Assignment

November 21, 2018

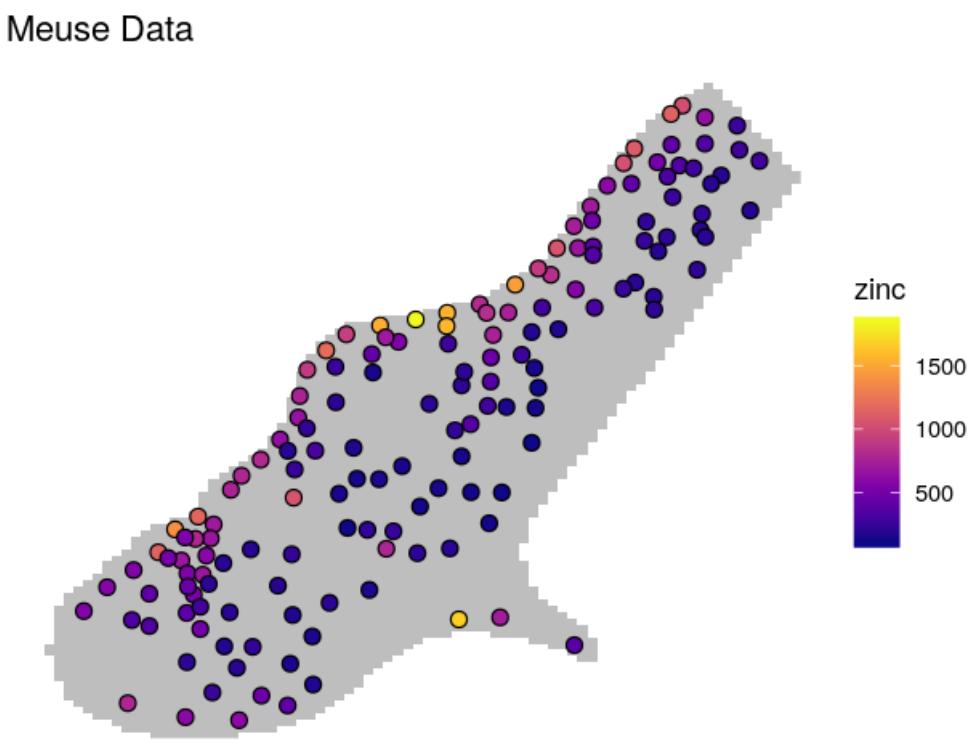
```
In [1]: data(meuse, package="sp")
```

1) Plot a histogram of the values of zinc. (3 pts)
2) What range of values fits better what we observe in the histogram: all real numbers, integers, continuous values between 0 and 1, non-negative numbers? (3 pts)
3) Using the method caret::createFolds(), define an index to split the meuse data into 2 subsets of data (train and test). (3 pts)
4) Using only the train set and fit a GLM to explain zinc with cadmium as a single covariate. Set the parameter: family = gaussian(link="log") (3 pts)
5) Compute the correlogram based on Moran's coefficient for the residuals (3 pts)
6) Is there any significant spatial autocorrelation in the residuals? (3 pts)
7) Now add a spatial effect with a GAM. (3 pts)
8) Compute the correlogram based on Moran's coefficient for the residuals. Is there significant spatial autocorrelation in the residuals? (3 pts)
9) Predict the values of the test set using the GLM and the GAM. (3 pts)
10) Compute the mean squared error for the predicted values and find which one has the smallest generalization error. (3 pts)

Total 30 pts

0.0.1 Visualization of Meuse data

```
In [2]: library(ggplot2)
        options(repr.plot.width=5, repr.plot.height=4)
        data(meuse, meuse.grid, package="sp")
        grid_df <- data.frame(rx = meuse.grid[,1], ry=meuse.grid[,2])
        plt <- ggplot() +
                geom_raster(data=grid_df, aes(rx,ry), fill="grey") +
                geom_point(data=meuse, aes(x, y, fill=zinc), pch=21, size=2.5) +
                viridis::scale_fill_viridis(option="plasma", na.value="darkblue") +
                ggtitle("Meuse Data") +
                theme_void()
        plt
```


Meuse Data
zinc
1500
1000
500