FINAL PROJECT DATA 606

BIKE RENTALS

Marjete Vucinaj

May 15, 2023

Abstract

Bike rentals have increased in the past few decades as a mean of transportation. This helps reduce traffic, carbon emissions and is beneficial for riders' health.

The goal of this project is to determine if temperature predicts quantity of bike rentals. To investigate this, a bike share dataset, from Washington DC, containing temperature and bike rental quantity data was analyzed using regression analysis.

The result of this analysis shows a strong linear relationship (R squared of 39%) between temperature and number of bike rentals, specifically, as the weather is warmer, the number of bike rentals increase.

It is more important than ever that city planners continue to update infrastructure and ensure bike lanes are accessible in more neighborhoods. Additionally, bike rental companies should closely monitor their supply of bike gear in warmer months. Business' can also leverage data like this project to update their subscriptions or offer free wintertime referrals.

Overview

The Bike Sharing Data Set was obtained from UCI Machine Learning Repository and is a observational study.

The company, Capital Bike Share, publishes their data each quarter, and the researchers aggregated the data and added the corresponding weather information.

The data has 731 observations representing number of days in 2011 and 2012.

Overview

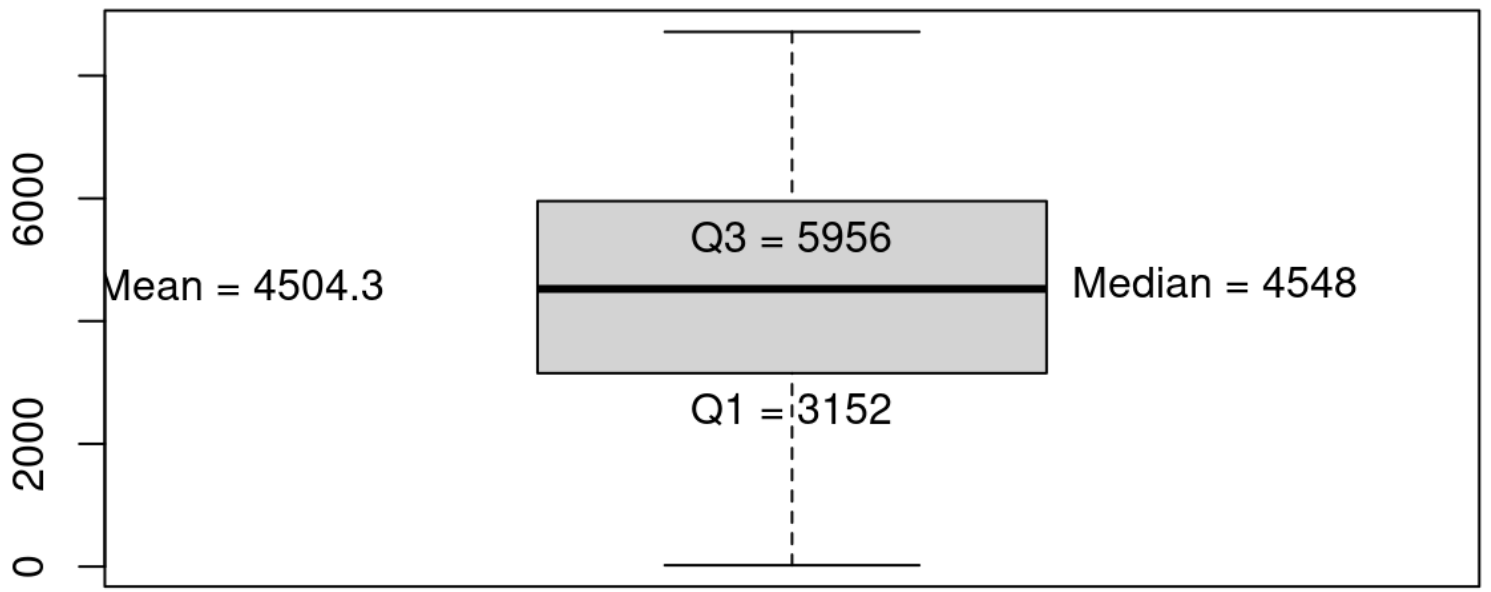
Research question: Does the temperature predict the quantity of daily bike rentals?

The dependent variable is the count of daily bike rentals

The independent variable is the temperature, this variable is normalized or rescaled so the range of these values is between 0-1

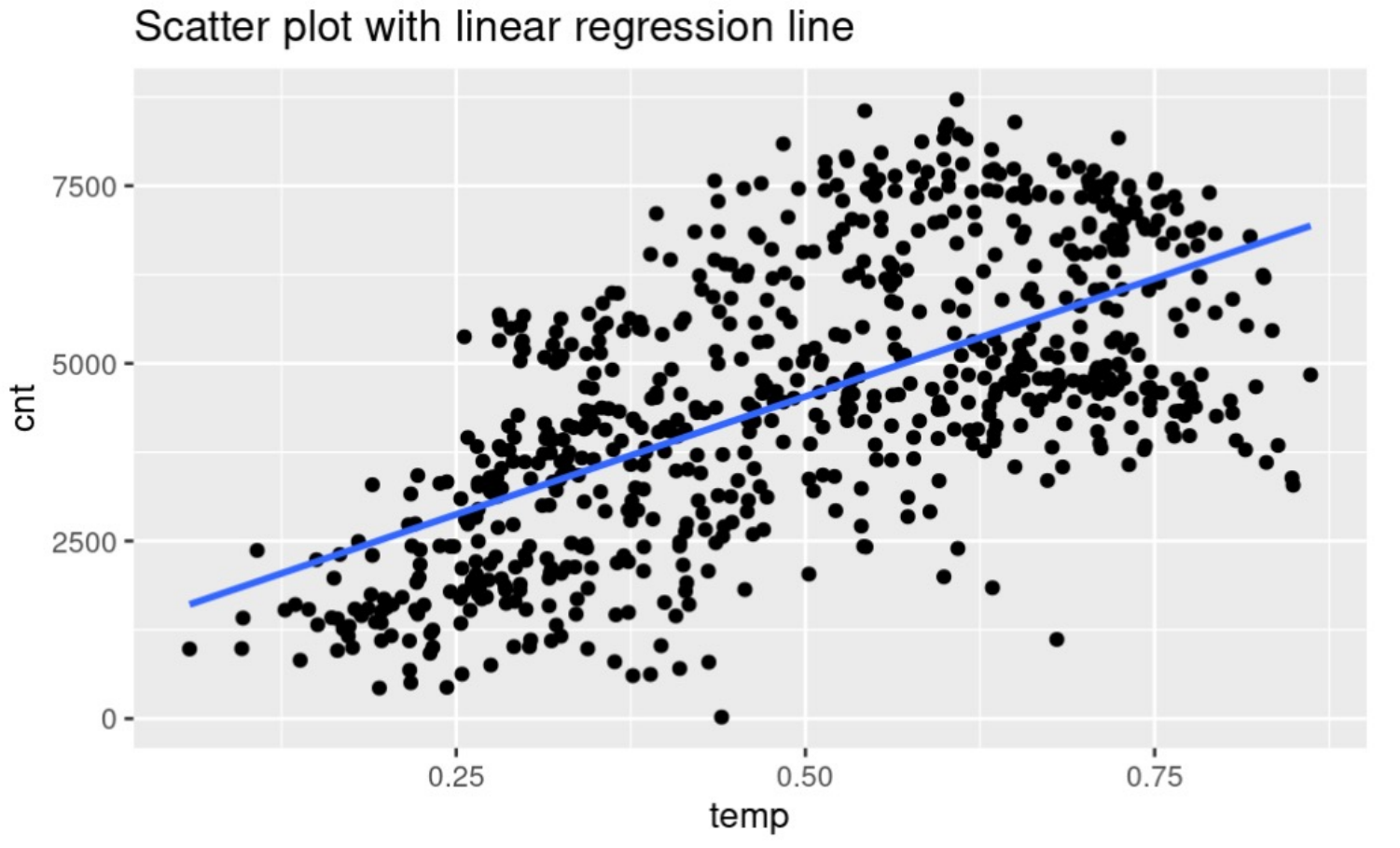
Summary statistics

cnt		temp	
Min.	: 22	Min.	:0.05913
1st Qu.	:3152	1st Qu.	:0.33708
Median	:4548	Median	:0.49833
Mean	:4504	Mean	:0.49538
3rd Qu.	:5956	3rd Qu.	:0.65542
Max.	:8714	Max.	:0.86167



BOX PLOT OF COUNT DATA

Data visualizations



Statistical output

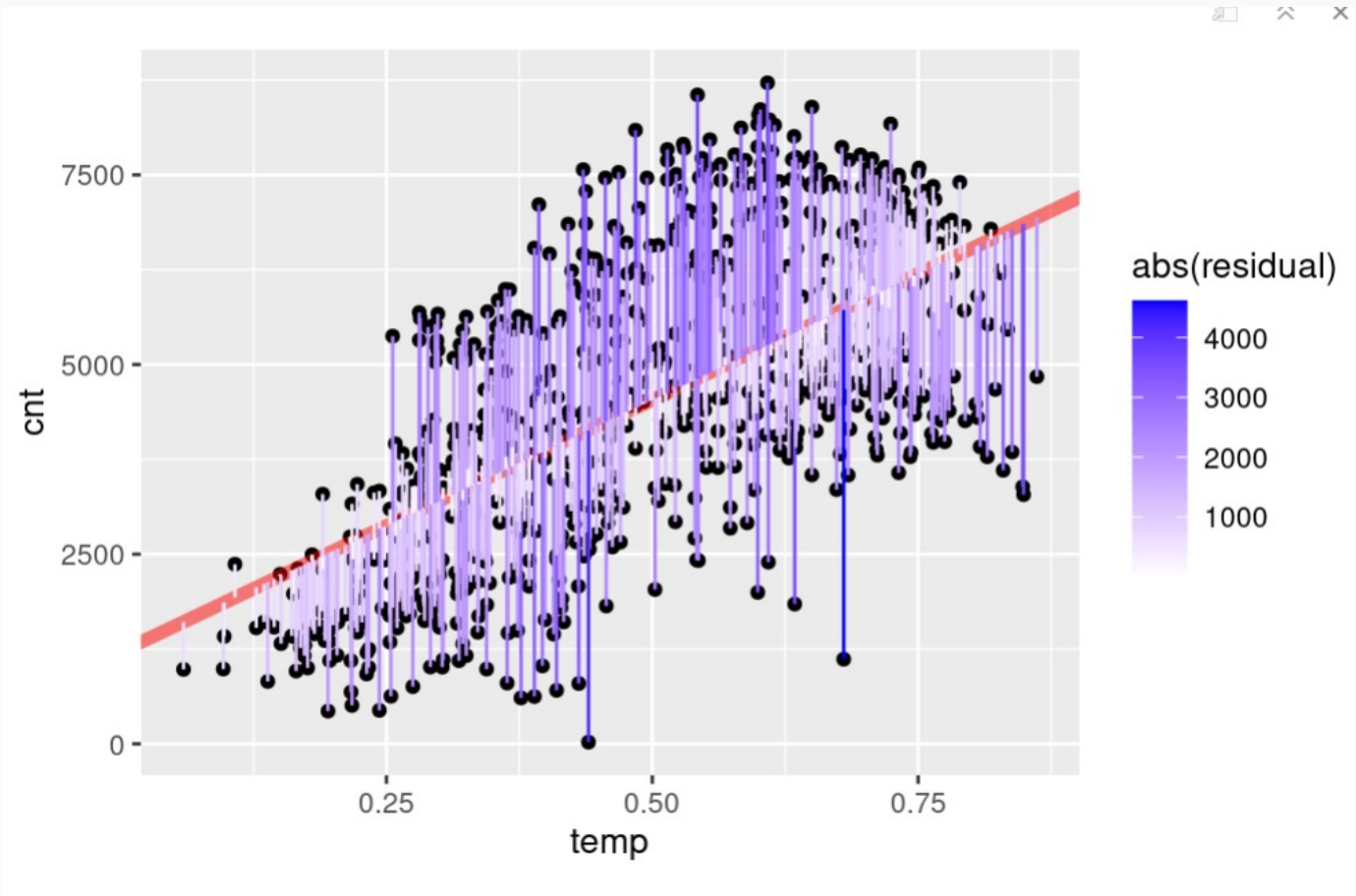
```{r}
mydata.lm <- lm(cnt ~ temp, data=mydata)
summary(mydata.lm)
```

```
Call:
lm(formula = cnt ~ temp, data = mydata)

Residuals:
    Min      1Q  Median      3Q     Max
-4615.3 -1134.9  -104.4  1044.3  3737.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1214.6      161.2   7.537 1.43e-13 ***
temp          6640.7      305.2  21.759  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1509 on 729 degrees of freedom
Multiple R-squared:  0.3937,    Adjusted R-squared:  0.3929
F-statistic: 473.5 on 1 and 729 DF,  p-value: < 2.2e-16
```

R squared is 0.39, meaning my model accounts for 39% of the variability of temperature predicting number of bike rentals

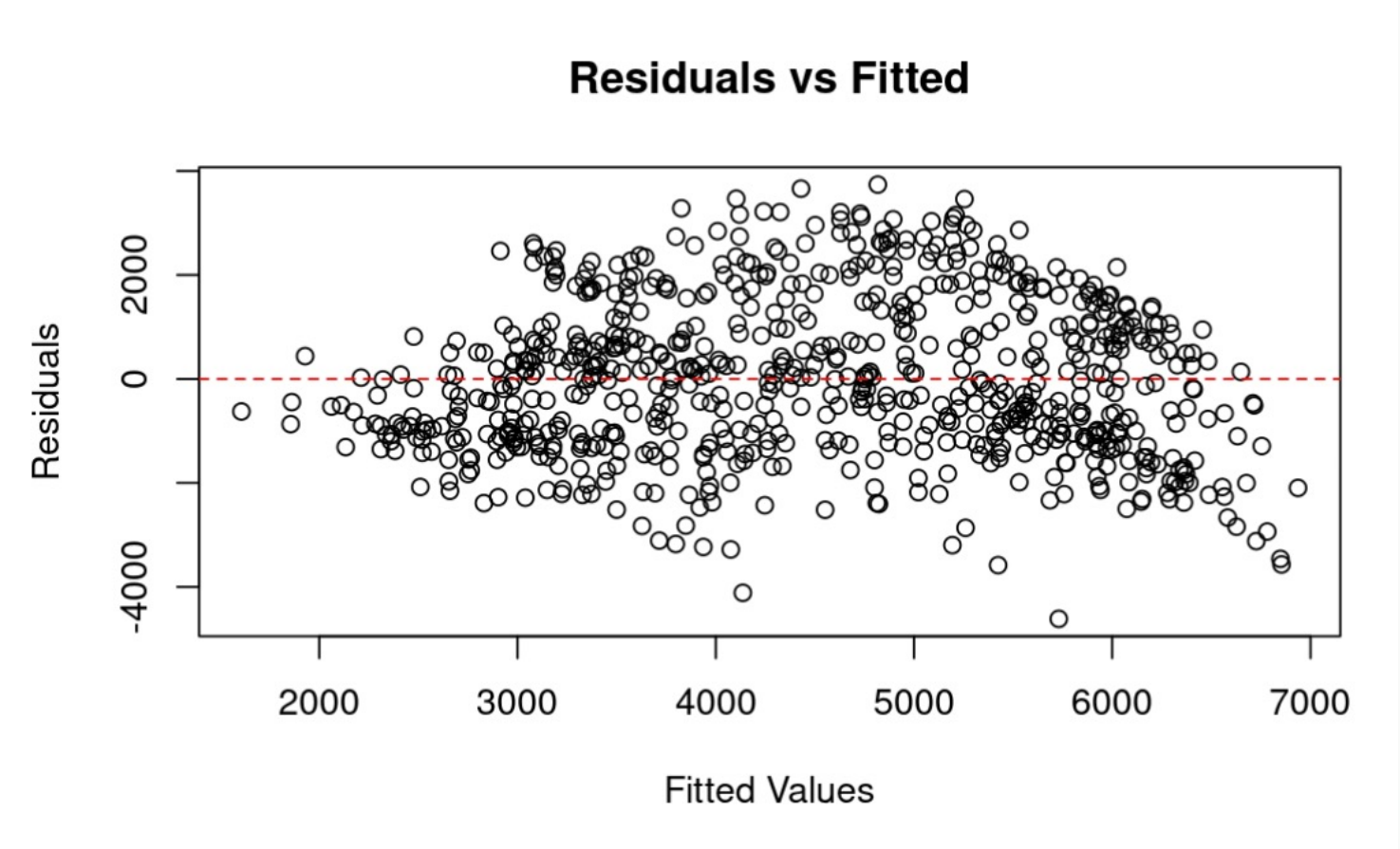
Statistical output

To assess whether the linear model is reliable, we need to check for
(1)linearity,
(2) nearly normal residuals, and
(3) constant variability.



Scatter plot with residuals

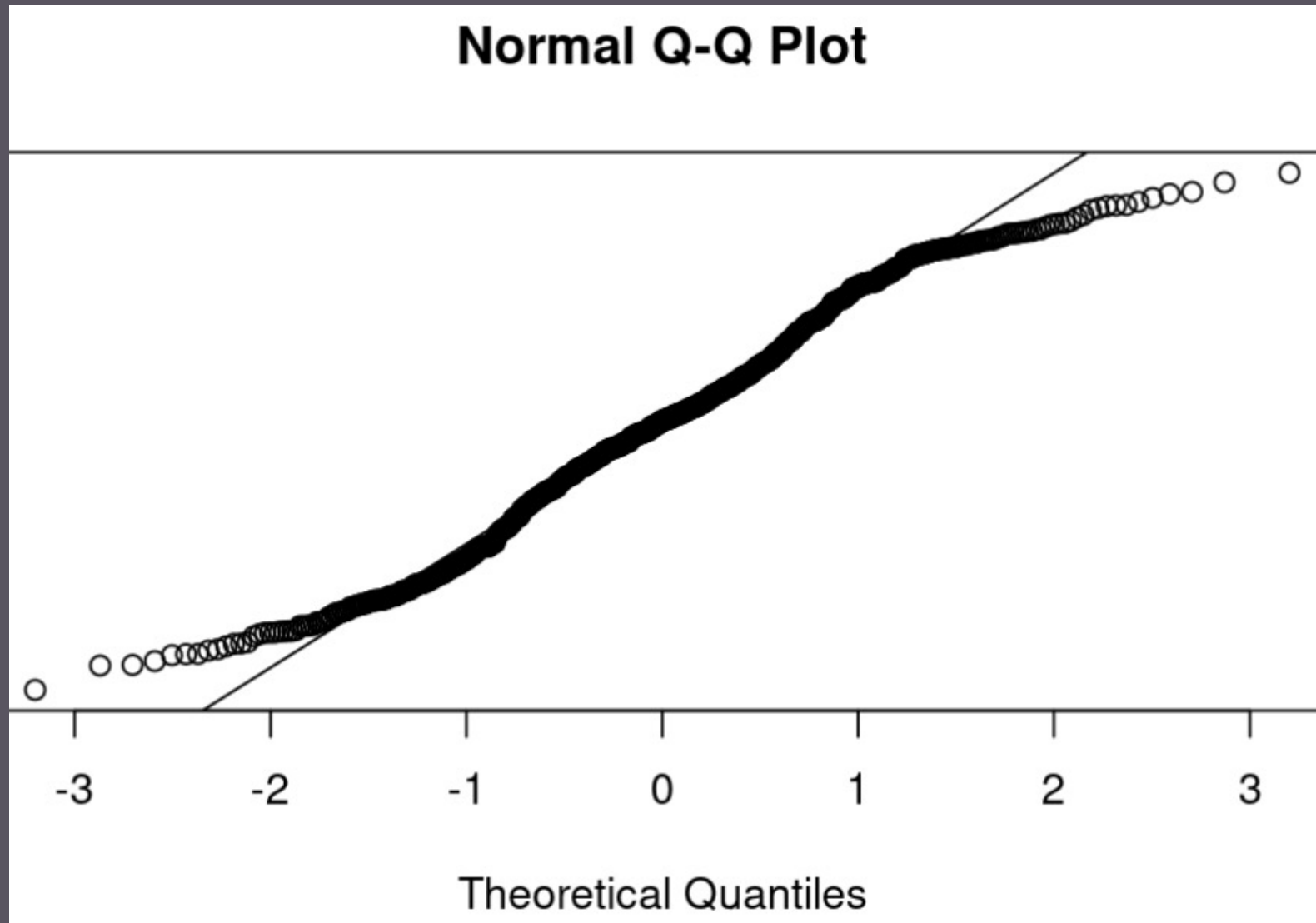
Statistical output



Condition of constant variability is met as there is no pattern around the zero, suggesting a linear relationship.

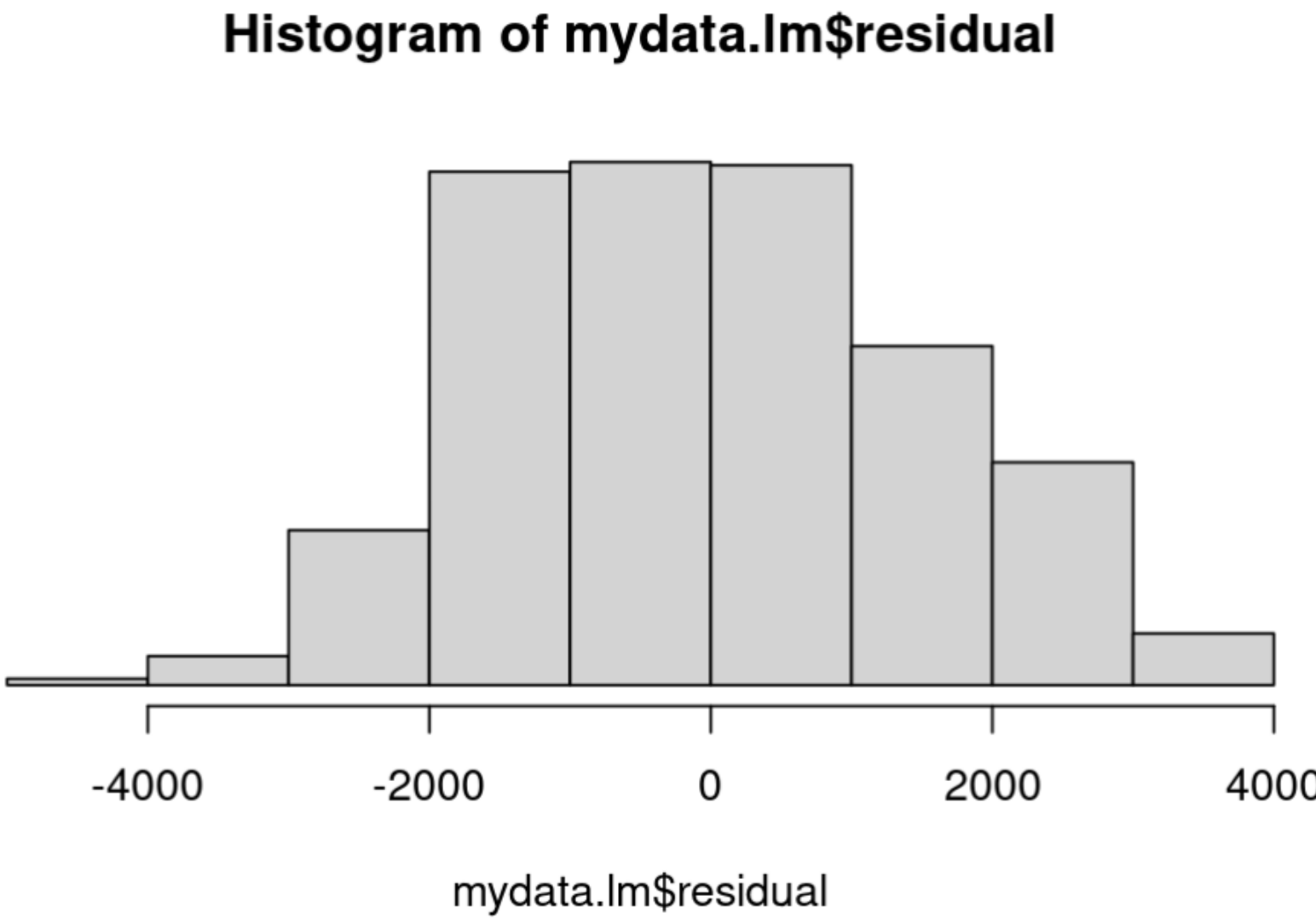
Statistical output

The normality assumption can be confirmed with qqplot as most points do not deviate from line.



Statistical output

The histogram of residuals look close to a normal distribution. It has a left skew.



Conclusion

39% of my model suggests that temperature predicts number of daily bike rental.

Importance: As bike rentals have become a more common mode of transportation, it is ideal to examine the factors that encourage one to rent a bike.

- It is helpful for city planners to know which neighborhoods have a high bike rentals to ensure bike lanes are accessible.

- Companies can ensure they have adequate supply of biking gear especially in warmer weather.

- While casual users can rent a bike with most companies, businesses' typically make more profit with subscription-based models.

- Hypothetically casual users can receive reduced price subscription for the winter/ colder months to increase company revenue

Conclusion

Limitations: One of the limits of this analysis is the data itself being 11 years old at this point.

Future steps would be to utilize a multiple regression model with current data to investigate if other variables can more strongly predict count of bike rentals.

Additionally, I would want analyze the categories of ‘registered users’ and ‘casual users’ to better understand how these categories impact the linear regression.

References

Source: Hadi Fanaee-T, Laboratory of Artificial Intelligence and Decision Support (LIAAD), University of Porto, INESC Porto, Campus da FEUP, Rua Dr. Roberto Frias, 378. 4200 - 465 Porto, Portugal

Thank you!!