Stat 215 A - Week 3

Zoe Vernon

Lab 1 check in

How is it going?

What are people finding difficult?

Any questions?

Source: http://www.redwoodhikes.com/JedSmith/BoyScout1.jpg

Lab 1 check in

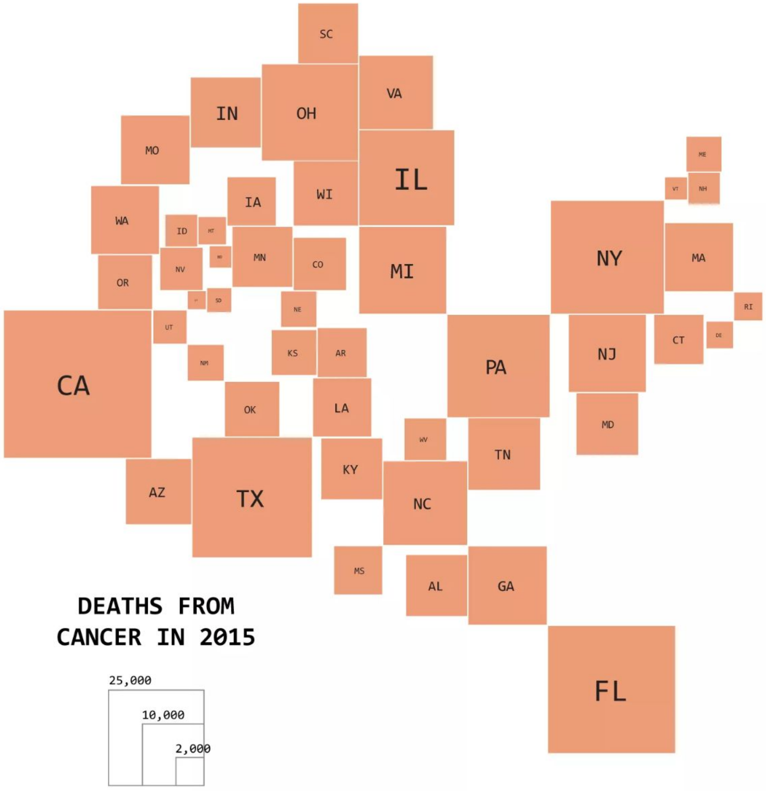
Some thoughts for coming up with findings

- ❑ Use your domain knowledge to come up with questions you may want to answer
- ❑ Look closer at the data
  - ❑ Zoom in on a specific day
  - ❑ Or specific times of day

Flowing data blog

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

FLOWINGDATA

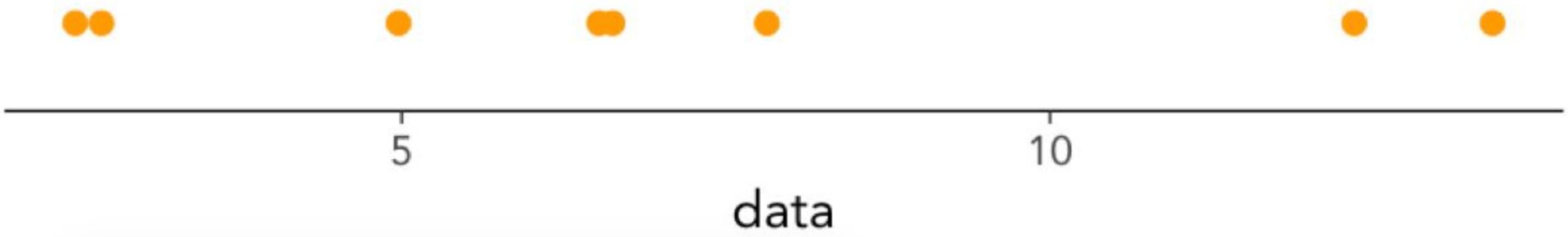


SOURCE: CENTERS FOR DISEASE CONTROL AND PREVENTION

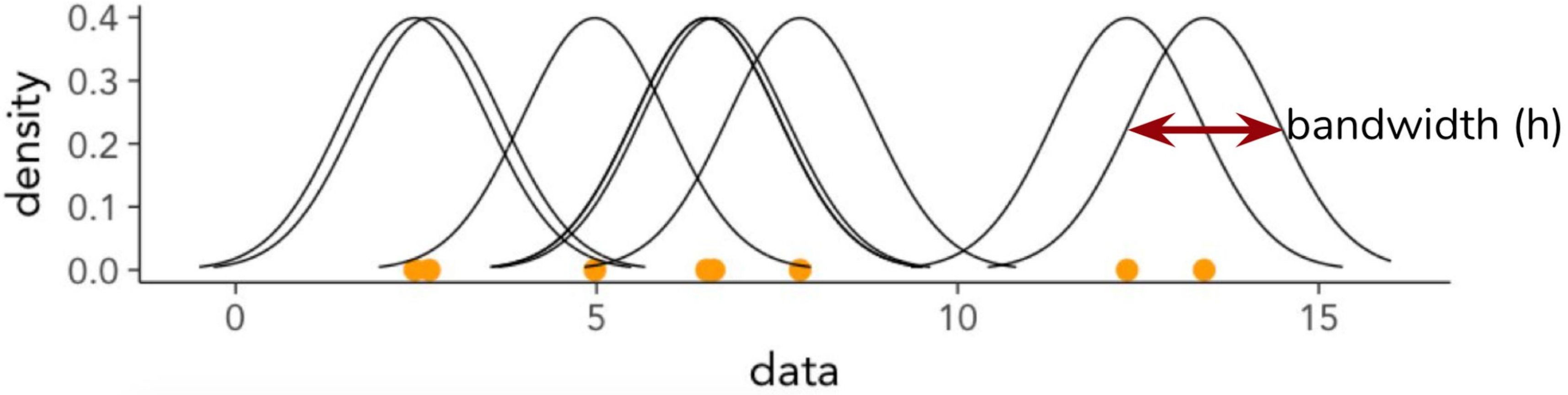
Kernel density estimation

Slides in this section from Rebecca Barter

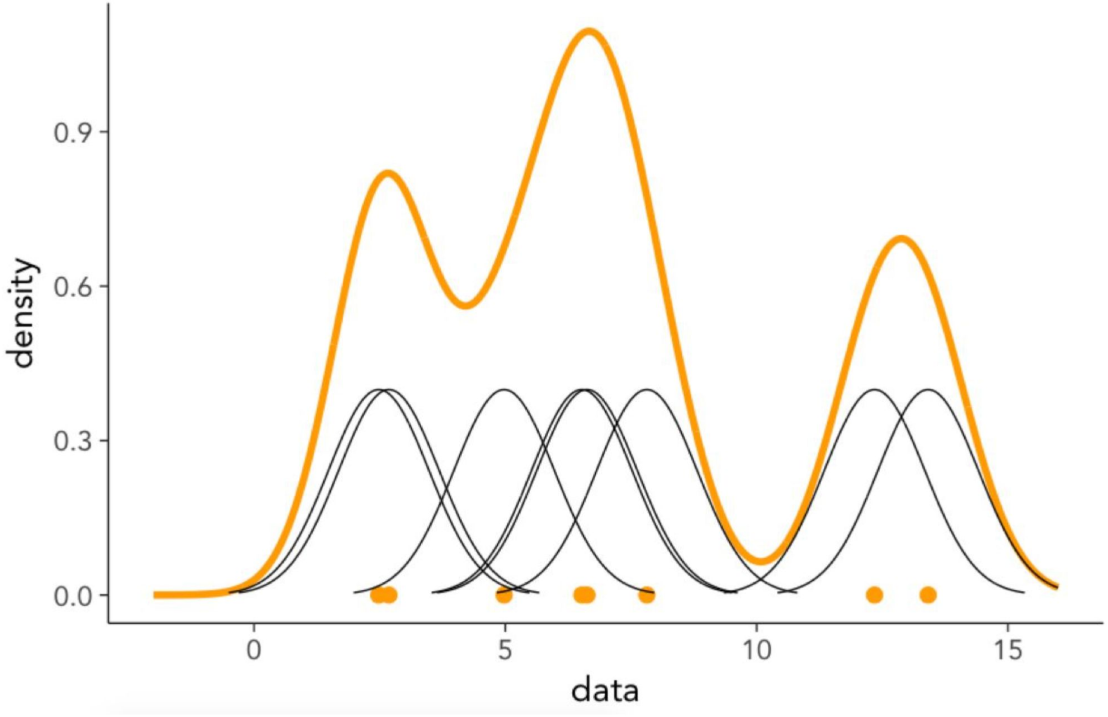
Kernel density estimation



Kernel density estimation



Kernel density estimation



Kernel density estimation

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

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

Kernel density estimation

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

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

Each kernel function is
centered at a data point

Kernel density estimation

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

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

Kernel density estimation

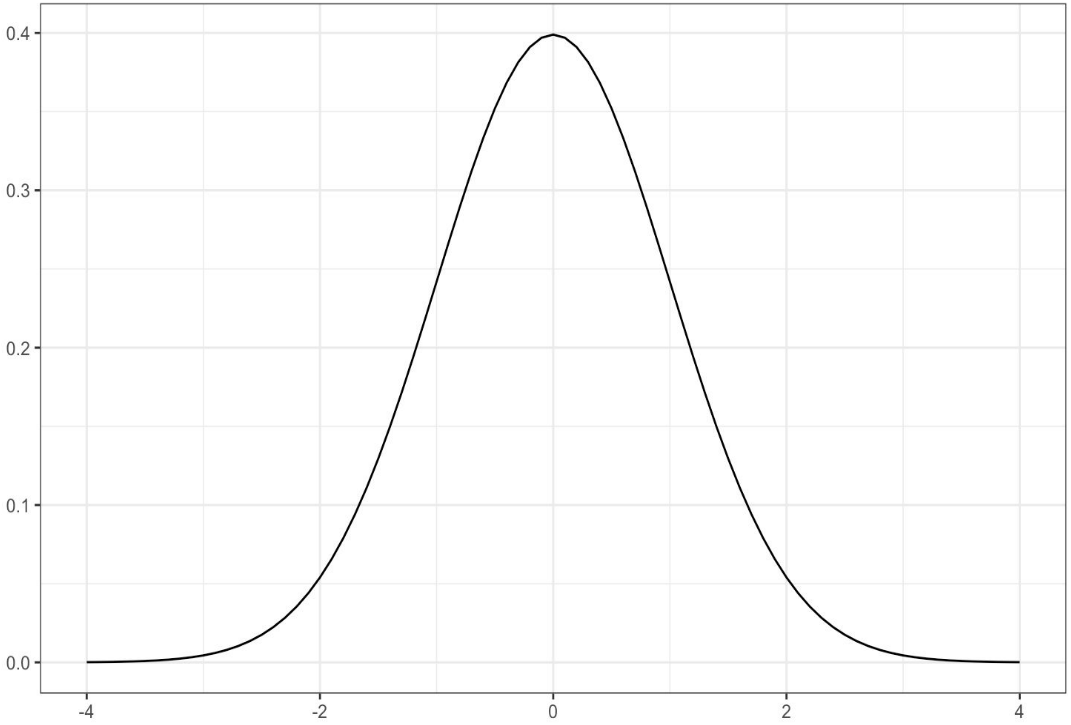
There are many possible Kernel functions that you could use

- Gaussian
- Uniform
- Triangular
- …

Gaussian Kernel

Support: $u \in \mathbb{R}$

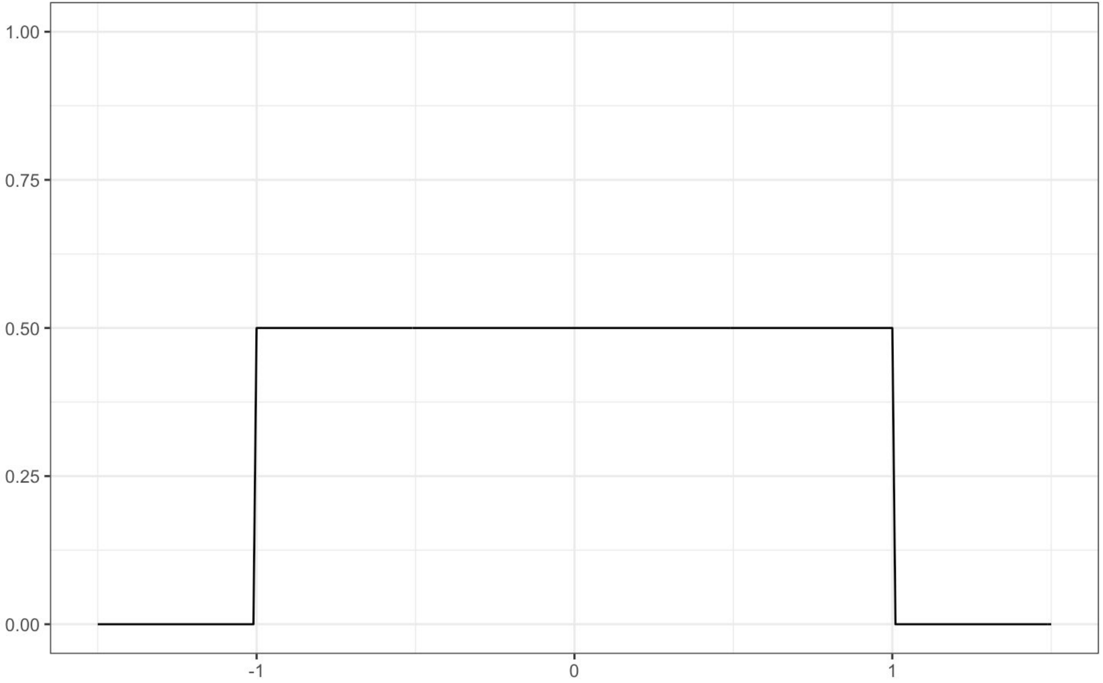
$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$



Uniform Kernel

Support: $|u| \leq 1$

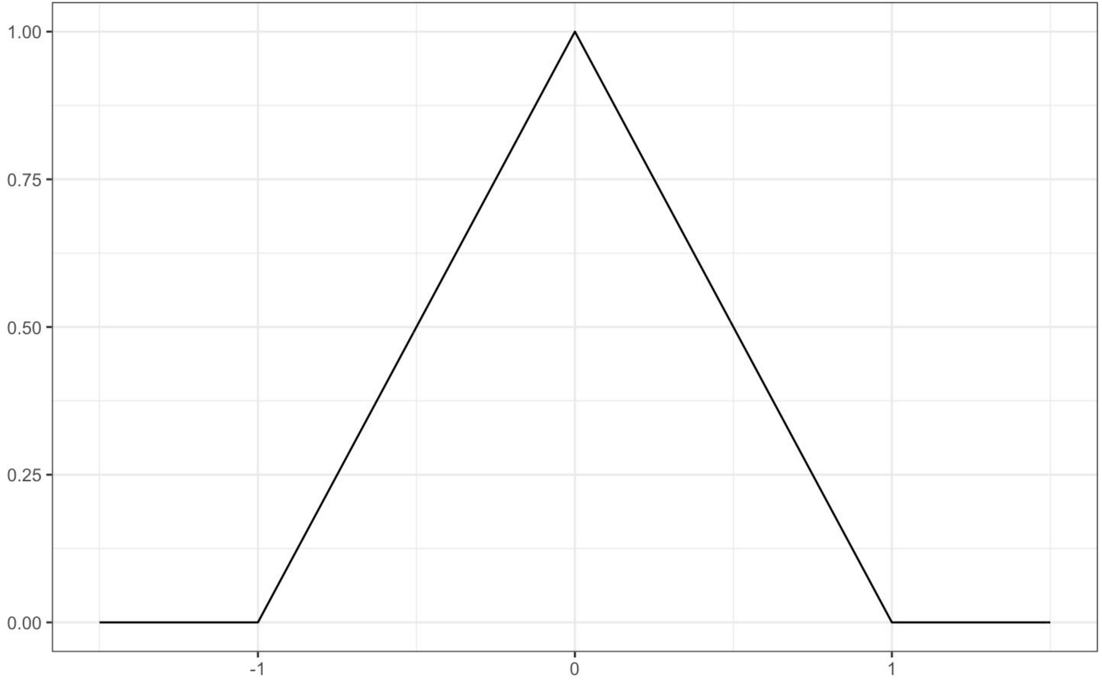
$$K(u) = \frac{1}{2}$$



Triangular Kernel

Support: $|u| \leq 1$

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



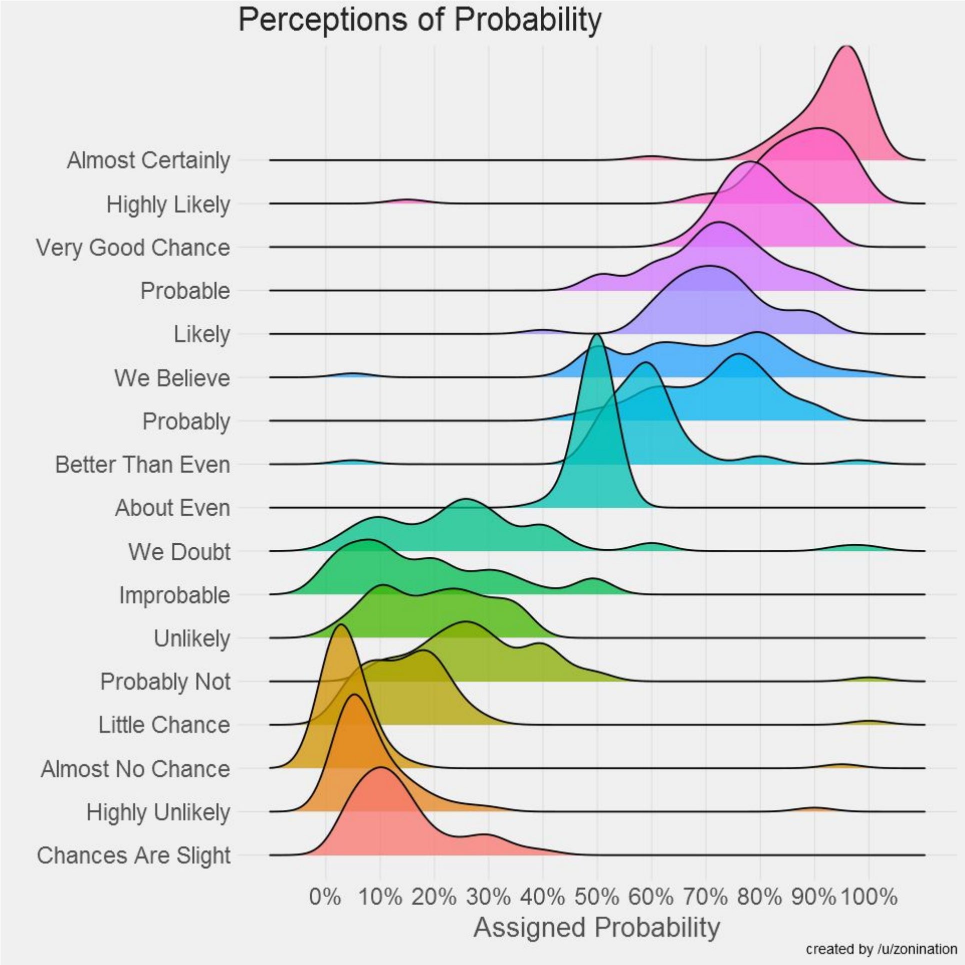
Probability Perceptions

Joyplot

Way to put multiple kernel density estimates on one plot

Is there anything you would change about this plot?

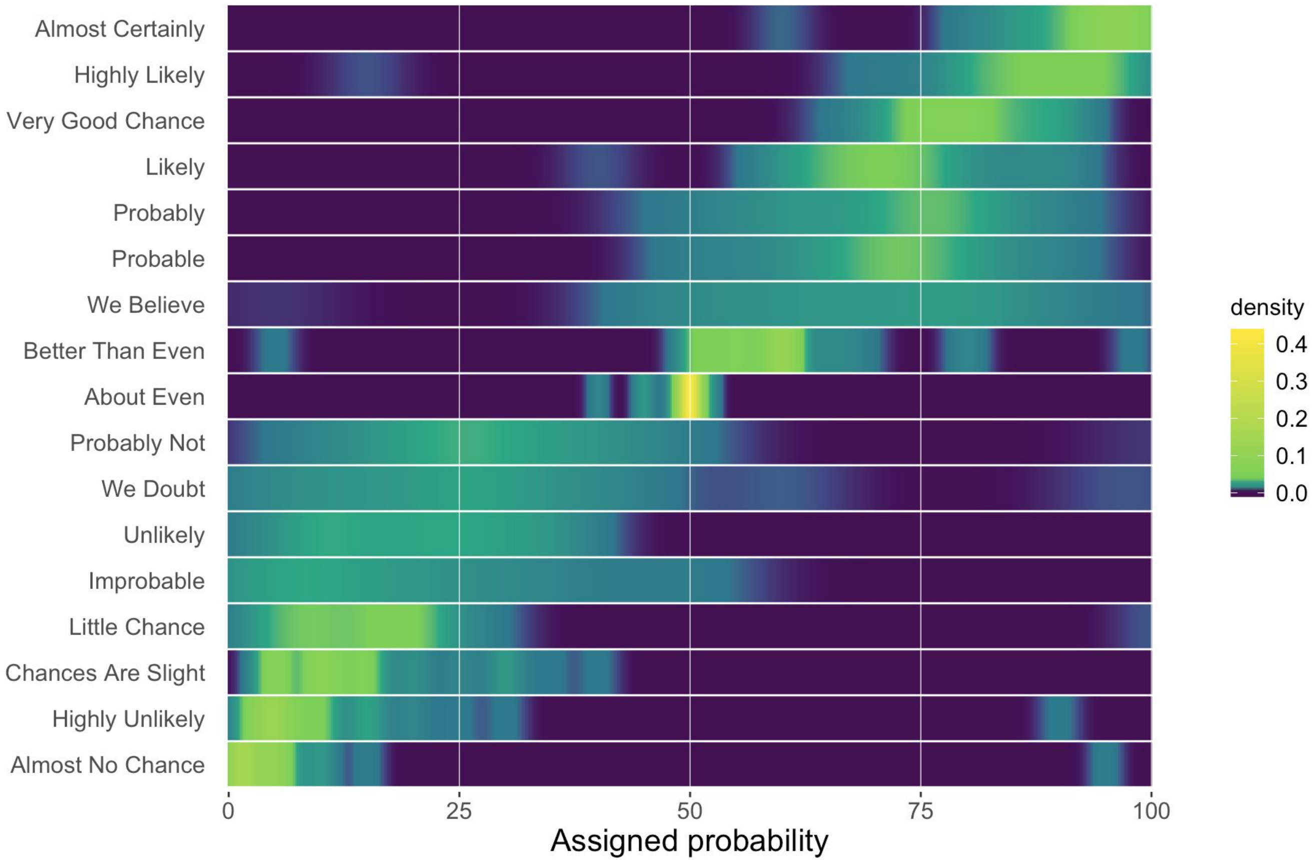
Image and data source: https://github.com/zonination/perceptions



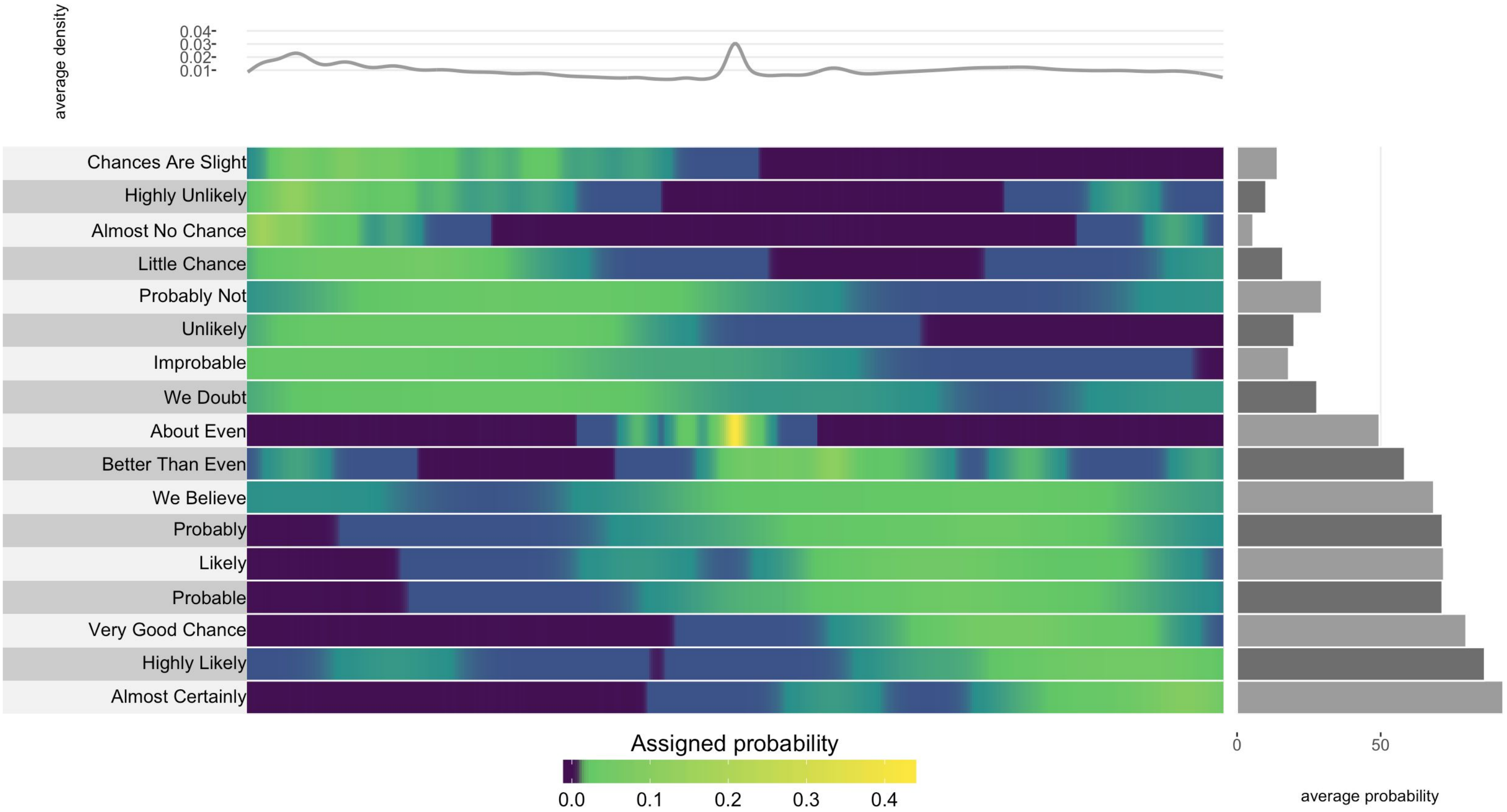
Exercise 1

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

Alternative view of data: heatmap



Alternative view of data: superheat



Color choice

Choosing color palettes

Don’t always use the default in R

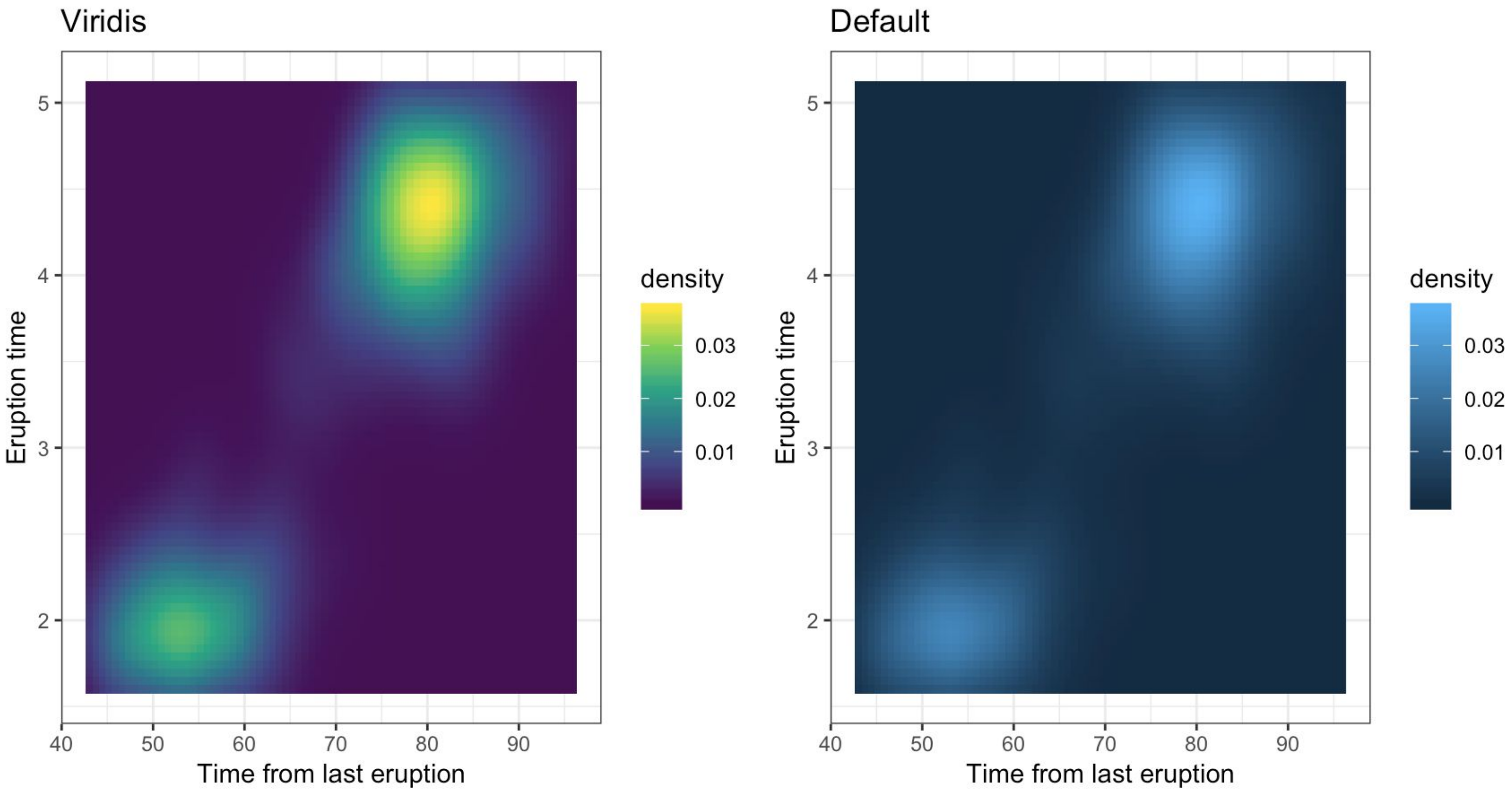
Think about what you are trying to convey in the plot

Color choices can affect the way we perceive the plot

Two helpful websites

1. https://coolors.co/app
2. http://colorbrewer2.org/?

Viridis color scheme



Viridis color scheme

Reasons to use viridis

1. Makes pretty plots!
2. Perceptually uniform colors (meaning changes in the data should be accurately decoded by our brains)
    a. Another colormap with the quality is ColorBrewer
3. Perceived by most common forms of color blindness

Exercise 2

Come up with your own visualization of the perception data.

Be creative!

Interactive plots

Exercise 3

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

Useful R packages

1. plotly: https://plot.ly/r/
2. crosstalk: https://rstudio.github.io/crosstalk/
3. Highcharther: http://jkunst.com/highcharter/
4. Shiny: https://shiny.rstudio.com/gallery/