Lab Report 2: OLS Model

Team #2

10/04/2021

Setup

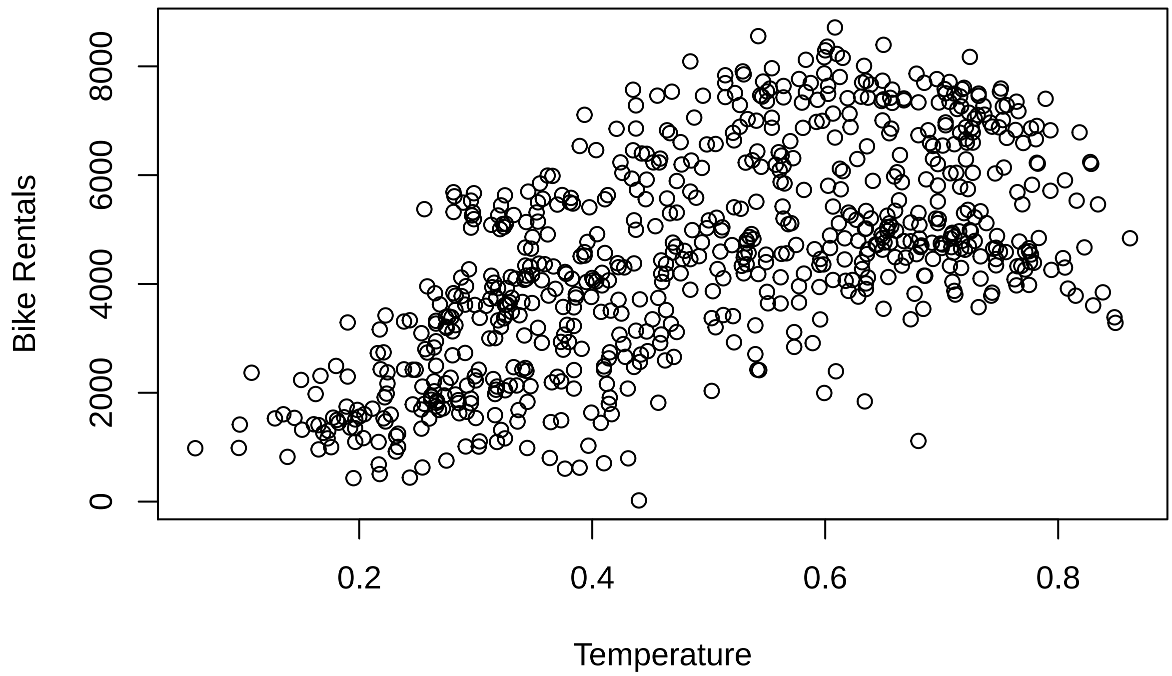
```r
# Load libraries
library(tidyverse)

# Read data
bike_data <- read.csv("../data/bike-day.csv",header=T)

# Clean data by fixing types
bike_data$temp <- as.numeric(bike_data$temp)
bike_data$cnt <- as.integer(bike_data$cnt)
```

Data Visualization

```r
# Scatterplot
plot(bike_data$temp, bike_data$cnt, ylab="Bike Rentals", xlab="Temperature")
```



Linear Model

Using the normalized temperature values as our predictor (t), we fit an ordinary least squares model of the daily bike usage count (c) in a quadratic model that takes the form:

$$c \sim \beta_0 + \beta_1 t + \beta_2 t^2$$

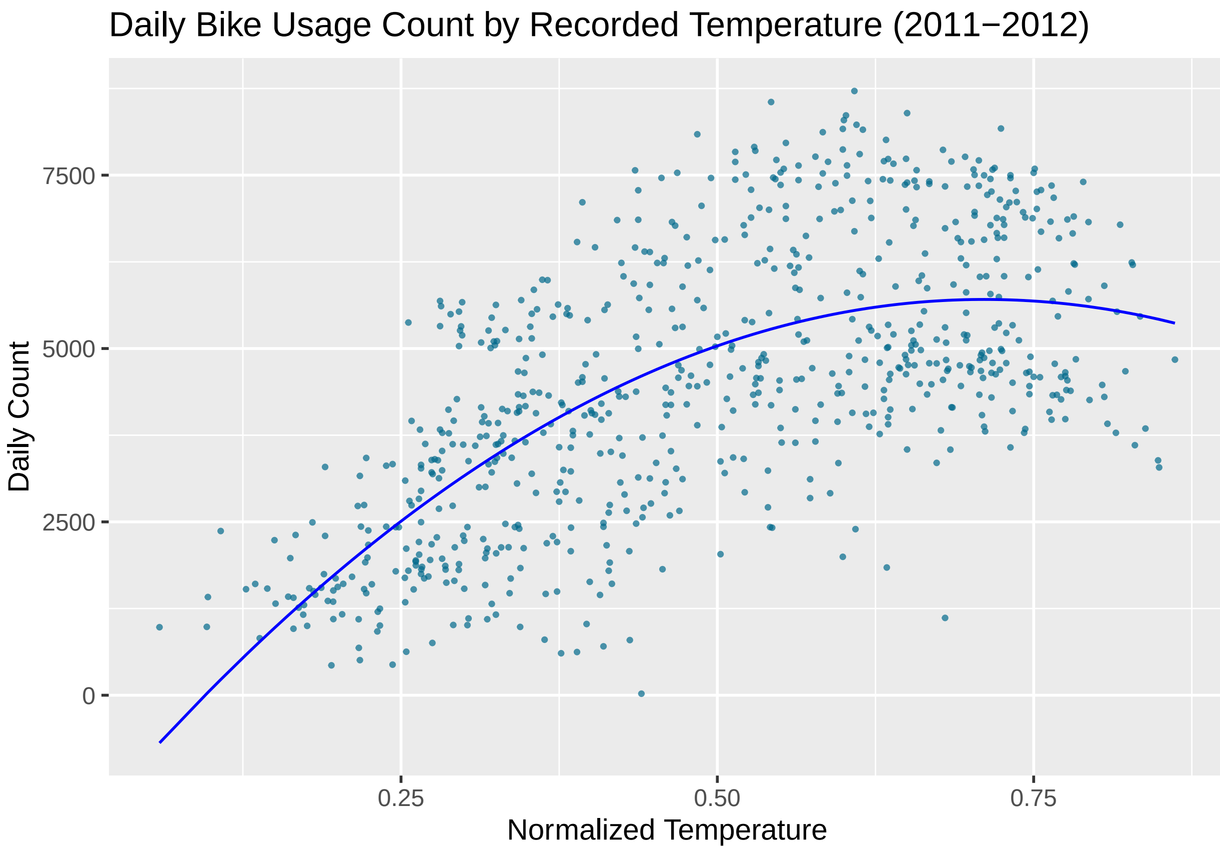
Model Results

```r
# Ordinary LS
m.ols <- lm(cnt~I(temp^2) + temp, data = bike_data)
summary(m.ols)
```

```
##
## Call:
## lm(formula = cnt ~ I(temp^2) + temp, data = bike_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4580.4 -1043.6   -79.1  1150.7  3274.5
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1902.0      382.4  -4.974 8.19e-07 ***
## I(temp^2)   -15055.0     1692.5  -8.895  < 2e-16 ***
## temp         21406.9     1685.2  12.703  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1434 on 728 degrees of freedom
## Multiple R-squared:  0.4532, Adjusted R-squared:  0.4517
## F-statistic: 301.7 on 2 and 728 DF,  p-value: < 2.2e-16
```

Model Visualization

```r
# Obtain the fitted values from the model
fitted_vals <- predict(m.ols)
# Plot the predicted values
bike_data %>%
  ggplot(aes(x=temp, y=cnt)) +
  geom_point(size = 0.5, color = "deepskyblue4", alpha = 0.7) +
  geom_line(mapping = aes(x = temp, y = fitted_vals), col= "blue") +
  labs(x = "Normalized Temperature", y = "Daily Count",
       title = "Daily Bike Usage Count by Recorded Temperature (2011-2012)")
```


Daily Bike Usage Count by Recorded Temperature (2011−2012)
Daily Count
7500
5000
2500
0
0.25
0.50
0.75
Normalized Temperature