临床研究SAS高级编程
——Graph
Block Chart

Total Sales

Block Chart of Sum

Atlanta: $66,687
Paris: $85,885
Sydney: $53,114
Horizontal Bar Chart

Total Sales

<table>
<thead>
<tr>
<th>Site</th>
<th>FREQ</th>
<th>Sales SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>12</td>
<td>$66,687</td>
</tr>
<tr>
<td>Paris</td>
<td>12</td>
<td>$85,885</td>
</tr>
<tr>
<td>Sydney</td>
<td>12</td>
<td>$53,114</td>
</tr>
</tbody>
</table>

Sales SUM

$0  $20,000  $40,000  $60,000  $80,000  $100,000
Vertical Bar Chart

Total Sales

Sales SUM

- $90,000
- $80,000
- $70,000
- $60,000
- $50,000
- $40,000
- $30,000
- $20,000
- $10,000
- $0

Atlanta | Paris | Sydney
Site

Sales SUM(b)
Pie Chart

Sales by Site and Department

- Atlanta
- Paris
- Sydney

Parts
Repairs
Tools

All Quarters
Star Chart

Rejected Parts

- May 64
- Apr 70
- Mar 61
- Jun 63
- Jul 72
- Aug 59
- Sep 48
- Oct 37
- Nov 59
- Feb 66
- Jan 53
- Dec 60

GCHDSCRT(a)
Scatter Plot

Study of Height vs Weight

- height
- weight

50 60 70 80 90 100 110 120 130 140 150
Regression Plot

Study of Height vs Weight

Regression Equation:
\[
\text{height} = 30.77829 + 0.50121 \times \text{weight} - 0.002375 \times \text{weight}^2 + 5.448 \times 10^{-6} \times \text{weight}^3
\]
High-Low Plot

Comparison of Grades by Section

Monday       Wednesday       Friday

GSYCMBP1(a)
Surface Plot

Surface Plot of HAT Data Set
3-D Scatter Plot

Iris Species Classification
Physical Measurement
Source: Fisher (1936) Iris Data

Petallen: Petal Length in mm.
Petalwid: Petal Width in mm.

Sepallen: Sepal Length in mm.
Sepal Width not shown

GTDSCAT
Contour Plot

Clay Content at Site A

pct_clay

GCTCLAY1
Hazardous Waste Site Installations (1997)
China Map
Choropleth Map

Hazardous Waste Site Installations (1997)
Surface Map

Hazardous Waste Site Installations (1997)
New Directions

Goals and strategies for the coming year

ABC Engineering, Inc
January 1998
Graphs Combination
Annotated Graph

Distribution Center Locations

Seattle
Chicago
Atlanta
Graph Elements

- **minor tick marks**
- **vertical axis label**
- **vertical axis line**
- **reference line**
- **plot symbols**
- **plot line**
- **major tick marks**
- **major tick mark values**
- **horizontal axis line**
- **horizontal axis label**
- **YEAR**

---

**Graph Details**

- **Rate**
  - 500
  - 400
  - 300
  - 200
  - 100

- **Year Values**
  - 1970
  - 1972
  - 1974
  - 1976
  - 1978
  - 1980
Graph Procedure

PROC statement
- proc gplot data=demo;

Subordinate statement
- plot ht*wt;

BY statement

NOTE statement
- Title;
- Footnote;

RUN statement
- quit;
Global Statements

- GOPTIONS
- SYMBOL
- AXIS
- LEGEND
- PATTERN
- TITLE & FOOTNOTE
Other Statements

LIBNAME
FILENAME
FORMAT
LABEL
ODS
OPTIONS
Size of Cell

HPOS=8
VPOS=12
HSIZE=4IN
VSIZE=6IN

HPOS=6
(specified)

VPOS=10
(specified)

VSIZE=7.5IN
(specified)
Default Location for Graph Elements
proc gdevice catalog=sashelp.devices nofs;
   list;
run;
quit;

goptions device=cgm;
Save Graph Files

GSFNAME
GSFMODE = Replace, Append
GSACCESS

filename grafout "&cd.slide.ps";

goptions reset=all
device=pscolor
gsfname=grafout
gsfmode=replace
ftext=swissb;

proc gslide border name='proposal';
title1 h=4 'Proposed Design Improvements:';
title2 h=3 ' * Increase Stability';
title3 h=3 ' * Increase Speed';
title4 h=3 ' * Reduce Weight';
footnote h=2 j=l 'ABC Company';
run;
quit;
Colors in SAS/GRAPH

SAS employs a number of different color schemes:

- HLS (hue lightness saturation)
- RGB (red green blue)
- HSV (hue saturation value)
- Gray-Scale
- CMY(K) (cyan magenta yellow(black))
- SAS color names (from the SAS Registry)
- the SAS Color Naming System (CNS).

Each color scheme uses different complex algorithms to construct a given color.
HLS color scheme

- Any color in the HLS scheme can be constructed by specifying values for hue, lightness and saturation.
- HLS color names are of the form Hhhhllss (hexadecimal format).
  - hhh is the hue component (000-168)
  - ll is the lightness component (0-FF)
  - ss is the saturation component (0-FF).
- The %HLS converts values of hue, lightness and saturation into the corresponding HLS code and circumvents the need for hexadecimal representations of colors.
  - E.g. %hls(360,50,100)=H16880FF, %hls(130,50,90)=H08280E6
  - Pattern color= %hls(130,50,90)
RGB color scheme

- Any given color can be defined as a combination of red, green and blue colors

- RGB color names are of the form CXrrggb (hexadecimal format)
  - rr is the red component (00-FF)
  - gg is the green component (00-FF)
  - bb is the blue component (00-FF).

- Like HLS, %RGB converts percentages to a hexadecimal number
  - E.g. %RGB(30,50,100)=CX4D80FF, %RGB(80,50,90)=CXCC80E6
  - Pattern color= %RGB(30,50,100)
RGB color scheme
GRAY-Scale color scheme

The Gray-scale color scheme may be used to create various shades of gray.

Gray-scale shades are formed by combining the word GRAY with the hexadecimal value of a color from 0 to 255.

Gray-scale color names are of the form GRAY‖ (hexadecimal format).

<table>
<thead>
<tr>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>A0</th>
<th>B0</th>
<th>C0</th>
<th>D0</th>
<th>E0</th>
<th>F0</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>14</td>
<td>1E</td>
<td>28</td>
<td>32</td>
<td>3C</td>
<td>46</td>
<td>50</td>
<td>5A</td>
<td>64</td>
<td>6E</td>
<td>78</td>
<td>82</td>
<td>8C</td>
<td>96</td>
<td>A0</td>
</tr>
</tbody>
</table>

E.g. grayAE, gray45
All of the SAS/GRAPH fonts are stored in the catalog SASHELP.FONTS.

The special symbols can be Greek letters, mathematical symbols, subscription, superscryption, underline and custom-designed symbols.

To preview the fonts/symbols and its corresponding character codes, we can use the following code:

```
proc gfont name=greek/marker/music/math/sperical
   nobuild
   height=3.7
   romcol=red
   romfont=swissl
     noromhex
   romht=2.7
   showroman;
run;
quit;
```
† This is only an example.
<table>
<thead>
<tr>
<th>Type Style</th>
<th>Font Name</th>
<th>Type Sample</th>
<th>Uniform Font</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brush</td>
<td>BRUSH</td>
<td>A B C a b c 1 2 3</td>
<td></td>
</tr>
<tr>
<td>Century</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bold</td>
<td>CENTB</td>
<td>A B C a b c 1 2 3</td>
<td>CENTBU</td>
</tr>
<tr>
<td>Bold Empty</td>
<td>CENTBE</td>
<td>A B C a b c 1 2 3</td>
<td>CENTBIU</td>
</tr>
<tr>
<td>BoldItalic</td>
<td>CENTBI</td>
<td>A B C a b c 1 2 3</td>
<td>CENTBIU</td>
</tr>
<tr>
<td>BoldItalic Empty</td>
<td>CENTBIE</td>
<td>A B C a b c 1 2 3</td>
<td></td>
</tr>
<tr>
<td>Expanded</td>
<td>CENTX</td>
<td>A B C a b c 1 2 3</td>
<td>CENTXU</td>
</tr>
<tr>
<td>Expanded Empty</td>
<td>CENTXE</td>
<td>A B C a b c 1 2 3</td>
<td></td>
</tr>
<tr>
<td>Expanded Italic</td>
<td>CENTXI</td>
<td>A B C a b c 1 2 3</td>
<td>CENTXIU</td>
</tr>
<tr>
<td>Expanded Italic Empty</td>
<td>CENTXIE</td>
<td>A B C a b c 1 2 3</td>
<td></td>
</tr>
<tr>
<td>□</td>
<td>⊥</td>
<td>∠</td>
<td>:</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>≤</td>
<td>≥</td>
<td>∞</td>
<td>~</td>
</tr>
<tr>
<td>L</td>
<td>M</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>←</td>
<td>↓</td>
<td>∂</td>
<td>∇</td>
</tr>
<tr>
<td>l</td>
<td>j</td>
<td>k</td>
<td>l</td>
</tr>
</tbody>
</table>
GOPTIONS

RESET=ALL | GLOBAL | statements
options that affect the appearance of the display area and the graphics output:

- ASPECT=`scaling-factor`
- AUTOSIZE=ON | OFF | DEFAULT
- BORDER | NOBORDER
- CELL | NOCELL
- GSIZE=`lines`
- HORIGIN=`horizontal-offset` <IN | CM>
- HPOS=`columns`
- HSIZE=`horizontal-size` <IN | CM>
- IBACK=`fileref` | `external-file`
- IMAGESTYLE = TILE | FIT
- IMAGEPRINT | NOIMAGEPRINT
- ROTATE=LANDSCAPE | PORTRAIT
- ROTATE | NORotate
- SWAP | NOSWAP
- TARGETDEVICE=`target-device-entry`
- VORIGIN=`vertical-offset` <IN | CM>
- VPOS=`rows`
- VSIZE=`vertical-size` <IN | CM>
- XMAX=`width` <IN | CM>
- XPIXELS=`width-in-pixels`
- YMAX=`height` <IN | CM>
- YPIXELS=`height-in-pixels`
GOPTIONS

options that affect color
   CBACK=background-color
   CBY=BY-line-color
   COLORS=<colors-list | NONE>
   CPATTERN=pattern-color
   CSYMBOL=symbol-color
   CTEXT=text-color
   CTITLE=title-color
   PENMOUNTS=active-pen-mounts
   PENSORT | NOPENSORT
options that control font selection or text appearance
   CHARTYPE=hardware-font-chartype
   FASTTEXT | NOFASTTEXT
   FBY=BY-line-font
   FCACHE=number-fonts-open
   FONTRES=NORMAL | PRESENTATION
   FTEXT=text-font
   FTITLE=title-font
   FTRACK=LOOSE | NONE | NORMAL | TIGHT | TOUCH | V5
   HBY=BY-line-height <units>
   HTEXT=text-height <units>
   HTITLE=title-height <units>
   RENDER=APPEND | DISK | MEMORY | NONE | READ
   RENDERLIB=libref
   SIMFONT=software-font
options that control how output is sent to devices or files

- ADMGDF | NOADMGDF
- DEVADDR= device-address
- DEVICE= device-entry
- DEVMAP= device-map-name | NONE
- EXTENSION= 'file-type'
- FILECLOSE= DRIVERTERM | GRAPHEND
- FILEONLY | NOFILEONLY
- GACCESS= output-format | 'output-format > destination'
- GEND= 'string' < 'string-n'
- GEPILLOG= 'string' < 'string-n'
- GOUTMODE= APPEND | REPLACE
- GPROLOG= 'string' < 'string-n'
- GPROTOCOL= module-name
- GSFLN= record-length
- GSFMODE= APPEND | PORT | REPLACE
- GSFNAME= fileref
- GSFPROMPT | NOGSFPROMPT
- GSTART= 'string' < 'string-n'
- HANDSHAKE= HARDWARE | NONE | SOFTWARE | XONXOFF
- KEYMAP= map-name | NONE
- POSTGEPILLOG= 'string'
goptions reset=all
cback=white
colors=(black)
  gunit=pct
rotate=landscape
  device=cgmof97l
gsfmode=replace
gsfname=plotgph;
Axis Syntax

axis scale options:

- INTERVAL=EVEN | UNEVEN | PARTIAL
- LOGBASE=base | E | PI
- LOGSTYLE=EXPAND | POWER
- ORDER=(value-list)

appearance options:

- COLOR=axis-color
- LENGTH=axis-length <units>
- NOBRACKETS
Axis Syntax

NOPLANE
OFFSET=\(<n1 >\:<,n2 >\)<units> | (\(<n1<units>>\:<,n2<units>>\))
ORIGIN=\(<x>,<y >\)<units> | (\(<x<units>>\:<,y<units>>\))
STYLE=\(line-type\)
WIDTH=\(thickness-factor\)
tick mark options:
  MAJOR=(\(tick-mark-suboption(s)\)) | NONE
  MINOR=(\(tick-mark-suboption(s)\)) | NONE

text options:
  LABEL=(\(text-argument(s)\)) | NONE
  REFLABEL=(\(text-argument(s)\)) | NONE
  SPLIT="split-char"
  VALUE=(\(text-argument(s)\)) | NONE
Axis Syntax

Unit

CELLS character cells
CM centimeters
IN inches
PCT percentage of the graphics output area
PT points

If you omit units, a unit specification is searched for in this order:
1 GUNIT= in a GOPTIONS statement
2 the default unit, CELLS.
### Axis Syntax

#### Color

<table>
<thead>
<tr>
<th>Option</th>
<th>Items Affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AXIS statement:</td>
<td>axis label</td>
</tr>
<tr>
<td>LABEL=(COLOR=color)</td>
<td>reference-line labels</td>
</tr>
<tr>
<td>REFLABEL=(COLOR=color)</td>
<td>major tick mark values</td>
</tr>
<tr>
<td>VALUE=(COLOR=color)</td>
<td></td>
</tr>
<tr>
<td>calling procedure:</td>
<td>all axis text (AXIS label and major tick mark</td>
</tr>
<tr>
<td></td>
<td>value descriptions)</td>
</tr>
<tr>
<td>CTEXT=</td>
<td></td>
</tr>
<tr>
<td>CAXIS=</td>
<td>axis line and major and minor tick marks</td>
</tr>
</tbody>
</table>
Axis Syntax

Text Description Suboptions

ANGLE=degrees
COLOR=text-color
FONT=font | NONE
HEIGHT=text-height <units>
JUSTIFY=LEFT | CENTER | RIGHT
ROTATE=degrees
Legend

`LEGEND<1...99><options>;;`

option(s) can be one or more options from any or all of the following categories:
- appearance options
- `ACROSS=number-of-columns`
- `CBLOCK=block-color`
- `CBORDER=frame-color`
- `CFRAME=background-color`
- `CSHADOW=shadow-color`
- `DOWN=number-of-rows`
- `FRAMEFWIDTH=thickness-factor`
- `SHAPE=BAR(width,height) <units> | LINE(length) <units> | SYMBOL(width,height) <units>`
position-options

- **MODE**=PROTECT | RESERVE | SHAREOFFSET=(<x><,y><units>) | (<x<units><,y<units>>)  
- **ORIGIN**=(<x><,y><units>) | (<x<units><,y<units>>)  
- **POSITION**=(<BOTTOM | MIDDLE | TOP><LEFT | CENTER | RIGHT><INSIDE | OUTSIDE>)

text-options

- **LABEL**=(text-argument(s)) | NONEORDER=(value-list)  
- **VALUE**=(text-argument(s)) | NONE
legend label=(justify=c 'Distribution'
     justify=c 'Centers')
value=(tick=1 justify=c 'Portland,'
     justify=c 'Maine'
     tick=2 justify=c 'Paris,'
     justify=c 'France'
     tick=3 justify=c 'Sydney,'
     justify=c 'Australia');
SYMBOL<1...255>
    <COLOR=symbol-color>
    <MODE=EXCLUDE | INCLUDE>
    <REPEAT=number-of-times>
    <STEP=distance<units>>
    <appearance-option(s)>
    <interpolation-option>
    <SINGULAR=n>;

appearance-options can be one or more of these:
BWIDTH=box-width
CI=line-color
CO=color
CV=value-color
FONT=font
HEIGHT=symbol-height<units>
LINE=line-type
POINTLABEL<=label-description(s)) | NONE>
VALUE=special-symbol | text-string | NONE
WIDTH=thickness-factor
The `interpolation-option` can be one of these:

- **general methods**
  - `INTERPOL=JOIN`
  - `INTERPOL=map/plot-pattern`
  - `INTERPOL=NEEDLE`
  - `INTERPOL=NONE`
  - `INTERPOL=STEP<placement><J><S>`

- **high-low interpolation methods**
  - `INTERPOL=BOX<option(s)><00...25>`
  - `INTERPOL=HILO<C<option(s)>`
  - `INTERPOL=STD<1 | 2 | 3><variance<option(s)>`

- **regression interpolation methods**
  - `INTERPOL=R?type><0><CLM | CLI<50...99>>`

- **spline interpolation methods**
  - `INTERPOL=L<degree><P><S>`
  - `INTERPOL=SM<nn><P><S>`
  - `INTERPOL=SPLINE<P><S>`
I=BOX <options> <00-25>

F        fills the box with the color specified by CV= and outlines the box with the color specified by CO=

J        joins the median points of the boxes with a line

T        draws tops and bottoms on the whiskers.

In addition, you can specify a percentile to control the length of the whiskers within the range 00 through 25. These are examples of percentile specifications and their effect:

00        high/low extremes. INTERPOL=BOX00 is not the same as the default, INTERPOL=BOX.

01        1st percentile low, 99th high

05        5th percentile low, 95th high

10        10th percentile low, 90th high

25        25th percentile low, 75th high; since the box extends from the 25th to the 75th percentile, no whiskers are produced.
临床研究
SAS
高级编程

I=BOX

图示：

Y轴：
0 10 20 30 40 50

X轴：
A B C

数据分布图
I=HILO <C><option>

C draws tick marks at the close value instead of at the mean value. Specifying C assumes that there are three values of Y (HIGH, LOW, and CLOSE) for every value of X. If more or fewer than three Y values are specified, the mean is ticked. The Y values can be in any order in the input data set.

In addition, you can specify one of these values for option:

B connects the minimum and maximum Y values with bars instead of lines. Use the BWIDTH= option to increase the width of the bars.

J joins the mean values or the close values (if HILOC is specified) with a line. This point is not marked with a tick mark. You cannot use the PLOT statement option AREAS= with INTERPOL=HILOJ.

T adds tops and bottoms to each line.

BJ connects maximum and minimum values with a bar and joins the mean or close values.

TJ adds tops and bottoms to the lines and joins the mean or close values.
CONNECTS data points with straight lines. Points are connected in the order they occur in the input dataset. Therefore, the data should be sorted by the independent variable.

If the data contain missing values, the observations are omitted. However, the plot line is not broken at missing values unless SKIPMISS option is used.
Type specifies the type of regression. Specify one of these values for type:

- **L** requests linear regression representing the regression equation
  \[ Y = \beta_0 + \beta_1 X \]

- **Q** requests quadratic regression representing the regression equation
  \[ Y = \beta_0 + \beta_1 X + \beta_2 X^2 \]

- **C** requests cubic regression representing the regression equation
  \[ Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \]
Other Interpol

- I=SM
- I=Needle
- I=STD <1 | 2 | 3>
- I=STEP
<table>
<thead>
<tr>
<th>VALUE=</th>
<th>Plot Symbol</th>
<th>VALUE=</th>
<th>Plot Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLUS</td>
<td>+</td>
<td>%</td>
<td>(percent)</td>
</tr>
<tr>
<td>X</td>
<td>×</td>
<td>&amp;</td>
<td>(ampersand)</td>
</tr>
<tr>
<td>STAR</td>
<td>*</td>
<td>'</td>
<td>(single quote)</td>
</tr>
<tr>
<td>SQUARE</td>
<td>□</td>
<td>=</td>
<td>(equals)</td>
</tr>
<tr>
<td>DIAMOND</td>
<td>◆</td>
<td>-</td>
<td>(hyphen)</td>
</tr>
<tr>
<td>TRIANGLE</td>
<td>△</td>
<td>@</td>
<td>(at)</td>
</tr>
<tr>
<td>HASH</td>
<td>#</td>
<td>*</td>
<td>(asterisk)</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>+</td>
<td>(plus)</td>
</tr>
<tr>
<td>Z</td>
<td>Z</td>
<td>&gt;</td>
<td>(greater than)</td>
</tr>
<tr>
<td>PAW</td>
<td>.</td>
<td>.</td>
<td>(period)</td>
</tr>
<tr>
<td>POINT</td>
<td>.</td>
<td>&lt;</td>
<td>(less than)</td>
</tr>
<tr>
<td>DOT</td>
<td>●</td>
<td>,</td>
<td>(comma)</td>
</tr>
</tbody>
</table>
PROC GCHART <DATA=input-data-set>
   <ANNOTATE=Annotate-data-set>
   <GOUT=<libref:output-catalog>
   <IMAGEMAP=output-data-set>;

BLOCK chart-variable(s) <option(s>);

HBAR | HBAR3D | VBAR | VBAR3D chart-variable(s) <option(s>);

PIE | PIE3D | DONUT chart-variable(s) <option(s>);

STAR chart-variable(s) <option(s>);
PROC GPLOT <DATA=input-data-set>
   <ANNOTATE=Annotate-data-set>
   <GOUT=<libref.>output-catalog>
   <IMAGEMAP=output-data-set >
   <UNIFORM>;

   BUBBLE plot-request(s) </option(s)>;
   BUBBLE2 plot-request(s) </option(s)>;

   PLOT plot-request(s) </option(s)>;
   PLOT2 plot-request(s) </option(s)>;
Excellence in statistical graphics consists of complex ideas communicated with **clarity**, **precision** and **efficiency** (Tufte, 1983)

Tufte’s graphics principles for accurate and retained information extraction

- Show all the data
- Induce the viewer to think about the substance rather than the graphic design – maximize the data-to-ink ratio
- Avoid distorting what the data are saying
- Make large data sets coherent
- Encourage the eye to compare different pieces of data - leverage investment by showing multiple plots of same type
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a clear purpose: description, exploration, tabulation
- Be closely integrated with the statistical and verbal descriptions of a data set
- Use gray scale and color sparingly
Gold Standards of Good Graphics

- Clarity
- Precision
- Efficiency
Which one is more clear?
Upper or Bottom?
Precision

Mean % change in Triglycerides

-5 (n=9)
-4 (n=24)
-2 (n=48)
0 (n=77)
2 (n=83)
4 (n=58)
>=5 (n=76)

Placebo
50 mg SR121463B
100 mg SR121463B

SBP (min HG) - (Sitting)

107
108
109
110
111
112
113
114
115
116
117

Day

15 30 45 60 75 90 105 120

# subjects

Placebo 110 104 99 94 95 99 95 96
50 mg 106 104 100 95 99 99 96 96
100 mg 115 114 107 100 100 100 100 100
Good Plotting Practices

- Using two-dimensional graphs to display one-dimensional data (such as three-dimensional histograms) is inefficient and often results in perceptual ambiguity;
- Important events or data points of particular interest should be labeled on the graph and not just as some reference in the legend;
- Always use clear, concise, detailed labeling. Avoid computer labels;
- Maximize the data-ink ratio, i.e., the proportion of a graph’s ink that is devoted to the display of non-redundant information;
If a grid is needed, do not make the grid color black or the line width the same width as other lines or symbols in the graph. Use a gray color instead to bring the data to the foreground;
- Avoid confusing legends;
- Plots should have greater length than height, the so-called golden rectangle, since our eye is naturally better at detecting trends on the horizontal than the vertical and it is easier to place labels in a wider graph than a thinner graph;
- Use colors to identify data properly.
Graphics Principles for Clinical Data

- Set a sound aspect ratio
- Include (feint) grid lines and reference lines as appropriate
- Sort categories e.g. on dot plots
- Use group and multi-panel displays (with same scale)
- Minimize use of legends by annotating the graph
- Don’t use more than 3 hues
- Use contrast hues (blue-orange/yellow); avoid color-blind hues (red-green)
- Use black and gray for 2 group comparisons (treatment and control)
- Don’t use too many dashed line types – they are difficult to distinguish
Good Plotting Practices

Leading Rice Producers 1995 and 1996

- China: 375
- India: 242
- Indonesia: 101

Leading Rice Producers 1995 and 1996

- China: 375
- India: 242
- Indonesia: 101
Good Plotting Practices (cont’ed)

Supine: Change from baseline on Day 1 – SAR407899 25 mg BID

Baseline is the Day 1 predose value.
Only planned values are included in the plot.

•: Value was measured with time deviations (<8 min in supine or not 2, 3 or 4 min in standing position).
Good Plotting Practices (cont’ed)

# subjects
Placebo 110 104  99  94  95
50 mg 106 104 100  95  99
100 mg 115 114 107 100  96

SBP (min HG) - (Sitting)

Day

Follow-up
Good Plotting Practices (cont’ed)

Heart Rate raw data in D1 - (Mean +/- SD)

Theoretical time (hours)

Heart Rate (bpm)

Treatment

Placebo
SAR110894 5 mg
SAR110894 15 mg
SAR110894 30 mg
SAR110894 60 mg
SAR110894 120 mg
SAR110894 300 mg
Good Plotting Practices (cont’ed)
Good Plotting Practices (cont’ed)

![Scatter plots showing the relationship between reading and math scores.](image)
Good Plotting Practices (cont’ed)
CDF plot using $i = \text{steplj}$ in SYMBOL statement

Cumulative Distribution Function for Heart rate (BPM)
Density plot using \( i = \text{join} \) in SYMBOL statement
Boxplot using `i = boxt` in SYMBOL statement

Heart Rate

<table>
<thead>
<tr>
<th>Day</th>
<th>Placebo (n=641)</th>
<th>0.5 mg (n=159)</th>
<th>5 mg (n=166)</th>
<th>Placebo (n=147)</th>
<th>0.5 mg (n=36)</th>
<th>5 mg (n=36)</th>
<th>Placebo (n=73)</th>
<th>0.5 mg (n=18)</th>
<th>5 mg (n=18)</th>
</tr>
</thead>
</table>
Needle plot using \( i = \text{needle} \) in SYMBOL statement.

Primary: Treatment=1 Gender=Male PntId=10001 SITE=

Days from first tablet taken
High-low plot using $i = \text{hilocjt}$ in SYMBOL statement
KM plot using \( i = \text{steplj} \) in SYMBOL statement

Symbols=Censors

TRT A

TRT B

Kaplan-Meier Estimate

log-rank \( p = 0.6195 \)

Number at Risk

<table>
<thead>
<tr>
<th>TRT A</th>
<th>TRT B</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>211</td>
</tr>
<tr>
<td>129</td>
<td>132</td>
</tr>
<tr>
<td>71</td>
<td>57</td>
</tr>
<tr>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

Time (Months)

0 3 6 9 12 15 18
What’s the Annotate Facility

- The annotate facility enables you to generate a special data set of graphics commands from which you can produce graphics output.
- This data set is referred to as an *Annotate data set*.
- Use it to generate custom graphics or to enhance graphics output from many SAS/GRAPH procedures, including GCHART, GCONTOUR, GMAP, GPLOT and so on.
- The Annotate Facility acts as a bridge between the procedure selected by the user and the user's desire to customize the graphics output.
The Annotate data set is not different from any other SAS data set. In an Annotate data set, each observation represents a command to draw a graphics element or to perform an action. The graphic elements drawn by these commands can be added to SAS/GRAPH output or displayed with the GANNO or GSLIDE procedure as a custom graphic.

The observations in an Annotate data set use a set of predefined Annotate variables. The values of the variables in the observation determine what is done and how it is done.

To create these observations, you assign values to the variables either explicitly with a DATA step or implicitly with Annotate macros.
What can the Annotate Facility do

Enhances output from SAS/GRAPH procedures by adding graphics elements to the output.

- label points on a map using map coordinates
- label bars on horizontal and vertical bar charts
- label points on a plot
- create a legend for a three-dimensional graph

Use an Annotate data set to create custom graphics

- create various types of graphs (including pie charts, bar charts, and plots)
- draw graphics elements such as lines, polygons, arcs, symbols, and text
Examples (1)

Distribution Center Locations

Seattle
Chicago
Atlanta

Flag of Micronesia
Examples(2)

<table>
<thead>
<tr>
<th># subjects</th>
<th>Placebo</th>
<th>50 mg SR121463B</th>
<th>100 mg SR121463B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 15</td>
<td>110</td>
<td>106</td>
<td>115</td>
</tr>
<tr>
<td>Day 30</td>
<td>104</td>
<td>104</td>
<td>114</td>
</tr>
<tr>
<td>Day 60</td>
<td>99</td>
<td>100</td>
<td>107</td>
</tr>
<tr>
<td>Day 90</td>
<td>94</td>
<td>95</td>
<td>100</td>
</tr>
<tr>
<td>Follow-up</td>
<td>95</td>
<td>99</td>
<td>96</td>
</tr>
</tbody>
</table>

SBP (min HG) - (Sitting)
The Annotate Process

The annotate is used primarily to enhance a graph. First, we must know three questions for use of the annotate.

- What is to be done?
- Where is it to be done?
- How is it to be done?

The annotate process is the same as answering the three questions above.

Some of the variables that are used to answer these three questions include:

- What: FUNCTION (this is the ONLY variable used to answer this question)
- How: COLOR, SIZE, STYLE, POSITION
- Where: X, Y, XSYS, YSYS
WHAT is to be done?

- An action variable FUNCTION tells what to do

- The character variable FUNCTION provides the information on WHAT is to be done. The FUNCTION variable accepts a set of predefined values (functions) that perform both graphics tasks and programming tasks

- Some of the common values of the variable FUNCTION
  - BAR: draw a rectangle
  - DRAW: draw a line
  - LABEL: places text or symbols on the graphic
  - MOVE: allows movement to a specific point on the graphic
  - SYMBOL: places a symbol on the graphic
  - POINT: places a single point
**HOW is it to be done?**

- Attribute variables tell how to do it

- The attribute variables specify the characteristics of the graphics element (for example, color, size, line style, text font)

- Some of the attribute variables include:
  - TEXT='string' add string to display
  - COLOR='color' specify the color of text
  - SIZE=n size of text
  - STYLE='font' select font for text string
WHERE is it to be done?

- Positioning variables tell where to do it

- The positioning variables specify the point at which to draw the graphics element

- some of the positioning variables include
  - X specifies a numeric horizontal coordinate
  - Y specifies a numeric vertical coordinate
  - XSYS specifies coordinate system for X or XC coordinates
  - YSYS specifies coordinate system for Y or YC coordinates
WHERE is it to be done?

The coordinates are usually placed using the numeric variables $X$ and $Y$.

How these coordinates are interpreted depends on the coordinate system, which is specified by the XSYS and YSYS variables?
Coordinate System(1)

The coordinate system itself can be selected by using the character variables XSYS and YSYS, these variables can take on one of twelve 'system' values, two of these values for XSYS and YSYS will satisfy most of your Annotate needs.

<table>
<thead>
<tr>
<th>Area</th>
<th>Unit</th>
<th>Coordinate System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>Absolute</td>
</tr>
<tr>
<td>Values</td>
<td>1</td>
<td>Relative</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Graphics Output Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>Absolute</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Relative</td>
</tr>
<tr>
<td>Cells</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Procedure Output Area</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>%</td>
<td>Absolute</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Relative</td>
</tr>
<tr>
<td>Cells</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A particular value of X or Y will be located depends on the value assigned to XSYS or YSYS.

The values of XSYS and YSYS need not be constant in the Annotate data set. XSYS and YSYS do not need to have the same values.

E.g.

- XSYS & YSYS='2'  ‘Absolute data value' places the point according to the values of the horizontal and vertical axes that are plotted on the graph.
- XSYS & YSYS='3'  'Absolute Graphics Output Area percent' uses percentages of the entire graphics area, which are measured from the lower left corner.
Annotate Example(1)

Sample Annotate Graphics

<table>
<thead>
<tr>
<th>function</th>
<th>style</th>
<th>color</th>
<th>text</th>
<th>hsys</th>
<th>xsys</th>
<th>ysys</th>
<th>x</th>
<th>y</th>
<th>position</th>
<th>size</th>
<th>line</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>swissb</td>
<td>green</td>
<td>Sample Annotate Graphics</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td>85</td>
<td>6</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>move</td>
<td>swissb</td>
<td>green</td>
<td>Sample Annotate Graphics</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>28</td>
<td>30</td>
<td>6</td>
<td>6</td>
<td>.</td>
</tr>
<tr>
<td>draw</td>
<td>swissb</td>
<td>red</td>
<td>Sample Annotate Graphics</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>68</td>
<td>30</td>
<td>6</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>draw</td>
<td>swissb</td>
<td>red</td>
<td>Sample Annotate Graphics</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>48</td>
<td>70</td>
<td>6</td>
<td>0.8</td>
<td>1</td>
</tr>
<tr>
<td>draw</td>
<td>swissb</td>
<td>red</td>
<td>Sample Annotate Graphics</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>28</td>
<td>30</td>
<td>6</td>
<td>0.8</td>
<td>1</td>
</tr>
</tbody>
</table>
Annotate Example(2)

Raw data from Day -3 to EOS - Mean (+/-SEM)

Baseline is Day -3

Visit
Day-3 Day7 Day15 Day22 EOS

Calcium (mmol/L)
2.35
2.40
2.45
2.50
2.55
2.60

Treatment: 
Placebo
SSR125543 20 mg
SSR125543 50 mg
SSR125543 100 mg

Baseline is Day -3

<table>
<thead>
<tr>
<th>STYLE</th>
<th>FUNCTION</th>
<th>COLOR</th>
<th>XSYS</th>
<th>YSYS</th>
<th>HSYS</th>
<th>WHEN</th>
<th>POSITION</th>
<th>X</th>
<th>Y</th>
<th>LINE</th>
<th>SIZE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MOVE</td>
<td>black</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>22</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>MOVE</td>
<td>black</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>-0.75</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DRAW</td>
<td>white</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>1.5</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>MOVE</td>
<td>white</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>-1.5</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>DRAW</td>
<td>black</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>0.15</td>
<td>7</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>6</td>
<td>MOVE</td>
<td>black</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>1.2</td>
<td>7</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>7</td>
<td>DRAW</td>
<td>black</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>a</td>
<td>5</td>
<td>0.15</td>
<td>7</td>
<td>1</td>
<td>2.5</td>
</tr>
</tbody>
</table>
The Annotate data set can be built:

- by using assignment statements
  - The Annotate data set can be built by using assignment statements.
  - This technique is best employed when the Annotate data set does not depend on incoming data and only a few observations are needed.

- from an existing SAS data set (or other data source)
  - When the graphics display depends on an established SAS data set, that data set can often be used to build the Annotate data set as well.
  - This technique is especially useful when you need to place labels or text strings at a location that is to be determined by the data itself.
Building the Annotate Dataset (2)

## OBS function style color xsys ysys when x y size
1 move solid color 5 2 a 1 50 .
2 bar solid color 5 2 a 8 500 .
3 move solid black 1 2 a 0 50 2
4 draw solid black 8 B a 4 4 2
5 draw solid black 8 B a -8 4 2
6 draw solid black 1 2 a 0 500 2
Building on an Existing Data Table

Supine: Change from baseline on Day 1 - SAR407899 15 mg BID

Baseline is the Day 1 predose value.
Only planned values are included in the plot.

• Value was measured with time deviations (<8 min in supine or not 2, 3 or 4 min in standing position).