

STAT021 Take Home Data Analysis Project

US National Park Visits

100/100
Perfect!

Yuchao Wang

December 5, 2018

Contents

1	Introduction	2
1.1	Exploratory Data Analysis	2
1.1.1	Further Investigating Campers and Lodging	3
2	Method	3
2.1	Model 1	3
2.1.1	Building Model	3
2.1.2	Selecting Model	4
2.2	Model 2	4
2.2.1	Building Model	4
2.2.2	Selecting Model	5
3	Results	6
3.1	Model 1	6
3.2	Model 2	7
3.3	Predictions	8
3.3.1	Model 1	8
3.3.2	Model 2	8
4	Conclusion and discussion	8
5	Appendix	9
5.1	Exploratory Data Analysis	9
5.1.1	Further Investigating Campers and Lodging	11
5.2	Model 1	13
5.2.1	CHOOSE	13
5.2.2	FIT	35
5.2.3	ASSESS	39
5.2.4	USE	41
5.3	Model 2	41
5.3.1	CHOOSE	41
5.3.2	FIT	52
5.3.3	ASSESS	53
5.3.4	USE	55

1 Introduction

US National Parks hit a historic high of 300 million visits in total in 2015, and there is a generally increasing trend for number of visits to US National Parks over time since 1904. To understand which factors contribute significantly to the number of visits, and thus see if we are able to build a model that accurately predicts number of visits, we will in this project use data from **National Park Service** website to build two models that best predicts number of visits to 3 national parks: Rocky Mountain, Grand Canyon, and Yellowstone, which have top total recreational visits in 2017. One model has no limitation on the number of variables used, while the other one needs to keep the number of variables under 20. We have data about these 3 national parks on *State*, *Year*, *Season*, *Lodging*, *Campers*, and *TotalVisits* from 1979 to 2017 (since 1979 was the first year NPS website had more detailed data on visits e.g. overnight stays). We will build regression models to predict *TotalVisits* for the 3 national parks in the following sections.

1.1 Exploratory Data Analysis

Summary statistics, visualization, and VIF test of 1.1 can be found in 5.1.

1. We can first disregard one variable, *State*, because each of the three national parks is from a different state and hence this variable does not add any new information to the model when we include *Park* as a predictor.
2. In barplot of *Lodging* according to *Park* we notice that there is no data (all 0) for *Lodging* for 1 park, which is Rocky Mountain. This is because Rocky Mountain offers no lodging and has only camping space in the park. We may therefore consider categorizing *Lodging* to separate out this group effect.
3. We see in exploratory data analysis matrix that distribution of *Lodging*, *Campers*, and *TotalVisits* are all left-skewed (the left skew for *Lodging* may be due to the 0 entries for Rocky Mountain). We can try log-transforming (*Lodging*+1) (to avoid $\log(0)$) and *Campers* to see if it gives a higher R^2 value, and we need to check scatterplot and Q-Q plot of residuals to determine if there is any violation of model assumptions.
4. From scatterplot of *TotalVisits* against *Year*, we can see different “strata”, which possibly indicate strong group effects caused by *Season* (indicated by the histogram of *TotalVisits* against *Season*). And since the strata seem almost equally spaced from each other, we may consider converting *Season* to a quantitative variable in building a simpler Model 2.
5. There is a general positive relationship between *Lodging* and *TotalVisits* as well as between *Campers* and *TotalVisits*, which makes sense because increase in number of campers and users of concessional lodging means more overall visitors. However, there are obvious clusters of data points which we can investigate further in 1.1.1.
6. From exploratory data analysis we see that there seem to be relatively strong correlations between *Lodging* and *TotalVisits* and between *Campers* and *TotalVisits*. However, the VIF values are all below $\sqrt{5}$ and therefore there is no multicollinearity problem.

1.1.1 Further Investigating Campers and Lodging

Plots can be found in 5.1.1.

Analyzing Campers variable: The scatterplot of *TotalVisits* against *Campers* indicates two apparent clusters, which do not differ by *Season* or *Park* (both belong to summer visits to Yellowstone). But a further scatterplot on the right mapping *Year* instead of *Season* onto the points shows two strong linear trends that differ by *Year*. Since we observe in scatterplot of *TotalVisits* against *Year* that the trend is non-linear, we therefore consider changing the quantitative variable *Year* arbitrarily to a categorical variable of 5 levels: before 1980 (1979), 1980-1989, 1990-1999, 2000-2009, and 2010-2017, which will be stored as *Year5*, and try if this categorical treatment captures the overall non-linear trend between *Year* and *TotalVisits* better. We may also consider to categorize *Campers* into two categories with a cutoff of 300 (thousand). This categorical variable is stored as *Campers2*.

Analyzing Lodging variable: The scatterplot of *TotalVisits* against *Lodging* highlights that the number of lodging for Rocky Mountain is always 0. The cluster of data points farther away from the origin all belong to summer but differ in years. We may consider categorizing *Lodging* into three categories: 0 (Rocky Mountain), 0.1-499.9, and over 500, labeled as *Lodging3*.

2 Method

2.1 Model 1

2.1.1 Building Model

From the above analysis, we have *Park*, (*Year* or *Year5*), *Season*, (*Lodging* or *logLodging* or *Lodging3*), and (*Campers* or *logCampers* or *Campers2*) as potential predictors. We can then build full models and carry out forward selection, backward elimination, and stepwise procedure for each full model, compare the three results from these procedures, and select the best model for a particular full model (with the highest R^2 value in this case, since we aim at best predicting the *TotalVisits*). The best models from all full models built can then be compared using the same criterion (highest R^2 value) and checking model assumptions to ultimately give Model 1.

In building the full models, up to three-way interaction terms were considered, as well as up to cubic terms for quantitative predictors. For example:

FullModel1: $TotalVisits \sim (Park + Year + Season + Campers + Lodging)^3 + I(Campers^2) + I(Lodging^2) + I(Year^2) + I(Campers^3) + I(Lodging^3) + I(Year^3)$

FullModel2: $TotalVisits \sim (Park + Year5 + Season + Campers + Lodging)^3 + I(Campers^2) + I(Lodging^2) + I(Campers^3) + I(Lodging^3)$

2.1.2 Selecting Model

We also tried other full models that differ from **FullModel1** by substituting one original variable with a transformed variable or categorized variable, e.g. *logLodging* instead of *Lodging*, as well as with fewer (up to two-way) interaction terms or only quadratic terms of quantitative predictors. But none of them gave a higher R^2 value than **bw2** produced by backward elimination of **FullModel2**, which has R^2 of 0.9881. The second highest R^2 is 0.9861, produced by **bw1** from **FullModel1**. **bw1** and **bw2** also have the highest adjusted R^2 values and lowest Mellow's C_p and AIC values among the three models generated by their respective full model. And we used nested F test to ensure that **bw1** and **bw2** are significantly different.

We then checked model assumptions in 5.2.3. The BP test shows that both **bw1** and **bw2** do not fulfill equal variance assumption. However, there is no significant pattern left in the Residual vs. Fitted Value scatterplot for both models (hence linearity fulfilled) and variance seems roughly constant (though there is a difference between when fitted values < 1000 and when fitted values > 1000), so we can ignore the BP test, especially given the extremely high R^2 values which are desirable for our objective of prediction. The Q-Q plot for **bw2** shows apparent deviation from the $y = x$ line, which indicates that the normal distribution assumption is violated, while the Q-Q plot for **bw1** does not show this deviation and fulfills normality assumption. 0 mean is automatically fulfilled. Independence of data is not fulfilled (since it is an observational data collection process rather than a controlled experiment), but we can ignore it for the purpose of generating a model for prediction. For unusual points, there are exceptionally unusual points judged by Cook's distance for **bw2**, but no unusual points in terms of Cook's distance for **bw1** (though there are 12 exceptionally usual points in terms of leverage and a few in terms of studentized residual). So **bw1** account for unusual points better and may give better predictions. We therefore chose **bw1** as our final Model 1 since it has the second highest R^2 value and fulfills model assumptions better than **bw2** (and it is also simpler than **bw2**). For details of **bw1** as Model 1, see 3.1.

2.2 Model 2

2.2.1 Building Model

We aim to have a much simpler Model 2 (with no more than 20 variables) which also predicts *TotalVisits* well (having a high R^2 value). From previous analyses in 1.1 and 2.1, we have a list of potential variables to choose from. And observing in **bw2** that there seems to be a linear increasing trend for the coefficient of main effects of *Year5* (see **bw2** in 5.2.1), we turned it into a quantitative variable *Year5num* by assigned each group in chronological order a value of 0, 1, 2, 3, and 4 respectively. There also seems to be a linear increasing trend for the coefficient of main effects of *Season* (as expected in pt. 4 of 1.1), so we converted *Season* into *Seasonnum* by assigning 0, 1, 2, and 3 to Winter, Spring, Fall, Summer respectively. The same aforementioned process was used: to build full models using 3 different selection procedures to obtain potential models and compare them.

In building the full models, up to two-way interaction terms were considered, as well as up to quadratic terms for quantitative predictors. For example:

FullModel3: $\text{TotalVisits} \sim (\text{Park} + \text{Year5num} + \text{Seasonnum} + \text{Campers} + \text{Lodging3})^2 + \text{I}(\text{Campers}^2)$

FullModel4: $\text{TotalVisits} \sim (\text{Park} + \text{Year} + \text{Seasonnum} + \text{Campers} + \text{Lodging3})^2 + \text{I}(\text{Year}^2) + \text{I}(\text{Campers}^2)$

2.2.2 Selecting Model

FullModel3 generates **st3** (by stepwise procedure), which is the same as **bw3** (by backward elimination). Though it also generates **fw3** by forward selection, which has a slightly higher R^2 value, nested F test shows that **fw3** is not significantly different from **st3** while using more variables. **st3** also has lower Mellow's C_p and AIC value than **fw3**. Therefore **st3**, which is the same as **bw3**, is the best model produced by FullModel3.

FullModel4 generates **st4** by stepwise procedure, which is the same as **fw4** by forward selection. Nested F test shows that **bw4**, though having a slightly higher R^2 value than **st4**, is not significantly different from **st4**. Since **st4** has less variables, it is the best model generated by FullModel4. By nested F test, we see that **st3** and **st4** are significantly different, suggesting that **st4** is better (since it has more variables). **st3** has R^2 of 0.9522 and **st4** has R^2 of 0.9569. **st4** also has higher adjusted R^2 values and lower Mellow's C_p and AIC values than **st3**. We also tried other full models that differ from these two by exchanging one original variable with a transformed variable or categorized variable, e.g. *Season* instead of *Seasonnum*. But none of them generated models with a higher R^2 value than those of **st3** and **st4** by using no more than 20 variables.

We then checked model assumptions in 5.3.3. The BP test shows that both **st3** and **st4** do not fulfill equal variance assumption. However, there is no significant pattern left in the Residual vs. Fitted Value scatterplot for both models (hence linearity fulfilled) and variance seems roughly constant (though there is a difference between when fitted values < 1000 and when fitted values > 1000), so we can ignore the BP test, especially given the high R^2 values which are desirable for our objective of prediction. The Q-Q plots for **st3** and **st4** fit the $y=x$ line well, which indicates that the normal distribution assumption is fulfilled. 0 mean is automatically fulfilled. Independence of data is not fulfilled (since it is an observational data collection process rather than a controlled experiment), but we can ignore it for the purpose of generating a model for prediction. For unusual points, there is no unusual point in terms of Cook's distance for both **st3** and **st4**. So overall **st3** and **st4** do equally well in terms of fulfilling model assumptions. We therefore chose **st4** as our final Model 2 since it has a higher R^2 value. For details of **st4** as Model 2, see 3.2.

3 Results

3.1 Model 1

Choose

For exploratory data analysis of variables used in Model 1, see 5.1, since we only used original variables. As previously mentioned, there is no multicollinearity problem.

Y indicates *TotalVisit* X_P indicates *Park*, where X_{Pr} : Rocky Mountain, X_{Py} : Yellowstone, X_{Pg} : Grand Canyon.

X_S indicates *Season*, where X_{Sp} : Spring, X_{Ss} : Summer, X_{Sf} : Fall.

X_Y indicates *Year*, X_C indicates *Campers* and X_L indicates *Lodging*.

Model 1: $Y = \beta_0 + \beta_1 X_{Py} + \beta_2 X_{Pg} + \beta_3 X_Y + \beta_4 X_{Sp} + \beta_5 X_{Sf} + \beta_6 X_{Ss} + \beta_7 X_C + \beta_8 X_L + \beta_9 X_C^2 + \beta_{10} X_Y^2 + \beta_{11} X_C^3 + \beta_{12} X_L^3 + \beta_{13} X_Y^3 + \beta_{14} X_{Py} X_Y + \beta_{15} X_{Pg} X_Y + \beta_{16} X_{Py} X_{Sp} + \beta_{17} X_{Pg} X_{Sp} + \beta_{17} X_{Py} X_{Sf} + \beta_{18} X_{Pg} X_{Sf} + \beta_{19} X_{Py} X_{Ss} + \beta_{20} X_{Pg} X_{Ss} + \beta_{21} X_{Py} X_C + \beta_{22} X_{Pg} X_C + \beta_{23} X_{Py} X_L + \beta_{24} X_Y X_{Sp} + \beta_{25} X_Y X_{Sf} + \beta_{26} X_Y X_{Ss} + \beta_{27} X_Y X_C + \beta_{28} X_Y X_L + \beta_{29} X_{Sp} X_C + \beta_{30} X_{Sf} X_C + \beta_{31} X_{Ss} X_C + \beta_{32} X_{Sp} X_L + \beta_{33} X_{Sf} X_L + \beta_{34} X_{Ss} X_L + \beta_{35} X_C X_L + \beta_{36} X_{Py} X_Y X_{Sp} + \beta_{37} X_{Pg} X_Y X_{Sp} + \beta_{38} X_{Py} X_Y X_{Sf} + \beta_{39} X_{Pg} X_Y X_{Sf} + \beta_{40} X_{Py} X_Y X_{Ss} + \beta_{41} X_{Pg} X_Y X_{Ss} + \beta_{42} X_{Py} X_Y X_C + \beta_{43} X_{Pg} X_Y X_C + \beta_{44} X_{Py} X_Y X_L + \beta_{45} X_{Py} X_{Sp} X_L + \beta_{46} X_{Py} X_{Sf} X_L + \beta_{47} X_{Py} X_{Ss} X_L + \beta_{48} X_Y X_{Sp} X_C + \beta_{49} X_Y X_{Sf} X_C + \beta_{50} X_Y X_{Ss} X_C + \beta_{51} X_Y X_{Sp} X_L + \beta_{52} X_Y X_{Sf} X_L + \beta_{53} X_Y X_{Ss} X_L + \beta_{54} * X_Y X_C X_L + \epsilon$, where $\epsilon \stackrel{iid}{\sim} N(0, \sigma)$

Fit

Estimated regression line:

$$\begin{aligned} \bar{y} = & -3.1*10^8 + 5.9*10^3 X_{Py} - 1.3*10^4 X_{Pg} + 4.6*10^5 X_Y - 830 X_{Sp} + 1.3*10^4 X_{Sf} - 1.8*10^3 X_{Ss} \\ & - 300 X_C - 150 X_L - 0.027 X_C^2 - 230 X_Y^2 + 3.7 * 10^{-5} X_C^3 + 4.5 * 10^{-6} X_L^3 + 0.038 X_Y^3 - 2.9 X_{Py} X_Y \\ & + 6.3 X_{Pg} X_Y - 3.2 * 10^4 X_{Py} X_{Sp} - 1.6 * 10^5 X_{Pg} X_{Sp} - 4.7 * 10^4 X_{Py} X_{Sf} - 8.4 * 10^4 X_{Pg} X_{Sf} - \\ & 1.7 * 10^5 X_{Py} X_{Ss} - 2.4 * 10^5 X_{Pg} X_{Ss} + 790 X_{Py} X_C + 920 X_{Pg} X_C - 460 X_{Py} X_L + 0.43 X_Y X_{Sp} - \\ & 6.4 X_Y X_{Sf} + 1.1 X_Y X_{Ss} + 0.15 X_Y X_C + 0.076 X_Y X_L - 310 X_{Sp} X_C - 480 X_{Sf} X_C + 97 X_{Ss} X_C \\ & + 900 X_{Sp} X_L + 740 X_{Sf} X_L + 860 X_{Ss} X_L - 1.4 X_C X_L + 16 X_{Py} X_Y X_{Sp} + 80 X_{Pg} X_Y X_{Sp} \\ & + 23 X_{Py} X_Y X_{Sf} + 42 X_{Pg} X_Y X_{Sf} + 88 X_{Py} X_Y X_{Ss} + 120 X_{Pg} X_Y X_{Ss} - 0.40 X_{Py} X_Y X_C - \\ & 0.46 X_{Pg} X_Y X_C + 0.23 X_{Py} X_Y X_L + 11 X_{Py} X_{Sp} X_L + 5.8 X_{Py} X_{Sf} X_L - 0.67 X_{Py} X_{Ss} X_L + \\ & 0.16 X_Y X_{Sp} X_C + 0.25 X_Y X_{Sf} X_C - 0.044 X_Y X_{Ss} X_C - 0.46 X_Y X_{Sp} X_L - 0.37 X_Y X_{Sf} X_L - \\ & 0.44 X_Y X_{Ss} X_L + 7.0 * 10^{-4} * X_Y X_C X_L \end{aligned}$$

Assess

Model 1 has R^2 of 0.9861, meaning that 98.61% of the variability in *TotalVisits* can be explained by this model. F test for the entire model has p value $< 2.2 * 10^{-16}$, hence Model 1 is significant and strong in explanatory power. For t tests for individual predictors, see 5.2.2.

For checking assumptions and searching for unusual points, see 2.1.2.

Use

Understanding a polynomial term with significant p value of t -test of slope:

Predictor $Year^3$ has a coefficient of 0.038 in Model 1 and the slope is significant. This means that, holding other variables constant, when $Year$ increases by 1 unit, there is a cubic increase in $TotalVisits$ with a coefficient of 0.038.

Understanding a three-way interaction term with significant p value of t -test of slope:

Predictor $Year:SeasonSpring:Lodging$ has a coefficient of -0.46 and the slope is significant. This means that, holding other variables constant, when $Year$ increases by 1 unit, there is a decrease by a coefficient of 0.46 of how the slope of $TotalVisits$ against $Lodging$ changes in response to $SeasonSpring$ changing from 0 to 1 (Season is now spring) (difference of differences of differences).

3.2 Model 2

Choose

For exploratory data analysis of variables used in Model 1, see 5.3.1. Here, the “0” level of $Lodging3$ gives the same categorization as Rocky Mountain in $Park$, as discussed in introduction. The VIF test therefore cannot be applied to all five variables (there is multicollinearity between $Lodging3$ and $Park$). However, we tested that using $Lodging3$ gives a better model fit than using $Lodging$. And we know that $Lodging3$ gives more information than what $Park$ supplies (they are not the same predictor). In view of our objective, we therefore kept $Lodging3$ in the model to obtain better predictions.

Y indicates $TotalVisit$ X_P indicates $Park$, where X_{Pr} : Rocky Mountain, X_{Py} : Yellowstone, X_{Pg} : Grand Canyon.

X_{Sn} indicates $Seasonnum$, where Winter = 0, Spring = 1, Fall = 2 and Summer = 3.

X_Y indicates $Year$, X_C indicates $Campers$ and X_{L3} indicates $Lodging3$ with default value 0, where $X_{L30} = 1$ for group with $Lodging = 0$, $X_{L3u} = 1$ for group with $0 < Lodging < 500$, $X_{L3o} = 1$ for group with $500 \leq Lodging < 1000$.

$$Y \sim \beta_0 + \beta_1 X_{Sn} + \beta_2 X_{L3u} + \beta_3 X_{L3o} + \beta_4 X_Y + \beta_5 X_{Py} + \beta_6 X_{Pg} + \beta_7 X_C + \beta_8 X_Y^2 + \beta_9 X_{L3u} X_Y + \beta_{10} X_{L3o} X_Y + \beta_{11} X_{L3u} X_C + \beta_{12} X_{L3o} X_C + \beta_{13} X_{Sn} X_Y + \beta_{14} X_Y X_{Py} + \beta_{15} X_Y X_{Pg} + \beta_{16} X_{Sn} X_C + \beta_{17} X_Y X_C + \epsilon, \text{ where } \epsilon \stackrel{iid}{\sim} N(0, \sigma)$$

Fit

Estimated regression line:

$$Y \sim -6.4 * 10^5 - 4.8 * 10^3 X_{Sn} + 1.0 * 10^4 X_{L3u} - 1.1 * 10^4 X_{L3o} + 630 X_Y - 2.2 * 10^4 X_{Py} - 57 X_C - 0.16 X_Y^2 - 4.9 X_{L3u} X_Y + 6.6 X_{L3o} X_Y - 3.0 X_{L3u} X_C - 6.9 X_{L3o} X_C + 2.5 X_{Sn} X_Y + 1.1 X_Y X_{Py} + 0.30 X_{Sn} X_C + 0.032 X_Y X_C$$

Assess

Model 2 has R^2 of 0.9569, meaning that 95.69% of the variability in $TotalVisits$ can be explained by this model. F test for the entire model has p value $< 2.2 * 10^{-16}$, hence Model 2 is significant and strong in explanatory power, and fulfills our requirement of having no

more than 20 predictors (16 in total as two predictors were discarded due to singularities). For t tests for individual predictors, see 5.3.2.

For checking assumptions and searching for unusual points, see 2.2.2.

Use

Understanding a polynomial term with significant p value of t -test of slope:

Predictor $Year^2$ has a coefficient of -1.6 in Model 2 and the slope is significant. This means that, holding other variables constant, when $Year$ increases by 1 unit, there is a quadratic decrease in $TotalVisits$ with a coefficient of 1.6.

Understanding a two-way interaction term with significant p value of t -test of slope:

Predictor $Year:Campers$ has a coefficient of 3.2 and the slope is significant. This means that, holding other variables constant, when $Year$ increases by 1 unit, the slope of $TotalVisits$ against $Campers$ increases by 3.2.

3.3 Predictions

3.3.1 Model 1

In 2017 Summer for the Yellowstone with $Lodging = 858.573$ (in thousand) and $Campers = 146.306$ (in thousand), $TotalVisits$ (in thousand) (individual response) is predicted by Model 1 as 3169.767, with lower bound of 2963.91 and upper bound of 3375.624 prediction interval. The actual $TotalVisits$ was 3233.820, which is relatively close to the predicted value and within the prediction interval. Therefore Model 1 seems to predict accurately and relatively precisely.

3.3.2 Model 2

In 2017 Summer for the Yellowstone with $Lodging = 858.573$ (in thousand) and $Campers = 146.306$ (in thousand), $TotalVisits$ (in thousand) (individual response) is predicted by Model 1 as 2969.06, with lower bound of 2652.899 and upper bound of 3285.221 prediction interval. The actual $TotalVisits$ was 3233.820, which is within the prediction interval, but not as close to the actual value as Model 1 predicts. Both Model 1 and Model 2 therefore seem to predict accurately, but Model 1 gives relatively a smaller prediction interval and thus more precise predictions than Model 2.

4 Conclusion and discussion

Both Model 1 and Model 2 give very high R^2 values, indicating that they have high explanatory power and thus predictive power. However, increasing the R^2 value from 0.9569 (Model 2)

to 0.9861 (Model 1) tremendously complicates the model and makes both understanding the model and applying the model by hand difficult (Model 1 has 49 predictors while Model 2 only has 16). Therefore it is important for us to rationally weigh the benefit and cost of adding predictors and adjust our strategy depending on the task at hand. In this project, our main goal was to increase predictive power and model application was done by R. Therefore it was worth it to try a large, complicated model like Model 1. However, Model 2 significantly reduced the difficulty of understanding the model and gave insights into the relationship between variables better than Model 1, while not compromising much on its predictive power. In real life, I would prefer using Model 2.

There are other reasons, based on the article *The National Parks Have Never Been More Popular* by Andrew Flowers, that could account for variability in *TotalVisits* but were not considered in this project. For example, price of gas and people's mentality (getting used to high price after a price surge) can affect people's willingness to go to the parks. Social movements like campaigns targeting millennials before rare events e.g. the centennial of national parks, location of the park and admission fee, as well as global financial situation could also affect *TotalVisits*. The number of overnight stays plummeted in 2008, which was possibly related to the financial crisis. If we include more parks into the model, it may not be the best practice to simply code each as a categorical variable, since it can significantly increase the number of variables. Accounting for the aforementioned factors, on the other hand, may elicit more general trends and give more insights into the reason in variability of *TotalVisits*.

5 Appendix

5.1 Exploratory Data Analysis

To go back to main text: 1.1.

```
head(NationalParkVisits)
```

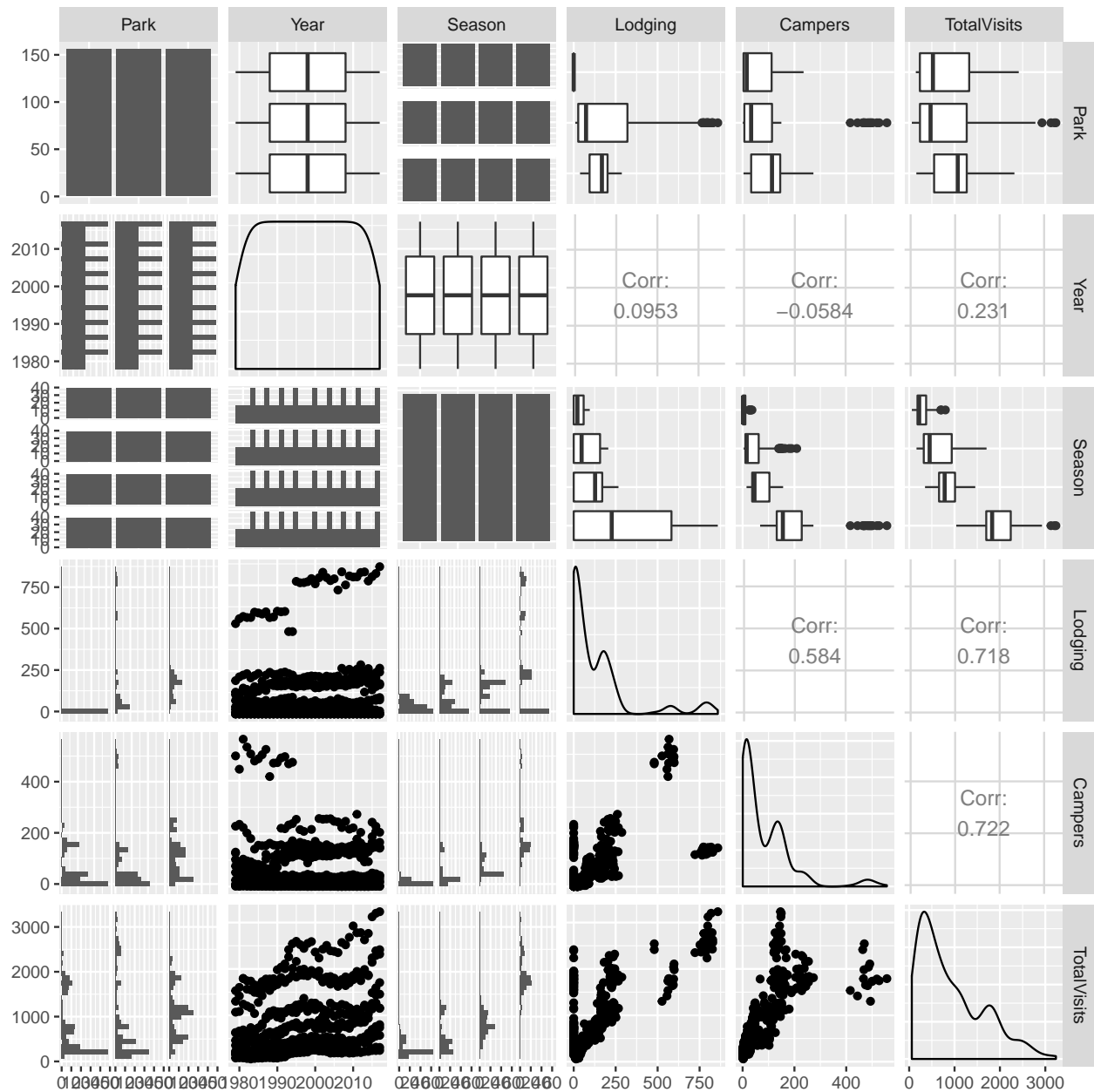
```
##           Park State Year Season Lodging Campers TotalVisits
## 1 Grand Canyon   AZ 1979   Fall 112.351  53.313    624.161
## 2 Grand Canyon   AZ 1979  Spring 126.996  77.839    460.096
## 3 Grand Canyon   AZ 1979  Summer 194.132 130.148   1067.756
## 4 Grand Canyon   AZ 1979  Winter  41.337   5.946    158.321
## 5 Grand Canyon   AZ 1980   Fall 120.403  47.863    625.547
## 6 Grand Canyon   AZ 1980  Spring 119.356  53.311    611.532
```

```
library(psych)
#Summary stats
describe(NationalParkVisits)[,2:4]
```

```
##           n      mean      sd
## Park*    468      2.00    0.82
```

```
## State*      468      2.00   0.82
## Year        468 1998.00  11.27
## Season*     468      2.50   1.12
## Lodging     468  128.16 192.88
## Campers     468   81.75 102.14
## TotalVisits 468  909.66 710.26
```

```
library(GGally)
#Summary matrix with Year
ggpairs(NationalParkVisits[,c("Park", "Year", "Season", "Lodging",
                              "Campers", "TotalVisits")])
```



```
library(car)
#VIF with Year
vif(lm(TotalVisits ~ Park + Year + Season + Lodging + Campers,
      data=NationalParkVisits))
```

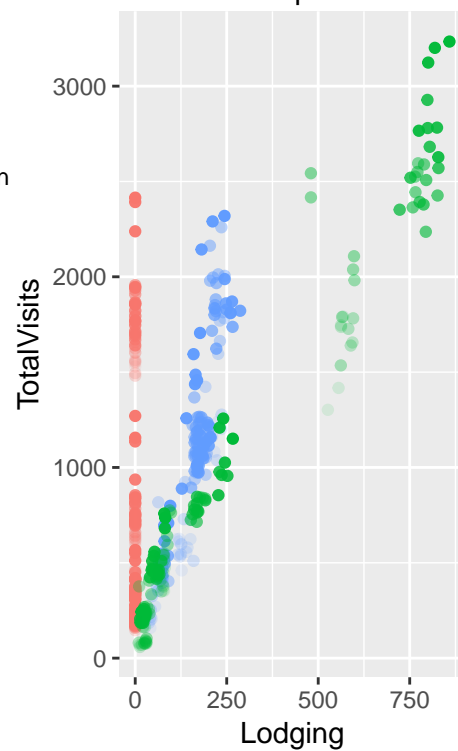
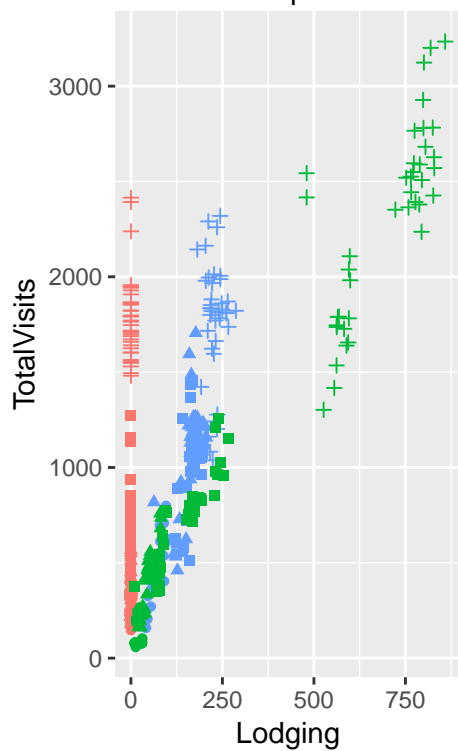
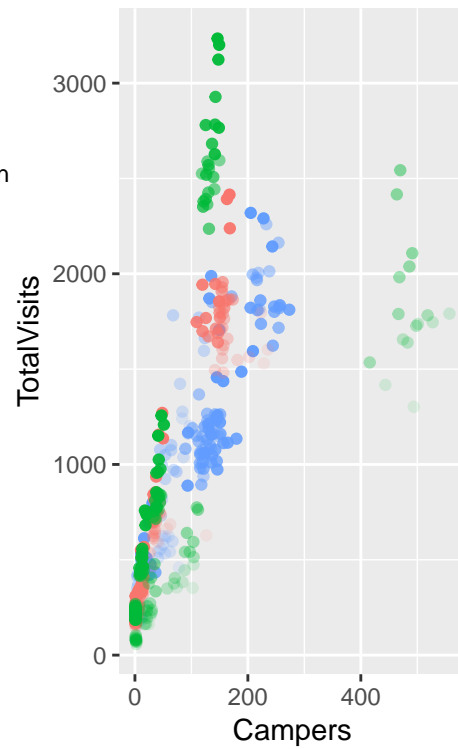
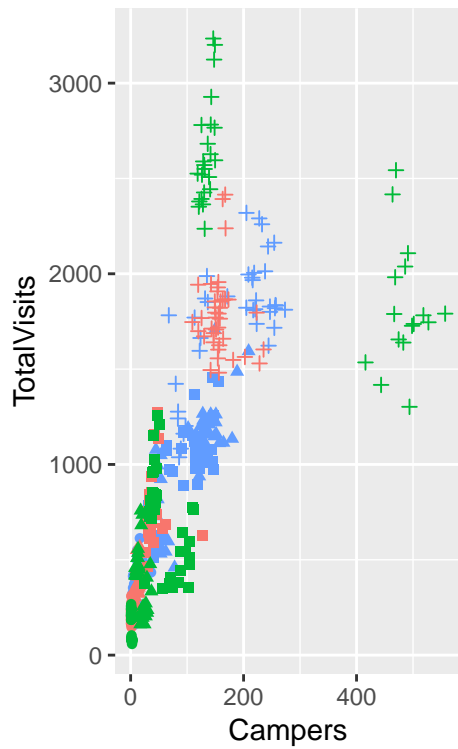
```
##           GVIF Df GVIF^(1/(2*Df))
## Park      1.634893 2      1.130765
## Year      1.036748 1      1.018208
## Season    2.722259 3      1.181648
## Lodging   2.415081 1      1.554053
## Campers   2.582164 1      1.606911
```

To go back to main text: 1.1.

5.1.1 Further Investigating Campers and Lodging

To go back to main text: 1.1.1.

```
#TotalVisits ~ Campers, by Park and Season
ggplot(data=NationalParkVisits, aes(x=Campers, y=TotalVisits, col=Park)) +
  geom_point(aes(shape=Season), position = "jitter")
#TotalVisits ~ Campers, by Park and Year
ggplot(data=NationalParkVisits, aes(x=Campers, y=TotalVisits, col=Park)) +
  geom_point(aes(alpha=Year), position = "jitter")
#TotalVisits ~ Lodging, by Park and Season
ggplot(data=NationalParkVisits, aes(x=Lodging, y=TotalVisits, col=Park)) +
  geom_point(aes(shape=Season), position = "jitter")
#TotalVisits ~ Lodging, by Park and Year
ggplot(data=NationalParkVisits, aes(x=Lodging, y=TotalVisits, col=Park)) +
  geom_point(aes(alpha=Year), position = "jitter")
```



To go back to main text: 1.1.1.

Beautiful!

5.2 Model 1

5.2.1 CHOOSE

To go back to main text: 2.1.1.

To see exploratory data analysis of Model 1, see ??exploratory).

```
#Year as categorical variable
Year5 <- cut(NationalParkVisits$Year, breaks=c(1978, 1979.5, 1989.5,
                                                1999.5, 2009.5, 2018),
            labels=c("79", "80s", "90s", "00s", "10s"))
#Lodging as categorical variable
Lodging3 <- cut(NationalParkVisits$Lodging, breaks=c(-0.5, 0.5, 500, 1000),
              labels=c("0", "under500", "over500"))
#Campers as categorical variable
Campers2 <- cut(NationalParkVisits$Campers, breaks=c(0, 299.5, 600),
              labels = c("under300", "over300"))
#logtransform Campers
logCampers <- log(NationalParkVisits$Campers)
#logtransform Lodging
logLodging <- log(NationalParkVisits$Lodging+1) # +1 to avoid error of log(0)
NationalParkVisits <- data.frame(NationalParkVisits, Year5, logCampers,
                                logLodging, Lodging3, Campers2)

# Zero model and full models
zeroModel <- lm(TotalVisits ~ 1, data=NationalParkVisits)
fullModel1 <- lm(TotalVisits ~ (Park + Year + Season + Campers + Lodging)^3
               + I(Campers^2) + I(Lodging^2) + I(Year^2) + I(Campers^3)
               + I(Lodging^3) + I(Year^3), data=NationalParkVisits)

fullModel2 <- lm(TotalVisits ~ (Park + Year5 + Season + Campers + Lodging)^3
               + I(Campers^2) + I(Lodging^2) + I(Campers^3) +
               I(Lodging^3), data=NationalParkVisits)

#With FullModel1
fw1 <- step(zeroModel, scope=list(upper=fullModel1), direction="forward", trace=0)
bw1 <- step(fullModel1, scope=list(lower=zeroModel), direction="backward", trace=0)
st1 <- step(zeroModel, scope=list(upper=fullModel1), direction="both", trace=0)
summary(fw1)

##
## Call:
## lm(formula = TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) +
##     I(Year^3) + Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) +
##     Season:Lodging + Lodging:Year + Season:Park + Season:Year +
```

```

##      Lodging:Campers + Year:Park + Season:Campers + Park:Campers +
##      Season:Lodging:Year + Season:Year:Park + Season:Park:Campers +
##      Season:Lodging:Campers, data = NationalParkVisits)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -313.44  -53.08   -1.95   53.76  406.77
##
## Coefficients:
##
##              Estimate Std. Error t value
## (Intercept)      -3.835e+08  3.496e+07 -10.970
## SeasonSpring      -3.941e+03  4.518e+03  -0.872
## SeasonFall       -1.726e+04  4.271e+03  -4.041
## SeasonSummer     -3.360e+04  4.751e+03  -7.073
## Lodging          -8.063e+01  2.530e+02  -0.319
## Year             5.757e+05  5.248e+04  10.970
## ParkYellowstone   1.007e+03  7.150e+03   0.141
## ParkGrand Canyon -1.286e+04  1.728e+04  -0.744
## I(Year^2)        -2.881e+02  2.626e+01 -10.970
## I(Year^3)         4.806e-02  4.381e-03  10.970
## Campers          4.147e+01  4.921e+01   0.843
## I(Campers^3)      4.434e-05  7.132e-06   6.217
## I(Campers^2)     -4.465e-02  7.026e-03  -6.356
## I(Lodging^2)      1.716e-03  9.683e-04   1.772
## SeasonSpring:Lodging 2.808e+02  2.701e+02   1.040
## SeasonFall:Lodging 1.769e+02  2.543e+02   0.695
## SeasonSummer:Lodging 3.379e+02  2.542e+02   1.329
## Lodging:Year      3.962e-02  1.275e-01   0.311
## SeasonSpring:ParkYellowstone -1.940e+04  1.090e+04  -1.780
## SeasonFall:ParkYellowstone -1.791e+04  1.306e+04  -1.372
## SeasonSummer:ParkYellowstone -2.120e+05  3.830e+04  -5.535
## SeasonSpring:ParkGrand Canyon -5.458e+04  2.567e+04  -2.126
## SeasonFall:ParkGrand Canyon 1.089e+04  1.957e+04   0.556
## SeasonSummer:ParkGrand Canyon -2.789e+04  2.080e+04  -1.341
## SeasonSpring:Year 1.996e+00  2.264e+00   0.881
## SeasonFall:Year   8.761e+00  2.123e+00   4.128
## SeasonSummer:Year 1.683e+01  2.339e+00   7.195
## Lodging:Campers   1.562e-01  1.731e-01   0.903
## Year:ParkYellowstone -4.529e-01  3.575e+00  -0.127
## Year:ParkGrand Canyon 6.651e+00  8.717e+00   0.763
## SeasonSpring:Campers -2.770e+01  4.961e+01  -0.558
## SeasonFall:Campers -3.067e+01  4.923e+01  -0.623
## SeasonSummer:Campers -2.524e+01  4.926e+01  -0.512
## ParkYellowstone:Campers -8.129e+01  5.985e+01  -1.358
## ParkGrand Canyon:Campers -5.528e+01  5.084e+01  -1.087

```

## SeasonSpring:Lodging:Year	-1.382e-01	1.362e-01	-1.015
## SeasonFall:Lodging:Year	-8.705e-02	1.282e-01	-0.679
## SeasonSummer:Lodging:Year	-1.669e-01	1.281e-01	-1.303
## SeasonSpring:Year:ParkYellowstone	9.678e+00	5.471e+00	1.769
## SeasonFall:Year:ParkYellowstone	8.838e+00	6.550e+00	1.349
## SeasonSummer:Year:ParkYellowstone	1.051e+02	1.900e+01	5.533
## SeasonSpring:Year:ParkGrand Canyon	2.710e+01	1.295e+01	2.093
## SeasonFall:Year:ParkGrand Canyon	-5.785e+00	9.873e+00	-0.586
## SeasonSummer:Year:ParkGrand Canyon	1.361e+01	1.046e+01	1.301
## SeasonSpring:ParkYellowstone:Campers	7.176e+01	6.017e+01	1.193
## SeasonFall:ParkYellowstone:Campers	7.810e+01	5.989e+01	1.304
## SeasonSummer:ParkYellowstone:Campers	9.054e+01	5.993e+01	1.511
## SeasonSpring:ParkGrand Canyon:Campers	5.925e+01	5.132e+01	1.155
## SeasonFall:ParkGrand Canyon:Campers	5.522e+01	5.093e+01	1.084
## SeasonSummer:ParkGrand Canyon:Campers	5.642e+01	5.083e+01	1.110
## SeasonSpring:Lodging:Campers	-2.137e-01	1.739e-01	-1.228
## SeasonFall:Lodging:Campers	-1.615e-01	1.738e-01	-0.929
## SeasonSummer:Lodging:Campers	-1.746e-01	1.731e-01	-1.008
##	Pr(> t)		
## (Intercept)	< 2e-16	***	
## SeasonSpring	0.3836		
## SeasonFall	6.34e-05	***	
## SeasonSummer	6.47e-12	***	
## Lodging	0.7501		
## Year	< 2e-16	***	
## ParkYellowstone	0.8881		
## ParkGrand Canyon	0.4571		
## I(Year^2)	< 2e-16	***	
## I(Year^3)	< 2e-16	***	
## Campers	0.3999		
## I(Campers^3)	1.23e-09	***	
## I(Campers^2)	5.47e-10	***	
## I(Lodging^2)	0.0772	.	
## SeasonSpring:Lodging	0.2991		
## SeasonFall:Lodging	0.4871		
## SeasonSummer:Lodging	0.1845		
## Lodging:Year	0.7562		
## SeasonSpring:ParkYellowstone	0.0757	.	
## SeasonFall:ParkYellowstone	0.1710		
## SeasonSummer:ParkYellowstone	5.51e-08	***	
## SeasonSpring:ParkGrand Canyon	0.0341	*	
## SeasonFall:ParkGrand Canyon	0.5783		
## SeasonSummer:ParkGrand Canyon	0.1808		
## SeasonSpring:Year	0.3786		
## SeasonFall:Year	4.43e-05	***	

```
## SeasonSummer:Year                2.95e-12 ***
## Lodging:Campers                   0.3673
## Year:ParkYellowstone              0.8993
## Year:ParkGrand Canyon             0.4459
## SeasonSpring:Campers              0.5769
## SeasonFall:Campers                0.5337
## SeasonSummer:Campers              0.6087
## ParkYellowstone:Campers           0.1751
## ParkGrand Canyon:Campers          0.2775
## SeasonSpring:Lodging:Year          0.3109
## SeasonFall:Lodging:Year           0.4975
## SeasonSummer:Lodging:Year          0.1934
## SeasonSpring:Year:ParkYellowstone 0.0777 .
## SeasonFall:Year:ParkYellowstone   0.1779
## SeasonSummer:Year:ParkYellowstone 5.59e-08 ***
## SeasonSpring:Year:ParkGrand Canyon 0.0370 *
## SeasonFall:Year:ParkGrand Canyon  0.5583
## SeasonSummer:Year:ParkGrand Canyon 0.1939
## SeasonSpring:ParkYellowstone:Campers 0.2337
## SeasonFall:ParkYellowstone:Campers  0.1929
## SeasonSummer:ParkYellowstone:Campers 0.1316
## SeasonSpring:ParkGrand Canyon:Campers 0.2489
## SeasonFall:ParkGrand Canyon:Campers  0.2789
## SeasonSummer:ParkGrand Canyon:Campers 0.2677
## SeasonSpring:Lodging:Campers         0.2200
## SeasonFall:Lodging:Campers           0.3533
## SeasonSummer:Lodging:Campers         0.3139
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.34 on 415 degrees of freedom
## Multiple R-squared:  0.9837, Adjusted R-squared:  0.9816
## F-statistic: 480.2 on 52 and 415 DF,  p-value: < 2.2e-16
```

```
summary(bw1)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Park + Year + Season + Campers + Lodging +
##      I(Campers^2) + I(Year^2) + I(Campers^3) + I(Lodging^3) +
##      I(Year^3) + Park:Year + Park:Season + Park:Campers + Park:Lodging +
##      Year:Season + Year:Campers + Year:Lodging + Season:Campers +
##      Season:Lodging + Campers:Lodging + Park:Year:Season + Park:Year:Campers +
##      Park:Year:Lodging + Park:Season:Lodging + Year:Season:Campers +
##      Year:Season:Lodging + Year:Campers:Lodging, data = NationalParkVisits)
```



```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -271.59  -47.86   -2.24   50.05  367.53
##
## Coefficients: (5 not defined because of singularities)
##
##              Estimate Std. Error t value
## (Intercept)    -3.058e+08  3.683e+07  -8.303
## ParkYellowstone    5.882e+03  5.613e+03   1.048
## ParkGrand Canyon  -1.263e+04  1.279e+04  -0.988
## Year             4.593e+05  5.529e+04   8.306
## SeasonSpring     -8.311e+02  3.852e+03  -0.216
## SeasonFall       1.299e+04  5.870e+03   2.213
## SeasonSummer    -1.843e+03  1.470e+04  -0.125
## Campers         -2.972e+02  4.914e+02  -0.605
## Lodging         -1.484e+02  2.059e+02  -0.721
## I(Campers^2)     -2.692e-02  8.140e-03  -3.307
## I(Year^2)       -2.299e+02  2.767e+01  -8.309
## I(Campers^3)     3.658e-05  8.452e-06   4.328
## I(Lodging^3)     4.503e-06  1.842e-06   2.445
## I(Year^3)        3.835e-02  4.615e-03   8.312
## ParkYellowstone:Year -2.942e+00  2.818e+00  -1.044
## ParkGrand Canyon:Year  6.365e+00  6.428e+00   0.990
## ParkYellowstone:SeasonSpring -3.188e+04  9.239e+03  -3.451
## ParkGrand Canyon:SeasonSpring -1.586e+05  2.395e+04  -6.622
## ParkYellowstone:SeasonFall -4.694e+04  2.028e+04  -2.315
## ParkGrand Canyon:SeasonFall -8.441e+04  2.140e+04  -3.945
## ParkYellowstone:SeasonSummer -1.710e+05  1.096e+05  -1.561
## ParkGrand Canyon:SeasonSummer -2.390e+05  3.744e+04  -6.383
## ParkYellowstone:Campers  7.933e+02  2.536e+02   3.128
## ParkGrand Canyon:Campers  9.191e+02  1.448e+02   6.349
## ParkYellowstone:Lodging -4.607e+02  1.062e+02  -4.339
## ParkGrand Canyon:Lodging      NA          NA      NA
## Year:SeasonSpring  4.323e-01  1.929e+00   0.224
## Year:SeasonFall   -6.449e+00  2.941e+00  -2.193
## Year:SeasonSummer  1.057e+00  7.351e+00   0.144
## Year:Campers       1.508e-01  2.453e-01   0.615
## Year:Lodging       7.574e-02  1.032e-01   0.734
## SeasonSpring:Campers -3.056e+02  4.807e+02  -0.636
## SeasonFall:Campers  -4.800e+02  4.779e+02  -1.004
## SeasonSummer:Campers  9.659e+01  4.825e+02   0.200
## SeasonSpring:Lodging  9.031e+02  2.241e+02   4.030
## SeasonFall:Lodging   7.358e+02  2.221e+02   3.313
## SeasonSummer:Lodging  8.636e+02  2.418e+02   3.572
## Campers:Lodging    -1.396e+00  4.491e-01  -3.108
```

## ParkYellowstone:Year:SeasonSpring	1.594e+01	4.640e+00	3.435
## ParkGrand Canyon:Year:SeasonSpring	7.994e+01	1.206e+01	6.629
## ParkYellowstone:Year:SeasonFall	2.349e+01	1.019e+01	2.306
## ParkGrand Canyon:Year:SeasonFall	4.227e+01	1.075e+01	3.931
## ParkYellowstone:Year:SeasonSummer	8.795e+01	5.531e+01	1.590
## ParkGrand Canyon:Year:SeasonSummer	1.202e+02	1.879e+01	6.396
## ParkYellowstone:Year:Campers	-4.012e-01	1.274e-01	-3.149
## ParkGrand Canyon:Year:Campers	-4.628e-01	7.269e-02	-6.367
## ParkYellowstone:Year:Lodging	2.285e-01	5.332e-02	4.286
## ParkGrand Canyon:Year:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSpring:Lodging	1.124e+01	3.391e+00	3.316
## ParkGrand Canyon:SeasonSpring:Lodging	NA	NA	NA
## ParkYellowstone:SeasonFall:Lodging	5.783e+00	3.116e+00	1.856
## ParkGrand Canyon:SeasonFall:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSummer:Lodging	-6.660e-01	3.883e+00	-0.172
## ParkGrand Canyon:SeasonSummer:Lodging	NA	NA	NA
## Year:SeasonSpring:Campers	1.558e-01	2.400e-01	0.649
## Year:SeasonFall:Campers	2.447e-01	2.385e-01	1.026
## Year:SeasonSummer:Campers	-4.450e-02	2.408e-01	-0.185
## Year:SeasonSpring:Lodging	-4.550e-01	1.124e-01	-4.049
## Year:SeasonFall:Lodging	-3.698e-01	1.114e-01	-3.320
## Year:SeasonSummer:Lodging	-4.347e-01	1.212e-01	-3.586
## Year:Campers:Lodging	7.001e-04	2.255e-04	3.105
##	Pr(> t)		
## (Intercept)	1.47e-15	***	
## ParkYellowstone	0.295308		
## ParkGrand Canyon	0.323704		
## Year	1.44e-15	***	
## SeasonSpring	0.829264		
## SeasonFall	0.027426	*	
## SeasonSummer	0.900285		
## Campers	0.545636		
## Lodging	0.471291		
## I(Campers^2)	0.001025	**	
## I(Year^2)	1.41e-15	***	
## I(Campers^3)	1.89e-05	***	
## I(Lodging^3)	0.014892	*	
## I(Year^3)	1.38e-15	***	
## ParkYellowstone:Year	0.297066		
## ParkGrand Canyon:Year	0.322676		
## ParkYellowstone:SeasonSpring	0.000616	***	
## ParkGrand Canyon:SeasonSpring	1.11e-10	***	
## ParkYellowstone:SeasonFall	0.021116	*	
## ParkGrand Canyon:SeasonFall	9.39e-05	***	
## ParkYellowstone:SeasonSummer	0.119349		

```

## ParkGrand Canyon:SeasonSummer      4.70e-10 ***
## ParkYellowstone:Campers              0.001882 **
## ParkGrand Canyon:Campers             5.73e-10 ***
## ParkYellowstone:Lodging              1.81e-05 ***
## ParkGrand Canyon:Lodging              NA
## Year:SeasonSpring                    0.822797
## Year:SeasonFall                      0.028894 *
## Year:SeasonSummer                    0.885721
## Year:Campers                         0.539058
## Year:Lodging                         0.463374
## SeasonSpring:Campers                 0.525252
## SeasonFall:Campers                  0.315731
## SeasonSummer:Campers                 0.841422
## SeasonSpring:Lodging                 6.64e-05 ***
## SeasonFall:Lodging                  0.001003 **
## SeasonSummer:Lodging                 0.000396 ***
## Campers:Lodging                     0.002015 **
## ParkYellowstone:Year:SeasonSpring    0.000653 ***
## ParkGrand Canyon:Year:SeasonSpring   1.06e-10 ***
## ParkYellowstone:Year:SeasonFall      0.021617 *
## ParkGrand Canyon:Year:SeasonFall     9.93e-05 ***
## ParkYellowstone:Year:SeasonSummer    0.112547
## ParkGrand Canyon:Year:SeasonSummer   4.34e-10 ***
## ParkYellowstone:Year:Campers          0.001756 **
## ParkGrand Canyon:Year:Campers         5.14e-10 ***
## ParkYellowstone:Year:Lodging          2.27e-05 ***
## ParkGrand Canyon:Year:Lodging         NA
## ParkYellowstone:SeasonSpring:Lodging  0.000995 ***
## ParkGrand Canyon:SeasonSpring:Lodging NA
## ParkYellowstone:SeasonFall:Lodging    0.064214 .
## ParkGrand Canyon:SeasonFall:Lodging   NA
## ParkYellowstone:SeasonSummer:Lodging  0.863900
## ParkGrand Canyon:SeasonSummer:Lodging NA
## Year:SeasonSpring:Campers             0.516516
## Year:SeasonFall:Campers               0.305511
## Year:SeasonSummer:Campers             0.853505
## Year:SeasonSpring:Lodging             6.15e-05 ***
## Year:SeasonFall:Lodging              0.000979 ***
## Year:SeasonSummer:Lodging            0.000376 ***
## Year:Campers:Lodging                  0.002037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.18 on 412 degrees of freedom
## Multiple R-squared:  0.9861, Adjusted R-squared:  0.9842

```

```
## F-statistic: 531.1 on 55 and 412 DF,  p-value: < 2.2e-16
```

```
summary(st1)
```

```
##
```

```
## Call:
```

```
## lm(formula = TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) +  
##      I(Year^3) + Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) +  
##      Season:Lodging + Lodging:Year + Season:Park + Season:Year +  
##      Lodging:Campers + Year:Park + Season:Campers + Park:Campers +  
##      Season:Lodging:Year + Season:Year:Park + Season:Park:Campers +  
##      Season:Lodging:Campers, data = NationalParkVisits)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -313.44  -53.08   -1.95   53.76  406.77
```

```
##
```

```
## Coefficients:
```

	Estimate	Std. Error	t value
## (Intercept)	-3.835e+08	3.496e+07	-10.970
## SeasonSpring	-3.941e+03	4.518e+03	-0.872
## SeasonFall	-1.726e+04	4.271e+03	-4.041
## SeasonSummer	-3.360e+04	4.751e+03	-7.073
## Lodging	-8.063e+01	2.530e+02	-0.319
## Year	5.757e+05	5.248e+04	10.970
## ParkYellowstone	1.007e+03	7.150e+03	0.141
## ParkGrand Canyon	-1.286e+04	1.728e+04	-0.744
## I(Year^2)	-2.881e+02	2.626e+01	-10.970
## I(Year^3)	4.806e-02	4.381e-03	10.970
## Campers	4.147e+01	4.921e+01	0.843
## I(Campers^3)	4.434e-05	7.132e-06	6.217
## I(Campers^2)	-4.465e-02	7.026e-03	-6.356
## I(Lodging^2)	1.716e-03	9.683e-04	1.772
## SeasonSpring:Lodging	2.808e+02	2.701e+02	1.040
## SeasonFall:Lodging	1.769e+02	2.543e+02	0.695
## SeasonSummer:Lodging	3.379e+02	2.542e+02	1.329
## Lodging:Year	3.962e-02	1.275e-01	0.311
## SeasonSpring:ParkYellowstone	-1.940e+04	1.090e+04	-1.780
## SeasonFall:ParkYellowstone	-1.791e+04	1.306e+04	-1.372
## SeasonSummer:ParkYellowstone	-2.120e+05	3.830e+04	-5.535
## SeasonSpring:ParkGrand Canyon	-5.458e+04	2.567e+04	-2.126
## SeasonFall:ParkGrand Canyon	1.089e+04	1.957e+04	0.556
## SeasonSummer:ParkGrand Canyon	-2.789e+04	2.080e+04	-1.341
## SeasonSpring:Year	1.996e+00	2.264e+00	0.881
## SeasonFall:Year	8.761e+00	2.123e+00	4.128

## SeasonSummer:Year	1.683e+01	2.339e+00	7.195
## Lodging:Campers	1.562e-01	1.731e-01	0.903
## Year:ParkYellowstone	-4.529e-01	3.575e+00	-0.127
## Year:ParkGrand Canyon	6.651e+00	8.717e+00	0.763
## SeasonSpring:Campers	-2.770e+01	4.961e+01	-0.558
## SeasonFall:Campers	-3.067e+01	4.923e+01	-0.623
## SeasonSummer:Campers	-2.524e+01	4.926e+01	-0.512
## ParkYellowstone:Campers	-8.129e+01	5.985e+01	-1.358
## ParkGrand Canyon:Campers	-5.528e+01	5.084e+01	-1.087
## SeasonSpring:Lodging:Year	-1.382e-01	1.362e-01	-1.015
## SeasonFall:Lodging:Year	-8.705e-02	1.282e-01	-0.679
## SeasonSummer:Lodging:Year	-1.669e-01	1.281e-01	-1.303
## SeasonSpring:Year:ParkYellowstone	9.678e+00	5.471e+00	1.769
## SeasonFall:Year:ParkYellowstone	8.838e+00	6.550e+00	1.349
## SeasonSummer:Year:ParkYellowstone	1.051e+02	1.900e+01	5.533
## SeasonSpring:Year:ParkGrand Canyon	2.710e+01	1.295e+01	2.093
## SeasonFall:Year:ParkGrand Canyon	-5.785e+00	9.873e+00	-0.586
## SeasonSummer:Year:ParkGrand Canyon	1.361e+01	1.046e+01	1.301
## SeasonSpring:ParkYellowstone:Campers	7.176e+01	6.017e+01	1.193
## SeasonFall:ParkYellowstone:Campers	7.810e+01	5.989e+01	1.304
## SeasonSummer:ParkYellowstone:Campers	9.054e+01	5.993e+01	1.511
## SeasonSpring:ParkGrand Canyon:Campers	5.925e+01	5.132e+01	1.155
## SeasonFall:ParkGrand Canyon:Campers	5.522e+01	5.093e+01	1.084
## SeasonSummer:ParkGrand Canyon:Campers	5.642e+01	5.083e+01	1.110
## SeasonSpring:Lodging:Campers	-2.137e-01	1.739e-01	-1.228
## SeasonFall:Lodging:Campers	-1.615e-01	1.738e-01	-0.929
## SeasonSummer:Lodging:Campers	-1.746e-01	1.731e-01	-1.008
##	Pr(> t)		
## (Intercept)	< 2e-16 ***		
## SeasonSpring	0.3836		
## SeasonFall	6.34e-05 ***		
## SeasonSummer	6.47e-12 ***		
## Lodging	0.7501		
## Year	< 2e-16 ***		
## ParkYellowstone	0.8881		
## ParkGrand Canyon	0.4571		
## I(Year^2)	< 2e-16 ***		
## I(Year^3)	< 2e-16 ***		
## Campers	0.3999		
## I(Campers^3)	1.23e-09 ***		
## I(Campers^2)	5.47e-10 ***		
## I(Lodging^2)	0.0772 .		
## SeasonSpring:Lodging	0.2991		
## SeasonFall:Lodging	0.4871		
## SeasonSummer:Lodging	0.1845		

```

## Lodging:Year                                0.7562
## SeasonSpring:ParkYellowstone                0.0757 .
## SeasonFall:ParkYellowstone                 0.1710
## SeasonSummer:ParkYellowstone               5.51e-08 ***
## SeasonSpring:ParkGrand Canyon              0.0341 *
## SeasonFall:ParkGrand Canyon               0.5783
## SeasonSummer:ParkGrand Canyon             0.1808
## SeasonSpring:Year                          0.3786
## SeasonFall:Year                          4.43e-05 ***
## SeasonSummer:Year                        2.95e-12 ***
## Lodging:Campers                          0.3673
## Year:ParkYellowstone                     0.8993
## Year:ParkGrand Canyon                   0.4459
## SeasonSpring:Campers                    0.5769
## SeasonFall:Campers                     0.5337
## SeasonSummer:Campers                   0.6087
## ParkYellowstone:Campers                 0.1751
## ParkGrand Canyon:Campers               0.2775
## SeasonSpring:Lodging:Year               0.3109
## SeasonFall:Lodging:Year                0.4975
## SeasonSummer:Lodging:Year              0.1934
## SeasonSpring:Year:ParkYellowstone       0.0777 .
## SeasonFall:Year:ParkYellowstone        0.1779
## SeasonSummer:Year:ParkYellowstone      5.59e-08 ***
## SeasonSpring:Year:ParkGrand Canyon     0.0370 *
## SeasonFall:Year:ParkGrand Canyon       0.5583
## SeasonSummer:Year:ParkGrand Canyon     0.1939
## SeasonSpring:ParkYellowstone:Campers   0.2337
## SeasonFall:ParkYellowstone:Campers     0.1929
## SeasonSummer:ParkYellowstone:Campers   0.1316
## SeasonSpring:ParkGrand Canyon:Campers  0.2489
## SeasonFall:ParkGrand Canyon:Campers    0.2789
## SeasonSummer:ParkGrand Canyon:Campers  0.2677
## SeasonSpring:Lodging:Campers            0.2200
## SeasonFall:Lodging:Campers              0.3533
## SeasonSummer:Lodging:Campers            0.3139
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 96.34 on 415 degrees of freedom
## Multiple R-squared:  0.9837, Adjusted R-squared:  0.9816
## F-statistic: 480.2 on 52 and 415 DF,  p-value: < 2.2e-16

c(summary(fw1)$adj, summary(bw1)$adj, summary(st1)$adj)

```

```
## [1] 0.9816029 0.9842355 0.9816029
```

```
anova(fw1, bw1, st1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) + I(Year^3) +  
##   Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) + Season:Lodging +  
##   Lodging:Year + Season:Park + Season:Year + Lodging:Campers +  
##   Year:Park + Season:Campers + Park:Campers + Season:Lodging:Year +  
##   Season:Year:Park + Season:Park:Campers + Season:Lodging:Campers
```

```
## Model 2: TotalVisits ~ Park + Year + Season + Campers + Lodging + I(Campers^2) +  
##   I(Year^2) + I(Campers^3) + I(Lodging^3) + I(Year^3) + Park:Year +  
##   Park:Season + Park:Campers + Park:Lodging + Year:Season +  
##   Year:Campers + Year:Lodging + Season:Campers + Season:Lodging +  
##   Campers:Lodging + Park:Year:Season + Park:Year:Campers +  
##   Park:Year:Lodging + Park:Season:Lodging + Year:Season:Campers +  
##   Year:Season:Lodging + Year:Campers:Lodging
```

```
## Model 3: TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) + I(Year^3) +  
##   Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) + Season:Lodging +  
##   Lodging:Year + Season:Park + Season:Year + Lodging:Campers +  
##   Year:Park + Season:Campers + Park:Campers + Season:Lodging:Year +  
##   Season:Year:Park + Season:Park:Campers + Season:Lodging:Campers
```

```
##   Res.Df      RSS Df Sum of Sq      F      Pr(>F)  
## 1      415 3851501  
## 2      412 3276508   3      574993 24.101 2.178e-14 ***  
## 3      415 3851501 -3     -574993 24.101 2.178e-14 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fw1, bw1, st1, test="Cp")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) + I(Year^3) +  
##   Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) + Season:Lodging +  
##   Lodging:Year + Season:Park + Season:Year + Lodging:Campers +  
##   Year:Park + Season:Campers + Park:Campers + Season:Lodging:Year +  
##   Season:Year:Park + Season:Park:Campers + Season:Lodging:Campers
```

```
## Model 2: TotalVisits ~ Park + Year + Season + Campers + Lodging + I(Campers^2) +  
##   I(Year^2) + I(Campers^3) + I(Lodging^3) + I(Year^3) + Park:Year +  
##   Park:Season + Park:Campers + Park:Lodging + Year:Season +  
##   Year:Campers + Year:Lodging + Season:Campers + Season:Lodging +  
##   Campers:Lodging + Park:Year:Season + Park:Year:Campers +  
##   Park:Year:Lodging + Park:Season:Lodging + Year:Season:Campers +  
##   Year:Season:Lodging + Year:Campers:Lodging
```

```
## Model 3: TotalVisits ~ Season + Lodging + Year + Park + I(Year^2) + I(Year^3) +
##   Campers + I(Campers^3) + I(Campers^2) + I(Lodging^2) + Season:Lodging +
##   Lodging:Year + Season:Park + Season:Year + Lodging:Campers +
##   Year:Park + Season:Campers + Park:Campers + Season:Lodging:Year +
##   Season:Year:Park + Season:Park:Campers + Season:Lodging:Campers
##   Res.Df      RSS Df Sum of Sq      Cp
## 1    415 3851501      4694486
## 2    412 3276508   3    574993 4167209
## 3    415 3851501  -3   -574993 4694486
```

```
AIC(fw1, bw1, st1)
```

```
##      df      AIC
## fw1 54 5655.383
## bw1 57 5585.714
## st1 54 5655.383
```

```
#With FullModel2 (with Year5)
```

```
fw2 <- step(zeroModel, scope=list(upper=fullModel2), direction="forward", trace=0)
bw2 <- step(fullModel2, scope=list(lower=zeroModel), direction="backward", trace=0)
st2 <- step(zeroModel, scope=list(upper=fullModel2), direction="both", trace=0)
summary(fw2)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Season + Lodging + Year5 + Park +
##   I(Lodging^3) + Campers + I(Campers^3) + I(Lodging^2) + Season:Lodging +
##   Lodging:Year5 + Season:Park + Season:Year5 + Lodging:Campers,
##   data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -410.57  -48.43   -2.05   44.72  468.90
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.504e+02  6.895e+01   2.181 0.029736 *
## SeasonSpring   -5.788e+01  9.738e+01  -0.594 0.552583
## SeasonFall     2.597e+02  1.005e+02   2.584 0.010086 *
## SeasonSummer    6.917e+02  1.358e+02   5.094 5.28e-07 ***
## Lodging         4.526e+00  1.394e+00   3.246 0.001264 **
## Year580s       -1.152e+01  6.903e+01  -0.167 0.867536
## Year590s        4.229e+01  7.141e+01   0.592 0.554006
## Year500s        7.948e+01  7.118e+01   1.117 0.264801
## Year510s        1.026e+02  7.221e+01   1.421 0.155957
## ParkYellowstone -1.415e+02  3.866e+01  -3.660 0.000284 ***
```



```
## ParkGrand Canyon      -1.158e+02  9.121e+01  -1.270  0.204907
## I(Lodging^3)           7.258e-06  2.664e-06   2.724  0.006711 **
## Campers                2.763e+00  3.366e-01   8.208  2.69e-15 ***
## I(Campers^3)           4.845e-06  1.914e-06   2.531  0.011729 *
## I(Lodging^2)          -6.921e-03  3.848e-03  -1.799  0.072778 .
## SeasonSpring:Lodging   -8.630e-01  1.522e+00  -0.567  0.571021
## SeasonFall:Lodging     -9.632e-01  1.439e+00  -0.669  0.503743
## SeasonSummer:Lodging   -3.457e+00  1.953e+00  -1.770  0.077470 .
## Lodging:Year580s       6.124e-01  2.996e-01   2.044  0.041537 *
## Lodging:Year590s       9.104e-01  3.114e-01   2.924  0.003641 **
## Lodging:Year500s       7.250e-01  3.208e-01   2.260  0.024352 *
## Lodging:Year510s       8.825e-01  3.256e-01   2.710  0.006999 **
## SeasonSpring:ParkYellowstone -6.306e+00  6.097e+01  -0.103  0.917666
## SeasonFall:ParkYellowstone -3.767e+02  8.313e+01  -4.532  7.59e-06 ***
## SeasonSummer:ParkYellowstone 1.050e+03  4.129e+02   2.543  0.011350 *
## SeasonSpring:ParkGrand Canyon 1.458e+02  1.520e+02   0.960  0.337799
## SeasonFall:ParkGrand Canyon -2.365e+02  1.240e+02  -1.908  0.057114 .
## SeasonSummer:ParkGrand Canyon 1.954e+02  2.778e+02   0.703  0.482204
## SeasonSpring:Year580s    1.093e+02  9.758e+01   1.120  0.263434
## SeasonFall:Year580s     1.182e+02  9.884e+01   1.196  0.232417
## SeasonSummer:Year580s    2.447e+02  1.189e+02   2.058  0.040178 *
## SeasonSpring:Year590s    1.631e+02  1.003e+02   1.625  0.104812
## SeasonFall:Year590s     2.046e+02  1.011e+02   2.025  0.043477 *
## SeasonSummer:Year590s    5.162e+02  1.209e+02   4.271  2.40e-05 ***
## SeasonSpring:Year500s    1.757e+02  1.010e+02   1.740  0.082630 .
## SeasonFall:Year500s     1.646e+02  1.018e+02   1.617  0.106600
## SeasonSummer:Year500s    4.034e+02  1.210e+02   3.333  0.000935 ***
## SeasonSpring:Year510s    2.287e+02  1.025e+02   2.230  0.026246 *
## SeasonFall:Year510s     2.647e+02  1.040e+02   2.546  0.011252 *
## SeasonSummer:Year510s    6.625e+02  1.236e+02   5.362  1.35e-07 ***
## Lodging:Campers        -8.147e-03  1.376e-03  -5.923  6.51e-09 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 112.8 on 427 degrees of freedom
```

```
## Multiple R-squared:  0.9769, Adjusted R-squared:  0.9748
```

```
## F-statistic: 452.2 on 40 and 427 DF,  p-value: < 2.2e-16
```

```
summary(bw2)
```

```
##
```

```
## Call:
```

```
## lm(formula = TotalVisits ~ Park + Year5 + Season + Campers +
```

```
##     Lodging + I(Lodging^2) + Park:Year5 + Park:Season + Park:Campers +
```

```
##     Park:Lodging + Year5:Season + Year5:Campers + Year5:Lodging +
```

```

##      Season:Campers + Season:Lodging + Campers:Lodging + Park:Year5:Season +
##      Park:Year5:Campers + Park:Year5:Lodging + Park:Season:Lodging +
##      Park:Campers:Lodging + Year5:Season:Campers + Year5:Season:Lodging +
##      Year5:Campers:Lodging + Season:Campers:Lodging, data = NationalParkVisits)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -322.35   -35.78     0.00    32.06   425.78
##
## Coefficients: (21 not defined because of singularities)
##
##              Estimate Std. Error t value
## (Intercept)      2.341e+02  9.934e+01   2.357
## ParkYellowstone  -1.080e+02  1.456e+02  -0.742
## ParkGrand Canyon   3.384e+02  3.741e+02   0.905
## Year580s          -4.328e+01  9.828e+01  -0.440
## Year590s          -3.859e+01  9.692e+01  -0.398
## Year500s          -1.224e+01  9.684e+01  -0.126
## Year510s           1.856e+01  9.697e+01   0.191
## SeasonSpring      -9.869e+01  1.378e+02  -0.716
## SeasonFall        -5.545e+02  5.300e+02  -1.046
## SeasonSummer       1.993e+03  7.335e+02   2.718
## Campers           -2.108e+01  2.654e+01  -0.794
## Lodging           -6.971e+00  7.545e+00  -0.924
## I(Lodging^2)       2.459e-02  6.763e-03   3.637
## ParkYellowstone:Year580s  6.942e+01  1.606e+02   0.432
## ParkGrand Canyon:Year580s -3.773e+02  3.280e+02  -1.150
## ParkYellowstone:Year590s  1.766e+02  1.776e+02   0.994
## ParkGrand Canyon:Year590s  4.477e+02  3.032e+02   1.477
## ParkYellowstone:Year500s  2.159e+02  1.752e+02   1.233
## ParkGrand Canyon:Year500s  3.279e+02  3.420e+02   0.959
## ParkYellowstone:Year510s -9.118e+00  1.482e+02  -0.062
## ParkGrand Canyon:Year510s -9.918e+01  3.591e+02  -0.276
## ParkYellowstone:SeasonSpring -2.027e+02  2.321e+02  -0.873
## ParkGrand Canyon:SeasonSpring  3.612e+03  9.260e+02   3.901
## ParkYellowstone:SeasonFall -7.367e+02  9.812e+02  -0.751
## ParkGrand Canyon:SeasonFall  2.751e+03  8.141e+02   3.380
## ParkYellowstone:SeasonSummer  1.248e+04  4.595e+03   2.717
## ParkGrand Canyon:SeasonSummer  5.579e+03  1.411e+03   3.954
## ParkYellowstone:Campers    1.642e+01  9.894e+00   1.660
## ParkGrand Canyon:Campers   -1.890e+01  5.854e+00  -3.228
## ParkYellowstone:Lodging     2.424e+00  6.066e+00   0.400
## ParkGrand Canyon:Lodging      NA         NA         NA
## Year580s:SeasonSpring      1.307e+02  1.399e+02   0.934
## Year590s:SeasonSpring      5.081e+01  1.387e+02   0.366
## Year500s:SeasonSpring     -7.019e+01  1.462e+02  -0.480

```

## Year510s:SeasonSpring	5.102e+01	1.342e+02	0.380
## Year580s:SeasonFall	7.598e+02	5.367e+02	1.416
## Year590s:SeasonFall	8.687e+02	5.476e+02	1.586
## Year500s:SeasonFall	7.236e+02	5.418e+02	1.336
## Year510s:SeasonFall	4.384e+02	4.846e+02	0.905
## Year580s:SeasonSummer	-8.109e+02	7.498e+02	-1.082
## Year590s:SeasonSummer	-8.780e+02	8.943e+02	-0.982
## Year500s:SeasonSummer	-1.196e+03	8.155e+02	-1.467
## Year510s:SeasonSummer	-1.755e+03	7.193e+02	-2.439
## Year580s:Campers	6.547e+00	1.677e+01	0.390
## Year590s:Campers	1.551e+01	1.343e+01	1.155
## Year500s:Campers	1.536e+01	1.110e+01	1.384
## Year510s:Campers	1.387e+01	3.300e+00	4.203
## Year580s:Lodging	9.684e+00	5.920e+00	1.636
## Year590s:Lodging	-2.997e+00	5.233e+00	-0.573
## Year500s:Lodging	8.345e-01	5.489e+00	0.152
## Year510s:Lodging	7.479e+00	2.125e+00	3.520
## SeasonSpring:Campers	3.072e+01	2.522e+01	1.218
## SeasonFall:Campers	2.858e+01	2.520e+01	1.134
## SeasonSummer:Campers	1.802e+01	2.515e+01	0.717
## SeasonSpring:Lodging	-1.543e+01	8.869e+00	-1.740
## SeasonFall:Lodging	-1.044e+01	8.354e+00	-1.250
## SeasonSummer:Lodging	-2.327e+01	9.115e+00	-2.553
## Campers:Lodging	2.834e-01	2.998e-01	0.945
## ParkYellowstone:Year580s:SeasonSpring	1.857e+02	2.741e+02	0.678
## ParkGrand Canyon:Year580s:SeasonSpring	-2.714e+03	7.733e+02	-3.510
## ParkYellowstone:Year590s:SeasonSpring	-5.738e+02	3.267e+02	-1.756
## ParkGrand Canyon:Year590s:SeasonSpring	-3.752e+03	7.455e+02	-5.032
## ParkYellowstone:Year500s:SeasonSpring	2.490e+02	3.520e+02	0.707
## ParkGrand Canyon:Year500s:SeasonSpring	-2.492e+03	7.884e+02	-3.161
## ParkYellowstone:Year510s:SeasonSpring	2.561e+02	2.162e+02	1.184
## ParkGrand Canyon:Year510s:SeasonSpring	-1.683e+02	5.248e+02	-0.321
## ParkYellowstone:Year580s:SeasonFall	6.663e+02	9.890e+02	0.674
## ParkGrand Canyon:Year580s:SeasonFall	-2.746e+03	6.902e+02	-3.979
## ParkYellowstone:Year590s:SeasonFall	7.618e+02	1.031e+03	0.739
## ParkGrand Canyon:Year590s:SeasonFall	-1.383e+03	7.514e+02	-1.840
## ParkYellowstone:Year500s:SeasonFall	1.462e+03	1.061e+03	1.378
## ParkGrand Canyon:Year500s:SeasonFall	-1.360e+03	8.273e+02	-1.644
## ParkYellowstone:Year510s:SeasonFall	-2.000e+02	6.209e+02	-0.322
## ParkGrand Canyon:Year510s:SeasonFall	-1.718e+03	3.712e+02	-4.627
## ParkYellowstone:Year580s:SeasonSummer	-4.555e+03	2.821e+03	-1.615
## ParkGrand Canyon:Year580s:SeasonSummer	-5.108e+03	1.139e+03	-4.484
## ParkYellowstone:Year590s:SeasonSummer	-2.324e+03	2.940e+03	-0.790
## ParkGrand Canyon:Year590s:SeasonSummer	-2.332e+03	1.147e+03	-2.033
## ParkYellowstone:Year500s:SeasonSummer	4.844e+03	2.270e+03	2.134

## ParkGrand Canyon:Year500s:SeasonSummer	-1.083e+02	1.352e+03	-0.080
## ParkYellowstone:Year510s:SeasonSummer	NA	NA	NA
## ParkGrand Canyon:Year510s:SeasonSummer	NA	NA	NA
## ParkYellowstone:Year580s:Campers	-1.781e+01	1.012e+01	-1.760
## ParkGrand Canyon:Year580s:Campers	1.056e+01	4.103e+00	2.575
## ParkYellowstone:Year590s:Campers	-2.008e+01	1.095e+01	-1.834
## ParkGrand Canyon:Year590s:Campers	8.062e+00	4.991e+00	1.615
## ParkYellowstone:Year500s:Campers	-3.072e+01	1.394e+01	-2.204
## ParkGrand Canyon:Year500s:Campers	-3.577e+00	5.623e+00	-0.636
## ParkYellowstone:Year510s:Campers	NA	NA	NA
## ParkGrand Canyon:Year510s:Campers	NA	NA	NA
## ParkYellowstone:Year580s:Lodging	-8.252e+00	4.441e+00	-1.858
## ParkGrand Canyon:Year580s:Lodging	NA	NA	NA
## ParkYellowstone:Year590s:Lodging	4.339e+00	3.442e+00	1.261
## ParkGrand Canyon:Year590s:Lodging	NA	NA	NA
## ParkYellowstone:Year500s:Lodging	-1.356e+00	2.965e+00	-0.457
## ParkGrand Canyon:Year500s:Lodging	NA	NA	NA
## ParkYellowstone:Year510s:Lodging	NA	NA	NA
## ParkGrand Canyon:Year510s:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSpring:Lodging	9.798e+00	6.660e+00	1.471
## ParkGrand Canyon:SeasonSpring:Lodging	NA	NA	NA
## ParkYellowstone:SeasonFall:Lodging	4.191e+00	5.107e+00	0.821
## ParkGrand Canyon:SeasonFall:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSummer:Lodging	-1.412e+01	7.787e+00	-1.813
## ParkGrand Canyon:SeasonSummer:Lodging	NA	NA	NA
## ParkYellowstone:Campers:Lodging	-4.653e-02	2.170e-02	-2.144
## ParkGrand Canyon:Campers:Lodging	NA	NA	NA
## Year580s:SeasonSpring:Campers	-1.452e+01	1.548e+01	-0.937
## Year590s:SeasonSpring:Campers	-9.085e+00	1.162e+01	-0.782
## Year500s:SeasonSpring:Campers	-3.754e-01	9.519e+00	-0.039
## Year510s:SeasonSpring:Campers	NA	NA	NA
## Year580s:SeasonFall:Campers	-8.360e+00	1.545e+01	-0.541
## Year590s:SeasonFall:Campers	-1.680e+01	1.170e+01	-1.436
## Year500s:SeasonFall:Campers	-1.236e+01	9.244e+00	-1.337
## Year510s:SeasonFall:Campers	NA	NA	NA
## Year580s:SeasonSummer:Campers	-2.249e+00	1.546e+01	-0.145
## Year590s:SeasonSummer:Campers	-9.255e+00	1.169e+01	-0.792
## Year500s:SeasonSummer:Campers	-7.163e+00	9.277e+00	-0.772
## Year510s:SeasonSummer:Campers	NA	NA	NA
## Year580s:SeasonSpring:Lodging	1.084e+01	6.605e+00	1.642
## Year590s:SeasonSpring:Lodging	2.295e+01	6.238e+00	3.680
## Year500s:SeasonSpring:Lodging	1.194e+01	6.325e+00	1.888
## Year510s:SeasonSpring:Lodging	NA	NA	NA
## Year580s:SeasonFall:Lodging	5.251e+00	6.175e+00	0.850
## Year590s:SeasonFall:Lodging	6.005e+00	5.963e+00	1.007

## Year500s:SeasonFall:Lodging	4.730e+00	6.700e+00	0.706
## Year510s:SeasonFall:Lodging	NA	NA	NA
## Year580s:SeasonSummer:Lodging	1.129e+01	6.540e+00	1.726
## Year590s:SeasonSummer:Lodging	9.884e+00	6.405e+00	1.543
## Year500s:SeasonSummer:Lodging	6.075e-01	7.165e+00	0.085
## Year510s:SeasonSummer:Lodging	NA	NA	NA
## Year580s:Campers:Lodging	2.197e-02	1.453e-02	1.511
## Year590s:Campers:Lodging	1.888e-02	1.532e-02	1.232
## Year500s:Campers:Lodging	4.593e-02	2.051e-02	2.239
## Year510s:Campers:Lodging	NA	NA	NA
## SeasonSpring:Campers:Lodging	-3.297e-01	2.923e-01	-1.128
## SeasonFall:Campers:Lodging	-2.652e-01	2.920e-01	-0.908
## SeasonSummer:Campers:Lodging	-2.547e-01	2.939e-01	-0.867
##	Pr(> t)		
## (Intercept)	0.018965	*	
## ParkYellowstone	0.458824		
## ParkGrand Canyon	0.366328		
## Year580s	0.659912		
## Year590s	0.690730		
## Year500s	0.899528		
## Year510s	0.848370		
## SeasonSpring	0.474208		
## SeasonFall	0.296167		
## SeasonSummer	0.006898	**	
## Campers	0.427581		
## Lodging	0.356162		
## I(Lodging^2)	0.000317	***	
## ParkYellowstone:Year580s	0.665853		
## ParkGrand Canyon:Year580s	0.250787		
## ParkYellowstone:Year590s	0.320712		
## ParkGrand Canyon:Year590s	0.140678		
## ParkYellowstone:Year500s	0.218518		
## ParkGrand Canyon:Year500s	0.338429		
## ParkYellowstone:Year510s	0.950991		
## ParkGrand Canyon:Year510s	0.782561		
## ParkYellowstone:SeasonSpring	0.383095		
## ParkGrand Canyon:SeasonSpring	0.000115	***	
## ParkYellowstone:SeasonFall	0.453270		
## ParkGrand Canyon:SeasonFall	0.000807	***	
## ParkYellowstone:SeasonSummer	0.006918	**	
## ParkGrand Canyon:SeasonSummer	9.31e-05	***	
## ParkYellowstone:Campers	0.097810	.	
## ParkGrand Canyon:Campers	0.001362	**	
## ParkYellowstone:Lodging	0.689691		
## ParkGrand Canyon:Lodging	NA		

## Year580s:SeasonSpring	0.350797
## Year590s:SeasonSpring	0.714367
## Year500s:SeasonSpring	0.631459
## Year510s:SeasonSpring	0.703996
## Year580s:SeasonFall	0.157745
## Year590s:SeasonFall	0.113549
## Year500s:SeasonFall	0.182513
## Year510s:SeasonFall	0.366328
## Year580s:SeasonSummer	0.280188
## Year590s:SeasonSummer	0.326895
## Year500s:SeasonSummer	0.143398
## Year510s:SeasonSummer	0.015218 *
## Year580s:Campers	0.696411
## Year590s:Campers	0.248822
## Year500s:Campers	0.167254
## Year510s:Campers	3.35e-05 ***
## Year580s:Lodging	0.102783
## Year590s:Lodging	0.567162
## Year500s:Lodging	0.879255
## Year510s:Lodging	0.000489 ***
## SeasonSpring:Campers	0.224019
## SeasonFall:Campers	0.257590
## SeasonSummer:Campers	0.474047
## SeasonSpring:Lodging	0.082731 .
## SeasonFall:Lodging	0.212037
## SeasonSummer:Lodging	0.011097 *
## Campers:Lodging	0.345100
## ParkYellowstone:Year580s:SeasonSpring	0.498437
## ParkGrand Canyon:Year580s:SeasonSpring	0.000507 ***
## ParkYellowstone:Year590s:SeasonSpring	0.079909 .
## ParkGrand Canyon:Year590s:SeasonSpring	7.75e-07 ***
## ParkYellowstone:Year500s:SeasonSpring	0.479786
## ParkGrand Canyon:Year500s:SeasonSpring	0.001711 **
## ParkYellowstone:Year510s:SeasonSpring	0.237089
## ParkGrand Canyon:Year510s:SeasonSpring	0.748600
## ParkYellowstone:Year580s:SeasonFall	0.500946
## ParkGrand Canyon:Year580s:SeasonFall	8.41e-05 ***
## ParkYellowstone:Year590s:SeasonFall	0.460503
## ParkGrand Canyon:Year590s:SeasonFall	0.066573 .
## ParkYellowstone:Year500s:SeasonFall	0.169158
## ParkGrand Canyon:Year500s:SeasonFall	0.101093
## ParkYellowstone:Year510s:SeasonFall	0.747613
## ParkGrand Canyon:Year510s:SeasonFall	5.23e-06 ***
## ParkYellowstone:Year580s:SeasonSummer	0.107193
## ParkGrand Canyon:Year580s:SeasonSummer	9.91e-06 ***

## ParkYellowstone:Year590s:SeasonSummer	0.429878	
## ParkGrand Canyon:Year590s:SeasonSummer	0.042834	*
## ParkYellowstone:Year500s:SeasonSummer	0.033569	*
## ParkGrand Canyon:Year500s:SeasonSummer	0.936177	
## ParkYellowstone:Year510s:SeasonSummer	NA	
## ParkGrand Canyon:Year510s:SeasonSummer	NA	
## ParkYellowstone:Year580s:Campers	0.079246	.
## ParkGrand Canyon:Year580s:Campers	0.010443	*
## ParkYellowstone:Year590s:Campers	0.067530	.
## ParkGrand Canyon:Year590s:Campers	0.107137	
## ParkYellowstone:Year500s:Campers	0.028203	*
## ParkGrand Canyon:Year500s:Campers	0.525088	
## ParkYellowstone:Year510s:Campers	NA	
## ParkGrand Canyon:Year510s:Campers	NA	
## ParkYellowstone:Year580s:Lodging	0.063967	.
## ParkGrand Canyon:Year580s:Lodging	NA	
## ParkYellowstone:Year590s:Lodging	0.208320	
## ParkGrand Canyon:Year590s:Lodging	NA	
## ParkYellowstone:Year500s:Lodging	0.647610	
## ParkGrand Canyon:Year500s:Lodging	NA	
## ParkYellowstone:Year510s:Lodging	NA	
## ParkGrand Canyon:Year510s:Lodging	NA	
## ParkYellowstone:SeasonSpring:Lodging	0.142110	
## ParkGrand Canyon:SeasonSpring:Lodging	NA	
## ParkYellowstone:SeasonFall:Lodging	0.412408	
## ParkGrand Canyon:SeasonFall:Lodging	NA	
## ParkYellowstone:SeasonSummer:Lodging	0.070663	.
## ParkGrand Canyon:SeasonSummer:Lodging	NA	
## ParkYellowstone:Campers:Lodging	0.032710	*
## ParkGrand Canyon:Campers:Lodging	NA	
## Year580s:SeasonSpring:Campers	0.349194	
## Year590s:SeasonSpring:Campers	0.434755	
## Year500s:SeasonSpring:Campers	0.968568	
## Year510s:SeasonSpring:Campers	NA	
## Year580s:SeasonFall:Campers	0.588689	
## Year590s:SeasonFall:Campers	0.151881	
## Year500s:SeasonFall:Campers	0.181971	
## Year510s:SeasonFall:Campers	NA	
## Year580s:SeasonSummer:Campers	0.884417	
## Year590s:SeasonSummer:Campers	0.429128	
## Year500s:SeasonSummer:Campers	0.440540	
## Year510s:SeasonSummer:Campers	NA	
## Year580s:SeasonSpring:Lodging	0.101561	
## Year590s:SeasonSpring:Lodging	0.000270	***
## Year500s:SeasonSpring:Lodging	0.059907	.

```
## Year510s:SeasonSpring:Lodging          NA
## Year580s:SeasonFall:Lodging             0.395692
## Year590s:SeasonFall:Lodging             0.314602
## Year500s:SeasonFall:Lodging             0.480619
## Year510s:SeasonFall:Lodging             NA
## Year580s:SeasonSummer:Lodging           0.085273 .
## Year590s:SeasonSummer:Lodging           0.123686
## Year500s:SeasonSummer:Lodging           0.932482
## Year510s:SeasonSummer:Lodging           NA
## Year580s:Campers:Lodging                0.131627
## Year590s:Campers:Lodging                0.218606
## Year500s:Campers:Lodging                0.025763 *
## Year510s:Campers:Lodging                NA
## SeasonSpring:Campers:Lodging            0.260134
## SeasonFall:Campers:Lodging              0.364377
## SeasonSummer:Campers:Lodging            0.386724
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.31 on 352 degrees of freedom
## Multiple R-squared:  0.9881, Adjusted R-squared:  0.9842
## F-statistic: 253.8 on 115 and 352 DF,  p-value: < 2.2e-16
```

```
summary(st2)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Season + Lodging + Year5 + Park +
##      I(Lodging^3) + Campers + I(Campers^3) + I(Lodging^2) + Lodging:Year5 +
##      Season:Park + Season:Year5 + Lodging:Campers, data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -418.53  -49.02   -1.71    44.73   441.33
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.488e+02  6.751e+01   2.204  0.02804 *
## SeasonSpring   -5.532e+01  9.531e+01  -0.580  0.56196
## SeasonFall      2.522e+02  9.916e+01   2.543  0.01135 *
## SeasonSummer    7.033e+02  1.355e+02   5.191 3.23e-07 ***
## Lodging         4.821e+00  8.252e-01   5.842 1.02e-08 ***
## Year580s       -1.164e+01  6.882e+01  -0.169  0.86581
## Year590s        4.424e+01  6.926e+01   0.639  0.52333
## Year500s        8.169e+01  6.920e+01   1.181  0.23843
```



```
## Year510s          1.048e+02  6.996e+01   1.498  0.13487
## ParkYellowstone  -1.457e+02  2.991e+01  -4.872  1.56e-06 ***
## ParkGrand Canyon -1.128e+02  4.824e+01  -2.337  0.01987 *
## I(Lodging^3)      9.870e-06  2.321e-06   4.253  2.58e-05 ***
## Campers          2.826e+00  3.320e-01   8.511  2.90e-16 ***
## I(Campers^3)      5.239e-06  1.883e-06   2.782  0.00565 **
## I(Lodging^2)     -1.241e-02  2.907e-03  -4.270  2.41e-05 ***
## Lodging:Year580s  6.286e-01  3.002e-01   2.094  0.03686 *
## Lodging:Year590s  9.337e-01  3.120e-01   2.993  0.00292 **
## Lodging:Year500s  7.467e-01  3.215e-01   2.323  0.02067 *
## Lodging:Year510s  9.249e-01  3.260e-01   2.837  0.00476 **
## SeasonSpring:ParkYellowstone -4.490e+01  3.928e+01  -1.143  0.25362
## SeasonFall:ParkYellowstone -4.308e+02  5.842e+01  -7.374  8.60e-13 ***
## SeasonSummer:ParkYellowstone 2.571e+02  2.280e+02   1.128  0.26011
## SeasonSpring:ParkGrand Canyon 8.952e+01  5.530e+01   1.619  0.10619
## SeasonFall:ParkGrand Canyon -3.092e+02  5.067e+01  -6.104  2.32e-09 ***
## SeasonSummer:ParkGrand Canyon -3.985e+02  6.487e+01  -6.143  1.85e-09 ***
## SeasonSpring:Year580s 1.103e+02  9.755e+01   1.131  0.25880
## SeasonFall:Year580s  1.234e+02  9.877e+01   1.250  0.21207
## SeasonSummer:Year580s 2.270e+02  1.188e+02   1.911  0.05671 .
## SeasonSpring:Year590s 1.581e+02  9.745e+01   1.622  0.10548
## SeasonFall:Year590s  2.069e+02  9.878e+01   2.095  0.03675 *
## SeasonSummer:Year590s 4.956e+02  1.195e+02   4.148  4.05e-05 ***
## SeasonSpring:Year500s 1.713e+02  9.753e+01   1.757  0.07965 .
## SeasonFall:Year500s  1.657e+02  9.944e+01   1.666  0.09647 .
## SeasonSummer:Year500s 3.793e+02  1.197e+02   3.167  0.00165 **
## SeasonSpring:Year510s 2.240e+02  9.865e+01   2.271  0.02366 *
## SeasonFall:Year510s  2.787e+02  1.011e+02   2.757  0.00608 **
## SeasonSummer:Year510s 6.273e+02  1.218e+02   5.150  3.98e-07 ***
## Lodging:Campers   -8.584e-03  1.362e-03  -6.301  7.31e-10 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 113.1 on 430 degrees of freedom
```

```
## Multiple R-squared:  0.9767, Adjusted R-squared:  0.9746
```

```
## F-statistic: 486.1 on 37 and 430 DF,  p-value: < 2.2e-16
```

```
c(summary(fw2)$adj, summary(bw2)$adj, summary(st2)$adj)
```

```
## [1] 0.9747778 0.9841880 0.9746422
```

```
anova(fw2, bw2, st2)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Season + Lodging + Year5 + Park + I(Lodging^3) +
```

```
## Campers + I(Campers^3) + I(Lodging^2) + Season:Lodging +
## Lodging:Year5 + Season:Park + Season:Year5 + Lodging:Campers
## Model 2: TotalVisits ~ Park + Year5 + Season + Campers + Lodging + I(Lodging^2) +
## Park:Year5 + Park:Season + Park:Campers + Park:Lodging +
## Year5:Season + Year5:Campers + Year5:Lodging + Season:Campers +
## Season:Lodging + Campers:Lodging + Park:Year5:Season + Park:Year5:Campers +
## Park:Year5:Lodging + Park:Season:Lodging + Park:Campers:Lodging +
## Year5:Season:Campers + Year5:Season:Lodging + Year5:Campers:Lodging +
## Season:Campers:Lodging
## Model 3: TotalVisits ~ Season + Lodging + Year5 + Park + I(Lodging^3) +
## Campers + I(Campers^3) + I(Lodging^2) + Lodging:Year5 + Season:Park +
## Season:Year5 + Lodging:Campers
## Res.Df RSS Df Sum of Sq F Pr(>F)
## 1 427 5433049
## 2 352 2807778 75 2625271 4.3883 < 2.2e-16 ***
## 3 430 5500637 -78 -2692859 4.3281 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(fw2, bw2, st2, test="Cp")
```

```
## Analysis of Variance Table
##
## Model 1: TotalVisits ~ Season + Lodging + Year5 + Park + I(Lodging^3) +
## Campers + I(Campers^3) + I(Lodging^2) + Season:Lodging +
## Lodging:Year5 + Season:Park + Season:Year5 + Lodging:Campers
## Model 2: TotalVisits ~ Park + Year5 + Season + Campers + Lodging + I(Lodging^2) +
## Park:Year5 + Park:Season + Park:Campers + Park:Lodging +
## Year5:Season + Year5:Campers + Year5:Lodging + Season:Campers +
## Season:Lodging + Campers:Lodging + Park:Year5:Season + Park:Year5:Campers +
## Park:Year5:Lodging + Park:Season:Lodging + Park:Campers:Lodging +
## Year5:Season:Campers + Year5:Season:Lodging + Year5:Campers:Lodging +
## Season:Campers:Lodging
## Model 3: TotalVisits ~ Season + Lodging + Year5 + Park + I(Lodging^3) +
## Campers + I(Campers^3) + I(Lodging^2) + Lodging:Year5 + Season:Park +
## Season:Year5 + Lodging:Campers
## Res.Df RSS Df Sum of Sq Cp
## 1 427 5433049 6087134
## 2 352 2807778 75 2625271 4658359
## 3 430 5500637 -78 -2692859 6106862
```

```
AIC(fw2, bw2, st2)
```

```
## df AIC
## fw2 42 5792.392
## bw2 117 5633.462
```

```
## st2 39 5792.179
```

```
#Nested F test for bw1 and bw2  
anova(bw1, bw2)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Park + Year + Season + Campers + Lodging + I(Campers^2) +  
## I(Year^2) + I(Campers^3) + I(Lodging^3) + I(Year^3) + Park:Year +  
## Park:Season + Park:Campers + Park:Lodging + Year:Season +  
## Year:Campers + Year:Lodging + Season:Campers + Season:Lodging +  
## Campers:Lodging + Park:Year:Season + Park:Year:Campers +  
## Park:Year:Lodging + Park:Season:Lodging + Year:Season:Campers +  
## Year:Season:Lodging + Year:Campers:Lodging
```

```
## Model 2: TotalVisits ~ Park + Year5 + Season + Campers + Lodging + I(Lodging^2) +  
## Park:Year5 + Park:Season + Park:Campers + Park:Lodging +  
## Year5:Season + Year5:Campers + Year5:Lodging + Season:Campers +  
## Season:Lodging + Campers:Lodging + Park:Year5:Season + Park:Year5:Campers +  
## Park:Year5:Lodging + Park:Season:Lodging + Park:Campers:Lodging +  
## Year5:Season:Campers + Year5:Season:Lodging + Year5:Campers:Lodging +  
## Season:Campers:Lodging
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 412 3276508  
## 2 352 2807778 60 468730 0.9794 0.5233
```

To go back to main text: 2.1.1.

5.2.2 FIT

To go back to main text: 3.1.

```
#Model 1  
summary(bw1)
```

```
##
```

```
## Call:
```

```
## lm(formula = TotalVisits ~ Park + Year + Season + Campers + Lodging +  
## I(Campers^2) + I(Year^2) + I(Campers^3) + I(Lodging^3) +  
## I(Year^3) + Park:Year + Park:Season + Park:Campers + Park:Lodging +  
## Year:Season + Year:Campers + Year:Lodging + Season:Campers +  
## Season:Lodging + Campers:Lodging + Park:Year:Season + Park:Year:Campers +  
## Park:Year:Lodging + Park:Season:Lodging + Year:Season:Campers +  
## Year:Season:Lodging + Year:Campers:Lodging, data = NationalParkVisits)
```

```
##
```

```
## Residuals:
```

```
## Min 1Q Median 3Q Max
```

```
## -271.59  -47.86   -2.24   50.05  367.53
##
## Coefficients: (5 not defined because of singularities)
##
##              Estimate Std. Error t value
## (Intercept)    -3.058e+08  3.683e+07  -8.303
## ParkYellowstone    5.882e+03  5.613e+03   1.048
## ParkGrand Canyon -1.263e+04  1.279e+04  -0.988
## Year             4.593e+05  5.529e+04   8.306
## SeasonSpring     -8.311e+02  3.852e+03  -0.216
## SeasonFall       1.299e+04  5.870e+03   2.213
## SeasonSummer    -1.843e+03  1.470e+04  -0.125
## Campers          -2.972e+02  4.914e+02  -0.605
## Lodging          -1.484e+02  2.059e+02  -0.721
## I(Campers^2)     -2.692e-02  8.140e-03  -3.307
## I(Year^2)        -2.299e+02  2.767e+01  -8.309
## I(Campers^3)      3.658e-05  8.452e-06   4.328
## I(Lodging^3)      4.503e-06  1.842e-06   2.445
## I(Year^3)        3.835e-02  4.615e-03   8.312
## ParkYellowstone:Year -2.942e+00  2.818e+00  -1.044
## ParkGrand Canyon:Year 6.365e+00  6.428e+00   0.990
## ParkYellowstone:SeasonSpring -3.188e+04  9.239e+03  -3.451
## ParkGrand Canyon:SeasonSpring -1.586e+05  2.395e+04  -6.622
## ParkYellowstone:SeasonFall -4.694e+04  2.028e+04  -2.315
## ParkGrand Canyon:SeasonFall -8.441e+04  2.140e+04  -3.945
## ParkYellowstone:SeasonSummer -1.710e+05  1.096e+05  -1.561
## ParkGrand Canyon:SeasonSummer -2.390e+05  3.744e+04  -6.383
## ParkYellowstone:Campers 7.933e+02  2.536e+02   3.128
## ParkGrand Canyon:Campers 9.191e+02  1.448e+02   6.349
## ParkYellowstone:Lodging -4.607e+02  1.062e+02  -4.339
## ParkGrand Canyon:Lodging      NA          NA      NA
## Year:SeasonSpring 4.323e-01  1.929e+00   0.224
## Year:SeasonFall -6.449e+00  2.941e+00  -2.193
## Year:SeasonSummer 1.057e+00  7.351e+00   0.144
## Year:Campers      1.508e-01  2.453e-01   0.615
## Year:Lodging      7.574e-02  1.032e-01   0.734
## SeasonSpring:Campers -3.056e+02  4.807e+02  -0.636
## SeasonFall:Campers -4.800e+02  4.779e+02  -1.004
## SeasonSummer:Campers 9.659e+01  4.825e+02   0.200
## SeasonSpring:Lodging 9.031e+02  2.241e+02   4.030
## SeasonFall:Lodging 7.358e+02  2.221e+02   3.313
## SeasonSummer:Lodging 8.636e+02  2.418e+02   3.572
## Campers:Lodging -1.396e+00  4.491e-01  -3.108
## ParkYellowstone:Year:SeasonSpring 1.594e+01  4.640e+00   3.435
## ParkGrand Canyon:Year:SeasonSpring 7.994e+01  1.206e+01   6.629
## ParkYellowstone:Year:SeasonFall 2.349e+01  1.019e+01   2.306
```

## ParkGrand Canyon:Year:SeasonFall	4.227e+01	1.075e+01	3.931
## ParkYellowstone:Year:SeasonSummer	8.795e+01	5.531e+01	1.590
## ParkGrand Canyon:Year:SeasonSummer	1.202e+02	1.879e+01	6.396
## ParkYellowstone:Year:Campers	-4.012e-01	1.274e-01	-3.149
## ParkGrand Canyon:Year:Campers	-4.628e-01	7.269e-02	-6.367
## ParkYellowstone:Year:Lodging	2.285e-01	5.332e-02	4.286
## ParkGrand Canyon:Year:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSpring:Lodging	1.124e+01	3.391e+00	3.316
## ParkGrand Canyon:SeasonSpring:Lodging	NA	NA	NA
## ParkYellowstone:SeasonFall:Lodging	5.783e+00	3.116e+00	1.856
## ParkGrand Canyon:SeasonFall:Lodging	NA	NA	NA
## ParkYellowstone:SeasonSummer:Lodging	-6.660e-01	3.883e+00	-0.172
## ParkGrand Canyon:SeasonSummer:Lodging	NA	NA	NA
## Year:SeasonSpring:Campers	1.558e-01	2.400e-01	0.649
## Year:SeasonFall:Campers	2.447e-01	2.385e-01	1.026
## Year:SeasonSummer:Campers	-4.450e-02	2.408e-01	-0.185
## Year:SeasonSpring:Lodging	-4.550e-01	1.124e-01	-4.049
## Year:SeasonFall:Lodging	-3.698e-01	1.114e-01	-3.320
## Year:SeasonSummer:Lodging	-4.347e-01	1.212e-01	-3.586
## Year:Campers:Lodging	7.001e-04	2.255e-04	3.105
##	Pr(> t)		
## (Intercept)	1.47e-15	***	
## ParkYellowstone	0.295308		
## ParkGrand Canyon	0.323704		
## Year	1.44e-15	***	
## SeasonSpring	0.829264		
## SeasonFall	0.027426	*	
## SeasonSummer	0.900285		
## Campers	0.545636		
## Lodging	0.471291		
## I(Campers^2)	0.001025	**	
## I(Year^2)	1.41e-15	***	
## I(Campers^3)	1.89e-05	***	
## I(Lodging^3)	0.014892	*	
## I(Year^3)	1.38e-15	***	
## ParkYellowstone:Year	0.297066		
## ParkGrand Canyon:Year	0.322676		
## ParkYellowstone:SeasonSpring	0.000616	***	
## ParkGrand Canyon:SeasonSpring	1.11e-10	***	
## ParkYellowstone:SeasonFall	0.021116	*	
## ParkGrand Canyon:SeasonFall	9.39e-05	***	
## ParkYellowstone:SeasonSummer	0.119349		
## ParkGrand Canyon:SeasonSummer	4.70e-10	***	
## ParkYellowstone:Campers	0.001882	**	
## ParkGrand Canyon:Campers	5.73e-10	***	

```

## ParkYellowstone:Lodging          1.81e-05 ***
## ParkGrand Canyon:Lodging          NA
## Year:SeasonSpring                0.822797
## Year:SeasonFall                  0.028894 *
## Year:SeasonSummer                0.885721
## Year:Campers                     0.539058
## Year:Lodging                     0.463374
## SeasonSpring:Campers             0.525252
## SeasonFall:Campers               0.315731
## SeasonSummer:Campers             0.841422
## SeasonSpring:Lodging             6.64e-05 ***
## SeasonFall:Lodging               0.001003 **
## SeasonSummer:Lodging             0.000396 ***
## Campers:Lodging                  0.002015 **
## ParkYellowstone:Year:SeasonSpring 0.000653 ***
## ParkGrand Canyon:Year:SeasonSpring 1.06e-10 ***
## ParkYellowstone:Year:SeasonFall   0.021617 *
## ParkGrand Canyon:Year:SeasonFall  9.93e-05 ***
## ParkYellowstone:Year:SeasonSummer 0.112547
## ParkGrand Canyon:Year:SeasonSummer 4.34e-10 ***
## ParkYellowstone:Year:Campers      0.001756 **
## ParkGrand Canyon:Year:Campers     5.14e-10 ***
## ParkYellowstone:Year:Lodging      2.27e-05 ***
## ParkGrand Canyon:Year:Lodging      NA
## ParkYellowstone:SeasonSpring:Lodging 0.000995 ***
## ParkGrand Canyon:SeasonSpring:Lodging NA
## ParkYellowstone:SeasonFall:Lodging 0.064214 .
## ParkGrand Canyon:SeasonFall:Lodging NA
## ParkYellowstone:SeasonSummer:Lodging 0.863900
## ParkGrand Canyon:SeasonSummer:Lodging NA
## Year:SeasonSpring:Campers         0.516516
## Year:SeasonFall:Campers           0.305511
## Year:SeasonSummer:Campers         0.853505
## Year:SeasonSpring:Lodging         6.15e-05 ***
## Year:SeasonFall:Lodging           0.000979 ***
## Year:SeasonSummer:Lodging         0.000376 ***
## Year:Campers:Lodging              0.002037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.18 on 412 degrees of freedom
## Multiple R-squared:  0.9861, Adjusted R-squared:  0.9842
## F-statistic: 531.1 on 55 and 412 DF,  p-value: < 2.2e-16

```

To go back to main text: 3.1.

5.2.3 ASSESS

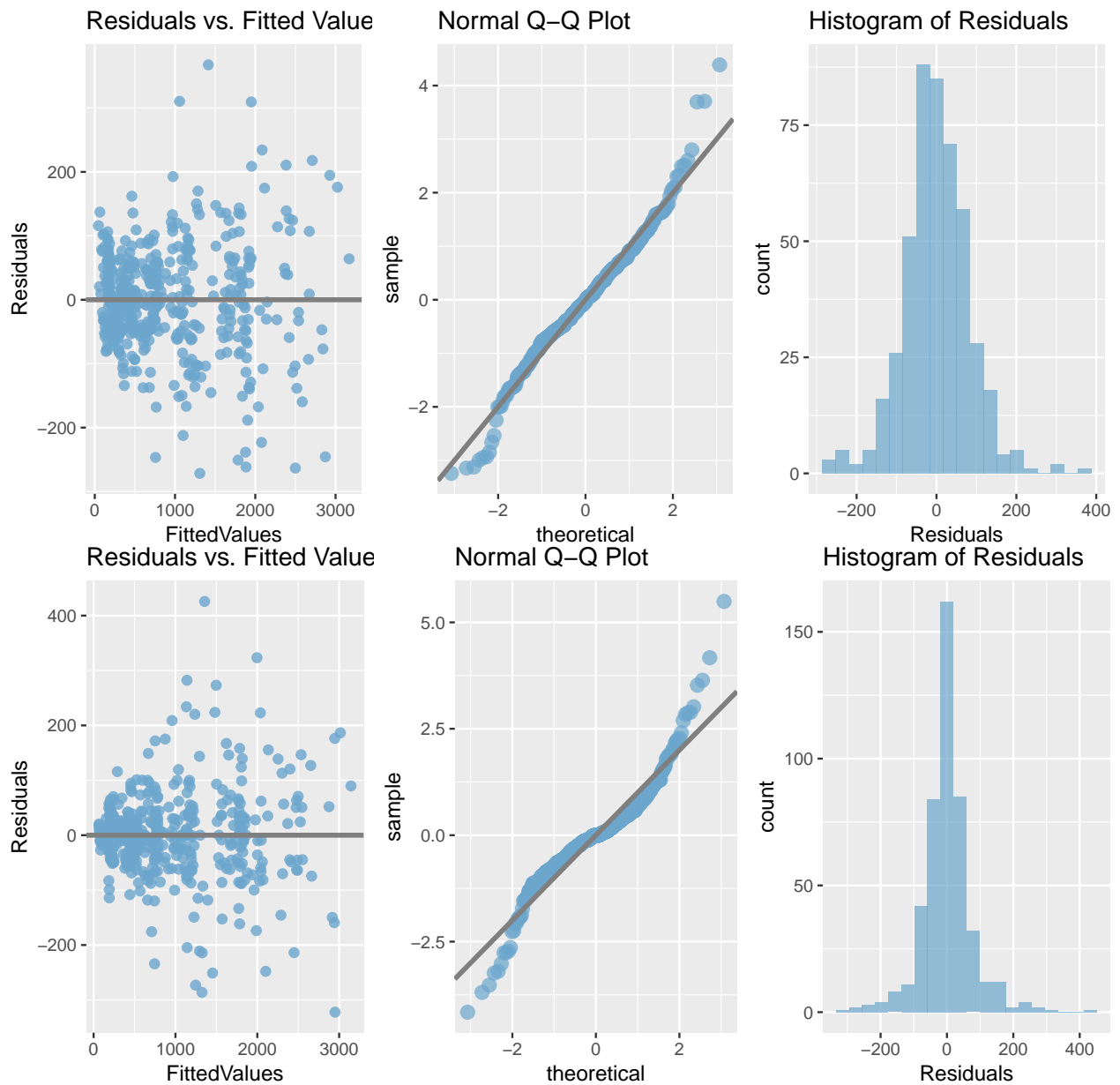
To go back to main text: 2.1.2.

```
CheckAssumptions(bw1)
```

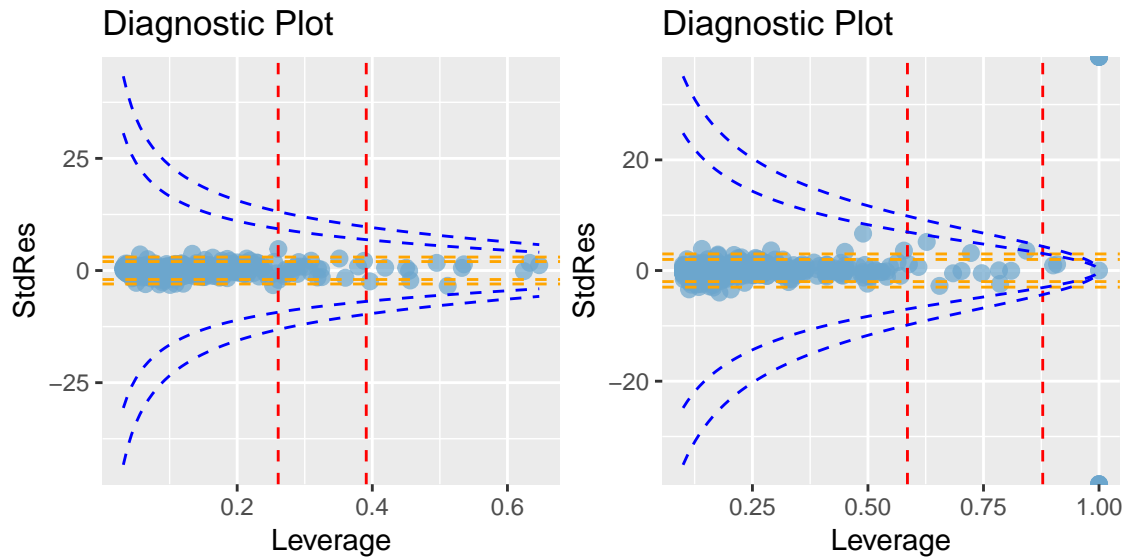
```
##  
## studentized Breusch-Pagan test  
##  
## data: model  
## BP = 130.75, df = 55, p-value = 4.125e-08
```

```
CheckAssumptions(bw2)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: model  
## BP = 235.24, df = 115, p-value = 2.814e-10
```



```
UnusualPoints(bw1, NationalParkVisits)
UnusualPoints(bw2, NationalParkVisits)
```

To go back to main text: 2.1.2.

5.2.4 USE

To go back to main text: 3.3.1.

```
predict(bw1, list(Lodging=858.573, Season="Summer", Year=2017,
                  Park="Yellowstone", Campers=146.306),
        interval = "confidence")
```

```
##          fit      lwr      upr
## 1 3169.767 3061.846 3277.688
```

```
predict(bw1, list(Lodging=858.573, Season="Summer", Year=2017,
                  Park="Yellowstone", Campers=146.306),
        interval = "prediction")
```

```
##          fit      lwr      upr
## 1 3169.767 2963.91 3375.624
```

5.3 Model 2

5.3.1 CHOOSE

To go back to main text: 2.2.1.

```
Seasonnum <- as.numeric(NationalParkVisits$Season)-1
Year5num <- as.numeric(Year5)-1
NationalParkVisits <- data.frame(NationalParkVisits, Seasonnum, Year5num)
```

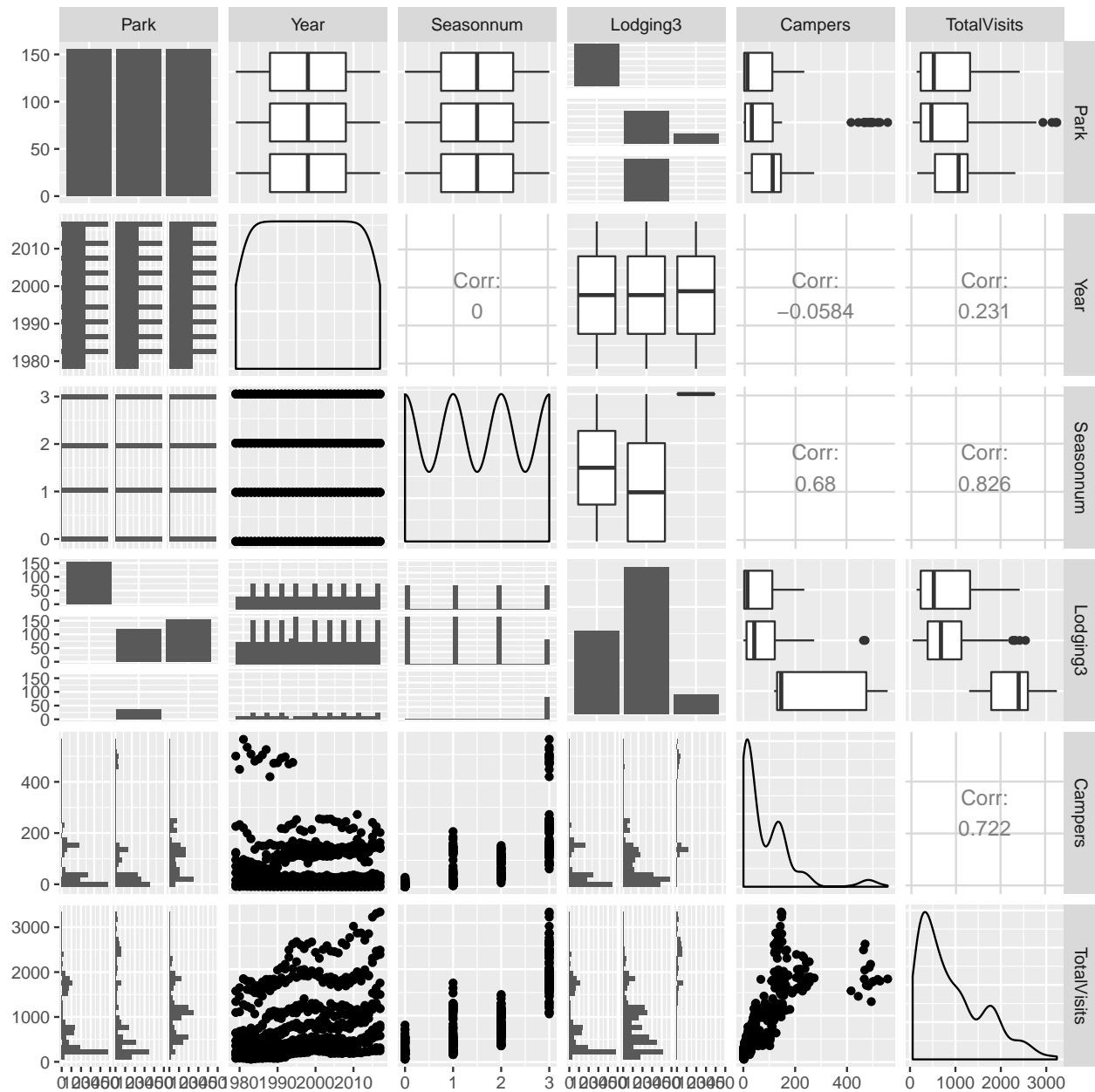
```
#Summary stats
```

```
describe(NationalParkVisits)[,2:4]
```

##	n	mean	sd
## Park*	468	2.00	0.82
## State*	468	2.00	0.82
## Year	468	1998.00	11.27
## Season*	468	2.50	1.12
## Lodging	468	128.16	192.88
## Campers	468	81.75	102.14
## TotalVisits	468	909.66	710.26
## Year5*	468	3.36	1.14
## logCampers	468	3.36	1.79
## logLodging	468	3.18	2.40
## Lodging3*	468	1.75	0.59
## Campers2*	468	1.03	0.18
## Seasonnum	468	1.50	1.12
## Year5num	468	2.36	1.14

```
#Summary matrix with Year
```

```
ggpairs(NationalParkVisits[,c("Park", "Year", "Seasonnum", "Lodging3",  
                              "Campers", "TotalVisits")])
```



```
#VIF
alias(lm(TotalVisits ~ Park + Year + Seasonnum + Lodging3 + Campers,
data=NationalParkVisits)) #check for aliased variables
```

```
## Model :
## TotalVisits ~ Park + Year + Seasonnum + Lodging3 + Campers
##
## Complete :
##              (Intercept) ParkYellowstone ParkGrand Canyon Year
## Lodging3over500 0          1          1          0
##              Seasonnum Lodging3under500 Campers
## Lodging3over500 0          -1          0
```

```
vif(lm(TotalVisits ~ Park + Year + Seasonnum + Campers,
      data=NationalParkVisits)) #Lodging3 taken out
```

```
##              GVIF Df GVIF^(1/(2*Df))
## Park          1.092535  2          1.022372
## Year           1.006978  1          1.003483
## Seasonnum      1.944711  1          1.394529
## Campers        2.044224  1          1.429764
```

```
# Zero model and full models
```

```
zeroModel <- lm(TotalVisits ~ 1, data=NationalParkVisits)
fullModel3 <- lm(TotalVisits ~ (Park + Year5num + Seasonnum + Campers
                               + Lodging3)^2 + I(Campers^2),
                data=NationalParkVisits)
```

```
fullModel4 <- lm(TotalVisits ~ (Park + Year + Seasonnum + Campers + Lodging3)^2
                + I(Year^2) + I(Campers^2), data=NationalParkVisits)
```

```
#With FullModel3 (Year5num, Seasonnum, Lodging3)
```

```
fw3 <- step(zeroModel, scope=list(upper=fullModel3), direction="forward", trace=0)
bw3 <- step(fullModel3, scope=list(lower=zeroModel), direction="backward", trace=0)
st3 <- step(zeroModel, scope=list(upper=fullModel3), direction="both", trace=0)
summary(fw3)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Seasonnum + Lodging3 + Year5num +
##     Park + Campers + Lodging3:Year5num + Lodging3:Campers + Seasonnum:Year5num +
##     Year5num:Park + Year5num:Campers + Seasonnum:Campers, data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -460.08  -85.84   -2.58   87.21  686.66
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      19.6446     39.0839   0.503  0.615471
## Seasonnum       129.4272     23.8877   5.418 9.80e-08 ***
## Lodging3under500  403.3817     44.8356   8.997 < 2e-16 ***
## Lodging3over500  2139.5903    174.6135  12.253 < 2e-16 ***
## Year5num         54.8850     14.7670   3.717 0.000227 ***
## ParkYellowstone -517.5675     46.3577 -11.165 < 2e-16 ***
## ParkGrand Canyon      NA         NA      NA      NA
## Campers           5.1966     0.5617   9.252 < 2e-16 ***
## Lodging3under500:Year5num -63.3874     17.9743  -3.527 0.000464 ***
```

```
## Lodging3over500:Year5num    -28.7026    50.4747   -0.569 0.569873
## Lodging3under500:Campers    -2.9304     0.2498  -11.731 < 2e-16 ***
## Lodging3over500:Campers     -7.4932     0.3740  -20.033 < 2e-16 ***
## Seasonnum:Year5num          19.7169     9.2936    2.122 0.034417 *
## Year5num:ParkYellowstone    117.6245    19.9679    5.891 7.51e-09 ***
## Year5num:ParkGrand Canyon      NA         NA         NA         NA
## Year5num:Campers             0.3721     0.1519    2.449 0.014691 *
## Seasonnum:Campers           0.2727     0.1295    2.106 0.035728 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 157.6 on 453 degrees of freedom
## Multiple R-squared:  0.9523, Adjusted R-squared:  0.9508
## F-statistic: 645.3 on 14 and 453 DF,  p-value: < 2.2e-16
```

```
summary(bw3)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Park + Year5num + Seasonnum + Campers +
##     Lodging3 + Park:Year5num + Year5num:Seasonnum + Year5num:Campers +
##     Seasonnum:Campers + Campers:Lodging3, data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -461.82  -83.92   -3.47    87.45   686.76
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      23.5680    38.7382   0.608 0.543233
## ParkYellowstone  1731.9711    82.7641  20.927 < 2e-16 ***
## ParkGrand Canyon  2260.7494    78.0130  28.979 < 2e-16 ***
## Year5num           53.2668    14.6125   3.645 0.000298 ***
## Seasonnum        129.7083    23.8745   5.433 9.06e-08 ***
## Campers           5.1238     0.5535   9.256 < 2e-16 ***
## Lodging3under500 -1854.3293    65.6885 -28.229 < 2e-16 ***
## Lodging3over500      NA         NA         NA         NA
## ParkYellowstone:Year5num    57.9753    16.5375   3.506 0.000501 ***
## ParkGrand Canyon:Year5num  -65.3395    17.7894  -3.673 0.000268 ***
## Year5num:Seasonnum    19.3306     9.2761   2.084 0.037728 *
## Year5num:Campers       0.4132     0.1423   2.903 0.003880 **
## Seasonnum:Campers      0.2714     0.1294   2.098 0.036497 *
## Campers:Lodging3under500  -2.9419     0.2493 -11.803 < 2e-16 ***
## Campers:Lodging3over500   -7.6179     0.3376 -22.562 < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 157.5 on 454 degrees of freedom
## Multiple R-squared:  0.9522, Adjusted R-squared:  0.9508
## F-statistic: 695.5 on 13 and 454 DF,  p-value: < 2.2e-16

summary(st3)

##
## Call:
## lm(formula = TotalVisits ~ Seasonnum + Lodging3 + Year5num +
##      Park + Campers + Lodging3:Campers + Seasonnum:Year5num +
##      Year5num:Park + Year5num:Campers + Seasonnum:Campers, data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -461.82  -83.92   -3.47   87.45  686.76
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      23.5680     38.7382   0.608 0.543233
## Seasonnum        129.7083     23.8745   5.433 9.06e-08 ***
## Lodging3under500   406.4201     44.6446   9.103 < 2e-16 ***
## Lodging3over500   2260.7494     78.0130  28.979 < 2e-16 ***
## Year5num          53.2668     14.6125   3.645 0.000298 ***
## ParkYellowstone  -528.7783     44.0277 -12.010 < 2e-16 ***
## ParkGrand Canyon      NA         NA      NA      NA
## Campers           5.1238      0.5535   9.256 < 2e-16 ***
## Lodging3under500:Campers -2.9419     0.2493 -11.803 < 2e-16 ***
## Lodging3over500:Campers -7.6179     0.3376 -22.562 < 2e-16 ***
## Seasonnum:Year5num    19.3306      9.2761   2.084 0.037728 *
## Year5num:ParkYellowstone 57.9753     16.5375   3.506 0.000501 ***
## Year5num:ParkGrand Canyon -65.3395     17.7894  -3.673 0.000268 ***
## Year5num:Campers      0.4132      0.1423   2.903 0.003880 **
## Seasonnum:Campers      0.2714      0.1294   2.098 0.036497 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 157.5 on 454 degrees of freedom
## Multiple R-squared:  0.9522, Adjusted R-squared:  0.9508
## F-statistic: 695.5 on 13 and 454 DF,  p-value: < 2.2e-16

c(summary(fw3)$adj, summary(bw3)$adj, summary(st3)$adj)

## [1] 0.9507749 0.9508180 0.9508180
```

```
anova(fw3, bw3, st3)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Seasonnum + Lodging3 + Year5num + Park + Campers +
```

```
##   Lodging3:Year5num + Lodging3:Campers + Seasonnum:Year5num +
```

```
##   Year5num:Park + Year5num:Campers + Seasonnum:Campers
```

```
## Model 2: TotalVisits ~ Park + Year5num + Seasonnum + Campers + Lodging3 +
```

```
##   Park:Year5num + Year5num:Seasonnum + Year5num:Campers + Seasonnum:Campers +
```

```
##   Campers:Lodging3
```

```
## Model 3: TotalVisits ~ Seasonnum + Lodging3 + Year5num + Park + Campers +
```

```
##   Lodging3:Campers + Seasonnum:Year5num + Year5num:Park + Year5num:Campers +
```

```
##   Seasonnum:Campers
```

```
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      453 11249121
```

```
## 2      454 11264062 -1      -14941 0.6017 0.4384
```

```
## 3      454 11264062  0           0
```

```
anova(fw3, bw3, st3, test="Cp")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Seasonnum + Lodging3 + Year5num + Park + Campers +
```

```
##   Lodging3:Year5num + Lodging3:Campers + Seasonnum:Year5num +
```

```
##   Year5num:Park + Year5num:Campers + Seasonnum:Campers
```

```
## Model 2: TotalVisits ~ Park + Year5num + Seasonnum + Campers + Lodging3 +
```

```
##   Park:Year5num + Year5num:Seasonnum + Year5num:Campers + Seasonnum:Campers +
```

```
##   Campers:Lodging3
```

```
## Model 3: TotalVisits ~ Seasonnum + Lodging3 + Year5num + Park + Campers +
```

```
##   Lodging3:Campers + Seasonnum:Year5num + Year5num:Park + Year5num:Campers +
```

```
##   Seasonnum:Campers
```

```
##   Res.Df      RSS Df Sum of Sq      Cp
```

```
## 1      453 11249121      11994096
```

```
## 2      454 11264062 -1      -14941 11959372
```

```
## 3      454 11264062  0           0 11959372
```

```
AIC(fw3, bw3, st3)
```

```
##      df      AIC
```

```
## fw3 16 6080.998
```

```
## bw3 15 6079.619
```

```
## st3 15 6079.619
```

```
#With FullModel4 (Seasonnum, Lodging3)
```

```
fw4 <- step(zeroModel, scope=list(upper=fullModel4), direction="forward", trace=0)
```

```
bw4 <- step(fullModel4, scope=list(lower=zeroModel), direction="backward", trace=0)
```

```
st4 <- step(zeroModel, scope=list(upper=fullModel4), direction="both", trace=0)
summary(fw4)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Seasonnum + Lodging3 + Year + Park +
##     Campers + I(Year^2) + Lodging3:Year + Lodging3:Campers +
##     Seasonnum:Year + Year:Park + Seasonnum:Campers + Year:Campers,
##     data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -443.26  -84.68    0.25   75.22  669.53
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.402e+05  2.491e+05  -2.570  0.01049 *
## Seasonnum     -4.752e+03  1.845e+03  -2.576  0.01030 *
## Lodging3under500  1.014e+04  3.495e+03   2.902  0.00388 **
## Lodging3over500 -1.119e+04  1.048e+04  -1.068  0.28619
## Year          6.356e+02  2.494e+02   2.549  0.01114 *
## ParkYellowstone -2.198e+04  3.894e+03  -5.644  2.94e-08 ***
## ParkGrand Canyon      NA         NA      NA      NA
## Campers        -5.719e+01  3.062e+01  -1.868  0.06245 .
## I(Year^2)      -1.577e-01  6.241e-02  -2.527  0.01184 *
## Lodging3under500:Year -4.944e+00  1.752e+00  -2.822  0.00499 **
## Lodging3over500:Year  6.560e+00  5.222e+00   1.256  0.20968
## Lodging3under500:Campers -2.983e+00  2.382e-01 -12.524 < 2e-16 ***
## Lodging3over500:Campers -6.915e+00  3.844e-01 -17.989 < 2e-16 ***
## Seasonnum:Year    2.468e+00  9.240e-01   2.671  0.00784 **
## Year:ParkYellowstone  1.088e+01  1.953e+00   5.570  4.37e-08 ***
## Year:ParkGrand Canyon      NA         NA      NA      NA
## Seasonnum:Campers  2.890e-01  1.233e-01   2.343  0.01954 *
## Year:Campers      3.163e-02  1.533e-02   2.063  0.03973 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 149.8 on 452 degrees of freedom
## Multiple R-squared:  0.9569, Adjusted R-squared:  0.9555
## F-statistic: 669.4 on 15 and 452 DF,  p-value: < 2.2e-16
```

```
summary(bw4)
```

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



```
## lm(formula = TotalVisits ~ Park + Year + Seasonnum + Campers +
##   Lodging3 + I(Year^2) + I(Campers^2) + Park:Year + Park:Seasonnum +
##   Year:Seasonnum + Year:Campers + Year:Lodging3 + Campers:Lodging3,
##   data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -453.67  -81.44    1.87   77.87  627.95
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -7.176e+05  2.504e+05  -2.866  0.004355 **
## ParkYellowstone -3.267e+04  9.215e+03  -3.546  0.000433 ***
## ParkGrand Canyon -1.261e+04  1.052e+04  -1.199  0.231334
## Year           7.132e+02  2.506e+02   2.846  0.004636 **
## Seasonnum     -5.358e+03  1.865e+03  -2.872  0.004266 **
## Campers       -5.607e+01  3.174e+01  -1.766  0.078028 .
## Lodging3under500  2.179e+04  9.471e+03   2.300  0.021891 *
## Lodging3over500           NA         NA         NA         NA
## I(Year^2)      -1.772e-01  6.272e-02  -2.825  0.004946 **
## I(Campers^2)    2.364e-03  1.074e-03   2.201  0.028213 *
## ParkYellowstone:Year  1.723e+01  4.584e+00   3.759  0.000193 ***
## ParkGrand Canyon:Year  7.305e+00  5.242e+00   1.393  0.164169
## ParkYellowstone:Seasonnum  3.480e+01  2.925e+01   1.190  0.234723
## ParkGrand Canyon:Seasonnum  6.053e+01  2.678e+01   2.260  0.024287 *
## Year:Seasonnum    2.763e+00  9.333e-01   2.960  0.003240 **
## Year:Campers      3.143e-02  1.590e-02   1.977  0.048629 *
## Year:Lodging3under500 -1.178e+01  4.713e+00  -2.499  0.012810 *
## Year:Lodging3over500           NA         NA         NA         NA
## Campers:Lodging3under500 -4.005e+00  4.717e-01  -8.490  3.03e-16 ***
## Campers:Lodging3over500 -8.232e+00  6.188e-01 -13.303  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 149.6 on 450 degrees of freedom
## Multiple R-squared:  0.9572, Adjusted R-squared:  0.9556
## F-statistic: 592.7 on 17 and 450 DF,  p-value: < 2.2e-16
```

```
summary(st4)
```

```
##
## Call:
## lm(formula = TotalVisits ~ Seasonnum + Lodging3 + Year + Park +
##   Campers + I(Year^2) + Lodging3:Year + Lodging3:Campers +
##   Seasonnum:Year + Year:Park + Seasonnum:Campers + Year:Campers,
```

```

##      data = NationalParkVisits)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -443.26   -84.68     0.25    75.22   669.53
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.402e+05  2.491e+05  -2.570  0.01049 *
## Seasonnum     -4.752e+03  1.845e+03  -2.576  0.01030 *
## Lodging3under500  1.014e+04  3.495e+03   2.902  0.00388 **
## Lodging3over500 -1.119e+04  1.048e+04  -1.068  0.28619
## Year          6.356e+02  2.494e+02   2.549  0.01114 *
## ParkYellowstone -2.198e+04  3.894e+03  -5.644  2.94e-08 ***
## ParkGrand Canyon      NA         NA      NA      NA
## Campers        -5.719e+01  3.062e+01  -1.868  0.06245 .
## I(Year^2)      -1.577e-01  6.241e-02  -2.527  0.01184 *
## Lodging3under500:Year -4.944e+00  1.752e+00  -2.822  0.00499 **
## Lodging3over500:Year  6.560e+00  5.222e+00   1.256  0.20968
## Lodging3under500:Campers -2.983e+00  2.382e-01 -12.524 < 2e-16 ***
## Lodging3over500:Campers -6.915e+00  3.844e-01 -17.989 < 2e-16 ***
## Seasonnum:Year    2.468e+00  9.240e-01   2.671  0.00784 **
## Year:ParkYellowstone  1.088e+01  1.953e+00   5.570  4.37e-08 ***
## Year:ParkGrand Canyon      NA         NA      NA      NA
## Seasonnum:Campers  2.890e-01  1.233e-01   2.343  0.01954 *
## Year:Campers      3.163e-02  1.533e-02   2.063  0.03973 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 149.8 on 452 degrees of freedom
## Multiple R-squared:  0.9569, Adjusted R-squared:  0.9555
## F-statistic: 669.4 on 15 and 452 DF,  p-value: < 2.2e-16
c(summary(fw4)$adj, summary(bw4)$adj, summary(st4)$adj)

## [1] 0.9554951 0.9556328 0.9554951
anova(fw4, bw4, st4)

## Analysis of Variance Table
##
## Model 1: TotalVisits ~ Seasonnum + Lodging3 + Year + Park + Campers +
##      I(Year^2) + Lodging3:Year + Lodging3:Campers + Seasonnum:Year +
##      Year:Park + Seasonnum:Campers + Year:Campers
## Model 2: TotalVisits ~ Park + Year + Seasonnum + Campers + Lodging3 +
##      I(Year^2) + I(Campers^2) + Park:Year + Park:Seasonnum + Year:Seasonnum +

```

```
##      Year:Campers + Year:Lodging3 + Campers:Lodging3
## Model 3: TotalVisits ~ Seasonnum + Lodging3 + Year + Park + Campers +
##      I(Year^2) + Lodging3:Year + Lodging3:Campers + Seasonnum:Year +
##      Year:Park + Seasonnum:Campers + Year:Campers
##  Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1      452 10147977
## 2      450 10071820  2      76157 1.7013 0.1836
## 3      452 10147977 -2      -76157 1.7013 0.1836
```

```
anova(fw4, bw4, st4, test="Cp")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Seasonnum + Lodging3 + Year + Park + Campers +
##      I(Year^2) + Lodging3:Year + Lodging3:Campers + Seasonnum:Year +
##      Year:Park + Seasonnum:Campers + Year:Campers
```

```
## Model 2: TotalVisits ~ Park + Year + Seasonnum + Campers + Lodging3 +
##      I(Year^2) + I(Campers^2) + Park:Year + Park:Seasonnum + Year:Seasonnum +
##      Year:Campers + Year:Lodging3 + Campers:Lodging3
```

```
## Model 3: TotalVisits ~ Seasonnum + Lodging3 + Year + Park + Campers +
##      I(Year^2) + Lodging3:Year + Lodging3:Campers + Seasonnum:Year +
##      Year:Park + Seasonnum:Campers + Year:Campers
```

```
##  Res.Df      RSS Df Sum of Sq      Cp
## 1      452 10147977      10864196
## 2      450 10071820  2      76157 10877565
## 3      452 10147977 -2      -76157 10864196
```

```
AIC(fw4, bw4, st4)
```

```
##      df      AIC
## fw4 17 6034.787
## bw4 19 6035.261
## st4 17 6034.787
```

```
#Nested F test for st1 and st2
```

```
anova(st3, st4)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: TotalVisits ~ Seasonnum + Lodging3 + Year5num + Park + Campers +
##      Lodging3:Campers + Seasonnum:Year5num + Year5num:Park + Year5num:Campers +
##      Seasonnum:Campers
```

```
## Model 2: TotalVisits ~ Seasonnum + Lodging3 + Year + Park + Campers +
##      I(Year^2) + Lodging3:Year + Lodging3:Campers + Seasonnum:Year +
##      Year:Park + Seasonnum:Campers + Year:Campers
```

```
##  Res.Df      RSS Df Sum of Sq      F      Pr(>F)
## 1      454 11264062
```

```
## 2      452 10147977  2      1116085 24.856 5.737e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To go back to main text: 2.2.1.

5.3.2 FIT

To go back to main text: 3.2.

```
#Model 2
summary(st4)

##
## Call:
## lm(formula = TotalVisits ~ Seasonnum + Lodging3 + Year + Park +
##      Campers + I(Year^2) + Lodging3:Year + Lodging3:Campers +
##      Seasonnum:Year + Year:Park + Seasonnum:Campers + Year:Campers,
##      data = NationalParkVisits)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -443.26  -84.68    0.25   75.22  669.53
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -6.402e+05  2.491e+05  -2.570  0.01049 *
## Seasonnum     -4.752e+03  1.845e+03  -2.576  0.01030 *
## Lodging3under500  1.014e+04  3.495e+03   2.902  0.00388 **
## Lodging3over500 -1.119e+04  1.048e+04  -1.068  0.28619
## Year          6.356e+02  2.494e+02   2.549  0.01114 *
## ParkYellowstone -2.198e+04  3.894e+03  -5.644 2.94e-08 ***
## ParkGrand Canyon      NA         NA      NA      NA
## Campers        -5.719e+01  3.062e+01  -1.868  0.06245 .
## I(Year^2)      -1.577e-01  6.241e-02  -2.527  0.01184 *
## Lodging3under500:Year -4.944e+00  1.752e+00  -2.822  0.00499 **
## Lodging3over500:Year  6.560e+00  5.222e+00   1.256  0.20968
## Lodging3under500:Campers -2.983e+00  2.382e-01 -12.524 < 2e-16 ***
## Lodging3over500:Campers -6.915e+00  3.844e-01 -17.989 < 2e-16 ***
## Seasonnum:Year    2.468e+00  9.240e-01   2.671  0.00784 **
## Year:ParkYellowstone  1.088e+01  1.953e+00   5.570 4.37e-08 ***
## Year:ParkGrand Canyon      NA         NA      NA      NA
## Seasonnum:Campers  2.890e-01  1.233e-01   2.343  0.01954 *
## Year:Campers      3.163e-02  1.533e-02   2.063  0.03973 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 149.8 on 452 degrees of freedom
## Multiple R-squared:  0.9569, Adjusted R-squared:  0.9555
## F-statistic: 669.4 on 15 and 452 DF,  p-value: < 2.2e-16
```

To go back to main text: 3.2.

5.3.3 ASSESS

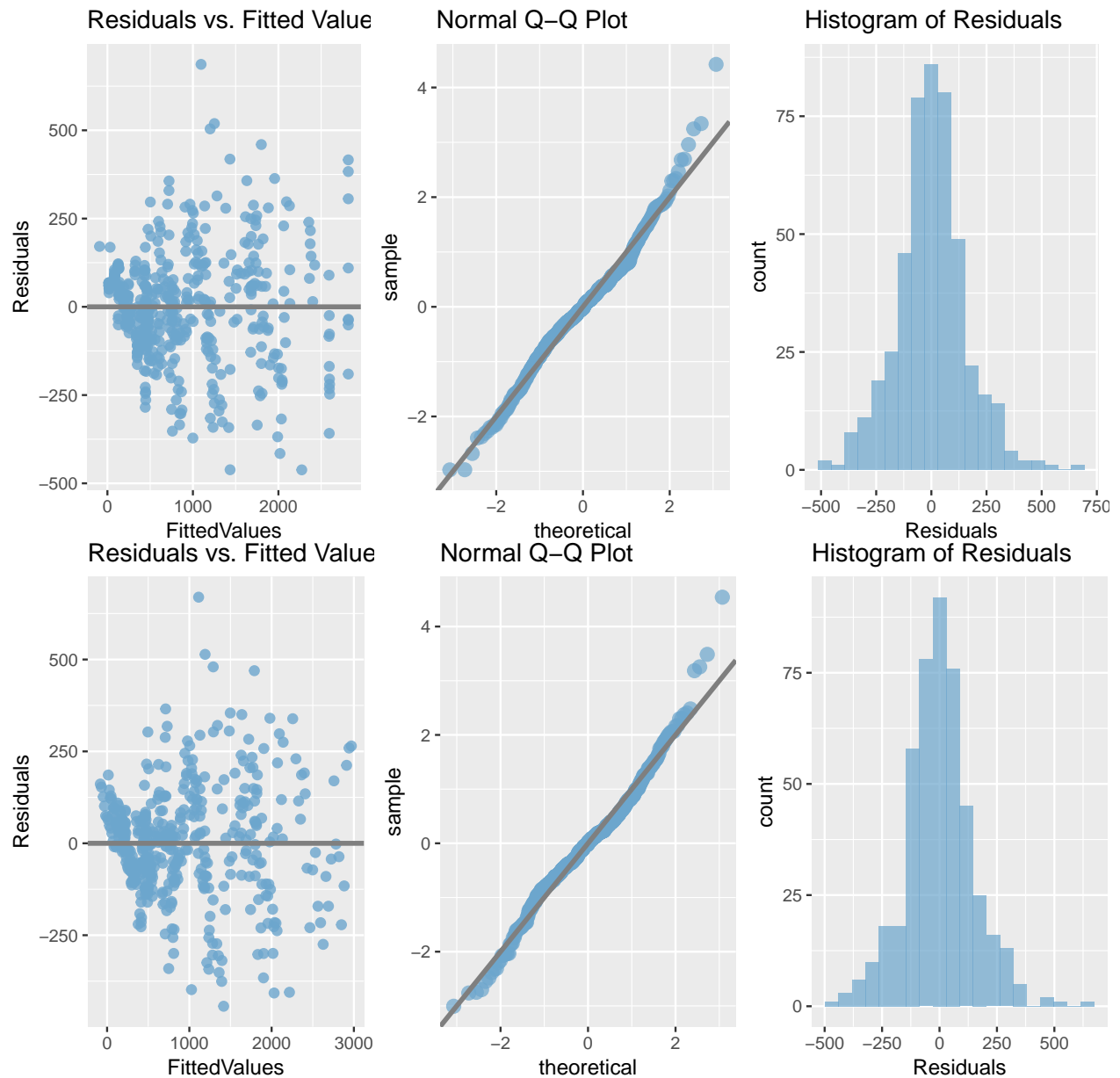
To go back to main text: 2.2.2.

```
CheckAssumptions(st3)
```

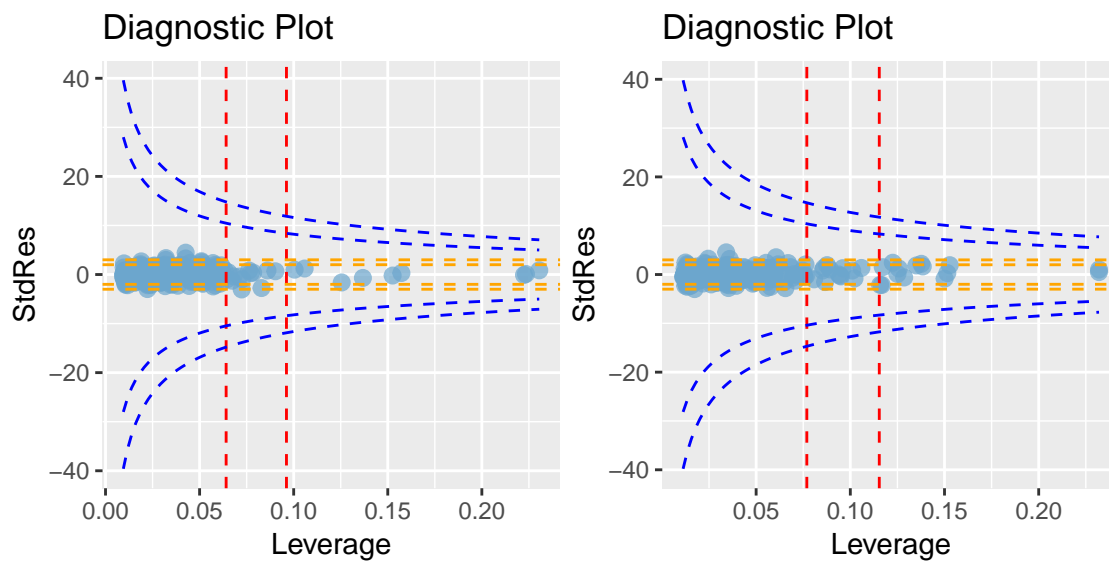
```
##
## studentized Breusch-Pagan test
##
## data: model
## BP = 97.857, df = 13, p-value = 4.312e-15
```

```
CheckAssumptions(st4)
```

```
##
## studentized Breusch-Pagan test
##
## data: model
## BP = 109.2, df = 15, p-value = 2.297e-16
```



```
UnusualPoints(st3, NationalParkVisits)
UnusualPoints(st4, NationalParkVisits)
```



To go back to main text: 2.2.2.

5.3.4 USE

To go back to main text: 3.3.2.

#This chunk could not be executed by R so I left it out.

```
predict(st4, list(Lodging3="over500", Seasonnum=3,
Year=2017, Park="Yellowstone", Campers=146.306),
interval = "confidence")
```

```
##      fit      lwr      upr
## 1 2969.06 2853.958 3084.161
```

```
predict(st4, list(Lodging3="over500", Seasonnum=3,
Year=2017, Park="Yellowstone", Campers=146.306),
interval = "prediction")
```

```
##      fit      lwr      upr
## 1 2969.06 2652.899 3285.221
```

To go back to main text: 3.3.2.