# How do you analyze non-linear relationships?
Height and number of tillers in wire grass
## *The use of real priors*

Non-linear relationships are common in ecological studies. Here we develop a Bayesian model of an allometric relationship for a grass species. On September 8, 2013, UCF graduate students measured the height and counted the tillers of hundreds of wiregrass (*Aristida beyrichiana*, formerly called *Aristida stricta*) clumps scheduled to be transplanted to the field. The analysis presented in this demo was used as a baseline in a study that evaluates the effect of microhabitat and ridge elevation on wiregrass growth and survival, with hopes of learning how to make successful reintroductions for restoration of native ecosystems. For this demo you will need to download two R scripts (`wire_grass 2019.R` and `wg_priorsscript.R`), and two data files (`wg_data.txt`, `wg_priors.txt`). You will also need the package `rstan` and `rethinking` and the Stan software.
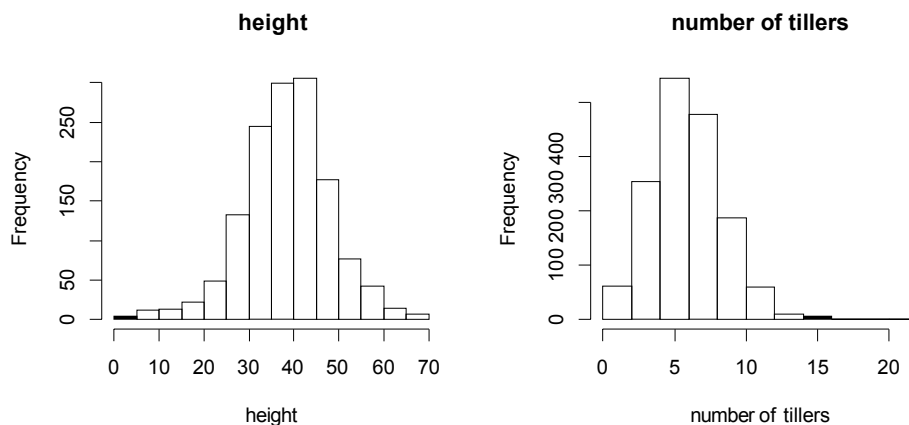
Figure 1. Histograms for plant height in cm and number of tillers of wire grass (data collected by Jennifer Navarra, UCF biology graduate students and Pedro Quintana-Ascencio).

We start in the `wire_grass.R` script by plotting the data in `wg_data.txt` (lg: height, lf: number of tillers). Figure 1 depicts the distribution of height and number of tillers, while the plot in Figure 2 indicates that a quadratic relationship is a reasonable hypothesis about their association.

```
reg_data <- read.table("wg_data.txt", header=T)
par(mfrow=c(1,2))
hist(reg_data$lg)
hist(reg_data$lf)
par(mfrow=c(1,1))
plot(reg_data$lf,reg_data$lg,pch=16,cex=0.55,xlab="number of tillers",ylab="height")
```
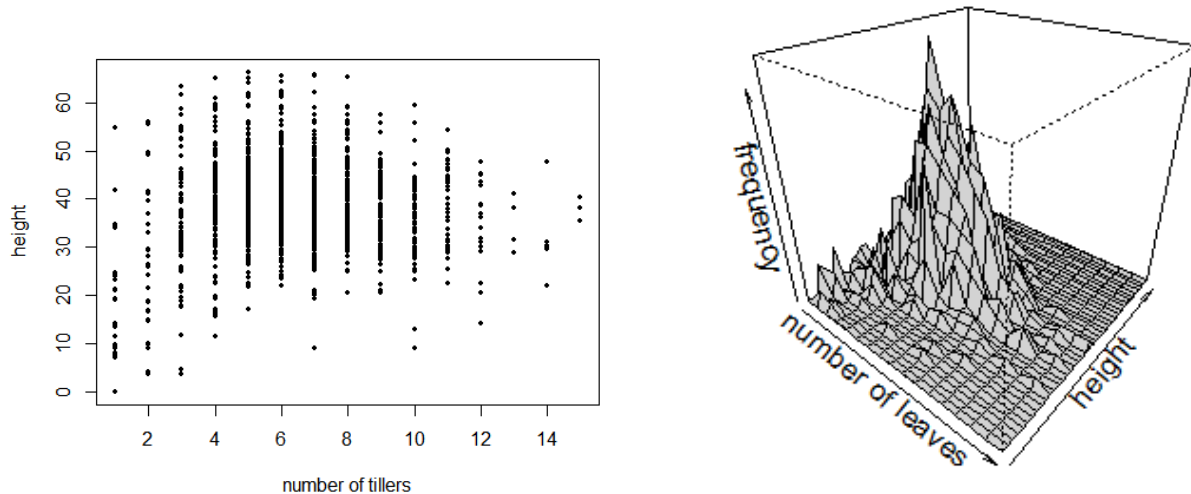

**Figure 2.** Plot of number of tillers vs. plant height in cm.

$$height = p_i + p_2 \times tillers + p_3 tillers^2 \quad Likelihood \sim N(\mu, \sigma)$$

We also recognize that there is a potential outlier beyond 16 tillers and decide to remove the data for this individual. Then, we create the quadratic values to evaluate our hypothesis and prepare an index to properly order the data in the plots.

```
max <- 17
reg_data$x<- reg_data$lf
reg_data$x2 <- reg_data$x^2
reg_data$y <- reg_data$lg
reg_data$x0 <- na.omit(reg_data$x)
ord <- order(reg_data$x0)
reg_data <-subset(reg_data,lf<max)
```

We evaluate two models. We use one additional predictor variable: whether there was one plant or two plants per pot (`tp`). We compare the models using WAIC. Notice that when you want to include an interaction with a quadratic variable, you must do it for both terms (x and x²).

```
m1 <- map2stan(
  alist(
    y ~ dnorm(mu,sigma),
    mu <- a + b1*x + b2*x2 + c*tp,
    a ~ dnorm(0,100),
    b1 ~ dnorm(0,10),
    b2 ~ dnorm(0,10),
    c ~ dnorm(0,10),
    sigma ~ dunif(0,10)
    ),
  data = reg_data,chains =3,
)
```
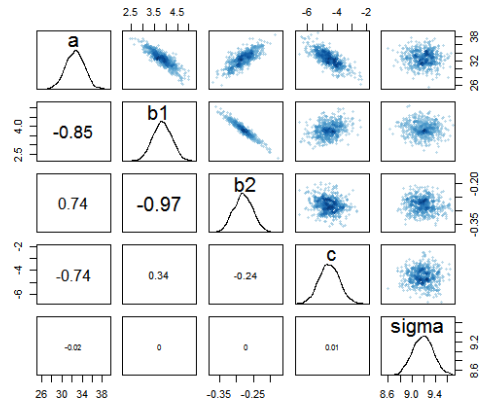


```
m2 <- map2stan(
    alist(
    y ~ dnorm(mu,sigma),
    mu <- a + b1*x +b2*x2 + c*tp + d1*x*tp + d2*x2*tp,
    a ~ dnorm(0,100),
    b1 ~ dnorm(0,10),
    b2 ~ dnorm(0,10),
    d1 ~ dnorm(0,10),
    d2 ~ dnorm(0,10),
    c ~ dnorm(0,10),
    sigma ~ dunif(0,10)
    ),
 data = reg_data,chains =3,
    )
```
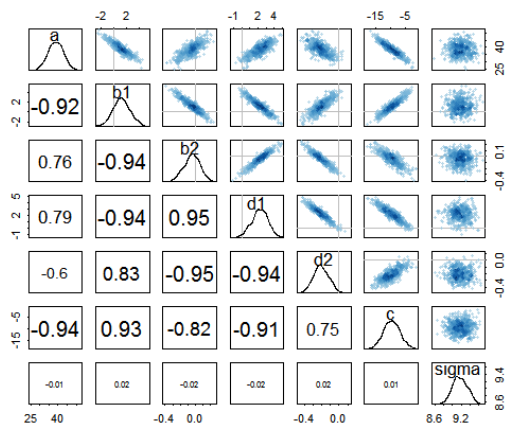

Notice the high correlations among the coefficients. We repeat the analysis this time scaling all the independent variables to allow for more efficient estimation of the coefficients.

```
reg_data$x_c <- scale(reg_data$x)
reg_data$x2_c <- scale(reg_data$x2)
reg_data$tp_c <- scale(reg_data$tp)
```
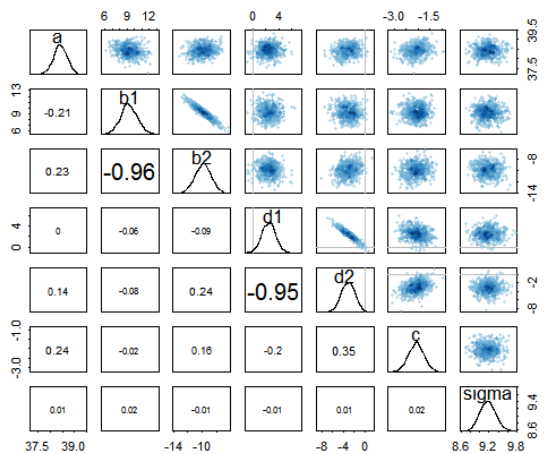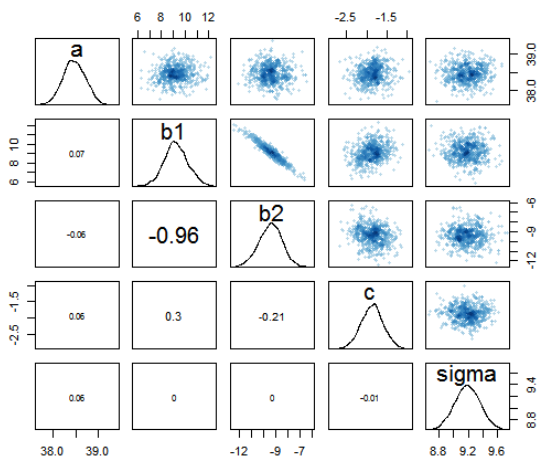
We use the function compare to evaluate their relative information.

```
         WAIC pWAIC dWAIC weight    SE   dSE
m2 10158.4   9.5   0.0   0.56 60.27   NA
m1 10158.9   6.3   0.5   0.44 60.21 6.46
```

The information in these two models was very similar. We use the simpler, the model with quadratic relationships between tillers and height specific to plants in pots with either one or more than one plant, but no interactions (Figure 3). Below is the summary of this model.

```
> precis(m1,digits=2)
        Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
a      38.47   0.25      38.08      38.87  2285    1
b1      9.22   1.01       7.65      10.88  1075    1
b2     -9.40   0.98     -10.99      -7.86  1124    1
c      -1.87   0.27      -2.31      -1.44  1939    1
sigma   9.19   0.17       8.91       9.47  2039    1
```

Next, we do a plot of our model (see code in the script). Figure 3 shows the data separated by the number of plants in the pot together with our two prediction lines and their 95 % credibility interval.
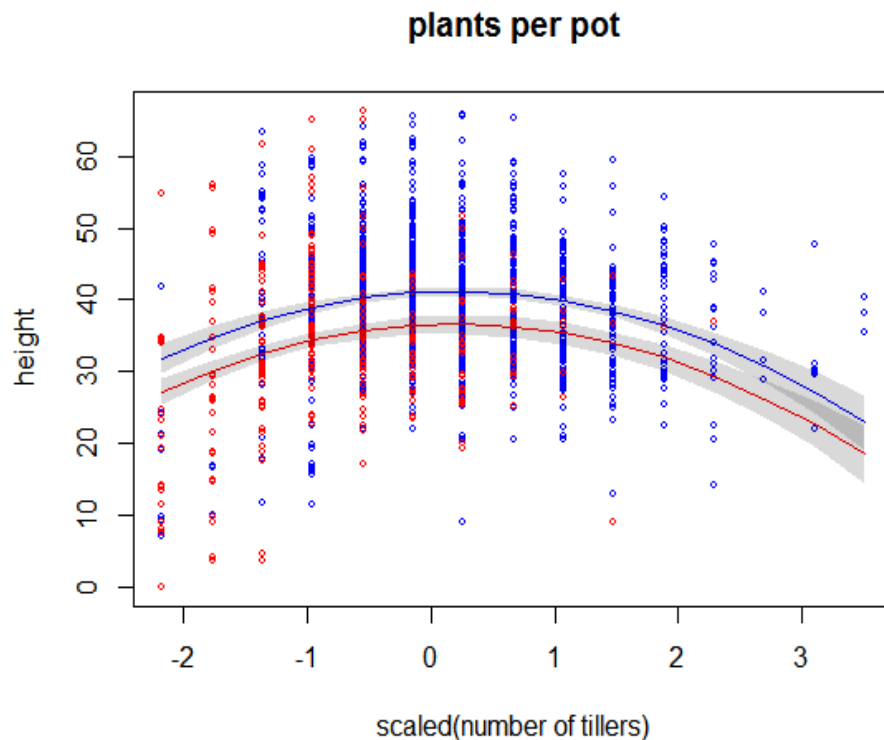
Figure 3. Plot of number of tillers vs. plant height in cm. Symbols in red are plants in pots with more than one plant per plot. In blue those in pots with one plant per pot. Notice the location of the mean number of tillers (0 in the x axis).

But what if we could get information from the literature to build informed priors? After all we must not be the first people to ever work with wiregrass! Gordon and Rice (1998) collected similar data on number of tillers and height of *Aristida beyrichiana* at seven different populations in Florida: Torreya State Park, Apalachicola Bluffs, Ravines Preserve, Wekiwa Springs (2 populations) and Rock Springs Run (2 populations - Figure 4). Researchers from the Jones Center in Georgia also collected similar data. They kindly allowed us to use their data (which is summarized in Table 2). See the script called `wg_priors.R` for how we got these figures.
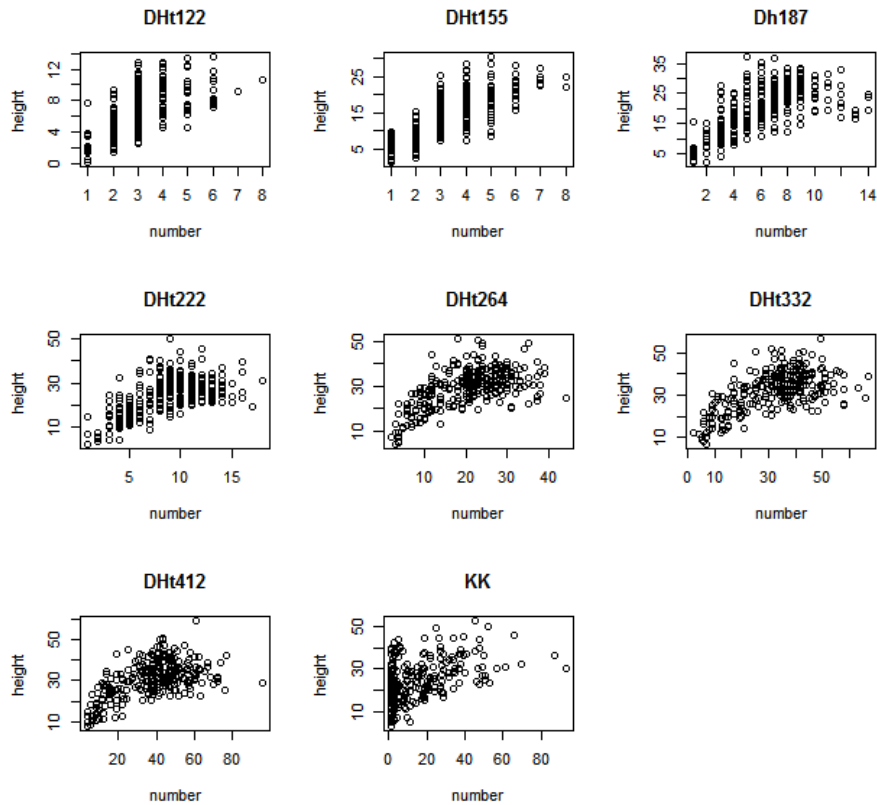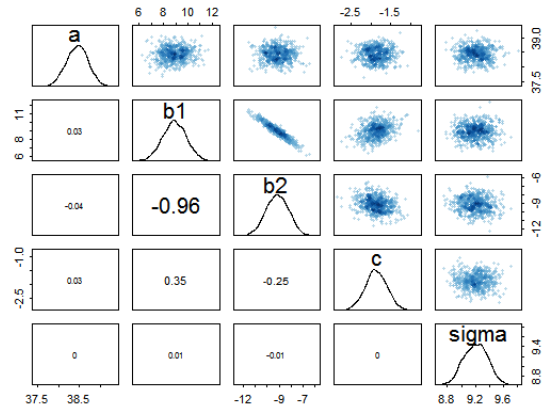

Figure 4. Plots of number of tillers vs. plant height in cm using data from Gordon and Rice (1998).

**Table 2.** Parameters, their mean and its standard errors calculated using available data.

	intercept	x	x^2	tp
Pop 1	6.69	4.02	-2.49	NA
Pop 2	14.52	8.00	-3.31	NA
Pop 3	20.21	16.21	-11.51	NA
Pop 4	24.62	16.72	-12.20	NA
Pop 5	29.43	17.23	-12.72	NA
Pop 6	32.44	18.15	-13.19	NA
Pop 7	31.13	15.75	-11.48	NA
KK	24.77	6.60	-2.59	NA
Mean	22.98	12.84	-8.69	NA
deviation standard	8.85	5.64	4.92	NA
Our study	38.46	8.64	-8.91	-1.89

We replace the priors in the above Bayesian Model with ones that reflect this new information as shown below (the rest of the code is the same). A plot of the realization of the Bayesian model with informative priors is shown in Figure 5, and a comparison among priors and posteriors in Figure 6.

```
# Informative priors
  a ~ dnorm(22.98,8.85),
  b1 ~ dnorm(12.84,5.64),
  b2 ~ dnorm(-8.69,4.92),
```



```
        Mean StdDev lower 0.89 upper 0.89 n_eff Rhat
a       38.45   0.25      38.05      38.87  2220    1
b1       8.90   0.96       7.45      10.52  1327    1
b2      -9.16   0.93     -10.59      -7.66  1336    1
c       -1.87   0.27      -2.30      -1.43  1722    1
sigma    9.20   0.17       8.95       9.47  2623    1
```
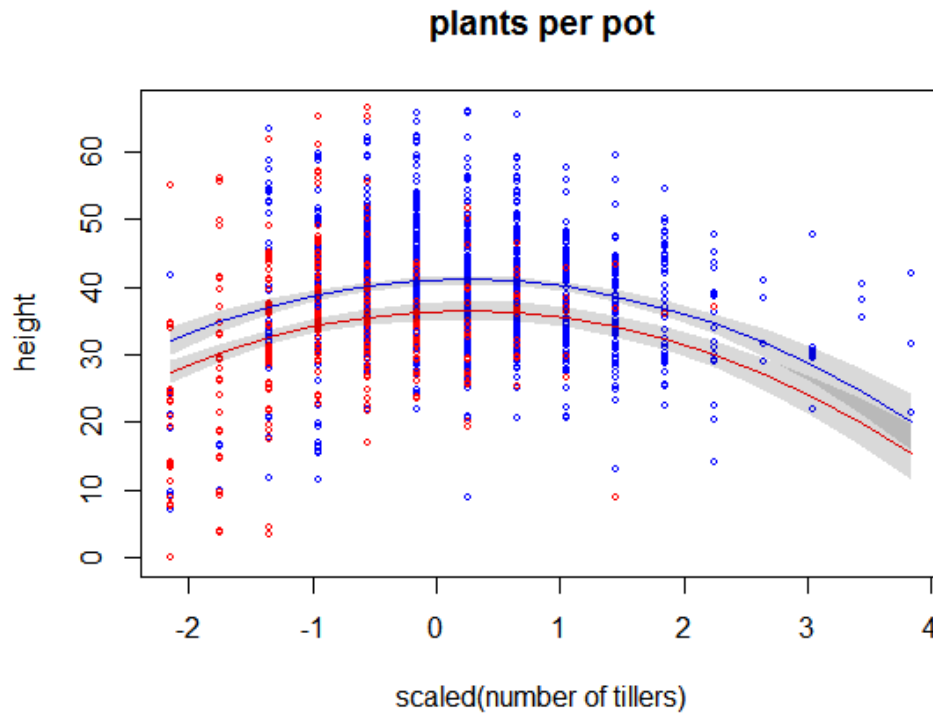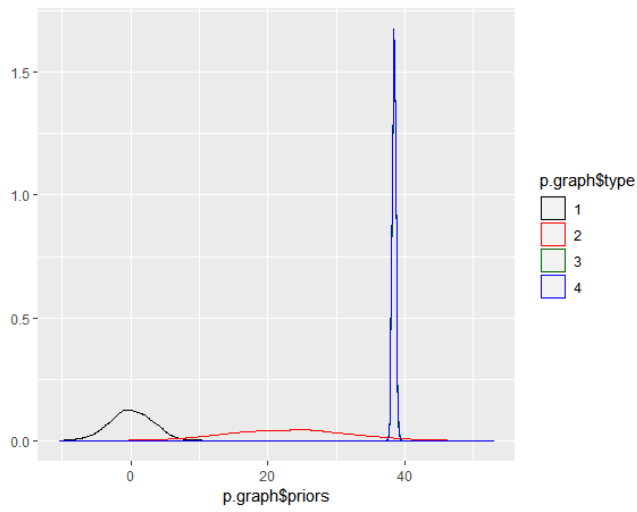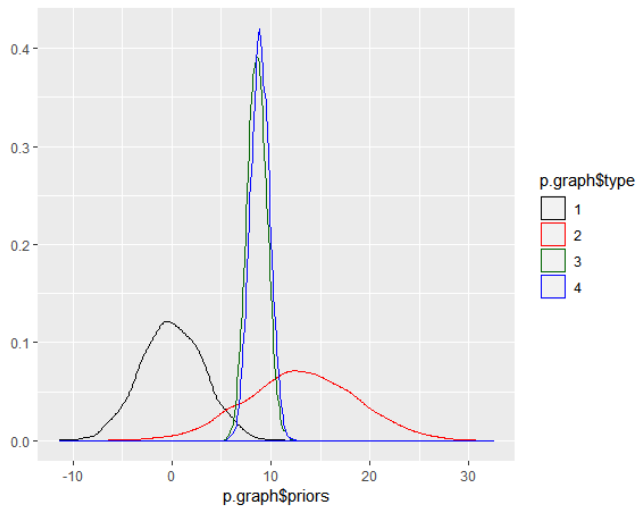

Figure 5. Plot of number of tillers vs. plant Height in cm. In light blue plants with more than one plant per pot. In dark blue plants with one plant per pot. Model with informative priors based on data from Gordon and Rice (1998).
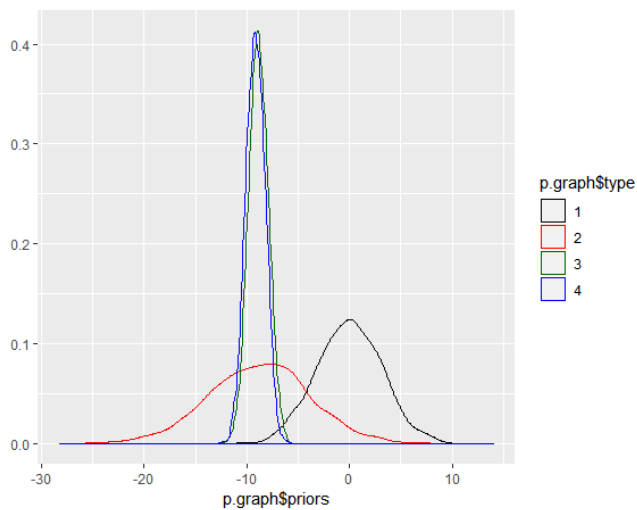
Intercept

Parameter b1

Parameter b2

Figure 6. Distribution of the diffuse (in black) and informed priors (in red) and distributions of the posteriors of models with diffuse (in green) and informed priors (in blue) for parameters β_0, β_1 and β_2.

NOTE: all the materials for this demo can be found at:

https://sciences.ucf.edu/biology/d4lab/methods-2/

References

Gordon, D.R. and K. J. Rice. 1998. Patterns of differentiation in Wiregrass (*Aristida beyrichiana*): implications for restoration efforts. Restoration Ecology 6: 166-174.

McElreath, R.M. 2016. Statistical Rethinking: a Bayesian course with examples in R and Stan. Chapman and Hall.