

SGupta_HW01

Download the tyrannosaurid growth curves data

```
# Read the data from a CSV file
growth_data <- read.csv("/Users/saurabhgupta/projects/statistics/ST540/GrowthCurvesData.csv")
# Display the first few rows of the data to verify
head(growth_data)
```

	Taxon.ID		Taxon	Age	Mass
1	1	Gorgosaurus	5	127.0	
2	1	Gorgosaurus	7	229.0	
3	1	Gorgosaurus	14	607.0	
4	1	Gorgosaurus	14	747.0	
5	1	Gorgosaurus	18	1105.0	
6	2	Albertosaurus	2	50.3	

(a) Write a loop to compute the mean and standard deviation of the mass for each taxon.

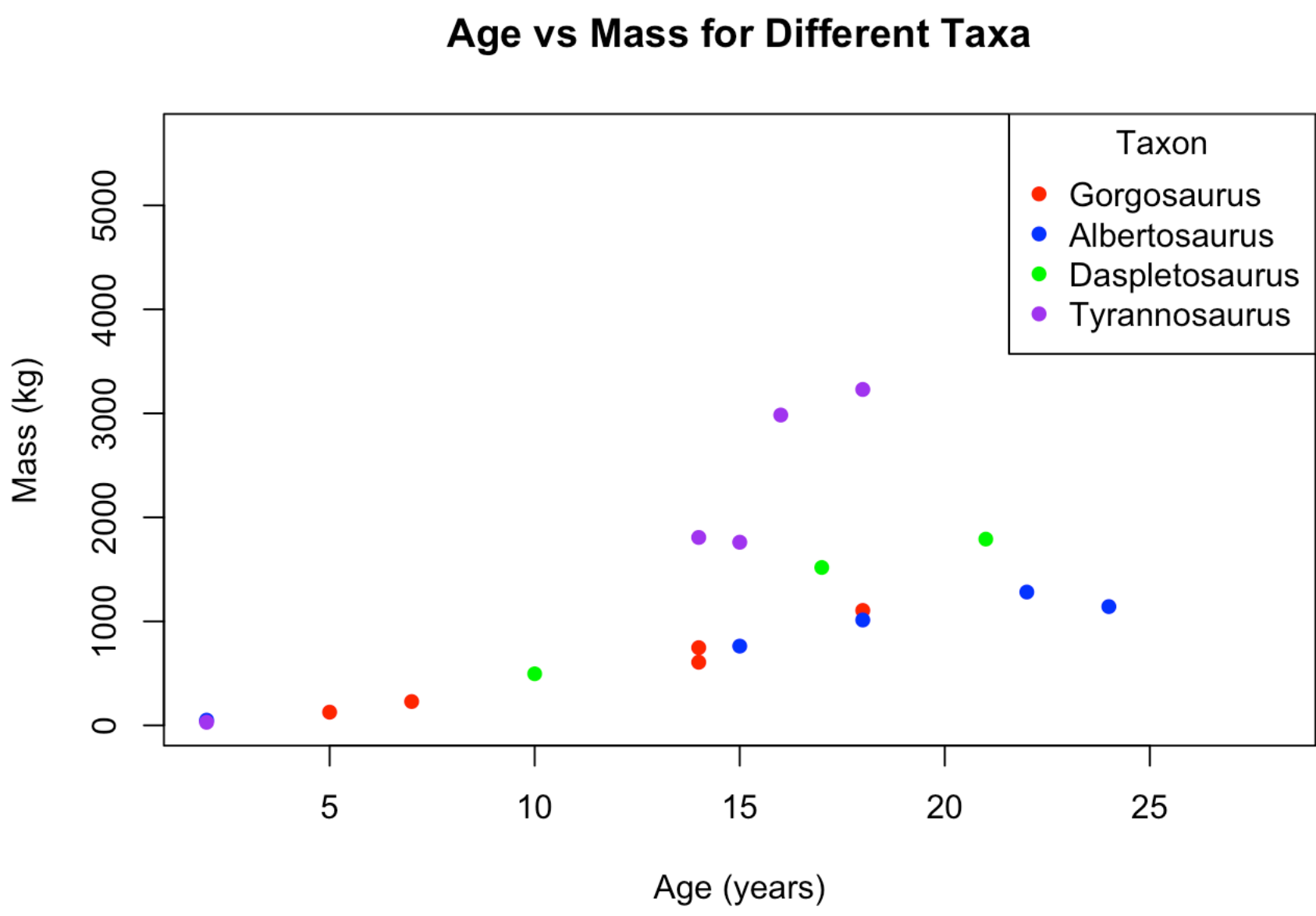
```
# Unique taxa
unique_taxon <- unique(growth_data$Taxon)
# Initialize a results dataframe
results <- data.frame(Taxon = character(), Mean_Mass = numeric(), Std_Dev_Mass = numeric())
# Loop through each taxon
for (taxon in unique_taxon) {
  # Filter data for the current taxon
  taxon_dataset <- growth_data[growth_data$Taxon == taxon, ]
  # Compute mean and standard deviation
  mean_mass <- mean(taxon_dataset$Mass)
  std_dev_mass <- sd(taxon_dataset$Mass)
  # Append the results
  results <- rbind(results, data.frame(Taxon = taxon, Mean_Mass = mean_mass, Std_Dev_Mass = std_dev_mass))
}
# Display the results
print(results)
```

	Taxon	Mean_Mass	Std_Dev_Mass
1	Gorgosaurus	563.000	397.2430
2	Albertosaurus	849.860	486.2409
3	Daspletosaurus	1268.333	682.6466
4	Tyrannosaurus	2929.414	1958.0373

(b) Make a plot of age versus mass that includes all observations but a different plotting symbol or color for each taxon.

```
# Assign colors for each taxon
colors <- c("red", "blue", "green", "purple")
taxon <- unique(growth_data$Taxon)
growth_data$Color <- colors[match(growth_data$Taxon, taxon)]

# Plot Age vs Mass
plot(
  growth_data$Age, growth_data$Mass,
  col = growth_data$Color, pch = 16,
  xlab = "Age (years)", ylab = "Mass (kg)",
  main = "Age vs Mass for Different Taxa"
)
# Add legend
legend(
  "topright", legend = taxon, col = colors, pch = 16,
  title = "Taxon"
)
```



(c) Perform a non-Bayesian linear regression of y=log(mass) onto x=log(age) (ignoring taxon), and interpret the results.

```
# Log-transform age and mass
growth_data$log_Age <- log(growth_data$Age)
growth_data$log_Mass <- log(growth_data$Mass)
# Perform linear regression
regression_model <- lm(log_Mass ~ log_Age, data = growth_data)
# Display summary of the regression
summary(regression_model)
```

Call:
lm(formula = log_Mass ~ log_Age, data = growth_data)

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

Regression model: $\log(\text{mass}) = \beta_0 + \beta_1 \log(\text{age}) + e$

β_0 is Intercept and β_1 - Slope which indicates the relationship between $\log(\text{age})$ and $\log(\text{mass})$. Specifically, it shows the proportional change in mass for a proportional change in age.

The residuals range from -0.8279 to 0.8437, with the median residual being -0.1350. This suggests a reasonably symmetric distribution of residuals

The intercept is 2.2946 which indicates when $\log(\text{age})=0$ the predicted $\log(\text{mass})$ is 2.2946 The slope is 1.7538 which means 1% increase in age correspond to approx a 1.75% increase in mass on average

R-squared: Indicates how well the model explains the variance in the data (higher values mean a better fit).It is 0.87 which means our model is a good fit.

p-values: Assess the statistical significance of the coefficients ($p<0.05$), suggests model is highly significant overall, indicating that $\log(\text{Age})$ is a strong predictor of $\log(\text{Mas})$.