

Midterm - 2025

The data

Load the data:

```
library(rethinking)
library(tidyverse)
# Set the web address where R will look for files from this repository
repo_url <- "https://raw.githubusercontent.com/elahi/elahi.github.io/master/"
dat <- read.csv(paste(repo_url, "data/meps_2012_survey_data.csv", sep = ""))
glimpse(dat)
```

Rows: 72

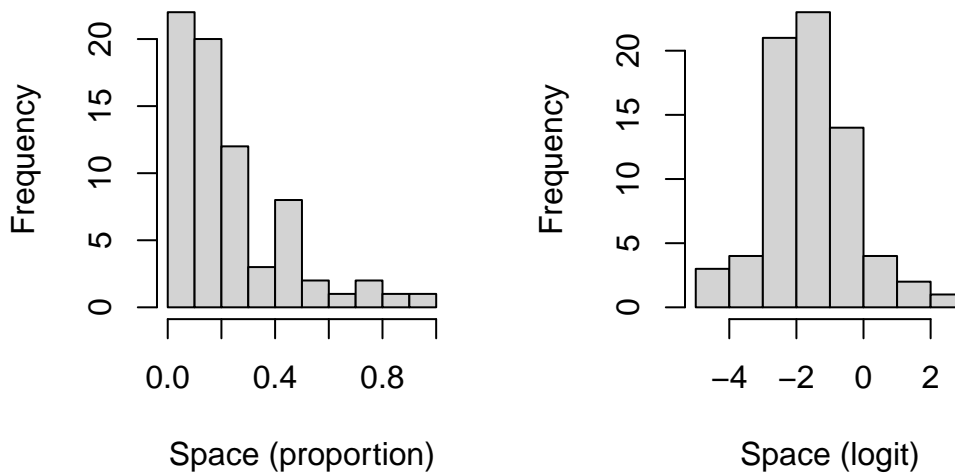
Columns: 9

```
$ X          <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ~
$ site       <chr> "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON", "ON"~
$ transect   <chr> "ON0", "ON0", "ON0", "ON0", "ON1", "ON1", "ON1", "ON1", "ON1", "ON1", "ON1"~
$ quadrat    <chr> "ON0_45", "ON0_46", "ON0_48", "ON0_49", "ON1_17", "ON1_84"~
$ chiton_mean <dbl> 0.0000000, 1.0000000, 0.0000000, 0.0000000, 1.0000000, 0.0~
$ space      <dbl> 0.14213721, 0.22878590, 0.12508222, 0.24288815, 0.73034557~
$ richness   <int> 10, 12, 8, 10, 10, 9, 9, 7, 8, 8, 9, 8, 10, 13, 10, 15, 13~
$ simpsondiv <dbl> 0.7004303, 0.7497803, 0.6160904, 0.7580468, 0.7172267, 0.7~
$ urchin_mean <dbl> 0.6666667, 0.6666667, 0.6666667, 0.6666667, 1.0000000, 1.0~
```

Transform the proportional cover of space so that we can model it using a normal distribution:

```
# Define logit function
logit <- function(p) log(p / (1 - p))
dat <- dat %>%
  mutate(space_logit = logit(space))
# Plot
```

```
par(mfrow = c(1,2))
hist(dat$space, main = "", xlab = "Space (proportion)")
hist(dat$space_logit, main = "", xlab = "Space (logit)")
```



Importantly, the response is no longer bounded between 0 and 1. It is *not* important that $\text{logit}(\text{space})$ appears normally distributed! It just so happens that this is the case here. For now, ignore the effect of site and transect.

Suggested workflow

You will skip the generative simulation for this exercise. Here is a suggested workflow:

1. Visualize the data to make sense of the patterns
2. Standardize the predictors
3. Come up with sensible priors by plotting prior predictive distributions (see HW4-key or SR2 Fig 5.8 for example). Justify your choice of priors.
4. Run the models using `ulam`. Diagnose your chains - what are the criteria you checked to ensure convergence? See SR2 9.4-9.5 for help.
5. Compare WAIC for the models.
6. Interpret in the context of the DAG and the published results.

7. Plot the counterfactual effect of one predictor, holding any other predictors constant. Justify your choice of the model used for these posterior predictions. See SR2 Fig. 5.9 for example of a counterfactual plot from a multiple regression. Interpret.