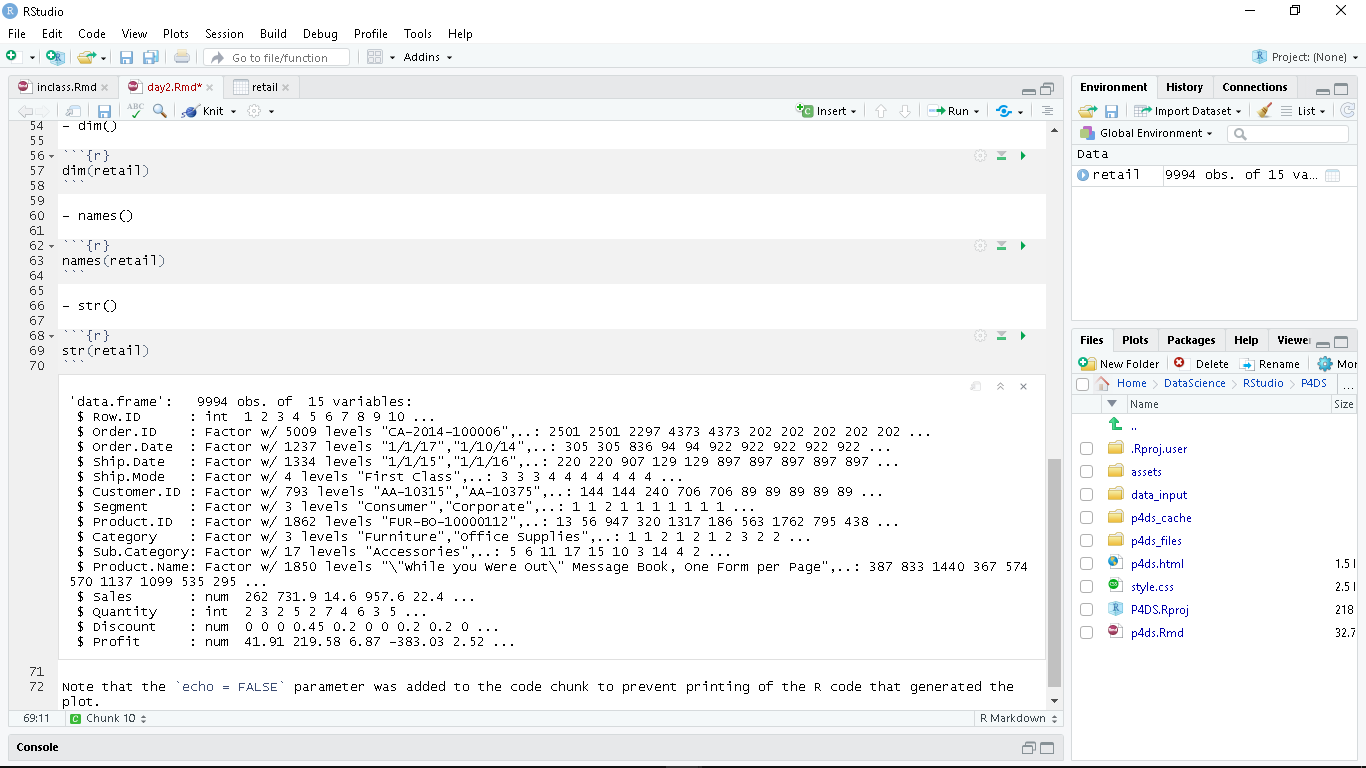
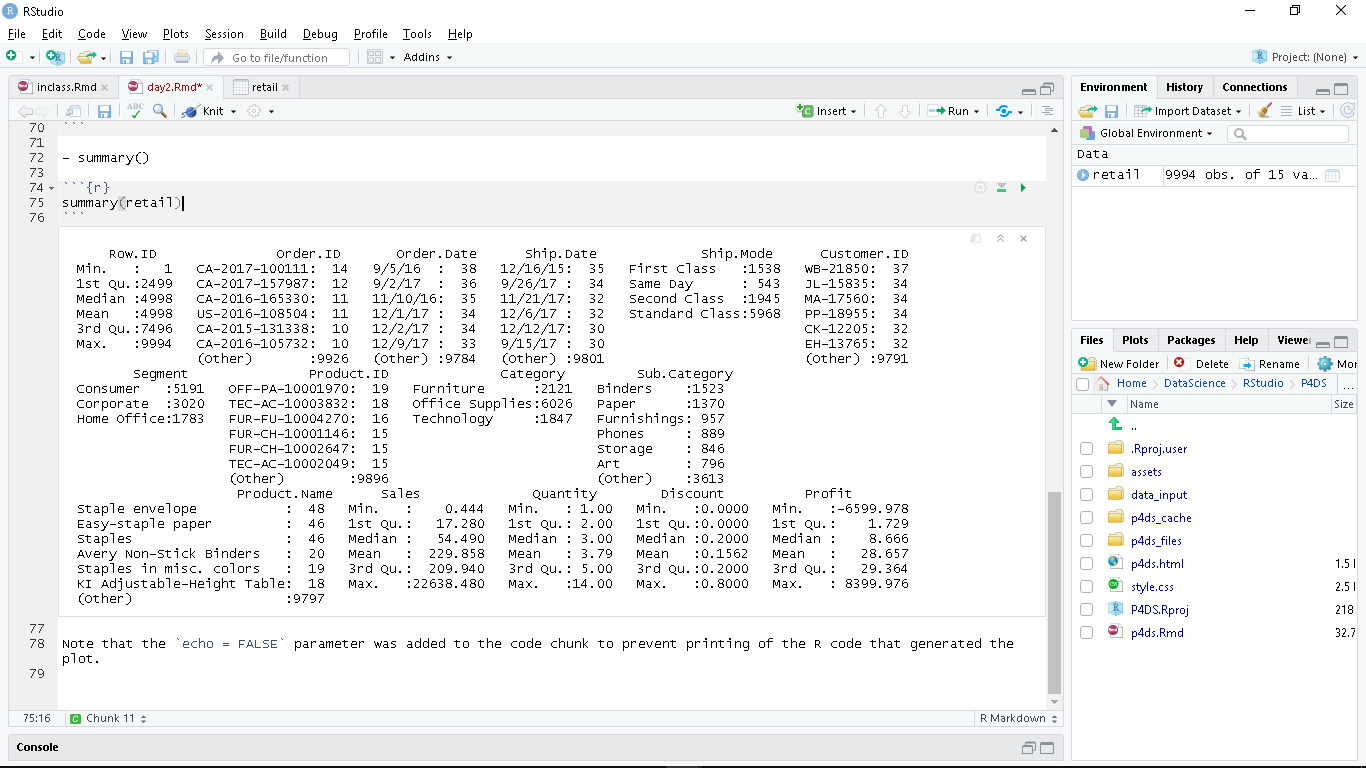
Data Science Notes (RStudio) (Day 2)

* Exploratory Data Analysis
  + *‘read.csv()’*
    - To insert a csv file into our environment
    - To set the separator, you can add parameter *‘sep’*
    - Other way to do it, click ‘import dataset’ and choose ‘read text(base)’
    - How to use:
      * **\*variable name\*** <- *read.csv(***\*directory of the data\****)*
  + *‘head()’*
    - To display several first rows of our current data
    - To control how many rows you want to display, you can add parameter with a number
    - How to use:
      * *head(***\*variable name\****)*
  + *‘tail()’*
    - To get several last rows of our current data
    - To control how many rows you want to display, you can add parameter with a number
    - How to use:
      * *tail(***\*variable name\****)*
  + *‘dim()’*
    - To get the dimension (row and column) of our current data
    - How to use:
      * *dim(***\*variable name\****)*
  + *‘names()’*
    - To display every column’s names
  + *‘str()’*
    - To display the structure of our current data.
    - Result of the function:
      * Dimensions
      * Variables’ names
      * Every data’s types
      * Some parts of its data
    - How to use:
      * *str(***\*variable name\****)*
    - How to look at *‘str()’* result (picture below):



* + ‘summary()’
    - To display the summary of the current data
    - Each type of data displays different type of summary
      * Numeric
        + Displays the minimum, 1st quarter, median, mean, 3rd quarter and maximum
      * Integer
        + Displays the minimum, 1st quarter, median, mean, 3rd quarter and maximum
      * Date
        + Displays the minimum, 1st quarter, median, mean, 3rd quarter and maximum
      * Factor
        + Displays top 6 levels’ total occurrences with ‘other’ is the rest of the levels’ occurrences.
      * Character
        + Only displays the length, class type and mode type
    - How to use:
      * *summary(***\*variable name\****)*



* Manipulation
  + ‘as.character()’
    - To manipulate a column’s data type into Character
    - How to use:
      * **\*variable name\***$**\*column name\*** <- *as.character(***\*variable name\***$**\*column name\****)*
  + ‘as.Date()’
    - To manipulate a column’s data type into Date
    - How to use:
      * **\*variable name\***$**\*column name\*** <- *as.Date(***\*variable name\***$**\*column name\***, **\*format date\****)*
      * Format date:
        + Month/date/year -> *“%m/%d/%y”*
        + Date/month/year -> *“%d/%d/%y”*
* Subset and Sampling Data
  + *‘[,]’*  -> ([ROW, COLUMN])
    - To display selected row and column
    - To display selected rows in specific, use function *c()* at the ROW section
    - To ‘erase’ a specific column, use subset function (*‘subset()’*)
    - How to use:
      * **\*current data\****[***\*selected row\****,***\*selected column\****]*
  + *‘table()’*
    - To display the total amount of a current data’s column
    - To display in percentage, use *‘prop.table()’* on the outside of the ‘table()’
    - **Always** execute *’prop.table()’* with table-type data as an input
    - How to use:
      * In Integer: *table(***\*current data\****$***\*current column\****)*
      * In percentage (Numeric): *prop.table(table(***\*current data\****$***\*current column\****))*
* Extra Notes
  + Commands
    - Check Working Directory -> *‘getwd()’*
    - Set Working Directory -> *‘setwd(***\*directory\****)’*
    - To look at the data in detail -> *‘View(***\*data at environment\****)’*
    - To exclude some part, use negative sign
  + Shortcuts:
    - To open compact console -> *CTRL + ALT + ‘I’*
* Exercise (With Answers)
  + How many ‘Technology’ category that used ‘Second Class’?
    - Answer(s):
      * Way #1:
        + *table(***retail***$***Category***)*
      * Way #2:
        + number1 <- retail[]
  + How many items that has profit > 100 that used ‘First Class’?
    - Answer(s):
      * Way #1:
        + Number2 <- retail[retail,Profit>100]
        + table(number2$Ship.Mode)
      * Way #2
        + Table(retail$Profit>100 & retail$Ship.Mode==”First Class”)