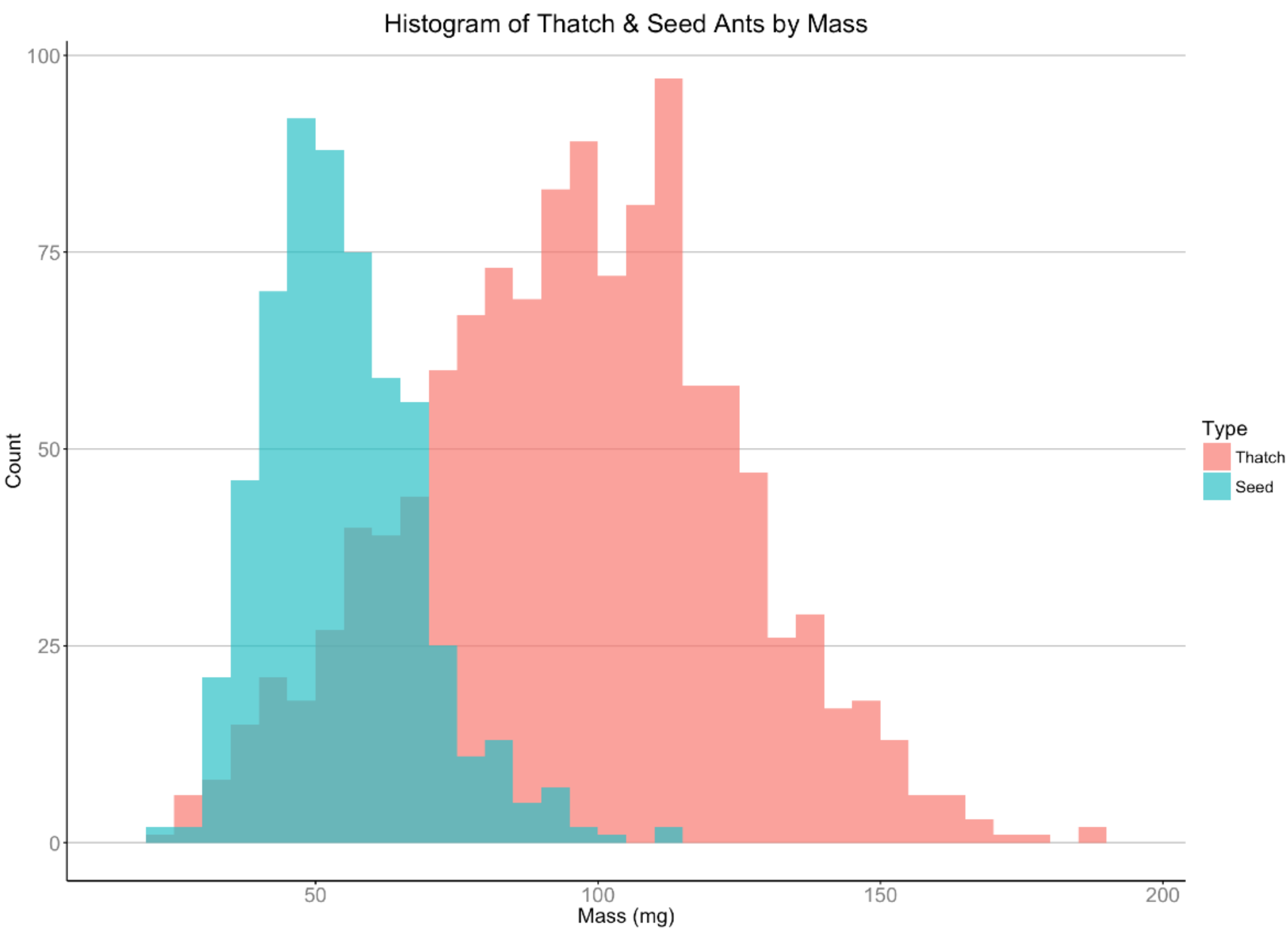
STA303
Assignment 1
by Justin SJ Lee

Let's investigate some data collected by Peter Nonacs from UCLA. The dataset and a data dictionary and some relevant background information can be found here. You'll need to do some mild cleaning first. In particular, you should set appropriate levels for the size factor, and make Colony a factor as well. Distance you can leave as numeric. For this assignment, you may ignore the fact that ants from the same colony are likely not independent.

1. Let's see if thatch ants are different from seed ants, on average.

(a) Present a layered histogram of mass (in mg) with one layer for Thatch ants and one for Seed ants.



(b) Do a simple analysis to determine if the observed difference is statistically significant. If it is, give the magnitude, direction and evidence of the effect.

We can see if the difference is statistically significant by implementing a Two Sample t-test regarding the mass of thatch ants and seed ants.

Results: t-statistic = 41.451 with 1770 degrees of freedom. The p-value is less than 2.2e-16.

Ho = True difference in mean mass of thatch ants and seed ants is equal to 0.
Reject Ho, since p-value is less than 0.0001 and we have irrefutable evidence against Ho.

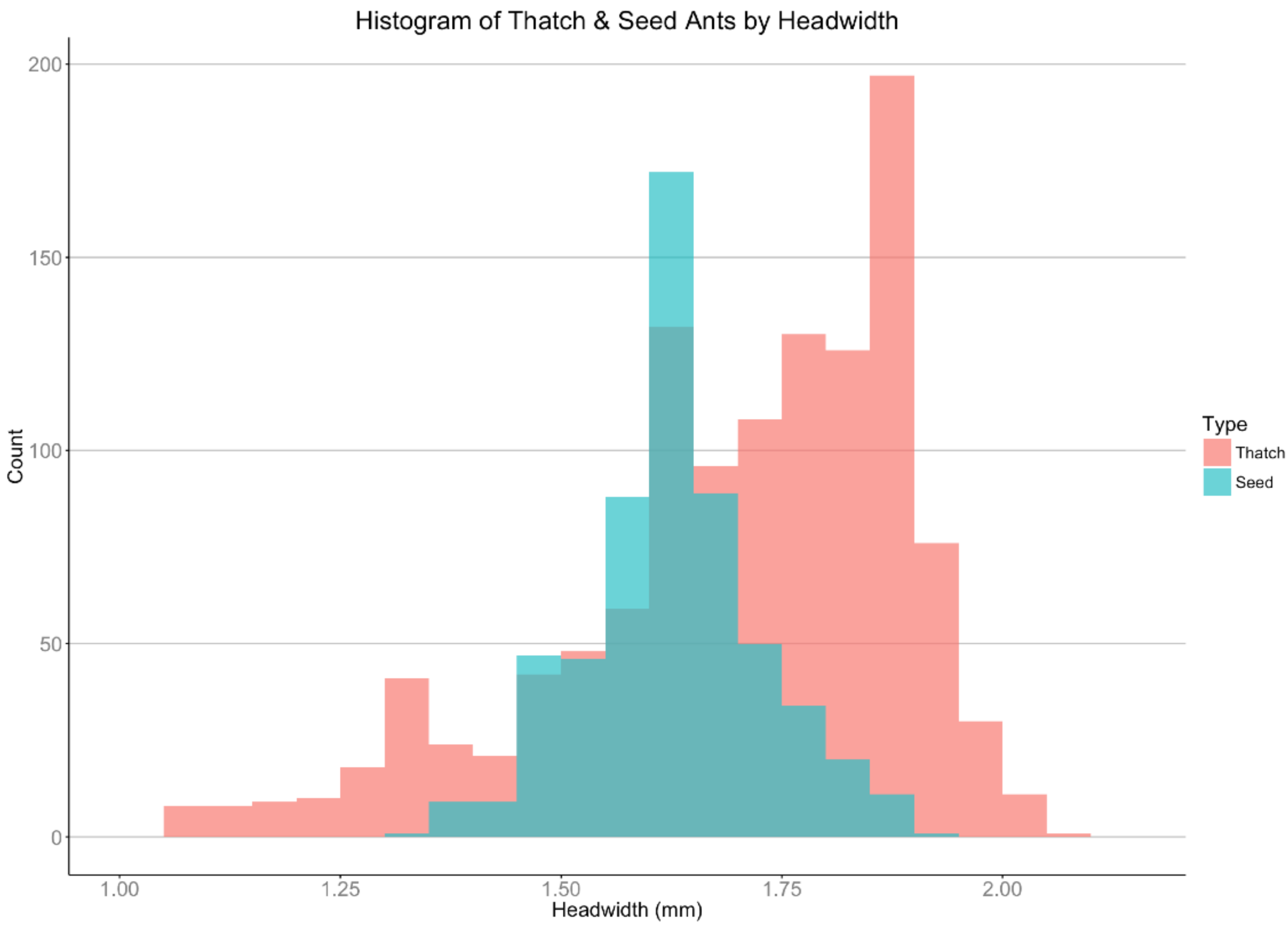
We conclude that true difference in mean mass of thatch ant and seed ant is not equal to 0.

Thus we can conclude that the observed difference is significantly significant: **Thatch ants have a mean mass 41.451 greater than that of seed ants.**

Out of 95% of trials, the observed difference between the means of thatch ant mass and seed ant mass will be between 39.04723 and 42.92584.

(c) Repeat the previous two parts using Headwidth in mm instead of Mass.

We can see if the difference is statistically significant by implementing a Two Sample t-test regarding the mass of thatch ants and seed ants.



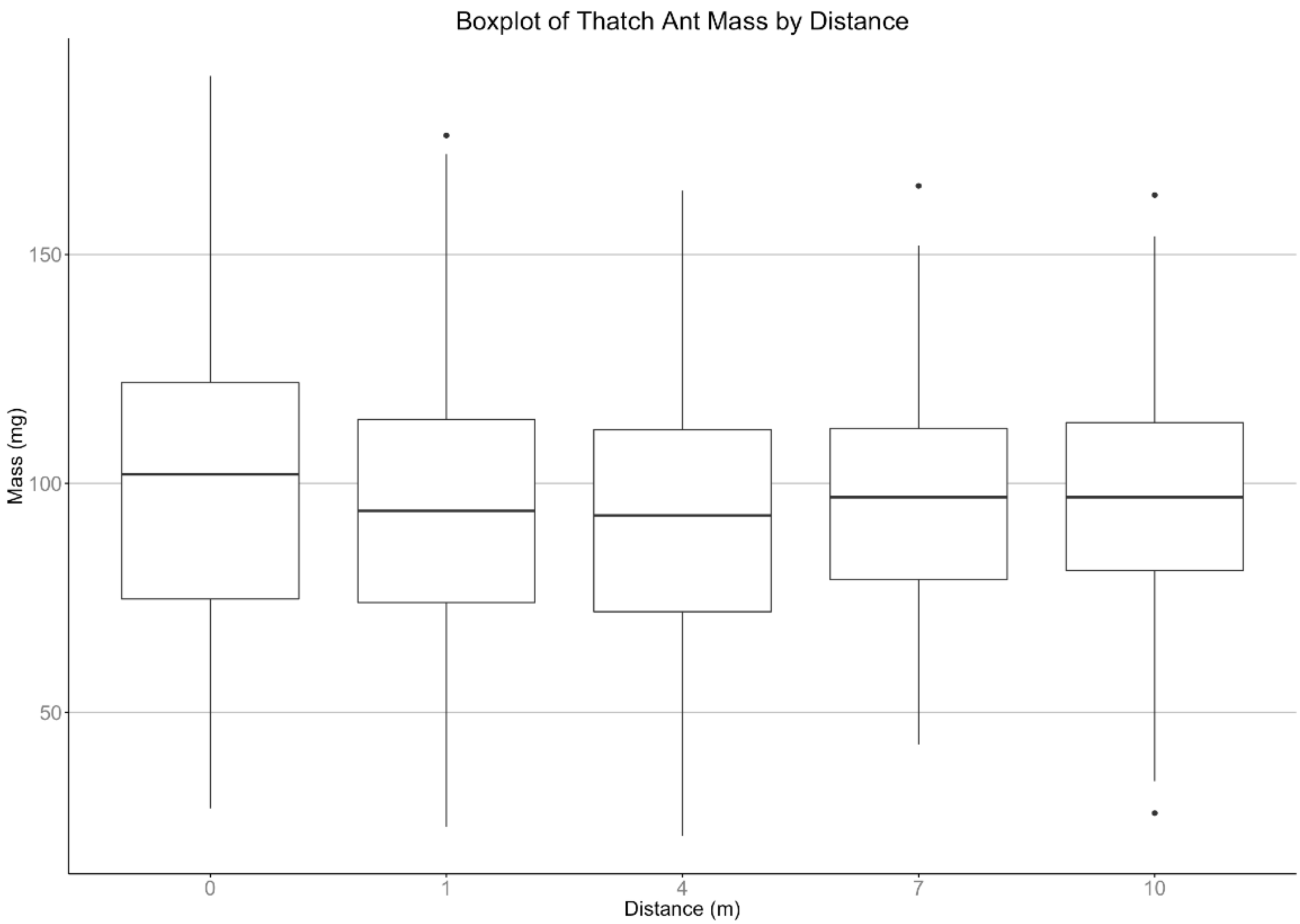
Results: t-statistic = 10.591 with 1761.2 degrees of freedom. The p-value is less than 2.2e-16.
Ho = True difference in mean headwidth of thatch ants and seed ants is equal to 0.
Reject Ho, since p-value is less than 0.0001 and we have irrefutable evidence against Ho.
We conclude that true difference in mean headwidth of thatch ant and seed ant is not equal to 0.

Thus we can conclude that the observed difference between the mean headwidth of thatch ants and seed ants is significantly significant.

Out of 95% of trials, the observed difference between the means of thatch ant mass and seed ant mass will be between 0.06082629 and 0.08847647.

2. Let's look for differences in Mass between several groups.

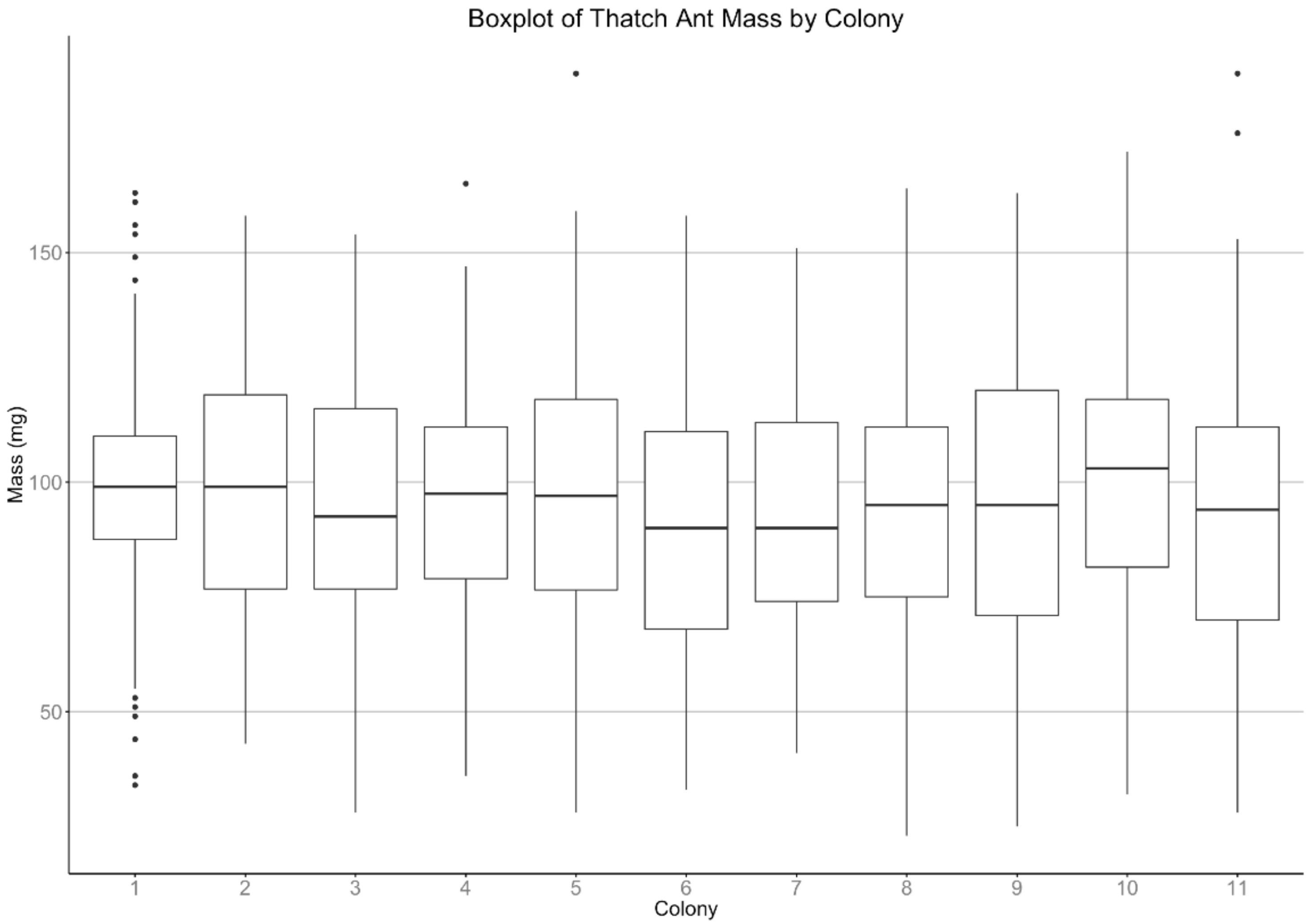
(a) Looking at the Thatch ants only and treating Distance as a factor, present a boxplot of Mass vs. Distance and do some analysis to determine if the average ant Mass differs based on distance traveled. If you decide that the groups are not all the same, follow the analysis up with some Tukey HSD corrected tests to determine which distances are different from each other.



We check out null hypothesis that the difference between means of each Distance group is 0 by using ANOVA comparisons of means with a 95% confidence level.

By analyzing the p-value from ANOVA: 0.1284, we see that it is well above 0.1 and can conclude that there is no evidence against Ho. Thus the mean mass of each (distance) group is not very different from the other groups. The boxplot also supports our conclusion since the means are all around the same level and they all lie in similar 25th & 75th quantiles.

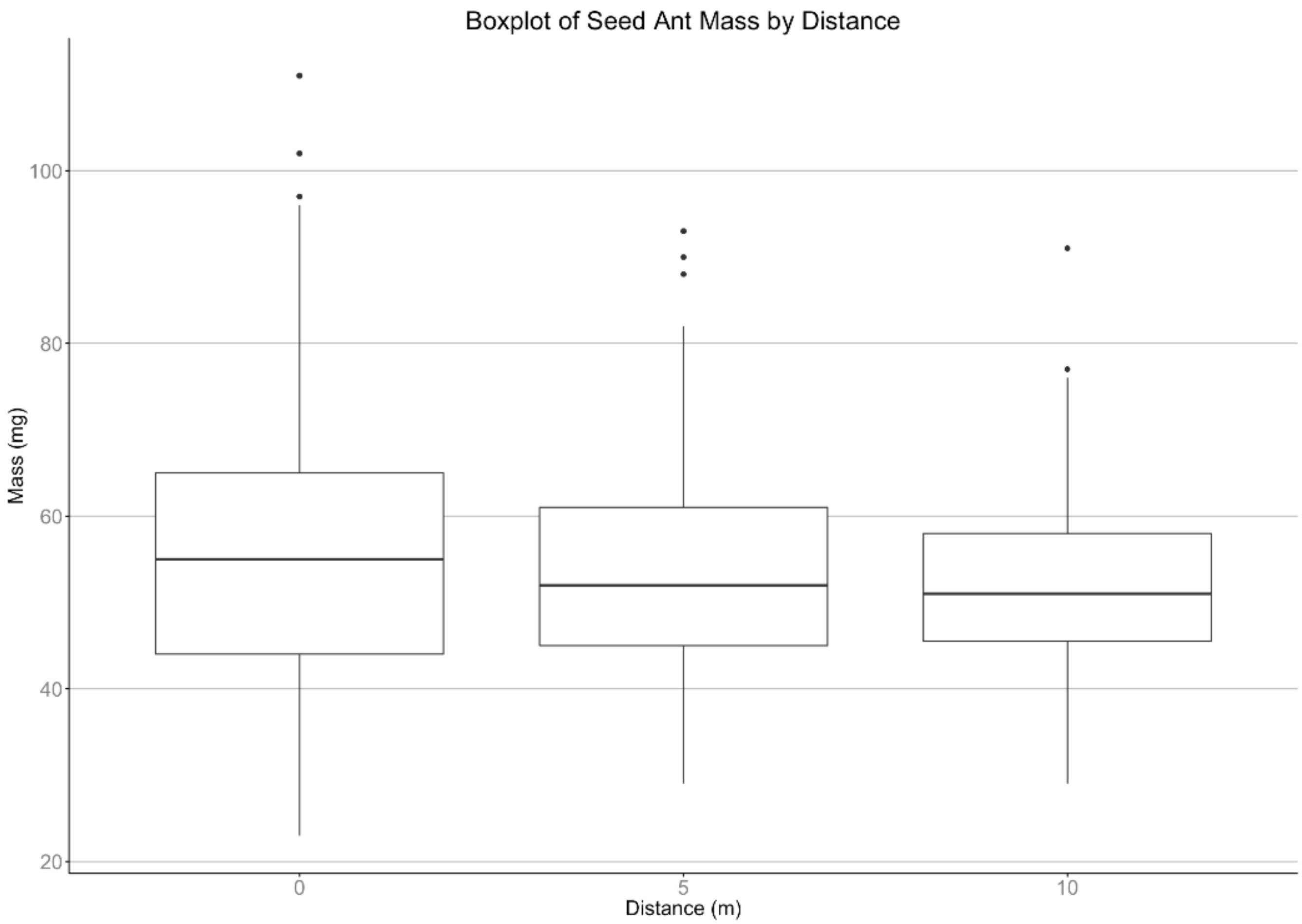
(b) Repeat the previous analysis using Colony as the independent variable instead



We check out null hypothesis that the difference between means of each colony is 0 by using ANOVA comparisons of means with a 95% confidence level.

By analyzing the p-value from ANOVA: 0.1493, we see that it is well above 0.1 and can conclude that there is no evidence against Ho. Thus the mean mass of each colony is not very different from the other groups. The boxplot also supports our conclusion since the means are all around the same level and they all lie in similar 25th & 75th quantiles.

(c) Repeat the previous two analyses, this time using the Seed ants instead



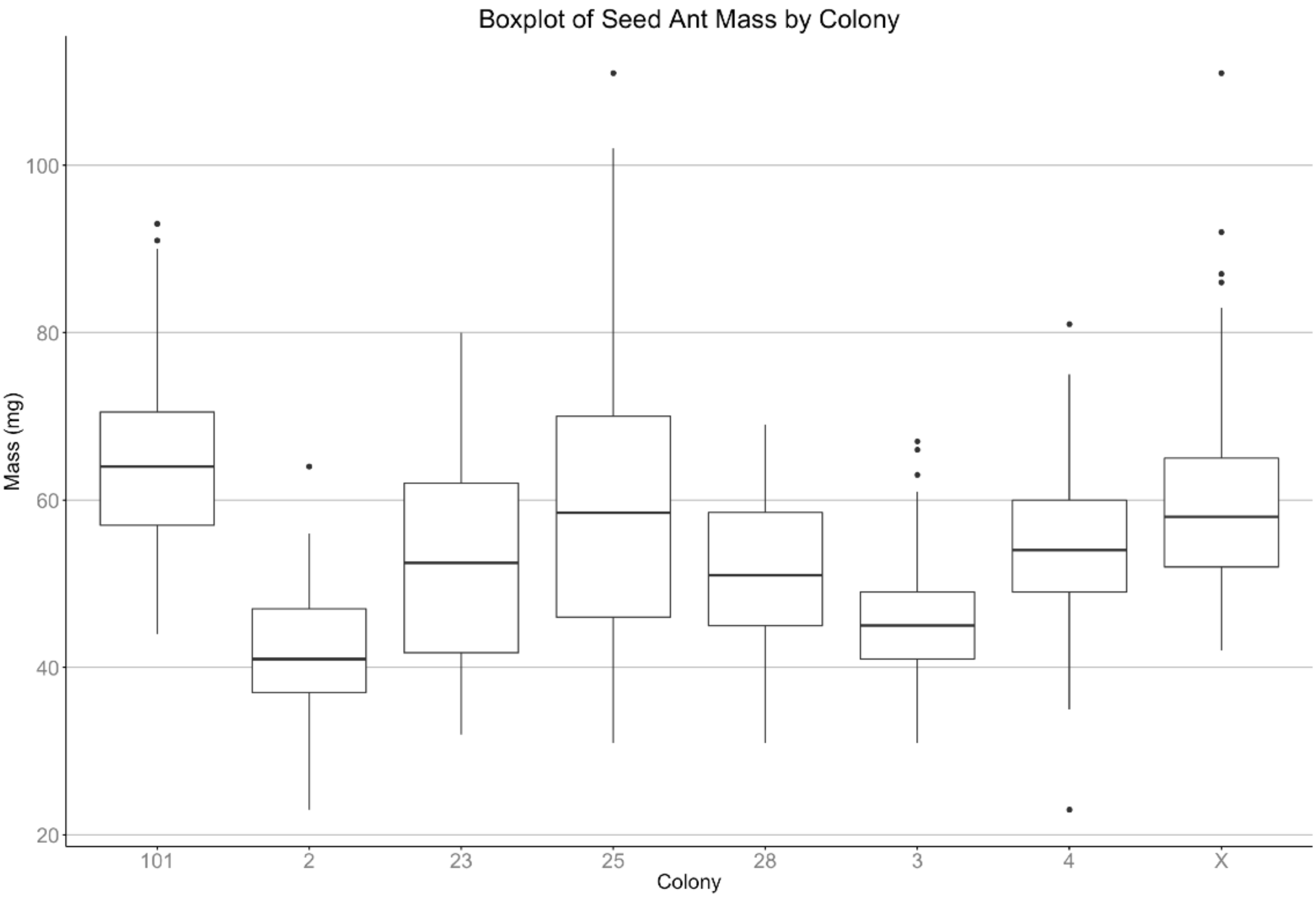
We check out null hypothesis that the difference between means of each distance group is 0 by using Tukey multiple comparisons of means with a 95% family-wise confidence level.

By analyzing the adjusted p - values for all pairs, we see that the adjusted p - value for the difference between mean mass from groups (distance = 5, distance = 0) and groups (distance = 10, distance = 0) have values 0.0265577 and 0.0098107 respectively.

From this we can conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from group 0 and group 5 is equal to 0 with moderate evidence.

We also conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from group 0 and group 10 is equal to 0 with strong evidence.

However, we fail to reject the hypothesis that the true difference in mean mass of seed ants from group 10 and group 5 is equal to 0.



We check out null hypothesis that the difference between means of each colony is 0 by using Tukey multiple comparisons of means with a 95% family-wise confidence level.

From this we can conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from colony pairs (2-101), (23-101), (28-101), (3-101), (4-101), (23-2), (25-2), (28-2), (4-2), (X-2), (3-25), (X-28), (X-3) with irrefutable evidence.

From this we can conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from colony pairs (4-3), (28-25), (X-23) is equal to zero with very strong evidence.

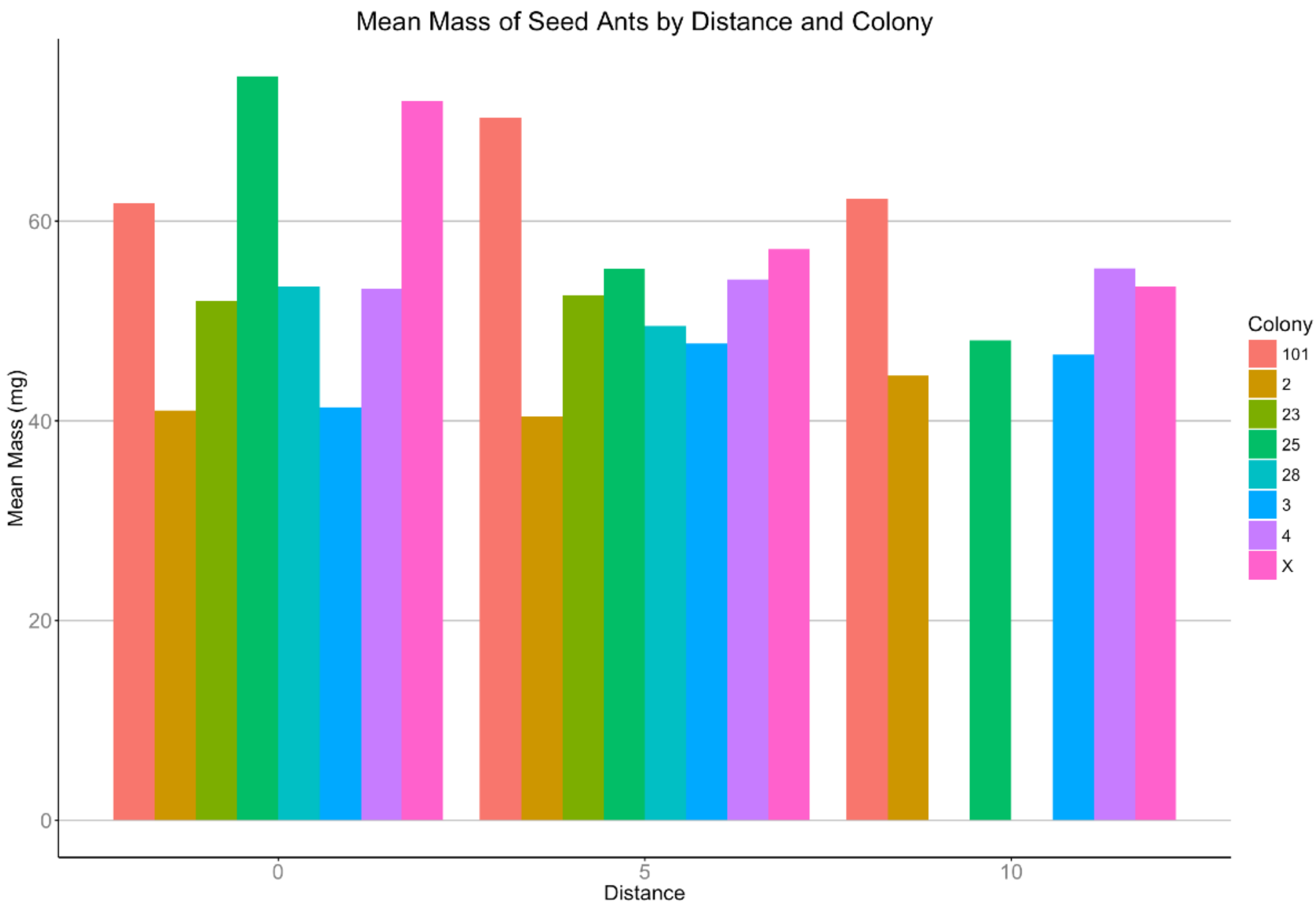
From this we can conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from colony pairs (X-4), (25-23) is equal to zero with strong evidence.

From this we can conclude that we can reject the null hypothesis that the true difference in mean mass of seed ants from colony pairs (3-23), (3-28), (4-25) is equal to zero with moderate evidence.

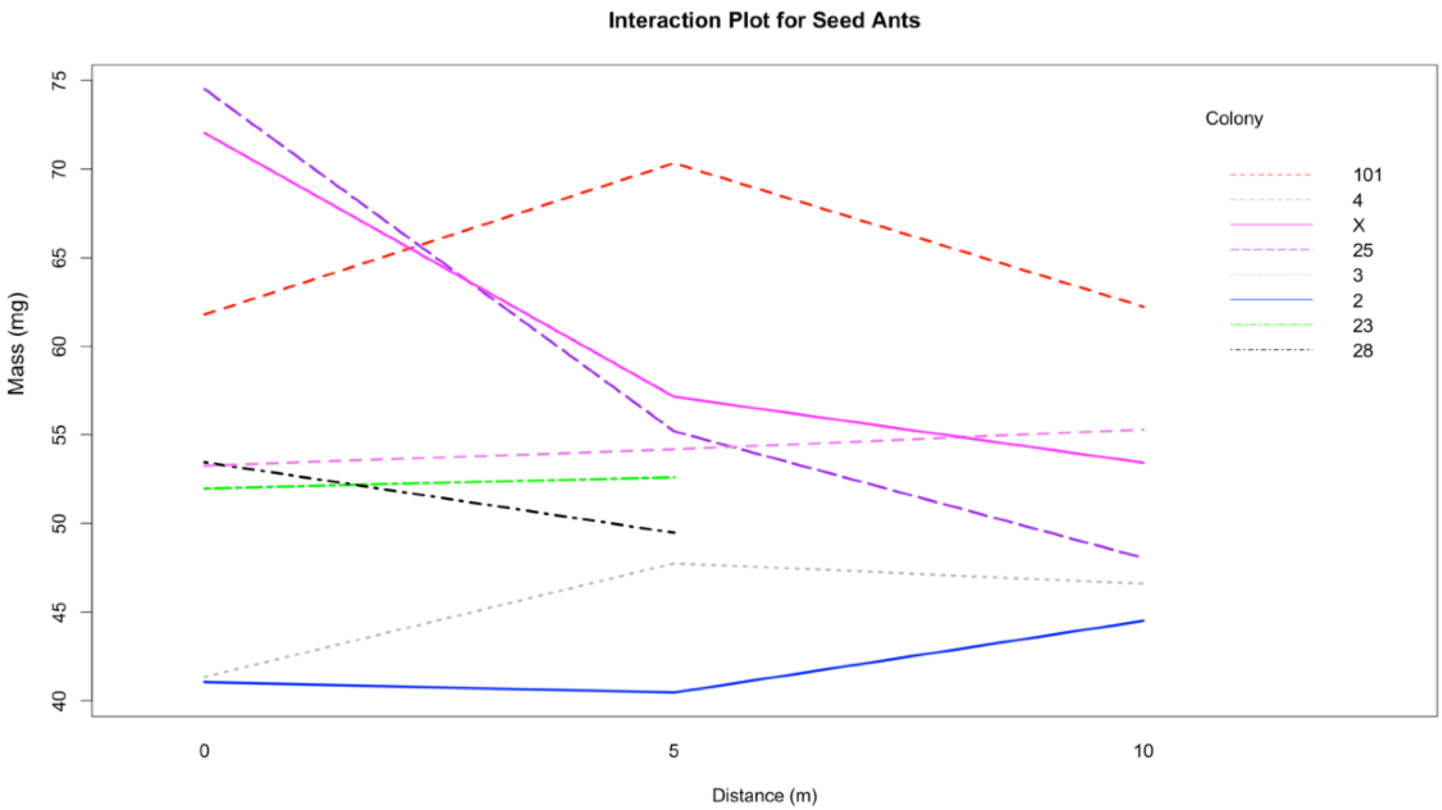
However, we fail to reject the hypothesis that the true difference in mean mass of seed ants from pairs from colony (25-101), (X-101), (3-2), (28-23), (4-23), (X-25), (4-28) is equal to zero.

3. The last part does not take into account both factors at the same time. Now, let's model the Mass as a function of both Distance and Colony and look at the Seed ants only.

(a) Present an appropriate plot showing how the means of Mass differ for different levels of Colony and Distance.



The bar plot of the mean mass by each distance level shows that the colonies have different mean mass. The interaction plot below also supports this statement by showing how the mean mass are different for each level of colony.



(b) Present a 2D table showing the group means of Mass for each combination of Colony and Distance. Round the output to one decimal place.

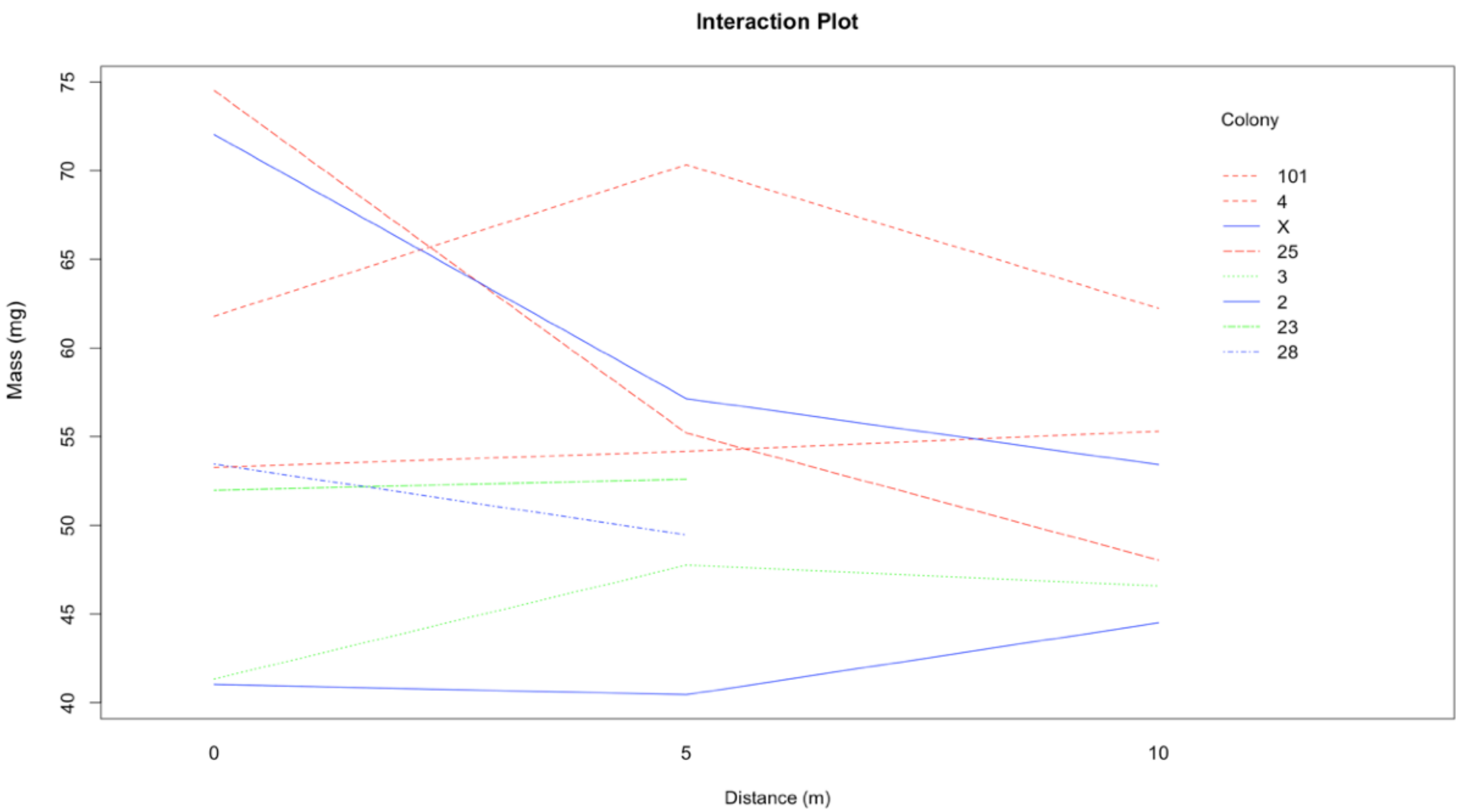
Colony\Distance	0m	5m	10m
101	61.8 mg	70.3 mg	62.2 mg
2	41.0 mg	40.4 mg	44.5 mg
23	52.0 mg	52.6 mg	
25	74.5 mg	55.2 mg	48.0 mg
28	53.4 mg	49.5 mg	
3	41.3 mg	47.8 mg	46.6 mg
4	53.3 mg	54.2 mg	55.3 mg
X	72.0 mg	57.1 mg	53.4 mg

The table above shows the **mean mass** of **seed ants** for distances 0m, 5m, 10m (measured from the entrance of ant mound) and the type of colony, the ant belongs to. The 2 occurrences of blank cells are due to the lack of data for ants from Colony 23 and 28 found at 10m from their mounds.

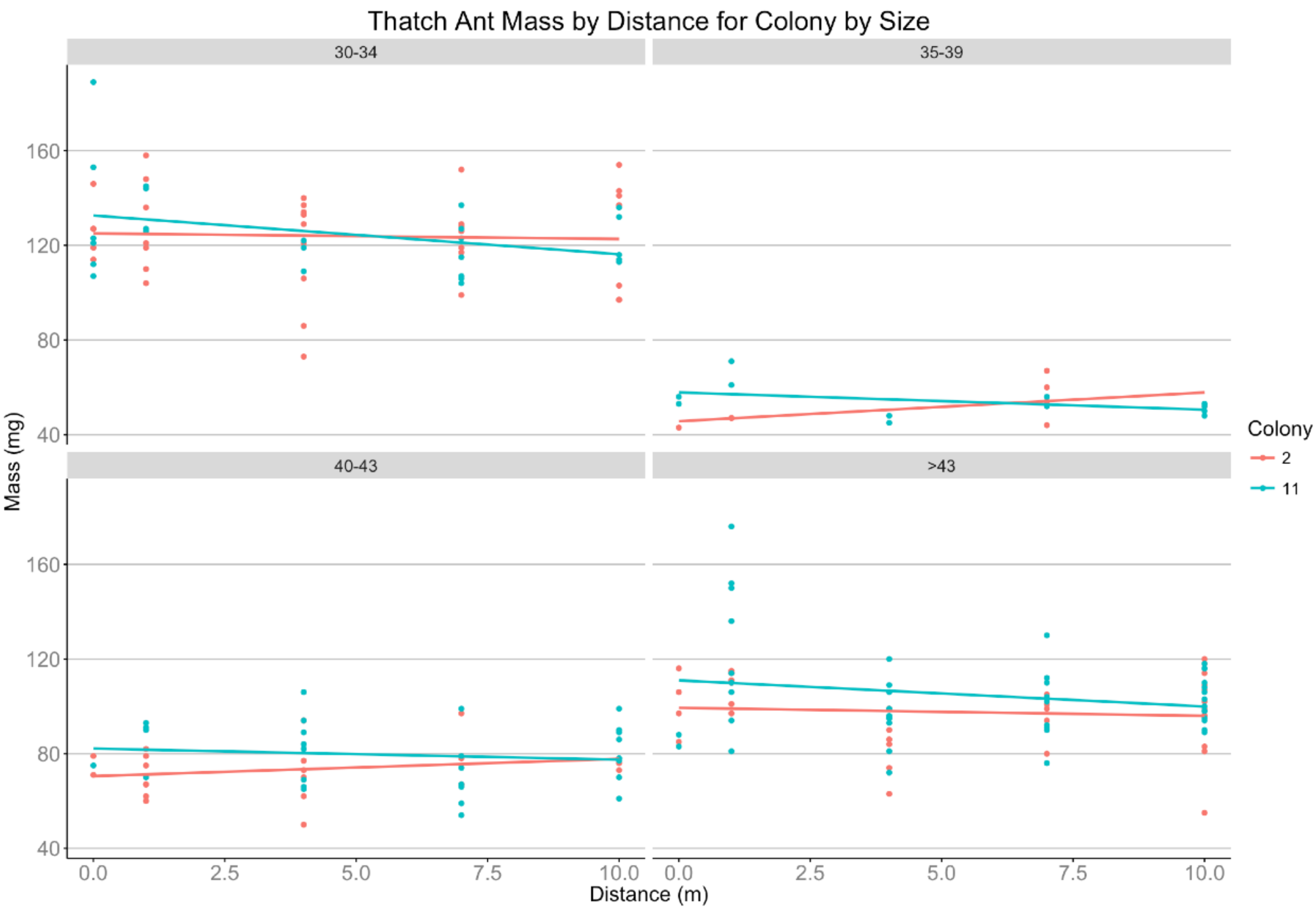
(c) Do some analysis to determine if Colony and Distance are useful factors to predict the Mass of a Seed ant. Give your conclusion in plain English.

By performing 2-way ANOVA on a linear regression model: Mass ~ Distance + Colony + Distance:Colony

The corresponding p-values for Distance, Colony and their interaction are 6.596e-05, below 2.2e-16 and below 2.2e-16. Thus we have irrefutable evidence that Distance, Colony and the interaction are useful predictors in our model where Mass is the response variable. The interaction.plot plot also supports this statement since from the shape of the graphs, there is a parallel relationship. Thus, an interaction effect exists in our model.



4. Using ggplot2, show the relationship between Mass and Distance (numeric) for each size classification factor level, for thatched ant colonies 2 and 11 only. Comment on the foraging strategy for each colony, being as detailed as your plots will allow.

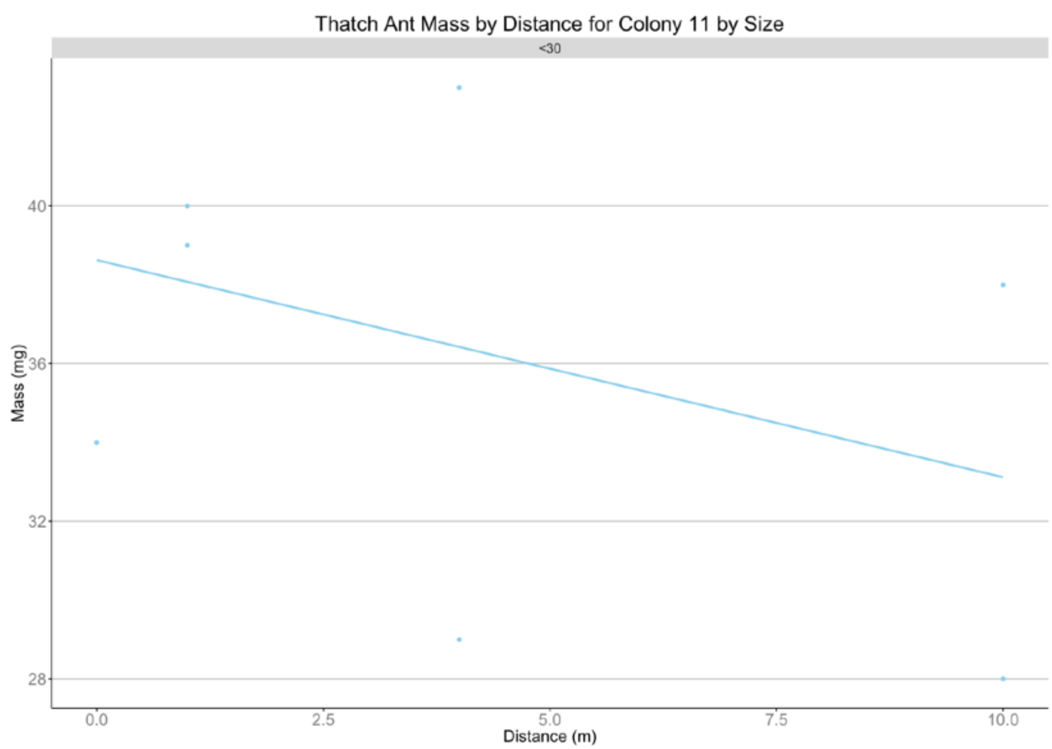


The above graph shows the relationship between mass and distance foraged from the mound's entrance for each colony. I've attached the graph for **<30** separately below since there is only data available for colony 2 and the main objective of our study is to see if different colony 2 and 11 had different strategies for sending workers out to gather food.

	Size < 30	Size: 30-34	Size: 35-39	Size: 40-43	Size > 43
Mass (Colony 11)	Decreasing	Decreasing	Decreasing	Decreasing	Decreasing
Mass (Colony 2)	No Data	Decreasing*	Increasing	Increasing	Decreasing*

The above table shows whether mass is increasing/decreasing for each class (size) for each colony as distance goes from 0m -> 10m. The * means that the slope (status) might change with more data since the slope for the graphs is very close to 0 for colony 2 at 30-34 and > 43. We can see that the foraging strategy for Colony 11 is consistent: Decreasing Mass for ants for ants travelling further from mound for every size class. If we think of mass as a good indicator for how much energy an ant is carrying (more mass ~ more energy), then we can conclude that Colony 11 of Thatch Ants have an **"Energy Centric"** mentality because they send out ants with less energy further away from the mound where they have a higher chance of not coming back. However, since they were not carrying that much energy in the first place, it could be better for the colony if they didn't come back (with no food/resources for the colony).

For Colony 2, the situation looks more complicated. Since our sample size is relatively small, if we had more data, I feel that we could get a better slope for the * values. However, from what we have, we can say that for small ants (Size: 30-34), the ants had fairly constant mass when they were further away from the mound. However, for (Size: 35-43), the ants had increasing mass when they were further away from the mound. From (Size: > 43), the slope is fairly flat again and the mass of ants is very slightly decreasing as they get further from the mound. We can conclude, that Colony 2 is **"Worker-Centric"** for ants in the (Size: 35-43) range, since they make ants with more energy (mass), forage further from the mound. With more mass, the ants would be more likely to survive, especially when they get further from the mound because they will likely be stronger and less likely to starve than ants with less mass. However, the strategy seems to change after Size > 43, when the slope is very slightly decreasing. We can say that Colony 2 shifts to a **"Energy-Centric"** strategy for Size > 43 ants. If we could increase our sample size, we could produce more accurate conclusions.



to the left is the graph for <30 for Colony 11. (added separately since there is no data for Colony 2) The shape of the line shows that mass is decreasing for colony 11 as distance increases for size < 32