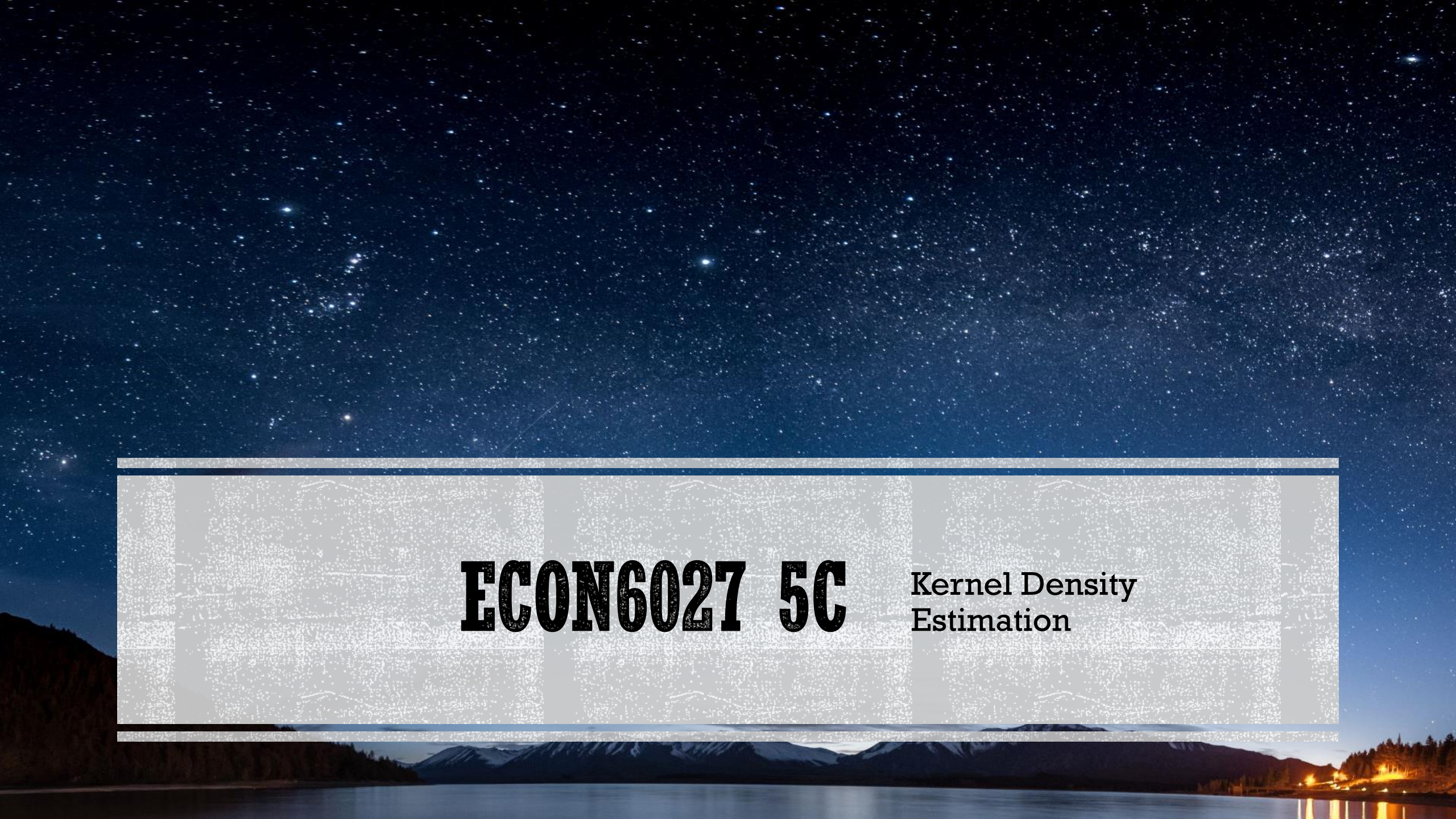


ECON6027 5C

Kernel Density Estimation

(NON-PARAMETRIC) FIRST ORDER ANALYSES

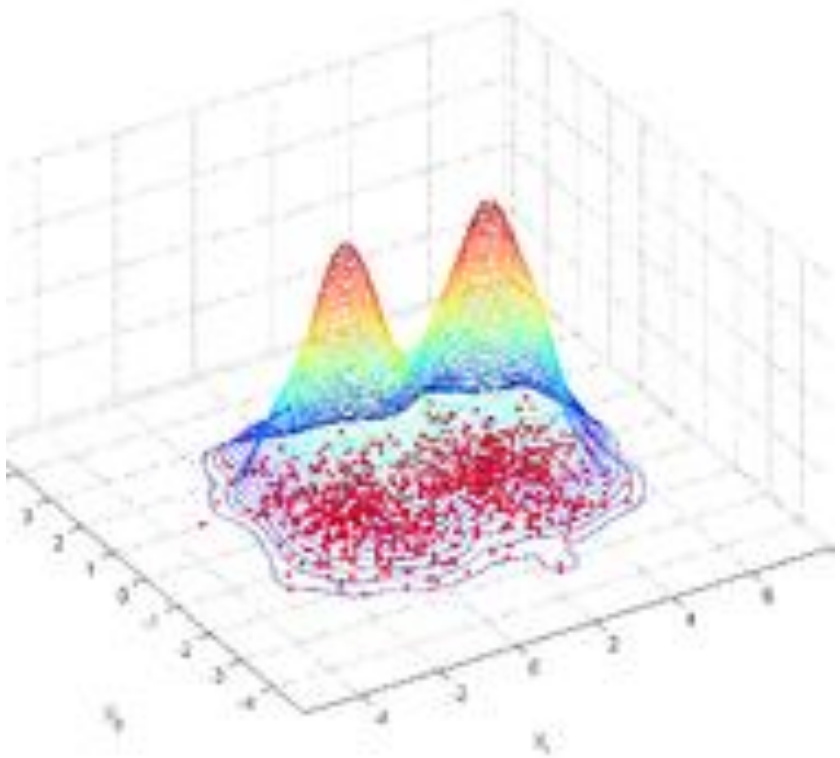
- Premise: each observation is drawn independently from an unknown distribution with the same probability density function (pdf) $f(x)$ (the classic iid assumption).
- This function maps a location x to a probability density.
- If we think of locations in space as a very fine pixel grid, and assume a value of probability density is assigned to each pixel, then summing the pixels making up an arbitrary region on the map gives the probability that an event occurs in that area.
- It is generally more practical to assume an unknown $f(.)$ (rather than say, Gaussian) and estimate it using non-parametric methods.
  - For example, pandemic outbreaks happen in several locations rather than a simplistic radial "bell curve" from an epicentre.
- However, given that $f(.)$ is estimated using "non-parametric" methods, there are no classic hypothesis tests compliant to this method.

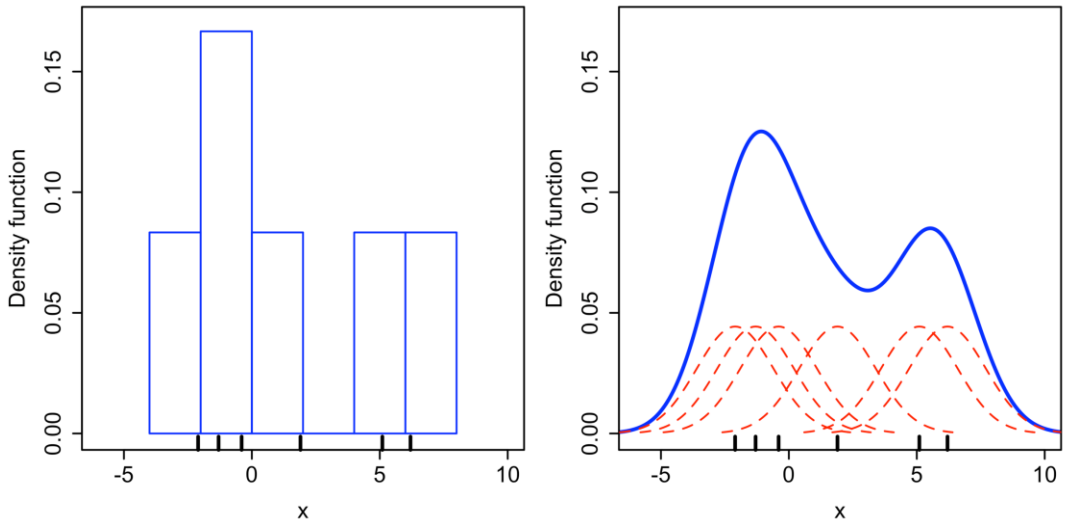
KERNEL DENSITY ESTIMATION

- A common technique to estimate $f(\boldsymbol{x})$ is the kernel density estimation (Silverman, 1986).
- Methodology: average (smoothen) a series of "small bumps" cantered around each observation.

$$\hat{f}(\boldsymbol{x}) = \hat{f}(x, y) = \frac{1}{n h_x h_y} \sum_i k\left(\frac{x - x_i}{h_x}, \frac{y - y_i}{h_y}\right)$$

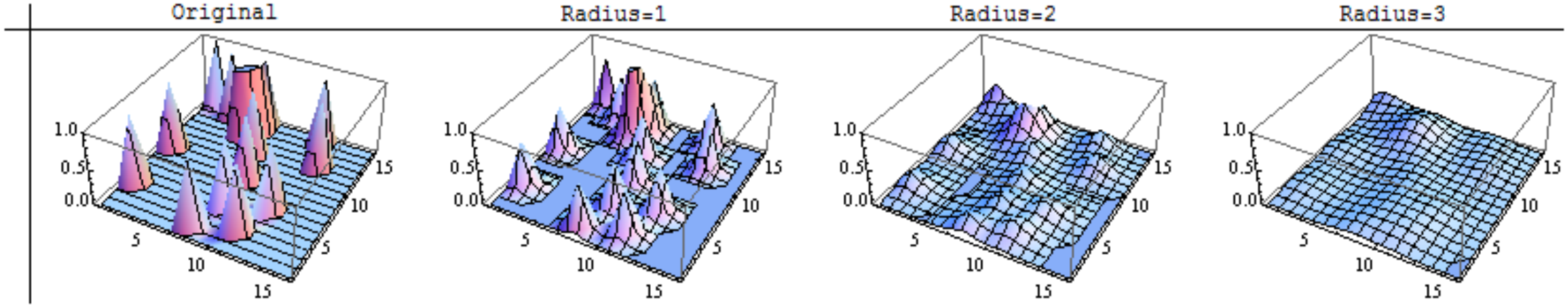
where $k(.)$ is the kernel function that describe the bump averaging process and h_x and h_y are smoothing parameters (bandwidths).





KERNEL DENSITY ESTIMATION

- Comparison of the histogram (left) and kernel density estimate (right) constructed using the same data. The six individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis.
- For the histogram, first the horizontal axis is divided into sub-intervals or bins which cover the range of the data: In this case, six bins each of width 2. Whenever a data point falls inside this interval, a box of height 1/12 is placed there. If more than one data point falls inside the same bin, the boxes are stacked on top of each other.
- For the kernel density estimate, normal kernels with standard deviation 2.25 (indicated by the red dashed lines) are placed on each of the data points x_i. The kernels are summed to make the kernel density estimate (solid blue curve).
- The smoothness of the kernel density estimate (compared to the discreteness of the histogram) illustrates how kernel density estimates converge faster to the true underlying density for continuous random variables.



CHOICE OF BANDWIDTH

- Low bandwidths lead to "spiky" estimates although less biased and large bandwidths lead to "smooth" estimates that are largely biased.
- Typically $h_x = h_y$
- There is no exact way to choose bandwidths, a common rule is what is known as the Scott's rule (Scott, 1992),

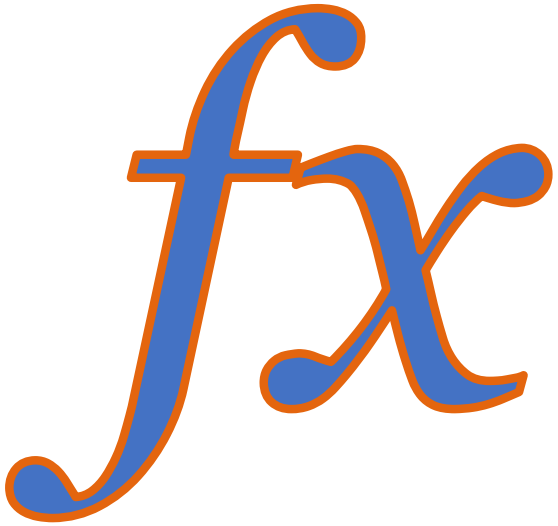
$$h_i = \sigma_i \left(\frac{2}{3n}\right)^{\frac{1}{6}}$$

- A mean square error based procedure to choose bandwidth parameter includes cross validated bandwidth selection for kernel density (Berman and Diggle, 1989)., and likelihood cross validated bandwidth selection for kernel density.

KERNEL FUNCTION

- Any function $k(u)$ with properties,
  - Normalisation: $\int k(u).du = 1$ (i.e. a valid pdf) and
  - Symmetry: $\int uk(u).du = 0$ this means $k(u) = k(-u)$ cantered around 0.

  can serve as a kernel function.
- Popular kernel functions:
  - Uniform kernel (rectangular, gives equal weight to all points in the neighbourhood): $k(t) = \begin{cases} 0 & |t_i - t_0| > h \\ 1 & o/w \end{cases}$
  - Gaussian kernel (default in R): $k(t) = \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{t^2}{2}\right\}}$
  - Quadratic (Epanechnikov) kernel: $k(t) = \begin{cases} \frac{3}{4}(1 - t^2) & |t| \leq 1 \\ 0 & o/w \end{cases}$
  - Minimum variance kernel: $\mathrm{k(t)} = \begin{cases} \frac{3}{8}(3 - 5t^2) & |t| \leq 1 \\ 0 & o/w \end{cases}$
- **The choice of the kernel function is less important than the choice of bandwidth bandwidth in practice.**



PACKAGES YOU NEED

- sp
- spatstat

ESTIMATE KERNEL DENSITY (PRESELECTED BANDWIDTH)

- KDE is highly sensitive to the choice of the bandwidth.
- Here we make two estimates for sigma=500 and sigma=1000 respectively.

```
> plot(density(breach.ppp, sigma=500))
> plot(density(breach.ppp, sigma=1000))
> ?density.ppp # for more information
```

ESTIMATE KERNEL DENSITY (OPTIMAL BANDWIDTH)

```
> (breach.bw = bw.diggle(breach.ppp))
```

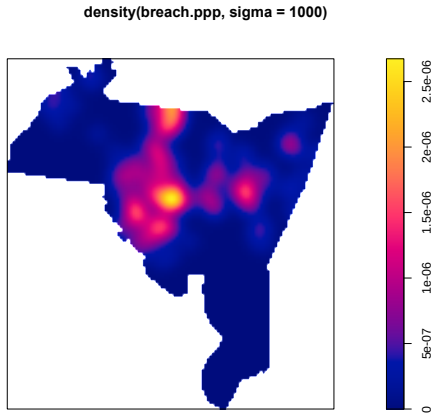
sigma

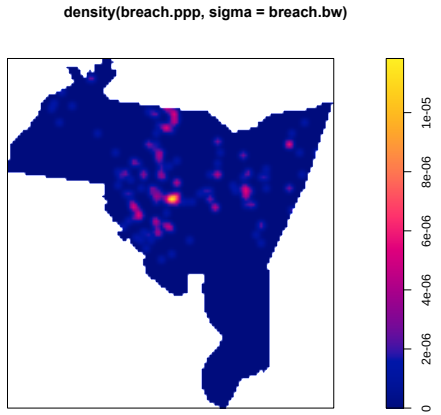
333.6992

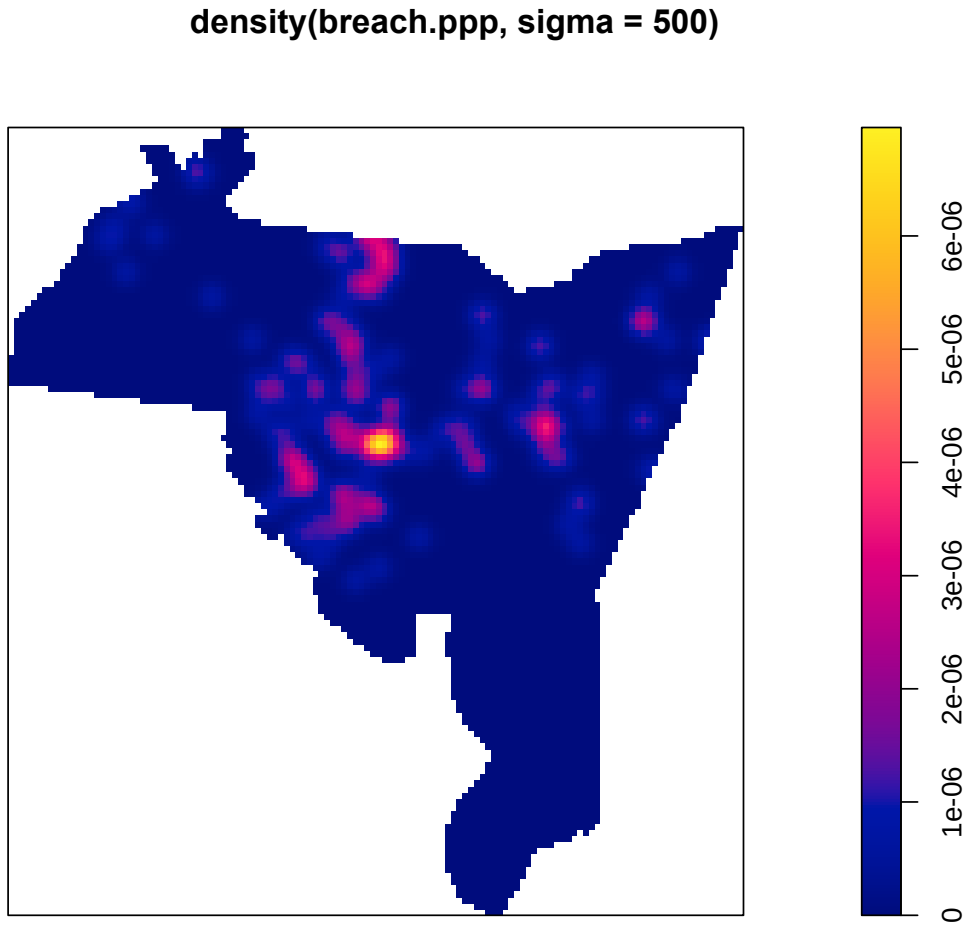
```
> breach.kde1 = density(breach.ppp,
sigma=bw.diggle(breach.ppp))
```

```
> plot(breach.kde1)
```

The function "density" is able apply weights in the kernel estimation and applies an automatic edge correction (which can be set to False).


density(breach.ppp, sigma = 1000)
0
5e-07
1e-06
1.5e-06
2e-06
2.5e-06


density(breach.ppp, sigma = breach.bw)
0
2e-06
4e-06
6e-06
8e-06
1e-05


density(breach.ppp, sigma = 500)
0
1e-06
2e-06
3e-06
4e-06
5e-06
6e-06

ACTIVITY A

- Prepare kernel density estimates for for the attributes: burglaries, forced vs. not ("burgres.f" vs. "burgres.n")
- Comment on the differences you observe.
- Estimate the KDE for "breach" using other methods of bandwidth selection (e.g.: bw.diggle, bw.CvL, bw.scott or bw.ppl) and compare.
- Estimate the KDE for "breach" using different kernel functions and compare.

TAKE HOME POINTS…

- Non-parametric fist order analysis of point patterns using kernel density estimates.
- Kernel functions
- Bandwidth selection: impact of the bandwidth on the KDE and methods of bandwidth selection

REFERENCES

- ***Spatial Analysis*** by Tonny Oyana, 2nd edition, Chapter 6.
- https://en.wikipedia.org/wiki/Kernel_density_estimation
- Silverman, B.W. (1986) Density estimation for statistics and data analysis. London: Chapman & Hall.
- https://bookdown.org/lexcomber/brunsdoncomber2e/Ch6.html
- https://cran.r-project.org/web/packages/spatstat/spatstat.pdf
- https://cran.r-project.org/web/packages/maptools/maptools.pdf