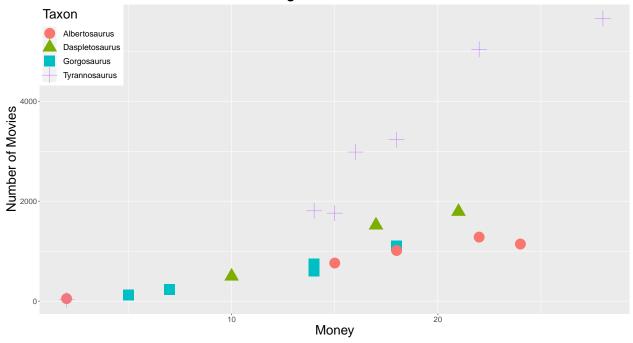# ST_540 Assigment 1

Tilekbek Zhoroev

2 (a) Write a loop to compute the mean and standard deviation of the mass for each taxon. You must write a loop to get full credit. Put the results in a table and make sure the rows and columns are labelled clearly.

```r
#read data
data<-read.csv("Growth curves data.csv")
# find the all unique Taxon Names
all_taxons <- unique(data$Taxon)
n<-length(all_taxons)
# define means and standard deviation and evaluate using for loop
means = numeric(n)
stds =numeric(n)
for(i in 1:n){ means[i]=mean(data$Mass[data$Taxon==all_taxons[i]])
             stds[i]=sd(data$Mass[data$Taxon==all_taxons[i]])}
# create data frame
table=data.frame(means, stds)
# define column and row names of the data frame
colnames(table)=c("Means", "Standar deviation")
rownames(table)=all_taxons
table
```

```
##                   Means Standar deviation
## Gorgosaurus     563.000          397.2430
## Albertosaurus   849.860          486.2409
## Daspletosaurus 1268.333          682.6466
## Tyrannosaurus  2929.414         1958.0373
```

b) Make a plot of age versus mass that includes all observations but a different plotting symbol (i.e., the pch option in plot) or color for each taxon. Make sure the axes and legends are clearly labeled.

```r
library(ggplot2)
ggplot(data, aes(x=Age, y=Mass, colour=Taxon, shape=Taxon))+
xlab("Money")+
ylab("Number of Movies")+
ggtitle("Ages versus mass")+
geom_point(size=9, alpha=1)+
theme(axis.title.x=element_text(colour="Black", size=24),
      axis.title.y=element_text(colour="Black", size=24),
    axis.text.x=element_text(size=14),
    axis.text.y=element_text(size=14),
    legend.title=element_text(size=24),
    legend.text=element_text(size=15),
    legend.position=c(0,1),
    legend.justification=c(0,1),
    plot.title=element_text(size=24,face = "bold", hjust=.5),
    )
```

Ages versus mass

```r
lm_model=lm(log(Mass)~log(Age),data = data[,3:4])
summary(lm_model)
```

```
##
## Call:
## lm(formula = log(Mass) ~ log(Age), data = data[, 3:4])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8279 -0.3692 -0.1350  0.4475  0.8437
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.2946     0.4206   5.456 3.50e-05 ***
## log(Age)      1.7538     0.1598  10.977 2.09e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5199 on 18 degrees of freedom
## Multiple R-squared:  0.87,   Adjusted R-squared:  0.8628
## F-statistic: 120.5 on 1 and 18 DF,  p-value: 2.09e-09
```

Firstly the call function show us the formula R used to fit the data to the linear regression model. Next the residuals are difference between actual data and model prediction. We have given 5 number summary of residuals. From this summary we can observe that these numbers are summetric with mean near to zero. It implies that the residuals are normally distributed or errors are iid, and we can conclude that overall the linear model predicts well. The next Coefficients give us information about linear model and using intercept(log(mass)) and slope (log(age)) we can estimate other points, it could be outside of the data. Each columns are informations about estimate, standart error, t-value and P(>|t|). Here we can note that in our model p-values are very close to zero. The Residual standard error tells about average error of the model and degree of freedom = 18 (all data points(20)- number of parameters(slope and intercept)). Next, we have given the other type of error metric forms.