EEEB UN3005/GR5005
Homework - Week 10 - Due 09 Apr 2019

USE YOUR NAME HERE

Homework Instructions: Complete this assignment by writing code in the code chunks provided. If required, provide written explanations below the relevant code chunks. Replace "USE YOUR NAME HERE" with your name in the document header. When complete, knit this document within RStudio to generate a pdf. Please review the resulting pdf to ensure that all content relevant for grading (i.e., code, code output, and written explanations) appears in the document. Rename your pdf document according to the following format: hw_week_10_firstname_lastname.pdf. Upload this final homework document to CourseWorks by 5 pm on the due date.

All the following homework problems will draw on the country-level dataset (`rugged`) that was discussed in the *Statistical Rethinking* book and lecture. In particular, we'll be interested in an African island nation, Seychelles, and how inclusion of data from this one country might affect our statistical inference.

Problem 1 (3 points)

Following the lecture code, import the `rugged` dataset, create a logged version of the year 2000 GDP variable (for use as our outcome variable), and subset the data down to only those countries that actually have GDP data.

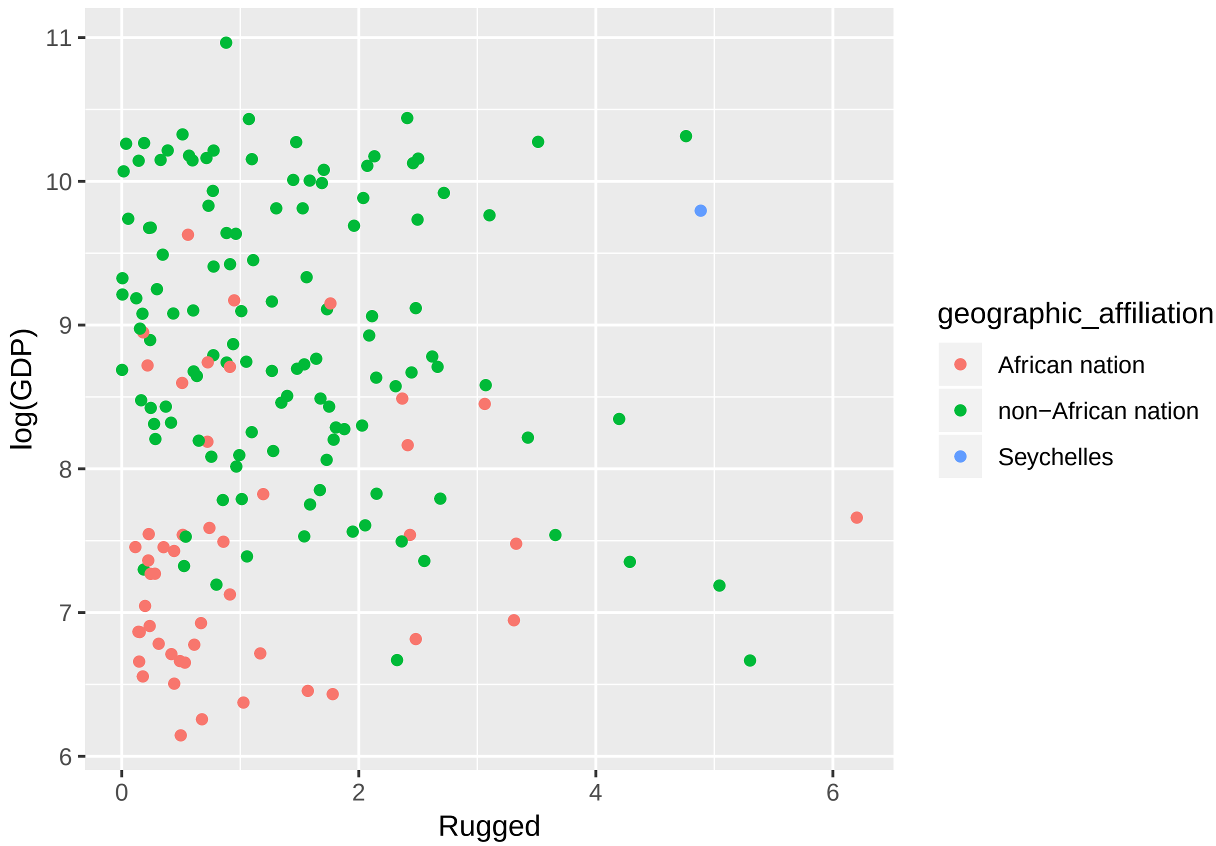
Now for something new: because we are interested in Seychelles, we'd like to visualize where Seychelles stands in relation to other African (and non-African countries). Therefore, using a method of your choice, create a new variable in the `rugged` data frame called `geographic_affiliation`. `geographic_affiliation` should have the value of "non-African nation" anywhere `cont_africa == 0`. Similarly, `geographic_affiliation` should have the value of "African nation" anywhere `cont_africa == 1` EXCEPT when `country == "Seychelles"`. There, `geographic_affiliation` should have a value of "Seychelles".

Using the `ggplot()` function, visualize the relationship between `rugged` (x-axis) and log GDP (y-axis) using a scatterplot. Assign the color of the points to `geographic_affiliation`. If all has gone correctly, you should end up with points of three different colors, corresponding to "African nation", "non-African nation", and "Seychelles".

Where does the GDP of Seychelles lie relative to most other African countries? Where does the terrain ruggedness of Seychelles lie relative to most other African countries?

```
data(rugged)
d = rugged[complete.cases(rugged$rgdppc_2000),]
d$logGDP = log(d$rgdppc_2000)
d$geographic_affiliation =
```

```r
    ifelse(d$country == 'Seychelles', 'Seychelles',
    ifelse(d$cont_africa == 0, 'non-African nation', 'African nation'))
graph = ggplot(d, aes(x = rugged, y = logGDP, color = geographic_affiliation)) +
    geom_point() +
    xlab('Rugged') +
    ylab('log(GDP)')
plot(graph)
```



Answer:

Its GDP is the highest of all African countries, but its ruggedness is also higher than most of African countries.

Problem 2 (4 points)

Now replicate the interaction model as given in lecture (m7.5b) using a dataset that excludes Seychelles. In addition, re-fit model m7.5b as we did in lecture, using the full dataset (by "full dataset" I mean the `rugged` dataset with all countries that have GDP data). Compare these two models using `precis()` to show the 97% PIs of model parameters. Interpret the change you see in the bAR parameter (the interaction term) in your new fit model relative to the estimate derived from m7.5b.

```
m.exclude = map(
    alist(
        logGDP ~ dnorm(mu, sigma),
        mu <- a + bR * rugged + bA * cont_africa + bAR * rugged * cont_africa,
        a ~ dnorm(8, 100),
        bR ~ dnorm(0, 1),
        bA ~ dnorm(0, 1),
        bAR ~ dnorm(0, 1),
        sigma ~ dunif(0, 10)),
    data = d[d$country != 'Seychelles',])
m.full = map(
    alist(
        logGDP ~ dnorm(mu, sigma),
        mu <- a + bR * rugged + bA * cont_africa + bAR * rugged * cont_africa,
        a ~ dnorm(8, 100),
        bR ~ dnorm(0, 1),
        bA ~ dnorm(0, 1),
        bAR ~ dnorm(0, 1),
        sigma ~ dunif(0, 10)),
    data = d)
precis(m.exclude, 0.97)
```

```
##         Mean StdDev  5.5% 94.5%
## a       9.19   0.14  8.97  9.40
## bR     -0.19   0.08 -0.31 -0.07
## bA     -1.78   0.22 -2.13 -1.43
## bAR     0.25   0.14  0.04  0.47
## sigma   0.93   0.05  0.85  1.01
```

```
precis(m.full, 0.97)
```

```
##         Mean StdDev  5.5% 94.5%
## a       9.18   0.14  8.97  9.40
## bR     -0.18   0.08 -0.31 -0.06
## bA     -1.85   0.22 -2.20 -1.50
## bAR     0.35   0.13  0.14  0.55
## sigma   0.93   0.05  0.85  1.01
```

Answer:

Since Seychelles is a exception of "more rugged, less GDP", if we exclude this exception, the African data will fit "more rugged, less GDP" better.

In the model, the interaction parameter b_{AR} increase the slope of GDP ~ Rugged in African countries:

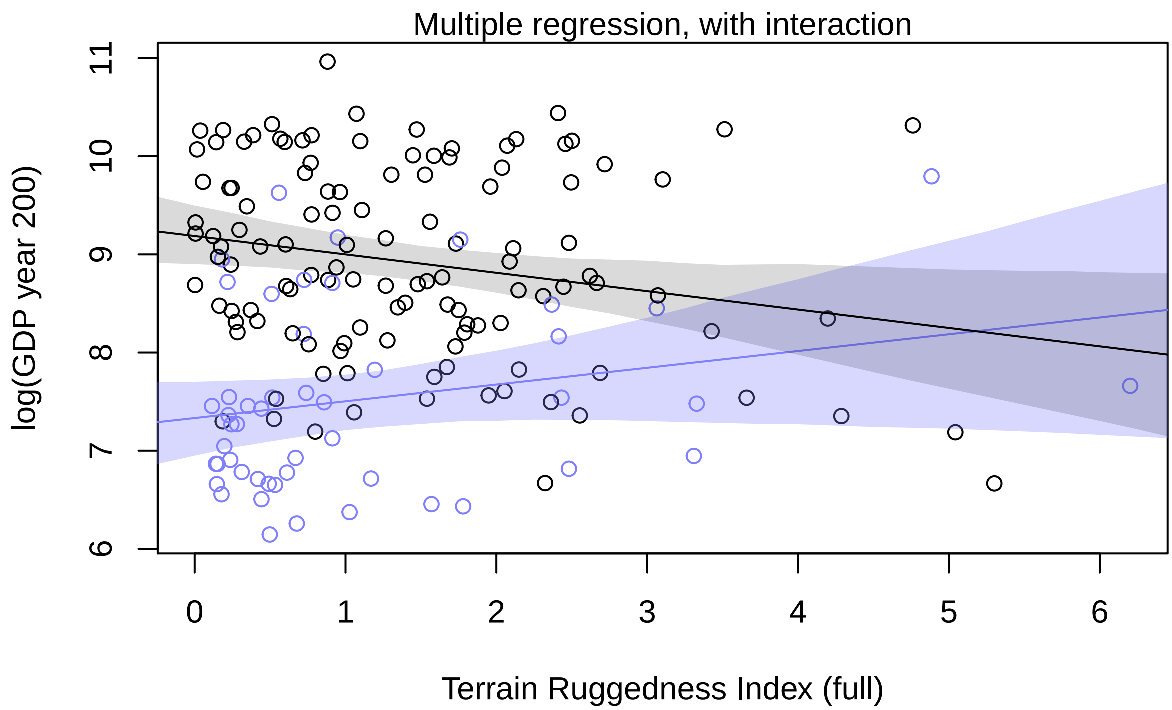
$$\log(\text{GDP}) = \begin{cases} (a + b_A) + (b_R + b_{AR})R \\ a + b_R R \end{cases}$$

After excluding Seychelles, African data is less positively correlated with ruggedness, which means the slope is smaller. So, b_{AR} is smaller.

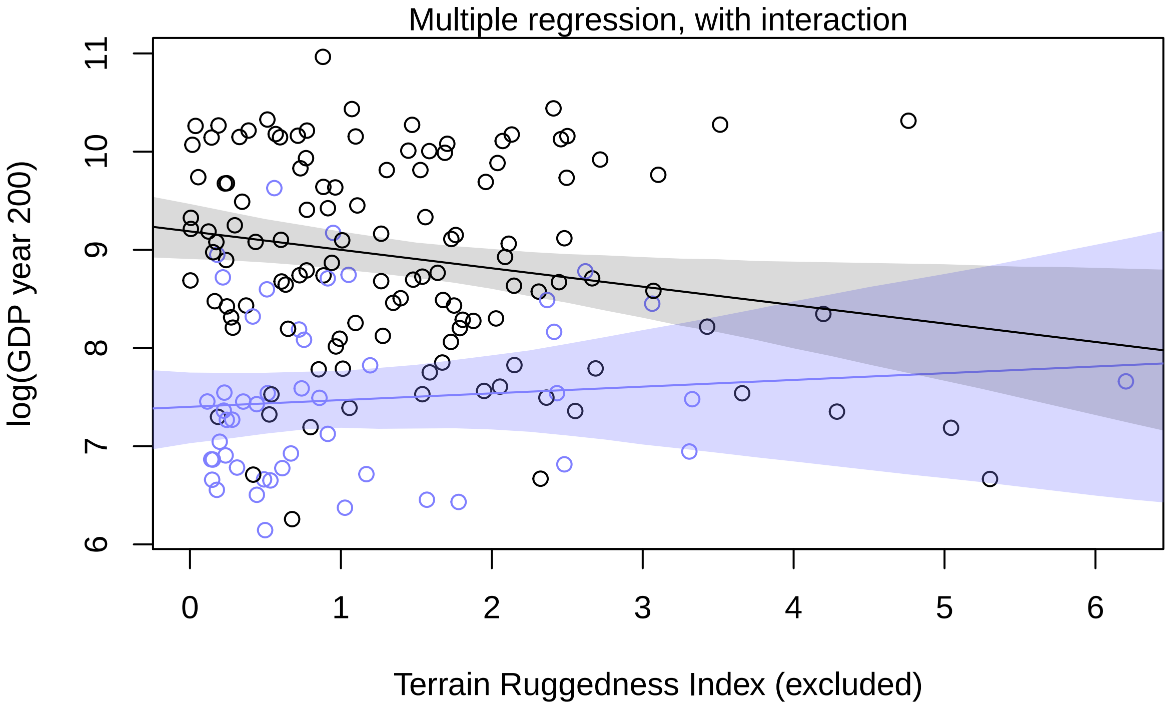
Problem 3 (3 points)

Using the lecture code as a guide, plot model-based predictions for both m7.5b and your new model that excludes Seychelles. For a given model, you can choose to show predictions for the ruggedness effect inside and outside of Africa in two separate panels or together on one plot. Both methods were demonstrated in lecture.

```
rugged.seq <- seq(from = -1, to = 8, by = 0.25)

mu.Africa <- link(m.full,
    data = data.frame(cont_africa = 1, rugged = rugged.seq))
mu.Africa.mean <- apply(mu.Africa, 2, mean)
mu.Africa.PI <- apply(mu.Africa, 2, PI, prob = 0.97)

mu.NotAfrica <- link(m.full,
    data = data.frame(cont_africa = 0, rugged = rugged.seq))
mu.NotAfrica.mean <- apply(mu.NotAfrica, 2, mean)
mu.NotAfrica.PI <- apply(mu.NotAfrica, 2, PI, prob = 0.97)

plot(logGDP ~ rugged, data = d,
     xlab = "Terrain Ruggedness Index (full)",
     ylab = "log(GDP year 200)",
     col = ifelse(d$cont_africa == 1, rangi2, "black")
)
mtext("Multiple regression, with interaction", 3)
lines(rugged.seq, mu.Africa.mean, col = rangi2)
shade(mu.Africa.PI, rugged.seq,
      col = col.alpha(rangi2, 0.3))
lines(rugged.seq, mu.NotAfrica.mean)
shade(mu.NotAfrica.PI, rugged.seq)
```




```
mu.Africa <- link(m.exclude,
    data = data.frame(cont_africa = 1, rugged = rugged.seq))
mu.Africa.mean <- apply(mu.Africa, 2, mean)
mu.Africa.PI <- apply(mu.Africa, 2, PI, prob = 0.97)

mu.NotAfrica <- link(m.exclude,
    data = data.frame(cont_africa = 0, rugged = rugged.seq))
mu.NotAfrica.mean <- apply(mu.NotAfrica, 2, mean)
mu.NotAfrica.PI <- apply(mu.NotAfrica, 2, PI, prob = 0.97)

plot(logGDP ~ rugged, data = d[d$country != 'Seychelles',],
     xlab = "Terrain Ruggedness Index (excluded)",
     ylab = "log(GDP year 200)",
     col = ifelse(d$cont_africa == 1, rangi2, "black")
)
mtext("Multiple regression, with interaction", 3)
lines(rugged.seq, mu.Africa.mean, col = rangi2)
shade(mu.Africa.PI, rugged.seq,
      col = col.alpha(rangi2, 0.3))
lines(rugged.seq, mu.NotAfrica.mean)
shade(mu.NotAfrica.PI, rugged.seq)
```


Multiple regression, with interaction
log(GDP year 200)
11
10
9
8
7
6
0
1
2
3
4
5
6
Terrain Ruggedness Index (excluded)