

Graphs Which Almost Knock Your Socks Off: Using Visualization Principles to Convey the Message

Kathy Shelley, Iowa State University

With the sophistication of current personal computer software languages, most graphic packages offer many bells and whistles which include a variety of bitmap images and options to make the graph more flashy. Taken to the extreme, this can produce what Edward Tufte has termed "chartjunk" ⁽¹⁾. In an attempt to provide guidelines for constructing data graphs, statisticians and psychologists have conducted experimental designs to see which visual decoding tasks are easier than others for the viewer to understand the data patterning in the graph. This paper will present one visual decoding hierarchy and will use SAS/GRAPH* program code to illustrate these concepts. Barchart design will also be addressed, as will the use of color as an enhancement tool in two and three-dimensional bar charts. The resulting graphs will not "knock off socks", but will provide the viewer with a reasonably, clear display of the data.

William Cleveland has conducted statistical experimental designs to ascertain which visual decoding tasks are easier for the viewer to evaluate with fewer errors. His research has resulted in this visual decoding hierarchy ⁽²⁾ :

- (1) Position Along A Common Scale
- (2) Position Along Identical, Nonaligned Scales
- (3) Length
- (4) Angle - Slope
- (5) Area
- (6) Volume
- (7) Color Hue, Color Saturation, and Density

This hierarchy can be used in graphic design. For example, stacked bar graphs are difficult to visually decode because they involve length judgments as opposed to position judgments. Comparing **Figure 1** to **Figure 2**, we can see that it is much easier to see that (Midwest 13 to <16% BA) is bigger than (Midwest 0 to <13% BA) from **Figure 2**.

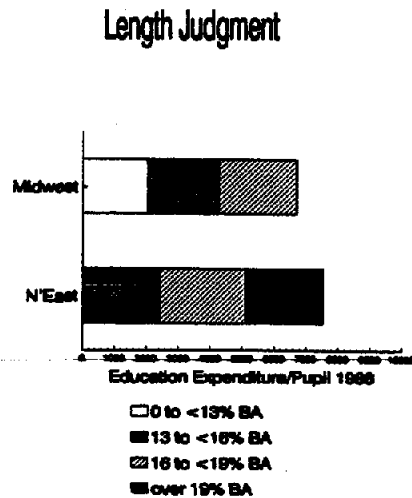


Figure 1

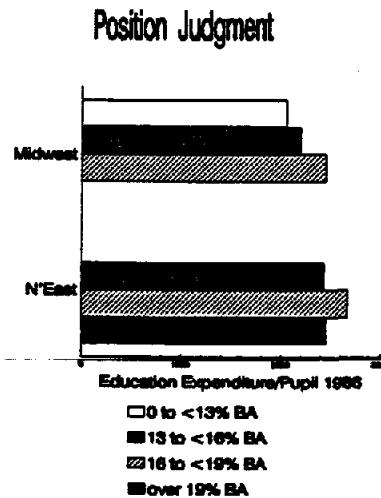


Figure 2

Another poor choice of graph design is the use of pie charts to compare the size of the various pie slices in contrast to using a bar chart to compare the length of the bars. As seen in **Figure 3** and **Figure 4** below, pie charts are difficult to visually decode because they involve angle judgments as opposed to position judgments.

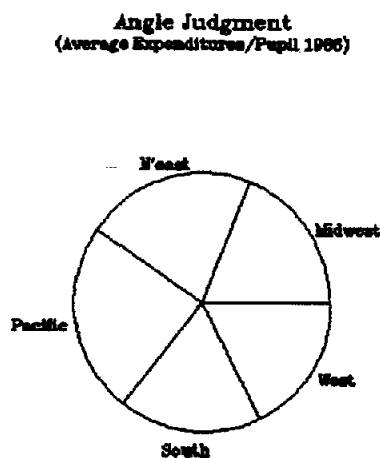


Figure 3

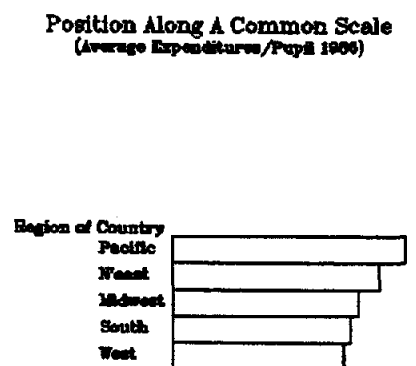


Figure 4

Optimally designed bar charts can be created using SAS/GRAPH* with the following properties: (1) Use HBAR instead of VBAR for better individual bar labelling, (2) use DESCENDING option for non-ordered data, (3) use ANNOTATE data sets to label bar values, and (4) use color for emphasis of a subset of the bars. Figure 5 was created from a SAS program which uses all four of these concepts. Highlights of the code will be printed following the figure.

Position Along A Common Scale in Descending Order
(Average Expenditure Is \$2290.39)

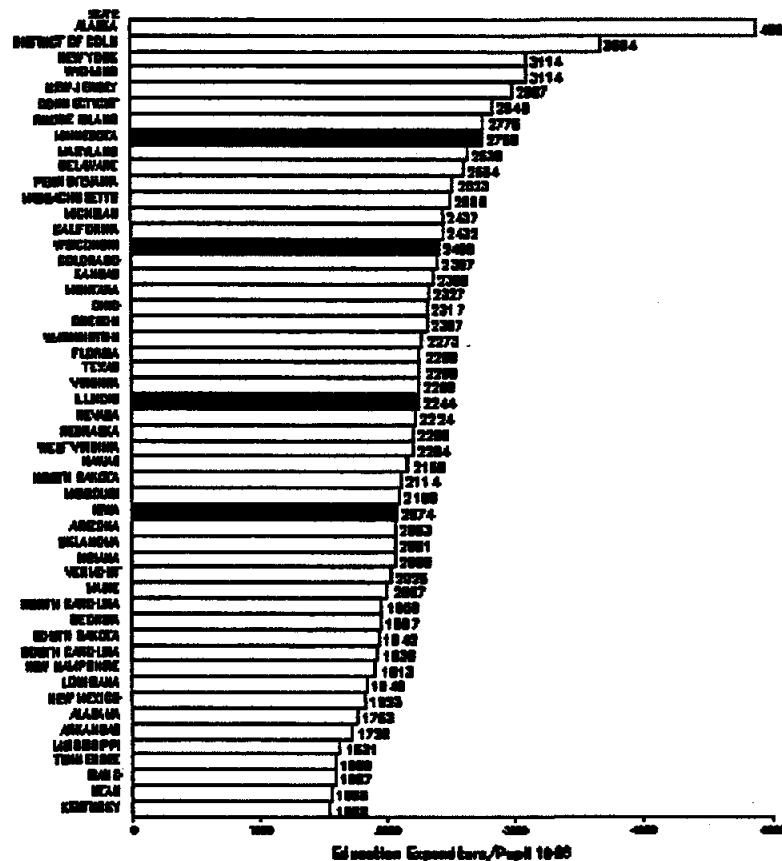


Figure 5

```
* this is statecol.sas;
data states;
infile 'may20.dat';
input #1 state 1-2 eastwest $ 30 region $ 23-29 pop 40-56
#2 #3 pctcoll 50-63 edex_sal 68-71 #4 #5;
name=fipstate(state);
...
if 0<pctcoll<13 then coll=1;
else if 13<=pctcoll<16 then coll=2;
else if 16<=pctcoll<19 then coll=3;
else if pctcoll>=19 then coll=4;
run;
```

This data step computes the mean expenditure and saves it to a macro variable so it can be passed to a title statement in the GCHART procedure:

```
data mean ;
  set states end=eof;
  * compute the mean and store it to a macro variable;
  n+1;
  sum+edex_sal;
  if eof then do;
    ave_sal=sum/n;
    output mean;
  end;
run;
```

```
data _null_;
  set mean; ave_sal=ave_sal*100;
  ave_sal=int(ave_sal); ave_sal=ave_sal/100;
  new_ave=compress("$"||ave_sal);
  call symput ('salmean',new_ave);
run;
```

This next data step creates a 2-valued variable, xcol, which will be used to highlight the bars with red for IA, MN, WI, or IL.:

```
data b;
  set states;
  name=substr(name,1,2);
  if name="IA" or name="MN"
    or name="WI" or name="IL" then xcol=1;
  else xcol=2;
  ...
```

This data step creates the annotate data set which will place the expenditure value at the end of the bar for each state:

```
data annot;
  length function color $ 8 TEXT $ 6 style $ 7;
  retain function 'LABEL'
  position '6' size 1.35 style 'simplex'
  when 'A'
  xsys '2' ysys '2' color 'BLACK';

  set b;
  x=.;
  y=edex_sal;
  MIDPNT=STATE;
  text=' '||compress(edex_sal)||' ';
  output;
run;
```

```
proc format; value xcolft
1="Midwest" 2="Other Regions"; run;
```

This is the GCHART procedure. Note that the pattern statements apply to the xcol variable, the annotate data set was incorporated with "ANNOTATE=ANNOX", and the &salmean macro variable was used in the title3 statement:

```
goptions rotate=landscape device=vga;
goptions vpos=100 hpos=100 ftext=simplex htext=1;
goptions cback=white colors=(black);
axis1 minor=none value=(f=simplex h=1.25);
axis2 minor=none / order=1500 to 5000 by 200 / length=50 pct
label=(h=1.5);
pattern1 c=red v=s;
pattern2 c=black v=s;
proc gchart data=b;
hbar state/ sumvar=edex_sal subgroup=xcol patternid=subgroup
DISCRETE type=mean descending nolegend ANNOTATE=ANNOX
/vref=ave_sal/ made=axis1 rade=axis2 space=0 nostats nozeros;
title1 h=1;
title2 h=3 f=triplex 'Position Along A Common Scale in Descending Order';
title3 h=2.5 f=complex "(Average Expenditure Is &salmean)";
footnote1 h=1;
format coll coll. xcol xcolft. state stateft.;
run;
```

Three-dimensional graphs can be made easier to comprehend by using color and bar labeling. In PROC G3D, selective bar coloring is built into the procedure and one can build an annotate data set to label the "bars". In Figure 6 below, the pillar shape has been used for the bars, magenta has been used to denote midwestern states, and 2-letter state abbreviations have been used to label each pillar. Note that Alaska and D.C. were excluded to allow the variation in states to be displayed.

Using Color to Add Emphasis in 3-D Graphs

Midwestern States Are From The Midwest
(Alaska and H.C. Excluded)

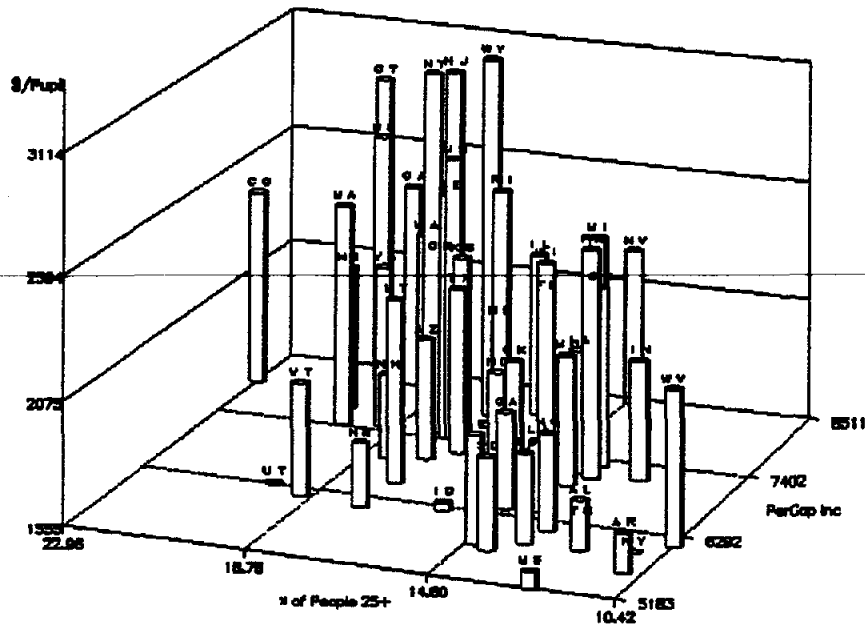


Figure 6

The SAS program code for Figure 6 follows:

```
* this is g3dstat1.sas;
data states;
infile 'may20.dat';
input #1 state 1-2 eastwest $ 30 region $ 23-29 pop 48-56
#2 #3 pcapinc 10-14 pctcoll 58-63 edex_sal 68-71 #4 #5;
...
```

Note that the color magenta in this annotate data step refers to the 2-letter state abbreviations above each pillar for Midwestern states:

```
data annex;
length function color $ 8 TEXT $ 5;
retain function 'LABEL'
position 'Z' size 1.5
when 'A'
xsys 'Z' ysys 'Z' zsys 'Z';
set b;
if region='Midwest' then color='magenta';
else color='black';
y=pctcoll;
x=pcapinc;
z=edex_sal;
text=compress(name);
output;
run;
```

In this data step, a variable named COLORVAL is created which will be used in the **color =** option in the **scatter** statement. The **where** statement eliminates Alaska and DC in the **PROC G3D** statement. The **anno=annox** option uses the state abbreviations above each pillar:

```
data b; set b;
if xcol=1 then colorval='magenta';
run;
options rotate=landscape device=vga;
options vpos=100 hpos=100 ftext=simplex htext=2;
options cback=white colors=(black);
pattern1 c=magenta v=s;
pattern2 c=black v=s;
proc g3d data=b anno=annox;
where name ne 'AK' and name ne 'DC';
scatter pctcoll*pcapinc=edex_sal/shape='pillar' color=colorval
grid;
format coll coll. xcol xcolft. pct_edex 3.2;
title1 f=swiss h=3
'Using Color to Add Emphasis in 3-D Graphs';
title2 f=triplex h=2 'Magenta States Are From The Midwest';
title3 f=triplex h=2 '(Alaska and D.C. Excluded)';
run;
```

Because of the complexity of 3-dimensional graphs, color can often be used as a clarifier and also can provide a 4th dimension to the graph. This is easily accomplished with SAS' data step capabilities coupled with annotate data sets and the **G3D** procedure.

* SAS, SAS/GRAPH are trademarks or registered trademarks of SAS Institute, Inc. in the USA and other countries.

References:

- (1) The Visual Display of Quantitative Information by Edward R. Tufte
Graphics Press, 1983
- (2) The Elements of Graphing Data by William S. Cleveland
Wadsworth Advanced Book Program, 1985
- (3) Envisioning Information by Edward R. Tufte
Graphics Press, 1990
- (4) SAS/GRAPH Software Volumes 1 and 2, Reference Version 6 First Edition, SAS Institute Inc., 1990