

RLadies New York

Data Visualization Workshop

Joyce Robbins

August 17, 2017

packages to install:

dplyr, ggplot2, gridExtra, pgmm, scales, tibble, tidyr,
vcdExtra, viridis

download this repo: www.github.com/jtr13/RLadies

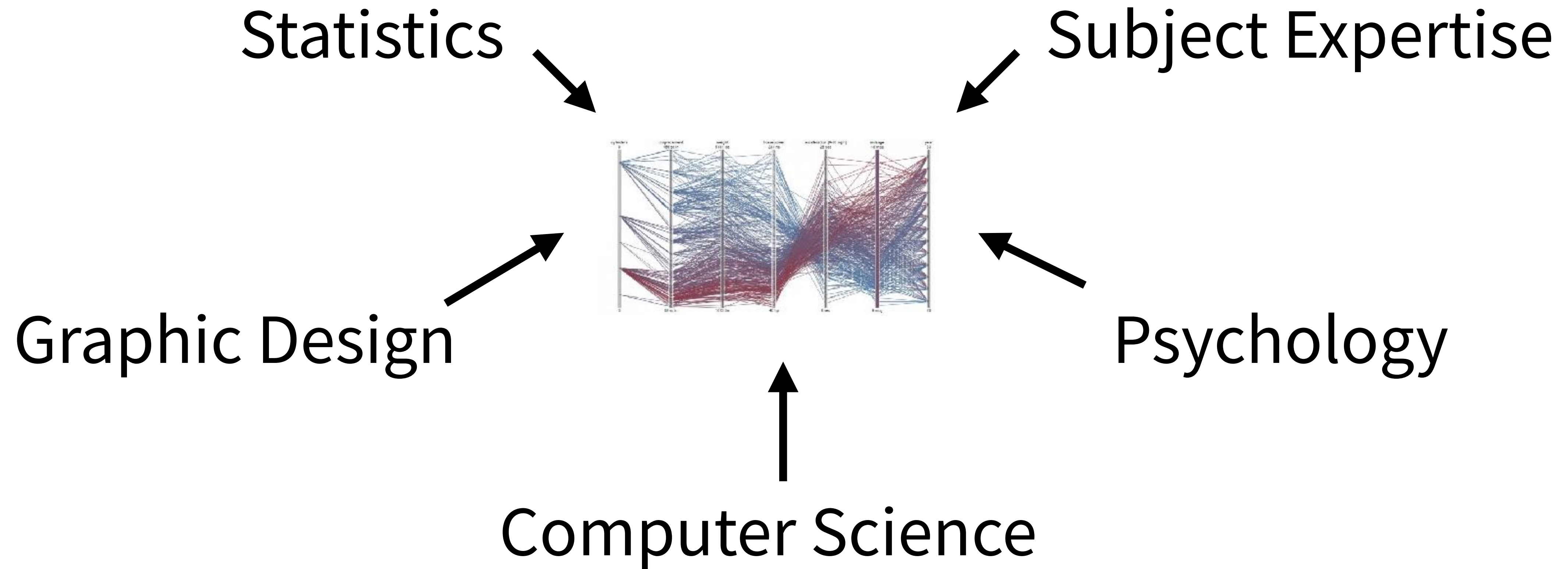
Agenda

- SLIDES: Datavis Intro & Heatmaps
- RSTUDIO: geom_tile, geom_rect, geom_raster
- PRACTICE: tiny data sets
- RSTUDIO: heatmaps
- PRACTICE: Yamaguchi87 dataset (vcdExtra)
- SLIDES: Color
- PRACTICE: Yamaguchi87 or wine dataset (pgmm)

What is data visualization?

- relatively new field (but long history)
- multidisciplinary
- lack of consensus

Interdisciplinary influences



Data visualization for data science

- detecting patterns
- finding outliers
- making comparisons
- identifying clusters

Why heatmaps?

- Employs color to add a third dimension
- Works well in cases in which there's an observation for every (x, y) pair, such as spatial data
- Works with categorical or numerical data, discrete or continuous data

Heat Map Examples

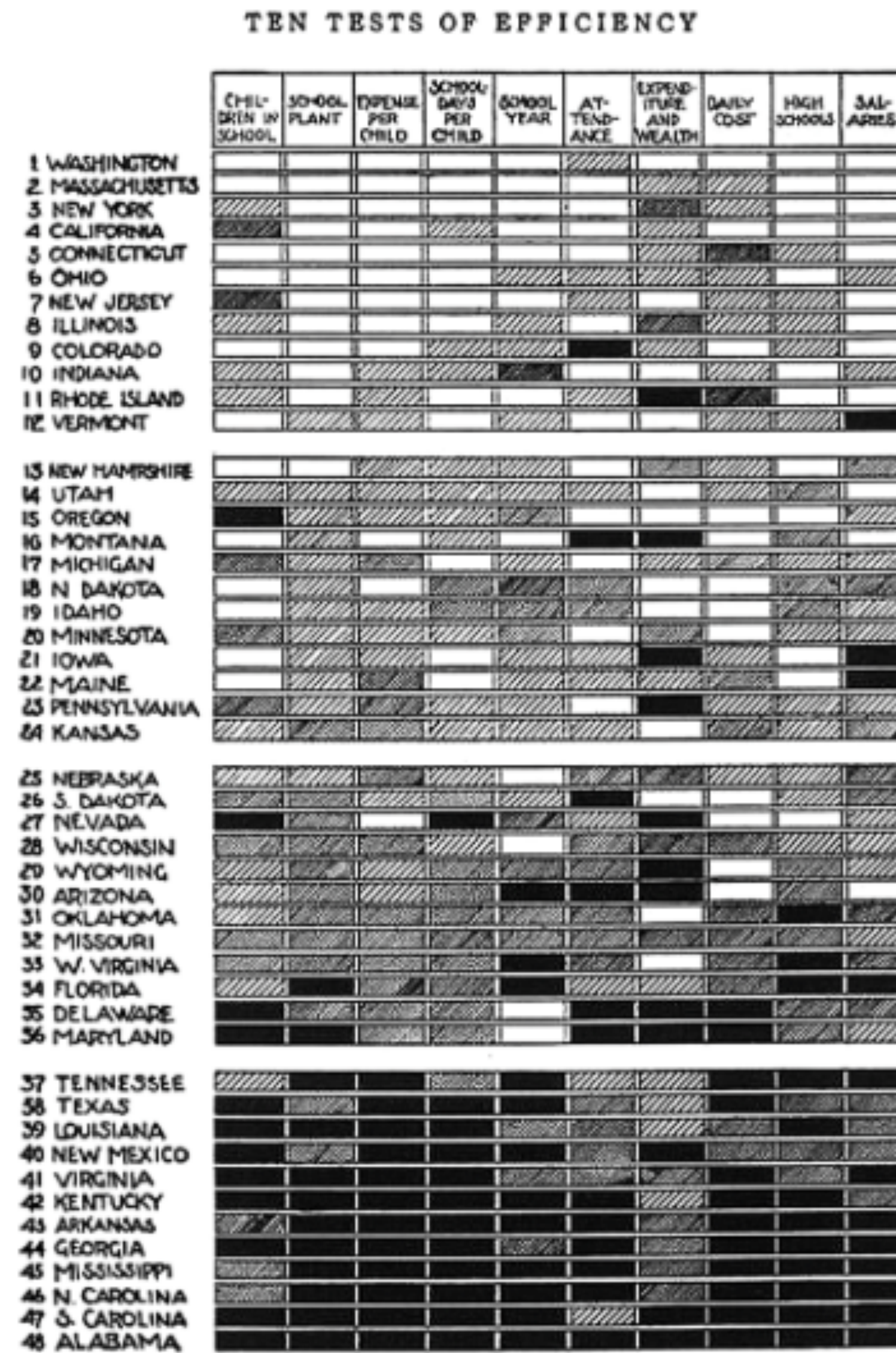


Figure 3. Sorted shaded display from Brinton (1914). The data are ranks of U.S. states on each of 10 educational features assessed in 1910. The matrix has been sorted by the row-marginal ranks.

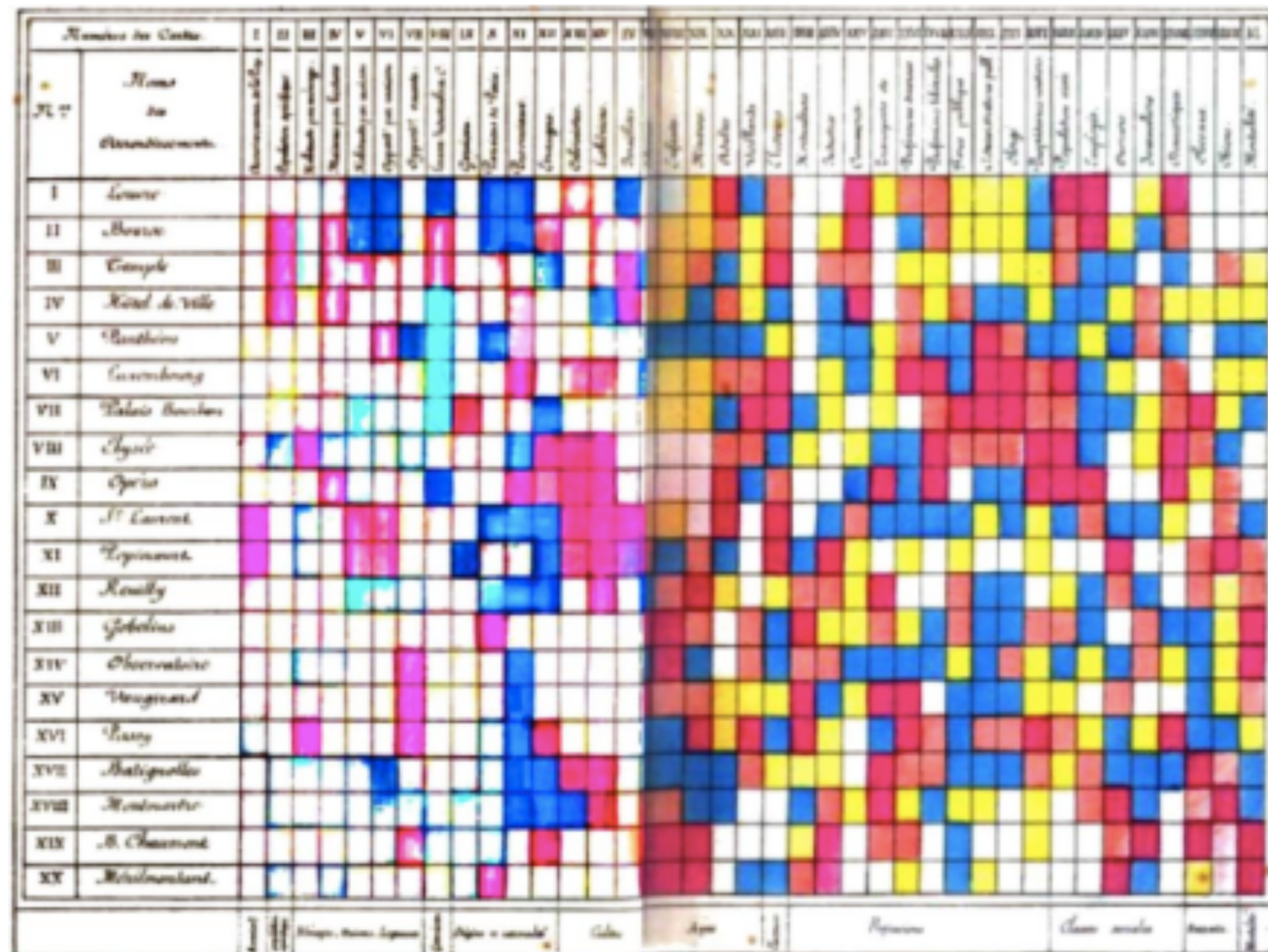
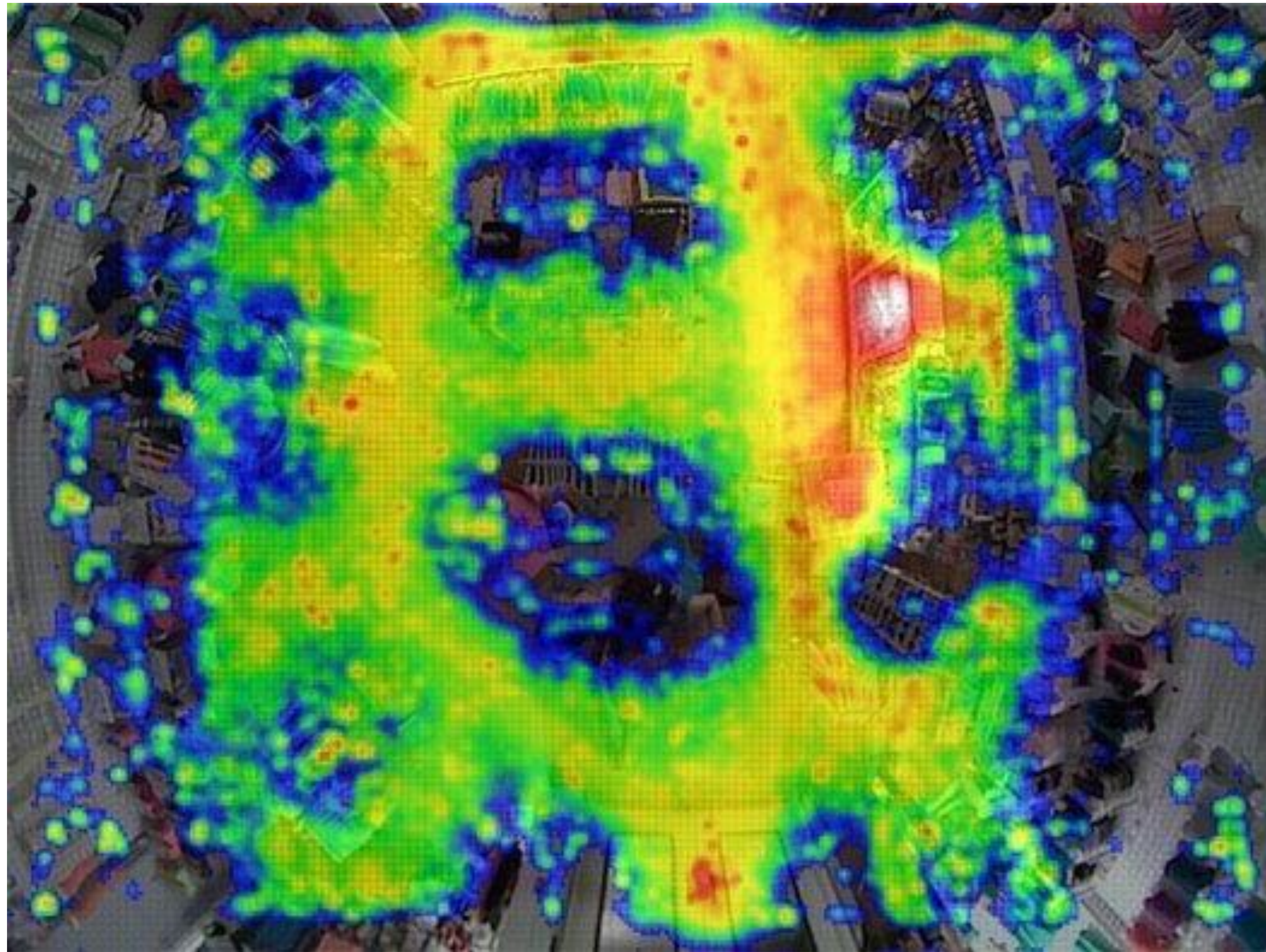
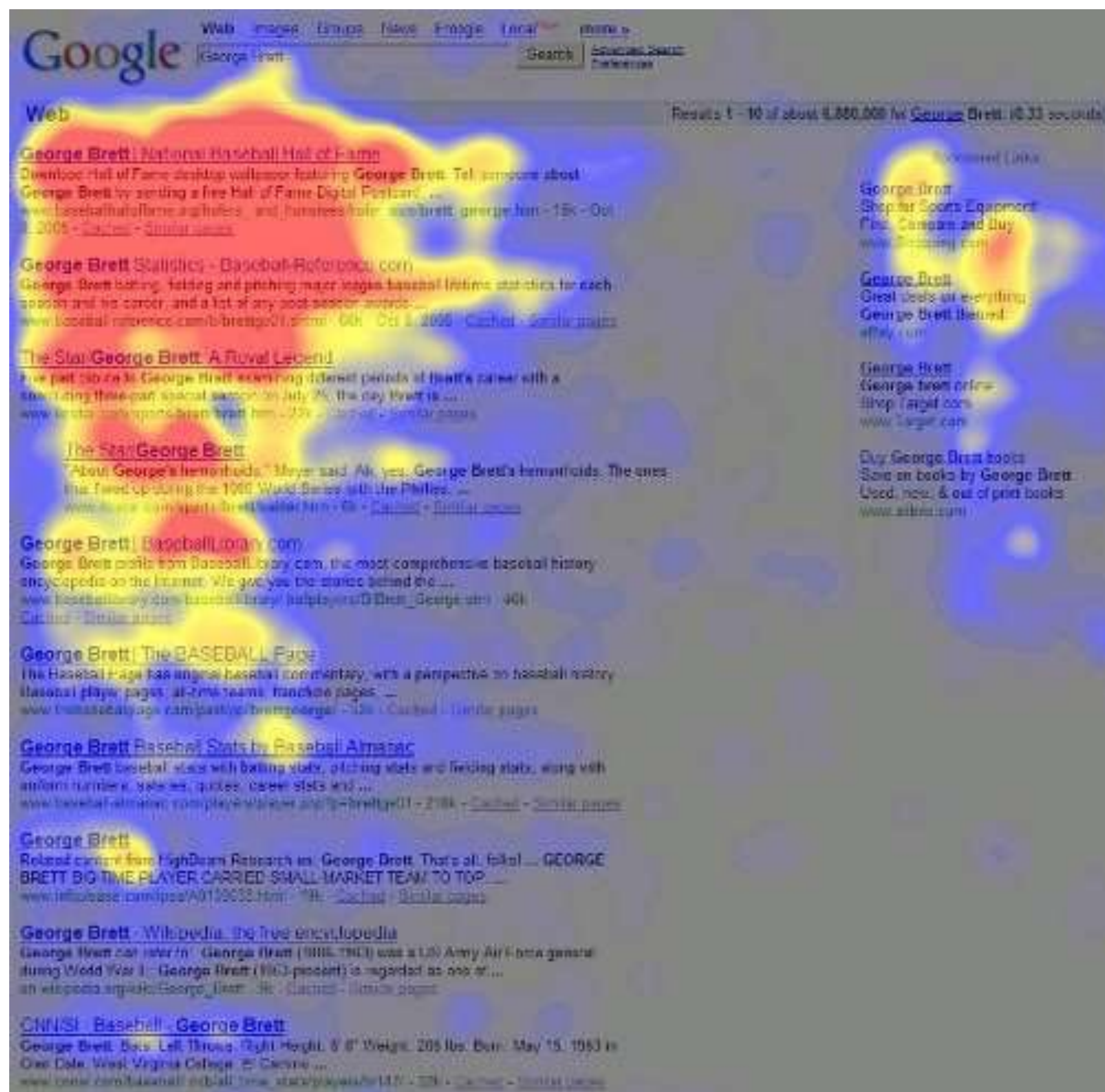


Figure 2. Shaded matrix display from Loua (1873), available online at <http://books.google.com/books/>. This was designed as a summary of 40 separate maps of Paris, showing the characteristics (e.g., national origin, professions, age, social classes) of 20 districts, using a color scale ranging from white (low) through yellow and blue to red (high).

In-store shopping pattern



physical
location



screen
location



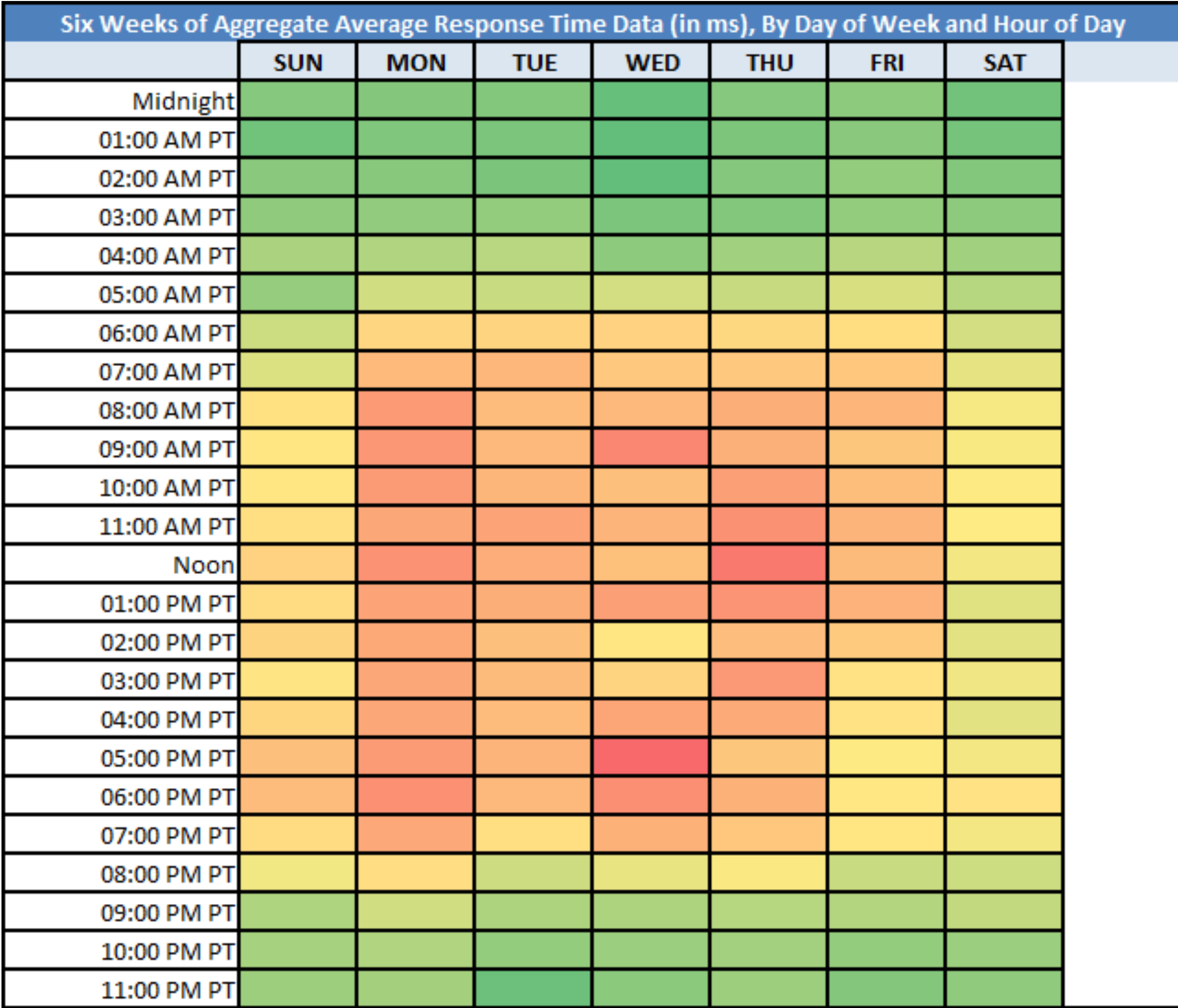
Extra gentle for the most sensitive skin.

So gentle for sensitive skin, add the chemicals and moisture of a diaper and you have diaper rash.

Baby Wipes's unique high-absorbency natural-blend cotton provides cotton-soft, extra thick, gel-free protection for your baby's sensitive skin. The chlorine-free materials and absorbent polymers is non-toxic and non-irritating. Clinically tested and pediatrician recommended for babies with allergies and sensitive skin.



Web site response



time

Gene expression

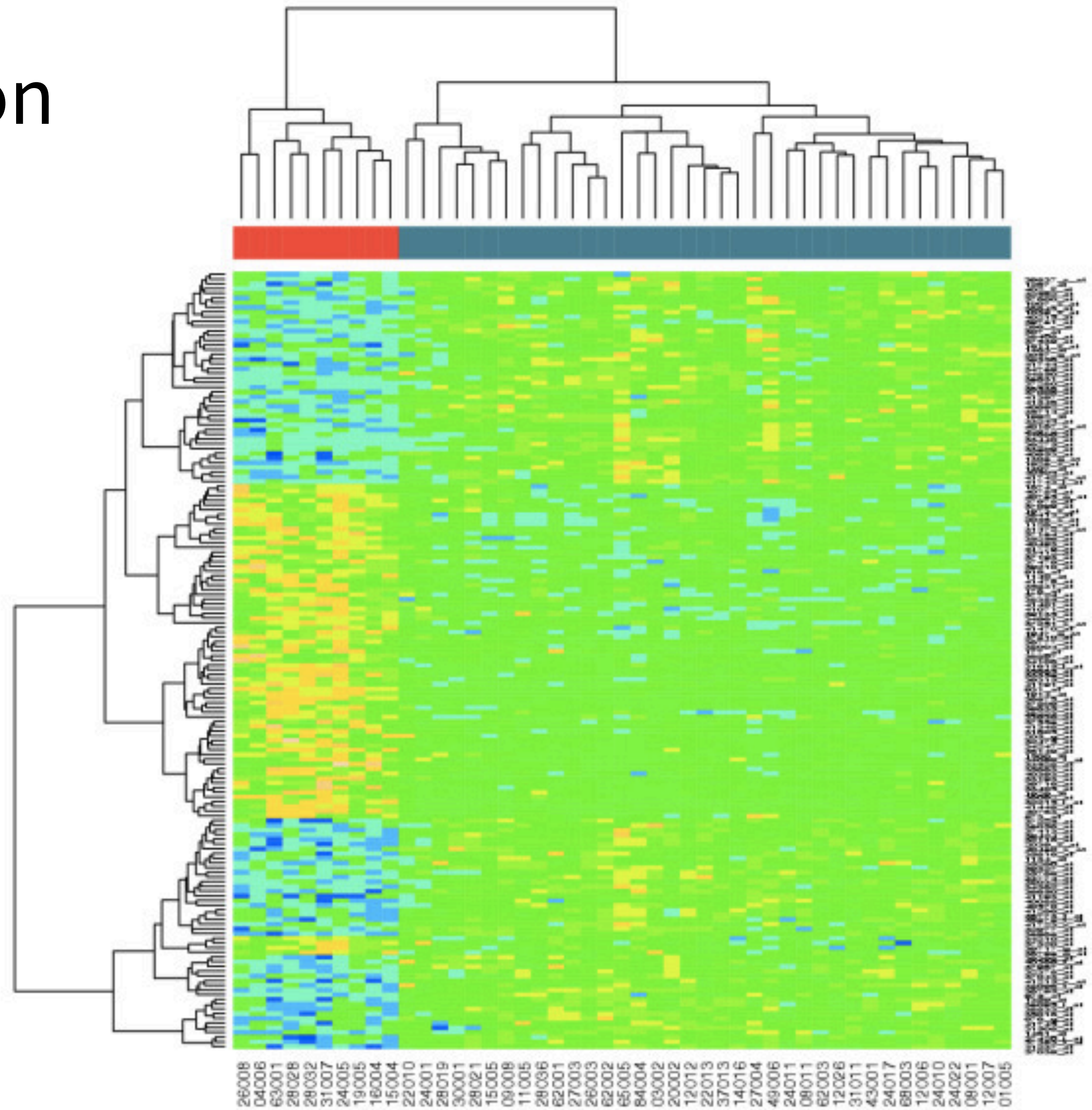
clustered
samples

clustered
genes



(normalized data)

Gene expression

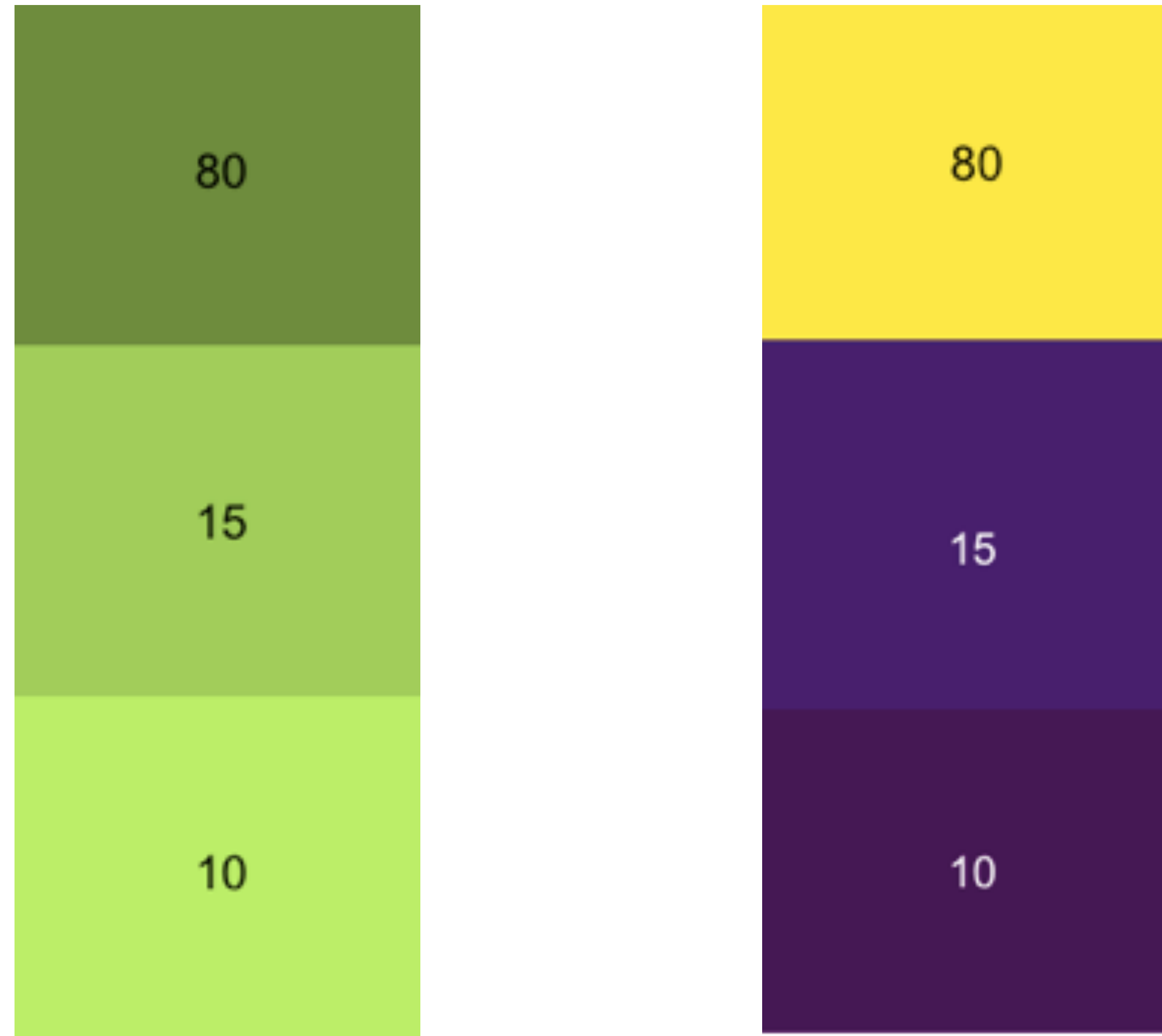


Treemap

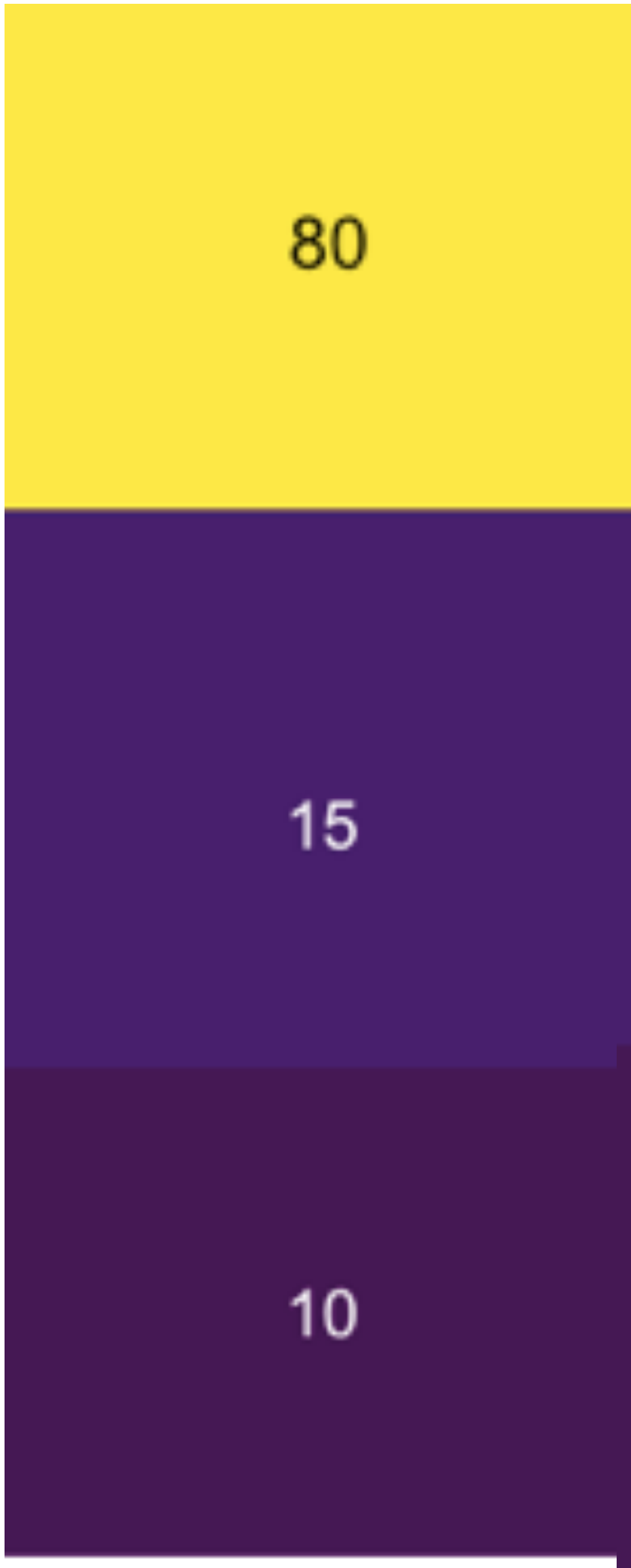


Time to code...

Perceptually uniform color space



Perceptually uniform color space



viridis

Perceptually uniform color spaces

- don't blur important distinctions in the data
- don't add distinctions that don't exist in the data

`viridis` package

tl;dr

Use the color scales in this package to make plots that are pretty, better represent your data, easier to read by those with colorblindness, and print well in grey scale.

<http://bids.github.io/colormap/>

<http://matplotlib.org/users/colormaps.html>

viridis



inferno



plasma



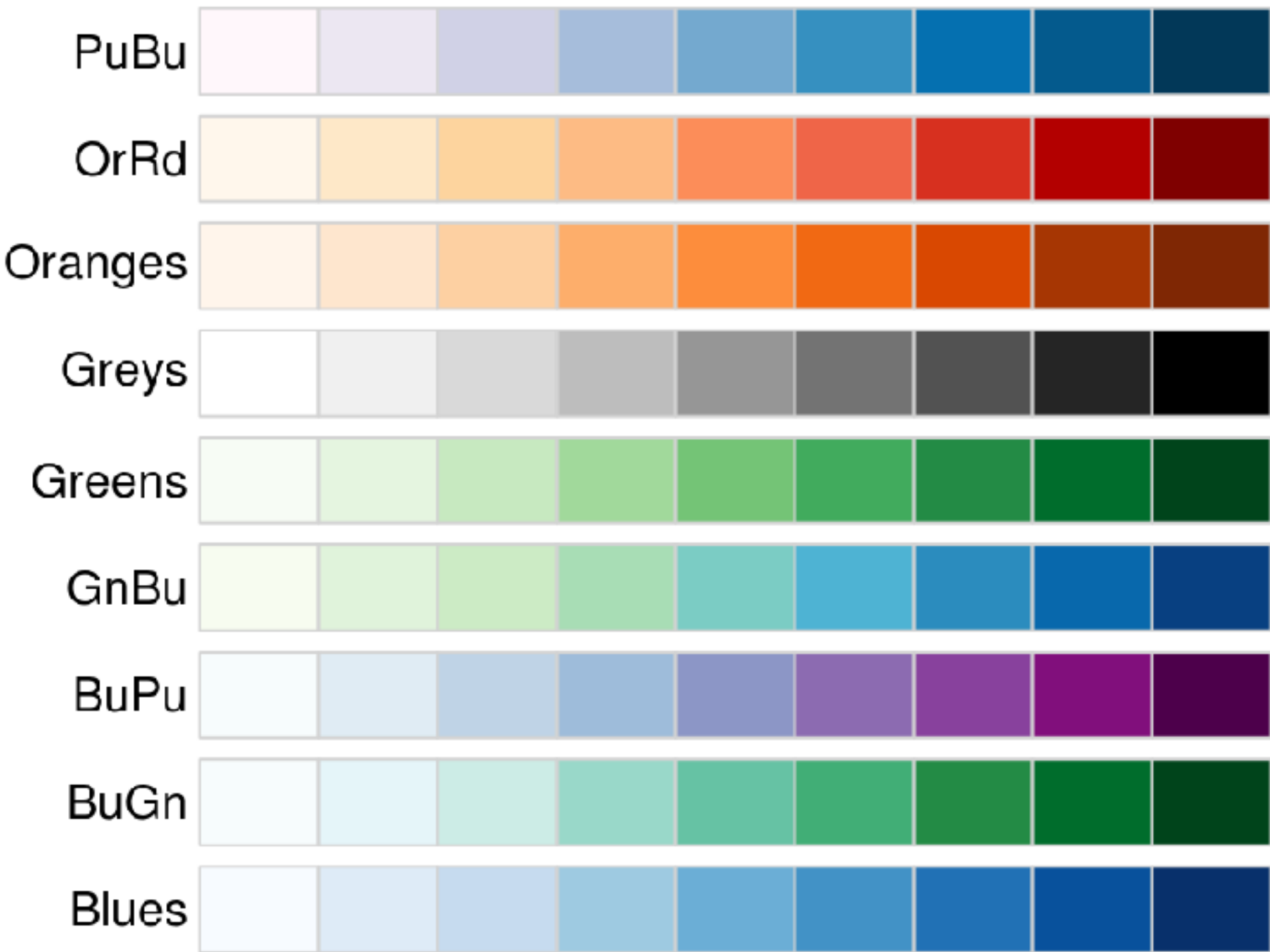
viridis
package

magma



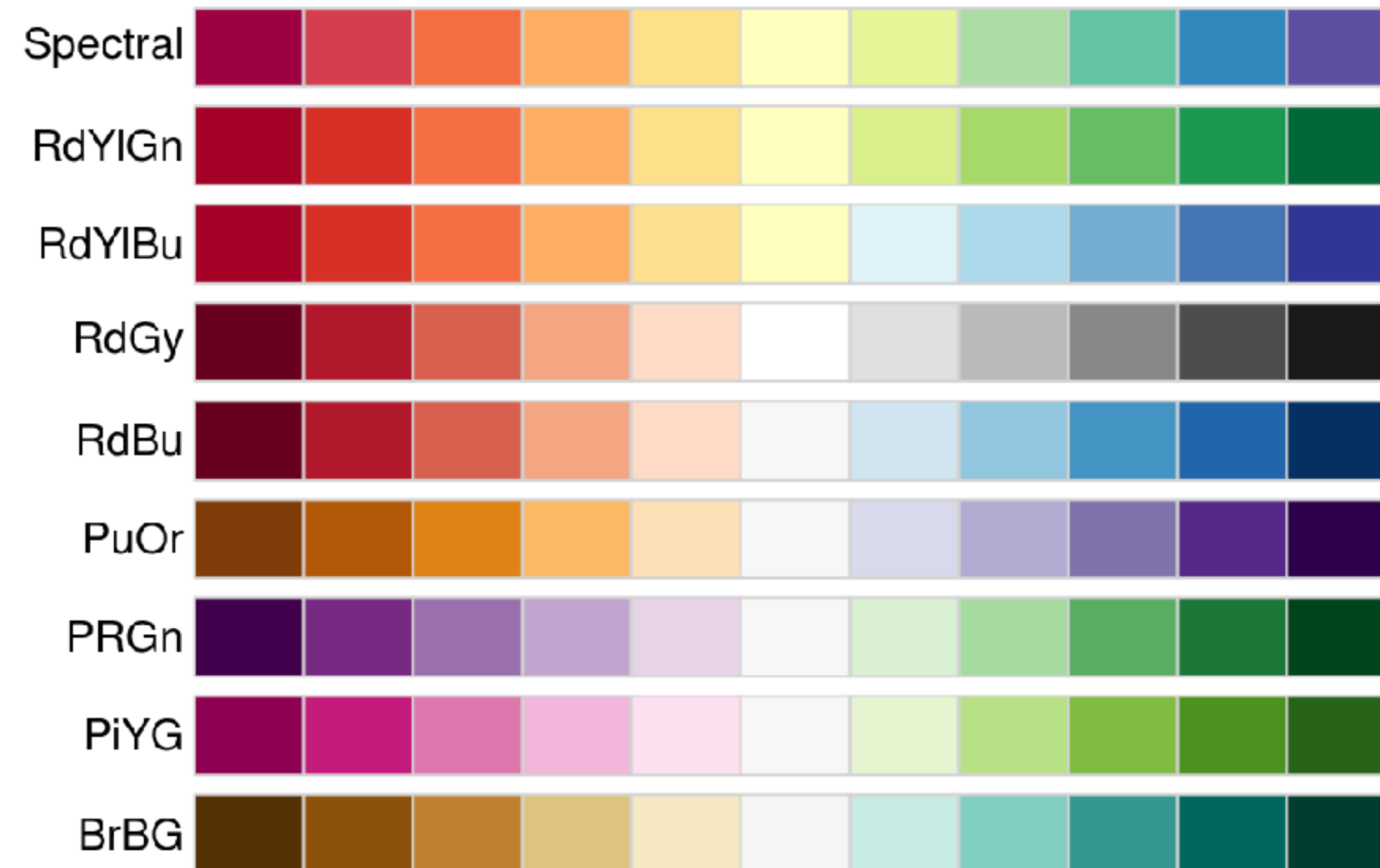
RColorBrewer Color Schemes

sequential



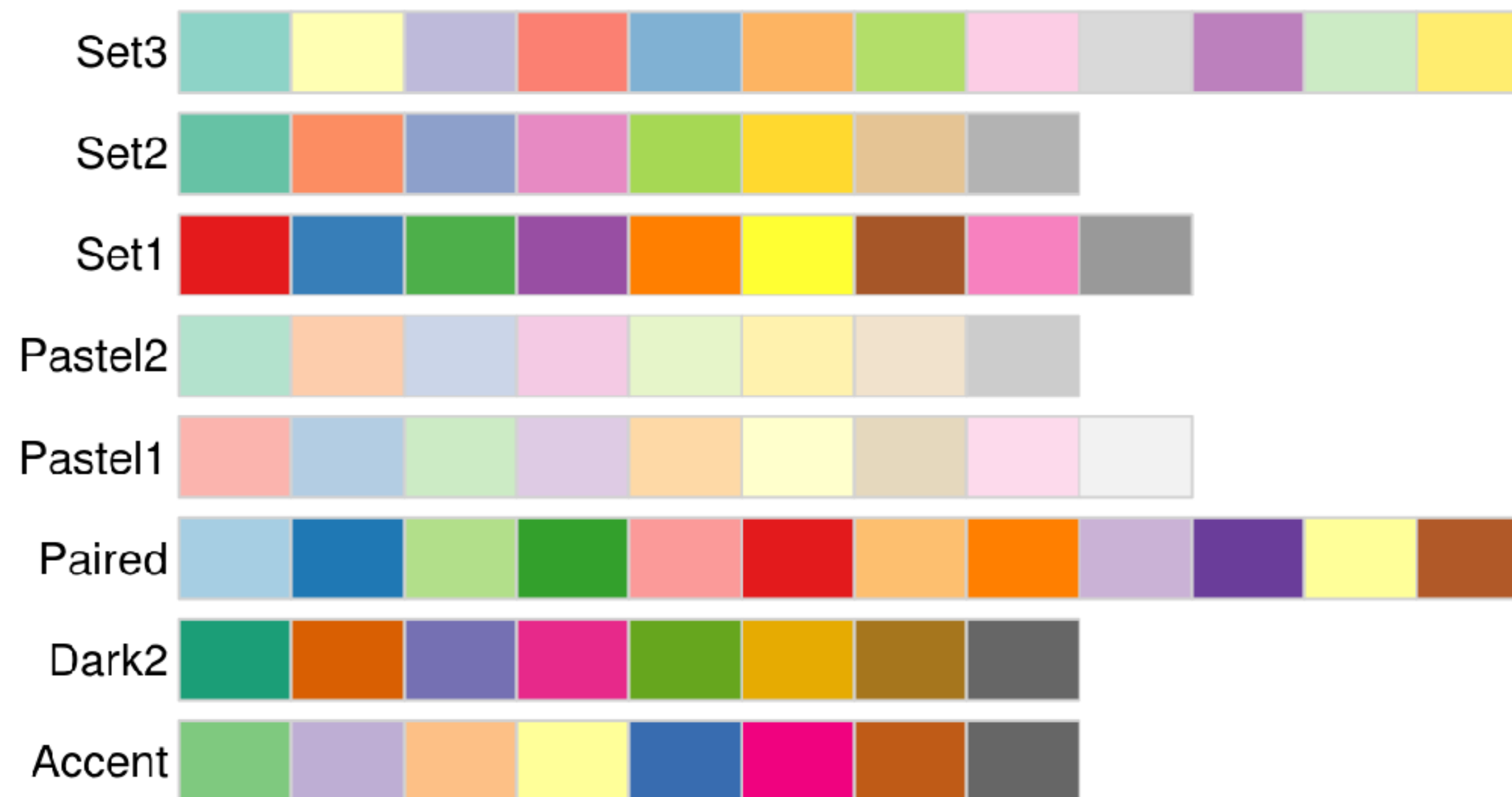
RColorBrewer Color Schemes

diverging



RColorBrewer Color Schemes

qualitative (for categorical data)



Continuous data

VIRIDIS

+ scale_color_viridis()

OR: `_fill_`

RCOLORBREWER

+ scale_color_distiller(palette = "PuBu")

~~[+scale_color_brewer(palette = "PuBu")]~~ Error: Continuous value supplied to discrete scale]

~~[+scale_color_continuous(palette = "PuBu")]~~ Error in f(..., self = self) : attempt to apply non-function]

CREATE YOUR OWN

+ scale_color_gradient(low = "white", high = "red")

+ scale_color_gradient2(low = "red", mid = "white", high = "blue", midpoint = 50)

+ scale_color_gradientn(colours = c("red", "pink", "lightblue", "blue"))

Discrete data

VIRIDIS

```
+ scale_color_viridis() Error: Discrete value supplied to continuous scale  
+ scale_color_viridis(discrete = TRUE)
```

RCOLORBREWER

```
+ scale_color_brewer(palette = "PuBu")  
[+ scale_color_discrete(palette = "PuBu") Error in f(..., self = self) : attempt to apply non-function]  
+ scale_fill_grey()
```

CREATE YOUR OWN

```
+ scale_color_manual(values = c("red", "yellow", "blue"))
```

Be careful...

```
> g + scale_fill_economist()
```

Error: Continuous value supplied to discrete scale

```
> g + scale_fill_viridis()
```

Error: Discrete value supplied to continuous scale

Color Vision Deficiency

approx. 8% of men, 0.5% of women have some form

missing or deficient cones:

protanopia (red)

deuteranopia (green)

tritanopia (blue)

How to make CVD friendly graphs

Use palettes that have already been tested
(see viridis help)

Use a CVD simulator

www.vischeck.com

<http://www.color-blindness.com/coblis-color-blindness-simulator/>

Use high contrast

Back to RStudio...