

Visualization, analysis and design

Jeffrey
West

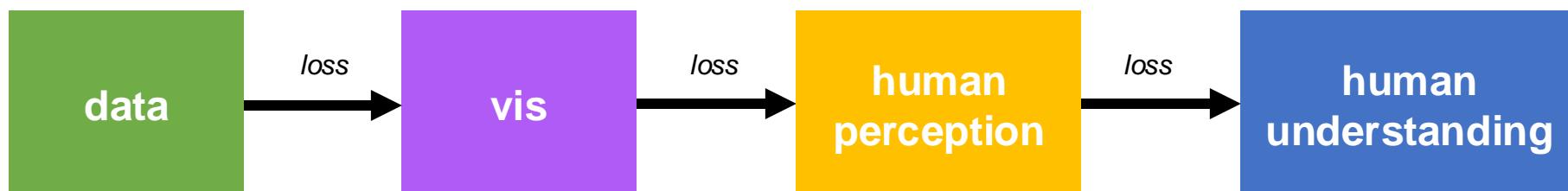
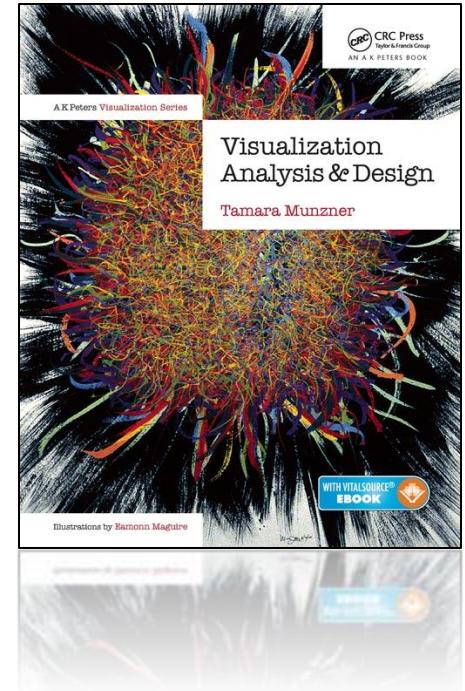
Assistant Professor
Integrated Mathematical Oncology
H. Lee Moffitt Cancer Center



Some intriguing thoughts on visualization



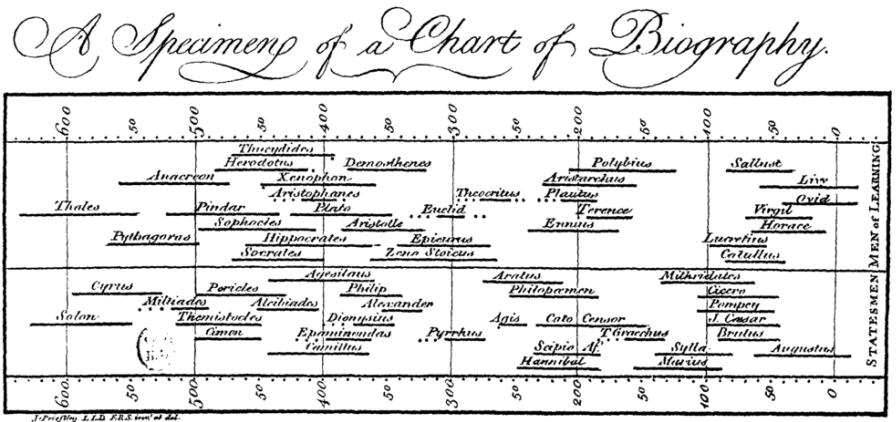
- **Not all visualization methods are created equal.**
 - A bad vis can lead to erroneous interpretations
 - A good vis can lead to insights otherwise missed
- **Not all visualization methods have been invented yet.**
 - The space of possible vis methods is astronomical
- **Visualization choice should match the purpose**
 - Vis can be used to consume, produce, or enjoy



A brief history of visualization....

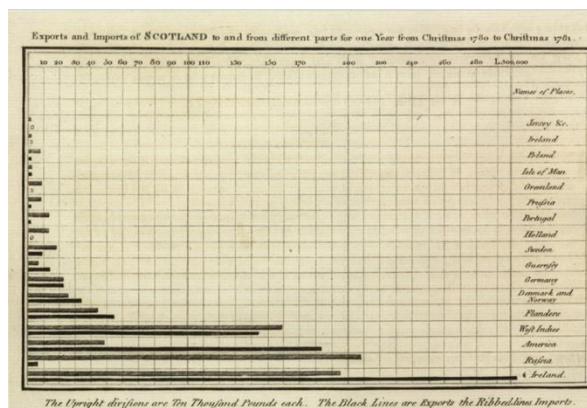


Timeline chart (1765)
Joseph Priestley



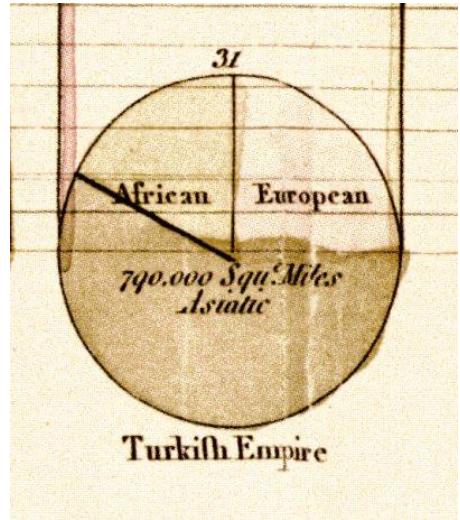
Timeline chart, used to visualize the birth, death, and lifespan of famous individuals.

Bar Chart (1780)
William Playfair



Bar chart Scotland's imports and exports from and to 17 countries in 1781 are represented

Pie Chart (1801)
William Playfair

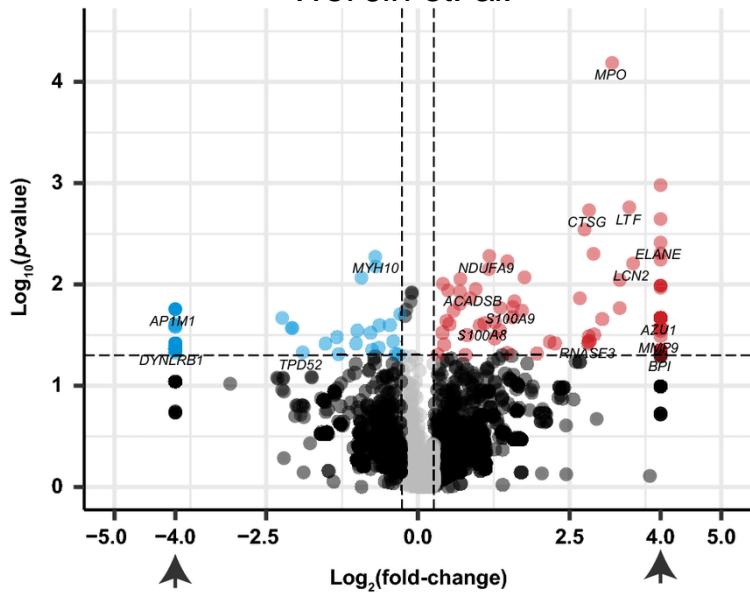


Pie chart from Playfair's Statistical Breviary, showing the proportions of the Turkish Empire located in Asia, Europe and Africa before 1789

Examples of novel visualization schema

Volcano Plot (2001)

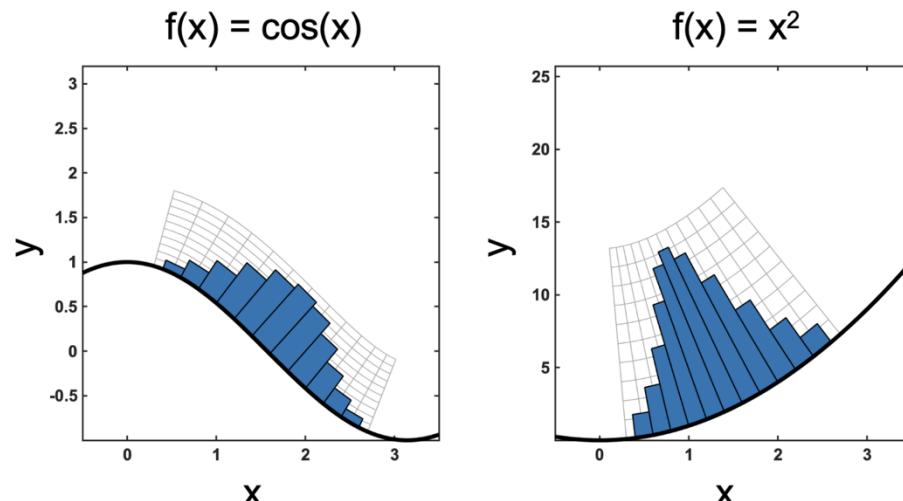
Wei Jin et. al.



A volcano plot combines a measure of statistical significance from a statistical test (e.g., a p value from an ANOVA model) with the magnitude of the change

Curvogram (2022)

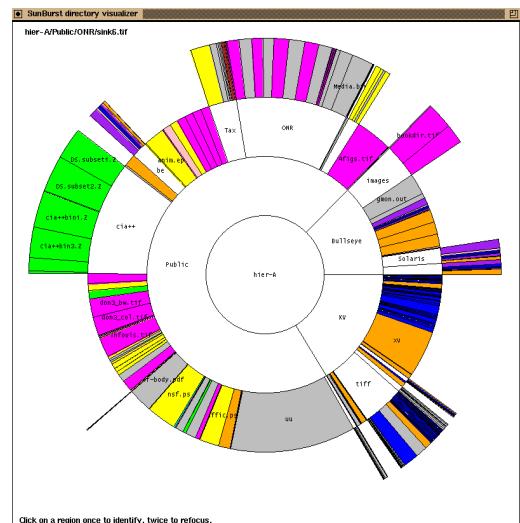
Ryan Schenck



A curvogram plot projects a discrete probability mass function (p.m.f.) onto a line plot for a given $f(x)$.

Sunburst (2000)

John Stasko

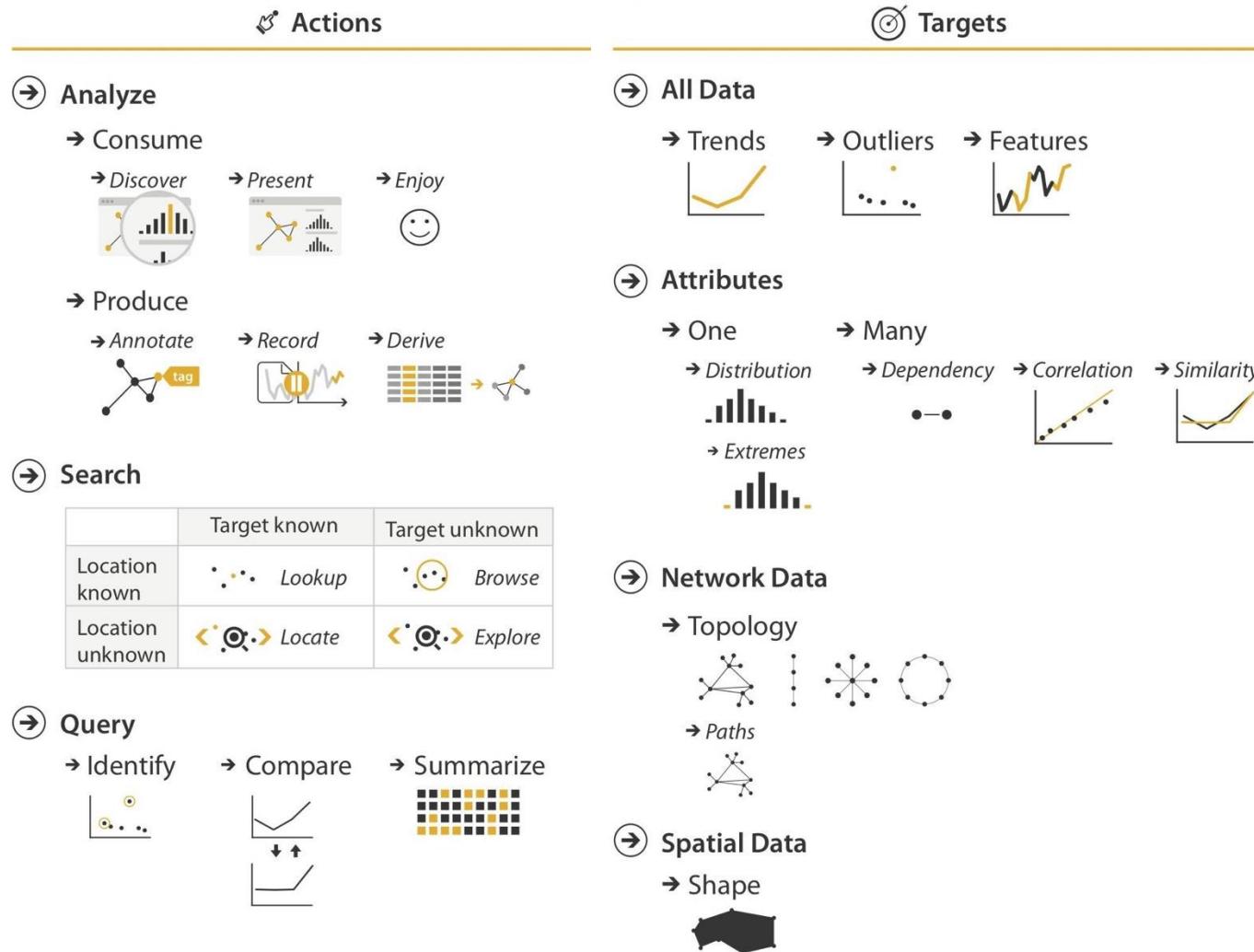


A sunburst diagram illustrates nested relationships as parts of a whole.

Analysis Tasks

Abstracting the possible tasks to manipulate / analysis data

Abstracting tasks to manipulate / analysis data



Actions (verb)

- Define the goals of the visualizer
- Analyze
 - Search
 - Query

Target (noun)

- Define the meaningful aspect of interest within dataset
- Trends
 - Outliers
 - Features
 - Summary statistics

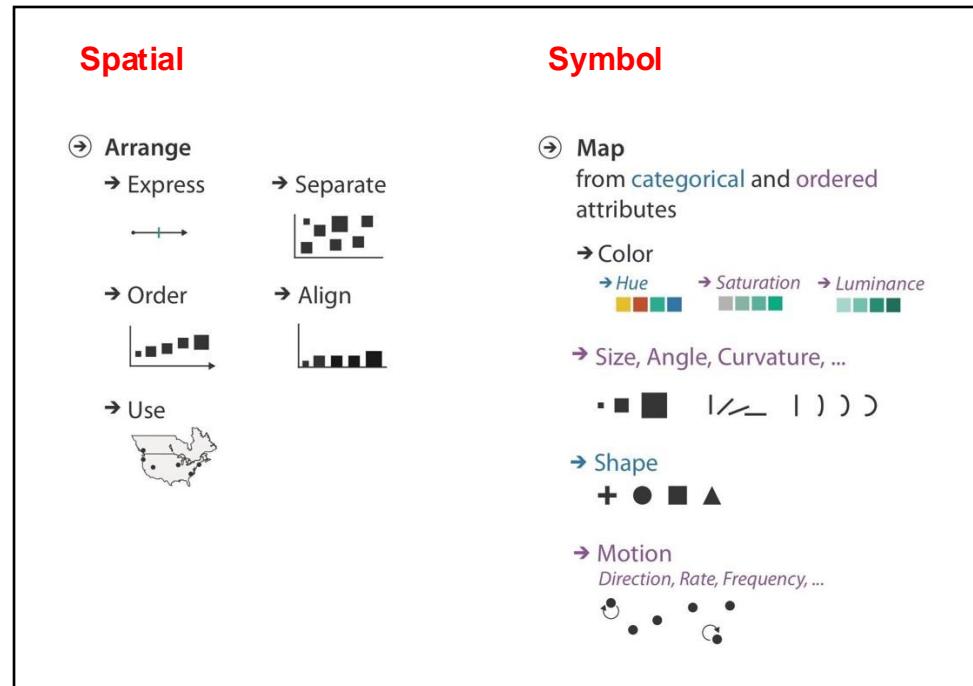
The process & state space of vis



For example ...

compare input/output + trend line approx. → **line plot with scatter plot**

What is the space of encoding?



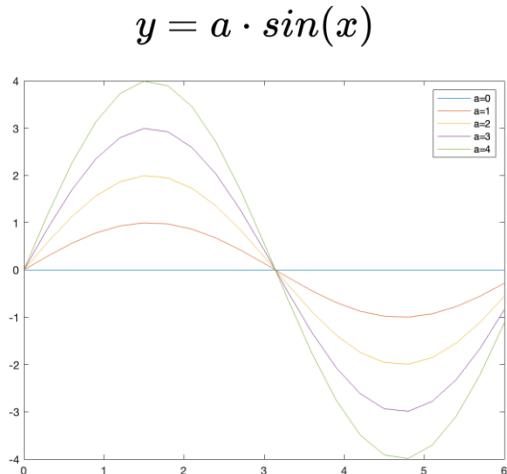
Shape, size, color

Encoding information

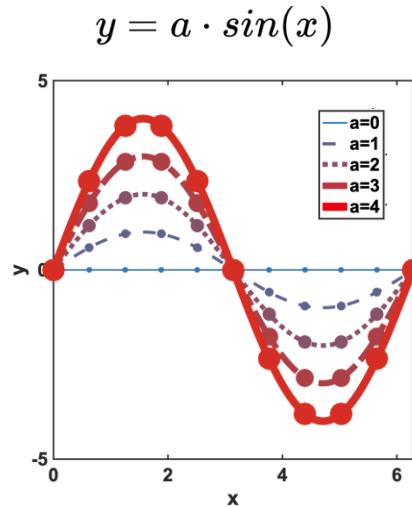
Encoding information with shape, size, color



Principle 1:
Default settings are universally bad



Principle 3:
A goldilocks zone for multi-encoding
(simplicity is aesthetic)



Principle 5:
Feedback is required



Principle 2:
Encoding types are confounded
(color encoding is difficult to interpret for low
size encoding)

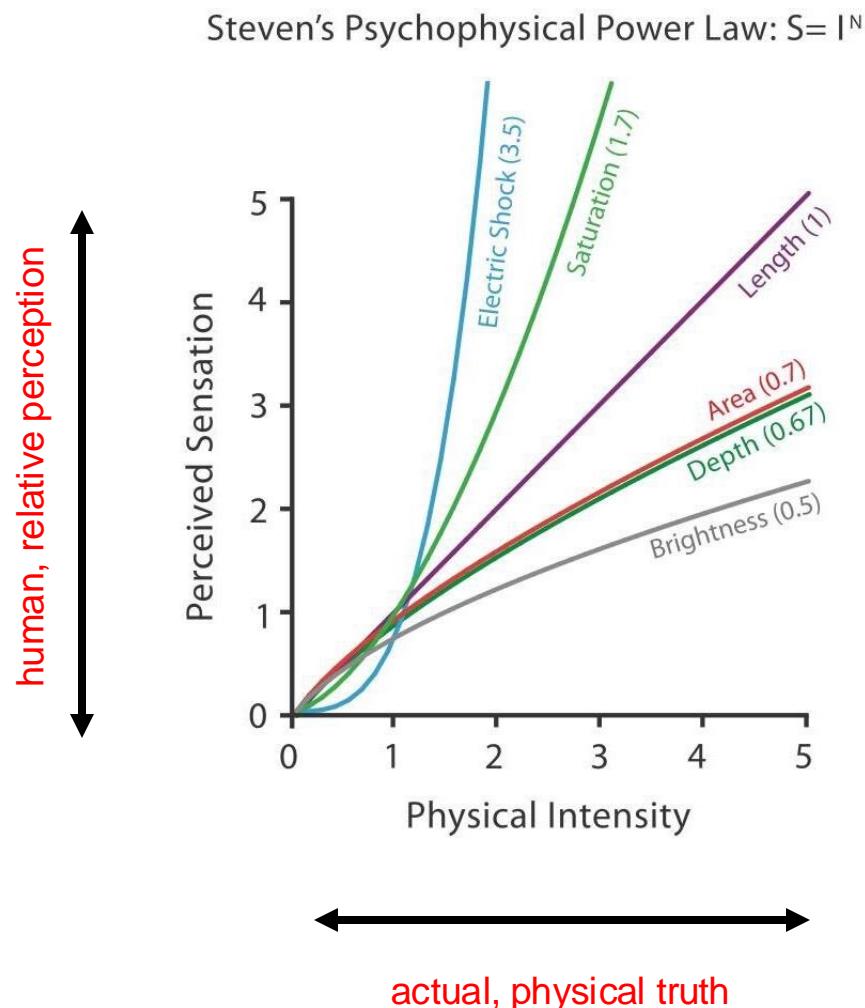
Principle 4:
What's good for publication may not be good
for presentation

women

men

xkcd Color Survey
<https://blog.xkcd.com/2010/05/03/color-survey-results/>

Principle: vis based on human perception



Channels: Expressiveness Types and Effectiveness Ranks

④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



Most

Effectiveness

Same

Least

Principle: vis based on human perception

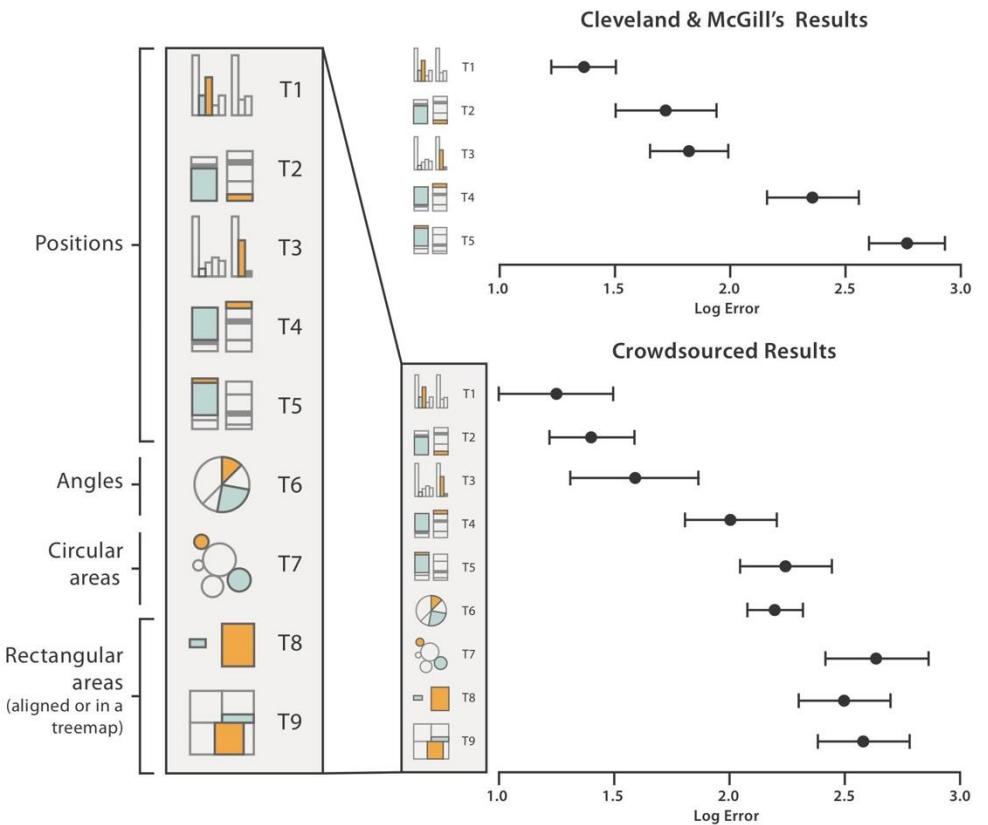
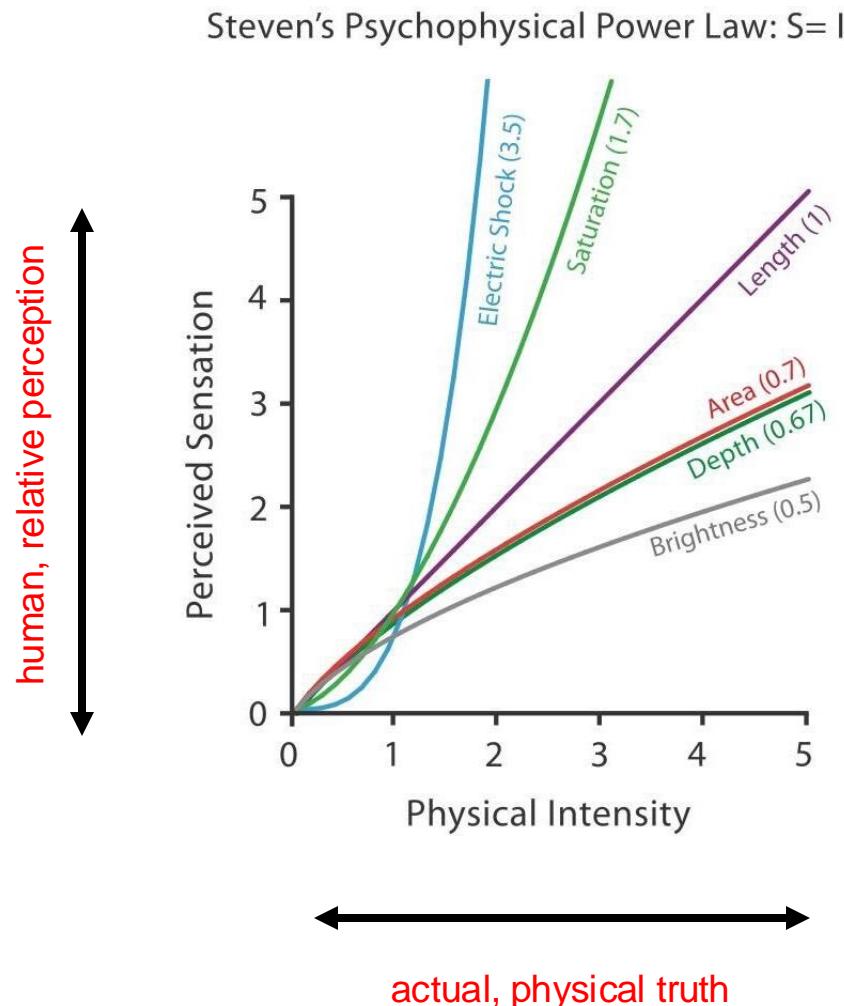


Figure 5.8. Error rates across visual channels, with recent crowdsourced results replicating and extending seminal work from Cleveland and McGill [Cleveland and McGill 84a]. After [Heer and Bostock 10, Figure 4].

Colors

Encoding information using color

Color as visual encoding



→ Color

→ Color Encoding

→ Hue → Saturation → Luminance



→ Color Map

→ Categorical



→ Ordered

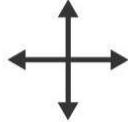
→ Sequential



→ Diverging



→ Bivariate

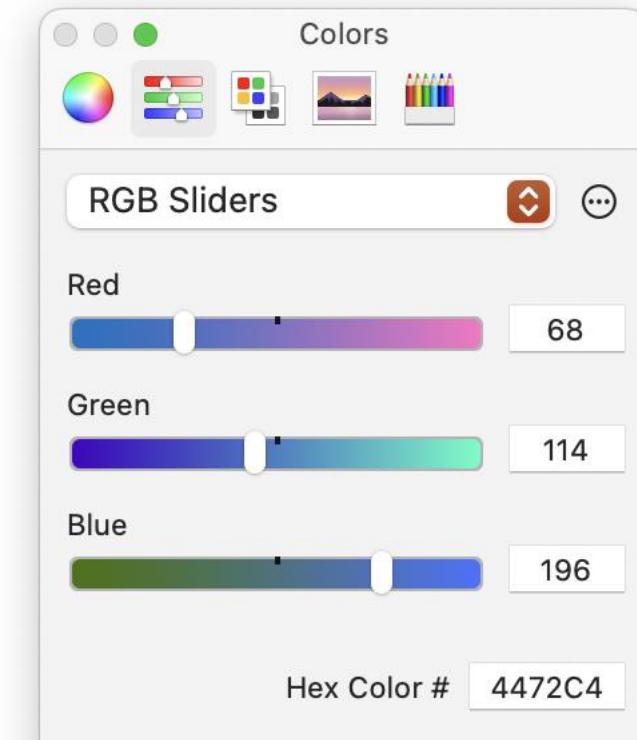


Human visual system's color detection is 3-dimensional



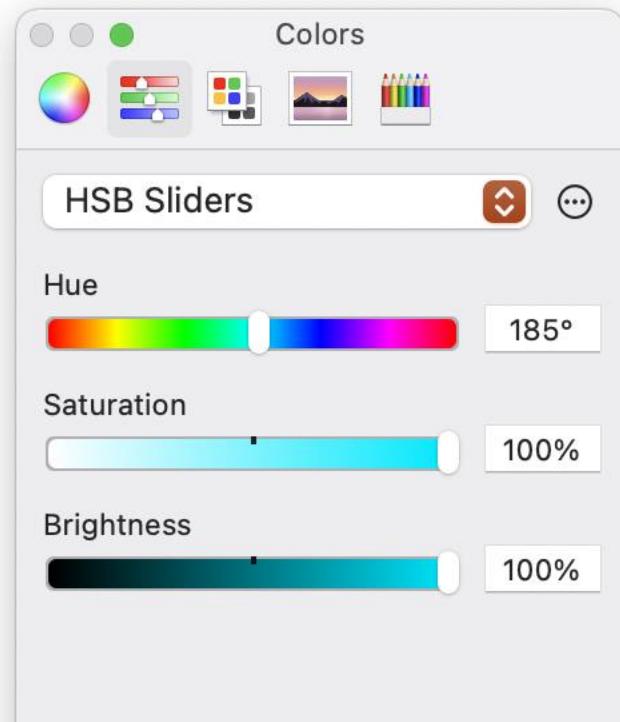
RGB system:

Red: Red-sensing cones (L cones)
Green: Green-sensing cones (M cones)
Blue: Blue-sensing cones (S cones)



HSL system:

Hue: pure colors, not mixed with white or black
Saturation: amount of white
Lightness: amount of black (or, Brightness)



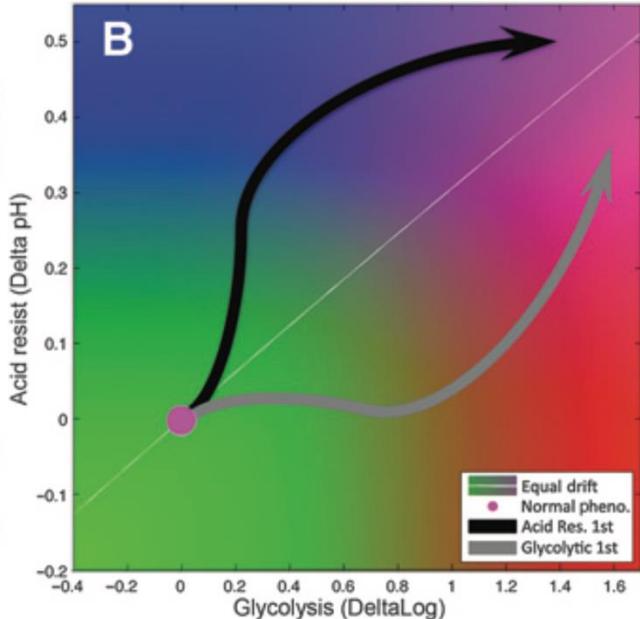
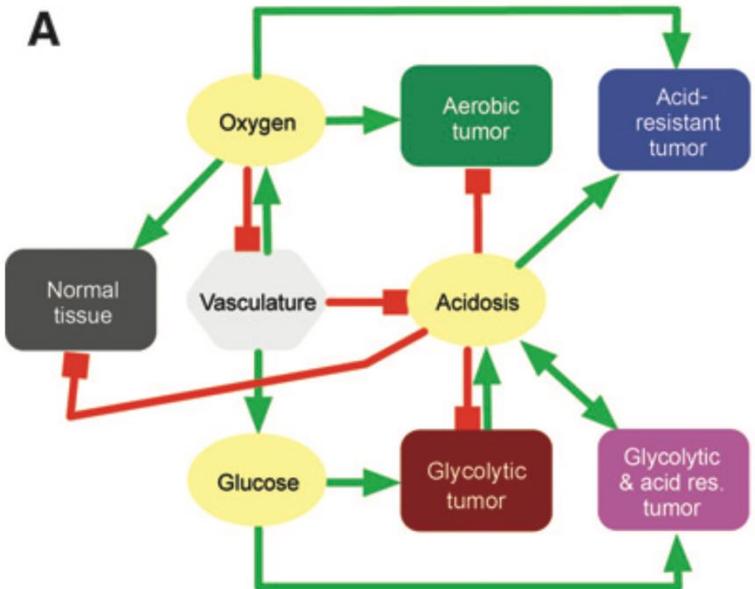
Lightness

Color is analogue, but encoding is digital



Colors are encoded digitally, which can cause problems for continuous colormaps

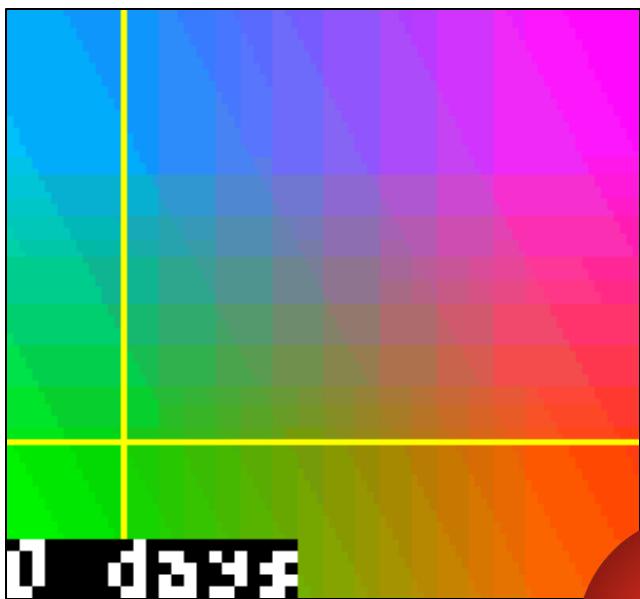
Example:



[Robertson-Tessi, Cancer Research 2015]

Example: gifs

- Gifs can only encode 24-bit RGB colorspace
- Only 256 colors!



Continuous colormap appears pixelated

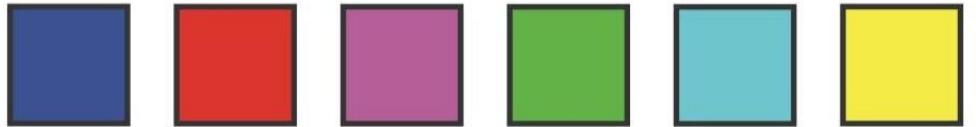


Perceptually linear luminance



Lightness

- The value of L in the HSL system



6 example colors

Luminance

- The photometric measure of *luminous intensity* per unit area of light travelling in a given direction



$L = 0.05$



Measured luminance

L^*

- Nonlinear transformation of luminance that accounts for human perception

Perceptually linear luminance



Lightness

- The value of L in the HSL system

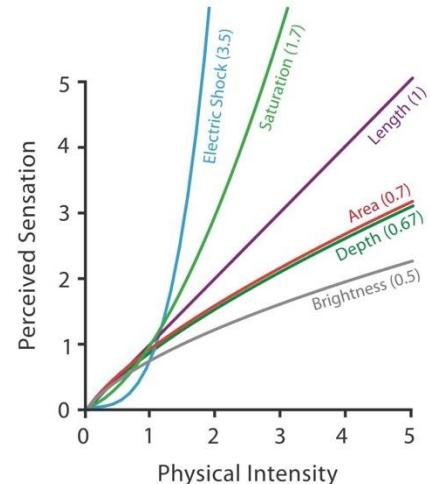
Luminance

- The photometric measure of *luminous intensity* per unit area of light travelling in a given direction

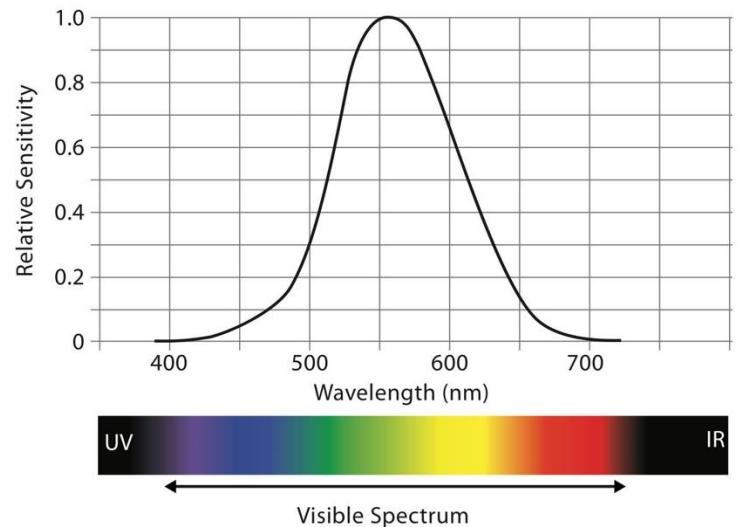
L*

- Nonlinear transformation of luminance that accounts for human perception

Perceived Brightness scales like Intensity^{0.5}



Sensitivity changes with Light Wavelength

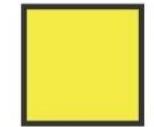
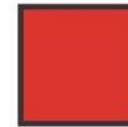


Perceptually linear luminance



Lightness

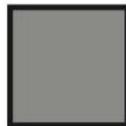
- The value of L in the HSL system



6 example colors

Luminance

- The photometric measure of *luminous intensity* per unit area of light travelling in a given direction



$L = 0.05$

L^*

- Nonlinear transformation of luminance that accounts for human perception



Measured luminance



Computed L^*

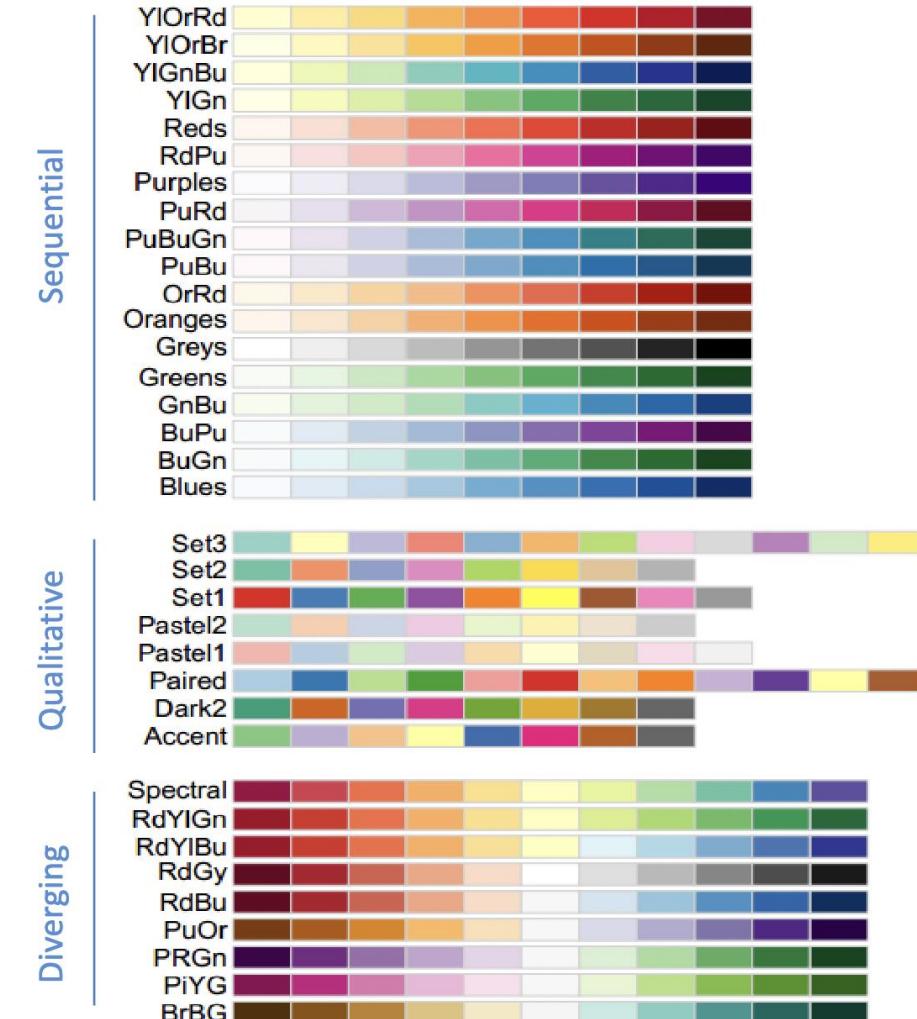
In summary: Colormaps



MatLab

Colormap Name	Color Scale
parula	
turbo	
hsv	
hot	
cool	
spring	
summer	
autumn	
winter	
gray	
bone	
copper	
pink	
sky (since R2023a)	
abyss (since R2023b)	
jet	
lines	
colorcube	
prism	
flag	
white	

RColorBrewer



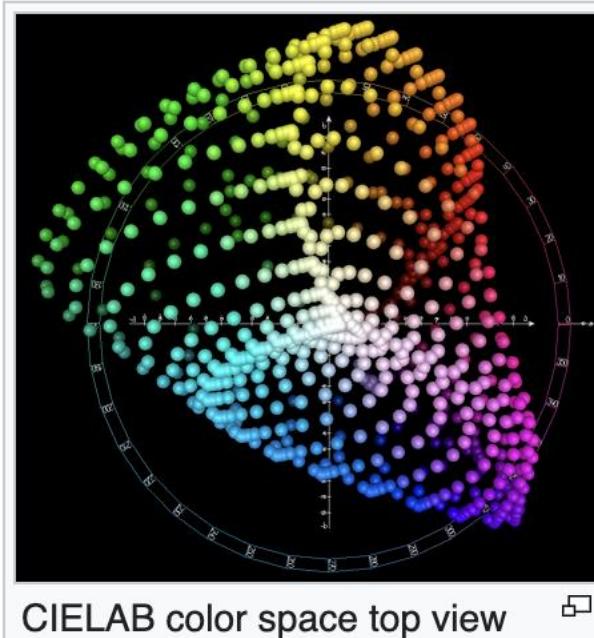
To display RColorBrewer palette: `display.brewer.all()`
 For interactive color selector: <http://colorbrewer2.org/>

In summary: Colormaps



MatLab

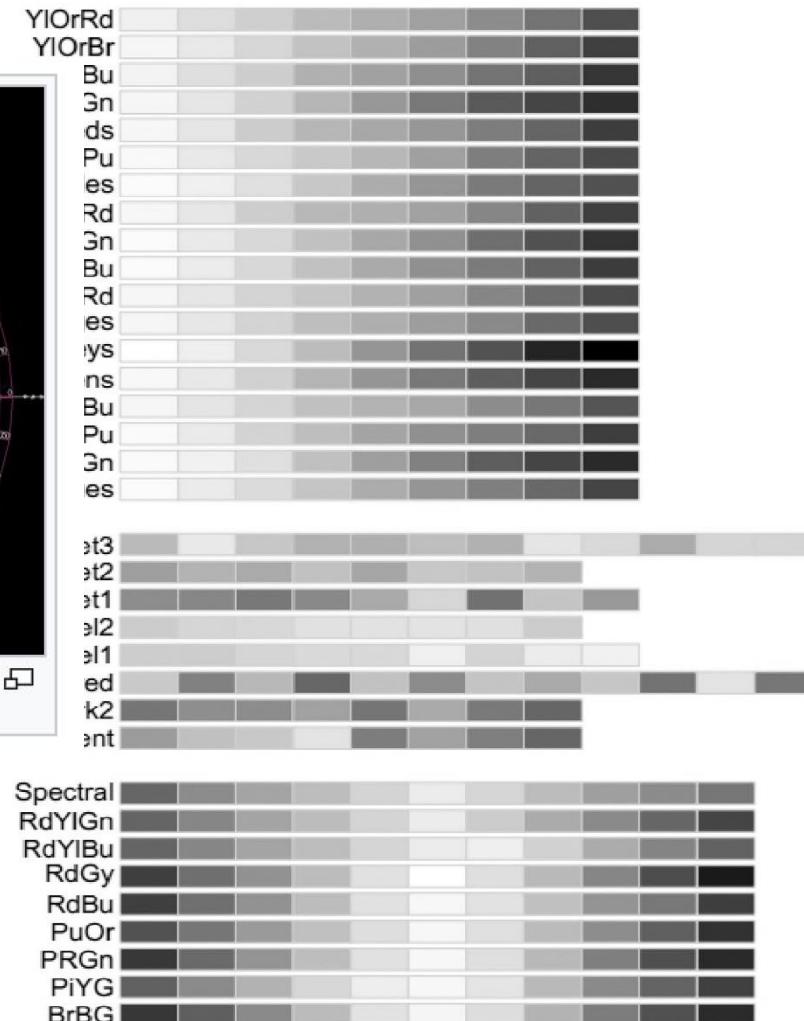
Colormap Name	Color Scale
parula	
turbo	
hsv	
hot	
cool	
spring	
summer	
autumn	
winter	
gray	
bone	
copper	
pink	
sky (since R2023a)	
abyss (since R2023b)	
jet	
lines	
colorcube	
prism	
flag	
white	



$L^*a^*b^*$

Diverging

RColorBrewer



To display RColorBrewer palette: `display.brewer.all()`
 For interactive color selector: <http://colorbrewer2.org/>

Color blindness



Photoreceptors

- Red-sensing cones (L cones): long wavelengths (around 560 nanometers).
- Green-sensing cones (M cones): middle wavelengths (around 530 nanometers).
- Blue-sensing cones (S cones): short wavelengths (around 420 nanometers).

Deutanopia

- Red-green color blindness
- Missing **M cones** (green)
- Mostly see shades of blues and golds

Protanopia

- Red-green color blindness
- Missing **L cones** (red)
- Mostly see shades of blues and golds

Tritanopia

- Blue-yellow color blindness
- Missing **S cones** (blue)
- Mostly see reds, light blues, pinks and lavender.

Deutanomaly

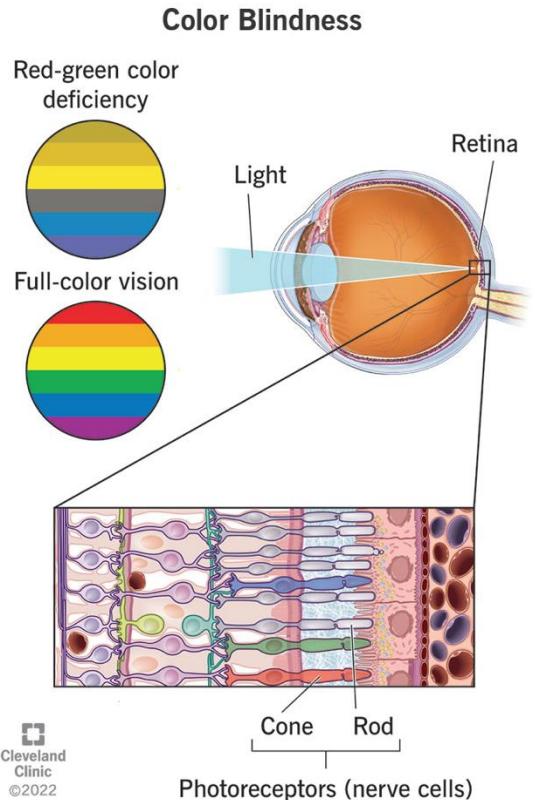
- Red-green color blindness
- All cones, but **M cones** (green) are less sensitive

Protanomaly

- Red-green color blindness
- All cones, but **L cones** (red) are less sensitive

Tritanomaly

- Blue-yellow color blindness
- All cones, but **S cones** (blue) are less sensitive



Color Oracle



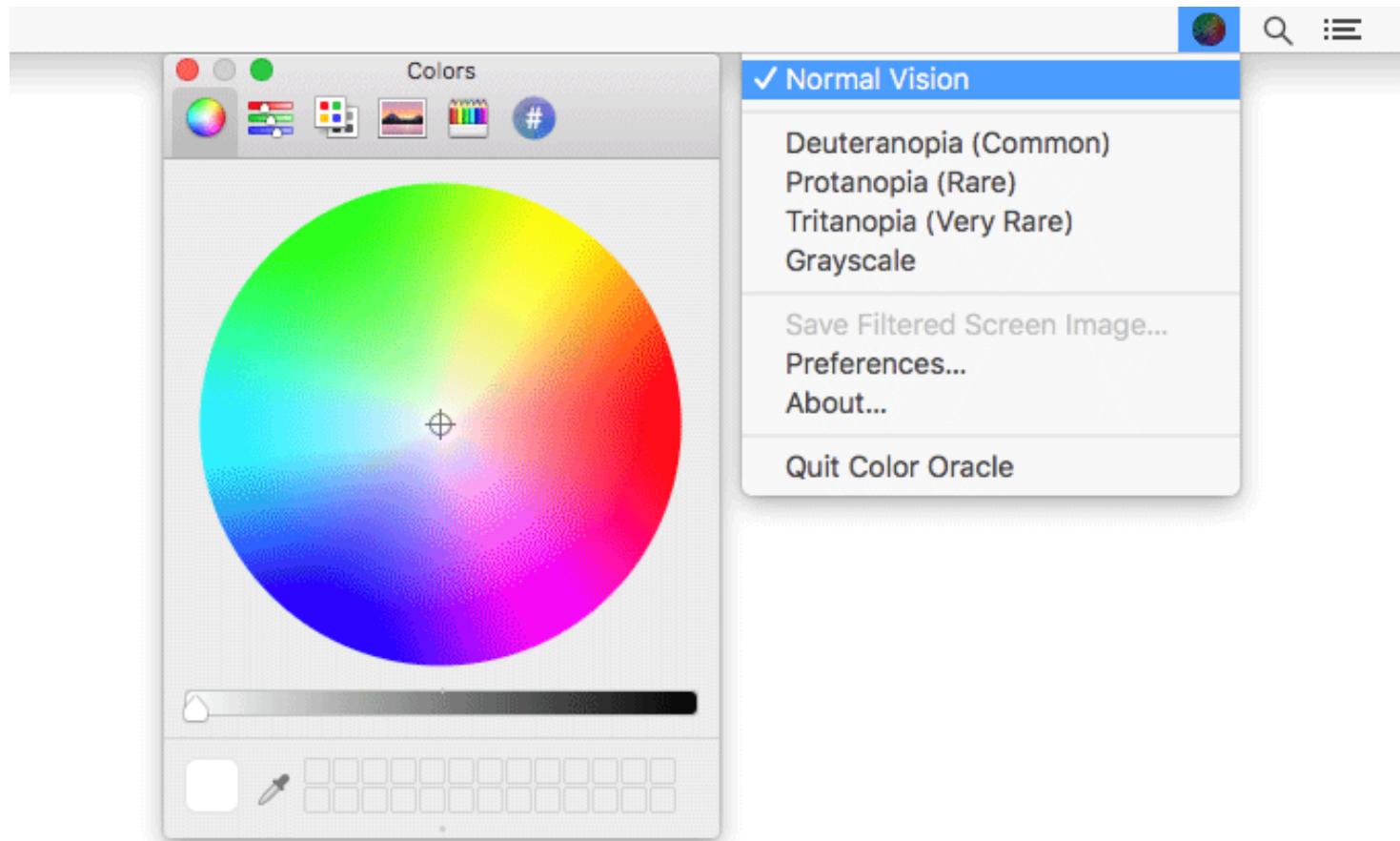
Color Oracle

Design for the Color Impaired

Color Oracle is a free color blindness simulator for Windows, Mac and Linux. It takes the guesswork out of designing for color blindness by showing you in real time what people with common color vision impairments will see.



Color Oracle applies a full screen color filter to art you are designing, independently of the software in use. Eight percent of all males are affected by color vision impairment – make sure that your graphical work is readable by the widest possible audience.



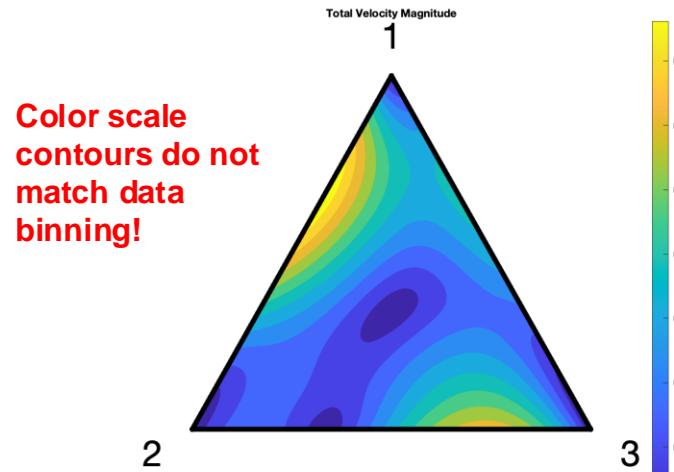
<https://colororacle.org/>
[Free]

Example color maps: the good, bad, caution



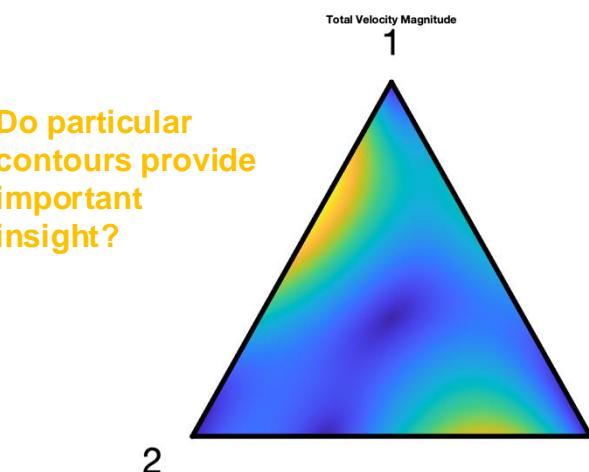
Replicator Equation

Colored by **total growth rate** where $g = [0, \infty]$



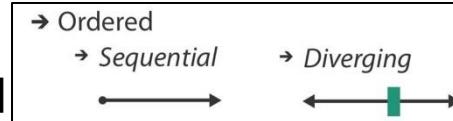
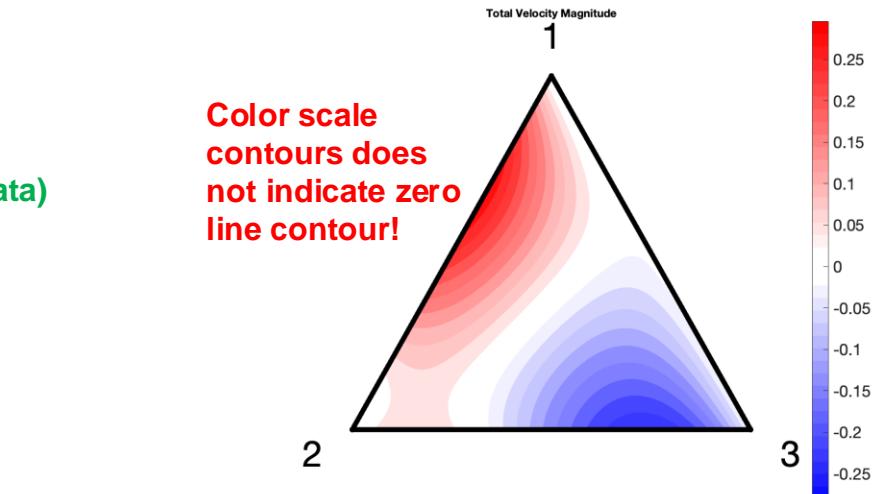
Color scale is ~monotonic luminance (sequential data)

Color scale includes meaningful bounds



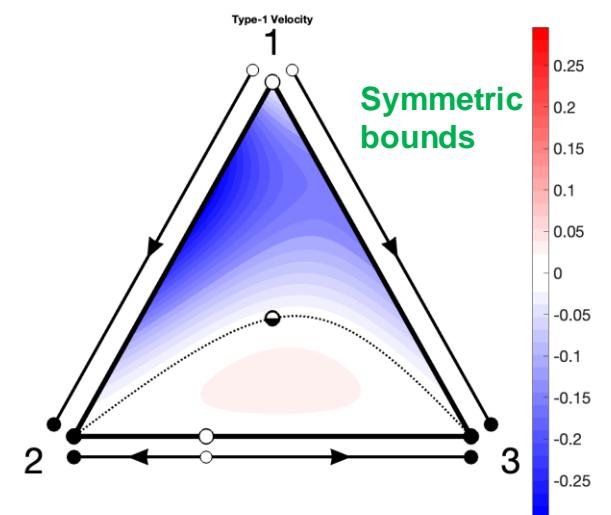
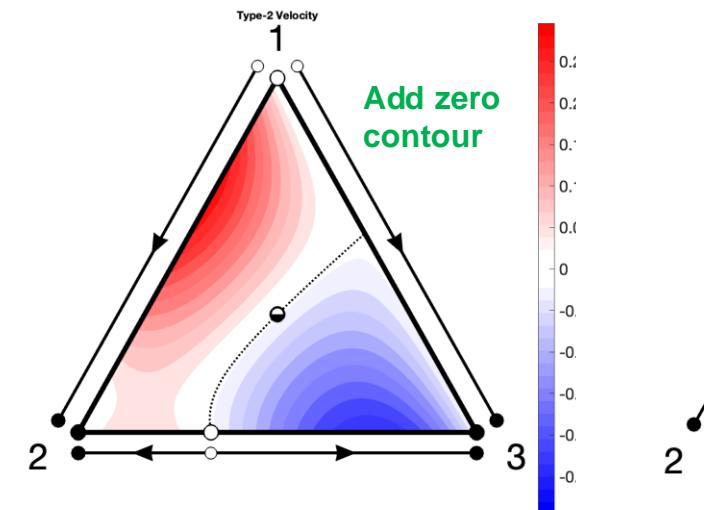
Replicator Equation

Colored by **relative growth rate** where $g = [-\infty, \infty]$



Color scale is ~diverging luminance (diverging data)

Color scale has symmetric bounds, eg $[-0.25, 0.25]$



Final thoughts



Create a library of colors,
colormaps across platforms



EDIT Lab Library

.m black.m
.m blue.m
.m categoricalColors.m
.m ColorRange.m
.m gold.m
.m green.m
.m pink.m
.m purple.m
.m red.m
.m white.m
.m ycbcrColors.m
.m yellow.m

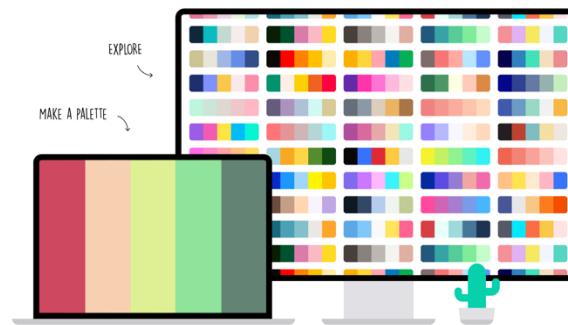
Ordinal color maps

<https://colors.co/>

The super fast color palettes generator!

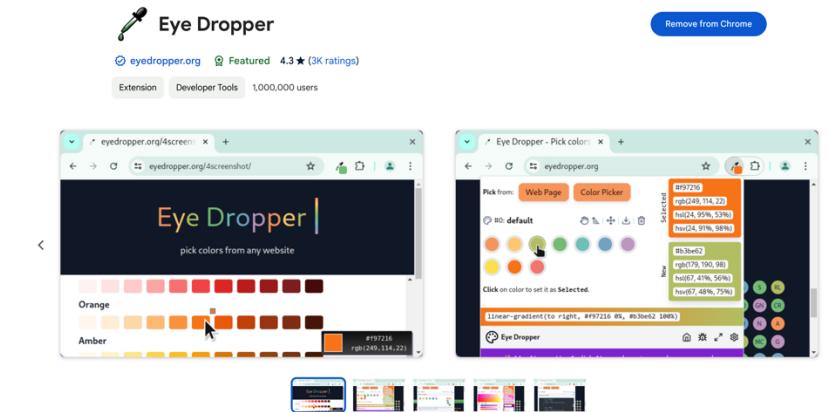
Create the perfect palette or get inspired by
thousands of beautiful color schemes.

[Start the generator!](#)
[Explore trending palettes](#)
[#1 Product of the Month](#)



Eyedropper on Chrome

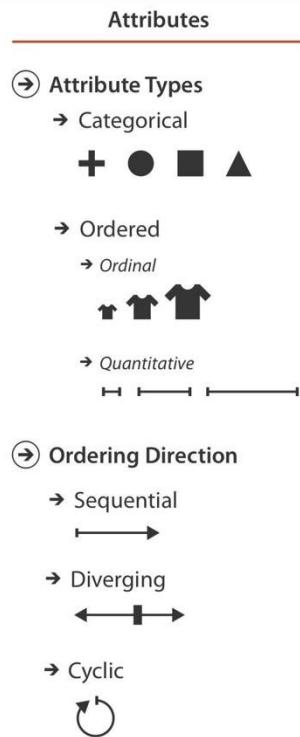
[link](#)



Data

Data Types and Attributes

Datasets, Data Types, and Attributes



Categorical

- No implicit order
- Often still hierarchical

Ordinal

- Implicit order
- Often not arithmetic (e.g. **shirt size, letter grade**)

Sequential

- Homogeneous/continuous range from min to max
- Arithmetic / quantitative data (e.g. **height**)

Diverging

- Two continuous sequences in opposite directions from zero
- Arithmetic / quantitative data (e.g. **elevation**, below/above sea level)

Cyclic

- Ordinal or Sequential data which is periodic
- Ordinal or Arithmetic / quantitative data (e.g. **time**)

Recommended Colormaps:

Categorical

- Constant Luminance (L^*), Varied Hue

Ordinal

- Constant Luminance (L^*), Varied Hue

Sequential

- Perceptually Linear Luminance (L^*)

Diverging

- Diverging, but Perceptually Linear Luminance (L^*)

Cyclic

- If ordinal:
 - Constant Luminance (L^*), Varied Hue
- If Arithmetic:
 - Perceptually Linear Luminance

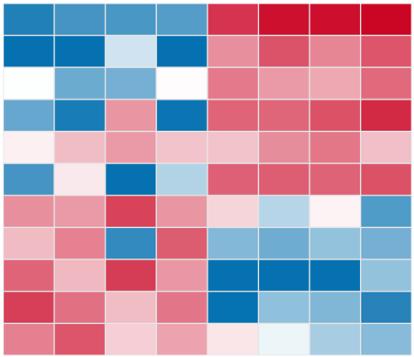
Example: Morpheus



Morpheus

<https://software.broadinstitute.org/morpheus/>

R package also available.



MORPHEUS

Versatile matrix visualization and analysis software

View your dataset as a heat map, then explore the interactive tools in Morpheus. Cluster, create new annotations, search, filter, sort, display charts, and more.

30,000+ users

100,000+ matrices analyzed.

If you use Morpheus for published work, please cite:

Morpheus, <https://software.broadinstitute.org/morpheus>

How to structure the data for input to Morpheus

Each column is a patient

ID	1	2	3	4	5	6	7	8	9	9	10	11	11	12	13	14	15
Age	77	33	55	50	25	60	45	46	52	52	20	51	51	20	67	55	55
DiseaseState	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer	Cancer
BCR::ABL1/like	BCR::ABL1	Negative	BCR::ABL1	Negative	BCR::ABL1	BCR::ABL1	Negative	BCR::ABL1	Negative	BCR::ABL1							
Sex	F	M	F	F	F	M	M	F	F	F	F	F	F	M	M	M	F
Specimen	PB	BM	BM	PB	BM	PB	PB	BM	BM	PB	BM	PB	PB	PB	PB	BM	BM
MRD(flow)	0	0	0	1	0	0	1	0	1	1	0	NaN	NaN	1	0	1	1
MRD(molecular)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	NaN	1	1
Relapse	1	1	NaN	NaN	NaN	NaN	0	0	NaN	0							
Kinetics	3	1	NaN	NaN	NaN	NaN	NaN	NaN	2	NaN	NaN	NaN	NaN	2	2	NaN	NaN
PersistantMRD	0	1	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	NaN	0	0	NaN	NaN	1
IKZF1	1	0	0	NaN	1	0	1	1	1	1	0	NaN	NaN	0	NaN	0	0
CDKN2A/B	1	0	0	NaN	0	1	0	1	0	0	0	NaN	NaN	0	NaN	1	0
RAS/RAF/MAPK or FLT3	0	0	0	NaN	0	0	0	0	0	0	0	NaN	NaN	0	NaN	1	0
TP53	0	0	0	NaN	0	0	0	0	0	0	0	NaN	NaN	0	NaN	1	0
x1	0.9726	0.0055	0.0224	0.9414	0.8821	0.6491	0.2805	0.7898	0.9987	0.7482	0.031	0.9685	0.5142	0.8164	0.0011	0.0135	0.999
x2	0.0152	0.2454	0.9115	0.0566	0.1013	0.0991	0.6897	0.2097	0	0.2501	0.9684	0.0311	0.4858	0.1831	0.1629	0.9744	0.000
x3	0.0029	0.7435	0.0661	0.0002	0.0087	0.1356	0.0128	0.0003	0	0.0005	0.0006	0.0002	0	0.0001	0.8272	0.0116	0.000
x4	0.0093	0.0056	0	0.0018	0.0079	0.1162	0.017	0.0002	0.0013	0.0012	0	0.0002	0	0.0004	0.0088	0.0005	0.000

1st column is the name of each data type

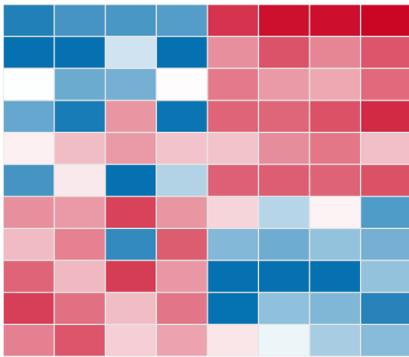
Example: Morpheus



Morpheus

<https://software.broadinstitute.org/morpheus/>

R package also available.



MORPHEUS

Versatile matrix visualization and analysis software

View your dataset as a heat map, then explore the interactive tools in Morpheus. Cluster, create new annotations, search, filter, sort, display charts, and more.

30,000+ users

100,000+ matrices analyzed.

If you use Morpheus for published work, please cite:

Morpheus, <https://software.broadinstitute.org/morpheus>

How to structure the data for input to Morpheus

Upon upload, user specifies **annotations** vs **data**

Open

Click the table cell containing the first data row and column.

■ Data Matrix
■ Column Annotations
■ Row Annotations

MRD(molecule)	1	1	1	1	1	1
Relapse	1	1	NaN	NaN	NaN	NaN
Kinetics	3	1	NaN	NaN	NaN	NaN
PersistantMR0	1	NaN	NaN	NaN	NaN	NaN
IKZF1	1	0	0	NaN	1	0
CDKN2A/B	1	0	0	NaN	0	1
RAS/RAF/MA0 or FLT3	0	0	NaN	0	0	0
TP53	0	0	0	NaN	0	0
x1	0.9726	0.0055	0.0224	0.9414	0.8821	0.6491
x2	0.0150	0.2454	0.9115	0.0566	0.1013	0.0991
x3	0.0029	0.7435	0.0661	0.0002	0.0087	0.1356
x4	0.0093	0.0056	0	0.0018	0.0079	0.1162

OK Cancel

Datasets, Data Types, and Attributes

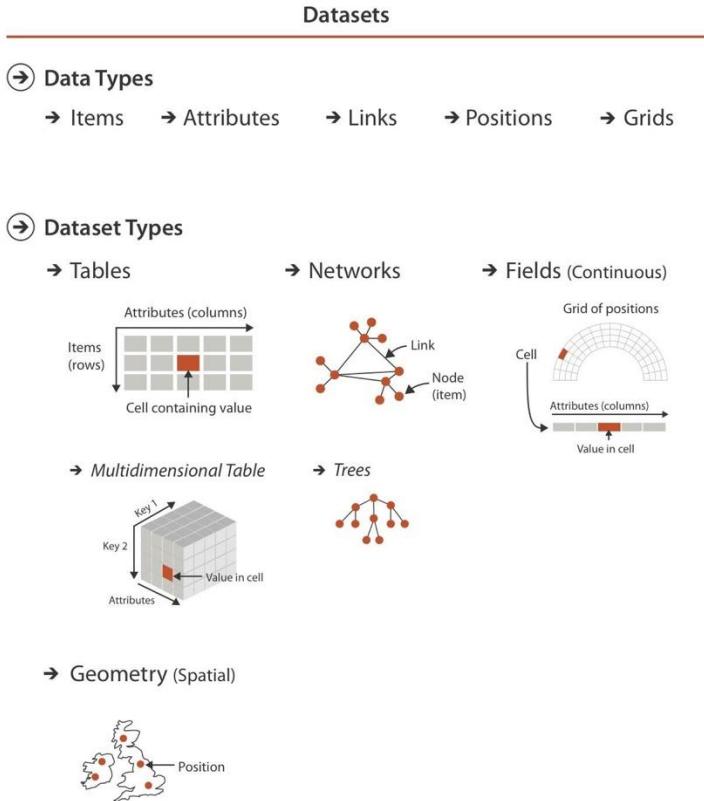


Table
➤ Row / Column

Network
➤ Node / Edge

Field
➤ Grid / Value

Treen
➤ Parent / Children

Geometry
➤ Position / Attribute

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box	0.72	7/17/07
32	7/16/07	2-High	Medium Box	0.6	7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69	6/4/05	4-Not Specified	Small Pack	0.44	6/6/05
69	6/4/05	4-Not Specified		0.6	6/6/05
70	12/18/06	5-Low		0.59	12/23/06
70	12/18/06	5-Low		0.82	12/23/06
96	4/17/05	2-High		0.55	4/19/05
97	1/29/06	3-Medium		0.38	1/30/06
129	11/19/08	5-Low		0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

quantitative
ordinal
categorical

Figure 2.9. The order table with the attribute columns colored by their type; none of them is a key.

Spatial

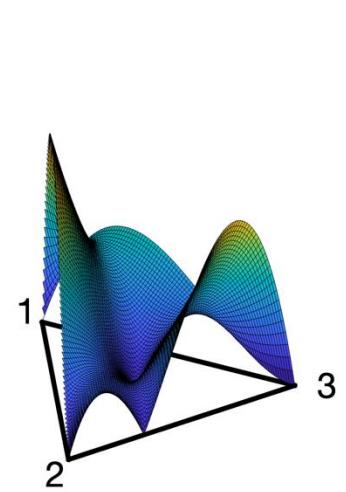
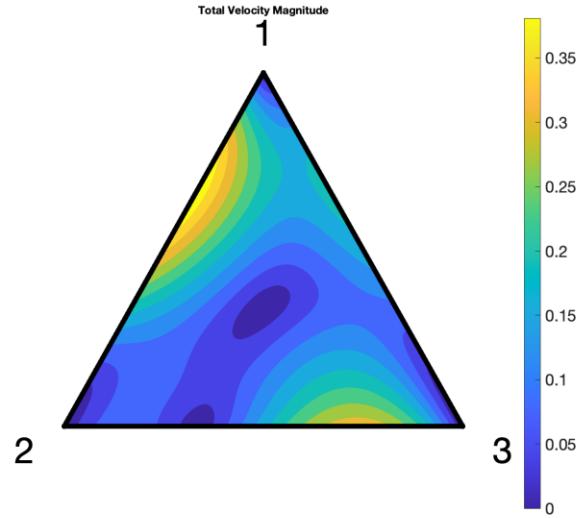
Encoding information using space

Rule of thumb: no unjustified 3d!



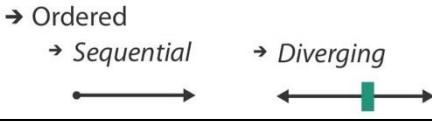
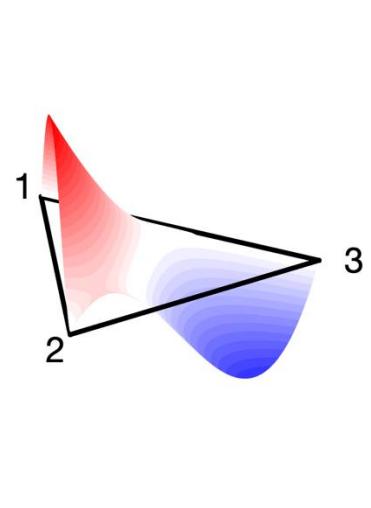
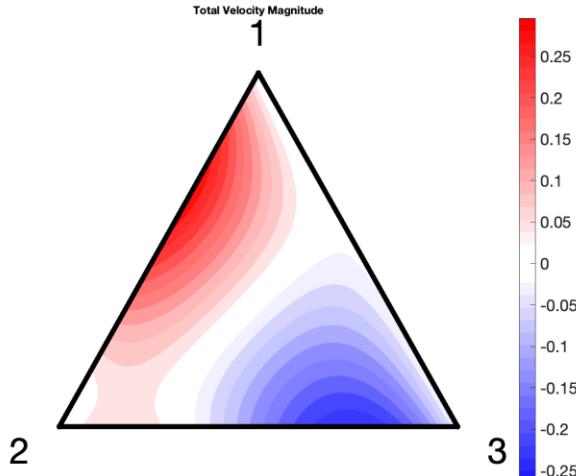
Replicator Equation

Colored by **total growth rate** where $g = [0, \infty]$



Replicator Equation

Colored by **relative growth rate** where $g = [-\infty, \infty]$

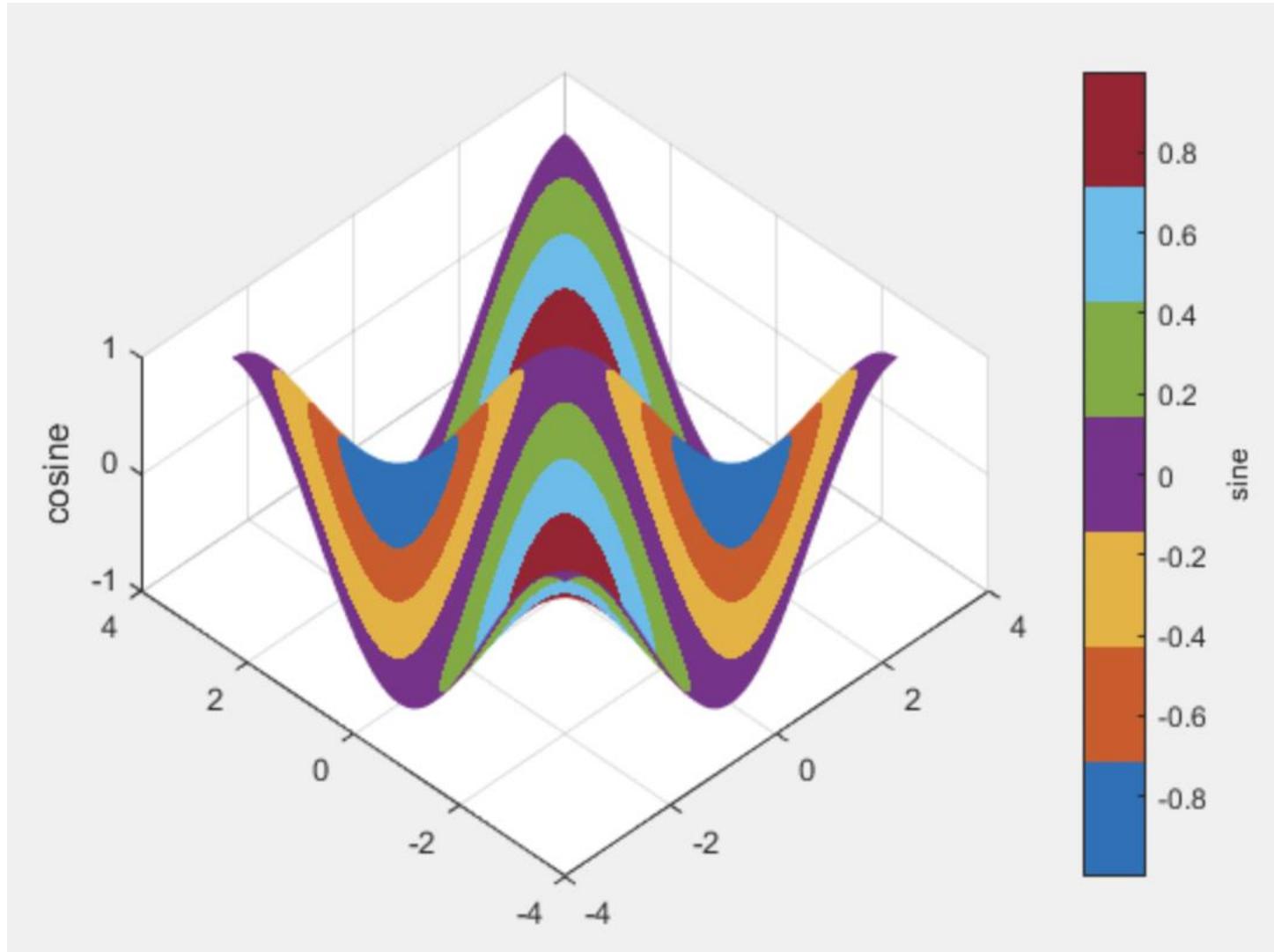


Any information gain?

Any information loss?

Duplicity:
height & color are
encoding same
information

Just for fun...



- Ordinal / Categorical color map for continuous data
- Unjustified 3d (height vs color)
- Binning is too sparse?
- Unlabelled x- & y- axes

Interactive

Visualization with real-time interaction

Visualization with real-time interaction



MathBox.js
Steven Wittens

Note: these slides use cutting edge CSS 3 and WebGL features. It is recommended to use Google Chrome to view them.

Making things with Maths

Steven Wittens

unconed
<http://acko.net>

<https://acko.net/files/fullfrontal/fullfrontal/wdcode/online.html>

Forecasting Health Site
Peter Kuhn

Dana and David Dornoff
College of Letters, Arts and Sciences

The Kuhn Laboratory
The Bridge@USC
Understand the Human Body
Improve the Human Condition

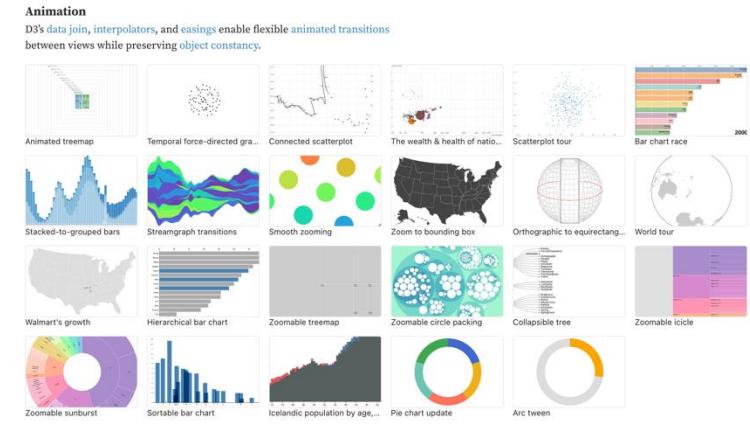
WEB-BASED PRESENTATION, ANALYSIS, & FORECASTING OF DISEASE PROGRESSION

Longitudinal datasets provide a unique view of the evolution of a disease over both time and space. With the use of datasets such as these, we are building complex data visualization tools that can be easily understood by all researchers, medical providers, and patients. The system that we are setting up will provide a way for various institutions to share and integrate their own data with others in an effort to better understand the dynamics of each disease. The forecasting models that we are building from these sets will help us to answer several questions such as:

1. Is cancer progression predictable? Is it random? Can we quantify the predictability/randomness?
2. How does treatment affect the pathways of progression? How does it affect the speed of progression?
3. What are the most common pathways from liver disease to liver cancer? Who is most at risk?
4. Will metastatic adrenal gland resection in lung cancer provide similar benefits as metastatic liver resection in colon cancer?

<https://kuhn.usc.edu/forecasting/>

ObservableHQ
runs on d3js library

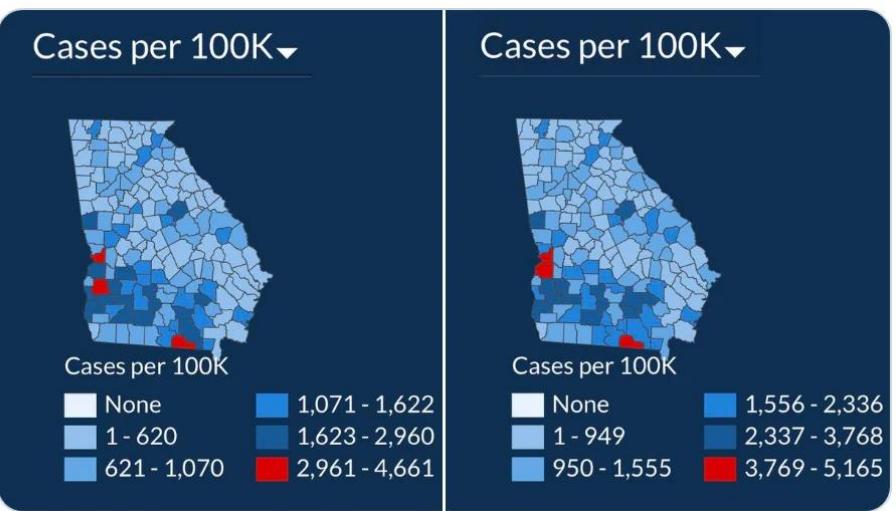
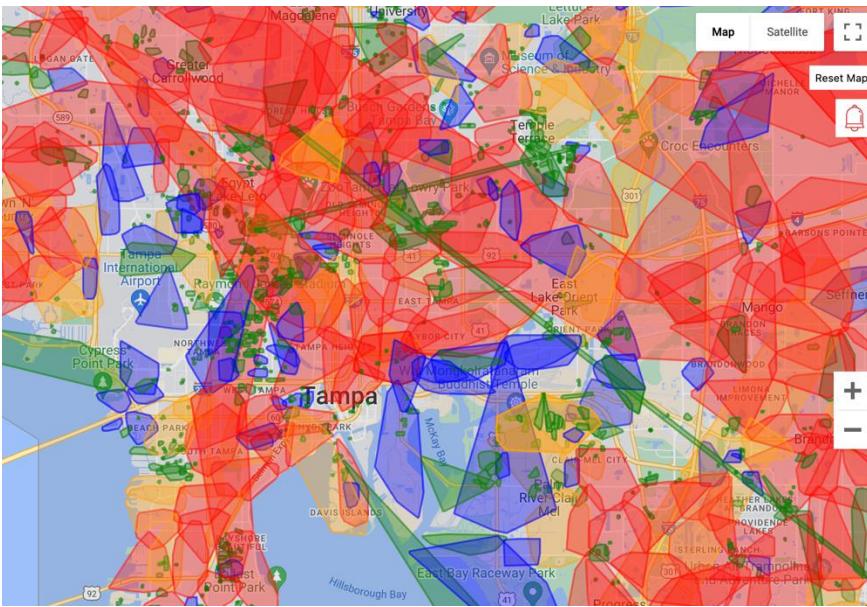


<https://observablehq.com/>
&
<https://d3js.org/>

Submitted figures

Let's judge some figures..

Submissions



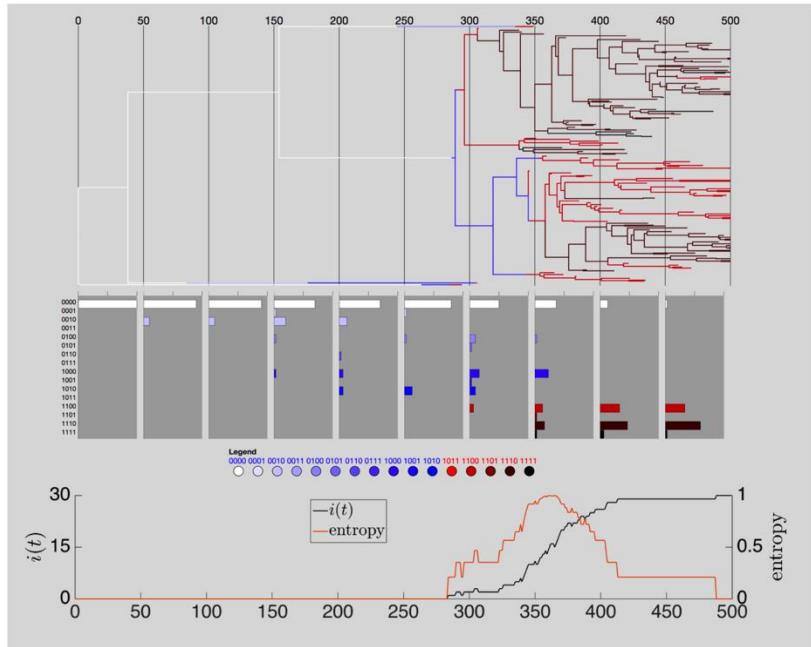
When Europeans fly nest

Average age at which young people leave the parental household, 2022

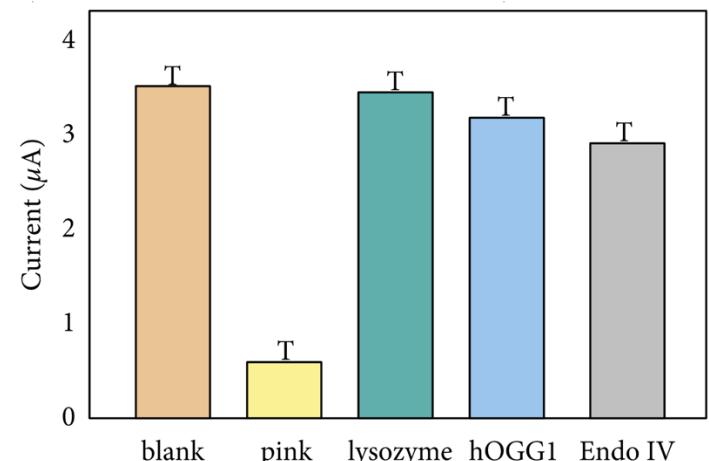
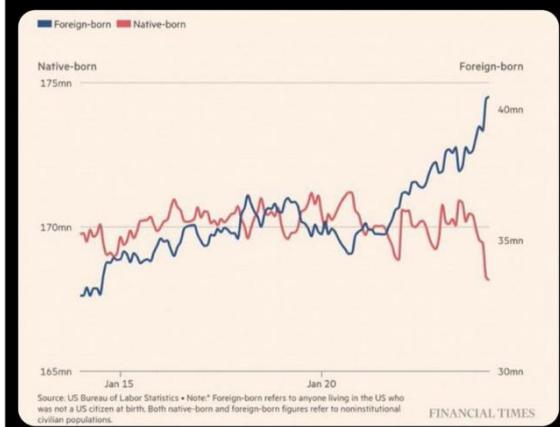
■ below 23 ■ 23-25 ■ 25-27 ■ 27-29 ■ 29-31 ■ over 31



E

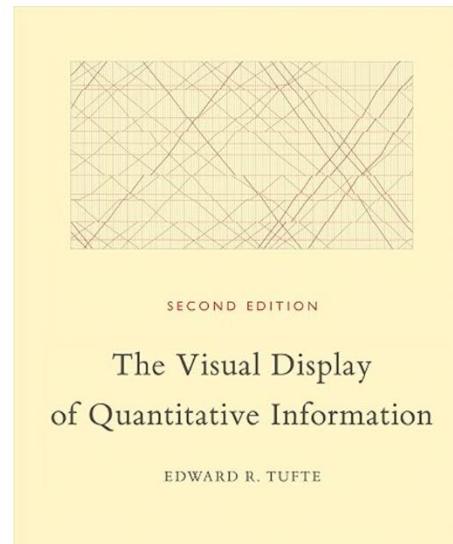
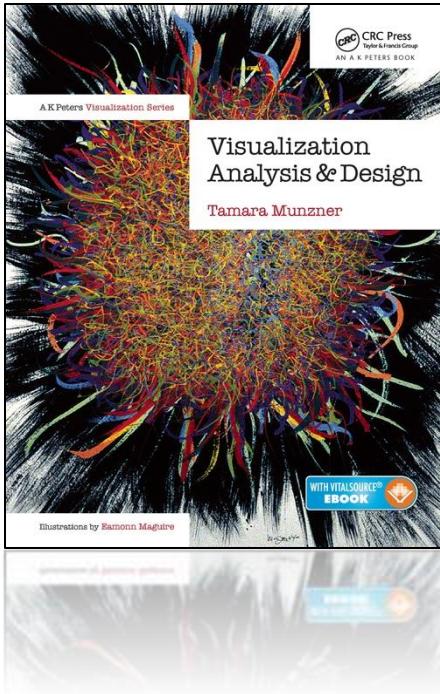


Americans are quite literally being replaced in the jobs market



Special thanks to Kit Gallagher, Kayode Olumoyin, Rick Miles

Recommended reading



Visualization, analysis and design

William Playfair, 1786

“As to the propriety and justness of representing sums of money, and time, by parts of space, tho' very readily agreed to by most men, yet a few seem to apprehend that there may possibly be some deception in it, of which they are not aware...”

