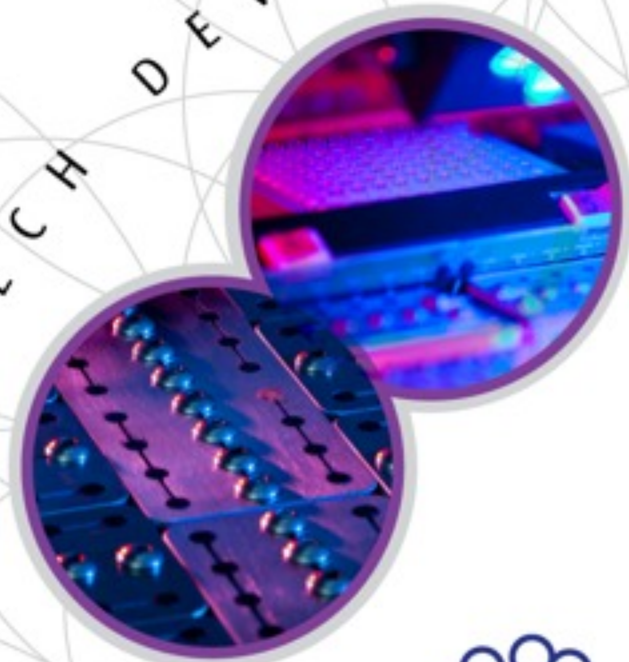
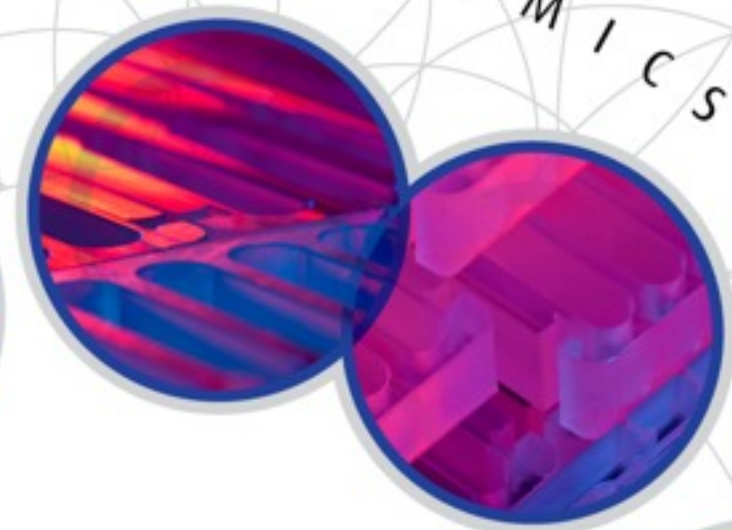
GENOMICS

TECH DEV



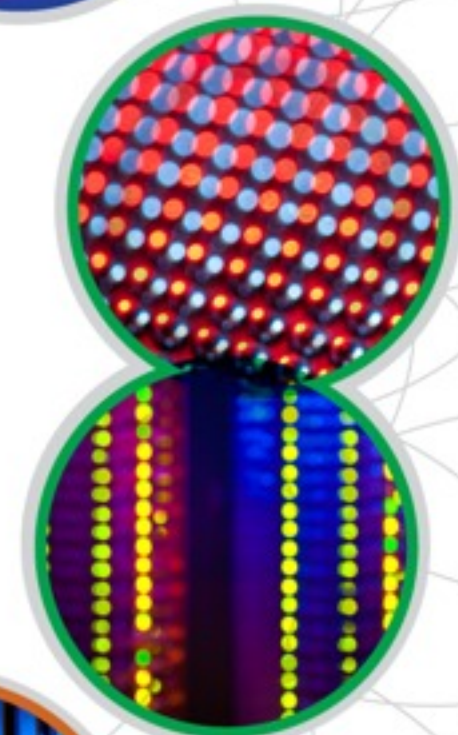


CANADA'S MICHAEL SMITH

GENOME SCIENCES CENTRE

WWW.BCGSC.CA

INFORMATICS



SEQUENCING

COMPUTING

AFFILIATIONS BC Cancer Research Center • BC Cancer Agency • BC Cancer Foundation • Genome BC • Simon Fraser University • University of Britsh Columbia • Genome Sciences Institute

behind every great visualization is a design principle

BCB SEMINAR SERIES

MARTIN KRZYWINSKI
Michael Smith Genome Sciences Center
BC Cancer Research Center
Vancouver, Canada

BIOINFORMATICS + COMPUTATIONAL BIOLOGY
Iowa State University
27 Feb 2012

BC Cancer Agency CARE + RESEARCH
GENOME SCIENCES CENTRE

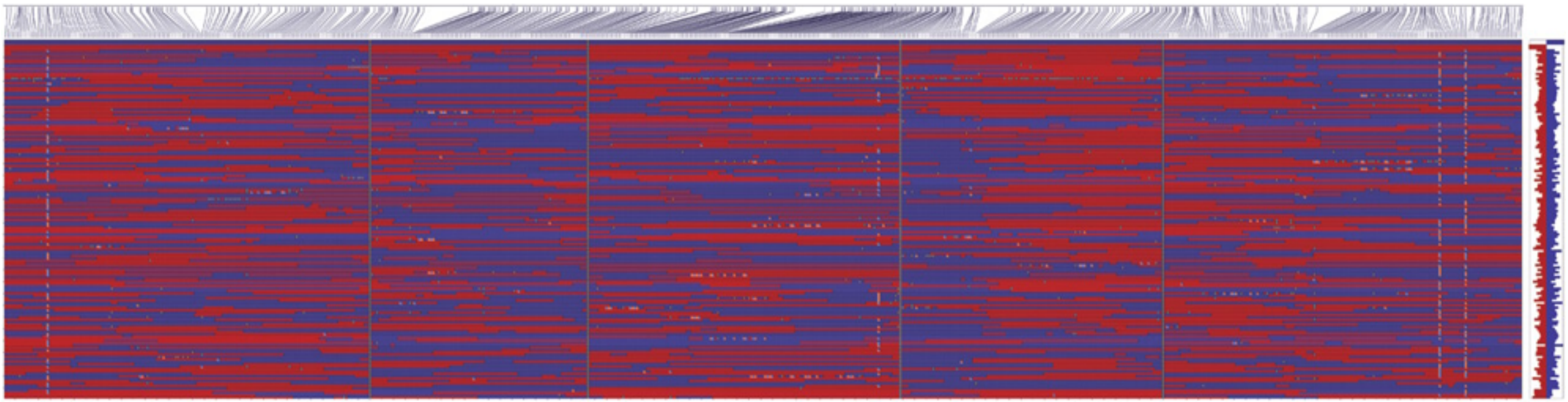
INEFFECTIVE VERBAL COMMUNICATION

individuals with Wernicke's aphasia speak in a 'word salad'

http://www.youtube.com/watch?v=67HMx-TdAZI

they are unaware that their communication is incoherent

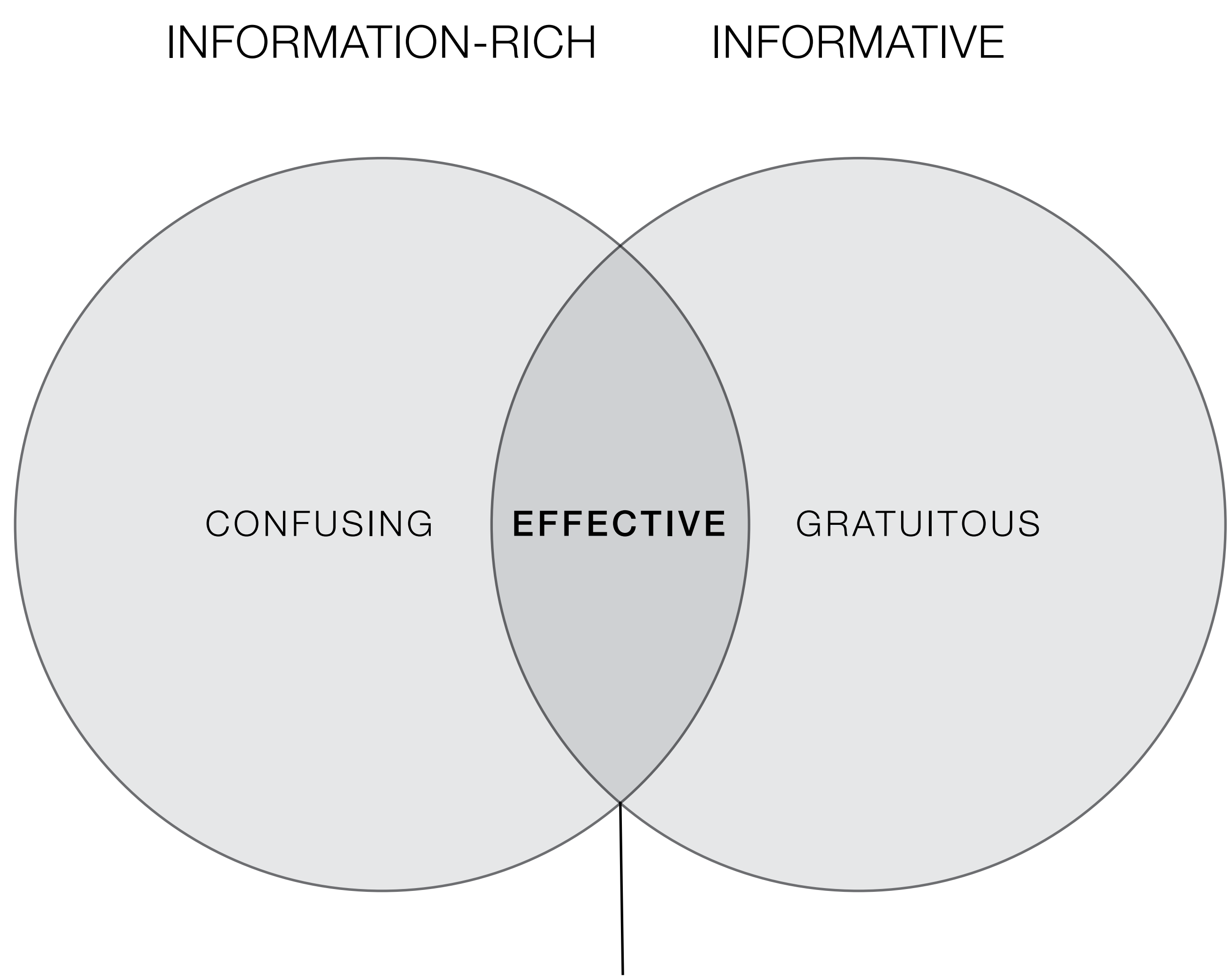
INEFFECTIVE VISUAL COMMUNICATION



West, M.A., et al., High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. Genome Res, 2006. 16(6): p. 787-95.

Is this image distinguishable from one created with random data? Unlikely.
You are looking at a visual ‘word salad’.

WHAT IS AN EFFECTIVE VISUALIZATION?



CLEAR MESSAGE

HIGH DATA-TO-INK RATIO

ACCESSIBLE COMPLEXITY

IS DESIGN SUBJECTIVE?

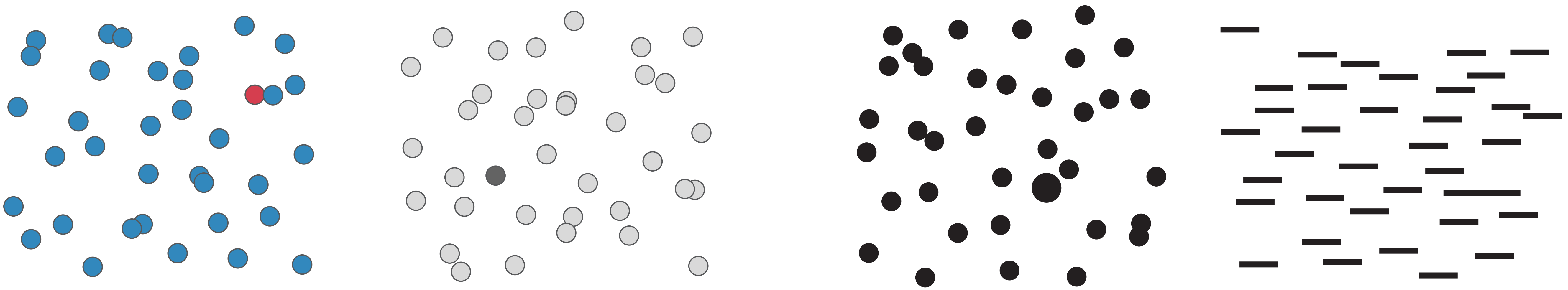
Attractive aspects of proportion, symmetry, color and texture have strong objective foundations.
Few of you would choose the creature on the right as the more attractive.

SALIENCE VS PERTINENCE

SALIENCE

HIGH

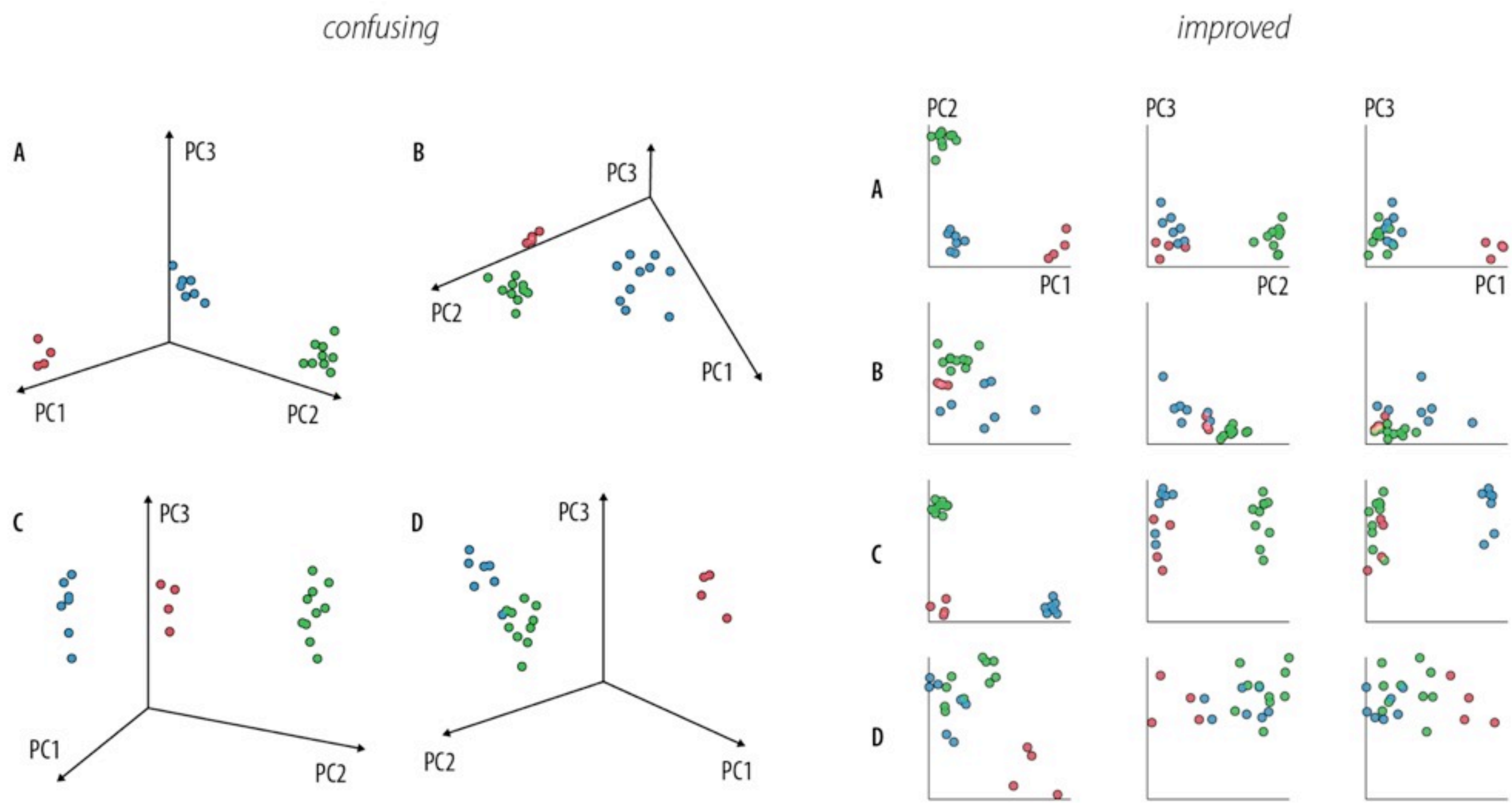
LOW



Fecteau JH, Munoz DP (2006) Salience, relevance, and firing: a priority map for target selection. Trends Cogn Sci 10: 382-390.
Yantis S (2005) How visual salience wins the battle for awareness. Nat Neurosci 8: 975-977.

When presented with a visual stimulus in which a small number of elements is different (color, shape, etc), your brain rapidly remaps the attentional capture of each element. Adjacent similar elements inhibit each other’s salience and your attention is quickly focused on the elements that are different.

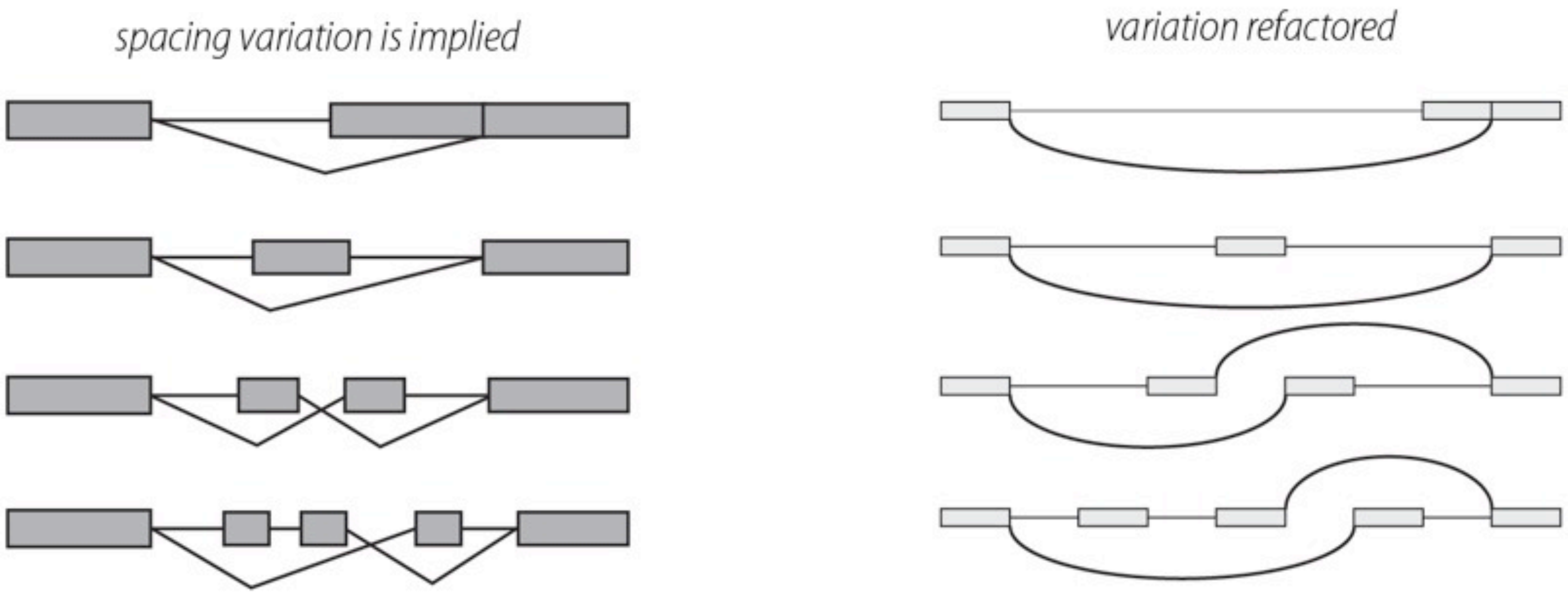
NO 3D



Son CG, Bilke S, Davis S, Greer BT, Wei JS, et al. (2005) Database of mRNA gene expression profiles of multiple human organs. Genome Res 15: 443-450.

3D views suffer from occlusion and imprecision. Without drop lines to each plane, you cannot uniquely determine a point's position.
In the 3D plots (A,C) the axes PC1/PC2 are reversed relative to (B,D). Did you see this immediately?

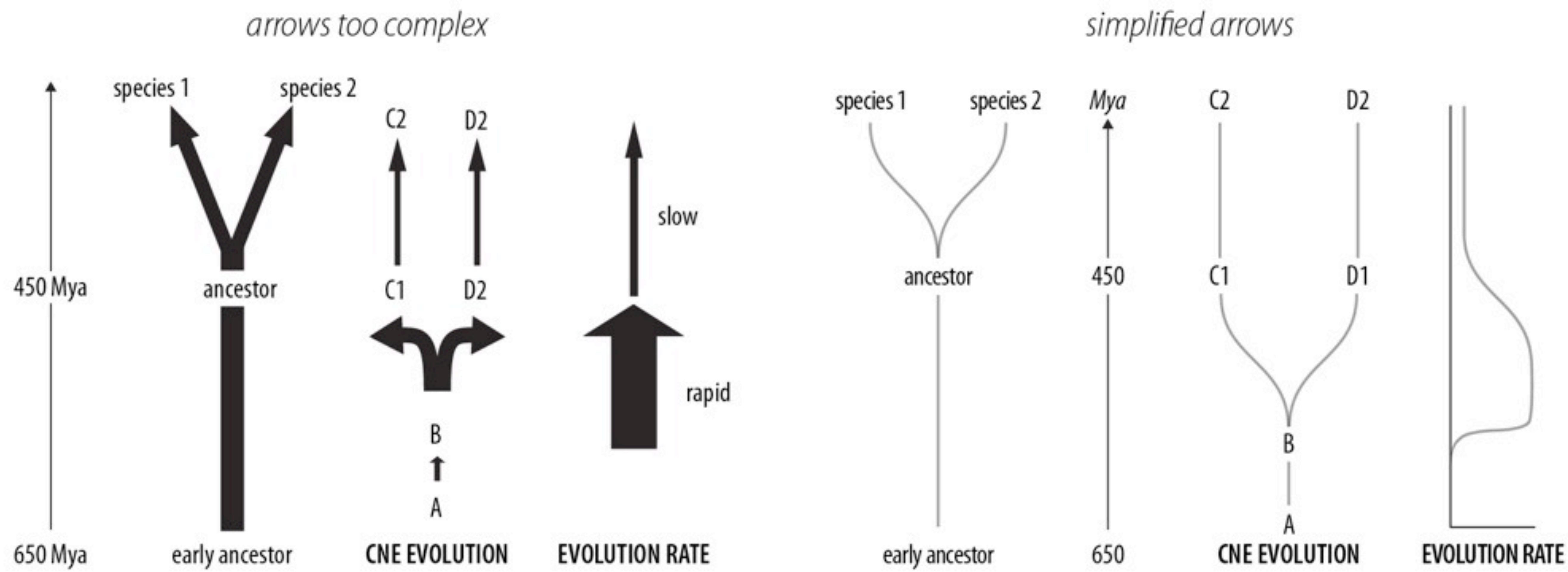
UNIFORM SPACING AND SIZING



Sharov AA, Dudekula DB, Ko MS (2005) Genome-wide assembly and analysis of alternative transcripts in mouse. Genome Res 15: 748-754.

Keep the size, spacing and alignment fixed of as many elements as possible. Any variation in the figure will be interpreted as important to its message. Splicing is an adjacency relationship - the emphasis should therefore be placed on the connections between exons. Recall the previous slide about salience. By limiting variation of non-critical elements, you are emphasizing the visual weight of the core message.

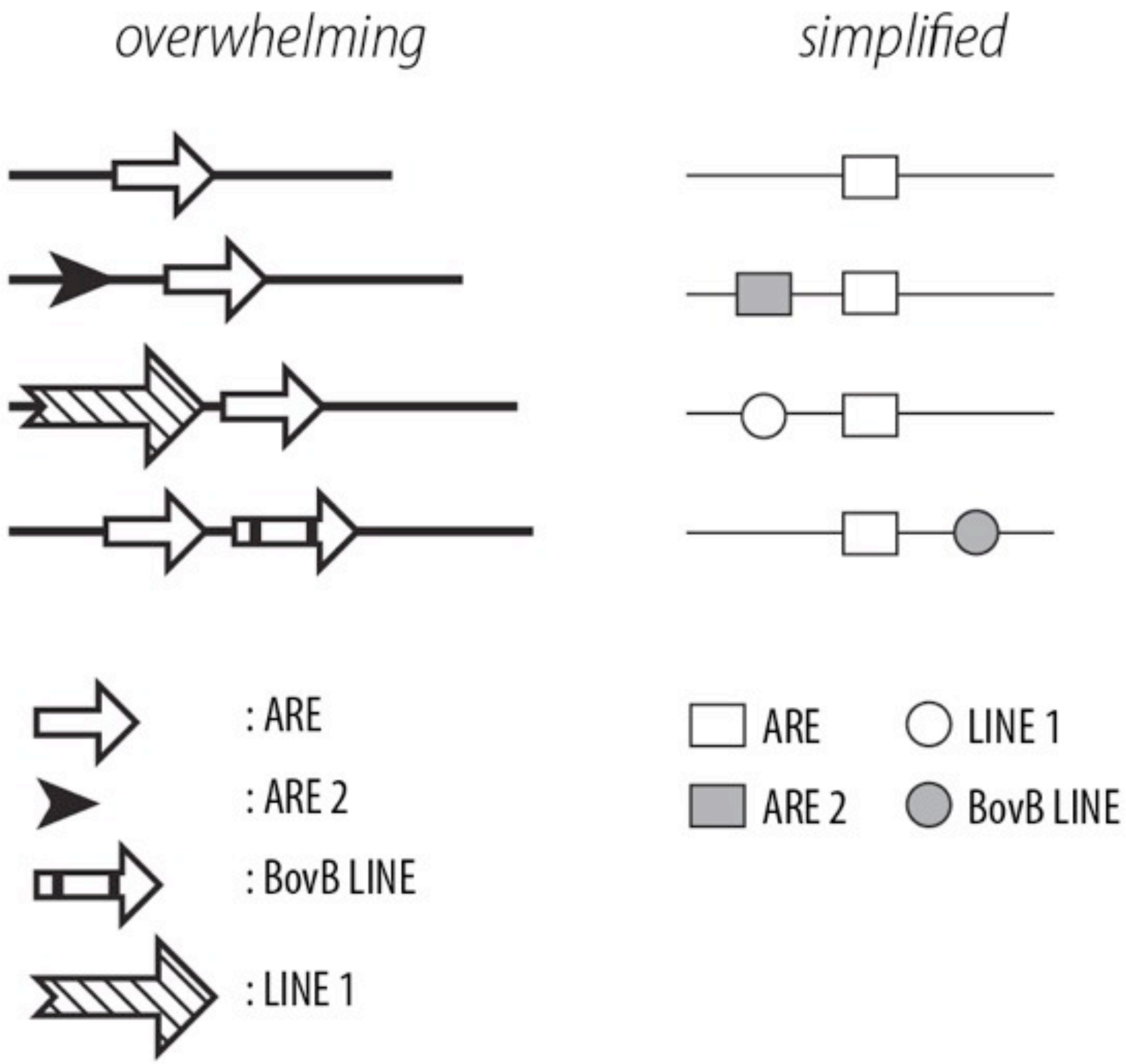
SIMPLIFICATION OF ELEMENTS



McEwen GK, Woolfe A, Goode D, Vavouri T, Callaway H, et al. (2006) Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. Genome Res 16: 451-465.

The reader must work hard to overlook the unnecessary variation in the arrows. Notice how the weight and size of arrows in the CNE EVOLUTION panel is irrelevant, but used to encode rate of evolution in the EVOLUTION RATE panel. The time axis already indicates the flow of time, making all other arrows redundant. Notice how the redesigned figure uses alignment and consistency to enhance the salience of critical elements.

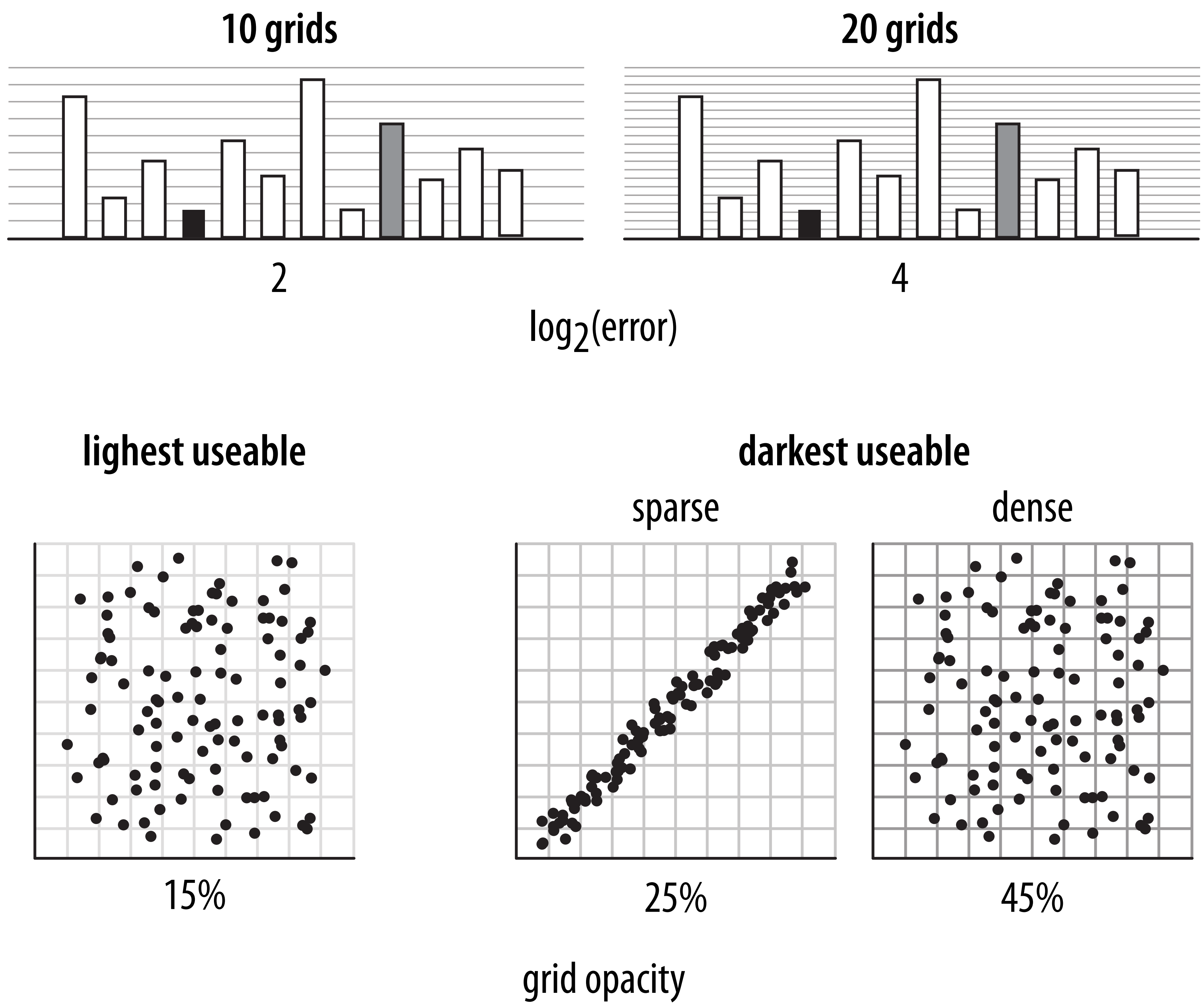
REMOVE REDUNDANCY AND FLOURISH



Nikaido M, Rooney AP, Okada N (1999) Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: Hippopotamuses are the closest extant relatives of whales. Proceedings of the National Academy of Sciences 96: 10261-10266.

The use of visually complex elements obscures the figure's message when patterns in spatial relationships are important. Here, because all the arrows point in the same direction, they are not useful. Again, alignment and simplification increases the salience of important patterns (e.g. relationship of elements to ARE) and attenuates unimportant characteristics, such as size of the elements and absolute distance between them (it is unclear whether the arrows' size and distance is meaningful in the original figure).

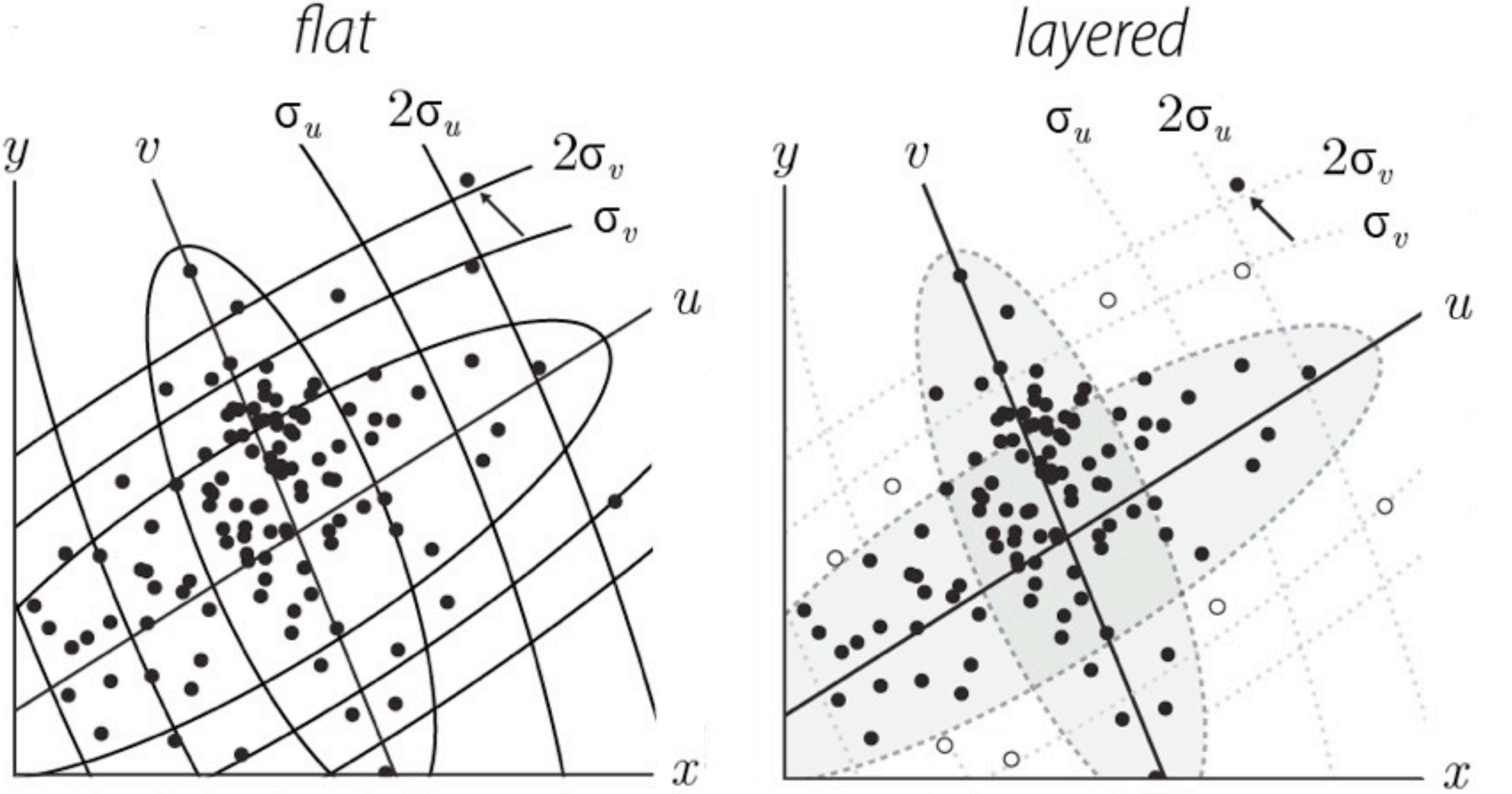
GRIDS



Heer J, Bostock M (2010) Crowdsourcing graphical perception: using mechanical turk to assess visualization design. Proceedings of the 28th international conference on Human factors in computing systems. Atlanta, Georgia, USA: ACM. pp. 203-212.

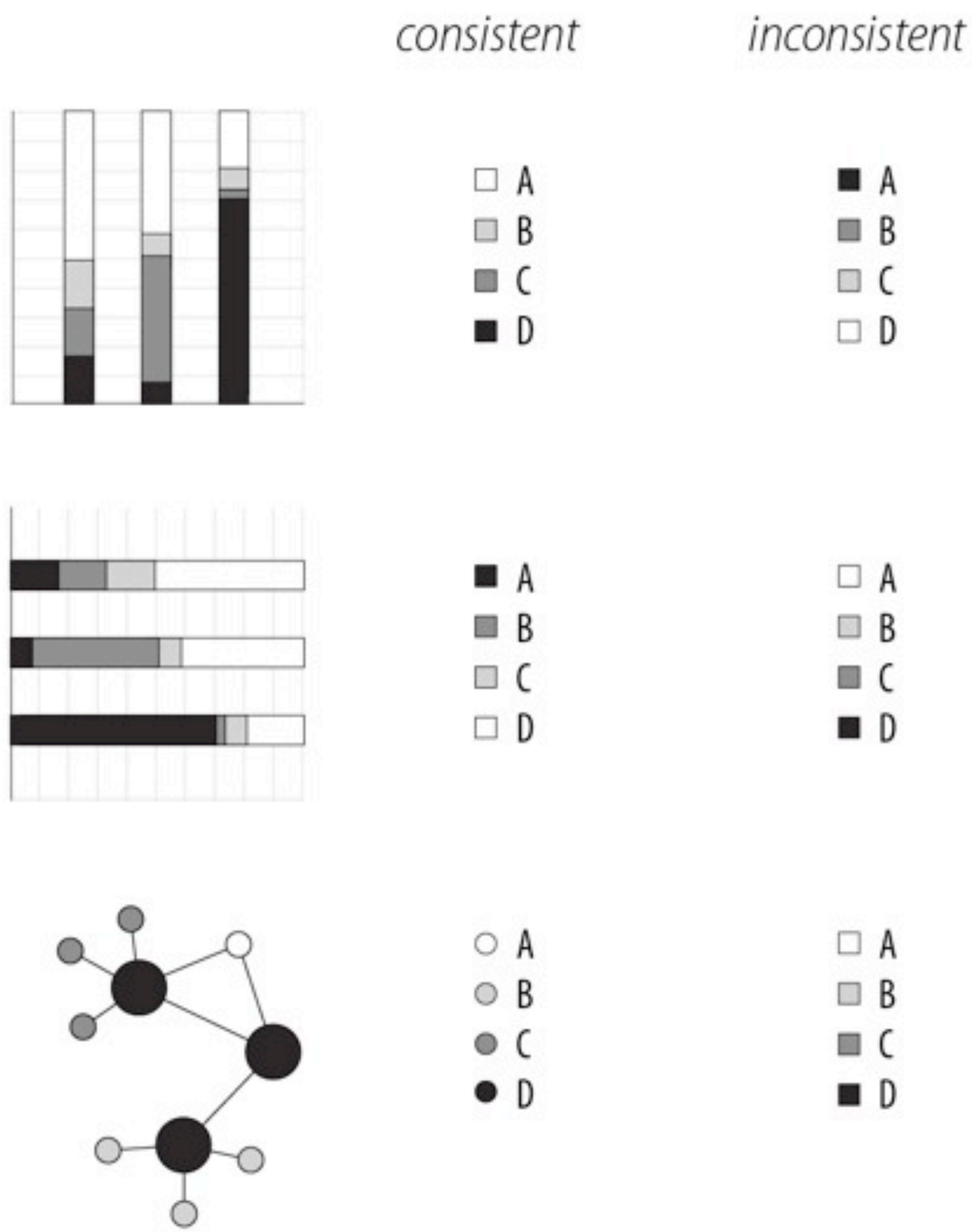
Use opacity to layer elements in the figure. For grids and navigational cues, 20% opacity is recommended.

VARIABLE OPACITY FOR LAYERING



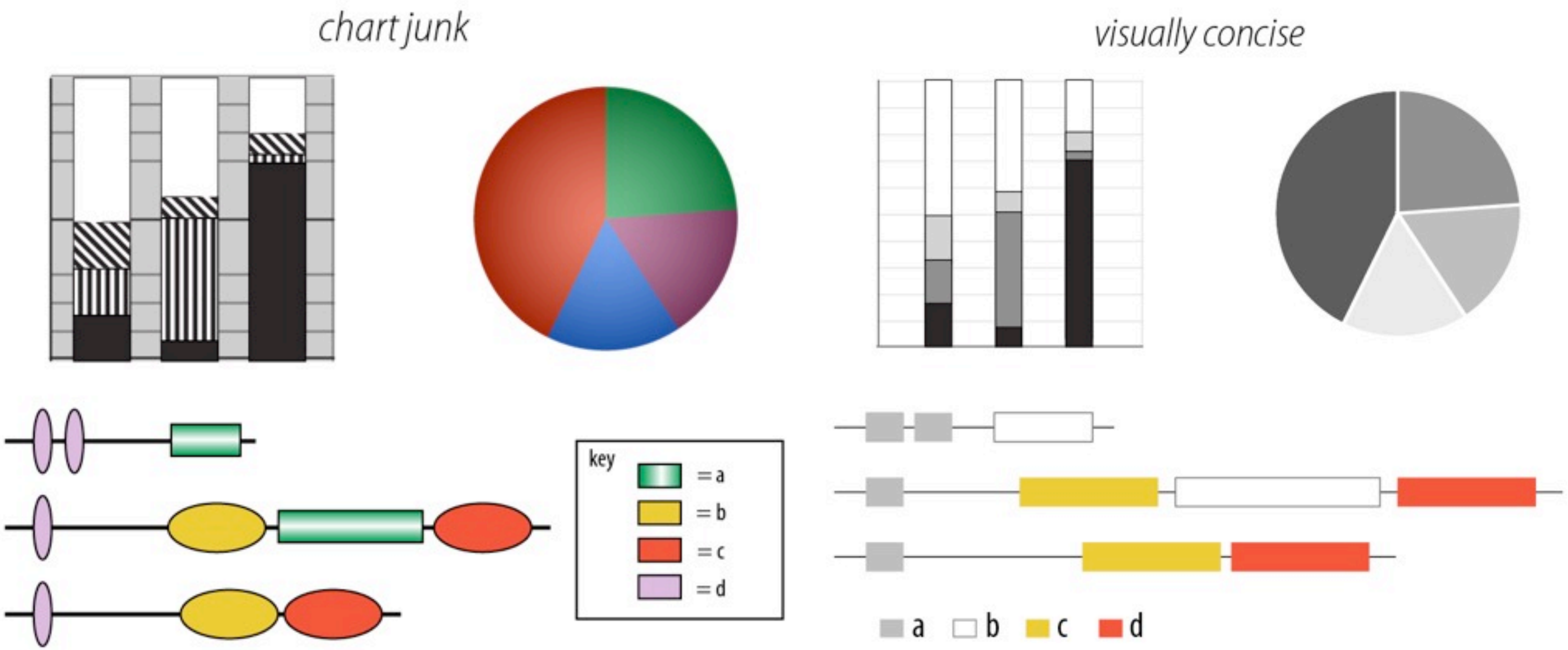
The Gestalt principle of similarity describes our tendency to treat similar objects as related. In this example, the similarity is achieved with opacity and stroke.

LEGEND ORDER



Order elements in the legend consistently with their appearance in the figure. A stack of elements will be perceived more visually balanced when darker tones are at the bottom. Where possible, use the same shape in the legend and the figure (e.g. do not use square symbols in a legend for a figure that uses circles).

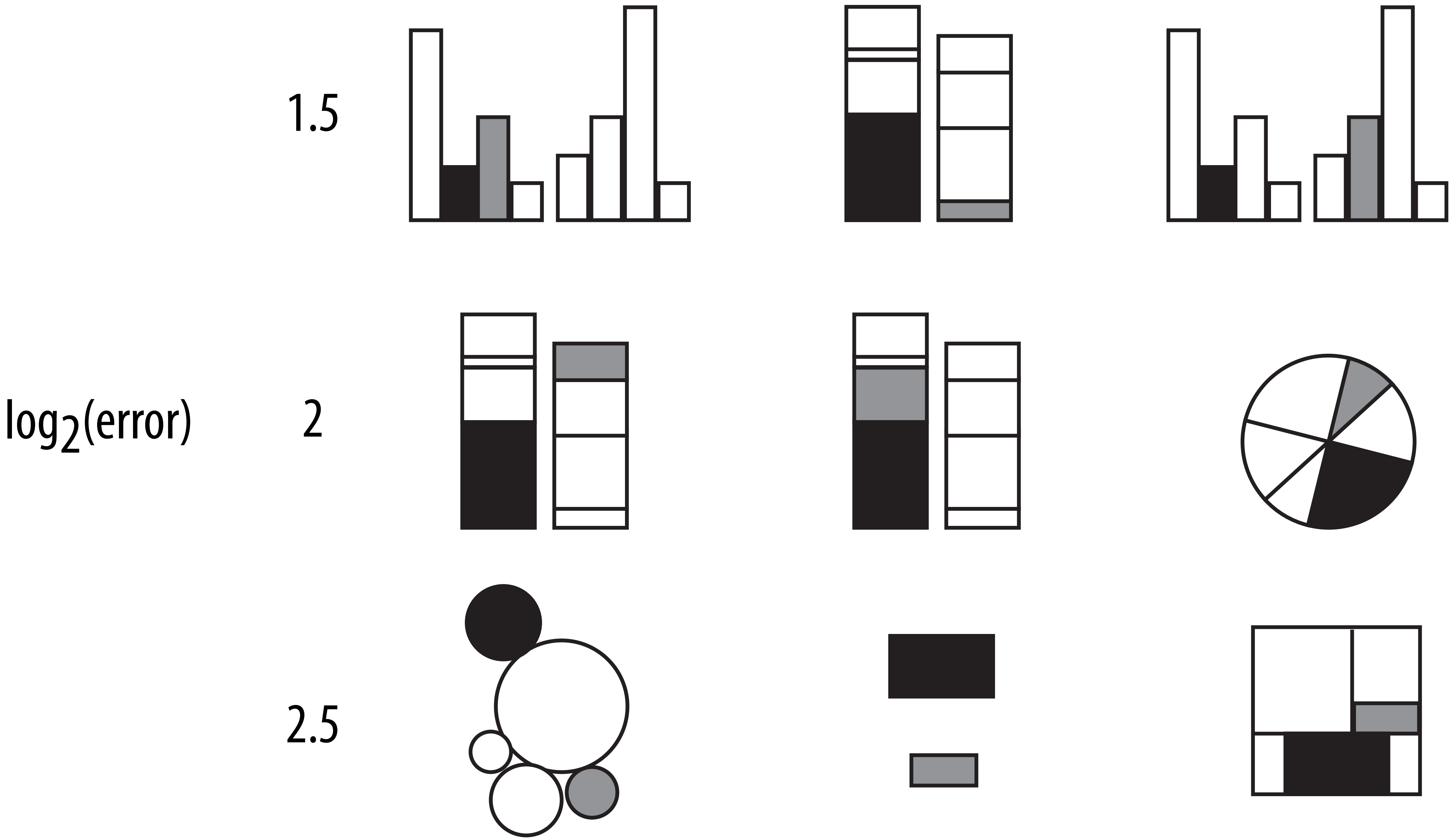
SIMPLIFY



Sharov AA, Malko DB, Makeev VJ, Mironov AA, Gelfand MS (2006) Evolution of exon-intron structure and alternative splicing in fruit flies and malarial mosquito genomes. Genome Res 16: 505-509. Peterson J, Garges S, Giovanni M, McInnes P, Wang L, et al. (2009) The NIH Human Microbiome Project. Genome Res 19: 2317-2323. Thomson NR, Yeats C, Bell K, Holden MT, Bentley SD, et al. (2005) The Chlamydophila abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation. Genome Res 15: 629-640. DB, Ko MS (2005) Genome-wide assembly and analysis of alternative transcripts in mouse. Genome Res 15: 748-754.

Resist the urge to decorate your figure. A simple and concise presentation will be valued by the reader. With each unnecessary variation the value of the figure is decreased.

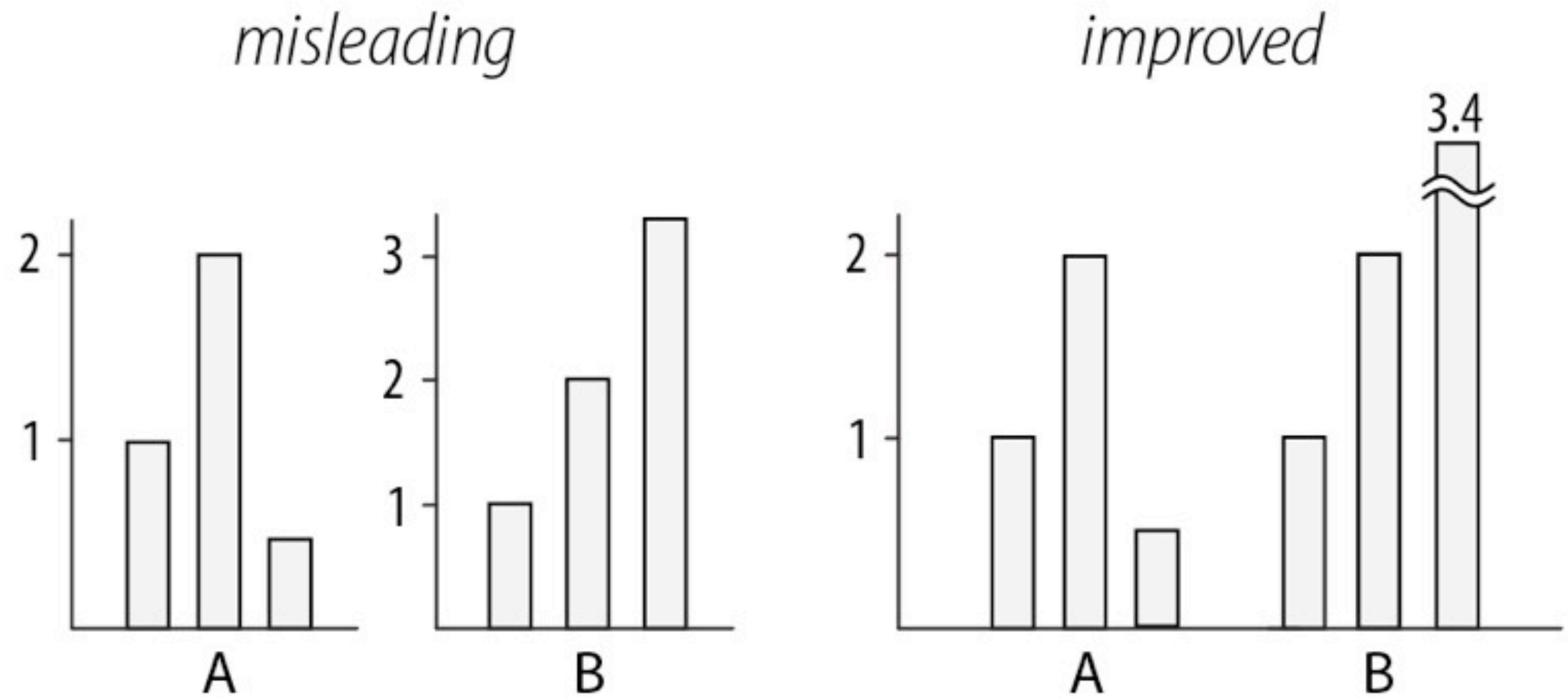
JUDGING LENGTHS AND AREAS



Heer J, Bostock M (2010) Crowdsourcing graphical perception: using mechanical turk to assess visualization design. Proceedings of the 28th international conference on Human factors in computing systems. Atlanta, Georgia, USA: ACM. pp. 203-212.

Our ability to judge lengths and areas is affected by proximity and alignment. We typically underestimate area because we use length as its proxy.

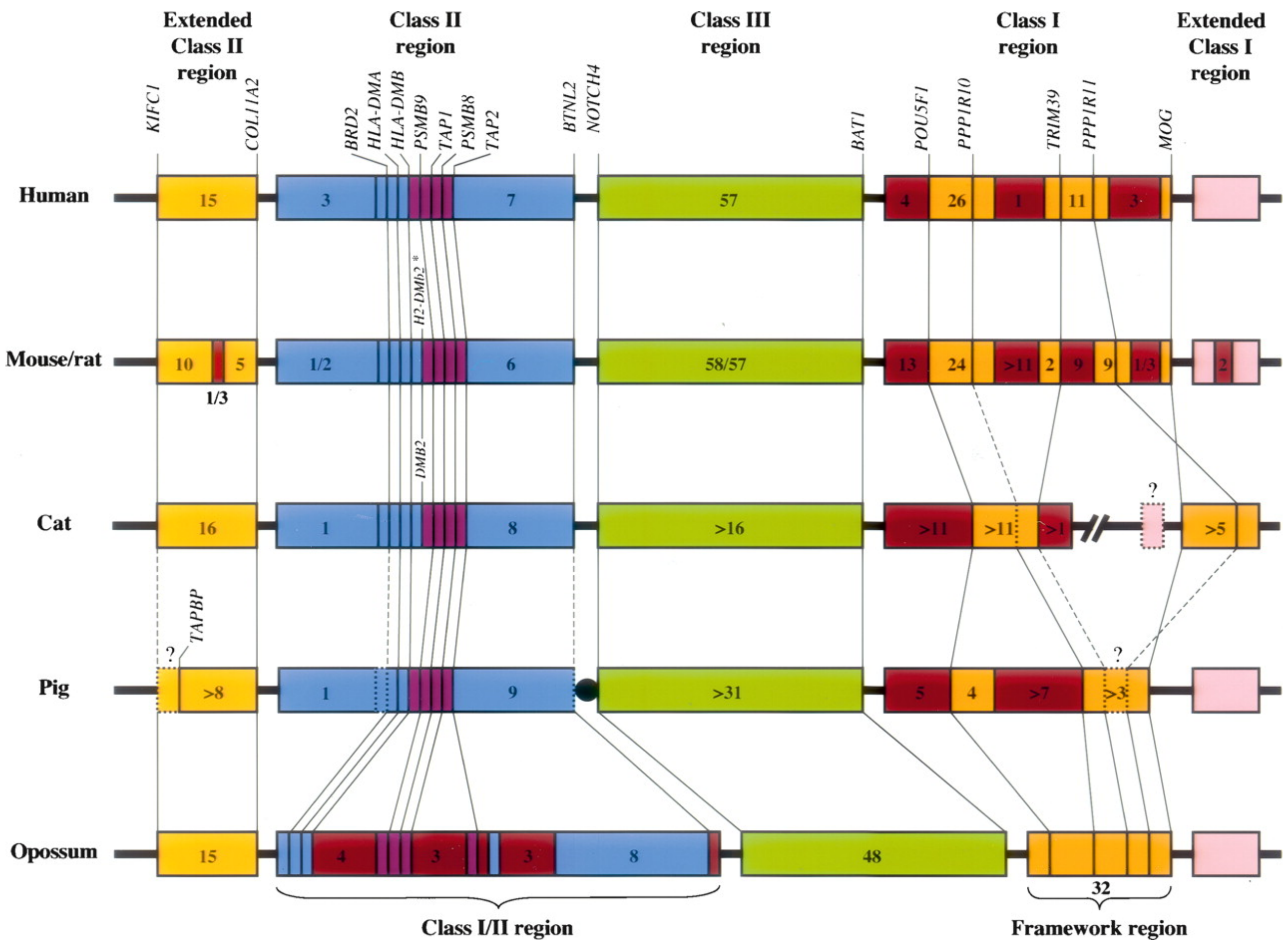
CONSISTENT AXES



e.g. Fig 1 in Raina SZ, Faith JJ, Disotell TR, Seligmann H, Stewart CB, et al. (2005) Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res 15: 665-673.

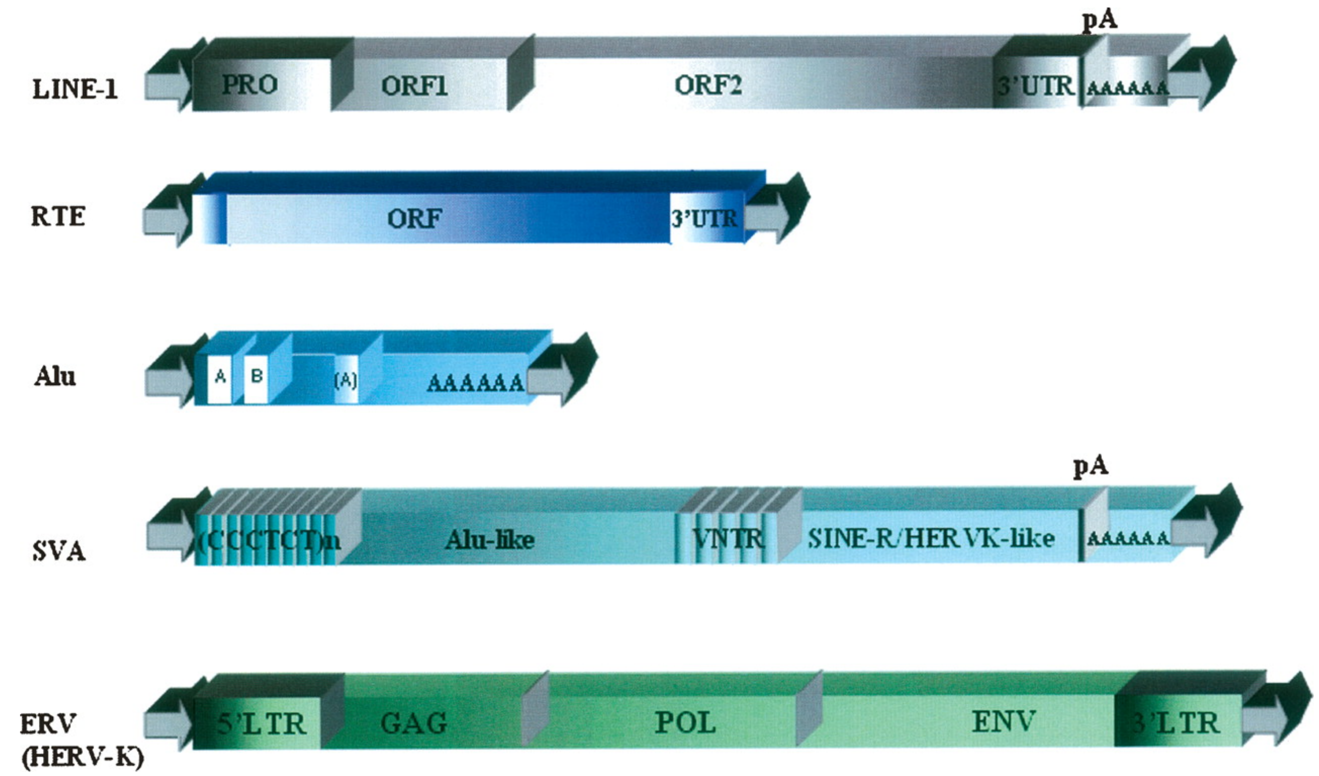
Avoid changing axis scales. You can save the reader from having to check the consistency of your axes by showing a single axis for multiple panels.

RESULTS OF USING DESIGN RULES



Samollow, P.B., The opossum genome: insights and opportunities from an alternative mammal. Genome Res, 2008. 18(8): p. 1199-215.

Excellent organization and consistency. Clear use of color. Light vertical lines cue continuity without overwhelming the figure.



Gentles, A.J., et al., Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. Genome Res, 2007. 17(7): p. 992-1004.

Chart junk. Illegible and inconsistently formatted text. Redundant elements. What are the patterns here?

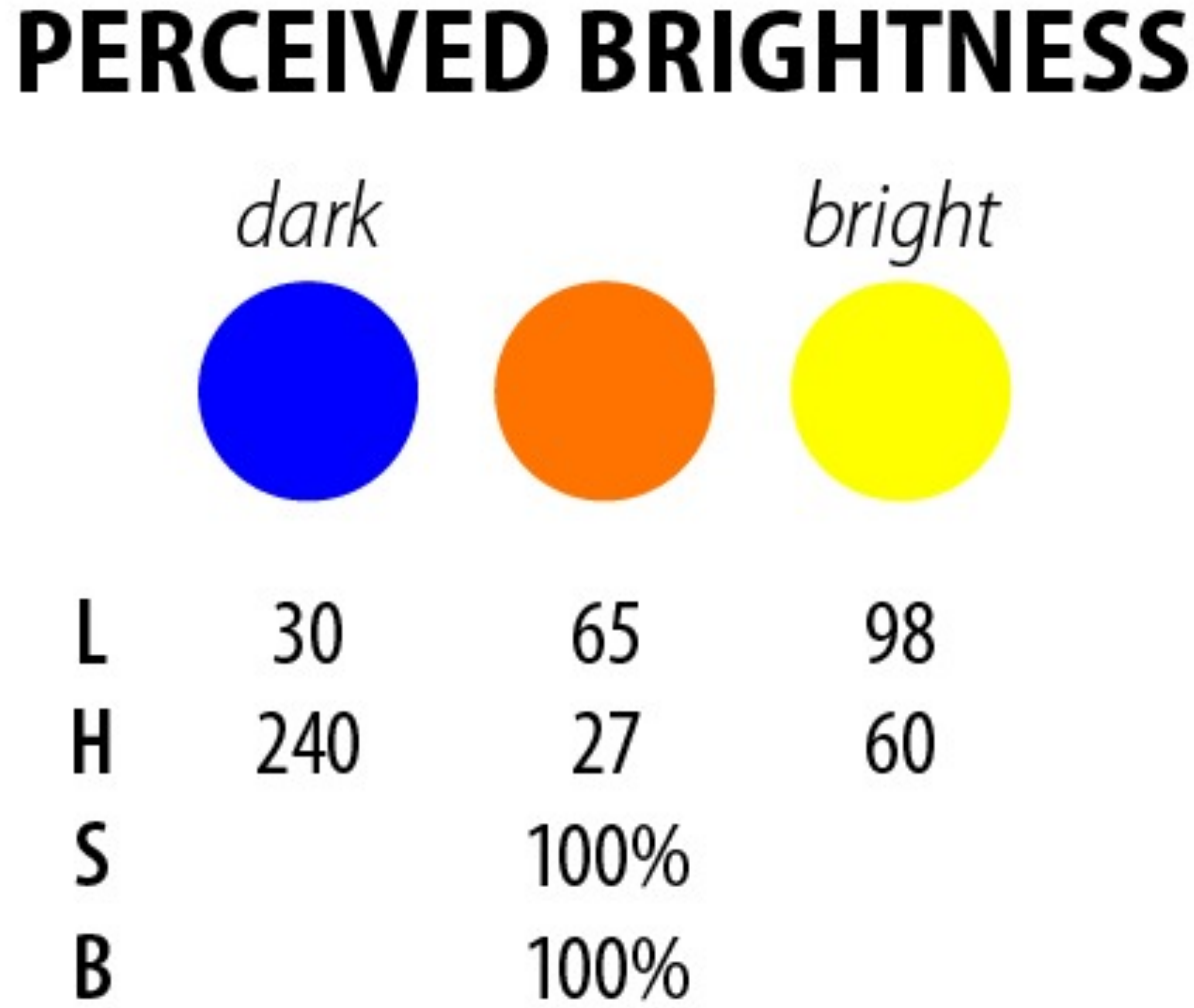
These two figures demonstrate the benefits of applying some of the design rules mentioned on the previous slides.

requirements for effective visual communication

LEGIBILITY
CLARITY
ATTRACTIVENESS

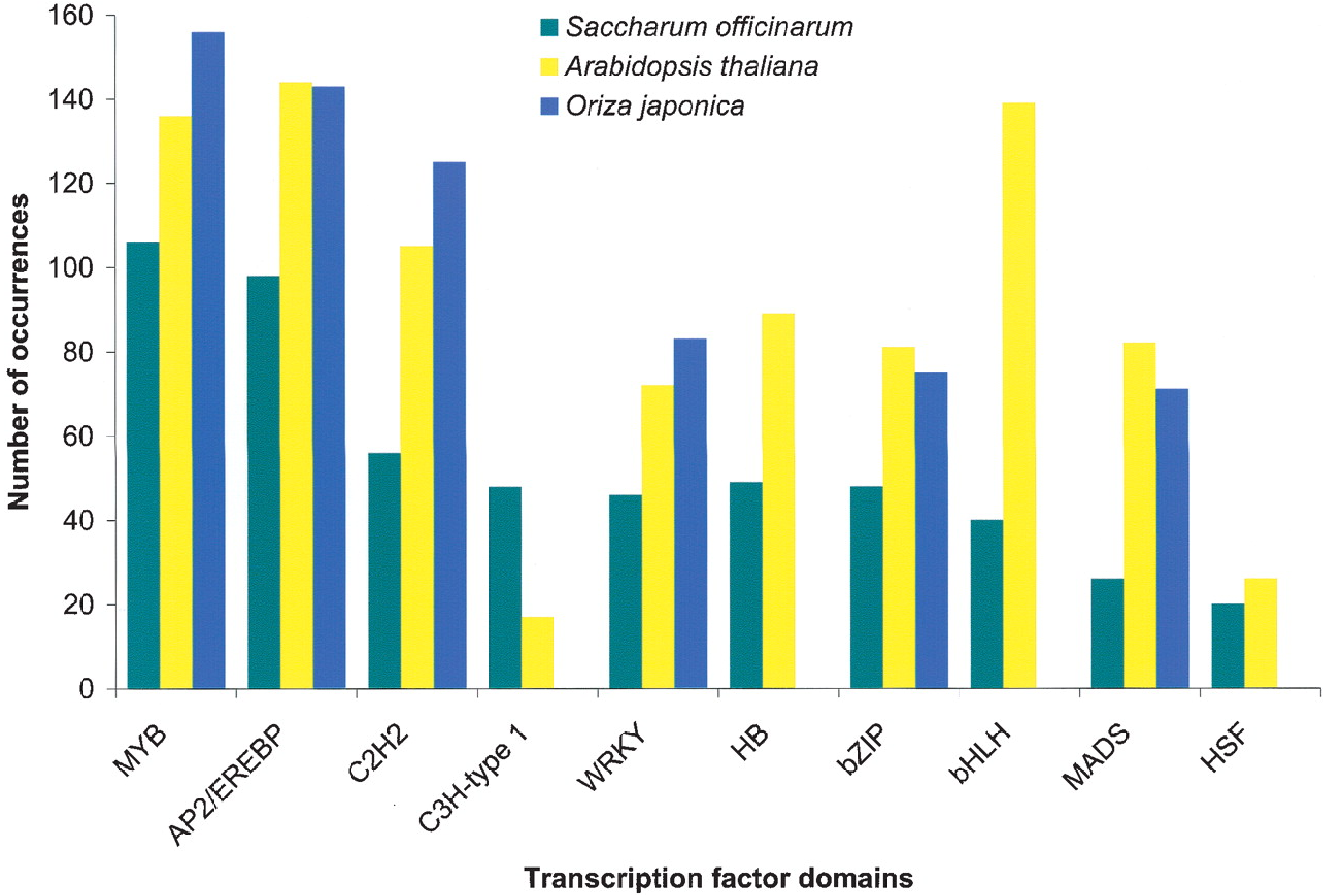
COLORCOLO
RCOLORCOL
ORCOLORCO
LORCOLORC

COLOR PERCEPTION



HSB is a common color space. Unfortunately, the way we perceive color is not fully captured by a color's HSV values.
For example, yellow appears brighter than blue even when their HSB brightness is the same. A better metric is the luminance (L), which measures the perceived brightness. The luminance of the yellow shown here is 3x that of the blue.

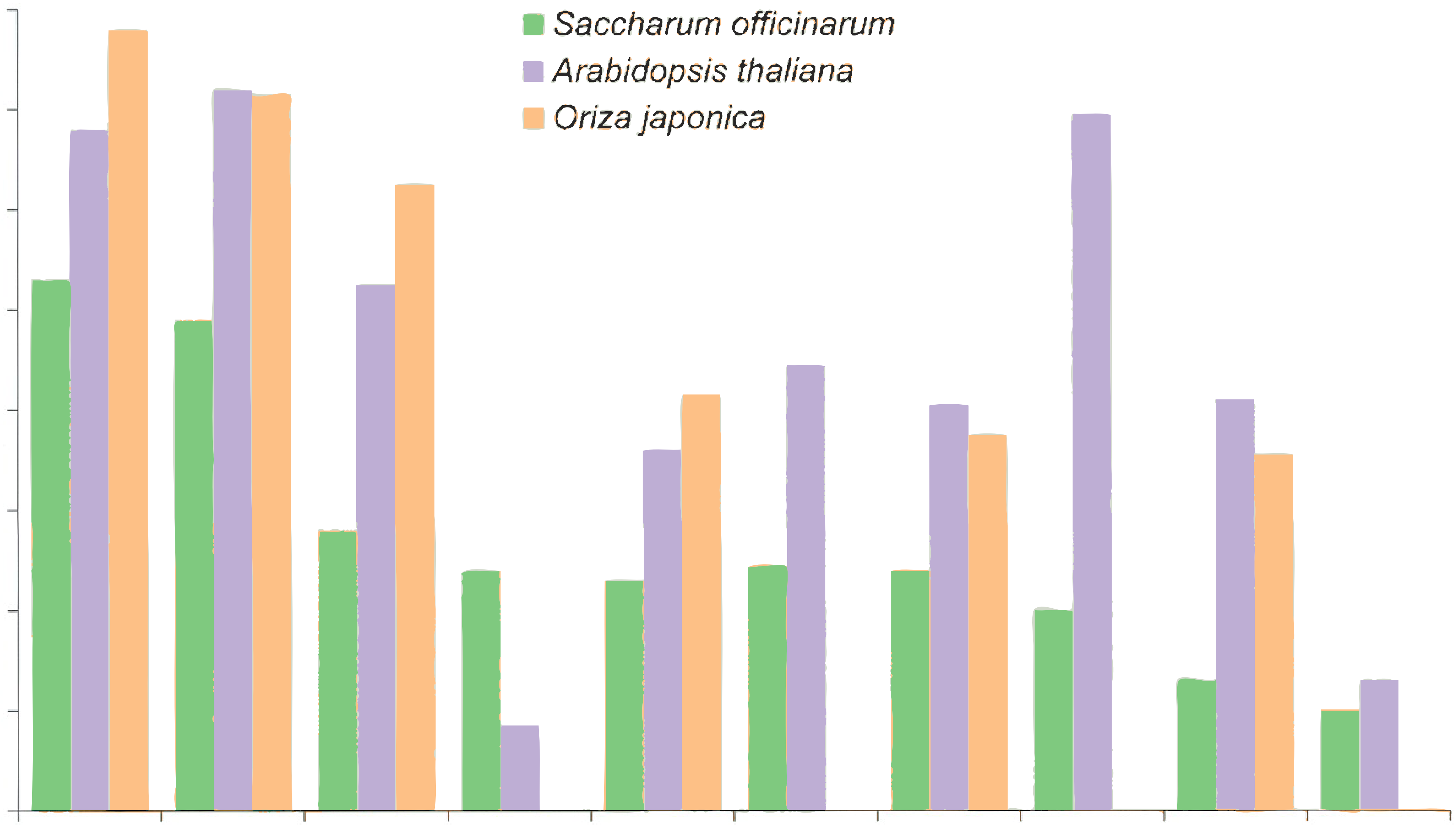
COLOR PERCEPTION



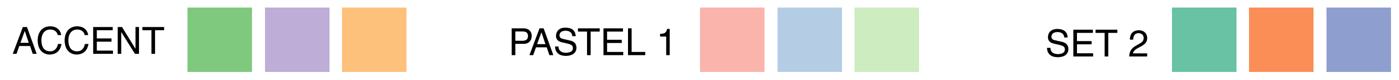
The 10 most common transcription factor Pfam domains in SAS proteins. Vettore, A.L., et al., Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res, 2003. 13(12): p. 2725-35.

When nominal data is shown (a nominal variable is one which is categorical and where the categories don't have a specific order), avoid the use of colors with large luminance differences. The bright colors (yellow) will dominate attention. In this example, the green and blue bars are seen as very similar, in contrast to the yellow, and form a group. This is misleading because each of the categories is independent.

COLOR PERCEPTION



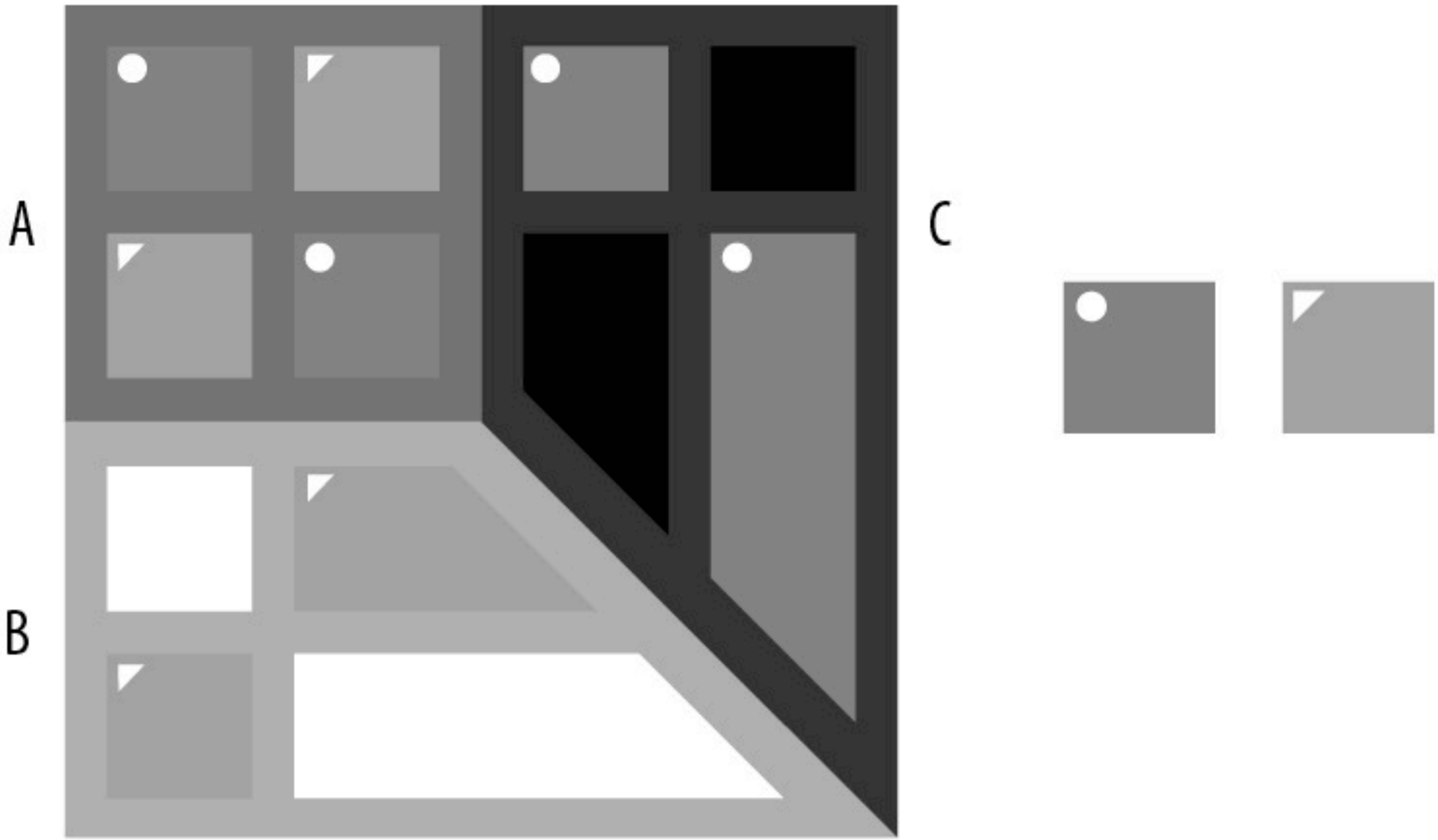
BREWER QUALITATIVE 3-COLOR PALETTES



W W W . C O L O R B R E W E R . O R G

Use the Brewer palettes for colors.

LUMINANCE EFFECT

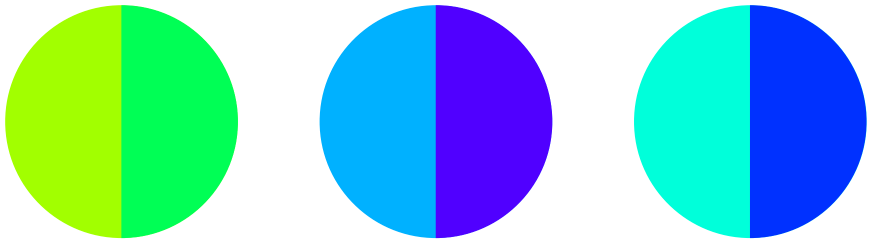


The perceived brightness of a tone is affected by its context. Note how the light grey in A appears lighter than the same tone in B. Try to persuade yourself while looking at the figure that the colors are indeed the same. It is impossible - the effect of the background is too strong.

PERCEPTUAL UNIFORMITY

PERCEIVED DIFFERENCE

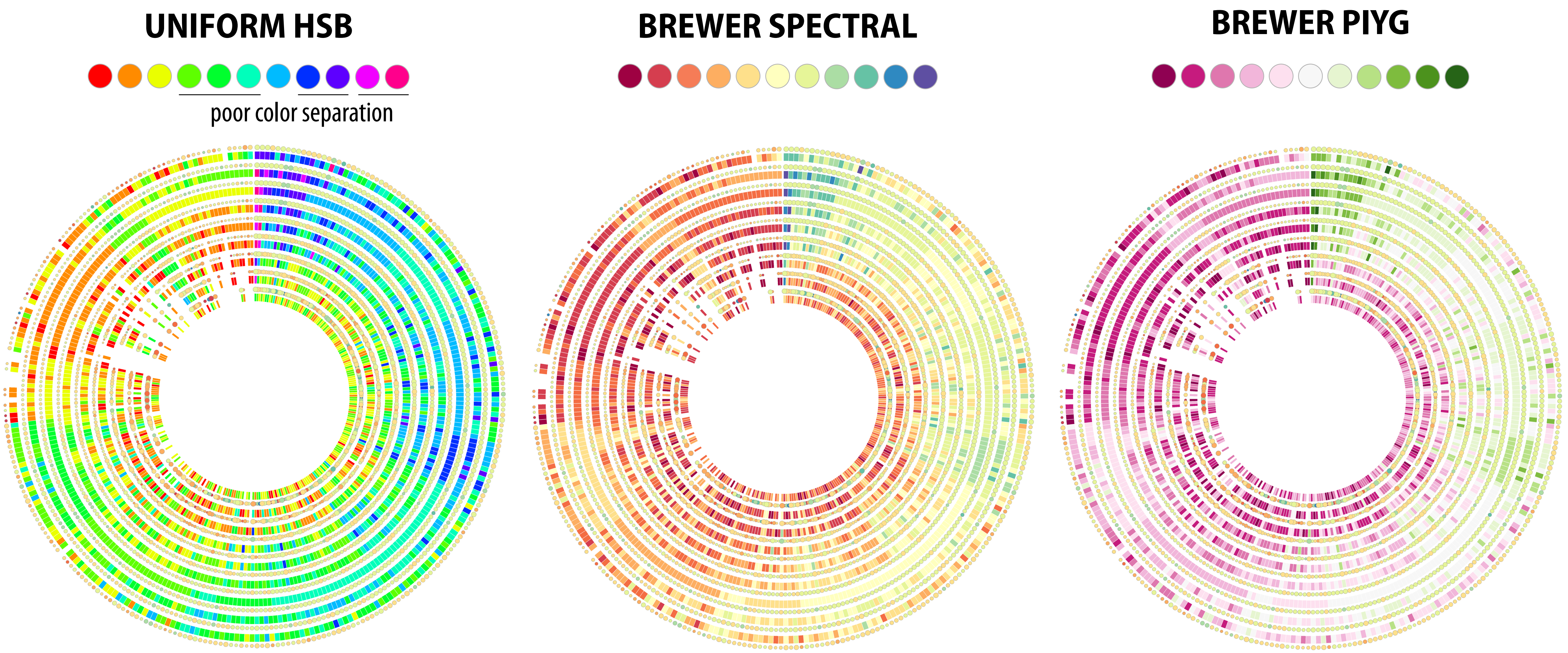
similar *different*



ΔE	36	104	173
H	83 143	200 260	170 230
ΔH		60	
ΔS		0	
ΔB		0	

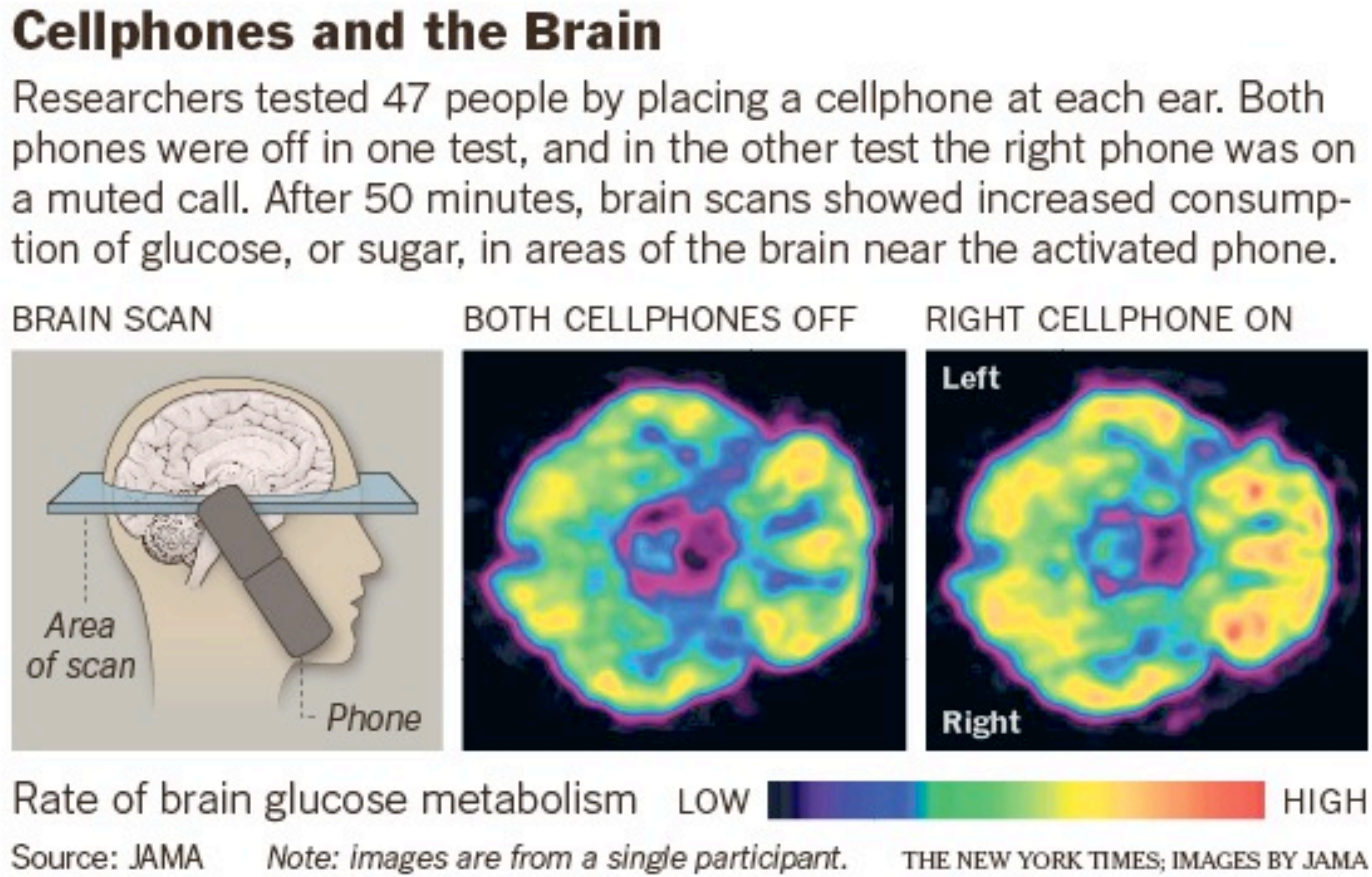
You should be aware that HSB is not perceptually uniform. This means that the perceived difference between colors is not proportional to their separation in HSB. For example, three color pairs whose components are each 60 hue units apart will have varying perceived difference depending on the hue. For this ΔH=60, the two greens shown here are the most similar HSB pair and the cyan/blue are the most different.

BREWER PALETTES



Sequential and diverging Brewer palettes have desirable perceptual qualities. The apparent difference between adjacent colors is uniform, making the palettes useful for heat maps. Generally, you should avoid the spectral palette to encode magnitude and instead choose two hue variants, such as the pink-yellow-green palette.

HUE AS MAGNITUDE

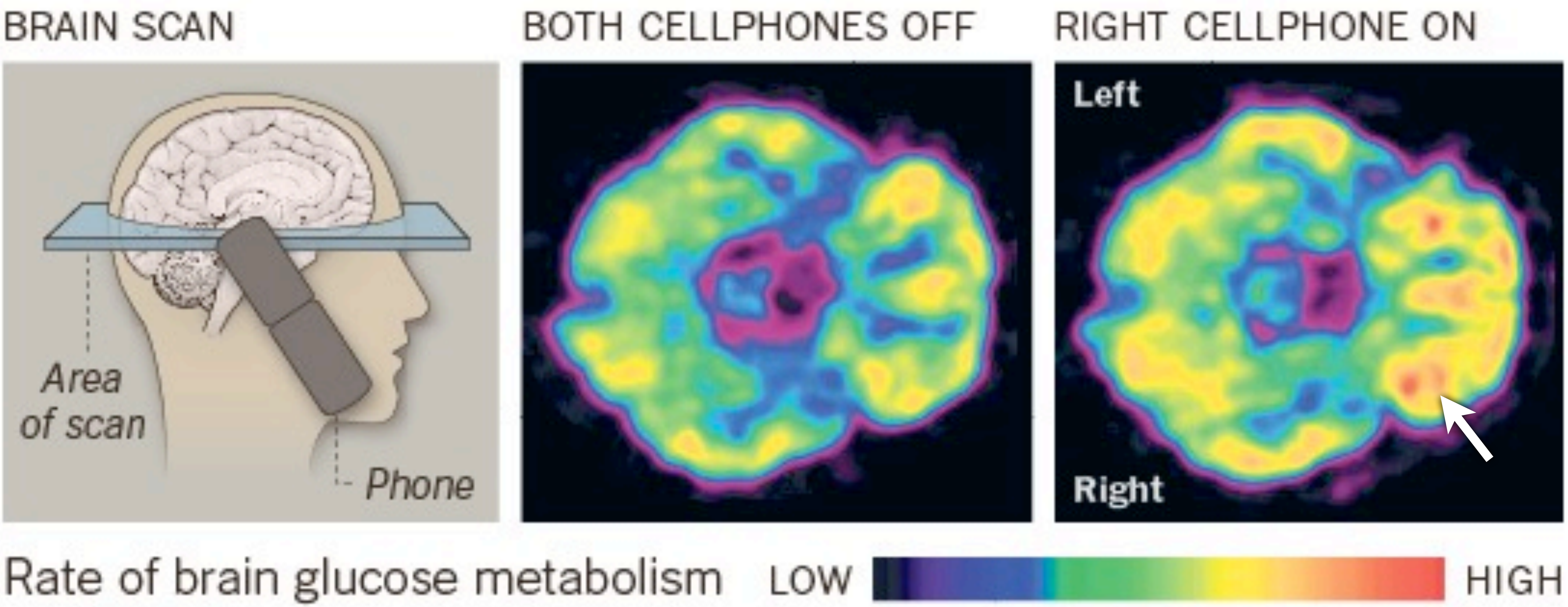


This image from the NYT confused readers, who interpreted the metabolism of the entire brain to be increased when the cellphone was on.

HUE AS MAGNITUDE

Cellphones and the Brain

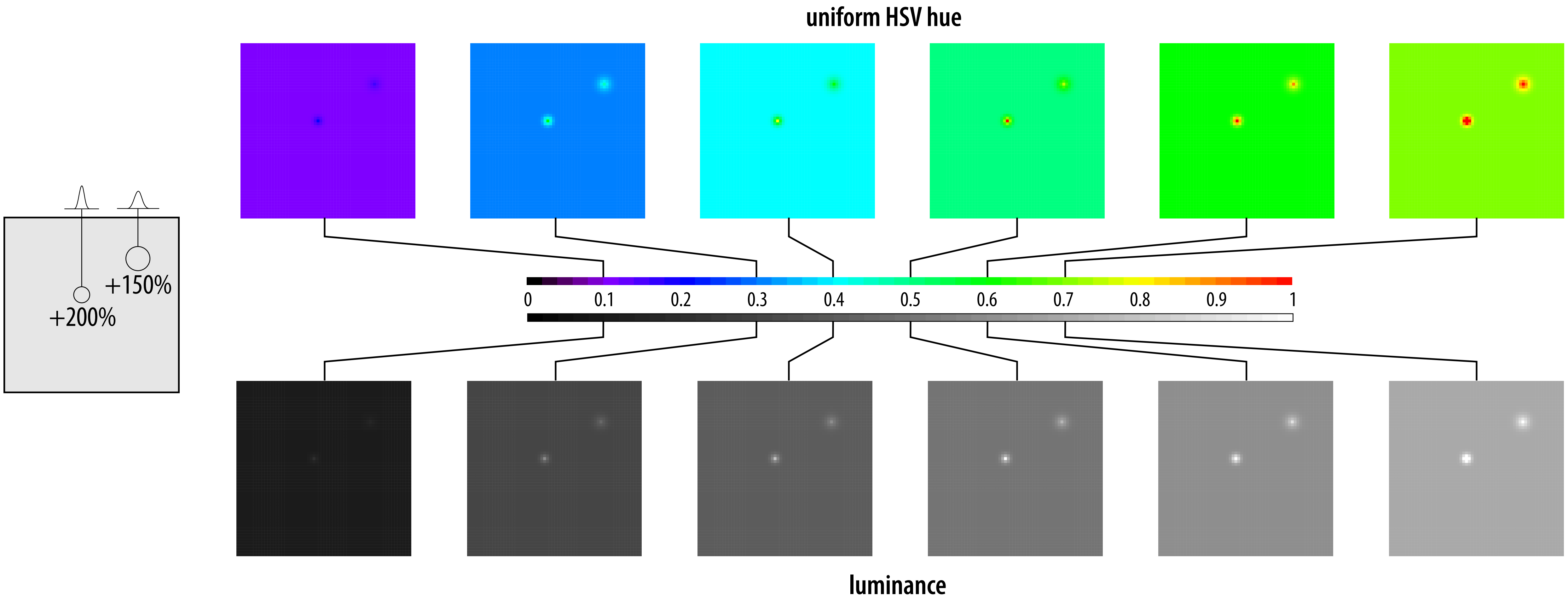
Researchers tested 47 people by placing a cellphone at each ear. Both phones were off in one test, and in the other test the right phone was on a muted call. After 50 minutes, brain scans showed increased consumption of glucose, or sugar, in areas of the brain near the activated phone.



Source: JAMA *Note: images are from a single participant.* THE NEW YORK TIMES; IMAGES BY JAMA

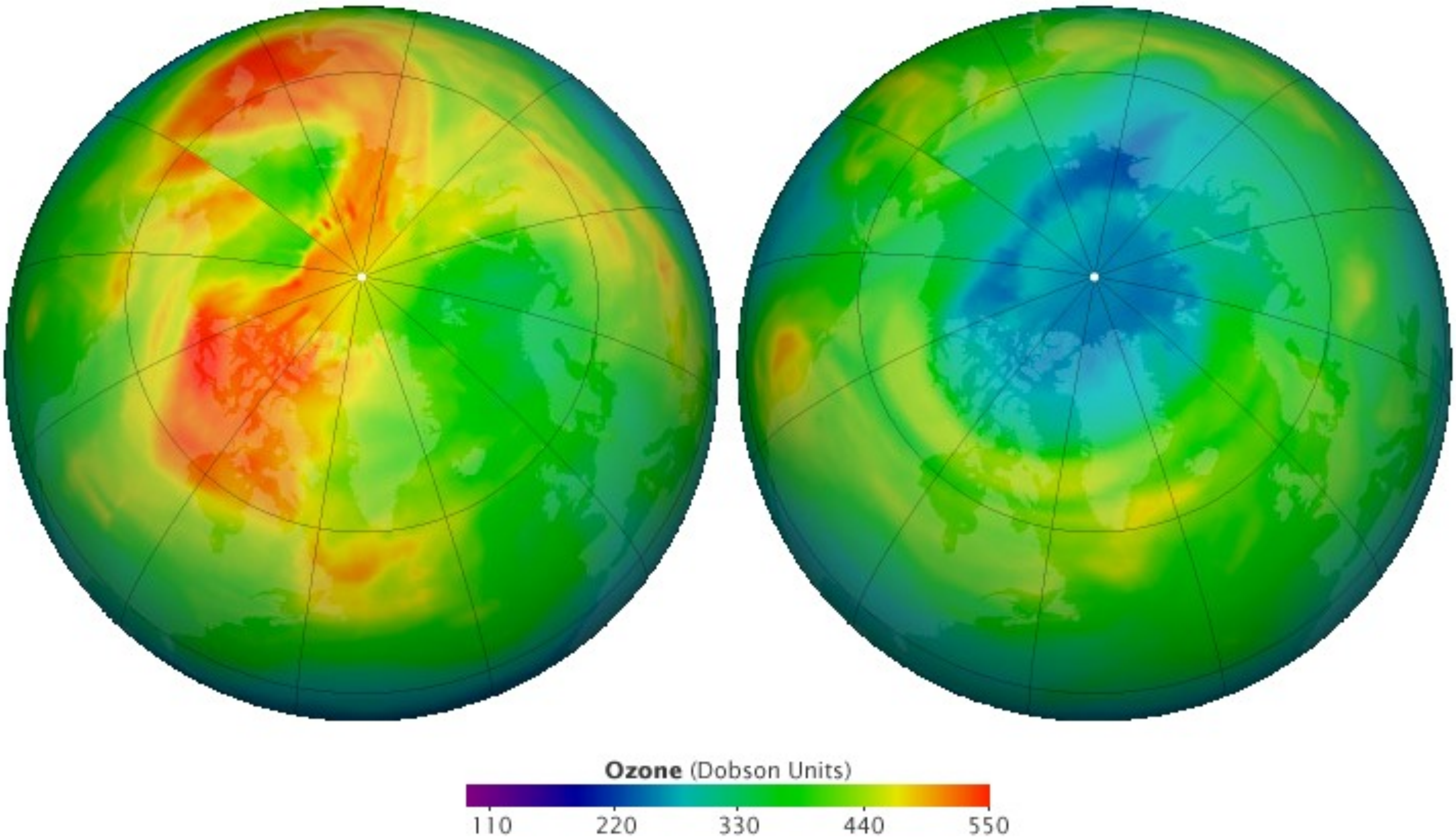
The editors added an arrow to the right panel to cue the reader to the important difference. The reason why the entire brain appears to be more active is a difference in baselines between the images.

HUE SHOULD NEVER ENCODE MAGNITUDE



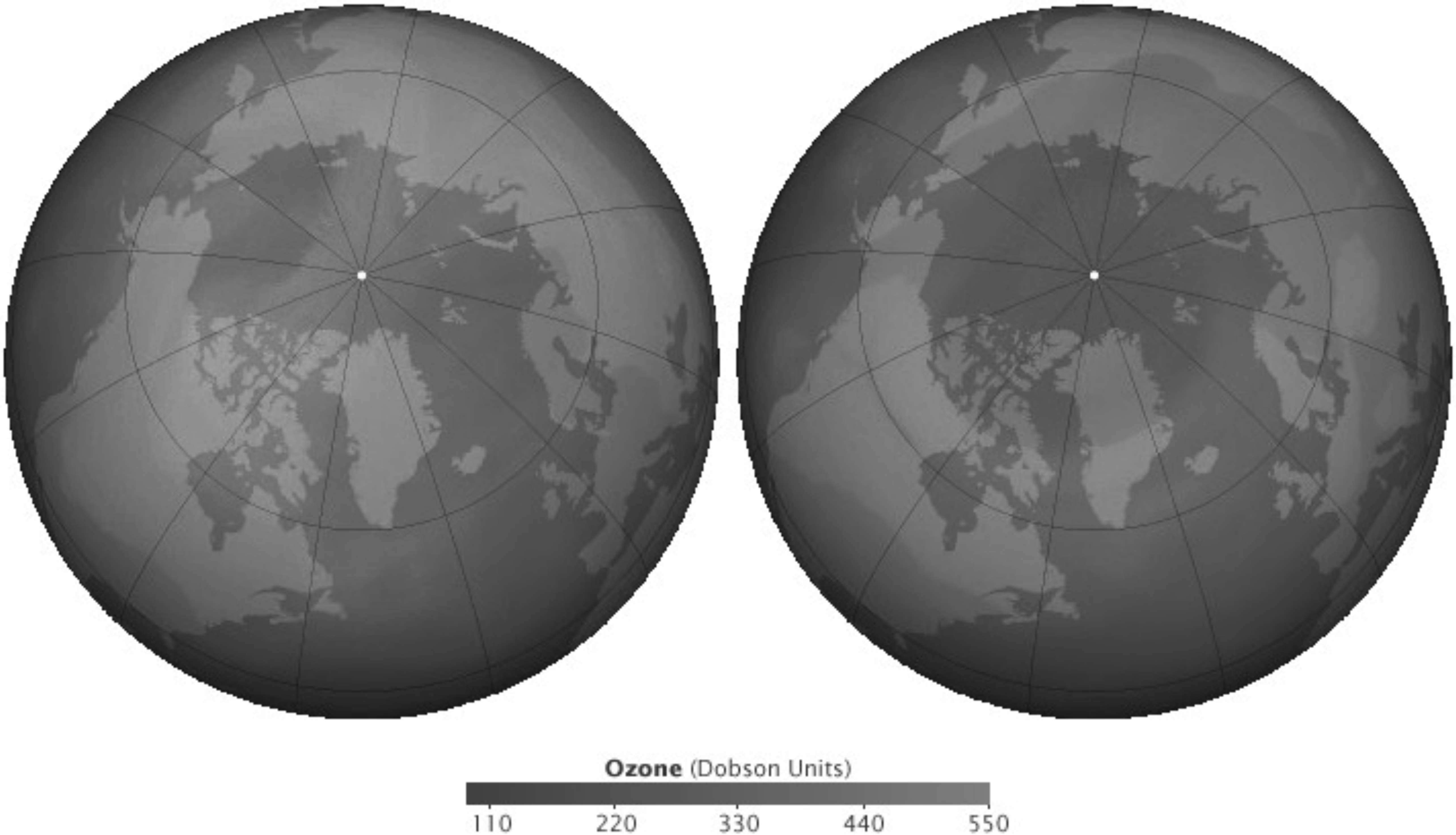
When hue is used to encode magintude, an increase in the baseline of a signal can be easily misinterpreted. It interferes with our ability to perceive the relative differences in the signal. This is caused by the perceptual non-uniformity of the underlying color scale. Notice on this figure how there is a large region of greens that are not distinguishable (0.5-0.7) but the same-sized region (0.8-1.0) spans three distinct colors (yellow, orange, red).

PRACTICAL CONSEQUENCES OF HUE MAPPING



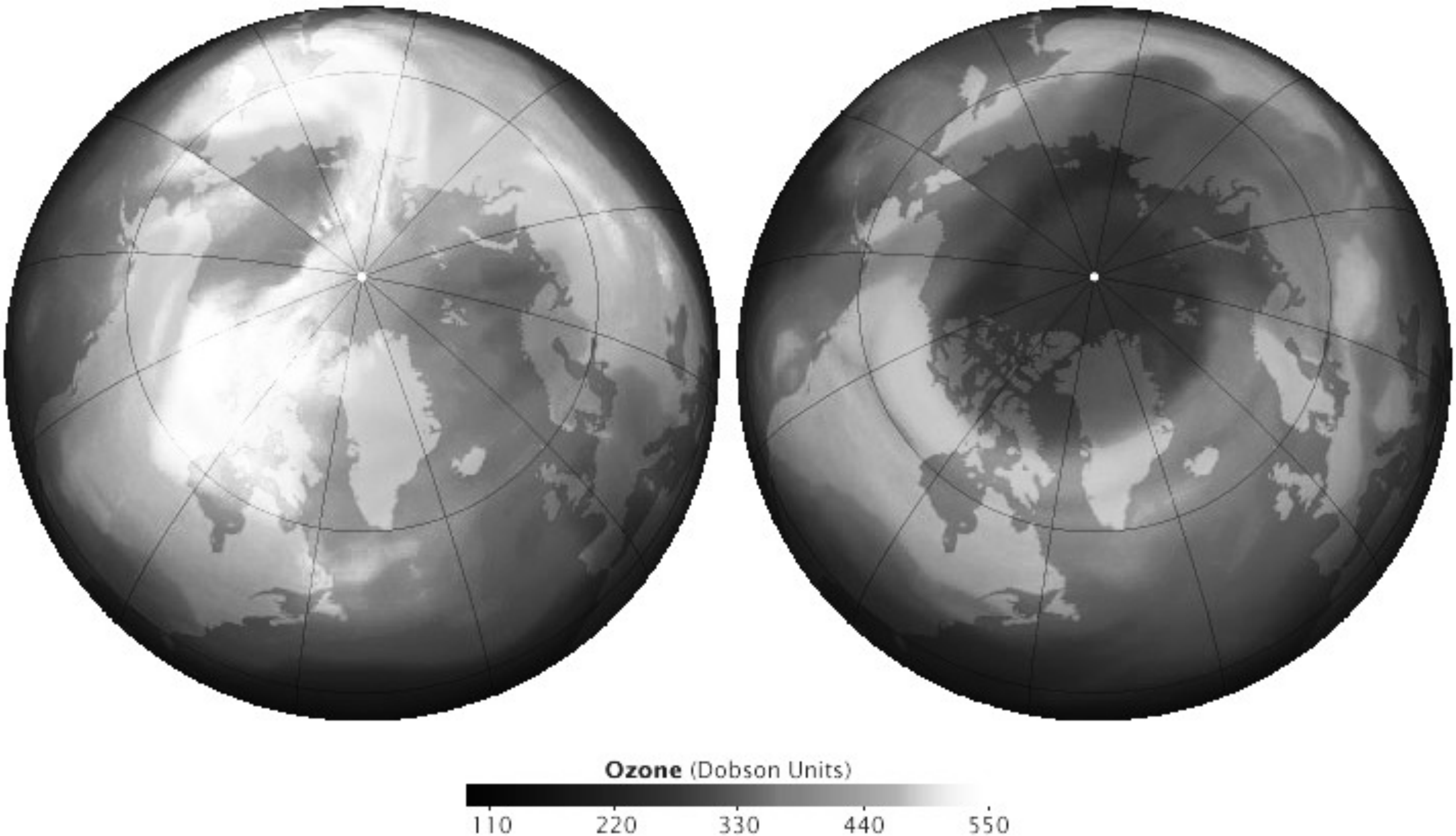
Recent observations from satellites and ground stations suggest that atmospheric ozone levels for March in the Arctic were approaching the lowest levels in the modern instrumental era. *http://earthobservatory.nasa.gov/IOTD/view.php?id=49874*

PRACTICAL CONSEQUENCES OF HUE MAPPING



Standard desaturation, which discards hue and saturation information and keeps only brightness, is inappropriate for images which use hue encoding.
This image did not make it into the New York Times because no interpretable b/w version was available.

MAP MAGNITUDE WITH TONE



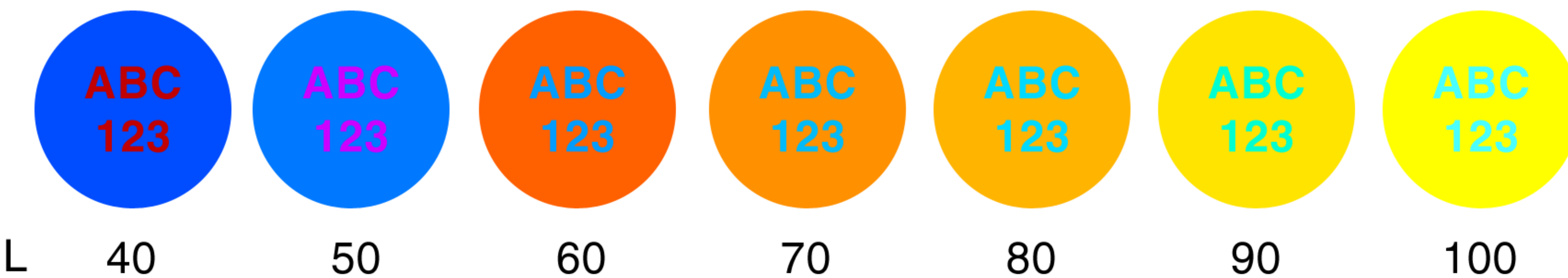
When hue is remapped to tone, the graphic is faithfully translated into black-and-white. This can be accomplished in Adobe Photoshop using the black-and-white adjustment layer and remapping each of the 6 channels (reds, yellows, greens, cyans, blues, purples) to uniformly increasing tone (0, 20, 40, 60, 80, 100%).

CONTRAST PITFALLS

AVOID ADJACENT PURE COLORS

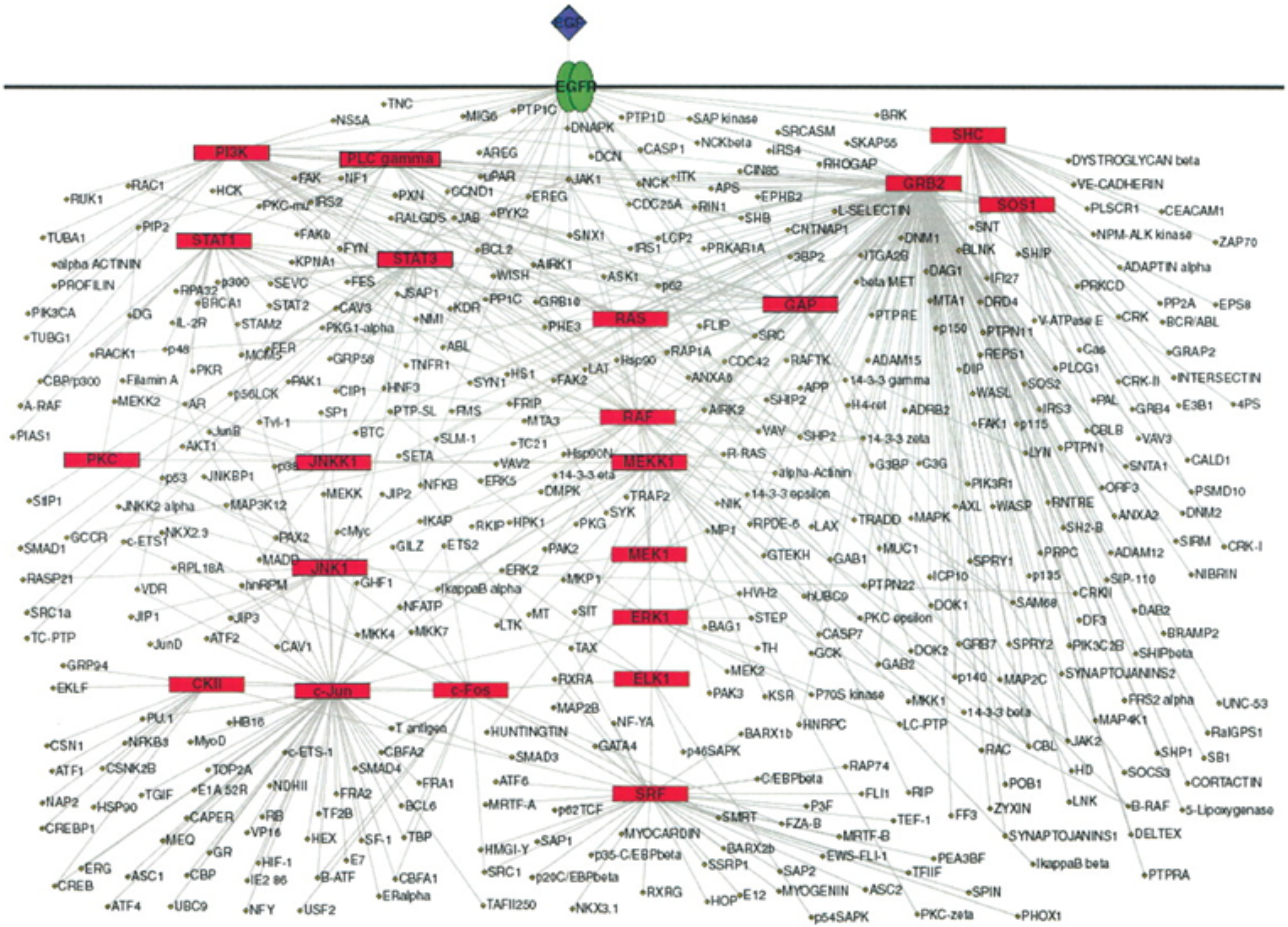
AVOID ADJACENT COLORS WITH SIMILAR LUMINANCE

$$\Delta L = 0$$



Simultaneous contrast occurs when two pure colors are adjacent. Poor contrast occurs when two colors have similar luminance (perceived brightness).

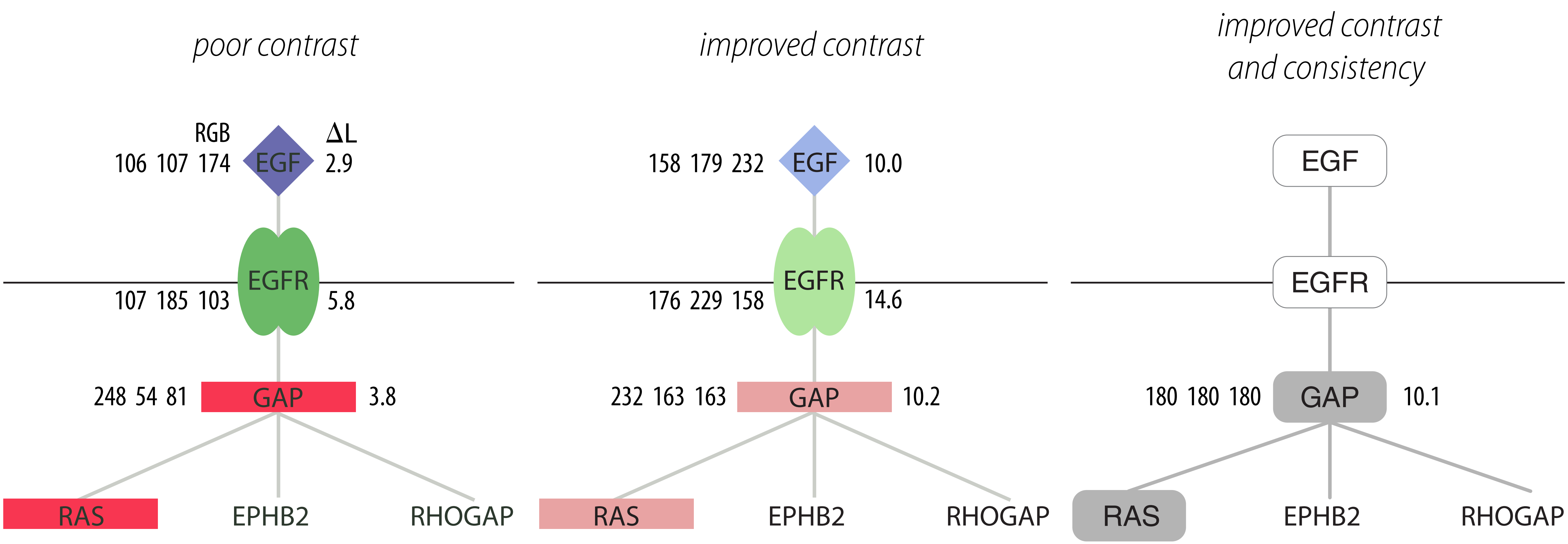
CONTRAST



Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003)
Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 13: 2363-2371.

Black text on dark colors is illegible due to insufficient luminosity contrast.

GOOD CONTRAST



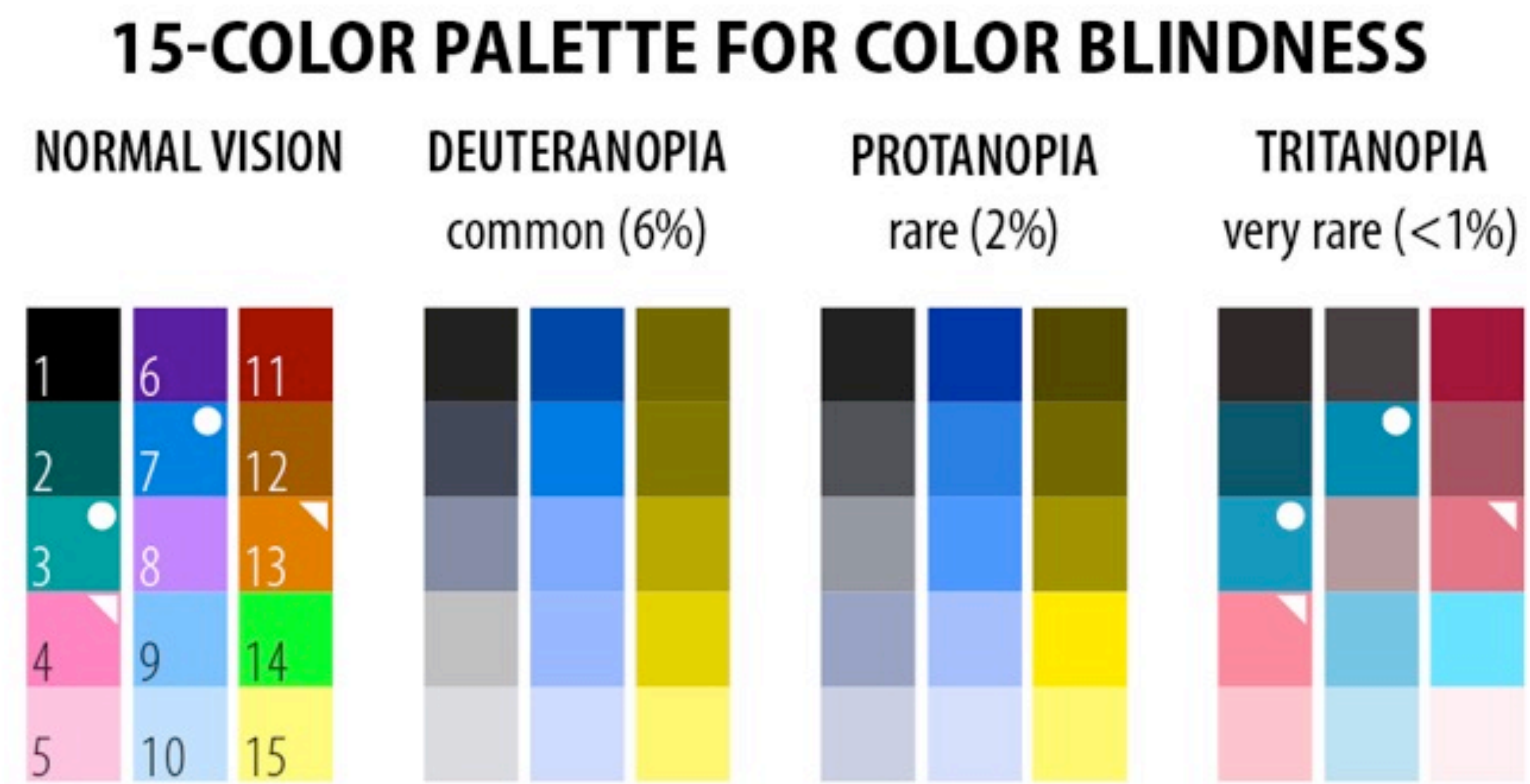
Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003)
Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 13: 2363-2371.

Where possible keep background colors light and desaturated. Only use color where it is unavoidable. Variation in shape, stroke and grey tone is usually sufficient to provide contrast.

COLOR BLINDNESS

Color blindness is not uncommon! In an audience of 8 men and 8 women the chance that someone is affected is 50%. Think about that.
The most common form is red-green color blindness (protanopia and deuteranopia, where the L and M color receptor sensitivity is reduced - the effect is similar). Where possible avoid figures that require the reader to discriminate between red/yellow/green.

COLOR BLINDNESS SAFE PALETTE



	R	G	B		R	G	B		R	G	B
1	0	0	0	6	73	0	146	11	146	0	0
2	0	73	73	7 ●	0	109	219	12	146	73	0
3 ●	0	146	146	8	182	109	255	13 ◥	219	209	0
4 ◥	255	109	182	9	109	182	255	14	36	255	36
5	255	182	119	10	182	219	255	15	255	255	109

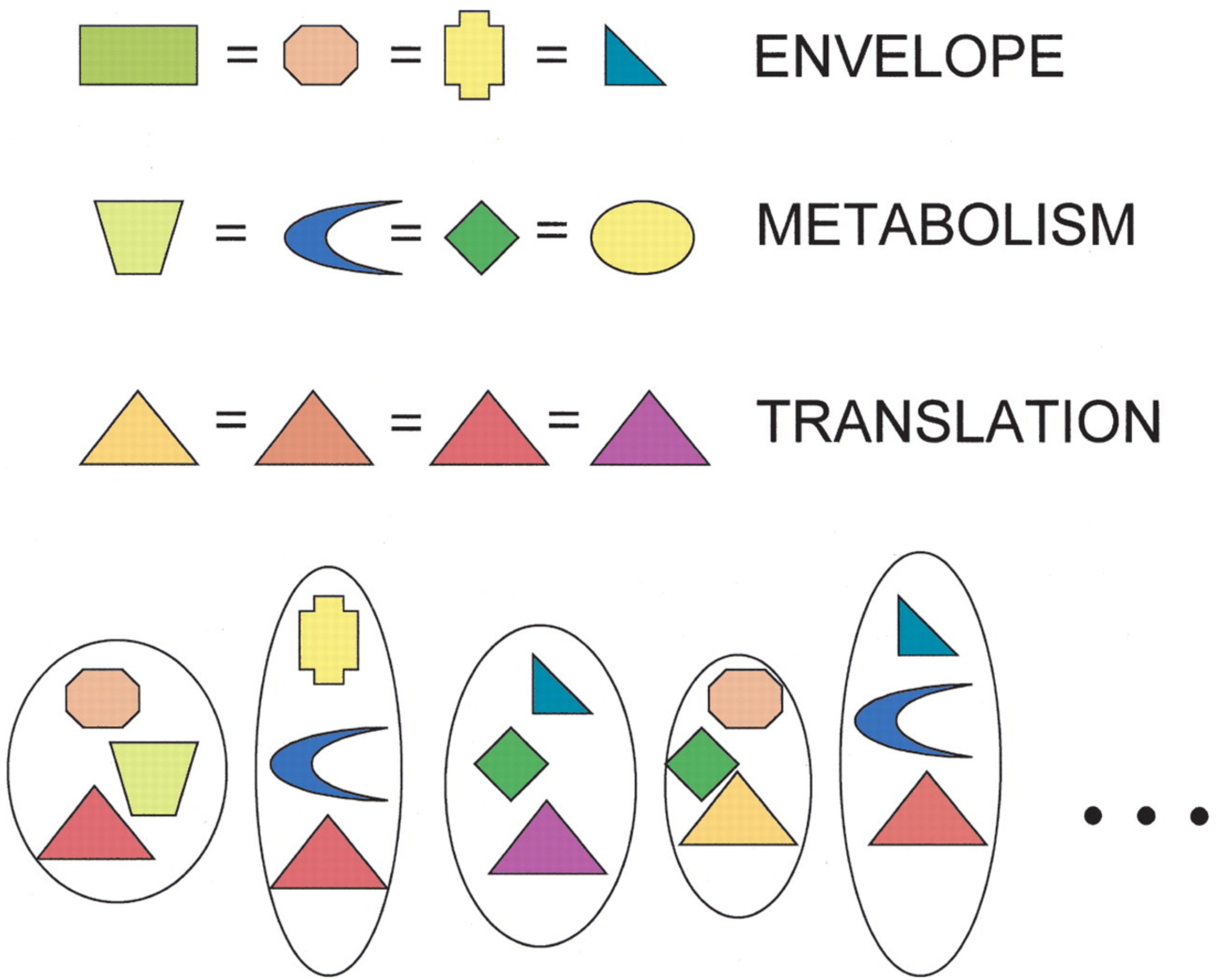
This palette of 15 colors is safe for common color blindness. Notice that each of the three 5-color groups will be perceived as a separate "color" with different tone. Two color pairs are indistinguishable by tritanopes.

IS YOUR MESSAGE NECESSARY?

http://lushlush.livejournal.com/190093.html

Is this a sign that warns you against the sign itself? When creating a figure, make sure that it doesn't merely paraphrase the text.

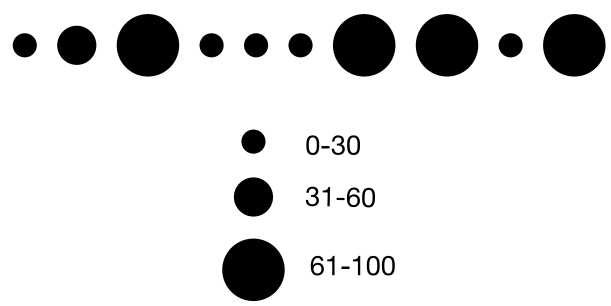
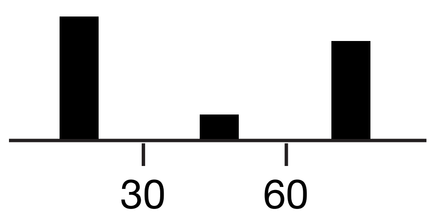
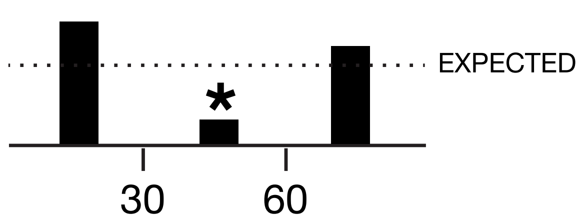
CLUES, ANYONE?



A mix-and-match model for prokaryotic genome evolution. Charlebois, R.L. and W.F. Doolittle, Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Res, 2004. 14(12): p. 2469-77.

Can you figure out what's going on here? Without reading the legend, it is very difficult to interpret this figure. Unfortunately, the concept is so simple that after reading the legend you no longer need to look at the figure.

CRAFTING A MESSAGE

WHAT IS SHOWN?		WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA	12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.
DISCRETIZED		SCALE	THREE RANGES ARE IMPORTANT. INDIVIDUAL VALUES WITHIN A RANGE ARE NOT.
BINNED		DISTRIBUTION	THERE ARE FEWER MEDIUM-SIZED VALUES.
TREND		SIGNIFICANCE	THERE ARE <u>SIGNIFICANTLY</u> FEWER MEDIUM-SIZED VALUES.

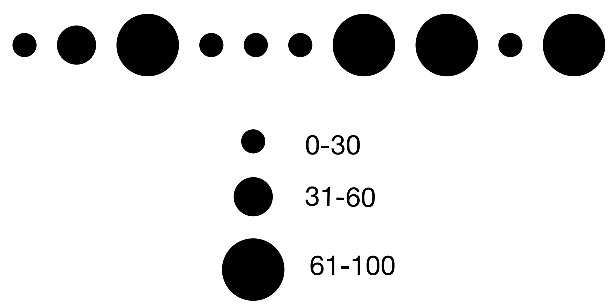
Remove and aggregate data to reduce the number of interpretations. Ideally, the figure should have a single unambiguous message.

CRAFTING A MESSAGE

WHAT IS SHOWN?		WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA	12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.

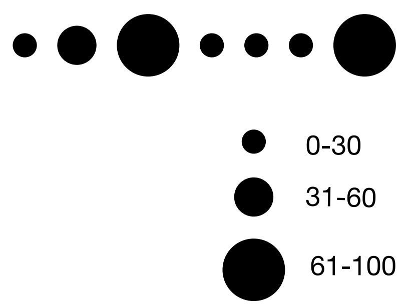
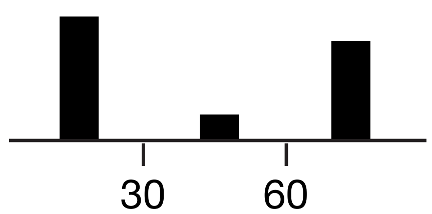
Remove and aggregate data to reduce the number of interpretations. Ideally, the figure should have a single unambiguous message.

CRAFTING A MESSAGE

	WHAT IS SHOWN?	WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA	12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.
DISCRETIZED		SCALE	THREE RANGES ARE IMPORTANT. INDIVIDUAL VALUES WITHIN A RANGE ARE NOT.

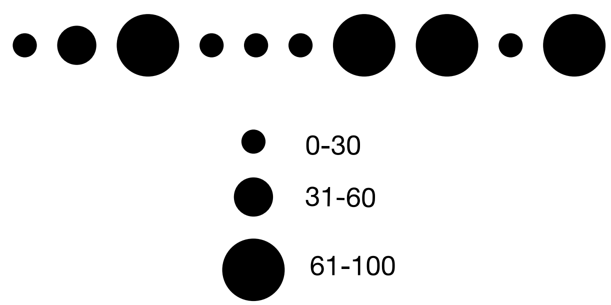
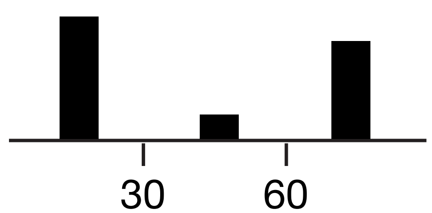
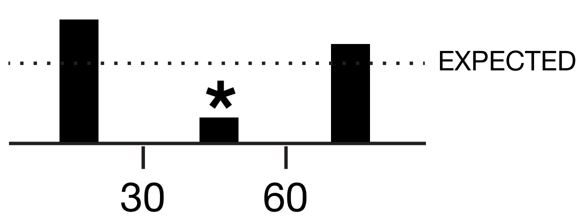
Remove and aggregate data to reduce the number of interpretations. Ideally, the figure should have a single unambiguous message.

CRAFTING A MESSAGE

	WHAT IS SHOWN?	WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA	12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.
DISCRETIZED		SCALE	THREE RANGES ARE IMPORTANT. INDIVIDUAL VALUES WITHIN A RANGE ARE NOT.
BINNED		DISTRIBUTION	THERE ARE FEWER MEDIUM-SIZED VALUES.

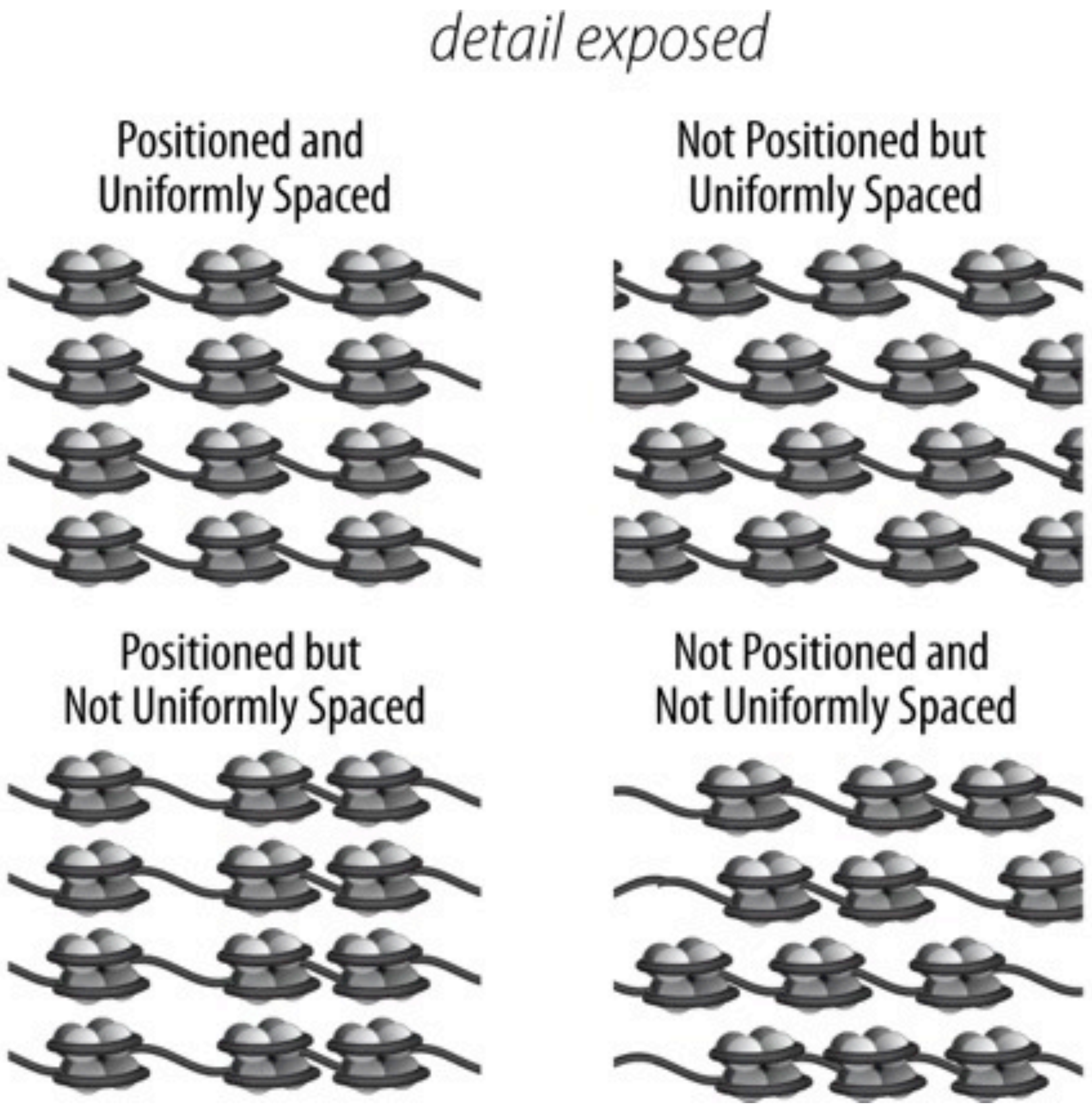
Remove and aggregate data to reduce the number of interpretations. Ideally, the figure should have a single unambiguous message.

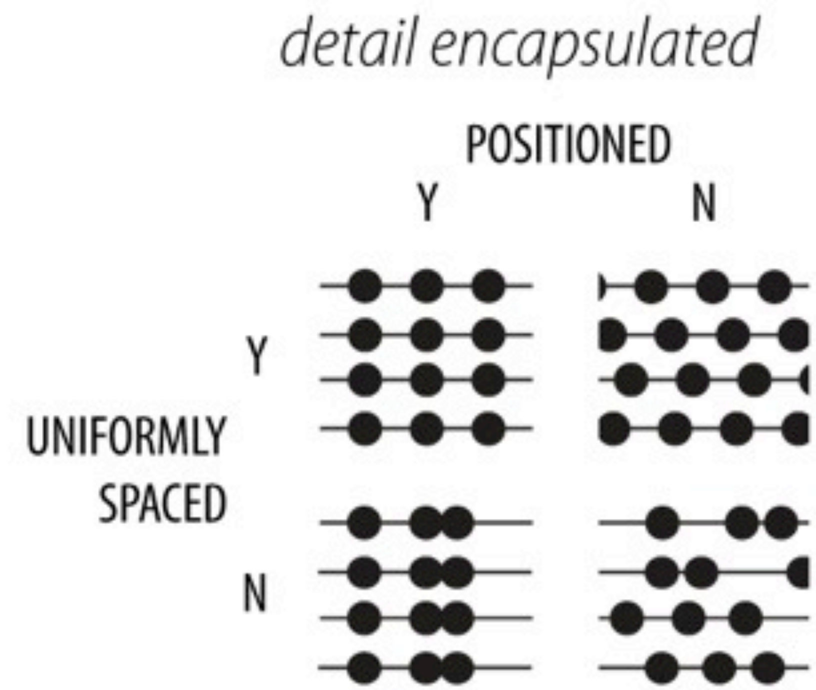
CRAFTING A MESSAGE

WHAT IS SHOWN?		WHAT IS COMMUNICATED?	WHAT IS INTERPRETED?
RAW DATA	12 54 82 29 25 22 67 61 23 79	NO CLEAR MESSAGE.	UNKNOWN. READER IS ON THEIR OWN.
DISCRETIZED		SCALE	THREE RANGES ARE IMPORTANT. INDIVIDUAL VALUES WITHIN A RANGE ARE NOT.
BINNED		DISTRIBUTION	THERE ARE FEWER MEDIUM-SIZED VALUES.
TREND		SIGNIFICANCE	THERE ARE SIGNIFICANTLY FEWER MEDIUM-SIZED VALUES.

Remove and aggregate data to reduce the number of interpretations. Ideally, the figure should have a single unambiguous message.

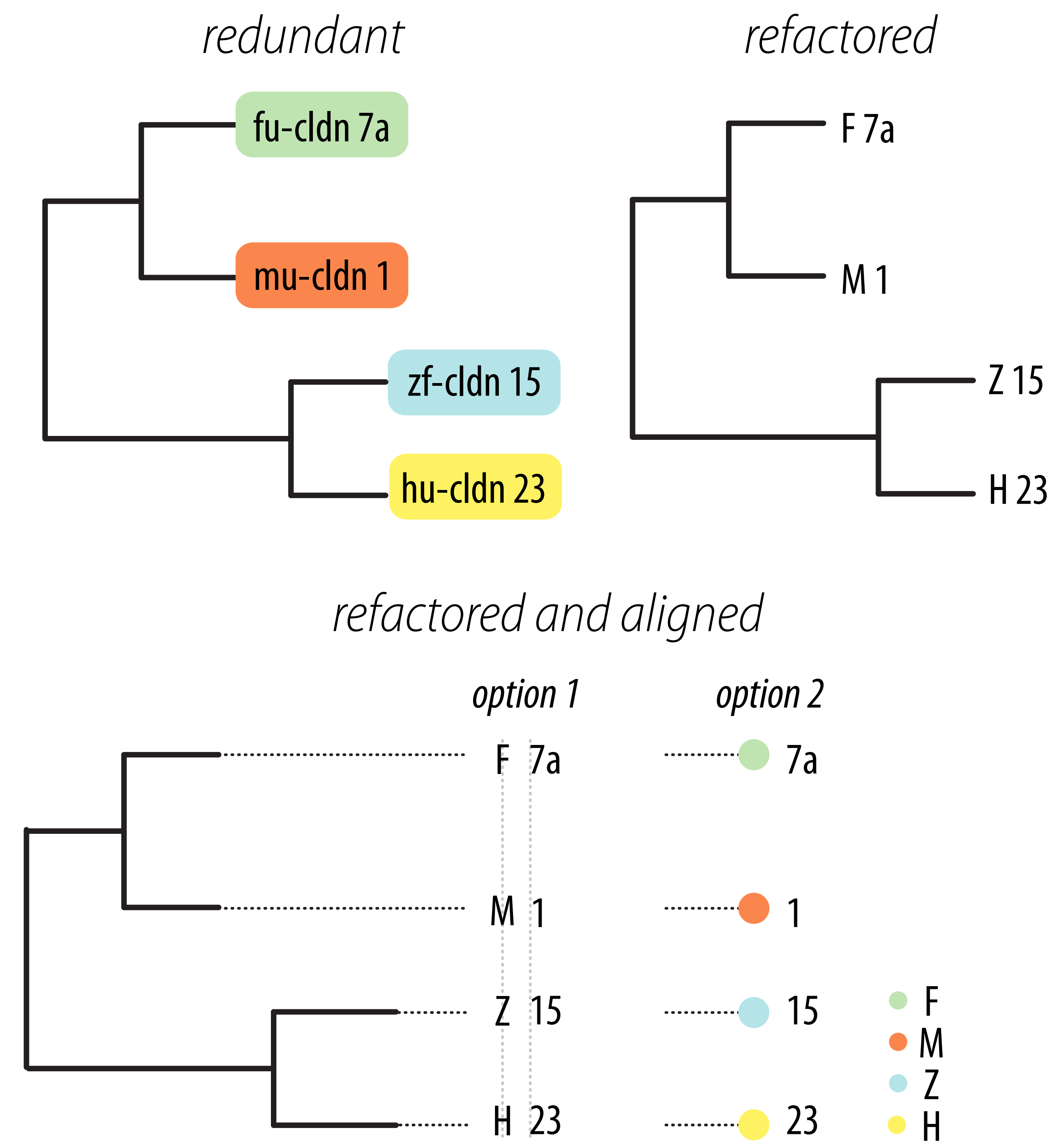
ENCAPSULATION





Reduce unnecessary detail.

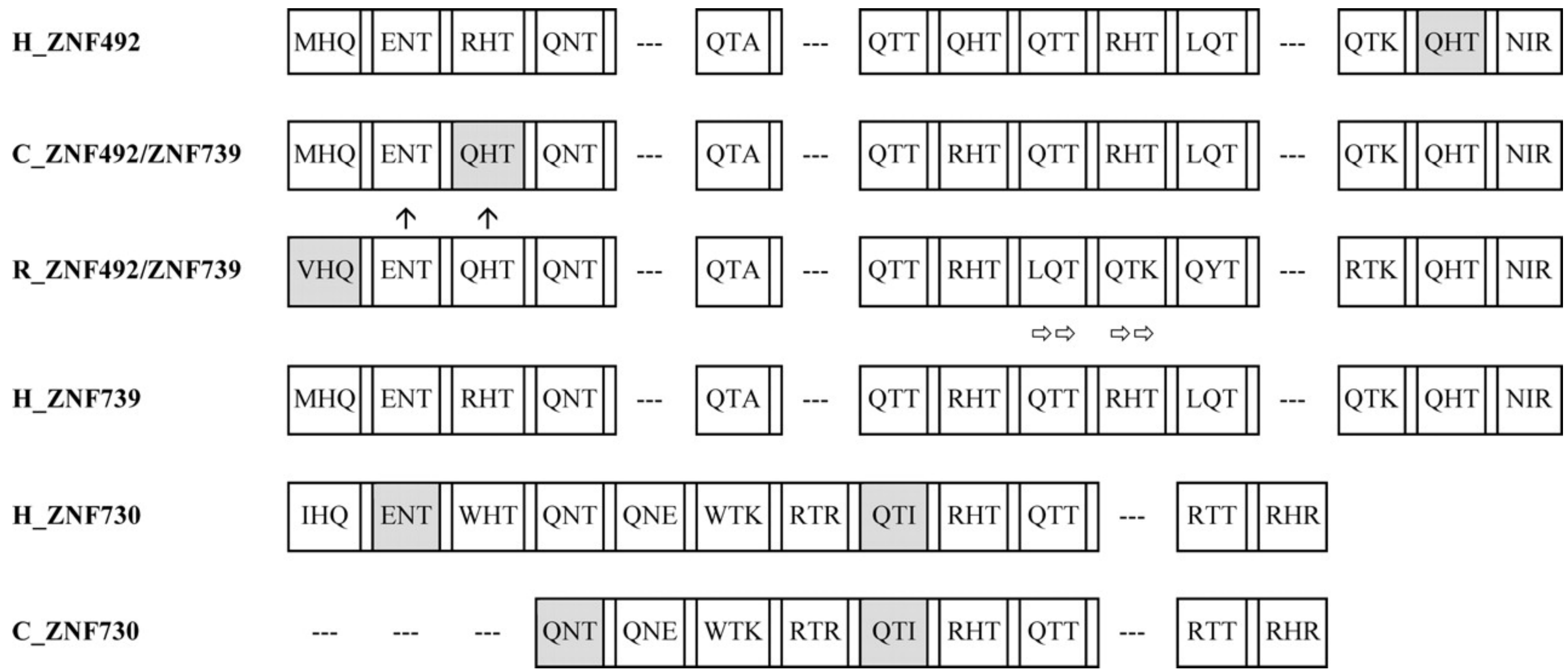
REDUNDANCY

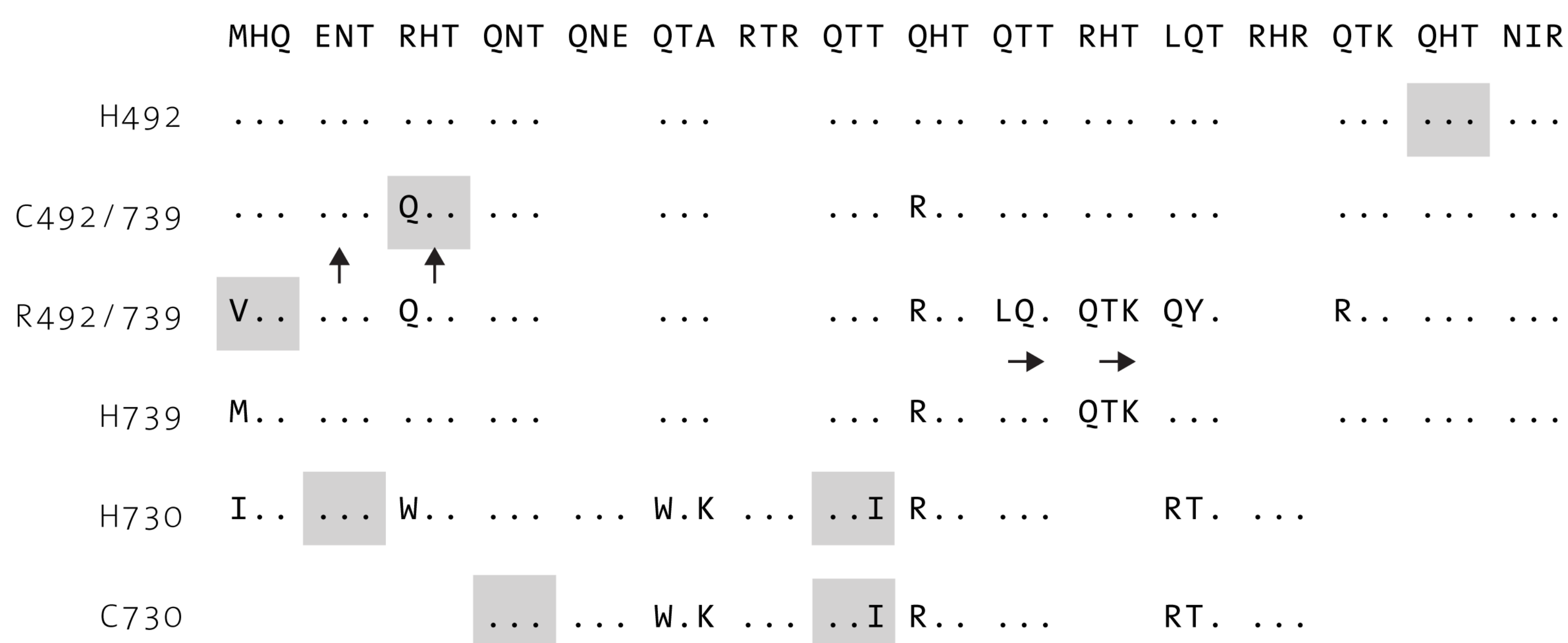


Y. H. Loh, A. Christoffels, S. Brenner, W. Hunziker, B. Venkatesh, Genome Res 14, 1248 (Jul, 2004).

Text labels should be shortened to emphasize uniqueness. In this case, each label shares the “-cldn” suffix which does not need to be repeated. The reader will find aligned text easier to compare.

EXCESS INK

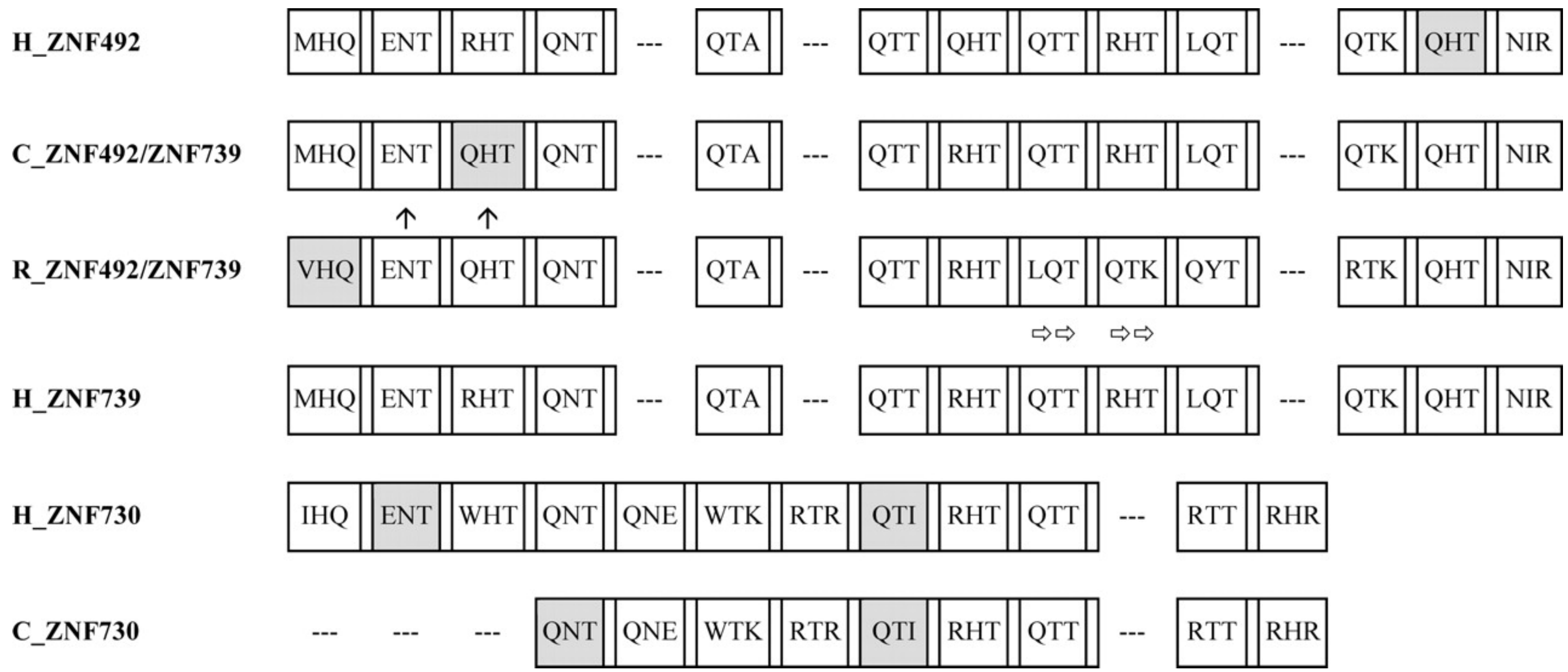




Zinc finger exon analysis for ZNF493 and ZNF738, two divergent genes from the ZNF431 clade. Hamilton, A.T., et al., Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. Genome Res, 2006. 16(5): p. 584-94.

When differences are important, emphasize what is changing by showing the reference data only once.

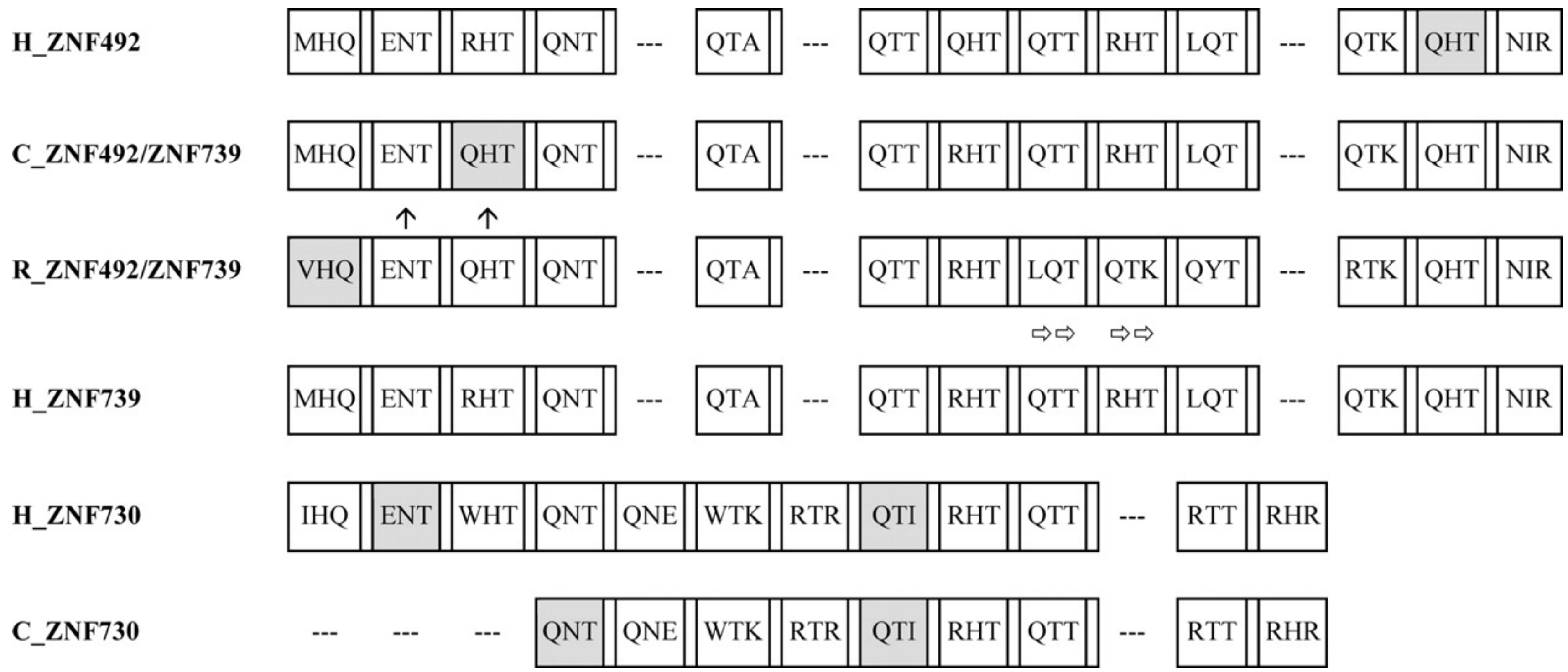
EXCESS INK

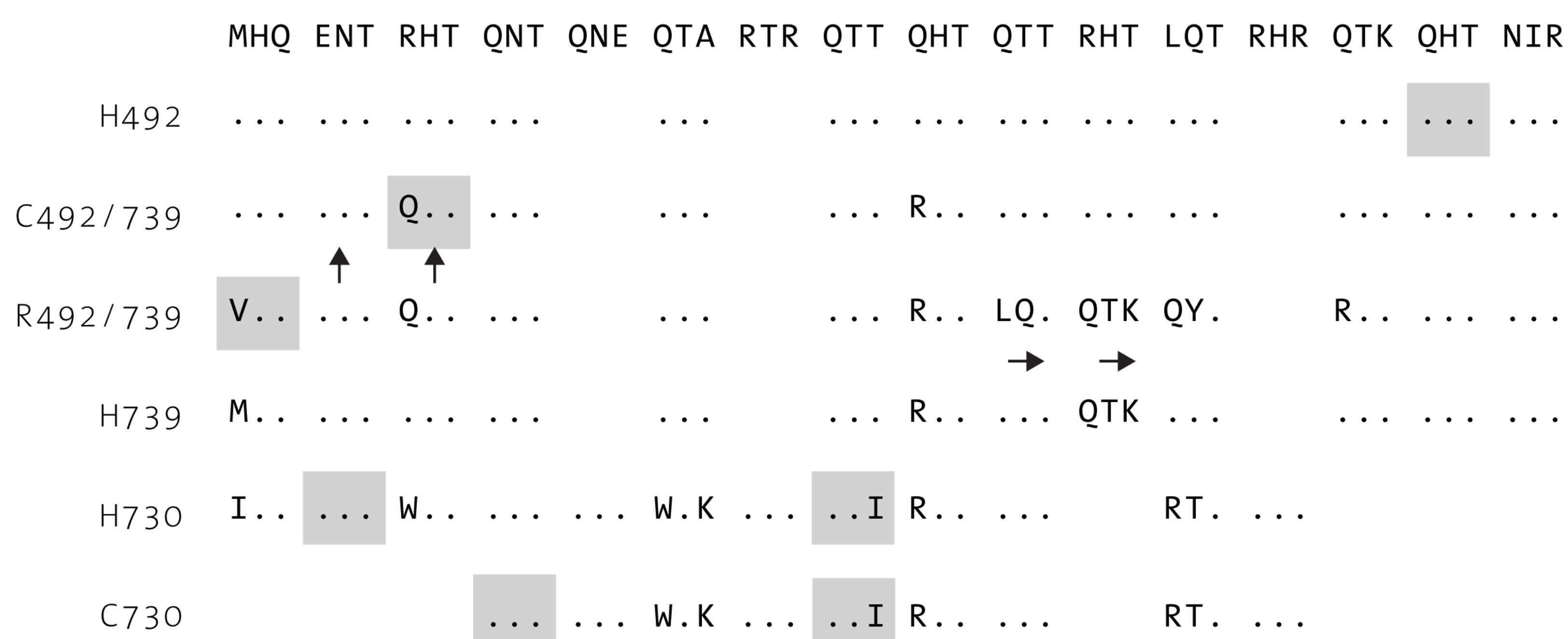


Zinc finger exon analysis for ZNF493 and ZNF738, two divergent genes from the ZNF431 clade. Hamilton, A.T., et al.,
Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. Genome Res, 2006. 16(5): p. 584-94.

When differences are important, emphasize what is changing by showing the reference data only once.

EXCESS INK

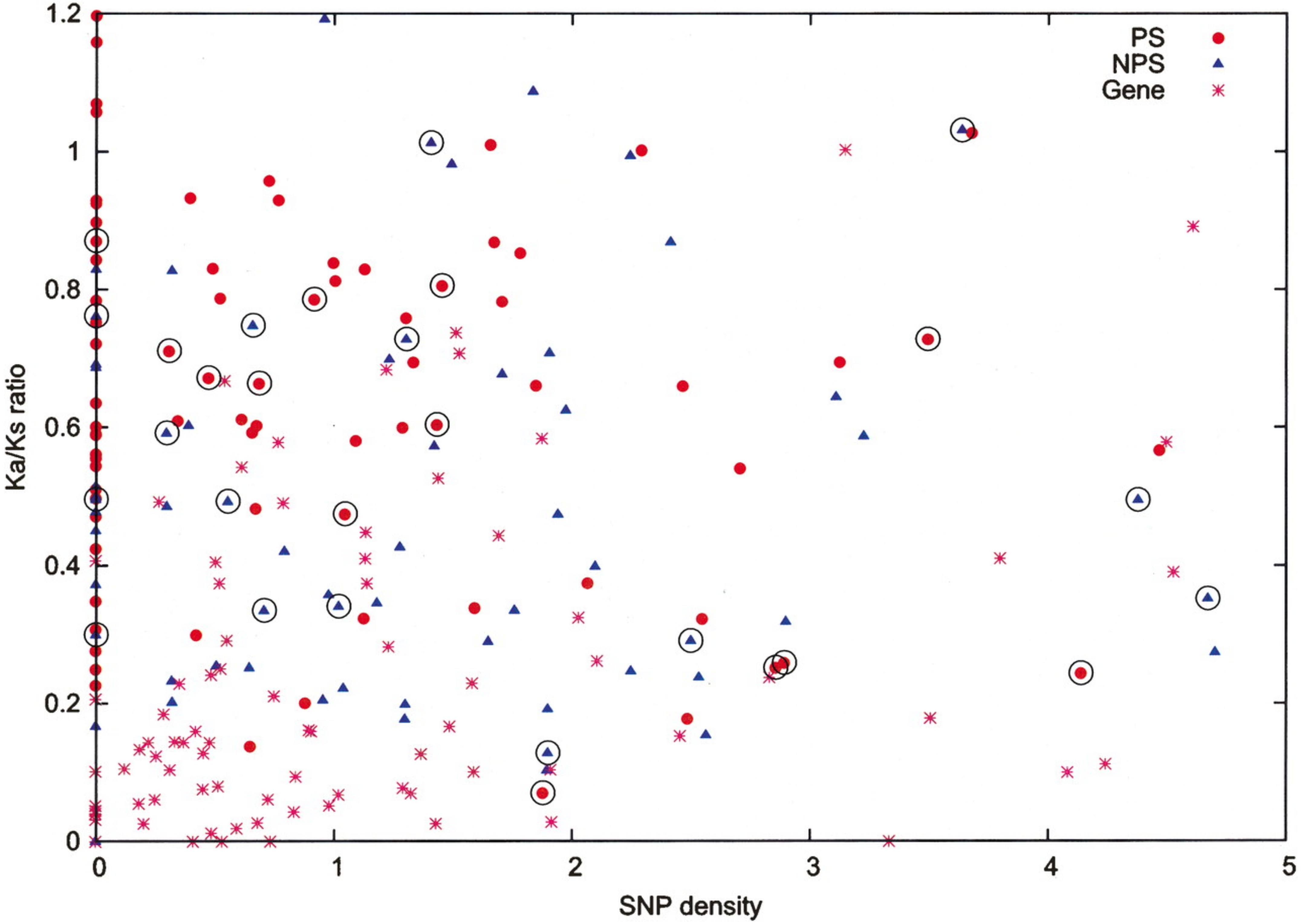




Zinc finger exon analysis for ZNF493 and ZNF738, two divergent genes from the ZNF431 clade. Hamilton, A.T., et al., Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. Genome Res, 2006. 16(5): p. 584-94.

When differences are important, emphasize what is changing by showing the reference data only once.

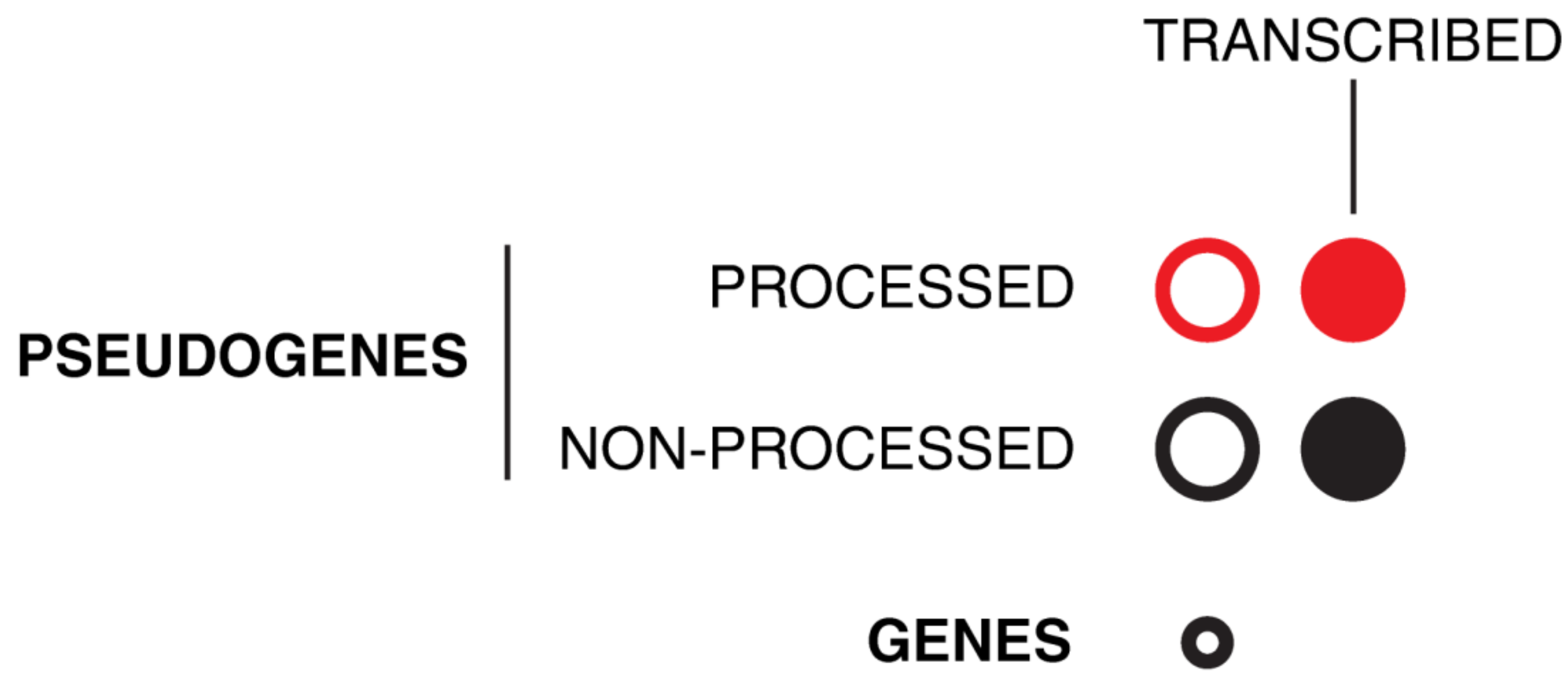
THE GLYPHS ARE CUES



Comparison of Ka/Ks ratio and SNP density for genes and pseudogenes. Zheng, D., et al., Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res, 2007. 17(6): p. 839-51.

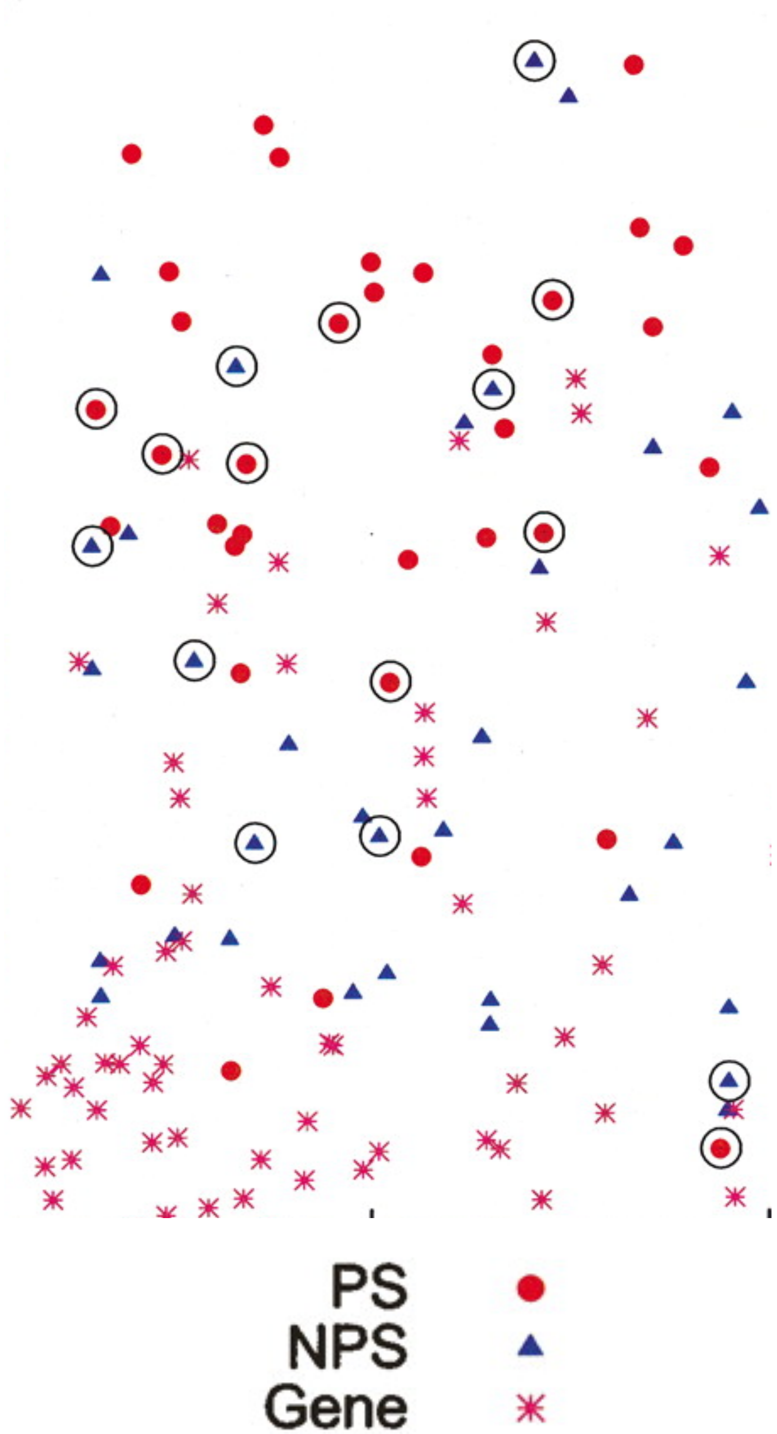
The symbols used here do not follow any meaningful hierarchy, despite the fact that their meaning is related.

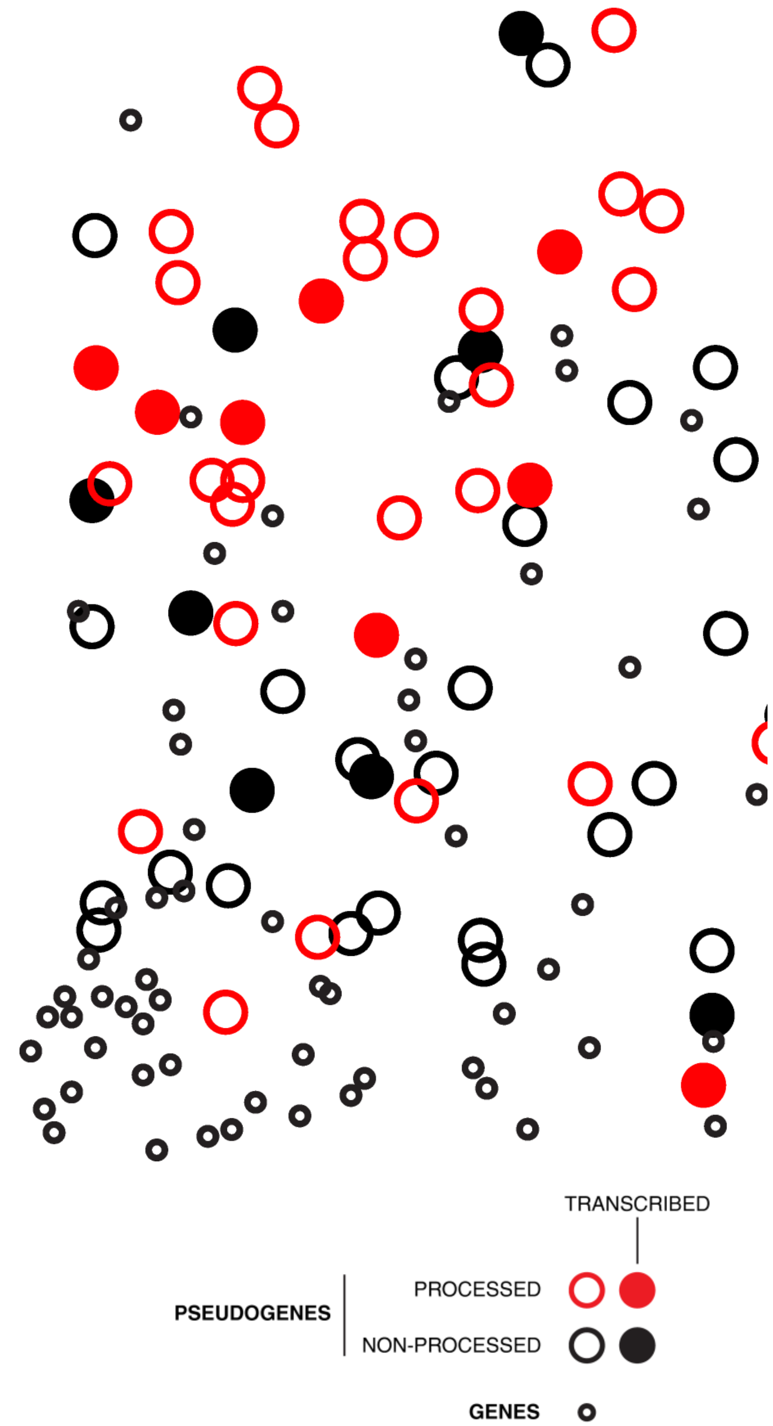
HIERARCHICAL GLYPHS



By varying shape and color meaningfully, a more memorable encoding is achieved.

HIERARCHICAL GLYPHS

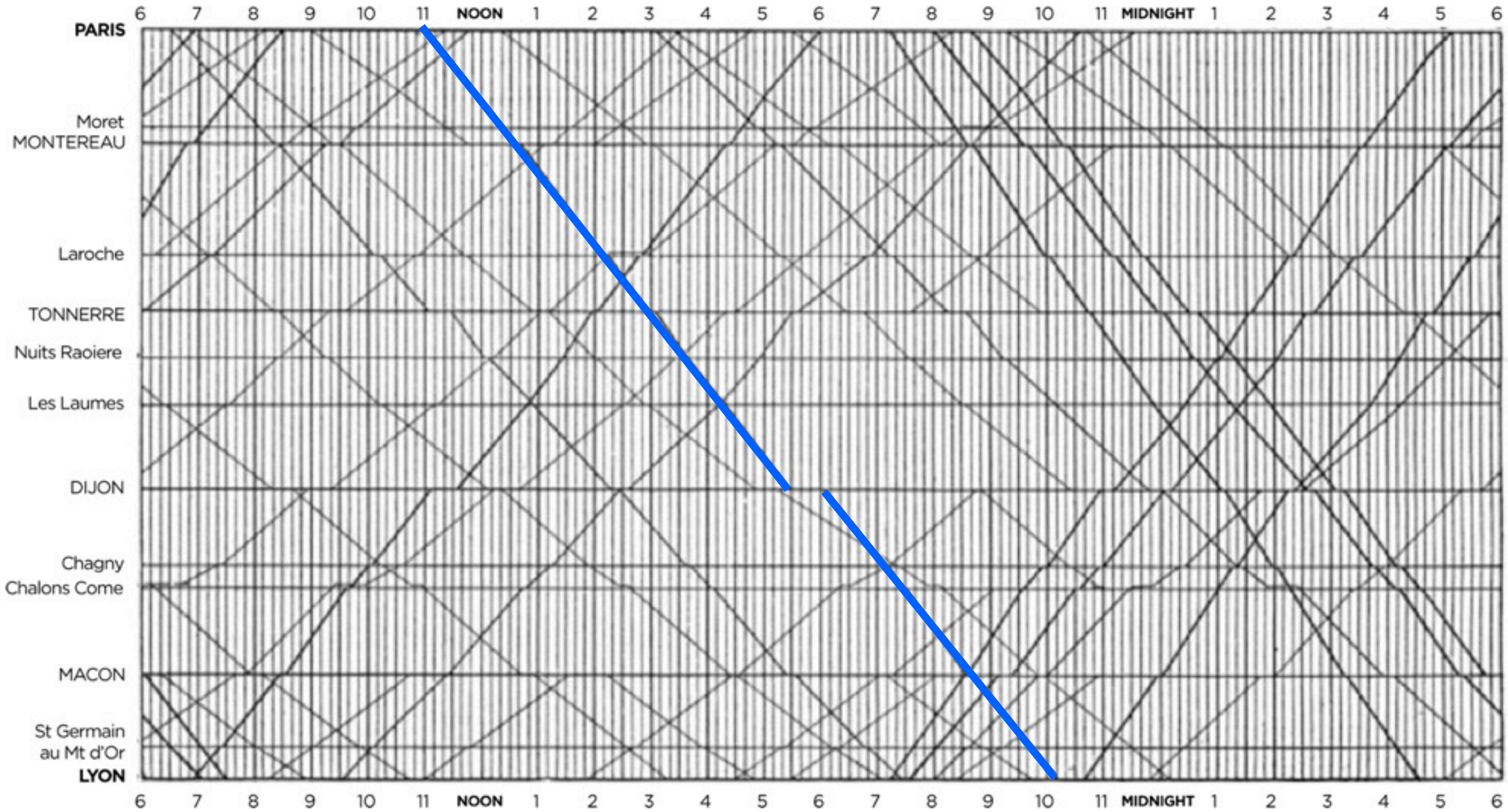




Comparison of Ka/Ks ratio and SNP density for genes and pseudogenes. Zheng, D., et al., Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res, 2007. 17(6): p. 839-51.

Try focusing on the blue circled triangles on the left. It's very difficult to keep them separated from the circled circles. This is Gestalt similarity principle at work - these encodings are too similar to be quickly differentiated into distinct groups. On the right, these are trivially separated (the filled red and black circles).

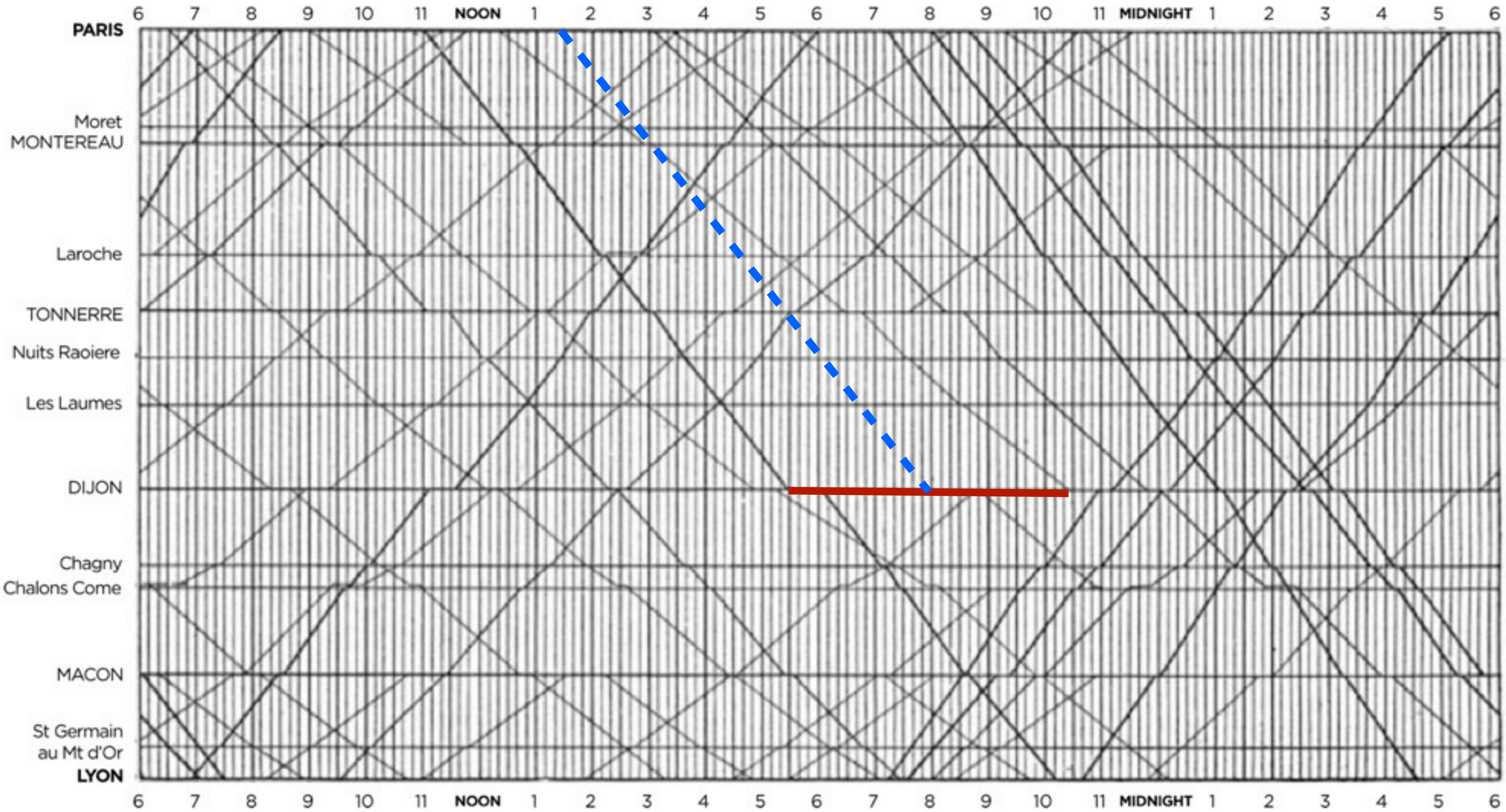
APPROPRIATE REPRESENTATION



Tufte E (1992) Visual Display of Quantitative Information: Graphics Press.
see *http://c82.net/posts.php?id=66*

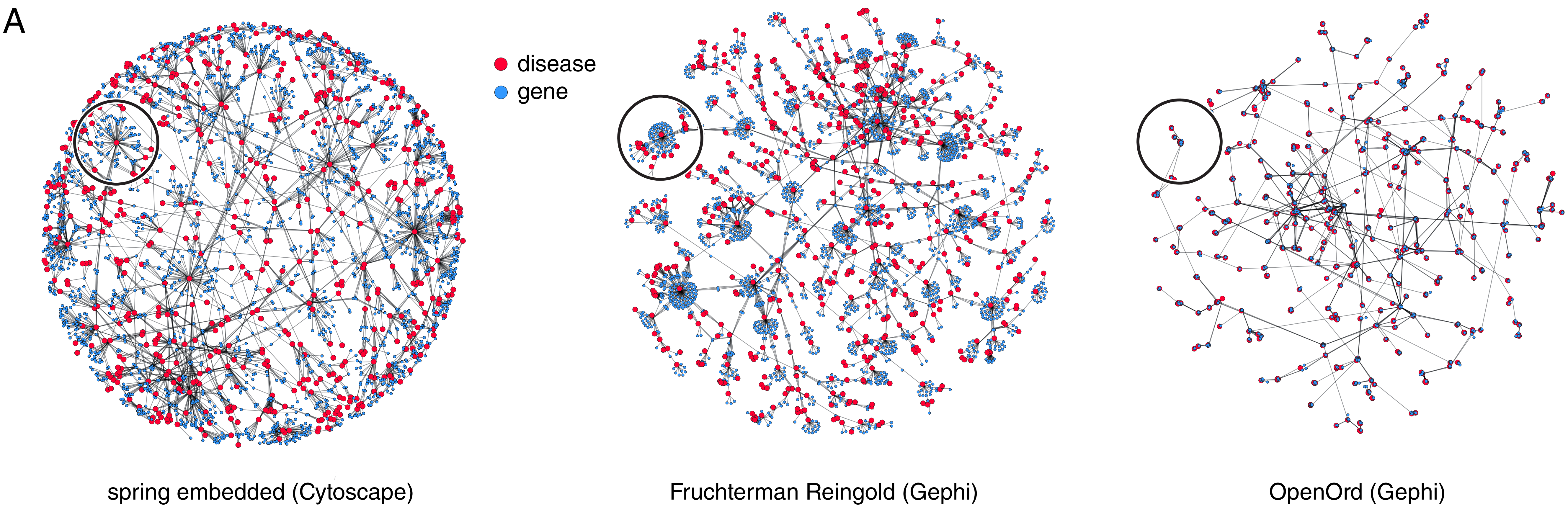
By representing a train schedule as a series of lines, it is trivial to plan a trip. For example, if you wish to get from Paris to Lyon and leave after 10am, the train choice is clear. Notice how the 30 minute layover in Dijon is emphasized by this format. If you knew that the distances between the towns on the right was proportional to their spacing on the figure, you could evaluate the relative differences in train speeds.

APPROPRIATE REPRESENTATION



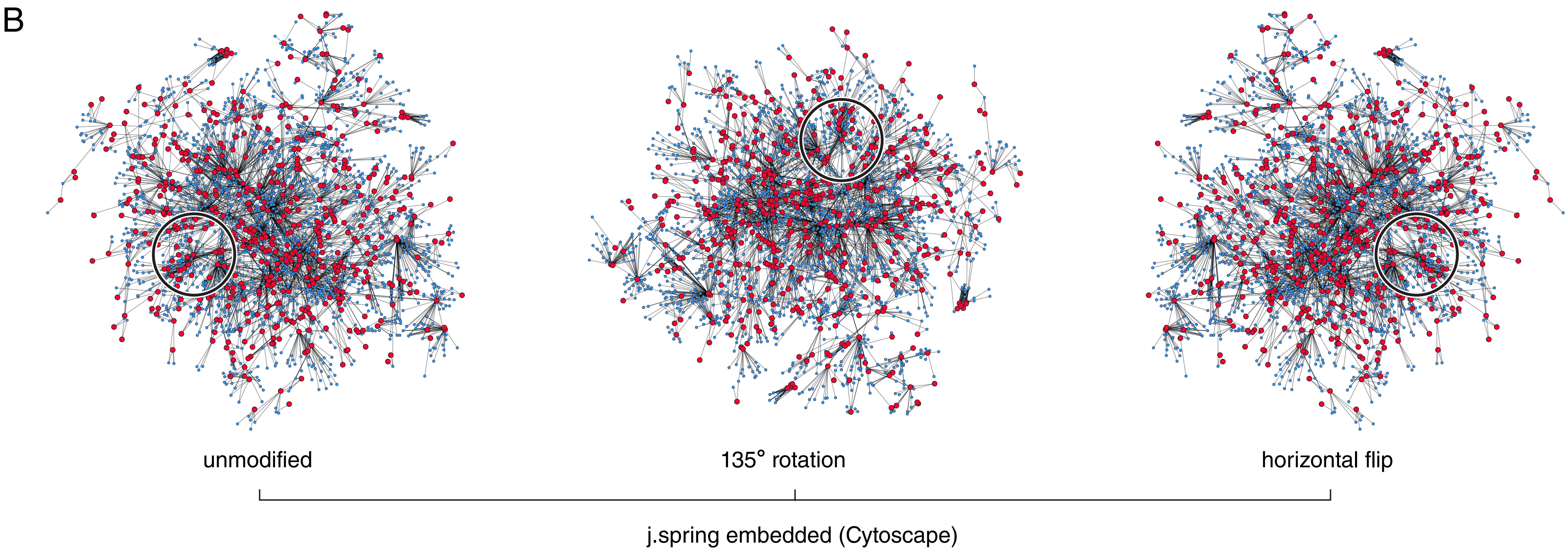
Consider the task of using a standard tabular schedule to answer the following question. Are there any blackout periods of service between Paris and Dijon? Simply looking at the lines the answer is quick to glean - there is a 5 hour period (5-10pm) in which no trains from Paris arrive at Dijon. By installing a train that leaves Paris at 1:30, we can offer riders a useful alterantive.

TROUBLE WITH NETWORK LAYOUTS



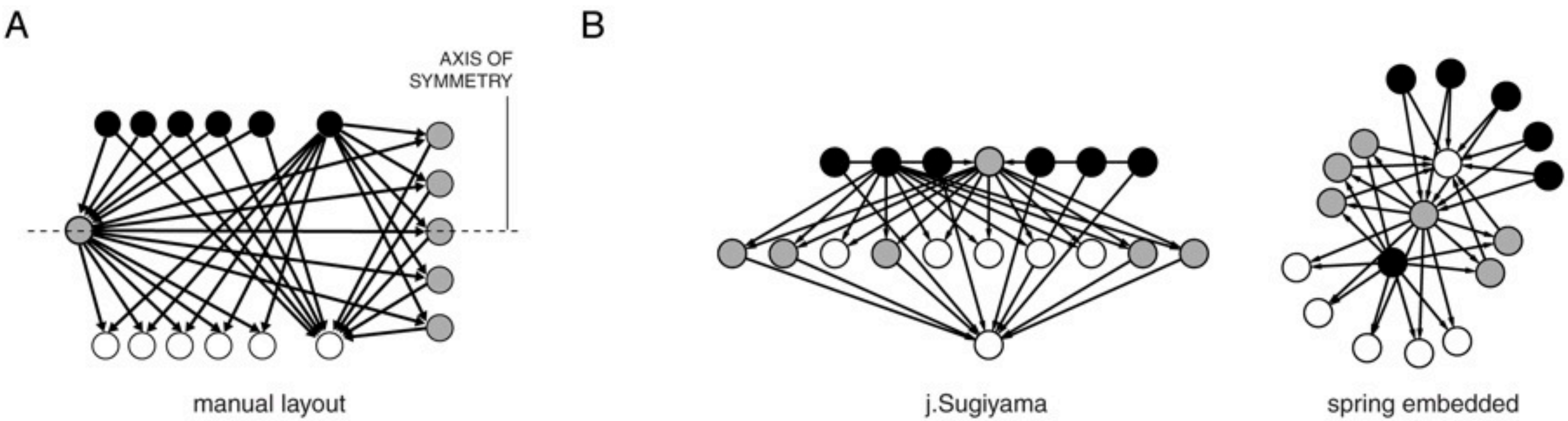
Network layouts can present very different views, depending on the algorithm. Familiarity with the layout mechanism is important for interpreting patterns. Unfortunately, the algorithms are often heuristic and complicated - their behaviour is difficult to understand.

SAME NETWORK?



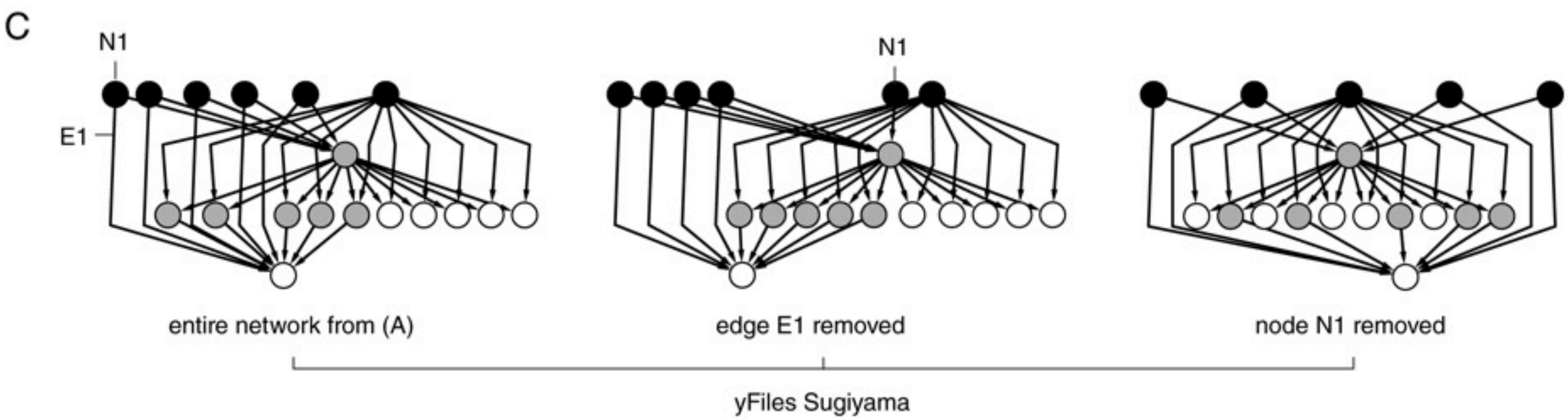
Even the same layout, when rotated or flipped, can appear as a different network.

PATTERNS ARE HIDDEN



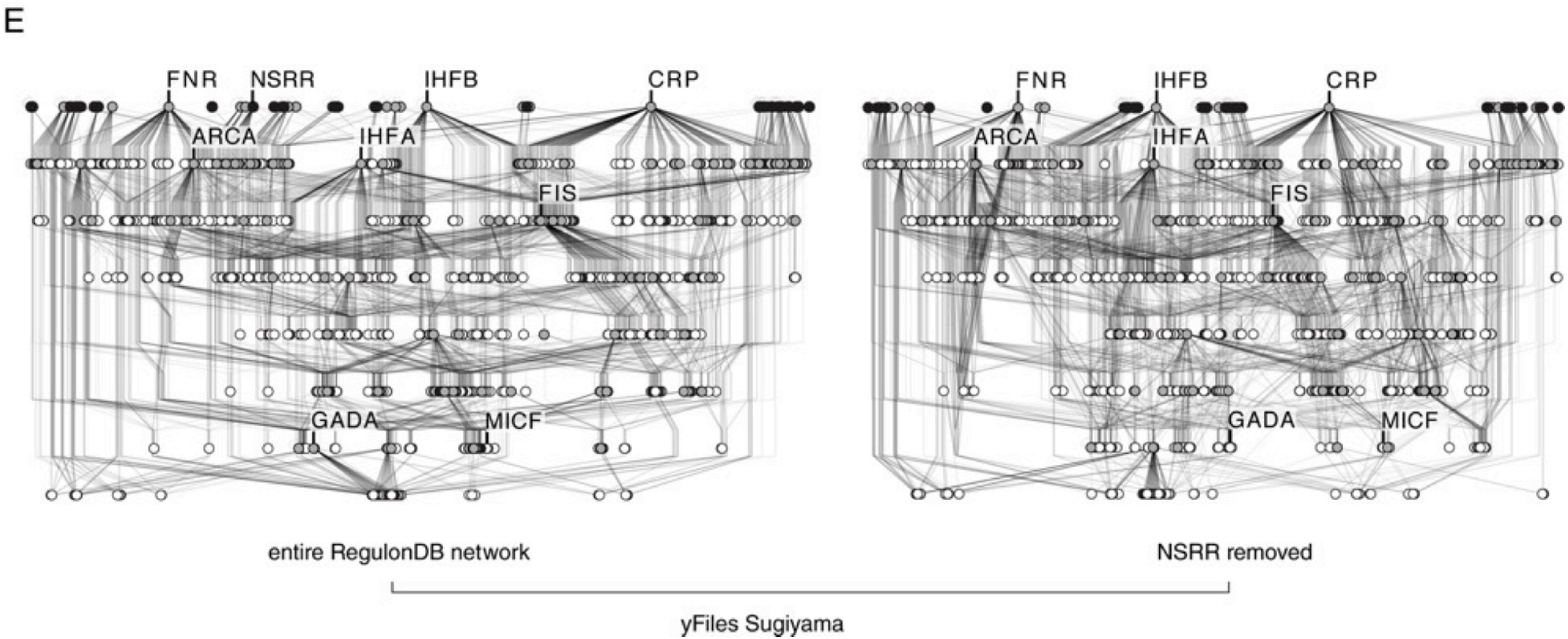
Patterns in very simple networks are commonly hidden in automated layouts.

NOT PERCEPTUALLY UNIFORM!



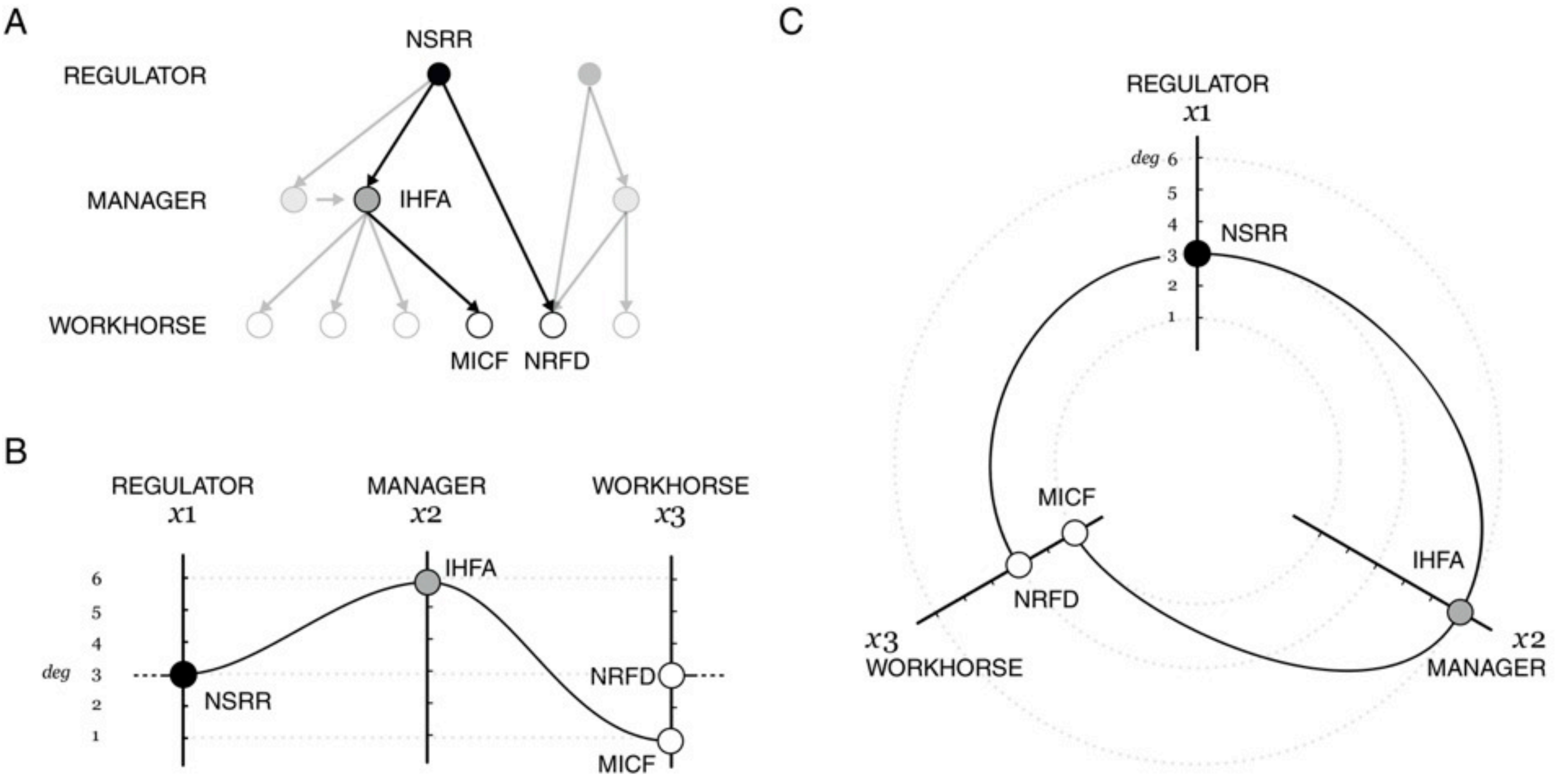
Just like color spaces that are not perceptually uniform, neither are automated layouts. A minor change in the network, here the deletion of an edge or node, results in a significantly different layout. It's very hard to tell what has changed.

WHAT HAS CHANGED?



When the network is large, the removal of a single node can have an enormous effect on the layout.

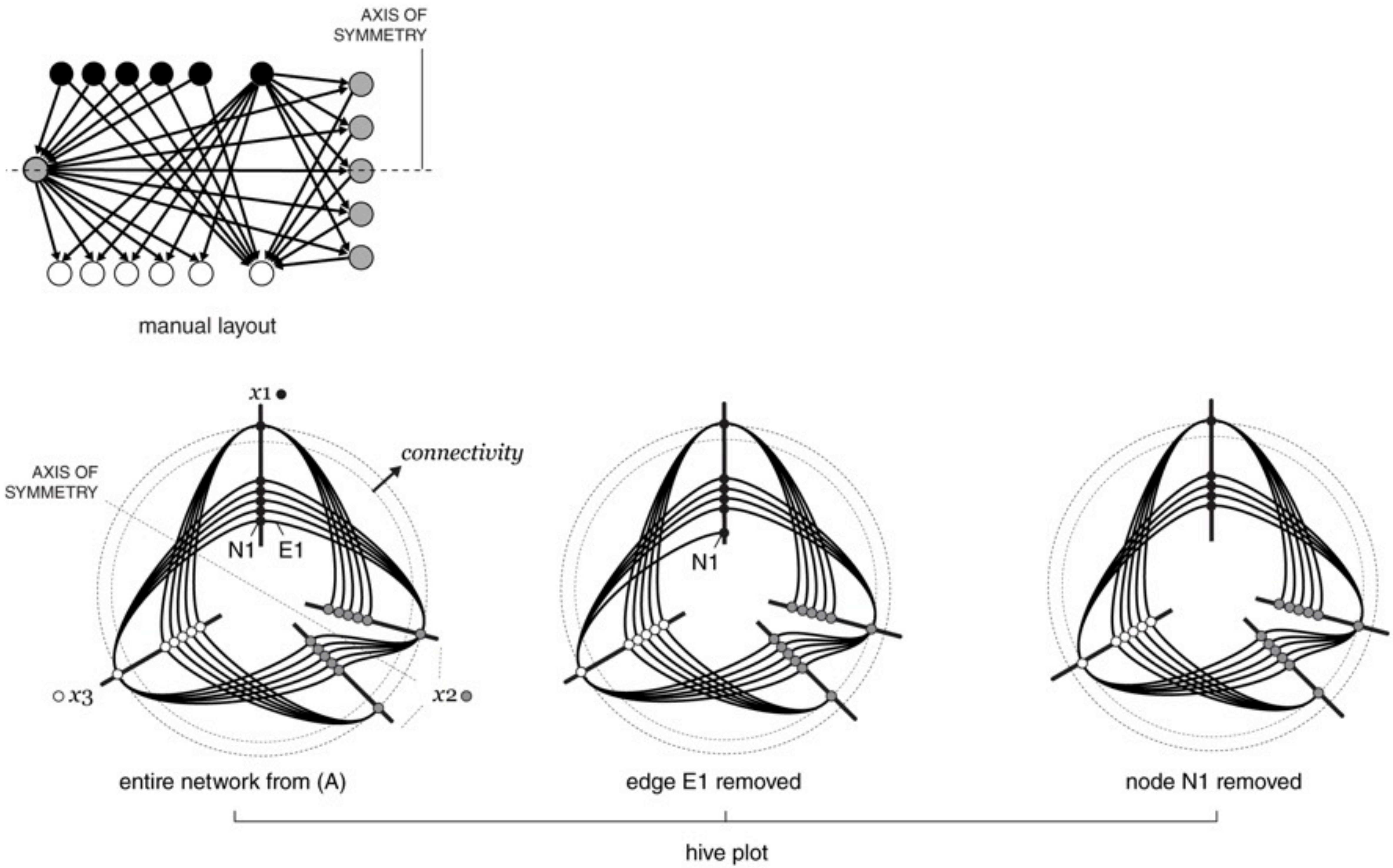
CONSTRUCTING A HIVE PLOT



Krzywinski M, Birol I, Jones S, Marra M (2011). Hive Plots — Rational Approach to Visualizing Networks. Briefings in Bioinformatics (early access 9 December 2011, doi: 10.1093/bib/bbr069)

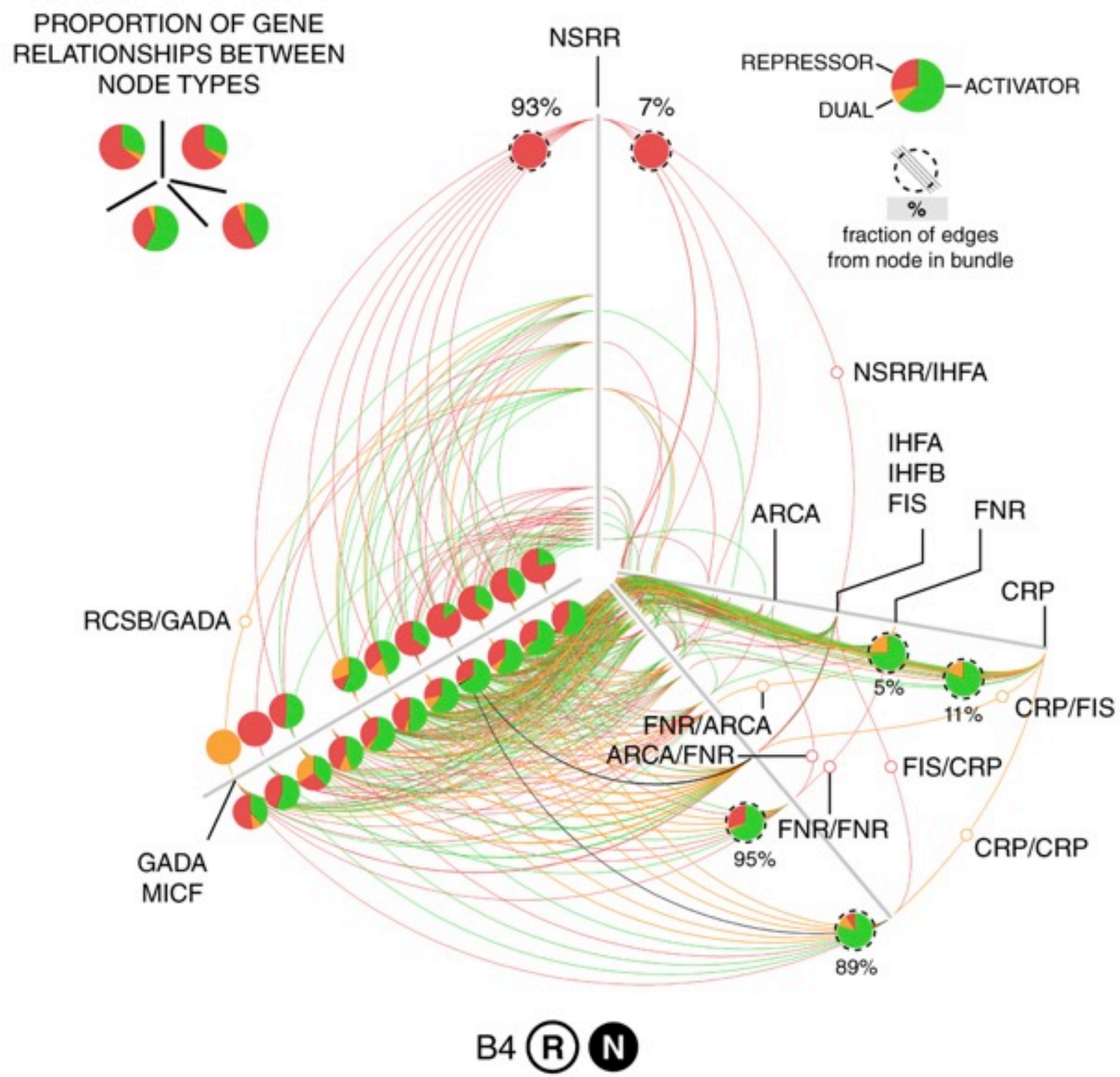
Hive plots address these issues by providing a rational and quantitative way of drawing networks.

HIVE PLOT IS PERCEPTUALLY UNIFORM



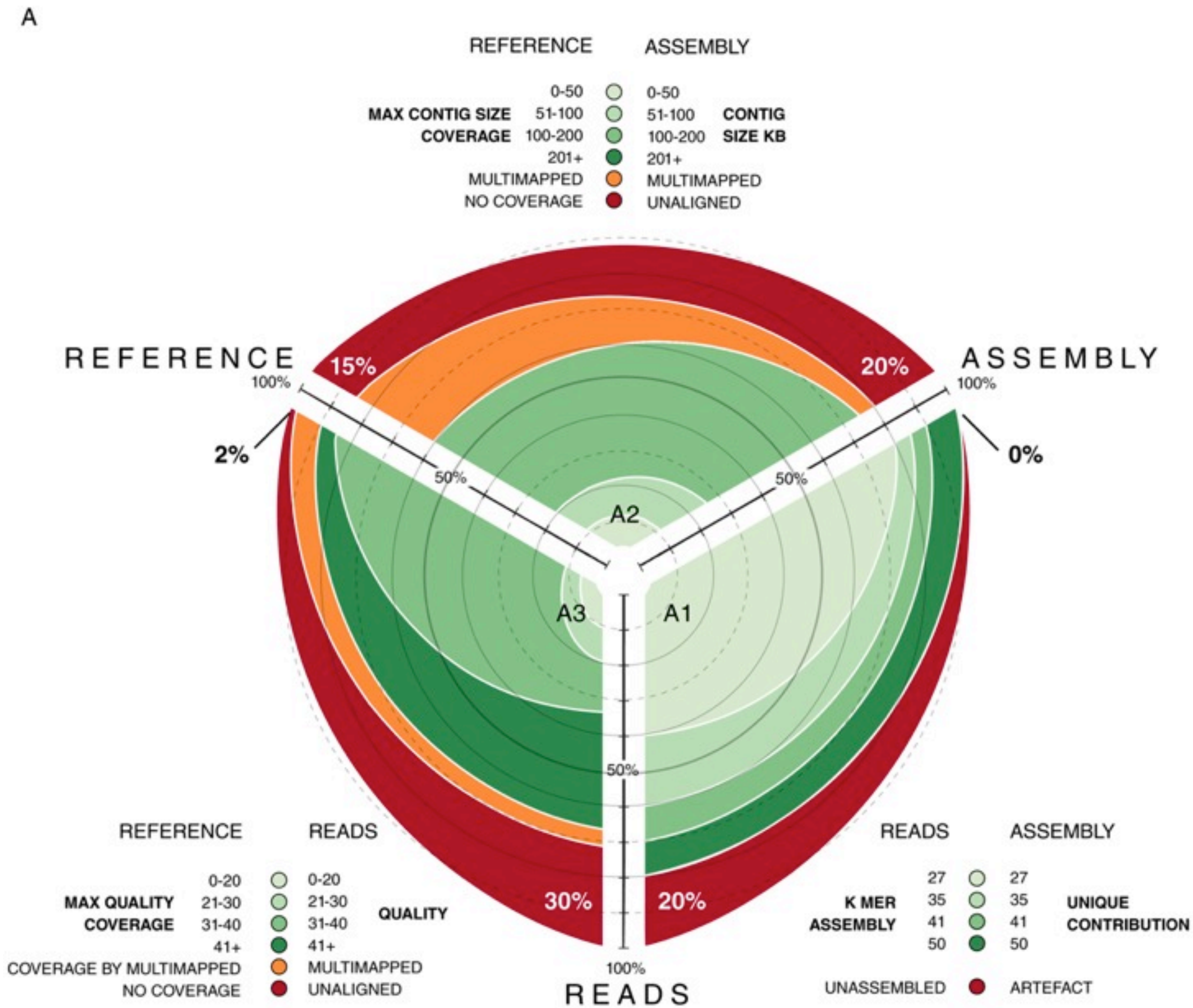
Hive plots are perceptually uniform. Changes in the hive plot are proportional the changes in the underlying data.

E COLI REGULATORY NETWORK



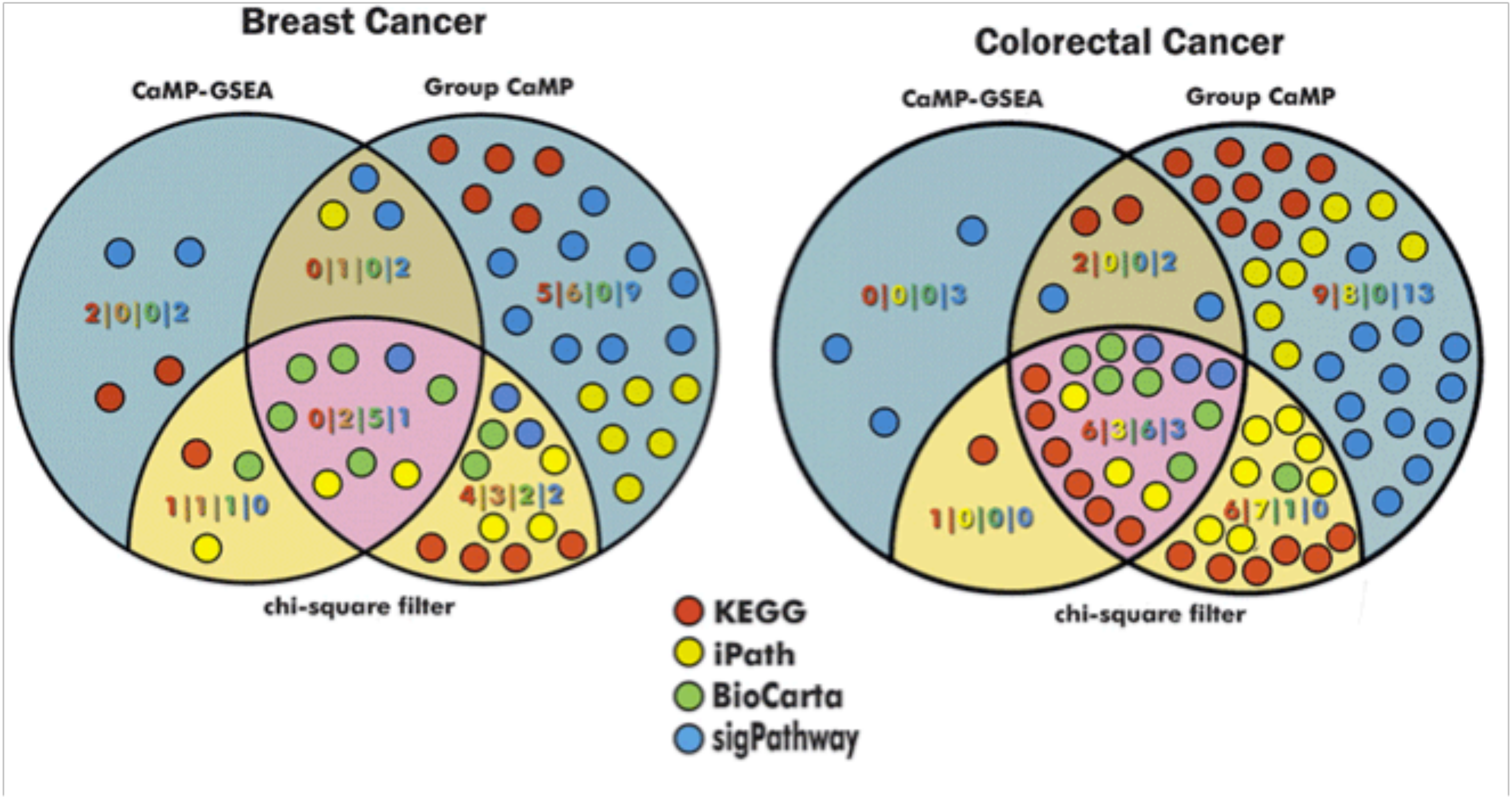
Recall the Sugiyama layout of RegulonDB in the “What Has Changed?” slide. Noticing that NSRR was removed would be trivial to detect from the hive plot. Also notice here that you can identify the following characteristics, of which none can be gleaned from any automated layout: (1) all edges from NSRR are red (2) NSRR does not connect to the two most connected managers (FNR, CRP) (3) connectivity between highly connected managers is repressive (red, orange).

HIVE PLOT FOR RATIOS



When connections between axes are drawn as ribbons, you can effectively use hive plots to illustrate differences in ratios.

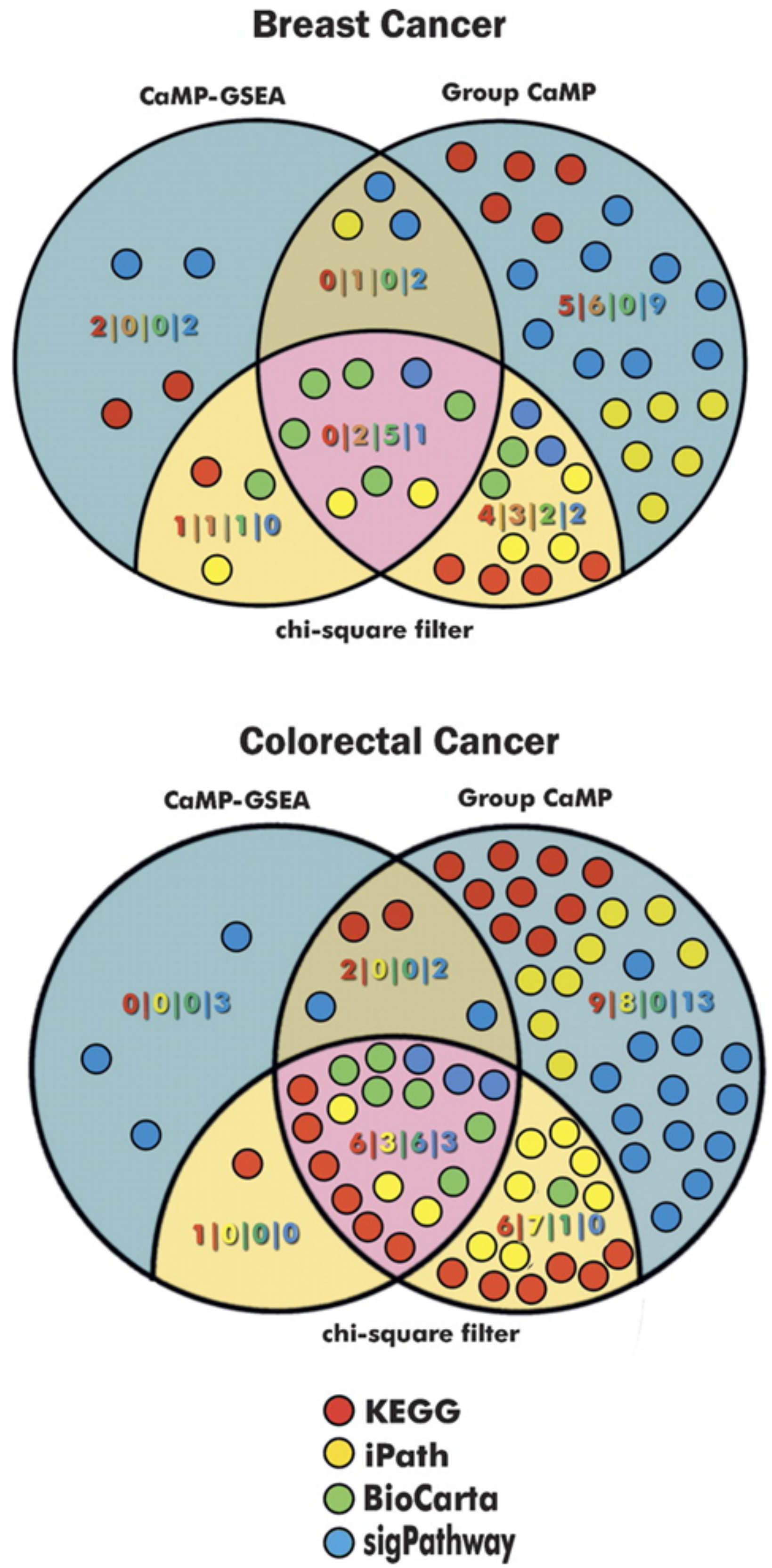
PARSABILITY LIMITS



Comparison of mutation enrichment in cellular pathways using complementary statistical approaches. Lin, J., et al.,
A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.

An impossible figure to interpret. Where are the patterns? The use of color is not intuitive - blues do not mix to give brown, yellow and purple simultaneously. The counts of the individual points are not consistently colored (orange numbers enumerate yellow points). Inconsistent spacing and multitude of numbers places too much cognitive load on the reader.

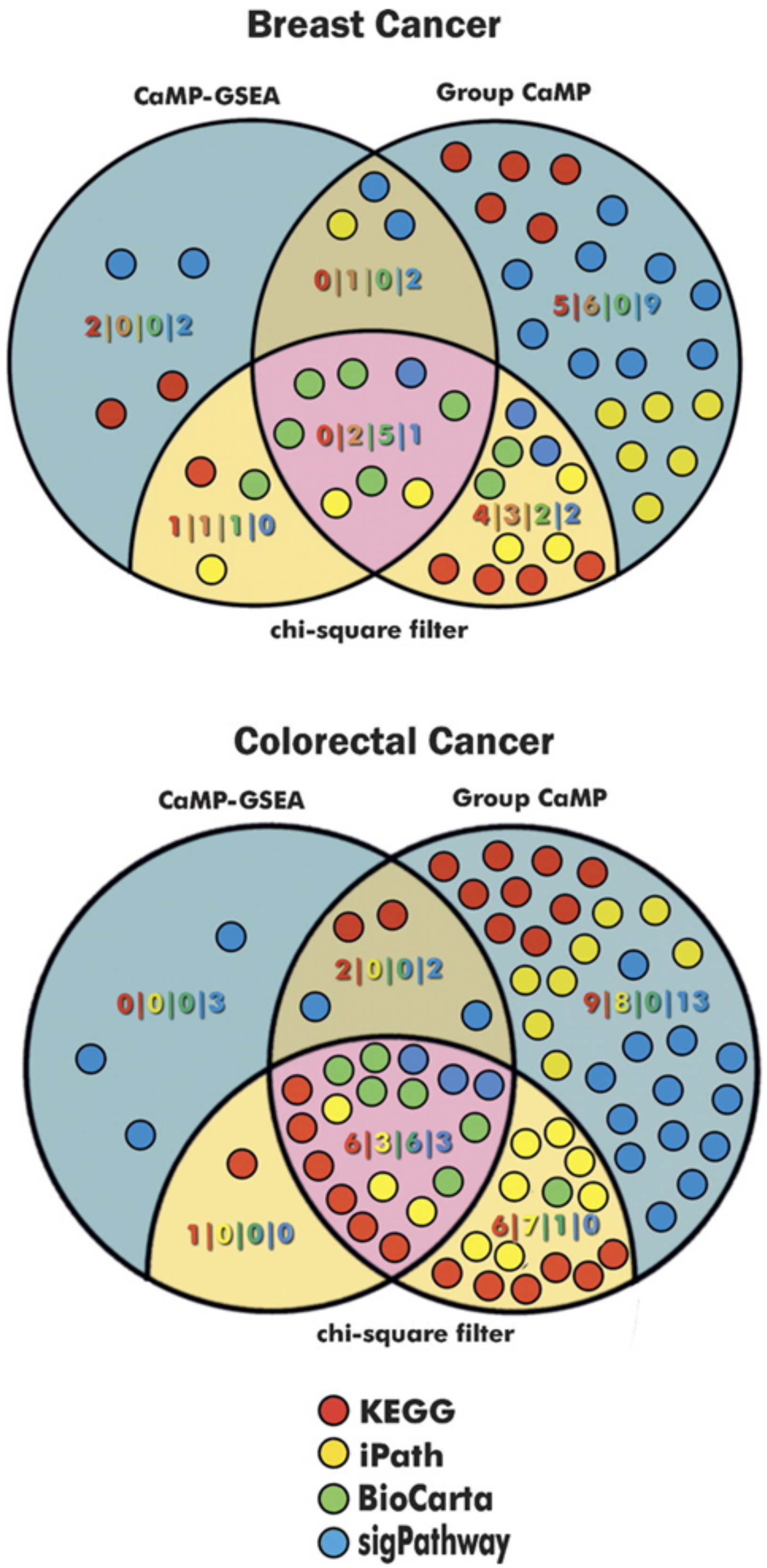
REFACTORING COMPLEXITY

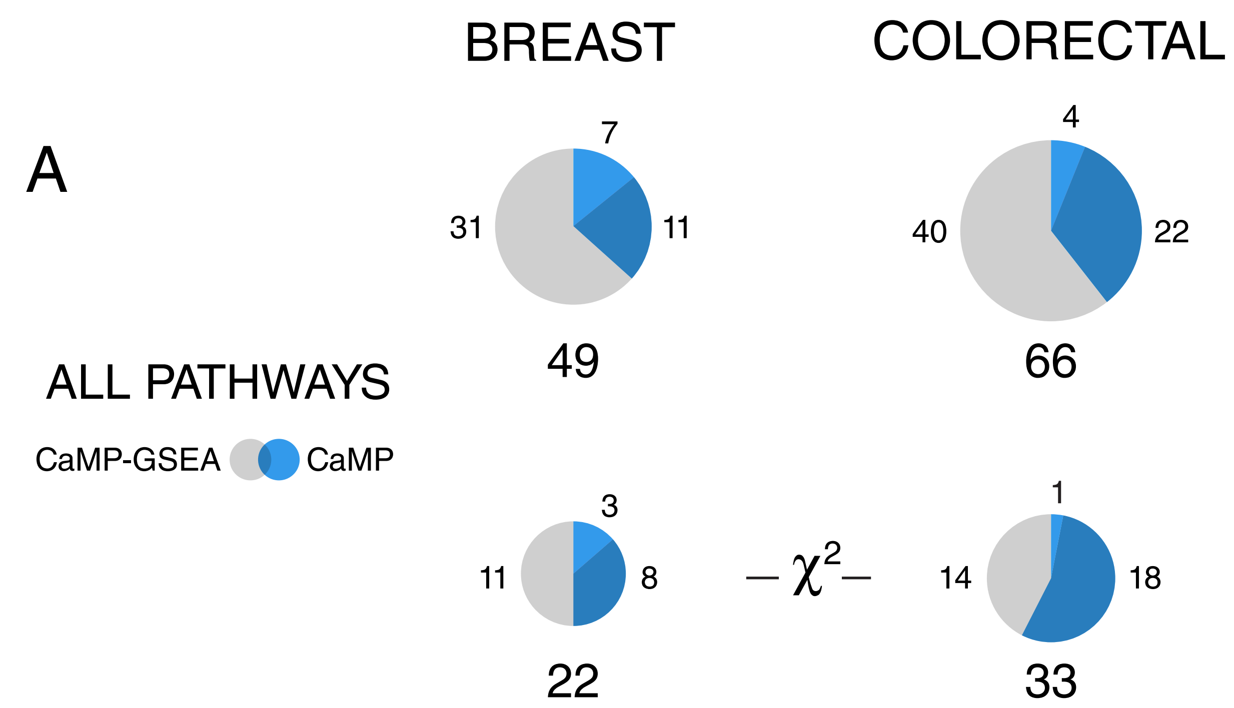


Comparison of mutation enrichment in cellular pathways using complementary statistical approaches. Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.

The figure is too complex. It can be salvaged by presenting the data more rationally.

REFACTORING COMPLEXITY

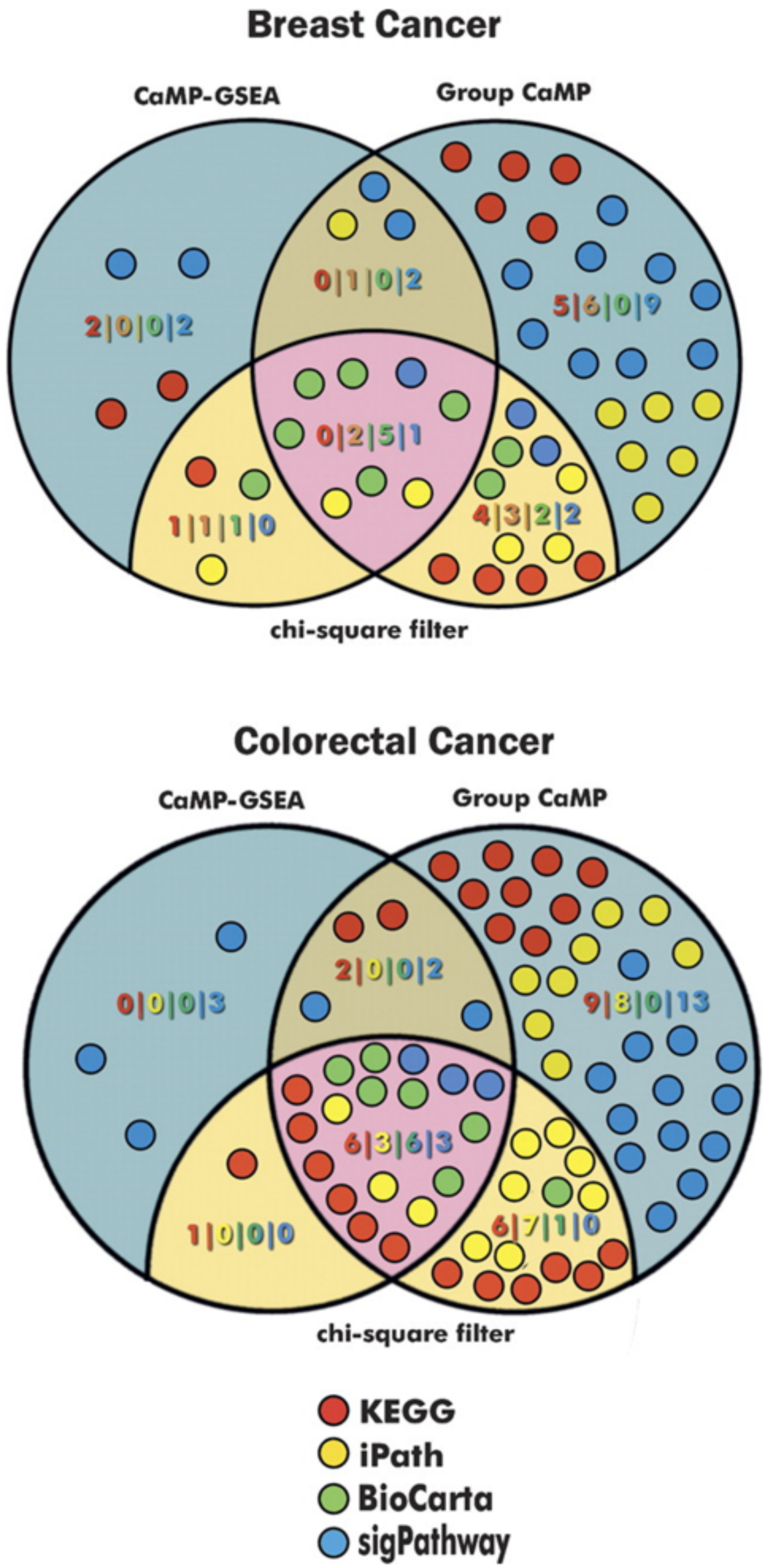


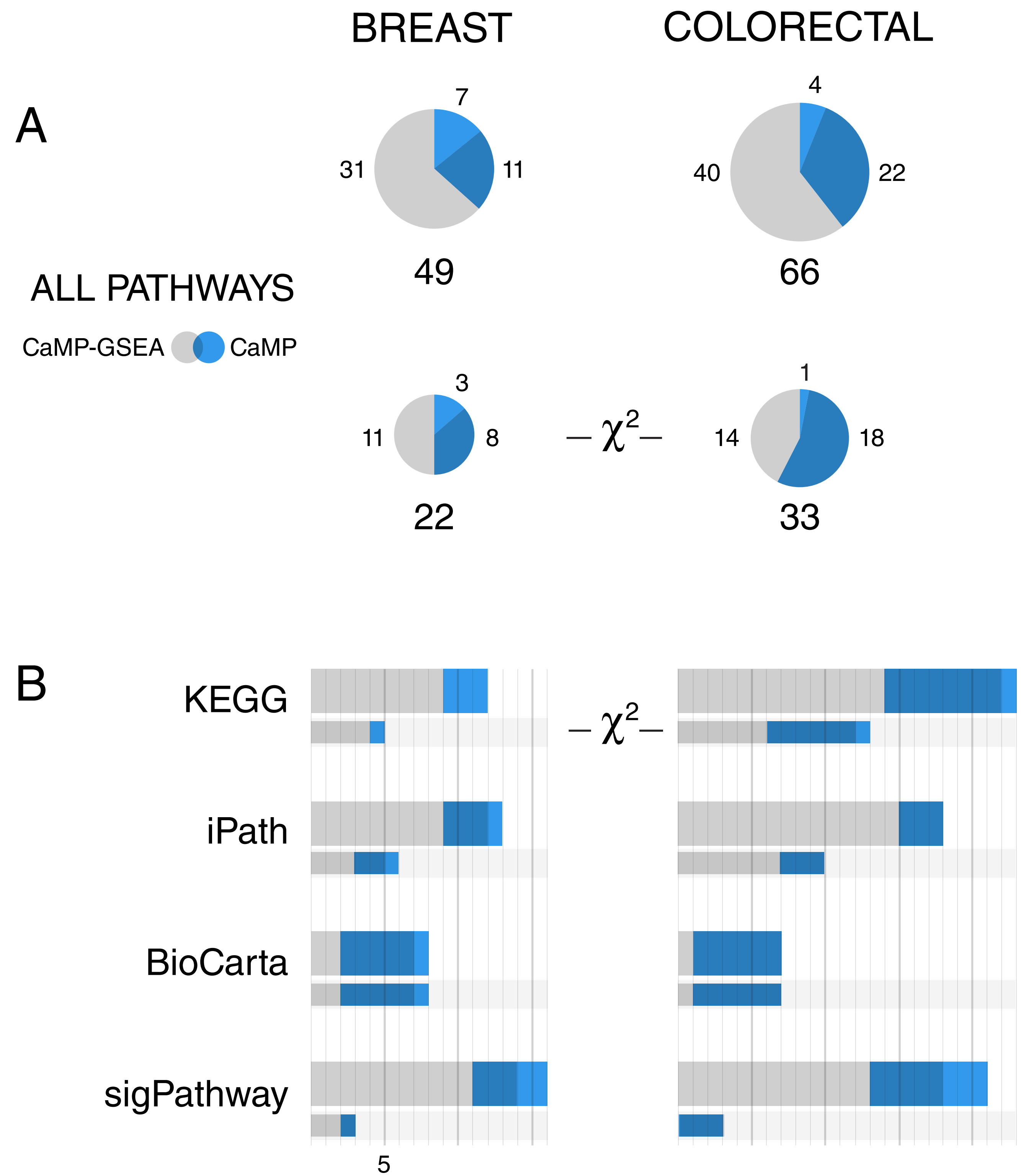


Comparison of mutation enrichment in cellular pathways using complementary statistical approaches. Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.

First, introduce the reader to the basic classifications and proportions. Already the reader can see that the chi-squared filter decreases the relative number of observations from the CaMP-GSEA group.

REFACTORING COMPLEXITY

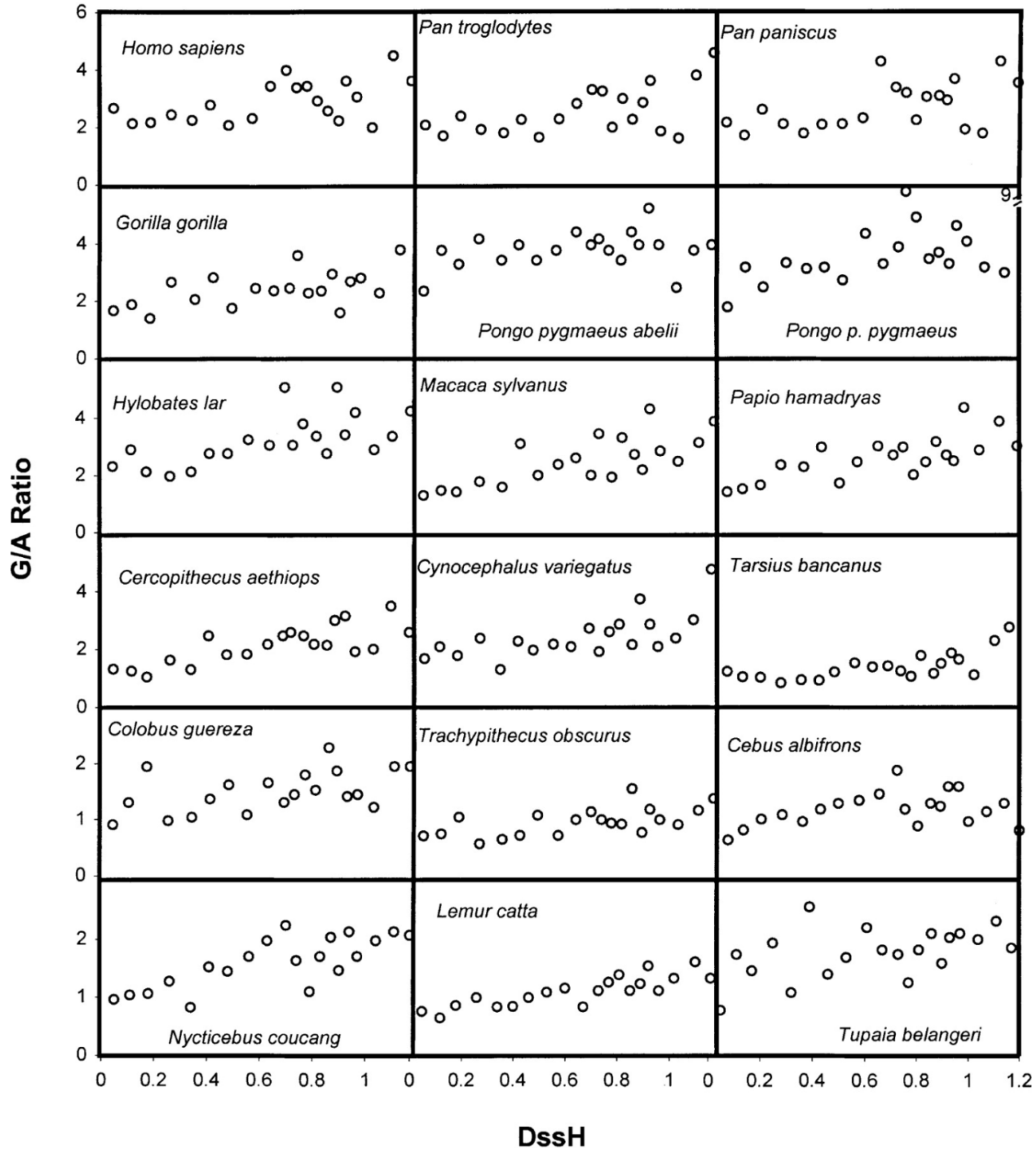




Comparison of mutation enrichment in cellular pathways using complementary statistical approaches. Lin, J., et al., A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res, 2007. 17(9): p. 1304-18.

Second, focus on the individual pathways. The reader's task is simplified by the consistency between the panels. Patterns are easily spotted. (1) Colorectal cancer has typically more observations (2) CaMP-GSEA observations are excluded by the chi-squared filter, except in the case of BioCarta. (3) Observations in both CaMP-GSEA and CaMP groups (dark blue) are relatively unaffected by the filter.

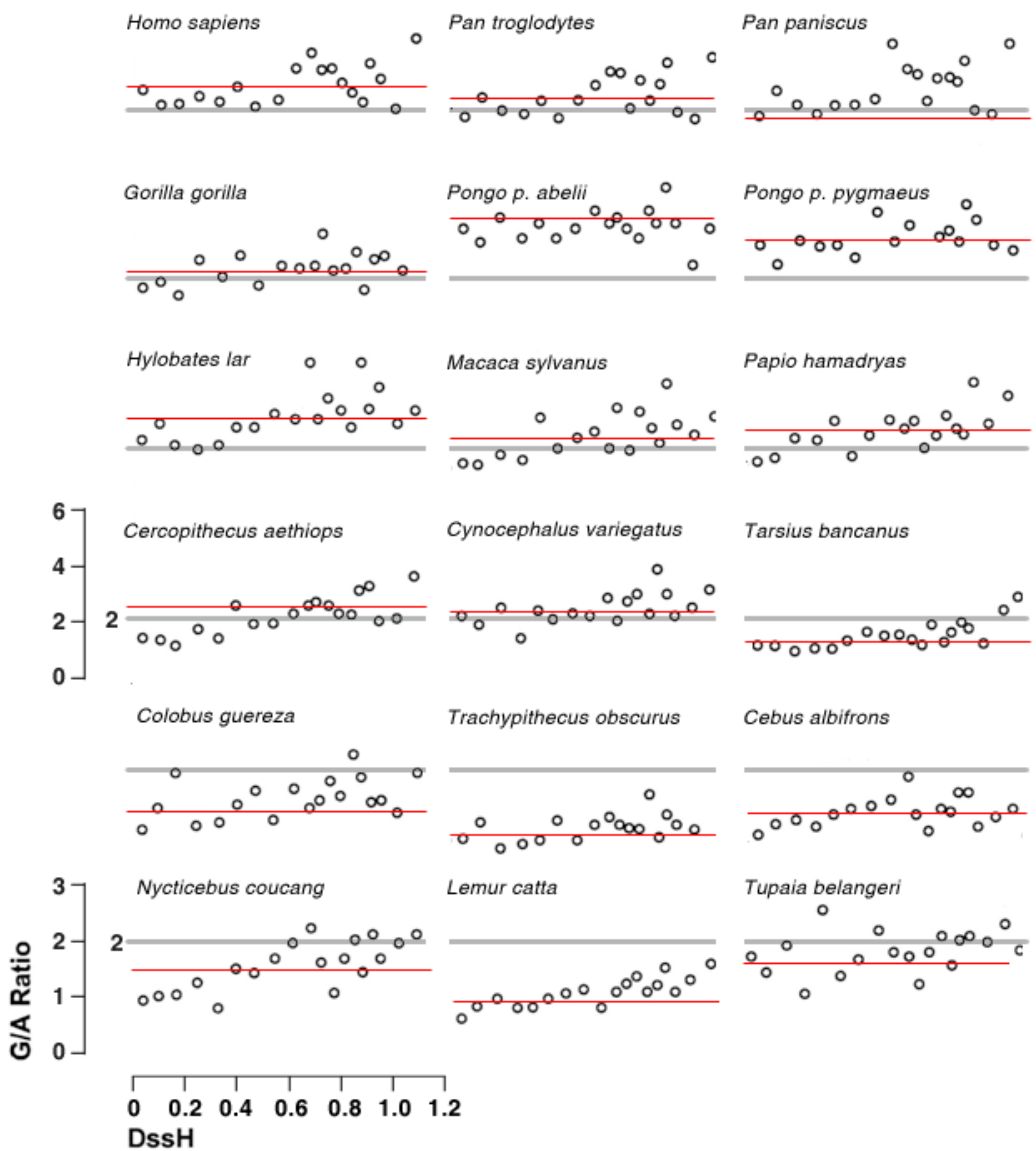
FOCUS ON DATA



G/A ratios for complete primate mitochondrial genomes and two near outgroups. Raina, S.Z., et al., Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Res, 2005. 15(5): p. 665-73.

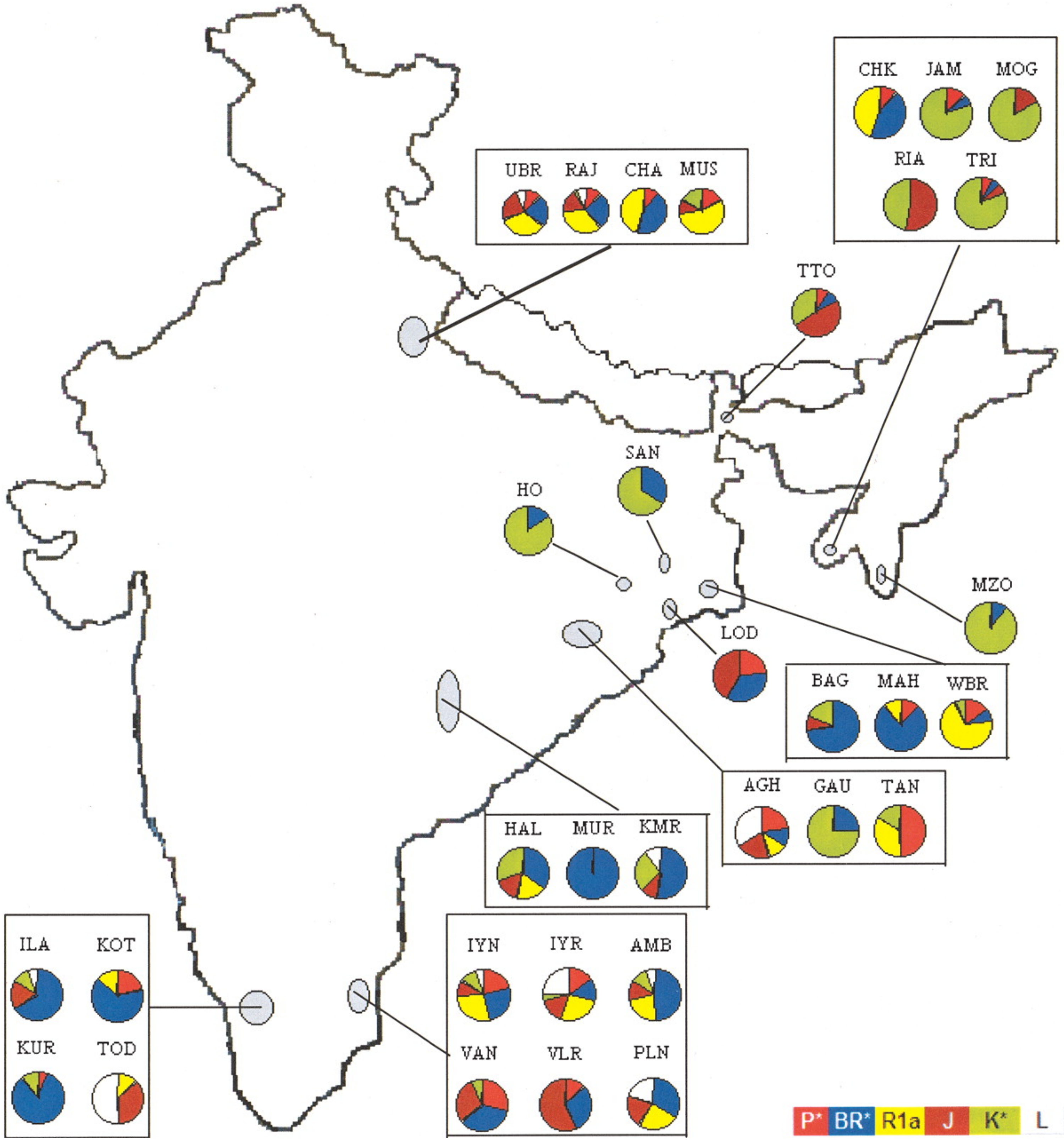
Lack of alignment, changing scales and thick boxes reduce the salience of the data.

FOCUS ON DATA



The data-to-ink ratio is increased by removing unnecessary elements. Notice how this version uses negative space (gutters between the columns) to achieve a high degree of visual organization.

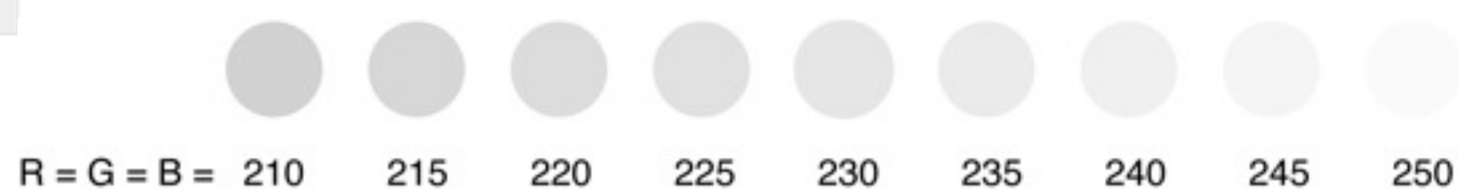
GRID LAYOUT



Frequencies (%) of Y-chromosomal haplogroups among ethnic populations. Basu, A., et al., Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res, 2003. 13(10): p. 2277-90.

The use of a grid layout can greatly improve a figure. Here, elements are organized haphazardly.

GRID LAYOUT

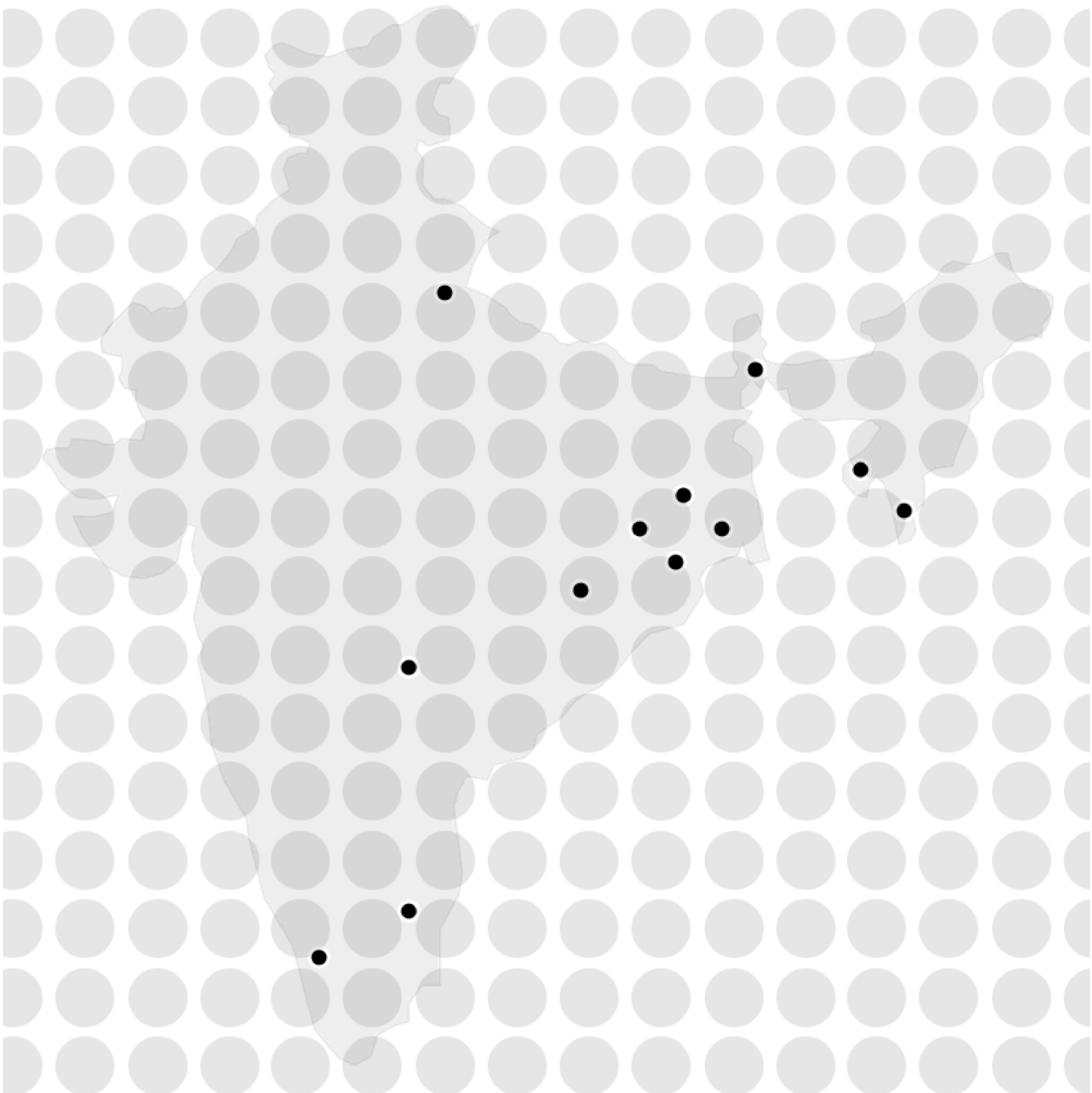


Be careful when using very light colors. They frequently do not render well on LCD projectors and appear washed out.

GRID LAYOUT

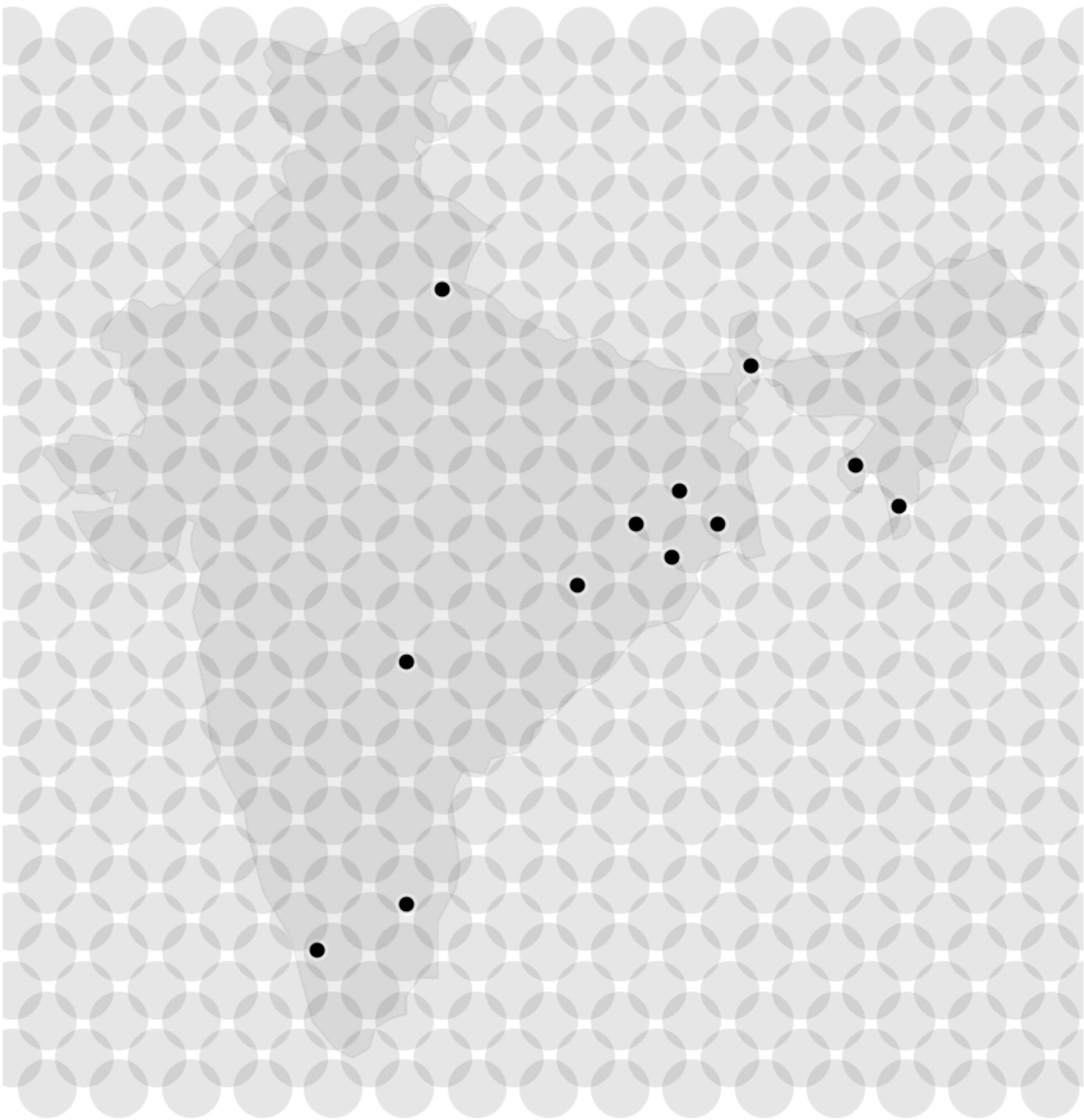
Remove variation in shape and size of locations.

GRID LAYOUT



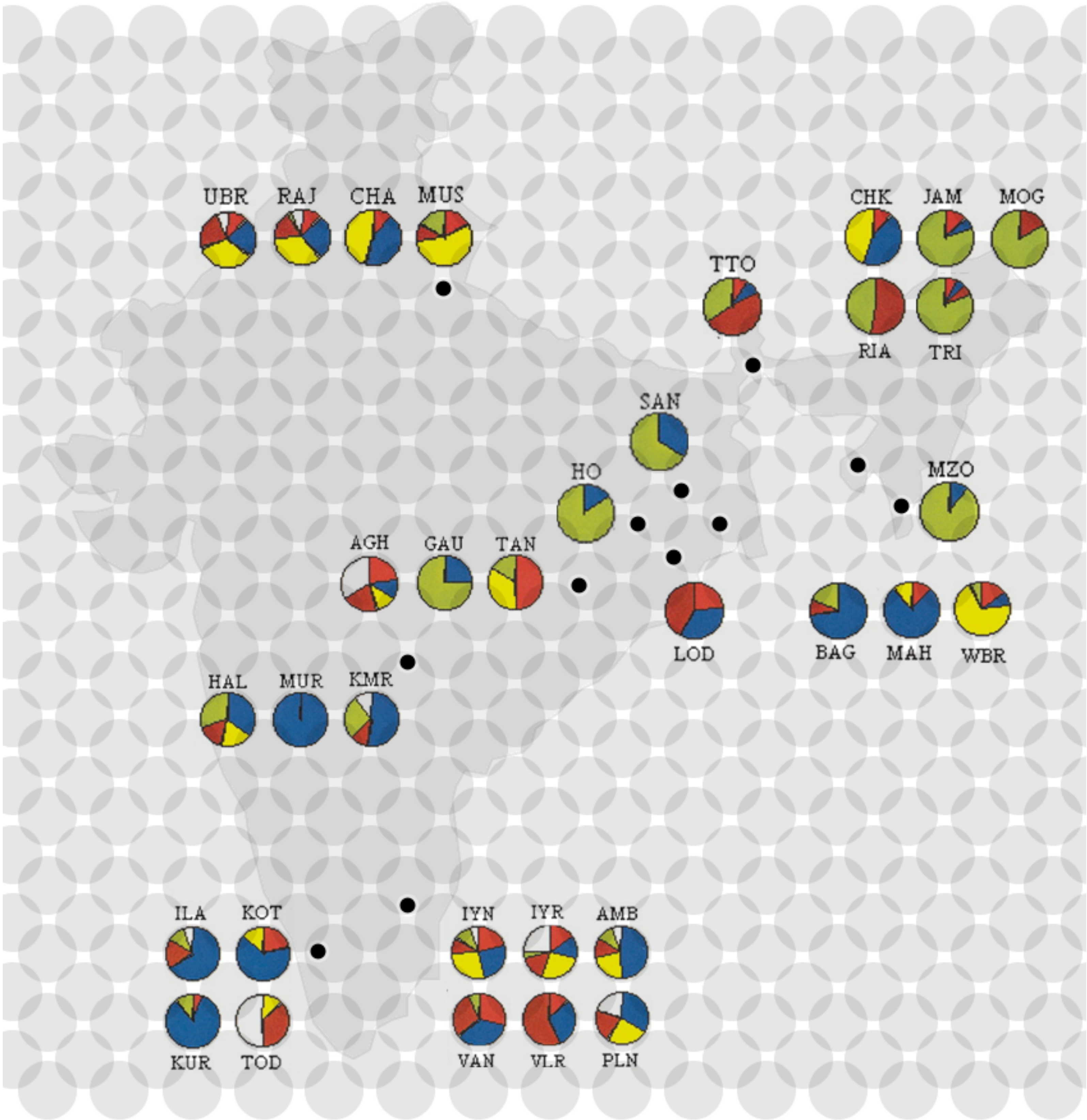
Construct a placement grid.

GRID LAYOUT



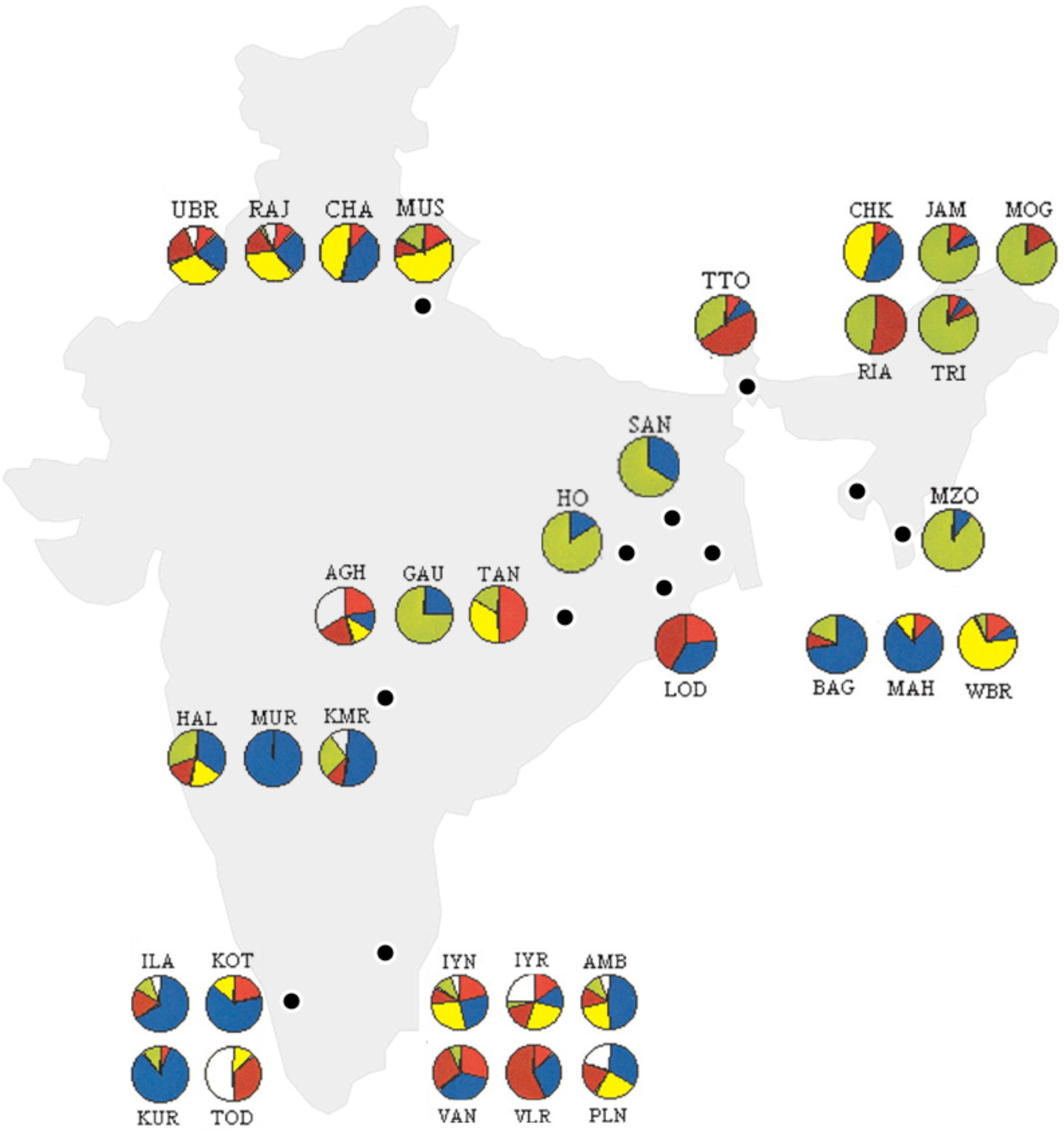
Double up the grid to increase number of positions for the pie charts.

GRID LAYOUT



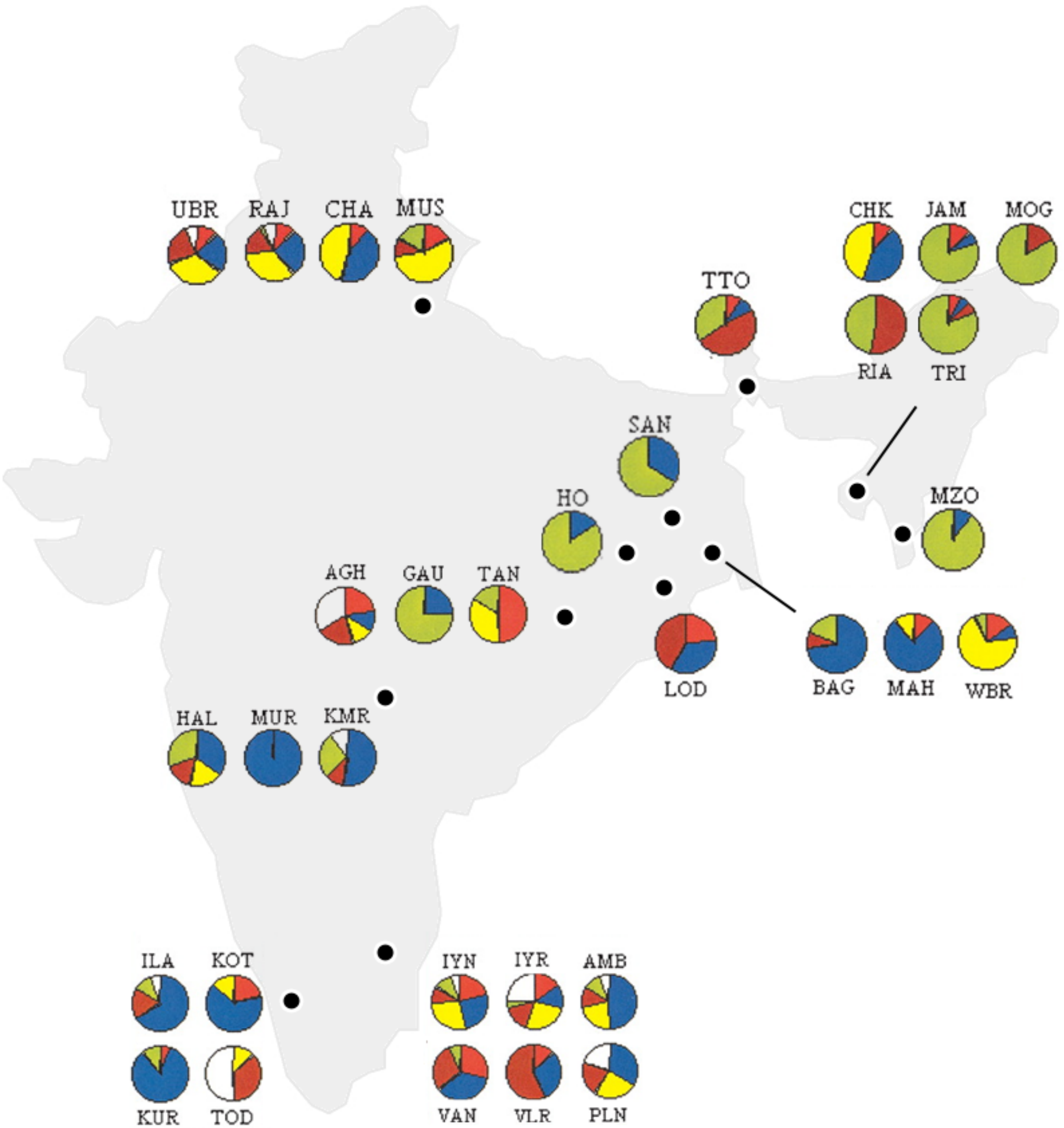
Place pie charts only on the grid circles, as close to their geographical locations as possible.

GRID LAYOUT



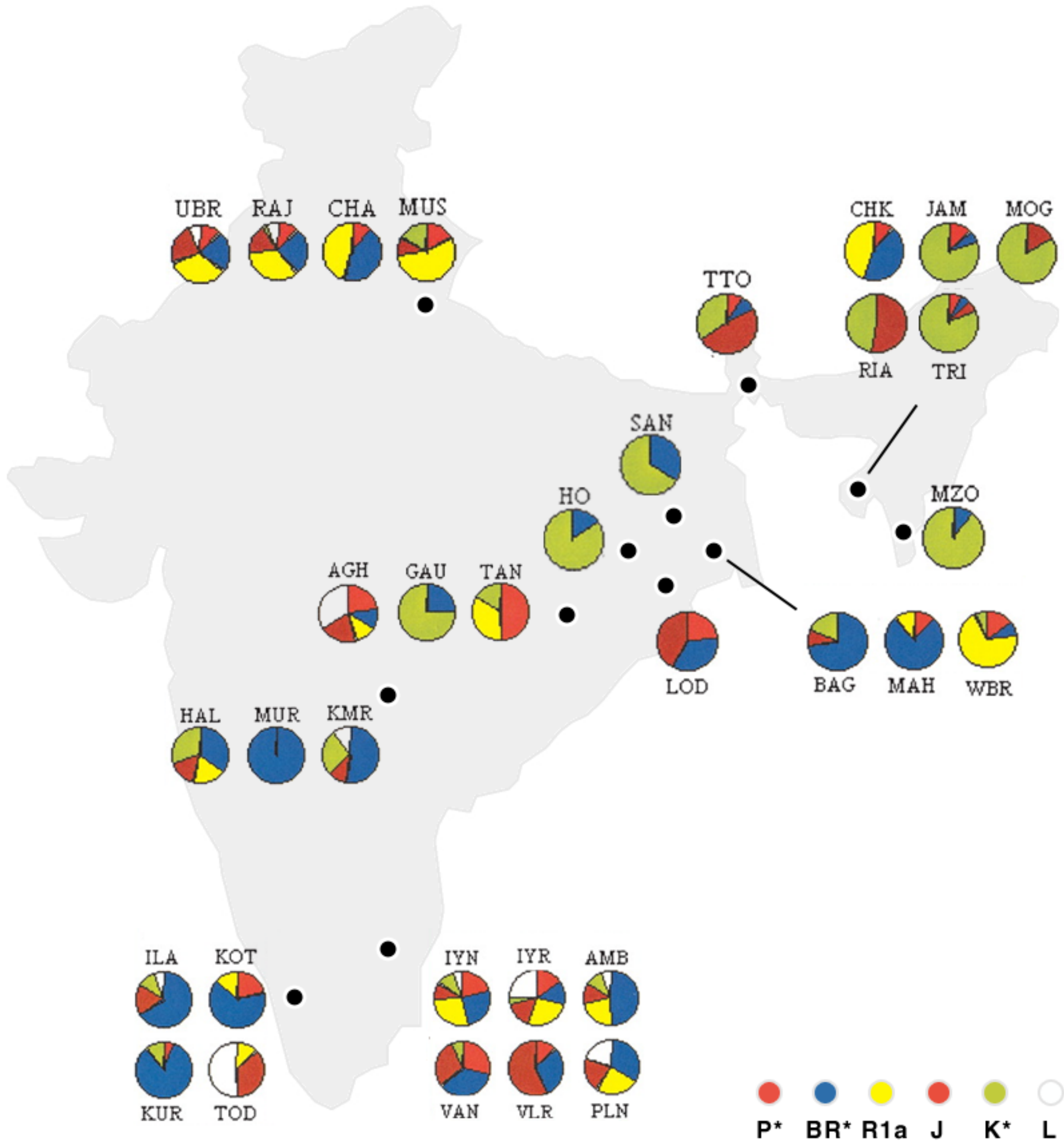
By doing this, the correspondence between the pie chart groups and locations is unambiguous even without any callout lines.

GRID LAYOUT



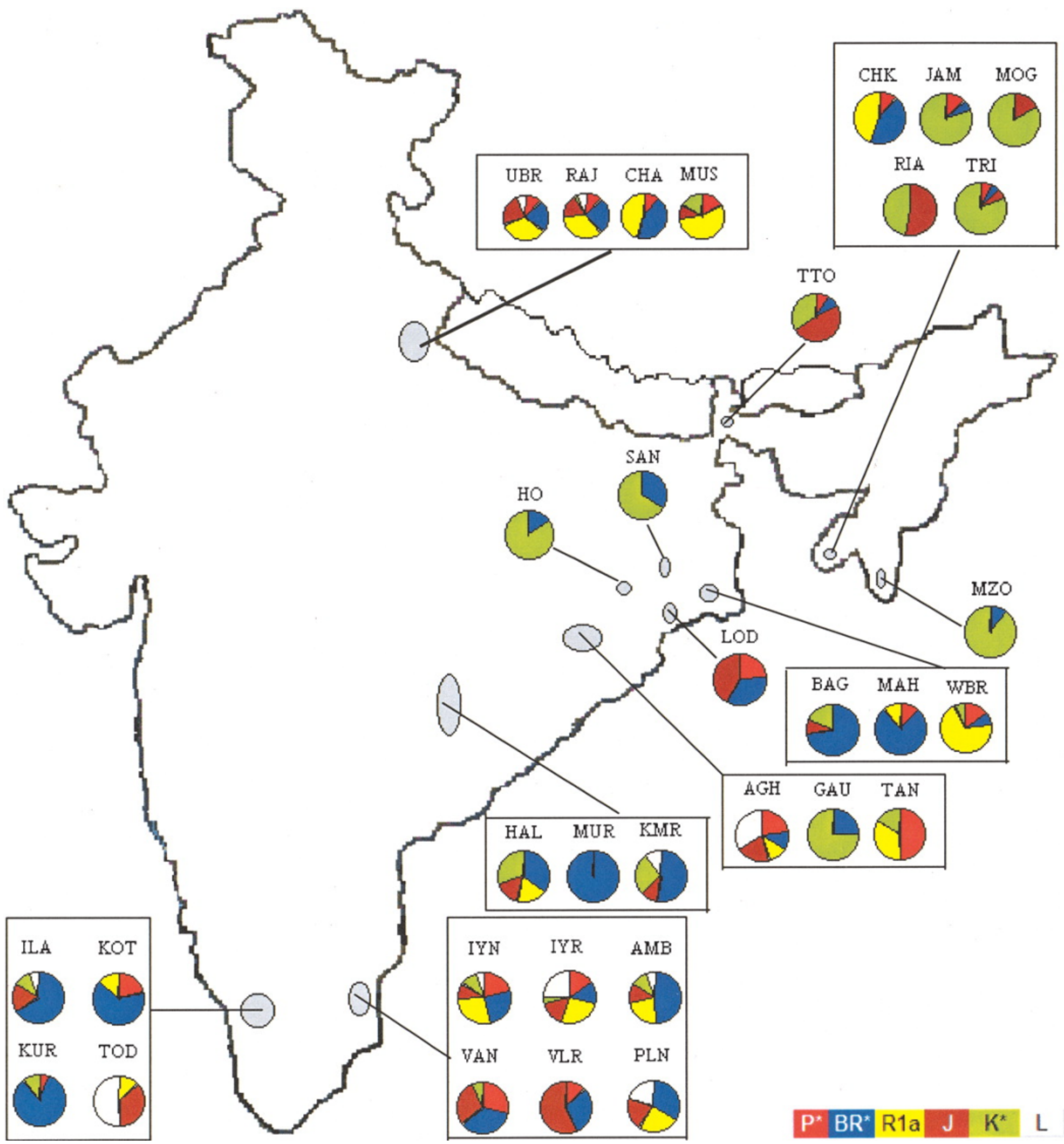
We need callouts only for two cases. Here, limit the variation of angles of callout lines (both are at 45 degrees).

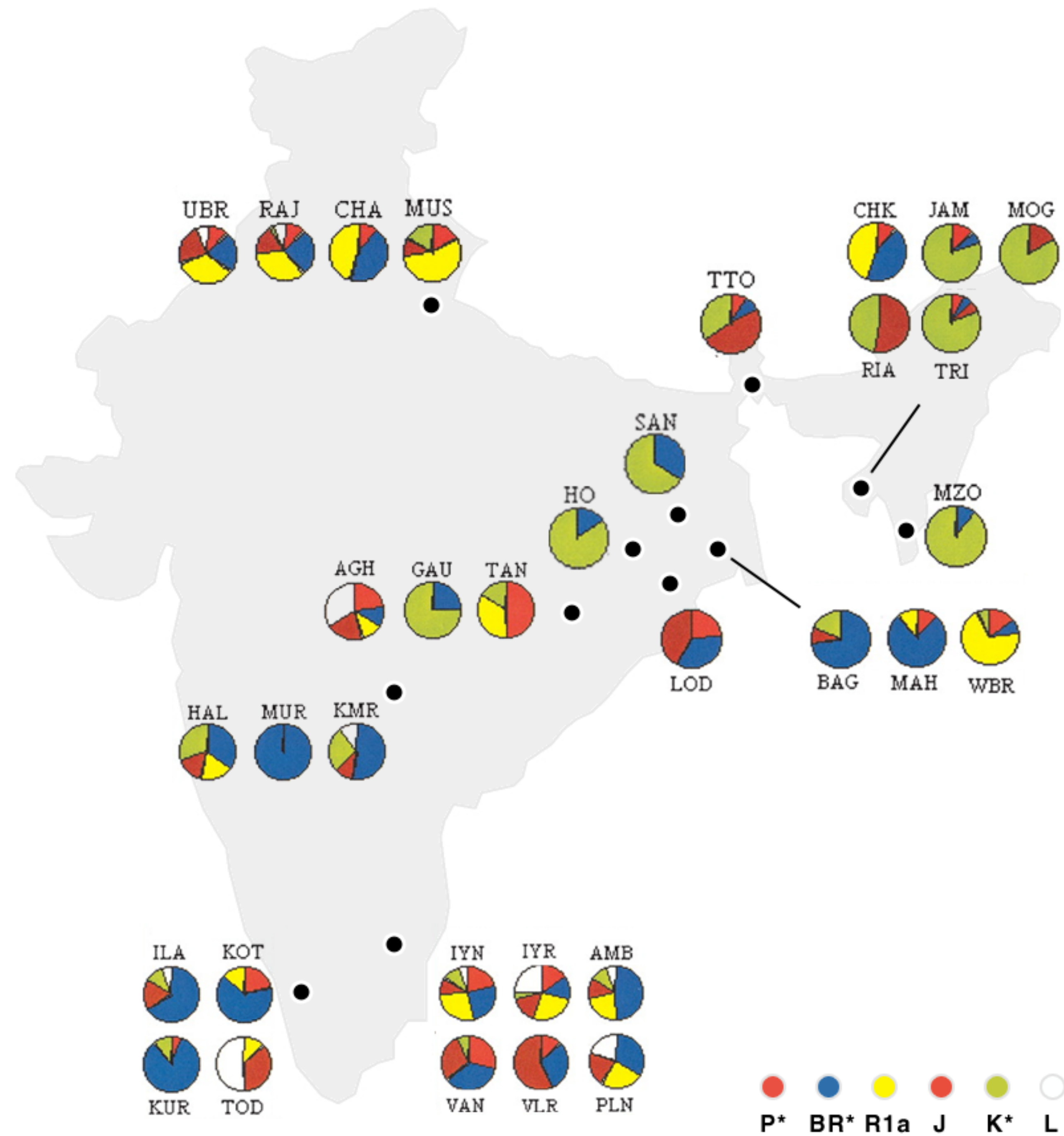
GRID LAYOUT



Maintain correspondence between the shape of data (circular pie chart) and legend (circles).

GRID LAYOUT





We have been able to remove the dependence on boxes and callouts by using a grid model. Clutter is reduced.

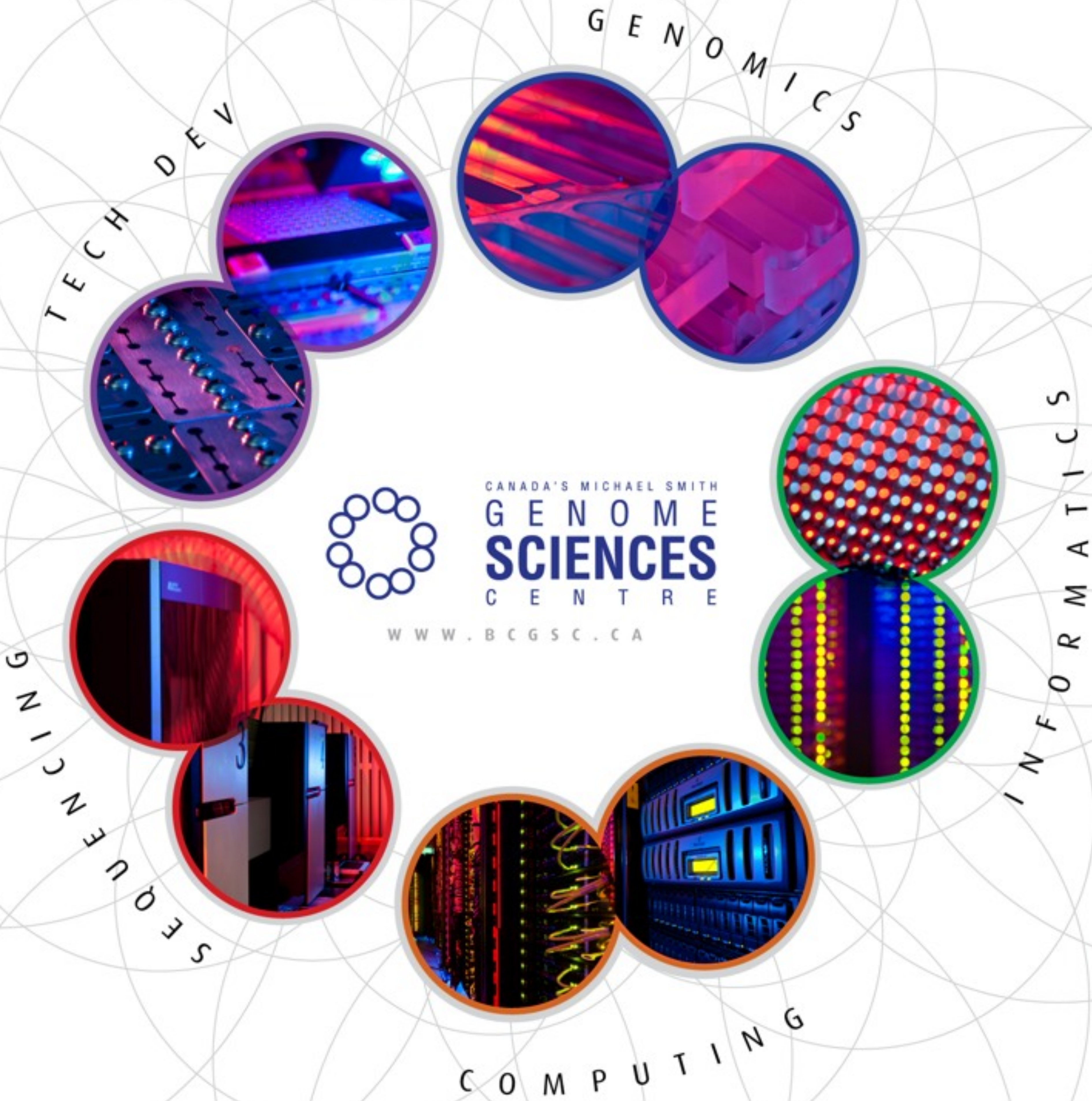
VISUALIZATION CHECK LIST

do I have a core message? is it salient?

have I used the simplest visual representation?

can I simplify or remove elements without impacting the message?

have I captured relevant hierarchies?



AFFILIATIONS BC Cancer Research Center • BC Cancer Agency • BC Cancer Foundation • Genome BC • Simon Fraser University • University of Britsh Columbia • Genome Sciences Institute