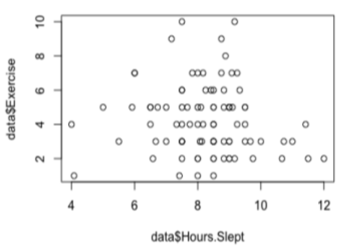
Task 1: Hypothesis

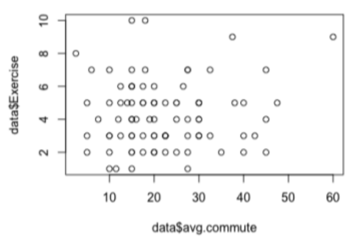
1. <u>Before forming a hypothesis, import and explore the data.</u>

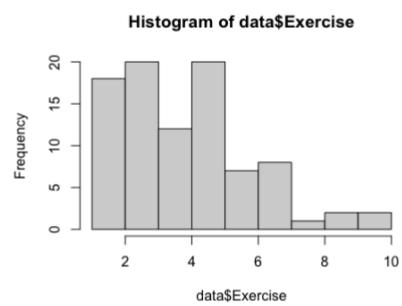
We want to know whether morning routine affects the amount that an individual exercises per week. Average daily commute, hours of sleep, and whether an individual has breakfast are the variables affecting morning routine.

2. <u>Adjust columns to make them the appropriate units and classes for each variable used and remove NAs.</u>

3. <u>Initial exploration of the data to inform the hypothesis: plot exercise against the continuous predictors, check means of each predictor against the categorical variable Breakfast (not shown), and check Exercise distribution.</u>







Hypothesis: We expect individuals to exercise more times per week if they sleep longer and have a shorter average commute time. We expect this relationship to be different when and individual either does or does not have breakfast.

Task 2: Model

1. <u>Predictors are hours slept, average commute, and whether an individual eats breakfast. The response is the number of times an individual exercises per week. Checking the distribution of the response from the earlier histogram: not normal, but it is count data, so we can use a Poisson distribution when creating the model.</u>
2. <u>Check for correlations or interactions between the predictors:</u>

```
print(aggregate(data$Hours.Slept, by=list(data$Breakfast), mean, na.rm =T))

##   Group.1        x
## 1      No 7.903333
## 2     Yes 8.359130 # people who eat breakfast may sleep longer.

#Check if continuous variables are correlated.
cor.test(data$Hours.Slept,data$avg.commute)
# p = 0.5643, cor = -0.06157292. Not correlated.
```

3. <u>Create the MAM</u>

The first, most complex model, with all predictors was created and reviewed with a summary and analysis of deviance:

```r
msi1<- glm(Exercise~ Hours.Slept + Hours.Slept*Breakfast+ I(Hours.Slept^2) + avg.commute, data=data,
family= poisson)


anova(msi1, test="Chisq")# p value of average commute = 0.68261  and is the highest p value so we ca
n remove average commute as a predictor. The AIC was 374.32.
```

Average commute was removed, so now the two final models will confirm whether an interaction term is needed:

```r
m.s<- glm(Exercise~Hours.Slept + Breakfast+ I(Hours.Slept^2), data=data, family= poisson)
# no interaction term.

m.s.i<- glm(Exercise~Hours.Slept+ Hours.Slept*Breakfast + I(Hours.Slept^2), data=data, family= poiss
on)
# interaction term.
```

Results from two remaining models using analysis of deviance: `anova(m.s, m.s.i, test="Chisq")`:

Models	AIC	Residual Deviance	Degrees of Freedom	Deviance	Pr(>Chi)
Hours.Slept + Breakfast + I(Hours.Slept^2)	374.53	76.877	86		
Hours.Slept + Hours.Slept*Breakfast + I(Hours.Slept^2)	372.53	72.879	85	3.9976	0.04557

4. <u>Checking diagnostic plots + final model</u>

Since we are using a GLM, it is not necessary to look at the diagnostic plots, as we do not have the assumptions of residual normality and homogeneity of variances that are necessary for a linear model. We will use the model with the interaction and quadratic terms as it is the best fit for the data. It has the lowest AIC, and the interaction term is significant in reducing the deviance of the model. The model and a table of anova() output are below:

```r
m.s.i<- anova(m.s.i, test="Chisq")
```

```
glm(Exercise~Hours.Slept*Breakfast + I(Hours.Slept^2), data = data, family =
poisson)
```

Coefficients	Degrees of freedom	Deviance	Residual degrees of freedom	Residual Dev.	P(>Chi)
NULL			89	84.306	
Hours.Slept	1	0.8624	88	83.444	0.35307
BreakfastYes	1	2.8629	87	80.581	0.090965
I(Hours.Slept^2)	1	3.7039	86	76.877	0.05429
Hours.Slept:BrekafastYes	1	3.9976	85	72.879	0.04557*

Task 3: Results

While there was no significant effect of hours slept on the number of times an individual exercised per week, the results still indicate a potential quadratic relationship between sleep and exercise (b±SE = -0.042 ± 0.018; $\chi2$ 1 = 3.7039, p = 0.05429). There was a significant effect of breakfast on the relationship between hours slept and exercise (b±SE = 0.189 ±0.097; $\chi2$ 1 = 3.9976, p = 0.04557). Note that the parameter estimates are on the log scale.

Task 4: Plot

Before plotting, two new data frames were created for the predicted values of the model to ensure a smooth line in the final plot and created two separate lines for each category in Breakfast. The plot is on the original scale after back-transforming from the log transformation which was done when using the poisson family in the glm.

```
newdata.Y <- data.frame(Breakfast=rep("Yes", 100),
                    Hours.Slept=seq(min(data$Hours.Slept[data$Breakfast=="Yes"]),
                               max(data$Hours.Slept[data$Breakfast=="Yes"]),
                                              length.out=100))

newdata.N <- data.frame(Breakfast=rep("No", 100),
                        Hours.Slept=seq(min(data$Hours.Slept[data$Breakfast=="No"]),
                                   max(data$Hours.Slept[data$Breakfast=="No"]),
                                               length.out=100))

predicted.Y2 <- predict(m.s.i, newdata.Y, type='response')
predicted.N2 <- predict(m.s.i, newdata.N, type='response')

plot(Exercise ~ Hours.Slept, data=data, pch=NA, xlab="Hours of Sleep", ylab="Exercise (Times per wee
k)")
points(Exercise ~ Hours.Slept, data=data[data$Breakfast=="Yes", ], pch=19, col="blue")
points(Exercise ~ Hours.Slept, data=data[data$Breakfast=="No", ], pch=19, col="red")
lines(predicted.Y2[order(newdata.Y$Hours.Slept)] ~
        sort(newdata.Y$Hours.Slept), lwd=1.5, col="blue")
lines(predicted.N2[order(newdata.N$Hours.Slept)] ~
        sort(newdata.N$Hours.Slept), lwd=1.5, col="red")
legend(x="topleft", legend=c("Yes", "No"), pch=19,
col=c("blue", "red"), lwd=c(1,1), title="Eats Breakfast", cex=0.8)
```

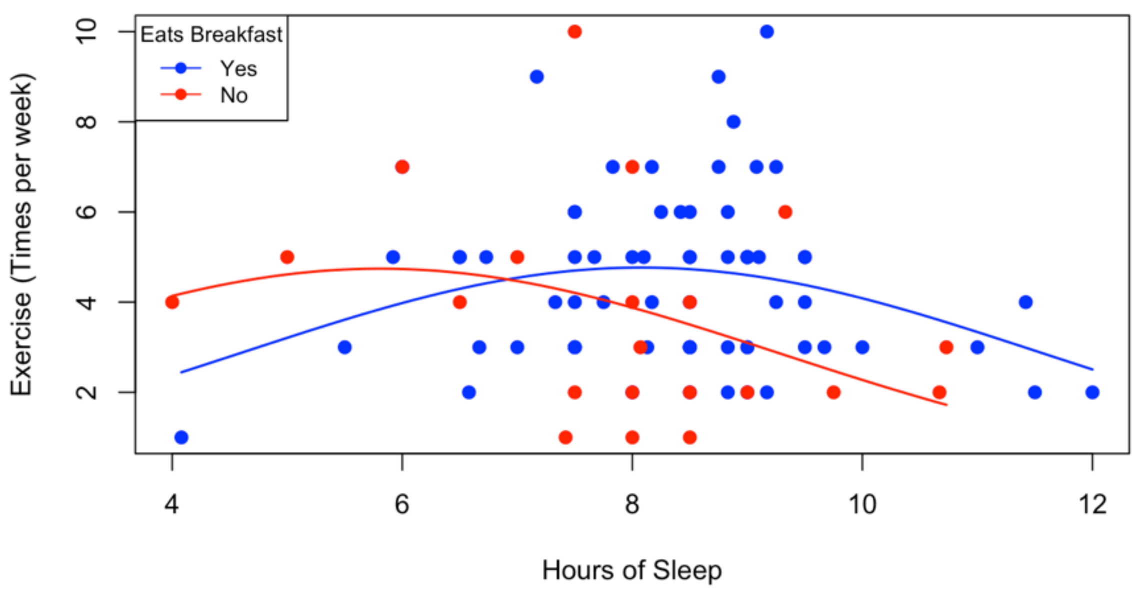


Figure 1. Effect of morning routing on exercise.

The model indicates a negative quadratic relationship between hours of sleep and number of times an individual exercises per week, however. There is also a significant difference in individuals who do and don't eat breakfast and their relationship between sleep and exercise, with each group having a peak number of times exercised for different hours of sleep. At the recommended 8 hours of sleep for adults who eat breakfast, the predicted number of times exercised is 4.76. Adults who don't eat breakfast exercise more with less sleep, and at the recommended 8 hours their predicted exercise is 3.88.