PART A: QUANTITATIVE QUESTIONS

QUESTION 1: The Long-Term Impact of the Slave Trade

a)

i. Countries that are compliers are those that are situated at a low distance from a major slave destination and have high levels of slave exports. These are also countries that are situated at a great distance from a major slave destination and have low levels of slave exports.

Countries that are always-takers are those that have a high high level of slave exports regardless of whether or not they are at a low distance from a major slave destination.

ii-iii. Table 1: ITT, proportion of compilers, CACE

ITT	-0.016
Proportion of compliers	0.058
CACE	-0.276

iv. Table 2: Estimating p-value of CACE

	Estimate	**Standard Error**	**t-value**	**p-value**
High slavery	-0.276	3.879	-0.071	0.944

v. The ITT is -0.016, meaning that being low distance from a major slave destination caused a 0.016 decrease in log real GDP per capita in 2000.The proportion of compliers is 0.058 meaning that only 5.778% of the countries are compliers. The CACE is -0.276 meaning that amongst the compliers, those who have high slavery, causes a decrease in logged real GDP per capita in the year 2000 by 0.276. The CACE is not statistically significant at 5% and 10% level with a p-value of 0.944.

b) Table 3: First-stage results

	Estimate (standard error)
Atlantic distance	-1.314*** (0.357)
Indian distance	-1.095** (0.380)
Saharan distance	-2.435** (0.823)
Red Sea distance	-0.002 (0.710)
** P<0.05 ** P<0.01 *** P<0.001* *Values were rounded off to three decimal places*	

The first stage tells us how strongly the instrument affects the treatment. As can be seen, we obtain the same results as Nunn, with all coefficients "generally negative" (Nunn, 2008, p162) and that only the red sea distance variable is not statistically significant.

c) The F statistic is 0.811 which is less than the minimum level to consider an instrument sufficiently strong (F-stat > 10). Resultantly, the instruments in this paper are subject to the weak instrument problem. This means that it may not strongly reflect the causal effect of the treatment towards the outcome variable.

Nunn reflected this in his paper when he acknowledged that the distances to slave markets may be correlated with distances to other locations that are important for economic development.

d) To satisfy the exclusion restriction, it should be that the only way that the sailing or overland distance of a country from the nearest slave trade (i.e. Atlantic, Indian, Saharan and Red Sea slave trades) affects the log real per capita GDP in 2000 is through its impact on the log total number of slaves exported divided by land area. To check for this, I regressed the log real per capita GDP in 2000 with the instrumental variables. The table illustrates that all the IVs are significant except for the Red Sea, which means that there is a direct correlation between the outcome variable and three of the instrumental variables apart through its treatment uptake, thus violating the exclusion restriction.

Table 4: Regressing log real per capita GDP in 2000 with Instrumental Variables

	Estimate (standard error)	**p-value**
Atlantic distance	0.328	3.45e-05***
Indian distance	0.307	0.000***
Saharan distance	0.594	0.001***
Red Sea distance	0.101	0.484

e) Table 5: Second-stage coefficients and standard errors

	Second stage models. Dependent variable is log income, y		
	(1)	**(2)**	**(3)**
In(exports/area)	-0.208*** (0.053)	-0.201*** (0.047)	-0.286* (0.153)
Colonizer fixed effects	No	Yes	Yes
Geography controls	No	No	Yes
F-stat	15.390	4.322	1.703

The estimated Local Average Treatment Effect (LATE) of the number of slaves exported on GDP per capita in 2000 is that having one unit increase in log total number of slaves exported divided by land area led to a 0.208 decrease in log real per capita GDP in 2000. When controlling for coloniser fixed effects, one unit increase in log total number of slaves exported divided by land area led to a 0.201 decrease in log real per capita GDP in 2000. Lastly, when controlling for coloniser fixed effects and geography controls, one unit increase in log total number of slaves exported divided by land area led to a 0.286 decrease in log real per capita GDP in 2000. The p-values for the first two are significant at the 5% level and the last being significant at the 10% level.

f) Nunn estimated the additional models because randomisation could be violated. For example, African countries that are colonised by a specific coloniser country may have higher economic development than others given that the coloniser may have assisted in ways such as public education. Or that being colonised by a specific coloniser country would make them significantly economically worse off than others because of resource exploitation. This is why Nunn may have controlled for coloniser fixed effects.

Secondly, geographic controls such as minimum monthly rainfall would also have an impact on the country's agriculture economy which might ultimately affect the outcome variable, the log real per capita GDP in 2000. This may be why Nunn controlled for geographic attributes.

QUESTION 2: A Simulated Experiment

a) T-value is -5.715 showing that it is statistically significant. X coefficient is -0.625. Even though it is relatively close to zero, it is nonetheless statistically significant and shows that randomization has failed. It shows that people under the treatment group were less likely to have the covariate (0.0 in comparison to 0.4) in comparison to those in the control group.

b) The true ATE is -7.495 and the observed ATE is -9.421. There is a 1.926 negative difference between the two ATEs. Thus, selection bias is negative.

c) There is a positive correlation between the outcome and the baseline covariate variables of 0.477. However, there is a negative correlation between the treatment and baseline covariate variable of -0.5. This means that there is a negative bias. This is further supported by question b, where the observed ATE is more negative than the true ATE.

d) By blocking and converting X to a factor variable, we get an ATE of -7.499 which is much closer to the true ATE at -7.495. It is statistically significant at the 5% level with p-value 1.33e-06.

Table 6: Coefficient for ATE regression with blocking

	Estimate	**Standard Error**	**t-value**	**p-value**
d	-7.499	1.454	-5.157	1.33e-06***

e) For units to be missing at random, the average potential outcome of the control group and that reports, should be equal to the average potential outcome of the control group for those that do not report. The same holds for the treatment group.

Table 7: Missing at random

	Average potential outcome	p-value
Control group		
Reporters	26.372	0.004***
Non-reporters	32.259	
Treatment group		
Reporters	19.851	1.842e-12***
Non-reporters	17.621	

From the table above, we see that for the control group, those that report have an average potential outcome of 26.372 whereas those that do not report have an average potential outcome of 32.259. Hence, the two do not equal each other and it is not missing at random for the control group. It is significant at the 5% level.

Likewise occurs for the treatment group, as those that report have an average potential outcome of 19.851 whereas those that do not report have an average potential outcome of 17.621. Hence, the two do not equal each other and it is not missing at random for the treatment group. It is significant at the 5% level.

f)

i) As seen in part e, the dataset is not missing at random. From the dataset, we can see that missingness (non-reporters) only occurs when treatment, d, equals 0 (i.e. control group). Thus they are no longer valid counterfactuals for each other. Whereas always-reporters come about when missingness occurs the same way in both treatment and control group. This contrasts missingness not at random and thus we cannot calculate the ATE for always-reporters as they do not exist in this dataset and the statement is true.

ii) For units to be missing-at-random within values of x, the mean of the baseline covariates for those who report should equal the mean of the baseline covariate for those who do not report. From Table 8, we can see that for those who do reports, the mean outcome for values of x is 0.091 and for those who do not, is 1. The results are statistically significant at the 5% level with p-value of less than 2.2e-16. Hence, they are not equal to each other and the statement is false.

Table 8: Missing-at-random within values of x

	Mean of baseline covariates	**p-value**
Reporters	0.091	<2.2e-16
Non-reporters	1	