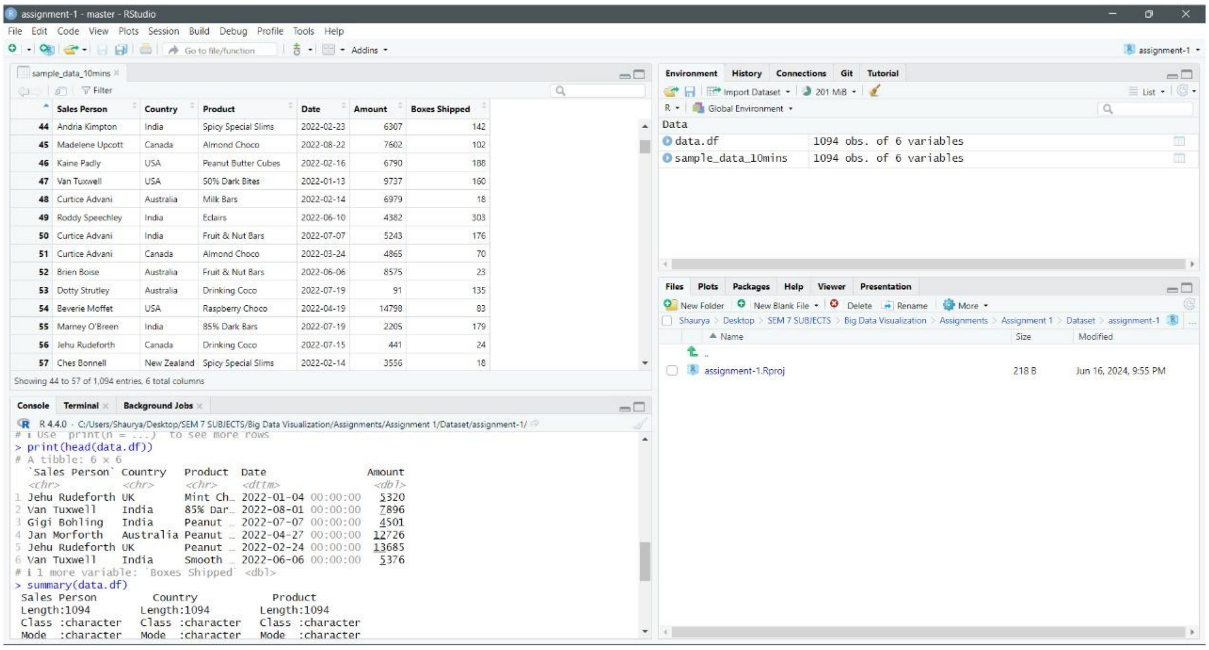
ASSIGNMENT 1

- **Shaurya Singh Srinet**

DATASET DETAILS- Sample Dataset taken from google

sample-data-10mins.xlsx

1. LOADING AND OVERVIEW OF THE DATA:



assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
201 MiB
Global Environment
List
Files Plots Packages Help Viewer Presentation
New Folder New Blank File Delete Rename More
Shaurya > Desktop > SEM 7 SUBJECTS > Big Data Visualization > Assignments > Assignment 1 > Dataset > assignment-1
Console Terminal Background Jobs

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70
52	Brien Boise	Australia	Fruit & Nut Bars	2022-06-06	8575	23
53	Dotty Strutley	Australia	Drinking Coco	2022-07-19	91	135
54	Beverie Moffet	USA	Raspberry Choco	2022-04-19	14798	83
55	Marney O'Breen	India	85% Dark Bars	2022-07-19	2205	179
56	Jehu Rudeforth	Canada	Drinking Coco	2022-07-15	441	24
57	Ches Bonnell	New Zealand	Spicy Special Slims	2022-02-14	3556	18

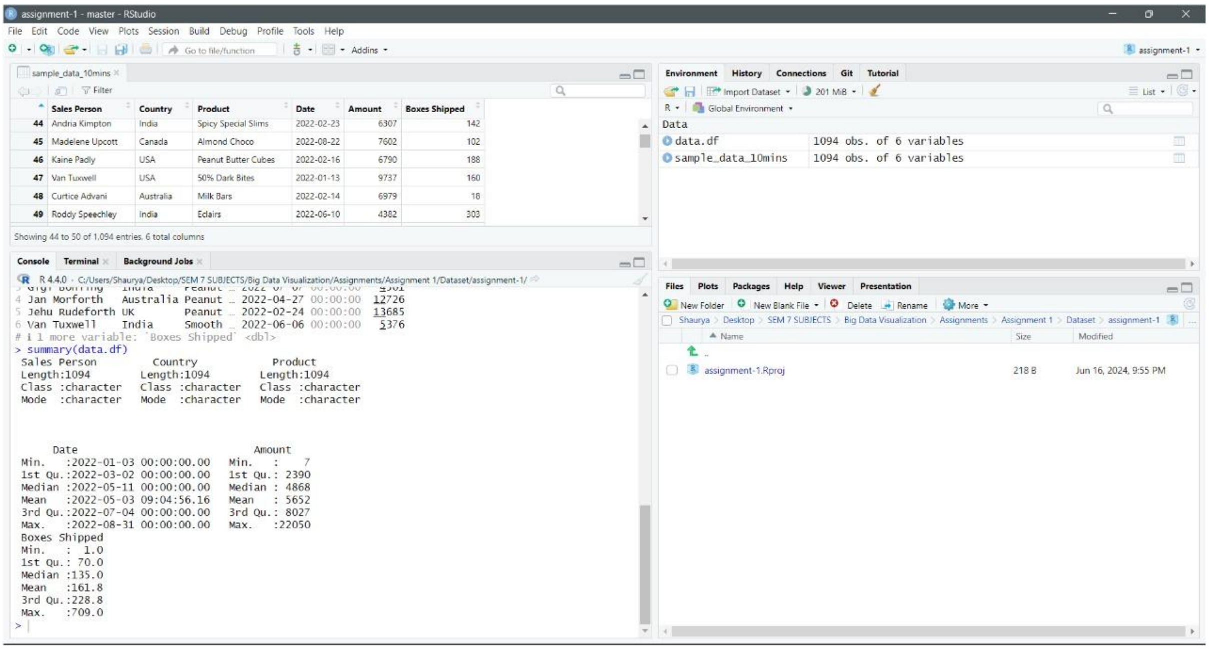
Showing 44 to 57 of 1,094 entries, 6 total columns

Data

data.df	1094 obs. of 6 variables	
sample_data_10mins	1094 obs. of 6 variables	

Name	Size	Modified
assignment-1.Rproj	218 B	Jun 16, 2024, 9:55 PM

```
R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/
# i Use `print(n = ...)` to see more rows
> print(head(data.df))
# A tibble: 6 × 6
  `Sales Person` Country   Product  Date                Amount
  <chr>          <chr>     <chr>    <dttm>               <dbl>
1 Jehu Rudeforth UK        Mint Ch… 2022-01-04 00:00:00   5320
2 Van Tuxwell    India     85% Dar… 2022-08-01 00:00:00   7896
3 Gigi Bohling   India     Peanut … 2022-07-07 00:00:00   4501
4 Jan Morforth   Australia Peanut … 2022-04-27 00:00:00  12726
5 Jehu Rudeforth UK        Peanut … 2022-02-24 00:00:00  13685
6 Van Tuxwell    India     Smooth … 2022-06-06 00:00:00   5376
# i 1 more variable: `Boxes Shipped` <dbl>
> summary(data.df)
 Sales Person         Country            Product         
 Length:1094        Length:1094        Length:1094       
 Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character  
```



	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303

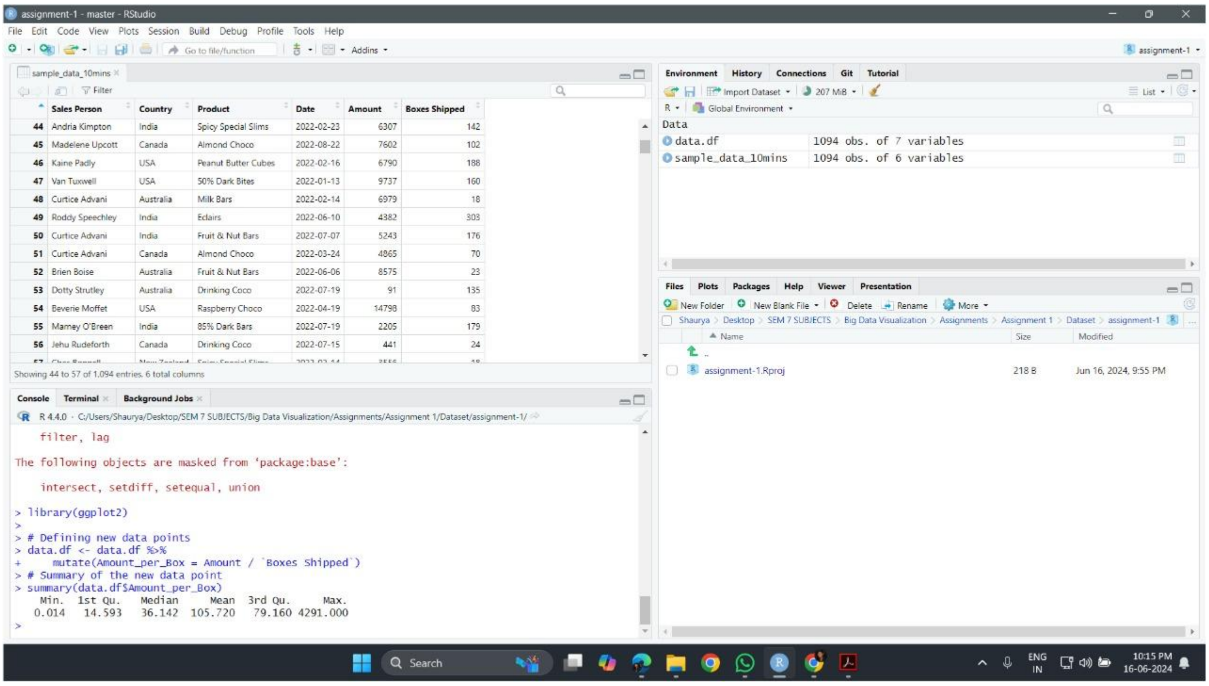
Showing 44 to 50 of 1,094 entries, 6 total columns

```
4 Jan Morforth   Australia Peanut … 2022-04-27 00:00:00  12726
5 Jehu Rudeforth UK        Peanut … 2022-02-24 00:00:00  13685
6 Van Tuxwell    India     Smooth … 2022-06-06 00:00:00   5376
# i 1 more variable: `Boxes Shipped` <dbl>
> summary(data.df)
 Sales Person         Country            Product         
 Length:1094        Length:1094        Length:1094       
 Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character  



      Date                          Amount     
 Min.   :2022-01-03 00:00:00.00   Min.   :    7  
 1st Qu.:2022-03-02 00:00:00.00   1st Qu.: 2390  
 Median :2022-05-11 00:00:00.00   Median : 4868  
 Mean   :2022-05-03 09:04:56.16   Mean   : 5652  
 3rd Qu.:2022-07-04 00:00:00.00   3rd Qu.: 8027  
 Max.   :2022-08-31 00:00:00.00   Max.   :22050  
 Boxes Shipped  
 Min.   :  1.0  
 1st Qu.: 70.0  
 Median :135.0  
 Mean   :161.8  
 3rd Qu.:228.8  
 Max.   :709.0  
> 
```

2. DEFINING NEW DATA POINTS:



sample_data_10mins

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70
52	Brien Boise	Australia	Fruit & Nut Bars	2022-06-06	8575	23
53	Dotty Strutley	Australia	Drinking Coco	2022-07-19	91	135
54	Beverie Moffet	USA	Raspberry Choco	2022-04-19	14798	83
55	Mamey O'Breen	India	85% Dark Bars	2022-07-19	2205	179
56	Jehu Rudeforth	Canada	Drinking Coco	2022-07-15	441	24

Showing 44 to 57 of 1,094 entries, 6 total columns

Data

```
data.df              1094 obs. of 7 variables
sample_data_10mins   1094 obs. of 6 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

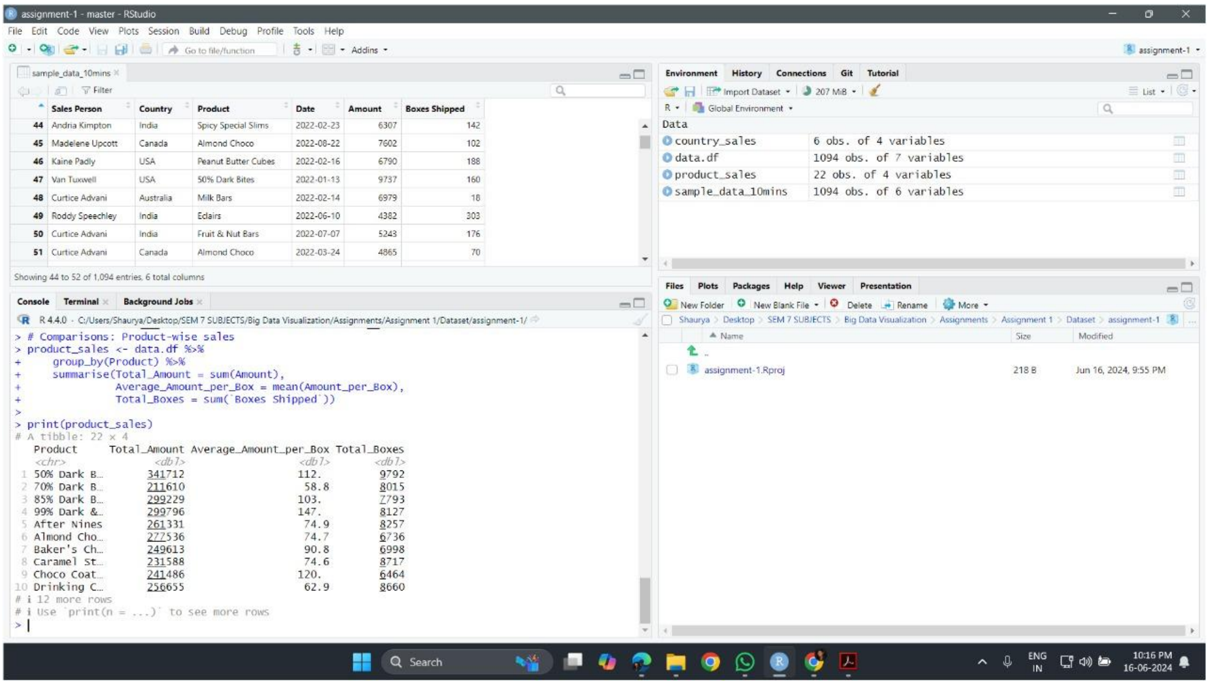
```
    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union

> library(ggplot2)
>
> # Defining new data points
> data.df <- data.df %>%
+     mutate(Amount_per_Box = Amount / `Boxes Shipped`)
> # Summary of the new data point
> summary(data.df$Amount_per_Box)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
   0.014   14.593   36.142  105.720   79.160 4291.000
>
```

3. COMPARISONS:



sample_data_10mins

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

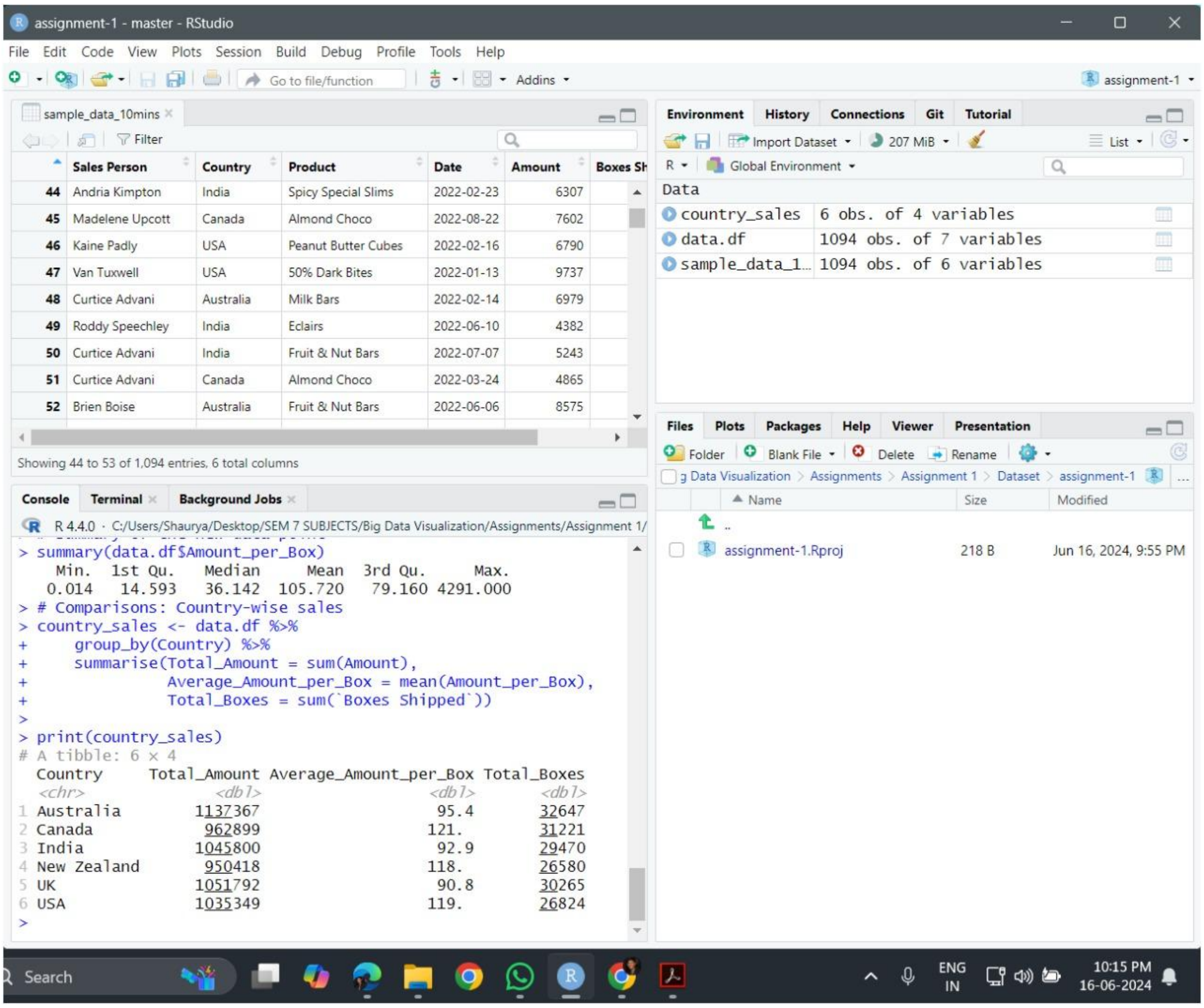
Showing 44 to 52 of 1,094 entries, 6 total columns

Data

```
country_sales        6 obs. of 4 variables
data.df              1094 obs. of 7 variables
product_sales        22 obs. of 4 variables
sample_data_10mins   1094 obs. of 6 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> # Comparisons: Product-wise sales
> product_sales <- data.df %>%
+     group_by(Product) %>%
+     summarise(Total_Amount = sum(Amount),
+               Average_Amount_per_Box = mean(Amount_per_Box),
+               Total_Boxes = sum(`Boxes Shipped`))
>
> print(product_sales)
# A tibble: 22 × 4
   Product      Total_Amount Average_Amount_per_Box Total_Boxes
   <chr>               <dbl>                  <dbl>       <dbl>
 1 50% Dark B…        341712                  112.         9792
 2 70% Dark B…        211610                   58.8        8015
 3 85% Dark B…        299229                  103.         7793
 4 99% Dark &…        299796                  147.         8127
 5 After Nines        261331                   74.9        8257
 6 Almond Cho…        277536                   74.7        6736
 7 Baker's Ch…        249613                   90.8        6998
 8 Caramel St…        231588                   74.6        8717
 9 Choco Coat…        241486                  120.         6464
10 Drinking C…        256655                   62.9        8660
# ℹ 12 more rows
# ℹ Use `print(n = ...)` to see more rows
> |
```



assignment-1 - master - RStudio

sample_data_10mins

	Sales Person	Country	Product	Date	Amount	Boxes Sh
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	
52	Brien Boise	Australia	Fruit & Nut Bars	2022-06-06	8575	

Showing 44 to 53 of 1,094 entries, 6 total columns

Global Environment

```
Data
country_sales      6 obs. of 4 variables
data.df            1094 obs. of 7 variables
sample_data_1…     1094 obs. of 6 variables
```

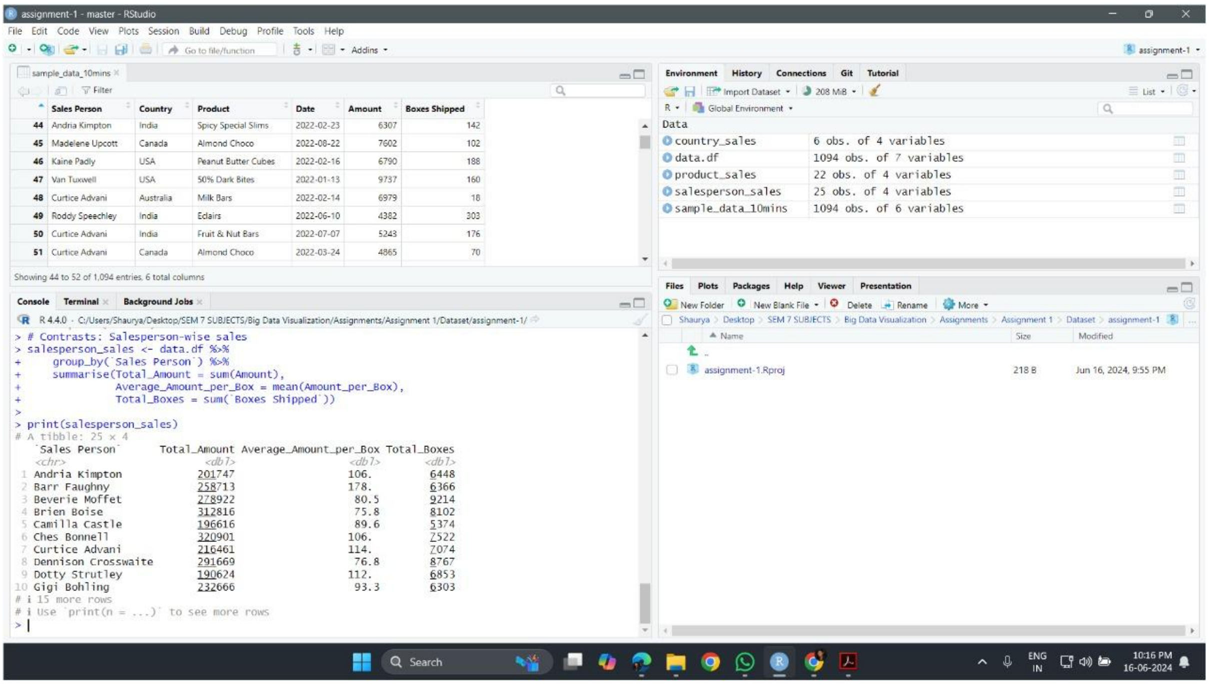
R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/

```
> summary(data.df$Amount_per_Box)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
   0.014   14.593   36.142  105.720   79.160 4291.000 
> # Comparisons: Country-wise sales
> country_sales <- data.df %>%
+     group_by(Country) %>%
+     summarise(Total_Amount = sum(Amount),
+               Average_Amount_per_Box = mean(Amount_per_Box),
+               Total_Boxes = sum(`Boxes Shipped`))
> 
> print(country_sales)
# A tibble: 6 × 4
  Country     Total_Amount Average_Amount_per_Box Total_Boxes
  <chr>              <dbl>                  <dbl>       <dbl>
1 Australia        1137367                   95.4       32647
2 Canada            962899                  121.        31221
3 India            1045800                   92.9       29470
4 New Zealand       950418                  118.        26580
5 UK               1051792                   90.8       30265
6 USA              1035349                  119.        26824
> 
```

Files: g Data Visualization > Assignments > Assignment 1 > Dataset > assignment-1

Name	Size	Modified
..		
assignment-1.Rproj	218 B	Jun 16, 2024, 9:55 PM

File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Filter
Environment History Connections Git Tutorial
Import Dataset
207 MiB
List
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Folder Blank File Delete Rename
Search
ENG IN
10:15 PM
16-06-2024



assignment-1 - master - RStudio

sample_data_10mins

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

Global Environment

```
Data
country_sales       6 obs. of 4 variables
data.df             1094 obs. of 7 variables
product_sales       22 obs. of 4 variables
salesperson_sales   25 obs. of 4 variables
sample_data_10mins  1094 obs. of 6 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

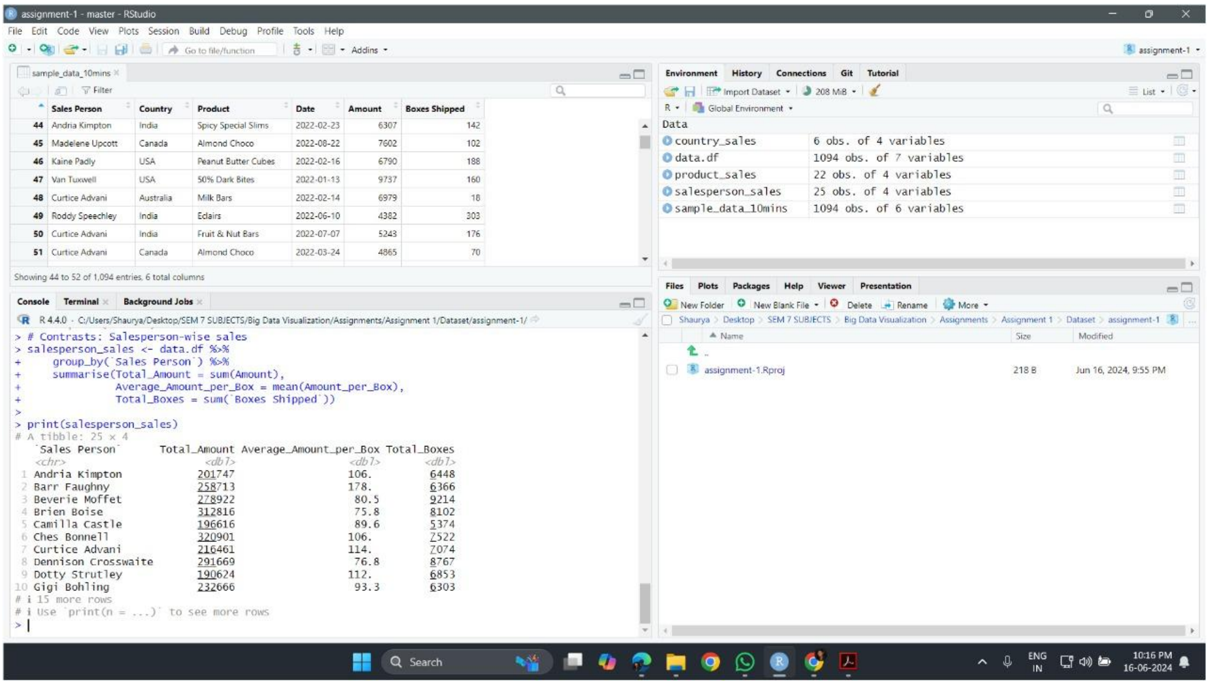
```
> # Contrasts: Salesperson-wise sales
> salesperson_sales <- data.df %>%
+     group_by(`Sales Person`) %>%
+     summarise(Total_Amount = sum(Amount),
+               Average_Amount_per_Box = mean(Amount_per_Box),
+               Total_Boxes = sum(`Boxes Shipped`))
> 
> print(salesperson_sales)
# A tibble: 25 × 4
   `Sales Person`      Total_Amount Average_Amount_per_Box Total_Boxes
   <chr>                      <dbl>                  <dbl>       <dbl>
 1 Andria Kimpton            201747                  106.         6448
 2 Barr Faughny              258713                  178.         6366
 3 Beverie Moffet            278922                   80.5        9214
 4 Brien Boise               312816                   75.8        8102
 5 Camilla Castle            196616                   89.6        5374
 6 Ches Bonnell              320901                  106.         7522
 7 Curtice Advani            216461                  114.         7074
 8 Dennison Crosswaite       291669                   76.8        8767
 9 Dotty Strutley            190624                  112.         6853
10 Gigi Bohling              232666                   93.3        6303
# ℹ 15 more rows
# ℹ Use `print(n = ...)` to see more rows
> 
```

Files: Shaurya > Desktop > SEM 7 SUBJECTS > Big Data Visualization > Assignments > Assignment 1 > Dataset > assignment-1

Name	Size	Modified
..		
assignment-1.Rproj	218 B	Jun 16, 2024, 9:55 PM

File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Filter
Environment History Connections Git Tutorial
Import Dataset
208 MiB
List
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
New Folder New Blank File Delete Rename More
Search
ENG IN
10:16 PM
16-06-2024

4. CONTRASTS:



assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
208 MiB
List
Global Environment
Files Plots Packages Help Viewer Presentation
New Folder
New Blank File
Delete
Rename
More
Shaurya > Desktop > SEM 7 SUBJECTS > Big Data Visualization > Assignments > Assignment 1 > Dataset > assignment-1
Console Terminal Background Jobs
Search
ENG IN
10:16 PM 16-06-2024

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

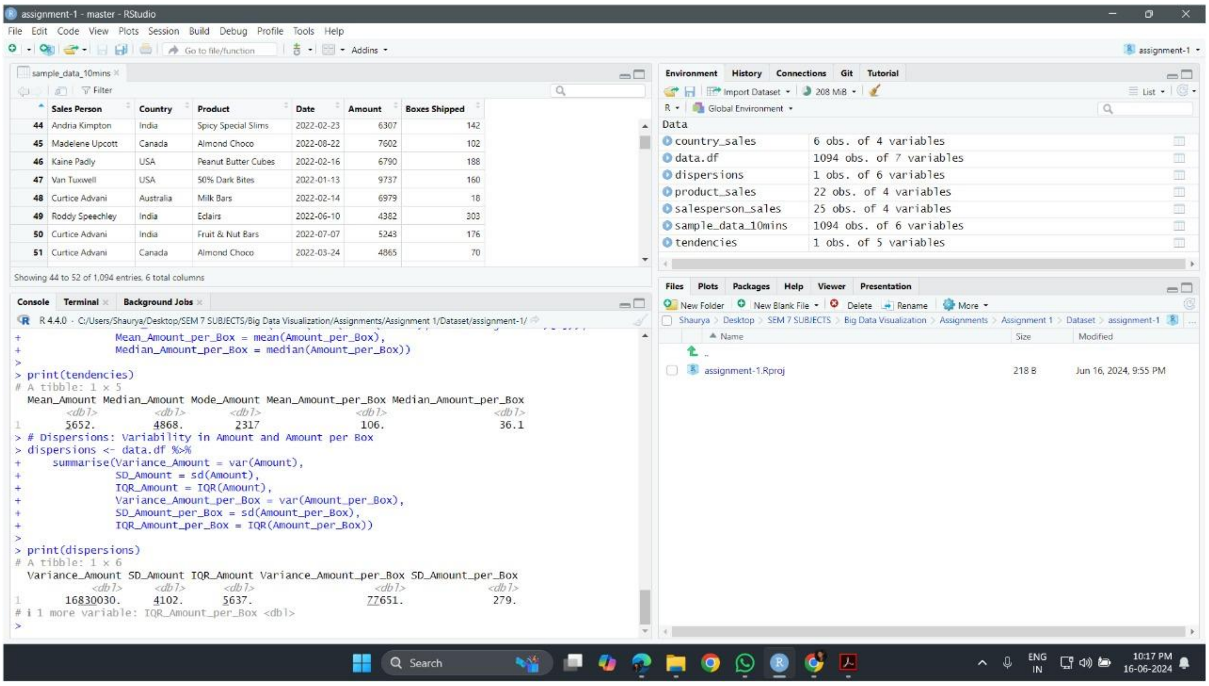
```
Data
country_sales        6 obs. of 4 variables
data.df              1094 obs. of 7 variables
product_sales        22 obs. of 4 variables
salesperson_sales    25 obs. of 4 variables
sample_data_10mins   1094 obs. of 6 variables
```

Name	Size	Modified
assignment-1.Rproj	218 B	Jun 16, 2024, 9:55 PM

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> # Contrasts: Salesperson-wise sales
> salesperson_sales <- data.df %>%
+     group_by(`Sales Person`) %>%
+     summarise(Total_Amount = sum(Amount),
+               Average_Amount_per_Box = mean(Amount_per_Box),
+               Total_Boxes = sum(`Boxes Shipped`))
>
> print(salesperson_sales)
# A tibble: 25 × 4
   `Sales Person`      Total_Amount Average_Amount_per_Box Total_Boxes
   <chr>                      <dbl>                  <dbl>       <dbl>
 1 Andria Kimpton            201747                  106.         6448
 2 Barr Faughny              258713                  178.         6366
 3 Beverie Moffet            278922                   80.5        9214
 4 Brien Boise               312816                   75.8        8102
 5 Camilla Castle            196616                   89.6        5374
 6 Ches Bonnell              320901                  106.         7522
 7 Curtice Advani            216461                  114.         7074
 8 Dennison Crosswaite       291669                   76.8        8767
 9 Dotty Strutley            190624                  112.         6853
10 Gigi Bohling              232666                   93.3        6303
# ℹ 15 more rows
# ℹ Use `print(n = ...)` to see more rows
> |
```

5. IDENTIFYING TENDENCIES:



assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
208 MiB
List
Global Environment
Files Plots Packages Help Viewer Presentation
New Folder
New Blank File
Delete
Rename
More
Shaurya > Desktop > SEM 7 SUBJECTS > Big Data Visualization > Assignments > Assignment 1 > Dataset > assignment-1
Console Terminal Background Jobs
Search
ENG IN
10:17 PM 16-06-2024

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

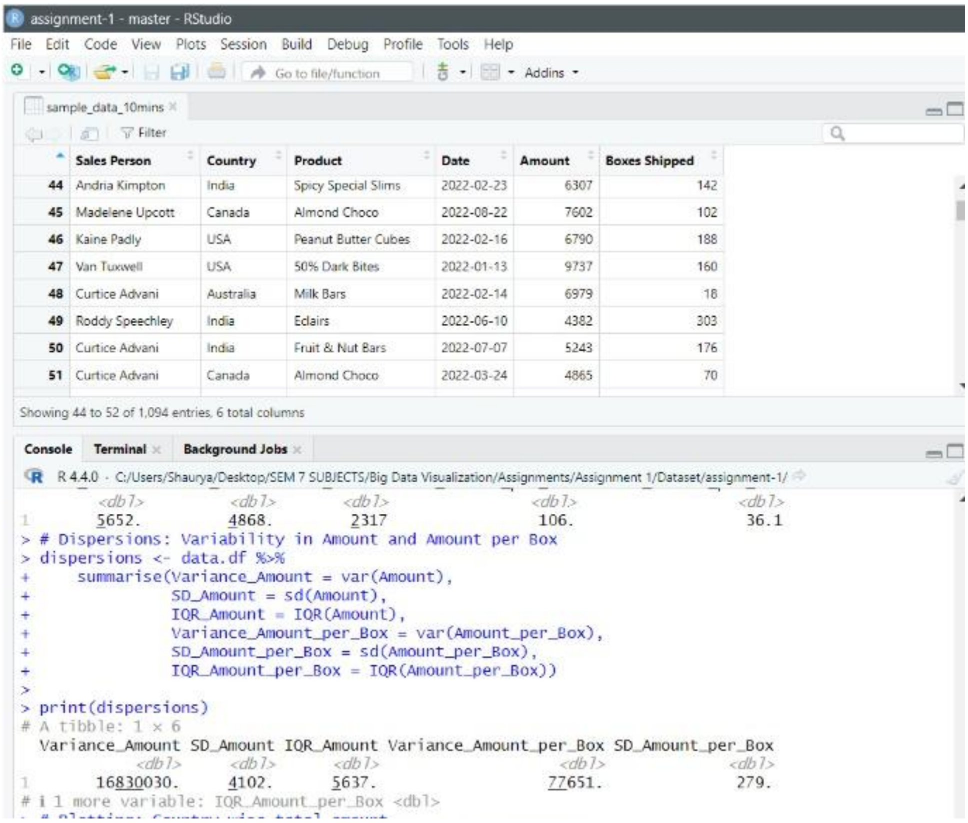
```
Data
country_sales        6 obs. of 4 variables
data.df              1094 obs. of 7 variables
dispersions          1 obs. of 6 variables
product_sales        22 obs. of 4 variables
salesperson_sales    25 obs. of 4 variables
sample_data_10mins   1094 obs. of 6 variables
tendencies           1 obs. of 5 variables
```

Name	Size	Modified
assignment-1.Rproj	218 B	Jun 16, 2024, 9:55 PM

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
+               Mean_Amount_per_Box = mean(Amount_per_Box),
+               Median_Amount_per_Box = median(Amount_per_Box))
>
> print(tendencies)
# A tibble: 1 × 5
  Mean_Amount Median_Amount Mode_Amount Mean_Amount_per_Box Median_Amount_per_Box
        <dbl>         <dbl>       <dbl>               <dbl>                 <dbl>
1       5652.         4868.        2317                106.                  36.1
> # Dispersions: Variability in Amount and Amount per Box
> dispersions <- data.df %>%
+     summarise(Variance_Amount = var(Amount),
+               SD_Amount = sd(Amount),
+               IQR_Amount = IQR(Amount),
+               Variance_Amount_per_Box = var(Amount_per_Box),
+               SD_Amount_per_Box = sd(Amount_per_Box),
+               IQR_Amount_per_Box = IQR(Amount_per_Box))
>
> print(dispersions)
# A tibble: 1 × 6
  Variance_Amount SD_Amount IQR_Amount Variance_Amount_per_Box SD_Amount_per_Box
            <dbl>     <dbl>      <dbl>                   <dbl>             <dbl>
1       16830030.     4102.      5637.                  77651.              279.
# ℹ 1 more variable: IQR_Amount_per_Box <dbl>
>
```

6. DISPERSIONS:



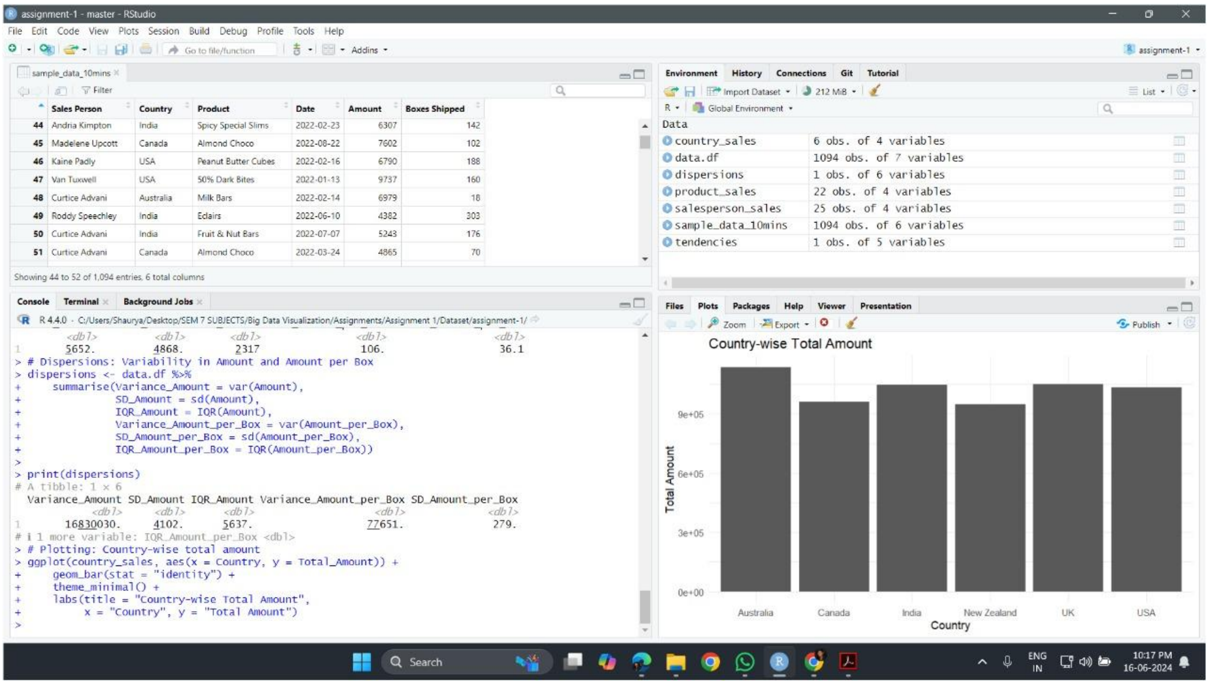
assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
sample_data_10mins
Filter
Console
Terminal
Background Jobs
R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
       <dbl>        <dbl>        <dbl>                   <dbl>                  <dbl>
1      5652.        4868.        2317                    106.                   36.1
> # Dispersions: Variability in Amount and Amount per Box
> dispersions <- data.df %>%
+     summarise(Variance_Amount = var(Amount),
+               SD_Amount = sd(Amount),
+               IQR_Amount = IQR(Amount),
+               Variance_Amount_per_Box = var(Amount_per_Box),
+               SD_Amount_per_Box = sd(Amount_per_Box),
+               IQR_Amount_per_Box = IQR(Amount_per_Box))
>
> print(dispersions)
# A tibble: 1 × 6
  Variance_Amount SD_Amount IQR_Amount Variance_Amount_per_Box SD_Amount_per_Box
            <dbl>     <dbl>      <dbl>                   <dbl>             <dbl>
1       16830030.     4102.      5637.                  77651.              279.
# ℹ 1 more variable: IQR_Amount_per_Box <dbl>
```

7. PLOTTING:



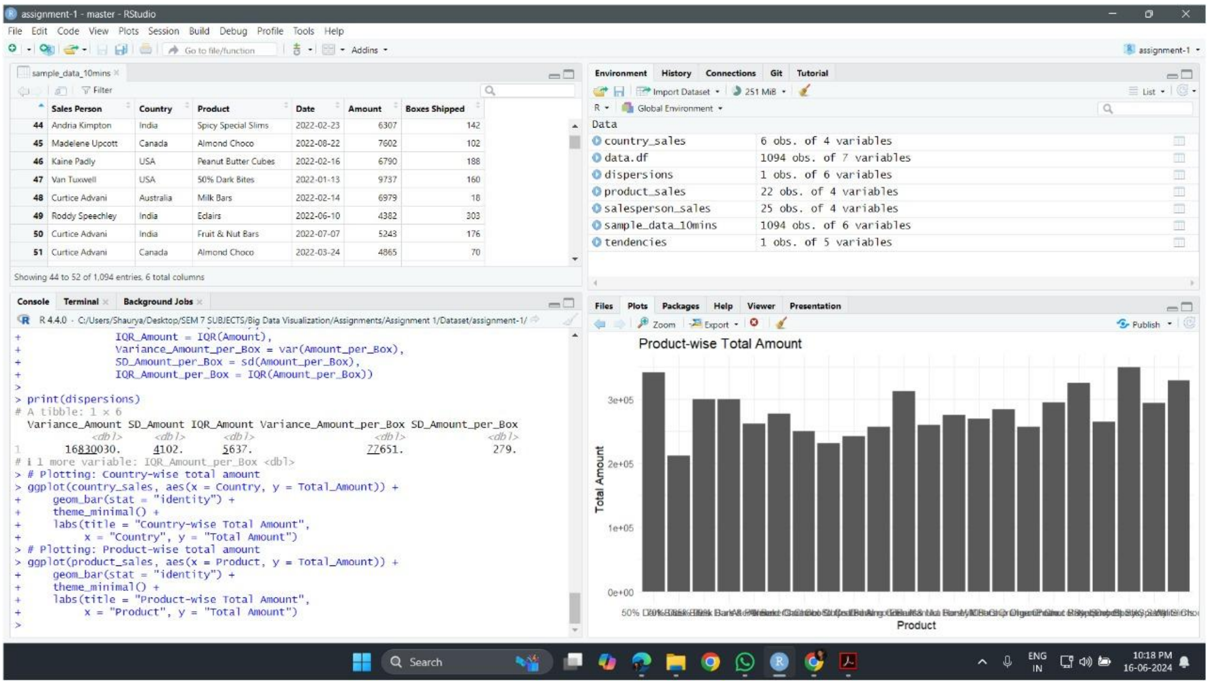
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

Data	
country_sales	6 obs. of 4 variables
data.df	1094 obs. of 7 variables
dispersions	1 obs. of 6 variables
product_sales	22 obs. of 4 variables
salesperson_sales	25 obs. of 4 variables
sample_data_10mins	1094 obs. of 6 variables
tendencies	1 obs. of 5 variables

```
       <dbl>        <dbl>        <dbl>                   <dbl>                  <dbl>
1      5652.        4868.        2317                    106.                   36.1
> # Dispersions: Variability in Amount and Amount per Box
> dispersions <- data.df %>%
+     summarise(Variance_Amount = var(Amount),
+               SD_Amount = sd(Amount),
+               IQR_Amount = IQR(Amount),
+               Variance_Amount_per_Box = var(Amount_per_Box),
+               SD_Amount_per_Box = sd(Amount_per_Box),
+               IQR_Amount_per_Box = IQR(Amount_per_Box))
>
> print(dispersions)
# A tibble: 1 × 6
  Variance_Amount SD_Amount IQR_Amount Variance_Amount_per_Box SD_Amount_per_Box
            <dbl>     <dbl>      <dbl>                   <dbl>             <dbl>
1       16830030.     4102.      5637.                  77651.              279.
# ℹ 1 more variable: IQR_Amount_per_Box <dbl>
```

```
> # Plotting: Country-wise total amount
> ggplot(country_sales, aes(x = Country, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Country-wise Total Amount",
+          x = "Country", y = "Total Amount")
>
```


assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
251 MiB
List
Global Environment
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Product-wise Total Amount
Total Amount
3e+05
2e+05
1e+05
0e+00
Product
Search
ENG IN
10:18 PM 16-06-2024

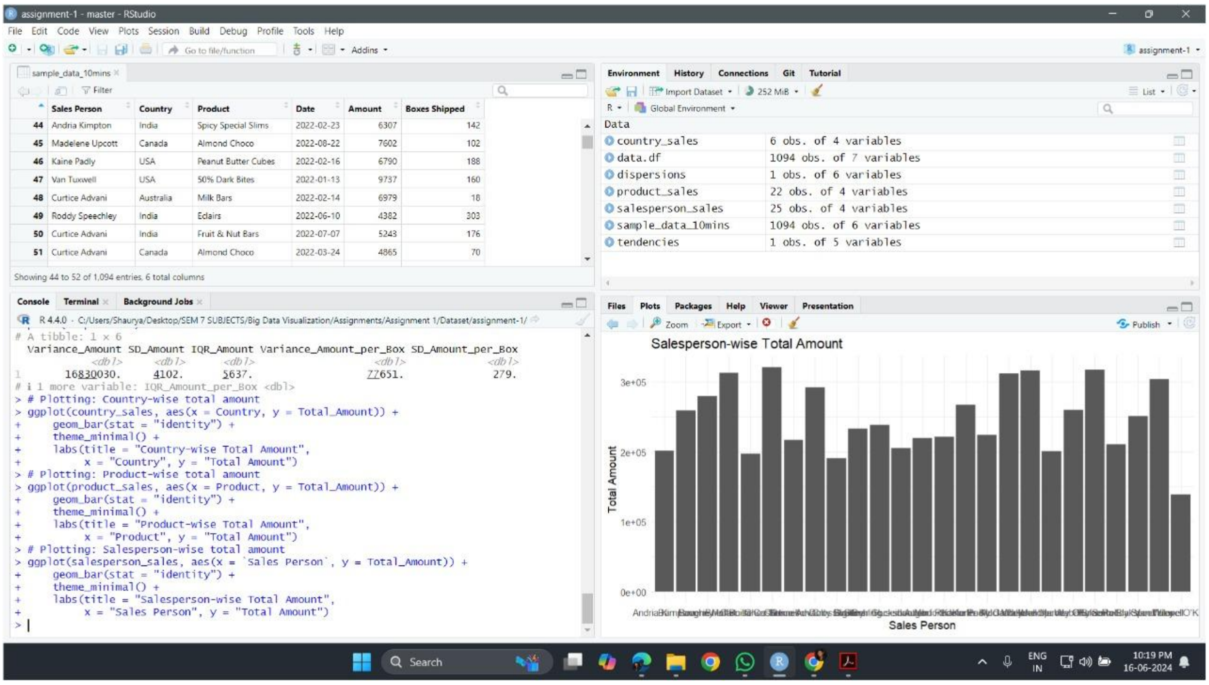
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales          6 obs. of 4 variables
data.df                1094 obs. of 7 variables
dispersions            1 obs. of 6 variables
product_sales          22 obs. of 4 variables
salesperson_sales      25 obs. of 4 variables
sample_data_10mins     1094 obs. of 6 variables
tendencies             1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
+                IQR_Amount = IQR(Amount),
+                Variance_Amount_per_Box = var(Amount_per_Box),
+                SD_Amount_per_Box = sd(Amount_per_Box),
+                IQR_Amount_per_Box = IQR(Amount_per_Box))
>
> print(dispersions)
# A tibble: 1 × 6
  Variance_Amount SD_Amount IQR_Amount Variance_Amount_per_Box SD_Amount_per_Box
            <dbl>     <dbl>      <dbl>                   <dbl>             <dbl>
1       16830030.     4102.      5637.                  77651.              279.
# ℹ 1 more variable: IQR_Amount_per_Box <dbl>
> # Plotting: Country-wise total amount
> ggplot(country_sales, aes(x = Country, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Country-wise Total Amount",
+          x = "Country", y = "Total Amount")
> # Plotting: Product-wise total amount
> ggplot(product_sales, aes(x = Product, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Product-wise Total Amount",
+          x = "Product", y = "Total Amount")
>
```


assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
252 MiB
List
Global Environment
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Salesperson-wise Total Amount
Total Amount
3e+05
2e+05
1e+05
0e+00
Sales Person
Search
ENG IN
10:19 PM 16-06-2024

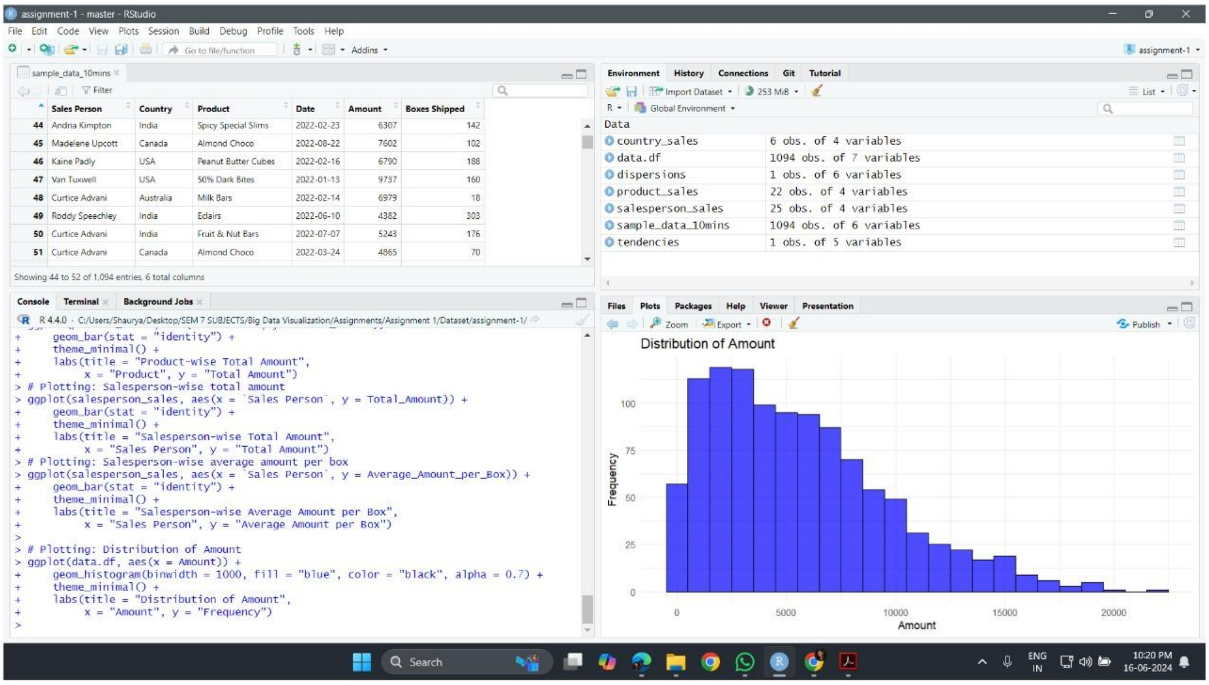
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales          6 obs. of 4 variables
data.df                1094 obs. of 7 variables
dispersions            1 obs. of 6 variables
product_sales          22 obs. of 4 variables
salesperson_sales      25 obs. of 4 variables
sample_data_10mins     1094 obs. of 6 variables
tendencies             1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
# A tibble: 1 × 6
  Variance_Amount SD_Amount IQR_Amount Variance_Amount_per_Box SD_Amount_per_Box
            <dbl>     <dbl>      <dbl>                   <dbl>             <dbl>
1       16830030.     4102.      5637.                  77651.              279.
# ℹ 1 more variable: IQR_Amount_per_Box <dbl>
> # Plotting: Country-wise total amount
> ggplot(country_sales, aes(x = Country, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Country-wise Total Amount",
+          x = "Country", y = "Total Amount")
> # Plotting: Product-wise total amount
> ggplot(product_sales, aes(x = Product, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Product-wise Total Amount",
+          x = "Product", y = "Total Amount")
> # Plotting: Salesperson-wise total amount
> ggplot(salesperson_sales, aes(x = `Sales Person`, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Total Amount",
+          x = "Sales Person", y = "Total Amount")
> 
```



	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

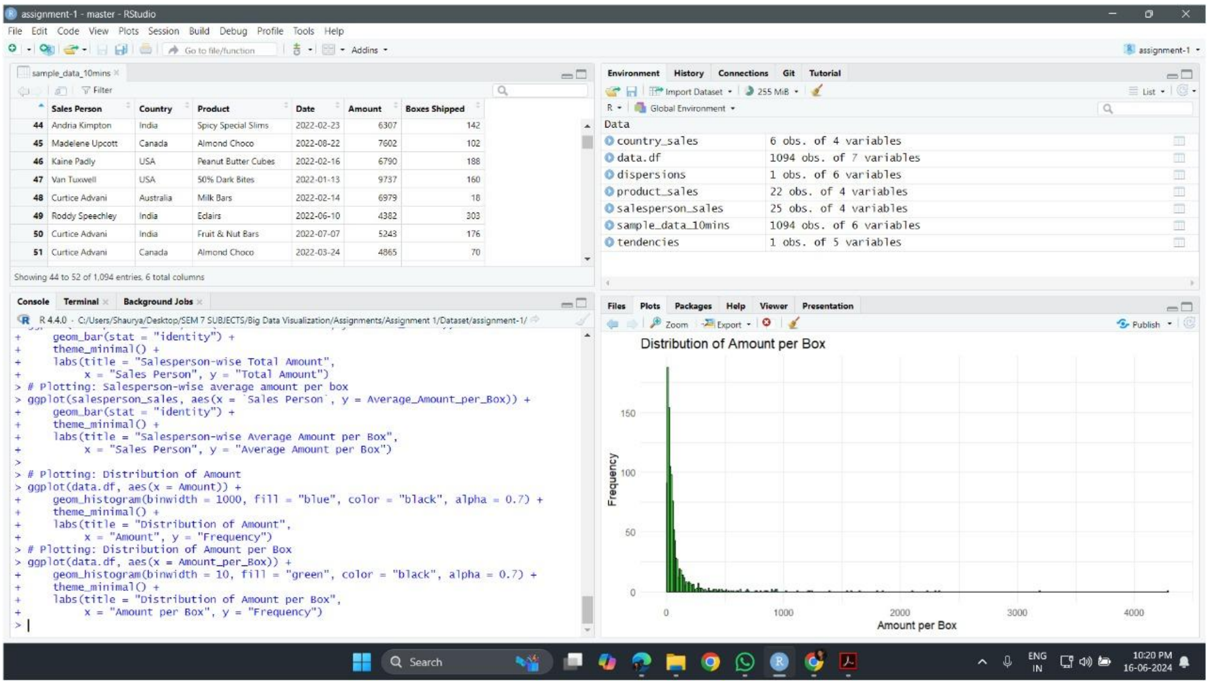
Showing 44 to 52 of 1,094 entries, 6 total columns

Data

country_sales	6 obs. of 4 variables
data.df	1094 obs. of 7 variables
dispersions	1 obs. of 6 variables
product_sales	22 obs. of 4 variables
salesperson_sales	25 obs. of 4 variables
sample_data_10mins	1094 obs. of 6 variables
tendencies	1 obs. of 5 variables

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Product-wise Total Amount",
+          x = "Product", y = "Total Amount")
> # Plotting: Salesperson-wise total amount
> ggplot(salesperson_sales, aes(x = `Sales Person`, y = Total_Amount)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Total Amount",
+          x = "Sales Person", y = "Total Amount")
> # Plotting: Salesperson-wise average amount per box
> ggplot(salesperson_sales, aes(x = `Sales Person`, y = Average_Amount_per_Box)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Average Amount per Box",
+          x = "Sales Person", y = "Average Amount per Box")
>
> # Plotting: Distribution of Amount
> ggplot(data.df, aes(x = Amount)) +
+     geom_histogram(binwidth = 1000, fill = "blue", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount",
+          x = "Amount", y = "Frequency")
>
```



assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
255 MiB
List
R
Global Environment
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Distribution of Amount per Box
Frequency
0
50
100
150
0
1000
2000
3000
4000
Amount per Box
Search
ENG IN
10:20 PM 16-06-2024

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

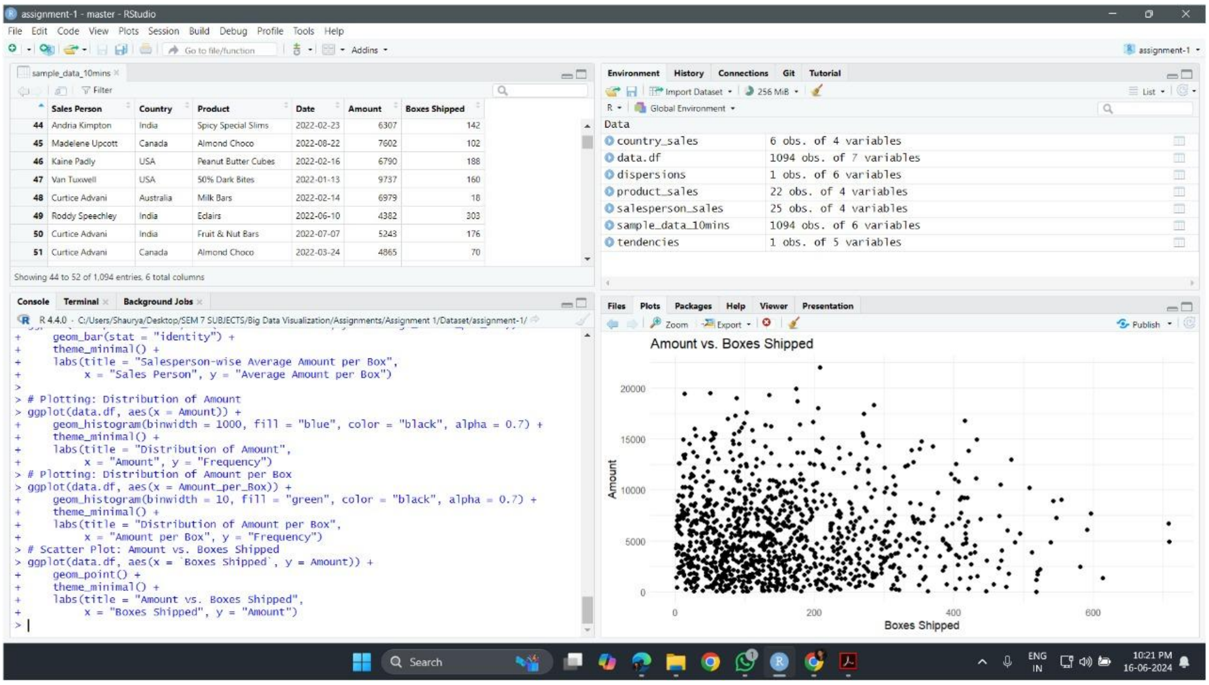
Showing 44 to 52 of 1,094 entries, 6 total columns

Data

country_sales	6 obs. of 4 variables
data.df	1094 obs. of 7 variables
dispersions	1 obs. of 6 variables
product_sales	22 obs. of 4 variables
salesperson_sales	25 obs. of 4 variables
sample_data_10mins	1094 obs. of 6 variables
tendencies	1 obs. of 5 variables

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Total Amount",
+          x = "Sales Person", y = "Total Amount")
> # Plotting: Salesperson-wise average amount per box
> ggplot(salesperson_sales, aes(x = `Sales Person`, y = Average_Amount_per_Box)) +
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Average Amount per Box",
+          x = "Sales Person", y = "Average Amount per Box")
>
> # Plotting: Distribution of Amount
> ggplot(data.df, aes(x = Amount)) +
+     geom_histogram(binwidth = 1000, fill = "blue", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount",
+          x = "Amount", y = "Frequency")
> # Plotting: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_histogram(binwidth = 10, fill = "green", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount per Box",
+          x = "Amount per Box", y = "Frequency")
> |
```



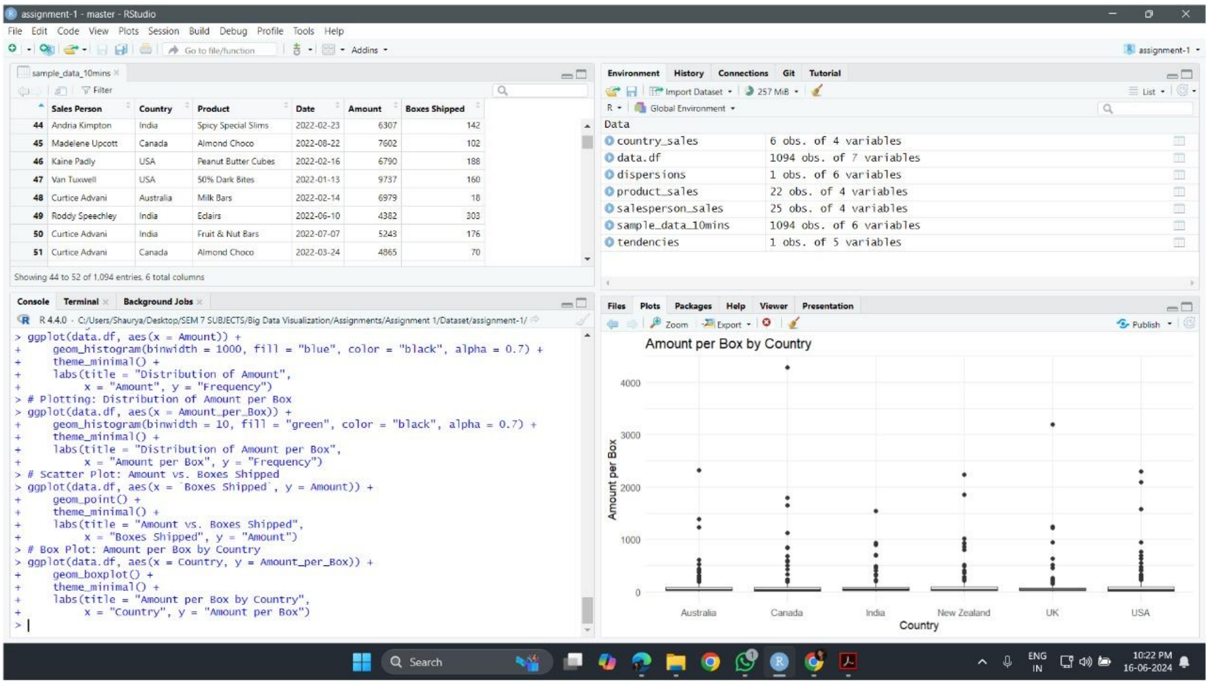
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales          6 obs. of 4 variables
data.df                1094 obs. of 7 variables
dispersions            1 obs. of 6 variables
product_sales          22 obs. of 4 variables
salesperson_sales      25 obs. of 4 variables
sample_data_10mins     1094 obs. of 6 variables
tendencies             1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
+     geom_bar(stat = "identity") +
+     theme_minimal() +
+     labs(title = "Salesperson-wise Average Amount per Box",
+          x = "Sales Person", y = "Average Amount per Box")
>
> # Plotting: Distribution of Amount
> ggplot(data.df, aes(x = Amount)) +
+     geom_histogram(binwidth = 1000, fill = "blue", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount",
+          x = "Amount", y = "Frequency")
> # Plotting: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_histogram(binwidth = 10, fill = "green", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount per Box",
+          x = "Amount per Box", y = "Frequency")
> # Scatter Plot: Amount vs. Boxes Shipped
> ggplot(data.df, aes(x = `Boxes Shipped`, y = Amount)) +
+     geom_point() +
+     theme_minimal() +
+     labs(title = "Amount vs. Boxes Shipped",
+          x = "Boxes Shipped", y = "Amount")
> 
```


assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
257 MiB
List
Global Environment
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
Amount per Box by Country
4000
3000
2000
1000
0
Amount per Box
Australia
Canada
India
New Zealand
UK
USA
Country
Search
ENG
IN
10:22 PM
16-06-2024

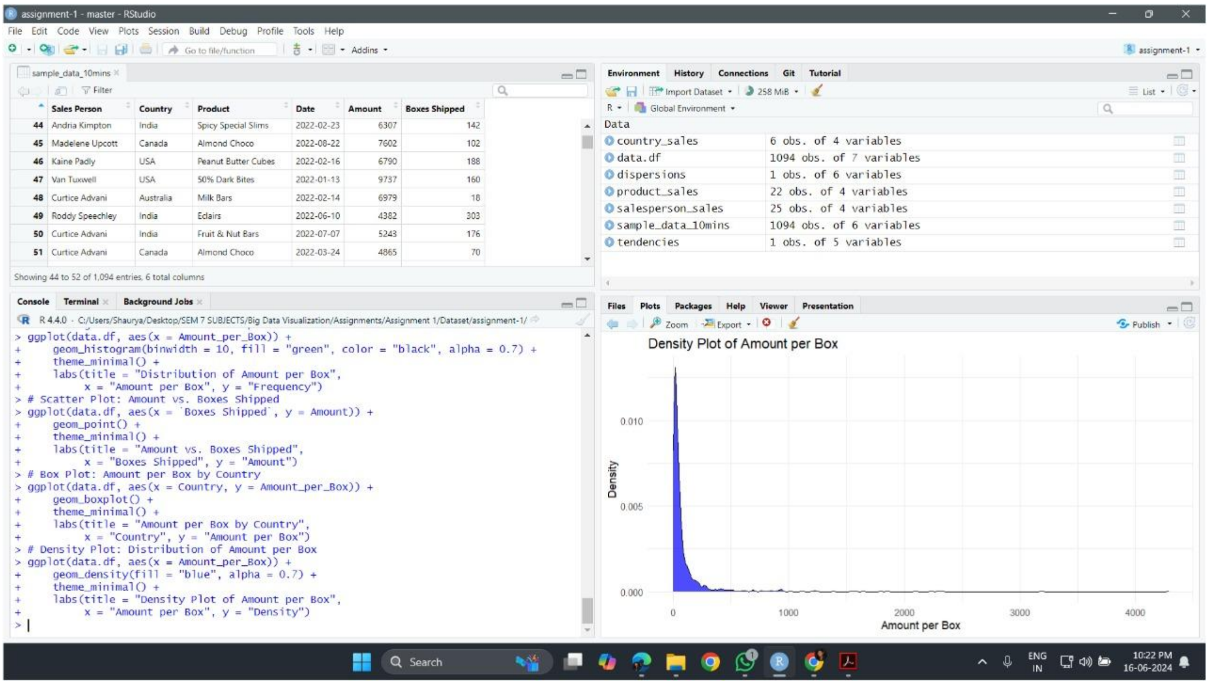
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales          6 obs. of 4 variables
data.df                1094 obs. of 7 variables
dispersions            1 obs. of 6 variables
product_sales          22 obs. of 4 variables
salesperson_sales      25 obs. of 4 variables
sample_data_10mins     1094 obs. of 6 variables
tendencies             1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> ggplot(data.df, aes(x = Amount)) +
+     geom_histogram(binwidth = 1000, fill = "blue", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount",
+          x = "Amount", y = "Frequency")
> # Plotting: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_histogram(binwidth = 10, fill = "green", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount per Box",
+          x = "Amount per Box", y = "Frequency")
> # Scatter Plot: Amount vs. Boxes Shipped
> ggplot(data.df, aes(x = `Boxes Shipped`, y = Amount)) +
+     geom_point() +
+     theme_minimal() +
+     labs(title = "Amount vs. Boxes Shipped",
+          x = "Boxes Shipped", y = "Amount")
> # Box Plot: Amount per Box by Country
> ggplot(data.df, aes(x = Country, y = Amount_per_Box)) +
+     geom_boxplot() +
+     theme_minimal() +
+     labs(title = "Amount per Box by Country",
+          x = "Country", y = "Amount per Box")
> 
```



assignment-1 - master - RStudio

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

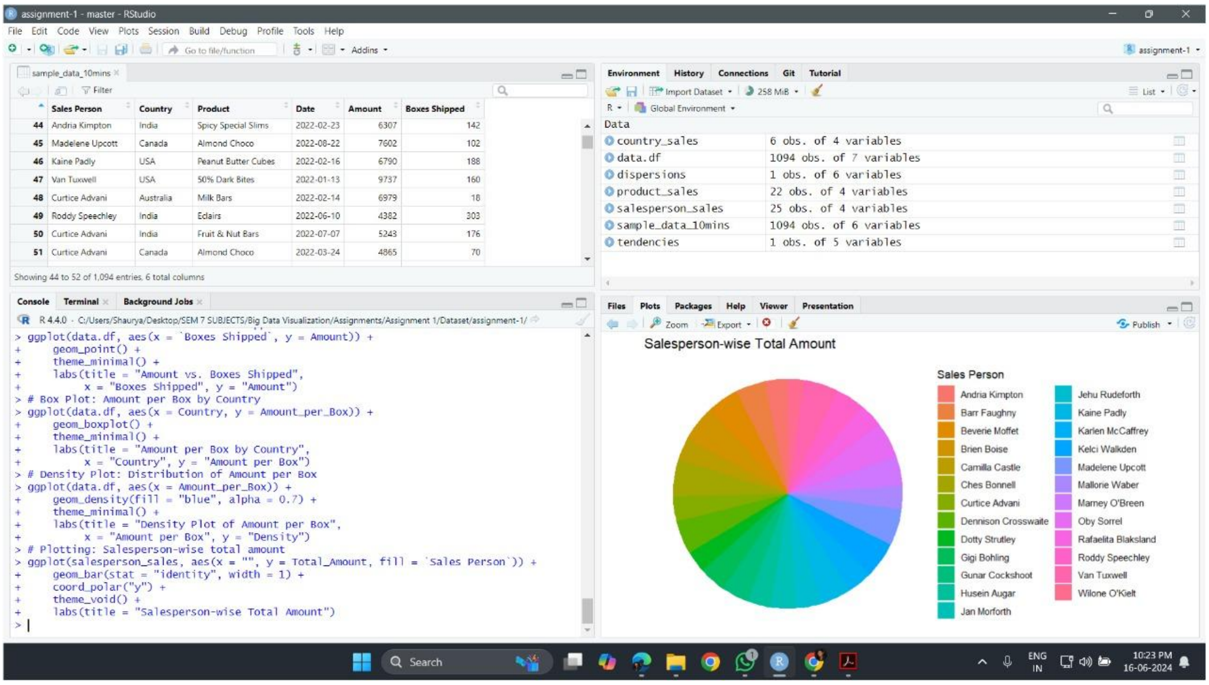
Showing 44 to 52 of 1,094 entries, 6 total columns

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_histogram(binwidth = 10, fill = "green", color = "black", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Distribution of Amount per Box",
+          x = "Amount per Box", y = "Frequency")
> # Scatter Plot: Amount vs. Boxes Shipped
> ggplot(data.df, aes(x = `Boxes Shipped`, y = Amount)) +
+     geom_point() +
+     theme_minimal() +
+     labs(title = "Amount vs. Boxes Shipped",
+          x = "Boxes Shipped", y = "Amount")
> # Box Plot: Amount per Box by Country
> ggplot(data.df, aes(x = Country, y = Amount_per_Box)) +
+     geom_boxplot() +
+     theme_minimal() +
+     labs(title = "Amount per Box by Country",
+          x = "Country", y = "Amount per Box")
> # Density Plot: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_density(fill = "blue", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Density Plot of Amount per Box",
+          x = "Amount per Box", y = "Density")
> 
```

Data

country_sales	6 obs. of 4 variables
data.df	1094 obs. of 7 variables
dispersions	1 obs. of 6 variables
product_sales	22 obs. of 4 variables
salesperson_sales	25 obs. of 4 variables
sample_data_10mins	1094 obs. of 6 variables
tendencies	1 obs. of 5 variables



assignment-1 - master - RStudio

	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

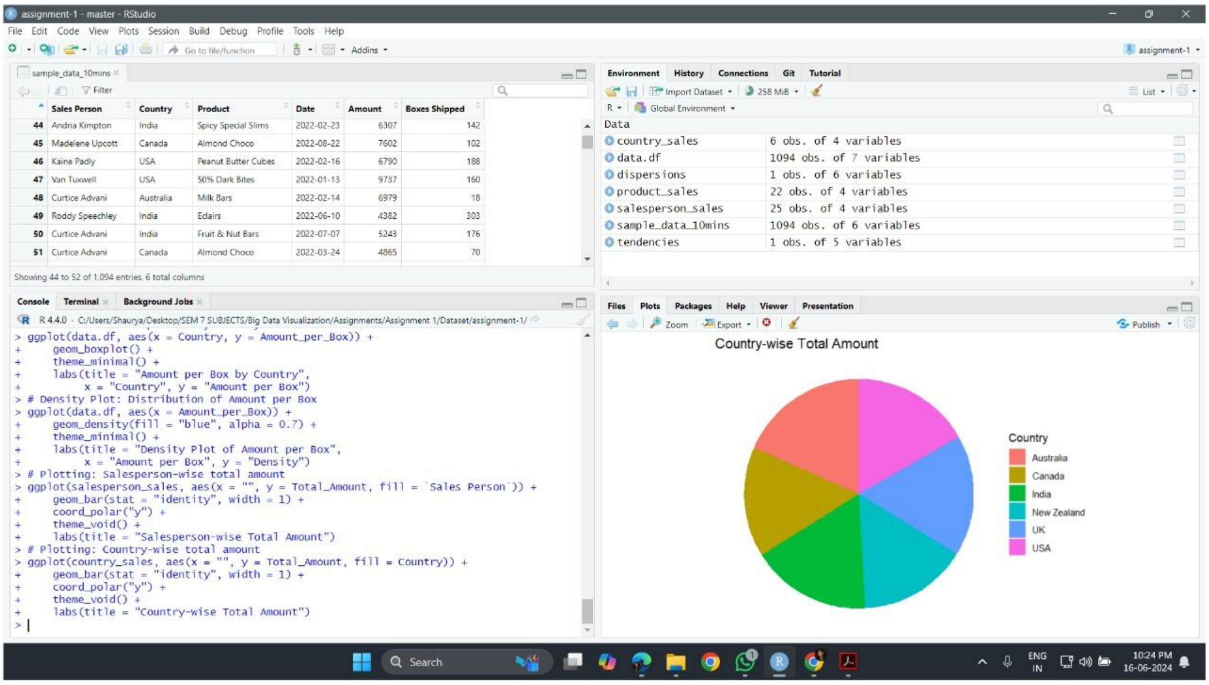
Showing 44 to 52 of 1,094 entries, 6 total columns

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> ggplot(data.df, aes(x = `Boxes Shipped`, y = Amount)) +
+     geom_point() +
+     theme_minimal() +
+     labs(title = "Amount vs. Boxes Shipped",
+          x = "Boxes Shipped", y = "Amount")
> # Box Plot: Amount per Box by Country
> ggplot(data.df, aes(x = Country, y = Amount_per_Box)) +
+     geom_boxplot() +
+     theme_minimal() +
+     labs(title = "Amount per Box by Country",
+          x = "Country", y = "Amount per Box")
> # Density Plot: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_density(fill = "blue", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Density Plot of Amount per Box",
+          x = "Amount per Box", y = "Density")
> # Plotting: Salesperson-wise total amount
> ggplot(salesperson_sales, aes(x = "", y = Total_Amount, fill = `Sales Person`)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Salesperson-wise Total Amount")
> 
```

Data

country_sales	6 obs. of 4 variables
data.df	1094 obs. of 7 variables
dispersions	1 obs. of 6 variables
product_sales	22 obs. of 4 variables
salesperson_sales	25 obs. of 4 variables
sample_data_10mins	1094 obs. of 6 variables
tendencies	1 obs. of 5 variables


assignment-1 - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
assignment-1
sample_data_10mins
Filter
Environment History Connections Git Tutorial
Import Dataset
258 MiB
List
Global Environment
Console Terminal Background Jobs
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Country-wise Total Amount
Country
🟥 Australia
🟨 Canada
🟩 India
New Zealand
🟦 UK
🟪 USA
Search
ENG IN
10:24 PM
16-06-2024

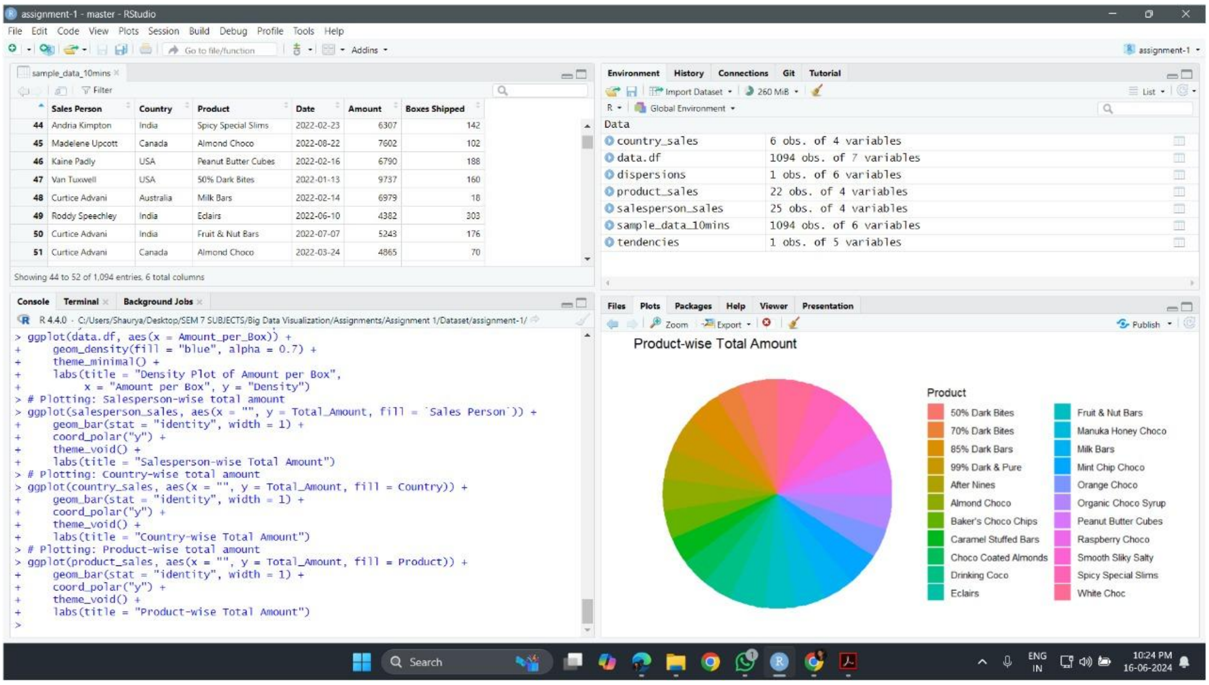
	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales         6 obs. of 4 variables
data.df               1094 obs. of 7 variables
dispersions           1 obs. of 6 variables
product_sales         22 obs. of 4 variables
salesperson_sales     25 obs. of 4 variables
sample_data_10mins    1094 obs. of 6 variables
tendencies            1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> ggplot(data.df, aes(x = Country, y = Amount_per_Box)) +
+     geom_boxplot() +
+     theme_minimal() +
+     labs(title = "Amount per Box by Country",
+          x = "Country", y = "Amount per Box")
> # Density Plot: Distribution of Amount per Box
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_density(fill = "blue", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Density Plot of Amount per Box",
+          x = "Amount per Box", y = "Density")
> # Plotting: Salesperson-wise total amount
> ggplot(salesperson_sales, aes(x = "", y = Total_Amount, fill = `Sales Person`)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Salesperson-wise Total Amount")
> # Plotting: Country-wise total amount
> ggplot(country_sales, aes(x = "", y = Total_Amount, fill = Country)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Country-wise Total Amount")
> 
```



	Sales Person	Country	Product	Date	Amount	Boxes Shipped
44	Andria Kimpton	India	Spicy Special Slims	2022-02-23	6307	142
45	Madelene Upcott	Canada	Almond Choco	2022-08-22	7602	102
46	Kaine Padly	USA	Peanut Butter Cubes	2022-02-16	6790	188
47	Van Tuxwell	USA	50% Dark Bites	2022-01-13	9737	160
48	Curtice Advani	Australia	Milk Bars	2022-02-14	6979	18
49	Roddy Speechley	India	Eclairs	2022-06-10	4382	303
50	Curtice Advani	India	Fruit & Nut Bars	2022-07-07	5243	176
51	Curtice Advani	Canada	Almond Choco	2022-03-24	4865	70

Showing 44 to 52 of 1,094 entries, 6 total columns

```
Data
country_sales         6 obs. of 4 variables
data.df               1094 obs. of 7 variables
dispersions           1 obs. of 6 variables
product_sales         22 obs. of 4 variables
salesperson_sales     25 obs. of 4 variables
sample_data_10mins    1094 obs. of 6 variables
tendencies            1 obs. of 5 variables
```

R 4.4.0 · C:/Users/Shaurya/Desktop/SEM 7 SUBJECTS/Big Data Visualization/Assignments/Assignment 1/Dataset/assignment-1/

```
> ggplot(data.df, aes(x = Amount_per_Box)) +
+     geom_density(fill = "blue", alpha = 0.7) +
+     theme_minimal() +
+     labs(title = "Density Plot of Amount per Box",
+          x = "Amount per Box", y = "Density")
> # Plotting: Salesperson-wise total amount
> ggplot(salesperson_sales, aes(x = "", y = Total_Amount, fill = `Sales Person`)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Salesperson-wise Total Amount")
> # Plotting: Country-wise total amount
> ggplot(country_sales, aes(x = "", y = Total_Amount, fill = Country)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Country-wise Total Amount")
> # Plotting: Product-wise total amount
> ggplot(product_sales, aes(x = "", y = Total_Amount, fill = Product)) +
+     geom_bar(stat = "identity", width = 1) +
+     coord_polar("y") +
+     theme_void() +
+     labs(title = "Product-wise Total Amount")
> 
```