Statistics 120
Pie Charts, Bar Charts & Dot Charts

A Single Categorical Variable

- We often need to display a set of values each of which is associated with a single category of a factor or ordered factor.
- Most commonly the values are counts or proportions.
- Here is a typical example.

New Zealand Meat Consumption (1997)

Lamb	Mutton	Pigmeat	Poultry	Beef
8%	10%	16%	25%	41%

- Proportions are often presented in pie charts.

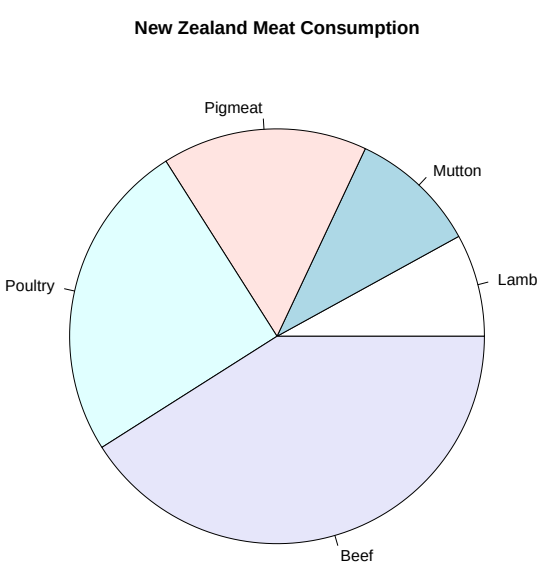
Producing a Pie Chart with R

A basic pie chart is produced from a vector of named values. such a vector can be created as follows:

```
> meat = c(8, 10, 16, 25, 41)
> names(meat) = c("Lamb", "Mutton", "Pigmeat",
                  "Poultry", "Beef")
```

Once the data vector is created, the plot is easy to create.

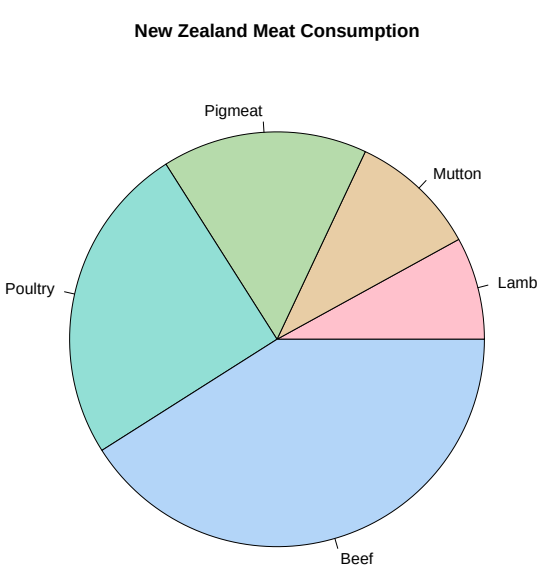
```
> pie(meat,
      main = "New Zealand Meat Consumption")
```


New Zealand Meat Consumption
Pigmeat
Mutton
Lamb
Poultry
Beef

Customizing Pie Chart Colours

Different colours can be specified for the pie slices with a `col=` argument. For example:

```
> pie(meat,
      main = "New Zealand Meat Consumption",
      col = hcl(seq(0, 240, by = 60)))
```

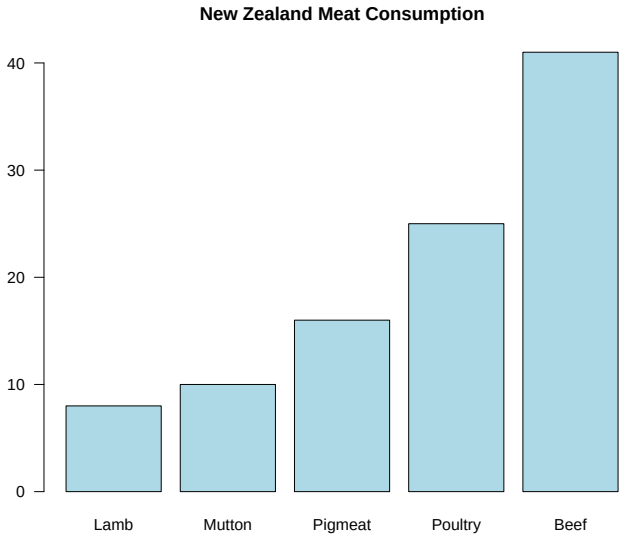

New Zealand Meat Consumption
Pigmeat
Mutton
Lamb
Poultry
Beef

Pie Charts

- Perceptual theory suggests that pie charts are not a good way of presenting information because they are based on angles or areas.
- Often pie charts are presented with the values as part of or close to the pie slice labels. This indicates that it is hard to decode the encoded values.
- Pie charts are only useful for proportions. They can’t be used for other kinds of values.

Bar Charts

- Bar charts provide a useful alternative way of presenting a set of values associated with the levels of a factor.
- Bar charts can be used for values other than proportions.
- Bar charts use position on a common scale to encode their values.
- This means they provide a better way of delivering information than pie charts.


New Zealand Meat Consumption
40
30
20
10
0
Lamb
Mutton
Pigmeat
Poultry
Beef

Producing a Bar Chart with R

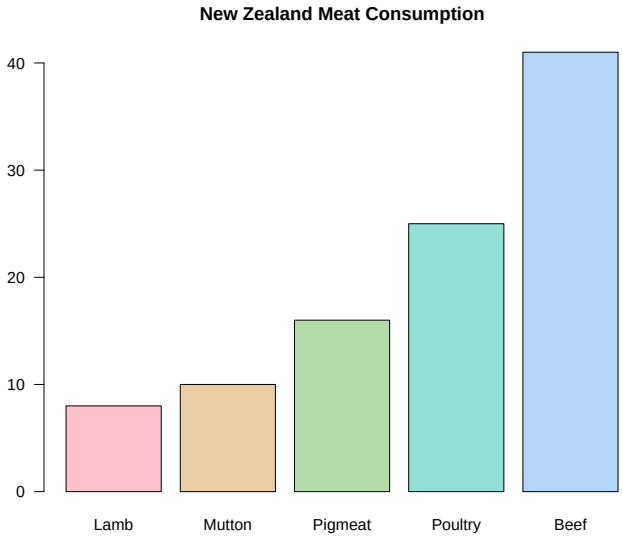
A basic bar chart is produced as follows:

```
> barplot(meat, col = "lightblue",
          main = "New Zealand Meat Consumption")
```

In pie charts it is very important to colour the slices different colours. In bar charts it is less important, but it can still be done.

```
> barplot(meat,
          main = "New Zealand Meat Consumption",
          col = hcl(h = seq(0, 240, by = 60)),
          las = 1)
```

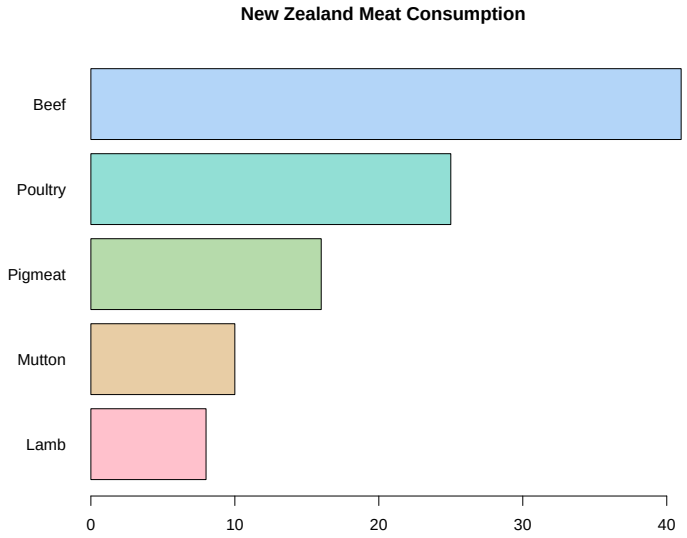
(Note the use of `las=1` here to rotate the *y* axis labels.)


New Zealand Meat Consumption
40
30
20
10
0
Lamb
Mutton
Pigmeat
Poultry
Beef

Horizontal Bars

It can be useful to draw the bars of bar-chart horizontally. In R, this is done by specifying `horiz=TRUE`.

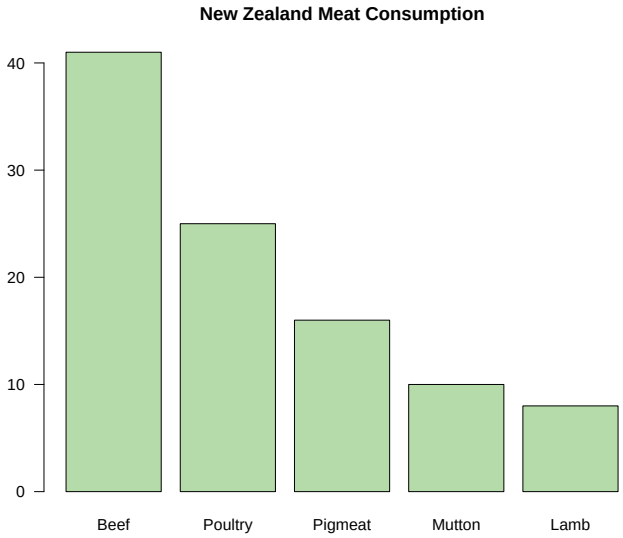
```
> barplot(meat,
          main = "New Zealand Meat Consumption",
          col = hcl(seq(0, 240, by = 60)),
          horiz = TRUE,
          las = 1)
```


New Zealand Meat Consumption
Beef
Poultry
Pigmeat
Mutton
Lamb
0
10
20
30
40

Pareto Charts

- Pareto charts are a bar plot variation which are used in quality control work.
- A Pareto chart sorts the bars of a barplot from biggest to smallest, left to right across the graphs.
- They are typically used to determine an importance ordering of faults detected in a manufacturing process.
- Pareto charts can be produced in R using the function `sort`.

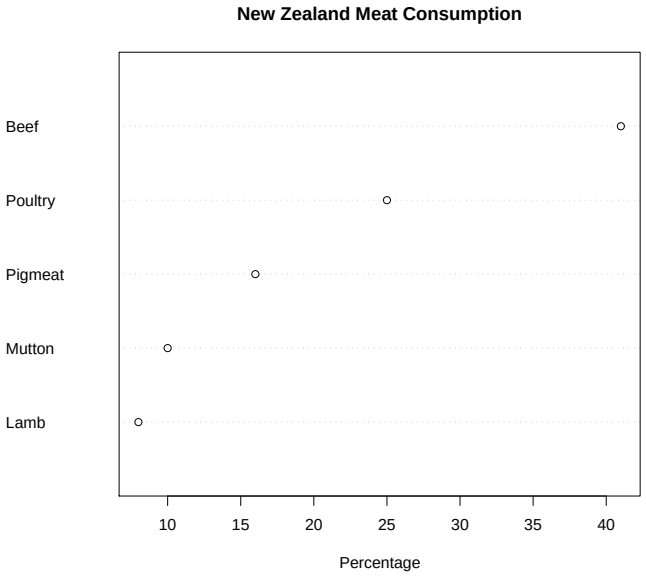
```
> barplot(sort(meat, decreasing = TRUE),
      main = "New Zealand Meat Consumption",
      col = hcl(120), las = 1)
```


New Zealand Meat Consumption
40
30
20
10
0
Beef
Poultry
Pigmeat
Mutton
Lamb

Dot Charts

- Dot charts are very similar to a horizontal bar chart.
- The bars are replaced by a single “dot” and a horizontal grid line.
- Dot charts have the advantage that many more bars can be plotted horizontally.
- In R, dotcharts are easy to produce.

```
> dotchart(meat,
      main = "New Zealand Meat Consumption",
      xlab = "Percentage")
```


New Zealand Meat Consumption
Beef
Poultry
Pigmeat
Mutton
Lamb
10
15
20
25
30
35
40
Percentage

Two Category Bar Charts

Here is a set of data which is *cross-classified* by two factors.

Percentage of the New Zealand Population
With a Potentially Hazardous Drinking Pattern

	Age			
Sex	*15-24*	25-*44*	*45-64*	*65+*
Male	41	28	21	9
Female	26	9	4	.5

This is exactly the form of data produced by `table`.
It is also possible to input this data directly as a matrix.

Data for Two Category Bar Charts

```
> alcohol = matrix(c(41, 26,
      28, 9, 21, 4, 9, 0.5),
      nr = 2)
> dimnames(alcohol) = list(Sex = c("Male",
      "Female"), Age = c("15-24",
      "25-44", "45-64", "65+"))
> alcohol
```

```
        Age
Sex      15-24 25-44 45-64 65+
  Male      41    28    21 9.0
  Female    26     9     4 0.5
```

This is a 2×4 matrix with row and column labels.

Bar Chart Layout

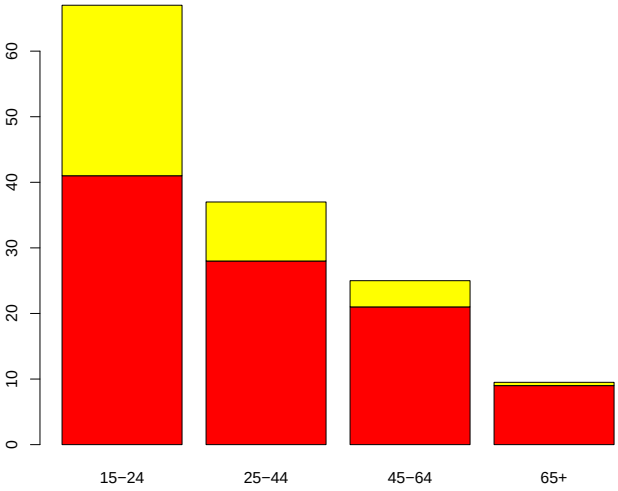
There are two ways in which a barchart for cross-classified variables can be produced:

- stacked bars.
- side-by-side bars.

Both of these layouts are possible in R.

The simplest bar chart command is:

```
> barplot(alcohol)
```


0
10
20
30
40
50
60
15−24
25−44
45−64
65+

Improving the Default Layout

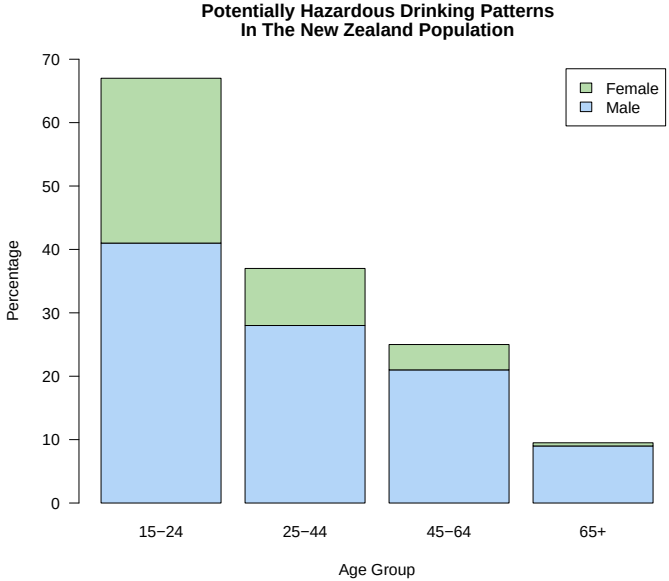
- There are clear problems with default barplot.
- The colours are awful!
- The plot needs a *legend* explaining what the two colours mean.
- The plot needs an overall title and labels for the x and y axes.
- The tick mark labels should be horizontal.
- The tick marks should span the full height of the bars.

Improving the Default Layout

```
> main = c("Potentially Hazardous Drinking Patterns",
           "In The New Zealand Population")

> barplot(alcohol,
          legend = rownames(alcohol),
          col = hcl(c(240, 120)),
          ylim = c(0, 70),
          las = 1,
          main = main,
          xlab = "Age Group",
          ylab ="Percentage")
```

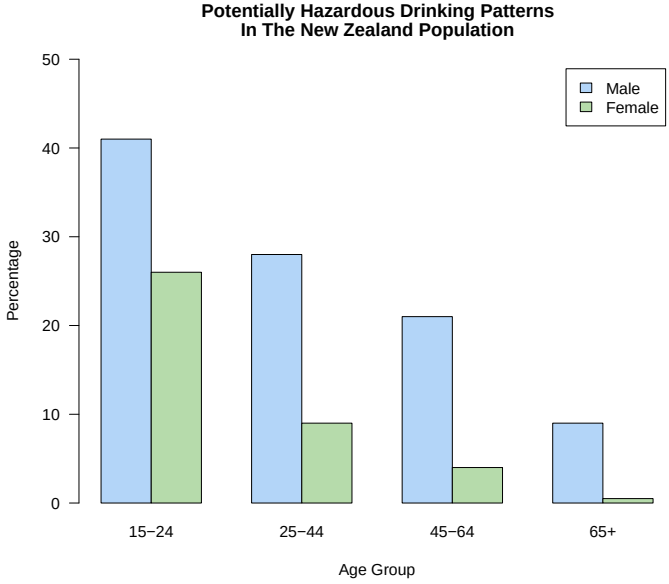
The legend could also be added with a separate `legend` statement. This is more flexible.


Potentially Hazardous Drinking Patterns
In The New Zealand Population
Female
Male
Percentage
70
60
50
40
30
20
10
0
15−24
25−44
45−64
65+
Age Group

Side-by-Side Bars

Instead of drawing the “stacked” form of bar chart, it is also possible to produce a “side-by-side” form.

```
> barplot(alcohol, beside = TRUE,
          legend = rownames(alcohol),
          col = hcl(c(240, 120)),
          ylim = c(0, 50),
          las = 1,
          main = main,
          xlab = "Age Group",
          ylab ="Percentage")
```


Potentially Hazardous Drinking Patterns
In The New Zealand Population
Male
Female
Percentage
0
10
20
30
40
50
15−24
25−44
45−64
65+
Age Group

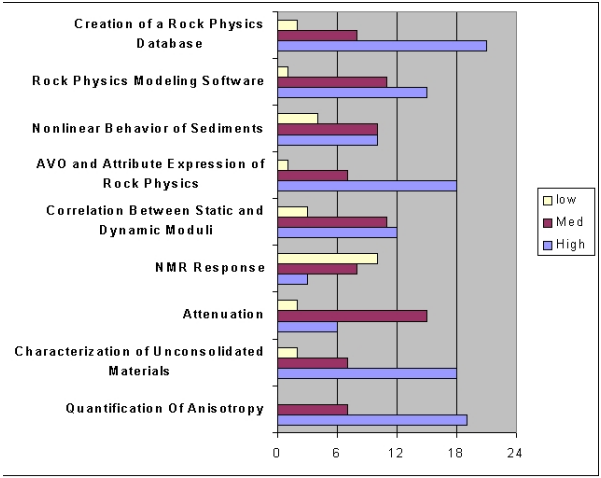
Choosing a Layout

- The side-by-side bar plot variant encodes all its values as positions on a common scale (i.e. the position of the top of bar).
- The stacked bar plot variant encodes some of its information in the lengths of the stacked bars.
- Perceptual theory tells us that the side-by-side version will work best.
- Occasionally the stacked version is useful — this is when the combined height of the bars is meaningful.

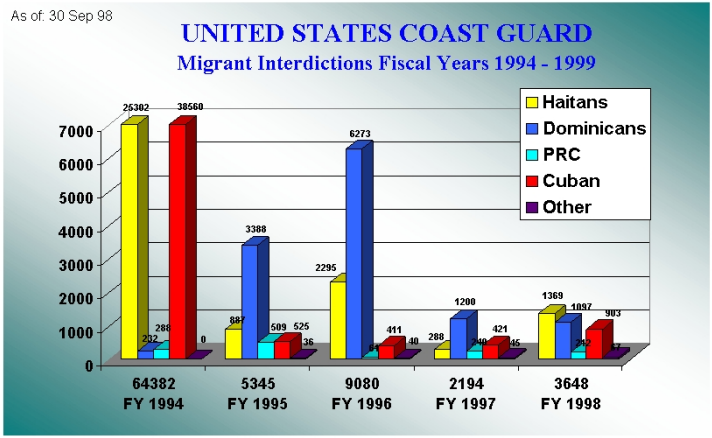
Some Examples

- There are thousands of examples of bar charts on the Web.
- To see a sample of them go to Google Images and search for “bar chart.”
- Here a few examples.

Student Rankings

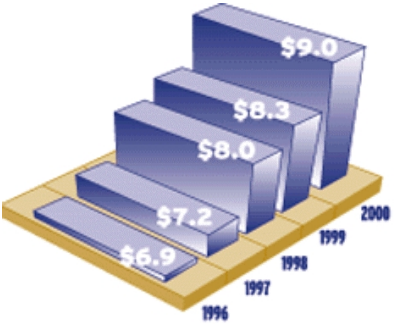


This is the basic bar chart produced by Excel.



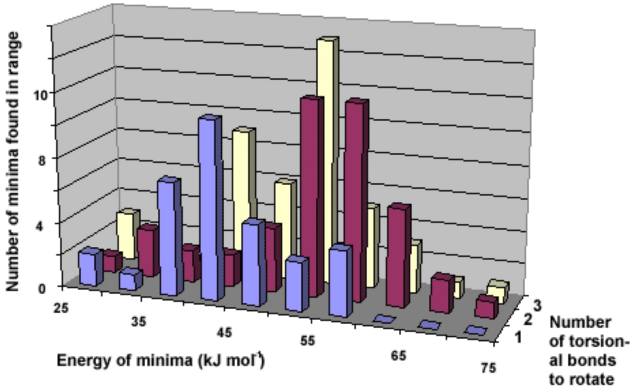
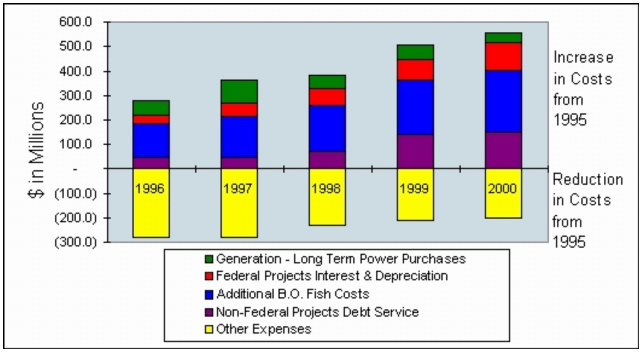
Gradient Background. Three dimensional bars.

Tourism-Related Gross Receipts/Sales
1996-2000 in Billions



Gross sales generated by the Minnesota tourism industry grew 31% from 1996 to 2000. Adjusted for inflation, this represents a real average annual growth rate of almost 4%.

A Two Dimensional Bar Chart



A Barplot Variation