

## Wickham's version (*ggplot2*)

(DATA and TRANS are already handled within R.)

- The layered grammar is a combination of
  - default dataset
  - set of mappings from variables to aesthetics
  - one of more layers composed of a geom and a stat
    - (+ position adjustment and a dataset and mappings)
  - one scale for each aesthetic mapping
  - a coordinate system
  - the faceting specification

## Scatterplots (*base graphics*)

```
data(movies)
plot(movies$votes, movies$rating)
plot(movies$votes, movies$rating, pch=20)
abline(h=9)
plot(movies$votes, movies
      $rating, pch=20, xlim=c(10000, 160000))
```

Lots of flexibility through parameters.  
Parameters are a rather unstructured list.  
Parameters can be amended and commands reentered.  
Limited layering.

## Scatterplots (*qplot*)

```
q1<-qplot(votes, rating, data=movies)
q2<-q1+geom_hline(yintercept=9)
q3<-q2+opts(title="MOVIES")
q4<-q3+facet_grid(. ~ mpaa)
```

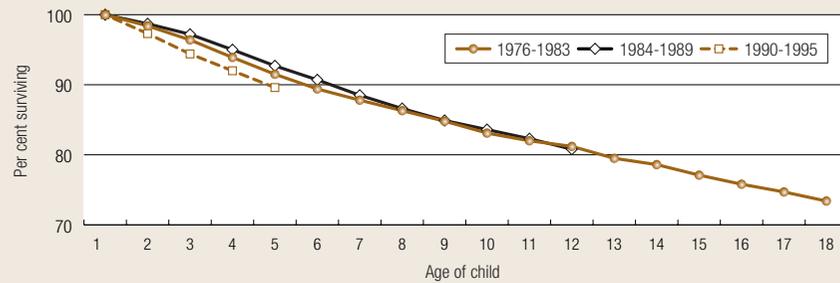
Primarily designed for using default parameter choices.  
Parameters can be amended and added.  
Defining plots as objects works well.  
Layering works.

## Scatterplots (*ggplot2*)

```
p1<-ggplot(movies, aes(x=votes, y=rating))
p1+geom_point()
p2<-p1+geom_point(size=1)
p3<-p2+geom_hline(yintercept=9)
p4<-p3+opts(title="MOVIES")
p5<-p4+facet_grid(. ~ mpaa)
```

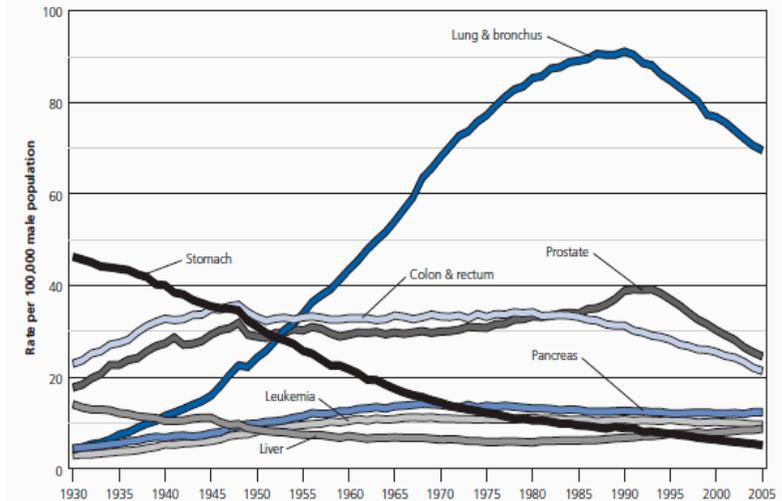
Users must specify what they want.  
Some parameter combinations produce rubbish.  
Parameters are well structured.  
Defining plots as objects is (almost) essential.  
Layering works.

Figure 15.11 Survival of parents' relationship by child's age by birth cohort, 1976-2000



Source: HILDA, 2001 (FaCS 2002a).  
Figure reproduced from de Vaus and Gray (2003)

Age-adjusted Cancer Death Rates,\* Males by Site, US, 1930-2005



\*Per 100,000, age adjusted to the 2000 US standard population.

Note: Due to changes in ICD coding, numerator information has changed over time. Rates for cancer of the liver, lung and bronchus, and colon and rectum are affected by these coding changes.

Source: US Mortality Data, 1960 to 2005, US Mortality Volumes, 1930 to 1959, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.

American Cancer Society, Surveillance and Health Policy Research, 2009

## Multiple plots (1)

- Small multiples are available using faceting.

```
p6<-p4+facet_grid(. ~ Comedy+Action+Romance)
p7<-p4+facet_grid(Comedy ~ Action+Romance)
```

- Conditioning variables to the left are vertical and to the right are horizontal. A point stop is used as a placeholder if no variable appears vertically /horizontally.
- Multiple plots on one page are not elegantly supported, though you do have flexibility.

## Multiple plots (2)

- Multiple plots in *ggplot2*

```
dev.new(width = 1250, height = 600)
grid.newpage()
pushViewport(viewport(layout = grid.layout(2,5)))
vplayout <- function(x, y)
{viewport(layout.pos.row = x, layout.pos.col = y)}

print(bm, vp = vplayout(1, 1))
print(bf, vp = vplayout(2, 1))
print(mf, vp = vplayout(1:2, 2))
print(fm, vp = vplayout(1:2,3))
print(fluc, vp = vplayout(1:2,4:5))
```

## Parallel coordinate plots

- Parallel coordinate plots are constructed with a special command (`ggpcp`), which first stacks the data into a new form.

```
g1<-ggpcp(mtcars) + geom_line()
str(g1)
ggpcp(mtcars, scale="var") + geom_line()
ggpcp(mtcars, scale="range") +
geom_boxplot(aes(group=variable))
```

- Parallel coordinate plots using a subset of the variables are stacked independently.

```
g2<-ggpcp(mtcars, vars=names(mtcars[2:6])) +
geom_line()
str(g2)
```

## Mosaicplots

- Mosaicplots are not yet available in `ggplot2`
  - multiple barcharts are available through facetting

```
(g1<-ggplot(movies, aes(factor(Action)))+geom_bar())
(g2<-g1+facet_grid(.~mpaa))
(g3<-g1+facet_grid(mpaa~.))
```

- there is a special command for fluctuation diagrams, though only as yet for 2-d

```
ggfluctuation(table(movies$Action, movies$Comedy))
ggfluctuation(table(movies$Action, movies$Comedy),
type="colour")
```

## Mosaicplots and faceting?

- The option `scales="free"` allows each facet to have its own scale (subject to row and column restrictions for `facet_grid`).
- There are restricted versions `scales="free_x"` and `scales="free_y"`.
- `facet_grid` has an additional subsidiary parameter `space="free"`, which allocates rows and columns space according to the range of their scales, e.g.,

```
qplot(cty, model, data=mpg) +
facet_grid(manufacturer ~ ., scales = "free",
space = "free")
```
- How is this related to mosaicplots?

## Wilkinson and mosaicplots

- Wilkinson (p343 2nd edition) suggests
  - graphics with equal-sized tiles coloured by the dependent variable
  - classical mosaicplots built up on his graph algebra of the operators *cross*, *nest* and *blend*
  - e.g. for the Titanic dataset `1*1*age*1*1*sex*class` with the cell rectangles colored by the variable *survival*