Data Analysis #2

Dustin Leatherman

5/24/2019

Introduction

It is common practice for industries that harvest natural resources, such as logging and mining, to replace trees after they have been removed. The impact of climate change on adolescent trees is not well understood. South Western Australia provides a unique test bed for assessing climate change impact since it is one of the few places that have already felt significant climate change. Mean rainfall has experienced a 17% reduction between 1975-2011 compared to the mean rainfall between 1900-1974. In parallel, the average daily highs and lows have been increasing which exacerbates droughts (Rachel J. Standish 2015; Bates, B.C 2008).

Jarrah trees (*Eucalyptus marginata*) are canopy trees in South Western Australia that thrive in the 650-1250mm rainfall zone (*Jarrah*, n.d.). Jarrah are resistant to termites and pests but particularly susceptible to Dieback or root rot. Dieback is caused by bacteria that travels in water and is absorbed through a root system (*Phytophthora Root Rot - Fact Sheet*, n.d.). One method in identifying wet and dry season is by using an Ombrothermic diagram. It summarizes trends in temperature and precipitation to allow for establishing a relationship between the two. The length of the wet season is determined by ombrothermic relationship where average rainfall exceeded average high temperatures (*Geography 11th Std* 2019). The dataset contains aggregate data so rebuilding ombrothermic diagrams for each year is not possible but it is still a valuable tool in understanding how the data was calculated.

Questions of Interest

- What is the effect of temperature, rain amount, rain consistency, and diebeck on the species richness of plot-level vegetation?
- Is the length of a wet season a factor in whether soil is afflicted with Diebeck?

Sample Collection

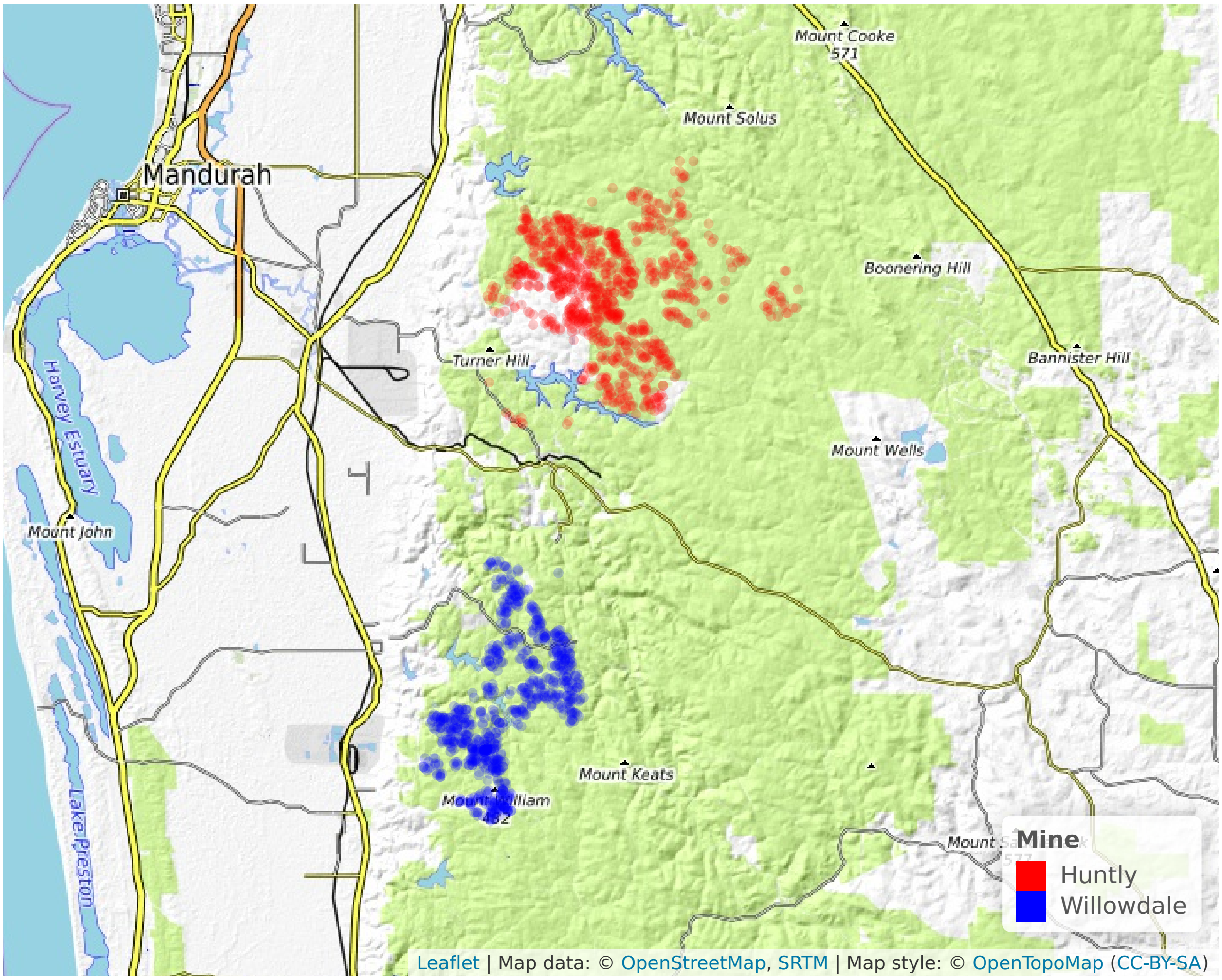
Seedlings were established in 1938 plots total between the restoration sites of the Huntly and Willowdale mines. The seedlings were monitored 15 months after planting as well as after the onset of the first wet season from 1992-2010. Data were recorded as species frequencies ranging from 0-100 based on their presence within five 4 x 4m quadrats nested within 20 x 20m plots. The data describes the species composition of the restored Jarrah fores, including 491 plant species in total (Rachel J. Standish 2015).

The Climate variables were measured from climate stations at the Huntly and Willowdale mining sites while additional air temperature statistics were taken from Dwellingup, the meteorological station nearest to both sites. Not captured in this particular dataset is historical data the researchers used to determine the surety of the effect of climate change in the area. Mediterranean climates are defined by long summer droughts when rainfall is less than twice the mean temperature. This relationship was used to define the wet season for each year (Rachel J. Standish 2015).

73.5% of the plots recorded a diebeck status so those will be the only ones used in analysis. This is a potential gap for accurately predicting diebeck.

```
climate <- read.csv("~/Downloads/Standish et al. JEcol-2014-0252R1.csv")
# friendly name for legends
climate$MineFull <- ifelse(climate$Mine == "HU", "Huntly", "Willowdale")

# x,y are required for project() to convert UTM coords to lat-long
projDf <- data.frame(x = climate$Easting, y = climate$Northing)

# UTM Coordinates for south western Australia where the study took place
proj4Config <- "+proj=utm +zone=50 +south +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs"
pj <- project(projDf, proj4Config, inverse = TRUE)
climate$lat <- pj$y
climate$lon <- pj$x

# Make a pretty Topographical map. colored by mines
pal <- colorFactor(c("red", "blue"), domain = c("Huntly","Willowdale"))
leaflet(data = climate, options = leafletOptions(zoomControl = FALSE)) %>%
  addProviderTiles(providers$OpenTopoMap) %>%
  addCircleMarkers(
    radius = 3,
    color = ~pal(MineFull),
    stroke = FALSE
  ) %>%
  addLegend("bottomright", pal = pal, values = ~MineFull,
            title = "Mine",
            opacity = 1
  )
```

```
## Assuming "lon" and "lat" are longitude and latitude, respectively
```


Mandurah
Mount Cooke
571
Mount Solus
Boonering Hill
Turner Hill
Bannister Hill
Mount Wells
Harvey Estuary
Mount John
Mount Keats
Lake Preston
Mine
Huntly
Willowdale
Leaflet | Map data: © OpenStreetMap, SRTM | Map style: © OpenTopoMap (CC-BY-SA)

Summary Statistics

```
climateCalcs <-
  climate %>%
    select_if(is.numeric) %>%
    select(-Easting, -Northing, -Elevation, -Slope, -Year) %>%
    summarize_all(
      funs(
        Min=min,
        Q25 = quantile(., 0.25, na.rm = TRUE),
        Median=median,
        Q75 = quantile(., 0.75, na.rm = TRUE),
        Max=max,
        Mean=mean,
        Stdev = sd
        )
      ) %>%
    gather(stat, val) %>%
    separate(stat, into = c("Variable", "stat"), sep = "_") %>%
    spread(stat, val) %>%
    column_to_rownames("Variable") %>%
    select(Min, Q25, Median, Q75, Max, Mean, Stdev)

climateCalcs %>%
  kable(
    digits = 4,
    caption = "Jarrah Restoration for 1992-2010 (n = 1938). Bolded rows indicate variables used in subse
    row.names = TRUE
  ) %>%
    kable_styling(full_width = T, bootstrap_options = "striped", latex_options = "hold_position") %>%
    row_spec(which(rownames(climateCalcs) %in% c("Rain30", "Rain60", "Rain90", "Rain120", "RainWS", "Ra
```

```
# Averages and Extreme
plot.temps <-
  climate %>%
  select(Year, `Average Low` = AvgMinTemp, `Average High` = AvgMaxTemp, `High` = HighestTemp, `Low` = L
  gather(c("Average High", "High", "Low", "Average Low"), key = "Temperature", value = "value") %>%
  mutate(variable = factor(Temperature, level = c("High", "Average High", "Average Low", "Low"))) %>%
  ggplot(aes(x = Year, color = Temperature, y = value)) +
    geom_line() +
    xlab("Year") + ylab("Temperature (C\u00B0)") +
    labs(
      title = "Estimated Annual Temperature",
      caption = "Homogenous results across both sites"
    )

climate.groups <-
  climate %>%
  select(Year, Mine=MineFull, `30 Days` = Rain30, `60 Days` = Rain60, `90 Days` = Rain90, `120 Days` = 
  distinct %>%
  gather(c(-Year,-Mine), key = "variable", value = "value") %>%
  # reorder factors for presentation purposes
```

Table 1: Jarrah Restoration for 1992-2010 (n = 1938). Bolded rows indicate variables used in subsequent analysis

	Min	Q25	Median	Q75	Max	Mean	Stdev
AvgMaxTemp	**21.4000**	**22.3000**	**22.5000**	**22.8000**	**23.6000**	**22.4844**	**0.5117**
AvgMinTemp	**9.3000**	**9.7000**	**9.9000**	**10.2000**	**10.5000**	**9.9324**	**0.3364**
DaysOver35	4.0000	7.0000	13.0000	15.0000	25.0000	12.4520	5.4887
Drought.Length	111.0000	137.0000	156.0000	173.0000	231.0000	158.2828	28.0015
False.Start	0.0000	0.0000	0.0000	0.0000	1.0000	0.2472	0.4315
FertK	12.5000	25.0000	50.0000	50.0000	50.0000	40.4541	12.9773
FertN	20.0000	40.0000	80.0000	80.0000	80.0000	64.7265	20.7637
FertP	20.0000	40.0000	80.0000	80.0000	80.0000	64.7265	20.7637
HighestTemp	**37.1000**	**38.5000**	**39.2000**	**40.0000**	**42.0000**	**39.3540**	**1.2394**
lat	-32.9586	-32.8800	-32.6514	-32.5874	-32.5073	-32.7238	0.1475
lon	115.9651	116.0292	116.0760	116.1162	116.2695	116.0772	0.0576
LowestTemp	**-3.0000**	**-2.0000**	**-1.0000**	**-0.7000**	**1.0000**	**-1.0699**	**0.9612**
Prop.Sp.shared	0.4460	0.6558	0.6872	0.7180	0.9233	0.6874	0.0486
Rain.Length	**133.0000**	**184.0000**	**211.0000**	**229.0000**	**281.0000**	**207.2817**	**35.6795**
Rain120	**442.3000**	**758.4000**	**817.4000**	**921.0000**	**1099.9000**	**810.5330**	**141.7416**
Rain15mth	**1484.0000**	**1975.0000**	**2253.1000**	**2330.0400**	**2544.1000**	**2150.3100**	**249.9018**
Rain30	**55.6000**	**113.5000**	**163.1000**	**227.7000**	**410.5000**	**186.2630**	**88.7971**
Rain60	**62.0000**	**310.4000**	**365.1000**	**477.7000**	**877.3000**	**392.8057**	**155.0203**
Rain90	**320.7000**	**554.3000**	**615.1000**	**680.6000**	**895.2000**	**617.2696**	**124.5693**
RainEven120	**0.5996**	**0.7195**	**0.7477**	**0.7724**	**0.8153**	**0.7384**	**0.0486**
RainEven15mth	**0.4591**	**0.6186**	**0.6971**	**0.7298**	**0.7793**	**0.6710**	**0.0733**
RainEven30	**0.2490**	**0.4521**	**0.5220**	**0.6213**	**0.7763**	**0.5324**	**0.1330**
RainEven60	**0.3454**	**0.5493**	**0.6375**	**0.7228**	**0.8634**	**0.6239**	**0.1223**
RainEven90	**0.4915**	**0.6521**	**0.7159**	**0.7564**	**0.7970**	**0.7013**	**0.0694**
RainEvenWS	**0.6474**	**0.7419**	**0.7576**	**0.7735**	**0.8017**	**0.7550**	**0.0303**
RainPrevWS	647.8000	1005.8000	1139.7000	1210.3000	1420.3000	1093.3995	184.4803
RainWS	**464.7000**	**990.4000**	**1136.5000**	**1192.6000**	**1420.3000**	**1065.9174**	**216.6387**
SeedMix.SR	30.0000	57.0000	66.0000	75.0000	106.0000	66.3297	14.6522
SeedMix.tt	1.0000	10.0000	20.0000	29.0000	38.0000	19.3437	10.6845
SeedWetstart.syn	NA	9.0000	NA	76.0000	NA	NA	NA
Similarity	0.6581	4.1034	5.5921	7.3303	23.2376	6.0426	2.8216
StripSpread.delay	NA	30.0000	NA	480.0000	NA	NA	NA
Topsoil.rateALL	NA	1.0000	NA	7.0000	NA	NA	NA
Veg.SR	**15.0000**	**42.0000**	**49.0000**	**59.0000**	**99.0000**	**50.6259**	**12.3086**

```
  mutate(variable = factor(variable, levels = c("30 Days", "60 Days", "90 Days", "120 Days", "15 Months

climate.groups.avg <-
  climate.groups %>%
  group_by(Mine, variable) %>%
  summarize_at("value", funs(avg=mean))

# Line charts for Rainfall
plot.scatter.rainfall <-
  climate.groups %>% inner_join(climate.groups.avg, by = c("Mine","variable")) %>%
  ggplot(aes(x = Year, y = value, color = Mine)) +
    geom_line() +
    geom_line(aes(y = avg), linetype = "dashed") +
```

```
    facet_wrap(~ variable , ncol = 3) +
    xlab("Year") + ylab("Rainfall (mm)") +
    labs(
      title = "Rainfall Throughout the Wet Season",
      caption = "Dashed lines indicate average"
    )

climate.raineven <-
  climate %>%
  select(Year, Mine=MineFull, `30 Days` = RainEven30, `60 Days` = RainEven60, `90 Days` = RainEven90, `
  gather(c(-Year,-Mine), key = "variable", value = "value") %>%
  # reorder factors for presentation purposes
  mutate(variable = factor(variable, levels = c("30 Days", "60 Days", "90 Days", "120 Days", "15 Months

climate.raineven.avg <-
  climate.raineven %>%
  group_by(Mine, variable) %>%
  summarize_at("value", funs(avg=mean))

# Line charts that show Rain Evenness
plot.scatter.raineveness <-
  climate.raineven %>% inner_join(climate.raineven.avg, by = c("Mine","variable")) %>%
  ggplot(aes(x = Year, y = value, color = Mine)) +
    geom_line() +
    geom_line(aes(y = avg), linetype = "dashed") +
    facet_wrap(~ variable , ncol = 3) +
    xlab("Year") + ylab("Shannon Diversity Index") +
    labs(
      title = "Rainfall Evenness Throughout the Wet Season",
      caption = "0 = complete unevenness. 1 = complete evenness. Dashed lines indicate average"
    )

climate.wet.season <-
  climate %>%
    select(Rain.Length, Year, Mine=MineFull) %>%
    group_by(Mine, Year) %>%
    distinct

climate.wet.season.mine <-
  climate.wet.season %>%
    group_by(Mine) %>%
    summarise_at("Rain.Length", funs(avg = mean))

plot.wetseason <-
  climate.wet.season %>%
  inner_join(climate.wet.season.mine, by = "Mine") %>%
  ggplot(aes(x = Year, y = Rain.Length, color = Mine)) +
    geom_line(show.legend = FALSE) +
    geom_line(aes(y = avg), linetype = "dashed", show.legend = FALSE) +
    xlab("Year") + ylab("Days") +
    labs(
      title = "Length of Wet Season",
      caption = "Dashed lines indicate average"
```

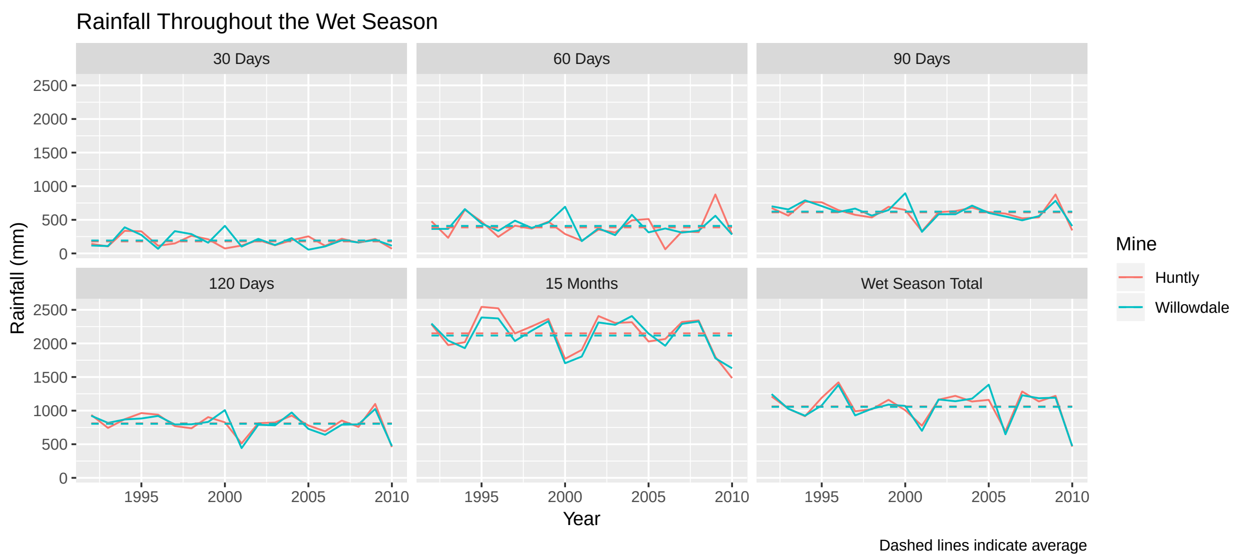
```r
    )

diebeck <-
  climate  %>%
  drop_na(Dieback) %>%
  mutate(isDieback = ifelse(Dieback == "DBF", 1, 0))

diebeck.by.year <-
  diebeck %>%
  group_by(MineFull) %>%
  summarize_at("Year", funs(cnt = n()))

diebeck.by.year.tot <-
  diebeck %>%
  select(isDieback, Year, MineFull) %>%
  group_by(MineFull, Year) %>%
  summarise_at("Year", funs(cnt=n())) %>%
  inner_join(diebeck.by.year, by = "MineFull") %>%
  mutate(prcnt.tot = cnt.x / cnt.y)

diebeck.by.mine <-
  diebeck.by.year.tot %>%
  group_by(MineFull) %>%
  summarize_at("prcnt.tot", funs(avg=mean))
# percent of isDiebeck
plot.diebeck <-
  diebeck.by.year.tot %>%
  inner_join(diebeck.by.mine, by = "MineFull") %>%
  rename(Mine = MineFull) %>%
  ggplot(aes(x = Year, y = prcnt.tot, color = Mine)) +
  geom_line() +
  geom_line(aes(y = avg), linetype = "dashed") +
  xlab("Year") + ylab("Percentage") +
  labs(
      title = "Trees Afflicted with Diebeck",
      caption = "Dashed lines indicate average"
    )

climate.veg.vars <- climate %>% select(Veg.SR, Year, Mine=MineFull)
climate.veg <-
  climate.veg.vars %>%
    group_by(Mine, Year) %>%
    summarize_at("Veg.SR", funs(avg = mean))

climate.veg.mine <-
    climate.veg.vars %>%
    group_by(Mine) %>%
    summarise_at("Veg.SR", funs(avg = mean))

plot.veg <-
  climate.veg.vars %>%
    inner_join(climate.veg, by = c("Year", "Mine")) %>%
    inner_join(climate.veg.mine, by = "Mine") %>%
```

```r
  ggplot(aes(x = Year, y = Veg.SR, color = Mine)) +
    geom_point(alpha = 0.6) +
    geom_line(aes(y = avg.x), show.legend = FALSE) +
    geom_line(aes(y = avg.y), linetype = "dashed", show.legend = FALSE) +
    xlab("Year") + ylab("Number of Species") +
    labs(
      title = "Number of Species Found",
      caption = "Dashed lines indicate average for all years. Solid lines indicate average across yea
    )
```

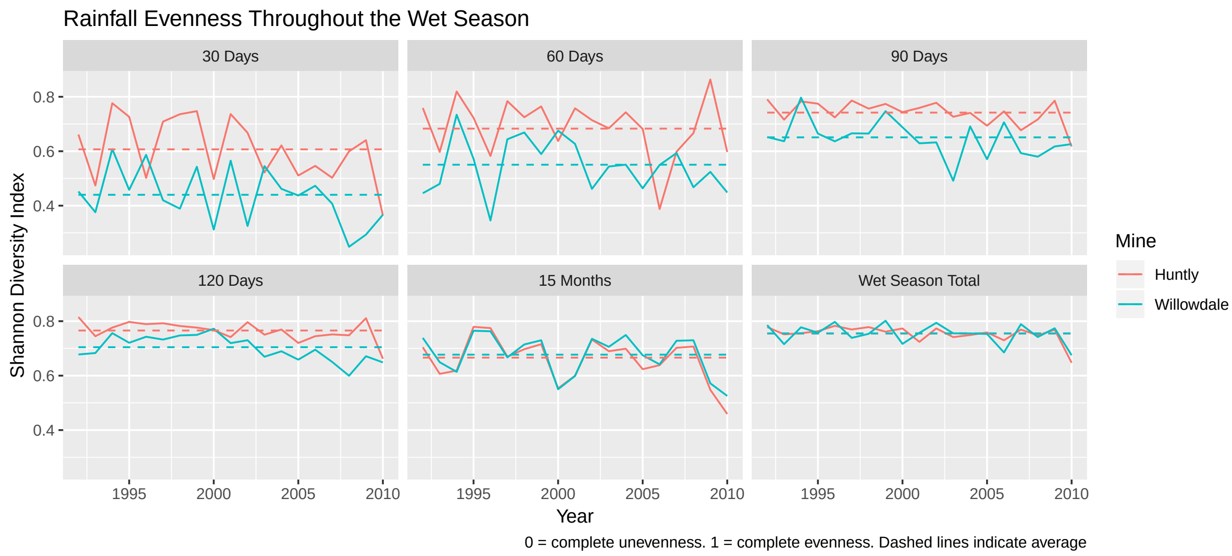
A Closer Look

Though not pictured, Rainfall N is positively correlated with with Rainfall (N+30). Rainfall is generally more consistent at the Huntly mining site. At the start of the wet season, the consistency at both sites is variable year-over-year but tends to flatten out as the season progresses. This is evident from the converging averages as well as the less erratic nature of measures for the Wet Season Total. For simplicity and to resolve multicollinearity, Wet Season Totals will be selected since they are similar for both mine sites and contain the least amount of variation. As expected, rainfall increases throughout the wet season but oddly enough, rainfall after 15 months is much higher than the season total. This is a curiousity that would be explored more if this variable were selected.

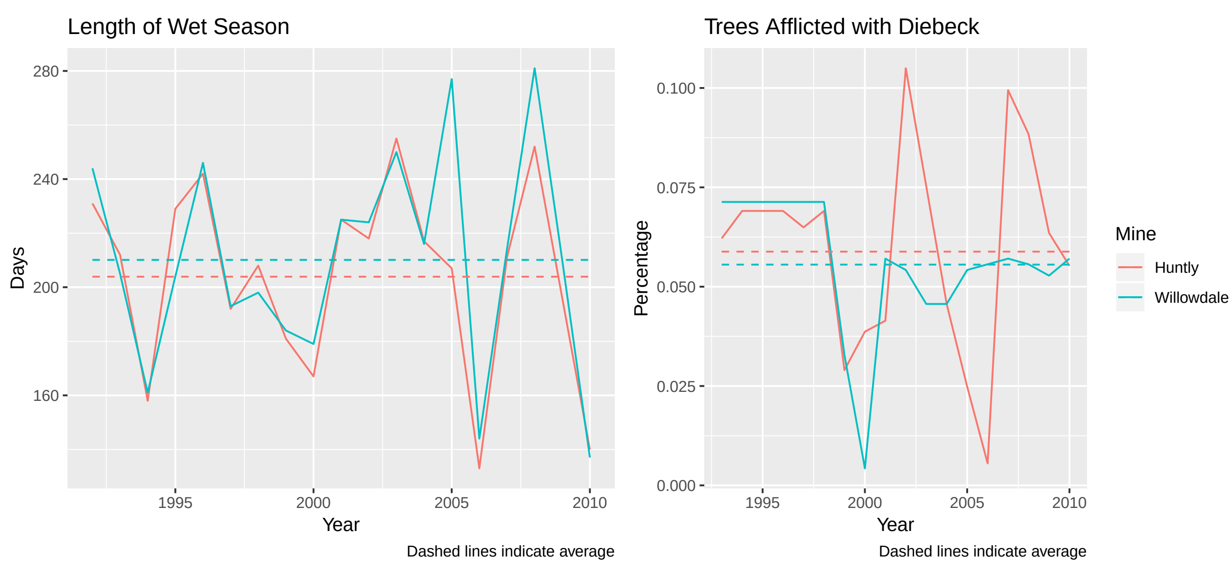
```r
# show the plots here. This is done so I can quickly change display order
plot.scatter.rainfall
```



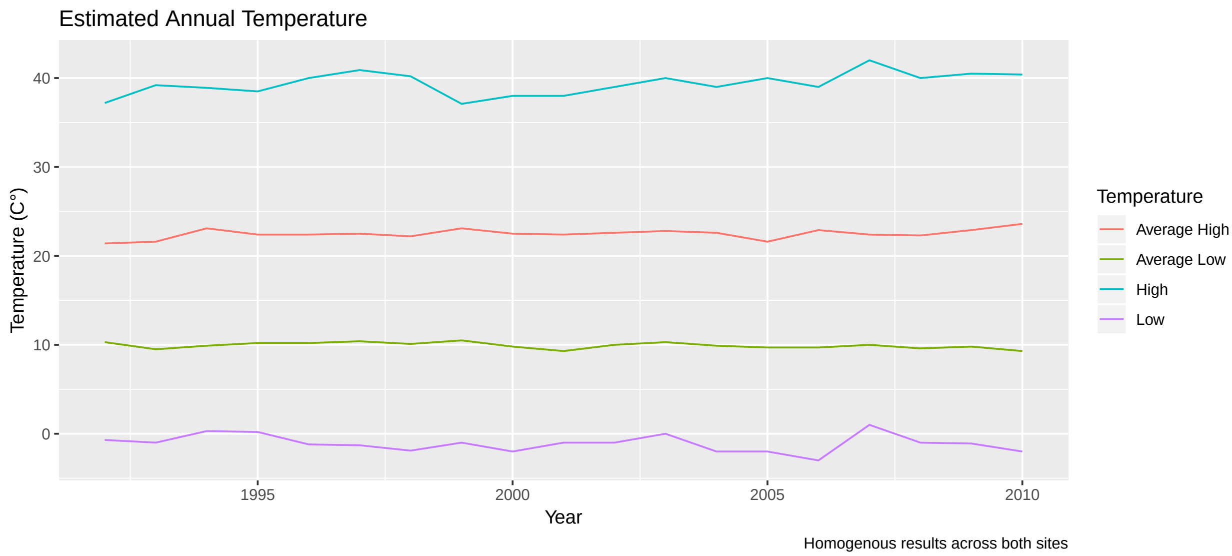

```r
plot.scatter.raineveness
```




```r
grid.arrange(plot.wetseason, plot.diebeck, ncol = 2)
```

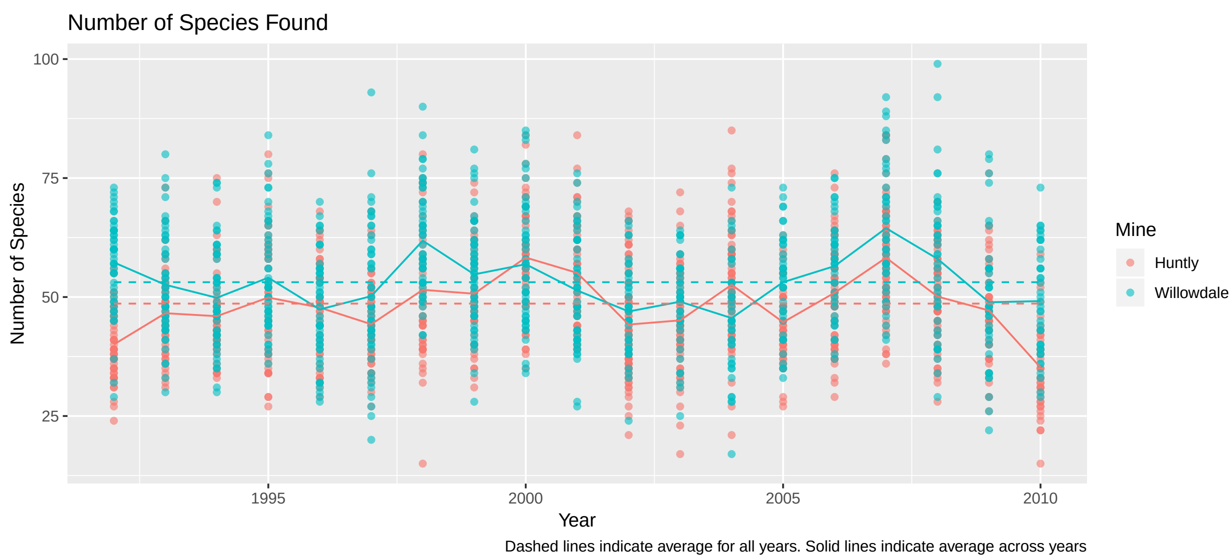



```r
plot.temps
```



The length of wet season and the trees afflicted with Diebeck are variable year-over-year. For a more consistent measure, serial correlation may be present.

```
plot.veg
```

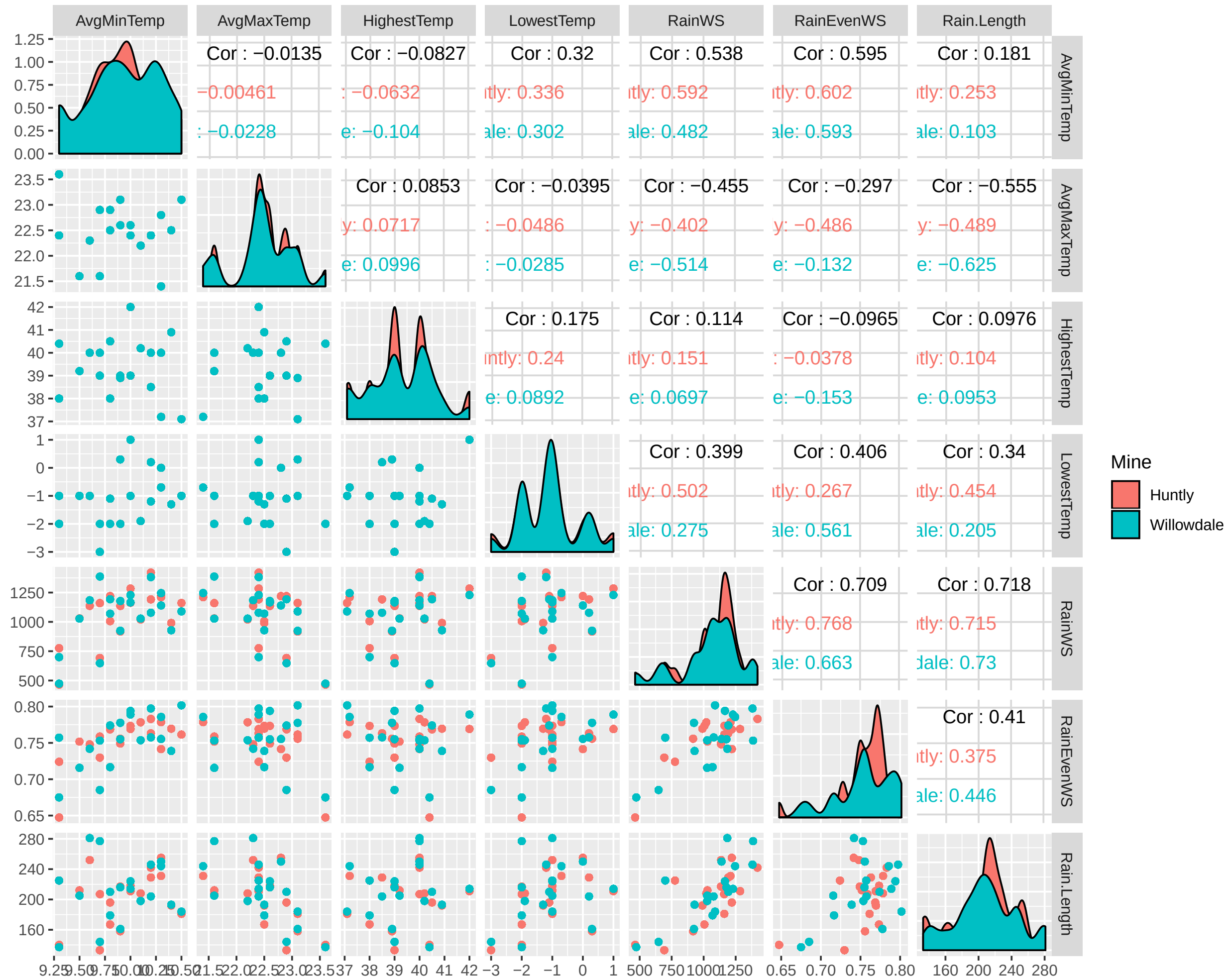


Willowdale in general appears to have a higher number of species found than Huntly. There also appear to be a few more large outliers at Willowdale compared to huntly which influence the overall average. The peaks and valleys of this chart do not seem to line up as nicely as the others indicating that perhaps climate variables are not the biggest influence in number of species found.

Assessing Transformations

The strongest correlation is between Wet Season Length and Rainfall for the Wet Season with a coefficient of 0.718 across both campsites. There appears to be a log-like relationship between Average Min Temperature and Rain Evenness but given that Rain Evenness is an index, a log transformation does not affect its shape.

```r
climate %>%
  select(Mine=MineFull, AvgMinTemp, AvgMaxTemp, HighestTemp, LowestTemp, RainWS, RainEvenWS, Rain.Lengt
  ggpairs(
    mapping = aes(color = Mine),
    columns = c("AvgMinTemp", "AvgMaxTemp", "HighestTemp", "LowestTemp", "RainWS", "RainEvenWS", "Rain.
    legend = c(1,1)
  )
```



2002 & 2006

2002 is the maxima for percentage of trees affected with Diebeck for the Huntly mine. It also appears to be a maxima for rainfall, rainfall evenness for Wet Season Totals, and a minor maxima for the length of Wet Season. 2006 is the minima for percentage of trees affected with Diebeck and similar minimas are apparent in the aforementioned variables. This indicates that there may be a relationship between trees being affected Diebeck and rainfall. 2006 yields a maxima for the average high and minimas for High and Low temperatures. No such relationship is apparent for 2002.

Inferential Procedures

Species Richness is a measure describing the number of vegetation species found in a given plot. If each species is considered a success, then the number of species in a given plot can be considered some number of successes during a period of time; time in this case is measured by year. This can be modeled with by Negative Binomial or a Poisson General Linear Model. First, Poisson and Quasipoisson models will be reviewed for adequacy, followed by a Negative Binomial Model. A Likelihood Ratio Test between the Poisson and Negative Binomial models will be run in order to determine which one is more suitable.

Ascertaining whether or not a given plot is afflicted with diebeck can be done with a logistic regression model. Selecting mostly climate-related variables, a logistic regression model can be created and its error rates analyzed to determine whether or not Wet Season Length is a significant factor in determining whether or not a plot has diebeck. Since the data is time series and therefore not independent, a quasibinomial model will be used to account for overdispersion.

Results

Species

Hypothesis: It is predicted that an increase in average high temperatures and an increase in rain precipitation will yield more vegetation species in a plot.

Ruling out Poisson

Extra-poisson variation is likely since there are known impactful explanatory variables available in the dataset that are not being included in the model. Fitting the model and examining the deviance goodness-of-fit test in conjunction helps determine whether extra-poisson variation is present. All mean values produced by the model are greater than 5 meaning that it is also possible. There is convincing evidence that the poisson model is not inadequate (Deviance Goodness-of-fit Test. p-value = 0). Running the same test for the quasipoisson version model yields the same result (p-value = 0). This indicates that the Poisson distribution is an inadequate model for the response.

```
model.species.poisson <- glm(Veg.SR ~ RainWS + RainEvenWS + AvgMaxTemp + AvgMinTemp + HighestTemp + Lowe

gof4 <- 1 - pchisq(model.species.poisson$deviance, model.species.poisson$df.residual)

model.species.poissonq <- glm(Veg.SR ~ RainWS + RainEvenWS + AvgMaxTemp + AvgMinTemp + HighestTemp + Low

gof5 <- 1 - pchisq(model.species.poissonq$deviance, model.species.poissonq$df.residual)
```

A Negative Binomial model will be reviewed as an alternative to the Quasi-poisson model.

Negative Binomial

The main negative binomial model in question:

$$log(\hat{Veg.SR}) = \beta_0 + \beta_1 RainWS + \beta_2 RainEvenWS + \beta_3 AvgMaxTemp + \beta_4 AvgMinTemp + \beta_5 HighestTemp$$

$$+\beta_6 LowestTemp + \beta_7 Rain.Length + \beta_8 isDiebeck$$

$$log(\hat{Veg.SR}) = 5.919 - 0.00001 RainWS + 0.16 RainEvenWS - 0.0867 AvgMaxTemp + 0.0094 AvgMinTemp$$

$$-0.0033 HighestTemp + 0.0203 LowestTemp - 0.0003 Rain.Length - 0.0359 isDiebeck$$

There is no evidence that this model is inadequate (Deviance Goodness-of-Fit Test. p-value = 0.3244). A Maximum Likelihood Ratio Test indicates that there is convincing evidence that a Negative Binomial GLM is a better fit than a Poisson GLM (p-value = 0.3244).

Testing to see if a model with interaction between each of the variables and the diebeck indicator yield no evidence that the interaction adds value (Drop-in-Deviance Test. p-value = 0.9991).

There is convincing evidence that log Number of Species is associated with Average Max Temp, Lowest Temp, and the Diebeck indicator (Two-tailed Wald Test. p-value = 2.31e-06, 0.0089, 0.0069 respectively).

It is estimated that the average number of species in a plot is 1.09 times higher for every degree celcius that the average max temperature decreases after fixing all other variables. With 95% confidence, a one degree celcius decrease in Average Max Temperature increases the average number of species in a plot by between 1.0515 and 1.1312 times.

It is estimated that the average number of species in a plot is 1.02 times higher for every degree celcius that the Lowest temperature increases after fixing all other variables. With 95% confidence, a one degree celcius increase in the Lowest Temperature increases the average number of species in a plot by between 1.0052 and 1.0359 times.

It is estimated that the average number of species in a plot is 1.03 times lower when diebeck is present after fixing all other variables. With 95% confidence, the presence of Diebeck decreases the average number of species in a plot by between 1.0099 and 1.064 times.

There is no evidence that Rainfall or Rain Evenness for the Wet Season have any impact on the number of species in the plot (Two-tailed Wald Test. p-value = 0.8505, 0.6629 respectively).

```
model.species.nb <- glm.nb(Veg.SR ~ RainWS + RainEvenWS + AvgMaxTemp + AvgMinTemp + HighestTemp + Lowes
#odTest(model.species.nb)

# 0.3244
gof3 <- 1 - pchisq(model.species.nb$deviance, model.species.nb$df.residual)

model.species.nb.full <- glm.nb(Veg.SR ~ (RainWS + RainEvenWS + AvgMaxTemp + AvgMinTemp + HighestTemp +

# 0.9991
#dind(model.species.nb.full, model.species.nb)
```

Diebeck

Hypothesis: It is expected that a longer wet season will yield a higher chance of a plot becoming diebeck.

The model with the parameters described above is:

$$logit(diabeck) = \beta_0 + \beta_1 RainWS - \beta_2 RainEvenWS + \beta_3 AvgMaxTemp - \beta_4 AvgMinTemp - \beta_5 HighestTemp$$

$$+\beta_6 LowestTemp + \beta_7 RainLength$$

There is convincing evidence that this model is inadequate (Deviance Goodness of Fit Test. p-value = 1.1102e-16).

Even though the model is inadequate, the Wet Season Length parameter can still be assessed for significance. There is no evidence that the log odds of diebeck are associated with the length of the Wet Season (Wald Logistic Regression on a Single Variable. p-value = 0.3227). Given that this inadequate model contains Wet Season Length and it is not significant, it is feasible that this is enough evidence to discount this relationship. Further analysis with more data would need to be done in order to gain more insight.

```
model.diebeck <- glm(isDieback ~ RainWS + RainEvenWS + AvgMaxTemp + AvgMinTemp + HighestTemp + LowestTem

gof1 <- 1 - pchisq(model.diebeck$deviance, model.diebeck$df.residual)
```

Conclusion

Species

The results surrounding Species prediction turned out to be different than anticipated. There was a small but significant positive multiplicative effect on Species richness when Average High Temperatures decreased and Low Temperatures increased. This is counter to the hypothesis where it was posited that the number of species would increase with average high temperatures and precipitation. Interestingly enough, precipitation had no effect on species. This may be due to the temperature increasing past the threshold when the Jarrah trees best take root. This would explain why an increase in Lowest Temperature also had an associated effect on the number of species taking root.

Diebeck

The hypothesis was that the length of the Wet Season would increase the chances of soil becoming diebeck. The chosen parameters were not sufficient to provide an adequate logistic regression model. For the inadequate model that was fit, Wet Season Length did not significantly affect the log odds of diebeck which could be construed as not being significant. This is a possibility but a tenuous one at best until further analysis is done with more data.

Recommendations

The goal of this analysis was to understand the impact of climate on diebeck in soil and the ability for vegetation species to thrive in the face of climate change. The original paper concludes that climate had a significant but small associated impact to number of species compared to the impact of healthy restoration practices. While this analysis did not include any of the restoration-centric explanatory variables (barring the diebeck indicator), it was able to provide similar conclusions.

This dataset has a wealth of possibilities but it requires a decent understanding in climate and forestry in order to draw conclusions. I reccommend appropriate time series analysis techniques and further research on the proper environmental conditions and growth practices for Jarrah trees.

Limitations

Time Series analysis is a must for proper analysis. The analysis provided in this report is interesting but the conclusions are weak since the modeling does not account for trends, seasonality, or spatial analysis between sites. The biggest limitation for this analysis was experience with the appropriate analysis techniques required to correctly analyze this data. Further research in Time Series Analysis is needed in order to adjust for serial correlation between Years appropriately. I would have used more of the restoration variables had I understood them better. The particular focus for this analysis was mostly on climate with a touch of restoration but that led to weak models due to climate only being part of the equation.

Appendix

Variables

Temporal variable

Name	Description	Type
Year	1992 to 2010	continuous

Spatial variables

Name	Description	Type
Mine	Huntly (HU) or Willowdale (WD)	categorical
Temp.Plot	Year of Monitoring, then plot number	identifier
Easting	In UTMs to 2 d.p.	continuous
Northing	In UTMs to 2 d.p.	continuous
Elevation	Elevation in metres asl to 2 d.p.	continuous
Slope	Slope in degrees to 4 d.p.	continuous
Aspect	Aspect in degrees to 4 d.p.	continuous

Climate variables

Name	Description	Type
AvgMinTemp	Annual estimate	continuous
AvgMaxTemp	Annual estimate	continuous
LowestTemp	Annual estimate	continuous
HighestTemp	Annual estimate	continuous
DaysOver35	Per year	continuous
Rain.Length	Wet seasonal length in days	continuous
Drought.Length	Length of first summer drought in days	continuous
False.Start	1 = yes. 0 = no.	factor
Rain30	Rainfall amount 30 days after onset of wet season	continuous
Rain60	Rainfall amount 60 days after onset of wet season	continuous
Rain90	Rainfall amount 90 days after onset of wet season	continuous
Rain120	Rainfall amount 120 days after onset of wet season	continuous
RainWS	Wet season total rainfall	continuous
RainPrevWS	Previous year's wet season total rainfall	continuous
RainEven30	Rainfall evenness at 30 days*	continuous
RainEven60	Rainfall evenness at 60 days*	continuous
RainEven90	Rainfall evenness at 90 days*	continuous
RainEven120	Rainfall evenness at 120 days*	continuous
RainEvenWS	Rainfall evenness at end of wet season*	continuous
RainEven15mth	Rainfall evenness at 15 months	continuous
Rain15mth	Total rainfall between onset of wet season and approximate date of vegetation monitoring	continuous

* Rainfall Evenness is calculated using the modified Shannon diversity index. Values vary from 0 to 1 where 0 implies complete unevenness (i.e. all rain recorded for the period falls in one event), and 1 implies complete

evenness (i.e. rain recorded for the period falls in equal amounts each day) (Rachel J. Standish 2015).

(all data sourced from Bureau of Meterology, Dwellingup climate station, the closest station to HU and WD)

Variables describing restoration practice

Name	Description	Type
Topsoil.rateALL	Ordinal scale for topsoil rtn	ordinal
StripSpread.delay	Delay between stripping the topsoil from the donor site and spreading it onto the restoration site: days.	continuous
Wet.start	Date of the start of the wet season according to Ombrothermic relationship.	continuous
SeedMix.tt	Seed mix treatment according to year of application and mine site	categorical
SeedMix.SR	Number of species in the seed mix per year of application and mine site	continuous
Dieback	Dieback-affected topsoil (DB) or dieback-free topsoil (DBF); NA = missing data	categorical
FertP	Expressed kg elemental per ha	continuous
FertN	Expressed kg elemental per ha	continuous
FertK	Expressed kg elemental per ha	continuous
SeedWetstart.syn	Number of days between seeding and the start of the wet season. Can be negative or positive	continuous

Response Variables

Name	Description	Type
Similarity	Bray-Curtis similarity between plot-level vegetation and seed mix applied at HU or WD for relevant year (n= 518 species)	continuous
Veg.SR	Species richness of plot-level vegetation (n= 491 species)	continuous
Prop.Sp.shared	Sp.shared.with.Speciose' divided by 'Veg.SR'	continuous

Topsoil.rateALL

Score	Description
Score 1	Fallow-stockpiled topsoil (F-STP); (Fresh)-stockpiled topsoil (STP); Fallow-sieved topsoil (F-SS); Stockpiled sieved topsoil (STPSS) and fallow stockpiled and screened topsoil (F-STPSS).
Score 2	Fallow-direct returned topsoil, includes all ratios of return rates (F-DRT).
Score 3	Fresh sieved topsoil at rate of 20 m2 (SS)
Score 4	Fresh sieved topsoil at rate of 30 m2 (SS)
Score 5	Fresh sieved topsoil at rate of 40 m2 (SS)
Score 6	Fresh direct-returned 1: 4 unsieved topsoil; includes all return ratios greater than 1: 4 too (DRT)
Score 7	Fresh direct-returned 1: 1 unsieved topsoil, 1: 2 unsieved topsoil, 1: 3 unsieved topsoil (DRT)

References

Bates, B.C. 2008. *Key Findings from the Indian Ocean Climate Initiative and Their Impact on Policy Development in Australia*. British Ecological Society.

Geography 11th Std. 2019. Department of School Education: Government of Tamil Nadu. https://www.brainkart.com/article/Ombrothermic-Diagram_33847/.

Jarrah. n.d. Forest Products Commission of Western Australia. http://www.fpc.wa.gov.au.

Phytophthora Root Rot - Fact Sheet. n.d. Mrs. Macquaries Rd Sydney MSW: Plant Disease Diagnostic Unit. https://web.archive.org/web/20050307053709/http://www.rbgsyd.gov.au/information_about_plants/pests_diseases/fact_sheets/phytophthora_root_rot.

Rachel J. Standish. 2015. *Long-term Data Suggest Jarrah-forest Establishment at Restored Mine Sites Is Resistant to Climate Variability*. British Ecological Society. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2745.12301.