My Graphics Inventions: You Can Use Them, Too

LeRoy Bessler PhD

Bessler Consulting and Research

Visual Data Insights™

Strong Smart Systems™

Le_Roy_Bessler@wi.rr.com

See also Anything Anywhere All At Once - Output Graphs, Tables, and Text As Any Composite You Wish

Visual Data Insights and Strong Smart Systems are trademarks of LeRoy Bessler PhD

These slides are about the book
Visual Data Insights Using SAS ODS Graphics:
A Guide to Communication-Effective Data Visualization
By LeRoy Bessler

The epigraph from my book

Let your computer draw a picture
to let viewers see the data
with image for an easy, immediate impression
of what's larger, what's smaller
of what the trend is
of what the relationship is
of what the distribution is
including geographically
and with precise numbers for correct, reliable understanding.

Let it paint a picture
that shows the viewer what's important.

**Closing Value of IBM Shares First Trading Day of Each Month in 1999
Critical Points and Last Change**

209,APR99

183,JAN99

108,DEC99

98,OCT99   +5

Epigraph Purpose and Content

► Two important design principles:
  Visual + Precise Numbers
  Show the Viewer What's Important

► A list of graph types

► An example of the two principles

Show the Viewer What's Important

► For time series, annotate essential data points, and the latest change

► For categorical data, use ranking and optionally subsetting

The book carries out its mission with dozens of design principles and hundreds of examples. Precise Numbers and Ranking are among the most frequent principles implemented.

These slides include only a few of the book's principles for communication-effective graphic design and communication-effective use of color.

With perhaps the exception of the fringe plot overlay on a histogram, it consists of examples of uncommon graphs that are, I think, unlikely to have been shown elsewhere. It's impossible to survey all graphic images ever created to prove no prior identical or similar graphic art exists.

Unremarkable Graphs Omitted Here:
All of the familiar graphs, done my way
Panels, Lattices, and Matrices of Graphs
Free-Form Composite of Graphs, Tables,
and Text

Uncommon Graphs Omitted Here:
Pac-Man Pie Chart (Extremes Of Other)
Triptych of Two-Bite Donut Charts
Visual Proof of the Danger of 3D Pie Charts
(I **like** pie charts just like the hundreds of
millions of people who **see** them
and **understand** them, if done well.)

The book tries to touch all the bases, i.e., show <u>all</u> of the kinds of graphs, plots, and charts that ODS Graphics can do, but tries to do them better, even when not as a new design.

It might be that more useful to more readers are the examples omitted here. But there is too much to show.
So I hope this preview will interest the reader in the book. End of Introduction

Certainty, Clarity, Completeness

Visuals for quick easy inference

Precise Values for correct inference

Common Obstacle to Certainty or Clarity
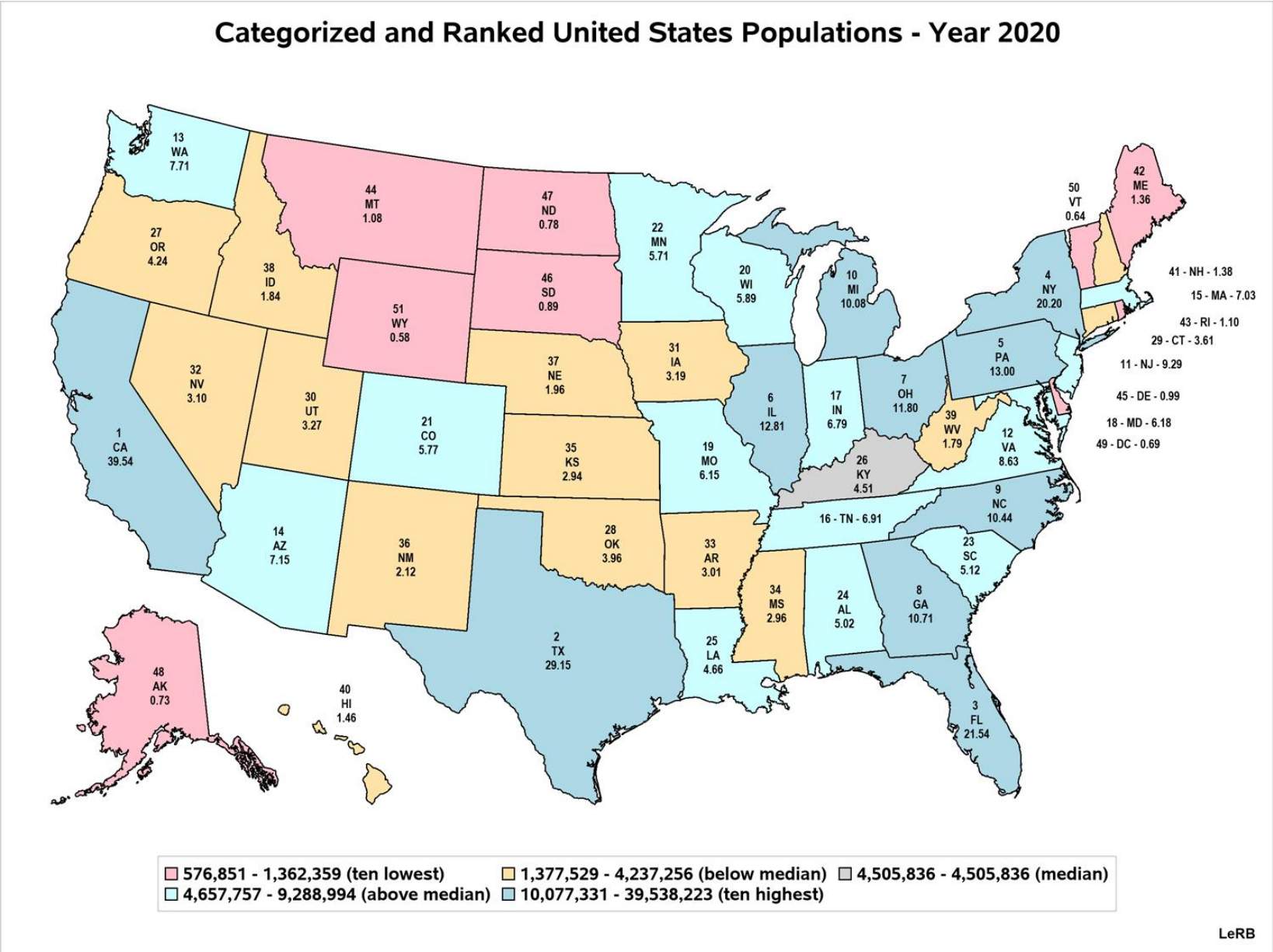
Poor Color Use

Text Readability

➢ **Sans Serif Font Big and Bold**

   11 pt Arial Bold (whenever it fits)

➢ **Maximize contrast between text color and background color**

Assure Color Distinguishability

➤ **Inability to distinguish Red and Green is commonest color blindness**

➤ **Thick Enough Text**

➤ **Thick Enough Lines**

➤ **Big Enough Plot Markers**

➤ **Big Enough Legend Color Swatches**

➤ **Discrete Legend Colors**

➤ **NOT a Continuous Color Gradient Legend**

As In the Book: Dull Color Palette for Sufficient Contrast with Annotation, But Legend Color Swatches Are Too Small
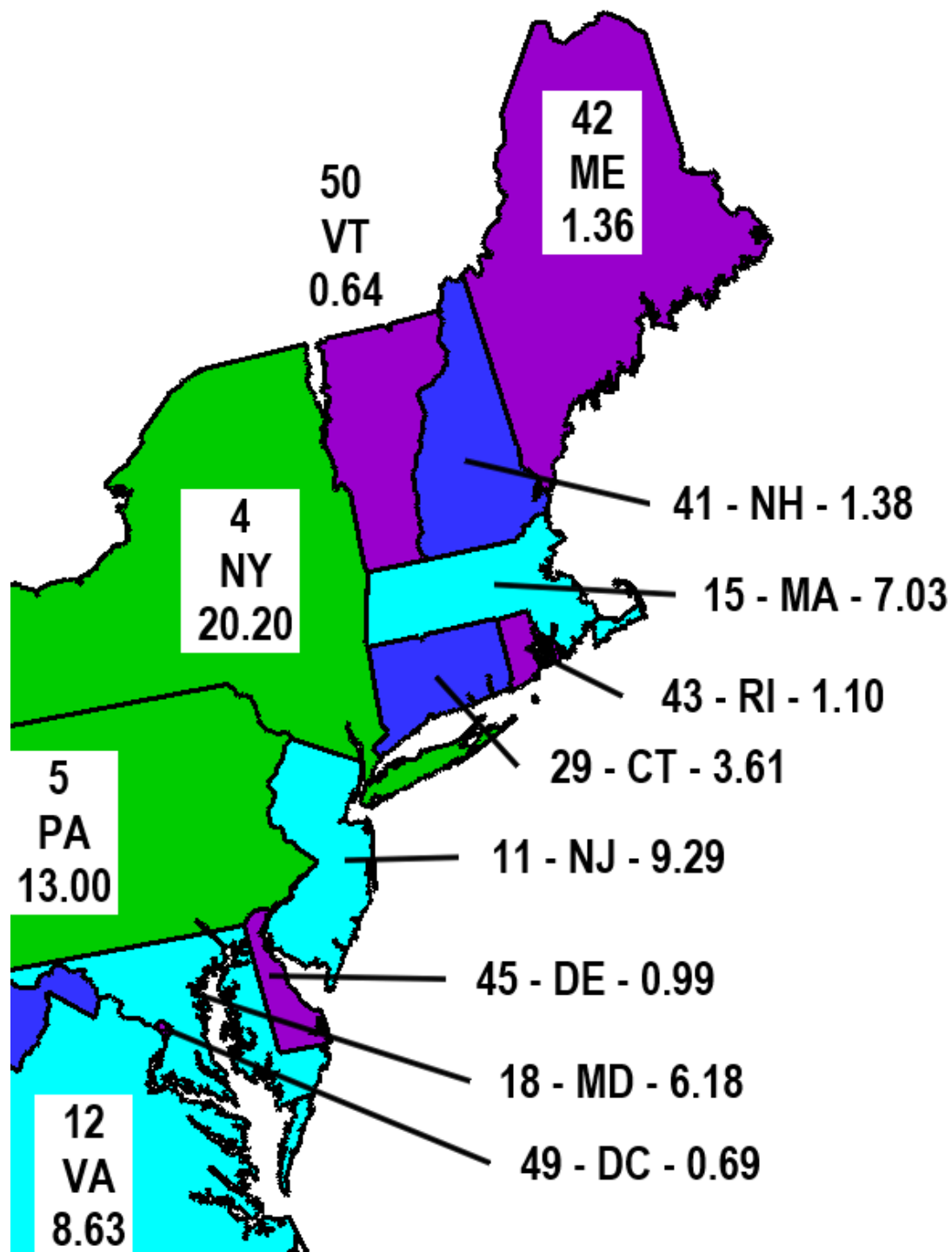

Categorized and Ranked United States Populations - Year 2020

Rank	State	Value
13	WA	7.71
44	MT	1.08
47	ND	0.78
42	ME	1.36
50	VT	0.64
27	OR	4.24
38	ID	1.84
51	WY	0.58
46	SD	0.89
22	MN	5.71
20	WI	5.89
10	MI	10.08
4	NY	20.20
32	NV	3.10
30	UT	3.27
37	NE	1.96
31	IA	3.19
5	PA	13.00
7	OH	11.80
1	CA	39.54
21	CO	5.77
35	KS	2.94
19	MO	6.15
6	IL	12.81
17	IN	6.79
39	WV	1.79
12	VA	8.63
26	KY	4.51
9	NC	10.44
14	AZ	7.15
36	NM	2.12
28	OK	3.96
33	AR	3.01
23	SC	5.12
34	MS	2.96
24	AL	5.02
8	GA	10.71
48	AK	0.73
40	HI	1.46
2	TX	29.15
25	LA	4.66
3	FL	21.54

41 - NH - 1.38
15 - MA - 7.03
43 - RI - 1.10
29 - CT - 3.61
11 - NJ - 9.29
45 - DE - 0.99
18 - MD - 6.18
49 - DC - 0.69
16 - TN - 6.91

Legend:
- 576,851 - 1,362,359 (ten lowest)
- 1,377,529 - 4,237,256 (below median)
- 4,505,836 - 4,505,836 (median)
- 4,657,757 - 9,288,994 (above median)
- 10,077,331 - 39,538,223 (ten highest)

LeRB

Categorized and Ranked United States Populations - Year 2020

SAS M08 enables better legend (using multi-color footnotes)
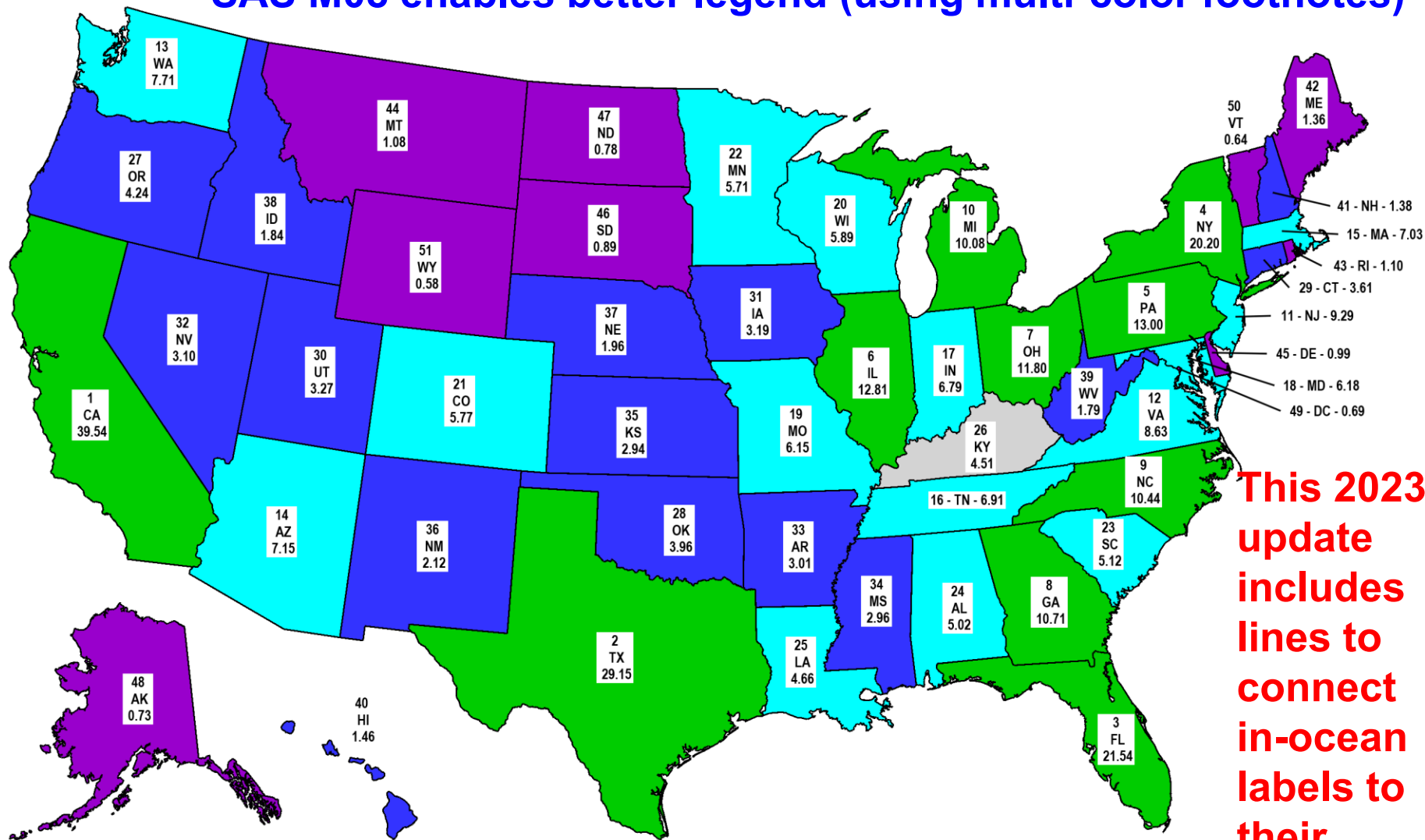
This 2023 update includes lines to connect in-ocean labels to their states.

576,851 - 1,362,359 (ten lowest)	**1,377,529 - 4,237,256 (below median)**	**4,505,836 - 4,505,836 (median)**
4,657,757 - 9,288,994 (above median)	**10,077,331 - 39,538,223 (ten highest)**	

LeRB

Data Label Limitations for Time Series Plots

With many points per line and/or many lines,
you can get:
label-line collisions
label-label collisions

The Safest Annotation
no line for labels to collide with

Labels Visually Show the Trend AND Provide Precise Numbers

Close Price for IBM Shares on First Trading Day Each Month - 1998

184
Dec

165
Nov

149
Oct

133
Jul

129
Sep

118
May

116
Apr

115
Jun

113
Aug

104
Feb

104
Mar

99
Jan

0
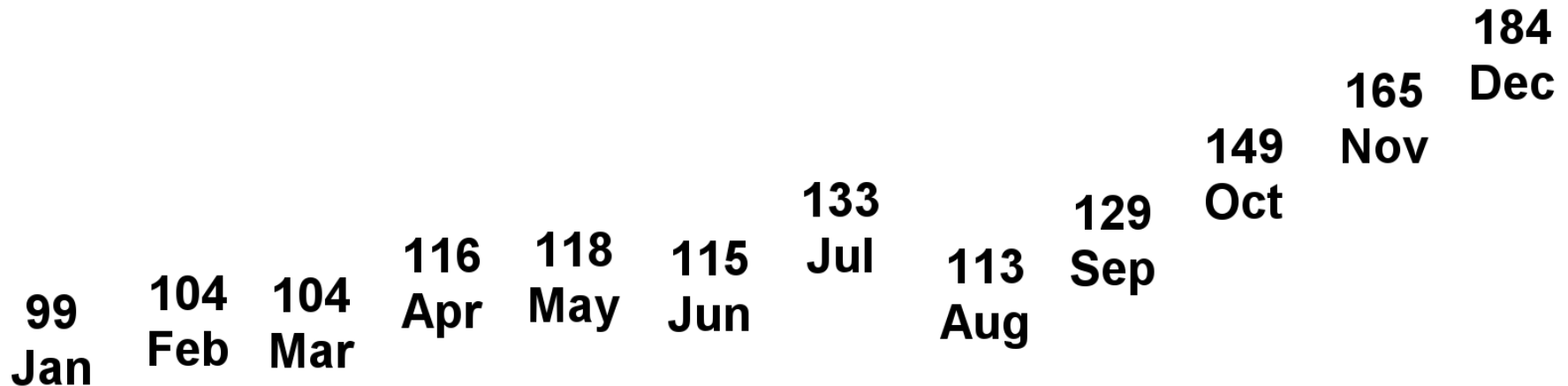
In Single-Line Time Series Plots
Unless minimum Y value is negative,
Start Y axis at 0***
to prevent magnification of what can
be changes of no great significance
*****Exception Examples Shown Later**

Close Price for IBM Shares on First Trading Day Each Month - 1998

184
Dec

165
Nov

149
Oct

133
Jul

129
Sep

116
Apr

118
May

115
Jun

113
Aug

99
Jan

104
Feb

104
Mar
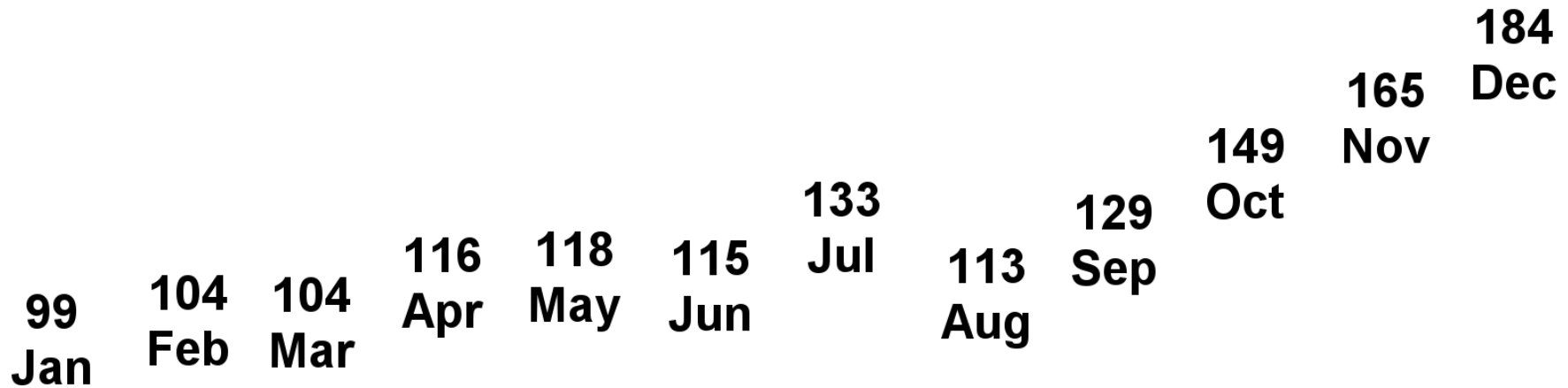
0

Start Y axis at Zero (and <u>Show</u> 0)
Present the trend truly,
with no magnification of change

Showing 0 on the Y axis when starting Y at 0 is not essential, but
its presence does no harm and
addresses any uncertainty that the viewer might have about fairness of the
visual presentation of the data.

For Safe Annotation

minimizing the number of labels minimizes possible collisions

Show Them What Is Important

For Time Series Plots, Annotate the Always Essential Information:

➤ **Latest Point-to-Point Change**

➤ **Critical Data Points: Start, End, Intermediate Max and/or Min**

➤ **If NO Max and NO Min, annotate a point where slope changes significantly and permanently, if that happens**

Show Them What Is Important
with Start, End, Min, Max, Key Change Point

Sparsely Annotating Start, End, Min, Max.

Annual USA Beer Consumption
Gallons per Capita

24.6

22.5

23.3

0

1977   1981   1989
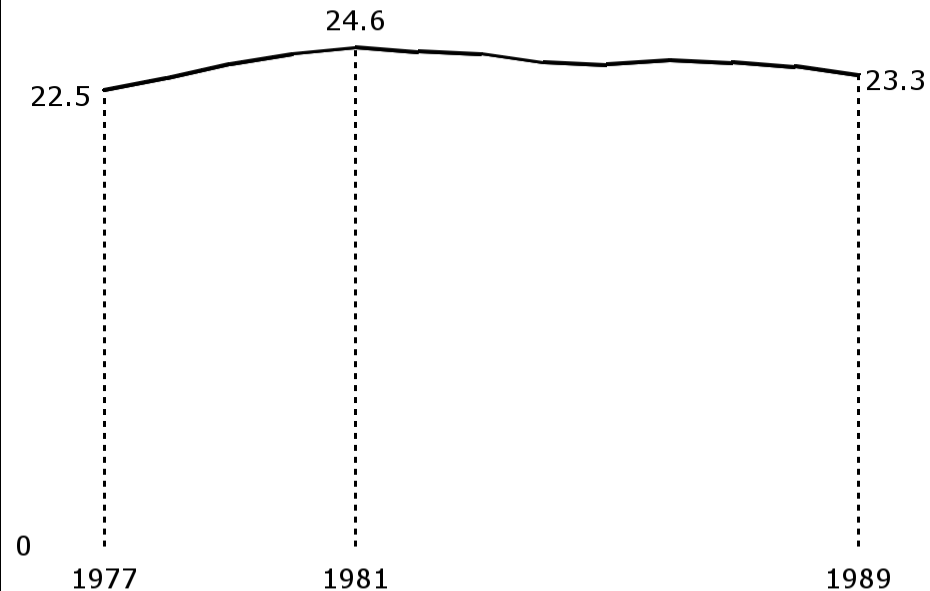
Data From: "Beverage Industry", February 1990

Sparsely Annotating Start, End, Growth Rate Change

Production of a Major Brand Lite Beer
Millions of Barrels

20.0

16.3

7.5

0

1977   1981   1989
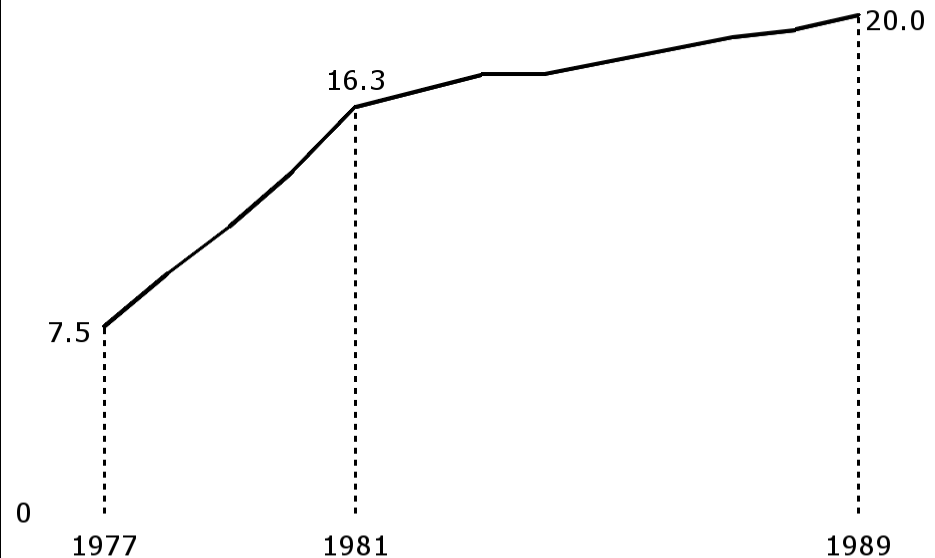
Data From: Estimates By LeRoy Bessler

This was LeRB's Sparse Line Annotation in 1992
Zelchenko's "Spark Line" appeared in 1998

Exception to Start Y at Zero On Sparse Lines (e.g., <u>Next</u> Slide)

The most important change is annotated, which prevents visual misinterpretation of change significance when Y axis starts at minimum Y, not at 0.

The Sparse Line

Closing Value of IBM Shares First Trading Day of Each Month in 1999
Critical Points and Last Change

209,APR99

183,JAN99

108,DEC99

98,OCT99

+5
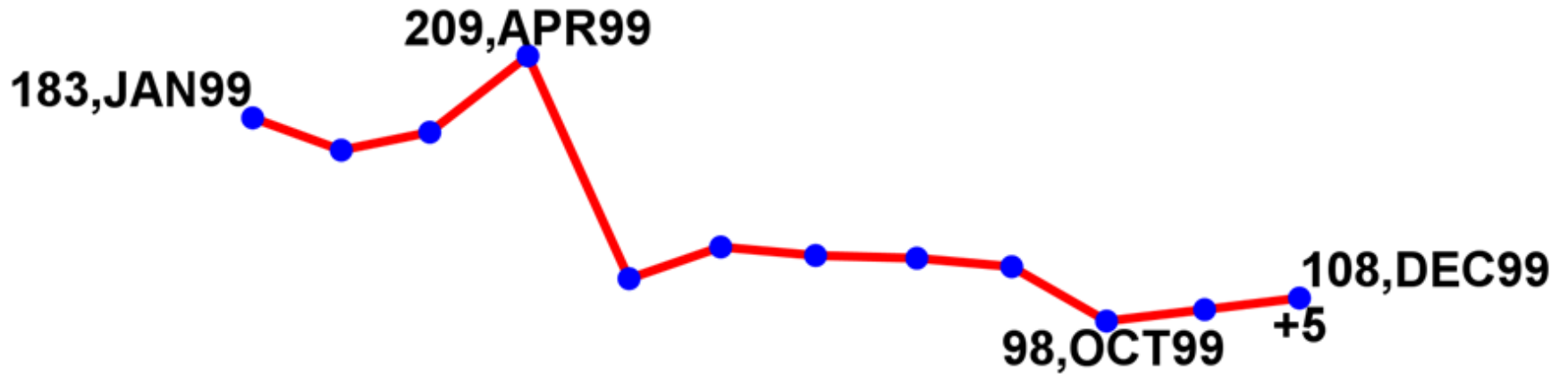
The Sparse Line Is <u>Informative</u>

➢ **Visual AND precise numbers:**
   quick easy overview of history
   AND precise knowledge
   of <u>Critical</u> Points

➢ <u>Sufficiently</u> Informative

The Sparse Line is not a SparkLine

➢ Zelchenko's SparkLine

 is a <u>bare</u> line

  ➢ No Y axis

  ➢ No annotation

➢ A visual accessory

 in a table column

➢ It has value in THAT context

**Exception to Start Y at Zero
On Multi-Line Plots (e.g., Next Slide)**
Start at 0 would cause needless squishing of the lines
AND data labels would be more likely to collide with other data labels or with other lines

Sparse Annotation for Overlay Multi-Line Plot
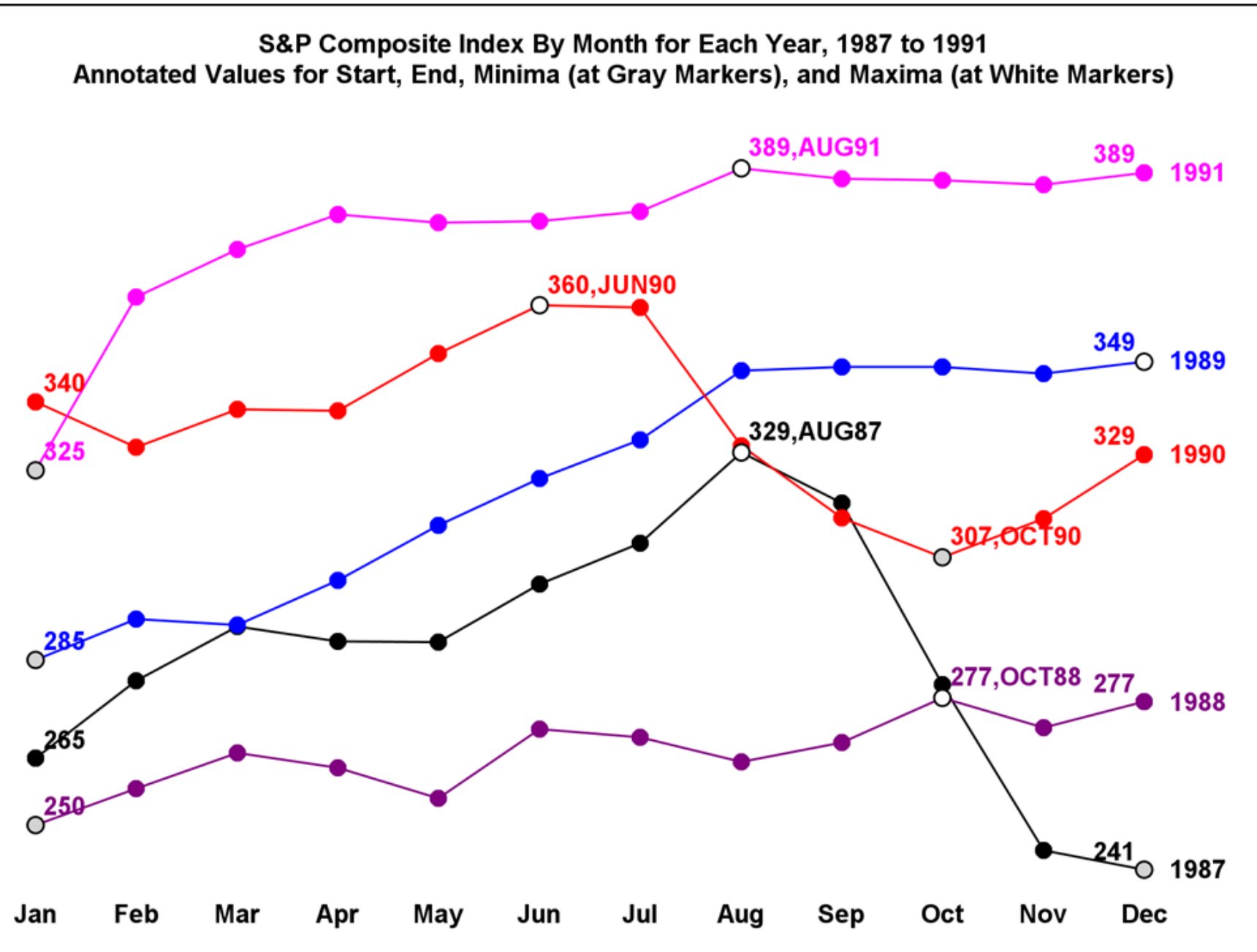
S&P Composite Index By Month for Each Year, 1987 to 1991
Annotated Values for Start, End, Minima (at Gray Markers), and Maxima (at White Markers)

S&P Composite Index By Month for Each Year, 1987 to 1991
Values for Start, End, Minima (at Gray Markers), and Maxima (at White Markers)

340

325

360,JUN90

Close Up of All 1990 points
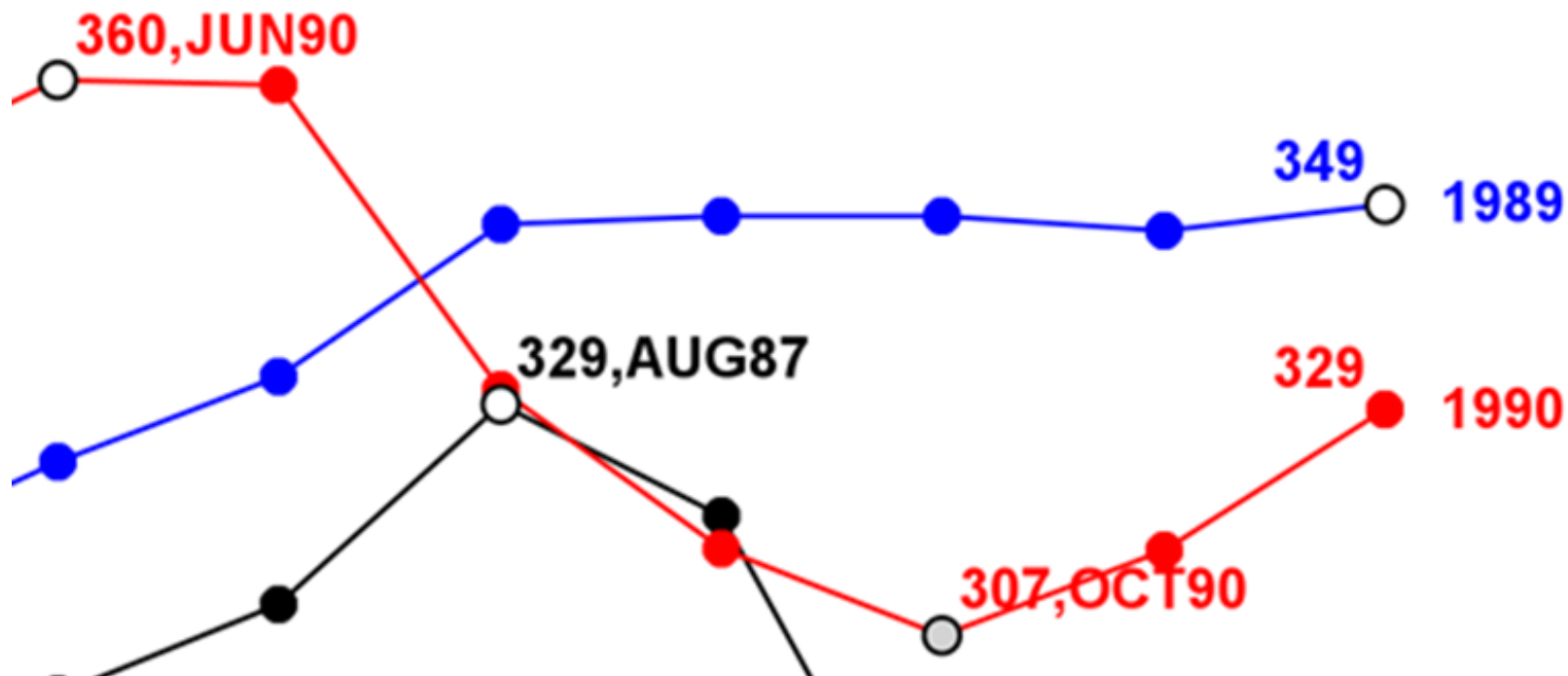
360,JUN90

349

1989

329,AUG87

329

1990

307,OCT90

S&P Composite Index By Month for Each Year, 1987 to 1991
Values for Start, End, Minima (at Gray Markers), and Maxima (at White Markers)

389 • 1991

Useful Potential Improvement:
+3 "manually" added here ►
Annotate below each end point
the latest point-to-point change
as in the Sparse Line

349 ○ 1989
+3

329 • 1990
OCT90

OCT88 277 • 1988

OCT88 • 1988

241 ○ 1987
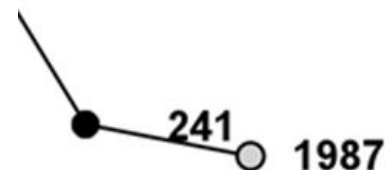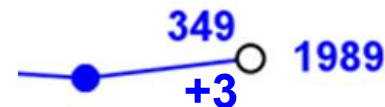
Sparse Annotation + X Axis Table = Everything
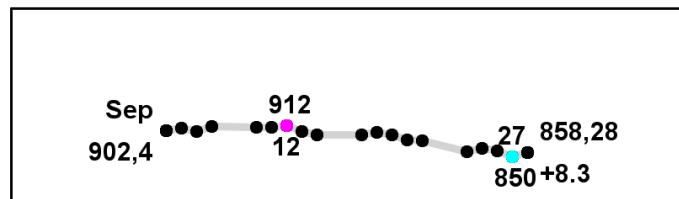## Nov & Dec Data: No Need for LastChange Anno

S&P Composite Index By Month for Each Year, 1987 to 1991
Annotated Values for Start, End, Minima (at Gray Markers), and Maxima (at White Markers)
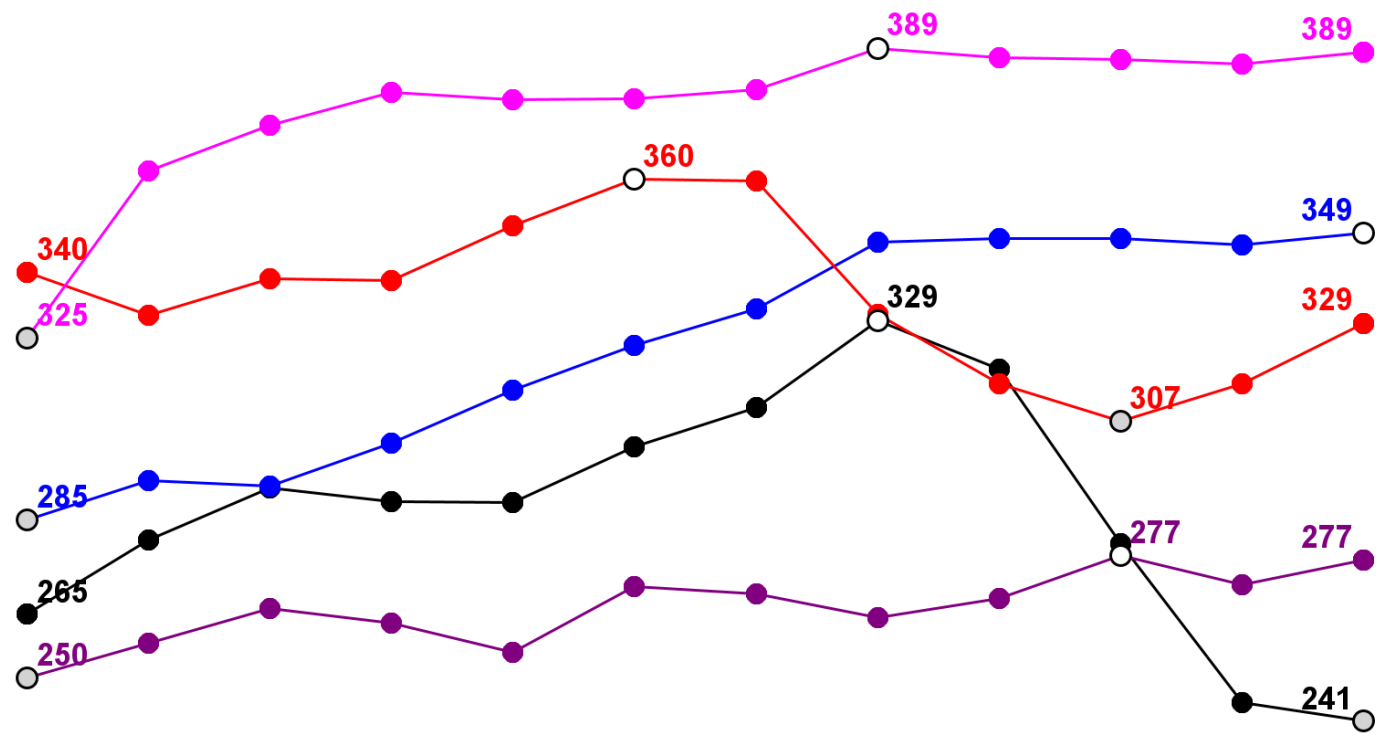
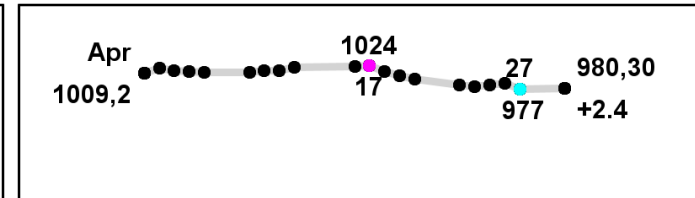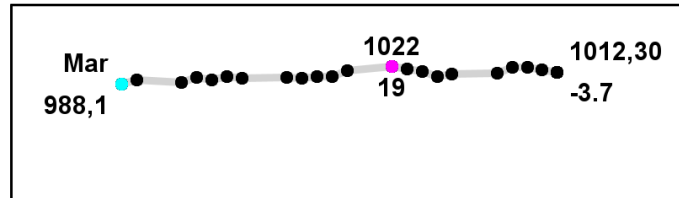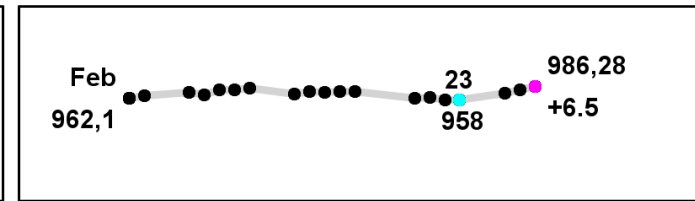|	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1991	325	362	372	380	378	378	380	389	387	387	386	389
1990	340	330	338	338	350	360	360	331	315	307	315	329
1989	285	294	293	302	314	324	332	347	347	347	346	349
1988	250	258	266	263	256	271	269	264	268	277	271	277
1987	265	281	292	289	289	301	310	329	319	280	245	241

Sparse Line Table for Twelve Months of Trading Months

Dow Jones Composite Index By Trading Day By Month in 1990
Maximum Color Is Magenta, Minimum Is Turquoise
Index and Day Values for First, Last, Min, Max
Last Annotation Includes Latest Change

Jan
1054,2
30   960,31
943   +16.9

Feb
962,1
23   986,28
958   +6.5

Mar
988,1
1022
19
1012,30
-3.7

Apr
1009,2
1024
17
27   980,30
977   +2.4

May
985,1
1040,31
+1.0

Jun
1049,1
1064
4
26   1031,29
1019   +1.5

Jul
1035,2
1064
16
1023,31
-8.5

Aug
1023,1
23   904,31
864   +5.9

Sep
902,4
912
12
27   858,28
850   +8.3

Oct
883,1
11
839
891
22
863,31
-0.1

Nov
869,1
8
861
897
14
893,30
+17.6

Dec
899,3
926
26
921,31
+0.2

a typical sparse line table cell

Why Gaps? The New York Stock Exchange is closed on Saturday, Sunday, and Thanksgiving Holiday

Simplest Vertical "Bar" Chart

Ranked Shoe Sales ($M)and Percent Share By Region - Total $33.85

$5.63
16.6%
Middle
East

$5.50
16.3%
United
States

$4.87
14.4%
Western
Europe

$4.26
12.6%
Canada

$3.66
10.8%
Central
America
Caribbean

$2.43
7.2%
South
America

$2.39
7.1%
Eastern
Europe

$2.34
6.9%
Africa

$2.30
6.8%
Pacific

$0.46
1.4%
Asia

0
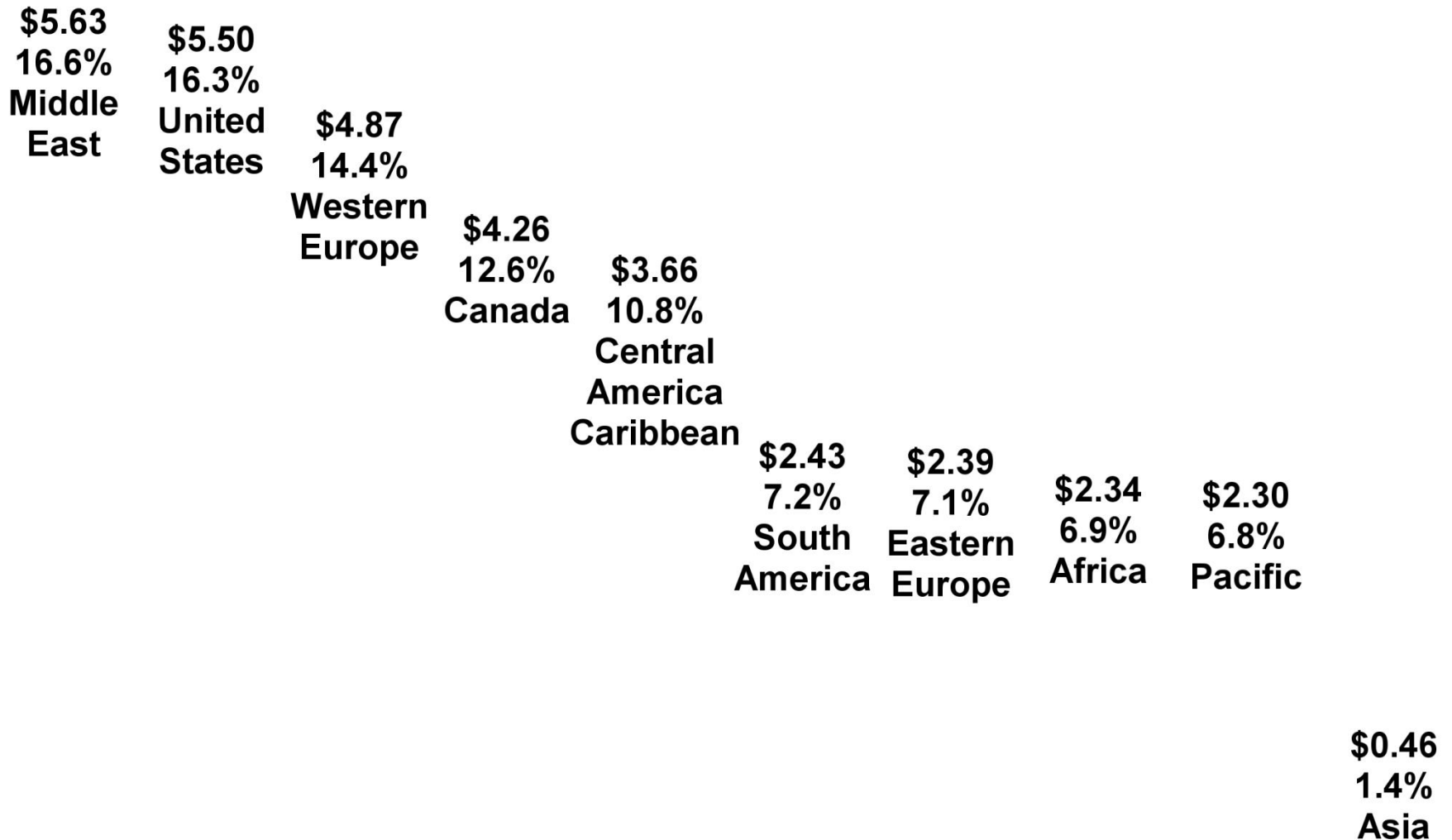
▼ Showing That Minimum Y = 0

Ranked Shoe Sales ($M)and Percent Share By Region - Total $33.85

$5.63
16.6%
Middle
East

$5.50
16.3%
United
States

$4.87
14.4%
Western
Europe

$4.26
12.6%
Canada

$3.66
10.8%
Central
America
Caribbean

$2.43
7.2%
South
America

$2.39
7.1%
Eastern
Europe

$2.34
6.9%
Africa

$2.30
6.8%
Pacific

$0.46
1.4%
Asia

0
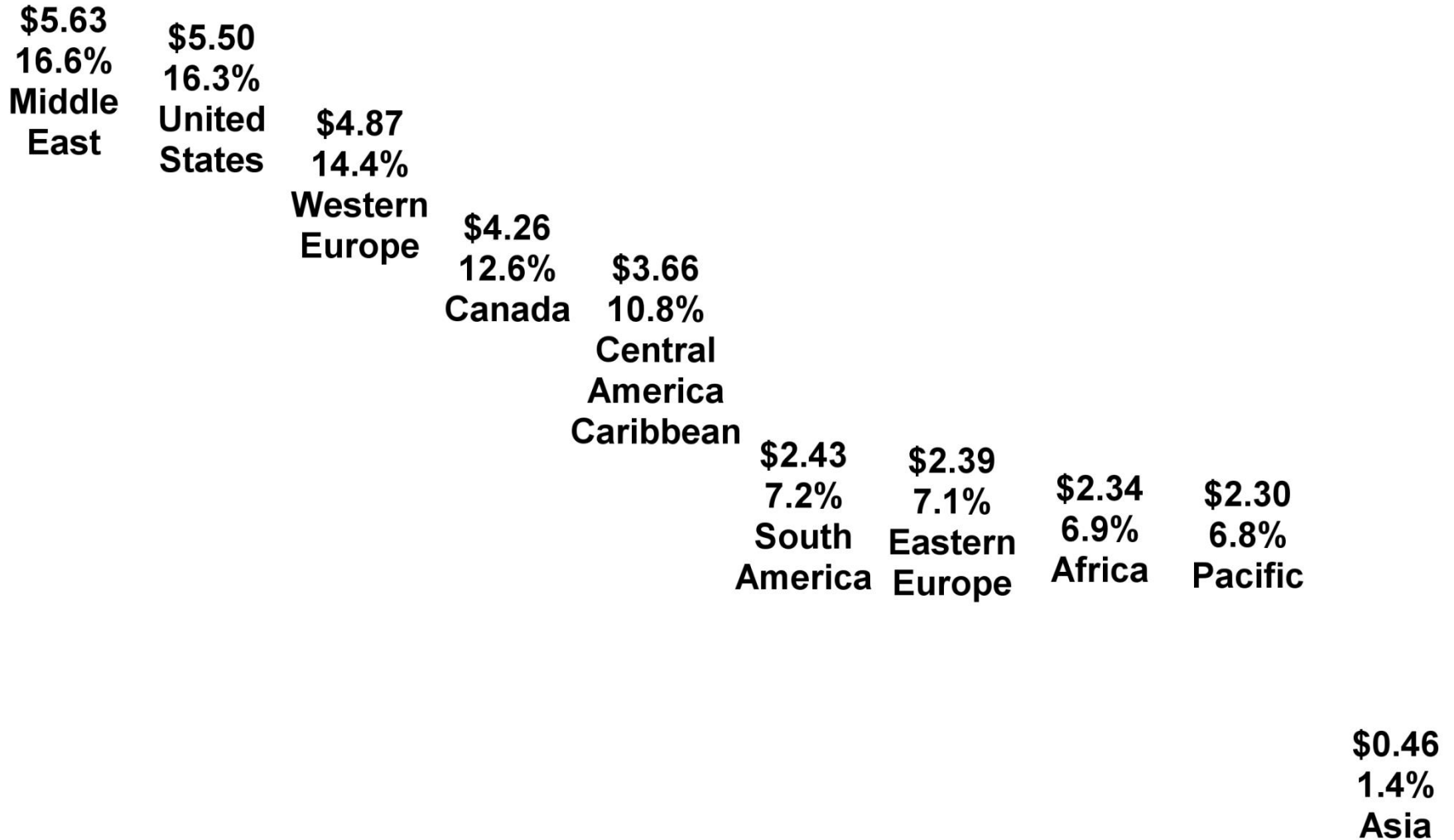
▼ Prefer This Axis Information?

Ranked Shoe Sales ($M)and Percent Share By Region - Grand Total $33.85

Sales

$5.63

$5.63
16.6%
Middle
East

$5.50
16.3%
United
States

$4.87
14.4%
Western
Europe

$4.26
12.6%
Canada

$3.66
10.8%
Central
America
Caribbean

$2.43
7.2%
South
America

$2.39
7.1%
Eastern
Europe

$2.34
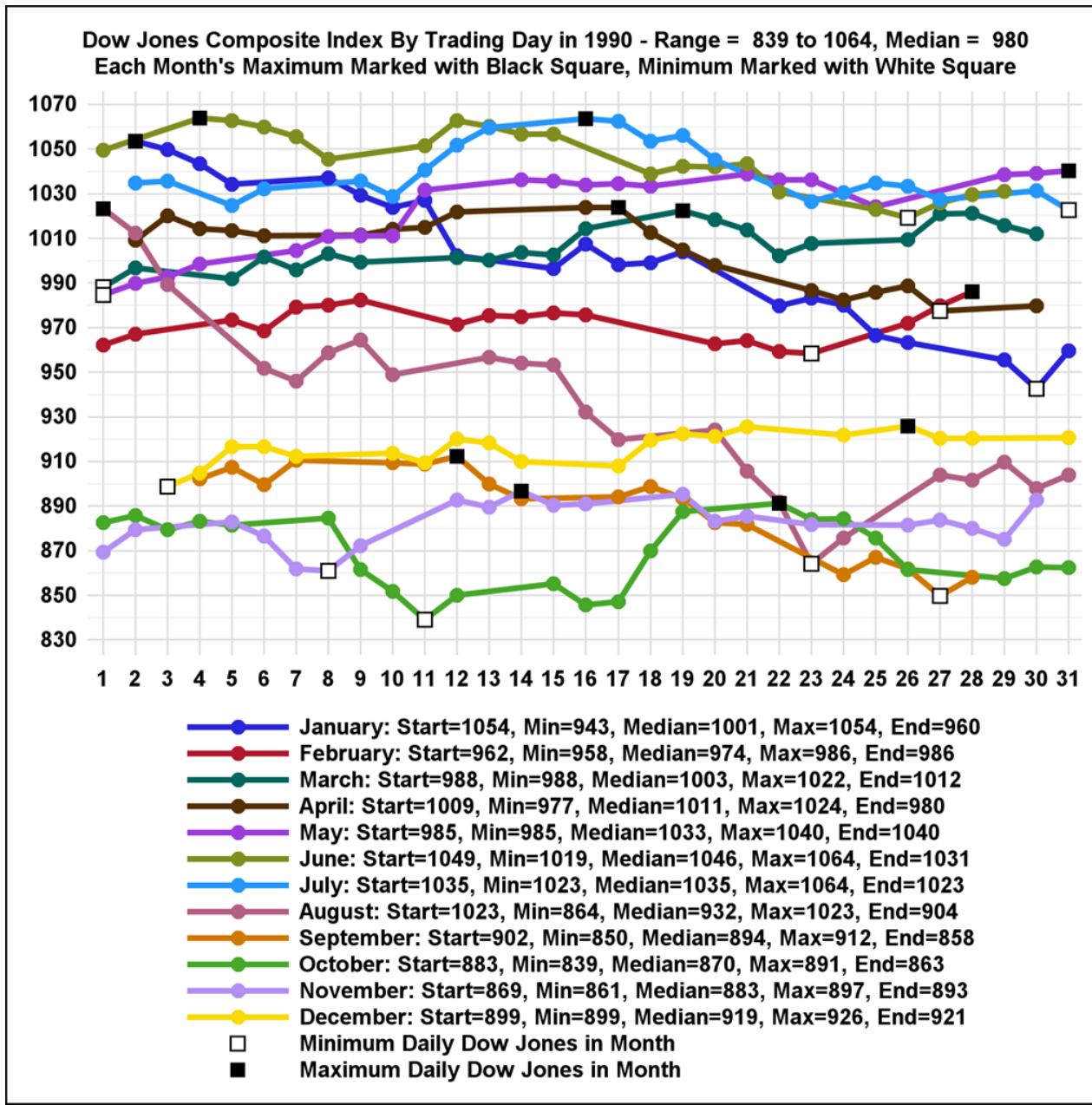6.9%
Africa

$2.30
6.8%
Pacific

$0.46
1.4%
Asia

0
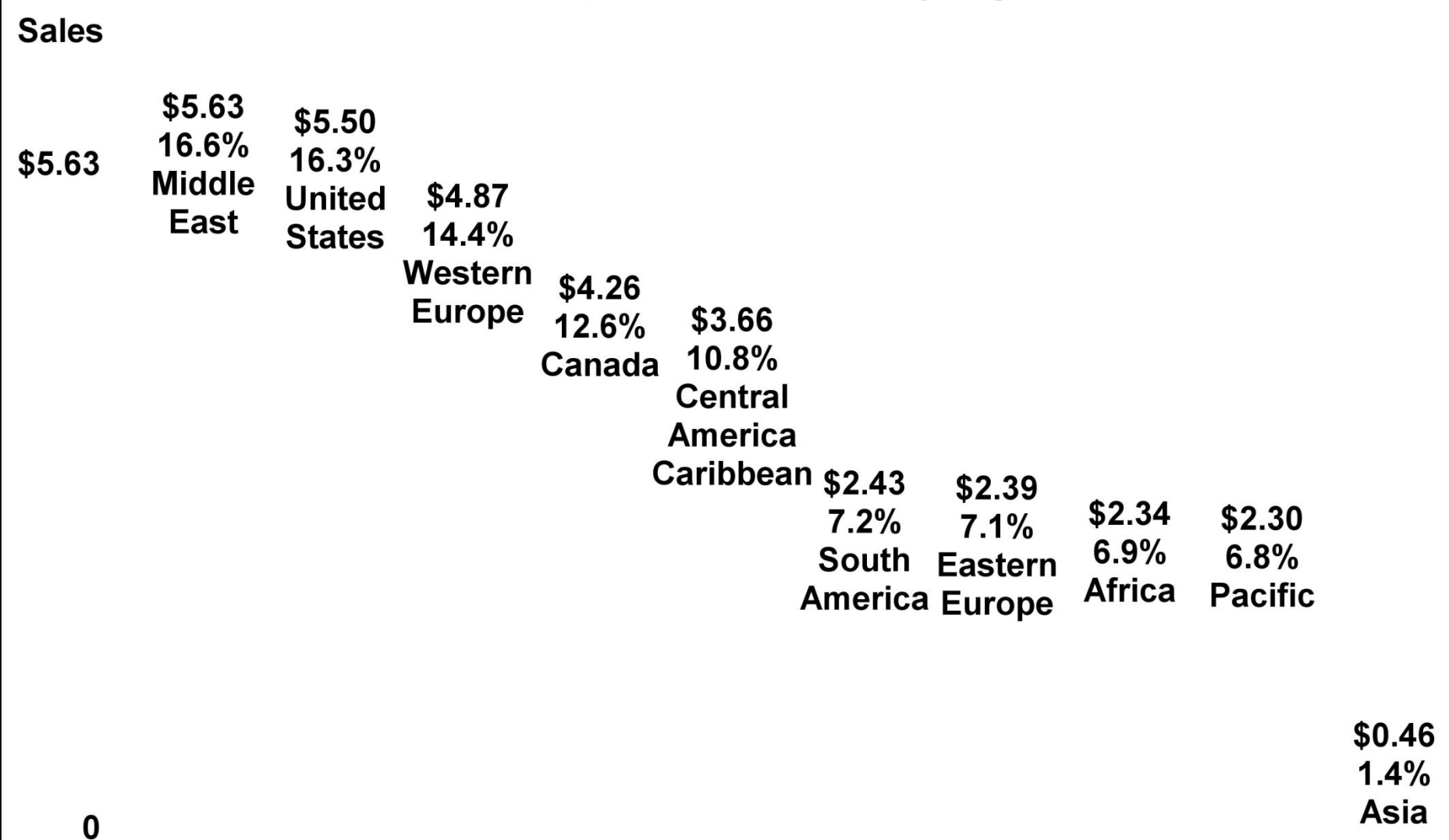
What to do when NO annotation is possible?

Of course, a web graph with mouseover text (aka data tips) is always an option, but data tips are transient, and a companion table is really mandatory.
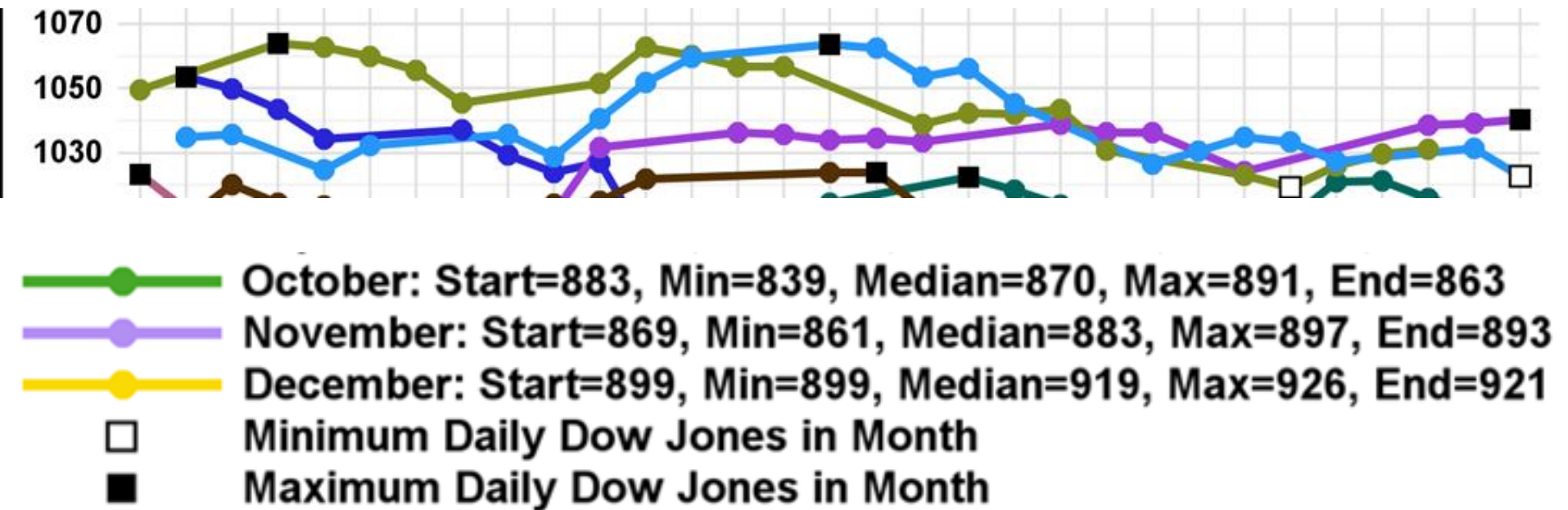
So what's a possible compromise alternative?

Maximally Informative Legend

Dow Jones Composite Index By Trading Day in 1990 - Range = 839 to 1064, Median = 980
Each Month's Maximum Marked with Black Square, Minimum Marked with White Square

January: Start=1054, Min=943, Median=1001, Max=1054, End=960
February: Start=962, Min=958, Median=974, Max=986, End=986
March: Start=988, Min=988, Median=1003, Max=1022, End=1012
April: Start=1009, Min=977, Median=1011, Max=1024, End=980
May: Start=985, Min=985, Median=1033, Max=1040, End=1040
June: Start=1049, Min=1019, Median=1046, Max=1064, End=1031
July: Start=1035, Min=1023, Median=1035, Max=1064, End=1023
August: Start=1023, Min=864, Median=932, Max=1023, End=904
September: Start=902, Min=850, Median=894, Max=912, End=858
October: Start=883, Min=839, Median=870, Max=891, End=863
November: Start=869, Min=861, Median=883, Max=897, End=893
December: Start=899, Min=899, Median=919, Max=926, End=921
□ Minimum Daily Dow Jones in Month
■ Maximum Daily Dow Jones in Month

Maximally Informative Legend

Each Month's Maximum Marked with Black Square,
Minimum Marked with White Square

October: Start=883, Min=839, Median=870, Max=891, End=863
November: Start=869, Min=861, Median=883, Max=897, End=893
December: Start=899, Min=899, Median=919, Max=926, End=921
☐ Minimum Daily Dow Jones in Month
■ Maximum Daily Dow Jones in Month

You Can Make a Legend Do Anything That You Wish

It need not be limited to identifying the graphic element entities.

ADD Any Information That You Wish

Categorical Data

➢ **Rank/Order the bars, pie slices, donut bites Exception: Alphabetic Order for Easy LookUp (include rank number)**

➢ **Make it Maximally Informative:**

    ➢ **Provide Category, Value, Percent, if many categories—Rank Number**

    ➢ **In title or subtitle, show Grand Total**

➢ **A Subsetted Bar Chart might suffice**

For Categorical Data:
Horizontal Bar Chart Is Best

Lots of space to fit the Rank Number, Category, Value, and Percent at left of each bar

Maximally Informative Horizontal Bar Chart, Subsets Can Suffice*

All have similar design and titles

Not Shown Here:
► Top Ten
► Cities with At Least $1M Sales

*Show Them What's Important

Top 15 Ranked Shoes Sales By City
Selecting Only Enough for At Least 60% of Total Sales
SubTotal Sales $21,046,712 is 62.2% of Total
All 53 Cities had Total Sales $33,851,566

Rank		Sales	Share	
1	Vancouver	$3,227,768	9.5%	
2	Tel Aviv	$2,567,568	7.6%	
3	Kingston	$2,235,204	6.6%	
4	Dubai	$1,910,544	5.6%	
5	Chicago	$1,565,585	4.6%	
6	New York	$1,489,207	4.4%	
7	Al-Khobar	$1,153,667	3.4%	
8	Minneapolis	$1,099,937	3.2%	
9	Heidelberg	$967,739	2.9%	
10	Lisbon	$898,345	2.7%	
11	Manila	$854,904	2.5%	
12	Caracas	$789,323	2.3%	
13	Warsaw	$786,714	2.3%	
14	London	$762,009	2.3%	
15	Cairo	$738,198	2.2%	

Showing Enough

Showing Enough

The particular data used for the prior example regrettably does not illustrate the phenomenon, BUT it often can be the case that, say, 80% or 90% of the measure of interest can be accounted for with a rather small number of categories.

Though a percent chosen can be arbitrary, I think, e.g., Top 10 is more arbitrary.

Dynamic Titles
Answer All Likely Questions

Top 15 Ranked Shoes Sales By City
Selecting Only Enough for At Least 60% of Total Sales
SubTotal Sales $21,046,712 is 62.2% of Total
All 53 Cities had Total Sales $33,851,566

All of the Need-To-Know Information

Rank		Sales	Share
1	Vancouver	$3,227,768	9.5%
2	Tel Aviv	$2,567,568	7.6%
3	Kingston	$2,235,204	6.6%
4	Dubai	$1,910,544	5.6%
5	Chicago	$1,565,585	4.6%

Alphabetic Order (Include Rank)

Shoe Sales, Percent Share, and Rank By Product

	Sales	Share	Rank	
Boot	$2,350,543	6.9%	6	
Men's Casual	$7,933,707	23.4%	1	
Men's Dress	$5,507,243	16.3%	4	
Sandal	$868,436	2.6%	7	
Slipper	$6,175,834	18.2%	3	
Sport Shoe	$651,467	1.9%	8	
Women's Casual	$4,137,861	12.2%	5	
Women's Dress	$6,226,475	18.4%	2	

Stacked Vertical Bar Chart (Pie Chart Alternative)

Ranked Sales & Percent Share By Region

$33,851,566

- 01:MiddleEast:$5,631,779:16.64%
- 02:UnitedStates:$5,503,986:16.26%
- 03:WesternEurope:$4,873,000:14.40%
- 04:Canada:$4,255,712:12.57%
- 05:CentralAmerica/Caribbean:$3,657,753:10.81%
- 06:SouthAmerica:$2,434,783:7.19%
- 07:EasternEurope:$2,394,940:7.07%
- 08:Africa:$2,342,588:6.92%
- 09:Pacific:$2,296,794:6.78%
- 10:Asia:$460,231:1.36%
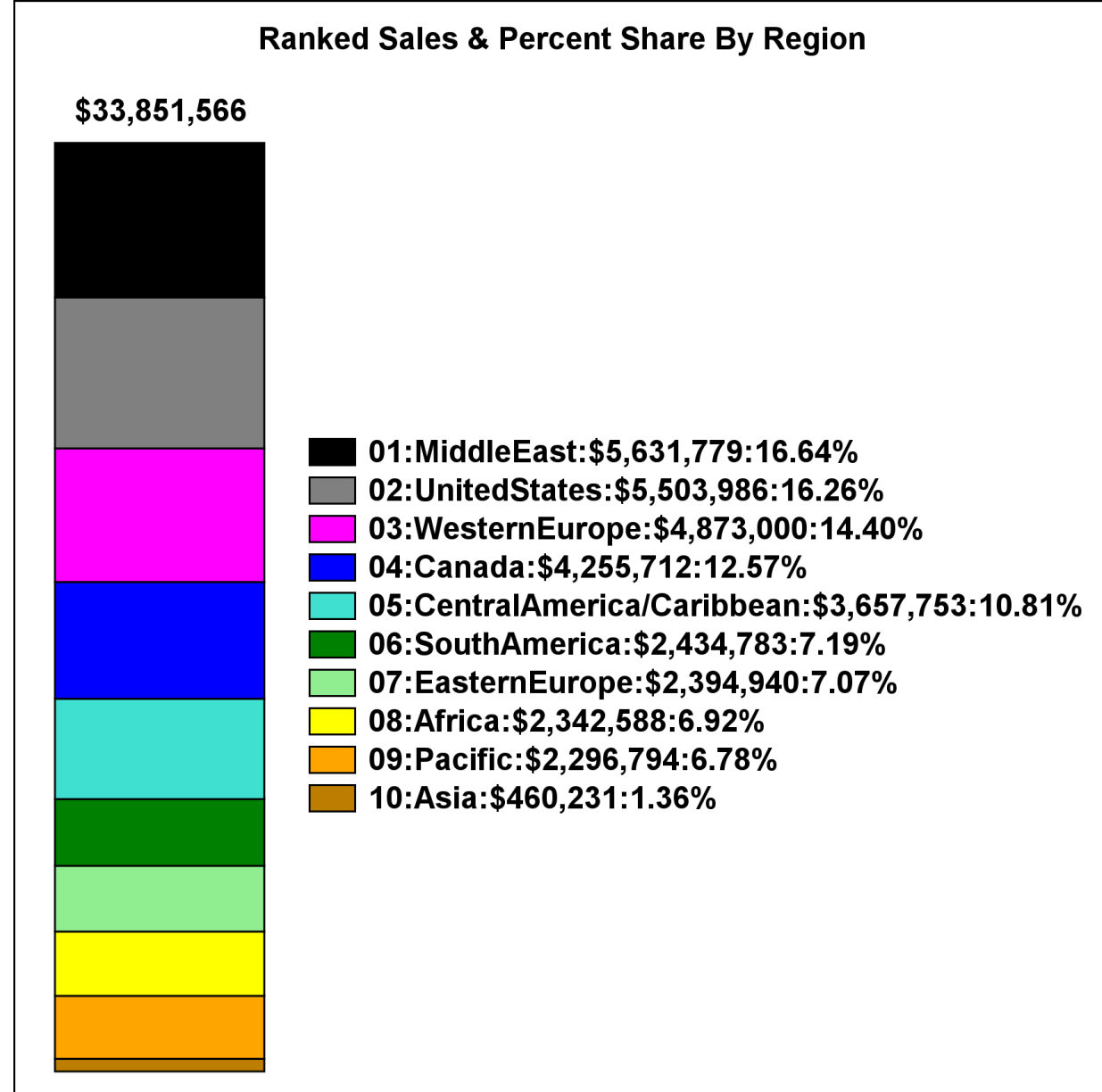
Rank-Ordered Table with a Visual Companion

A **Real** Pie Chart
Maximally Informative and **Never**-Fail

Shoe Sales and Percent Share By Region - Total = $33,851,566

- 1 Middle East $5,631,779 16.6%
- 2 United States $5,503,986 16.3%
- 3 Western Europe $4,873,000 14.4%
- 4 Canada $4,255,712 12.6%
- 5 Central Americ $3,657,753 10.8%
- 6 South America $2,434,783 7.2%
- 7 Eastern Europe $2,394,940 7.1%
- 8 Africa $2,342,588 6.9%
- 9 Pacific $2,296,794 6.8%
- 10 Asia $460,231 1.4%

Rank-Ordered Table with a Visual Companion

Clipping of the Graph Title
Plus Everything Needed to Know

By Region - Total = $33,851,566

- 1 Middle East $5,631,779 16.6%
- 2 United States $5,503,986 16.3%
- 3 Western Europe $4,873,000 14.4%
- 4 Canada $4,255,712 12.6%
- 5 Central Americ $3,657,753 10.8%
- 6 South America $2,434,783 7.2%
- 7 Eastern Europe $2,394,940 7.1%
- 8 Africa $2,342,588 6.9%
- 9 Pacific $2,296,794 6.8%
- 10 Asia $460,231 1.4%

Rank-Ordered Table Companion for Visuals

Simplest Pie Charts

I usually discourage using "Other" for Pie Charts,
to prevent the question:
What is in "Other"?

The Extremes of Other

Simplest, Most Powerful, and Crisply Informative Pie Chart
Other Is Insignificant

Simplest Pie Chart Has Maximum Impact
Here, the Extreme Other is Very Small

SAS Software Dominates 1993
Mainframe Data Analysis Market

SAS - 91%

Other

Simplest Pie Chart Has Maximum Impact
Here, the Extreme Other is Very Small

SAS Software Dominates 1993
Mainframe Data Analysis Market

SAS - 91%

Other

Simplest, Most Powerful, and Crisply Informative Pie Chart Here, Directing Attention <u>TO</u> Other
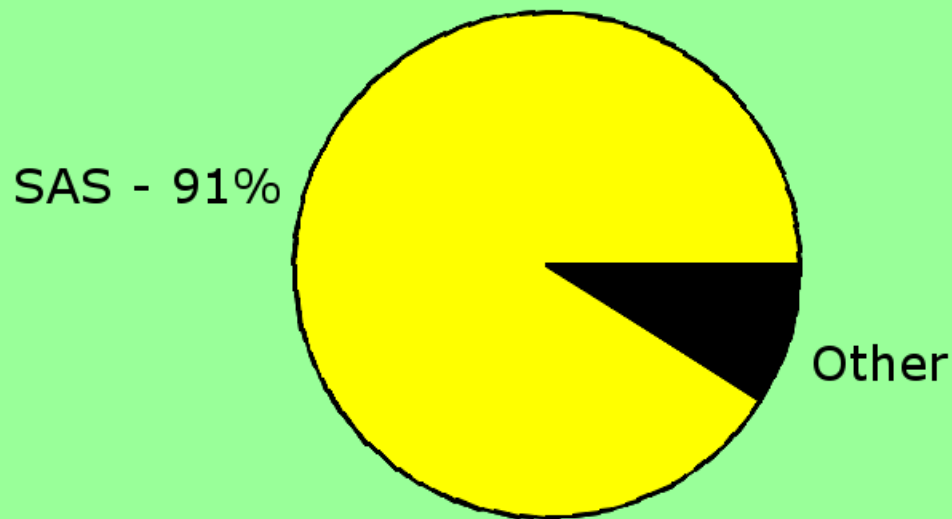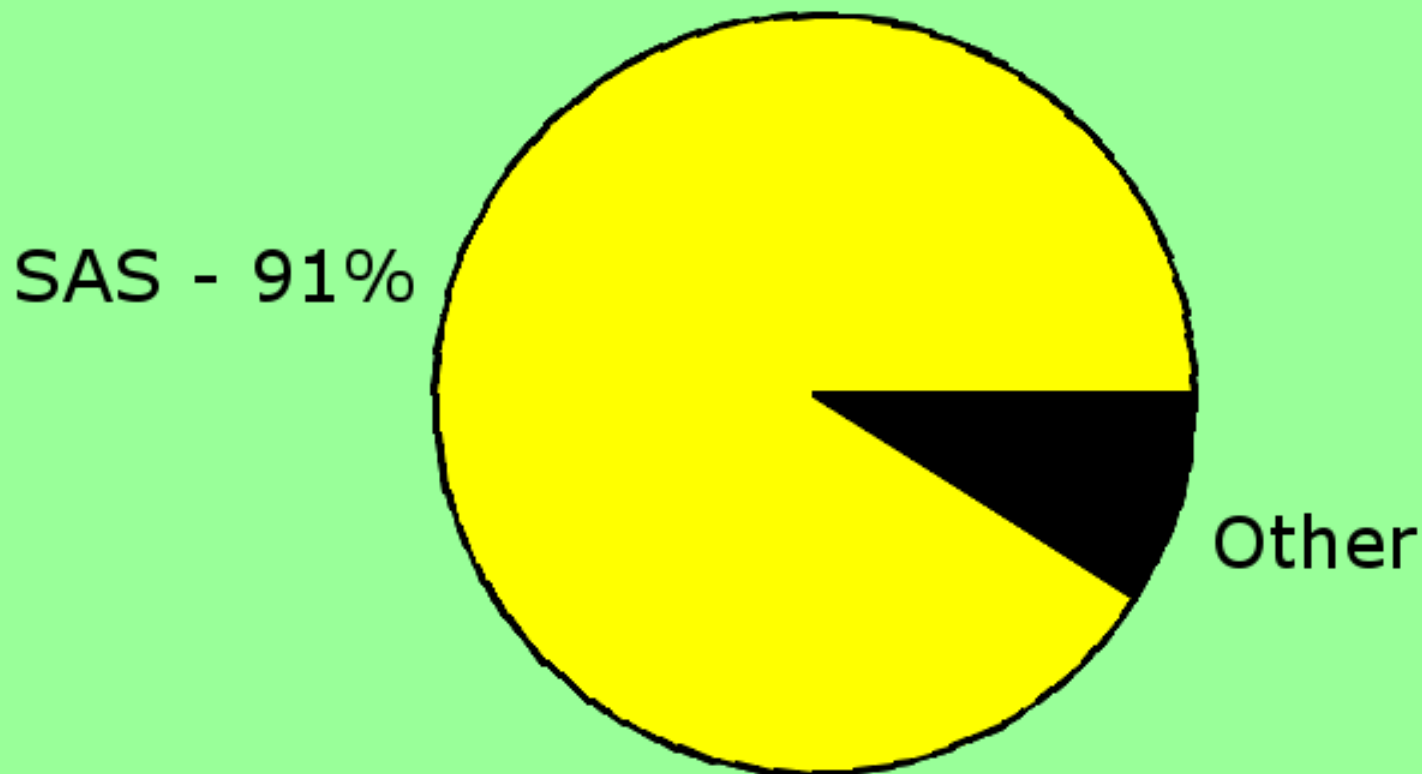
Simplest Pie Chart Has Maximum Impact
Here, the Extreme Other is Very Large

Most of Your Property Tax
is NOT for Local Government

Other Govt.

Village - 21%
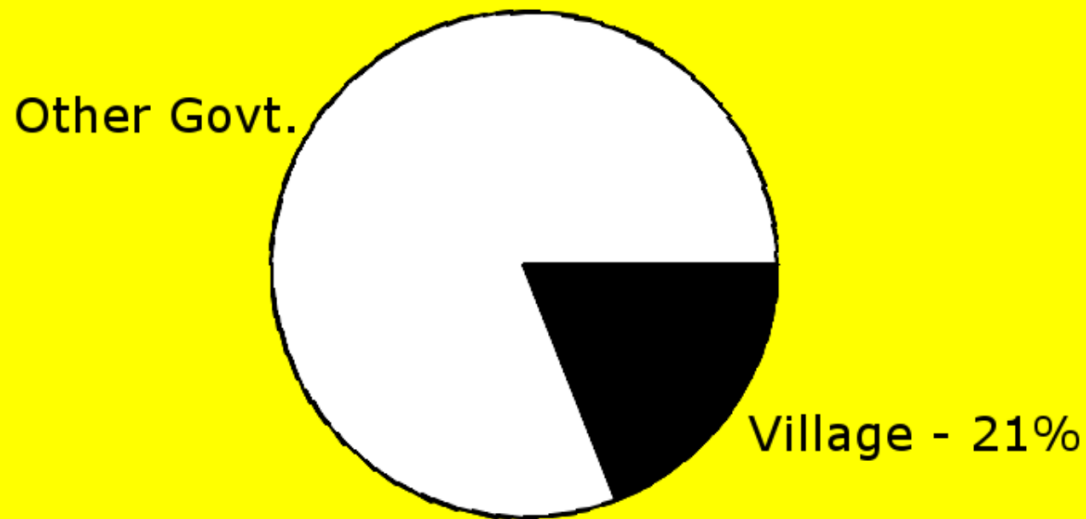
A Triptych of Donut Charts
(Simplest Donut Charts)
Imitative, Not Inventive
(my derivative of a found example)

**City with Largest
Share of Shoe Sales**

Vancouver - $3,227,768

9.5%

Other Cities

**Region with Largest
Share of Shoe Sales**

Middle East - $5,631,779

16.6%
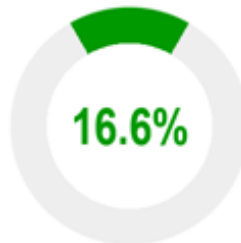
Other Regions

**Product with Largest
Share of Shoe Sales**

Men's Casual - $7,933,707

23.4%
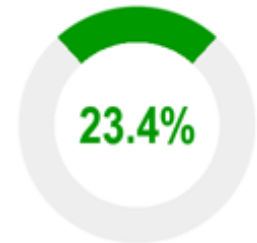
Other Products

For Two Categorical Variables (when NOT using ODS Graphics GROUP option in a Bar Chart)

Heat Maps
Bubble Plots
(but Bubble Plots can also be used for ONE Categorical Variable)
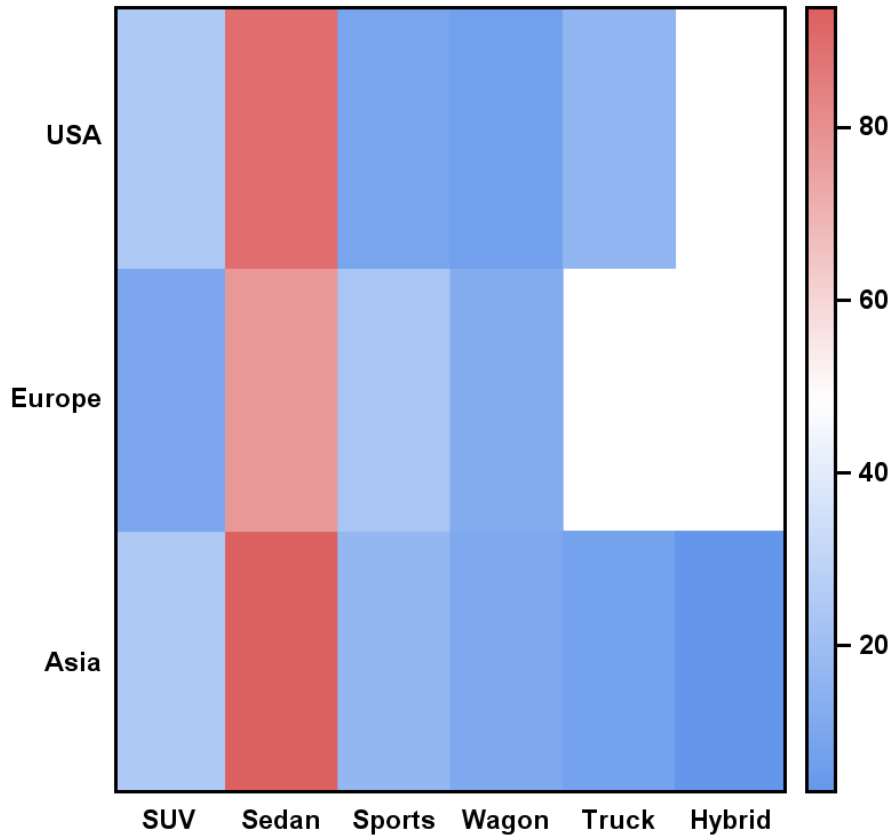
Counts of Car Origin and Type (Default Heat Map)
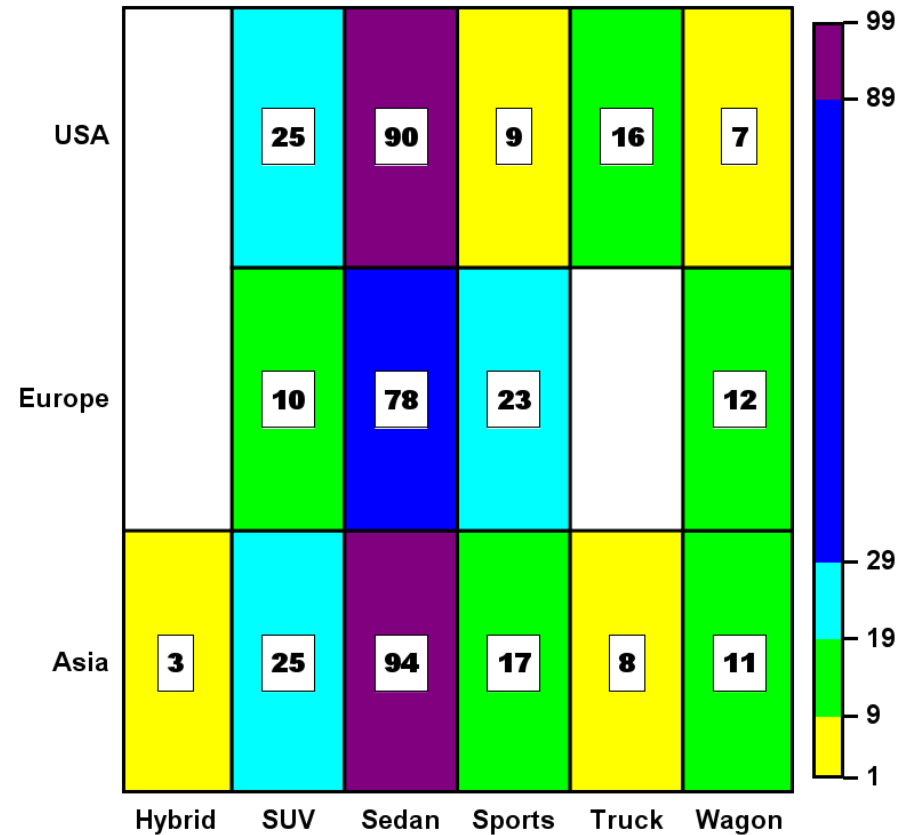
Default Heat Map
Useless Legend

Counts of Car Origin and Type (Custom Heat Map)

	Hybrid	SUV	Sedan	Sports	Truck	Wagon
USA		25	90	9	16	7
Europe		10	78	23		12
Asia	3	25	94	17	8	11

Custom Heat Map
Annotated Cells
Legend SubRanges

Continuous Color Gradient Legends do not provide knowledge of values

Continuous Color Gradient Legends do not provide knowledge of values

➢ **Matching between area color and legend color point is impossible.**

➢ **Value labels for each color point on the legend is impossible.**

➢ **Legend <u>Subrange</u> <u>Gradients</u> provide a visual aid for quick easy categorization and comparison.**

Heat Map of Blood Pressure for Weight-Height Combinations

Two Measurement Variables:
Average BP for Wgt-Hgt Cell
Frequency of Obs Within Cell

Average Diastolic Blood Pressure (BP)
By Weight (pounds) and Height (inches)
In 12 Weight Bins and 13 Height Bins
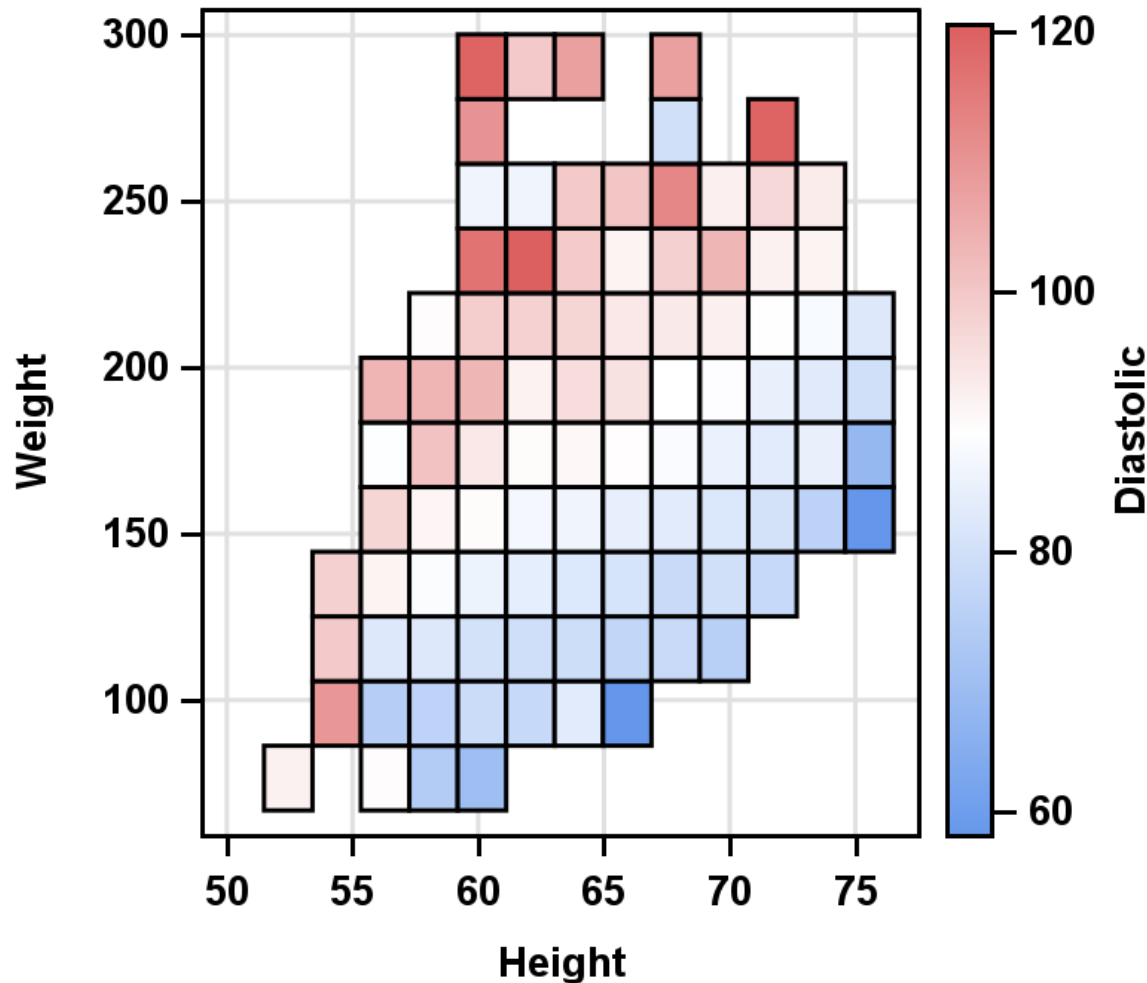
Default Heat Map: Useless for Precise Numbers
- Cannot match area to legend color
- Cannot label all colors on the legend

Custom Heat Map

- Annotate with Precise Numbers for Avg BP & Freq
- Subranges Gradient Legend for Quick Easy Category ID

Average Diastolic Blood Pressure
By Average Weight in pounds & Average Height in inches
For 5199 Observations in 12 weight Bins, 13 height Bins, & 94 Cells
Range of Avg Diastolic is 54-120 with Mean 89.4 & Standard Deviation 12.1
Labels Frequency above BP in Weight-Height Cells, Color Gradient for BP
All bins are equal width, but axis values are averages, not bin midpoints.
So the increment between axis values along an axis can vary.

Avg BP

	52	55	57	59	61	63	65	67	69	71	72	75	76
294					1 / 120	1 / 100	1 / 108	1 / 108					
273					3 / 111				1 / 80		1 / 120		
249		1 / 86		1 / 86	2 / 103	1 / 96	2 / 113	2 / 93	5 / 95	1 / 90			
231				3 / 118	7 / 113	6 / 100	8 / 95	12 / 100	10 / 98	12 / 92	2 / 95		
210			2 / 95	11 / 97	13 / 100	23 / 98	33 / 93	50 / 93	38 / 90	23 / 90	6 / 82	2 / 90	
192		2 / 104	6 / 91	18 / 106	54 / 93	54 / 96	85 / 93	117 / 89	93 / 87	35 / 83	6 / 87		
174		4 / 98	15 / 98	58 / 92	117 / 91	169 / 92	233 / 87	213 / 88	108 / 84	38 / 86	5 / 73	1 / 70	
154		8 / 95	38 / 92	131 / 90	239 / 86	302 / 86	281 / 84	186 / 82	75 / 81	22 / 77	1 / 58		
135	3 / 99	18 / 90	89 / 87	225 / 86	364 / 83	336 / 83	165 / 79	91 / 79	22 / 78	4 / 78			
117	3 / 79	12 / 85	93 / 83	186 / 80	249 / 80	142 / 78	47 / 76	13 / 77	5 / 76				
100	1 / 110	11 / 74	32 / 77	46 / 79	28 / 77	8 / 82	1 / 54						
77	1 / 92		1 / 90	2 / 74	2 / 70								

120
89
79
54
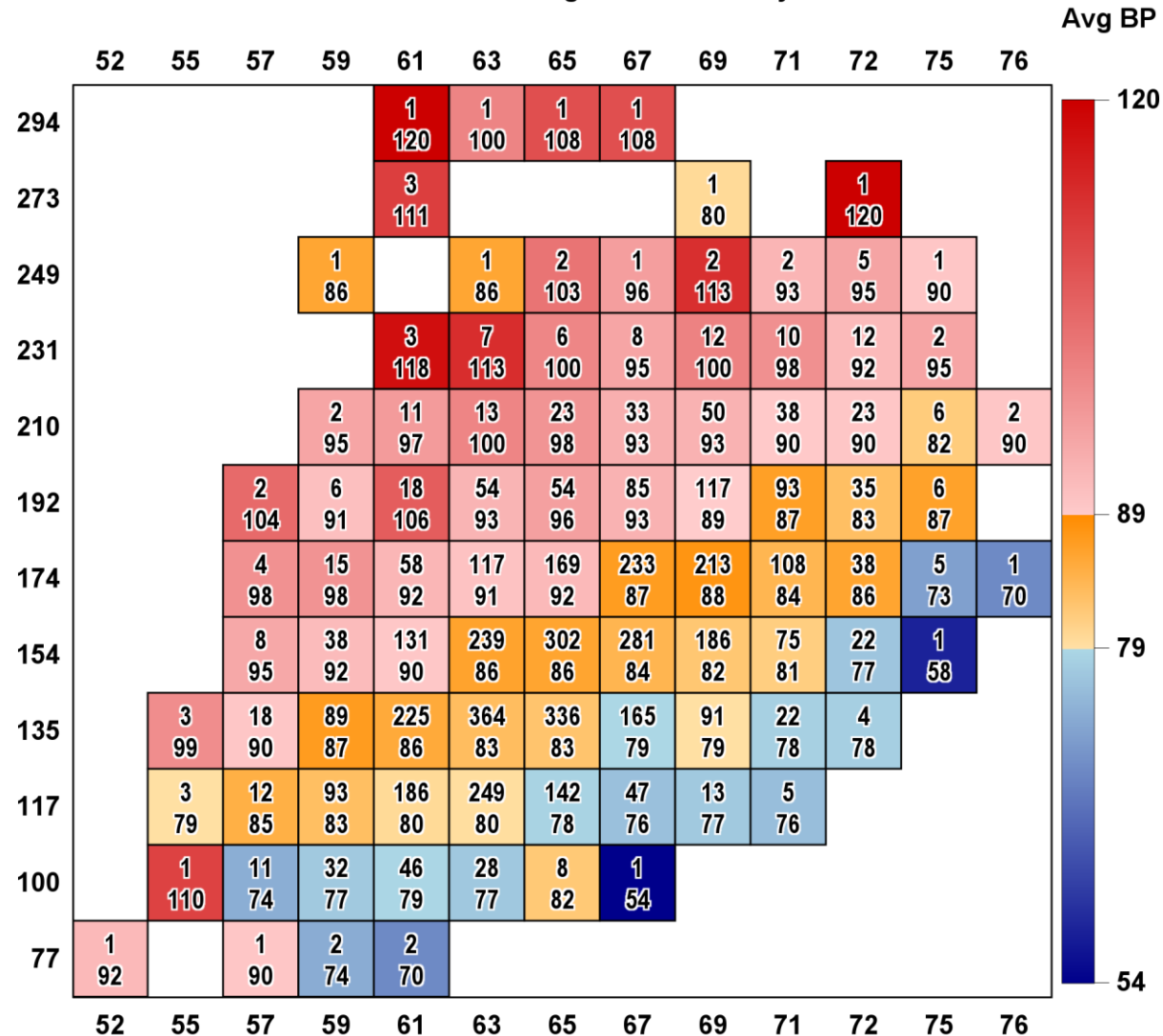
Blue [to 79] is Normal Diastolic BP & Orange [to 89] is Hypertension Stage 1
Red [to 120] is Hypertension Stage 2 & Purple [> 120] is Hypertensive Crisis

Average Diastolic Blood Pressure
By Average Weight in pounds & Average Height in inches
For 5199 Observations in 4 weight Bins, 13 height Bins, & 42 Cells
Range of Avg Diastolic is 70-113 with Mean 88.8 & Standard Deviation 9.6
Labels Frequency above BP in Weight-Height Cells, Color Gradient for BP
All bins are equal width, but axis values are averages, not bin midpoints.
So the increment between axis values along an axis can vary.

Avg BP

Each cell shows Frequency (top) and Avg BP (bottom):

Weight \ Height	52	55	57	59	61	63	65	67	69	71	72	75	76
261				1 / 86	4 / 113	2 / 93	3 / 104	2 / 102	3 / 102	2 / 93	6 / 99	1 / 90	
200			2 / 104	8 / 92	32 / 104	74 / 96	83 / 97	126 / 93	179 / 91	141 / 88	70 / 87	14 / 86	2 / 90
152		3 / 99	30 / 92	142 / 89	414 / 88	720 / 86	807 / 86	679 / 84	490 / 84	205 / 83	64 / 82	6 / 70	1 / 70
114	1 / 92	4 / 87	24 / 80	127 / 82	234 / 80	277 / 79	150 / 79	48 / 76	13 / 77	5 / 76			

Color scale (Avg BP): 120 — 89 — 79 — 70

Blue [to 79] is Normal Diastolic BP & Orange [to 89] is Hypertension Stage 1
Red [to 120] is Hypertension Stage 2 & Purple [> 120] is Hypertensive Crisis

Let's Look More Closely at the Heat Map ►►►

Average Diastolic Blood Pressure
By Average Weight in pounds & Average Height in inches
For 5199 Observations in 4 weight Bins, 13 height Bins, & 42 Cells
Range of Avg Diastolic is 70-113 with Mean 88.8 & Standard Deviation 9.6
Labels Frequency above BP in Weight-Height Cells, Color Gradient for BP
All bins are equal width, but axis values are averages, not bin midpoints.
So the increment between axis values along an axis can vary.

▲Dynamic Titles Static Footnotes▼

Blue [to 79] is Normal Diastolic BP & Orange [to 89] is Hypertension Stage 1
Red [to 120] is Hypertension Stage 2 & Purple [> 120] is Hypertensive Crisis

ODS Graphics Bubble Plots

➢ **Bubble Sizing Default is peculiar, misleading, and counterintuitive**

➢ **Other defaults are peculiar also**

Software Bubble Sizing

"LINEAR" Default
2 is over-represented versus 1

2

1
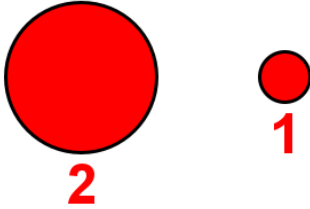
PROPORTIONAL Option
area for 2 is twice area for 1

2 1
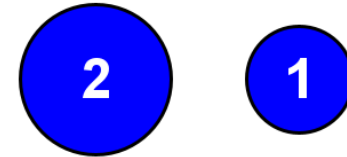
Examples Using Proportional Sizing

Categorical Variables:

Automobile Place of Origin

Vehicle Type

Measurement Variables:

Average Vehicle Price (Bubble Size)

Miles Per Gallon (Initially Only Legend)

Legend color swatches too small
Useless color match of bubble data labels
NonBlack bubble outlines—no value add

Avg Price and Avg Miles Per Gallon (City) by Car Origin and Type

DEFAULTS

	Hybrid	SUV	Sedan	Sports	Truck	Wagon
USA		$34,589	$25,639	$45,257	$27,220	$22,346
Europe		$48,346	$42,992	$71,999		$37,851
Asia	$19,920	$29,569	$22,764	$32,511	$20,384	$23,144

MPG (City)

55 17 15 16 23 20 21 18 22 19
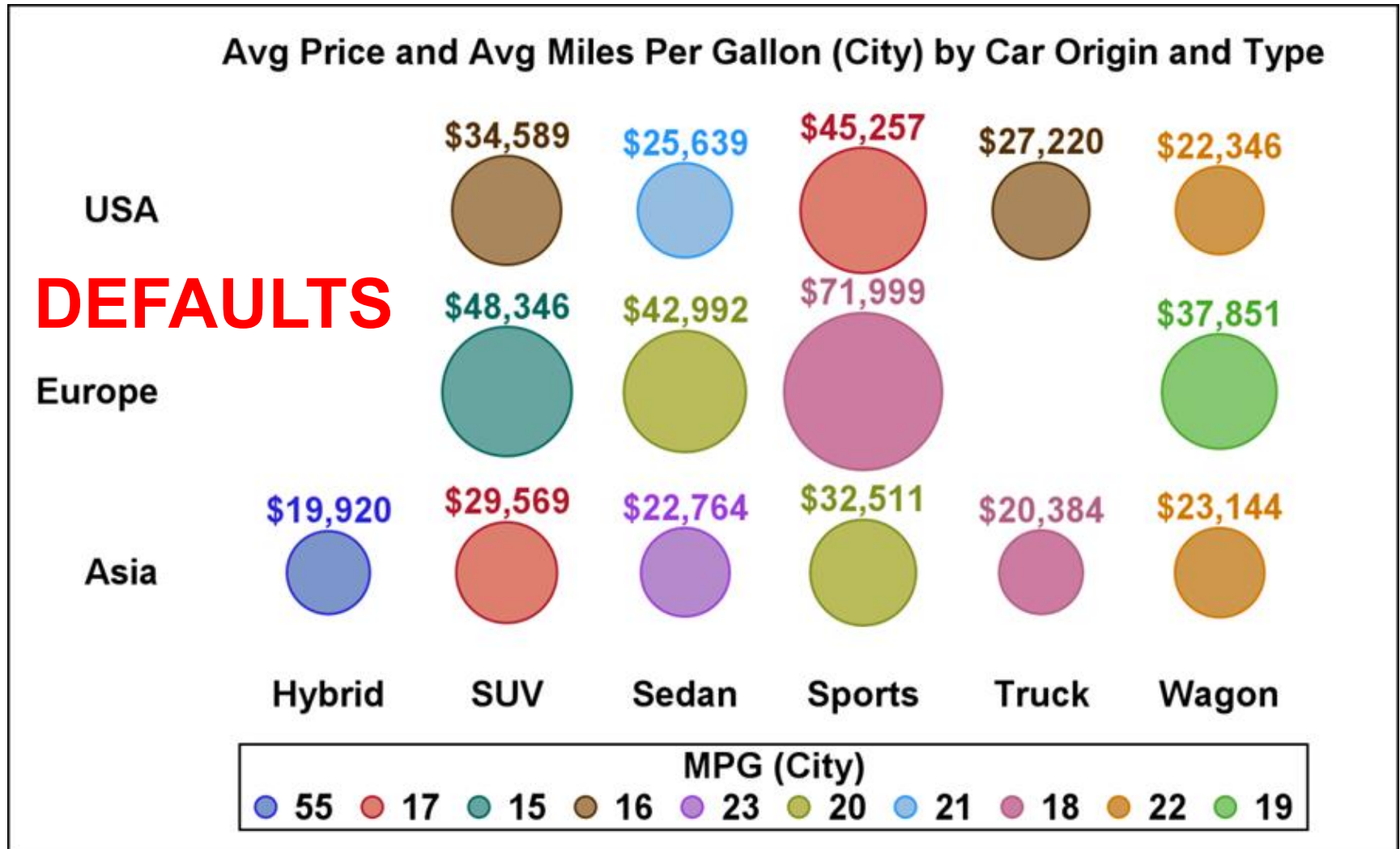
Using bubble color for bubble labels can make labels hard to read if a faint enough color palette is used. Here, no crisis.

Avg Price and Avg Miles Per Gallon (City) by Car Origin and Type

DEFAULTS

	Hybrid	SUV	Sedan	Sports	Truck	Wagon
USA	$34,589	$25,639	$45,257	$27,220	$22,346
Europe	$48,346	$42,992	$71,999		$37,851
Asia	$19,920	$29,569	$22,764	$32,511	$20,384	$23,144

MPG (City)
55 17 15 16 23 20 21 18 22 19
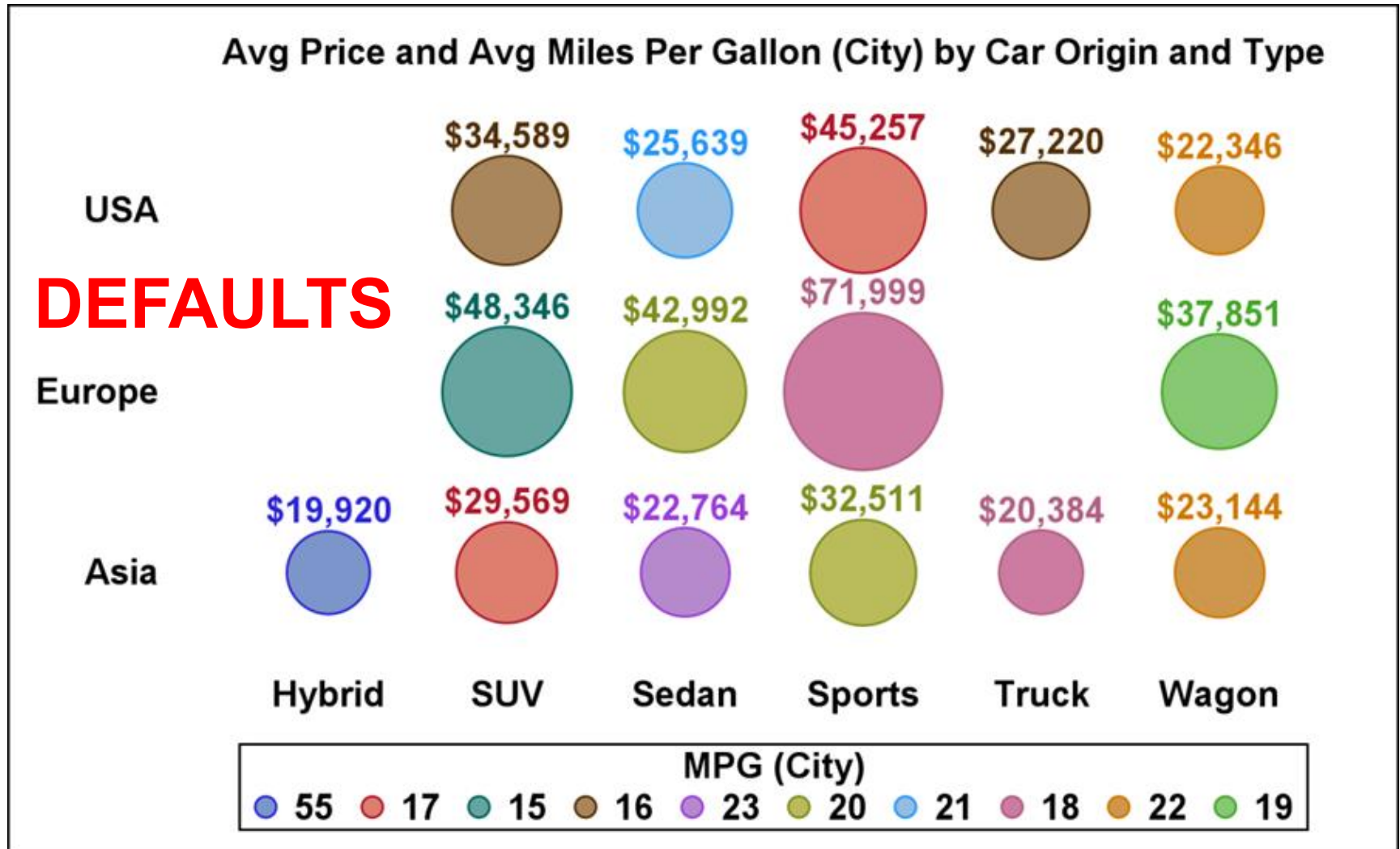
Distinguishable color swatches
Maximally readable black data labels
Brighter colors, Black outline for bubbles

Avg Price and Avg Miles Per Gallon (City) by Car Origin and Type
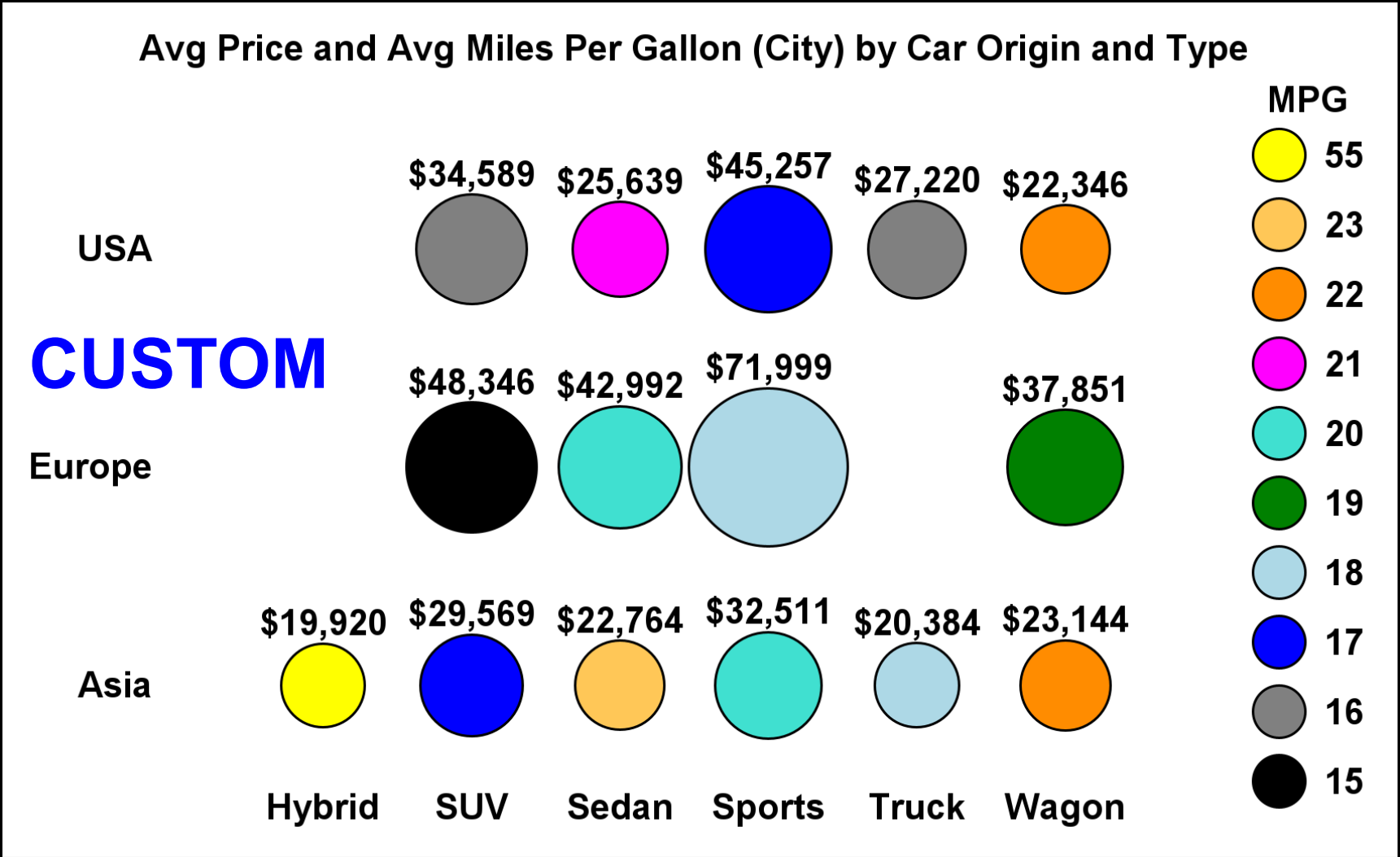
Annotate both Dollars in K and MPG
Retain Legend As a Quick Visual Ranker

Avg Price and Avg Miles Per Gallon (City) by Car Origin and Type
Bubble Size Proportional To Price. MPG & Its Color Are Inside the Bubble.
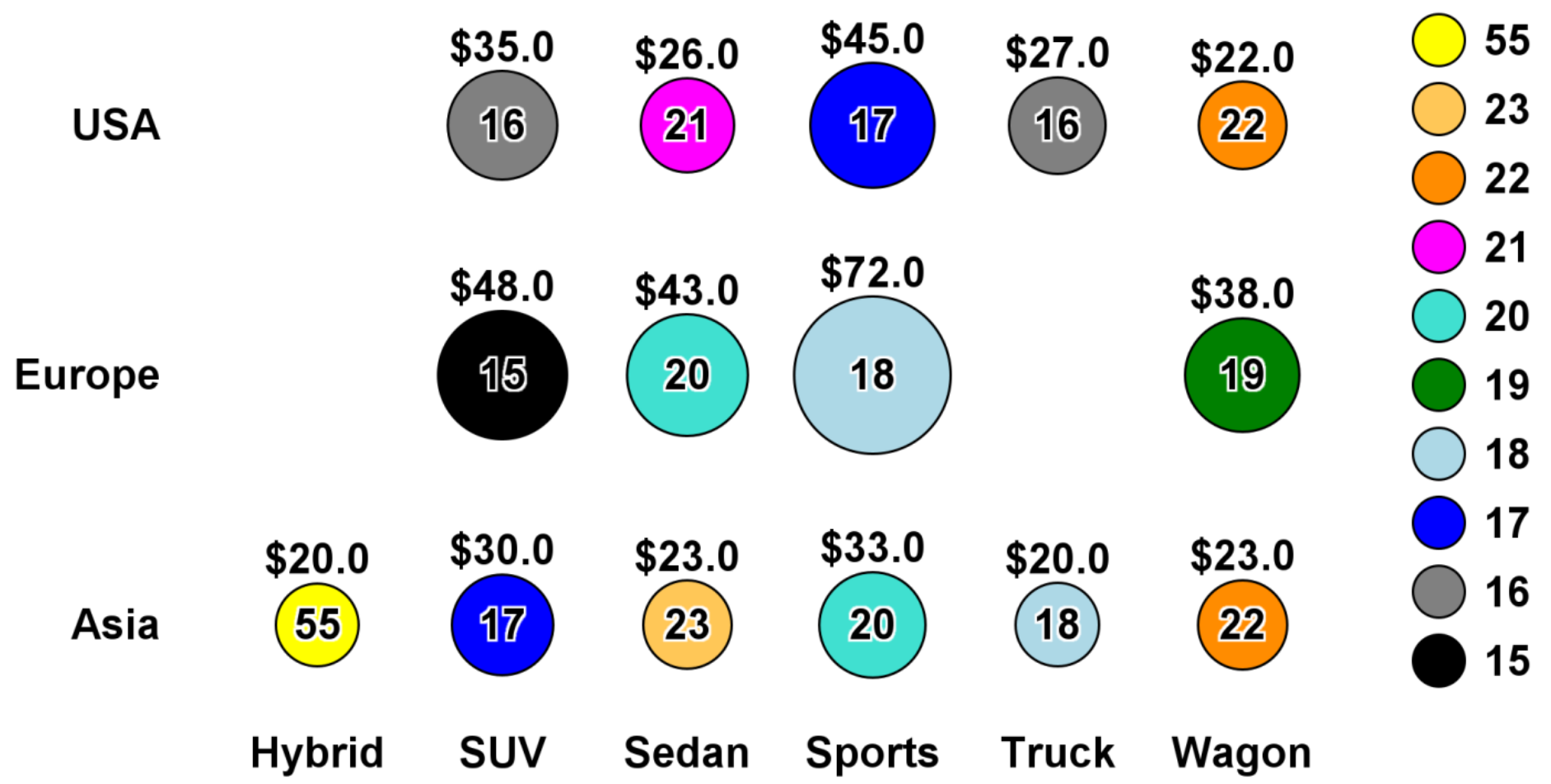
BackFill can be used instead of BackLight
Due to Sufficient Space Inside Bubbles

Avg Price and Avg Miles Per Gallon (City) by Car Origin and Type
Bubble Size Proportional To Price. MPG & Its Color Are Inside the Bubble.

USA: Hybrid —, SUV $35.0 (16), Sedan $26.0 (21), Sports $45.0 (17), Truck $27.0 (16), Wagon $22.0 (22)

Europe: SUV $48.0 (15), Sedan $43.0 (20), Sports $72.0 (18), Wagon $38.0 (19)

Asia: Hybrid $20.0 (55), SUV $30.0 (17), Sedan $23.0 (23), Sports $33.0 (20), Truck $20.0 (18), Wagon $23.0 (22)

Legend: 55, 23, 22, 21, 20, 19, 18, 17, 16, 15

Hybrid   SUV   Sedan   Sports   Truck   Wagon
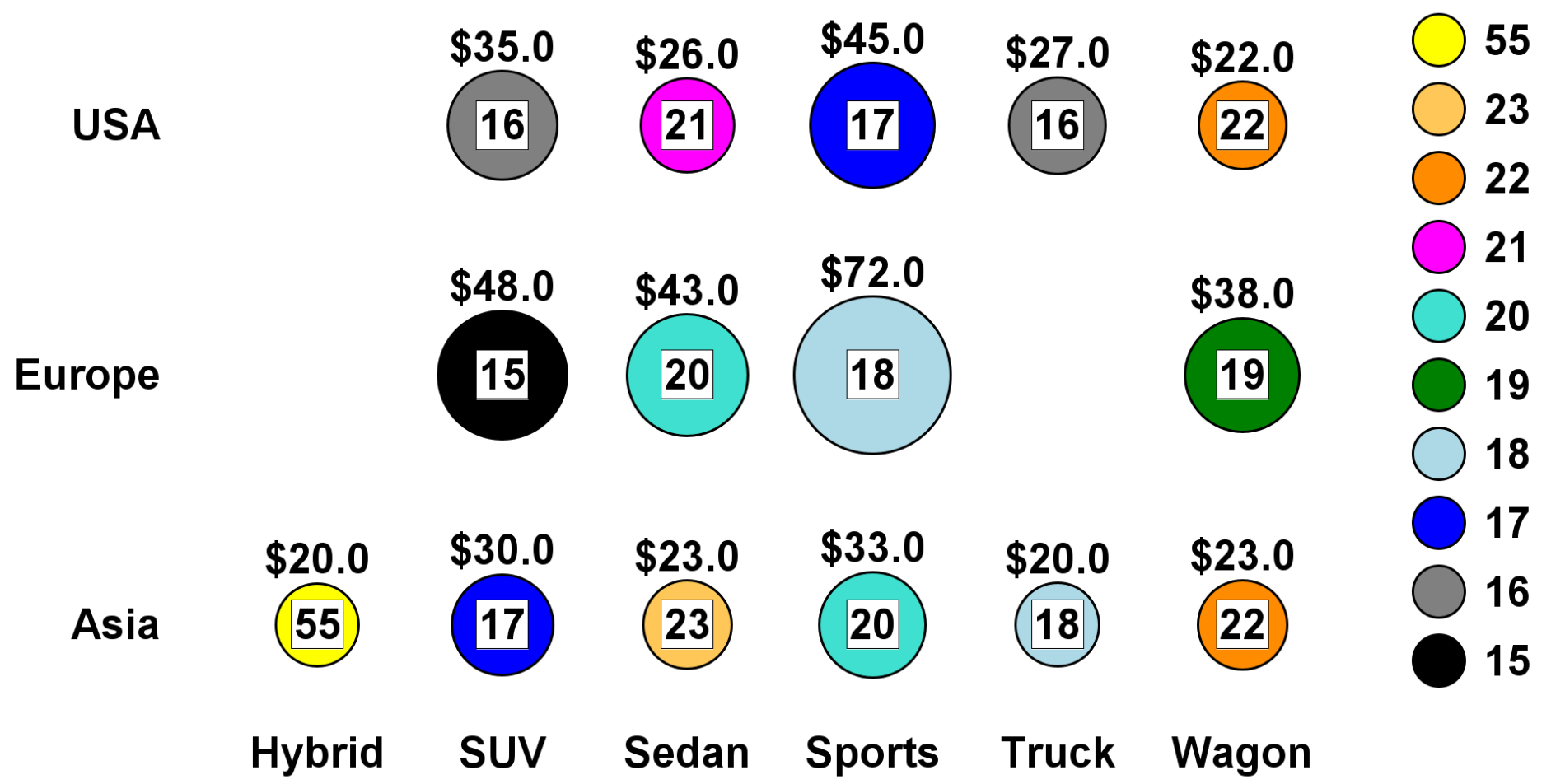
BackFill usually better than BackLight

If using BackFill instead of BackLight, be sure enough color fill is available to be able to distinguish it reliably. With both var values annotated, the legend is superfluous, but does no harm, and IS a visual comparator.

Squeezing More into a Bubble Plot

Shoe Sales (in $M) by Region and Product with Region Totals and Product Totals
Bubble Color for Count of Cities Selling Product in Region
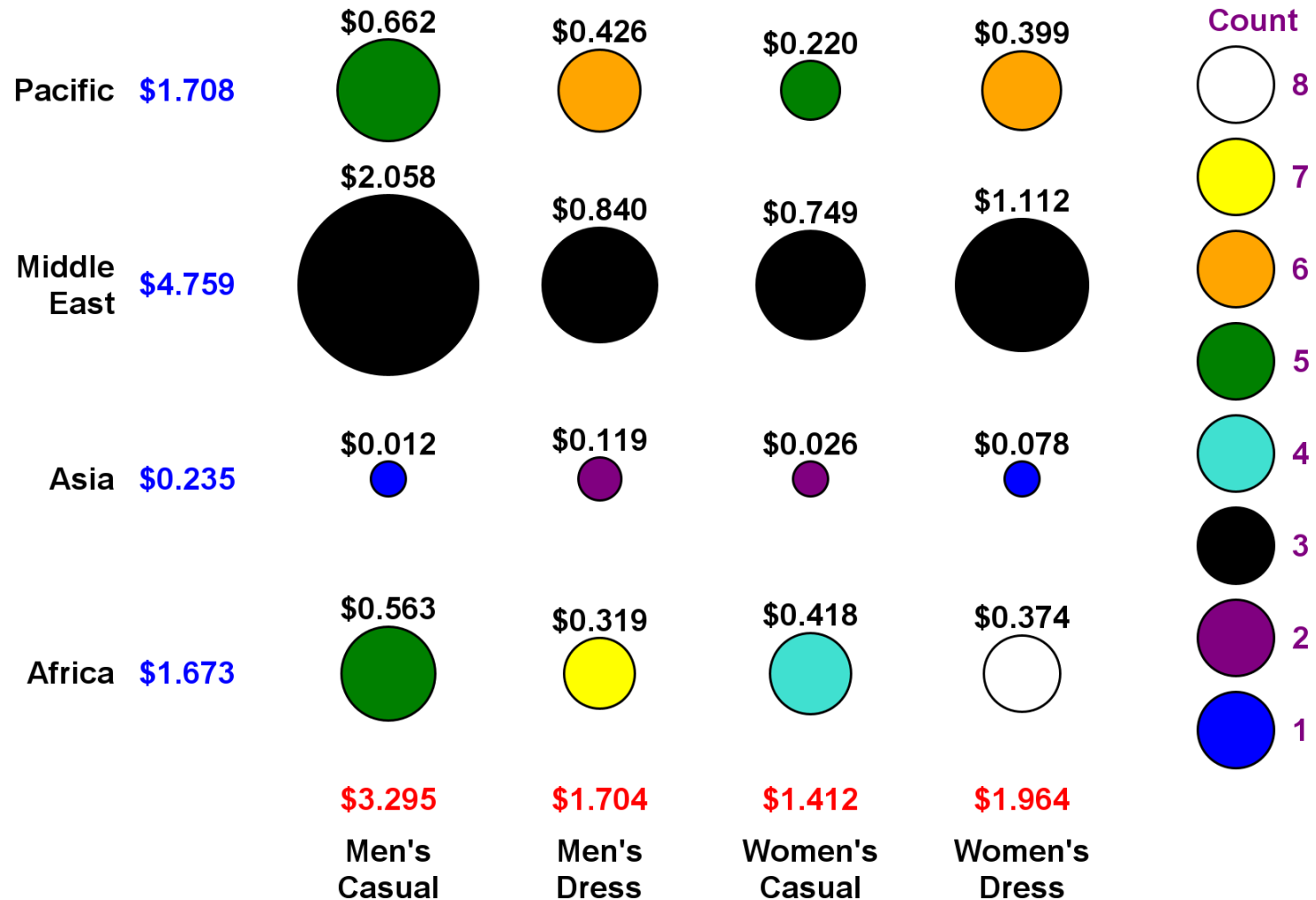
Histograms are
Usually Vague Summary Images

Histogram Bins with a genuine information/insight purpose
Title Line 2 is dynamic

Distribution of Cholesterol in sashelp.heart Data Set
Count:5057, Min:96, Max:568, Mean:227, StdDev:44.94
Values: Less Than 200 - Desirable, 200 to 240 - Borderline, Over 240 - High

1405   1901   1751

96   199   240   568

Cholesterol By Range
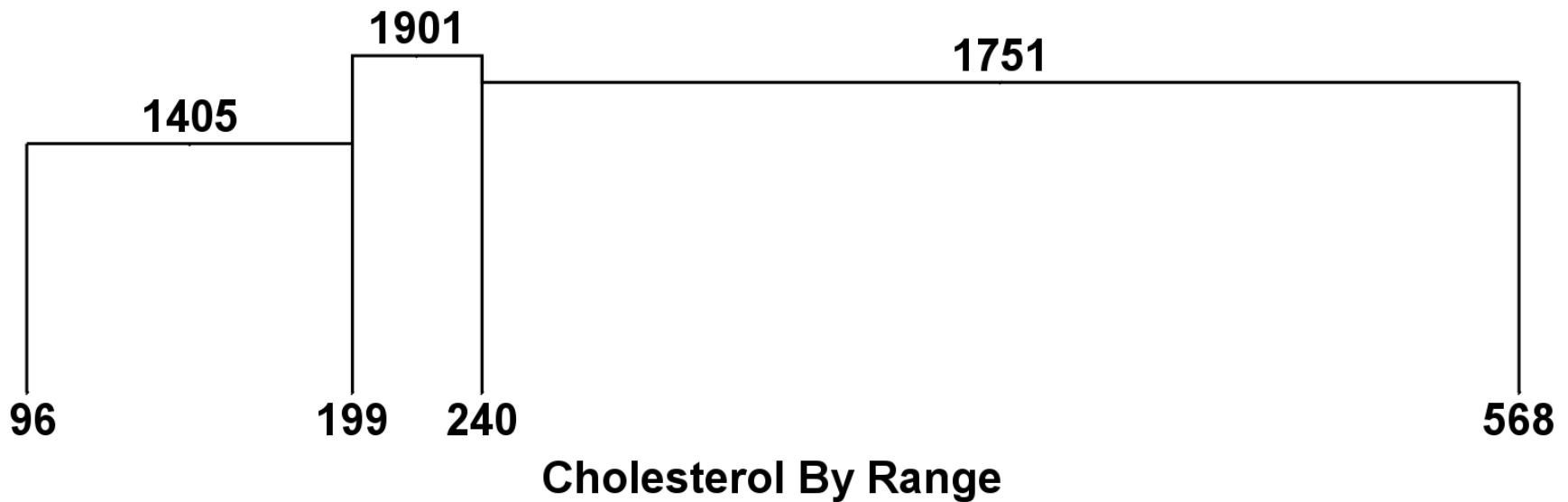
7 Quantiles Distribution of Cholesterol in sashelp.heart
5057 Values in 6 Bins of 723 values each and Last Bin with 719 values
Min:96,Q1:181,Q2:200,Q3:217,Q4:231,Q5:248,Q6:275,Q7/Max:568
Red: Min, First Quantile, Second from Last Quantile, Max/Last Quantile
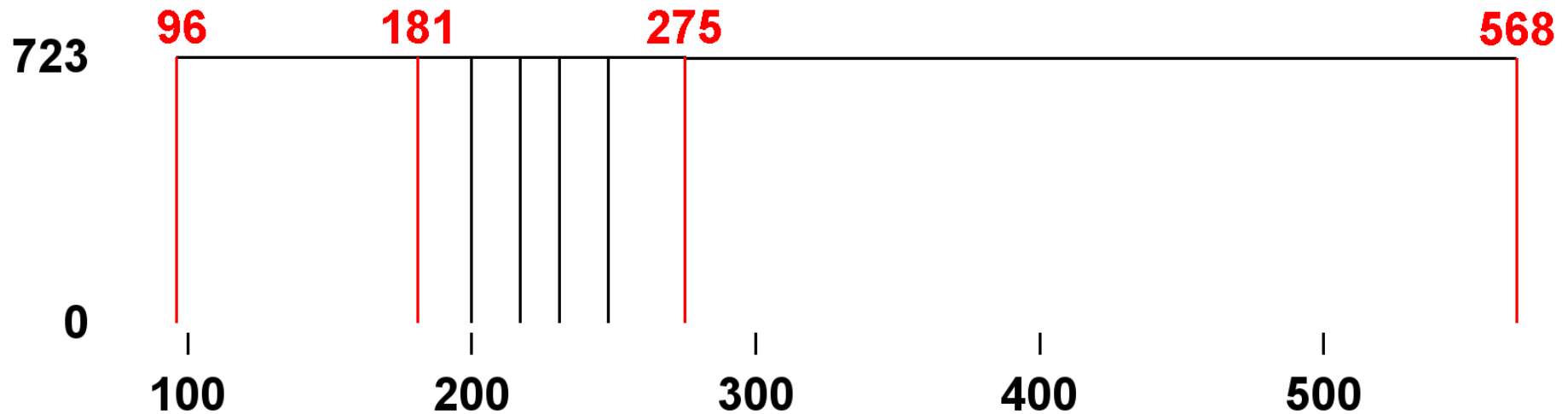
Quantiles As Bins
Last Bin is leftovers unless
observations count / quantiles count
has no remainder

Statistical Bins and Statistics Dynamically Delivered by Code ►►

Height By Standard Deviation Range in sashelp.heart Data Set
where Sex EQ 'Female'
Count:2869,1 STD:1993(69.47%), 2 STD:2730(95.16%), 3 STD:2860(99.69%)
Mean:62.57, Median:62.5, Mode:62.50 Max Freq: 155@62.50
Min:51.5, Max:70.75, Mean:62.57, STD:2.45

Title Lines 3,4,5 are dynamic

1993

395

342

74

6

56

3

52   55   58   60   65   67   70 71

Height in Inches Rounded
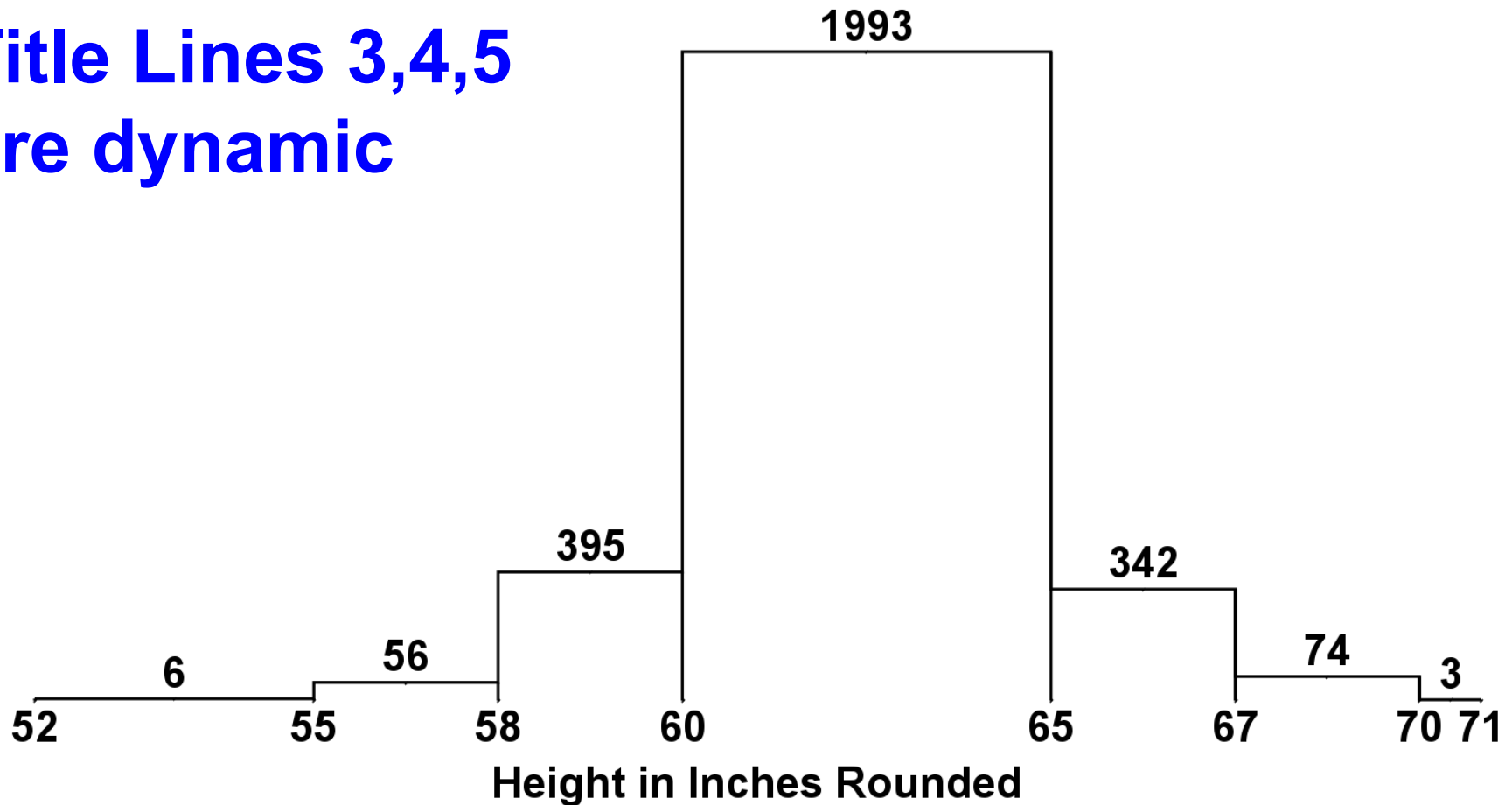
If Normal Distribution: Mean = Median = Mode, and
68.27% of counts within 1 STD, 95.45% within 2 STD, 99.73% within 3 STD

Actual and Normal Dist Overlay
Also, Normality Test Explained ►
- Needle Plot for Freq Distribution
- PDF function for Normal Curve

Frequency Distribution of Height in SASHELP.HEART
Overlaid with output from the SAS PDF function for a Normal Distribution
2869 Height Values with 66 Frequencies
▲ Max. Freq: 155@62.50 | StdDev: 2.45 | Values for Dots & Squares:
● Minimum:51.50, Median:62.50, Maximum:70.75, Mean:62.57, Mode:62.50
■ Percentile 1:56.75, 10:59.50, 25:61.00, 75:64.25, 90:65.75, 99:68.25
Data Selection Filter: Sex EQ 'Female'

Frequency

Probability Density

Height in Inches

If Normal Distribution: Mean = Median = Mode (i.e., value with max freq)

Frequency Distribution of Height in SASHELP.HEART
Overlaid with output from the SAS PDF function for a Normal Distribution
2869 Height Values with 66 Frequencies
▲ Max. Freq: 155@62.50 | StdDev: 2.45 | Values for Dots & Squares:
● Minimum:51.50, Median:62.50, Maximum:70.75, Mean:62.57, Mode:62.50
■ Percentile 1:56.75, 10:59.50, 25:61.00, 75:64.25, 90:65.75, 99:68.25
Data Selection Filter: Sex EQ 'Female'

Frequency

Probability Density

160
150

0.15

Title Lines 3-7 are dynamic, Footnote is static

10
0

0.00

51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71

Height in Inches

If Normal Distribution: Mean = Median = Mode (i.e., value with max freq)

When the Data is nowhere nearly a Normal Distribution,
Use This Highly Informative Actual Distribution ► ► ►
It's Laden with Statistics and Helpful Markers for Them.
I ended up here in a quest to get more than any histogram would deliver for me.

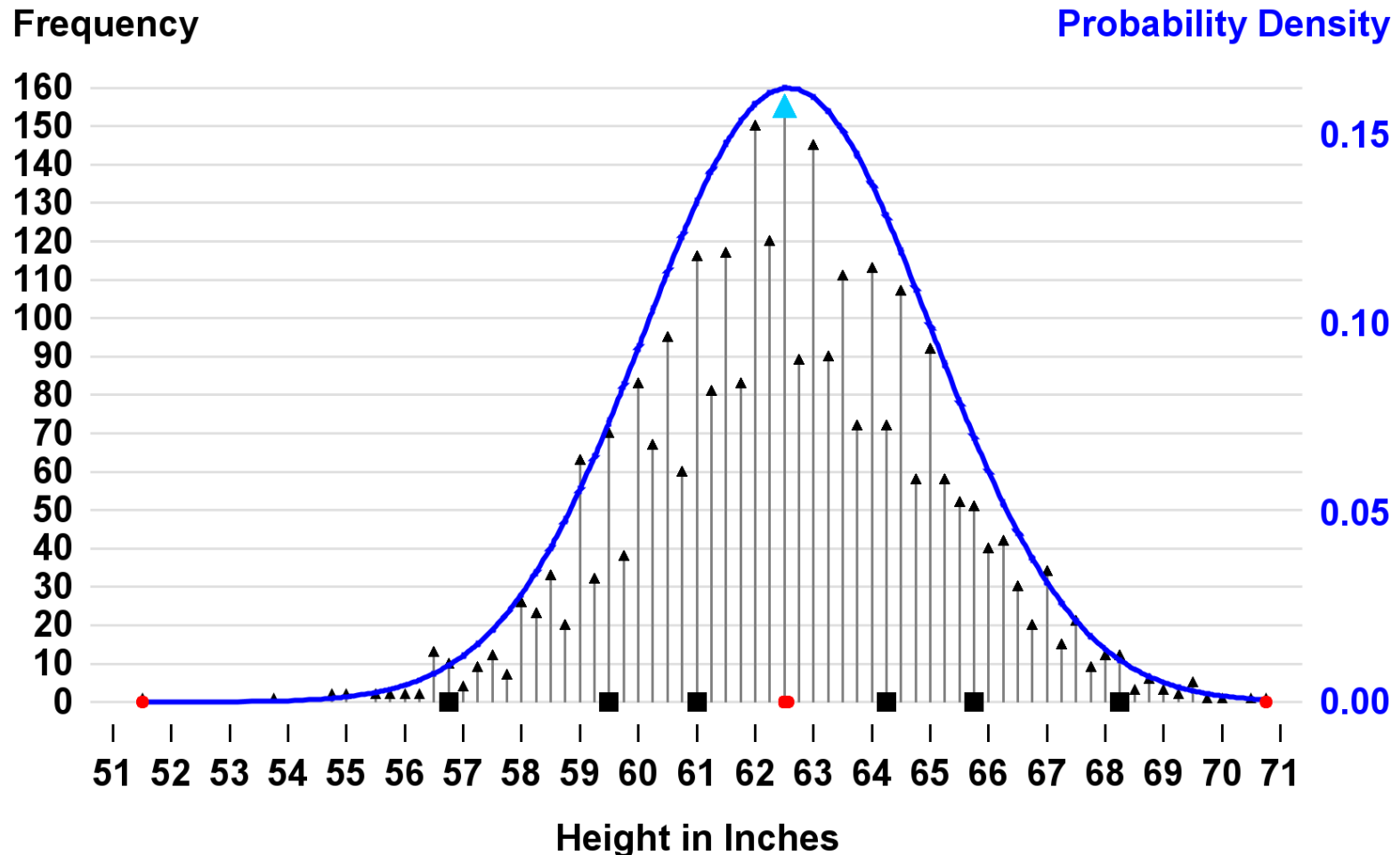Frequency Distribution of Height in SASHELP.HEART
2869 Height Values with 66 Frequencies
▲ Max. Freq: 155@62.50 | StdDev: 2.45 | Values for Dots & Squares:
● Minimum:51.50, Median:62.50, Maximum:70.75, Mean:62.57, Mode:62.50
■ Percentile 1:56.75, 10:59.50, 25:61.00, 75:64.25, 90:65.75, 99:68.25
Data Selection Filter: Sex EQ 'Female'

Height in Inches
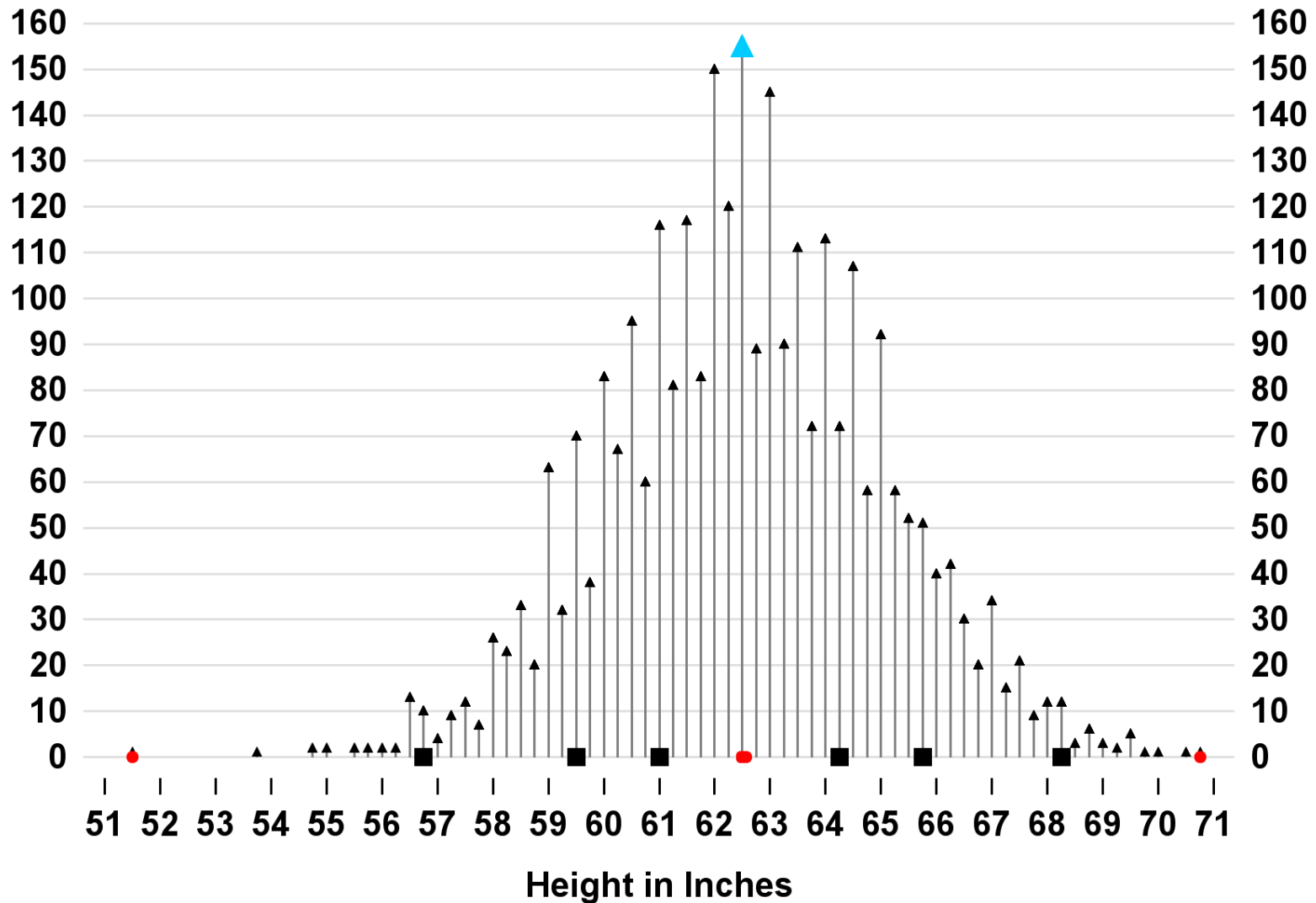
Fringe Plot Adds Value to the Otherwise Habitually Underinformative

Count and Percent of Height Ranges for Population of 19 Students

Height Instance (two are overlaid pairs)
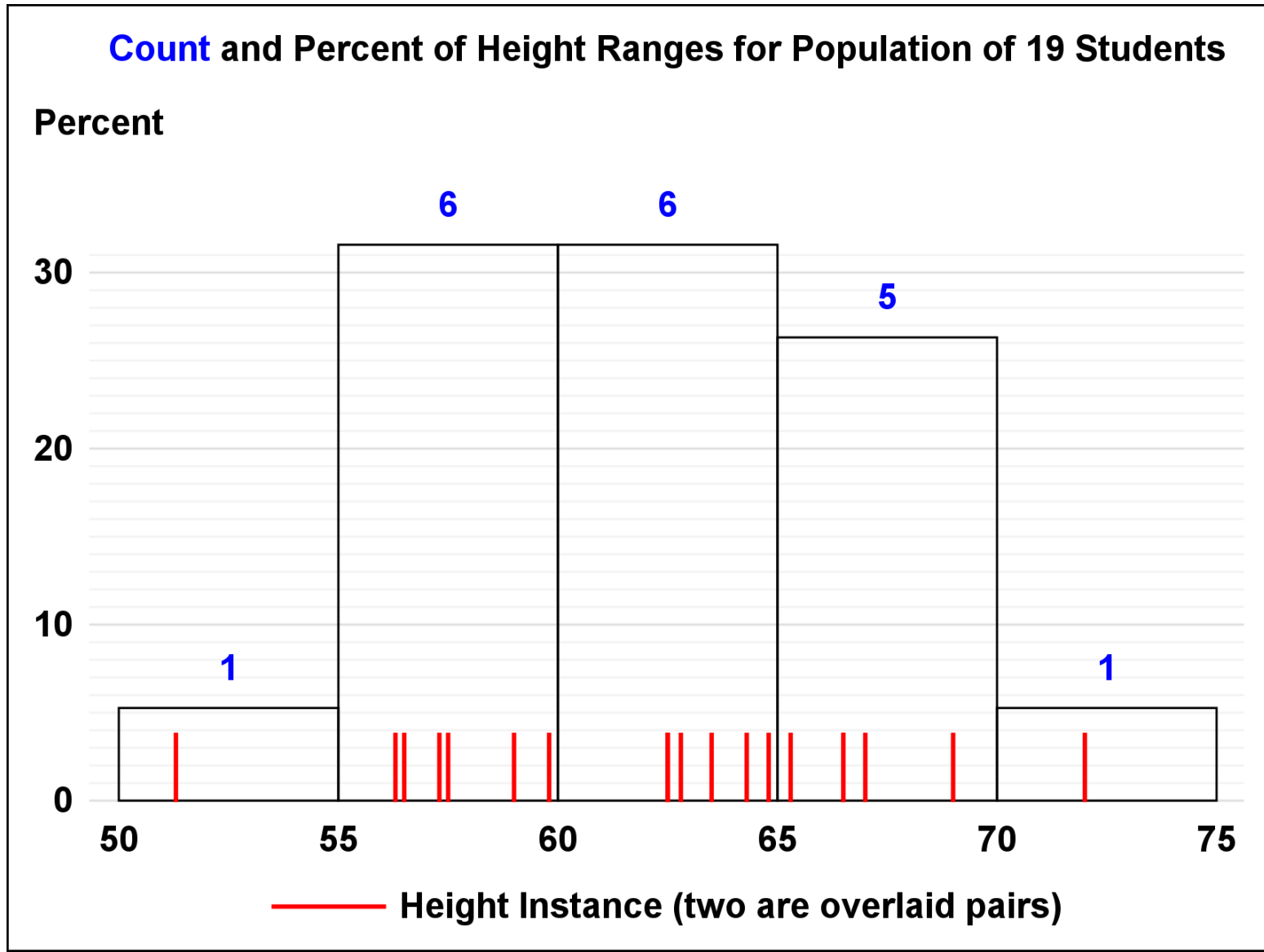
**Typical Box Plots Are Only
for Viewers Already in the Know**

Bar charts, pie charts, and time series plots need NO explanation.

For Box Plots, a SAS programmer can find a diagram in the
SAS ODS Graphics Procedures Guide to decode it.
(Box Height has NO meaning!)

What is this???

Distribution of Female Height (inches) in SASHELP.HEART Data Set
(outliers not shown)

50   55   60   65   70

Distribution of Female Height (inches) in SASHELP.HEART Data Set

52 54 55 55 56 56   69 69 70 70 7171

55 55 56 56 56 56   70 70 70 70 70

50   55   60   65   70
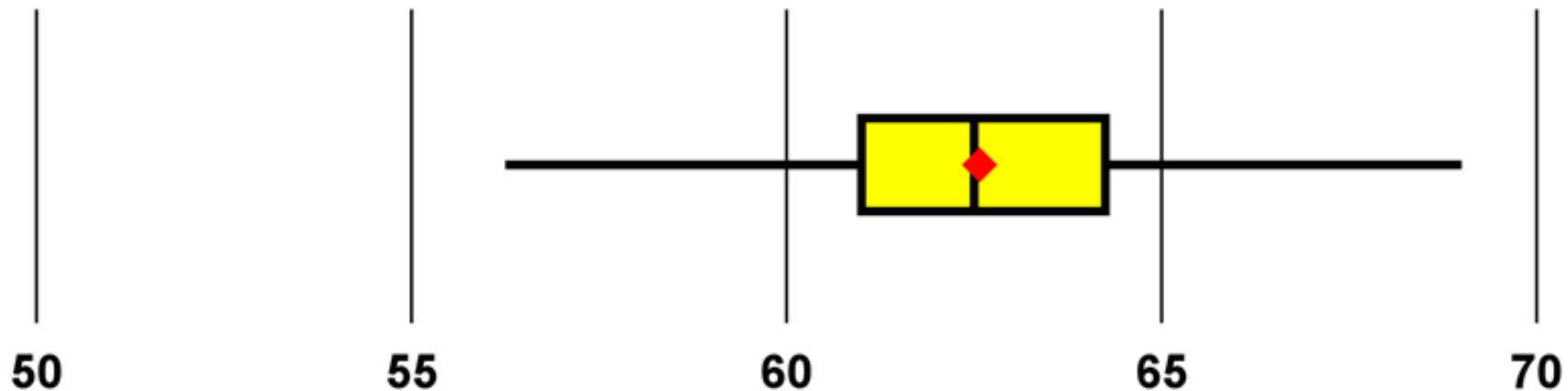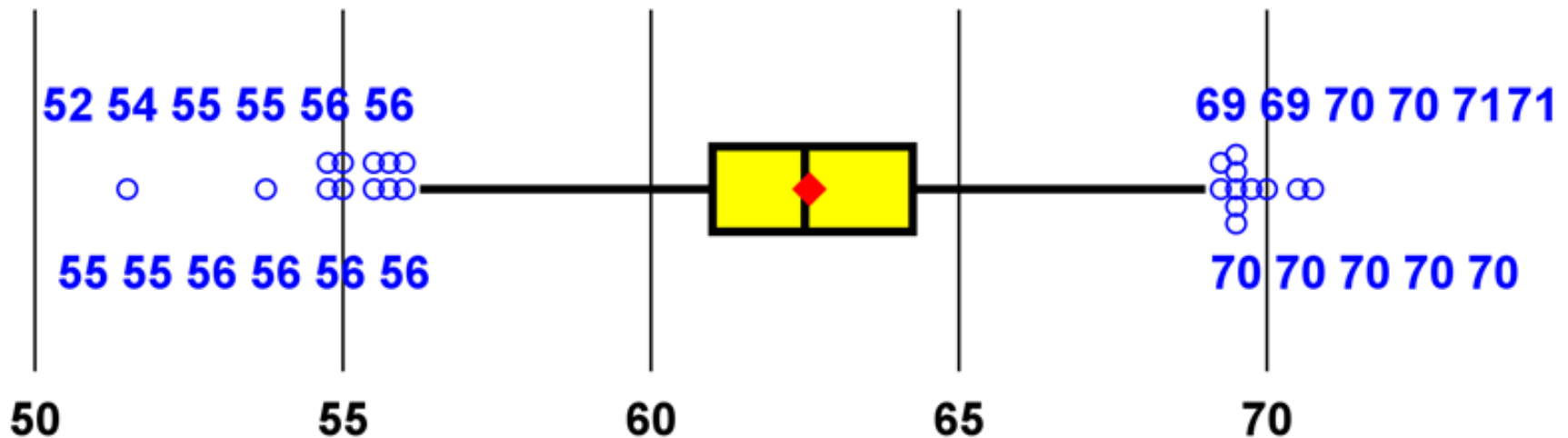
Very Useful for Visual Comparison (IF you understand it)
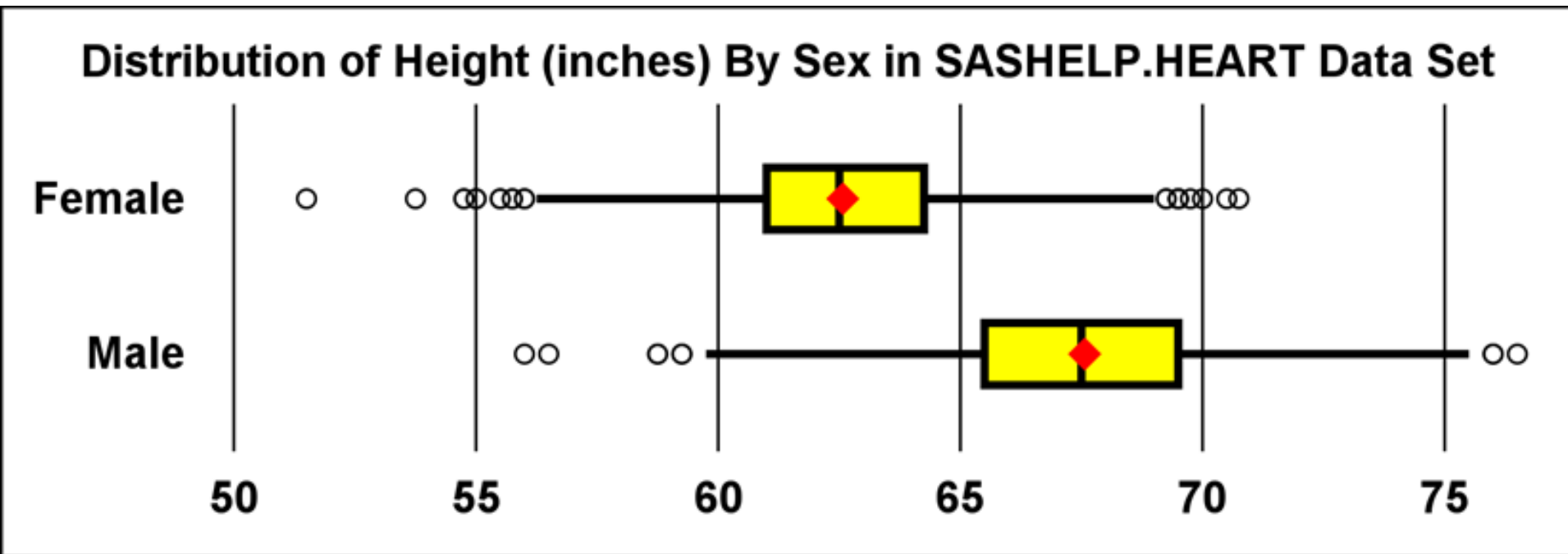
Distribution of Height (inches) By Sex in SASHELP.HEART Data Set

Distribution of Height (inches) in sashelp.heart Data Set
Data Selection Filter: Sex EQ 'Female'

51.50   70.75
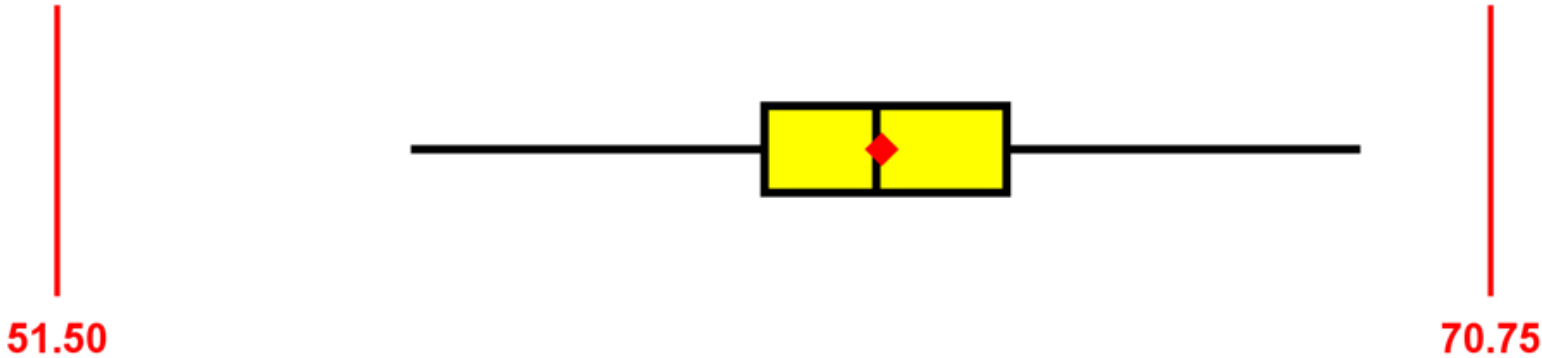
2869 Values, Standard Deviation 2.45
Maximum Frequency: 155@62.50
Minimum 51.50 and Maximum 70.75
Left End of Lower Whisker 56.25
First Quartile (left edge of box) 61.00
Median (vertical bar in box) 62.50
Mean (diamond in box) 62.57
Third Quartile (right edge of box) 64.25
Right End of Upper Whisker 69.00
Outliers (not shown): 51.50, 53.75, 54.75, 54.75, 55.00, 55.00, 55.50, 55.50,
55.75, 55.75, 56.00, 56.00, 69.25, 69.25, 69.50, 69.50, 69.50, 69.50, 69.50, 69.75,
70.00, 70.50, 70.75

Self-Explanation (The Full Story)
Is Dynamically Delivered by Code

Distribution of Height (inches) in sashelp.heart Data Set
Data Selection Filter: Sex EQ 'Female'

52 54 55 55 56 56   69 69 70 70 71 71

55 55 56 56 56 56   70 70 70 70 70

2869 Values, Standard Deviation 2.45
Maximum Frequency: 155@62.50
Minimum 51.50 and Maximum 70.75
Left End of Lower Whisker 56.25
First Quartile (left edge of box) 61.00
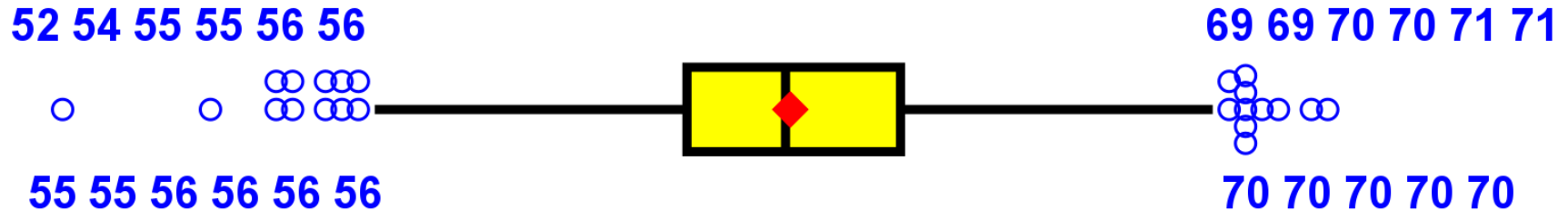**Median (vertical bar in box) 62.50**
Mean (diamond in box) 62.57
Third Quartile (right edge of box) 64.25
Right End of Upper Whisker 69.00

Self-Explanation
AND The <u>Complete</u> Picture

Outlaw Data Art

Alternative Vertical Box Plot [omits the box](#) (Outliers could be added, if required.)

Distribution of Height (inches) in sashelp.heart Data Set
Data Selection Filter: Sex EQ 'Female'
2869 Height Values, Standard Deviation 2.45
Maximum Frequency: 155@62.50

■ Maximum=70.75

● Pctl_99=68.25

● Pctl_95=66.75
● Pctl_90=65.75

● Quartile_3=64.25

Median=62.50 ◄ Mean=62.57

● Quartile_1=61.00

● Pctl_10=59.50
● Pctl_05=58.50

● Pctl_01=56.75

■ Minimum=51.50
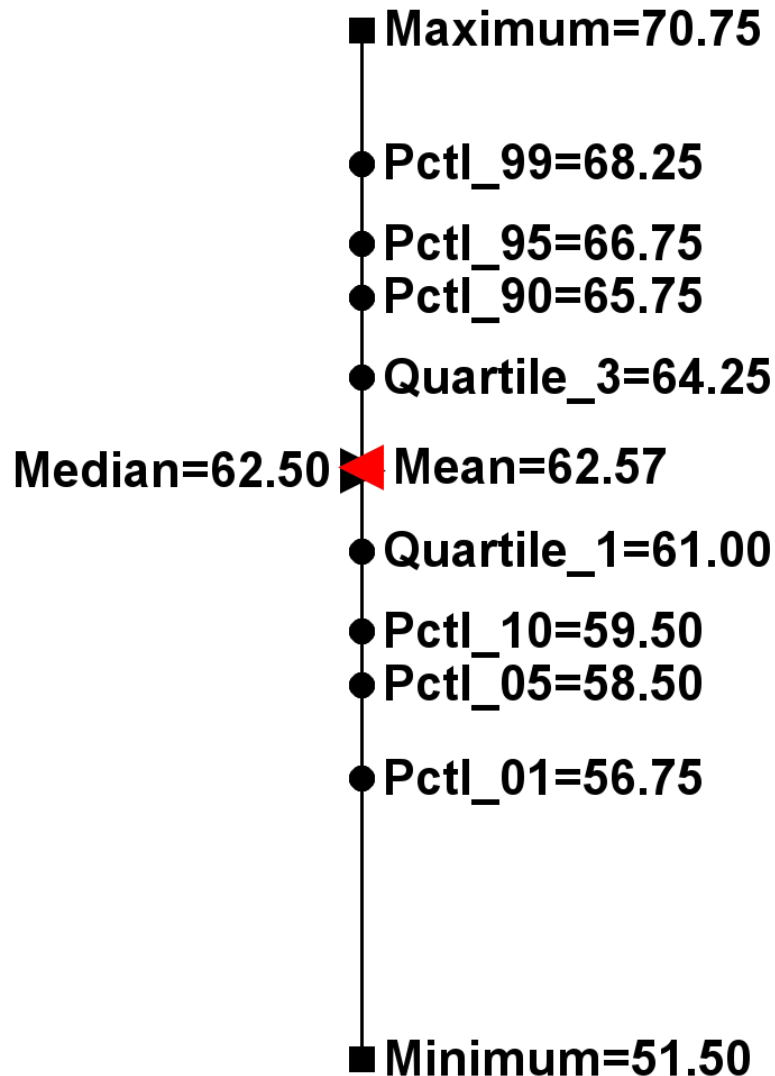
Needle Plots

Student Ranked Weight (pounds) with Height (inches), Age, and Sex

	Janet	John	Judy	Jane	Jeffrey	James	Joyce
Weight	112.5	99.5	90	84.5	84	83	50.5
Height	62.5	59	64.3	59.8	62.5	57.3	51.3
Age	15	12	14	12	13	12	11
Sex	F	M	F		M		F
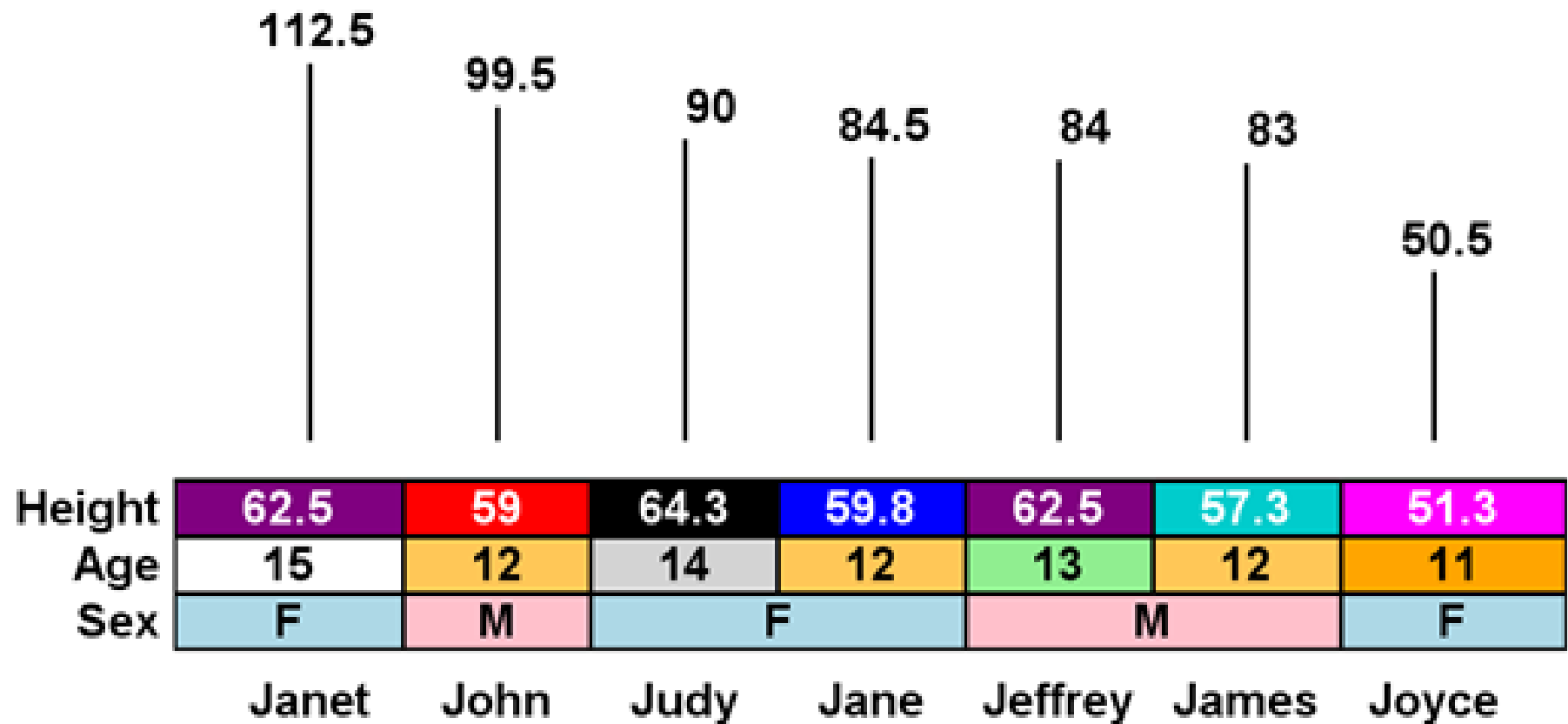
FIVE variables in TWO dimensions
Needle Plot + Group-ColorCoded Block Chart
More value than X axis tables

Needle Plots

That Compare Categories With Each Other & With The Total

Tree Chart Like VBAR Chart But Total Bar to Compare

**Ranked Shoe Sales and Percent Share By Region and Total
for Boots, Sandals, Slippers, and Sport Shoes (in Four Regions)**

Total - $2,356,373 - 100%

Middle East - $872,955 - 37.0%

Africa - $669,470 - 28.4%
Pacific - $588,908 - 25.0%

Asia - $225,040 - 9.6%

Flag Chart Like VBAR Chart But Total Bar to Compare

Ranked Shoe Sales and Percent Share
By Region and Total for Boots,
Sandals, Slippers, and Sport Shoes
(in Four Regions)

● Total - $2,356,373 - 100%

● Middle East - $872,955 - 37.0%

● Africa - $669,470 - 28.4%
● Pacific - $588,908 - 25.0%

● Asia - $225,040 - 9.6%

CrossRoads SignPost Chart,
Like Vbar Chart But Total Bar

**Ranked Shoe Sales and Percent Share By Region and Total
for Boots, Sandals, Slippers, and Sport Shoes (in Four Regions)**

● Total - $2,356,373 - 100%
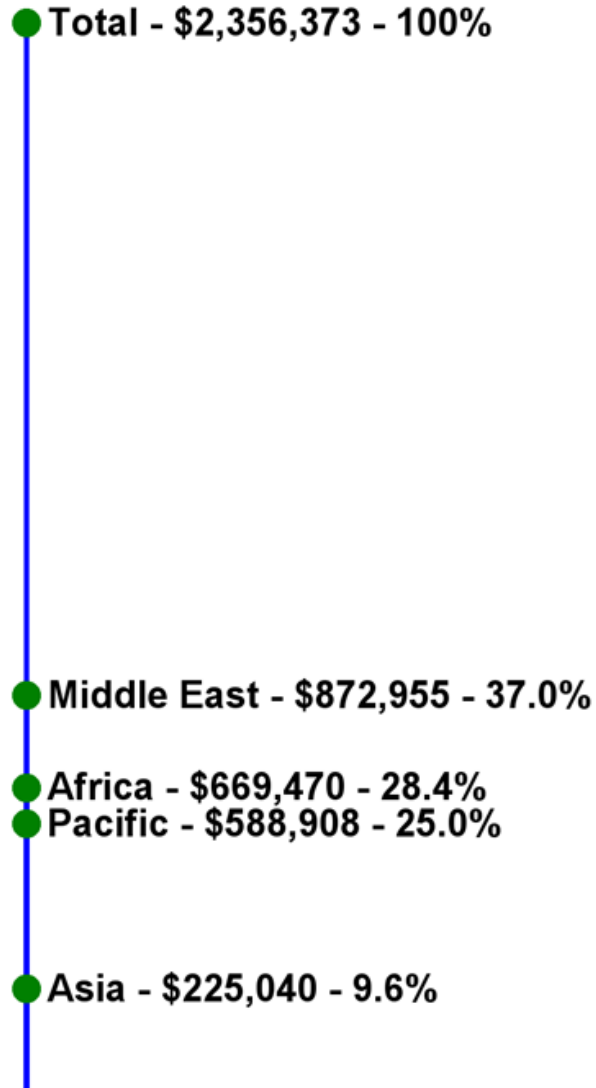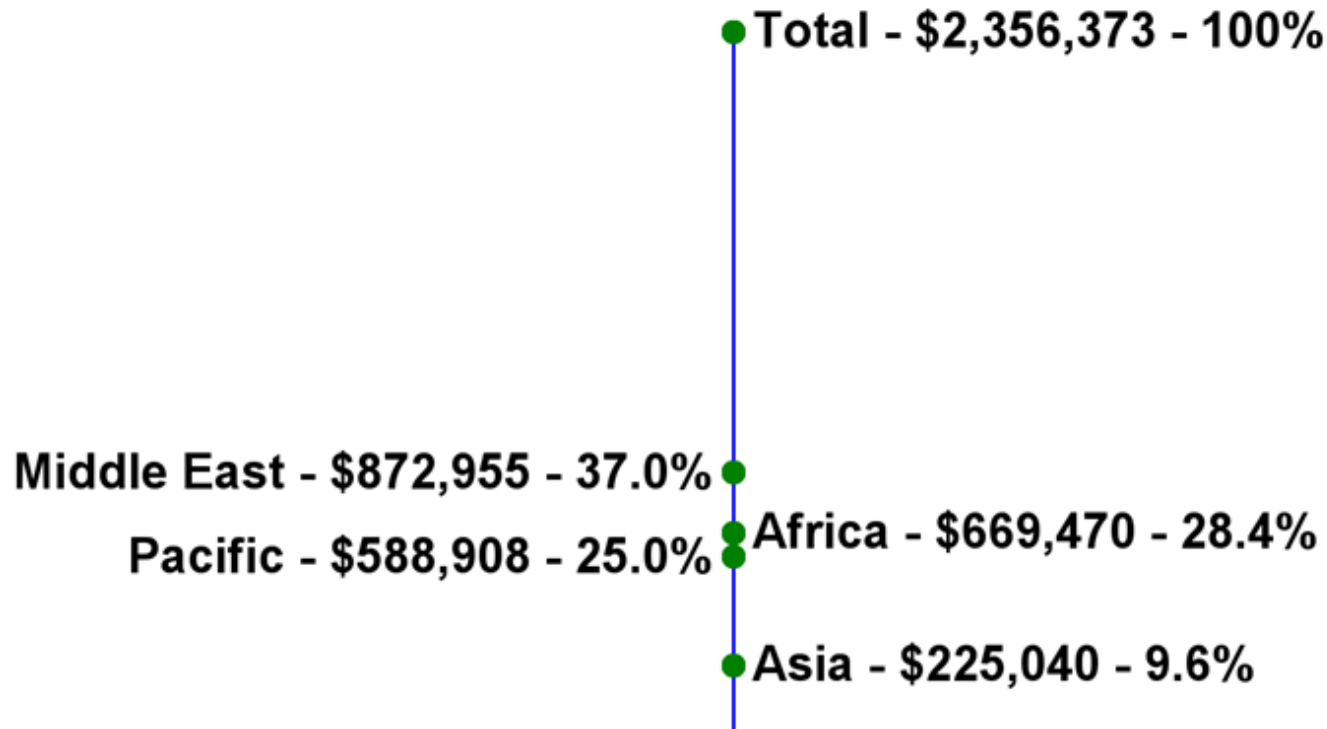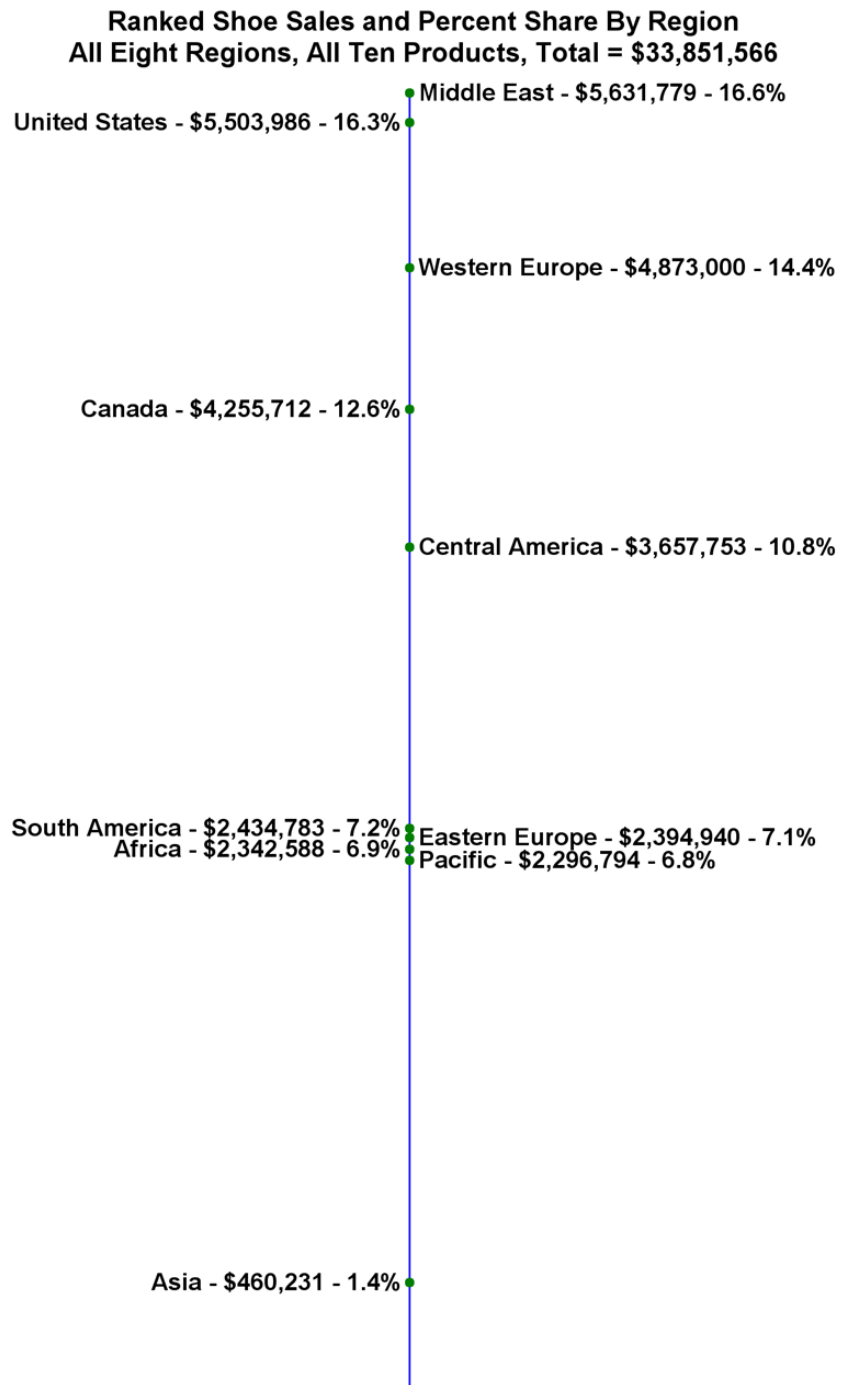
Middle East - $872,955 - 37.0% ●

Pacific - $588,908 - 25.0% ● Africa - $669,470 - 28.4%

● Asia - $225,040 - 9.6%

CrossRoads SignPost Chart (All Regions, No Total)

Like an OverLaid Vertical Bar Chart

Ranked Shoe Sales and Percent Share By Region
All Eight Regions, All Ten Products, Total = $33,851,566

Middle East - $5,631,779 - 16.6%

United States - $5,503,986 - 16.3%

Western Europe - $4,873,000 - 14.4%

Canada - $4,255,712 - 12.6%

Central America - $3,657,753 - 10.8%

South America - $2,434,783 - 7.2%   Eastern Europe - $2,394,940 - 7.1%
Africa - $2,342,588 - 6.9%   Pacific - $2,296,794 - 6.8%

Asia - $460,231 - 1.4%

Pie Chart Alternative

Lelia McConnell at SAS helped me with the color folderol.

Ranked Shoe Sales and Percent Share By Region (Decreasing Upwards)

Total - $33,851,566 - 100%

Asia - $460,231 - 1.4%

Pacific - $2,296,794 - 6.8%

Africa - $2,342,588 - 6.9%

Eastern Europe - $2,394,940 - 7.1%

South America - $2,434,783 - 7.2%

Central America - $3,657,753 - 10.8%

Canada - $4,255,712 - 12.6%

Western Europe - $4,873,000 - 14.4%

United States - $5,503,986 - 16.3%

Middle East - $5,631,779 - 16.6%

Ranked Shoe Sales and Percent Share By Region (Decreasing Upwards)

Total - $33,851,566 - 100%

Asia - $460,231 - 1.4%

Pacific - $2,296,794 - 6.8%

Africa - $2,342,588 - 6.9%

Eastern Europe - $2,394,940 - 7.1%

South America - $2,434,783 - 7.2%

Central America - $3,657,753 - 10.8%

Canada - $4,255,713 - 12.6%

Clipping of Top

Canada - $4,255,712 - 12.6%

Western Europe - $4,873,000 - 14.4%

United States - $5,503,986 - 16.3%

Middle East - $5,631,779 - 16.6%

Clipping of Bottom

There are, of course, alternatives to the preceding needle plots, but these can be **functional** and **communication-effective** alternatives to the familiar.

If interested in any images in these slides that are not found in the book, send me an email request for the code. Such cases are the SAS M08 map and both doubly-annotated bubble plots.
NOTE: In the book, "the vertical bar chart without bars" for Shoe Sales By Region omits grand total, omits percents, and omits the 0 that IS used to start the Y axis. Also, that near-analogue in the book is By Product, not By Region.

Omitted Here
All of the code
Dozens of graph & color design principles
Hundreds of examples
See the book . . .

Your comments and questions are welcome. Contact me at: Le_Roy_Bessler@wi.rr.com

LeRoy Bessler PhD
Bessler Consulting and Research
Visual Data Insights™
Strong Smart Systems™