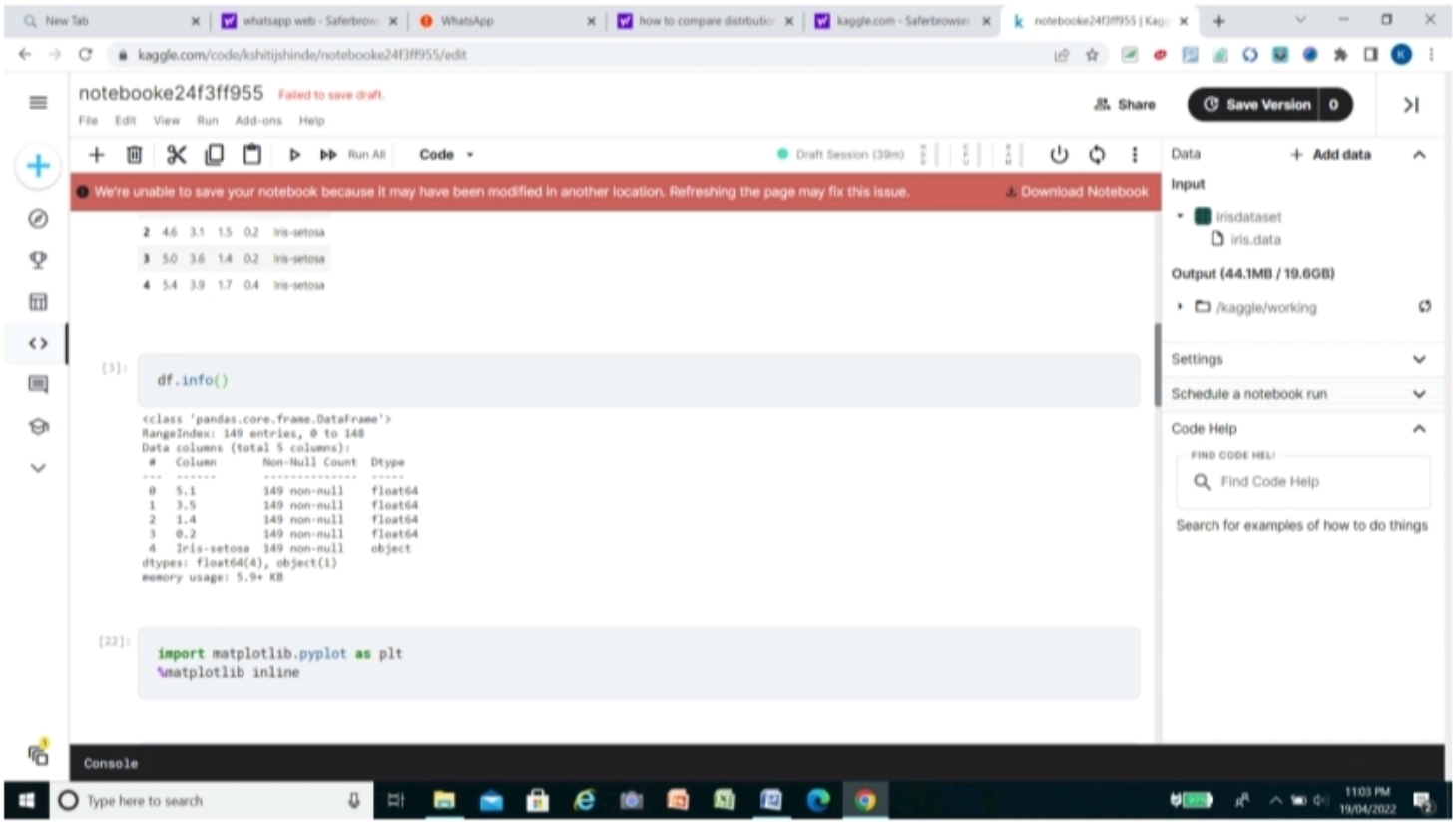
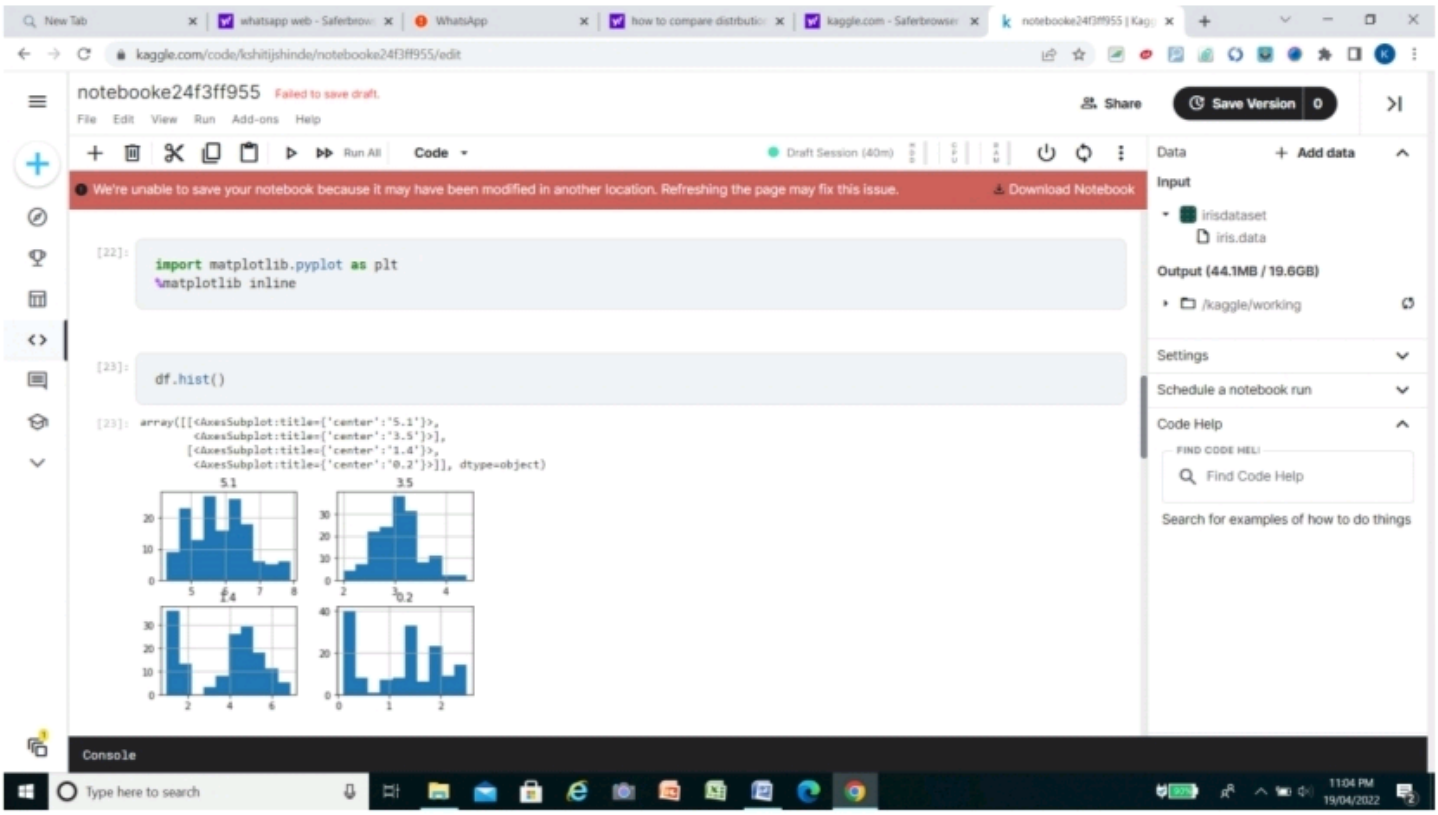
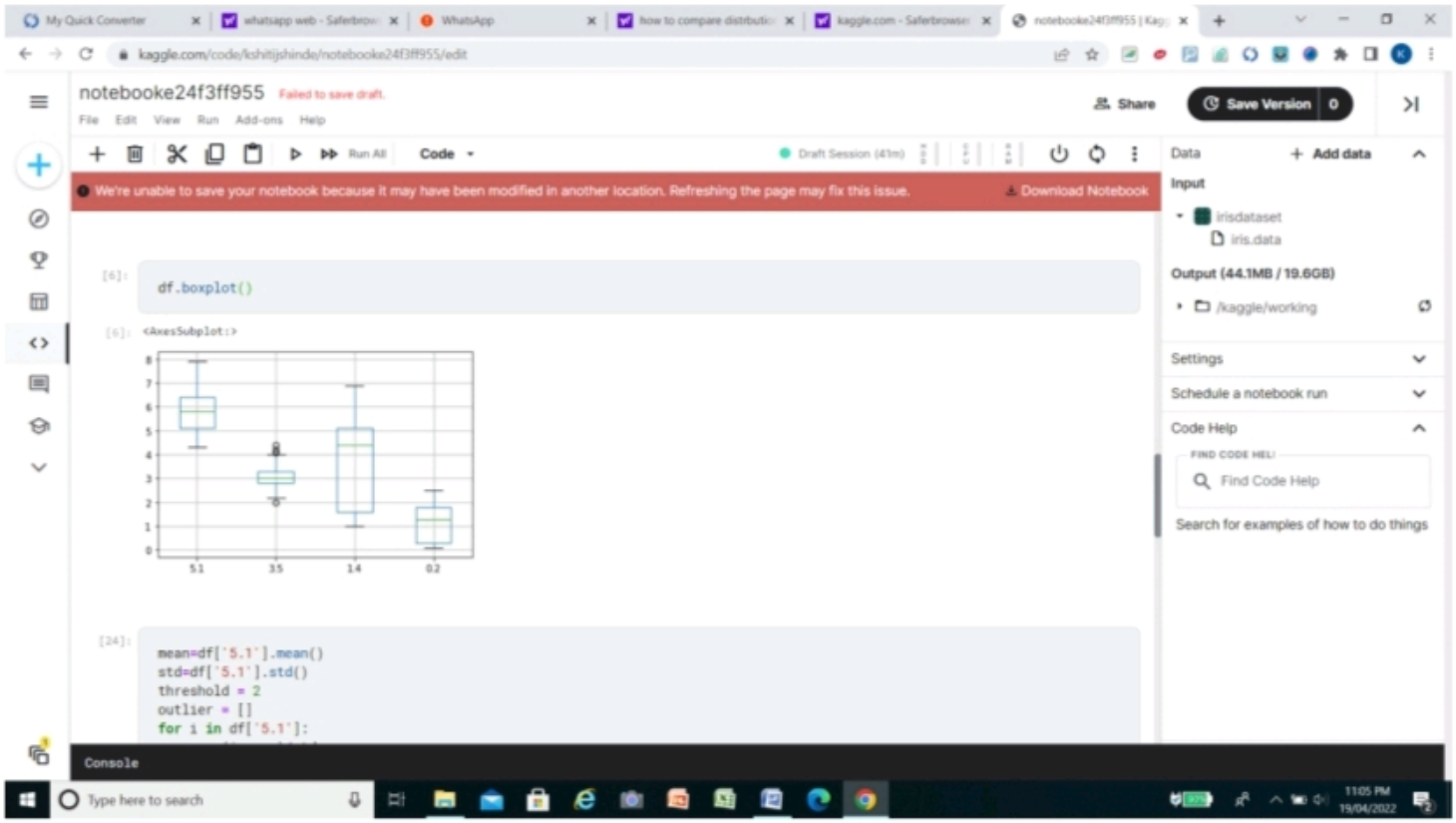
1. List down the features and their types (e.g., numeric, nominal) available in the dataset.



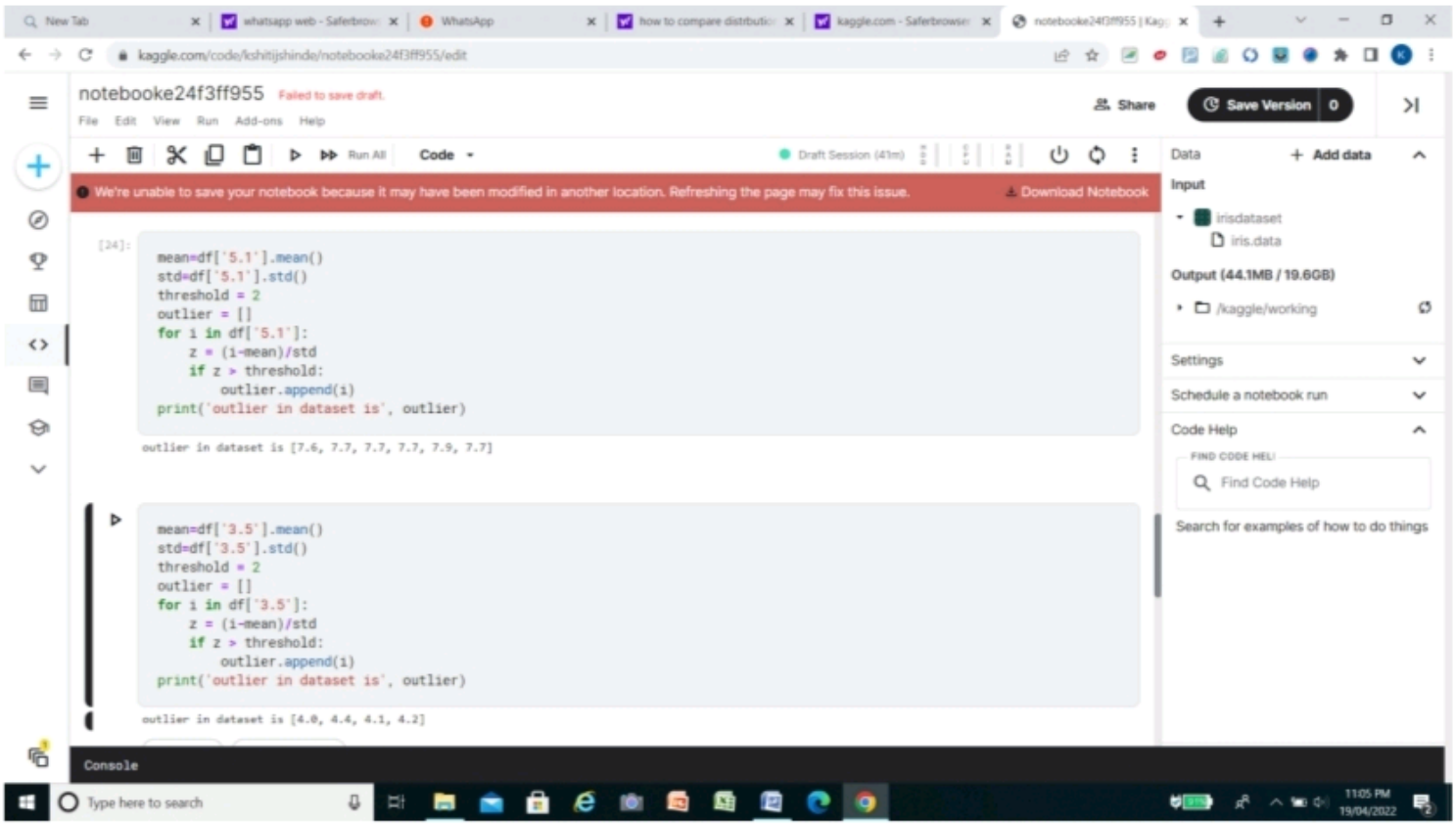
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.

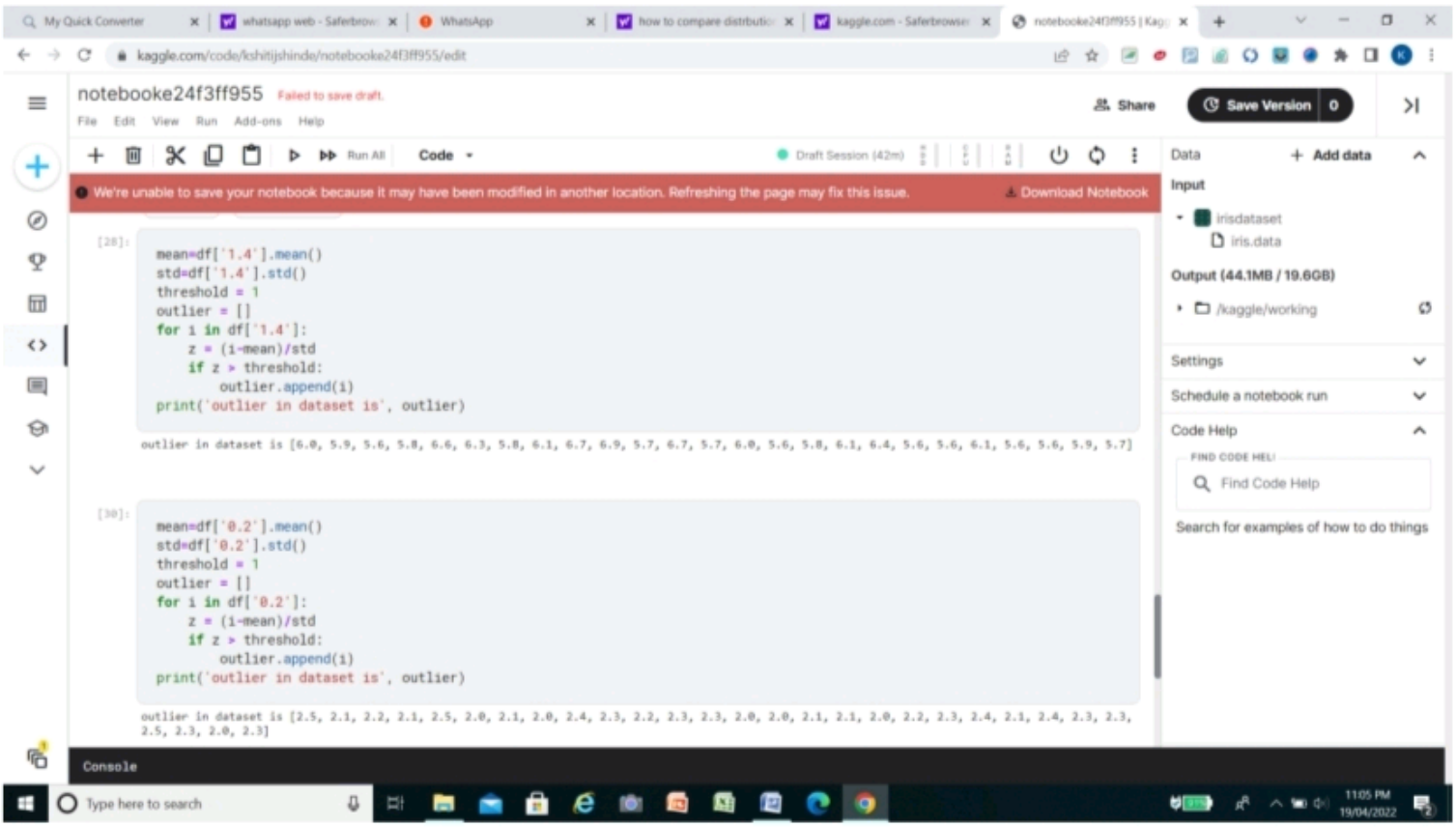


3. Create a box plot for each feature in the dataset.

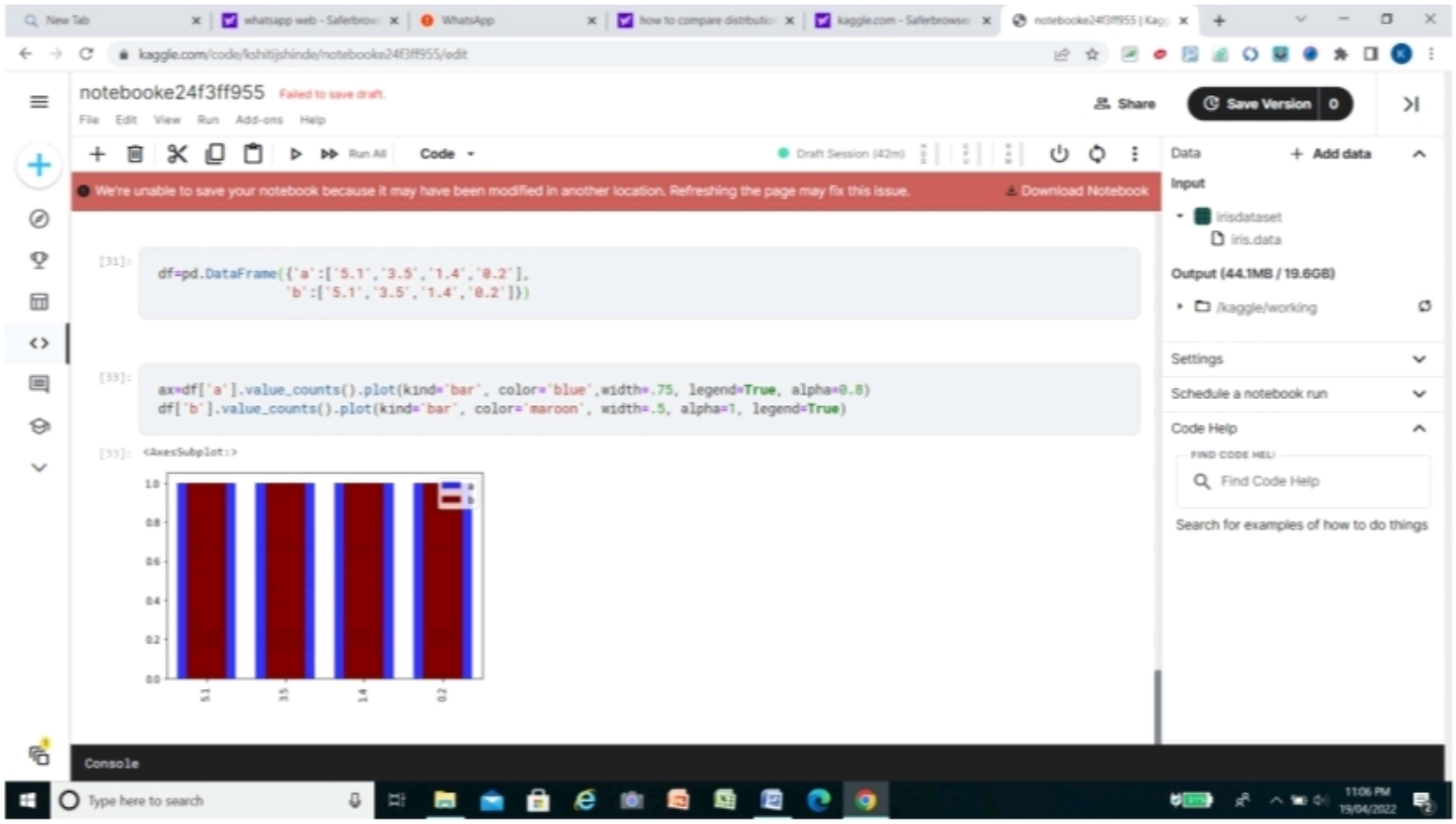


4. Compare distributions and identify outliers.
   Identify outliers :



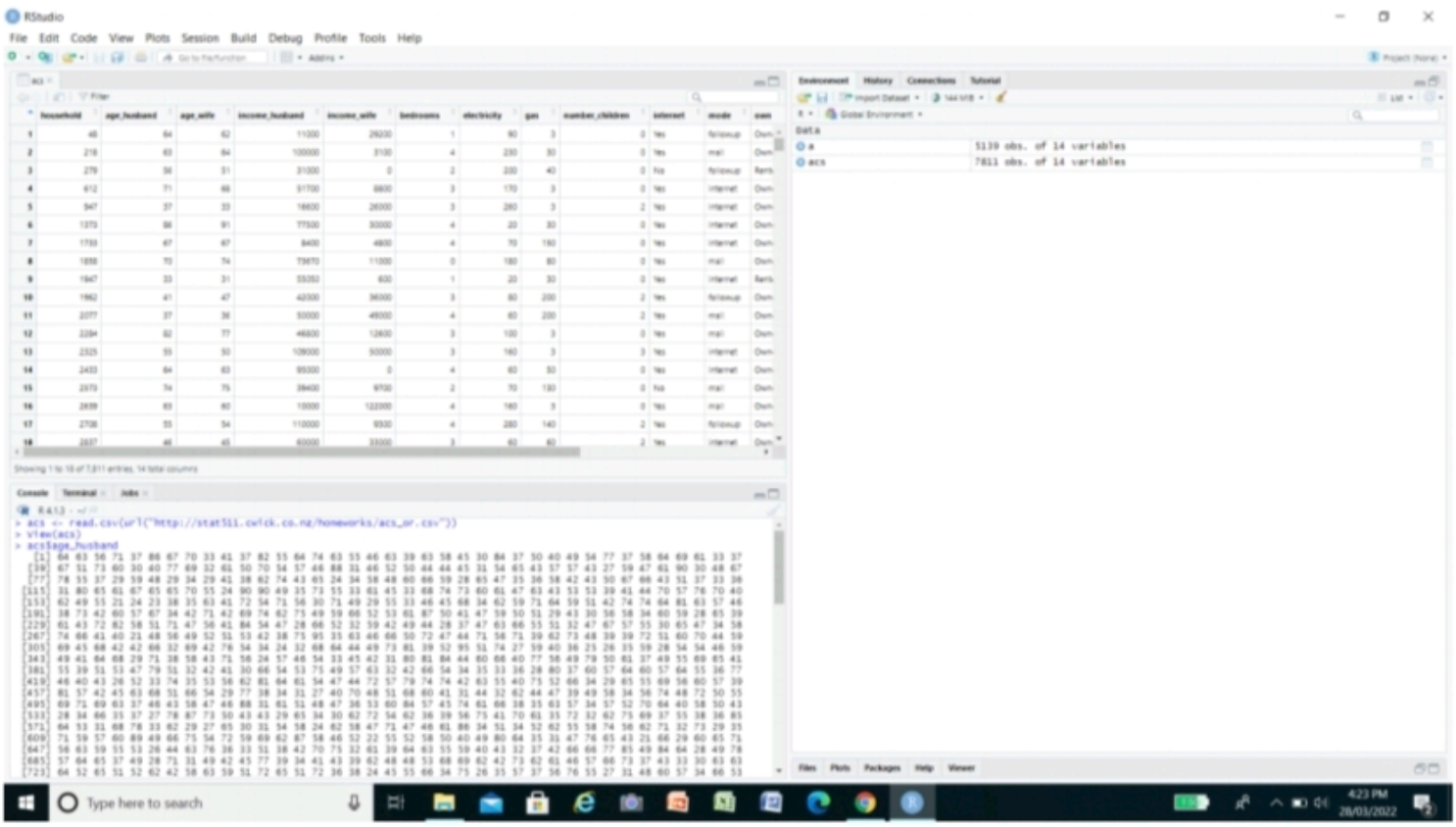

kaggle.com/code/kshitijshinde/notebooke24f3ff955/edit
notebooke24f3ff955
Failed to save draft.
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (42m)
We're unable to save your notebook because it may have been modified in another location. Refreshing the page may fix this issue.
Download Notebook
[28]:
mean=df['1.4'].mean()
std=df['1.4'].std()
threshold = 1
outlier = []
for i in df['1.4']:
z = (i-mean)/std
if z > threshold:
outlier.append(i)
print('outlier in dataset is', outlier)
outlier in dataset is [6.0, 5.9, 5.6, 5.8, 6.6, 6.3, 5.8, 6.1, 6.7, 6.9, 5.7, 6.7, 5.7, 6.0, 5.6, 5.8, 6.1, 6.4, 5.6, 5.6, 6.1, 5.6, 5.6, 5.9, 5.7]
[30]:
mean=df['0.2'].mean()
std=df['0.2'].std()
threshold = 1
outlier = []
for i in df['0.2']:
z = (i-mean)/std
if z > threshold:
outlier.append(i)
print('outlier in dataset is', outlier)
outlier in dataset is [2.5, 2.1, 2.2, 2.1, 2.5, 2.0, 2.1, 2.0, 2.4, 2.3, 2.2, 2.3, 2.3, 2.0, 2.0, 2.1, 2.1, 2.0, 2.2, 2.3, 2.4, 2.1, 2.4, 2.3, 2.3, 2.5, 2.3, 2.0, 2.3]
Console
Data
Add data
Input
irisdataset
iris.data
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
Type here to search
11:05 PM
19/04/2022

Comparing distributions :

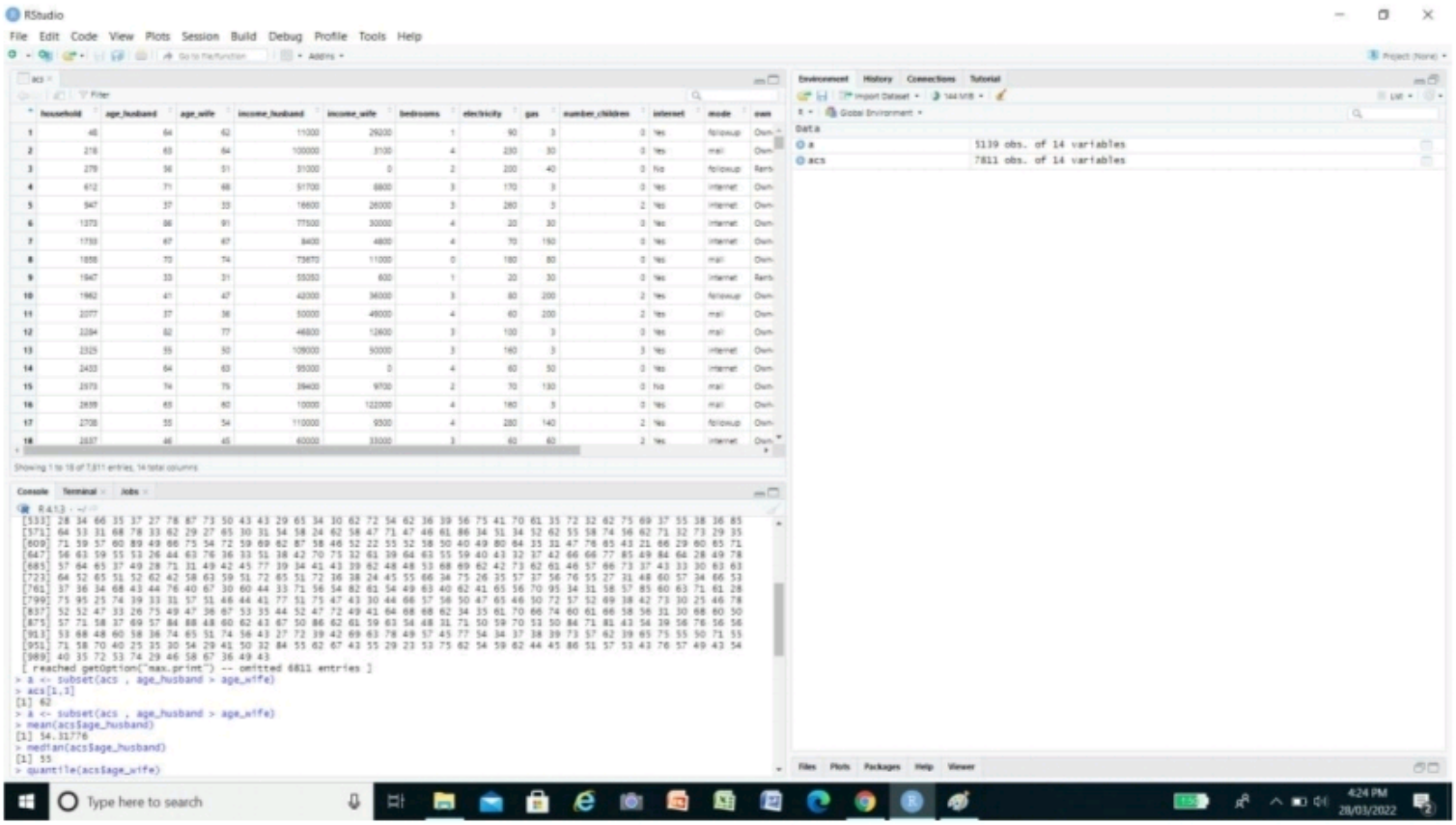


OUTPUT IN RSTUDIO :

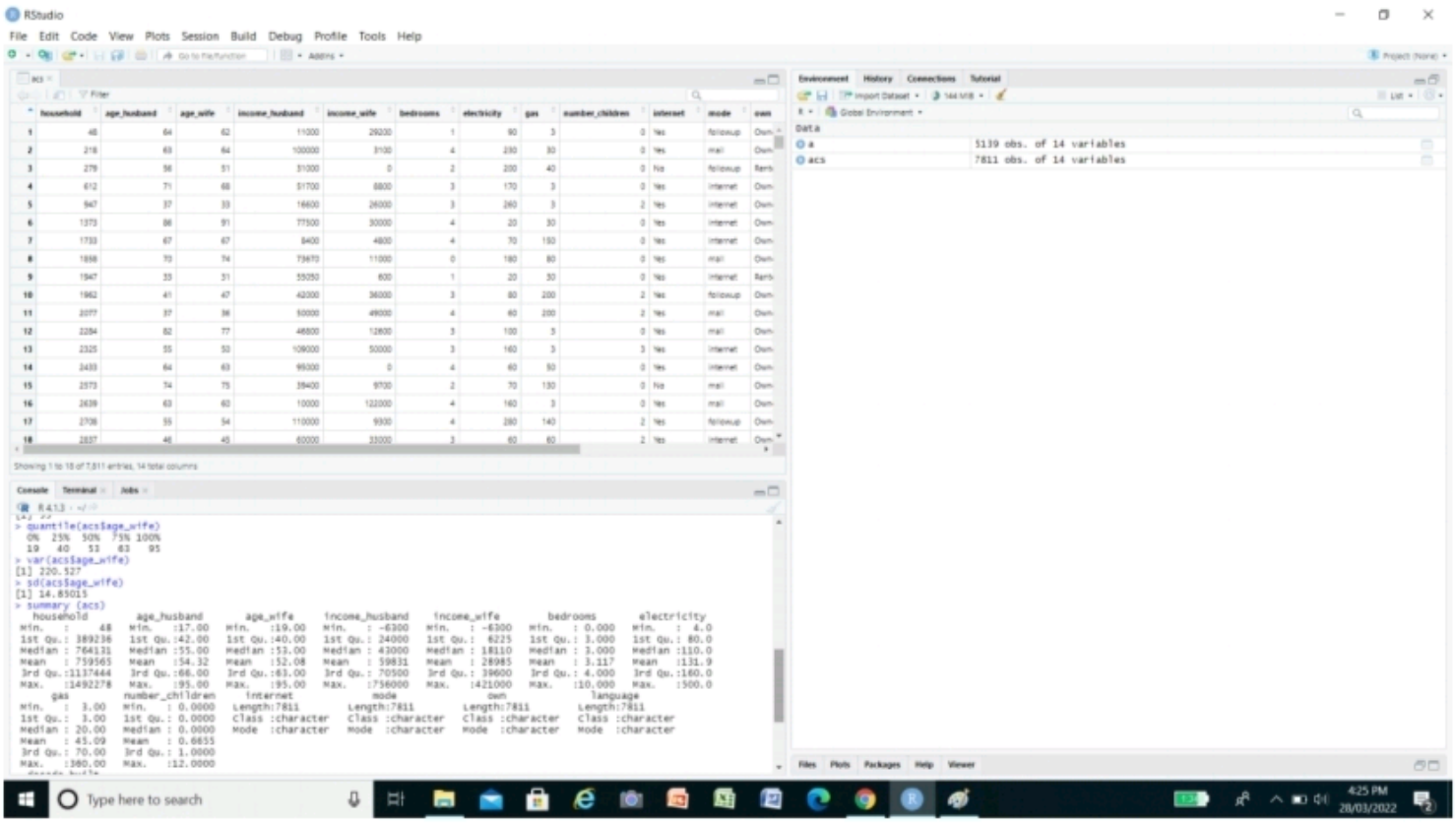
1. Importing dataset :



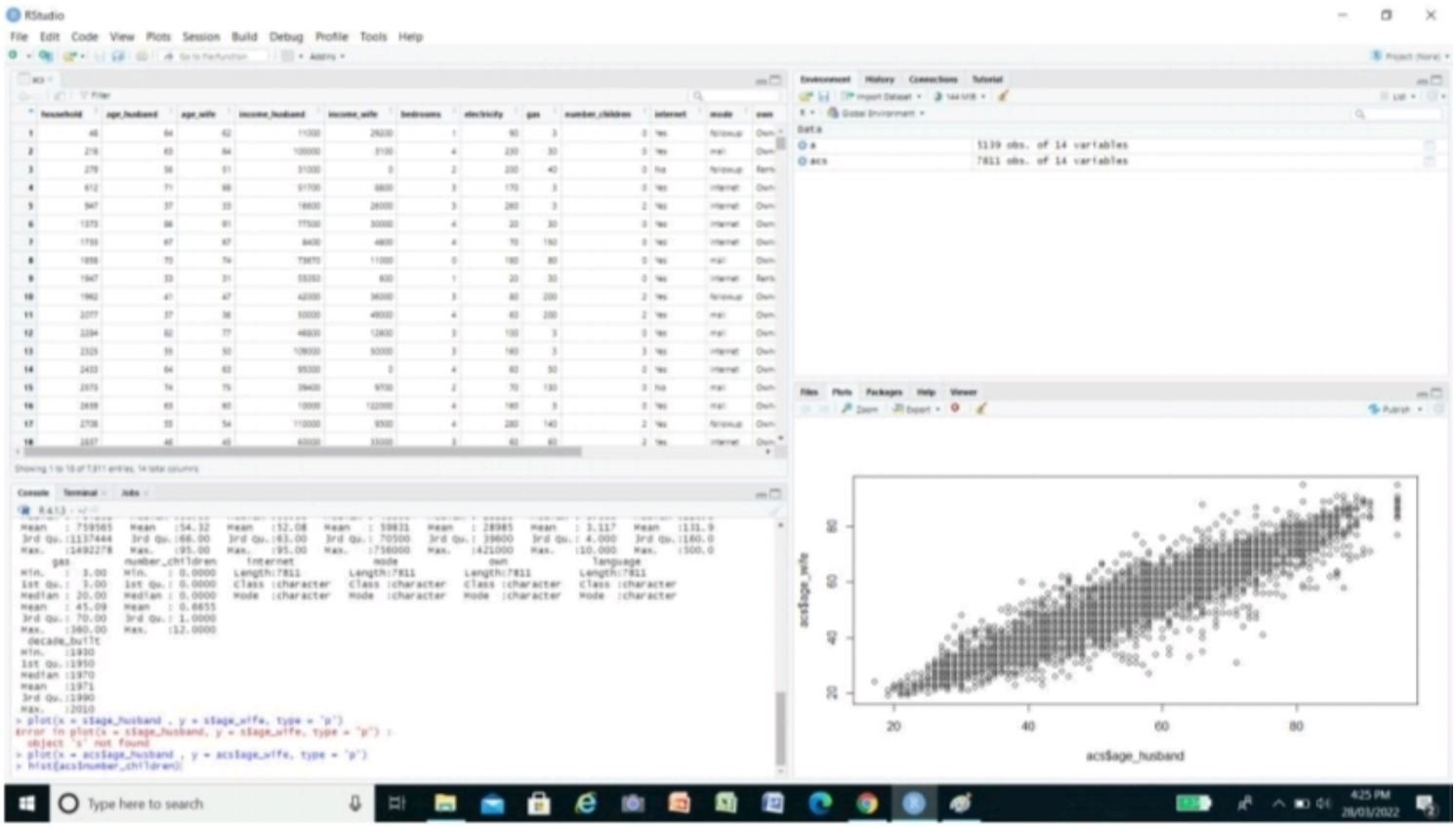
2. Transforming Data :

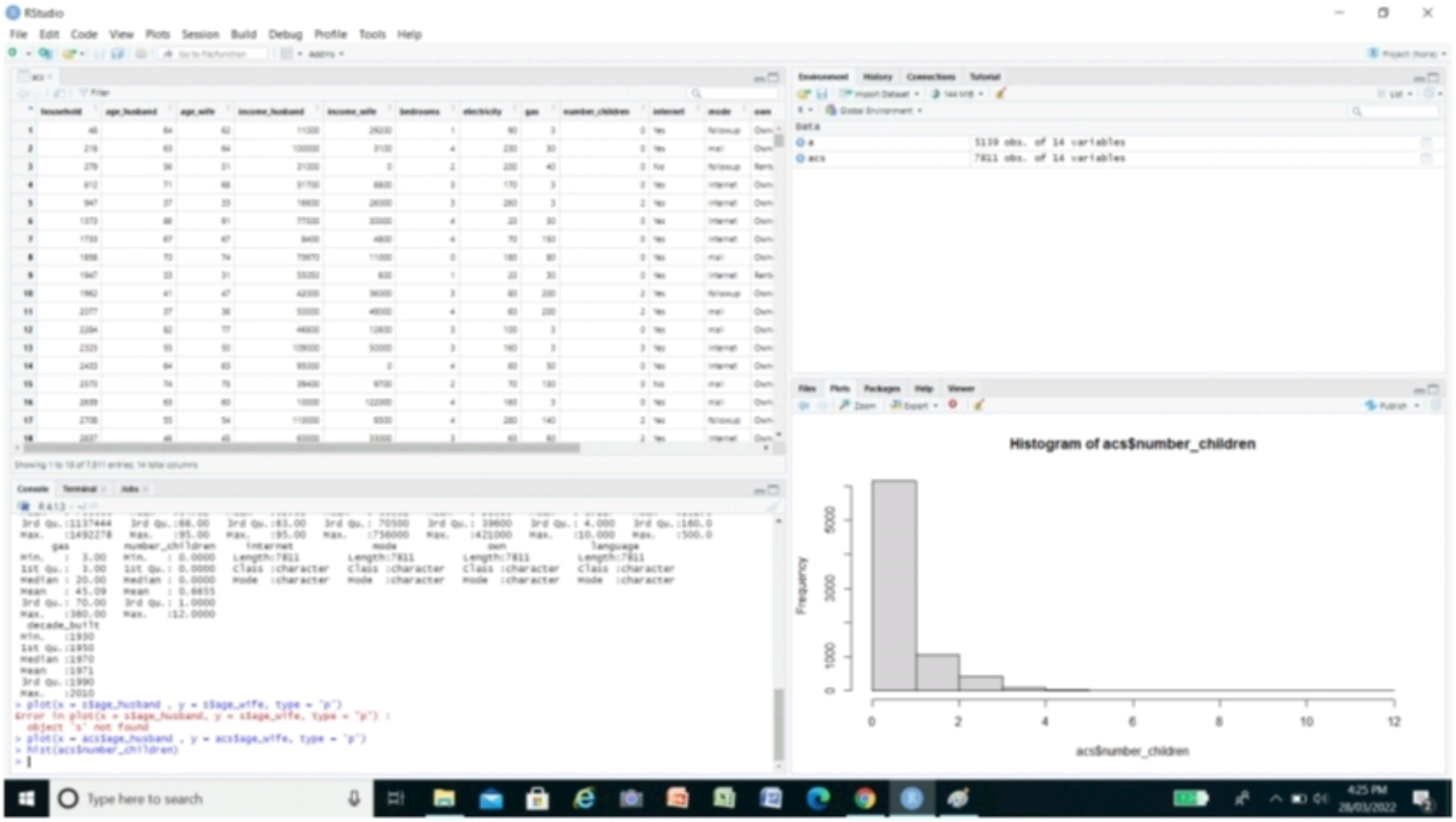


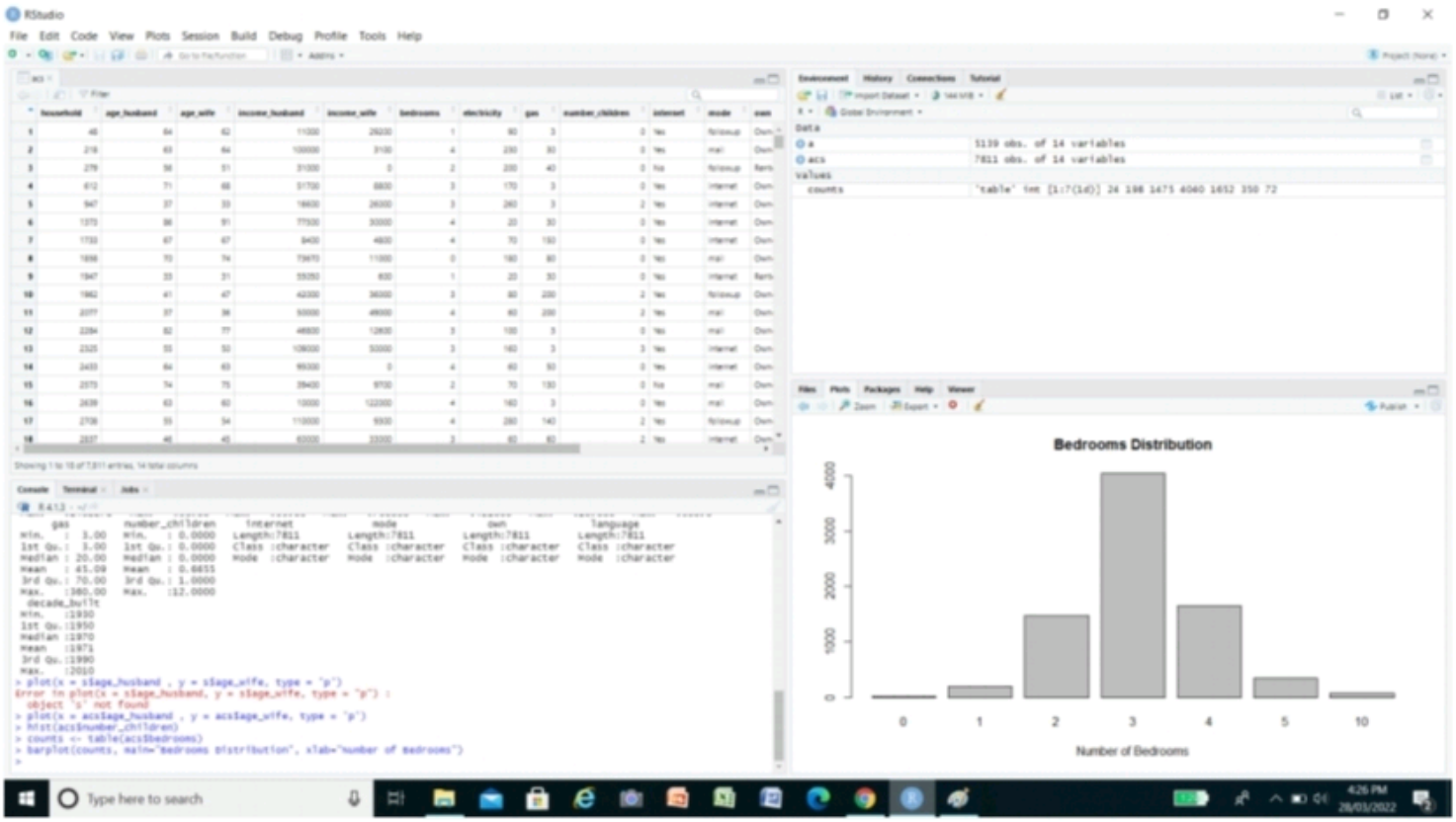
3. Getting Statistical Averages from data :



4. Plotting Data :




RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Histogram of acs$number_children
Frequency
acs$number_children
> 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)
Type here to search
4:25 PM
26/03/2022


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Bedrooms Distribution
Number of Bedrooms
> 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")
Type here to search
4:26 PM
26/03/2022