STAT 217 Final Exam

Instructions:

This document contains the questions for the final exam. Note that you do not need to answer all of these questions, but you should spend some time before November 18 familiarizing yourself with the questions and organizing materials that you will need.

On November 18, you will be randomly assigned questions from the list below. See D2L for your randomized questions. Do not assume that you have been assigned the same questions as others you may frequently work with.

After preparing answers for your assigned questions, you are to record a video (no more than six minutes long) that addresses each question. You should start your video by giving your name. Then state the part and question number before giving your answer and explanation. Make sure all assigned questions are addressed. End your video by addressing anyone that you worked with on this exam. Failure to acknowledge collaboration will result in a zero and may lead to a report to the Dean of Students.

TechSmith must be used to record videos. A link to your recorded video is due to Gradescope by November 25 at noon. Your submission will be graded on four parts (two questions, addressing collaborative efforts, and time).

Resources: You are allowed all provided materials (videos, notes, textbook, etc.) and access to discussions with other students, the MLC, other tutors, and can ask questions of your instructor. You must document the discussions/help that you have (other than with your instructor) in preparing your answer.

See full set of instructions along with information for recording with TechSmith in Final Exam Announcement posted on D2L.

Part I (15 points):

Choose the most appropriate analysis from the following list of options: two-sample mean test, One-Way ANOVA, Two-Way ANOVA, Chi-squared independence, Chi-squared homogeneity, Simple Linear Regression, Multiple Linear Regression with quantitative predictors, or Multiple Linear Regression with a mix of quantitative and categorical predictors. In your recorded video, **explain your choice of analysis in context** *and how the variables will be used in the model/analysis (explanatory/response, categorical/quantitative)*. Then discuss the **scope of inference for the study in context**.

1) A researcher is interested in the impact of the pea leaf weevil on pea yields. He designs a study with three levels of an insecticidal seed treatment (none, 30 grams/100 kg seed, and 50 grams/100 kg of seed) and three times at which a foliar insecticide is sprayed (none, early season spray, late season spray). A single field in Gallatin County is chosen for the study and the combinations of the treatments (seed and foliar) are randomly assigned to plots within the field in a completely randomized design. At the end of the season, the yield in kg per hectare is measured for each plot. There are 10 replicates at each combination of treatments, so n=90, which covers the entire field in plots. He does not expect the insecticide to have different impacts on yield based on the foliar spray timing, but will check that first.

2) A researcher is interested in studying milfoil presence in a set of four lakes (Ham, North Arm, Otter, and Schmidt) and how its abundance may differ across them. She measures 40 plots in each lake, with the location of the plots randomly selected from a grid arrangement of all possible plots in each lake

(so 40 plots in four lakes and n=160). The response is measured as pounds of dry weight of milfoil harvested at each plot.

3) A researcher is studying fish use of a fishway that can be used to bypass a diversion dam on a large river and whether release location is related to use of the bypass in the Yellowstone River. Fish are electroshocked (all fish caught are used but the fish to catch are based on what happened to be in an area when the machine was turned on) and released with an electronic tag. There are six different release locations on the river (the locations start upriver of the bypass and include a location around a river bend from the bypass downriver). They roll a die to decide which location each fish will be released at. He records whether the fish actually used the bypass or not after being released. There were n=1800 fish measured with approximately 300 released at each location; each fish either used the bypass or didn't.

4) A researcher is studying willow heights in Yellowstone National Park. She has a data set to analyze that is based on a long term study. Four different sites were chosen for study 30 years ago. In the initial year of the study, 20 trees were randomly sampled from all of the trees in each site and tagged for monitoring. Every year researchers revisit the willow trees that were previously tagged and measure their height (in cm). The total data set includes n= 2400 height measurements (20 trees, 30 years, 4 sites). The researchers are interested in assessing changes over time and whether the changes over time are different in the four locations.

5) A researcher is interested in studying factors that might relate to falls in elderly subjects, and in particular medical costs associated with those falls. He obtains a list of all the residents in a large (3,000 resident) nursing home complex in Florida and takes a random sample of n=100 of the residents. He obtains the age, height, weight, and number of falls in the previous five years for each selected resident. Then he follows up after a year and obtains the total medical costs (in dollars) for each participant related to new falls over that year. He is interested in building a model to predict yearly costs from new falls based on these variables.

6) A researcher is studying impacts of insecticides on bees. She randomly samples n=30 different fields from 300 different fields that could be suitable for doing her study in Northwestern Iowa. In each field she measures the bee abundance by putting out a trap to get a count of bees on the day the study is conducted (same day for all fields). The counts ranged from 3 to 250 and were usually over 100. She also obtains information on whether the field was sprayed with an insecticide or not (it ended up that 10 fields were sprayed and 20 were not). She also measures the type of crop planted in the field (wheat, soybeans, or peas). She suspects that bee abundance would be impacted differently by insecticide in the different types of crops.

7) A snow science researcher is interested in study aspects of Bridger Bowl that are related to more or less dense snow. He creates a grid of points across all the areas that are not out of bounds and are safely skiable. Then he randomly samples n=300 points to go to out of 17,000 possible locations. He measures the snow density (in kg/m^3) and snow depth (inches) when he gets to each location during the course of one day. He suspects that the elevation (in meters) of the locations, slope angle (in degrees), and snow depth (inches) might all be related to snow density.

8) A researcher is interested in assessing the relationship between the modulus (elasticity in Newtons per square meter) and thickness (in millimeters) of the thorax of a moth, under the assumption that differences in thickness might be related to differences in the modulus. A single moth is obtained for measurements and locations for testing across the thorax area are randomly selected from all possible testing areas using a random number generator for the locations. A total of n=480 different locations are selected.

Part II (15 points): Predicting Power Output in a Power Plant

- Read Tufekci (2014) to prepare for the questions below. They are interested in predicting power output using a suite of variables. With some minor edits, here is their "metadata" discussion. This is complemented by the discussion on p. 127 that contains units for the variables:

This dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (PE) of the plant. A combined cycle power plant (CCPP) is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the Vacuum is collected from and has an effect on the Steam Turbine, the other three of the ambient variables affect the GT performance.

Features consist of hourly average ambient variables:

- Temperature (`AT`) in the range 1.81°C and 37.11°C,
- Ambient Pressure (`AP`) in the range 992.89-1033.30 millibar,
- Relative Humidity (`RH`) in the range 25.56% to 100.16%,
- Exhaust Vacuum (`V`) in the range 25.36-81.56 cm Hg, and
- Net hourly electrical energy output (`PE`) 420.26-495.76 MW.

The averages are taken from various sensors located around the plant that record the ambient variables every second.

- Tufekci, P. (2014) "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods". *Electrical Power and Energy Systems*, 60, 126-140. https://www.sciencedirect.com/science/article/abs/pii/S0142061514000908?via%3Dihub

The author considers a wide array of models and methods (some we've never heard of) to compete with linear regression techniques. We will focus on the linear regression and linear model methods that might be applicable to modeling this data set.

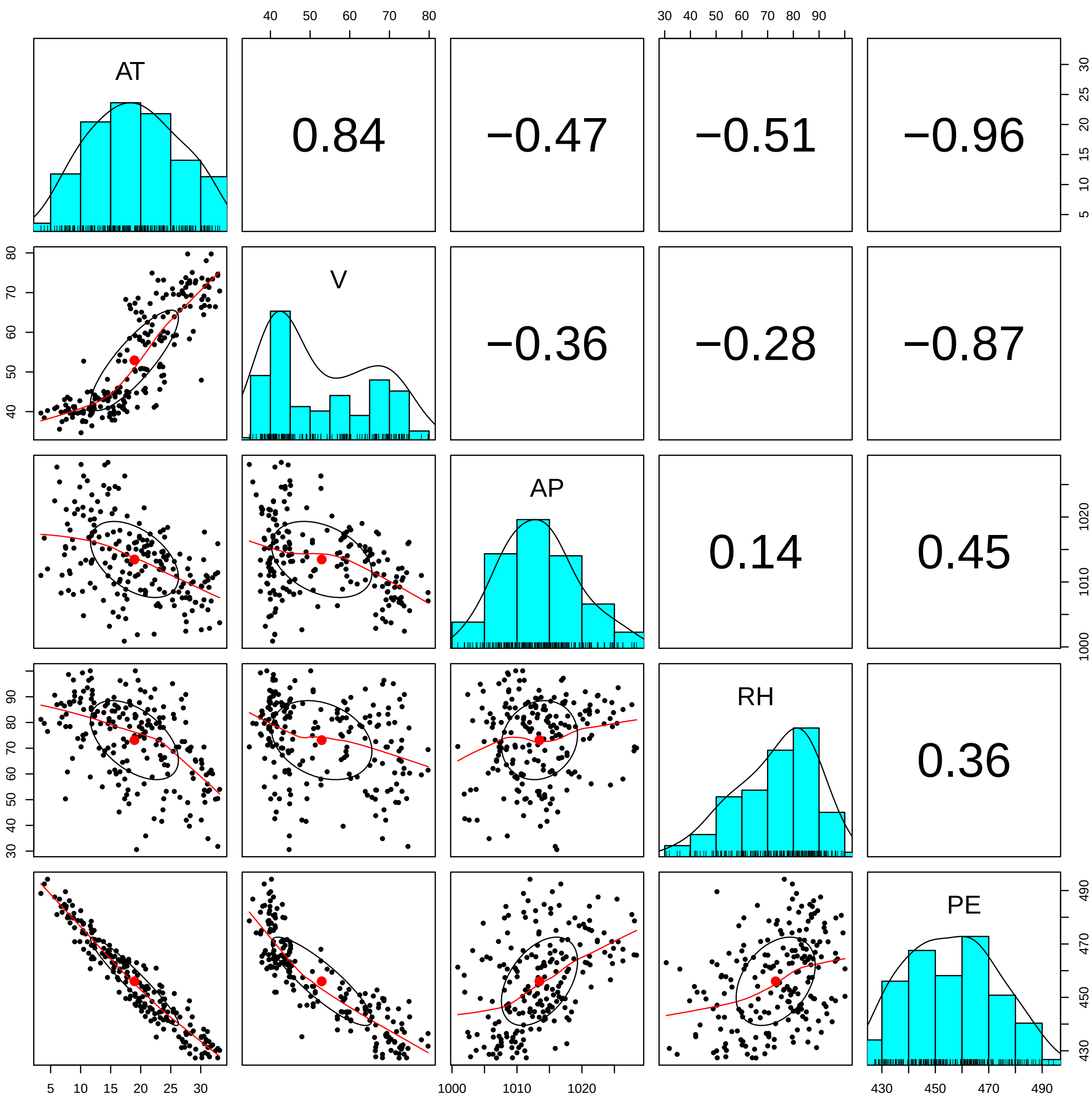
Note: We are surprised that this was published in its current state. We expect more from you in understanding and explaining regression models and the notation is not correct and actually quite confusing. A PDF of the paper has been posted on D2L. If you read the paper, take it with a grain of salt and use the lens of our discussions to critically assess and focus on just their Sections 1, 2.1, 2.2, 2.3 through the first column of page 129, and Section 3.1. We will make our own conclusions about the models we consider based on the subset of the data you will analyze.

For each question below, you will work with a random sample of all the observations collected over the 5 years.

```
suppressMessages(library(readxl))
fulldata <- read_excel("Full_powerplant.xlsx", sheet = 1)
ids <- 1:length(fulldata$PE)
set.seed(1234)
Q1 <- fulldata[sample(ids, 200),]
Q2 <- fulldata[sample(ids, 200),]
Q3 <- fulldata[sample(ids, 200),]
```

For 9-14, we would like to build a model that predicts the energy output with all four predictors as defined above. This analysis will use the `Q1` subset of the full data to fit the model stored in `lm1`.

```
library(psych)
pairs.panels(Q1)
```




```
lm1 <- lm(PE ~ AT + V + AP + RH, data= Q1)
summary(lm1)
```

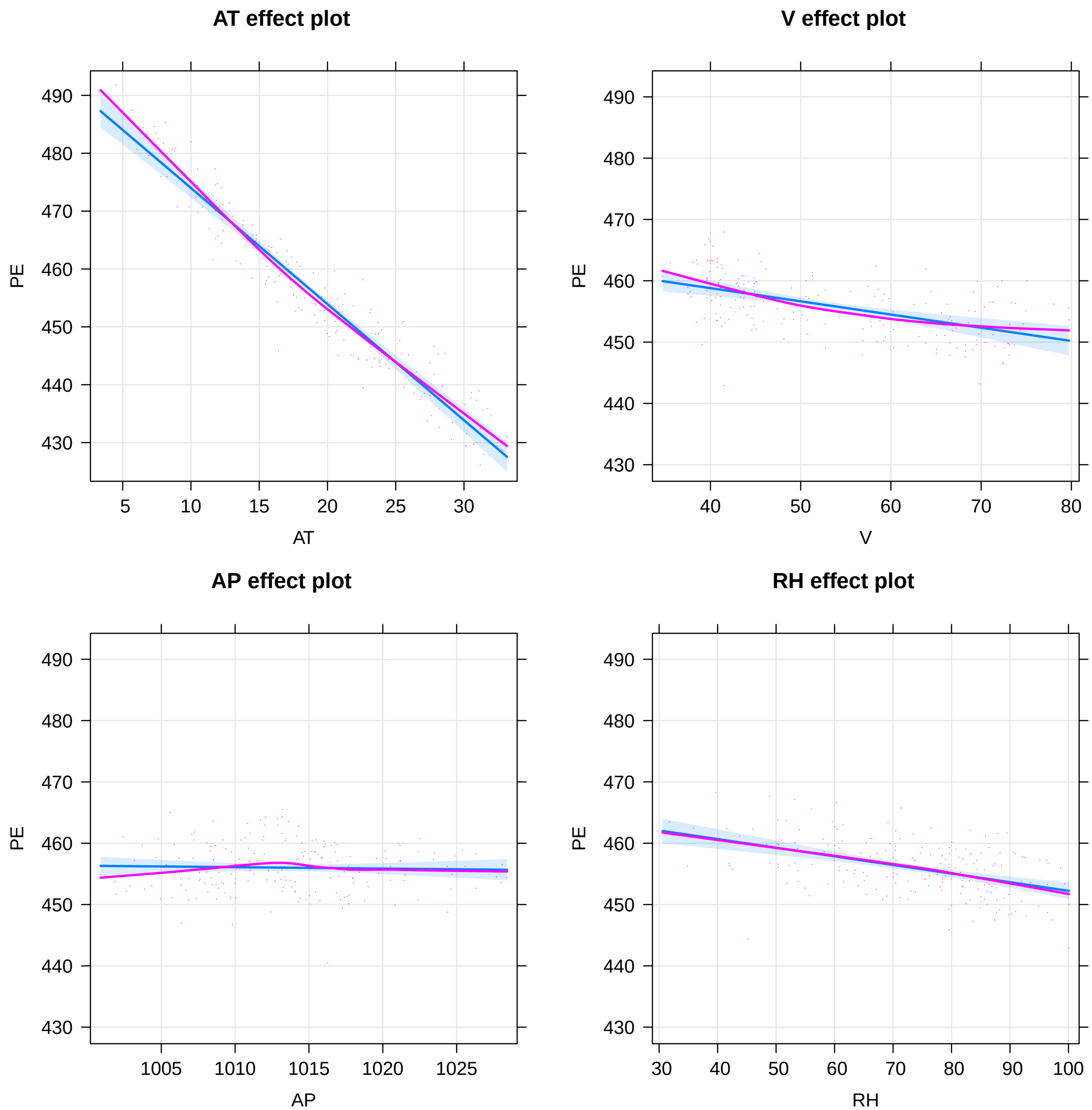
```
##
## Call:
## lm(formula = PE ~ AT + V + AP + RH, data = Q1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.5350  -2.7742   0.0714   2.6765   9.4827
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 538.95186   57.10308   9.438  < 2e-16
```

```
## AT           -2.00658    0.09066 -22.134  < 2e-16
## V            -0.21538    0.04414  -4.879 2.21e-06
## AP           -0.02290    0.05570  -0.411    0.681
## RH           -0.14057    0.02289  -6.140 4.53e-09
##
## Residual standard error: 3.998 on 195 degrees of freedom
## Multiple R-squared:  0.9429, Adjusted R-squared:  0.9417
## F-statistic: 804.4 on 4 and 195 DF,  p-value: < 2.2e-16
```

```r
library(car)
vif(lm1)
```

```
##       AT        V       AP       RH
## 5.494155 3.892240 1.328219 1.542985
```

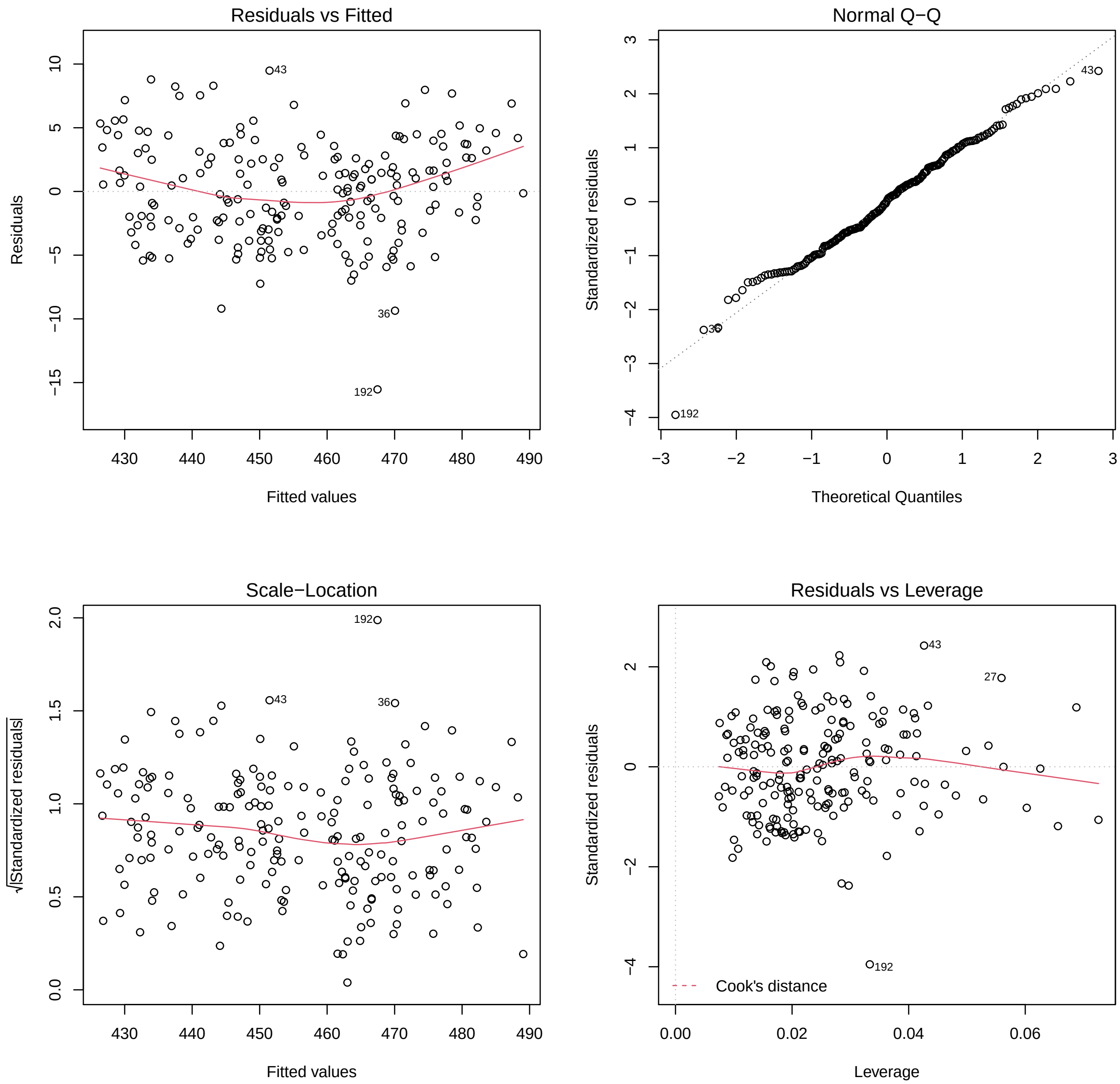
```r
library(effects)
plot(allEffects(lm1, residuals=T), grid=T, partial.residuals=list(pch=16, cex=.1))
```




```r
confint(lm1)
```

```
##                   2.5 %       97.5 %
## (Intercept) 426.3329314 651.57078392
## AT           -2.1853693  -1.82778875
## V            -0.3024392  -0.12832479
## AP           -0.1327522   0.08695266
## RH           -0.1857185  -0.09541902
```

```r
par(mfrow=c(2,2))
plot(lm1)
```




```r
summary(Q1)
```

```
##        AT              V               AP              RH        
##  Min.   : 3.38   Min.   :34.69   Min.   :1001   Min.   : 30.59  
##  1st Qu.:13.25   1st Qu.:41.26   1st Qu.:1009   1st Qu.: 61.50  
##  Median :19.01   Median :48.56   Median :1013   Median : 76.31  
##  Mean   :18.96   Mean   :52.93   Mean   :1013   Mean   : 73.14  
##  3rd Qu.:23.97   3rd Qu.:65.20   3rd Qu.:1017   3rd Qu.: 85.05  
##  Max.   :33.15   Max.   :79.74   Max.   :1028   Max.   :100.09  
##        PE       
##  Min.   :427.3  
##  1st Qu.:443.6  
##  Median :455.6  
##  Mean   :456.0  
```

```
##   3rd Qu.:467.4
##   Max.   :494.2
```

9) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. For that model, discuss the potential issues with multicollinearity, referencing plots and numerical results. Interpret the VIF for the most impacted variable in context. Find R-squared for this model and give an interpretation in context.

10) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. For that model, interpret one of the slope coefficients in context and report the confidence interval for that slope coefficient. Is 0 a plausible value for this coefficient? Why? Find R-squared for this model and give an interpretation in context.

11) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. Assess all validity conditions for this model except for independence. Make sure you discuss which plots/information you are using to assess each assumption and include a discussion of the effects plots in this assessment.

12) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. For `AT`, `V`, and `AP` at their mean values and `RH` of 40 and 41, what are the predicted mean `PE`s? Explain how you calculated those. Find R-squared for this model, give an interpretation in context, and discuss whether this seems like a good model based on this.

13) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. Give a conclusion, in context, for `AP` in this model. Based on this result, what would you do next? If you dropped it and re-fit the model, would you expect this to increase or decrease R-squared? Why?

```r
library(MuMIn)
options(na.action = "na.fail")
dredge(lm1, rank="AIC", extra = c("R^2"))
```

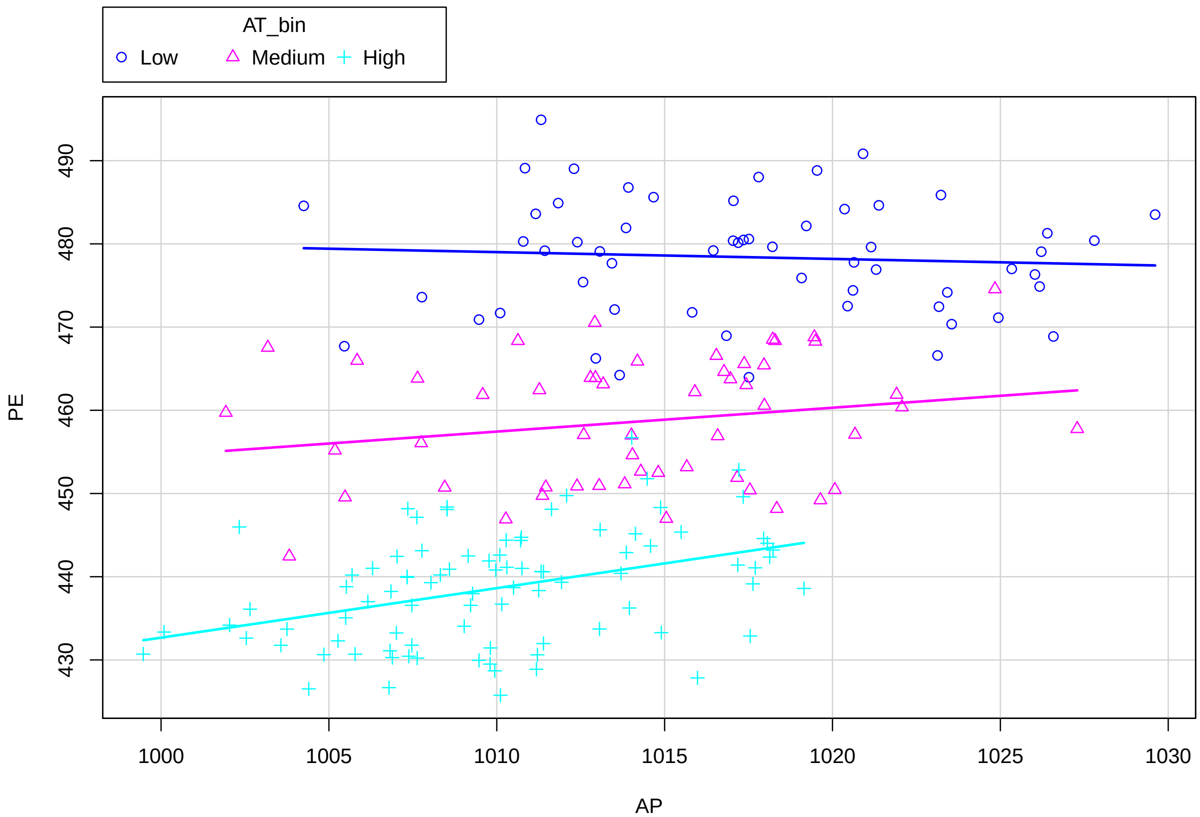
```
## Global model call: lm(formula = PE ~ AT + V + AP + RH, data = Q1)
## ---
## Model selection table
##    (Intrc)        AP     AT      RH       V    R^2 df   logLik    AIC  delta
## 15  515.50           -1.993 -0.1390 -0.2177 0.9428  5 -558.505 1127.0   0.00
## 16  539.00 -0.022900 -2.007 -0.1406 -0.2154 0.9429  6 -558.418 1128.8   1.83
## 7   513.40           -2.348 -0.1752         0.9356  4 -570.442 1148.9  21.87
## 8   572.90 -0.058110 -2.373 -0.1782         0.9359  5 -569.937 1149.9  22.87
## 11  504.70           -1.719         -0.3047 0.9317  4 -576.265 1160.5  33.52
## 12  469.40  0.034640 -1.703         -0.3067 0.9318  5 -576.094 1162.2  35.18
## 3   497.00           -2.162                 0.9158  3 -597.131 1200.3  73.25
## 4   494.10  0.002829 -2.161                 0.9158  4 -597.130 1202.3  75.25
## 14   64.14  0.430400         0.1359 -1.0250 0.7993  5 -684.049 1378.1 251.09
## 10   60.93  0.445600                -1.0680 0.7846  4 -691.114 1390.2 263.22
## 13  503.50                   0.1425 -1.0940 0.7791  4 -693.652 1395.3 268.30
## 9   516.40                          -1.1420 0.7629  3 -700.722 1407.4 280.44
## 6  -736.60  1.153000         0.3308         0.2952  4 -809.655 1627.3 500.30
## 2  -831.70  1.271000                        0.2025  3 -822.009 1650.0 523.01
## 5   427.50                   0.3905         0.1315  3 -830.539 1667.1 540.07
## 1   456.00                                  0.0000  2 -844.640 1693.3 566.27
##    weight
## 15  0.714
## 16  0.286
## 7   0.000
```

```
## 8   0.000
## 11  0.000
## 12  0.000
## 3   0.000
## 4   0.000
## 14  0.000
## 10  0.000
## 13  0.000
## 9   0.000
## 6   0.000
## 2   0.000
## 5   0.000
## 1   0.000
## Models ranked by AIC(x)
```

14) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lm1`) for `PE`. Examine the AIC results provided by `dredge` with this model. List the variables in the top model according to AIC. Describe any models with similar support as the top model and explain how you are identifying them. If there are models with similar support, which should we consider as our final model? Find R-squared for the top AIC model and give an interpretation in context.

For 15-19, we would like to build a model that allows the relationship between energy output and ambient pressure to vary by temperature, where temperature (`AT_bin`) has been categorized as "Low," "Medium," and "High." This analysis will use the `Q2` subset of the full data to fit the model stored in `lmI_P`.

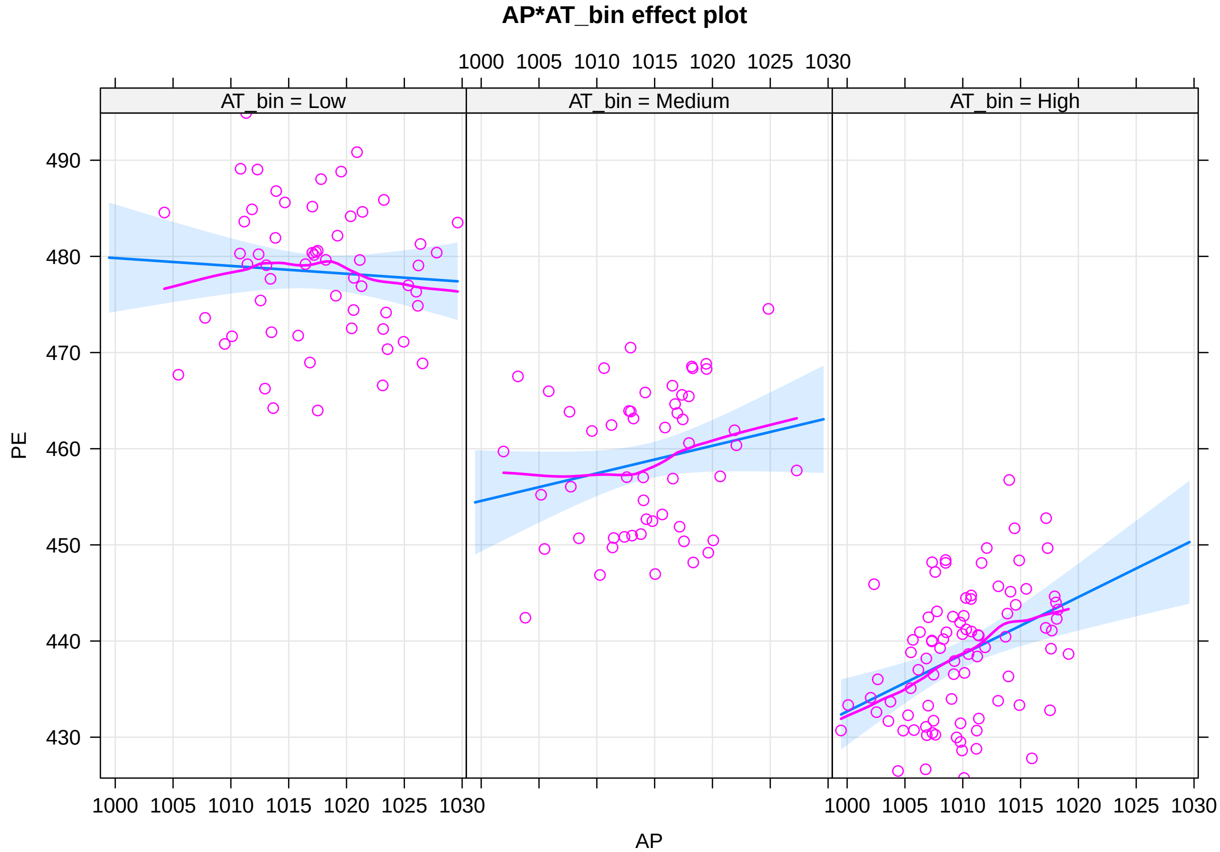
```r
Q2$AT_bin <- factor(with(Q2, cut(AT, breaks = c(4.32, 14, 21, 34.30), include.lowest=T)))
levels(Q2$AT_bin) <- c("Low", "Medium", "High")

scatterplot(PE ~ AP|AT_bin, data=Q2, smooth=F)
```




```r
lmI_P <- lm(PE ~ AP*AT_bin, data=Q2)
summary(lmI_P)
```

```
##
## Call:
## lm(formula = PE ~ AP * AT_bin, data = Q2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.4814  -5.6025   0.4655   4.8650  16.0077
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       561.4244   155.3073   3.615 0.000383
## AP                 -0.0816     0.1526  -0.535 0.593507
## AT_binMedium     -393.6049   235.4805  -1.671 0.096236
## AT_binHigh       -723.5599   225.1018  -3.214 0.001531
## AP:AT_binMedium    0.3684     0.2318   1.589 0.113716
## AP:AT_binHigh      0.6764     0.2221   3.046 0.002643
##
## Residual standard error: 6.824 on 194 degrees of freedom
## Multiple R-squared:  0.8632, Adjusted R-squared:  0.8597
## F-statistic: 244.9 on 5 and 194 DF,  p-value: < 2.2e-16
```

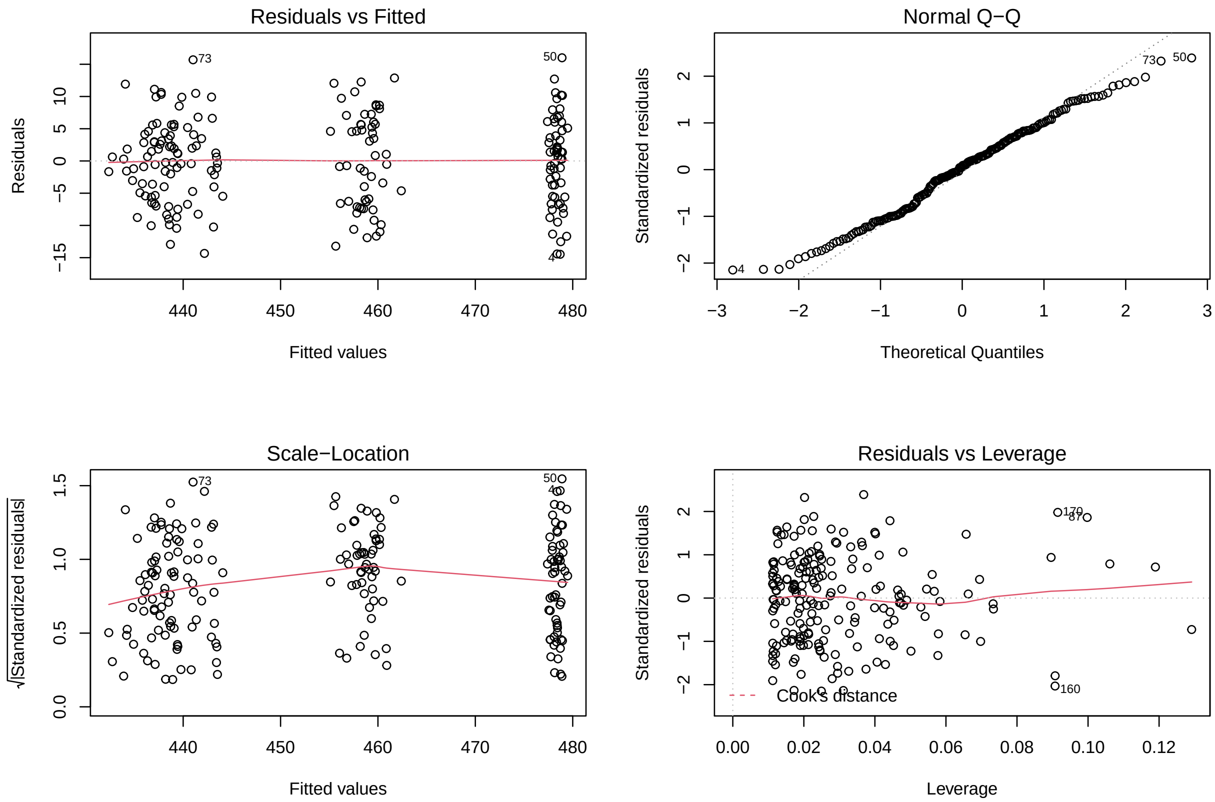
```r
plot(allEffects(lmI_P, residuals=T), grid=T)
```




```r
Anova(lmI_P)
```

```
## Anova Table (Type II tests)
##
## Response: PE
##            Sum Sq  Df  F value    Pr(>F)
## AP            338   1   7.2474  0.007721
## AT_bin      36700   2 394.0239 < 2.2e-16
## AP:AT_bin     435   2   4.6664  0.010487
## Residuals    9035 194
```

```r
par(mfrow=c(2,2))
plot(lmI_P)
```



15) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lmI_P`) for `PE`, defining any indicator variables used. Report the simplified models for any two of the `AT_bin` levels and explain how you did the simplification. Examine the scatterplot for these data. Does it appear that there is an interaction present? How can you tell?

16) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lmI_P`) for `PE`, defining any indicator variables used. Report the simplified models for any two of the `AT_bin` levels and explain how you did the simplification. Provide a conclusion for the test of the term, in context, that explores whether the slopes should differ in the model. What does this test suggest about the inclusion of this term?

17) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lmI_P`) for `PE`, defining any indicator variables used. Report the simplified models for any two of the `AT_bin` levels and explain how you did the simplification. Assess all validity conditions for this model except for independence and multicollinearity. Make sure you discuss which plots/information you are using to assess each assumption.

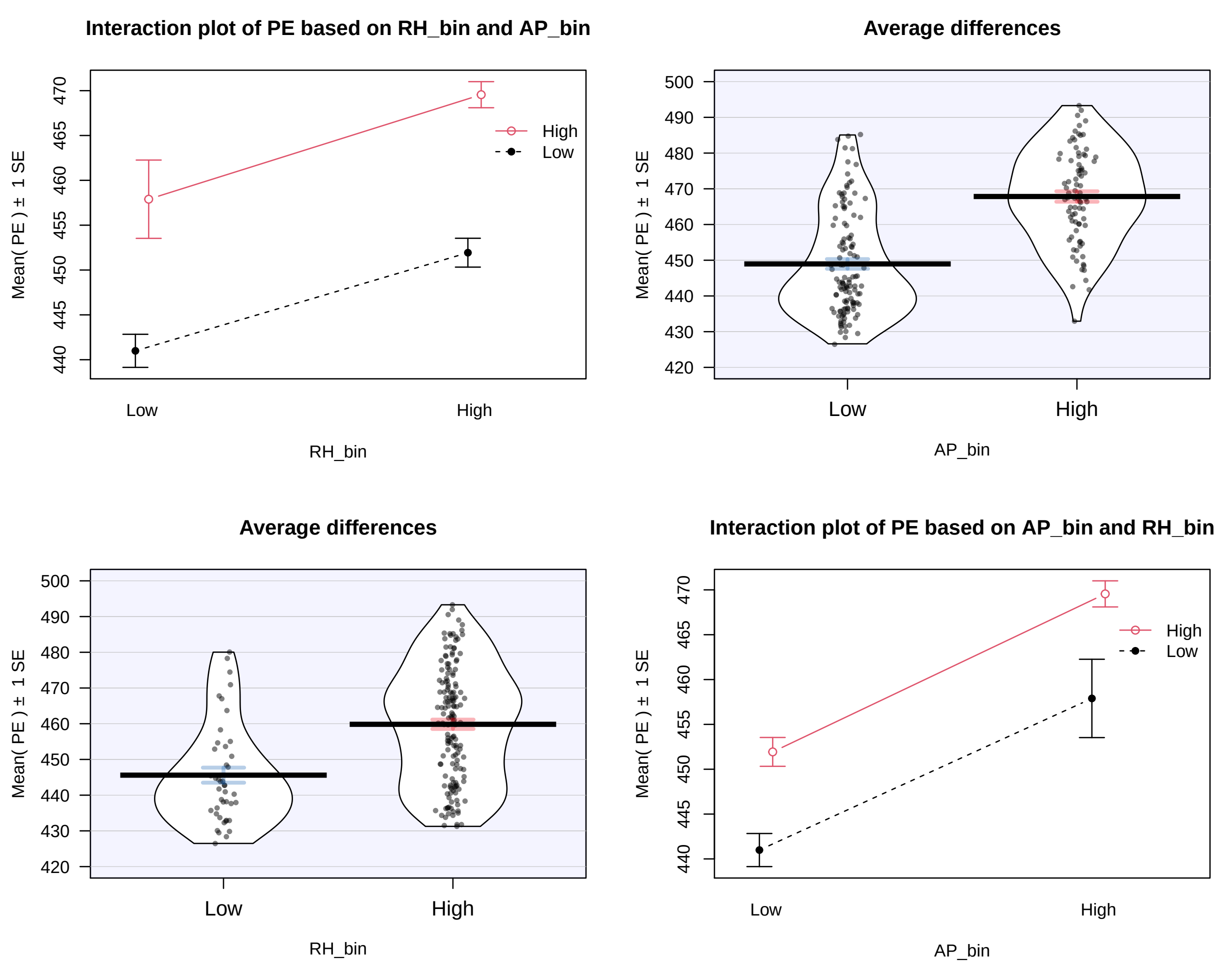
18) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lmI_P`) for `PE`, defining any indicator variables used. Report the simplified models for "Low" and "High" `AT_bin` levels and explain how you did the simplification. Suppose that you observed an AP of 1010 and `AT_bin` of "High." Calculate the predicted `PE` for this observation. What is the residual if `PE` was 430 MW for this observation? Describe all calculations and give values. Find R-squared for the full model `lmI_P` and give an interpretation in context.

19) State the method of analysis needed for these data, whether the fitted model is additive or has an interaction, and then report the estimated model (`lmI_P`) for `PE`, defining any indicator variables used.

Report the simplified models for any two of the `AT_bin` levels and explain how you did the simplification. Find and report R-squared for the full model `lmI_P`. Based on these results, does it seem like a power plant manager could use this model to understand what energy output to expect? Use the term plots to discuss the conditions for `AP` and `AT_bin` where the highest mean output would be expected.

For 20-25, we would like to build a model to explore differences in energy output based on relative humidity and atmospheric pressure, where relative humidity (`RH_bin`) has been categorized as "Low" or "High" and atmospheric pressure (`AP_bin`) is categorized as "Low" or "High". This analysis will use the `Q3` subset of the full data to fit the models stored in `lm2_APbyRH` or `lm2_AP_RH` depending on the question.

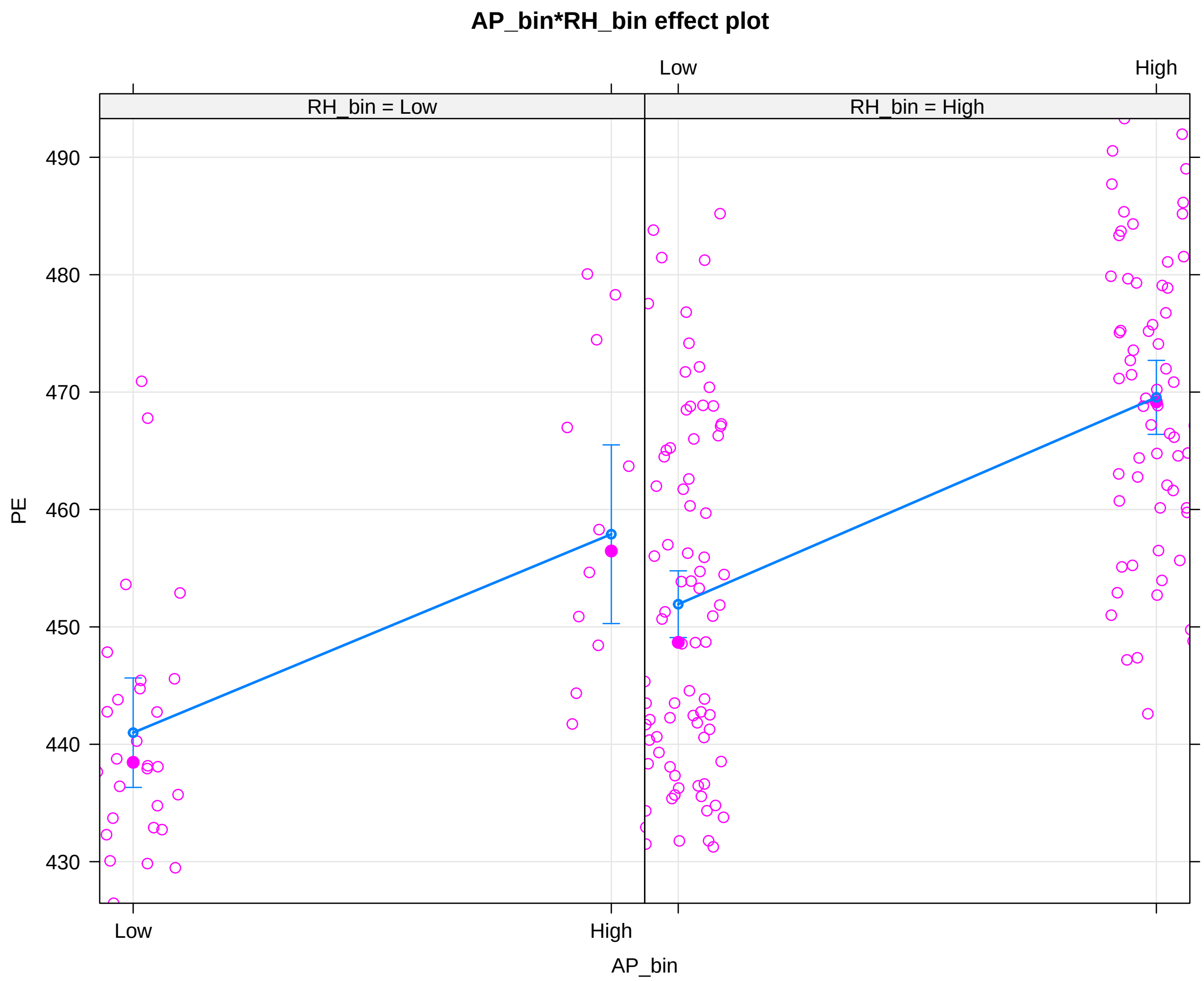
```
Q3$AT_bin <- factor(with(Q3, cut(AT, breaks = c(1.81, 14, 21, 33.41), include.lowest=T)))
levels(Q3$AT_bin) <- c("Low", "Medium", "High")
Q3$RH_bin <- factor(with(Q3, cut(RH, breaks = c(40.47, 65, 100.09), include.lowest=T)))
levels(Q3$RH_bin) <- c("Low", "High")
Q3$AP_bin <- factor(with(Q3, cut(AP, breaks = c(1000, 1015, 1033), include.lowest=T)))
levels(Q3$AP_bin) <- c("Low", "High")

library(catstats)
intplotarray(PE ~ AP_bin*RH_bin, data=Q3)
```




```
lm2_APbyRH <- lm(PE ~ AP_bin*RH_bin, data=Q3)
summary(lm2_APbyRH)
```

```
##
## Call:
## lm(formula = PE ~ AP_bin * RH_bin, data = Q3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.951  -9.809  -1.671   9.761  33.268
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              440.9884     2.3628 186.636  < 2e-16
## AP_binHigh                16.9066     4.5245   3.737 0.000245
## RH_binHigh                10.9438     2.7677   3.954 0.000107
## AP_binHigh:RH_binHigh      0.7119     5.0100   0.142 0.887145
##
## Residual standard error: 13.37 on 196 degrees of freedom
## Multiple R-squared:  0.3797, Adjusted R-squared:  0.3702
## F-statistic: 39.99 on 3 and 196 DF,  p-value: < 2.2e-16
```

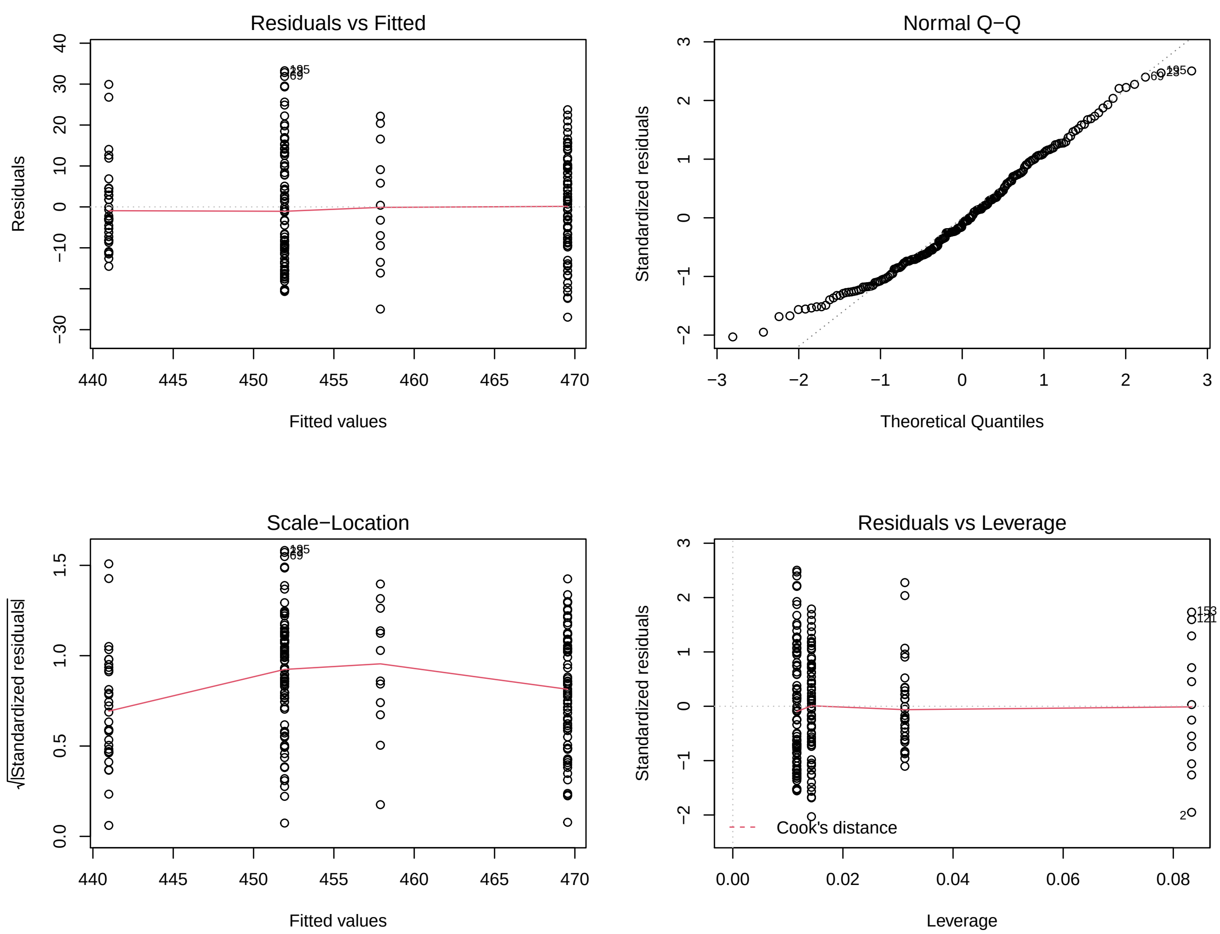
```
plot(allEffects(lm2_APbyRH, residuals=T), grid=T)
```



```r
Anova(lm2_APbyRH)
```

```
## Anova Table (Type II tests)
##
## Response: PE
##                Sum Sq  Df F value    Pr(>F)
## AP_bin          14470   1 80.9920 < 2.2e-16
## RH_bin           4181   1 23.4042 2.652e-06
## AP_bin:RH_bin       4   1  0.0202    0.8871
## Residuals       35016 196
```

```r
par(mfrow=c(2,2))
plot(lm2_APbyRH)
```

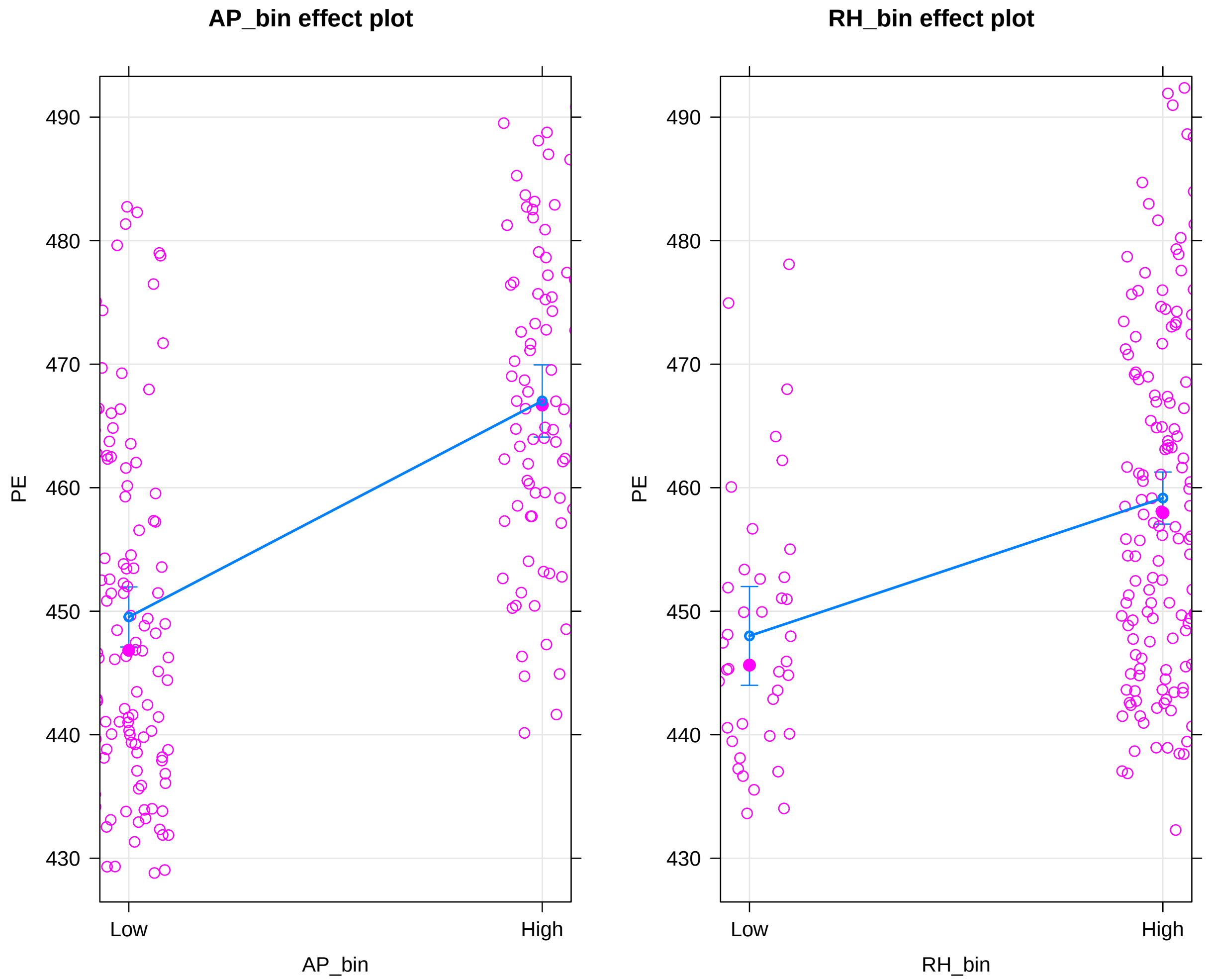



```r
lm2_AP_RH <- lm(PE ~ AP_bin + RH_bin, data=Q3)
summary(lm2_AP_RH)
```

```
##
## Call:
## lm(formula = PE ~ AP_bin + RH_bin, data = Q3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -26.878  -9.888  -1.665   9.760  33.209
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  440.830      2.078 212.104  < 2e-16
## AP_binHigh    17.487      1.938   9.022  < 2e-16
## RH_binHigh    11.161      2.301   4.850  2.5e-06
##
## Residual standard error: 13.33 on 197 degrees of freedom
## Multiple R-squared:  0.3796, Adjusted R-squared:  0.3733
## F-statistic: 60.27 on 2 and 197 DF,  p-value: < 2.2e-16
```

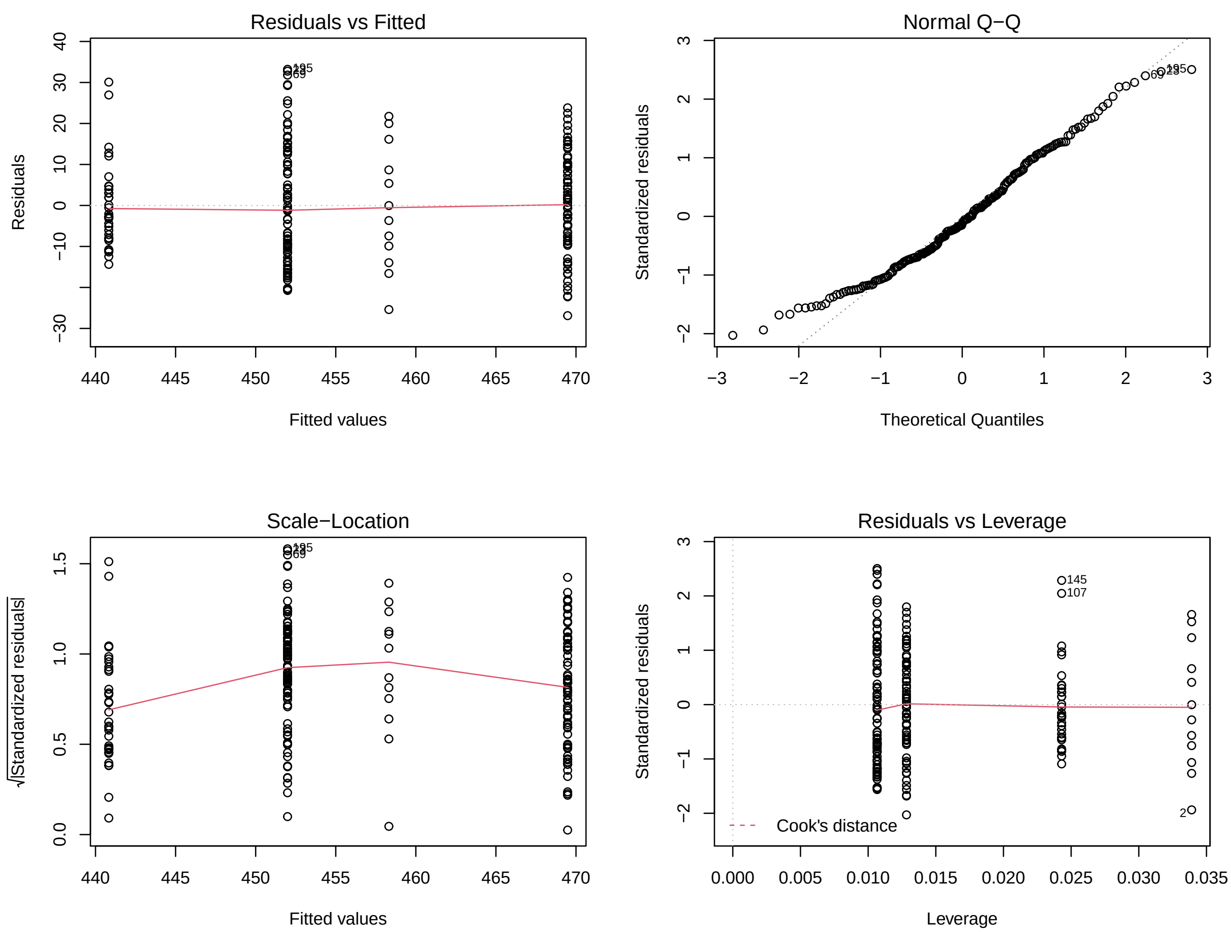
```r
plot(allEffects(lm2_AP_RH, residuals=T), grid=T)
```




```r
Anova(lm2_AP_RH)
```

```
## Anova Table (Type II tests)
##
## Response: PE
##           Sum Sq  Df F value    Pr(>F)
## AP_bin     14470   1  81.397 < 2.2e-16
## RH_bin      4181   1  23.521 2.503e-06
## Residuals  35020 197
```

```r
par(mfrow=c(2,2))
plot(lm2_AP_RH)
```



20) Discuss the interaction plot for `AP_bin` and `RH_bin` from the `Q3` data set. Does it appear that there is an interaction present? How can you tell? State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_APbyRH` for `PE`. Then report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the high levels of both variables and explain how you found those values. Find R-squared for this model and give an interpretation in context.

21) State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_APbyRH` for `PE`. Then report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the high levels of both variables and explain how you found those values. Provide a conclusion for the test of the term, in context, that explores whether the impacts of `AP_bin` on `PE` change by `RH_bin` levels. What does this test suggest about the inclusion of this term?

22) State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_APbyRH` for `PE`. Then report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the high levels of both variables and explain how you found those values. Assess the validity conditions for that model except for independence. Make sure you discuss which plots/information you are using to assess each assumption. Find R-squared for this model and give an interpretation in context.

23) State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_AP_RH` for `PE`. Report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the high level of `AP_bin` and low level of `RH_bin` and explain how you found those values. Interpret the coefficient in the `AP_binHigh` row in the context of the problem. Find R-squared for this model and give an interpretation in context.

24) State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_AP_RH` for `PE`. Report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the low level of `AP_bin` and high level of `RH_bin` and explain how you found those values. Interpret the coefficient in the `RH_binHigh` row in the context of the problem. Find R-squared for this model and give an interpretation in context.

25) State the method of analysis needed for these data and whether the fitted model is additive or has an interaction in the model `lm2_AP_RH` for `PE`. Report the estimated mean for the low levels of `AP_bin` and `RH_bin` and for the low level of `AP_bin` and high level of `RH_bin` and explain how you found those values. Find and report the R-squared for this model. Based on these results, does it seem like a power plant manager could use this model to understand what power output to expect? Based on this model, under what conditions for `AP_bin` and `RH_bin` would the highest mean output be expected?

Part III (5 points): Collaboration

- Discuss any support, discussions, or other feedback you used in preparing your answer. Failure to address this in your recording will result in losing all 5 points. You are allowed to discuss and get feedback on your answers but you must be the speaker in the video and your answers must be your own words.

Part IV (5 points): Time limit

- You must stay within the allotted 6 minutes for your recording. Any exceedance of that time may result in a loss of points.