Homework 2

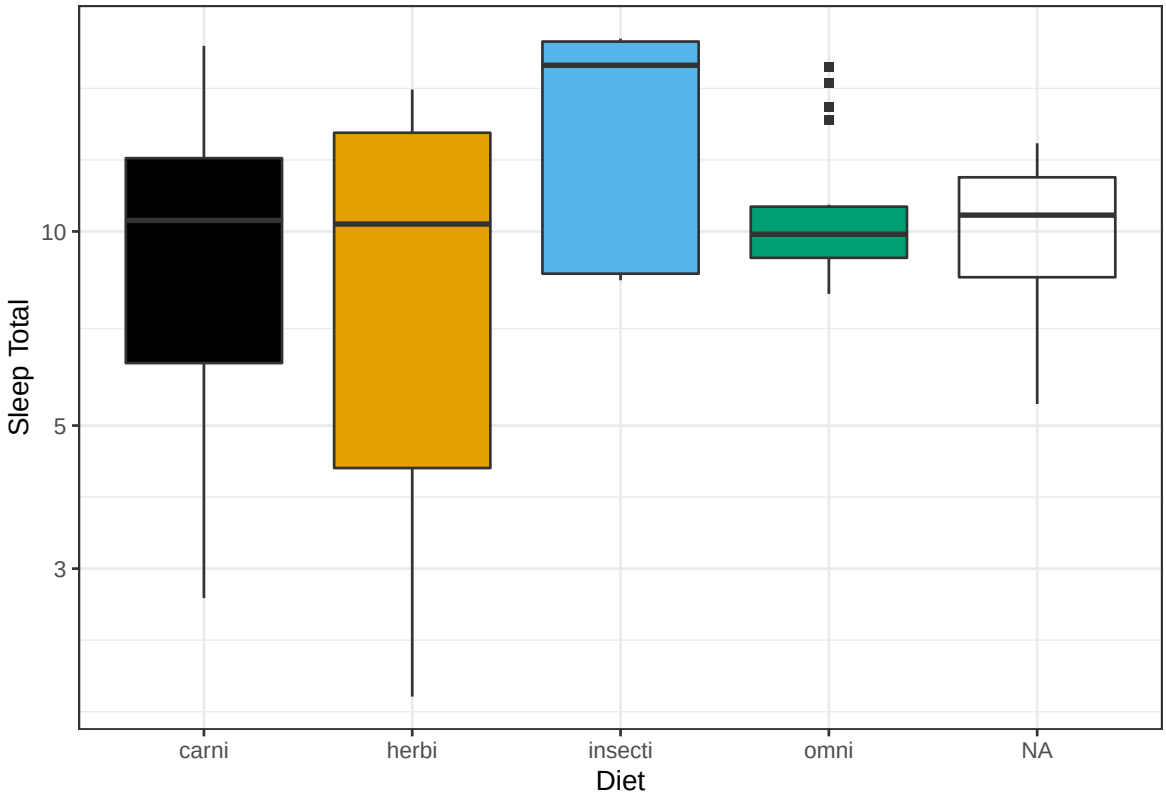
Your Name Here

Due Date Here

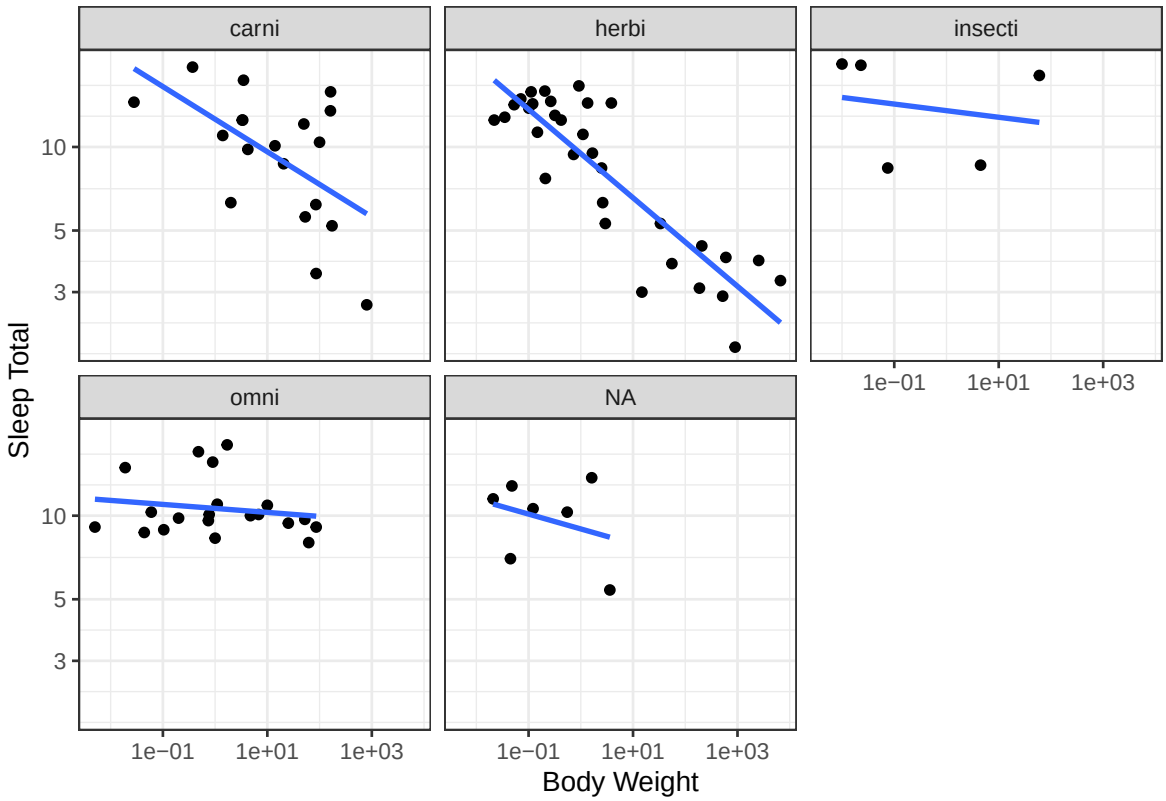
Exercise 1 (STAT 612: 10 points; STAT 412: 15 points)

Complete the following exercises.

1. Load `msleep` data frame into R.
2. How many mammals are in the `msleep` data frame? How many variables? Use two R functions to get this information.
3. Plot body weight against the total amount of sleep. Hint: think of the appropriate plot to explore their association. Hint: What is the response variable and what is the explanatory variable?
4. When you see a curved relationship in a plot, you can often get rid of these curves by taking a log transformation of either the explanatory or the response variable (or a log transformation of both). Do this until you get an appropriately linear relationship and plot the results.
5. Color code the plot in part 4 by the diet of the animals (`vore`). Make the axis labels nice, change the theme to black and white, and add a title.
6. In the plot from part 5, add the OLS line (**without** standard errors) to each `vore` category. Does the effect of body weight on sleep total appear larger for some diets?
7. Also add the overall OLS line (**without** standard errors) to the above plot (when you include all observational units). Make sure this line is black and dashed.
8. Change the title of the legend to "Diet".
9. Reproduce the following plot (hint: I used the colorblind safe palette):



10. Reproduce the following plot:



Exercise 2: Midwest Data (STAT 612: 10 points; STAT 412: 15 points)

- For this exercise, we'll use the `midwest` data from ggplot2.

1. Load these data into R.
2. What are the observational units?
3. Calculate the total population in each state. Display the data in an appropriate plot
4. Make a scatterplot of population density against the percent college educated on a scale that is appropriately linear. You should exclude counties in the bottom tenth percentile of total population. Color code the counties by state. Add the overall OLS line (the one that takes all points into account). Make the OLS line black and dotted.
5. Make an appropriate plot to explore the association between the state and the percentage adults. Make sure the axis labels are nice and use the black-and-white theme.
6. Use an R function to determine the possible values of `state`.
7. In the `state` variable, replace `IL` with `Illinois`. Replace `IN` with `Indiana`, etc...