

STA490 Exploratory Data Analysis

Vivian Ngo

10/11/2018

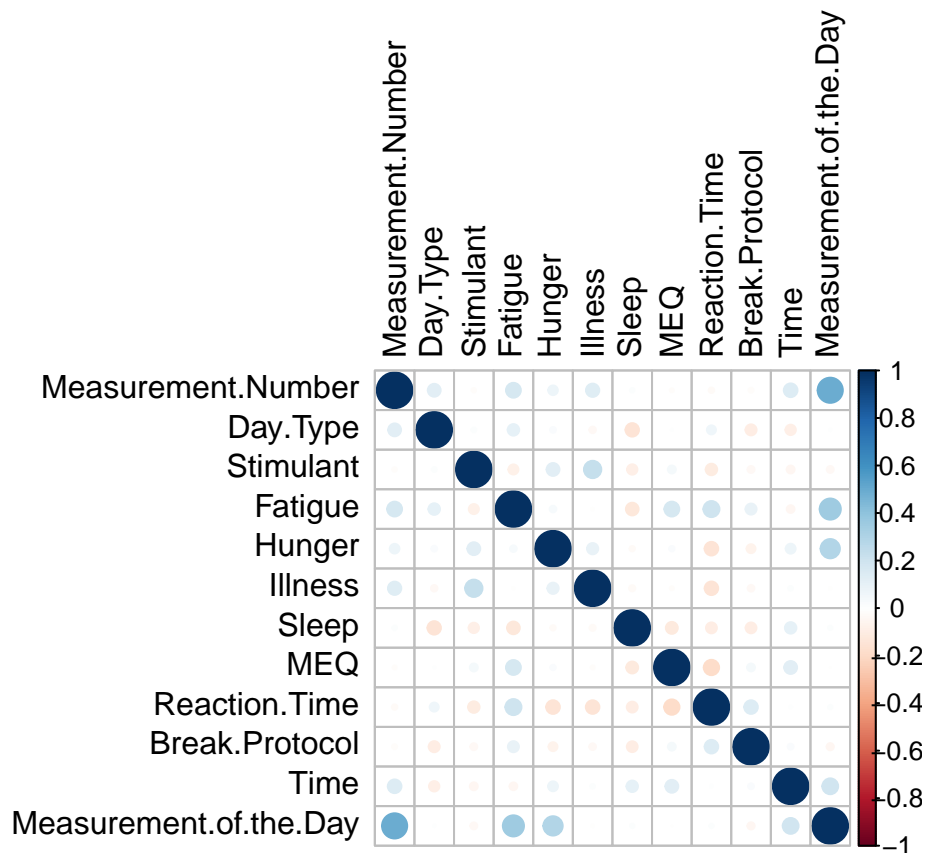
Variable Definitions

To begin, let's define what the variables are.

Variable	Description
ID	The unique ID number for every student
Measurement Number	Number of the measurement. Ranges from 1 to 8
Raw Time	The 24 hour format time when the data was recorded
Day Type	1 indicates busy day, 0 indicates light day
Stimulant	1 indicates a stimulant was taken, 0 indicates otherwise
Fatigue	Fatigue measured on the Samn-Perelli 7-pt scale. Ranges from 1 to 7
Hunger	Hunger measured on the hunger scale, value decreasing with hunger level
Illness	1 indicates illness, 0 indicates otherwise
Sleep	The hours of sleep that was had the night before the observation
MEQ	The Morningness Eveningness Questionnaire score, ranging from 16 to 86
Reaction Time	The reaction time that was measured (in seconds)
Break Protocol	1 indicates protocol was broken, 0 indicates otherwise
MEQ Category	Classification of MEQ score into Evening, Morning, or Neither.
Measurement of the Day	If the observation was first, second, third, or fourth observation of the day

Correlation Between Variables

Let's see if there is any correlation between the variables.



According to the correlation plot above, Measurement Number and Measurement of the Day are closely correlated. This is expected because Measurement of the Day is obtained from the Measurement Number. There are also other correlated variables, such as Fatigue and Measurement of the Day, Hunger and Measurement of the Day, and Illness and Stimulant. Reaction Time appears to have almost no correlation to Time and Measurement of the Day. However, it appears to be slightly correlated to Fatigue, Hunger, Illness, and MEQ.

A Look at Reaction Time

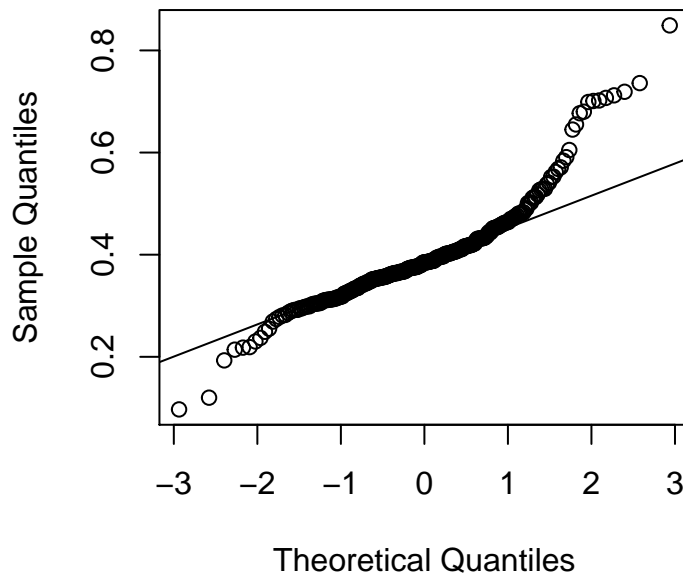
Summary Statistics

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0970	0.3470	0.3850	0.3976	0.4320	0.8490	18

Is reaction time normal?

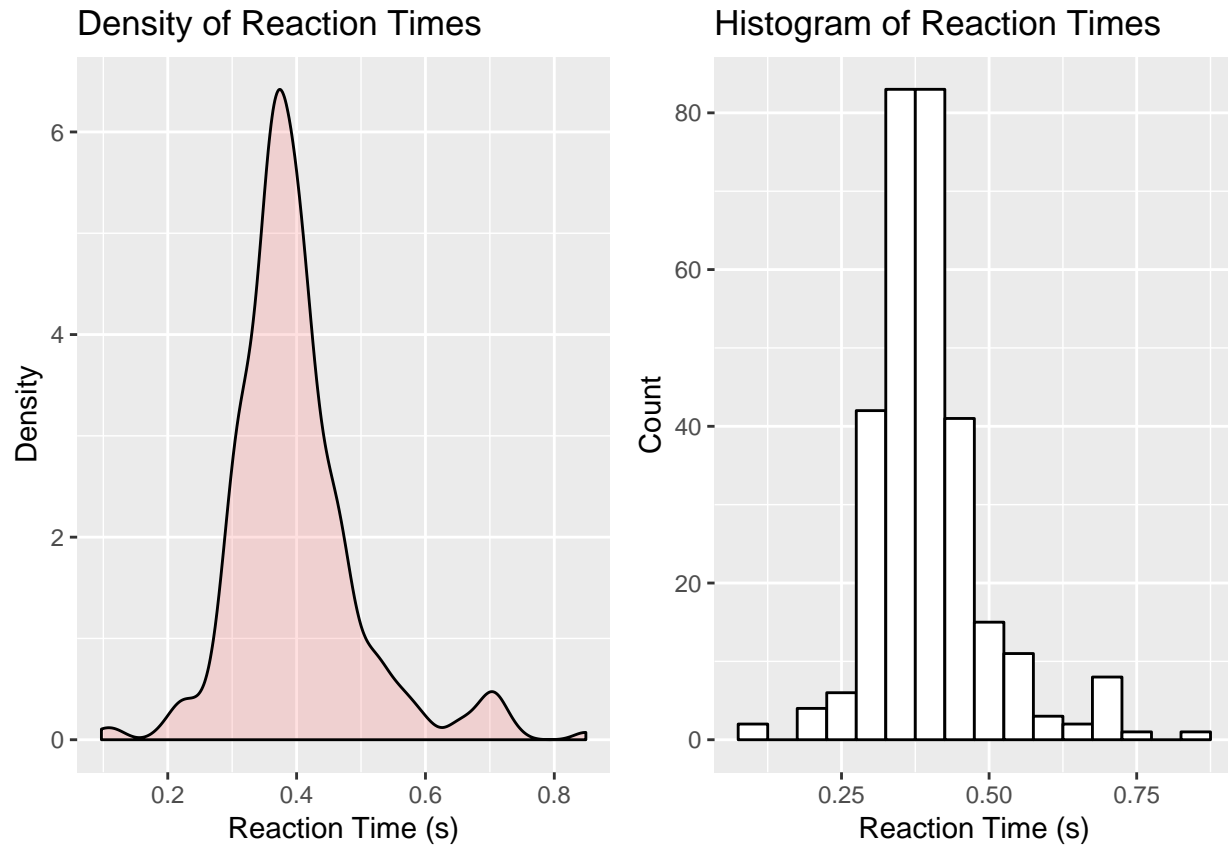
Let's check with a normal Q-Q plot:

Normal Q-Q Plot



```
##  
## Shapiro-Wilk normality test  
##  
## data: rxn$Reaction.Time  
## W = 0.90255, p-value = 4.808e-13
```

The Q-Q plot suggested non-normality and a Wilks-Shapiro test further supports that hypothesis. It appears that reaction time may not be normal.

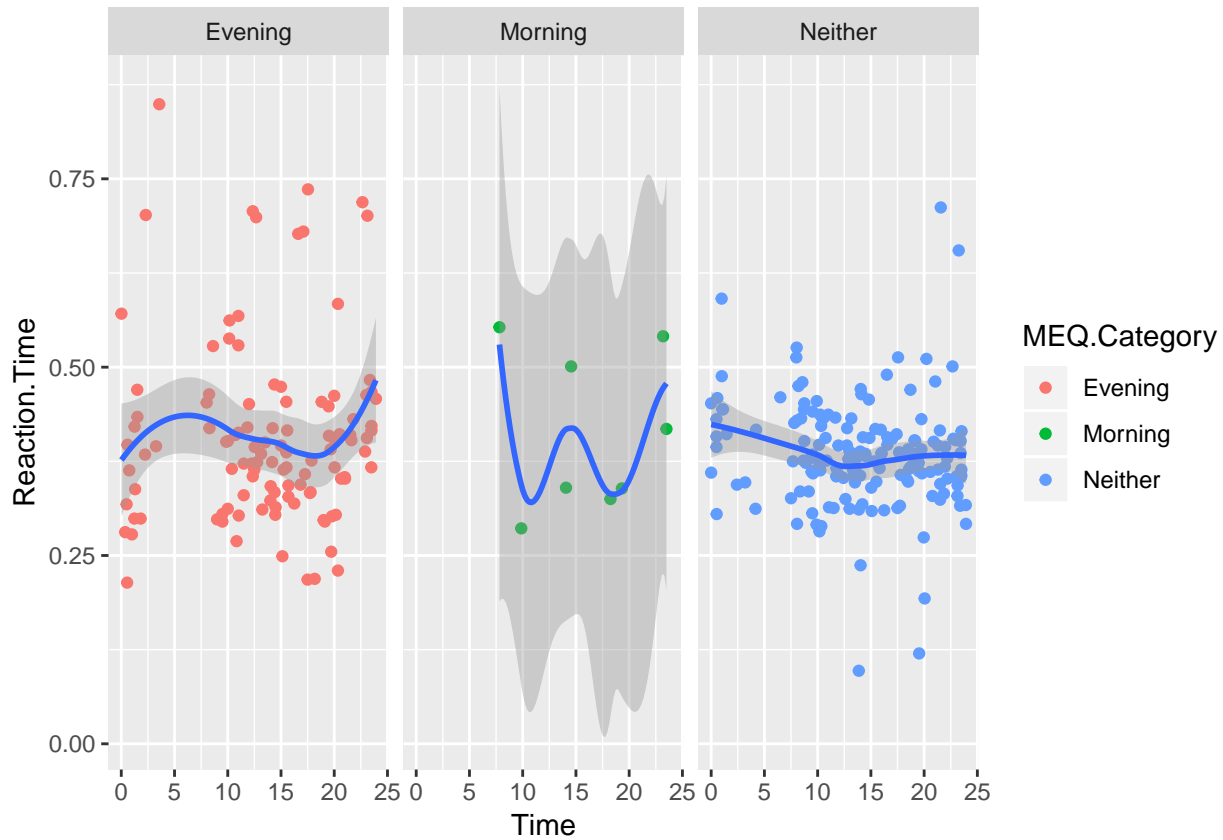


Reaction time seems to be concentrated around 0.37 seconds.

Correlation between Reaction Time and other Variables

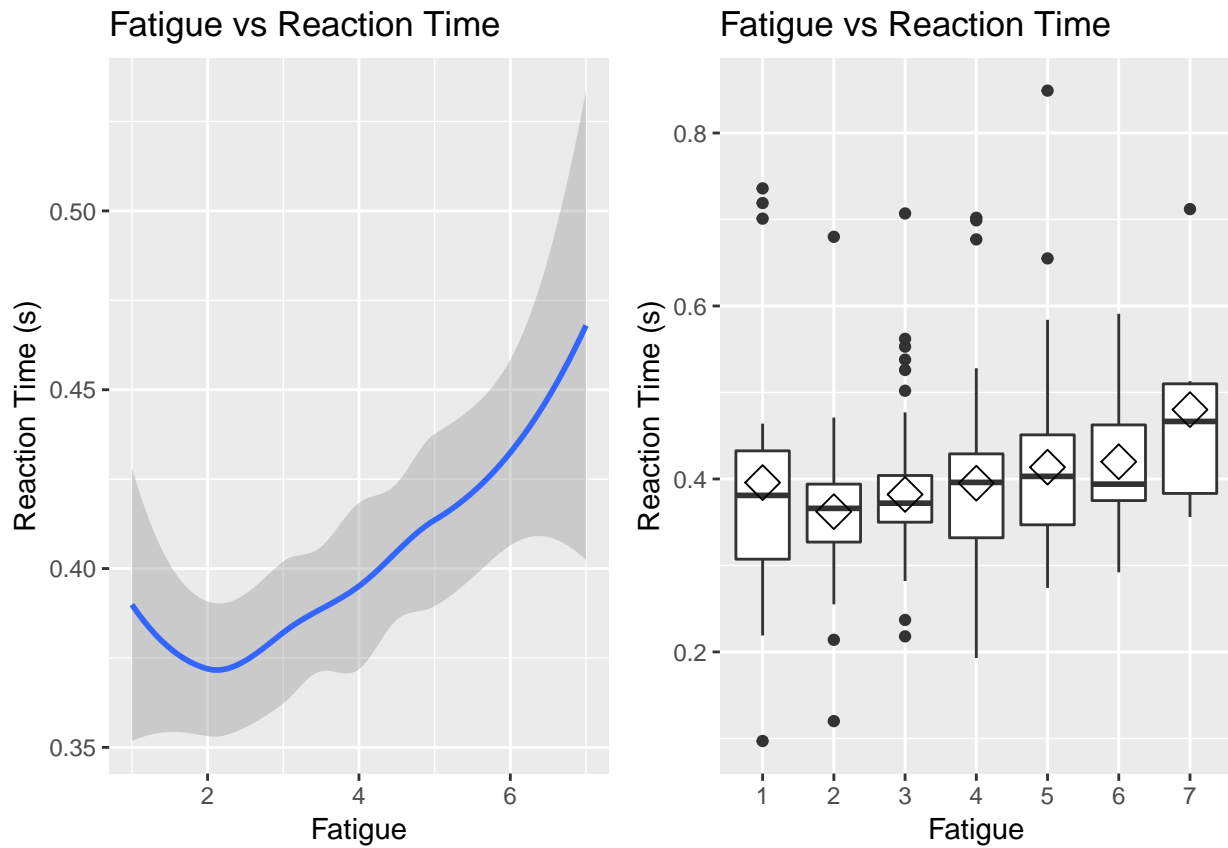
Variable	Correlation Coefficient (with Reaction Time)
MEQ	-0.184
Hunger	-0.146
Illness	-0.144
Stimulant	-0.103
Sleep	-0.092
Measurement.Number	-0.023
Time	0.006
Measurement.of.the.Day	0.015
Day.Type	0.063
Break.Protocol	0.144
Fatigue	0.202
Reaction.Time	1.000

Time of the Day vs Reaction Time



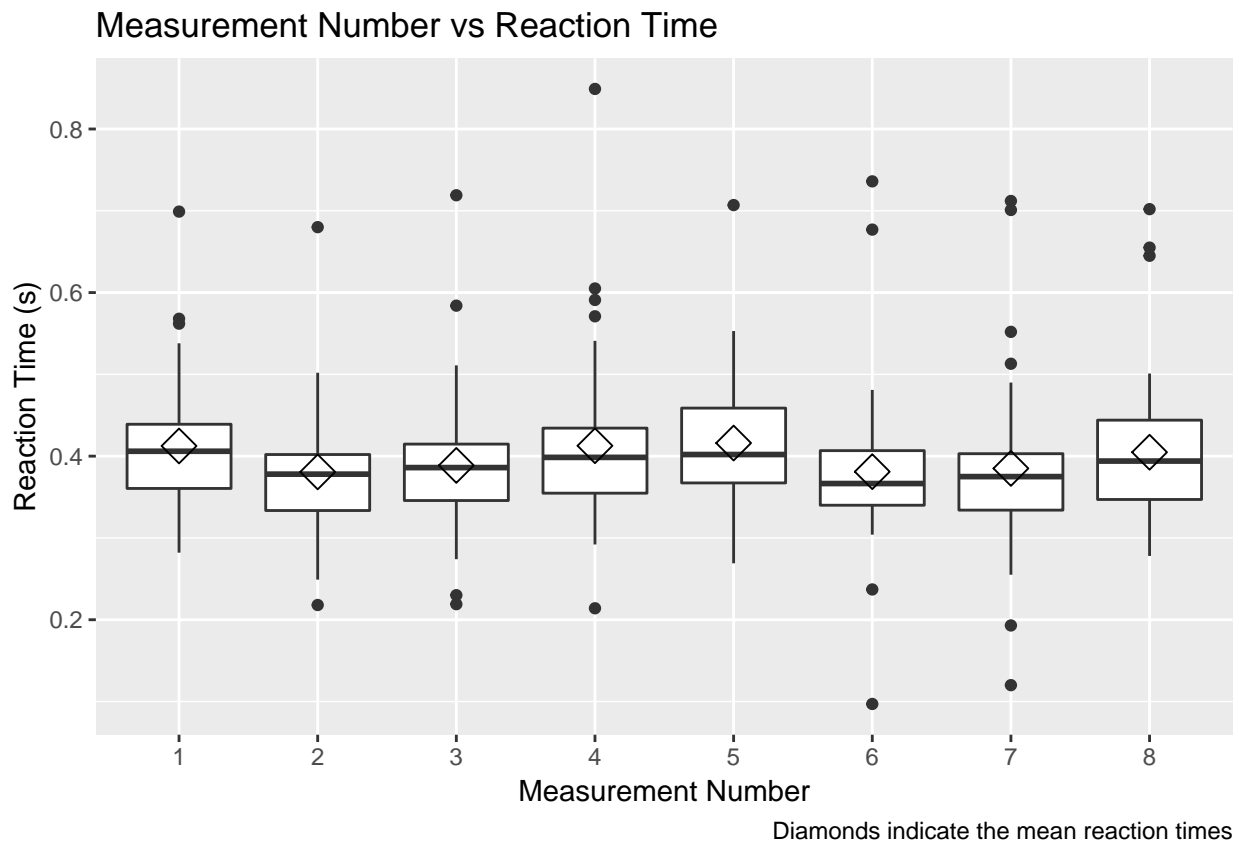
From the previous correlation plot, Time of Day and Reaction Time had close to no correlation. However, by looking at each MEQ Category separately, it appears that each category has its own trend. Evening persons seem to react slower near the end of the day, morning people seem to react slower at the beginning and end of the day, and those who are neither perform slightly better in the first ten hours of the day and then remain constant. However, there was only one morning person recorded so this data should be analyzed carefully.

Fatigue vs Reaction Time

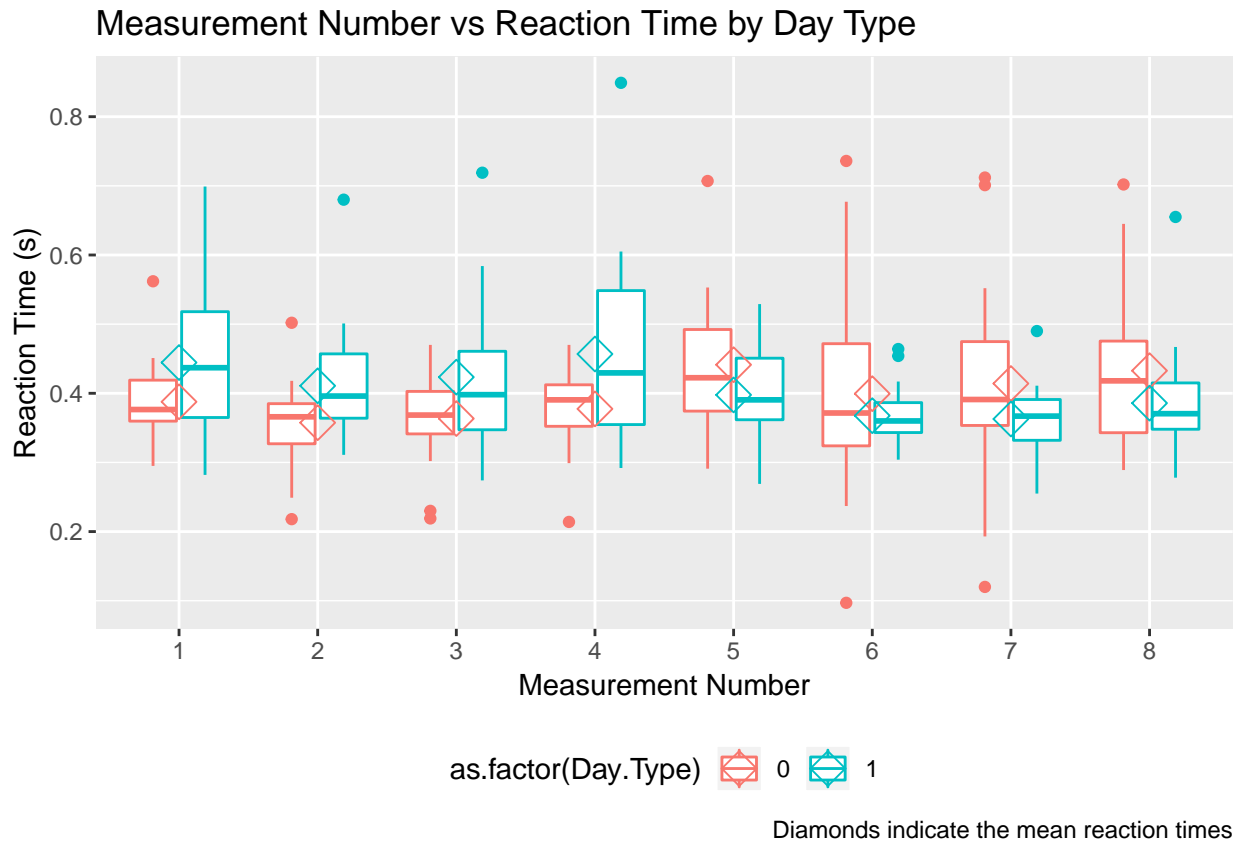


The diamonds in the boxplots indicate mean reaction time for each fatigue level. Although the previous plot showed a positive correlation between fatigue and reaction time, there is some overlap of the interquartile ranges in this boxplot. Further tests should be done to see if the mean reaction times of the fatigue levels are different.

Measurement Number vs Reaction Time

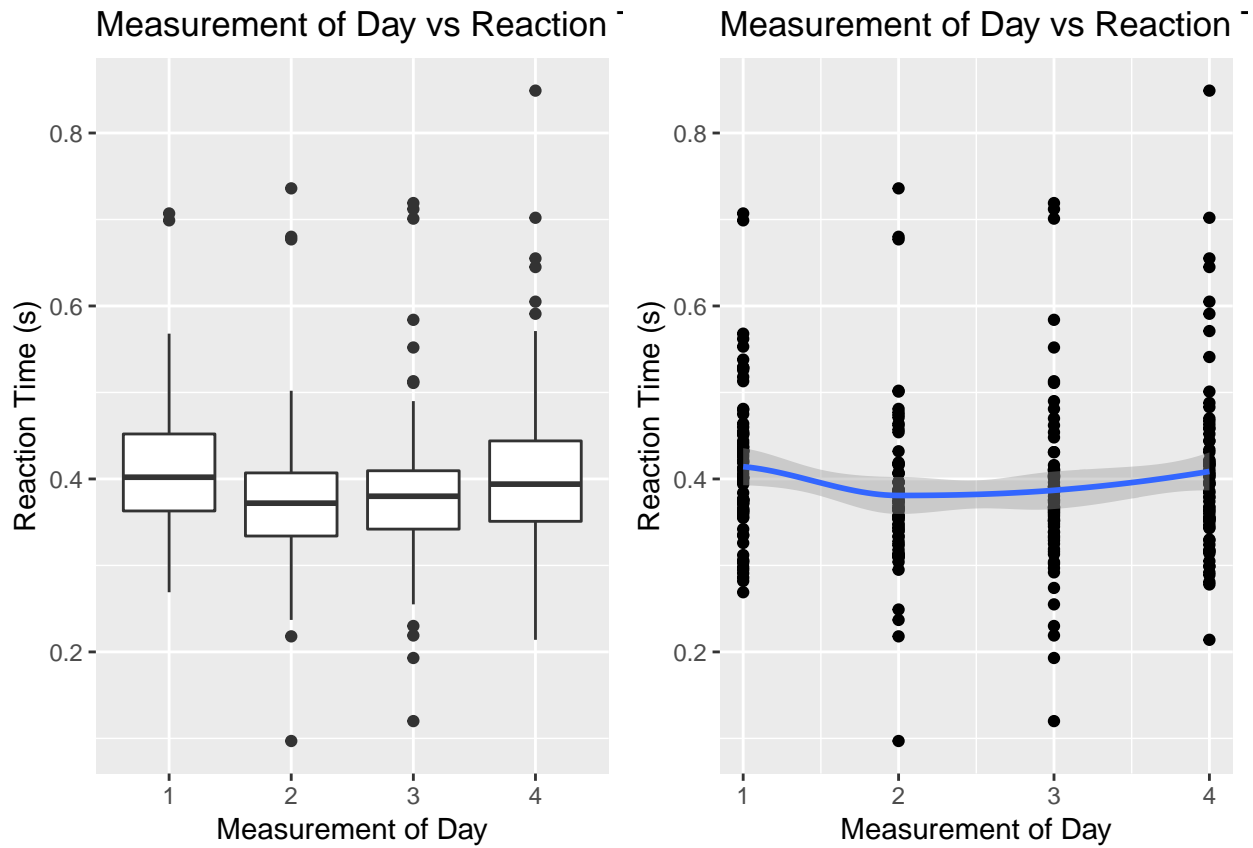


In the boxplot above, it appears that reaction time is slightly lower in the middle of the day, compared to the beginning and end of the day. We can further investigate this by looking at the Measurement of the Day.



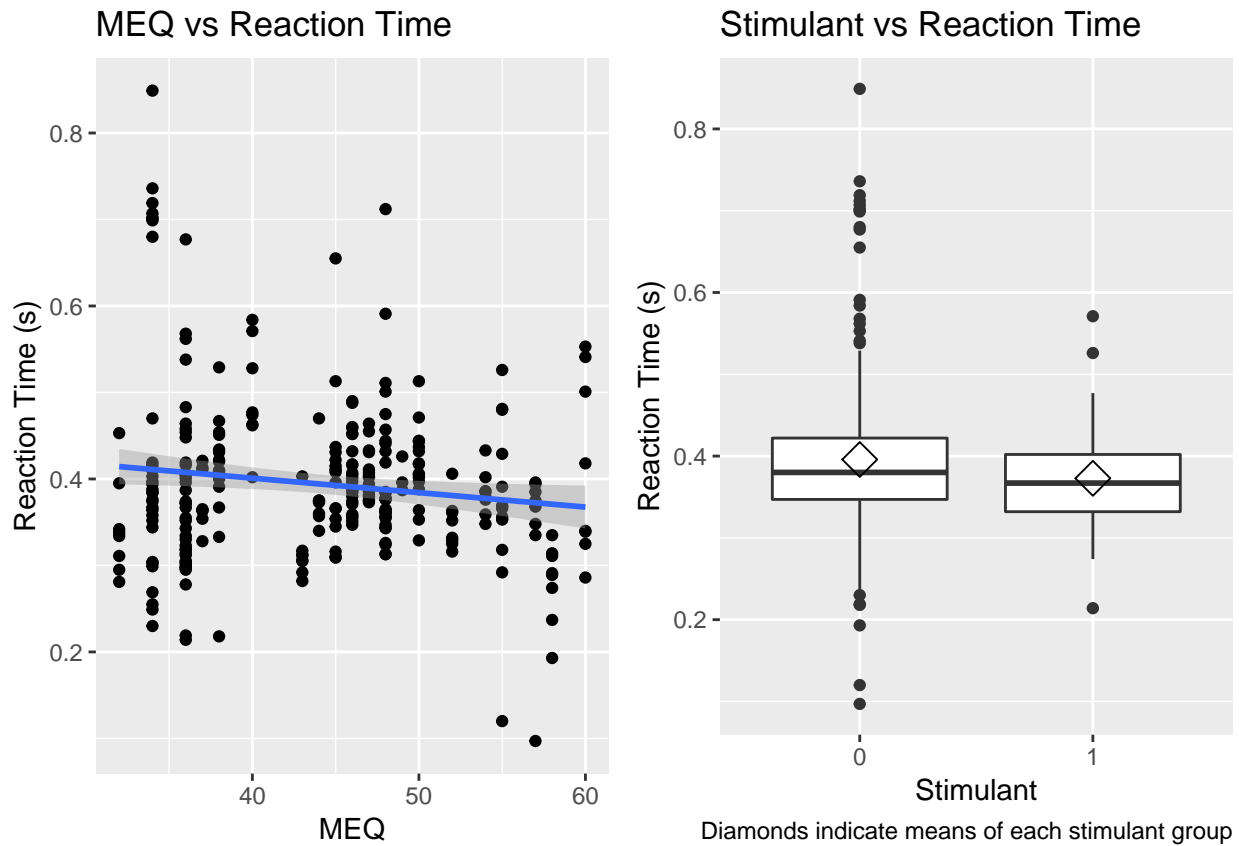
This is another boxplot of Measurement Number against Reaction Time, but with the Type of Day also included (0 = light, 1 = busy). This graph is interesting because the reaction time for one type of day seems to differ depending on whether a busy day was observed first or a light day was observed first.

Measurement of the Day vs Reaction Time



It appears that there is a nonlinear relationship between the measurement of the day and the reaction time. It is possible that reaction times are smaller in the middle of the day, and longer at the beginning and end of the day.

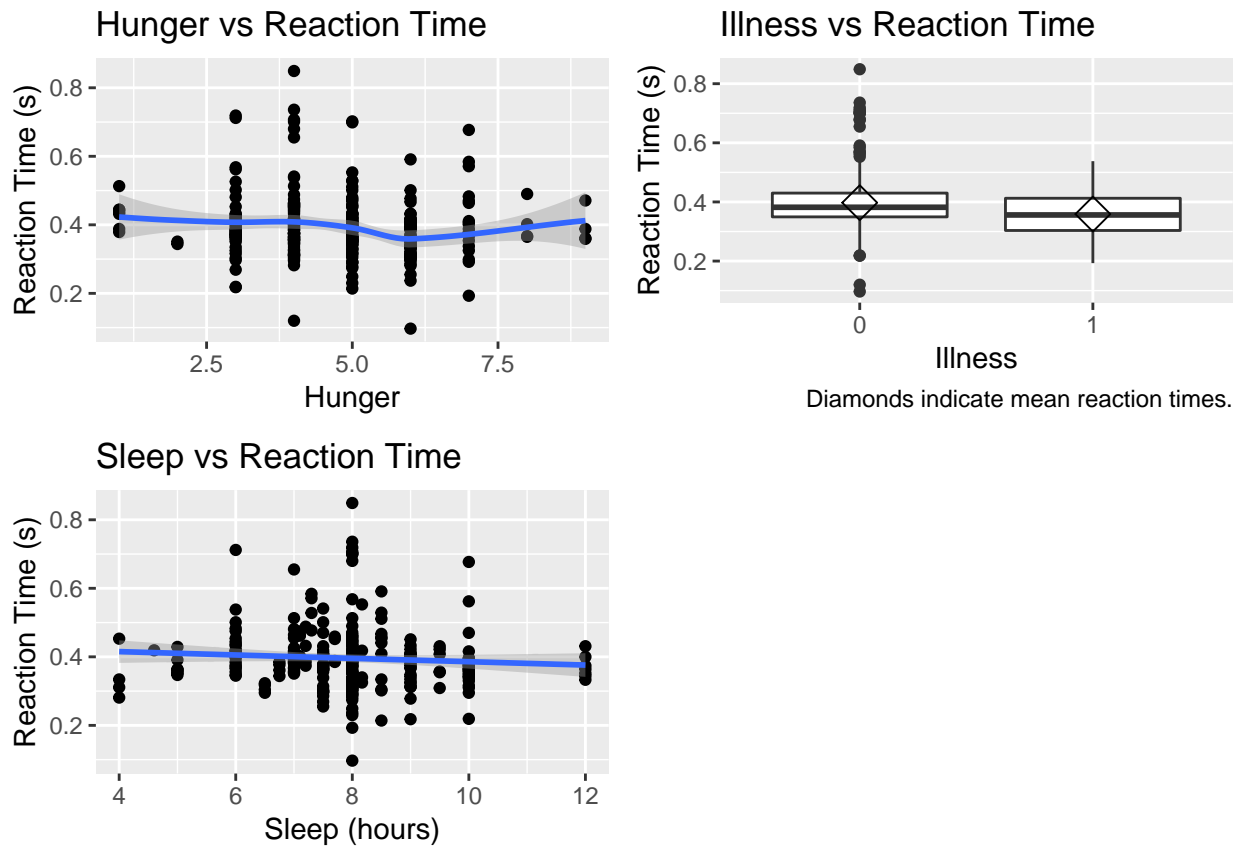
MEQ, Stimulant vs Reaction Time



The downwards sloping curve in the first graph indicates that there may be some negative correlation between MEQ and Reaction Time.

In the second graph, the mean of reaction time appears slightly lower for measurements that were taken after a stimulant. However, there is overlap in interquartile ranges so this difference may or may not be significant.

Hunger, Illness, Sleep vs Reaction Time



Hunger vs Reaction Time:

Although there is a slight correlation between hunger and reaction time, the trend is hard to see because the line of best fit looks quite flat. However, there could possibly be a non linear relationship between hunger and reaction time.

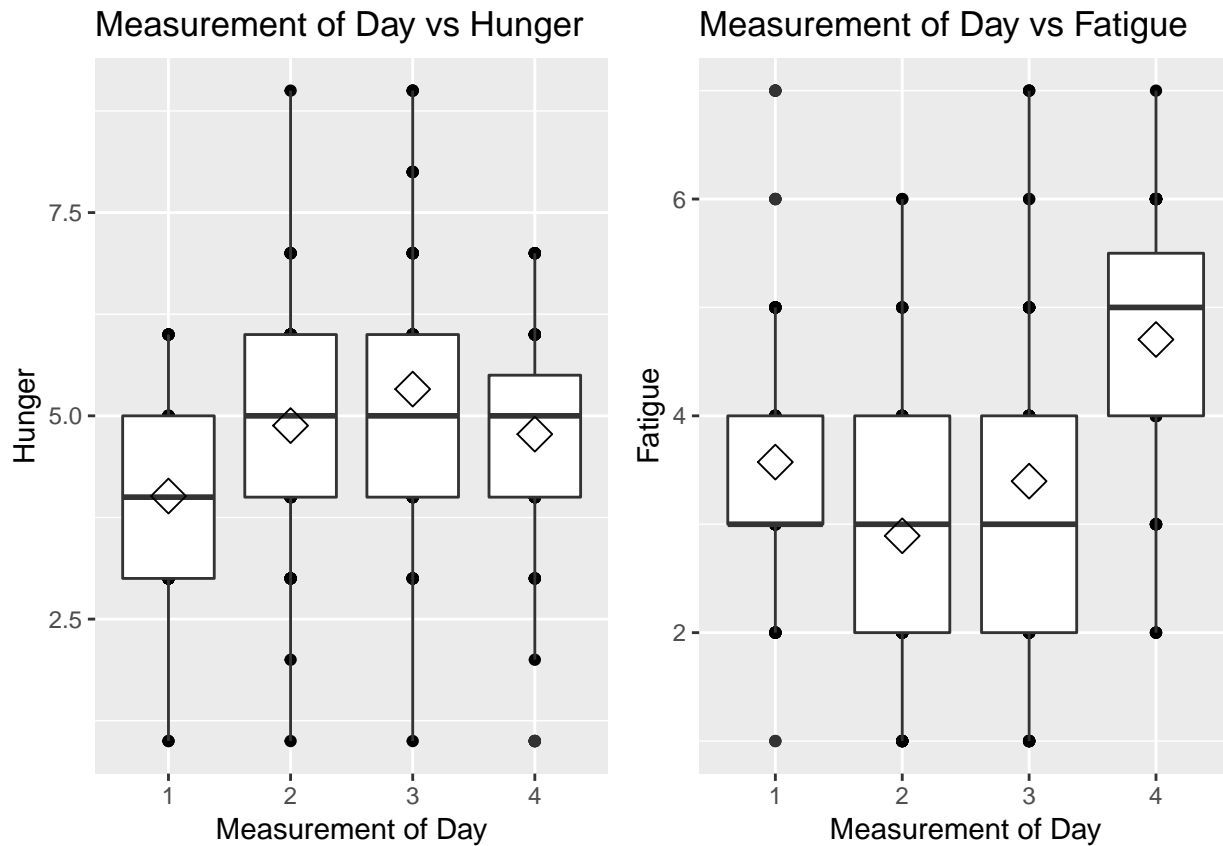
Illness vs Reaction Time:

It appears that reaction time is negatively correlated with illness. However, special care should be taken because 32 observations were taken with an illness and 259 were not so the sizes of the groups are not equal.

Sleep vs Reaction Time:

There appears to be a negative correlation between hours of sleep and reaction time.

Fatigue, Hunger vs Measurement of the Day



Recall that higher hunger score corresponds to being full. By the first plot, it appears that the hunger score is lower in the beginning and end of the day and higher during the middle of the day.

Recall that higher fatigue score corresponds to being more fatigued. From the second plot, it appears that fatigue is highest at the last observation.

Summary Statistics for all variables

Finally, to wrap up, here are the summary statistics for the numerical variables:

Time

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.00	10.01	14.20	14.00	19.48	23.93	33

Sleep

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	4.000	7.000	8.000	7.897	8.500	12.000	44

MEQ

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	32.00	36.00	46.00	44.38	49.00	60.00	24

Fatigue as a continuous variable

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	1.000	3.000	3.500	3.629	5.000	7.000	26

Fatigue as a categorical variable

##	1	2	3	4	5	6	7	NA's
##	20	43	84	61	57	23	6	26

Hunger as a continuous variable

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	1.000	4.000	5.000	4.745	6.000	9.000	26

Hunger as a categorical variable

##	1	2	3	4	5	6	7	8	9	NA's
##	7	2	42	83	76	54	22	4	4	26

MEQ Category

##		Evening	Morning	Neither
##	24	112	8	176

Day Type

##	0	1	NA's
##	152	155	13

Illness

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.00	0.00	0.00	0.11	0.00	1.00	29

Stimulant

##	0	1	NA's
##	247	41	32

Break Protocol

##	0	1	NA's
##	270	6	44

Reaction Time

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	0.0970	0.3470	0.3850	0.3976	0.4320	0.8490	18