

Best Practices for Graphic Design & Color Use,
Plus Uncommon Data Art for Data Science
You Can Make Data
Quickly Easily Completely Understood

LeRoy Bessler PhD
Bessler Consulting and Research
Visual Data Insights™
Strong Smart Systems™
BesslerLeRoy2024@gmail.com
Valparaiso, Indiana, USA

SAS® is a registered trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.
Other product and brand names are trademarks or registered trademarks of their respective owners.
Visual Data Insights™ and Strong Smart Systems™ are trademarks of LeRoy Bessler PhD.

LeRoy Bessler PhD is a data artist, the world's longest serving and most frequent advisor to SAS users on best practices for graphic design and use of color, and the author of Visual Data Insights Using SAS® ODS Graphics: A Guide to Communication-Effective Data Visualization.

He also has specialized in Software-Intelligent Application Development for reliability, reusability, maintainability, flexibility, and extendibility, as well as ODS tools (Excel, PDF, HTML5, LAYOUT) to deliver SAS output and reports, tabular and/graphic. LeRoy is a data analyst and SAS programmer, who has supported SAS servers, SAS software, SAS analytic data, and hundreds of SAS users at employers and client sites.

He is a quarterly contributor to the VIEWS News online newsletter. He has presented at SAS users conferences in the USA, Canada, Ireland, England, and Germany, and has provided pre- and post-conference training. An advocate for, and provider of, SAS user mutual education opportunities, LeRoy was event organizer for the Wisconsin Illinois SAS Users Group for most years from 1989 to 2016. For the MidWest SAS Users Group he served as Conference Chair in 1991, Assistant to Co-chairs in 1993, and Program Chair in 2010.

Acknowledgements In Brief

Thanks to

Susan McDermott

Philip R. Holland

Marcia Surratt

Lelia McConnell

Martin Mincey

Kathryn McLawhorn

Amber Elam

Cyrus Bradford

Liz Edwards

Acknowledgements In Full

The book which serves as a resource for many more guidelines and examples, and for the code related to some of this presentation was created between the Summer 2020 and December 2022. For the complete Acknowledgements pertinent, please see Page xiii in the book. Here I will say . . . I am thankful to Apress Executive Editor Susan McDermott for engaging me to write the book, and to Philip R. Holland who did the technical review. It was Phil who got me connected to Susan McDermott. Without Susan and Phil, there would have been no book.

I am grateful to Marcia Surratt, Lelia McConnell, Martin Mincey, Kathryn McLawhorn, Amber Elam, Cyrus Bradford, and Liz Edwards at SAS Technical Support who handled my problems and questions during that project. Since 1981 I have been striving to get the best from SAS graphics software, and writing and speaking about that topic since 1991. I appreciate all of the excellent SAS Technical Support help that I have received since 1981 from too many people to recall and enumerate. Thanks to All of You.

Creators of the Most Frequently Used Data Graphics

Pie Chart - William S. Playfair

Time Series Plot - William S. Playfair

Bar Chart - Philippe Buache and Guillaume de L'Isle

Thank You

NOTE: As one of its earliest users,
Playfair is often credited as the creator of the Bar Chart.

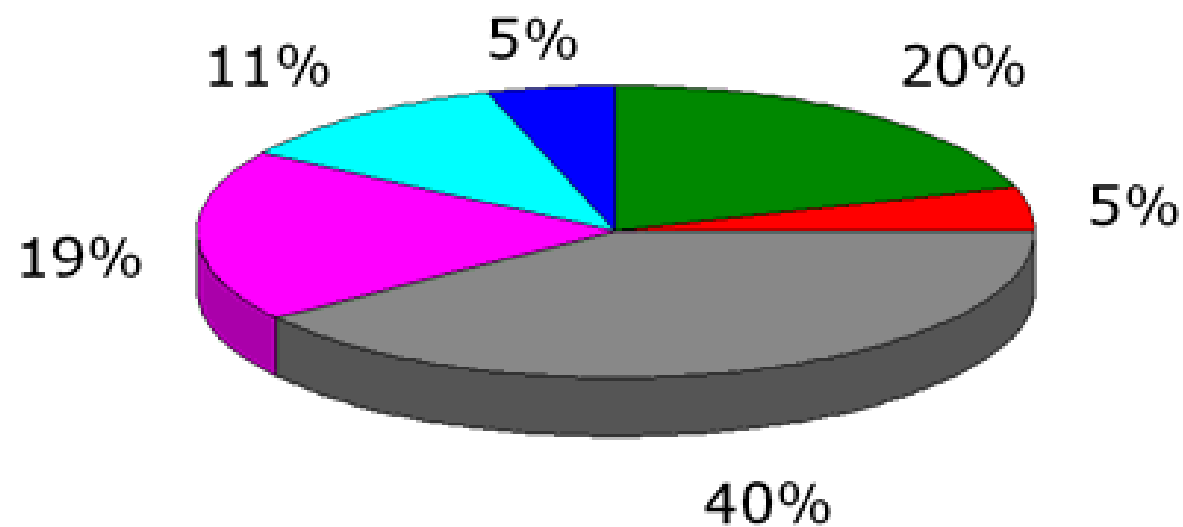
3D

Is Best Used for THREE Variables

The Only Real Danger with a Pie Chart

Figure 1: 3D Pie Chart.

Always Distorts Apparent Relative Size of Slices.

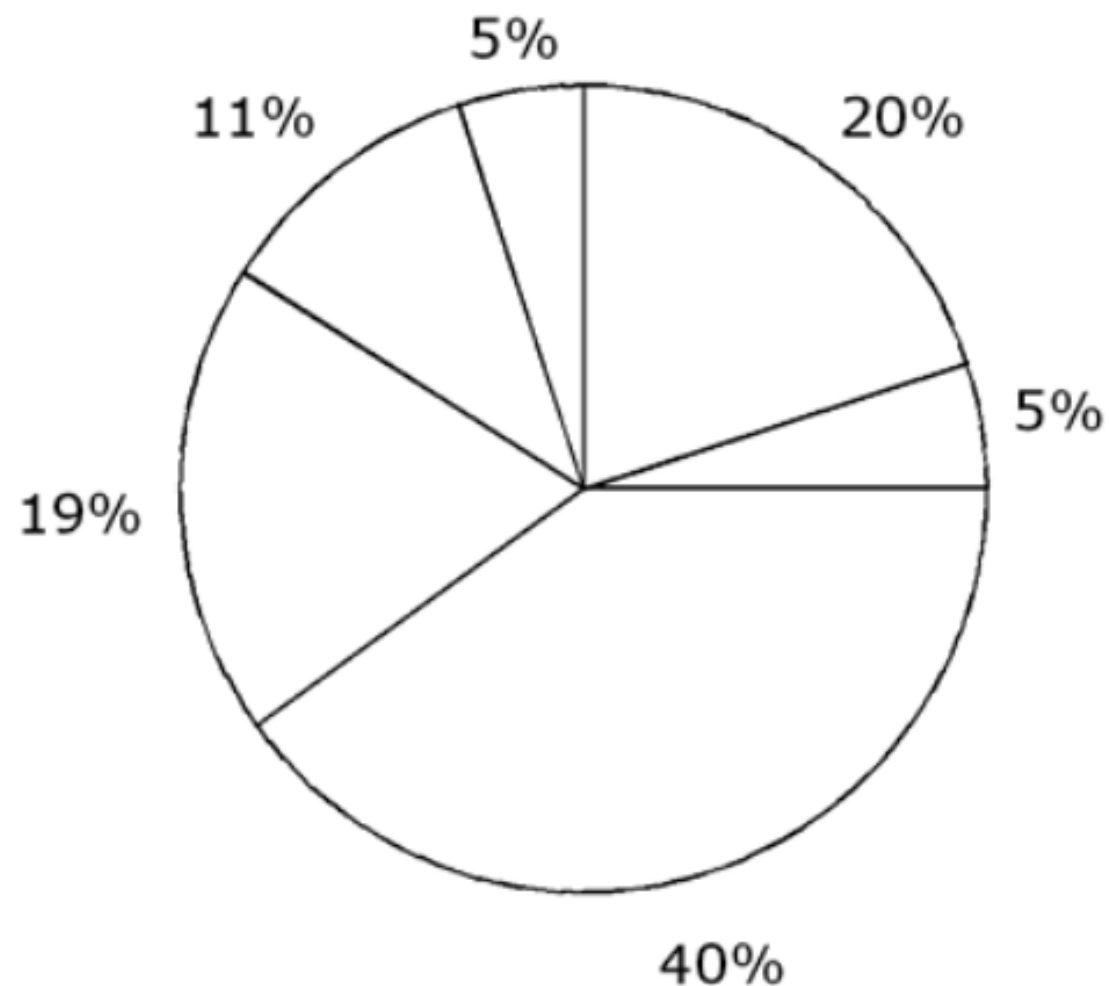


5% slice at rear is twice as big as 5% slice at right
11% slice is almost as big as 19% slice

The Solution

Figure 2: 2D Pie Chart. Simpler is Better.

Always Presents Accurate Relative Size of Slices.



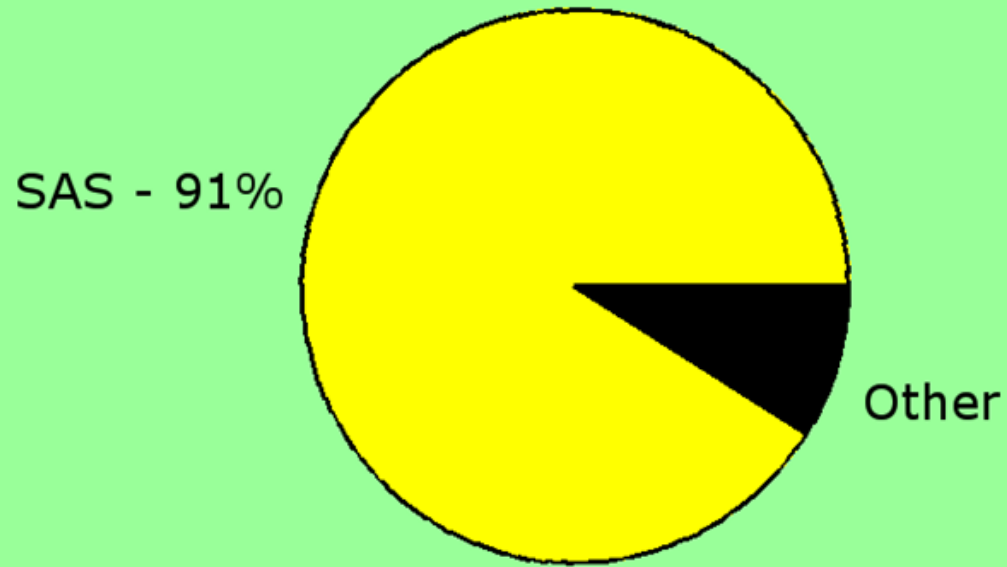
Show Them What's Important

A Design Principle Easily Implemented

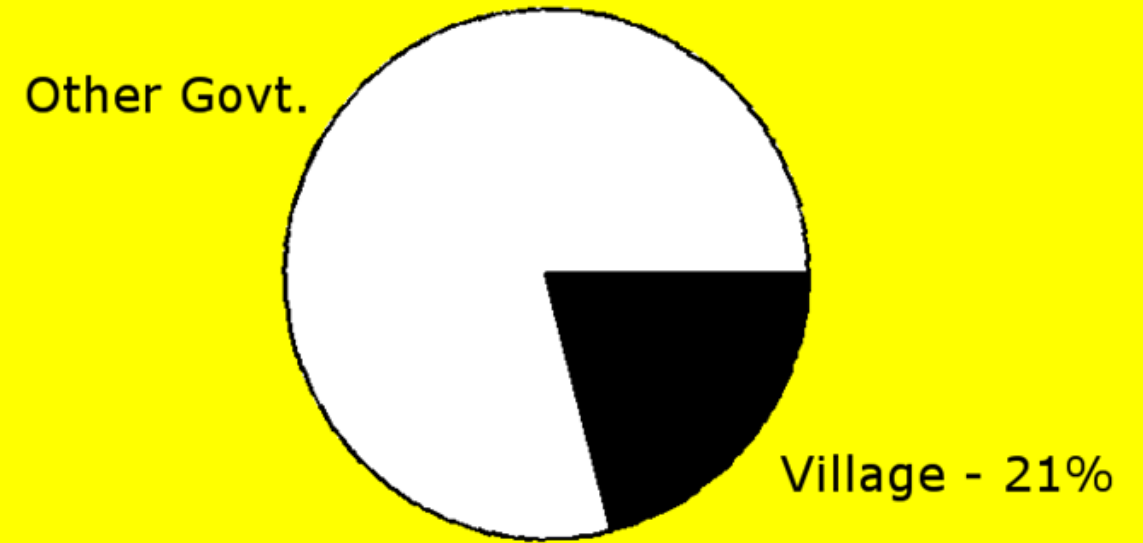
With the Simplest Ever, But Powerful, Graph

I called it "The Pac-Man Pie Chart"

SAS Software Dominates 1993 Mainframe Data Analysis Market



Most of Your Property Tax is NOT for Local Government



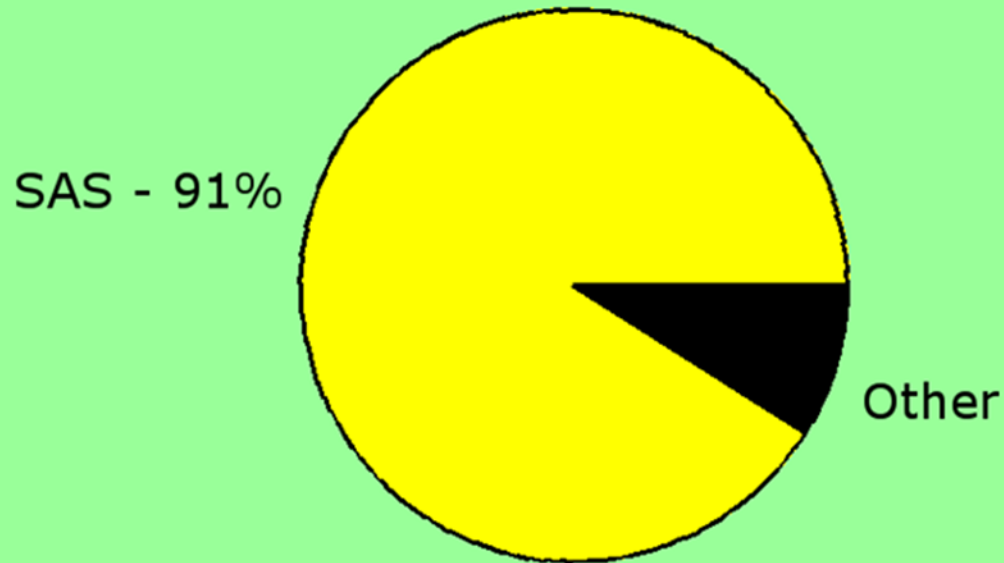
Show Them What Is Important with The Extremes of 'Other'

Tip: Make your graph title a headline

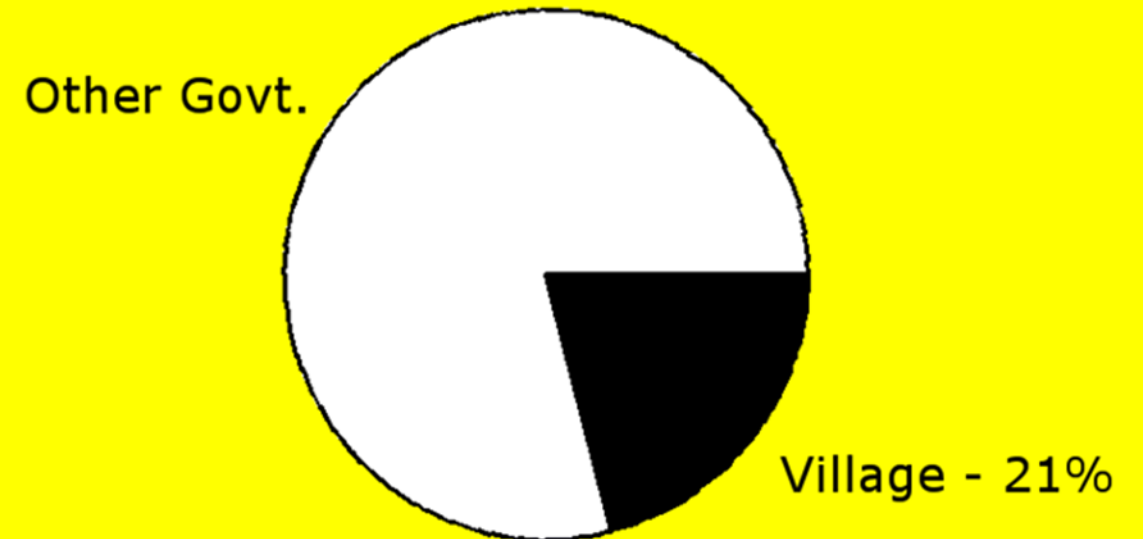
Tell the viewer the conclusion or revelation.

Do not simply describe it. Do not force the viewer to guess it.

SAS Software Dominates 1993
Mainframe Data Analysis Market



Most of Your Property Tax
is NOT for Local Government



Make Data Quickly Easily Completely Understood

- Deliver **image** + **precise numbers**
(for **quick easy** inference + **reliable accurate inference**)
- Assure text readability
- Assure color distinguishability
- Provide a headline for your graph if possible
(TELL the meaning/conclusion)

Provide Image AND Precise Numbers

Image – For Impression

- Immediate
- Easy
- Memorable (“Sticky”)

Precise Numbers – For Knowledge

- Clear
- Certain
- Correct (“The Right Answer”)

Getting to Precise Numbers

- Axes, Tick Marks, and Grid Lines **cannot deliver precise numbers**
- They are tools to ESTIMATE where a plot point coordinates or a bar end value MIGHT BE along the axis
- Then doing mental arithmetic, with the difference between embracing tickmarks and the approximate distance from the axis point to nearest tick mark
- **NET: a guess, not certainty**

For Readable Text and Clarity of Graphic Detail

- Bold Sans Serif fonts are Best
11 pt Arial Bold is an always readable standard
- Provide a plain solid color background
NO color gradient, texture, photo, or frame as background
- White is always a satisfactory background
- Black is always a satisfactory text color
- As discussed in an addendum,
black is not always the software default text color.

Assure color distinguishability

- Use red/blue, NOT red/green, for color coding bad/good
- Make lines and text thick enough
- Make markers and legend color swatches big enough
- No continuous color gradient legends
- In your KEYLEGEND statement
if you cannot use FILLASPECT=GOLDEN, then use AUTOITEMSIZE

Graph OR Table?

The answer often is
both Graph and Companion Table,
which ALWAYS delivers Image AND Precise Numbers

Watch For:

- Graph with Built-In Table
- Graph, in effect, AS Table

Categorical Data

Categorical Data

- If bar chart, use horizontal (to fit category, rank, value, and percent)
- Show Them What's Important
 - Use Ranking
 - Physically/Visually order the bars, slices, etc.
 - Provide Rank number for each category
 - Consider Subsetting
 - Top N
 - Cutoff / Filter (for goal to achieve, or threshold for concern)
 - Enough of Top Ranked for X Percent of Grand Total
 - If Subsetting, also provide an All Categories Chart (Include Rank)
- Clarify the context for the response values
 - Provide percent of grand total for each category
 - List the grand total in the chart title or subtitle

Show Them What's Important

Show them what's important:

Present graphic elements (e.g., bars) in RANKED order

Clarify the context with:

Rank of the Response Value (Relative Importance)

Percent of Grand Total for the Response Value

Grand Total of the Response Value

➤ On the Internet Commentariat
there **CONTINUES** to be
this tiresomely repeated criticism:
“Bar lengths are easier to compare than
sizes of pie slices”

➤ My long recommended and simple solution:

**Let the pie chart DO the comparison
and GIVE YOU the precise numbers**

Pie Charts are **EVERYWHERE**

because people **UNDERSTAND** them **AT A GLANCE**

A Pie Chart That Never Fails!

The legend eliminates any possible pie slice label collisions

Ranked Sales & Percent Share By Region - Total = \$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%

Is this really just a Rank-Ordered Table with a Visual Companion?

Never-Fail Pie Chart is Ranked and Fully Informative It Always Informs Even For Vanishingly Small Slices [Here all slices ARE visible.]

Ranked Sales & Percent Share By Region - Total = \$33,851,566

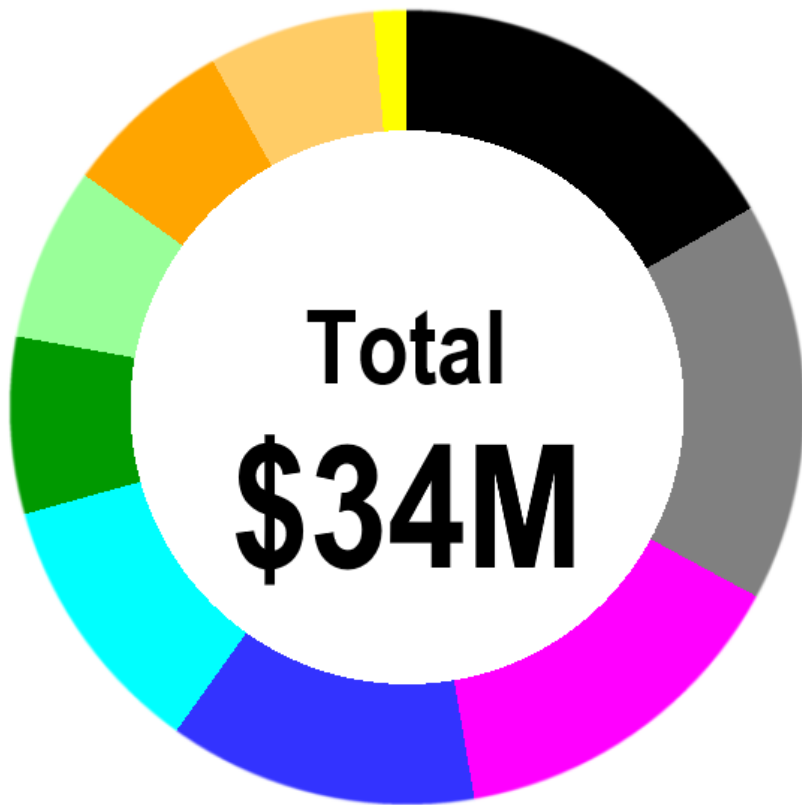


01:MiddleEast:\$5,631,779:16.64%
02:UnitedStates:\$5,503,986:16.26%
03:WesternEurope:\$4,873,000:14.40%
04:Canada:\$4,255,712:12.57%
05:CentralAmerica/Caribbean:\$3,657,753:10.81%
06:SouthAmerica:\$2,434,783:7.19%
07:EasternEurope:\$2,394,940:7.07%
08:Africa:\$2,342,588:6.92%
09:Pacific:\$2,296,794:6.78%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

Donut Chart – Pie With Hole In The Middle

For Total as \$33,851,566: Omit Hole Label and put Total in the title

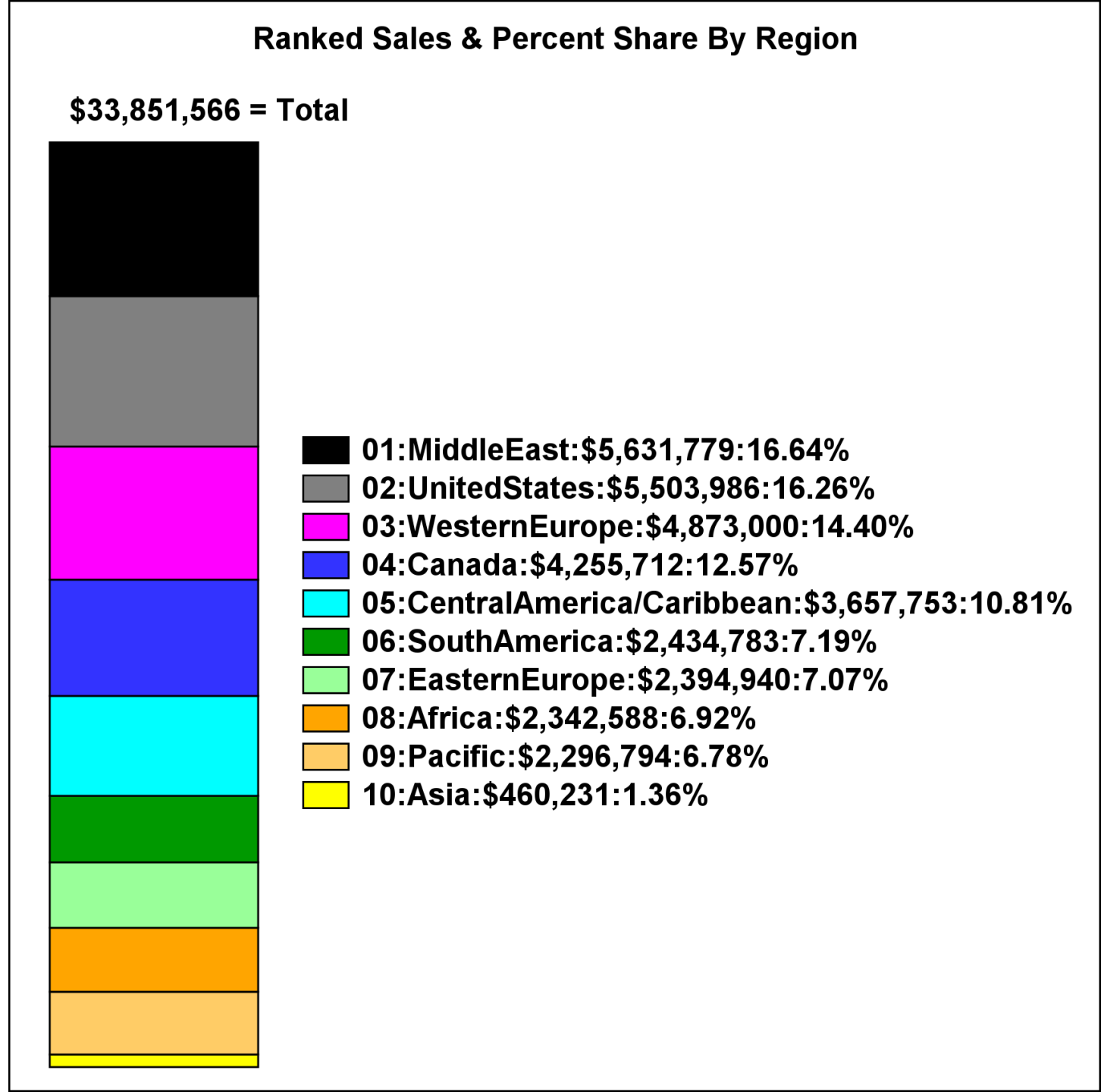
Ranked Sales & Percent Share By Region



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%

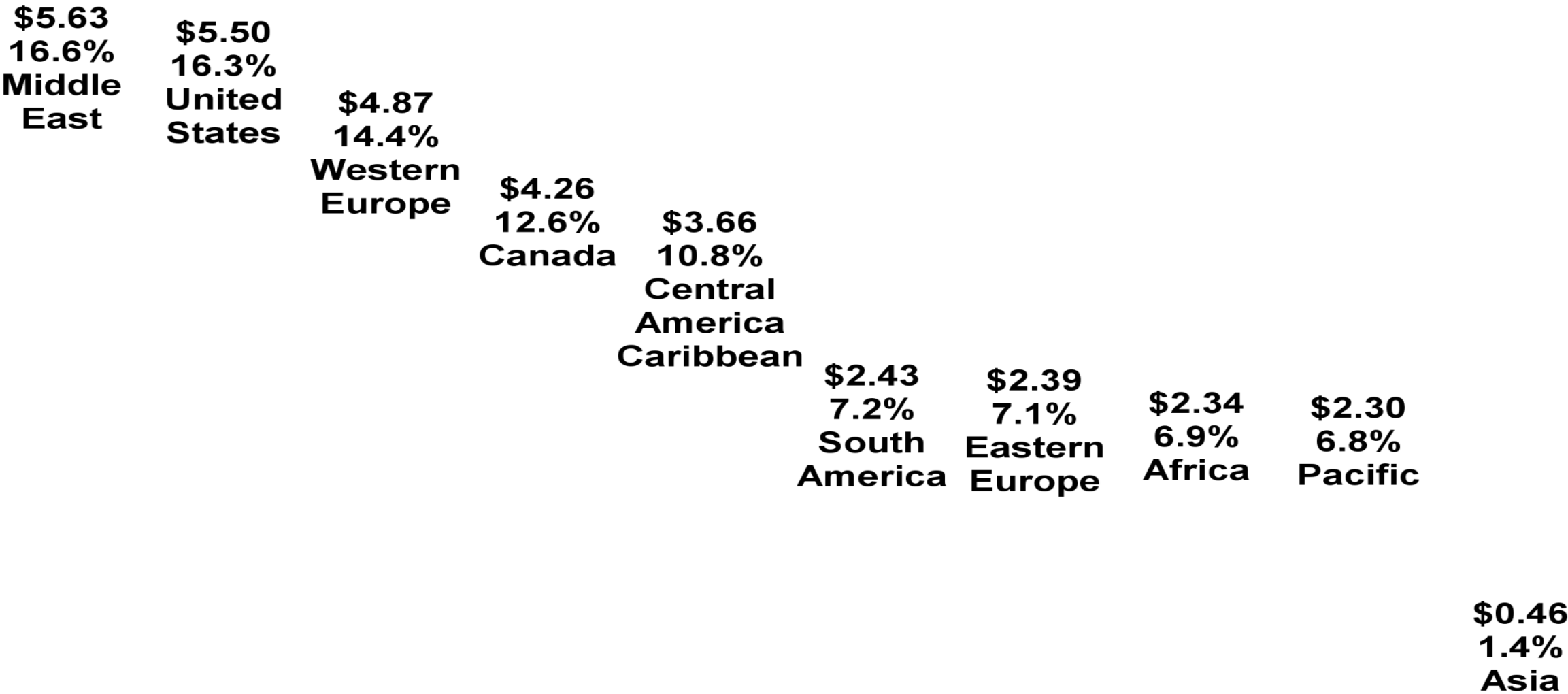
Another Rank-Ordered Table with a Visual Companion

Stacked Vertical Bar Chart (As a Pie Chart Alternative) A Rank-Ordered Table with a Visual Companion



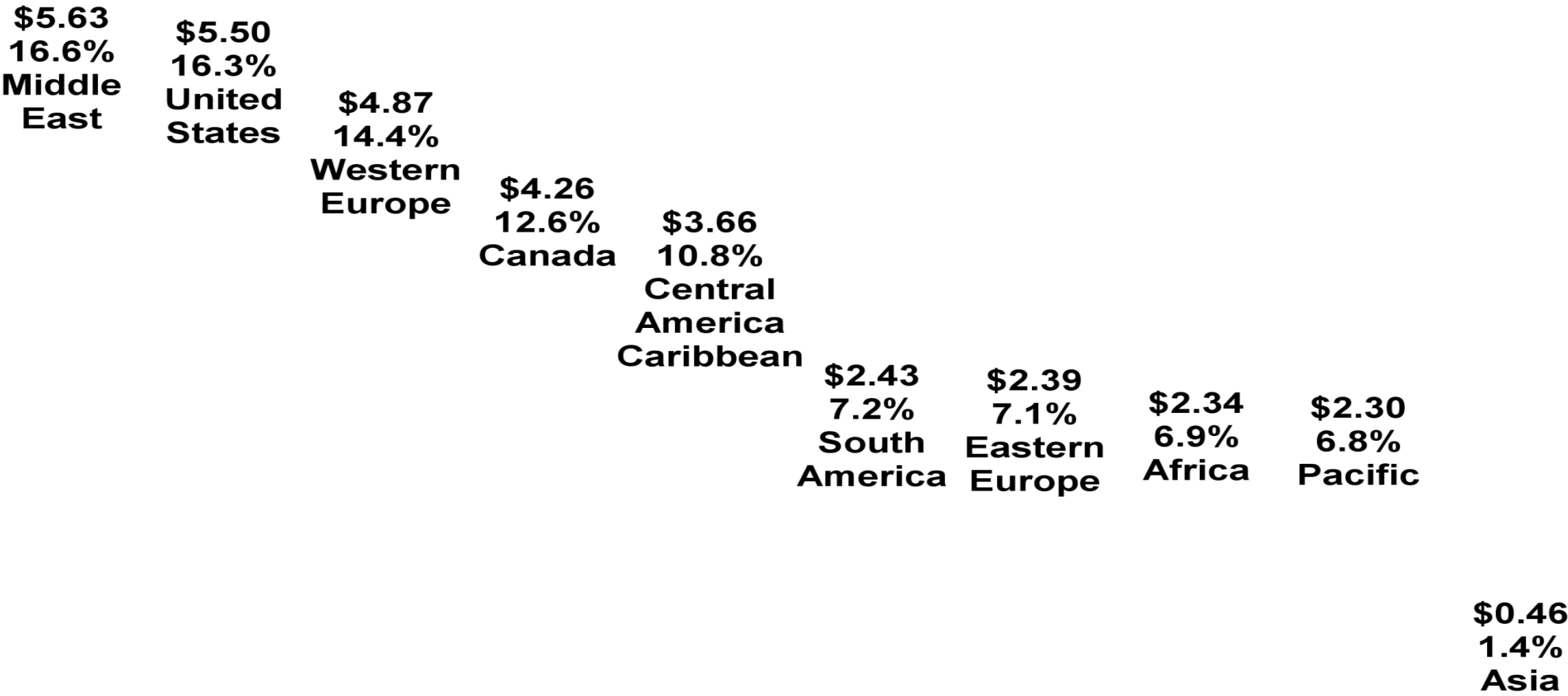
Simplest Vertical Bar Chart Delivers Maximum Information

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



For Bar Demanders: "This is a bar chart with white bars."

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

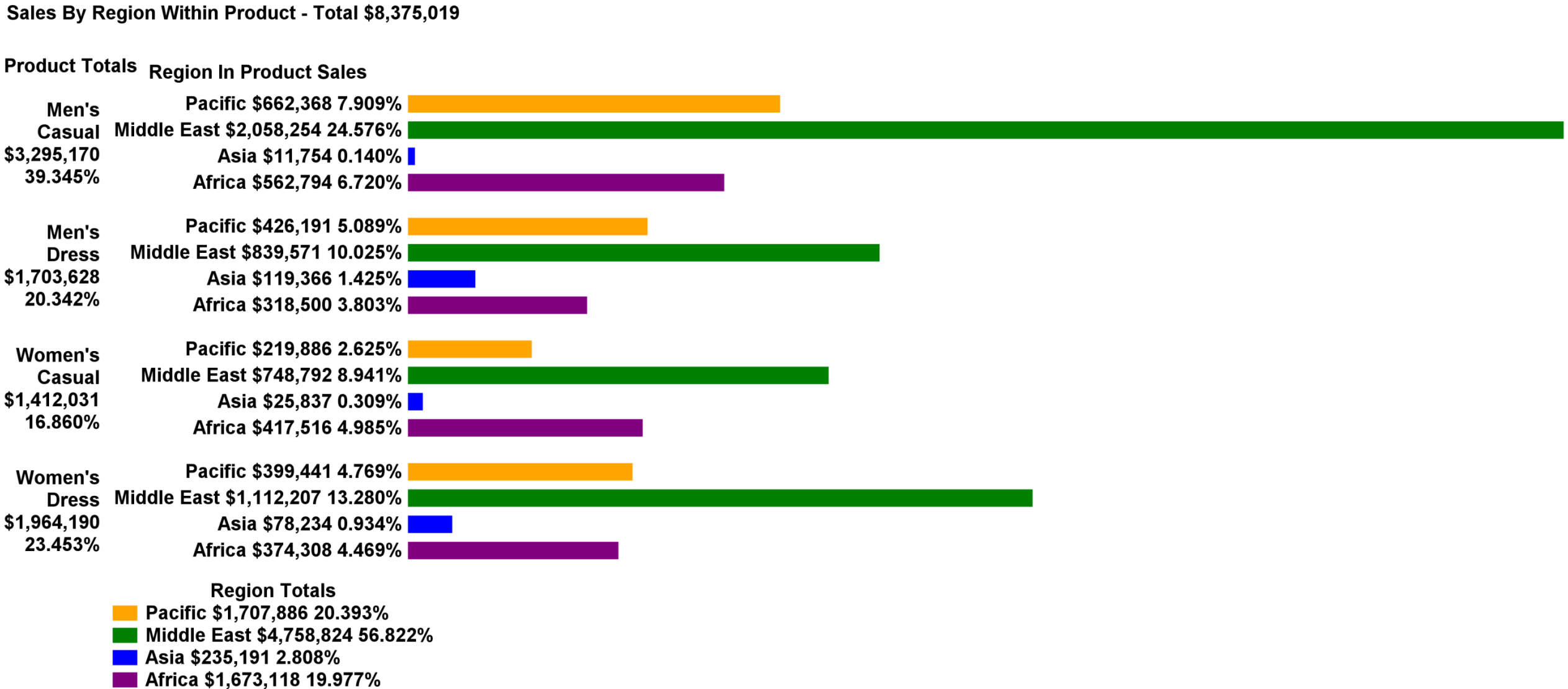


Categorical Data with a Group Variable

- For data with Two Descriptive Variables (Categories), in SAS Language,
one variable is assigned as Category =
the other variable is assigned as Group =
- Each category value or group value
has a Subtotal of Response
and each of those Subtotals has a Percent of Grand Total.
With data preprocessing, you **can** manage to deliver them.
- See my book for why Dot Plots are poor for grouped data

The Ultimate Clustered Bar Chart Has EVERYTHING!

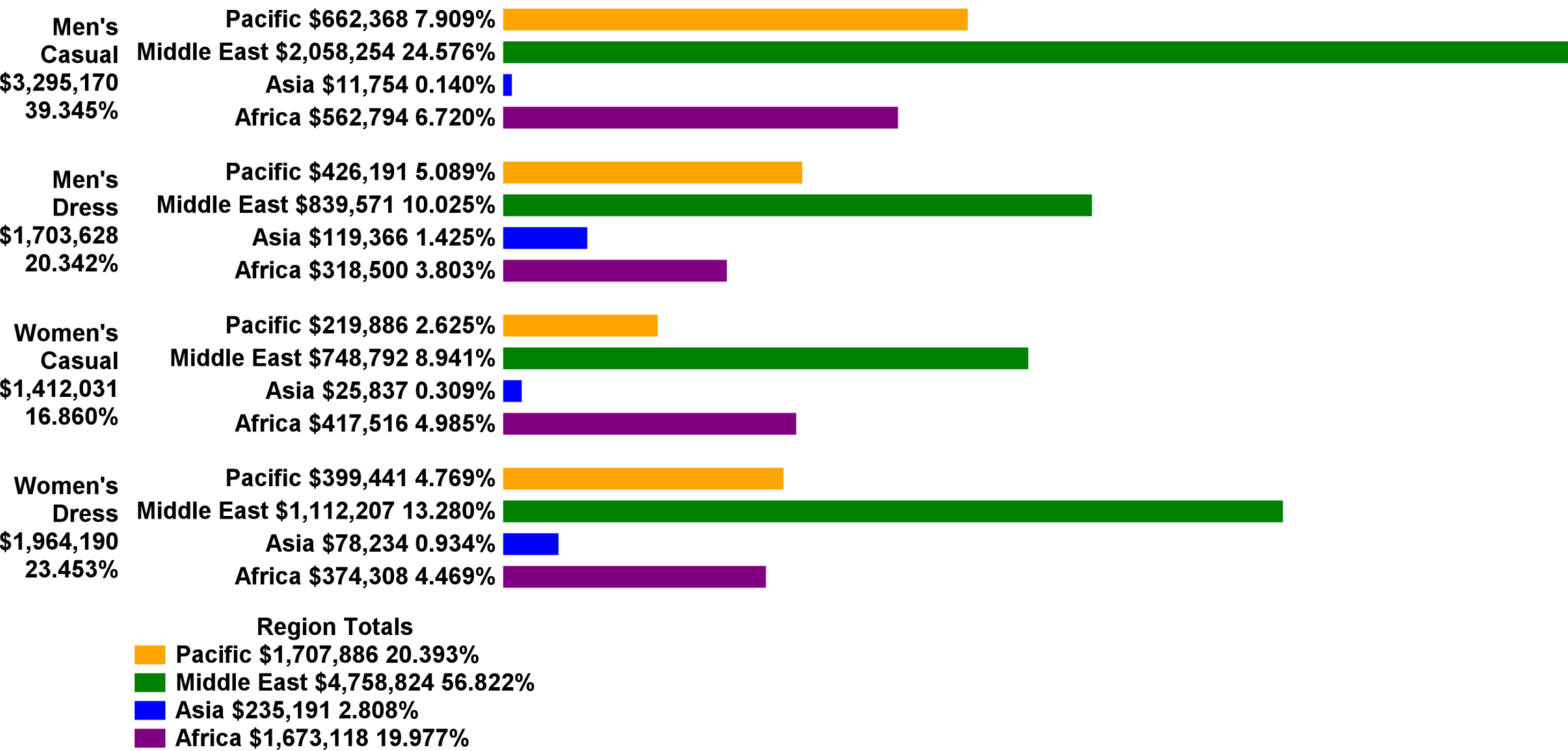
For each Region, Product, and Region-Product: the Category, Value, and Percent of Total are provided, and the Total is in the Title.



Sales By Region Within Product - Total \$8,375,019

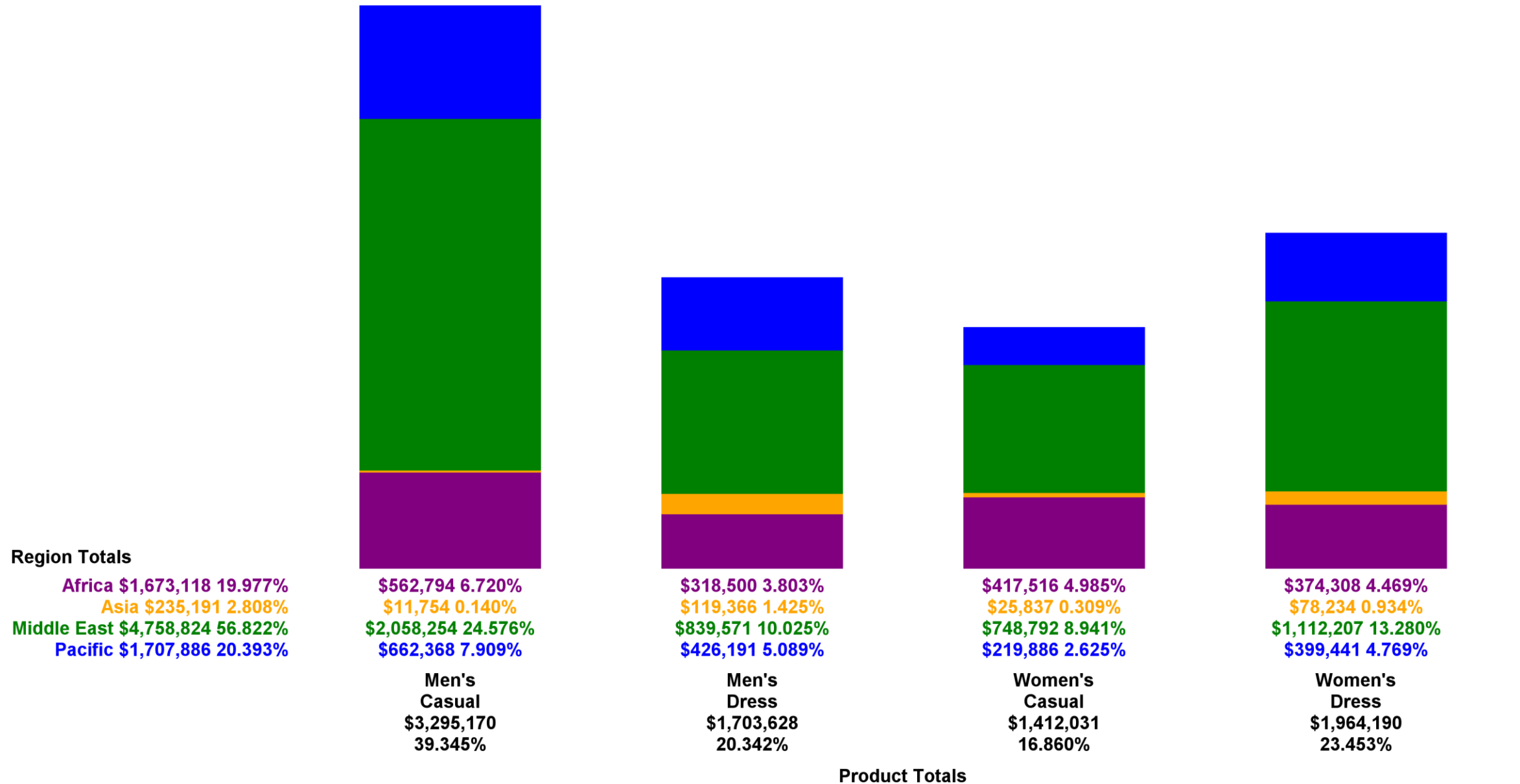
Image Magnified, Longest Bar Clipped

Product Totals Region In Product Sales



The Ultimate Stacked Bar Chart Has EVERYTHING!

Sales By Region Within Product - Total \$8,375,019



A Closer Look at the Lower Left of the Image

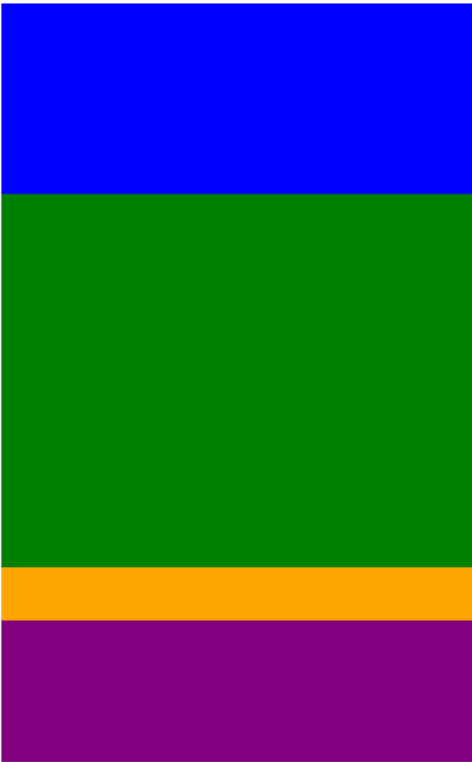
Chart works
even when
segments are
vanishingly
small ►

Region Totals

Africa	\$1,673,118	19.977%
Asia	\$235,191	2.808%
Middle East	\$4,758,824	56.822%
Pacific	\$1,707,886	20.393%



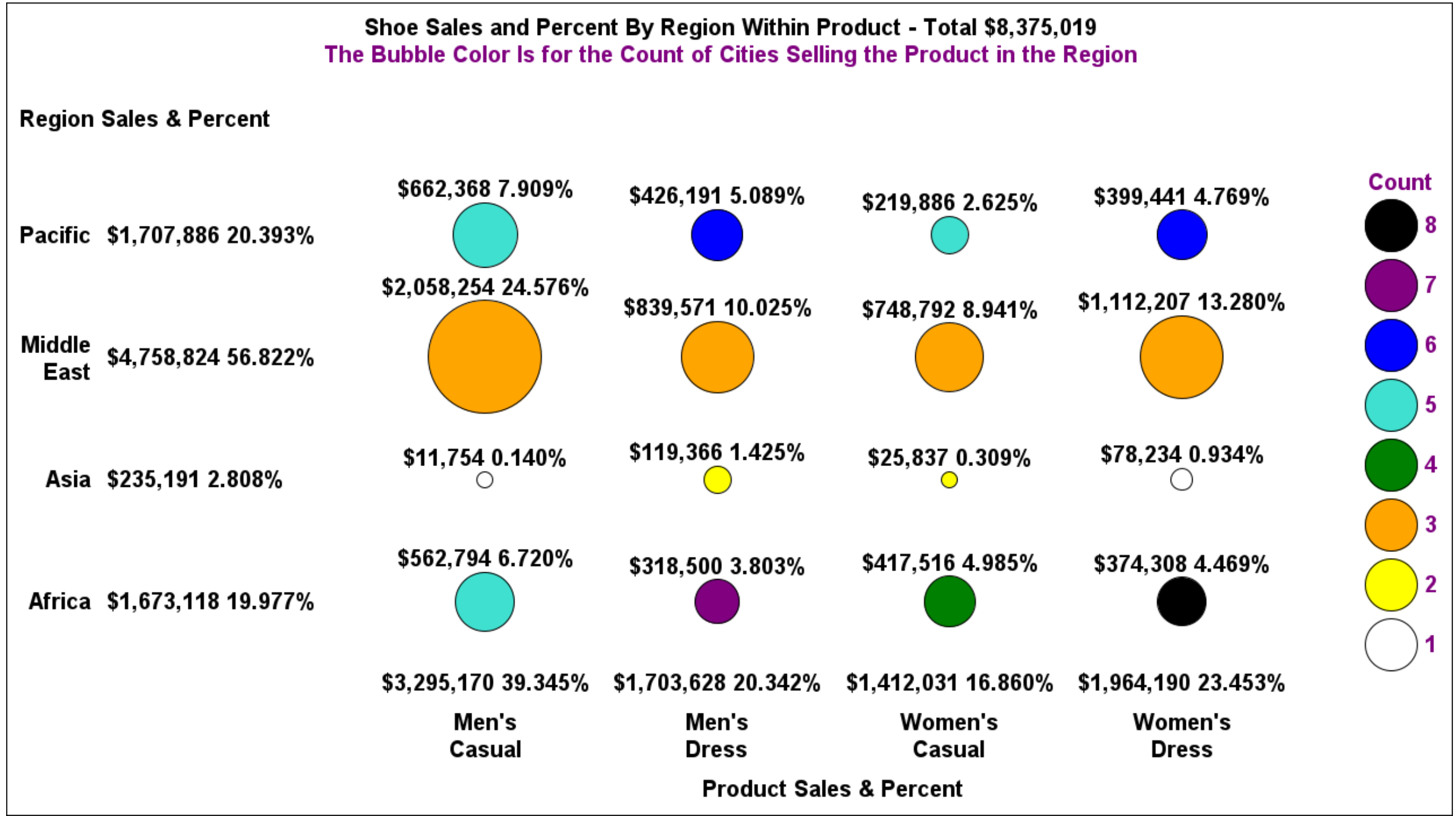
Men's
Casual
\$3,295,170
39.345%



Men's
Dress
\$1,703,628
20.342%

Product Totals

The Ultimate Bubble Plot Has EVERYTHING + MORE!

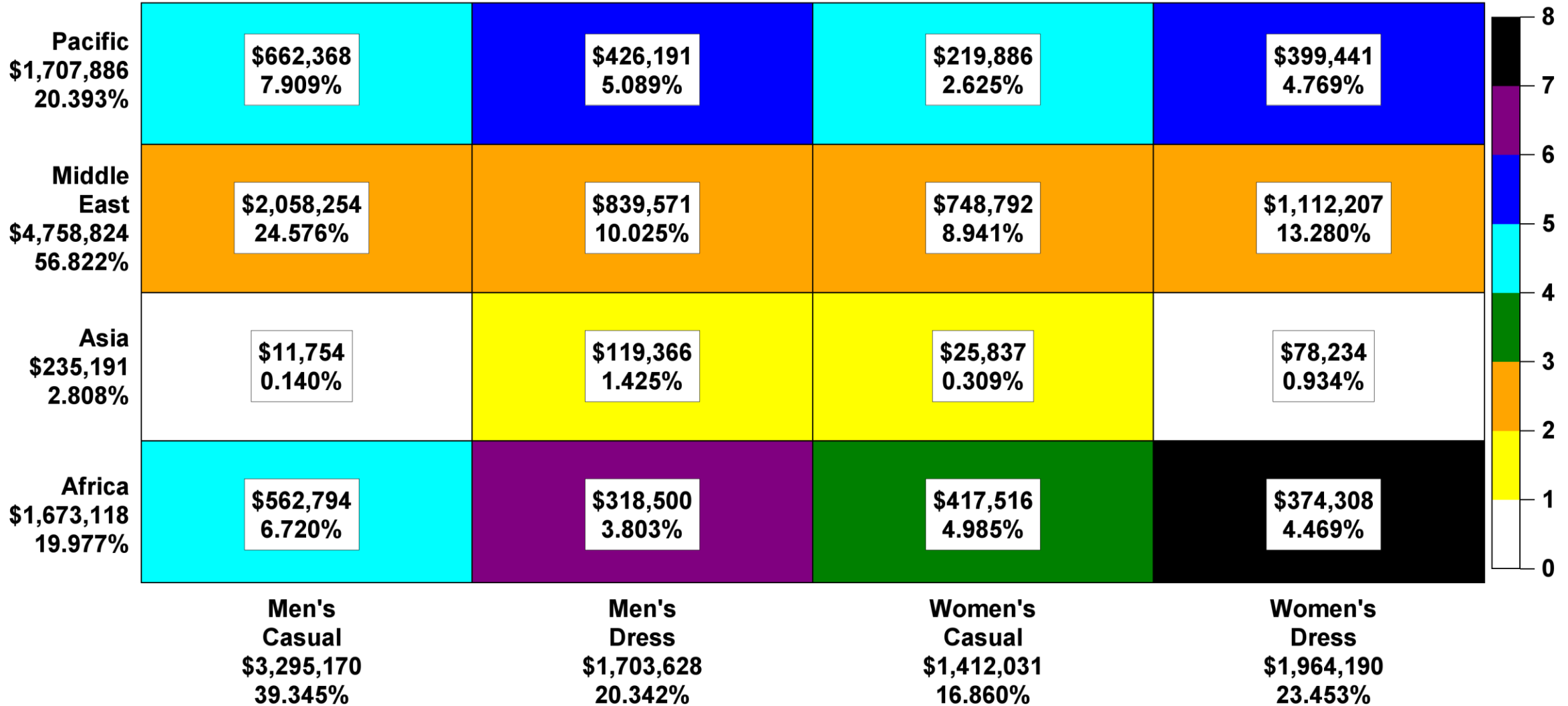


A Heat Map Analogue for The Ultimate Bubble Plot

Shoe Sales and Percent By Region Within Product - Total \$8,375,019

Area Color Is for the Count of Cities (see the legend) Selling the Product in the Region

Region & Sales & Percent



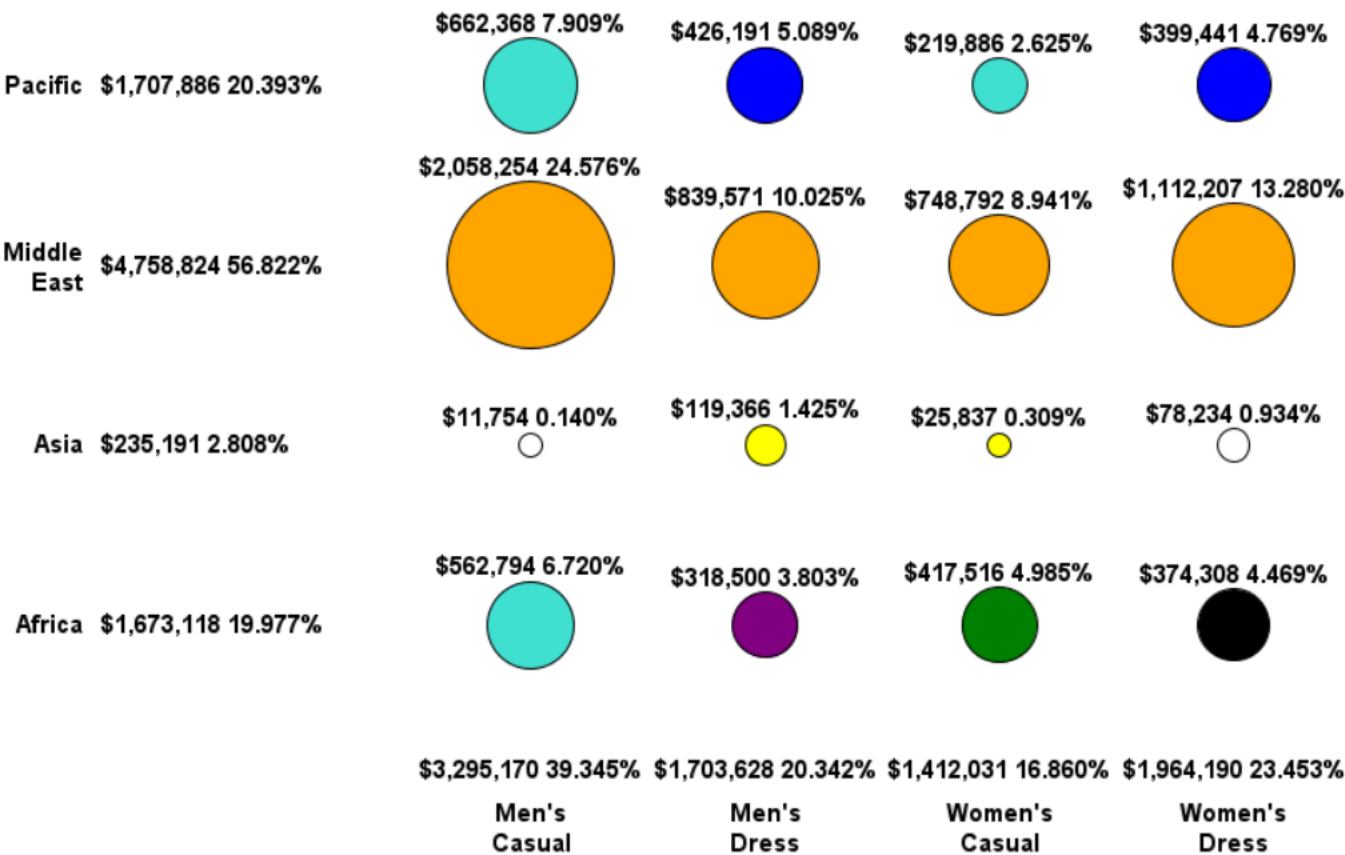
Product & Sales & Percent

Web-Enabled Ultra-Ultimate Bubble Plot + Listing of Cities in Alphabetic Order

Shoe Sales and Percent By Region Within Product - Total \$8,375,019
The Bubble Color Is for the Count of Cities Selling the Product in the Region
[Go To This Same Bubble Plot with a List of Cities By Region](#)

All Cities Rank Value Percent::
Addis Ababa : 7 : \$304,518 : 3.636%
Al-Khobar : 3 : \$1,069,980 : 12.776%
Algiers : 8 : \$277,597 : 3.315%
Auckland : 15 : \$65,841 : 0.786%
Bangkok : 19 : \$8,422 : 0.101%
Cairo : 4 : \$706,829 : 8.440%
Canberra : 16 : \$65,482 : 0.782%
Dubai : 2 : \$1,412,913 : 16.871%
Jakarta : 6 : \$583,959 : 6.973%
Johannesburg : 17 : \$42,682 : 0.510%
Khartoum : 14 : \$94,910 : 1.133%
Kinshasa : 11 : \$108,538 : 1.296%
Kuala Lumpur : 9 : \$259,542 : 3.099%
Luanda : 12 : \$100,942 : 1.205%
Manila : 5 : \$633,626 : 7.566%
Nairobi : 18 : \$37,102 : 0.443%
Seoul : 10 : \$226,769 : 2.708%
Singapore : 13 : \$99,436 : 1.187%
Tel Aviv : 1 : \$2,275,931 : 27.175%

Region Sales & Percent



Product Sales & Percent

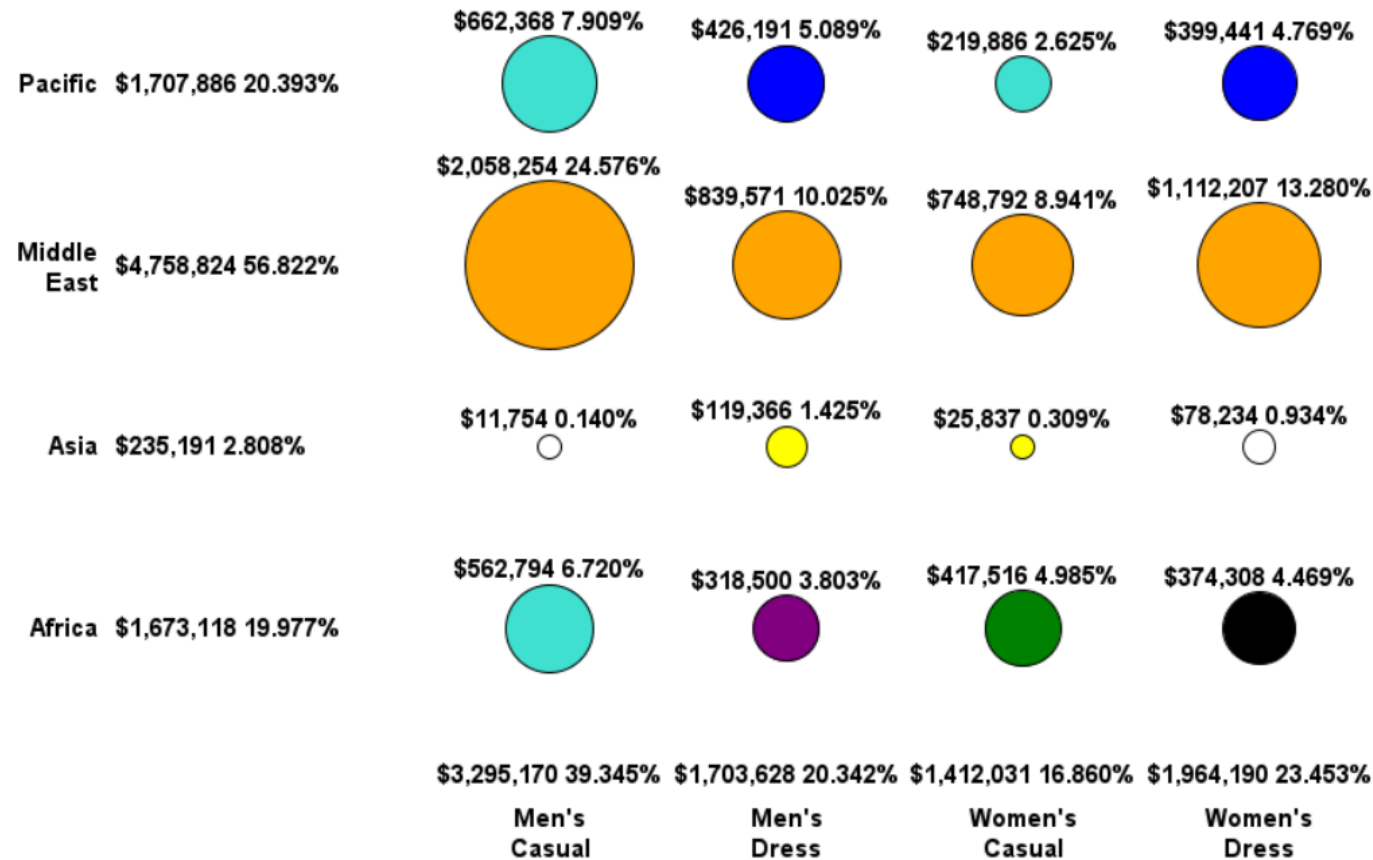
Web-Enabled Ultra-Ultimate Bubble Plot + Listing of Cities By Region

Shoe Sales and Percent By Region Within Product - Total \$8,375,019

The Bubble Color Is for the Count of Cities Selling the Product in the Region

[Go To This Same Bubble Plot with a List of Cities in Alphabetic Order](#)

Region Sales & Percent



Count



Africa Cities Rank Value Percent:

Cairo : 4 : \$706,829 : 8.440%
Addis Ababa : 7 : \$304,518 : 3.636%
Algiers : 8 : \$277,597 : 3.315%
Kinshasa : 11 : \$108,538 : 1.296%
Luanda : 12 : \$100,942 : 1.205%
Khartoum : 14 : \$94,910 : 1.133%
Johannesburg : 17 : \$42,682 : 0.510%
Nairobi : 18 : \$37,102 : 0.443%

Asia Cities Rank Value Percent:

Seoul : 10 : \$226,769 : 2.708%
Bangkok : 19 : \$8,422 : 0.101%

Middle East Cities Rank Value Percent:

Tel Aviv : 1 : \$2,275,931 : 27.175%
Dubai : 2 : \$1,412,913 : 16.871%
Al-Khobar : 3 : \$1,069,980 : 12.776%
Pacific Cities Rank Value Percent:
Manila : 5 : \$633,626 : 7.566%
Jakarta : 6 : \$583,959 : 6.973%
Kuala Lumpur : 9 : \$259,542 : 3.099%
Singapore : 13 : \$99,436 : 1.187%
Auckland : 15 : \$65,841 : 0.786%
Canberra : 16 : \$65,482 : 0.782%

Product Sales & Percent

Realization while preparing these slides

To Show Them What's Important,
I emphasize the importance of Ranking . . .

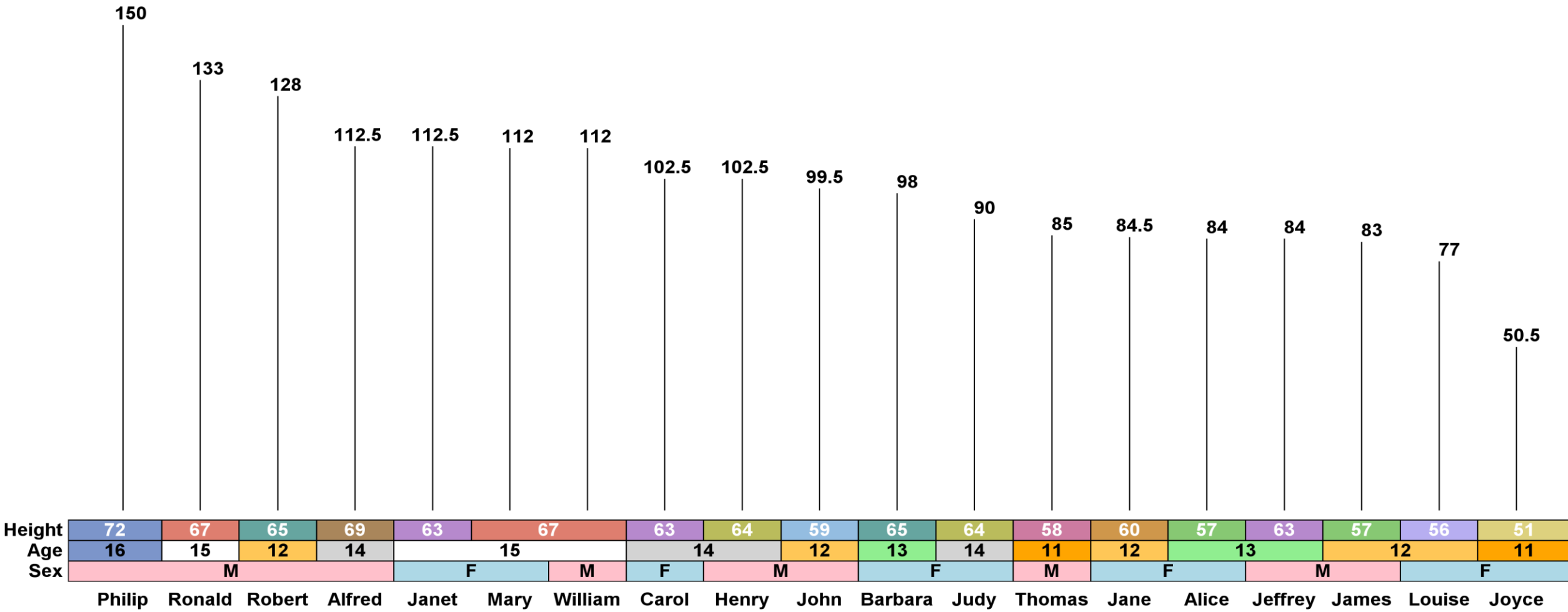
So, a viewer MIGHT wonder:

What is the rank of Sales for Each Product-Region pair?

- ▶ Easy to see, visually, on a clustered hbar chart
- ▶ More difficult for stacked vbar chart & bubble plot
- ▶ Impossible for heat map

**For My Next Edition of Charts with Category & Group
I must add Rank to each Response-Percent label to
REALLY deliver EVERYTHING**

Students Ranked By Weight (Pounds) with Height (Inches Rounded), Age, and Sex



One Category Variable and Three Group Variables
FIVE variables in TWO dimensions
Needle Plot + Three Group Color-Coded Block Charts

For Categorical Data
Show the Viewer What's Important
With:

- **Ranking**
- **Subsetting**
- **Enough**
- **But All of It as an optional companion**

Horizontal Bar Chart Is Best

Lots of space to fit the

Rank Number, Category, Value,
and Percent at the left of each bar

They are Essential Information for Every Bar

For Categorical Data, To Show Them What's Important

- Display categories/bars/pie slices/donut bites in ranked response value order

	Rank	Cases		% of World Cases
United States of America	1	103,436,829		13.3727
China	2	99,322,525		12.8408
India	3	45,010,891		5.8192

Put values in context:
































- Provide **percents of grand total**, not only the response values
- In title or subtitle, **display the grand total**
- If many categories, provide an alphabetic order companion for easy lookup of any category of special interest to viewer

A Real Data Example Using Subsetted Ranked Bar Charts
































Top 30 Countries By Total Cases & By Total Deaths

[Go To Alphabetic Order by Country Table of All COVID-19 Information WorldWide](#)

Top 30 Countries Ranked by COVID-19 Total Cases - 2020 to 2023
This SubTotal of 657,229,080 Cases Is 85.0% of the World Total
All 232 Countries had a Total of 773,490,126 Cases

	Rank	Cases		% of World Cases	Case Rate
United States of America	1	103,436,829		13.3727	30.5764
China	2	99,322,525		12.8408	6.9657
India	3	45,010,891		5.8192	3.1761
France	4	38,997,490		5.0418	57.5074
Germany	5	38,437,756		4.9694	46.1051
Brazil	6	37,519,960		4.8507	17.4257
South Korea	7	34,571,873		4.4696	66.7207
Japan	8	33,803,572		4.3703	27.2716
Italy	9	26,517,706		3.4283	44.9167
United Kingdom	10	24,843,591		3.2119	36.8004
Russia	11	23,608,723		3.0522	16.3141
Turkey	12	17,004,677		2.1984	19.9255
Spain	13	13,980,340		1.8074	29.3960
Australia	14	11,710,216		1.5139	44.7341
Viet Nam	15	11,624,000		1.5028	11.8387
Argentina	16	10,075,242		1.3026	22.1384
Netherlands	17	8,627,693		1.1154	49.1214
Mexico	18	7,702,809		0.9959	6.0412
Iran	19	7,625,463		0.9859	8.6114
Indonesia	20	6,820,926		0.8818	2.4758
Poland	21	6,609,092		0.8545	16.5820
Colombia	22	6,385,539		0.8255	12.3097
Austria	23	6,081,287		0.7862	68.0263
Portugal	24	5,637,114		0.7288	54.8846
Greece	25	5,533,964		0.7155	53.2882
Ukraine	26	5,520,483		0.7137	13.9049
Chile	27	5,329,103		0.6890	27.1841
Malaysia	28	5,206,724		0.6731	15.3418
Belgium	29	4,841,934		0.6260	41.5405
Israel	30	4,841,558		0.6259	51.2388

Top 30 Countries Ranked by COVID-19 Total Deaths - 2020 to 2023
This SubTotal of 5,906,988 Mortalities Is 84.5% of the World Total
All 226 Countries had a Total of 6,994,144 Deaths

	Rank	Deaths		% of World Deaths	Mortality Rate
United States of America	1	1,144,877		16.3691	1.1068
Brazil	2	702,116		10.0386	1.8713
India	3	533,346		7.6256	1.1849
Russia	4	400,967		5.7329	1.6984
Mexico	5	334,958		4.7891	4.3485
United Kingdom	6	232,112		3.3187	0.9343
Peru	7	221,583		3.1681	4.8842
Italy	8	193,886		2.7721	0.7312
Germany	9	174,979		2.5018	0.4552
France	10	167,985		2.4018	0.4308
Indonesia	11	161,965		2.3157	2.3745
Iran	12	146,757		2.0983	1.9246
Colombia	13	142,727		2.0407	2.2352
Argentina	14	130,687		1.8685	1.2971
China	15	121,889		1.7427	0.1227
Spain	16	121,852		1.7422	0.8716
Poland	17	120,018		1.7160	1.8160
Ukraine	18	109,918		1.5716	1.9911
South Africa	19	102,595		1.4669	2.5191
Turkey	20	101,419		1.4501	0.5964
Japan	21	74,694		1.0680	0.2210
Romania	22	68,601		0.9808	1.9559
Philippines	23	66,795		0.9550	1.6004
Chile	24	64,482		0.9219	1.2100
Canada	25	53,595		0.7663	1.1281
Hungary	26	48,917		0.6994	2.1997
Czechia	27	43,225		0.6180	0.9139
Viet Nam	28	43,206		0.6177	0.3717
Bulgaria	29	38,613		0.5521	2.9227
Greece	30	38,224		0.5465	0.6907

**Top 30 Ranked Countries account for
85% of total COVID-19 cases
and
85% of total COVID-19 deaths**

Enough?

**Do you really need to initially see a chart with
approximately 200 more countries included?**

Subset Charts for Categorical Data

Show Them Enough Data!

- But provide access also to the full package
- Three Ways to **Enough**:
 - Top N
 - A Cutoff: All categories that satisfy a condition, e.g., a goal to achieve or danger threshold to avoid
 - Enough of Top Ranked Categories to Accumulate X Percent of Grand Total

Why ONLY Enough Data?

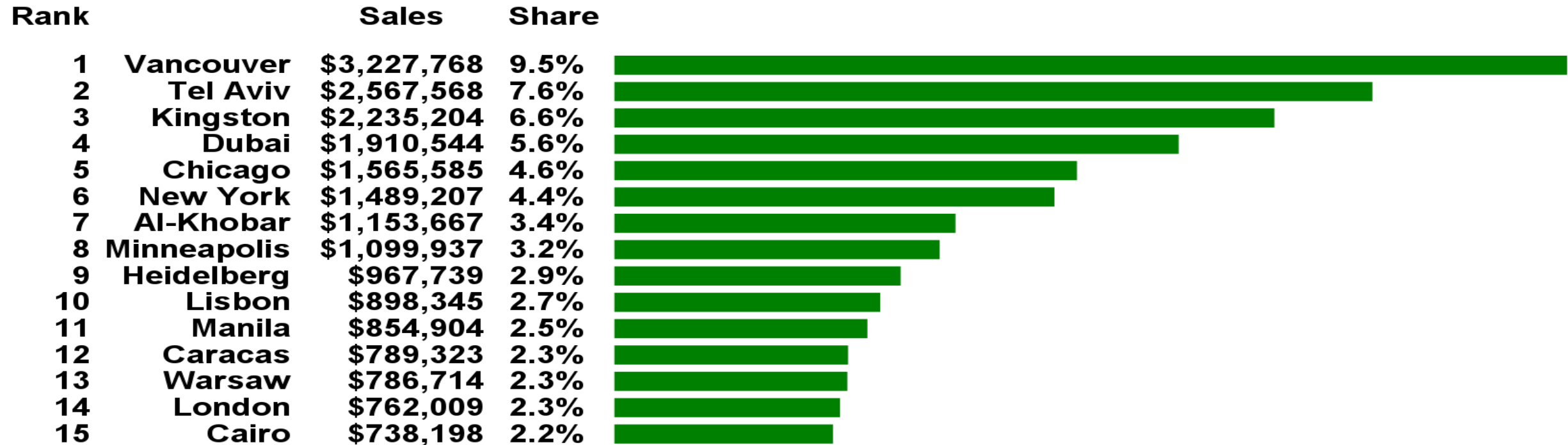
Not Every Category IS Important

Especially if the subsetting criterion

(Cutoff or Enough of the Top Ranked Values)

defines a key business/operational objective

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



Maximally Informative Horizontal Bar Chart Showing Enough
Rank, Sales, & Share columns are Y axis tables.

In Real Life, I'd pick a higher percent, but I needed a chart to fit on this slide.

Dynamic Titles

Can 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

- ▶ **Good** graphs **ANTICIPATE** and **ANSWER** questions
- ▶ Any graph that **PROMPTS** a question that could have been answered in the image itself has **failed an opportunity**
- ▶ And, the important question might not be asked!

Other Enough (focus on Top N)

Top 10 Ranked Shoes Sales By City
Selecting Only Top 10
SubTotal Sales \$17,115,564 is 50.6% of Total
All 53 Cities had Total Sales \$33,851,566

Rank		Sales	Share	
1	Vancouver	\$3,227,768	9.5%	<div></div>
2	Tel Aviv	\$2,567,568	7.6%	<div></div>
3	Kingston	\$2,235,204	6.6%	<div></div>
4	Dubai	\$1,910,544	5.6%	<div></div>
5	Chicago	\$1,565,585	4.6%	<div></div>
6	New York	\$1,489,207	4.4%	<div></div>
7	Al-Khobar	\$1,153,667	3.4%	<div></div>
8	Minneapolis	\$1,099,937	3.2%	<div></div>
9	Heidelberg	\$967,739	2.9%	<div></div>
10	Lisbon	\$898,345	2.7%	<div></div>
















Other Enough (use of a cutoff/goal/threshold)

Top 8 Ranked Shoes Sales By City
Selecting Only Cities with Sales At Least \$1000000
SubTotal Sales \$15,249,480 is 45.0% of Total
All 53 Cities had Total Sales \$33,851,566

Rank		Sales	Share	
1	Vancouver	\$3,227,768	9.5%	<div></div>
2	Tel Aviv	\$2,567,568	7.6%	<div></div>
3	Kingston	\$2,235,204	6.6%	<div></div>
4	Dubai	\$1,910,544	5.6%	<div></div>
5	Chicago	\$1,565,585	4.6%	<div></div>
6	New York	\$1,489,207	4.4%	<div></div>
7	Al-Khobar	\$1,153,667	3.4%	<div></div>
8	Minneapolis	\$1,099,937	3.2%	<div></div>

Also provide the full package (and include Rank)
In Alphabetic Order for Easy LookUp
Bar Chart is truncated here to fit in the slide

By City Shoe Sales, Percent, and Rank

	Sales	Share	Rank	
Addis Ababa	\$467,429	1.38%	26	
Al-Khobar	\$1,153,667	3.41%	7	
Algiers	\$395,600	1.17%	31	
Auckland	\$124,424	0.37%	45	
Bangkok	\$16,667	0.05%	52	
Bogota	\$206,234	0.61%	38	
Budapest	\$410,529	1.21%	30	
Buenos Aires	\$118,283	0.35%	46	
Cairo	\$738,198	2.18%	15	
Calgary	\$61,403	0.18%	51	
Canberra	\$155,547	0.46%	42	
Caracas	\$789,323	2.33%	12	
Chicago	\$1,565,585	4.62%	5	
Copenhagen	\$693,116	2.05%	17	
Dubai	\$1,910,544	5.64%	4	

Time Series Data

Time Series Data

- Annotate y and x at each point or display Y value above X value with X axis table
- OR use a web-enabled plot with mouseover text (aka "data tips") for on-point information AND, for non-vanishing inspection also provide a companion table (see the options and resources suggested in prior slides)
- OR use a static plot-table composite created with ODS LAYOUT
- For simplicity, use my Sparse Line or my Array of Multiple Sparse Lines
- Facilitate seasonality detection for time series with overlay plots
- For overlay plots create a custom maximally informative legend
- On a single-line plot, unless negative Y values, start the Y axis at zero to prevent magnifying visual effect of not really significant increases or decreases in Y value
- To fit more X axis tick mark values, use FITPOLICY=STAGGER on XAXIS statement
- When software mistakenly concludes that not all X or Y axis tick mark values can fit, use FITPOLICY=NONE on the XAXIS or YAXIS statement

In Time Series Plots

- Start the Y axis at zero, unless negative values.
- **Prevent over-reaction** to ups and downs amplified when the Y axis range is minimum to maximum.
- What is important about any Up or Down is the actual significance of that amount, not its visual appearance.
- For an overlay of multiple time series, it may be necessary to use a Y axis from min Y to max Y, especially for annotated plots that risk collisions of point label with point label, or with another line

The Fundamental Question with Time Series Plots

➤ What **ARE** the X & Y values?

In Time Series Plots

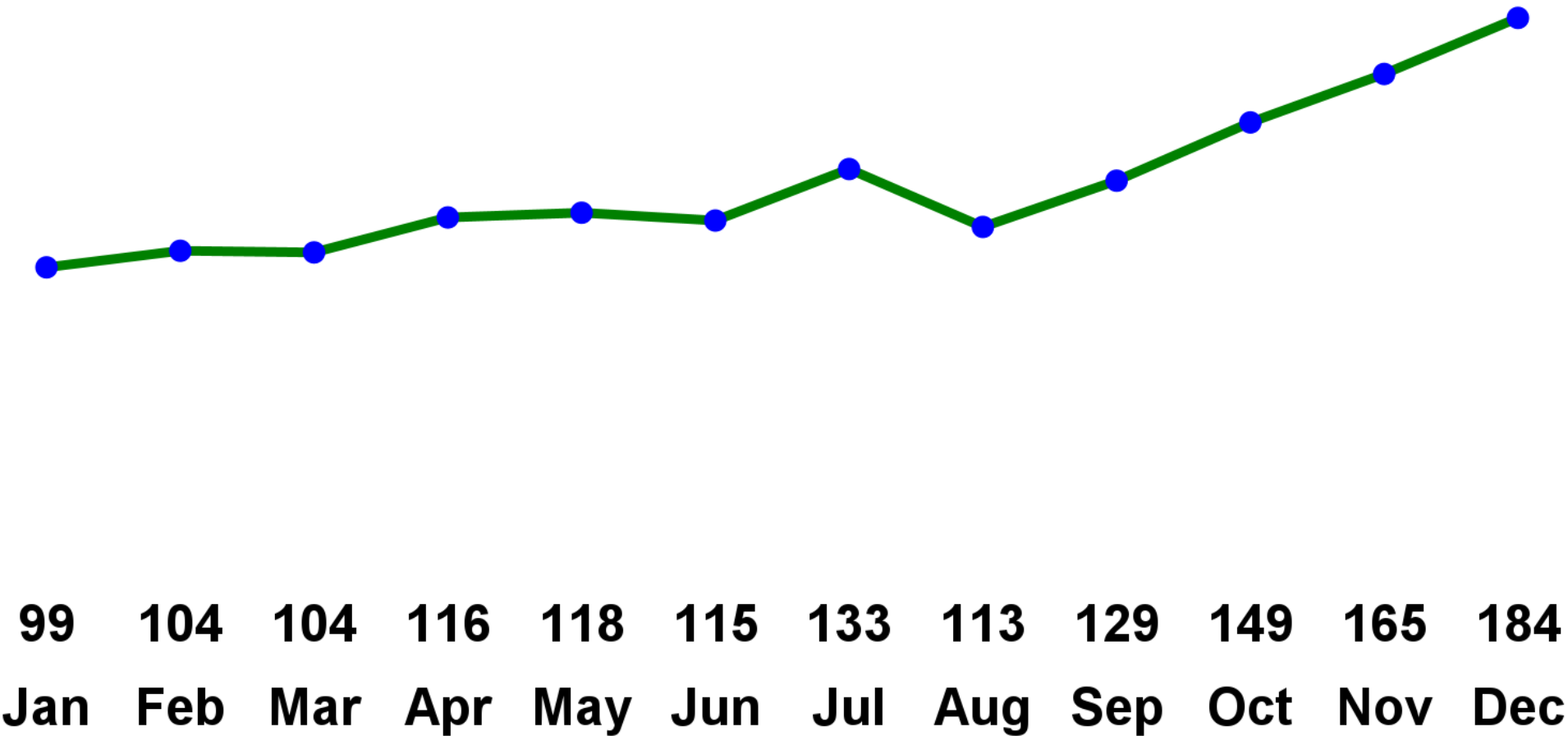
- Provide X values for each data point if possible.
If not, use STAGGER or THIN option or smaller font.

In Time Series Plots

- **Annotate the plot points with Y values** if it can be done without collisions, between labels and/or between labels and lines.
- If done, omit the Y axis.
- **If feasible, annotate X values also**, & omit the X axis.
- **When annotation is infeasible, try an X axis table**, which puts each Y value above (or below—it's your choice) its X value.

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

184.38



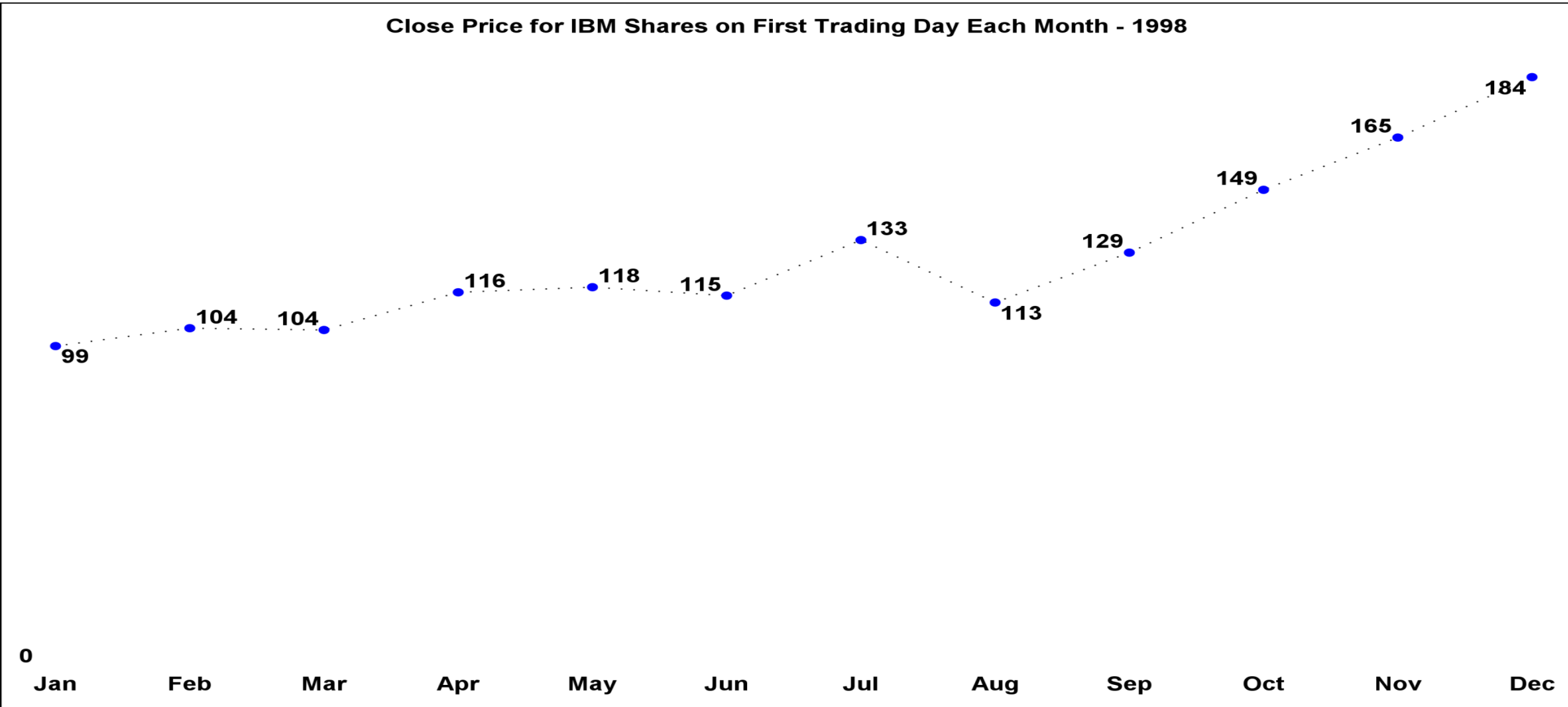
0

X Axis Table puts Y values immediate to X values
Accelerates Acquisition and Prevents Ambiguity

Y Values As Data Labels.

Light dotted line de-emphasizes collision for Y = 184.

But Y values are distant from X values.



The SAFEST Annotation:

No Line & No Markers - for labels to collide with

Y and X Values Immediate To Each Other – NO Ambiguity

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

99 Jan	104 Feb	104 Mar	116 Apr	118 May	115 Jun	133 Jul	113 Aug	129 Sep	149 Oct	165 Nov	184 Dec
-----------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

0

The Y-value-over-X-value labels Visually Show the Trend
and **ARE** the Precise Numbers

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



Month	Close Price
Jan	99
Feb	104
Mar	104
Apr	116
May	118
Jun	115
Jul	133
Aug	113
Sep	129
Oct	149
Nov	165
Dec	184

0

This IS The Trend

Omitted Trend Line and Plot Points Are Superfluous

For Safe Annotation:

- ▶ Minimizing the number of labels
MINIMIZES collision opportunities
- ▶ What should be minimum number of labels?

Sparse Annotation

- Most interesting, and often sufficient—
Date (or Time or DateTime) and Y value for:
 - Start
 - End
 - Intermediate Maximum, If Any
 - Intermediate Minimum, If Any
 - Amount and Direction of Change From the Second Last Data Point To the End Point
 - Point of Inflection, if no inner max or min

Point of Inflection

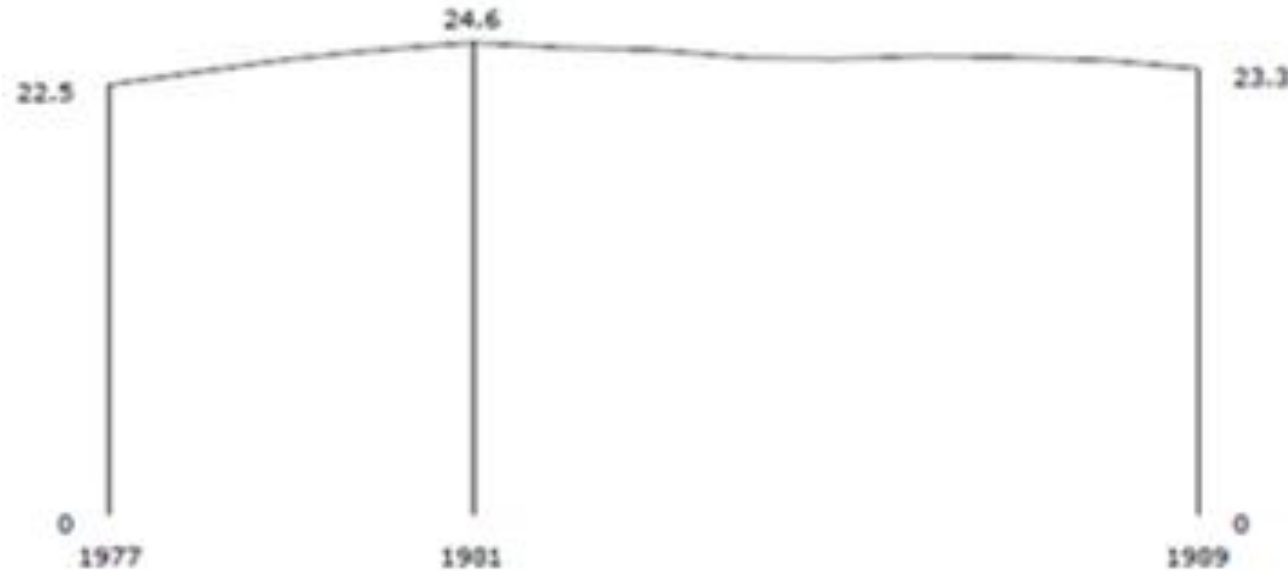
If no intermediate maximum and/or minimum, very significant would be a point of inflection. I.e., the point of a persistent trend change, in any of four ways:

- Slowly Increasing to Rapidly Increasing
- Rapidly Increasing to Slowly Increasing
- Slowly Decreasing to Rapidly Decreasing
- Rapidly Decreasing to Slowly Decreasing

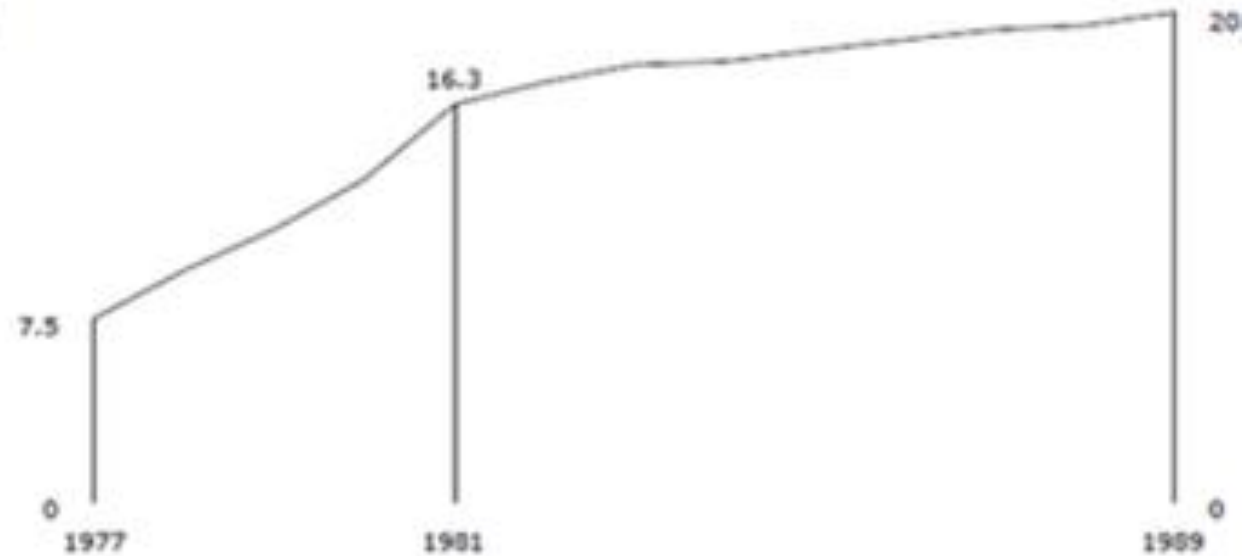
What Inspired My Sparse Annotation in 1991

Real Life Data where precise values at other points were Of No Interest

Annual U.S. Beer Consumption Peaked in 1981
Gallons per Capita



Annual Miller Lite Production Growth Slowed in 1981
Millions of Barrels



Sparse Line Annotation

Shows Them **What's Important**

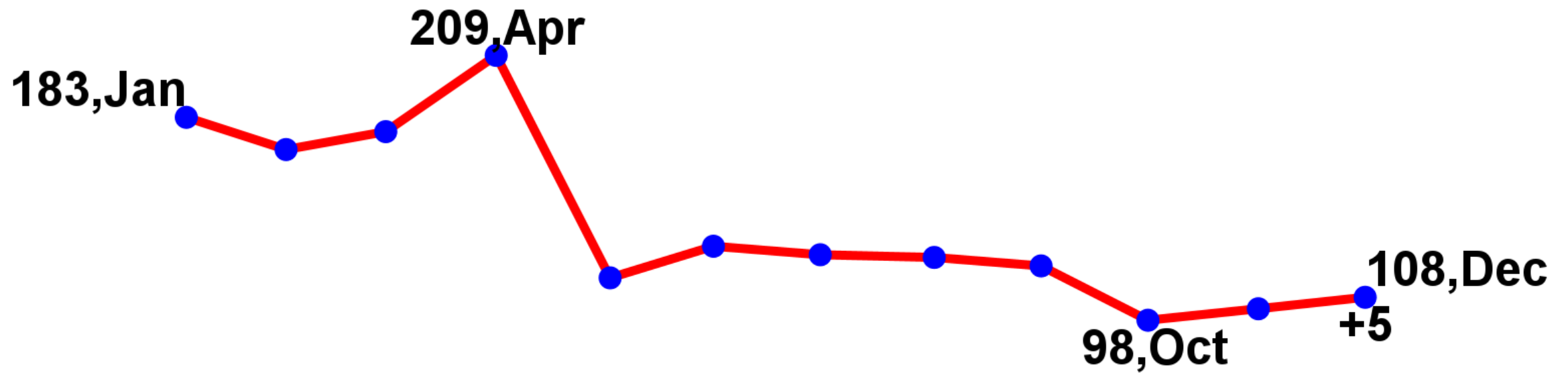
Shows Them **Enough**

Delivers both **Necessary** and **Sufficient!**

Sparse Line Annotation: Axes-Free Plot Line

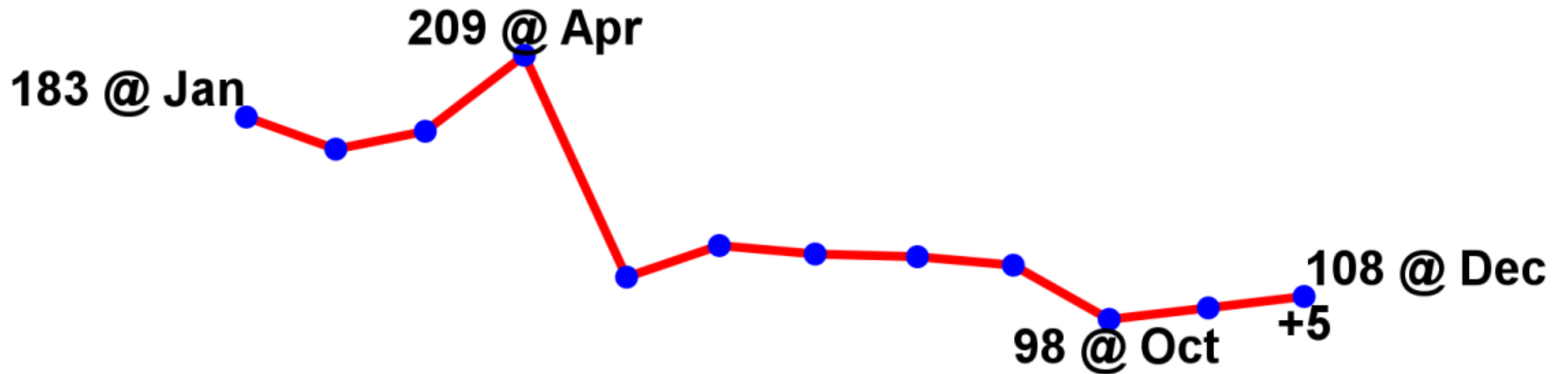
The simplest, but importantly informative, time series plot that you can possibly make.

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















If the Y values had commas,
@ would be a better Y-X connector

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



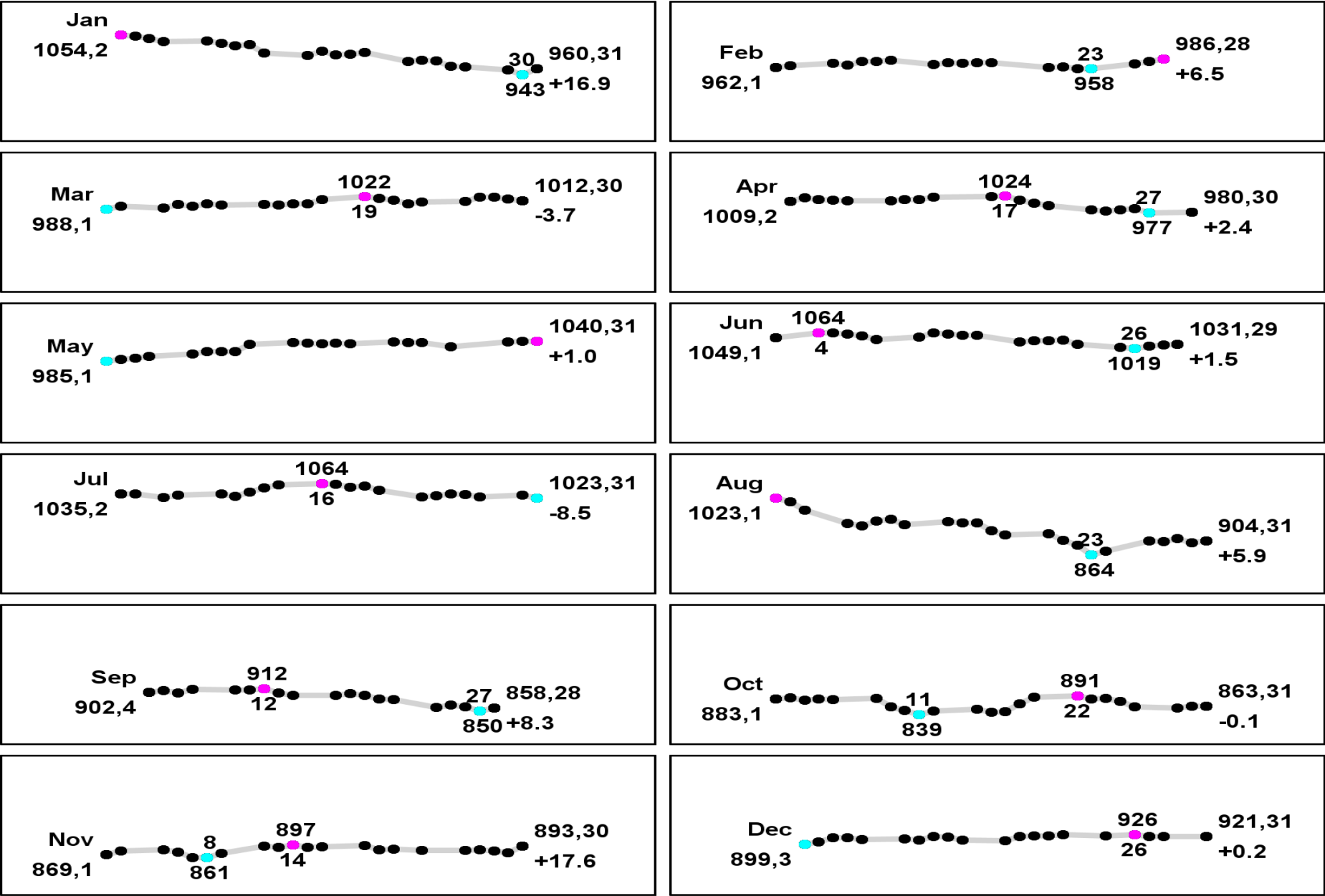
To Avoid Confusion:

Years After My Sparse Line Annotation,
Came Peter Zelchenko's **Spark Line** (with no annotation):
It came in a column of plots in a table
which includes values for the plots.

Symbol	Bid	Ask	Last	Change	T	Chart	Volume	High	Low	Value Change		Value	Gain	
DELL	89 3/4	89 13/16	89 3/4	+ 1 1/4	↑		10,310,100	90 1/8	88 1/2	+1.41%	250	17,950	+273.72%	13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16			25,628,700	51 1/4	1/4	-1.65%	-81	4,844	+60.79%	1,831
SDTI	26 1/4	26 3/8	26 3/8	+ 1/2	↓		504,600	27 3/8	25 5/8	+1.93%	250	13,188	+133.15%	7,531
COMS	46 1/2	46 9/16	46 9/16	- 25/32	↓		3,191,100	47 15/16	45 3/4	-1.65%	-102	6,053	+29.79%	1,389
LU	111 5/8	111 11/16	111 9/16	+ 1 9/16			5,104,600	112 5/8	110	+1.42%	78	5,578	+22.76%	1,034
YHOO	368 1/16	368 1/2	368 1/2	+ 17 1/4	↓		3,787,800	381 3/16	280	+4.91%	431	9,213	-0.41%	-38
AOL	162 13/16	163	163	+ 8			10,008,500	164	158 1/2	+5.16%	280	5,705	+73.06%	2,408
CMGI	97 3/8	97 1/2	97 1/2	+ 5 7/8	↓		1,323,800	98 1/2	93	+6.41%	705	11,700	+186.76%	7,620
SPLN	33 13/16	33 15/16	33 13/16	+ 7/16	↓		300,200	34 3/4	33 5/8	+1.31%	88	6,763	+94.60%	3,288
BEAS	13 1/2	13 5/8	13 5/8	- 7/16	↓		389,200	14 1/4	13 1/8	-3.11%	-44	1,363	-9.17%	-138
GNET	102	103 3/16	101 5/16	+ 6 1/8	↑		307,600	108	97	+6.43%	613	10,131	+130.26%	5,731
RNWK	67	67 1/4	67	+ 2 3/4	↓		1,233,900	69	64 15/16	+4.28%	275	6,700	+79.87%	2,975
MSFT	173 1/8	173 1/4	173 5/16	+ 1 3/4	↓		13,284,500	174 7/16	170	+1.02%	175	17,331	+54.74%	6,131
INTC	133 3/4	133 13/16	133 13/16	- 3 1/8	↓		8,094,300	137 1/2	133 3/8	-2.28%	-625	26,763	+65.20%	10,563
TOTAL					↑			205,302	80,993	+1.63%	2,293	143,280	+79.41%	63,377

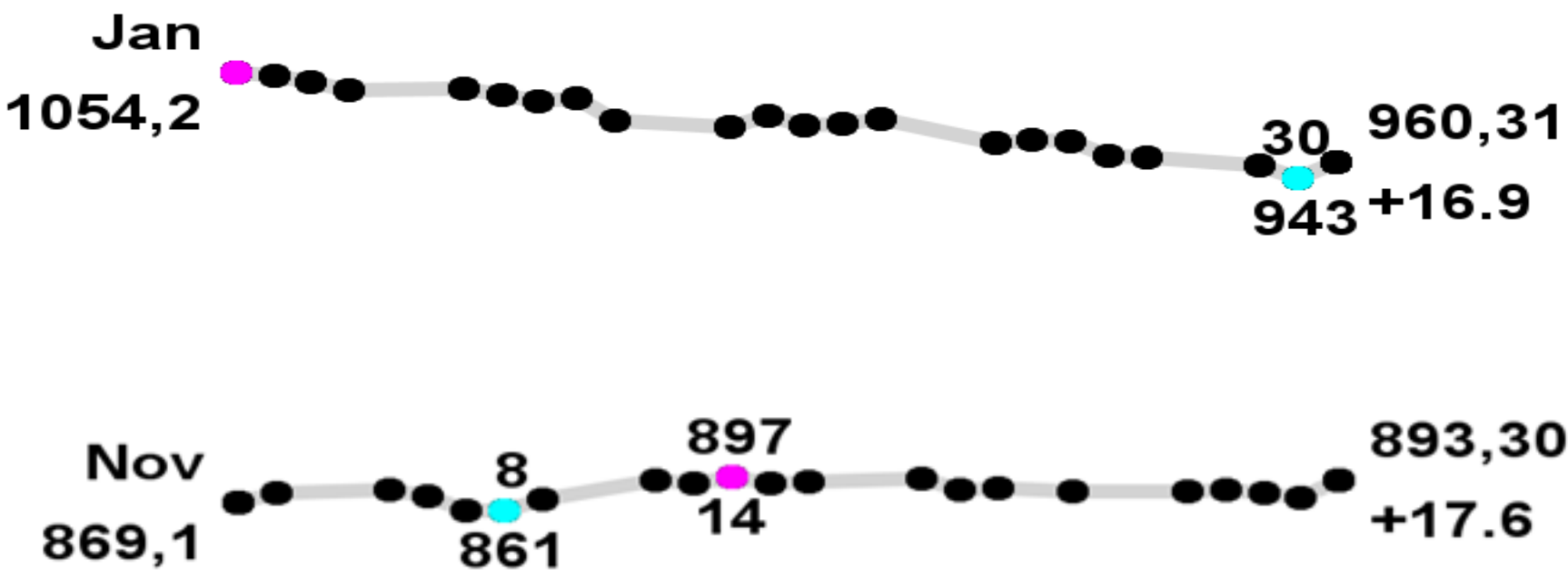
Sparse Line Array for Twelve Months of Trading Days

Color coding the
minimum and
maximum ALWAYS
identifies the
critical point, even
when it occurs at a
Start or End point,
rather than at an
intermediate point.



Sparse Line Array
A Closer Look
at Two Months

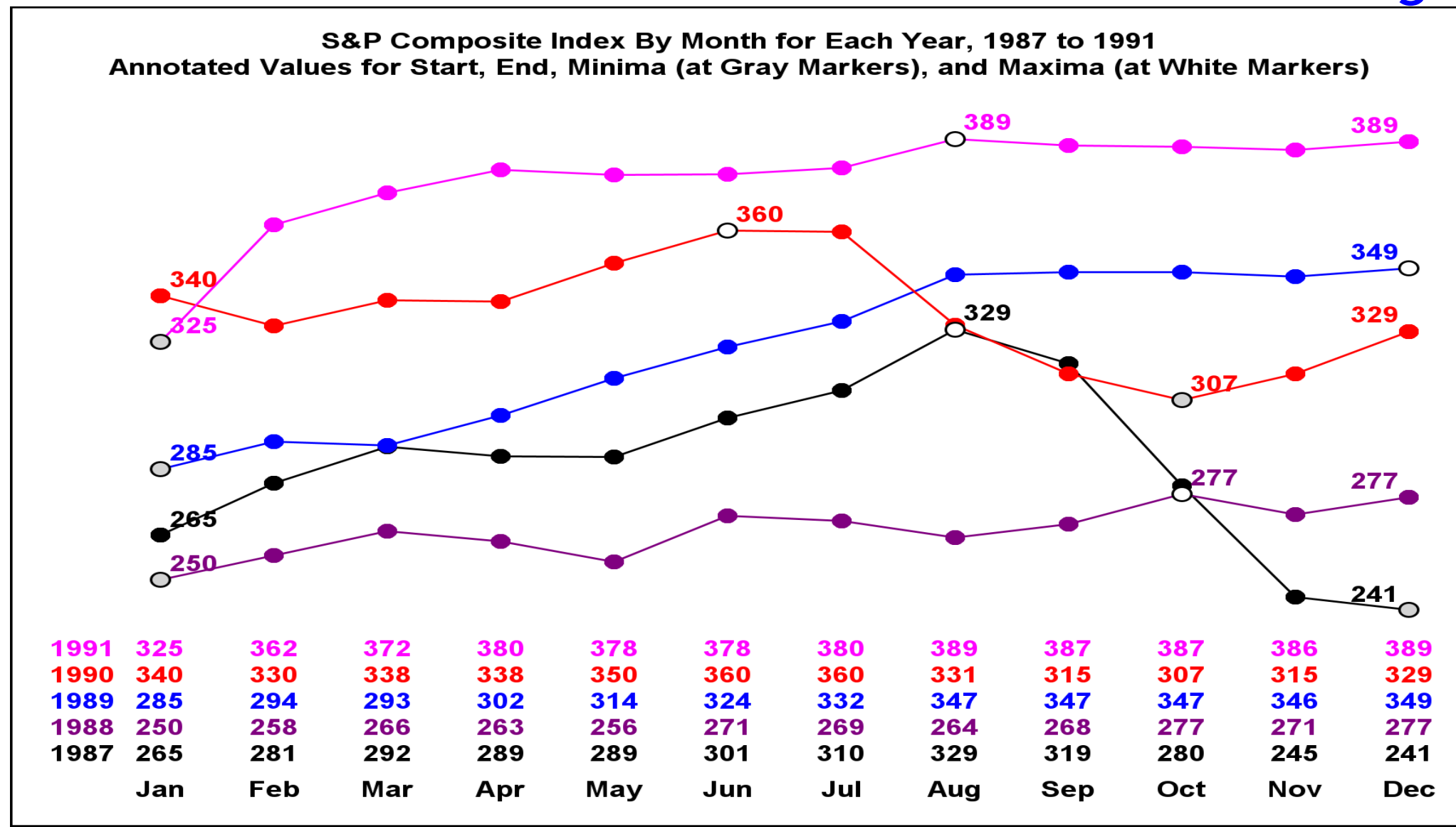
Color coding the
minimum and
maximum ALWAYS
identifies the
critical point, even
when it occurs at a
Start or End point,
rather than at an
intermediate point.



Sparse Annotation + X Axis Table = Everything!

Each minimum & maximum highlighted: The Whole Story is Told.

With Nov & Dec Data: No Need To Annotate the Last Change

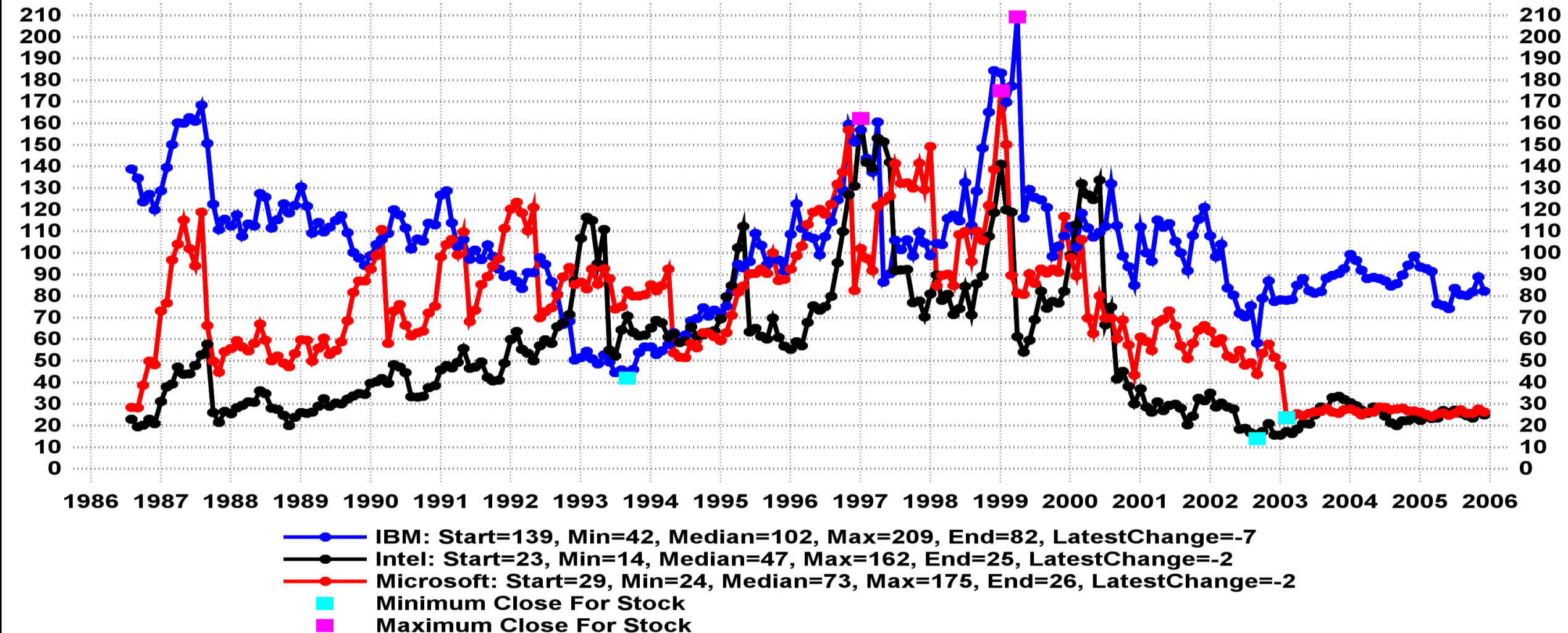


What to do when NO annotation is possible?
Provide a Maximally Informative Legend

You Can Make a Legend Do **Anything** That You Wish

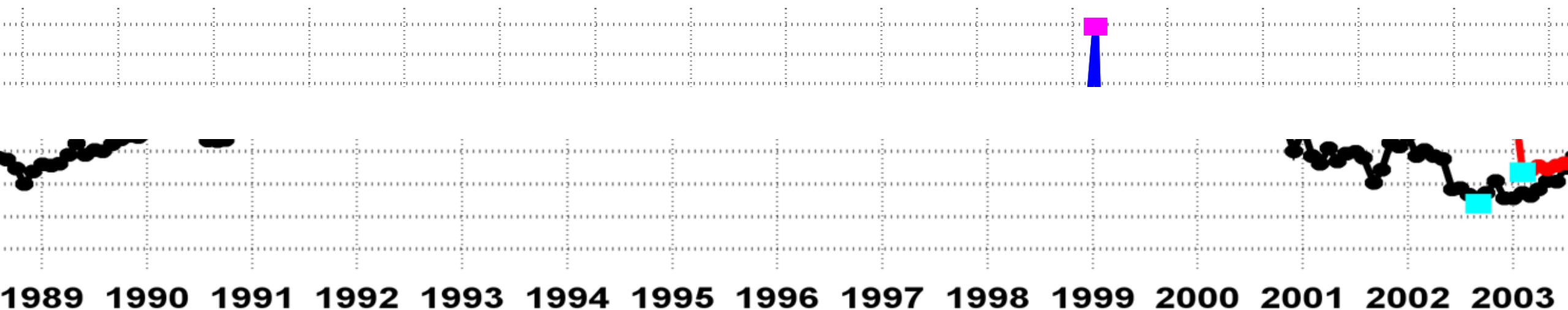
It need not be limited to providing only line identity.

1986 to 2005 Stock Close Each Month - Range = 14 to 209, Median = 77
Each Stock's Maximum Marked with **Magenta Square**, Minimum Marked with **Turquoise Square**



A Closer Look at Titles and at The Maximally Informative Legend:

1986 to 2005 Stock Close Each Month - Range = 14 to 209, Median = 77
Each Stock's Maximum Marked with Magenta Square, Minimum Marked with Turquoise Square



- IBM: Start=139, Min=42, Median=102, Max=209, End=82, LatestChange=-7
- Intel: Start=23, Min=14, Median=47, Max=162, End=25, LatestChange=-2
- Microsoft: Start=29, Min=24, Median=73, Max=175, End=26, LatestChange=-2
- Minimum Close For Stock
- Maximum Close For Stock

If annotation & X axis table are infeasible, some options:

- web page plot with mouseover text ("data tips") and table, web linked, or below on same web page (Use ODS HTML5) See Chapter 14 of my book
- plot with table in same or separate Excel worksheet See my *Powerful SAS Output Delivery with ODS EXCEL*
<https://www.lexjansen.com/mwsug/2019/SP/MWSUG-2019-SP-072.pdf>
- graph, table, text on same page using ODS LAYOUT for some ODS destinations

Look for my presentation on ODS LAYOUT
(and ODS PDF and ODS HTML5 and ODS EXCEL)
in the forthcoming proceedings for MWSUG 2025

If annotation and X axis table are infeasible,
other options are:
companion table on the same or a separate page
using ODS PDF, ODS Word, ODS PowerPoint, etc.

FINALLY, after all of those **PICTURES**, let's "switch gears" and look/re-look at, and apply, some of my principles for communication-effective text and color.

For All Charts, Plots, and Graphs

- If axis identity is in the title or subtitle, omit the axis label to maximize your drawing space
- Dates need no axis label, unless special dates
- When software wrongly concludes that it cannot fit all of the axis values UNdo "thinning" with FITPOLICY=NONE on YAXIS/XAXIS statement
- Keep text horizontal—it's how we read!
 - UNDO the common default phenomenon of vertical text for the Y axis label
 - Put axis labels in the graph title/subtitle
 - Or, for the Y axis, use YAXIS LABELLOC=TOP
 - If the software tilts your X axis values to make them fit, on the XAXIS statement instead try FITPOLICY=STAGGER or VALUEATTRS=(SIZE=Best Smaller Size)

Assure Text Readability

- Colored text must be thick for distinguishability
- Thin Color: **Arial** | **Calibri** | **Calibri Light** | **Courier New**
- Maximize contrast of text and background color
- Black and White = maximum readability
- Black and Yellow seize visual attention

Assure Text Readability

➤ Readability Obstructions I have encountered:

Yellow on White, Black On Blue

Absurd: Gray on White for emails, or web pages

Do you like this better? If so, WHY?

Do you like this better? If so, WHY?

Do you like this better? If so, WHY?

Do you like this better? If so, WHY?

Do you like this better? If so, WHY?

The only thing sillier than Gray on White
is White on White

Use Color with a Purpose

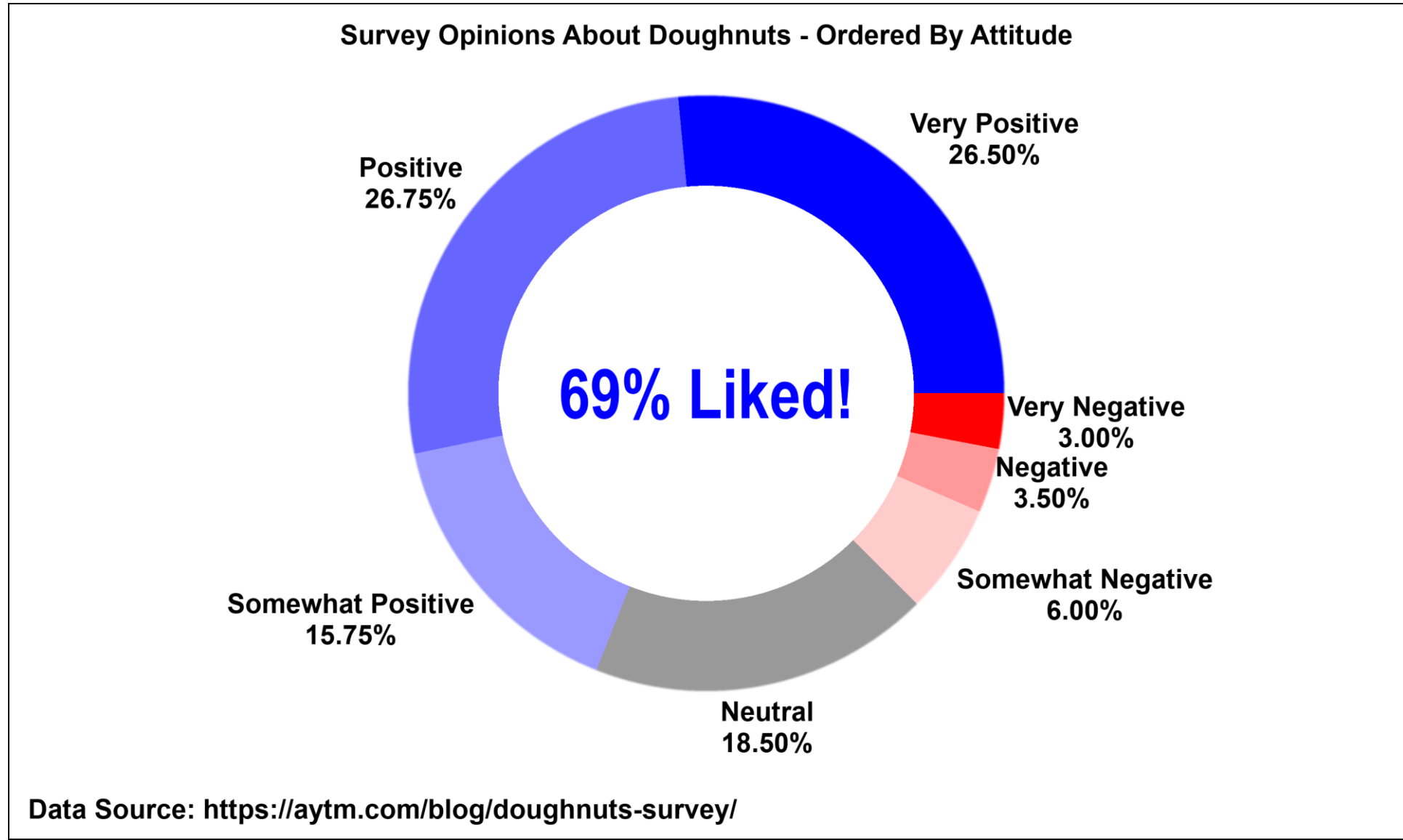
- Use it to communicate, not decorate
- When color is used, viewer ASSUMES a purpose
- If NONE, you MYSTIFY or CONFUSE your viewer

Assure **Color** Distinguishability

- Inability to distinguish Red & Green is the most common color blindness, as I have been explaining since 1995
- **Do not use so-called “Traffic Lighting”**
i.e., Red versus Green to signal Bad versus Good

My FIRST Donut Chart (uses REAL data . . . about donuts)

These Shades of Blue and Red are Safe & Effective



Prevent color communication failures

If you use color, make it distinguishable.

Provide sufficient color mass with:

- Lines & Text thick enough
- Plot Markers big enough
- Legend Color Swatches big enough

If you use color, make it **usable** for the viewer

- Abstain from continuous color gradient legends

Bigger Legend Color Swatches Are Better

Use rectangles—they are bigger than squares

- Get rectangles with **FILLASPECT=GOLDEN**
- If squares, make bigger with **AUTOITEMSIZE**

Above are **KEYLEGEND** statement options

- **FILLASPECT IS NOT** available for PROC SGMAP
- **AUTOITEMSIZE IS** available for PROC SGMAP

PERSISTENT, POPULAR, POOR color decision

A Color Gradient Legend is Under-Communicative

- **IMPOSSIBLE** to match any color in the graph with the precise same color in the legend
- **IMPOSSIBLE** to label every legend color
- **IMPOSSIBLE** to determine exactly how much two graph colors differ
- **DIFFICULT** to be certain whether two parts of the graph are the same color

Typical Color Gradient Map (unhelpful legend omitted)



My Ranges have **a Rationale** and a **Usable** Legend

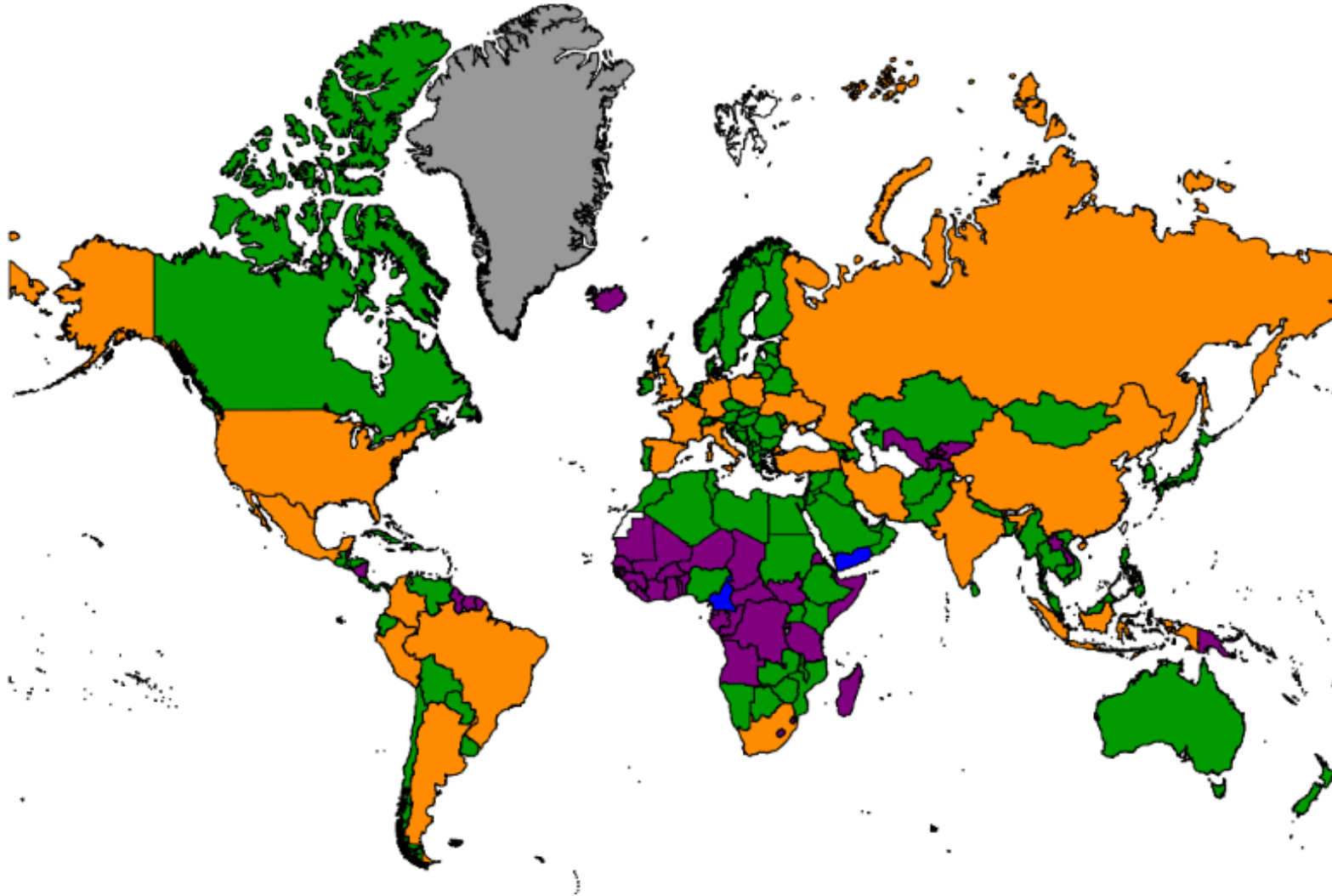
Map Of Total Deaths In World With Alphabetic Order FootNote List Of Countries

Grand Total = 6,994,144 for 226 Countries in Ranges: ■ 1 - 37 (20 lowest) ■ 40 - 1,971 (below median)

■ 1,974 - 2,159 (median) ■ 2,250 - 74,694 (above median) ■ 101,419 - 1,144,877 (20 highest) □ No Data

[Bar Chart](#) [Population](#) [Total Cases](#) [Case Rate](#) [Mortality Rate](#) [Same Map With Range Footnotes](#) [Index for Population Maps](#)

Use the mouse for more information. For very small countries, find the rank in the footnote, and click that Rank button in the map.



My Ranges have a Rationale and a Usable Legend

Map Of Total Deaths In World With Alphabetic Order FootNote List Of Countries

Grand Total = 6,994,144 for 226 Countries in Ranges: ■ ■ 1 - 37 (20 lowest) ■ ■ 40 - 1,971 (below median)

■ ■ 1,974 - 2,159 (median) ■ ■ 2,250 - 74,694 (above median) ■ ■ 101,419 - 1,144,877 (20 highest) □ No Data

My Ranges have a Rationale and a Usable Legend

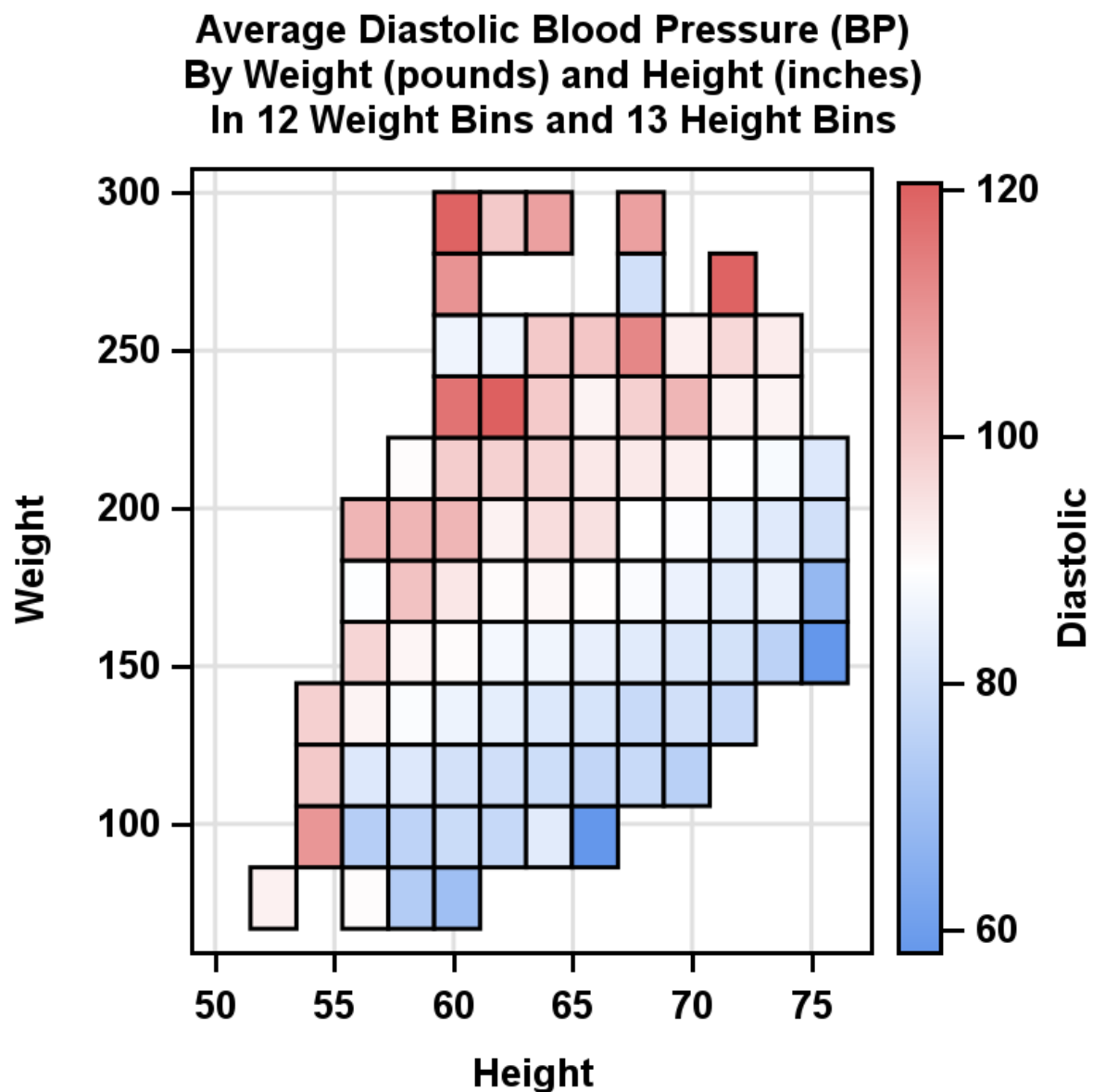
Map Of Total Deaths In World With Alphabetic Order FootNote List Of Countries

Grand Total = 6,994,144 for 226 Countries in Ranges: ■ ■ 1 - 37 (20 lowest) ■ ■ 40 - 1,971 (below median)

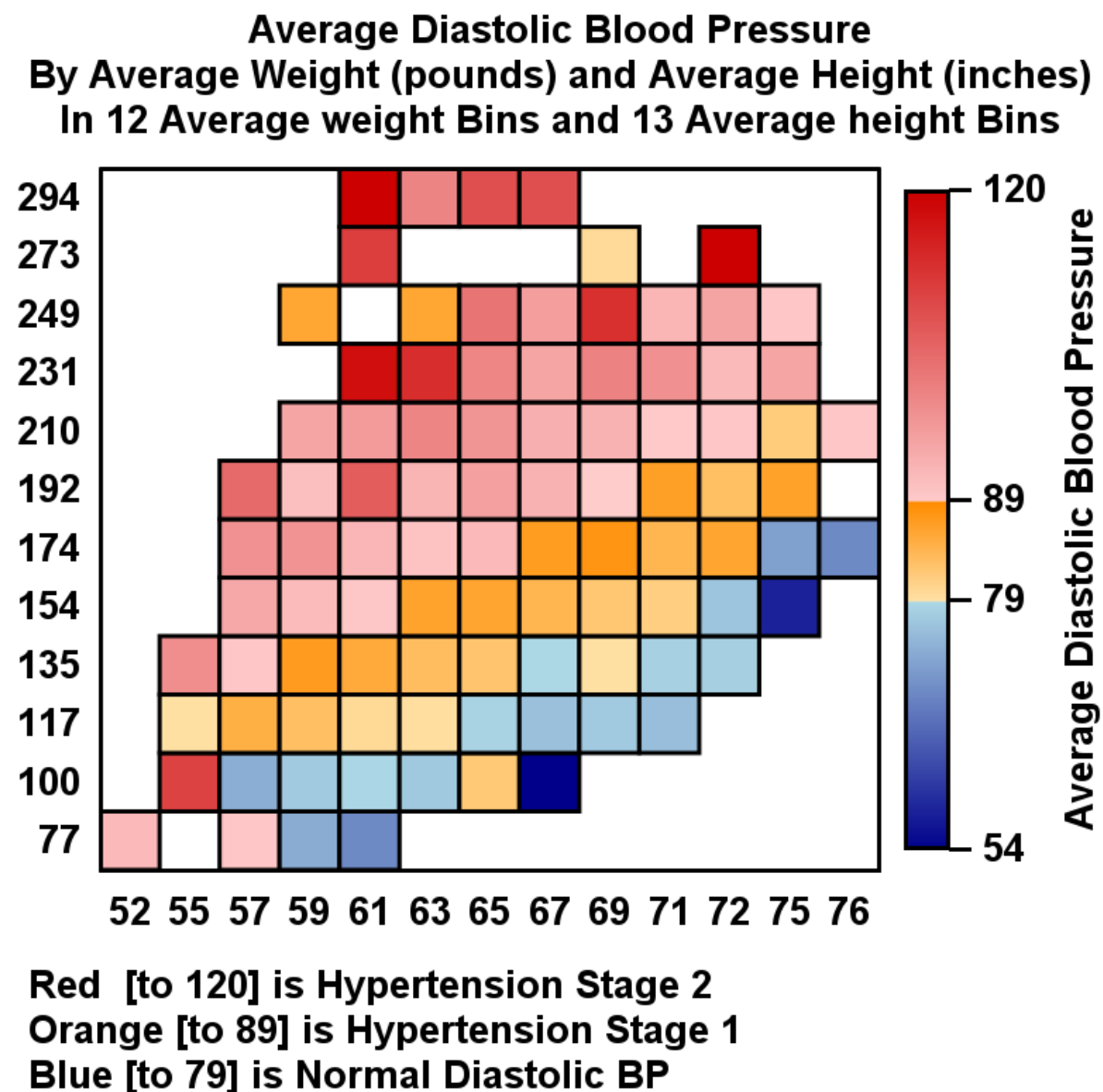
■ ■ 1,974 - 2,159 (median) ■ ■ 2,250 - 74,694 (above median) ■ ■ 101,419 - 1,144,877 (20 highest) □ No Data

IF PROC SGMAP supported FILLASPECT=GOLDEN, like other ODS Graphics Procedures and like SAS/GRAPH PROC GMAP, I would have used those rectangles, instead of the double squares used here to supply more color mass.

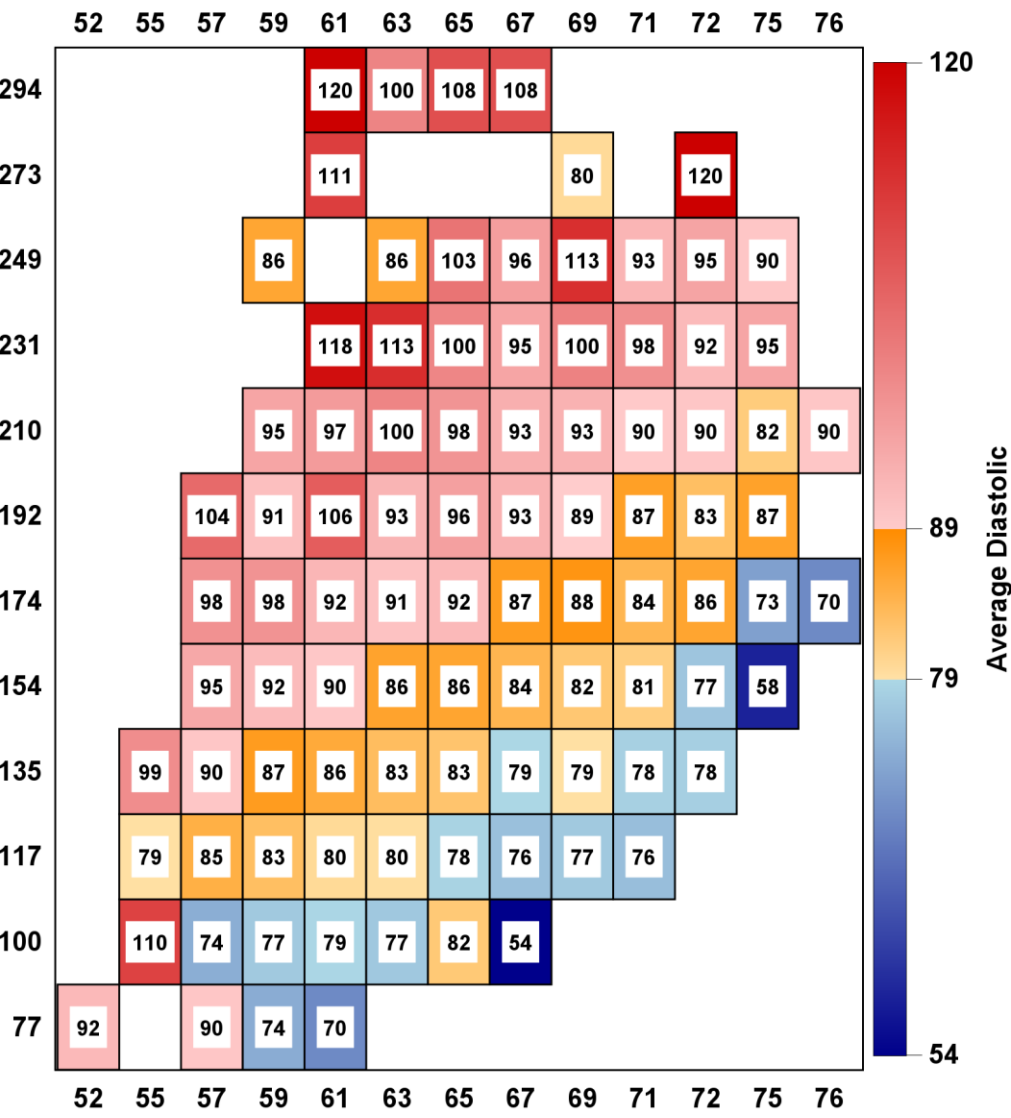
Default Gradient Legend



Significance-Based SubGradients But Annotate for USABILITY ►



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
All bins are equal width, but axis values are averages, not bin midpoints.
So the increment between axis values along an axis can vary.

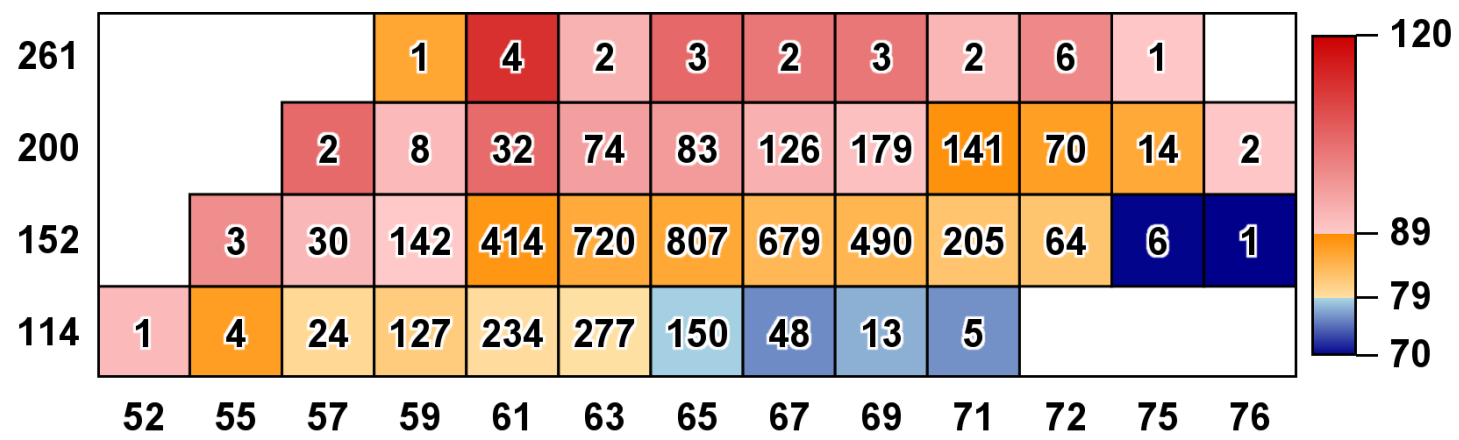


Blue [to 79] is Normal Diastolic BP & Orange [to 89] is Hypertension Stage 1
Red [to 120] is Hypertension Stage 2 & Purple [> 120] is Hypertensive Crisis

Actually Usable Significance-Based SubGradients Are Aided by Annotation

To improve readability, annotation option at left, is BACKFILL (OUTLINE is not used) at bottom, is BACKLIGHT

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
Frequency Counts in Weight-Height Cells and 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.



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.

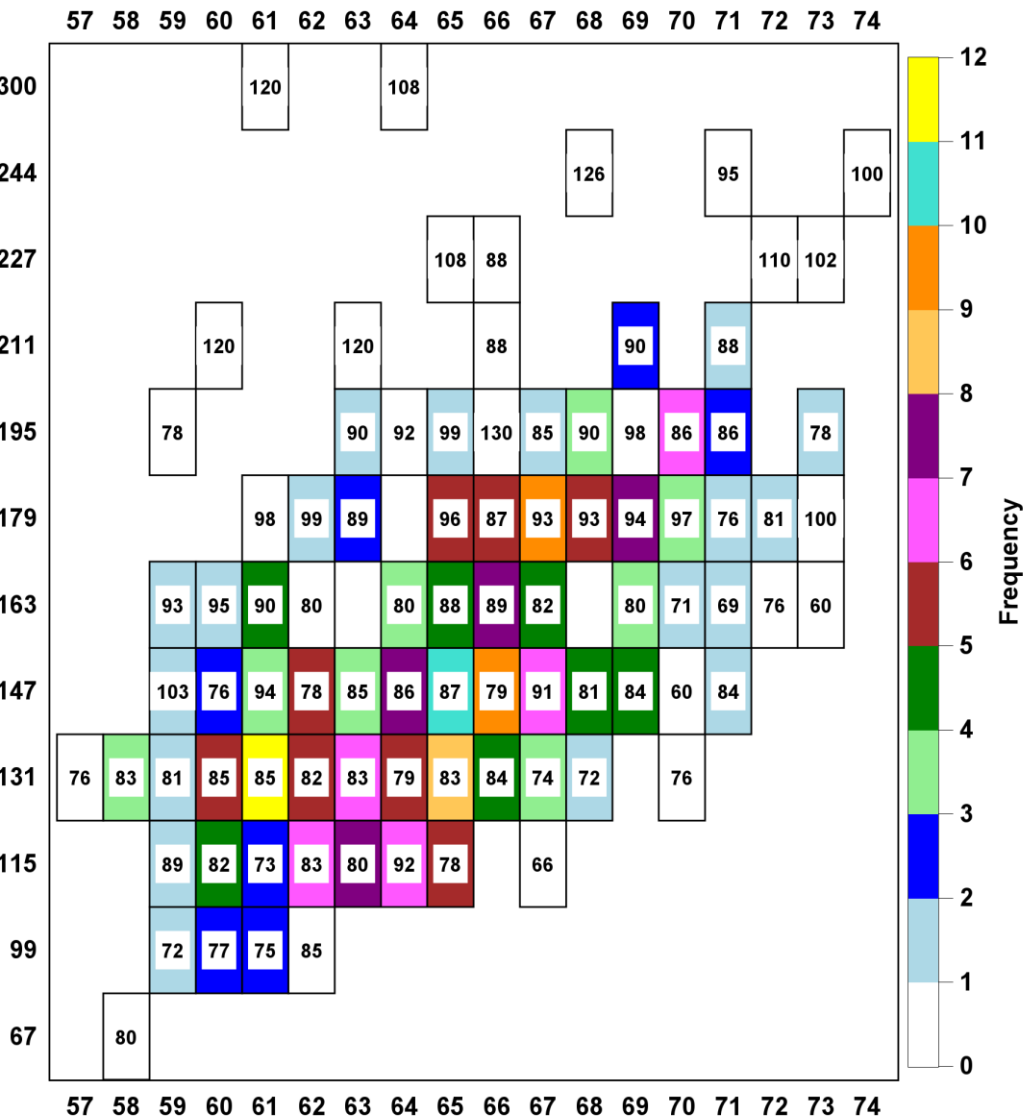
▲ Dynamic Data-Specific Titles

▼ 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

Average Diastolic Blood Pressure
By Weight in pounds & Height in inches
For 318 Observations in 12 weight Bins, 18 height Bins, & 89 Cells
Range of Avg Diastolic is 60-130 with Mean 87.5 & Standard Deviation 13.1
BP in Weight-Height Cells and Color Gradient for Frequency Counts
Each bin is for only one rounded value of height or weight
Each cell is only one height-weight, but may be for multiple data points.



Under 80 is Normal BP and 80-89 is Hypertension Stage 1
90-120 is Hypertension Stage 2 and Over 120 is Hypertensive Crisis

A Heat Map with a Discrete Color Legend is not only possible, but **USEFUL**

If you can, use data preprocessing to create macro variables whose values are dynamically supplied to titles and/or footnotes to make your graph self-explanatory and maximally informing.

**Brevity enables quicker reading,
but presenting more text
in pursuit of communication
IS Safe and Effective.**

Let's review some of the principles

Make Data Quickly Easily Completely Understood

- Deliver **image** + **precise numbers**
(for **quick easy** inference + **reliable accurate inference**)
- Assure text readability
- Assure color distinguishability
- Provide a headline for your graph if possible
(TELL the meaning/conclusion)

Categorical Data

- If bar chart, use horizontal (to fit category, rank, value, and percent)
- Show Them What's Important
 - Use Ranking
 - Physically/Visually order the bars, slices, etc.
 - Provide Rank number for each category
 - Consider Subsetting
 - Top N
 - Cutoff / Filter (for goal to achieve, or threshold for concern)
 - Enough of Top Ranked for X Percent of Grand Total
 - If Subsetting, also provide an All Categories Chart (Include Rank)
- Clarify the context for the response values
 - Provide percent of grand total for each category
 - List the grand total in the chart title or subtitle

Time Series Data

- Annotate y and x at each point or display Y value above X value with X axis table
- OR use a web-enabled plot with mouseover text (aka "data tips") for on-point information AND, for non-vanishing inspection also provide a companion table (see the options and resources suggested in prior slides)
- OR use a static plot-table composite created with ODS LAYOUT
- For simplicity, use my Sparse Line or my Array of Multiple Sparse Lines
- Facilitate seasonality detection for time series with overlay plots
- For overlay plots create a custom maximally informative legend
- On a single-line plot, unless negative Y values, start the Y axis at zero to prevent magnifying visual effect of not really significant increases or decreases in Y value
- To fit more X axis tick mark values, use FITPOLICY=STAGGER on XAXIS statement
- When software mistakenly concludes that not all X or Y axis tick mark values can fit, use FITPOLICY=NONE on the XAXIS or YAXIS statement

Univariate Data

- See my book for designs for which there was no showing time today

Consider These Widely Applicable Plots for When You Have a Category Variable and a Group Variable

- Ultimate Stacked Vertical Bar Chart
- Ultimate Clustered Horizontal Bar Chart
- Needle Plot Plus Block Plots

In my book *Visual Data Insights Using SAS® ODS Graphics: A Guide to Communication-Effective Data Visualization*, see Chapters 1 and 2 for the comprehensive lists of principles for graphic design and use of color. (Much more numerous than those shown today.)

(Some post January 2023 principles were included here today. I cannot promise to abstain from adding more. Watch for my further work. Since 1981, I'm still on my quest to get the best data graphics.)

The book includes 327 illustrations for examples and code to create them. It's meant for people with no prior ODS Graphics experience.

Most principles presented there and here are actually software-independent. The book covers every kind of chart or plot that ODS graphics can create except vector plots.

Addenda / Tips

Prevent Unexpected Effects

For Some Procedures (SGPLOT, some SGPANEL charts, some SGSCATTER charts),

Default Title Placement is NOT middle of image width, but above the middle of the graph drawing area.

(Y axis tick mark values usually push the drawing area to the right.)

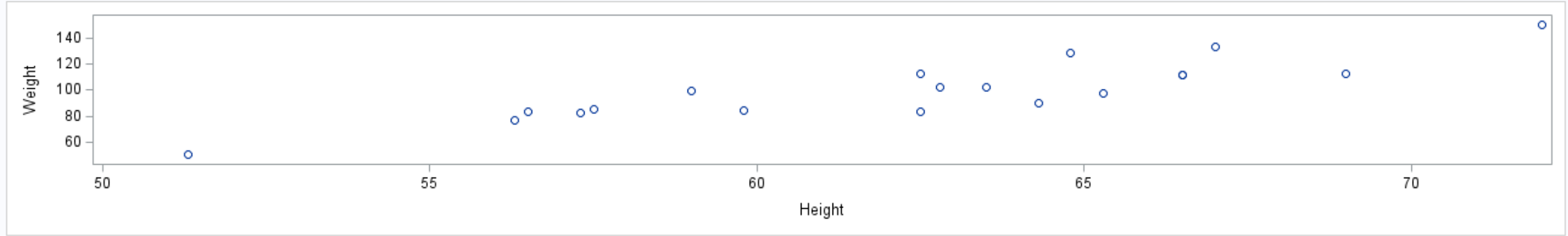
IF you want titles at the top center of the image width, get that with `TITLE JUSTIFY=CENTER . . . ;`

Do Not Assume Black is the default text color

- It often is, but not ALWAYS (e.g., next slide)
- I have a macro to assure black on white
- It allows choice of font (Family) and Weight
- Macro defaults: Arial and Bold
- Parent style is HTMLblue
 - You could change that
 - Or make Parent a macro invocation option

Default HTMLblue versus Custom ODS Style

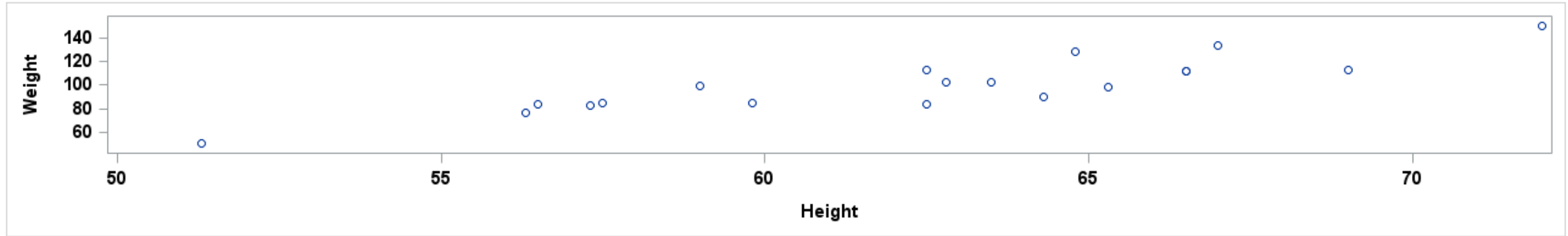
NOGTITLE TITLE1 for: NOGTITLE NOGFOOTNOTE Test of text and background colors using default ODS Style HTMLblue



NOGFOOTNOTE FOOTNOTE1 for: NOGTITLE NOGFOOTNOTE Test of text and background colors using default ODS Style HTMLblue

ODS TEXT for: NOGTITLE NOGFOOTNOTE Test of text and background colors using default ODS Style HTMLblue **Note the faint blue no-value-add background!**

NOGTITLE TITLE1 for: NOGTITLE NOGFOOTNOTE Test of text and background colors using Custom ODS Style WhiteWebPageAndAllBlackOnWhiteForTextArial11ptBold



NOGFOOTNOTE FOOTNOTE1 for: NOGTITLE NOGFOOTNOTE Test of text and background colors using Custom ODS Style WhiteWebPageAndAllBlackOnWhiteForTextArial11ptBold

ODS TEXT for: NOGTITLE NOGFOOTNOTE Test of text and background colors using Custom ODS Style WhiteWebPageAndAllBlackOnWhiteForTextArial11ptBold

You Can Make Data Quickly Easily Completely Understood
Send Questions, Requests for Code or Macro,
Suggestions, Comments, etc. to:

LeRoy Bessler PhD
Bessler Consulting and Research
Visual Data Insights™
Strong Smart Systems™
BesslerLeRoy2024@gmail.com
Valparaiso, Indiana, USA