

情報可視化ソフトウェアHiddenによるデータ分析

[概要]

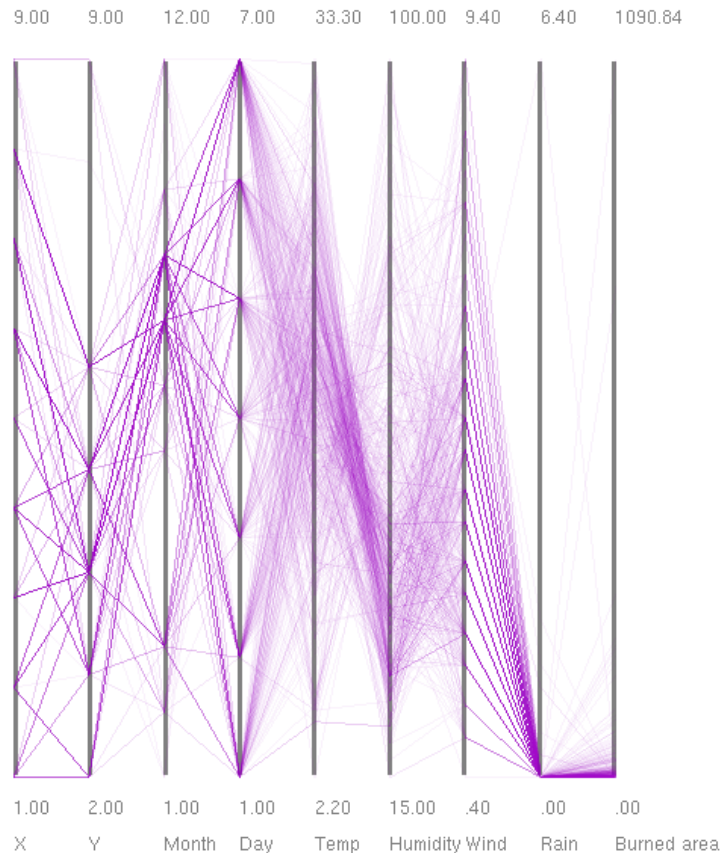
サンプルデータを参考にして自分だけのデータを作り、それを可視化して得られた可視化結果から得られる知識を解説。

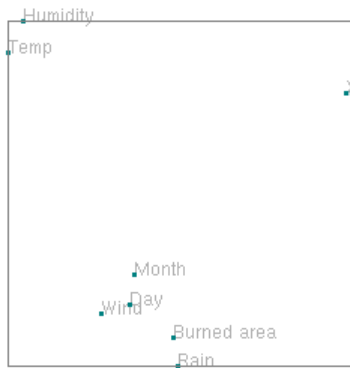
[使用データ]

UCI Machine Learning Repositoryから気になるデータを取得。[ポルトガル北東部の山火事のデータ](#)を変更を加えて[新しいデータ](#)を作成。横軸縦軸の説明は以下の通りである。(データの数は517個)

X	...	Montesinho公園マップ内のx軸空間座標：1から9
Y	...	Montesinho公園マップ内のx軸空間座標：2から9
Month	...	1月から12月：1から12
Day	...	月曜日から日曜日：1から7
Temp	...	摂氏度の温度：2.2から33.30
Humidity	...	相対湿度(%)：15.0~100
Wind	...	風速(km / h)：0.40~9.40
Rain	...	降水量(mm / m2)：0.0~6.4
Burned Area...	...	森林の焼失面積 (ha)：0.00~1090.84

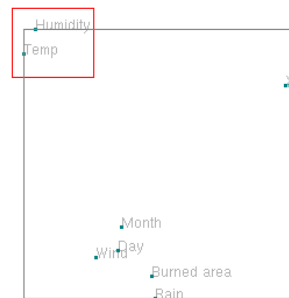
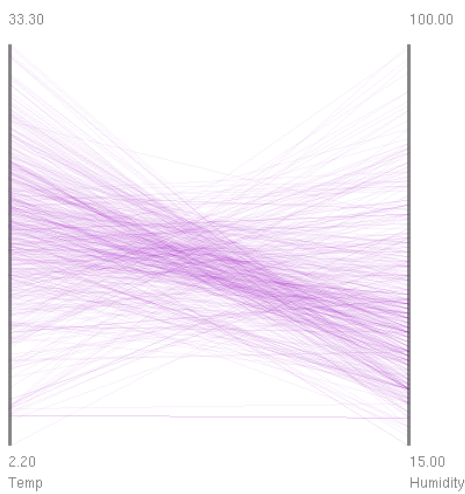
[Hiddenを使った全体データ]



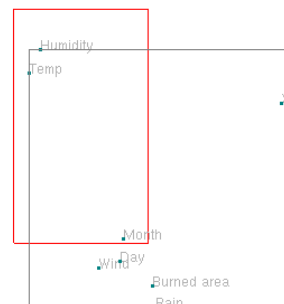
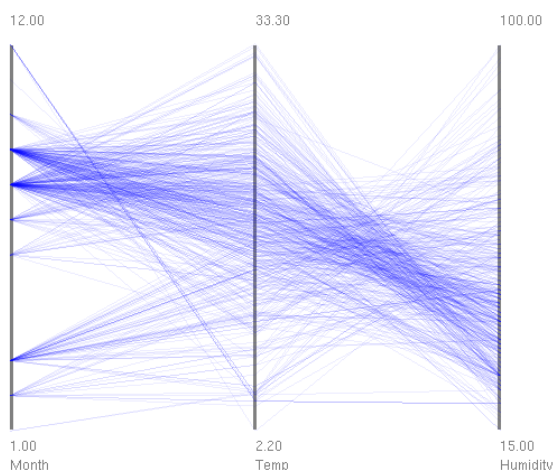


[可視化によりわかったこと]

上の散布図から、おおよそ4つのグループに分けることができる。
(1.Humidity&Temp)

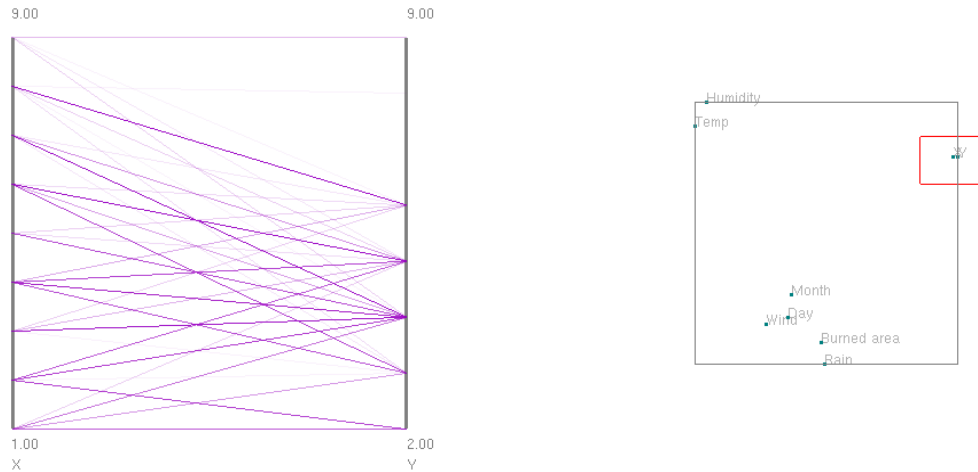


気温が高く湿度が低い、または、気温が低く湿度が高い時に山火事は起こっている。この中にも人災(放火等)も含まれてはいるが、気温が高く湿度が低い場合は乾燥による自然発火、気温が低く湿度が高い場合は嵐&雷による落雷が原因だとわかる。しかし、線の傾きや重なり具合の濃さなどから乾燥による自然発火の割合が多い。そこで気になった気温&湿度&月の3値のデータが以下である。



やはり8月9月あたり(ポルトガルは北半球なので夏)から出ている線の重なりが濃く現れている。

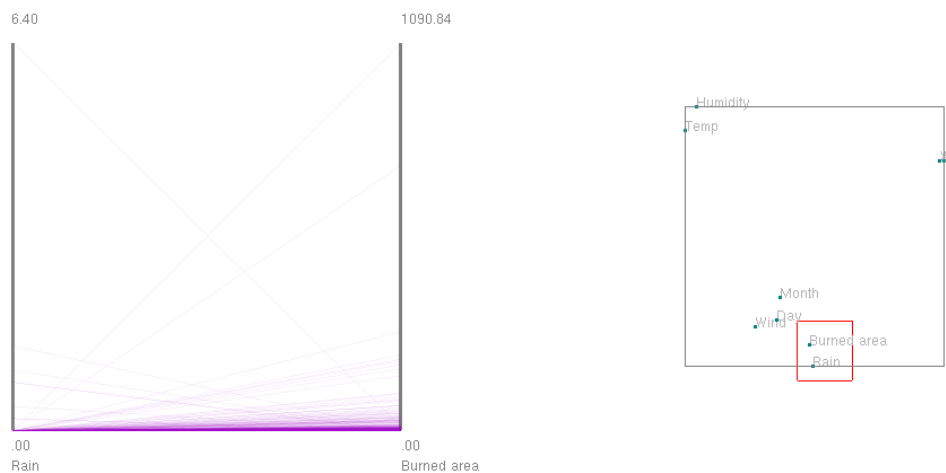
(2.X&Y)



UCI Machine Learning Repositoryが扱っているXY座標データはMontesinho公園マップ内の座標をX:1~9、Y:2~9で表したものとなる。x座標、y座標とも平均的に出ている線の数と同じであるため、Montesinho公園では全体的に偏ることはなく山火事が起こっていることがわかる。

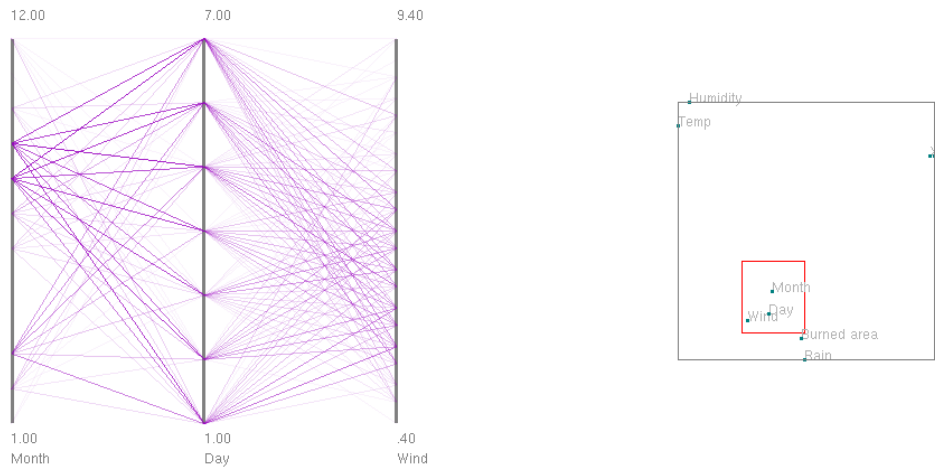


(3.Burned area&Rain)



降水量と森林の焼失面積の関係性は偏っており、互いに0に近い値を取っている。降水量が多いにも関わらず山火事が発生しているということは湿度が高いため落雷か放火が原因だとわかる。降水量が少ない時に山火事が起こっているのは乾燥が原因であることは明らかで、山火事の被害が小さい時が大半であるが1、2度ほど焼失面積が大きい大災害が起こっていることがわかる。

(4.Month&Day&Wind)



月別で出ている線の重なるの濃さと曜日別に集まっている線の重なるの濃さが大きく偏っていることもなく、また風速もあまり偏りが無い。つまり山火事には風があまり影響ない(?)だが、風速の値まで伸びている線が全体的に中心から上の方へ偏っているため、少なからず風の影響はあるのかもしれない。

[感想]

あらかじめ情報可視化ソフトウェアが用意されていたので、今回の課題は自分が気になるデータを拾ってきて入力、そして出力された結果をもとにレポートを書く簡単なものだと思っていたが意外にも自分が一番知りたい&興味があるデータを取ってくることに苦戦した。私の趣味は旅行&漫画を読むことだったのでそれらに関連するデータを取ってきたかったがインターネットに出回っているデータが全然無い。今回は時間がなく、スクレイピングをして目的のデータを取得することができなかったが大量のデータを持つことはとても大変なことであるのだなと感じた。