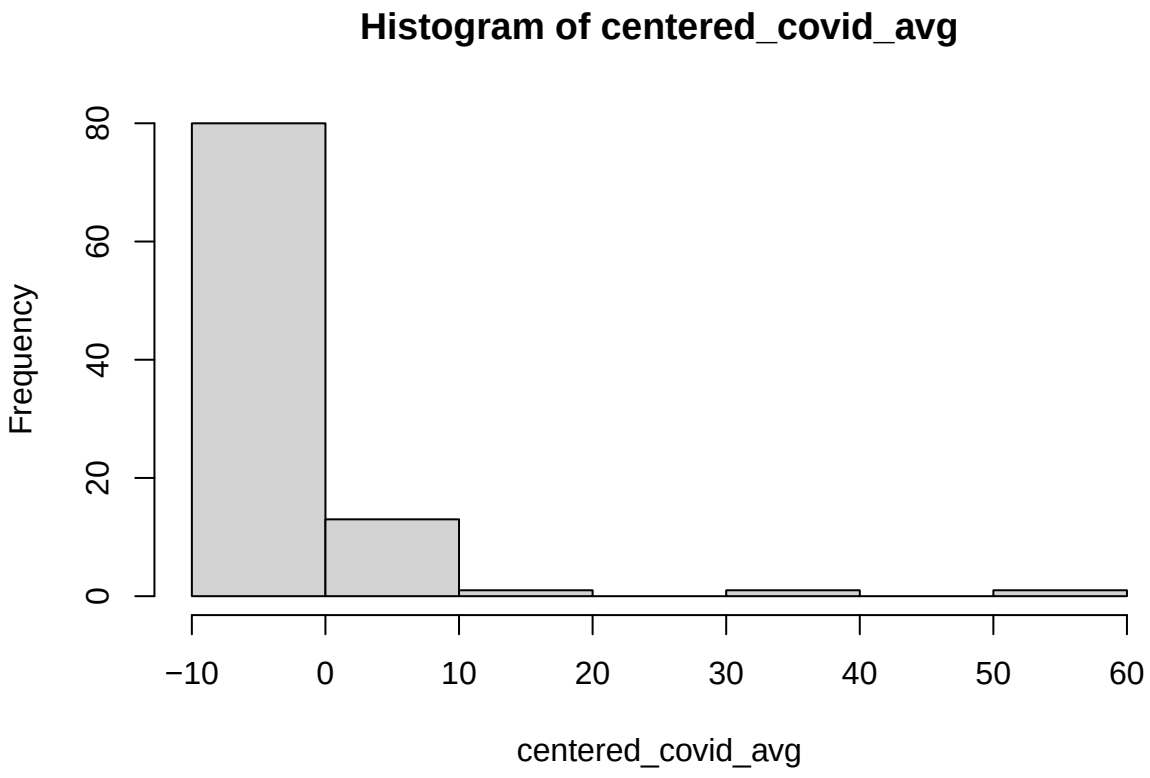
FinalReport

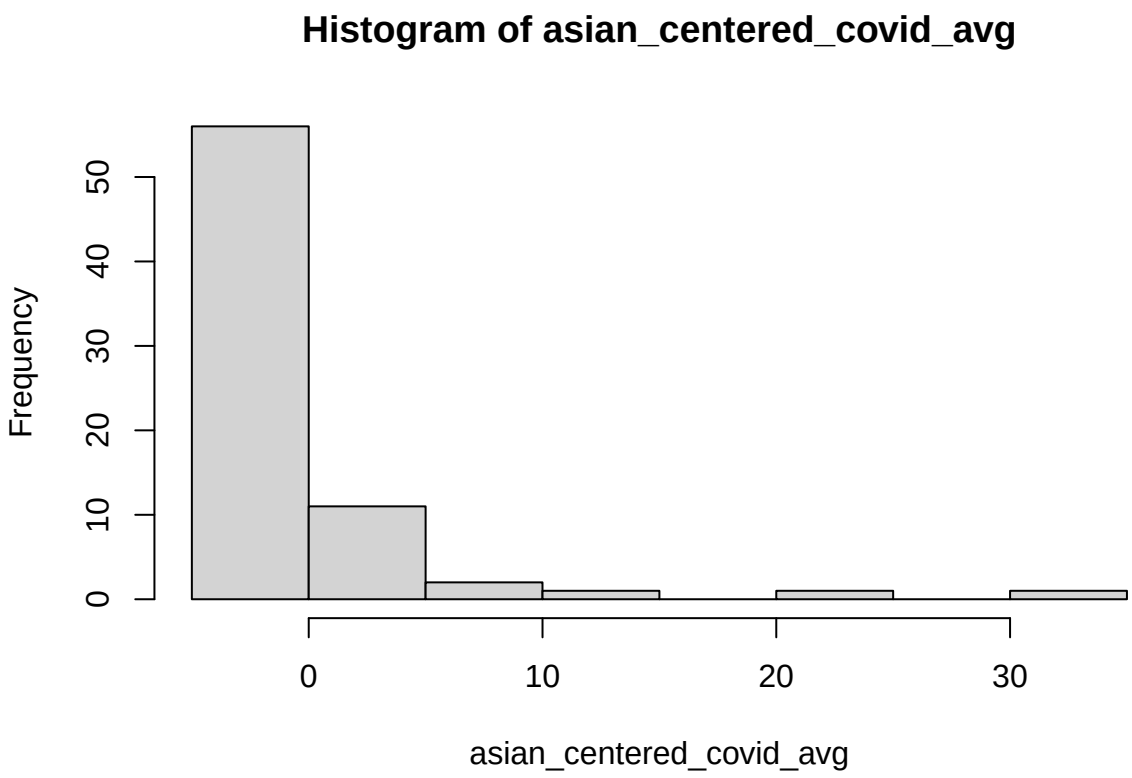
Gautam Gireesh, Saakshi Shah

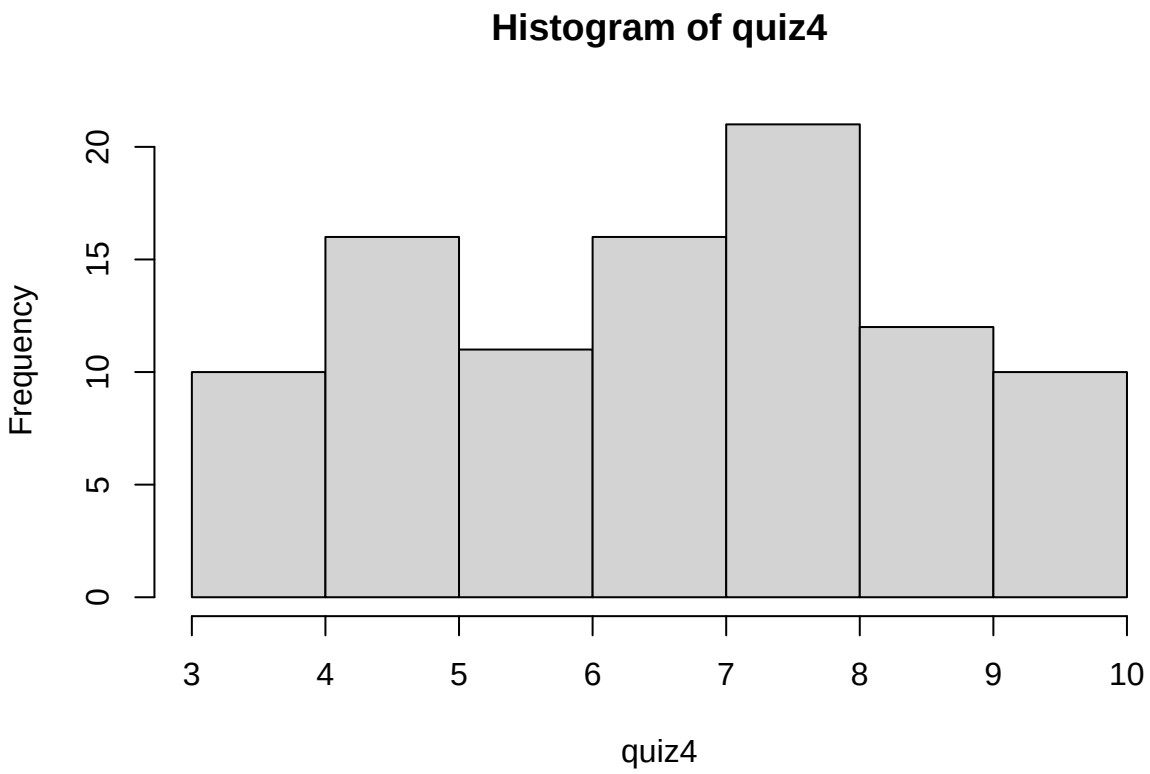
22/08/2021

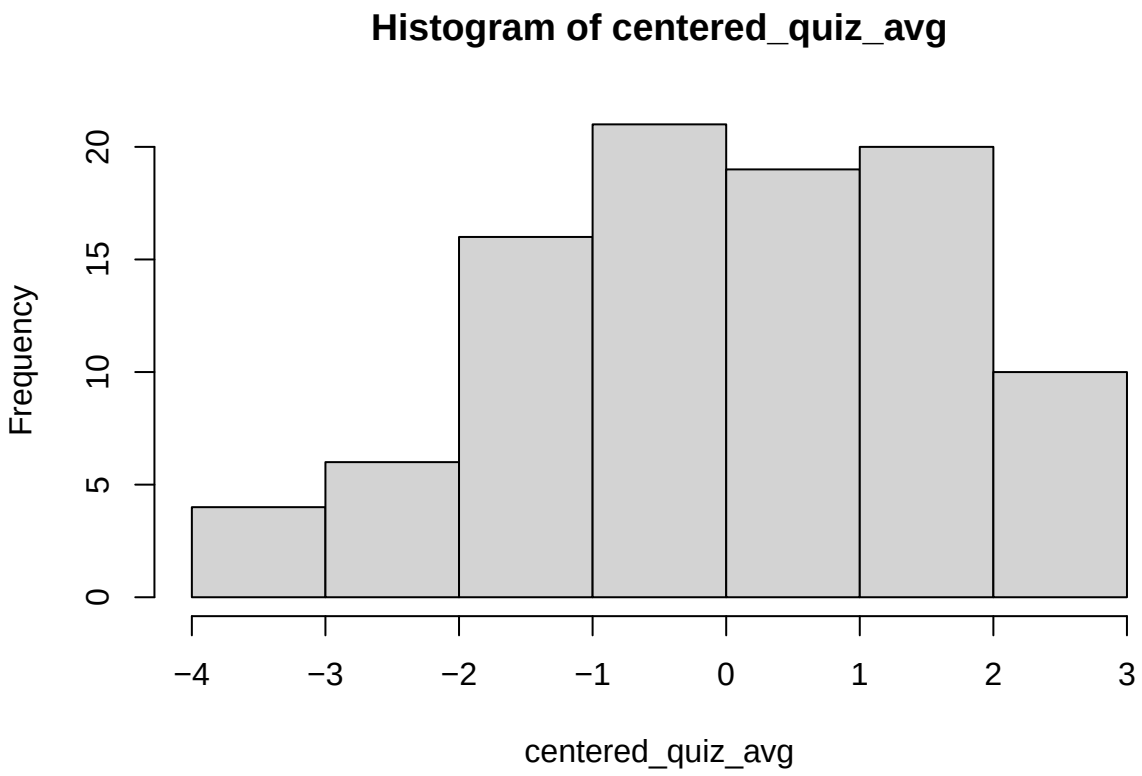
Introduction

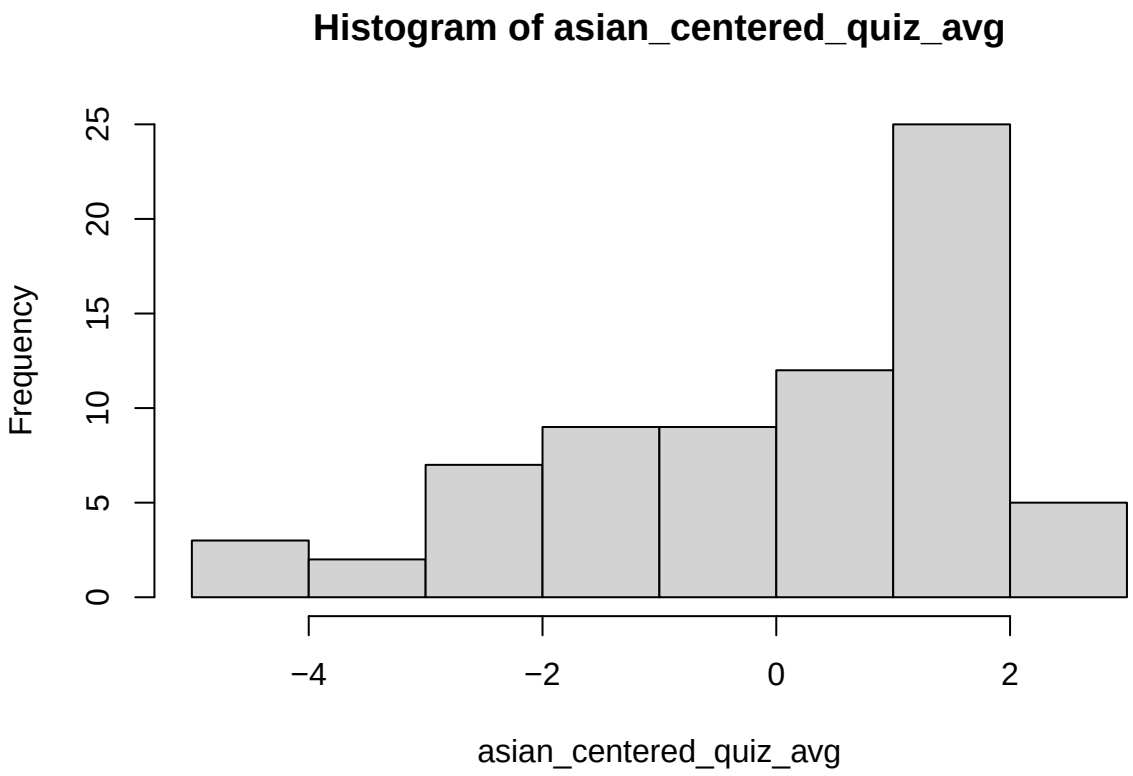
As aspiring statisticians, the goal of this study for us was to develop a reasonable regression model using the techniques we have learnt over the course of this semester. Through this, we would be able to understand and determine what factors do in fact impact a student's performance on quiz 4. At a glance, it would be easy for us to assume that all the aspects - country, quiz grades, hours spent on COVID and stats - must make an impact. But having been students of this course, we know better and will put our intuition to the test to determine whether or not all the factors make an impact. The following report will conduct an in-depth analysis of the data and explain how and why we chose the model.

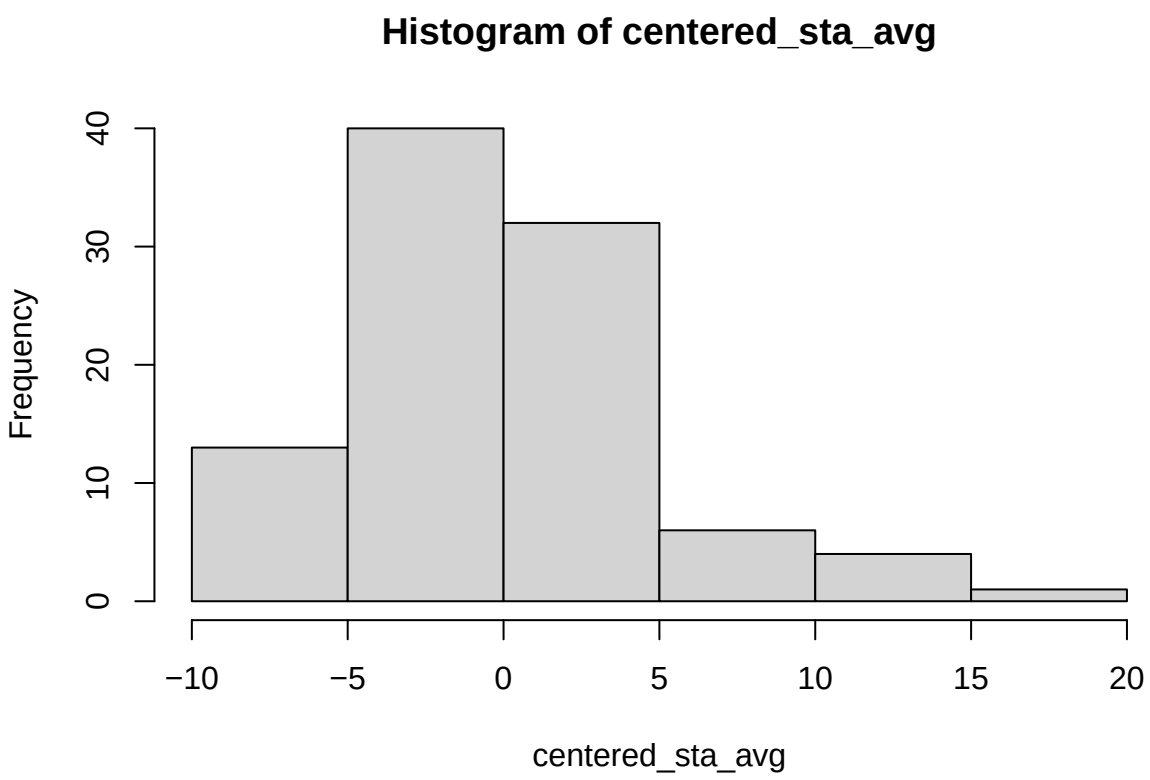


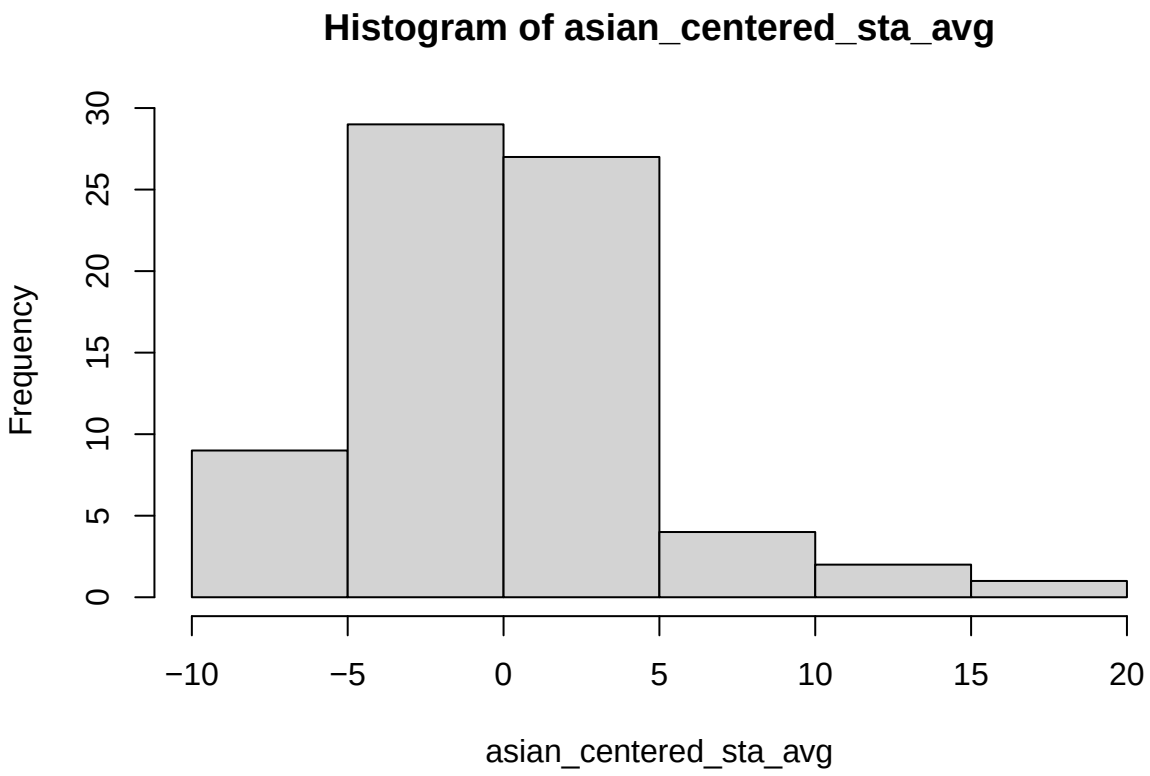

Histogram of asian_centered_covid_avg
Frequency
0
10
20
30
40
50
0
10
20
30
asian_centered_covid_avg

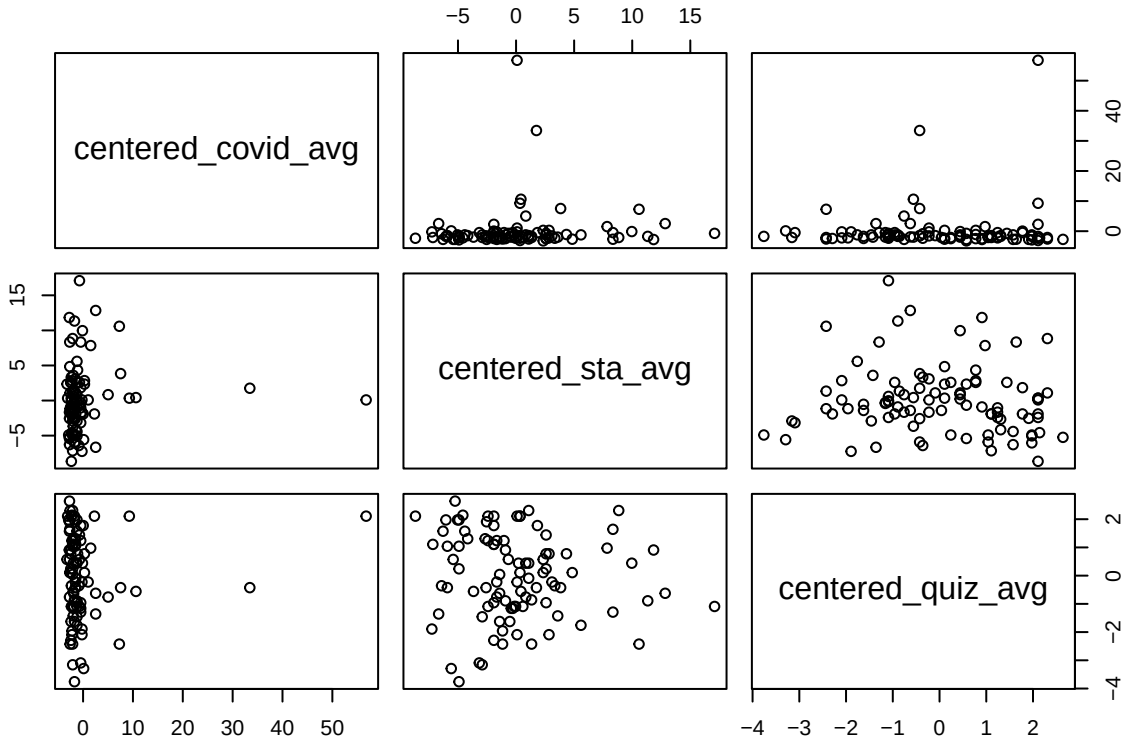

Histogram of quiz4
Frequency
0
5
10
15
20
3
4
5
6
7
8
9
10
quiz4

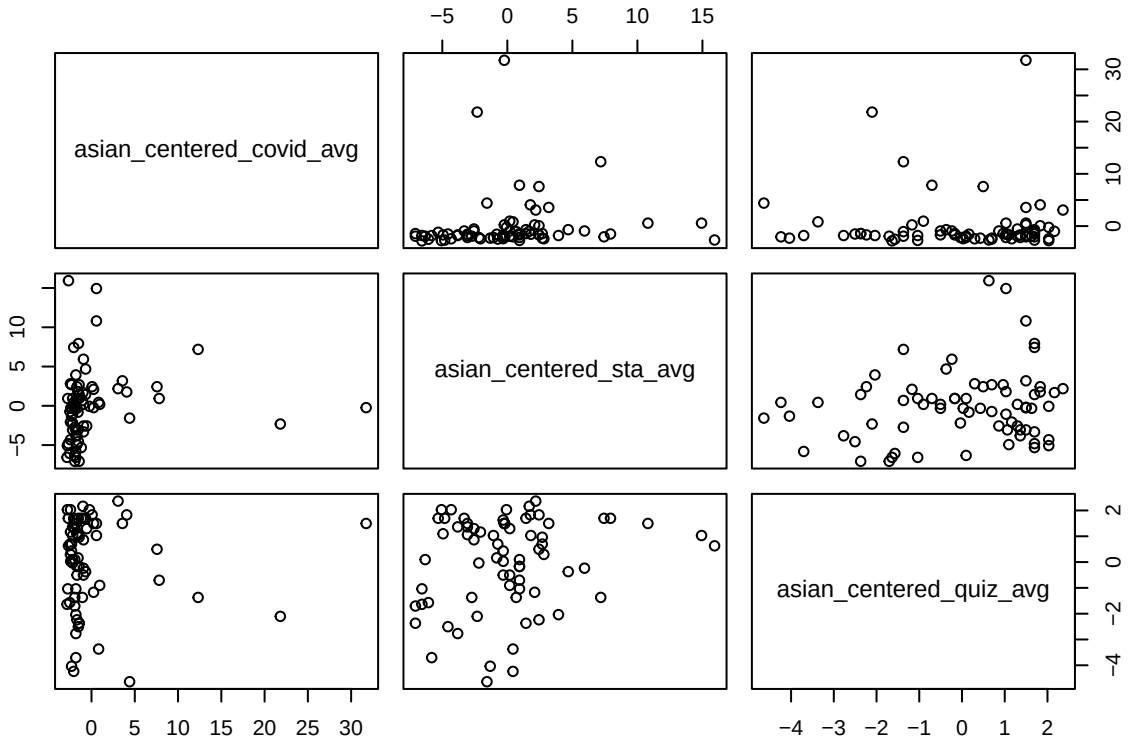

Histogram of centered_quiz_avg
Frequency
0
5
10
15
20
−4
−3
−2
−1
0
1
2
3
centered_quiz_avg


Histogram of asian_centered_quiz_avg
Frequency
0
5
10
15
20
25
−4
−2
0
2
asian_centered_quiz_avg


Histogram of centered_sta_avg
Frequency
0
10
20
30
40
−10
−5
0
5
10
15
20
centered_sta_avg


Histogram of asian_centered_sta_avg
Frequency
0
5
10
15
20
25
30
−10
−5
0
5
10
15
20
asian_centered_sta_avg


centered_covid_avg
centered_sta_avg
centered_quiz_avg
−5
0
5
10
15
0
20
40
−5
5
15
0
10
20
30
40
50
−4
−3
−2
−1
0
1
2
−4
−2
0
2



Choice of Method

Variable Selection

Description of Data

Processing of Obtaining Final Model

Goodness of Final Model

Final Model Interpretation and Importance:

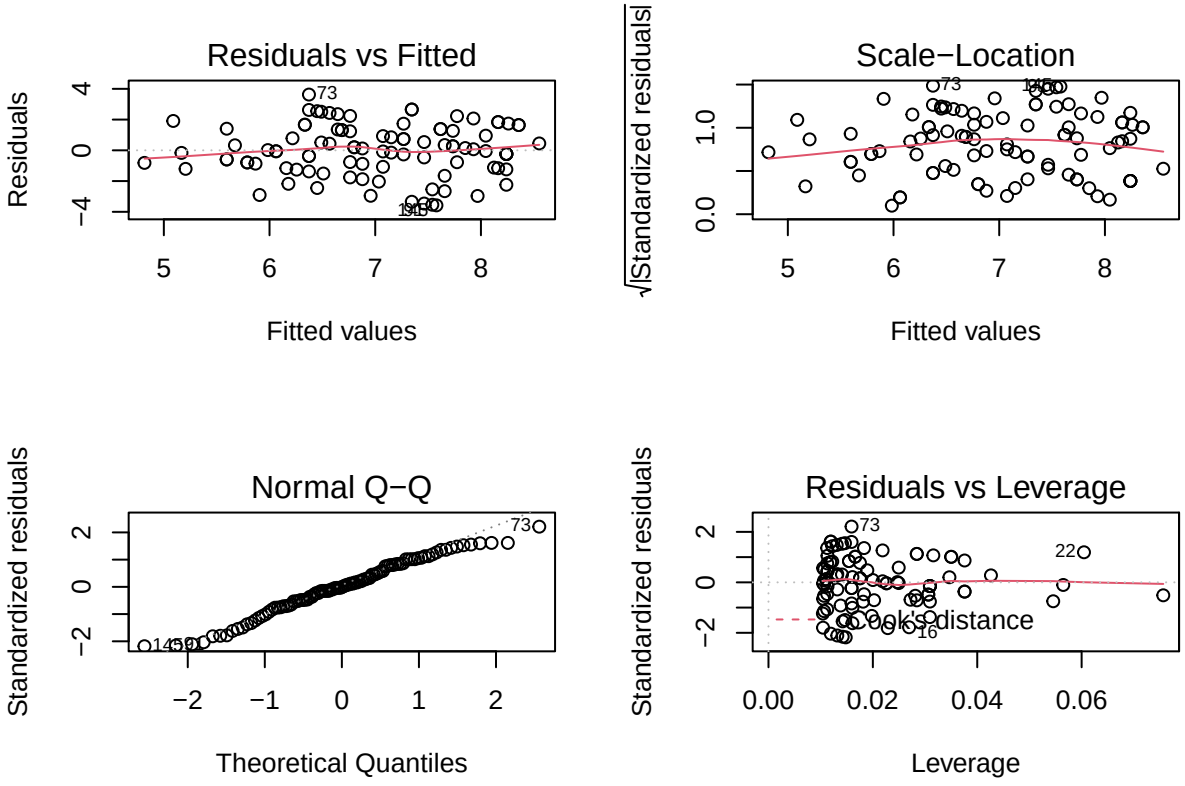
Limitations of Analysis:

Appendix

American Dataset modelling

```
n<-length(quiz4)

full_model<-lm(quiz4 ~ centered_covid_avg + centered_sta_avg + centered_quiz_avg, data=american_data)
```

```
model1<-lm(quiz4~centered_covid_avg, data=american_data)
p_prime1 <- length(model1$coefficients)
summary(model1)
SSres1<-deviance(model1)
AIC1<-n*log(SSres1) -n*log(n) + 2*p_prime1
print(AIC1)
mallow_cp1<-ols_mallows_cp(model1, full_model)

model2<-lm(quiz4~centered_sta_avg, data=american_data)
p_prime2 <- length(model2$coefficients)
summary(model2)
SSres2<-deviance(model2)
AIC2<-n*log(SSres2) -n*log(n) + 2*p_prime2
print(AIC2)
mallow_cp2<-ols_mallows_cp(model2, full_model)

model3<-lm(quiz4~centered_quiz_avg, data=american_data)
p_prime3 <- length(model3$coefficients)
summary(model3)
SSres3<-deviance(model3)
AIC3<-n*log(SSres3) -n*log(n) + 2*p_prime3
print(AIC3)
mallow_cp3<-ols_mallows_cp(model3, full_model)
layout(matrix(c(1,2,3,4),2,2))
plot(model3)
```

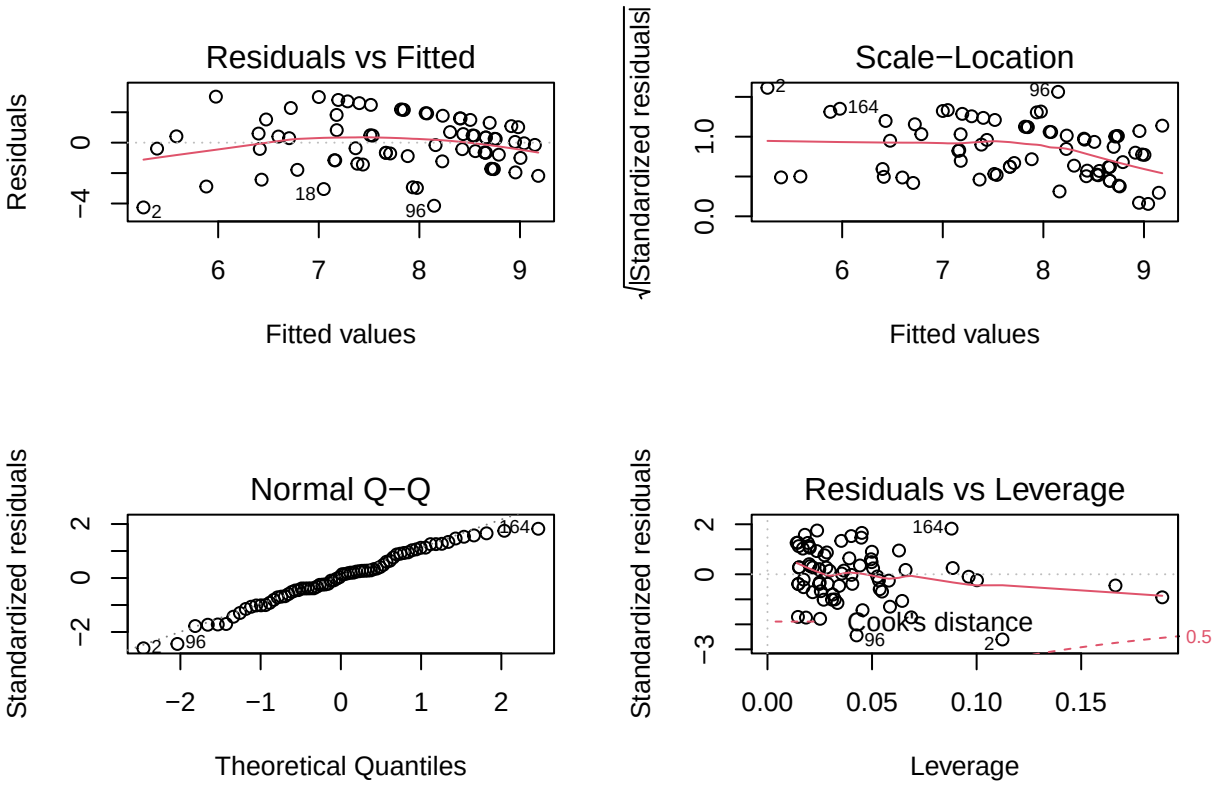


```
model4<-lm(quiz4~centered_covid_avg + centered_quiz_avg, data=american_data)
p_prime4 <- length(model4$coefficients)
summary(model4)
SSres4<-deviance(model4)
AIC4<-n*log(SSres4) -n*log(n) + 2*p_prime4
print(AIC4)
mallow_cp4<-ols_mallows_cp(model4, full_model)

model5<-lm(quiz4~centered_sta_avg + centered_quiz_avg, data=american_data)
p_prime5 <- length(model5$coefficients)
summary(model5)
ssres5<-deviance(model5)
aic5<-n*log(ssres5) -n*log(n) + 2*p_prime5
print(aic5)
mallow_cp5<-ols_mallows_cp(model5, full_model)


model6<-lm(quiz4~centered_sta_avg + centered_covid_avg, data=american_data)
p_prime6 <- length(model6$coefficients)
summary(model6)
SSres6<-deviance(model6)
AIC6<-n*log(SSres6) -n*log(n) + 2*p_prime6
print(AIC6)
mallow_cp6<-ols_mallows_cp(model6, full_model)

model7<-lm(quiz4~centered_quiz_avg + centered_covid_avg + centered_sta_avg, data=american_data)
p_prime7 <- length(model7$coefficients)
summary(model7)
SSres7<-deviance(model7)
AIC7<-n*log(SSres7) -n*log(n) + 2*p_prime7
print(AIC7)
mallow_cp7<-ols_mallows_cp(model7, full_model)
```

Asian Dataset modelling

```
asian_n<-length(asian_quiz4)

asian_full_model<-lm(asian_quiz4~asian_centered_covid_avg + asian_centered_sta_avg + asian_centered_qui

asian_model1<-lm(asian_quiz4~asian_centered_covid_avg, data=asian_data)
asian_p_prime1 <- length(asian_model1$coefficients)
summary(asian_model1)
asian_SSres1<-deviance(asian_model1)
asian_AIC1<-asian_n*log(asian_SSres1) -asian_n*log(asian_n) + 2*asian_p_prime1
print(asian_AIC1)
asian_mallow_cp1<-ols_mallows_cp(asian_model1, asian_full_model)

asian_model2<-lm(asian_quiz4~asian_centered_sta_avg, data=asian_data)
asian_p_prime2 <- length(asian_model2$coefficients)
summary(asian_model2)
asian_SSres2<-deviance(asian_model2)
```

```
asian_AIC2<-asian_n*log(asian_SSres2) -asian_n*log(asian_n) + 2*asian_p_prime2
print(asian_AIC2)
asian_mallow_cp2<-ols_mallows_cp(asian_model2, asian_full_model)

asian_model3<-lm(asian_quiz4~asian_centered_quiz_avg, data=asian_data)
asian_p_prime3 <- length(asian_model3$coefficients)
summary(asian_model3)
asian_SSres3<-deviance(asian_model3)
asian_AIC3<-asian_n*log(asian_SSres3) -asian_n*log(asian_n) + 2*asian_p_prime3
print(asian_AIC3)
asian_mallow_cp3<-ols_mallows_cp(asian_model3, asian_full_model)

asian_model4<-lm(asian_quiz4~asian_centered_covid_avg + asian_centered_quiz_avg, data=asian_data)
asian_p_prime4 <- length(asian_model4$coefficients)
summary(asian_model4)
asian_SSres4<-deviance(asian_model4)
asian_AIC4<-asian_n*log(asian_SSres4) -asian_n*log(asian_n) + 2*asian_p_prime4
print(asian_AIC4)
asian_mallow_cp4<-ols_mallows_cp(asian_model4, asian_full_model)

asian_model5<-lm(asian_quiz4~asian_centered_sta_avg + asian_centered_quiz_avg, data=asian_data)
asian_p_prime5 <- length(asian_model5$coefficients)
summary(asian_model5)
asian_SSres5<-deviance(asian_model5)
asian_AIC5<-asian_n*log(asian_SSres5) -asian_n*log(asian_n) + 2*asian_p_prime5
print(asian_AIC5)
asian_mallow_cp5<-ols_mallows_cp(asian_model5, asian_full_model)
layout(matrix(c(1,2,3,4),2,2))
plot(asian_model5)
```




```
asian_model6<-lm(asian_quiz4~asian_centered_sta_avg + asian_centered_covid_avg, data=asian_data)
asian_p_prime6 <- length(asian_model6$coefficients)
summary(asian_model6)
asian_SSres6<-deviance(asian_model6)
asian_AIC6<-asian_n*log(asian_SSres6) -asian_n*log(asian_n) + 2*asian_p_prime6
print(asian_AIC6)
asian_mallow_cp6<-ols_mallows_cp(asian_model6, asian_full_model)

asian_model7<-lm(asian_quiz4~asian_centered_quiz_avg + asian_centered_covid_avg + asian_centered_sta_avg
asian_p_prime7 <- length(asian_model7$coefficients)
summary(asian_model7)
asian_SSres7<-deviance(asian_model7)
asian_AIC7<-asian_n*log(asian_SSres7) -asian_n*log(asian_n) + 2*asian_p_prime7
print(asian_AIC7)
asian_mallow_cp7<-ols_mallows_cp(asian_model7, asian_full_model)
```