Exercise 3

Samuel Higgins

4/14/2020

Question 1: Predictive Model Building

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(MASS)
```

```
##
## Attaching package: 'MASS'
## The following object is masked from 'package:dplyr':
##
##     select
```

```
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
green_b <- read.csv("https://github.com/jgscott/SDS323/raw/master/data/greenbuildings.csv")


green_b$Energystar2 <- factor(green_b$Energystar, levels = c(0,1), labels = c("no", "yes"))
green_b$LEED2 <- factor(green_b$LEED, levels = c(0,1), labels = c("no", "yes"))
green_b$green_rating2 <- factor(green_b$green_rating, levels = c(0,1), labels = c("no", "yes"))

lmbasic_gr <- lm(Rent ~ green_rating2, data = green_b)
summary(lmbasic_gr)
```

```
##
## Call:
```

```
## lm(formula = Rent ~ green_rating2, data = green_b)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.287  -9.044  -3.267   5.733 221.733
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       28.2668     0.1775 159.275   <2e-16 ***
## green_rating2yes   1.7493     0.6025   2.903   0.0037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.07 on 7892 degrees of freedom
## Multiple R-squared:  0.001067,   Adjusted R-squared:  0.0009405
## F-statistic:  8.43 on 1 and 7892 DF,  p-value: 0.003701
```

```r
t.test(Rent ~ green_rating2, data = green_b, var.eq = T)
```

```
##
##  Two Sample t-test
##
## data:  Rent by green_rating2
## t = -2.9035, df = 7892, p-value = 0.003701
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.9302447 -0.5682584
## sample estimates:
##  mean in group no mean in group yes
##          28.26678          30.01603
```

```r
## Stepwise Selection
fit1 <- lm(Rent ~ green_rating2 + ., data = green_b)
step1 <- stepAIC(fit1, direction = "backward")
```

```
## Start:  AIC=35088.2
## Rent ~ green_rating2 + (CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + green_rating + net + amenities + cd_total_07 +
##     hd_total07 + total_dd_07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent + Energystar2 + LEED2 + green_rating2)
##
##
## Step:  AIC=35088.2
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + green_rating + net + amenities + cd_total_07 +
##     hd_total07 + total_dd_07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent + Energystar2
##
##
## Step:  AIC=35088.2
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + green_rating + net + amenities + cd_total_07 +
```

```
##     hd_total07 + total_dd_07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent
##
##
## Step:  AIC=35088.2
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + green_rating + net + amenities + cd_total_07 +
##     hd_total07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent
##
##
## Step:  AIC=35088.2
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + net + amenities + cd_total_07 + hd_total07 +
##     Precipitation + Gas_Costs + Electricity_Costs + cluster_rent
##
##                     Df Sum of Sq     RSS   AIC
## - Energystar         1         0  690931 35086
## - green_rating2      1         3  690934 35086
## - LEED               1        24  690955 35086
## - renovated          1        27  690958 35087
## - cd_total_07        1        64  690995 35087
## <none>                            690931 35088
## - leasing_rate       1       279  691209 35089
## - CS_PropertyID      1       313  691244 35090
## - stories            1       408  691339 35091
## - age                1       622  691552 35093
## - cluster            1       623  691554 35093
## - amenities          1       627  691558 35093
## - Precipitation      1       796  691727 35095
## - class_b            1      1062  691993 35098
## - empl_gr            1      1276  692206 35101
## - net                1      1651  692581 35105
## - Gas_Costs          1      1825  692756 35107
## - hd_total07         1      3155  694086 35122
## - class_a            1      3816  694747 35129
## - Electricity_Costs  1      5068  695999 35143
## - size               1      9356  700286 35191
## - cluster_rent       1    446007 1136938 38981
##
## Step:  AIC=35086.2
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + net + amenities + cd_total_07 + hd_total07 + Precipitation +
##     Gas_Costs + Electricity_Costs + cluster_rent
##
##                     Df Sum of Sq     RSS   AIC
## - renovated          1        27  690958 35085
## - cd_total_07        1        64  690995 35085
## - green_rating2      1       122  691053 35086
## <none>                            690931 35086
## - LEED               1       208  691139 35087
```

```
## - leasing_rate       1       279  691210 35087
## - CS_PropertyID      1       313  691244 35088
## - stories            1       408  691339 35089
## - age                1       622  691553 35091
## - cluster            1       623  691554 35091
## - amenities          1       627  691558 35091
## - Precipitation      1       797  691728 35093
## - class_b            1      1062  691993 35096
## - empl_gr            1      1276  692207 35099
## - net                1      1651  692582 35103
## - Gas_Costs          1      1826  692757 35105
## - hd_total07         1      3155  694086 35120
## - class_a            1      3816  694747 35127
## - Electricity_Costs  1      5070  696001 35141
## - size               1      9355  700286 35189
## - cluster_rent       1    446070 1137001 38979
##
## Step:  AIC=35084.5
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + class_a + class_b + LEED +
##     net + amenities + cd_total_07 + hd_total07 + Precipitation +
##     Gas_Costs + Electricity_Costs + cluster_rent
##
##                     Df Sum of Sq     RSS   AIC
## - cd_total_07        1        60  691018 35083
## - green_rating2      1       125  691083 35084
## <none>                            690958 35085
## - LEED               1       206  691164 35085
## - leasing_rate       1       273  691231 35086
## - CS_PropertyID      1       342  691300 35086
## - stories            1       417  691375 35087
## - amenities          1       618  691576 35089
## - cluster            1       628  691586 35090
## - Precipitation      1       793  691751 35091
## - age                1       927  691885 35093
## - class_b            1      1039  691997 35094
## - empl_gr            1      1285  692243 35097
## - net                1      1653  692611 35101
## - Gas_Costs          1      1843  692801 35103
## - hd_total07         1      3285  694243 35120
## - class_a            1      3789  694747 35125
## - Electricity_Costs  1      5151  696109 35141
## - size               1      9351  700309 35188
## - cluster_rent       1    452555 1143513 39022
##
## Step:  AIC=35083.19
## Rent ~ green_rating2 + CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + class_a + class_b + LEED +
##     net + amenities + hd_total07 + Precipitation + Gas_Costs +
##     Electricity_Costs + cluster_rent
##
##                     Df Sum of Sq     RSS   AIC
## - green_rating2      1       119  691137 35083
## <none>                            691018 35083
```

```
## - LEED                1       215  691233 35084
## - leasing_rate        1       285  691303 35084
## - CS_PropertyID       1       334  691352 35085
## - stories             1       420  691438 35086
## - amenities           1       595  691614 35088
## - cluster             1       637  691655 35088
## - Precipitation       1       756  691774 35090
## - age                 1       879  691897 35091
## - class_b             1      1060  692078 35093
## - empl_gr             1      1286  692304 35096
## - net                 1      1705  692723 35100
## - Gas_Costs           1      2397  693415 35108
## - class_a             1      3934  694953 35126
## - hd_total07          1      4323  695341 35130
## - Electricity_Costs   1      5610  696629 35144
## - size                1      9292  700310 35186
## - cluster_rent        1    467175 1158193 39120
##
## Step:  AIC=35082.53
## Rent ~ CS_PropertyID + cluster + size + empl_gr + leasing_rate +
##     stories + age + class_a + class_b + LEED + net + amenities +
##     hd_total07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent
##
##                     Df Sum of Sq     RSS   AIC
## <none>                            691137 35083
## - leasing_rate       1       308  691445 35084
## - LEED               1       335  691472 35084
## - CS_PropertyID      1       335  691472 35084
## - stories            1       457  691594 35086
## - amenities          1       613  691750 35087
## - cluster            1       651  691788 35088
## - Precipitation      1       761  691899 35089
## - age                1       933  692070 35091
## - class_b            1      1067  692204 35093
## - empl_gr            1      1279  692416 35095
## - net                1      1698  692835 35100
## - Gas_Costs          1      2454  693591 35108
## - class_a            1      4150  695287 35127
## - hd_total07         1      4255  695392 35129
## - Electricity_Costs  1      5617  696755 35144
## - size               1      9396  700533 35186
## - cluster_rent       1    467879 1159017 39123
```

```
step1$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Rent ~ green_rating2 + (CS_PropertyID + cluster + size + empl_gr +
##     leasing_rate + stories + age + renovated + class_a + class_b +
##     LEED + Energystar + green_rating + net + amenities + cd_total_07 +
##     hd_total07 + total_dd_07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent + Energystar2 + LEED2 + green_rating2)
```

```
##
## Final Model:
## Rent ~ CS_PropertyID + cluster + size + empl_gr + leasing_rate +
##     stories + age + class_a + class_b + LEED + net + amenities +
##     hd_total07 + Precipitation + Gas_Costs + Electricity_Costs +
##     cluster_rent
##
##
##                Step Df     Deviance Resid. Df Resid. Dev      AIC
## 1                                       7798   690930.8 35088.20
## 2            - LEED2  0    0.0000000      7798   690930.8 35088.20
## 3      - Energystar2  0    0.0000000      7798   690930.8 35088.20
## 4      - total_dd_07  0    0.0000000      7798   690930.8 35088.20
## 5     - green_rating  0    0.0000000      7798   690930.8 35088.20
## 6       - Energystar  1    0.2750184      7799   690931.1 35086.20
## 7        - renovated  1   26.8887588      7800   690957.9 35084.50
## 8      - cd_total_07  1   60.3567620      7801   691018.3 35083.19
## 9 - green_rating2  1 119.0018140      7802   691137.3 35082.53
```

```r
coeftest(step1)
```

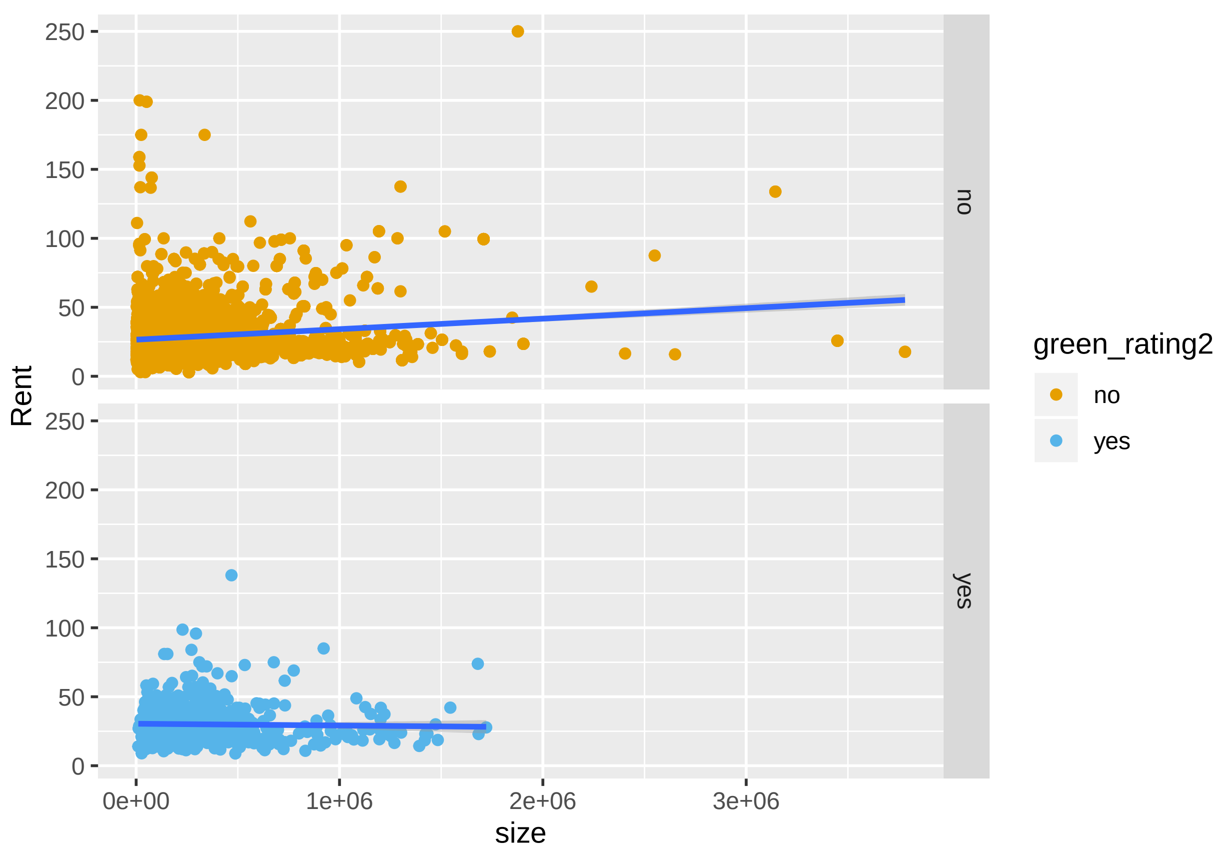
```
##
## t test of coefficients:
##
##                       Estimate  Std. Error t value  Pr(>|t|)
## (Intercept)        -8.4121e+00  9.9765e-01 -8.4319 < 2.2e-16 ***
## CS_PropertyID       3.0330e-07  1.5606e-07  1.9435 0.0519958 .
## cluster             7.6870e-04  2.8366e-04  2.7100 0.0067436 **
## size                6.7337e-06  6.5383e-07 10.2989 < 2.2e-16 ***
## empl_gr             5.8731e-02  1.5459e-02  3.7992 0.0001463 ***
## leasing_rate        9.9095e-03  5.3142e-03  1.8647 0.0622611 .
## stories            -3.6583e-02  1.6108e-02 -2.2711 0.0231651 *
## age                -1.3524e-02  4.1675e-03 -3.2450 0.0011794 **
## class_a             2.9480e+00  4.3072e-01  6.8444 8.259e-12 ***
## class_b             1.1828e+00  3.4083e-01  3.4705 0.0005223 ***
## LEED                2.5142e+00  1.2937e+00  1.9435 0.0519961 .
## net                -2.5894e+00  5.9141e-01 -4.3783 1.212e-05 ***
## amenities           6.5993e-01  2.5092e-01  2.6300 0.0085557 **
## hd_total07          5.6537e-04  8.1579e-05  6.9303 4.533e-12 ***
## Precipitation       4.6948e-02  1.6014e-02  2.9316 0.0033816 **
## Gas_Costs          -3.8462e+02  7.3080e+01 -5.2629 1.456e-07 ***
## Electricity_Costs  1.9386e+02  2.4344e+01  7.9631 1.915e-15 ***
## cluster_rent        1.0103e+00  1.3901e-02 72.6755 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
fit2 <- lm(Rent ~ green_rating2 + size + (green_rating2*size), data = green_b)
summary(fit2)
```

```
##
## Call:
## lm(formula = Rent ~ green_rating2 + size + (green_rating2 * size),
##     data = green_b)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -37.605  -9.130  -2.955   6.052 209.167
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)            2.655e+01  2.208e-01 120.243  < 2e-16 ***
## green_rating2yes       3.887e+00  8.857e-01   4.388 1.16e-05 ***
## size                   7.611e-06  5.918e-07  12.860  < 2e-16 ***
## green_rating2yes:size -8.893e-06  2.055e-06  -4.328 1.53e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.91 on 7890 degrees of freedom
## Multiple R-squared:  0.02163,    Adjusted R-squared:  0.02126
## F-statistic: 58.14 on 3 and 7890 DF,  p-value: < 2.2e-16
```

```
ggplot(data = green_b, aes(x = size, y = Rent, group = green_rating2)) +
  geom_point(aes(color = green_rating2)) +
  geom_smooth(method = "lm") +
  facet_grid("green_rating2") +
  scale_color_manual(values = c("no"="#E69F00","yes"="#56B4E9"))
```

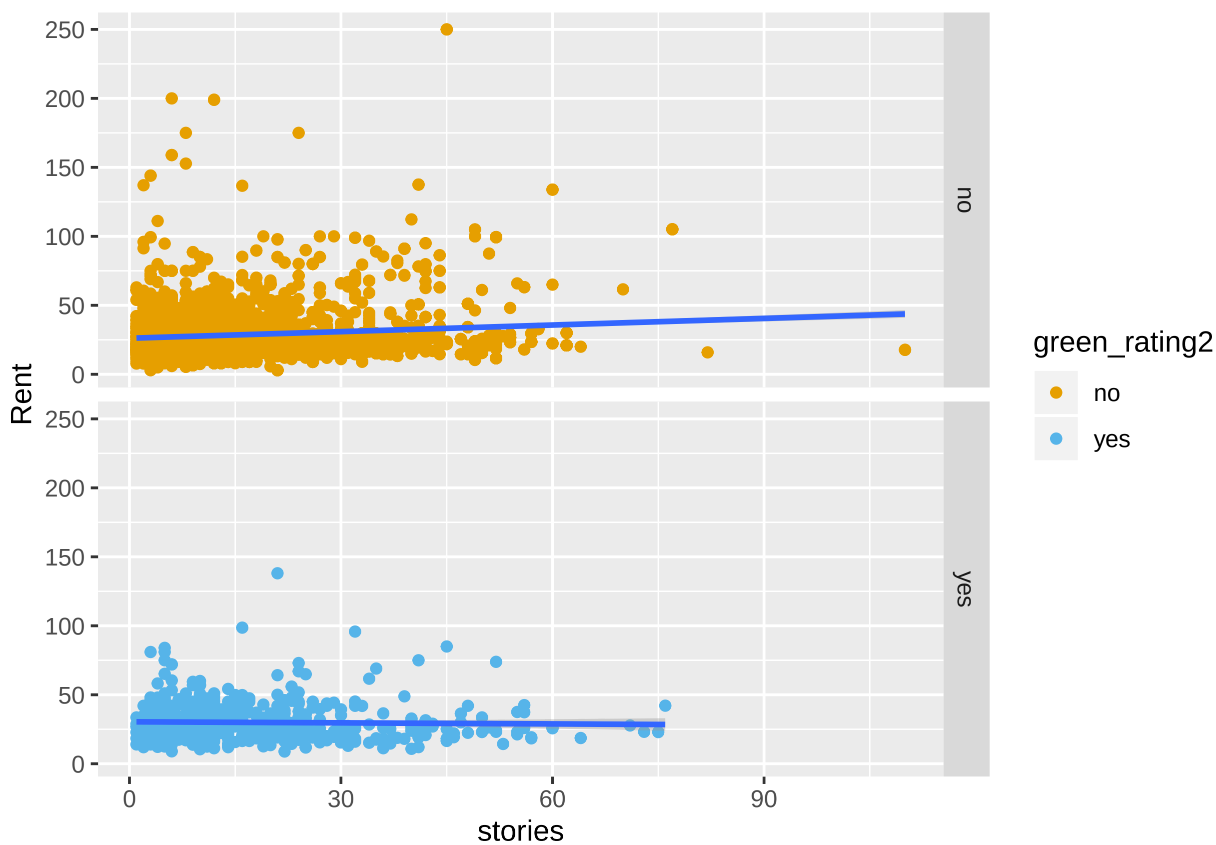



```
fit3 <- lm(Rent ~ green_rating2 + stories + (green_rating2*stories), data = green_b)
summary(fit3)
```

```
##
## Call:
```

```
## lm(formula = Rent ~ green_rating2 + stories + (green_rating2 *
##     stories), data = green_b)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.498  -9.282  -3.042   6.079 216.687
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                26.12250    0.26212  99.660  < 2e-16 ***
## green_rating2yes            4.28804    0.91148   4.704 2.59e-06 ***
## stories                     0.15980    0.01447  11.046  < 2e-16 ***
## green_rating2yes:stories -0.18553    0.04542  -4.085 4.45e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.95 on 7890 degrees of freedom
## Multiple R-squared:  0.01632,    Adjusted R-squared:  0.01595
## F-statistic: 43.65 on 3 and 7890 DF,  p-value: < 2.2e-16
```

```r
ggplot(data = green_b, aes(x = stories, y = Rent)) +
  geom_point(aes(color = green_rating2)) +
  geom_smooth(method = "lm") +
  facet_grid("green_rating2") +
  scale_color_manual(values = c("no"="#E69F00","yes"="#56B4E9"))
```



Controlling for size, there is a significant effect of green rating on Rent. When a building is green certified (either with LEED or Energystar certification), rent increases 3.887 dollars per square foot on average, t = 4.388, df = 7890, p < .001. Controlling for stories, there is also a significant effect of green rating on rent. When a building is green certified, rent increases 4.288 dollars per square foot on average, t= 4.704, df = 7890, p < .001. Stepwise selection was used to obtain the best predicted model for price.

Question 2: What Causes What?

2.1

It would be tricky to run a regression on "Crime" and "Police" for a sample of cities because some cities may have different reasons for putting more cops on the street, unrelated to that city's crime rate. In the podcast, they say that Washington D.C is a "high value" target for terrorist attacks. Based on a terror detection system, they may put more cops in public places and on the streets based on a potential terroristic threat, and when this happens, crime rate tends to be lower.

2.2

The researchers discovered the terrorist alert system in D.C, which gives a good example of an increase in police in the city, unrelated to crime. Controlling for Metro ridership, there is a significant effect of high-alert days on total daily crime. For every 1-unit increase in the high-alert system, total daily crime decreases by 7.316 on average. p < .05.

2.3

They had to control for Metro ridership because they wanted to know if tourists and civilians were less likely to be on the streets or in public as a result of the alert system. They found that the number of victims in the public remained unchanged on "high-terror" days.

2.4

On High-alert days, the total number of crimes decreases by 2.621 for District 1, relative to other police districts. Likewise, the total number of crimes decreases by .571 for other police districts, relative to District 1. Controlling for interactions, the log midday-ridership increases by 2.477.

Question 3: Clustering and PCA

```r
library(cluster)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

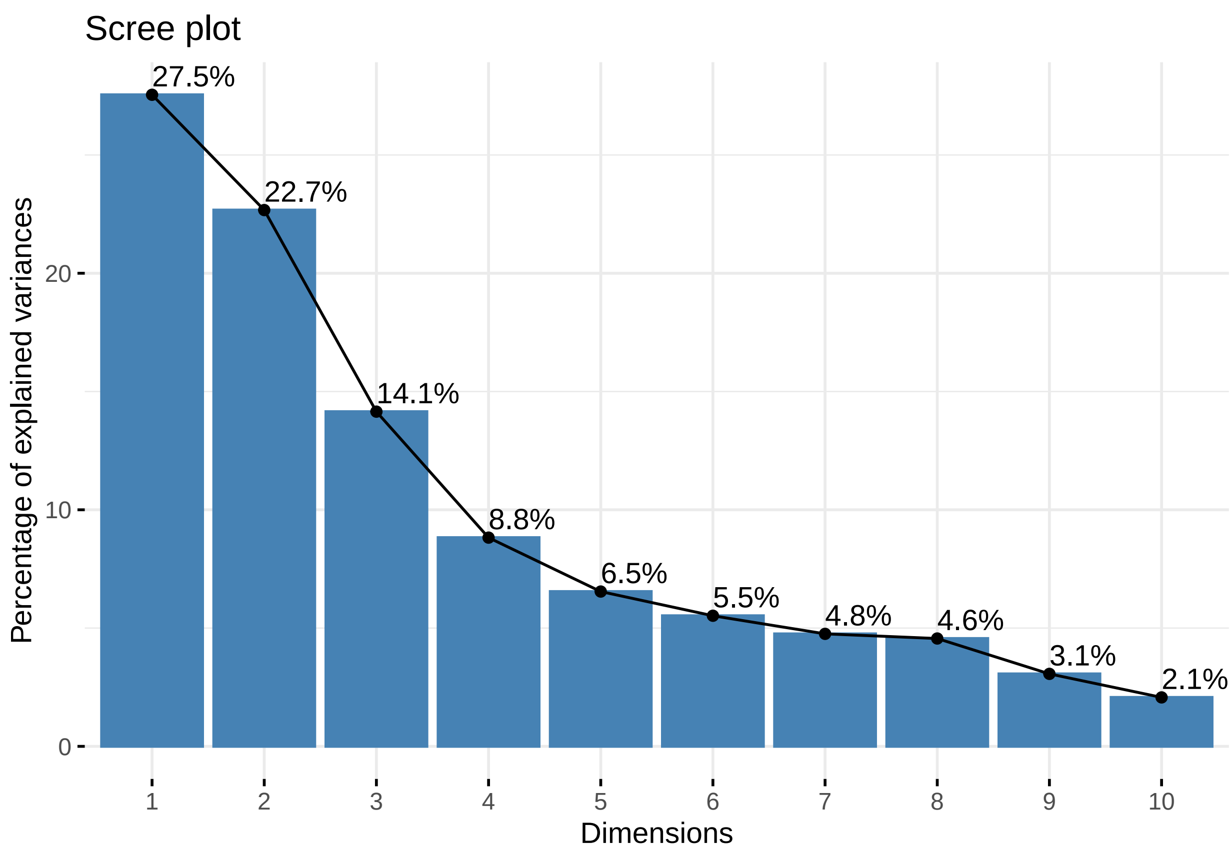
```r
wine <- read.csv("https://github.com/jgscott/SDS323/raw/master/data/wine.csv")


wine1 <- wine %>% dplyr::select(., -color, -quality)
wine_nums <- wine1 %>% select_if(is.numeric) %>% scale
wine_pca <- princomp(wine_nums)
names(wine_pca)
```

```
## [1] "sdev"     "loadings" "center"   "scale"    "n.obs"    "scores"   "call"
```

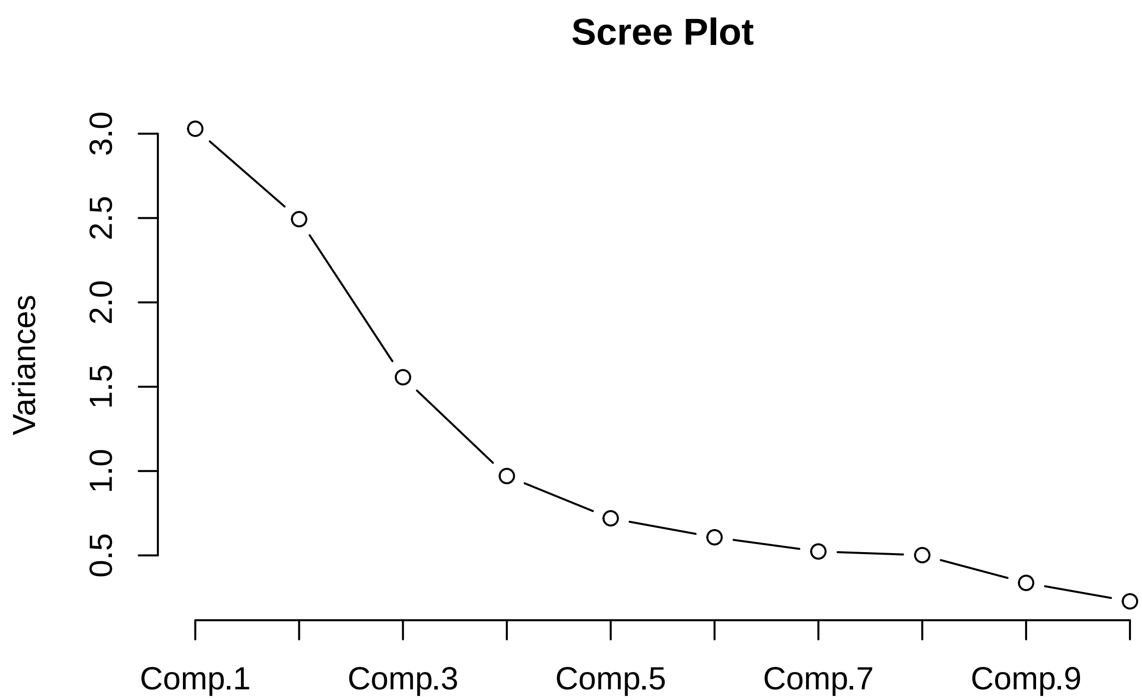
```r
summary(wine_pca, loadings = T)
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     1.7405178 1.5790637 1.2474403 0.98509020 0.84838913
## Proportion of Variance 0.2754426 0.2267115 0.1414861 0.08823201 0.06544317
## Cumulative Proportion  0.2754426 0.5021541 0.6436401 0.73187216 0.79731533
##                            Comp.6     Comp.7     Comp.8     Comp.9    Comp.10
## Standard deviation     0.77924209 0.72324148 0.70811941 0.58049304 0.47713805
## Proportion of Variance 0.05521016 0.04755989 0.04559184 0.03063855 0.02069961
## Cumulative Proportion  0.85252548 0.90008537 0.94567722 0.97631577 0.99701538
##                            Comp.11
## Standard deviation     0.181178776
## Proportion of Variance 0.002984618
## Cumulative Proportion  1.000000000
##
## Loadings:
##                      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
## fixed.acidity         0.239  0.336  0.434  0.164  0.147  0.205  0.283  0.401
## volatile.acidity      0.381  0.118 -0.307  0.213 -0.151  0.492  0.389
## citric.acid          -0.152  0.183  0.591 -0.264  0.155 -0.228  0.381 -0.293
## residual.sugar       -0.346  0.330 -0.165  0.167  0.353  0.233 -0.218 -0.525
## chlorides             0.290  0.315        -0.245 -0.614 -0.161        -0.472
## free.sulfur.dioxide  -0.431        -0.134 -0.357 -0.224  0.340  0.299  0.208
## total.sulfur.dioxide -0.487        -0.107 -0.208 -0.158  0.151  0.139  0.129
## density                      0.584 -0.176         0.307
## pH                    0.219 -0.156 -0.455 -0.415  0.453 -0.297  0.419
## sulphates             0.294  0.192        -0.641  0.137  0.297 -0.525  0.166
## alcohol               0.106 -0.465  0.261 -0.107  0.189  0.518  0.104 -0.399
##                      Comp.9 Comp.10 Comp.11
## fixed.acidity         0.344  0.281   0.335
## volatile.acidity     -0.497 -0.152
## citric.acid          -0.403 -0.234
## residual.sugar        0.108          0.450
## chlorides             0.296  0.197
## free.sulfur.dioxide   0.367 -0.480
## total.sulfur.dioxide -0.321  0.714
## density               0.113         -0.715
## pH                    0.128  0.141   0.206
## sulphates            -0.208
## alcohol               0.252  0.205  -0.336
```

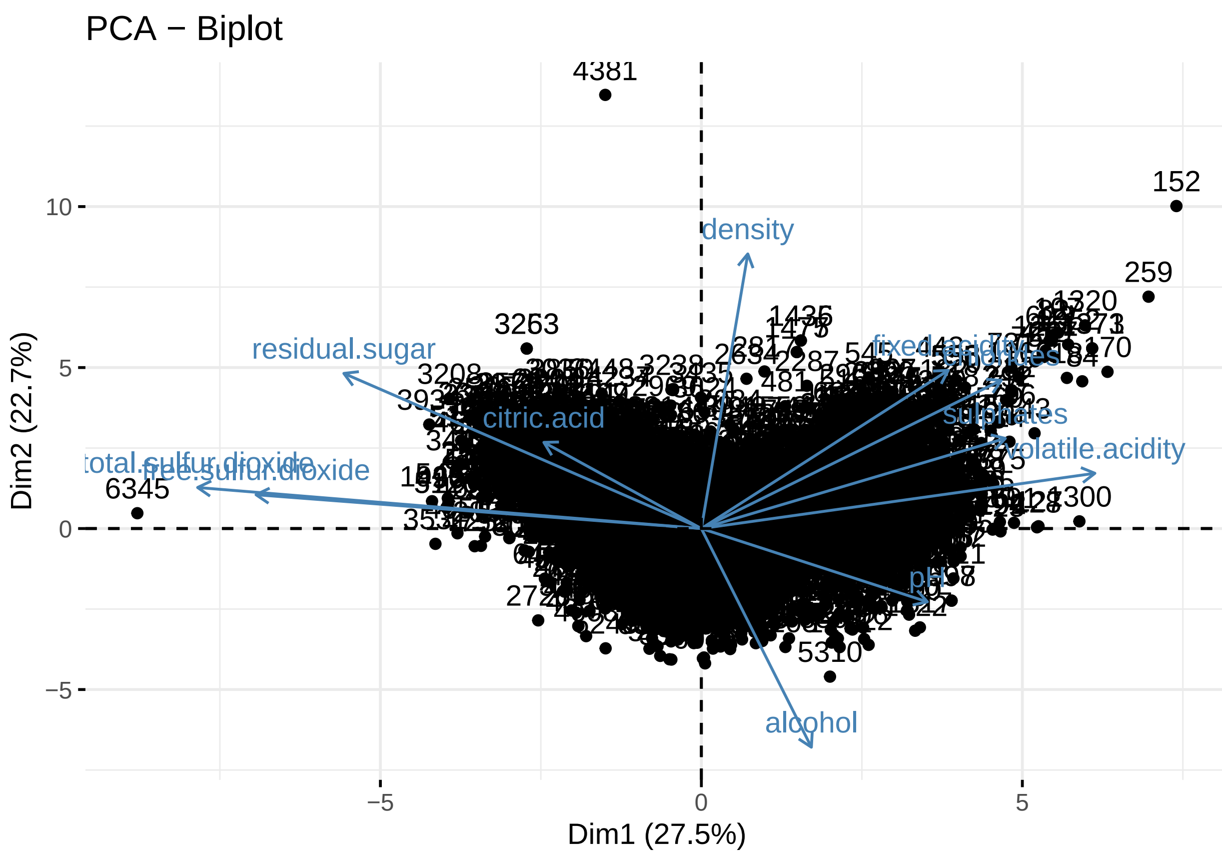
```r
## Scree plot of eigenvalues
fviz_screeplot(wine_pca, addlabels =T)
```



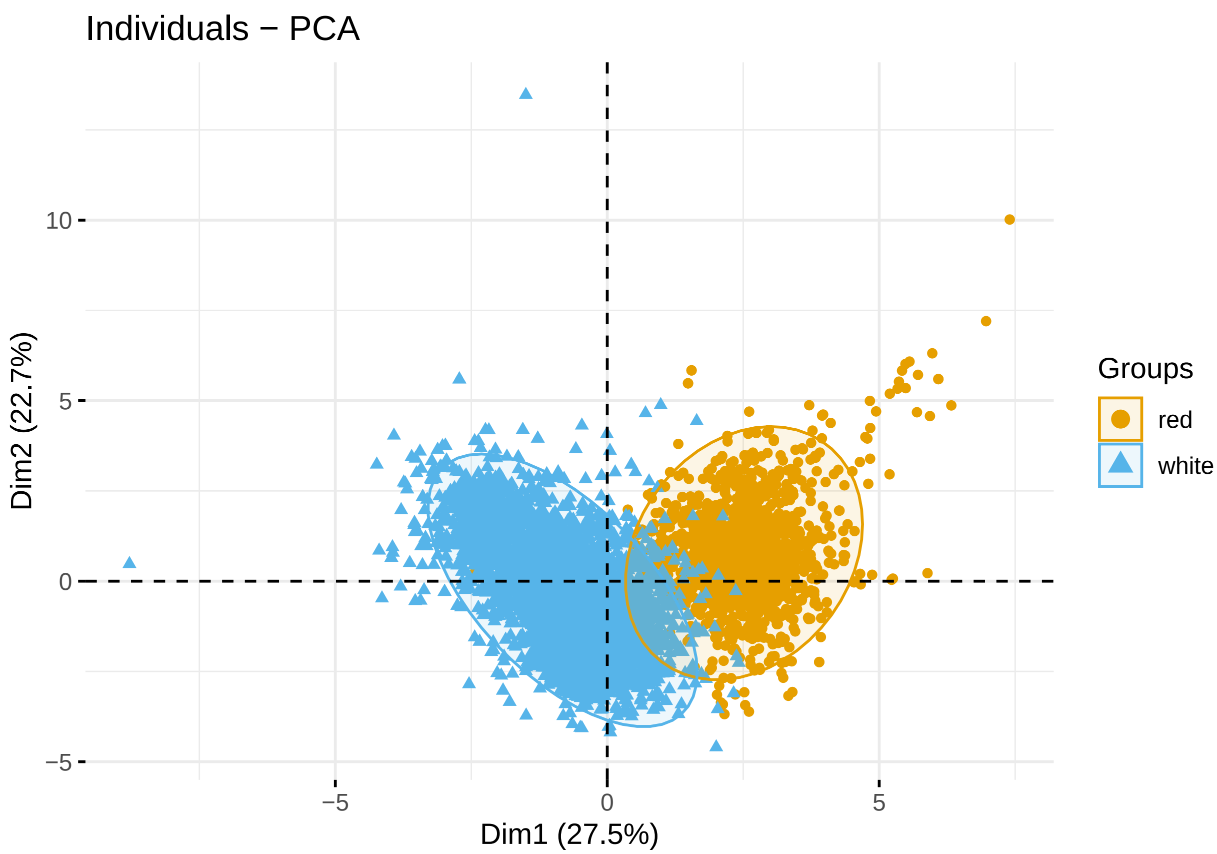

```
screeplot(wine_pca, type = "line", main = "Scree Plot") # Looks like we should retain prin. comps. 1-3
```




```r
## Biplot
fviz_pca_biplot(wine_pca)
```

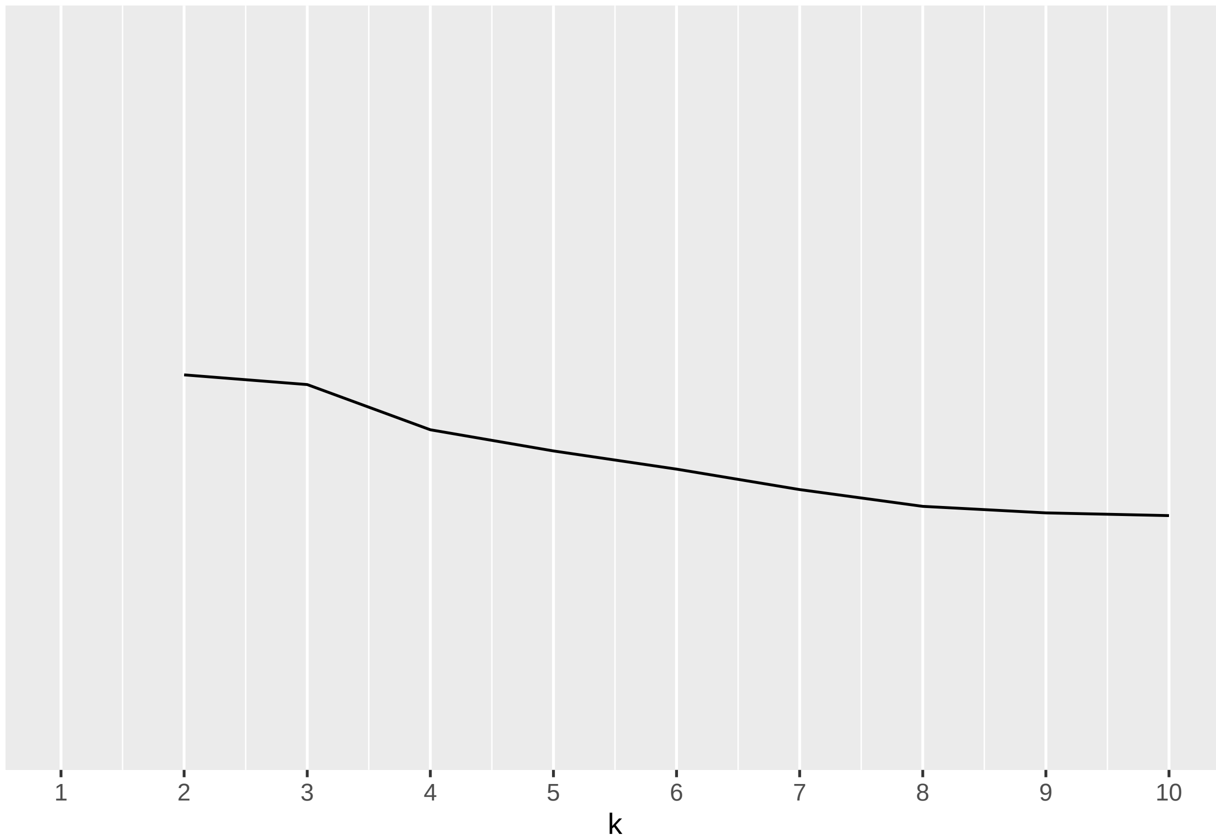



```r
## Plot individuals
fviz_pca_ind(wine_pca, label = "none", habillage = wine$color, addEllipses = T, palette = c("#E69F00",
```

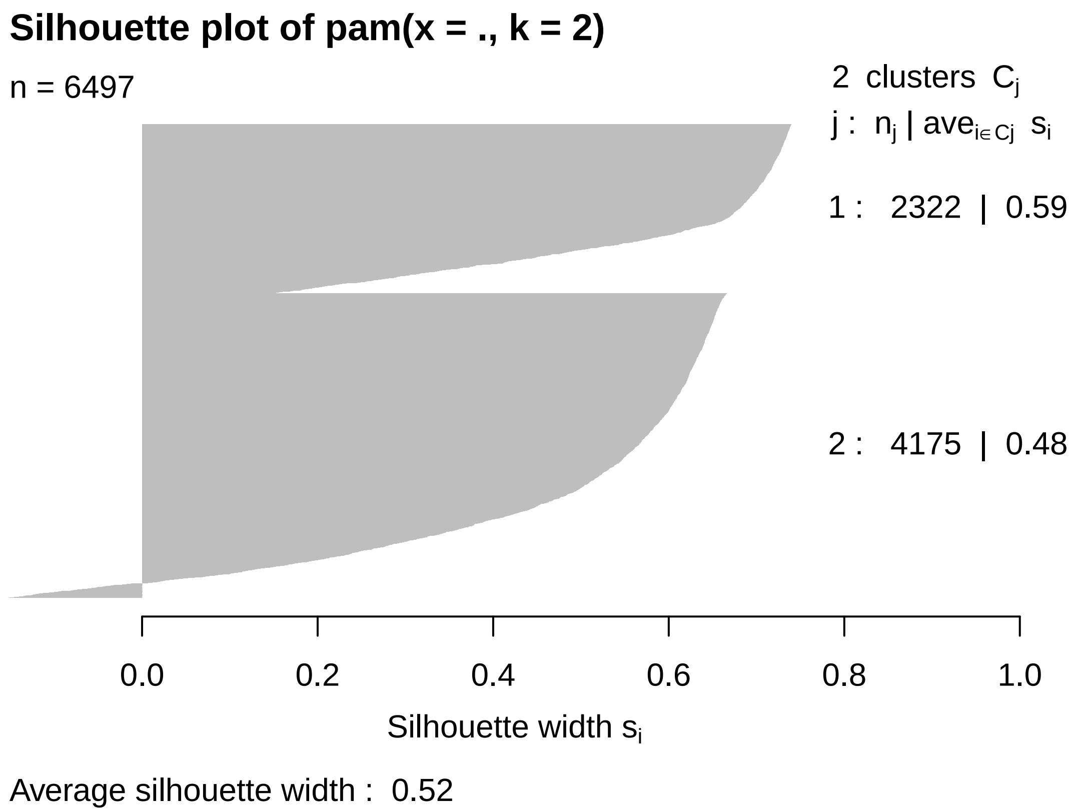



```r
### PAM Clustering
pam_wine <- wine %>% pam(2)

sil_width <- vector()
for(i in 2:10){
  pam_fit <- wine %>% pam(i)
  sil_width[i] <- pam_fit$silinfo$avg.width
}

ggplot() + geom_line(aes(x = 1:10), y = sil_width) + scale_x_continuous(name = "k", breaks = 1:10)
```

```
## Warning: Removed 1 rows containing missing values (geom_path).
```

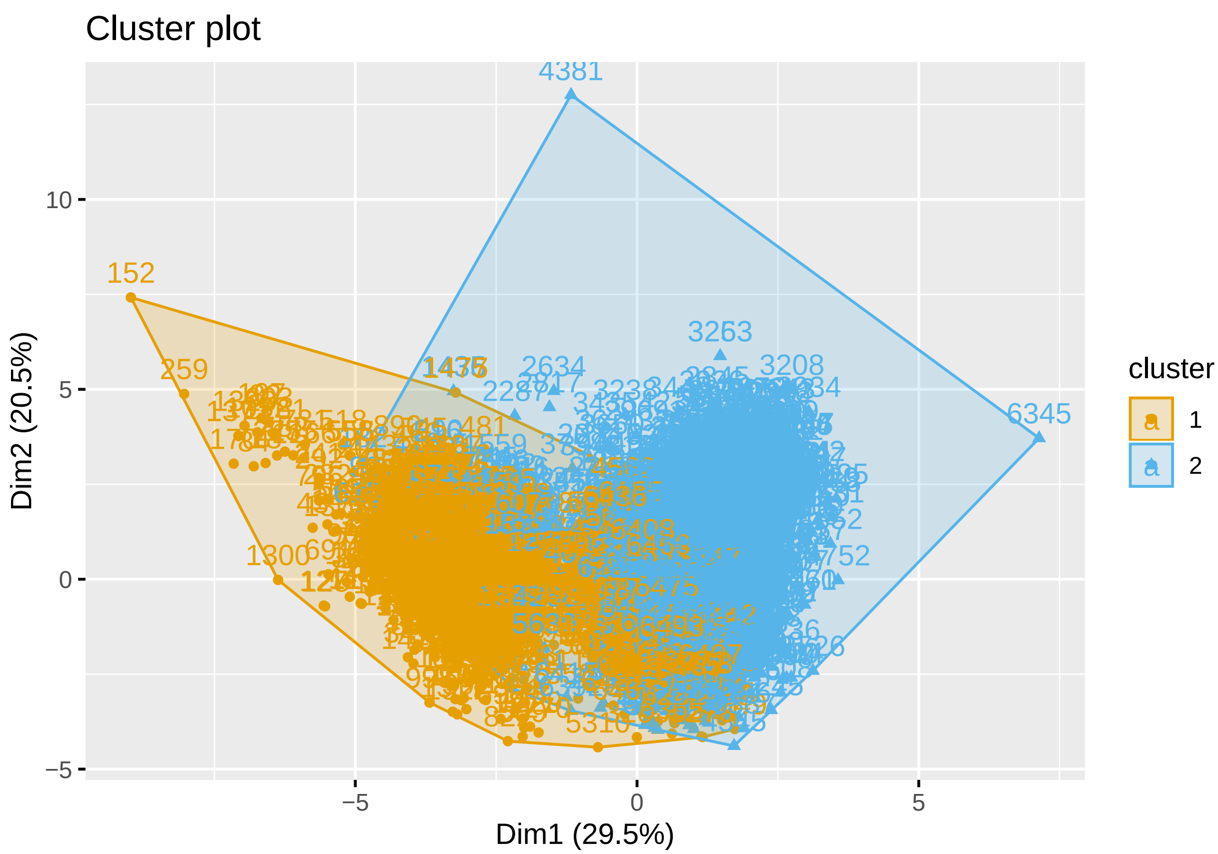



```
plot(pam_wine, which = 2) # 0.52 - Reasonable structure found, but really weak
```




```
pamclust <- wine %>% mutate(cluster = as.factor(pam_wine$clustering))
pamclust %>%
  ggplot(aes(residual.sugar, volatile.acidity, color = cluster)) +
  geom_point() +
  scale_color_manual(values = c("1" = "#E69F00", "2" = "#56B4E9"))
```


```r
fviz_cluster(pam_wine, palette = c("#E69F00", "#56B4E9"))
```



Conducting PAM clustering on the wine data set, we see that individual observations are clustered around wine color. From the silhouette plot, the average silhouette width is 0.52, indicating that a "barely" reasonable structure has been found after setting the clusters to k = 2.

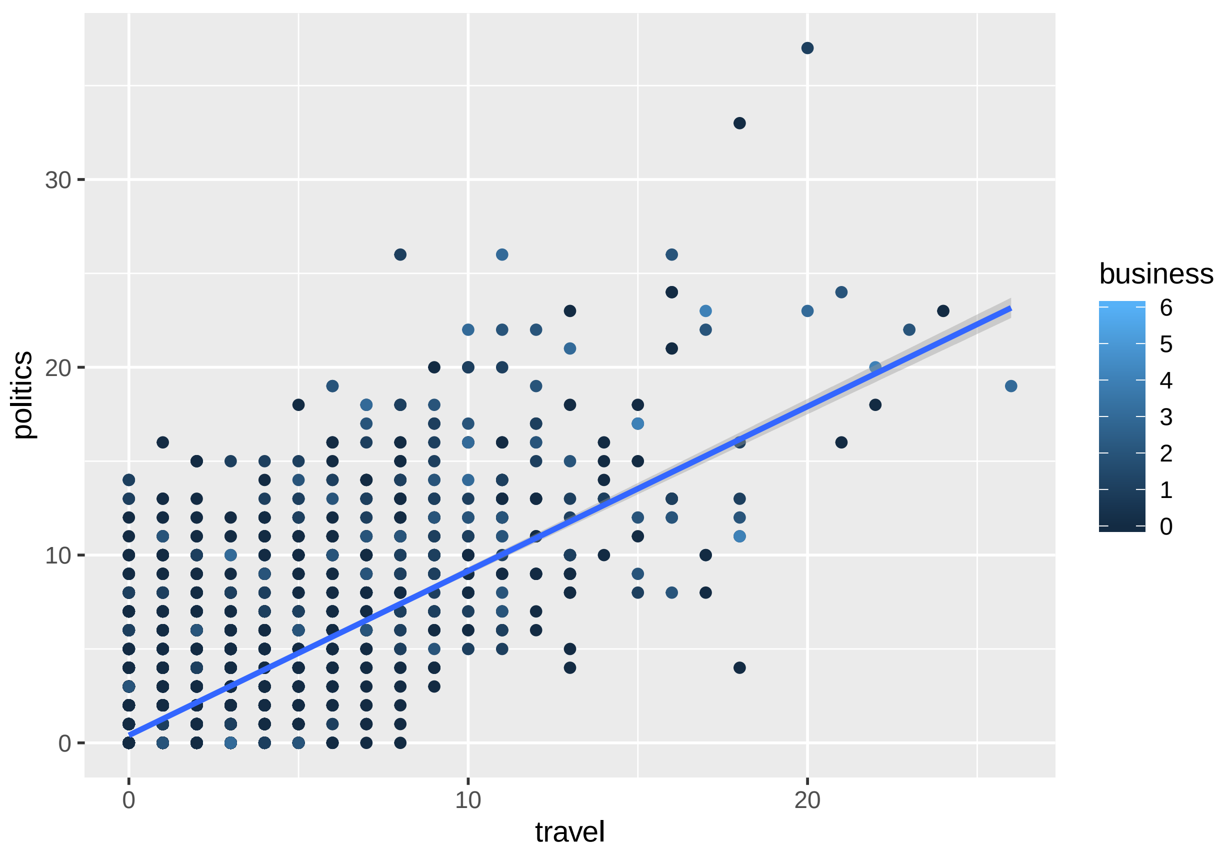
Question 4: Market Segmentation

```
sm <- read.csv("https://github.com/jgscott/SDS323/raw/master/data/social_marketing.csv")

sm_fit2 <- lm(politics ~ tv_film + travel + business + (tv_film*travel) + (travel*business) +
                (tv_film*business), data = sm)
summary(sm_fit2)
```

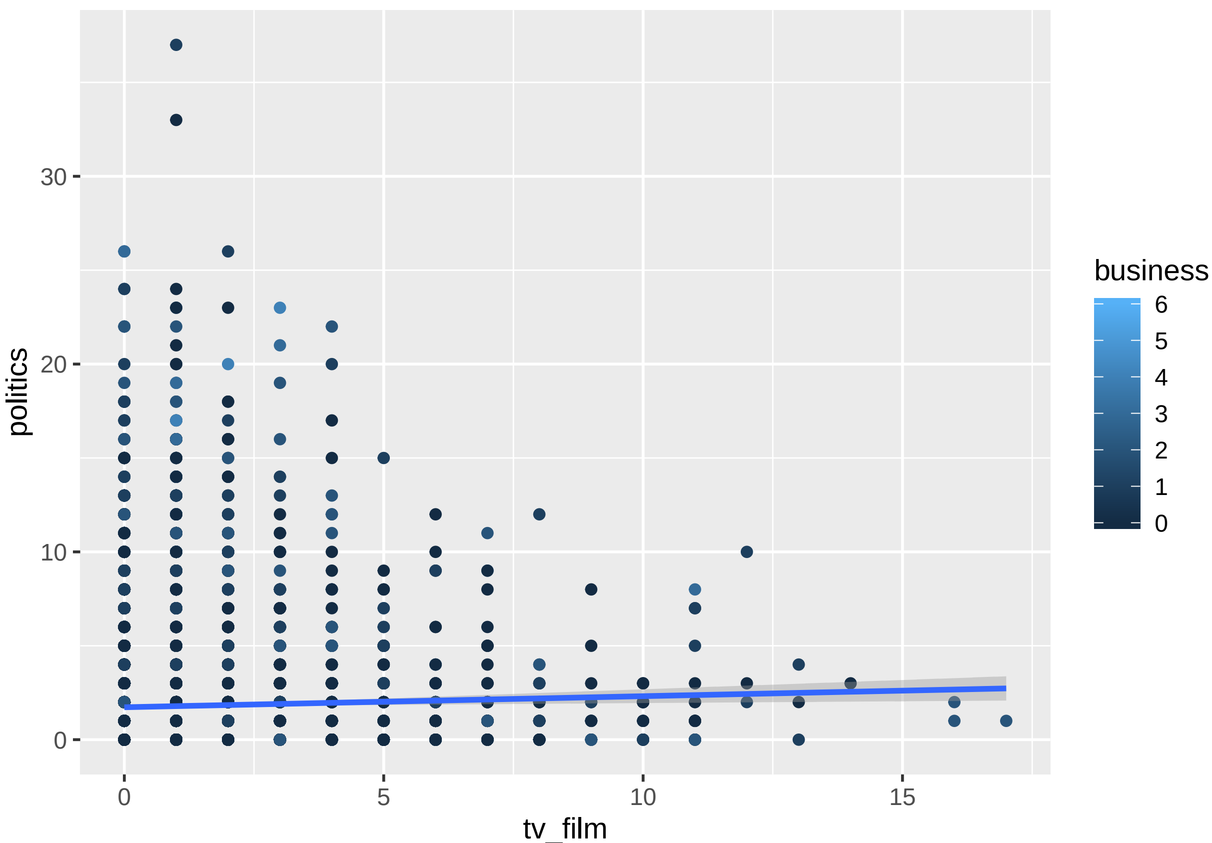
```
##
## Call:
## lm(formula = politics ~ tv_film + travel + business + (tv_film *
##     travel) + (travel * business) + (tv_film * business), data = sm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.3368  -1.2447  -0.4185   0.5815  18.7035
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.418479   0.041937   9.979  < 2e-16 ***
```

```
## tv_film           0.028075   0.023582   1.191  0.23386
## travel            0.826221   0.016528  49.989  < 2e-16 ***
## business          0.074455   0.049839   1.494  0.13524
## tv_film:travel   -0.028652   0.007181  -3.990 6.67e-05 ***
## travel:business   0.093730   0.010992   8.527  < 2e-16 ***
## tv_film:business -0.050446   0.019227  -2.624  0.00872 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.258 on 7875 degrees of freedom
## Multiple R-squared:  0.4455, Adjusted R-squared:  0.4451
## F-statistic:  1055 on 6 and 7875 DF,  p-value: < 2.2e-16
```

```r
ggplot(data = sm, aes(x = travel, y = politics)) +
  geom_point(aes(color = business)) +
  geom_smooth(method = "lm")
```




```r
ggplot(data = sm, aes(x = tv_film, y = politics)) +
  geom_point(aes(color = business)) +
  geom_smooth(method = "lm")
```



Controlling for businness and tv/film, there is a significant effect of travel on politics. When a tweet is categorized as travel, political tweets increase by 0.826 on average, t = 49.89, df = 7875, p < .001. The effect of tv/film tweets on political tweets is different for different values of travel tweets t = -3.990, df = 7875, p < .001. Likewise, the effect of travel tweets on political tweets is different for different values of business tweets, t = 8.527, p <.001. Finally, the effect of tv/film tweets on political tweets is different for different values of business tweets, t = -2.624, p < .05.