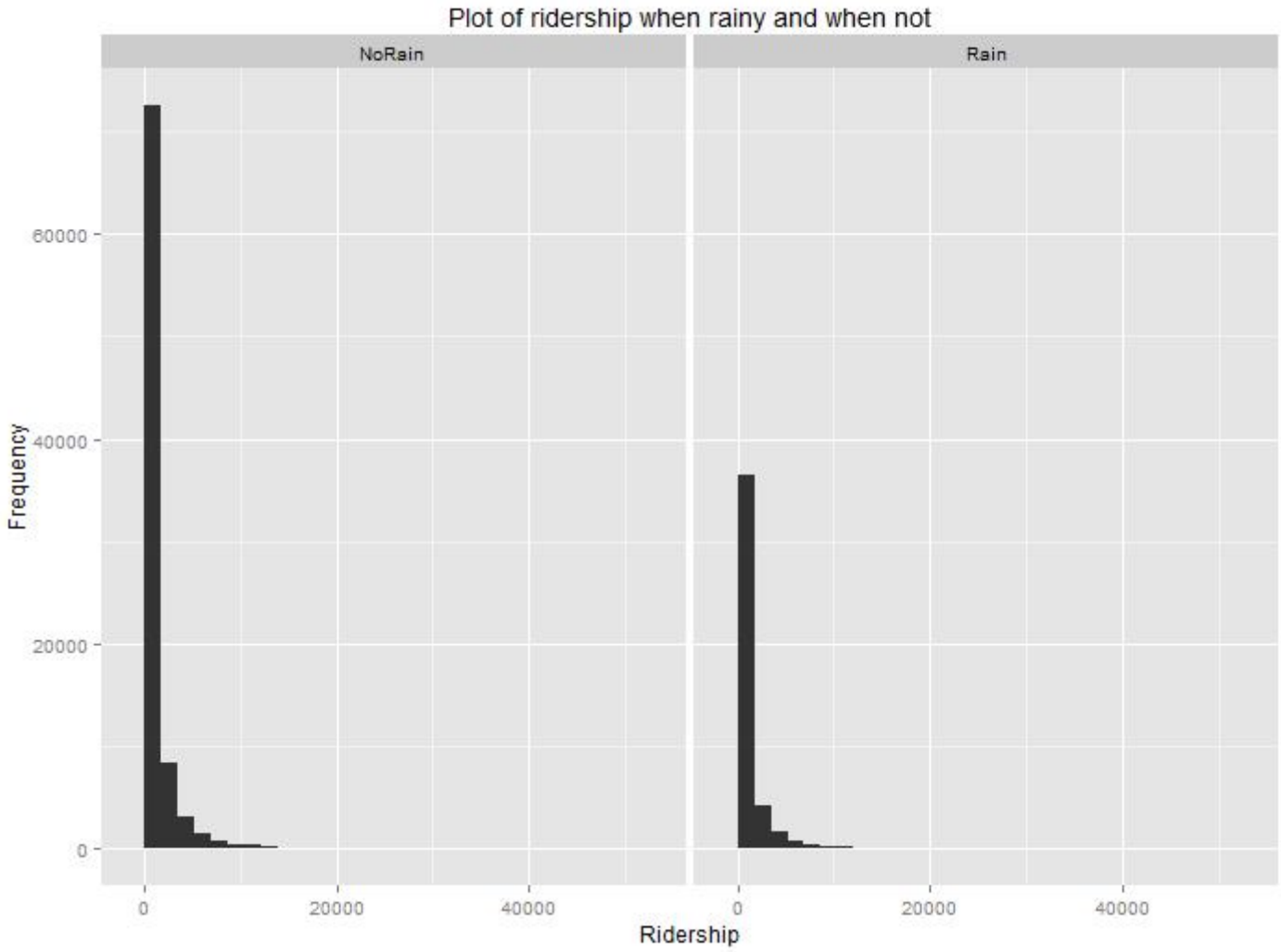
Analyzing the NYC Subway Project



Section 1. Statistical Test

1.1 Which statistical test did you use to analyze the NYC subway data? Did you use a one-tail or a two-tail P value? What is the null hypothesis? What is your p-critical value?

The Mann Whitney U-test was used on this as it does not necessarily require that the assumption of normality be met. The plot above shows both ridership on when rainy and not are not normally distributed also below is a result of a

normailty test. Also the "scipy.stats.shapiro" test for normality returns a pvalue of 0.0 for both. This means we reject the null hypothesis that that either of them was drawn from a normally distributed population. Two tailed test was used because the question was to check if the population means are significantly different from one another and not if one is higher or lower than the other. The Null hyptothesis is that there isn't a statistically significant different between the population means or that the difference between both means is 0. H0 = HA or H0 - HA = 0 95% implied confidence interval therefore P <= 0.05

1.2 Why is this statistical test applicable to the dataset? In particular, consider the assumptions that the test is making about the distribution of ridership in the two samples.

As stated earlier, this test is more appropriate as it is a non-parametric test and has a greater efficiency when it comes to non-normal distributions. It is applicable because we are comparing two independent population means. Using Python SciPy, P-Value is 0.024999912793489721

1.3 What results did you get from this statistical test? These should include the following numerical values: p-values, as well as the means for each of the two samples under test.

Using R's wilcox.test(td$ENTRIESn_hourly[td$rain == 0], td$ENTRIESn_hourly[td$rain == 1]), P-Value is 0.04988. Mean for entries with rain is 1105.45(rounded to 2 decimal places) and mean for entries without rain is 1090.29(rounded 2 dec places)

1.4 What is the significance and interpretation of these results?

In both instances, P-Value is lower than alpha of 0.05 The difference in population means is significant and not due to chance and as a result, we reject the Null Hypothesis.

Section 2. Linear Regression

2.1 What approach did you use to compute the coefficients theta and produce prediction for ENTRIESn_hourly in your regression model:
OLS Using statsmodels Station("UNIT").

Gradient descent (as implemented in exercise 3.5)
OLS using Statsmodels
Or something different?

2.2 What features (input variables) did you use in your model? Did you use any dummy variables as part of your features?

Rain, hour, actual day of week and peak/off-peak(rush hour) and a few of the other pressure, windspeed and temperature related features . Yes, I created a 3 dummy variables "newunit, dow and peak".

2.3 Why did you select these features in your model? We are looking for specific reasons that lead you to believe that the selected features will contribute to the predictive power of your model.

Your reasons might be based on intuition. For example, response for fog might be: “I decided to use fog because I thought that when it is very foggy outside people might decide to use the subway more often.”

Your reasons might also be based on data exploration and experimentation, for example: “I used feature X because as soon as I included it in my model, it drastically improved my R2 value.”

2.4 What are the coefficients (or weights) of the non-dummy features in your linear regression model?

It kind of makes sense that people who would otherwise have walked or cycled to their destinations would jump on the tube instead if it started to rain, if it got foggy or if the temperature dropped. these granted are not as particularly powerful predictors as one would expect. Most were selected on intuition some however, were not. My experience with public transportation in London reminded me of how more people tend to ride the tube during peak hours (Rush hour to our cousins across the pond) and it's more likely to find a seat at weekends than it is during the week.

Several coefficients for the stations, day of week also has a coefficient for each day and peak has -453.9744

and off-peak has -804.0951.

2.5 What is your model’s R2 (coefficients of determination) value?

2.6 What does this R2 value mean for the goodness of fit for your regression model? Do you think this linear model to predict ridership is appropriate for this dataset, given this R2 value?

This number is an indicator of how well our data fits the linear model, it also indicates the proportion of variance accounted for in Y(values) is due to variation in X (features) 0.5 R^2 implies that about 50% change in entries houly can be attributed to a change in the
features so while it's not the best, it is appropriate.

OLS Regression Results

Dep. Variable:	ENTRIESn_hourly	R-squared:	0.500
Model:	OLS	Adj. R-squared:	0.474
Method:	Least Squares	F-statistic:	19.76
Date:	Sun, 05 Apr 2015	Prob (F-statistic):	0.00
Time:	07:18:09	Log-Likelihood:	-88001.
No. Observations:	10000	AIC:	1.770e+05
Df Residuals:	9518	BIC:	1.804e+05
Df Model:	481		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]	
Hour	55.9731	2.570	21.784	0.000	50.936	61.010
fog	181.8301	68.614	2.650	0.008	47.332	316.328
rain	-108.1835	61.211	-1.767	0.077	-228.169	11.802
meanwindspdi	18.1004	12.439	1.455	0.146	-6.282	42.483
meantempi	-12.1503	5.260	-2.310	0.021	-22.461	-1.839
maxpressurei	612.3285	568.348	1.077	0.281	-501.755	1726.412
maxdewpti	16.1335	15.447	1.044	0.296	-14.145	46.412
mindewpti	-3.1301	16.875	-0.185	0.853	-36.209	29.949
minpressurei	-544.9574	519.089	-1.050	0.294	-1562.483	472.568
meandewpti	-9.3358	29.475	-0.317	0.751	-67.113	48.441
unit_R001	4247.4167	411.793	10.314	0.000	3440.214	5054.619
unit_R002	-1008.5852	475.022	-2.123	0.034	-1939.729	-77.441
unit_R003	-1271.7921	495.924	-2.564	0.010	-2243.910	-299.675

unit_R004	-1112.5988	496.438	-2.241	0.025	-2085.723	-139.475
unit_R005	-744.0667	456.501	-1.630	0.103	-1638.905	150.772
unit_R006	-972.9248	581.712	-1.673	0.094	-2113.205	167.356
unit_R007	-1365.0433	411.875	-3.314	0.001	-2172.407	-557.680
unit_R008	-1235.7542	439.884	-2.809	0.005	-2098.021	-373.487
unit_R009	-1256.4151	377.623	-3.327	0.001	-1996.636	-516.194
unit_R010	5011.2602	496.404	10.095	0.000	4038.202	5984.319
unit_R011	8402.0032	548.529	15.317	0.000	7326.770	9477.236
unit_R012	9138.0658	496.024	18.423	0.000	8165.753	1.01e+04
unit_R013	763.0365	456.597	1.671	0.095	-131.991	1658.064
unit_R014	2469.2156	439.947	5.613	0.000	1606.826	3331.605
unit_R015	-1015.8173	581.433	-1.747	0.081	-2155.550	123.915
unit_R016	-609.8500	456.575	-1.336	0.182	-1504.834	285.134
unit_R017	2210.9835	388.356	5.693	0.000	1449.722	2972.245
unit_R018	3323.3654	399.375	8.321	0.000	2540.505	4106.226
unit_R019	2101.9061	548.479	3.832	0.000	1026.771	3177.041
unit_R020	2427.9564	440.166	5.516	0.000	1565.138	3290.775
unit_R021	2687.7823	378.033	7.110	0.000	1946.758	3428.807
unit_R022	6545.6265	456.538	14.338	0.000	5650.715	7440.538
unit_R023	6196.2237	351.137	17.646	0.000	5507.920	6884.528
unit_R024	912.6585	548.442	1.664	0.096	-162.404	1987.722
unit_R025	2020.6347	456.661	4.425	0.000	1125.482	2915.787
unit_R027	808.8489	520.237	1.555	0.120	-210.928	1828.625
unit_R028	1466.8297	377.888	3.882	0.000	726.088	2207.571
unit_R029	4126.3982	368.439	11.200	0.000	3404.178	4848.618
unit_R030	689.6972	377.912	1.825	0.068	-51.091	1430.485
unit_R031	2437.3992	440.267	5.536	0.000	1574.382	3300.416
unit_R032	2525.7974	425.081	5.942	0.000	1692.548	3359.046
unit_R033	3486.3871	581.666	5.994	0.000	2346.198	4626.576
unit_R034	-504.8928	399.727	-1.263	0.207	-1288.443	278.657
unit_R035	1964.0497	520.669	3.772	0.000	943.427	2984.672
unit_R036	-753.8133	411.437	-1.832	0.067	-1560.318	52.691
unit_R037	-792.5165	671.214	-1.181	0.238	-2108.239	523.206
unit_R038	-1457.6237	475.143	-3.068	0.002	-2389.004	-526.243
unit_R039	-1071.6159	520.730	-2.058	0.040	-2092.359	-50.873
unit_R040	-212.1882	456.455	-0.465	0.642	-1106.938	682.562
unit_R041	1408.7761	388.366	3.627	0.000	647.496	2170.056
unit_R042	-604.0650	520.315	-1.161	0.246	-1623.993	415.863
unit_R043	-931.2385	219.192	-4.249	0.000	-1360.901	-501.576
unit_R044	1990.0204	411.862	4.832	0.000	1182.682	2797.358
unit_R045	756.1166	581.479	1.300	0.194	-383.707	1895.940
unit_R046	7455.8358	440.211	16.937	0.000	6592.929	8318.743
unit_R047	3809.5238	399.366	9.539	0.000	3026.682	4592.366
unit_R048	481.6085	822.445	0.586	0.558	-1130.559	2093.776
unit_R049	1109.6424	496.541	2.235	0.025	136.316	2082.968
unit_R050	1299.5385	475.373	2.734	0.006	367.705	2231.372
unit_R051	3289.9988	388.246	8.474	0.000	2528.954	4051.043
unit_R052	-503.9901	439.887	-1.146	0.252	-1366.262	358.282
unit_R053	2049.5926	388.044	5.282	0.000	1288.944	2810.241
unit_R054	-597.6718	343.873	-1.738	0.082	-1271.736	76.392
unit_R055	6064.4006	581.627	10.427	0.000	4924.287	7204.514
unit_R056	631.0518	581.822	1.085	0.278	-509.444	1771.548
unit_R057	3061.2510	411.634	7.437	0.000	2254.360	3868.142
unit_R058	-734.0976	411.693	-1.783	0.075	-1541.104	72.909
unit_R059	-254.5084	377.886	-0.674	0.501	-995.246	486.229
unit_R060	-779.1082	520.783	-1.496	0.135	-1799.955	241.738
unit_R061	-983.8297	425.183	-2.314	0.021	-1817.279	-150.380
unit_R062	1414.2709	671.180	2.107	0.035	98.615	2729.927
unit_R063	-564.2528	425.198	-1.327	0.185	-1397.731	269.225
unit_R064	-570.4453	581.583	-0.981	0.327	-1710.473	569.582
unit_R065	-1011.7269	440.167	-2.299	0.022	-1874.548	-148.906
unit_R066	-1240.8845	411.587	-3.015	0.003	-2047.684	-434.085
unit_R067	-342.4722	548.532	-0.624	0.532	-1417.712	732.767
unit_R068	-963.7569	388.425	-2.481	0.013	-1725.153	-202.361
unit_R069	-820.3824	411.421	-1.994	0.046	-1626.856	-13.909
unit_R070	-576.2211	475.362	-1.212	0.225	-1508.032	355.590
unit_R079	870.5199	359.503	2.421	0.015	165.817	1575.223
unit_R080	3272.6519	399.674	8.188	0.000	2489.206	4056.098

unit_R081	782.8124	621.642	1.259	0.208	-435.738	2001.363
unit_R082	152.2073	456.693	0.333	0.739	-743.008	1047.423
unit_R083	1698.5040	336.259	5.051	0.000	1039.365	2357.643
unit_R084	5330.1617	377.750	14.110	0.000	4589.691	6070.633
unit_R085	590.2979	368.358	1.603	0.109	-131.763	1312.359
unit_R086	735.5714	440.137	1.671	0.095	-127.190	1598.333
unit_R087	-633.7528	456.701	-1.388	0.165	-1528.984	261.478
unit_R088	-995.8531	621.798	-1.602	0.109	-2214.709	223.003
unit_R089	-953.6532	425.085	-2.243	0.025	-1786.911	-120.395
unit_R090	-1105.2342	474.900	-2.327	0.020	-2036.139	-174.329
unit_R091	-69.1597	399.474	-0.173	0.863	-852.213	713.894
unit_R092	499.1709	439.942	1.135	0.257	-363.209	1361.551
unit_R093	242.1109	520.726	0.465	0.642	-778.623	1262.845
unit_R094	243.7409	388.363	0.628	0.530	-517.533	1005.015
unit_R095	830.1985	425.287	1.952	0.051	-3.455	1663.852
unit_R096	253.5271	439.883	0.576	0.564	-608.738	1115.792
unit_R097	1357.2064	622.015	2.182	0.029	137.924	2576.488
unit_R098	1148.3842	388.346	2.957	0.003	387.143	1909.626
unit_R099	220.4501	411.737	0.535	0.592	-586.643	1027.543
unit_R100	-1114.3824	440.081	-2.532	0.011	-1977.036	-251.729
unit_R101	1939.0026	735.106	2.638	0.008	498.037	3379.968
unit_R102	3023.4121	411.522	7.347	0.000	2216.741	3830.084
unit_R103	-434.6133	425.225	-1.022	0.307	-1268.146	398.919
unit_R104	685.9804	548.052	1.252	0.211	-388.318	1760.279
unit_R105	1944.2675	399.343	4.869	0.000	1161.470	2727.065
unit_R106	-468.6879	440.387	-1.064	0.287	-1331.940	394.564
unit_R107	-625.7513	425.131	-1.472	0.141	-1459.099	207.597
unit_R108	4118.6326	475.415	8.663	0.000	3186.718	5050.548
unit_R109	-101.6273	411.822	-0.247	0.805	-908.887	705.632
unit_R110	1566.3225	621.520	2.520	0.012	348.011	2784.635
unit_R111	872.4646	456.345	1.912	0.056	-22.069	1766.998
unit_R112	13.8433	440.455	0.031	0.975	-849.543	877.230
unit_R113	1708.6440	475.271	3.595	0.000	777.012	2640.276
unit_R114	-587.2204	411.732	-1.426	0.154	-1394.303	219.863
unit_R115	-324.8453	388.392	-0.836	0.403	-1086.177	436.486
unit_R116	329.6396	439.839	0.749	0.454	-532.539	1191.818
unit_R117	-814.0578	475.104	-1.713	0.087	-1745.362	117.247
unit_R118	328.9278	548.170	0.600	0.548	-745.602	1403.458
unit_R119	-0.4195	425.084	-0.001	0.999	-833.674	832.835
unit_R120	152.0480	456.507	0.333	0.739	-742.802	1046.898
unit_R121	-66.0625	581.834	-0.114	0.910	-1206.580	1074.455
unit_R122	571.7046	425.127	1.345	0.179	-261.636	1405.045
unit_R123	823.9414	411.610	2.002	0.045	17.099	1630.784
unit_R124	-914.2114	496.433	-1.842	0.066	-1887.326	58.903
unit_R125	-924.6290	456.740	-2.024	0.043	-1819.937	-29.321
unit_R126	113.5832	475.112	0.239	0.811	-817.737	1044.903
unit_R127	2639.2257	377.848	6.985	0.000	1898.562	3379.889
unit_R128	-996.3955	456.704	-2.182	0.029	-1891.632	-101.159
unit_R129	-134.3844	439.786	-0.306	0.760	-996.458	727.689
unit_R130	-820.2592	440.173	-1.863	0.062	-1683.092	42.574
unit_R131	1330.9936	456.931	2.913	0.004	435.310	2226.677
unit_R132	2423.7579	581.732	4.166	0.000	1283.439	3564.077
unit_R133	-290.9618	388.127	-0.750	0.453	-1051.774	469.851
unit_R134	-1116.4884	475.122	-2.350	0.019	-2047.828	-185.149
unit_R135	-7.5899	411.640	-0.018	0.985	-814.492	799.312
unit_R136	-279.8927	424.988	-0.659	0.510	-1112.960	553.175
unit_R137	992.2219	411.902	2.409	0.016	184.807	1799.637
unit_R138	3736.7731	439.875	8.495	0.000	2874.524	4599.022
unit_R139	-123.1051	581.497	-0.212	0.832	-1262.963	1016.752
unit_R140	-86.2290	475.097	-0.181	0.856	-1017.521	845.063
unit_R141	2817.3959	378.400	7.446	0.000	2075.650	3559.141
unit_R142	1904.1544	425.484	4.475	0.000	1070.114	2738.194
unit_R143	786.3009	368.438	2.134	0.033	64.084	1508.517
unit_R144	2308.7172	622.137	3.711	0.000	1089.197	3528.237
unit_R145	-868.9330	377.763	-2.300	0.021	-1609.429	-128.437
unit_R146	-140.9747	581.723	-0.242	0.809	-1281.276	999.327
unit_R147	1112.9317	425.325	2.617	0.009	279.203	1946.660
unit_R148	-907.8702	425.683	-2.133	0.033	-1742.300	-73.440

unit_R149	-220.8601	581.443	-0.380	0.704	-1360.612	918.891
unit_R150	-611.0058	475.227	-1.286	0.199	-1542.553	320.541
unit_R151	325.9710	496.164	0.657	0.511	-646.616	1298.558
unit_R152	-34.5665	359.804	-0.096	0.923	-739.860	670.727
unit_R153	604.8135	456.458	1.325	0.185	-289.942	1499.569
unit_R154	-413.6191	496.370	-0.833	0.405	-1386.610	559.372
unit_R155	-150.4016	388.401	-0.387	0.699	-911.750	610.946
unit_R156	-174.4161	475.557	-0.367	0.714	-1106.609	757.777
unit_R157	-77.4746	399.353	-0.194	0.846	-860.291	705.342
unit_R158	2494.3584	548.544	4.547	0.000	1419.094	3569.622
unit_R159	-719.8775	548.603	-1.312	0.189	-1795.256	355.501
unit_R160	1102.4180	388.117	2.840	0.005	341.625	1863.211
unit_R161	144.5630	399.490	0.362	0.717	-638.522	927.648
unit_R163	2418.2343	377.780	6.401	0.000	1677.705	3158.764
unit_R164	1323.8464	388.363	3.409	0.001	562.572	2085.121
unit_R165	-1637.2260	377.768	-4.334	0.000	-2377.732	-896.720
unit_R166	1180.0787	425.044	2.776	0.006	346.903	2013.255
unit_R167	1439.4690	399.558	3.603	0.000	656.251	2222.687
unit_R168	4331.9261	496.349	8.728	0.000	3358.976	5304.877
unit_R169	49.6469	440.472	0.113	0.910	-813.772	913.066
unit_R170	8844.8312	359.643	24.593	0.000	8139.855	9549.808
unit_R171	-321.3438	440.348	-0.730	0.466	-1184.520	541.833
unit_R172	-622.8712	520.452	-1.197	0.231	-1643.068	397.325
unit_R173	-637.3306	411.827	-1.548	0.122	-1444.599	169.938
unit_R174	216.0967	475.023	0.455	0.649	-715.050	1147.243
unit_R175	3974.4761	456.572	8.705	0.000	3079.498	4869.454
unit_R176	1848.6840	474.966	3.892	0.000	917.649	2779.719
unit_R177	3138.7326	456.424	6.877	0.000	2244.045	4033.420
unit_R178	5194.7852	425.118	12.220	0.000	4361.463	6028.107
unit_R179	6548.6954	425.336	15.397	0.000	5714.946	7382.445
unit_R180	1111.4427	456.570	2.434	0.015	216.467	2006.418
unit_R181	217.7085	399.615	0.545	0.586	-565.621	1001.038
unit_R182	942.2188	475.681	1.981	0.048	9.782	1874.656
unit_R183	-573.9176	475.643	-1.207	0.228	-1506.278	358.443
unit_R184	-798.1677	439.889	-1.814	0.070	-1660.443	64.108
unit_R185	-662.6830	440.327	-1.505	0.132	-1525.818	200.452
unit_R186	-176.0162	388.368	-0.453	0.650	-937.300	585.268
unit_R187	-108.4011	411.765	-0.263	0.792	-915.548	698.745
unit_R188	684.2728	496.364	1.379	0.168	-288.707	1657.253
unit_R189	-153.0489	399.260	-0.383	0.701	-935.684	629.586
unit_R190	659.8596	399.590	1.651	0.099	-123.423	1443.142
unit_R191	129.2909	411.837	0.314	0.754	-677.997	936.578
unit_R192	296.4455	425.526	0.697	0.486	-537.676	1130.567
unit_R193	-22.5569	457.211	-0.049	0.961	-918.788	873.674
unit_R194	174.6841	548.571	0.318	0.750	-900.633	1250.001
unit_R195	3425.7496	359.862	9.520	0.000	2720.344	4131.155
unit_R196	-315.7476	424.860	-0.743	0.457	-1148.564	517.068
unit_R197	433.5688	496.057	0.874	0.382	-538.809	1405.946
unit_R198	1067.1504	456.517	2.338	0.019	172.279	1962.021
unit_R199	-762.3537	400.180	-1.905	0.057	-1546.791	22.084
unit_R200	280.3695	411.692	0.681	0.496	-526.634	1087.373
unit_R201	862.5115	411.653	2.095	0.036	55.584	1669.439
unit_R202	982.3263	399.355	2.460	0.014	199.506	1765.147
unit_R203	-417.1587	496.702	-0.840	0.401	-1390.800	556.482
unit_R204	98.8944	368.508	0.268	0.788	-623.460	821.249
unit_R205	392.7387	399.506	0.983	0.326	-390.378	1175.855
unit_R206	186.9290	475.011	0.394	0.694	-744.194	1118.052
unit_R207	258.2752	496.096	0.521	0.603	-714.180	1230.730
unit_R208	2401.3838	475.402	5.051	0.000	1469.494	3333.273
unit_R209	-414.0944	548.873	-0.754	0.451	-1490.004	661.815
unit_R210	-820.9826	456.617	-1.798	0.072	-1716.048	74.083
unit_R211	305.8391	440.004	0.695	0.487	-556.663	1168.341
unit_R212	245.3989	474.881	0.517	0.605	-685.469	1176.267
unit_R213	-43.8330	520.634	-0.084	0.933	-1064.386	976.720
unit_R214	-711.3316	399.647	-1.780	0.075	-1494.724	72.061
unit_R215	-416.9184	368.448	-1.132	0.258	-1139.155	305.318
unit_R216	-893.9152	388.274	-2.302	0.021	-1655.014	-132.816
unit_R217	21.1529	440.059	0.048	0.962	-841.456	883.762

unit_R218	-442.5440	411.493	-1.075	0.282	-1249.158	364.070
unit_R219	-891.8103	456.347	-1.954	0.051	-1786.348	2.727
unit_R220	-343.3468	496.130	-0.692	0.489	-1315.867	629.174
unit_R221	-548.4816	475.108	-1.154	0.248	-1479.795	382.832
unit_R222	1938.8013	440.480	4.402	0.000	1075.367	2802.236
unit_R223	1021.8834	671.238	1.522	0.128	-293.887	2337.653
unit_R224	-694.4826	496.273	-1.399	0.162	-1667.283	278.318
unit_R225	-984.6765	474.912	-2.073	0.038	-1915.606	-53.747
unit_R226	-806.6127	456.675	-1.766	0.077	-1701.793	88.567
unit_R227	-406.7194	377.857	-1.076	0.282	-1147.399	333.961
unit_R228	70.0295	399.356	0.175	0.861	-712.793	852.852
unit_R229	-1110.3481	456.880	-2.430	0.015	-2005.931	-214.765
unit_R230	-804.2307	456.494	-1.762	0.078	-1699.057	90.596
unit_R231	-328.3517	496.085	-0.662			

0.508	-1300.785	644.082				
unit_R232	-360.8958	475.112	-0.760	0.448	-1292.217	570.426
unit_R233	-587.9393	520.140	-1.130	0.258	-1607.524	431.646
unit_R234	-1211.8542	475.161	-2.550	0.011	-2143.271	-280.437
unit_R235	1943.1947	377.774	5.144	0.000	1202.677	2683.712
unit_R236	159.4755	399.655	0.399	0.690	-623.933	942.884
unit_R237	-752.7916	425.287	-1.770	0.077	-1586.445	80.862
unit_R238	897.4107	456.652	1.965	0.049	2.276	1792.546
unit_R239	-687.5619	456.682	-1.506	0.132	-1582.756	207.632
unit_R240	1833.4393	399.576	4.588	0.000	1050.185	2616.694
unit_R241	-1121.3302	388.146	-2.889	0.004	-1882.180	-360.481
unit_R242	-1008.3337	520.343	-1.938	0.053	-2028.318	11.650
unit_R243	185.4442	424.713	0.437	0.662	-647.084	1017.973
unit_R244	-409.4252	387.854	-1.056	0.291	-1169.701	350.850
unit_R246	-1124.6999	496.059	-2.267	0.023	-2097.082	-152.318
unit_R247	-1132.1191	399.459	-2.834	0.005	-1915.145	-349.093
unit_R248	2292.6540	456.591	5.021	0.000	1397.638	3187.670
unit_R249	-283.0949	456.571	-0.620	0.535	-1178.072	611.882
unit_R250	-585.8663	520.345	-1.126	0.260	-1605.854	434.121
unit_R251	-835.3005	520.265	-1.606	0.108	-1855.131	184.530
unit_R252	-945.6192	548.694	-1.723	0.085	-2021.176	129.938
unit_R253	-919.8643	456.644	-2.014	0.044	-1814.984	-24.745
unit_R254	-92.6411	399.382	-0.232	0.817	-875.516	690.233
unit_R255	-634.6625	520.221	-1.220	0.222	-1654.406	385.081
unit_R256	-657.4107	475.162	-1.384	0.167	-1588.829	274.007
unit_R257	214.3564	329.887	0.650	0.516	-432.293	861.006
unit_R258	412.3335	440.132	0.937	0.349	-450.419	1275.086
unit_R259	-493.9007	475.205	-1.039	0.299	-1425.403	437.602
unit_R260	82.5342	411.908	0.200	0.841	-724.893	889.961
unit_R261	-330.8284	456.348	-0.725	0.469	-1225.367	563.710
unit_R262	-648.4079	548.648	-1.182	0.237	-1723.875	427.059
unit_R263	-1400.7221	399.700	-3.504	0.000	-2184.219	-617.225
unit_R264	-1047.5055	440.529	-2.378	0.017	-1911.037	-183.974
unit_R265	-370.2771	621.468	-0.596	0.551	-1588.487	847.932
unit_R266	-318.5888	496.390	-0.642	0.521	-1291.619	654.442
unit_R267	-122.5216	378.244	-0.324	0.746	-863.961	618.918
unit_R268	1036.9909	439.929	2.357	0.018	174.637	1899.345
unit_R269	-536.4755	388.315	-1.382	0.167	-1297.655	224.704
unit_R270	-909.9501	425.034	-2.141	0.032	-1743.107	-76.793
unit_R271	-1089.2992	496.150	-2.196	0.028	-2061.859	-116.740
unit_R272	97.7531	388.321	0.252	0.801	-663.438	858.945
unit_R273	-413.4655	548.832	-0.753	0.451	-1489.293	662.362
unit_R274	-826.9078	425.779	-1.942	0.052	-1661.525	7.710
unit_R275	-771.6459	456.647	-1.690	0.091	-1666.772	123.480
unit_R276	63.5109	412.063	0.154	0.878	-744.221	871.243
unit_R277	-820.0034	671.156	-1.222	0.222	-2135.613	495.606
unit_R278	-1115.3208	411.397	-2.711	0.007	-1921.747	-308.895
unit_R279	-752.2167	377.962	-1.990	0.047	-1493.102	-11.331
unit_R280	-864.5223	411.941	-2.099	0.036	-1672.014	-57.031
unit_R281	-144.8332	323.148	-0.448	0.654	-778.273	488.607
unit_R282	-108.4469	495.838	-0.219	0.827	-1080.395	863.501
unit_R283	-638.7703	496.365	-1.287	0.198	-1611.751	334.211
unit_R284	-481.7945	399.168	-1.207	0.227	-1264.249	300.660

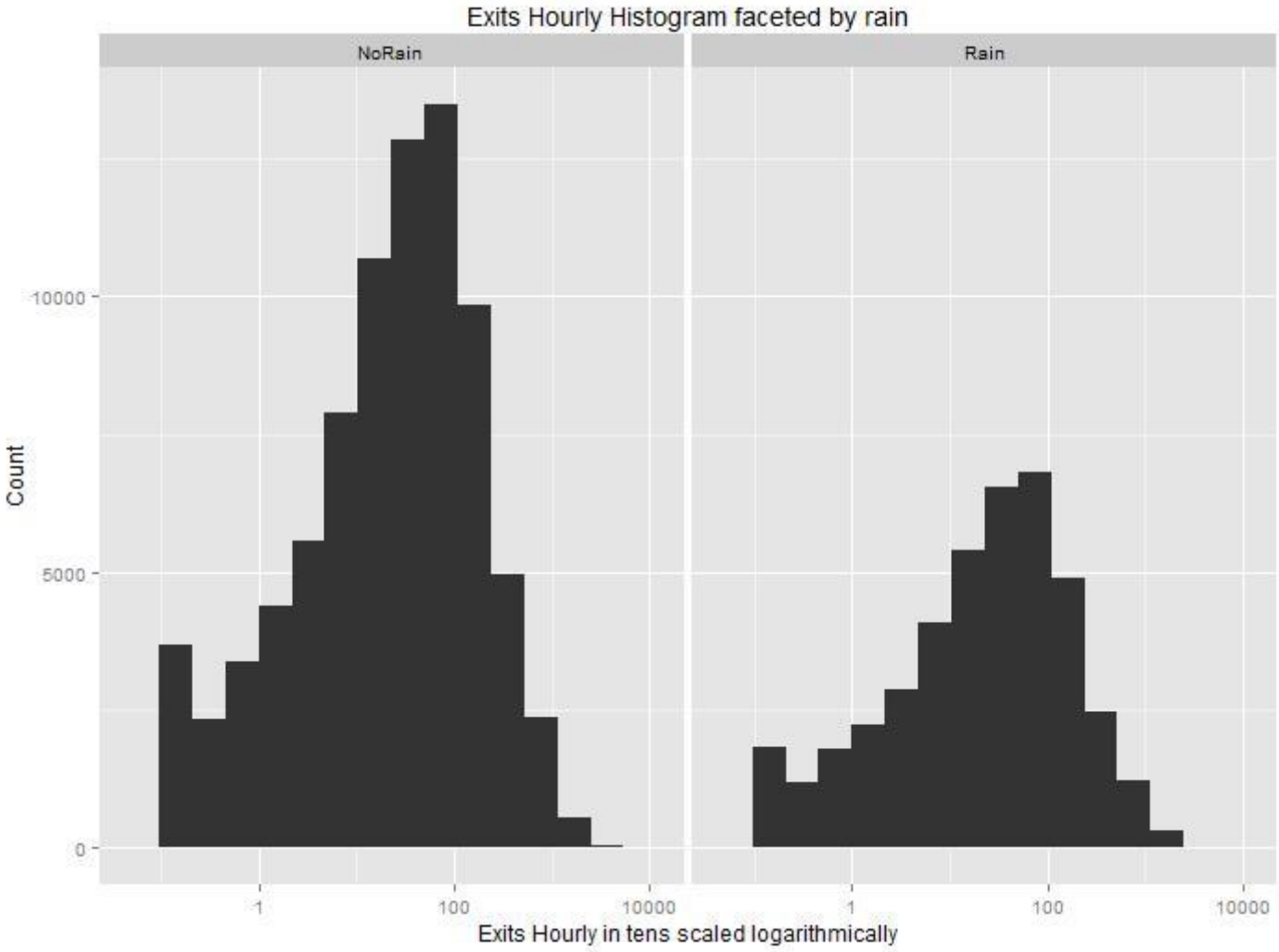
unit_R285	-920.7110	425.554	-2.164	0.031	-1754.887	-86.535
unit_R286	-884.3196	388.369	-2.277	0.023	-1645.607	-123.033
unit_R287	-671.7751	581.902	-1.154	0.248	-1812.428	468.878
unit_R288	298.1506	411.380	0.725	0.469	-508.241	1104.542
unit_R289	-824.8379	456.337	-1.808	0.071	-1719.356	69.680
unit_R290	141.0524	377.915	0.373	0.709	-599.741	881.845
unit_R291	-656.6881	548.717	-1.197	0.231	-1732.290	418.914
unit_R292	-977.6766	399.289	-2.449	0.014	-1760.368	-194.985
unit_R293	6110.5807	440.344	13.877	0.000	5247.412	6973.749
unit_R294	-554.0566	399.679	-1.386	0.166	-1337.513	229.400
unit_R295	-1190.9613	520.581	-2.288	0.022	-2211.411	-170.511
unit_R296	-962.2667	475.228	-2.025	0.043	-1893.814	-30.719
unit_R297	-555.3349	368.765	-1.506	0.132	-1278.193	167.523
unit_R298	-597.7547	368.728	-1.621	0.105	-1320.540	125.031
unit_R299	-1206.6233	520.415	-2.319	0.020	-2226.747	-186.500
unit_R300	1279.2126	368.243	3.474	0.001	557.378	2001.047
unit_R301	-570.3552	456.562	-1.249	0.212	-1465.313	324.603
unit_R302	557.2626	377.935	1.474	0.140	-183.571	1298.096
unit_R303	374.5017	377.972	0.991	0.322	-366.405	1115.408
unit_R304	-527.8284	399.748	-1.320	0.187	-1311.420	255.763
unit_R306	-1151.4233	411.960	-2.795	0.005	-1958.953	-343.894
unit_R307	-1036.5726	399.829	-2.593	0.010	-1820.322	-252.823
unit_R308	-544.3155	456.702	-1.192	0.233	-1439.548	350.917
unit_R309	-564.0156	548.196	-1.029	0.304	-1638.597	510.566
unit_R310	588.1809	475.065	1.238	0.216	-343.049	1519.411
unit_R311	-1219.7550	496.428	-2.457	0.014	-2192.859	-246.651
unit_R312	-1153.9750	440.255	-2.621	0.009	-2016.968	-290.982
unit_R313	-1269.8884	621.761	-2.042	0.041	-2488.673	-51.103
unit_R314	-1083.6383	440.395	-2.461	0.014	-1946.906	-220.370
unit_R315	-984.9147	456.557	-2.157	0.031	-1879.864	-89.966
unit_R316	-1192.8714	456.232	-2.615	0.009	-2087.184	-298.559
unit_R317	-866.4081	412.233	-2.102	0.036	-1674.472	-58.344
unit_R318	-1070.3501	548.217	-1.952	0.051	-2144.972	4.272
unit_R319	682.9164	496.253	1.376	0.169	-289.845	1655.678
unit_R320	-778.0804	351.151	-2.216	0.027	-1466.412	-89.749
unit_R321	-437.4097	456.370	-0.958	0.338	-1331.993	457.173
unit_R322	631.1405	377.834	1.670	0.095	-109.495	1371.776
unit_R323	-214.8680	440.276	-0.488	0.626	-1077.903	648.167
unit_R324	177.5780	377.859	0.470	0.638	-563.106	918.262
unit_R325	-1217.0549	520.585	-2.338	0.019	-2237.513	-196.597
unit_R326	-1097.0782	496.464	-2.210	0.027	-2070.254	-123.903
unit_R327	-32.7196	475.305	-0.069	0.945	-964.419	898.980
unit_R328	-1022.0827	456.593	-2.238	0.025	-1917.103	-127.063
unit_R329	-1262.5587	735.897	-1.716	0.086	-2705.074	179.957
unit_R330	-186.9360	411.709	-0.454	0.650	-993.974	620.102
unit_R331	-1182.9301	456.913	-2.589	0.010	-2078.577	-287.283
unit_R332	-887.9156	475.131	-1.869	0.062	-1819.274	43.443
unit_R333	-576.5074	425.333	-1.355	0.175	-1410.251	257.236
unit_R334	-421.6793	411.990	-1.024	0.306	-1229.268	385.910
unit_R335	-1133.5592	496.794	-2.282	0.023	-2107.381	-159.737
unit_R336	-1359.8013	548.480	-2.479	0.013	-2434.939	-284.663
unit_R337	-1475.2201	495.995	-2.974	0.003	-2447.477	-502.964
unit_R338	-1533.4235	425.085	-3.607	0.000	-2366.682	-700.165
unit_R339	-1121.5260	475.181	-2.360	0.018	-2052.982	-190.070
unit_R340	-964.2014	399.671	-2.412	0.016	-1747.641	-180.761
unit_R341	-815.9784	411.654	-1.982	0.047	-1622.908	-9.049
unit_R342	-884.4871	388.814	-2.275	0.023	-1646.645	-122.329
unit_R343	-966.1352	425.032	-2.273	0.023	-1799.288	-132.983
unit_R344	-1161.8241	520.643	-2.232	0.026	-2182.396	-141.252
unit_R345	-1221.8252	520.278	-2.348	0.019	-2241.681	-201.969
unit_R346	-412.1614	495.882	-0.831	0.406	-1384.196	559.873
unit_R347	-564.0068	399.608	-1.411	0.158	-1347.323	219.310
unit_R348	-1370.7245	424.772	-3.227	0.001	-2203.367	-538.082
unit_R349	-618.0033	440.043	-1.404	0.160	-1480.582	244.575
unit_R350	-1239.4428	456.490	-2.715	0.007	-2134.260	-344.626
unit_R352	-993.9226	621.820	-1.598	0.110	-2212.823	224.978
unit_R353	-801.9378	495.900	-1.617	0.106	-1774.007	170.131
unit_R354	-1162.5266	496.359	-2.342	0.019	-2135.496	-189.557

unit_R355	-1116.1497	424.970	-2.626	0.009	-1949.182	-283.118
unit_R356	-1005.4456	520.298	-1.932	0.053	-2025.341	14.450
unit_R357	-1357.8913	475.381	-2.856	0.004	-2289.739	-426.044
unit_R358	-1317.0164	520.386	-2.531	0.011	-2337.083	-296.950
unit_R359	699.7606	377.538	1.853	0.064	-40.295	1439.816
unit_R360	-1455.8192	520.479	-2.797	0.005	-2476.068	-435.570
unit_R361	-634.2879	456.629	-1.389	0.165	-1529.377	260.801
unit_R362	-783.8720	496.135	-1.580	0.114	-1756.402	188.658
unit_R363	-901.6974	548.188	-1.645	0.100	-1976.263	172.868
unit_R364	-564.8315	671.135	-0.842	0.400	-1880.400	750.737
unit_R365	-891.4670	520.603	-1.712	0.087	-1911.960	129.026
unit_R366	-786.8471	475.324	-1.655	0.098	-1718.584	144.889
unit_R367	-873.5665	440.567	-1.983	0.047	-1737.172	-9.961
unit_R368	-1359.3808	582.347	-2.334	0.020	-2500.906	-217.856
unit_R369	-1034.8887	581.578	-1.779	0.075	-2174.907	105.129
unit_R370	-970.6207	496.465	-1.955	0.051	-1943.798	2.556
unit_R371	-692.7803	411.526	-1.683	0.092	-1499.458	113.898
unit_R372	-991.9233	411.728	-2.409	0.016	-1798.998	-184.848
unit_R373	-636.1047	496.474	-1.281	0.200	-1609.299	337.089
unit_R374	-427.3722	548.431	-0.779	0.436	-1502.414	647.669
unit_R375	-1141.6781	399.730	-2.856	0.004	-1925.234	-358.123
unit_R376	-1109.0340	399.353	-2.777	0.005	-1891.851	-326.217
unit_R377	-242.0262	368.591	-0.657	0.511	-964.543	480.491
unit_R378	-192.3603	496.109	-0.388	0.698	-1164.840	780.119
unit_R379	-1011.6803	496.609	-2.037	0.042	-1985.140	-38.221
unit_R380	-817.3492	496.131	-1.647	0.099	-1789.872	155.174
unit_R381	-671.4949	399.194	-1.682	0.093	-1454.001	111.011
unit_R382	-815.6208	474.957	-1.717	0.086	-1746.638	115.397
unit_R383	-849.4389	377.840	-2.248	0.025	-1590.086	-108.792
unit_R384	-1316.8339	399.483	-3.296	0.001	-2099.905	-533.763
unit_R385	-448.3592	456.393	-0.982	0.326	-1342.986	446.267
unit_R386	-602.8497	671.354	-0.898	0.369	-1918.847	713.148
unit_R387	-362.0827	456.458	-0.793	0.428	-1256.838	532.672
unit_R388	-271.1288	439.955	-0.616	0.538	-1133.534	591.276
unit_R389	-1000.6080	425.256	-2.353	0.019	-1834.201	-167.015
unit_R390	353.7092	411.674	0.859	0.390	-453.259	1160.678
unit_R391	-298.5876	475.504	-0.628	0.530	-1230.676	633.501
unit_R392	-572.2830	411.934	-1.389	0.165	-1379.761	235.195
unit_R393	-878.8324	456.662	-1.924	0.054	-1773.986	16.322
unit_R394	-316.2271	456.505	-0.693	0.489	-1211.074	578.620
unit_R395	-680.0490	440.158	-1.545	0.122	-1542.852	182.754
unit_R396	-914.6027	440.095	-2.078	0.038	-1777.282	-51.923
unit_R397	-1074.9514	411.871	-2.610	0.009	-1882.306	-267.597
unit_R398	-711.3744	475.006	-1.498	0.134	-1642.488	219.739
unit_R399	-1099.9907	368.357	-2.986	0.003	-1822.050	-377.932
unit_R400	-1167.6813	399.428	-2.923	0.003	-1950.645	-384.718
unit_R401	-1323.5182	496.284	-2.667	0.008	-2296.340	-350.696
unit_R402	-1192.6613	456.468	-2.613	0.009	-2087.436	-297.886
unit_R403	-1155.0074	368.526	-3.134	0.002	-1877.397	-432.617
unit_R404	-451.5187	399.404	-1.130	0.258	-1234.436	331.398
unit_R405	-869.1415	475.194	-1.829	0.067	-1800.624	62.341
unit_R406	-391.6186	412.053	-0.950	0.342	-1199.331	416.093
unit_R407	-1215.1863	399.478	-3.042	0.002	-1998.248	-432.125
unit_R408	36.6981	388.165	0.095	0.925	-724.189	797.585
unit_R409	-857.3911	496.242	-1.728	0.084	-1830.131	115.349
unit_R411	-1553.0327	621.815	-2.498	0.013	-2771.923	-334.143
unit_R412	-1330.9960	425.128	-3.131	0.002	-2164.337	-497.655
unit_R413	-1046.1371	411.762	-2.541	0.011	-1853.279	-238.995
unit_R414	-1155.6072	496.086	-2.329	0.020	-2128.041	-183.173
unit_R415	-1655.5071	475.083	-3.485	0.000	-2586.771	-724.243
unit_R416	-1267.5180	581.805	-2.179	0.029	-2407.980	-127.056
unit_R417	-1413.9908	520.155	-2.718	0.007	-2433.605	-394.376
unit_R418	-1331.3025	425.053	-3.132	0.002	-2164.496	-498.109
unit_R419	-1534.7173	456.609	-3.361	0.001	-2429.768	-639.667
unit_R420	-684.6160	456.842	-1.499	0.134	-1580.124	210.892
unit_R421	-1232.0247	388.297	-3.173	0.002	-1993.170	-470.879
unit_R422	-1135.7598	411.668	-2.759	0.006	-1942.717	-328.803
unit_R423	-973.7299	388.293	-2.508	0.012	-1734.867	-212.593

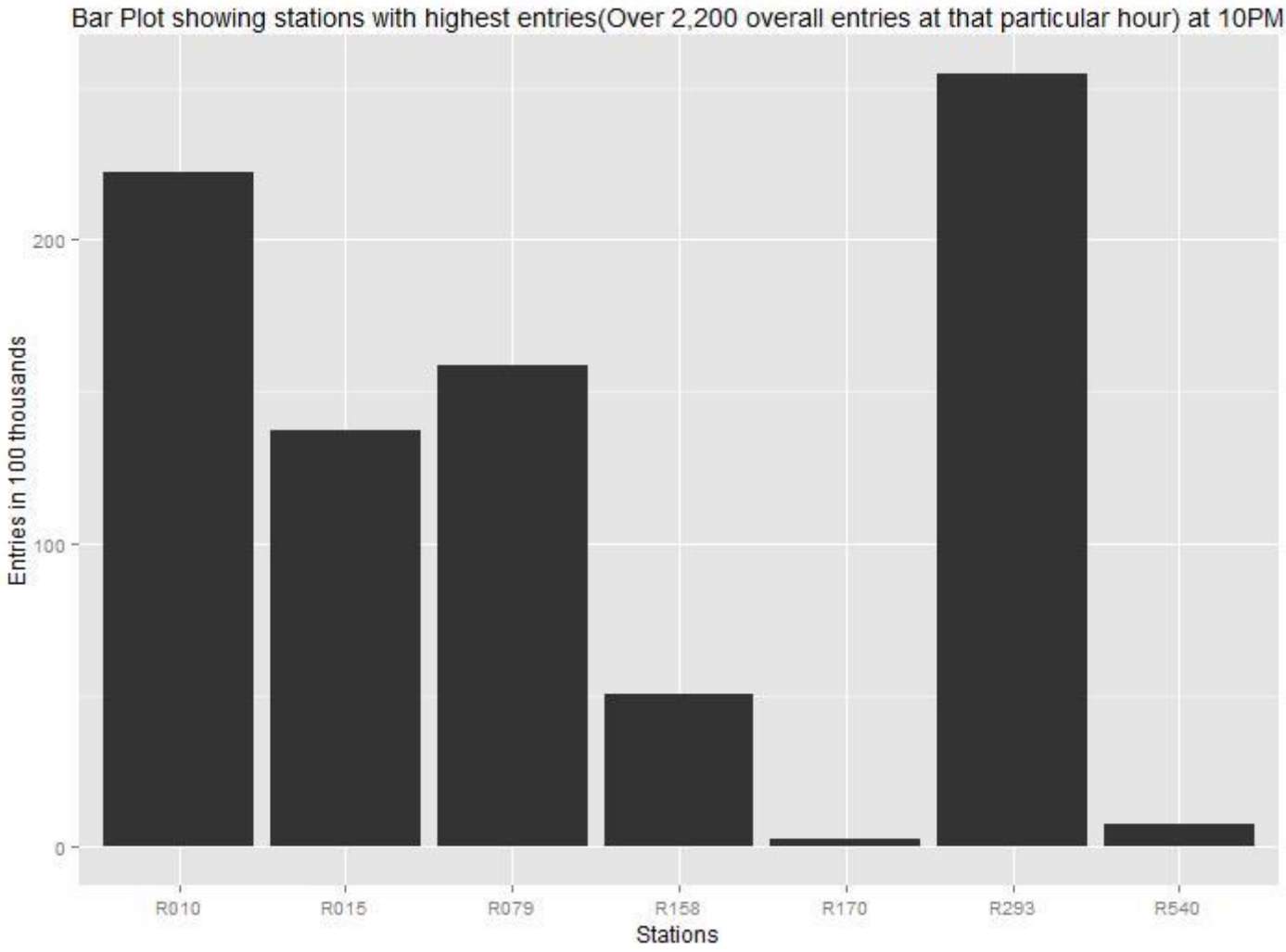
unit_R424	-1143.8728	520.214	-2.199	0.028	-2163.603	-124.143
unit_R425	-944.7946	399.763	-2.363	0.018	-1728.415	-161.174
unit_R426	-1238.3410	456.806	-2.711	0.007	-2133.779	-342.904
unit_R427	-1172.4795	456.435	-2.569	0.010	-2067.189	-277.770
unit_R428	-1472.9757	496.400	-2.967	0.003	-2446.025	-499.926
unit_R429	-777.7690	411.644	-1.889	0.059	-1584.680	29.142
unit_R430	-999.2836	440.112	-2.271	0.023	-1861.996	-136.571
unit_R431	-1254.3685	581.858	-2.156	0.031	-2394.935	-113.802
unit_R432	-1021.6438	520.481	-1.963	0.050	-2041.898	-1.390
unit_R433	-1163.6444	520.557	-2.235	0.025	-2184.047	-143.242
unit_R434	-829.9134	581.733	-1.427	0.154	-1970.234	310.407
unit_R435	-1113.4778	359.579	-3.097	0.002	-1818.330	-408.626
unit_R436	-1182.3496	496.294	-2.382	0.017	-2155.191	-209.508
unit_R437	-1028.5673	581.975	-1.767	0.077	-2169.363	112.228
unit_R438	-737.1490	581.801	-1.267	0.205	-1877.603	403.305
unit_R439	-792.8949	368.356	-2.153	0.031	-1514.950	-70.839
unit_R440	-1122.7481	519.981	-2.159	0.031	-2142.022	-103.474
unit_R441	-959.3190	581.638	-1.649	0.099	-2099.454	180.816
unit_R442	-1496.8398	440.059	-3.401	0.001	-2359.449	-634.231
unit_R443	-352.9818	671.360	-0.526	0.599	-1668.991	963.027
unit_R444	-1294.1439	581.788	-2.224	0.026	-2434.572	-153.716
unit_R445	-461.2630	548.405	-0.841	0.400	-1536.254	613.728
unit_R446	-826.3043	440.017	-1.878	0.060	-1688.832	36.224
unit_R447	-1173.7163	496.403	-2.364	0.018	-2146.772	-200.661
unit_R448	-1559.0179	360.129	-4.329	0.000	-2264.948	-853.088
unit_R449	-780.2681	425.368	-1.834	0.067	-1614.080	53.544
unit_R450	-1111.1738	399.914	-2.779	0.005	-1895.090	-327.258
unit_R451	-566.6814	495.934	-1.143	0.253	-1538.817	405.454
unit_R452	4781.7471	425.293	11.243	0.000	3948.082	5615.412
unit_R453	202.4112	548.537	0.369	0.712	-872.838	1277.661
unit_R454	-1375.5438	475.125	-2.895	0.004	-2306.890	-444.198
unit_R455	-1479.7772	475.694	-3.111	0.002	-2412.238	-547.316
unit_R456	-1153.6095	411.977	-2.800	0.005	-1961.173	-346.046
unit_R459	-2034.0553	549.197	-3.704	0.000	-3110.599	-957.512
unit_R460	124.2209	475.078	0.261	0.794	-807.032	1055.474
unit_R461	2660.6283	425.336	6.255	0.000	1826.880	3494.377
unit_R462	156.3849	548.402	0.285	0.776	-918.601	1231.370
unit_R463	1495.5701	411.584	3.634	0.000	688.777	2302.363
unit_R464	-1531.1817	475.097	-3.223	0.001	-2462.473	-599.890
unit_R468	-892.5093	496.417	-1.798	0.072	-1865.592	80.573
unit_R469	-1371.7545	496.482	-2.763	0.006	-2344.964	-398.545
unit_R535	-1005.8235	323.319	-3.111	0.002	-1639.597	-372.050
unit_R536	-1004.0111	351.355	-2.858	0.004	-1692.743	-315.280
unit_R540	-1314.7963	93.125	-14.119	0.000	-1497.341	-1132.251
unit_R541	-1417.9310	81.364	-17.427	0.000	-1577.423	-1258.439
unit_R542	-1484.2656	158.428	-9.369	0.000	-1794.819	-1173.712
unit_R543	-1495.2424	96.572	-15.483	0.000	-1684.545	-1305.940
unit_R544	-1535.2830	134.343	-11.428	0.000	-1798.624	-1271.942
unit_R545	-1492.0257	134.371	-11.104	0.000	-1755.421	-1228.630
unit_R546	-1408.8097	144.179	-9.771	0.000	-1691.431	-1126.188
unit_R547	-1395.2567	198.137	-7.042	0.000	-1783.647	-1006.866
unit_R548	-1419.4080	227.312	-6.244	0.000	-1864.988	-973.828
unit_R549	-1547.2362	59.490	-26.009	0.000	-1663.848	-1430.624
unit_R550	-1539.9940	75.202	-20.478	0.000	-1687.405	-1392.583
unit_R551	-1464.0866	112.821	-12.977	0.000	-1685.239	-1242.934
unit_R552	-1435.5675	108.553	-13.			

225	0.000	-1648.354	-1222.781			
day_0	-610.5771	1119.012	-0.546	0.585	-2804.079	1582.925
day_1	-190.5324	1143.041	-0.167	0.868	-2431.136	2050.071
day_2	-89.0160	1127.591	-0.079	0.937	-2299.334	2121.302
day_3	-24.7867	1117.621	-0.022	0.982	-2215.562	2165.989
day_4	-23.5253	1116.289	-0.021	0.983	-2211.690	2164.640
day_5	104.6999	1105.910	0.095	0.925	-2063.120	2272.520
day_6	-424.3319	1104.748	-0.384	0.701	-2589.873	1741.210
peak_off-peak	-453.9744	3913.289	-0.116	0.908	-8124.856	7216.907
peak_peak	-804.0951	3913.218	-0.205	0.837	-8474.836	6866.646

Section 3.
Visualization



This plot shows ridership by when it is raining and when it is not. As expected, fewer people exit the stations when it's raining as opposed to when it is not. Exits both when it is raining and when it is not does have the same almost normal distribution shape which seems to imply that exits follow a similar pattern though different in the actual number of people.



This is a bar plot of stations with over 2,200 entries overall at 10PM throughout the month. The entries is scaled to the 100 thousands meaning R293 has over 250thousand entries overall at 10pm throughout the time period of the dataset. These are the 7 station with the highest number of entries.

Section 4

More people ride the subway when it is raining as opposed to when it is not, this is confirmed both visually using ggplot as well as by a breakdown of the summary

```
some$rain: Not rainy
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0      38     278    1090    1111   43200
------------------------------------------------------------------------
some$rain: Rainy
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      0      41     282    1105    1103   51840
```

Using just the rain variable produces a R^2 of 0.000103179765343, this is a positive although very weak correlation. However, the MannWhitneyU test confirms that there is a statistically significant difference between the population

means.

Section 5

Since the dataset only spanned a period of a month, whatever deductions or insights gained might not be apply in general, if the one month is considered a sample, then there are multiple sampling biases that can be attributed to it and as a result, some mean observations could possibly be either under or over exaggerated. For example, if this data represents a time with very low precipitation, it is possible that other times of the year with high or very precipitation would indeed drive volume of ridership. The linear model did not seem to fit the data and wasn't very good at predicting ridership, R^2 only seemed to go up when EXITSn_hourly is added to the features but the exits is dependent on entries and as such should not be included.

Reflection

The weather it seems is not a particularly strong predictor of ridership, upon reflection, this somewhat makes sense as most people who generally take the train always planned to, and there are other options for people who don't want to be in the rain namely, waiting it out, taking the bus or staying at home. It is possible that more extreme precipitation at possibly different times of the year could force more people to plan to take the subway, the dataset is based around May

which is generally spring time.

List of Websites used

http://docs.ggplot2.org/0.9.3.1/geom_point.html

http://docs.scipy.org/doc/numpy/reference/generated/numpy.polyfit.html

http://en.wikipedia.org/wiki/Ordinary_least_squares

http://statsmodels.sourceforge.net/devel/

http://docs.ggplot2.org/0.9.3.1/geom_freqpoly.html

http://web.mta.info/lirr/about/TicketInfo/#Types