

# Expanding Freight Business Unit

*Demond Love*

## Business Use

Customer is looking to extend their freight business unit and are looking for the areas with the most success thus far, and would customer segmentation analysis also. Ultimately settling on 4 questions:

1. Which are the most profitable freight surcharges?
2. Who are the most profitable freight customers?
3. Which customers have the highest proportion of freight to total revenue?
4. Which counties have the highest freight per load in Indiana, Illinois, Kentucky, Tennessee, & Missouri?

## Gather Data from a SQL Database

```
ProfitableSurcharges = sqlQuery(con, "
    SELECT
        TaxID AS Surcharge
        , CAST(SUM(TaxAmt) AS Money) AS TotalSurchargeAmt
    FROM ARInvoiceItemCompTax
    WHERE CreatedBy = 'F'
    GROUP BY TaxID
        HAVING SUM(TaxAmt) != 0
    ORDER BY TotalSurchargeAmt DESC
    ")
ProfitableCustomers = sqlQuery(con, "
    SELECT TOP 50
        via.StandardAcctNo
        , CAST(SUM(arict.TaxAmt) AS Money) AS TotalSurchargeAmt
    FROM ARInvoiceItemCompTax arict
        JOIN vInvoiceAll via ON arict.SysTrxNo = via.SysTrxNo
    GROUP BY via.StandardAcctNo
        HAVING SUM(arict.TaxAmt) != 0
    ORDER BY TotalSurchargeAmt DESC
    ")
ProportionFreight = sqlQuery(con, "
    SELECT TOP 50
        inv.StandardAcctNo
        , SUM(TotalFrtAmt) AS TotalFrtAmt
        , SUM(InvoiceTotal) AS InvoiceTotal
        , (SUM(TotalFrtAmt)/SUM(InvoiceTotal)) * 100 AS ProportionFrt
    FROM vInvoiceAll inv
        JOIN ARStandardAcct asa ON inv.StandardAcctID = asa.StandardAcctID
    WHERE asa.Active = 'Y'
    GROUP BY inv.StandardAcctNo
        HAVING SUM(TotalFrtAmt) != 0
    ORDER BY ProportionFrt DESC
    ")
```

```

FrtPerLoad = sqlQuery(con, "
SELECT
    sp.Code AS State
    , County.Code AS County
    , COUNT(via.SysTrxNo) as NumberofInvoices
    , CAST(SUM(arict.TaxAmt) AS Money) AS TotalSurchargeAmt
    , CAST(SUM(arict.TaxAmt)/COUNT(via.SysTrxNo) AS decimal(24,2)) AS FrtPerLoad
FROM ARInvoiceItemCompTax arict
JOIN vInvoiceAll via ON arict.SysTrxNo = via.SysTrxNo
JOIN ARShipToAddress on via.ShipToID = ARShipToAddress.ShipToID
JOIN StateProv sp on ARShipToAddress.StateProvID = sp.ID
JOIN County on ARShipToAddress.CountyID = County.ID
WHERE Status != 'A'
AND Status = 'C'
AND sp.Code IN ('IN', 'IL', 'KY', 'TN', 'MO')
GROUP BY sp.Code, county.Code
HAVING SUM(arict.TaxAmt) != 0
AND COUNT(via.SysTrxNo) > 8
ORDER BY FrtPerLoad DESC
")

```

## Analysis/Visualization

Which are the most profitable freight surcharges?

```

ProfitableSurcharges$Surcharge = as.character(ProfitableSurcharges$Surcharge)
ProfitableSurcharges$TotalSurchargeAmt = currency(ProfitableSurcharges$TotalSurchargeAmt,
                                                    digits = 0L)
ProfitableSurchargesTOP10 = ProfitableSurcharges[1:10,]
kable(ProfitableSurchargesTOP10, "latex", caption = "Profitable Surcharges", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))

```

Table 1: Profitable Surcharges

Surcharge	TotalSurchargeAmt
1654	\$32,767
1677	\$28,718
4268	\$5,180
1658	\$2,605
4287	\$657
2017	\$595
4215	\$350
1670	\$246
1656	\$159
4213	\$150

Who are the most profitable freight customers?

```
# str(ProfitableCustomers)
ProfitableCustomers$StandardAcctNo = as.character(ProfitableCustomers$StandardAcctNo)
ProfitableCustomers$TotalSurchargeAmt = currency(ProfitableCustomers$TotalSurchargeAmt,
                                                  digits = 0L)
ProfitableCustomersTOP10 = ProfitableCustomers[1:10,]
kable(ProfitableCustomersTOP10, "latex", caption = "Profitable Customers", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

Table 2: Profitable Customers

StandardAcctNo	TotalSurchargeAmt
101087	\$9,440,871
101113	\$8,356,297
103733	\$6,961,498
101103	\$5,801,152
101140	\$5,470,746
101098	\$5,161,752
104612	\$4,266,046
103528	\$3,965,452
103789	\$3,785,234
102565	\$3,432,772

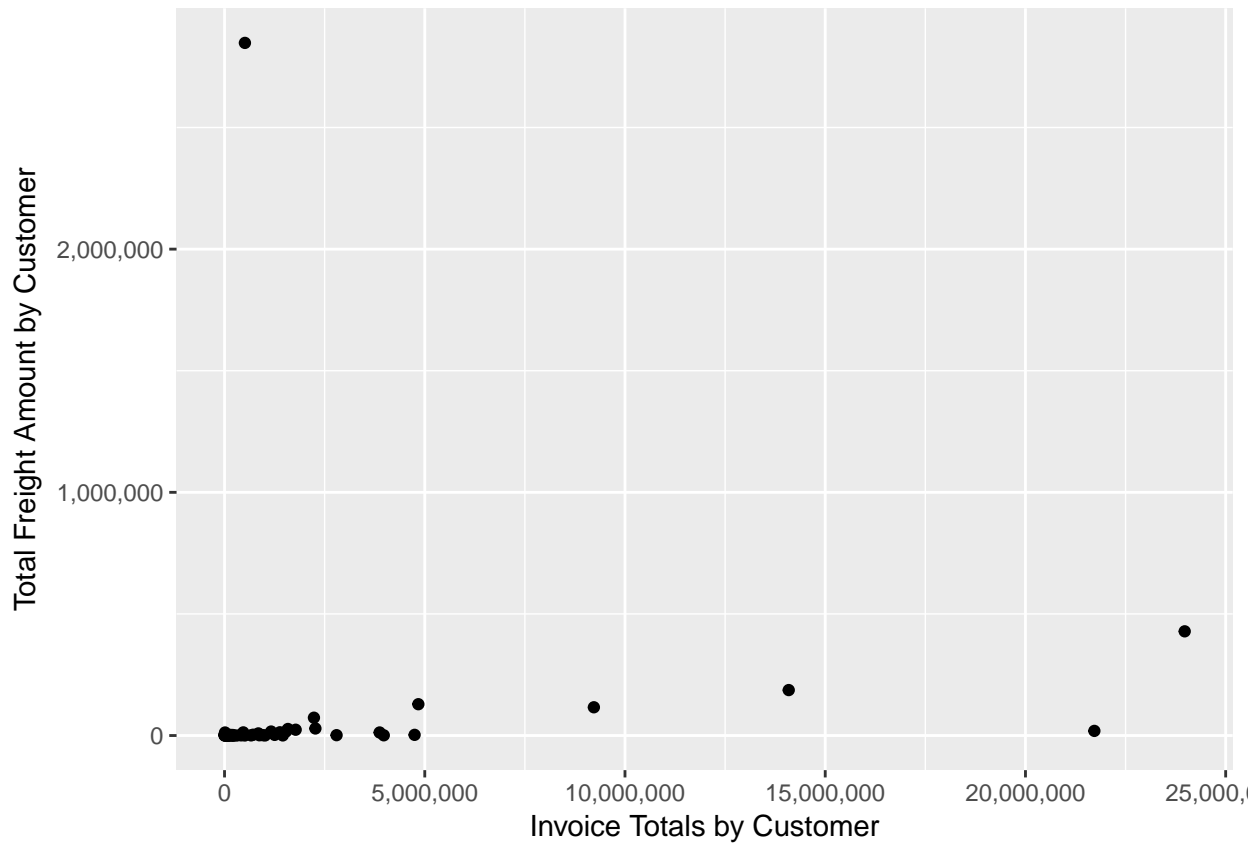
Which customers have the highest proportion of freight to total revenue?

```
# str(ProportionFreight)
ProportionFreight$TotalFrftAmt = currency(ProportionFreight$TotalFrftAmt,
                                           digits = 0L)
ProportionFreight$InvoiceTotal = currency(ProportionFreight$InvoiceTotal,
                                           digits = 0L)
ProportionFreight$ProportionFrft = gsub("\\\\.\\.", "", ProportionFreight$ProportionFrft)
ProportionFreight$ProportionFrft = paste(ProportionFreight$ProportionFrft, '%', sep = "")
ProportionFreightTOP10 = ProportionFreight[1:10,]
kable(ProportionFreightTOP10, "latex", caption = "Highest Proportion Freight", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

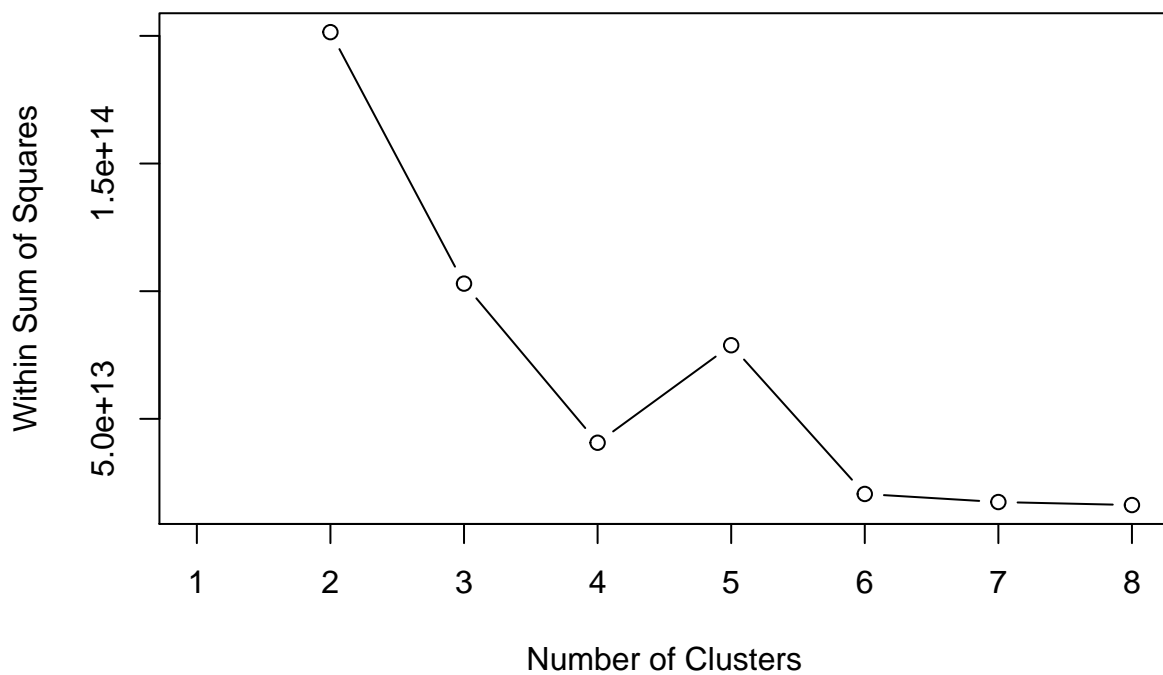
Table 3: Highest Proportion Freight

StandardAcctNo	TotalFrftAmt	InvoiceTotal	ProportionFrft
201029	\$2,847,896	\$509,208	559%
201020	\$5,955	\$5,939	100%
201013	\$925	\$924	100%
201001	\$12,422	\$13,087	94%
201003	\$2,989	\$3,723	80%
105544	\$73,261	\$2,232,792	3%
101676	\$12,932	\$468,729	2%
101096	\$128,813	\$4,839,895	2%
103769	\$111	\$4,183	2%
101649	\$2,835	\$132,057	2%

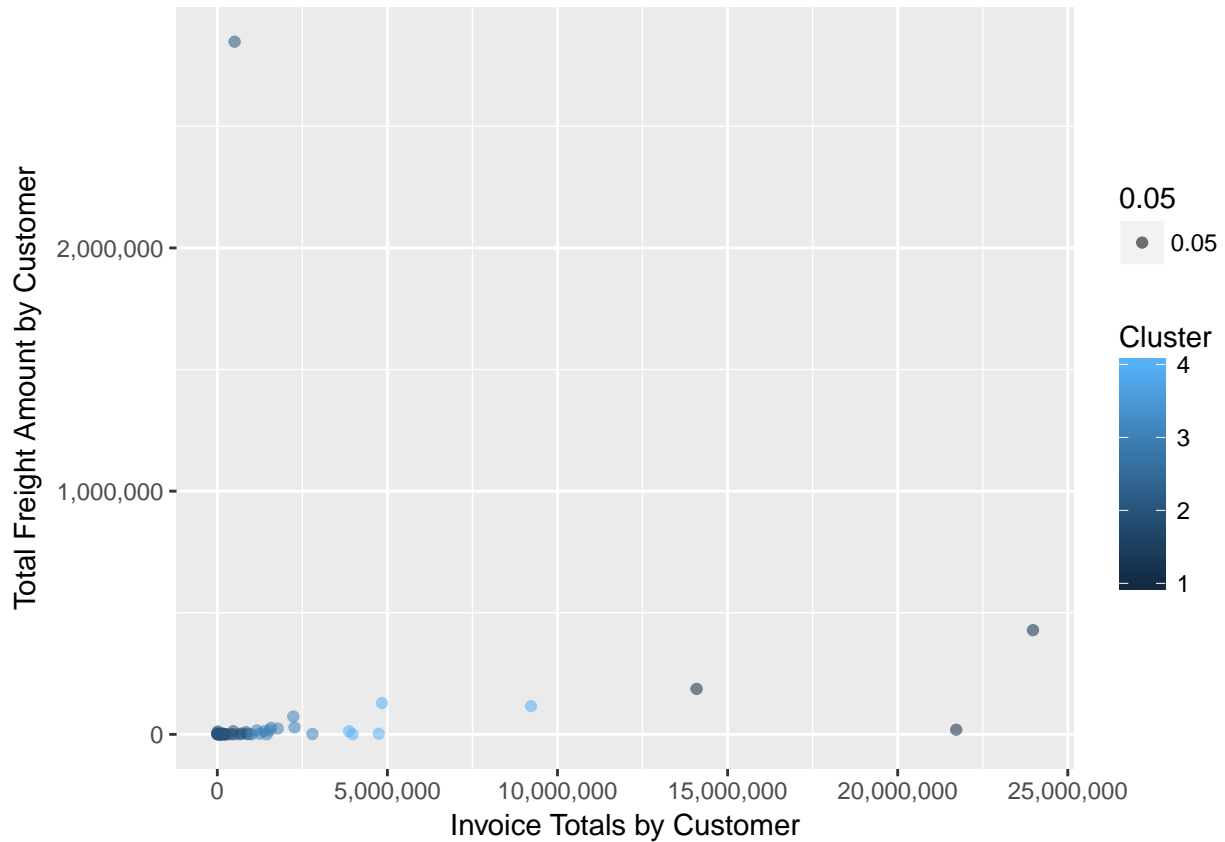
```
ggplot(ProportionFreight, aes(x = InvoiceTotal, y = TotalFrftAmt)) +
  geom_point() +
  scale_y_continuous(name="Total Freight Amount by Customer", labels = scales::comma) +
  scale_x_continuous(name="Invoice Totals by Customer", labels = scales::comma)
```



```
kmeansdata = data.frame(ProportionFreight$InvoiceTotal, ProportionFreight$TotalFrtAmt)
wss = (nrow(file)-1)*sum(apply(kmeansdata,2,var))
for (i in 2:8) wss[i] = sum(kmeans(kmeansdata, centers = i)$withinss)
plot(1:8, wss, type = 'b', xlab="Number of Clusters", ylab = "Within Sum of Squares")
```



```
clusters = kmeans(kmeansdata, centers = 4)
kmeansdata$Cluster = clusters$cluster
names(kmeansdata) = c('InvoiceTotal', 'TotalFrftAmt', 'Cluster')
ggplot(kmeansdata, aes(x = InvoiceTotal, y = TotalFrftAmt)) +
  geom_point(aes(alpha = 0.05, color = Cluster)) +
  scale_y_continuous(name="Total Freight Amount by Customer", labels = scales::comma) +
  scale_x_continuous(name="Invoice Totals by Customer", labels = scales::comma)
```



```
clusters$size
```

```
## [1] 3 30 12 5
```

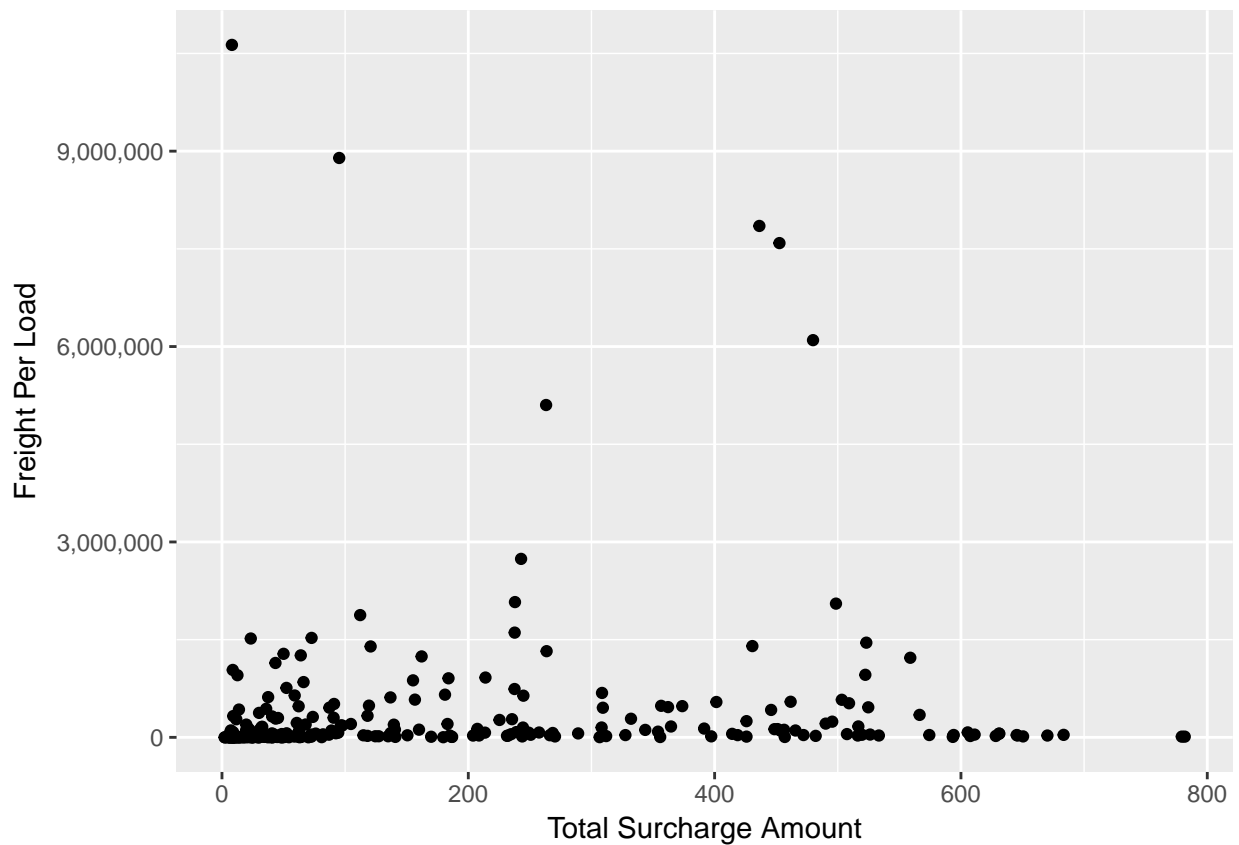
Which counties have the highest freight per load between Indiana, Illinois, Kentucky, Tennessee, & Missouri?

```
# str(FrtPerLoad)
FrtPerLoad$NumberOfInvoices = as.numeric(FrtPerLoad$NumberOfInvoices)
FrtPerLoad$TotalSurchargeAmt = currency(FrtPerLoad$TotalSurchargeAmt,
                                         digits = 0L)
FrtPerLoad$FrtPerLoad = currency(FrtPerLoad$FrtPerLoad,
                                  digits = 0L)
FrtPerLoadTOP20 = FrtPerLoad[1:20,]
kable(FrtPerLoadTOP20, "latex", caption = "Highest Freight Per Load", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"))
```

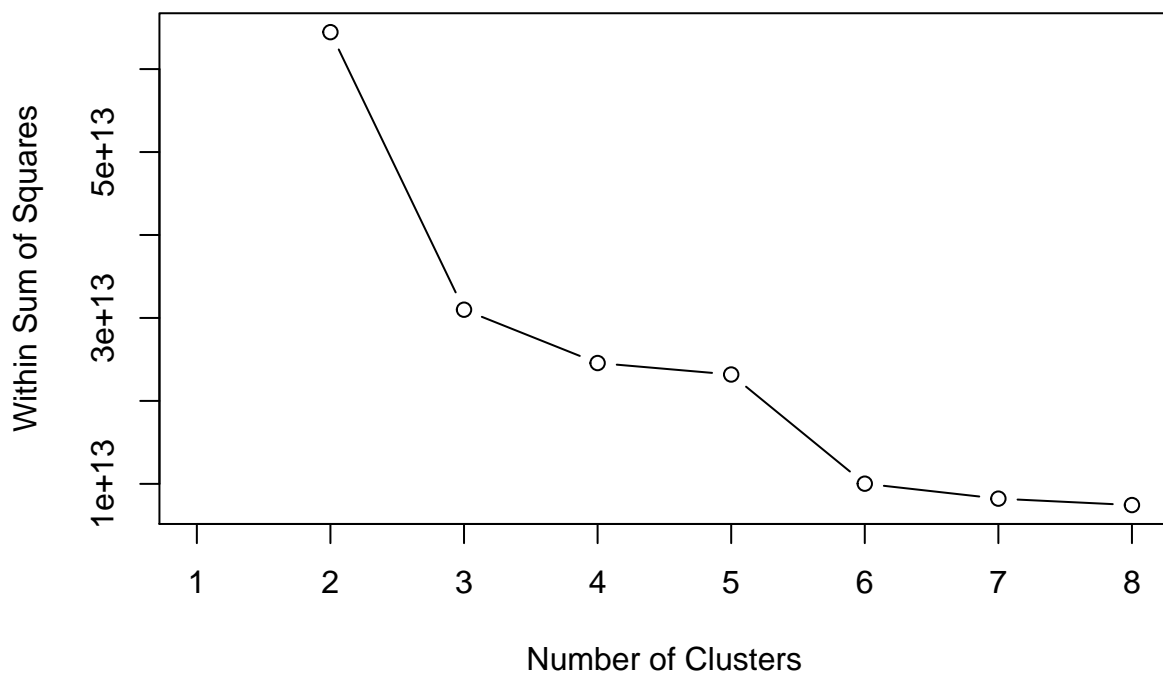
Table 4: Highest Freight Per Load

State	County	NumberOfInvoices	TotalSurchargeAmt	FrtPerLoad
IN	Franklin	14	\$10,944	\$782
IN	Cass	14	\$10,910	\$779
KY	Madison	56	\$38,269	\$683
IN	Washington	42	\$28,149	\$670
KY	Marion	22	\$14,310	\$650
KY	Anderson	44	\$28,421	\$646
KY	Mercer	55	\$35,489	\$645
KY	Carroll	90	\$56,804	\$631
KY	Jessamine	90	\$56,779	\$631
KY	Barren	34	\$21,362	\$628
KY	Harrison	69	\$42,166	\$611
KY	Grayson	34	\$20,659	\$608
IL	Kendall	125	\$75,690	\$606
IL	Whiteshield	61	\$36,242	\$594
IN	Tipton	43	\$25,528	\$594
IN	Steuben	12	\$7,121	\$593
KY	Shelby	62	\$35,608	\$574
IN	Clay	609	\$344,955	\$566
IN	Johnson	2186	\$1,221,686	\$559
KY	Simpson	53	\$28,268	\$533

```
ggplot(FrtPerLoad, aes(x = FrtPerLoad, y = TotalSurchargeAmt)) +
  geom_point() +
  scale_y_continuous(name="Freight Per Load", labels = scales::comma) +
  scale_x_continuous(name="Total Surcharge Amount", labels = scales::comma)
```

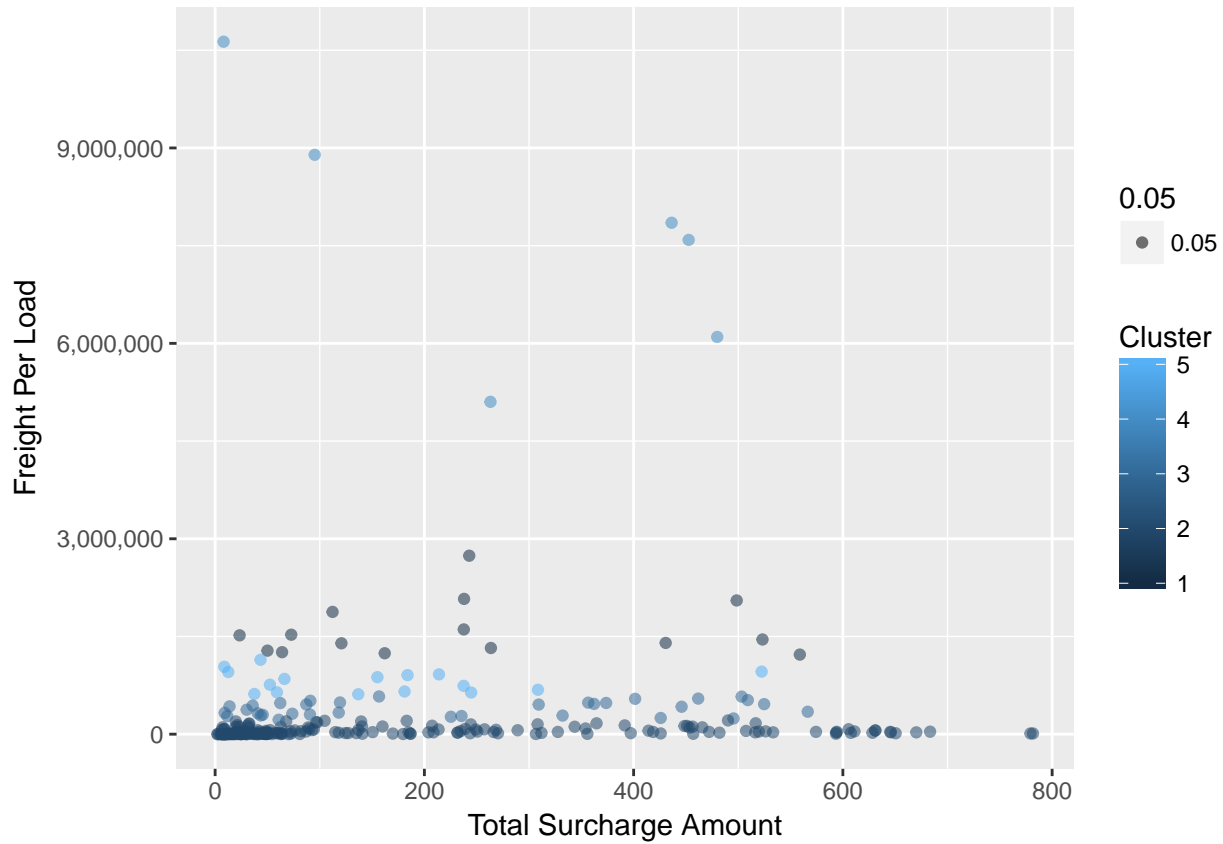


```
kmeansdata = data.frame(FrtPerLoad$FrtperLoad, FrtPerLoad$TotalSurchargeAmt)
wss = (nrow(file)-1)*sum(apply(kmeansdata,2,var))
for (i in 2:8) wss[i] = sum(kmeans(kmeansdata, centers = i)$withinss)
plot(1:8, wss, type = 'b', xlab="Number of Clusters", ylab = "Within Sum of Squares")
```





```
clusters = kmeans(kmeansdata, centers = 5)
kmeansdata$Cluster = clusters$cluster
names(kmeansdata) = c('FrtPerLoad', 'TotalSurchargeAmt', 'Cluster')
ggplot(kmeansdata, aes(x = FrtPerLoad, y = TotalSurchargeAmt)) +
  geom_point(aes(alpha = 0.05, color = Cluster)) +
  scale_y_continuous(name="Freight Per Load", labels = scales::comma) +
  scale_x_continuous(name="Total Surcharge Amount", labels = scales::comma)
```



```
clusters$size
```

```
## [1] 15 194 33 6 16
```