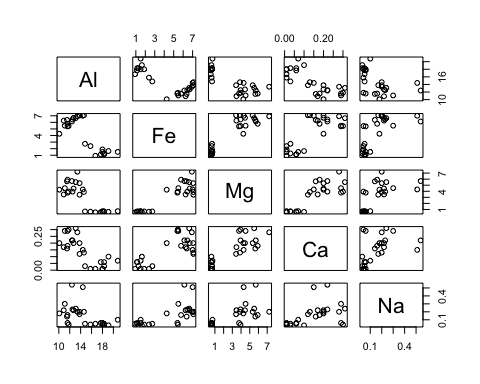
3R\_Esercitazione

Francesca Pecorari

2024-01-19

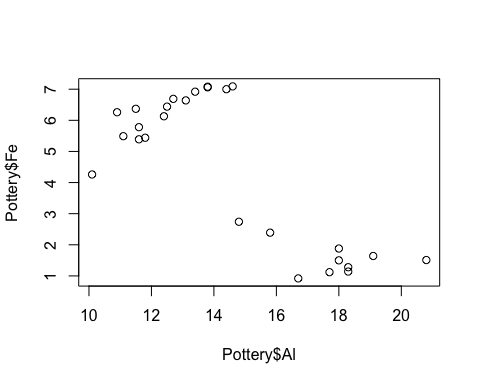
# Commento ai grafici ottenuti dal dataset Pottery

## Plots

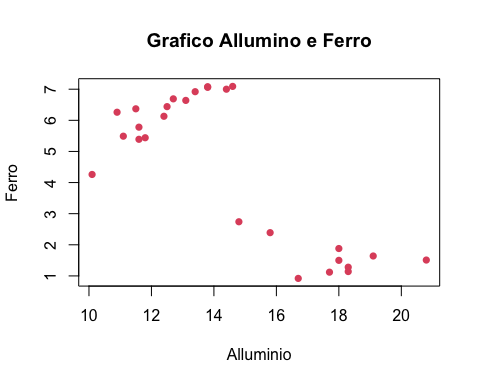


Il grafico prodotto mostra il comportamento di ognuna delle variabili rispetto ad ognuna delle altre in un unico grafico.

È possibile ottenere anche un grafico specifico per una coppia di variabili, in questo caso si visualizza quello delle variabili Al e Fe.

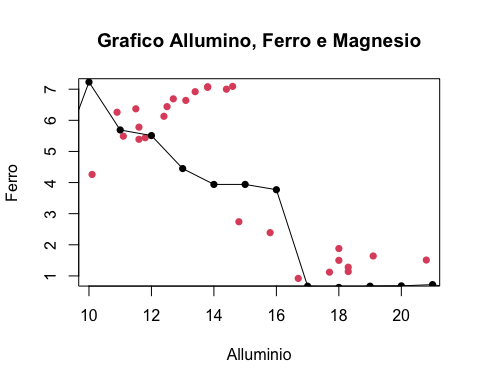


È possibile avere una visualizzazione più efficace aggiungendo anche colori ed etichette ai grafici.



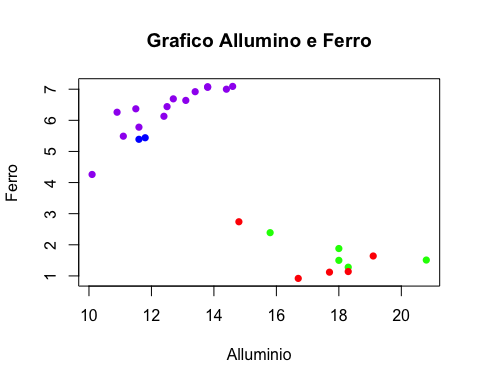
Si può affermare che la variabile Al avrà probabilmente una correlazione negativa con la variabile Fe, vista la distribuzione dei punti nel grafico.

È possibile aggiungere un terzo set di dati al grafico.

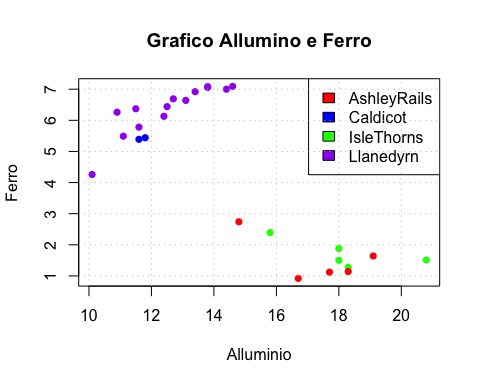


Si nota che probabilmente anche la variabile Al e la variabile Mg sono correlate negativamente.

Nel grafico sottostante è possibile vedere, oltre il comportamento della variabile Fe rispetto alla variabile Al, anche la divisione dei dati nelle varie categorie della variabile Site.

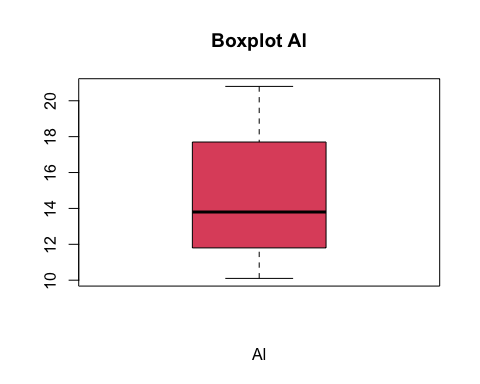


Nel seguente grafico sono state aggiunte una griglia e una legenda per poter comprendere meglio l’appartenenza dei punti a una determinata categoria.

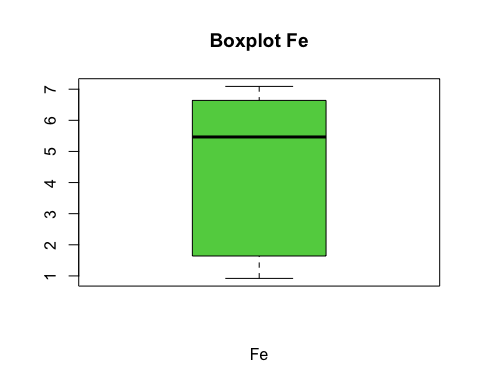


## Boxplot

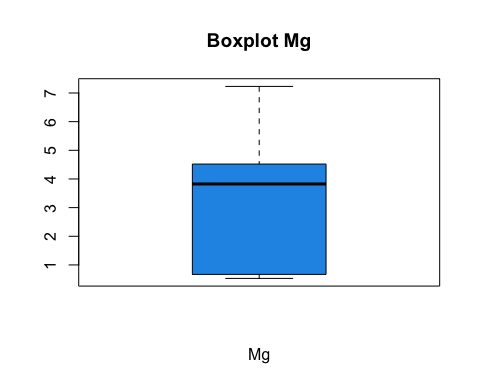
I boxplot sono stati realizzati per ognuna delle variabili numeriche e ci indicano la mediana, i quartili e gli eventuali outliers.



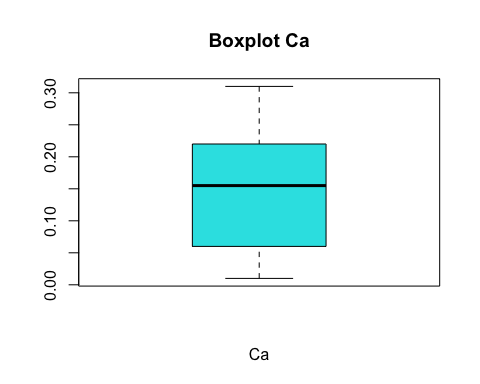
Dal grafico notiamo che la variabile Al ha una mediana di circa 14 e che il primo e il terzo quartile hanno valori intorno a 12 e 21 rispettivamente.



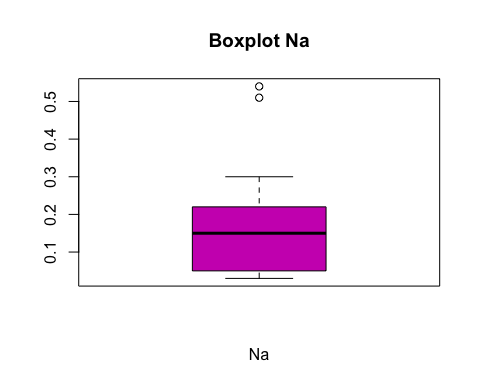
Dal grafico notiamo che la variabile Fe ha una mediana leggermente superiore a 5 e che il primo e il terzo quartile hanno valori leggermente inferiore a 2 e leggermente superiore a 6 rispettivamente.



Dal grafico notiamo che la variabile Mg ha una mediana di circa 4 e che il primo e il terzo quartile hanno valori inferiore a 1 e leggermente superiore di 4 rispettivamente.

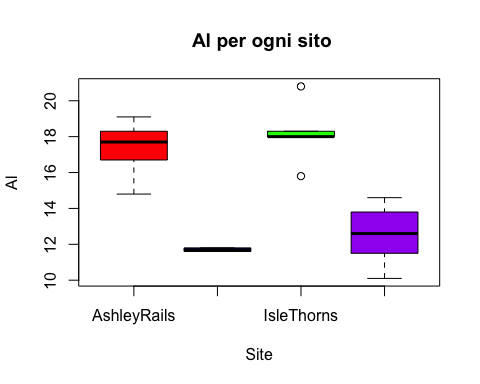


Dal grafico notiamo che la variabile Ca ha una mediana di circa 0.15 e che il primo e il terzo quartile hanno valori intorno a 0.05 e 0.20 rispettivamente.



Dal grafico notiamo che la variabile Na ha una mediana leggermente inferiore a 0.2 e che il primo e il terzo quartile hanno valori inferiore a 0.1 e leggermente maggiore di 0.2 rispettivamente. In questo caso, si nota anche la presenza di alcuni outliers.

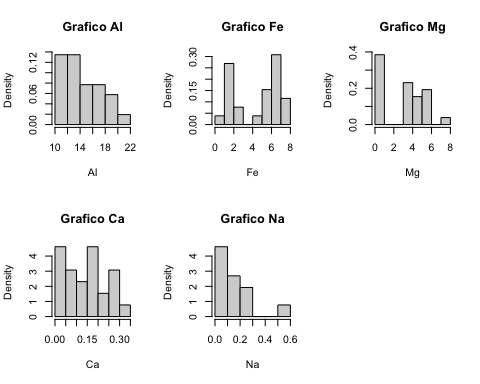
Nel seguente grafico è possibile visualizzare i valori della variabile Al rispetto ad ogni categoria della variabile Site.



Notiamo che le categorie in cui la variabile Al ha mediana più alta sono AshleyRails e IsleThorns, con una differenza nei valori dei quartili, per le categorie Caldicot e Llanedyrn la mediana ha valori più bassi e anche in questo caso, la differenza principale sta nei quartili; i quartili primo e terzo della distribuzione di Llanedyrn sono molto più lontani dal valore della mediana rispetto a quelli di Caldicot.

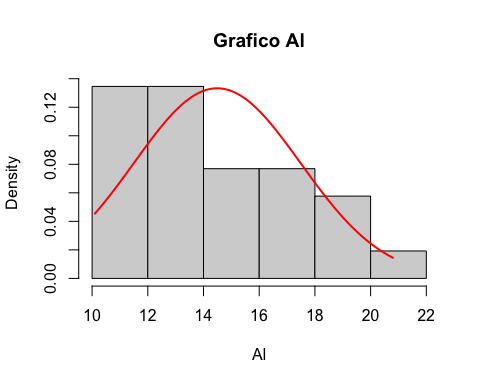
## Istogrammi

Gli istogrammi ci permettono di visualizzare la densità di probabilità delle variabili.



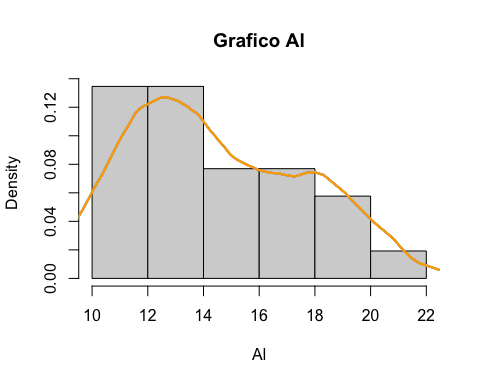
Nei grafici riportati sopra è possibile vedere le diverse distribuzioni per ognuna delle variabili del nostro dataset. Si nota che nessuna di queste sembra avere una distribuzione gaussiana, per questo, se si dovesse procedere in un’analisi che richiede come ipotesi la gaussianità dei dati, potrebbe essere opportuno applicare delle trasformazioni.

È possibile sovrapporre all’istogramma una funzione gaussiana di densità di probabilità.



In questo caso si nota che la curva non delinea in modo soddisfacente la nostra distribuzione di dati, questo perché, come detto sopra, i dati sono asimmetrici.

Un’ulteriore possibilità è quella di sovrapporre all’istogramma una funzione di densità kernel di probabilità, in questo caso, anche utilizzando diversi tipi di funzioni kernel il risultato ottenuto era sempre il medesimo.

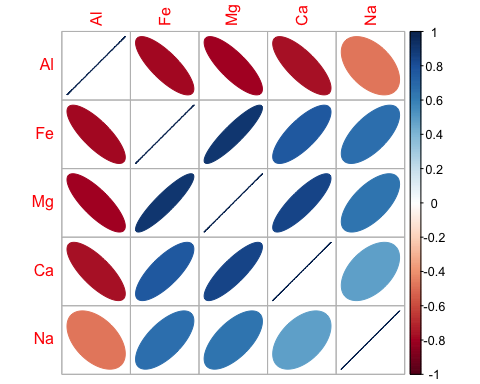


A differenza della funzione gaussiana, qui la curva accompagna meglio l’andamento dei dati.

## Corrplot

Otteniamo il grafico di correlazione.

## corrplot 0.92 loaded



Dall’ interpretazione di questo grafico si può dedurre che:

* la variabile Al ha un’alta correlazione negativa con le variabili Fe, Mg e Ca e una correlazione leggermente minore con la variabile Na
* la variabile Fe ha una forte correlazione positiva con la variabile Mg e una leggermente più lieve con Ca e Na
* la variabile Mg ha una forte correlazione positiva con la variabile Ca e una più lieve con la variabile Na