slide 1
Exploratory Graphics: Topics

- One variable
  - Aim: explore distribution of values
  - Plots: Histograms, Kernel density estimators, QQ plots

- Two variables
  - Aim: explore relationship of variables
  - Both continuous: Scatter plot
  - One of each: Side-by-side box plots or violin plots
  - Both categorical: Mosaic plots (see Chapter 5)

- Three or more variables
  - Aim: Identify pairs of relationships, visualise GLMs
  - Plots: Pairs plots, Rotating plots, coplots, 3D plots, contour plots
Data: daily changes in log(exchange rate) for USD/NZD
- Daily data from June 1986 to May 2014
- Source: Reserve Bank

Questions:
- What is the distribution of the daily changes in the logged exchange rate?
- Is it normal? If not, how is it different?
\[ \delta_t = \log(y_t) - \log(y_{t-1}) = \log \left( \frac{y_t}{y_{t-1}} \right) \]

Suppose we have the data \((\delta_t, t = 2, \ldots, 6980)\), in an \(\mathbb{R}\) vector, diff.in.logs

```r
# Draw histogram
hist(diff.in.logs, nclass=100, freq=FALSE)

# Add density estimates
lines(density(diff.in.logs), col="blue", lwd=2)

# Add fitted normal density
xvec <- seq(-0.2, 0.1, length=100)
lines(xvec, dnorm(xvec, mean=mean(diff.in.logs),
                sd=sd(diff.in.logs)), col="red", lwd=2)
```
slide 8
slide 8
Normal QQ Plot

- Normal data? NO
- QQ plot indicates that the differences have longer tails than normal
- Plotted points are below line for smaller and above line for larger values.
Two variables: Rats!

Given: growth rates of 16 rats, i.e. relationship between weight and time.
- Want to explore the relationship graphically.
- Each rat was measured (roughly) every week for 10 weeks.
- For weeks 1–5, all rats were on a fixed diet.
- Diet was changed after week 6.
Two variables: Rats!

Dataset rats.df has variables
growth weight in grams
group litter, labelled 1–3
rat individual rat, labelled 1–16
change labelled 1–2. Diet was changed after 6 weeks, diet 1 for weeks 1–5, diet 2 for weeks 6–10.
day: day since start of study, 11 values, approximately weekly intervals.
Rats! Simple visualisation
Rats! More sophisticated visualisation

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>Weight (grams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>200</td>
</tr>
<tr>
<td>20</td>
<td>240</td>
</tr>
<tr>
<td>40</td>
<td>280</td>
</tr>
<tr>
<td>60</td>
<td>320</td>
</tr>
</tbody>
</table>

Growth rates for rats

Litter 1
Litter 2
Litter 3
### Rats! And more sophistication

<table>
<thead>
<tr>
<th>Time (days)</th>
<th>Weight (grams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>400</td>
<td>400</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>600</td>
<td>600</td>
</tr>
</tbody>
</table>

**Slide 16**
Rats! And more sophistication

slide 16
One continuous and one categorical variable

- Measurement of mouse body temperature every 15 minutes over duration of 25 days
- Wish to visualise relationship between day time (categorical) and body temperature (continuous)
- Side-by-side boxplots
The fever of mice

slide 20
More than two variables

- If all variables are continuous, we can explore the relationships between them using a pairs plot.
- If we have three variables, a rotating plot is a very useful tool.
- Example: Cherry trees.
Pairs Plot for Cherry Trees

slide 23
The challenge: to represent a 3-dimensional object on a 2-dimensional surface (current screen types)

Traditional method uses projection, perspective

A powerful idea is to use motion, looking at the 3D scene from different angles
Projection

slide 26
By dynamically changing the angle of view, we get a better impression of the 3-dimensional structure of the data. "Dynamic graphics" is a very powerful tool.
A powerful idea: Coplots

- Coplots show relationship between $x$ and $y$ for selected values of $z$ (usually a narrow range of $z$).
- By showing separate plots for different $z$ ranges, we can see how the relationship between $x$ and $y$ changes as $z$ changes.
- Coplot: conditioning plot, shows relationship between $x$ and $y$ conditional on $z$ (i.e., for fixed $z$).
To show the relationship between height and volume for different values of diameter:

- Divide the range of diameter (8.3 to 20.6) up into 6 subranges: 8 − 11, 10.5 − 11.5 etc.
- Draw 6 plots, the first using all data whose diameter is between 8 and 11, the second using all data whose diameter is between 10.5 and 11.5, and so on.
Cherry trees: coplots