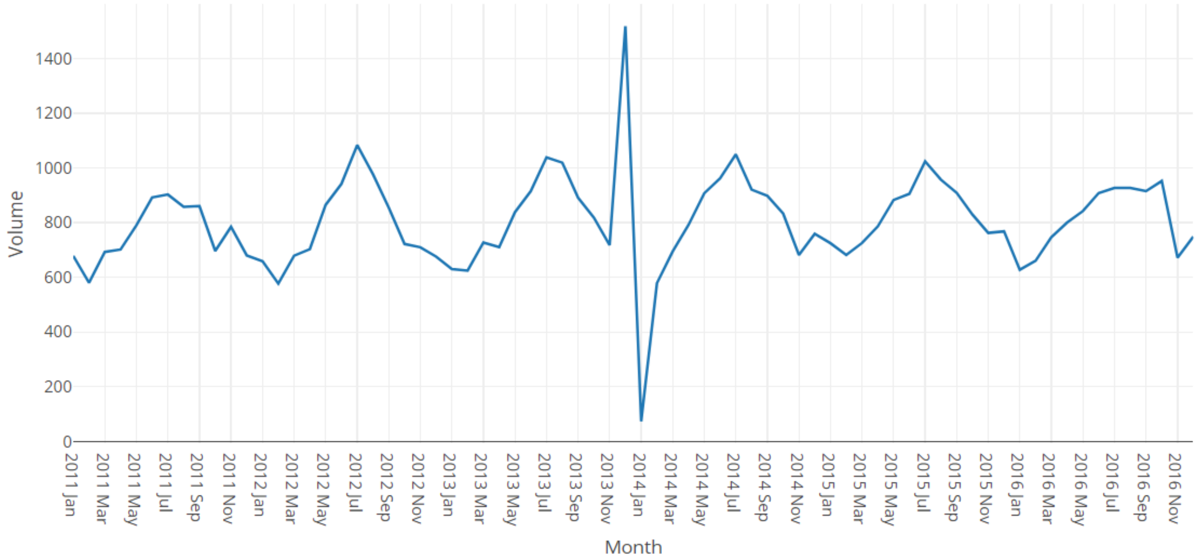
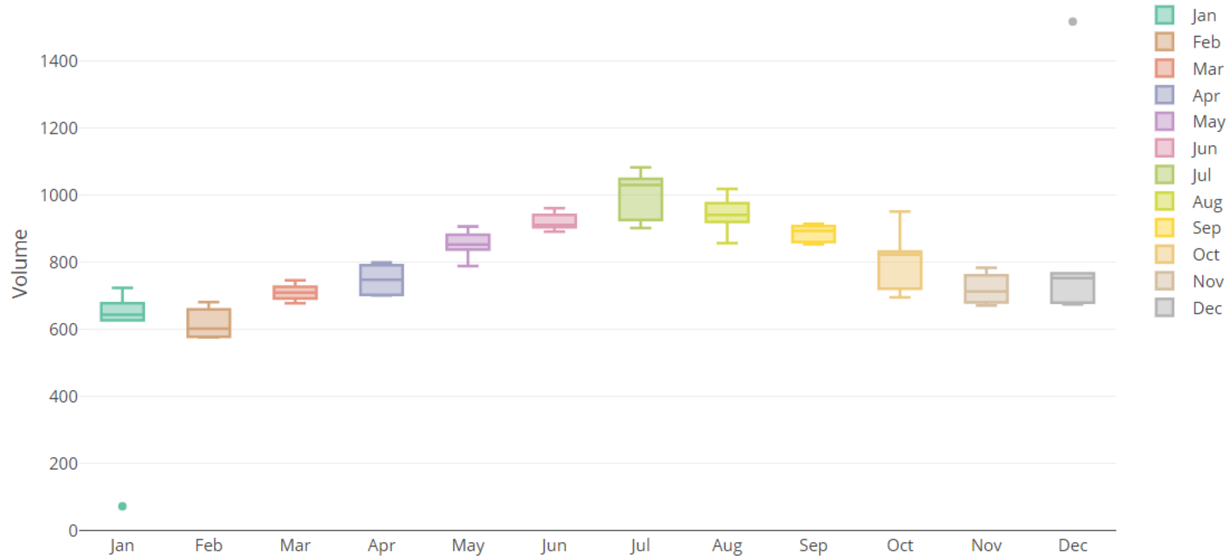
Outliers



Identification

- In the above example, outliers can be clearly identified by observing. However, observation cannot be scaled to thousands of records
- Common practices include checking percentiles and box plots

```
# Plotting the series
quantile(volume_ts, seq(0,1,by = 0.05))
##          0%          5%         10%         15%         20%         25%
##    73.17155   604.66837   658.51287   677.44359   681.49081   695.91958
##         30%         35%         40%         45%         50%         55%
##   710.03465   724.15420   747.55286   767.70870   790.53039   829.46234
##         60%         65%         70%         75%         80%         85%
##   849.55585   867.37696   895.95465   907.46436   914.80318   932.19810
##         90%         95%        100%
##   961.21952  1030.37763  1518.07595
# Creating boxplots to understand high and low variation months
data$Month1 <- factor(substr(as.character(data$Month), 6,
nchar(as.character(data$Month))), levels = month.abb)
plot_ly(data, x = ~Month1, y = ~Volume, type = "box", color = ~Month1) %>%
layout(autosize=F, width=900, height = 450, xaxis = list(title = ""))
```



- We see two clear outlier points in January and December
- A possible reason for this is an aritifical inventory push at the end of 2013 to meet shipment targets

Treatment

- Since the effect of shipments are cascading, exceptions have to be handled by redistributing volumes in proportion with historical values
- In the above case, January and December seem to be the only ones affected

```
proportion2012 <- c(data$Volume[12]/(data$Volume[12]+data$Volume[13]),
                    data$Volume[13]/(data$Volume[12]+data$Volume[13]))
proportion2012
## [1] 0.5080608 0.4919392
proportion2013 <- c(data$Volume[24]/(data$Volume[24]+data$Volume[25]),
                    data$Volume[25]/(data$Volume[24]+data$Volume[25]))
proportion2013
## [1] 0.51745 0.48255
proportion2014 <- c(data$Volume[36]/(data$Volume[36]+data$Volume[37]),
                    data$Volume[37]/(data$Volume[36]+data$Volume[37]))
proportion2014
## [1] 0.95401623 0.04598377
proportion2015 <- c(data$Volume[48]/(data$Volume[48]+data$Volume[49]),
                    data$Volume[49]/(data$Volume[48]+data$Volume[49]))
proportion2015
## [1] 0.5115344 0.4884656
```

- Re-proportioning 2013 December and 2014 January we arrive at the following plot :

```
total <- (data$Volume[36]+data$Volume[37])
data$Volume[36] <- total*0.51
data$Volume[37] <- total*0.49
```

```
plot_ly(data, x = ~Month, y = ~Volume, type = 'scatter', mode = 'lines') %>%
layout(autosize=F, width=900, height = 450, margin = list(b = 100))
```