6-17-2021

Grizzly Bears Mortalities and The Survival of The Species

Courtney Swanson
University of Minnesota - Morris, swan2187@morris.umn.edu

Recommended Citation

Swanson, Courtney, "Grizzly Bears Mortalities and The Survival of The Species" (2021). *Senior Seminars and Capstones*. 4.
https://digitalcommons.morris.umn.edu/capstone/4

Grizzly Bears Mortalities and The Survival of The Species.

Courtney Swanson

June 17, 2021

Abstract

In this paper we aim to understand what is happening in the grizzly bear population mortalities from the year 2010 to 2020. We are performing Classical and Regression Tree (CART) methods and Correspondence Analysis on data provided by the U.S. Geological Survey (USGS). We found certain variables in the data set to be important through CART methods. Correspondence Analysis then allowed us to compare these variables to determine their relationships and association to one another. Most of the grizzly bear deaths are human caused and mainly over land and resources such as food and habitat. This aligns with some of the issues mentioned in the past papers. Though grizzly population have improved, their cause of death and human-bear interaction remains relatively the same. Solutions on how to continue conserving this population will need to focus around the human-bear relationship.

Advisers and Key Words

- Primary Adviser: Engin Sungur
- Secondary Adviser: Peter Dolan
- Key word 1: Grizzly bear mortalities
- Key word 2: Classification and Regression Tree (CART)
- Key word 3: Correspondence Analysis (CA)

Introduction

In this project, we will be studying the grizzly bear mortalities in western areas of the United States. Grizzly bears are currently a threatened species on the endangered species list. We know the grizzly bear population has decreased due to loss of habitat, climate warming, human-bear interactions, and many other reasons (Mattson and Reid, 1991). The media also has influenced the human perspective of grizzly bears (Hughes *et al.*, 2020). We will take the results from data we have collected between 2010 and 2020 and compare the results from the more current data with older data to see how the mortality for bears are changing and what types of factors could be causing these deaths. Investigating factors such as reason for removal, sex, and age class will help us better understand what is happening to the grizzly bear population, how the efforts of humans have affected the mortality of grizzly bears, and what can be done to continue rebuilding their population.

Background

Grizzly bears once lived all over the western United States and Mexico, but now they are reduced to few places such as Yellowstone National Park, Teton National Park, Alaska, and Northern Canada. Most of the grizzly bear habitat has been overtaken by humans (The National Wildlife Federation). This species was recently added again to the endangered species list due to several concerns including logging, human-bear encounters, and habitat devastation (U.S. Fish & Wildlife Service, 2020). Though the grizzly bear population has been able to recover in some of the national parks and national forests, their numbers are not high enough to sustain their population and keep them off the endangered species list; currently, they are considered a threatened species (The National Wildlife Federation). Studies from the late 1900s helps to formulate why this is.

The papers we looked at contained a wide variety of what affects the grizzly bear population. In the 1991, Mattson and Reid concluded that some issues grizzly bears face are that their food supply could disappear as the climate gets warmer, they may have a genetic setback due to the small population, cub mortality, and human management of grizzly bears (Mattson and Reid, 1991). Though their population has increased, grizzly bears are endangered because they lack the numbers to survive without human protection, and they are losing land and resources quickly. In the late 1900s, people were working on a Grizzly Bear Recovery Plan (Servheen, 1993) which was later implemented; it focused on the management of the population. McLellan *et. al* analyzed data collected between 1975 and 1997; they concluded that the mortalities of grizzly bears must be documented to track and understand why grizzly bears are dying as well as unreported kills. McLellan *et. al* also asserted that providing grizzly bears conservation locations with less human settlement would be best to prevent human-grizzly bear interactions(McLellan *et al.*, 1999). Researchers Mattson and Reid also agreed with the previous conclusion (Mattson and Reid, 1991). This strategy might help to decrease the mortality of grizzly bears since (at the time), many grizzly bear deaths seemed to be related to their interactions with humans; McLellan *et. al* specifically found that young males were dying more often than female bears and male bears of other ages (McLellan *et al.*, 1999). These observations are something we will look into in to with our data.

Media portrayal of grizzly bears plays a role in how human's view grizzly bears. In 2008, Hughes *et.al* discussed finding about how the media portrays bears. They determined much of the writing was written to captivate viewers. However, bears tended to be drawn in a negative light. This is mostly in part by the large number of stories about negative human-bear conflict though many encounters are not threatening to humans (Hughes *et al.*, 2020). This strengthens the idea that the relationship between bears and humans, either indirectly or directly, is one of the most important relationships for grizzly bear survival.

With this information, we can investigate how the mortalities of grizzly bears have changed knowing that people have been trying to recover the grizzly bear's population through land management and hunting regulations. We can also better understand the challenges grizzly bears are currently facing and are continuing to face. In order to do this, we must look at some techniques used to study the data.

Classification and Regression Trees

CART Method

Classification and Regression Trees (CART) is a method which is used to build a model which predicts our response variable based on a group of variables which describe our prediction best. CART contains two main parts: splitting and pruning. We will focus first on the splitting aspect of CART.

Let's say we start with a data set. Inside this data set we have a categorical response Y containing classes. Our data set also contains either categorical or numeric predictor variables X_m, where m is a positive integer. Lets start with a sample space S: a set of all possible values of X_m and Y. We are going to split S recursively based on each of our predictor variables X_m. We split based on the concept of how many individuals in our data contain the same response variable in each of the groups we recursively make. Every split is made based off of a decision making process. Splitting takes place until we have K disjoint groups (Loh, 2011). The primary objective of a split is to have S contain K disjoint groups where the individuals in each groups have as many individuals with the same response level as possible (Patel, 2003). Note that $K \geq 1$. Also, since our K groups were made recursively, they are disjoint (Patel, 2003). When each of these K groups contain individuals with the same response level, this is known as homogeneity. The splitting is working towards the purity in the response that is achieved when the data set contains K homogeneous groups.

Let's say that we are looking at a cause of death variable with two levels: Human Caused (HC) and Natural (N). Let's say we will use the explanatory variables sex and age to describe them, where sex contains the levels Male, Female, and Unknown, and age contains the levels old and young. Let's say the variable old contains individuals older than 5 and the variable young contains individuals five and younger. In this case, Y is cause of death, X_1 is age, and X_2 is sex. We can then use CART to recursively split our sample space into K groups.

The splitting is determined by calculating the homogeneity of the data with K groups and with $K+1$ groups. We can calculate the homogeneity of classification tree groups using the Gini Index equation. The Gini Index equation is as follows

$$G_m = \sum_{k=1}^{K} \hat{p}_{mk} \cdot (1 - \hat{p}_{mk}), \quad 0 \leq \hat{p}_{mk} \leq 1,$$

where K is the number of classes and $\hat{p}_{mk}$ is the proportion of observations that are in the kth class and the mth group (James *et al.*, 2013). In this case k is referring to the classes that Y are split up into. Also note that the Gini Index is bound since $\hat{p}_{mk}$ is. Therefore, the Gini Index G_m is in the interval $[0, 1/4]$. Most importantly, as $\hat{p}_{mk}$ gets close to either zero or one, the Gini Index G_m gets closer to zero. The closer the Gini Index G_m is to zero, the more pure the group is.

In RStudio, R uses the Gini Index to determine the homogeneity of the data. Since it calculates the homogeneity before and after a split, we can take the difference between those homogeneity values. This tells us which split divided S into the most homogeneous version of the data. This is known as the "improvement" from the split. The split that causes the most "improvement" is chosen over the others because the data gains the most homogeneity from the split.

Using our previous example, we could get a tree such as our example tree in Figure 1. In that case, we can see that the tree has been split in two areas. Thus, we can see there are $K = 3$ groups. The

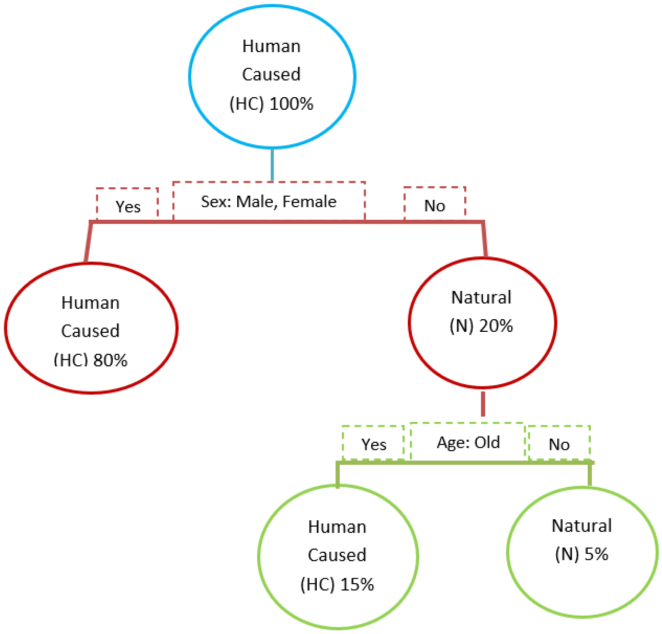


Figure 1: Example Classification Tree

homogeneity would have been calculated before and after the split. We can see that comparing sex (X_2) and age (X_1), sex was chosen. This means that the homogeneity before and after produced the highest improvement in comparison to the improvement age would provide. This process is continued until the remaining splits are not significant in describing the response variable: purifying the data (Loh, 2011). This is known as the stopping point. Then the predictor variable Y can take a value if its corresponding X belongs in one of the disjoint sets (Loh, 2011). Thus, each response Y is classified into some group which was created by the predictors X.

The other way one could determine the partitions is using an entropy equation. This equation works in a similar process, The entropy equation is as follows

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \dot{\log} \hat{p}_{mk}, \quad 0 \leq \hat{p}_{mk} \leq 1$$

where K is the number of classes and p_{mk} is the proportion of observations that are in the kth class and the mth group. Similar to the Gini index, a small entropy value D means the mth node is pure (James *et al.*, 2013). The values the entropy gives can be interpreted like the Gini index values, so the smaller the entropy value, the more "pure" the data has become.

CART methods use the Gini Index to calculate the homogeneity before and after the split occurs. We can use the Gini Index before and after the split to calculate how much the tree has improved in predicting the response of an individual using our data. Recall that a split splits one group into two subgroups. Let's denote the the Gini Index before the split as I_B, the Gini Index of the left subgroup after the split as I_L, and the Gini Index of the right subgroup after the split as I_R. We also need the proportion of the individuals in each of the sub groups to calculate the weighted averages of each of the subgroups. We denote the left subgroup proportion as P_L and the right subgroup proportion as P_R. This is known as "improvement" or the "information gain." We can calculate the information gain (IG) using the equation below:

$$IG = I_B - P_L \cdot I_L - P_R \cdot I_R,$$

where $P_L \cdot I_L$ and $P_R \cdot I_R$ are the weighted averages of the Gini Indexes of each of the sub groups (Emmery, 2018). Information gains for every kind of split using each of the explanatory variables is done. The split with the highest information gain is chosen. If two or more of the information gains is the same, the split is chosen at random from those information gains.

The other main feature of cart is pruning. Classification trees can be pruned using data which suggests which predictors are important, or unimportant, in describing the response (Patel, 2003). Once we have grown our tree, we can use it to predict other individuals. However, if our tree explains our data well and misclassifies predictions using other data, this is known as overfitting. We would like to produce a model which describes our data to the best of its abilities, and applies well to other data. This is where pruning comes in. Pruning is done to prevent the user from overfitting the model to the data.

In order to prune, we must have criterion to decide where to prune the tree. Patel describes a "cost complexity criterion" when pruning Classification and Regression Trees, which uses a misclassification error and a penalty for the size of the tree to prune off the overfitting splits (Patel, 2003). This is also know as cost criterion. The misclassification error is the number of observations the tree misclassifies. For example, we have a bear with an unknown sex and a young age. We would then use our tree and predict the individual would die a natural death. Say this individual ends up dying a human caused death. This would be a misclassification.

Penalty balances the effects of lowering the misclassification rate. The penalty for the size of the tree is calculated depending on the number of nodes in the tree (Patel, 2003). The cost of removing a leaf is in the explanation of the response. The more leaves which are removed, the less the model will explain the data set. Therefore, there is a balance between finding a tree which how a low misclassification rate without overfitting the model. Finishing our example using Figure 1, say that splitting by age brings a high penalty in exchange for a low misclassification of individuals. Then we would prune off where the tree splits by age.

The process for regression tress is similar to classification tree, but regression trees are slightly different. Regression trees have a shorter process in terms of looking for splits. Regression trees predict the sample mean of the response (Loh, 2011). Unlike classification trees, regression trees cannot use the Gini index or the Entropy equation to make the decisions for splits. Instead, regression trees must use a residual sum of squares equation. This equation is looking for one of the disjoint sets mentioned above which minimizes the residual sum of squares value. Splits are variables that cause the most improvement, and hence are most significant when performing a test of independence.

One of the advantages to the CART method is that since the data is presented in a tree structure, the model can be applied considering more than one variable (Loh, 2011). The CART method is relatively easy to use because it does not require transformations, parameter editing, and handles missing data without requiring the data point to be edited or removed (Patel, 2003). However, a flaw of this method is that there is some selection bias within this method (Loh, 2011). It is important to note that this happens rarely, and this is not an issue with the analysis we will conduct. Another flaw to CART methods is that it can be computationally time consuming. Therefore, we use what is referred to as a *recursive binary splitting* approach. This approach splits at each node without considering the tree as a whole (James *et al.*, 2013). This means that we might need to make several trees in order to find one that picks the best splits and cohesively makes the best tree.

A difference between classification and regression trees is that regression trees predict a quantitative response. Classification trees predict a qualitative response (James *et al.*, 2013). Regression trees are easy to read and understand, but they do not have the best accuracy in predicting and smoothness compared to other models. Also, this model is forced to compute some unnecessary computations when using regression models for piecewise, linear models (Loh, 2011). This will not be an issue for our data set since our response is categorical.

Correspondence Analysis

We will also be using correspondence analysis to inspect the relationship between the variables which we find important in explaining cause of death from the classification and regression trees. Correspondence analysis allows us to understand and display associations between categorical variables in a data set. The coordinates for each category are derived using the chi-squared distance. The chi-squared distance looks like the Pythagorean distance by equation, but the difference of each term is divided by the variable's average proportion.

First, we need to start with a two-way frequency table. This is also known as a contingency table. Let's say this table is i by j, where i is the number or rows and j is the number of columns. Also, let $n_{i\cdot}$ be the row totals and $n_{\cdot j}$ be the column totals, where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J$. We would then perform the chi-squared test for independence on the contingency table. To do this, we will also need the estimated expected value which is found using the marginal totals:

$$E_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n},$$

where $n_{i\cdot}$ is the row totals and $n_{\cdot j}$ is the column totals (Everitt and Dunn, 2001). We can then apply the chi-squared test for independence on the contingency table using the following equation:

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - E_{ij})^2}{E_{ij}},$$

where E_{ij} is as stated above under the assumption that the hypothesis of this test is that the row variable and column variable are independent (Everitt and Dunn, 2001). This would then tell us whether we can perform correspondence analysis. Specifically, if there is reason to doubt the null hypothesis, we can then proceed to use correspondence analysis.

When performing correspondence analysis, we must find either the row or column profile. This is a matrix that contains the row proportions calculated from the contingency table. For simplicity, we will show only the row profiles. To calculate the row profiles, we need a matrix P which is the correspondence matrix containing relative frequencies:

$$P = (p_{ij}) = \left(\frac{n_{ij}}{n_{\cdot\cdot}} \right),$$

where n_{ij} is a frequency in row i and column j and $n_{\cdot\cdot}$ is the sum of all the frequencies across i and j (Everitt and Dunn, 2001). We can use this equation to get the row profiles. We use the matrix P and a diagonal matrix with the row vector values on the diagonal, D_r to calculate the row profile:

$$R = D_r^{-1} \cdot P.$$

Note that the row profiles sum approximately to 1, or 100 percent, for each row. We can then use this row profile matrix to get the coordinates for plotting (Everitt and Dunn, 2001). We do this by creating a distance matrix which gives coordinate for each of the variables and their levels with 2 dimensions. These coordinates take into consideration the distance calculated between each of the variables and the levels in the variables.

Since we will be using more than three categorical variables, we must use multiple correspondence analysis. Multiple correspondence analysis is just an extension of simple correspondence analysis. Instead, the matrix for the contingency table is a k dimensional matrix, where $k \geq 3$. We would then proceed through the steps necessary to get the profiles and make the dimension coordinates for each of the categorical variables. This process is simple in R as long as the variables are all factors and do not contain NA values. Our data set has NA values, so we will have to remove the values in order to perform the analysis.

The main purpose of correspondence analysis is to investigate the relationship between categorical variables. Correspondence analysis allows us to display the categorical variable's relationships so that we can understand their similarities and differences. Some advantages to this type of analysis are that correspondence analysis can be used in many areas where we cannot implement other multivariate techniques; correspondence analysis also allows us to view the relationship between the categorical variables by their association.

Some disadvantages to correspondence analysis are that we cannot apply hypothesis testing, we cannot definitively determine the number of dimensions to use, we may miss attributes which are important, it is affected heavily by outliers. Before we apply these techniques, we must understand the data we are working with.

Data Collection

The qualitative data we are collecting includes information about their habitats, where grizzly bears reside currently, and their tendencies. We have a couple papers which study grizzly bears which we have talked about in the background section. The first paper we found was written by McLellan *et. al* (McLellan *et al.*, 1999). They took data from several places such as British Columbia, Alberta, Idaho, Washington, and Montana. They used this data to determine the mortality rates of grizzly bears in those locations. The second paper was written by Mattson and Reid which focuses their study on the Yellowstone grizzly bear populations. This paper considers qualitative data which gives some cause to the mortality rates. Some of these factors include climate warming, food source, human-bear interaction, and management of grizzly bear habitat (Mattson and Reid, 1991). We can use this information to better inform ourselves about potential factors and the results of human efforts through studying the more current data set.

We have also collected data from 2010 to 2020 on the mortalities of grizzly bears from the U.S Geological Survey website. After we compiled the data into a spreadsheet, we analyzed the cleaned data to see what predictors best explain the cause of death in grizzly bears. The results and conclusion from this data analysis will be used to compare the mortalities of grizzly bears in the past ten years and the late 1900s which we will do using the qualitative data we have collected and continue to collect.

We obtained the data by year from the U.S. Geological Survey. Each data set obtained from a specific year was presented in a table. This table was copied and pasted into a spread sheet. The data obtained in 2020 is contains all of the reported or found mortalities discovered that year (US Geology Survey, 2020); the data obtained in 2019 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2019); the data obtained from 2018 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2018); the data obtained from 2017 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2017); the data is obtained from 2016 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2016); the data is obtained from 2015 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2015); the data is obtained from 2014 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2014); the data is obtained from 2013 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2013); the data is obtained from 2012 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2012); the data is obtained from 2011 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2011); the data is obtained from 2010 contains all of the reported or found mortalities discovered that year (US Geology Survey, 2010).

The data we have compiled was collected by the Interagency Grizzly Bear Study Team (IGBST) (Interagency Grizzly Bear Study Team, 2021), a team within the Northern Rocky Mountain Science Center (NOROCK) and formed by the Department of the Interior (DOI). These subgroups are within the U.S. Geological Survey. The main objective of the IGBST is to observe and research the actions done in the Greater Yellowstone Ecosystem for the grizzly bear species. The IGBST specifically monitors radio collared females, the whitebark pine, and females which have cubs born in the current year: COY. The females which have cubs born in the current year (COY) are studied in order to get a better sense of changes in the size of the grizzly bear population. Radio collared bears are tracked in order to help monitor the fluctuations of the population size, help indicate of a healthy population, and help understand the potential of survival. Lastly, whitebark pine trees are monitored because their seeds are a high calorie food in the grizzly bear diet, especially helpful in preparing for hibernation (Interagency Grizzly Bear Study Team, 2021). The specific ways in which they collect this data is not provided.

Data Information and Data Cleaning

We compiled all the data from each year into one file. Then we began to clean and manipulate the data.

Count

The first variable is called *count*. Recall that each data set is data collected in a given year. This variable acts as an index for each bear found in a given data set. No cleaning was done to this variable.

Unique

The second variable is called *unique*. *Unique* is a number the bear is given when it is found. No cleaning was done to this variable.

Sex

The third variable is *sex*, which separates in to three categories: male (M), female (F), and unknown (Unk). Since the formatting of the data has changed over the ten years we have collected, there are two ways unknown was recorded: UNK and Unk. Therefore, we had to compile these two categories together under unknown (Unk) since they are the same.

Ageclass

The next variable is *ageclass*. *Ageclass* is separate into several categories: Adult, COY, Coy, Cub, Subadult, Unk, UNK, and Yearling. An adult bear is a full-grown bear. Coy and a COY stands for cub-of-the-year bears which are cubs born in the current year. Subadult bears are bears that are growing, but they are not yet full grown; these bears range anywhere from 2 years to 8 years. Unk and UNK make up the bears for which we do not know their sex. Finally, yearlings are bears between the ages of 1 and 2 years old. Because the *ageclass* variable has the repeated variable due to change in presentation of data over time, we combined the data of UNK with Unk and COY with Coy respectively.

Known Mortality Date (*kmd*)

The variable Known Mortality Date (*kmd*) includes entries which are dates giving day, month, and year, giving the season and the year, the year, or no data. Because of this, we have split up the *kmd* into three variables: *day*, *month*, and *year*. These dates are reflecting when the estimated time of death was for a particular grizzly bear.

Month We set the *month* variable to represent the months by season. Month is separated into the levels Summer, Fall, and Spring. Because there was only one observation in the winter, it was included in the Fall months. Fall contains the months September, October, November, and January. Spring contains the months March, April, and May. Summer contains the months June, July, and August. Since the data was provided in numeric and season form, we changed the numeric values to their respective months and separated them as indicated above. For data that was not provided, an NA (not available) was provided.

Year *Year* includes years 2010 to 2020, so there are 11 levels. For data that was not provided, an NA (not available) was provided. Also, two individuals were estimated to have died in 2009, but were found in the 2010 data set, so these individuals were included in the year 2010.

Day

Day is represented in numbers for each day ranging from 1 to 31. For data that was not provided, an NA (not available) was provided.

Location

The next variable is *location*. This variable gives us information about where the bear was found or spotted last. We split this variable up into three different categories: general area (*area*), type of land (*type*), and *state*. These categories can be broken down and understood as follows:

Area The variable *area* refers to a general area where the bear was found. This variable contains many different areas, so we combined many of them in terms of bodies of water, land, and manmade/general areas. We split them as follows: Creek, River, Lake, other, and natural land (Nland). Note that the classification other refers to unknown areas and manmade/general areas, including counties, roads, gulches, reservoirs, canals, and towns.

Type The variable *type* describes the land the bear was found on. Since there were many different types of land, we compiled them into national forest (NF), national park (NP), state-owned land (ST), private land (PR), and Bureaus (Bur). The national forest variable includes Beaverhead-Deerlodge National Forest, Bridger-Tenton National Forest, Custer Gallatin National Forest, Caribou-Targhee National Forest, Gallatin National Forest, and Shoshone National Forest. The national park variable includes Grand Teton National Park, Yellowstone National Park, and Wind River Indian Reservation. Note that some of these parks were not explicitly given, but their acronyms were. This is reflected in the original data sets; these data sets can be seen in the grizzlybearraw.csv file found in the Data Sets section of the appendix. These parks were decided based through reasonable conclusions.

State The variable *state* refers to which state the bear was found in Montana (MT), Wyoming (WY), and Idaho (ID). For individuals which did not provide a state, an NA was provided.

Latitude and *Longitude*

Along with the location, we added a *latitude* and *longitude* variable. We made these two categories in order to add an approximated coordinate location to where the bodies were found. In order to do this, we took each of the locations in the location variable, sorted them uniquely and alphabetically in Excel. We then used the information given from the location variable to search for the coordinates of each location. Since many places were rivers and creeks, we estimated the middle of the river, or lake. We also used information given in the location to narrow the distance. For example, if a river covered the majority of a state, but the location includes a national forest, we chose the middle of the river spanning over the national forest. We searched for the coordinates on the internet. For information on the coordinates, please see the attachment marked *Websites for Coordinates*.

Demographic Monitoring Area (*dma*)

The next variable is Demographic Monitoring Area (*dma*). There are two categories in this: Inside and Outside. This variable determines if the bear died within the monitoring area. Note that though no changes need to be made to this variable, this variable was not tracked before the 2015 data (US Geology Survey, 2015). Since dma was not tracked before 2015, we had to add this variable in to the data sets from 2010 to 2014. Though this was added, we did not insert any information into them. Therefore, they are marked with an NA.

Loss

Loss is the last variable in the original data sets. This column tells us if we know how the bear died, if the bear likely died, if the bear likely did not die or if the bear was under investigation. These are labeled known, probable, and under investigation, respectively. This variable also tells us the cause of death: human caused, natural, undetermined, or under investigation. In addition to these two pieces of information, loss also includes a small description of the reason the bear died or the suspected reason the bear died. We are splitting this section into three categories: *body*, cause of death (*cod*), and *specifics*.

Body The variable *body* has two categories: Known and Probable/Not. This section of the variable loss was originally split into three categories: known, probable, and not-probable. Known means that a body was found; probable means a body was not found, but there is strong evidence to suggest the bear died and what it died of; not-probable means that no body was found and there is not enough evidence to suggest what happened to the bear and if it died. We combined the probable and not-probable since these levels were deduced from the scene.

Cause of Death (*cod*) Cause of death (*cod*) is either human caused, natural, likely natural, under investigation, or undetermined cause in the original data set. To simplify this variable, we designated HC as human caused, N as natural and likely natural, and UD as undetermined cause and under investigation. Note that we include likely natural with natural deaths because it was deduced from other evidence that the bear died of natural causes.

Specifics The last aspect of loss is called *specifics*. This variable tells us what happened in detail. Because of this, we also split the specifics variable in to two variables: *krb* and *rfr*.

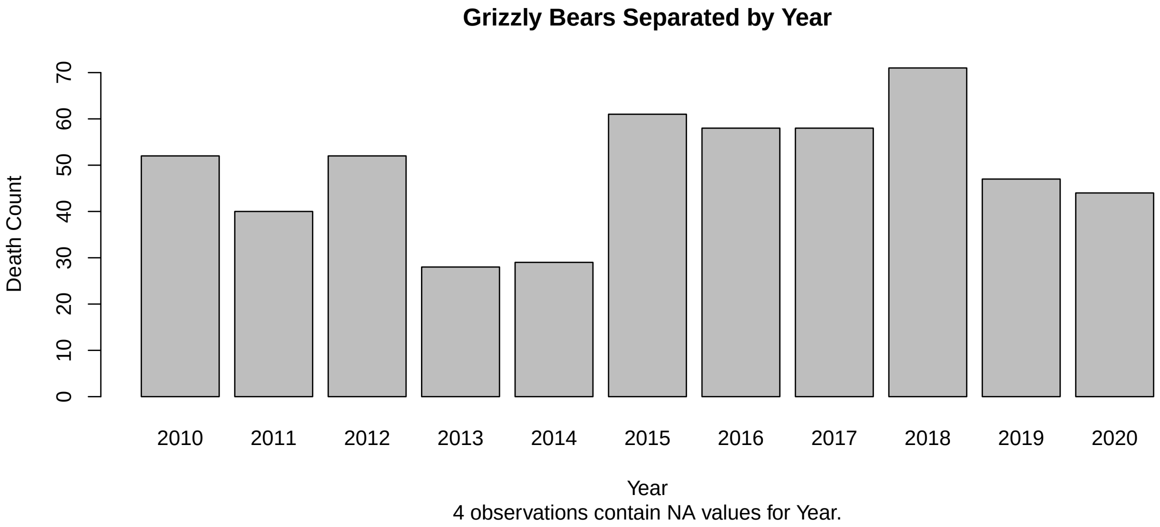
Killed/Removed By (*krb*) The variable *krb* stands for killed/removed by; this variable is categorical and tells us who the bear was killed by or removed by. We designated mgmt as management and Yellowstone National Park (YNP), CHV as civilian, hunter, and vehicle, SWA as survival or wild animal attack, and UD as undetermined for the groupings for who the bear was killed or removed by. Note that the wild animal attack in the variable SWA means the bear was in a fight with a wild animal.

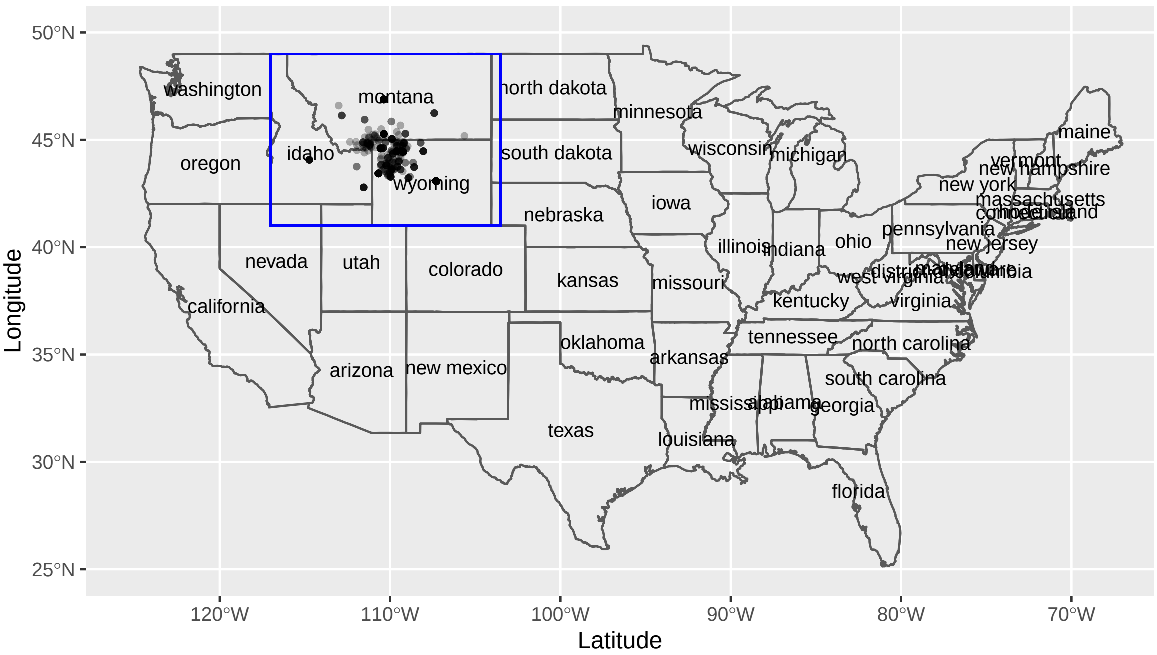
Reason For Removal (*rfr*) The variable *rfr* stands for reason for removal; this variable contains the most severe reason a bear was removed. Some of the bears had multiple reasons for removal, but the most serious reason was chosen for this category. Since there were many reasons for the bears to be removed, we categorized them as the following: bear human conflict (BHC), food depredation (FD), undetermined (UD), safety concern (SC), purposeful kill (PK), and other (OTH). Bear human conflict (BHC) encompasses all encounters between humans and bears in residential areas, mistakenly killed by hunters, property damage, killed on the road, hit with a vehicle, obtaining food, handling the movement of a bear, cubs whose mothers were killed in self-defense, and bears killed in self-defense. Undetermined deaths (UD) included deaths that are under investigation and undetermined reasons for death. Safety concern (SC) encompasses safety concerns among people, bears with bold behavior, bears with aggressive behavior, and bears who have killed a human. The category purposefully kill (PK) contains all the deaths from illegal activity, mistaken identity in which the hunter took the body, killed or shot maliciously, and bears that were killed because they were a nuisance. Other (OTH) contains deaths from accidents, bears that were found dead, predation, injured bears, old age, cubs that lost their mothers, and bear-bear confrontation.

We must note that many of these data pieces under *loss*, *location*, and *kmd* overlap quite a bit because we broke these variables down further. We are splitting these variables into other variables to simplify the output so that we get and better our understanding of what the data is telling us about the cause of death. Please see full table of the variables contained in this data set in the Summary Table of Grizzly Bear Data section of the appendix.

Data Observations (descriptive, exploratory)

The *year* variable tells us specifically how many bears were reportedly found within the Yellowstone Ecosystem area. This area is where Wyoming, Montana, and Idaho meet, and it contains many national forests and parks. We can see specifically that the amount of bear that died each year ranges between 29 and 71 bears between 2010 and 2020. Note that the lowest number of deaths was in 2013, and the highest number of deaths was in 2018.





3 observations have missing values.

Figure 2: US Map of Grizzly Bear Observations

In Figure 2, the data observations are individual bears represented as black dots. Montana, Idaho, and Wyoming are the states where all of the bear individuals died. These are the states where the Yellowstone Ecosystem is, and where our grizzly bear individuals reside. This is also the only place which grizzly bears can live as their population has severely decreased. Also notice that the darker areas are referring to a larger density of deaths. These death are particularly around where Montana, Wyoming, and Idaho meet. In order to better see the individuals contained in Figure 2, we created Figure 3: a map of the area of the map contained in the blue rectangle.

In Figure 3, many of the deaths appeared to be either adult or subadult grizzly bears. Many of the deaths also appear to be male. As mentioned before, many of these individuals are near where Idaho, Wyoming, and Montana meet. This is because there are several national parks or national forests in this area. These are places in which bears are protected in order to maintain or increase their populations. Since these places are also US landmarks, people often visit these areas or live in these areas as well, making it the perfect area for grizzly bear territory to meet human territory.

Determining the ratio of male, female and unknown bears is rather difficult. In order to combat this, we have created a mosaic plot.

The mosaic plot, Figure 4, shows the relationships between the variables separated by every type of grouping. Most notable is that there appears to be a relationship between age and sex. We can see as indicated in red that there are less male and female cubs, male COYs, unknown subadults, and unknown adults than expected. Also, there are more unknown COYs and cubs than expected. From 2010-2020, less appears to be known about the cubs in terms of sex.

Next, we focus on the relationship between who the bear was killed/removed by and their reason for removal. Figure 5 shows just that and is also contained inside the blue square in Figure 2. The green markers indicate bears that were removed because they died of other causes such as

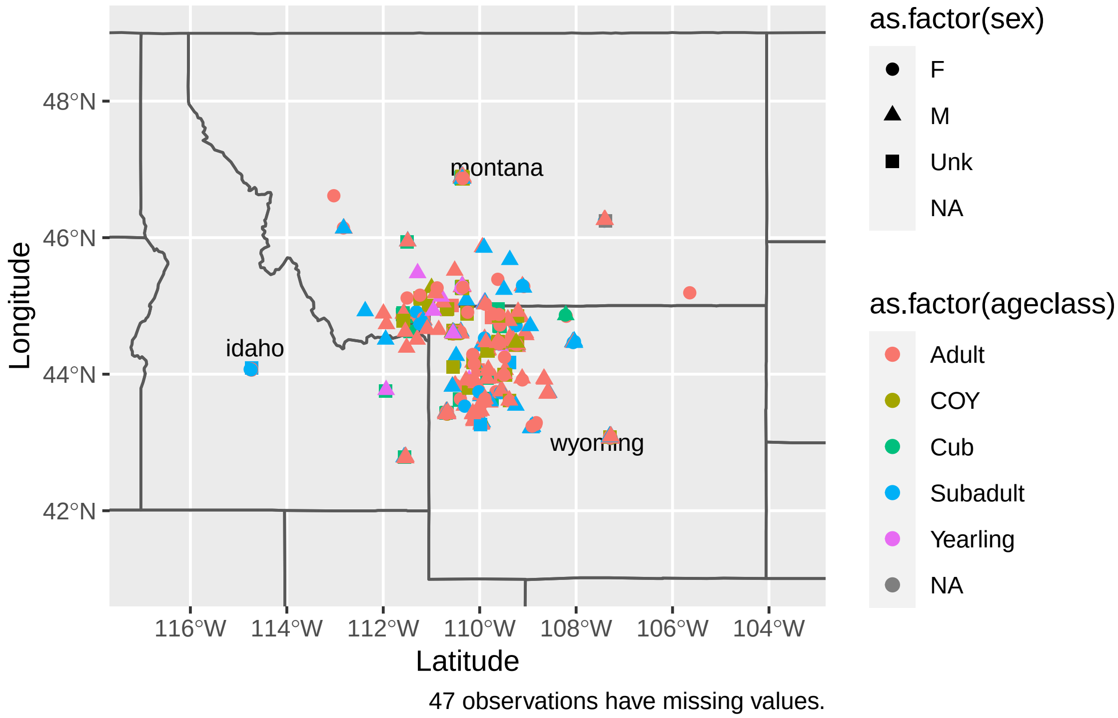


Figure 3: Grizzly Bear Mortalities Based on Sex and Age Class

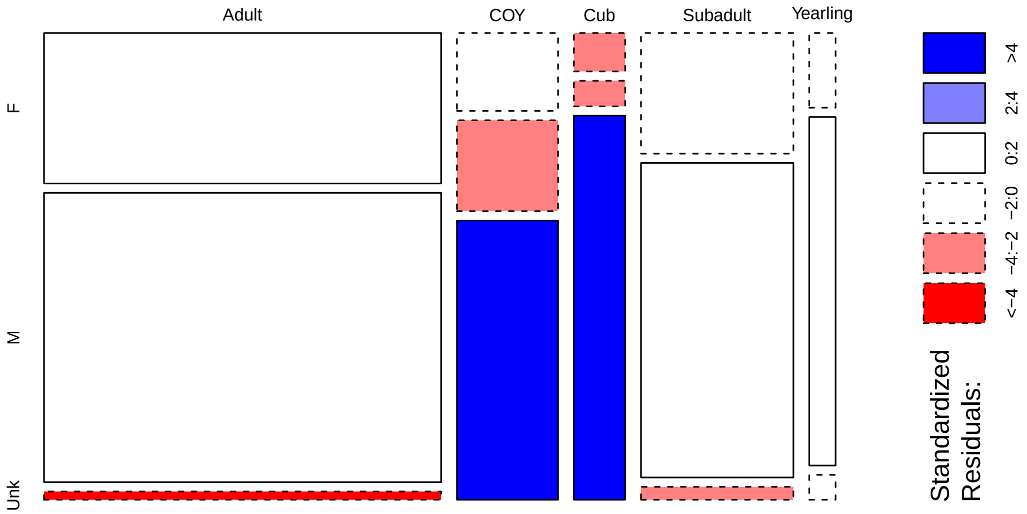


Figure 4: Mosaic Plot Comparing sex and age class.

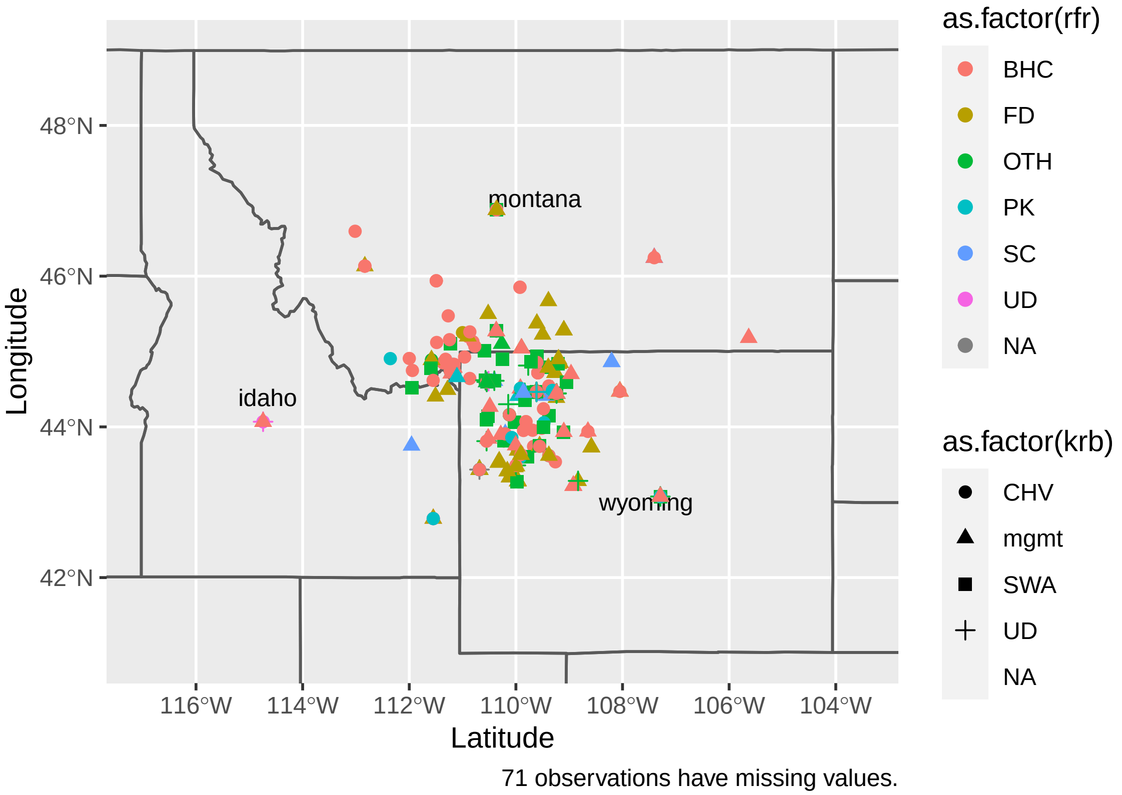


Figure 5: Grizzly Bear Mortalities Based on Reason For Removal and Killed/Removed By

drowning, attacked by a wild animal, sickness, and old age. We can see that these are sprinkled throughout the areas of the national parks and forests though some could be in residential areas. The most prominent reasons for removal appear to be bear-human contact, food depredation and other. These are sprinkled all throughout the map with bear-human contact being the most wide ranging. It appears that the food depredation and other removals were removed by management whereas the reason for removal bear-human contact seems to be killed/removed by either management or a civilian/hunter/vehicle.

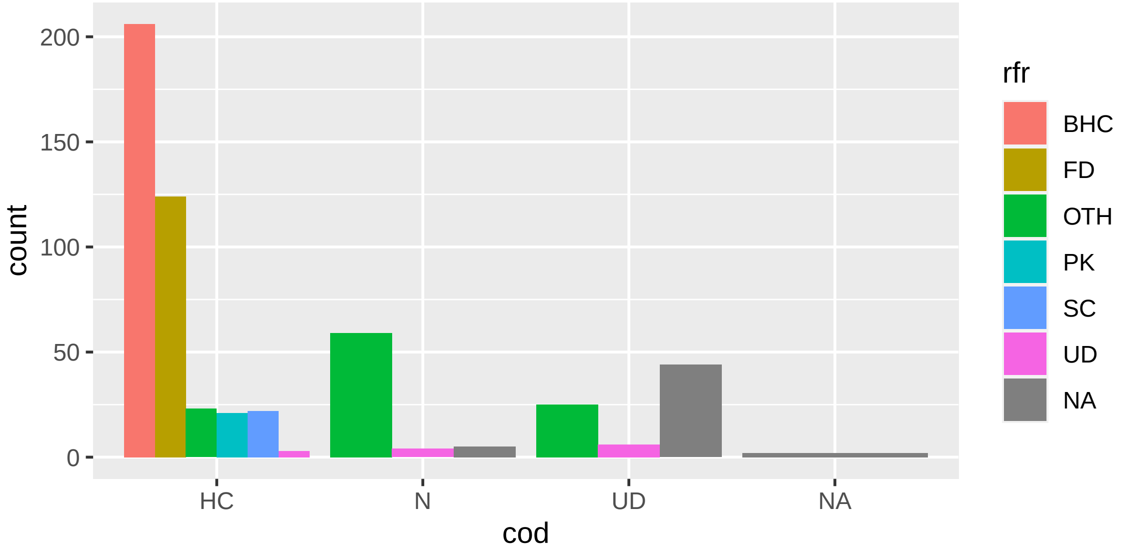


Figure 6: Cause of Death Described by Reason for Removal

The break down of cause of death in terms of reason for removal is displayed in Figure 6. We can see than most of the human caused deaths are because of bear-human conflict. The second highest reason was because of food depredation. In both of these instances, these deaths occurred where the bear territory and human territory overlap. Notice that most deaths that were natural or undetermined were because of causes under the other category, meaning the bear died of a natural accident, attacked by a wild animal, drown, etc.

We also look at the cause of death broken up by *sex*. In Figure 7, male grizzly bears are killed the most in human caused grizzly bear deaths. Most of the natural deaths are unknown. Also, there appears to be about the same number of deaths that are female or unknown in the undetermined cases.

```
## , ,  = F
##
##
##       Adult COY Cub Subadult Yearling
##    HC    72   9   3       26        2
##    N      9   1   0        2        0
##    UD    10   1   0        0        1
##
## , ,  = M
##
```

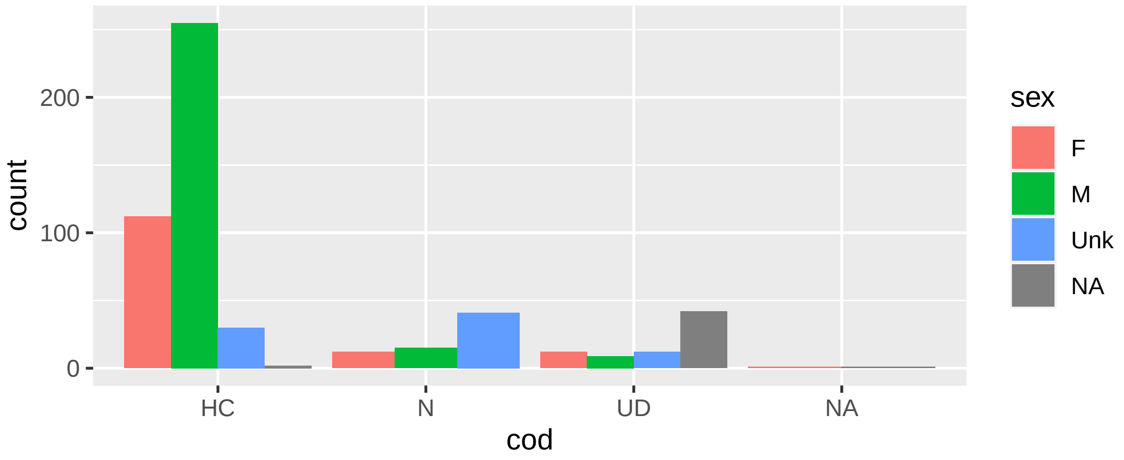


Figure 7: Cause of Death in Terms of the Sex of the Grizzly Bear

```
##
##       Adult COY Cub Subadult Yearling
##    HC   163  10   2       66       13
##    N      6   4   0        4        1
##    UD     6   0   0        3        0
##
## , ,  = Unk
##
##
##       Adult COY Cub Subadult Yearling
##    HC     0  16  13        0        0
##    N      2  24  13        2        0
##    UD     3   3   4        1        1
```

For male and female bears, most cases are adult bears with human caused deaths. The second highest is subadult bears with human caused deaths. However, when the sex is unknown, cubs and cubs born in that year are tending to die. This table is showing that knowing about the cause of death of the bear seems to be easier across age classes for bears where we know the sex of the bear. A relationship between *ageclass* and *sex* might be occurring where age depends on sex. Also, There is consistently undetermined age classes of the bears when the age is undetermined.

Research Data Analysis

Recall that the goal of this paper is to investigate the mortality of grizzly bear populations. We will look at the similarities and differences of the mortality rates of grizzly bears from 2010 to 2020 based on several variables such as age, sex, and location. We will look at these predictors to develop an understanding about the cause of grizzly bear deaths. Through this project, we hope to answer the following questions: How have the grizzly bear mortalities changed compared to the 1990s? What steps can we as humans take to improve the grizzly bear population? We will also explore what is affecting grizzly bears the most and some solutions that have been proposed to lessen the amount of grizzly bear deaths.

Classification and Regression Tree Analysis

So far, we have collected both quantitative data and qualitative data to further understand the behavior of grizzly bears as well as their current population and mortality rates. We will now use this data to build information about the grizzly bear population and their mortalities currently. Then we can compare these results to results found in the late 1900s.

In order to find the variables in the grizzlybear data set, we use CART methods in order to see which variables best describe the response: cause of death (*cod*). Below we are fitting the data for a model. The following tree is fitting the cause of death using all of the variables we are interested in. These variables are *sex*, *ageclass*, month the bear was found (*month*), day the bear was found (*day*), year the bear died (*year*), general area the bear was found (*area*), the type of land where the bear was found (*type*), *state*, demographic monitoring area (*dma*), the whereabouts of the body (*body*), killed/removed by (*krb*), and reason for removal (*rfr*).

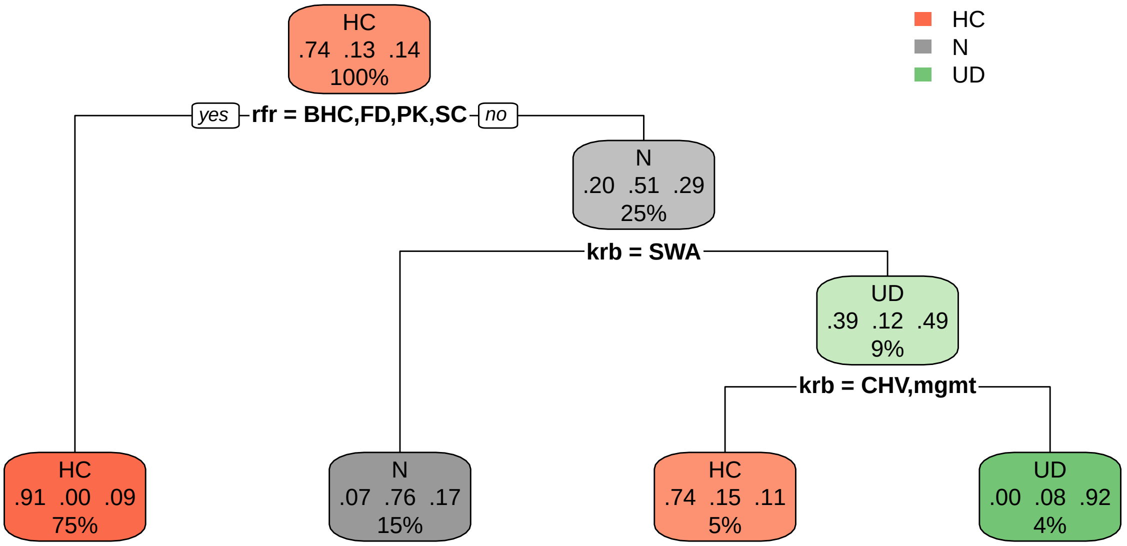


Figure 8: Classification Regression Tree Containing Variables sex, ageclass, month, day, year, area, type, state, dma, body, krb, and rfr.

From the Figure 8, we see that the primary splits occur with the variables reason for removal (*rfr*) and killed/removed by (*krb*). This means that the reason the bear was removed and the who/what killed or removed the bear is prominent in explaining the cause of death. For the fitting of this tree, we included the variables *sex*, *ageclass*, *month*, *day*, *year*, *area*, *type*, *state*, *dma*, *body*, *krb*, and *rfr*. However, with all of these variables, reason for removal (*rfr*) and killed/removed by (*krb*) are best in describing the response.

We can see from the tree that if the reason for removal was from bear-human contact (BHC), farming disturbance (FD), purposeful kill (PK), or safety concern (SC), these deaths were mainly human caused. If the reason for removal (*rfr*) was not one of the previous reasons, it was determined that the bear died of natural causes (N). The next primary split occurs at node three; for the group which was natural after the first split, the split within this group is split by the variable killed/removed by (*krb*). Note that the reason for removal of these bears are other (OTH) or undetermined (UD). If the bear died trying to survive or was attacked by a wild animal (SWA), the cause of death is considered natural (N). The bears who are not in the group SWA are considered Undetermined (UD). Of those undetermined deaths, if they were killed/removed by management (mgmt) or a citizen/hunter/vehicle (CHV), the cause fo death is human caused. Those which do not have one of the previously mentioned *krb* values is ruled as undetermined. It is also important to note that as we move farther down the tree, the characteristics of the bear are determined by the previous splits.

For more information on the splitting decisions, we can use commands such as the summary command and the printcp command. For this model fitting, R is calculating the amount of improvement that is made by splitting according to each of the variables. We can also see the surrogate splits, which also help to explain the cause of death. However, these surrogate splits do not make the most improvement, so they are not used. This information is the background behind making the nodes of the tree. The values at the top of what is produced in the summary command are made into a plot using the plotcp command.

The plotcp command gives us a graphical depiction of the x-val relative error. We can use this graphic to see where to prune our tree. For this model, *R* is suggesting we prune at the third split. This means that we could remove the split of the undetermined cause of death by killed/removed by without significantly losing explanation in the response. To see the summary command and print command of the model fit, see the Classification and Regression Tree Code section in the appendix.

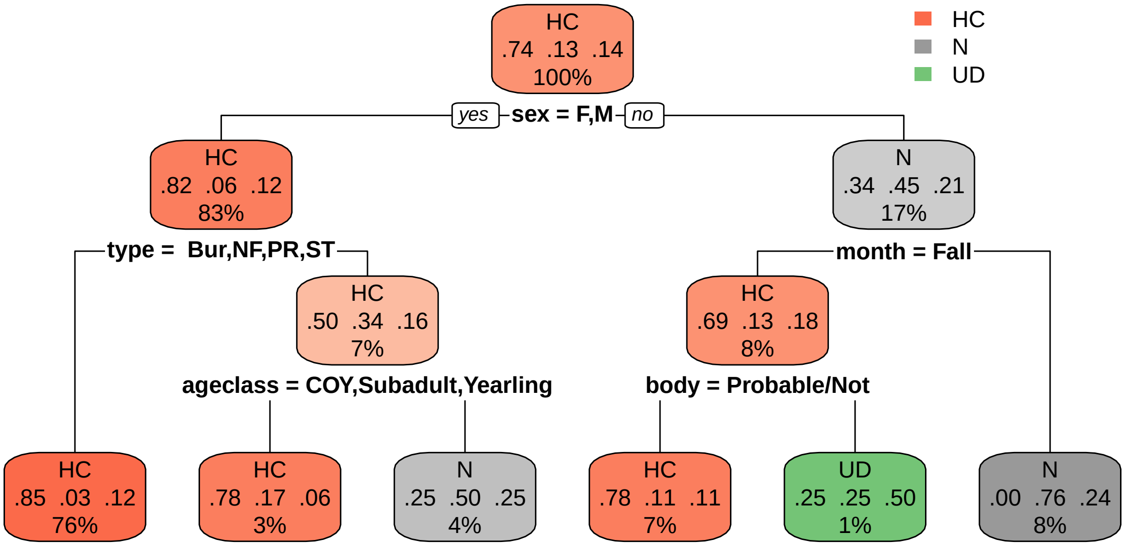


Figure 9: Classification Tree Containing Variables sex, ageclass, month, day, year, area, type, state, dma, and body.

The tree in Figure 9 removes the two variables which seemed to most overtly explain the cause of death: reason for removal (*rfr*) and killed/removed by (*krb*). This tree uses the variables *month*, *sex*, *type*, and *year* to describe cause of death. The data states that the most important variable is *sex*.

We can see this reflected in the tree and the data as it is the first split. At node 1, sex improves our understanding of cause of death the most. If the sex is known, male (M) or female (F), the main cause of death is human caused. If sex is unknown, the cause of death is considered natural.

At node 2, the most important split is by *type*. We can see from the classification tree that if the type of land is state owned, private land, a national forest, or a bureau, cause of death of male and female bears is human caused. If not, it is also human caused, but there is a split in this cause of death, which is discussed in node 3. Adult bears who were handled outside of direct management were mainly human caused deaths. The group that includes management includes a natural portion.

At node 3, the most important split is *month*. If the month is in the fall, the cause of death of a bear with an unknown sex is human caused. However, if the bear died in the Spring or Summer, the cause of death is natural. This is not unsurprising since hunting season for other animals is mainly in the winter, though there are some summer or spring seasons.

The fifth node shows a split by *ageclass*. Individuals who were male or female, were found on the type of land state owned, private land, a national forest, or a bureau are split by whether they are young or old. Bears who were cub of the years, yearlings, or subadults had deaths which were human caused. Bears who were adults or cubs die a natural death. Cubs appear to die a natural death in comparison to the ages around theirs: cub of the years (COYs) and subadults. This could be because they are old and stroung enough to keep up with their mothers, but still under their mother's protection.

Lastly at node 6, a split was determined from the variable *body*. If the bears death was probable or not probable, it was deemed human caused, under the constraints that their deaths were undetermined, and they died in the fall. If these bears death was known, the bear usually died in an undetermined manner.

We can see that most of the deaths are human caused and very few deaths are undetermined. Also, we get the values of the CP and relative error as well as the error in this approximation which are reflected in the graph of cp and x-value relative error. Please look at the Classification and Regression Tree Code section of the appendix for these graphics. Important nodes in this tree are the first and second, but the tree can be pruned at the first node. This means for these variables, the most important factor in explaining cause of death is the sex of the bear.

The fit that occurs in Figure 10 takes out the important variables from our second fitting as well as the first fitting. The remaining variables are *ageclass*, *day*, *area*, *state*, *dma*, and *body*. These variables alone have a hard time explaining the cause of death without splitting many times. This fitting has such a problem, that none of these splits at each of the nodes we see helps to better explain what is happening in terms of the cause of death. The X-val Relative Error plot in the Classification and Regression Tree Code section of the appendix shows that none of these nodes are worth splitting.

However, when we look at the splits that were created, we can see that many of the splits are by day for cubs and cub of the years. These splits could be due to hunting season. When hunting season occurs, grizzly bear can be mistaken for black bears. Female grizzly bears who are killed and who have cubs of their own often leave them behind to fend for themselves. A cub without a protector are more susceptible to getting killed. Therefore, there are two takeaways from this fitting; we must include the variables that have been removed to better explain cause of death; *day* might be splitting from hunting season through the the day cubs die after the loss of their protectors.

This fourth fit in Figure 11 reintroduces the variable killed/removed by (*krb*) back into the second fit. We can see that this addition changes increase the importance of splits. The first split is

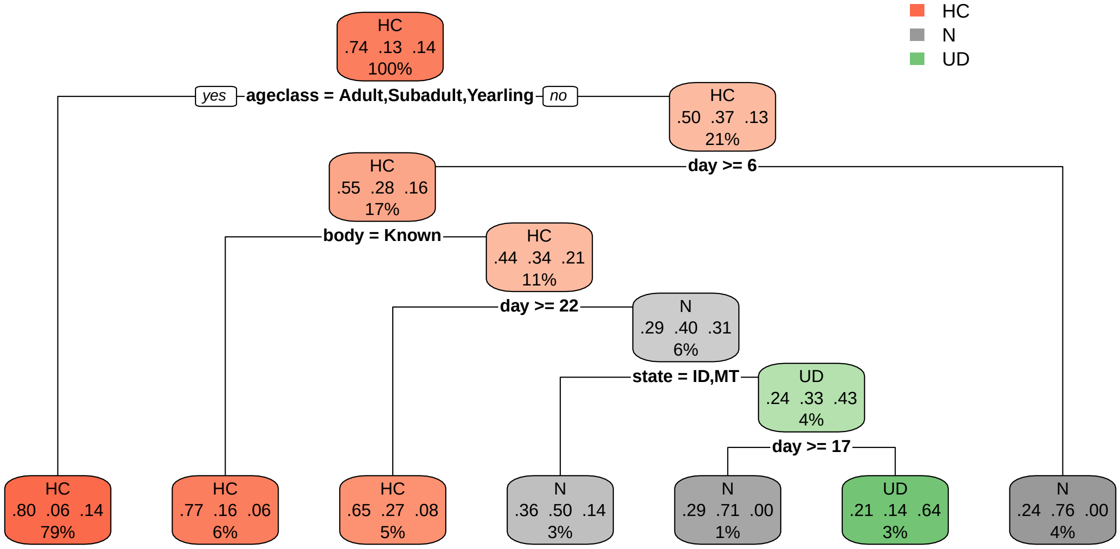


Figure 10: Classification Regression Tree Containing Variables ageclass, day, area, state, dma and body.

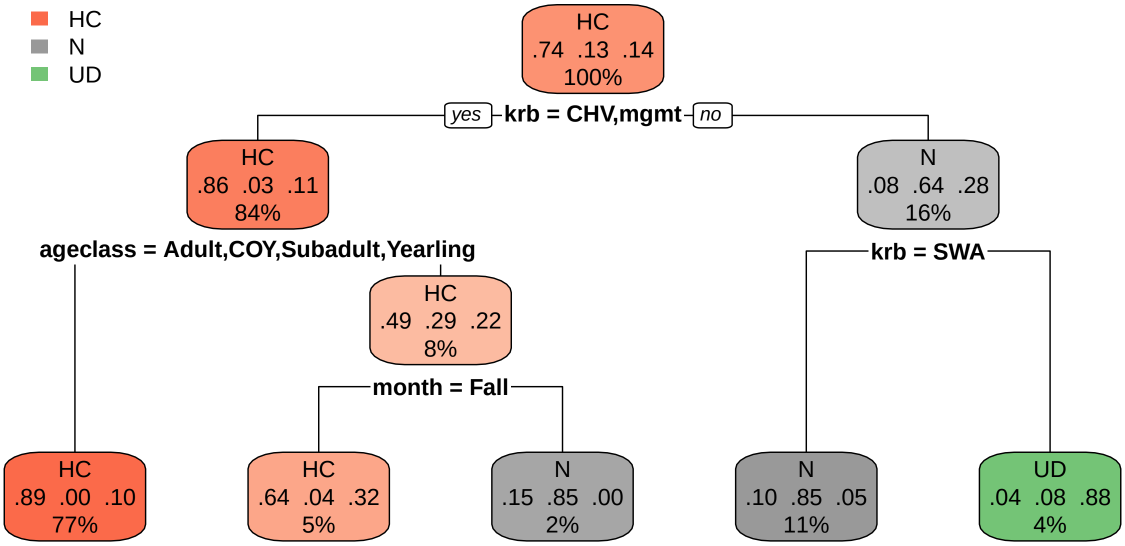


Figure 11: Classification Regression Tree Containing Variables krb, sex, ageclass, month, day, year, area, type, state, dma, and body.

in *krb*, depending on whether the bear was killed/removed by management (mgmt) or a civilian/hunter/vehicle (CHV). If the bear was killed/removed by one of these two categories, the cause of death is human caused. If not, the cause of death is natural.

The next most important split is by age class at node 2. If the *ageclass* level is Adult, COY, Subadult, or Yearling, the death is human caused. If the age class is not one of those groups, the cause of death is still human caused, but this node can be split further. We will get into this split when we talk about the split at node 5. Before that, the next most important split is at node 3. This split is from the variable *krb* again. If the *krb* is in the survival/wild animal (SWA) category, the cause of death is natural. If *krb* is not in the categories SWA, mgmt, or CHV, the cause of death is undetermined.

The last important split is at node 5. This explains that if the bear died between August and December, the death is human caused. If the bear dies in the Fall, the cause of death is natural. Note that the cp and X-value relative error suggest that the split occurring at node 3 is not necessary. We could prune off this split in our classification tree. Note that the cp and x-value relative error information can be found in the Classification and Regression Tree Code section of the appendix.

When we are predicting the number of deaths in the grizzly bear populations, we will need to consider the variables reason for removal (*rfr*), killed/removed by (*krb*), sex of the bear (*sex*), month the bear died (*month*), and how old the bear was (*ageclass*). We will also include the year to see how the years differ in terms of death.

We can then compare our results to the results found in the late 1900s. Finally, we can use the comparison to the results in the late 1900s to investigate what could be done to better the grizzly bear population and minimize the amount of grizzly bear mortalities in the future. First, we must analyze the variables relationship with one another.

Correspondence Analysis

In this section, we will be performing correspondence analysis on the variables deemed important from the previous section: reason for removal (*rfr*), killed/removed by (*krb*), sex of the bear (*sex*), month the bear died (*month*), year the bear died (*year*), and how old the bear was (*ageclass*). In order to do so, we must prepare a section of the data and ensure that each of the variables is a factor variable.

From Figure 12, we will look at the row and column profiles before talking about associations between variable levels. The mass of data which has congregated in the red square will be analyzed with the rest of the data, but we will zoom into this area to talk about it further later. First, we will look at the row profiles. We can see that Spring and 2014 have similar profiles over the columns. The undetermined reason for removal has a similar profile to 2012 over the columns. Cub, 2011, and 2015 have similar profiles over the columns. Some of the larger groups who have similar profiles over the columns are yearling adult, and 2016, food depredation, 2020, management, and safety concern, human caused death, bear-human contact, 2019, 2018, and sub adult, and Summer, 2010, 2013, fall, 2017, and civilian/hunter/vehicle.

The column profiles are over the rows. Undetermined cause of death and undetermined killed/removed by have similar profiles. Food depredation, 2020, and Summer, management, safety concern, 2019, yearling, and other reason for removal, 2018, Subadult, Adult, 2015, and Spring, purposeful kill, 2016, 2015, 2011, survival/wild animal, and natural cause of death have

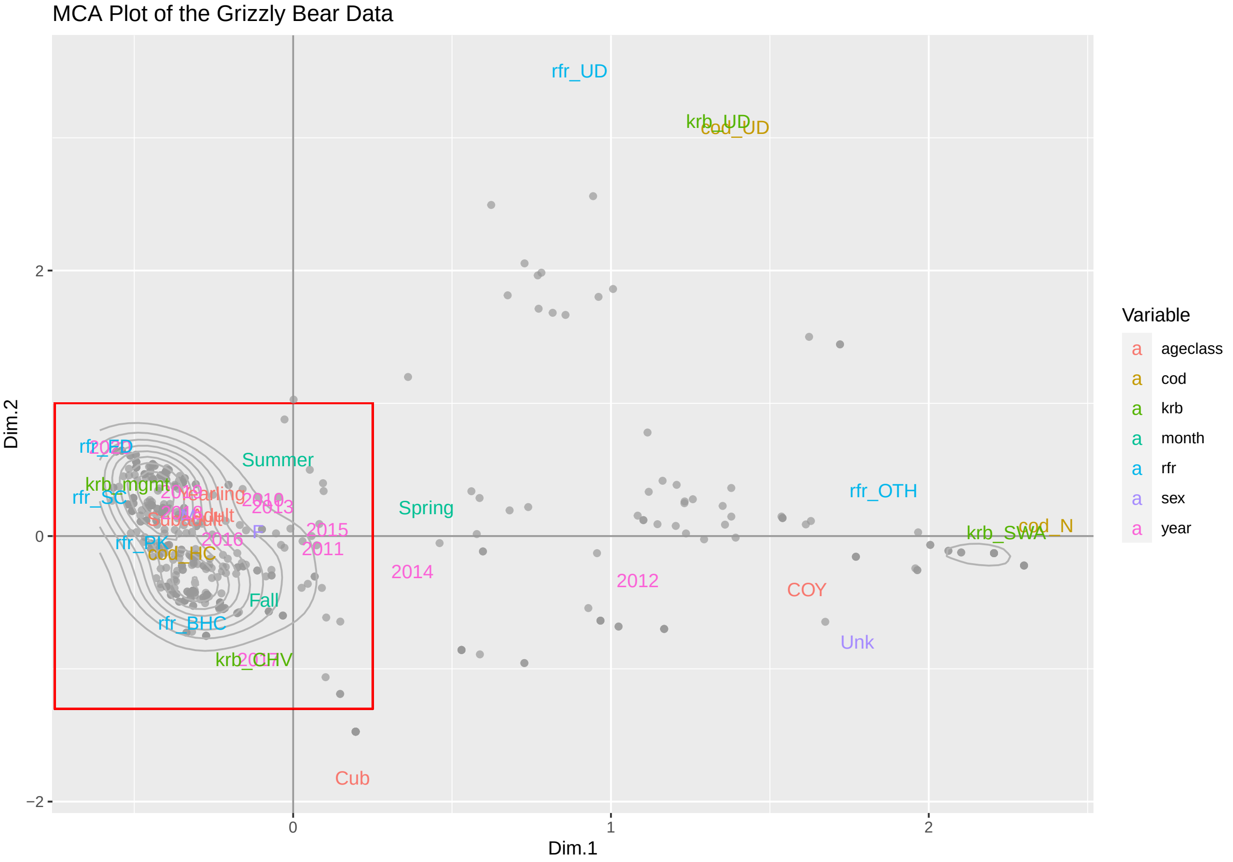


Figure 12: Multiple Correspondence Plot of the Variables ageclass, cod, krb, month, rfr, sex, and year.

similar profiles over the rows. Also, 2014, 2012 and COY, bear-human contact and unknown *ageclass*, and civilian/hunter/vehicle and 2017 have similar profiles over the rows.

The Undetermined level in killed/removed by (*krb*) is closely associated with the Undetermined level in cause of death (*cod*). Undetermined cause of death and undetermined killed/removed by are closely related with the undetermined level in reason for removal. We can also see that the natural level in cause of death is closely associated with the survival/wild animal level in killed/removed by. These are also associated with the other level in reason for removal(*rfr*), COY *ageclass*, and Unknown *sex*.

Another important separation is the year. Notice that 2012 is the farthest from all of the other years. Age classes Cub and COY are also not as associated with the other age classes. We can also see some individuals which are forming a cluster near survival/wild animal (SWA) and natural cause of death. Each of these comparisons is important to note as we move to the area contained in the red square because it still applies to Figure 13. Since Figure 13 is just a subplot of Figure 12. Therefore, the scaling of this subplot is different, and everything contained in the plot is considered close in Figure 12.

For the dimensions of Figure 12, it appears that dimension 1 is the cause of death for grizzly bears. The data is naturally separated over the cause of death variable with human cause in the left center, natural death on the right center, and undetermined cause of death in the top middle. Dimension 2 appears to be whether we knew how the bear died or not. It could be how much information we have on the bears.

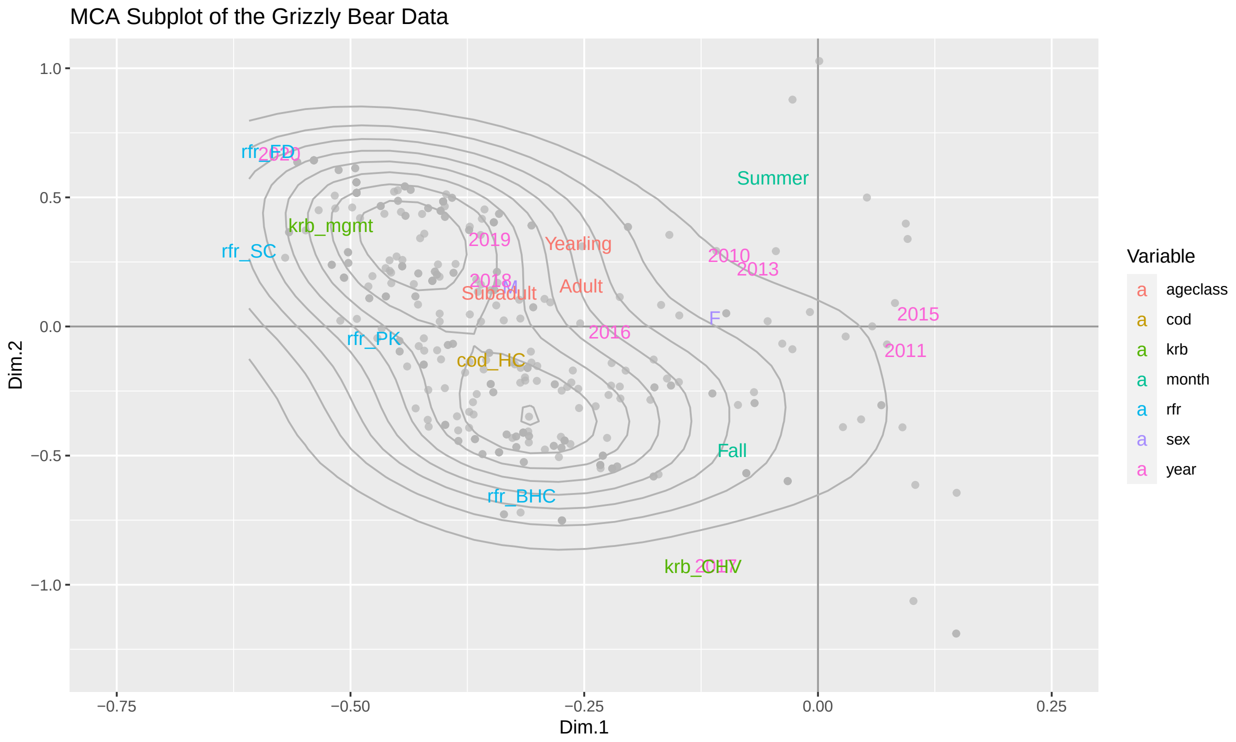


Figure 13: Multiple Correspondence Subplot of the Variables ageclass, cod, krb, month, rfr, sex, and year.

Figure 13 is a subplot of Figure 12 as reference by the red box. Note that the comparisons over the rows and columns have already been done, so we will move on to the associations. In the subplot, we can see that the years are spread across the cluster. Human cause of death is at the center of this cluster. We can see that the sex male is closely associated with the years 2018, 2019, and 2016, the age classes Subadult, Adult, and Yearling, and these are all associated with human caused death. We can see that as the years progress, there is fluctuation between how close they are to human caused death. For example, through 2016, 2018, and 2019 are closely associated, the years 2010, 2011, 2013, 2015, 2017, and 2020 are farther away. The years that are not mentioned do not appear in this subplot, so they are not closely associated with human caused death. Also, it appears reason for removal bear-human caused (BHC), purposeful kill (PK), and safety concern (SC) are most closely related with human cause of death. Also, the killed/removed by (*krb*) variable seems to change throughout the years as well. Notably, 2017 is closely associated with civilian/hunter/vehicle (CHV), meaning in 2017, there were many civilian, hunter, or vehicle killings. On the other hand, 2018 through 2020 are more closely associated with the removal of bears by management.

When looking at the month, we can see that the summer months are more associated with 2010 and 2013. The fall months are closely related with 2017 and the bear-human contact reason for removal. Also, the sex female is close to the years 2010, 2013, and 2016. The sex Female is also close to the Subadult and Adult *ageclass* levels. We emphasize that this is a subplot of the full plot. Because of this, the scaling is much smaller. Therefore, though we have said some of these points are not as close as others in Figure 13, they are all close in comparison to the others in Figure 12.

More information on the coordinates of the correspondence analysis plots can be found in the Correspondence Analysis Code section of the paper. The coordinates correspond to where the variables are place on the map. We also provide information on the eigen values. In order to completely explain the variation occurring, we need 28 dimensions as well as the amount of variance which is explained by adding each dimension.

Summary of Results

In this study, we wanted to study a data set on the mortalities of the endangered grizzly bears. We decided to use CART methods and Correspondence Analysis to better understand the relationship of other variables to cause of death from 2010 to 2020. Better understanding what is happening to the grizzly bear population will help us to better preserve and protect their species and environment. Using the grizzly bear data collected by the IGBST, we used CART methods to deduce which variables would be most important in explaining the response cause of death. As we reduced the variables that were most important in explaining cause of death, we also noticed that there may be some kinds of relationships between some of the variables such as sex and age class. Variables reason for removal (rfr), killed/removed by (krb), sex, ageclass, and month.

We then used Correspondence Analysis in order to determine the associations between each of the variables and their levels. From this analysis, the variables for more recent years, 2015 to 2020, are closely associated with human caused death. Ageclasses Adult, Subadult, and Yearling are also associated with these years. It appears that these human caused deaths are associated with the Fall and Summer months, but Spring is not too far away either. Management and civilian/hunter/vehicle killed/removed by levels are closely associated with these years, ageclasses, and months. Food depredation, bear-human contact, safety concern, and purposeful kill reason for removal are also closely associated with the previously mentioned ageclasses, years, months, and killed/removed by levels. On the other end, each of the cause of death variables appear to be unrelated and are related to certain groups which they were killed or removed by. The natural cause of death is closely associated with the killed/removed level survival/wild animal attack and the reason for removal level other. The undetermined cause of death is closely associated with the killed/removed level undetermined and the reason for removal level undetermined.

Conclusion and Future Work

In this study, we used CART methods and Correspondence analysis to better understand the relationship between the variables in the IGBST collected grizzly bear data. We found that the years 2015-2020 have been closely associated with human caused death for grizzly bears, though years like 2012 and 2014 were being pulled towards the natural cause of death.

Many of the reasons bears were killed by humans was over land and resources. Though bears were removed by management, they were also removed by civilians, hunters, or vehicles. It is clear that the main cause of grizzly bear deaths have been human caused, and these deaths occur when bears are looking for food, have contact with humans, are near camp sites, or are killed purposefully. Most of the encounters people have with grizzly bears are outside of their hibernation period, including the hunting seasons in summer and fall. Also, about half of the population in this study was a male grizzly bear. Of the deaths that were human caused, about half of them were male. Also, more adult bears were killed compared to subadults. It is hard to know if there was an improvement in the number of deaths from the two studies since the nature of the data is different. Also, the access to information about grizzly bears was few and far between then.

When comparing this information to McLellan *et. al*'s (McLellan *et al.*, 1999) work, the recommended that minimizing mortalities, especially adult females, was key to conserving their species. As we can see from the study, just over a fourth of the deaths were females and a little over half were males. Males in the past 10 years have had the most deaths though most of them are adults. Subadult bears trail behind the adults. Note that this cannot be compared to show improvement because the subpopulations by age are not known. According to the Interagency Grizzly Bear Committee (IGBC), the population is estimated to be over 700 grizzly bear now; this rebound is wonderful in comparison to the population of as low as 136 in 1975 (Interagency Grizzly Bear Committee (IGBC), 2021). In our study, there were 544 individuals spanning over 11 years. We saw from previous that the number of individuals who dies in 1 year is no greater than 71. In 2019, this data set contains 47 cases of grizzly bear mortalities. Let's the population is exactly 700; About 6.71 percent of the population died in 2019.

Grizzly bear mortalities seem to center around their encounters and relationships with humans. This data shows that most of the recorded mortalities are human caused. Though not all of them are negative and most of them appear where human activities meets bear habitats. Protections through the endangered species list has provided a new kind of management of grizzly bear habitats, improving the control over the grizzlybear mortalities McLellan *et. al* (McLellan *et al.*, 1999) mentioned. Though accidents and illegal acts cannot always be prevented, this removes a direct predator for the moment. As the status of the grizzly bear fluctuates between endangered and threatened, we must provide more analysis before we can conclude whether or not the grizzly bear population will survive without human help. At this point, the grizzly bear population threatened, though it has moved back and forth between threatened and endangered in the 2000s. This might indicate a relationship between the protection of species and if their population can remain viable without aid from humans. With more analysis, we may be able to determine if this is so.

As the grizzly bear population continues to grow, the space in which human-bear interactions occur will only increase. Unfortunately, this could aggravate the bear-bear contact (Mattson and Reid, 1991). A possible solution McLellan *et. al* emphasized was the need to control the mortalities (McLellan *et al.*, 1999) through managing their land. Bears tend to be solitary creatures other than females taking care of their young. Since bears want to avoid humans as much as possible, they are forced into a smaller area where they feel comfortable, but then their territories run into one another. Territory relationships is a key factor because as we help increase the grizzly bear population, but the protected space stays the same, fighting among grizzly bears and grizzly bears in human territories may increase. In our data, we are seeing many human caused death one in particular being food depredation. Food depredation is when grizzly bears seek food from farm animals or crops. It appears that the human-bear interactions might suggest these possibilities: grizzly bears are getting more comfortable going into residential areas and interacting with human spaces, grizzly bear food supplies is decreasing, or grizzly bears need more space to live.

In order to help the grizzly bear population be conserved and healthy, control and management over their habitat and hunting season is crucial since some of the changes in their habitat are not controllable. Also, monitoring of areas where grizzly bear and humans intersect is necessary. Minimizing the deaths caused by humans would be the most beneficial for this population. Areas this can be improved might be providing information about the proper way to handle grizzly bears for those who live near their habitats, limiting the closeness of human and bear habitats, and providing other means of bear removal: tranquilizing and moving when possible. Another important aspect of understanding the grizzly bear population is continuing to collect data on their whereabouts.

If we were to continue this study, we would want to look into some of the other aspects of the necessities grizzly bears need to survive. Necessities include their food source, their habitat space, global warming, their population, and their genetic make up. We could then use this data set to explore these changes and how this could be reflected in their mortalities. We would also like to begin modeling the data to explore and perform other, numeric testing.

Appendix

Data Sets

```
# The original compiled data sets from the USGS website.
library(readr)
grizzlybearraw <- read_csv("grizzlybearraw.csv")
#View(grizzlybear)
```

```
# Data set used in the Data Observations section and the Classification
# and Regression Tree Analysis
library(readr)
grizzlybear <- read_csv("grizzlybear.csv")
#View(grizzlybear)
```

Note that the code for the data transformation and final data set used for the Correspondence Analysis can be found in the Correspondence Analysis Code section of the Appendix.

Key to Data Set

- BDNF = Beaverhead-Deerlodge National Forest
- BHC = bear human conflict
- BLM = Bureau of Land Management
- BOR = Bureau of Reclamation
- BTNF = Bridger-Teton National Forest
- CGNF = Custer Gallatin National Forest
- CHV = civilian, hunter, and vehicle kills
- Crk = creek
- cod = cause of death
- COY = cub of the year
- CTNF = Caribou-Targhee National Forest
- dma = demographic monitoring area
- GNF Gallatin National Forest
- GTNP = Grand Teton National Park
- HC = human-caused
- IDFG = Idaho fish and Game
- kmd = known mortality date
- Known = grizzly bear body found

- mgmt = management which handles bears
- NTFWP = Montana Fish, Wildlife and Parks
- N = natural
- PK = purposeful kill
- PR = private land
- Probable = body not found but strong evidence of death
- rfr = reason for removal
- SC = safety concern
- SNF = Shoshone National Forest
- ST = state owned land
- UD = undetermined
- UNK = Unknown
- WRIR = Wind River Indian Reservation
- YNP = Yellowstone National Park

Summary Table of Grizzly Bear Data

```
complete_data_description <- read.csv(file = "completedatainformation.csv",
                                      header = T)
```

```
library(pander)
pandoc.table(complete_data_description,
             justify = c("left", "left", "left", "left", "left"),
             split.cells = c(7, 9, 9, 16, 25),
             caption = "Summary Table of Grizzly Bear Data")
```

```
##
## ---------------------------------------------------------------------------------------
## Variable    Type        Type.in.R   Levels             Description
## ----------- ----------- ----------- ------------------ ----------------------------
## count       category    numeric     1 to 61            This variable acts as an
##                                                        indexing variable for
##                                                        each original dataset.
##
## unique      category    numeric     544 levels         A specific number given
##                                                        to the bear when it was
##                                                        found including the year
##                                                        as the first four digits
##                                                        followed by two more
##                                                        digits.
```

```
##
## sex         category    character   male (M), female   This variable indicates
##                                     (F), unknown       the sex of the bear.
##                                     (Unk)
##
## ageclass    category    character   Adult, Cub of      This variable indicates
##                                     the Year (COY),    the age of the bear.
##                                     Cub, Subadult,     Note: COY means a cub
##                                     Yearling           born in the spring; a
##                                                        yearling is a cub between
##                                                        1 and 2 years old.
##
## kmd         character   character                      Known mortality date
##                                                        (kmd) was the estimated
##                                                        time of death of the
##                                                        bear. This variable was
##                                                        broken into three new
##                                                        variables: day, month,
##                                                        and year.
##
## month       category    character   Spring, Summer,    The months the bear's
##                                     Fall               body was estimated to
##                                                        have died.
##
## day         numeric     numeric     1 to 31            The day the bear's body
##                                                        was estimated to have
##                                                        died.
##
## year        numeric &   numeric     2010 to 2020       The year the bear's body
##             category                                   was estimated to have
##                                                        died between 2010 and
##                                                        2020.
##
## location    character   character                      The approximate location
##                                                        of the bear's body. This
##                                                        variable was broken into
##                                                        three new variables:
##                                                        area, type, and state.
##
## area        category    character   Creek, Lake,       The type of land the bear
##                                     Natural Land       was found near. Note that
##                                     (Nland), other,    other contains man-made
##                                     River              land.
##
## type        category    character   Bureau (Bur),      Range of land in terms of
##                                     National Forest    juridiction.
##                                     (NF), National
##                                     Park (NP),
```

```
##                                     Private Land
##                                     (PR), State (ST)
##
## state       category    character   Montana (MT),      The state the bear's body
##                                     Wyoming (WY),      was found in.
##                                     Idaho (ID)
##
## latitude    numeric     numeric                        The East-West coordinate
##                                                        portion.
##
## longitude   numeric     numeric                        The North-South
##                                                        coordinate portion.
##
## dma         category    character   Inside, Outside    Demographic Monitoring
##                                                        Area is areas which are
##                                                        monitored regularly.
##                                                        This variable tells us if
##                                                        the bear was inside this
##                                                        monitoring area. Note:
##                                                        this variable was added
##                                                        by the Interagency
##                                                        Grizzly Bear Study Team
##                                                        (IGBST).  beginning in
##                                                        2015.
##
## loss        character   character                      Information on how the
##                                                        bear died including if
##                                                        the bear died, how the
##                                                        bear died, and why the
##                                                        bear died. Note: This
##                                                        variable is broken into
##                                                        3: body, cod, and
##                                                        specifics.
##
## body        category    character   Known,             The ruling of the bears
##                                     Probable/Not       death depending on if the
##                                                        bears carcass was present
##                                                        or missing.
##
## cod         category    character   human caused       The cause of death as
##                                     (HC), natural      indicated by the scene.
##                                     (N),
##                                     undetermined
##                                     (UD)
##
## specifics   category    character                      This variable contains
##                                                        part of the loss
##                                                        variable, specifically
```

```
##                                                          why the bear died. This
##                                                          variable contains a
##                                                          string.
##
## killed/     category    character   Civilian/           This variable describes
## removed                             Hunter/ Vehicle     who, or what, the bear
## by                                  (CHV),              was Killed/Removed By
##                                     management          (krb).
##                                     (mgmt),
##                                     survival/wild
##                                     animal (SWA),
##                                     undetermined
##                                     (UD)
##
## reason      category    character   bear-human          The most negative reason
## for                                 conflict (BHC),     the bear was removed from
## removal                             food depredation    a location. Note: other
##                                     (FD), other         (OTH) contains deaths
##                                     (OTH),              from natural causes or
##                                     purposeful kill     accidents such as
##                                     (PK), safety        drowning.
##                                     concern (SC),
##                                     undetermined
##                                     (UD)
## ----------------------------------------------------------------------------------
##
## Table: Summary Table of Grizzly Bear Data
```

Data Observation Code

```r
library(readr)
grizzlybear <- read.csv("grizzlybear.csv")
set.seed(123456)
```

```r
# Code for barplot separating grizzly bears by year.
barplot(table(grizzlybear$year), main = "Grizzly Bears Separated by Year",
        sub = "4 observations contain NA values for Year.",
        ylab = "Count", xlab = "Year")
```

```r
library(tidyverse)
library(ggplot2)
library(car)
library("rnaturalearth")
library("rnaturalearthdata")
library("sf")
world <- ne_countries(scale = "medium", returnclass = "sf")
```

```
library("maps")
states <- st_as_sf(map("state", plot = FALSE, fill = TRUE))
states <- cbind(states, st_coordinates(st_centroid(states)))
```

Code for Figure 2.

```
# Code for US Map indiating grizzly bears with a blue box
ggplot(data = world) +
  geom_sf(data = states, fill = NA) +
  geom_text(data = states, aes(X, Y, label = ID), size = 3) +
  geom_rect(xmin = -117, xmax = -103.5, ymin = 41, ymax = 49, fill = NA,
            colour = "blue") +
  coord_sf(xlim = c(-125, -68), ylim = c(25, 50)) +
  geom_point(data = grizzlybear, aes(x = latitude, y = longitude), size = 1) +
  xlab("Latitude") +
  ylab("Longitude")
```

Code for Figure 3.

```
# Code for US Map subplot comparing sex and ageclass
ggplot(data = world) +
  geom_sf(data = states, fill = NA) +
  geom_text(data = states, aes(X, Y, label = ID), size = 3) +
  coord_sf(xlim = c(-117, -103.5), ylim = c(41, 49)) +
  geom_point(data = grizzlybear, aes(x = jitter(latitude), y = jitter(longitude),
                                     colour = as.factor(ageclass),
                                     shape = as.factor(sex)), size = 2) +
  xlab("Latitude") + ylab("Longitude")
```

For more information on the map code, please view these pages: part1, part 2, and part 3.

Code for Figure 4.

```
# Mosaic plot code for plot on sex and ageclass
mosaicplot(table(grizzlybear$ageclass, grizzlybear$sex), main = "", shade = TRUE)
```

Code for Figure 5.

```
# Code for US Map Subplot of killed/removed by and reason for removal
ggplot(data = world) +
  geom_sf(data = states, fill = NA) +
  geom_text(data = states, aes(X, Y, label = ID), size = 3) +
  coord_sf(xlim = c(-117, -103.5), ylim = c(41, 49)) +
  geom_point(data = grizzlybear,
             aes(x = latitude, y = longitude, colour = as.factor(rfr),
                 shape = as.factor(krb)), size = 2) +
  xlab("Latitude") +
  ylab("Longitude")
```

Code for Figure 6.

```r
# Code for barplot of cause of death separated by reason for removal
ggplot(data=grizzlybear, aes(x = cod, fill = rfr)) + geom_bar(position = "dodge")
```

Code for Figure 7.

```r
# Code for barplot of cause of death separated by sex
ggplot(data=grizzlybear, aes(x = cod, fill = sex)) + geom_bar(position = "dodge")
```

Code for table separating cause of death by ageclass and sex.

```r
# table understanding cause of death, ageclass, and sex
table(grizzlybear$cod, grizzlybear$ageclass, grizzlybear$sex)
```

Classification and Regression Tree Code

Fit

Code for Figure 8.

```r
set.seed(123456)
# Classification Tree
library(rpart)
library(rpart.plot)
# grow tree
fit <- rpart(cod ~ sex + ageclass + month + day + year + area
             + type + state + dma + body + krb + rfr, data=grizzlybear)
#plotcp(fit) # visualize cross-validation results
#summary(fit) # detailed summary of splits
# plot tree
rpart.plot(fit)
```

```r
set.seed(123456)
#fit <- rpart(cod ~ sex + ageclass + month + day + year + area
#             + type + state + dma + body + krb + rfr, data=grizzlybear)
printcp(fit) # display the results
```

```
##
## Classification tree:
## rpart(formula = cod ~ sex + ageclass + month + day + year + area +
##     type + state + dma + body + krb + rfr, data = grizzlybear)
##
## Variables actually used in tree construction:
## [1] krb rfr
##
## Root node error: 143/542 = 0.26384
##
## n=542 (2 observations deleted due to missingness)
##
##        CP nsplit rel error  xerror     xstd
```

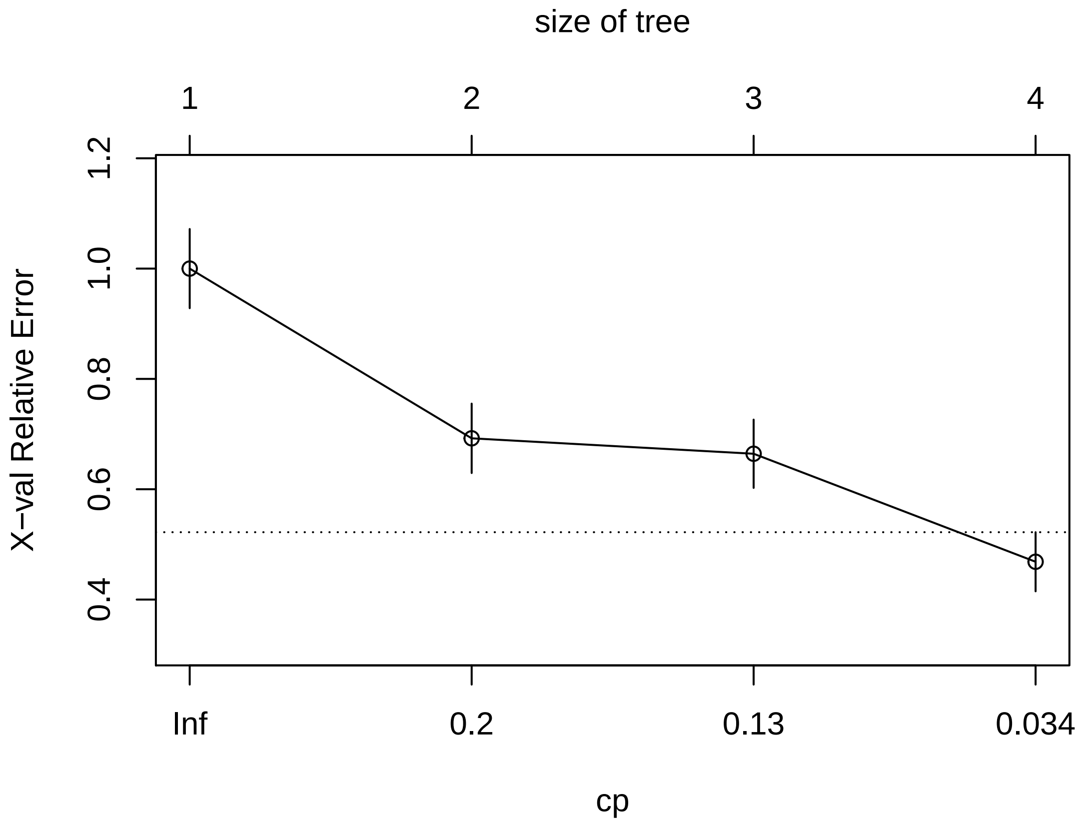
```
## 1 0.29371      0   1.00000 1.00000 0.071749
## 2 0.13287      1   0.70629 0.69231 0.062905
## 3 0.11888      2   0.57343 0.66434 0.061898
## 4 0.01000      3   0.45455 0.46853 0.053586
```

```
summary(fit) # detailed summary of splits
```

```
## Call:
## rpart(formula = cod ~ sex + ageclass + month + day + year + area +
##     type + state + dma + body + krb + rfr, data = grizzlybear)
##   n=542 (2 observations deleted due to missingness)
##
##           CP nsplit rel error    xerror       xstd
## 1 0.2937063      0 1.0000000 1.0000000 0.07174948
## 2 0.1328671      1 0.7062937 0.6923077 0.06290480
## 3 0.1188811      2 0.5734266 0.6643357 0.06189847
## 4 0.0100000      3 0.4545455 0.4685315 0.05358565
##
## Variable importance
##      krb      rfr      sex     body ageclass     type     year     area
##       33       31       11        9        8        5        2        1
##
## Node number 1: 542 observations,    complexity param=0.2937063
##   predicted class=HC  expected loss=0.2638376  P(node) =1
##     class counts:   399    68    75
##    probabilities: 0.736 0.125 0.138
##   left son=2 (409 obs) right son=3 (133 obs)
##   Primary splits:
##       rfr      splits as  LLRLLR, improve=86.79375, (49 missing)
##       krb      splits as  LLRR,   improve=86.50246, (68 missing)
##       sex      splits as  LLR,    improve=32.24659, (44 missing)
##       ageclass splits as  LRRLL,  improve=19.90285, (45 missing)
##       body     splits as  LR,     improve=19.55254, (3 missing)
##   Surrogate splits:
##       krb      splits as  LLRR,  agree=0.913, adj=0.642, (1 split)
##       sex      splits as  LLR,   agree=0.832, adj=0.308, (6 split)
##       body     splits as  LR,    agree=0.807, adj=0.208, (40 split)
##       ageclass splits as  LRLLL, agree=0.799, adj=0.175, (0 split)
##       type     splits as  LLRLL, agree=0.793, adj=0.150, (0 split)
##
## Node number 2: 409 observations
##   predicted class=HC  expected loss=0.08801956  P(node) =0.7546125
##     class counts:   373     0    36
##    probabilities: 0.912 0.000 0.088
##
## Node number 3: 133 observations,    complexity param=0.1328671
##   predicted class=N   expected loss=0.4887218  P(node) =0.2453875
##     class counts:    26    68    39
```

```
##    probabilities: 0.195 0.511 0.293
##   left son=6 (82 obs) right son=7 (51 obs)
##   Primary splits:
##       krb      splits as  RRLR,       improve=24.677190, (32 missing)
##       month    splits as  LRR,        improve= 5.170890, (20 missing)
##       day      < 6.5    to the right, improve= 4.446322, (26 missing)
##       ageclass splits as  RLLLR,      improve= 4.123147, (8 missing)
##       dma      splits as  RL,         improve= 3.215216, (66 missing)
##   Surrogate splits:
##       rfr      splits as  --L--R,     agree=0.624, adj=0.156, (20 split)
##       sex      splits as  RLL,        agree=0.594, adj=0.089, (5 split)
##       year     < 2011.5 to the right, agree=0.594, adj=0.089, (7 split)
##       ageclass splits as  LLLLR,      agree=0.574, adj=0.044, (0 split)
##       type     splits as  RLLLR,      agree=0.574, adj=0.044, (0 split)
##
## Node number 6: 82 observations
##   predicted class=N   expected loss=0.2439024  P(node) =0.1512915
##     class counts:     6    62    14
##    probabilities: 0.073 0.756 0.171
##
## Node number 7: 51 observations,    complexity param=0.1188811
##   predicted class=UD  expected loss=0.5098039  P(node) =0.09409594
##     class counts:    20     6    25
##    probabilities: 0.392 0.118 0.490
##   left son=14 (27 obs) right son=15 (24 obs)
##   Primary splits:
##       krb      splits as  LL-R,       improve=17.776110, (6 missing)
##       ageclass splits as  RLRRR,      improve= 4.826667, (1 missing)
##       day      < 16.5   to the right, improve= 3.451754, (13 missing)
##       body     splits as  RL,         improve= 3.398099, (0 missing)
##       area     splits as  R--RL,      improve= 2.733761, (6 missing)
##   Surrogate splits:
##       ageclass splits as  RLLRR,      agree=0.733, adj=0.455, (6 split)
##       body     splits as  RL,         agree=0.711, adj=0.409, (0 split)
##       area     splits as  R--RL,      agree=0.622, adj=0.227, (0 split)
##       year     < 2012.5 to the left,  agree=0.600, adj=0.182, (0 split)
##       sex      splits as  RRL,        agree=0.578, adj=0.136, (0 split)
##
## Node number 14: 27 observations
##   predicted class=HC  expected loss=0.2592593  P(node) =0.0498155
##     class counts:    20     4     3
##    probabilities: 0.741 0.148 0.111
##
## Node number 15: 24 observations
##   predicted class=UD  expected loss=0.08333333  P(node) =0.04428044
##     class counts:     0     2    22
##    probabilities: 0.000 0.083 0.917
```

```
plotcp(fit)
```



Fit 2

Code for Figure 9.

```
set.seed(123456)
fit2 <- rpart(cod ~ sex + ageclass + month + day + year + area
              + type + state + dma + body, method = "class", data=grizzlybear)
#
#plotcp(fit2) # visualize cross-validation results
#summary(fit2) # detailed summary of splits
#
rpart.plot(fit2)
```

```
set.seed(123456)
#fit2 <- rpart(cod ~ sex + ageclass + month + day + year + area
#              + type + state + dma + body, method = "class", data=grizzlybear)
#
printcp(fit2) # display the results
```

```
##
## Classification tree:
```

```
## rpart(formula = cod ~ sex + ageclass + month + day + year + area +
##     type + state + dma + body, data = grizzlybear, method = "class")
##
## Variables actually used in tree construction:
## [1] ageclass body     month    sex      type
##
## Root node error: 143/542 = 0.26384
##
## n=542 (2 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.122378      0   1.00000 1.00000 0.071749
## 2 0.017483      2   0.75524 0.76224 0.065256
## 3 0.013986      4   0.72028 0.79720 0.066350
## 4 0.010000      5   0.70629 0.79021 0.066135
```

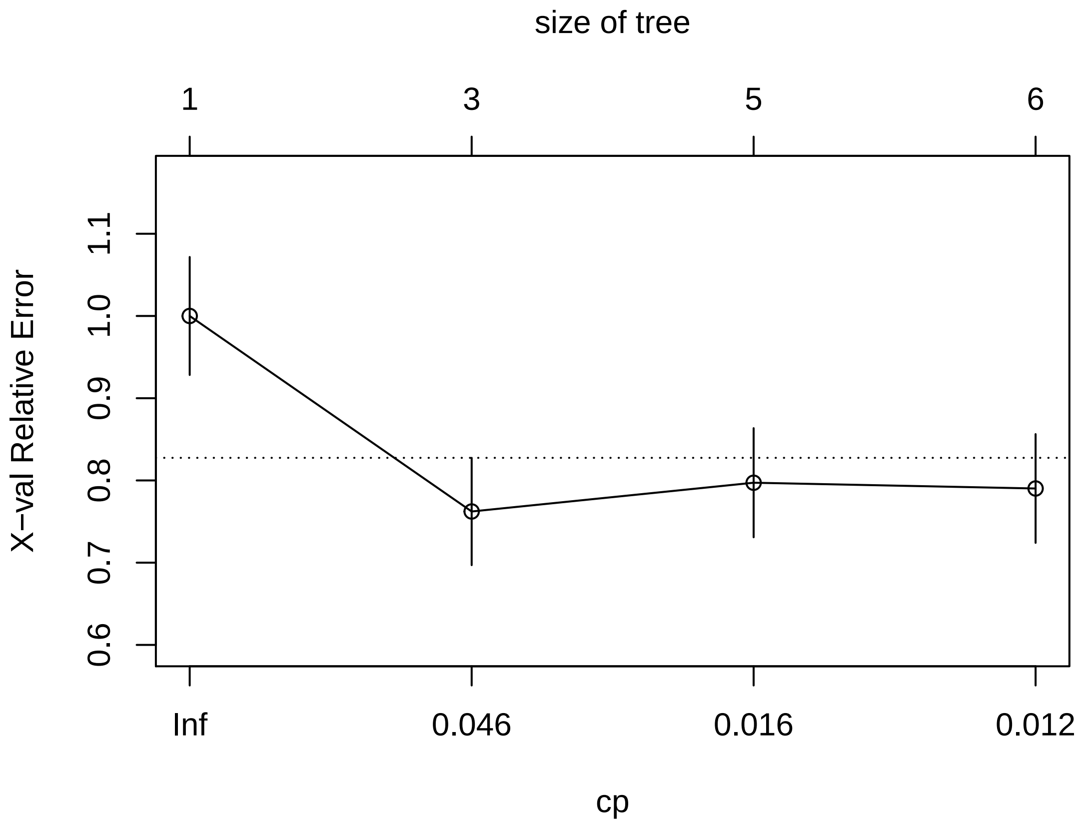
```r
summary(fit2) # detailed summary of splits
```

```
## Call:
## rpart(formula = cod ~ sex + ageclass + month + day + year + area +
##     type + state + dma + body, data = grizzlybear, method = "class")
##   n=542 (2 observations deleted due to missingness)
##
##           CP nsplit rel error    xerror       xstd
## 1 0.12237762      0 1.0000000 1.0000000 0.07174948
## 2 0.01748252      2 0.7552448 0.7622378 0.06525616
## 3 0.01398601      4 0.7202797 0.7972028 0.06634965
## 4 0.01000000      5 0.7062937 0.7902098 0.06613513
##
## Variable importance
##      sex     body ageclass    month     type      day     year
##       25       22       19       18       11        3        2
##
## Node number 1: 542 observations,    complexity param=0.1223776
##   predicted class=HC   expected loss=0.2638376  P(node) =1
##     class counts:   399    68    75
##    probabilities: 0.736 0.125 0.138
##   left son=2 (451 obs) right son=3 (91 obs)
##   Primary splits:
##       sex      splits as  LLR,   improve=32.246590, (44 missing)
##       ageclass splits as  LRRLL, improve=19.902850, (45 missing)
##       body     splits as  LR,    improve=19.552540, (3 missing)
##       type     splits as  LLRLL, improve=16.968620, (53 missing)
##       month    splits as  LRL,   improve= 6.385425, (57 missing)
##   Surrogate splits:
##       body     splits as  LR,    agree=0.950, adj=0.699, (42 split)
##       ageclass splits as  LRRLL, agree=0.918, adj=0.506, (0 split)
##
```

```
## Node number 2: 451 observations,    complexity param=0.01748252
##   predicted class=HC  expected loss=0.1840355  P(node) =0.8321033
##     class counts:   368    27    56
##    probabilities: 0.816 0.060 0.124
##   left son=4 (413 obs) right son=5 (38 obs)
##   Primary splits:
##       type  splits as  LLRLL,      improve=10.200490, (43 missing)
##       year  < 2014.5 to the left,  improve= 2.209477, (2 missing)
##       body  splits as  LR,         improve= 1.508574, (3 missing)
##       month splits as  LRL,        improve= 1.375533, (45 missing)
##       dma   splits as  RL,         improve= 1.324726, (172 missing)
##
## Node number 3: 91 observations,    complexity param=0.1223776
##   predicted class=N   expected loss=0.5494505  P(node) =0.1678967
##     class counts:    31    41    19
##    probabilities: 0.341 0.451 0.209
##   left son=6 (45 obs) right son=7 (46 obs)
##   Primary splits:
##       month    splits as  LRR,        improve=22.899630, (12 missing)
##       type     splits as  RLRLR,      improve= 4.919416, (10 missing)
##       day      < 21.5   to the right, improve= 3.720000, (16 missing)
##       ageclass splits as  RLLRR,      improve= 2.988566, (9 missing)
##       year     < 2012.5 to the left,  improve= 1.707692, (0 missing)
##   Surrogate splits:
##       type     splits as  RLRLL,      agree=0.608, adj=0.205, (4 split)
##       ageclass splits as  RRLR-,      agree=0.570, adj=0.128, (0 split)
##       day      < 5.5    to the right, agree=0.570, adj=0.128, (0 split)
##       year     < 2017.5 to the left,  agree=0.570, adj=0.128, (8 split)
##       body     splits as  RL,         agree=0.570, adj=0.128, (0 split)
##
## Node number 4: 413 observations
##   predicted class=HC  expected loss=0.1549637  P(node) =0.7619926
##     class counts:   349    14    50
##    probabilities: 0.845 0.034 0.121
##
## Node number 5: 38 observations,    complexity param=0.01748252
##   predicted class=HC  expected loss=0.5  P(node) =0.0701107
##     class counts:    19    13     6
##    probabilities: 0.500 0.342 0.158
##   left son=10 (18 obs) right son=11 (20 obs)
##   Primary splits:
##       ageclass splits as  RL-LL,      improve=4.0497080, (0 missing)
##       sex      splits as  RL-,        improve=2.0270020, (0 missing)
##       year     < 2016.5 to the left,  improve=1.7872910, (0 missing)
##       month    splits as  RRL,        improve=1.6592110, (3 missing)
##       day      < 23.5   to the right, improve=0.6740741, (3 missing)
##   Surrogate splits:
##       sex   splits as  RL-,        agree=0.658, adj=0.278, (0 split)
```

```
##       month splits as  RRL,        agree=0.579, adj=0.111, (0 split)
##       day   < 10.5   to the left,  agree=0.579, adj=0.111, (0 split)
##       year  < 2010.5 to the left,  agree=0.553, adj=0.056, (0 split)
##       area  splits as  RL--R,      agree=0.553, adj=0.056, (0 split)
##
## Node number 6: 45 observations,    complexity param=0.01398601
##   predicted class=HC  expected loss=0.3111111  P(node) =0.08302583
##     class counts:    31     6     8
##    probabilities: 0.689 0.133 0.178
##   left son=12 (37 obs) right son=13 (8 obs)
##   Primary splits:
##       body     splits as  RL,          improve=3.0168170, (0 missing)
##       ageclass splits as  RLRLR,      improve=2.1335950, (2 missing)
##       day      < 7.5    to the right, improve=1.9881880, (8 missing)
##       type     splits as  -RRLR,      improve=1.2100840, (3 missing)
##       year     < 2012.5 to the left,  improve=0.8767677, (0 missing)
##   Surrogate splits:
##       ageclass splits as  RLLRR, agree=0.889, adj=0.375, (0 split)
##
## Node number 7: 46 observations
##   predicted class=N   expected loss=0.2391304  P(node) =0.08487085
##     class counts:     0    35    11
##    probabilities: 0.000 0.761 0.239
##
## Node number 10: 18 observations
##   predicted class=HC  expected loss=0.2222222  P(node) =0.03321033
##     class counts:    14     3     1
##    probabilities: 0.778 0.167 0.056
##
## Node number 11: 20 observations
##   predicted class=N   expected loss=0.5  P(node) =0.03690037
##     class counts:     5    10     5
##    probabilities: 0.250 0.500 0.250
##
## Node number 12: 37 observations
##   predicted class=HC  expected loss=0.2162162  P(node) =0.06826568
##     class counts:    29     4     4
##    probabilities: 0.784 0.108 0.108
##
## Node number 13: 8 observations
##   predicted class=UD  expected loss=0.5  P(node) =0.01476015
##     class counts:     2     2     4
##    probabilities: 0.250 0.250 0.500
```

```r
plotcp(fit2)
```



Fit 3

Code for Figure 10.

```
set.seed(123456)
fit3 <- rpart(cod ~ ageclass + day + area + state + dma
              + body, method = "class", data=grizzlybear)
#
#plotcp(fit3) # visualize cross-validation results
#summary(fit3) # detailed summary of splits
#
rpart.plot(fit3)
```

```
set.seed(123456)
#fit3 <- rpart(cod ~ ageclass + day + area + state + dma
#              + body, method = "class", data=grizzlybear)
#
printcp(fit3) # display the results
```

```
##
## Classification tree:
## rpart(formula = cod ~ ageclass + day + area + state + dma + body,
##     data = grizzlybear, method = "class")
```

```
##
## Variables actually used in tree construction:
## [1] ageclass body     day      state
##
## Root node error: 143/542 = 0.26384
##
## n=542 (2 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.038462      0   1.00000 1.00000 0.071749
## 2 0.013986      2   0.92308 0.96503 0.070924
## 3 0.010000      6   0.84615 0.93706 0.070234
```

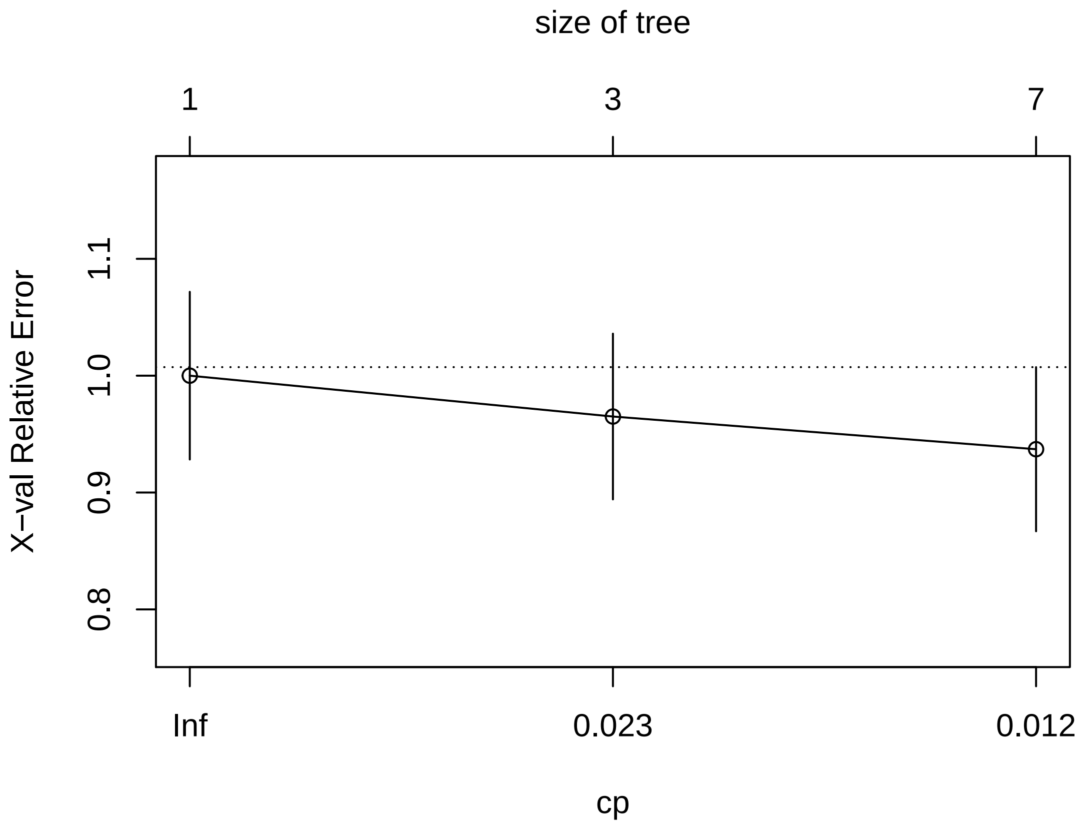
```r
summary(fit3) # detailed summary of splits
```

```
## Call:
## rpart(formula = cod ~ ageclass + day + area + state + dma + body,
##     data = grizzlybear, method = "class")
##   n=542 (2 observations deleted due to missingness)
##
##           CP nsplit rel error    xerror       xstd
## 1 0.03846154      0 1.0000000 1.0000000 0.07174948
## 2 0.01398601      2 0.9230769 0.9650350 0.07092422
## 3 0.01000000      6 0.8461538 0.9370629 0.07023390
##
## Variable importance
## ageclass     body      day    state     area
##       41       32       20        5        2
##
## Node number 1: 542 observations,    complexity param=0.03846154
##   predicted class=HC  expected loss=0.2638376  P(node) =1
##     class counts:   399    68    75
##    probabilities: 0.736 0.125 0.138
##   left son=2 (429 obs) right son=3 (113 obs)
##   Primary splits:
##       ageclass splits as  LRRLL,    improve=19.902850, (45 missing)
##       body     splits as  LR,       improve=19.552540, (3 missing)
##       dma      splits as  RL,       improve= 4.735420, (206 missing)
##       day      < 2.5  to the right, improve= 2.904946, (64 missing)
##       state    splits as  RLL,      improve= 1.841624, (62 missing)
##   Surrogate splits:
##       body splits as  LR, agree=0.926, adj=0.641, (43 split)
##
## Node number 2: 429 observations
##   predicted class=HC  expected loss=0.2004662  P(node) =0.7915129
##     class counts:   343    26    60
##    probabilities: 0.800 0.061 0.140
##
```

```
## Node number 3: 113 observations,    complexity param=0.03846154
##   predicted class=HC  expected loss=0.5044248  P(node) =0.2084871
##     class counts:    56    42    15
##    probabilities: 0.496 0.372 0.133
##   left son=6 (92 obs) right son=7 (21 obs)
##   Primary splits:
##       day   < 5.5  to the right, improve=5.703392, (11 missing)
##       body  splits as  LR,       improve=4.397321, (0 missing)
##       dma   splits as  RL,       improve=3.461905, (43 missing)
##       area  splits as  L-RRR,    improve=1.429801, (10 missing)
##       state splits as  RRL,      improve=1.379904, (20 missing)
##
## Node number 6: 92 observations,    complexity param=0.01398601
##   predicted class=HC  expected loss=0.4456522  P(node) =0.1697417
##     class counts:    51    26    15
##    probabilities: 0.554 0.283 0.163
##   left son=12 (31 obs) right son=13 (61 obs)
##   Primary splits:
##       body  splits as  LR,       improve=3.4017310, (0 missing)
##       dma   splits as  RL,       improve=3.3800000, (32 missing)
##       state splits as  RRL,      improve=2.0287630, (15 missing)
##       day   < 23.5 to the right, improve=1.7991290, (11 missing)
##       area  splits as  L-LRR,    improve=0.7497811, (10 missing)
##
## Node number 7: 21 observations
##   predicted class=N   expected loss=0.2380952  P(node) =0.03874539
##     class counts:     5    16     0
##    probabilities: 0.238 0.762 0.000
##
## Node number 12: 31 observations
##   predicted class=HC  expected loss=0.2258065  P(node) =0.05719557
##     class counts:    24     5     2
##    probabilities: 0.774 0.161 0.065
##
## Node number 13: 61 observations,    complexity param=0.01398601
##   predicted class=HC  expected loss=0.557377  P(node) =0.1125461
##     class counts:    27    21    13
##    probabilities: 0.443 0.344 0.213
##   left son=26 (26 obs) right son=27 (35 obs)
##   Primary splits:
##       day      < 21.5 to the right, improve=2.7194990, (10 missing)
##       state    splits as  LLR,      improve=1.8084850, (6 missing)
##       area     splits as  L-LRR,    improve=1.4176470, (10 missing)
##       ageclass splits as  -LR--,    improve=0.2418301, (10 missing)
##   Surrogate splits:
##       ageclass splits as  -RL--, agree=0.667, adj=0.292, (2 split)
##       area     splits as  R-LRL, agree=0.627, adj=0.208, (0 split)
##
```

```
## Node number 26: 26 observations
##   predicted class=HC  expected loss=0.3461538  P(node) =0.04797048
##     class counts:    17     7     2
##    probabilities: 0.654 0.269 0.077
##
## Node number 27: 35 observations,    complexity param=0.01398601
##   predicted class=N   expected loss=0.6  P(node) =0.06457565
##     class counts:    10    14    11
##    probabilities: 0.286 0.400 0.314
##   left son=54 (14 obs) right son=55 (21 obs)
##   Primary splits:
##       state    splits as  LLR,      improve=2.379762, (5 missing)
##       area     splits as  L--RR,    improve=1.814815, (8 missing)
##       day      < 17.5 to the left,  improve=1.633862, (8 missing)
##       ageclass splits as  -LR--,    improve=1.369578, (9 missing)
##
## Node number 54: 14 observations
##   predicted class=N   expected loss=0.5  P(node) =0.02583026
##     class counts:     5     7     2
##    probabilities: 0.357 0.500 0.143
##
## Node number 55: 21 observations,    complexity param=0.01398601
##   predicted class=UD  expected loss=0.5714286  P(node) =0.03874539
##     class counts:     5     7     9
##    probabilities: 0.238 0.333 0.429
##   left son=110 (7 obs) right son=111 (14 obs)
##   Primary splits:
##       day < 16.5 to the right, improve=1.740079, (5 missing)
##   Surrogate splits:
##       area splits as  L--RR, agree=0.625, adj=0.143, (0 split)
##
## Node number 110: 7 observations
##   predicted class=N   expected loss=0.2857143  P(node) =0.01291513
##     class counts:     2     5     0
##    probabilities: 0.286 0.714 0.000
##
## Node number 111: 14 observations
##   predicted class=UD  expected loss=0.3571429  P(node) =0.02583026
##     class counts:     3     2     9
##    probabilities: 0.214 0.143 0.643
```

```
plotcp(fit3)
```



Fit 4

Code for Figure 11.

```
# Fit 4 code for
set.seed(123456)
fit4 <- rpart(cod ~ krb + sex + ageclass + month + day + year + area
              + type + state + dma + body, method = "class", data=grizzlybear)
#
#plotcp(fit4) # visualize cross-validation results
#summary(fit4) # detailed summary of splits
#
rpart.plot(fit4)
```

```
set.seed(123456)
#fit4 <- rpart(cod ~ krb + sex + ageclass + month + day + year + area
#              + type + state + dma + body, method = "class", data=grizzlybear)
#
printcp(fit4) # display the results
```

```
##
## Classification tree:
## rpart(formula = cod ~ krb + sex + ageclass + month + day + year +
```

```
##     area + type + state + dma + body, data = grizzlybear, method = "class")
##
## Variables actually used in tree construction:
## [1] ageclass krb      month
##
## Root node error: 143/542 = 0.26384
##
## n=542 (2 observations deleted due to missingness)
##
##         CP nsplit rel error  xerror     xstd
## 1 0.328671      0   1.00000 1.00000 0.071749
## 2 0.132867      1   0.67133 0.67133 0.062154
## 3 0.031469      2   0.53846 0.53846 0.056838
## 4 0.010000      4   0.47552 0.46853 0.053586
```

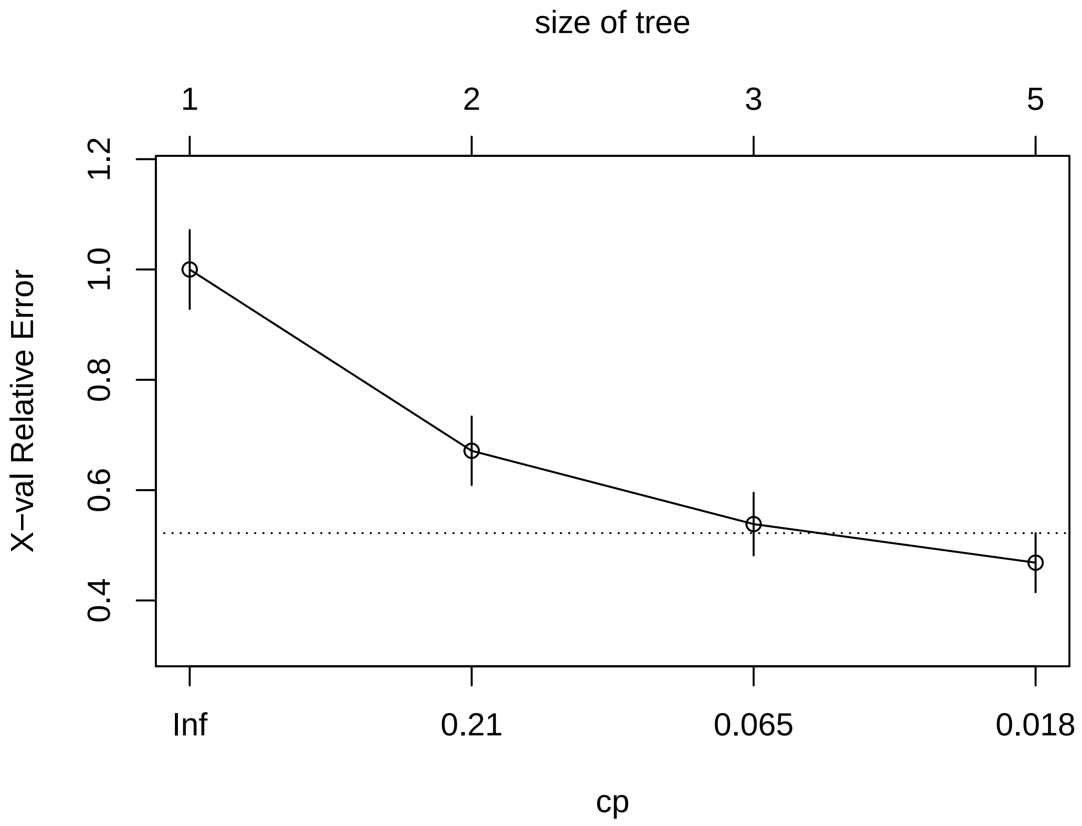
```r
summary(fit4) # detailed summary of splits
```

```
## Call:
## rpart(formula = cod ~ krb + sex + ageclass + month + day + year +
##     area + type + state + dma + body, data = grizzlybear, method = "class")
##   n=542 (2 observations deleted due to missingness)
##
##           CP nsplit rel error    xerror       xstd
## 1 0.32867133      0 1.0000000 1.0000000 0.07174948
## 2 0.13286713      1 0.6713287 0.6713287 0.06215375
## 3 0.03146853      2 0.5384615 0.5384615 0.05683766
## 4 0.01000000      4 0.4755245 0.4685315 0.05358565
##
## Variable importance
##      krb ageclass    month     type     year      sex      day    state
##       72        7        6        6        3        2        1        1
##     body
##        1
##
## Node number 1: 542 observations,    complexity param=0.3286713
##   predicted class=HC  expected loss=0.2638376  P(node) =1
##     class counts:   399    68    75
##    probabilities: 0.736 0.125 0.138
##   left son=2 (457 obs) right son=3 (85 obs)
##   Primary splits:
##       krb      splits as  LLRR,  improve=86.50246, (68 missing)
##       sex      splits as  LLR,   improve=32.24659, (44 missing)
##       ageclass splits as  LRRLL, improve=19.90285, (45 missing)
##       body     splits as  LR,    improve=19.55254, (3 missing)
##       type     splits as  LLRLL, improve=16.96862, (53 missing)
##   Surrogate splits:
##       type splits as  LLRLL, agree=0.844, adj=0.075, (21 split)
##       sex  splits as  LLR,   agree=0.833, adj=0.013, (5 split)
```

```
##
## Node number 2: 457 observations,    complexity param=0.03146853
##   predicted class=HC  expected loss=0.1422319  P(node) =0.8431734
##     class counts:   392    14    51
##    probabilities: 0.858 0.031 0.112
##   left son=4 (416 obs) right son=5 (41 obs)
##   Primary splits:
##       ageclass splits as  LLRLL,      improve=9.146321, (45 missing)
##       year     < 2015.5 to the left,  improve=7.817965, (0 missing)
##       body     splits as  LR,         improve=7.424024, (3 missing)
##       sex      splits as  LLR,        improve=7.202653, (44 missing)
##       dma      splits as  RL,         improve=2.201306, (155 missing)
##   Surrogate splits:
##       sex  splits as  LLR, agree=0.944, adj=0.281, (2 split)
##       body splits as  LR,  agree=0.939, adj=0.219, (41 split)
##
## Node number 3: 85 observations,    complexity param=0.1328671
##   predicted class=N   expected loss=0.3647059  P(node) =0.1568266
##     class counts:     7    54    24
##    probabilities: 0.082 0.635 0.282
##   left son=6 (61 obs) right son=7 (24 obs)
##   Primary splits:
##       krb      splits as  --LR,       improve=22.944050, (5 missing)
##       ageclass splits as  RLRLR,      improve= 4.664543, (0 missing)
##       day      < 7.5    to the left,  improve= 2.972542, (14 missing)
##       area     splits as  R--LR,      improve= 2.803140, (2 missing)
##       year     < 2014.5 to the left,  improve= 2.159438, (2 missing)
##   Surrogate splits:
##       ageclass splits as  LLLLR,      agree=0.713, adj=0.042, (5 split)
##       year     < 2010.5 to the right, agree=0.713, adj=0.042, (0 split)
##       type     splits as  RLLL-,      agree=0.713, adj=0.042, (0 split)
##
## Node number 4: 416 observations
##   predicted class=HC  expected loss=0.1057692  P(node) =0.7675277
##     class counts:   372     2    42
##    probabilities: 0.894 0.005 0.101
##
## Node number 5: 41 observations,    complexity param=0.03146853
##   predicted class=HC  expected loss=0.5121951  P(node) =0.07564576
##     class counts:    20    12     9
##    probabilities: 0.488 0.293 0.220
##   left son=10 (28 obs) right son=11 (13 obs)
##   Primary splits:
##       month splits as  LRR,        improve=8.891142, (8 missing)
##       type  splits as  RR-LR,      improve=3.335758, (8 missing)
##       day   < 20     to the right, improve=1.778765, (8 missing)
##       year  < 2018.5 to the right, improve=1.374153, (0 missing)
##       state splits as  LLR,        improve=1.013240, (0 missing)
```

```
##   Surrogate splits:
##       year  < 2018.5 to the left,  agree=0.758, adj=0.385, (8 split)
##       day   < 27     to the left,  agree=0.697, adj=0.231, (0 split)
##       state splits as  RLL,        agree=0.697, adj=0.231, (0 split)
##       type  splits as  RL-LR,      agree=0.667, adj=0.154, (0 split)
##
## Node number 6: 61 observations
##   predicted class=N   expected loss=0.147541  P(node) =0.1125461
##     class counts:     6    52     3
##    probabilities: 0.098 0.852 0.049
##
## Node number 7: 24 observations
##   predicted class=UD  expected loss=0.125  P(node) =0.04428044
##     class counts:     1     2    21
##    probabilities: 0.042 0.083 0.875
##
## Node number 10: 28 observations
##   predicted class=HC  expected loss=0.3571429  P(node) =0.05166052
##     class counts:    18     1     9
##    probabilities: 0.643 0.036 0.321
##
## Node number 11: 13 observations
##   predicted class=N   expected loss=0.1538462  P(node) =0.02398524
##     class counts:     2    11     0
##    probabilities: 0.154 0.846 0.000
```

```r
plotcp(fit4)
```



Correspondence Analysis Code

```
library(FactoMineR)
library(ggplot2)
```

```
newbear = grizzlybear[,c("sex", "ageclass", "month", "year", "cod", "krb", "rfr")]
newbear$sex <- as.factor(newbear$sex)
newbear$ageclass <- as.factor(newbear$ageclass)
newbear$month <- as.factor(newbear$month)
newbear$year <- as.factor(newbear$year)
newbear$cod <- as.factor(newbear$cod)
newbear$krb <- as.factor(newbear$krb)
newbear$rfr <- as.factor(newbear$rfr)
newbearNA<- na.omit(newbear)
cats = apply(newbearNA, 2, function(x) nlevels(as.factor(x)))
mcabear = MCA(newbearNA, graph=FALSE)
```

```
mcabear_vars_df = data.frame(mcabear$var$coord, Variable=rep(names(cats), cats))
mcabear_obs_df = data.frame(mcabear$ind$coord)
```

```
# eigen values for the correspondence analysis
mcabear$eig
```

```
##        eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.49860555             12.4651388                          12.46514
## dim 2  0.30889535              7.7223838                          20.18752
## dim 3  0.29491062              7.3727655                          27.56029
## dim 4  0.21799178              5.4497945                          33.01008
## dim 5  0.20041616              5.0104041                          38.02049
## dim 6  0.19054224              4.7635559                          42.78404
## dim 7  0.17138271              4.2845676                          47.06861
## dim 8  0.16664474              4.1661184                          51.23473
## dim 9  0.16311196              4.0777991                          55.31253
## dim 10 0.15691621              3.9229054                          59.23543
## dim 11 0.14763281              3.6908204                          62.92625
## dim 12 0.14474267              3.6185667                          66.54482
## dim 13 0.14130100              3.5325249                          70.07735
## dim 14 0.13710610              3.4276526                          73.50500
## dim 15 0.13099952              3.2749879                          76.77999
## dim 16 0.12963977              3.2409942                          80.02098
## dim 17 0.11600773              2.9001932                          82.92117
## dim 18 0.10308269              2.5770672                          85.49824
## dim 19 0.10163375              2.5408438                          88.03908
## dim 20 0.09432732              2.3581830                          90.39727
## dim 21 0.08676133              2.1690332                          92.56630
## dim 22 0.08469212              2.1173030                          94.68360
## dim 23 0.07323678              1.8309195                          96.51452
## dim 24 0.04923839              1.2309596                          97.74548
## dim 25 0.03415384              0.8538459                          98.59933
## dim 26 0.02668073              0.6670181                          99.26635
## dim 27 0.01924241              0.4810602                          99.74741
## dim 28 0.01010374              0.2525934                         100.00000
```

```
# coordinates for the variables and their levels
mcabear$var$coord
```

```
##                Dim 1       Dim 2       Dim 3        Dim 4       Dim 5
## F        -0.11030457  0.03504407  0.26424884 -0.018934782 -0.49397378
## M        -0.32903224  0.15455228 -0.19962465 -0.143150144  0.11276008
## Unk       1.77513259 -0.79680615  0.34611729  0.708175706  0.56523140
## Adult    -0.25330105  0.15845167 -0.09760162 -0.089844212 -0.30666297
## COY       1.61710735 -0.40145047 -0.11102085 -0.038723065 -0.33441971
## Cub       0.18672076 -1.81962741  0.84401967  2.636369808  2.05422617
## Subadult -0.34099713  0.13084576  0.14560409 -0.543954041  0.82224691
## Yearling -0.25629217  0.32339666  0.08633688  1.757942034 -1.25893650
## Fall     -0.09165216 -0.47852831  0.32238040  0.119107167 -0.23666503
## Spring    0.41908987  0.21544802 -0.26819190 -0.665897857  0.61351118
## Summer   -0.04805385  0.57714156 -0.33716618  0.113350830  0.07263912
## 2010     -0.09508638  0.27689292  0.65093171 -0.119388560 -1.06744459
## 2011      0.09375024 -0.09141473 -0.03377963  0.057707193 -1.12755241
## 2012      1.08457654 -0.33400261 -0.40945147 -0.384653649 -0.11901810
```

```
## 2013     -0.06430053  0.22480690  0.20274564 -0.814631316  0.76142816
## 2014      0.37532334 -0.26416535 -0.35633134 -0.548699580 -0.49415173
## 2015      0.10736950  0.04940737  0.38953996 -0.073673299 -0.46152689
## 2016     -0.22270853 -0.01918006 -0.24721994 -0.906471936  0.31682979
## 2017     -0.10898513 -0.92647209  0.54809223  0.901665970  1.18433703
## 2018     -0.35006474  0.17938040 -0.61131957  1.414958645 -0.30308011
## 2019     -0.35110751  0.33836786 -0.12724299 -0.028531127  1.31112788
## 2020     -0.57608773  0.66988960 -0.06551519 -0.359571198  0.78323709
## cod_HC   -0.34941815 -0.12806062 -0.01272249 -0.004099897 -0.01161930
## cod_N     2.36953235  0.08151624 -0.96379149 -0.132655995  0.06043273
## cod_UD    1.39098864  3.08036442  3.48086649  0.540486878  0.10623758
## krb_CHV  -0.12295296 -0.92797976  0.64821311 -0.308113075 -0.06866360
## krb_mgmt -0.52108213  0.39274627 -0.43797747  0.213163663  0.03128189
## krb_SWA   2.24414426  0.02641324 -1.01141724 -0.158779960  0.04474708
## krb_UD    1.33795291  3.12494377  3.48931535  0.387351129  0.05711998
## rfr_BHC  -0.31690866 -0.65437128  0.42113453 -0.230173048 -0.04561685
## rfr_FD   -0.58806910  0.67876874 -0.73623566  0.198968460  0.32561355
## rfr_OTH   1.85781371  0.34582890 -0.12429415  0.023647647 -0.10961181
## rfr_PK   -0.47489815 -0.04559279  0.51196428 -1.407388485 -0.52685920
## rfr_SC   -0.60847852  0.29379021 -0.46154564  2.152174216 -0.88845790
## rfr_UD    0.90156657  3.50582396  4.16757598 -0.135203554  1.91198298
```

Code for Figure 12.

```r
# Multiple Correpsondence Analysis Plot
ggplot(data=mcabear_obs_df, aes(x=Dim.1, y=Dim.2)) +
  geom_hline(yintercept=0, colour="gray60") +
  geom_vline(xintercept=0, colour="gray60") +
  geom_rect(xmin = -0.75, xmax = 0.25, ymin = -1.3, ymax = 1, fill = NA,
            colour = "red") +
  geom_point(colour="gray60", alpha=0.7) +
  geom_density2d(colour="gray70") +
  geom_text(data=mcabear_vars_df,
            aes(x=Dim.1, y=Dim.2, label=rownames(mcabear_vars_df),
                colour=Variable)) +
  ggtitle("MCA Plot of the Grizzly Bear Data") +
  scale_colour_discrete(name="Variable")
```

Code for Figure 13.

```r
# Multiple Correspondence Analysis Subplot
ggplot(data=mcabear_obs_df, aes(x=Dim.1, y=Dim.2)) +
  geom_hline(yintercept=0, colour="gray60") +
  geom_vline(xintercept=0, colour="gray60") +
  geom_point(colour="gray70", alpha=0.7) +
  geom_density2d(colour="gray70") +
  geom_text(data=mcabear_vars_df,
            aes(x=Dim.1, y=Dim.2, label=rownames(mcabear_vars_df),
                colour=Variable)) +
```

```
  ggtitle("MCA Subplot of the Grizzly Bear Data") +
  scale_colour_discrete(name="Variable") +
  coord_cartesian(xlim = c(-0.75, 0.25), ylim = c(-1.3, 1))
```

References

Emmery C (2018). “Full (CaRT)Decision Tree Example.” URL https://cmry.github.io/notes/trees.

Everitt B, Dunn G (2001). *Applied Multivariate Data Analysis*. 2 edition. ISBN 9780470711170.

Hughes C, Foote L, Yarmey NT, Hwang C, Thorlakson J, Nielsen S (2020). “From human invaders to problem bears: A media content analysis of grizzly bear conservation.” *Conservation Science and Practice*, **2**(4). ISSN 2578-4854, 2578-4854. doi:10.1111/csp2.176. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/csp2.176.

Interagency Grizzly Bear Committee (IGBC) (2021). “Yellowstone Ecosystem.” URL http://igbconline.org/yellowstone/.

Interagency Grizzly Bear Study Team (2021). “Interagency Grizzly Bear Study Team - Overview.” URL https://www.usgs.gov/science/interagency-grizzly-bear-study-team?qt-science_center_objects=0#qt-science_center_objects.

James G, Witten D, Hastie T, Tibshirani R (2013). *An Introduction to Statistical Learning*, volume 103 of *Springer Texts in Statistics*. Springer New York, 1 edition. ISBN 9781461471370. doi:10.1007/978-1-4614-7138-7. URL http://link.springer.com/10.1007/978-1-4614-7138-7.

Loh W (2011). “Classification and regression trees.” *WIREs Data Mining and Knowledge Discovery*, **1**(1), 14–23. ISSN 1942-4787, 1942-4795. doi:10.1002/widm.8.

Mattson DJ, Reid MM (1991). “Conservation of the Yellowstone Grizzly Bear.” *Conservation Biology*, **5**(3), 364–372. ISSN 0888-8892.

McLellan BN, Hovey FW, Mace RD, Woods JG, Carney DW, Gibeau ML, Wakkinen WL, Kasworm WF (1999). “Rates and Causes of Grizzly Bear Mortality in the Interior Mountains of British Columbia, Alberta, Montana, Washington, and Idaho.” *The Journal of Wildlife Management*, **63**(3), 911–920. ISSN 0022-541X. doi:10.2307/3802805.

Patel N (2003). *Lecture 3 Classification Trees*. MIT Open Courseware, Wallingford, CT.

Servheen C (1993). “Grizzly Bear Recovery Plan.” URL https://www.fws.gov/mountain-prairie/es/species/mammals/grizzly/Grizzly_bear_recovery_plan.pdf.

The National Wildlife Federation (????). “Grizzly Bear.” URL https://www.nwf.org/Home/Educational-Resources/Wildlife-Guide/Mammals/Grizzly-Bear.

US Fish & Wildlife Service (2020). “Getting started with mixed effect models in r.” URL https://www.fws.gov/mountain-prairie/es/grizzlybear.php.

US Geology Survey (2010). URL https://www.usgs.gov/data-tools/2010-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2011). URL https://www.usgs.gov/data-tools/2011-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2012). URL https://www.usgs.gov/data-tools/2012-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2013). URL https://www.usgs.gov/data-tools/2013-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2014). URL https://www.usgs.gov/data-tools/2014-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2015). URL https://www.usgs.gov/data-tools/2015-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2016). URL https://www.usgs.gov/data-tools/2016-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2017). URL https://www.usgs.gov/data-tools/2017-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2018). URL https://www.usgs.gov/data-tools/20218-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2019). URL https://www.usgs.gov/data-tools/2019-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.

US Geology Survey (2020). URL https://www.usgs.gov/data-tools/2020-known-and-probable-grizzly-bear-mortalities-greater-yellowstone-ecosystem.