Prediction of Online Product Sales

Dong Kim

March 4, 2018

**Load the data from csv file.**

#Load the Data set   
setwd("C:/Users/Dongsung/Documents/")  
OnlineProductSale <- read.csv("OnlineProductSale.csv",sep=",",header = TRUE)  
summary(OnlineProductSale)

## Outcome\_M1 Outcome\_M2 Outcome\_M3 Outcome\_M4   
## Min. : 2000 Min. : 500 Min. : 500 Min. : 500   
## 1st Qu.: 2000 1st Qu.: 2000 1st Qu.: 500 1st Qu.: 500   
## Median : 5000 Median : 3000 Median : 2000 Median : 2000   
## Mean : 20621 Mean : 10838 Mean : 5194 Mean : 3495   
## 3rd Qu.: 16000 3rd Qu.: 8000 3rd Qu.: 5000 3rd Qu.: 3000   
## Max. :795000 Max. :315000 Max. :108000 Max. :74000   
## NA's :3 NA's :5 NA's :5   
## Outcome\_M5 Outcome\_M6 Outcome\_M7 Outcome\_M8   
## Min. : 500 Min. : 500 Min. : 500 Min. : 500   
## 1st Qu.: 500 1st Qu.: 500 1st Qu.: 500 1st Qu.: 500   
## Median : 2000 Median : 2000 Median : 2000 Median : 500   
## Mean : 2852 Mean : 2018 Mean : 1751 Mean : 1493   
## 3rd Qu.: 2000 3rd Qu.: 2000 3rd Qu.: 2000 3rd Qu.: 2000   
## Max. :81000 Max. :27000 Max. :21000 Max. :21000   
## NA's :24 NA's :46 NA's :57 NA's :64   
## Outcome\_M9 Outcome\_M10 Outcome\_M11 Outcome\_M12   
## Min. : 500 Min. : 500 Min. : 500 Min. : 500   
## 1st Qu.: 500 1st Qu.: 500 1st Qu.: 500 1st Qu.: 500   
## Median : 500 Median : 500 Median : 500 Median : 500   
## Mean : 1411 Mean : 1321 Mean : 1191 Mean : 1072   
## 3rd Qu.: 2000 3rd Qu.: 2000 3rd Qu.: 2000 3rd Qu.: 2000   
## Max. :23000 Max. :23000 Max. :10000 Max. :10000   
## NA's :74 NA's :95 NA's :106 NA's :112   
## Date\_1 Date\_2 Quan\_1 Quan\_2   
## Min. :1420 Min. : 365 Min. : 31600 Min. :1600   
## 1st Qu.:2254 1st Qu.:1584 1st Qu.: 922700 1st Qu.:6100   
## Median :2737 Median :2234 Median :1864300 Median :7200   
## Mean :2703 Mean :2177 Mean :2015039 Mean :6987   
## 3rd Qu.:3220 3rd Qu.:2716 3rd Qu.:2789600 3rd Qu.:8000   
## Max. :3682 Max. :3508 Max. :5268500 Max. :9700   
## NA's :3 NA's :12   
## Quan\_3 Quan\_4 Quan\_5 Quan\_6   
## Min. : 1.00 Min. : 2 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 37.00 1st Qu.: 67368 1st Qu.: 4.00 1st Qu.: 1.000   
## Median : 60.00 Median : 245364 Median : 16.00 Median : 3.000   
## Mean : 62.17 Mean : 438943 Mean : 29.66 Mean : 8.603   
## 3rd Qu.: 87.00 3rd Qu.: 600178 3rd Qu.: 39.00 3rd Qu.:11.000   
## Max. :185.00 Max. :5417231 Max. :313.00 Max. :73.000   
## NA's :5 NA's :79 NA's :474 NA's :474   
## Quan\_7 Quan\_8 Quan\_9 Quan\_10   
## Min. : 0.00 Min. : 2.00 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.00 1st Qu.: 10.00 1st Qu.: 2.00 1st Qu.: 1.000   
## Median : 4.00 Median : 25.00 Median : 7.00 Median : 2.000   
## Mean : 10.43 Mean : 41.39 Mean : 15.97 Mean : 6.025   
## 3rd Qu.: 11.00 3rd Qu.: 43.00 3rd Qu.: 19.00 3rd Qu.: 8.000   
## Max. :273.00 Max. :544.00 Max. :288.00 Max. :46.000   
## NA's :474 NA's :522 NA's :474 NA's :474   
## Quan\_11 Quan\_12 Quan\_13 Quan\_14   
## Min. : 0.00 Min. : 2 Min. : 3 Min. : 2.000   
## 1st Qu.: 0.00 1st Qu.: 728 1st Qu.: 493 1st Qu.: 4.000   
## Median : 1.00 Median : 4096 Median : 3072 Median : 8.000   
## Mean : 3.56 Mean : 497433 Mean : 512329 Mean : 9.447   
## 3rd Qu.: 2.00 3rd Qu.: 31667 3rd Qu.: 16486 3rd Qu.: 12.000   
## Max. :265.00 Max. :7340032 Max. :16777216 Max. :256.000   
## NA's :474 NA's :245 NA's :11 NA's :308   
## Quan\_15 Quan\_16 Cat\_1 Cat\_2   
## Min. : 2.00 Min. : 2.00 Min. :1.000 Min. :0.0000   
## 1st Qu.: 8.00 1st Qu.: 8.00 1st Qu.:1.000 1st Qu.:0.0000   
## Median : 24.00 Median : 26.00 Median :1.000 Median :0.0000   
## Mean : 52.81 Mean : 55.49 Mean :1.033 Mean :0.2117   
## 3rd Qu.: 43.25 3rd Qu.: 46.00 3rd Qu.:1.000 3rd Qu.:0.0000   
## Max. :1039.00 Max. :1039.00 Max. :2.000 Max. :1.0000   
## NA's :407 NA's :438   
## Cat\_3 Cat\_4 Cat\_5 Cat\_6   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :1.0000 Median :0.0000 Median :0.0000   
## Mean :0.9933 Mean :0.8469 Mean :0.1704 Mean :0.3036   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## Cat\_7 Cat\_8 Cat\_9 Cat\_10 Cat\_11   
## Min. :1 Min. :1 Min. :0.0000 Min. :0 Min. :0.0000   
## 1st Qu.:1 1st Qu.:1 1st Qu.:0.0000 1st Qu.:0 1st Qu.:0.0000   
## Median :1 Median :1 Median :0.0000 Median :0 Median :0.0000   
## Mean :1 Mean :1 Mean :0.1212 Mean :0 Mean :0.3515   
## 3rd Qu.:1 3rd Qu.:1 3rd Qu.:0.0000 3rd Qu.:0 3rd Qu.:1.0000   
## Max. :1 Max. :1 Max. :1.0000 Max. :0 Max. :1.0000   
##   
## Cat\_12 Cat\_13 Cat\_14 Cat\_15   
## Min. :0.0000 Min. :1 Min. :0.0000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:1 1st Qu.:0.0000 1st Qu.:0.00000   
## Median :0.0000 Median :1 Median :0.0000 Median :0.00000   
## Mean :0.3023 Mean :1 Mean :0.2783 Mean :0.09587   
## 3rd Qu.:1.0000 3rd Qu.:1 3rd Qu.:1.0000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1 Max. :1.0000 Max. :1.00000   
##   
## Cat\_16 Cat\_17 Cat\_18 Cat\_19   
## Min. :0.00000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.0000   
## Median :0.00000 Median :0.0000 Median :0.0000 Median :1.0000   
## Mean :0.01332 Mean :0.1704 Mean :0.2597 Mean :0.9228   
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :2.00000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
##   
## Cat\_20   
## Min. :0.0000   
## 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.4154   
## 3rd Qu.:1.0000   
## Max. :1.0000   
##

**Dataset:** 751 Observations and 50 variables

**List of Libraries:**

library(tidyr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(plyr)

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

library(ridge)  
library(glmnet)

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':  
##   
## expand

## Loading required package: foreach

## Loaded glmnet 2.0-13

library(broom)  
library(tidyverse)

## -- Attaching packages --------- tidyverse 1.2.1 --

## v tibble 1.4.2 v purrr 0.2.4  
## v readr 1.1.1 v stringr 1.2.0  
## v tibble 1.4.2 v forcats 0.3.0

## -- Conflicts ------------ tidyverse\_conflicts() --  
## x purrr::accumulate() masks foreach::accumulate()  
## x plyr::arrange() masks dplyr::arrange()  
## x purrr::compact() masks plyr::compact()  
## x plyr::count() masks dplyr::count()  
## x Matrix::expand() masks tidyr::expand()  
## x plyr::failwith() masks dplyr::failwith()  
## x dplyr::filter() masks stats::filter()  
## x plyr::id() masks dplyr::id()  
## x dplyr::lag() masks stats::lag()  
## x plyr::mutate() masks dplyr::mutate()  
## x plyr::rename() masks dplyr::rename()  
## x plyr::summarise() masks dplyr::summarise()  
## x plyr::summarize() masks dplyr::summarize()  
## x purrr::when() masks foreach::when()

**CLEANING THE DATA**

**Adding a new column to the Data Set**

* Each row represents different consumer product but added new column called “Product\_Number” in the front of the data set.

OnlineProductSale["Product\_Number"] <- NA  
OnlineProductSale$Product\_Number <- 1:nrow(OnlineProductSale)  
OnlineProductSale <- OnlineProductSale[colnames(OnlineProductSale)[c(51,1:50)]]

**Variables with Missing Values**

* There are many missing values “Nan” throughout the data set in every columns.
* Replaced all the “Nan” with mean of each column

for(i in 1:ncol(OnlineProductSale))  
{  
 OnlineProductSale[is.na(OnlineProductSale[,i]), i] <- colMeans(OnlineProductSale[,i, drop = FALSE], na.rm = TRUE)  
 OnlineProductSale[]<-lapply(OnlineProductSale, as.integer)  
}

**Saving Cleaned Data Set**

write.csv(OnlineProductSale,"OnlineProductSale(Clean).csv",row.names = FALSE)

* Separately saved the cleaned dataset for better visualization of the data.

**Statistical Analysis**

* Calculated the minimum, median, maximum, and mean of outcome columns

dfnew1 <-OnlineProductSale[,c(2,3,4,5,6,7,8,9,10,11,12,13)]  
minimum\_outcome <- apply(dfnew1,2,min)  
minimum\_outcome

## Outcome\_M1 Outcome\_M2 Outcome\_M3 Outcome\_M4 Outcome\_M5 Outcome\_M6   
## 2000 500 500 500 500 500   
## Outcome\_M7 Outcome\_M8 Outcome\_M9 Outcome\_M10 Outcome\_M11 Outcome\_M12   
## 500 500 500 500 500 500

median\_outcome <- apply(dfnew1,2,median)  
median\_outcome

## Outcome\_M1 Outcome\_M2 Outcome\_M3 Outcome\_M4 Outcome\_M5 Outcome\_M6   
## 5000 3000 2000 2000 2000 2000   
## Outcome\_M7 Outcome\_M8 Outcome\_M9 Outcome\_M10 Outcome\_M11 Outcome\_M12   
## 1751 500 500 500 500 500

maximum\_outcome <- apply(dfnew1,2,max)  
maximum\_outcome

## Outcome\_M1 Outcome\_M2 Outcome\_M3 Outcome\_M4 Outcome\_M5 Outcome\_M6   
## 795000 315000 108000 74000 81000 27000   
## Outcome\_M7 Outcome\_M8 Outcome\_M9 Outcome\_M10 Outcome\_M11 Outcome\_M12   
## 21000 21000 23000 23000 10000 10000

mean\_outcome <- apply(dfnew1,2,mean)  
mean\_outcome

## Outcome\_M1 Outcome\_M2 Outcome\_M3 Outcome\_M4 Outcome\_M5 Outcome\_M6   
## 20620.506 10837.565 5194.368 3494.634 2852.128 2018.413   
## Outcome\_M7 Outcome\_M8 Outcome\_M9 Outcome\_M10 Outcome\_M11 Outcome\_M12   
## 1751.407 1492.660 1411.337 1320.772 1190.599 1071.840

* Minimum outcome of each sales months are represented. Outcome\_M1 has the highest mimimum sales outcome compared to other months.
* Outcome\_M1 has the products that were sold the most throughout the year.
* Outcome\_M1 has the highest average outcome.

**Statistical Analysis**

* What products are being sold below median throughout the year? Low Outcome Products

minim1 <- data.frame(OnlineProductSale$Product\_Number)  
LowOutcome\_Prod <-minim1[OnlineProductSale$Outcome\_M1<=5000 & OnlineProductSale$Outcome\_M2<=3000   
 & OnlineProductSale$Outcome\_M3<=2000 & OnlineProductSale$Outcome\_M4<=2000  
 & OnlineProductSale$Outcome\_M5<=2000 & OnlineProductSale$Outcome\_M6<=3000  
 & OnlineProductSale$Outcome\_M7<=1751 & OnlineProductSale$Outcome\_M8<=500   
 & OnlineProductSale$Outcome\_M9<=500 & OnlineProductSale$Outcome\_M10<=500   
 & OnlineProductSale$Outcome\_M11<=500 & OnlineProductSale$Outcome\_M12<=500,]  
LowOutcome\_Prod

## [1] 8 10 12 14 16 18 21 23 30 39 47 50 67 72 73 74 75  
## [18] 76 80 84 87 90 102 103 105 106 116 117 118 119 120 121 122 126  
## [35] 127 129 133 135 139 140 161 170 175 177 183 184 186 189 190 191 192  
## [52] 207 209 214 215 217 224 225 229 234 237 240 241 250 253 257 266 273  
## [69] 280 287 297 299 310 311 312 316 321 328 329 330 351 362 365 367 371  
## [86] 385 387 397 407 409 412 423 438 439 452 456 459 474 478 480 483 484  
## [103] 486 495 498 499 504 508 516 526 527 528 530 533 539 540 541 543 546  
## [120] 547 551 552 554 555 561 564 566 572 574 575 580 581 584 588 593 597  
## [137] 599 600 610 611 612 615 616 621 624 625 626 628 632 636 637 640 644  
## [154] 646 647 653 656 657 660 668 670 676 677 679 685 686 687 693 694 695  
## [171] 702 703 714 733 743

* Total of 175 products were sold below median
* What products are being sold above median throughout the year? BEST SELLERS: High Outcome Products

HighOutcome\_Prod <-minim1[OnlineProductSale$Outcome\_M1>=5000 & OnlineProductSale$Outcome\_M2>=3000   
 & OnlineProductSale$Outcome\_M3>=2000 & OnlineProductSale$Outcome\_M4>=2000  
 & OnlineProductSale$Outcome\_M5>=2000 & OnlineProductSale$Outcome\_M6>=3000  
 & OnlineProductSale$Outcome\_M7>=1751 & OnlineProductSale$Outcome\_M8>=500   
 & OnlineProductSale$Outcome\_M9>=500 & OnlineProductSale$Outcome\_M10>=500   
 & OnlineProductSale$Outcome\_M11>=500 & OnlineProductSale$Outcome\_M12>=500,]  
HighOutcome\_Prod

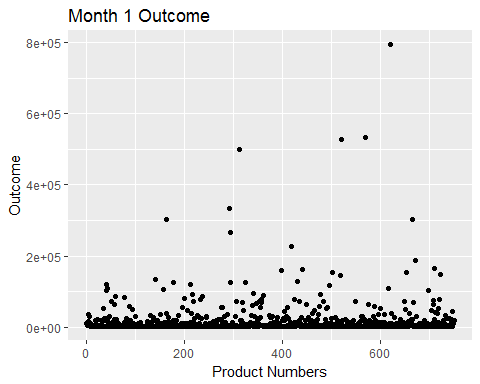
## [1] 4 32 36 41 43 44 57 58 59 83 85 89 91 94 101 109 131  
## [18] 137 142 146 157 164 165 176 179 182 197 206 216 218 220 222 231 238  
## [35] 252 260 276 294 295 298 307 313 320 322 324 325 340 341 344 348 353  
## [52] 355 356 357 358 361 391 395 404 411 417 419 440 442 460 462 465 477  
## [69] 479 485 493 496 503 513 518 520 538 542 545 549 562 563 573 578 587  
## [86] 591 601 603 609 619 620 630 641 643 648 650 654 655 659 661 664 666  
## [103] 673 689 690 698 705 708 709 710 711 719 722 729 734 735 739

* Total of 117 products were sold above median throughout the year.

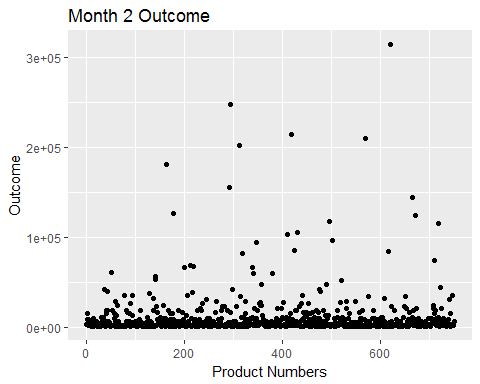
**Data Visualization**

* Graphs of each sales months

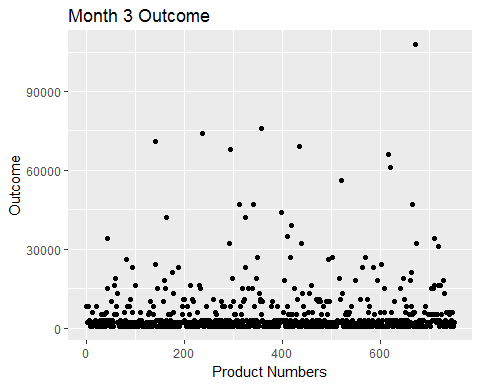
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M1))+geom\_point()+  
 labs(title= "Month 1 Outcome",x = "Product Numbers", y = "Outcome")



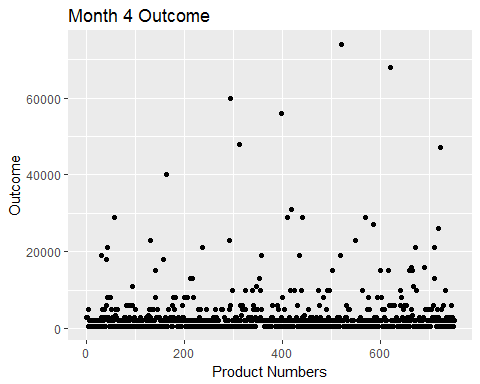
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M2))+geom\_point()+  
 labs(title= "Month 2 Outcome",x = "Product Numbers", y = "Outcome")



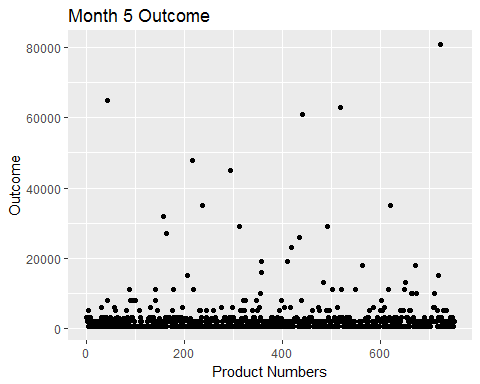
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M3))+geom\_point()+  
 labs(title= "Month 3 Outcome",x = "Product Numbers", y = "Outcome")



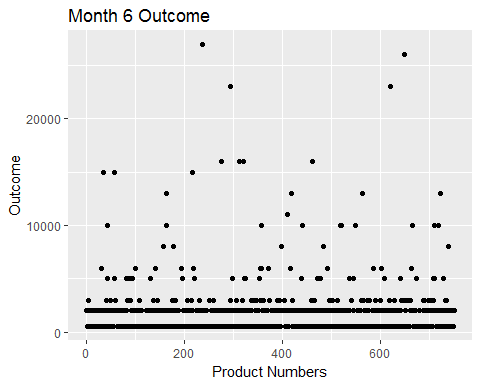
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M4))+geom\_point()+  
 labs(title= "Month 4 Outcome",x = "Product Numbers", y = "Outcome")



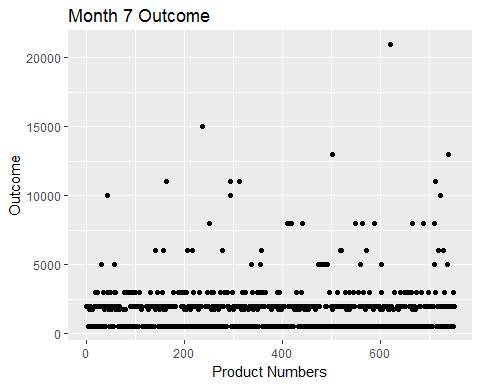
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M5))+geom\_point()+  
 labs(title= "Month 5 Outcome",x = "Product Numbers", y = "Outcome")



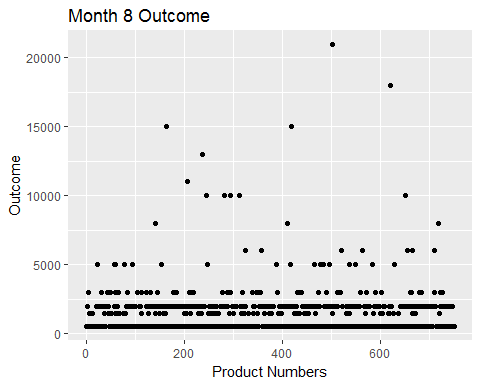
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M6))+geom\_point()+  
 labs(title= "Month 6 Outcome",x = "Product Numbers", y = "Outcome")



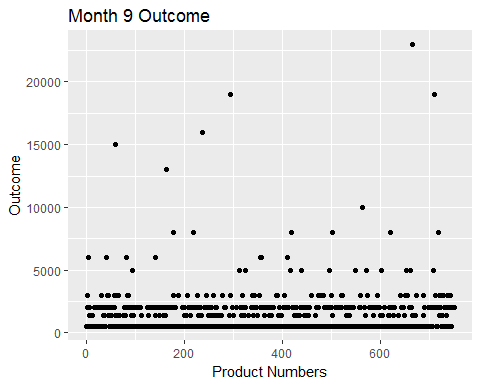
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M7))+geom\_point()+  
 labs(title= "Month 7 Outcome",x = "Product Numbers", y = "Outcome")



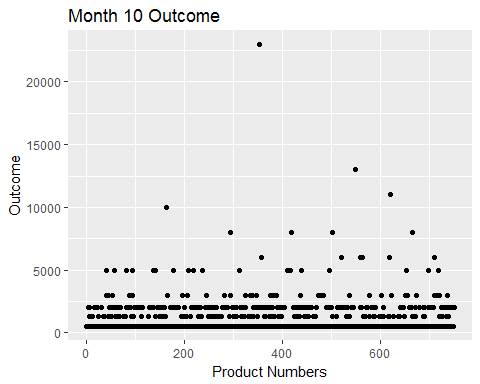
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M8))+geom\_point()+  
 labs(title= "Month 8 Outcome",x = "Product Numbers", y = "Outcome")



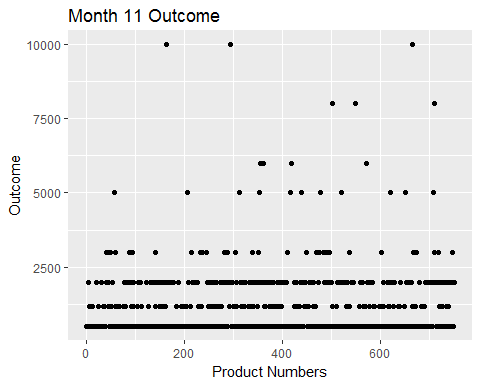
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M9))+geom\_point()+  
 labs(title= "Month 9 Outcome",x = "Product Numbers", y = "Outcome")



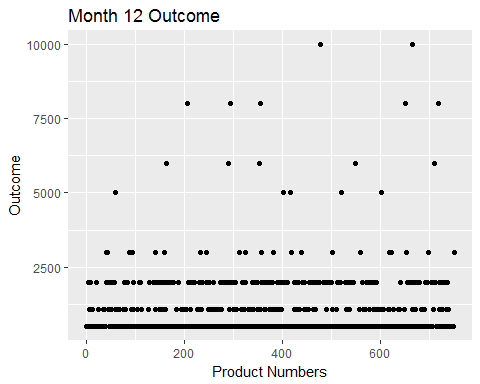
ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M10))+geom\_point()+  
 labs(title= "Month 10 Outcome",x = "Product Numbers", y = "Outcome")



ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M11))+geom\_point()+  
 labs(title= "Month 11 Outcome",x = "Product Numbers", y = "Outcome")



ggplot(OnlineProductSale,aes(x=Product\_Number, y=Outcome\_M12))+geom\_point()+  
 labs(title= "Month 12 Outcome",x = "Product Numbers", y = "Outcome")



* Compared to low and high values, it’s very challenging to find linear for these graphs with 751 products. Most of the plots were at very bottom of the graph. High selling products are highly scattered compared to low selling products. High selling products are more distinguishable compare to other products (outliers). Limitation for these graphs are there are too many products in the dataset which it is very challenging to distinguish between different products.

**Filtering Advertising Campaign**

* Calculated minimum, maximum, median, and mean of advertising campaign days.
* Date\_1 is the day number the major advertising campaign began and the product launched.
* Date\_2 is the day number the product was announced and a pre-release advertising campaign began.

dfnew2 <-OnlineProductSale[,c(14,15)]  
minimum\_adver <- apply(dfnew2,2,min)  
minimum\_adver

## Date\_1 Date\_2   
## 1420 365

max\_adver <- apply(dfnew2,2,max)  
max\_adver

## Date\_1 Date\_2   
## 3682 3508

median\_adver <-apply(dfnew2,2,median)  
median\_adver

## Date\_1 Date\_2   
## 2737 2233

mean\_adver <-apply(dfnew2,2,mean)  
mean\_adver

## Date\_1 Date\_2   
## 2702.860 2177.269

* Minimum: Date\_1: 1420 Date\_2:365 \*Maximum: Date\_1: 3682 Date\_2:3508
* Median: Date\_1:2737 Date\_2:2233
* Mean: Date\_1:2702.9 Date\_2:2177.3
* Filtered what products have been majorly advertised and launched (Date\_1) the shortest based on median number.

minim1 <- data.frame(OnlineProductSale$Product\_Number)  
Least\_adver <-minim1[OnlineProductSale$Date\_1<2737,]   
Least\_adver

## [1] 1 2 3 4 12 14 15 18 19 20 21 22 23 24 26 28 29  
## [18] 32 34 36 38 41 42 44 45 47 48 53 54 55 56 58 64 69  
## [35] 70 71 75 79 85 86 88 89 91 93 94 95 98 99 101 102 107  
## [52] 108 109 110 112 115 117 123 124 125 126 129 132 133 135 136 137 138  
## [69] 142 146 147 150 151 157 160 161 163 165 166 167 169 170 172 173 174  
## [86] 176 178 180 181 182 183 187 189 190 191 192 193 194 198 199 203 210  
## [103] 211 212 215 218 221 222 224 226 228 234 235 237 241 242 245 246 247  
## [120] 248 253 257 259 260 264 265 266 267 268 273 277 278 280 282 284 285  
## [137] 286 288 290 291 294 295 297 299 300 302 303 304 305 307 308 310 314  
## [154] 315 321 324 325 327 329 330 331 333 336 337 339 340 343 345 346 347  
## [171] 350 351 352 353 354 356 357 358 360 362 364 367 370 371 372 373 376  
## [188] 378 381 382 386 387 390 391 393 394 395 404 405 408 409 410 412 415  
## [205] 416 417 418 419 421 422 423 424 426 427 429 430 432 434 435 437 439  
## [222] 442 443 446 447 448 451 452 457 458 459 460 461 465 469 471 472 473  
## [239] 476 478 480 482 484 486 488 489 490 493 495 497 498 500 504 505 506  
## [256] 507 508 509 512 515 518 519 521 523 524 527 528 529 530 534 538 539  
## [273] 543 544 545 547 549 550 552 553 555 556 559 563 564 570 571 573 575  
## [290] 577 578 582 583 584 588 590 591 596 599 600 602 603 608 609 610 612  
## [307] 613 615 619 622 628 629 630 631 632 633 636 640 641 643 648 649 652  
## [324] 654 655 658 660 661 664 666 667 668 669 673 674 675 679 680 681 685  
## [341] 686 692 694 695 696 697 698 700 701 702 704 705 706 707 708 711 712  
## [358] 714 715 716 723 730 732 735 738 740 741 742 743 745 746 748 749 750  
## [375] 751

* Total of 375 products were advertised less than the median number of days.

Most\_adver <-minim1[OnlineProductSale$Date\_1>2737,]   
Most\_adver

## [1] 5 6 7 8 9 10 11 13 16 17 25 27 30 31 33 35 37  
## [18] 39 40 43 46 49 50 51 52 57 59 60 61 62 63 65 66 67  
## [35] 68 72 73 74 76 77 78 80 81 82 83 84 87 90 92 96 97  
## [52] 100 103 104 105 106 111 113 114 116 118 119 120 121 122 127 128 130  
## [69] 131 134 139 140 141 143 145 148 149 152 153 154 155 156 158 159 162  
## [86] 164 168 171 175 177 179 184 185 186 188 195 196 197 200 201 202 204  
## [103] 205 206 207 208 209 213 214 216 217 219 220 223 225 227 229 230 231  
## [120] 232 233 236 238 239 240 243 244 249 250 251 252 254 255 256 258 261  
## [137] 262 263 269 270 271 272 274 275 276 279 281 283 287 289 292 293 296  
## [154] 298 301 306 309 311 312 313 316 317 318 319 320 322 323 326 328 332  
## [171] 334 335 338 341 342 344 348 349 355 359 361 363 365 366 368 369 374  
## [188] 375 377 379 380 383 384 385 388 389 392 396 397 398 399 400 401 402  
## [205] 403 406 407 411 413 414 420 425 428 431 433 436 438 440 441 444 445  
## [222] 449 450 453 454 455 456 462 463 464 466 467 468 470 474 475 477 479  
## [239] 481 483 485 487 491 492 494 496 499 501 502 503 510 511 513 514 516  
## [256] 517 520 522 525 526 531 532 533 535 536 537 540 541 542 546 548 551  
## [273] 554 557 558 560 561 562 565 566 567 568 569 572 574 579 580 581 585  
## [290] 586 587 589 592 593 594 595 597 598 601 604 605 606 607 611 614 616  
## [307] 617 618 620 621 623 624 625 626 627 634 635 637 638 639 642 644 645  
## [324] 646 647 650 651 653 656 657 659 662 663 665 670 671 672 676 677 678  
## [341] 682 683 684 687 688 689 690 691 693 699 703 709 710 717 718 719 720  
## [358] 721 722 724 725 726 727 728 729 731 733 734 736 737 739 744 747

* Total of 373 products were advertised more than the median number of days.
* About 1:1 ratio of products were advertised more than another products
* Filtered what products have been announced and pre-release advertised (Date\_2) the shortest.

minim1 <- data.frame(OnlineProductSale$Product\_Number)  
LeastPre\_adver <-minim1[OnlineProductSale$Date\_2<2233,]   
LeastPre\_adver

## [1] 1 2 3 4 5 9 10 12 14 15 16 18 20 21 22 23 24  
## [18] 27 28 29 32 34 36 38 41 42 44 45 47 48 53 54 55 56  
## [35] 58 63 64 69 70 71 75 78 79 81 85 86 88 89 90 91 93  
## [52] 94 97 98 99 101 107 108 109 111 112 117 122 123 124 125 129 132  
## [69] 133 135 137 138 142 145 146 147 150 151 157 161 163 165 166 167 169  
## [86] 171 172 173 176 178 180 181 182 183 184 189 190 192 193 194 198 199  
## [103] 201 203 204 210 212 215 218 222 226 228 232 234 235 236 237 242 244  
## [120] 245 246 247 248 250 257 259 265 266 267 268 273 274 277 278 279 280  
## [137] 281 282 284 285 286 288 289 290 291 294 295 297 299 300 304 305 307  
## [154] 308 310 314 315 324 325 326 327 329 330 331 333 334 335 336 337 338  
## [171] 339 340 343 344 345 346 347 348 350 351 352 353 354 356 357 358 360  
## [188] 362 363 366 370 371 372 373 376 379 386 387 388 390 391 393 395 403  
## [205] 404 405 408 409 410 412 413 415 417 418 419 421 422 423 424 426 427  
## [222] 429 430 432 435 437 439 442 443 444 446 448 451 452 456 460 461 465  
## [239] 467 470 471 472 473 476 477 478 482 486 488 490 493 495 497 500 503  
## [256] 504 505 507 508 509 512 513 515 516 518 519 521 523 524 526 528 530  
## [273] 534 538 539 540 543 544 545 547 548 549 550 553 555 556 559 563 565  
## [290] 568 570 571 573 575 577 578 580 584 588 590 591 596 599 600 601 603  
## [307] 608 611 612 615 617 619 621 623 629 630 631 633 640 641 643 648 649  
## [324] 652 654 655 657 658 660 666 667 668 669 672 673 674 675 676 678 679  
## [341] 680 681 685 692 696 697 698 701 702 704 705 706 707 708 711 714 715  
## [358] 716 720 721 723 729 735 737 738 740 741 742 743 745 746 748 749 750

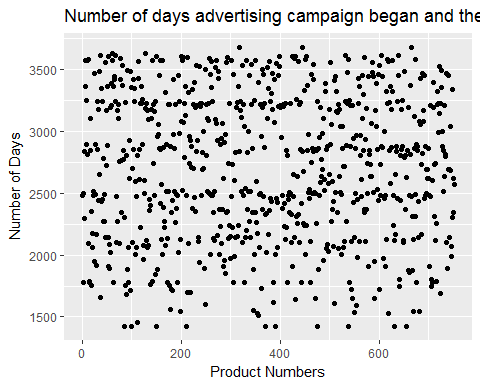
* Filtered what products have been announced and pre-release advertised (Date\_2) the shortest.

minim1 <- data.frame(OnlineProductSale$Product\_Number)  
MostPre\_adver <-minim1[OnlineProductSale$Date\_2>2233,]   
MostPre\_adver

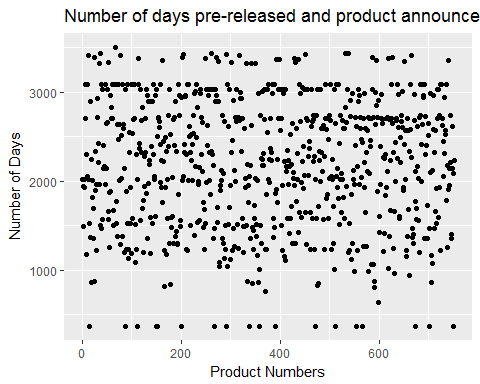
## [1] 6 7 8 11 13 17 19 25 26 30 31 33 35 37 39 40 43  
## [18] 46 49 50 51 52 57 59 60 61 62 65 66 67 68 72 73 74  
## [35] 76 77 80 82 83 84 87 92 95 96 100 102 103 104 105 106 110  
## [52] 113 114 115 116 118 119 120 121 126 127 128 130 131 134 136 139 140  
## [69] 141 143 144 148 149 152 153 154 155 156 158 159 160 162 164 168 170  
## [86] 174 175 177 179 185 186 187 188 191 195 196 197 200 202 205 206 207  
## [103] 208 209 211 213 214 216 217 219 220 221 223 224 225 227 229 230 231  
## [120] 233 238 239 240 241 243 249 251 252 253 254 255 256 258 260 261 262  
## [137] 263 264 269 270 271 272 275 276 283 287 292 293 296 298 301 302 303  
## [154] 306 309 311 312 313 316 317 318 319 320 321 322 323 328 332 341 342  
## [171] 349 355 359 361 364 365 367 368 369 374 375 377 378 380 381 382 384  
## [188] 385 389 392 394 396 397 398 399 400 401 402 406 407 411 414 416 420  
## [205] 425 428 431 433 434 436 438 440 441 445 447 449 450 453 454 455 457  
## [222] 458 459 462 463 464 466 468 469 474 475 480 481 483 484 485 487 489  
## [239] 491 492 494 496 498 499 501 502 506 510 511 514 517 520 522 525 527  
## [256] 529 531 532 533 535 536 537 541 542 546 551 552 554 557 558 560 561  
## [273] 562 564 566 567 569 572 574 576 579 581 582 583 585 586 587 589 592  
## [290] 593 594 595 597 598 602 604 605 606 607 609 610 613 614 616 618 620  
## [307] 624 625 626 627 628 632 634 635 636 637 638 639 642 644 645 646 647  
## [324] 650 651 653 656 659 661 662 663 664 665 670 671 677 682 683 684 686  
## [341] 687 688 689 690 691 693 694 695 699 700 703 709 710 712 713 717 718  
## [358] 719 722 724 725 726 727 728 730 731 732 733 734 736 739 744 747 751

* Scatterplot of High seller products advertising campaign

ggplot(OnlineProductSale,aes(x=Product\_Number, y=Date\_1))+geom\_point()+  
 labs(title= "Number of days advertising campaign began and the product launched",x = "Product Numbers", y = "Number of Days")



ggplot(OnlineProductSale,aes(x=Product\_Number, y=Date\_2))+geom\_point()+  
 labs(title= "Number of days pre-released and product announced",x = "Product Numbers", y = "Number of Days")



* Both pre-release advertising and post advertising not comparable. Number of days were very different from each products.
* From the figures above, it was found that products were advertised longer period of time when the products were launched than number of days which products were pre-released and products were announced.
* Overall, more advertisements before the products were launched had more effect on the sales outcome.