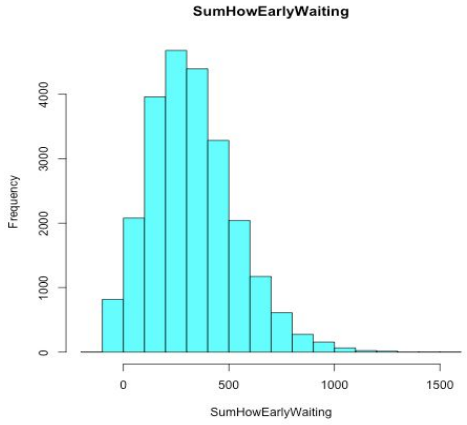
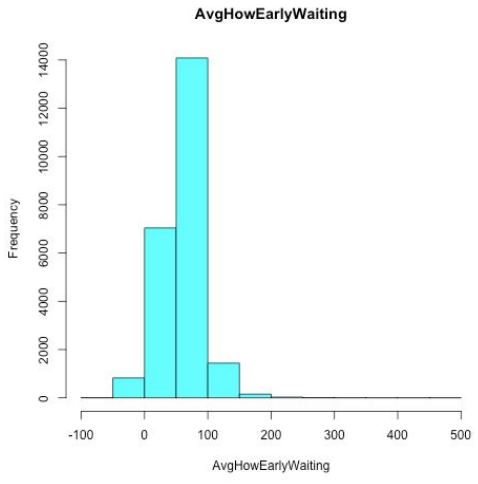
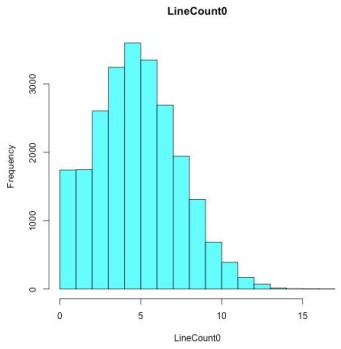
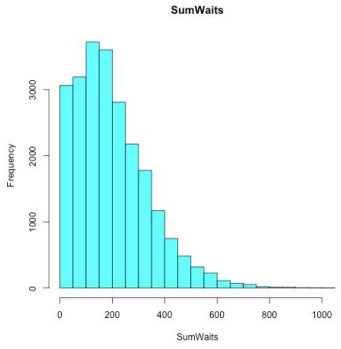
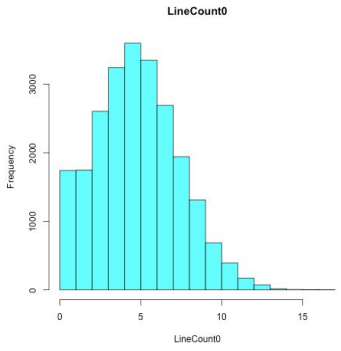
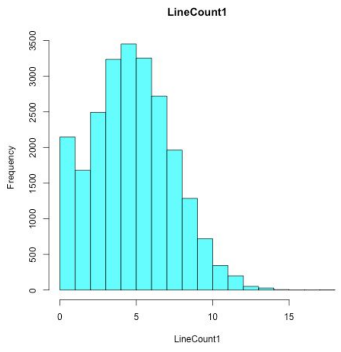
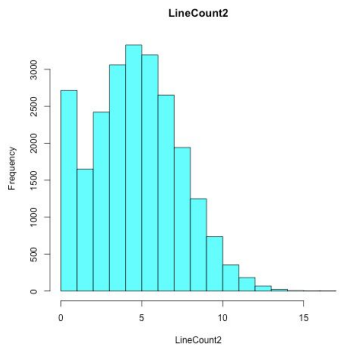
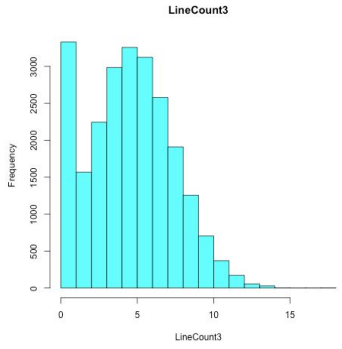
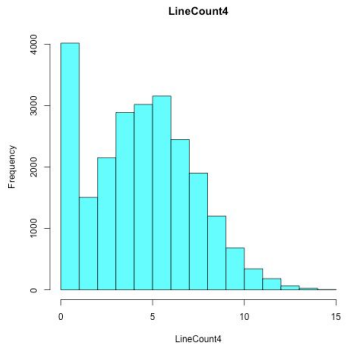
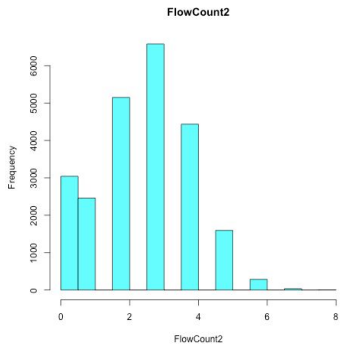
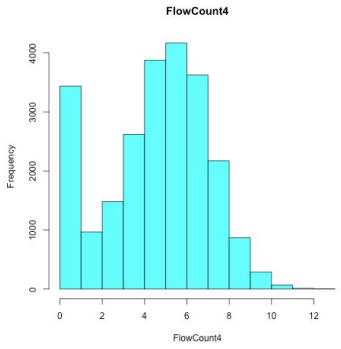
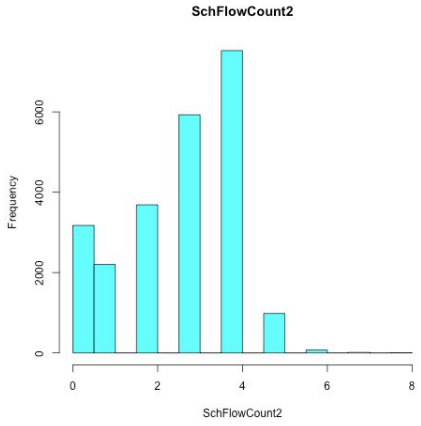
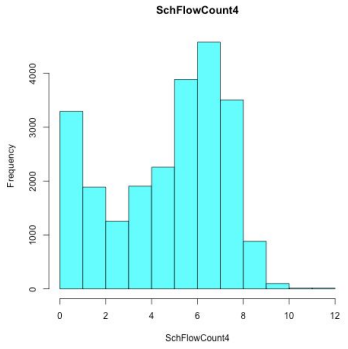
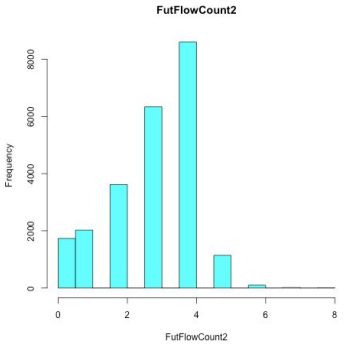
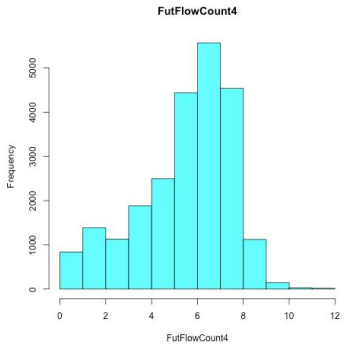
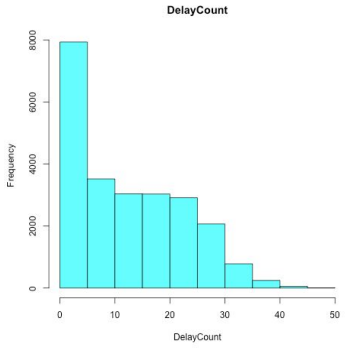
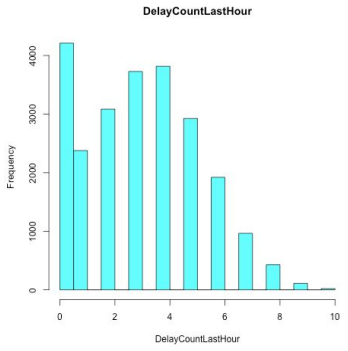
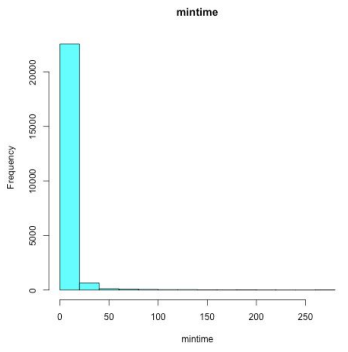
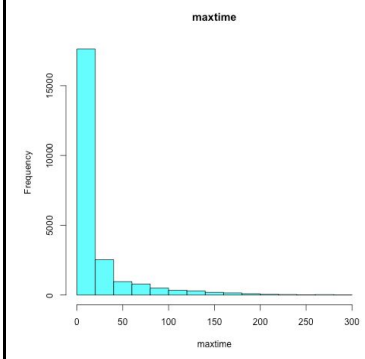
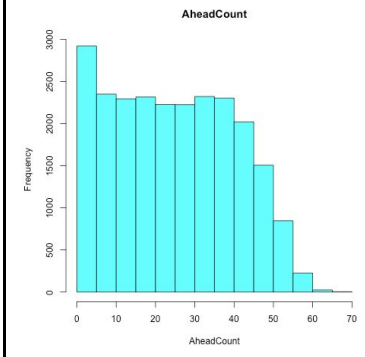
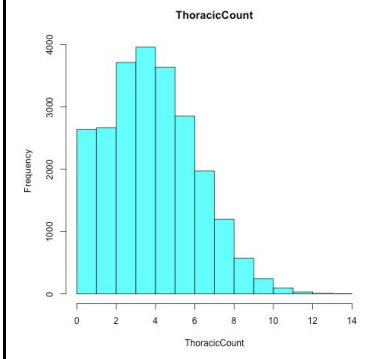
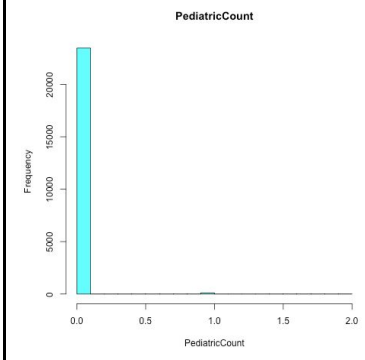
Attribute Name	Histogram	Comments	Should We Use It? Yes/No
x_ArrivalDTTM Arrival time of patient.			
x_ScheduledDTTM Scheduled time of patient's exam.			
x_BeginDTTM Begin time for the patient's exam.			
SumHowEarlyWaiting Sum of how early the patients in line are for their appointment.		Slightly Right-skewed-thi s graph has a nice distribution. Shows he sum of	Yes,worth keeping - I would like to get rid of this variable. I don't think its necessary to have the sum and the average of how early waiting.-which pts is most significant?
AvgHowEarlyWaiting Average of how early the patients in line are for their appointment.		Left-skewed- not much variation in this data.	worth keeping for calculations

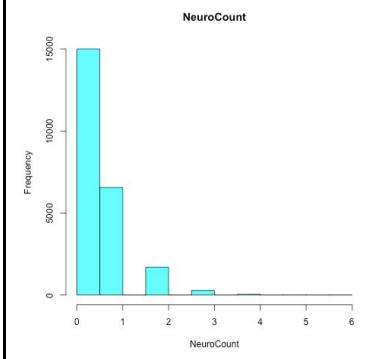
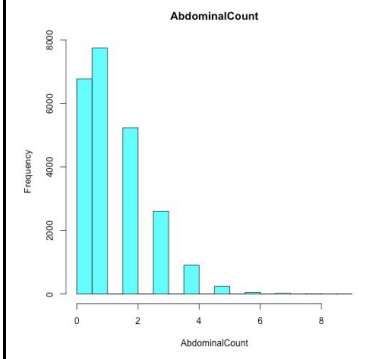
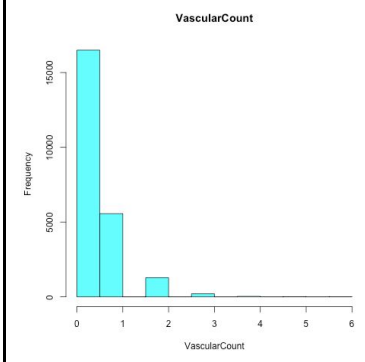
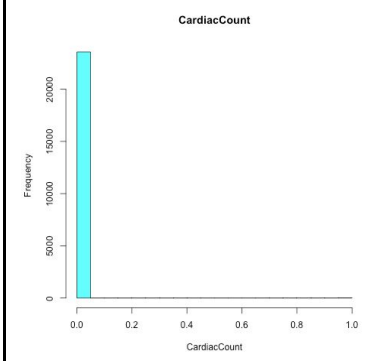
LineCount0Strict Number of patients in line with scheduled times after current time.		Right-skewed-	
SumWaits Sum of the wait times for patients in line.		Right-skewed -	
LineCount0 Number of patients in line measured when a patient arrives.		Right-Skewed Most commonly there is between 4-5 patients ahead at arrival	
LineCount1 Number of patients in line measured 15 minutes before a patient arrives.		Right -Skewed	

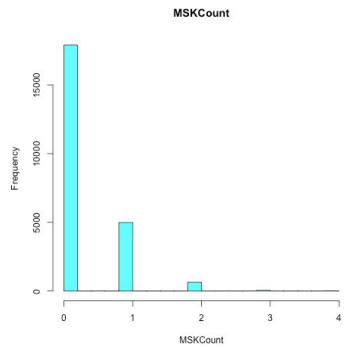
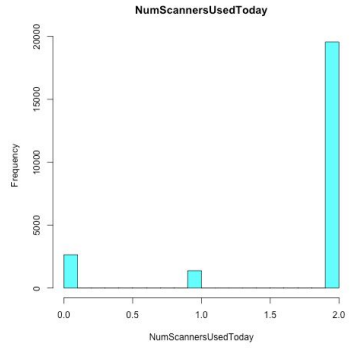
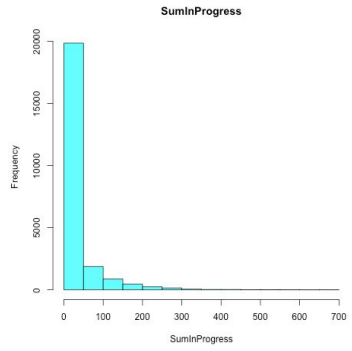
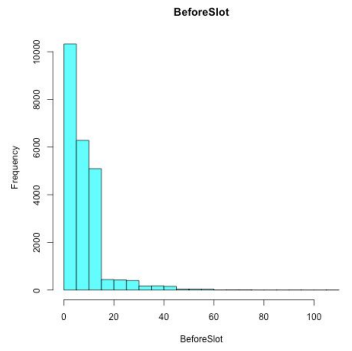
LineCount2 Number of patients in line measured 30 minutes before a patient arrives.		Right-Skewed	
LineCount3 Number of patients in line measured 45 minutes before a patient arrives.		Right-Skewed	
LineCount4 Number of patients in line measured 60 minutes before a patient arrives.		Right-Skewed	
FlowCount2 Number of patients starting exams in the 30-minute window before patient arrived.		Closest to normal curve	

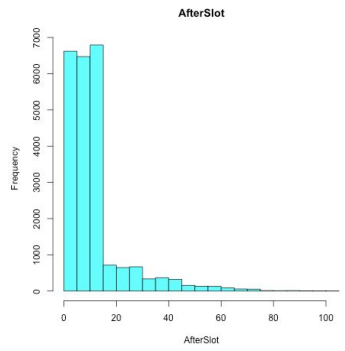
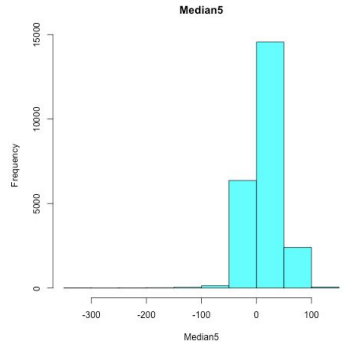
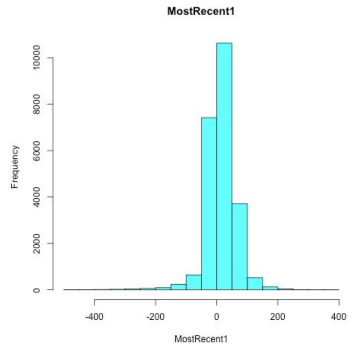
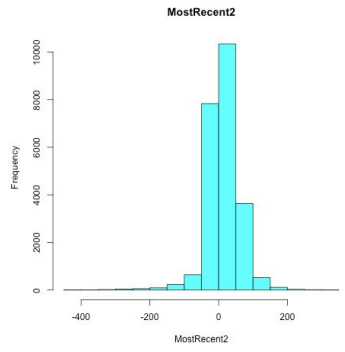
FlowCount4 Number of patients starting exams in the 60-minute window before patient arrived.			
SchFlowCount2 Number of patients scheduled in the 30-minute window before patient arrived.			
SchFlowCount4 Number of patients scheduled in the 60-minute window before patient arrived.			
FutFlowCount2 Number of patients scheduled in the 30-minute window after patient arrived.			

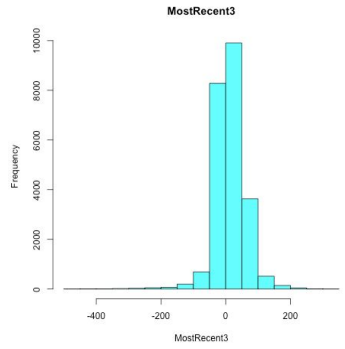
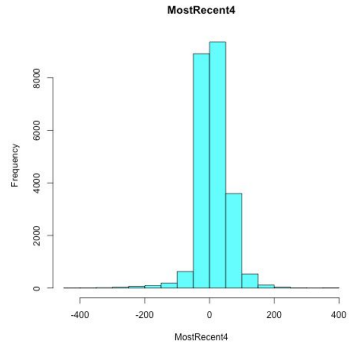
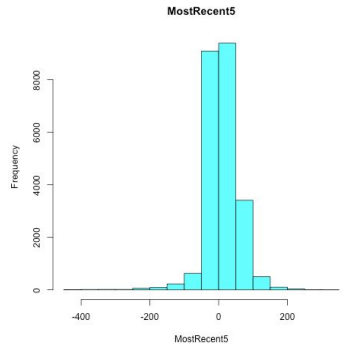
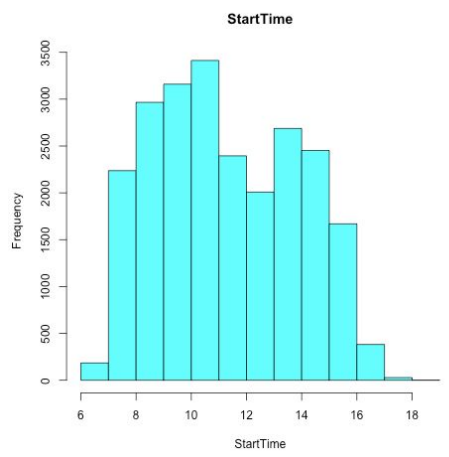
FutFlowCount4 Number of patients scheduled in the 60-minute window after patient arrived.			
DelayCount Number of delayed exams up to current time of day.			
DelayCountLastHour Number of delayed exams in last hour.			
Mintime Minimum wait time for the day.		This graph does not offer any significant findings.	No

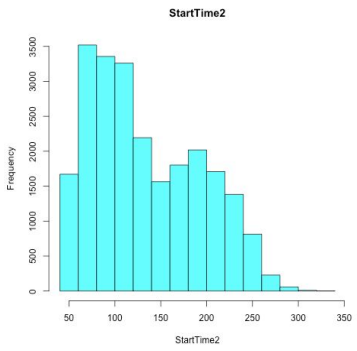
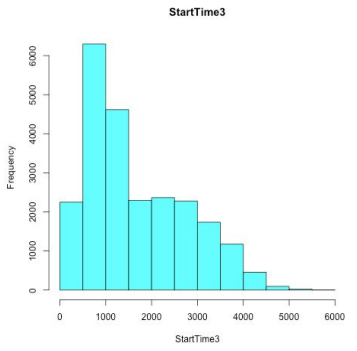
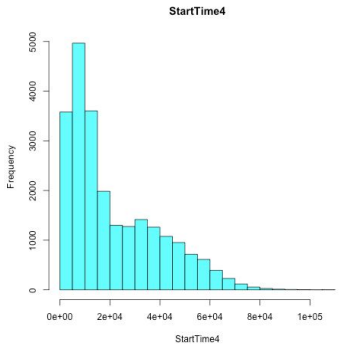
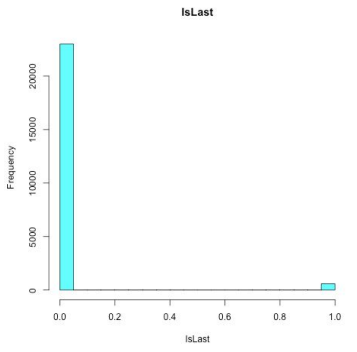
Maxtime Maximum wait time for the day.			Np
AheadCount Number of patients scheduled before current patient for the day.			NO
ThoracicCount Number of patients waiting for thoracic exam.		By far the most common special exam type, this variable is affecting the size of the line more often than not	Yes
PediatricCount Number of patients waiting for pediatric exam.		Pediatric patients are super rare, enough that it probably isn't worth including them as a factor	No, unless as part of a calculated field

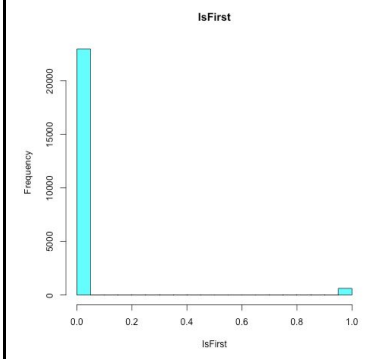
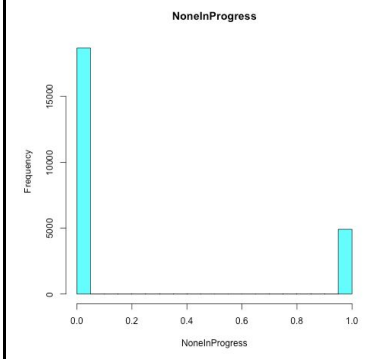
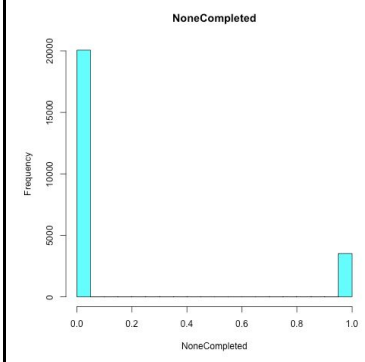
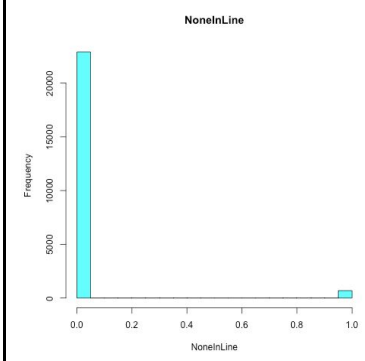
NeuroCount Number of patients waiting for neuro exam.		Usually zero, but can range up to 6	Probably only as part of a calculated field
AbdominalCount Number of patients waiting for abdominal exam.		More often nonzero than zero, can range up to 9	Yes; I think it could be worth adding a calculated field to include all of the special exam types together
VascularCount Number of patients waiting for vascular exam.		Usually zero, can range up to 6	Calculated field
CardiacCount Number of patients waiting for cardiac exam.		Either zero or 1, and very rarely 1	Calculated field

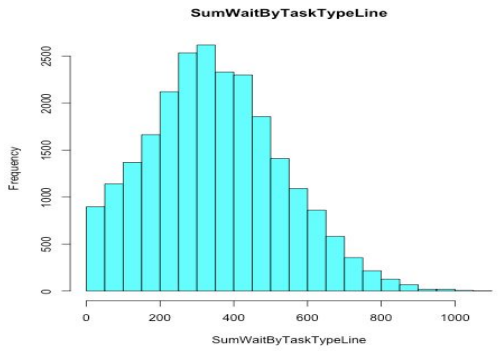
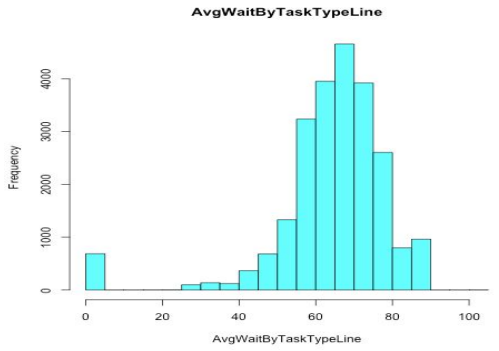
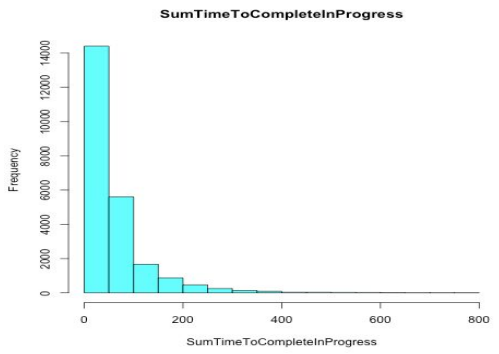
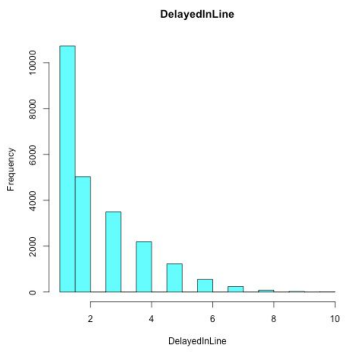
MSKCount Number of patients waiting for musculoskeletal exam.		Usually zero, can range up to 4	Calculated field
NumScannersUsedToday Number of scanners in facility that have been used on that day.		Only two total scanners available, and they are usually both used. Doesn't give us much information about anything in particular	No
SumInProgress Sum of length of time exams have been in progress.		Overwhelmingly 0-50 min of in-progress exams, but range is up to 671. I would be curious to see if this correlates closely with number of special exams for the day.	No, but could help us decide whether the number of special exams is worth using or not
BeforeSlot Time since last appointment slot.		Usually 0-5 minutes, ranges up to 106. I don't think this is super useful because it doesn't actually tell us much about how many patients have been coming in	No

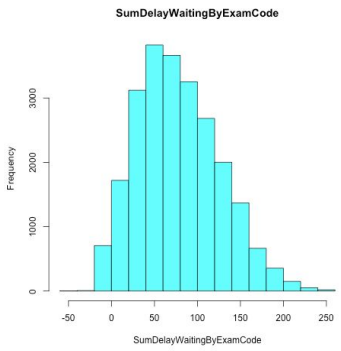
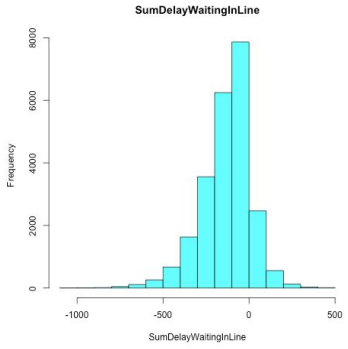
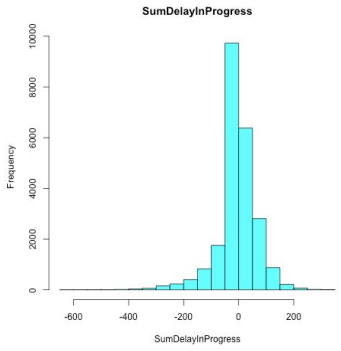
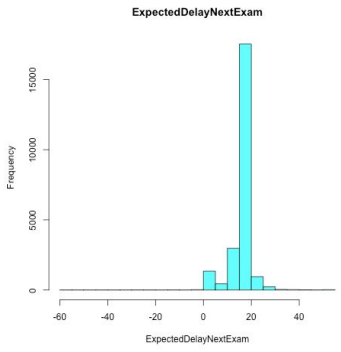
AfterSlot Time until next appointment slot.		Slightly more spread between 0-15 minutes than BeforeSlot, but otherwise a similar distribution	Could be useful to have another calculated field here- average of BeforeSlot and AfterSlot, to capture how much “cushion” a patient has between other scheduled people coming in
Median5 Median delay/wait time for 5 most recent customers.		Could be better represented with more bins, but I think this will be useful as a factor showing in general what kind of wait times are being experienced	Yes
MostRecent1 Delay/wait time for most recent patient.		All of these MostRecent columns have the same distribution as Median5 and even more so the same as each other	No
MostRecent2 Delay/wait time for 2nd most recent patient.		“”	No

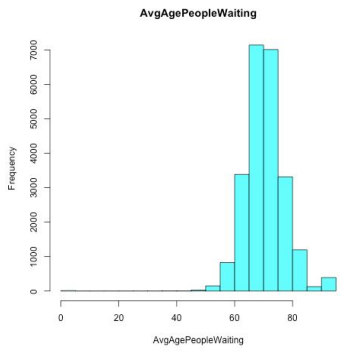
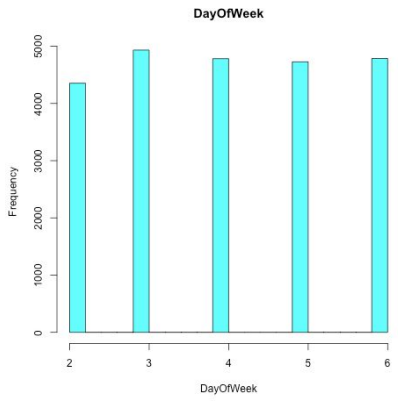
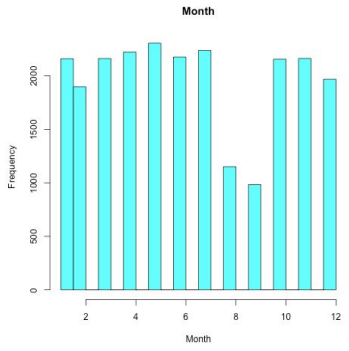
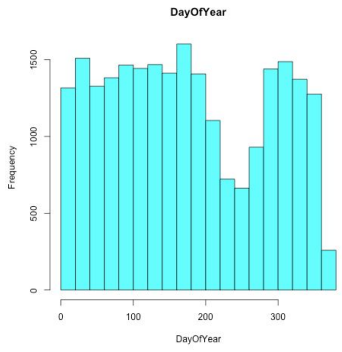
MostRecent3 Delay/wait time for 3rd most recent patient.		“”	No
MostRecent4 Delay/wait time for 4th most recent patient.		“”	No
MostRecent5 Delay/wait time for 5th most recent patient.		“”	No
StartTime Hour of arrival.		Makes sense that patients are pretty spread throughout the day, but this will help us determine which times of day are busiest	Yes

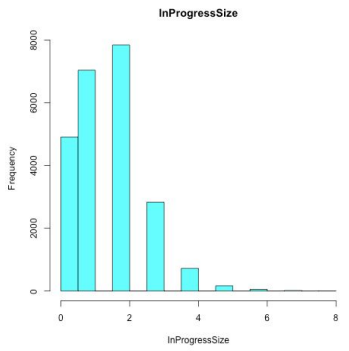
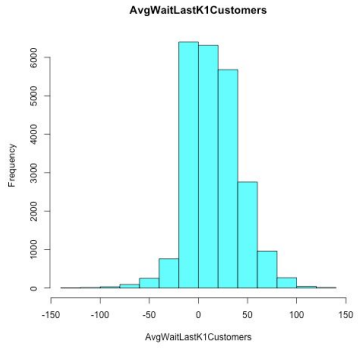
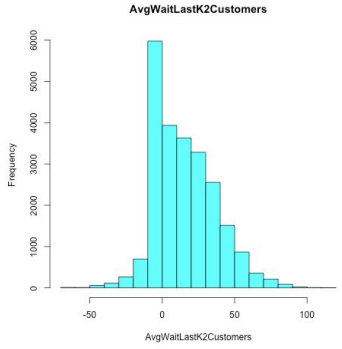
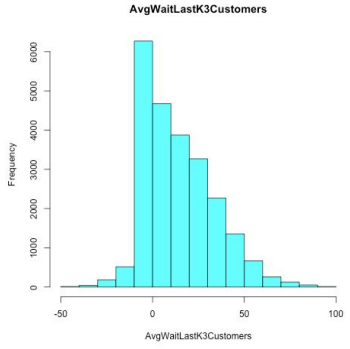
StartTime2 2nd power of hour of arrival to account for nonlinear trends.		Probably not useful unless we use a model that explicitly requires it	No
StartTime3 3rd power of hour of arrival to account for nonlinear trends.		“”	No
StartTime4 4th power of hour of arrival to account for nonlinear trends.		“”	No
IsLast Last scheduled patient for the day.		Boolean value- I don’t think we need to remove last patients, because they will still have a wait time affected by the people ahead of them- just need to keep in mind that we cannot use AfterSlot for these patients	No

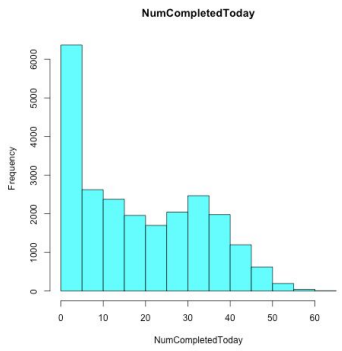
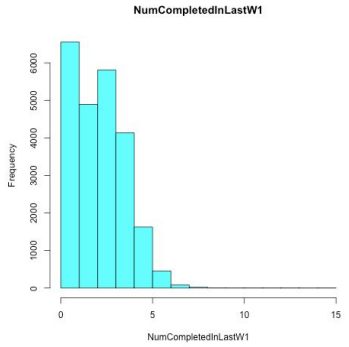
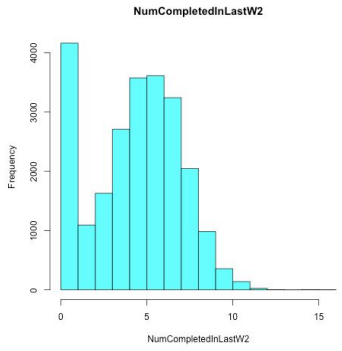
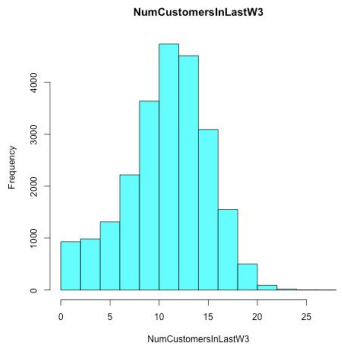
IsFirst First scheduled patient for the day.		Boolean value- I think these are outliers that we need to remove. The first patients of the day will almost always have unusually low wait times	No
NoneInProgress No exams in progress.		Boolean value- interesting that there are so many more instances of this than of the patient being the first of the day. I think this should be a useful indicator	Yes
NoneCompleted No exams completed that day.		This histogram summarizes discrete data (patient complete the exam = 1 and the patient does not complete the exam =0)	Yes, from this histogram we can see that the number of patients who have not completed the exam that day approximately 20000, which may request patients to come again on other days leading to an increase in the scheduled patients on these days, thus the number of patients in line of those days will increase.
NoneInLine No patients in line.		This histogram summarizes discrete data (patient non-inline=1 and patient inline =0)	Yes, from the histogram we can see that the number of patients non-inline and also the patients who are inline, which is extremely high, and this will affect the wait time

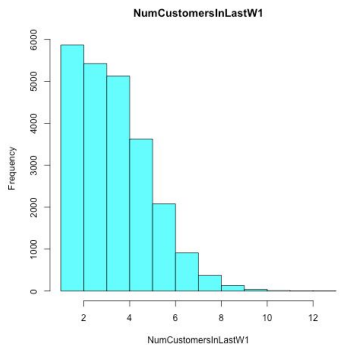
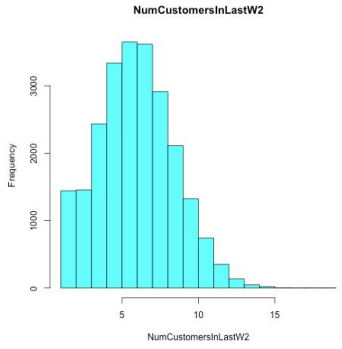
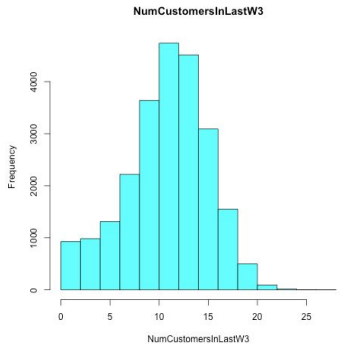
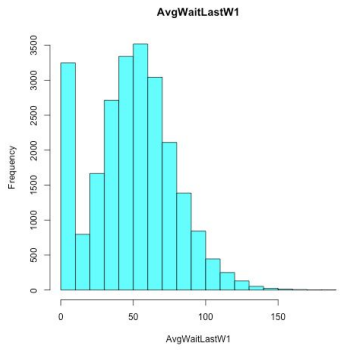
SumWaitByTaskTypeLine Sum of waits of patients in line by exam type.		This histogram summarizes continuous data, and it almost has a bell shape which tells us the majority of patients wait for between 200 and 400.	Yes
AvgWaitByTaskTypeLine Average waits of patients in line by exam type.		This histogram summarizes continuous data, and it has a random distribution.	No, we can get the average from the previous column (SumWaitByTaskTypeLine)
SumTimeToCompleteInProgress The sum of the expected times to complete of the exams in progress		This histogram summarizes continuous data, and it has a right-skewed distribution.	No
DelayedInLine The number of patients in line who are delayed.		This histogram summarizes continuous data, and it has a right-skewed distribution.	Yes, this increases the wait time for an individual.

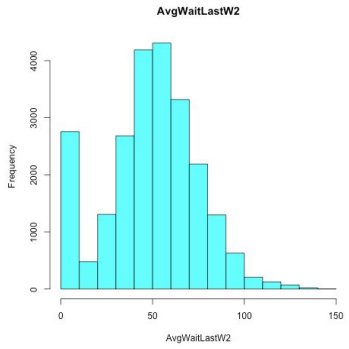
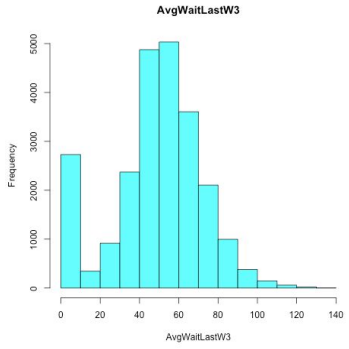
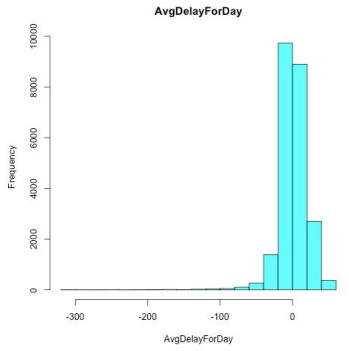
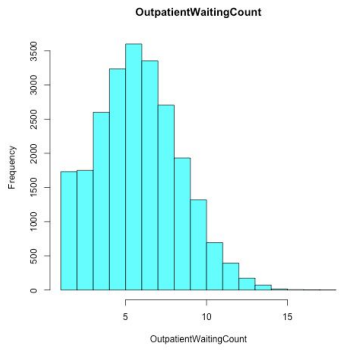
SumDelayWaitingByExamCode Sum of delays of patients in line by exam type.		This histogram summarizes continuous data, and it almost has a bell shape which tells us the majority of delay between 50 and 100	No,it seems like we have similar data showing on the column (SumWaitByTaskTypeLine)
SumDelayWaitingInLine Sum of delays/waits of patients in line.		This histogram summarizes continuous data, and it almost has a bell shape	Yes
SumDelayInProgress Sum of delays/waits of exams in progress.		This histogram summarizes continuous data, and it almost has a bell shape	No, it seems related to the exam in progress not to the patient waiting time in the waiting room.
ExpectedDelayNextExam Expected delay of the next scheduled exam.		The histogram has a random distribution.	Yes, estimates expected delay for all patients may be significant under the assumption that different examinations require different wait times

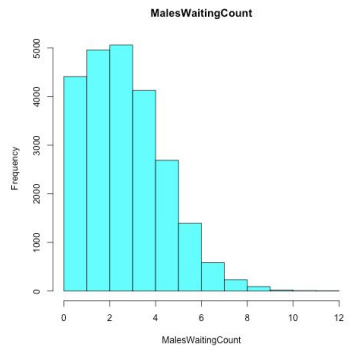
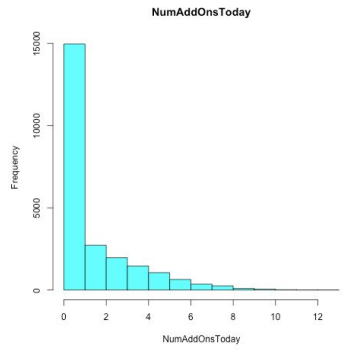
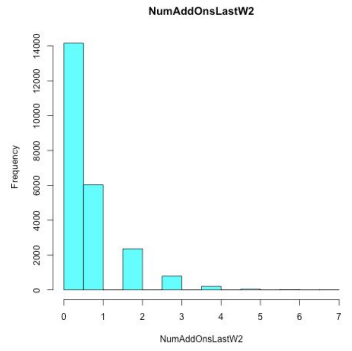
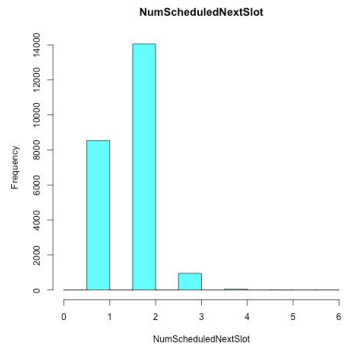
AvgAgePeopleWaiting Average age of the patients in line.		This histogram summarizes continuous data, and it almost has a bell shape	Yes,it is important to recognize the waiting time depending on the patient's age.Elderly patients may take longer with certain tasks (walking to the examination room, changing)
DayOfWeek The day of the week the exam is scheduled.		This histogram looks like uniform distribution , it shows that the number of patients evenly spread out among all values of the day.	Yes, the day of the week can reflect temporal trends in patient processing
Month The month of the year the exam is scheduled.		This histogram is almost uniform distribution ,	No
DayOfYear The day of the year the exam is scheduled.		This histogram is almost uniform distribution	Yes, the day of the year can reflect temporal trends in patient processing

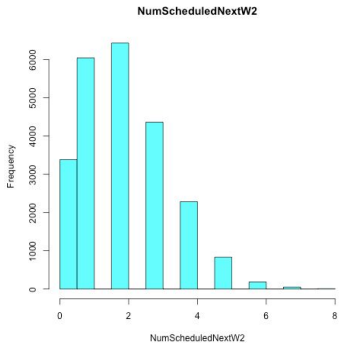
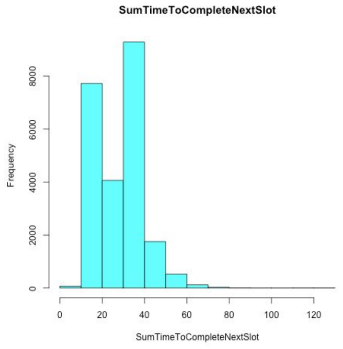
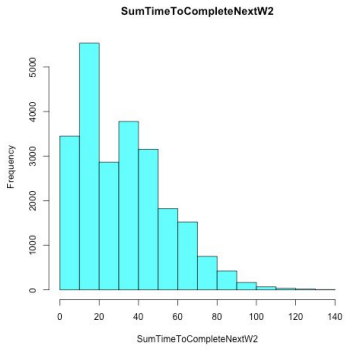
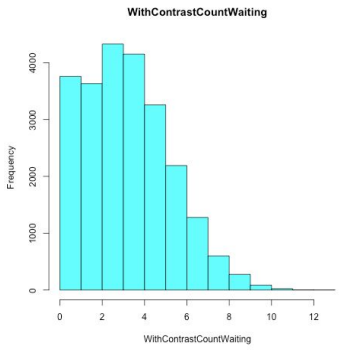
InProgressSize Number of exams in progress for facility.		This histogram is almost a right-skewed distribution	No
AvgWaitLastK1Customers Average wait for the last 2 customers.		This histogram summarizes continuous data, and it almost has a bell shape	No
AvgWaitLastK2Customers Average wait for the last 4 customers.		This histogram summarizes continuous data, and it almost has a right-skewed distribution	Yes
AvgWaitLastK3Customers Average wait for the last 8 customers.		This histogram summarizes continuous data, and it almost has a right-skewed distribution	No

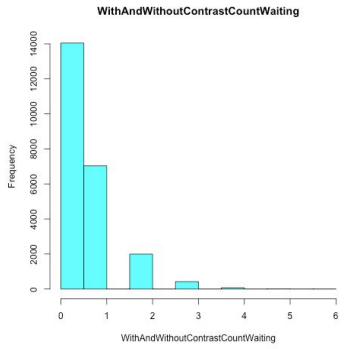
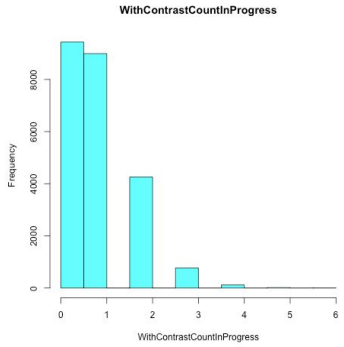
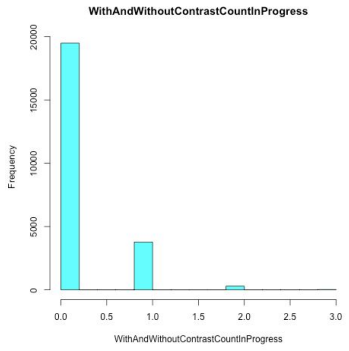
NumCompletedToday Number of exams completed up to current of day.		Continuous, left skewed	no
NumCompletedInLastW1 Number of exams completed in last 30 minutes.		Continuous, left skewed	yes
NumCompletedInLastW2 Number of exams completed in last 60 minutes.		Continuous, slightly left skewed	yes
NumCompletedInLastW3 Number of exams completed in last 120 minutes.		Continuous, slightly right skewed	yes

NumCustomersInLastW1 Number of customers who have arrived in the last 30 minutes.		Continuous, left skewed	Yes, very
NumCustomersInLastW2 Number of customers who have arrived in the last 60 minutes.		Less relevant than above	yes
NumCustomersInLastW3 Number of customers who have arrived in the last 120 minutes.		Less relevant than above	yes
AvgWaitLastW1 Average delay/wait time last 30 minutes.		Potentially very important	yes

AvgWaitLastW2 Average delay/wait time last 60 minutes.		Less relevant than above	yes
AvgWaitLastW3 Average delay/wait time last 120 minutes.		Less relevant than above	yes
AvgDelayForDay Average delay/wait for patients for that day.		Low correlation to wait time	no
OutpatientWaitingCount Number of outpatients waiting in line.		Probably very important	yes

MalesWaitingCount Number of male patients waiting in line.			no
NumAddOnsToday Number of people who have been added to the schedule for today.		Could be important part of model	Yes
NumAddOnsLastW2 Number of people who have been added to the schedule in last 60 minutes.		“”	yes
NumScheduledNextSlot Number of patients scheduled in next slot.		Probably has little to do with wait time	no

NumScheduledNextW2 Number of people scheduled in next 60 minutes.		“”	no
SumTimeToCompleteNextSlot Expected time to completion of exams in next slot.		What are meant by ‘next’ and ‘slot’	
SumTimeToCompleteNextW2 Expected time to completion of exams scheduled in next hour.		“”	
WithContrastCountWaiting Number of patients waiting for an exam with contrast.		What does this mean?	

WithAndWithoutContrastCountWaiting Number of patients waiting for an exam with and without contrast.		Doesn't make sense	
WithContrastCountInProgress Number of exams in progress with contrast.		See 2 above	
WithAndWithoutContrastCountInProgress Number of exams in progress with and without contrast.		""	

We Definitely Want To Remove:
The three timestamp columns
SumHowEarlyWaiting
LineCount1/2/3/4
DelayCount
Mintime
Seven special exam types
NumScannersUsedToday
SumInProgress
*Replace BeforeSlot and AfterSlot with averaged values
MostRecent1/2/3/4/5
StartTime2/3/4
IsFirst
IsLast

NoneInProgress
NoneCompleted
NoneInLine
SumWaitByTaskType
SumDelayWaitingByExamCode
SumDelayInProgress
Month
InProgressSize
AvgWaitLastK1Customers
AvgWaitLastK3Customers
NumCompletedToday
NumCompletedLastW1/2/3
NumCustomersLastW2/3
AvgWaitLast2/3
AvgDelayForDay
MalesWaitingCount
NumAddOnsLastW2
NumScheduledNextSlot
NumScheduledNextW2
SumTimeToCompleteNextW2