# Project-Tidying

Santosh Saranyan

11/30/2021

```
dir<-"Project"
path=file.path(dir,"data_relationship.csv")
df<-read_csv(path, na=c("-1", " ", "Other", "", "NA"))
```

```
## Rows: 328672 Columns: 34
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (7): rptownername, relationship, date, transactioncode, Quarter, ticker...
## dbl (13): Column1, X, fid, cik, rptownercik, transactionshares, transactionp...
## lgl (14): isCMO, isCTO, isCFO, isCEO, isCAO, isCRO, isCIO, isCSO, isEVP, isP...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
df<-na.omit(df)
df<-filter(df,priceafteroneweek>=0)
df2<-df[,-c(1,2)]
df2$date<-as.POSIXct(df2$date, format = "%m/%d/%Y")
head(df2,10)
```

```
## # A tibble: 10 x 32
##        fid     cik rptownercik rptownername    relationship  date
##      <dbl>   <dbl>       <dbl> <chr>           <chr>         <dttm>
##  1 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2018-12-31 00:00:00
##  2 1.57e14 1327318     1607879 Krafcik John    Director      2018-12-31 00:00:00
##  3 1.57e14 1327318     1688816 Gunsagar Neeraj ['EVP, Chief~ 2018-12-31 00:00:00
##  4 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2018-12-31 00:00:00
##  5 1.57e14 1327318     1609465 Pierantoni John ['Interim CF~ 2018-12-31 00:00:00
##  6 1.57e14 1327318     1607879 Krafcik John    Director      2019-01-01 00:00:00
##  7 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2019-01-02 00:00:00
##  8 1.57e14 1327318     1607879 Krafcik John    Director      2019-02-01 00:00:00
##  9 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2019-02-15 00:00:00
## 10 1.57e14 1327318     1721059 Darrow Michael  ['EVP, Partn~ 2019-02-15 00:00:00
## # ... with 26 more variables: transactioncode <chr>, transactionshares <dbl>,
## #   transactionpricepershare <dbl>, sharesownedfollowingtransaction <dbl>,
## #   Quarter <chr>, ticker <chr>, priceafteroneweek <dbl>,
## #   priceafteronemonth <dbl>, priceafteraquarter <dbl>,
## #   priceaftersixmonths <dbl>, priceafteryear <dbl>, new_relationship <chr>,
## #   isCMO <lgl>, isCTO <lgl>, isCFO <lgl>, isCEO <lgl>, isCAO <lgl>,
## #   isCRO <lgl>, isCIO <lgl>, isCSO <lgl>, isEVP <lgl>, ...
```

```
write_csv(head(df2,10),"Snippet.csv")
```

```
df2$percentchangeafterweek<-round(((df2$priceafteroneweek-df2$transactionpricepershare)/df2$transaction
df2$percentchangeafteronemonth<-round(((df2$priceafteronemonth-df2$transactionpricepershare)/df2$transa
df2$percentchangeafterquarter<-round(((df2$priceafteraquarter-df2$transactionpricepershare)/df2$transac
df2$percentchangeaftersixmonths<-round(((df2$priceaftersixmonths-df2$transactionpricepershare)/df2$trans
df2$percentchangeafteryear<-round(((df2$priceafteryear-df2$transactionpricepershare)/df2$transactionpri
df2
```

```
## # A tibble: 248,942 x 37
##        fid     cik rptownercik rptownername    relationship  date
##      <dbl>   <dbl>       <dbl> <chr>           <chr>         <dttm>
##  1 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2018-12-31 00:00:00
##  2 1.57e14 1327318     1607879 Krafcik John    Director      2018-12-31 00:00:00
##  3 1.57e14 1327318     1688816 Gunsagar Neeraj ['EVP, Chief~ 2018-12-31 00:00:00
##  4 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2018-12-31 00:00:00
##  5 1.57e14 1327318     1609465 Pierantoni John ['Interim CF~ 2018-12-31 00:00:00
##  6 1.57e14 1327318     1607879 Krafcik John    Director      2019-01-01 00:00:00
##  7 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2019-01-02 00:00:00
##  8 1.57e14 1327318     1607879 Krafcik John    Director      2019-02-01 00:00:00
##  9 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2019-02-15 00:00:00
## 10 1.57e14 1327318     1721059 Darrow Michael  ['EVP, Partn~ 2019-02-15 00:00:00
## # ... with 248,932 more rows, and 31 more variables: transactioncode <chr>,
## #   transactionshares <dbl>, transactionpricepershare <dbl>,
## #   sharesownedfollowingtransaction <dbl>, Quarter <chr>, ticker <chr>,
## #   priceafteroneweek <dbl>, priceafteronemonth <dbl>,
## #   priceafteraquarter <dbl>, priceaftersixmonths <dbl>, priceafteryear <dbl>,
## #   new_relationship <chr>, isCMO <lgl>, isCTO <lgl>, isCFO <lgl>, isCEO <lgl>,
## #   isCAO <lgl>, isCRO <lgl>, isCIO <lgl>, isCSO <lgl>, isEVP <lgl>, ...
```

```
df2<-df2 %>% mutate(changeaftersixmonths=ifelse((priceaftersixmonths-transactionpricepershare)>0, "Incr
df2<-df2 %>% mutate(changeafteramonth=ifelse((priceafteronemonth-transactionpricepershare)>0, "Increase
df2<-df2 %>% mutate(changeafteraweek=ifelse((priceafteroneweek-transactionpricepershare)>0, "Increase",
df2$changeafteronemonth<-round((df2$priceafteronemonth-df2$transactionpricepershare), digits=1)
df2
```

```
## # A tibble: 248,942 x 41
##        fid     cik rptownercik rptownername    relationship  date
##      <dbl>   <dbl>       <dbl> <chr>           <chr>         <dttm>
##  1 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2018-12-31 00:00:00
##  2 1.57e14 1327318     1607879 Krafcik John    Director      2018-12-31 00:00:00
##  3 1.57e14 1327318     1688816 Gunsagar Neeraj ['EVP, Chief~ 2018-12-31 00:00:00
##  4 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2018-12-31 00:00:00
##  5 1.57e14 1327318     1609465 Pierantoni John ['Interim CF~ 2018-12-31 00:00:00
##  6 1.57e14 1327318     1607879 Krafcik John    Director      2019-01-01 00:00:00
##  7 1.57e14 1327318     1741415 McClung Robert  ['EVP ', ' C~ 2019-01-02 00:00:00
##  8 1.57e14 1327318     1607879 Krafcik John    Director      2019-02-01 00:00:00
##  9 1.57e14 1327318     1688838 Swart Jeff      ['EVP, Gen. ~ 2019-02-15 00:00:00
## 10 1.57e14 1327318     1721059 Darrow Michael  ['EVP, Partn~ 2019-02-15 00:00:00
## # ... with 248,932 more rows, and 35 more variables: transactioncode <chr>,
## #   transactionshares <dbl>, transactionpricepershare <dbl>,
## #   sharesownedfollowingtransaction <dbl>, Quarter <chr>, ticker <chr>,
```

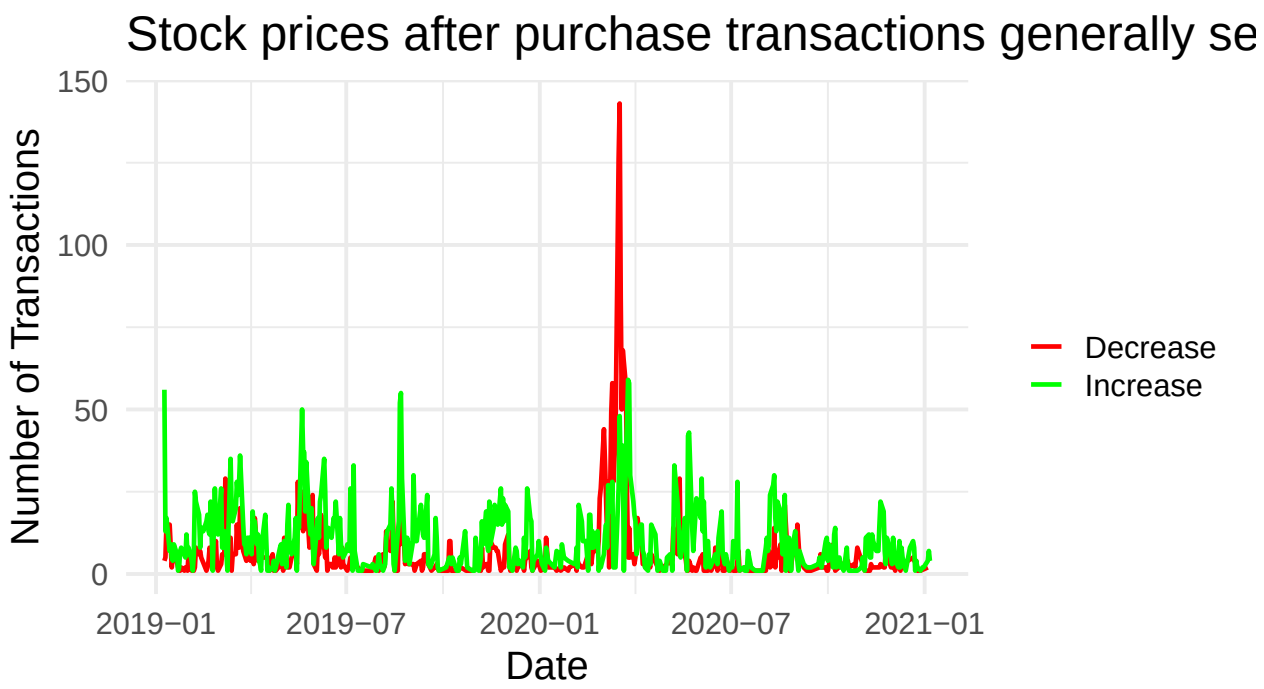
```
## #   priceafteroneweek <dbl>, priceafteronemonth <dbl>,
## #   priceafteraquarter <dbl>, priceaftersixmonths <dbl>, priceafteryear <dbl>,
## #   new_relationship <chr>, isCMO <lgl>, isCTO <lgl>, isCFO <lgl>, isCEO <lgl>,
## #   isCAO <lgl>, isCRO <lgl>, isCIO <lgl>, isCSO <lgl>, isEVP <lgl>, ...
```

## Correct graphs from this chunk onwards:

### After a week

```
#png("oneweekprices.png", width= 960, height= 540, unit= "px")
plot_data10 <- df2 %>%
  filter(date >= "2019-01-01" & date <= "2020-12-31" & transactioncode %in% c("P"))
plot_data10<-distinct(plot_data10, rptownercik, changeafteraweek, .keep_all = TRUE)
plot_data10$date<-plot_data10$date + days(7)
ggplot(plot_data10, aes(x= date, color= changeafteraweek)) +
  stat_count(geom= "line", aes(y= ..count..), position= "dodge", size= 1) +
  theme_minimal(base_size = 18) +
  labs(x= "Date", y= "Number of Transactions",
       title= "Stock prices after purchase transactions generally seem to increase", color="") +
  scale_color_manual(values= c("Red", "green"))
```

```
## Warning: Width not defined. Set with `position_dodge(width = ?)`
```

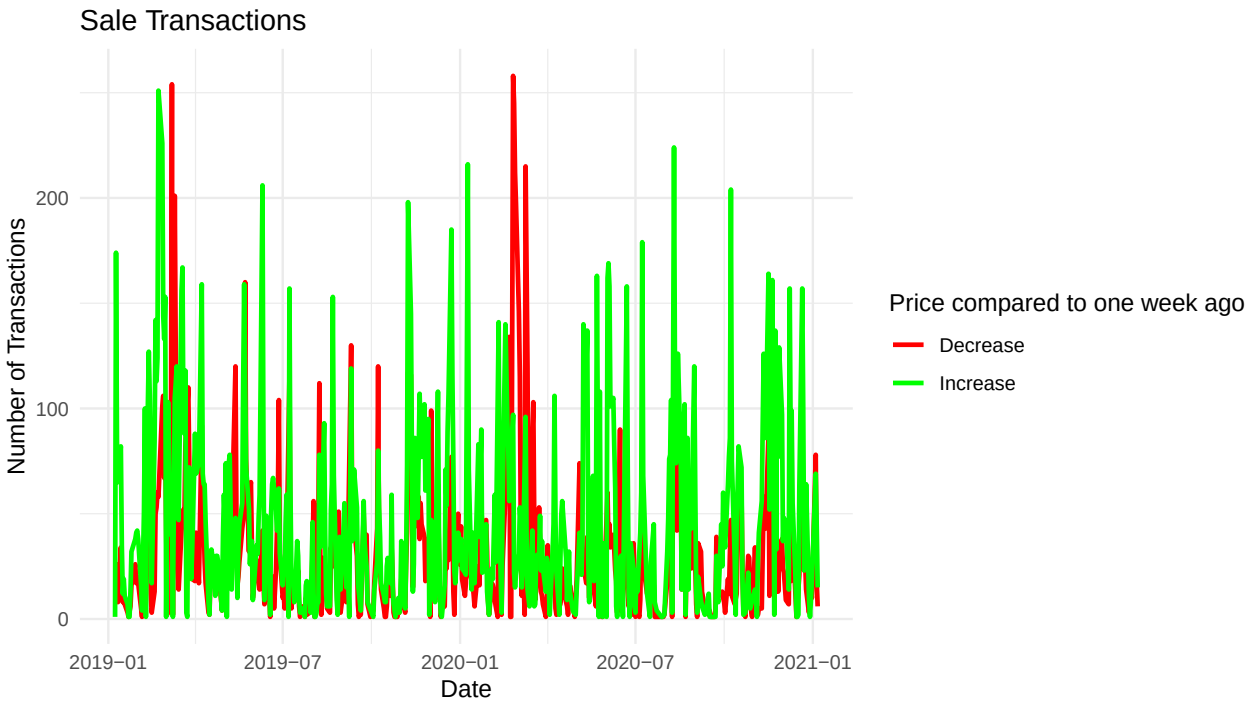



```
#dev.off()
```

```
plot_data10 <- df2 %>%
  filter(date >= "2019-01-01" & date <= "2020-12-31" & transactioncode %in% c("S"))
plot_data10<-distinct(plot_data10, rptownercik, changeafteraweek, date, .keep_all = TRUE)
```

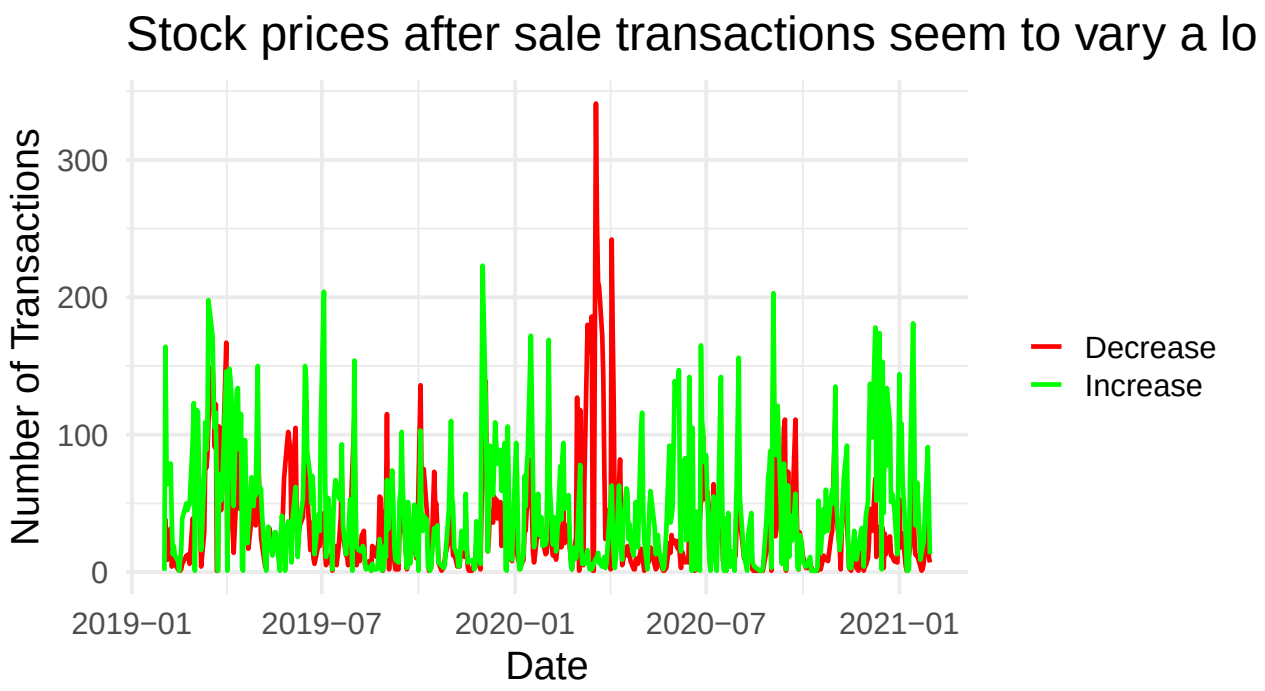
```
plot_data10$date<-plot_data10$date + days(7)
ggplot(plot_data10, aes(x= date, color= changeafteraweek)) +
  stat_count(geom= "line", aes(y= ..count..), position= "dodge", size= 1) +
  theme_minimal() +
  labs(x= "Date", y= "Number of Transactions",
       title= "Sale Transactions", color="Price compared to one week ago") +
  scale_color_manual(values= c("Red", "green"))
```

```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```




```
#png("onemonthpricessales.png", width= 960, height= 540, unit= "px")
plot_data <- df2 %>%
  filter(date >= "2019-01-01" & date <= "2020-12-31" & transactioncode %in% c("S"))
plot_data<-distinct(plot_data, rptownercik, changeafteramonth, date, .keep_all = TRUE)
plot_data$date<-plot_data$date %m+% months(1)
ggplot(plot_data, aes(x= date, color= changeafteramonth)) +
  stat_count(geom= "line", aes(y= ..count..), position= "dodge", size= 1) +
  theme_minimal(base_size = 18) +
  labs(x= "Date", y= "Number of Transactions",
       title= "Stock prices after sale transactions seem to vary a lot over time", color="") +
  scale_color_manual(values= c("Red", "green"))
```

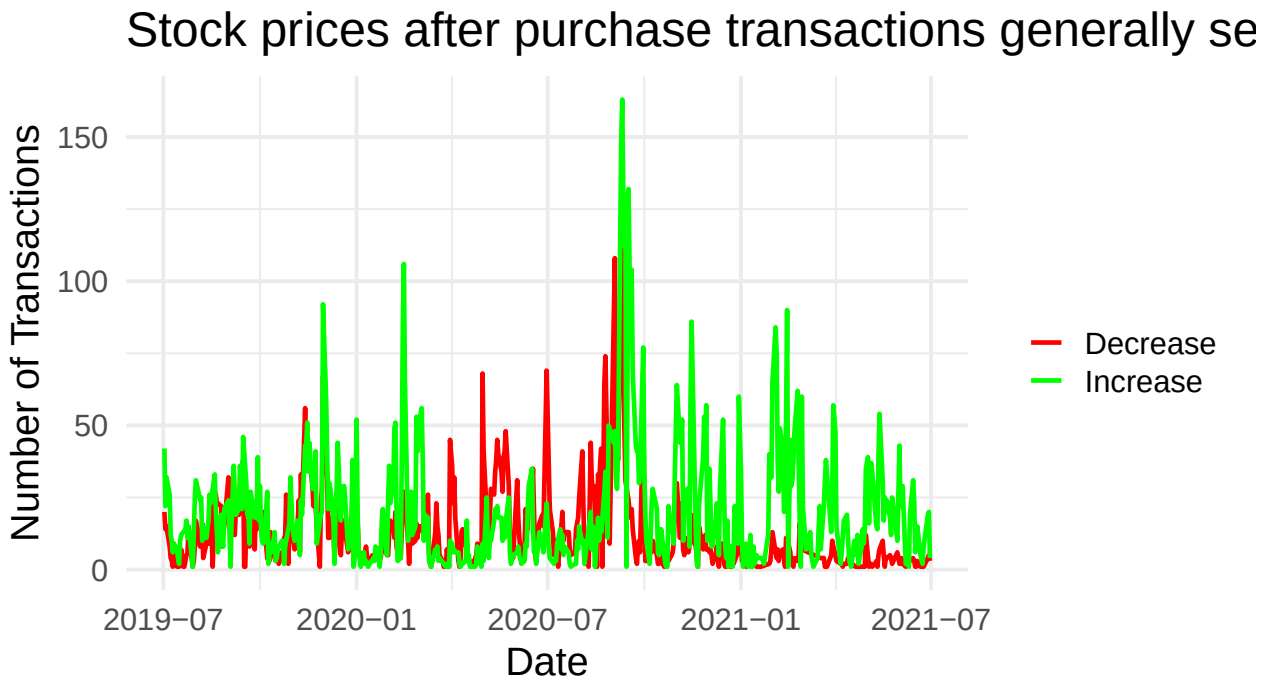
```
## Warning: Width not defined. Set with 'position_dodge(width = ?)'
```




```
#dev.off()
```

```
#png("sixmonthprices.png", width= 960, height= 540, unit= "px")
plot_data <- df2 %>%
  filter(date >= "2019-01-01" & date <= "2020-12-31" & transactioncode %in% c("P"))
plot_data<-distinct(plot_data, rptownercik, changeaftersixmonths, date, .keep_all = TRUE)
plot_data$date<-plot_data$date %m+% months(6)
ggplot(plot_data, aes(x= date, color= changeaftersixmonths)) +
  stat_count(geom= "line", aes(y= ..count..), position= "dodge", size= 1) +
  theme_minimal(base_size = 18) +
  labs(x= "Date", y= "Number of Transactions",
       title= "Stock prices after purchase transactions generally seem to increase after six months", c
  scale_color_manual(values= c("Red", "green"))
```

```
## Warning: Width not defined. Set with `position_dodge(width = ?)`
```




```
#dev.off()
```

```
df3<- df2 %>% filter(transactionshares>=10000 & transactioncode %in% c("P"))
df3<-distinct(df3, rptownercik, changeafteramonth, date, .keep_all = TRUE)
df3
```

```
## # A tibble: 5,535 x 41
##        fid     cik rptownercik rptownername           relationship date
##      <dbl>   <dbl>       <dbl> <chr>                  <chr>        <dttm>
##  1 1.57e14 1327318     1721059 Darrow Michael         Director     2020-03-11 00:00:00
##  2 8.99e13 1209028     1229249 TONKEL J ROCK JR       Director     2020-09-21 00:00:00
##  3 8.99e13 1209028     1229249 TONKEL J ROCK JR       Director     2020-09-22 00:00:00
##  4 8.99e13 1209028     1636595 Konzmann Richard Ernst ['EVP, CFO ~ 2020-09-22 00:00:00
##  5 8.99e13 1209028     1229249 TONKEL J ROCK JR       Director     2020-09-23 00:00:00
##  6 8.99e13 1209028     1636595 Konzmann Richard Ernst ['EVP, CFO ~ 2020-09-24 00:00:00
##  7 1.23e14    6201     1249552 PARKER W DOUGLAS       Director     2019-06-04 00:00:00
##  8 1.23e14    6201     1300047 Isom Robert D Jr       ['President~ 2019-06-04 00:00:00
##  9 1.23e14    6201     1197000 CAHILL JOHN T          Director     2019-06-04 00:00:00
## 10 1.23e14    6201     1182045 ALBAUGH JAMES F        Director     2019-10-28 00:00:00
## # ... with 5,525 more rows, and 35 more variables: transactioncode <chr>,
## #   transactionshares <dbl>, transactionpricepershare <dbl>,
## #   sharesownedfollowingtransaction <dbl>, Quarter <chr>, ticker <chr>,
## #   priceafteroneweek <dbl>, priceafteronemonth <dbl>,
## #   priceafteraquarter <dbl>, priceaftersixmonths <dbl>, priceafteryear <dbl>,
## #   new_relationship <chr>, isCMO <lgl>, isCTO <lgl>, isCFO <lgl>, isCEO <lgl>,
## #   isCAO <lgl>, isCRO <lgl>, isCIO <lgl>, isCSO <lgl>, isEVP <lgl>, ...
```

```
dftable<-table(df3$transactioncode, df3$changeafteramonth)
dftable
```

```
##
```

```
##     Decrease Increase
##   P     1779     3756
```

```
prop.table(dftable,1)
```

```
##
##      Decrease  Increase
##   P 0.3214092 0.6785908
```

```
chisq.test(dftable)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  dftable
## X-squared = 706.15, df = 1, p-value < 2.2e-16
```

```
dftb<-table(df3$is10percentowner, df3$changeafteraweek)
dftb
```

```
##
##         Decrease Increase
##   FALSE     1094     2943
##   TRUE       317     1181
```

```
prop.table(dftb,1)
```

```
##
##          Decrease  Increase
##   FALSE 0.2709933 0.7290067
##   TRUE  0.2116155 0.7883845
```

```
chisq.test(dftb)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb
## X-squared = 19.97, df = 1, p-value = 7.868e-06
```

```
y<-c("is10percentowner", "isCEO", "isCAO", "isCFO", "isChairman", "isDirector", "isCIO", "isCMO","isCRO
rel<-vector()
inc<-vector()
dec<-vector()
pos<-c("10percentOwner", "CEO", "CAO", "CFO", "Chairman", "Director", "CIO", "CMO","CRO", "CSO", "CTO",
for (i in y){
  print(i)
  dftb2<-table(pull(df3,i), df3$changeafteramonth)
  print(dftb2)
  if(length(as.numeric(table(dftb2)))>2){
```

```
  #inc<-append(inc,as.numeric(table(dftb2))[[4]])
  #dec<-append(dec,as.numeric(table(dftb2))[[3]])
  rel<-append(rel,as.numeric(prop.table(dftb2,1))[[4]]*100)
  }
else{
  rel<-append(rel,NA)
  #inc<-append(inc,0)
  #dec<-append(dec,0)
  }
  print(prop.table(dftb2,1))
  #chi<-append(chi,round(as.numeric(as.list(chisq.test(dftb2))[[1]]),digits=2))
  print(chisq.test(dftb2))
}
```

```
## [1] "is10percentowner"
##
##         Decrease Increase
##   FALSE     1392     2645
##   TRUE       387     1111
##
##          Decrease  Increase
##   FALSE 0.3448105 0.6551895
##   TRUE  0.2583445 0.7416555
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 37.057, df = 1, p-value = 1.147e-09
##
## [1] "isCEO"
##
##         Decrease Increase
##   FALSE     1760     3700
##   TRUE        19       56
##
##          Decrease  Increase
##   FALSE 0.3223443 0.6776557
##   TRUE  0.2533333 0.7466667
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 1.3146, df = 1, p-value = 0.2516
##
## [1] "isCAO"
##
##         Decrease Increase
##   FALSE     1776     3752
##   TRUE         3        4
##
##          Decrease  Increase
##   FALSE 0.3212735 0.6787265
##   TRUE  0.4285714 0.5714286
```

```
## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.041033, df = 1, p-value = 0.8395
##
## [1] "isCFO"
##
##         Decrease Increase
##   FALSE     1730     3631
##   TRUE        49      125
##
##          Decrease  Increase
##   FALSE 0.3227010 0.6772990
##   TRUE  0.2816092 0.7183908
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 1.1231, df = 1, p-value = 0.2892
##
## [1] "isChairman"
##
##         Decrease Increase
##   FALSE     1772     3752
##   TRUE         7        4
##
##          Decrease  Increase
##   FALSE 0.3207820 0.6792180
##   TRUE  0.6363636 0.3636364

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 3.6704, df = 1, p-value = 0.05539
##
## [1] "isDirector"
##
##         Decrease Increase
##   FALSE      559     1525
##   TRUE      1220     2231
##
##          Decrease  Increase
##   FALSE 0.2682342 0.7317658
##   TRUE  0.3535207 0.6464793
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
```

```
## data:  dftb2
## X-squared = 42.943, df = 1, p-value = 5.636e-11
##
## [1] "isClO"
##
##         Decrease Increase
##   FALSE     1775     3752
##   TRUE         4        4
##
##          Decrease  Increase
##   FALSE 0.3211507 0.6788493
##   TRUE  0.5000000 0.5000000

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.49505, df = 1, p-value = 0.4817
##
## [1] "isCMO"
##
##         Decrease Increase
##   FALSE     1779     3753
##   TRUE         0        3
##
##          Decrease  Increase
##   FALSE 0.3215835 0.6784165
##   TRUE  0.0000000 1.0000000

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.32954, df = 1, p-value = 0.5659
##
## [1] "isCRO"
##
##         Decrease Increase
##   FALSE     1778     3756
##   TRUE         1        0
##
##          Decrease  Increase
##   FALSE 0.3212866 0.6787134
##   TRUE  1.0000000 0.0000000

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
```

```
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.14626, df = 1, p-value = 0.7021
##
## [1] "isCSO"
##
##         Decrease Increase
##   FALSE     1779     3756
##
##          Decrease  Increase
##   FALSE 0.3214092 0.6785908
##
##  Chi-squared test for given probabilities
##
## data:  dftb2
## X-squared = 706.15, df = 1, p-value < 2.2e-16
##
## [1] "isCTO"
##
##         Decrease Increase
##   FALSE     1778     3752
##   TRUE         1        4
##
##         Decrease Increase
##   FALSE 0.321519 0.678481
##   TRUE  0.200000 0.800000

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.010517, df = 1, p-value = 0.9183
##
## [1] "isEVP"
##
##         Decrease Increase
##   FALSE     1754     3702
##   TRUE        25       54
##
##          Decrease  Increase
##   FALSE 0.3214809 0.6785191
##   TRUE  0.3164557 0.6835443
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 2.4937e-27, df = 1, p-value = 1
##
## [1] "isPresident_vp"
##
##         Decrease Increase
```

```
##   FALSE     1744     3677
##   TRUE        35       79
##
##          Decrease  Increase
##   FALSE 0.3217119 0.6782881
##   TRUE  0.3070175 0.6929825
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.053428, df = 1, p-value = 0.8172
##
## [1] "isSecretary"
##
##         Decrease Increase
##   FALSE     1774     3748
##   TRUE         5        8
##
##          Decrease  Increase
##   FALSE 0.3212604 0.6787396
##   TRUE  0.3846154 0.6153846

## Warning in chisq.test(dftb2): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dftb2
## X-squared = 0.036581, df = 1, p-value = 0.8483
```

```
#reldf<-data.frame(Position=pos,PercentIncrease=rel)
#reldf[10,3]=0
#reldf
```

```
summary(df$transactionshares)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##         0       653      2575     86044     10000 689874047
```

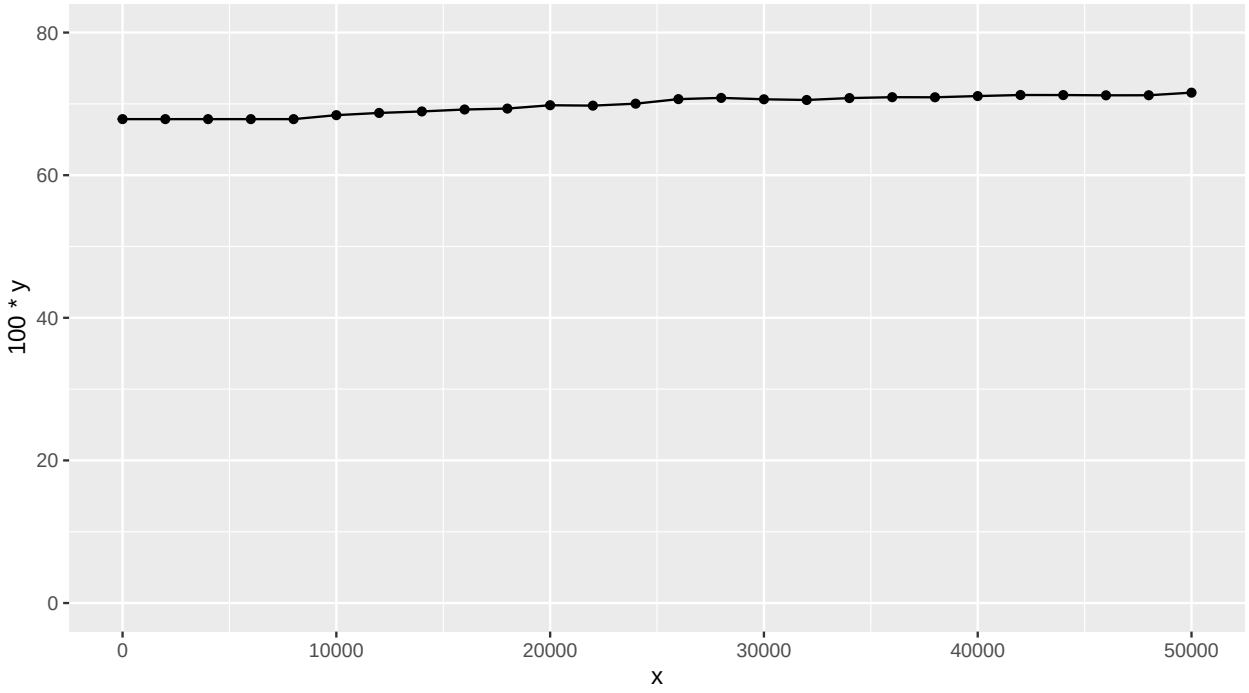
```
prop_list <- vector()
j= 0
for (i in seq(0,50000, 2000)){
  df_temp <- filter(df3, transactioncode %in% c("P") & transactionshares > i)
  dftable<-table(df_temp$transactioncode, df_temp$changeafteramonth)
  prop_list <- append(prop_list, as.numeric(prop.table(dftable,1))[[2]])
}

prop_data= data.frame(x=seq(0, 50000, 2000), y= prop_list)
prop_data
```

```
##        x         y
## 1      0 0.6785908
```

```
## 2   2000 0.6785908
## 3   4000 0.6785908
## 4   6000 0.6785908
## 5   8000 0.6785908
## 6  10000 0.6841668
## 7  12000 0.6872663
## 8  14000 0.6893816
## 9  16000 0.6921120
## 10 18000 0.6934075
## 11 20000 0.6980914
## 12 22000 0.6975216
## 13 24000 0.7002758
## 14 26000 0.7066622
## 15 28000 0.7083617
## 16 30000 0.7064286
## 17 32000 0.7054545
## 18 34000 0.7081162
## 19 36000 0.7093653
## 20 38000 0.7092338
## 21 40000 0.7109471
## 22 42000 0.7124374
## 23 44000 0.7123519
## 24 46000 0.7119965
## 25 48000 0.7120947
## 26 50000 0.7157385
```

```
ggplot(prop_data, aes(x=x, y= 100*y)) +
  geom_line() +
  geom_point() +
  ylim(0, 80)
```



```
prop_data= data.frame(x=seq(0, 50000, 2000), y= prop_list)
prop_data
```

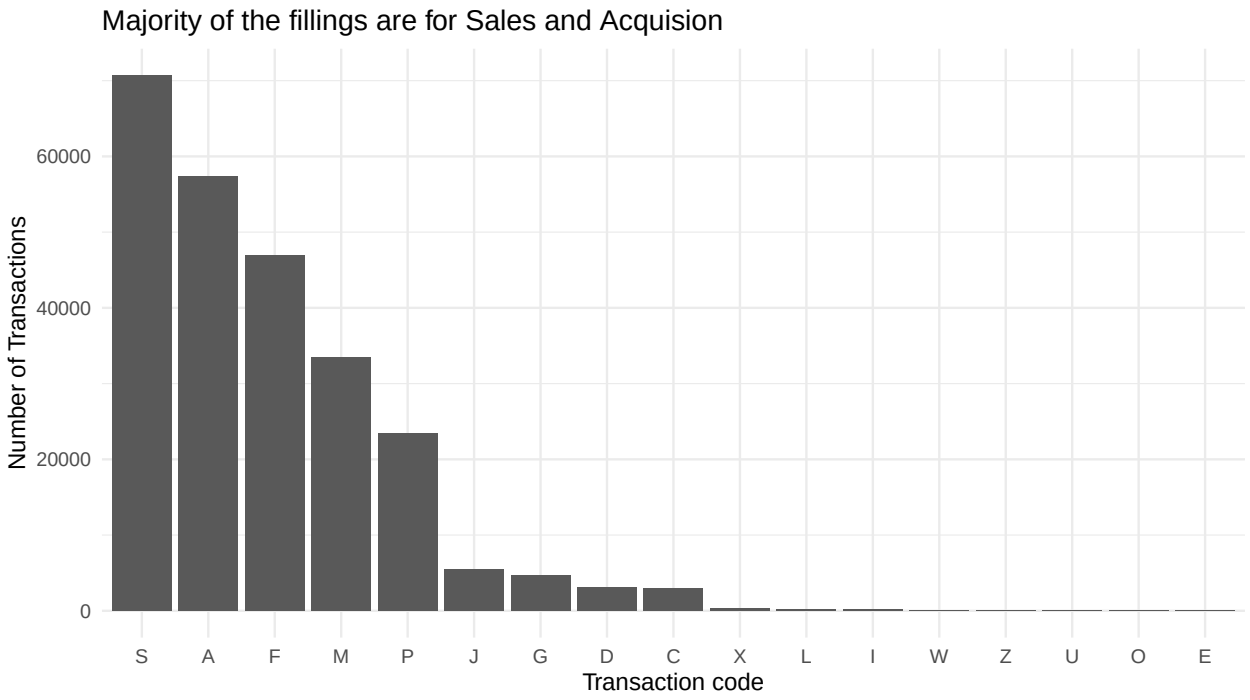
```
##        x         y
## 1      0 0.6785908
## 2   2000 0.6785908
## 3   4000 0.6785908
## 4   6000 0.6785908
## 5   8000 0.6785908
## 6  10000 0.6841668
## 7  12000 0.6872663
## 8  14000 0.6893816
## 9  16000 0.6921120
## 10 18000 0.6934075
## 11 20000 0.6980914
## 12 22000 0.6975216
## 13 24000 0.7002758
## 14 26000 0.7066622
## 15 28000 0.7083617
## 16 30000 0.7064286
## 17 32000 0.7054545
## 18 34000 0.7081162
## 19 36000 0.7093653
## 20 38000 0.7092338
## 21 40000 0.7109471
## 22 42000 0.7124374
## 23 44000 0.7123519
## 24 46000 0.7119965
## 25 48000 0.7120947
## 26 50000 0.7157385
```

```
png("shareinc.png", width= 960, height= 540, unit= "px")
ggplot(prop_data, aes(x=x, y= 100*y)) +
  geom_point() + labs(title="Proportion of shares that increase, increases with more shares per transac
dev.off()
```

```
## pdf
##   2
```

```
#png("plot1.png", width= 1280, height= 800, unit= "px")

ggplot(df2, aes(x= factor(transactioncode,levels= names(sort(table(transactioncode),
                                                             decreasing =  TRUE))))) +
  geom_bar() +
  theme_minimal() +
  labs(x= "Transaction code",
       y= "Number of Transactions",
       title= "Majority of the fillings are for Sales and Acquision")
```


Majority of the fillings are for Sales and Acquision
Number of Transactions
60000
40000
20000
0
S
A
F
M
P
J
G
D
C
X
L
I
W
Z
U
O
E
Transaction code

```
#dev.off()
```