

Grammar of Graphics

- is not
 - a command language
 - a taxonomy
 - a drawing package
 - a book of virtues
 - a heuristic system
 - a GIS
 - a visualization system

Wilkinson's 7 Components

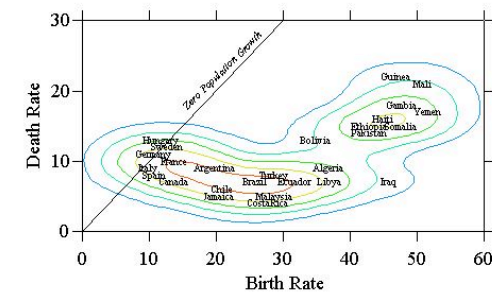
- DATA — variables from datasets
 - TRANSformations — variable transformations
 - FRAME — a set of variables, combined with operators, defining a space
 - SCALE — scale transformations
 - COORDinates — a coordinate system
 - GRAPH — graphs and their aesthetic attributes
 - GUIDE — axes, legends
- (FRAME and GRAPH are always necessary, the others not)

Three Phases

- Specification
 - the seven components
- Assembly
 - modifiable (or not)
 - dynamic or static
 - linked or separate
 - interactive (or not)
- Display
 - Rendering
 - Interface for interactivity

Example: Figure 1.1

Leland Wilkinson, *The Grammar of Graphics*, Springer-Verlag, 1999, Figure 1.1



```
FRAME: birth*death
GRAPH: point(size(0), label(country))
GRAPH: contour.density.kernel.epanechnikov:joint(color.hue())
GUIDE: form.line(position((0,0),(30,30)), label("Zero Population Growth"))
GUIDE: axis1(label("Birth Rate"))
GUIDE: axis2(label("Death Rate"))
```

Example: Figure 1.1 Notes

- Specification
 - Data in a data frame, so no DATA component
 - Data already as rates, so no TRANS component
 - Default scales, so no SCALE component
 - Default coordinates, so no COORD component
- Comments
 - Some country names are plotted, why those ones?
 - No colour legend

Wilkinson's recipe

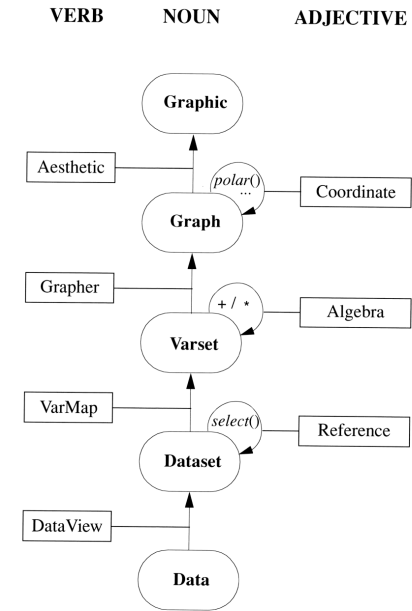
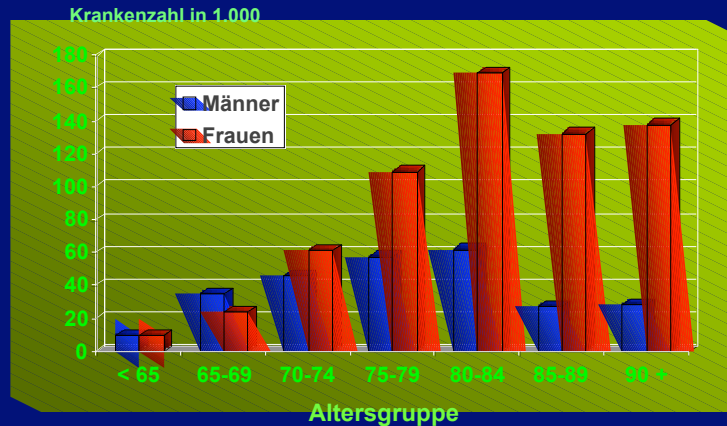


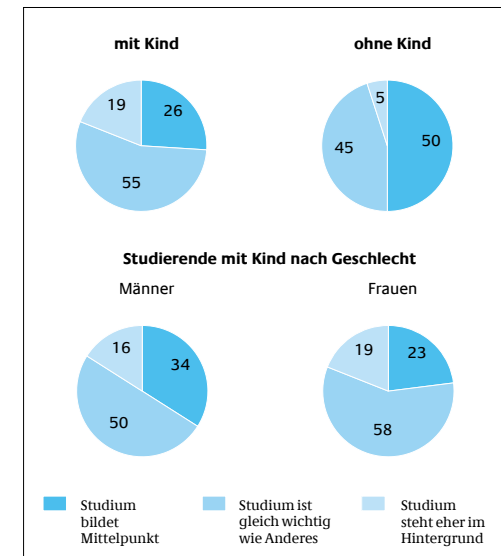
Figure 2.1 From data to graphic

Geschätzte Verteilung der Demenzkranken in Deutschland zum Ende des Jahres 2002 nach Geschlecht und Alter *



* EURODEM-Daten; Lobo et al. (2000) Neurology 54, Suppl. 5: 4-9

Bild 1.12 Zentralität des Studiums – Vergleich Studierender mit und ohne Kind
Studierende im Erststudium, in %



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

Transforming

- transform the data
 - `qplot(log10(price), data=diamonds)`
- transform the scale
 - `qplot(price, data=diamonds) + scale_x_log10()`
- transform the coordinate system
 - `qplot(price, data=diamonds) + coord_trans(x="log10")`

(Except that in this case it doesn't work because of zero.)

ggplot2 (had.co.nz/ggplot2/)

- Only works on dataframes
- Scales map data to aesthetics
- aesthetic mappings include
 - x, y
 - size, shape
 - colour, fill
 - group
 - weight, alpha
- Guides (axes, legends) map aesthetics to data

ggplot2 2

- *geoms* may be individual or collective
- Every *geom* has a default *statistic* and vice versa
- *qplot* is a quick form of *ggplot*
 - has mostly excellent defaults
 - layers can be added to a *qplot*
 - but
 - uses "data=" instead of "dataset" as first parameter
- `xlim` and `ylim` are not quite like in *base* and *lattice*
 - within `scale` they exclude data outside the limits
 - within `coord` they only visually exclude off-limits data