

Data Exploration & Linear Regression Modeling - Seoul Bike Sharing

Dataset from <https://archive.ics.uci.edu/ml/datasets/Seoul+Bike+Sharing+Demand>

Christine Kuhn

Data columns:

1 - Date : year-month-day, 2 - Rented Bike count - Count of bikes rented at each hour; 3 - Hour - Hour of the day; 4 - Temperature-Temperature in Celsius; 5 - Humidity - %; 6 - Windspeed - m/s; 7 - Visibility - 10m; 8 - Dew point temperature - Celsius; 9 - Solar radiation - MJ/m²; 10 - Rainfall - mm; 11 - Snowfall - cm; 12 - Seasons - Winter, Spring, Summer, Autumn; 13 - Holiday - Holiday/No holiday, 14 - Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

Output variable:

Column 2 (Rented bike count)

```
clear all  
close all
```

Read in dataset

```
dataset=readtable('SeoulBikeData.csv');
```

Warning: Column headers from the file were modified to make them valid MATLAB identifiers before creating variable names for the table. The original column headers are saved in the VariableDescriptions property. Set 'VariableNamingRule' to 'preserve' to use the original column headers as table variable names.

```
dataset = convertvars(dataset, "Date", 'datenum');  
dataset = convertvars(dataset, "FunctioningDay", 'string');  
dataset = convertvars(dataset, "Holiday", 'string');  
dataset = convertvars(dataset, "Seasons", 'string');  
seasonnum = double(categorical(dataset.Seasons));  
fdaynum = double(categorical(dataset.FunctioningDay));  
hdaynum = double(categorical(dataset.Holiday));  
datamat = table2array(dataset(:,1:11));  
% Note: Season, Holiday, and FunctioningDay data have been converted to  
% numbers, but those numbers still represent categories.  
% Subtract 1 from hdaynum and fdaynum to turn those into booleans (0 or 1)  
% Create 3 new columns to hold the season data: Winter Y/N, Spring Y/N,  
% Summer Y/N. (Fall will be represented by Winter=Spring=Summer=0)  
winter = logical(seasonnum==4);  
spring = logical(seasonnum==2);  
summer = logical(seasonnum==3);  
datamat(:,12:16) = [winter spring summer -1*(hdaynum-2) 1+(fdaynum-2)];
```

Divide dataset into inputs and outputs

```
% inputs are all of datamat except col 2  
datasetInputs=datamat;
```

```

datasetInputs(:,2)=[];
datasetOutputs=datamat(:,2);

onedayavg = movmean(datasetOutputs,[23,0]);
sevendayavg = movmean(datasetOutputs,[(7*24-1) 0]); % 7-day average up til now
thirtydayavg = movmean(datasetOutputs,[(30*24-1) 0]); % 30-day avg up til now

onedaymax = movmax(datasetOutputs,[23,0]);
sevendaymax = movmax(datasetOutputs,[(7*24-1) 0]); % 7-day max up til now
thirtydaymax = movmax(datasetOutputs,[(30*24-1) 0]); % 30-day max up til now

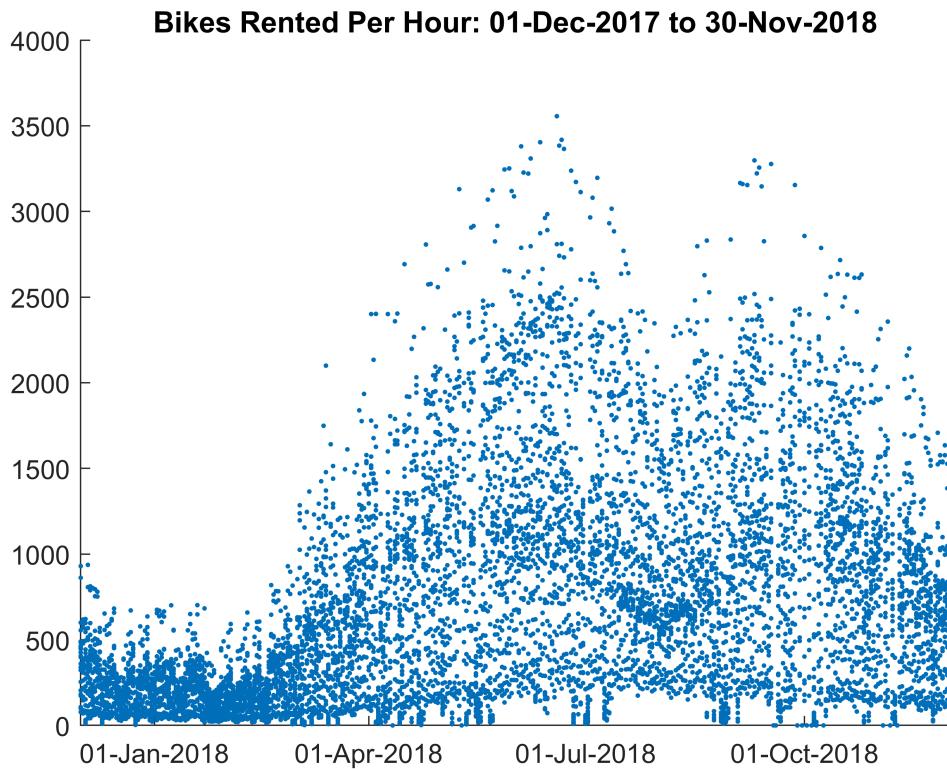
for ii=1:15
    onedayavg_in(:,ii) = movmean(datasetInputs(:,ii),[23,0]);
    sevendayavg_in(:,ii) = movmean(datasetInputs(:,ii),[(7*24-1) 0]); % 7-day average up til now
    thirtydayavg_in(:,ii) = movmean(datasetInputs(:,ii),[(30*24-1) 0]); % 30-day avg up til now

    onedaymax_in(:,ii) = movmax(datasetInputs(:,ii),[23,0]);
    sevendaymax_in(:,ii) = movmax(datasetInputs(:,ii),[(7*24-1) 0]); % 7-day max up til now
    thirtydaymax_in(:,ii) = movmax(datasetInputs(:,ii),[(30*24-1) 0]); % 30-day max up til now
end

dateslist = datetime(datamat(:,1),'ConvertFrom','datenum');

figure
scatter(dateslist, datasetOutputs,'.')
xtickformat('dd-MMM-yyyy')
title(['Bikes Rented Per Hour: ' datestr(dateslist(1)) ' to ' datestr(dateslist(end))])

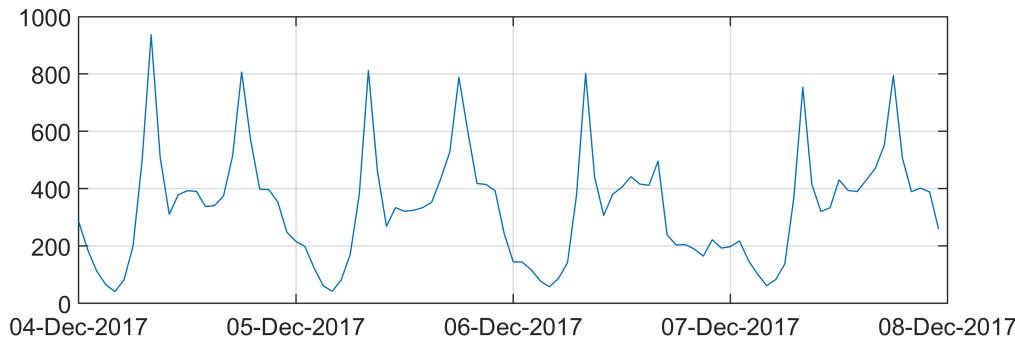
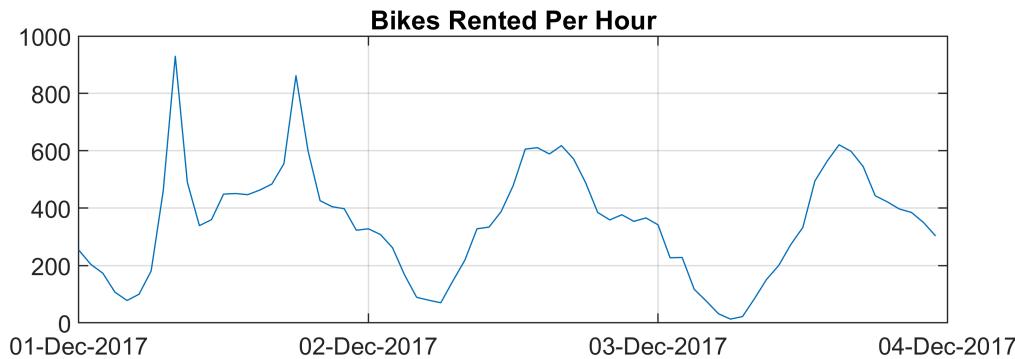
```



```

figure
subplot(2,1,1)
plot(dateslist(1:(3*24))+(1/24)*datamat(1:(3*24),3), datasetOutputs(1:(3*24)))
xtickformat('dd-MMM-yyyy')
grid on
title('Bikes Rented Per Hour')
subplot(2,1,2)
plot(dateslist((3*24+1):(7*24))+(1/24)*datamat((3*24+1):(7*24),3), datasetOutputs((3*24+1):(7*24)))
xtickformat('dd-MMM-yyyy')
grid on

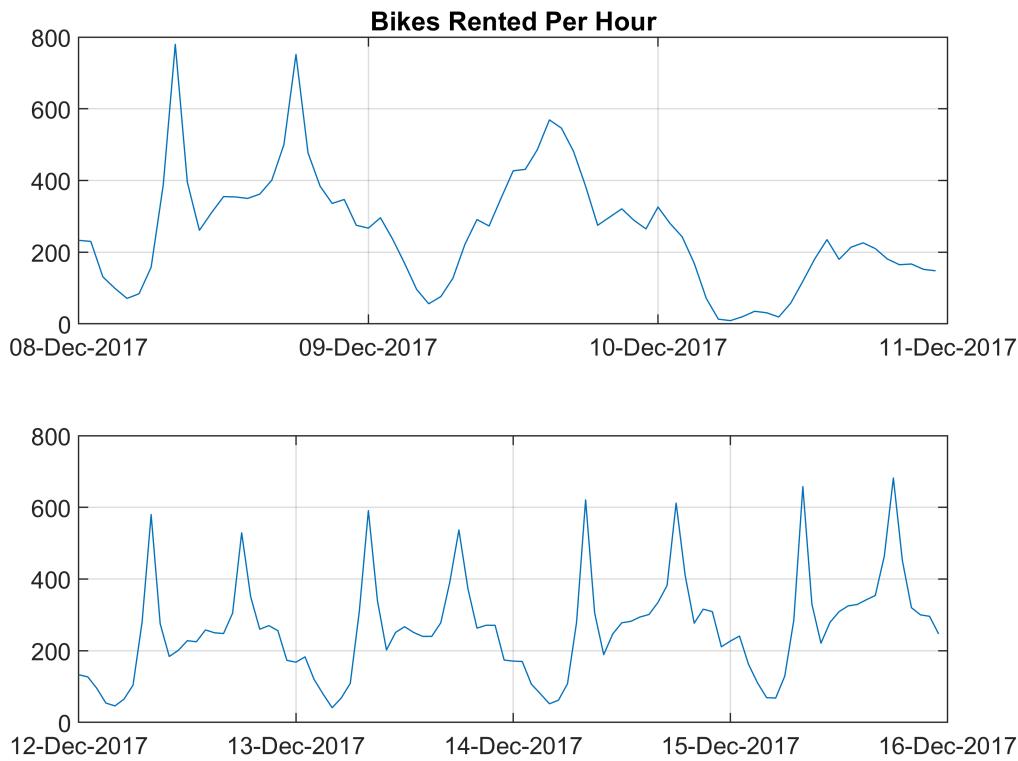
```



```

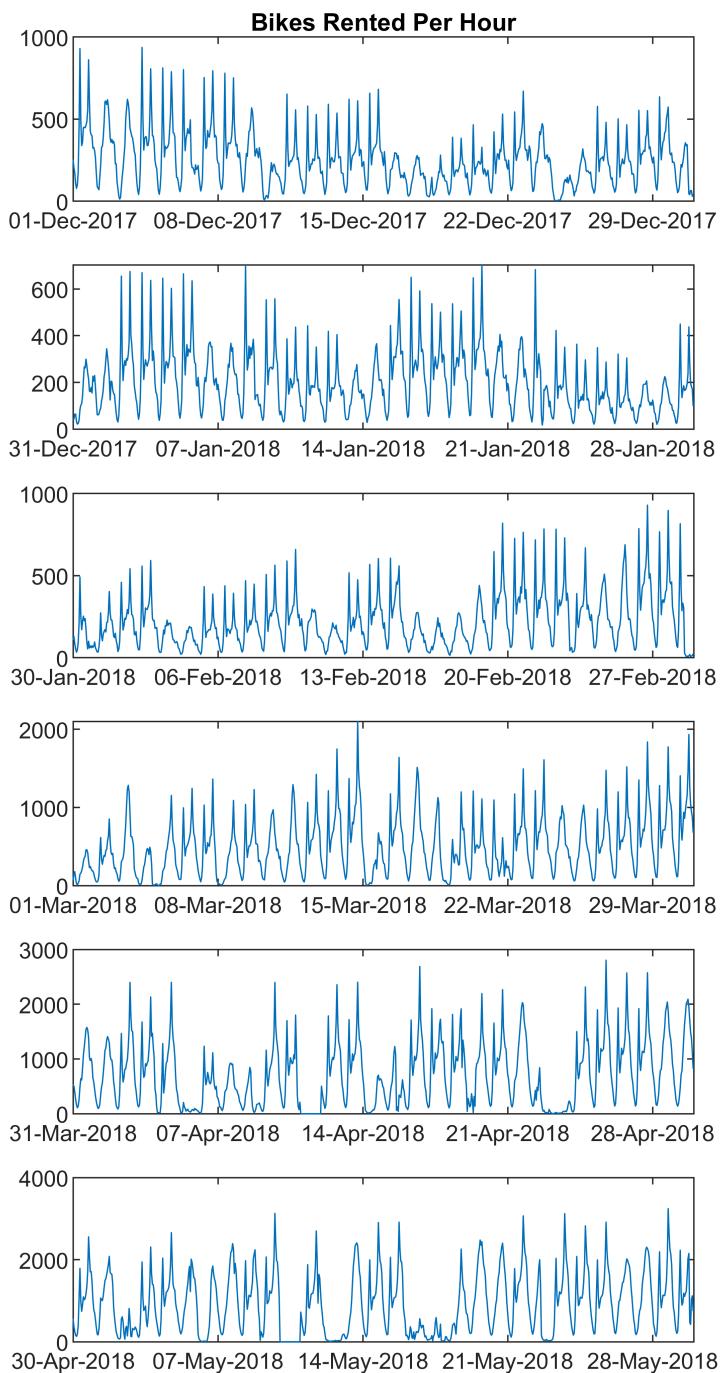
figure
subplot(2,1,1)
plot(dateslist((7*24+1):(10*24))+(1/24)*datamat((7*24+1):(10*24),3), datasetOutputs((7*24+1):(10*24)))
xtickformat('dd-MMM-yyyy')
grid on
title('Bikes Rented Per Hour')
subplot(2,1,2)
plot(dateslist((11*24+1):(15*24))+(1/24)*datamat((11*24+1):(15*24),3), datasetOutputs((11*24+1):(15*24)))
xtickformat('dd-MMM-yyyy')
grid on

```



```
% Weekdays have two peaks: one in the morning and one in the evening;
% weekend days have one peak in the afternoon. Similar patterns are seen in
% the plots below, except for Sept 18-26. Minimal rentals in the early
% morning, but not 0. Unclear whether the data is how many bike rentals are
% initiated during each hour or whether the count includes bikes which were
% rented earlier for multiple hours.
```

```
figure('Renderer', 'painters', 'Position', [10 10 800 1600])
for i=1:6
subplot(6,1,i)
im=i-1;
plot(dateslist((im*30*24+1):(i*30*24))+(1/24)*datamat((im*30*24+1):(i*30*24),3), datasetOutputs
xtickformat('dd-MMM-yyyy')
if i==1
title('Bikes Rented Per Hour')
end
end
```

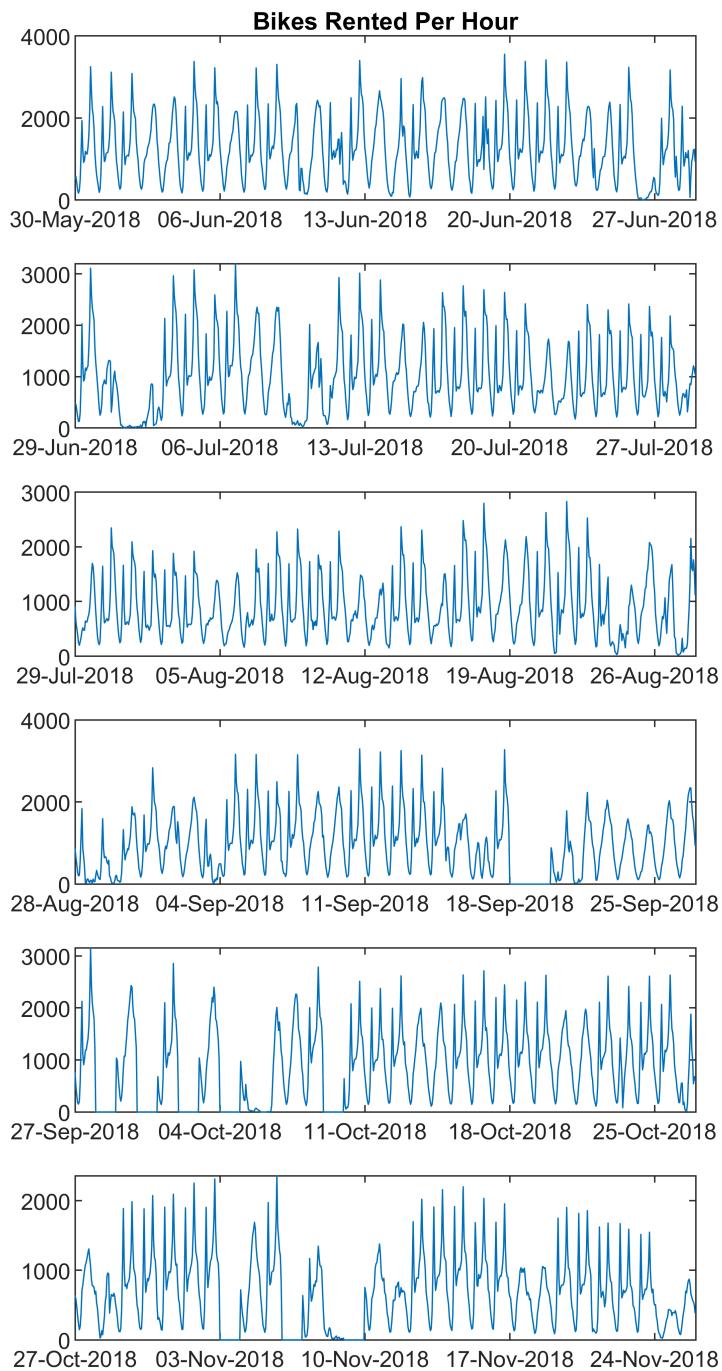


```
figure('Renderer', 'painters', 'Position', [10 10 800 1600])
for i=7:12
subplot(6,1,(i-6))
im=i-1;
```

```

plot(dateslist((im*30*24+1):(i*30*24)+(1/24)*datamat((im*30*24+1):(i*30*24),3), datasetOutputs
xtickformat('dd-MMM-yyyy')
if i==7
    title('Bikes Rented Per Hour')
end
end

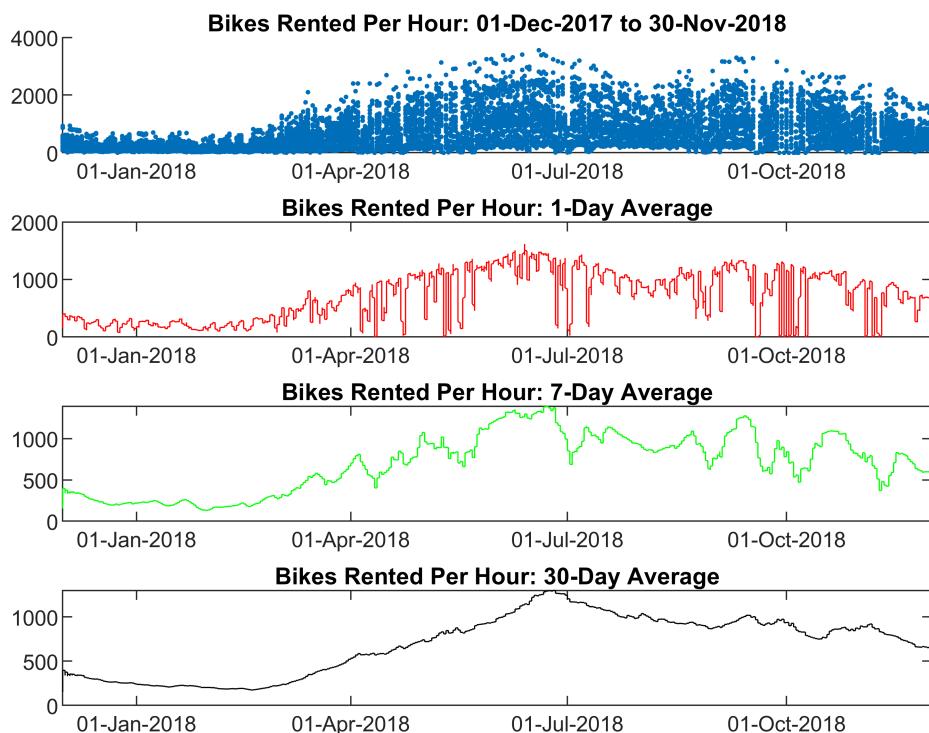
```



```

figure
subplot(4,1,1)
scatter(dateslist, datasetOutputs, '.')
hold on
xtickformat('dd-MMM-yyyy')
title(['Bikes Rented Per Hour: ' datestr(dateslist(1)) ' to ' datestr(dateslist(end))])
subplot(4,1,2)
plot(dateslist, onedayavg, 'r')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 1-Day Average')
subplot(4,1,3)
plot(dateslist, sevendayavg, 'g')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 7-Day Average')
subplot(4,1,4)
plot(dateslist, thirtydayavg, 'k')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 30-Day Average')

```



```

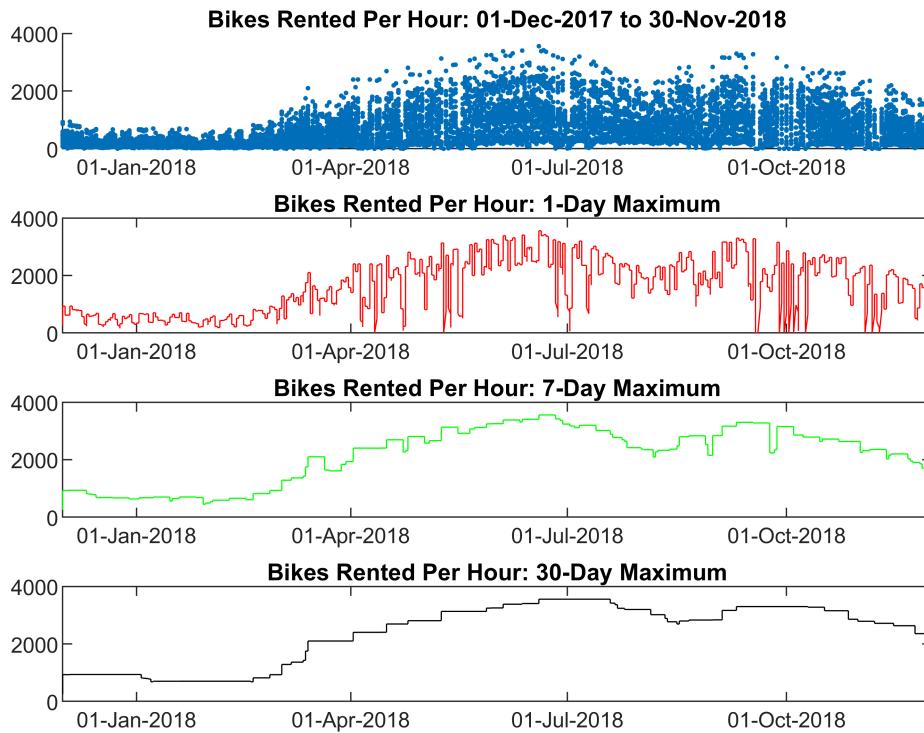
figure
subplot(4,1,1)
scatter(dateslist, datasetOutputs, '.')
hold on
xtickformat('dd-MMM-yyyy')
title(['Bikes Rented Per Hour: ' datestr(dateslist(1)) ' to ' datestr(dateslist(end))])

```

```

subplot(4,1,2)
plot(dateslist, onedaymax, 'r')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 1-Day Maximum')
subplot(4,1,3)
plot(dateslist, sevendaymax, 'g')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 7-Day Maximum')
subplot(4,1,4)
plot(dateslist, thirtydaymax, 'k')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 30-Day Maximum')

```

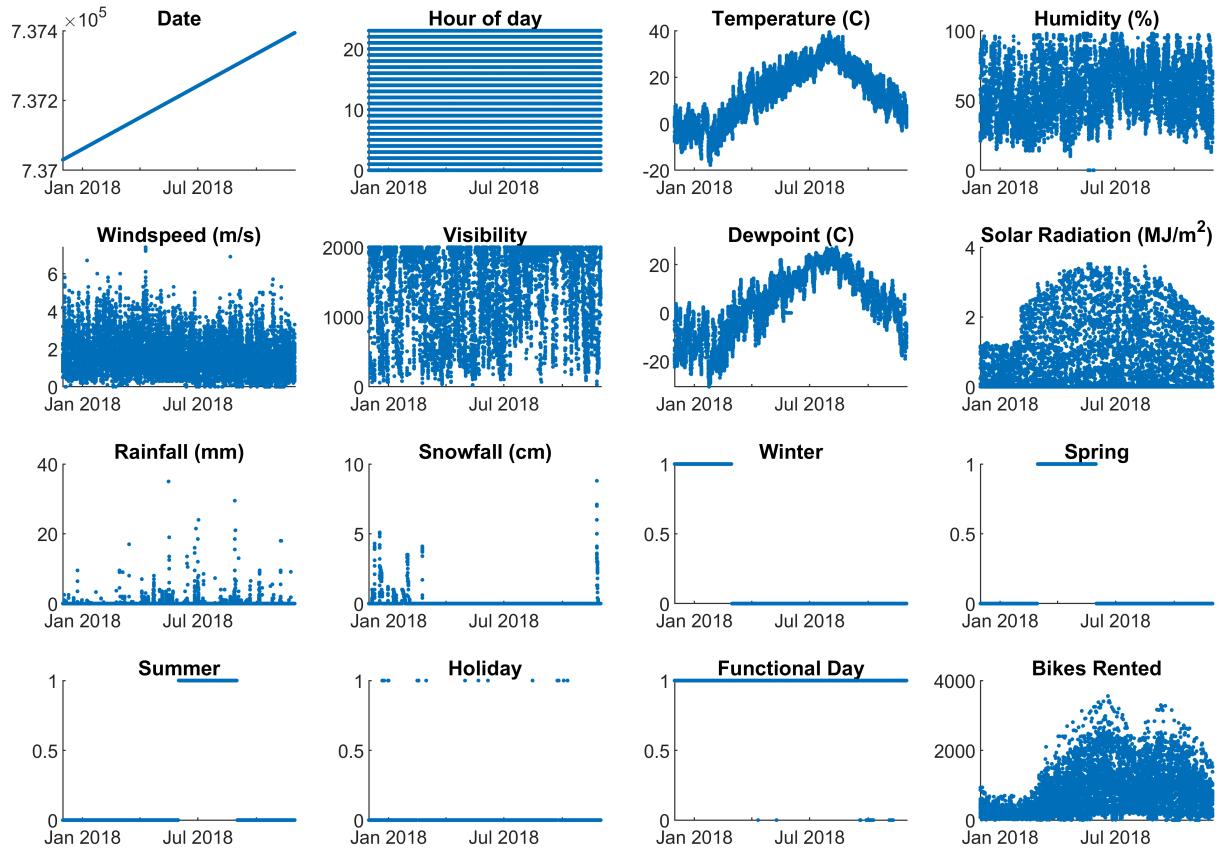


```

figure('Renderer', 'painters', 'Position', [10 10 900 600])
subplot(4,4,1)
scatter(dateslist, datasetInputs(:,1), '.')
title('Date')
subplot(4,4,2)
scatter(dateslist, datasetInputs(:,2), '.')
title('Hour of day')
subplot(4,4,3)
scatter(dateslist, datasetInputs(:,3), '.')
title('Temperature (C)')
subplot(4,4,4)
scatter(dateslist, datasetInputs(:,4), '.')
title('Humidity (%)')
subplot(4,4,5)
scatter(dateslist, datasetInputs(:,5), '.')

```

```
title('Windspeed (m/s)')
subplot(4,4,6)
scatter(dateslist, datasetInputs(:,6),'.')
title('Visibility')
subplot(4,4,7)
scatter(dateslist, datasetInputs(:,7),'.')
title('Dewpoint (C)')
subplot(4,4,8)
scatter(dateslist, datasetInputs(:,8),'.')
title('Solar Radiation (MJ/m^2)')
subplot(4,4,9)
scatter(dateslist, datasetInputs(:,9),'.')
title('Rainfall (mm)')
subplot(4,4,10)
scatter(dateslist, datasetInputs(:,10),'.')
title('Snowfall (cm)')
subplot(4,4,11)
scatter(dateslist, datasetInputs(:,11),'.')
title('Winter')
subplot(4,4,12)
scatter(dateslist, datasetInputs(:,12),'.')
title('Spring')
subplot(4,4,13)
scatter(dateslist, datasetInputs(:,13),'.')
title('Summer')
subplot(4,4,14)
scatter(dateslist, datasetInputs(:,14),'.')
title('Holiday')
subplot(4,4,15)
scatter(dateslist, datasetInputs(:,15),'.')
title('Functional Day')
subplot(4,4,16)
scatter(dateslist, datasetOutputs,'.')
title('Bikes Rented')
```

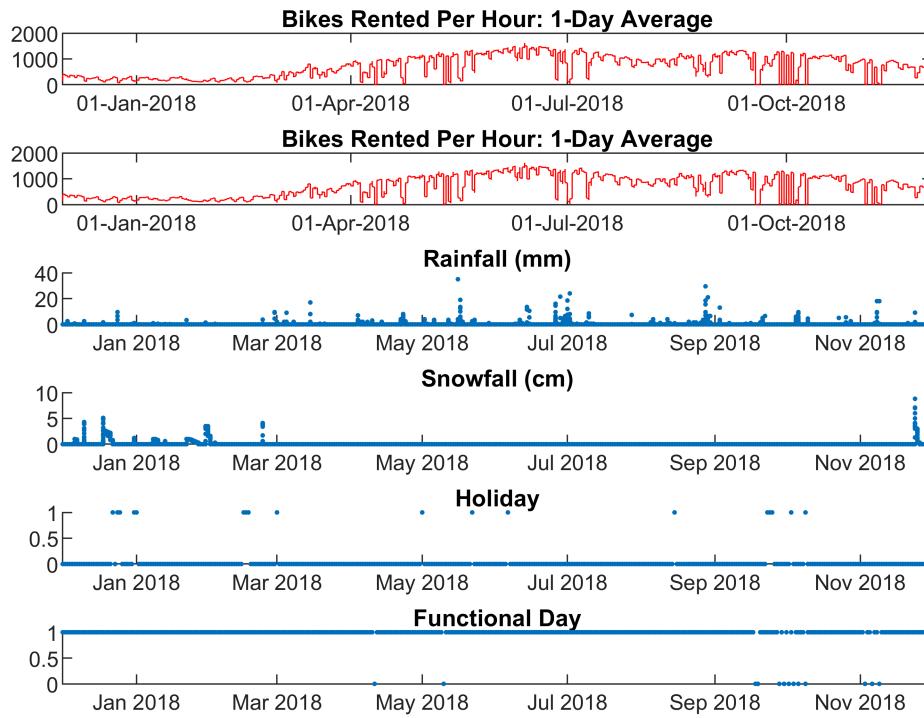


```

varnames={'Date', 'Hour', 'Temp', 'Hum', 'Windsp', 'Vis', 'Dewpt', 'SolRad', 'Rain', 'Snow', 'Winter', 'Sp'
          'Summer', 'Holiday', 'Functional Day' }

figure
subplot(6,1,1)
plot(dateslist, onedayavg, 'r')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 1-Day Average')
subplot(6,1,2)
plot(dateslist, onedayavg, 'r')
xtickformat('dd-MMM-yyyy')
title('Bikes Rented Per Hour: 1-Day Average')
subplot(6,1,3)
scatter(dateslist, datasetInputs(:,9), '.')
title('Rainfall (mm)')
subplot(6,1,4)
scatter(dateslist, datasetInputs(:,10), '.')
title('Snowfall (cm)')
subplot(6,1,5)
scatter(dateslist, datasetInputs(:,14), '.')
title('Holiday')
subplot(6,1,6)
scatter(dateslist, datasetInputs(:,15), '.')
title('Functional Day')

```



% Nonfunctional days and spring/summer rainfall days seem to correlate with % spikiness of bike rental distribution.

```
% Normalize the inputs
datasetInputs_orig = datasetInputs;
datasetInputs = mapminmax(datasetInputs_orig)';
testmax=max(datasetInputs');
testmin=min(datasetInputs');

% Linear Regression on Bikes Rented vs. Solar Radiation
mdl_linear = fitlm(datasetInputs(:,8),datasetOutputs)
```

```
mdl_linear =
Linear regression model:
y ~ 1 + x1

Estimated Coefficients:
              Estimate      SE      tStat     pValue
  (Intercept)    936.11    11.286    82.942    0
    x1            342.14    13.476    25.39    2.7804e-137
```

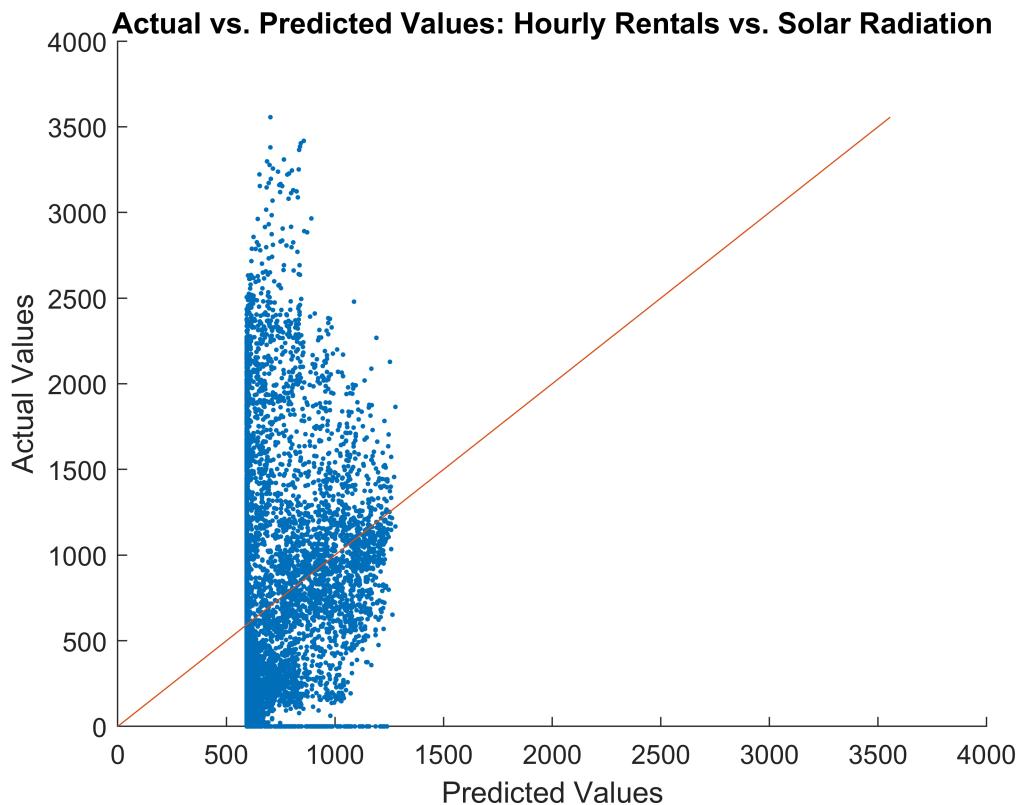
Number of observations: 8760, Error degrees of freedom: 8758
Root Mean Squared Error: 623
R-squared: 0.0686, Adjusted R-Squared: 0.0685
F-statistic vs. constant model: 645, p-value = 2.78e-137

```
predicted = predict(mdl_linear,datasetInputs(:,8));
```

```
RMSE_linearModel = sqrt(mean((predicted-datasetOutputs).^2))
```

```
RMSE_linearModel = 622.4593
```

```
figure  
scatter(predicted, datasetOutputs, '.')  
hold on  
plot(datasetOutputs,datasetOutputs)  
title('Actual vs. Predicted Values: Hourly Rentals vs. Solar Radiation')  
ylabel('Actual Values')  
xlabel('Predicted Values')
```



```
figure  
scatter(dateslist,datasetOutputs,'.')
```

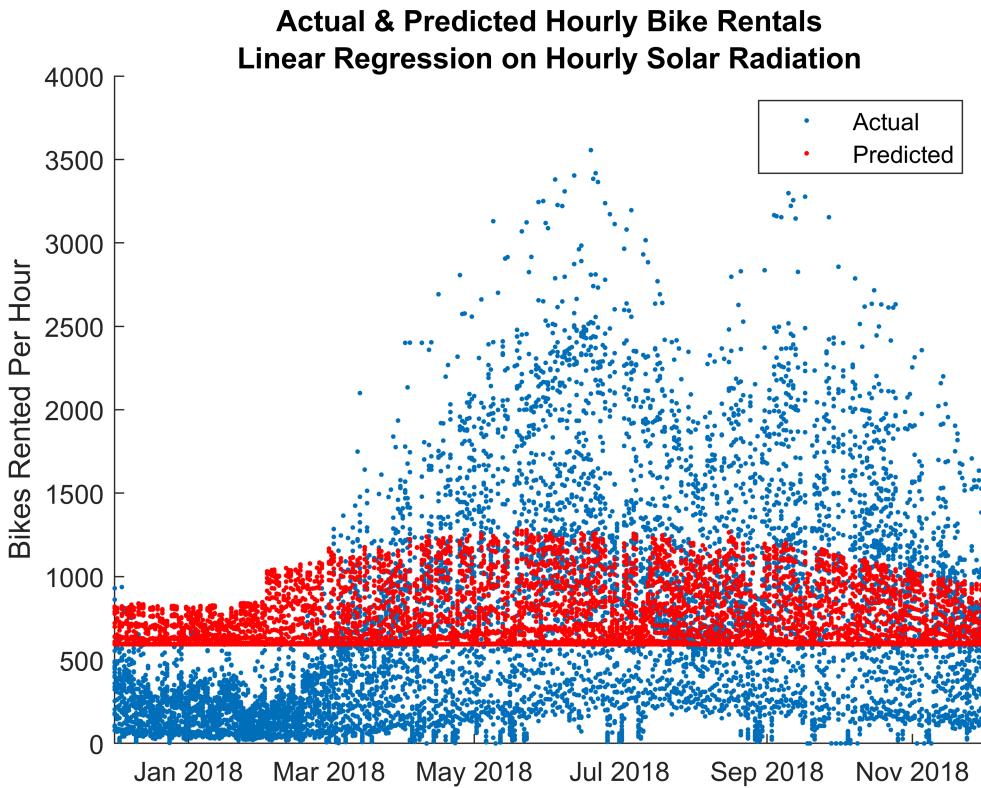
```
hold on
```

```
scatter(dateslist,predicted,'.r')
```

```
title({'Actual & Predicted Hourly Bike Rentals','Linear Regression on Hourly Solar Radiation'})
```

```
ylabel('Bikes Rented Per Hour')
```

```
legend({'Actual','Predicted'})
```

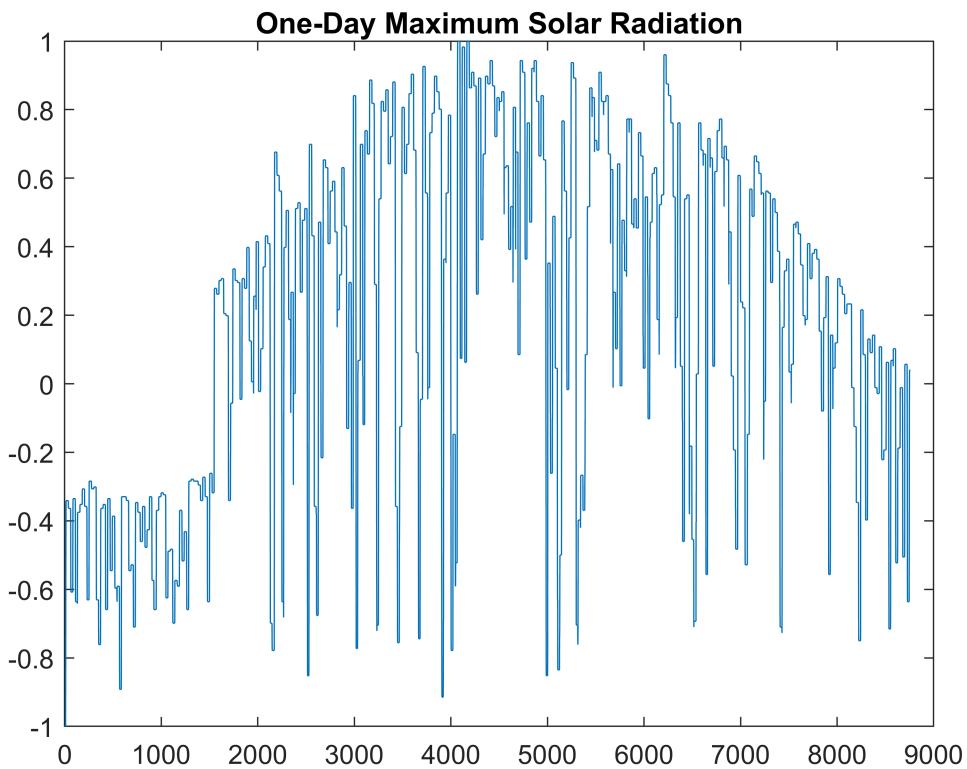


```

onedaymax_sol = movmax(datasetInputs(:,8),[23,0]);
sevendaymax_sol = movmax(datasetInputs(:,8),[(7*24-1) 0]); % 7-day max up til now
onedaymax_temp = movmax(datasetInputs(:,3),[23,0]);
sevendaymax_temp = movmax(datasetInputs(:,3),[(7*24-1) 0]); % 7-day max up til now
sevendaymean_vis = movmean(datasetInputs(:,6),[(7*24-1) 0]); % 7-day max up til now
sevendaymean_hum = movmean(datasetInputs(:,4),[(7*24-1) 0]); % 7-day max up til now

figure
plot(onedaymax_sol)
title('One-Day Maximum Solar Radiation')

```



```
mdl_linear2 = fitlm(onedaymax_sol,onedaymax)
```

```
mdl_linear2 =
Linear regression model:
y ~ 1 + x1
```

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1452.2	8.1682	177.78	0
x1	1169.7	15.128	77.323	0

Number of observations: 8760, Error degrees of freedom: 8758

Root Mean Squared Error: 711

R-squared: 0.406, Adjusted R-Squared: 0.406

F-statistic vs. constant model: 5.98e+03, p-value = 0

```
predicted2 = predict(mdl_linear2,onedaymax_sol);
RMSE_linearmodel2 = sqrt(mean((predicted2-onedaymax).^2))
```

RMSE_linearmodel2 = 710.8259

figure

scatter(predicted2, onedaymax, '.')

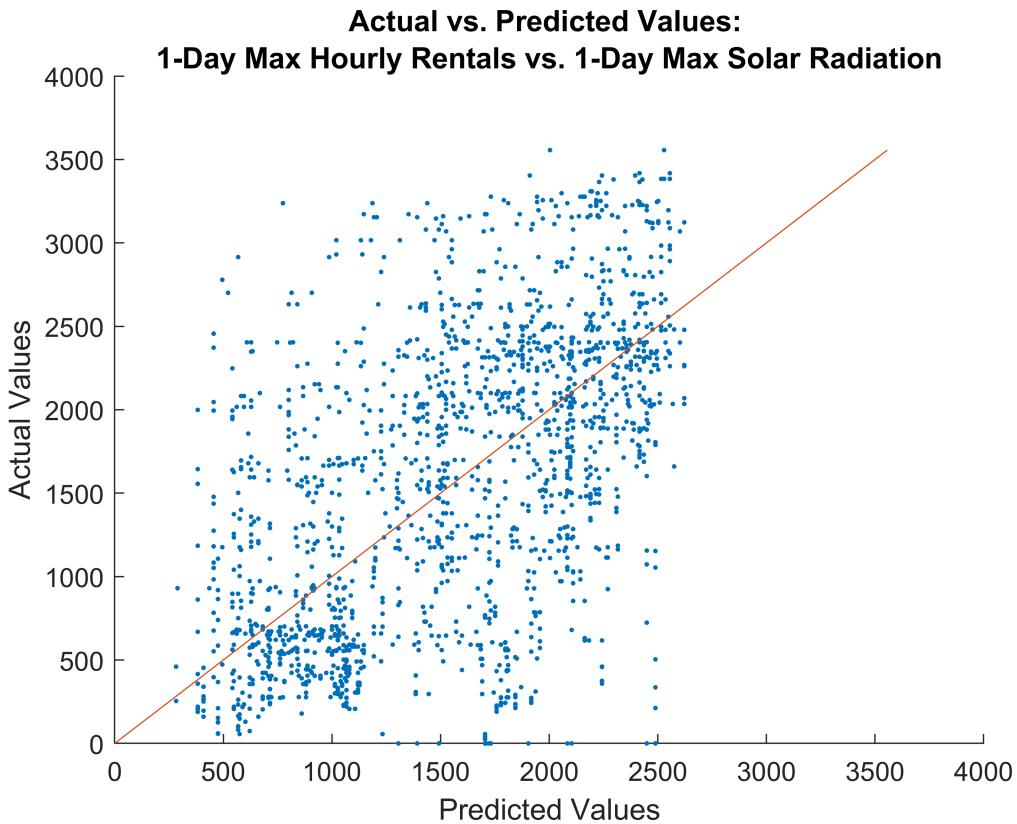
hold on

plot(onedaymax,onedaymax)

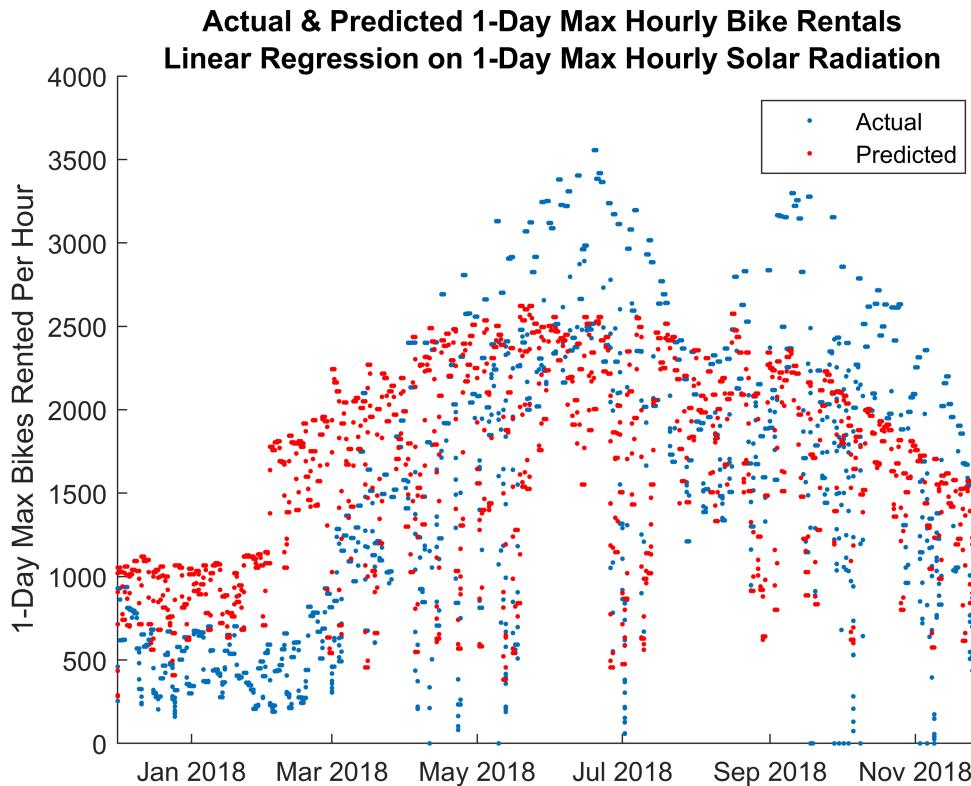
title({'Actual vs. Predicted Values: ';'1-Day Max Hourly Rentals vs. 1-Day Max Solar Radiation'})

ylabel('Actual Values')

xlabel('Predicted Values')



```
figure
scatter(dateslist,onedaymax,'.')
hold on
scatter(dateslist,predicted2,'.r')
title({'Actual & Predicted 1-Day Max Hourly Bike Rentals','Linear Regression on 1-Day Max Hourly Bike Rentals'})
ylabel('1-Day Max Bikes Rented Per Hour')
legend({'Actual','Predicted'})
```



```
mdl_linear3 = fitlm(sevendaymax_sol,sevendaymax)
```

```
mdl_linear3 =
Linear regression model:
y ~ 1 + x1
```

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1285.2	9.0889	141.4	0
x1	1795.6	13.662	131.43	0

Number of observations: 8760, Error degrees of freedom: 8758

Root Mean Squared Error: 566

R-squared: 0.664, Adjusted R-Squared: 0.664

F-statistic vs. constant model: 1.73e+04, p-value = 0

```
predicted3 = predict(mdl_linear3,sevendaymax_sol);
RMSE_linearmodel3 = sqrt(mean((predicted3-sevendaymax).^2))
```

RMSE_linearmodel3 = 565.8367

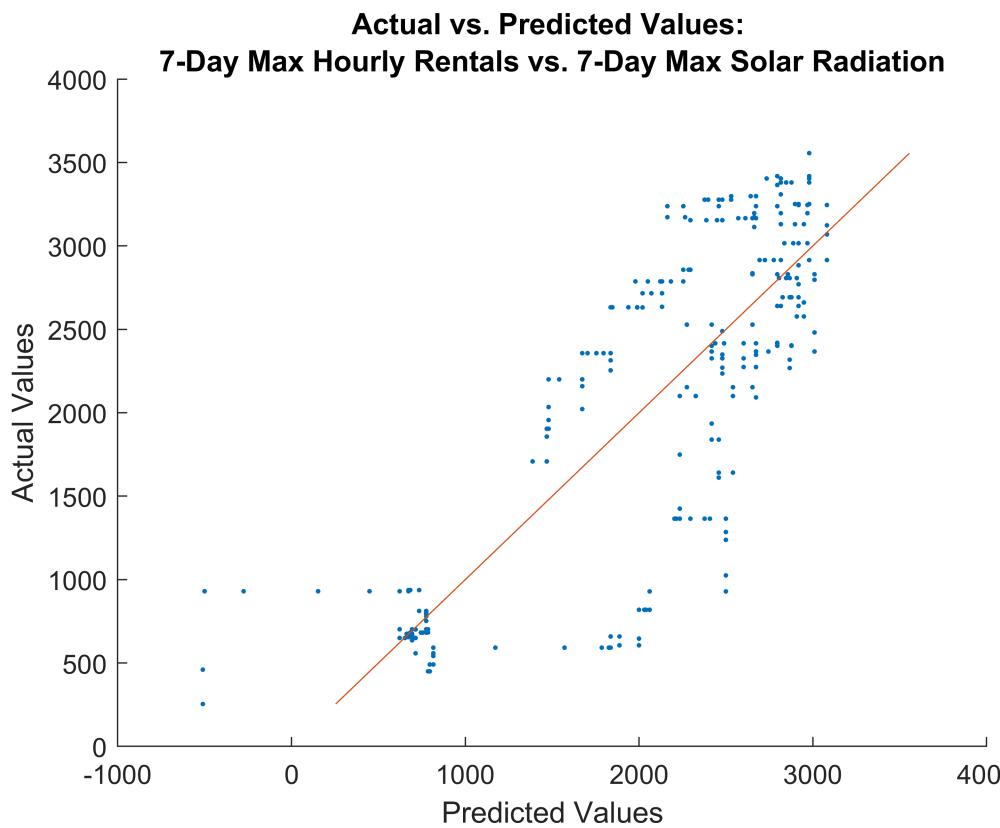
```
figure
```

```
scatter(predicted3, sevendaymax,'.')
hold on
```

```
plot(sevendaymax,sevendaymax)
```

```
title({'Actual vs. Predicted Values: ';'7-Day Max Hourly Rentals vs. 7-Day Max Solar Radiation'});
ylabel('Actual Values')
```

```
xlabel('Predicted Values')
```



```
figure
```

```
scatter(dateslist,sevendaymax,'.')
```

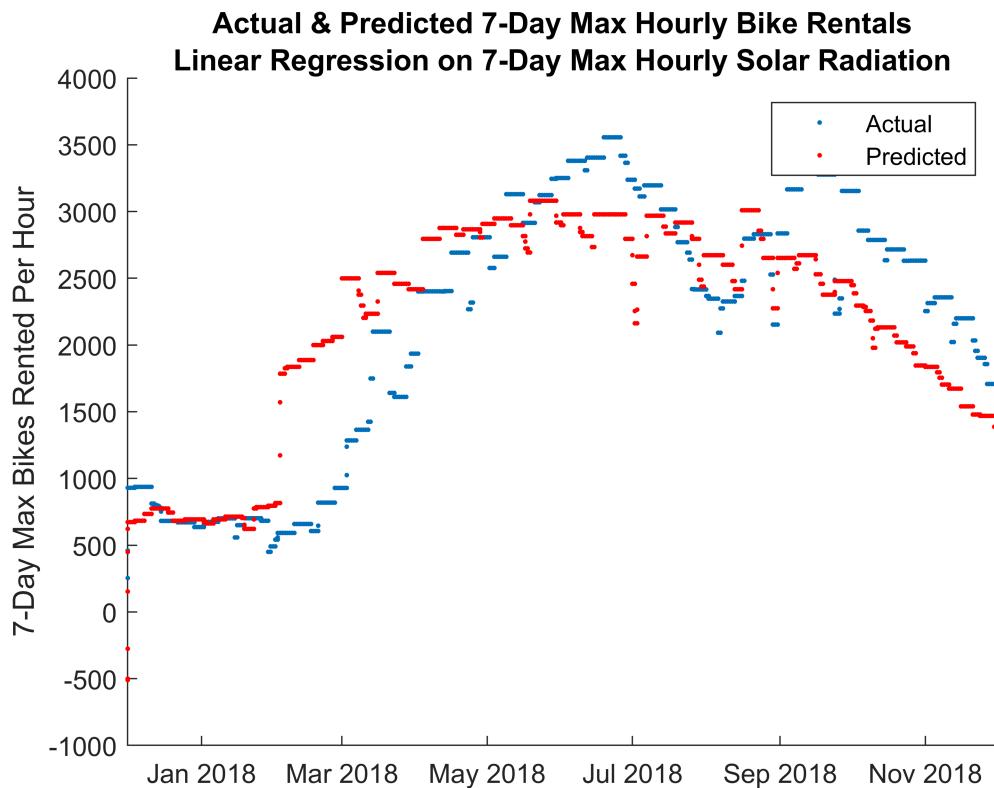
```
hold on
```

```
scatter(dateslist,predicted3,'.r')
```

```
title({'Actual & Predicted 7-Day Max Hourly Bike Rentals','Linear Regression on 7-Day Max Hourly Bike Rentals'})
```

```
ylabel('7-Day Max Bikes Rented Per Hour')
```

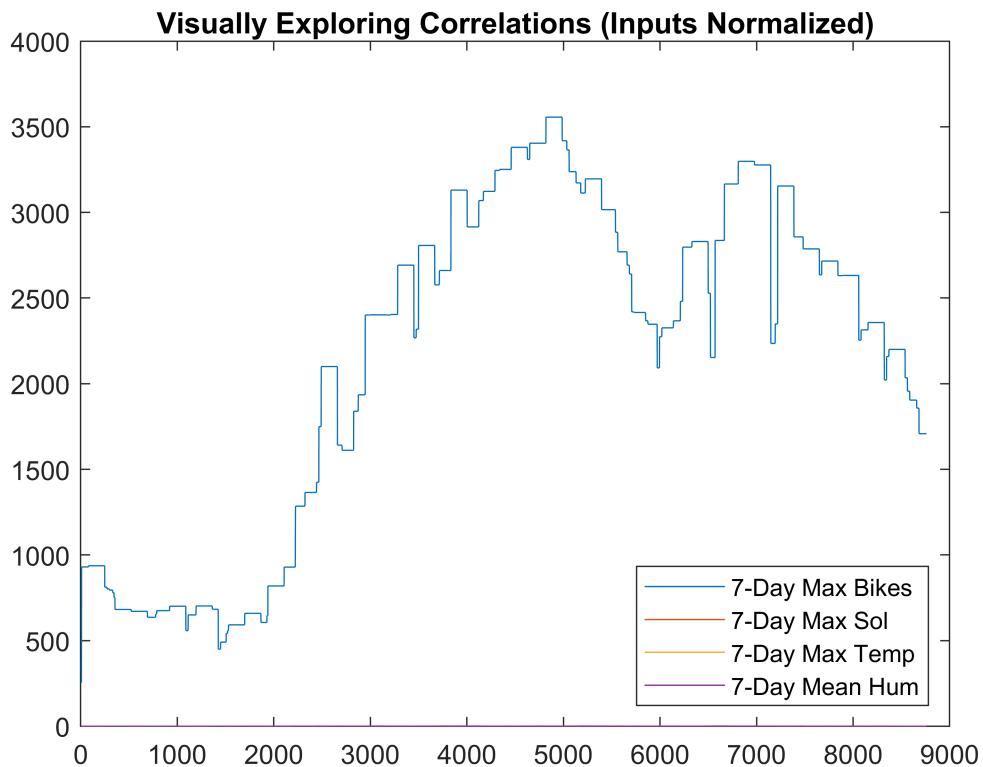
```
legend({'Actual','Predicted'})
```



```

figure
plot(sevendaymax)
hold on
plot(sevendaymax_sol)
plot(sevendaymax_temp)
%plot(2*sevendaymean_vis)
plot(sevendaymean_hum)
title('Visually Exploring Correlations (Inputs Normalized)')
legend({'7-Day Max Bikes','7-Day Max Sol','7-Day Max Temp', '7-Day Mean Hum'},'Location','South')

```



```
mdl_linear4 = fitlm([sevendaymax_sol, sevendaymax_temp, sevendaymean_hum], sevendaymax)
```

```
mdl_linear4 =
Linear regression model:
y ~ 1 + x1 + x2 + x3
```

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	1208.5	7.9017	152.94	0
x1	717.8	20.165	35.597	2.5041e-259
x2	1409.7	26.829	52.543	0
x3	485.74	31.517	15.412	6.6783e-53

Number of observations: 8760, Error degrees of freedom: 8756

Root Mean Squared Error: 448

R-squared: 0.789, Adjusted R-Squared: 0.789

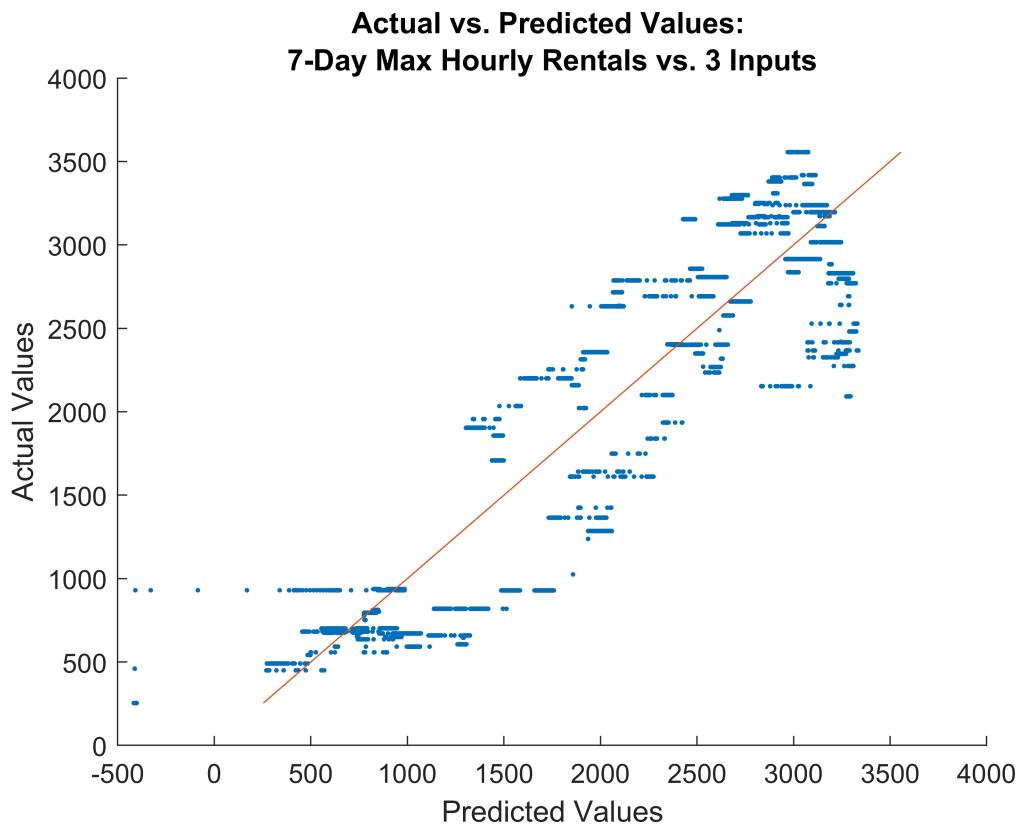
F-statistic vs. constant model: 1.09e+04, p-value = 0

```
predicted4 = predict(mdl_linear4, [sevendaymax_sol, sevendaymax_temp, sevendaymean_hum]);
RMSE_linearmodel4 = sqrt(mean((predicted4-sevendaymax).^2))
```

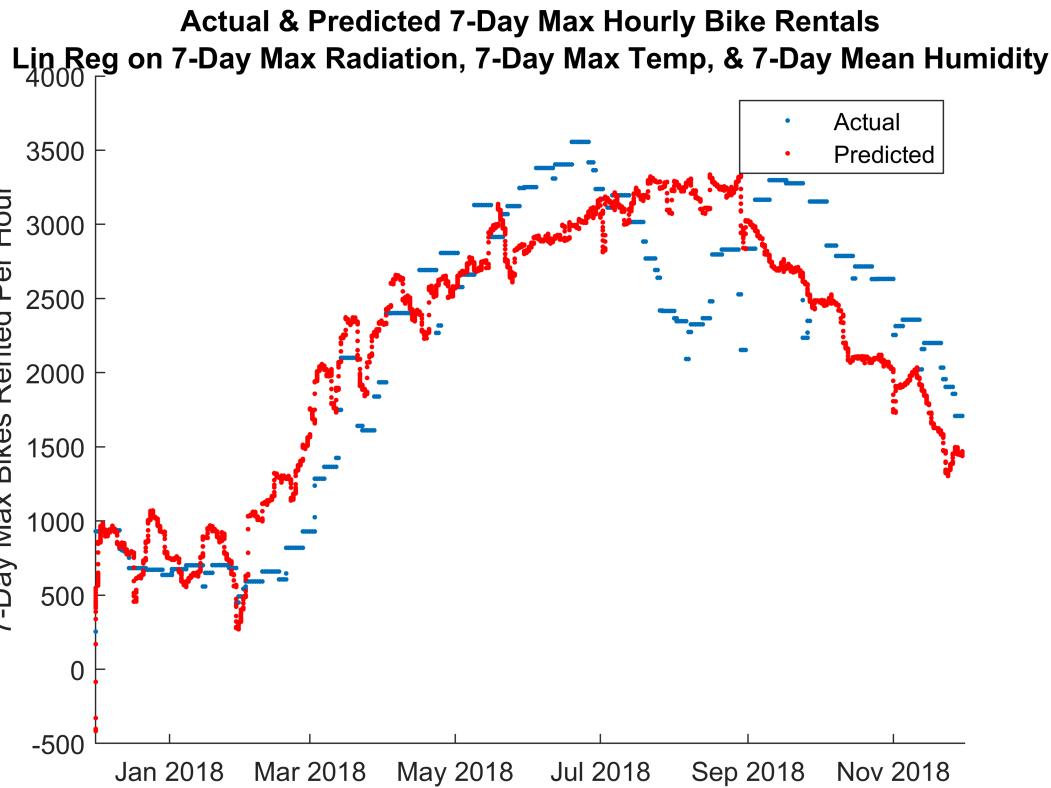
RMSE_linearmodel4 = 448.1754

```
figure
scatter(predicted4, sevendaymax, '.')
hold on
plot(sevendaymax, sevendaymax)
title({'Actual vs. Predicted Values: ';'7-Day Max Hourly Rentals vs. 3 Inputs'})
```

```
ylabel('Actual Values')  
xlabel('Predicted Values')
```



```
figure  
scatter(dateslist,sevendaymax,'.')  
hold on  
scatter(dateslist,predicted4,'r')  
title({'Actual & Predicted 7-Day Max Hourly Bike Rentals','Lin Reg on 7-Day Max Radiation, 7-Da  
ylabel('7-Day Max Bikes Rented Per Hour')  
legend({'Actual','Predicted'})
```



```
% Linear regressions with all available variables
```

```
% raw hourly counts
mdl_linear_raw = fitlm(datasetInputs,datasetOutputs)
```

```
mdl_linear_raw =
Linear regression model:
y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15
```

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-1052.2	93.778	-11.22	5.1332e-29
x1	-266.48	32.09	-8.3042	1.1546e-16
x2	314.19	8.4216	37.308	7.8268e-283
x3	483.6	104.38	4.6332	3.6523e-06
x4	-540.23	50.274	-10.746	9.1165e-27
x5	70.763	18.78	3.7681	0.0001656
x6	6.0309	9.7223	0.62031	0.53507
x7	311.86	110.41	2.8246	0.0047454
x8	-143.88	13.345	-10.781	6.2557e-27
x9	-1000.6	74.493	-13.433	9.8627e-41
x10	156.69	49.137	3.1888	0.0014336
x11	-382.49	25.948	-14.741	1.3467e-48
x12	-202.43	17.641	-11.475	2.8939e-30
x13	-146.39	11.95	-12.25	3.2189e-34
x14	-62.838	10.772	-5.8336	5.6163e-09
x15	468.8	13.279	35.305	2.2874e-255

Number of observations: 8760, Error degrees of freedom: 8744
 Root Mean Squared Error: 431
 R-squared: 0.554, Adjusted R-Squared: 0.553
 F-statistic vs. constant model: 724, p-value = 0

```
% 1-day average
mdl_linear_1avg = fitlm(onedayavg_in,onedayavg)
```

`mdl_linear_1avg =`
 Linear regression model:
 $y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15}$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	7.6489e+05	52398	14.598	1.0549e-47
x1	-1.0359	0.071065	-14.576	1.4343e-47
x2	-6.9071	5.2283	-1.3211	0.18651
x3	-35.609	3.0002	-11.869	3.0258e-32
x4	-15.954	0.87052	-18.327	1.1921e-73
x5	-39.01	3.4417	-11.334	1.4231e-29
x6	-0.035788	0.005066	-7.0643	1.738e-12
x7	54.089	3.1678	17.075	2.5149e-64
x8	460.1	10.554	43.597	0
x9	-114	4.9437	-23.06	2.8697e-114
x10	-11.8	5.4583	-2.1618	0.030664
x11	-651.59	21.06	-30.939	1.5274e-199
x12	-410.22	14.435	-28.419	5.2497e-170
x13	-268.41	10.032	-26.755	1.2081e-151
x14	-134.95	9.7702	-13.813	6.0106e-43
x15	951.17	13.057	72.85	0

Number of observations: 8760, Error degrees of freedom: 8744
 Root Mean Squared Error: 170
 R-squared: 0.832, Adjusted R-Squared: 0.832
 F-statistic vs. constant model: 2.9e+03, p-value = 0

```
% 7-day avg
mdl_linear_7avg = fitlm(sevendayavg_in,sevendayavg)
```

`mdl_linear_7avg =`
 Linear regression model:
 $y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15}$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	9.7489e+05	38407	25.383	3.317e-137
x1	-1.3202	0.052078	-25.351	7.0785e-137
x2	-17.869	3.3656	-5.3093	1.1278e-07
x3	-56.986	3.8756	-14.704	2.294e-48
x4	-16.531	1.1839	-13.963	7.7093e-44
x5	-160.58	5.5438	-28.966	3.0855e-176
x6	-0.022649	0.0057505	-3.9386	8.2586e-05
x7	66.506	4.1163	16.157	6.9561e-58
x8	898.68	13.854	64.869	0
x9	66.395	7.1918	9.232	3.2813e-20
x10	5.5389	6.8681	0.80646	0.42
x11	-705.52	15.961	-44.202	0
x12	-500.1	10.983	-45.534	0
x13	-301.4	7.9666	-37.833	3.2153e-290

x14	-121.47	12.613	-9.6305	7.6314e-22
x15	936.48	16.098	58.174	0

Number of observations: 8760, Error degrees of freedom: 8744
Root Mean Squared Error: 109
R-squared: 0.904, Adjusted R-Squared: 0.904
F-statistic vs. constant model: 5.49e+03, p-value = 0

% 30-day avg

```
mdl_linear_30avg = fitlm(thirtydayavg_in,thirtydayavg)
```

mdl_linear_30avg =
Linear regression model:
 $y \sim 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	2.0713e+06	30885	67.065	0
x1	-2.8052	0.041861	-67.012	0
x2	-2.7316	2.0329	-1.3437	0.17908
x3	-48.314	3.5637	-13.557	1.8811e-41
x4	-34.621	1.1103	-31.182	1.7066e-202
x5	-342.57	6.2549	-54.768	0
x6	-0.33586	0.0085179	-39.43	4.0915e-313
x7	71.684	3.751	19.111	8.6114e-80
x8	1115.1	14.227	78.383	0
x9	583.26	8.7326	66.791	0
x10	136.57	9.3554	14.598	1.0444e-47
x11	-905.89	14.039	-64.525	0
x12	-879.48	10.189	-86.315	0
x13	-660.02	8.7296	-75.608	0
x14	-692.24	14.86	-46.583	0
x15	787.96	24.911	31.63	5.3658e-208

Number of observations: 8760, Error degrees of freedom: 8744
Root Mean Squared Error: 53.9
R-squared: 0.974, Adjusted R-Squared: 0.974
F-statistic vs. constant model: 2.19e+04, p-value = 0

% 1-day max

```
mdl_linear_1max = fitlm(onedaymax_in,onedaymax)
```

mdl_linear_1max =
Linear regression model:
 $y \sim 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	8.4179e+05	1.356e+05	6.208	5.6095e-10
x1	-1.1434	0.18392	-6.2171	5.2949e-10
x2	-20.386	7.7117	-2.6434	0.0082214
x3	50.018	2.5369	19.716	1.0677e-84
x4	4.3136	0.69235	6.2304	4.8677e-10
x5	-14.424	5.9198	-2.4365	0.01485
x6	0.016246	0.016711	0.97216	0.331
x7	-16.95	2.4489	-6.9216	4.7806e-12
x8	215.48	10.983	19.619	6.6131e-84
x9	-18.515	1.5644	-11.836	4.4815e-32

x10	-30.288	7.5378	-4.0181	5.9159e-05
x11	-969.88	54.153	-17.91	1.7758e-70
x12	-425.08	36.226	-11.734	1.4658e-31
x13	-256.49	26.155	-9.8065	1.3801e-22
x14	-194.45	20.11	-9.6696	5.2302e-22
x15	2263.8	85.248	26.555	1.7002e-149

Number of observations: 8760, Error degrees of freedom: 8744
Root Mean Squared Error: 499
R-squared: 0.708, Adjusted R-Squared: 0.707
F-statistic vs. constant model: 1.41e+03, p-value = 0

```
% 7-day max
mdl_linear_7max = fitlm(sevendaymax_in,sevendaymax)
```

Warning: Regression design matrix is rank deficient to within machine precision.

mdl_linear_7max =
Linear regression model:
 $y \sim 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-19515	71823	-0.27172	0.78585
x1	0.024908	0.097382	0.25577	0.79813
x2	-62.226	5.8839	-10.576	5.5452e-26
x3	30.865	2.0078	15.372	1.2197e-52
x4	13.664	0.89131	15.33	2.3004e-52
x5	-34.949	4.7475	-7.3616	1.9813e-13
x6	1.2599	0.29544	4.2645	2.0246e-05
x7	-7.8118	1.9792	-3.947	7.9769e-05
x8	388.84	11.957	32.52	4.2586e-219
x9	1.5645	0.62543	2.5015	0.012385
x10	-38.13	2.8951	-13.171	3.0548e-39
x11	-1005.8	29.645	-33.93	3.5642e-237
x12	-412.46	19.769	-20.864	2.2134e-94
x13	-233.57	17.198	-13.581	1.3699e-41
x14	54.351	9.2475	5.8773	4.322e-09
x15	0	0	NaN	NaN

Number of observations: 8760, Error degrees of freedom: 8745
Root Mean Squared Error: 346
R-squared: 0.875, Adjusted R-Squared: 0.875
F-statistic vs. constant model: 4.36e+03, p-value = 0

```
% 30-day max
mdl_linear_30max = fitlm(thirtydaymax_in,thirtydaymax)
```

Warning: Regression design matrix is rank deficient to within machine precision.

mdl_linear_30max =
Linear regression model:
 $y \sim 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12 + x13 + x14 + x15$

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	0	0	NaN	NaN
x1	2.165	0.055961	38.689	2.0964e-302
x2	-77.31	7.7162	-10.019	1.6812e-23
x3	-24.829	1.8692	-13.283	7.0321e-40

x4	15.776	2.1327	7.3969	1.5227e-13
x5	-93.404	3.0229	-30.898	4.7537e-199
x6	-796.7	20.624	-38.629	1.5103e-301
x7	64.559	2.0965	30.793	8.9507e-198
x8	51.611	8.7801	5.8782	4.2987e-09
x9	0.45552	0.387	1.177	0.23921
x10	-46.869	1.736	-26.999	2.8321e-154
x11	-679.55	14.608	-46.52	0
x12	393.01	10.058	39.075	5.6849e-308
x13	140.76	10.289	13.68	3.608e-42
x14	16.687	5.6765	2.9397	0.0032937
x15	0	0	NaN	NaN

Number of observations: 8760, Error degrees of freedom: 8746

Root Mean Squared Error: 220

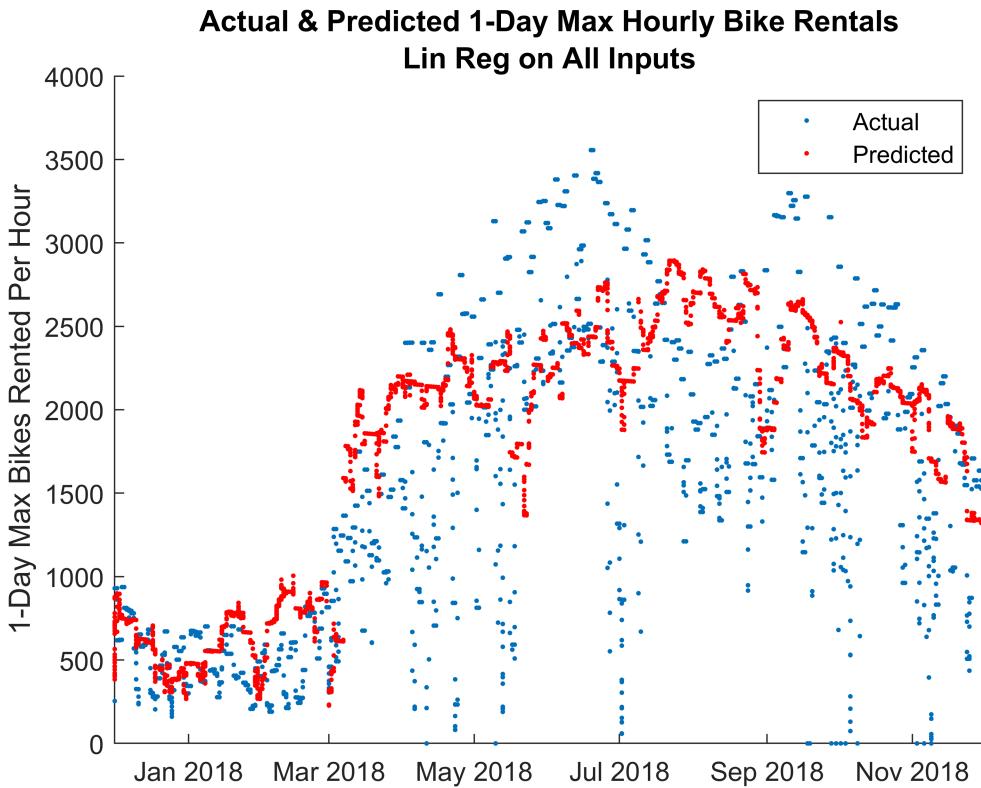
R-squared: 0.954, Adjusted R-Squared: 0.954

F-statistic vs. constant model: 1.39e+04, p-value = 0

```

predicted_1avg = predict(mdl_linear_1avg,sevendayavg_in);
predicted_7avg = predict(mdl_linear_7avg,sevendayavg_in);
predicted_30avg = predict(mdl_linear_30avg,thirtydayavg_in);
predicted_1max = predict(mdl_linear_1max,sevendaymax_in);
predicted_7max = predict(mdl_linear_7max,sevendaymax_in);
predicted_30max = predict(mdl_linear_30max,thirtydaymax_in);

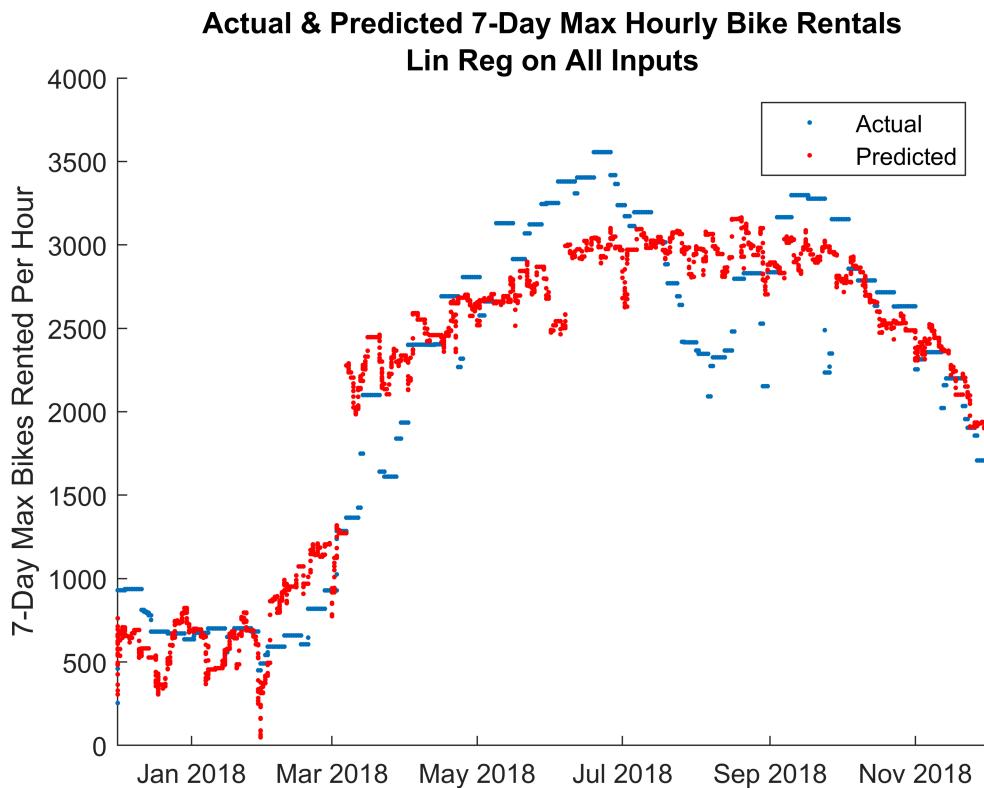
figure
scatter(dateslist,onedaymax,'.')
hold on
scatter(dateslist,predicted_1max,'.r')
title({'Actual & Predicted 1-Day Max Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('1-Day Max Bikes Rented Per Hour')
legend({'Actual','Predicted'})
```



```

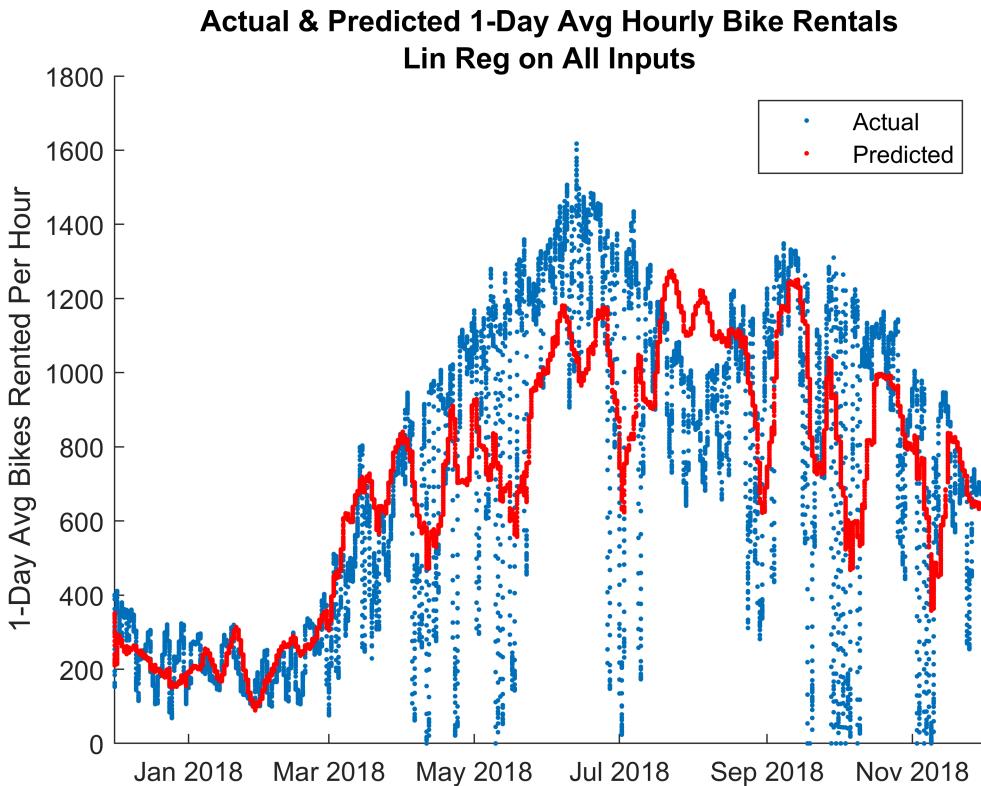
figure
scatter(dateslist,sevendaymax,'.')
hold on
scatter(dateslist,predicted_7max,'r')
title({'Actual & Predicted 7-Day Max Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('7-Day Max Bikes Rented Per Hour')
legend({'Actual','Predicted'})

```



```

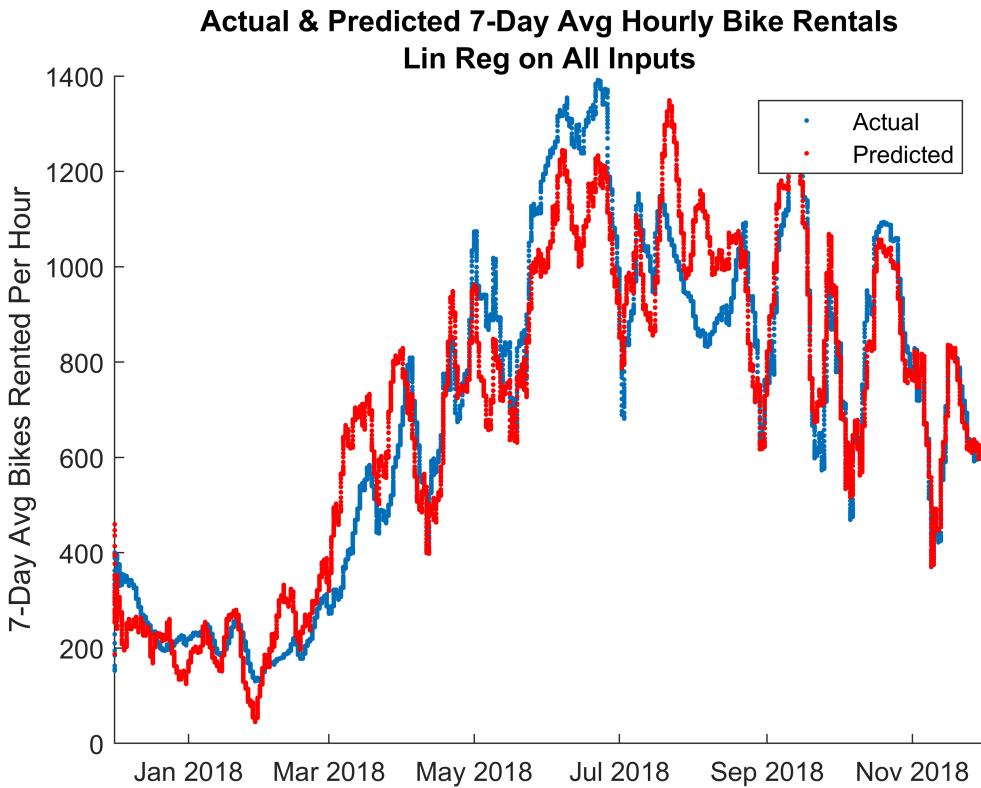
figure
scatter(dateslist,onedayavg,'.')
hold on
scatter(dateslist,predicted_1avg,'r')
title({'Actual & Predicted 1-Day Avg Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('1-Day Avg Bikes Rented Per Hour')
legend({'Actual','Predicted'})
```



```

figure
scatter(dateslist,sevendayavg,'.')
hold on
scatter(dateslist,predicted_7avg,'r')
title({'Actual & Predicted 7-Day Avg Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('7-Day Avg Bikes Rented Per Hour')
legend({'Actual','Predicted'})

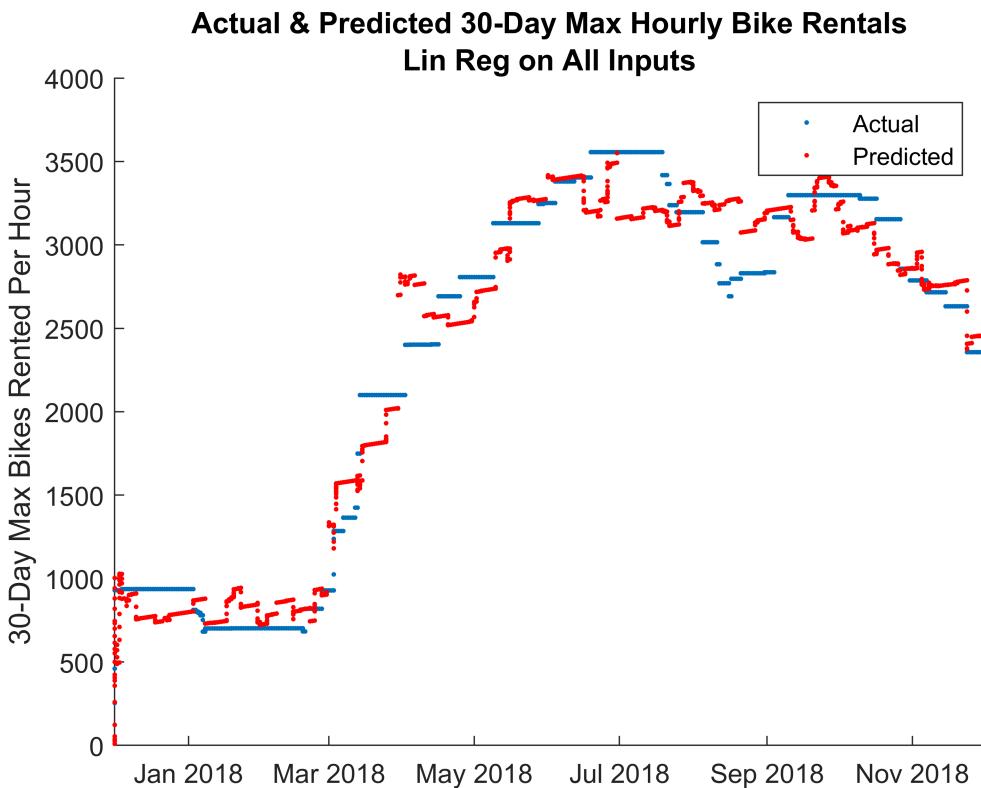
```



```

figure
scatter(dateslist,thirtydaymax,'.')
hold on
scatter(dateslist,predicted_30max,'r')
title({'Actual & Predicted 30-Day Max Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('30-Day Max Bikes Rented Per Hour')
legend({'Actual','Predicted'})

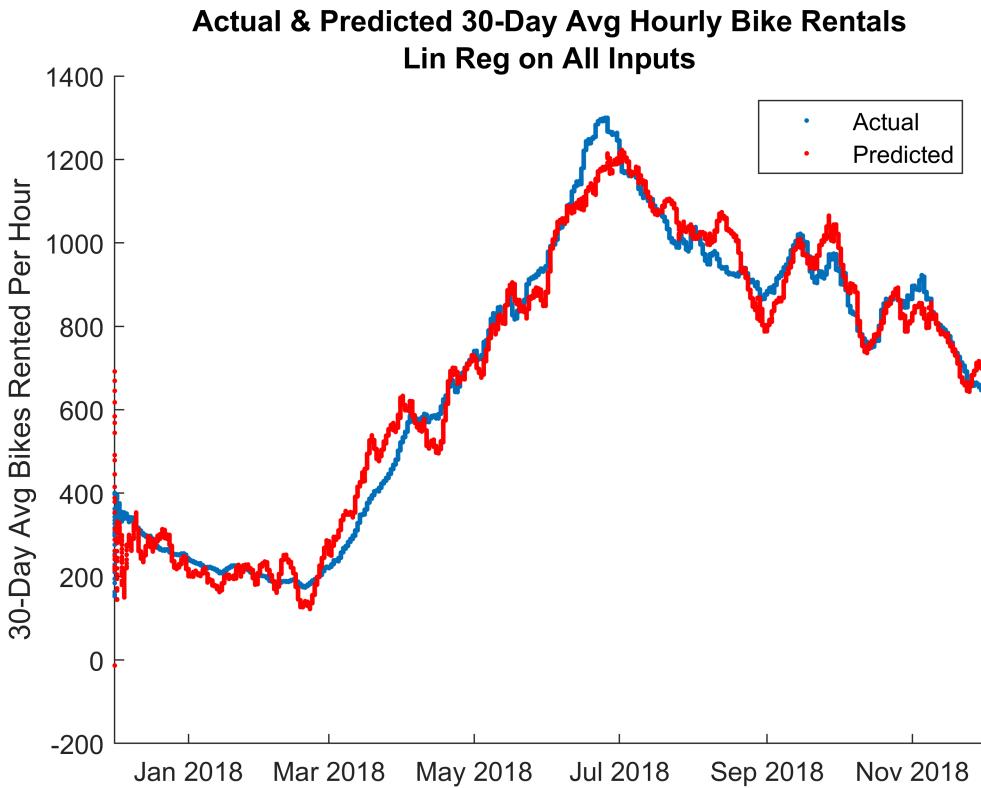
```



```

figure
scatter(dateslist,thirtydayavg,'.')
hold on
scatter(dateslist,predicted_30avg,'.r')
title({'Actual & Predicted 30-Day Avg Hourly Bike Rentals','Lin Reg on All Inputs'})
ylabel('30-Day Avg Bikes Rented Per Hour')
legend({'Actual','Predicted'})

```



```
% Compare coefficients
coeff_raw = mdl_linear_raw.Coefficients.Estimate(2:end);
coeff_1max = mdl_linear_1max.Coefficients.Estimate(2:end);
coeff_7max = mdl_linear_7max.Coefficients.Estimate(2:end);
coeff_30max = mdl_linear_30max.Coefficients.Estimate(2:end);
coeff_1avg = mdl_linear_1avg.Coefficients.Estimate(2:end);
coeff_7avg = mdl_linear_7avg.Coefficients.Estimate(2:end);
coeff_30avg = mdl_linear_30avg.Coefficients.Estimate(2:end);

figure
h=bar([coeff_raw,coeff_1avg,coeff_1max,coeff_7avg,coeff_7max,coeff_30avg,coeff_30max]);
h(1).FaceColor='#8e24aa';
h(2).FaceColor='#f4511e';
h(3).FaceColor='#ffab91';
h(4).FaceColor='#1e88e5';
h(5).FaceColor='#90caf9';
h(6).FaceColor='#43a047';
h(7).FaceColor='#a5d6a7';
legend({'Raw Counts','1-Day Avg','1-Day Max','7-Day Avg','7-Day Max','30-Day Avg','30-Day Max'})
title('Linear Regression Coefficients')
xlabel('Input Variable Index')
ylabel('Coefficient of Normalized Data')
set(gca,'XTickLabel',varnames)
xtickangle(45)
set(gca,'XTick',1:15)
```

