HW 4

DO NOT INCLUDE NAMES - Just add names in gradescope

# Rules

In groups of 2, 3, or 4, complete the following. Individual submissions are not allowed and will get a 0 without prior approval to work alone.

# Instructions

For this homework, you will use the tools we’ve been discussing to do the work that came into a statistical draft report I provide you (Demonstrationreport\_2024.docx). This report is based on work of students in a previous semester but has been heavily edited. Each of the highlighted or changed colored fonts in the draft report are noted for you to either edit the text or insert the described figures, tables, or citations. It is your task to figure what to do based on the written report (e.g., what model to fit, what sorts of plots to make, etc.).

The work relates to analyzing the data set from Que et al. (2017) but excludes the Indus river observations that they mentioned.

Edit the provided Word document to submit the report portion with your modifications. There are four sentences you need to edit based on the results. And you need to complete the References, Figures with captions (copy the figures from the knitted work), and Tables with captions (copy from knitted work) sections of the report based on your work.

Then you must combine the report word document that contains the report with references, figures, and tables (only) with the knitted version of this document where you will create the results to finish the report. You have access to Adobe Acrobat and can use it to combine PDF files.

In gradescope, identify the page of the edited sentences and added citations, figures, and table, and knitted Rmd pages.

library(readxl)  
riverdata <- read\_excel("Qu\_etal\_2017.xlsx")  
riverdataR <- riverdata %>% dplyr::filter(River != "Indus") %>%  
 mutate(logpCO2 = log(pCO2))

enhanced\_stripchart(pCO2 ~ River, data = riverdataR) + labs(y = "Partial CO2 pressure (\u03BCatm)")

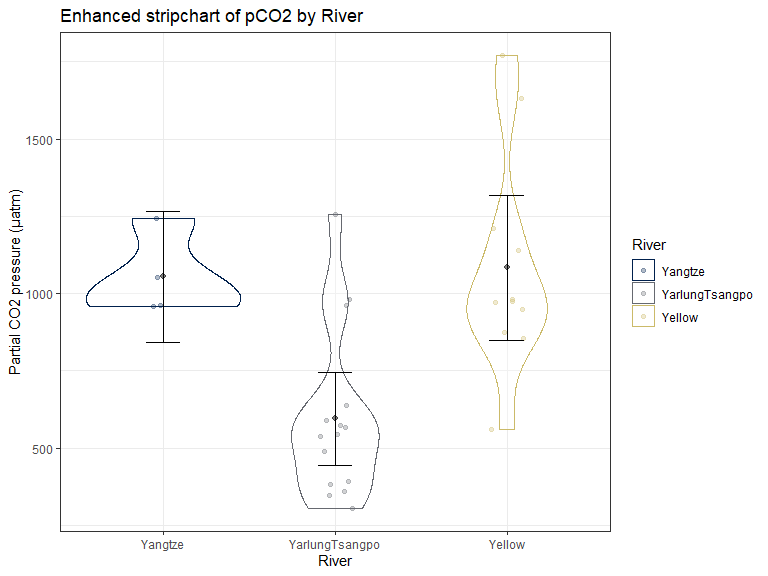


Table 1: Summary stats for pCO2 and elevation by river

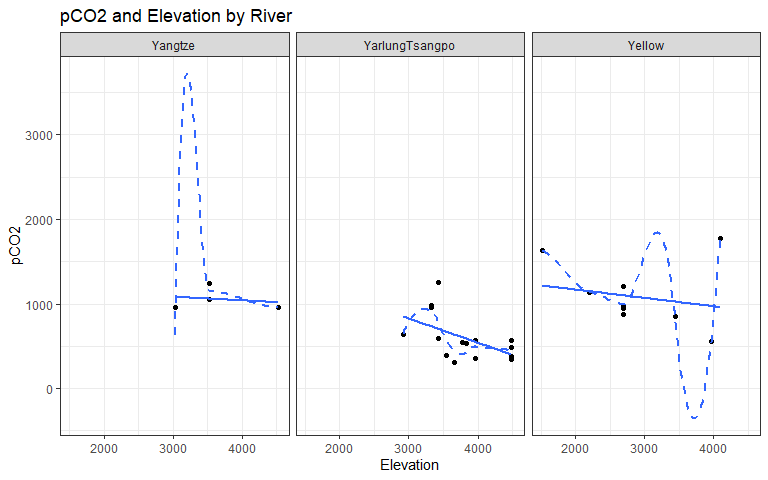
datasummary\_balance(Elevation + pCO2 ~ River, data = riverdataR, output = 'markdown',  
 title = 'Elevation + pCO2 ~ River')

Elevation + pCO2 ~ River

|  | Yangtze (N=4) | | YarlungTsangpo (N=15) | | Yellow (N=11) | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Std. Dev. | Mean | Std. Dev. | Mean | Std. Dev. |
| Elevation | 3652.8 | 627.1 | 3804.5 | 498.2 | 2848.5 | 739.7 |
| pCO2 | 1054.2 | 133.4 | 595.4 | 271.1 | 1083.4 | 348.3 |

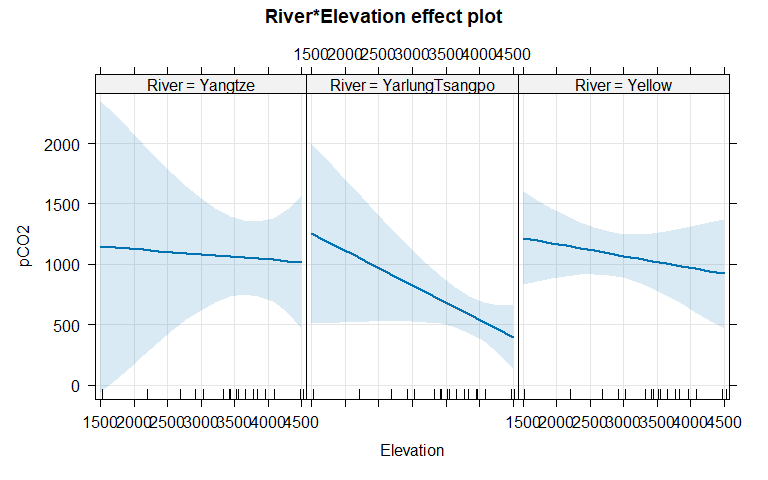
Figure 2: Scatter plot of pCO2 and elevation by river

ggplot(riverdataR, aes(x = Elevation, y = pCO2))+  
 geom\_point()+  
 facet\_wrap(~River)+  
 geom\_smooth(method = 'lm' , se = F)+  
 geom\_smooth(se = F, lty = 2)+  
 labs(title = 'pCO2 and Elevation by River')

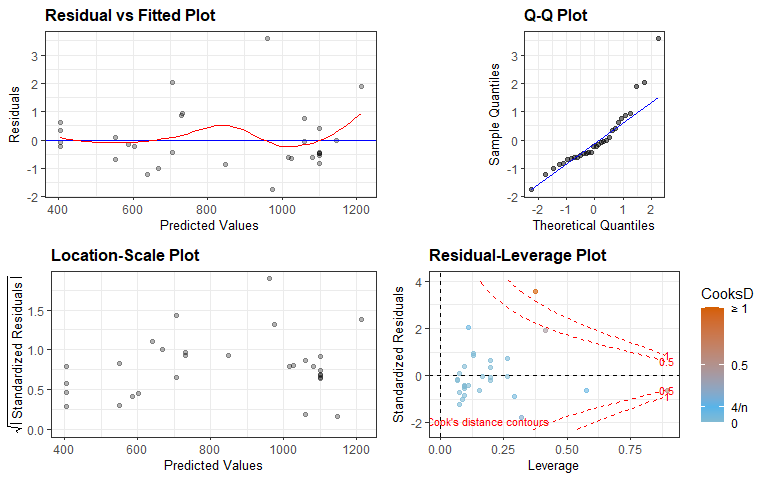


Models:

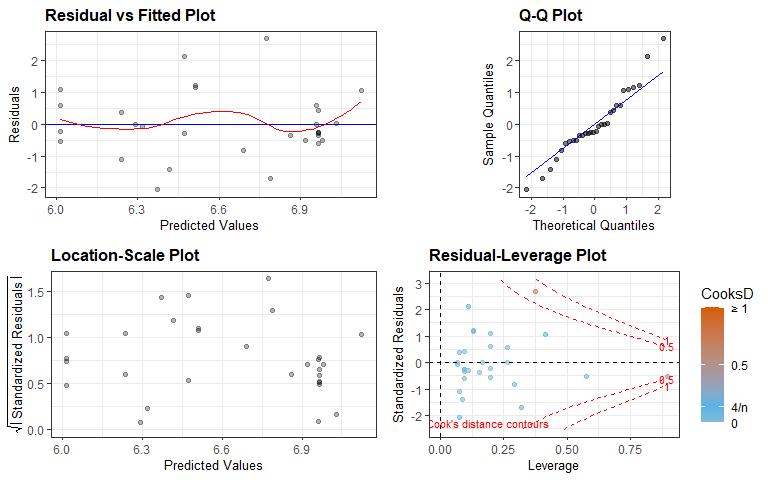
m3 <- lm(pCO2 ~ River\*Elevation, data = riverdataR)  
plot(allEffects(m3), grid = T)



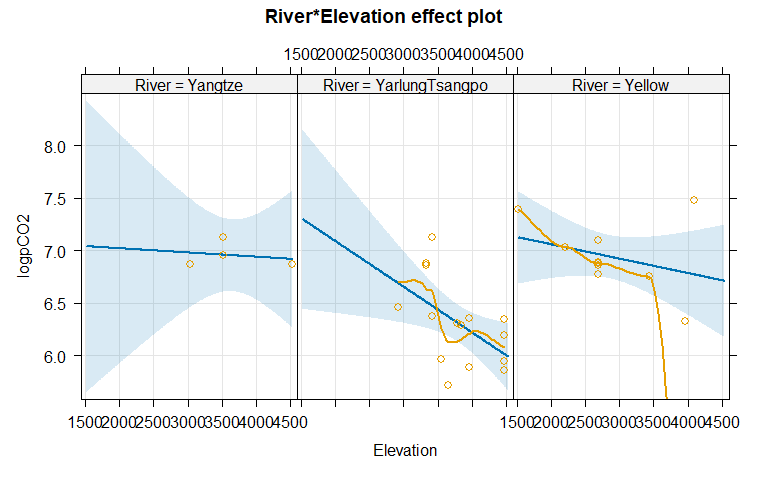
#fig 3  
resid\_panel(m3, "R", alpha =0.3)



m4 <- lm(logpCO2 ~ River\*Elevation, data = riverdataR)  
#fig 4  
resid\_panel(m4, "R", alpha = 0.3)



#Question 1: strong negative relationship for YT  
#fig 5  
plot(allEffects(m4, residuals = T), grid = T)



F-Test:

Anova(m4)

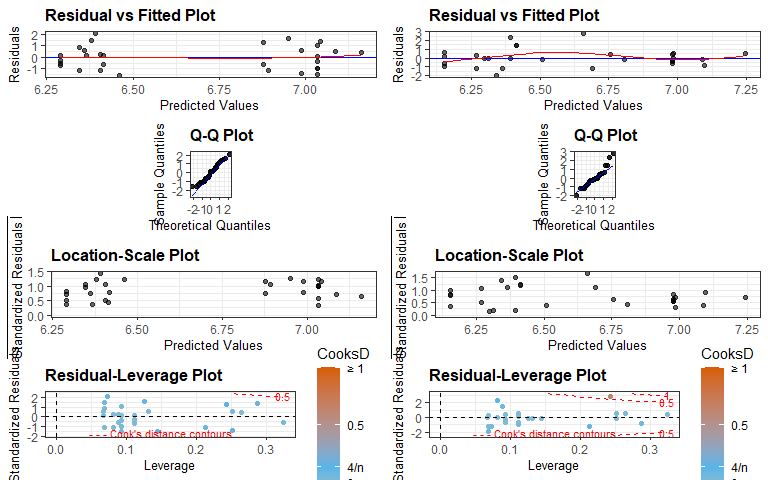
## Anova Table (Type II tests)  
##   
## Response: logpCO2  
## Sum Sq Df F value Pr(>F)  
## River 1.48826 2 6.7866 0.004613  
## Elevation 0.52664 1 4.8031 0.038350  
## River:Elevation 0.23595 2 1.0759 0.356863  
## Residuals 2.63153 24

# Weak evidence against the null of no interaction between river and elevation

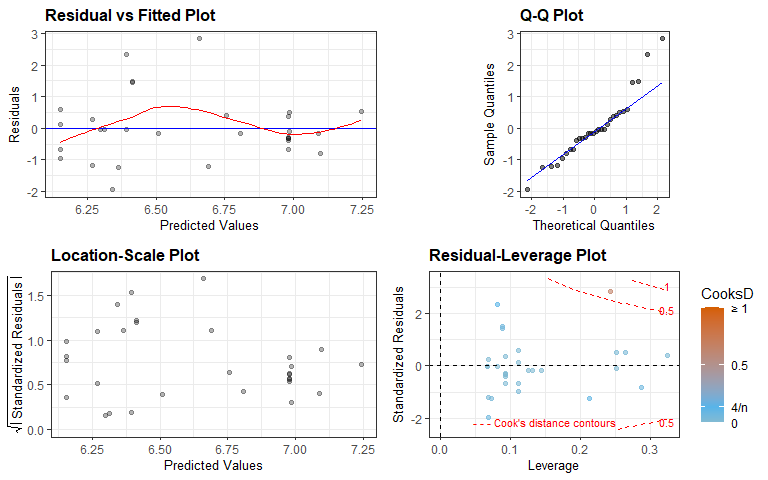
Figure 6:

m5 <- lm(logpCO2 ~ River + Elevation, data = riverdataR)  
# resid\_interact(lm\_add, "R")  
resid\_calibrate(m5, "R")

## [1] "Real residuals are in column 2"



resid\_panel(m5, "R", alpha = 0.3)



summary(m5)

##   
## Call:  
## lm(formula = logpCO2 ~ River + Elevation, data = riverdataR)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.62352 -0.18609 -0.03705 0.11648 0.81985   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 7.7878867 0.4157939 18.730 < 2e-16  
## RiverYarlungTsangpo -0.6145064 0.1875508 -3.276 0.00298  
## RiverYellow -0.1954865 0.2112858 -0.925 0.36336  
## Elevation -0.0002280 0.0001044 -2.185 0.03808  
##   
## Residual standard error: 0.3321 on 26 degrees of freedom  
## Multiple R-squared: 0.5567, Adjusted R-squared: 0.5056   
## F-statistic: 10.89 on 3 and 26 DF, p-value: 8.184e-05

Anova(m5)

## Anova Table (Type II tests)  
##   
## Response: logpCO2  
## Sum Sq Df F value Pr(>F)  
## River 1.48826 2 6.7472 0.004362  
## Elevation 0.52664 1 4.7752 0.038076  
## Residuals 2.86748 26

confint(m5)

## 2.5 % 97.5 %  
## (Intercept) 6.9332100615 8.642563e+00  
## RiverYarlungTsangpo -1.0000226612 -2.289902e-01  
## RiverYellow -0.6297906462 2.388176e-01  
## Elevation -0.0004425625 -1.353463e-05

Figure 7:

plot(allEffects(m5, residuals = T),grid = T)

