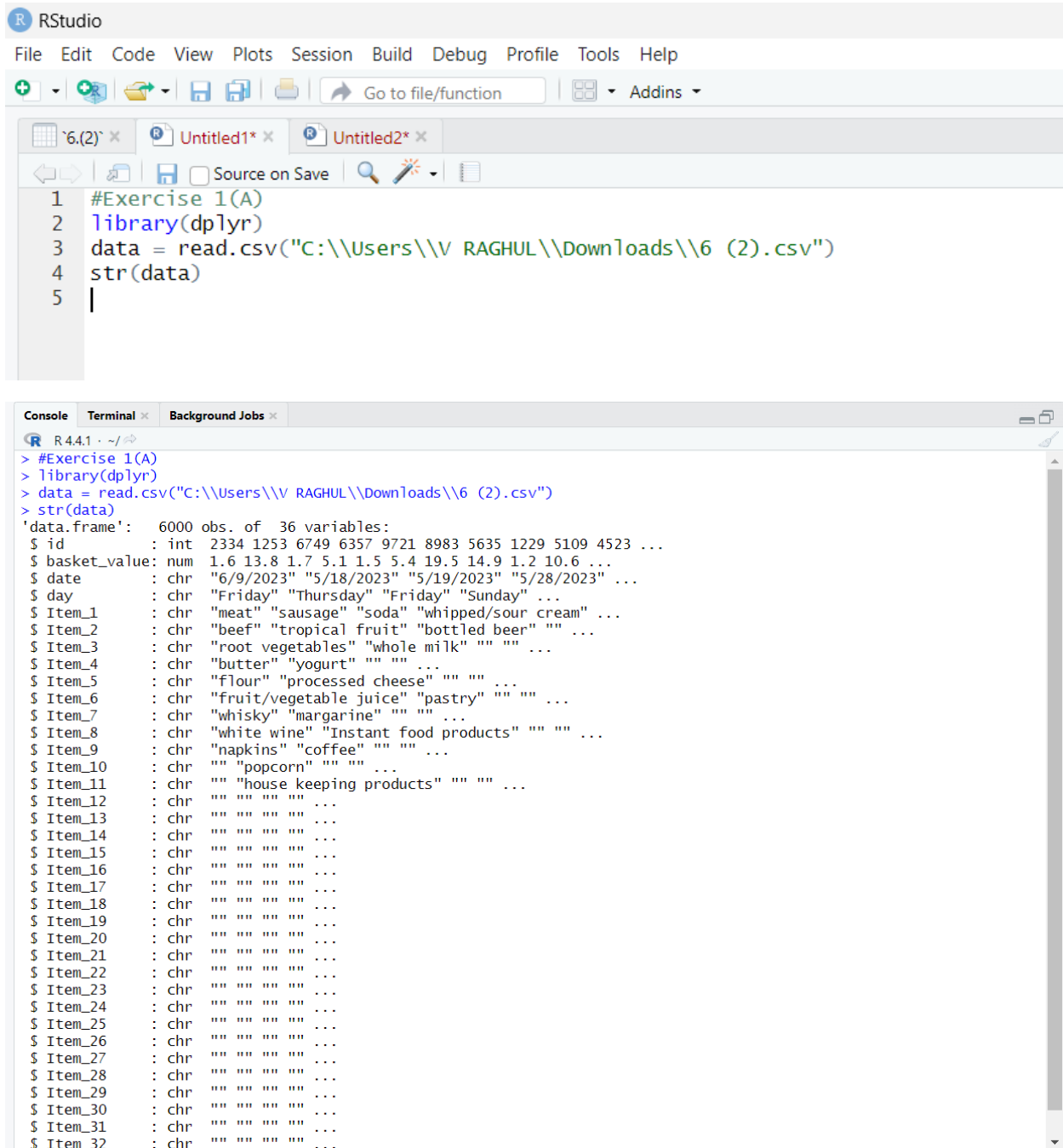


Homework-Group Assignment Week 9

Ragul Velmurugan, Brighton Matibe, Vijay Athithyaa, Venkatesh Prasad,
Harish Kumar

Exercise 1(A)



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
6.(2) x Untitled1* x Untitled2* x
Source on Save
1 #Exercise 1(A)
2 library(dplyr)
3 data = read.csv("C:\\Users\\V RAGHUL\\Downloads\\6 (2).csv")
4 str(data)
5 |

Console Terminal x Background Jobs x
R 4.4.1 ~ /
> #Exercise 1(A)
> library(dplyr)
> data = read.csv("C:\\Users\\V RAGHUL\\Downloads\\6 (2).csv")
> str(data)
'data.frame': 6000 obs. of 36 variables:
 $ id : int 2334 1253 6749 6357 9721 8983 5635 1229 5109 4523 ...
 $ basket_value: num 1.6 13.8 1.7 5.1 1.5 5.4 19.5 14.9 1.2 10.6 ...
 $ date : chr "6/9/2023" "5/18/2023" "5/19/2023" "5/28/2023" ...
 $ day : chr "Friday" "Thursday" "Friday" "Sunday" ...
 $ Item_1 : chr "meat" "sausage" "soda" "whipped/sour cream" ...
 $ Item_2 : chr "beef" "tropical fruit" "bottled beer" "" ...
 $ Item_3 : chr "root vegetables" "whole milk" "" "" ...
 $ Item_4 : chr "butter" "yogurt" "" "" ...
 $ Item_5 : chr "flour" "processed cheese" "" "" ...
 $ Item_6 : chr "fruit/vegetable juice" "pastry" "" "" ...
 $ Item_7 : chr "whisky" "margarine" "" "" ...
 $ Item_8 : chr "white wine" "Instant food products" "" "" ...
 $ Item_9 : chr "napkins" "coffee" "" "" ...
 $ Item_10 : chr "" "popcorn" "" "" ...
 $ Item_11 : chr "" "house keeping products" "" "" ...
 $ Item_12 : chr "" "" "" "" ...
 $ Item_13 : chr "" "" "" "" ...
 $ Item_14 : chr "" "" "" "" ...
 $ Item_15 : chr "" "" "" "" ...
 $ Item_16 : chr "" "" "" "" ...
 $ Item_17 : chr "" "" "" "" ...
 $ Item_18 : chr "" "" "" "" ...
 $ Item_19 : chr "" "" "" "" ...
 $ Item_20 : chr "" "" "" "" ...
 $ Item_21 : chr "" "" "" "" ...
 $ Item_22 : chr "" "" "" "" ...
 $ Item_23 : chr "" "" "" "" ...
 $ Item_24 : chr "" "" "" "" ...
 $ Item_25 : chr "" "" "" "" ...
 $ Item_26 : chr "" "" "" "" ...
 $ Item_27 : chr "" "" "" "" ...
 $ Item_28 : chr "" "" "" "" ...
 $ Item_29 : chr "" "" "" "" ...
 $ Item_30 : chr "" "" "" "" ...
 $ Item_31 : chr "" "" "" "" ...
 $ Item_32 : chr "" "" "" "" ...
```

Exercise 1(B)

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source on Save

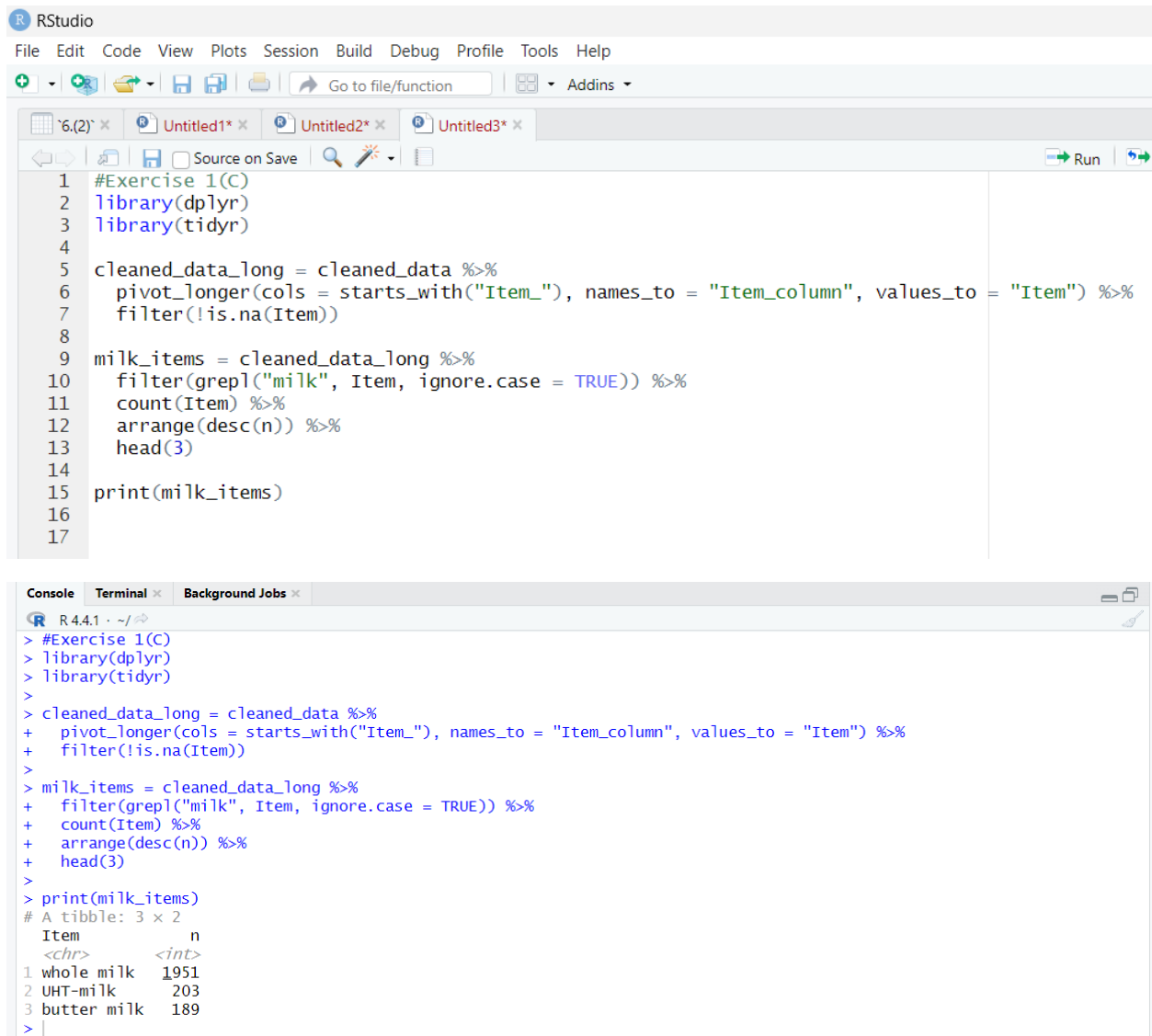
```
1 #Exercise 1(B)
2 cleaned_data = data %>%
3   filter_all(all_vars(!is.na(.))) %>%
4   distinct()
5 head(cleaned_data)
6 |
```

Console Terminal Background Jobs

R 4.4.1 · ~/

```
> #Exercise 1(B)
> cleaned_data = data %>%
+   filter_all(all_vars(!is.na(.))) %>%
+   distinct()
> head(cleaned_data)
  id basket_value   date    day      Item_1      Item_2      Item_3 Item_4
1 2334          1.6 6/9/2023 Friday      meat      beef root vegetables butter
2 1253         13.8 5/18/2023 Thursday  sausage tropical fruit  whole milk yogurt
3 6749          1.7 5/19/2023 Friday      soda  bottled beer
4 6357          5.1 5/28/2023 Sunday whipped/sour cream
5 9721          1.5 7/21/2023 Friday      yogurt
6 8983          5.4 5/23/2023 Tuesday citrus fruit  whole milk  bottled water  soda
  Item_5      Item_6      Item_7      Item_8 Item_9 Item_10
1      flour fruit/vegetable juice  whisky      white wine napkins
2 processed cheese      pastry margarine Instant food products  coffee popcorn
3
4
5
6 fruit/vegetable juice      bottled beer
  Item_11 Item_12 Item_13 Item_14 Item_15 Item_16 Item_17 Item_18 Item_19 Item_20 Item_21 Item_22
1
2 house keeping products
3
4
5
6
  Item_23 Item_24 Item_25 Item_26 Item_27 Item_28 Item_29 Item_30 Item_31 Item_32
1
2
3
4
5
6
> |
```

Exercise 1(C)



```
#Exercise 1(C)
library(dplyr)
library(tidyr)

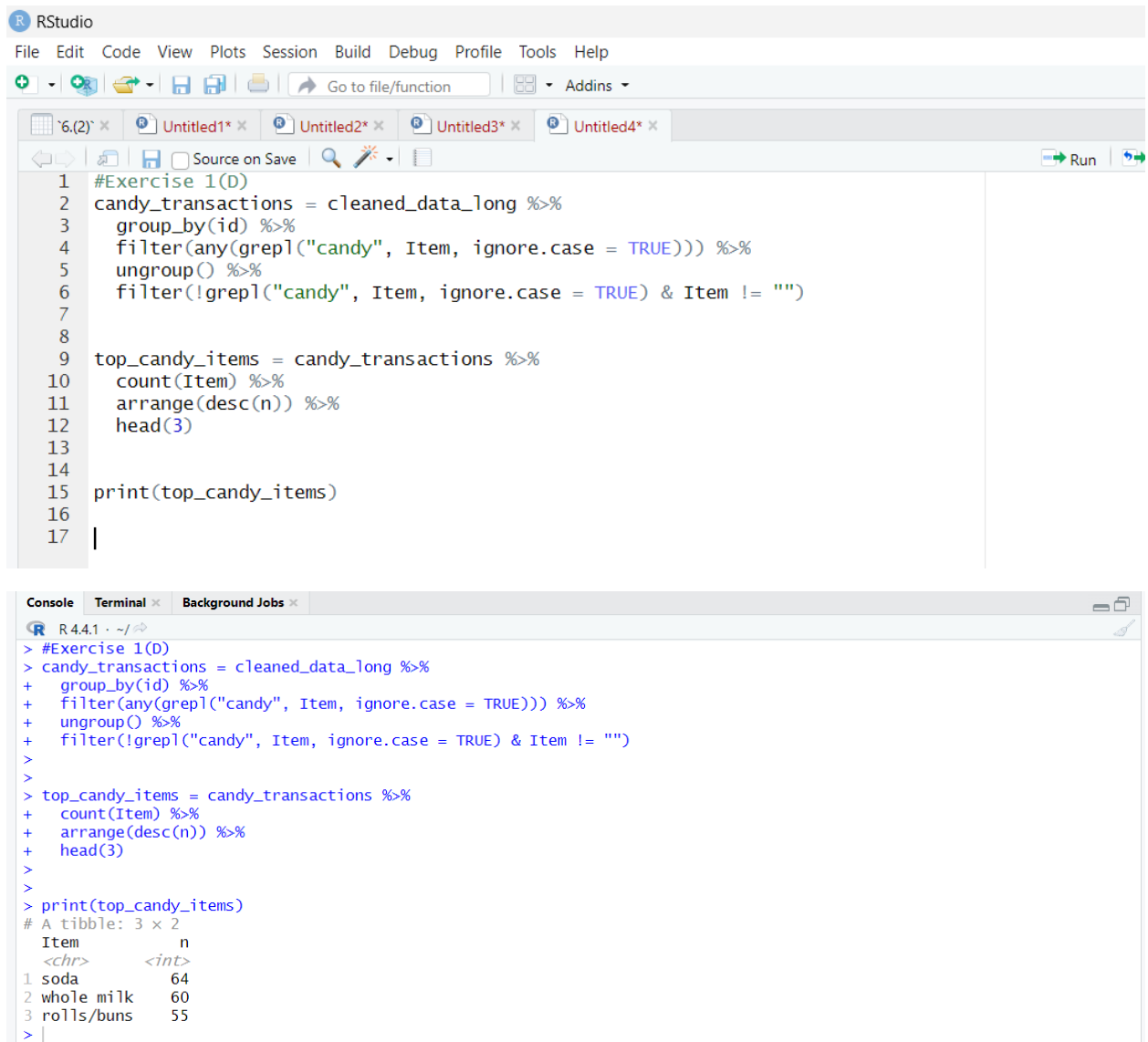
cleaned_data_long = cleaned_data %>%
  pivot_longer(cols = starts_with("Item_"), names_to = "Item_column", values_to = "Item") %>%
  filter(!is.na(Item))

milk_items = cleaned_data_long %>%
  filter(grepl("milk", Item, ignore.case = TRUE)) %>%
  count(Item) %>%
  arrange(desc(n)) %>%
  head(3)

print(milk_items)
```

```
R 4.4.1 ~ /
> #Exercise 1(C)
> library(dplyr)
> library(tidyr)
>
> cleaned_data_long = cleaned_data %>%
+   pivot_longer(cols = starts_with("Item_"), names_to = "Item_column", values_to = "Item") %>%
+   filter(!is.na(Item))
>
> milk_items = cleaned_data_long %>%
+   filter(grepl("milk", Item, ignore.case = TRUE)) %>%
+   count(Item) %>%
+   arrange(desc(n)) %>%
+   head(3)
>
> print(milk_items)
# A tibble: 3 x 2
  Item          n
<chr>      <int>
1 whole milk    1951
2 UHT-milk      203
3 butter milk   189
> |
```

Exercise 1(D)



```
1 #Exercise 1(D)
2 candy_transactions = cleaned_data_long %>%
3   group_by(id) %>%
4   filter(any(grepl("candy", Item, ignore.case = TRUE))) %>%
5   ungroup() %>%
6   filter(!grepl("candy", Item, ignore.case = TRUE) & Item != "")
7
8
9 top_candy_items = candy_transactions %>%
10  count(Item) %>%
11  arrange(desc(n)) %>%
12  head(3)
13
14
15 print(top_candy_items)
16
17 |
```

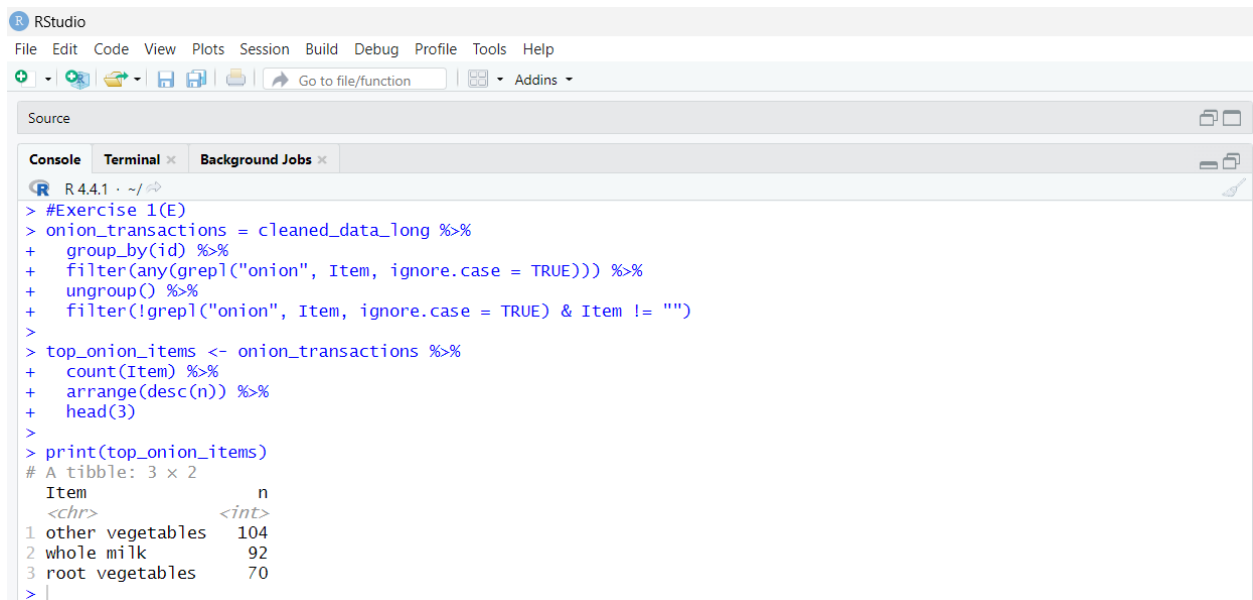
```
R 4.4.1 ~|
> #Exercise 1(D)
> candy_transactions = cleaned_data_long %>%
+   group_by(id) %>%
+   filter(any(grepl("candy", Item, ignore.case = TRUE))) %>%
+   ungroup() %>%
+   filter(!grepl("candy", Item, ignore.case = TRUE) & Item != "")
>
>
> top_candy_items = candy_transactions %>%
+   count(Item) %>%
+   arrange(desc(n)) %>%
+   head(3)
>
>
> print(top_candy_items)
# A tibble: 3 x 2
  Item      n
<chr> <int>
1 soda      64
2 whole milk 60
3 rolls/buns 55
> |
```

Exercise 1(E)



The screenshot shows the RStudio source editor. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for file operations and a search bar. The source editor shows the following R code:

```
1 #Exercise 1(E)
2 onion_transactions = cleaned_data_long %>%
3   group_by(id) %>%
4   filter(any(grepl("onion", Item, ignore.case = TRUE))) %>%
5   ungroup() %>%
6   filter(!grepl("onion", Item, ignore.case = TRUE) & Item != "")
7
8 top_onion_items <- onion_transactions %>%
9   count(Item) %>%
10  arrange(desc(n)) %>%
11  head(3)
12
13 print(top_onion_items)
14 |
```



The screenshot shows the RStudio console. The menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for file operations and a search bar. The console shows the following R code and its output:

```
> #Exercise 1(E)
> onion_transactions = cleaned_data_long %>%
+   group_by(id) %>%
+   filter(any(grepl("onion", Item, ignore.case = TRUE))) %>%
+   ungroup() %>%
+   filter(!grepl("onion", Item, ignore.case = TRUE) & Item != "")
>
> top_onion_items <- onion_transactions %>%
+   count(Item) %>%
+   arrange(desc(n)) %>%
+   head(3)
>
> print(top_onion_items)
# A tibble: 3 x 2
  Item          n
  <chr>      <int>
1 other vegetables 104
2 whole milk      92
3 root vegetables  70
> |
```

Exercise 1(F)

```
6.2) x  Untitled1* x  Untitled2* x  Untitled3* x  Untitled4* x  Untitled5* x  Untitled6* x
Source on Save  Run

1 #Exercise 1(F)
2 str(cleaned_data_long)
3 sample_data <- head(cleaned_data_long, 10)
4 print(sample_data)
5 top_items <- cleaned_data_long %>%
6   filter(Item != "") %>%
7   count(Item) %>%
8   arrange(desc(n)) %>%
9   head(3)
10 print(top_items)
11 related_turtle_items = cleaned_data_long %>%
12   filter(grepl("turt", Item, ignore.case = TRUE)) %>%
13   select(Item) %>%
14   distinct()
15
16 print(related_turtle_items)
17 turtle_transactions = cleaned_data_long %>%
18   filter(grepl("turtle", Item, ignore.case = TRUE)) %>%
19   group_by(id) %>%
20   ungroup() %>%
21   filter(Item != "") %>% |
22   filter(!grepl("turtle", Item, ignore.case = TRUE))
23
24 top_turtle_items = turtle_transactions %>%
25   count(Item) %>%
26   arrange(desc(n)) %>%
27   head(3)
28
29 print(top_turtle_items)
30
```

```
Console  Terminal x  Background Jobs x
R 4.4.1 ~ /
> #Exercise 1(F)
> str(cleaned_data_long)
tibble [182,240 x 6] (S3: tbl_df/tbl/data.frame)
 $ id      : int [1:182240] 2334 2334 2334 2334 2334 2334 2334 2334 2334 2334 ...
 $ basket_value: num [1:182240] 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 1.6 ...
 $ date      : chr [1:182240] "6/9/2023" "6/9/2023" "6/9/2023" "6/9/2023" ...
 $ day       : chr [1:182240] "Friday" "Friday" "Friday" "Friday" ...
 $ Item_column: chr [1:182240] "Item_1" "Item_2" "Item_3" "Item_4" ...
 $ Item      : chr [1:182240] "meat" "beef" "root vegetables" "butter" ...
> sample_data <- head(cleaned_data_long, 10)
> print(sample_data)
# A tibble: 10 x 6
   id basket_value date      day      Item_column Item
<int>      <dbl> <chr>    <chr>    <chr>    <chr>
1 2334      1.6 6/9/2023 Friday Item_1    "meat"
2 2334      1.6 6/9/2023 Friday Item_2    "beef"
3 2334      1.6 6/9/2023 Friday Item_3    "root vegetables"
4 2334      1.6 6/9/2023 Friday Item_4    "butter"
5 2334      1.6 6/9/2023 Friday Item_5    "flour"
6 2334      1.6 6/9/2023 Friday Item_6    "fruit/vegetable juice"
7 2334      1.6 6/9/2023 Friday Item_7    "whisky"
8 2334      1.6 6/9/2023 Friday Item_8    "white wine"
9 2334      1.6 6/9/2023 Friday Item_9    "napkins"
10 2334      1.6 6/9/2023 Friday Item_10   ""
> top_items <- cleaned_data_long %>%
+   filter(Item != "") %>%
+   count(Item) %>%
+   arrange(desc(n)) %>%
+   head(3)
> print(top_items)
# A tibble: 3 x 2
   Item      n
<chr>    <int>
1 whole milk 1951
2 other vegetables 1470
3 rolls/buns 1360
> related_turtle_items = cleaned_data_long %>%
+   filter(grepl("turt", Item, ignore.case = TRUE)) %>%
+   select(Item) %>%
+   distinct()
>
```

Exercise 1(G)

```
Source
Console Terminal Background Jobs
R 4.4.1 · ~/
> #Exercise 1(G)
> GetRecommendation = function(Cart_List, day, date) {
+   if (length(Cart_List) == 0) {
+     return("Cart_List is empty. Please add items to get recommendations.")
+   }
+
+   recommendations = cleaned_data_long %>%
+     filter(!Item %in% Cart_List & Item != "") %>%
+     filter(day == day & date == date) %>%
+     count(Item) %>%
+     arrange(desc(n)) %>%
+     head(3) %>%
+     pull(Item)
+
+   if (length(recommendations) == 0) {
+     return("No recommendations available based on the current cart.")
+   }
+
+   return(recommendations)
+ }
> |
```

Exercise 1(H)

```
6(2) x Untitled1 x Untitled2 x Untitled3 x Untitled4 x Untitled5 x Untitled6 x Untitled7 x Untitled8 x
Source on Save Run
1 #Exercise 1(H)
2 recommend_list1 <- GetRecommendation(c("tropical fruit", "sliced cheese", "rolls/buns"), "Friday", "2023-07-21")
3 recommend_list2 <- GetRecommendation(c("whipped/sour cream", "brown bread"), "Tuesday", "2023-05-30")
4 print(recommend_list1)
5 print(recommend_list2)
6 |
```

```
Console Terminal Background Jobs
R 4.4.1 · ~/
> #Exercise 1(H)
> # Get recommendation list for the first set of inputs
> recommend_list1 <- GetRecommendation(c("tropical fruit", "sliced cheese", "rolls/buns"), "Friday", "2023-07-21")
>
> # Get recommendation list for the second set of inputs
> recommend_list2 <- GetRecommendation(c("whipped/sour cream", "brown bread"), "Tuesday", "2023-05-30")
>
> # Display the results
> print(recommend_list1)
[1] "whole milk"      "other vegetables"
[3] "soda"
> print(recommend_list2)
[1] "whole milk"      "other vegetables"
[3] "rolls/buns"
> |
```