
Graphs and Tables

Purposes

- Exploratory — intended to help your analysis
- Presentation — functions as a paragraph of your report

If you have a graph or table in your report you should be able to explain why it is there and what point it makes. It's important that tables and graphs are designed to be readable.

Tables

- “Tables should not be set to look like nets with every number enclosed”
- “Vertical rules are needed only when the space between columns is so narrow that mistakes will occur in reading without rules”

Tischold *Asymmetric Typography*, quoted in Tufte *Envisioning Information*

Bad table

	M1	M2	M3	F1	F2	F3
AGEBL	72.37	74.12	73.65	71.79	73.08	73.76
EVERSM	0.67	0.67	0.7	0.43	0.42	0.46
PKYRS	23.71	24.54	29.75	11.81	11.34	12.86
alcoh	3.88	3.56	2.68	1.55	1.57	0.91
BPHI	0.4	0.42	0.51	0.42	0.51	0.66
diabetes	0.14	0.2	0.25	0.1	0.15	0.21
IFG	0.16	0.17	0.12	0.11	0.14	0.14
bmi	26.45	26.42	26.31	26.49	27.23	27.31
AVZMSYS	135.18	136.56	134.12	134.98	136.58	139.44
AVZMDIA	72.87	72.15	69.79	69.97	69.04	68.23

M1-healthy men, M2 -impaired men, M2 - CVD women, F1-healthy women, M2 -impaired women, M2 - CVD women

Better table

	Men			Women		
	Healthy	Impaired	CVD	Healthy	Impaired	CVD
Age	72	74	74	73	73	74
Ever smoked	67%	67%	70%	43%	42 %	46%
Pack-years	24	25	30	12	11	13
Drinks/wk	3.9	3.6	2.7	1.6	1.6	0.9
Hypertension	40%	42%	51%	42%	51%	66%
Diabetes	14%	20%	25%	10%	15%	21%
Glucose intolerance	16%	17%	12%	11%	14%	14%
BMI	26	26	26	26	27	27
Systolic BP	135	137	134	135	137	139
Diastolic BP	73	72	70	70	69	68

3.1415926535897932384626

- It is rarely worth having more than two significant digits in a table.
- Insignificant digits are even less worthwhile — if a regression coefficient is 0.00001 then you should probably change the units on the variable.
- p-values should be given to two or three digits: 0.35, 0.03, 0.01, <0.01 (not 0.00)

Graphs

A lot is known about graph useability: eg books by Tufte, Cleveland, Wilkinson

- continuous vs continuous: Scatterplots (lines/smoothers usually helpful)
- discrete vs continuous: barplots (not stacked, not piecharts), line plots, boxplots
- discrete vs discrete: this is hard. Consider four-fold plots, mosaic plots, barplots
- three-dimensional: levelplots, contour plots

S-PLUS makes it easy to add lines, points, text, legend, choose colors, customise axis labels, etc.

Colour

Colour is very useful for making extra distinctions in a graph and for emphasizing information. Consider

- Type of output: screen, projector, colour printer
- Need for photocopying to grayscale?
- Red/green colour blindness (about 10% of men)
- Cultural associations (white/death, red/happiness)

For points or lines use fairly saturated colors, for areas use light, unsaturated colors (www.colorbrewer.org)

Conditioning plots

- Plot X vs Y for different values of Z
- Illustrates confounding or interaction
- `coplot()`, the Trellis system.