Data transformation 2 (Oct 1)

Rename function is used to change a variable's name. Doesn't change the dataset but creates a new one.
rename(data, new var name = old var name)

Pipeline operator: %>%
The pipeline operator is used when you need to use multiple functions. Ex. select, arrange, filter
Data %>% filter(...) will insert that result into the dataset
Using multiple functions: data %>% filter(...)%>%select(...)%>%arrange(...)
You can use the pipeline operator to plot specific data: data %>%filter(...)%>%ggplot()+...

Summarize function, doesn't change dataset but creates new dataset with summary statistics
Creates new variables that can summarize the data
data%>% summarize(var1 = sum(n), var2 = max(n))
This will output var1 as the sum of all n and var2 as the max n value

Lag function returns the value before another, essentially shifts the data over by 1
Lead function does the opposite of lag, it will return the value after another
Ex. a=1:10
lag(a) = NA 1 2 3 4 5 6 7 8 9
lead(a) = 2 3 4 5 6 7 8 9 10 NA

Cumsum function will find cumulative sum
Ex. a = 1:3
cumsum(a) = 1 3 6

Group by function will help to group data before you use another function
Data %>% group_by(variable)
Goes well with summarize as it groups the summary by each variable.

Mutate function helps add new variables to original dataset
mutate(new var = (old var * 100) Creates new variable which is 100xold variable

Transmute function works the same as mutate but only outputs the new variables, not the old