dplyr - working with data

filter() - pick observations by their values
select() - pick columns by name
arrange() - reorder rows
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(), order(), sort(), 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)
```

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

How many trees with known status and mortality are missing a diameter in 2013?

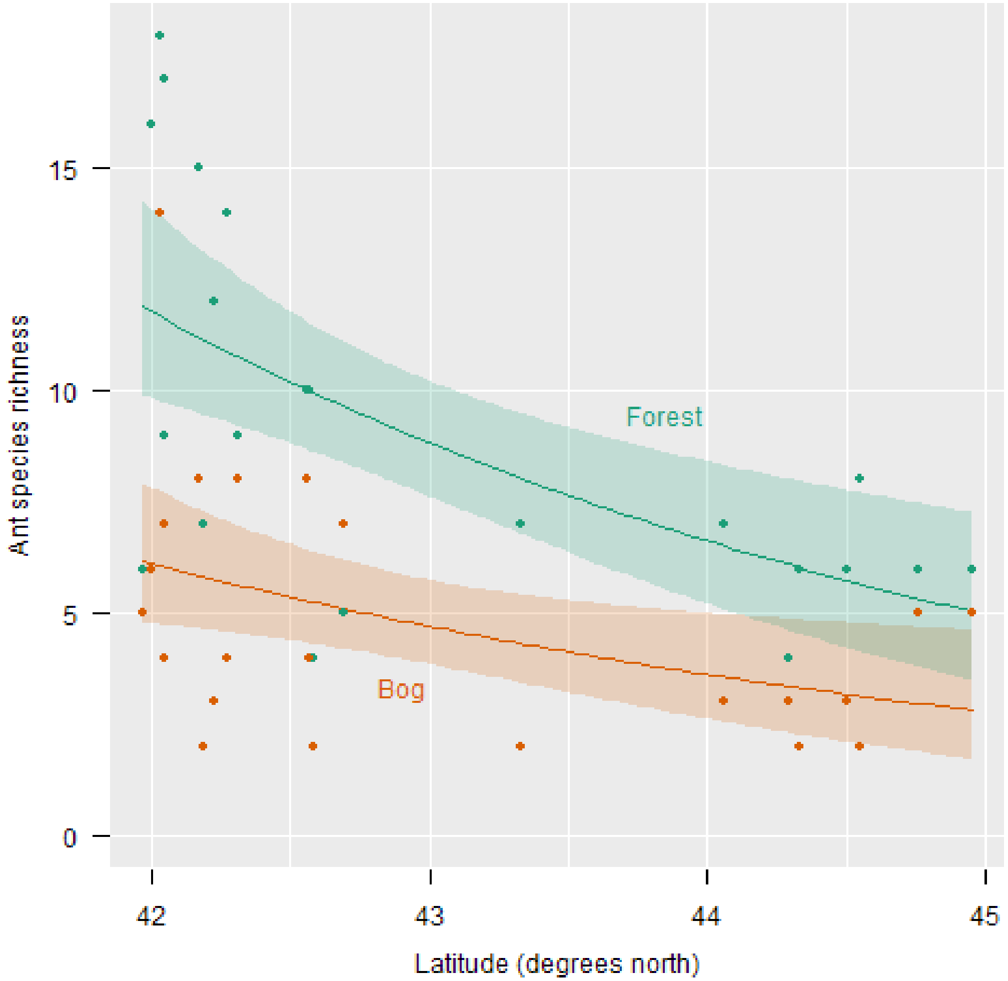
```
tree_dat %>%
    filter(status13==1) %>%
    filter(!is.na(mortality)) %>%
    mutate(diam_missing=is.na(diam13)) %>%
    summarize(sum(diam_missing))

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

GLM frequentist



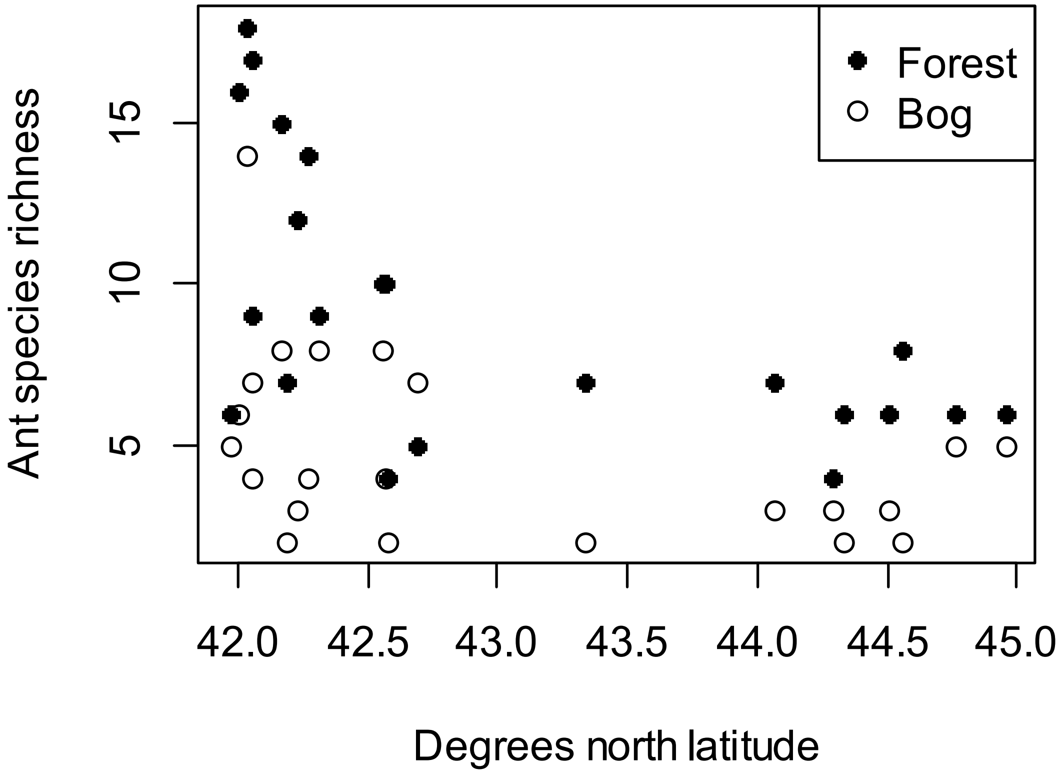
Scientific questions:

How different is species richness between habitats?

How does species richness vary with latitude?

Is this relationship different between habitats?

Bayesian model



Could you get inferences?
Where did you have problems?

Bayesian model - ants

- We started looking at the code
- Up to and including the summary output
- Next week we’ll look at priors and working with the samples