Programming - Conditions

Data Science and Business Analytics

Exercise 1 – Functions and Conditions

a) Write a function with input parameter a numeric value:
   - If the input parameter is NA (is.na()), return a warning()-message
   - If it is positive, print “positive”
   - If negative, print “negative”
   - If 0, print “exactly 0”

b) Try some values for the input parameter. Specifically, try
   sqrt(7)^2 – 7
   What is the expected answer and what is the answer of your function in a)?

Exercise 2 – Iteration and Conditions

a) Write a function with a vector as input and with return value:
   - the mean value if the vector is numeric
   - the mode if the vector is character
     note that there is no standard mode()*-function in R! How to find the value of the input vector which has the most observations?*

   Give the function a convenient name

b) Use the animal_species.csv data-set as available on Canvas to execute the function in a) for all columns of the data frame. In what format would you return these values (what map()-function would you use)?

Exercise 3 – Conditions within dplyr

a) Use the animal_species.csv data-set. Create a new column which divides the dataset into 5 groups:
   - a group where year is smaller than or equal to 1989 and weight is smaller than 37
   - a group where year is smaller than or equal to 1989 and weight is larger than or equal to 37
   - a group where year is larger than 1989 and weight is smaller than 37
   - a group where year is larger than 1989 and weight is larger than or equal to 37
   - a group where weight is missing

   Which function would you like to use?

b) Create a boxplot using the new grouping variable of a) to see the differences in distribution of hindfoot_length over the groups created in a)