Statistical Graphics:

Non Data Components of Graphs

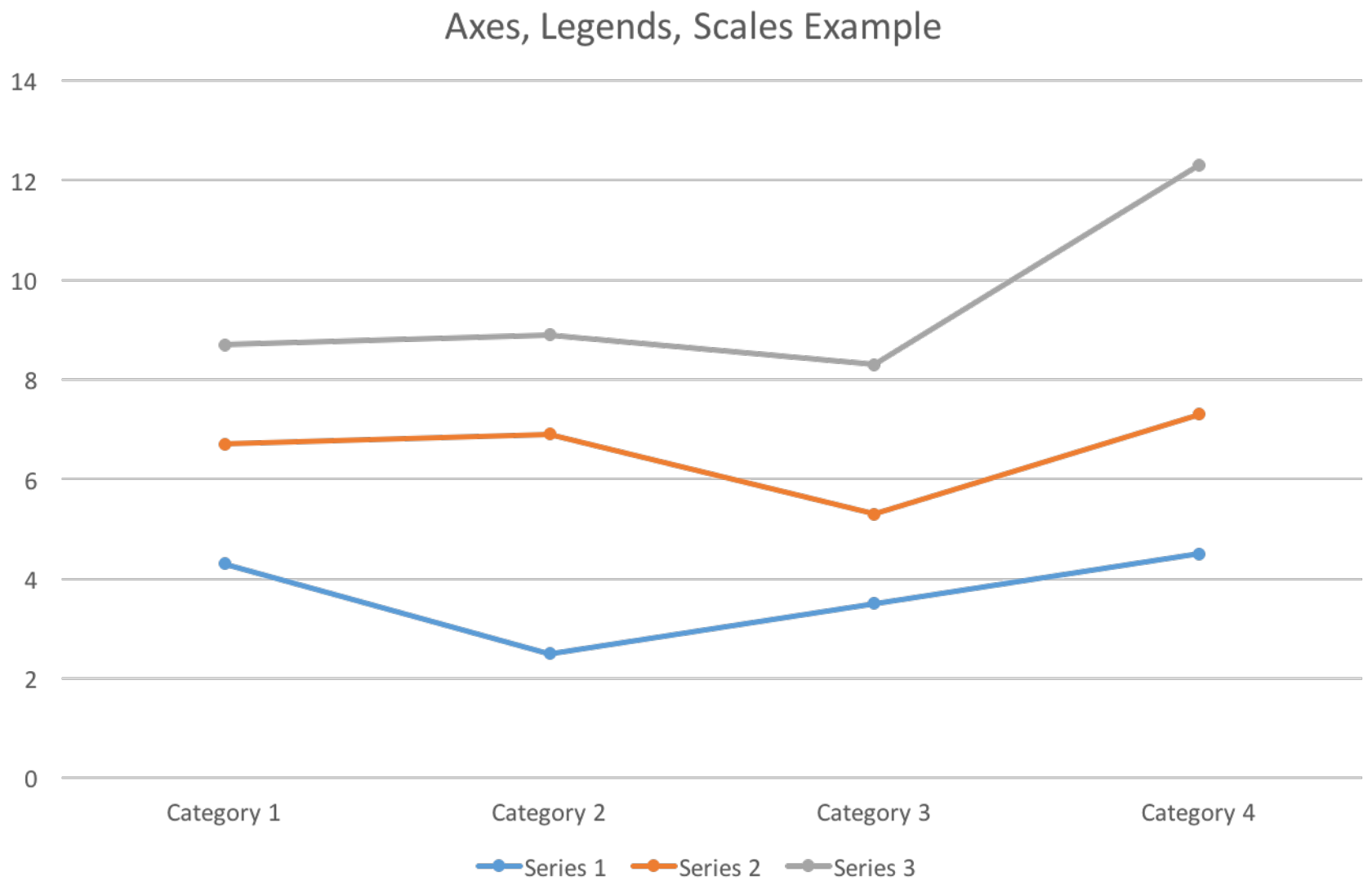
Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Define design principles for bar charts and line charts

Axes, Legends and Scales



Heckbert’s Labeling Algorithm

Data range

105 - 543

Data range

2.03-2.17

Nice Numbers

```
const ntick ← 5;                          desired number of tick marks

loose_label: label the data range from min to max loosely.
   (tight method is similar)

procedure loose_label(min, max: real);
nfrac: int;
d: real;                                  tick mark spacing
graphmin, graphmax: real;                 graph range min and max
range, x: real;
begin
   range ← nicenum(max − min, false);
   d ← nicenum(range/(ntick − 1), true);
   graphmin ← floor(min/d)*d;
   graphmax ← ceiling(max/d)*d;
   nfrac ← max( − floor(log10(d)), 0);    number of fractional digits to show

   for x ← graphmin to graphmax + .5*d step d do
      put tick mark at x, with a numerical label showing nfrac fraction digits
      endloop;
   endproc loose_label;

nicenum: find a "nice" number approximately equal to x.
Round the number if round = true, take ceiling if round = false.

function nicenum(x: real; round: boolean): real;
exp: int;                                 exponent of x
f: real;                                  fractional part of x
nf: real;                                 nice, rounded fraction
begin
   exp ← floor(log10(x));
   f ← x/expt(10., exp);                  between 1 and 10

if round then
   if f < 1.5 then nf ← 1.;
   else if f < 3. then nf ← 2.;
   else if f < 7. then nf ← 5.;
   else nf ← 10.;
else
   if f ≤ 1. then nf ← 1.;
   else if f ≤ 2. then nf ← 2.;
   else if f ≤ 5. then nf ← 5.;
   else nf ← 10.;
return nf*expt(10., exp);
endfunc nicenum;
```

P. Heckbert. Nice numbers for graph labels. In A. Glassner, editor, Graphics Gems, pages 61–63 657–659. Academic Press, Boston, 1990.

Heckbert's Labeling Algorithm

Problem

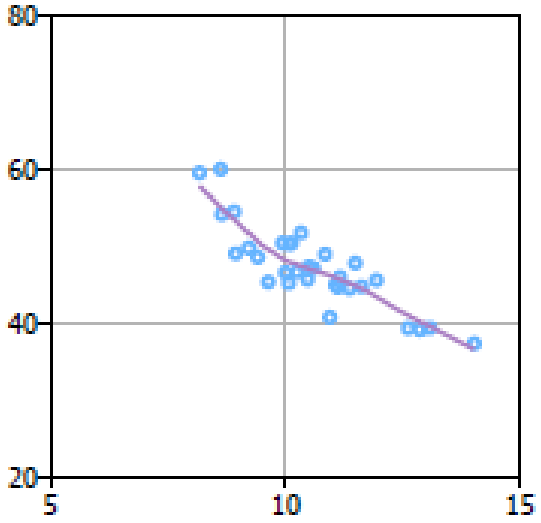
For small numbers, the range of labels can be much larger than the data range.

Solution

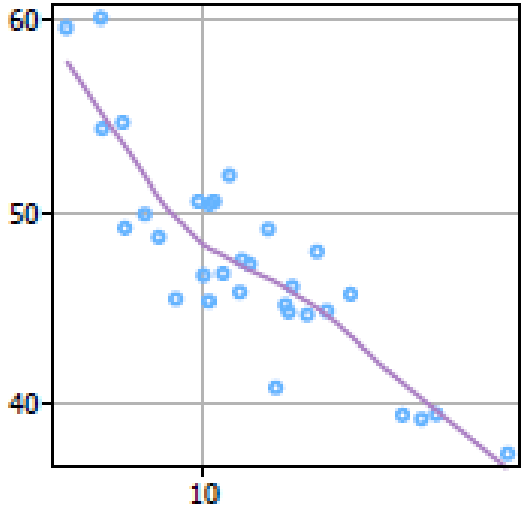
Drop labels which overlap or fall outside the data range

This leads to unevenly spaced labels or axes with only one label

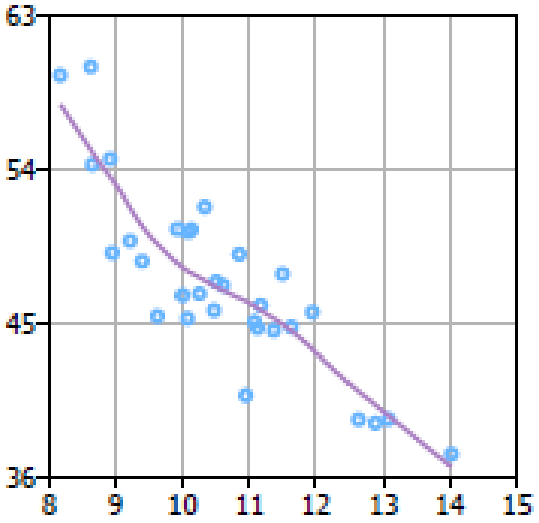
Extension of Wilkinson's Algorithm



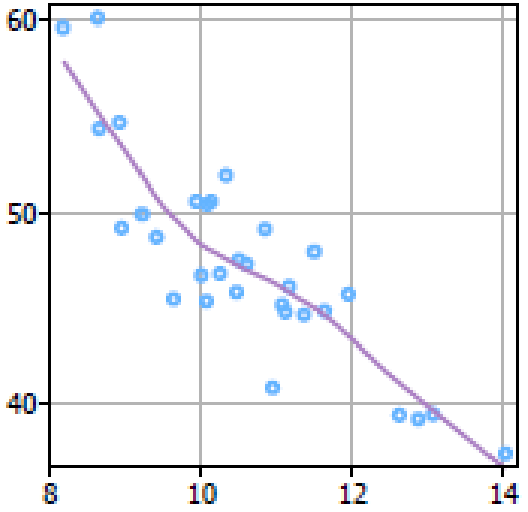
(a) Heckbert



(b) R's pretty



(c) Wilkinson



(d) Extended

Extension of Wilkinson's Algorithm

$$\text{Coverage} = 1 - \frac{1}{2}\frac{(d_{max} - l_{max})^2 + (d_{min} - l_{min})^2}{[.1(d_{max} - d_{min})]^2}$$

$$\text{Legibility} = \frac{format + font_{size} + orientation + overlap}{4}$$