warmXtrophic Project: Specific Leaf Area (SLA) Analyses

Phoebe Zarnetske

June 29, 2021

Load in packages & data

```r
# Clear all existing data
rm(list = ls())

# Load packages
library(tidyverse)
library(bbmle)
library(lmerTest)
library(fitdistrplus)
library(lme4)
library(car)


# Set ggplot2 plotting This code for ggplot2 sets the theme to mostly black and
# white (Arial font, and large font, base size=24)
theme_set(theme_bw(14))
theme_update(axis.text.x = element_text(size = 12, angle = 90), axis.text.y = element_text(size = 12))

# Set working directory to Google Drive
setwd("/Volumes/GoogleDrive/Shared drives/SpaCE_Lab_warmXtrophic/data/")

# Read in data
sla <- read.csv("./L1/SLA/SLA_L1.csv")
```

Check what species are measured at each site

```r
unique(sla$species)
```

```
## [1] "Dasp" "Popr" "Cest" "Hisp" "Soca" "Ruac"
```

```r
with(sla, table(sla$site, sla$species))
```

```
##
##         Cest Dasp Hisp Popr Ruac Soca
##   kbs      0    0   41    0    0   73
##   umbs   233  205    0  219  185    0
```

Each species is only at one site, so for now don't make separate dataframes (just
use species as fixed effect)

## Data exploration - some different ways of visualizing these data

```r
ggplot(sla, aes(area_cm2, fill = species)) + geom_histogram(binwidth = 0.5) + facet_grid(year ~
    site, margins = TRUE, scales = "free")
```
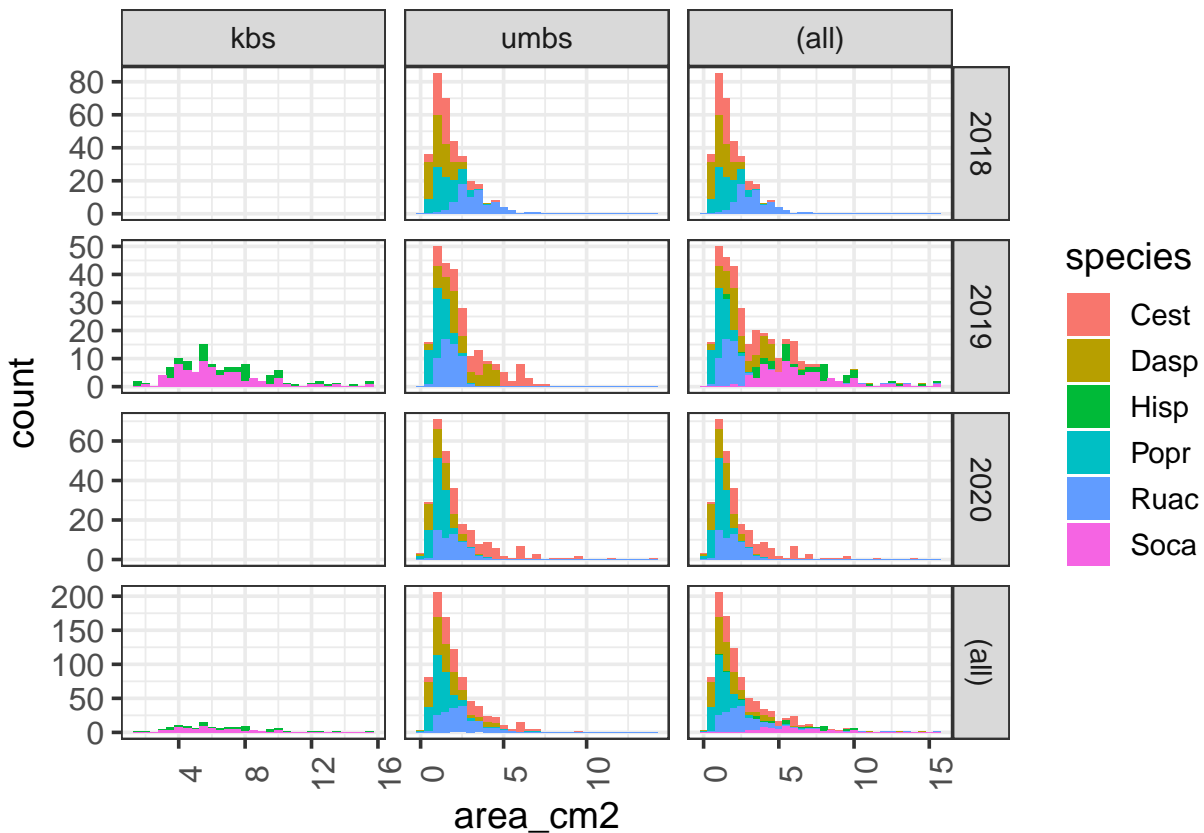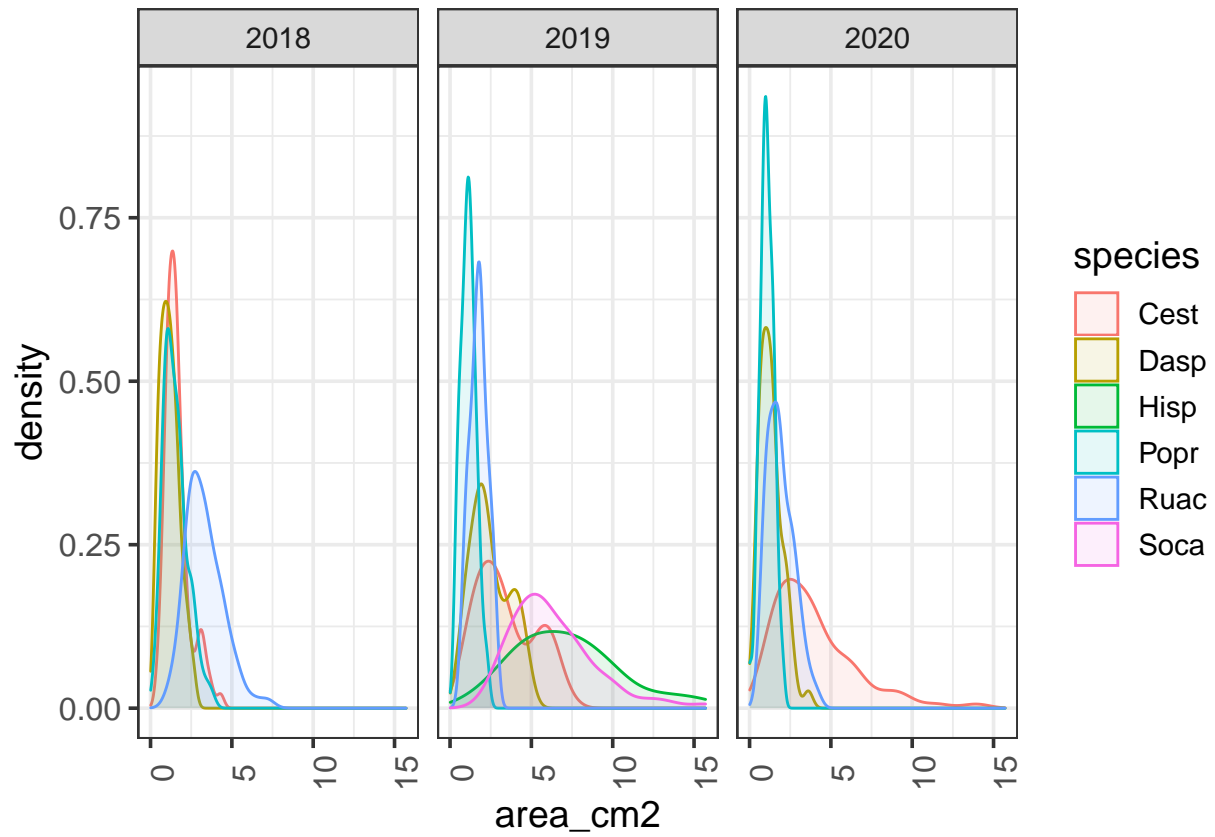


```r
ggplot(sla, aes(area_cm2, fill = as.factor(species))) + geom_histogram(binwidth = 0.5) +
    facet_grid(species ~ year, margins = TRUE, scales = "free")
```
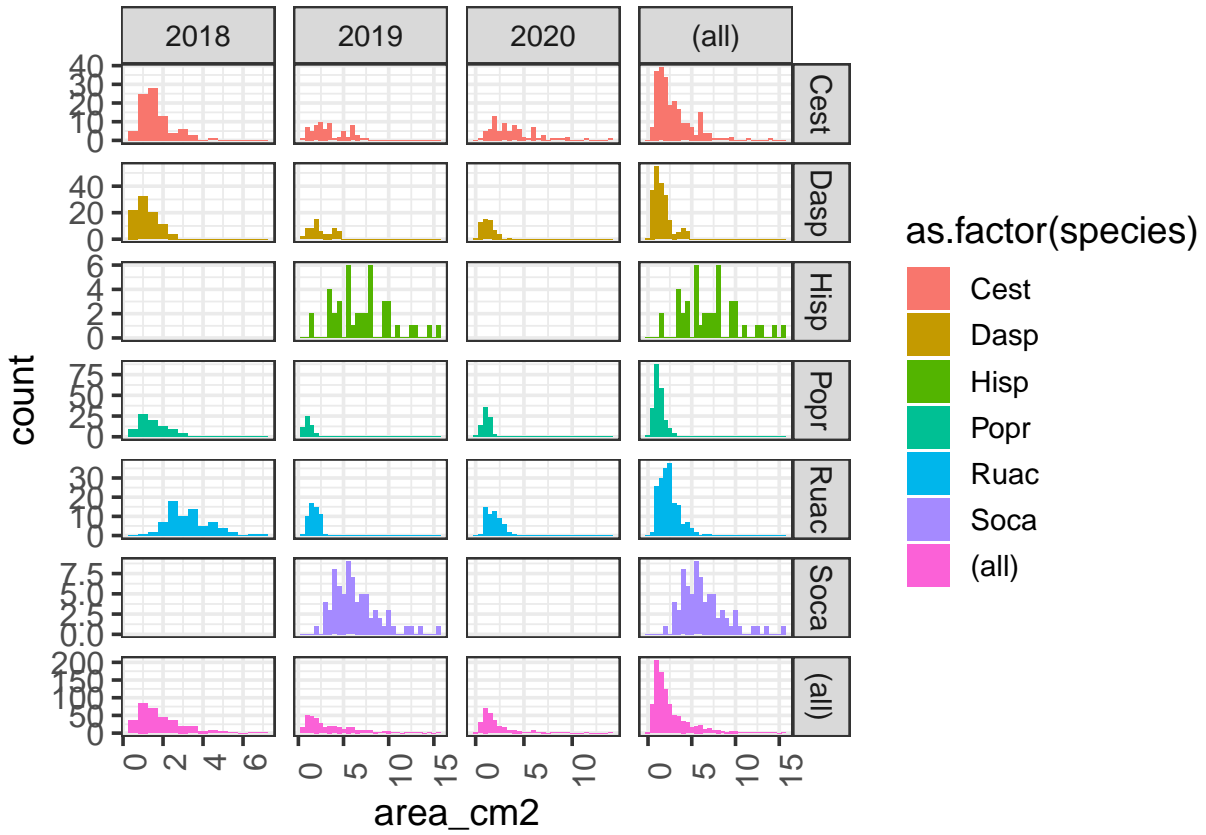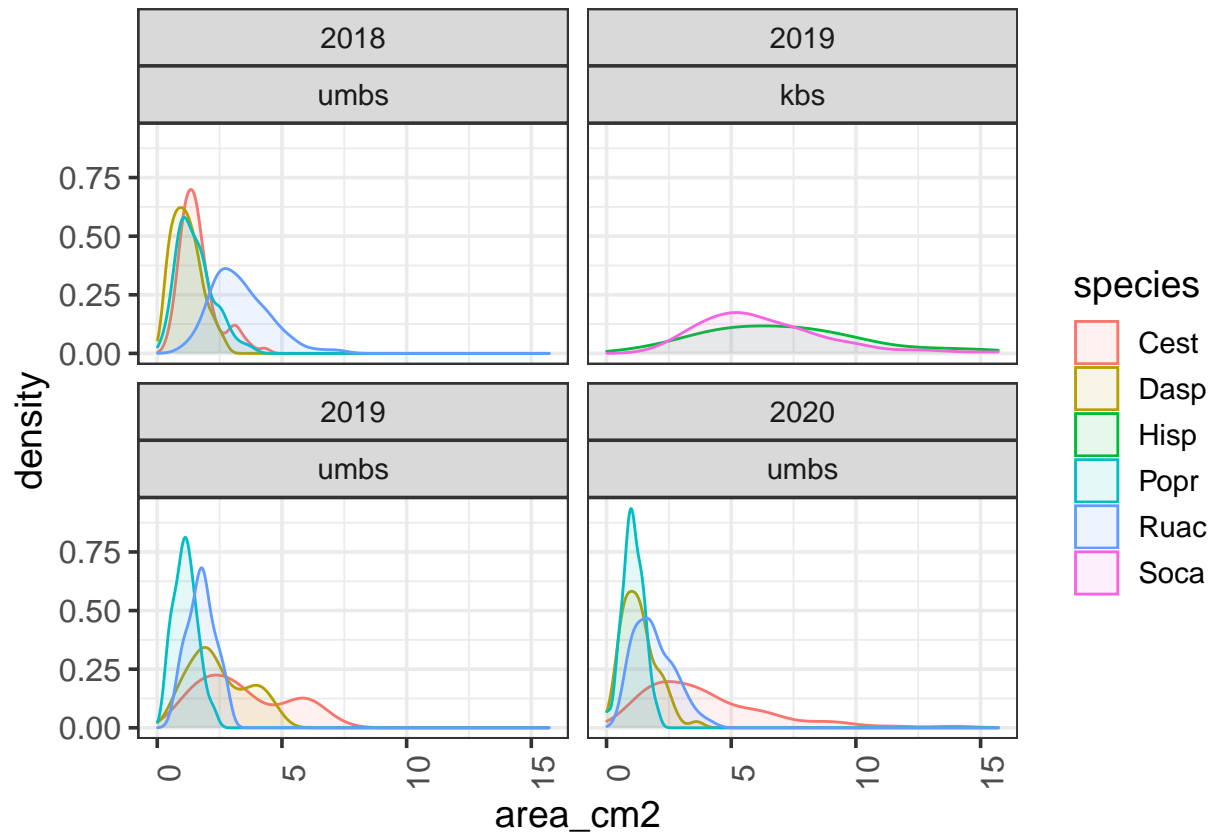
```
ggplot(sla, aes(area_cm2, fill = species, color = species)) + geom_density(alpha = 0.1)
```
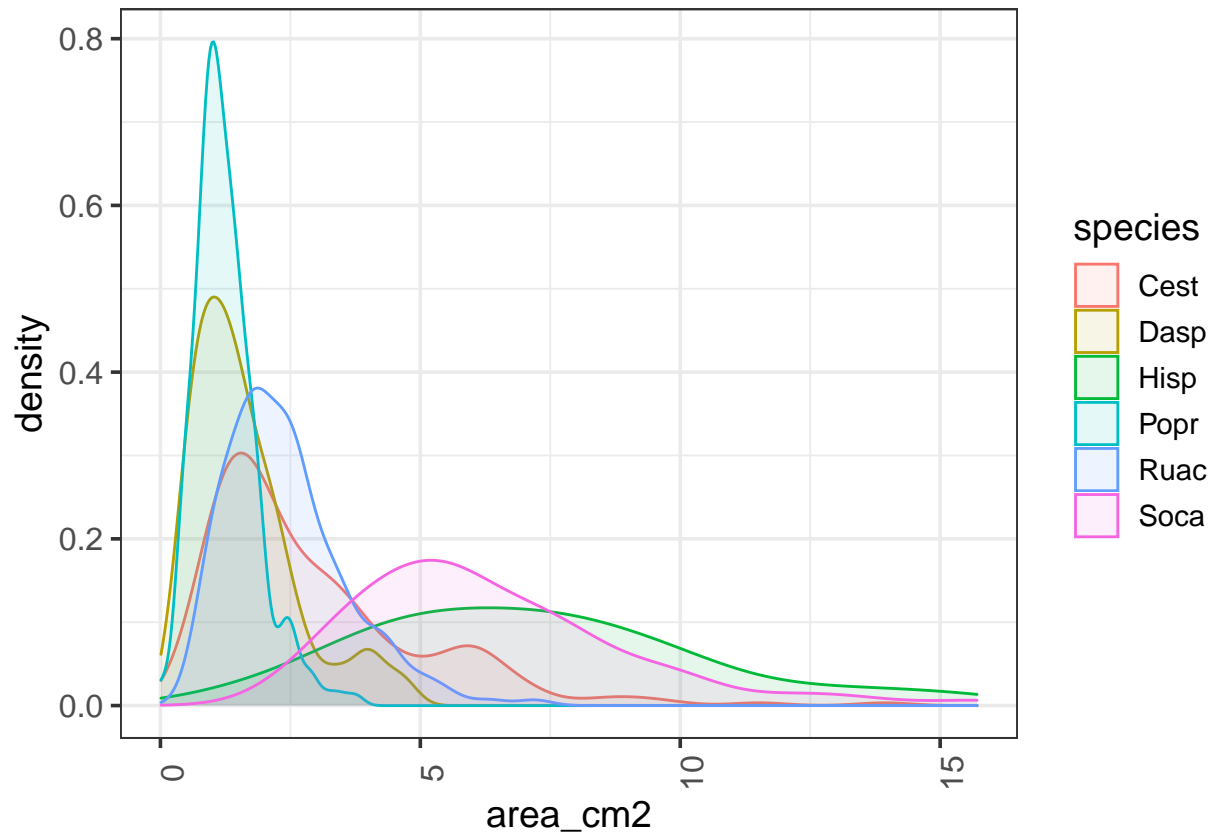
```
ggplot(sla, aes(area_cm2, fill = species, color = species)) + geom_density(alpha = 0.1) +
    facet_wrap(~year)
```
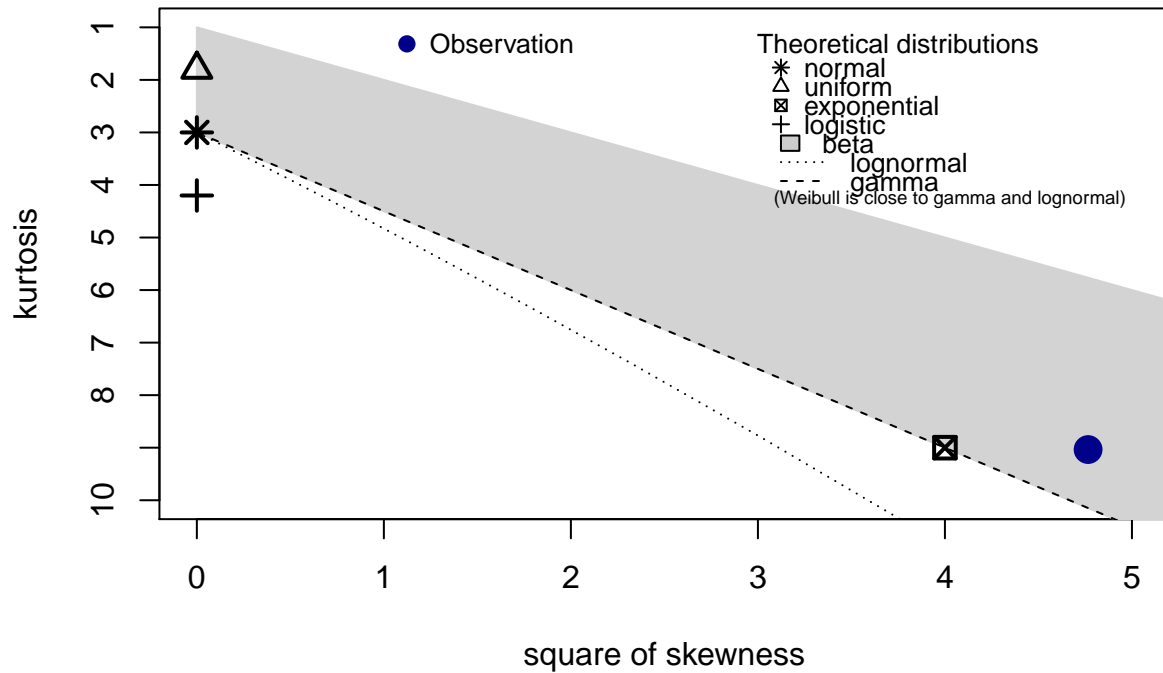
```
ggplot(sla, aes(area_cm2, fill = species, color = species)) + geom_density(alpha = 0.1) +
    facet_wrap(~year + site)
```

```
# Exploring distributions for these right-skewed data: See:
# http://www.di.fc.ul.pt/~jpn/r/distributions/fitting.html
descdist(sla$area_cm2, discrete = FALSE)
```
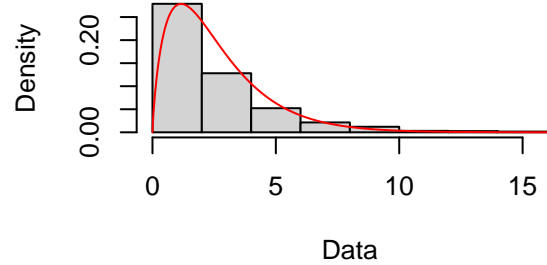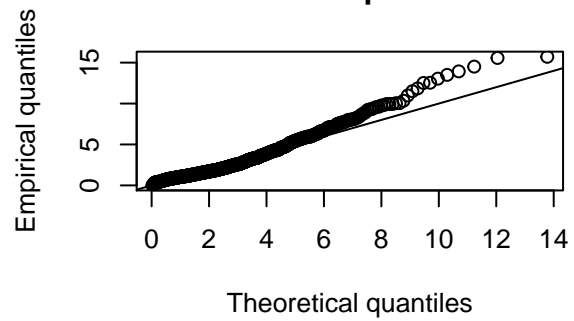
Cullen and Frey graph


```
## summary statistics
## ------
## min:  0.01    max:  15.71
## median:  1.8
## mean:  2.607322
## estimated sd:  2.277138
## estimated skewness:  2.183
## estimated kurtosis:  9.035438
```

```r
# Gamma distribution is an option
fit.gamma <- fitdist(sla$area_cm2, "gamma")
plot(fit.gamma)
```

Empirical and theoretical dens.

Q-Q plot

Empirical and theoretical CDFs
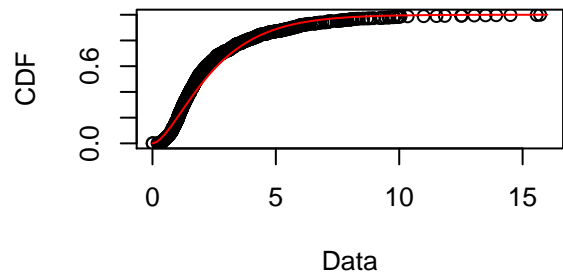
P-P plot

```r
# Exponential distribution is another option
fit.exp <- fitdist(sla$area_cm2, "exp")
plot(fit.exp)
```

Empirical and theoretical dens.
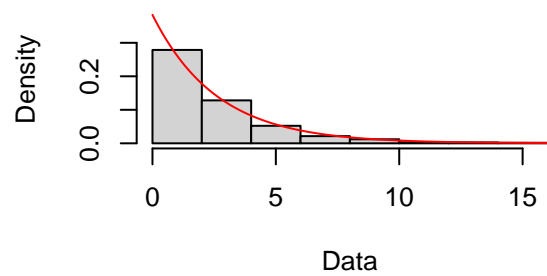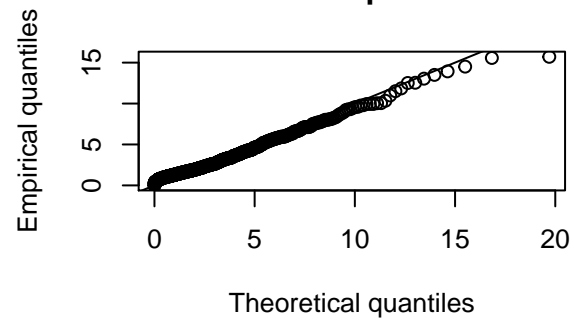
Q–Q plot

Empirical and theoretical CDFs
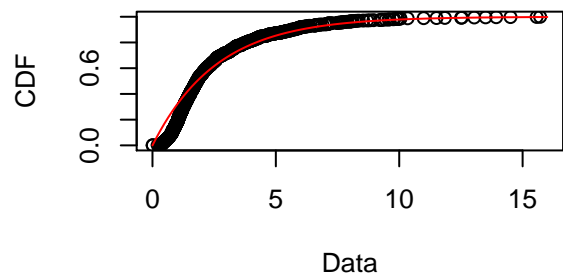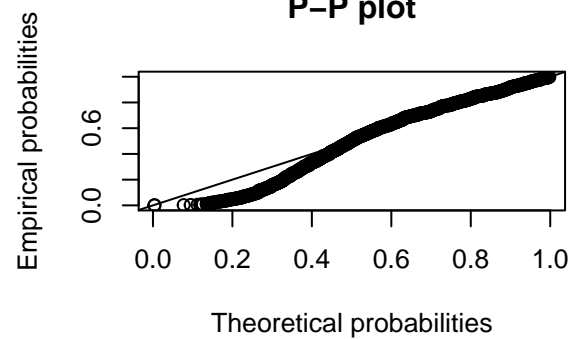
P–P plot


```r
# Lognormal distribution is another option
fit.ln <- fitdist(sla$area_cm2, "lnorm")
plot(fit.ln)
```
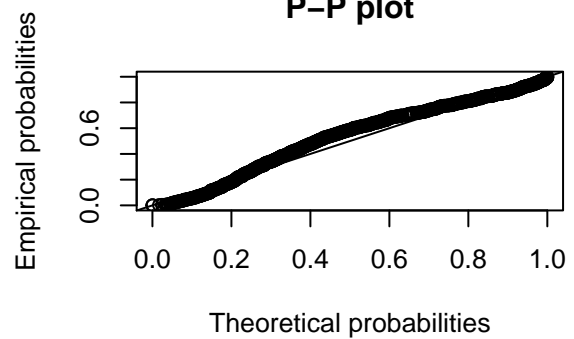
Empirical and theoretical dens.
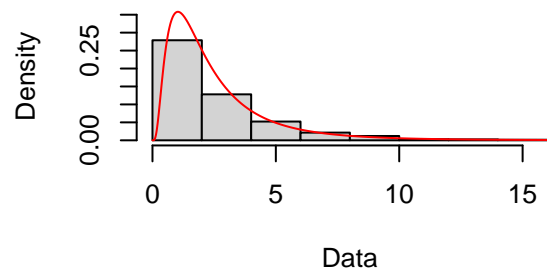
Q–Q plot
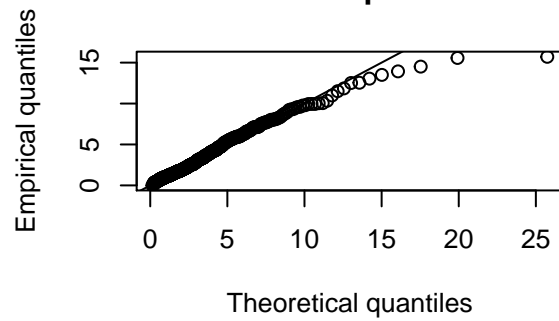
Empirical and theoretical CDFs
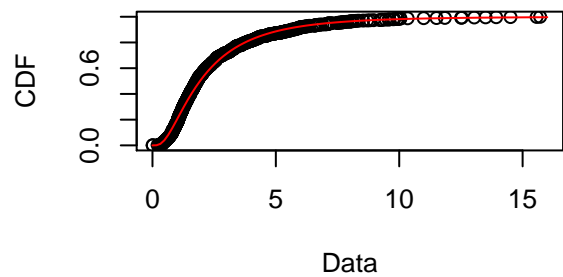
P–P plot
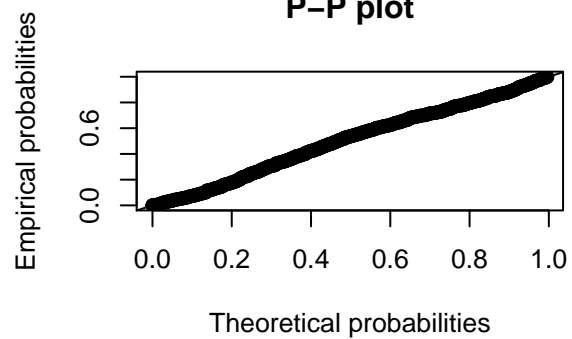
```r
# Square root transformed normal distribution
fit.sr <- fitdist(sqrt(sla$area_cm2), "norm")
plot(fit.sr)
```
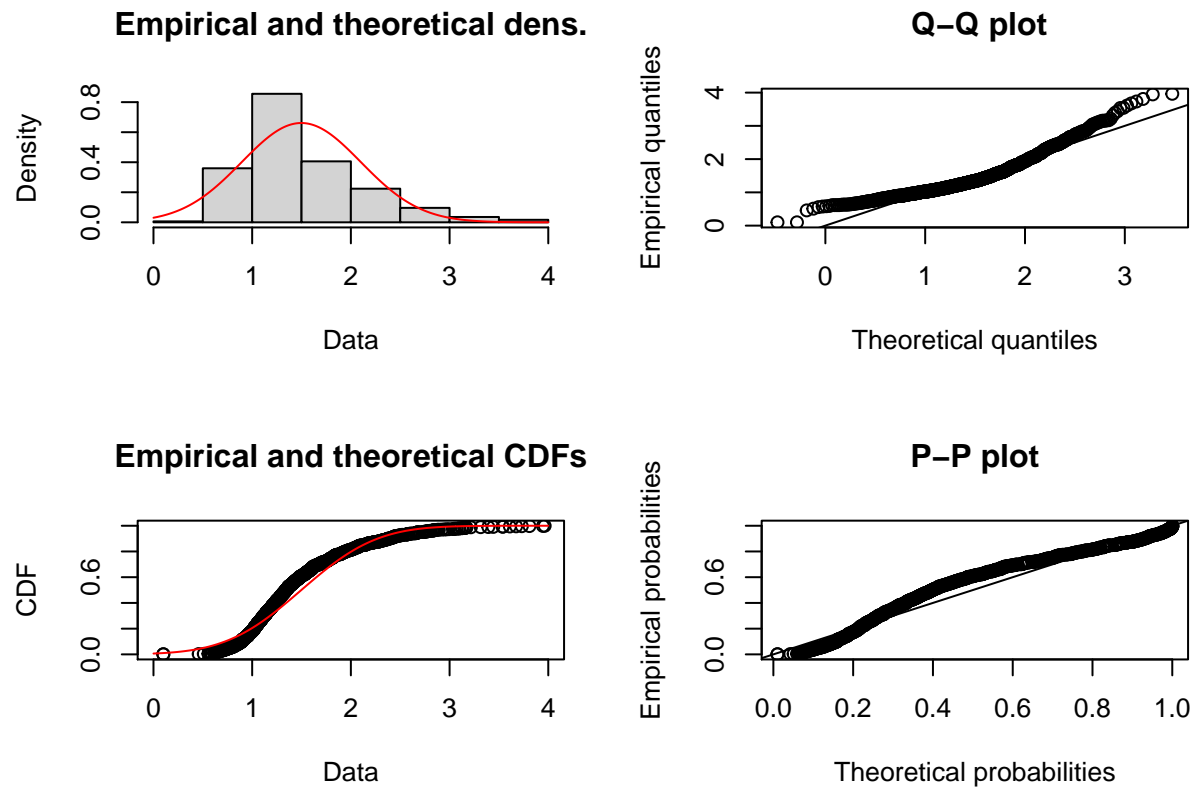
```
par(mfrow = c(2, 2))
plot.legend <- c("Gamma", "Exponential", "Log Normal")
denscomp(list(fit.gamma, fit.exp, fit.ln), legendtext = plot.legend)
cdfcomp(list(fit.gamma, fit.exp, fit.ln), legendtext = plot.legend)
qqcomp(list(fit.gamma, fit.exp, fit.ln), legendtext = plot.legend)
ppcomp(list(fit.gamma, fit.exp, fit.ln), legendtext = plot.legend)
```

Histogram and theoretical densities
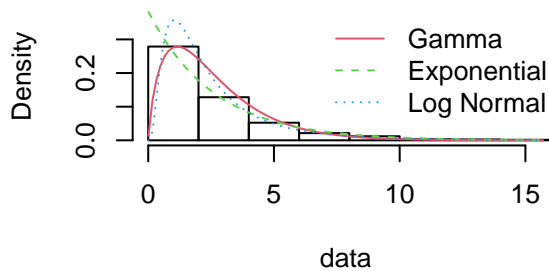
Empirical and theoretical CDFs
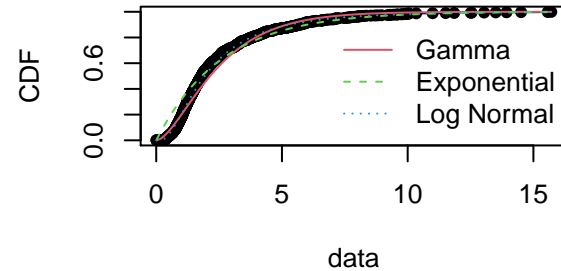
Q–Q plot
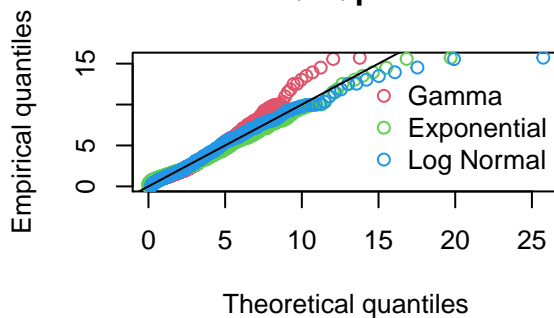
P–P plot
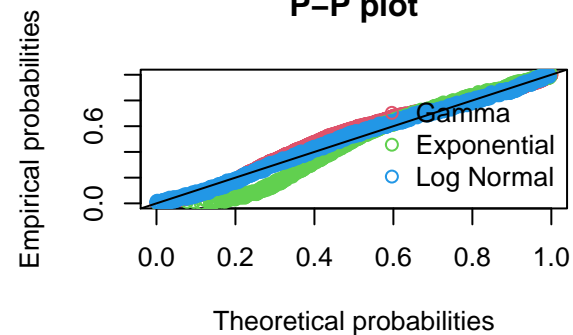
```
# Goodness of fit comparisons across fits (can't include the sqrt normal bc it
# becomes diff response values)
gofstat(list(fit.gamma, fit.exp, fit.ln), fitnames = c("Gamma", "Exponential", "Log Normal"))
```

```
## Goodness-of-fit statistics
##                              Gamma Exponential Log Normal
## Kolmogorov-Smirnov statistic  0.09305212    0.1658967 0.04205516
## Cramer-von Mises statistic    2.45538794    5.9902005 0.39592631
## Anderson-Darling statistic   13.42273156   37.9638617 2.46661286
##
## Goodness-of-fit criteria
##                              Gamma Exponential Log Normal
## Akaike's Information Criterion 3582.421     3746.315    3521.768
## Bayesian Information Criterion 3592.147     3751.178    3531.493
```

```
# Lognormal is a winner so far
```

Determining appropriate distribution in mixed effects model

```
# See Ben Bolker's site for details on fitting glmms:
# https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html Fixed effects: warming,
# year, insecticide Random effect = plot (bc observations are nested within plot)
# Fit with ML bc we are interested in estimating fixed effects more than random
# effects See
# https://uoftcoders.github.io/rcourse/lec08-linear-mixed-effects-models.html for
```

```
# more details
m1 <- lmer(log(area_cm2) ~ state + species + year + insecticide + (1 | plot), data = sla,
    REML = FALSE)

# Check Assumptions: (1) Linearity: if covariates are not categorical (year
# isn't) (2) Homogeneity: Need to Check by plotting residuals vs predicted
# values.
par(mfrow = c(1, 2))
plot(m1, main = "Log Area (cm2)")
```
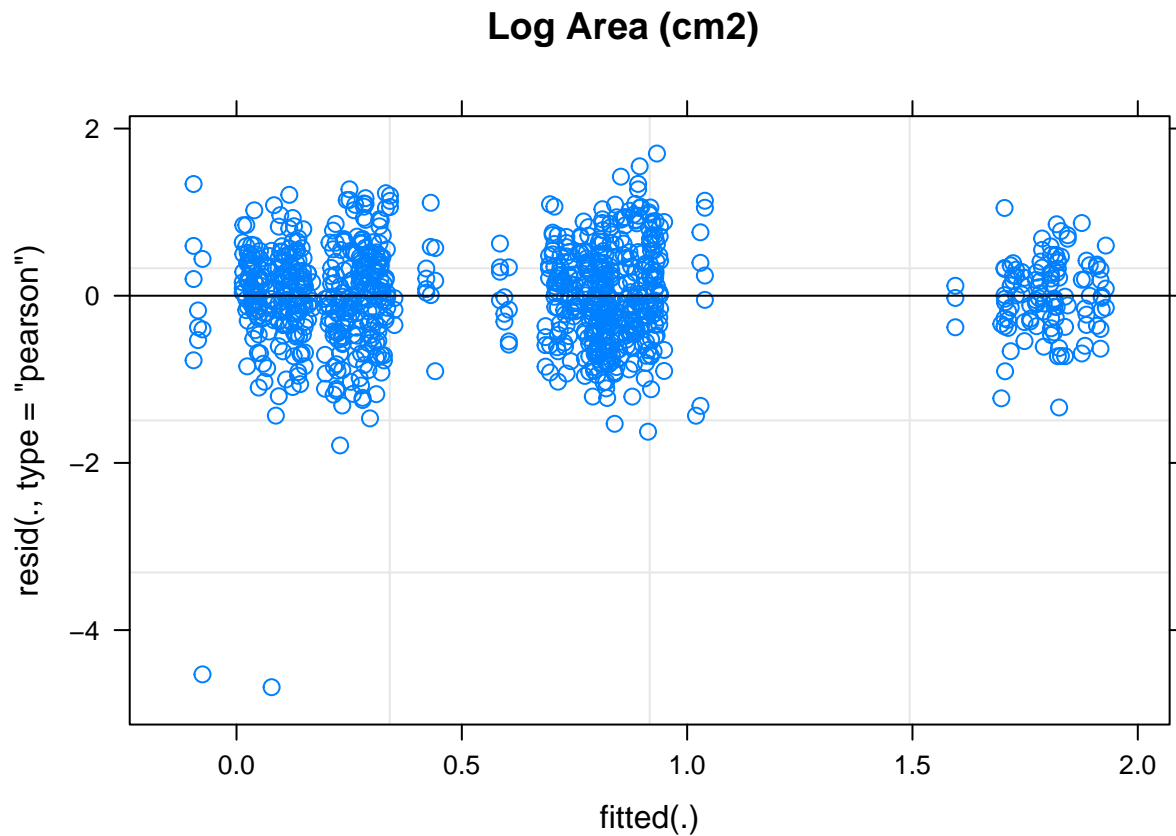
Log Area (cm2)


```
# Homogeneity of variance is ok here (increasing variance in resids is not
# increasing with fitted values) Check for homogeneity of variances (true if
# p>0.05). If the result is not significant, the assumption of equal variances
# (homoscedasticity) is met (no significant difference between the group
# variances).
leveneTest(residuals(m1) ~ sla$state)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value  Pr(>F)
## group    1  6.5179 0.01083 *
##        954
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Assumption not met
leveneTest(residuals(m1) ~ sla$species)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value    Pr(>F)
## group    5  6.4297 6.918e-06 ***
##        950
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Assumption not met - not surprising given the differences among species
leveneTest(residuals(m1) ~ sla$insecticide)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    1  0.1397 0.7087
##        954
```

```r
# Assumption met
leveneTest(residuals(m1) ~ sla$plot)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group   23  0.9514  0.528
##        932
```

```r
# Assumption met

# (3) Normality of error term: need to check by histogram, QQplot of residuals,
# could do Kolmogorov-Smirnov test.  Check for normal residuals
qqPlot(resid(m1), main = "Area (cm2)")
```

```
## [1] 801 377
```

```r
hist(residuals(m1), main = "Area (cm2)")
```

Area (cm2)   **Area (cm2)**

norm quantiles   residuals(m1)
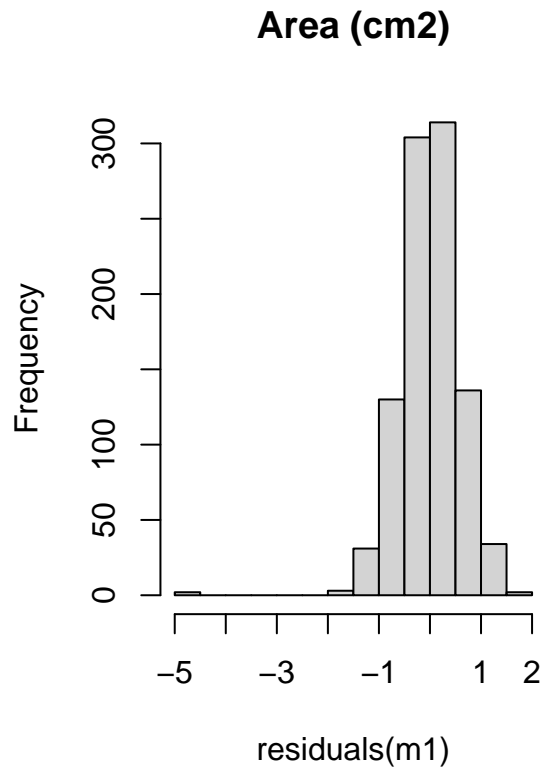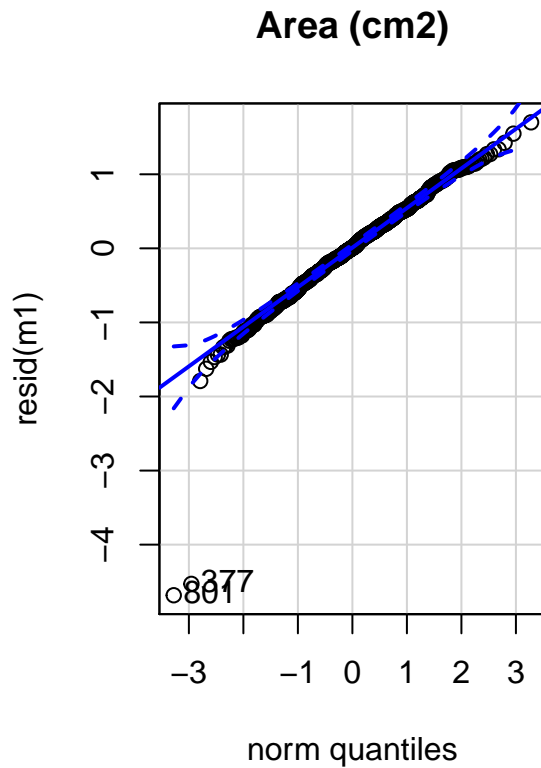
```r
shapiro.test(resid(m1))  # not normally distributed resids bc p>0.05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(m1)
## W = 0.9494, p-value < 2.2e-16
```

```r
# Outliers
sla[377, ]
```

```
##     plot site      date julian species plant_number area_cm2 mass_g  state
## 377   B3 umbs 7/23/2020    205    Popr            2     0.01    0.6 warmed
##     treatment_key insecticide day month year
## 377            WI     insects  23     7 2020
```

```r
sla[801, ]
```

```
##     plot site      date julian species plant_number area_cm2 mass_g  state
## 801   D2 umbs 7/23/2020    205    Popr            4     0.01   0.97 warmed
##     treatment_key insecticide day month year
## 801            W0  no_insects  23     7 2020
```

```r
# Outlier test - yes, these are outliers bc p<0.05
outlierTest(m1)
```

```
##      rstudent unadjusted p-value Bonferroni p
## 801 -8.049862         2.4816e-15   2.3724e-12
## 377 -7.788077         1.7865e-14   1.7079e-11
```

```r
# Remove these 2 points
sla1 <- sla[-c(377, 801), ]
m1.nooutlier <- lmer(log(area_cm2) ~ state + species + year + insecticide + (1 |
    plot), data = sla1, REML = FALSE)
shapiro.test(resid(m1.nooutlier))  # normally distributed resids - good!
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(m1.nooutlier)
## W = 0.99863, p-value = 0.6775
```

```r
# (4) Normality of random effect: Get the estimate of random effect (e.g., random
# intercepts), and check them as you would check the residual. [***need to do for
# final model]
```

Model exploration determined log normal distribution for SLA and dropping 2 outliers. Proceed with re-checking the distribution and assumptions as above with the outliers removed.

```r
# Fixed effects: warming, year, insecticide Random effect = plot (bc observations
# are nested within plot) Fit with ML bc we are interested in estimating fixed
# effects more than random effects See
# https://uoftcoders.github.io/rcourse/lec08-linear-mixed-effects-models.html for
# more details
m2 <- lmer(log(area_cm2) ~ state + species + year + insecticide + (1 | plot), data = sla1,
    REML = FALSE)

# Check Assumptions: (1) Linearity: if covariates are not categorical (year
# isn't) (2) Homogeneity: Need to Check by plotting residuals vs predicted
# values.
par(mfrow = c(1, 2))
plot(m2, main = "Log Area (cm2)")
```
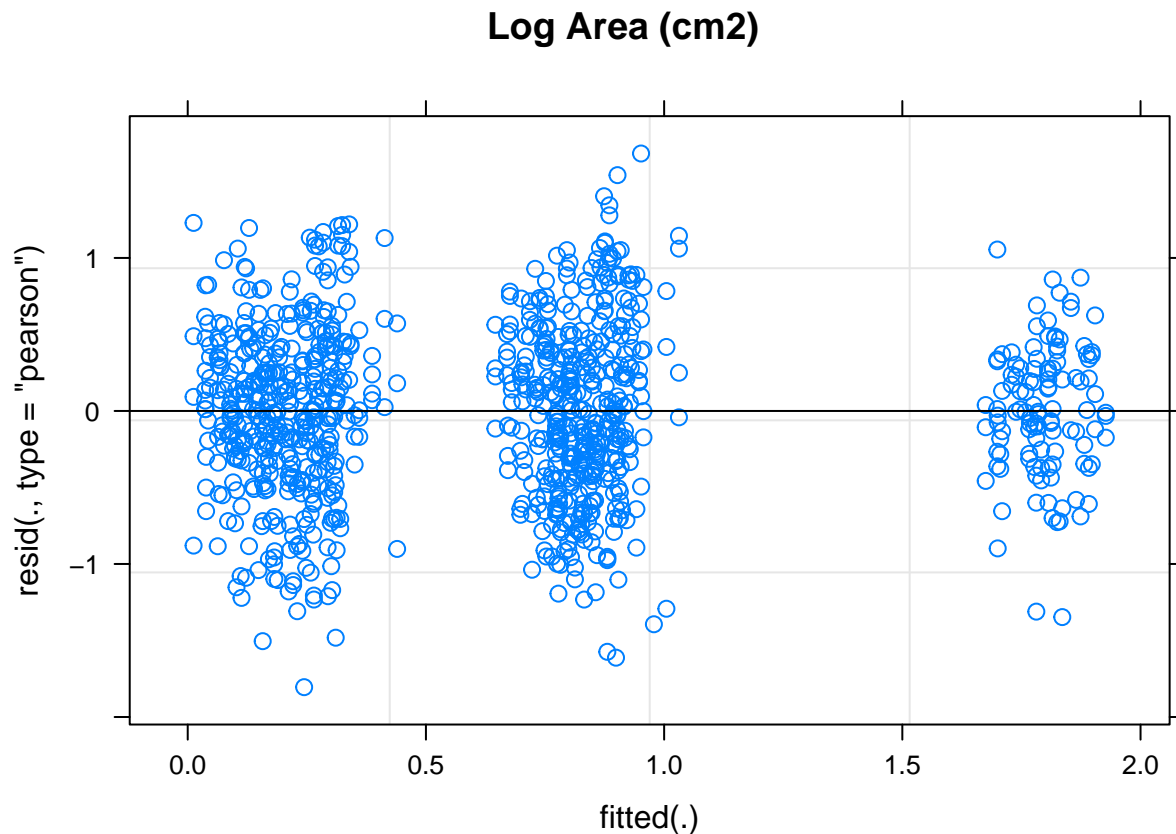
Log Area (cm2)


```r
# Homogeneity of variance is ok here (increasing variance in resids is not
# increasing with fitted values) Check for homogeneity of variances (true if
# p>0.05). If the result is not significant, the assumption of equal variances
# (homoscedasticity) is met (no significant difference between the group
# variances).
leveneTest(residuals(m2) ~ sla1$state)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value  Pr(>F)
## group   1  4.3454 0.03738 *
##       952
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Assumption not met
leveneTest(residuals(m2) ~ sla1$species)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value     Pr(>F)
## group   5  9.4353 8.614e-09 ***
##       948
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Assumption not met - not surprising given the differences among species
leveneTest(residuals(m2) ~ sla1$insecticide)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group    1  0.2575  0.612
##        952
# Assumption met
leveneTest(residuals(m2) ~ sla1$plot)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##         Df F value Pr(>F)
## group   23  0.6489 0.8956
##        930
# Assumption met

# (3) Normality of error term: need to check by histogram, QQplot of residuals,
# could do Kolmogorov-Smirnov test.  Check for normal residuals
qqPlot(resid(m2), main = "Area (cm2)")
```

```
## [1] 106 208
hist(residuals(m2), main = "Area (cm2)")
```

Area (cm2)   **Area (cm2)**
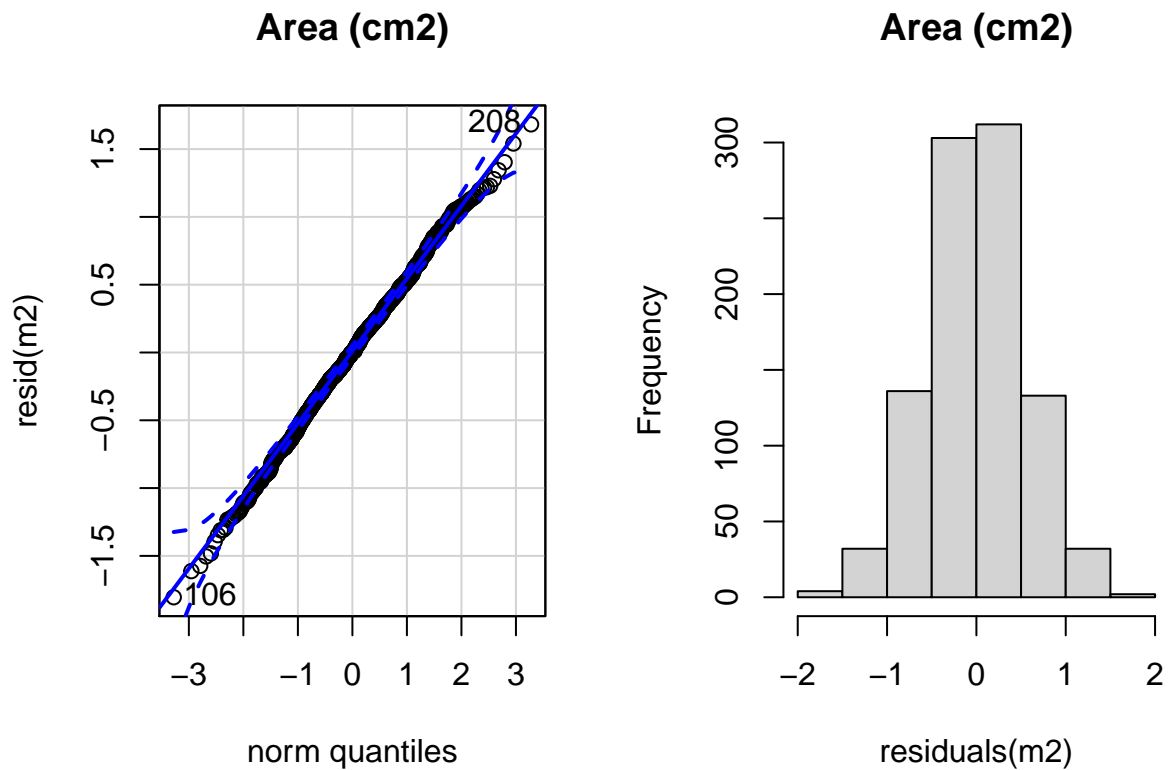
```r
shapiro.test(resid(m2))  # normally distributed resids bc p>0.05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  resid(m2)
## W = 0.99863, p-value = 0.6775
```

```r
# Outlier test - no, these are no outliers bc p>0.05
outlierTest(m2)
```

```
## No Studentized residuals with Bonferroni p < 0.05
## Largest |rstudent|:
##     rstudent unadjusted p-value Bonferroni p
## 106 -3.316548         0.00094621      0.90268
```

```r
# (4) Normality of random effect: Get the estimate of random effect (e.g., random
# intercepts), and check them as you would check the residual.
require(lme4)
r_int <- ranef(m2)$plot$`(Intercept)`
qqnorm(r_int)
qqline(r_int)
shapiro.test(r_int)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  r_int
```

```
## W = 0.92594, p-value = 0.07912
```
```
# Yes, normally distributed random effect.
```

Normal Q–Q Plot
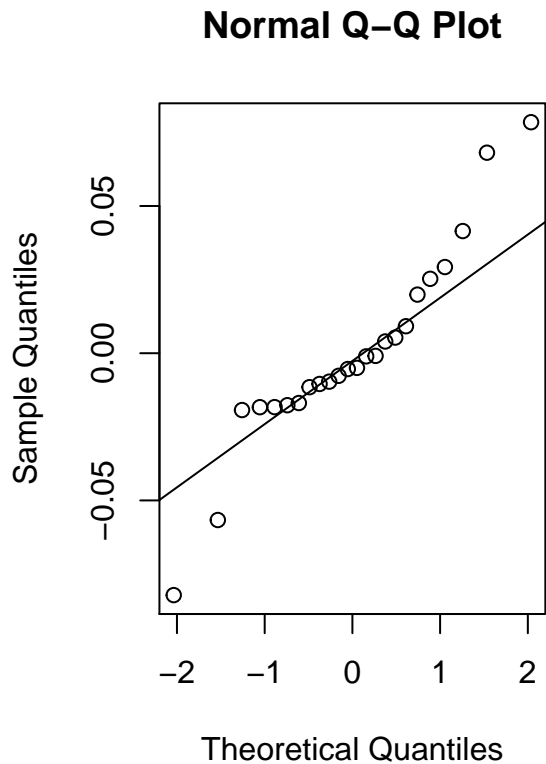
According to the results above, the outlier removal helps the model meed the assumptions. Proceed with model comparison. Start with investigating the model summary to determine which coviarates to drop.

```
summary(m2)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: log(area_cm2) ~ state + species + year + insecticide + (1 | plot)
##    Data: sla1
##
##      AIC      BIC   logLik deviance df.resid
##   1592.4   1645.9   -785.2   1570.4      943
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2911 -0.6385  0.0076  0.6769  3.0677
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  plot     (Intercept) 0.003531 0.05942
##  Residual             0.300796 0.54845
## Number of obs: 954, groups:  plot, 24
##
## Fixed effects:
```

```
##                       Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)          -52.23690   45.59534 941.48314  -1.146   0.2522
## statewarmed            0.04006    0.04376  24.86614   0.915   0.3688
## speciesDasp           -0.59150    0.05290 945.29635 -11.181   <2e-16 ***
## speciesHisp            1.00608    0.09376 946.00772  10.730   <2e-16 ***
## speciesPopr           -0.73661    0.05198 948.62269 -14.171   <2e-16 ***
## speciesRuac           -0.10312    0.05440 953.99284  -1.896   0.0583 .
## speciesSoca            0.89983    0.07362 935.35538  12.223   <2e-16 ***
## year                   0.02628    0.02258 941.48103   1.164   0.2449
## insecticideno_insects  0.06901    0.04342  23.96707   1.589   0.1251
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) sttwrm spcsDs spcsHs spcsPp spcsRc spcsSc year
## statewarmed -0.014
## speciesDasp -0.065  0.053
## speciesHisp  0.004  0.099  0.267
## speciesPopr -0.005  0.036  0.474  0.270
## speciesRuac -0.023 -0.035  0.450  0.249  0.460
## speciesSoca  0.007  0.014  0.332  0.189  0.338  0.322
## year        -1.000  0.013  0.064 -0.004  0.005  0.022 -0.007
## insctcdn_ns  0.019 -0.017 -0.044  0.016 -0.044 -0.031  0.004 -0.019
```

```
# Looks like year and warmed aren't significant - I'm leaving it in for now bc
# the data will change. With fuller data we'd begin model comparison here (will
# pick up here when those data are available). Then run the code below for diff
# models:

# AICctab(m2, m3, weights = T)
```