Create Codebook

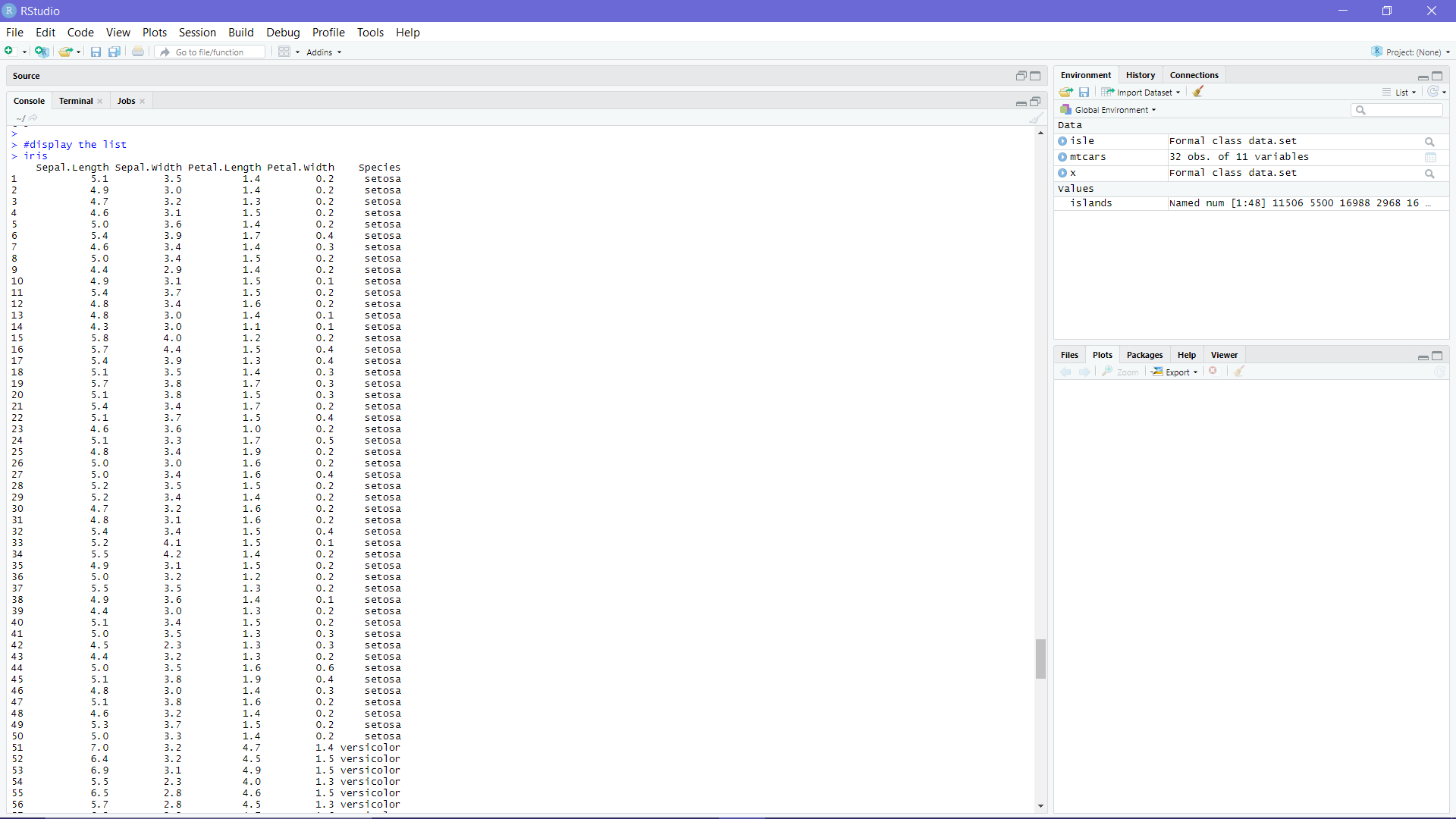
Methods used

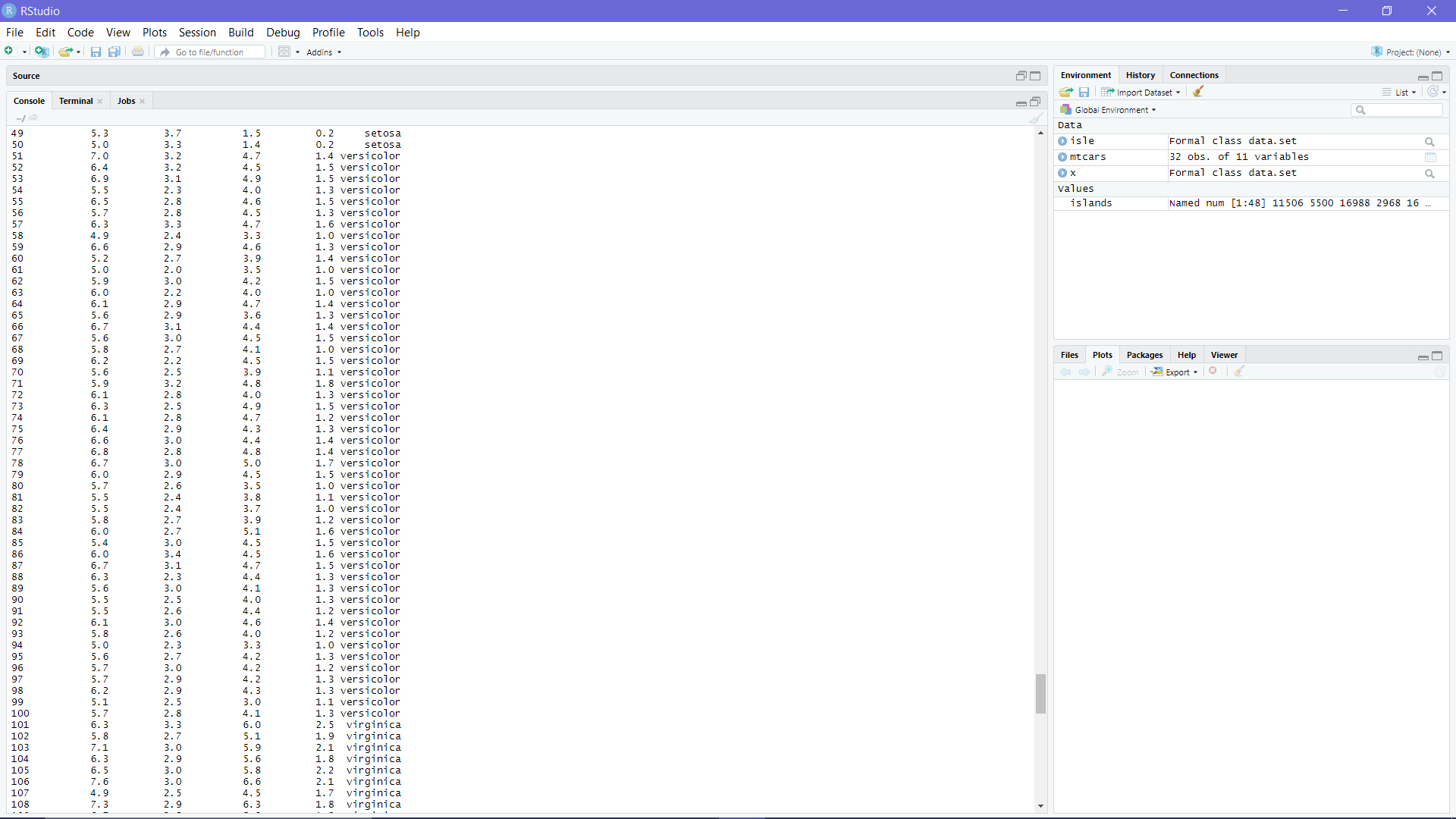
**1) Creating the codebook using standard R API functionality**

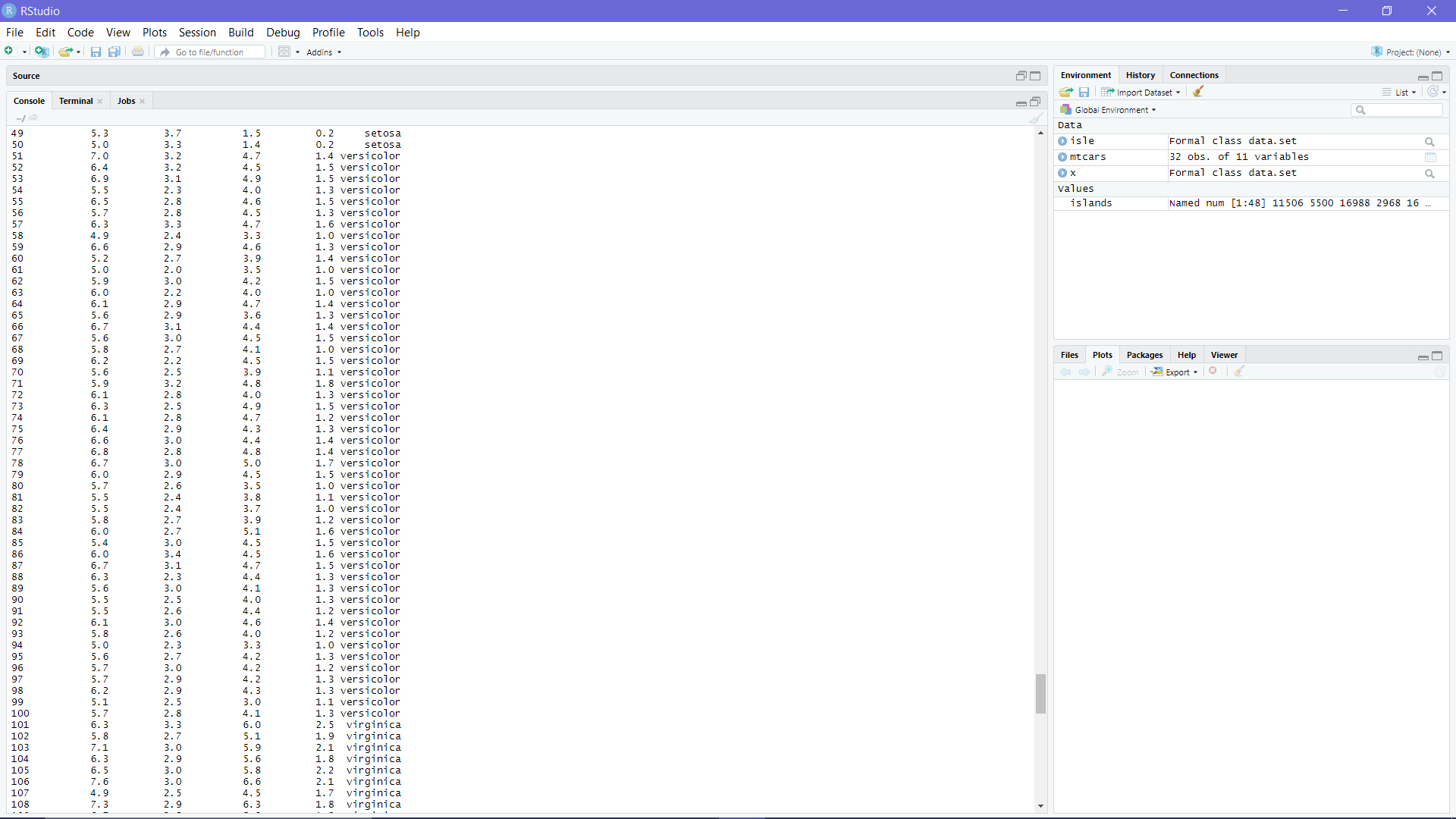
Below is the code used in RStudio to generate the Codebook

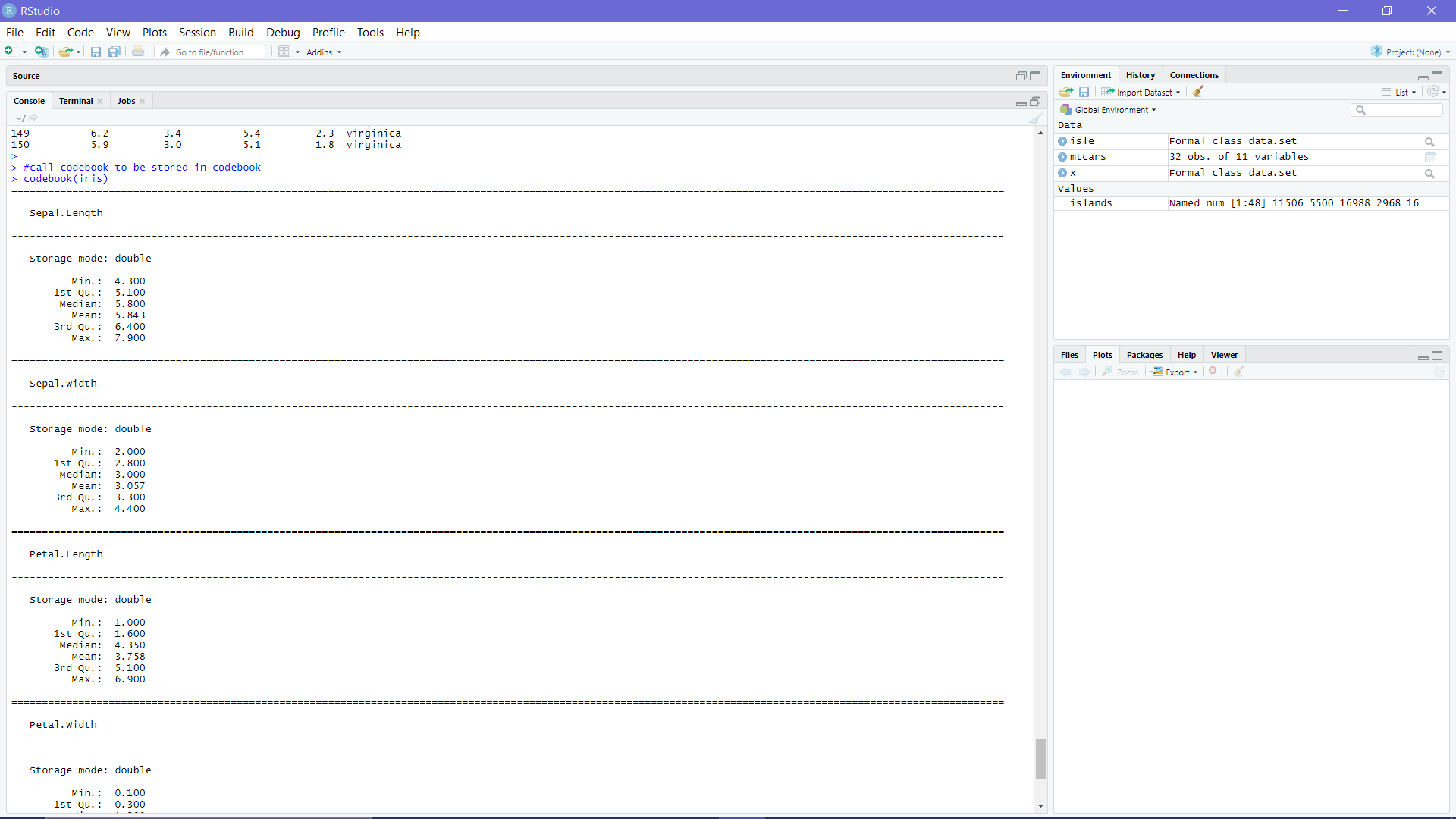
|  |
| --- |
| #install package  install.packages("memisc")  #call the package used  library(memisc)  #declare iris as a list of data set for iris  #iris is about Areas of the World's Major Landmasses  iris <- data.set(iris)  #check the variable of iris  typeof(iris)  #display the list  iris  #call codebook to be stored in codebook  codebook(iris)  #summary of the whole data stored in codebook  summary(iris) |

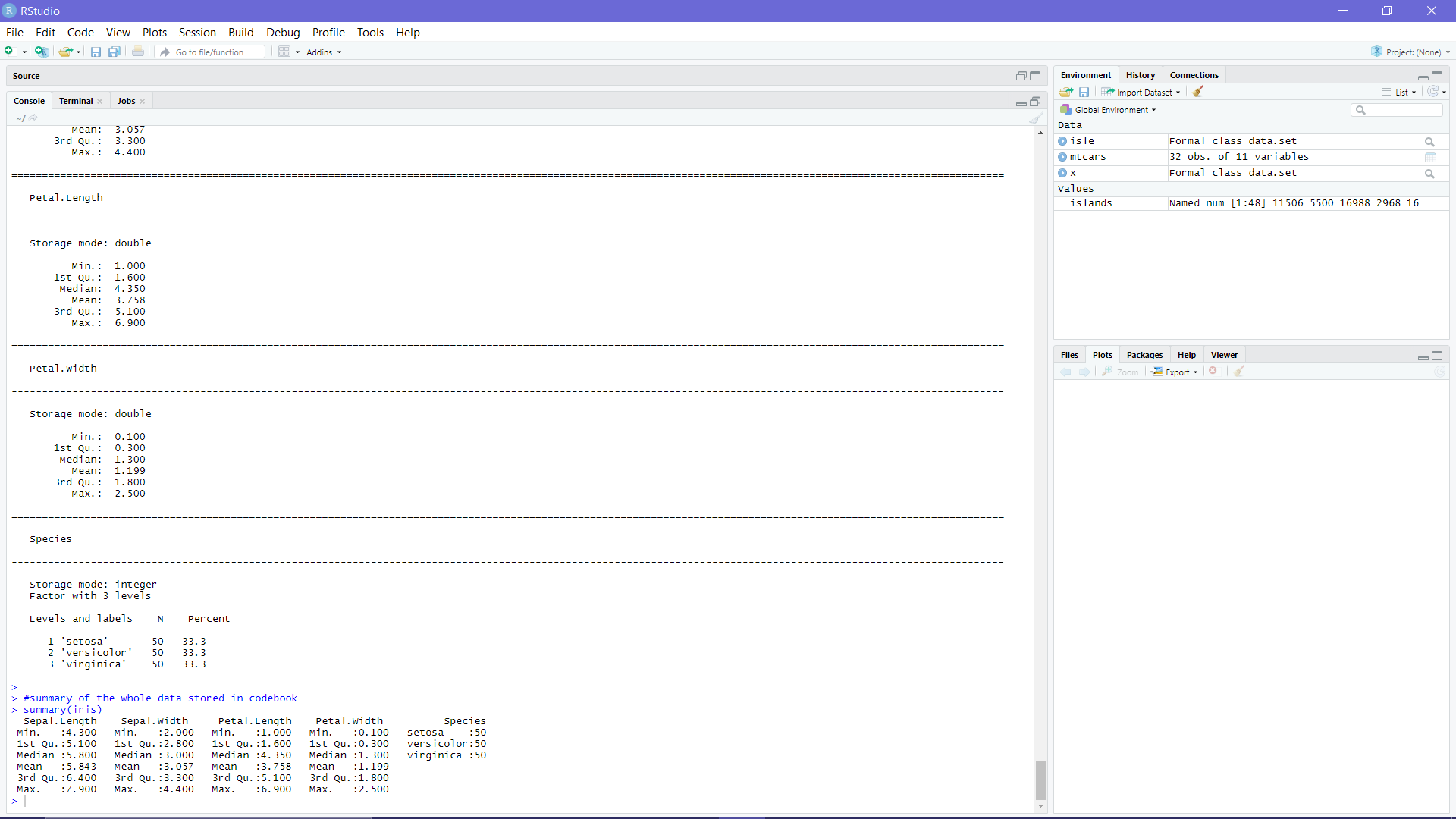
Below is the result displayed by RStudio regarding the data set “iris”











Description of Codebook:

* While displaying the iris using code

|  |
| --- |
| #display the list  iris |

RStudio will display **all** the **raw** data of the surface area of the World's Major Landmasses in list form. The data include the Sepal Length, Sepal Width, Petal Length, Petal Width and Species

* while calling for codebook and storing in code

|  |
| --- |
| #call codebook to be stored in codebook  codebook(iris) |

RStudio will display the data is **transformed into statistic about the data** including

Min.,1st Quartile, Median, Mean, 3rd Quartile and Max. These statistics are all classified by the category stated on list.

* while summary for codebook

|  |
| --- |
| #summary of the whole data stored in codebook  summary(iris) |

RStudio will display the data in a **summary form or table form of all the data stores in codebook.**

**1) Code to automatically generate a codebook - Using dataMaid Package**

Below is the code used in RStudio to generate the Codebook

|  |
| --- |
| #install the package  install.packages("dataMaid")  #call library to be used to make codebook  library(dataMaid)  #declare which data set to be used  data("iris")  #make a codebook for iris  makeCodebook(iris) |

Below is the result displayed by RStudio regarding the data set “iris”

Codebook for iris

Autogenerated data summary from dataMaid

2020-04-28 10:19:51

# Data report overview

The dataset examined has the following dimensions:

|  |  |
| --- | --- |
| Feature | Result |
| Number of observations | 150 |
| Number of variables | 5 |

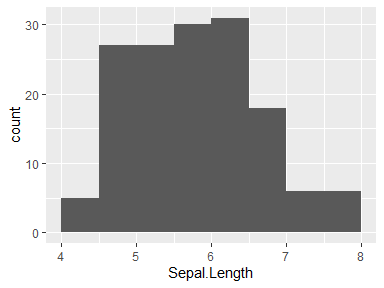
# Codebook summary table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Label | Variable | Class | # unique values | Missing | Description |
|  | [**Sepal.Length**](#sepal.length) | numeric | 35 | 0.00 % |  |
|  | [**Sepal.Width**](#sepal.width) | numeric | 23 | 0.00 % |  |
|  | [**Petal.Length**](#petal.length) | numeric | 43 | 0.00 % |  |
|  | [**Petal.Width**](#petal.width) | numeric | 22 | 0.00 % |  |
|  | [**Species**](#species) | factor | 3 | 0.00 % |  |

# Variable list

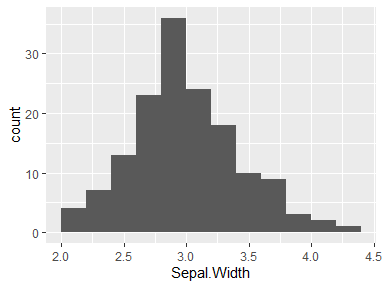
## Sepal.Length

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 35 |
| Median | 5.8 |
| 1st and 3rd quartiles | 5.1; 6.4 |
| Min. and max. | 4.3; 7.9 |



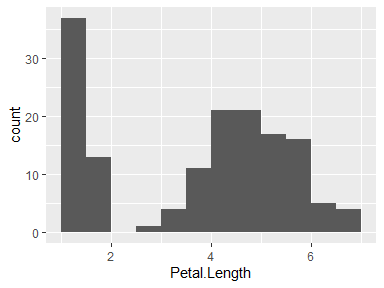
## Sepal.Width

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 23 |
| Median | 3 |
| 1st and 3rd quartiles | 2.8; 3.3 |
| Min. and max. | 2; 4.4 |



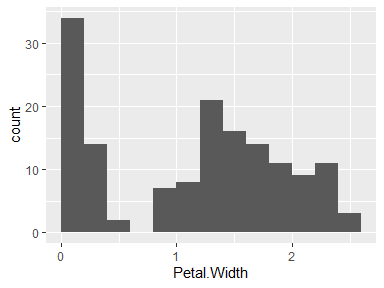
## Petal.Length

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 43 |
| Median | 4.35 |
| 1st and 3rd quartiles | 1.6; 5.1 |
| Min. and max. | 1; 6.9 |



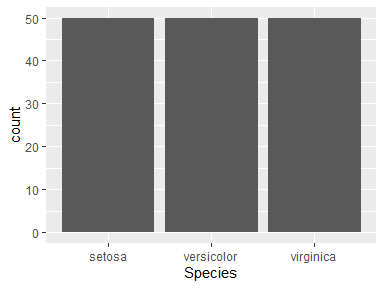
## Petal.Width

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | numeric |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 22 |
| Median | 1.3 |
| 1st and 3rd quartiles | 0.3; 1.8 |
| Min. and max. | 0.1; 2.5 |



## Species

|  |  |
| --- | --- |
| Feature | Result |
| Variable type | factor |
| Number of missing obs. | 0 (0 %) |
| Number of unique values | 3 |
| Mode | “setosa” |
| Reference category | setosa |



* Observed factor levels: "setosa", "versicolor", "virginica".

Report generation information:

* Created by: Could not determine from system (username: Unknown)
* Report creation time: Tue Apr 28 2020 10:19:51
* Report was run from directory: C:/Users/user/OneDrive/Documents
* dataMaid v1.4.0 [Pkg: 2019-12-10 from CRAN (R 3.6.3)]
* R version 3.6.2 (2019-12-12).
* Platform: x86\_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 18362)).
* Function call: dataMaid::makeDataReport(data = iris, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook\_iris.Rmd", checks = list( character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven\_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for iris")

Description of Codebook:

* While displaying the iris using code

|  |
| --- |
| #call library to be used to make codebook  library(dataMaid) |

RStudio will call library and use package “dataMaid” to be used to make a codebook.

* while calling for codebook and storing in code

|  |
| --- |
| #declare which data set to be used  data("iris") |

RStudio will compile all data regarding data set iris into the dataMaid’s library. This data will be stored in dataMaid as set of data to be used.

* while summary for codebook

|  |
| --- |
| #make a codebook for iris  makeCodebook(iris) |

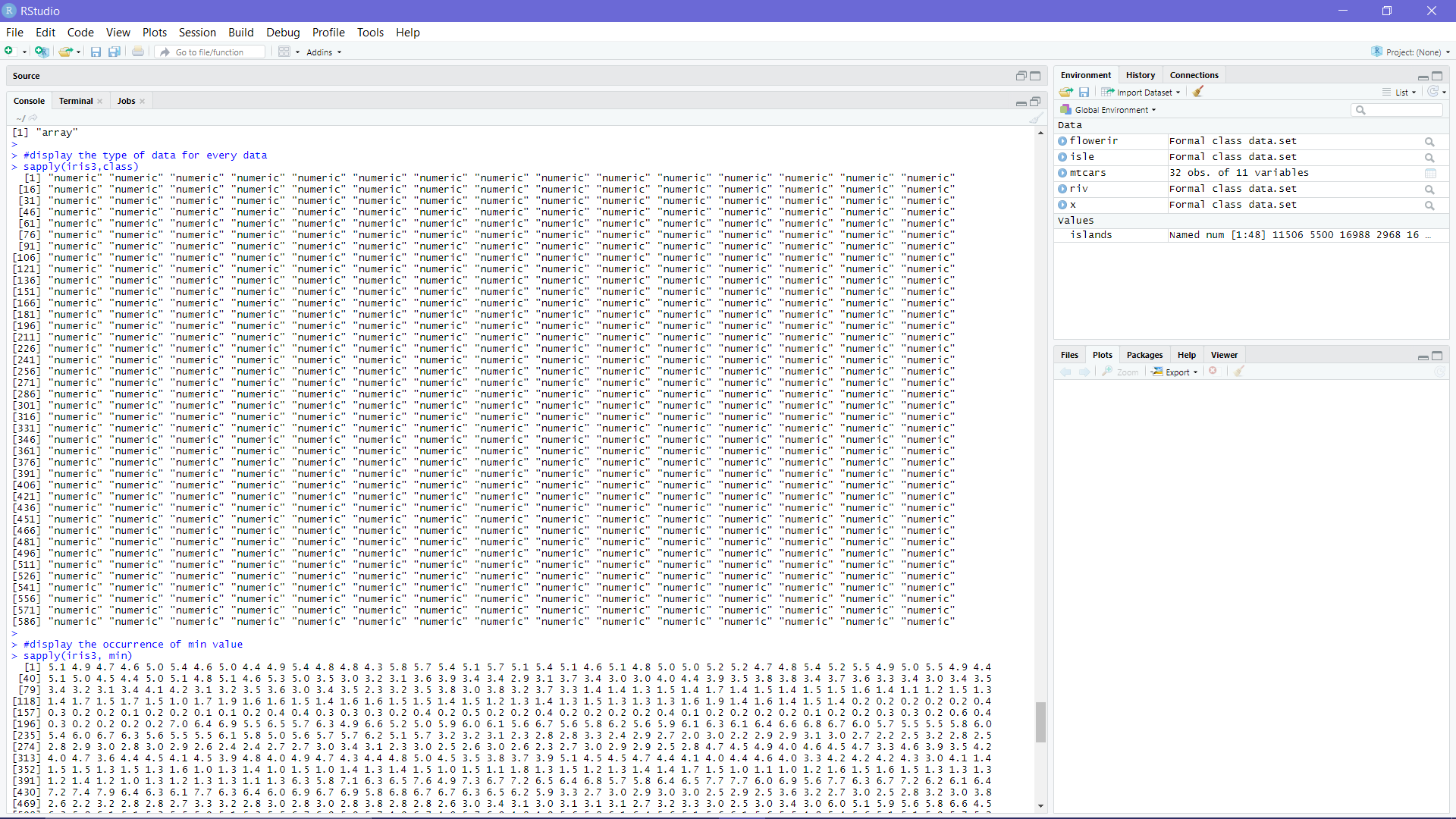
RStudio will make the codebook for the data set declared. In this case, it will make the codebook base on data set iris. The data set will be displayed in a word doc format.

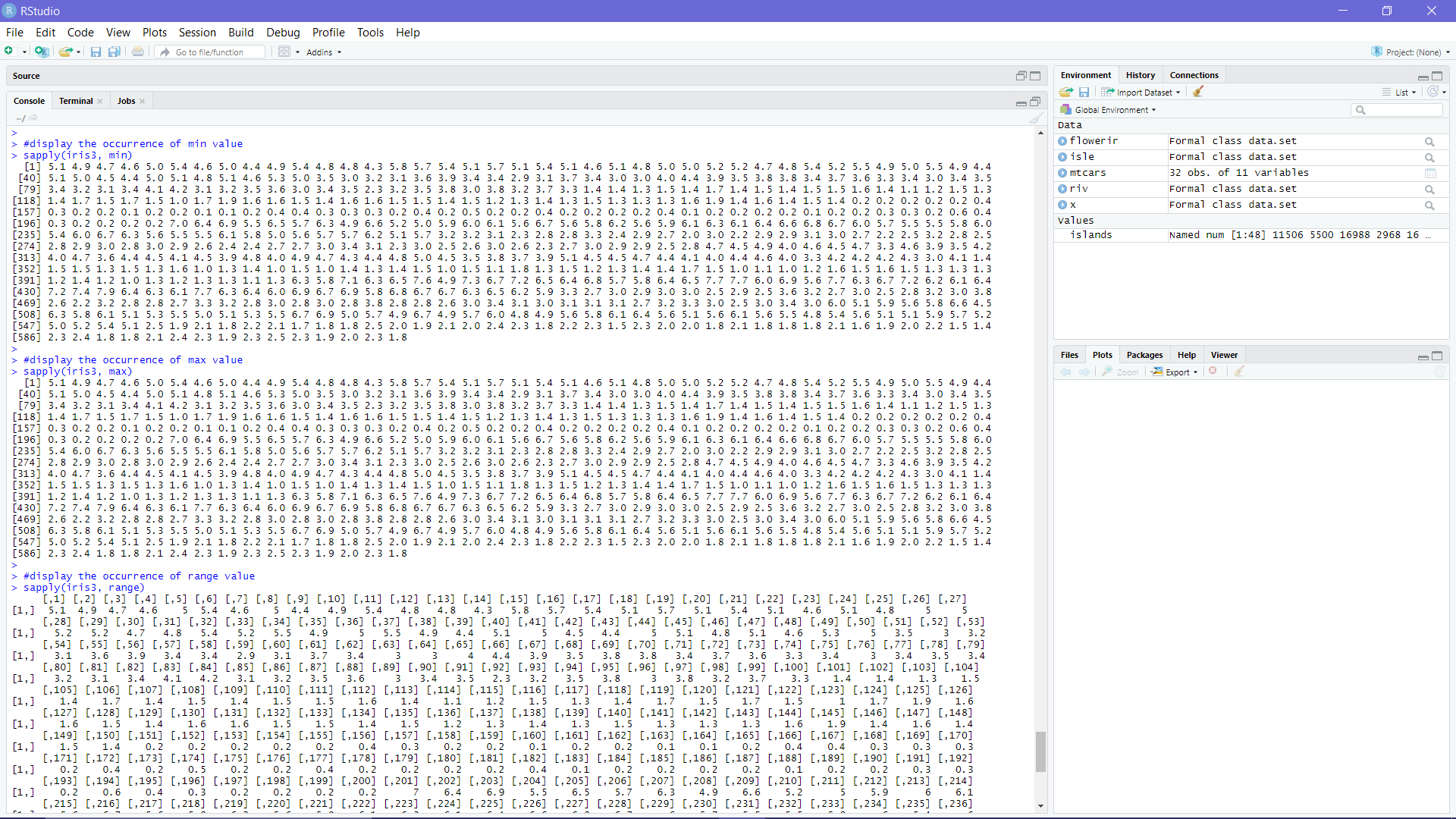
**3) Manually create a custom codebook.**

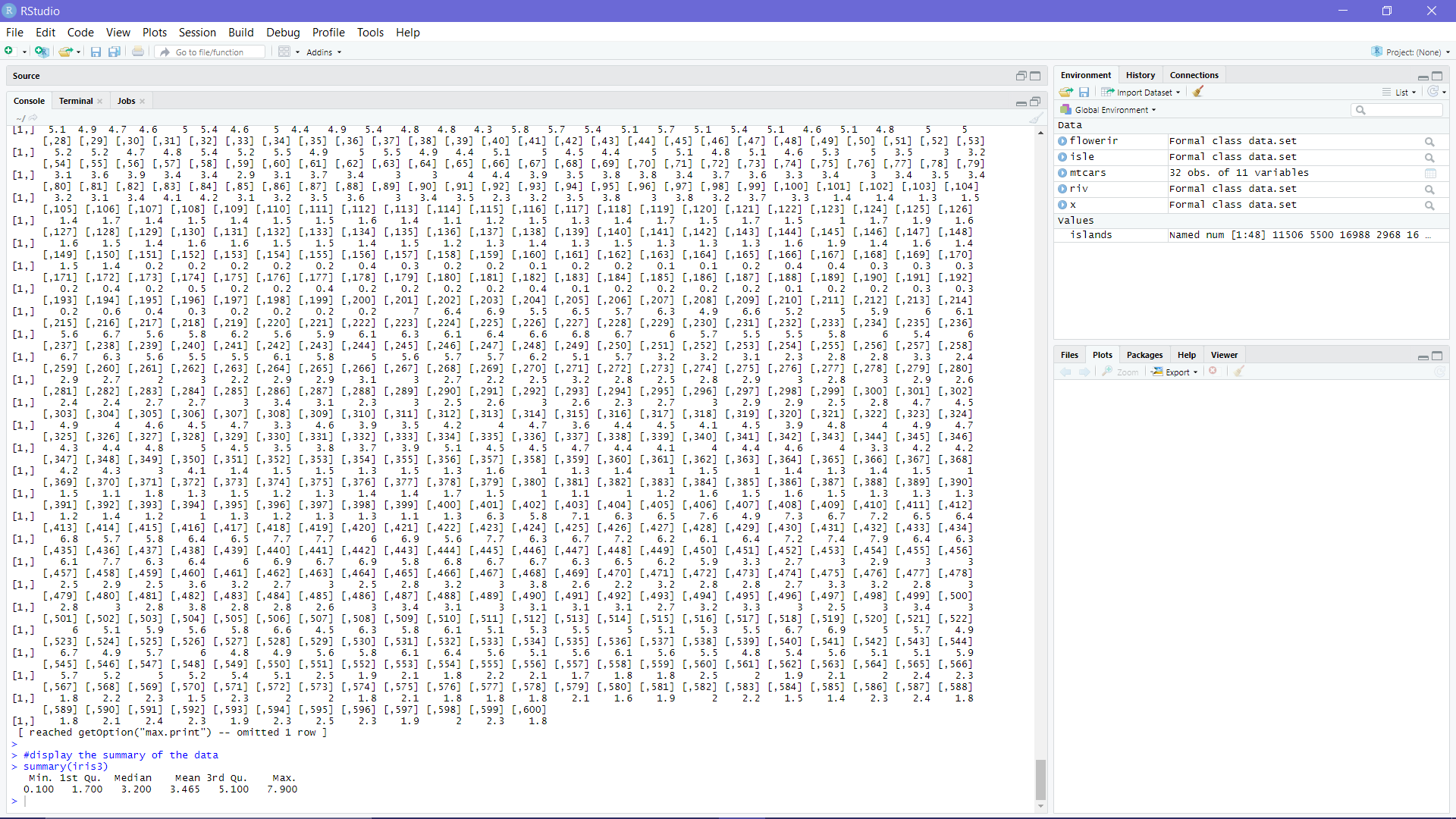
Below is the code used in RStudio to generate the Codebook

|  |
| --- |
| #display format of the data provided in data set  class(iris3)  #display the type of data for every data  sapply(iris3,class)  #display the occurrence of min value  sapply(iris3, min)  #display the occurrence of max value  sapply(iris3, max)  #display the occurrence of range value  sapply(iris3, range)  #display the summary of the data  summary(iris3) |

Below is the result displayed by RStudio regarding the data set “iris”







Description of Codebook:

* While displaying class or format

|  |
| --- |
| #display format of the data provided in data set  class(iris3) |

RStudio will display the format of the data set which is in array.

* while applying of data into specific fields

|  |
| --- |
| #display the type of data for every data  sapply(iris3,class)  #display the occurrence of min value  sapply(iris3, min)  #display the occurrence of max value  sapply(iris3, max)  #display the occurrence of range value  sapply(iris3, range) |

RStudio will display the occurrence of class, minimum, maximum and range of data. The reason why the data is in abundant is the redundancy of a single number as a minimum number or maximum number etc.

* while summary for codebook

|  |
| --- |
| #display the summary of the data  summary(iris3) |

RStudio will make the codebook for the data set declared. In this case, it will make the codebook base on data set iris. The data set will be displayed in a word doc format.