dplyr - working with data

filter() - pick observations by their values
arrange() - reorder rows
select() - pick columns by name
mutate() - create new variables from existing variables
summarise() - collapse values to a summary statistic
group_by() - all the above split by group

%>% pipe to combine

Base R: subset(), table(), aggregate(), |>

Tibbles – data frames with different display behavior

tibbles

Printing more of tibbles
?print.tbl > options

We want to inspect all the data by default:
options(tibble.width=Inf)
options(tibble.print_max=Inf)
options(max.print=1500) #dataset is < 1500 rows

You probably spent a lot of time collecting data. Wouldn’t you want to spend a few minutes to inspect each row?

dplyr vs base

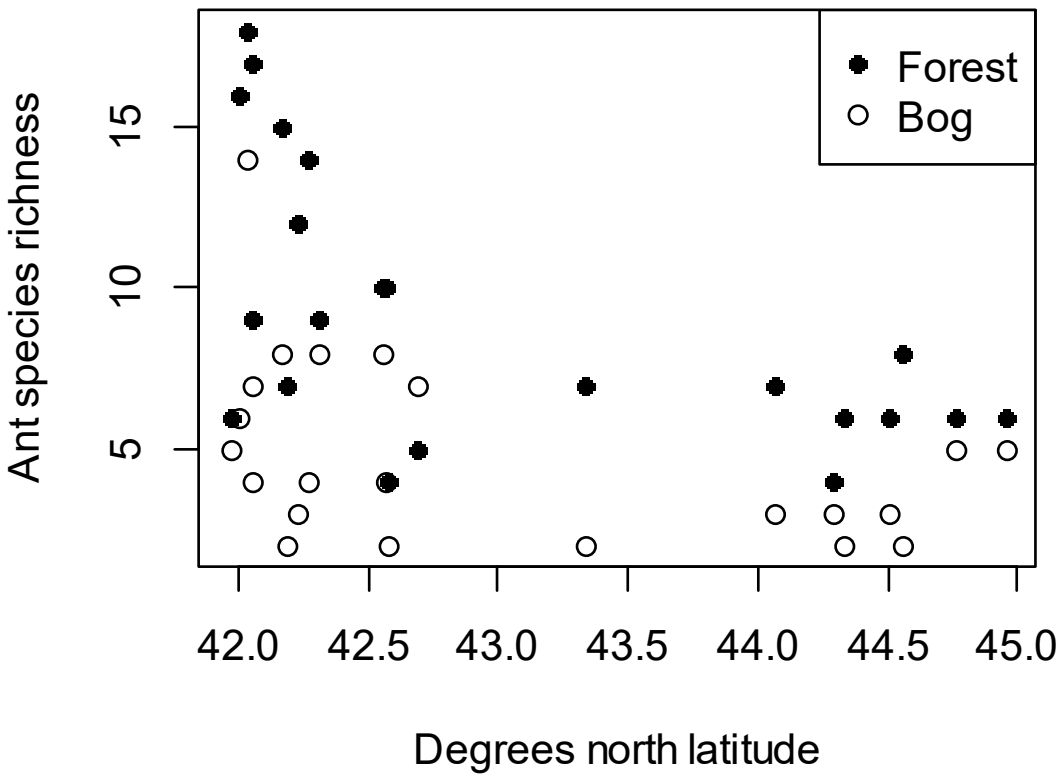
```
tree_dat %>%
    filter(status13==1) %>%
    filter(!is.na(mortality)) %>%
    mutate(diam_missing=is.na(diam13)) %>%
    summarize(sum(diam_missing))

sum(is.na(subset(tree_dat,status13==1 &
    !is.na(mortality))$diam13))
```

Independent project

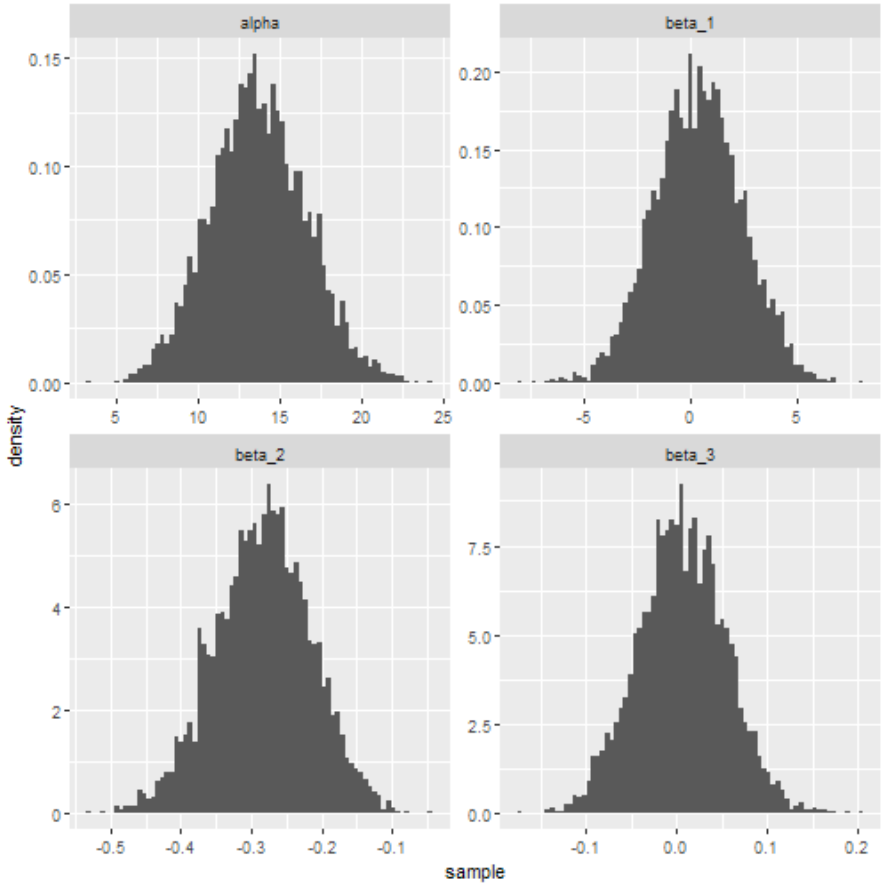
- Complete analysis (EDA through inference & conclusions)
- ggplot, dplyr
- Preferably hierarchical model:
  - rstanarm: stan_glmer or stan_lmer
- Submit .md from .R or .Rmd
- Due end of semester

Bayesian model

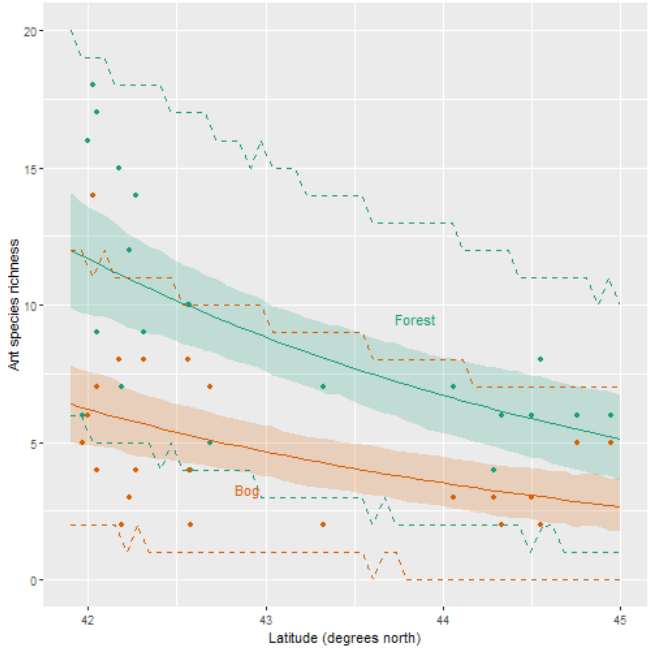


Could you get inferences?
Where did you have problems?

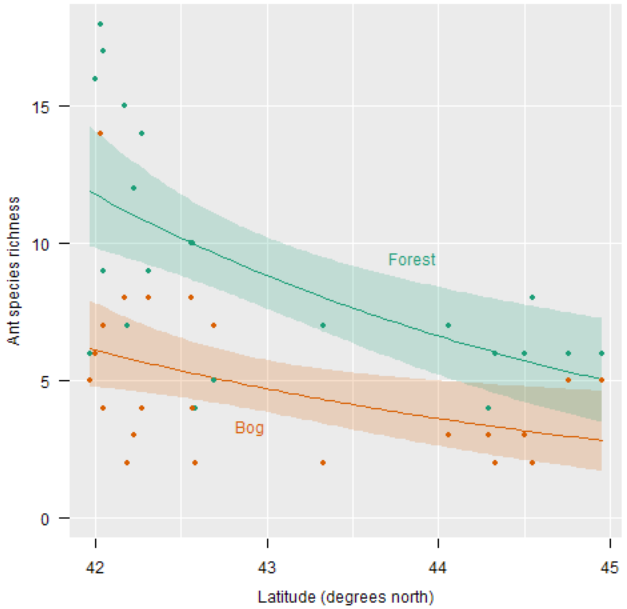
Bayesian model - ants



Bayesian model - ants



Bayesian



Frequentist
(w approx intervals)

Bayesian model

We discussed priors.

We discussed how samples are stored in a matrix by rows.

We discussed how to derive samples of new quantities (such as predicted y at values of x) that are combinations of the original samples of parameters.

See notes for the above.