Lab 12: Quadratic Regression

Electricity Demand

We have data from the Australian Energy Market Operator and the Australian Bureau of Meteorology with daily electricity demand for Victoria, Australia, in 2014. For each day, we have:

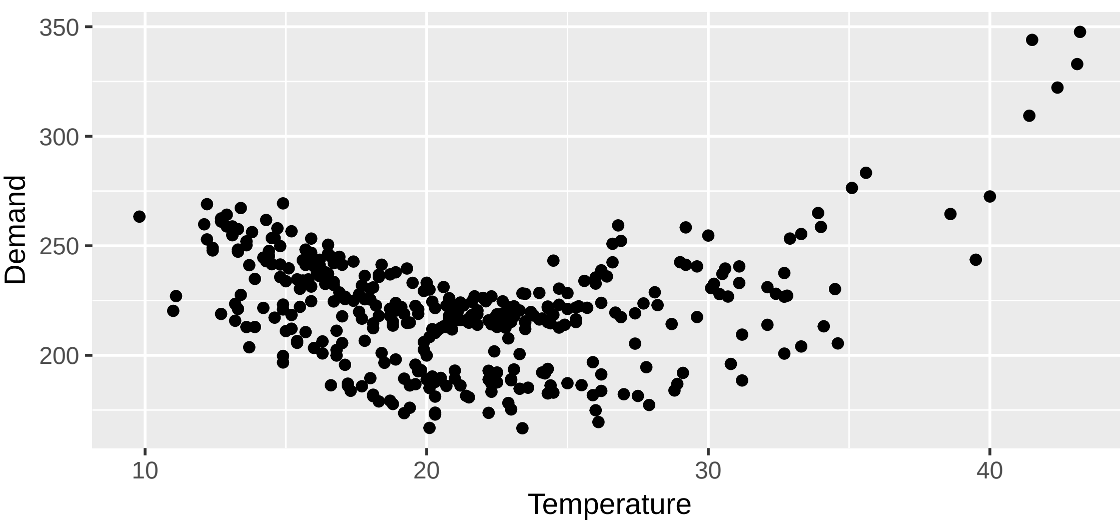
- `Demand`: Total electricity demand in GW for Victoria, Australia
- `WorkDay`: "WorkDay" for work days, and "Other" for non work days
- `Temperature`: The daily high temeprature in degrees Celsius

```
## # A tibble: 6 x 3
##   Demand WorkDay Temperature
##    <dbl> <chr>         <dbl>
## 1   175. Other          26
## 2   189. WorkDay        23
## 3   189. WorkDay        22.2
## 4   174. Other          20.3
## 5   170. Other          26.1
## 6   196. WorkDay        19.6
```

As always with data collected over time, we should be suspicious of the condition of independence. For today, let's set that aside and focus on an analysis of the relationships between these variables.

1. Make a plot of the data, treating Demand as the response and Temperature as the explanatory variable.

```r
ggplot(data = elecdaily, mapping = aes(x = Temperature, y = Demand)) +
  geom_point()
```



2. Fit an appropriate regression model using Temperature as an explanatory variable. Print a summary of your model fit.

```r
qf <- lm(Demand ~ poly(Temperature, degree = 2, raw = TRUE), data = elecdaily)
summary(qf)
```

```
##
## Call:
## lm(formula = Demand ~ poly(Temperature, degree = 2, raw = TRUE),
##     data = elecdaily)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -45.42 -18.03   6.11  14.43  48.72
##
## Coefficients:
##                                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                 387.69194   10.73907   36.10   <2e-16 ***
## poly(Temperature, degree = 2, raw = TRUE)1 -15.28379    0.91830  -16.64   <2e-16 ***
## poly(Temperature, degree = 2, raw = TRUE)2   0.32371    0.01861   17.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.62 on 362 degrees of freedom
## Multiple R-squared:  0.4602, Adjusted R-squared:  0.4573
## F-statistic: 154.3 on 2 and 362 DF,  p-value: < 2.2e-16
```

3. How good are the predictions of electricity demand from this model? Answer based on the residual standard error.

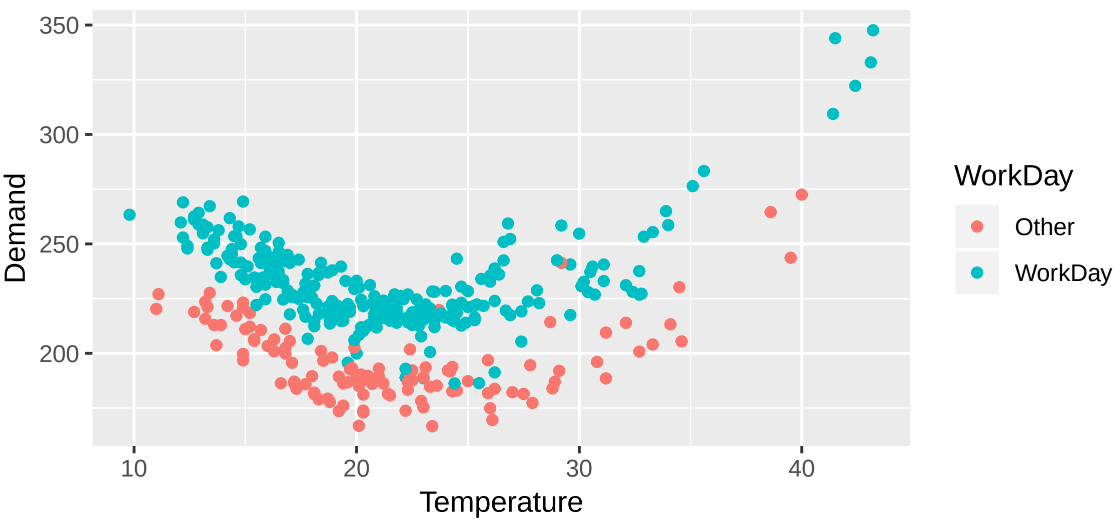
Approximately 95% of the fitted/predicted values for electricity demand are within plus or minus 40 GW of the actual electricity demand.

4. Write down the equation for the estimated mean electricity demand as a function of temperature.

$\hat{\mu}(Demand|Temperature) = 387.69 - 15.28Temperature + 0.32Temperature^2$

5. Make another plot of the data, this time coloring each day according to whether it is a work day or not.

```r
ggplot(data = elecdaily, mapping = aes(x = Temperature, y = Demand, color = WorkDay)) +
  geom_point()
```



6. Fit an appropriate model that uses both Temperature and WorkDay as explanatory variables. Do the data indicate any need for an interaction between the explanatory variables?

```r
lm_fit <- lm(Demand ~ WorkDay * poly(Temperature, degree = 2, raw = TRUE), data = elecdaily)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = Demand ~ WorkDay * poly(Temperature, degree = 2,
##     raw = TRUE), data = elecdaily)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.236  -5.859  -0.268   6.418  47.840
```

```
##
## Coefficients:
##                                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                                             353.25142   11.37302  31.060  < 2e-16 ***
## WorkDayWorkDay                                           46.21782   13.33144   3.467  0.00059 ***
## poly(Temperature, degree = 2, raw = TRUE)1              -14.34867    0.98389 -14.584  < 2e-16 ***
## poly(Temperature, degree = 2, raw = TRUE)2                0.30401    0.02014  15.096  < 2e-16 ***
## WorkDayWorkDay:poly(Temperature, degree = 2, raw = TRUE)1  -1.08591    1.14815  -0.946  0.34489
## WorkDayWorkDay:poly(Temperature, degree = 2, raw = TRUE)2   0.02466    0.02341   1.053  0.29286
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.83 on 359 degrees of freedom
## Multiple R-squared:  0.8371, Adjusted R-squared:  0.8348
## F-statistic:   369 on 5 and 359 DF,  p-value: < 2.2e-16
```

```r
lm_fit2 <- lm(Demand ~ WorkDay + poly(Temperature, degree = 2, raw = TRUE), data = elecdaily)
anova(lm_fit2, lm_fit)
```

```
## Analysis of Variance Table
##
## Model 1: Demand ~ WorkDay + poly(Temperature, degree = 2, raw = TRUE)
## Model 2: Demand ~ WorkDay * poly(Temperature, degree = 2, raw = TRUE)
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    361 42232
## 2    359 42069  2    163.23 0.6965  0.499
```

The data do not provide evidence of a need for an interaction between WorkDay and Temperature or Temperature squared.

7. How good are the predictions of electricity demand from this model? Answer based on the residual standard error.

Approximately 95% of the fitted/predicted values for electricity demand are within plus or minus 21.6 GW of the actual electricity demand.