



STAT 215A Fall 2017

Week 3

Rebecca Barter
09/08/2017



Lab 1 check in

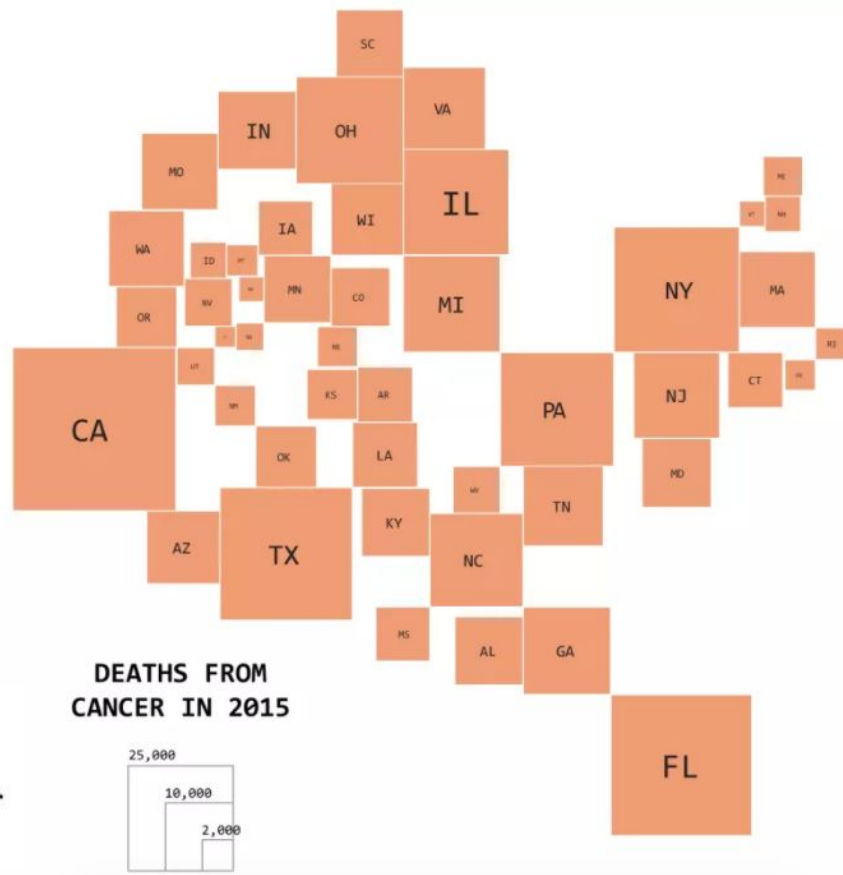
How do you explore the data (for understanding)?

How do you find an interesting story to tell about the data?

Where does creativity come in when doing data analysis?

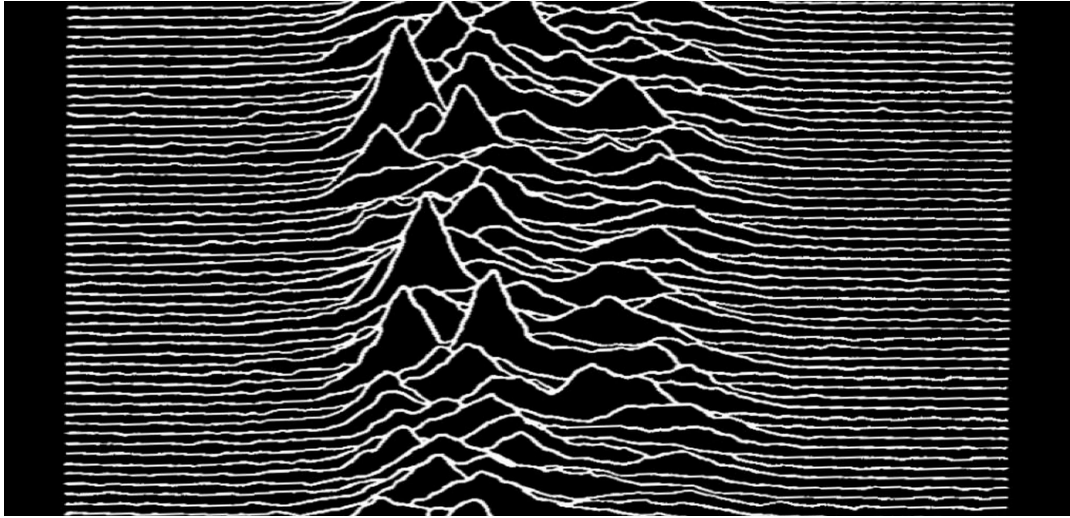
A blog to follow: Flowing Data

<http://flowingdata.com/2017/08/15/useless-points-of-comparison/>



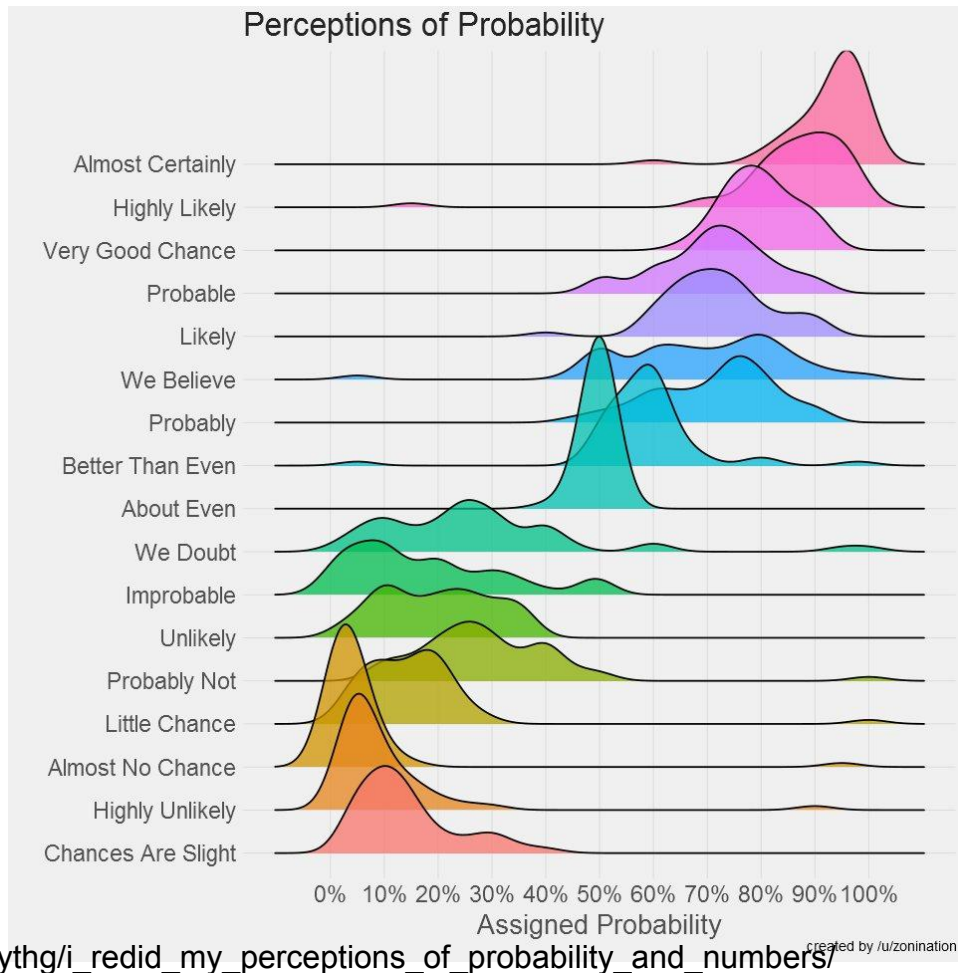
FLOWINGDATA

Modern kernel density estimation



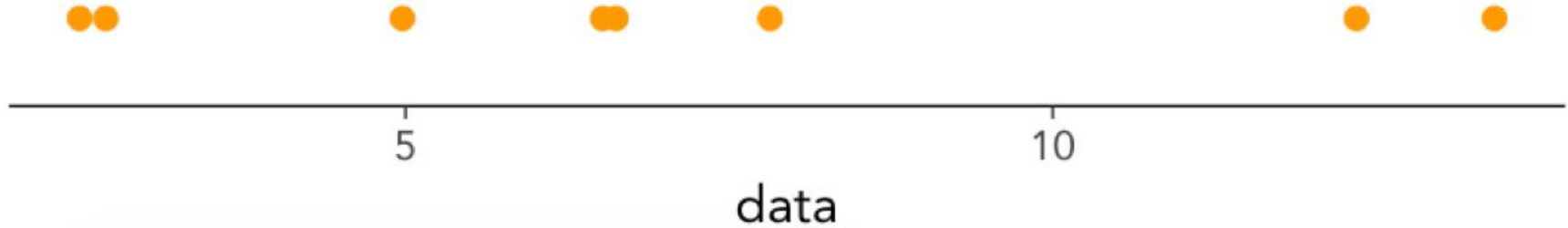
Known on twitter as the “joyplot”

<https://twitter.com/hashtag/joyplot>

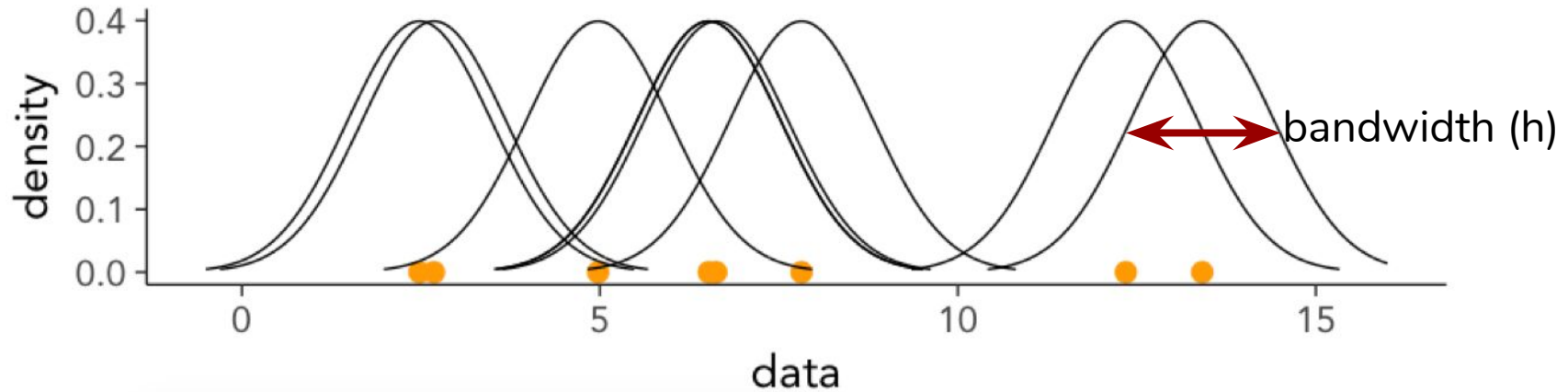


https://www.reddit.com/r/dataisbeautiful/comments/6vythg/i_redid_my_perceptions_of_probability_and_numbers/

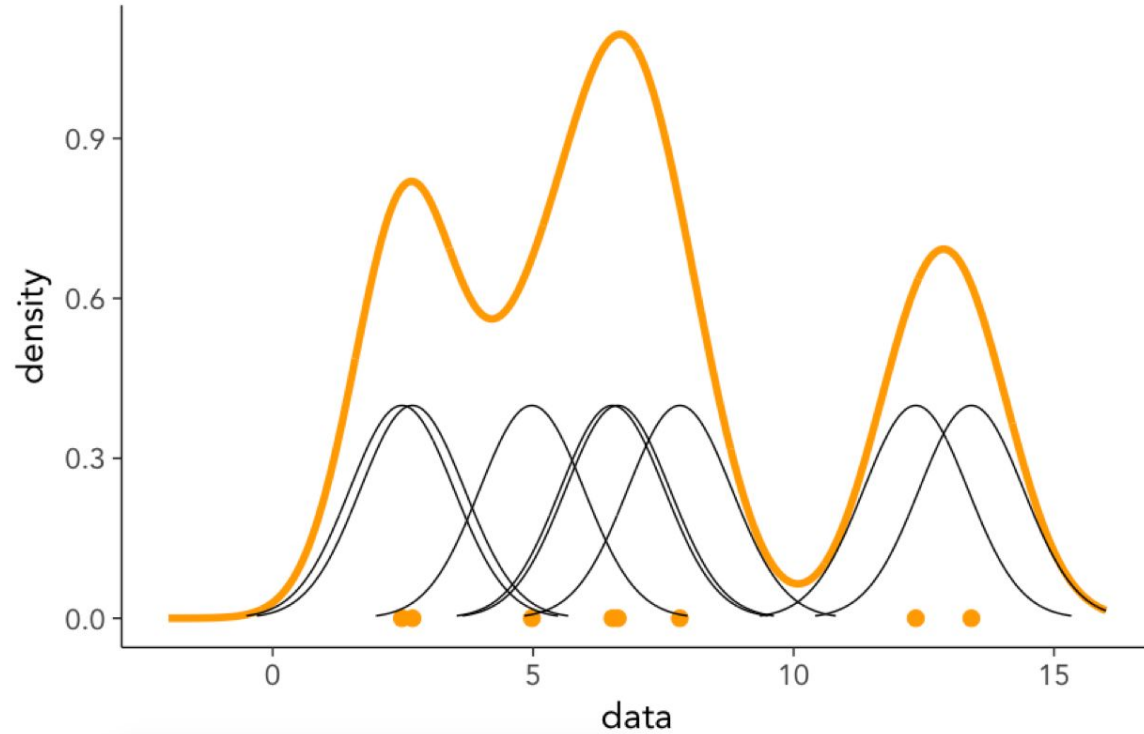
Kernel Density Estimation



Kernel Density Estimation



Kernel Density Estimation




Kernel Density Estimation

Estimate the density, f , by adding together individual kernel functions:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

Kernel Density Estimation

Estimate the density, f , by adding together individual kernel functions:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$


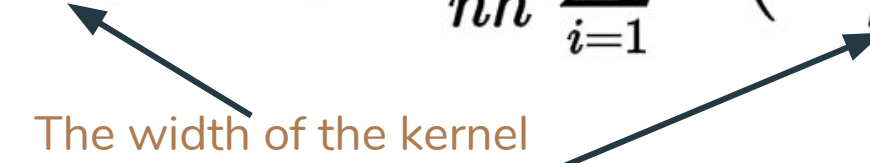
Each kernel function is
centered at a data point

Kernel Density Estimation

Estimate the density, f , by adding together individual kernel functions:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - x_i) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

The width of the kernel function is defined by the bandwidth, h

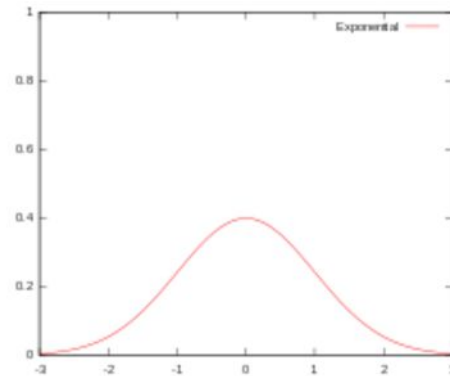


Kernel Density Estimation

There are many possible Kernel function that you could use

Gaussian

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$



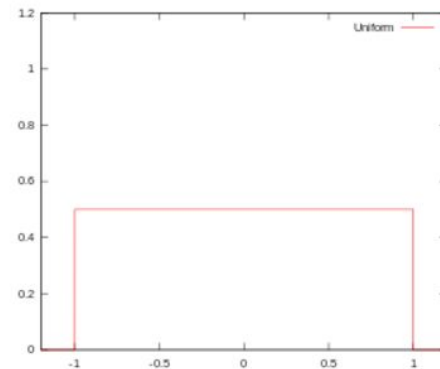
Kernel Density Estimation

There are many possible Kernel function that you could use

**Uniform
("rectangular
window")**

$$K(u) = \frac{1}{2}$$

$$\text{Support: } |u| \leq 1$$



"Boxcar function"

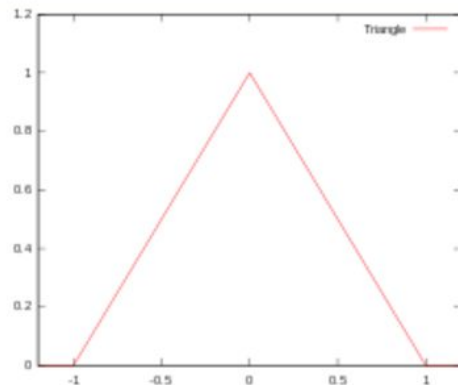
Kernel Density Estimation

There are many possible Kernel function that you could use

Triangular

$$K(u) = (1 - |u|)$$

$$\text{Support: } |u| \leq 1$$



Kernel Density Estimation

There are many possible Kernel function that you could use

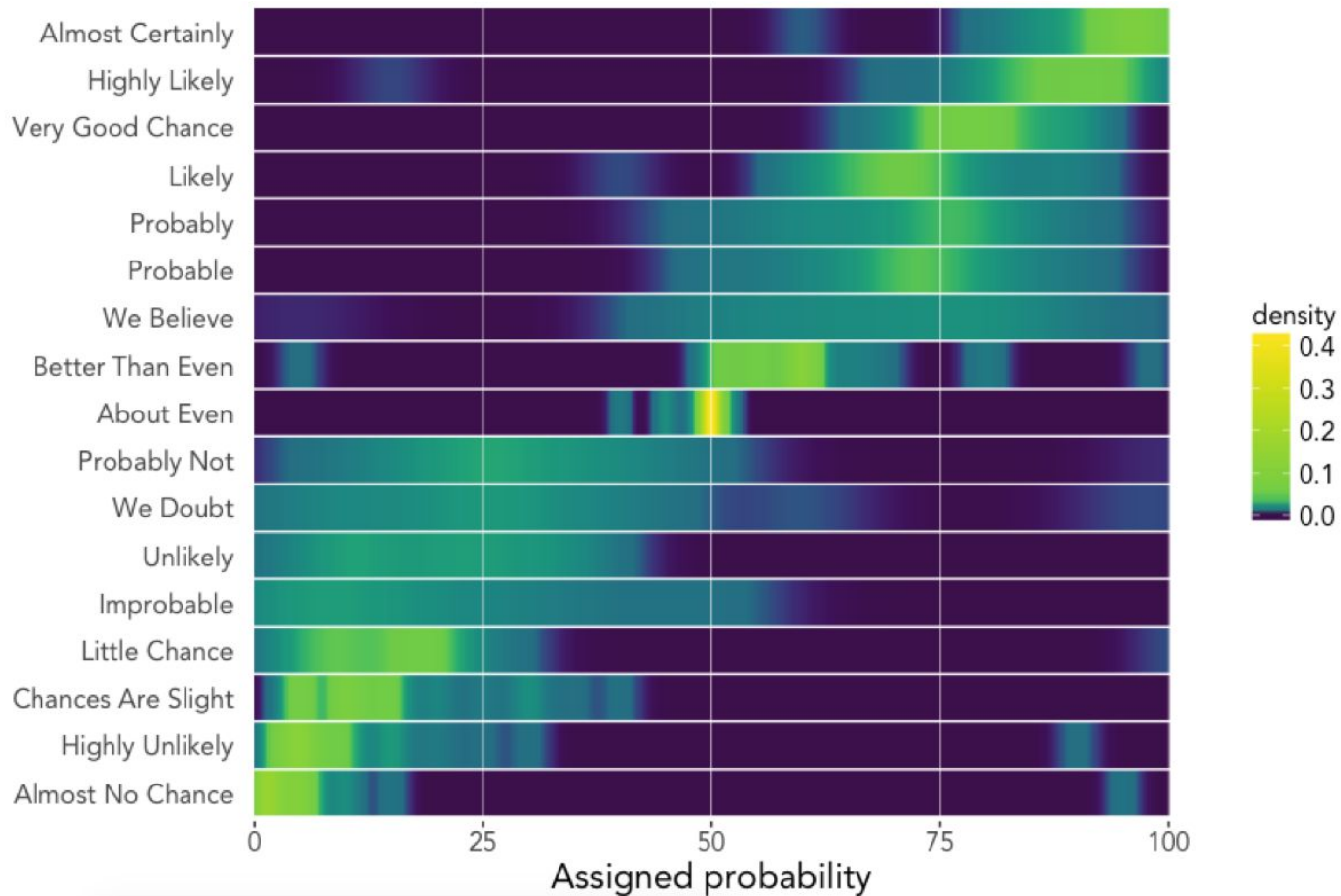
Exercise 1

1. Write a function that calculates the density of a vector of numbers.
2. Plot the density of the “chances are slight” probability interpretations.
3. Make a plot that displays the bias-variance tradeoff.

To download the data, clone Zonation’s github repository: <https://github.com/zonation/perceptions>

Alternative
view of the
same data:

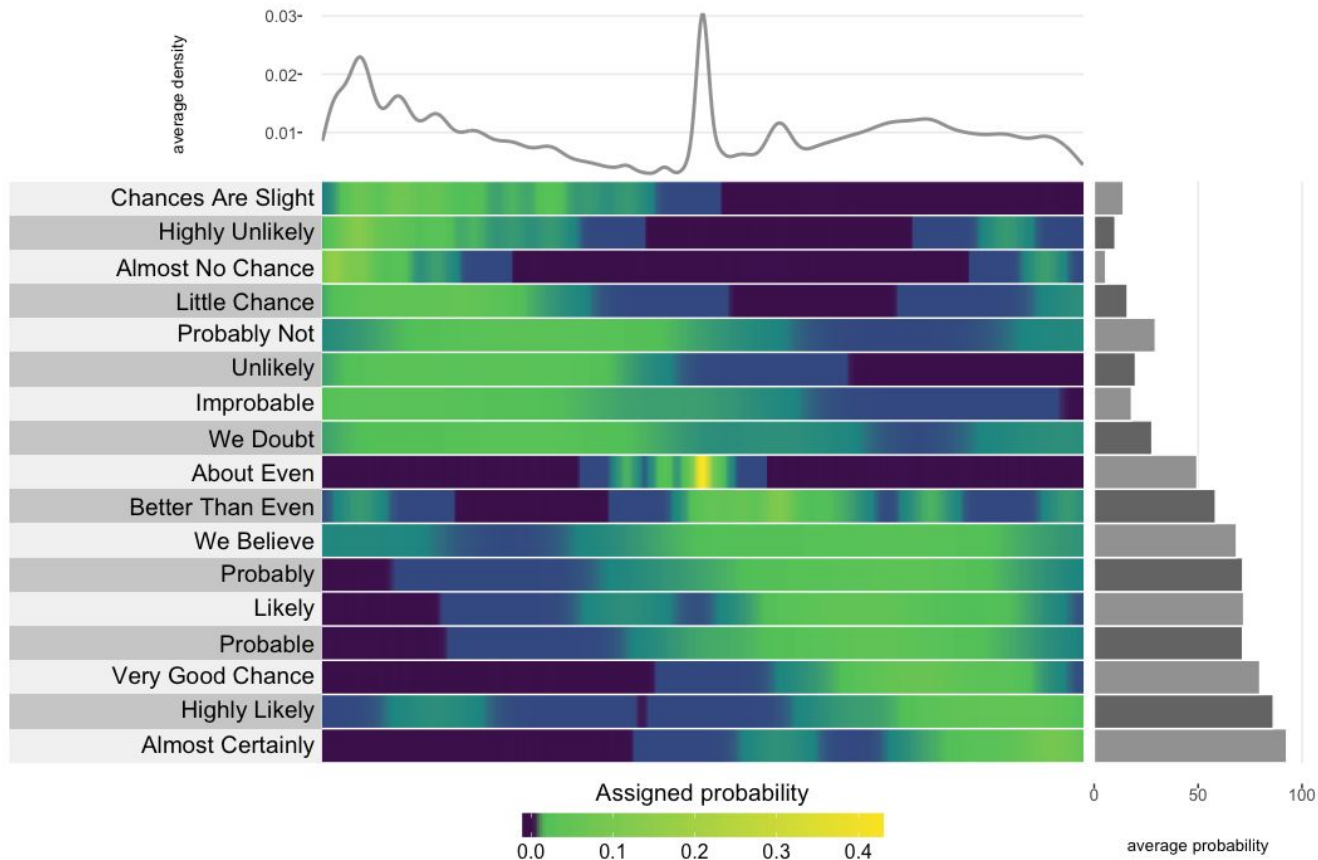
A heatmap
using the
viridis color
scheme



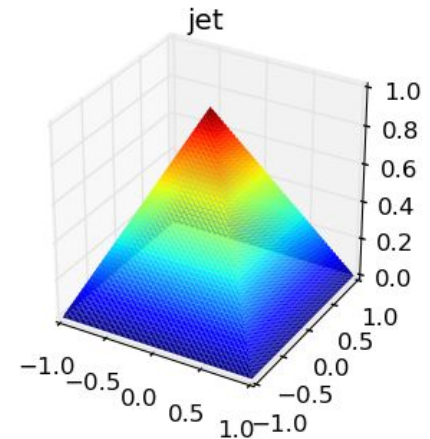
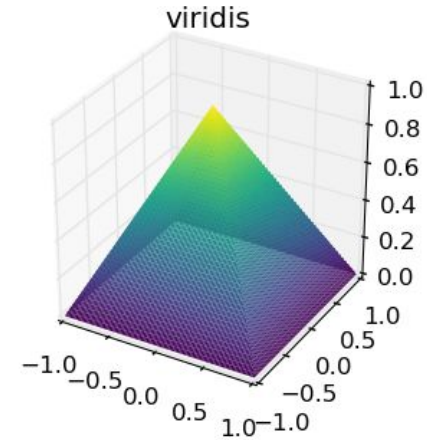
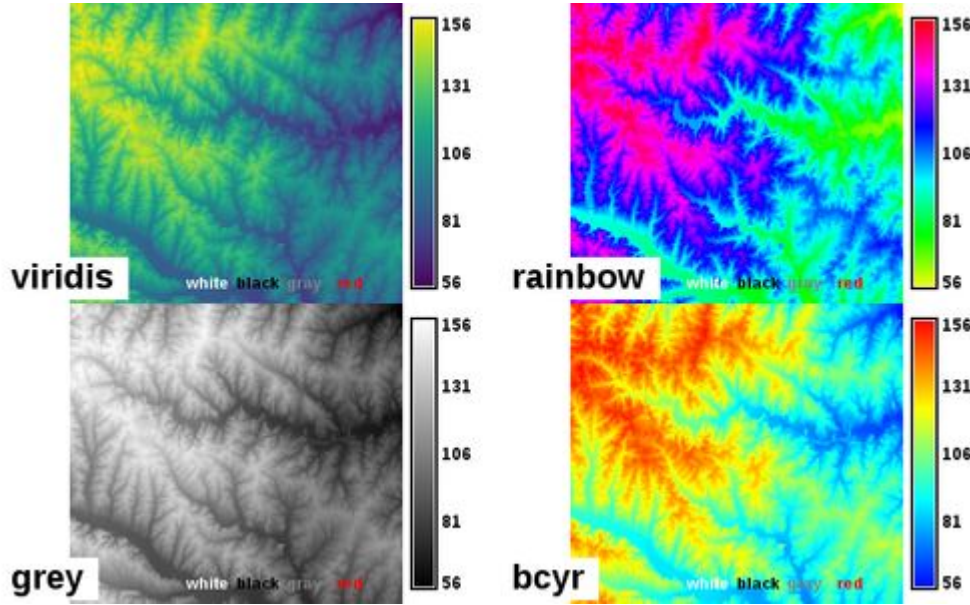
Data: <https://github.com/zonination/perceptions>

Superheat version

- Automatically selects color transitions based on data quantiles
- Automatically arranges rows in median order

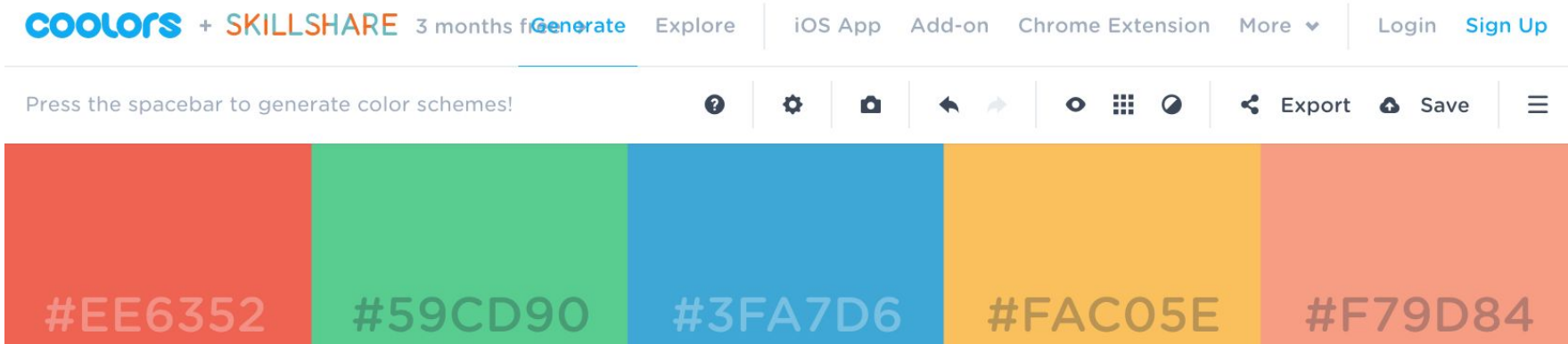


Viridis: the perceptually uniform colour palette



Choosing color palettes:

<https://coolors.co/app>



Choosing color palettes:

[http://http://colorbrewer2.org/](http://colorbrewer2.org/)

The screenshot displays the ColorBrewer 2.0 web application interface. On the left, a sidebar contains controls for selecting a color scheme. The 'Number of data classes' is set to 3. Under 'Nature of your data', the 'qualitative' option is selected. The 'Pick a color scheme' section shows a grid of 12 color palette options. The main area features a map of the United States with a qualitative color scheme applied, showing two primary colors (teal and orange) and a small area of blue. The top right of the interface displays the 'COLORBREWER 2.0' logo and the tagline 'color advice for cartography'. Navigation links for 'how to use', 'updates', 'downloads', and 'credits' are visible in the top left of the main area.

Number of data classes: 3

Nature of your data:
☐ sequential ☐ diverging ☒ qualitative

Pick a color scheme:

how to use | updates | downloads | credits

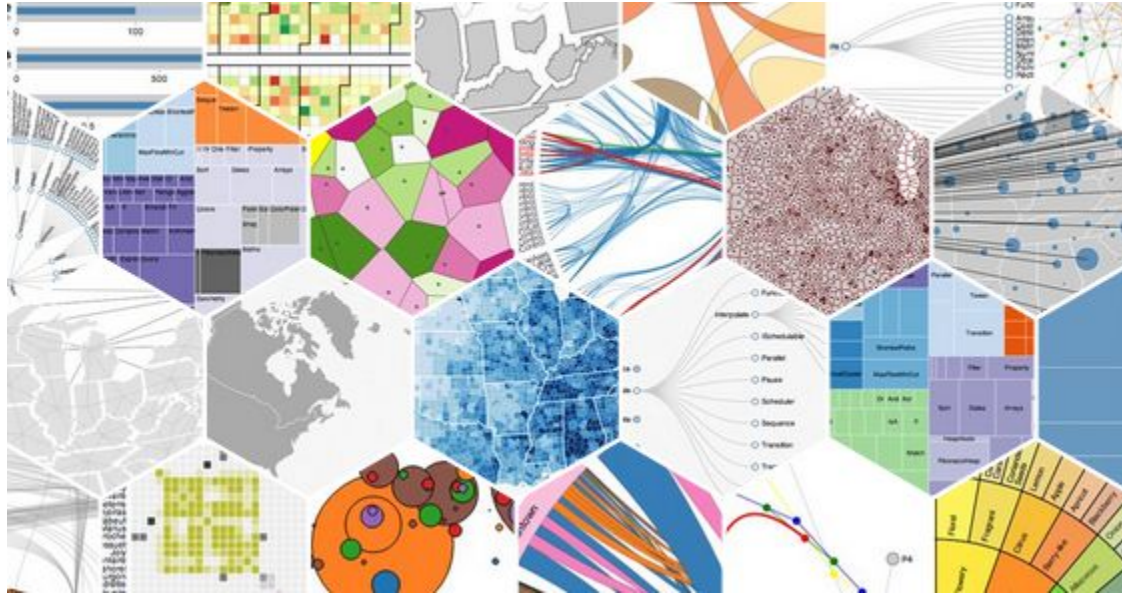
COLORBREWER 2.0
color advice for cartography

Exercise 2

Come up with your own visualization of the perception data.

The only requirement is: **be creative!**

Interactivity



Exercise 3

Come up with your own ***interactive*** visualization of the perception data.

Useful R packages:

- Plotly
- Crosstalk
- Highcharter
- Shiny

Next week.... PCA