# Tree survival modeling

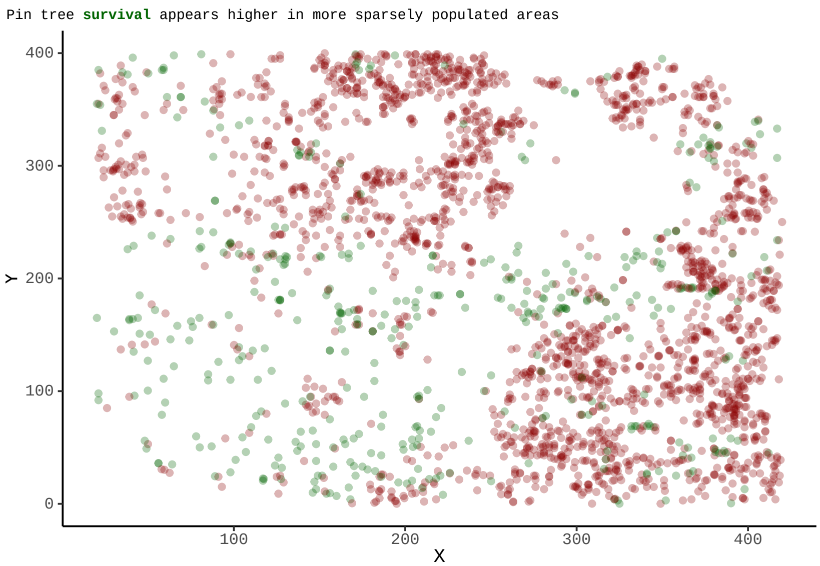
Erin Franke

May 2022

## Exploratory Data Analysis

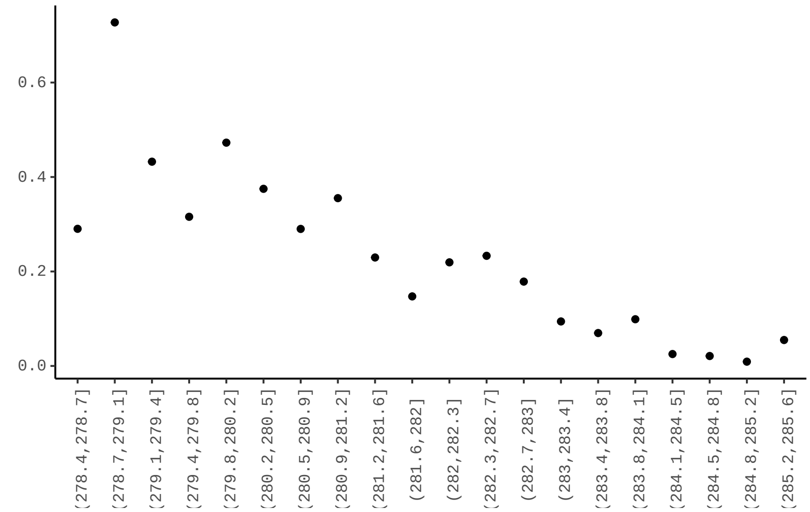
### Pin trees

It also appears that in more densely populated areas of pin trees, more trees die while in more sparsely populated regions more trees survive.



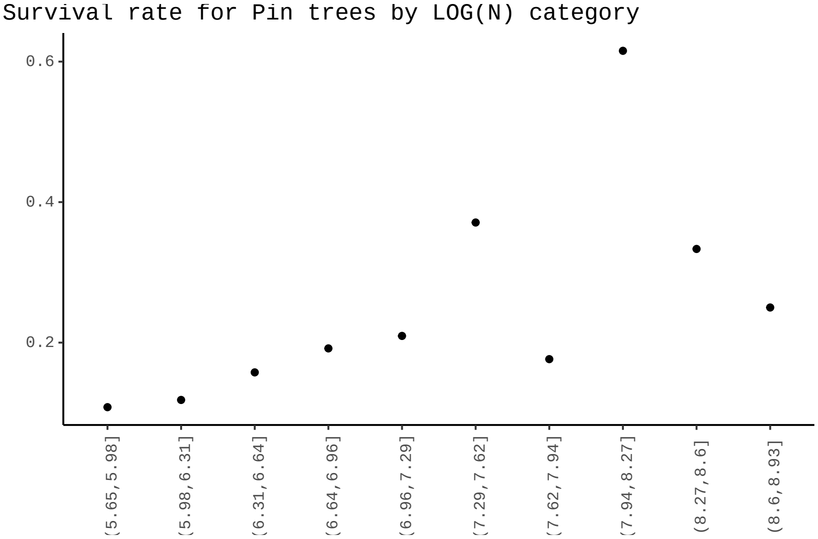

There is a pretty strong negative relationship between elevation and survival for pin trees.

Survival rate for Pin trees by elevation category

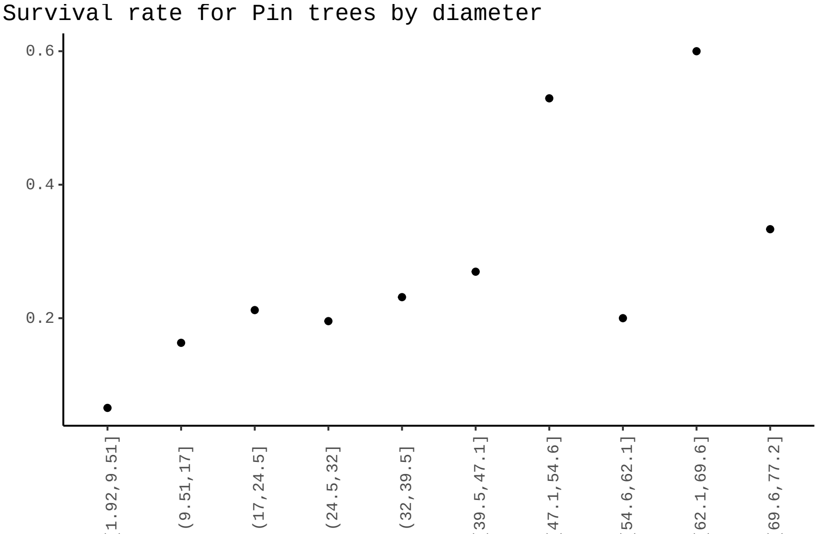



There is a positive moderate relationship between LOG(N) and pin tree survival, as well as diameter and pin tree survival.

Survival rate for Pin trees by LOG(N) category



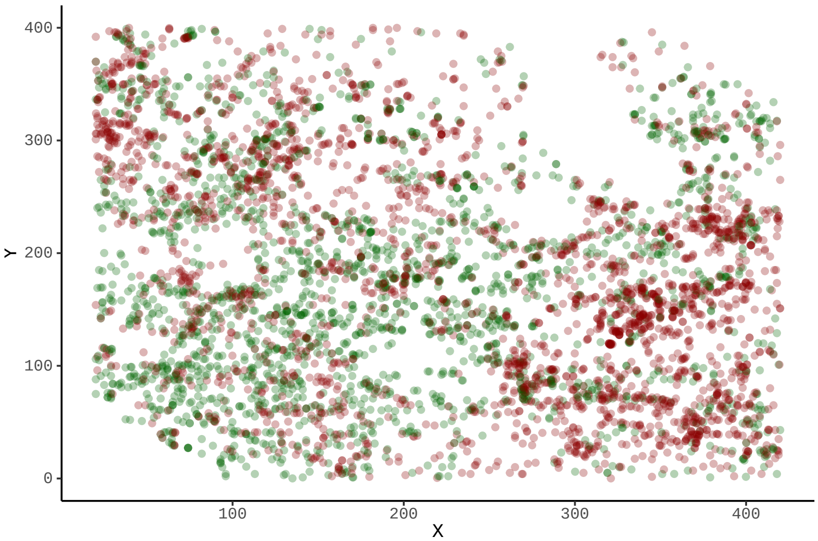

Survival rate for Pin trees by diameter



## Bur trees

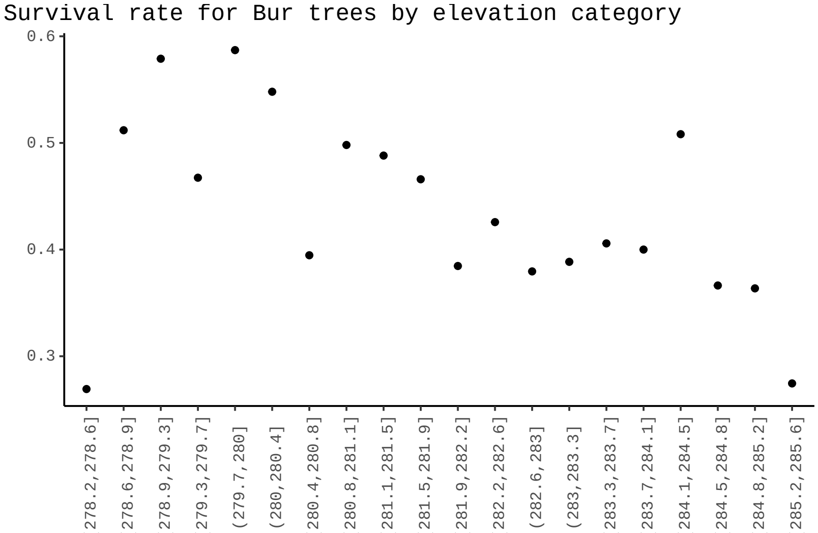
In contrast to pin trees, we do not see as strong of a pattern related to tree density and survival for bur trees.

Bur tree **survival** appears less dependent on tree density than pin tree survival

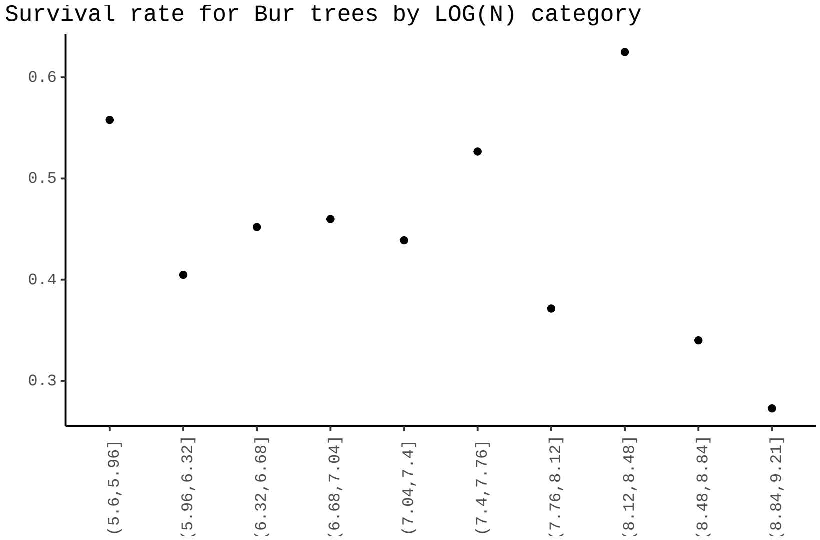


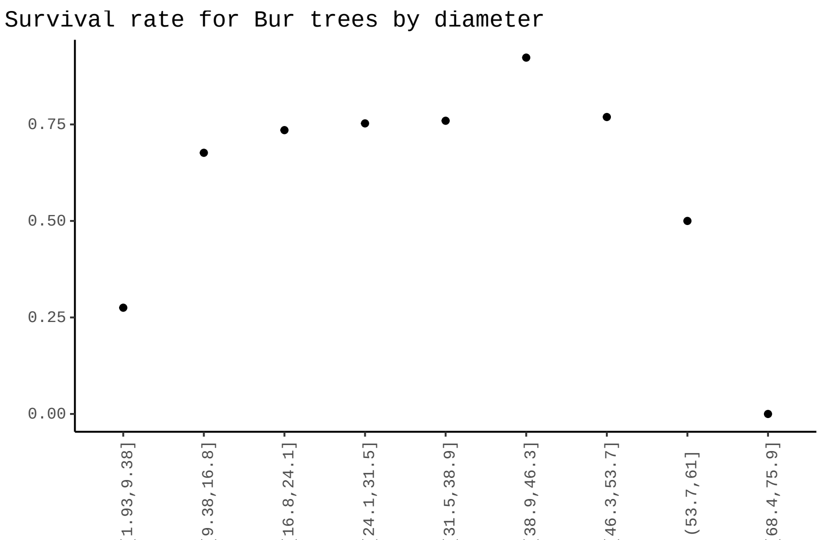

There is also a negative relationship between elevation and survival for bur trees, though it is not quite as strong as that for pin trees.

Survival rate for Bur trees by elevation category




There does not appear to be a distinct relationship between LOG(N) and bur tree survival. The relationship between diameter and bur tree survival looks somewhat quadratic, with survival being highest at medium diameters and much lower for skinnier or very large trees.




# Modeling

## Pin trees

Due to computational issues, we had to make this data areal and divided the data into 81 different regions of about equal size. Each region has between 1 and 106 trees, obviously depending on how densely the trees are populated in that region. We then fit a model that accounts for the spatial correlation of the trees using a GEE model. GEE provides robust SE's that are valid even if we specify the wrong correlation structure (which summarizes the correlation between trees in the same region). This means our standard errors will be valid and allow us to create valid confidence intervals and properly conduct hypothesis tests.

In the pin model, elevation and diameter are both significant coefficients on the $\alpha = 0.001$ level. LOG(N) is insignificant with a p-value of 0.88. The elevation coefficient tells us that holding diameter and LOG(N) constant, we expect the odds of survival to decrease by a multiplicative factor of exp(-0.52) = 0.6 for every one unit increase in pin tree elevation. Additionally, for every one unit increase in pin tree diameter we expect the odds of pin tree survival to increase by a multiplicative factor of exp(0.04) = 1.04 holding elevation and LOG(N) constant.

```r
pin_survival$region <- paste0('X',cut(pin_survival$X,breaks=seq(20,420,length = 10)),'Y',cut(pin_surviva

pinmod <- pin_survival %>%
  arrange(region) %>%
geeM::geem(survived ~ ELEVA + DBH95 + `LOG(N)`, family = binomial(),id = region, data = .,corstr='indepe

summary(pinmod)
```

```
##             Estimates  Model SE Robust SE    wald      p
## (Intercept) 145.80000 12.510000 24.200000  6.0250 0.0000
## ELEVA        -0.52420  0.042670  0.084110 -6.2320 0.0000
## DBH95         0.04034  0.004273  0.005608  7.1930 0.0000
## `LOG(N)`     -0.03925  0.141300  0.269400 -0.1457 0.8842
##
##  Estimated Correlation Parameter:  0
##  Correlation Structure:  independence
##  Est. Scale Parameter:  0.9852
##
##  Number of GEE iterations: 2
##  Number of Clusters:  81    Maximum Cluster Size:  106
##  Number of observations with nonzero weight:  2538
```

### Bur trees

The same type of model can be used to model bur survival.

For bur trees, elevation and diameter are both significant coefficients on the $\alpha = 0.01$ level. LOG(N) is insignificant with a p-value of 0.22. The elevation coefficient tells us that holding diameter and LOG(N) constant, we expect the odds of survival to decrease by a multiplicative factor of exp(-0.186) = 0.8 for every one unit increase in bur tree elevation. Additionally, for every one unit increase in bur tree diameter we expect the odds of pin tree survival to increase by a multiplicative factor of exp(0.15) = 1.16 holding elevation and LOG(N) constant.

```
##             Estimates Model SE Robust SE   wald        p
## (Intercept)   52.2600 17.32000  19.94000  2.621 0.008767
## ELEVA         -0.1864  0.05913   0.06931 -2.689 0.007165
## DBH95          0.1501  0.01552   0.01355 11.070 0.000000
## `LOG(N)`      -0.2193  0.17830   0.17960 -1.221 0.222200
##
##  Estimated Correlation Parameter:  0
##  Correlation Structure:  independence
##  Est. Scale Parameter:  5.948
##
##  Number of GEE iterations: 2
##  Number of Clusters:  36    Maximum Cluster Size:  244
##  Number of observations with nonzero weight:  5000
```