A Bridge Too Far: Negative Binomial Distribution

Name: Mona B.
Date: December 11, 2023

Introduction

Angelia Dyer, representing the New York Department of Transportation, has reached out for assistance in understanding cycling behavior over the four major east river bridges (Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge and Queensboro Bridge). The department is in the process of allocating resources for roadway improvement projects, particularly those related to bicycling trails. With daily data spanning from April to October 2017, the team aims to investigate the impact of weather variables and temporal factors on ridership, considering the unique characteristics of each bridge.

In addition to identifying the factors influencing cyclists, the department is interested in uncovering trends in ridership over time and understanding how rain and temperature affect cycling behavior. Of particular interest is whether these effects vary across the four bridges. To facilitate the analysis, the team will provide a data dictionary and a Stata dataset. The insights gained from this analysis will inform decisions on resource allocation, improvement projects, and policy adjustments for the benefit of the city's cycling infrastructure.

Research Question

Primary Research Question:

- ➢ What weather and time-related characteristics significantly influence the ridership of cyclists over the four major east river bridges in New York City?

Subsidiary Research Questions:

- ➢ What is the trend in ridership over time? Is there some point during the year when cycling behavior really starts to peak?
- ➢ How does rain and temperature affect ridership? Is there a certain amount of rain that really causes cycling behavior to decline?
- ➢ For the above questions, are the results the same for all four bridges?

Data Collection

The data that was provided by the New York Department of Transportation was collected on a daily basis from April to October 2017, providing a comprehensive record of the number of cyclists traversing each of the four major east river bridges in New York City, with a focus on factors such as temperature, precipitation, day of the week, and month of the year.

Data Cleaning

In the data cleaning process for our analysis on cycling behavior over the four major east river bridges, several key steps were undertaken. Initially, the relevant time-related data was converted into a DateTime format in R for consistent representation. Subsequently, a new column was created to extract and store the month information from the DateTime variable, providing additional temporal granularity. A thorough check for missing values was conducted to ensure data completeness. The precipitation column was re-coded, handling specific cases such as converting "Trace" to 0.001 inches of rain and converting the snow days to NA. Additionally, to streamline temperature data, reported high and low temperatures were mutated to calculate the average temperature for each day. High and low temperatures were related so taking the average temperature was appropriate. These data cleaning procedures were implemented to ensure the dataset's accuracy, reliability, and readiness for subsequent analysis of cycling patterns over the four bridges. The decision to condense days (weekend vs. not the weekend) was based on the weekends having the lowest overall cycling compared to weekedays.

Since the New York Department of Transportation is interested in rain, temperature, temporal data (month and days), these variables will be included and considered when predicting the number of cyclists.

Data Exploration

Fig 1.1 Codebook

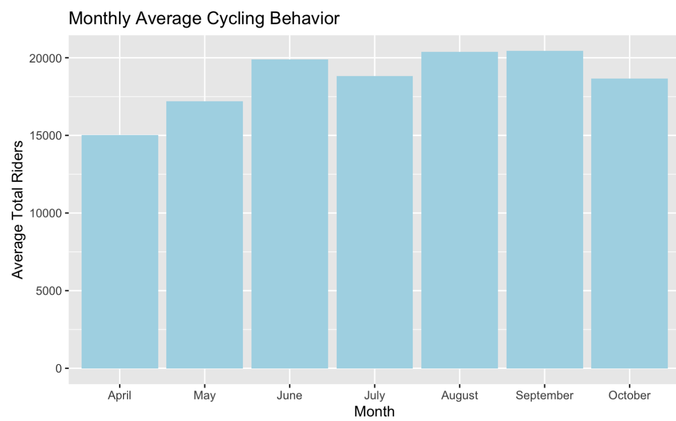
Variable_Name	Variable_Type	Description	Measurement_Unit	Missing_Values
date	Date	Month/Day	YYYY-MM-DD	None
weekend	Categorical	Weekend/Not Weekend	0/1	None
month	Categorical	Month	April/May/June...	None
average_temperature	Numerical	Average daily temperature	Fahrenheit	None
rain	Numerical	Daily precipitation	Inches	None
brooklynbridge	Numerical	# of riders over Brooklyn Bridge	Count	None
manhattanbridge	Numerical	# of riders over Manhattan Bridge	Count	None
williamsburgbridge	Numerical	# of riders over Williamsburg Bridge	Count	None
total	Numerical	Total riders over these four bridges	Count	None

Fig 1.2 Summary Statistic

	TRUE (N=214)
Brooklyn Bridge	
Mean (SD)	2680 (855)
Median [Min, Max]	2860 [151, 4960]
Manhattan Bridge	
Mean (SD)	5350 (1750)
Median [Min, Max]	5610 [484, 8240]

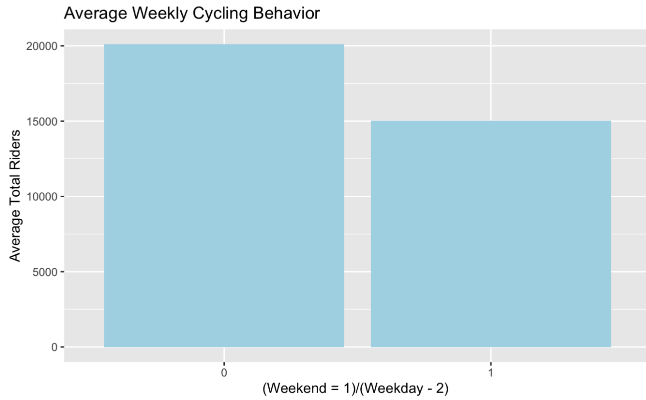
Williamsburg Bridge	
Mean (SD)	6050 (1760)
Median [Min, Max]	6290 [874, 8870]
Queensboro Bridge	
Mean (SD)	4550 (1310)
Median [Min, Max]	4680 [865, 6580]
Total	
Mean (SD)	18600 (5540)
Median [Min, Max]	19400 [2370, 27000]
Rain	
Mean (SD)	0.132 (0.394)
Median [Min, Max]	0 [0, 3.03]
Average Temperature (°F)	
Mean (SD)	68.1 (9.57)
Median [Min, Max]	70.0 [41.5, 86.0]

Fig 1.3 Monthly Average Cycling Behavior



The monthly analysis (Fig. 1.3) of bike ridership across all four bridges show that from April to September, there is a consistent increasing trend, with peak values observed in June, September, and August. However, in October, there is a slight decrease in the mean count. Seasonal patterns are evident, with higher cycling during the warmer months (May to August) and potential fluctuations associated with varying weather conditions.

Fig 1.4 Average Weekly Cycling Behavior



The analysis suggests that bike ridership tends to be on average higher on weekdays compared to weekends.

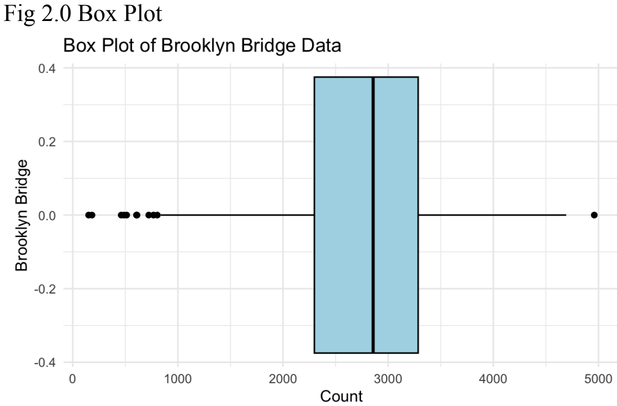
The provided data contains count data (the number of cyclists on each bridge). The **Poisson regression distribution** is well-suited for count data.

Bridge 1: Brooklyn Bridge

The Brooklyn Bridge, also known as the most famous and UNESCO-protected bridge in New York, experiences high foot and bike traffic due to its popularity. Connecting Lower Manhattan and Brooklyn, the bridge's pedestrian walkway and cycling lanes attract thousands of visitors, tourists, and locals alike. Its panoramic views of the city skyline make it an attraction for those looking for both a practical route and a memorable experience, whether by foot or bike.

Looking at Count Data:

Fig 2.0 Box Plot



There are a few outliers in this data set. The outliers can be caused by external factors such as potential road closures, holidays, bridge events, weather warnings and more.

Relationship Between Week Type and Number of Cyclists on Brooklyn Bridge.

Fig 2.1 Mean, Total, Standard Deviation of Cyclist on Brooklyn Bridge by Type of Week

weekend <dbl>	n <int>	mean <dbl>	sd <dbl>
0	152	2782.717	833.4486
1	62	2428.323	860.6327

For weekdays (weekend = 0), there are 152 observations with an average cyclist count of approximately 2783 and a standard deviation of around 833.45 for the Brooklyn Bridge. On the weekends (weekend = 1), there are 62 observations with an average cyclist count of about 2428.32 and a standard deviation of approximately 860.63. This suggests that, on average, there is a slightly higher cyclist count on weekdays compared to weekends.

Relationship Between Months and Number of Cyclists on Brooklyn Bridge.

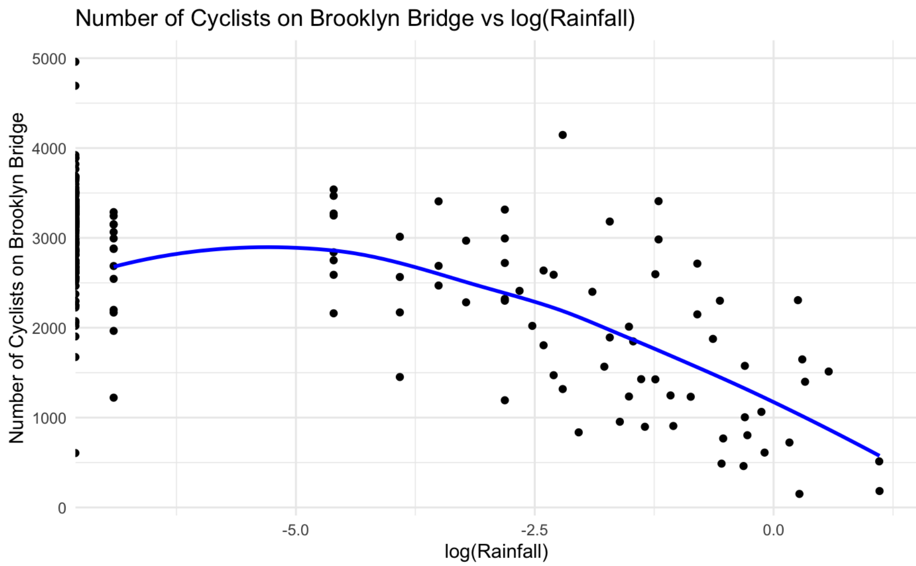
Fig. 2.2 Table of Average Monthly Counts of Cyclists on Brooklyn Bridge

month <fctr>	n <int>	mean <dbl>	sd <dbl>
April	30	2250.100	980.3012
May	31	2451.323	971.7178
June	30	2756.867	683.2127
July	31	2755.968	659.3642
August	31	3060.484	821.6654
September	30	2895.633	753.5339
October	31	2585.484	864.5973

In April, the mean count was 2,250.1, with an increase in May to 2,451.32. June sees a further rise to 2,756.87, and July has a similar count. August increased to 3,060.48, highlighting potential seasonal factors influencing cycling behavior. There is a slight decrease in September to an average of 2,895.63, and October exhibits a further decrease to 2,585.48.

Relationship Between Rain (log) and Number of Cyclists on Brooklyn Bridge.

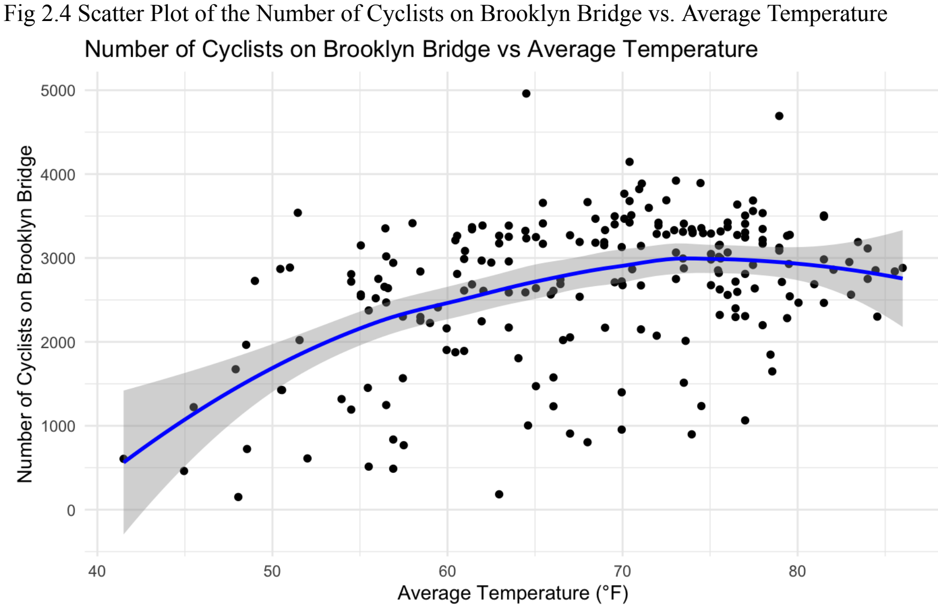
Fig 2.3 Scatter Plot of the Number of Cyclists on Brooklyn Bridge vs. the Log of Rain



A log transformation was used on rainfall to visually see the patterns between rainfall and the number of cyclists on Brooklyn Bridge. As the inches of rainfall increases, the number of cyclists on Brooklyn Bridge decreases.

Relationship between Temperature and Number of Cyclists on Brooklyn Bridge.

Fig 2.4 Scatter Plot of the Number of Cyclists on Brooklyn Bridge vs. Average Temperature



As the average temperature increases in Fahrenheit, the number of cyclists on Brooklyn Bridge increases. There seems to be potential outliers.

The Null Model:

Fig 2.5 Summary of the Null Model

```
Call:
glm(formula = brooklynbridge ~ 1, family = poisson(link = "log"),
    data = centered_brooklynbridge_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  7.89359    0.00132    5978   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 70021  on 213  degrees of freedom
Residual deviance: 70021  on 213  degrees of freedom
AIC: 72087
```

The null model for the Brooklyn Bridge Bridge has no predictors. The estimated intercept is 7.89359, with a standard error of 0.001, resulting in a z-value of 5,978 and a highly significant p-value (p <0.01). The null deviance, representing the goodness of fit for the null model, is 70,021 on 213 degrees of freedom. The Akaike Information Criterion (AIC) for this null model is 72,087. Further exploration of potential predictors will be examined further in the analysis.

Modeling Predictors:

Fig 2.6 ANOVA For the Four Models

```
Analysis of Deviance Table

Model 1: brooklynbridge ~ 1
Model 2: brooklynbridge ~ average_temp + as.numeric(rain)
Model 3: brooklynbridge ~ average_temp + as.numeric(rain) + factor(month)
Model 4: brooklynbridge ~ average_temp + as.numeric(rain) + factor(month) +
    factor(weekend)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       213      70021
2       211      33806  2    36216 < 2.2e-16 ***
3       205      32863  6      943 < 2.2e-16 ***
4       204      31677  1     1185 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance Table compares different models predicting the count of cyclists on the Brooklyn Bridge. Model 1, which includes only the intercept, is the null model. Subsequent models incorporate additional predictors: Model 2 includes the average temperature and inches of rain, Model 3 adds months and Model 4 further includes weekends. Model 2 significantly improves the fit, as evidenced by a deviance reduction of 36,216 with a p-value less than 0.01. Model 3 and Model 4 continue to enhance the fit significantly, with p-values indicating highly significant improvements. These results suggest adding temperature (centered at the mean), rainfall (centered at the mean), month, and weekend variables is the best fit. Model 4 will be considered, but further investigation of the underlying assumptions must be done.

Assessing Interaction:

Fig 2.7 ANOVA to Analyze Interaction

```
Analysis of Deviance Table

Model 1: brooklynbridge ~ factor(month) + factor(weekend) + average_temp + 
    as.numeric(rain)
Model 2: brooklynbridge ~ factor(month) + factor(weekend) + average_temp + 
    as.numeric(rain) + average_temp * as.numeric(rain)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)    
1       204      31677                          
2       203      28887  1     2790 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance Table provides supporting evidence that the inclusion of an interaction term in the predictive model for cyclists counts on the Brooklyn Bridge. Model 1, which includes the main effects of temperature (centered at the mean), rainfall (centered at the mean), month, and day, served as the baseline. Model 2, which is an extension Model 1 includes addition of an interaction term between temperature and rainfall. There is a significant improvement between Model 1 and Model 2. The chi-squared test for the difference in deviance between the two models has an extremely low p-value ($p < 0.01$). In conclusion, this test strongly supports the presence of an interaction effect. Therefore, it will be added to the model. Since there is interaction, numeric variables will be centered at its mean for the purpose of interpretation. Since effect modification has been confirmed, confounder will not be evaluated.

Assessing the Assumptions for Poisson Distribution:

Fig 2.8 Overdispersion Tests for Brooklyn Bridge

```
	Overdispersion test

data:  bb_interaction_model
z = 5.318, p-value = 5.245e-08
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion 
  142.8918 
```

The overdispersion test on the Model 4 (Fig 2.6) reveals significant evidence of overdispersion ($z = 5.318$, $p < 0.01$). The estimated dispersion parameter is 142.8918, indicating that the variability in cyclist counts on the Brooklyn Bridge does not satisfy the assumptions for a Poisson distribution. Therefore, the negative binomial regression is required.

Goodness of Fit on the Final Model:

Fig 2.9 GOF for Brooklyn Bridge

```
Model: Negative Binomial(12.6054), link: log
Response: brooklynbridge
```

```
Terms added sequentially (first to last)

                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                213     460.89
month              6   24.454       207     436.43 0.0004308 ***
weekend            1   10.078       206     426.36 0.0015007 **
average_temp       1   35.195       205     391.16 2.983e-09 ***
rain               1  162.807       204     228.35 < 2.2e-16 ***
average_temp:rain  1   10.205       203     218.15 0.0014003 *
```

The Analysis of Deviance Table for the Poisson regression model on Brooklyn Bridge ridership demonstrates the sequential inclusion of predictors. Average temperature ($p < 0.01$), rainfall ($p < 0.01$), monthly variations ($p < 0.01$), and interaction term each significantly contribute to explaining variability in ridership, making this model a good fit. The Poisson Deviance Goodness-of-fit ($p = 0.2216194$) provides additional evidence that the final model fits the data.

Final Model:

Fig 2.10 Final Model for Brooklyn Bridge

```
Call:
glm.nb(formula = brooklynbridge ~ factor(month) + factor(weekend) +
    average_temp + as.numeric(rain) + average_temp * as.numeric(rain),
    data = centered_brooklynbridge_data, init.theta = 12.60541684,
    link = log)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       7.846776   0.062089 126.380  < 2e-16 ***
August            0.090407   0.086915   1.040  0.29826
July             -0.023677   0.091860  -0.258  0.79660
June              0.041951   0.085509   0.491  0.62371
May               0.076232   0.073747   1.034  0.30128
October           0.048764   0.075006   0.650  0.51560
September         0.073456   0.082240   0.893  0.37176
weekend          -0.120589   0.042815  -2.817  0.00485 **
average_temp      0.014798   0.002873   5.151 2.59e-07 ***
rain             -0.630129   0.057495 -10.960  < 2e-16 ***
average_temp:rain 0.016346   0.005478   2.984  0.00285 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(12.6054) family taken to be 1)

    Null deviance: 460.89  on 213  degrees of freedom
Residual deviance: 218.15  on 203  degrees of freedom
AIC: 3438.9

Number of Fisher Scoring iterations: 1


              Theta:  12.61
          Std. Err.:  1.22

 2 x log-likelihood:  -3414.874
```

Results:

The negative binomial regression model was fitted to predict the counts of cyclists on the Brooklyn Bridge. The model includes predictors such as the month of the year, whether it's a weekend, the average temperature, rain, and the interaction between average temperature and rain. Average temperature and rain were centered at its mean.

Centering average temperature and rain at their means, the estimated intercept of 7.85 signifies the expected log count when all other predictors are at their mean values. The reference group is non-weekend days in April. Month-specific effects for August, July, June, May, October, and September were estimated, but none are statistically significant. The "weekend" variable has a significant negative effect, indicating that on weekends, the expected log count decreases by 0.12 compared to weekdays. Average temperature shows a significant positive effect, with a 0.015 increase in the expected log count for each one-unit increase. Rain has a significant negative effect, with a 0.63 decrease in the expected log count for each one-unit increase. The interaction between average temperature and rain is also significant, suggesting a modification in the effect of temperature on the log count based on the level of rain.

The model's goodness of fit is assessed through the deviance, and the residual deviance is 218.15 on 203 degrees of freedom. The AIC is 34,38.9, indicating a reasonably good fit compared to the reduced model (Fig 2.5).

The interaction term suggests that the relationship between average temperature and the log count of cyclists is influenced by the level of rain. The positive coefficient for the interaction indicates that the positive effect of average temperature on the log count when centered at its mean is strengthened when it's rainy.

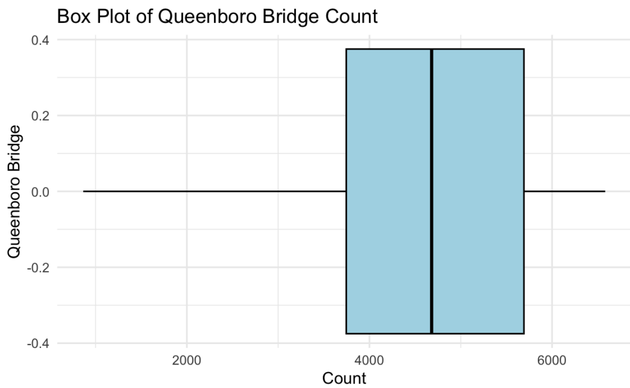
In conclusion, the number of cyclists on the Brooklyn Bridge is influenced by various factors, including the month, whether it's a weekend, average temperature, rain, and the interaction between average temperature and rain. Rain and average temperature were centered at their means to enhance the interpretation of the interaction term.

Bridge 2: Queensboro Bridge

The Queensboro Bridge, connecting Manhattan and Queens, is a transportation link. Across the bridge there are transportation stations (train, buses, etc). With bike paths and scenic views, it's a favorite route for cyclists commuting between boroughs. This bridge is closest to the infamous park called Central Park.

Looking at Count Data:

Fig 3.0 Box Plot



There are no outliers present.

Relationship Between Week Type and Number of Cyclists on Queensboro Bridge.

Fig 3.1 Mean, Total, Standard Deviation of Cyclist on Queensboro Bridge by Type of Week

weekend <dbl>	n <int>	mean <dbl>	sd <dbl>
0	152	4922.875	1271.4723
1	62	3637.548	877.1759

For weekdays (weekend = 0), there are 152 observations with an average cyclist count of approximately 4922.88 and a standard deviation of around 1,271.47 for the Queensboro Bridge. On the weekends (weekend = 1), there are 62 observations with an average cyclist count of about 3,637.55 and a standard deviation of approximately 877.18. This suggests that, on average, fewer cyclists cross the Queensboro Bridge on weekends compared to weekdays, and the counts are generally less variable during weekends

Relationship Between Months and Number of Cyclists on Queensboro Bridge.

Fig. 3.2 Table of Average Monthly Counts of Cyclists on Queensboro Bridge

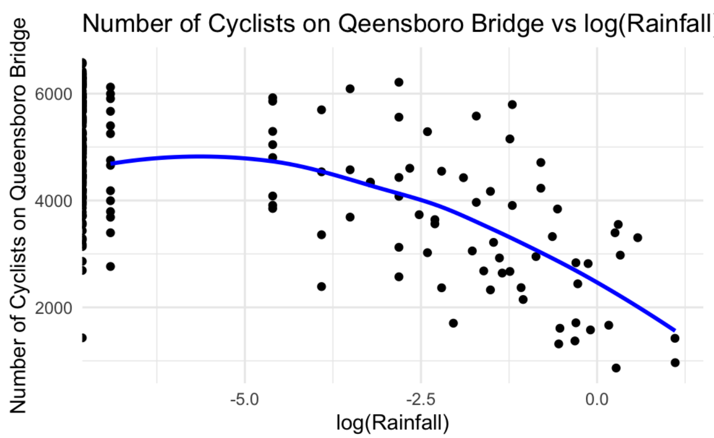
month <fctr>	n <int>	mean <dbl>	sd <dbl>
April	30	3482.800	1146.583
May	31	4074.839	1438.256
June	30	4736.533	1139.239

July	31	4550.581	1073.174
August	31	5169.387	1074.926
September	30	5053.833	1158.474
October	31	4773.258	1348.852

Fig. 3.2 shows the relationship between months and the number of cyclists on the Queensboro Bridge. There is an increase in average cyclist counts from April to August, with August exhibiting the highest counts. September and October show slightly lower counts, maintaining relatively high ridership.The standard deviation for May and October fluctuate.

Relationship Between Rain (log) and Number of Cyclists on Queensboro Bridge.

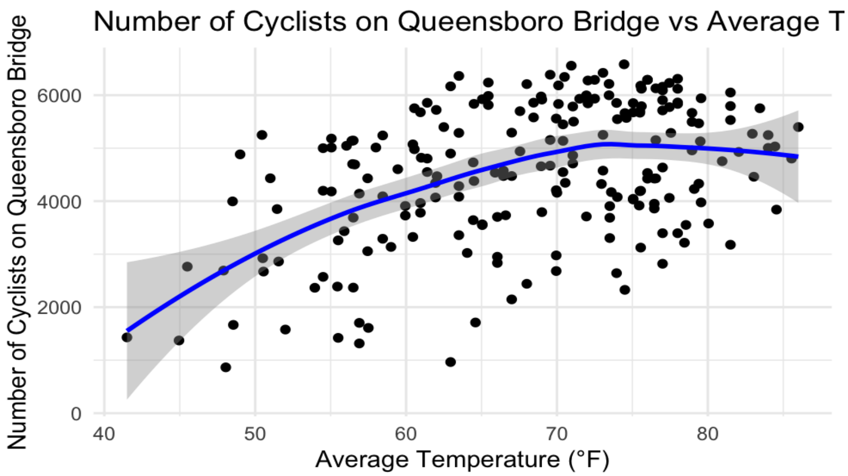
Fig 3.3 Scatter Plot of the Number of Cyclists on Queensboro Bridge vs. the Log of Rain



A log transformation was used on rainfall to visually see the patterns between rainfall and the number of cyclists on Queensboro Bridge. As the inches of rainfall increases, the number of cyclists on Queensboro Bridge decreases.

Relationship between Temperature and Number of Cyclists on Queensboro Bridge.

Fig 3.4 Scatter Plot of the Number of Cyclists on Queensboro Bridge vs. Average Temperature



As the average temperature increases in Fahrenheit, the number of cyclists on Queensboro Bridge increases. The scatter plot does not show extreme data points.

The Null Model:

Fig 3.5 Summary of the Null Model

```
Call:
glm(formula = queensborobridge ~ 1, family = poisson(link = "log"),
    data = centered_queensborobridge_data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 8.422990   0.001013    8312   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 89871  on 213  degrees of freedom
Residual deviance: 89871  on 213  degrees of freedom
AIC: 92057

Number of Fisher Scoring iterations: 4
```

The null model for the Queensboro Bridge has no predictors. The estimated intercept is 8.423, with a standard error of 0.001, resulting in a z-value of 8,312 and a highly significant p-value ($p < 0.01$). The null deviance, representing the goodness of fit for the null model, is 89,871 on 213 degrees of freedom. The Akaike Information Criterion (AIC) for this null model is 92,057. Further exploration of potential predictors will be examined further in the analysis.

Modeling Predictors:

Fig 3.6 ANOVA For the Four Models

```
Analysis of Deviance Table

Model 1: queensborobridge ~ 1
Model 2: queensborobridge ~ average_temp + as.numeric(rain)
Model 3: queensborobridge ~ average_temp + as.numeric(rain) + factor(month)
Model 4: queensborobridge ~ average_temp + as.numeric(rain) + factor(month) +
    factor(weekend)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       213      89871
2       211      47806  2    42065 < 2.2e-16 ***
3       205      42077  6     5729 < 2.2e-16 ***
4       204      28862  1    13215 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Model 1, with only the intercept, serves as the baseline, and subsequent models introduce average temperature, numeric rain, month, and weekend as predictors. The deviance decreases consistently across models, signifying a better fit. The p-values associated with the chi-square test indicate that all models with predictors are highly significant ($p < 0.01$). The inclusion of average temperature and rain (Model 2), followed by the addition of month (Model 3), and finally, the incorporation of weekend (Model 4), all contribute significantly to explaining the variability in cyclist counts. Model 4 will be considered, but further investigation of the underlying assumptions must be done.

Assessing Interaction:

Fig 3.7 ANOVA to Analyze Interaction

```
Analysis of Deviance Table

Model 1: queensborobridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain)
Model 2: queensborobridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain) + average_temp * as.numeric(rain)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       204      28862
2       203      26986  1   1875.5 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The difference in deviance between the two models is 1,875.5 with 1 degree of freedom, resulting in a highly significant p-value ($p < 0.01$). This indicates that Model 2, which includes the interaction term between average temperature and rain, provides a significantly better fit to the data compared to Model 1. In conclusion, this test strongly supports the presence of an interaction effect. Therefore, it will be added to the model. Since there is interaction, numeric variables will be centered at its mean for the purpose of interpretation. Because effect modification has been confirmed, confounder will not be evaluated.

Assessing the Assumptions for Poisson Distribution:

Fig 3.8 Overdispersion Tests for Queensboro Bridge

```
       Overdispersion test

data:  qb_full_poisson_model
z = 8.0441, p-value = 4.343e-16
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  127.1055
```

The overdispersion test on the Model 4 (Fig. 3.6) reveals significant evidence of overdispersion ($z = 8.0441$, $p < 0.01$). The estimated dispersion parameter is 127.1055, indicating that the variability in cyclist counts on the Queensboro Bridge does not satisfy the assumptions for a Poisson distribution. Therefore, the negative binomial regression is required.

Goodness of Fit of the Final Model:

Fig 3.9 GOF for Queensboro Bridge

```
Analysis of Deviance Table

Model: Negative Binomial(26.7105), link: log
Response: queensborobridge
Terms added sequentially (first to last)

                   Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                          213     633.42
month              6   88.118       207     545.31 < 2.2e-16 ***
weekend            1   95.466       206     449.84 < 2.2e-16 ***
average_temp       1   53.646       205     396.19 2.401e-13 ***
rain               1  169.809       204     226.39 < 2.2e-16 ***
average_temp:rain  1   10.471       203     215.91  0.001213 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance Table for the Poisson regression model on Queensboro Bridge ridership demonstrates the sequential inclusion of predictors. Average temperature ($p < 0.01$), rainfall ($p < 0.01$), monthly variations ($p < 0.01$), and interaction term each significantly contribute to explaining variability in ridership, making this model a good fit. The Poisson Deviance Goodness-of-fit ($p = 0.2543887$) provides additional evidence that the final model fits the data

Final Model:

Fig 3.10 Final Model for Queensboro Bridge

```
Call:
glm.nb(formula = queensborobridge ~ factor(month) + factor(weekend) +
    average_temp + as.numeric(rain) + average_temp * as.numeric(rain),
    data = centered_queensborobridge_data, init.theta = 26.71052056,
    link = log)
```

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       8.340899   0.042686 195.401  < 2e-16 ***
August            0.182421   0.059749   3.053 0.002265 **
July              0.052082   0.063149   0.825 0.409517
June              0.133452   0.058782   2.270 0.023191 *
May               0.117206   0.050696   2.312 0.020783 *
October           0.226185   0.051554   4.387 1.15e-05 ***
September         0.195243   0.056535   3.453 0.000553 ***
weekend          -0.272855   0.029434  -9.270  < 2e-16 ***
average_temp      0.012305   0.001975   6.232 4.62e-10 ***
rain             -0.418544   0.039411 -10.620  < 2e-16 ***
average_temp:rain 0.011667   0.003756   3.106 0.001894 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(26.7105) family taken to be 1)

    Null deviance: 633.42  on 213  degrees of freedom
Residual deviance: 215.91  on 203  degrees of freedom
AIC: 3514.1

Number of Fisher Scoring iterations: 1


              Theta:  26.71
          Std. Err.:  2.59

 2 x log-likelihood:  -3490.13
```

Results:

The negative binomial regression model was fitted to predict the counts of cyclists on the Queensboro Bridge. The model includes predictors such as the month of the year, whether it's a weekend, the average temperature, rain, and the interaction between average temperature and rain. Average temperature and rain were centered at its mean.

Centering average temperature and rain at their means, the intercept is 8.3409, representing the expected log count of bicycle trips when all other predictors are zero and centered at the mean, with the reference group being non-weekend days in April. The months May, June, July, August, September, October show positive coefficients, suggesting higher counts in these months compared to the reference month (April). The weekend has a negative coefficient of -0.2729, indicating that weekends are associated with lower bicycle counts. The coefficients for average temperature and rain are 0.0123 and -0.4185, indicating that higher temperatures and rainy conditions are associated with increased and decreased bicycle counts. All variables except the month of July are statistically significant at p-value is less than 0.05. The interaction term between average temperature and rain is positive (0.0117), suggesting that the effect of temperature on counts may vary under different rain conditions.

The dispersion parameter (Theta) is estimated to be 26.71, indicating substantial overdispersion. The model provides a good fit to the data, with a residual deviance of 215.91 on 203 degrees of freedom. Overall, the negative binomial model provides a more accurate understanding of the factors influencing Queensboro Bridge bicycle counts.

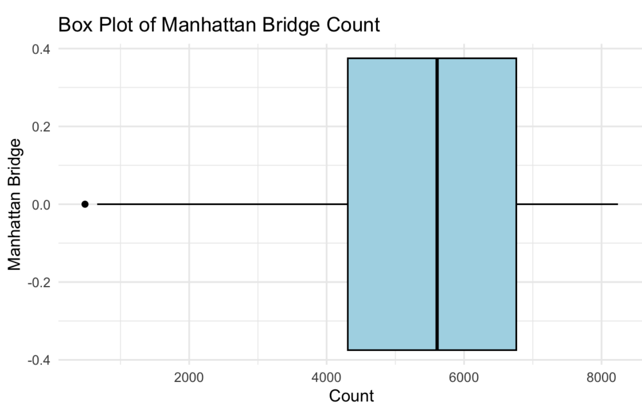
In conclusion, the number of cyclists on the Queensboro Bridge is influenced by various factors, including the month, whether it's a weekend, average temperature, rain, and the interaction between average temperature and rain. Rain and average temperature were centered at their means to enhance the interpretation of the interaction term.

Bridge 3: Manhattan Bridge

The Manhattan Bridge, which spans to the East River in New York City, is a popular route for commuters and cyclists. Connecting Manhattan and Brooklyn, the bridge not only serves as a key transportation link but provides a panoramic view of the city's skyline, making it a popular route for cyclists. Just like Brooklyn Bridge, Manhattan bridge also connects to the part of New York where most jobs are located. Manhattan Bridge is the second most popular bridge in New York.

Looking at Count Data:

Fig 4.0 Box Plot



There is an outlier present in this data set. This outlier can be caused by external factors such as potential road closures, holidays, bridge events, weather warnings and more.

Relationship Between Week Type and Number of Cyclists on Manhattan Bridge.

Fig 4.1 Mean, Total, Standard Deviation of Cyclist on Manhattan Bridge by Type of Week

weekend <dbl>	n <int>	mean <dbl>	sd <dbl>
0	152	5859.013	1699.743

1	62	4086.516	1109.780

For weekdays (weekend = 0), there are 152 observations with an average cyclist count of approximately 5,859.013 and a standard deviation of around 1,699.74 for the Manhattan Bridge. On the weekends (weekend = 1), there are 62 observations with an average cyclist count of about 4086.516 and a standard deviation of approximately 1,109.780. This suggests that, on average, fewer cyclists cross the Manhattan Bridge on weekends compared to weekdays, and the counts are generally less variable during weekends.

Relationship Between Months and Number of Cyclists on Manhattan Bridge.

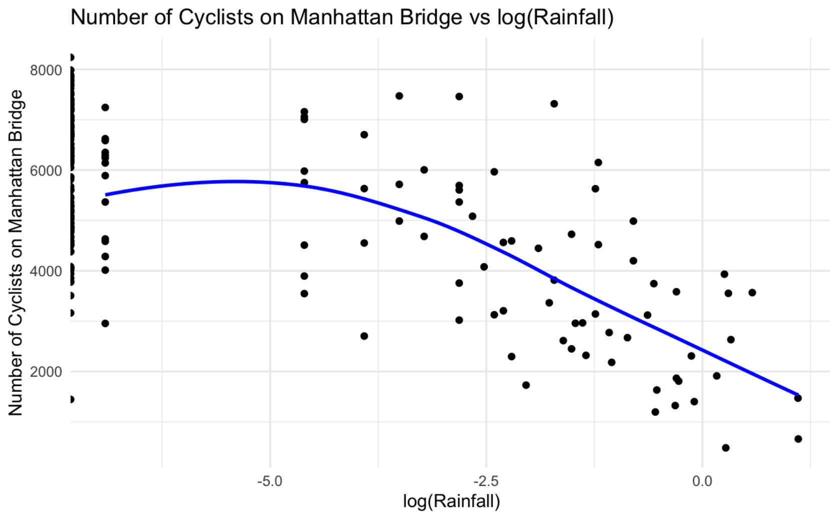
Fig. 4.2 Table of Average Monthly Counts of Cyclists on Manhattan Bridge

month <fctr>	n <int>	mean <dbl>	sd <dbl>
April	30	4353.900	1692.502
May	31	4959.226	2064.093
June	30	5780.433	1663.637
July	31	5424.613	1566.338
August	31	5730.484	1514.544
September	30	5871.033	1613.759
October	31	5297.710	1709.847

In June, the mean count reached its highest at 5,780.43, followed closely by September with 5,871.03. These months show higher counts compared to the other months, suggesting increased cycling activity during the warmer seasons. April and May show lower mean counts at 4,353.90 and 4,959.23, respectively.

Relationship Between Rain (log) and Number of Cyclists on Manhattan Bridge.

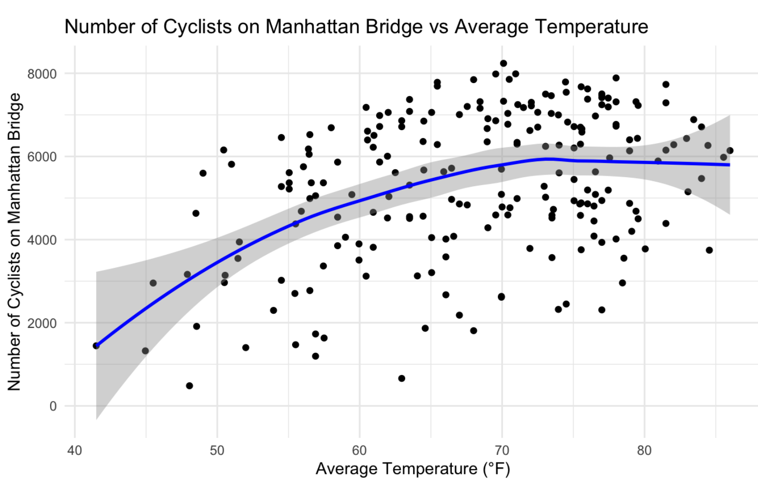
Fig 4.3 Scatter Plot of the Number of Cyclists on Manhattan Bridge vs. the Log of Rain



A log transformation was used on rainfall to visually see the patterns between rainfall and the number of cyclists on Brooklyn Bridge. As the inches of rainfall increases, the number of cyclists on Manhattan Bridge decreases.

Relationship between Temperature and Number of Cyclists on Manhattan Bridge.

Fig 4.4 Scatter Plot of the Number of Cyclists on Manhattan Bridge vs. Average Temperature



As the average temperature increases in Fahrenheit, the number of cyclists on Manhattan Bridge increases.

The Null Model:

Fig 4.5 Summary of the Null Model

```
Call:
glm(formula = manhattanbridge ~ 1, family = poisson(link = "log"),
    data = centered_manhattanbridge_data)
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 8.584008   0.000935    9181   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 139331  on 213  degrees of freedom
Residual deviance: 139331  on 213  degrees of freedom
AIC: 141546

Number of Fisher Scoring iterations:
```

The null model for the Manhattan Bridge has no predictors. The estimated intercept is 8.584008, with a standard error of 0.0009, resulting in a z-value of 9,181 and a highly significant p-value (p <0.01). The null deviance, representing the goodness of fit for the null model, is 13,9331 on 213 degrees of freedom. The Akaike Information Criterion (AIC) for this null model is 141,546. Further exploration of potential predictors will be examined further in the analysis.

Modeling Predictors:

Fig 4.6 ANOVA For the Four Models Analysis of Deviance Table

```
Model 1: manhattanbridge ~ 1
Model 2: manhattanbridge ~ average_temp + as.numeric(rain)
Model 3: manhattanbridge ~ average_temp + as.numeric(rain) + factor(month)
Model 4: manhattanbridge ~ average_temp + as.numeric(rain) + factor(month) +
    factor(weekend)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       213     139331
2       211      73891  2    65440 < 2.2e-16 ***
3       205      71226  6     2665 < 2.2e-16 ***
4       204      49048  1    22178 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The analysis of the deviance table for Manhattan Bridge reveals the impact of including additional predictors in the models. The null model (Model 1) serves as a baseline with only the intercept, resulting in a residual deviance of 13,9331. Model 2 introduces average temperature and rain as predictors, leading to a significant reduction in deviance (65,440) and a highly significant p-value, indicating that these weather-related variables contribute significantly to explaining the variability in bicycle counts.

The inclusion of months (Model 3) results in an additional reduction in deviance (2,665) compared to Model 2, indicating that the month of the year also plays a significant role in predicting bicycle counts on Manhattan Bridge. Finally, Model 4 includes the weekend variable, resulting in further reduction in deviance (22,178) and a highly significant p-value, highlighting the influence of weekends on bicycle counts. Model 4 will be considered, but further investigation of the underlying assumptions must be done

Assessing Interaction:

Fig 4.7 ANOVA to Analyze Interaction

```
Analysis of Deviance Table

Model 1: manhattanbridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain)
Model 2: manhattanbridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain) + average_temp * as.numeric(rain)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       204      49048
2       203      46481  1   2566.8 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of the Deviance table compares two models for predicting bicycle counts on Manhattan Bridge. Model 1 includes month, weekend, average temperature, and rain as main effects. Model 2 extends Model 1 by incorporating an interaction term between average temperature and rain. The comparison shows that Model 2 significantly improves the fit, with a deviance reduction of 2,566.8 and a highly significant p-value ($P < 0.01$).

The significant improvement in deviance and the low p-value indicate that this interaction term contributes significantly to the model's ability to explain the variability in bicycle counts on Manhattan Bridge. Therefore, Model 2, with the added interaction term, is preferred as it provides a better fit to the observed data. Since there is interaction, numeric variables will be centered at its mean for the purpose of interpretation. Because effect modification has been confirmed, confounder will not be evaluated.

Assessing the Assumptions of a Poisson Distribution :

Fig 4.8 Overdispersion Tests for Manhattan Bridge

```
       Overdispersion test

data:  mb_interaction_model
z = 5.8533, p-value = 2.41e-09
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  218.9714
```

The overdispersion test on the Model 4 (Fig. 4.6) reveals significant evidence of overdispersion ($z = 5.8533$, $p < 0.01$). The estimated dispersion parameter is 218.9714, indicating that the variability in cyclist counts on the Manhattan Bridge does not satisfy the assumptions for a Poisson distribution. Therefore, the negative binomial regression is required.

Goodness of Fit for Final Model:

Fig 4.9 GOF for Manhattan Bridge

```
Analysis of Deviance Table
Model: Negative Binomial(16.4546), link: log
Response: manhattanbridge
Terms added sequentially (first to last)

                   Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                           213     546.42
month               6   32.384       207     514.04 1.377e-05 ***
weekend             1   84.754       206     429.28 < 2.2e-16 ***
average_temp        1   43.113       205     386.17 5.167e-11 ***
rain                1  164.130       204     222.04 < 2.2e-16 ***
average_temp:rain   1    5.358       203     216.68   0.02062 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance Table for the Poisson regression model on Manhattan Bridge ridership demonstrates the sequential inclusion of predictors. Average temperature ($p < 0.01$), rainfall ($p < 0.01$), monthly variations ($p < 0.01$), and interaction term each significantly contribute to explaining variability in ridership, making this model a good fit. The Poisson Deviance Goodness-of-fit ($p = 0.2428504$) provides additional evidence that the final model fits the data.

Final Model:

Fig 4.10 Final Model for Manhattan Bridge

```
Call:
glm.nb(formula = manhattanbridge ~ factor(month) + factor(weekend) +
    average_temp + as.numeric(rain) + average_temp * as.numeric(rain),
    data = centered_manhattanbridge_data, init.theta = 16.45457534,
    link = log)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       8.586228   0.054290 158.156  < 2e-16 ***
August            0.032358   0.076011   0.426   0.6703
July             -0.019426   0.080330  -0.242   0.8089
June              0.090065   0.074774   1.204   0.2284
May               0.075805   0.064473   1.176   0.2397
October           0.094629   0.065580   1.443   0.1490
September         0.093956   0.071918   1.306   0.1914
weekend          -0.333854   0.037442  -8.917  < 2e-16 ***
average_temp      0.014117   0.002512   5.620 1.91e-08 ***
rain             -0.547218   0.050117 -10.919  < 2e-16 ***
average_temp:rain 0.010468   0.004773   2.193   0.0283 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(16.4546) family taken to be 1)

    Null deviance: 546.42  on 213  degrees of freedom
```

```
Residual deviance: 216.68  on 203  degrees of freedom
AIC: 3678.2

Number of Fisher Scoring iterations: 1


              Theta:  16.45
          Std. Err.:  1.58

 2 x log-likelihood:  -3654.23
```

Results:

The negative binomial regression model was fitted to predict the counts of cyclists on the Manhattans bridge. The model includes predictors such as the month of the year, whether it's a weekend, the average temperature , rain, and the interaction between average temperature and rain. Average temperature and rain were centered at its mean.

The model includes predictors such as the month of the year, whether it's a weekend, centered average temperature, rain, and the interaction between average temperature and rain. Centering average temperature and rain at their means, the intercept is set at 8.5862, representing the baseline count of bike crossings when all other predictors are zero, with the reference group being non-weekend days in April. While the coefficients for each month lack statistical significance, the weekend indicator (weekend =1) shows a significant negative effect, indicating fewer bike crossings on weekends than weekdays ($p < 0.001$).

Average temperature positively influences bike crossings, with a coefficient of 0.0141, meaning that for each one-unit increase in temperature, the log of expected crossings increases by 0.0141. Rain has a negative impact, as indicated by the coefficient of -0.5472. The interaction term between average temperature and rain is statistically significant ($p = 0.0283$), suggesting that the effect of temperature on bike crossings is influenced by rain.

The dispersion parameter (Theta) is estimated at 16.45, revealing overdispersion in the data, indicating higher variability than expected under a Poisson distribution. Despite this, the model fits well, reflected in the AIC value of 3,678.2, significantly improving fit compared to the null deviance (Fig 4.5).

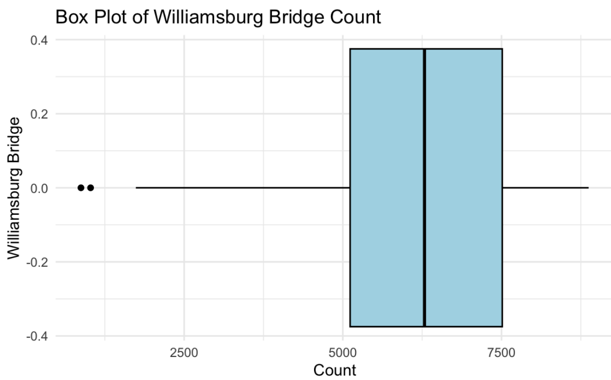
In conclusion, the number of cyclists on the Manhattan Bridge is influenced by various factors, including the month, whether it's a weekend, average temperature, rain, and the interaction between average temperature and rain. Rain and average temperature were centered at their means to enhance the interpretation of the interaction term.

Bridge 4: Williamsburg Bridge

The Williamsburg Bridge connects the neighborhoods of Brooklyn and Manhattan. The bridge provides a pathway for commuters, cyclists, and pedestrians. It is in between Queensboro Bridge and Manhattan Bridge. Just like Brooklyn and Manhattan Bridge, the Williamsburg Bridge also connects to Lower Manhattan (where most jobs are located).

Looking at Count Data:

Fig 5.0 Box Plot



There are a few outliers in this data set. The outliers can be caused by external factors such as potential road closures, holidays, bridge events, weather warnings and more.

Relationship Between Week Type and Number of Cyclists on Williamsburg Bridge.

Fig 5.1 Mean, Total, Standard Deviation of Cyclist on Williamsburg Bridge by Type of Week

weekend <dbl>	n <int>	mean <dbl>	sd <dbl>
0	152	6534.559	1725.141
1	62	4867.774	1179.876

For weekdays (weekend = 0), there are 152 observations with an average cyclist count of approximately 6,534.56 and a standard deviation of around 1,725.141 for the Williamsburg Bridge. On the weekends (weekend = 1), there are 62 observations with an average cyclist count of about 4,867.774 and a standard deviation of approximately 1,179.876. This suggests that, on average, fewer cyclists cross the Williamsburg Bridge on weekends compared to weekdays, and the counts are generally less variable during weekends

Relationship Between Months and Number of Cyclists on Williamsburg Bridge.

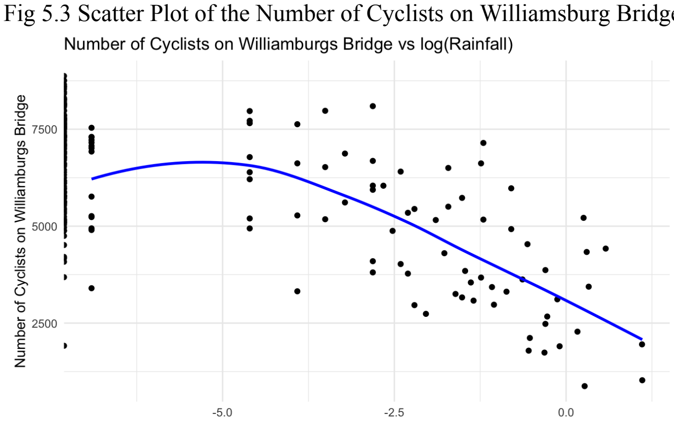
Fig. 5.2 Table of Average Monthly Counts of Cyclists on Williamsburg Bridge

month <fctr>	n <int>	mean <dbl>	sd <dbl>
April	30	4942.267	1733.685
May	31	5710.161	2007.357
June	30	6620.000	1598.698
July	31	6073.677	1575.697
August	31	6406.806	1489.961
September	30	6614.333	1641.935
October	31	5995.065	1746.135

In April, the mean count of cyclists was 4,942.27, with a standard deviation of 1,733.69. There was an increase in May with a mean of 5,710.16 and a standard deviation of 2,007.36. June exhibits the highest average daily bike crossings among the observed months, with a mean of 6,620.00 and a standard deviation of 1,598.70. July experiences a slight dip with a mean of 6,073.68, followed by August (6,406.81) and September (6,614.33) with relatively stable figures. October sees a decrease with a mean of 5,995.07.

Relationship Between Rain (log) and Number of Cyclists on Williamsburg Bridge.

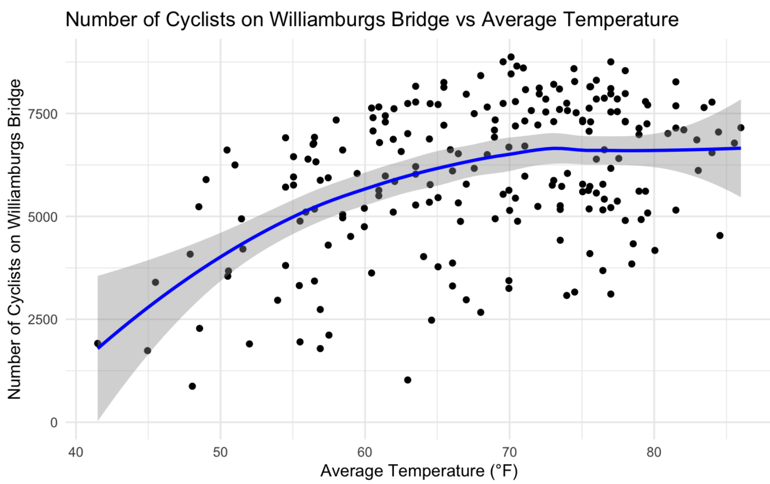
Fig 5.3 Scatter Plot of the Number of Cyclists on Williamsburg Bridge vs. the Log of Rain



A log transformation was used on rainfall to visually see the patterns between rainfall and the number of cyclists on Williamsburg Bridge. As the inches of rainfall increases, the number of cyclists on Williamsburg Bridge decreases.

Relationship between Temperature and Number of Cyclists on Williamsburg Bridge.

Fig 5.4 Scatter Plot of the Number of Cyclists on Williamsburg Bridge vs. Average Temperature



As the average temperature increases in Fahrenheit, the number of cyclists on Williamsburg Bridge increases.

The Null Model:

Fig 5.5 Summary of the Null Model

```
Call:
glm(formula = williamsburgbridge ~ 1, family = poisson(link = "log"),
    data = centered_williamsburgbridge_data)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) 8.7080877  0.0008787    9910   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 123257  on 213  degrees of freedom
Residual deviance: 123257  on 213  degrees of freedom
AIC: 125503

Number of Fisher Scoring iterations: 4
```

The null model for the Williamsburg Bridge has no predictors. The estimated intercept is 8.7080877, with a standard error of 0.0009, resulting in a z-value of 9,910 and a highly significant p-value ($p < 0.01$). The null deviance, representing the goodness of fit for the null model, is 12,3257 on 213 degrees of freedom. The Akaike Information Criterion (AIC) for this null model is 12,5503. Further exploration of potential predictors will be examined further in the analysis.

Modeling Predictors:

Fig 5.6 ANOVA For the Four Models Analysis of Deviance Table

```
Analysis of Deviance Table

Model 1: williamsburgbridge ~ 1
Model 2: williamsburgbridge ~ average_temp + as.numeric(rain)
Model 3: williamsburgbridge ~ average_temp + as.numeric(rain) + factor(month)
Model 4: williamsburgbridge ~ average_temp + as.numeric(rain) + factor(month) +
    factor(weekend)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       213     123257
2       211      64038  2    59220 < 2.2e-16 ***
3       205      60339  6     3699 < 2.2e-16 ***
4       204      43805  1    16534 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 1, with only the intercept, serves as the baseline, and subsequent models introduce additional predictors. As each model incorporates more predictors, the deviance decreases, indicating improved model fit. The p-values associated with the chi-square test assess the significance of the additional predictors. In this case, all the models with predictors are highly significant ($p < 0.001$), suggesting that including temperature, rain, month, and weekend information significantly improves the models compared to the baseline. The decreasing deviance and significant p-values support the added predictors in explaining the variability in Williamsburg Bridge bike crossings. Model 4 (Fig 5.6) will be considered, but further investigation of the underlying assumptions must be done.

Assessing Interaction:

Fig 5.7 ANOVA to Analyze Interaction

```
Analysis of Deviance Table

Model 1: williamsburgbridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain)
Model 2: williamsburgbridge ~ factor(month) + factor(weekend) + average_temp +
    as.numeric(rain) + average_temp * as.numeric(rain)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       204      43805
2       203      41193  1   2611.8 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance table compares two models predicting daily bike crossings on the Williamsburg Bridge. Model 1 includes month, weekend, average temperature, and rain, while

Model 2 incorporates an interaction term between average temperature and rain. The chi-square test assesses the significance of the added interaction term. Model 2 has a significantly lower deviance than Model 1 ($p < 0.001$), indicating that the interaction term improves the model fit. Therefore, the interaction between average temperature and rain is a valuable addition in explaining the variability in Williamsburg Bridge bike crossings. Since there is interaction, numeric variables will be centered at its mean for the purpose of interpretation. Because effect modification has been confirmed, confounder will not be evaluated.

Assessing the Assumptions of a Poisson Distribution :

Fig 5.8 Overdispersion Tests for Williamsburg Bridge

```
Overdispersion test

data:  wb_interaction_model
z = 6.3159, p-value = 1.343e-10
alternative hypothesis: true dispersion is greater than 1
sample estimates:
dispersion
  189.7187
```

The overdispersion test on the Model 4 reveals significant evidence of overdispersion ($z = 6.3159$, $p < 0.01$). The estimated dispersion parameter is 189.7187, indicating that the variability in cyclist counts on the Brooklyn Bridge does not satisfy the assumptions for a Poisson distribution. Therefore, the negative binomial regression is required.

Goodness of Fit for Final Model:

Fig 5.9 GOF for Williamsburg Bridge

```
Analysis of Deviance Table
Model: Negative Binomial(22.6132), link: log
Response: williamsburgbridge
Terms added sequentially (first to last)

                   Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                                 213     566.08
month              6   41.745       207     524.34 2.064e-07 ***
weekend            1   76.868       206     447.47 < 2.2e-16 ***
average_temp       1   51.747       205     395.72 6.312e-13 ***
rain               1  171.615       204     224.11 < 2.2e-16 ***
average_temp:rain  1    8.086       203     216.02  0.004461 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Analysis of Deviance Table for the Poisson regression model on Williamsburg Bridge ridership demonstrates the sequential inclusion of predictors. Average temperature ($p < 0.01$), rainfall ($p < 0.01$), monthly variations ($p < 0.01$), and interaction term each significantly contribute to explaining variability in ridership, making this model a good fit. The Poisson Deviance Goodness-of-fit ($p = 0.2527467$) provides additional evidence that the final model fits the data.

Final Model:

Fig 5.10 Final Model for Williamsburg Bridge

```
Call:
glm.nb(formula = williamsburgbridge ~ factor(month) + factor(weekend) +
    average_temp + as.numeric(rain) + average_temp * as.numeric(rain),
    data = centered_williamsburgbridge_data, init.theta = 22.61322593,
    link = log)

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        8.693054   0.046327 187.646  < 2e-16 ***
August             0.033580   0.064861   0.518  0.60466
July              -0.022322   0.068547  -0.326  0.74469
June               0.111304   0.063804   1.744  0.08108 .
May                0.106213   0.055015   1.931  0.05353 .
October            0.099706   0.055960   1.782  0.07479 .
September          0.101551   0.061368   1.655  0.09797 .
weekend            -0.267363   0.031948  -8.369  < 2e-16 ***
average_temp       0.013237   0.002144   6.175 6.62e-10 ***
rain              -0.463885   0.042748 -10.852  < 2e-16 ***
average_temp:rain 0.011032   0.004072   2.709  0.00675 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(22.6132) family taken to be 1)

    Null deviance: 566.08  on 213  degrees of freedom
Residual deviance: 216.02  on 203  degrees of freedom
AIC: 3670.4

Number of Fisher Scoring iterations: 1


              Theta:  22.61
          Std. Err.:  2.18

 2 x log-likelihood:  -3646.364
```

Results:

The negative binomial regression model was fitted to predict the counts of cyclists on the Williamsburg Bridge. The model includes predictors such as the month of the year, whether it's a weekend, the average temperature , rain, and the interaction between average temperature and rain. Average temperature and rain were centered at its mean.

The intercept is estimated at 8.6931, indicating the expected log count of cyclists when all other predictors are zero and centered at their means, with the reference group being non-weekend days in April. While month-specific effects are not statistically significant, the weekend variable exhibits a significant negative effect, suggesting a decrease in expected log counts by 0.2674 on weekends compared to weekdays ($p < 0.001$). Average temperature demonstrates a significant positive effect, with a 0.0132 increase in the expected log count for each one-unit temperature rise ($p < 0.001$). Rain has a significant negative effect, with a 0.4639 decrease in the expected log count for each additional unit of rain ($p < 0.001$).

The interaction term between average temperature and rain is also significant ($p = 0.00675$), indicating a modification in the effect of temperature on log counts based on rain levels. The model's dispersion parameter (Theta) is estimated at 22.61, signifying overdispersion, and the goodness of fit is supported by the residual deviance of 216.02 on 203 degrees of freedom, yielding an AIC of 3,670.4.

In conclusion, the number of cyclists on the Williamsburg Bridge is influenced by various factors, including the month, whether it's a weekend, average temperature, rain, and the interaction between average temperature and rain. Rain and average temperature were centered at their means to enhance the interpretation of the interaction term.

Conclusion:

Equation:

Any four of the bridges~Negative Binomial(θ) =

$$\beta0 + \beta1 * (august) + \beta2 * (july) + \beta3 * (june) + \beta4 * (may) + \beta5 * (october) + \beta6 * (september) + \beta7 * (weekend) + \beta8 * (average\ rain\ centered) + \beta9 * (average\ temperature\ centered) + ((\beta8 * (average\ rain\ centered)) ** \beta9 * (average\ temperature\ centered)))$$

Brooklyn Bridge:

- The model revealed that weekend days are associated with a significant decrease in ridership (Coefficient: -0.1206, p-value: 0.00485).
- Temperature had a positive effect on ridership (Coefficient: 0.0148, p-value: < 0.001), while rain had a negative impact (Coefficient: -0.6301, p-value: < 0.001).
- The interaction between average temperature and rain was also statistically significant (Coefficient: 0.0163, p-value: 0.00285).

Queensboro Bridge:

- The model revealed that weekend days are associated with a significant decrease in ridership (Coefficient: -0.2729, p-value: < 0.001).
- Temperature had a positive effect on ridership (Coefficient: 0.0123, p-value: < 0.001) and rain (Coefficient: -0.4185, p-value: < 0.001) while rain had a negative effect.

- The interaction term, average temperature and rain, showed statistical significance (Coefficient: 0.0117, p-value: 0.00189).

Manhattan Bridge:

- The model revealed that weekend days are associated with a significant decrease in ridership (Coefficient: -0.3339, p-value: < 0.001).
- Temperature had a positive effect on ridership(Coefficient: 0.0141, p-value: < 0.001), while rain had a negative effect (Coefficient: -0.5472, p-value: < 0.001).
- The interaction term between average temperature and rain was also statistically significant (Coefficient: 0.0105, p-value: 0.0283).

Williamsburg Bridge:

- The model revealed that weekend days are associated with a significant decrease in ridership (Coefficient: -0.2674, p-value: < 0.001).
- Temperature had a positive effect on ridership (Coefficient: 0.0132, p-value: <0.001) and rain (Coefficient: -0.4639, p-value: < 0.001) while rain had a negative effect.
- The interaction term, average temperature and rain, showed statistical significance (Coefficient: 0.0110, p-value: 0.00675).

Overall Comparisons:

- Weekend days consistently showed a negative influence on ridership across all bridges.
- Temperature had a positive impact on ridership for all bridges, emphasizing its role in encouraging cycling behavior.
- Rain consistently had a negative impact on ridership, indicating that adverse weather conditions discourage cycling.
- The interaction term between average temperature and rain was statistically significant for all bridges.

The mean ridership (Fig 1.2) for the four major East River bridges, from lowest to highest, is as follows: Brooklyn Bridge (2,680), Queensboro Bridge (4,550), Manhattan Bridge (5,350), and Williamsburg Bridge (6,050). The lower mean ridership may be associated with the bridge's popularity and the challenges cyclists face during high foot traffic periods. Weekends consistently show a negative impact on ridership across all bridges, potentially attributed to increased congestion. Increasing temperature positively affects ridership on all bridges. Conversely, rain consistently shows a negative influence, suggesting that adverse weather conditions discourage cycling across the bridges. The statistically significant interaction term between average temperature and rain emphasizes the relationship between these weather variables.

Further Considerations:

To increase model accuracy, Multivariable Polynomial Functions (MPF) can be used. Experimenting with MPF, especially for variables like average temperature and rainfall, may lead to improved predictive accuracy. The decision to not include MFP was due to its lack of interoperability for the predictor variables on its outcome. However, the code for MFP can be found in the R Quarto notebook.

Additionally, assessing whether high foot traffic areas correlate with decrease bike ridership can provide more information about the interconnected dynamics of how the bridge is being used.