Hw problem in a couple of weeks (capstone problem, includes nearly everything we've learned):

- There is a premium associated with green buildings
    - Is it worth it?
        - Costs money to go through the certification process, do it, etc
        - Given information about every green building in the United States and ones that are comparable without this certification
        - Task? – action item - should you pursue certification or not?

<u>Peak Demand data set:</u>

```
library(mosaic)

 #Build a model to forecast peak demand

 sunday <-- subset(peakdemand,Sun="1")
 saturday <--subset(peakdeamnd,Sat="1")

#Scrub missing values
 peakdeamnd=na.omit(peakdemand)

#there is definitely a season effect because the graph is showing almost a cyclical effect.
plot(peakdemand$PeakDemand, main= "Peak Demand over Time")
plot(PeakDemand~DailyTemp,data=peakdemand, main = "Peak Demand vs Temperature")
plot(PeakDemand~factor(Month),data=peakdemand)
#One possibility: in the hotter months there is more peak demands


#Step1: Add a time value
N= nrow(peakdemand)
peakdemand$period = 1:nrow(peakdemand)
peakdemand$period
head(peakdemand)

lm1=lm(PeakDemand~period,data=peakdemand)
plot(PeakDemand ~ period,data=peakdemand)
abline(lm1, col='red',lwd=4)

lm2=lm(PeakDemand~period+DailyTemp + I(DailyTemp^2) + Sat + Sun, data=peakdemand)
#dummy variable of temperature
plot(resid(lm1),type='l')
```

```
lines(resid(lm2),col='red')
#the red residuals are much smaller showing you are getting closer to the truth

#lm3 is with seasonal dummies
lm3=lm(PeakDemand~period+DailyTemp+I(DailyTemp^2)+Sat+Sun+factor(Month
),data=peakdemand)

plot(resid(lm3),type='l')
summary(lm3)
anova(lm3)
#how to assess whether a variable helps: Neyman pearson command , the shuffle.
Gives us the notion
#whether adding varibales added a significant amount to our R^2 or if it was simply
to chance

lmstep=step(lm3,direction='backward')
#You want to minimize AIC, every variable i could delete makes AIC bigger. AIC
would suggest that we
have a decent model there is nothing we can add.

plot(resid(lm3),type='l')

#comparefits
plot(PeakDemand~period,data=peakdemand,type='l')
lines(fitted(lm3)~period,data=peakdemand,col='red')
#seems like a very good forecasting model because you can barely see the blacks
behind the reds
summary(lm3)
#pretty well predicting model

plot(resid(lm3),type='l')
#suggests that there is a little bit of predictability still left
```

<u>Probability</u>

- Rules (Kolmogorov's Axioms):
  1) Probabilities for mutually exclusive events must sum to 1
  2) Probabilities for disjoint events must add together
     a. Disjoint = can't occur at once. Ex: probability that someone is attending UT or OU. Student can't be attending both
  3) Probabilities are numbers between 0 and 1

- More complex rules:
    1) Addition Rule (Union rule)
        a. $P(A \cup B) = P(A) + P(B) - P(A,B)$
           "A or B" "A &B" – joint event – both occur

    2) Multiplication Rule
        a. $P(A,B) = P(A) * P(B|A)$
            i. The vertical bar means conditional upon or given.
            ii. Probability of event A and B is the probability of A times the probability of B given that A occurs. Ex: the probability that I will get sued and lose is equal to the probability that I will get sued times the probability that I will lose given that I get sued

<u>"But what does it all mean"</u>

1) Frequency interpretation
    a. Vegas interpretation
        i. P(Black 31) = # Times Black 31 comes up/# spins of roulette wheel
    b. What are the chances that its going to rain today?
        i. Hard to answer that question under the frequency interpretation
2) Degree of belief interpretation
    a. There is a formal mathematic argument that leads to this in the notes. Don't expect you to know the derivation, just know it exists
    b. The amount you would pay in order to be written $100 contract that pays out if the event is true. Ex: If I believe there is a 10% chance of it raining before 5 pm then I would be willing to pay you $10 for that $100 contract.
    c. The "Wall Street Interpretation" of probability

<u>Bayes Rule</u>

$P(A,B) = P(A) * P(B|A)$ but you could've written this the opposite way
$= P(B) * P(A|B)$

so $P(A|B) = [\ P(A)\ P(B|A)\ ] / P(B)$

"Learning" Rule / "Bayenian updating" rule

Updating your probability that you believe something after getting new information.

G: accused in a trial is guilty
D: accused DNA "matches" the DNA at crime scene

$P(G) \rightarrow P(G|D)$
- Bayes rule tells us how to do that