Appendix: Principle Components Analysis (PCA)

Principle Components Analysis

This is the appendix of principle components analysis plot output. This section is organized as follows:

- Numerical Tables for Plot Generation
- Scree Plots
- Individual Plots
- Variable Plots
- Biplot graphs
- Individual Plots with Overlaid Categorical Ellipses
- R Code

Numerical Tables for Plot Generation

Three types of tables are calculated in order to derive the points for plot generation. These are:

- **Coordinates** - The coordinates of each variable that are used to create a scatter plot. Note, there are seven values for each variable, meaning we can plot them into seven dimensions. For ease of interpretation, we will limit ourselves to just two dimensions.
- **Quality of Represenation** - This is the square of the coordinate of each variable. This number is the sum of square distance of the observation when the original data point is projected onto a component.
- **Contribution** - For each relative principal component, this is a variable's percent contribution to that specific dimension.

PCA Coordinates

Table 1: Pre-Construction Variable Coordinates

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	-0.1229866	0.7842005	-0.4306890	0.2521996	-0.2768383	-0.0384252	0.2066162
count	0.2551599	0.0721583	0.6594669	0.6639405	-0.1864401	0.1384550	0.0065798
hour	0.3897387	0.4127161	0.3268349	-0.6341311	-0.2316353	0.3391538	0.0120669
pcip	-0.4220547	0.6568835	0.2182412	0.0396880	0.5743218	0.0903102	-0.0562858
tmax	0.8088372	-0.2189521	-0.1757404	0.0937489	0.3428593	0.2513561	0.2782739
tmin	0.7972627	0.2474817	-0.3807353	0.1933024	0.0179767	0.0915389	-0.3348092

Table 2: Post-Construction Variable Coordinates

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	-0.3918502	0.3414729	-0.5260843	0.1940435	-0.6425560	-0.0458815	-0.0211850
count	-0.0862071	0.5312588	-0.6146900	-0.2258718	0.4462131	0.2868939	0.0075080
hour..	0.1341212	0.6304476	0.5365995	0.2002354	-0.1474247	0.4837964	-0.0268525
pcip	-0.5007996	0.1874130	0.2684132	-0.7732041	-0.1778896	-0.0716370	-0.0860778
tmax	0.9313850	-0.0141383	-0.1679939	-0.0989815	-0.0957820	-0.0193320	-0.2911269
tmin	0.8617055	0.0889783	-0.0770919	-0.3329779	-0.2568937	0.0190124	0.2576301

Numerical Tables for Plot Generation (Continued)

Variable Quality of Representation

Table 3: Pre-Construction Variable Quality of Representation

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	0.0151257	0.6149704	0.1854930	0.0636047	0.0766395	0.0014765	0.0426903
count	0.0651066	0.0052068	0.4348967	0.4408170	0.0347599	0.0191698	0.0000433
hour	0.1518963	0.1703346	0.1068210	0.4021222	0.0536549	0.1150253	0.0001456
pcip	0.1781302	0.4314960	0.0476292	0.0015751	0.3298455	0.0081559	0.0031681
tmax	0.6542177	0.0479400	0.0308847	0.0087889	0.1175525	0.0631799	0.0774363
tmin	0.6356278	0.0612472	0.1449594	0.0373658	0.0003232	0.0083794	0.1120972

Table 4: Post-Construction Variable Quality of Representation

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	0.1535466	0.1166037	0.2767647	0.0376529	0.4128782	0.0021051	0.0004488
count	0.0074317	0.2822359	0.3778438	0.0510181	0.1991061	0.0823081	0.0000564
hour..	0.0179885	0.3974641	0.2879391	0.0400942	0.0217340	0.2340590	0.0007211
pcip	0.2508002	0.0351236	0.0720456	0.5978446	0.0316447	0.0051319	0.0074094
tmax	0.8674780	0.0001999	0.0282220	0.0097973	0.0091742	0.0003737	0.0847549
tmin	0.7425363	0.0079171	0.0059432	0.1108743	0.0659944	0.0003615	0.0663733

Variable Contribution

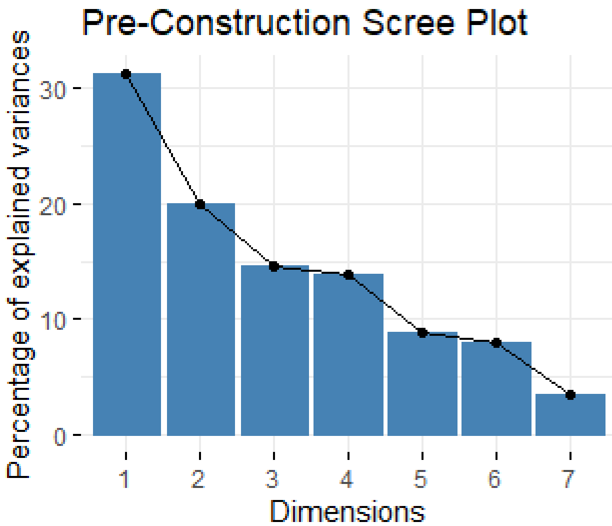
Table 5: Pre-Construction Variable Dimension Contribution (%)

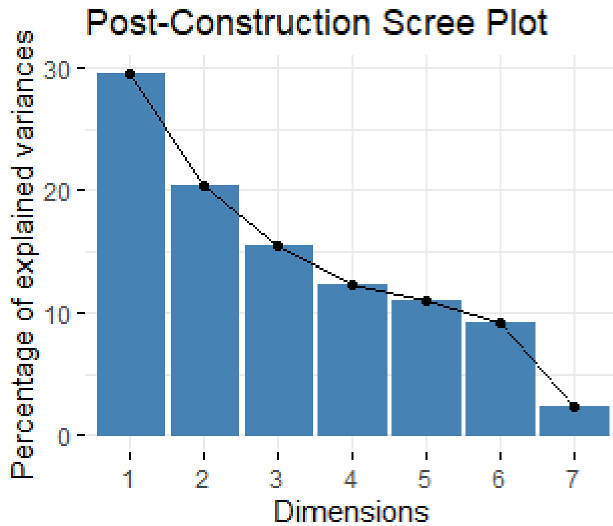
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	0.6915456	44.1313877	18.111798	6.5124343	12.3835031	0.2631166	17.9062149
count	2.9766660	0.3736513	42.463924	45.1349260	5.6165492	3.4161245	0.0181594
hour	6.9446840	12.2235186	10.430157	41.1729987	8.6696295	20.4979366	0.0610755
pcip	8.1440944	30.9649315	4.650583	0.1612770	53.2968557	1.4534173	1.3288410
tmax	29.9107713	3.4402623	3.015624	0.8998848	18.9942870	11.2588932	32.4802774
tmin	29.0608467	4.3952106	14.154039	3.8258599	0.0522169	1.4932337	47.0185931

Table 6: Post-Construction Variable Dimension Contribution (%)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	7.4295347	8.1907856	25.7628526	4.354842	53.653097	0.3282287	0.2806589
count	0.3595900	19.8255560	35.1718859	5.900627	25.873632	12.8334585	0.0352512
hour..	0.8703948	27.9197238	26.8030303	4.637201	2.824316	36.4944207	0.4509127
pcip	12.1352683	2.4672459	6.7064228	69.145262	4.112198	0.8001575	4.6334461
tmax	41.9739552	0.0140413	2.6270633	1.133137	1.192177	0.0582712	53.0013553
tmin	35.9285045	0.5561370	0.5532236	12.823452	8.575897	0.0563602	41.5064255

Scree Plots

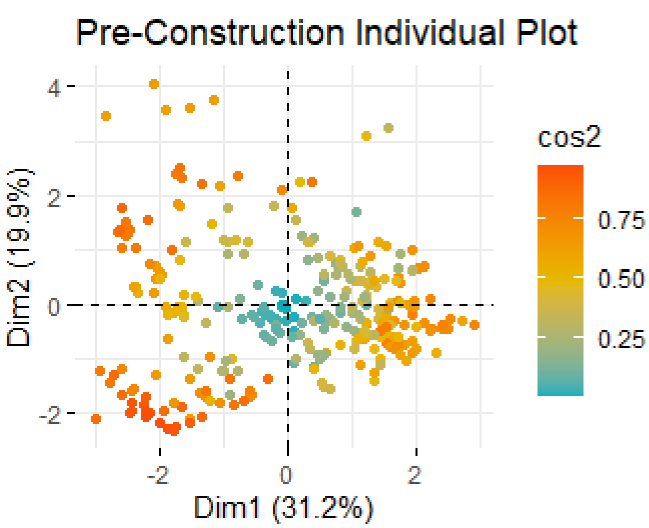


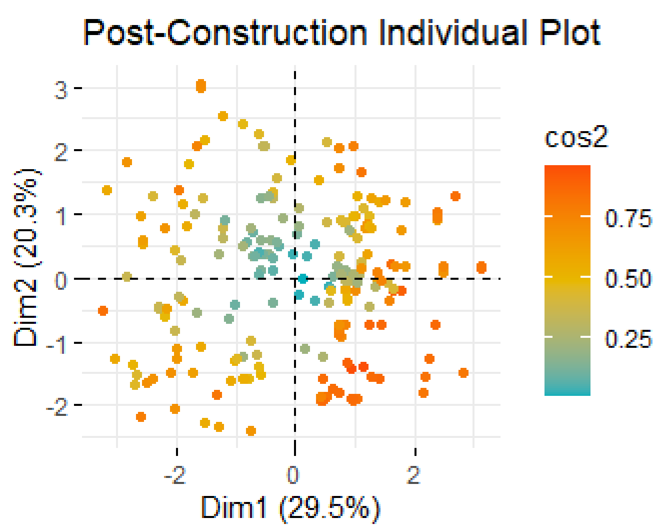


As seen above, the scree plots for both pre and post-construction show many principle dimensions. This is not ideal as we cannot readily visualize k=4 or k=6 on a two-dimensional plane. Although there will exist a fair amount of noise in our dataset, we can still utilize PCA to visualize variance clustering.

The following plots will only be shown in two dimensions for ease of interpretation. For three-dimensional principle components plots, please refer to the R-code at the end of this section.

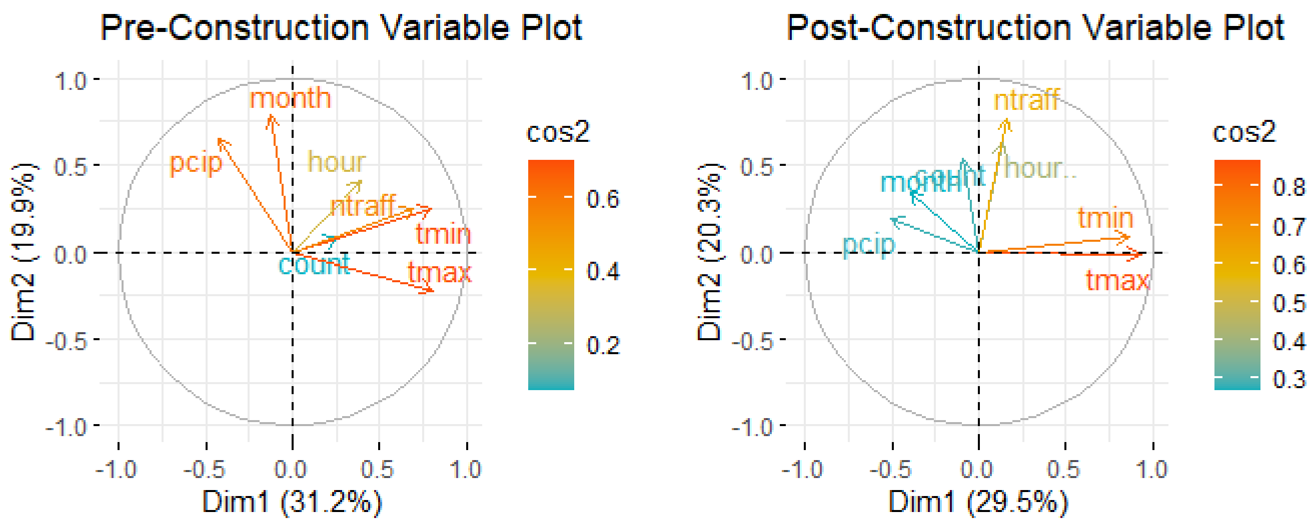
Principal Components Individual Plots





These are the principle components individual contribution plots. The color indicates greater contribution to the dimensions in question. In this case, red means that an individual observation contributes a greater percentage of variation to a particular dimension. Conversely, the turqoise coloring indicates that a particular observation does not contribute much at all.

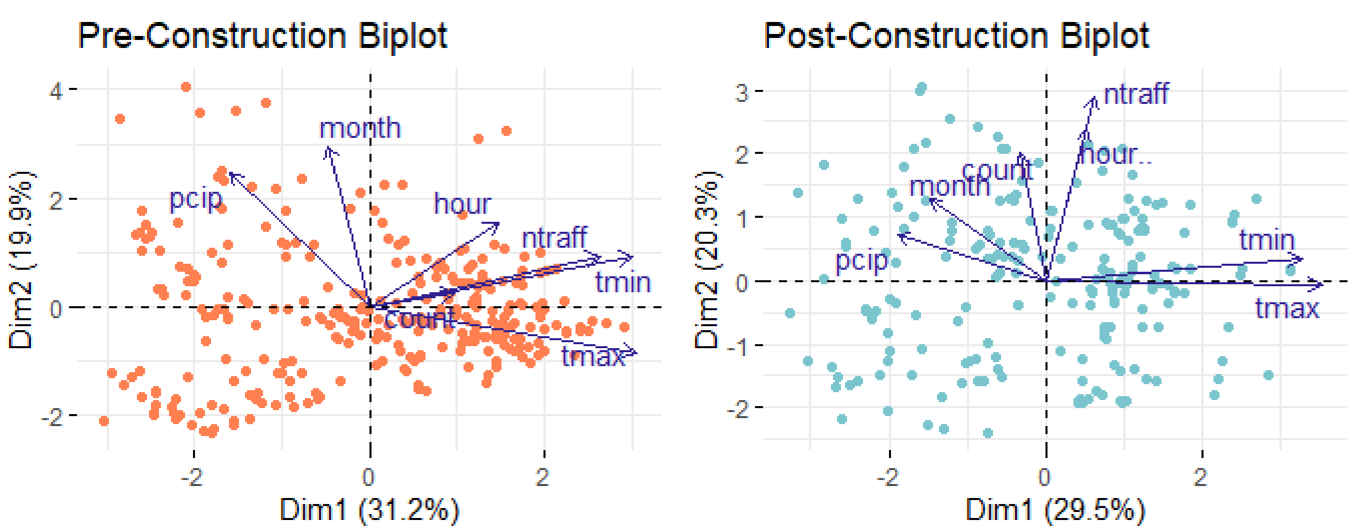
Principal Components Variable Plots



Similar to the individual contribution plots above, these variable plots measure only the contribution that each collected variable provides relative to each other. Arrows that point away from each other indicate non-correlated data. Arrows that point in the same direction indicate more correlated data.

In this particular case, we note that precipitation and month are closely correlated. This makes sense give that precipitation amounts are often seasonal and go hand-in-hand with months. Additionally, we see strong correlation between daily minimum and maximum temperature. This is not surprising as we expect to see fairly consistent differences between daily highs and lows. Traffic appears to correlated with hour in both plots, again, this is to be expected.

Principal Components Biplots



These two biplots overlay variable contributions on top of the indivdual observations into a singular plot. This provides a succinct method to quickly evaluate how each observation corresponds with each variable.

Individual Plots with Overlaid Categorical Ellipses

The following plots show the relative grouping the several categorical variables onto the individual principle components plot. Each plot will have an ellipses generated on top of a pre or post-construction individiual pca plot. Each generated ellipse has a center that is the mean of all relative points.

For example, in the 'Pre-Construction: Species Groupings' plot, we see that deer and coyote have overlapping ellipses. This is useful for a snap shot of how the species interact. Upon closer examination, we see more dense groupings of deer and coyotes that are further away from each other. The means of the ellipsis are sensitive outliers and further examination of the data points would be prudent before finalizing any conclusions.

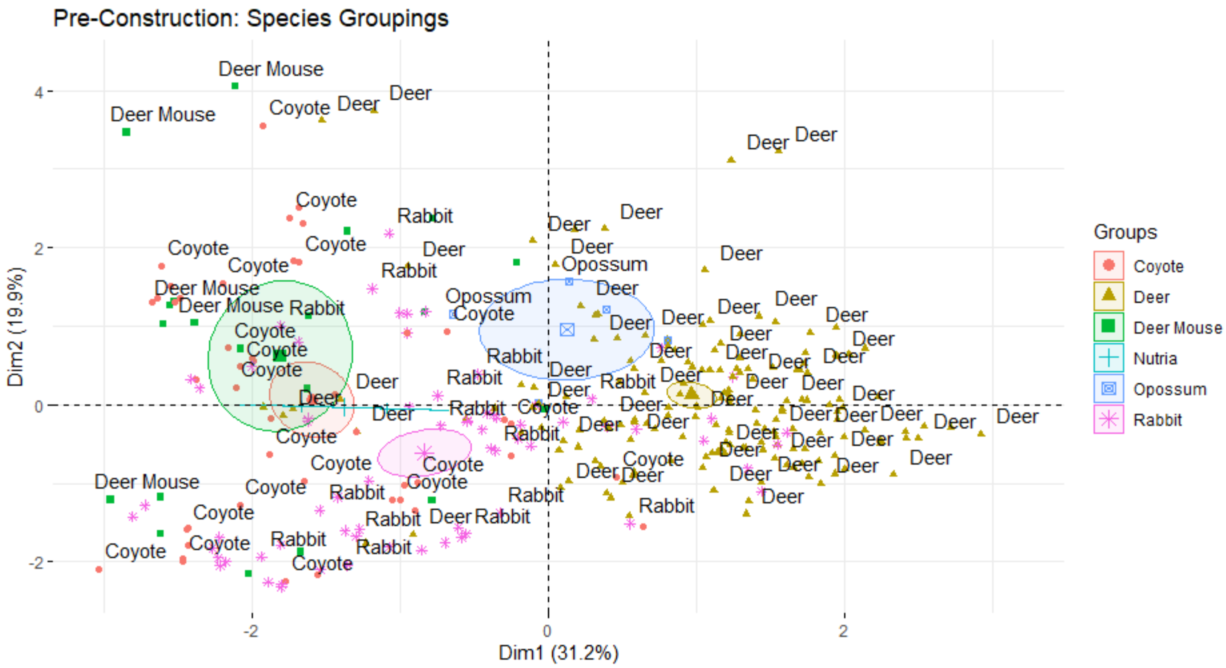
Each page will have pre-construction and post-construction plots adjacent to each other. Please note that due to variation in sampling methodology, any changes due to the road construction cannot be reliably determined. In that light, they can serve to highlight how the activity of the populations in the two datasets behave.

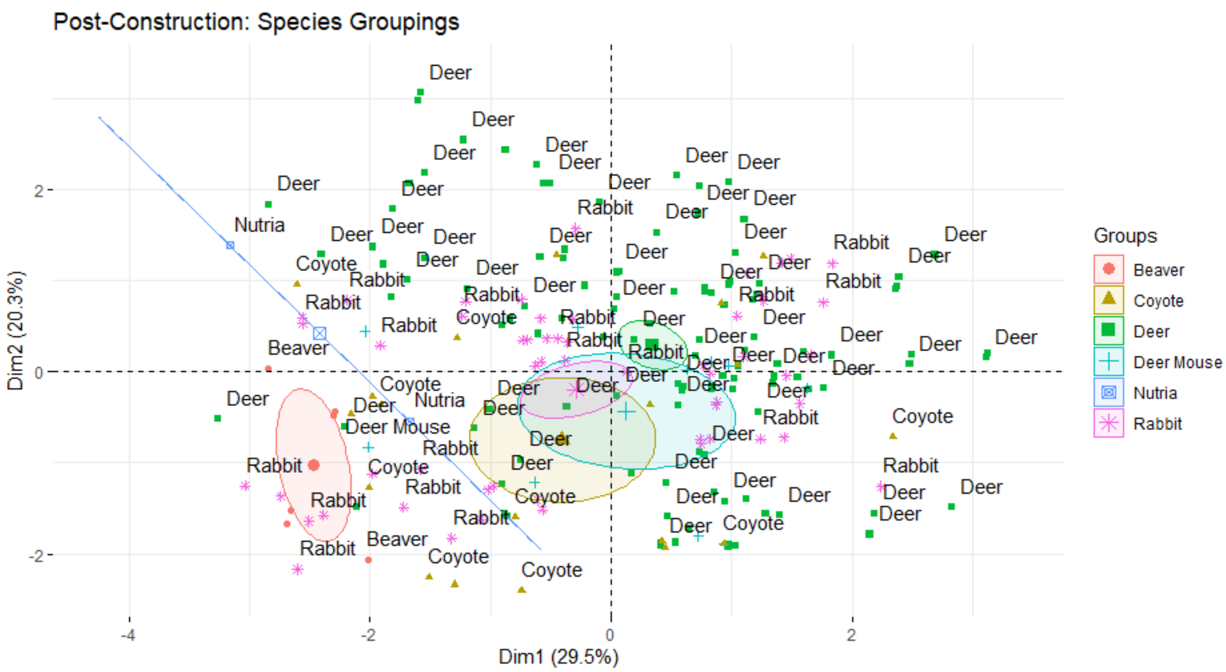
The categorical plots are generated in this order:

- Species Groupings
- Mammlian Groupings
- Crepuscular/Nocturnal Groupings
- Traffic Preference

Categorical Plots: Species

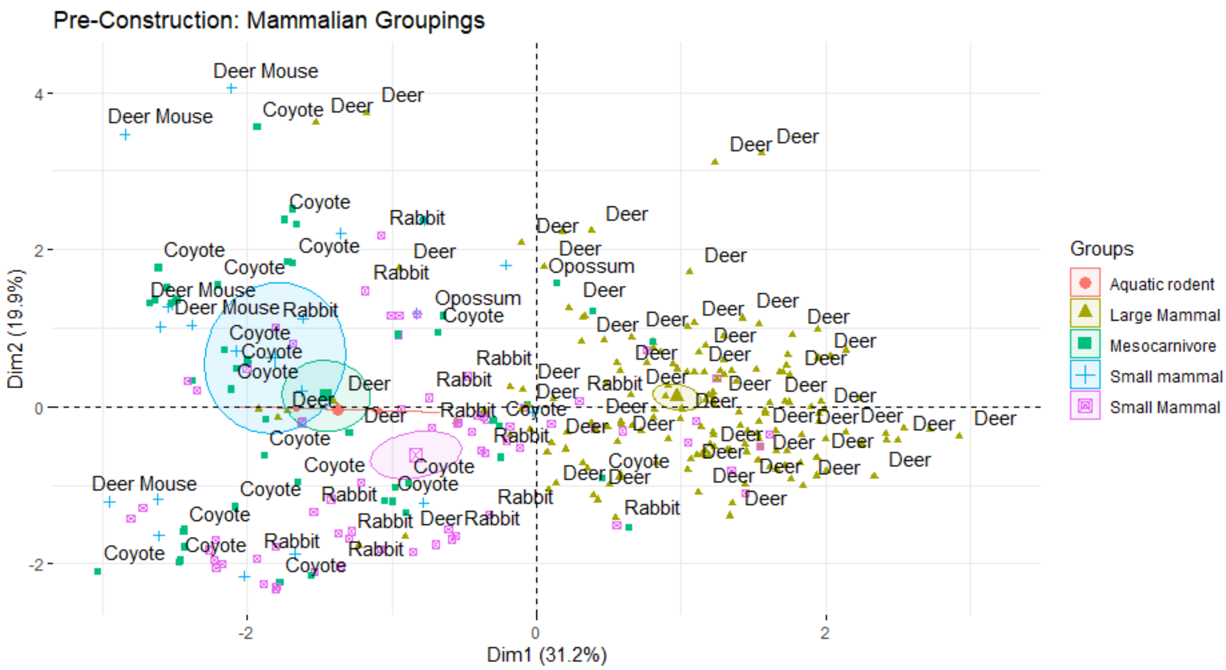
The plots, below, show how each species interacts with each other. This could provide a useful framework to approach future culvert design to better improve usability for specific groups instead of specific species. Generally speaking, we can see that deer and coyote form two of the larger, more distinct groups. Additionally, we see that deer mice and coyote are similarly grouped. This specific interaction could suggest a predator-prey relationship as well as culvert preference due to their smaller size.

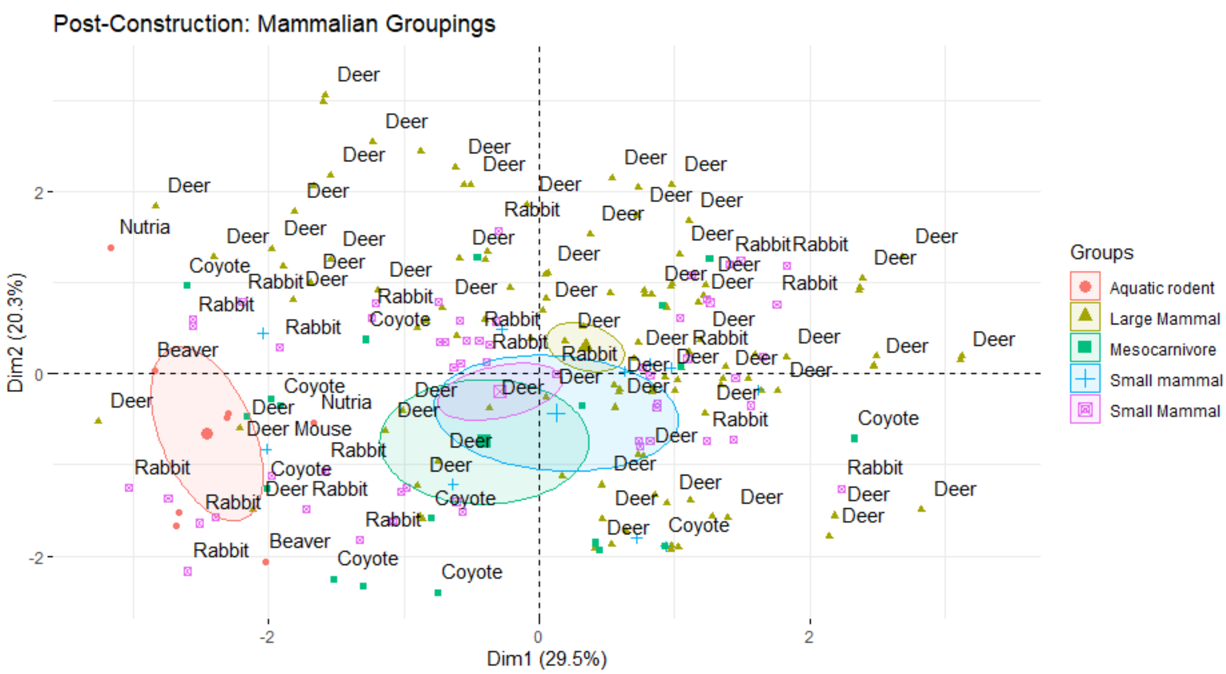




Categorical Plots: Mammalian Groups

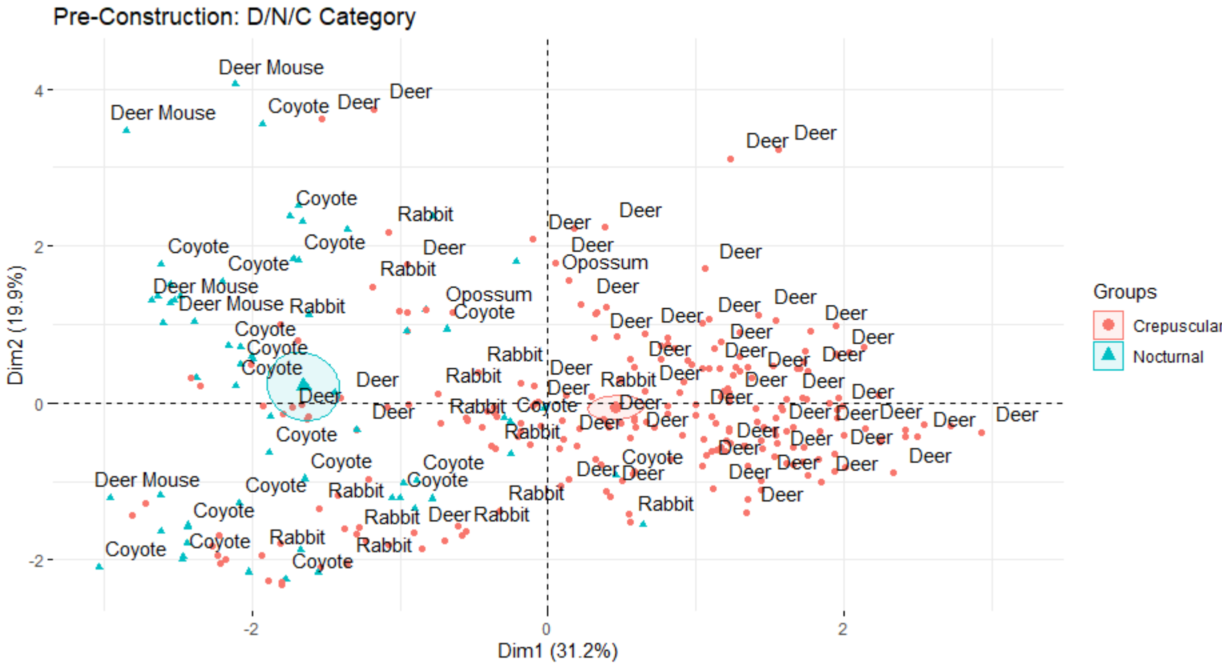
Simiar to the species plots above, we broaden our scope of definition and sort our species counts by mammalian groups. Each species observation is still noted, but the colors here refer to the key seen on the right-hand side of the plots. Pre-construction shows a clear grouping of deer while every other mammalian group are more similar to each other. In post-construction, we see the variations between the groups are more clustered together. This could indicate that the culverts presents a funneling effect.

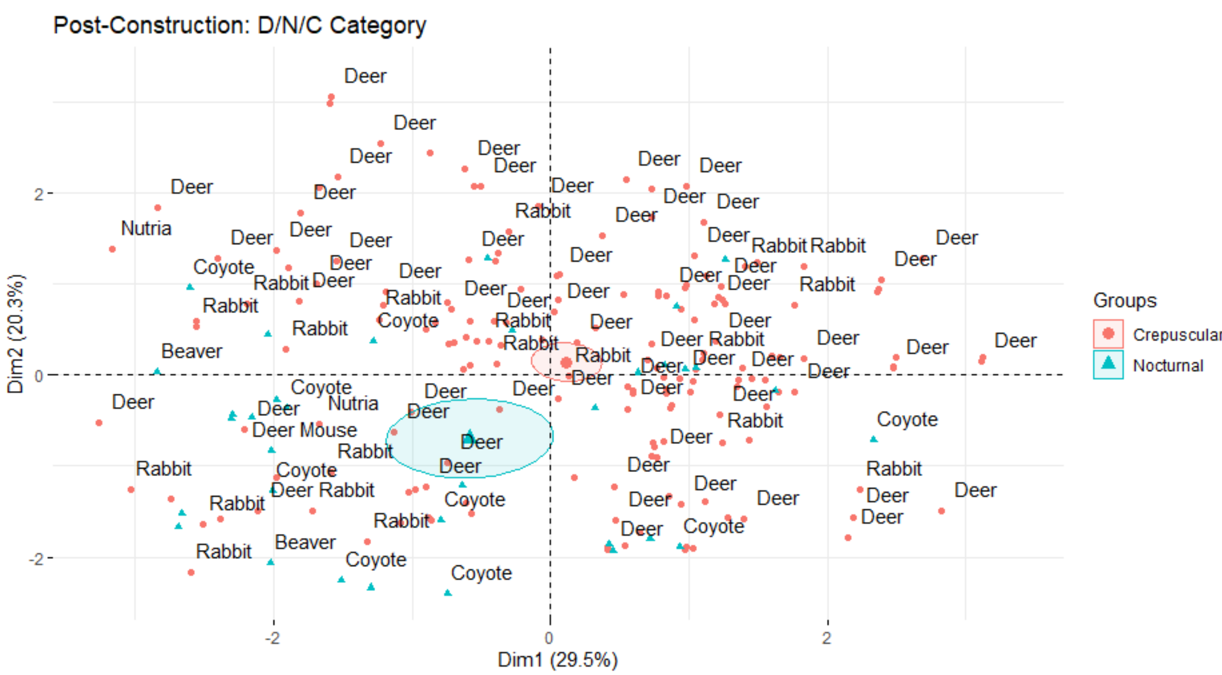




Categorical Plots: Crepuscular/Nocturnal

Here, we can further our scope from mammalian groups into specific hourly activity preference that species are known for. There are two groups - crepuscular and nocturnal. Crepuscular animals are more active during the dawn and evening hours while nocturnal animals are more active during night time hours. In pre-construction, we see the two categories are distinctly grouped - as is expected. These groupings still exist in post-construction, yet the variability is more evenly mixed and centered. If these two groups are merging, then it could further suppport the notion that the culverts present a funneling effect.



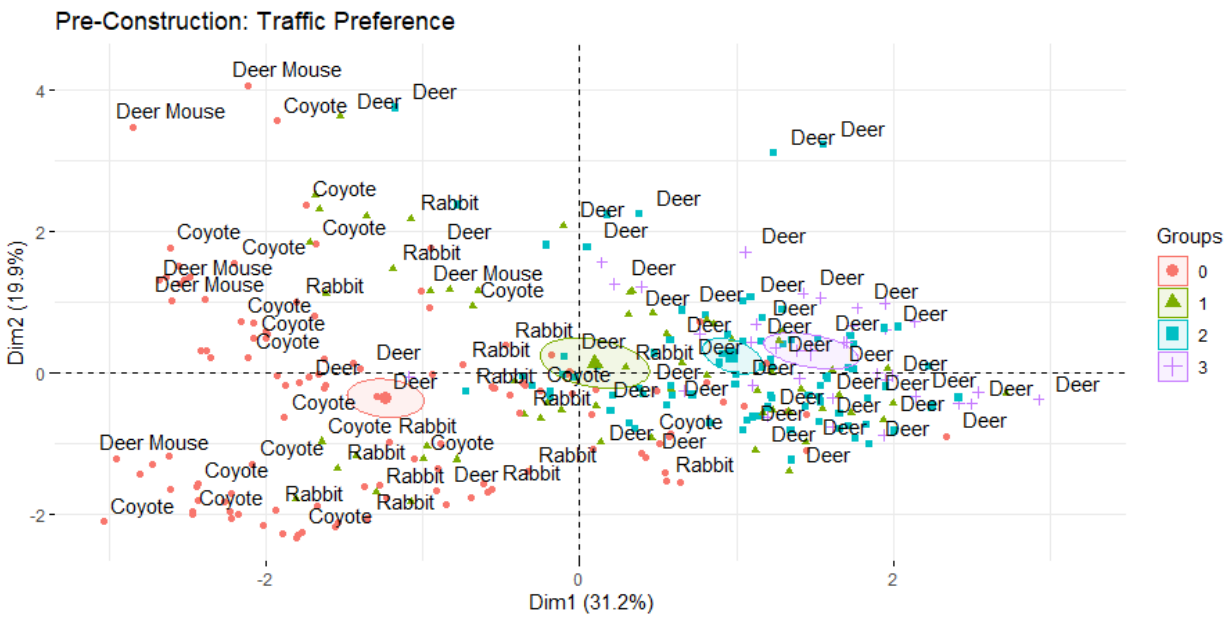


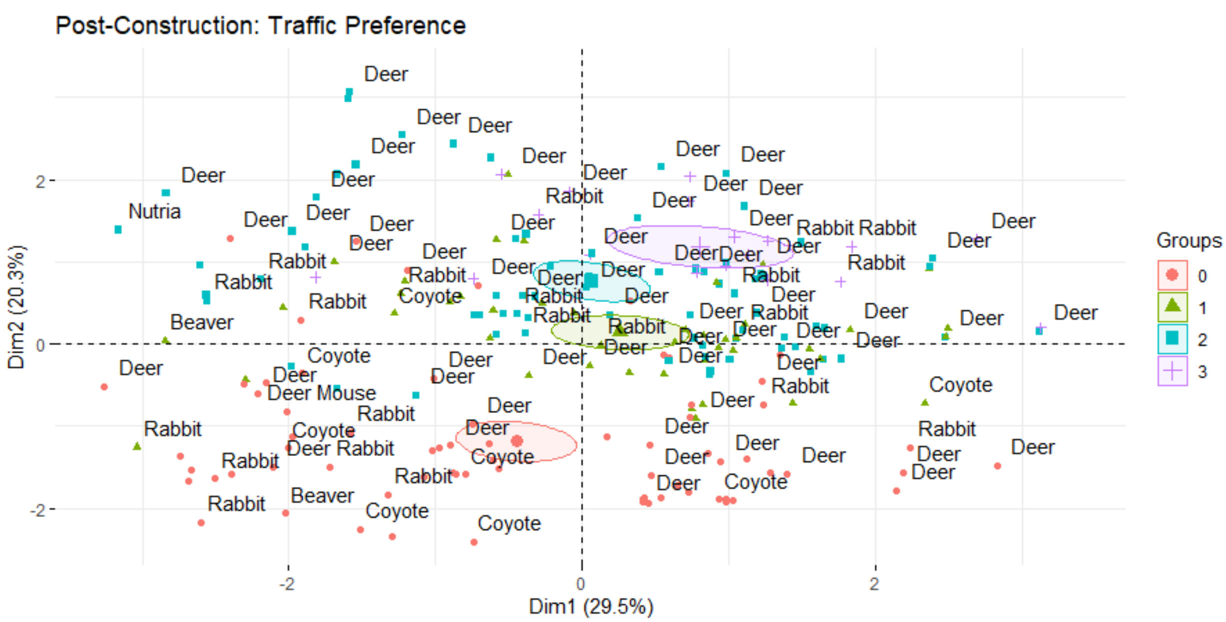
Categorical Plots: Traffic

When overlaid with recent traffic data onto our historical data, we can see some relationship with the presence of traffic. Pre-construction shows that only deer have almost no traffic preference and will be present regardless of activity. In post-construction, we see a similar trend. However, it seems that only coyotes have a strong preference for traffic while every other species does not seem affected.

It should strongly noted that the datasets used for the traffic analysis are severely mismatched. The observational data sets are from the year 2015-2016 and 2018-2019 while the traffic data is from 2020 and in a location that is due approximately 300 meters due north at Boeckman Road. It could be assumed that populations that access the Boeckman Road and Kinsman Road crossings are the same populations. However, there is no empirical basis for this assumption.

The results here should only be used to encourage the need for additional study with improved sampling methods.





R Code (1 of 8)

```
# Import data set from github
# Github Access URL: https://github.com/Kinsman-Road/rcode.git
library(readxl)
pre <- read_excel("PCA/pca.xlsx", sheet = "pca.pre.c")
post <- read_excel("PCA/pca.xlsx", sheet = "pca.post.c")

# Prepare dataframes
pre <- data.frame(pre)
post <- data.frame(post)

pre.n <- pre[1:7]       #create dataframes with only numerical columns from pre
post.n <- post[1:7]     #create dataframs with only numerical columns from post

# Performing principal components analysis
library(factoextra)
library(FactoMineR)

pca.pre <- prcomp(pre.n, scale = TRUE)   #singular value decomposition
pca.post <- prcomp(post.n, scale = TRUE)   #singular value decomposition
pre.eig <- get_eigenvalue(pca.pre)
post.eig <- get_eigenvalue(pca.post)

# Calculate PCA coordinates

# Pre-Construction PCA Coordinates
pre.vcf <- function(pre.load, comp.sdev){pre.load*comp.sdev}
pre.load <- pca.pre$rotation
pre.sdev <- pca.pre$sdev
pre.vcoord <- t(apply(pre.load, 1, pre.vcf, pre.sdev ))
pre.vc <- head(pre.vcoord[,1:7])   #1:7 number of dimensions

# Post-Construction PCA Coordinates
post.vcf <- function(post.load, comp.sdev){post.load*comp.sdev}
post.load <- pca.post$rotation
post.sdev <- pca.post$sdev
post.vcoord <- t(apply(post.load, 1, post.vcf, post.sdev))
post.vc <- head(post.vcoord[,1:7])   #1:7 number of dimensions

pre.vc       #table of pre pca coords
post.vc      #table of post pca coords

# PCA Qualtity of Representation
pre.cos2 <- pre.vcoord^2
post.cos2 <- post.vcoord^2

pre.cos2     #table of contribution to each dimension
post.cos2    #table of contribution to each dimension
```

R Code (2 of 8)

```r
# PCA Contributions to Each Given Component
pre.cc2 <- apply(pre.cos2, 2, sum)
contrib <- function(pre.cos2, pre.cc2){pre.cos2*100/pre.cc2}
pre.varc <- t(apply(pre.cos2, 1, contrib, pre.cc2))
pre.vcontrib <- head(pre.varc[,1:7])   #1:7 number of dimensions

post.cc2 <- apply(post.cos2, 2, sum)
contrib <- function(post.cos2, post.cc2){post.cos2*100/post.cc2}
post.varc <- t(apply(post.cos2, 1, contrib, post.cc2))
post.vcontrib <- head(post.varc[,1:7])   #1:7 number of dimensions

pre.vcontrib
post.vcontrib

# Creating a scree plot to determine principal dimensions of interest
pre.scree <- fviz_eig(pca.pre,
                      title = "Pre-Construction Scree Plot")

post.scree <- fviz_eig(pca.post,
                       title = "Post-Construction Scree Plot")
pre.scree
post.scree

# Creating contribution plot for individual observations
pre.ind <- fviz_pca_ind(pca.pre,
                        col.ind = "cos2",
                        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                        repel = TRUE,
                        label = "none",
                        title = "Pre-Construction Individual Plot")

post.ind <- fviz_pca_ind(pca.post,
                        col.ind = "cos2",
                        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                        repel = TRUE,
                        label = "none",
                        title = "Post-Construction Individual Plot")
pre.ind
post.ind
```

R Code (3 of 8)

```r
# Creating contribution plot for variable contributions
pre.var <- fviz_pca_var(pca.pre,
                        col.var = "cos2",
                        gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                        repel = TRUE,
                        title = "Pre-Construction Variable Plot")

post.var <- fviz_pca_var(pca.post,
                         col.var = "cos2",
                         gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
                         repel = TRUE,
                         title = "Post-Construction Variable Plot")
pre.var
post.var

# Creating a biplot(combination of ind + var plots)
pre.bp <- fviz_pca_biplot(pca.pre,
                         col.ind = "coral",
                         col.var = "#2f2091",
                         label = "var",
                         repel = TRUE,
                         title = "Pre-Construction Biplot")

post.bp <- fviz_pca_biplot(pca.post,
                           col.ind = "cadetblue3",
                           col.var = "#2f2091",
                           label = "var",
                           repel = TRUE,
                           title = "Post-Construction Biplot")
pre.bp
post.bp

# Creating 3D Observation Plots
library(pca3d)

pre3d.species <- pca3d(pca.pre, group=pre$species)
pre3d.mcat <- pca3d(pca.pre, group=pre$category)
pre3d.dnc <- pca3d(pca.pre, group=pre$dnc)
pre3d.traffic <- pca3d(pca.pre, group=pre$traffic)

post3d.species <- pca3d(pca.post, group=post$species)
post3d.mcat <- pca3d(pca.post, group=post$category)
post3d.dnc <- pca3d(pca.post, group=post$dnc)
post3d.traffic <- pca3d(pca.post, group=post$traffic)
```

R Code (4 of 8)

```r
# Creating an individual PCA plot with ellipses for categories
# (1) First define categories as factors

# (1a) Pre categories
pre.g.species <- as.factor(pre$species[1:351])
pre.g.solar <- as.factor(pre$solar[1:351])
pre.g.cat <- as.factor(pre$category[1:351])
pre.g.cam <- as.factor(pre$camera[1:351])
pre.g.traffic <- as.factor(pre$traffic[1:351])
pre.g.dnc <- as.factor(pre$dnc[1:351])

# (1b) Post categories
post.g.species <- as.factor(post$species[1:221])
post.g.solar <- as.factor(post$solar[1:221])
post.g.cat <- as.factor(post$category[1:221])
post.g.cam <- as.factor(post$camera[1:221])
post.g.traffic <- as.factor(post$traffic[1:221])
post.g.dnc <- as.factor(post$dnc[1:221])

# (2) Produce ellipses PCA graphs for every factor

# (2a) Pre-Construction Ellipses PCA categories
pre.species <- fviz_pca_ind(pca.pre,
                            col.ind = pre.g.species,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Pre-Construction: Species Groupings") +
                   geom_text(
                            label=pre$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)

pre.solar <- fviz_pca_ind(pca.pre,
                            col.ind = pre.g.solar,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Pre-Construction: Daylight Preference") +
                   geom_text(
                            label=pre$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)
```

R Code (5 of 8)

```
# (2a) Pre-Construction Ellipses PCA categories (Continued)
pre.cat <- fviz_pca_ind(pca.pre,
                            col.ind = pre.g.cat,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Pre-Construction: Mammalian Groupings") +
                   geom_text(
                            label=pre$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)

pre.cam <- fviz_pca_ind(pca.pre,
                            col.ind = pre.g.cam,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Pre-Construction: Camera Preference") +
                   geom_text(
                            label=pre$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)

pre.traffic <- fviz_pca_ind(pca.pre,
                           col.ind = pre.g.traffic,
                           palette = c( ),
                           addEllipses = TRUE,
                           ellipse.type = "confidence",
                           legend.title = "Groups",
                           repel = TRUE,
                           label = "none",
                           title = "Pre-Construction: Traffic Preference") +
                  geom_text(
                           label=pre$species,
                           nudge_x = 0.25, nudge_y = 0.25,
                           check_overlap = T)
```

R Code (6 of 8)

```
# (2a) Pre-Construction Ellipses PCA categories (Continued)
pre.dnc <- fviz_pca_ind(pca.pre,
                            col.ind = pre.g.dnc,
                            palette = c(""),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "groups",
                            repel = TRUE,
                            label = "none",
                            title = "Pre-Construction: D/N/C Category") +
                   geom_text(
                            label=pre$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)

# (2b) Post-Construction Ellipses PCA categories
post.species <- fviz_pca_ind(pca.post,
                            col.ind = post.g.species,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Post-Construction: Species Groupings") +
                   geom_text(
                            label=post$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)

post.solar <- fviz_pca_ind(pca.post,
                          col.ind = post.g.solar,
                          palette = c( ),
                          addEllipses = TRUE,
                          ellipse.type = "confidence",
                          legend.title = "Groups",
                          repel = TRUE,
                          label = "none",
                          title = "Post-Construction: Daylight preference") +
                 geom_text(
                          label=post$species,
                          nudge_x = 0.25, nudge_y = 0.25,
                          check_overlap = T)
```

R Code (7 of 8)

```r
# (2b) Post-Construction Ellipses PCA categories (Continued)
post.cat <- fviz_pca_ind(pca.post,
                        col.ind = post.g.cat,
                        palette = c( ),
                        addEllipses = TRUE,
                        ellipse.type = "confidence",
                        legend.title = "Groups",
                        repel = TRUE,
                        label = "none",
                        title = "Post-Construction: Mammalian Groupings") +
               geom_text(
                        label=post$species,
                        nudge_x = 0.25, nudge_y = 0.25,
                        check_overlap = T)

post.cam <- fviz_pca_ind(pca.post,
                        col.ind = post.g.cam,
                        palette = c( ),
                        addEllipses = TRUE,
                        ellipse.type = "confidence",
                        legend.title = "Groups",
                        repel = TRUE,
                        label = "none",
                        title = "Post-Construction: Camera Preference") +
               geom_text(
                        label=post$species,
                        nudge_x = 0.25, nudge_y = 0.25,
                        check_overlap = T)

post.traffic <- fviz_pca_ind(pca.post,
                            col.ind = post.g.traffic,
                            palette = c( ),
                            addEllipses = TRUE,
                            ellipse.type = "confidence",
                            legend.title = "Groups",
                            repel = TRUE,
                            label = "none",
                            title = "Post-Construction: Traffic Preference") +
                   geom_text(
                            label=post$species,
                            nudge_x = 0.25, nudge_y = 0.25,
                            check_overlap = T)
```

R Code (8 of 8)

```
# (2b) Post-Construction Ellipses PCA categories (Continued)
post.dnc <- fviz_pca_ind(pca.post,
                             col.ind = post.g.dnc,
                             palette = c( ),
                             addEllipses = TRUE,
                             ellipse.type = "confidence",
                             legend.title = "Groups",
                             repel = TRUE,
                             label = "none",
                             title = "Post-Construction: D/N/C Category") +
                    geom_text(
                             label=post$species,
                             nudge_x = 0.25, nudge_y = 0.25,
                             check_overlap = T)

# Generate Plots
pre.scree
pre.ind
pre.var
pre.bp
pre.species
pre.solar
pre.cat
pre.cam
pre.traffic
pre.dnc

post.scree
post.ind
post.var
post.bp
post.species
post.solar
post.cat
post.cam
post.traffic
post.dnc
```