Bessler's Principles of Communication-Effective Data Visualization

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

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 advisor to SAS[®] users on best practices for graphic design and use of color, and author of Visual Data Insights Using SAS ODS Graphics: A Guide to Communication-Effective Data Visualization. His principles explained and demonstrated in the book are useful for any graphics software.

LeRoy is a data analyst and SAS programmer, who has supported SAS servers, SAS software, SAS analytic data, and hundreds of SAS users.

"If you can't do it with SAS software, you probably don't really need to do it."—LeRB

Scope of Presentation

- Focus on 4+ principles from my approx. 80 principles See the handout for the FULL list
- Presenting examples Only/Mainly the Uncommon Ones from the book's 320+ examples PLUS New or Enhanced
- Code for most of them is in the book Other code on-request via email
- Questions As We Go Through Slides

A Fundamental Requirement for Graphic Communication Effectiveness is →

Visuals for Quick, Easy Inference Numbers for Accurate Inference

- **Visuals Impression**
- Immediate
- Easy
- Memorable ("Sticky")
- Numbers Knowledge
- Clear
- Certain
- Correct ("The Answer")

Getting to the Numbers

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

For Categorical Data:

- For bar charts, provide category,
- response, percent of total, and rank
- all adjacent to each bar

For Time Series Data:

If possible, annotate each plot point with both Y and X value

For Time Series Data:

- If possible, annotate each plot point with both Y and X value
- If not, display Y value immediately above each X value, using an X axis table

For Time Series Data:

- If possible, annotate each plot point with both Y and X value
- If not, display Y value immediately above each X value, using an X axis table
- If not, provide a table either below the plot, or in a linked-to web table or in a linked-to Excel worksheet

For Time Series Data:

- If possible, annotate each plot point with both Y and X value
- If not, display Y value immediately above each X value, using an X axis table
- If not, provide a table either below the plot, or in a linked-to web table or in a linked-to Excel worksheet
- A web graph with data tips is helpful, but they vanish when the mouse moves A static table is better

Show Them What's Important

Show Them What's Important

So They Need To Look No Further

- Show Them What's Important for Categorical Data:
- Order It (By Descending Response) Rank It (Number it 1 to N)
- **N** = Count of Categories

- For Categorical Data Provide:
- Fotal Response in Title/Subtitle
- For Each Category, display:
 - Rank (Order Sequence Number)
 - Category
 - Response
 - Percent of Total

- For Categorical Data,
- if possible, display all four values (Rank*, Category, Response, Percent of Total) adjacent to each other for each bar, pie slice, donut bite, dot, or needle—
- immediate association of all four items
- eliminates visual hunting for the pieces and the need to mentally assemble this package of essential pieces for each category
 *Rank can be unnecessary when categories are few enough

Y Axis Start Principle for Bar Charts Start the Y axis of bar charts at zero to prevent misleading comparison

- Show Them What's Important For Time Series Data: with Sparse Line Annotation whenever it suffices—which may be often
- or as a prelude or companion
- to the complete picture

- Show Them What's Important with Sparse Line Annotation Annotate Y value and Date OR DateTime OR Time (whichever X variable is appropriate) FOR
- Start,
- End,
- any intermediate minimum/maximum and
- Annotate the Latest Change of Y

The Sparse Line



Sparse Line Panel for Twelve Months Of Trading **Months**





Dow Jones Composite Index by Day of Month in Nov. 1990 No Trading on Saturday, Sunday, and Holidays

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



- **Show Them What's Important**
- with Sparse Line Annotation
- Annotate Y value and Date OR DateTime OR Time)
- FOR any point where slope permanently significantly changes
- Slow increase to Rapid increase
- Rapid increase to Slow increase
- Slow decrease to Rapid decrease
- Rapid decrease to Slow decrease and
- Annotate the Latest Change of Y

- **Y Axis Start Principle for Time Series Plots**
- Start the Y axis of time series plots at zero
- to prevent needless elation or anxiety
- about changes that may not really be
- of great significance
- but there are exceptions
- (a) My Sparse Line Annotation, aka Sparse
- Line, NOT to be confused with the
- underinformative so-called "spark line"
- (b) a multi-line overlay time series plot if it
- becomes too collision-plagued by starting the
- Y axis at zero

Make Text Readable

Usually Assumed, But Delivered?

Make Text Readable

- The Most Readable Graph Text Is
- Big
- Bold
- Black (on light backgrounds)
- White (on dark backgrounds)
- Sans Serif (Serif is suitable for long narrative, novels, essays, bios, . . .)
 11pt Arial Bold is my preference

Make Text Readable For me, The Most Readable Graph Text Is **Uniform Throughout the Graph** > 11pt Arial Bold is my preference > If you are a SAS ODS Graphics user, ask me for my AllGraphTextSetUp macro AllTextSetUp macro if you want a shortcut to uniformity, using YOUR preferred font, size (height), and weight (Bold vs Normal)

Make Text Readable

- **Use Non-Disruptive Non-Distracting**
- **Backgrounds for Graph and Its Text**
- Solid Color Only (Best is White)
- No texture
- No gradient
- No image or photo

Make Text Readable When Text Is Over Color, Maximize Contrast of Text Color and Background Color

Assure color distinguishability/usability Very frequent design error: misuse of color

Problems are demonstrated later

Assure **Color** Distinguishability/Usability

- Inability to distinguish Red and Green is the commonest color blindness
- Thick Enough Text
- Thick Enough Lines
- Big Enough Plot Markers
- Big Enough Legend Color Swatches
 If for area fill (bar/pie/donut chart, heat map, infogeographic, or bubble plot, use FILLASPECT=GOLDEN
- Discrete Legend Colors NOT a Continuous Color Gradient Legend (you will see and understand Why later)

Assure **Color** Distinguishability

- Inability to distinguish Red and Green is the commonest color blindness
- When color coding became a popular tool, there was rush to and a persistent fad of

"Traffic Lighting"—

Red versus Green to signal

Bad versus Good

4+ out of nearly 80 Principles

- Visuals PLUS Precise Numbers
- Show Them What's Important
- Assure Text Readability
- Assure Color Distinguishability/Usability

Shades of Blue and Red are Safe & Effective


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

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



- 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 America \$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%

Pie Chart for Sales by Product converted to Donut Chart

Space for the hole label for Total Sales Label Size and Format are controlled only by software, NO coding options.

Ranked Shoe Sales and Percent Share By Product



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

Collision-Free Data Labels Outside By Consolidating Two Smallest Bites



Donut Chart Converted to Pie Chart Total in title suffix via dynamic macro variable





Arial 13pt **Bold** Labels Fit Inside. Retaining merger of two smallest slices.

3D Pie Charts Are Always Misleading



"The Extremes of Other" My SAS/GRAPH Parents of My ODS Graphics Pac-Man Pie Chart



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

> SAS Software Dominates 1993 Mainframe Data Analysis Market



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

> Most of Your Property Tax is NOT for Local Government



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



Anti-Communicative 3D Gimmick

Sufficient & Simple 2D Solution

Shoe Sales By Region Shoe Sales By Region 3D Complexity - Viewer Must Guess Sales 2D Simplicity - Viewer KNOWS Sales Africa Total Sales Asia Middle East \$5,631,779 Canada \$5,503,986 United States Central America \$4,873,000 Western Europe Eastern Europe \$4,255,712 Middle East Canada Pacific \$3,657,753 Central America South America \$2,434,783 South America United States \$2,394,940 Eastern Europe Western Europe \$2,342,588 Africa \$2,296,794 5 n 2 3 4 6 Pacific \$460,231 Asia Total Sales in \$M

Horizontal Bar Charts Can Fit More Information Than Vertical Bar Charts

Show Them What's Important with Ranking/Order

Upper: ALWAYS Create This Ranked Chart Lower: Supplement for Easy Alphabetic LookUp When There Are Many Bars

Region Shoe Ranked By Sales and Percent Share - Grand Total \$33,851,566



Alphabetic Order Region with Shoe Sales, Percent, and Rank - Grand Total \$33,851,566

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

Calas

Chara Dank



Show Them What's Important With a Subset

- A Subset Can Be Enough
- Also can serve as a useful companion
- to a complete bar chart, especially when All Bars are numerous

Subset with The Top N (Using N=10)

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



Subset with a Minimum Filter

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



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

пк		Sales	Snare
1 Vanco	uver	\$3,227,768	9.5%
2 Tel	Aviv	\$2,567,568	7.6%
3 King	ston	\$2,235,204	6.6%
4 D)ubai	\$1,910,544	5.6%
5 Chie	cago	\$1,565,585	4.6%
6 New	York	\$1,489,207	4.4%
7 Al-Kh	obar	\$1,153,667	3.4%
8 Minnea	polis	\$1,099,937	3.2%
9 Heidel	berg	\$967,739	2.9%
10 Lis	sbon	\$898,345	2.7%
11 M	anila	\$854,904	2.5%
12 Car	acas	\$789,323	2.3%
13 Wai	rsaw	\$786,714	2.3%
14 Lor	ndon	\$762,009	2.3%
15 C	Cairo	\$738,198	2.2%
14 Lo 15 (ndon Cairo	ndon \$762,009 Cairo \$738,198

Enough. For OTHER data, it can happen that 80%, 90%, or more of the total comes from a shorter list. See the next slide for an example with real data.

COVID-19 Top 30 Countries By Total Cases & By Total Deaths

~	3	Top 30 CC	OVID-19 Cas	e and De	a × +						_	
~	\rightarrow	C	<u>ن</u>	File C:	/WUSS%202	2024/!%20COVID-	19/Output/	BarChartsAndTable/PNG_Cour	ntsRank	edSubsetH 🕁	Ď	
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						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					
An 202 Countries had a Total Of 773,450,120 Case				Casaa	% of World	Case	All 220 Countries had a T	Bank	Dootho	% of World	Mortality	
.	nited	States of	f Amorica	капк	Cases	Cases	Rate	United States of America	капк			Rate
	meu	States	China	2	99,322,52	12.8408	6.9657	Brazil	2	702,116	10.0386	1.8713
			India	3	45,010,89	5.8192	3.1761	India	3	533,346	7.6256	1.1849
			France	4	38,997,49	0 5.0418	57.5074	Russia	4	400,967	5.7329	1.6984
			Germany	5 6	38,437,75		46.1051	Mexico	с 6	334,958	4.7891	4.3485
		So	uth Korea	7	34.571.87	3 4.4696	66.7207	Peru	7	221.583	3.1681	4.8842
			Japan	8	33,803,57	4.3703	27.2716	Italy	8	193,886	2.7721	0.7312
			Italy	9	26,517,70	6 📕 3.4283	44.9167	Germany	9	174,979 📕	2.5018	0.4552
		United	Kingdom	10	24,843,59	3.2119	36.8004	France	10	167,985	2.4018	0.4308
			Russia	11	23,608,72	3 3.0522	16.3141	Indonesia	11	161,965	2.3157	2.3745
			Snain	12	13 980 34	0 1 8074	29 3960	Colombia	12	140,757	2.0903	2 2352
			Australia	14	11.710.21	6 1.5139	44.7341	Argentina	14	130.687	1.8685	1.2971
			Viet Nam	15	11,624,00	0 1.5028	11.8387	China	15	121,889	1.7427	0.1227
			Argentina	16	10,075,24	2 1.3026	22.1384	Spain	16	121,852 📕	1.7422	0.8716
		Ne	therlands	17	8,627,69	3 1.1154	49.1214	Poland	17	120,018	1.7160	1.8160
			Mexico	18	7,702,80	0.9959	6.0412	Ukraine	18		1.5716	1.9911
			Iran Indonesia	19	6 820 02	0.9859	0.0114 2.4758	South Africa	19	102,595	1.4669	2.5191
			Poland	20	6,609.09	2 0.8545	16.5820	Japan	20	74.694	1.0680	0.2210
			Colombia	22	6,385.53	9 0.8255	12.3097	Romania	22	68,601	0.9808	1.9559
			Austria	23	6,081,28	0.7862	68.0263	Philippines	23	66,795	0.9550	1.6004
			Portugal	24	5,637,11	4 0.7288	54.8846	Chile	24	64,482	0.9219	1.2100
			Greece	25	5,533,96	0.7155	53.2882	Canada	25	53,595	0.7663	1.1281
			Ukraine	26	5,520,48	0.7137	13.9049	Hungary	26	48,917	0.6994	2.1997
			Chile	27	5,329,10	0.6890	27.1841	Czechia	27	43,225	0.6180	0.9139
			Relatium	28	5,206,72		15.3418	Viet Nam Bulgaria	28	43,206	0.51//	0.3/1/
			Israel	29	4,841,93	8 0.6250	51.2388	Greece	29	38,224	0.5465	0.6907
			101401		4,041,00	0.0200	0112000			00,224	0.0400	0.0007

COVID-19 2020 to 2023 Top 30 Countries Suffered 85% of the World's Cases & Deaths

Three Best Ranked Vertical Bar Charts

Stacked Vertical Bar Chart (Pie Chart Alternative)



Rank-Ordered Table with a Visual Companion

The Fully Informative Vertical Bar Chart



Ranked Vertical Bar Chart Without Bars



Categorical Data with a Group Variable



Segment Labels identify region values for each product. Each bar has a Data Label that displays the Product Total. The legend identifies the **Region for each** segment, and is a Region Totals table. This is the best possible Vbar **Chart for Grouped Categorical Data**, but NO Percents.

The Best Bar Chart for (unranked) Categorical Data with a Group Variable

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



Percent of Whole for Every Part Legend serves as a table but is not needed as a bar identifier

 Women's
 Pacific \$399,441 4.769%

 Dress
 Middle East \$1,112,207 13.280%

 \$1,964,190
 Asia \$78,234 0.934%

 23.453%
 Africa \$374,308 4.469%

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

Time Series Plots

Annotated Y values. Use light dotted line to deemphasize any line-label collisions.

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



Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

0

Jan

X Axis Table. Y values AT the X values. Most useful on a denser plot.



Move Y and X values in a stack to be AT the plot point.

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



Time Series Trend Line With No Line A line or markers would add NO value. Middle of Y & X positions is AT the point.

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

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



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

If my Principle to Start Y Axis at Zero is disregarded here, the gentle before-peak and after-peak changes would meaninglessly greatly expand from gentle to severe.





Show Them What Is Important with Start, End, Min and/or Max or Significant Persistent Change Point



LeRB's Sparse Line Annotation in 1992 (Zelchenko's "Spark Line" appeared in 1998)
The Sparse Line



Sparse Line Panel for Twelve Months Of Trading **Months**





Dow Jones Composite Index by Day of Month in Nov. 1990 No Trading on Saturday, Sunday, and Holidays

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



Sparkline invented by Peter Zelchenko Screenshot of SparkLines used in the Medved QuoteTracker, 1998 https://en.wikipedia.org/wiki/Sparkline The SparkLine is a valuable visual accessory/element inside a table, but has no standalone informative value.

Symbol	Bid	Ask	Last	Change	Т	Chart	Volume	High	Low	Value C	hange	Value	Ga	in
DELL	89 3/4	89 13/16	89 3/4	+11/4	î	~	10,310,100	90 1/8	88 1/2	+1.41%	250	17,950	+273.729	13,147
CPQ	48 7/16	48 9/16	48 7/16	- 13/16		V-	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	î	Charles and the second	504,600	27 3/8	25 5/8	+1.93%	250	13,188	+133.159	7,531
COMS	46 1/2	46 9/16	46 9/16	- 25/32	î	The second and the se	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	+19/16		مسيد ميدينين	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	î	and the second s	3,787,800	381 3/16	280	+4.91%	431	9,213	-0.41%	-38
AOL	162 13/16	163	163	+ 8		and the second second	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	î	and a superior	1,323,800	98 1/2	93	+6.41%	705	11,700	+186.769	7,620
SPLN	33 13/16	33 15/16	33 13/16	+ 7/16	î	a and the house of the second s	300,200	34 3/4	33 5/8	+1.31%	88	6,763	+94.60%	3,288
BEAS	131/2	13 5/8	13 5/8	- 7/16	î	The second second	389,200	14 1/4	131/8	-3.11%	-44	1,363	-9.17%	-138
GNET	102	103 3/16	101 5/16	+61/8	Û	CI INTEL CONTRACTOR	307,600	108	97	+6.43%	613	10,131	+130.269	5,731
RNWK	67	67 1/4	67	+ 2 3/4	î	And and a state of the state of	1,233,900	69	64 15/16	+4.28%	275	6,700	+79.87%	2,975
MSFT	1731 <i>1</i> 8	1731/4	173 5/16	+ 1 3/4	î	and the second	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	Ŧ	Marine	8,094,300	137 1/2	133 3/8	-2.28%	-625	26,763	+65.20%	10,563
TOTAL					Û	and a second		205,302	80,993	+1.63%	2,293	143,280	+79.41%	63,377

Left is a narrowed Spark Plot. Right is Sparse Line Table created with ODS HTML5 and ODS LAYOUT GRIDDED. GRIDDED LAYOUT can be web-enabled and easier code than ABSOLUTE LAYOUT, but provides limited layout controls.



Monthly First Day Close Price 2004-2005 for Three Stocks Maximum Color Is Magenta, Minimum Is Turquoise Index and Month Values for First, Last, Min, Max Last Annotation Includes Latest Change







Data Source: SASHELP.STOCKS

Sparse Line Table at the right was created with ODS PRINTER and ODS LAYOUT ABSOLUTE.



When Collision-Free Annotation Is Infeasible

Use Maximally Informative Title Use Maximally Informative Legend Start, End, Median; Min & Max Values and Dates



- Microsoft: Start=26, Min=24 on FEB2003, Median=73, Max=175 on JAN1999, End=29
- Minimum Close For Stock
- Maximum Close For Stock

Rationale-Based Histograms

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



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, but Last Bin is leftovers unless count of observations / count of quantiles has no remainder

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



Run-Time Self-Explaining Distributions

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'



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



Show Them Where It's At with a Ranked COVID-19 InfoGeographic (Partially Annotated)

MortalityRate for 226 Countries in World With Top and Bottom 10 Annotated

0.0185 - 0.0808 (10 lowest)

0.0887 - 0.8508 (below median)

0.8658 - 0.8716 (median)

0.8749 - 3.3792 (above median) See MouseOver Text for More Information 3.4622 - 18.0745 (10 highest) 🛛 🔲 No Data

Go To Bar Chart Go To Map of CaseRate By Country



Top 10 MortalityRate Countries (highest to lowest): 1:Yemen | 2:Sudan | 3:Syria | 4:Somalia | 5:Peru | 6:Egypt | 7:Mexico | 8:Bosnia and Herzegovina | 9:Liberia | 10:Afghanistan Bottom 10 MortalityRate Countries (lowest to highest): 226:Nauru | 225:Burundi | 224:Cook Islands | 223:Bhutan | 222:Tuvalu | 221:Brunei | 220:Saint Pierre and Miquelon | 219:Singapore | 218:Tonga | 217:Faroe Islands No MortalityRate Data for: Falkland Islands | Niue | North Korea | Pitcairn | Saint Helena, Ascension Island, & Tristan da Cunha | Taiwan | Tokelau | Turkmenistan | Vatican City State | Western Sahara

The Top of Its Linked-to Bar Chart

226 Countries in World Go Back To The Map

			Total	Total			
	Rank	Population	Cases	Deaths	CaseRate	MortalityRate	
Yemen	1	33,696,612	11,945	2,159	0.0354	18.0745	
Sudan	2	46,874,200	63,993	5,046	0.1365	7.8852	
Syria	3	22,125,242	57,423	3,163	0.2595	5.5082	
Somalia	4	17,597,508	27,334	1,361	0.1553	4.9791	
Peru	5	34,049,588	4,536,733	221,583	13.3239	4.8842	
Egypt	6	110,990,096	516,023	24,830	0.4649	4.8118	
Mexico	7	127,504,120	7,702,809	334,958	6.0412	4.3485	
Bosnia and Herzegovina	8	3,233,530	403,468	16,378	12.4776	4.0593	
Liberia	9	5,302,690	8,090	294	0.1526	3.6341	
Afghanistan	10	41,128,772	230,290	7,973	0.5599	3.4622	
Ecuador	11	18,001,002	1,066,409	36,036	5.9242	3.3792	
Niger	12	26,207,982	9,515	315	0.0363	3.3106	
Myanmar	13	54,179,312	641,422	19,494	1.1839	3.0392	
Malawi	14	20,405,318	89,168	2,686	0.4370	3.0123	
Gambia	15	2,705,995	12,626	372	0.4666	2.9463	
Bulgaria	16	6,781,955	1,321,120	38,613	19.4799	2.9227	
North Macedonia	17	2,093,606	350,197	9,963	16.7270	2.8450	
Paraguay	18	6,780,745	736,182	19,945	10.8569	2.7092	
Tunisia	19	12,356,116	1,153,361	29,423	9.3343	2.5511	
Algeria	20	44,903,228	272,010	6,881	0.6058	2.5297	
Chad	21	17,723,312	7,698	194	0.0434	2.5201	
South Africa	22	59,893,884	4,072,636	102,595	6.7998	2.5191	
Sri Lanka	23	21,832,150	672,685	16,888	3.0812	2.5105	
Haiti	24	11,585,003	34,509	860	0.2979	2.4921	
Namibia	25	2,567,024	172,239	4,103	6.7097	2.3822	
Indonesia	26	275,501,344	6,820,926	161,965	2.4758	2.3745	
Honduras	27	10,432,858	472,743	11,114	4.5313	2.3510	_
Jamaica	28	2,827,382	156,584	3,590	5.5381	2.2927	
Trinidad and Tobago	29	1,531,043	191,496	4,390	12.5076	2.2925	
Mali	30	22,593,598	33,164	743	0.1468	2.2404	
Colombia	31	51,874,028	6,385,539	142,727	12.3097	2.2352	_

Making Every Island Discoverable in the Continent Oceania, Which Is Mostly Islands

CaseRate for 24 Countries in Oceania With All Annotated By Rank Of CaseRate 0.4621 - 4.2261 (5 lowest) 7.4263 - 23.2505 (below median) 25.8665 - 25.9638 (median) 27.6122 - 38.7998 (above median) 42.2910 - 48.0020 (5 highest) See MouseOver Text for More Information Go To Bar Chart Go To Map of MortalityRate By Country 10 9 7 12 20 185 22 16 11 17

Top 5 CaseRate Countries (highest to lowest): 1:Niue | 2:New Zealand | 3:Australia | 4:Nauru | 5:Cook Islands Other Countries (highest to lowest Rank): 6:Marshall Islands | 7:Palau | 8:Wallis and Futuna | 9:Guam | 10:Northern Mariana Islands | 11:New Caledonia | 12:Tuvalu | 13:French Polynesia | 14:Federated States of Micronesia | 15:American Samoa | 16:Tonga | 17:Pitcairn | 18:Samoa | 19:Fiji Bottom 5 CaseRate Countries (lowest to highest): 24:Papua New Guinea | 23:Solomon Islands | 22:Vanuatu | 21:Kiribati | 20:Tokelau

Mouse on the World's Least Populous Country (Mainly Bounty Mutiny Descendants)



Bottom 5 CaseRate Countries (lowest to highest): 24:Papua New Guinea | 23:Solomon Islands | 22:Vanuatu | 21:Kiribati | 20:Tokelau

Pitcairn Island No Data for Total Deaths

- Country = Pitcairn
- Rank of CaseRate in Continent = 17 of 24
- Rank of Population in World = 236 of 236
- Population = 47
- Rank of Case Rate in World = 141 of 232
- Cases As Percent of Population = 8.5106
- Rank of Mortality Rate in World =
- Mortality Rate (Deaths As Percent of Cases) =
- Rank of Total Cases in World = 232 of 232
- Total Cases = 4
- Rank of Total Deaths in World =
- Total Deaths =
- Continent = Oceania

- Planned Enhancement for the InfoGeographic for COVID-19 Data and other applications of the design
- Append the Case Rate (or whatever is the response) to each entry in the Top N and Bottom N footnotes
- Depending on what is the Total Number of Countries in the Map (maximum is 226 for the World), consider including the response in the optional footnote for Other Countries. For some maps, a full annotation creates an unusable collection annotation collisions.
 A fully informative optional Other footnote
- reduces map drawing space. Maybe Harmless

Nomenclature

- What I have called an InfoGeographic for the last few decades is usually called a Choropleth Map or a Thematic Statistical Map.
- It is different from what is popularly called an InfoGraphic: a mixture of graphs, text, and sometimes tables. I have always called this combination simply a composite.
- A Composite is needed if there is no other way to provide static precise values for the graphic content. Web graph data tips vanish.
 Typical infographics require the viewer to endure Information Overload.

An InfoGraphic Built with SAS ODS Graphics

ODS LAYOUT ABSOLUTE

used in my book's Chapter 13

can put Anything Anywhere All At Once

WARNING: A Mercifully Uncomplicated, But Functional, Example Follows

- A Composite of Graph and Table
 ALWAYS Visual PLUS Precise Numbers
- Static, No Vanishing Data Tips

This simple 2 BY 2 layout is NOT the Maximum Capability. **ODS LAYOUT** ABSOLUTE **Supports** Anything **Anywhere** All At Once

Student Average Height & Detail, By Gender



Male Avg Height By Age



Female Student Information

Name Age Height 56.5 Alice 13 Barbara 13 65.3 62.8 Carol 14 12 59.8 Jane 62.5 15 Janet 51.3 Joyce 11 14 64.3 Judy Louise 56.3 12 15 66.5 Mary

Male Student Information

Name	Age	Height
Alfred	14	69.0
Henry	14	63.5
James	12	57.3
Jeffrey	13	62.5
John	12	59.0
Philip	16	72.0
Robert	12	64.8
Ronald	15	67.0
Thomas	11	57.5
William	15	66.5

Data Source: SASHELP.CLASS Data Set

ODS LAYOUT GRIDDED

- Usable in more ODS destinations
- Easier to Code
- Much, Much Less Control
- May Be Adequate to Your Needs
- Can be used with ODS HTML5 for Composite Image with Data Tips (mouseover text)

Usable Heat Maps



Default Heat Map Useless Legend

Custom Heat Map Annotated Cells Legend SubRanges

Using a Range Attribute Map enables better color strategy AND useful legend Legend for comparison, not for precise information

Shoe Sales by Region and Product										
Western Europe	\$296,031	\$946,248	\$747,918	\$11,349	\$857,298	\$201,030	\$985,647	\$827,479		- \$2,058,254
United States	\$448,296	\$1,372,527	\$969,271	\$12,039	\$967,927	\$104,403	\$541,536	\$1,087,987		
South America	\$245,675	\$544,950	\$425,669	\$165,925	\$462,651	\$33,061	\$179,227	\$377,625		
Pacific	\$123,575	\$662,368	\$426,191	\$48,424	\$390,740	\$26,169	\$219,886	\$399,441		
Middle East	\$171,282	\$2,058,254	\$839,571	\$35,186	\$662,480	\$4,007	\$748,792	\$1,112,207		
Eastern Europe	Eastern Europe \$306,785		\$335,761	\$3,716	\$509,698	\$91,202	\$209,256	\$362,126		- \$1,000,000
Central America Caribbean	\$190,743	\$756,513	\$404,895	\$378,382	\$883,181	\$26,964	\$399,357	\$617,718		
Canada	\$385,613	\$441,903	\$920,101	\$14,798	\$952,751	\$140,389	\$410,807	\$989,350		
Asia	\$62,708	\$11,754	\$119,366	\$8,208	\$152,032	\$2,092	\$25,837	\$78,234		
Africa	\$119,835	\$562,794	\$318,500	\$190,409	\$337,076	\$22,150	\$417,516	\$374,308		- \$2,092
	Boot	Men's Casual	Men's Dress	Sandal	Slipper	Sport Shoe	Women's Casual	Women's Dress		



Default **Heat Map: Useless for** Precise Numbers - Cannot match area to legend color - Cannot label all colors on the legend

Useful 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.



Ava BP

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

Fewer OBS in bins: for actual values, not averages

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.



90-120 is Hypertension Stage 2 and Over 120 is Hypertensive Crisis

Making Bubble Plots Better

Probably the Oddest Default in SAS Software All of my examples use PROPORTIONAL



PROPORTIONAL Option area for 2 is twice area for 1


Distinguishable color swatches Maximally readable black data labels Brighter colors, Black outline for bubbles



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.



Squeezing More into a Bubble Plot



Needle Plots with Overlaid Needles (Four Pie Chart Alternatives)

Tree Chart

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

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

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



- **Needle Plot with**
- the Appearance of Stacked Needles
- Better than pie chart, stacked vbar chart: Precise Numbers ATTACHED
- to their needles,
- not in a separate legend for lookup

Pie Chart Alternative Lelia McConnell at SAS helped me with this.

Ranked Shoe Sales and Percent Share By Region (Decreasing Upwards)	
Asia - \$460,231 - 1.4%	Total - \$33,851,566 - 100%
Africa - \$2,342,588 - 6.9%	Pacific - \$2,296,794 - 6.8%
South America \$2,424,782, 7,20/	Eastern Europe - \$2,394,940 - 7.1%
500th America - \$2,454,705 - 7.276	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)
             Asia - $460,231 - 1.4% Total - $33,851,566 - 100%
                                   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%
                                        Clipping of Top
```



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

Clipping of Bottom

Ranked Needle Plot with Help Three Numeric Variables Needle Plot + 2 Block Charts + 2 X Axis Tables Five Variables in Two Dimensions



4+ out of nearly 80 Principles

- Visuals PLUS Precise Numbers
- Show Them What's Important
- Assure Text Readability
- Assure Color Distinguishability/Usability

For ALL of the Principles

Find all of them in:

- Chapter 1: Principles of Communication-Effective Graphic Design
- Chapter 2: Principles of Communication-Effective Use of Color
- Principles are demonstrated in those and in other chapters with 320+ examples Find everything in:
- Visual Data Insights Using SAS® ODS Graphics: A Guide to Communication-Effective Data Visualization

Comments, questions, and requests are welcome

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

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