main.R

enzoz

Sun May 07 23:38:17 2017

#Zheng Zhou  
#MA415  
#Final Project Code  
  
## Question: Is there a significant trend in Residential Fire Fatalities from 2006 to 2015?  
  
## Explainations and Results are at the end of this document.  
  
## This script will run for about 5 to 10 minutes.  
## If there are zipped folders under the working directory, unzip to their current folder before running this script.  
  
options(useFancyQuotes = FALSE)  
  
#Load packages  
library(foreign)  
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.3.3

##   
## 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

library(tidyr)

## Warning: package 'tidyr' was built under R version 3.3.3

library(data.table)

## -------------------------------------------------------------------------

## data.table + dplyr code now lives in dtplyr.  
## Please library(dtplyr)!

## -------------------------------------------------------------------------

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.3

library(stringr)

## Warning: package 'stringr' was built under R version 3.3.3

library(magrittr)

## Warning: package 'magrittr' was built under R version 3.3.3

##   
## Attaching package: 'magrittr'

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

library(lubridate)

## Warning: package 'lubridate' was built under R version 3.3.3

##   
## Attaching package: 'lubridate'

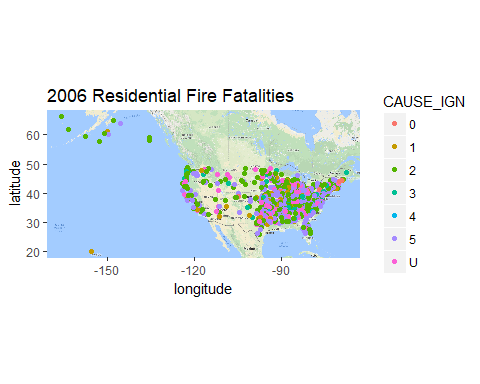
## The following objects are masked from 'package:data.table':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday,  
## week, yday, year

## The following object is masked from 'package:base':  
##   
## date

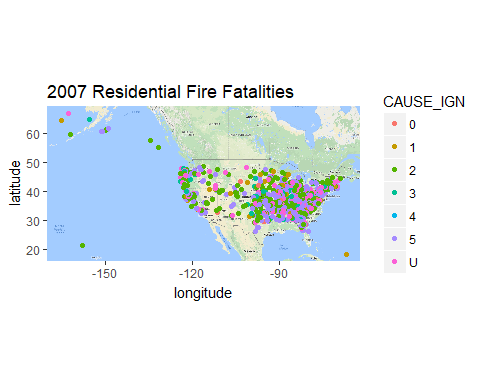
library(zipcode)  
library(png)  
  
#Read "Civilian Casualty" data of 2006 to 2015. (ps: as the diagram suggests, civilian casuality only happens when there is a fire incident.)  
civiliancasualty2006 <- read.dbf("data/2006/civiliancasualty.dbf")  
civiliancasualty2007 <- read.dbf("data/2007/civiliancasualty.dbf")  
civiliancasualty2008 <- read.dbf("data/2008/civiliancasualty.dbf")  
civiliancasualty2009 <- read.dbf("data/2009/civiliancasualty.dbf")  
civiliancasualty2010 <- read.dbf("data/2010/civiliancasualty.dbf")  
civiliancasualty2011 <- read.dbf("data/2011/civiliancasualty.dbf")  
civiliancasualty2012 <- read.dbf("data/2012/civiliancasualty.dbf")  
civiliancasualty2013 <- read.dbf("data/2013/civiliancasualty.dbf")  
civiliancasualty2014 <- read.dbf("data/2014/civiliancasualty.dbf")  
civiliancasualty2015 <- read.dbf("data/2015/civiliancasualty.dbf")  
  
#Read "Fire Incident" data of 2006 to 2015.  
fireincident2006 <- read.dbf("data/2006/fireincident.dbf")  
fireincident2007 <- read.dbf("data/2007/fireincident.dbf")  
fireincident2008 <- read.dbf("data/2008/fireincident.dbf")  
fireincident2009 <- read.dbf("data/2009/fireincident.dbf")  
fireincident2010 <- read.dbf("data/2010/fireincident.dbf")  
fireincident2011 <- read.dbf("data/2011/fireincident.dbf")  
fireincident2012 <- read.dbf("data/2012/fireincident.dbf")  
fireincident2013 <- read.dbf("data/2013/fireincident.dbf")  
fireincident2014 <- read.dbf("data/2014/fireincident.dbf")  
fireincident2015 <- read.dbf("data/2015/fireincident.dbf")  
  
#Read "Incident Address" data of 2006 to 2015.  
incidentaddress2006 <- read.dbf("data/2006/incidentaddress.dbf")  
incidentaddress2007 <- read.dbf("data/2007/incidentaddress.dbf")  
incidentaddress2008 <- read.dbf("data/2008/incidentaddress.dbf")  
incidentaddress2009 <- read.dbf("data/2009/incidentaddress.dbf")  
incidentaddress2010 <- read.dbf("data/2010/incidentaddress.dbf")  
incidentaddress2011 <- read.dbf("data/2011/incidentaddress.dbf")  
incidentaddress2012 <- read.csv2("data/2012/incidentaddress.txt",sep = "^",stringsAsFactors = F)  
incidentaddress2013 <- read.csv2("data/2013/incidentaddress.txt",sep = "^",stringsAsFactors = F)

## Warning in scan(file = file, what = what, sep = sep, quote = quote, dec =  
## dec, : embedded nul(s) found in input

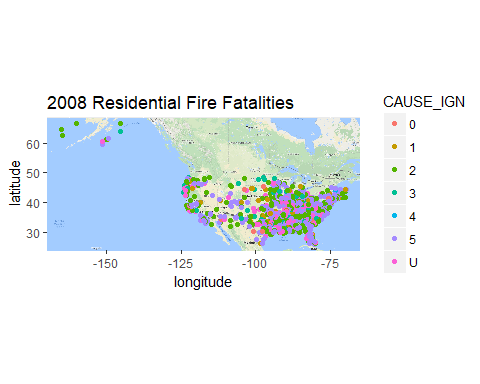
incidentaddress2014 <- read.csv2("data/2014/incidentaddress.txt",sep = "^",stringsAsFactors = F)  
incidentaddress2015 <- read.csv2("data/2015/incidentaddress.txt",sep = "^",stringsAsFactors = F)  
  
  
##########After loading data, Clean the data of year 2006  
  
 #1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2006 <- subset(civiliancasualty2006, grepl("5",civiliancasualty2006$SEV))  
 # There are 2079 civilian fatalities from a total of 11879 casualities.  
  
 #2. find out which fire are residential  
residentialfireincident2006 <- subset(fireincident2006,grepl("N",fireincident2006$NOT\_RES))  
 # There are 281695 residential fire incidents from a total of 636062 fire incident.  
  
 #3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2006 <- merge(civilianfatality2006,residentialfireincident2006,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1559 residential fire fatalities of year 2006.  
  
 #4. merge cfrfrfi2006 with incidentaddress2006  
cfrfrfi2006address <- merge(cfrfrfi2006,incidentaddress2006,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
zipcode\_1 <- data(zipcode)  
  
clean2006 <- data.frame(STATE = cfrfrfi2006address$STATE ,FDID = cfrfrfi2006address$FDID, INC\_DATE = cfrfrfi2006address$INC\_DATE,INC\_NO = cfrfrfi2006address$INC\_NO,EXP\_NO = cfrfrfi2006address$EXP\_NO, zip =cfrfrfi2006address$ZIP5,CAUSE\_INJ = cfrfrfi2006address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2006address$CAUSE\_IGN )  
  
clean2006location <- merge(clean2006,zipcode,by = c("zip"))  
  
  
usa <- readPNG("usa.png")  
  
## plot 2006 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2006location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2006 Residential Fire Fatalities")



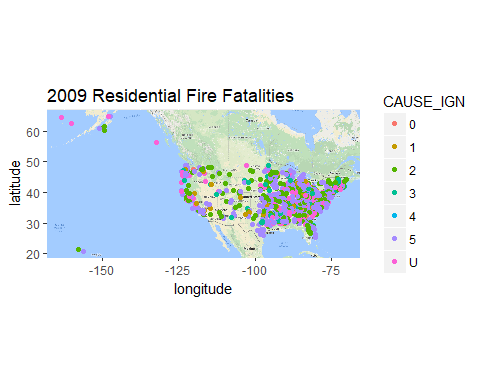
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
  
  
# Follow the same precedures of year 2006 with year 2007 ~ 2015.  
  
########## Clean the data of year 2007  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2007 <- subset(civiliancasualty2007, grepl("5",civiliancasualty2007$SEV))  
# There are 2027 civilian fatalities from a total of 11890 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2007 <- subset(fireincident2007,grepl("N",fireincident2007$NOT\_RES))  
# There are 284181 residential fire incidents from a total of 648495 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2007 <- merge(civilianfatality2007,residentialfireincident2007,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1583 residential fire fatalities of year 2007.  
  
#4. merge cfrfrfi2007 with incidentaddress2007  
cfrfrfi2007address <- merge(cfrfrfi2007,incidentaddress2007,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
  
clean2007 <- data.frame(STATE = cfrfrfi2007address$STATE ,FDID = cfrfrfi2007address$FDID, INC\_DATE = cfrfrfi2007address$INC\_DATE,INC\_NO = cfrfrfi2007address$INC\_NO,EXP\_NO = cfrfrfi2007address$EXP\_NO, zip =cfrfrfi2007address$ZIP5,CAUSE\_INJ = cfrfrfi2007address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2007address$CAUSE\_IGN)  
  
clean2007location <- merge(clean2007,zipcode,by = c("zip"))  
  
## plot 2007 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2007location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2007 Residential Fire Fatalities")



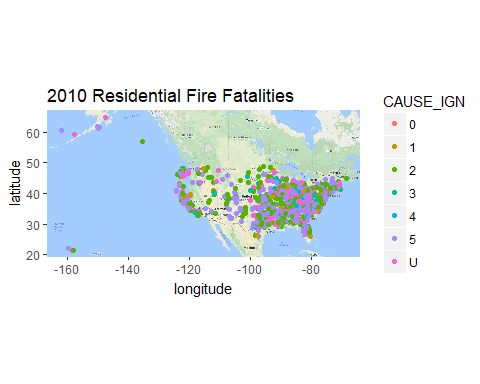
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
########## Clean the data of year 2008  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2008 <- subset(civiliancasualty2008, grepl("5",civiliancasualty2008$SEV))  
# There are 2087 civilian fatalities from a total of 11930 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2008 <- subset(fireincident2008,grepl("N",fireincident2008$NOT\_RES))  
# There are 293745 residential fire incidents from a total of 637653 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2008 <- merge(civilianfatality2008,residentialfireincident2008,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1617 residential fire fatalities of year 2008.  
  
#4. merge cfrfrfi2008 with incidentaddress2008  
cfrfrfi2008address <- merge(cfrfrfi2008,incidentaddress2008,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
  
clean2008 <- data.frame(STATE = cfrfrfi2008address$STATE ,FDID = cfrfrfi2008address$FDID, INC\_DATE = cfrfrfi2008address$INC\_DATE,INC\_NO = cfrfrfi2008address$INC\_NO,EXP\_NO = cfrfrfi2008address$EXP\_NO, zip =cfrfrfi2008address$ZIP5,CAUSE\_INJ = cfrfrfi2008address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2008address$CAUSE\_IGN)  
  
clean2008location <- merge(clean2008,zipcode,by = c("zip"))  
  
## plot 2008 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2008location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2008 Residential Fire Fatalities")



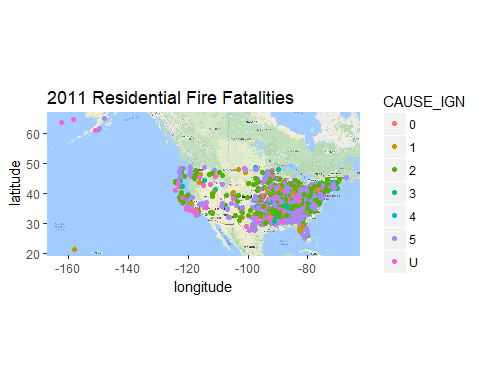
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
########## Clean the data of year 2009  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2009 <- subset(civiliancasualty2009, grepl("5",civiliancasualty2009$SEV))  
# There are 1931 civilian fatalities from a total of 11601 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2009 <- subset(fireincident2009,grepl("N",fireincident2009$NOT\_RES))  
# There are 288145 residential fire incidents from a total of 606409 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2009 <- merge(civilianfatality2009,residentialfireincident2009,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1577 residential fire fatalities of year 2009.  
  
#4. merge cfrfrfi2009 with incidentaddress2009  
cfrfrfi2009address <- merge(cfrfrfi2009,incidentaddress2009,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
  
clean2009 <- data.frame(STATE = cfrfrfi2009address$STATE ,FDID = cfrfrfi2009address$FDID, INC\_DATE = cfrfrfi2009address$INC\_DATE,INC\_NO = cfrfrfi2009address$INC\_NO,EXP\_NO = cfrfrfi2009address$EXP\_NO, zip =cfrfrfi2009address$ZIP5,CAUSE\_INJ = cfrfrfi2009address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2009address$CAUSE\_IGN)  
  
clean2009location <- merge(clean2009,zipcode,by = c("zip"))  
  
## plot 2009 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2009location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2009 Residential Fire Fatalities")



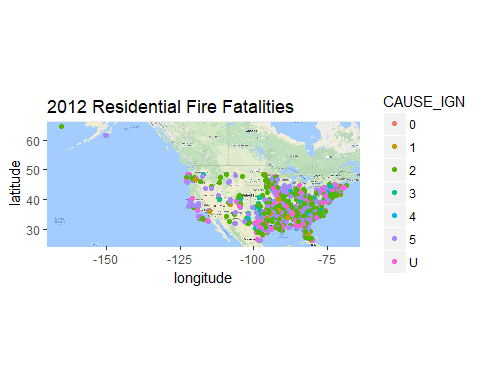
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
########## Clean the data of year 2010  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2010 <- subset(civiliancasualty2010, grepl("5",civiliancasualty2010$SEV))  
# There are 2068 civilian fatalities from a total of 12535 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2010 <- subset(fireincident2010,grepl("N",fireincident2010$NOT\_RES))  
# There are 317074 residential fire incidents from a total of 663333 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2010 <- merge(civilianfatality2010,residentialfireincident2010,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1651 residential fire fatalities of year 2010.  
  
#4. merge cfrfrfi2010 with incidentaddress2010  
cfrfrfi2010address <- merge(cfrfrfi2010,incidentaddress2010,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
  
clean2010 <- data.frame(STATE = cfrfrfi2010address$STATE ,FDID = cfrfrfi2010address$FDID, INC\_DATE = cfrfrfi2010address$INC\_DATE,INC\_NO = cfrfrfi2010address$INC\_NO,EXP\_NO = cfrfrfi2010address$EXP\_NO, zip =cfrfrfi2010address$ZIP5,CAUSE\_INJ = cfrfrfi2010address$CAUSE\_INJ, CAUSE\_IGN = cfrfrfi2010address$CAUSE\_IGN)  
  
clean2010location <- merge(clean2010,zipcode,by = c("zip"))  
  
## plot 2010 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2010location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2010 Residential Fire Fatalities")



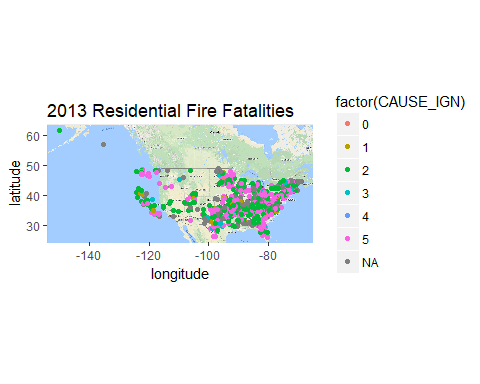
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
########## Clean the data of year 2011  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2011 <- subset(civiliancasualty2011, grepl("5",civiliancasualty2011$SEV))  
# There are 2087 civilian fatalities from a total of 12611 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2011 <- subset(fireincident2011,grepl("N",fireincident2011$NOT\_RES))  
# There are 321499 residential fire incidents from a total of 671329 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2011 <- merge(civilianfatality2011,residentialfireincident2011,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1659 residential fire fatalities of year 2011.  
  
#4. merge cfrfrfi2011 with incidentaddress2011  
cfrfrfi2011address <- merge(cfrfrfi2011,incidentaddress2011,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
  
clean2011 <- data.frame(STATE = cfrfrfi2011address$STATE ,FDID = cfrfrfi2011address$FDID, INC\_DATE = cfrfrfi2011address$INC\_DATE,INC\_NO = cfrfrfi2011address$INC\_NO,EXP\_NO = cfrfrfi2011address$EXP\_NO, zip =cfrfrfi2011address$ZIP5,CAUSE\_INJ = cfrfrfi2011address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2011address$CAUSE\_IGN)  
  
clean2011location <- merge(clean2011,zipcode,by = c("zip"))  
  
## plot 2011 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2011location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2011 Residential Fire Fatalities")



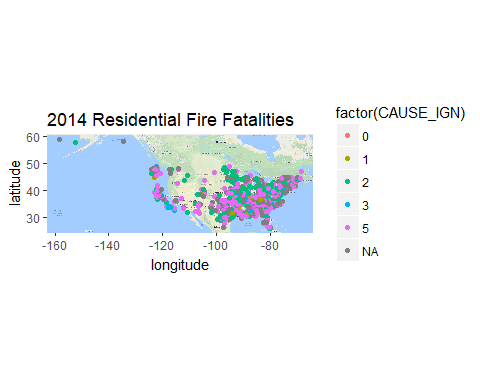
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
  
########## Clean the data of year 2012  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2012 <- subset(civiliancasualty2012, grepl("5",civiliancasualty2012$SEV))  
# There are 1966 civilian fatalities from a total of 11344 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2012 <- subset(fireincident2012,grepl("N",fireincident2012$NOT\_RES))  
# There are 281115 residential fire incidents from a total of 599879 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2012 <- merge(civilianfatality2012,residentialfireincident2012,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1515 residential fire fatalities of year 2012.  
  
#4. merge cfrfrfi2012 with incidentaddress2012  
cfrfrfi2012address <- merge(cfrfrfi2012,incidentaddress2012,by =c("INC\_DATE","INC\_NO","EXP\_NO"))  
#####cfrfrfi2012address is wrong when using all 5 identification variables, because of txt ?, so instead I use the last 3 identification variables.  
##### maybe some states changed the fire station id in their incidentaddress?  
  
clean2012 <- data.frame(INC\_DATE = cfrfrfi2012address$INC\_DATE,INC\_NO = cfrfrfi2012address$INC\_NO,EXP\_NO = cfrfrfi2012address$EXP\_NO, zip =cfrfrfi2012address$ZIP5,CAUSE\_INJ = cfrfrfi2012address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2012address$CAUSE\_IGN)  
  
clean2012location <- merge(clean2012,zipcode,by = c("zip"))  
  
## plot 2012 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2012location, mapping=aes(x=longitude, y=latitude, color=CAUSE\_IGN )) + coord\_equal(ratio = 1) + ggtitle("2012 Residential Fire Fatalities")



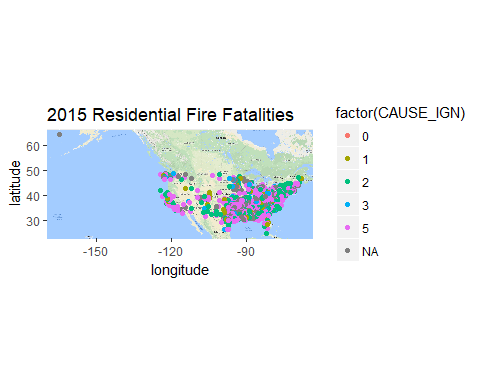
#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
########## Clean the data of year 2013  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2013 <- subset(civiliancasualty2013, grepl("5",civiliancasualty2013$SEV))  
# There are 1940 civilian fatalities from a total of 10891 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2013 <- subset(fireincident2013,grepl("N",fireincident2013$NOT\_RES))  
# There are 272564 residential fire incidents from a total of 554671 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2013 <- merge(civilianfatality2013,residentialfireincident2013,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1548 residential fire fatalities of year 2013.  
  
#4. merge cfrfrfi2013 with incidentaddress2013  
cfrfrfi2013address <- merge(cfrfrfi2013,incidentaddress2013,by =c("INC\_DATE","INC\_NO","EXP\_NO"))  
#####cfrfrfi2013address is wrong when using all 5 identification variables, because of txt ?, so instead I use the last 3 identification variables.  
  
clean2013 <- data.frame(INC\_DATE = cfrfrfi2013address$INC\_DATE,INC\_NO = cfrfrfi2013address$INC\_NO,EXP\_NO = cfrfrfi2013address$EXP\_NO, zip =cfrfrfi2013address$ZIP5,CAUSE\_INJ = cfrfrfi2013address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2013address$CAUSE\_IGN)  
  
clean2013location <- merge(clean2013,zipcode,by = c("zip"))  
  
## plot 2013 Residential Fire Fatalities  
usa2013 <- readPNG("usa2013.png")  
  
ggplot() + annotation\_raster(usa2013, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2013location, mapping=aes(x=longitude, y=latitude, color=factor(CAUSE\_IGN))) + coord\_equal(ratio = 1) + ggtitle("2013 Residential Fire Fatalities")



#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
  
########## Clean the data of year 2014  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2014 <- subset(civiliancasualty2014, grepl("5",civiliancasualty2014$SEV))  
# There are 2151 civilian fatalities from a total of 11386 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2014 <- subset(fireincident2014,grepl("N",fireincident2014$NOT\_RES))  
# There are 289628 residential fire incidents from a total of 596521 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2014 <- merge(civilianfatality2014,residentialfireincident2014,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1611 residential fire fatalities of year 2014.  
  
#4. merge cfrfrfi2014 with incidentaddress2014  
cfrfrfi2014address <- merge(cfrfrfi2014,incidentaddress2014,by =c("INC\_DATE","INC\_NO","EXP\_NO"))  
#####cfrfrfi2014address is wrong when using all 5 identification variables, because of txt ?, so instead I use the last 3 identification variables.  
  
clean2014 <- data.frame(INC\_DATE = cfrfrfi2014address$INC\_DATE,INC\_NO = cfrfrfi2014address$INC\_NO,EXP\_NO = cfrfrfi2014address$EXP\_NO, zip =cfrfrfi2014address$ZIP5,CAUSE\_INJ = cfrfrfi2014address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2014address$CAUSE\_IGN)  
  
clean2014location <- merge(clean2014,zipcode,by = c("zip"))  
  
## plot 2014 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2014location, mapping=aes(x=longitude, y=latitude, color= factor(CAUSE\_IGN) )) + coord\_equal(ratio = 1) + ggtitle("2014 Residential Fire Fatalities")



#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
###### background does not fit due to different "outliers" results in different locations of X,Y ######  
  
  
########## Clean the data of year 2015  
  
#1. find out which fire incidents cause fatalities(civilian).  
civilianfatality2015 <- subset(civiliancasualty2015, grepl("5",civiliancasualty2015$SEV))  
# There are 2174 civilian fatalities from a total of 11356 casualities.  
  
#2. find out which fire are residential  
residentialfireincident2015 <- subset(fireincident2015,grepl("N",fireincident2015$NOT\_RES))  
# There are 288002 residential fire incidents from a total of 599361 fire incident.  
  
#3. merge the above tables to construct table of civilian fatality results from residential fire incident(cfrfrfi).  
cfrfrfi2015 <- merge(civilianfatality2015,residentialfireincident2015,by =c("STATE","FDID","INC\_DATE","INC\_NO","EXP\_NO"))  
  
# There are 1636 residential fire fatalities of year 2015.  
  
#4. merge cfrfrfi2015 with incidentaddress2015  
cfrfrfi2015address <- merge(cfrfrfi2015,incidentaddress2015,by =c("INC\_DATE","INC\_NO","EXP\_NO"))  
#####cfrfrfi2015address is wrong when using all 5 identification variables, because of txt ?, so instead I use the last 3 identification variables.  
  
clean2015 <- data.frame(INC\_DATE = cfrfrfi2015address$INC\_DATE,INC\_NO = cfrfrfi2015address$INC\_NO,EXP\_NO = cfrfrfi2015address$EXP\_NO, zip =cfrfrfi2015address$ZIP5,CAUSE\_INJ = cfrfrfi2015address$CAUSE\_INJ,CAUSE\_IGN = cfrfrfi2015address$CAUSE\_IGN)  
  
clean2015location <- merge(clean2015,zipcode,by = c("zip"))  
  
## plot 2015 Residential Fire Fatalities  
ggplot() + annotation\_raster(usa, -Inf, Inf, -Inf, Inf, interpolate = TRUE) + geom\_point(data=clean2015location, mapping=aes(x=longitude, y=latitude, color=factor(CAUSE\_IGN) )) + coord\_equal(ratio = 1) + ggtitle("2015 Residential Fire Fatalities")



#The above plot of location is not very accurate with the google map of USA because google map simulate  
# the projection of the surface of a sphere(Earth) while the X,Y axies are flat on 2 dimension space.  
  
  
  
# The plots for Residential Fire Fatalities of year 2006~2015 are saved in folder "plots" under the working directory.  
  
  
  
###### Explanations ######  
# Q1: Why do I use CivilianCasualty, FireIncident, IncidentAddress as my data for the project?   
# A1: CivilianCasualty contains the objective civilian fatalities. FireIncident provides a variable that distinguishes Residential Fire Incident From Others. IncidentAddress provides address for incidents for ploting.  
  
# Q2: How do I use these data?  
# A2: Merge all 3 data together by the common first 5 identification variables, join with zipcode, and reshape the results only containing necessary variables.   
  
###### Results ######  
results <- read.dbf("Results.dbf")  
results

## Year Death Incident  
## 1 2006 1559 281695  
## 2 2007 1583 284181  
## 3 2008 1617 293745  
## 4 2009 1577 288145  
## 5 2010 1651 317074  
## 6 2011 1659 321499  
## 7 2012 1515 281115  
## 8 2013 1548 272564  
## 9 2014 1611 289628  
## 10 2015 1636 288002

# There is no significant trend of the number of Civilian Residential Fire Fatalities.  
# There is no significant trend of the number of Residential Fire Incidents.  
  
# After oberserving the plots of Residential Fire Fatalities of 2006 to 2015, there is no significant trend of the distrubution of Civilian Residential Fire Fatalities.  
# As all the plots show, most of fatalities occured in Mid-east and East Coast, with small amount occured in the West Coast.  
# Almost all the fatalities occured in mainland.  
  
# End of Document