## Deuxième partie : Introduction

Dans la première partie de ce rapport nous avons explicité le contexte général des données sur lesquelles nous comptons effectuer nos analyses. En effet, nous y avons exposé en détails le jeu de données « *Forest Fires »*. Cette explication du contexte s’est conclue avec quelques questions ouvertes qui permettraient de mieux comprendre les feux de forêts qu’on peut observé dans le parc de Montesinho. Dans cette deuxième partie du rapport il sera question de soutirer de l’information de ces données afin de mieux comprendre ce qui caractérisent ces feux de forêts. Par la suite, nous utiliserons ces notions afin de répondre aux questionnements proposés dans la première partie du rapport. Afin de bien comprendre le phénomène à l’étude il est essentiel de débuter par une phase d’analyse statistique descriptive. En effet, les 13 variables qui composent notre jeu de donnée, ne sont pas intelligibles, c’est-à-dire qu’il est difficile de ressentir les tendances qui émanent de celle-ci lorsqu’on les visualise dans une matrice. Autrement dit, il nous faut décrire les données en les résumant et en les structurant [4 p.192]. À partir de ces descriptions il nous sera possible de remarquer certaines propriétés qui caractérisent une forêt favorisant les incendies. Ensuite, nous exposerons les potentielles distributions probabilistes sous-jacentes à certaines de ces variables. Pour ce faire, nous utiliserons d’abord notre instinct visuel. En effet, certaines de ces variables semblent d’emblée exprimée des signes de distribution normale. Nous utiliserons ensuite les tests du chi carré et de afin de vérifier si il est probable ou non que ces variables soient effectivement issues d’une distribution normale. Ces vérifications seront fait sous la proposition d’une hypothèse nulle stipulant qu’elles suivent effectivement une loi normale. Comme nous le verrons, ces hypothèse s’avèreront improbable selon la *p-value* choisi. Afin de renforcir cette conclusion, nous utiliserons ensuite le teste de *Shapiro Wilk* selon la même p-value. Nous poursuivrons notre investigation concernant les distributions potentielles de certaines variables à l’aide d’une batterie de test un peu plus élaboré. En effet, afin de cibler certains candidats potentiels, nous analyserons un graph le graph de *Cullen et Frey* qui compare les propriétés d’aplatissement et d’asymétrie de certaines distributions avec les variables qui nous intéresse. À partir de ces résultats nous calculerons le critère d'information d'Akaike afin de trancher sur quel modèle continuer notre investigation. Enfin, nous passerons le test de Kolmogorov–Smirnov afin de rendre compte de la véracité probabiliste de notre hypothèse sur les variables à l’étude. Nous verrons que certaines variables comme la température semble être exprimer une distribution de type Weibull. Une fois que nous aurons cerné les distributions sous-jacentes à ces variables, il nous sera possible d’estimer certains paramètres de la population d’où proviennent ces variables. Enfin, nous terminerons notre analyse statistique par une proposition d’un modèle de régression sur deux variables qui semblent partager une interaction. Finalement, nous conclurons en soulignant les résultats qui nous semblent pertinents et en posant quelques ouvertes.

## 1) Analyse descriptive des données

Dans le but de bien cerner certaine tendance de l’état de la forêt lors d’incendie dans le parc de Montesinho, nous avons généré des histogrammes pour les variables qui composent le jeu de données. Nous avons joint à ces figures des légendes contenant certaines statistiques descriptives des variables à l’étude. Ces statistiques sont la moyenne, l’écart-type, le coefficient de symétrie et le coefficient d’aplatissement. Nous avons ignoré certaines variables comme le jour et les mois et les précipitations. Les deux premières car elle ne nous semblent d’aucun intérêt pour le moment et la dernière, car comme on s’en doute, les précipitation lors les incendies de forêts sont toujours nulle.

## 1.1) L’indice d’humidité de l’humus (*DMC)*

Comme mentionné dans la première partie, le *DMC* représente l’indice d’humidité du humus, c’est-à-dire le facteur de combustion de la couche organique situé entre 5cm et 10cm sous le sol. Il est à noter que cette section du sol n’est pas affectée par les pluies de moins de 20mm. Cette section du sol est aussi protégée des périodes de sècheresse rapide. On remarque sans surprise que les instances de feux de forêts sur un sol ayant un indice de *DMC* très élevé se font plus rare. Par contre, on remarque une certaine tendance : le plus grand nombre d’instance de feux on une métrique de *DMC* qui semblent au milieu de l’échelle. On remarque aussi un écart type assez grand par rapport à l’échelle en question. Les index d’humidité du humus sont donc plutôt dispersés dans un scénario de feux de forêts. Cet indice sera utilisé dans les sections qui suivent. En effet, comme mentionné dans à la fin de la section 1, nous aimerions savoir si il existe une relation entre le DMC d’une forêt et son indice ISI (voir page 2). L’histogramme de fréquence des métrique observées de cet indice en plus de sa boite à moustache.

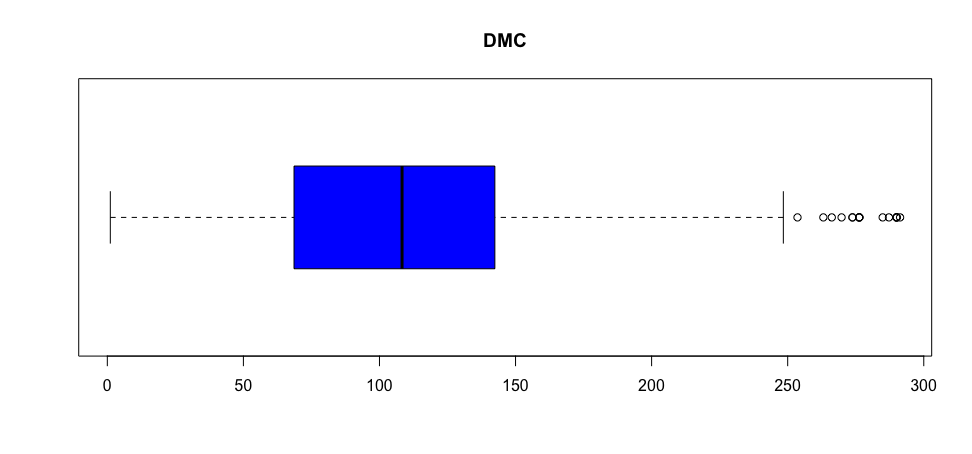
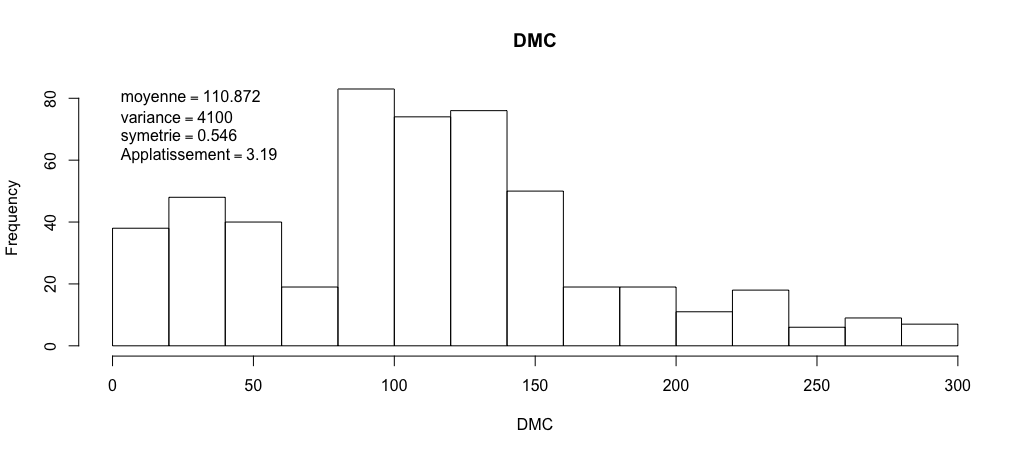
**

Figure 1 Histogramme et boite à moustache de l'indice DMC

## 1.2) Indice de combustible léger (*FFMC*)

L’indice de combustible léger correspond au niveau d’humidité de végétation de moins de 1cm de diamètre (foin, herbes, petites branches). On s’intéresse alors ici à une couche de la végétation un petit peu plus en surface que celle précédemment étudié. Contrairement au *DMC,* le *FFMC* est beaucoup sensible au changement de température et aux précipitations du au fait que le matériel organique se retrouve en surface. On remarque dans l’histogramme du FFMC que les feux de forêts sont très similaires quant à leur indice de combustible léger. En effet, on remarque que l’écart-type est de 5.52 ce qui est très petit compte tenu de l’échelle en question. La boite à moustache confirme cette idée de par sa forme compacte.

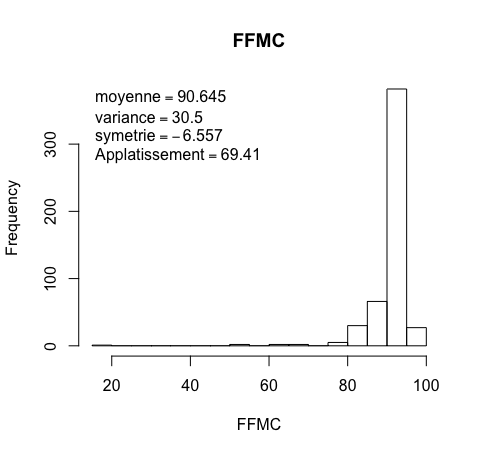
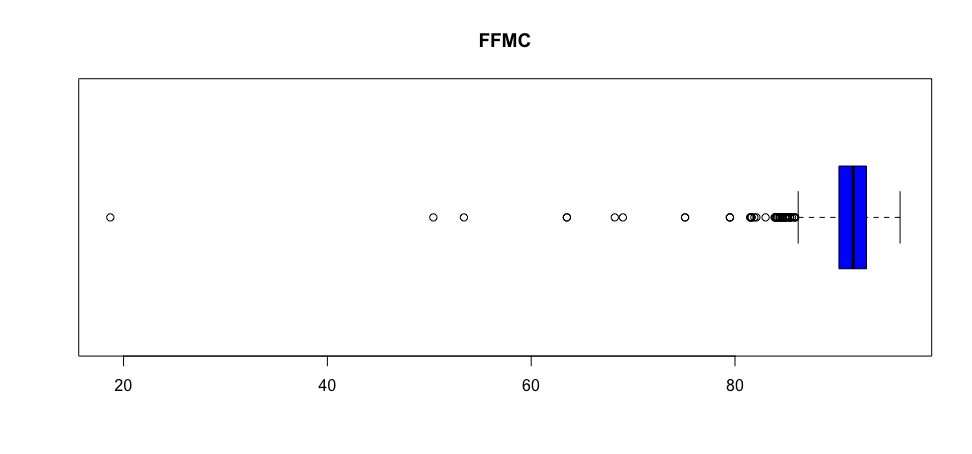
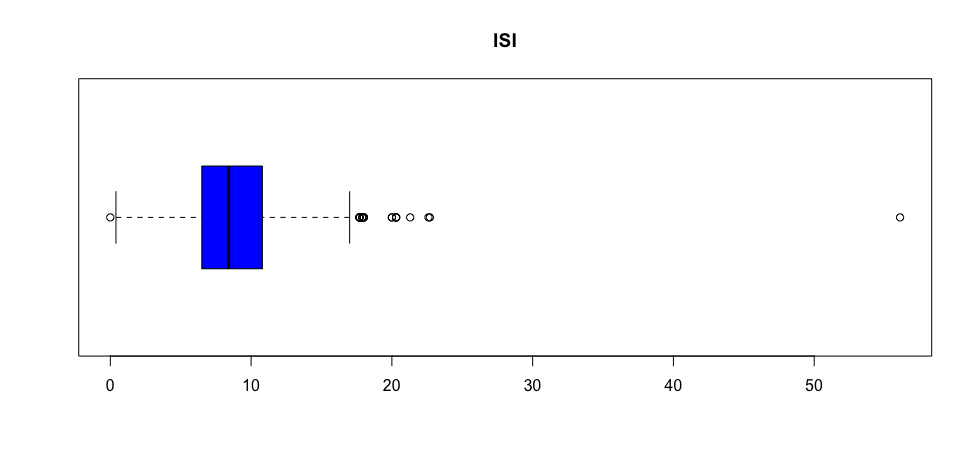


Figure 2 Histogramme et boite à moustache du FFMC

## 1.3) Indice de propagation initiale (*ISI)*

L’indice de propagation initiale (*ISI)* correspond à une métrique servant à évaluer à quel point un feu gagne en envergure par rapport au temps. Encore une fois, on remarque une certaine tendance dans début de l’histogramme. La boite a moustache montre bien qu’un feu typique semble avoir un ISI qui gravite autour de 10.



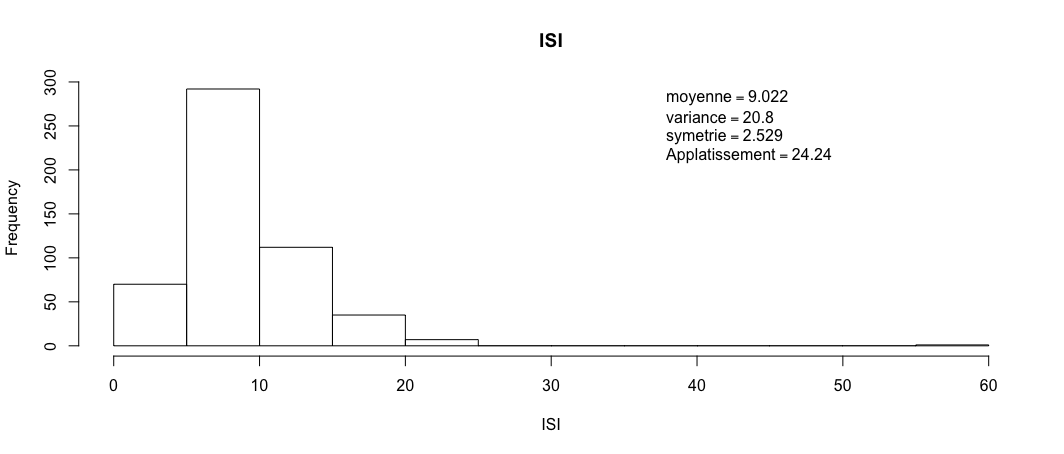
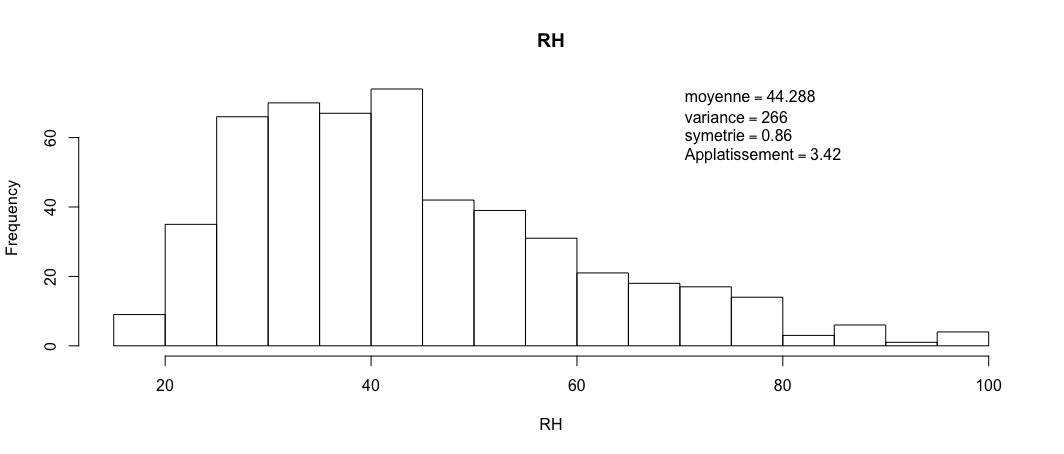
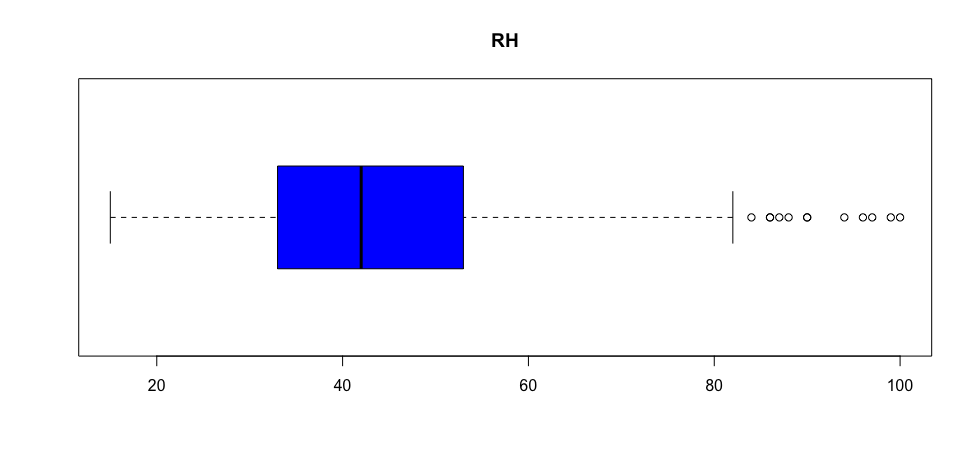


Figure 3 Histogramme *ISI* et sa boite à moustache

## 1.4) Humidité relative (*RH*)

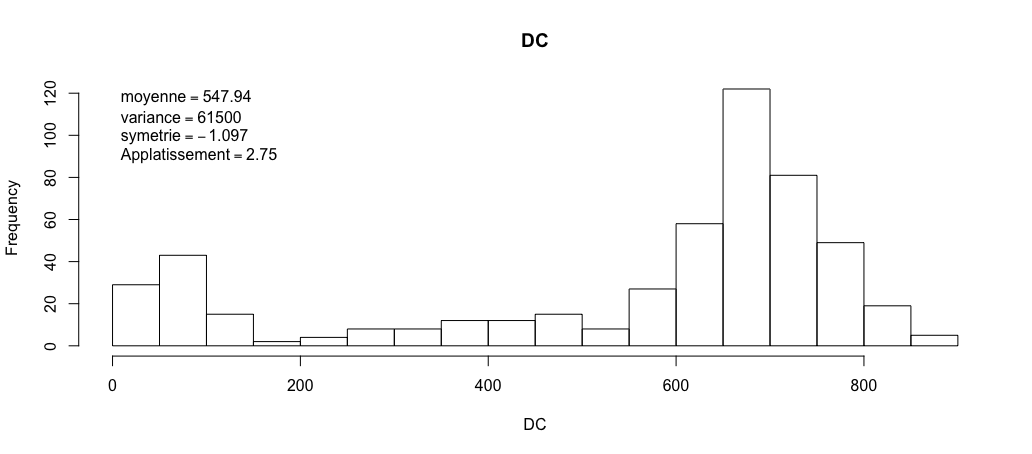
En ce qui concerne le niveau d’Humidité relative, on observe une progression vers le bas au niveau des fréquence plus le niveau d’humidité augmente. Il est intéressant de remarquer que les feux typiques ne profitent pas particulièrement d’un niveau d’humidité relative en dessous de 30. Ceci est relativement surprenant. En effet, on s’aurait attendu à ce que le plus grand nombre d’instance de feux de forêts se produisent lorsque l’indice d’humidité est au plus bas. La boite à moustache et l’histogramme témoigne tout les deux





## 1.5) Indice de sècheresse (*DC*)

L’indice de sècheresse correspond à un indice d’humidité du contenu en profondeur de la matière organique compacte [5]. Le DC est indicatif des périodes d’humidité du contenu organique sur de longues périodes étant donné que ce contenu est en profondeur. L’histogramme montre une tendance normale à la fin de la distribution ; ce à quoi on s’attendrait de façon intuitive. En effet, il fait du sens de penser qu’une forêt qui favorise les feux est généralement plus sèche. La boite à moustache confirme encore une fois cette idée et montre bien les quelques *outlier* au début de la distribution sont potentiellement des données aberrantes.



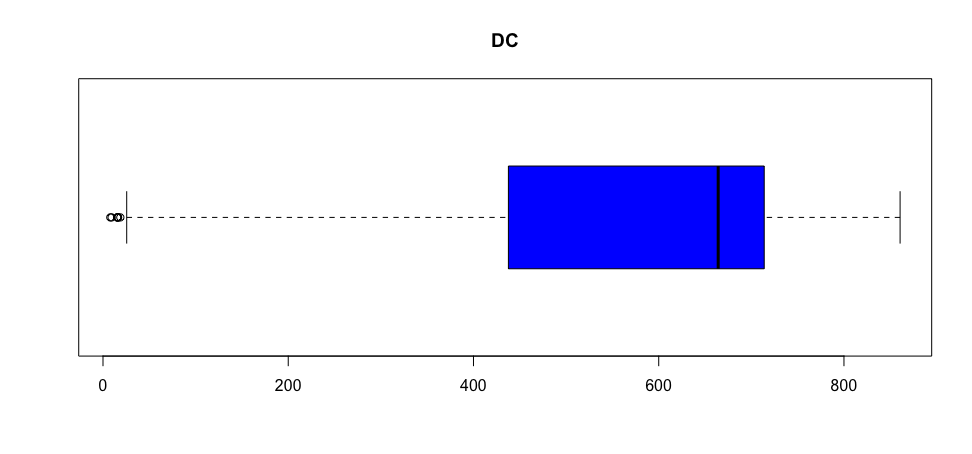


Figure 4Histogramme et boite à moustache de la métrique *DC*

## 1.6) Indice de température

À l’œil, l’indice de température nous suggère une distribution normale. En effet, l’histogramme dénote des signes de distribution plutôt uniforme. Nous étudierons d’ailleurs cette hypothèse dans la prochaine section. Qui plus est, on remarque que le début de la distribution parait un peu curieux pour une distribution normale.

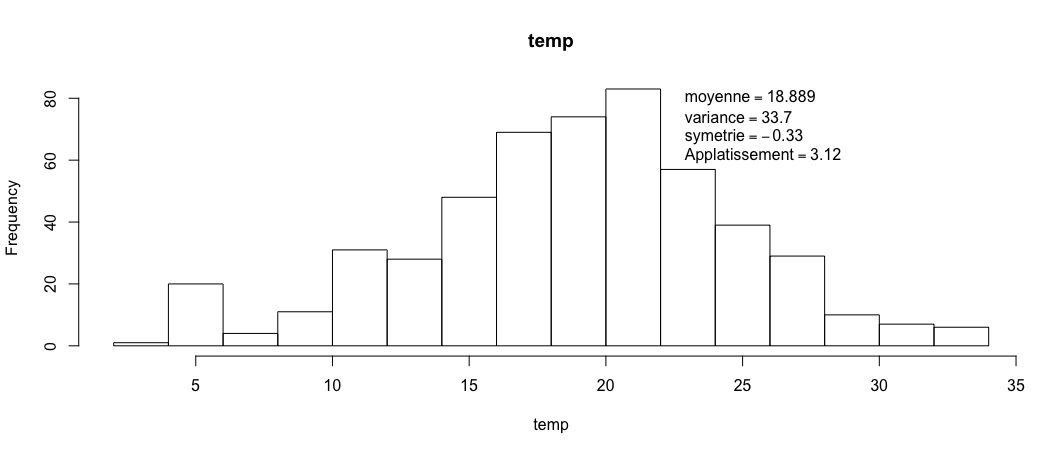
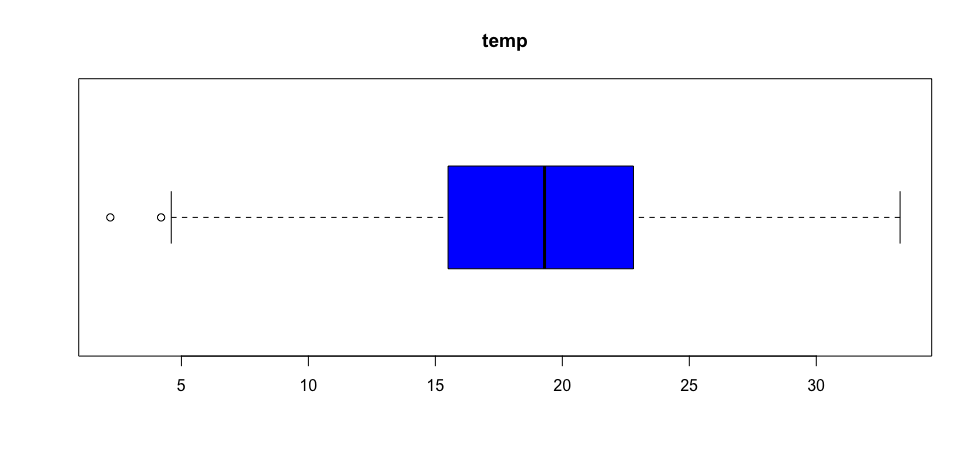
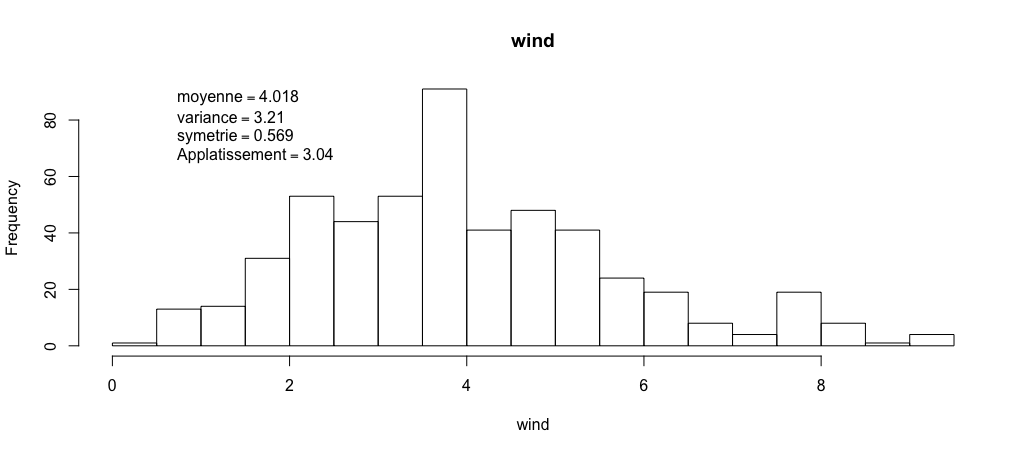
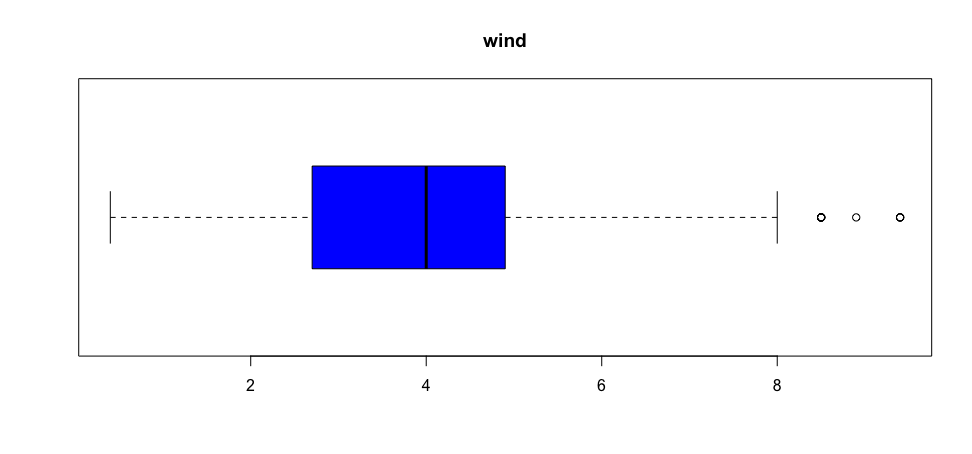


Figure 5 Histogramme et boite à moustache de la température

## 1.6) Indice du vent

L’indice du vent dénote lui aussi certain signe de distribution normale. Ceci semble dire qu’il existe une vitesse du vent qui favorise les feux forestiers. Cependant, encore une fois, une analyse visuelle un peu plus rigoureuse fait remarquée que la fin de la distribution est une peu curieuse pour une distribution normale.





## 1.7) Surface incendié

La surface incendiée n’est pas intéressante à visualiser à l’aide d’un histogramme. Cependant, on peut quand même utiliser la boite à moustache afin de bien comprendre que les feux qui attaquent la forêt de Montesinho sont généralement très petits. En effet, la majorité d’entre eux font en dessous de 100m carré [2].

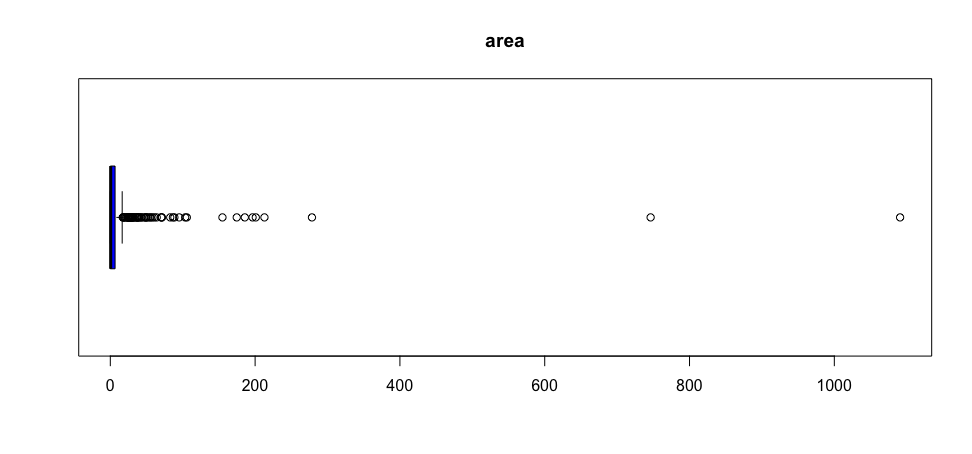


Figure 6 Boite à moustache représentant la surface incendiée

[4] W.W. Hines, D.C. Montgomery, D.M Goldsman, C. M. Borror, Probability and Statistic pour ingénieurs, 3ème éd. Montréal, Canada, Chanelière-Éducation, 2017.

[5] W.J. De Groot, Interpreting the Canadian Forest Fire Weather Iinde (FWI) System[EN LIGNE]. Disponible: http://www.dnr.state.mi.us/WWW/FMD/WEATHER/Reference/FWI\_Background.pdf