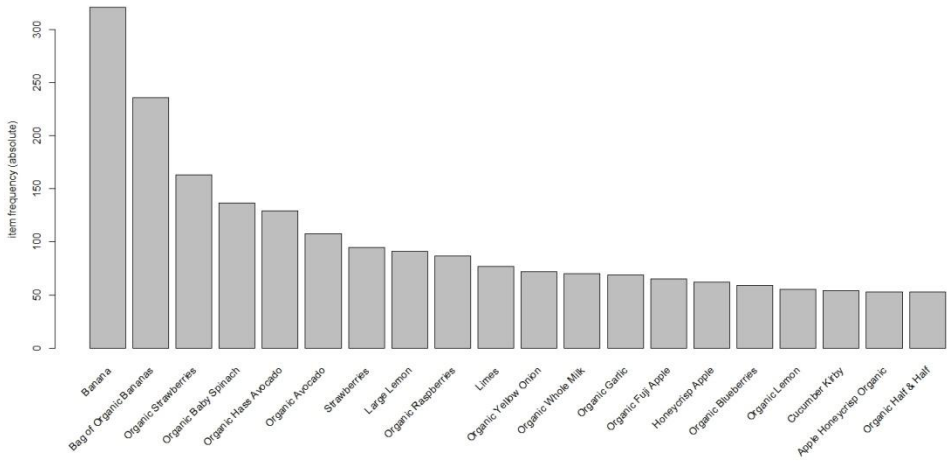
Sriprasath Gujuluva Parthasarathy (sxg180154)

InstaCart Market Basket Analysis:

Question 1: Frequent itemsets for products in orders dataset. You have to output product names and not just product id:

Item Frequency plot:



Hence it seems that Banana has the highest item frequency from the taken dataset.

Question 2: Association rules for products in orders dataset. You have to output product names and not just product id

Association Rules:

```
> inspect(head(rules,5))
    lhs                                          rhs             support confidence lift count
[1] {Sliced Peaches}                          => {Banana}        0.0015  1           6.3 3
[2] {Chunk Light Tuna}                        => {Strawberries}  0.0015  1          21.4 3
[3] {Apple Cider Vinegar}                     => {Banana}        0.0015  1           6.3 3
[4] {Cheddar Snack Crackers Cheddar Bunnies} => {Banana}        0.0015  1           6.3 3
[5] {Coho Salmon}                             => {Banana}        0.0015  1           6.3 3
```

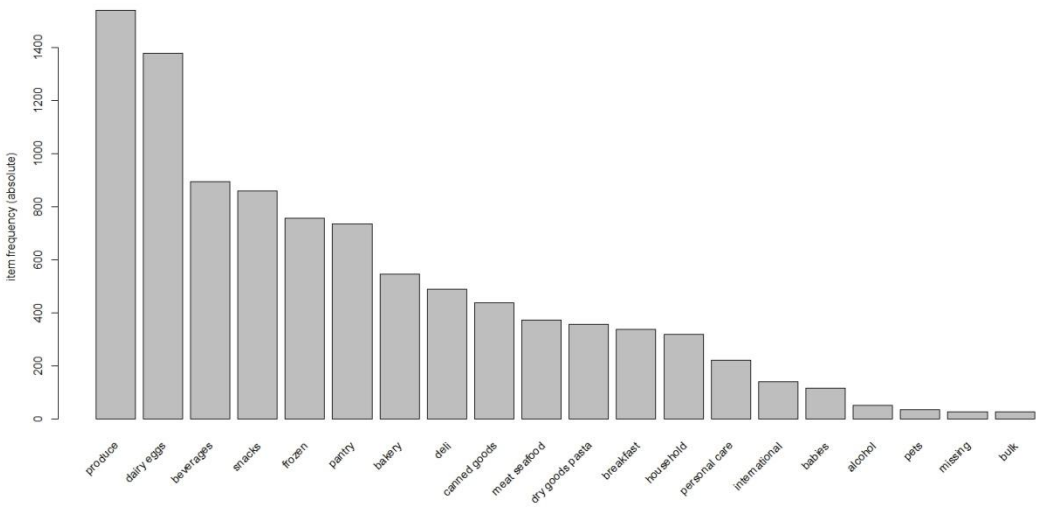
Association Rules after sorting by lift:

```
> rules<-sort(rules, by="lift", decreasing=TRUE)
> inspect(head(rules,5))
    lhs                                          rhs                                                support confidence lift count
[1] {Greek Whole Milk Blended Blueberry Yogurt}          => {Organic Greek Whole Milk Blended Strawberry Yogurt}  0.0015          1  677     3
[2] {Organic Greek Whole Milk Blended Strawberry Yogurt} => {Greek Whole Milk Blended Blueberry Yogurt}           0.0015          1  677     3
[3] {Greek Whole Milk Blended Blueberry Yogurt}          => {Organic Blended Raspberry Whole Milk Greek Yogurt}   0.0015          1  677     3
[4] {Organic Blended Raspberry Whole Milk Greek Yogurt}  => {Greek Whole Milk Blended Blueberry Yogurt}           0.0015          1  677     3
[5] {Organic Greek Whole Milk Blended Strawberry Yogurt} => {Organic Blended Raspberry Whole Milk Greek Yogurt}   0.0015          1  677     3
```

So from above it is clear that the Association rule {Greek Whole Milk Blended Blueberry Yogurt} => {Organic Greek Whole Milk Blended Strawberry Yogurt} has the highest lift value in the taken dataset. The support and confidence value used is 0.001 and 0.8 respectively.

3. Frequent itemsets for departments in orders dataset (i.e which departments have highest number of orders). You have to output department names and not just department id

Item Frequency plot:



So from above it is clear that produce department has the highest frequency in the taken dataset.

4. Association rules for departments in orders dataset (e.g. frozen -> groceries). You have to output department names and not just department id:

Association Rules:

```
> inspect(head(rules,5))
    lhs           rhs          support confidence lift count
[1] {other}   => {dairy eggs} 0.0089  0.86       1.3  18
[2] {bulk}    => {dairy eggs} 0.0098  0.80       1.2  20
[3] {bulk}    => {produce}    0.0118  0.96       1.3  24
[4] {missing} => {dairy eggs} 0.0108  0.81       1.2  22
[5] {missing} => {produce}    0.0108  0.81       1.1  22
```

Association Rules after sorting by lift:

```
> rules<-sort(rules, by="lift", decreasing=TRUE)
> inspect(head(rules,5))
    lhs                                                                rhs      support confidence lift count
[1] {canned goods,deli,dry goods pasta,frozen,household,meat seafood}          => {babies} 0.0015  1          18   3
[2] {canned goods,dry goods pasta,frozen,household,meat seafood,pantry}        => {babies} 0.0015  1          18   3
[3] {canned goods,dry goods pasta,household,meat seafood,pantry,snacks}        => {babies} 0.0020  1          18   4
[4] {canned goods,deli,dry goods pasta,frozen,household,meat seafood,snacks}   => {babies} 0.0015  1          18   3
[5] {canned goods,dairy eggs,deli,dry goods pasta,frozen,household,meat seafood} => {babies} 0.0015  1          18   3
```

From above it is clear that the association rule {canned goods,deli,dry goods pasta, frozen, household, meat seafood} => {babies} has the highest lift value in the taken dataset. The support and confidence value used is 0.001 and 0.8 respectively.