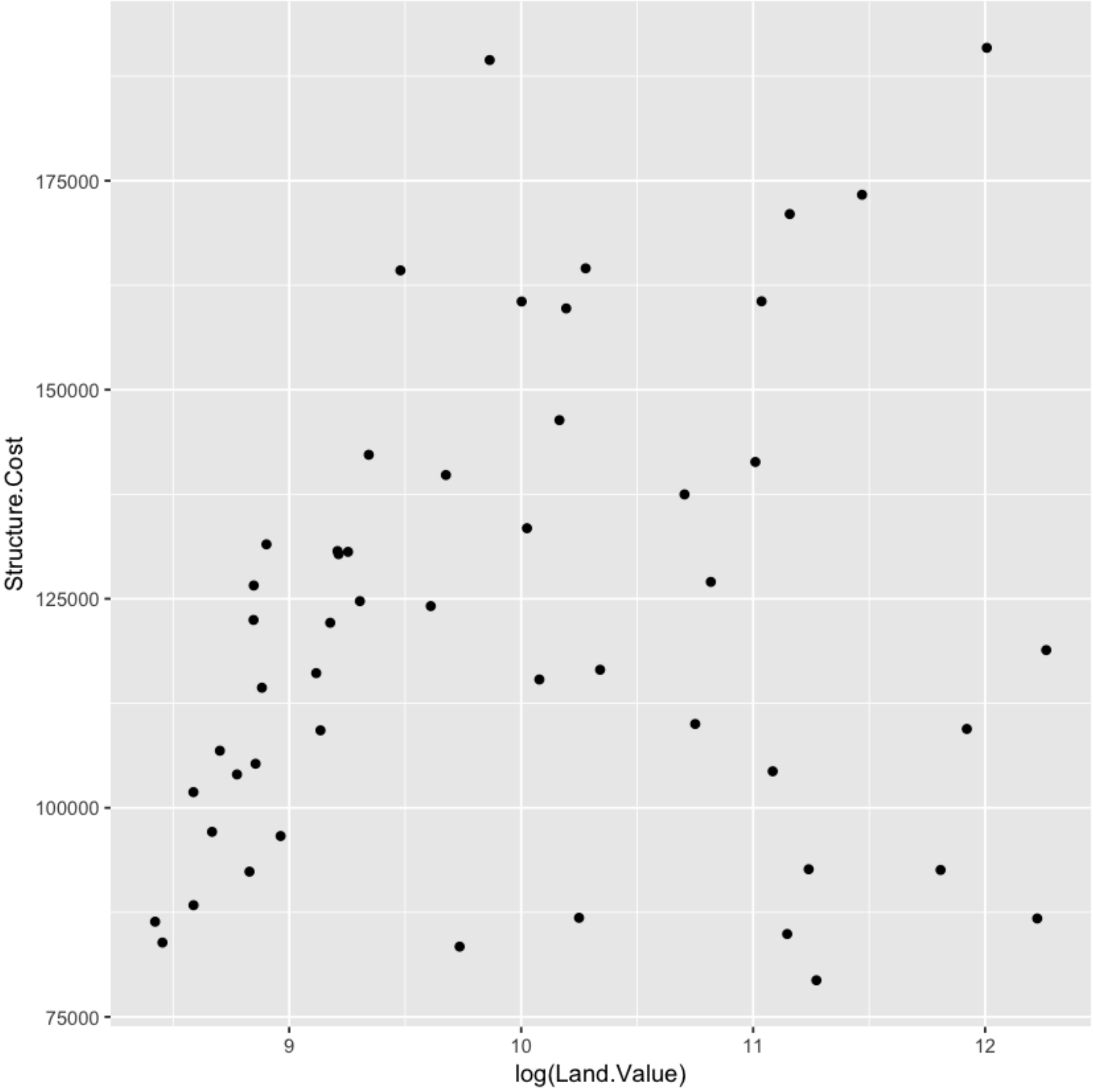
lab3q1

March 20, 2017

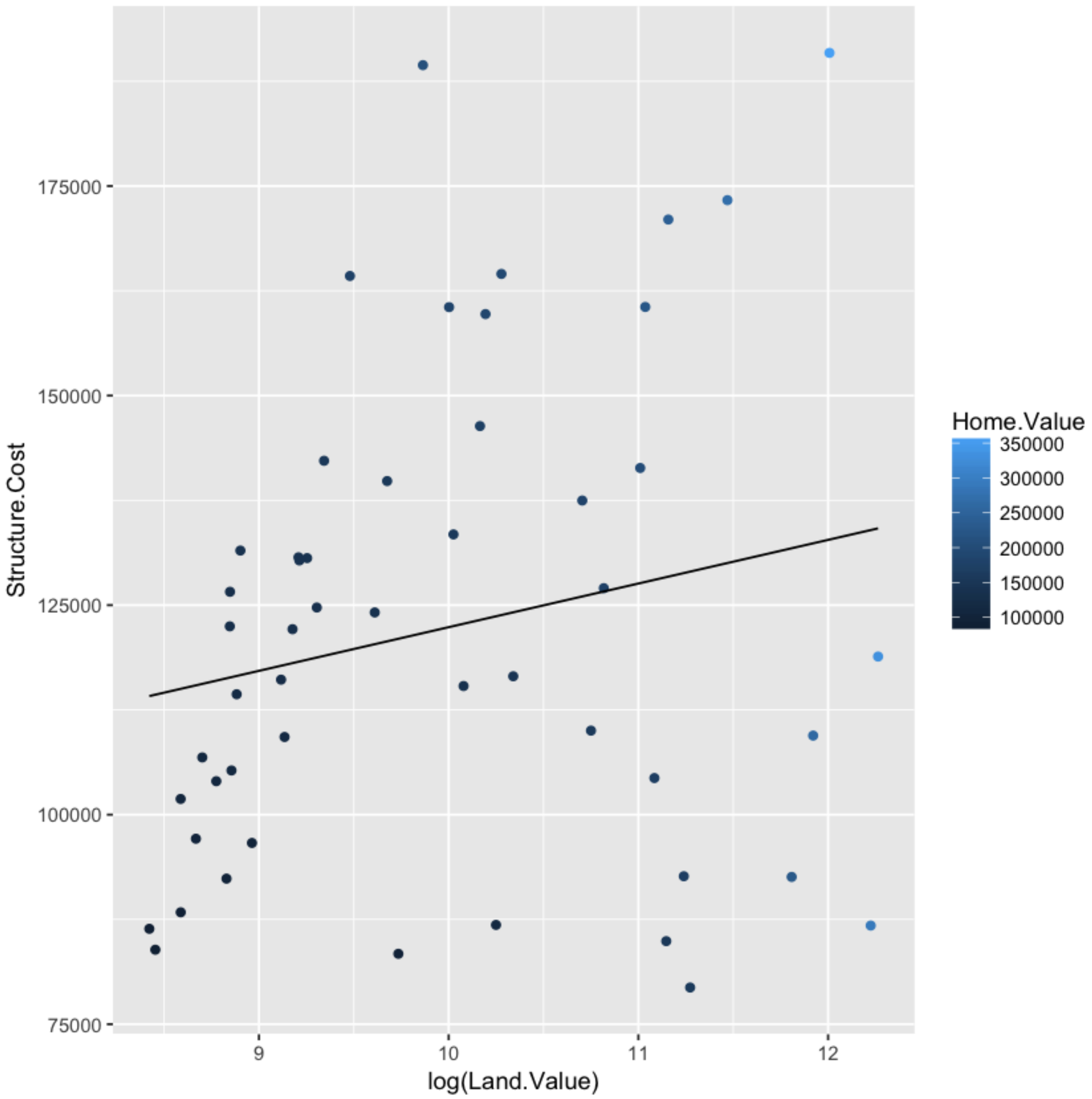
```
In [1]: ## Question 1
        library(ggplot2)
        housing <- read.csv("landdata-states.csv")

        hp2001Q1 <- subset(housing, Date == 2001.25)


        ggplot(hp2001Q1,
               aes(y = Structure.Cost, x = log(Land.Value))) +
            geom_point()

        ## Linear model
        hp2001Q1$pred.SC <- predict(lm(Structure.Cost ~ log(Land.Value), data = hp2

        p1 <- ggplot(hp2001Q1, aes(x = log(Land.Value), y = Structure.Cost))

        p1 + geom_point(aes(color = Home.Value)) +
            geom_line(aes(y = pred.SC))

        ## Kmeans
        km = kmeans(hp2001Q1[,4:6],3)
        km
        clusters = as.factor(km$cluster)
        ggplot(hp2001Q1,aes(x = Home.Value, y = Land.Value, color = clusters)) + ge

        ## knn
        data(iris)
        library(class)

        iris$Species = as.character(iris$Species)
        iris$Species[iris$Species == "setosa"] = 1
        iris$Species[iris$Species == "versicolor"] = 2
        iris$Species[iris$Species == "virginica"] = 3

        iris$Species = as.factor(iris$Species)

        training <- sample( 1:nrow(iris), 0.7*nrow(iris),replace = F)
```

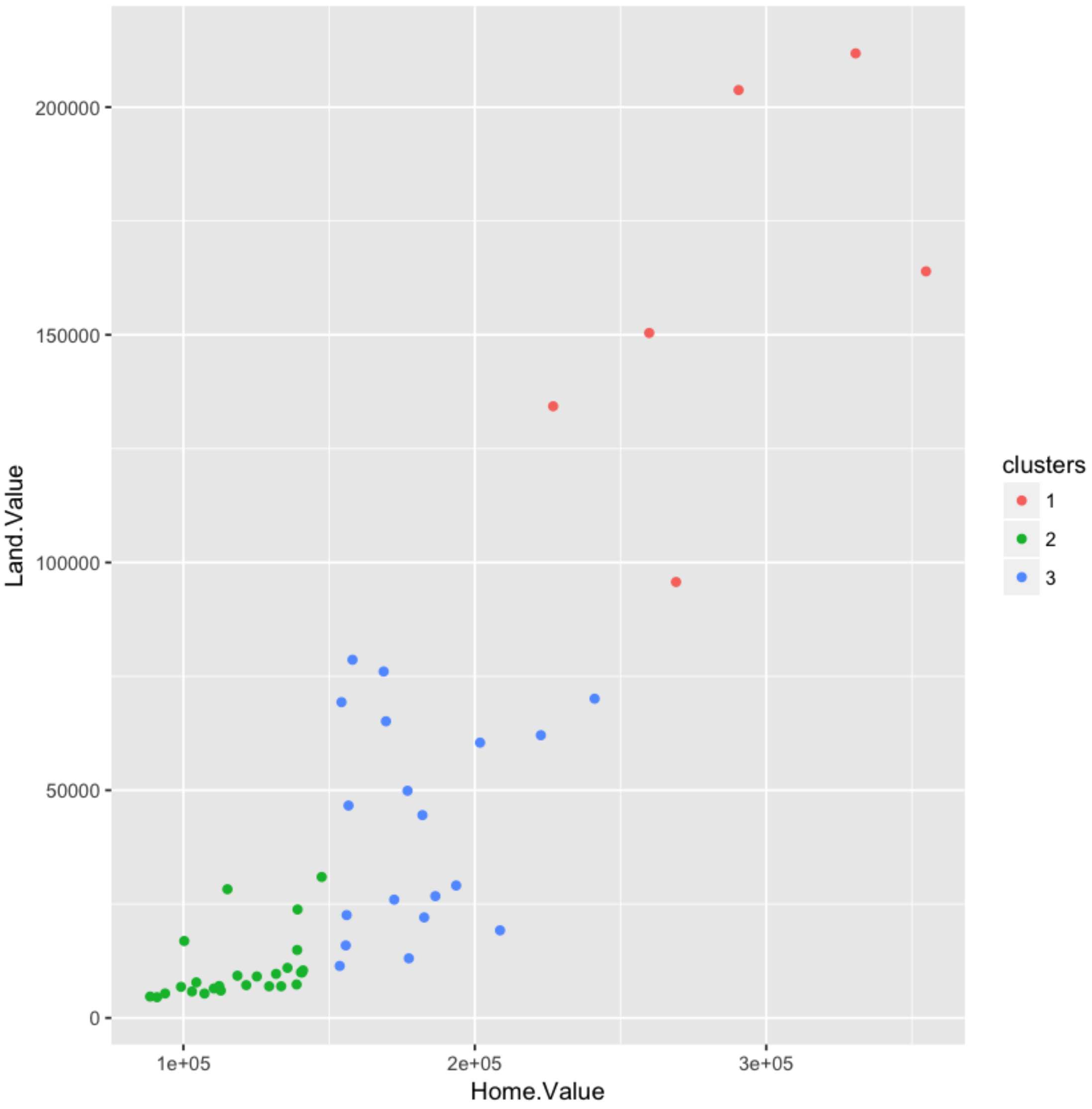
```
train.iris <- iris[training,]
test.iris <- iris[-training,]
cl = train.iris$Species

knn.model = knn(train.iris,test.iris,cl,k = 3)
```



```
K-means clustering with 3 clusters of sizes 6, 26, 19

Cluster means:
  Home.Value Structure.Cost Land.Value
1   288639.2       128640.2  159998.67
2   120017.2       109528.7   10488.35
3   179902.1       137320.8   42581.11
```

```
Clustering vector:
 143  226  380  532  685  907  991 1144 1329 1482 1635 1788 1941 2094 2247 2400 
   3    2    2    3    1    1    1    3    3    3    1    2    2    3    2    2 
2553 2706 2859 3012 3165 3318 3471 3624 3777 3930 4083 4236 4389 4542 4695 4848 
   2    2    1    3    2    3    3    2    2    2    2    2    2    3    3    2 
5001 5154 5307 5460 5613 5766 5919 6072 6225 6378 6531 6684 6837 6990 7143 7296 
   3    3    2    2    3    2    3    2    2    2    2    3    3    3    3    2 
7449 7602 7765 
   2    2    1 

Within cluster sum of squares by cluster:
[1] 30063376761 15906513733 37833925012
 (between_SS / total_SS =  76.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"      
```


Structure.Cost
175000
150000
125000
100000
75000
9
10
11
12
log(Land.Value)
Home.Value
350000
300000
250000
200000
150000
100000


200000
150000
100000
50000
0
Land.Value
1e+05
2e+05
3e+05
Home.Value
clusters
● 1
● 2
● 3