Interactions

EES 5891-03

Bayesian Statistical Methods

Jonathan Gilligan

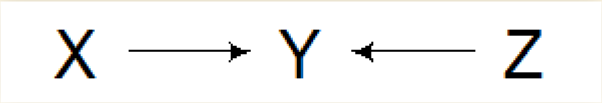
Class #10: Tuesday, September 27 2022

Interactions

- Multiple regression models may or may not include interactions:
  - Non-interacting models are *separable*: The effect of each variable is independent of all the other variables. \[\mu = \alpha + \beta_x x + \beta_x z \]
    - Doubling *x* has the same effect on \(\mu\) regardless what *z* is.
  - Interacting models are *non-separable*: The effect of one variable depends on another. \[\mu = \alpha + \beta_x x + \beta_z z + \beta_{xz} x z \]
    - The effect of doubling *x* depends on *z*.
    - Example:
      - Smoking increases lung cancer risks 10.3 times
      - Exposure to asbestos increases lung cancer risks 7.4 times in non-smokers.
      - Exposure to asbestos increases lung cancer risks 36.8 times in smokers.
        - If there was no interaction, the increase would be 17.7 times.

Interactions and DAGs

- A DAG can’t tell you about an interaction.



- This figure could represent a model with or without interactions: \[\begin{align} \mu &= \alpha + \beta_x x + \beta_z z \\ & \text{or} \\ \mu &= \alpha + \beta_x x + \beta_z z + \beta_{xz} x z \\ \end{align} \]
- The DAG says that *x* and *z* influence *y*, but it doesn’t say **how** they influence *y*.

Worked Example: Africa

Geography and Economy in Africa

- Is there a relationship between topographic roughness and per-capita GDP?

```
data(rugged)
d <- rugged
```

  - Columns include:

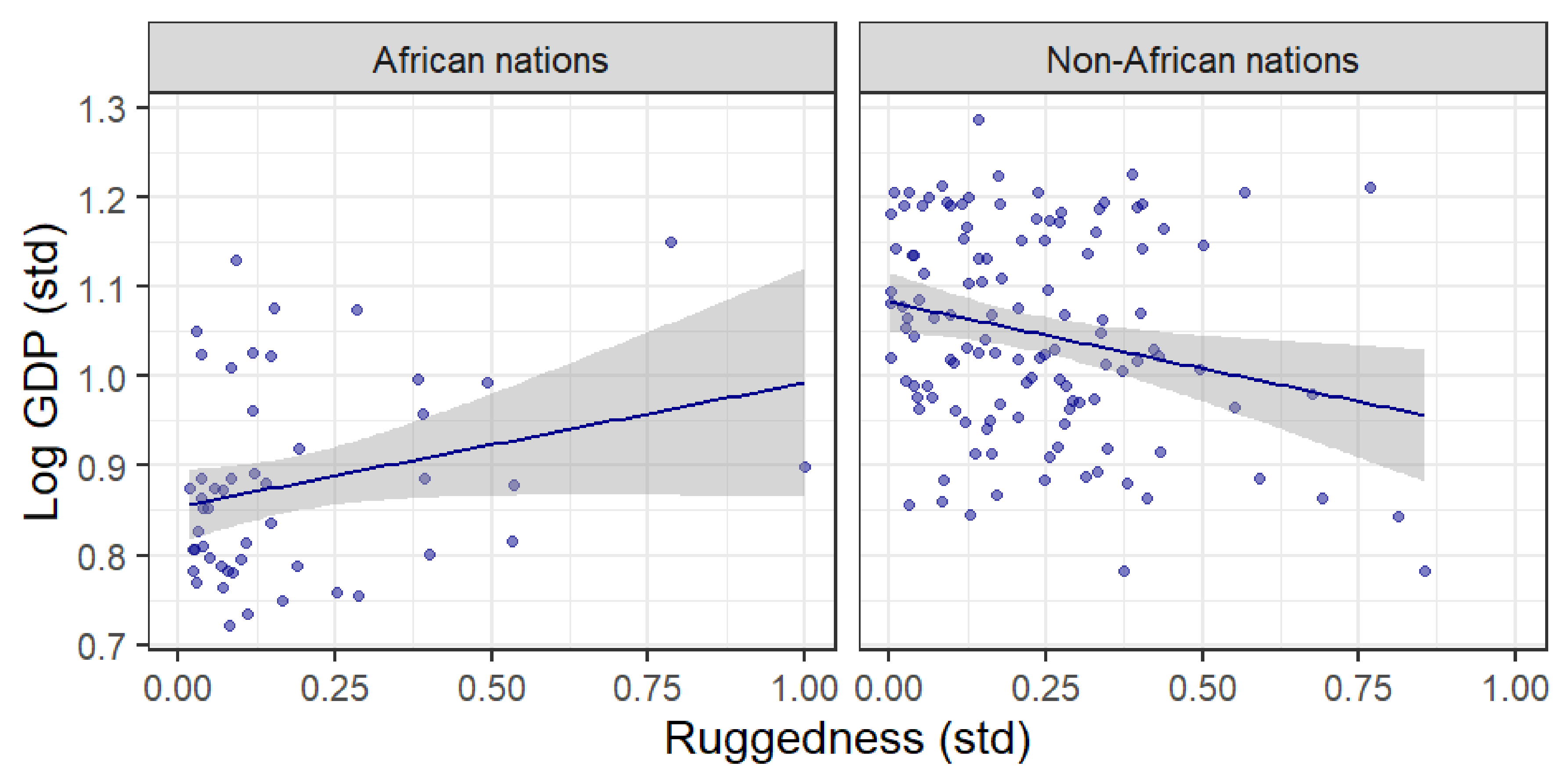
    rgdppc_2000
    GDP per-capita in 2000

    rugged
    Terrain ruggedness.

    cont_africa
    Indicator variable if continent is Africa

  - Transform data

```
dd <- d %>%
# drop missing values for GDP
filter(complete.cases(rgdppc_2000)) %>%
# Convert GDP to log-scale and standardize
mutate(
log_gdp = log(rgdppc_2000),
log_gdp_std = log_gdp / mean(log_gdp),
rugged_std = rugged / max(rugged)
)
```



```
dd %>%
  mutate(Africa = ifelse(cont_africa,"African nations",
                                     "Non-African nations")) %>%
  ggplot(aes(x = rugged_std, y = log_gdp_std)) +
  geom_point(size = 2, alpha = 0.5, color = "darkblue") +
  geom_smooth(method = "lm", color = "darkblue") +
  labs(x = "Ruggedness (std)", y = "Log GDP (std)") +
  facet_wrap(~Africa)
```

Model

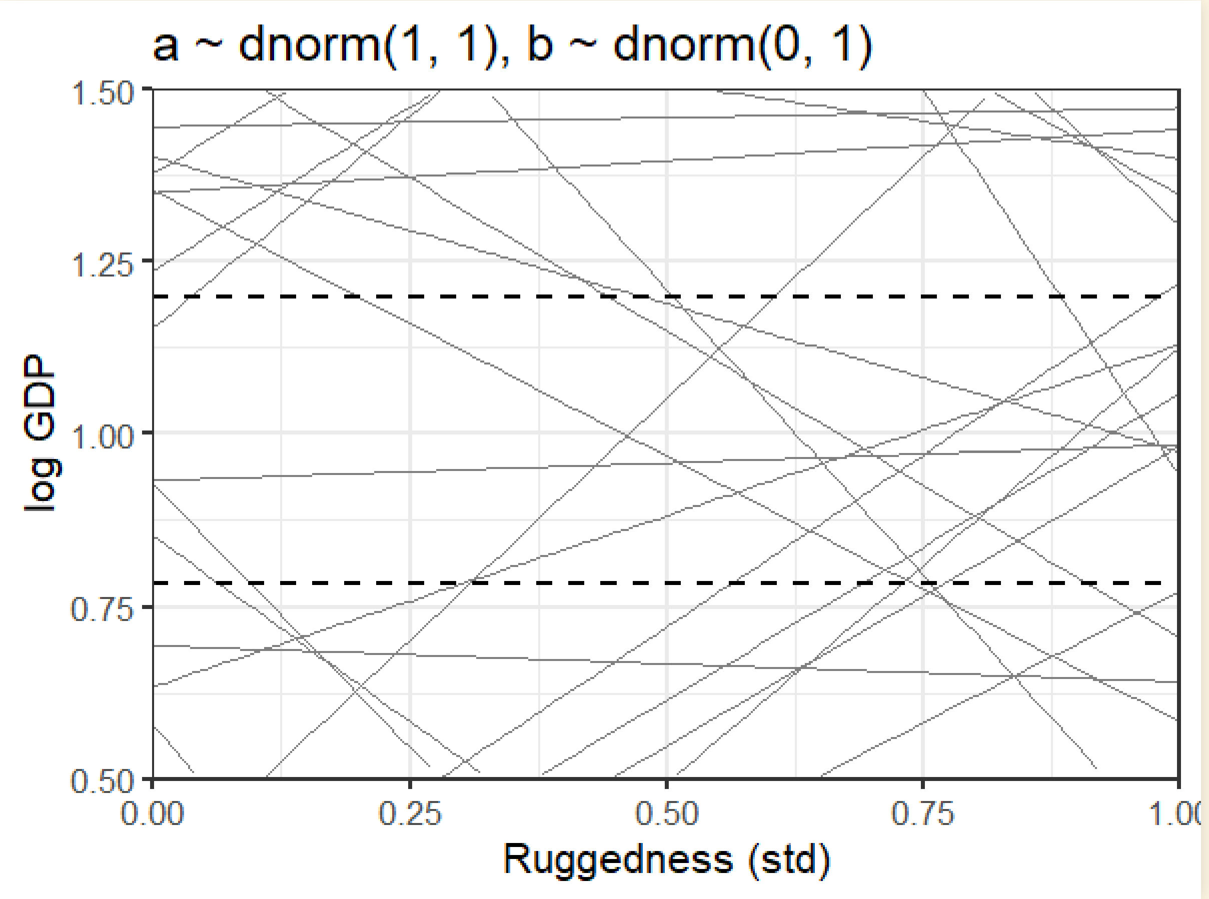
```
mdl_simple <- quap(
  alist(
    log_gdp_std ~ dnorm(mu , sigma),
    mu <- a + b * (rugged_std - 0.215),
    a ~ dnorm(1, 1),
    b ~ dnorm(0, 1),
    sigma ~ dexp(1)
  ), data = dd)
```

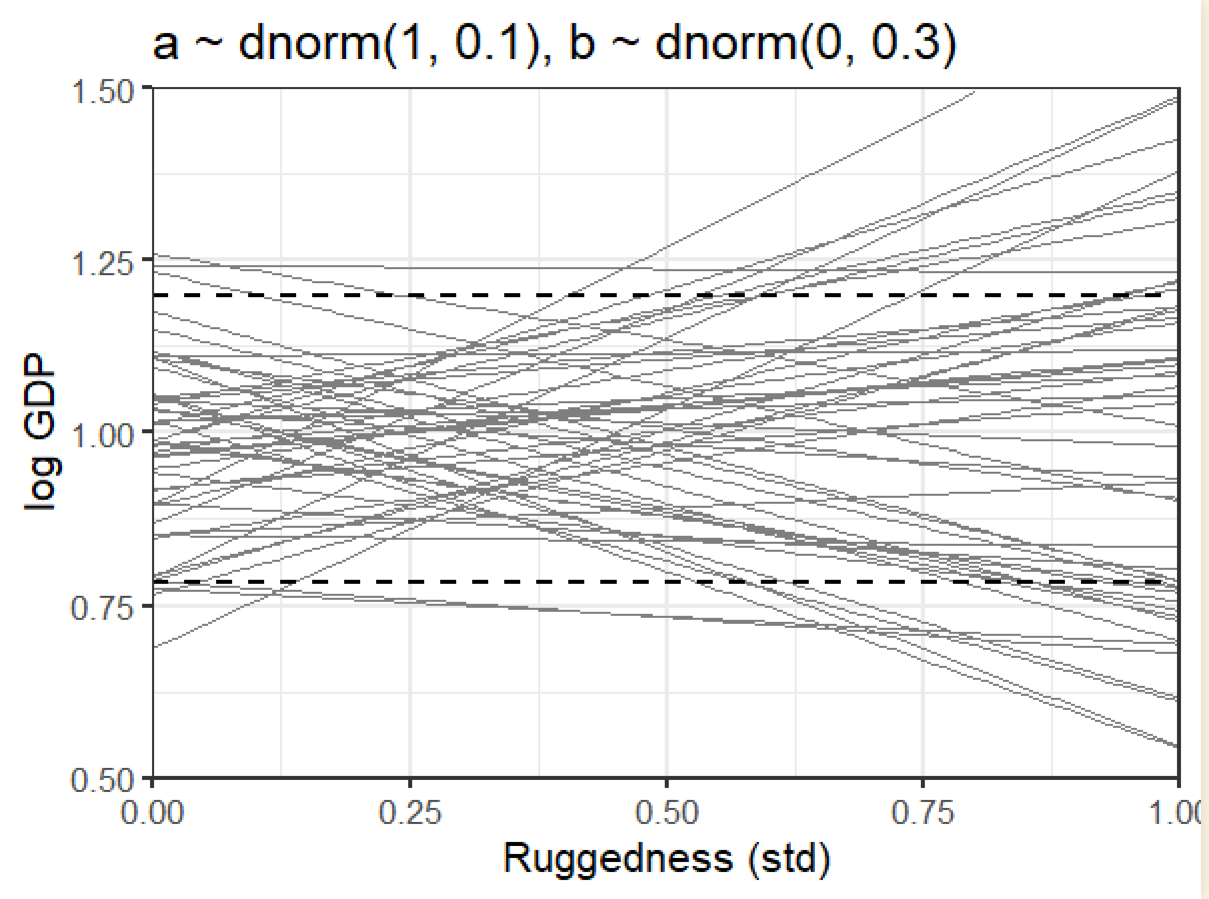
- The `0.215` is the average roughness for all countries.
- Now check the priors

```
prior <- extract.prior(mdl_simple, n = 50)
new_data <- tibble(rugged_std = seq(0, 1, 0.01))
prior_samples <- linpred_draws(
  mdl_simple, new_data, post = prior, value = "mu")
```

- Plot the priors

```
ggplot(prior_samples, aes(x = rugged_std, y = mu, group =
        .draw)) +
  geom_line(color = "gray50") +
  geom_hline(yintercept = quantile(dd$log_gdp_std, c(0.055,
        0.945)),
             linetype = "dashed", size = 1) +
  labs(x = "Ruggedness (std)", y = "log GDP", )
```





Examine the model

```
mdl_simple <- quap(
  alist(
    log_gdp_std ~ dnorm(mu , sigma),
    mu <- a + b * (rugged_std - 0.215),
    a ~ dnorm(1, 0.1),
    b ~ dnorm(0, 0.3),
    sigma ~ dexp(1)
  ), data = dd)
```

```
precis_show(precis(mdl_simple, digits = 2))
```

```
##        mean   sd  5.5% 94.5%
## a      1.00 0.01  0.98  1.02
## b      0.00 0.05 -0.09  0.09
## sigma  0.14 0.01  0.12  0.15
```

- Create an interaction:
  - Intercept depends on the continent
  - Create an indicator variable:

```
# Make an indicator variable
dd <- dd %>% mutate(cid = ifelse( dd$cont_africa==1 , 1 ,
        2 ))
```

```
mdl_inter <- quap(
    alist(
        log_gdp_std ~ dnorm(mu, sigma) ,
        mu <- a[cid] + b * (rugged_std - 0.215) ,
        a[cid] ~ dnorm(1, 0.1) ,
        b ~ dnorm(0, 0.3) ,
        sigma ~ dexp(1)
    ), data=dd)
```

```
compare(mdl_simple, mdl_inter) %>% round(2)
```

```
##                WAIC    SE dWAIC   dSE pWAIC weight
## mdl_inter  -252.28 15.31  0.00    NA  4.21      1
## mdl_simple -188.82 13.32 63.46 15.28  2.66      0
```

```
precis_show(precis(mdl_inter, depth = 2, digits = 2))
```

```
##         mean   sd  5.5% 94.5%
## a[1]    0.88 0.02  0.85  0.91
## a[2]    1.05 0.01  1.03  1.07
## b      -0.05 0.05 -0.12  0.03
## sigma   0.11 0.01  0.10  0.12
```

How does the model do?

Better Interactions

```
mdl_int_2 <- quap(
    alist(
        log_gdp_std ~ dnorm(mu, sigma) ,
        mu <- a[cid] + b[cid] * (rugged_std - 0.215) ,
        a[cid] ~ dnorm(1, 0.1) ,
        b[cid] ~ dnorm(0, 0.3) ,
        sigma ~ dexp(1)
    ), data=dd)
```

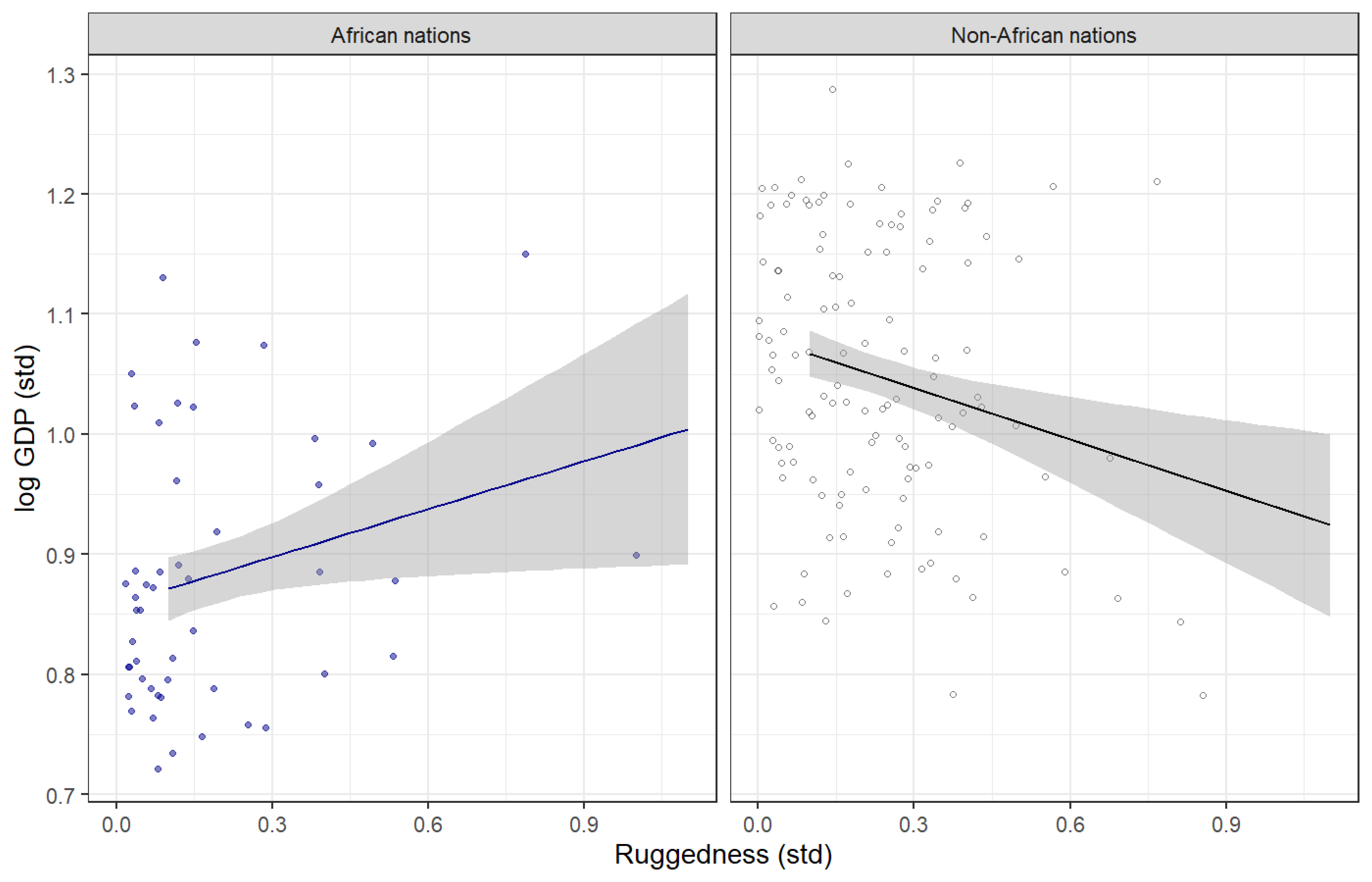
```
precis_show(precis(mdl_int_2, depth = 2, digits = 2))
```

```
##        mean   sd  5.5% 94.5%
## a[1]   0.89 0.02  0.86  0.91
## a[2]   1.05 0.01  1.03  1.07
## b[1]   0.13 0.07  0.01  0.25
## b[2]  -0.14 0.05 -0.23 -0.06
## sigma  0.11 0.01  0.10  0.12
```

```
compare(mdl_simple, mdl_inter, mdl_int_2) %>% round(2)
```

```
##               WAIC    SE dWAIC   dSE pWAIC weight
## mdl_int_2  -259.19 15.17  0.00    NA  5.13   0.97
## mdl_inter  -252.12 15.29  7.07  6.85  4.33   0.03
## mdl_simple -189.08 13.24 70.11 15.40  2.52   0.00
```

How does the model do?



Continuous Interactions

A Winter Flower

- Tulips grown in greenhouses
  - Soil
  - Light
  - Water

```
data(tulips)
d <- tulips
head(tulips)
```

```
##   bed water shade blooms
## 1   a     1     1   0.00
## 2   a     1     2   0.00
## 3   a     1     3 111.04
## 4   a     2     1 183.47
## 5   a     2     2  59.16
## 6   a     2     3  76.75
```

```
levels(tulips$bed)
```

```
## [1] "a" "b" "c"
```

```
d <- d %>% mutate(
  blooms_std = blooms / max(blooms),
  water_cent = water - mean(water),
  shade_cent = shade - mean(shade)
  )
```

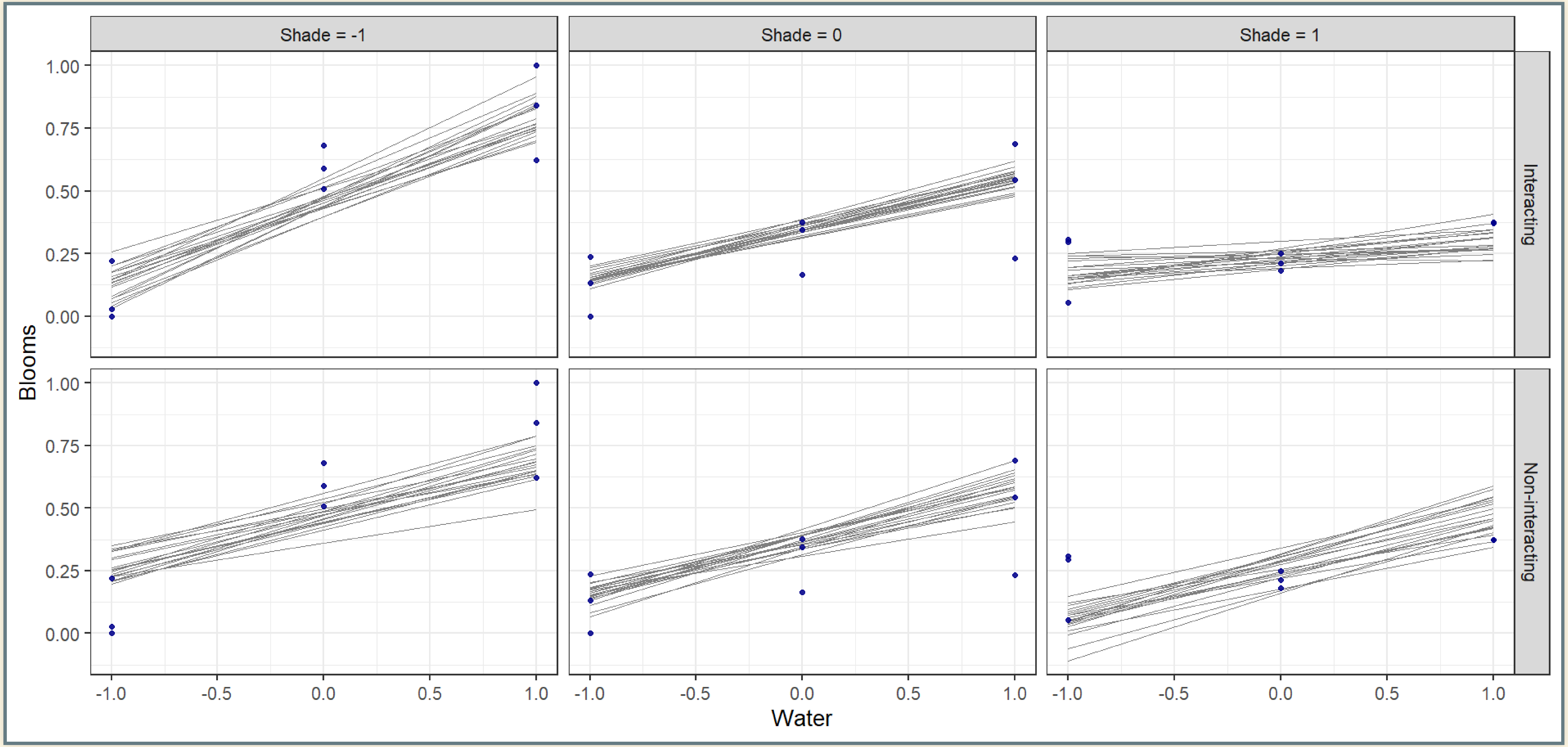
- Predict `blooms` from `water`, and `shade`.
- Two models:
  - **Non-interacting:**

```
mdl_tulip_non <- quap(
  alist(
    blooms_std ~ dnorm(mu, sigma),
    mu <- a + bw * water_cent + bs * shade_cent,
    a ~ dnorm(0.5, 0.25),
    bw ~ dnorm(0, 0.25),
    bs ~ dnorm(0, 0.25),
    sigma ~ dexp(1)
  ), data=d)
```

  - **Interacting:**

```
mdl_tulip_inter <- quap(
  alist(
    blooms_std ~ dnorm(mu, sigma) ,
    mu <- a + bw * water_cent + bs * shade_cent +
          bws * water_cent * shade_cent,
    a ~ dnorm(0.5, 0.25),
    bw ~ dnorm(0, 0.25),
    bs ~ dnorm(0, 0.25),
    bws ~ dnorm(0, 0.25),
    sigma ~ dexp(1)
  ), data=d)
```

Interpreting the models


Shade = -1
Shade = 0
Shade = 1
Interacting
Non-interacting
Blooms
Water