

Chip Analysis Task 2

Alif Safwan

2025-01-06

Contents

1	Import Library and Data	1
2	Data Preparation and Exploratory Data Analysis	1
3	Finding Control Store	4
4	Analysis of Trial Store 77 compare to Control Store 233	5
5	Analysis of Trial Store 86 compare to Control Store 155	11
6	Analysis of Trial Store 88 compare to Control Store 237	17

1 Import Library and Data

```
library(tidyverse)
library(ggplot2)
library(readxl)
library(tidyr)
setwd("C:/Users/User/Documents/Project Forage/Quantium")
```

```
Chip <- read_excel("C:/Users/User/Documents/Project Forage/Quantium/chip_data.xlsx")
```

2 Data Preparation and Exploratory Data Analysis

```
summary(Chip)
```

```
##          DATE                STORE_NBR  LYLTY_CARD_NBR
## Min.      :2018-07-01 00:00:00.00  Min.    : 1  Min.    : 1000
## 1st Qu.:2018-09-30 00:00:00.00  1st Qu.: 70  1st Qu.: 70017
## Median :2018-12-30 00:00:00.00  Median :130  Median : 130352
## Mean    :2018-12-30 02:57:08.07  Mean    :135  Mean    : 135506
## 3rd Qu.:2019-03-31 00:00:00.00  3rd Qu.:203  3rd Qu.: 203076
## Max.    :2019-06-30 00:00:00.00  Max.    :272  Max.    :2373711
##      TXN_ID      PROD_NBR      PROD_NAME      PROD_QTY
## Min.      :      1  Min.      : 1.00  Length:251158  Min.      :1.000
## 1st Qu.: 67575  1st Qu.: 27.00  Class :character  1st Qu.:2.000
## Median : 135110  Median : 52.00  Mode  :character  Median :2.000
## Mean    : 135111  Mean    : 56.17                Mean    :1.906
## 3rd Qu.: 202619  3rd Qu.: 86.00                3rd Qu.:2.000
## Max.    :2415841  Max.    :114.00                Max.    :5.000
##      TOT_SALES  packed_size      product_brand      product_name
## Min.      : 1.500  Length:251158  Length:251158  Length:251158
## 1st Qu.: 5.600  Class :character  Class :character  Class :character
## Median : 7.400  Mode  :character  Mode  :character  Mode  :character
## Mean    : 7.262
## 3rd Qu.: 8.800
## Max.    :29.500
##      LIFESTAGE      PREMIUM_CUSTOMER
## Length:251158  Length:251158
## Class :character  Class :character
## Mode  :character  Mode  :character
##
##
##
```

```
colSums(is.na(Chip))
```

```
##          DATE      STORE_NBR  LYLTY_CARD_NBR      TXN_ID
##          0          0          0          0
##      PROD_NBR      PROD_NAME      PROD_QTY      TOT_SALES
##          0          0          0          0
##      packed_size  product_brand  product_name  LIFESTAGE
##          0          0          0          0
## PREMIUM_CUSTOMER
##          0
```

```
Chip %>%
  filter(STORE_NBR == 77)
```

```
## # A tibble: 543 x 13
##   DATE                STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME
##   <dtm>                <dbl>      <dbl>  <dbl>   <dbl> <chr>
```

```
## 1 2019-06-17 00:00:00      77      77069 74987      70 Tyrrells Crisps~
## 2 2019-03-28 00:00:00      77      77000 74911      18 Cheetos Chs & B~
## 3 2019-04-13 00:00:00      77      77000 74912      69 Smiths Chip Thi~
## 4 2018-07-17 00:00:00      77      2330211 236744      94 Burger Rings 22~
## 5 2019-03-03 00:00:00      77      77063 74977     112 Tyrrells Crisps~
## 6 2019-02-20 00:00:00      77      77069 74985      98 NCC Sour Cream ~
## 7 2019-03-08 00:00:00      77      77069 74986       8 Smiths Crinkle ~
## 8 2019-02-28 00:00:00      77      77310 75254       9 Kettle Tortilla~
## 9 2018-10-02 00:00:00      77      77502 75463      94 Burger Rings 22~
## 10 2019-02-20 00:00:00      77      77502 75464      69 Smiths Chip Thi~
## # i 533 more rows
## # i 7 more variables: PROD_QTY <dbl>, TOT_SALES <dbl>, packed_size <chr>,
## #   product_brand <chr>, product_name <chr>, LIFESTAGE <chr>,
## #   PREMIUM_CUSTOMER <chr>
```

```
Chip %>%
  filter(STORE_NBR == 86)
```

```
## # A tibble: 1,479 x 13
##   DATE                STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME
##   <dtm>                <dbl>      <dbl> <dbl>      <dbl> <chr>
## 1 2019-01-11 00:00:00      86      86245 85639       21 WW Sour Cream &~
## 2 2019-02-27 00:00:00      86      86245 85640       10 RRD SR Slow Rst~
## 3 2019-03-31 00:00:00      86      86245 85641       81 Pringles Origin~
## 4 2019-05-27 00:00:00      86      86245 85642       72 WW Crinkle Cut ~
## 5 2018-11-04 00:00:00      86      86248 85655       26 Pringles Sweet&~
## 6 2018-11-12 00:00:00      86      86248 85656       42 Doritos Corn Ch~
## 7 2018-12-24 00:00:00      86      86248 85658       84 GrnWves Plus Bt~
## 8 2019-03-15 00:00:00      86      86248 85659       61 Smiths Crinkle ~
## 9 2019-04-30 00:00:00      86      86248 85660      102 Kettle Mozzarel~
## 10 2019-05-22 00:00:00      86      86248 85661       84 GrnWves Plus Bt~
## # i 1,469 more rows
## # i 7 more variables: PROD_QTY <dbl>, TOT_SALES <dbl>, packed_size <chr>,
## #   product_brand <chr>, product_name <chr>, LIFESTAGE <chr>,
## #   PREMIUM_CUSTOMER <chr>
```

```
Chip %>%
  filter(STORE_NBR == 88)
```

```
## # A tibble: 1,786 x 13
##   DATE                STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME
##   <dtm>                <dbl>      <dbl> <dbl>      <dbl> <chr>
## 1 2019-05-20 00:00:00      88      88320 87811      113 Twisties Chicke~
## 2 2018-09-12 00:00:00      88      88000 86220       4 Dorito Corn Chp~
## 3 2018-10-26 00:00:00      88      88000 86221      31 Infzns Crn Crnc~
## 4 2019-02-02 00:00:00      88      88000 86222      46 Kettle Original~
```

```
## 5 2019-05-01 00:00:00      88      88000 86223      47 Doritos Corn Ch~
## 6 2019-05-02 00:00:00      88      88000 86224     102 Kettle Mozzarel~
## 7 2018-08-25 00:00:00      88      88074 86580      9 Kettle Tortilla~
## 8 2018-11-02 00:00:00      88      88074 86581     26 Pringles Sweet&~
## 9 2018-12-13 00:00:00      88      88089 86650      9 Kettle Tortilla~
## 10 2018-12-26 00:00:00      88      88089 86652     102 Kettle Mozzarel~
## # i 1,776 more rows
## # i 7 more variables: PROD_QTY <dbl>, TOT_SALES <dbl>, packed_size <chr>,
## #   product_brand <chr>, product_name <chr>, LIFESTAGE <chr>,
## #   PREMIUM_CUSTOMER <chr>
```

3 Finding Control Store

```
modified_chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  summarise(monthly_sales = sum(TOT_SALES, na.rm = TRUE),
            monthly_customer = length(unique(LYLT_Y_CARD_NBR)),
            monthly_avg_trans_customer = n()/length(unique(LYLT_Y_CARD_NBR)),
            monthly_avg_chip_trans = sum(PROD_QTY, na.rm = TRUE)/n()) %>%
  ungroup()
```

`summarise()` has grouped output by 'YMcode'. You can override using the
`.groups` argument.

```
colSums(is.na(modified_chip))
```

```
##           YMcode           STORE_NBR
##           0           0
##   monthly_sales   monthly_customer
##           0           0
## monthly_avg_trans_customer monthly_avg_chip_trans
##           0           0
```

```
pre_trial <- modified_chip %>%
  filter(YMcode < "2019-02" )
```

```
metrics_wide <- pre_trial %>%
  dplyr::select(STORE_NBR, YMcode, monthly_sales, monthly_customer, monthly_avg_trans_customer)
  pivot_wider(names_from = YMcode, values_from = c(monthly_sales, monthly_customer, monthly_avg_trans_customer))
  drop_na()
```

```

# Extract trial store data (Store 77, 86 and 88)
trial_store <- c(77,86,88)
for (trial_store_id in trial_store) {

trial_data <- metrics_wide %>% filter(STORE_NBR == trial_store_id) %>% dplyr::select(-STORE_NBR)

# Step 3: Compute Euclidean Distance for Each Store
compute_euclidean <- function(x, y) {
  sqrt(sum((x - y) ^ 2, na.rm = TRUE)) # Standard Euclidean formula
}

distance_results <- metrics_wide %>%
  filter(STORE_NBR != trial_store_id) %>%
  rowwise() %>%
  mutate(euclidean_dist = compute_euclidean(c_across(-STORE_NBR), unlist(trial_data))) %>%
  arrange(euclidean_dist) %>%
  ungroup()

best_control_store <- distance_results %>% slice(1)

print(paste("Best Control Store for", trial_store_id, "is", best_control_store$STORE_NBR))
}

```

```

## [1] "Best Control Store for 77 is 233"
## [1] "Best Control Store for 86 is 155"
## [1] "Best Control Store for 88 is 237"

```

4 Analysis of Trial Store 77 compare to Control Store 233

a) Total Sales

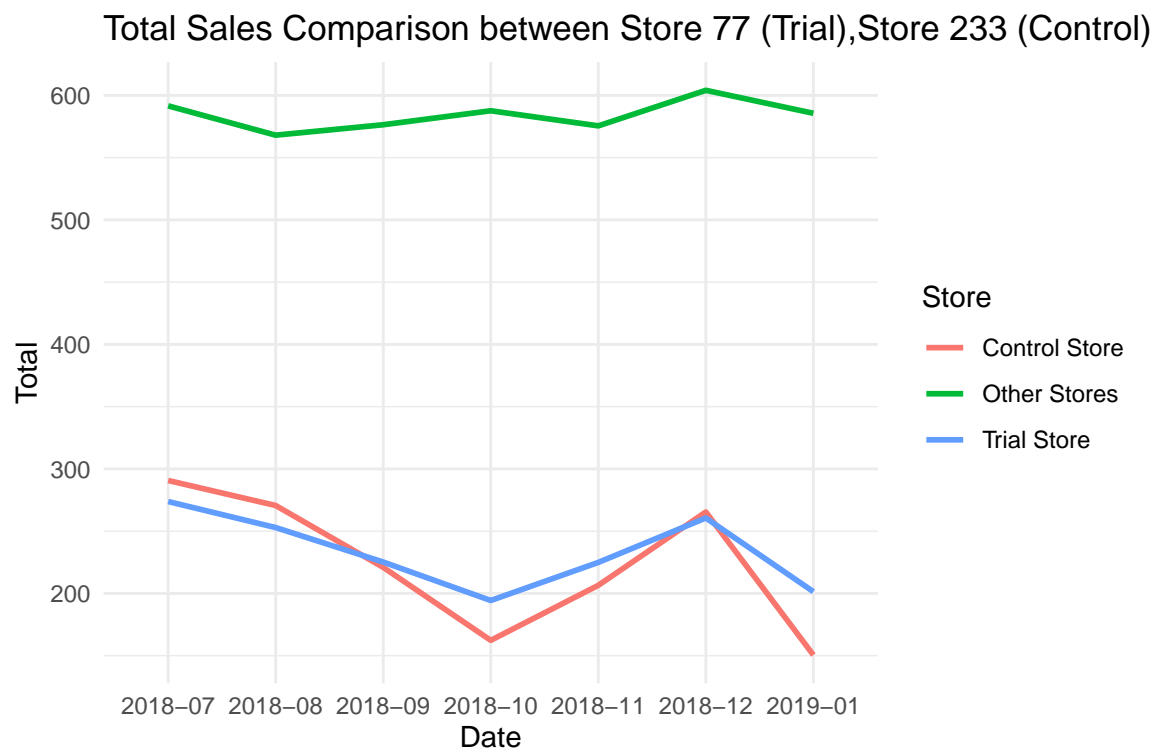
```

Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 77 ~ "Trial Store",
    STORE_NBR == 233 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +

```

```
geom_line(size = 1) +
labs(title = "Total Sales Comparison between Store 77 (Trial),Store 233 (Control) and Average",
     x = "Date",
     y = "Total",
     color = "Store") +
theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 77 | STORE_NBR == 233 )) %>%
  summarise( total = sum(TOT_SALES)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_233 = STORE_NBR_233*sum(STORE_NBR_77)/sum(STORE_NBR_233),
         Abs_diff_tot_with_control = abs(STORE_NBR_77 - Controlled_Store_NBR_233))
```

```

trial <- Mod_Chip %>%
  filter(YMcode < "2019-02") %>%
  pull(Abs_diff_tot_with_control)
Mean <- Mod_Chip %>%
  filter(YMcode == i) %>%
  pull(Abs_diff_tot_with_control)
SD <- sd(trial)
t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
if(t > qt(p = 0.025,6, lower.tail = FALSE))
{ print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
else
{
  print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
}
}

```

```

## [1] "Trial Month 2019-02 is not significant different compared to pre-trial month"
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"
## [1] "Trial Month 2019-04 is significant different compared to pre-trial month"

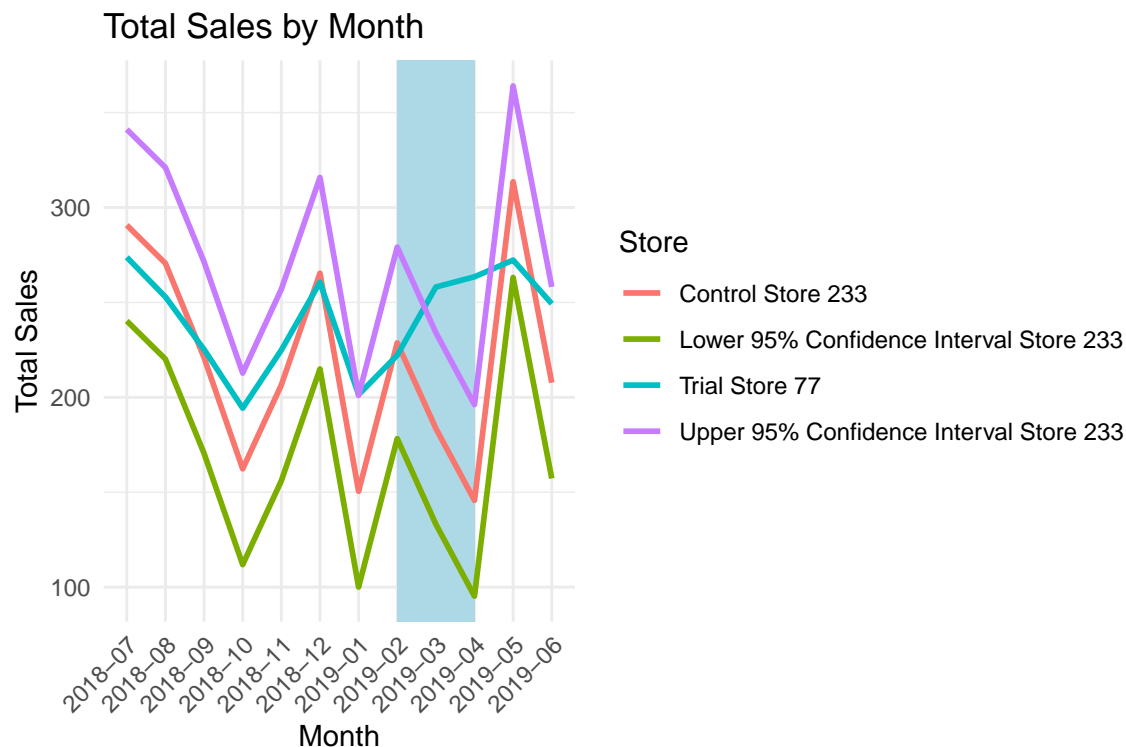
```

```

Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%
  filter(STORE_NBR == 77 | STORE_NBR == 233) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%
  mutate(upper = Store_Nbr233+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr233[YMcode < "2019-02-01"]),
         lower = Store_Nbr233+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr233[YMcode < "2019-02-01"]),
  pivot_longer(cols = c(Store_Nbr77, Store_Nbr233,upper,lower),
               names_to = "Series",
               values_to = "Values") %>%
  mutate(Series = case_when(Series == "Store_Nbr233" ~ "Control Store 233",
                           Series == "Store_Nbr77" ~ "Trial Store 77",
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 233",
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 233")) %>%
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +
  geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),
                 xmax = format(as.Date("2019-04-30"), "%Y-%m"),
                 ymin = -Inf, ymax = Inf),
            fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +
  geom_line(size = 1) +
  labs(title = "Total Sales by Month",
       x = "Month",
       y = "Total Sales",
       color = "Store") +
  scale_y_continuous(breaks = seq(0, 400, by = 100)) +

```

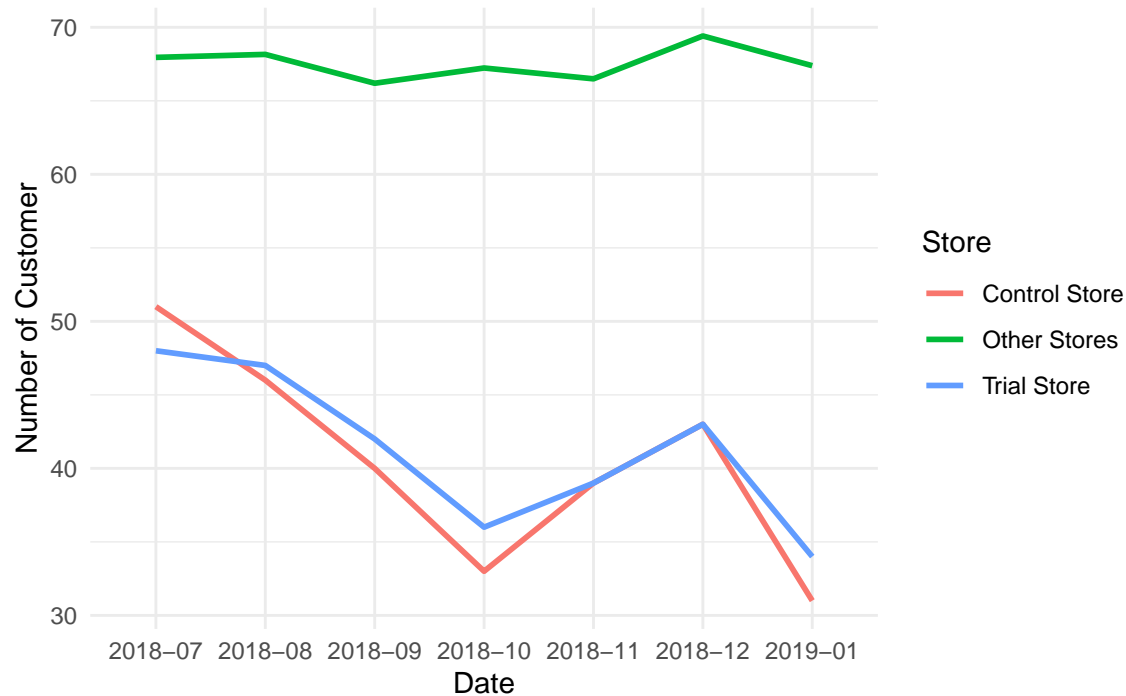
```
scale_x_discrete(guide = guide_axis(angle = 45)) +
theme_minimal()
```



b) Number of Customer

```
Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = n_distinct(LYLT_CARD_NBR), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 77 ~ "Trial Store",
    STORE_NBR == 233 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +
  geom_line(size = 1) +
  labs(title = "Number of Customer Comparison between Store 77 (Trial),Store 233 (Control) and",
       x = "Date",
       y = "Number of Customer",
       color = "Store") +
  theme_minimal()
```


Number of Customer Comparison between Store 77 (Trial),Store 233



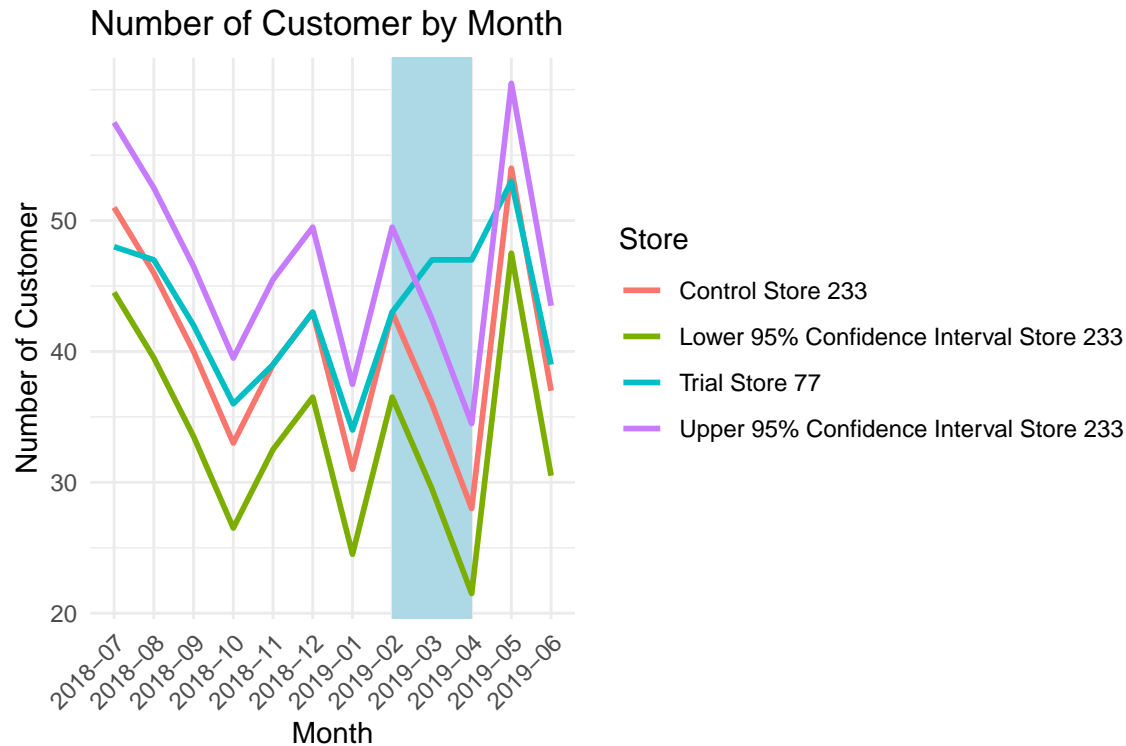
```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 77 | STORE_NBR == 233 )) %>%
  summarise(total = n_distinct(LYLT_CARD_NBR)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_233 = STORE_NBR_233*sum(STORE_NBR_77)/sum(STORE_NBR_233),
         Abs_diff_tot_with_control = abs(STORE_NBR_77 - Controlled_Store_NBR_233))
  trial <- Mod_Chip %>%
    filter(YMcode < "2019-02") %>%
    pull(Abs_diff_tot_with_control)
  Mean <- Mod_Chip %>%
    filter(YMcode == i) %>%
    pull(Abs_diff_tot_with_control)
  SD <- sd(trial)
  t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
  if(t > qt(p = 0.025,6, lower.tail = FALSE))
  { print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
  else
  {
    print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
  }
}
```

```
}
```

```
## [1] "Trial Month 2019-02 is not significant different compared to pre-trial month"  
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-04 is significant different compared to pre-trial month"
```

```
Chip %>%
```

```
  mutate(YMcode = format(DATE, "%Y-%m")) %>%  
  group_by(YMcode, STORE_NBR) %>%  
  summarise(total = n_distinct(LYLT_CARD_NBR), .groups = "drop") %>%  
  filter(STORE_NBR == 77 | STORE_NBR == 233) %>%  
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%  
  mutate(upper = Store_Nbr233+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr233[YMcode < "2019-04-01"]),  
         lower = Store_Nbr233+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr233[YMcode < "2019-04-01"]),  
  pivot_longer(cols = c(Store_Nbr77, Store_Nbr233,upper,lower),  
               names_to = "Series",  
               values_to = "Values") %>%  
  mutate(Series = case_when(Series == "Store_Nbr233" ~ "Control Store 233",  
                           Series == "Store_Nbr77" ~ "Trial Store 77",  
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 233",  
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 233")) %>%  
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +  
    geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),  
                  xmax = format(as.Date("2019-04-30"), "%Y-%m"),  
                  ymin = -Inf, ymax = Inf),  
              fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +  
    geom_line(size = 1) +  
    labs(title = "Number of Customer by Month",  
         x = "Month",  
         y = "Number of Customer",  
         color = "Store") +  
    scale_y_continuous(breaks = seq(0, 50, by = 10)) +  
    scale_x_discrete(guide = guide_axis(angle = 45)) +  
    theme_minimal()
```

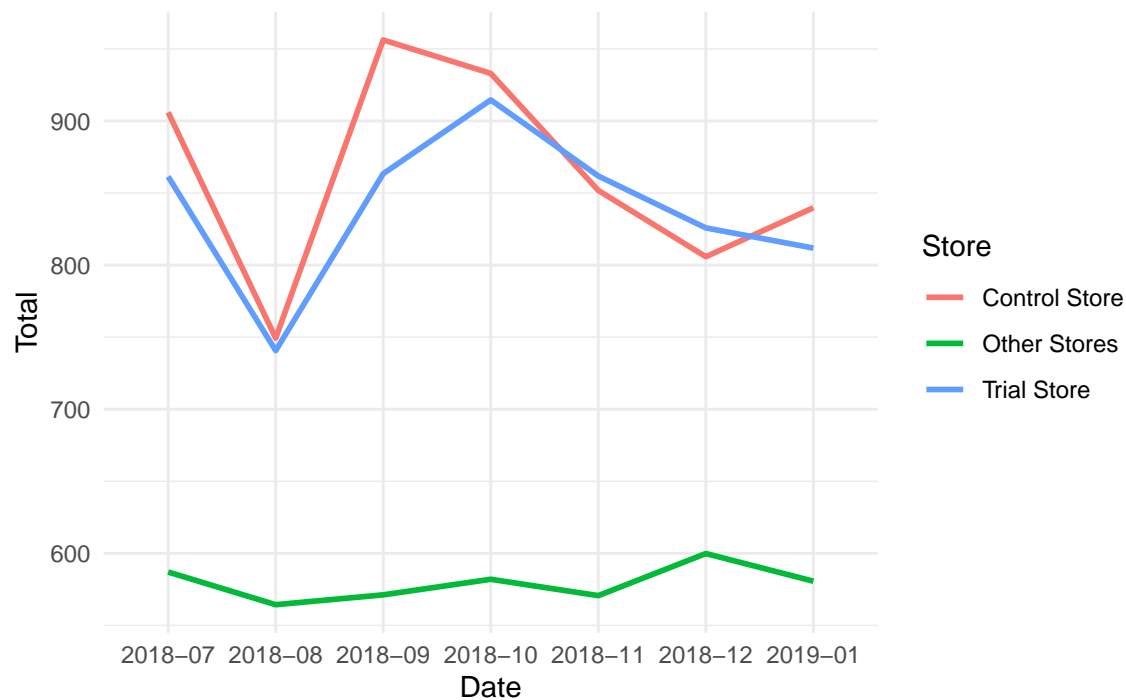


5 Analysis of Trial Store 86 compare to Control Store 155

a) Total Sales

```
Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 86 ~ "Trial Store",
    STORE_NBR == 155 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +
  geom_line(size = 1) +
  labs(title = "Total Sales Comparison between Store 86 (Trial),Store 155 (Control) and Average",
       x = "Date",
       y = "Total",
       color = "Store") +
  theme_minimal()
```

Total Sales Comparison between Store 86 (Trial),Store 155 (Control)



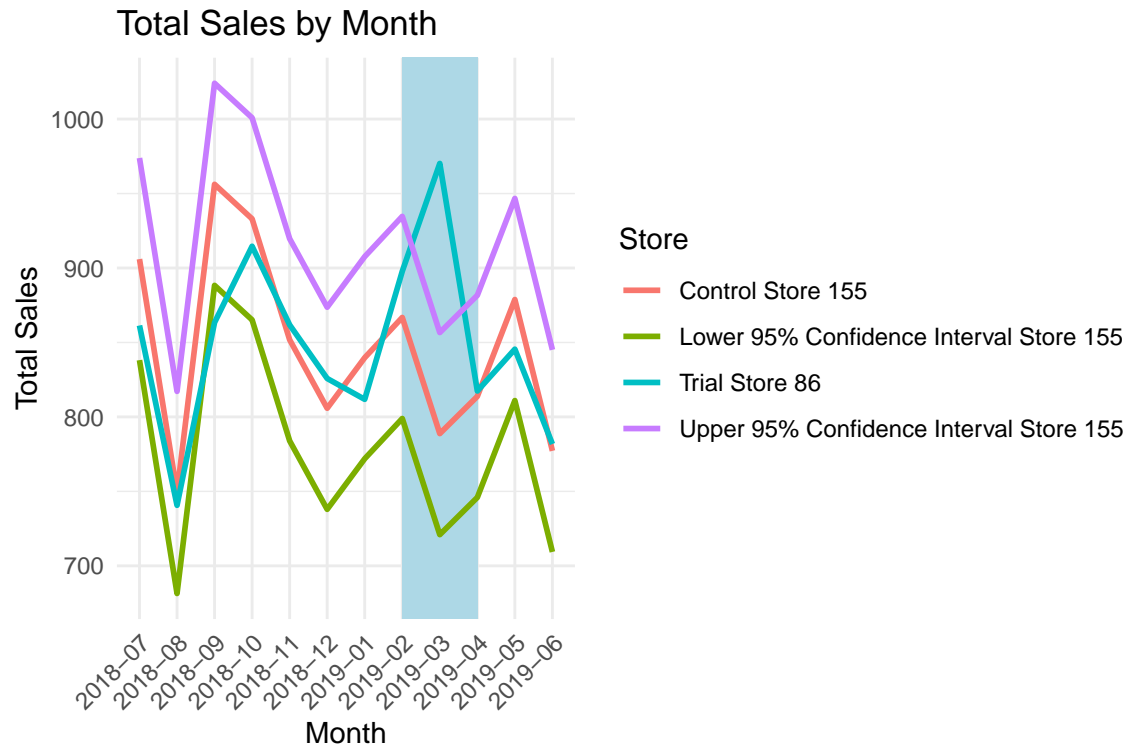
```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 86 | STORE_NBR == 155 )) %>%
  summarise( total = sum(TOT_SALES)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_155 = STORE_NBR_155*sum(STORE_NBR_86)/sum(STORE_NBR_155),
         Abs_diff_tot_with_control = abs(STORE_NBR_86 - Controlled_Store_NBR_155))
  trial <- Mod_Chip %>%
    filter(YMcode < "2019-02") %>%
    pull(Abs_diff_tot_with_control)
  Mean <- Mod_Chip %>%
    filter(YMcode == i) %>%
    pull(Abs_diff_tot_with_control)
  SD <- sd(trial)
  t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
  if(t > qt(p = 0.025,6, lower.tail = FALSE))
  { print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
  else
  {
    print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
  }
}
```

```
}
```

```
## [1] "Trial Month 2019-02 is not significant different compared to pre-trial month"  
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-04 is not significant different compared to pre-trial month"
```

```
Chip %>%
```

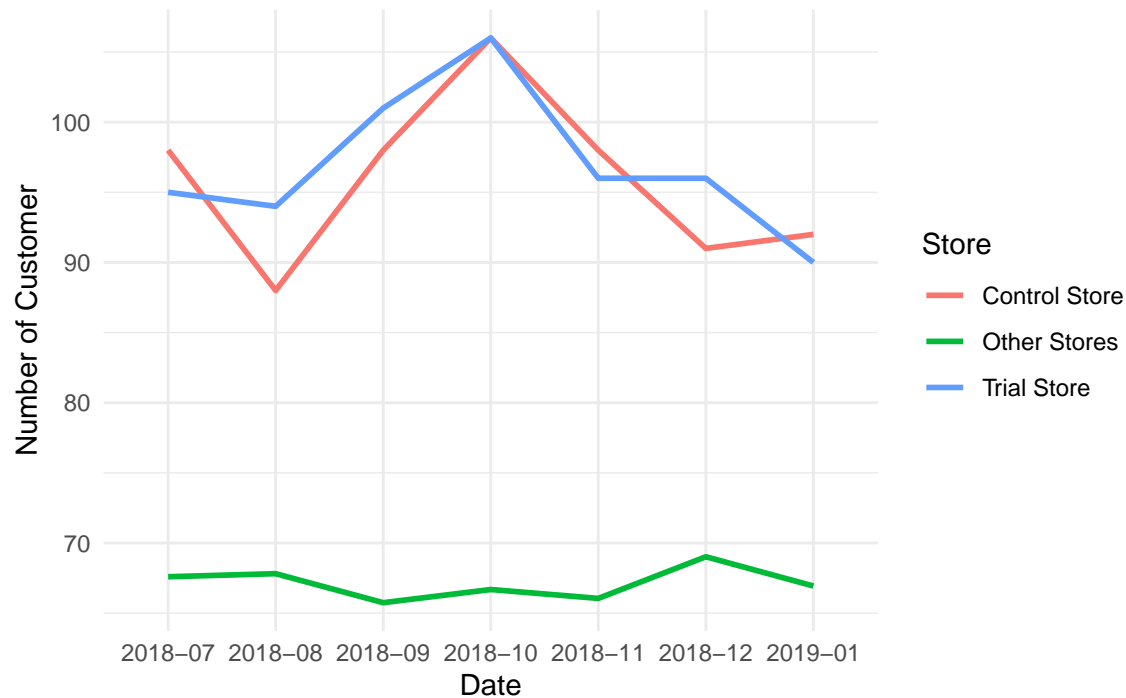
```
  mutate(YMcode = format(DATE, "%Y-%m")) %>%  
  group_by(YMcode, STORE_NBR) %>%  
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%  
  filter(STORE_NBR == 86 | STORE_NBR == 155) %>%  
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%  
  mutate(upper = Store_Nbr155+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr155[YMcode < "2019-04-01"]),  
         lower = Store_Nbr155+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr155[YMcode < "2019-04-01"]),  
  pivot_longer(cols = c(Store_Nbr86, Store_Nbr155,upper,lower),  
               names_to = "Series",  
               values_to = "Values") %>%  
  mutate(Series = case_when(Series == "Store_Nbr155" ~ "Control Store 155",  
                           Series == "Store_Nbr86" ~ "Trial Store 86",  
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 155",  
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 155")) %>%  
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +  
    geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),  
                  xmax = format(as.Date("2019-04-30"), "%Y-%m"),  
                  ymin = -Inf, ymax = Inf),  
              fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +  
    geom_line(size = 1) +  
    labs(title = "Total Sales by Month",  
         x = "Month",  
         y = "Total Sales",  
         color = "Store") +  
    scale_y_continuous(breaks = seq(0, 1000, by = 100)) +  
    scale_x_discrete(guide = guide_axis(angle = 45)) +  
    theme_minimal()
```



b) Number of Customer

```
Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = n_distinct(LYLT_CARD_NBR), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 86 ~ "Trial Store",
    STORE_NBR == 155 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +
  geom_line(size = 1) +
  labs(title = "Number of Customer Comparison between Store 86 (Trial),Store 155 (Control) and",
       x = "Date",
       y = "Number of Customer",
       color = "Store") +
  theme_minimal()
```

Number of Customer Comparison between Store 86 (Trial),Store 155



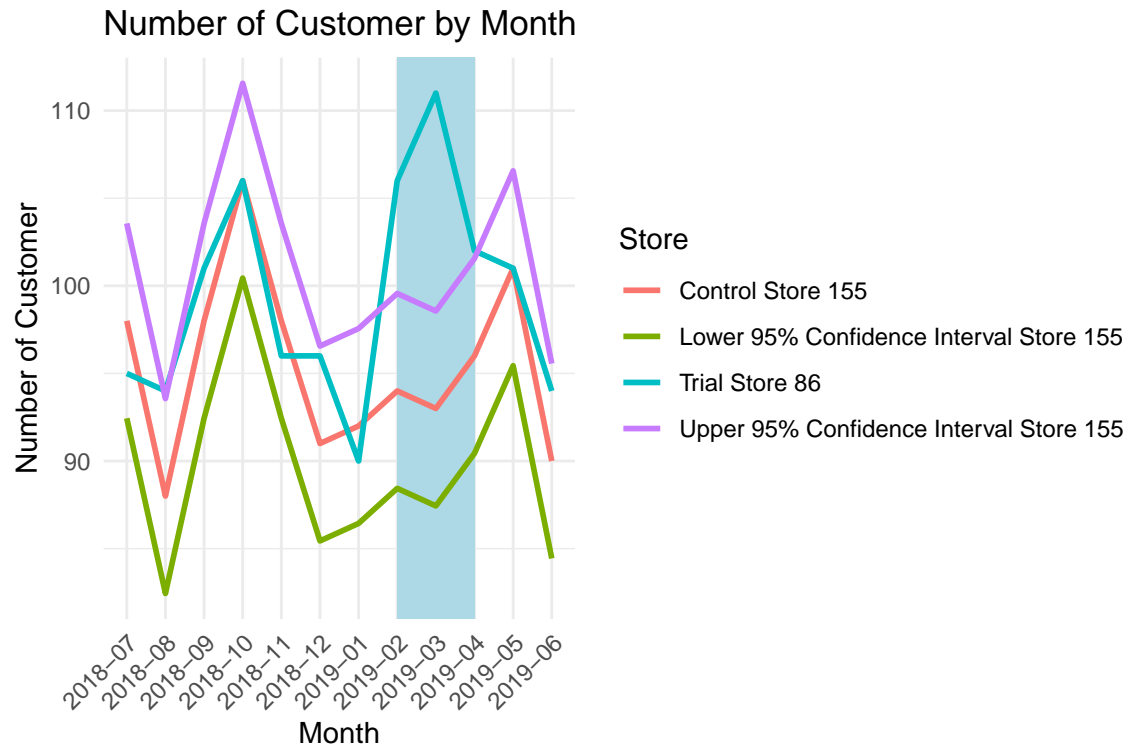
```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 86 | STORE_NBR == 155 )) %>%
  summarise( total = n_distinct(LYLT_CARD_NBR)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_155 = STORE_NBR_155*sum(STORE_NBR_86)/sum(STORE_NBR_155),
         Abs_diff_tot_with_control = abs(STORE_NBR_86 - Controlled_Store_NBR_155))
  trial <- Mod_Chip %>%
    filter(YMcode < "2019-02") %>%
    pull(Abs_diff_tot_with_control)
  Mean <- Mod_Chip %>%
    filter(YMcode == i) %>%
    pull(Abs_diff_tot_with_control)
  SD <- sd(trial)
  t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
  if(t > qt(p = 0.025,6, lower.tail = FALSE))
  { print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
  else
  {
    print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
  }
}
```

```
}
```

```
## [1] "Trial Month 2019-02 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-04 is not significant different compared to pre-trial month"
```

```
Chip %>%
```

```
  mutate(YMcode = format(DATE, "%Y-%m")) %>%  
  group_by(YMcode, STORE_NBR) %>%  
  summarise(total = n_distinct(LYLT_CARD_NBR), .groups = "drop") %>%  
  filter(STORE_NBR == 86 | STORE_NBR == 155) %>%  
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%  
  mutate(upper = Store_Nbr155+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr155[YMcode < "2019-04-01"]),  
         lower = Store_Nbr155+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr155[YMcode < "2019-04-01"]),  
  pivot_longer(cols = c(Store_Nbr86, Store_Nbr155,upper,lower),  
               names_to = "Series",  
               values_to = "Values") %>%  
  mutate(Series = case_when(Series == "Store_Nbr155" ~ "Control Store 155",  
                           Series == "Store_Nbr86" ~ "Trial Store 86",  
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 155",  
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 155")) %>%  
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +  
    geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),  
                  xmax = format(as.Date("2019-04-30"), "%Y-%m"),  
                  ymin = -Inf, ymax = Inf),  
              fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +  
    geom_line(size = 1) +  
    labs(title = "Number of Customer by Month",  
         x = "Month",  
         y = "Number of Customer",  
         color = "Store") +  
    scale_y_continuous(breaks = seq(0, 150, by = 10)) +  
    scale_x_discrete(guide = guide_axis(angle = 45)) +  
    theme_minimal()
```

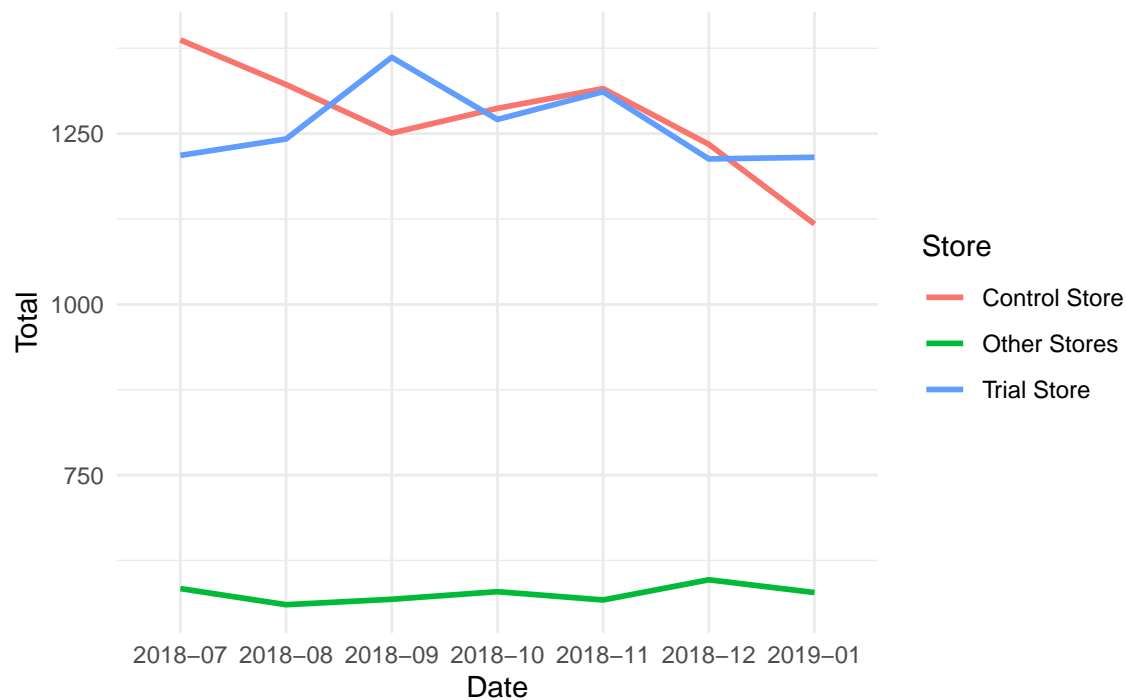



6 Analysis of Trial Store 88 compare to Control Store 237

a) Total Sales

```
Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 88 ~ "Trial Store",
    STORE_NBR == 237 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +
  geom_line(size = 1) +
  labs(title = "Total Sales Comparison between Store 88 (Trial), Store 237 (Control) and Average",
       x = "Date",
       y = "Total",
       color = "Store") +
  theme_minimal()
```

Total Sales Comparison between Store 88 (Trial),Store 237 (Control)



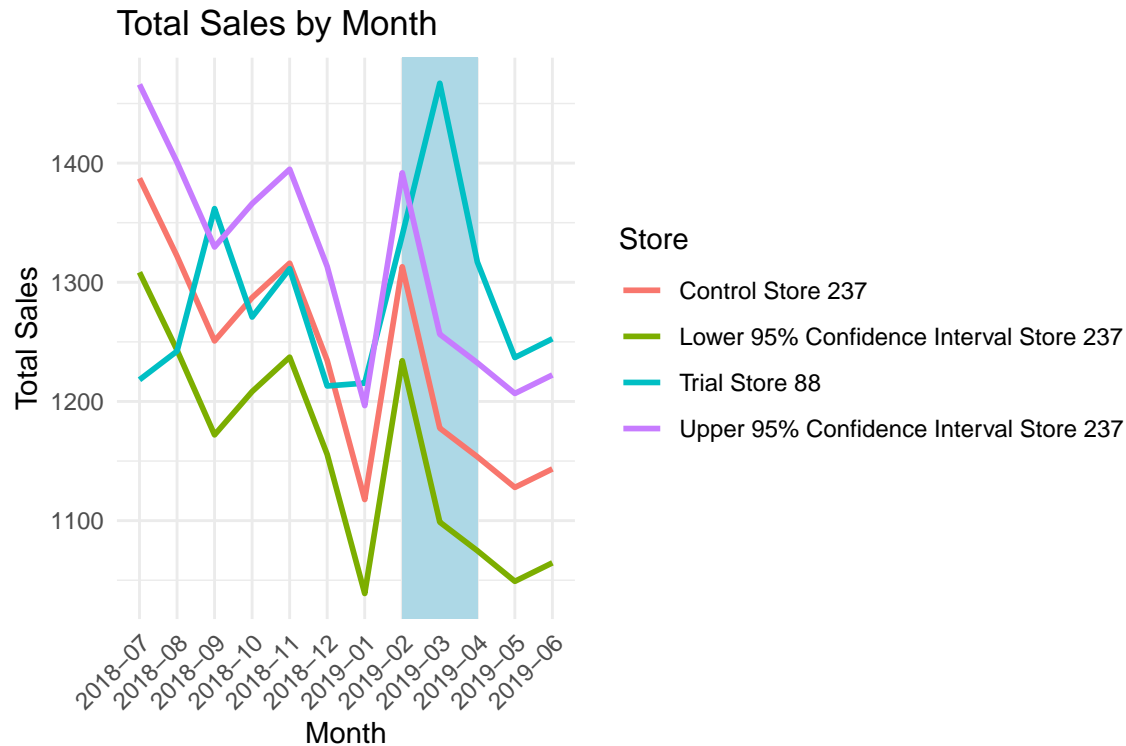
```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 88 | STORE_NBR == 237 )) %>%
  summarise( total = sum(TOT_SALES)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_237 = STORE_NBR_237*sum(STORE_NBR_88)/sum(STORE_NBR_237),
         Abs_diff_tot_with_control = abs(STORE_NBR_88 - Controlled_Store_NBR_237))
  trial <- Mod_Chip %>%
    filter(YMcode < "2019-02") %>%
    pull(Abs_diff_tot_with_control)
  Mean <- Mod_Chip %>%
    filter(YMcode == i) %>%
    pull(Abs_diff_tot_with_control)
  SD <- sd(trial)
  t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
  if(t > qt(p = 0.025,6, lower.tail = FALSE))
  { print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
  else
  {
    print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
  }
}
```

```
}
```

```
## [1] "Trial Month 2019-02 is not significant different compared to pre-trial month"  
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-04 is significant different compared to pre-trial month"
```

```
Chip %>%
```

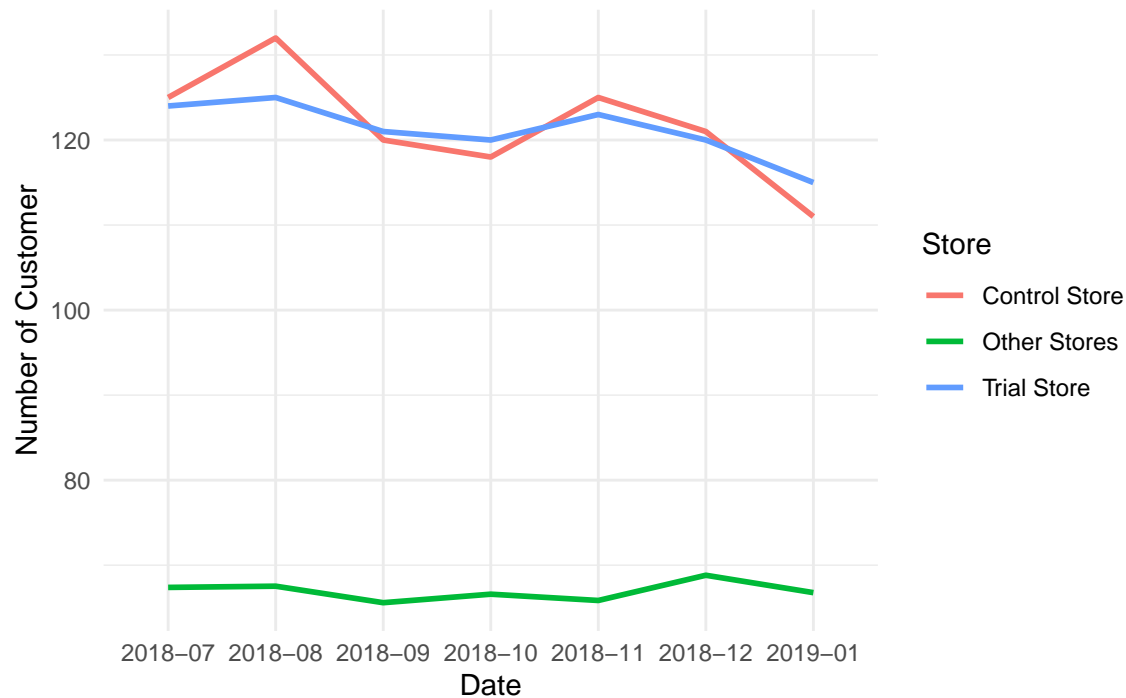
```
  mutate(YMcode = format(DATE, "%Y-%m")) %>%  
  group_by(YMcode, STORE_NBR) %>%  
  summarise(total = sum(TOT_SALES), .groups = "drop") %>%  
  filter(STORE_NBR == 88 | STORE_NBR == 237) %>%  
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%  
  mutate(upper = Store_Nbr237+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr237[YMcode < "2019-04-01"]),  
         lower = Store_Nbr237+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr237[YMcode < "2019-04-01"]),  
  pivot_longer(cols = c(Store_Nbr88, Store_Nbr237,upper,lower),  
               names_to = "Series",  
               values_to = "Values") %>%  
  mutate(Series = case_when(Series == "Store_Nbr237" ~ "Control Store 237",  
                           Series == "Store_Nbr88" ~ "Trial Store 88",  
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 237",  
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 237")) %>%  
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +  
    geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),  
                  xmax = format(as.Date("2019-04-30"), "%Y-%m"),  
                  ymin = -Inf, ymax = Inf),  
              fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +  
    geom_line(size = 1) +  
    labs(title = "Total Sales by Month",  
         x = "Month",  
         y = "Total Sales",  
         color = "Store") +  
    scale_y_continuous(breaks = seq(0, 1500, by = 100)) +  
    scale_x_discrete(guide = guide_axis(angle = 45)) +  
    theme_minimal()
```



b) Number of Customer

```
Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter(YMcode < "2019-02") %>%
  summarise(total = n_distinct(LYLT_CARD_NBR), .groups = "drop") %>%
  mutate(timeline = case_when(
    STORE_NBR == 88 ~ "Trial Store",
    STORE_NBR == 237 ~ "Control Store",
    TRUE ~ "Other Stores"
  )) %>%
  group_by(YMcode, timeline) %>%
  summarise(total = mean(total), .groups = "drop") %>%
  ggplot(aes(x = YMcode, y = total, color = timeline, group = timeline)) +
  geom_line(size = 1) +
  labs(title = "Number of Customer Comparison between Store 88 (Trial),Store 237 (Control) and",
    x = "Date",
    y = "Number of Customer",
    color = "Store") +
  theme_minimal()
```

Number of Customer Comparison between Store 88 (Trial),Store 237



```
trial_month <- c("2019-02","2019-03","2019-04")
for (i in trial_month) {
  Mod_Chip <- Chip %>%
  mutate(YMcode = format(DATE, "%Y-%m")) %>%
  group_by(YMcode, STORE_NBR) %>%
  filter((YMcode < "2019-02" | YMcode == i) & (STORE_NBR == 88 | STORE_NBR == 237 )) %>%
  summarise( total = n_distinct(LYLT_CARD_NBR)) %>%
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "STORE_NBR_") %>%
  ungroup() %>%
  mutate(Controlled_Store_NBR_237 = STORE_NBR_237*sum(STORE_NBR_88)/sum(STORE_NBR_237),
         Abs_diff_tot_with_control = abs(STORE_NBR_88 - Controlled_Store_NBR_237))
  trial <- Mod_Chip %>%
    filter(YMcode < "2019-02") %>%
    pull(Abs_diff_tot_with_control)
  Mean <- Mod_Chip %>%
    filter(YMcode == i) %>%
    pull(Abs_diff_tot_with_control)
  SD <- sd(trial)
  t <- abs(mean(trial)-Mean)/(SD/sqrt(7))
  if(t > qt(p = 0.025,6, lower.tail = FALSE))
  { print(paste("Trial Month",i, "is significant different compared to pre-trial month"))}
  else
  {
    print(paste("Trial Month",i, "is not significant different compared to pre-trial month"))
  }
}
```

```
}
```

```
## [1] "Trial Month 2019-02 is not significant different compared to pre-trial month"  
## [1] "Trial Month 2019-03 is significant different compared to pre-trial month"  
## [1] "Trial Month 2019-04 is not significant different compared to pre-trial month"
```

```
Chip %>%
```

```
  mutate(YMcode = format(DATE, "%Y-%m")) %>%  
  group_by(YMcode, STORE_NBR) %>%  
  summarise(total = n_distinct(LYLTY_CARD_NBR), .groups = "drop") %>%  
  filter(STORE_NBR == 88 | STORE_NBR == 237) %>%  
  pivot_wider(names_from = STORE_NBR, values_from = total, names_prefix = "Store_Nbr") %>%  
  mutate(upper = Store_Nbr237+qt(p = 0.025,6, lower.tail = FALSE)*sd(Store_Nbr237[YMcode < "2019-04-01"]),  
         lower = Store_Nbr237+qt(p = 0.025,6, lower.tail = TRUE)*sd(Store_Nbr237[YMcode < "2019-04-01"]),  
  pivot_longer(cols = c(Store_Nbr88, Store_Nbr237,upper,lower),  
               names_to = "Series",  
               values_to = "Values") %>%  
  mutate(Series = case_when(Series == "Store_Nbr237" ~ "Control Store 237",  
                           Series == "Store_Nbr88" ~ "Trial Store 88",  
                           Series == "upper" ~ "Upper 95% Confidence Interval Store 237",  
                           Series == "lower" ~ "Lower 95% Confidence Interval Store 237")) %>%  
  ggplot(aes(x = YMcode, y = Values, color = Series, group = Series)) +  
    geom_rect(aes(xmin = format(as.Date("2019-02-01"), "%Y-%m"),  
                  xmax = format(as.Date("2019-04-30"), "%Y-%m"),  
                  ymin = -Inf, ymax = Inf),  
              fill = "lightblue", alpha = 0.3, inherit.aes = FALSE) +  
    geom_line(size = 1) +  
    labs(title = "Number of Customer by Month",  
         x = "Month",  
         y = "Number of Customer",  
         color = "Store") +  
    scale_y_continuous(breaks = seq(0, 150, by = 10)) +  
    scale_x_discrete(guide = guide_axis(angle = 45)) +  
    theme_minimal()
```

