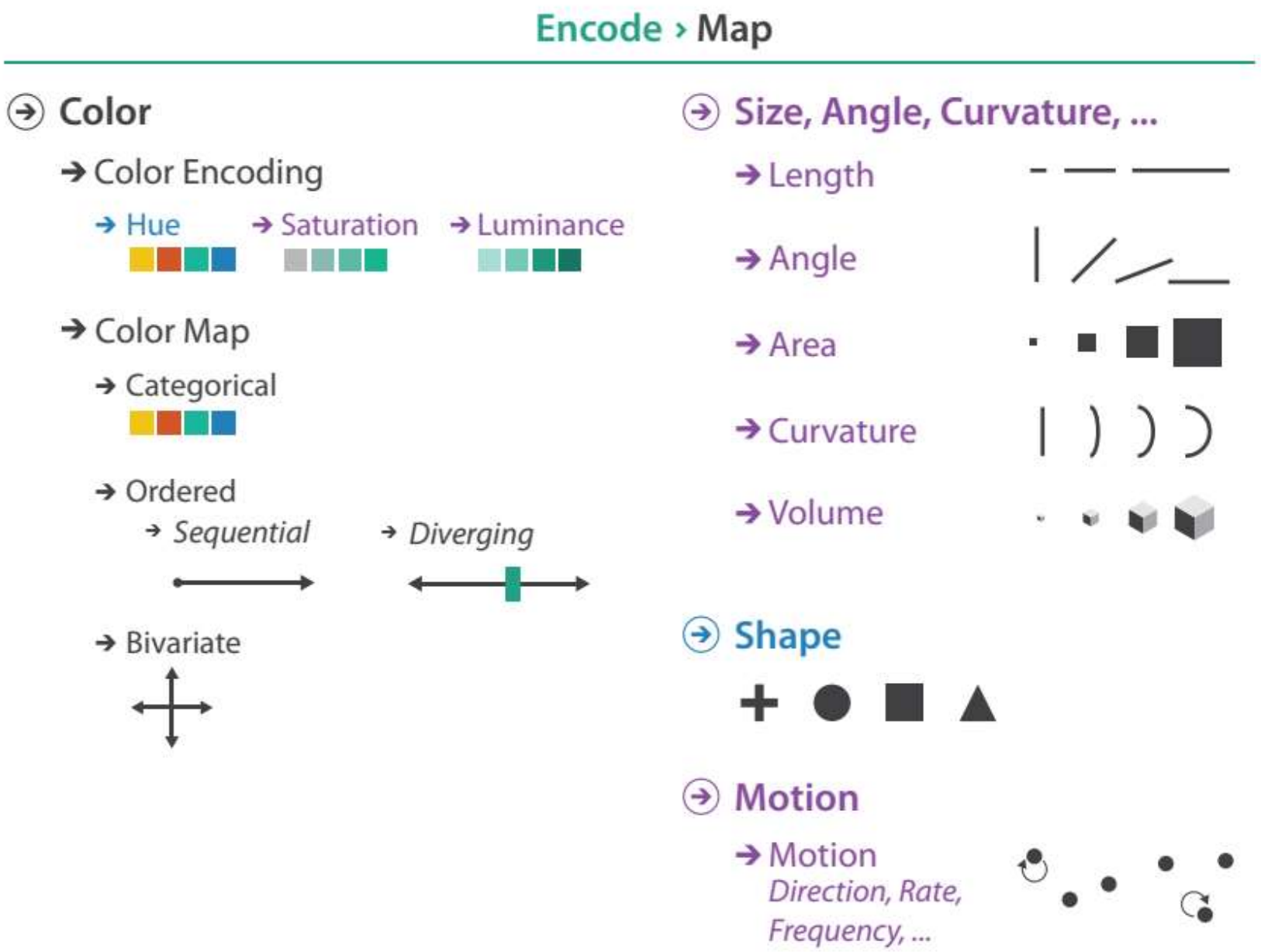
Data Visualization

Chapter 06

Color

Khoa Công Nghệ Thông Tin
Trường Đại Học Khoa Học Tự Nhiên

Big picture


Encode › Map
Color
Color Encoding
Hue
Saturation
Luminance
Color Map
Categorical
Ordered
Sequential
Diverging
Bivariate
Size, Angle, Curvature, ...
Length
Angle
Area
Curvature
Volume
Shape
Motion
Motion
Direction, Rate,
Frequency, ...

Color spaces

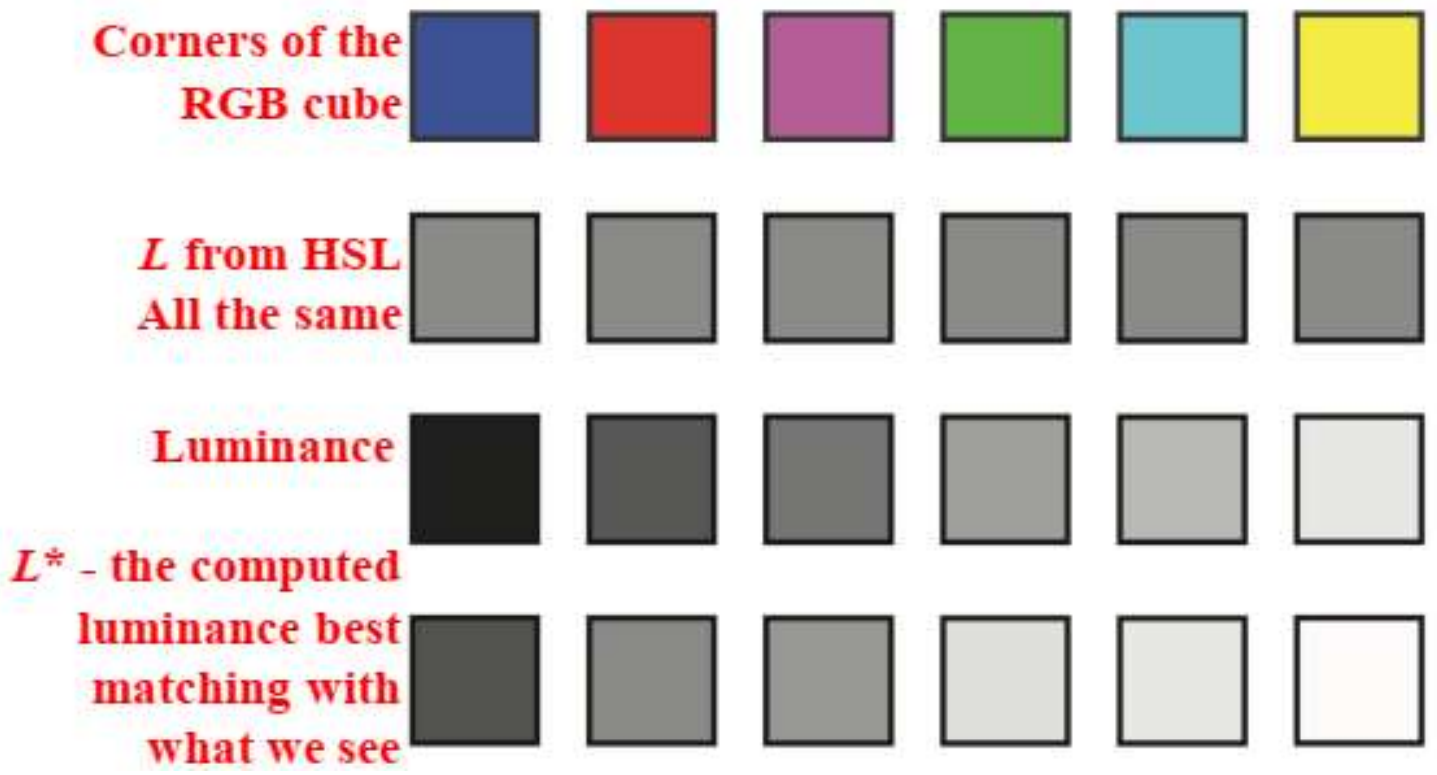
- The **color space** of what colors the human visual system can detect is three-dimensional
- Some of these color spaces for computer manipulation, and the others are a better match with the characteristic of human vision
- The most common color space is the **RGB** system
  - Each color is represented by the weights assigned to the **red**, **green** and **blue** channels.

Color spaces

- HSL system is
- HSL system
  - Requires HUE, Saturation, and Lightness axes
  - Each color is defined by the weights assigned to the hue, saturation and lightness axes
    - Hue axis captures what we normally think of as pure color
      - Not mixed with pure white and pure black
    - Saturation axis is the amount of white mixed with a pure color
      - Pink = red + pure white
    - Lightness axis is the amount of black mixed with a color
- HSV space is similar, where V is for grayscale and is linear to lightness L

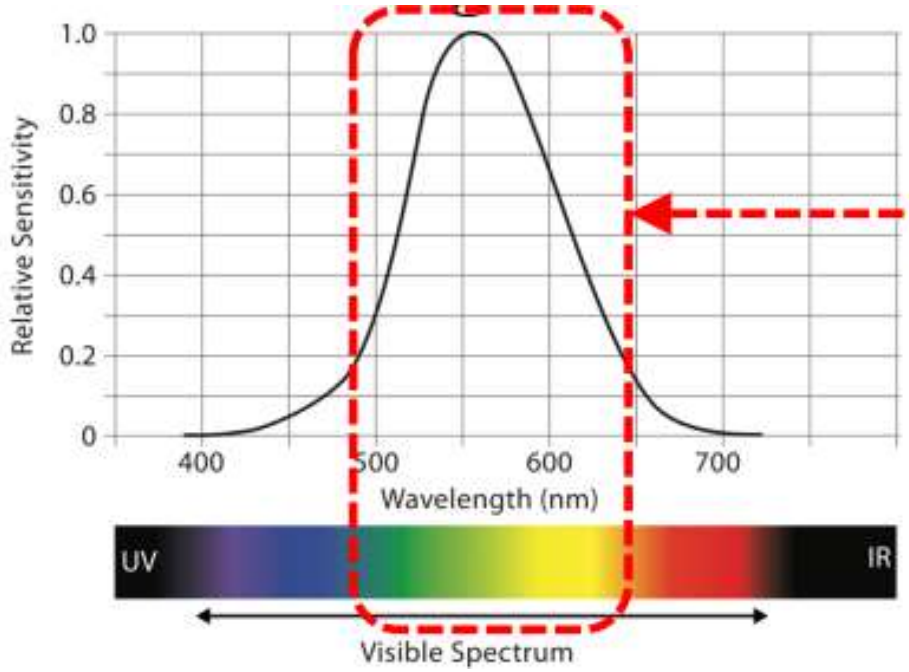
Color spaces

- HSL space does not truly reflect how we perceive color
  - Lightness L is wildly different from how we perceive luminance



Color spaces

- The amount of luminance that human perceives depends on the wavelength.
- Human are much more sensitive to middle wavelengths of green and yellow than to the outer wavelengths of red and blue



We are more sensitive to middle wavelengths of green and yellow.

Luminance, Saturation and Hue Channels

- ***Luminance***, ***saturation*** and ***hue*** can be used as a ***magnitude*** channel and as an ***identity*** channel
  - Magnitude channel
    - Luminance
    - Saturation
  - Identify channel
    - HUE

Luminance, Saturation and Hue Channels

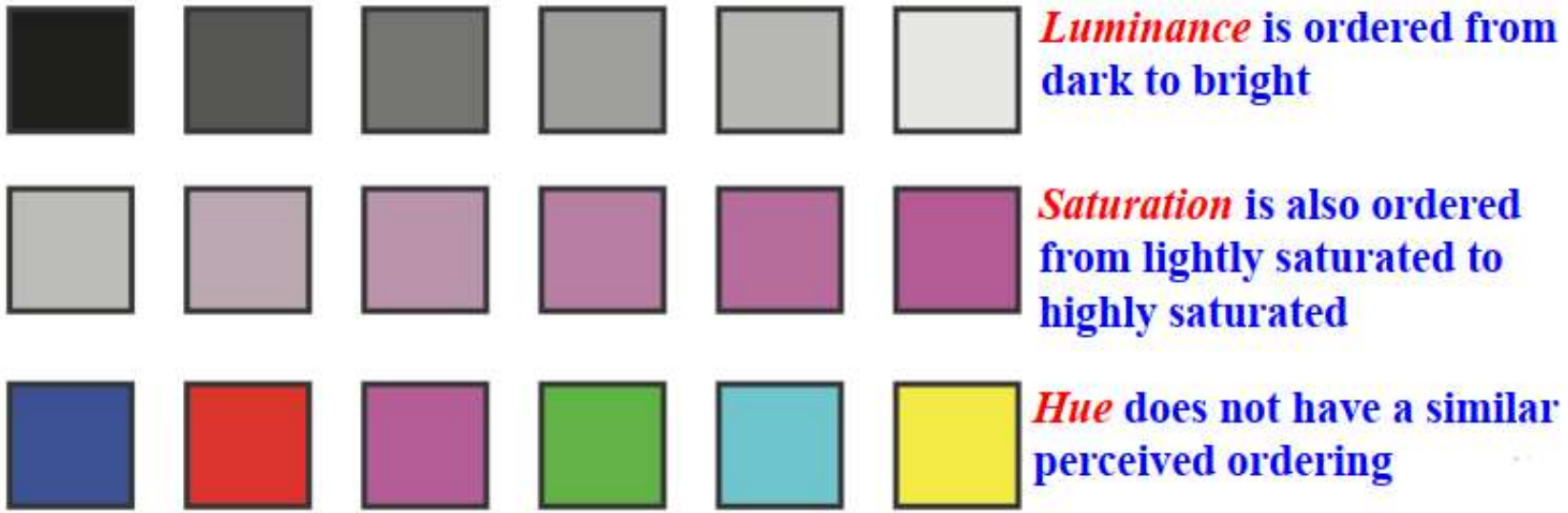
- ***Magnitude channel***
  - ***Luminance***
    - Suitable for ordered data
    - Low-accuracy channel in terms of perceiving whether noncontiguous regions have the same luminance because of contrast effects
    - Limit grayscale bins to no more than 4
  - Saturation
- ***Identify channel***
  - HUE

Luminance, Saturation and Hue Channels

- ***Magnitude channel***
  - ***Luminance***
  - ***Saturation***
    - Suitable for ordered data
    - Low-accuracy for noncontiguous regions
    - The number of discriminate steps for saturation is low: around 3 bins
    - Interact with the size channel
    - Use bright, highly saturated colors so that small regions are more distinguishable
- ***Identify channel***
  - ***HUE***

Luminance, Saturation and Hue Channels

- ***Identify channel***
  - ***HUE***
    - Very effective for categorical data and showing groups (*highest ranked after spatial position*)
    - HUE and size can interact: hue is harder to distinguish in small regions
    - Make fine distinctions in hue for contiguous regions
    - Very limited discriminability between separate regions
    - Hue does not have an implicit perceptual ordering



Transparency

- ***Transparency*** is a channel related to the other three color channels
- ***Transparency*** coding *interacts strongly* with ***luminance*** and ***saturation***
- It can be used with hue encoding with a very small number of discriminate steps, most frequently just two
- Most often ***transparency*** is used with superimposed layers.

Colormaps

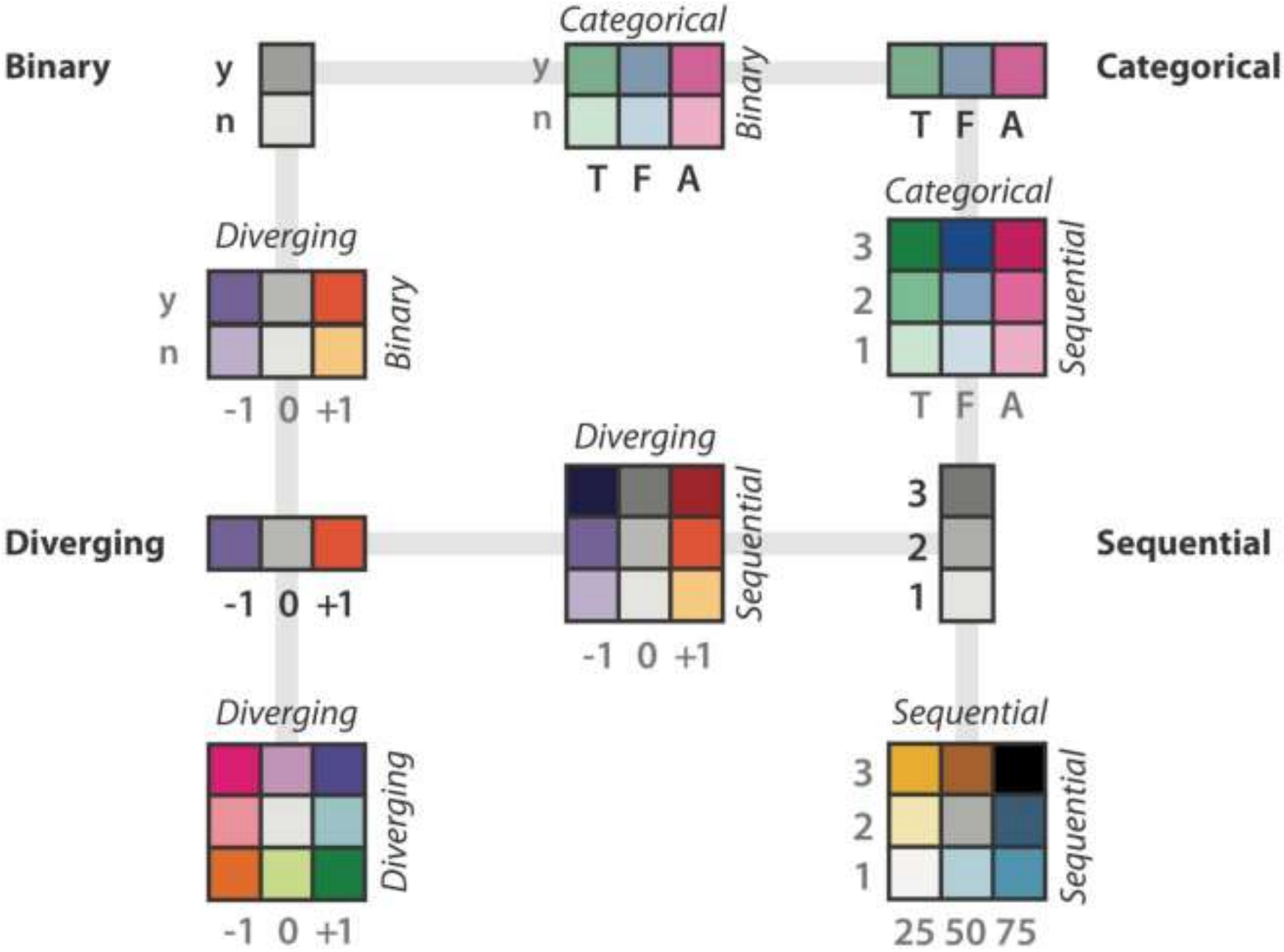
Colormaps

- Categorical colormaps
- Ordered colormaps
- Bivariate colormaps
- Colorblind-safe colormaps design

Colormaps

- A ***colormap*** defines a mapping between colors and data values (*visual encoding with colors*)
- ***Colormap***s can be *categorical* or *ordered*
  - Ordered colormaps can be *sequential* or *diverging*
- ***Colormap***s can also be a *continuous* range of values or *segmented* into discrete bins of colors
  - *Continuous* ***colormap***s are used for showing *quantitative* attributes
  - *Segmented* ***colormap***s are suitable for *categorical* data
  - For *ordinal* data
    - *Segmented colormaps* will emphasize their *discrete* nature
    - While *continuous colormaps* will emphasize their *ordered* nature

Colormaps

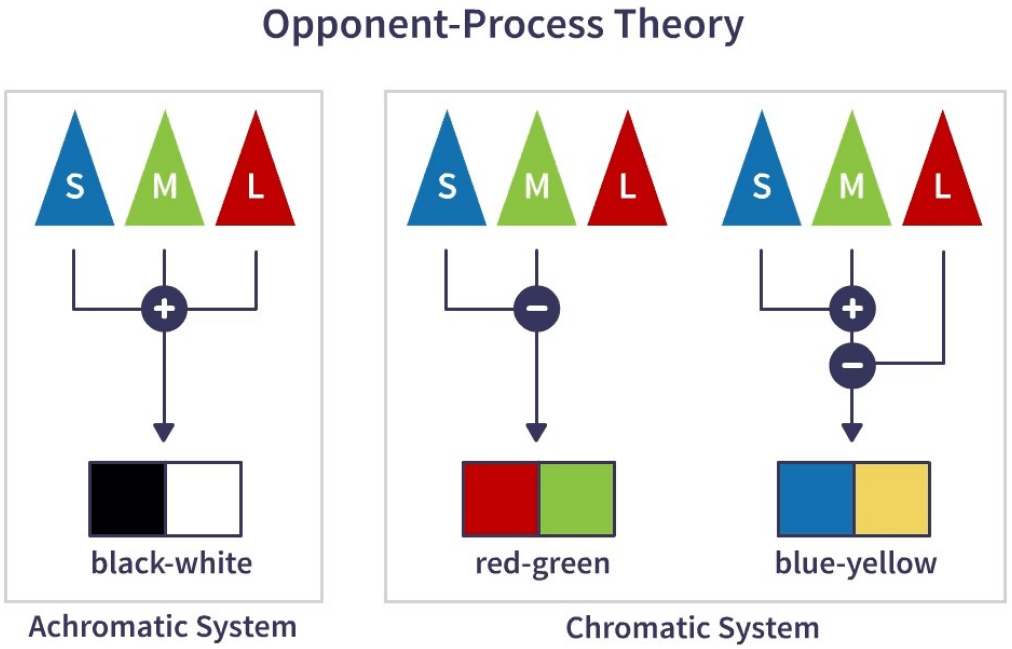

Binary
y
n
Categorical
y
n
Binary
T F A
T F A
Categorical
Diverging
y
n
Binary
-1 0 +1
Categorical
3
2
1
Sequential
T F A
Diverging
Diverging
-1 0 +1
Sequential
-1 0 +1
3
2
1
Sequential
Diverging
Diverging
-1 0 +1
Sequential
3
2
1
Sequential
25 50 75

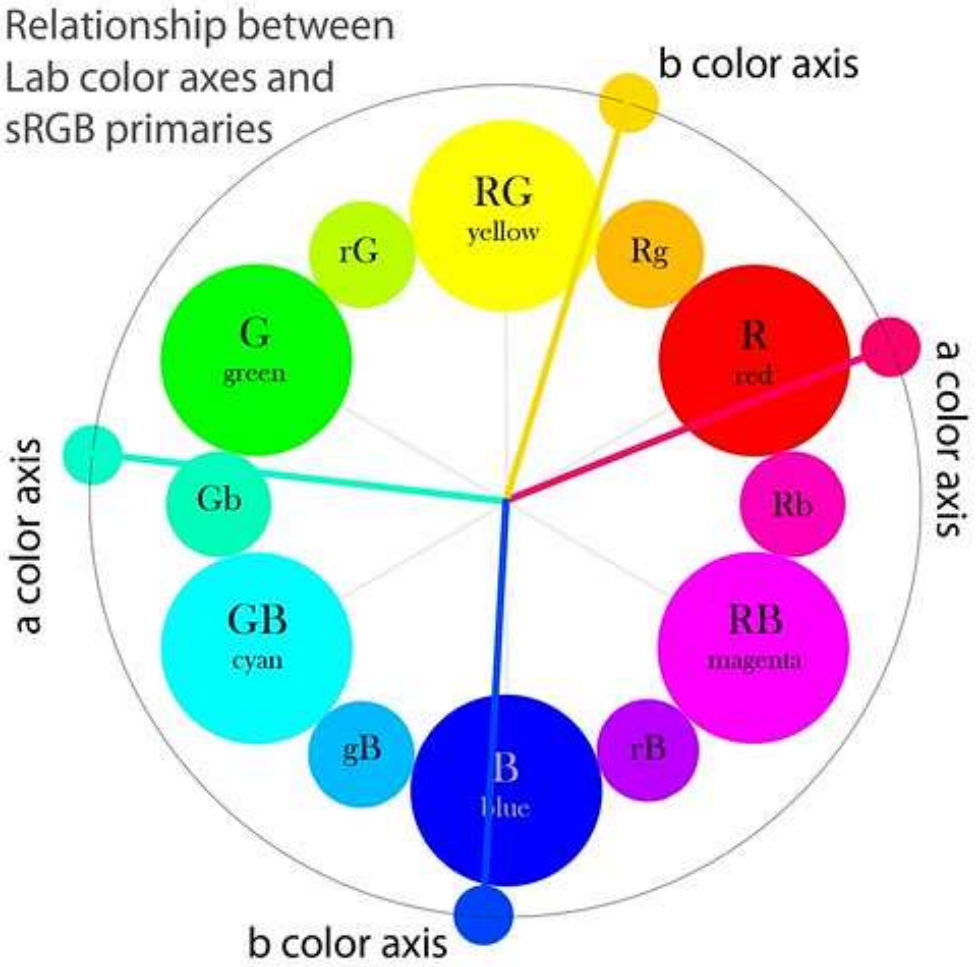
Categorical Colormaps

- *Categorical **colormap*** uses color to encode categories and groupings
- *Categorical **colormap***s are typically designed by using color as an integral identity channel to encode a single attribute, rather than to encode three completely separate attributes with the three channels of HUE, saturation and luminance.
- The number of discriminable colors for coding *small separated regions* is limited to between *6 and 12 bins*

Categorical Colormaps

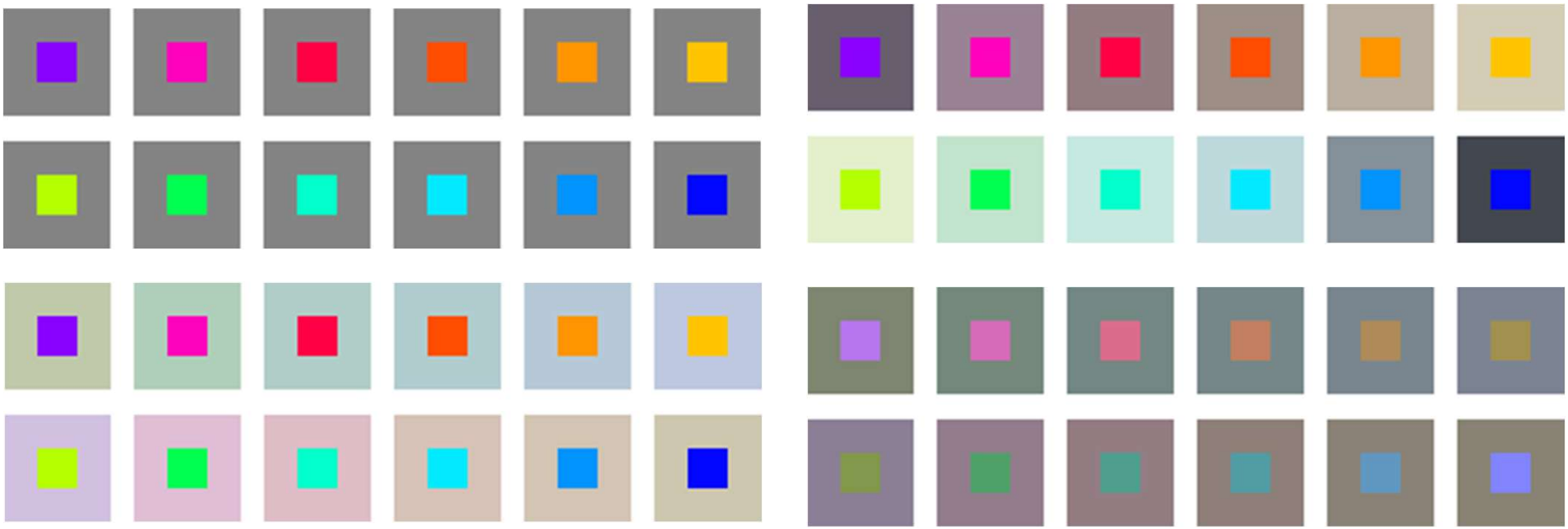
- A good set of initial choices are the fully saturated and easily nameable colors, which are also the opponent color axes: red, blue, green, and yellow.





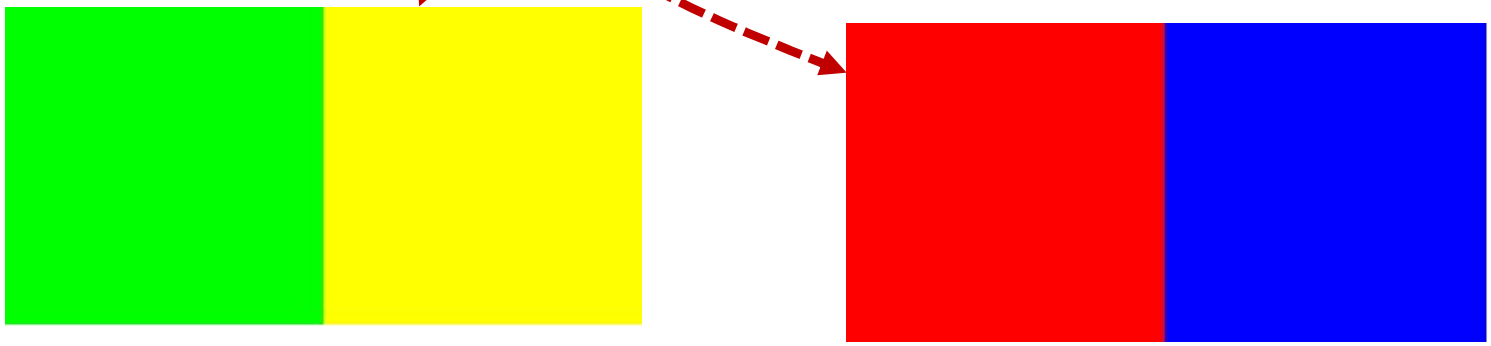
Categorical Colormaps

- Be careful with ***luminance contrast***: for some uses, the colors should be close in luminance to avoid major differences in salience and to ensure that all can be seen against the background.



Categorical Colormaps

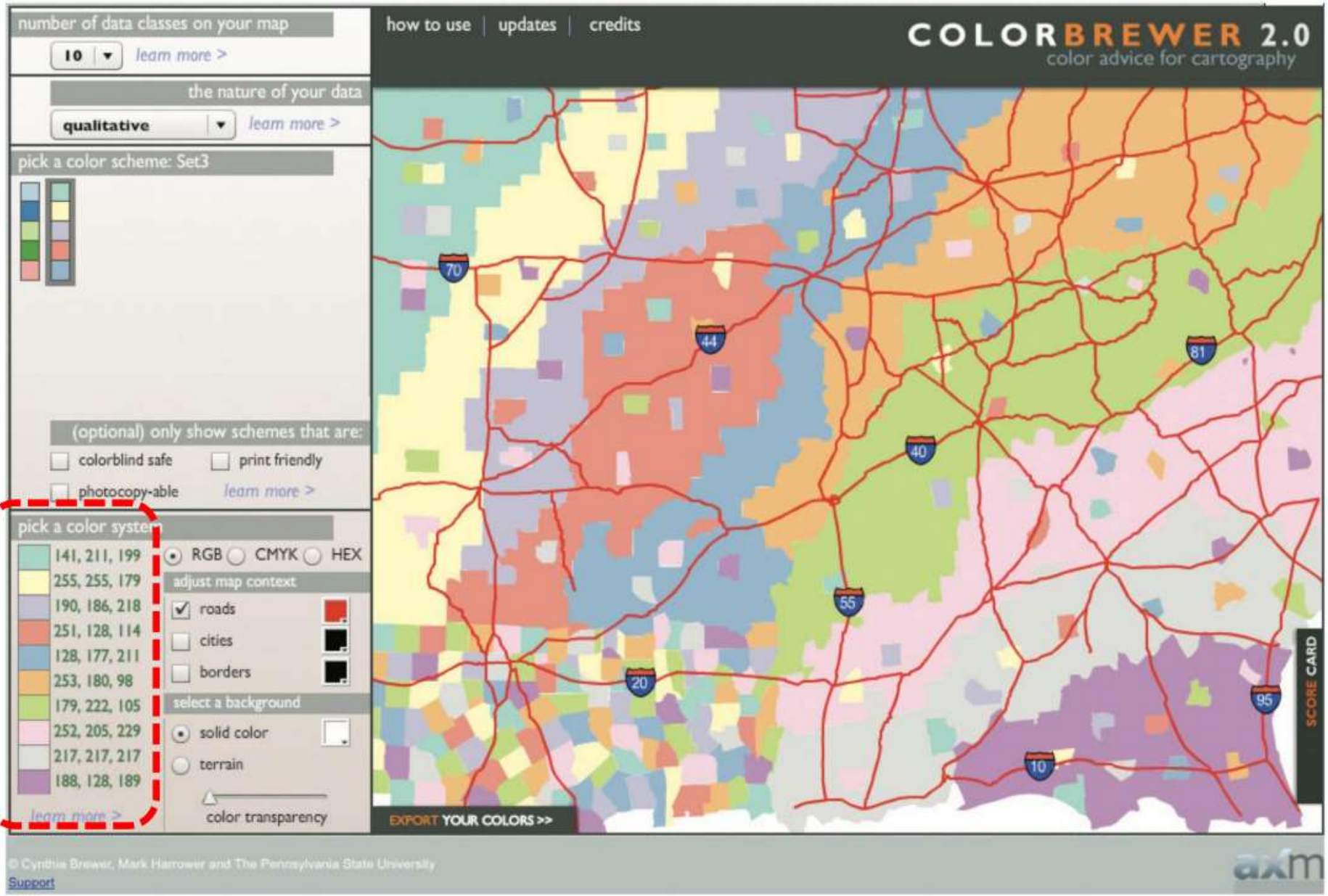
- **Example**:
  - Fully saturated green and yellow will have much less *luminance contrast* against a white background than red and blue



Categorical Colormaps

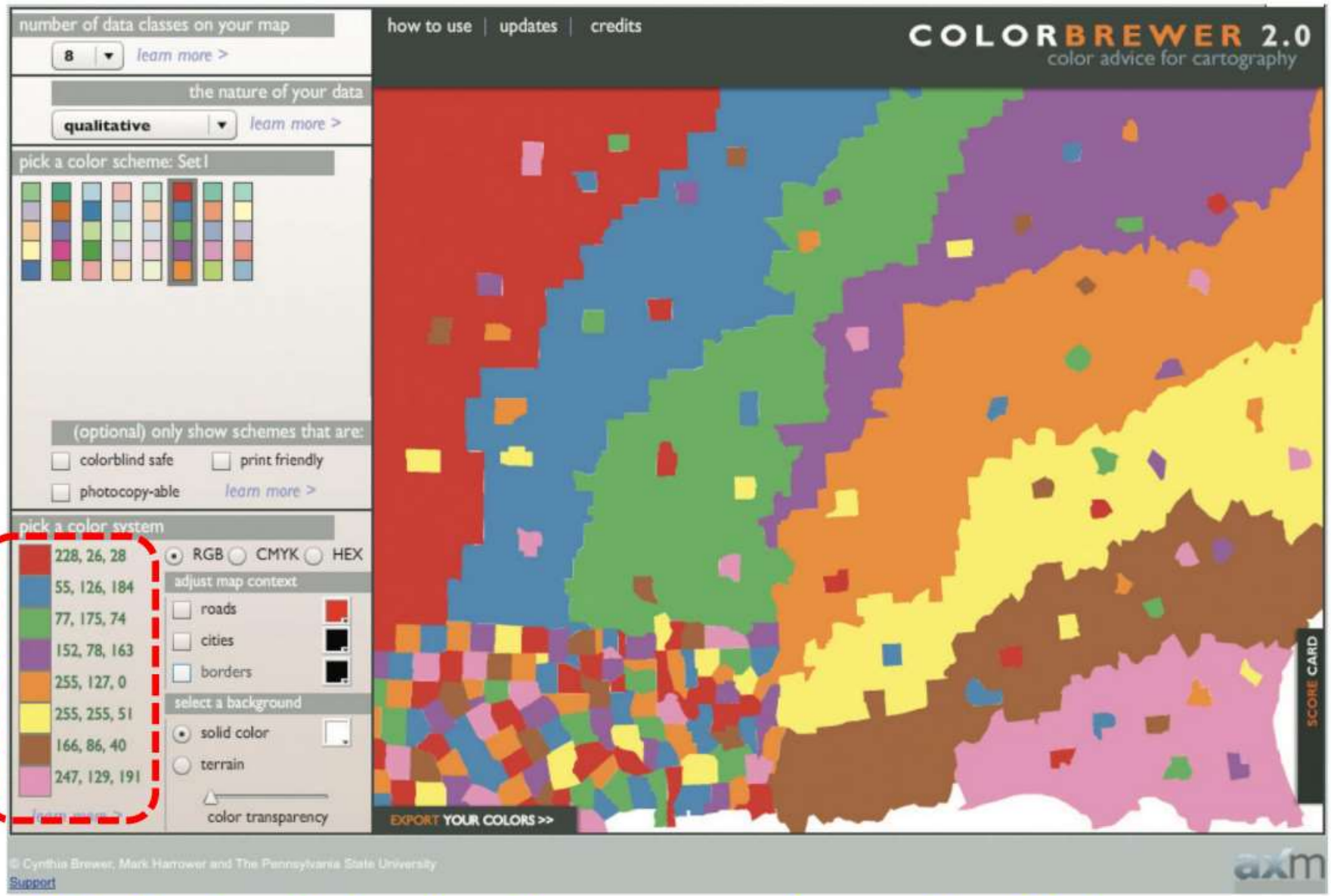
- For other uses, colors should be sufficiently different in luminance that they can be distinguished even in black and white
- ***Colormap***s for
  - *Small region*s such as lines should be highly saturated
  - *Large region*s such as areas should have low saturation
- An appropriate ***colormap*** may depend on the *mark type* being used.

Categorical Colormaps - Example



This 10-color low saturation map works well with large areas.

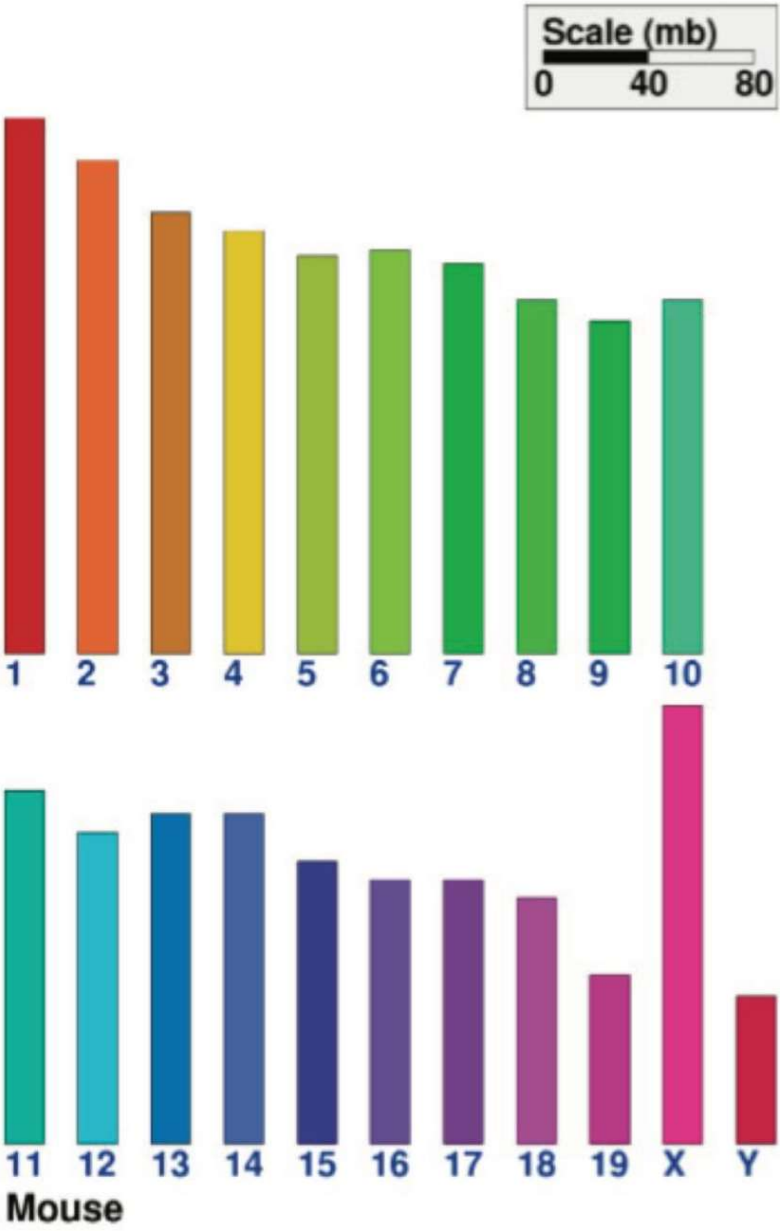
Categorical Colormaps - Example



This 8-color high saturation map works poorly with large areas. It would be better suited for small regions.

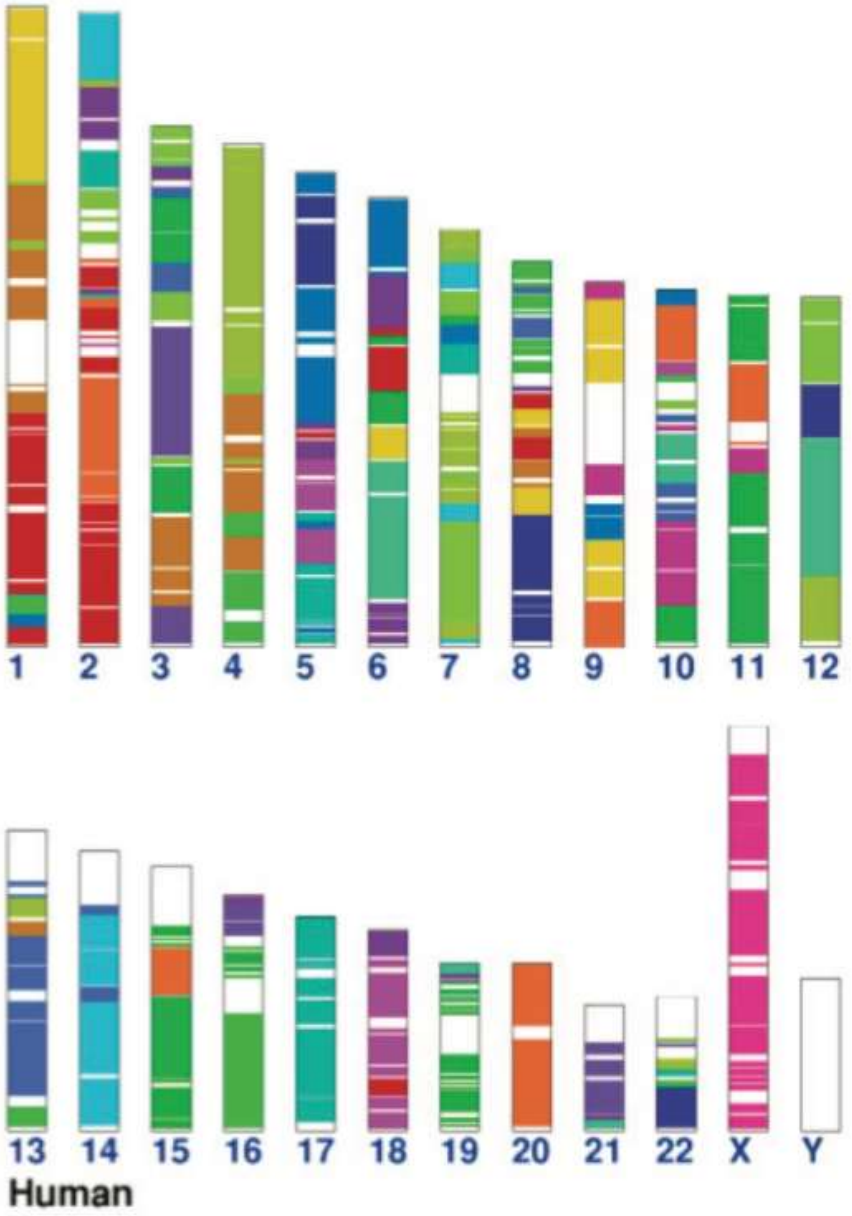
Categorical Colormaps - Example

- The 21 colors used as an index for each ***mouse chromosome*** can indeed be distinguished in large regions next to each other
- These colors act as a legend and an index because:
  - Regions are large
  - The most subtle differences are between regions that are right next to each other.

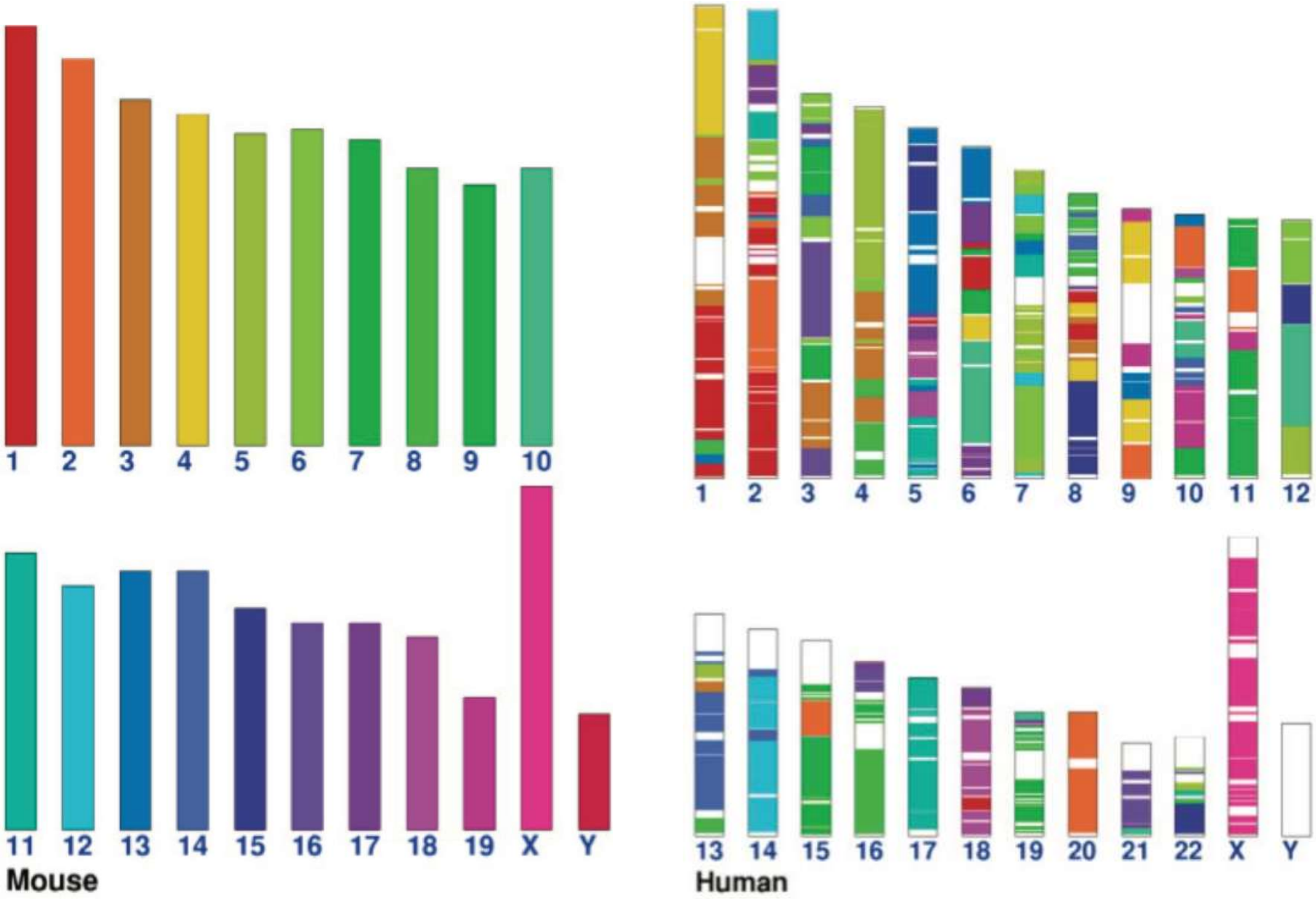


Categorical Colormaps - Example

- The regions of the *human chromosomes* that correspond to those in the *mouse chromosomes* have been colored to illustrate how genomic regions have moved around as the species evolved independently of each other after diverging from a common ancestor.



Categorical Colormaps - Example



Categorical Colormaps - Example

- ***Solve this problem***
  - *Transform the data* that take into account the nature of the data and task, so that each bin can be encoded with a distinguishable color
    - Find a possible ***hierarchical structure*** that can be used to derive meaningful aggregated groups
    - Filter the attributes to only encode a small set of the most important ones with color, and aggregate all the reset into a new category
  - *Use a different encoding idiom* that uses other visual channels instead of, or in addition to, the color channel alone

Ordered Colormaps

- An ***ordered*** colormap is appropriate for encoding *ordinal* or *quantitative* attributes
- The two major variants of ***continuous colormap***s for ordered data have expressiveness characteristics that should match up with the attribute type.
  - Sequential
  - Diverging

Ordered Colormaps

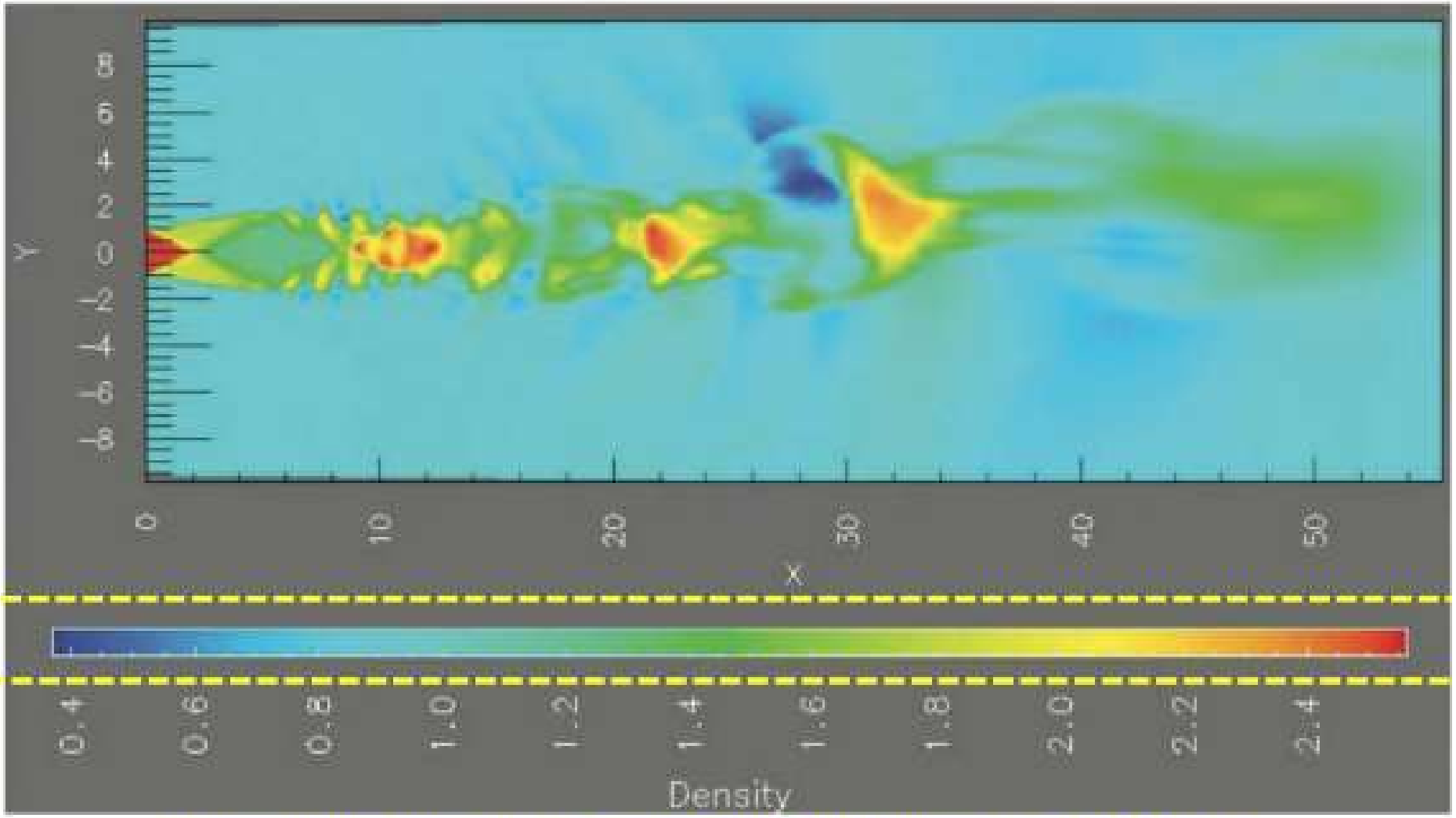
- A ***sequential*** colormap
  - Ranges from a minimum value to a maximum value
  - If only the *luminance* channel is used, the result is a grayscale ramp
    - When *hue* is added, one end of the map is a specific hue at full saturation and brightness
  - If *saturation* is a variable, the other end is pale or white
    - If *luminance* is the varying quantity, the other end is dark or black
- A ***diverging*** colormap
  - has two hues at the endpoints and a neutral color as a midpoint
  - This neutral color may be white, gray, or black, or a high-luminance color such as yellow.

Ordered Colormaps

- The number of unique HUEs to be used in ***continuous colormap***s depends on what level of structure should be emphasized:
  - The high-level structure
  - The middle range of local neighborhoods
  - Fine-grained detail

Ordered Colormaps

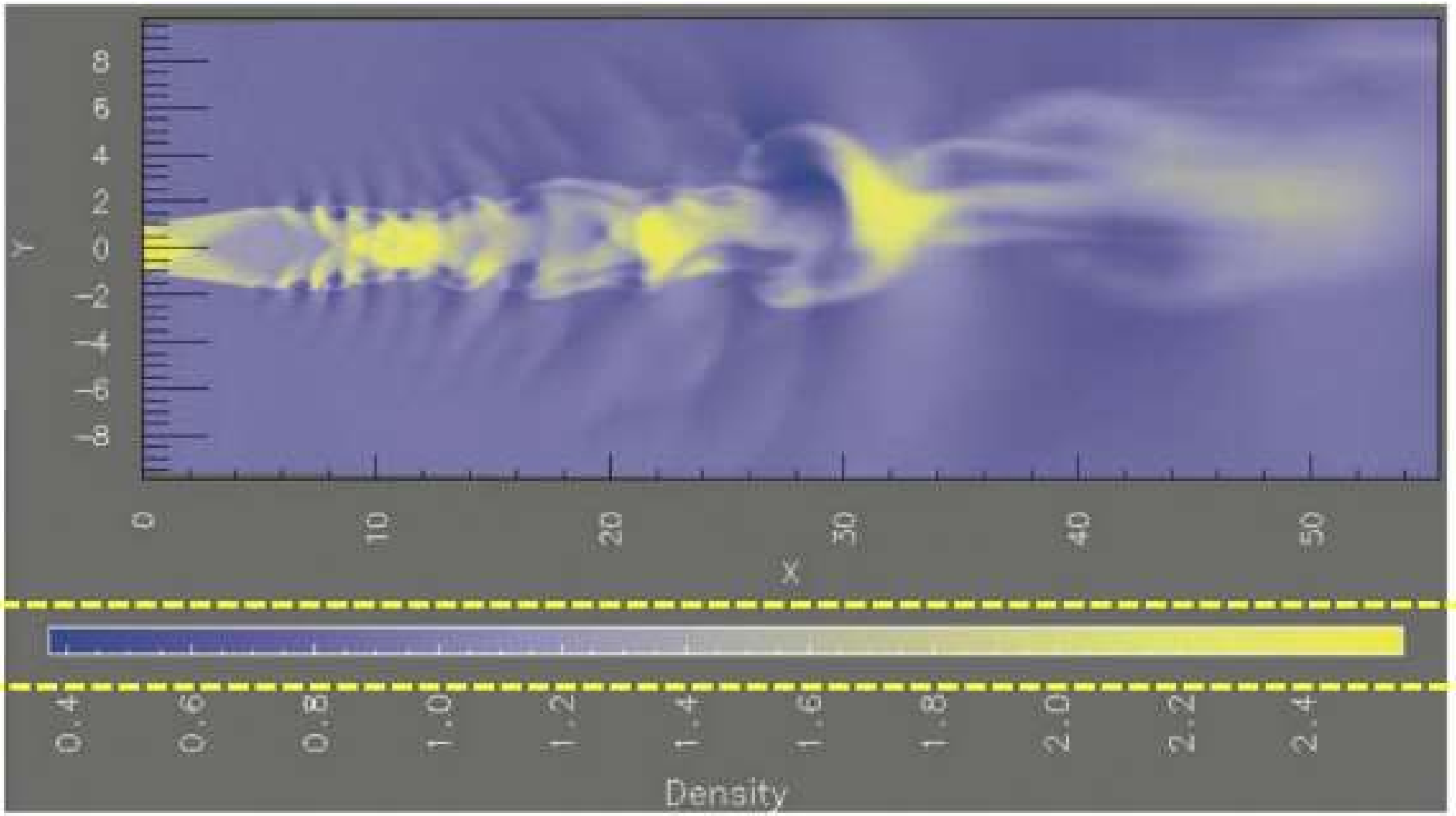
- Rainbow colormap



Use many hues as in a rainbow colormap to emphasize mid-scale neighborhood

Ordered Colormaps

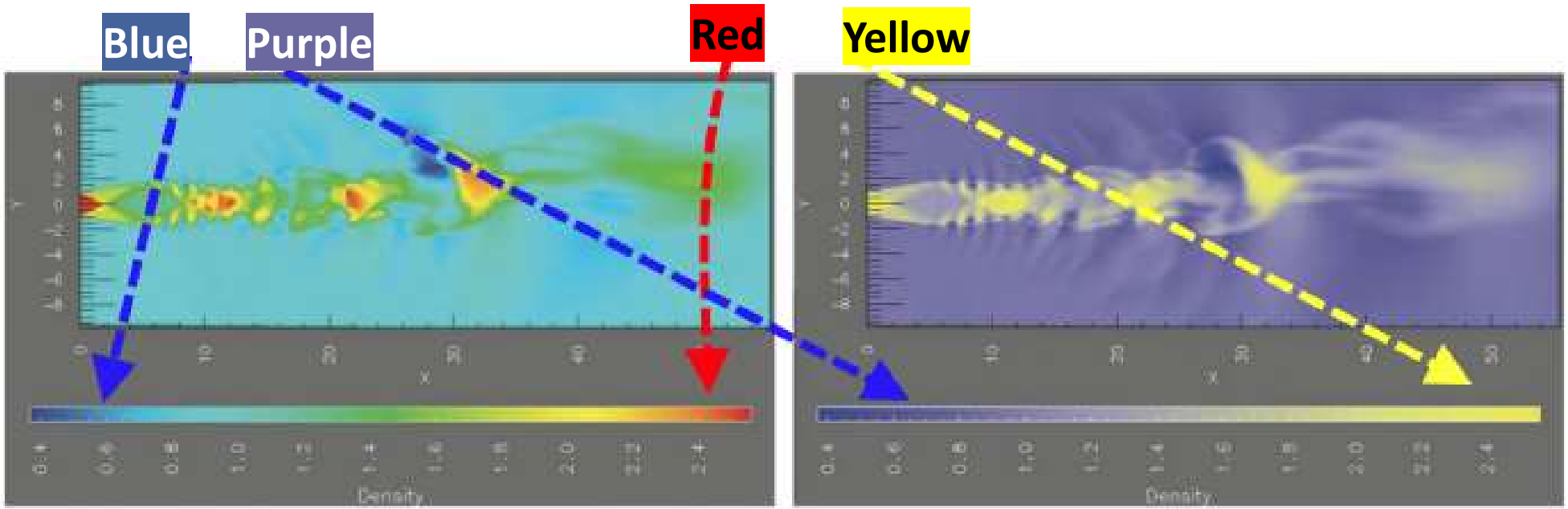
- Blue - yellow colormap



Use a 2-hue blue-yellow colormap to emphasize large-scale

Ordered colormaps

- With a ***rainbow*** colormap, people can easily discuss specific subranges because the differences are easily nameable
- With a colormap only using *saturation* or *luminance* changes, people can only talk about the blue side and yellow side.

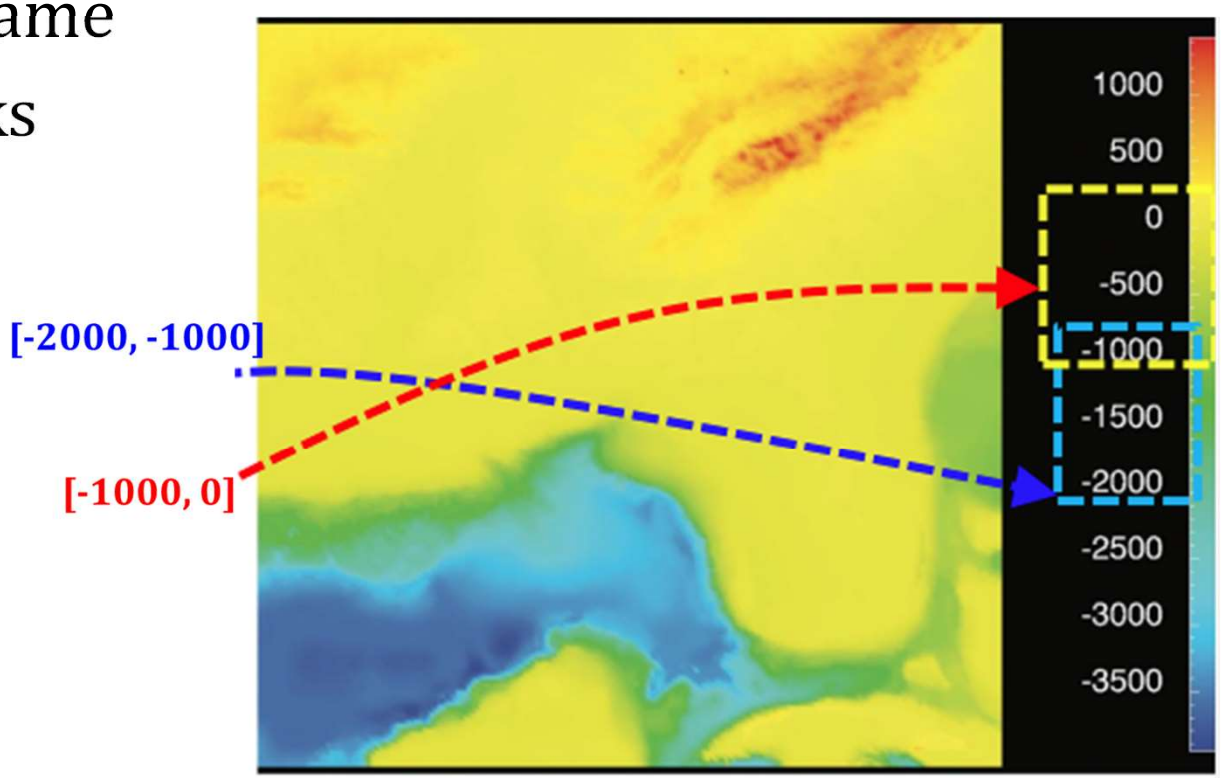


Ordered colormaps

- ***Rainbow*** colormaps, which are usually a default choice in many software packages, suffer from three serious problems.
    1. *Hue* is used to indicate order, despite being an identity channel
       → *without an implicit perceptual ordering*
    2. The scale is not perceptually linear: steps of the same size at different points in the colormap range are not perceived equally by our eyes
       → *Expressively mismatch*
    3. Fine details cannot be perceived with the hue channel; the luminance channel would be a better choice
       → *Accuracy mismatch*

Ordered colormaps

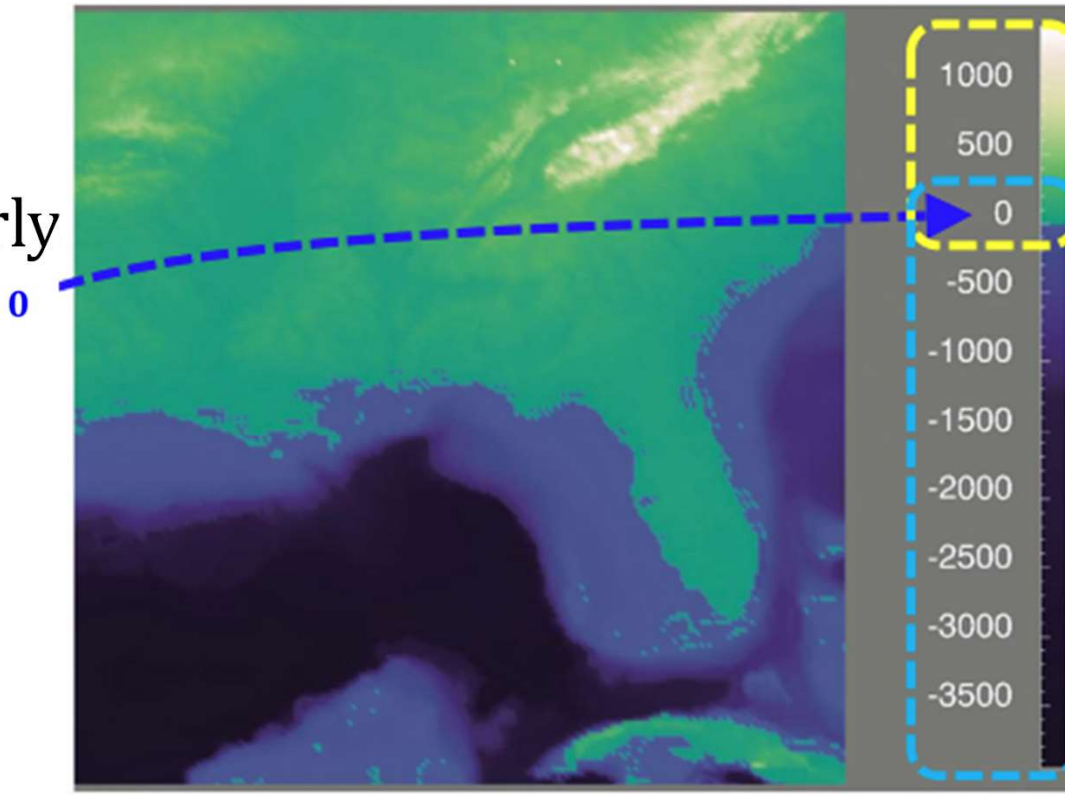
- A range of 1000 units has different characteristics depending on where within the colormap it falls
  - The range [-2000, -1000] has three distinct colors (i.e., cyan, green, and yellow)
  - The range of the same size [-1000, 0] looks all yellow!



What and where is it?

Ordered colormaps

- A colormap that combines ***monotonically increasing luminance*** with multiple hues for semantic categories with a clear segmentation at the 0 point
  - In doing so, it shows high-level, mid-level, and low-level structures clearly



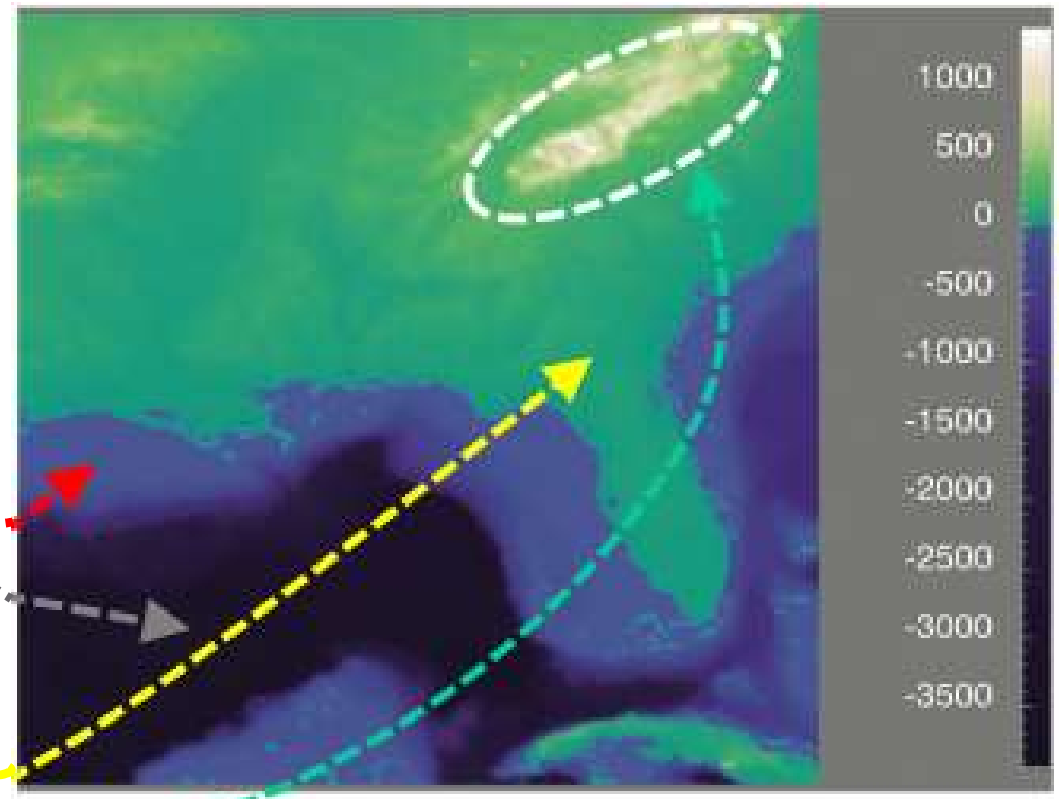
South-East of the US!

Ordered colormaps

- One way to address all three problems is to design **monotonically increasing luminance** colormaps
  - More precisely, the multiple hues being used are ordered according to their luminance from lowest to highest
  - The varying hues allow easy segmentation into categorical regions
  - Luminance is a magnitude channel, that provides perceptual ordering.

Ordered colormaps

- HUE is used to creating a semantically meaningful categorization:
  - 0 point matches with sea level
  - dark *blue sea*
  - cyan *continental shelf*
  - green *lowlands*
  - white *mountains*



South-East of the US!

Ordered colormaps

- It is possible to create a perceptually linear rainbow colormap, but at the cost of losing part of the dynamic range because the fully saturated colors are not available for use.

The standard rainbow colormap is perceptually nonlinear

Perceptually linear rainbow colormap is possible,
but they are less bright with a decreased dynamic range

Segmented rainbows work well for categorical data
when the number of categories is small

Ordered colormaps

segmented

diverging

Spectral
RdYlGn
RdYlBu
RdGy
RdBu
PuOr
PRGn
PiYG
BrBG

sequential

YlOrRd
YlOrBr
YlGnBu
YlGn
Reds
RdPu
Purples
PuRd
PuBuGn
PuBu
OrRd
Oranges
Greys
Greens
GnBu
BuPu
BuGn
Blues

categorical

Set3
Set2
Set1
Pastel2
Pastel1
Paired
Dark2
Accent

continuous

sequential single hue

diverging two hue

sequential multihue

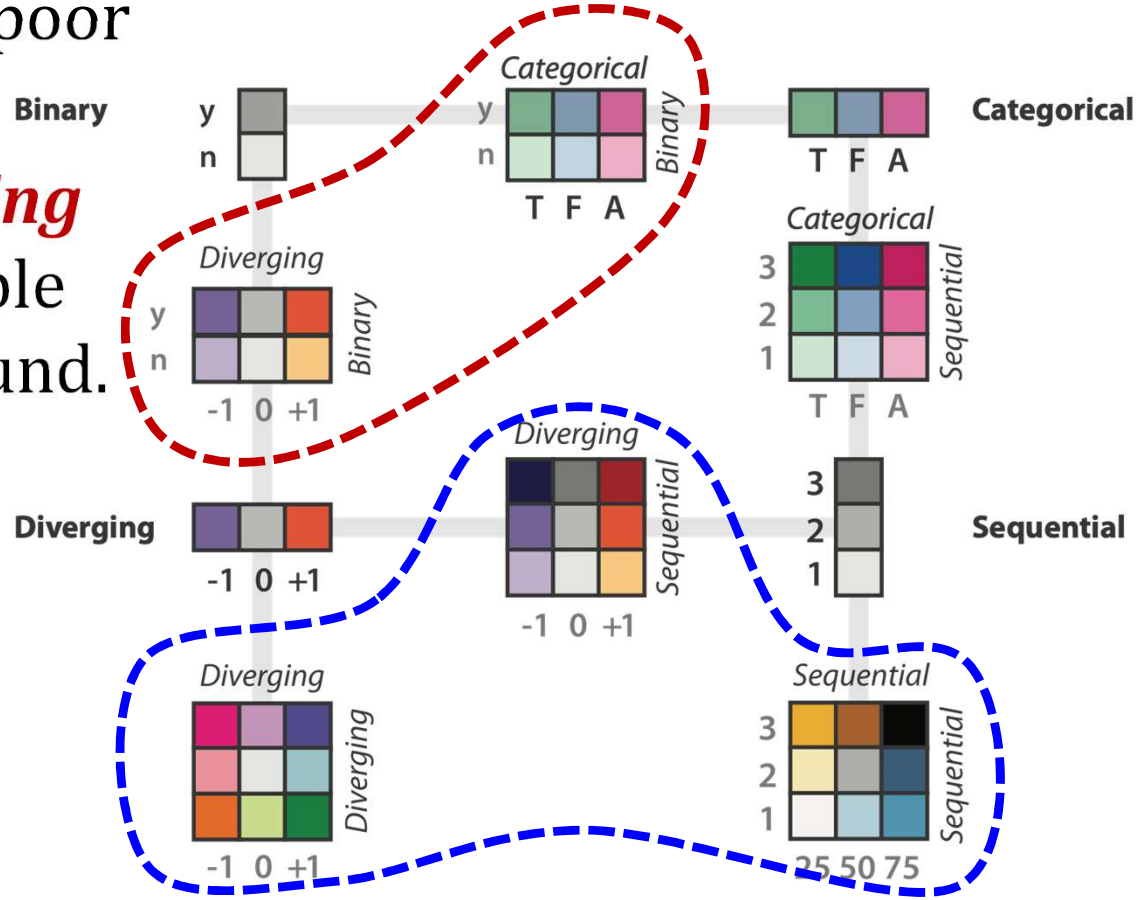
cyclic multihue

Bivariate Colormaps

- The safest use of the color channel is to visually encode a single attribute: these colormaps are known as **univariate**
- Colormaps that encode two separate attributes are **bivariate**

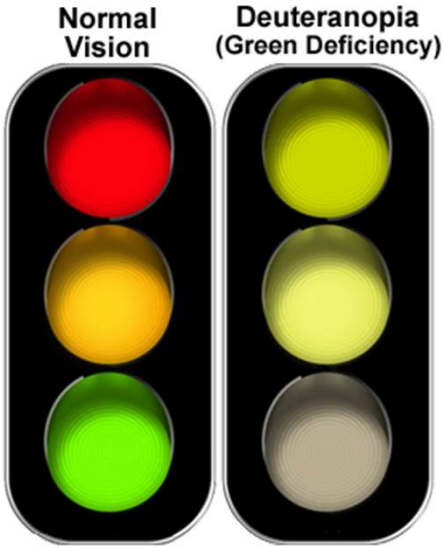
Bivariate Colormaps

- Fixing a base set of hues and varying their saturation
  - When both attributes are categorical with multiple levels, results will be poor
  - Combinations of two ***sequential*** or ***diverging*** attributes with multiple levels is a middle ground.



Colorblind-Safe Colormaps

- **Color blindness** is a sex-linked inherited trait that affects 8% of males and 0.5% of females.
  - In the common forms of color blindness, the ability to sense along the ***red-green*** color axis is limited or absent.
  - Many pairs that are discriminable to people with normal color vision are confused
    - Including *red-black*, *blue-purple*, *light green-white*, and *brown-green*.



Colorblind-Safe Colormaps

- Avoid the use of only the HUE channel to encode information
  - In addition to HUE design, categorical colormaps vary in luminance or saturation
- If possible, avoid colormaps that emphasize red-green
  - Especially in divergent red-green ramps
- Check
  - http://www.color-blindness.com/

Q&A