NCHS

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

NCHS\_New1 <- read.csv("NCHS.csv")  
str(NCHS\_New1)

summary(NCHS\_New1)

sum(is.na(NCHS\_New1))

aft\_mis\_NCHS\_New1<-na.omit(NCHS\_New1)  
sum(is.na(aft\_mis\_NCHS\_New1))

aft\_mis\_NCHS\_New2<- aft\_mis\_NCHS\_New1[!(aft\_mis\_NCHS\_New1$State == "United States"),]  
dat <- aft\_mis\_NCHS\_New2  
head(dat)

library(noncensus)

data("states")

Health\_data <- merge(dat,Reg,by.x = "State.FIPS.Code",by.y = "state")  
head(Health\_data)

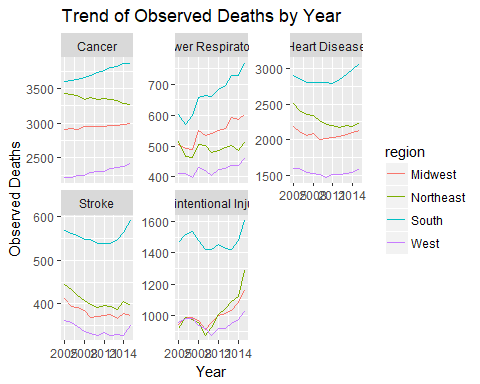
# write.csv(Health\_data,"Health data.csv")  
  
# 1.trend of all the five leading cause of Observed deaths over time for each region  
  
Data\_all <- Health\_data[,c(2:3,9,14)]  
head(Data\_all)

library(dplyr)

Data\_allNew <- Data\_all %>%  
 group\_by(region, Year, Cause.of.Death) %>%  
 summarise\_all(funs(mean))  
Data\_allNew

library(ggplot2)

ggplot(data = Data\_allNew, aes(x=Year, y = Observed.Deaths, colour = region, group = region)) + labs(title="Trend of Observed Deaths by Year", x="Year",y="Observed Deaths") +  
 geom\_line()+scale\_x\_continuous(breaks=seq(2005, 2015, 3)) + facet\_wrap(~Cause.of.Death,scales = "free\_y")



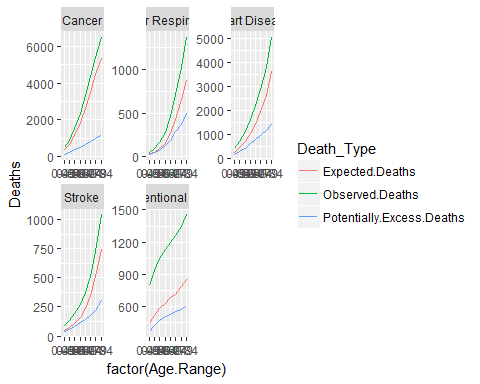
# 2. Max and Min value of all types of deaths in Each age group

library(dplyr)  
Data\_sec\_New <- Data\_sec %>%  
 group\_by(Age.Range, Cause.of.Death) %>%  
 summarise\_all(funs(mean))  
Data\_sec\_New

library(tidyr)

long\_data\_sec\_New <- gather(Data\_sec\_New,"Death\_Type", "Deaths", Observed.Deaths, Expected.Deaths, Potentially.Excess.Deaths)  
long\_data\_sec\_New

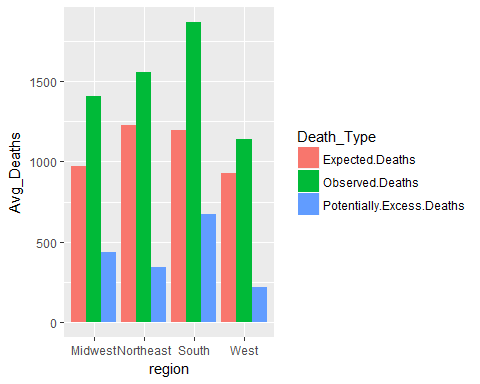
library(ggplot2)  
ggplot(data = long\_data\_sec\_New, aes(x=factor(Age.Range), y = Deaths, colour = Death\_Type, group = Death\_Type)) +   
 geom\_line() + facet\_wrap(~Cause.of.Death,scales = "free")



# 3. Max and Min Deaths by region  
library(dplyr)  
Data\_region <- Health\_data[,c(9,11,12,14)] %>%  
 group\_by(region) %>%  
 summarise\_all(funs(mean))

library(tidyr)  
long\_data\_region <- gather(Data\_region,"Death\_Type", "Avg\_Deaths", Observed.Deaths, Expected.Deaths, Potentially.Excess.Deaths)  
long\_data\_region

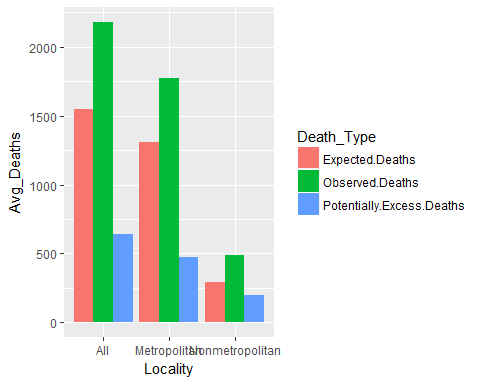
library(ggplot2)  
ggplot(long\_data\_region,aes(x = region,y = Avg\_Deaths,fill = Death\_Type)) + geom\_bar(position = "dodge", stat = "identity")



#4.  
library(dplyr)  
Data\_Locality <- Health\_data[,c(8,9,11,12)] %>%  
 group\_by(Locality) %>%  
 summarise\_all(funs(mean))

library(tidyr)  
long\_Data\_Locality <- gather(Data\_Locality,"Death\_Type", "Avg\_Deaths", Observed.Deaths, Expected.Deaths, Potentially.Excess.Deaths)  
long\_Data\_Locality

library(ggplot2)  
ggplot(long\_Data\_Locality,aes(x = Locality,y = Avg\_Deaths,fill = Death\_Type)) + geom\_bar(position = "dodge", stat = "identity")

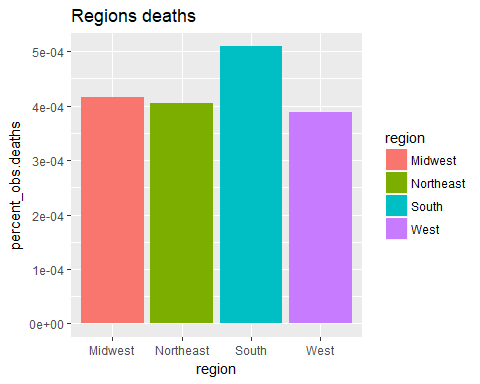


#5  
Health\_data$percent\_obs.deaths <- Health\_data$Observed.Deaths / Health\_data$Population  
head(Health\_data)

Obs.Deaths\_region <- Health\_data[,c(14:15)] %>%  
 group\_by(region) %>%  
 summarise\_all(funs(mean))  
Obs.Deaths\_region

## # A tibble: 4 x 2  
## region percent\_obs.deaths  
## <fctr> <dbl>  
## 1 Midwest 0.0004149184  
## 2 Northeast 0.0004044605  
## 3 South 0.0005085527  
## 4 West 0.0003889081

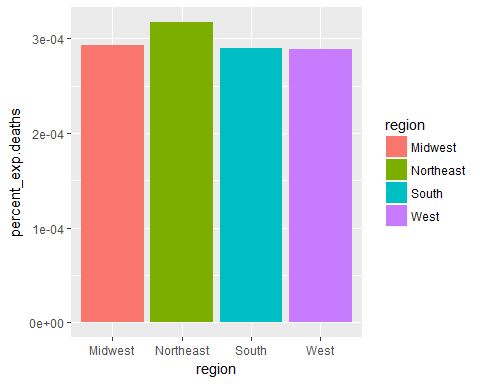
library(ggplot2)  
ggplot(Obs.Deaths\_region,aes(x = region,y = percent\_obs.deaths,fill = region)) + labs(title="Regions deaths") + geom\_bar(stat = "identity")



#6  
Health\_data$percent\_exp.deaths <- Health\_data$Expected.Deaths / Health\_data$Population  
head(Health\_data)

exp.Deaths\_region <- Health\_data[,c(14,16)] %>%  
 group\_by(region) %>%  
 summarise\_all(funs(mean))  
exp.Deaths\_region

library(ggplot2)  
ggplot(exp.Deaths\_region,aes(x = region,y = percent\_exp.deaths,fill = region)) + geom\_bar(stat = "identity")



#Standard diviation:  
  
sd(Health\_data$Potentially.Excess.Deaths)

sd(Health\_data$Observed.Deaths)

sd(Health\_data$Expected.Deaths)

#Anova table  
Anova.aov= aov(Health\_data$Expected.Deaths~Health\_data$Cause.of.Death)  
Anova.aov

summary(Anova.aov)

Anova.aov= aov(Health\_data$Population~Health\_data$Observed.Death)  
Anova.aov

summary(Anova.aov)

Anova.aov= aov(Health\_data$Observed.Deaths~Health\_data$Cause.of.Death)  
Anova.aov

summary(Anova.aov)

#Linear Regression  
Regression<-lm(Health\_data$Expected.Deaths ~ Health\_data$Cause.of.Death + Health\_data$Population)  
summary(Regression)

#Correlation  
cor.test(Health\_data$percent\_obs.deaths,Health\_data$Population)

# partial correlations  
library(ggm)

pcor(c("Observed.Deaths", "percent\_obs.deaths"), var(Health\_data))

## Warning in var(Health\_data): NAs introduced by coercion

## [1] 0.4312337