RADY CHILDREN'S HOSPITAL

CAPSTONE REPORT

Bihan Zhu, Xuehan Zhang, Linping Yu, Derrick Buntin, Feiyang Chen

Table of Contents

EXECUTIVE SUMMARY

This Rady School of Management Master of Science in Business Analytics team has taken on the client Rady Children's Hospital (RCH) in order to improve an anticipated demand model for their Emergency Department. This is a particularly important aspect of the hospital system to be able to estimate demand since bed and staff resources are limited and there are costs or losses to have too many or too little staff on the floor at any given time. In order to assist in the decision-making process for allocating the proper amount of staff months in advance, our team has performed data exploration and analysis on individual patients and staff scheduling to best address this issue.

In our analysis, we have found that the best predictors of patient volume tend to be day of the week, week number of the year, and month. Adding external data factors such as historical CDC flu data for San Diego county or weather data did not change results significantly. A second important finding is based on discovering the bottleneck of operations during busy times as revealed by the data. Upon filtering data to show only times during which all available hospital beds are occupied, we sought to verify whether beds or physicians tended to be the rate-limiting factor for allowing maximal patient throughput. After this patient flow analysis, we have found that there is an ideal number of physicians to be staffed during peak hours. There is evidence that low numbers of physicians will cause a bottleneck in patient throughput while excessively high numbers will have a plateau effect and will have no effect on increasing patient throughput. In this project paper, we will discuss the motivation for conducting analysis on these topics, elaborate on various techniques and methodology, provide evidence in a thorough analysis section, and finally give actionable recommendations for the decision makers of the RCH Emergency Department.

MODEL

We would like to recommend two models here to predict the patient volume demand on a daily and hourly scale.

DAILY MODEL

In this daily linear model, we use the weather data and flu data four months ago as well as calendar data to predict the daily patient volume.

For example, if we would like to predict the daily patient volume on October 1st, 2018, we already have the daily patient volume on June 1st, 2018. The daily patient volume on June 1st, 2018 is one of the input in the daily prediction model.

The function is like below:

$$daily.patient.arrival.number = \alpha_1 * weeknumber + \alpha_2 * activity.level + \alpha_3 * activity.level.ny + \alpha_4 * avg.temp + \alpha_5 * avg.relh + \alpha_6 * avg.sknt + \alpha_7 * avg.vsby + \alpha_8 * panic + \alpha_9 * weekday + \alpha_{10} * month + \alpha_{11} * number$$

Variables included in the model:

Weather Data:

avg_temp: average temperature for the day four months ago;

avg_relh: average humidity for the day four months ago;

avg_sknt: wind Speed in knots

avg_vsby: average wind degrees from north for the day four months ago.

**The weather data is from the Iowa Environment Mesonet.*

Flu Data:

activity_level: average flu level (0-10) in California for the day four months ago; *activity_level_ny:* average flu level(0-10) in New York for the day four months ago.

*The flu data is from ILI net. Since the flu report is on a weekly basis, so the activity level of the seven days in one certain week should be all the same.

Calendar Data:

weeknumber: the week number of the day that we want to predict;

month: the month of the day that we want to predict;

weekday: the weekday of the day that we want to predict;

panic: if the day that we want to predict is among mid-term, final weeks, the beginning of one semester, workdays right after a holiday, then we make it a panic day, which might increase the patient volume of that day.

Patient Data:

number: the patient volume for the day four months ago.

Model Performance and Interpretation

The summary of the daily linear model with 4-month-lag is as below:

Linear regression (OLS)

Data : test_lag

Response variable : actual_number_that_day Explanatory variables: weeknumber, activity_level, avg_temp, avg_relh, avg_sknt, avg_vsby, free, number, activity_level_ny, weekday, month

Null hyp.: the effect of x on actual_number_that_day is zero

Alt. hyp.: the effect of x on actual_number_that_day is not zero

	coefficient	std.error	t.value	p.value	
(Intercept)	184.729	26.079	7.083	<0.001	***
weeknumber|2	-9.31	8.024	-1.16	0.246	
weeknumber|3	13.214	8.026	1.646	0.1	.
weeknumber|4	36.578	8.352	4.379	<0.001	***
weeknumber|5	27.944	13.63	2.05	0.041	*
weeknumber|6	40.426	15.823	2.555	0.011	*
weeknumber|7	42.948	15.861	2.708	0.007	**
weeknumber|8	16.256	15.852	1.026	0.305	
weeknumber|9	10.652	17.614	0.605	0.545	
weeknumber|10	18.625	19.17	0.972	0.331	
weeknumber|11	25.991	19.256	1.35	0.177	
weeknumber|12	10.403	19.305	0.539	0.59	
weeknumber|13	-19.519	19.243	-1.014	0.311	
weeknumber|14	-8.435	21.217	-0.398	0.691	
weeknumber|15	-1.219	21.443	-0.057	0.955	

weeknumber|16	-0.14	21.479	-0.007	0.995	
weeknumber|17	-7.553	21.886	-0.345	0.73	
weeknumber|18	-0.391	22.241	-0.018	0.986	
weeknumber|19	-12.703	22.312	-0.569	0.569	
weeknumber|20	-6.437	22.303	-0.289	0.773	
weeknumber|21	-18.541	22.469	-0.825	0.409	
weeknumber|22	-10.707	21.817	-0.491	0.624	
weeknumber|23	-20.511	22.045	-0.93	0.352	
weeknumber|24	-26.7	21.935	-1.217	0.224	
weeknumber|25	-23.218	21.924	-1.059	0.29	
weeknumber|26	-24.17	21.391	-1.13	0.259	
weeknumber|27	-25.05	21.314	-1.175	0.24	
weeknumber|28	-23.248	21.306	-1.091	0.275	
weeknumber|29	-22.613	21.279	-1.063	0.288	
weeknumber|30	-20.414	21.181	-0.964	0.335	
weeknumber|31	-21.116	20.395	-1.035	0.301	
weeknumber|32	-22.53	20.617	-1.093	0.275	
weeknumber|33	-19.605	20.592	-0.952	0.341	
weeknumber|34	-16.925	20.639	-0.82	0.412	
weeknumber|35	-10.171	19.722	-0.516	0.606	
weeknumber|36	-0.004	19.579	0	1	
weeknumber|37	2.721	19.606	0.139	0.89	
weeknumber|38	6.302	19.576	0.322	0.748	
weeknumber|39	14.842	18.979	0.782	0.434	
weeknumber|40	4.298	17.917	0.24	0.81	
weeknumber|41	13.405	17.928	0.748	0.455	
weeknumber|42	13.802	17.884	0.772	0.44	
weeknumber|43	14.125	17.917	0.788	0.431	
weeknumber|44	-10.846	16.062	-0.675	0.5	
weeknumber|45	-7.304	15.988	-0.457	0.648	
weeknumber|46	-7.278	15.982	-0.455	0.649	
weeknumber|47	-19.315	15.993	-1.208	0.227	
weeknumber|48	-15.339	14.456	-1.061	0.289	
weeknumber|49	-10.138	13.525	-0.75	0.454	
weeknumber|50	-3.335	13.535	-0.246	0.805	
weeknumber|51	-10.39	13.536	-0.768	0.443	
weeknumber|52	15.912	12.837	1.24	0.215	
weeknumber|53	-6.816	12.985	-0.525	0.6	
activity_level	-1.83	0.876	-2.089	0.037	*
avg_temp	-0.377	0.262	-1.441	0.15	

avg_relh	0.172	0.078	2.212	0.027	*
avg_sknt	1.806	0.457	3.954	<0.001	***
avg_vsby	1.714	0.779	2.199	0.028	*
free	-4.746	3.801	-1.249	0.212	
number	0.378	0.025	15.111	<0.001	***
activity_level_ny	1.309	0.493	2.658	0.008	**
weekday|2	-4.95	2.721	-1.819	0.069	.
weekday|3	-25.643	2.734	-9.378	<0.001	***
weekday|4	-21.782	2.737	-7.96	<0.001	***
weekday|5	-21.389	2.69	-7.952	<0.001	***
weekday|6	-20.483	2.685	-7.629	<0.001	***
weekday|7	-16.279	2.7	-6.029	<0.001	***
month|2	12.899	13.607	0.948	0.343	
month|3	13.983	17.353	0.806	0.421	
month|4	3.04	19.148	0.159	0.874	
month|5	-19.905	20.334	-0.979	0.328	
month|6	-42.719	20.254	-2.109	0.035	*
month|7	-49.224	19.891	-2.475	0.013	*
month|8	-39.704	19.203	-2.068	0.039	*
month|9	-30.279	18.245	-1.66	0.097	.
month|10	-24.619	16.563	-1.486	0.137	
month|11	-1.956	14.454	-0.135	0.892	
month|12	5.207	11.762	0.443	0.658	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1

R-squared: 0.626, Adjusted R-squared: 0.603

F-statistic: 27.137 df(77,1248), p.value < .001 Nr obs: 1,326

Prediction error (RMSE): 25.383 Residual st.dev (RSD): 26.164

$$RMSE_{fo} = \sqrt{[\sum(z_{fi} - z_{oi})^2/N]}$$

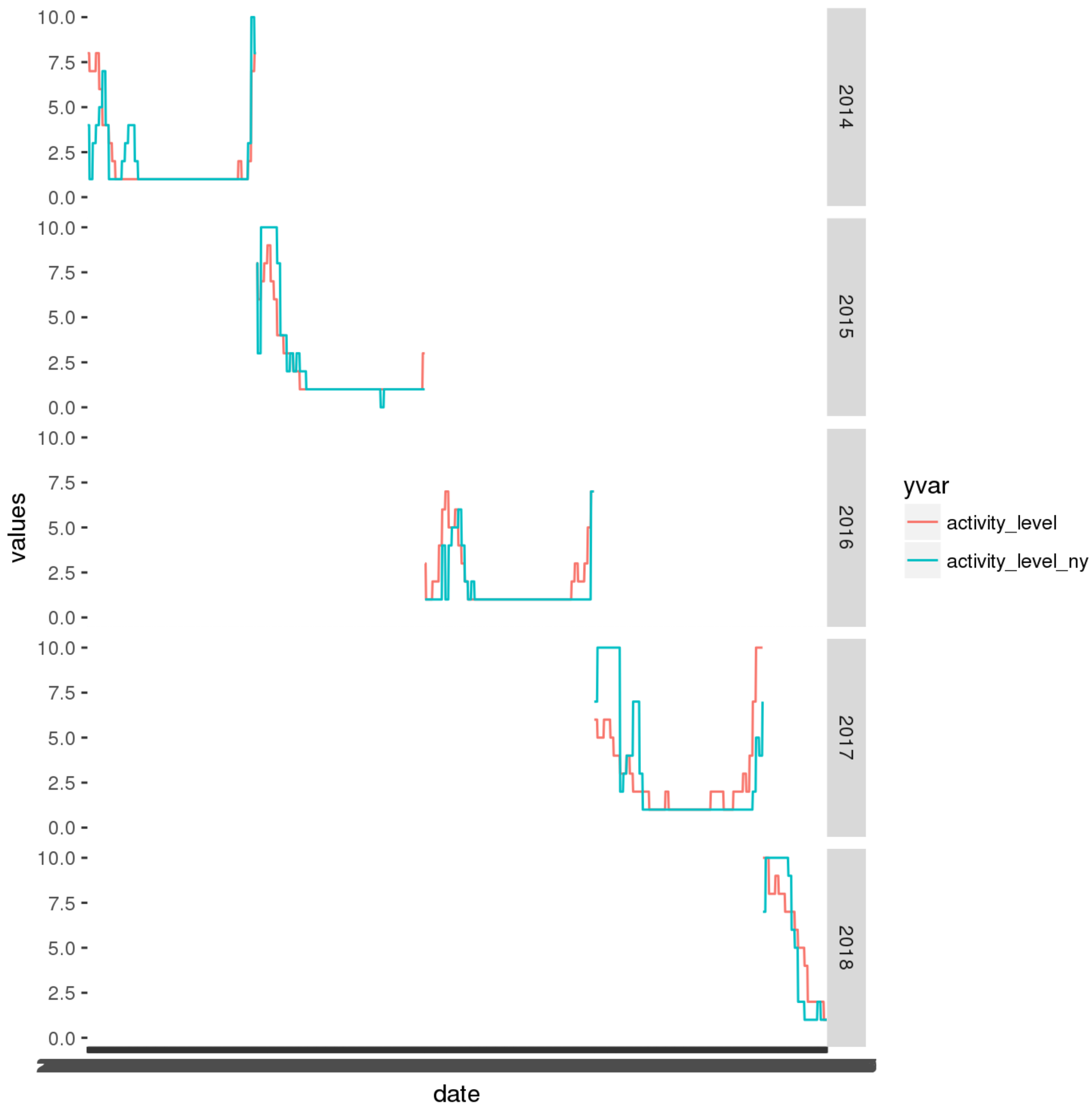
f = forecasts (expected values or unknown results); o = observed values (known results)

The RMSE can be interpreted as the error level. The RMSE of the currently-undertaken model is 84.55 a day. The new model decreases the RMSE level by 70%.

For the week number, we can see that from week 3 to week 12, the coefficients are all positive, which aligns to the flu season pattern. Month variable can also catch that pattern, but it is not as accurate as week number as week number is on a smaller granularity compared to months.

For the weekday, all the coefficients of other days compared to Monday is negative, indicating Monday tends to have the most patient volume within a week. The second largest patient volume weekday is Tuesday. From Wednesday to Sunday are on the same volume scale.

For the flu data, the flu activity levels in California and New York are both significant. However, the coefficitent of activity level in California four months ago is negative, while the coefficient of activity level in New York four months ago is positive. This means the flu condition four months ago in New York tend to increase the daily patient volume of ED department today, indicating the lag of flu spread from east to west coast. This can also be proved by the graph below.



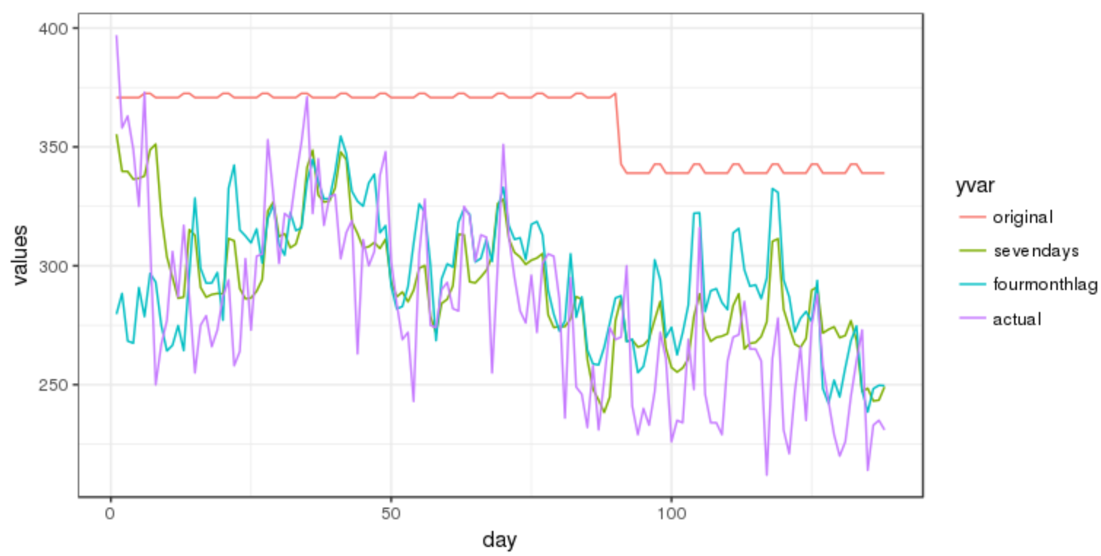
For the weather data, the most intuitive variable, average temperature of the day, is not significant. A unit increase in humidity, wind speed and visibility four mounth ago seem to increase the patient volume of ED department today. But these variables are difficult to interpret the internal relationship. In general, the weather data is not that helpful in predicting the daily patient volume four months later. It might be helpful to a more recent prediction.

Model Prediction and Validation

The RMSE for the validation set (2018) is 36.1. Since the flu season in 2018 comes significantly earlier than other years, which might be an outlier for this model, the RMSE after removing the first two weeks is 32.2.

Model Improvement & Comparison

If we have more flexibility on the prediction time window, let's say, we can use the average number in the past seven days to make a prediction, the in-sample RMSE will drop from 25 to 22 (12%), the validation sample RMSE will drop to 27.8(23%).



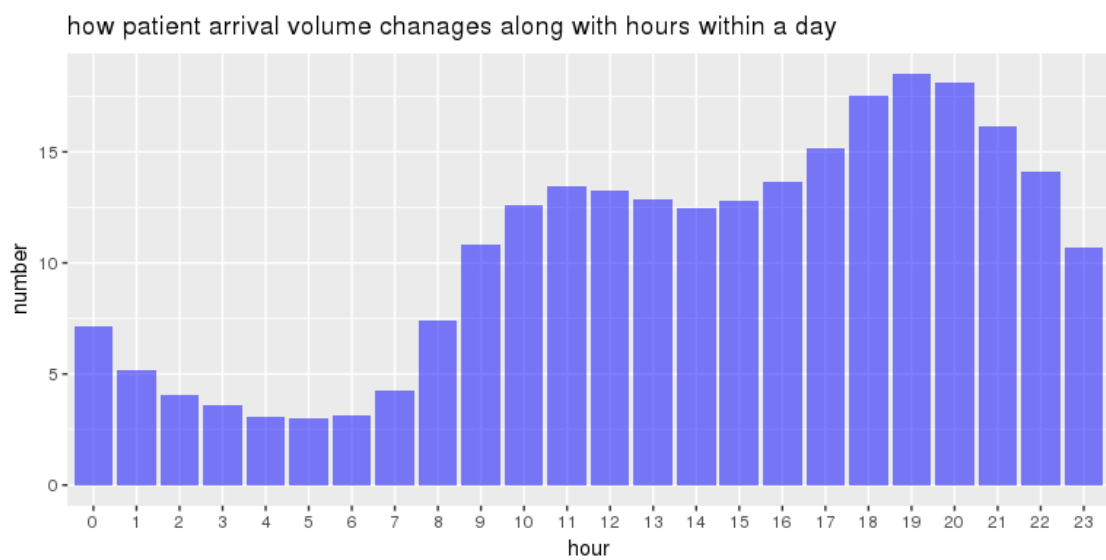
HOURLY MODEL

Linear Regression Model

The linear regresssion model is used to predict the hourly patient volume for a certain day with a time window of 4 months in advance.

From plot below, we can see that there is a distribution pattern of the hourly patient volume within a day.

We can see that there will be two peaks within a day, which starts at 11 am and 19 pm.



From intuition, we think that season and weekday could possibly affect the hourly pattern. For instance, the peak patient arrival might be earlier in the summer and in weekdays because parents have to work later. Thus, we add weekday and season to the model to see whether this intuition makes sense.

The function is as below.

$$hourly.patient.arrival.number = \alpha_1 * hour + \alpha_2 * weeknumber + \alpha_3 * weekday + \alpha_4 * total\_specimens + \alpha_5 * percent\_positive + \alpha_6 * percent\_positive{:}total\_specimens \ + \alpha_7 * total\_specimens{:}weeknumber \ + \alpha_8 * percent\_positive{:}weeknumber$$

Variables included in the model:

CDC flu Data: We use the flu data of California from ILI net as an indicator of the flu condition.

Total specimens: total specimen tested that week.

Positive percentage: the positive rate of the test result.

Calendar Data: the feature of the date itself, including hour weekday, week number.

Patient Arrival Data: patient arrival of the predicted day

Model Performance

Suppose we can get all those information mentioned above, the R-square of the linear regression model is 71.3%.

The in-sample rmse is 3.39, compared to the original model's RMSE as 4.8 (drops 29.3%). The validation rmse drops from 4.9 in the original model to 3.61 (27%).

Interpretation

Linear regression (OLS)
Data: test_regression
Response variable : number
Explanatory variables: hour, weeknumber, weekday, total_specimens, percent_positive
Null hyp.: the effect of x on number is zero
Alt. hyp.: the effect of x on number is not zero

	coefficient	std.error	t.value	p.value	
(Intercept)	3.105	0.219	14.15	<0.001	***
hour|1	-1.943	0.127	-15.312	<0.001	***
hour|2	-3.162	0.127	-24.922	<0.001	***
hour|3	-3.671	0.127	-28.936	<0.001	***
hour|4	-4.231	0.127	-33.349	<0.001	***
hour|5	-4.335	0.127	-34.169	<0.001	***
hour|6	-4.185	0.127	-32.993	<0.001	***
hour|7	-2.94	0.127	-23.174	<0.001	***
hour|8	0.255	0.127	2.007	0.045	*
hour|9	3.504	0.127	27.625	<0.001	***
hour|10	5.354	0.127	42.203	<0.001	***
hour|11	6.195	0.127	48.834	<0.001	***
hour|12	6.058	0.127	47.755	<0.001	***
hour|13	5.652	0.127	44.556	<0.001	***
hour|14	5.285	0.127	41.658	<0.001	***
hour|15	5.576	0.127	43.957	<0.001	***
hour|16	6.484	0.127	51.111	<0.001	***
hour|17	7.962	0.127	62.765	<0.001	***
hour|18	10.304	0.127	81.223	<0.001	***
hour|19	11.264	0.127	88.793	<0.001	***
hour|20	10.883	0.127	85.788	<0.001	***

hour|21	8.945	0.127	70.508	<0.001	***
hour|22	6.99	0.127	55.098	<0.001	***
hour|23	3.568	0.127	28.127	<0.001	***
weeknumber|2	-0.564	0.191	-2.954	0.003	**
weeknumber|3	-0.122	0.191	-0.639	0.523	
weeknumber|4	0.244	0.191	1.277	0.202	
weeknumber|5	0.565	0.191	2.962	0.003	**
weeknumber|6	0.92	0.191	4.813	<0.001	***
weeknumber|7	1.156	0.191	6.039	<0.001	***
weeknumber|8	0.918	0.192	4.773	<0.001	***
weeknumber|9	0.804	0.193	4.161	<0.001	***
weeknumber|10	1.312	0.195	6.714	<0.001	***
weeknumber|11	1.873	0.196	9.539	<0.001	***
weeknumber|12	1.859	0.198	9.395	<0.001	***
weeknumber|13	1.283	0.2	6.416	<0.001	***
weeknumber|14	1.541	0.202	7.641	<0.001	***
weeknumber|15	1.992	0.202	9.84	<0.001	***
weeknumber|16	2.239	0.205	10.905	<0.001	***
weeknumber|17	2.219	0.208	10.669	<0.001	***
weeknumber|18	2.509	0.208	12.069	<0.001	***
weeknumber|19	2.283	0.208	10.968	<0.001	***
weeknumber|20	2.631	0.209	12.61	<0.001	***
weeknumber|21	2.431	0.209	11.61	<0.001	***
weeknumber|22	2.64	0.213	12.41	<0.001	***
weeknumber|23	2.039	0.213	9.585	<0.001	***
weeknumber|24	1.784	0.216	8.247	<0.001	***
weeknumber|25	1.791	0.216	8.287	<0.001	***
weeknumber|26	1.848	0.217	8.51	<0.001	***
weeknumber|27	1.703	0.22	7.726	<0.001	***
weeknumber|28	1.727	0.219	7.883	<0.001	***
weeknumber|29	1.58	0.22	7.184	<0.001	***
weeknumber|30	1.661	0.22	7.533	<0.001	***
weeknumber|31	1.87	0.222	8.425	<0.001	***
weeknumber|32	1.956	0.22	8.88	<0.001	***
weeknumber|33	1.933	0.221	8.73	<0.001	***
weeknumber|34	2.011	0.221	9.085	<0.001	***
weeknumber|35	2.362	0.22	10.755	<0.001	***
weeknumber|36	2.858	0.219	13.031	<0.001	***

weeknumber|37	2.856	0.219	13.068	<0.001 ***
weeknumber|38	2.867	0.216	13.265	<0.001 ***
weeknumber|39	2.938	0.215	13.688	<0.001 ***
weeknumber|40	2.018	0.214	9.452	<0.001 ***
weeknumber|41	2.255	0.211	10.666	<0.001 ***
weeknumber|42	2.246	0.211	10.622	<0.001 ***
weeknumber|43	2.156	0.21	10.242	<0.001 ***
weeknumber|44	1.419	0.207	6.859	<0.001 ***
weeknumber|45	1.679	0.206	8.133	<0.001 ***
weeknumber|46	1.628	0.205	7.935	<0.001 ***
weeknumber|47	1.019	0.203	5.024	<0.001 ***
weeknumber|48	0.821	0.2	4.1	<0.001 ***
weeknumber|49	1.111	0.197	5.651	<0.001 ***
weeknumber|50	0.974	0.193	5.032	<0.001 ***
weeknumber|51	0.025	0.193	0.127	0.899
weeknumber|52	0.767	0.192	4.002	<0.001 ***
weeknumber|53	-0.135	0.3	-0.451	0.652
weekday|2	0.089	0.069	1.298	0.194
weekday|3	-0.683	0.069	-9.967	<0.001 ***
weekday|4	-0.831	0.068	-12.142	<0.001 ***
weekday|5	-0.866	0.068	-12.652	<0.001 ***
weekday|6	-0.89	0.068	-13.001	<0.001 ***
weekday|7	-0.699	0.068	-10.206	<0.001 ***
total_specimens	0.003	0	40.486	<0.001 ***
percent_positive	0.019	0.008	2.417	0.016 *
total_specimens:percent_positive	0	0	-9.821	<0.001 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-squared: 0.707, Adjusted R-squared: 0.707

F-statistic: 1006.072 df(84,34979), p.value < .001

Nr obs: 35,064

The hour variable has 24 factors from 0 to 23. All of then are significant. 8 am has a similar number of patient arrivals as the midnight. The lowest patient arrival occurs at 5 am. The patient arrival starts to increase since 6 am and slow down and decrease for a little bit and starts to increase again since 4 pm. These patterns all align with previous exploratory analysis and what we heard about from ED departments.

The coefficient of total specimen is positive, which means all else equal, when there is more specimen (indicating a flu season is around), there will be more patient arrival. However, the coefficient of the percentage of positive results from these specimens is negative. We think that's because the total specimen is in a large scale, but the percent positive is less than 100. They two have interactions with each other.

ARIMA Model

The linear model requires total specimen number and positive percent of the day that we would like to predict, as well as the daily patient volume of the same day, but these two variables are not available at the time. We would like to predict it first and than use it as explanatory variables in the linear regression model.

ARIMA Model for Daily Patient Volume Prediction

The only variable we input here is the daily patient volume in the time series order. The RMSE is 24.89, the MAE is 16.9

Auto arima model gives us several choices when choosing the model. Here we choose the 95% confidence interval at the lower bound.

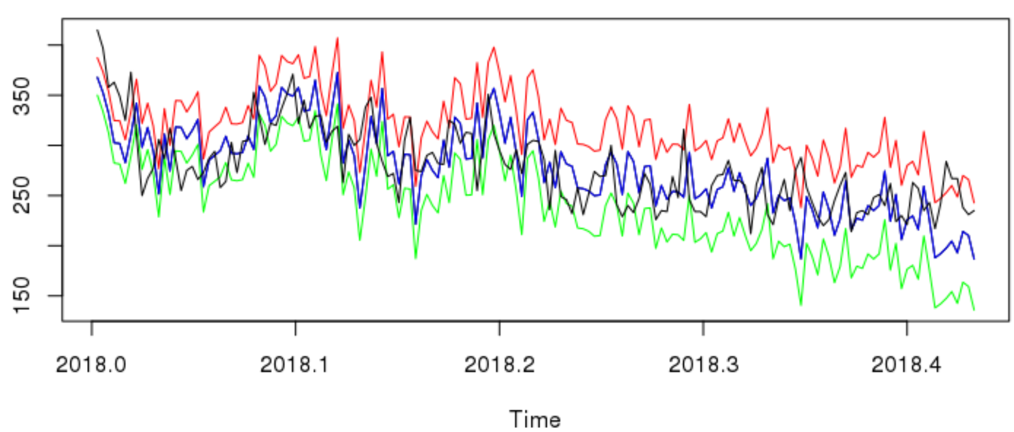
We can see from the plot beneath that the 95% has the closest line with the actual line.

95% confidence interval: black line,

99% confidence interval: green line,

80% confidence interval: red line,

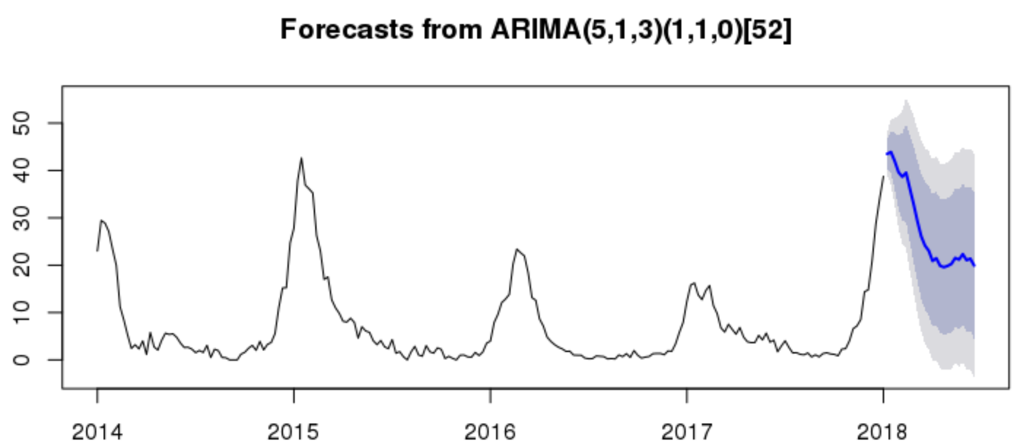
actual: blue line.



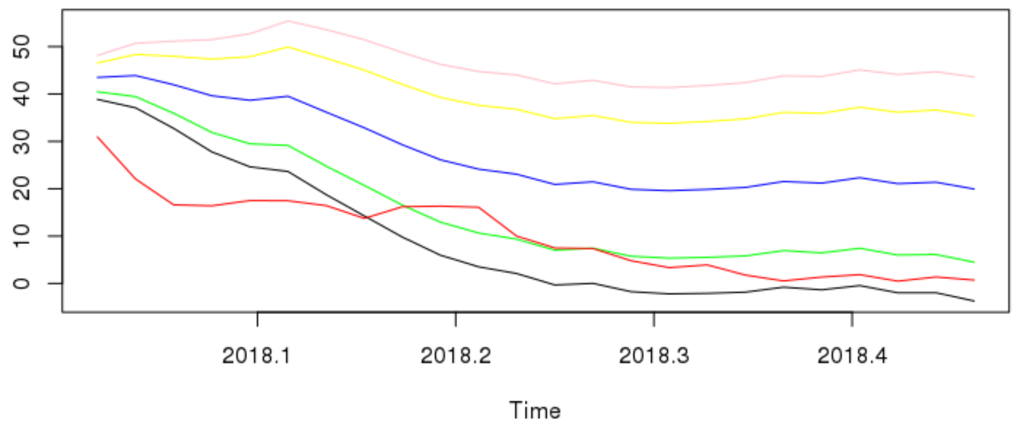
The rmse of the daily patient volume arima model is 33.9137362

ARIMA Model for Percent Positive

```
Series: test_cdc_arima
ARIMA(5,1,3)(1,1,0)[52]

Coefficients:
         ar1      ar2     ar3      ar4      ar5      ma1     ma2      ma3
      1.1564  -0.2203  0.1661  -0.2438  -0.0213  -1.0851  0.4028  -0.1877
s.e.  0.0901   0.1866  0.1620   0.1111   0.0532   0.1067  0.1931   0.1047
         sar1
      -0.4332
s.e.   0.0256

sigma^2 estimated as 5.612:  log likelihood=-357.11
AIC=734.22   AICc=735.73   BIC=764.71

Training set error measures:
                     ME     RMSE      MAE MPE MAPE      MASE          ACF1
Training set 0.07354113 1.986835 1.237927 NaN  Inf 0.2453353 -0.000432888
```



We also tried using ARIMA model to predict weekly percent positive. The trend seems correct but when we look at it on a smaller time scale, the difference between predictions and actual values seems to be unacceptable.



The rmse is 12.0814648.

Keras Model

Since ARIMA model doesn't work quite well in predicting the variables we need in the linear regression model, we think of Keras neural network model. The model we use here is (LSTM, long-short-term-model), which is a neural network model for time series data especially. The input here is just the time series number of the variable that we would like to predict. The train dataset is from 2014-2017, the validation data is in year 2018.

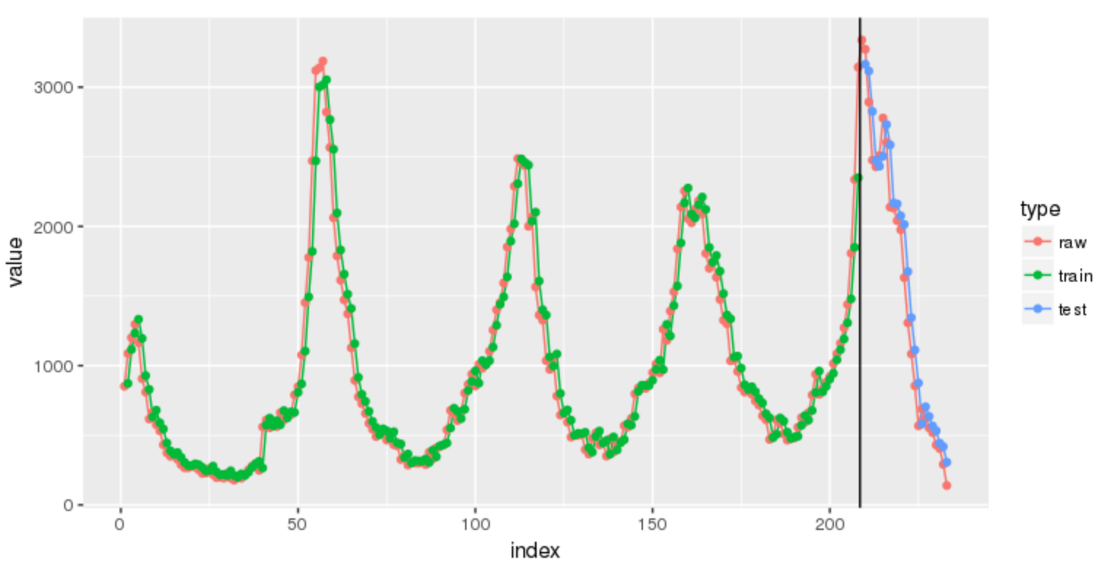
Keras Model for daily patient volume

```
1460 159

[1] "Train Score: 623.2420 MSE (24.9648 RMSE)"

[1] "Test Score: 709.4144 MSE (26.6348 RMSE)"
```

The RMSE is **RMSE_comparison_daily_patient_volume_keras**.

Keras model for total specimens

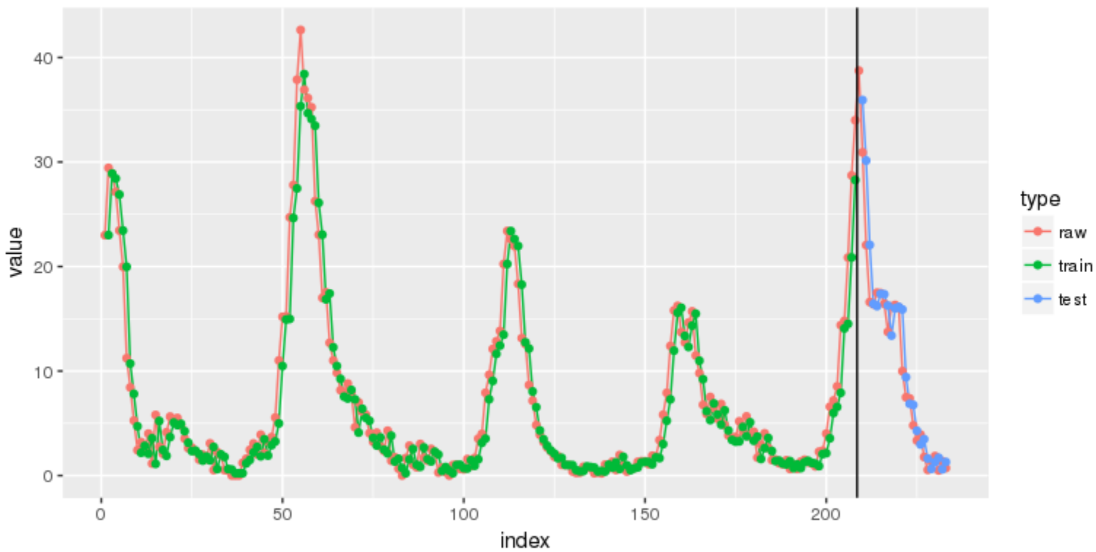
```
208 25

[1] "Train Score: 26132.3925 MSE (161.6552 RMSE)"

[1] "Test Score: 45944.2234 MSE (214.3460 RMSE)"
```



The RMSE is **RMSE_total_specimen**.

Keras Model for Percent Positive

```
208 25
[1] "Train Score: 5.9623 MSE (2.4418 RMSE)"
[1] "Test Score: 7.8683 MSE (2.8051 RMSE)"
```



Linear Model Performance with variables that we predicted

Beforehand we are using the validation data in 2018, which already gives us the total specimen and percent positive data. But when we use the model for prediction in reality, which is out of availability. Thus, we have to use the predicted value of total specimen, percent positive and daily volume to get the prediction of the hourly patient volume.

The RMSE changes from 3.61 to 3.67. The slight change in RMSE verifies our model. We can be confident to use the prediction of necessary variables on a daily base to predict the hourly patient volume distribution.

Operational Analysis

Bottleneck:

In our analysis, we calculate hourly total number of physicians and total number of patient per hour.

Patient Output Number:

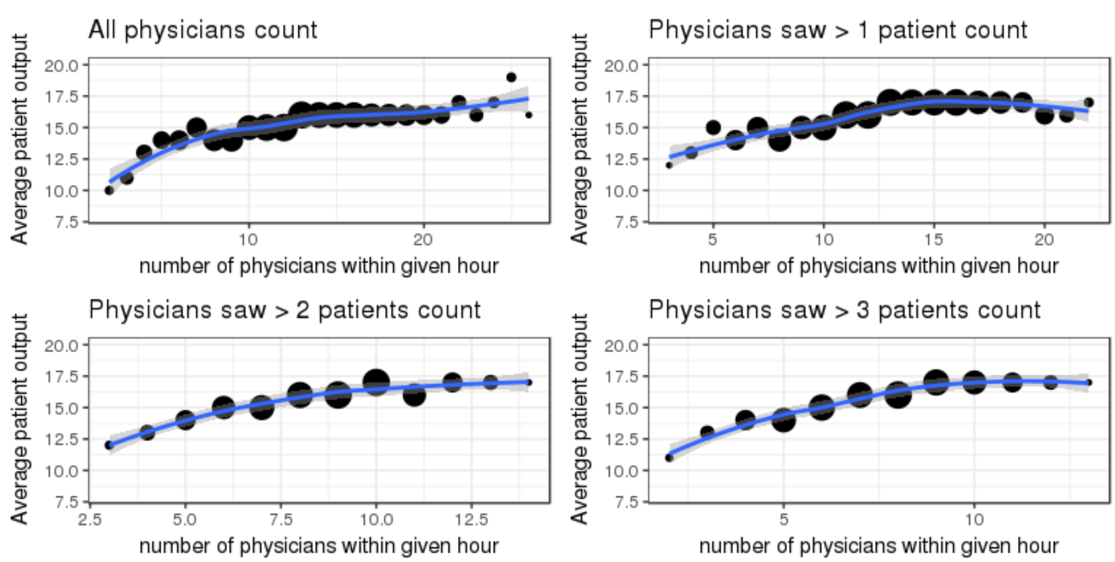
- The number of patient output is defined by calculating how many patients encountered the event **`Patient transferred to Checkout`**.

Physician Number Calculation:

- Group by the name of physician and date.
- For a particular physician, we find out the time of both first and last show up time.
- Calculate the census of all physician: the number of physicians working within each hour.

The assumption of the calculation:

- Physician do not leave the hospital early in the day, and return late in the same day (i.e. go home at 1am and back to work at 11pm)
- During the time period between fist and last show up time, the physician was working.



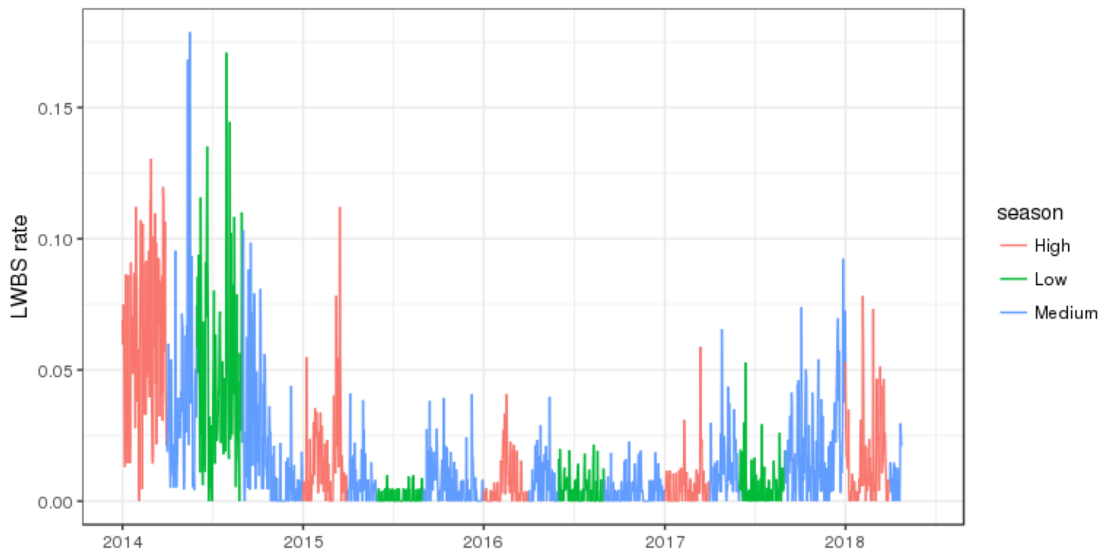
From above plots, as the number of physicians increases, the average patient output increases as well until number of physicians reach to 10. When there are more than 10 physicians working in peak hours, it will not increase number of patients get out of ED. Therefore, based on our analysis, during peak hour (from 6pm to midnight), 10 physicians will be enought.

LEFT WITHOUT BEING SEEN (LWBS)

In this section, we will take a look at the characteristics of patients labeled as Left Without Being Seen (LWBS).

We first take a look at how the LWBS rate is distributed across years.

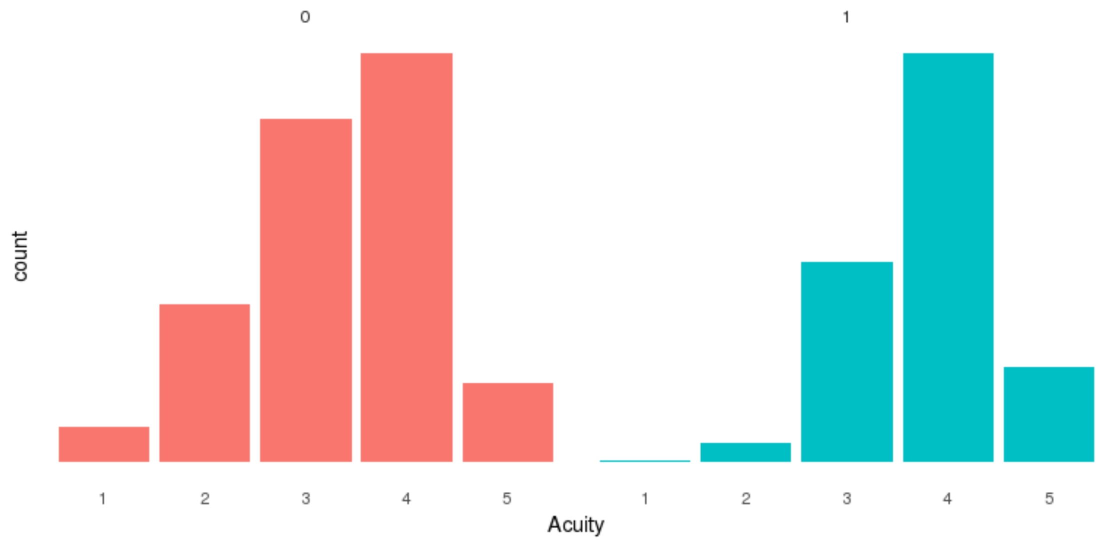
LWBS rate across years



The plot revealed that LWBS rate had great variations across the year. It was generally aroung 10% in year 2014, and then drop to around 4-5% throughout year 2015 to year 2017. But recently in the end of 2017 and beginning of 2018, the LWBS rate began to increase. Also, there was not a huge difference in LWBS rate between high and medium demand season, but the LWBS rate seemed to be lower on average during low demand season. Since the LWBS rate was significantly higher in year 2014 compared to that in other years, in the following analysis, we will exclude the data in year 2014.

How did acuity level affect the likelihood of leaving?

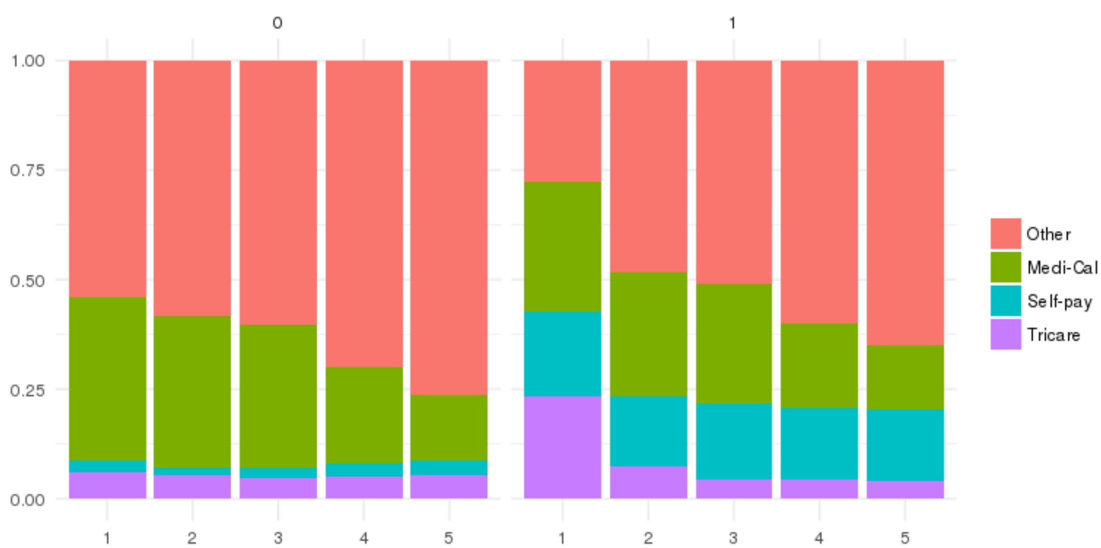
Generally, we think patient with high acuities will choose not to leave until they got fully cured.



The above graph showed the different distribution pattern of regular patient and patient with LWBS label. Low acuity patients take more percentage in the total distribution. In other words, lower acuity patients have a stronger likelihood to leave without been seen by a physician (get fully cured).

How did financial class affect the likelihood of leaving?

The financial class mostly represents the type of insurance that the patient had. Because the insurance will have direct impact on how much the patient would pay for the service, it will also impact the decision to leave because of too much a payment or lack of ability to pay.



The above graph revealed that, patients with **Tricare** as their insurance are more likely to leave when they have extreme high acuity. But this did not happen too much since we only observed very few LWBS with acuity level 1 from the previous graph. But what's more significant is, patients with no insurance, namely **self-pay** patients, are much more likely to leave in any situation.

How many patients LWBS before roomed in ED?

We want to see if people would leave because they have been waiting in the waiting room for too long.

```
          0           1 
0.001952489 0.998047511
```

The label 0 represents that there was not an event named "Patient roomed in ED" in the `event` records, label 1 represents the opposite.

There is only 0.2 percent LWBS not roomed in ED. So, it is not the case that a long wait in the waiting room cause the leave, as long as the patient had already registered. But this percentage analysis has limitations. One major of those is that we cannot catch and count those who did not registered but leave after seeing so many people waiting in the waiting room.

How many patients LWBS after seen by a physician?

```
        0         1 
0.8957045 0.1042955
```

The label 0 represents that there was not an event named "Assign Attending" in the `event` records, label 1 represents the opposite.

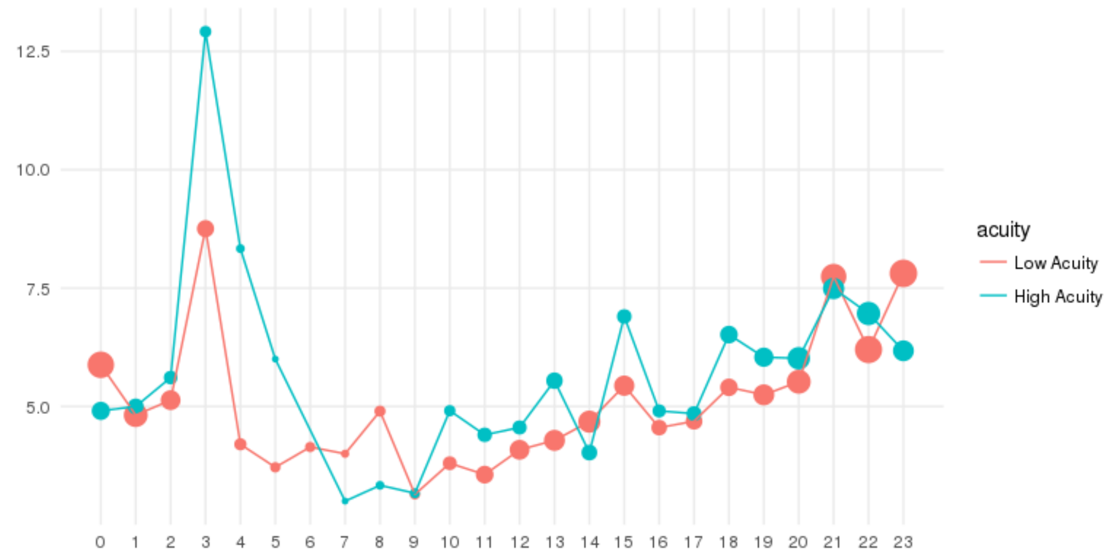
Around 10% LWBS seen patients left after seen by a physician. This percentage is siginificantly larger. One possibile reason is that patient leave because they spent too long time in bed waiting for a physician to come. But when they felt their conditions were not that bad at all, they just left to save time and money.

In the following section, we will dig into the timeline of patients labeled as LWBS.

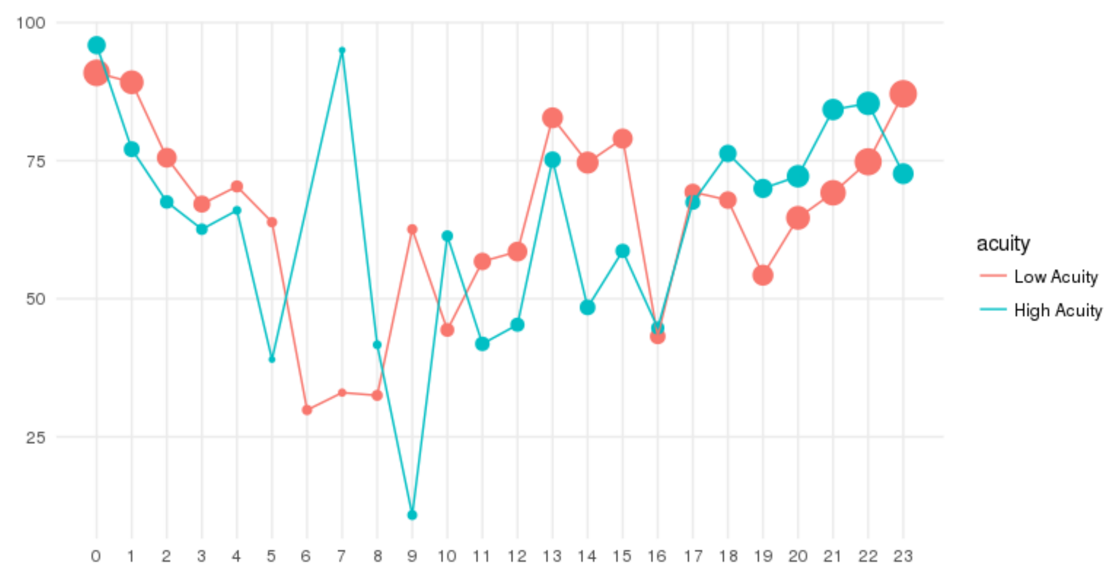
Timeline

In plots below, the size of the point represents the number of observations. The larger the points, the more confident we are to generalize it.

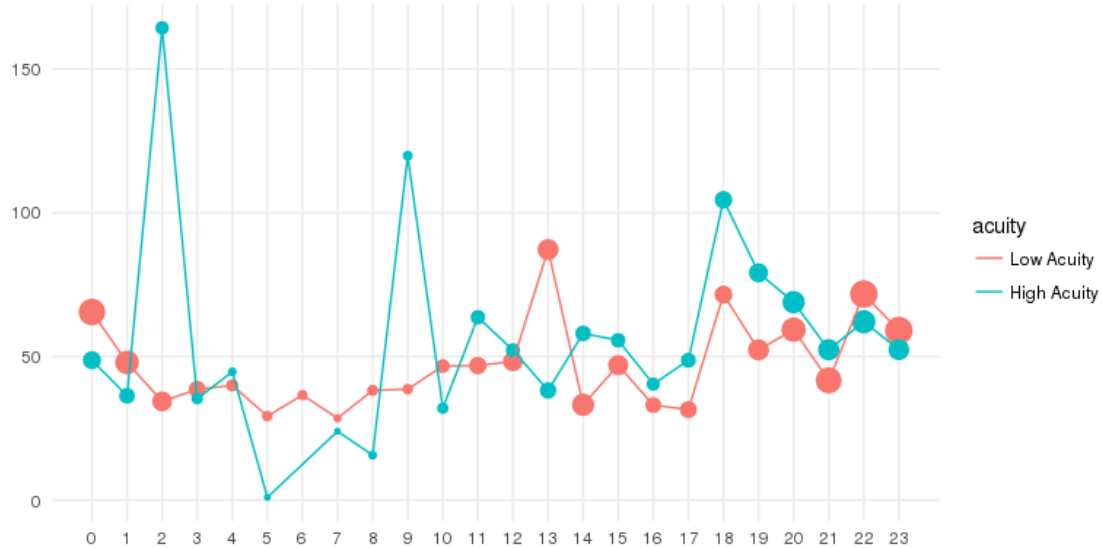
HOW MUCH TIME ON AVERAGE LWBSS SPENT BETWEEN ARRIVAL AND TRIAGE COMPLETE?



How much time on average they spent between triage complete and roomed



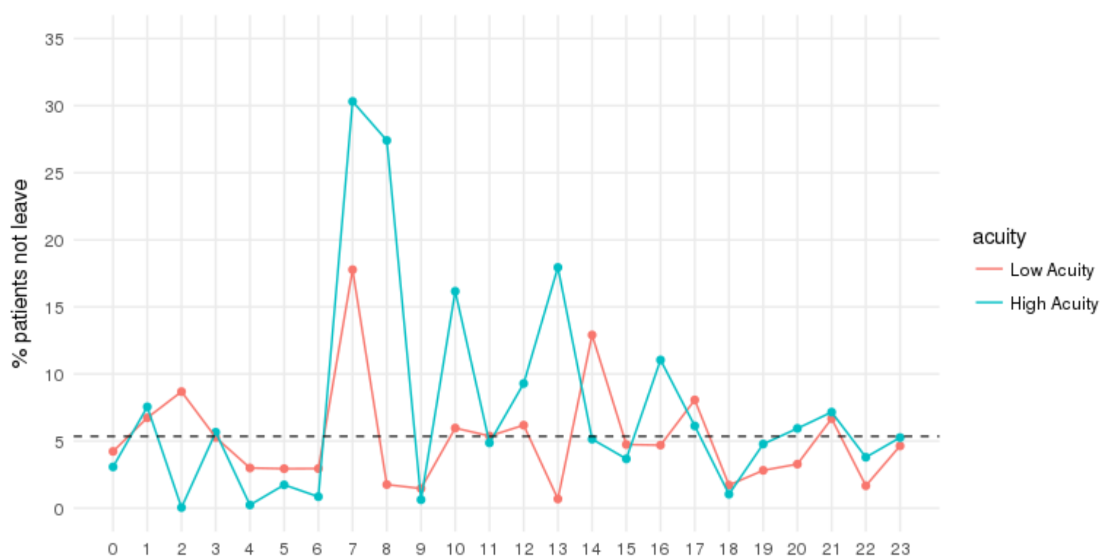
How much time on average they spent between roomed and left



The three plots shown above revealed that the majority of time a patient labeled with LWBS spent on is waiting to be roomed. Since only 0.2% LWBS patients left before roomed, this was not a huge problem.

How about compare the time spent waiting for a physician of regular patients and patients labeled as LWBS?

Percentage patient (did not leave) wait longer than LWBS



We assume the time a patient labeled with LWBS left to be the time the physician was assigned. Then we are able to compare the length of two different time gaps.

- The time gap between roomed and assigned physician for a patient not labeled with LWBS. For simplicity, we called this type of patients NL patients

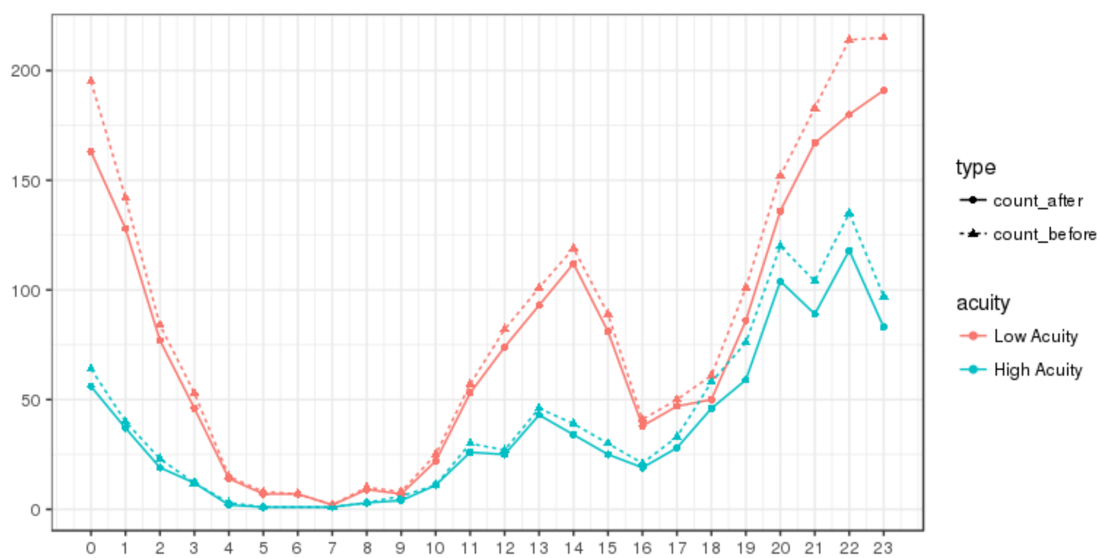
• The time gap between roomed and left for a patient labeled with LWBS. We called this type of patients L patients.

The black dotted line represents the average percentage that NL patients waited for a physician longer than L patients waited to leave. Only 6% NL patients did not see an actual physician until the average time gap for a L patient to leave after roomed. This implies that, if we could manage to reduce the time gap between roomed and got to see a physician to a certain amount, we can potentially decrease the number of LWBS patients.

Threshold

120 minutes

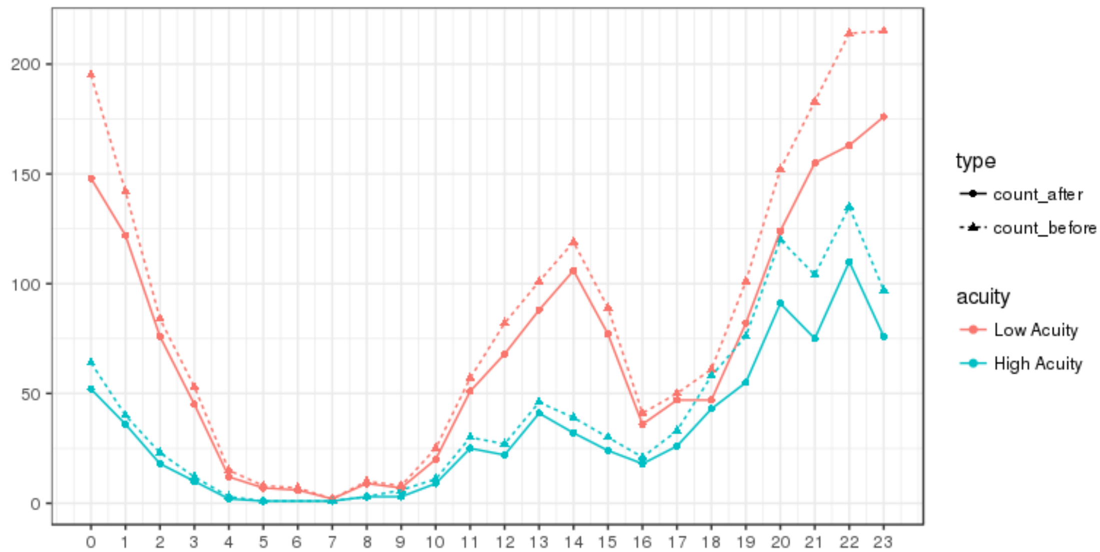
If we made a policy that every patient should have at least a physician assigned within 120 minutes after roomed:



We can expect a 11.99% reduction in number of LWBS patients.

90 minutes

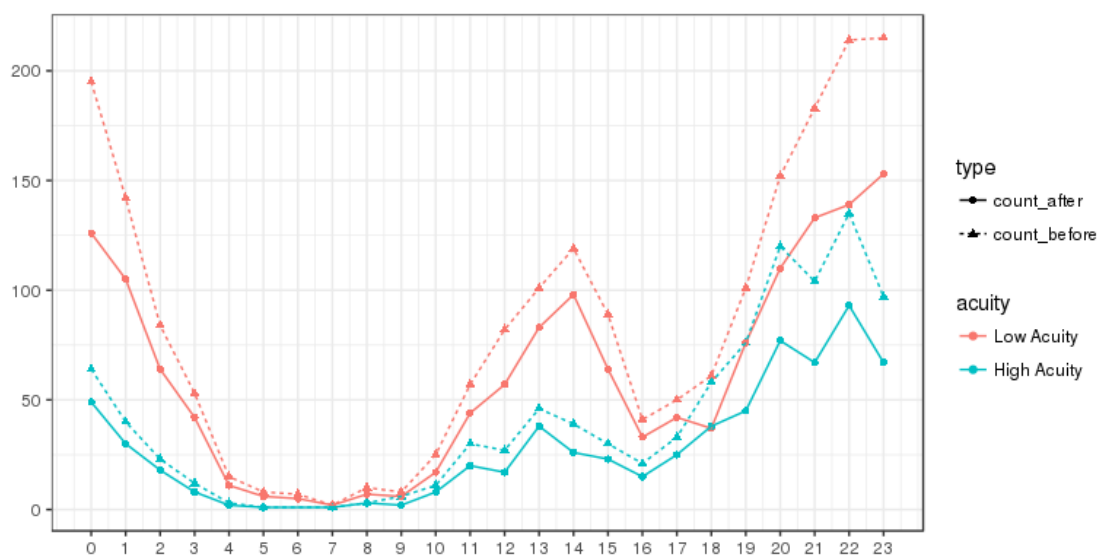
If we made a policy that every patient should have at least a physician assigned within 90 minutes after roomed:



We can expect a 18.27% reduction in number of LWBS patients.

60 minutes

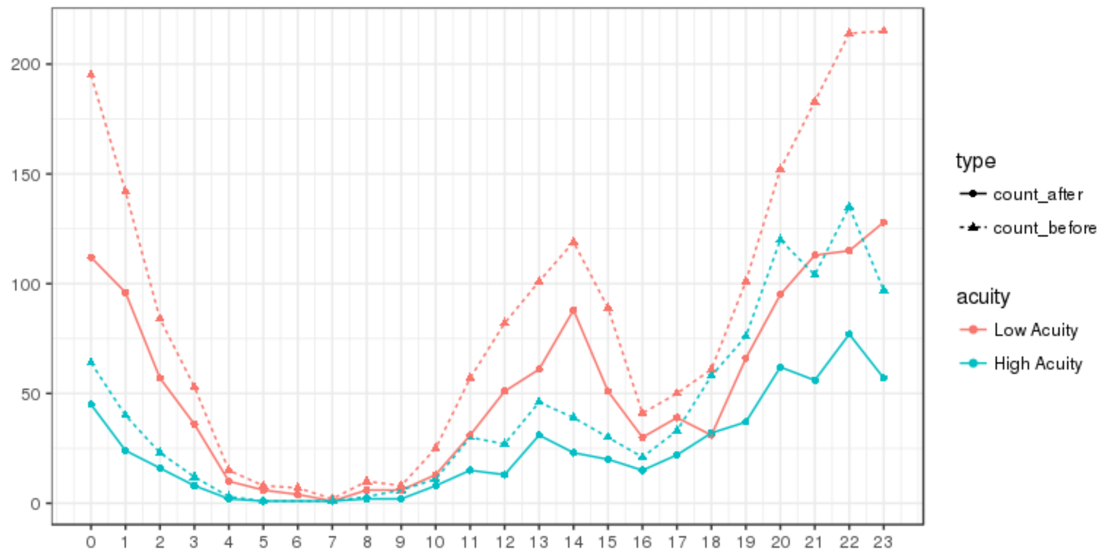
If we made a policy that every patient should have at least a physician assigned within 60 minutes after roomed:



We can expect a 28.76% reduction in number of LWBS patients.

Average minutes waiting of the non-LWBS pateints

If we made a policy that every patient should have at least a physician assigned within the average wait time for a NP patient to be seen by a physician after roomed, which is around 39.09 minutes:



We can expect a 39.38% reduction in number of LWBS patients.

-End-