Title: Impact of Food Security due to COVID-19

Objective: Exploring the dataset, finding out the insights from the data and solving the problems

In [2]:
```
data=read.csv("FOOD PRICE MONITORING EXPORT INDIA DATA UPDATED 2.csv")
data
```

FAQ) daily prices in the market Tamluk (Medinipur E)	11552	03/13/20	2500	2400	2600	USD/Ton	rice
India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	03/12/20	2500	2400	2600	USD/Ton	rice
India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	03/11/20	2500	2400	2600	USD/Ton	rice
India Rice (Common							

In [8]:
```
#Structure of the dataset
str(data)
```

```
'data.frame':   158420 obs. of  19 variables:
 $ price.series               : Factor w/ 4986 levels "India Apple (American, L
arge)  daily prices  in the market Sangrur",..: 3371 3371 3371 3371 3371 3371 3
371 3371 3371 3371 3371 ...
 $ priceTypeID                : int  11552 11552 11552 11552 11552 11552 11552
11552 11552 11552 11552 ...
 $ date                       : Date, format: "2020-03-20" "2020-03-19" ...
 $ price                      : num  2500 2500 2500 2500 2500 2500 2500 2500 25
00 2500 ...
 $ minPrice                   : num  2400 2400 2400 2400 2400 2400 2400 2400 24
00 2400 ...
 $ maxPrice                   : num  2600 2600 2600 2600 2600 2600 2600 2600 26
00 2600 ...
 $ unit                       : Factor w/ 1 level "USD/Ton": 1 1 1 1 1 1 1 1 1
1 ...
 $ commodity                  : Factor w/ 21 levels "apple","banana",..: 17 17
17 17 17 17 17 17 17 17 ...
 $ CountryCode                : Factor w/ 1 level "IND": 1 1 1 1 1 1 1 1 1 1
...
 $ market                     : Factor w/ 1335 levels "A lot","Abohar",..: 1241
1241 1241 1241 1241 1241 1241 1241 1241 1241 ...
 $ marketLevel                : Factor w/ 1 level "Wholesale": 1 1 1 1 1 1 1 1
1 1 ...
 $ lat                        : num  22.3 22.3 22.3 22.3 22.3 ...
 $ long                       : num  87.9 87.9 87.9 87.9 87.9 ...
 $ Country                    : Factor w/ 1 level "India": 1 1 1 1 1 1 1 1 1 1
...
 $ PercentageChange           : num  0 0 0 0 0 0 0 0 0 0 ...
 $ percentageChangeAlert      : int  0 0 0 0 0 0 0 0 0 0 ...
 $ percentageChange95Threshold: int  15 15 15 15 15 15 15 15 15 15 ...
 $ percentageChange90Threshold: int  5 5 5 5 5 5 5 5 5 5 ...
 $ district                   : Factor w/ 389 levels "","Agra","Ahmedabad",..:
245 245 245 245 245 245 245 245 245 245 ...
```

In [9]:
```
# Summary of dataset
summary(data)
```

```
                                                                    price.s
eries
 India Banana (Banana - Ripe, Medium)  daily prices  in the market Wazirganj:
159
 India Tomato (Other, FAQ)  daily prices  in the market Sealdah Koley Market:
159
 India Banana (Medium, Medium)  daily prices  in the market Muradabad       :
156
 India Banana (Banana - Ripe, Medium)  daily prices  in the market Jafarganj:
153
 India Banana (Banana - Ripe, Medium)  daily prices  in the market Puranpur :
151
 India Banana (Nendra Bale, Medium)  daily prices  in the market Kayamkulam :
151
 (Other)                                                                    :
157491
  priceTypeID         date                price          minPrice
 Min.   :11552   Min.   :2019-12-01   Min.   :  100   Min.   :    0
 1st Qu.:12567   1st Qu.:2020-02-19   1st Qu.: 1300   1st Qu.: 1200
 Median :13184   Median :2020-03-30   Median : 1925   Median : 1800
 Mean   :13639   Mean   :2020-03-19   Mean   : 2666   Mean   : 2474
 3rd Qu.:14309   3rd Qu.:2020-04-27   3rd Qu.: 3200   3rd Qu.: 3000
 Max.   :18693   Max.   :2020-05-27   Max.   :50000   Max.   :45000

    maxPrice          unit          commodity     CountryCode
 Min.   :    0   USD/Ton:158420   banana :35305   IND:158420
 1st Qu.: 1400                    potato :29608
 Median : 2000                    tomato :28029
 Mean   : 2846                    apple  :25620
 3rd Qu.: 3500                    onion  :14199
 Max.   :60000                    wheat  : 8906
                                  (Other):16753
                          market          marketLevel          lat
 Binny Mill (F&V), Bangalore:   826   Wholesale:158420   Min.   : 8.343
 Anchal                     :   821                      1st Qu.:22.036
 Palakkad                   :   756                      Median :26.561
 Thodupuzha                 :   715                      Mean   :24.078
 Chala                      :   674                      3rd Qu.:28.625
 Kayamkulam                 :   659                      Max.   :34.104
 (Other)                    :153969
     long        Country       PercentageChange    percentageChangeAlert
 Min.   :69.63   India:158420   Min.   : -94.6700   Min.   :-1.0000
 1st Qu.:76.53                  1st Qu.:  -1.2700   1st Qu.:-1.0000
 Median :78.01                  Median :   0.0000   Median : 0.0000
 Mean   :79.17                  Mean   :   0.9442   Mean   : 0.1416
 3rd Qu.:80.77                  3rd Qu.:   1.0100   3rd Qu.: 1.0000
 Max.   :94.82                  Max.   :1775.0000   Max.   : 2.0000

 percentageChange95Threshold percentageChange90Threshold         district
 Min.   :15                  Min.   :5                   Bangalore   :  2754
 1st Qu.:15                  1st Qu.:5                   Agra        :  2515
 Median :15                  Median :5                   Kottayam    :  2503
 Mean   :15                  Mean   :5                   Bulandshahar:  2428
```

```
 3rd Qu.:15          3rd Qu.:5          Alappuzha   :  2198
 Max.   :15          Max.   :5          Malappuram  :  2024
                                        (Other)     :143998
```

We are splitting the above dataset into into parts. The splitting is done according to the date when the lockdown was initialised by the Government of India (i.e. 16/03/2020).

In [6]:
```
# Converting to date format
data$date=as.Date(data$date,"%m/%d/%y")
```

Before intialising Lockdown

In [7]:
```
before_data=subset(data, date <= as.Date("2020-03-15"))
before_data
```

	price.series	priceTypeID	date	price	minPrice	maxPrice	unit	commodity
6	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-15	2500	2400	2600	USD/Ton	rice
7	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-14	2500	2400	2600	USD/Ton	rice
8	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-13	2500	2400	2600	USD/Ton	rice
	India Rice (Common, FAQ)		2020-					

In [13]:
```
# Summary of before lockdown dataset
summary(before_data)
```

```
                                                                                price.s
eries
 India Apple (Delicious, Medium)  daily prices  in the market Wazirganj     :
104
 India Banana (Banana - Ripe, Medium)  daily prices  in the market Wazirganj:
104
 India Apple (Delicious, Medium)  daily prices  in the market Tanda         :
103
 India Apple (Simla, Medium)  daily prices  in the market Purwa             :
103
 India Tomato (Other, FAQ)  daily prices  in the market Sealdah Koley Market:
103
 India Apple (Delicious, Medium)  daily prices  in the market Jafarganj     :
101
 (Other)                                                                    :
63623
  priceTypeID         date                price          minPrice
 Min.   :11552   Min.   :2019-12-01   Min.   :  150   Min.   :    0
 1st Qu.:12584   1st Qu.:2020-01-02   1st Qu.: 1670   1st Qu.: 1500
 Median :12928   Median :2020-02-04   Median : 2650   Median : 2500
 Mean   :13303   Mean   :2020-01-31   Mean   : 3374   Mean   : 3122
 3rd Qu.:13313   3rd Qu.:2020-03-04   3rd Qu.: 5000   3rd Qu.: 4650
 Max.   :18499   Max.   :2020-03-15   Max.   :19000   Max.   :18000

    maxPrice          unit              commodity     CountryCode
 Min.   :    0   USD/Ton:64241   banana      :23451   IND:64241
 1st Qu.: 1800                   apple       :18692
 Median : 3000                   potato      : 6341
 Mean   : 3609                   tomato      : 5994
 3rd Qu.: 5290                   french beans: 3220
 Max.   :20000                   beans       : 2466
                                 (Other)     : 4077
                         market         marketLevel         lat
 Chala                      :  589   Wholesale:64241   Min.   : 8.343
 Thodupuzha                 :  546                     1st Qu.:18.563
 Binny Mill (F&V), Bangalore:  527                     Median :26.476
 Mechua                     :  469                     Mean   :23.178
 Kattappana                 :  465                     3rd Qu.:28.724
 Anchal                     :  398                     Max.   :34.104
 (Other)                    :61247
     long        Country      PercentageChange   percentageChangeAlert
 Min.   :69.63   India:64241   Min.   :-90.5300   Min.   :-1.0000
 1st Qu.:76.39                 1st Qu.: -0.4900   1st Qu.:-1.0000
 Median :77.57                 Median :  0.0000   Median :-1.0000
 Mean   :78.74                 Mean   :  0.8641   Mean   :-0.1343
 3rd Qu.:80.13                 3rd Qu.:  0.8500   3rd Qu.: 0.0000
 Max.   :94.82                 Max.   :900.0000   Max.   : 2.0000

 percentageChange95Threshold percentageChange90Threshold
 Min.   :15                  Min.   :5
 1st Qu.:15                  1st Qu.:5
 Median :15                  Median :5
 Mean   :15                  Mean   :5
```

```
3rd Qu.:15                3rd Qu.:5
Max.   :15                Max.   :5

             district
Bangalore         : 1587
Thiruvananthapuram: 1485
Idukki            : 1337
Kottayam          : 1276
Alappuzha         : 1247
Kollam            : 1133
(Other)           :56176
```

During Lockdown

In [8]:
```
after_data=subset(data, date > as.Date("2020-03-15"))
after_data
```

	market Tamluk (Medinipur E)							
2	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-19	2500	2400	2600	USD/Ton	rice
3	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-18	2500	2400	2600	USD/Ton	rice
4	India Rice (Common, FAQ) daily prices in the market Tamluk (Medinipur E)	11552	2020-03-17	2500	2400	2600	USD/Ton	rice

In [12]:
```
# Summary of during lockdown dataset
summary(after_data)
```

```
                                                                    price.ser
ies
 India Potato (Jyoti, FAQ)  daily prices  in the market Sheoraphuly       :
63
 India Potato (Jyoti, FAQ)  daily prices  in the market Bishnupur(Bankura):
62
 India Potato (Jyoti, FAQ)  daily prices  in the market Khatra            :
62
 India Potato (Jyoti, FAQ)  daily prices  in the market Panisagar         :
62
 India Potato (Potato, FAQ)  daily prices  in the market Hindol           :
62
 India Potato (Red, FAQ)  daily prices  in the market Gadaura             :
62
 (Other)                                                                  :93
806
  priceTypeID         date                price          minPrice
 Min.   :11552   Min.   :2020-03-16   Min.   :  100   Min.   :    0
 1st Qu.:12507   1st Qu.:2020-04-06   1st Qu.: 1200   1st Qu.: 1080
 Median :13760   Median :2020-04-21   Median : 1625   Median : 1500
 Mean   :13869   Mean   :2020-04-21   Mean   : 2184   Mean   : 2032
 3rd Qu.:15224   3rd Qu.:2020-05-12   3rd Qu.: 2200   3rd Qu.: 2000
 Max.   :18693   Max.   :2020-05-27   Max.   :50000   Max.   :45000

    maxPrice          unit         commodity     CountryCode
 Min.   :    0   USD/Ton:94179   potato :23267   IND:94179
 1st Qu.: 1300                   tomato :22035
 Median : 1750                   onion  :13725
 Mean   : 2326                   banana :11854
 3rd Qu.: 2400                   wheat  : 8751
 Max.   :60000                   apple  : 6928
                                 (Other): 7619
           market         marketLevel         lat              long
 Anchal       :  423   Wholesale:94179   Min.   : 8.343   Min.   :69.63
 Palakkad     :  368                     1st Qu.:23.049   1st Qu.:76.64
 Fatehabad    :  359                     Median :26.716   Median :78.28
 Kayamkulam   :  341                     Mean   :24.692   Mean   :79.47
 Muzzafarnagar:  321                     3rd Qu.:28.590   3rd Qu.:81.28
 Haathras     :  316                     Max.   :34.104   Max.   :94.82
 (Other)      :92051
  Country      PercentageChange    percentageChangeAlert
 India:94179   Min.   : -94.6700   Min.   :-1.0000
               1st Qu.:  -2.0000   1st Qu.:-1.0000
               Median :   0.0000   Median : 0.0000
               Mean   :   0.9987   Mean   : 0.3298
               3rd Qu.:   1.2700   3rd Qu.: 2.0000
               Max.   :1775.0000   Max.   : 2.0000

 percentageChange95Threshold percentageChange90Threshold          district
 Min.   :15                  Min.   :5                   Agra         : 1494
 1st Qu.:15                  1st Qu.:5                   Bulandshahar : 1458
 Median :15                  Median :5                   Muzaffarnagar: 1254
 Mean   :15                  Mean   :5                   Kottayam     : 1227
```

```
3rd Qu.:15          3rd Qu.:5          Kanpur       : 1204
Max.   :15          Max.   :5          Bangalore    : 1167
                                       (Other)      :86375
```

>Maximum commidty purchased:

```
Before lockdown: Banana
During lockdown: Potato
```

>Price Range(min,max):

```
Before lockdown: 150,19000
During Lockdown: 100,50000
```

Analyzing of the data

Q1. Maximum commodity purchased before and during lockdown.

From the summary of before_data(before lockdown) dataset the maximum commodity purchased was banana

In [5]:
```
banana_ds=subset(before_data,commodity=="banana")
banana_ds
```

	price.series	priceTypeID	date	price	minPrice	maxPrice	unit	commodity	Cou
52477	India Banana (Banana - Ripe, Medium) daily prices in the market Rajnandgaon	12717	2020-03-15	2500	2500	2500	USD/Ton	banana	
52478	India Banana (Banana - Ripe, Medium) daily prices in the market Rajnandgaon	12717	2020-03-14	2500	2500	2500	USD/Ton	banana	
52479	India Banana (Banana - Ripe, Medium)	12717	2020-03-13	2500	2500	2500	USD/Ton	banana	

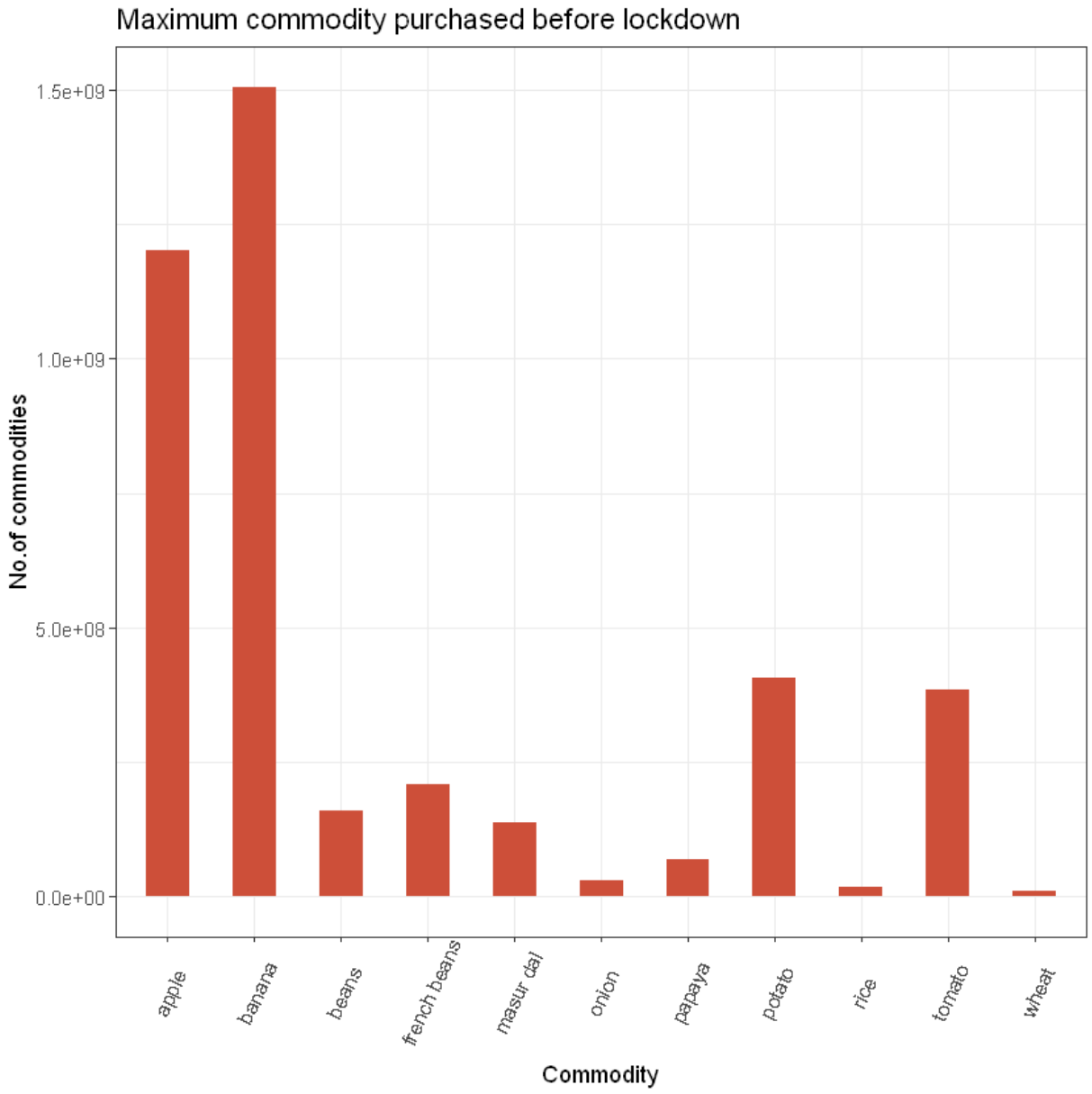
In [6]:
```
nrow(banana_ds)
```

23451

In [12]:
```
library(ggplot2)
```

In [73]:
```
library(scales)
theme_set(theme_bw())
ggplot(before_data, aes(x=commodity,y=nrow(before_data))) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(title="Maximum commodity purchased before lockdown") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))+xlab("Commodity")+ylab('
```

Warning message:
"package 'scales' was built under R version 3.6.3"



From the summary of after_data(during lockdown) dataset the maximum commodity purchased was potato.

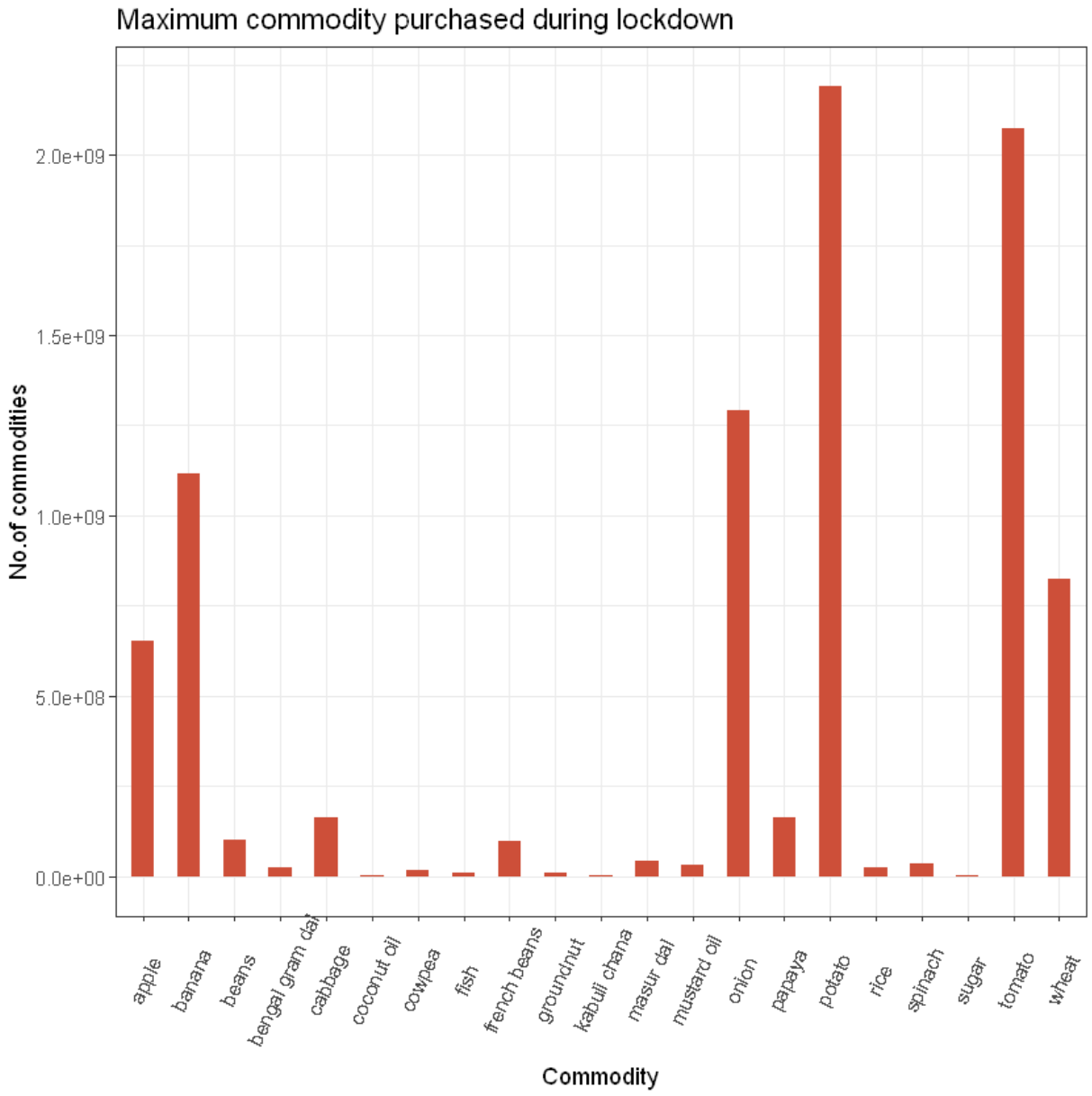
In [17]:
```
potato_ds=subset(after_data,commodity=="potato")
potato_ds
```

	price.series	priceTypeID	date	price	minPrice	maxPrice	unit	commodity	Co
91862	India Potato (Other, FAQ) daily prices in the market Durg	13538	2020-04-12	1850	1800	1900	USD/Ton	potato	
91863	India Potato (Other, FAQ) daily prices in the market Durg	13538	2020-04-11	1800	1600	2000	USD/Ton	potato	
91864	India Potato (Other, FAQ) daily prices in the market Durg	13538	2020-04-10	1900	1800	2000	USD/Ton	potato	
	India Potato (Other, FAQ)		2020-						

In [18]:
```
nrow(potato_ds)
```

23267

In [24]:
```
ggplot(after_data, aes(x=commodity,y=nrow(after_data))) +
  geom_bar(stat="identity", width=.5, fill="tomato3") +
  labs(title="Maximum commodity purchased during lockdown") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))+xlab("Commodity")+ylab('
```



>The maximum commodity purchased before lockdown was banana and during lockdown was potato.

Q2. What was the difference in the prices of food before and during lockdown?

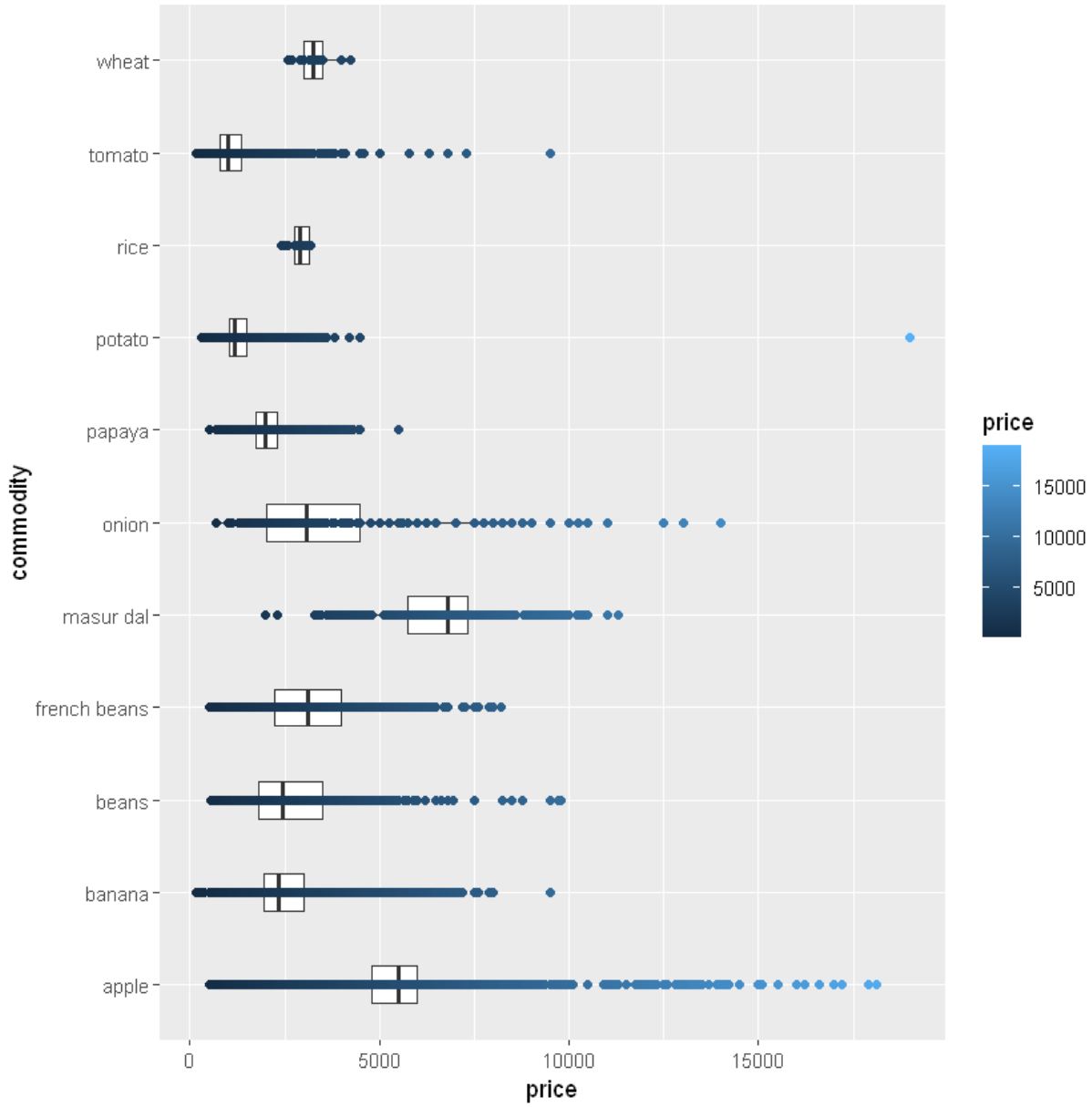
In [25]:
```
#maximum price before lockdown
max(before_data$price)

#maximum price after lockdown
max(after_data$price)
```

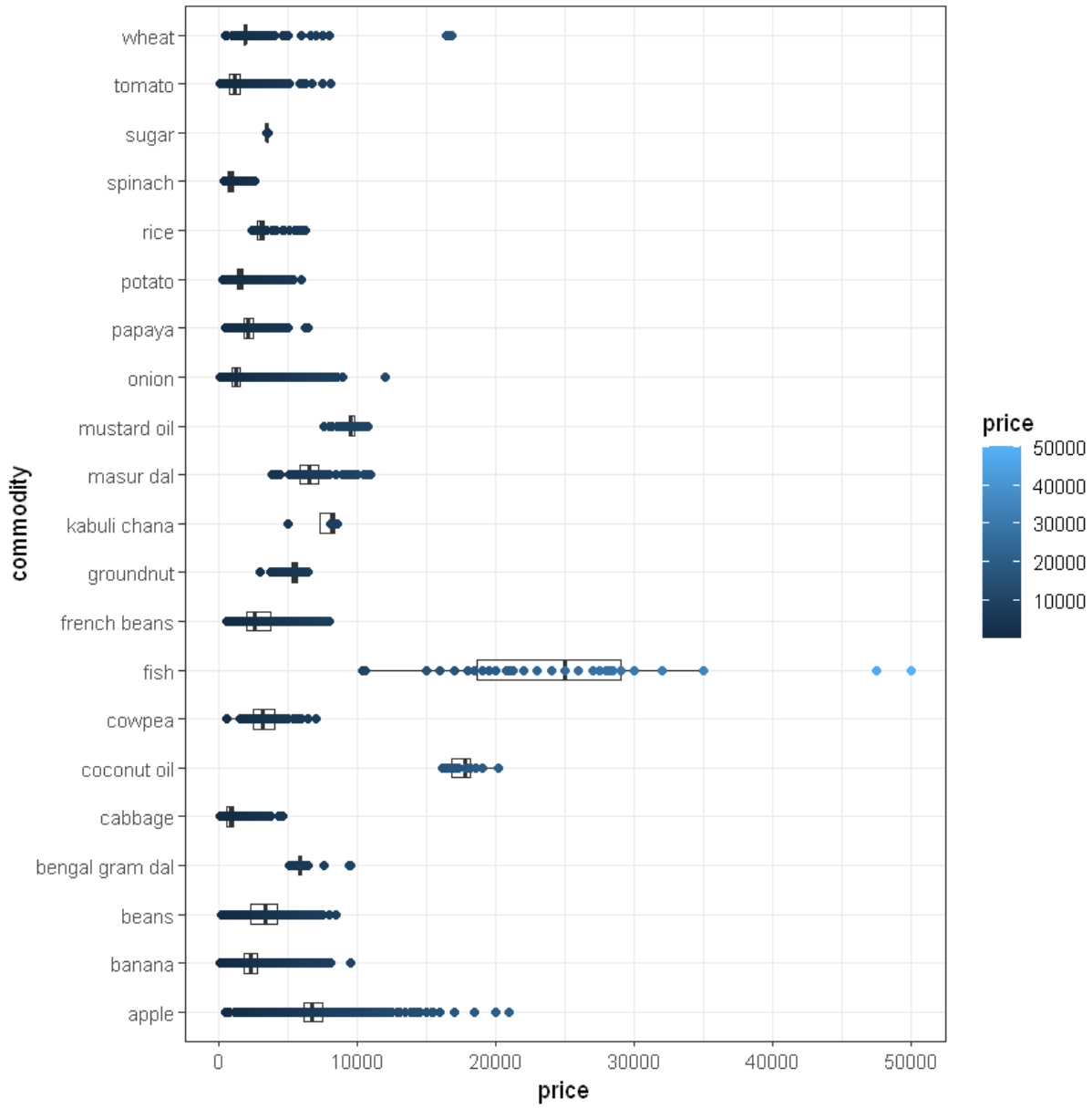
19000

50000

In [38]:
```
ggplot(before_data,aes(x=price,y=commodity))+geom_boxplot(width=0.4)+geom_point(
```



In [28]:
```
ggplot(after_data,aes(x=price,y=commodity))+geom_boxplot(width=0.4)+geom_point(a
```



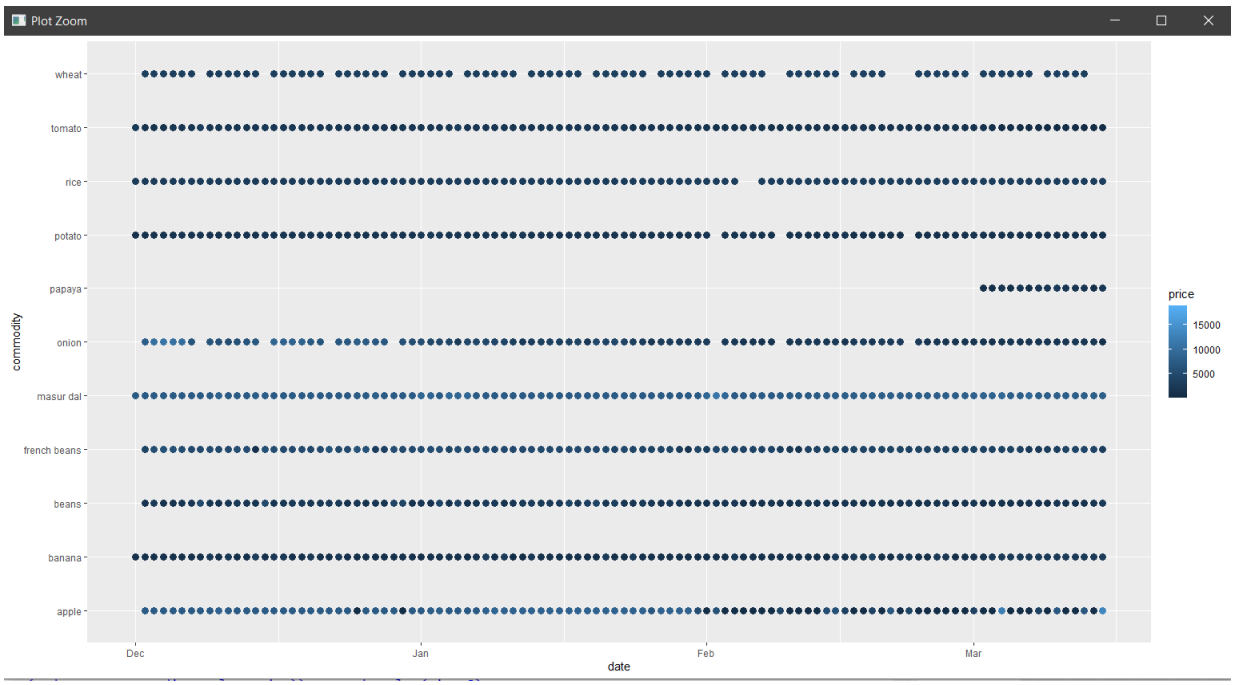
After comparing above two boxplots

>the spread of data of onion in before lockdown graph is more , so that the variation in the price is more and in after lockdown graph the variation was low.
>During lockdown graph , the variation in the price of fish was more but before lockdown fish was less (data was inelig ible)
>This happened due to the increase and decrease of the supply chain

Q3. Price distributions of commodity in different months.

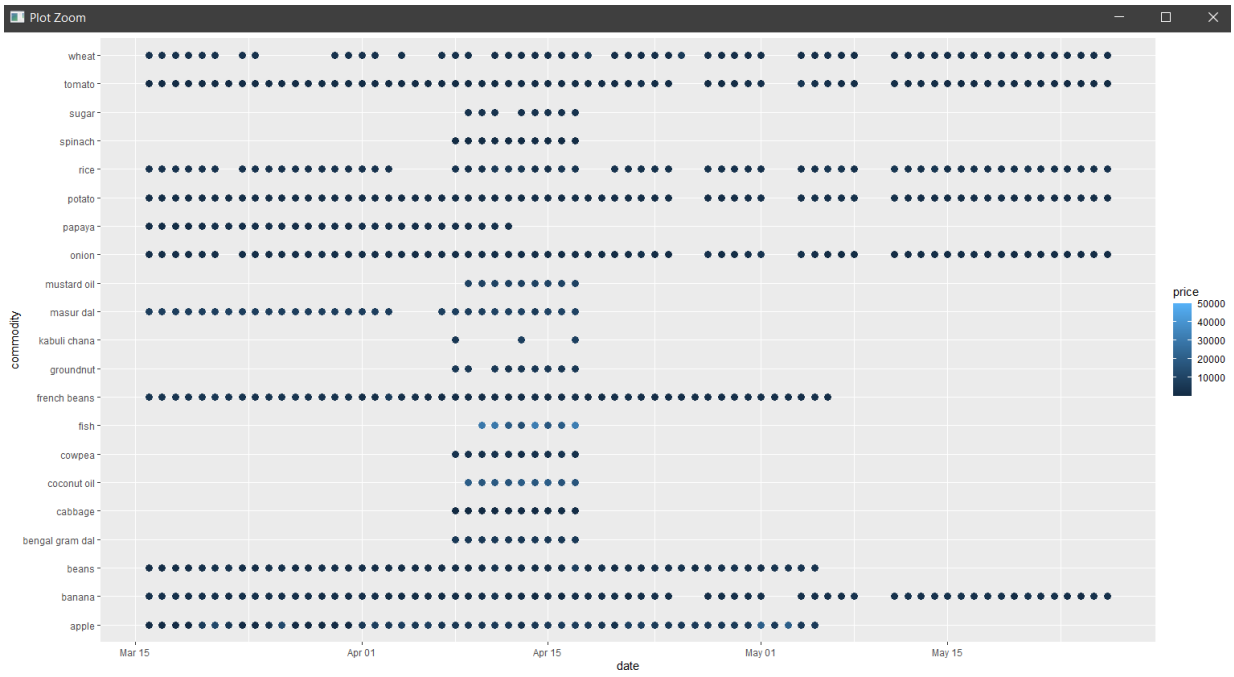
In [ ]:
```
ggplot(before_data,aes(x=date,y = commodity,color=price))+ geom_point(size=3)
```

In [131]:
```
library("IRdisplay")
display_png(file="Q3 before.png")
```



In [ ]:
```
ggplot(after_data,aes(x=date,y = commodity,color=price))+ geom_point(size=3)
```

In [132]:
```
display_png(file="Q3 after.png")
```



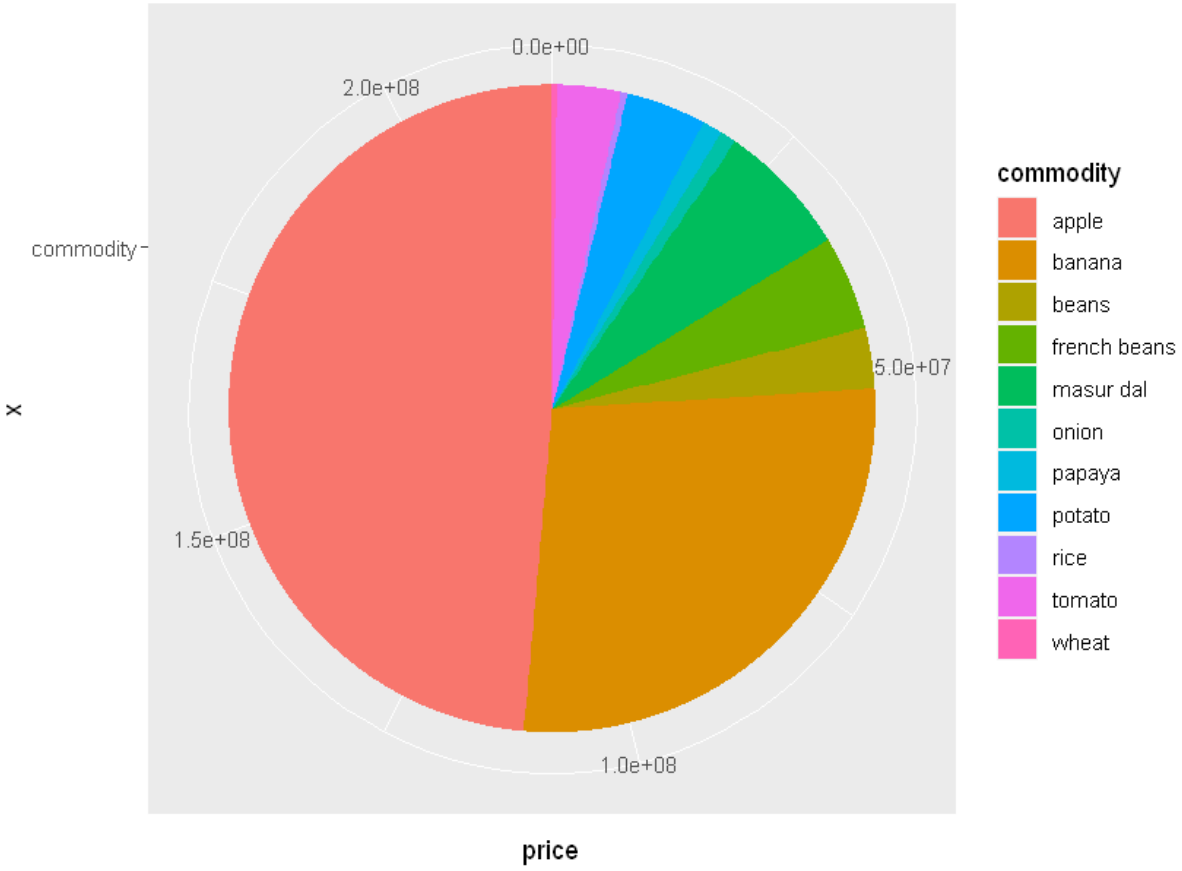
The above plots was done in R studio for better clarity of visualization.

According to before data plot there was continuous purchasing of commodities each day.

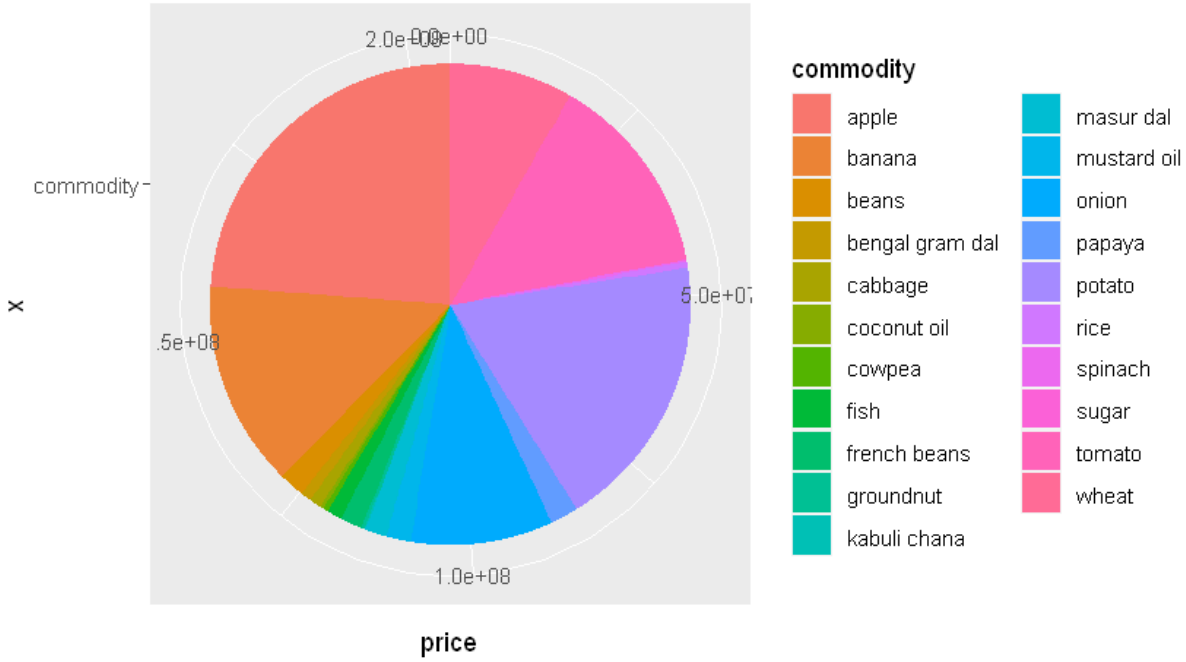
Then during lockdown the some commodities like sugar, spinach, wheat, rice, tomato, potato purchasing were largely disturbed.

Q4. Due to covid 19, did the supply chain increased or decreased?

In [59]:
```
ggplot(before_data,aes(x="commodity",y=price,fill=commodity))+geom_bar(stat="ide
```



In [27]:
```
ggplot(after_data,aes(x="commodity",y=price,fill=commodity))+geom_bar(stat="iden
```



The supply of rice, sugar, wheat, tomato, onion, potato was increased during lockdown and rest were moderately increased.

Q5. Is there any loss generated on commodity due to COVID-19?

We took the sum of prices of during lockdown dataset and subtracted with the sum of prices of before lockdown dataset to check for any loss or profit.

In [66]:
```
#subsets from before lockdown dataset
banana_ds=subset(before_data,commodity=="banana")
apple_ds=subset(before_data,commodity=="apple")
rice_ds=subset(before_data,commodity=="rice")
wheat_ds=subset(before_data,commodity=="wheat")
```

In [67]:
```
#subsets from during lockdown datasets
banana_dsb=subset(after_data,commodity=="banana")
apple_dsb=subset(after_data,commodity=="apple")
rice_dsb=subset(after_data,commodity=="rice")
wheat_dsb=subset(after_data,commodity=="wheat")
```

In [91]:
```
#checking for banana
sum(banana_dsb$price)-sum(banana_ds$price)
```

-30985917

In [92]:
```
#checking for apple
sum(apple_dsb$price)-sum(apple_ds$price)
```

-56379859

From the above calculations we can tell the prices of apple and banana decreased during lockdown.

In [93]:
```
#checking for rice
sum(rice_dsb$price)-sum(rice_ds$price)
```

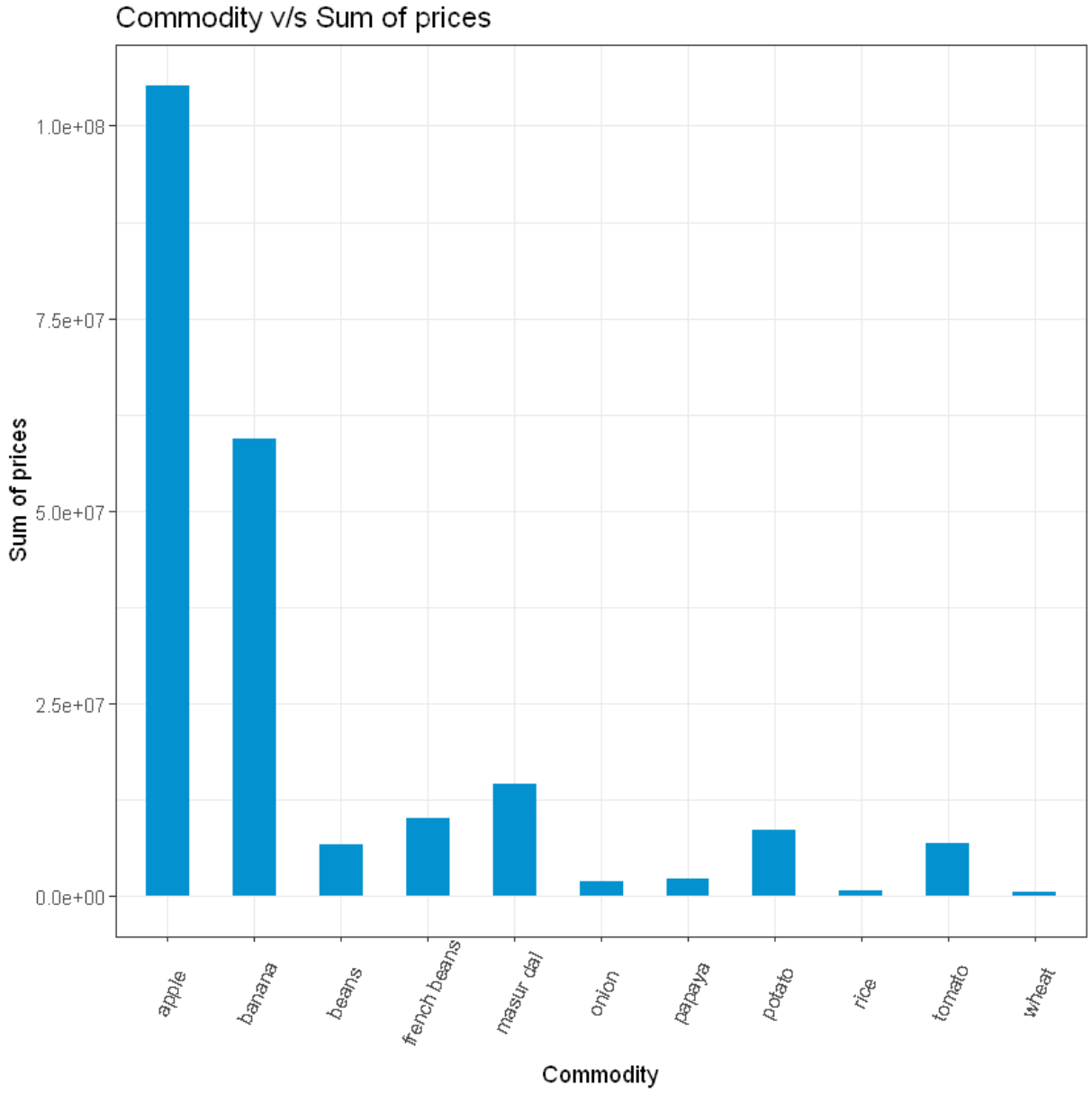
75363

In [95]:
```
#checking for wheat
sum(wheat_dsb$price)-sum(wheat_ds$price)
```

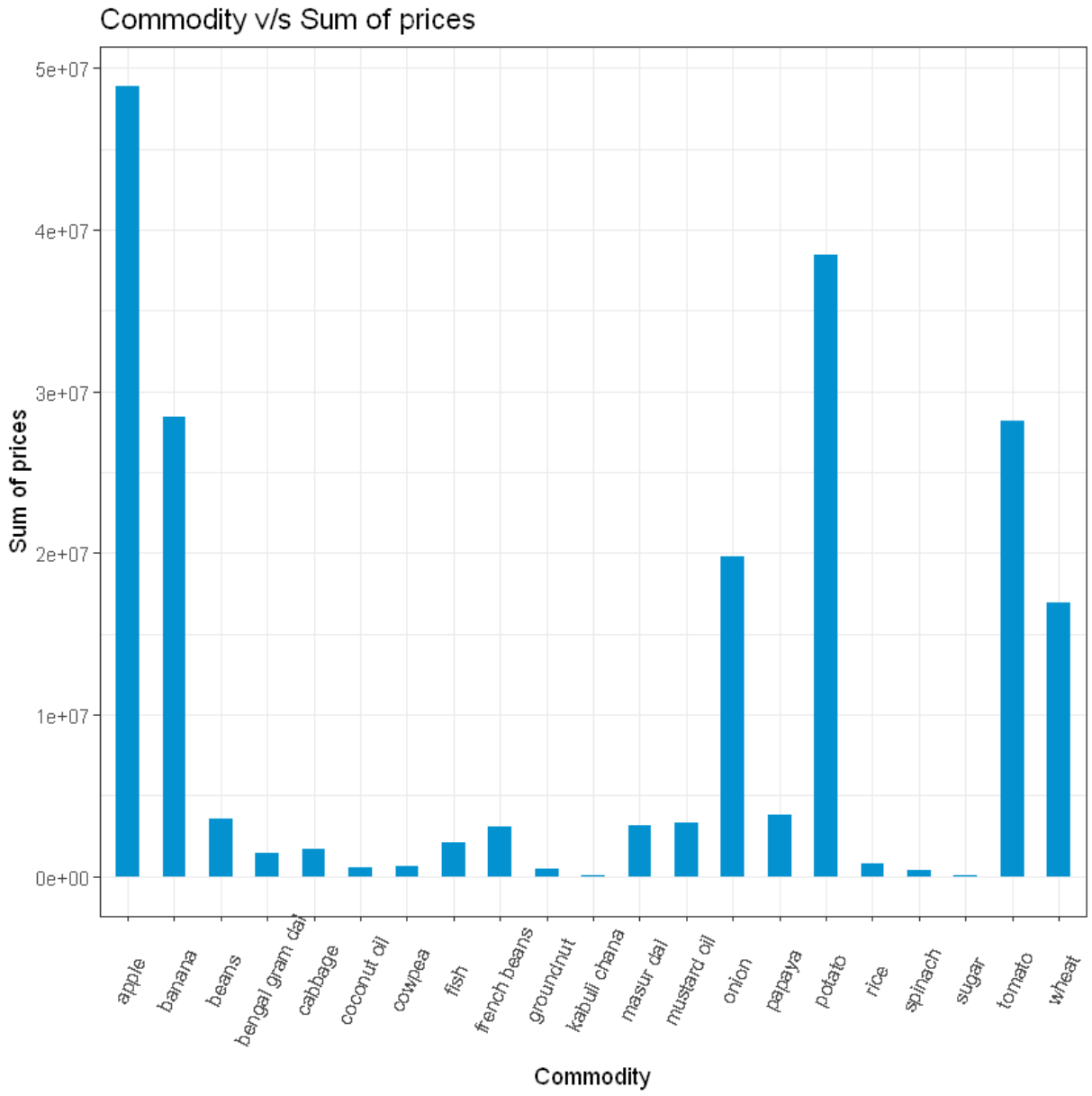
16449767.67

From the above calculations we can tell the prices of apple and banana increased during lockdown.

In [89]:
```
library(scales)
theme_set(theme_bw())
ggplot(before_data, aes(x=commodity,y=price)) +
  geom_bar(stat="identity", width=.5, fill="#0392cf") +
  labs(title="Commodity v/s Sum of prices") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))+xlab("Commodity")+ylab('
```



In [90]:
```
ggplot(after_data, aes(x=commodity,y=price)) +
  geom_bar(stat="identity", width=.5, fill="#0392cf") +
  labs(title="Commodity v/s Sum of prices") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))+xlab("Commodity")+ylab('
```

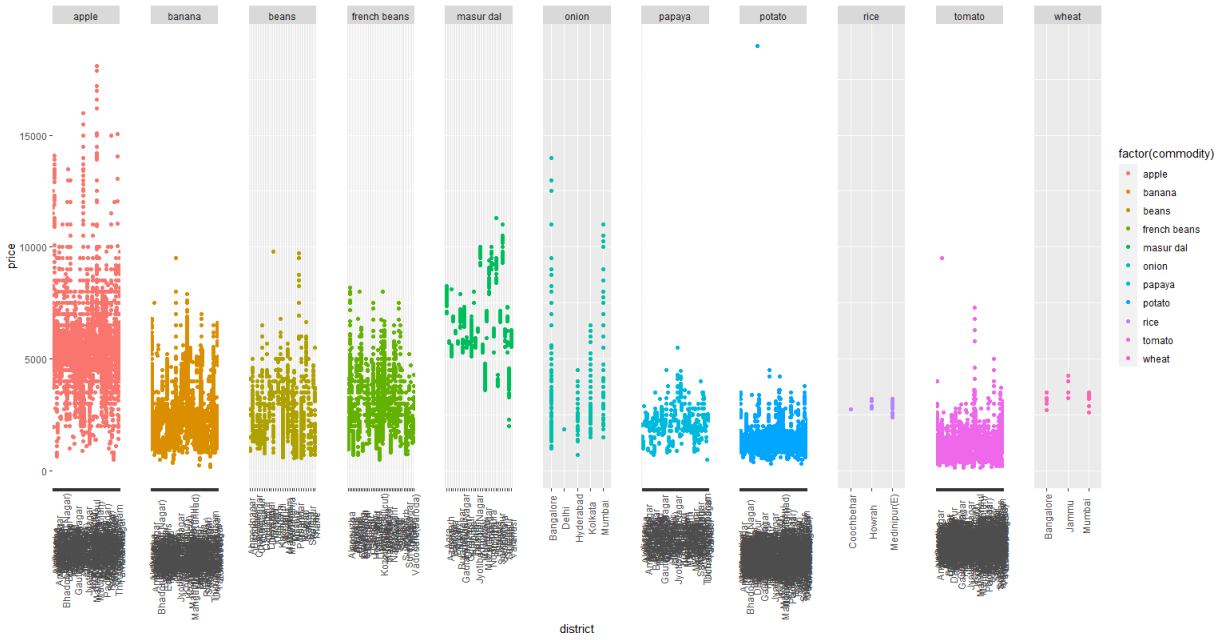


From the above calculations , we can conclude that there was a loss on some commodities during lockdown and there were some commodities with profit during lockdown

Q6. Price distribution of commodities district-wise

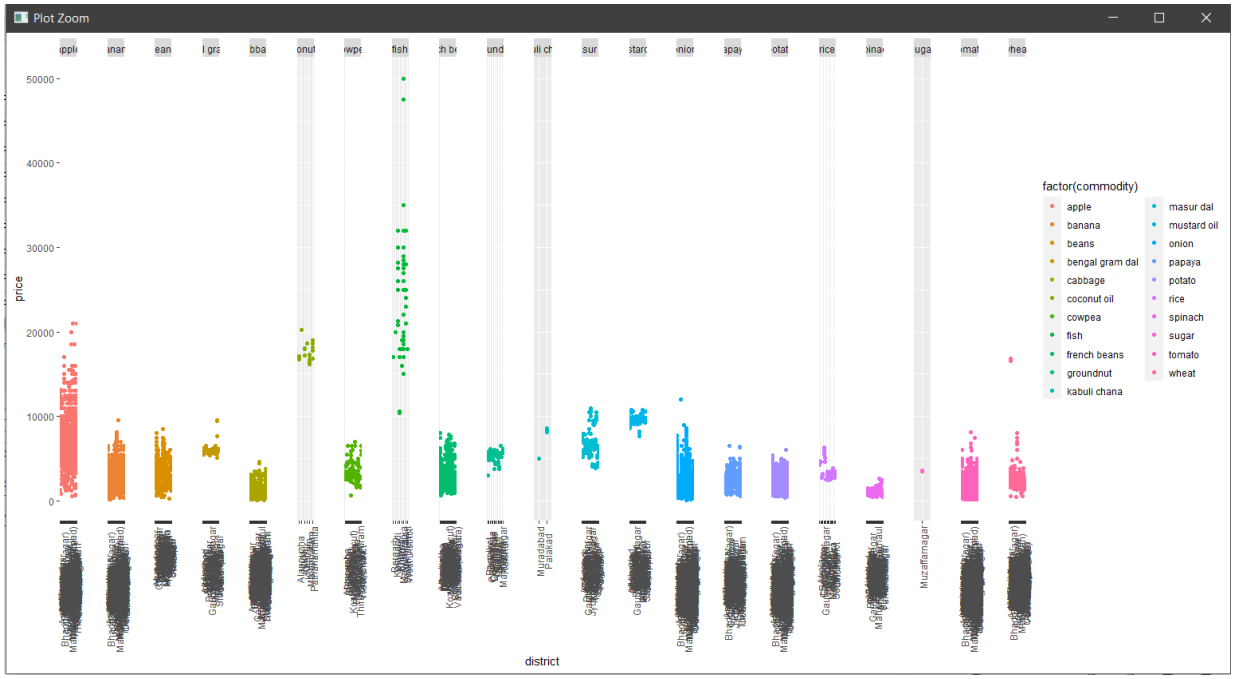
In [ ]:
```
ggplot(before_data, aes(district,price, colour = factor(commodity))) + geom_poin
+ facet_grid(cols =vars(commodity), scales = "free")
+ theme(axis.text.x = element_text(angle=90, vjust=0.6))
+ theme(panel.spacing.x =unit(2,"lines"),panel.spacing.y= unit(2,"lines"))
```

In [133]:
```
display_png(file="before.png")
```



In [ ]:
```
ggplot(after_data, aes(district,price, colour = factor(commodity))) + geom_point
+ facet_grid(cols =vars(commodity), scales = "free")
+ theme(axis.text.x = element_text(angle=90, vjust=0.6))
+ theme(panel.spacing.x =unit(2,"lines"),panel.spacing.y= unit(2,"lines"))
```

In [134]:
```
display_png(file="after.PNG")
```



The above plots was done in R studio for better clarity of visualization.

Inferential Analysis

One sample T-test

In [30]:
```
mean(data$price)
```

2666.38277155662

In [31]:
```
t.test(data$price,mu = 2666.38,alternative = "two.sided")
```

```
        One Sample t-test

data:  data$price
t = 0.00051475, df = 158419, p-value = 0.9996
alternative hypothesis: true mean is not equal to 2666.38
95 percent confidence interval:
 2655.830 2676.936
sample estimates:
mean of x
 2666.383
```

p value is greater than 0.05, so we can't reject the Null Hypothesis

confidence limits: (2655.830 , 2676.936)

CHI square test

In [52]:
```
options(scipen = 0)
```

In [53]:
```
test=chisq.test(data$minPrice,data$maxPrice)
test
```

```
Warning message in chisq.test(data$minPrice, data$maxPrice):
"Chi-squared approximation may be incorrect"

        Pearson's Chi-squared test

data:  data$minPrice and data$maxPrice
X-squared = 88587050, df = 4974592, p-value < 2.2e-16
```

In [39]:
```
test$p.value
```

0

As the p value is less than the significance level ,we can conclude that the observed distributions is not same as the expected distribution . therefore the relationship exist between the categorical variables.

Anova Test

In [54]:
```
res.aov=aov(minPrice~maxPrice,data=data)
res.aov
```

```
Call:
   aov(formula = minPrice ~ maxPrice, data = data)

Terms:
                    maxPrice    Residuals
Sum of Squares  608007387944  54917632840
Deg. of Freedom            1       158418

Residual standard error: 588.7808
Estimated effects may be unbalanced
```

In [55]:
```
summary(res.aov)
```

```
                Df    Sum Sq   Mean Sq F value Pr(>F)
maxPrice         1 6.080e+11 6.080e+11 1753887 <2e-16 ***
Residuals   158418 5.492e+10 3.467e+05
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p value is less than 0.05, then this data didn't occured by chance and is repeatable.

In [ ]:
```

```

Thanks to IBM