HW3

Mandi Zhu (U49566596)

2/23/2020

The total number of orders by last-touch ("converter") and by first-touch ("originator") attribution

The corresponding share of credit for the two attribution models

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------------

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts -------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

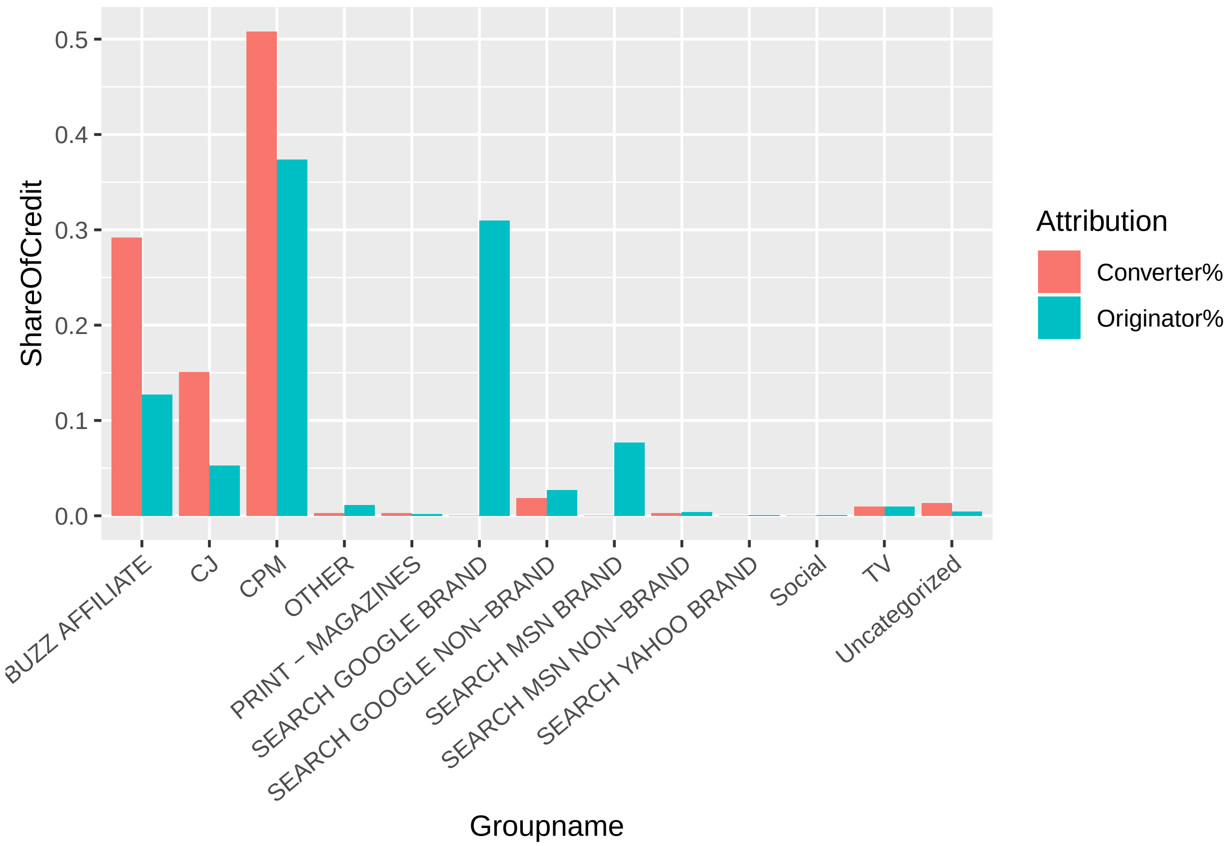
```r
load("Winters_Attribution-W20-MSBA-MW.RData")
data1<-data[data$Positionname=='CONVERTER'|data$Positionname=='ORIGINATOR',]
data1<-data1 %>% group_by(Groupname,Positionname) %>% count()
data2<-data1 %>% pivot_wider(names_from=Positionname,values_from=n)
data2[is.na(data2)]<-0
data2<-data2 %>% mutate('Converter%' = CONVERTER/sum(data2$CONVERTER),
                 'Originator%' = ORIGINATOR/sum(data2$ORIGINATOR))


Groupname='TOTAL'
CONVERTER=sum(data2$CONVERTER,na.rm = T)
ORIGINATOR=sum(data2$ORIGINATOR,na.rm = T)
Converter_per=sum(data2$`Converter%`,na.rm = T)
Originator_per=sum(data2$`Originator%`,na.rm = T)
total=data.frame(Groupname,CONVERTER,ORIGINATOR,Converter_per,Originator_per)
colnames(total)<-c('Groupname','CONVERTER','ORIGINATOR','Converter%','Originator%')
data2<-bind_rows(data2,total)
data2<-data2[c(1:5,7,9,12:13,6,8,10:11,14),]

as.table(as.matrix(data2))
```

```
##   Groupname                CONVERTER ORIGINATOR Converter%  Originator%
## A BUZZ AFFILIATE            443       193        0.291831357 0.1271409750
## B CJ                        229        80        0.150856390 0.0527009223
## C CPM                       771       567        0.507905138 0.3735177866
## D OTHER                       4        17        0.002635046 0.0111989460
## E PRINT - MAGAZINES           4         3        0.002635046 0.0019762846
## F SEARCH GOOGLE NON-BRAND    28        41        0.018445323 0.0270092227
## G SEARCH MSN NON-BRAND        4         6        0.002635046 0.0039525692
## H TV                         15        15        0.009881423 0.0098814229
## I Uncategorized              20         7        0.013175231 0.0046113307
```

```
## J SEARCH GOOGLE BRAND        0       470       0.000000000 0.3096179183
## K SEARCH MSN BRAND           0       117       0.000000000 0.0770750988
## L SEARCH YAHOO BRAND         0         1       0.000000000 0.0006587615
## M Social                     0         1       0.000000000 0.0006587615
## N TOTAL                   1518      1518       1.000000000 1.0000000000
```

The share of credit (in percentage) for the first- and last-touch attribution models by marketing channel

```
data2b<-data2 %>% pivot_longer(cols=c(`Converter%`,`Originator%`),names_to ='Attribution',values_to = 'S
data2b<-data2b[c(-27,-28),]
ggplot(data2b,aes(x=Groupname,y=ShareOfCredit,fill=Attribution)) +
  geom_col(position='dodge') +
  theme(axis.text.x = element_text(angle = 40, hjust = 1))
```



Compare and contrast the two attribution model results

Last-touch (converter credit) attribution indicates that almost all the credit goes to display ads (CPM) and affiliates (Buzz Affiliates and CJ). In particular, last-touch attribution gives little credit to search. By contrast, first-touch attribution credits the search channel (Google & MSN, especially brand search) as playing an important role. Still, display ads (CPM) are the most important channel for both attribution methods. Thus, the data indicates that search plays a more important role at the beginning of the customer journey and both display and affiliate play an important role both at the beginning and end of the journey.

If Winters relied solely on last-touch attribution, it would overlook the role of search and under-invest in search advertising.

The average number of days that it takes for a new/old customer to convert

```
dataQ2a<-data %>% filter(Positionname=='ORIGINATOR') %>%
  select(c(Positionname,Newcustomer,DaysToConvert))
dataQ2a<-dataQ2a %>% group_by(Newcustomer) %>% summarise(mean_DaysToConvert=mean(DaysToConvert))
dataQ2a
```

```
## # A tibble: 2 x 2
##   Newcustomer mean_DaysToConvert
##   <chr>                    <dbl>
## 1 N                        29.0
## 2 Y                         6.14
```

The average number of touchpoints by new versus old customer's orders

```
dataQ2b<-data %>% add_count(Orderid) %>% select(Orderid,n,Newcustomer)
dataQ2b<-dataQ2b[!duplicated(dataQ2b),]
dataQ2b<-dataQ2b %>% group_by(Newcustomer) %>% summarise(mean_Touches=mean(n))
dataQ2b
```

```
## # A tibble: 2 x 2
##   Newcustomer mean_Touches
##   <chr>              <dbl>
## 1 N                   5.18
## 2 Y                   4.25
```

The average order sales amount by new versus old customer's orders

```
dataQ2c<-data %>% select(Orderid,Newcustomer,Saleamount)
dataQ2c<-dataQ2c[!duplicated(dataQ2c),] %>%
  group_by(Newcustomer) %>% summarise(mean_sales=mean(Saleamount))
dataQ2c
```

```
## # A tibble: 2 x 2
##   Newcustomer mean_sales
##   <chr>            <dbl>
## 1 N                 207.
## 2 Y                 272.
```

How new and old customers differ

Relative to old customers, new customers on average take almost a quarter as much time to convert, require fewer marketing touchpoints (~20%), and spend more per order (~30%).

First-touch attribution: the average sales per order as well as the total revenue by originator channel

```
dataQ3a<-data %>% filter(Positionname=='ORIGINATOR') %>% group_by(Groupname) %>%
  summarise(meanSales=mean(Saleamount),totalRevenue=sum(Saleamount))
Groupname='TOTAL'
meanSales=sum(dataQ3a$meanSales)
totalRevenue=sum(dataQ3a$totalRevenue)
total=data.frame(Groupname,meanSales,totalRevenue)
dataQ3a<-bind_rows(dataQ3a,total)
```

```
dataQ3a<-dataQ3a[c(1:5,7,9,12:13,6,8,10:11,14),]
as.table(as.matrix(dataQ3a))
```

```
##   Groupname                meanSales totalRevenue
## A BUZZ AFFILIATE            258.0513  49803.90
## B CJ                        262.3790  20990.32
## C CPM                       242.2739 137369.32
## D OTHER                     227.4729   3867.04
## E PRINT - MAGAZINES         324.6533    973.96
## F SEARCH GOOGLE NON-BRAND   234.5817   9617.85
## G SEARCH MSN NON-BRAND      274.9000   1649.40
## H TV                        239.2540   3588.81
## I Uncategorized             200.4157   1402.91
## J SEARCH GOOGLE BRAND       250.3493 117664.16
## K SEARCH MSN BRAND          229.0900  26803.53
## L SEARCH YAHOO BRAND        258.4900    258.49
## M Social                    165.0000    165.00
## N TOTAL                    3166.9112 374154.69
```

The total incremental gross revenue accruing to Winters by originator channel

Assume that Winters has a gross margin of 40%. Also assume an incrementality factor of 5% for branded search and 10% for the remaining channels. Note: An incrementality factor refers to the share of sales that are assumed to be incremental or caused by the channel. For instance, an incrementality factor of 20% implies that 0.20 of every 1 in sales is incremental

```
dataQ3b<-data %>% filter(Positionname=='ORIGINATOR') %>% select(Groupname,Saleamount,Brand) %>%
  group_by(Groupname,Brand) %>% summarise(totalRevenue=sum(Saleamount))  #Null in the Brand col?? (yes?
dataQ3b<-dataQ3b %>% mutate(IncFactor=ifelse(Brand=='Y',0.05,0.1),
                            IncRevenue=totalRevenue*IncFactor*0.4)
Groupname='TOTAL'
Brand='--'
totalRevenue=sum(dataQ3b$totalRevenue)
IncFactor=0
IncRevenue=sum(dataQ3b$IncRevenue)
total<-data.frame(Groupname,Brand,totalRevenue,IncFactor,IncRevenue)
dataQ3b<-bind_rows(dataQ3b,total)
as.table(as.matrix(dataQ3b))
```

```
##   Groupname               Brand totalRevenue IncFactor IncRevenue
## A BUZZ AFFILIATE          N      49803.90    0.10       1992.1560
## B CJ                      N      20990.32    0.10        839.6128
## C CPM                     N     137369.32    0.10       5494.7728
## D OTHER                   N        630.32    0.10         25.2128
## E OTHER                   NULL     677.13    0.10         27.0852
## F OTHER                   Y       2559.59    0.05         51.1918
## G PRINT - MAGAZINES       N        973.96    0.10         38.9584
## H SEARCH GOOGLE BRAND     Y     117664.16    0.05       2353.2832
## I SEARCH GOOGLE NON-BRAND N       9617.85    0.10        384.7140
## J SEARCH MSN BRAND        Y      26803.53    0.05        536.0706
## K SEARCH MSN NON-BRAND    N       1649.40    0.10         65.9760
## L SEARCH YAHOO BRAND      Y        258.49    0.05          5.1698
## M Social                  N        165.00    0.10          6.6000
## N TV                      NULL    3588.81    0.10        143.5524
## O Uncategorized           N       1402.91    0.10         56.1164
```

```
## P TOTAL                   --     374154.69    0.00      12020.4722
```

Q3 c

```r
sum(dataQ3b$IncRevenue[dataQ3b$Brand=='Y'])
```

```
## [1] 2945.715
```

If we assue that Winters search ad team spent 4,200 on the branded search advertising: Because incremental gross profit 2945.715 < branded search advertising spent $4,200, based on the pure number, the company should spend less on the branded search advertising and ideally not spend more than 2945 dollar on it. However, we may need more information to make this decision even if branded search advertising seems to be a loss. It is likely that our competitors will steal our market once we stop the branded search advertising.

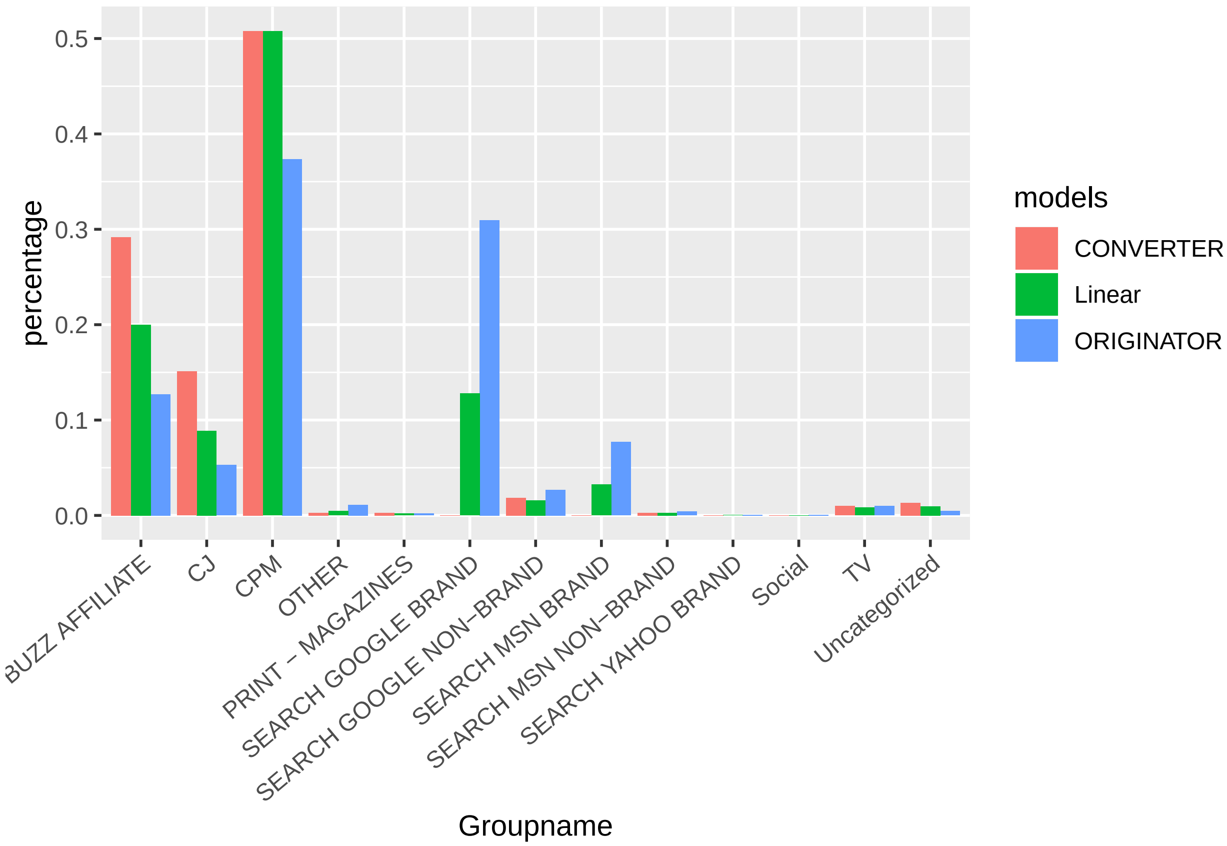
Linear attribution model

```r
dataQ4a<-data %>% add_count(Orderid) %>% mutate(LinearAttributionShare=1/n) %>%
  group_by(Groupname) %>% summarise(totLinearShares=sum(LinearAttributionShare)) %>%
  mutate(Credits=totLinearShares/sum(totLinearShares))

Groupname='TOTAL'
totLinearShares=sum(dataQ4a$totLinearShares)
Credits=sum(dataQ4a$Credits)
total<-data.frame(Groupname,totLinearShares,Credits)
dataQ4a<-bind_rows(dataQ4a,total)
dataQ4a<-dataQ4a[c(1:5,7,9,12:13,6,8,10:11,14),]
as.table(as.matrix(dataQ4a))
```

```
##   Groupname                totLinearShares Credits
## A BUZZ AFFILIATE            303.4789683    0.1999202689
## B CJ                        134.6825397    0.0887236757
## C CPM                       770.7567460    0.5077448920
## D OTHER                       6.9976190    0.0046097622
## E PRINT - MAGAZINES           2.9261905    0.0019276617
## F SEARCH GOOGLE NON-BRAND    24.0896825    0.0158693561
## G SEARCH MSN NON-BRAND        4.1055556    0.0027045821
## H TV                         12.3952381    0.0081655060
## I Uncategorized              14.5773810    0.0096030178
## J SEARCH GOOGLE BRAND       194.0670635    0.1278439153
## K SEARCH MSN BRAND           49.1194444    0.0323580003
## L SEARCH YAHOO BRAND          0.4285714    0.0002823264
## M Social                      0.3750000    0.0002470356
## N TOTAL                    1518.0000000    1.0000000000
```

Share of credit (in percentage) for all three attribution models: first-touch, last-touch and linear/uniform.

```r
dataQ4b<-data.frame(data2$Groupname,data2$`Converter%`,data2$`Originator%`,dataQ4a$Credits)[-14,]
colnames(dataQ4b)<-c('Groupname','CONVERTER','ORIGINATOR','Linear')
dataQ4b<-dataQ4b %>% pivot_longer(cols=c(CONVERTER,ORIGINATOR,Linear),
                                  names_to = 'models',values_to = 'percentage')
dataQ4b %>% ggplot(aes(x=Groupname,y=percentage,fill=models))+
  geom_col(position='dodge')+
  theme(axis.text.x = element_text(angle = 40, hjust = 1))
```



Compare the linear model to the first-touch and last-touch models

The linear model achieves a balance between first-touch and last-touch models. As we can see from the graph, no matter when some channels will be overestimated or underestimated by first-touch or last-touch models, their credits based on linear model will always between credits they get from first-touch and last-touch models.

Examine the role of the intermediate (Roster and Assist) touch points

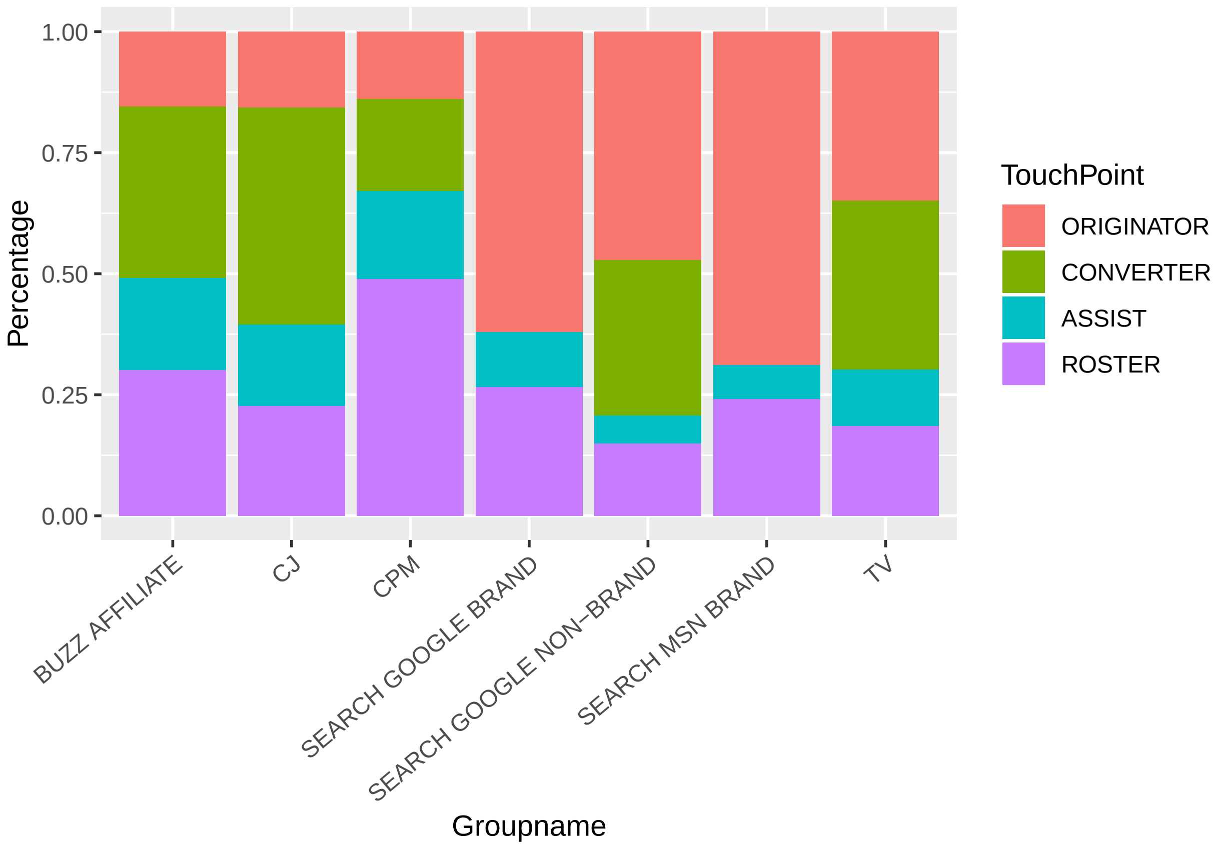
```
topChannel<-c('BUZZ AFFILIATE', 'CJ', 'CPM', 'SEARCH GOOGLE BRAND',
              'SEARCH GOOGLE NON-BRAND', 'SEARCH MSN BRAND', 'TV')
dataQ5a<-data %>% filter(Groupname %in% topChannel) %>% select(Groupname,Positionname) %>%
  group_by(Groupname,Positionname) %>% count()
dataQ5a<-pivot_wider(dataQ5a,names_from = Positionname,values_from = n,values_fill=list(n = 0))
dataQ5a1<-mutate(dataQ5a,total=sum(ASSIST,CONVERTER,ORIGINATOR,ROSTER),
                 ORIGINATOR=ORIGINATOR/total,ROSTER=ROSTER/total,
                 ASSIST=ASSIST/total,CONVERTER=CONVERTER/total) %>%
  mutate(total=sum(ASSIST,CONVERTER,ORIGINATOR,ROSTER)) %>%
  select(Groupname,ORIGINATOR,ROSTER,ASSIST,CONVERTER,total)
as.table(as.matrix(dataQ5a1))
```

```
##   Groupname               ORIGINATOR ROSTER    ASSIST     CONVERTER total
## A BUZZ AFFILIATE          0.1542766  0.3013589 0.19024780 0.3541167 1
## B CJ                      0.1565558  0.2270059 0.16829746 0.4481409 1
## C CPM                     0.1392778  0.4890690 0.18226480 0.1893884 1
## D SEARCH GOOGLE BRAND     0.6200528  0.2664908 0.11345646 0.0000000 1
## E SEARCH GOOGLE NON-BRAND 0.4712644  0.1494253 0.05747126 0.3218391 1
```

```
## F SEARCH MSN BRAND        0.6882353  0.2411765 0.07058824 0.0000000 1
## G TV                      0.3488372  0.1860465 0.11627907 0.3488372 1
```

The share in percentage (y-axis) of touchpoint types by marketing channels (x-axis)

```r
dataQ5b<-dataQ5a1 %>% select(-total) %>%
  pivot_longer(cols=c(ASSIST,CONVERTER,ORIGINATOR,ROSTER),values_to = 'Percentage',names_to = 'TouchPoi

dataQ5b$TouchPoint<-factor(dataQ5b$TouchPoint,
                           levels = c('ORIGINATOR','CONVERTER','ASSIST','ROSTER'))
ggplot(dataQ5b,aes(x=Groupname,y=Percentage,fill=TouchPoint))+
  geom_col()+
  theme(axis.text.x = element_text(angle = 40, hjust = 1))
```



Summary

Channels seem to have relatively MORE as ASSIST AND ROSTER: CPM,BUZZ AFFILIATE. Channels seem to have relatively LESS as ASSIST AND ROSTER:SEARCH GOOGLE NON-BRAND

Receive TOO LITTLE credit under first- and last-touch attribution: CPM,BUZZ AFFILIATE. Receive TOO MUCH under first- and last-touch attribution: SEARCH GOOGLE NON-BRAND

The following stands out: 1. TV has fewer intermediate (Roster & Assist) touchpoints than Originator or Converter touchpoints. 2. The search channels have fewer intermediate touchpoints than Originator touchpoints. 3. Though first-touch and last-touch favor display (CPM) as a top category, still more display ads arise in the Assist and especially the Roster positions. 4. In the affiliate channels (Buzz & CJ), the intermediate shares are similar to the Originator shares and a little less than the Converter shares. As such, first- and last-touch models may put give too much credit to TV and too little to search ads and especially display ads relative to a multi-touch attribution model.