Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Question 1

Right Click on Forest_NonForest, then properties, symbology, and change the yellow (value 1) to green. Then press ok, apply, and close.

Question 2

628 – 601 = 27. 27 villagers were removed.

Question 3

“ZS_SUM” represents the total number of forested pixels under that buffer zone. Zs_Mean is the total number of forested pixels divided by the total number of pixels (forest + non-forest) within that buffer zone.

Question 4

The expression used to create the new column was (1 – ZS__Mean) * 745.29. The buffer with the smallest amount of unforested area had a cluster ID of 30106331, and an area of 30m^2.

Question 5

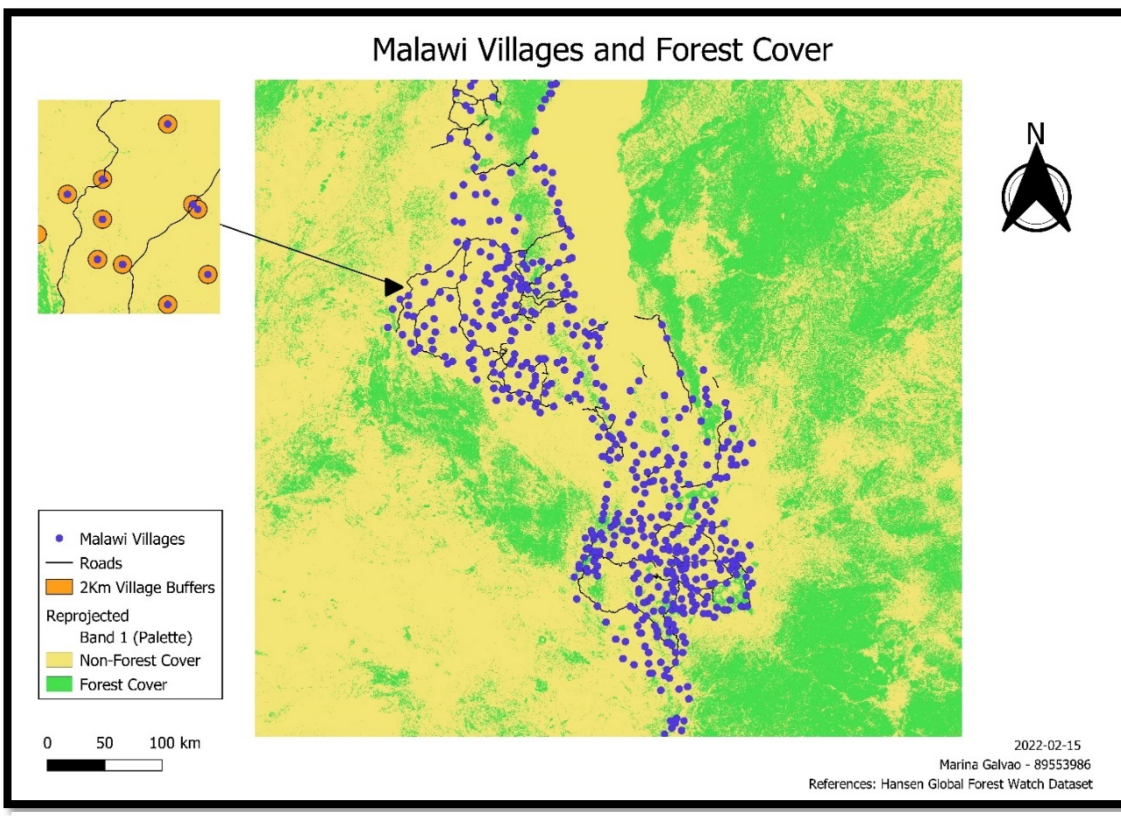

Figure 1, Map showing Malawi Villages, spatial forest cover and non-forested regions, as well as 2Km village buffers.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Question 6

There are 10037 households in the dataset. “Annual Temperature” is a non-dietary variable that might affect dietary diversity because temperatures that are too extreme (whether cold or warm) will inhibit some crop species from growing. A hypothesis for this may be “years with annual temperatures that are significantly different from the usual range of annual temperature mean will have a smaller variety of crops, lowering dietary diversity.”

Question 7

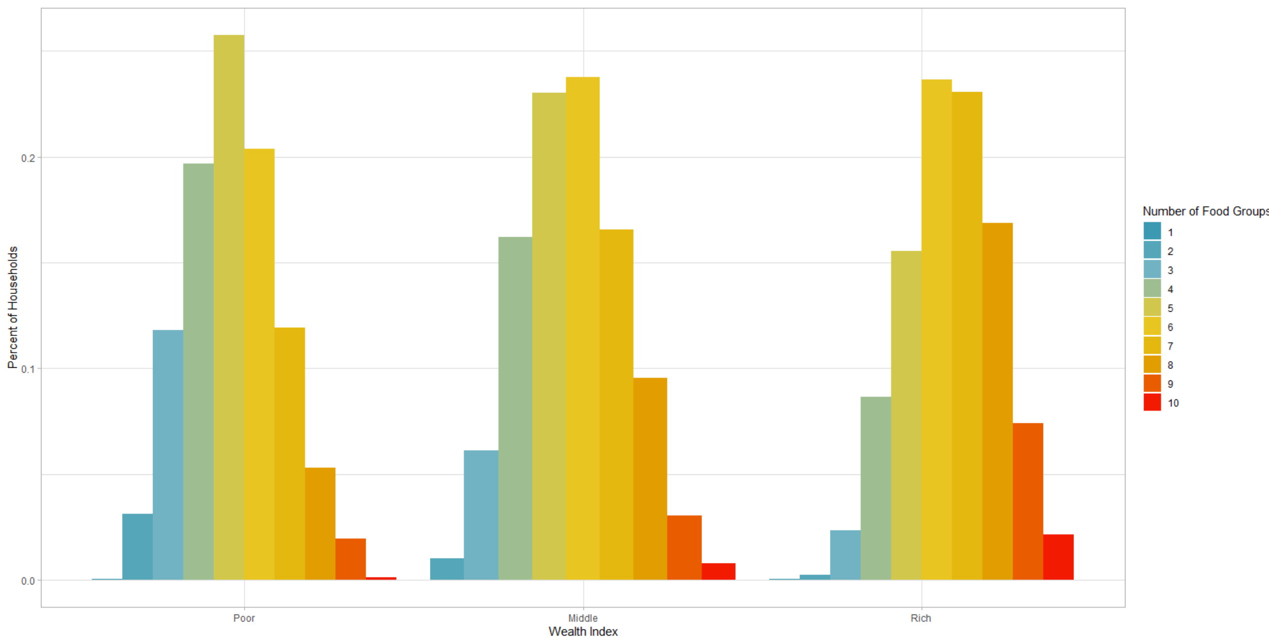

Figure 2, Malawi Population Food Groups Diversity. A bar plot of the Malawi population as percentage of households (y-axis) and categorized into Wealth Indexes (x-axis, “Poor”, “Middle”, and “Rich” classes). Number of food groups are shown in the legend. This dataset encompasses 1-10 food groups.

For communities of a Poor Class Wealth Index, most households consume between 4-6 food groups. For communities of a Middle-Class Wealth Index, the trend shows most of the consumption occurs between 4-7 food groups, showing an increase in relation to the Poor Wealth Index. Lastly, the Rich Wealth Index shows large amounts of consumption of 6 and 7 food groups, followed by 5 and 8 food groups.

There is a relationship present in Figure 2. The relationship shown in the bar chart is one in which as Wealth Index increases (from Poor, to Middle, to Rich), increases are also seen in number of food groups consumed. Percent of households that consume 1-5 food groups decrease as the graph moves from poor to rich. Percent of households which consume 6-10 food groups, on the other hand, increase. The figure shows that wealth is a factor which is strong enough to cause incredibly high increases in food group diversity in relation to consumption.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Question 8

Table 1, Malawi Healthy DDS Score mean respective to Poor, Middle, and Rich Wealth Groups. The Poor Wealth Group shows the lowest Healthy DDS Score mean, followed by the Middle Wealth Group, and the Rich Wealth Group.

Mean for Poor Wealth Group's Healthy DDS Score	Mean for Middle Wealth Group's Healthy DDS Score	Mean for the Rich Wealth Group's Healthy DDS Score
5.156828	5.701811	6.465139

Table 2, Malawi Combined DDS Score mean respective to Poor, Middle, and Rich Wealth Groups. The Poor Wealth Group shows the lowest Combined DDS Score mean, followed by the Middle Wealth Group and the Rich Wealth Group.

Mean for Poor Wealth Group's Combined DDS Score	Mean for Middle Wealth Group's Combined DDS Score	Mean for Rich Wealth Group's Combined DDS Score
7.098049	8.208137	9.376494

Visually analyzing the means calculated, the Poor Wealth Group possesses the lowest Healthy DDS Score and Combined DDS Score, followed by the Middle Wealth Group and the Rich Wealth Group. The same pattern being observed in both Healthy DDS (healthy food groups) and Combined DDS (healthy and unhealthy with slight difference in vegetable and fruit grouping) shows that the richer you are, the more diversity in healthier foods you eat, but also, the more food diversity is present in your diet in general.

To begin, ANOVA tests were performed, first between Healthy DDS amongst Wealth Indexes, and then between Combined DDs amongst Wealth Indexes. Both ANOVAs resulted in a p-value of $< 2.2e-16$, presenting a statistical significance between Wealth Group means for both Healthy DDs and Combined DDS.

t-tests were also performed to test for specificity in the significant differences. The first t-test analysed Healthy DDS between the Middle Wealth Group and the Poor Wealth Group. Its alternate hypothesis stated that the true difference in means between the Middle Wealth Group and the Poor Wealth Group is not equal to 0. The p value of $< 2.2e-16$ lead to the conclusion that the means (5.16 for the Poor Wealth Group, and 5.7 for the Middle Wealth Group) are statistically significant. The second t-test analysed Healthy DDS between the Middle Wealth Group and the Rich Wealth Group. With a p-value of $< 2.2e-16$, this test also showed a statistically significant difference. The last t-test performed looked at Healthy DDS means between the Poor Wealth Group and Rich Wealth Group. With a p-value of $<2.2e-16$, the conclusion was also that the difference in means is significant. Details on codes and results of t-tests can be found in the appendix along with supplementary figures for tables 1 and 2.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Overall, all t-tests showed a statistically significant difference between groups. Numerous possibilities for the presence of the statistically significant differences are possible. For example, the rich are likely to have greater accessibility to markets and areas which are further away, thus expanding their dietary diversity. Moreover, as you get richer and possess more purchasing power, you are able to buy a greater quantity of different products. Further, poorer people may tend to eat from their own croplands, whereas richer households might work different jobs and purchase food instead, increasing their variety as the market is likely to possess more options than a household farm.

Question 9

Forest cover is positively associated with Healthy DDS and Combined DDS, however, the strength of this relationship is minor (Figures 3 and 4). This is likely because Malawi does not possess a lot of forest cover. The Global Forest Watch data also backs up this lab's findings, as Malawi's spatial representation shows very limited forest cover. The article which this lab is based on also corroborates these findings, stating that Malawi was the only country in the study that did not contain clusters in heavily forested landscapes, and had extremely lower patch sizes (Rasmussen et al., 2020). This extremely low amount of forest cover in the country impacts the results because it does not show a strong positive or negative correlation given the lack of forest cover present to test the possibility of a relationship with food security.

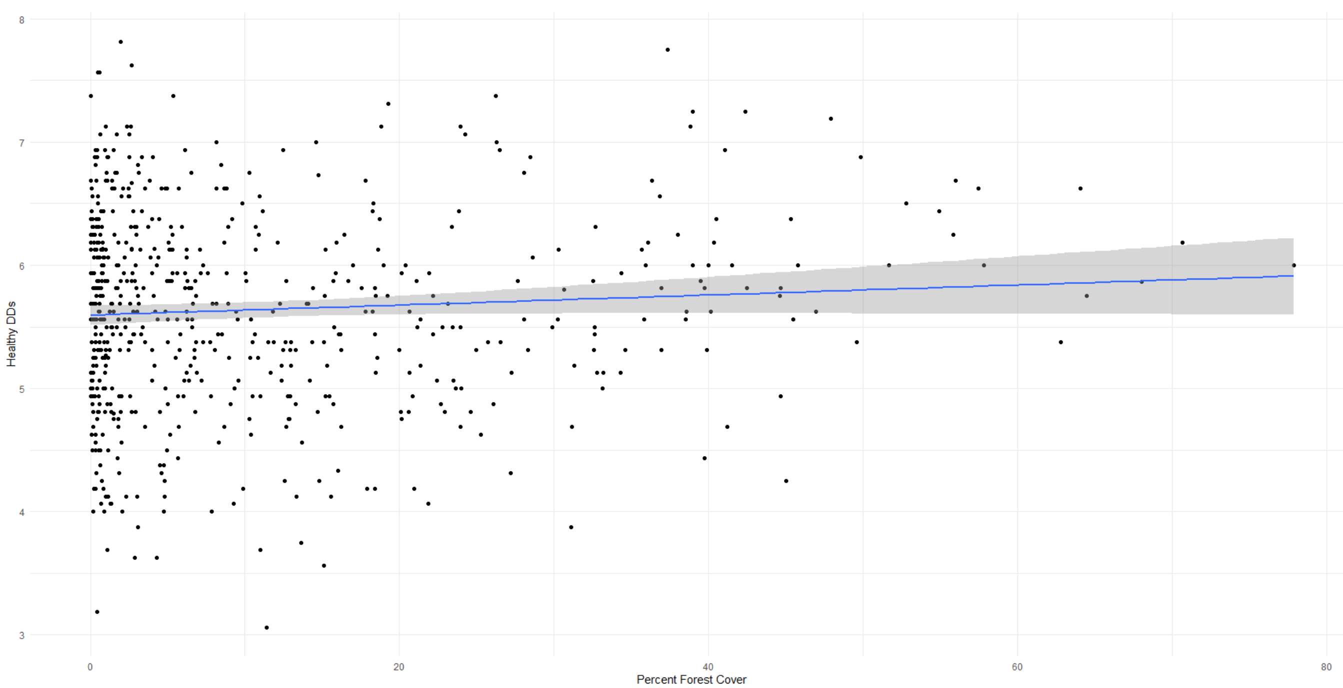

Figure 3, Malawi's Percent Forest Cover (x-axis) as a function of Healthy DDS (y-axis). A linear regression line has been included.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

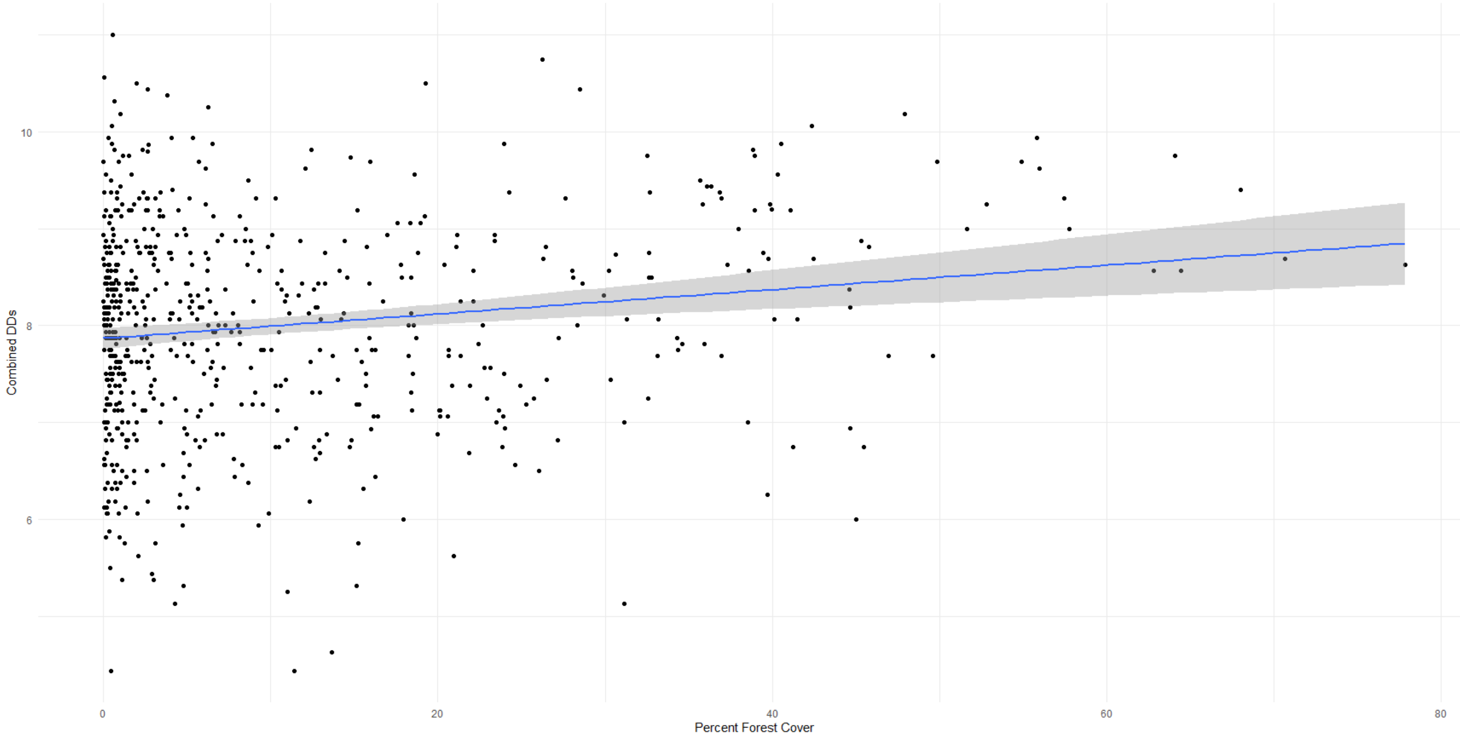

Figure 4, Malawi's Percent Forest Cover (x-axis) as a function of Combined DDS (y-axis). A linear regression line has been included.

Question 10

Box plots were created for Wealth Index and Combined DDS (Figure 5), as well as Wealth Index and Healthy DDS (Figure 6). The Combined DDS (healthy and unhealthy with slight difference in vegetable and fruit grouping) box plot (Figure 5) shows lowest DDS mean is <30% Forest Cover, followed by 30-60% Forest Cover and very closely to >60% Forest Cover. The Healthy DDS (healthy food groups) box plot (Figure 6) shows lowest DDS mean is found in <30% Forest Cover, followed by >60% Forest Cover and closely followed by 30-60% Forest Cover.

With these findings in mind, it is evident that higher forest cover results in both greater Combined DDS and greater Healthy DDS. In other words, food security is linked to forests as higher forest coverage leads to higher diversity of food (Figures 5 and 6). This is likely because an increase in forest area would lead to greater food supply quantities and an increased source of various nutrients. However, it is important to keep in mind Malawi's low forest cover which was previously discussed. In the figures (5 and 6), the mean values appear very close together between groups "30-60% Forest Cover" and ">60% Forest Cover". This may be because of a certain point where the bar graphs begin to level off in relation to forest cover effects. Linking this discussion to Figures 3 and 4, fewer data points are seen in high percent forest cover, which may explain the relationship being observed here – one in which effects are not large after a certain point.

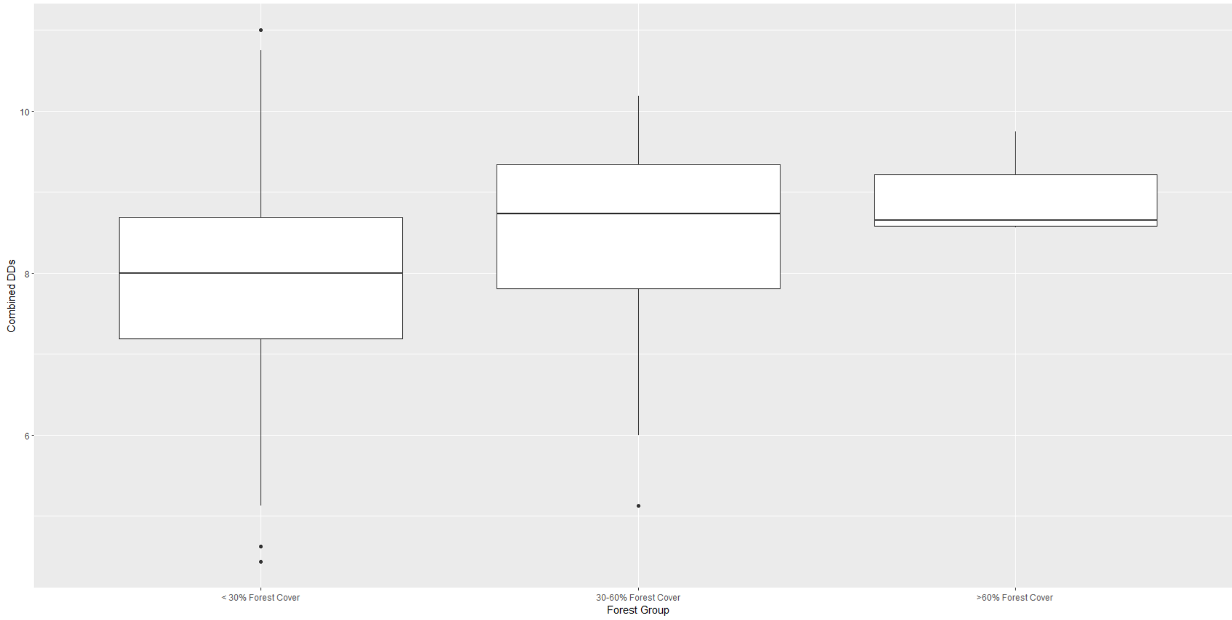

Figure 5, Box Plot of Forest Groups (x-axis, "<30% forest cover", "30-60% Forest Cover" and ">60% Forest Cover") and respective combined DDS (y-axis).

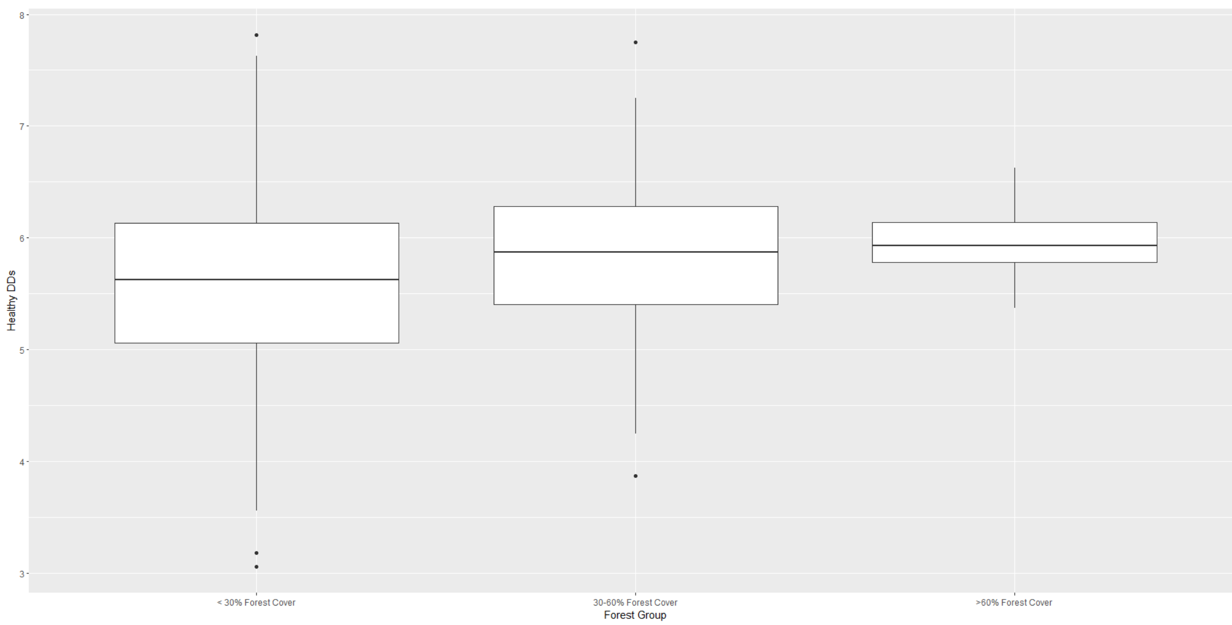

Figure 6, Box Plot of Forest Groups (x-axis, "<30% forest cover", "30-60% Forest Cover" and ">60% Forest Cover") and respective Healthy DDS (y-axis).

To test the potential for significant differences present in the figures, ANOVAs were conducted. An ANOVA was conducted for Healthy DDS and Forest Group. The resulting p-value was 0.03, thus being statistically significant. An ANOVA was also conducted for Combined DDS and Forest Group. The p-value was statistically significant at 2.096e-05.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

To further understand the differences present, t-tests were conducted. The first t-test conducted tested the relationship between 30-60% Forest Cover and >60% Forest Cover means for Combined DDs. The p-value was 0.1394, showing no statistically significant difference. The second t-test conducted was between <30% Forest Cover and >60% Forest Cover means for Combined DDS. The p-value was 0.00433, showing a statistically significant difference. The last t-test conducted tested groups <30% Forest Cover and 30-60% Forest Cover. The p-value was 8.524e-05, showing a statistically significant difference.

t-tests were also conducted to test Healthy DDS. In this case, the first t-test tested the relationship between Healthy DDs and Forest Group between groups < 30% Forest Cover and 30-60% Forest Cover. The alternative hypothesis stated that true difference in means between group <30% Forest Cover and group 30-60% Forest Cover is not equal to 0. The resulting p-value of 0.01164 showed that the difference is statistically significant. The second t-test tested the difference between Healthy DDs and Forest Group for groups 30-60% Forest Cover and >60% Forest Cover. The resulting p-value was 0.6257, showing differences are not statistically significant. Lastly, the last t-test tested Healthy DDs and Forest Group differences between >60% Forest Cover and < 30% Forest Cover. With a p-value of 0.09201, this test showed no statistically significant difference.

Therefore, three tests showed a statistically significant difference. For Combined DDS, these were between groups <30% Forest Cover and >60%, as well as groups <30% Forest Cover and 30-60% Forest Cover. This makes sense visually as well (Figure 5), as 30-60% and 60% are extremely close to one another. In terms of Healthy DDS, the groups which showed statistically significant difference were < 30% Forest Cover and 30-60% Forest Cover. One would expect <30% and >60% to also be statistically significant here, however, this was not the case, albeit a p-value of 0.09 was present, that which is close to <0.05. Moreover, an important factor to reiterate is that deforested land is so prominent in this region that despite some effects between forest cover and dietary diversity being observed, these are not extremely impactful. Regressions were also run to better understand the context of landscape characteristics and dietary diversity in Malawi:

Table 3, Regressions run and respective p-values.

Regression 1 = Healthy DDS and Number of Patches	p = 0.07896
Regression 2 = Combined DDS and Distance to Roads	p = < 2.2e-16
Regression 3 = Combined DDS and Number of Patches	p-value = 0.7663
Regression 4 = Healthy DDS and Distance to Roads	p-value = 4.001e-14

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Regressions 2 and 4 showed statistically significant p-values. This shows that the number of forest patches has no effect on dietary diversity in Malawi. This is likely because of the high deforestation in this area. Albeit patches are likely to increase dietary diversity and resulting food security in general, it appears that Malawi's proportion of patches does not corroborate with this fact. Distance to roads shows a statistically significant difference to Combined DDS and Healthy DDS (Table 2), linking back to that which was previously mentioned in relation to richer households having better accessibility (Figure 2). In conclusion, the data implies that forested areas and patches' effects on food security are minimal, whereas distance to roads shows a significant difference to DDS. However, the relationship of distance to roads would have to be further analysed (perhaps through a visual representation of DDS values and distance to roads) to fully conclude what may be occurring within this context.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

References

Rasmussen, L. V., Fagan, M. E., Ickowitz, A., Wood, S. L. R., Kennedy, G., Powell, B., Baudron, F., Gergel, S., Jung, S., Smithwick, E. A. H., Sunderland, T., Wood, S., & Rhemtulla, J. M. (2020). Forest pattern, not just amount, influences dietary quality in five african countries. *Global Food Security, 25*, 100331. https://doi.org/10.1016/j.gfs.2019.100331

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Appendix

Question 8:

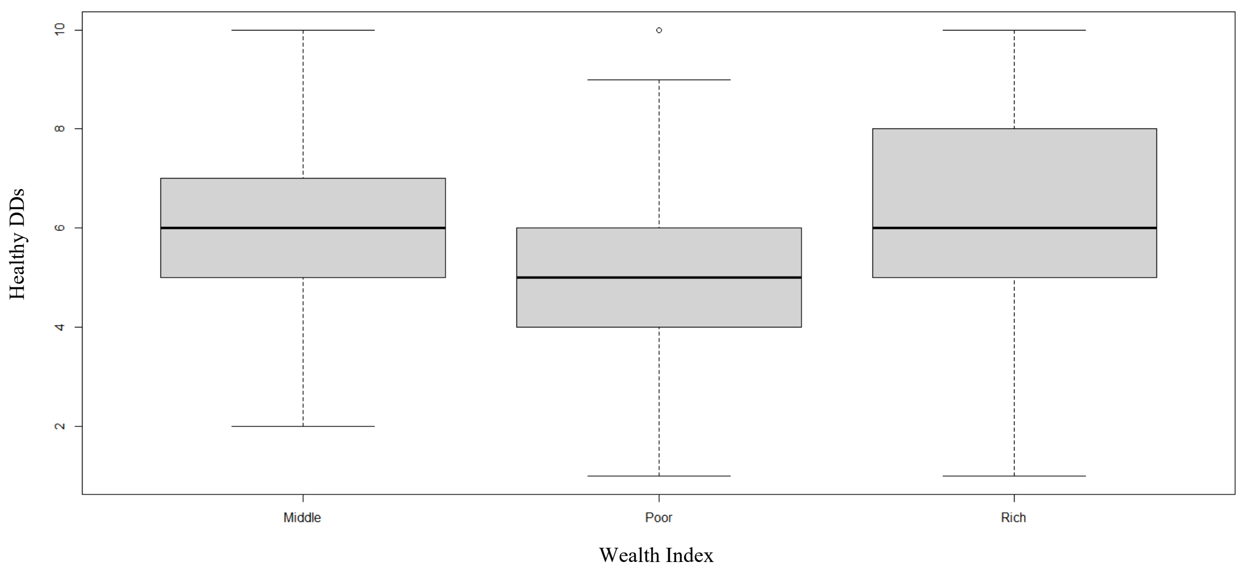

Figure 7, Wealth Index (x-axis, Middle, Poor, and Rich) as a function of Healthy DDs. The data shows that the rich possess a higher Healthy DDS, followed by the Middle Class and the Poor

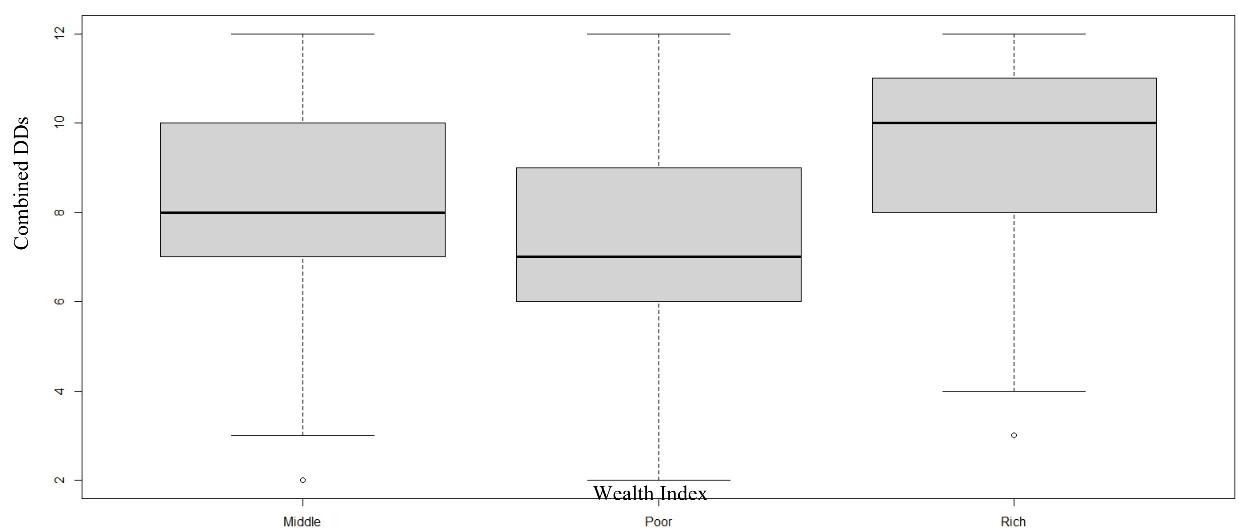

Figure 8, Wealth Index (x-axis, Middle, Poor, and Rich) as a function of Combined DDs. The data shows that the rich possess a higher Combined DDS, followed by the Middle Class and the Poor.

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022
Welch Two Sample t-test

data: Healthy_DDS by wealth.index

t = 15.645, df = 8026.8, p-value < 2.2e-16

alternative hypothesis: true difference in means between group Middle and group Poor is not equal to 0

95 percent confidence interval:

0.4766964 0.6132687

sample estimates:

mean in group Middle   mean in group Poor

5.701811   5.156828

Second t-test

Welch Two Sample t-test

data: Healthy_DDS by wealth.index

t = -17.823, df = 3992.4, p-value < 2.2e-16

alternative hypothesis: true difference in means between group Middle and group Rich is not equal to 0

95 percent confidence interval:

-0.8472951 -0.6793618

sample estimates:

mean in group Middle   mean in group Rich

5.701811   6.465139

Third t-test

Welch Two Sample t-test

data: Healthy_DDS by wealth.index

t = -30.537, df = 3992.8, p-value < 2.2e-16

alternative hypothesis: true difference in means between group Poor and group Rich is not equal to 0

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022
95 percent confidence interval:

-1.392309 -1.224313

sample estimates:

mean in group Poor mean in group Rich

5.156828 6.465139

ANOVA Tests

```
> ES_ANOVA3 = lm(Healthy_DDS ~ wealth.index, data=HH_Malawi)
> ES_ANOVA4 = lm(Combined_DDS ~ wealth.index, data=HH_Malawi)
> anova(ES_ANOVA3)
Analysis of Variance Table

Response: Healthy_DDS
               Df  Sum Sq Mean Sq F value    Pr(>F)
wealth.index    2  2315.8 1157.92  474.19 < 2.2e-16 ***
Residuals   10034 24501.8    2.44
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(ES_ANOVA4)
Analysis of Variance Table

Response: Combined_DDS
               Df Sum Sq Mean Sq F value    Pr(>F)
wealth.index    2   7232  3615.8  923.87 < 2.2e-16 ***
Residuals   10034  39271     3.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 10

(Regression 1)

Coefficients:

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

```
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       5.706e+00  5.037e-02  113.28   <2e-16 ***
Number.of.patches -7.052e-05  4.007e-05   -1.76    0.079 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7987 on 626 degrees of freedom
Multiple R-squared:  0.004922,	Adjusted R-squared:  0.003332
F-statistic: 3.096 on 1 and 626 DF,  p-value: 0.07896
```

(Regression2)

```
Call:
lm(formula = CombinedDDS ~ dist_road, data = ClustersCombined)

Residuals:
    Min      1Q  Median      3Q     Max
-3.8744 -0.6518  0.0396  0.7178  2.8585

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.342291   0.056469 147.733   <2e-16 ***
dist_road   -0.034614   0.003902  -8.871   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.034 on 626 degrees of freedom
Multiple R-squared:  0.1117,	Adjusted R-squared:  0.1103
F-statistic: 78.69 on 1 and 626 DF,  p-value: < 2.2e-16
```

(Regression3)

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

Call:

lm(formula = CombinedDDS ~ Number.of.patches, data = ClustersCombined)

Residuals:

```
    Min      1Q  Median      3Q     Max
-3.5635 -0.7585  0.0562  0.7556  3.0123
```

Coefficients:

```
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       7.985e+00  6.922e-02 115.352   <2e-16 ***
Number.of.patches 1.637e-05  5.507e-05   0.297    0.766
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.097 on 626 degrees of freedom

Multiple R-squared: 0.0001412,	Adjusted R-squared: -0.001456

F-statistic: 0.08839 on 1 and 626 DF, p-value: 0.7663

(Regression4) ##

Call:

lm(formula = HealthyDDS ~ dist_road, data = ClustersHealthy)

Residuals:

```
     Min       1Q   Median       3Q      Max
-2.65106 -0.51318  0.03029  0.52885  2.09697
```

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
```

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

(Intercept) 5.858156   0.041754 140.30   <2e-16 ***

dist_road   -0.022331   0.002885   -7.74   4e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7649 on 626 degrees of freedom

Multiple R-squared: 0.08734,   Adjusted R-squared: 0.08589

F-statistic: 59.91 on 1 and 626 DF, p-value: 4.001e-14

t-tests

T_test4 = subset(ForestCoverAnova, ForestCoverAnova$ForestGroup == "< 30% Forest Cover" | ForestCoverAnova$ForestGroup == "30-60% Forest Cover")

> t.test(HealthyDDS ~ ForestGroup, data=T_test4)

Welch Two Sample t-test

data: HealthyDDS by ForestGroup

t = -2.583, df = 78.917, p-value = 0.01164

alternative hypothesis: true difference in means between group < 30% Forest Cover and group 30-60% Forest Cover is not equal to 0

95 percent confidence interval:

-0.45986195 -0.05957327

sample estimates:

mean in group < 30% Forest Cover mean in group 30-60% Forest Cover

5.608140   5.867857

> T_test5 = subset(ForestCoverAnova, ForestCoverAnova$ForestGroup == "30-60% Forest Cover" | ForestCoverAnova$ForestGroup == ">60% Forest Cover")

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022

```
> t.test(Healthy_DDS ~ ForestGroup, data=T_test5)

Error in eval(predvars, data, env) : object 'Healthy_DDS' not found

> t.test(HealthyDDS ~ ForestGroup, data=T_test5)


	Welch Two Sample t-test


data:  HealthyDDS by ForestGroup

t = -0.50627, df = 8.4103, p-value = 0.6257

alternative hypothesis: true difference in means between group 30-60% Forest Cover and group >60% Forest Cover is not equal to 0

95 percent confidence interval:

 -0.5489155  0.3499076

sample estimates:

mean in group 30-60% Forest Cover   mean in group >60% Forest Cover

                         5.867857                          5.967361




> T_test6 = subset(ForestCoverAnova, ForestCoverAnova$ForestGroup == ">60% Forest Cover" | ForestCoverAnova$ForestGroup == "< 30% Forest Cover")

> t.test(HealthyDDS ~ForestGroup, data=T_test6)


	Welch Two Sample t-test


data:  HealthyDDS by ForestGroup

t = -2.0457, df = 5.3975, p-value = 0.09201

alternative hypothesis: true difference in means between group < 30% Forest Cover and group >60% Forest Cover is not equal to 0

95 percent confidence interval:

 -0.8007861  0.0823429
```

Cons 452, Lab 4
Marina Galvao – [redacted]
February 28th, 2022
sample estimates:

mean in group < 30% Forest Cover	mean in group >60% Forest Cover
5.608140	5.967361