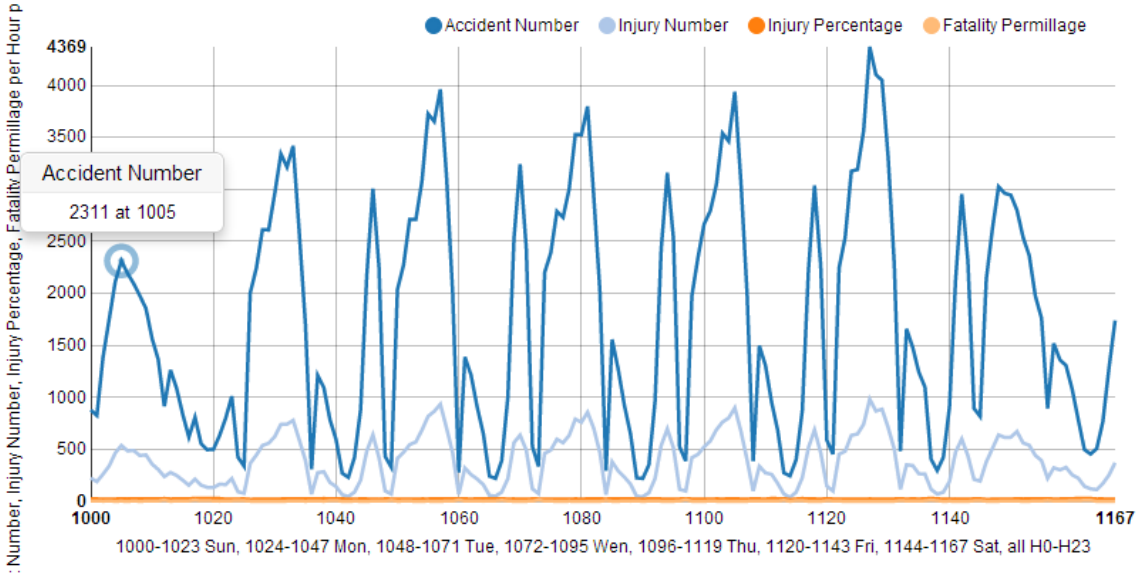
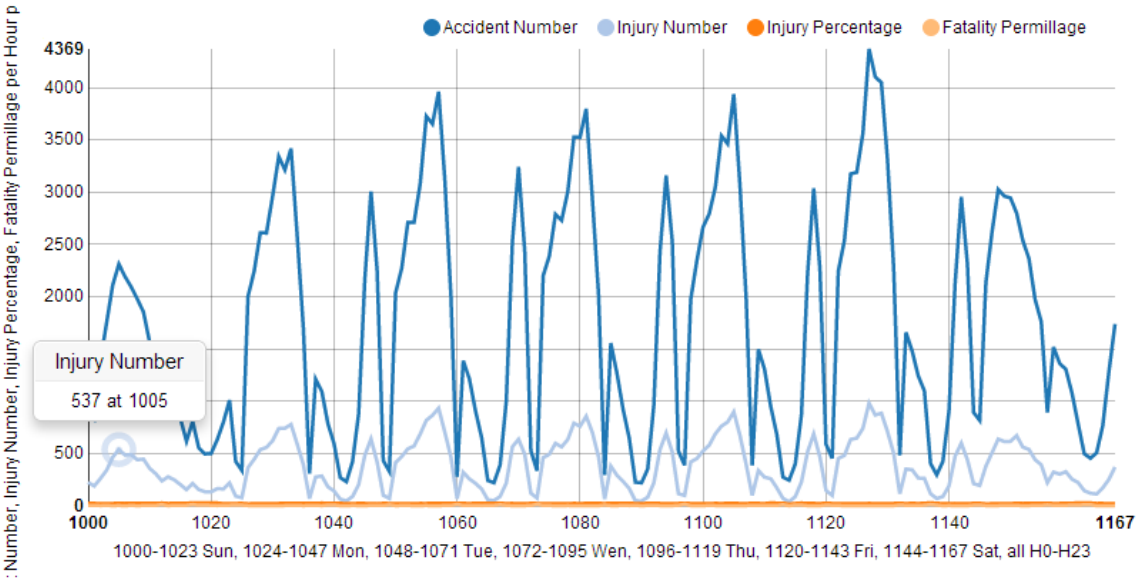
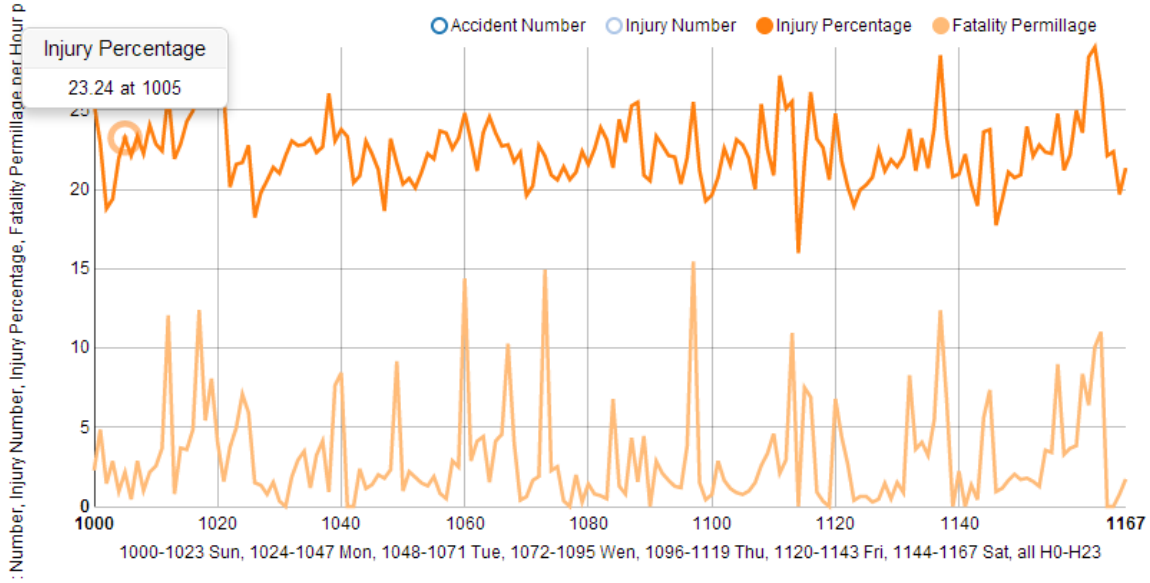
Introduction to R

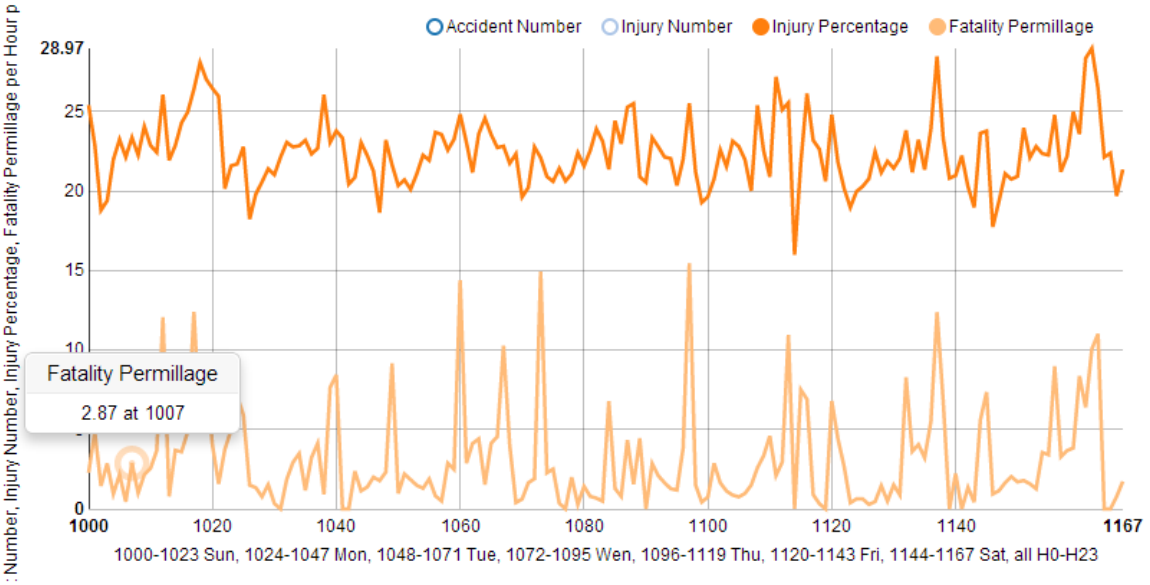
1. Accident Distribution

Plots:









Codes:

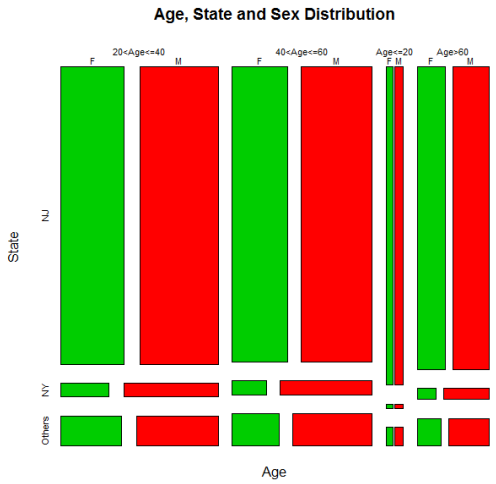
```
setwd("C:/users/gang/desktop")
require(devtools)
require(rCharts)
Accidents<-read.delim("NewJersey2011Accidents.txt",header=FALSE,sep=",")
```

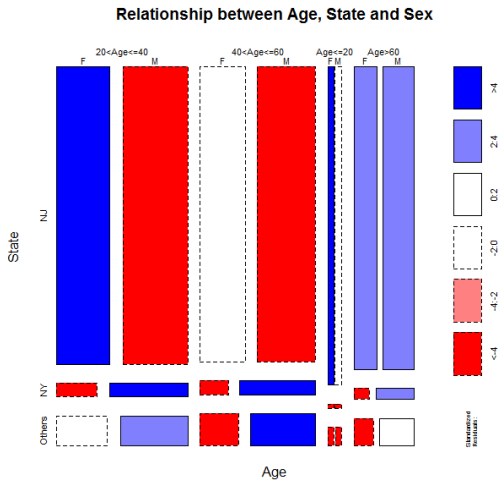
```
Accidents<- Accidents[which(Accidents$V6!="    "),]
Hour<- Accidents$V6
count <- nchar(Hour)
for (i in 1:291462){
  if (count[i]== 4){
    Hour[i]<-substr(Accidents[i, ]$V6, 1,2)
  }
}
for (i in 1:291462){
  if (count[i]== 3){
    Hour[i]<-substr(Accidents[i, ]$V6, 1,1)
  }
}
for (i in 1:291462){
  if (count[i]<= 2){
    Hour[i]<-0
  }
}
Accidents<- cbind(Accidents, Hour)
temp <- table(Accidents$V5, Accidents$Hour)
temp <- temp[2:8,]
sample1 <- table(Accidents$Hour, Accidents$V5, Accidents$V14)
Fatality <- sample1[1:24,2:8,1:1]
Injury <- sample1[1:24,2:8,2:2]
Accidents$Hour<-as.integer(Accidents$Hour)
anum <- c(temp[3,], temp[2,], temp[6,],temp[7,],temp[5,],temp[1,],temp[4,])
inum <- c(Injury[,3],Injury[,2], Injury[,6],Injury[,7],Injury[,5],Injury[,1],Injury[,4])
fnum <-
c(Fatality[,3],Fatality[,2],Fatality[,6],Fatality[,7],Fatality[,5],Fatality[,1],Fatality[,4])
iinum <- inum
inum <- inum/anum*100
```

```
fnum <- fnum/anum*1000
fnum <- round(fnum, 2)
inum <- round(inum, 2)
Result <- c(anum, iinum, inum, fnum)
Number <- c(1000:1167)
Number <- c(Number, Number, Number, Number)
df = data.frame(Result, Number)
aa <- rep("Accident Number", 168)
nn <- rep("Injury Number", 168)
ii <- rep("Injury Percentage", 168)
ff <- rep("Fatality Permillage", 168)
type <- c(aa, nn, ii, ff)
df$Type <-type
dfPlot <- nPlot(
  Result ~ Number,
  data = df,
  group = "Type",
  type = "lineChart")
dfPlot$yAxis(axisLabel = "Accident Number, Injury Number, Injury Percentage, Fatality
Permillage per Hour per Weekday", width = 62)
dfPlot$xAxis(axisLabel = "1000-1023 Sun, 1024-1047 Mon, 1048-1071 Tue, 1072-1095
Wen, 1096-1119 Thu, 1120-1143 Fri, 1144-1167 Sat, all H0-H23")
dfPlot
```

2. Relationship

Plots:





The plot on the right is based on the model which Sex is jointly independent of State and Age. Residuals show the extent to which Sex is related to the combination of State and Age. Above zero residual means there's more than would be if State and Age were jointly independent of Sex, and below zero residual means there's less than would be if State and Age were jointly independent of Sex.

Codes:

```
setwd("C:/users/gang/desktop")
Drivers<-read.delim("NewJersey2011Drivers.txt",header=FALSE,sep=",")
Vehicles<-read.delim("NewJersey2011Vehicles.txt",header=FALSE,sep=",")
Drivers <- Drivers[which(Drivers$V7!="          "),]
age <- substr(Drivers$V7, 7,10)
age <- as.integer(age)
age <- 2014 - age
age[which(age<=21)]<- 1
age[which(age>60)]<-4
age[which(age>21&age<=40)]<-2
age[which(age>40&age<=60)]<-3
Drivers$V7 <- age
Drivers <- Drivers[which(Drivers$V8!=" "),]
```

```
Drivers <- data.frame(lapply(Drivers, as.character), stringsAsFactors=FALSE)
Vehicles <- data.frame(lapply(Vehicles, as.character), stringsAsFactors=FALSE)
mm <- Vehicles$V4
mm <- as.character(mm)
for (i in 1:449111){
  if (mm[i]!= "NJ"){
    if (mm[i]!= "NY"){
      if (mm[i]!= "PA"){
        mm[i] <- "Others"
      }
    }
  }
}
Vehicles <- data.frame(lapply(Vehicles, as.character), stringsAsFactors=FALSE)
Vehicles$V4 <- mm
Drivers <- data.frame(lapply(Drivers, as.character), stringsAsFactors=FALSE)
keep <- c("V1","V2","V4")
Vehicles <- Vehicles[keep]
keeps <- c("V1","V2","V7","V8")
Drivers<- Drivers[keeps]
m1 <- merge(Drivers, Vehicles, by.x =c("V1","V2"), by.y=c("V1","V2") )
age <- as.integer(m1$V7)
sex <- as.character(m1$V8)
state <- as.character(m1$V4)
df = data.frame(age, sex, state)
mm <- rep("N", 449111)
for (i in 1:449111){
  if (state[i]=="NY"){
    mm[i]<-"NY"
  }
  if (state[i]=="NJ"){
```

```
   mm[i]<-"NJ"
  }
  if (state[i]=="Others"){
   mm[i]<-"Others"
  }
}
Age <- df$age
Sex <- df$sex
for (i in 1:449111){
  if (Age[i]==1){
    Age[i]<- "Age<=20"
    }
  if (Age[i]==2){
    Age[i]<- "20<Age<=40"
    }
  if (Age[i]==3){
    Age[i]<- "40<Age<=60"
  }
  if (Age[i]==4){
    Age[i]<- "Age>60"
  }
}
State<- mm
temp <- table(Age, State, Sex)
par(mfrow=c(1,2))
mosaicplot(temp, main = "Age, State and Sex Distribution" , col = c(3, 2) )
mosaicplot(temp, main = "Relationship between Age, State and Sex" , shade = TRUE)
```