Foundations of Marketing Analytics: Module 0: Introduction

Stefan Avey

2017-01-02

Contents

1 Load the Data

```
## Load data
purchases <- read.delim("data/purchases.txt", header = FALSE)

## Add column names
colnames(purchases) <- c("customer_id", "purchase_amount", "date_of_purchase")

head(data)
```

customer_id	purchase_amount	date_of_purchase	year_of_purchase
860	50	2012-09-28	2012
1200	100	2005-10-25	2005
1420	50	2009-07-09	2009
1940	70	2013-01-25	2013
1960	40	2013-10-29	2013
2620	30	2006-03-09	2006

```
summary(data)
```

```
##   customer_id     purchase_amount   date_of_purchase     year_of_purchase
##  Min.   :    10   Min.   :   5.00   Min.   :2005-01-02   Min.   :2005
##  1st Qu.: 57722   1st Qu.:  25.00   1st Qu.:2009-01-17   1st Qu.:2009
```

```
##  Median :102440   Median :  30.00   Median :2011-11-23   Median :2011
##  Mean   :108937   Mean   :  62.34   Mean   :2011-07-14   Mean   :2011
##  3rd Qu.:160528   3rd Qu.:  60.00   3rd Qu.:2013-12-29   3rd Qu.:2013
##  Max.   :264200   Max.   :4500.00   Max.   :2015-12-31   Max.   :2015
```

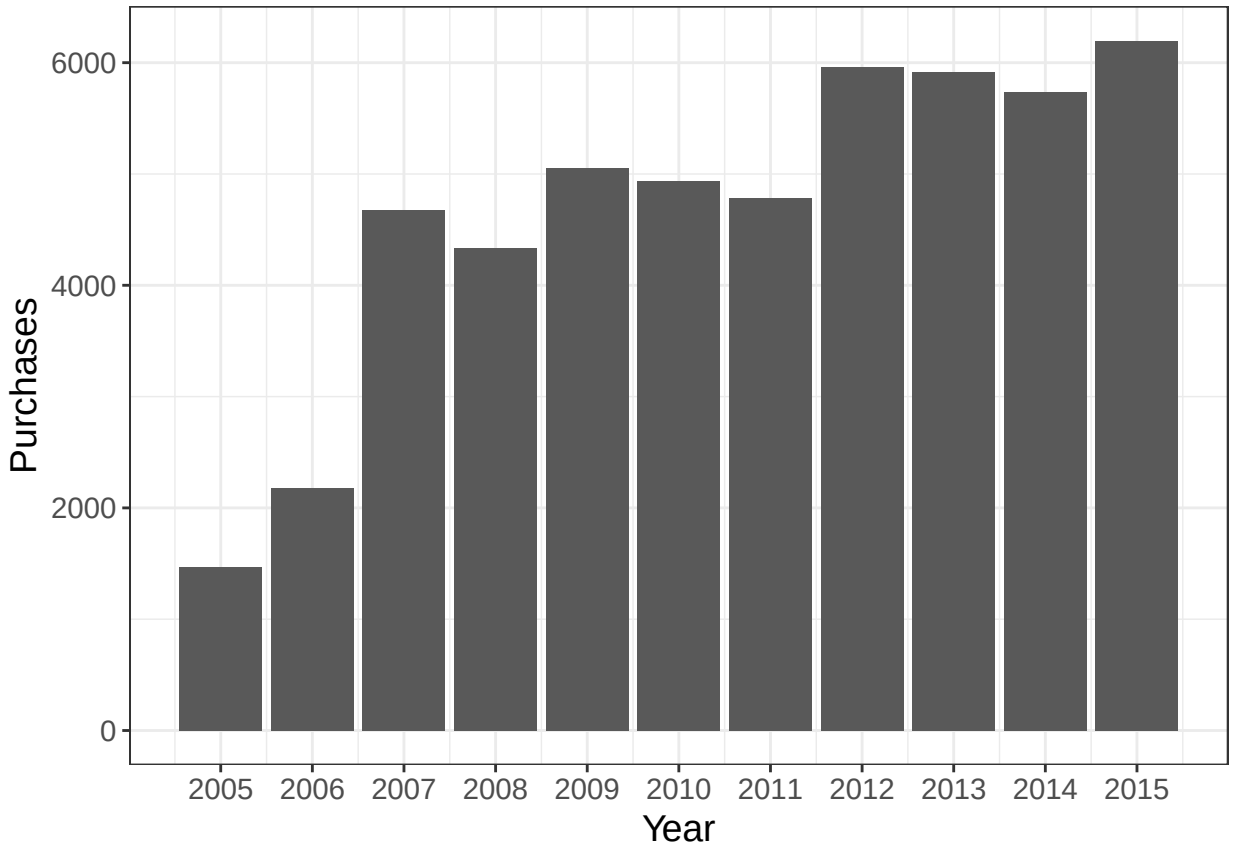
```
## Convert date and add column for purchase year
purchases <- purchases %>%
    mutate(date_of_purchase = ymd(date_of_purchase)) %>%
    mutate(year_of_purchase = year(date_of_purchase))
```

2 Explore the Data

```
plotDat <- purchases %>%
    group_by(year_of_purchase) %>%
    summarize(counter = n(),
              avg_amount = mean(purchase_amount),
              sum_amount = sum(purchase_amount),
              med_amount = median(purchase_amount)) %>%
    ungroup() %>%
    arrange(year_of_purchase)
```

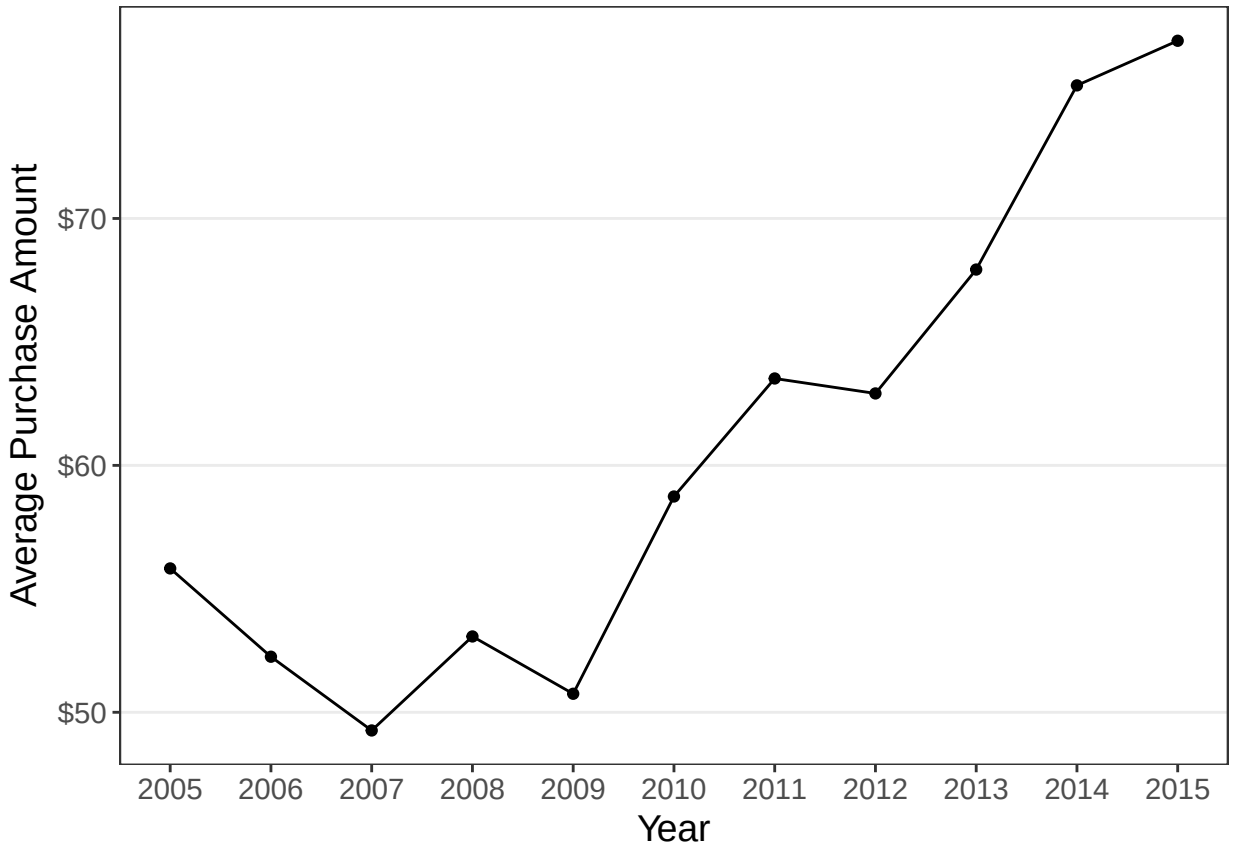
2.1 How many purchases are made each year?

```
ggplot(data = plotDat) +
    geom_bar(stat = "identity", aes(x = year_of_purchase, y = counter)) +
    scale_x_continuous(breaks = plotDat$year_of_purchase,
                       labels = plotDat$year_of_purchase) +
    xlab("Year") +
    ylab("Purchases") +
    getBaseTheme()
```

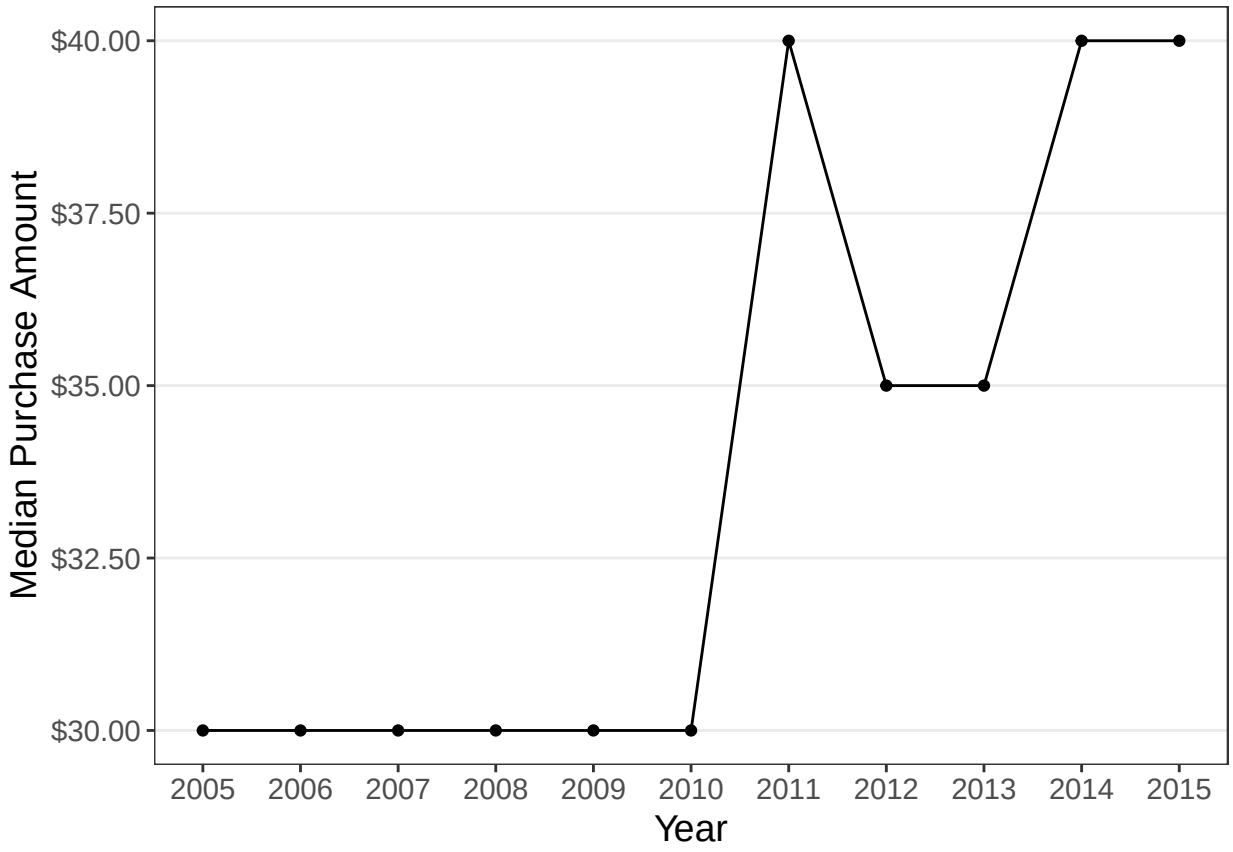


2.2 How has the purchase amount changed over time?

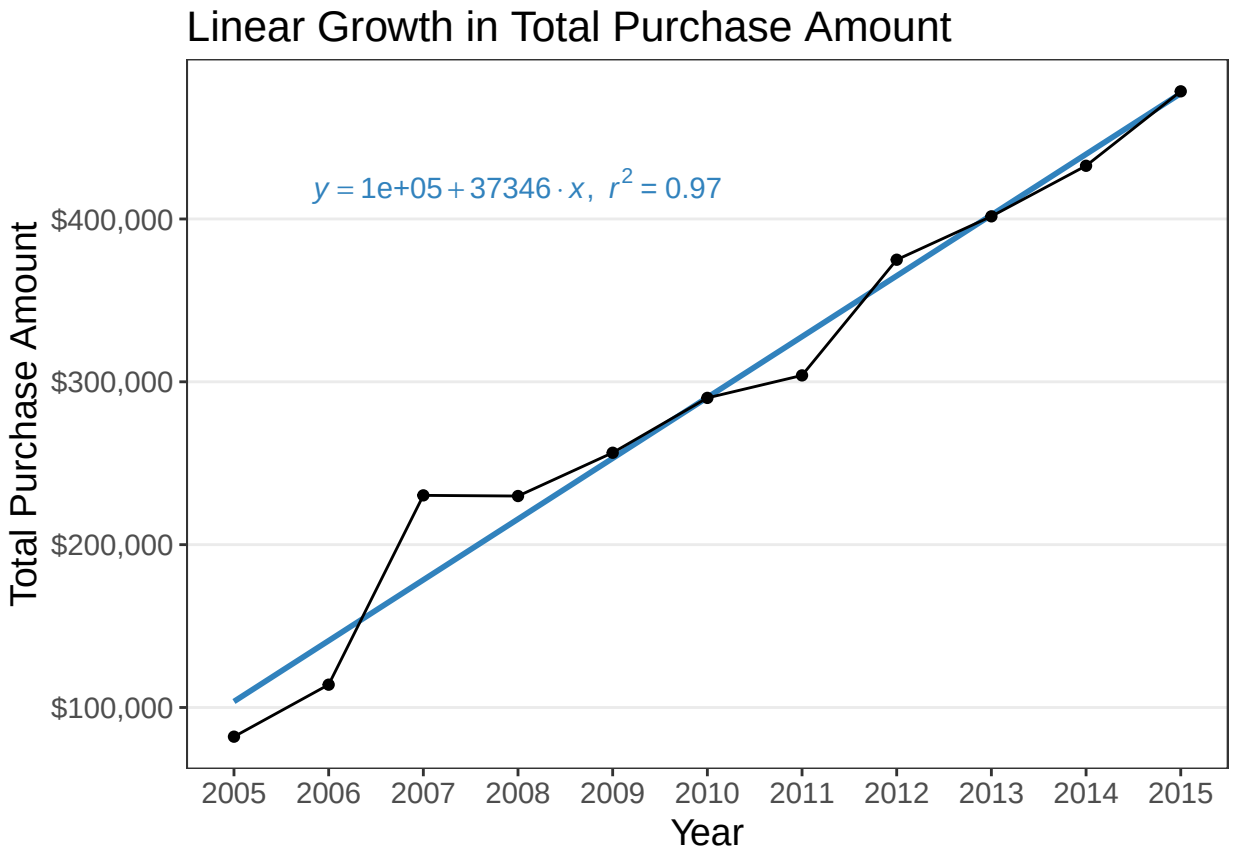
```
ggplot(data = plotDat, aes(x = year_of_purchase, y = avg_amount)) +
    geom_point() +
    geom_line() +
    scale_x_continuous(breaks = plotDat$year_of_purchase,
                       labels = plotDat$year_of_purchase) +
    scale_y_continuous(labels = scales::dollar) +
    xlab("Year") +
    ylab("Average Purchase Amount") +
    getBaseTheme() +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.x = element_blank())
```




```
ggplot(data = plotDat, aes(x = year_of_purchase, y = med_amount)) +
    geom_point() +
    geom_line() +
    scale_x_continuous(breaks = plotDat$year_of_purchase,
                       labels = plotDat$year_of_purchase) +
    scale_y_continuous(labels = scales::dollar) +
    xlab("Year") +
    ylab("Median Purchase Amount") +
    getBaseTheme() +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.x = element_blank())
```



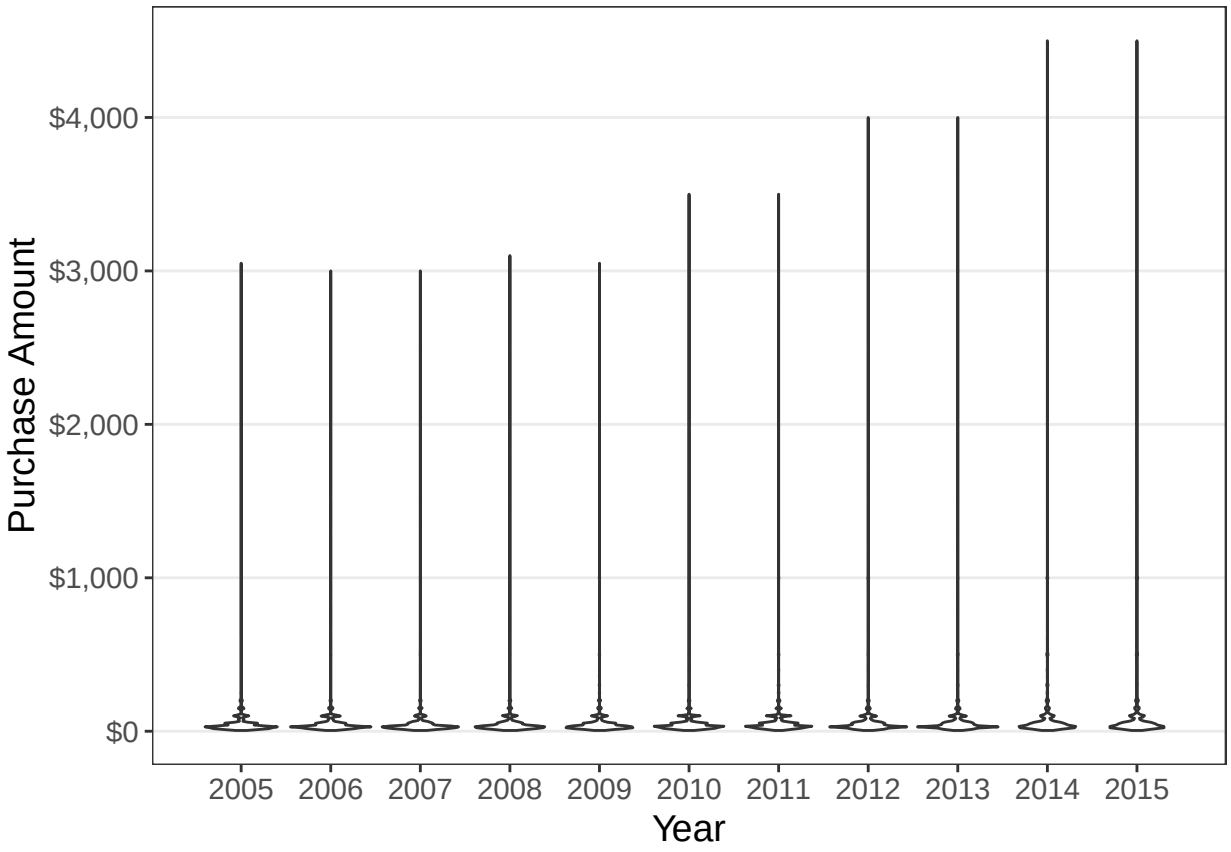

```
lmod <- lm(data = plotDat %>% mutate(year = year_of_purchase - 2005),
           sum_amount ~ year)
ggplot(data = plotDat, aes(x = year_of_purchase, y = sum_amount)) +
    geom_smooth(method = lm, color = "#3182bd", se = FALSE) +
    geom_point() +
    geom_line() +
    annotate("text", x = 2008, y = 420000, color = "#3182bd",
             label = GetEqn(lmod), parse = TRUE) +
    scale_x_continuous(breaks = plotDat$year_of_purchase,
                       labels = plotDat$year_of_purchase) +
    scale_y_continuous(labels = scales::dollar) +
    xlab("Year") +
    ylab("Total Purchase Amount") +
    getBaseTheme() +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.x = element_blank()) +
    ggtitle("Linear Growth in Total Purchase Amount")
```



2.3 What is the purchase distribution each year?

While looking at the number of purchases, average amount sold, and total amount sold are informative, it is helpful to explore the whole distribution of purchases per year.

```
ggplot(data = purchases, aes(x = year_of_purchase, y = purchase_amount)) +
    geom_violin(aes(group = year_of_purchase)) +
    scale_x_continuous(breaks = plotDat$year_of_purchase,
                       labels = plotDat$year_of_purchase) +
    scale_y_continuous(labels = scales::dollar) +
    xlab("Year") +
    ylab("Purchase Amount") +
    getBaseTheme() +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.x = element_blank())
```



3 Conclusions

- From 2005–2015, the total purchase ammount has increased linearly at a rate of approximately $37,000 / year
- While the growth in total sales in 2005–2007 was due mainly to an increase in sales volume, the growth in 2012–2015 was due mainly to an increase in purchase amount.