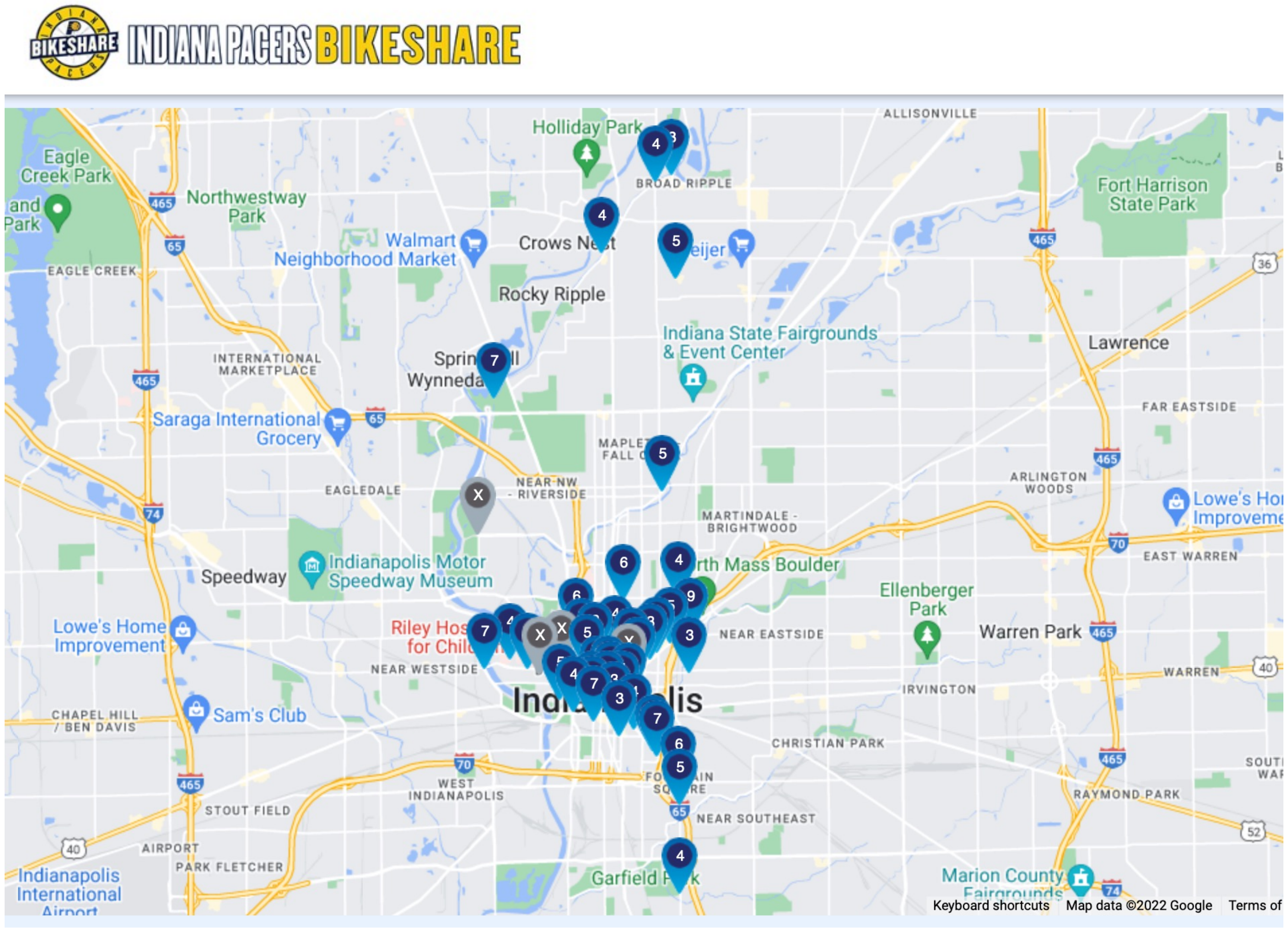
Negative Binomial Modeling on Seoul Bike Sharing Demand

Minmin Pan

Mark Gottermeier

Yao Chen

Introduction



Introduction

- Benefits of Bike sharing system
  - Convenient first/last mile link to other transportations
  - Pleasant city sightseeing approach
  - Boosting public health
  - Reduce greenhouse gas emission
  - ……

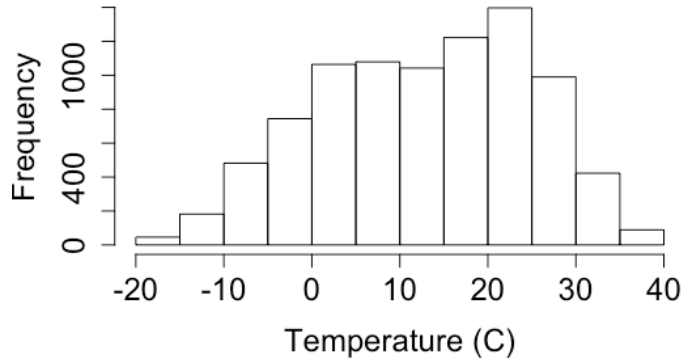
  -> Worldwide rising, even from urban to rural

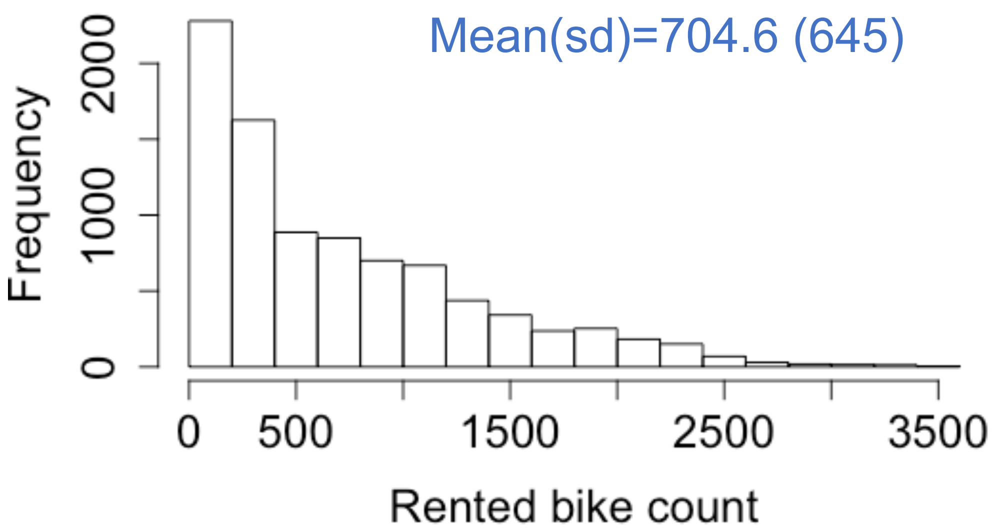
Introduction

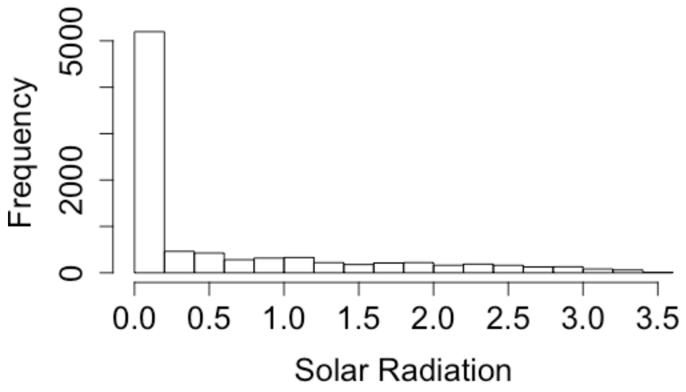
- Need to predict the demand of bikes in advance
  - A model between the influencing factors and the demand of bikes

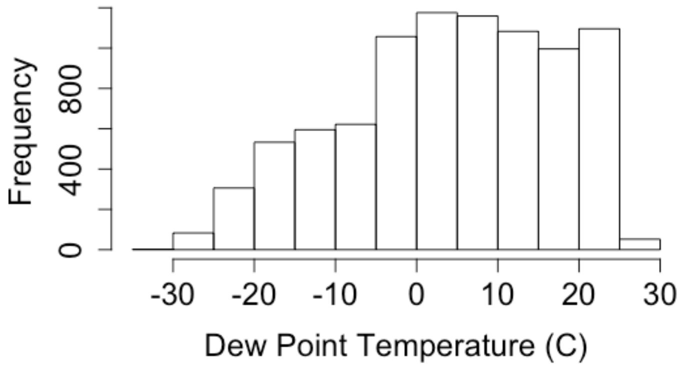
- Our work
  - Dataset: hourly Seoul bike share demand data for a whole year
  - Factors: weather(temperature, humidity etc.), hour of the day, holiday
  - GLM tools: utilizing negative binomial regression, identifying significant factors and interactions, evaluating the model

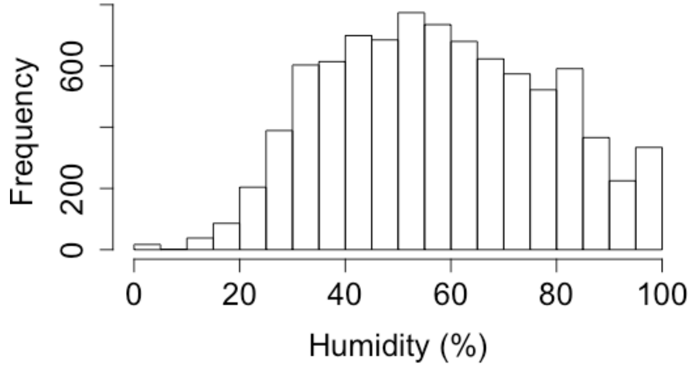
Characteristics



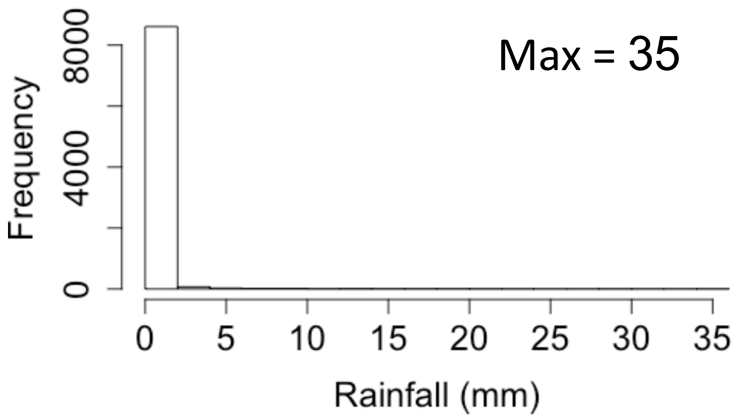


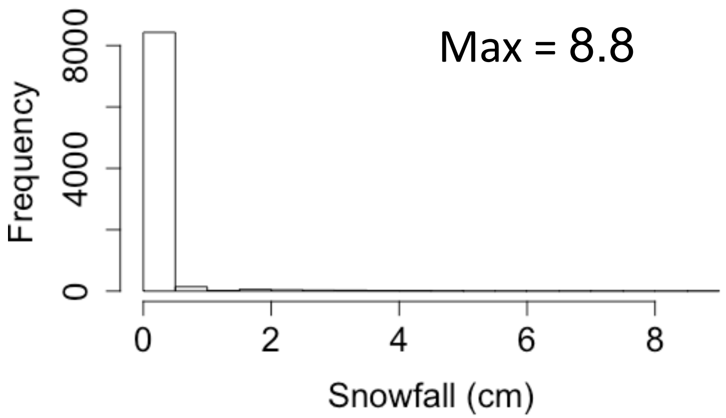


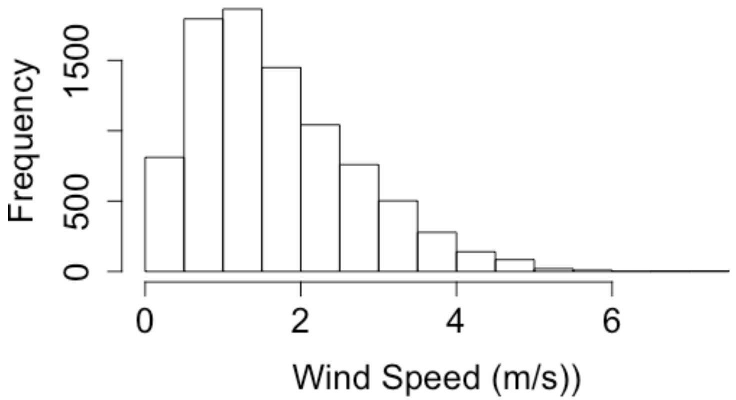


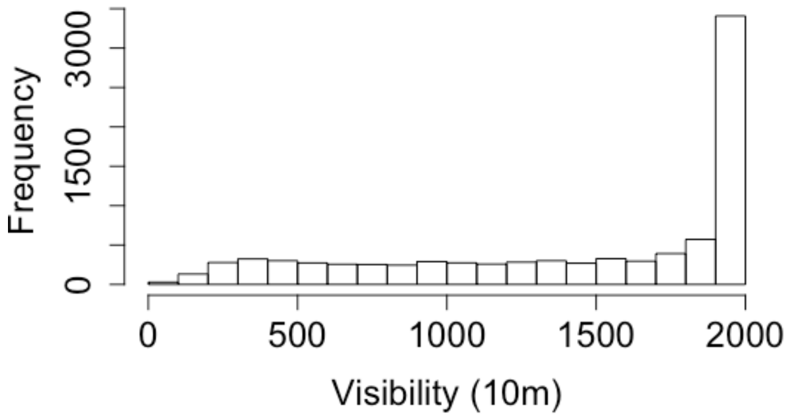


Characteristics

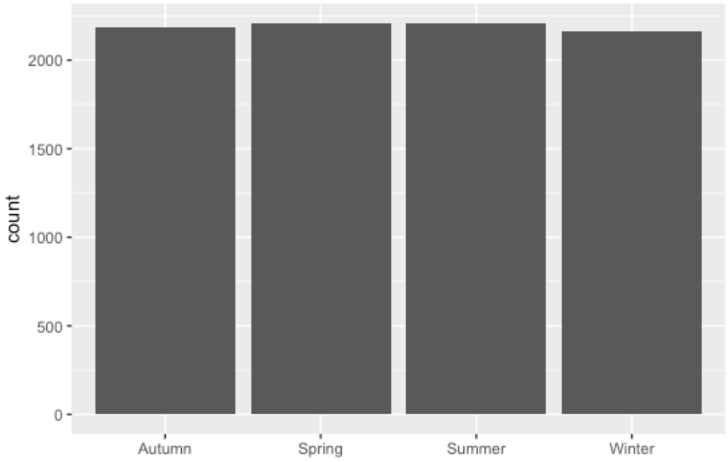




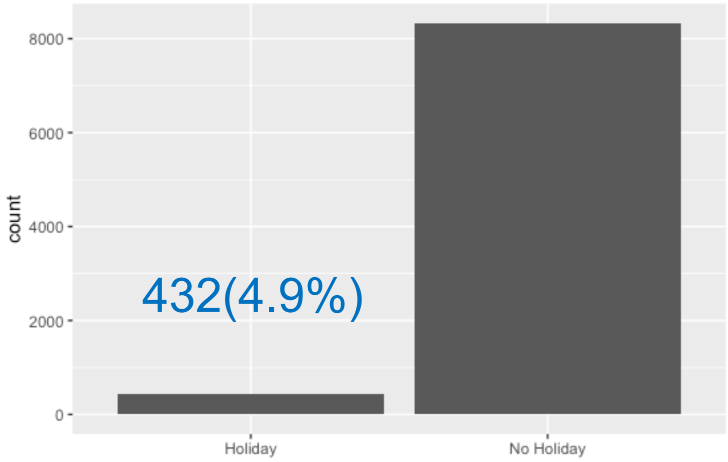




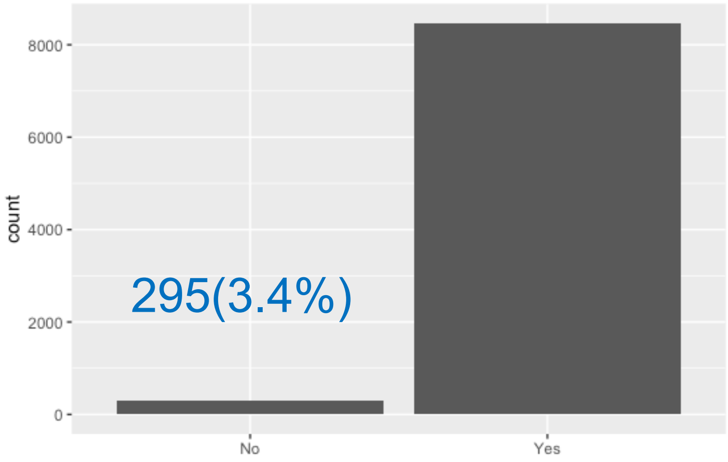
Characteristics



Season



Holiday



Functioning day

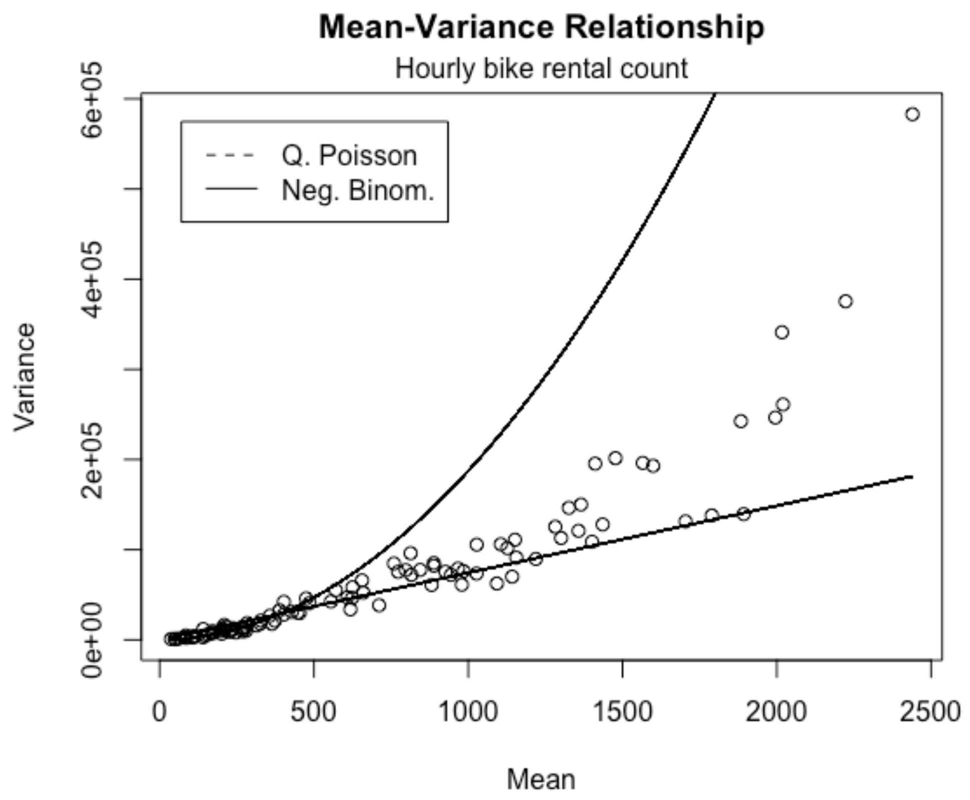
Correlation Matrix

Poisson Model

- Response variable is a rate (number of bikes rented per hour)
- All explanatory variables significant from univariate analysis
- Stepwise variable selection
- Full model deviance of 993,729 on 8,429 degrees of freedom
  - Overdispersion
  - Negative binomial model more appropriate
  - Poisson model has an AIC of 704,657
  - Negative binomial model has an AIC of 113,915

Poisson, Quasi-Poisson and Negative binomial Models

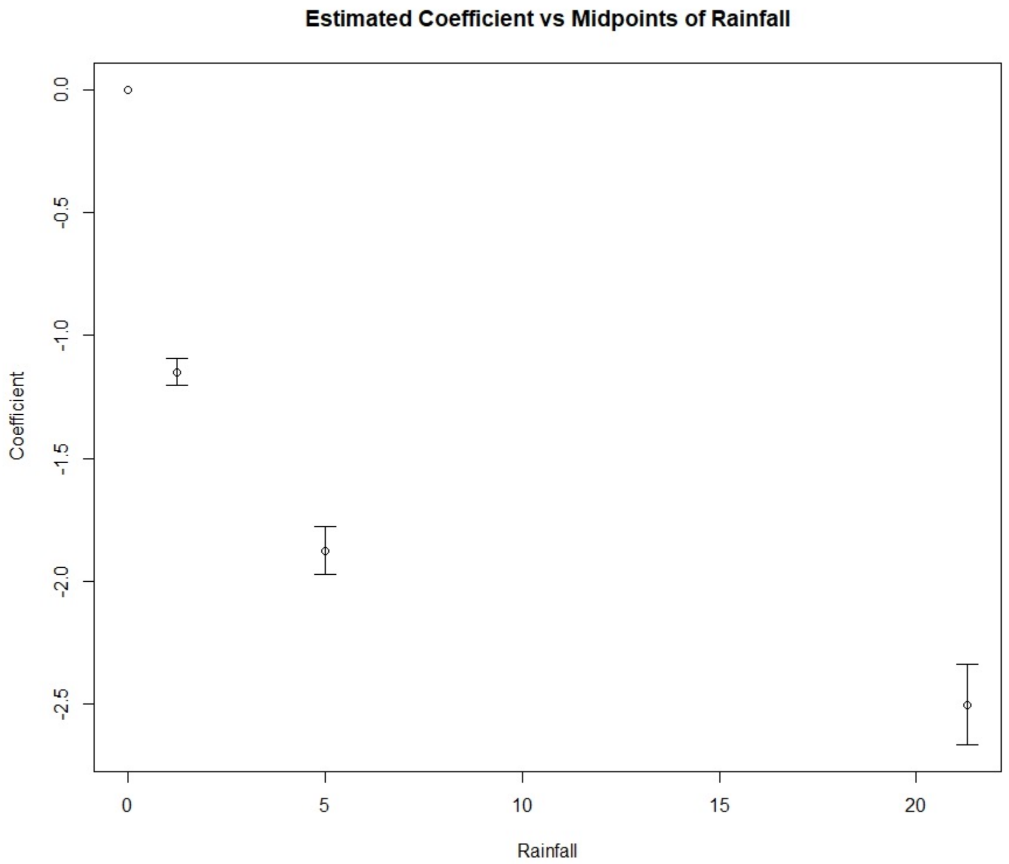
- The linear quasi-Poisson variance function performs very well for most of the data points but fail to capture the large variances of the highest demands of bike rental.
- The quadratic negative binomial variance function rises drastically as mean increases; however, it deviates from the true variances by over estimation.



Negative Binomial Model

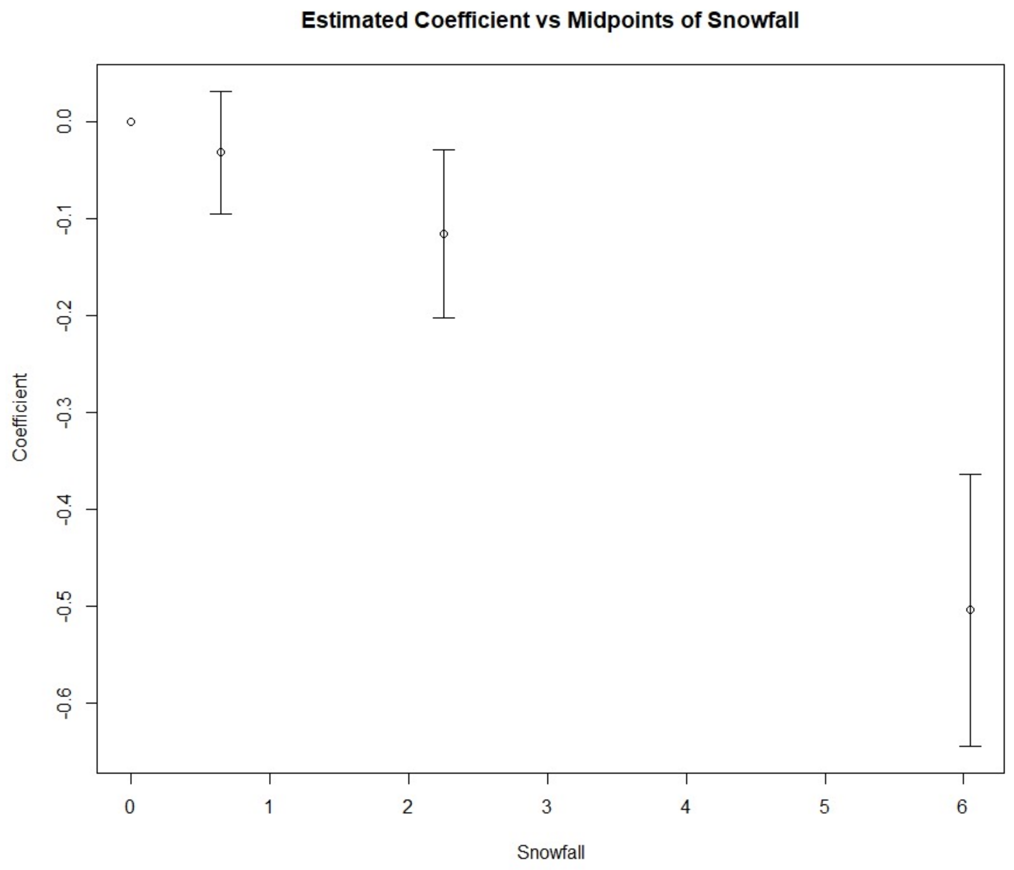
- In the presence of other explanatory variables
  - Visibility excluded (p-value 0.56)
  - Solar radiation excluded (p-value 0.058)
- Temperature
  - Adding dew point increases standard error
  - Highly correlated with dew point (0.91)
  - Drop dew point to avoid multicollinearity

Rainfall



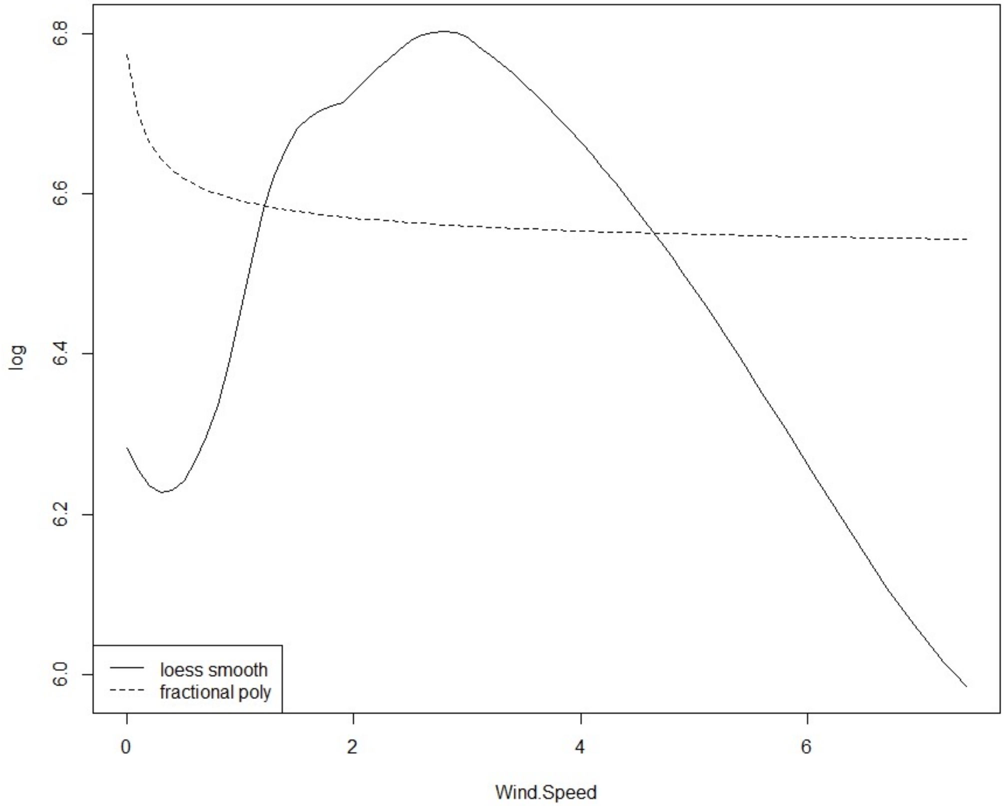
- Categories
  - None
  - Light – less than 2.5 mm/hour
  - Moderate – 2.5 to 7.6 mm/hour
  - Heavy – 7.6 mm/hour or greater
- No overlap between levels
- Fractional polynomials didn’t appear to make sense
- Rental rate decreases as rainfall increases

Snowfall



- Liquid Water Equivalent
- Categories
  - None
  - Light – less than 1 mm/hour
  - Moderate – 1 to 5 mm/hour
  - Heavy – 5 mm/hour or greater
- Lower AIC than fractional polynomials
- Rental rate decreases as snowfall increases

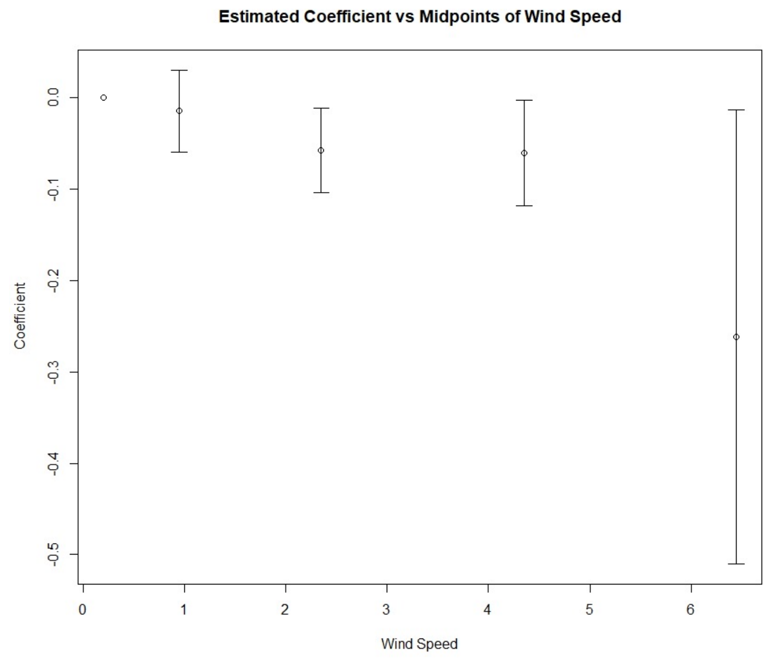
Wind Speed

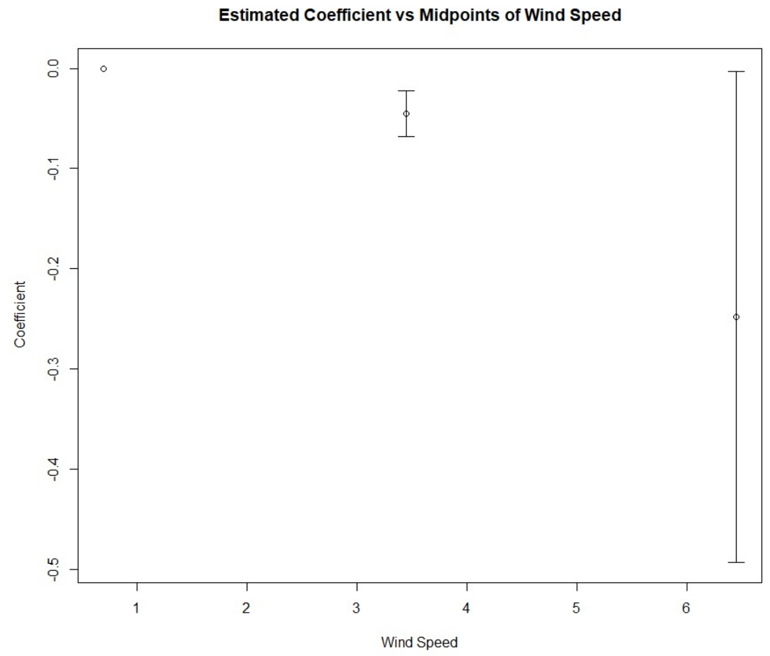


- Beaufort scale
- Categories
  - 0 (calm) – less than 0.5 m/s
  - 1 (light air) – 0.5-1.5 m/s
  - 2 (light breeze) – 1.6-3.3 m/s
  - 3 (gentle breeze) – 3.4-5.5 m/s
  - 4 (moderate breeze) – 5.5-7.9 m/s
- Rental rate should decrease at higher wind speeds

Wind Speed

- Level 1 appears no different than level 0
- Level 3 appears no different than level 2
- Likelihood ratio test supports combining
- High variance in level 4



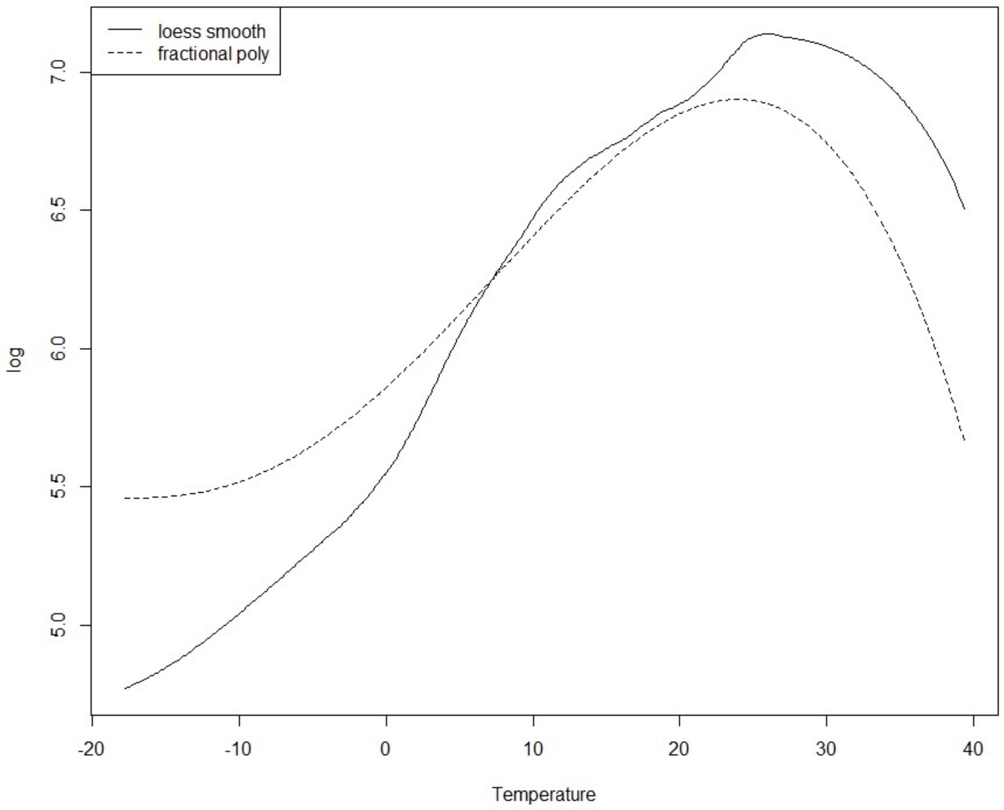


Temperature

- Fractional polynomials
- Rental rate increases up to 24 degrees Celsius and then decreases
- J = 2 is the best model

$$FP1 = I\left(\left(\frac{Temperature + 17.9}{10}\right)^3\right)$$

$$FP2 = I\left(\left(\frac{Temperature + 17.9}{10}\right)^3 * log\left(\left(\frac{Temperature + 17.9}{10}\right)\right)\right)$$

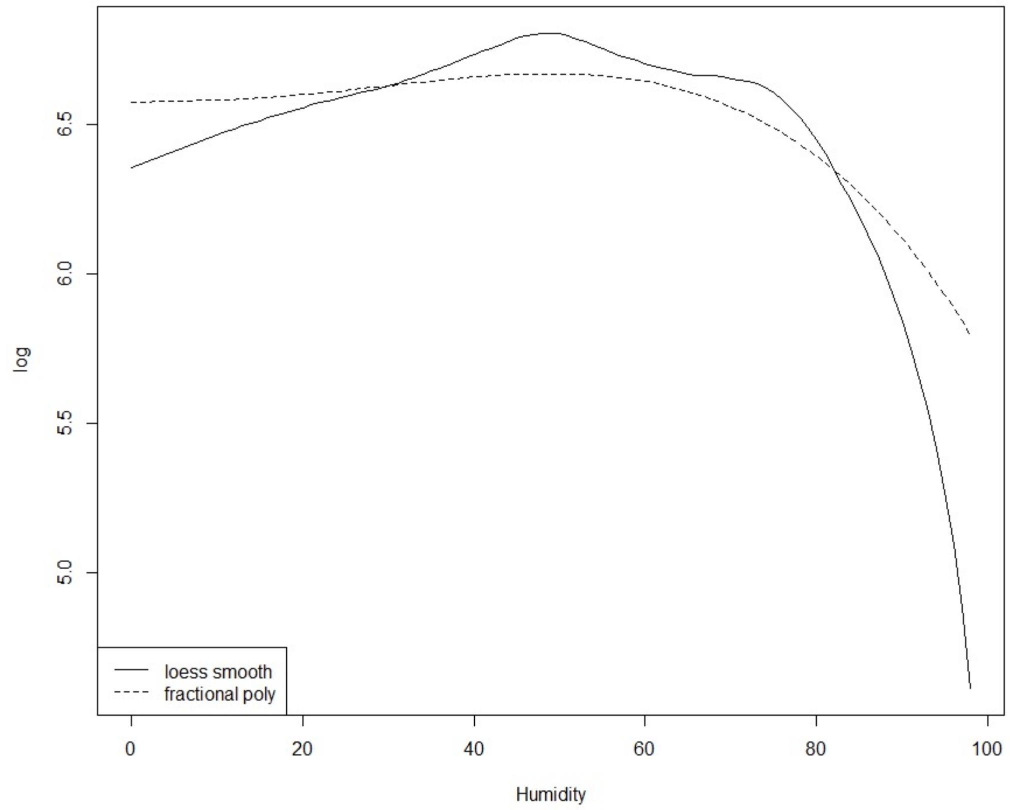


Humidity

- Fractional polynomials
- Rental rate flat up until 60%
- J = 2 is the best model

$$FP1 = I\left(\left(\frac{Humidity + 1}{100}\right)^3\right)$$

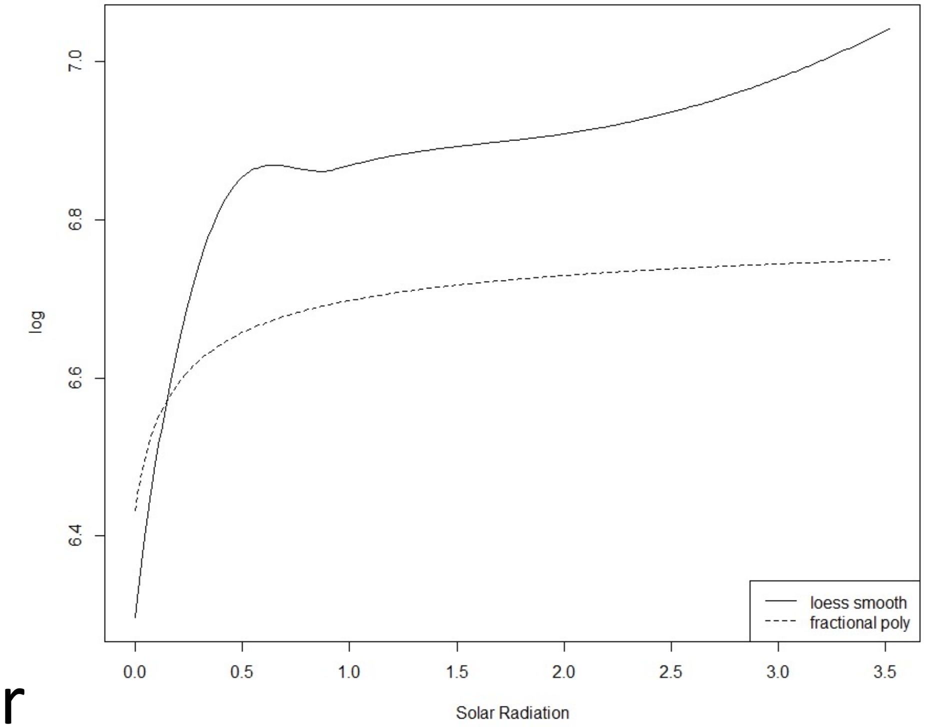
$$FP2 = I\left(\left(\frac{Humidity + 1}{100}\right)^3 * log\left(\left(\frac{Humidity + 1}{100}\right)\right)\right)$$



Visibility & Solar Radiation

- Check main effects not included
  - Visibility still not significant (p-value 0.21)
  - Solar radiation significant
- Add solar radiation back into the model
- Rental rate should decrease when its darker
- J = 1 is the best model

$$FP1 = I\left((Solar.Radiation + 0.1)^{-0.5}\right)$$



Interaction Effects

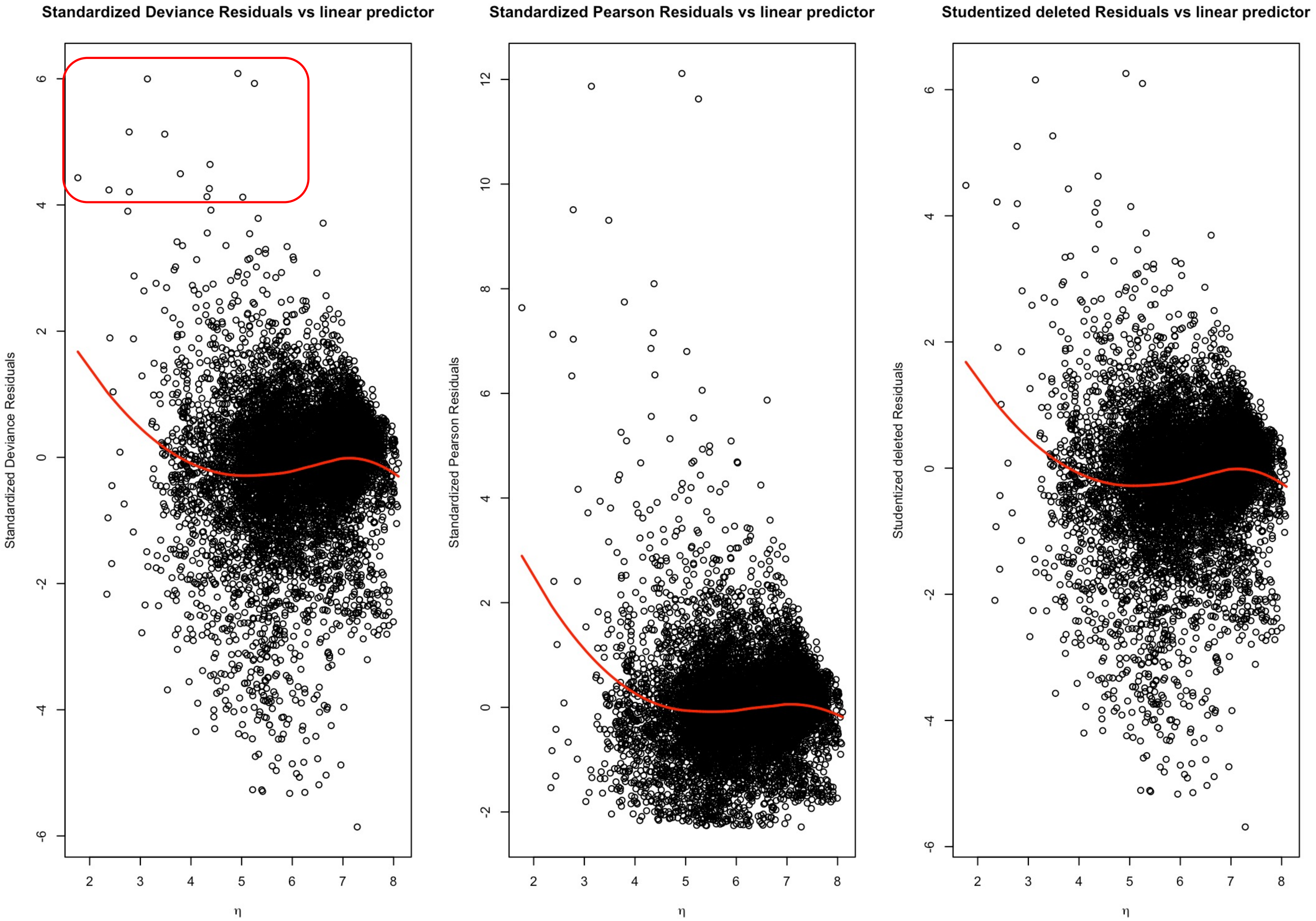
- Temperature, rainfall, and snowfall are the most important variables
- Temperature FP1 highly correlated with FP2 term
- Humidity FP1 is not highly correlated with FP2 term
- Interaction effects included in the model:
    - Temperature & Rainfall, Temperature & Snowfall, Temperature & Humidity, Temperature & Wind Speed, Temperature & Solar Radiation, Humidity & Wind Speed, Humidity & Solar Radiation
- Main effects included:
    - Hour, Temperature, Humidity, Wind Speed, Rainfall, Snowfall, Season, Holiday, Solar Radiation

Residual plot

Most outliers have high humidity.

humidity in outliers with |standard deviance residuals |> 3:

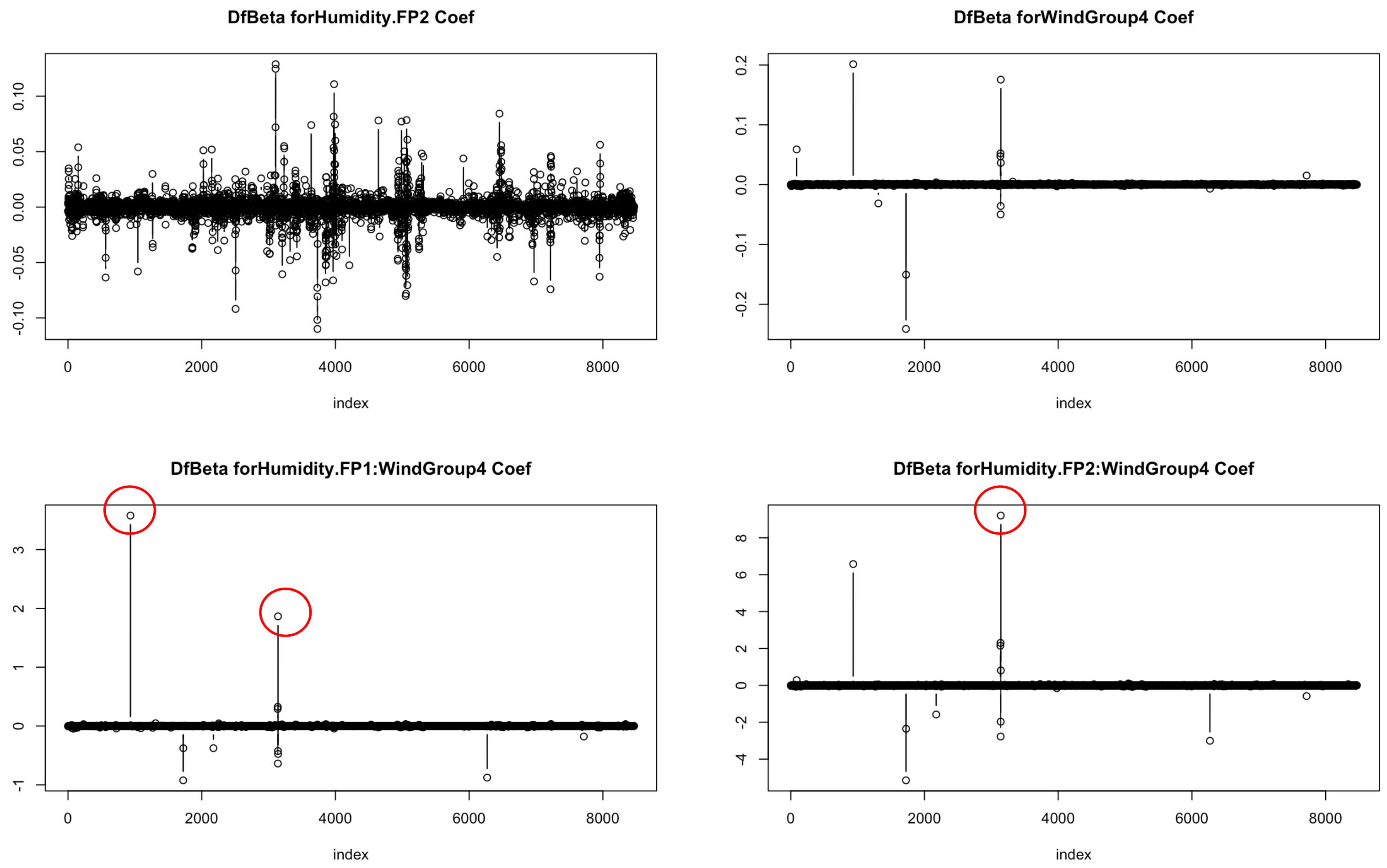
Min. : 39.0
1st Qu. : 87.0
Median : 92.0
Mean : 88.8
3rd Qu. : 97.0
Max. :98.0



DFBETAS

- Indicates the effect that deleting each observation has on the estimates for the regression coefficients.
- Most DFBETA <0.05

- Humidity FP2: Wind.grp.4
- Humidity FP1: Wind.grp.4
- Humidity FP2
- Windgroup4



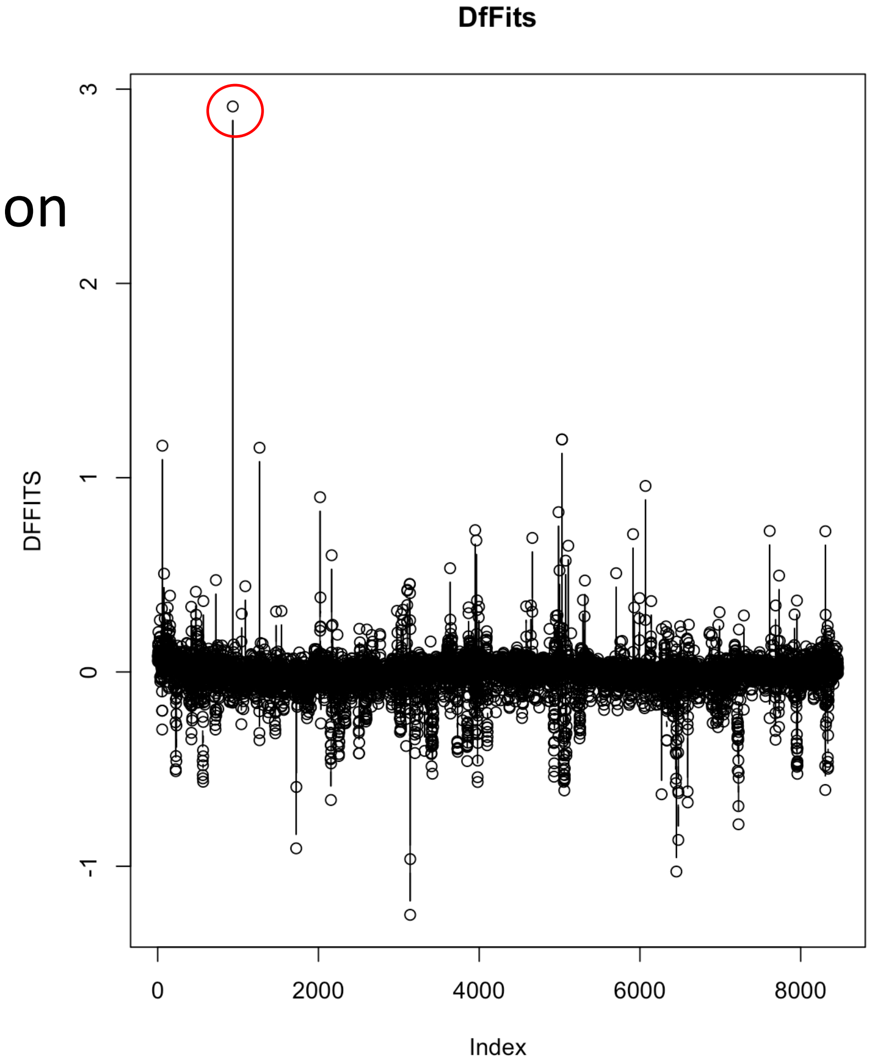
Few observations in Wind.4

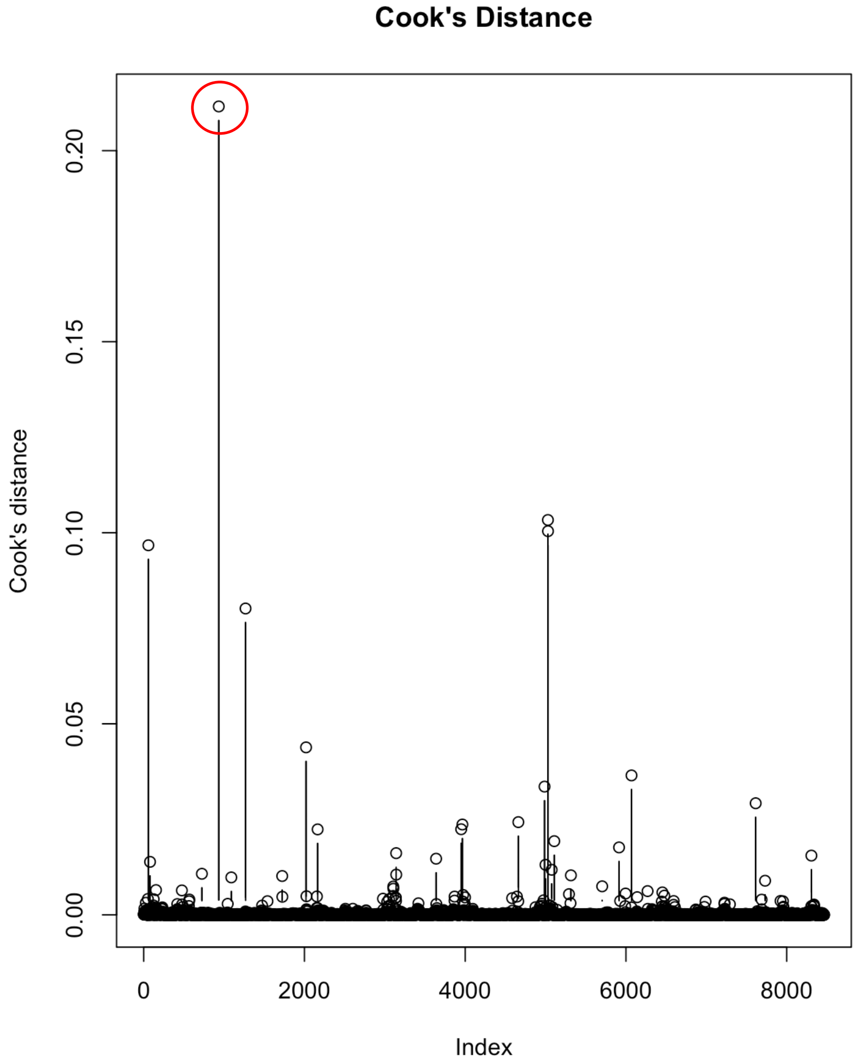
Wind01	Wind23	Wind4
3982	4468	15

DIFFITS

- Influential points
  in a statistical regression

- Few observations
  in Wind.4





VIF

- Multicollinearity

Variable Name	GVIF	Df	GVIF^(1/(2*Df))
Hour	15.33	23	1.06
Temperature.FP1	402.31	1	20.06
Temperature.FP2	344.80	1	18.57
Humidity.FP1	26.77	1	5.17
Humidity.FP2	17.08	1	4.13
Wind Group	58.85	2	2.77
Rain	225.75	3	2.47
Snow	50.75	3	1.92
Season	9.81	3	1.46
Holiday	1.04	1	1.02

Variable Name (Interactions)	GVIF	Df	
Solar.Radiation.FP1	19.99	1	4.47
Temperature.FP1:Rain	228.46	3	2.47
Temperature.FP1:Snow	47.10	3	1.90
Temperature.FP1:Humidity.FP1	9.83	1	3.14
Temperature.FP1:Humidity.FP2	19.06	1	4.37
Temperature.FP1:WindGroup	30.43	2	2.35
Temperature.FP1: Solar.Radiation.FP1	9.45	1	2.47
Humidity.FP1:WindGroup	46.05	2	2.61
Humidity.FP2:WindGroup	376.90	2	4.41
Humidity.FP1:Solar.Radiation.FP1	21.92	1	4.68
Humidity.FP2:Solar.Radiation.FP1	23.88	1	4.89

Model Interpretation

Rush hour –

more bicycles need

$$\exp(0.6965) = 2.01$$
$$\exp(0.7955) = 2.22$$

Coefficients	Estimate	S.E.	z-value	p-value
(Intercept)	5.5676	0.0705	78.9577	0.0000
Hour1	-0.1878	0.0329	-5.7099	0.0000
Hour2	-0.4906	0.0330	-14.8673	0.0000
Hour3	-0.7994	0.0331	-24.1232	0.0000
Hour4	-1.2624	0.0333	-37.8574	0.0000
Hour5	-1.2169	0.0334	-36.4448	0.0000
Hour6	-0.4764	0.0332	-14.3335	0.0000
Hour7	0.2219	0.0343	6.4731	0.0000
Hour8	0.6965	0.0377	18.4530	0.0000
Hour9	0.1108	0.0424	2.6154	0.0089
Hour10	-0.2873	0.0445	-6.4531	0.0000
Hour11	-0.2276	0.0451	-5.0415	0.0000
Hour12	-0.0860	0.0455	-1.8908	0.0587
Hour13	-0.0568	0.0456	-1.2473	0.2123
Hour14	-0.0543	0.0453	-1.1977	0.2310
Hour15	0.0294	0.0447	0.6590	0.5099
Hour16	0.1274	0.0434	2.9348	0.0033
Hour17	0.3736	0.0411	9.0964	0.0000
Hour18	0.7955	0.0371	21.4713	0.0000
Hour19	0.5752	0.0343	16.7879	0.0000
Hour20	0.4762	0.0332	14.3501	0.0000
Hour21	0.4957	0.0330	15.0428	0.0000
Hour22	0.3918	0.0328	11.9410	0.0000
Hour23	0.1297	0.0328	3.9532	0.0001

Interpretation

Coefficients	Estimate	S.E.	z-value	p-value
Temperature.FP1	0.0956	0.0025	38.3230	0.0000
Temperature.FP2	-0.0549	0.0014	-38.7724	0.0000
Humidity.FP1	-1.0601	0.0979	-10.8285	0.0000
Humidity.FP2	-1.6354	0.5812	-2.8137	0.0049
WindGroup23	**-0.1270**	**0.0313**	**-4.0509**	**0.0001**
WindGroup4	**-0.1416**	**0.2629**	**-0.5387**	**0.5901**
RainLight	**-1.0263**	**0.0508**	**-20.1855**	**0.0000**
RainModerate	**-1.9849**	**0.1115**	**-17.8017**	**0.0000**
RainHeavy	**-2.5205**	**0.2067**	**-12.1942**	**0.0000**
SnowLight	**-0.1426**	**0.0378**	**-3.7675**	**0.0002**
SnowModerate	**-0.0186**	**0.0676**	**-0.2751**	**0.7833**
SnowHeavy	**-0.6015**	**0.1934**	**-3.1101**	**0.0019**
SeasonSpring	**0.3612**	**0.0207**	**17.4177**	**0.0000**
SeasonSummer	**0.5794**	**0.0290**	**19.9946**	**0.0000**
SeasonAutumn	**0.6081**	**0.0209**	**29.1089**	**0.0000**
Holiday Holiday	**-0.2425**	**0.0226**	**-10.7226**	**0.0000**
Solar.Radiation.FP1	**-0.2449**	**0.0198**	**-12.3888**	**0.0000**

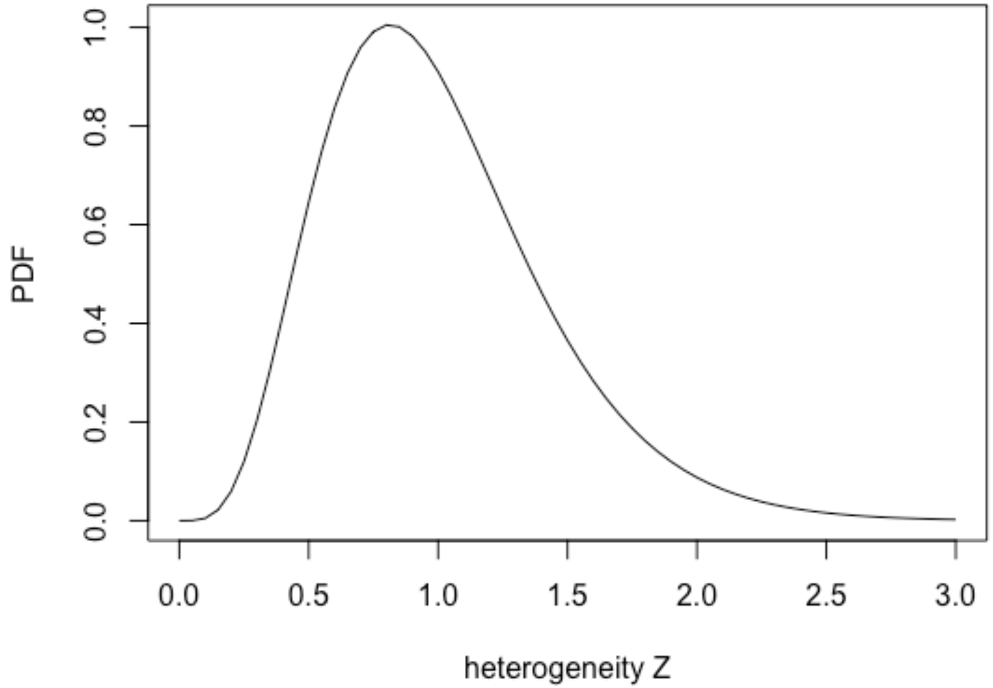
Interpretation

- Interaction terms are all associated with FP terms.
- Fractional Polynomial is a more analytical method which we cannot interpretate intuitively.

Coefficients	Estimate	S.E.	z-value	p-value
Temperature.FP1:RainLight	0.0021	0.0009	2.2877	0.0222
Temperature.FP1:RainModerate	0.0094	0.0021	4.4635	0.0000
Temperature.FP1:RainHeavy	0.0088	0.0039	2.2285	0.0259
Temperature.FP1:SnowLight	0.0353	0.0071	4.9661	0.0000
Temperature.FP1:SnowModerate	-0.0151	0.0112	-1.3422	0.1795
Temperature.FP1:SnowHeavy	0.0567	0.0306	1.8546	0.0637
Temperature.FP1:Humidity.FP1	-0.0022	0.0010	-2.2638	0.0236
Temperature.FP1:Humidity.FP2	0.0194	0.0055	3.5229	0.0004
Temperature.FP1:WindGroup23	0.0017	0.0003	6.0106	0.0000
Temperature.FP1:WindGroup4	-0.0052	0.0065	-0.7945	0.4269
Temperature.FP1:Solar.Radiation.FP1	0.0015	0.0002	8.3548	0.0000
Humidity.FP1:WindGroup23	-0.1872	0.0415	-4.5116	0.0000
Humidity.FP1:WindGroup4	-0.5850	2.2223	-0.2633	0.7924
Humidity.FP2:WindGroup23	-0.8170	0.3177	-2.5711	0.0101
Humidity.FP2:WindGroup4	-0.5464	9.3099	-0.0587	0.9532
Humidity.FP1:Solar.Radiation.FP1	0.1438	0.0292	4.9302	0.0000
Humidity.FP2:Solar.Radiation.FP1	-0.6395	0.1775	-3.6021	0.0003

Heterogeneity Z

- Contributes to individual's mean unobserved characteristics.
- Bike rental count
  - at Q1 of the distribution of unobserved heterogeneity Z is **31% lower** than expected from their observed characteristics
  - Median is **6% lower**
  - At Q3, **25% higher** than expected

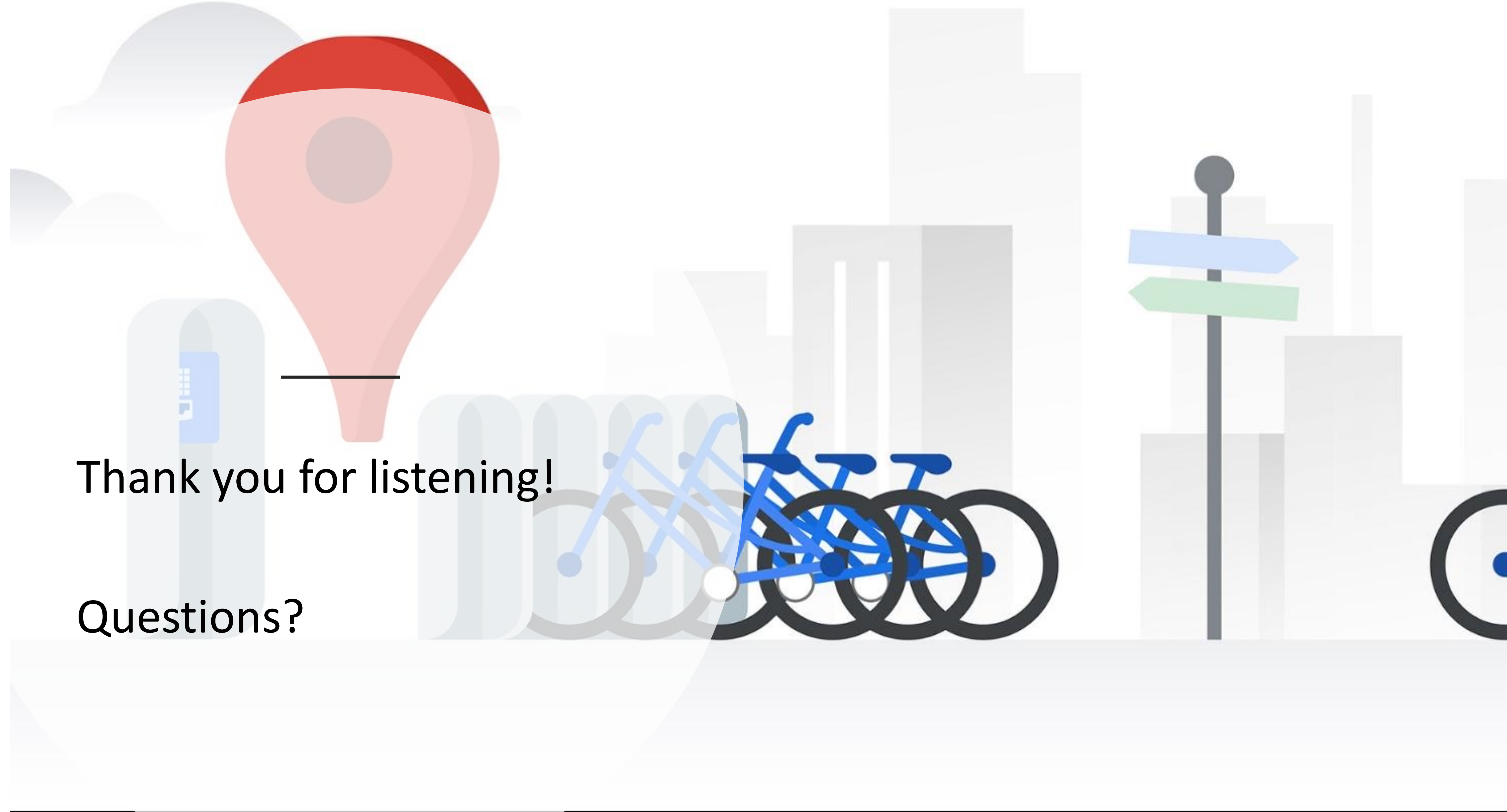


Summary

- Our best model includes
  - the hour of the day, temperature, humidity, wind, rain, snow, season, holiday, solar radiation and
  - interaction terms
    - between temperature and rain, snow, humidity, wind, solar radiation,
    - between humidity and wind, solar radiation
- Fractional polynomial
- Multicollinearity

References

[1] https://www.baranidesign.com/faq-articles/2020/1/19/rain-rate-intensity-classification
[2] https://fpaw.aero/sites/default/files/128/baker-snowfall-intensity-table-a4a-fpaw-summer-brief-v3-0.pdf
[3] https://www.nssl.noaa.gov/education/svrwx101/winter/faq/
[4] https://windy.app/blog/wind-speed-beaufort-scale.html
[5] https://journals.aau.dk/index.php/djtr/article/view/3560/3106
[6] https://link-springer-com.proxy.ulib.uits.iu.edu/article/10.1007/s11116-014-9540-7



Thank you for listening!

Questions?