In [30]: ```
#Data analysis of the Fatality Analysis Reporting System (FARS)
#from the National Highway Traffic Safety Administration
#(1975-1981)
```

In [2]: ```
accidents <- read.csv("/class/datamine/data/fars/7581.csv")
```

In [ ]: ```
#The "head" command gives us a preview of the data in this dataset.
#It has 45 different variables reported for 43,445 seperate
#accidents.
```

In [38]: ```
head(accidents)
```

A data.frame: 6 x 45

	STATE	COUNTY	MONTH	DAY	YEAR	HOUR	MINUTE	VE_FORMS	PERSONS
	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
9955	17	31	1	2	1975	17	45	4	6
9956	17	43	1	2	1975	19	46	2	3
9957	17	31	3	10	1975	19	55	1	2
9958	17	117	1	4	1975	4	20	1	1
9959	17	31	1	5	1975	17	2	1	5
9960	17	113	1	5	1975	3	0	1	2

In [ ]: ```
#If we wanted to look at accidents involving drunk drivers, we can
#load a table of all accidents involving drunk drivers sorted by
#the number of drunk drivers involved.
```

In [46]: ```
table(accidents$DRUNK_DR)
```

```
    0     1     2     3     4     6
28956 13575   887    24     2     1
```

In [ ]: ```
#i.e. there were 28,956 accidents that did not involve
#drunk drivers, 13,575 accidents that involved 1 drunk driver...,
#and 1 accident that involved 6 drunk drivers.
```

In [ ]: ```
#We can look at two simultameous variables from our dataset.
#For instance, we can see how many accidents involved both
#drunk drivers and school buses.
```

```
In [48]: table(accidents$DRUNK_DR, accidents$SCH_BUS)
```

```
        0     1
0  19019   124
1  11233     7
2    823     0
3     23     0
4      2     0
6      1     0
```

```
In [ ]: #i.e. this table shows there were 124 cases where 1 school bus
        #crashed and no drunk drivers were involved, and 7 where
        #1 school bus and 1 drunk driver were involved.
```

```
In [60]: #We can easily sort the number of accidents involving
         #drunk drivers and school buses by year.
```

```
In [52]: table(accidents$YEAR[which(accidents$DRUNK_DR > 0 & accidents$SCH_B
         US==1)])
```

```
1977 1978 1979 1981
   1    2    2    2
```

```
In [62]: #To see an example with more data points, we can look at drunk
         #drivers and number of fatalities
```

```
In [64]: table(accidents$DRUNK_DR, accidents$FATALS)
```

```
        1      2     3     4     5    6    7    8    9   12
0   26394   2057   350   117    25    6    3    2    1    1
1   12032   1245   222    47    22    5    2    0    0    0
2     642    189    40    10     4    1    1    0    0    0
3      18      2     3     0     1    0    0    0    0    0
4       1      0     0     1     0    0    0    0    0    0
6       1      0     0     0     0    0    0    0    0    0
```

```
In [ ]: #We can also sort this data by year.
```

```
In [61]: table(accidents$YEAR[which(accidents$DRUNK_DR > 0 & accidents$FATAL
         S)])
```

```
1975 1976 1977 1978 1979 1980 1981
1039 1361 2235 2400 2568 2465 2421
```

```
In [ ]: #We can look to see if weather impacts the number of accidents.
        #One way is to look at Indiana and its surronding states to see
        #whether inclement winter weather increases the number of
        #accidents per month.
```

```
In [4]: dat <- accidents[accidents$STATE %in% c(17, 18, 26, 39),]
        barplot(table(dat$STATE, dat$MONTH))
```
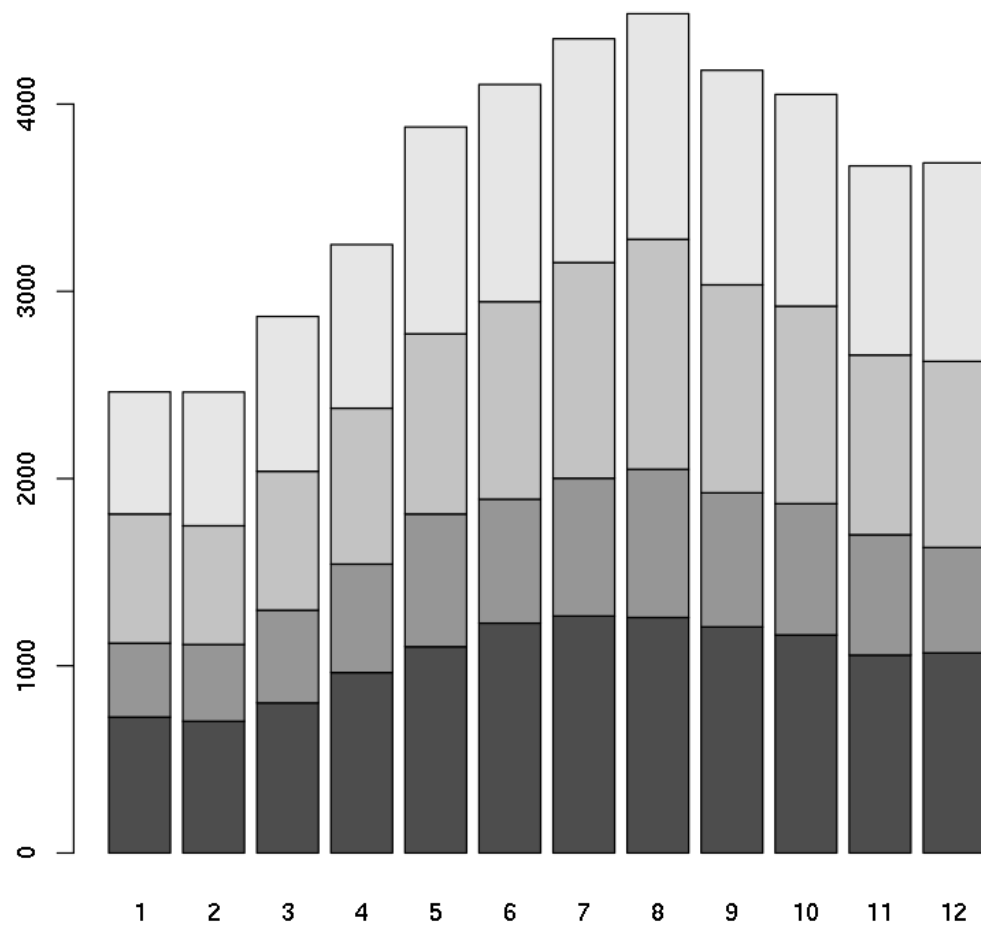


```
In [ ]: #Interestingly, we see the winter months actually have less
        #accidents!
```

```
In [ ]: #Now lets add colors and labels to better explain the information
        #given.
```

```
In [36]: library(RColorBrewer)
         coul <- brewer.pal(5, "Set2")
         dat <- accidents[accidents$STATE %in% c(17, 18, 26, 39),]
         dat$STATE <- factor(dat$STATE, labels=c("Illinois", "Indiana", "Mic
         higan", "Ohio"))
         barplot(table(dat$STATE, dat$MONTH),
                 col=coul,
                 xlab="Months of the year",
                 ylab="Number of accidents",
                 main="Number of accidents per month by state",
                 xlim=c(0, 20),
                 args.legend=list(x=19.3, y=4600),
                 legend=T
                 )
```


Number of accidents per month by state
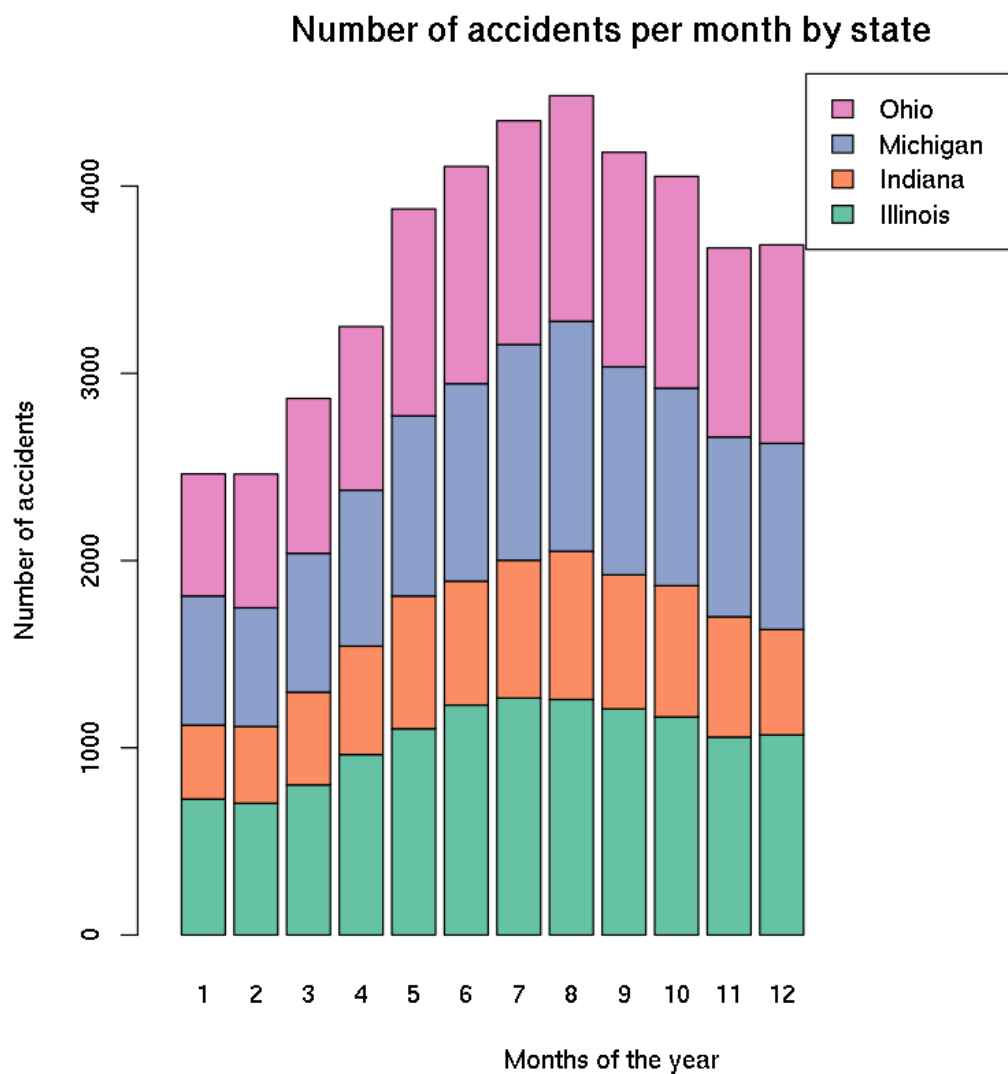
```
In [ ]: #The following graph shows which day of the week has the most
        #cummulative fatal accidents.
```

```
In [100]: library(RColorBrewer)
          coul <- brewer.pal(5, "Set2")
          accidents$DAY_WEEK <- factor(accidents$DAY_WEEK, labels=c("Sun", "M
          on", "Tues", "Wed", "Thurs", "Fri", "Sat", "Unknown"))
          dat <- tapply(accidents$FATALS, accidents$DAY_WEEK, sum)
          barplot((dat),
                  col = coul
                  )
```
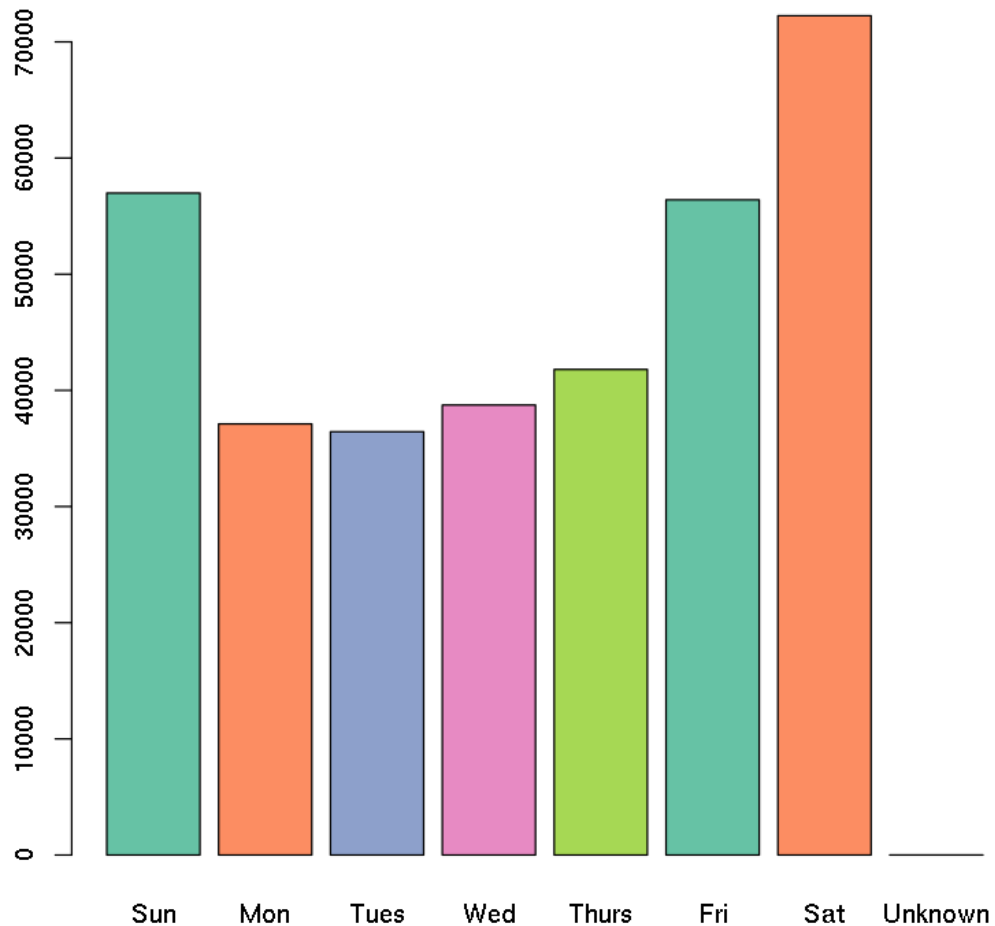


```
In [89]: #This graph shows that more traffic fatalities occured on the
         #weekends than on weekdays.
```

```
In [103]: #We can also look at the number of accidents based upon the time of
          #day.
```

In [3]:
```
library(RColorBrewer)
coul <- brewer.pal(5, "Set2")
this <- tapply(accidents$FATALS, cut(accidents$HOUR, breaks=c(0,6,1
2,18,24,99), include.lowest=T), sum)
barplot((this),
        col = coul
        )
```
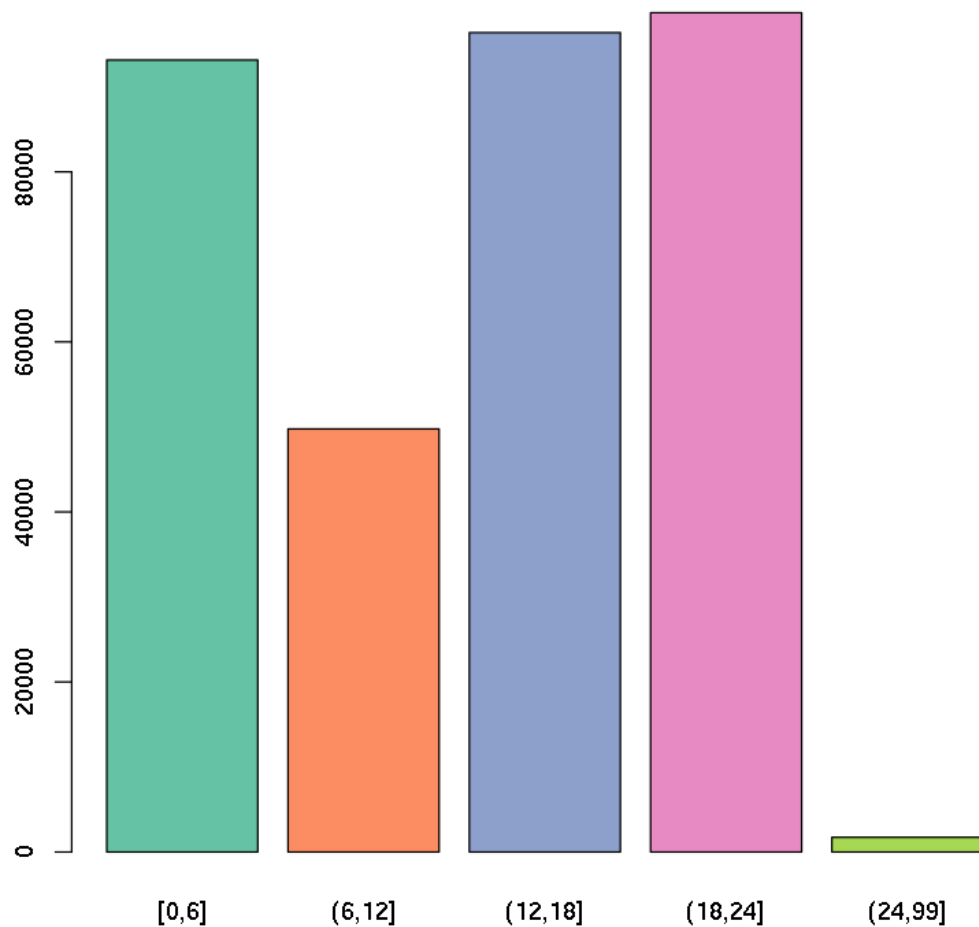
In [ ]:
```
#In this we used the "cut" function to parse the number of
#accidents per time of day. Specifically we looked at from
#midnight to 6a.m., 6a.m. to noon, noon to 6p.m., 6p.m. to
#midnight, and there is an unknown category.
```