Visualization of Regional Variation in Forest Fire Data

Purpose - to answer the following questions:

Is there regional variation in the fires?

How best can this regional variation be modelled. Should we use factor variables for each area or use the X and Y grid values. Perhaps if the X and Y values are used we should tensor splines rather than simply assuming linearity?

Methods

A new graphical method is developed. We examine a two-way analysis of variance, treating X and Y as separate factors. We use *Mathematica* package ANOVA.

In[1]:=
```
Needs["ANOVA`"]
```

Basic data visualization using ListDensityPlot

In[2]:=
```
GetQuantiles[x_] := Module[{i1, i2, i3},
  i1 = If[# > Quantile[x, 0.25], 1, 0] & /@ x;
  i2 = If[# > Quantile[x, 0.5], 1, 0] & /@ x;
  i3 = If[# > Quantile[x, 0.75], 1, 0] & /@ x;
  {i1 + i2 + i3} // Flatten]
```

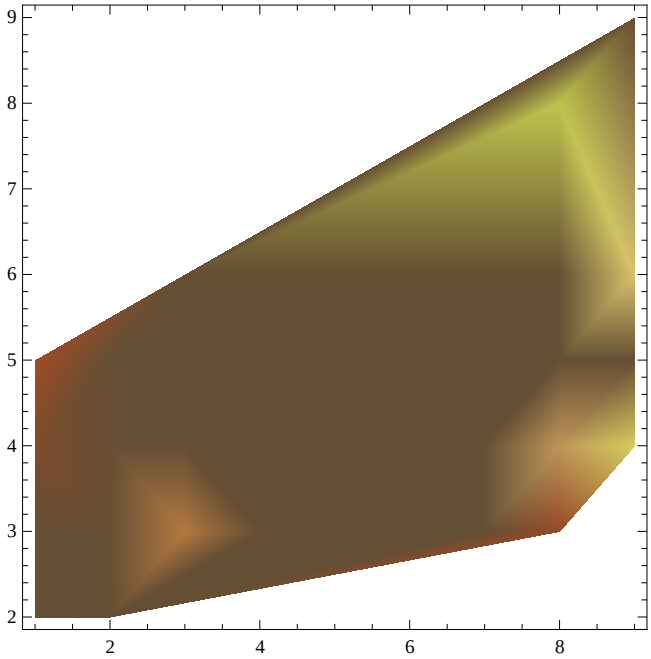
In[3]:=
```
GetQuantiles::usage =
  "GetQuantiles[x] returns 0, 1, 2, 3 according as the corresponding value
    in x is in the first, second, third or fourth quartile.";
```

In[4]:=
```
f = ReadList["d:/r/2009/fires/fires.dat", {Real, Real, Real}];
```

This is a pretty plot but not really helpful.

In[5]:=
```
ListDensityPlot[f, ColorFunction → "SouthwestColors", Mesh → Full]
```

Out[5]=



The ANOVA suggests that there is no interaction, so an additive model with X and Y may be suitable.

In[6]:=
```
ANOVA /. ANOVA[f, {x, y, All}, {x, y}, CellMeans → False]
```

Out[6]//TableForm=

	DF	SumOfSq	MeanSq	FRatio	PValue
x	8	28.3402	3.54253	1.89292	0.0590973
y	6	39.8564	6.64273	3.54949	0.00189229
x y	21	40.733	1.93967	1.03645	0.416704
Error	481	900.172	1.87146		
Total	516	1009.1			

Mathematica's ANOVA function doesn't work for 'xyarea' because some cells have only one observation. We use R.

```
> A<-factor(paste("A",X,Y,sep=""))
> outA<-lm(lburned~A)
> anova(outA)
Analysis of Variance Table

Response: lburned
           Df Sum Sq Mean Sq F value Pr(>F)
A          35    109       3    1.66  0.011 *
Residuals 481    900       2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

As expected the 'xyarea' is highly significant. Let's compare with simple linear model,

```
> outB<-lm(lburned~X+Y)
> anova(outB)
Analysis of Variance Table

Response: lburned
           Df Sum Sq Mean Sq F value Pr(>F)
X           1      4       4    1.98   0.16
Y           1  0.041   0.041    0.02   0.88
Residuals 514   1005       2
```

This doesn't work. Regression is not signficant. We could try polynomials in X and Y but splines are better!

```
> dfX <- dfY <- 7
> XS<-ns(X, df=dfX)
> YS<-ns(Y, df=dfY)
> XYS<-cbind(XS, YS)
> outXYS<-lm(lburned~XYS)
> anova(outXYS)
Analysis of Variance Table

Response: lburned
           Df Sum Sq Mean Sq F value  Pr(>F)
XYS        13     68       5    2.79 0.00074 ***
Residuals 503    941       2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

It seems that natural splines with 7 df will work work. This saves 21 df.

Notes

Standard deviation is redefined to be 0 when there is only 1 observation.

RGB color encoding of quartile information. Black (lowest, no light!) for observations less than or equal to the first quartile. Then Red (lowest frequency), Green and Blue (highest frequency) corresponding the next quartiles.

Improved Data Plots

Number of Observations in each Region -- Figure1.pdf

All points with same X, Y coordinates in separate lists

In[7]:=
```
out = Gather[f, Most[#1] == Most[#2] &];
```

Cells coordinates

In[8]:=
```
xy = Most /@ (First /@ out)
```

Out[8]=
```
{{7., 5.}, {7., 4.}, {8., 6.}, {6., 5.}, {5., 5.}, {8., 5.}, {6., 4.}, {5., 4.}, {6., 3.},
 {7., 3.}, {4., 4.}, {5., 6.}, {6., 6.}, {4., 3.}, {2., 2.}, {4., 5.}, {9., 9.}, {1., 3.},
 {1., 2.}, {3., 4.}, {2., 4.}, {3., 5.}, {2., 5.}, {4., 6.}, {3., 6.}, {1., 4.}, {8., 3.},
 {1., 5.}, {9., 4.}, {7., 6.}, {3., 3.}, {2., 3.}, {9., 6.}, {8., 4.}, {9., 5.}, {8., 8.}}
```

Number of observations in each cell

In[9]:=
```
NumObs = Length /@ out
```

Out[9]=
```
{11, 45, 52, 49, 3, 4, 9, 23, 25, 2, 36, 4, 3, 22, 25,
 25, 6, 10, 19, 43, 27, 7, 20, 8, 4, 15, 3, 4, 4, 2, 1, 1, 1, 1, 2, 1}
```

In[10]:=
```
f2 = MapThread[Append[#1, #2] &, {xy, NumObs}];
```

Table is interesting but hard to read. We redo to present this information graphically.

In[11]:=
```
TableForm[Partition[f2, 4], TableDepth → 2]
```

Out[11]//TableForm=

{7., 5., 11}	{7., 4., 45}	{8., 6., 52}	{6., 5., 49}
{5., 5., 3}	{8., 5., 4}	{6., 4., 9}	{5., 4., 23}
{6., 3., 25}	{7., 3., 2}	{4., 4., 36}	{5., 6., 4}
{6., 6., 3}	{4., 3., 22}	{2., 2., 25}	{4., 5., 25}
{9., 9., 6}	{1., 3., 10}	{1., 2., 19}	{3., 4., 43}
{2., 4., 27}	{3., 5., 7}	{2., 5., 20}	{4., 6., 8}
{3., 6., 4}	{1., 4., 15}	{8., 3., 3}	{1., 5., 4}
{9., 4., 4}	{7., 6., 2}	{3., 3., 1}	{2., 3., 1}
{9., 6., 1}	{8., 4., 1}	{9., 5., 2}	{8., 8., 1}

We will color-code according to quantile

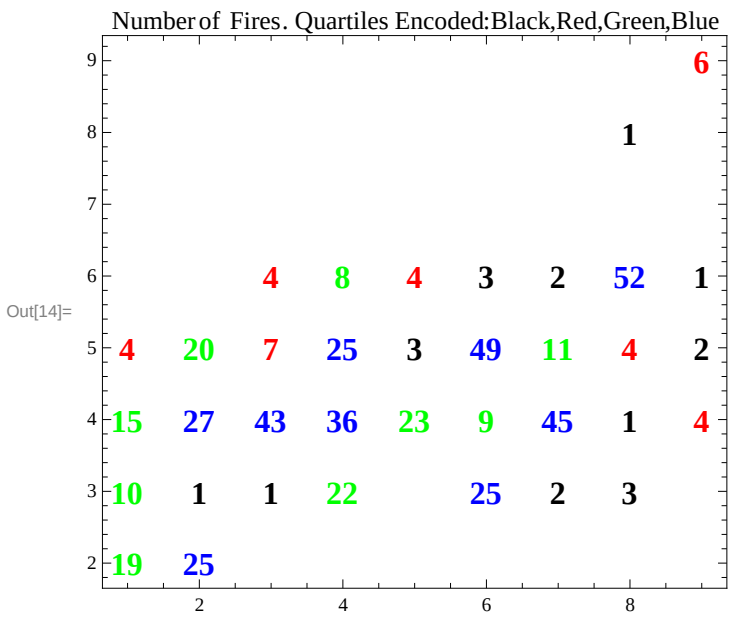
In[12]:=
```
q = GetQuantiles[NumObs]
```

Out[12]=
```
{2, 3, 3, 3, 0, 1, 2, 2, 3, 0, 3, 1, 0, 2, 3,
 3, 1, 2, 2, 3, 3, 1, 2, 2, 1, 2, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0}
```

In[13]:=
```
f2 = Append[Transpose[f2], q] // Transpose;
```

In[14]:=
```
g2 = Graphics[Text[Style[#[[3]], FontSize → 15, FontWeight → "Bold", FontColor → Which[
          #[[4]] == 0, RGBColor[0, 0, 0],
          #[[4]] == 1, RGBColor[1, 0, 0],
          #[[4]] == 2, RGBColor[0, 1, 0],
          #[[4]] == 3, RGBColor[0, 0, 1]
         ]
        ],
        {#[[1]], #[[2]]}] & /@ f2, Frame → True, PlotRangePadding → 0.35,
  PlotLabel → "Number of Fires. Quartiles Encoded:Black,Red,Green,Blue"]
```

Out[14]=



In[16]:=
```
Export["d:/r/2009/Fires/Figure1NumFires.pdf", g2];
```

In[17]:=
```
ANOVA /. ANOVA[Most /@ f2, {x, y}, {x, y}, CellMeans → False]
```

Out[17]//TableForm=

	DF	SumOfSq	MeanSq	FRatio	PValue
x	8	1145.51	143.188	0.551178	0.804883
y	6	1427.3	237.883	0.91569	0.503136
Error	21	5455.5	259.786		
Total	35	8028.31			

Number of fires does not depend in a simple way on X and Y.

■ Oneway ANOVA for Regions

Mathematica fails when there is only 1 observation, so we need to eliminate all regions with only 1 factor

■ Median in each Region. Figure 2

In[18]:=
```
meds = Median /@ Map[Last, out, {2}]
```

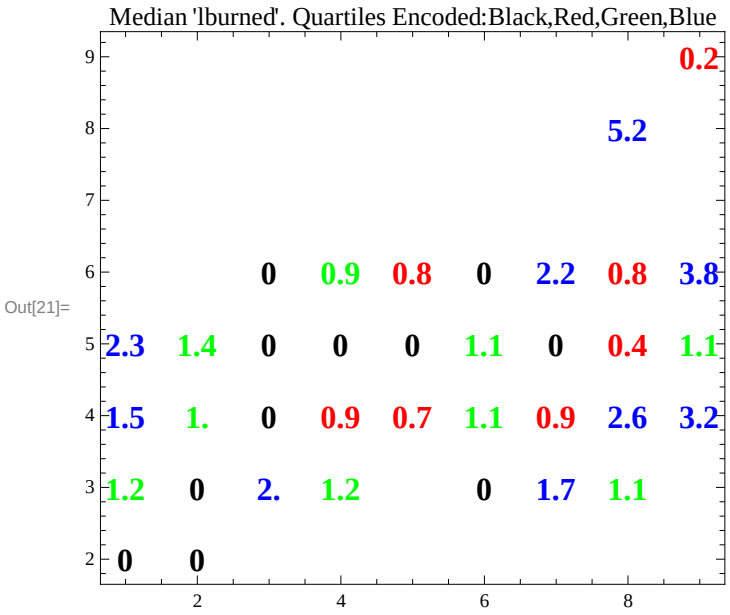
Out[18]=
```
{0., 0.8671, 0.776709, 1.14422, 0., 0.366953, 1.06471, 0.667829,
 0., 1.67231, 0.881577, 0.766278, 0., 1.19194, 0., 0., 0.153742, 1.19381,
 0., 0., 0.970779, 0., 1.42062, 0.926584, 0., 1.50408, 1.07841, 2.32219,
 3.2322, 2.23553, 2.02551, 0., 3.78123, 2.5787, 1.10742, 5.22982}
```

In[19]:=
```
f3 = MapThread[Append[#1, #2] &, {xy, meds}];
```

In[20]:=
```
f3 = Append[Transpose[f3], GetQuantiles[meds]] // Transpose;
```

In[21]:=
```
g3 = Graphics[
  Text[Style[Round[#[[3]], .1], FontSize → 15, FontWeight → "Bold", FontColor → Which[
        #[[4]] == 0, RGBColor[0, 0, 0],
        #[[4]] == 1, RGBColor[1, 0, 0],
        #[[4]] == 2, RGBColor[0, 1, 0],
        #[[4]] == 3, RGBColor[0, 0, 1]
       ]
    ],
    {#[[1]], #[[2]]} & /@ f3, Frame → True, PlotRangePadding → 0.35,
  PlotLabel → "Median 'lburned'. Quartiles Encoded:Black,Red,Green,Blue"]
```

Out[21]=



In[22]:=
```
Export["d:/r/2009/Fires/Figure2Median.pdf", g2];
```

In[23]:=
```
ANOVA /. ANOVA[Most /@ f3, {x, y}, {x, y}, CellMeans → False]
```

Out[23]//TableForm=

	DF	SumOfSq	MeanSq	FRatio	PValue
x	8	13.2531	1.65664	2.30126	0.0602522
y	6	22.1498	3.69163	5.12809	0.0021953
Error	21	15.1176	0.719884		
Total	35	50.5204			

Median 'lburned' depends on X and Y. Low X, Y tends to have low 'lburned'

■ Mean in each Region. Figure 3

In[24]:=
```
mens = Mean /@ Map[Last, out, {2}]
```

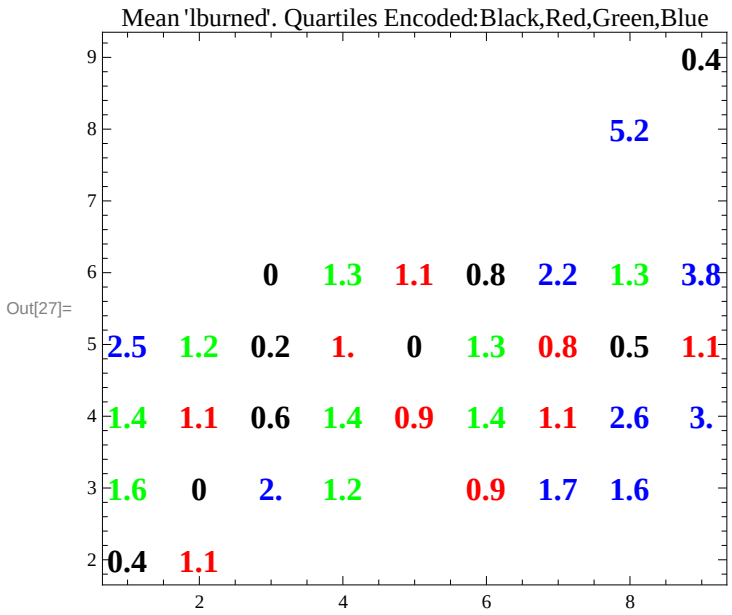
Out[24]=
```
{0.844601, 1.0631, 1.27383, 1.29981, 0., 0.458961, 1.44663, 0.853836, 0.918182,
 1.67231, 1.40058, 1.05999, 0.753574, 1.24744, 1.10722, 1.0287, 0.41465, 1.61779,
 0.434583, 0.632903, 1.10566, 0.165269, 1.23837, 1.33236, 0., 1.38342, 1.64878,
 2.52769, 3.02526, 2.23553, 2.02551, 0., 3.78123, 2.5787, 1.10742, 5.22982}
```

In[25]:=
```
f4 = MapThread[Append[#1, #2] &, {xy, mens}];
```

In[26]:=
```
f4 = Append[Transpose[f4], GetQuantiles[mens]] // Transpose;
```

In[27]:=
```
g4 = Graphics[
  Text[Style[Round[#[[3]], .1], FontSize → 15, FontWeight → "Bold", FontColor → Which[
        #[[4]] == 0, RGBColor[0, 0, 0],
        #[[4]] == 1, RGBColor[1, 0, 0],
        #[[4]] == 2, RGBColor[0, 1, 0],
        #[[4]] == 3, RGBColor[0, 0, 1]
       ]
    ],
    {#[[1]], #[[2]]} & /@ f4, Frame → True, PlotRangePadding → 0.35,
  PlotLabel → "Mean 'lburned'. Quartiles Encoded:Black,Red,Green,Blue"]
```

Out[27]=



In[28]:=
```
Export["d:/r/2009/Fires/Figure3Mean.pdf", g3];
```

In[29]:=
```
ANOVA /. ANOVA[Most /@ f4, {x, y}, {x, y}, CellMeans → False]
```

Out[29]//TableForm=

	DF	SumOfSq	MeanSq	FRatio	PValue
x	8	10.6159	1.32698	2.3289	0.0576934
y	6	17.0159	2.83599	4.97726	0.00258727
Error	21	11.9656	0.56979		
Total	35	39.5974			

Mean 'lburned' depends on X and Y. Low X, Y tends to have low 'lburned'

■ Sd in each Region. Figure 4

Note: we redefine sd so it is 0 if there is only 1 observation!

In[30]:=
```
SD[x_] := If[Length[x] < 2, 0, StandardDeviation[x]];
```

In[31]:=
```
sds = SD /@ Map[Last, out, {2}]
```

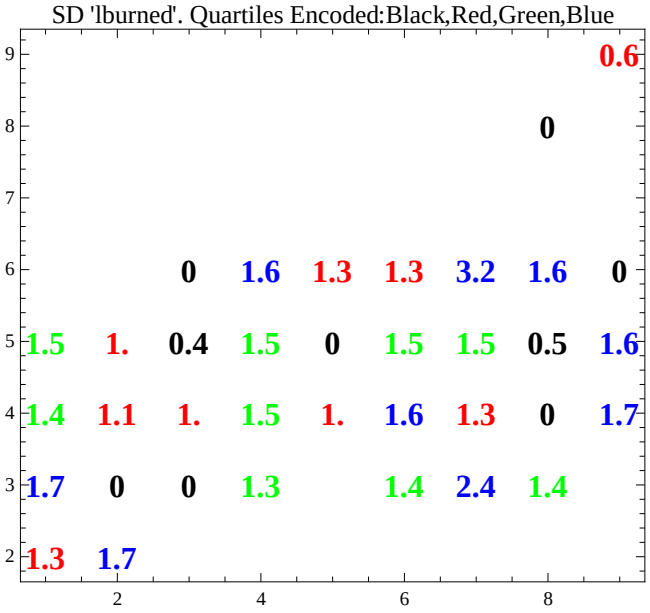
Out[31]=
```
{1.48435, 1.31657, 1.55245, 1.45974, 0., 0.478578, 1.64559, 1.00272,
 1.42298, 2.36501, 1.51111, 1.31458, 1.30523, 1.33014, 1.65591, 1.50879,
 0.556287, 1.70267, 1.27701, 1.0273, 1.1278, 0.43726, 1.03824, 1.57612,
 0., 1.36905, 1.35432, 1.52466, 1.74819, 3.16152, 0, 0, 0, 0, 1.56613, 0}
```

In[32]:=
```
f5 = MapThread[Append[#1, #2] &, {xy, sds}];
```

In[33]:=
```
f5 = Append[Transpose[f5], GetQuantiles[sds]] // Transpose;
```

In[34]:=
```
g5 = Graphics[
  Text[Style[Round[#[[3]], .1], FontSize → 15, FontWeight → "Bold", FontColor → Which[
        #[[4]] == 0, RGBColor[0, 0, 0],
        #[[4]] == 1, RGBColor[1, 0, 0],
        #[[4]] == 2, RGBColor[0, 1, 0],
        #[[4]] == 3, RGBColor[0, 0, 1]
       ]
     ],
    {#[[1]], #[[2]]}] & /@ f5, Frame → True, PlotRangePadding → 0.35,
  PlotLabel → "SD 'lburned'. Quartiles Encoded:Black,Red,Green,Blue"]
```

Out[34]=



In[35]:=
```
Export["d:/r/2009/Fires/Figure4SD.pdf", g4];
```

In[36]:=
```
ANOVA /. ANOVA[Most /@ f5, {x, y}, {x, y}, CellMeans → False]
```

Out[36]//TableForm=

	DF	SumOfSq	MeanSq	FRatio	PValue
x	8	8.97769	1.12221	2.75698	0.0298157
y	6	1.1866	0.197766	0.48586	0.811427
Error	21	8.54791	0.407043		
Total	35	18.7122			

SD 'lburned' depends on X but not on Y.