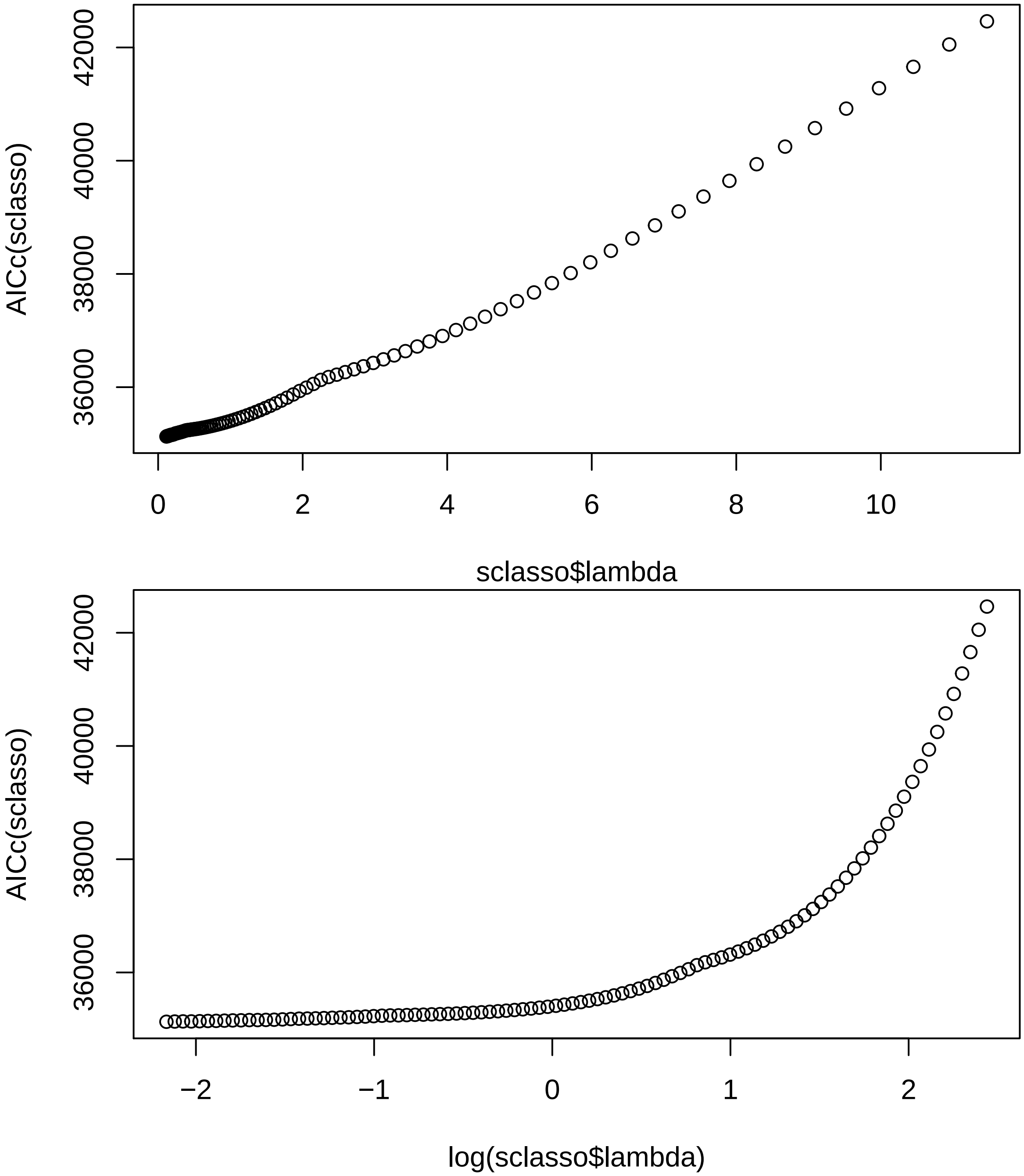
exercise3

Question 1 (collaboration on code with Linh Nguyen)

Using a singlo lasso and a sparse matrix was determined to be the most effective way to model price. Using these methods, the resulting graphs were created (used to optimize AIC):



In order to to quantify the average change in rental income per square foot associated with green certification, and to assess whether the "green certification" effect is different for different buildings, we display the coefficients from the lasso model:

```
## 50 x 1 sparse Matrix of class "dgCMatrix"
##                                               seg100
## intercept                              -5.001999e+00
## CS_PropertyID                            .
## cluster                                 5.467304e-04
## size                                    5.839770e-06
## empl_gr                                 1.304135e-02
## leasing_rate                            6.175159e-03
## stories                                  .
## age                                    -1.029233e-02
## renovated                              -7.579708e-02
## class_a                                 1.860125e+00
## class_b                                 2.354109e-01
## LEED                                    6.404668e-01
## Energystar                               .
## green_rating                             .
## net                                    -1.817368e+00
## amenities                               4.396783e-01
## cd_total_07                            -9.643548e-05
## hd_total07                              2.827175e-04
## total_dd_07                              .
## Precipitation                            .
## Gas_Costs                              -1.169901e+02
## Electricity_Costs                       9.112882e+01
## cluster_rent                            1.028235e+00
## paymentnet contract                    -3.997372e-03
## green_typeNon-green                      .
## data$green_rating                        .
## CS_PropertyID:data$green_rating          .
## cluster:data$green_rating               3.737705e-05
## size:data$green_rating                 -2.864685e-07
## empl_gr:data$green_rating                .
## leasing_rate:data$green_rating           .
## stories:data$green_rating                .
## age:data$green_rating                    .
## renovated:data$green_rating              .
## class_a:data$green_rating                .
## class_b:data$green_rating               3.899673e-01
## LEED:data$green_rating                  1.062111e-01
## Energystar:data$green_rating             .
## green_rating:data$green_rating           .
## net:data$green_rating                    .
## amenities:data$green_rating            -2.055393e-01
## cd_total_07:data$green_rating            .
## hd_total07:data$green_rating             .
## total_dd_07:data$green_rating            .
## Precipitation:data$green_rating          .
## Gas_Costs:data$green_rating              .
## Electricity_Costs:data$green_rating      .
## cluster_rent:data$green_rating          2.748034e-02
## paymentnet contract:data$green_rating    .
## green_typeNon-green:data$green_rating    .
```

Suprisingly, using the coefficient of 0 from “data$green_rating”, it appears rent does not differ as a whole between green and non-green apartments. However, we can see small differences in rent based upon certain characteristics. For example, rent increases by .39 units when the apartment is a green class b apartment. The rest of these effects are displayed on the above matrix. Every coefficient within the "green_rating" section shows how much rent differs based upon every individual attribute when the apartment is green.

Question 2

1. This method will not work because cities of different sizes will only have different amounts of cops due to differing crime rates. Since bigger cities with high crime rates need more cops, this cannot prove that a higher amount of cops help suppress crime more effectively.

2. In order to isolate the effect, the researchers found a specific example where the amount of cops increased for reasons besides crime rates. The example was having increased police force in Washington D.C. when the threat of terrorism was heightened, known as a high-alert day. When the data was examined, it was revealed that crime did go down on these high-alert days. Holding all else fixed, when it was a high-alert day, crime decreased by approximately 6.046 units, controlling for the METRO ridership.

3. The researchers needed to control for METRO ridership in order to determine if less people would be out and about (hence less victims for criminals) due to the elevated terror level. They found that the same numbers of people were riding the METRO on high-alert days compared to any other day.

4. Table 4 explores how the number of crimes on high-alert days decreased in a specific area, district 1. This was achieved via two interaction variables: high-alert*district 1, and high-alert*other districts. The dummy variables would only be activated if the data took place in the district of interest. So, this allowed for comparison between how much crime decreased in district 1 versus all of the other districts. It appears district 1 saw crime decrease by more than the other districts as a whole, as on high-alert days crime in district 1 decreased by 2.621 units and in all other districts on high-alert day crime decreased by .571 units (holding all else fixed).