Analysis of bicycle rentals in Washington D.C. and Seoul

Student 200019804

2023-10-17

Introduction

In this project we investigate bicycle rentals in two cities, Seoul, South Korea, and Washington D.C., USA by examining locally collected data sets and through the use of statistical modelling.

Climate

To begin to understand bike rental numbers in the two cities, it is important to understand each city's climate.

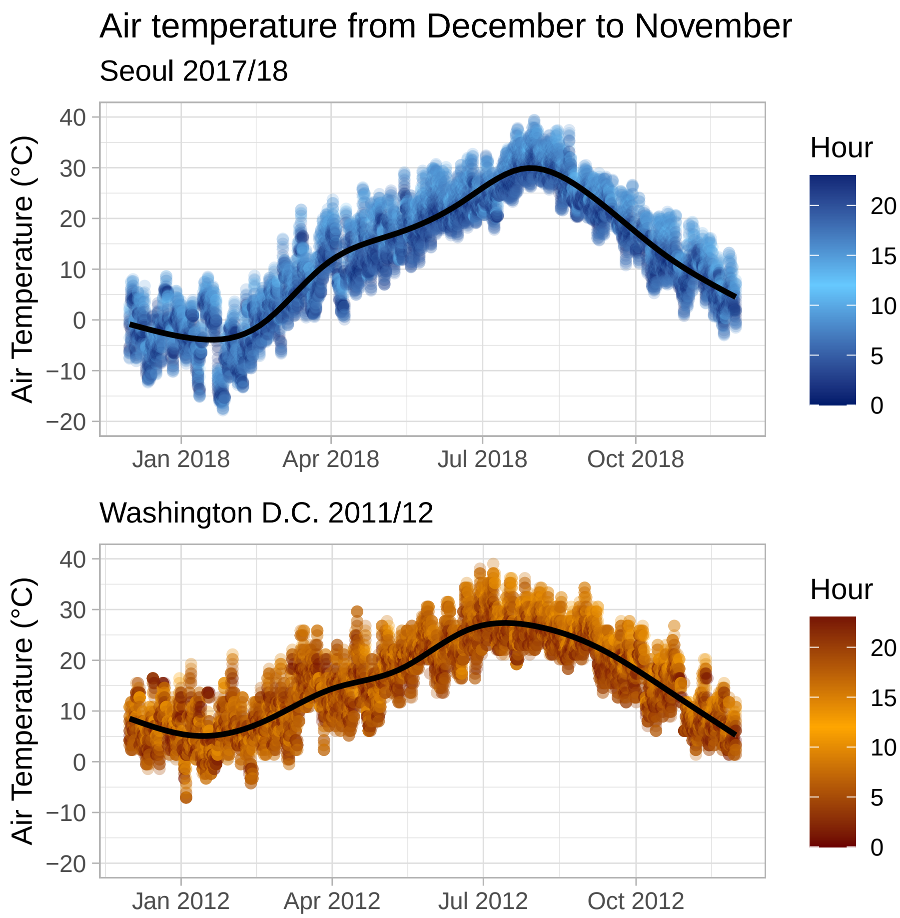

Figure 1: Air Temperature in Seoul and Washington D.C.

Figure 1 shows the air temperature (°C) throughout a year in Washington D.C. and Seoul. The temperature peaks between July and August in both cities with Seoul experiancing a shorter, more pronounced period of peak temperature. Visibly the climate in Seoul is more extreme than in Washington D.C. in the chosen year, with a hotter summer and colder winter.

Table 1: Temperature (°C)

	Max	Min	Mean
Seoul	39.4	-17.8	12.8
Washington D.C.	39.0	-7.1	15.4

Seoul experiences a slightly higher maximum temperature and a much colder minimum temperature, resulting in Washington D.C. having the higher mean temperature.

Bicycle Rentals

Figure 3 shows the impact of Seoul's colder winter on bike rentals. The bike rentals in winter are noticeably lower than those in Washington D.C. when compared to the other seasons.

Figure 3 also shows that the bike rentals in each season are related to the temperature with the highest average rentals in summer, the hottest month for both cities, and the lowest in winter. The bike rentals vary less in Washington D.C. than in Seoul, this is likely due to Washington's less extreme climate exhibited in Figure 1.

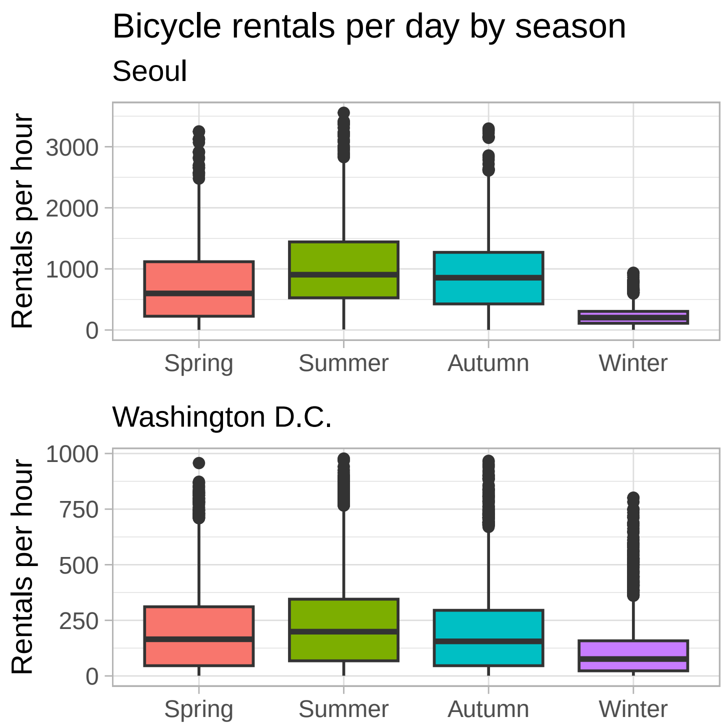

Figure 2: Bike rentals in Seoul and Washington D.C.

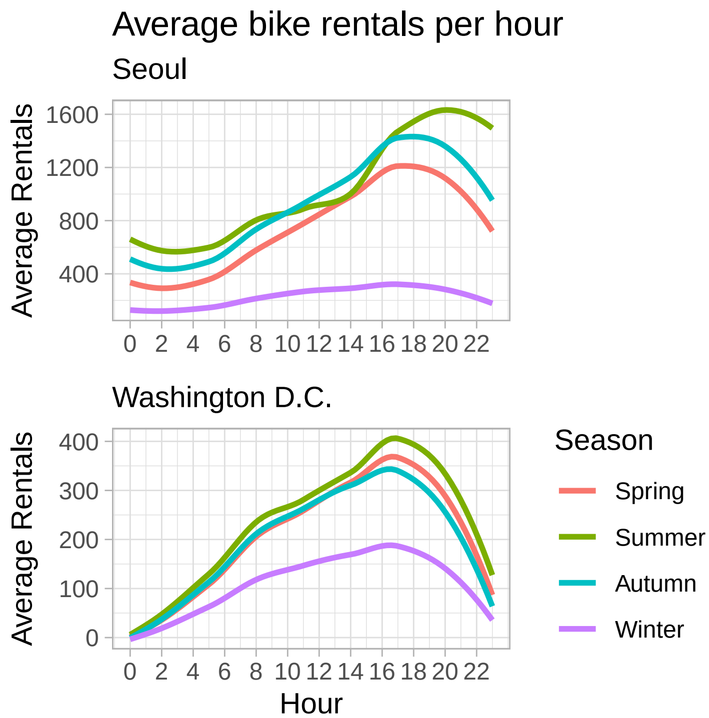

Figure 3: Placeholder

Figure 3 shows the relationship between time of day and bike rentals in each season in the two cities. In Washington D.C. the average bike rentals peak between 16:00 and 17:00 year round, with summer peaking the latest.

The average bike rentals per hour in Seoul vary more between seasons than Washington D.C.. Firstly the the average bike rentals in Seoul peak later in the day than in Washington D.c., most noticeably in summer and autumn peaking at 20:00 and 18:00 respectively. Visibly the the average bike rentals in Seoul stay at their maximum values for longer than in D.C..

Table 2: Holiday Bike Rentals in Seoul

Holiday	Count
Yes	529.2
No	739.3

Table X and Y show the impact of holidays on bike rentals in Seoul and Washington D.c. respectively. On average, the bike rentals in Seoul on a holiday are 28% lower than a non-holiday. This is likely due to the lack of commuters renting bikes.

Similarly, the average bike rentals in Washington D.C. on a holiday are 17% lower than a non-holiday.

The difference between bike rentals on holidays in the two cities suggests a cultural difference in the way bicycles are used for leisure and work.

Table 3: Holiday Bike Rentals in Washington D.C.

Holiday	Count
Yes	156.9
No	190.4

Meteorological effects

Next we can analyse the effect of different meteorological factors on the number of bike rentals in the two cities.

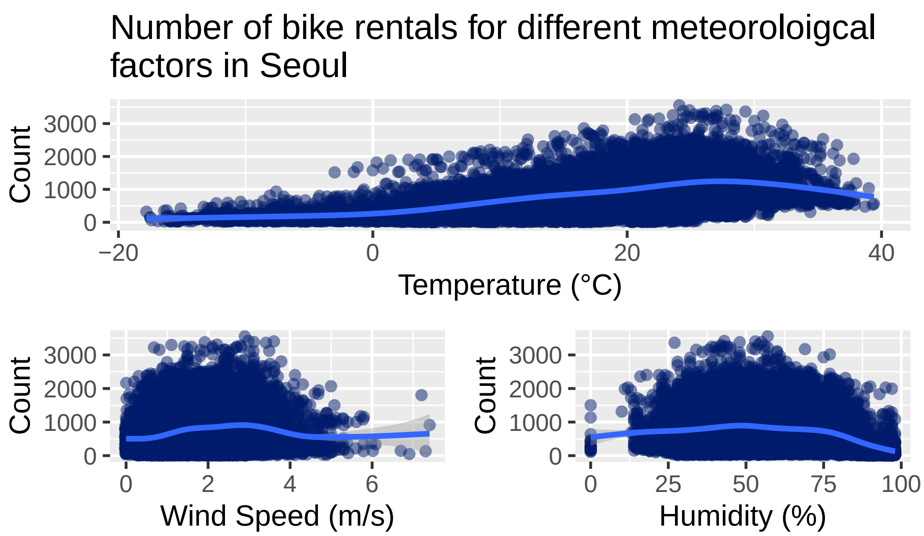

Figure 4: Seoul bike rentals

Figure 5 shows the impact of air temperature, humidity, and wind speed on bicycle rentals in Seoul. Bike rentals increase with air temperature up to ~30°C before starting to decrease, suggesting people in Seoul are more likely to rent a bicycle on warm days but not on the hottest days. This is expected as cycling in low temperatures presents dangers like ice as well as exposure to the cold temperature itself.

The impact of wind speed appears to be lesser, only showing a small decrease for speeds above 4m/s, showing that only high wind speeds have any meaningful effect on the number of bike rentals, due to high winds presenting danger for cyclists.

The impact of the three meteorological factors is similar in Washington D.C., shown in Figure 6. One noticeable difference is that the bike rentals in Washington D.C. do not see such a decline at the higher temperatures close to 40°C. The decline in bike rentals in Washington D.C. with increasing humidity is more pronounced than in Seoul, starting at ~25% compared to ~50% in Seoul.

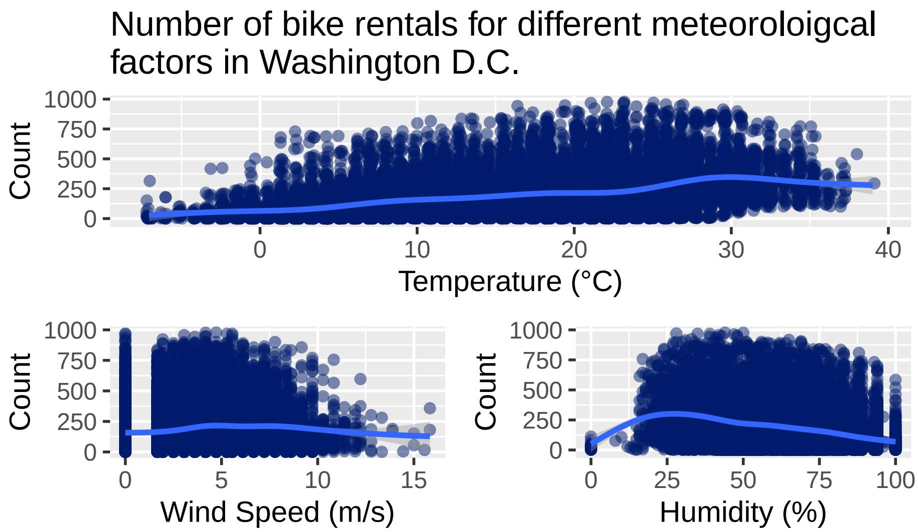

Figure 5: Washington D.C. bike rentals

Modelling

In order to model and predict bike rentals in Seoul and Washington D.C. we apply linear regression to the logarithm of the number of bicycle rentals per hour, $log(Count)$, with season, air temperature, humidity, and wind speed as independent variables. The model can be represented as

$$log(Count) \sim Season + Temperature + Humidity + WindSpeed.$$

Checking Model Assumptions

To assess the effectiveness, reliability, and validity of the models, it is important to check the assumptions about both sets of data. Firstly we check the assumption of normality of residuals for both models.

Figures 7 and 8 show Normal Quantile-Quantile plots for the residuals of the models for Seoul and Washington D.C. respectively.

The data for Seoul, shown above in Figure 6, exhibits a slight left-skew for its tails however, the main body of the data lies on the reference line, indicating the assumption of normality of residuals is fair.

The data for Washington D.C.,Figure 7, is light tailed although the upper part of the data lies well on the reference line, again suggesting normality of residuals is a fair assumption.

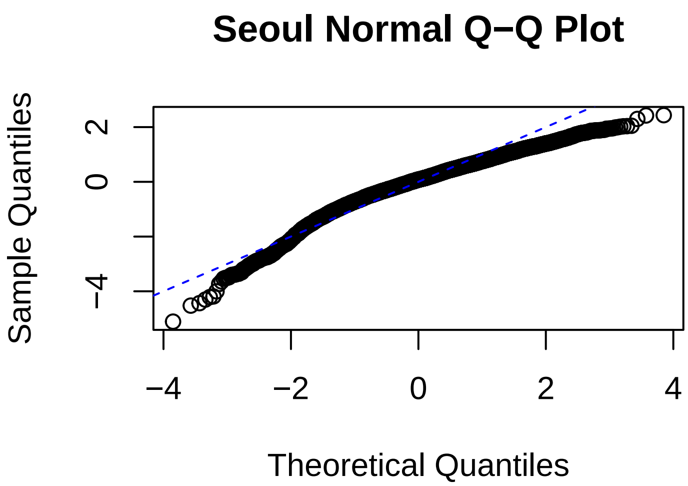

Figure 6: Seoul normal Q-Q plot

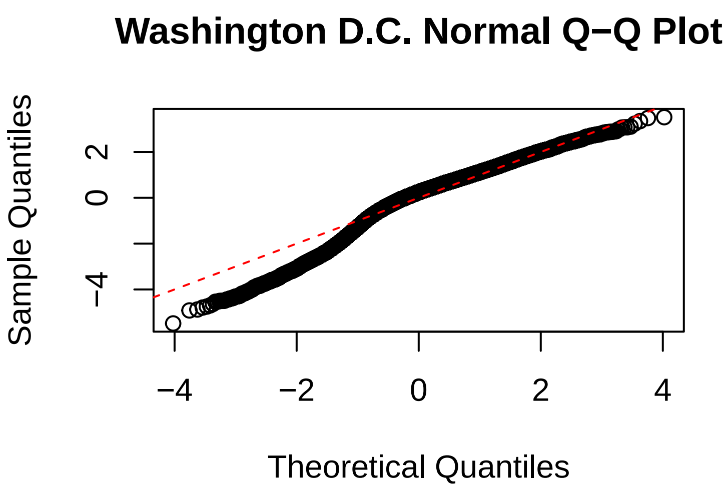

Figure 7: Washington D.C. normal Q-Q plot

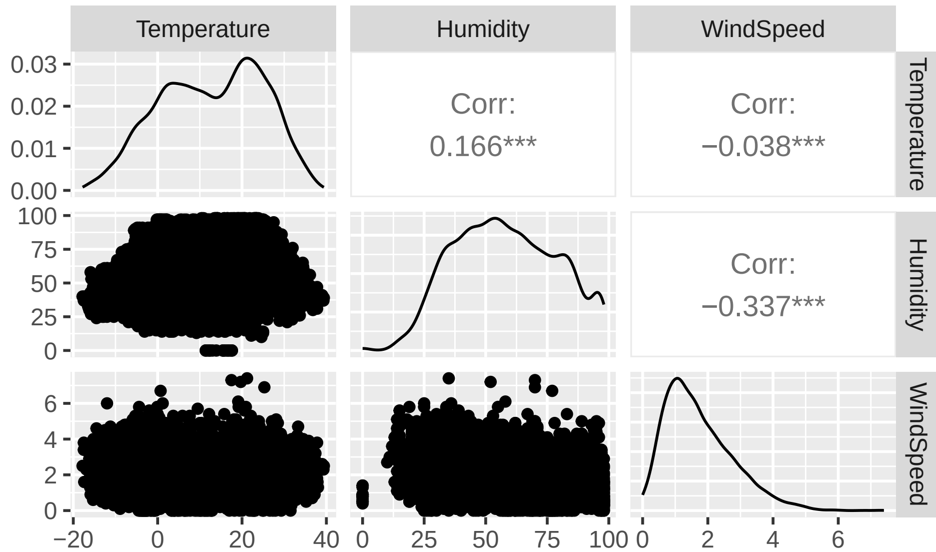

Figure 8: Seoul scatterplot matrix

Figure 8 shows that the three meteorological factors selected as predictors exhibit strong multicollinearity, meaning the assumption of independence between variables isn't shown in the data, which may negatively effect the accuracy and validity of the model for Seoul.

The data for Washington D.C., shown in Figure 9, suffers from the same issue.

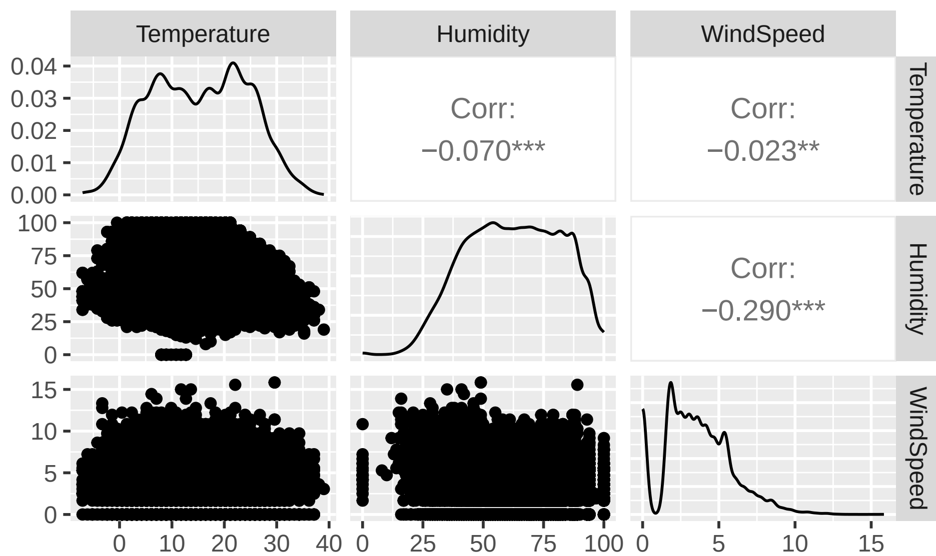

Figure 9: Washington D.C. scatterplot matrix

Model Analysis

```
##
## Call:
## lm(formula = log(Count) ~ Season + Temperature + Humidity + WindSpeed,
##     data = seoul3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.1073 -0.4281  0.0812  0.5493  2.4352
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.7336965  0.0467062 144.171  < 2e-16 ***
## SeasonSummer  0.0036038  0.0327843   0.110  0.91247
## SeasonAutumn  0.3733211  0.0261578  14.272  < 2e-16 ***
## SeasonWinter -0.3830362  0.0349918 -10.946  < 2e-16 ***
## Temperature   0.0492700  0.0015053  32.732  < 2e-16 ***
## Humidity     -0.0224974  0.0004844 -46.441  < 2e-16 ***
## WindSpeed     0.0253809  0.0093544   2.713  0.00668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8276 on 8458 degrees of freedom
## Multiple R-squared:  0.4941, Adjusted R-squared:  0.4937
## F-statistic:  1377 on 6 and 8458 DF,  p-value: < 2.2e-16
```

For the Seoul model, all independent variables are statistically significant at the 1% level, other than than summer not being significantly different to the intercept, spring. A multiple R-squared value of 0.49 indicates that the model is not a very good fit as there is still much variation in bike rentals not explained by the predictors. This may be due to the string multicollinearity between the predictors, and important factors such as time of day and holiday status being left out of the model.

```
##
## Call:
## lm(formula = log(Count) ~ Season + Temperature + Humidity + WindSpeed,
##     data = washington3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4834 -0.6069  0.2458  0.8440  3.5203
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.6264010  0.0576892  80.195  < 2e-16 ***
## SeasonSummer -0.3651680  0.0300276 -12.161  < 2e-16 ***
## SeasonAutumn  0.5361839  0.0289332  18.532  < 2e-16 ***
## SeasonWinter  0.1046103  0.0341346   3.065  0.00218 **
## Temperature   0.0797914  0.0017401  45.856  < 2e-16 ***
## Humidity     -0.0233425  0.0005317 -43.901  < 2e-16 ***
## WindSpeed     0.0245022  0.0044358   5.524 3.37e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 1.263 on 17372 degrees of freedom
## Multiple R-squared:  0.278,  Adjusted R-squared:  0.2777
## F-statistic:  1115 on 6 and 17372 DF,  p-value: < 2.2e-16
```

For the Washington D.C. model, all predictors are statistically significant at the 1% level. The low multiple R-Squared value of 0.28 suggests the model is not a good fit for the data, for similar reasons as the Seoul model.

Coefficients

Table 4: Seoul model parameters 97% CI

	1.5 %	98.5 %
Intercept	6.6323227	6.8350703
Summer	-0.0675531	0.0747607
Autumn	0.3165466	0.4300955
Winter	-0.4589844	-0.3070880
Temperature	0.0460029	0.0525372
Humidity	-0.0235488	-0.0214459
Wind Speed	0.0050777	0.0456842

The Seoul model has winter (against intercept season spring) and humidity as its only (>97%) negative coefficients; factors that decrease expected rentals when present. This is supported by the data shown in Figure 3 and Figure 5.

Table 5: Washington D.C. model parameters 97% CI

	1.5 %	98.5 %
Intercept	4.5012000	4.7516020
Summer	-0.4303359	-0.3000002
Autumn	0.4733911	0.5989767
Winter	0.0305290	0.1786916
Temperature	0.0760151	0.0835678
Humidity	-0.0244964	-0.0221885
Wind Speed	0.0148754	0.0341290

The D.C. model has (>97%) negative coefficients summer and humidity. This is unexpected as summer is the season with most bike rentals, as seen in Figure 3. This may be because the Washinton D.C. model is a worse fit for the data than the Seoul model. The humidity coefficient being negative is supported by Figure 5.

Prediction

Using the models for Seoul and Washington, we can make predictions for the average number of bicycle rentals per hour on a unspecified day in a given season with given temperature, humidity, and wind speed. We predict for freezing (0°C) temperature, with 0.5m/s wind speed, humidity of 20%, and in winter.

Table 6: Bike rental prediction

	Prediction	5%	95%
Seoul	370	352.8	387.9
Washington D.C.	72	67.7	76.5

The model predicts 370 bike rentals per hour in Seoul and 72 in Washington D.C..