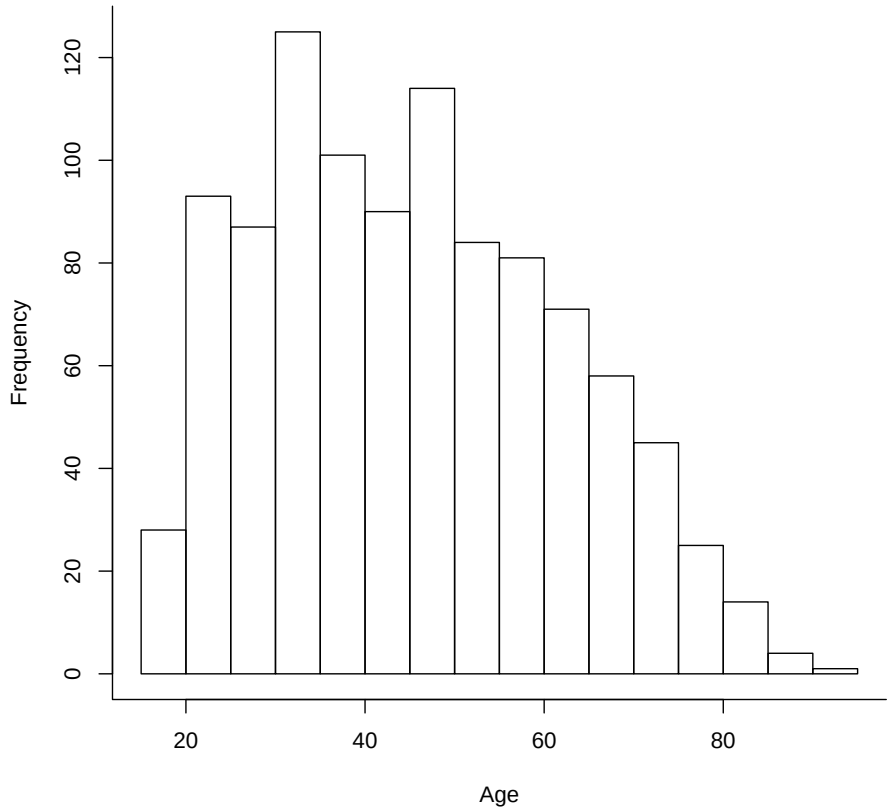
More bins?

```
hist(issp.df$Age, breaks = 40)
```



Always use meaningful axis labels

```
hist(issp.df$Age, xlab = "Age", main = NULL)
box(bty="l")
```

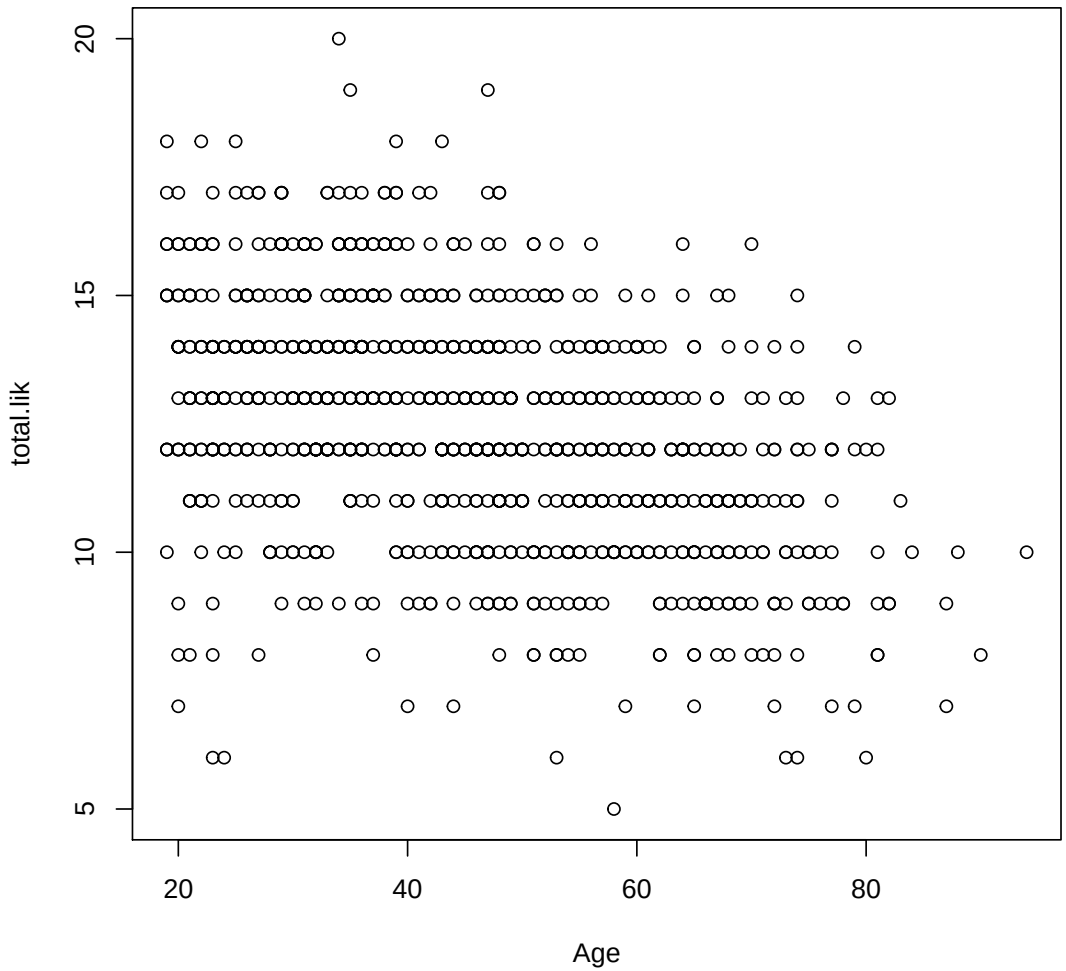


Scatter plots

- Not often used in the social sciences??
- Introduced briefly here.
- Used to display paired quantitative (numeric) data, e.g. age and score per respondent.
- Make sure they are numeric.
- Use `plot()` function, e.g.

```
with(issp.df, plot(Age, total.lik))
```

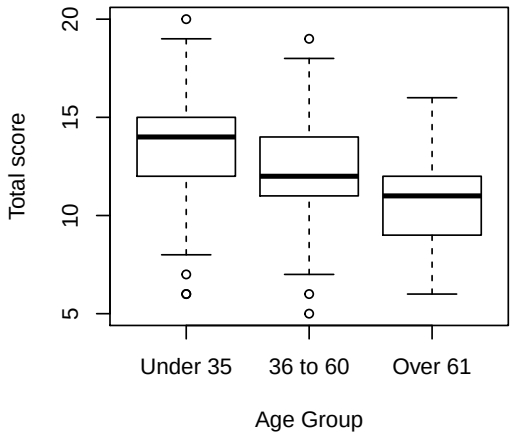
Scatter plots

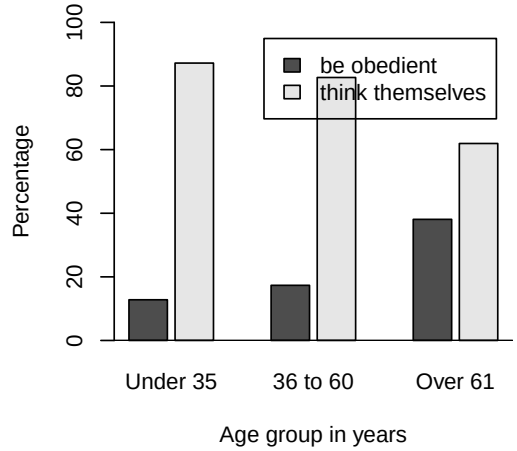


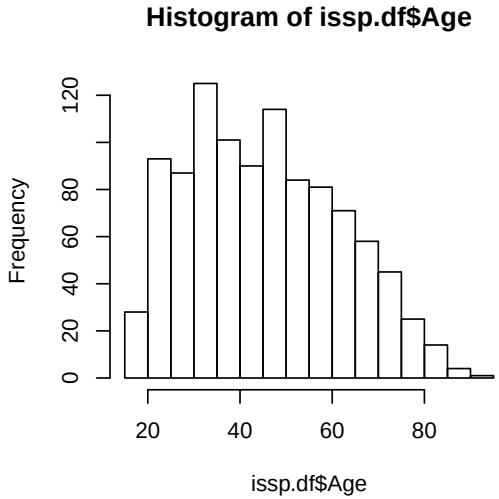
Graphical Parameters

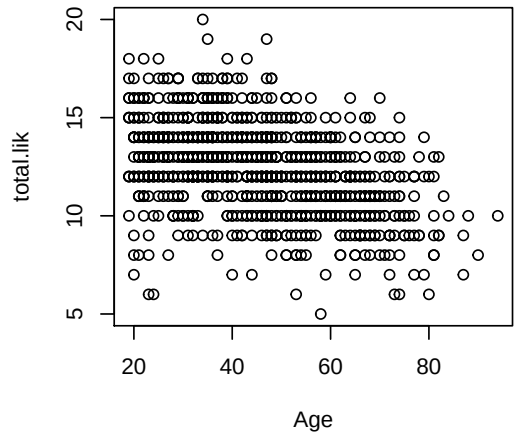
```
?par
op <- par(mfrow = c(2, 2))
  with(issp.df, boxplot(total.lik~age.group,
      xlab = "Age Group", ylab = "Total score"))
  barplot(q5.age, beside = TRUE, ylab = "Percentage",
      xlab = "Age group in years", ylim = c(0, 100),
      space = c(0.2, 1.5), legend.text = TRUE)
      abline(h = 0)
  hist(issp.df$Age)
  with(issp.df, plot(Age, total.lik))
## At end of plotting, reset to previous settings:
par(op)
```

Graphical Parameters









Summary

- Boxplots
- Histograms/Barplots
- Scatterplots