Homework 5

```
## Loading required package: ggplot2
## Loading required package: magrittr
```

Sorry about all that preliminary stuff. The function above makes Mosiac plots very similar to jmp.

Problem 39

Part a

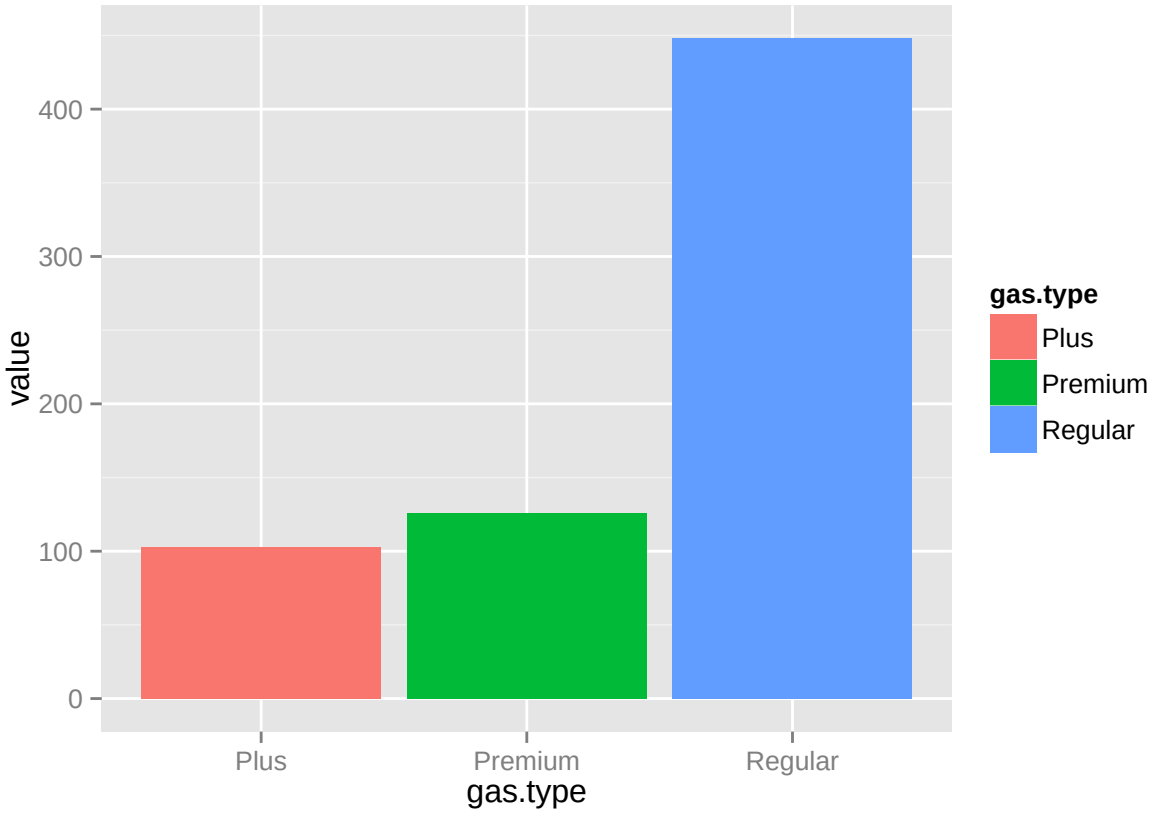
Contingency table for Gas Sales

```
gas.table <- table(gas$Grade.of.Gasoline, gas$Type.of.Day)
row.margin <- margin.table(gas.table, 1)

plot.table <- gas.table  %>% as.data.frame # we'll use this for graphing later
names(plot.table) <- c("gas.type", "day.type", "value")

gas.table <- gas.table %>%
  cbind(row.margin)

column.margin <- margin.table(gas.table, 2)

gas.table <- gas.table %>%
  rbind(column.margin) %>%
  as.data.frame %>%
  print
```

```
##                Weekday Weekend row.margin
## Plus               103      29        132
## Premium            126      63        189
## Regular            448     115        563
## column.margin      677     207        884
```

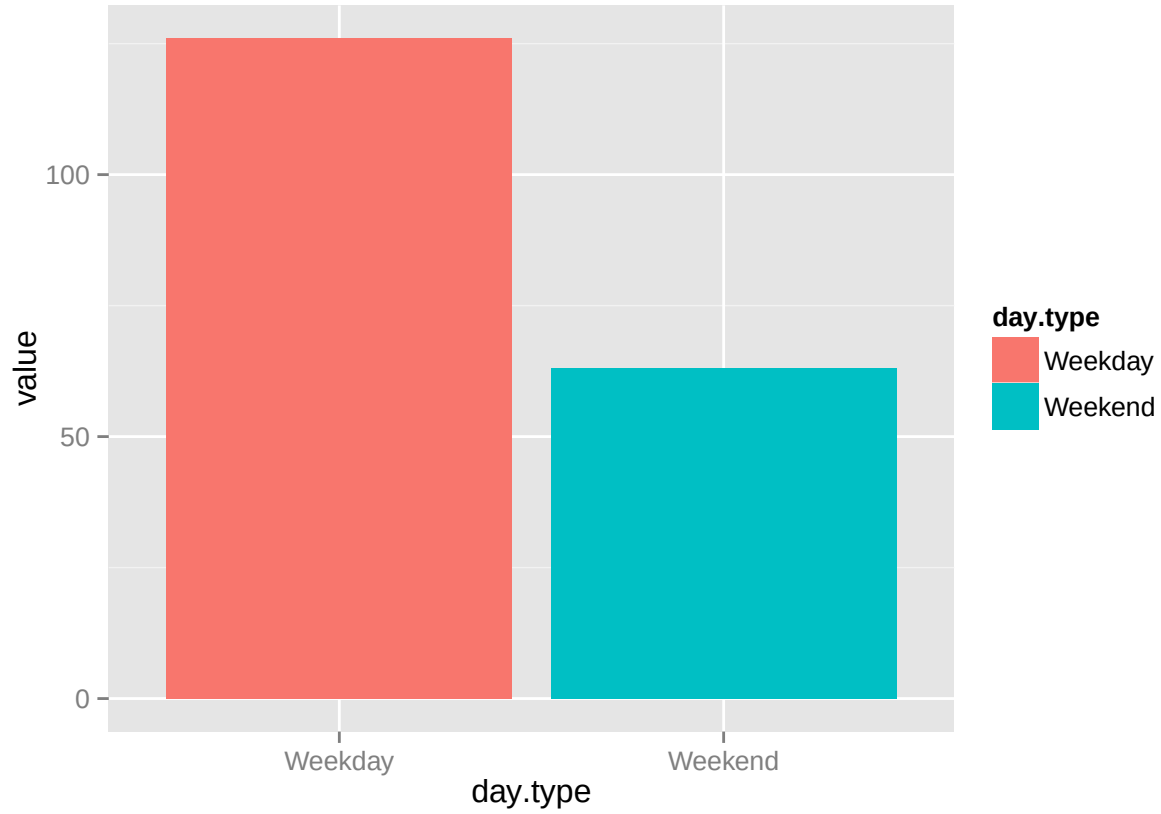
Part b

```
ggplot(data = plot.table[plot.table$day.type == "Weekday",], aes(x = gas.type, y = value, fill = gas.ty
  geom_bar(stat = "identity")
```



part c

```
ggplot(data = plot.table[plot.table$gas.type == "Premium",], aes(x = day.type, y = value, fill = day.ty
  geom_bar(stat = "identity")
```



part d

Yes, this is probably because vehicles that require premium gas are more closely tied to recreation and leasure. People typiclaly people indulge in these practises on the weekend.

Problem 43

part a

```
# get the data
########

owners <- read.table("C:\\Users\\Jonathan\\Google Drive\\Stats Camp\\Stine&Foster\\Data by Chapter\\Cha

owner.table <- table(owners$Satisfaction, owners$Question.Wording)

owner.table <- cbind(owner.table, margin.table(owner.table, 1))

owner.table <- rbind(owner.table, margin.table(owner.table, 2))

print(owner.table)
```

```
##                         Dissatisfied Satisfied
## Somewhat dissatisfied           20        12  32
## Somewhat satisfied              69        82 151
## Very dissatisfied               23        10  33
## Very satisfied                 128       139 267
##                                240       243 483
```

part b

The table indicates that the questions that were asked in terms of satisfaction generally got the most possitive results. It appears that the way the question was ask primed a certain responce. The somewhat satisfied and very satisfied rows are the only two that have a higher proportion of people in the satisfied column.

part c

The company should word the question in terms of customer satisfaction.

Problem 54

part a

In this instance we want to use column percentages because we are more interested percentages relitive to the company.

```
status <- c("On Time", "Delayed")
american <- c(1536, 416)
delta <- c(11769, 3343)
flight.delays <- data.frame(american, delta)
rownames(flight.delays) <- status

flight.delays$american <- flight.delays$american / sum(flight.delays$american)
flight.delays$delta <- flight.delays$delta / sum(flight.delays$delta)

print(flight.delays)
```

```
##          american     delta
## On Time 0.7868852 0.7787851
## Delayed 0.2131148 0.2212149
```

Based off this percentage table American Airlines arrives ontime a greater percentage of the time.

part b

```
Atlanta <- c(11512, 3334)
Las_Vegas <- c(1007, 244)
San_Diego <- c(601, 366)

city.delay <- data.frame(Atlanta, Las_Vegas, San_Diego)
```

```
rownames(city.delay) <- status
city.delay$Atlanta <- city.delay$Atlanta / sum(city.delay$Atlanta)
city.delay$Las_Vegas <- city.delay$Las_Vegas /sum(city.delay$Las_Vegas)
city.delay$San_Diego <- city.delay$San_Diego / sum(city.delay$San_Diego)

Atlanta <- c(653, 14193)
Las_Vegas <- c(698, 553)
San_Diego <- c(601, 366)
airline <- c("American", "Delta")

city.airline <- data.frame(Atlanta, Las_Vegas, San_Diego, row.names = airline)

city.airline$Atlanta <- city.airline$Atlanta / sum(city.airline$Atlanta)
city.airline$Las_Vegas <- city.airline$Las_Vegas /sum(city.airline$Las_Vegas)
city.airline$San_Diego <- city.airline$San_Diego / sum(city.airline$San_Diego)

print(city.delay)
```

```
##           Atlanta Las_Vegas San_Diego
## On Time 0.7754277  0.804956 0.6215098
## Delayed 0.2245723  0.195044 0.3784902
```

```
print(city.airline)
```

```
##             Atlanta Las_Vegas San_Diego
## American 0.04398491 0.5579536 0.6215098
## Delta    0.95601509 0.4420464 0.3784902
```

Based on these two tables, we suspect the presents of a lurking variable. This is pricipally due to the discrepancy the frequency that different airlines fly certain routes. If you fly to Las Vegas the probablity that you'll be delayed is much smaller than if you fly to San Diego. In our case the majority of Delta's flights go into Atlanta. Hence their percentage in part a looks most like the Atlanta conditional distribution.

part c

```
Atlanta <- c(497, 11015)
Las_Vegas <- c(561, 446)
San_Diego <- c(478, 308)

ontime <- data.frame(Atlanta, Las_Vegas, San_Diego, row.names = airline)
ontime$Atlanta <- ontime$Atlanta / sum(ontime$Atlanta)
ontime$Las_Vegas <- ontime$Las_Vegas /sum(ontime$Las_Vegas)
ontime$San_Diego <- ontime$San_Diego / sum(ontime$San_Diego)

print(ontime)
```

```
##             Atlanta Las_Vegas San_Diego
## American 0.04317234 0.5571003 0.6081425
## Delta    0.95682766 0.4428997 0.3918575
```

Upon examining the table it looks like destination isn't actually a lurking variable. The percentages on the ontime table and the city.airline table are about the same.