

# Appendix C: 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 Representation** - 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	-0.1222332	0.7930852	-0.3339131	0.3493888	-0.2800734	-0.0174022	0.2091898
count	0.2521509	0.0682524	0.8222814	0.4482944	-0.1826160	0.1457663	0.0071085
hour	0.3903734	0.4109103	0.1449199	-0.7110076	-0.2118863	0.3274154	0.0114516
pcip	-0.4196659	0.6550756	0.2199545	-0.0163340	0.5813983	0.0688288	-0.0578650
tmax	0.8084495	-0.2232759	-0.1375800	0.1301557	0.3501975	0.2440556	0.2801556
tmin	0.7980485	0.2484833	-0.3088037	0.2870622	0.0155446	0.1068322	-0.3345973

Table 2: Post-Construction Variable Coordinates

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	-0.5953325	0.1339423	-0.2446354	0.2545993	-0.6520869	0.2788229	0.0034427
count	-0.1635017	0.4729479	-0.7934043	-0.0129382	0.1704482	-0.3014121	0.0052746
hour..	0.1534500	0.6811972	0.4448833	0.2547053	-0.2556553	-0.4278270	-0.0350987
pcip	-0.4110667	0.3534902	0.1360185	-0.8127083	-0.1156309	0.0382359	-0.1106373
tmax	0.8991581	-0.0246318	-0.2205253	0.0012199	-0.1858991	0.1027143	-0.3117152
tmin	0.8431608	0.1551327	-0.1183800	-0.2896895	-0.2850856	0.0589638	0.2869308

# 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	0.0149409	0.6289842	0.1114980	0.1220725	0.0784411	0.0003028	0.0437604
count	0.0635801	0.0046584	0.6761467	0.2009679	0.0333486	0.0212478	0.0000505
hour	0.1523914	0.1688473	0.0210018	0.5055318	0.0448958	0.1072008	0.0001311
pcip	0.1761195	0.4291240	0.0483800	0.0002668	0.3380240	0.0047374	0.0033484
tmax	0.6535905	0.0498521	0.0189283	0.0169405	0.1226383	0.0595631	0.0784872
tmin	0.6368815	0.0617440	0.0953597	0.0824047	0.0002416	0.0114131	0.1119554

Table 4: Post-Construction Variable Quality of Representation

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	0.3544208	0.0179405	0.0598465	0.0648208	0.4252174	0.0777422	0.0000119
count	0.0267328	0.2236797	0.6294905	0.0001674	0.0290526	0.0908493	0.0000278
hour..	0.0235469	0.4640296	0.1979212	0.0648748	0.0653596	0.1830359	0.0012319
pcip	0.1689758	0.1249553	0.0185010	0.6604948	0.0133705	0.0014620	0.0122406
tmax	0.8084853	0.0006067	0.0486314	0.0000015	0.0345585	0.0105502	0.0971663
tmin	0.7109202	0.0240662	0.0140138	0.0839200	0.0812738	0.0034767	0.0823293

## Variable Contribution

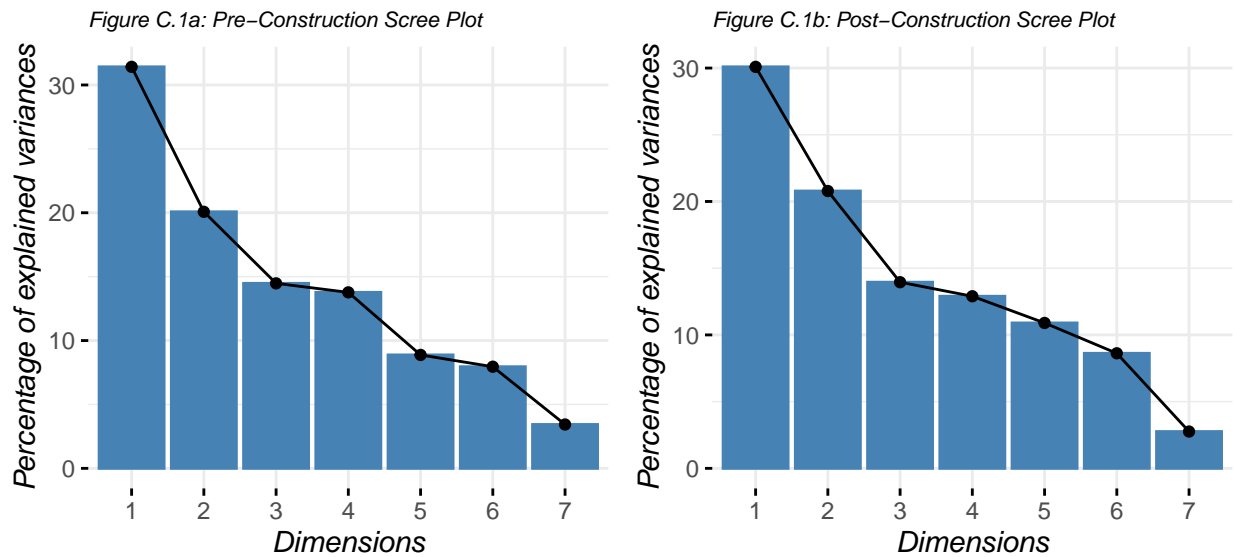
Table 5: Pre-Construction Variable Dimension Contribution (%)

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	0.6793669	44.7593734	11.002785	12.663911	12.6258465	0.0544015	18.2143155
count	2.8909955	0.3314975	66.723157	20.848585	5.3677741	3.8169585	0.0210323
hour	6.9292564	12.0154029	2.072487	52.444311	7.2264046	19.2575547	0.0545841
pcip	8.0081760	30.5370504	4.774208	0.027678	54.4081782	0.8510277	1.3936819
tmax	29.7188505	3.5475448	1.867869	1.757424	19.7397963	10.6999218	32.6685943
tmin	28.9590880	4.3937848	9.410239	8.548737	0.0388935	2.0502512	46.5990066

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

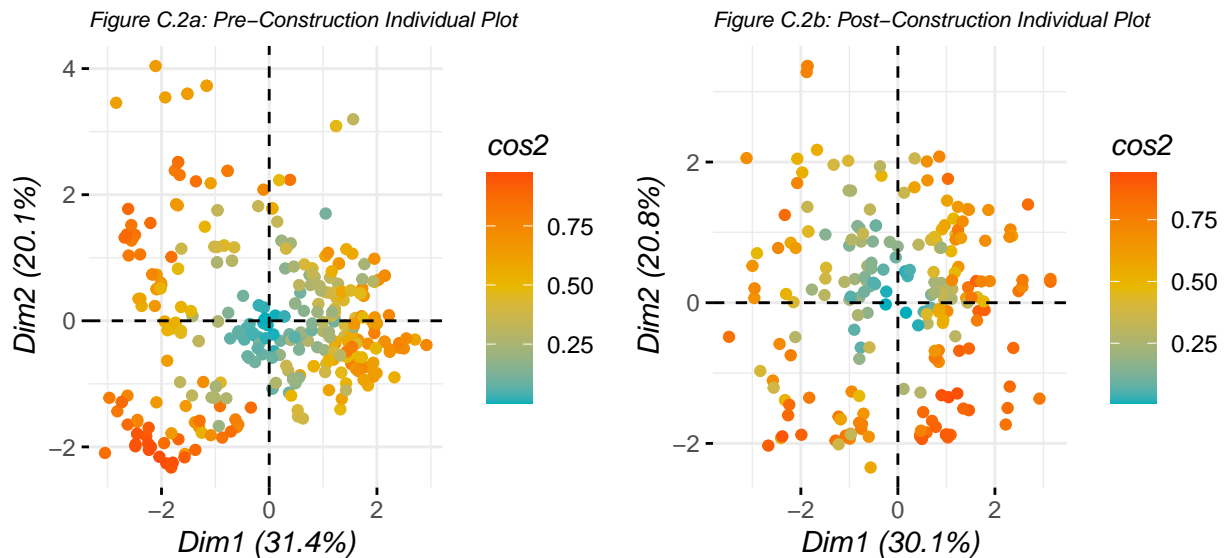
	PC1	PC2	PC3	PC4	PC5	PC6	PC7
month	16.826306	1.2331263	6.128477	7.1799845	55.744827	12.8811410	0.0061375
count	1.269153	15.3744133	64.461918	0.0185419	3.808714	15.0528534	0.0144071
hour..	1.117902	31.8946421	20.267789	7.1859651	8.568466	30.3273030	0.6379336
pcip	8.022211	8.5886875	1.894567	73.1608310	1.752836	0.2422366	6.3386307
tmax	38.383251	0.0417026	4.980019	0.0001648	4.530520	1.7480709	50.3162724
tmin	33.751295	1.6541654	1.435062	9.2955420	10.654775	0.5760611	42.6330991

## Scree Plots



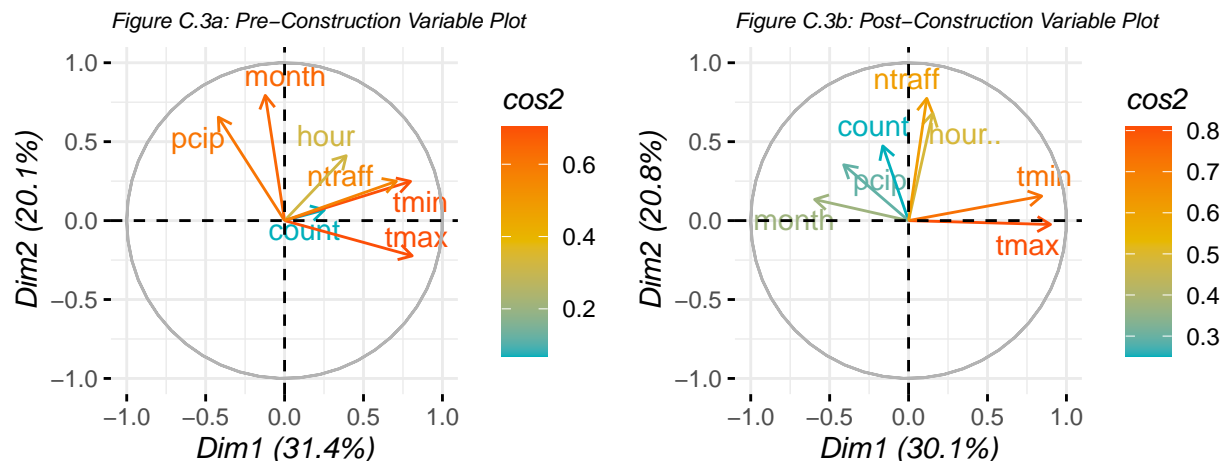
As seen above, the scree plots for both pre and post-construction show many principle dimensions. This is not ideal as we cannot readily visualize 4 or more dimensions. Though we miss a great deal of detail by examining fewer plots, we can still gain a sense of understanding since there is already so much noise in the datasets. 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.

## 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 turquoise coloring indicates that a particular observation does not contribute much at all.

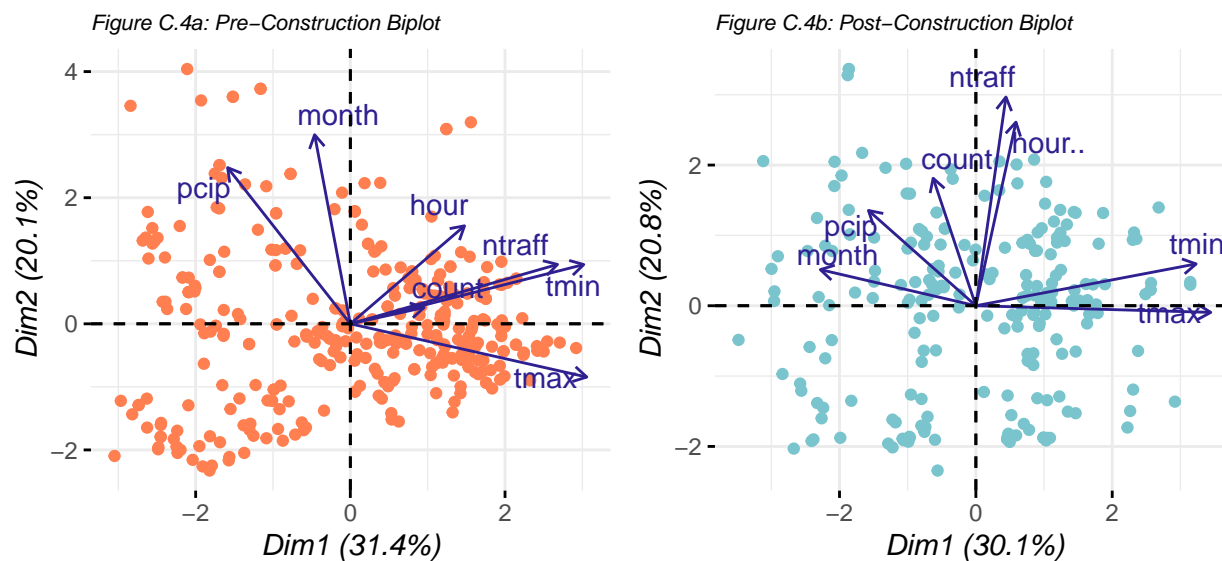
## Variable Plots



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 given 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 be correlated with hour in both plots, again, this is to be expected.

## Biplots



These two biplots overlay variable contributions on top of the individual 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 individual 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
- Mammalian Groupings
- Diet Activity 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.

Figure C.5a: Pre-Construction Species Groupings

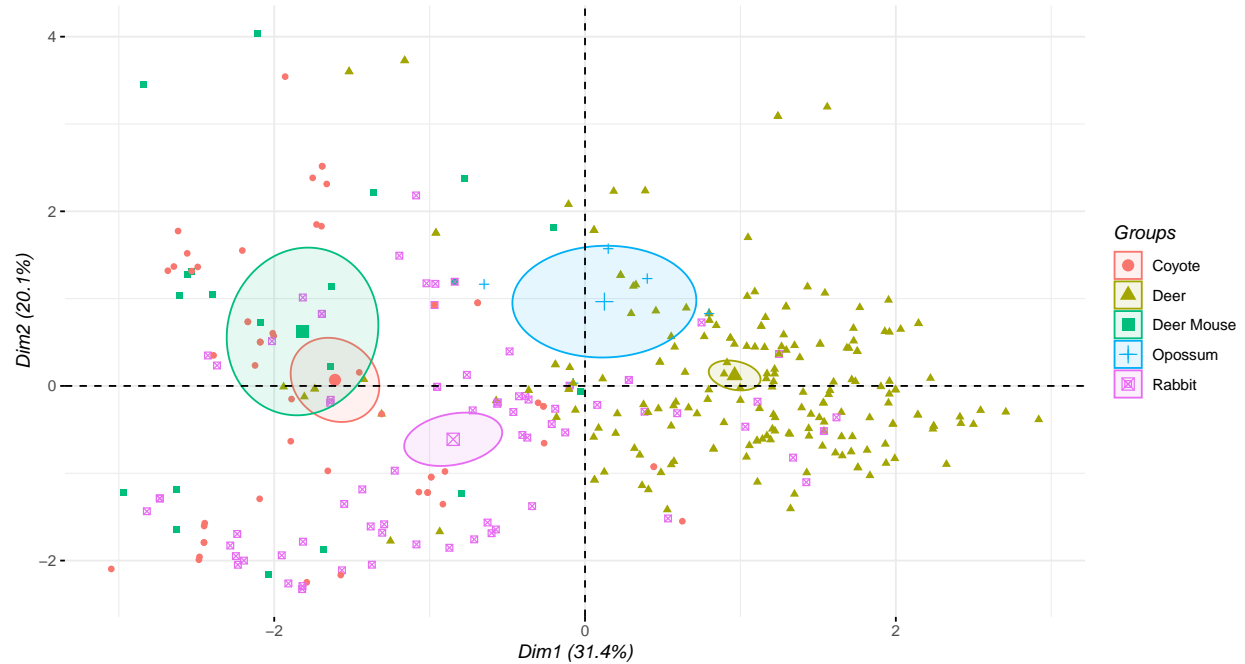
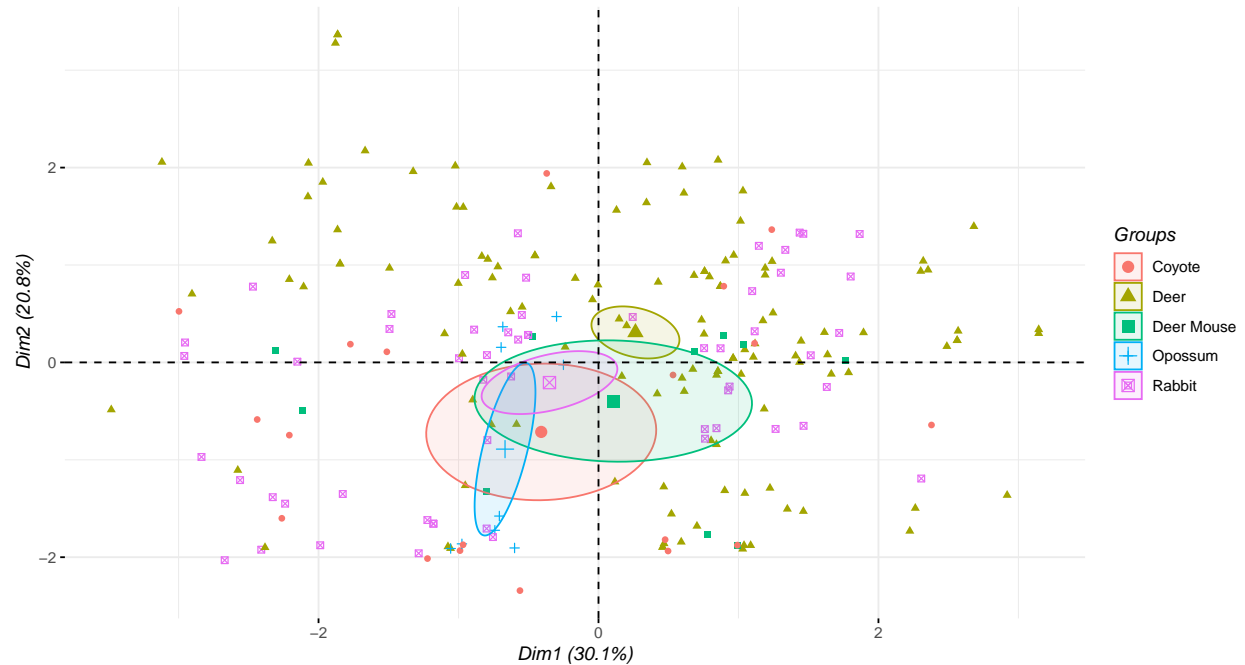


Figure C.5b: Post-Construction Species Groupings



## Categorical Plots: Mammalian Groups

Similar 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.

Figure C.6a: Pre-Construction Mammalian Groupings

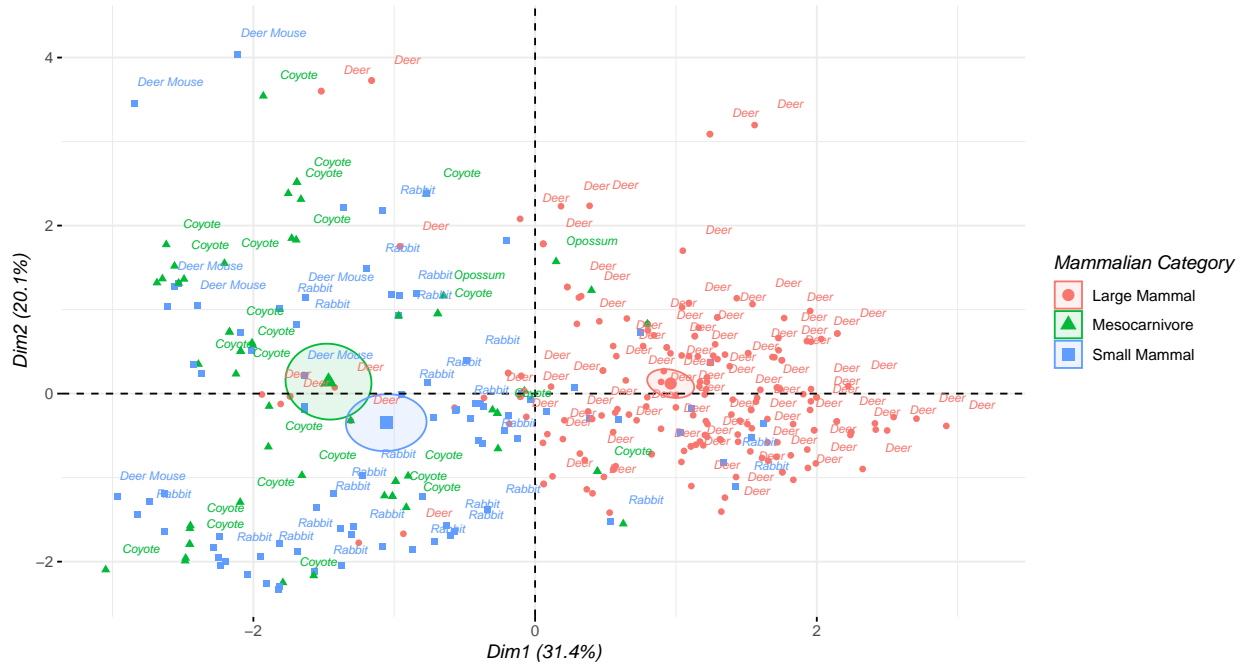


Figure C.6b: Post-Construction Mammalian Groupings



## Categorical Plots: Diel Activity Groups

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 support the notion that the culverts present a funneling effect.

Figure C.7a: Pre-Construction Diel Activity

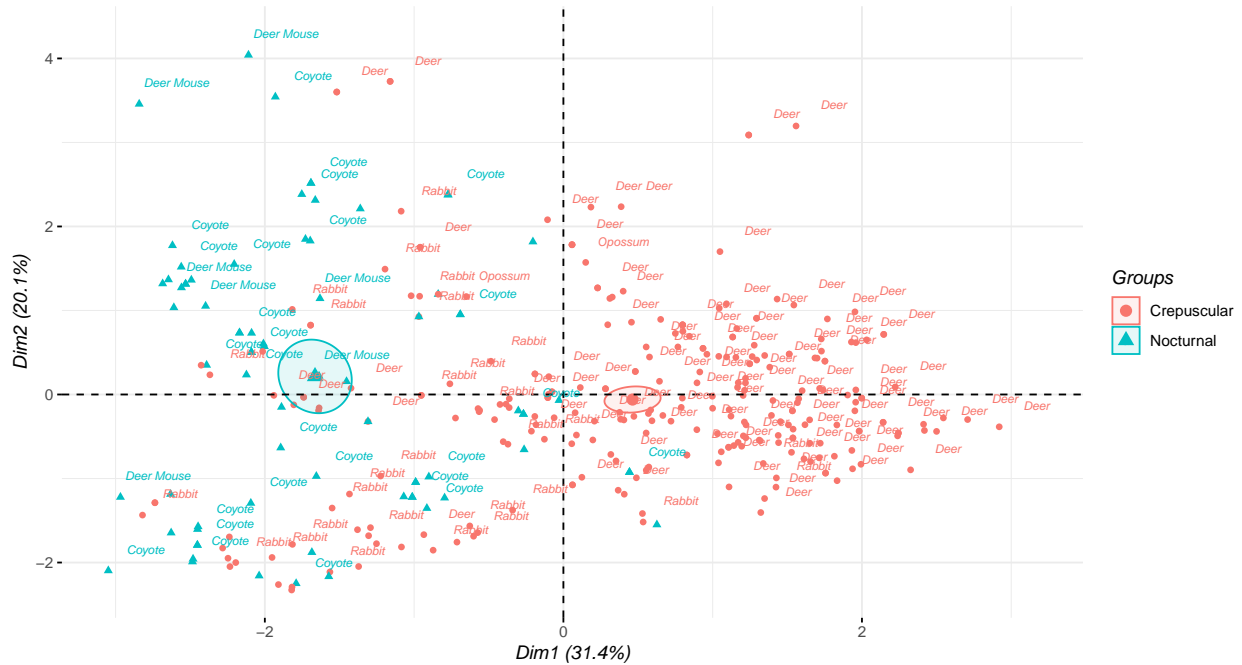
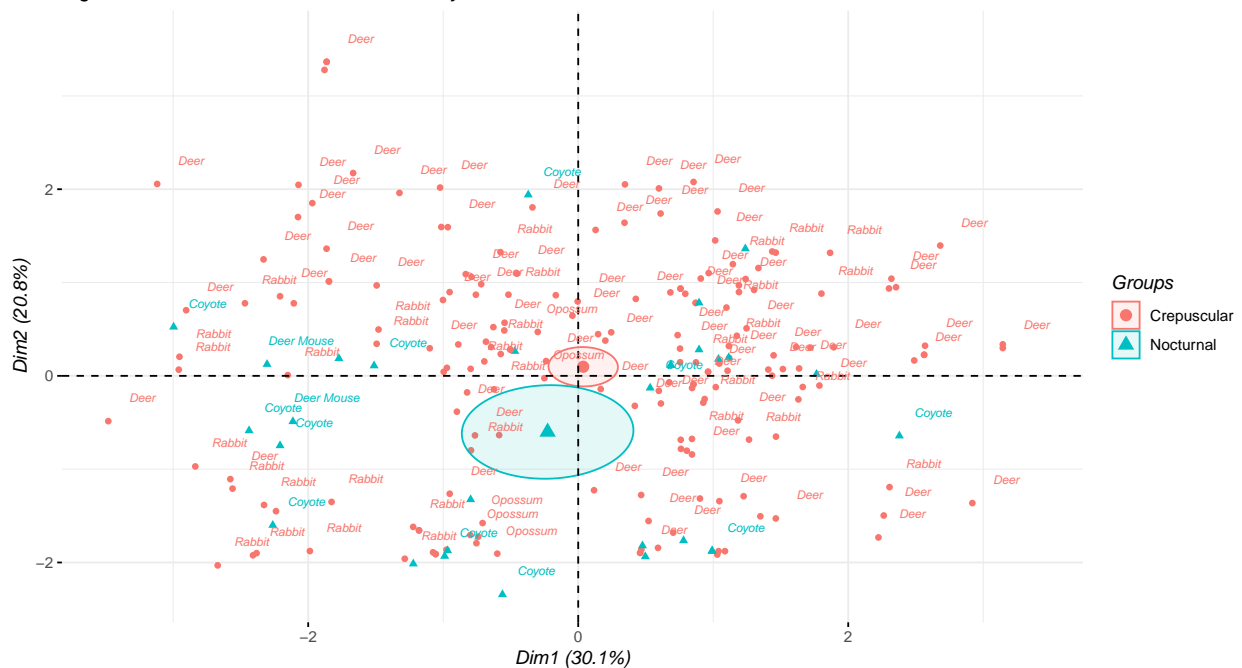


Figure C.7b: Post-Construction Diel Activity





## Categorical Plots: Camera Preference

When examining the general groupings for camera preference, we notice some intriguing changes in behavior. In pre-construction observations, there are two distinct groupings. The first grouping involves camera 2 North and 4 South while the second group involves camera 1 North and camera 3 South. In pre-construction, this particular grouping is likely due to the angle of the cameras. In post-construction, however, all of the cameras seem to have more overlap. This could support the notion that the culverts present a funneling effect.

Figure C.8a: Pre-Construction Camera Preference

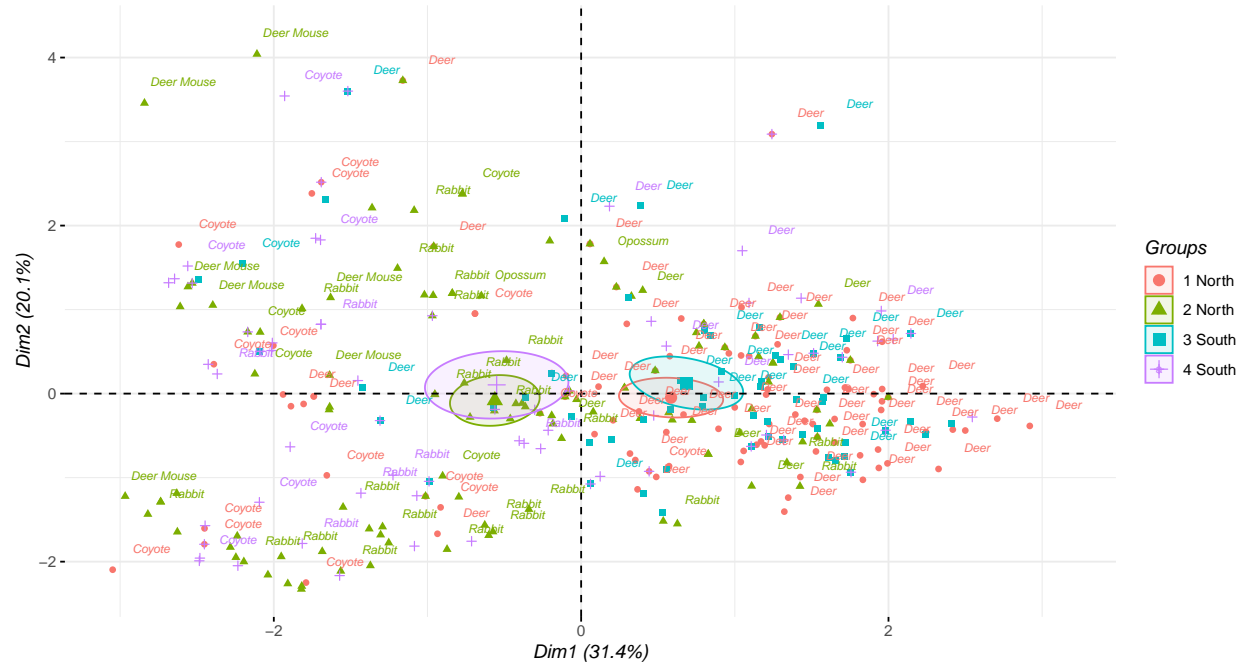


Figure C.8b: Post-Construction Camera Preference



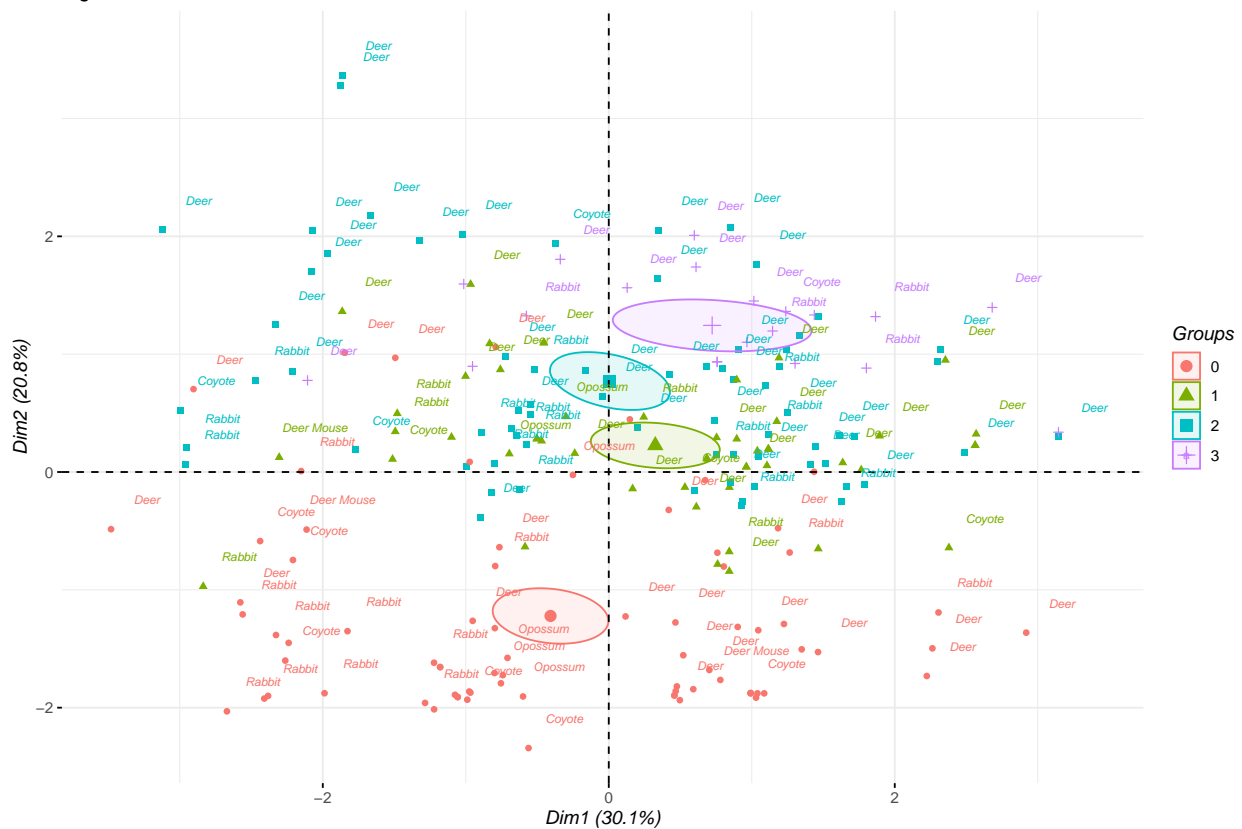
## Categorical Plots: Traffic

When overlaid with recent traffic data onto our historical data, we can see some relationship with the presence of traffic. Only post-construction data can be examined here due to the lack of a constructed roadway in pre-construction settings. In post-construction, it's observed that coyotes have a preference for traffic while every other species does not seem affected. This could present an issue with coyote's natural hunting patterns as traffic is presenting as a barrier to prey access.

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.

Figure C.9b: Post-Construction Traffic Preference



## 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 dataframes 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 Quality 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)

```
# 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)

```
# 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)

```
# 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)

```
# (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
```