HW2: Short suggested answers

11 March 2019

Section 1.8 Exercise 1

Case 3: the following predictor variables may be useful, assuming the relevant data are available

a. Model and make of the vehicles
b. Odometer reading (mileage)
c. Conditions of the vehicles
d. Company the vehicle was leased to
e. Color of the vehicles

Case 4:

a. Day of the week,
b. Day of the year,
c. Is the day before the long weekend?
d. Is the day before or in the beginning of the school holidays (one variable per state)?
e. Is the day in the end of school holidays (one variable per state)?
f. Is the day before or in the beginning of a major sport event?
g. Is the day after a major sport event?
h. competitors' rpices (relative to the airline in question)
i. Is there a pilot strike?

Section 5.10 Ex 1

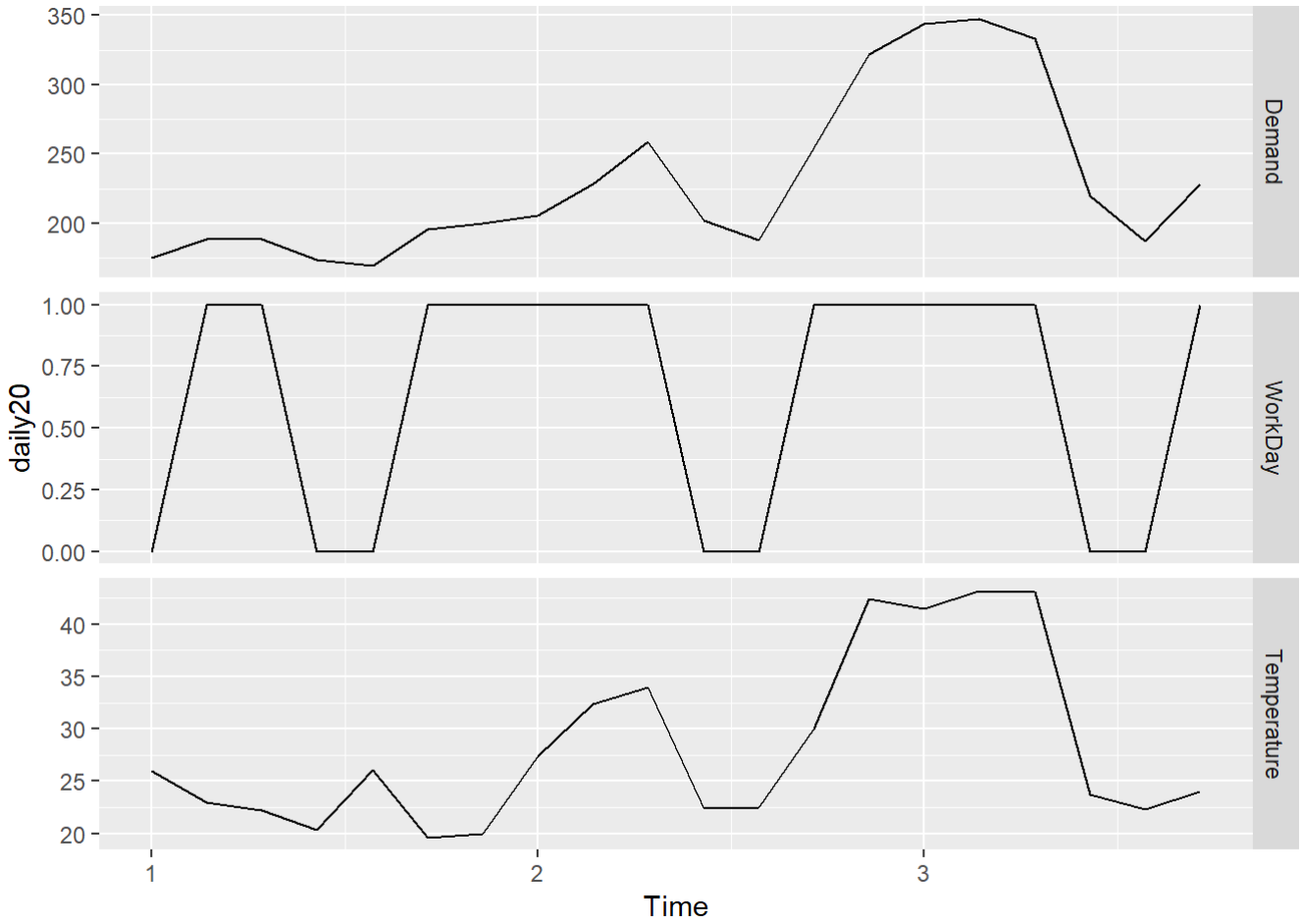
```
library(fpp2)
```

Daily electricity demand for Victoria, Australia, during 2014 is contained in `elecdaily`. The data for the first 20 days can be obtained as follows.

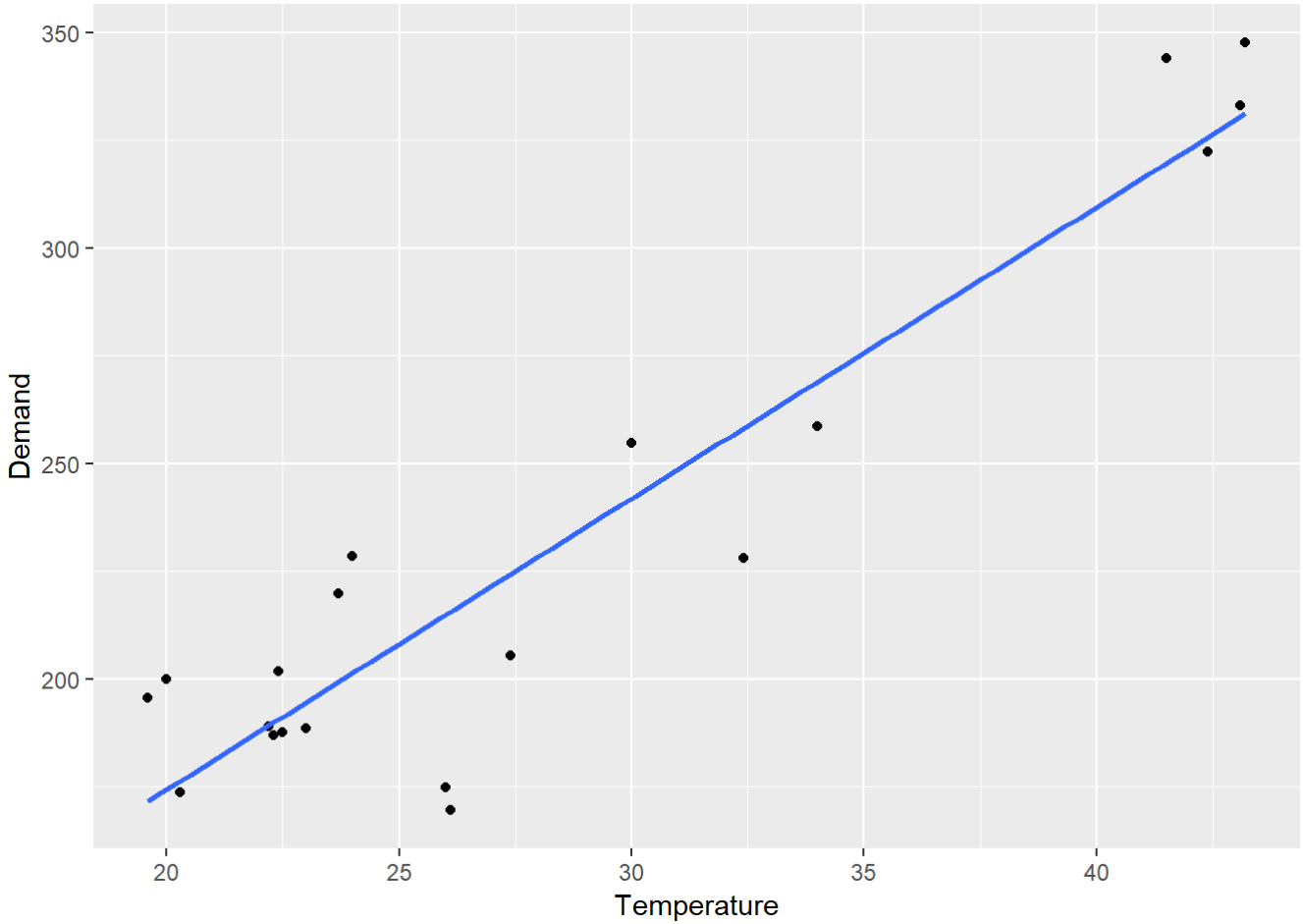
```
daily20 <- head(elecdaily,20)
```

a. Plot the data and find the regression model for Demand with temperature as an explanatory variable. Why is there a positive relationship?

```
autoplot(daily20, facets=TRUE)
```




```
daily20 %>%
  as.data.frame() %>%
  ggplot(aes(x = Temperature, y = Demand)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE)
```




```
(fit <- tslm(Demand ~ Temperature, data = daily20))
```

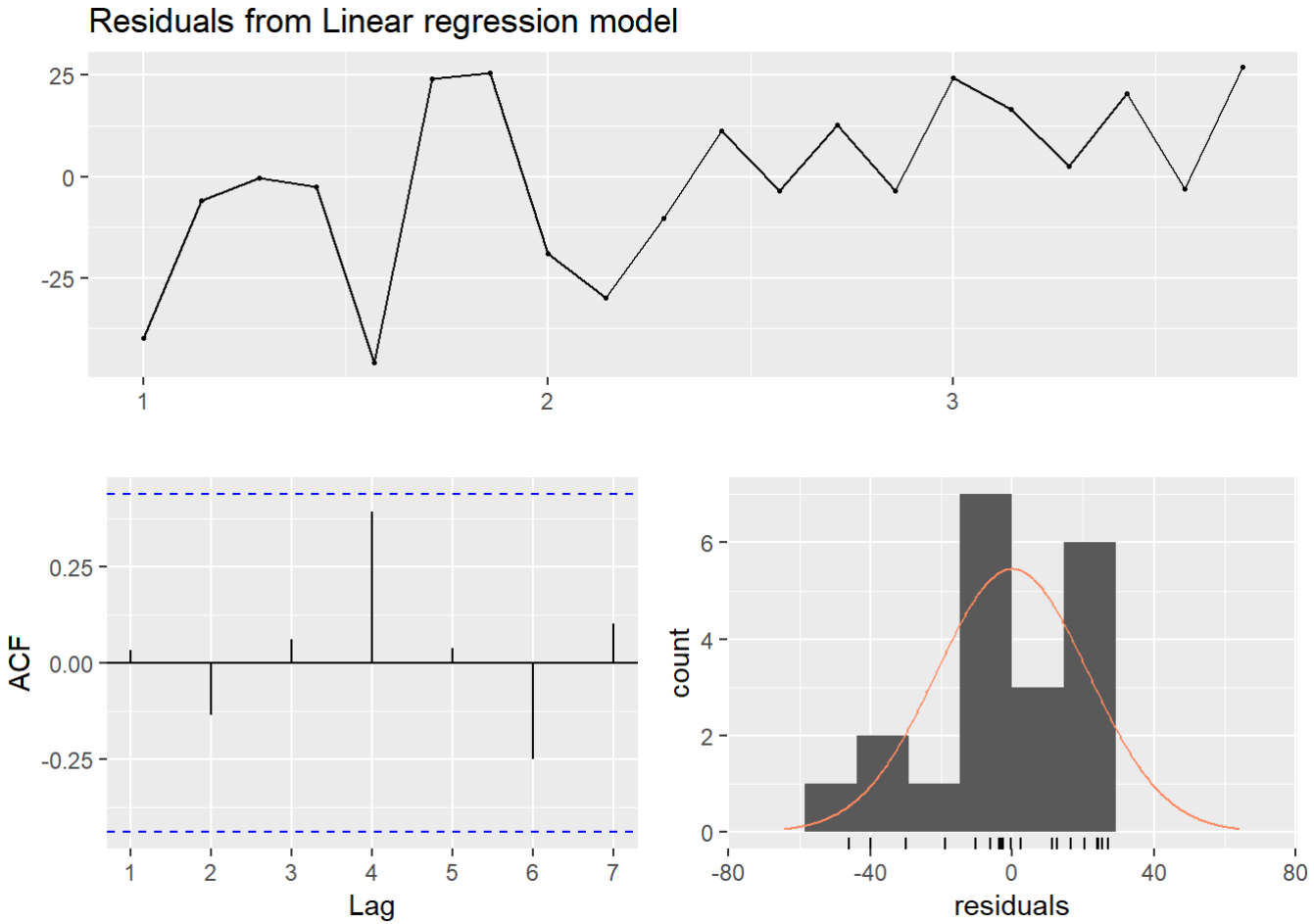
```
##
## Call:
## tslm(formula = Demand ~ Temperature, data = daily20)
##
## Coefficients:
## (Intercept)  Temperature
##      39.212        6.757
```

There is a positive relationship between temperature and electricity consumption.

Given the time of year, and the recorded temperature values, it is likely that electricity is being used for air conditioners. Since higher temperatures mean a higher demand for cooling, this leads to a positive relationship between temperature and electricity consumption.

b. Produce a residual plot. Is the model adequate? Are there any outliers or influential observations?

```
checkresiduals(fit)
```



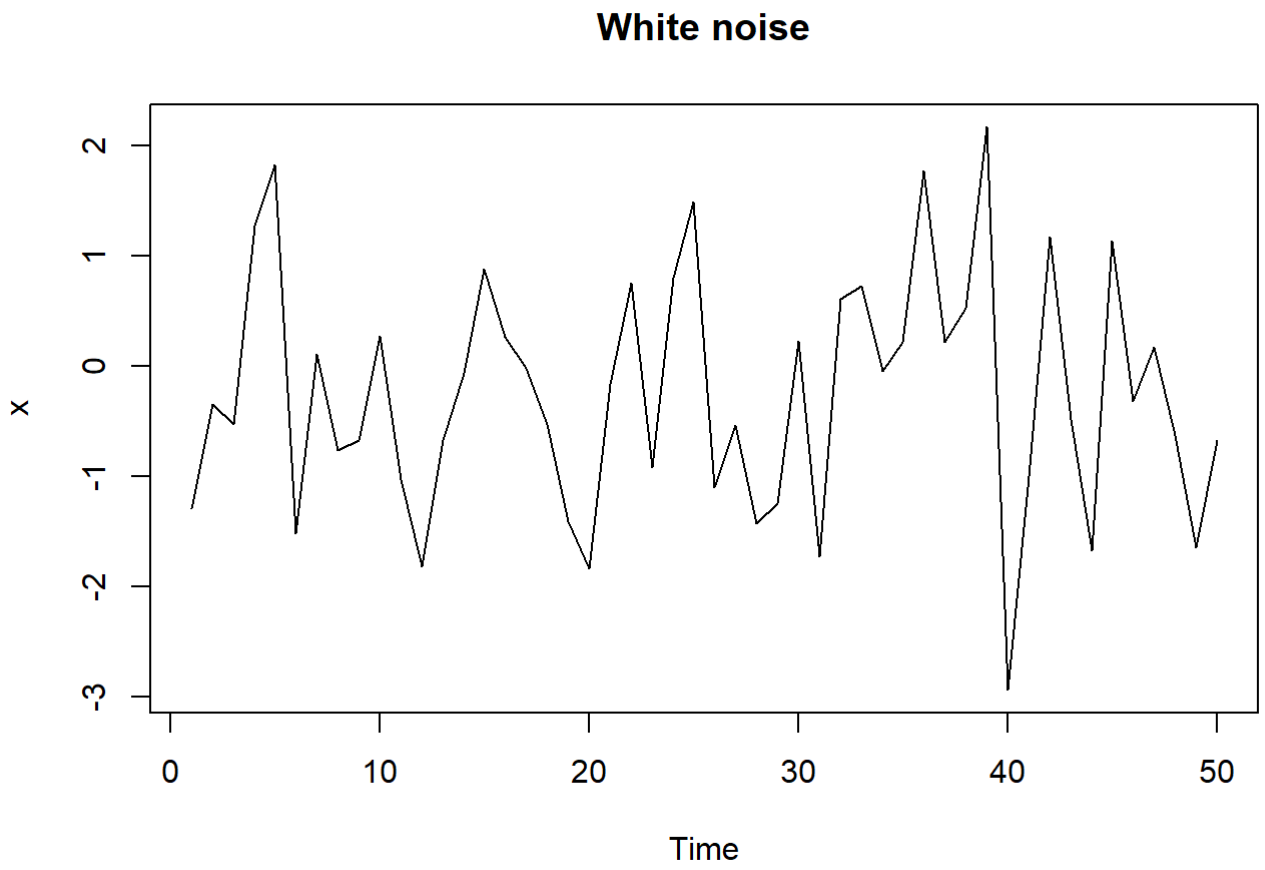
```
##
##  Breusch-Godfrey test for serial correlation of order up to 5
##
## data:  Residuals from Linear regression model
## LM test = 3.8079, df = 5, p-value = 0.5774
```

Although the ACF tests are passed, there is a linear trend in the residuals. So the model looks inadequate.

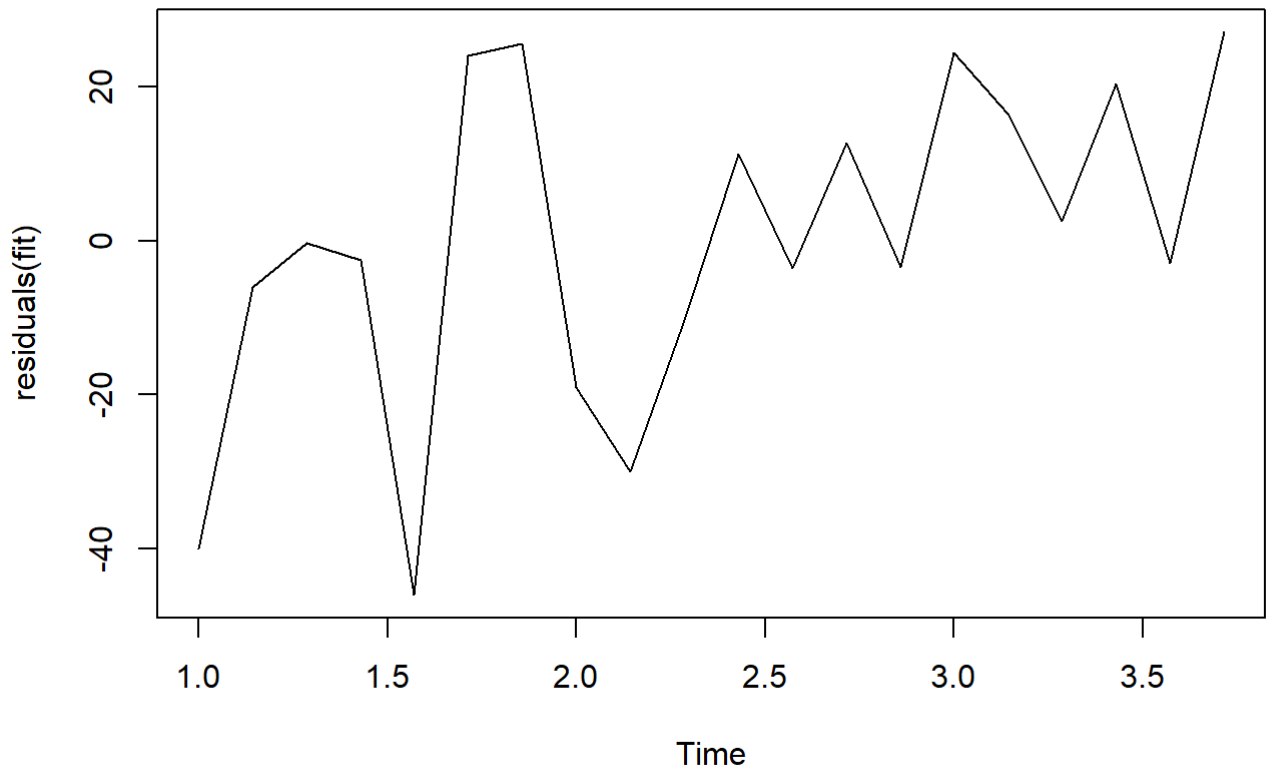
Note that majority answered that the model can no longer be improved as the errors are white noise.

For comparison, white noise errors are shown below:

```
set.seed(30); x <- ts(rnorm(50)); plot(x, main="White noise")
```




```
plot(residuals(fit))
```



c. Use the model to forecast the electricity demand that you would expect for the next day if the maximum temperature was 15° and compare it with the forecast if the with maximum temperature was 35°. Do you believe these forecasts?

```
(fc <- forecast(fit, newdata=data.frame(Temperature=c(15,35))))
```

```
##          Point Forecast    Lo 80    Hi 80     Lo 95    Hi 95
## 3.857143       140.5701 108.6810 172.4591  90.21166 190.9285
## 4.000000       275.7146 245.2278 306.2014 227.57056 323.8586
```

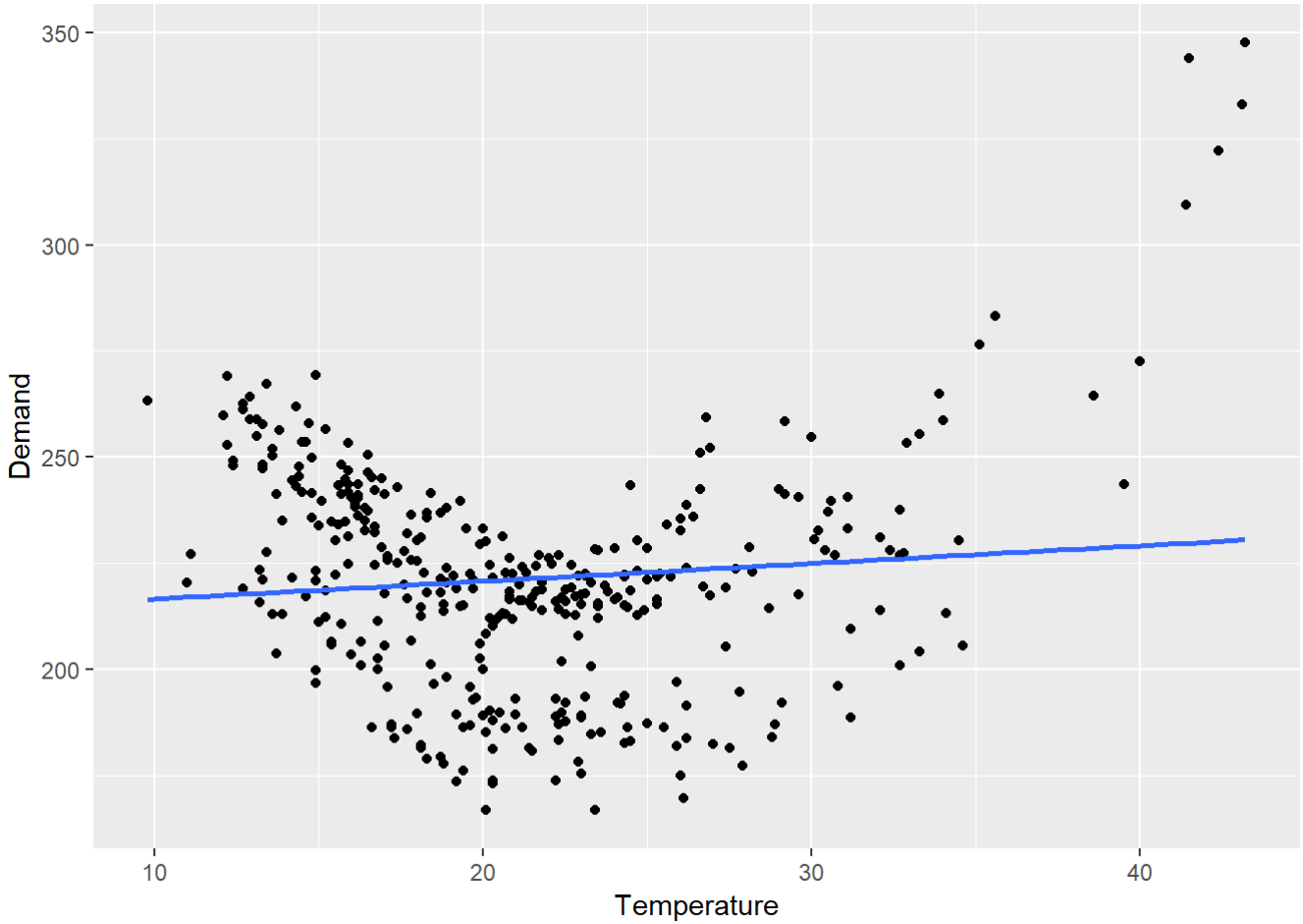
The prediction for 35° looks reasonable, but the one for 15° assumes the trend continues to decrease for temperature values lower than 20, which is unlikely. Heating will mean it will increase for lower temperatures.

> d. Give prediction intervals for your forecasts.

See above.

> e. Plot Demand vs Temperature for all of the available data in `elecdaily`. What does this say about your model?

```
elecdaily %>%
  as.data.frame() %>%
  ggplot(aes(x = Temperature, y = Demand)) +
    geom_point() +
    geom_smooth(method = "lm", se = FALSE)
```



This shows the non-linear relationship clearly. Even limiting the data to above 20, there is a nonlinear relationship between demand and temperature. The model is inadequate.