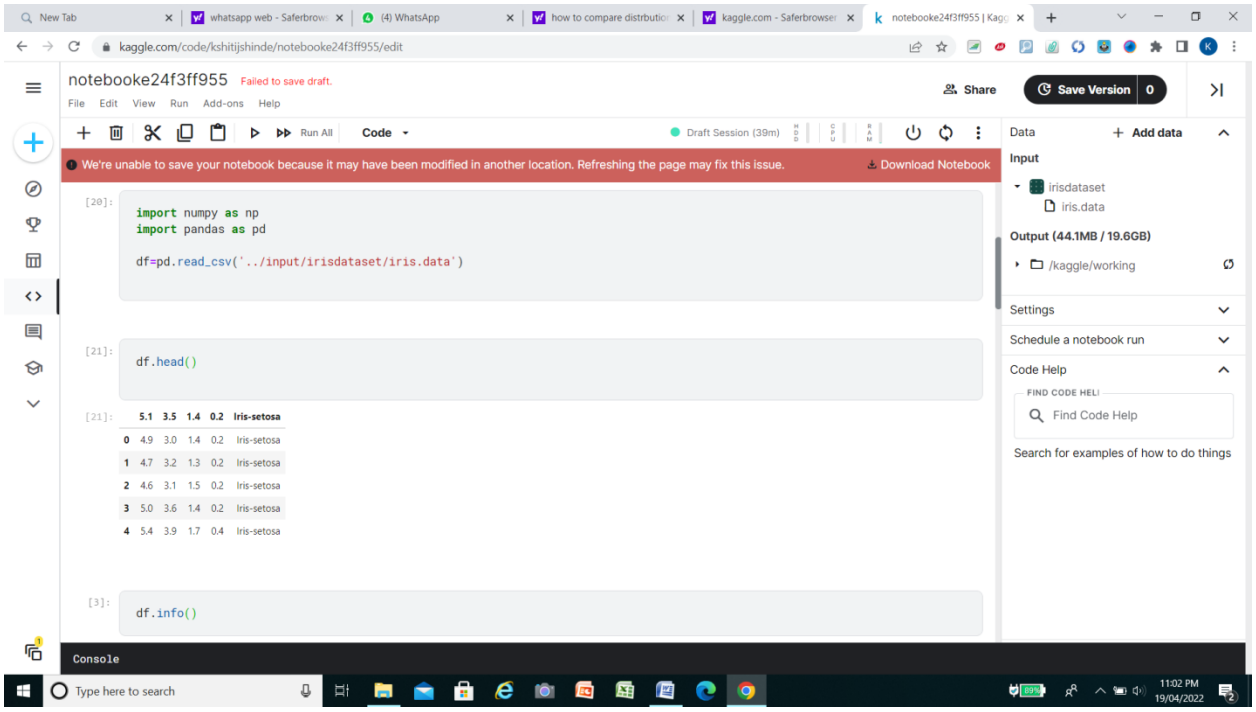
Group A
Assignment No. 10

Aim : Data Visualization III

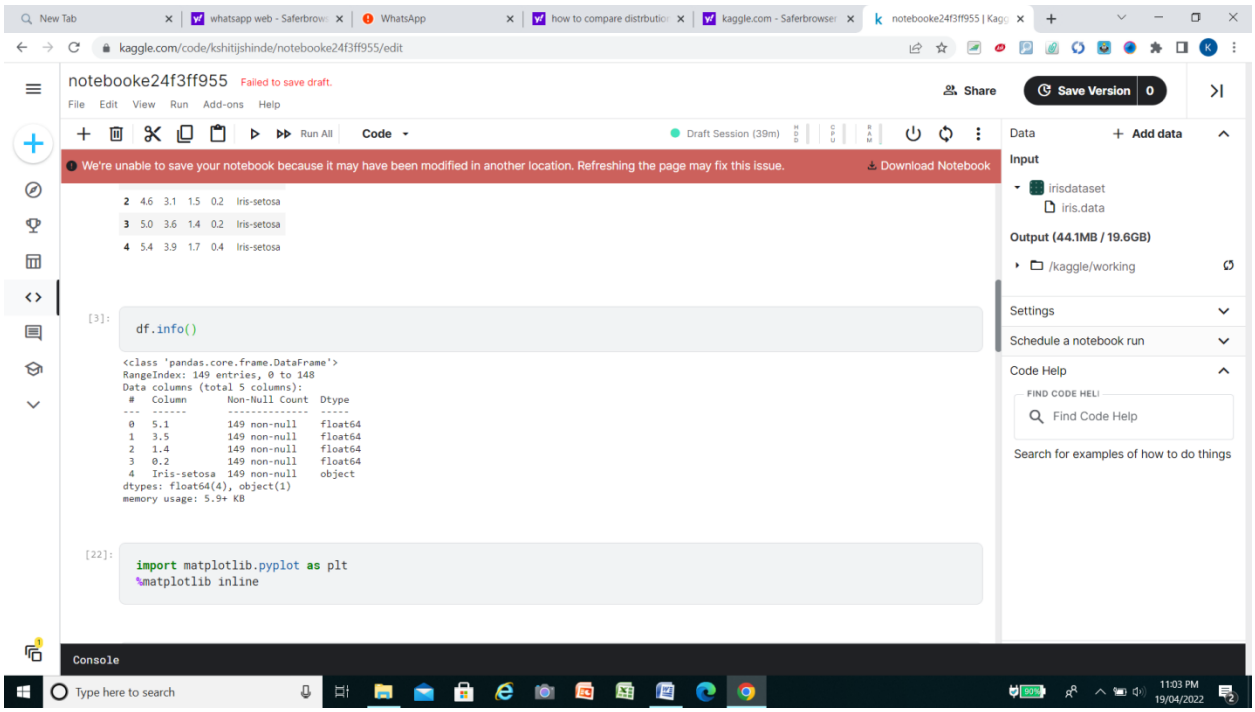
Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the inference as:
1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a box plot for each feature in the dataset.
4. Compare distributions and identify outliers.

Sol^{n} :

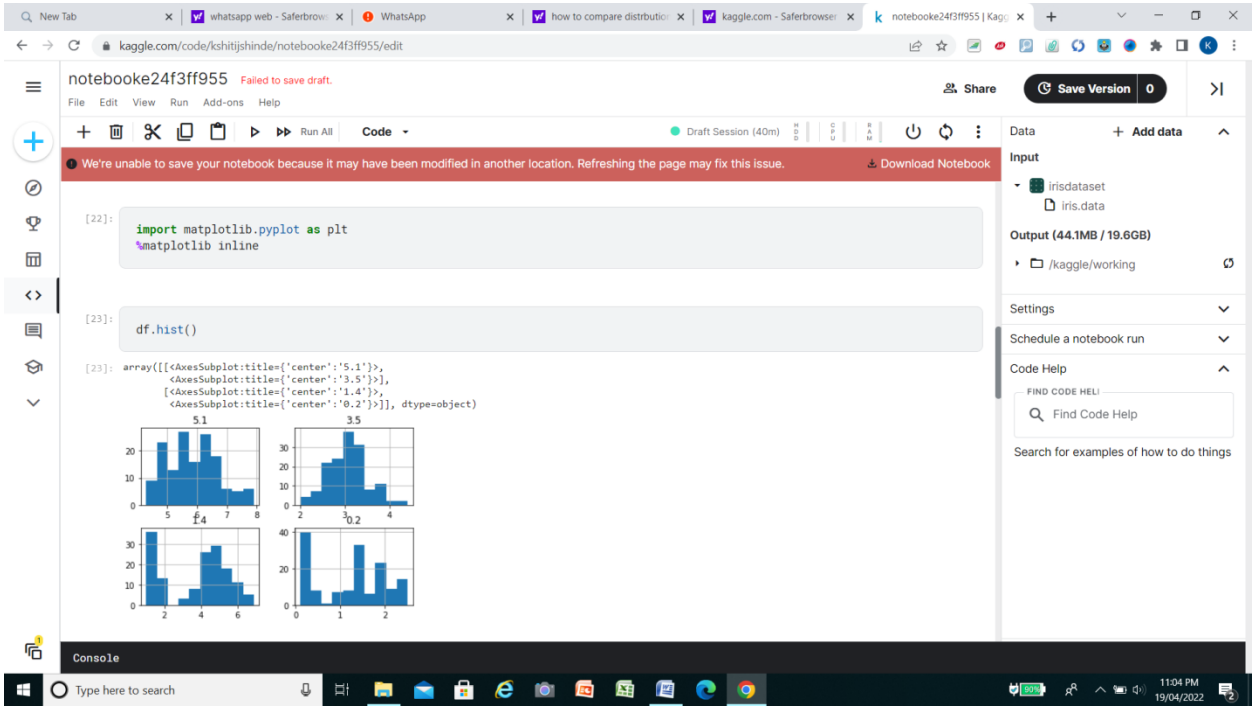
OUTPUT :



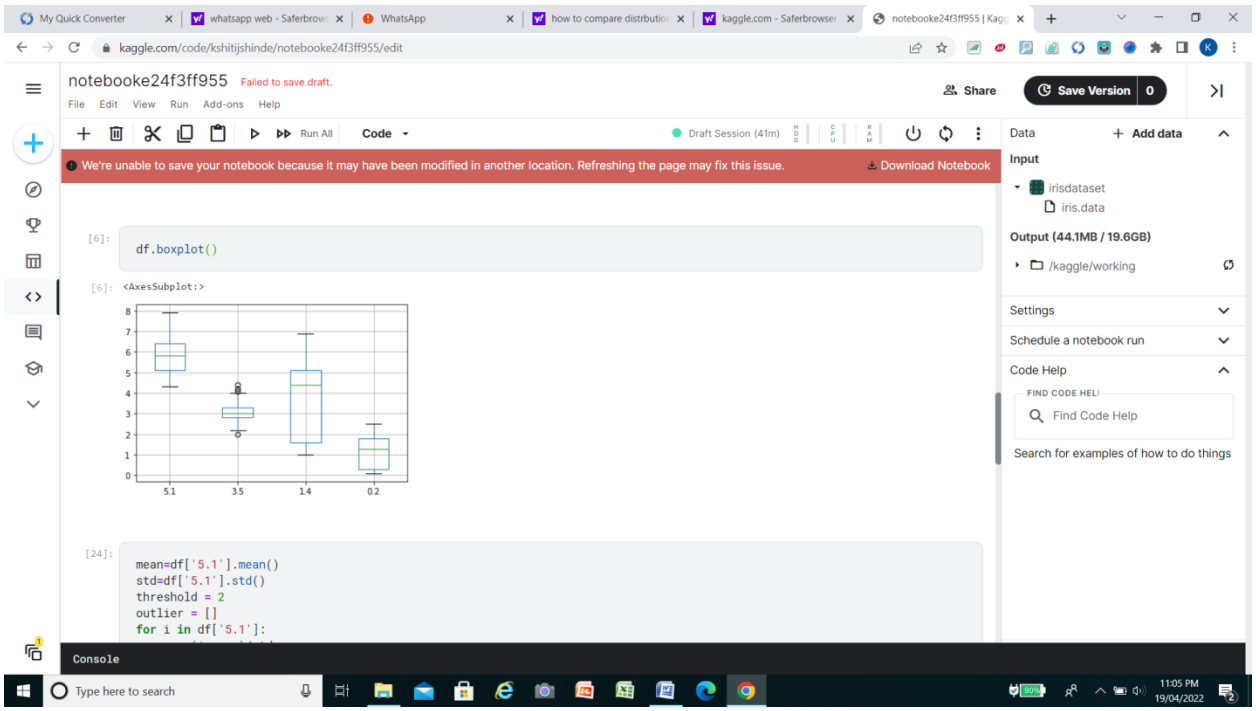
1. List down the features and their types (e.g., numeric, nominal) available in the dataset.



2. Create a histogram for each feature in the dataset to illustrate the feature distributions.

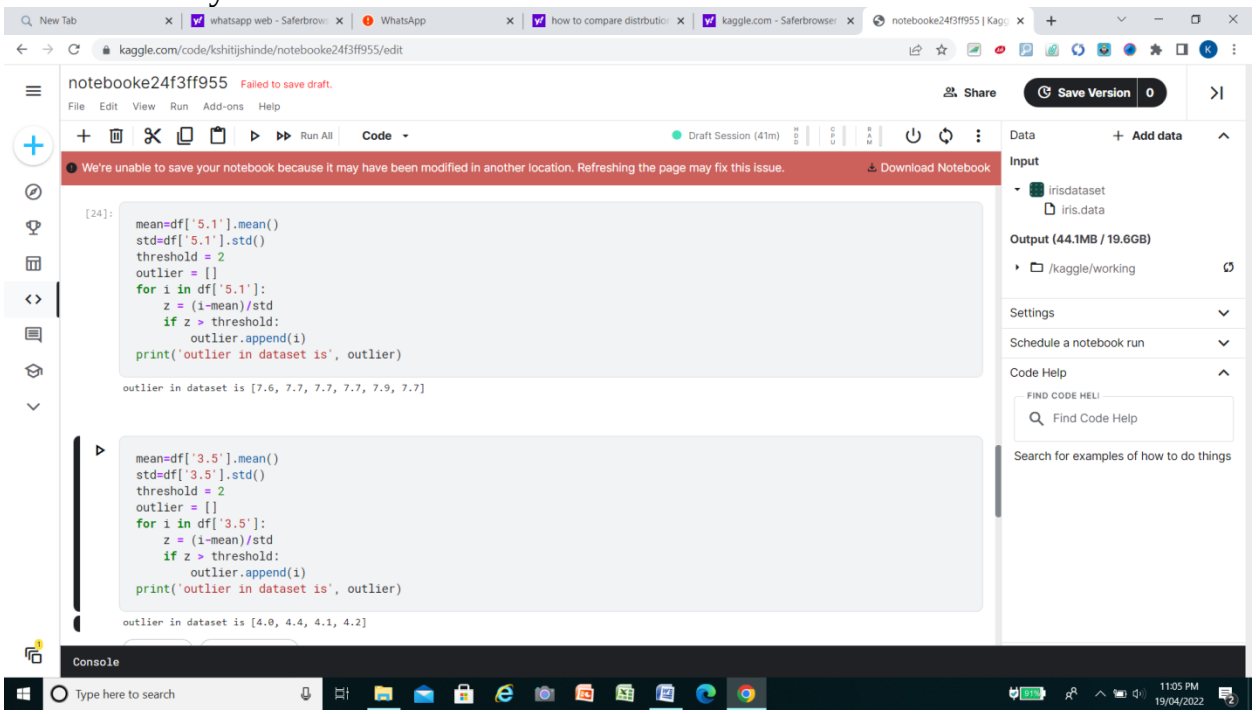


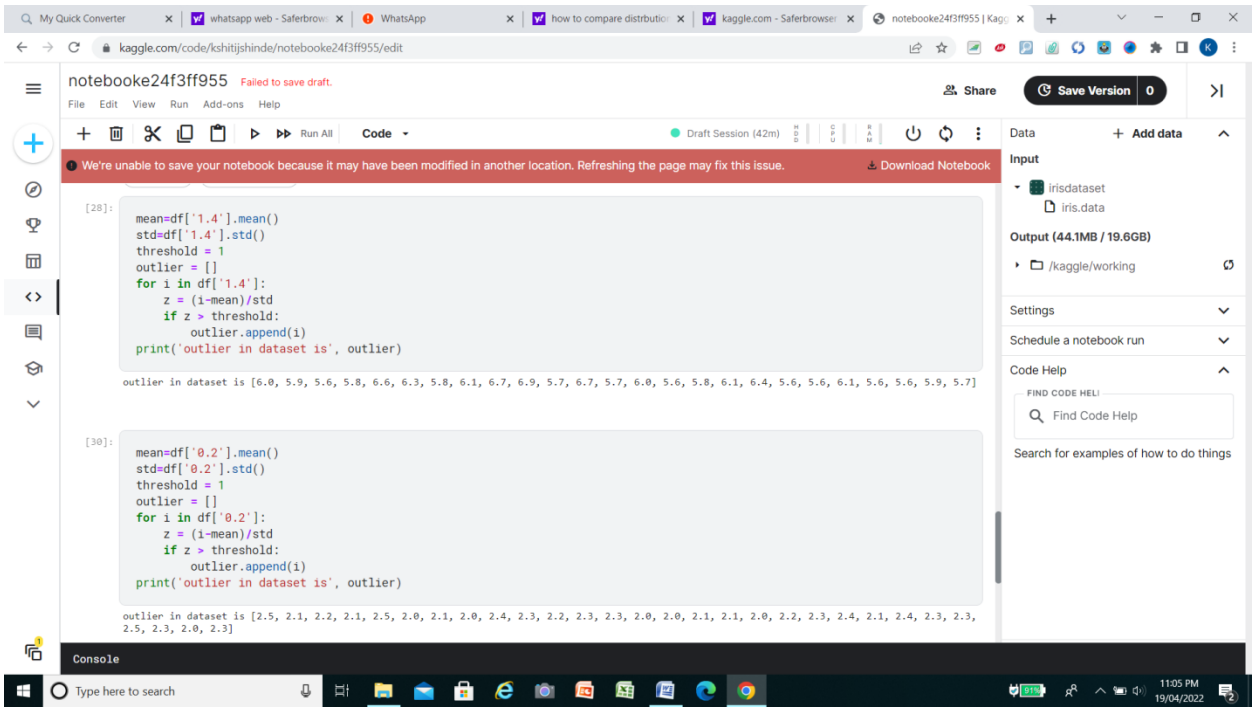
3. Create a box plot for each feature in the dataset.



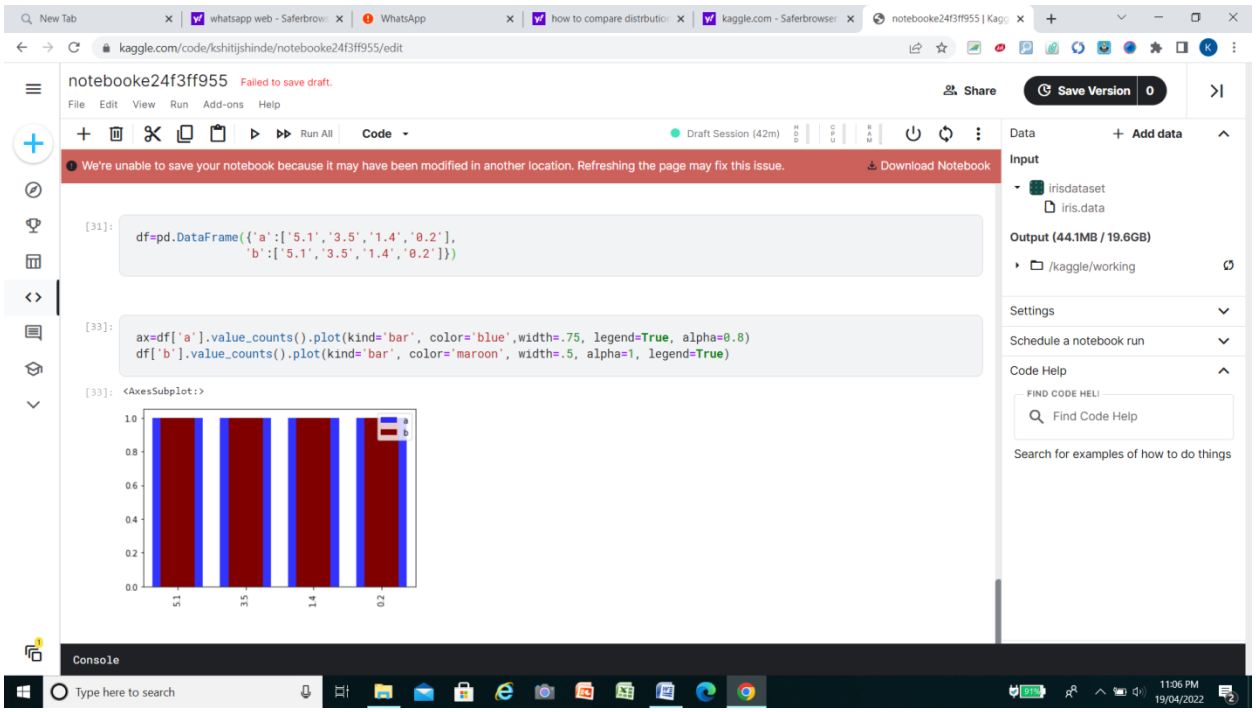
4. Compare distributions and identify outliers.

Identify outliers :



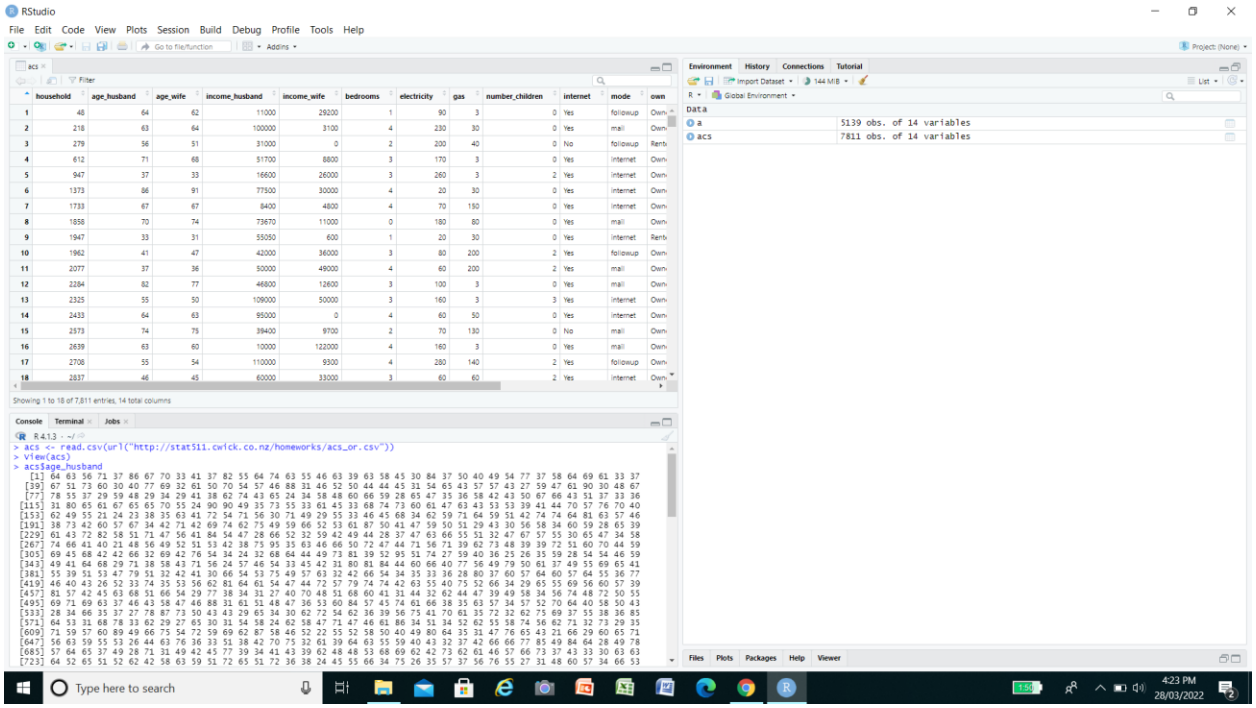


Comparing distributions :

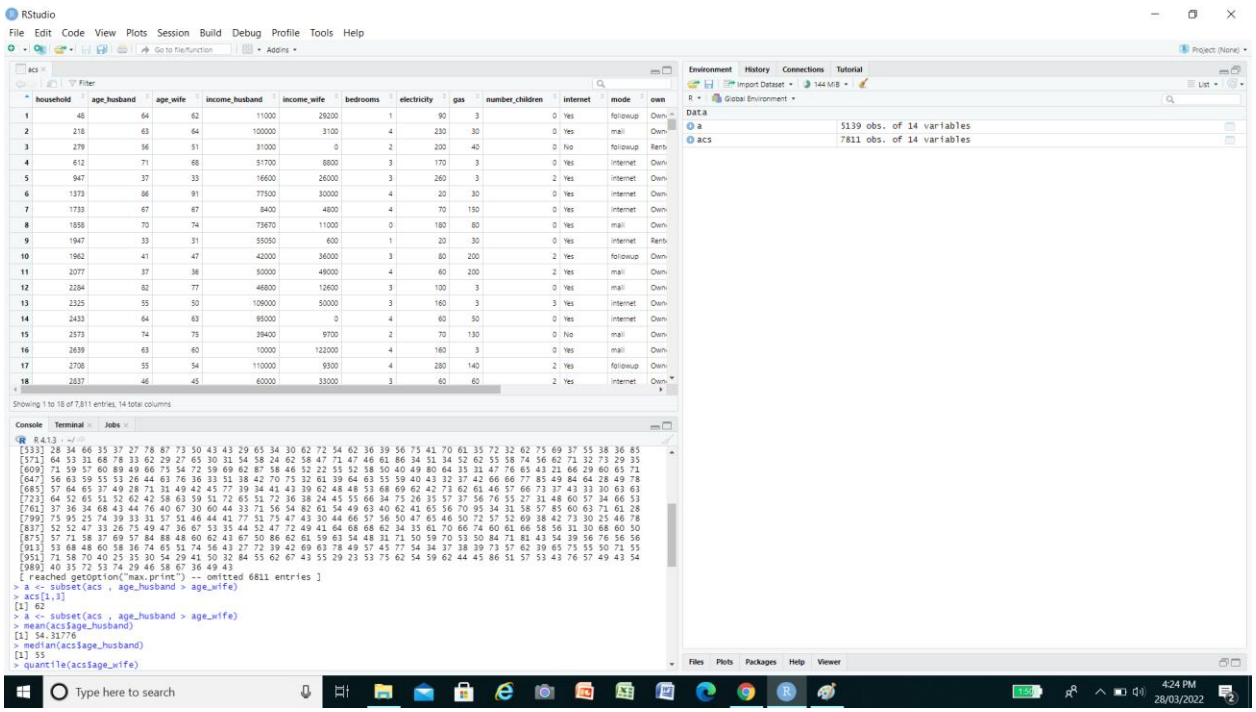


OUTPUT IN RSTUDIO :

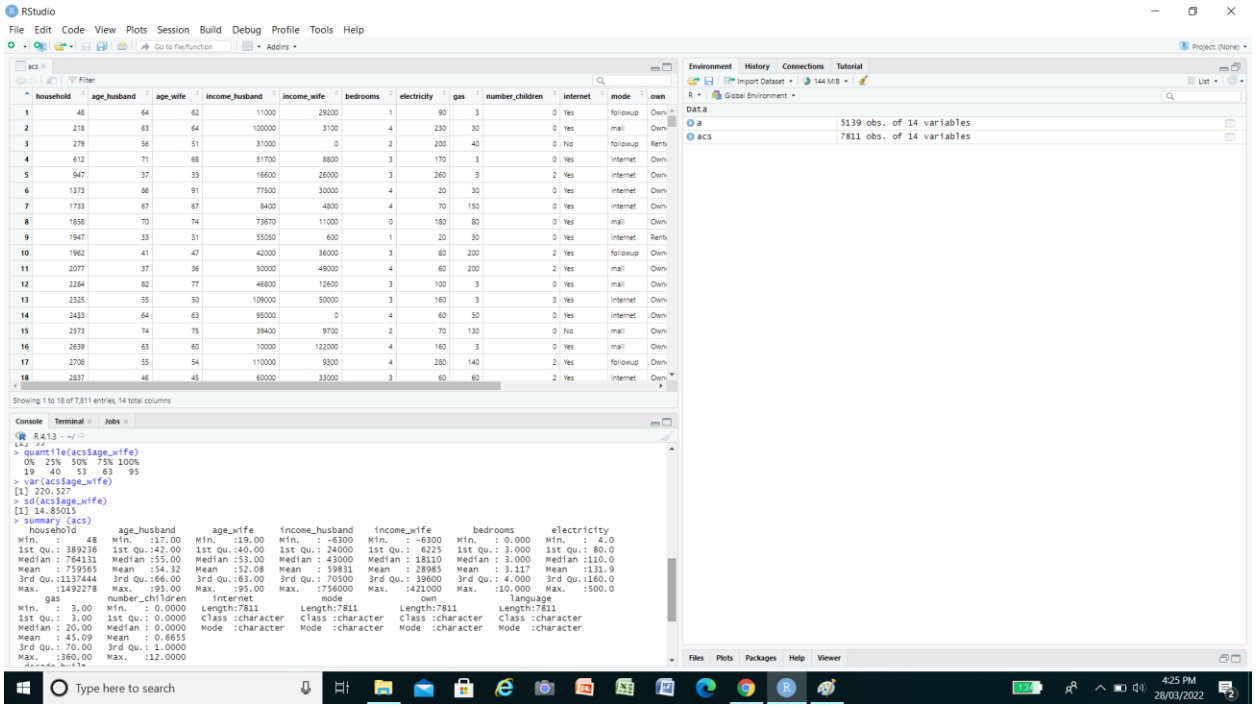
1. Importing dataset :



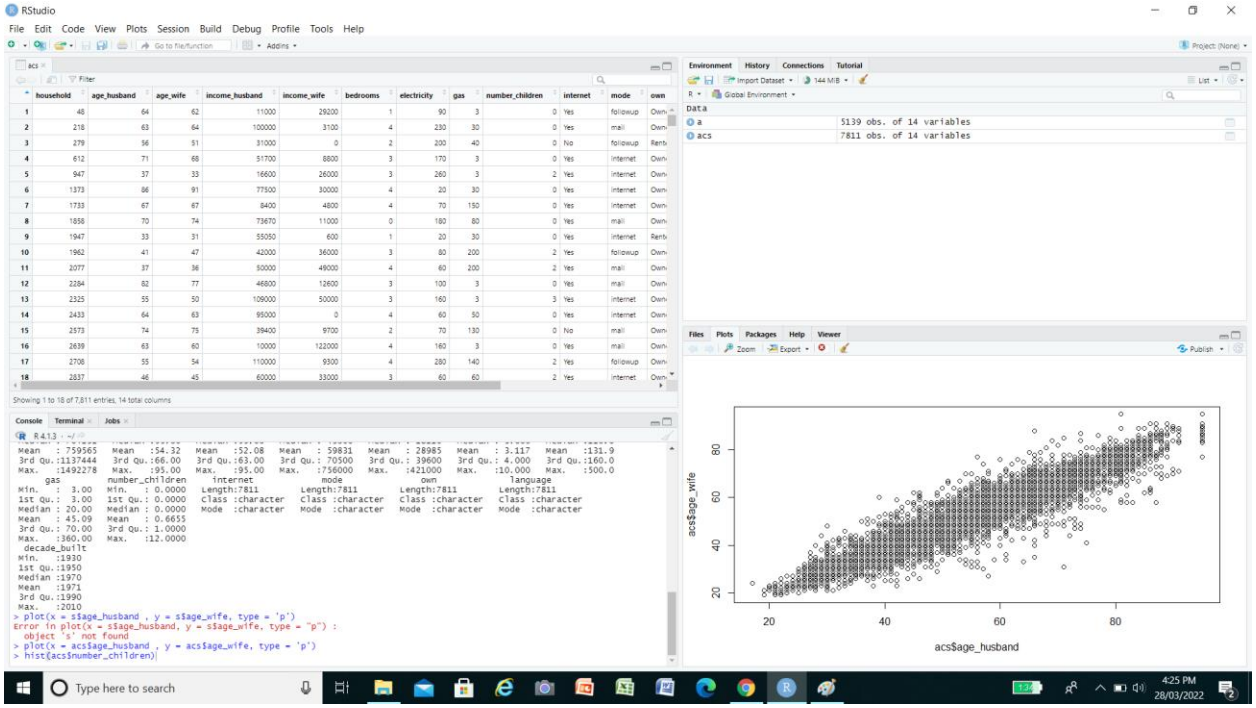
2. Transforming Data :



3. Getting Statistical Averages from data :



4. Plotting Data :



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

	household	age_husband	age_wife	income_husband	income_wife	bedrooms	electricity	gas	number_children	internet	mode	own
1	48	64	62	11000	29200	1	90	3	0	Yes	followup	Own
2	218	63	64	100000	3100	4	230	30	0	Yes	mail	Own
3	279	56	51	31000	0	2	200	40	0	No	followup	Rent
4	612	71	68	51700	8800	3	170	3	0	Yes	internet	Own
5	947	37	33	16600	26000	3	260	3	2	Yes	internet	Own
6	1373	86	91	77500	30000	4	20	30	0	Yes	internet	Own
7	1733	67	67	8400	4800	4	70	150	0	Yes	internet	Own
8	1858	70	74	73670	11000	0	180	80	0	Yes	mail	Own
9	1947	33	31	55050	600	1	20	30	0	Yes	internet	Rent
10	1962	41	47	42000	36000	3	80	200	2	Yes	followup	Own
11	2077	37	36	50000	49000	4	60	200	2	Yes	mail	Own
12	2284	82	77	46800	12600	3	100	3	0	Yes	mail	Own
13	2325	55	50	109000	50000	3	160	3	3	Yes	internet	Own
14	2433	64	63	95000	0	4	60	50	0	Yes	internet	Own
15	2573	74	75	39400	9700	2	70	130	0	No	mail	Own
16	2639	63	60	10000	122000	4	160	3	0	Yes	mail	Own
17	2708	55	54	110000	9300	4	280	140	2	Yes	followup	Own
18	2837	46	45	60000	33000	3	60	60	2	Yes	internet	Own

Showing 1 to 18 of 7,811 entries, 14 total columns

```
3rd Qu.:1137444   3rd Qu.:66.00   3rd Qu.:63.00   3rd Qu.: 70500   3rd Qu.: 39600   3rd Qu.: 4.000   3rd Qu.:160.0
Max.   :1492278   Max.   :95.00   Max.   :95.00   Max.   :756000   Max.   :421000   Max.   :10.000   Max.   :500.0
      gas          number_children     internet            mode              own             language
 Min.   :  3.00   Min.   : 0.0000   Length:7811        Length:7811        Length:7811        Length:7811
 1st Qu.:  3.00   1st Qu.: 0.0000   Class :character   Class :character   Class :character   Class :character
 Median : 20.00   Median : 0.0000   Mode  :character   Mode  :character   Mode  :character   Mode  :character
 Mean   : 45.09   Mean   : 0.6655
 3rd Qu.: 70.00   3rd Qu.: 1.0000
 Max.   :360.00   Max.   :12.0000
  decade_built
 Min.   :1930
 1st Qu.:1950
 Median :1970
 Mean   :1971
 3rd Qu.:1990
 Max.   :2010
> plot(x = s$age_husband , y = s$age_wife, type = 'p')
Error in plot(x = s$age_husband, y = s$age_wife, type = "p") :
  object 's' not found
> plot(x = acs$age_husband , y = acs$age_wife, type = 'p')
> hist(acs$number_children)
>
```

Environment History Connections Tutorial

Import Dataset 144 MiB

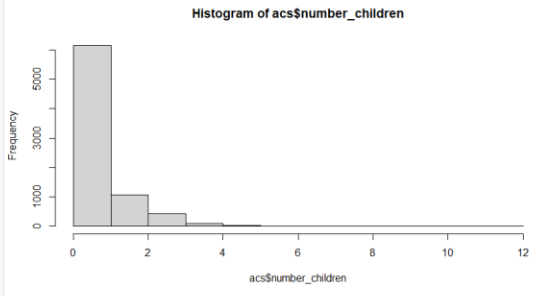
Global Environment

Data

a	5139 obs. of 14 variables
acs	7811 obs. of 14 variables

Files Plots Packages Help Viewer

Zoom Export Publish



Type here to search

4:25 PM
28/03/2022

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

	household	age_husband	age_wife	income_husband	income_wife	bedrooms	electricity	gas	number_children	internet	mode	own
1	48	64	62	11000	29200	1	90	3	0	Yes	followup	Own
2	218	63	64	100000	3100	4	230	30	0	Yes	mail	Own
3	279	56	51	31000	0	2	200	40	0	No	followup	Rent
4	612	71	68	51700	8800	3	170	3	0	Yes	internet	Own
5	947	37	33	16600	26000	3	260	3	2	Yes	internet	Own
6	1373	86	91	77500	30000	4	20	30	0	Yes	internet	Own
7	1733	67	67	8400	4800	4	70	150	0	Yes	internet	Own
8	1858	70	74	73670	11000	0	180	80	0	Yes	mail	Own
9	1947	33	31	55050	600	1	20	30	0	Yes	internet	Rent
10	1962	41	47	42000	36000	3	80	200	2	Yes	followup	Own
11	2077	37	36	50000	49000	4	60	200	2	Yes	mail	Own
12	2284	82	77	46800	12600	3	100	3	0	Yes	mail	Own
13	2325	55	50	109000	50000	3	160	3	3	Yes	internet	Own
14	2433	64	63	95000	0	4	60	50	0	Yes	internet	Own
15	2573	74	75	39400	9700	2	70	130	0	No	mail	Own
16	2639	63	60	10000	122000	4	160	3	0	Yes	mail	Own
17	2708	55	54	110000	9300	4	280	140	2	Yes	followup	Own
18	2837	46	45	60000	33000	3	60	60	2	Yes	internet	Own

Showing 1 to 18 of 7,811 entries, 14 total columns

```
      gas          number_children     internet            mode              own             language
 Min.   :  3.00   Min.   : 0.0000   Length:7811        Length:7811        Length:7811        Length:7811
 1st Qu.:  3.00   1st Qu.: 0.0000   Class :character   Class :character   Class :character   Class :character
 Median : 20.00   Median : 0.0000   Mode  :character   Mode  :character   Mode  :character   Mode  :character
 Mean   : 45.09   Mean   : 0.6655
 3rd Qu.: 70.00   3rd Qu.: 1.0000
 Max.   :360.00   Max.   :12.0000
  decade_built
 Min.   :1930
 1st Qu.:1950
 Median :1970
 Mean   :1971
 3rd Qu.:1990
 Max.   :2010
> plot(x = s$age_husband , y = s$age_wife, type = 'p')
Error in plot(x = s$age_husband, y = s$age_wife, type = "p") :
  object 's' not found
> plot(x = acs$age_husband , y = acs$age_wife, type = 'p')
> hist(acs$number_children)
> counts <- table(acs$bedrooms)
> barplot(counts, main="Bedrooms Distribution", xlab="Number of Bedrooms")
>
```

Environment History Connections Tutorial

Import Dataset 144 MiB

Global Environment

Data

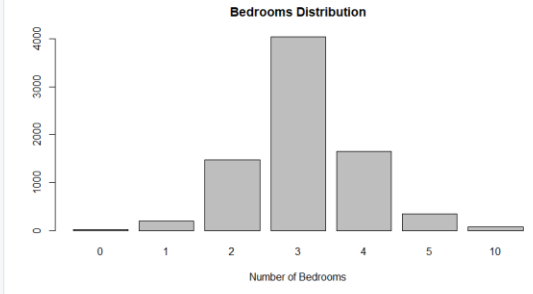
a	5139 obs. of 14 variables
acs	7811 obs. of 14 variables

Values

counts	'table' int [1:7(1d)] 24 198 1475 4040 1652 350 72

Files Plots Packages Help Viewer

Zoom Export Publish



Type here to search

4:26 PM
28/03/2022