Customer Segments Analysis

2025-03-08

Load Data

```r
# Load transaction data
transaction_data <- read_excel("C:/Users/udayu/OneDrive/Documents/chips/QVI_transaction_data.xlsx")

# Load customer segmentation data
customer_data <- read_csv("C:/Users/udayu/OneDrive/Documents/chips/QVI_purchase_behaviour.csv")
```

```
## Rows: 72637 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): LIFESTAGE, PREMIUM_CUSTOMER
## dbl (1): LYLTY_CARD_NBR
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Data Exploration

```r
# Check for missing values
sum(is.na(transaction_data))
```

```
## [1] 0
```

```r
sum(is.na(customer_data))
```

```
## [1] 0
```

```r
# Display first few rows
head(transaction_data)
```

```
## # A tibble: 6 x 8
##     DATE STORE_NBR LYLTY_CARD_NBR TXN_ID PROD_NBR PROD_NAME   PROD_QTY TOT_SALES
##    <dbl>     <dbl>          <dbl>  <dbl>    <dbl> <chr>          <dbl>     <dbl>
## 1 43390         1           1000      1        5 Natural Chi~       2       6
## 2 43599         1           1307    348       66 CCs Nacho C~       3       6.3
## 3 43605         1           1343    383       61 Smiths Crin~       2       2.9
## 4 43329         2           2373    974       69 Smiths Chip~       5      15
## 5 43330         2           2426   1038      108 Kettle Tort~       3      13.8
## 6 43604         4           4074   2982       57 Old El Paso~       1       5.1
```

```
head(customer_data)
```

```
## # A tibble: 6 x 3
##   LYLTY_CARD_NBR LIFESTAGE             PREMIUM_CUSTOMER
##            <dbl> <chr>                 <chr>
## 1           1000 YOUNG SINGLES/COUPLES  Premium
## 2           1002 YOUNG SINGLES/COUPLES  Mainstream
## 3           1003 YOUNG FAMILIES         Budget
## 4           1004 OLDER SINGLES/COUPLES  Mainstream
## 5           1005 MIDAGE SINGLES/COUPLES Mainstream
## 6           1007 YOUNG SINGLES/COUPLES  Budget
```

```
# Structure and summary
str(customer_data)
```

```
## spc_tbl_ [72,637 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ LYLTY_CARD_NBR  : num [1:72637] 1000 1002 1003 1004 1005 ...
##  $ LIFESTAGE       : chr [1:72637] "YOUNG SINGLES/COUPLES" "YOUNG SINGLES/COUPLES" "YOUNG FAMILIES"
##  $ PREMIUM_CUSTOMER: chr [1:72637] "Premium" "Mainstream" "Budget" "Mainstream" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   LYLTY_CARD_NBR = col_double(),
##   ..   LIFESTAGE = col_character(),
##   ..   PREMIUM_CUSTOMER = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
summary(customer_data)
```

```
##  LYLTY_CARD_NBR     LIFESTAGE         PREMIUM_CUSTOMER
##  Min.   :   1000   Length:72637      Length:72637
##  1st Qu.:  66202   Class :character   Class :character
##  Median : 134040   Mode  :character   Mode  :character
##  Mean   : 136186
##  3rd Qu.: 203375
##  Max.   :2373711
```

```
# Distribution of customer segments
table(customer_data$LIFESTAGE)
```

```
##
## MIDAGE SINGLES/COUPLES          NEW FAMILIES        OLDER FAMILIES
##                  7275                  2549                  9780
##  OLDER SINGLES/COUPLES              RETIREES        YOUNG FAMILIES
##                 14609                 14805                  9178
##  YOUNG SINGLES/COUPLES
##                 14441
```

```
table(customer_data$PREMIUM_CUSTOMER)
```

```
##
##      Budget Mainstream    Premium
##       24470      29245      18922
```

Merge Datasets

```
merged_data <- merge(transaction_data, customer_data, by = "LYLTY_CARD_NBR")

# Check structure and missing values
str(merged_data)
```

```
## 'data.frame':    264836 obs. of  10 variables:
##  $ LYLTY_CARD_NBR  : num  1000 1002 1003 1003 1004 ...
##  $ DATE            : num  43390 43359 43531 43532 43406 ...
##  $ STORE_NBR       : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ TXN_ID          : num  1 2 3 4 5 6 8 7 9 10 ...
##  $ PROD_NBR        : num  5 58 52 106 96 86 10 49 20 51 ...
##  $ PROD_NAME       : chr  "Natural Chip        Compny SeaSalt175g" "Red Rock Deli Chikn&Garlic Aioli
##  $ PROD_QTY        : num  2 1 1 1 1 1 1 1 1 2 ...
##  $ TOT_SALES       : num  6 2.7 3.6 3 1.9 2.8 2.7 3.8 5.7 8.8 ...
##  $ LIFESTAGE       : chr  "YOUNG SINGLES/COUPLES" "YOUNG SINGLES/COUPLES" "YOUNG FAMILIES" "YOUNG FAI
##  $ PREMIUM_CUSTOMER: chr  "Premium" "Mainstream" "Budget" "Budget" ...
```

```
sum(is.na(merged_data))
```

```
## [1] 0
```

Sales Analysis

```
# Total sales by customer segment
sales_by_segment <- merged_data %>%
  group_by(PREMIUM_CUSTOMER) %>%
  summarise(total_sales = sum(TOT_SALES, na.rm = TRUE))

# Plot total sales by customer segment
ggplot(sales_by_segment, aes(x = PREMIUM_CUSTOMER, y = total_sales, fill = PREMIUM_CUSTOMER)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Sales by Customer Segment",
       x = "Customer Type",
       y = "Total Sales ($)") +
  theme_minimal()
```
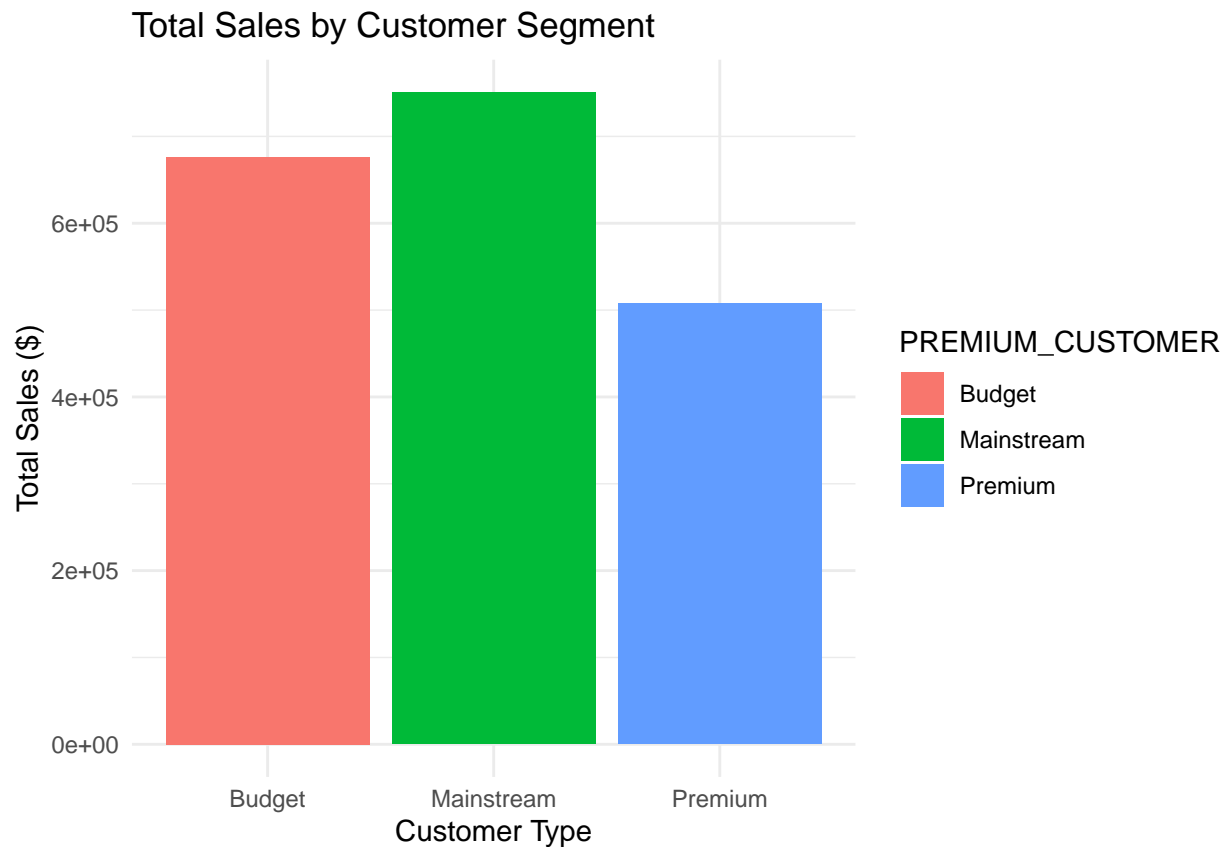
Total Sales by Customer Segment

Sales by Life Stage

```
sales_by_lifestage <- merged_data %>%
  group_by(LIFESTAGE) %>%
  summarise(total_sales = sum(TOT_SALES, na.rm = TRUE)) %>%
  arrange(desc(total_sales))

# Plot sales by life stage
ggplot(sales_by_lifestage, aes(x = reorder(LIFESTAGE, -total_sales), y = total_sales, fill = LIFESTAGE)
  geom_bar(stat = "identity") +
  labs(title = "Total Sales by Life Stage",
       x = "Life Stage",
       y = "Total Sales ($)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Total Sales by Life Stage
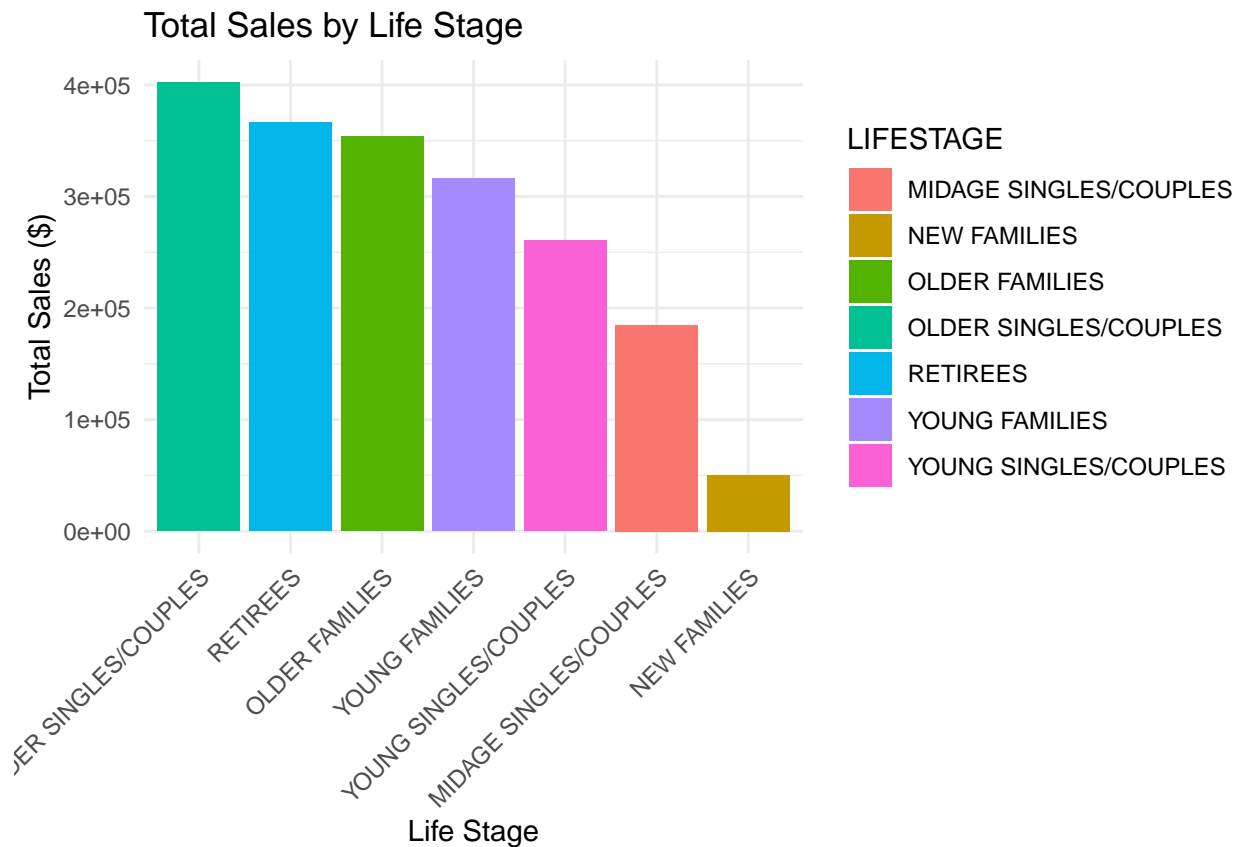
Brand Analysis

```r
# Extract brand name from product description
merged_data$BRAND <- str_extract(merged_data$PROD_NAME, "^[^ ]+")

# Find top brands
top_brands <- merged_data %>%
  group_by(BRAND) %>%
  summarise(total_sales = sum(TOT_SALES, na.rm = TRUE)) %>%
  arrange(desc(total_sales))

# Plot top 10 brands
ggplot(top_brands[1:10, ], aes(x = reorder(BRAND, -total_sales), y = total_sales, fill = BRAND)) +
  geom_bar(stat = "identity") +
  labs(title = "Top 10 Best-Selling Chip Brands",
       x = "Brand",
       y = "Total Sales") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 10 Best–Selling Chip Brands
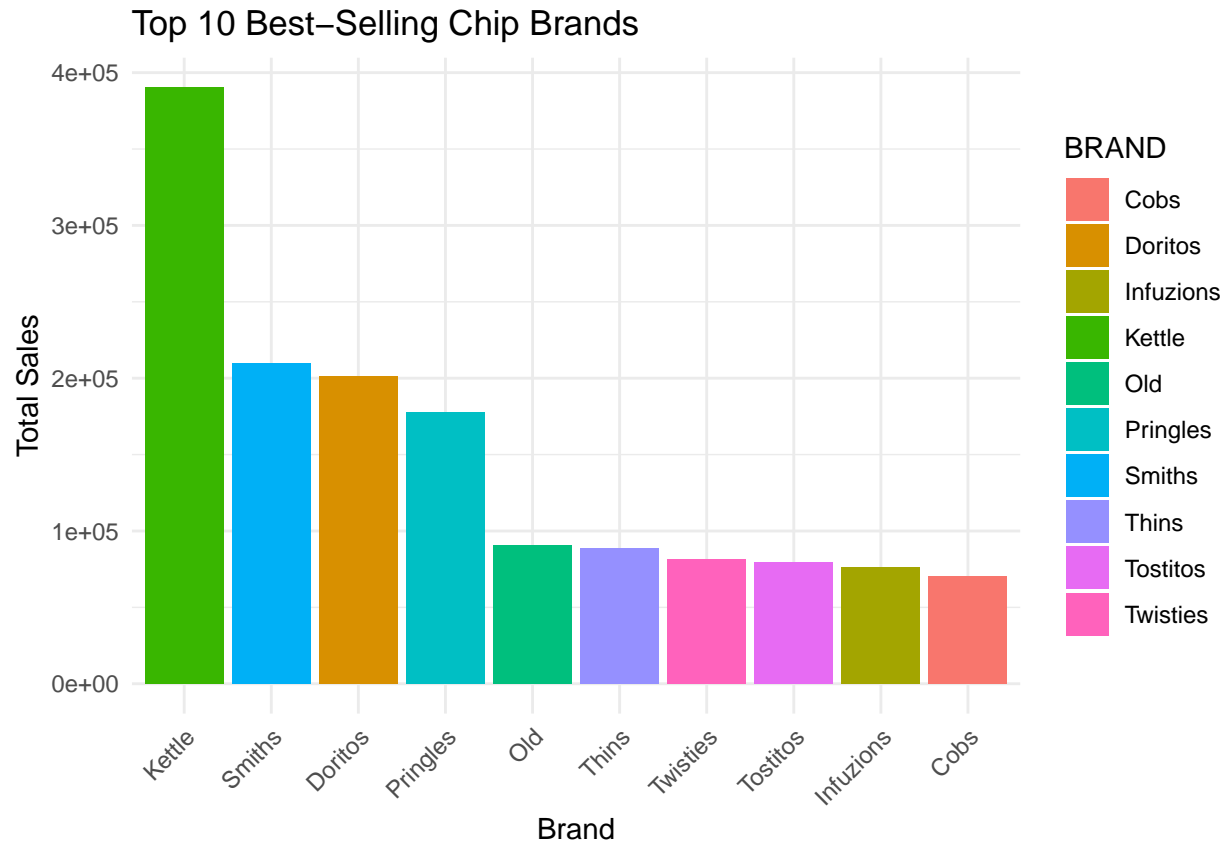
Pack Size Analysis

```r
# Extract pack size
merged_data$PACK_SIZE <- as.numeric(str_extract(merged_data$PROD_NAME, "\\d+"))

# Sales by pack size
pack_size_sales <- merged_data %>%
  group_by(PACK_SIZE) %>%
  summarise(Total_Sales = sum(TOT_SALES, na.rm = TRUE)) %>%
  arrange(desc(Total_Sales))

# Plot preferred pack sizes
ggplot(pack_size_sales, aes(x = PACK_SIZE, y = Total_Sales)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(title = "Preferred Pack Sizes",
       x = "Pack Size (grams)",
       y = "Total Sales") +
  theme_minimal()
```

Preferred Pack Sizes
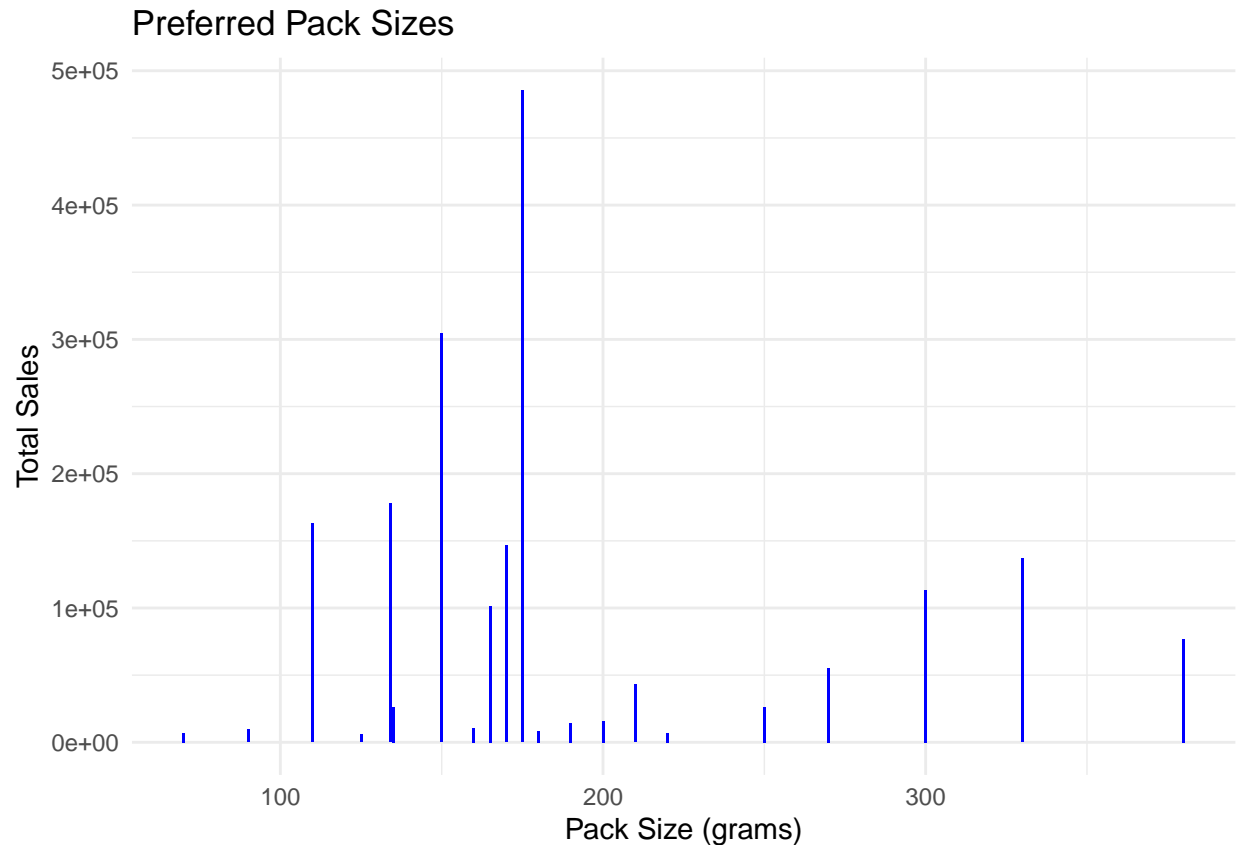
Conclusion & Recommendations

1. High-Spending Customer Segments

- Young Singles/Couples (Premium) and Mid-Age Families (Premium) spend the most per transaction.
- Older Families (Budget) buy the highest quantity but spend less per transaction.

Recommendation: - Focus marketing campaigns on Young Singles/Couples (Premium) and Mid-Age Families (Premium) as they are more likely to spend more. - Introduce loyalty programs or premium chip varieties to retain high-value customers.

2. Budget vs. Premium Customers

- Premium customers spend more per transaction than budget customers.
- Budget customers buy in larger quantities but at lower price points.

Recommendation: - Introduce bulk-buy discounts or value packs to attract budget-conscious shoppers. - Offer limited-edition flavors or bundles for premium customers to increase spending.

3. Most Popular Pack Sizes

- Mid-sized packs (150g - 200g) have the highest sales.
- Smaller pack sizes (<100g) contribute less to overall sales.

Recommendation: - Continue stocking and promoting mid-sized packs. - Offer "value packs" (multiple mid-sized packs) to encourage bulk purchases.

4. Brand Performance & Promotional Strategy

- Certain brands dominate in premium and budget segments.
- High sales from mainstream brands, but premium customers prefer niche brands.

Recommendation: - Partner with leading brands for exclusive promotions (e.g., "Buy 2 Get 1 Free"). - Introduce in-store sampling for premium brands to increase trial purchases.

Strategic Action Plan

Increase customer retention in high-spending segments. Expand mid-sized packs and introduce premium flavors. Target budget-conscious families with combo deals and value packs. Run seasonal promotions with top-performing brands to maximize sales.

By implementing these strategies, the chips category can drive higher revenue and stronger customer engagement over the next half-year.

```
## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS W

When you click the **Knit** button a document will be generated that includes both content as well as tl


``` r
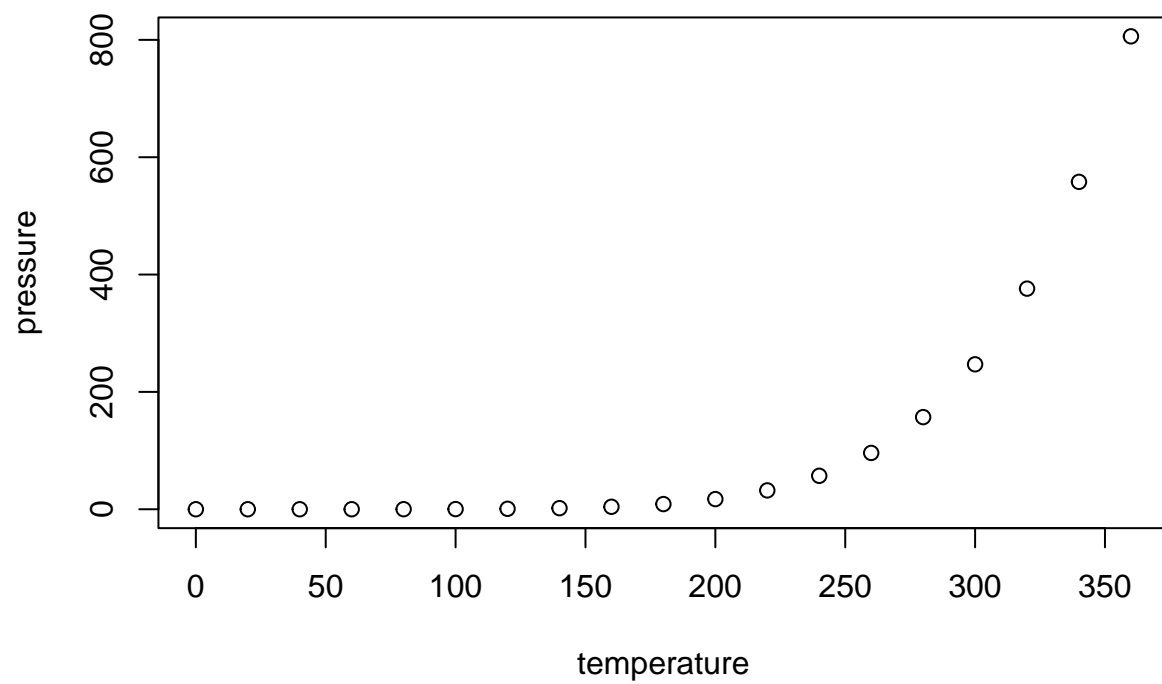summary(cars)

##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.