CKME 136 Capstone Forest Fires

1 Load packages.

#Make sure to use libraries below  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

#install.packages("ggplot2")  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

#install.packages("ggpubr")  
library(ggpubr)

## Loading required package: magrittr

#install.packages("tidyr")  
library(tidyr)

## Warning: package 'tidyr' was built under R version 3.5.3

##   
## Attaching package: 'tidyr'

## The following object is masked from 'package:magrittr':  
##   
## extract

#install.packages("scales")  
library(scales)

## Warning: package 'scales' was built under R version 3.5.3

#install.packages("wesanderson")  
#library(wesanderson)  
#install.packages("viridis") # Install  
library("viridis") # Load

## Warning: package 'viridis' was built under R version 3.5.3

## Loading required package: viridisLite

##   
## Attaching package: 'viridis'

## The following object is masked from 'package:scales':  
##   
## viridis\_pal

#theme\_set(theme\_pubclean())  
library("ggrepel")

## Warning: package 'ggrepel' was built under R version 3.5.3

#install.packages("janitor")  
library(janitor)

## Warning: package 'janitor' was built under R version 3.5.3

##   
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

#install.packages("formattable")  
#library(formattable)  
#install.packages("DT")  
#library(DT)  
#install.packages("FSelector")  
library(FSelector)

## Warning: package 'FSelector' was built under R version 3.5.3

#??FSelector

1. Read the “fores fire .csv” file from the following website.

fires <-read.csv('CKME 136 Forest Fire Data.csv',header=T)

1. Have a look at the data set. View(train.data)

# Fires  
  
head(fires)

## Année Cause Data.qualifier Juridiction Jurisdiction  
## 1 1990 Forest industry a Alberta Alberta  
## 2 1991 Forest industry a Alberta Alberta  
## 3 1992 Forest industry a Alberta Alberta  
## 4 1993 Forest industry a Alberta Alberta  
## 5 1994 Forest industry a Alberta Alberta  
## 6 1995 Forest industry a Alberta Alberta  
## Niveau.d.intervention Number Origine Protection.zone  
## 1 Normale 22 Industrie forestiere Intensive  
## 2 Normale 14 Industrie forestiere Intensive  
## 3 Normale 12 Industrie forestiere Intensive  
## 4 Normale 11 Industrie forestiere Intensive  
## 5 Normale 13 Industrie forestiere Intensive  
## 6 Normale 14 Industrie forestiere Intensive  
## Response.category Year Zone.de.protection  
## 1 Full 1990 Intensive  
## 2 Full 1991 Intensive  
## 3 Full 1992 Intensive  
## 4 Full 1993 Intensive  
## 5 Full 1994 Intensive  
## 6 Full 1995 Intensive

tail(fires)

## Année Cause Data.qualifier  
## 19866 2018 Unspecified human activities p  
## 19867 2018 Unspecified human activities p  
## 19868 2018 Unspecified human activities p  
## 19869 2018 Unspecified human activities p  
## 19870 2018 Unspecified human activities p  
## 19871 2018 Unspecified human activities p  
## Juridiction Jurisdiction Niveau.d.intervention  
## 19866 Nouvelle-Écosse Nova Scotia Modulée  
## 19867 Ontario Ontario Modulée  
## 19868 Île-du-Prince-Édouard Prince Edward Island Modulée  
## 19869 Québec Quebec Modulée  
## 19870 Saskatchewan Saskatchewan Modulée  
## 19871 Yukon Yukon Modulée  
## Number Origine Protection.zone  
## 19866 0 Activités humaines indéterminées Unspecified  
## 19867 11 Activités humaines indéterminées Unspecified  
## 19868 0 Activités humaines indéterminées Unspecified  
## 19869 6 Activités humaines indéterminées Unspecified  
## 19870 13 Activités humaines indéterminées Unspecified  
## 19871 6 Activités humaines indéterminées Unspecified  
## Response.category Year Zone.de.protection  
## 19866 Modified 2018 Indéterminée  
## 19867 Modified 2018 Indéterminée  
## 19868 Modified 2018 Indéterminée  
## 19869 Modified 2018 Indéterminée  
## 19870 Modified 2018 Indéterminée  
## 19871 Modified 2018 Indéterminée

str(fires)

## 'data.frame': 19871 obs. of 12 variables:  
## $ Année : int 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...  
## $ Cause : Factor w/ 10 levels "Forest industry",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Data.qualifier : Factor w/ 8 levels "a","e","E","n",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Juridiction : Factor w/ 13 levels "Alberta","Colombie-Britannique",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Jurisdiction : Factor w/ 13 levels "Alberta","British Columbia",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Niveau.d.intervention: Factor w/ 4 levels "Aucune","Indéterminée",..: 4 4 4 4 4 4 4 4 4 4 ...  
## $ Number : int 22 14 12 11 13 14 8 29 10 20 ...  
## $ Origine : Factor w/ 10 levels "Activités humaines indéterminées",..: 8 8 8 8 8 8 8 8 8 8 ...  
## $ Protection.zone : Factor w/ 3 levels "Intensive","Limited",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Response.category : Factor w/ 4 levels "Full","Modified",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Year : int 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...  
## $ Zone.de.protection : Factor w/ 3 levels "Indéterminée",..: 2 2 2 2 2 2 2 2 2 2 ...

# Check data types of attributes  
sapply(fires, class)

## Année Cause Data.qualifier   
## "integer" "factor" "factor"   
## Juridiction Jurisdiction Niveau.d.intervention   
## "factor" "factor" "factor"   
## Number Origine Protection.zone   
## "integer" "factor" "factor"   
## Response.category Year Zone.de.protection   
## "factor" "integer" "factor"

sapply(fires, typeof)

## Année Cause Data.qualifier   
## "integer" "integer" "integer"   
## Juridiction Jurisdiction Niveau.d.intervention   
## "integer" "integer" "integer"   
## Number Origine Protection.zone   
## "integer" "integer" "integer"   
## Response.category Year Zone.de.protection   
## "integer" "integer" "integer"

1. Extract relevant columns.

# Fires  
  
new\_fire <- fires[, c("Cause", "Jurisdiction", "Number", "Protection.zone", "Response.category", "Year")]  
#new\_fire

1. Check for missing values.

# Fire  
  
sum(is.na(new\_fire$Cause) == TRUE) # 0 Missing values.

## [1] 0

length(new\_fire$Cause)

## [1] 19871

sum(is.na(new\_fire$Jurisdiction) == TRUE) # 0 Missing values.

## [1] 0

length(new\_fire$Jurisdiction)

## [1] 19871

sum(is.na(new\_fire$Number) == TRUE) # 8352 initial missing values for "Number" field.

## [1] 8352

length(new\_fire$Number)

## [1] 19871

sum(is.na(new\_fire$Protection.zone) == TRUE) # 0 Missing values.

## [1] 0

length(new\_fire$Protection.zone)

## [1] 19871

sum(is.na(new\_fire$Response.category) == TRUE) # 0 Missing values.

## [1] 0

length(new\_fire$Response.category)

## [1] 19871

sum(is.na(new\_fire$Year) == TRUE) # 0 Missing values.

## [1] 0

length(new\_fire$Year)

## [1] 19871

1. Only “Number” has missing rows. Remove all rows with missing values.

# Fires  
  
# Remove remaining records with missing values.  
FireClean <- new\_fire[complete.cases(new\_fire),]  
  
nrow(FireClean) #11519 rows remaining

## [1] 11519

1. Check attributes after missing rows have been removed.

# Fire  
  
attach(FireClean)  
  
head(FireClean)

## Cause Jurisdiction Number Protection.zone Response.category  
## 1 Forest industry Alberta 22 Intensive Full  
## 2 Forest industry Alberta 14 Intensive Full  
## 3 Forest industry Alberta 12 Intensive Full  
## 4 Forest industry Alberta 11 Intensive Full  
## 5 Forest industry Alberta 13 Intensive Full  
## 6 Forest industry Alberta 14 Intensive Full  
## Year  
## 1 1990  
## 2 1991  
## 3 1992  
## 4 1993  
## 5 1994  
## 6 1995

tail(FireClean)

## Cause Jurisdiction Number  
## 19866 Unspecified human activities Nova Scotia 0  
## 19867 Unspecified human activities Ontario 11  
## 19868 Unspecified human activities Prince Edward Island 0  
## 19869 Unspecified human activities Quebec 6  
## 19870 Unspecified human activities Saskatchewan 13  
## 19871 Unspecified human activities Yukon 6  
## Protection.zone Response.category Year  
## 19866 Unspecified Modified 2018  
## 19867 Unspecified Modified 2018  
## 19868 Unspecified Modified 2018  
## 19869 Unspecified Modified 2018  
## 19870 Unspecified Modified 2018  
## 19871 Unspecified Modified 2018

str(FireClean)

## 'data.frame': 11519 obs. of 6 variables:  
## $ Cause : Factor w/ 10 levels "Forest industry",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Jurisdiction : Factor w/ 13 levels "Alberta","British Columbia",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Number : int 22 14 12 11 13 14 8 29 10 20 ...  
## $ Protection.zone : Factor w/ 3 levels "Intensive","Limited",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Response.category: Factor w/ 4 levels "Full","Modified",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Year : int 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...

dim(FireClean) # 11519 rows, 6 columns

## [1] 11519 6

# Check data types of attributes  
sapply(FireClean, class)

## Cause Jurisdiction Number Protection.zone   
## "factor" "factor" "integer" "factor"   
## Response.category Year   
## "factor" "integer"

levels(FireClean$Cause)

## [1] "Forest industry" "Incendiary"   
## [3] "Lightning" "Miscellaneous known causes"   
## [5] "Other industry" "Railways"   
## [7] "Recreation" "Residents"   
## [9] "Unspecified" "Unspecified human activities"

levels(FireClean$Jurisdiction)

## [1] "Alberta" "British Columbia"   
## [3] "Manitoba" "National parks"   
## [5] "New Brunswick" "Newfoundland and Labrador"  
## [7] "Northwest Territories" "Nova Scotia"   
## [9] "Ontario" "Prince Edward Island"   
## [11] "Quebec" "Saskatchewan"   
## [13] "Yukon"

levels(FireClean$Protection.zone)

## [1] "Intensive" "Limited" "Unspecified"

levels(FireClean$Response.category)

## [1] "Full" "Modified" "None" "Unspecified"

levels(FireClean$Year)

## NULL

summary(FireClean) # Only the "number" attribute maybe usefull with the summary

## Cause Jurisdiction   
## Lightning :1325 Quebec :1516   
## Unspecified :1290 Newfoundland and Labrador:1486   
## Miscellaneous known causes:1279 Ontario :1368   
## Recreation :1274 Manitoba :1304   
## Incendiary :1271 Yukon :1059   
## Residents :1268 Northwest Territories : 773   
## (Other) :3812 (Other) :4013   
## Number Protection.zone Response.category Year   
## Min. : 0.00 Intensive :8190 Full :4204 Min. :1990   
## 1st Qu.: 0.00 Limited :3281 Modified :3521 1st Qu.:1997   
## Median : 0.00 Unspecified: 48 None :3681 Median :2004   
## Mean : 18.64 Unspecified: 113 Mean :2004   
## 3rd Qu.: 4.00 3rd Qu.:2011   
## Max. :2913.00 Max. :2018   
##

1. Shorten “Jurisdiction” name

FireClean$Juris\_Long <- FireClean$Jurisdiction # Duplicate Jurisdiction column  
  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "British Columbia"] <- "BC"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Alberta"] <- "AB"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "National parks"] <- "NP"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Northwest Territories"] <- "NT"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Prince Edward Island"] <- "PE"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Yukon"] <- "YT"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "New Brunswick"] <- "NB"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Nova Scotia"] <- "NS"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Quebec"] <- "QC"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Manitoba"] <- "MB"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Newfoundland and Labrador"] <- "NL"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Ontario"] <- "ON"  
levels(FireClean$Jurisdiction)[levels(FireClean$Jurisdiction) == "Saskatchewan"] <- "SK"  
levels(FireClean$Jurisdiction)

## [1] "AB" "BC" "MB" "NP" "NB" "NL" "NT" "NS" "ON" "PE" "QC" "SK" "YT"

1. create new column: Cause\_Grouped People vs Lightning

FireClean <- FireClean %>%  
 mutate(Cause\_Grouped = case\_when(  
 Cause == "Lightning" ~ "Lightning",  
 TRUE ~ "People"  
 )  
 )

1. Create new column: Time1 Group Years

FireClean <- FireClean %>%  
 mutate(Time1 = case\_when(  
 Year <= 1995 ~ "Early 90s",  
 Year >= 1996 & Year <= 2000 ~ "Late 90s",  
 Year >= 2001 & Year <= 2005 ~ "Early 10s",  
 Year >= 2006 & Year <= 2010 ~ "Late 10s",  
 Year >= 2011 & Year <= 2015 ~ "Early 20s",  
 Year >= 2016 ~ "Late 20s"  
 )  
 )

1. Create new column: Time2 Group Years

FireClean <- FireClean %>%  
 mutate(Time2 = case\_when(  
 Year >= 1990 & Year <= 1999 ~ "1990s",  
 Year >= 2000 & Year <= 2009 ~ "2000s",  
 Year >= 2010 & Year <= 2018 ~ "2010s"  
 )  
 )

1. Group Provinces into regions

FireClean <- FireClean %>%   
 mutate(Region = case\_when(  
 Jurisdiction %in% c("AB", "MB", "SK") ~ "Prairie Region",  
 Jurisdiction %in% c("BC") ~ "Pacific Region",  
 Jurisdiction %in% c("NP") ~ "National Parks",  
 Jurisdiction %in% c("NB", "NL", "NS", "PE") ~ "Atlantic Region",  
 Jurisdiction %in% c("ON", "QC") ~ "Central Region",  
 Jurisdiction %in% c("YT", "NT") ~ "North Region"  
 )  
 )

1. Check structure of attributes again.

attach(FireClean)

## The following objects are masked from FireClean (pos = 3):  
##   
## Cause, Jurisdiction, Number, Protection.zone,  
## Response.category, Year

head(FireClean)

## Cause Jurisdiction Number Protection.zone Response.category  
## 1 Forest industry AB 22 Intensive Full  
## 2 Forest industry AB 14 Intensive Full  
## 3 Forest industry AB 12 Intensive Full  
## 4 Forest industry AB 11 Intensive Full  
## 5 Forest industry AB 13 Intensive Full  
## 6 Forest industry AB 14 Intensive Full  
## Year Juris\_Long Cause\_Grouped Time1 Time2 Region  
## 1 1990 Alberta People Early 90s 1990s Prairie Region  
## 2 1991 Alberta People Early 90s 1990s Prairie Region  
## 3 1992 Alberta People Early 90s 1990s Prairie Region  
## 4 1993 Alberta People Early 90s 1990s Prairie Region  
## 5 1994 Alberta People Early 90s 1990s Prairie Region  
## 6 1995 Alberta People Early 90s 1990s Prairie Region

tail(FireClean)

## Cause Jurisdiction Number Protection.zone  
## 11514 Unspecified human activities NS 0 Unspecified  
## 11515 Unspecified human activities ON 11 Unspecified  
## 11516 Unspecified human activities PE 0 Unspecified  
## 11517 Unspecified human activities QC 6 Unspecified  
## 11518 Unspecified human activities SK 13 Unspecified  
## 11519 Unspecified human activities YT 6 Unspecified  
## Response.category Year Juris\_Long Cause\_Grouped Time1  
## 11514 Modified 2018 Nova Scotia People Late 20s  
## 11515 Modified 2018 Ontario People Late 20s  
## 11516 Modified 2018 Prince Edward Island People Late 20s  
## 11517 Modified 2018 Quebec People Late 20s  
## 11518 Modified 2018 Saskatchewan People Late 20s  
## 11519 Modified 2018 Yukon People Late 20s  
## Time2 Region  
## 11514 2010s Atlantic Region  
## 11515 2010s Central Region  
## 11516 2010s Atlantic Region  
## 11517 2010s Central Region  
## 11518 2010s Prairie Region  
## 11519 2010s North Region

str(FireClean)

## 'data.frame': 11519 obs. of 11 variables:  
## $ Cause : Factor w/ 10 levels "Forest industry",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Jurisdiction : Factor w/ 13 levels "AB","BC","MB",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Number : int 22 14 12 11 13 14 8 29 10 20 ...  
## $ Protection.zone : Factor w/ 3 levels "Intensive","Limited",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Response.category: Factor w/ 4 levels "Full","Modified",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Year : int 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 ...  
## $ Juris\_Long : Factor w/ 13 levels "Alberta","British Columbia",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Cause\_Grouped : chr "People" "People" "People" "People" ...  
## $ Time1 : chr "Early 90s" "Early 90s" "Early 90s" "Early 90s" ...  
## $ Time2 : chr "1990s" "1990s" "1990s" "1990s" ...  
## $ Region : chr "Prairie Region" "Prairie Region" "Prairie Region" "Prairie Region" ...

dim(FireClean) # 11519 rows, 6 columns

## [1] 11519 11

# Check data types of attributes  
sapply(FireClean, class)

## Cause Jurisdiction Number Protection.zone   
## "factor" "factor" "integer" "factor"   
## Response.category Year Juris\_Long Cause\_Grouped   
## "factor" "integer" "factor" "character"   
## Time1 Time2 Region   
## "character" "character" "character"

levels(FireClean$Cause)

## [1] "Forest industry" "Incendiary"   
## [3] "Lightning" "Miscellaneous known causes"   
## [5] "Other industry" "Railways"   
## [7] "Recreation" "Residents"   
## [9] "Unspecified" "Unspecified human activities"

levels(FireClean$Jurisdiction)

## [1] "AB" "BC" "MB" "NP" "NB" "NL" "NT" "NS" "ON" "PE" "QC" "SK" "YT"

levels(FireClean$Protection.zone)

## [1] "Intensive" "Limited" "Unspecified"

levels(FireClean$Response.category)

## [1] "Full" "Modified" "None" "Unspecified"

levels(FireClean$Year)

## NULL

summary(FireClean) # Only the "number" attribute maybe usefull with the summary

## Cause Jurisdiction Number   
## Lightning :1325 QC :1516 Min. : 0.00   
## Unspecified :1290 NL :1486 1st Qu.: 0.00   
## Miscellaneous known causes:1279 ON :1368 Median : 0.00   
## Recreation :1274 MB :1304 Mean : 18.64   
## Incendiary :1271 YT :1059 3rd Qu.: 4.00   
## Residents :1268 NT : 773 Max. :2913.00   
## (Other) :3812 (Other):4013   
## Protection.zone Response.category Year   
## Intensive :8190 Full :4204 Min. :1990   
## Limited :3281 Modified :3521 1st Qu.:1997   
## Unspecified: 48 None :3681 Median :2004   
## Unspecified: 113 Mean :2004   
## 3rd Qu.:2011   
## Max. :2018   
##   
## Juris\_Long Cause\_Grouped Time1   
## Quebec :1516 Length:11519 Length:11519   
## Newfoundland and Labrador:1486 Class :character Class :character   
## Ontario :1368 Mode :character Mode :character   
## Manitoba :1304   
## Yukon :1059   
## Northwest Territories : 773   
## (Other) :4013   
## Time2 Region   
## Length:11519 Length:11519   
## Class :character Class :character   
## Mode :character Mode :character   
##   
##   
##   
##

summary(FireClean$Number)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00 0.00 0.00 18.64 4.00 2913.00

summary(FireClean$Year)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1990 1997 2004 2004 2011 2018

1. Visulization, Barplot, Boxplot

data <- tbl\_df(FireClean)  
  
head(data)

## # A tibble: 6 x 11  
## Cause Jurisdiction Number Protection.zone Response.catego~ Year  
## <fct> <fct> <int> <fct> <fct> <int>  
## 1 Fore~ AB 22 Intensive Full 1990  
## 2 Fore~ AB 14 Intensive Full 1991  
## 3 Fore~ AB 12 Intensive Full 1992  
## 4 Fore~ AB 11 Intensive Full 1993  
## 5 Fore~ AB 13 Intensive Full 1994  
## 6 Fore~ AB 14 Intensive Full 1995  
## # ... with 5 more variables: Juris\_Long <fct>, Cause\_Grouped <chr>,  
## # Time1 <chr>, Time2 <chr>, Region <chr>

########################################################################################################################################  
#11   
  
pivot3 <- data %>%  
select(Year, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Year Number Cause\_Grouped  
## <int> <int> <chr>   
## 1 1990 22 People   
## 2 1991 14 People   
## 3 1992 12 People   
## 4 1993 11 People   
## 5 1994 13 People   
## 6 1995 14 People

pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Cause\_Grouped, Number, Year, Jurisdiction) %>%   
 filter(Jurisdiction == "BC") %>%   
 group\_by(Year, Cause\_Grouped) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
  
pivot3 %>%   
 spread(Year, sum\_Number)

## # A tibble: 2 x 30  
## Cause\_Grouped `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <chr> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 Lightning 2015 759 2344 609 2913 342 723 675  
## 2 People 1240 1254 1461 888 1175 1132 623 486  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

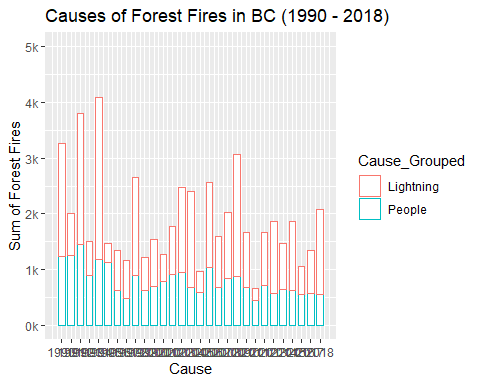
############  
#SUMMING COLUMNS AND ROWS  
pivot4 <- pivot3 %>%   
 spread(Year, sum\_Number)  
  
pivot4

## # A tibble: 2 x 30  
## Cause\_Grouped `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <chr> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 Lightning 2015 759 2344 609 2913 342 723 675  
## 2 People 1240 1254 1461 888 1175 1132 623 486  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

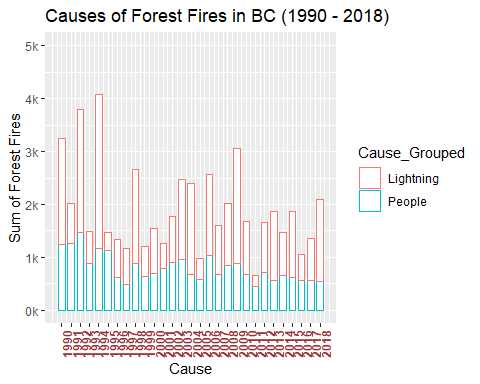
pivot5 <- pivot4 %>%  
 adorn\_totals("row") %>%   
 adorn\_totals("col")   
  
pivot5

## Cause\_Grouped 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001  
## Lightning 2015 759 2344 609 2913 342 723 675 1773 585 842 479  
## People 1240 1254 1461 888 1175 1132 623 486 889 629 697 785  
## Total 3255 2013 3805 1497 4088 1474 1346 1161 2662 1214 1539 1264  
## 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015  
## 870 1513 1716 384 1536 912 1175 2184 992 209 944 1296 816 1237  
## 911 959 682 587 1033 682 848 880 681 446 718 569 649 621  
## 1781 2472 2398 971 2569 1594 2023 3064 1673 655 1662 1865 1465 1858  
## 2016 2017 2018 Total  
## 486 782 1537 32648  
## 563 569 549 23206  
## 1049 1351 2086 55854

##################################################  
# Test 1  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Year,  
 y = sum\_Number,  
 color = Cause\_Grouped)  
 ) +  
 geom\_bar(stat="identity", width = 0.7, fill="white") +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 scale\_x\_continuous(breaks=1990:2018)+  
 #stat\_summary(fun.y = sum, aes(label = ..y.., group = Year), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 5000))+  
 labs(title = "Causes of Forest Fires in BC (1990 - 2018)")  
  
p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 0 ,size = 9, angle = 90))



########################################################################################################################################  
#10-B Practice sum column and rows  
  
pivot3 <- data %>%  
select(Region, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Region Number Cause\_Grouped  
## <chr> <int> <chr>   
## 1 Prairie Region 22 People   
## 2 Prairie Region 14 People   
## 3 Prairie Region 12 People   
## 4 Prairie Region 11 People   
## 5 Prairie Region 13 People   
## 6 Prairie Region 14 People

# attach(pivot3)  
# detach(pivot3)  
  
pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Region, Number, Cause\_Grouped) %>%   
 group\_by(Cause\_Grouped, Region) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))  
   
pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)

## # A tibble: 6 x 3  
## Region Lightning People  
## <chr> <int> <int>  
## 1 Atlantic Region 2120 21073  
## 2 Central Region 25023 30116  
## 3 National Parks 1421 1096  
## 4 North Region 8549 2163  
## 5 Pacific Region 32648 23206  
## 6 Prairie Region 31234 36116

############  
#SUMMING COLUMNS AND ROWS  
pivot4 <- pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)  
  
pivot4

## # A tibble: 6 x 3  
## Region Lightning People  
## <chr> <int> <int>  
## 1 Atlantic Region 2120 21073  
## 2 Central Region 25023 30116  
## 3 National Parks 1421 1096  
## 4 North Region 8549 2163  
## 5 Pacific Region 32648 23206  
## 6 Prairie Region 31234 36116

pivot5 <- pivot4 %>%  
 adorn\_totals("row") %>%   
 adorn\_totals("col")   
  
pivot5

## Region Lightning People Total  
## Atlantic Region 2120 21073 23193  
## Central Region 25023 30116 55139  
## National Parks 1421 1096 2517  
## North Region 8549 2163 10712  
## Pacific Region 32648 23206 55854  
## Prairie Region 31234 36116 67350  
## Total 100995 113770 214765

##################################################  
#CORRELATION USE FOR pivot3 ONLY  
CorDataFrame <- pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)  
  
CorDataFrame

## # A tibble: 6 x 3  
## Region Lightning People  
## <chr> <int> <int>  
## 1 Atlantic Region 2120 21073  
## 2 Central Region 25023 30116  
## 3 National Parks 1421 1096  
## 4 North Region 8549 2163  
## 5 Pacific Region 32648 23206  
## 6 Prairie Region 31234 36116

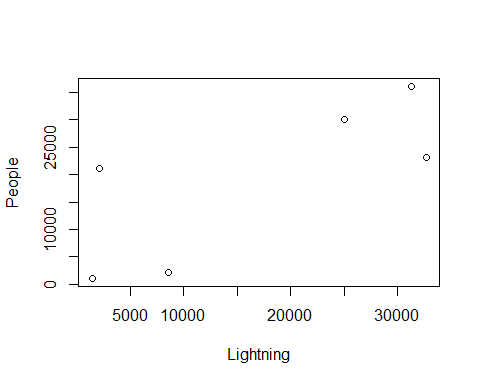
sapply(CorDataFrame, is.numeric) # Which columns are numeric?

## Region Lightning People   
## FALSE TRUE TRUE

my\_num\_data <- CorDataFrame[, sapply(CorDataFrame, is.numeric)] # Subset numeric columns  
my\_num\_data

## # A tibble: 6 x 2  
## Lightning People  
## <int> <int>  
## 1 2120 21073  
## 2 25023 30116  
## 3 1421 1096  
## 4 8549 2163  
## 5 32648 23206  
## 6 31234 36116

plot(my\_num\_data) # Works



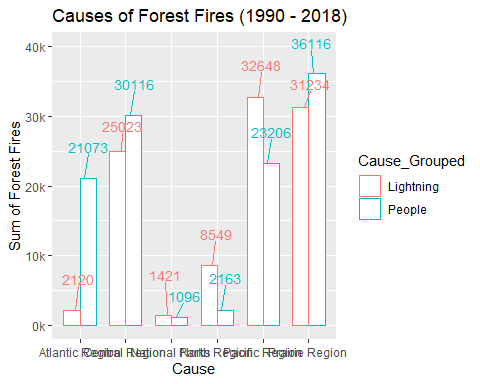
cor(my\_num\_data)

## Lightning People  
## Lightning 1.0000000 0.7550789  
## People 0.7550789 1.0000000

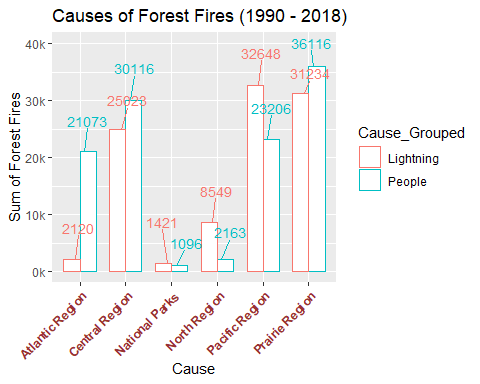
##################################################  
  
# Test 1  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3,  
 aes(x = Region, y = sum\_Number, color=Cause\_Grouped)) +  
 geom\_bar(position = "dodge", stat="identity", width = 0.7, fill="white") +  
 geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 40000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")

## Warning: `show\_guide` has been deprecated. Please use `show.legend`  
## instead.

p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 45))



########################################################################################################################################  
#10   
  
pivot3 <- data %>%  
select(Region, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Region Number Cause\_Grouped  
## <chr> <int> <chr>   
## 1 Prairie Region 22 People   
## 2 Prairie Region 14 People   
## 3 Prairie Region 12 People   
## 4 Prairie Region 11 People   
## 5 Prairie Region 13 People   
## 6 Prairie Region 14 People

# attach(pivot3)  
# detach(pivot3)  
  
pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Region, Number, Cause\_Grouped) %>%   
 group\_by(Cause\_Grouped, Region) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
  
pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)

## # A tibble: 6 x 3  
## Region Lightning People  
## <chr> <int> <int>  
## 1 Atlantic Region 2120 21073  
## 2 Central Region 25023 30116  
## 3 National Parks 1421 1096  
## 4 North Region 8549 2163  
## 5 Pacific Region 32648 23206  
## 6 Prairie Region 31234 36116

##################################################  
#CORRELATION  
CorDataFrame <- pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)  
  
CorDataFrame

## # A tibble: 6 x 3  
## Region Lightning People  
## <chr> <int> <int>  
## 1 Atlantic Region 2120 21073  
## 2 Central Region 25023 30116  
## 3 National Parks 1421 1096  
## 4 North Region 8549 2163  
## 5 Pacific Region 32648 23206  
## 6 Prairie Region 31234 36116

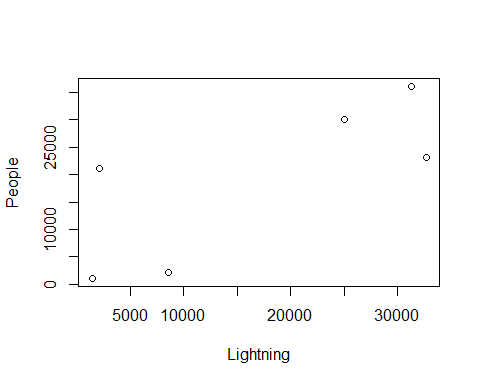
sapply(CorDataFrame, is.numeric) # Which columns are numeric?

## Region Lightning People   
## FALSE TRUE TRUE

my\_num\_data <- CorDataFrame[, sapply(CorDataFrame, is.numeric)] # Subset numeric columns  
my\_num\_data

## # A tibble: 6 x 2  
## Lightning People  
## <int> <int>  
## 1 2120 21073  
## 2 25023 30116  
## 3 1421 1096  
## 4 8549 2163  
## 5 32648 23206  
## 6 31234 36116

plot(my\_num\_data) # Works



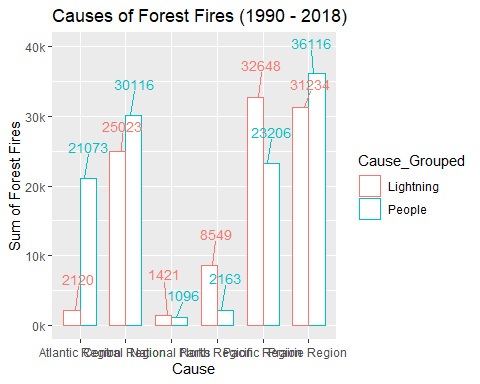
cor(my\_num\_data)

## Lightning People  
## Lightning 1.0000000 0.7550789  
## People 0.7550789 1.0000000

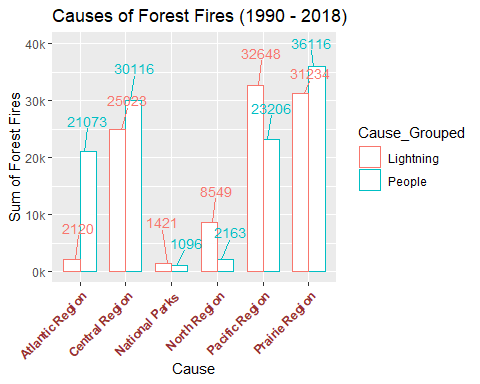
##################################################  
  
# Test 1  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3,  
 aes(x = Region, y = sum\_Number, color=Cause\_Grouped)) +  
 geom\_bar(position = "dodge", stat="identity", width = 0.7, fill="white") +  
 geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 40000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")

## Warning: `show\_guide` has been deprecated. Please use `show.legend`  
## instead.

p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 45))



########################################################################################################################################  
#9  
  
pivot3 <- data %>%  
select(Year, Number, Jurisdiction)  
head(pivot3)

## # A tibble: 6 x 3  
## Year Number Jurisdiction  
## <int> <int> <fct>   
## 1 1990 22 AB   
## 2 1991 14 AB   
## 3 1992 12 AB   
## 4 1993 11 AB   
## 5 1994 13 AB   
## 6 1995 14 AB

pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Year) %>%   
 group\_by(Year, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
  
pivot3 %>%   
 spread(Year, sum\_Number)

## # A tibble: 13 x 30  
## Jurisdiction `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <fct> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 AB 1296 923 1055 848 872 803 376 456  
## 2 BC 3255 2013 3805 1497 4088 1474 1346 1161  
## 3 MB 570 676 298 239 555 660 424 373  
## 4 NP 128 56 72 63 166 62 NA 55  
## 5 NB 377 656 576 430 518 547 367 368  
## 6 NL 197 166 109 83 143 103 148 110  
## 7 NT 236 331 285 469 627 215 350 105  
## 8 NS 498 733 299 317 245 408 272 371  
## 9 ON 1614 2560 960 743 1053 2122 1245 1636  
## 10 PE 38 48 27 29 43 29 0 34  
## 11 QC 851 1216 765 543 499 1265 1250 876  
## 12 SK 897 762 701 646 699 650 422 491  
## 13 YT 154 187 116 136 255 148 149 112  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

#Produces matrix with zero  
pivot4 <- pivot3 %>%   
 spread(Year, sum\_Number)  
  
pivot4

## # A tibble: 13 x 30  
## Jurisdiction `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <fct> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 AB 1296 923 1055 848 872 803 376 456  
## 2 BC 3255 2013 3805 1497 4088 1474 1346 1161  
## 3 MB 570 676 298 239 555 660 424 373  
## 4 NP 128 56 72 63 166 62 NA 55  
## 5 NB 377 656 576 430 518 547 367 368  
## 6 NL 197 166 109 83 143 103 148 110  
## 7 NT 236 331 285 469 627 215 350 105  
## 8 NS 498 733 299 317 245 408 272 371  
## 9 ON 1614 2560 960 743 1053 2122 1245 1636  
## 10 PE 38 48 27 29 43 29 0 34  
## 11 QC 851 1216 765 543 499 1265 1250 876  
## 12 SK 897 762 701 646 699 650 422 491  
## 13 YT 154 187 116 136 255 148 149 112  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

#Remove NAs from matrix, replace with zero  
pivot4[is.na(pivot4)] <- 0  
#Matrix now has NAs removed, now replaced with zero  
pivot4

## # A tibble: 13 x 30  
## Jurisdiction `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <fct> <int> <int> <int> <int> <int> <int> <dbl> <int>  
## 1 AB 1296 923 1055 848 872 803 376 456  
## 2 BC 3255 2013 3805 1497 4088 1474 1346 1161  
## 3 MB 570 676 298 239 555 660 424 373  
## 4 NP 128 56 72 63 166 62 0 55  
## 5 NB 377 656 576 430 518 547 367 368  
## 6 NL 197 166 109 83 143 103 148 110  
## 7 NT 236 331 285 469 627 215 350 105  
## 8 NS 498 733 299 317 245 408 272 371  
## 9 ON 1614 2560 960 743 1053 2122 1245 1636  
## 10 PE 38 48 27 29 43 29 0 34  
## 11 QC 851 1216 765 543 499 1265 1250 876  
## 12 SK 897 762 701 646 699 650 422 491  
## 13 YT 154 187 116 136 255 148 149 112  
## # ... with 21 more variables: `1998` <dbl>, `1999` <dbl>, `2000` <dbl>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

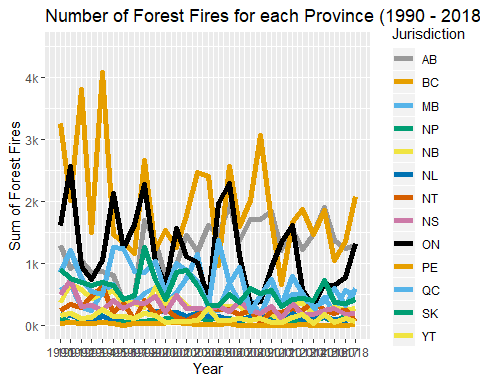
############  
#SUMMING COLUMNS AND ROWS  
  
pivot5 <- pivot4 %>%  
 adorn\_totals("row") %>%   
 adorn\_totals("col")   
  
pivot5

## Jurisdiction 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000  
## AB 1296 923 1055 848 872 803 376 456 1698 1355 783  
## BC 3255 2013 3805 1497 4088 1474 1346 1161 2662 1214 1539  
## MB 570 676 298 239 555 660 424 373 515 613 354  
## NP 128 56 72 63 166 62 0 55 0 0 0  
## NB 377 656 576 430 518 547 367 368 286 607 333  
## NL 197 166 109 83 143 103 148 110 192 228 219  
## NT 236 331 285 469 627 215 350 105 399 170 275  
## NS 498 733 299 317 245 408 272 371 348 464 212  
## ON 1614 2560 960 743 1053 2122 1245 1636 2278 1016 644  
## PE 38 48 27 29 43 29 0 34 27 34 0  
## QC 851 1216 765 543 499 1265 1250 876 854 1037 516  
## SK 897 762 701 646 699 650 422 491 1266 735 419  
## YT 154 187 116 136 255 148 149 112 198 160 55  
## Total 10111 10327 9068 6043 9763 8486 6349 6148 10723 7633 5349  
## 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014  
## 989 1447 1188 1612 1448 1954 1349 1712 1710 1840 1218 1568 1226 1470  
## 1264 1781 2472 2398 971 2569 1594 2023 3064 1673 655 1662 1865 1465  
## 537 754 1148 234 248 682 382 397 184 581 312 497 494 245  
## 128 86 112 101 107 118 65 108 134 121 76 86 93 84  
## 490 317 228 253 306 308 316 168 197 179 81 344 354 178  
## 202 143 191 153 145 96 87 139 176 61 53 198 101 124  
## 127 85 160 297 261 261 184 241 42 224 207 279 248 385  
## 486 267 272 258 302 234 393 248 198 313 116 352 171 171  
## 1561 1111 1012 424 1961 2298 1124 341 385 931 1334 1619 580 303  
## 46 29 14 20 13 36 8 4 8 4 4 8 10 4  
## 1003 895 716 319 1374 683 935 222 483 737 329 795 515 292  
## 857 880 640 329 323 501 370 599 511 571 302 422 430 403  
## 63 66 77 282 83 80 110 76 118 56 56 126 177 34  
## 7753 7861 8230 6680 7542 9820 6917 6278 7210 7291 4743 7956 6264 5158  
## 2015 2016 2017 2018 Total  
## 1898 1376 1244 1288 37002  
## 1858 1049 1351 2086 55854  
## 459 202 559 477 13669  
## 122 65 169 140 2517  
## 221 285 245 282 9817  
## 128 91 80 132 3998  
## 245 189 262 59 7218  
## 247 274 175 190 8834  
## 668 645 776 1327 34271  
## 5 8 4 10 544  
## 384 602 319 593 20868  
## 720 364 353 416 16679  
## 185 53 115 67 3494  
## 7140 5203 5652 7067 214765

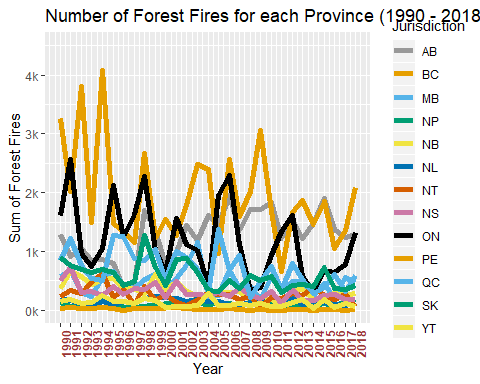
##################################################  
# Test 4  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Year,  
 y = sum\_Number,  
 color = Jurisdiction)  
 ) +  
 geom\_line(stat="identity", width = 0.7, fill="white", size = 2) +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 scale\_color\_manual(values = c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7",  
 "#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442")) +  
 scale\_x\_continuous(breaks=1990:2018)+  
 #stat\_summary(fun.y = sum, aes(label = ..y.., group = Year), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Year") +  
 ylab("Number of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 4500))+  
 labs(title = "Number of Forest Fires for each Province (1990 - 2018)")

## Warning: Ignoring unknown parameters: width, fill

p



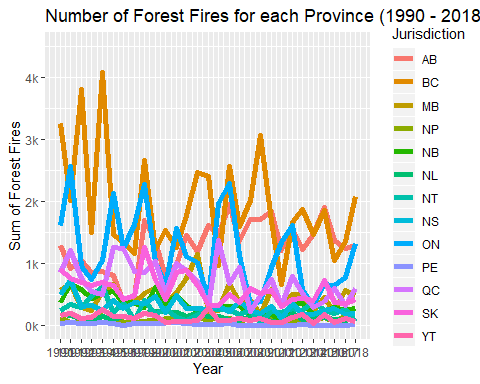
p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 0 ,size = 9, angle = 90))



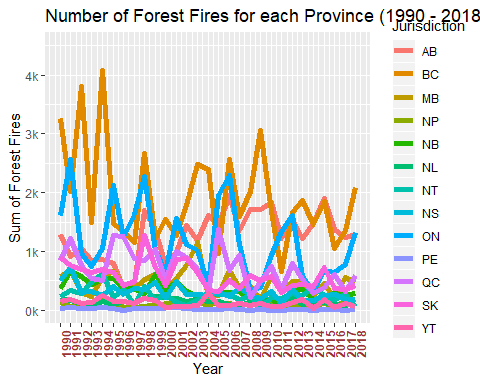
##################################################  
# Test 3  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Year,  
 y = sum\_Number,  
 color = Jurisdiction)  
 ) +  
 geom\_line(stat="identity", width = 0.7, fill="white", size = 2) +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 scale\_x\_continuous(breaks=1990:2018)+  
 #stat\_summary(fun.y = sum, aes(label = ..y.., group = Year), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Year") +  
 ylab("Number of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 4500))+  
 labs(title = "Number of Forest Fires for each Province (1990 - 2018)")

## Warning: Ignoring unknown parameters: width, fill

p



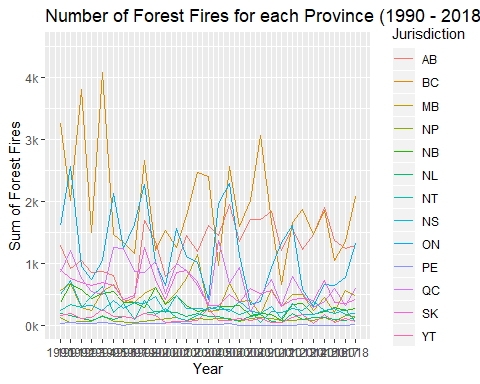
p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 0 ,size = 9, angle = 90))



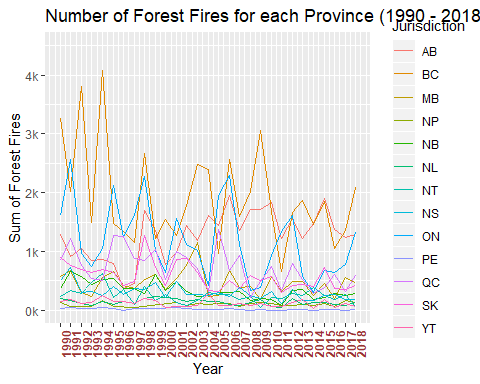
##################################################  
# Test 2  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Year,  
 y = sum\_Number,  
 color = Jurisdiction)  
 ) +  
 geom\_line(stat="identity", width = 0.7, fill="white") +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 scale\_x\_continuous(breaks=1990:2018)+  
 #stat\_summary(fun.y = sum, aes(label = ..y.., group = Year), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Year") +  
 ylab("Number of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 4500))+  
 labs(title = "Number of Forest Fires for each Province (1990 - 2018)")

## Warning: Ignoring unknown parameters: width, fill

p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 0 ,size = 9, angle = 90))



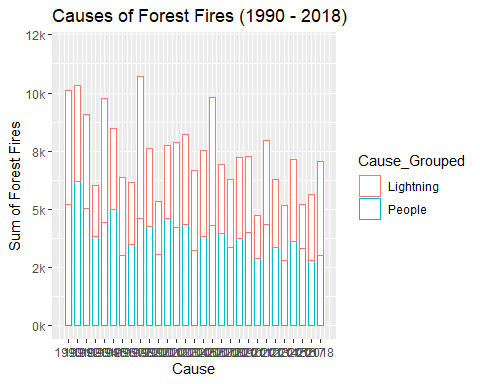
########################################################################################################################################  
#8  
  
pivot3 <- data %>%  
select(Year, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Year Number Cause\_Grouped  
## <int> <int> <chr>   
## 1 1990 22 People   
## 2 1991 14 People   
## 3 1992 12 People   
## 4 1993 11 People   
## 5 1994 13 People   
## 6 1995 14 People

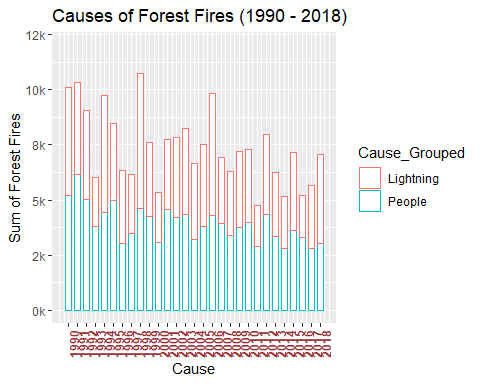
pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Cause\_Grouped, Number, Year) %>%   
 group\_by(Year, Cause\_Grouped) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
  
pivot3 %>%   
 spread(Year, sum\_Number)

## # A tibble: 2 x 30  
## Cause\_Grouped `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <chr> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 Lightning 4895 4146 4021 2229 5324 3504 3317 2652  
## 2 People 5216 6181 5047 3814 4439 4982 3032 3496  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

##################################################  
# Test 8  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Year,  
 y = sum\_Number,  
 color = Cause\_Grouped)  
 ) +  
 geom\_bar(stat="identity", width = 0.7, fill="white") +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 scale\_x\_continuous(breaks=1990:2018)+  
 #stat\_summary(fun.y = sum, aes(label = ..y.., group = Year), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 12000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")  
  
p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 0 ,size = 9, angle = 90))



########################################################################################################################################  
#7  
  
pivot3 <- data %>%  
select(Jurisdiction, Number, Year)  
head(pivot3)

## # A tibble: 6 x 3  
## Jurisdiction Number Year  
## <fct> <int> <int>  
## 1 AB 22 1990  
## 2 AB 14 1991  
## 3 AB 12 1992  
## 4 AB 11 1993  
## 5 AB 13 1994  
## 6 AB 14 1995

pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Year) %>%   
 group\_by(Year, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
#Produces matrix with zero  
pivot4 <- pivot3 %>%   
 spread(Year, sum\_Number)  
  
pivot4

## # A tibble: 13 x 30  
## Jurisdiction `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <fct> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 AB 1296 923 1055 848 872 803 376 456  
## 2 BC 3255 2013 3805 1497 4088 1474 1346 1161  
## 3 MB 570 676 298 239 555 660 424 373  
## 4 NP 128 56 72 63 166 62 NA 55  
## 5 NB 377 656 576 430 518 547 367 368  
## 6 NL 197 166 109 83 143 103 148 110  
## 7 NT 236 331 285 469 627 215 350 105  
## 8 NS 498 733 299 317 245 408 272 371  
## 9 ON 1614 2560 960 743 1053 2122 1245 1636  
## 10 PE 38 48 27 29 43 29 0 34  
## 11 QC 851 1216 765 543 499 1265 1250 876  
## 12 SK 897 762 701 646 699 650 422 491  
## 13 YT 154 187 116 136 255 148 149 112  
## # ... with 21 more variables: `1998` <int>, `1999` <int>, `2000` <int>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

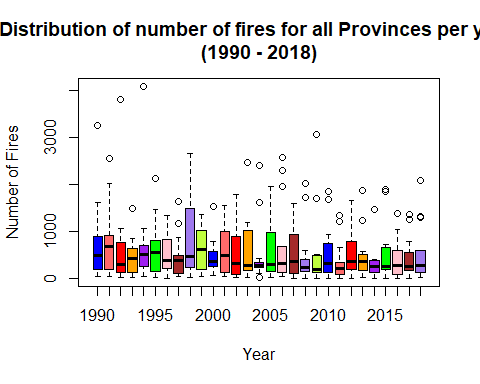
#Remove NAs from matrix, replace with zero  
pivot4[is.na(pivot4)] <- 0  
#Matrix now has NAs removed, now replaced with zero  
pivot4

## # A tibble: 13 x 30  
## Jurisdiction `1990` `1991` `1992` `1993` `1994` `1995` `1996` `1997`  
## <fct> <int> <int> <int> <int> <int> <int> <dbl> <int>  
## 1 AB 1296 923 1055 848 872 803 376 456  
## 2 BC 3255 2013 3805 1497 4088 1474 1346 1161  
## 3 MB 570 676 298 239 555 660 424 373  
## 4 NP 128 56 72 63 166 62 0 55  
## 5 NB 377 656 576 430 518 547 367 368  
## 6 NL 197 166 109 83 143 103 148 110  
## 7 NT 236 331 285 469 627 215 350 105  
## 8 NS 498 733 299 317 245 408 272 371  
## 9 ON 1614 2560 960 743 1053 2122 1245 1636  
## 10 PE 38 48 27 29 43 29 0 34  
## 11 QC 851 1216 765 543 499 1265 1250 876  
## 12 SK 897 762 701 646 699 650 422 491  
## 13 YT 154 187 116 136 255 148 149 112  
## # ... with 21 more variables: `1998` <dbl>, `1999` <dbl>, `2000` <dbl>,  
## # `2001` <int>, `2002` <int>, `2003` <int>, `2004` <int>, `2005` <int>,  
## # `2006` <int>, `2007` <int>, `2008` <int>, `2009` <int>, `2010` <int>,  
## # `2011` <int>, `2012` <int>, `2013` <int>, `2014` <int>, `2015` <int>,  
## # `2016` <int>, `2017` <int>, `2018` <int>

############  
#SUMMING COLUMNS AND ROWS  
  
pivot5 <- pivot4 %>%  
 adorn\_totals("row") %>%   
 adorn\_totals("col")   
  
pivot5

## Jurisdiction 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000  
## AB 1296 923 1055 848 872 803 376 456 1698 1355 783  
## BC 3255 2013 3805 1497 4088 1474 1346 1161 2662 1214 1539  
## MB 570 676 298 239 555 660 424 373 515 613 354  
## NP 128 56 72 63 166 62 0 55 0 0 0  
## NB 377 656 576 430 518 547 367 368 286 607 333  
## NL 197 166 109 83 143 103 148 110 192 228 219  
## NT 236 331 285 469 627 215 350 105 399 170 275  
## NS 498 733 299 317 245 408 272 371 348 464 212  
## ON 1614 2560 960 743 1053 2122 1245 1636 2278 1016 644  
## PE 38 48 27 29 43 29 0 34 27 34 0  
## QC 851 1216 765 543 499 1265 1250 876 854 1037 516  
## SK 897 762 701 646 699 650 422 491 1266 735 419  
## YT 154 187 116 136 255 148 149 112 198 160 55  
## Total 10111 10327 9068 6043 9763 8486 6349 6148 10723 7633 5349  
## 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014  
## 989 1447 1188 1612 1448 1954 1349 1712 1710 1840 1218 1568 1226 1470  
## 1264 1781 2472 2398 971 2569 1594 2023 3064 1673 655 1662 1865 1465  
## 537 754 1148 234 248 682 382 397 184 581 312 497 494 245  
## 128 86 112 101 107 118 65 108 134 121 76 86 93 84  
## 490 317 228 253 306 308 316 168 197 179 81 344 354 178  
## 202 143 191 153 145 96 87 139 176 61 53 198 101 124  
## 127 85 160 297 261 261 184 241 42 224 207 279 248 385  
## 486 267 272 258 302 234 393 248 198 313 116 352 171 171  
## 1561 1111 1012 424 1961 2298 1124 341 385 931 1334 1619 580 303  
## 46 29 14 20 13 36 8 4 8 4 4 8 10 4  
## 1003 895 716 319 1374 683 935 222 483 737 329 795 515 292  
## 857 880 640 329 323 501 370 599 511 571 302 422 430 403  
## 63 66 77 282 83 80 110 76 118 56 56 126 177 34  
## 7753 7861 8230 6680 7542 9820 6917 6278 7210 7291 4743 7956 6264 5158  
## 2015 2016 2017 2018 Total  
## 1898 1376 1244 1288 37002  
## 1858 1049 1351 2086 55854  
## 459 202 559 477 13669  
## 122 65 169 140 2517  
## 221 285 245 282 9817  
## 128 91 80 132 3998  
## 245 189 262 59 7218  
## 247 274 175 190 8834  
## 668 645 776 1327 34271  
## 5 8 4 10 544  
## 384 602 319 593 20868  
## 720 364 353 416 16679  
## 185 53 115 67 3494  
## 7140 5203 5652 7067 214765

##################################################  
  
boxplot(sum\_Number~Year,  
 data=pivot3,  
 main="Distribution of number of fires for all Provinces per year \n(1990 - 2018)",  
 xlab="Year",  
 ylab="Number of Fires",  
 col=c("blue", "indianred1","red","orange","purple","green",  
 "pink","brown", "mediumpurple2","olivedrab1"),  
 border="black"  
)



########################################################################################################################################  
#6 Boxplot  
pivot3 <- data %>%  
select(Jurisdiction, Number, Year)  
head(pivot3)

## # A tibble: 6 x 3  
## Jurisdiction Number Year  
## <fct> <int> <int>  
## 1 AB 22 1990  
## 2 AB 14 1991  
## 3 AB 12 1992  
## 4 AB 11 1993  
## 5 AB 13 1994  
## 6 AB 14 1995

pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Year) %>%   
 group\_by(Year, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
  
pivot3 %>%   
 spread(Jurisdiction, sum\_Number)

## # A tibble: 29 x 14  
## # Groups: Year [29]  
## Year AB BC MB NP NB NL NT NS ON PE QC  
## <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 1990 1296 3255 570 128 377 197 236 498 1614 38 851  
## 2 1991 923 2013 676 56 656 166 331 733 2560 48 1216  
## 3 1992 1055 3805 298 72 576 109 285 299 960 27 765  
## 4 1993 848 1497 239 63 430 83 469 317 743 29 543  
## 5 1994 872 4088 555 166 518 143 627 245 1053 43 499  
## 6 1995 803 1474 660 62 547 103 215 408 2122 29 1265  
## 7 1996 376 1346 424 NA 367 148 350 272 1245 0 1250  
## 8 1997 456 1161 373 55 368 110 105 371 1636 34 876  
## 9 1998 1698 2662 515 NA 286 192 399 348 2278 27 854  
## 10 1999 1355 1214 613 NA 607 228 170 464 1016 34 1037  
## # ... with 19 more rows, and 2 more variables: SK <int>, YT <int>

#Produces matrix with zero  
pivot4 <- pivot3 %>%   
 spread(Jurisdiction, sum\_Number)  
  
pivot4

## # A tibble: 29 x 14  
## # Groups: Year [29]  
## Year AB BC MB NP NB NL NT NS ON PE QC  
## <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 1990 1296 3255 570 128 377 197 236 498 1614 38 851  
## 2 1991 923 2013 676 56 656 166 331 733 2560 48 1216  
## 3 1992 1055 3805 298 72 576 109 285 299 960 27 765  
## 4 1993 848 1497 239 63 430 83 469 317 743 29 543  
## 5 1994 872 4088 555 166 518 143 627 245 1053 43 499  
## 6 1995 803 1474 660 62 547 103 215 408 2122 29 1265  
## 7 1996 376 1346 424 NA 367 148 350 272 1245 0 1250  
## 8 1997 456 1161 373 55 368 110 105 371 1636 34 876  
## 9 1998 1698 2662 515 NA 286 192 399 348 2278 27 854  
## 10 1999 1355 1214 613 NA 607 228 170 464 1016 34 1037  
## # ... with 19 more rows, and 2 more variables: SK <int>, YT <int>

#Remove NAs from matrix, replace with zero  
pivot4[is.na(pivot4)] <- 0  
#Matrix now has NAs removed, now replaced with zero  
pivot4

## # A tibble: 29 x 14  
## # Groups: Year [29]  
## Year AB BC MB NP NB NL NT NS ON PE QC  
## <int> <int> <int> <int> <dbl> <int> <int> <int> <int> <int> <dbl> <int>  
## 1 1990 1296 3255 570 128 377 197 236 498 1614 38 851  
## 2 1991 923 2013 676 56 656 166 331 733 2560 48 1216  
## 3 1992 1055 3805 298 72 576 109 285 299 960 27 765  
## 4 1993 848 1497 239 63 430 83 469 317 743 29 543  
## 5 1994 872 4088 555 166 518 143 627 245 1053 43 499  
## 6 1995 803 1474 660 62 547 103 215 408 2122 29 1265  
## 7 1996 376 1346 424 0 367 148 350 272 1245 0 1250  
## 8 1997 456 1161 373 55 368 110 105 371 1636 34 876  
## 9 1998 1698 2662 515 0 286 192 399 348 2278 27 854  
## 10 1999 1355 1214 613 0 607 228 170 464 1016 34 1037  
## # ... with 19 more rows, and 2 more variables: SK <int>, YT <int>

############  
#SUMMING COLUMNS AND ROWS  
  
pivot5 <- pivot4 %>%  
 adorn\_totals("row") %>%   
 adorn\_totals("col")   
  
pivot5

## Year AB BC MB NP NB NL NT NS ON PE QC SK  
## 1990 1296 3255 570 128 377 197 236 498 1614 38 851 897  
## 1991 923 2013 676 56 656 166 331 733 2560 48 1216 762  
## 1992 1055 3805 298 72 576 109 285 299 960 27 765 701  
## 1993 848 1497 239 63 430 83 469 317 743 29 543 646  
## 1994 872 4088 555 166 518 143 627 245 1053 43 499 699  
## 1995 803 1474 660 62 547 103 215 408 2122 29 1265 650  
## 1996 376 1346 424 0 367 148 350 272 1245 0 1250 422  
## 1997 456 1161 373 55 368 110 105 371 1636 34 876 491  
## 1998 1698 2662 515 0 286 192 399 348 2278 27 854 1266  
## 1999 1355 1214 613 0 607 228 170 464 1016 34 1037 735  
## 2000 783 1539 354 0 333 219 275 212 644 0 516 419  
## 2001 989 1264 537 128 490 202 127 486 1561 46 1003 857  
## 2002 1447 1781 754 86 317 143 85 267 1111 29 895 880  
## 2003 1188 2472 1148 112 228 191 160 272 1012 14 716 640  
## 2004 1612 2398 234 101 253 153 297 258 424 20 319 329  
## 2005 1448 971 248 107 306 145 261 302 1961 13 1374 323  
## 2006 1954 2569 682 118 308 96 261 234 2298 36 683 501  
## 2007 1349 1594 382 65 316 87 184 393 1124 8 935 370  
## 2008 1712 2023 397 108 168 139 241 248 341 4 222 599  
## 2009 1710 3064 184 134 197 176 42 198 385 8 483 511  
## 2010 1840 1673 581 121 179 61 224 313 931 4 737 571  
## 2011 1218 655 312 76 81 53 207 116 1334 4 329 302  
## 2012 1568 1662 497 86 344 198 279 352 1619 8 795 422  
## 2013 1226 1865 494 93 354 101 248 171 580 10 515 430  
## 2014 1470 1465 245 84 178 124 385 171 303 4 292 403  
## 2015 1898 1858 459 122 221 128 245 247 668 5 384 720  
## 2016 1376 1049 202 65 285 91 189 274 645 8 602 364  
## 2017 1244 1351 559 169 245 80 262 175 776 4 319 353  
## 2018 1288 2086 477 140 282 132 59 190 1327 10 593 416  
## Total 37002 55854 13669 2517 9817 3998 7218 8834 34271 544 20868 16679  
## YT Total  
## 154 10111  
## 187 10327  
## 116 9068  
## 136 6043  
## 255 9763  
## 148 8486  
## 149 6349  
## 112 6148  
## 198 10723  
## 160 7633  
## 55 5349  
## 63 7753  
## 66 7861  
## 77 8230  
## 282 6680  
## 83 7542  
## 80 9820  
## 110 6917  
## 76 6278  
## 118 7210  
## 56 7291  
## 56 4743  
## 126 7956  
## 177 6264  
## 34 5158  
## 185 7140  
## 53 5203  
## 115 5652  
## 67 7067  
## 3494 214765

##################################################  
  
  
##################################################  
#CORRELATION  
CorDataFrame <- pivot3 %>%   
 spread(Jurisdiction, sum\_Number)  
  
  
CorDataFrame

## # A tibble: 29 x 14  
## # Groups: Year [29]  
## Year AB BC MB NP NB NL NT NS ON PE QC  
## <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 1990 1296 3255 570 128 377 197 236 498 1614 38 851  
## 2 1991 923 2013 676 56 656 166 331 733 2560 48 1216  
## 3 1992 1055 3805 298 72 576 109 285 299 960 27 765  
## 4 1993 848 1497 239 63 430 83 469 317 743 29 543  
## 5 1994 872 4088 555 166 518 143 627 245 1053 43 499  
## 6 1995 803 1474 660 62 547 103 215 408 2122 29 1265  
## 7 1996 376 1346 424 NA 367 148 350 272 1245 0 1250  
## 8 1997 456 1161 373 55 368 110 105 371 1636 34 876  
## 9 1998 1698 2662 515 NA 286 192 399 348 2278 27 854  
## 10 1999 1355 1214 613 NA 607 228 170 464 1016 34 1037  
## # ... with 19 more rows, and 2 more variables: SK <int>, YT <int>

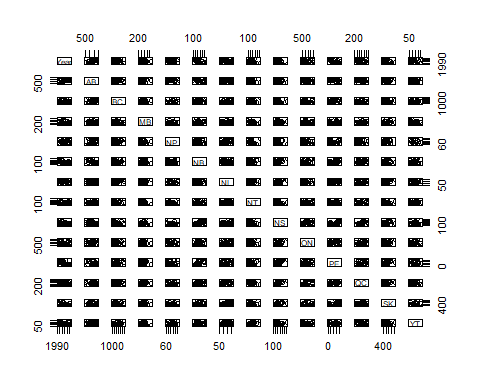
sapply(CorDataFrame, is.numeric) # Which columns are numeric?

## Year AB BC MB NP NB NL NT NS ON PE QC SK YT   
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE

my\_num\_data <- CorDataFrame[, sapply(CorDataFrame, is.numeric)] # Subset numeric columns  
my\_num\_data

## # A tibble: 29 x 14  
## # Groups: Year [29]  
## Year AB BC MB NP NB NL NT NS ON PE QC  
## <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int> <int>  
## 1 1990 1296 3255 570 128 377 197 236 498 1614 38 851  
## 2 1991 923 2013 676 56 656 166 331 733 2560 48 1216  
## 3 1992 1055 3805 298 72 576 109 285 299 960 27 765  
## 4 1993 848 1497 239 63 430 83 469 317 743 29 543  
## 5 1994 872 4088 555 166 518 143 627 245 1053 43 499  
## 6 1995 803 1474 660 62 547 103 215 408 2122 29 1265  
## 7 1996 376 1346 424 NA 367 148 350 272 1245 0 1250  
## 8 1997 456 1161 373 55 368 110 105 371 1636 34 876  
## 9 1998 1698 2662 515 NA 286 192 399 348 2278 27 854  
## 10 1999 1355 1214 613 NA 607 228 170 464 1016 34 1037  
## # ... with 19 more rows, and 2 more variables: SK <int>, YT <int>

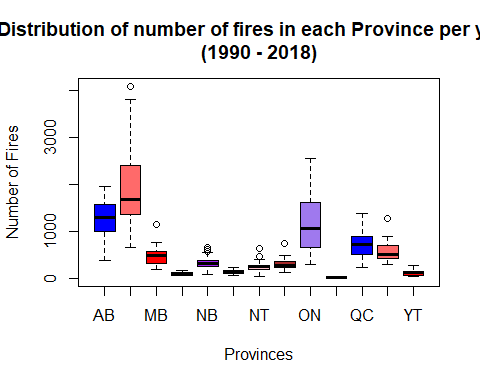
plot(my\_num\_data) # Works



cor(my\_num\_data)

## Year AB BC MB NP NB  
## Year 1.0000000 0.546335923 -0.36110450 -0.185100904 NA -0.7000538  
## AB 0.5463359 1.000000000 0.15023454 -0.007571187 NA -0.5189530  
## BC -0.3611045 0.150234537 1.00000000 0.161751543 NA 0.2114017  
## MB -0.1851009 -0.007571187 0.16175154 1.000000000 NA 0.2038883  
## NP NA NA NA NA 1 NA  
## NB -0.7000538 -0.518953042 0.21140166 0.203888335 NA 1.0000000  
## NL -0.3317171 -0.041129119 0.22401619 0.247988400 NA 0.3037542  
## NT -0.3474347 -0.172973137 0.30415171 -0.120227227 NA 0.2320419  
## NS -0.6080368 -0.232706662 0.01817127 0.286415878 NA 0.6993488  
## ON -0.4188543 -0.148060585 -0.01483963 0.375743689 NA 0.4197618  
## PE NA NA NA NA NA NA  
## QC -0.5246978 -0.367004187 -0.16781797 0.281679330 NA 0.5951643  
## SK -0.5412805 0.062870086 0.41185714 0.418406166 NA 0.3824341  
## YT -0.3859273 -0.076749198 0.44794406 0.015002077 NA 0.3890247  
## NL NT NS ON PE QC  
## Year -0.33171713 -0.347434654 -0.608036773 -0.41885428 NA -0.52469782  
## AB -0.04112912 -0.172973137 -0.232706662 -0.14806059 NA -0.36700419  
## BC 0.22401619 0.304151713 0.018171267 -0.01483963 NA -0.16781797  
## MB 0.24798840 -0.120227227 0.286415878 0.37574369 NA 0.28167933  
## NP NA NA NA NA NA NA  
## NB 0.30375423 0.232041878 0.699348814 0.41976184 NA 0.59516430  
## NL 1.00000000 -0.056363050 0.378076786 0.12265643 NA 0.25919831  
## NT -0.05636305 1.000000000 0.003673412 0.02984490 NA -0.11646900  
## NS 0.37807679 0.003673412 1.000000000 0.58423893 NA 0.67343110  
## ON 0.12265643 0.029844896 0.584238930 1.00000000 NA 0.70744952  
## PE NA NA NA NA 1 NA  
## QC 0.25919831 -0.116469001 0.673431099 0.70744952 NA 1.00000000  
## SK 0.42672134 0.134702176 0.490365233 0.36364534 NA 0.26994050  
## YT 0.20183017 0.487585257 0.253303187 0.09305185 NA 0.03992442  
## SK YT  
## Year -0.54128047 -0.38592735  
## AB 0.06287009 -0.07674920  
## BC 0.41185714 0.44794406  
## MB 0.41840617 0.01500208  
## NP NA NA  
## NB 0.38243408 0.38902468  
## NL 0.42672134 0.20183017  
## NT 0.13470218 0.48758526  
## NS 0.49036523 0.25330319  
## ON 0.36364534 0.09305185  
## PE NA NA  
## QC 0.26994050 0.03992442  
## SK 1.00000000 0.27080987  
## YT 0.27080987 1.00000000

##################################################  
# Test 1  
  
boxplot(sum\_Number~Jurisdiction,  
 data=pivot3,  
 main="Distribution of number of fires in each Province per year \n(1990 - 2018)",  
 xlab="Provinces",  
 ylab="Number of Fires",  
 col=c("blue", "indianred1","red","orange","purple","green",  
 "pink","brown", "mediumpurple2","olivedrab1"),  
 border="black"  
)



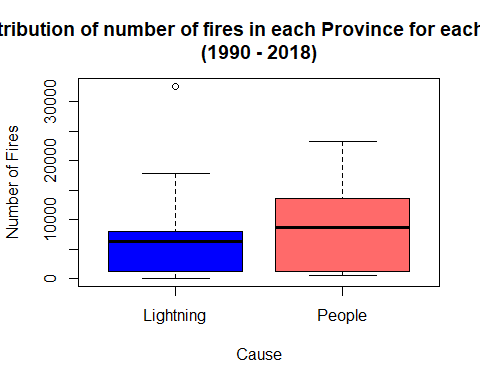
##################################################  
  
########################################################################################################################################  
#5 Boxplot  
  
pivot3 <- data %>%  
select(Jurisdiction, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Jurisdiction Number Cause\_Grouped  
## <fct> <int> <chr>   
## 1 AB 22 People   
## 2 AB 14 People   
## 3 AB 12 People   
## 4 AB 11 People   
## 5 AB 13 People   
## 6 AB 14 People

pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Cause\_Grouped) %>%   
 group\_by(Cause\_Grouped, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)

## # A tibble: 13 x 3  
## Jurisdiction Lightning People  
## <fct> <int> <int>  
## 1 AB 16457 20545  
## 2 BC 32648 23206  
## 3 MB 6838 6831  
## 4 NP 1421 1096  
## 5 NB 1156 8661  
## 6 NL 765 3233  
## 7 NT 6274 944  
## 8 NS 198 8636  
## 9 ON 17782 16489  
## 10 PE 1 543  
## 11 QC 7241 13627  
## 12 SK 7939 8740  
## 13 YT 2275 1219

##################################################  
  
# Test 2  
   
  
boxplot(sum\_Number~Cause\_Grouped,  
 data=pivot3,  
 main="Distribution of number of fires in each Province for each cause \n(1990 - 2018)",  
 xlab="Cause",  
 ylab="Number of Fires",  
 col=c("blue", "indianred1","red","orange","purple","green",  
 "pink","brown", "mediumpurple2","olivedrab1"),  
 border="black"  
)



########################################################################################################################################  
#4  
  
pivot3 <- data %>%  
select(Jurisdiction, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Jurisdiction Number Cause\_Grouped  
## <fct> <int> <chr>   
## 1 AB 22 People   
## 2 AB 14 People   
## 3 AB 12 People   
## 4 AB 11 People   
## 5 AB 13 People   
## 6 AB 14 People

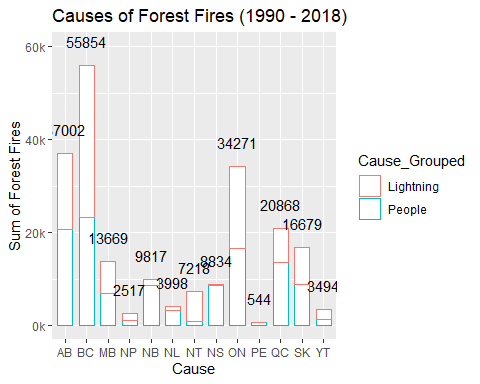
pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Cause\_Grouped) %>%   
 group\_by(Cause\_Grouped, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)

## # A tibble: 13 x 3  
## Jurisdiction Lightning People  
## <fct> <int> <int>  
## 1 AB 16457 20545  
## 2 BC 32648 23206  
## 3 MB 6838 6831  
## 4 NP 1421 1096  
## 5 NB 1156 8661  
## 6 NL 765 3233  
## 7 NT 6274 944  
## 8 NS 198 8636  
## 9 ON 17782 16489  
## 10 PE 1 543  
## 11 QC 7241 13627  
## 12 SK 7939 8740  
## 13 YT 2275 1219

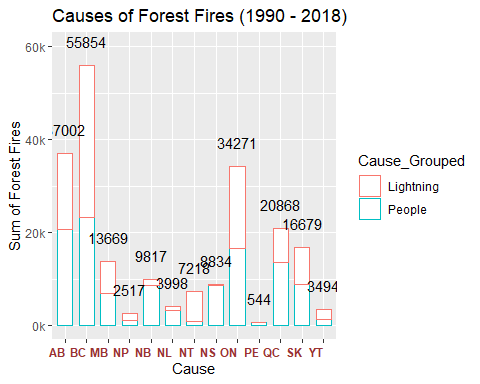
##################################################  
# Test 5  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3, aes(x = Jurisdiction,  
 y = sum\_Number,  
 color = Cause\_Grouped)  
 ) +  
 geom\_bar(stat="identity", width = 0.7, fill="white") +  
 # geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 # vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 stat\_summary(fun.y = sum, aes(label = ..y.., group = Jurisdiction), geom = "text", vjust= -1.5, show\_guide = F)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 60000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")

## Warning: `show\_guide` has been deprecated. Please use `show.legend`  
## instead.

p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 0))



########################################################################################################################################  
#3  
pivot3 <- data %>%  
select(Jurisdiction, Number, Cause\_Grouped)  
head(pivot3)

## # A tibble: 6 x 3  
## Jurisdiction Number Cause\_Grouped  
## <fct> <int> <chr>   
## 1 AB 22 People   
## 2 AB 14 People   
## 3 AB 12 People   
## 4 AB 11 People   
## 5 AB 13 People   
## 6 AB 14 People

# attach(pivot3)  
# detach(pivot3)  
  
pivot3 <- data %>% #Groups Cause together and sums Number  
 select(Jurisdiction, Number, Cause\_Grouped) %>%   
 group\_by(Cause\_Grouped, Jurisdiction) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))   
   
pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)

## # A tibble: 13 x 3  
## Jurisdiction Lightning People  
## <fct> <int> <int>  
## 1 AB 16457 20545  
## 2 BC 32648 23206  
## 3 MB 6838 6831  
## 4 NP 1421 1096  
## 5 NB 1156 8661  
## 6 NL 765 3233  
## 7 NT 6274 944  
## 8 NS 198 8636  
## 9 ON 17782 16489  
## 10 PE 1 543  
## 11 QC 7241 13627  
## 12 SK 7939 8740  
## 13 YT 2275 1219

##################################################  
#CORRELATION  
CorDataFrame <- pivot3 %>%   
 spread(Cause\_Grouped, sum\_Number)  
  
CorDataFrame

## # A tibble: 13 x 3  
## Jurisdiction Lightning People  
## <fct> <int> <int>  
## 1 AB 16457 20545  
## 2 BC 32648 23206  
## 3 MB 6838 6831  
## 4 NP 1421 1096  
## 5 NB 1156 8661  
## 6 NL 765 3233  
## 7 NT 6274 944  
## 8 NS 198 8636  
## 9 ON 17782 16489  
## 10 PE 1 543  
## 11 QC 7241 13627  
## 12 SK 7939 8740  
## 13 YT 2275 1219

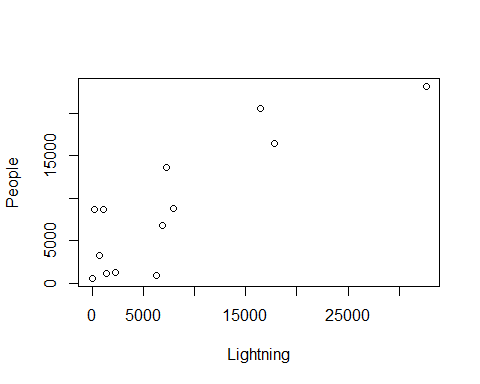
sapply(CorDataFrame, is.numeric) # Which columns are numeric?

## Jurisdiction Lightning People   
## FALSE TRUE TRUE

my\_num\_data <- CorDataFrame[, sapply(CorDataFrame, is.numeric)] # Subset numeric columns  
my\_num\_data

## # A tibble: 13 x 2  
## Lightning People  
## <int> <int>  
## 1 16457 20545  
## 2 32648 23206  
## 3 6838 6831  
## 4 1421 1096  
## 5 1156 8661  
## 6 765 3233  
## 7 6274 944  
## 8 198 8636  
## 9 17782 16489  
## 10 1 543  
## 11 7241 13627  
## 12 7939 8740  
## 13 2275 1219

plot(my\_num\_data) # Works



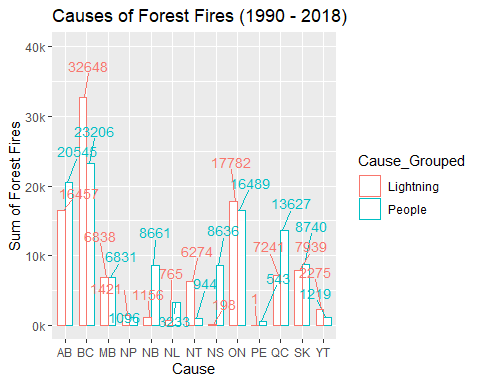
cor(my\_num\_data)

## Lightning People  
## Lightning 1.0000000 0.8506535  
## People 0.8506535 1.0000000

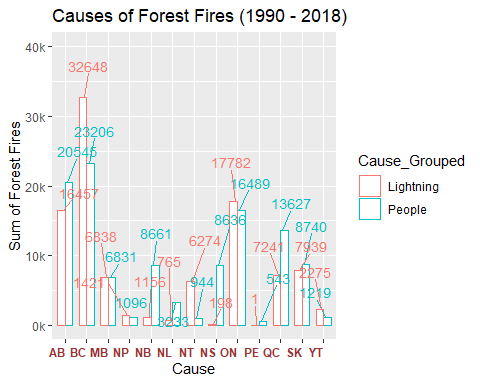
##################################################  
  
# Test 11  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
  
p <- ggplot(data = pivot3,  
 aes(x = Jurisdiction, y = sum\_Number, color=Cause\_Grouped)) +  
 geom\_bar(position = "dodge", stat="identity", width = 0.7, fill="white") +  
 geom\_text\_repel(aes(label=sum\_Number), show\_guide = F, position=position\_dodge(width=0.4),  
 vjust= -2.4, hjust = 0.4, size = 3.8, angle = 0)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 40000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")

## Warning: `show\_guide` has been deprecated. Please use `show.legend`  
## instead.

p



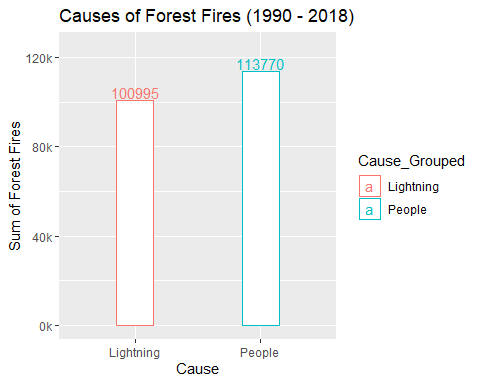
p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 0))



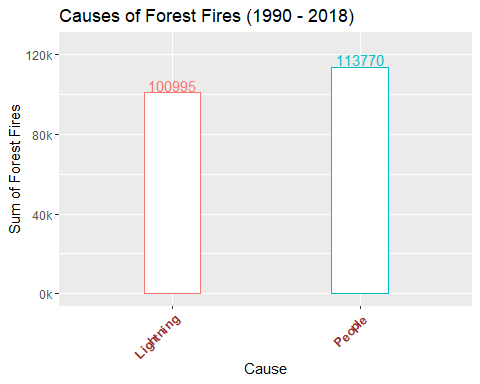
########################################################################################################################################  
#2  
#Below works with percentage totals  
GCause\_HvsL <- data %>% #Groups Cause together and sums Number  
 group\_by(Cause\_Grouped) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE)) %>%   
 mutate(rel.freq = paste0(round(100 \* sum\_Number/sum(sum\_Number),0),"%"))  
  
GCause\_HvsL

## # A tibble: 2 x 3  
## Cause\_Grouped sum\_Number rel.freq  
## <chr> <int> <chr>   
## 1 Lightning 100995 47%   
## 2 People 113770 53%

##################################################  
  
# Boxplot GCause 11, testing  
   
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
p <- ggplot(data = GCause\_HvsL,  
 aes(x = Cause\_Grouped, y = sum\_Number, color=Cause\_Grouped)) +  
 geom\_bar(stat="identity", width = 0.3, fill="white") +  
 geom\_text(aes(label=sum\_Number), position=position\_dodge(width=0.9), vjust=-0.15)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 125000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")  
  
p



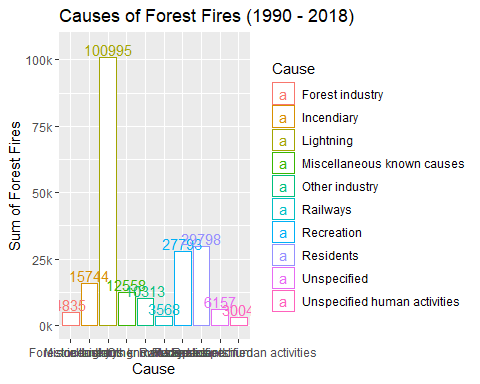
p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 45), legend.position = "none")



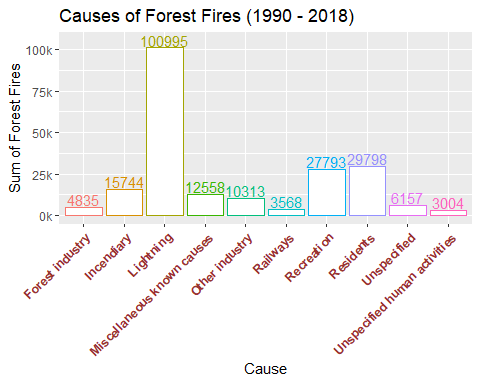
########################################################################################################################################  
#1  
#Below works with percentage totals  
GCause <- data %>% #Groups Cause together and sums Number  
 group\_by(Cause) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE)) %>%   
 mutate(rel.freq = paste0(round(100 \* sum\_Number/sum(sum\_Number),0),"%"))  
  
GCause

## # A tibble: 10 x 3  
## Cause sum\_Number rel.freq  
## <fct> <int> <chr>   
## 1 Forest industry 4835 2%   
## 2 Incendiary 15744 7%   
## 3 Lightning 100995 47%   
## 4 Miscellaneous known causes 12558 6%   
## 5 Other industry 10313 5%   
## 6 Railways 3568 2%   
## 7 Recreation 27793 13%   
## 8 Residents 29798 14%   
## 9 Unspecified 6157 3%   
## 10 Unspecified human activities 3004 1%

##################################################  
  
ks <- function (x) { number\_format(accuracy = 1,  
 scale = 1/1000,  
 suffix = "k",  
 big.mark = ",")(x) }  
  
p <- ggplot(data = GCause,  
 aes(x = Cause, y = sum\_Number, color=Cause)) +  
 geom\_bar(stat="identity", fill="white") +  
 geom\_text(aes(label=sum\_Number), position=position\_dodge(width=0.9), vjust=-0.15)+  
 xlab("Cause") +  
 ylab("Causes of Forest Fires (1990 - 2018)") +  
 scale\_y\_continuous(name="Sum of Forest Fires", labels = ks)+  
 coord\_cartesian(ylim = c(0, 105000))+  
 labs(title = "Causes of Forest Fires (1990 - 2018)")  
  
p



p + theme(  
 axis.text.x = element\_text(face = "bold", color = "#993333", hjust = 1,size = 9, angle = 45), legend.position = "none")



########################################################################################################################################  
  
#Sample code for various statistics  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Cause, Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 262 x 3  
## # Groups: Cause [10]  
## Cause Year sum\_Number  
## <fct> <int> <int>  
## 1 Forest industry 1990 355  
## 2 Forest industry 1991 299  
## 3 Forest industry 1992 258  
## 4 Forest industry 1993 240  
## 5 Forest industry 1994 222  
## 6 Forest industry 1995 309  
## 7 Forest industry 1996 282  
## 8 Forest industry 1997 276  
## 9 Forest industry 1998 151  
## 10 Forest industry 1999 239  
## # ... with 252 more rows

#Below code works  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Cause, Jurisdiction, Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 3,219 x 4  
## # Groups: Cause, Jurisdiction [129]  
## Cause Jurisdiction Year sum\_Number  
## <fct> <fct> <int> <int>  
## 1 Forest industry AB 1990 22  
## 2 Forest industry AB 1991 14  
## 3 Forest industry AB 1992 12  
## 4 Forest industry AB 1993 11  
## 5 Forest industry AB 1994 13  
## 6 Forest industry AB 1995 14  
## 7 Forest industry AB 1996 8  
## 8 Forest industry AB 1997 29  
## 9 Forest industry AB 1998 10  
## 10 Forest industry AB 1999 20  
## # ... with 3,209 more rows

#Below code works  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Jurisdiction, Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 372 x 3  
## # Groups: Jurisdiction [13]  
## Jurisdiction Year sum\_Number  
## <fct> <int> <int>  
## 1 AB 1990 1296  
## 2 AB 1991 923  
## 3 AB 1992 1055  
## 4 AB 1993 848  
## 5 AB 1994 872  
## 6 AB 1995 803  
## 7 AB 1996 376  
## 8 AB 1997 456  
## 9 AB 1998 1698  
## 10 AB 1999 1355  
## # ... with 362 more rows

#Below code works  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Year, Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 372 x 3  
## # Groups: Year [29]  
## Year Jurisdiction sum\_Number  
## <int> <fct> <int>  
## 1 1990 AB 1296  
## 2 1990 BC 3255  
## 3 1990 MB 570  
## 4 1990 NP 128  
## 5 1990 NB 377  
## 6 1990 NL 197  
## 7 1990 NT 236  
## 8 1990 NS 498  
## 9 1990 ON 1614  
## 10 1990 PE 38  
## # ... with 362 more rows

#Below code works  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 13 x 2  
## Jurisdiction sum\_Number  
## <fct> <int>  
## 1 AB 37002  
## 2 BC 55854  
## 3 MB 13669  
## 4 NP 2517  
## 5 NB 9817  
## 6 NL 3998  
## 7 NT 7218  
## 8 NS 8834  
## 9 ON 34271  
## 10 PE 544  
## 11 QC 20868  
## 12 SK 16679  
## 13 YT 3494

#Below code works  
data %>% #Groups Cause, Year together and sums Number  
 group\_by(Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 13 x 2  
## Jurisdiction sum\_Number  
## <fct> <int>  
## 1 AB 37002  
## 2 BC 55854  
## 3 MB 13669  
## 4 NP 2517  
## 5 NB 9817  
## 6 NL 3998  
## 7 NT 7218  
## 8 NS 8834  
## 9 ON 34271  
## 10 PE 544  
## 11 QC 20868  
## 12 SK 16679  
## 13 YT 3494

#######################################################  
#Below code works  
data %>% #Groups Cause, Year, Jurisdiction and sums Number  
 group\_by(Cause, Year, Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE)) %>%   
 filter(Jurisdiction == "BC") %>%  
 filter(Cause == "Lightning")

## # A tibble: 29 x 4  
## # Groups: Cause, Year [38]  
## Cause Year Jurisdiction sum\_Number  
## <fct> <int> <fct> <int>  
## 1 Lightning 1990 BC 2015  
## 2 Lightning 1991 BC 759  
## 3 Lightning 1992 BC 2344  
## 4 Lightning 1993 BC 609  
## 5 Lightning 1994 BC 2913  
## 6 Lightning 1995 BC 342  
## 7 Lightning 1996 BC 723  
## 8 Lightning 1997 BC 675  
## 9 Lightning 1998 BC 1773  
## 10 Lightning 1999 BC 585  
## # ... with 19 more rows

#Below code works  
data %>% #Groups Cause together and sums Number  
 group\_by(Cause, Year, Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE)) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning")

## # A tibble: 29 x 4  
## # Groups: Cause, Year [38]  
## Cause Year Jurisdiction sum\_Number  
## <fct> <int> <fct> <int>  
## 1 Lightning 1990 AB 914  
## 2 Lightning 1991 AB 466  
## 3 Lightning 1992 AB 626  
## 4 Lightning 1993 AB 517  
## 5 Lightning 1994 AB 499  
## 6 Lightning 1995 AB 357  
## 7 Lightning 1996 AB 217  
## 8 Lightning 1997 AB 242  
## 9 Lightning 1998 AB 1192  
## 10 Lightning 1999 AB 890  
## # ... with 19 more rows

#Below works great, show max number of fires and year for Alberta  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 top\_n(n=1) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning")

## Selecting by max\_Number

## # A tibble: 1 x 4  
## # Groups: Jurisdiction, Cause [130]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 1998 1192

#Below works great, Top 10 years for Lightning in Alberta  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 top\_n(n=10) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning") %>%   
 arrange(desc(max\_Number))

## Selecting by max\_Number

## # A tibble: 10 x 4  
## # Groups: Jurisdiction, Cause [1]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 1998 1192  
## 2 AB Lightning 1990 914  
## 3 AB Lightning 1999 890  
## 4 AB Lightning 2002 868  
## 5 AB Lightning 2008 779  
## 6 AB Lightning 2015 772  
## 7 AB Lightning 2006 746  
## 8 AB Lightning 2010 741  
## 9 AB Lightning 2004 732  
## 10 AB Lightning 1992 626

#Below works great, Top 10 years for Lightning in Alberta  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 top\_n(n=10) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning")

## Selecting by max\_Number

## # A tibble: 10 x 4  
## # Groups: Jurisdiction, Cause [130]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 1990 914  
## 2 AB Lightning 1992 626  
## 3 AB Lightning 1998 1192  
## 4 AB Lightning 1999 890  
## 5 AB Lightning 2002 868  
## 6 AB Lightning 2004 732  
## 7 AB Lightning 2006 746  
## 8 AB Lightning 2008 779  
## 9 AB Lightning 2010 741  
## 10 AB Lightning 2015 772

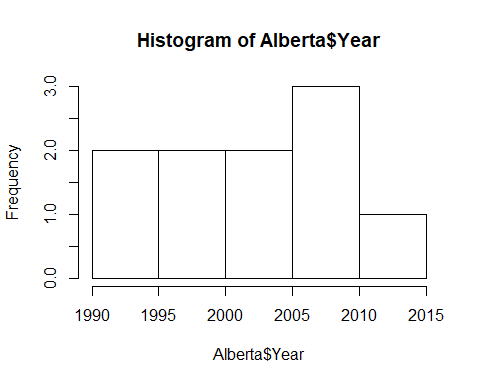
#Below works great, Top 10 years for Lightning in Alberta  
Alberta <- data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 top\_n(n=10) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning")

## Selecting by max\_Number

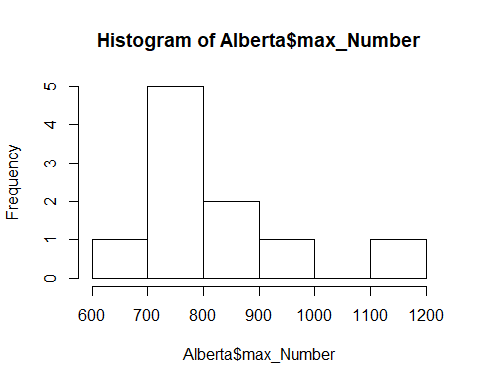
Alberta

## # A tibble: 10 x 4  
## # Groups: Jurisdiction, Cause [130]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 1990 914  
## 2 AB Lightning 1992 626  
## 3 AB Lightning 1998 1192  
## 4 AB Lightning 1999 890  
## 5 AB Lightning 2002 868  
## 6 AB Lightning 2004 732  
## 7 AB Lightning 2006 746  
## 8 AB Lightning 2008 779  
## 9 AB Lightning 2010 741  
## 10 AB Lightning 2015 772

hist(Alberta$Year)



hist(Alberta$max\_Number)



#Below works  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning") %>%   
 filter(Year > "2003")

## # A tibble: 15 x 4  
## # Groups: Jurisdiction, Cause [130]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 2004 732  
## 2 AB Lightning 2005 432  
## 3 AB Lightning 2006 746  
## 4 AB Lightning 2007 513  
## 5 AB Lightning 2008 779  
## 6 AB Lightning 2009 566  
## 7 AB Lightning 2010 741  
## 8 AB Lightning 2011 215  
## 9 AB Lightning 2012 436  
## 10 AB Lightning 2013 287  
## 11 AB Lightning 2014 563  
## 12 AB Lightning 2015 772  
## 13 AB Lightning 2016 514  
## 14 AB Lightning 2017 416  
## 15 AB Lightning 2018 511

#Below works, Top 10 after 2003  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning") %>%   
 filter(Year > "2003") %>%   
 arrange(desc(max\_Number))

## # A tibble: 15 x 4  
## # Groups: Jurisdiction, Cause [1]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 2008 779  
## 2 AB Lightning 2015 772  
## 3 AB Lightning 2006 746  
## 4 AB Lightning 2010 741  
## 5 AB Lightning 2004 732  
## 6 AB Lightning 2009 566  
## 7 AB Lightning 2014 563  
## 8 AB Lightning 2016 514  
## 9 AB Lightning 2007 513  
## 10 AB Lightning 2018 511  
## 11 AB Lightning 2012 436  
## 12 AB Lightning 2005 432  
## 13 AB Lightning 2017 416  
## 14 AB Lightning 2013 287  
## 15 AB Lightning 2011 215

data %>%   
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 3,219 x 4  
## # Groups: Jurisdiction, Cause [129]  
## Jurisdiction Cause Year sum\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Forest industry 1990 22  
## 2 AB Forest industry 1991 14  
## 3 AB Forest industry 1992 12  
## 4 AB Forest industry 1993 11  
## 5 AB Forest industry 1994 13  
## 6 AB Forest industry 1995 14  
## 7 AB Forest industry 1996 8  
## 8 AB Forest industry 1997 29  
## 9 AB Forest industry 1998 10  
## 10 AB Forest industry 1999 20  
## # ... with 3,209 more rows

#Below works, Top 10 after 2003  
data %>% #Groups Cause together and sums Number  
 group\_by(Jurisdiction, Cause, Year) %>%   
 summarize(max\_Number = max(Number, na.rm = TRUE)) %>%   
 filter(Jurisdiction == "AB") %>%  
 filter(Cause == "Lightning") %>%   
 filter(Year > "2003") %>%   
 arrange(desc(max\_Number))

## # A tibble: 15 x 4  
## # Groups: Jurisdiction, Cause [1]  
## Jurisdiction Cause Year max\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Lightning 2008 779  
## 2 AB Lightning 2015 772  
## 3 AB Lightning 2006 746  
## 4 AB Lightning 2010 741  
## 5 AB Lightning 2004 732  
## 6 AB Lightning 2009 566  
## 7 AB Lightning 2014 563  
## 8 AB Lightning 2016 514  
## 9 AB Lightning 2007 513  
## 10 AB Lightning 2018 511  
## 11 AB Lightning 2012 436  
## 12 AB Lightning 2005 432  
## 13 AB Lightning 2017 416  
## 14 AB Lightning 2013 287  
## 15 AB Lightning 2011 215

################################################  
  
data %>% #Groups Cause together and sums Number. Also provides avg for Number of fires.  
 group\_by(Cause) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE),  
 count = n(),  
 average\_fire = mean(Number, na.rm = TRUE))

## # A tibble: 10 x 4  
## Cause sum\_Number count average\_fire  
## <fct> <int> <int> <dbl>  
## 1 Forest industry 4835 1262 3.83  
## 2 Incendiary 15744 1271 12.4   
## 3 Lightning 100995 1325 76.2   
## 4 Miscellaneous known causes 12558 1279 9.82  
## 5 Other industry 10313 1261 8.18  
## 6 Railways 3568 1265 2.82  
## 7 Recreation 27793 1274 21.8   
## 8 Residents 29798 1268 23.5   
## 9 Unspecified 6157 1290 4.77  
## 10 Unspecified human activities 3004 24 125.

data %>% #Groups Cause together and sums Number. Also provides avg for Number of fires.  
 group\_by(Cause) %>%   
 summarize(  
 sum\_Number = sum(Number, na.rm = TRUE),  
 count = n(),  
 average\_fire = mean(Number, na.rm = TRUE),  
 total = sum(sum\_Number)  
 )

## # A tibble: 10 x 5  
## Cause sum\_Number count average\_fire total  
## <fct> <int> <int> <dbl> <int>  
## 1 Forest industry 4835 1262 3.83 4835  
## 2 Incendiary 15744 1271 12.4 15744  
## 3 Lightning 100995 1325 76.2 100995  
## 4 Miscellaneous known causes 12558 1279 9.82 12558  
## 5 Other industry 10313 1261 8.18 10313  
## 6 Railways 3568 1265 2.82 3568  
## 7 Recreation 27793 1274 21.8 27793  
## 8 Residents 29798 1268 23.5 29798  
## 9 Unspecified 6157 1290 4.77 6157  
## 10 Unspecified human activities 3004 24 125. 3004

data %>% #Groups Year and sums Number  
 group\_by(Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 29 x 2  
## Year sum\_Number  
## <int> <int>  
## 1 1990 10111  
## 2 1991 10327  
## 3 1992 9068  
## 4 1993 6043  
## 5 1994 9763  
## 6 1995 8486  
## 7 1996 6349  
## 8 1997 6148  
## 9 1998 10723  
## 10 1999 7633  
## # ... with 19 more rows

data %>% #Groups Jurisdiction and sums Number  
 group\_by(Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 13 x 2  
## Jurisdiction sum\_Number  
## <fct> <int>  
## 1 AB 37002  
## 2 BC 55854  
## 3 MB 13669  
## 4 NP 2517  
## 5 NB 9817  
## 6 NL 3998  
## 7 NT 7218  
## 8 NS 8834  
## 9 ON 34271  
## 10 PE 544  
## 11 QC 20868  
## 12 SK 16679  
## 13 YT 3494

data %>%   
 group\_by(Year) %>%  
 group\_by(Jurisdiction) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 13 x 2  
## Jurisdiction sum\_Number  
## <fct> <int>  
## 1 AB 37002  
## 2 BC 55854  
## 3 MB 13669  
## 4 NP 2517  
## 5 NB 9817  
## 6 NL 3998  
## 7 NT 7218  
## 8 NS 8834  
## 9 ON 34271  
## 10 PE 544  
## 11 QC 20868  
## 12 SK 16679  
## 13 YT 3494

data %>%   
 group\_by(Jurisdiction) %>%  
 group\_by(Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 29 x 2  
## Year sum\_Number  
## <int> <int>  
## 1 1990 10111  
## 2 1991 10327  
## 3 1992 9068  
## 4 1993 6043  
## 5 1994 9763  
## 6 1995 8486  
## 7 1996 6349  
## 8 1997 6148  
## 9 1998 10723  
## 10 1999 7633  
## # ... with 19 more rows

data %>%   
 group\_by(Jurisdiction) %>%  
 group\_by(Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 29 x 2  
## Year sum\_Number  
## <int> <int>  
## 1 1990 10111  
## 2 1991 10327  
## 3 1992 9068  
## 4 1993 6043  
## 5 1994 9763  
## 6 1995 8486  
## 7 1996 6349  
## 8 1997 6148  
## 9 1998 10723  
## 10 1999 7633  
## # ... with 19 more rows

data %>%   
 group\_by(Jurisdiction) %>%  
 group\_by(Year) %>%   
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 29 x 2  
## Year sum\_Number  
## <int> <int>  
## 1 1990 10111  
## 2 1991 10327  
## 3 1992 9068  
## 4 1993 6043  
## 5 1994 9763  
## 6 1995 8486  
## 7 1996 6349  
## 8 1997 6148  
## 9 1998 10723  
## 10 1999 7633  
## # ... with 19 more rows

data %>%   
 group\_by(Jurisdiction, Cause, Year) %>%  
 summarize(sum\_Number = sum(Number, na.rm = TRUE))

## # A tibble: 3,219 x 4  
## # Groups: Jurisdiction, Cause [129]  
## Jurisdiction Cause Year sum\_Number  
## <fct> <fct> <int> <int>  
## 1 AB Forest industry 1990 22  
## 2 AB Forest industry 1991 14  
## 3 AB Forest industry 1992 12  
## 4 AB Forest industry 1993 11  
## 5 AB Forest industry 1994 13  
## 6 AB Forest industry 1995 14  
## 7 AB Forest industry 1996 8  
## 8 AB Forest industry 1997 29  
## 9 AB Forest industry 1998 10  
## 10 AB Forest industry 1999 20  
## # ... with 3,209 more rows

###################

.Misc code to be entered.