

# Viz5 - Dashboards

Hayden Ginman

2023-10-01

```
#Step 1 - Load requisite libraries into coding environment
options(repos = c(CRAN = "https://cloud.r-project.org/"))
#install.packages("scales")
#install.packages("lubridate")
library("lubridate")
```

```
##
## Attaching package: 'lubridate'

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

```
library("scales")
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("reshape2")
library("readxl")
library("viridis")
```

```
## Loading required package: viridisLite
```

```
##
## Attaching package: 'viridis'

## The following object is masked from 'package:scales':
##
##    viridis_pal
```

```
#Step 2 - Read clean data into coding environment
```

```
iowa_sales <- read_xlsx("Smaller_Iowa_Liquor_Sales.xlsx", sheet = 1, .name_repair = "universal")
```

```
#Step 3 - Summarize the top categories and group them by category name. Summarize total sales. Do the s
```

```

top_categories <- iowa_sales %>%
  group_by(Category.Name) %>%
  summarize(TotalSales = sum(State.Bottle.Retail * Bottles.Sold))
top_categories <- top_categories %>%
  arrange(-TotalSales) %>%
  head(10)

worst_categories <- iowa_sales %>%
  group_by(Category.Name) %>%
  summarize(TotalSales = sum(State.Bottle.Retail * Bottles.Sold)) %>%
  arrange(TotalSales) %>%
  head(10)

```

```
top_categories
```

```

## # A tibble: 10 x 2
##   Category.Name      TotalSales
##   <chr>             <dbl>
## 1 AMERICAN VODKAS    21007688.
## 2 CANADIAN WHISKIES  16438968.
## 3 SPICED RUM         9780703.
## 4 IMPORTED VODKAS    8235447.
## 5 STRAIGHT BOURBON WHISKIES 7942365.
## 6 WHISKEY LIQUEUR    6817983.
## 7 TENNESSEE WHISKIES 5895243.
## 8 100% AGAVE TEQUILA 4924266.
## 9 AMERICAN FLAVORED VODKA 4627030.
## 10 IMPORTED BRANDIES 4564791.

```

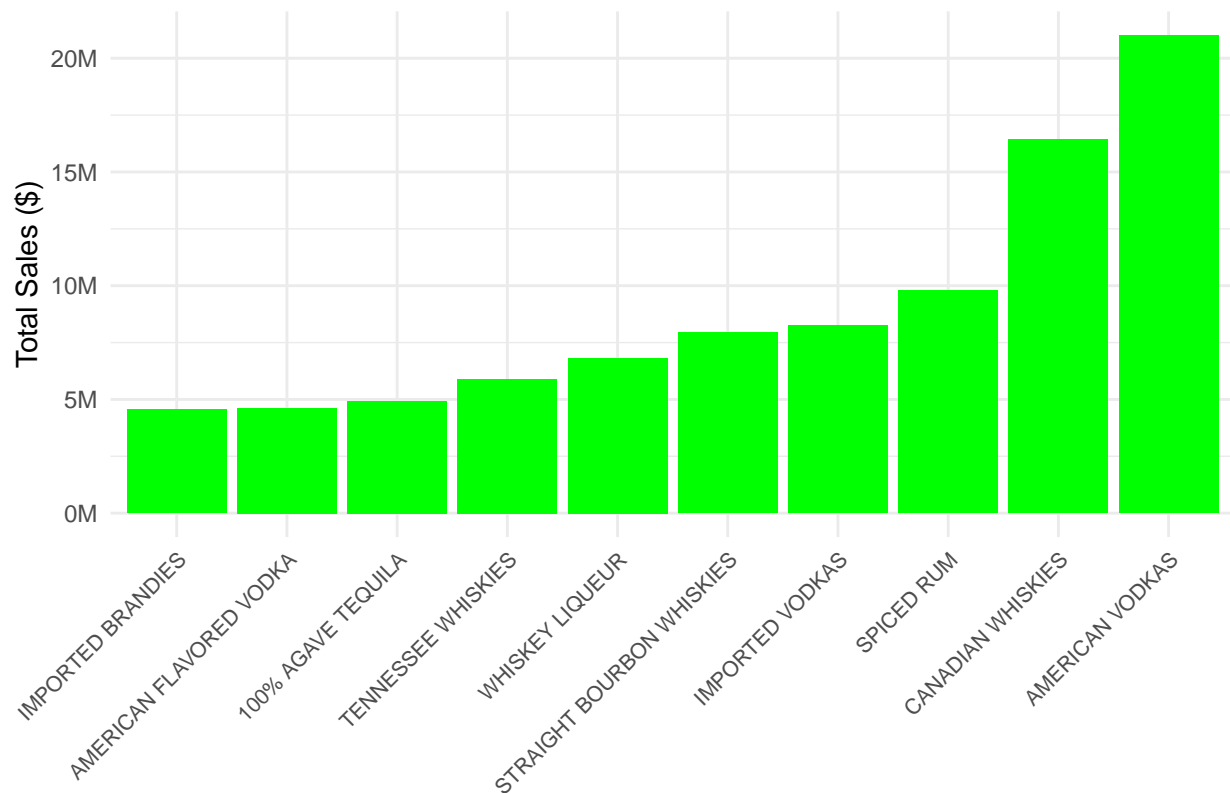
*#Step 4 - Visualize the Top 10 Total sales by category*

```

ggplot(top_categories, aes(x = reorder(Category.Name, TotalSales), y = TotalSales)) +
  geom_bar(stat = "identity", fill = "Green") +
  labs(title = "Top 10 Alcohol Categories by Total Sales",
       x = NULL,
       y = "Total Sales ($)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8)) +
  scale_y_continuous(labels = scales::label_number(scale = 1e-6, suffix = "M"))

```

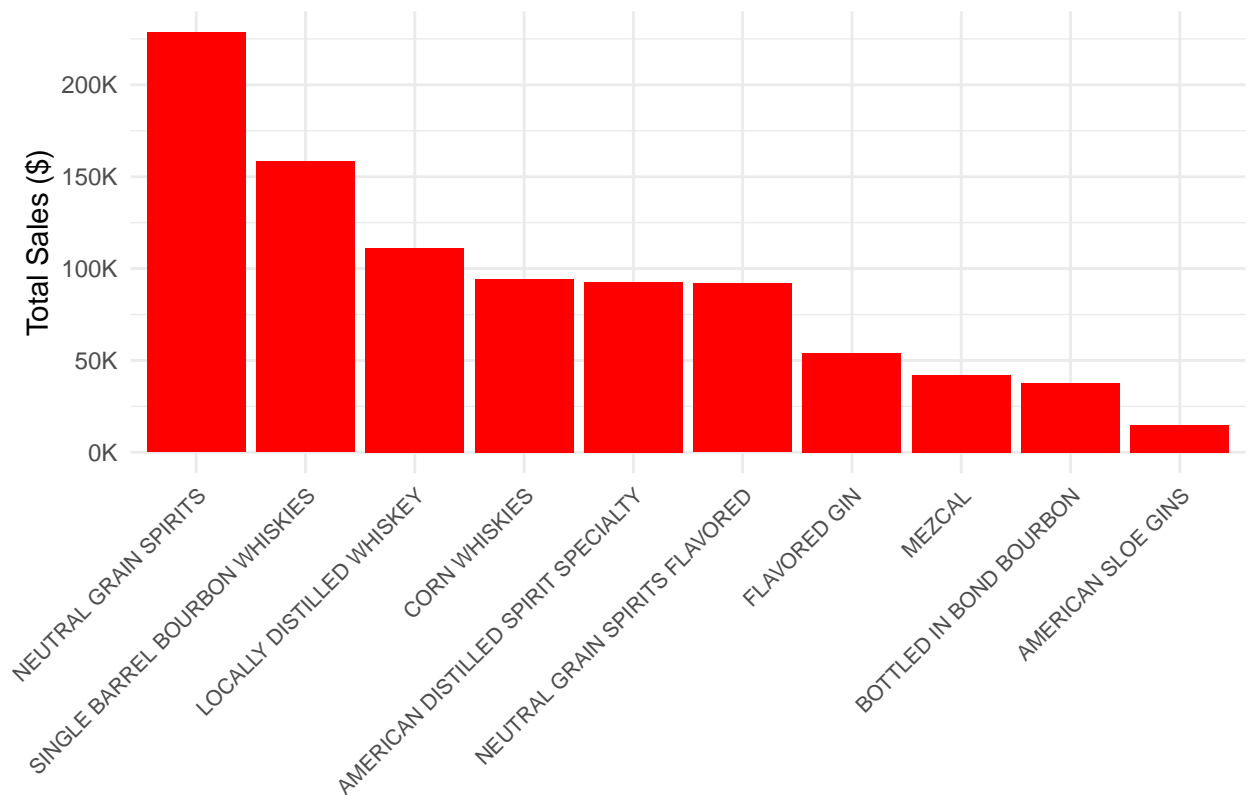
Top 10 Alcohol Categories by Total Sales



*# Visualize the bottom 10 selling alcohol categories*

```
ggplot(worst_categories, aes(x = reorder(Category.Name, -TotalSales), y = TotalSales)) +
  geom_bar(stat = "identity", fill = "Red") +
  labs(title = "Bottom 10 Alcohol Categories by Total Sales",
       x = NULL,
       y = "Total Sales ($)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 8)) +
  scale_y_continuous(labels = scales::label_number(scale = 1e-3, suffix = "K"))
```

Bottom 10 Alcohol Categories by Total Sales



*#Step 5 - Make line plot for sales over time for top selling categories*

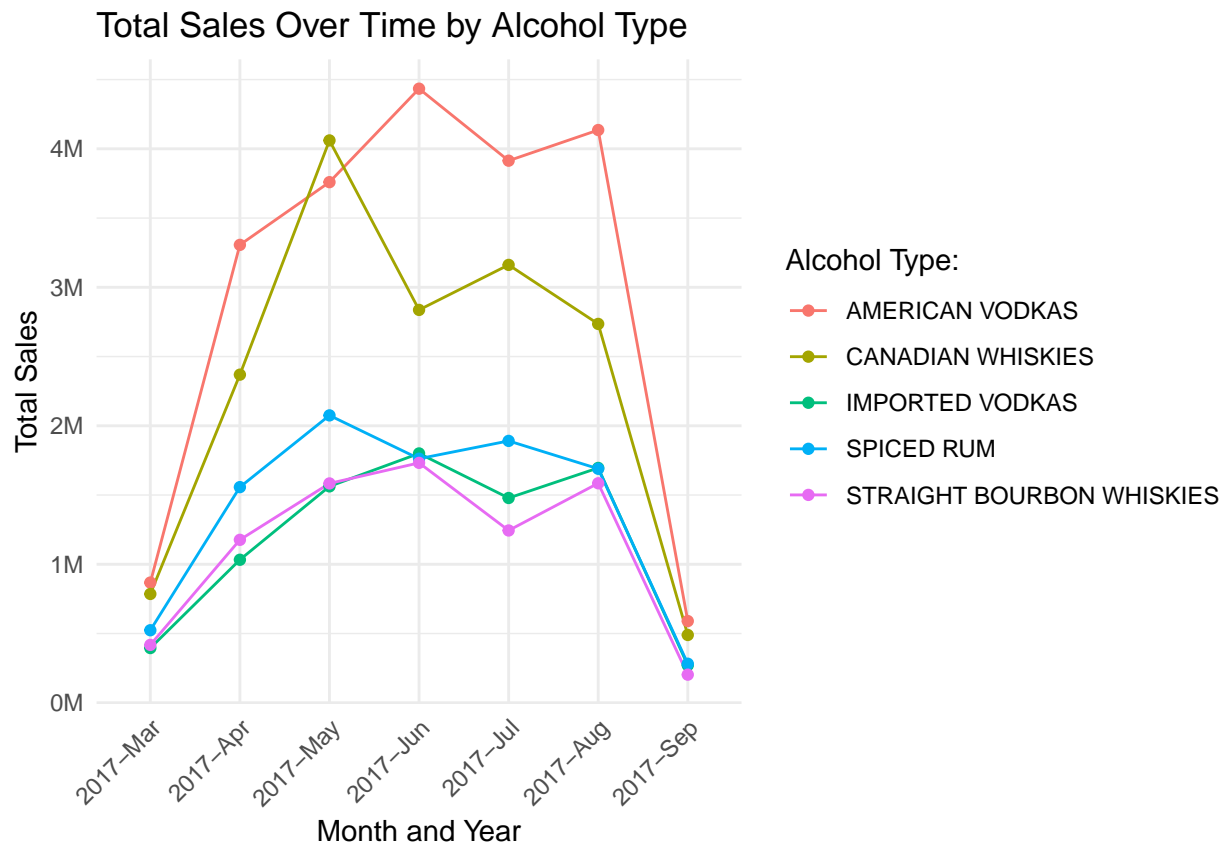
```
filtered_data <- iowa_sales %>%
  filter(Category.Name %in% c("AMERICAN VODKAS", "CANADIAN WHISKIES", "SPICED RUM", "IMPORTED VODKAS", "

filtered_data$Total.Sales <- filtered_data$State.Bottle.Retail * filtered_data$Bottles.Sold

summed_sales <- filtered_data %>%
  mutate(Year = year(Date), Month = month(Date, label = TRUE)) %>%
  group_by(Year, Month, Category.Name) %>%
  summarise(Total.Sales = sum(Total.Sales))

## `summarise()` has grouped output by 'Year', 'Month'. You can override using the
## `.groups` argument.

ggplot(summed_sales, aes(x = interaction(Year, Month, sep = "-"), y = Total.Sales, color = Category.Name)) +
  geom_line(aes(group = Category.Name)) +
  geom_point() +
  labs(title = "Total Sales Over Time by Alcohol Type", x = "Month and Year", y = "Total Sales", color = "Category") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::label_number(scale = 1e-6, suffix = "M"))
```



*# Make visualization for best selling American Vodkas*

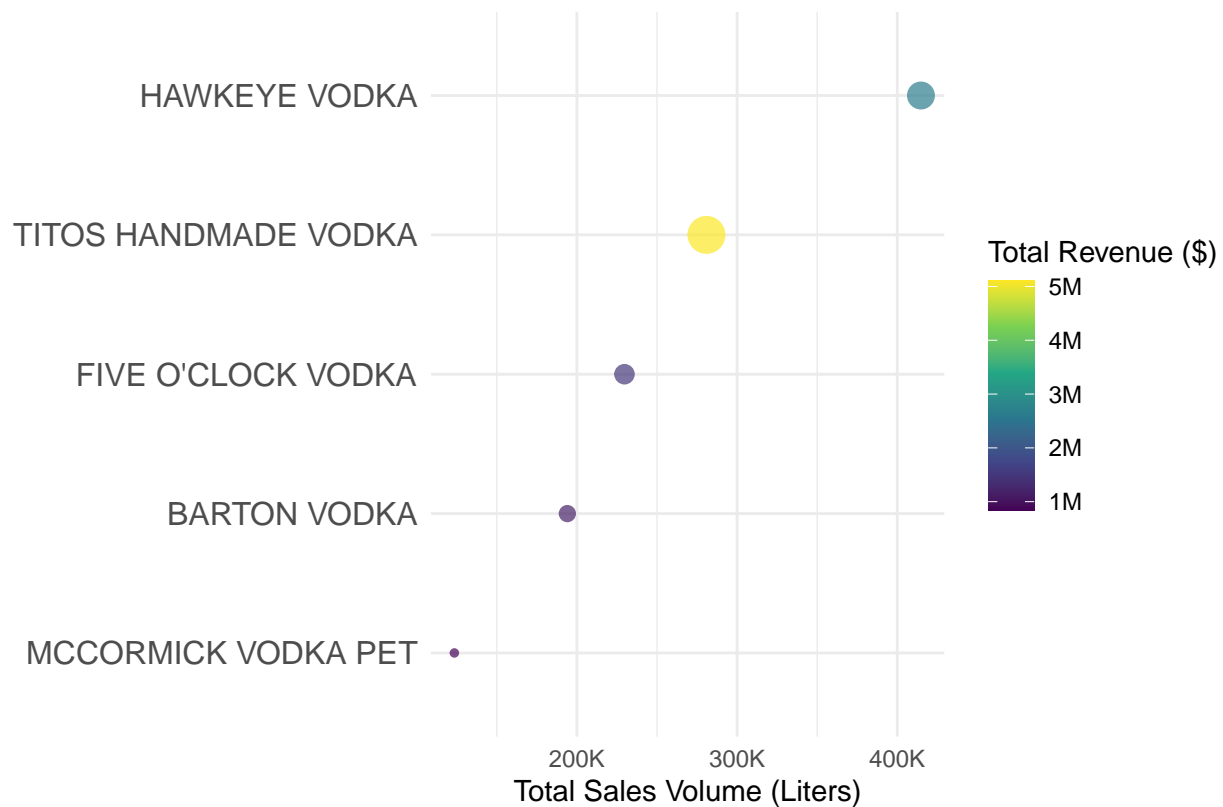
```
american_vodka_sales <- iowa_sales %>%
  filter(Category.Name == "AMERICAN VODKAS") %>%
  group_by(Item.Description) %>%
  summarize(TotalVolume = sum(Volume.Sold.Liters), TotalRevenue = sum(Bottles.Sold * State.Bottle.RetailPrice))
  arrange(-TotalVolume)

selected_vodkas <- c("HAWKEYE VODKA", "TITOS HANDMADE VODKA", "FIVE O'CLOCK VODKA", "BARTON VODKA", "MCCORMICK VODKA")

filtered_vodka_sales <- american_vodka_sales %>%
  filter(Item.Description %in% selected_vodkas)

ggplot(filtered_vodka_sales, aes(x = reorder(Item.Description, TotalVolume), y = TotalVolume)) +
  geom_point(aes(size = TotalRevenue, color = TotalRevenue), alpha = 0.7) +
  coord_flip() +
  theme_minimal() +
  labs(title = "Revenue and Sales Volume of Top Selling American Vodka Brands", x = NULL, y = "Total Sales Volume (Liters)") +
  theme(axis.text.y = element_text(size = 12), plot.title = element_text(size = 10)) +
  guides(size = "none") +
  scale_color_viridis_c(name = "Total Revenue ($)", labels = scales::label_number(scale = 1e-6, suffix = "K")) +
  scale_y_continuous(labels = scales::label_number(scale = 1e-3, suffix = "K"))
```

Revenue and Sales Volume of Top Selling American Vodka Brands



*# Calculate profit margin for best selling vodkas and visualize it*

```
profit_margin_data <- iowa_sales %>%
  filter(Item.Description %in% selected_vodkas) %>%
  group_by(Item.Description) %>%
  summarize(
    TotalRevenue = sum(State.Bottle.Retail * Bottles.Sold),
    TotalCost = sum(State.Bottle.Cost * Bottles.Sold) %>%
    mutate(ProfitMargin = (TotalRevenue - TotalCost) / TotalRevenue * 100)

profit_margin_data
```

```
## # A tibble: 5 x 4
##   Item.Description      TotalRevenue TotalCost ProfitMargin
##   <chr>                <dbl>      <dbl>      <dbl>
## 1 BARTON VODKA          1163141.    775014.    33.4
## 2 FIVE O'CLOCK VODKA    1491609.    993805.    33.4
## 3 HAWKEYE VODKA         2637605.   1757743.    33.4
## 4 MCCORMICK VODKA PET    843175.    562006.    33.3
## 5 TITOS HANDMADE VODKA  5111072.   3407003.    33.3
```

*#Make plot for alcohol that generates the least amount of revenue*

```
iowa_sales <- iowa_sales %>%
  mutate(Revenue = State.Bottle.Retail * Bottles.Sold)

least_revenue_alcohol <- iowa_sales %>%
  group_by(Item.Description) %>%
```

```

summarize(TotalRevenue = sum(Revenue)) %>%
arrange(TotalRevenue)

least_revenue_alcohol <- head(least_revenue_alcohol, 10)

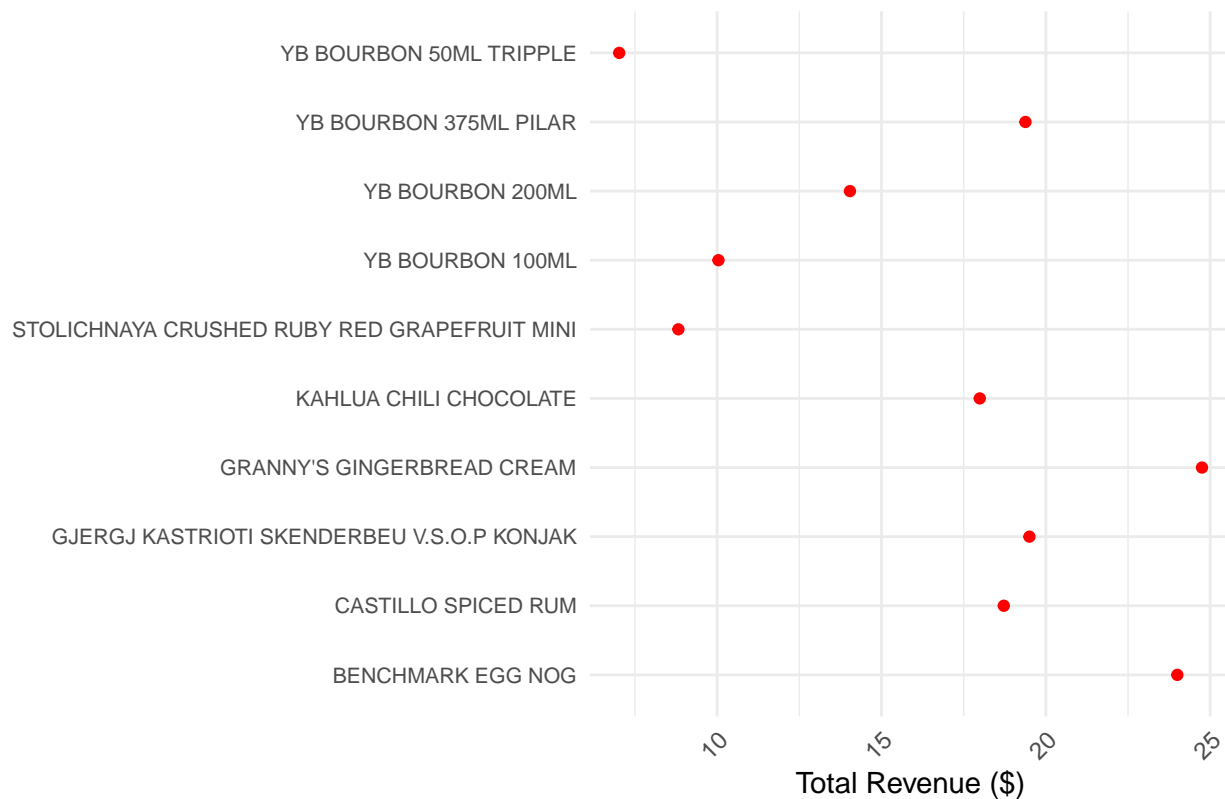
least_revenue_alcohol

## # A tibble: 10 x 2
##   Item.Description      TotalRevenue
##   <chr>                <dbl>
## 1 YB BOURBON 50ML TRIPPLE      7.02
## 2 STOLICHNAYA CRUSHED RUBY RED GRAPEFRUIT MINI      8.82
## 3 YB BOURBON 100ML           10.0
## 4 YB BOURBON 200ML           14.0
## 5 KAHLUA CHILI CHOCOLATE      18.0
## 6 CASTILLO SPICED RUM         18.7
## 7 YB BOURBON 375ML PILAR      19.4
## 8 GJERGJ KASTRIOTI SKENDERBEU V.S.O.P KONJAK      19.5
## 9 BENCHMARK EGG NOG          24
## 10 GRANNY'S GINGERBREAD CREAM      24.8

ggplot(least_revenue_alcohol, aes(x = Item.Description, y = TotalRevenue)) +
  geom_point(color = "red") +
  coord_flip() +
  labs(title = "Bottom 10 Products by Revenue", x = NULL, y = "Total Revenue ($)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme(axis.text.y = element_text(size = 8))

```

## Bottom 10 Products by Revenue



*#Visualize worst performing cities by total revenue*

```
city_sales <- iowa_sales %>%
  group_by(City) %>%
  summarize(TotalRevenue = sum(State.Bottle.Retail * Bottles.Sold)) %>%
  arrange(TotalRevenue)

worst_cities <- head(city_sales, 10)

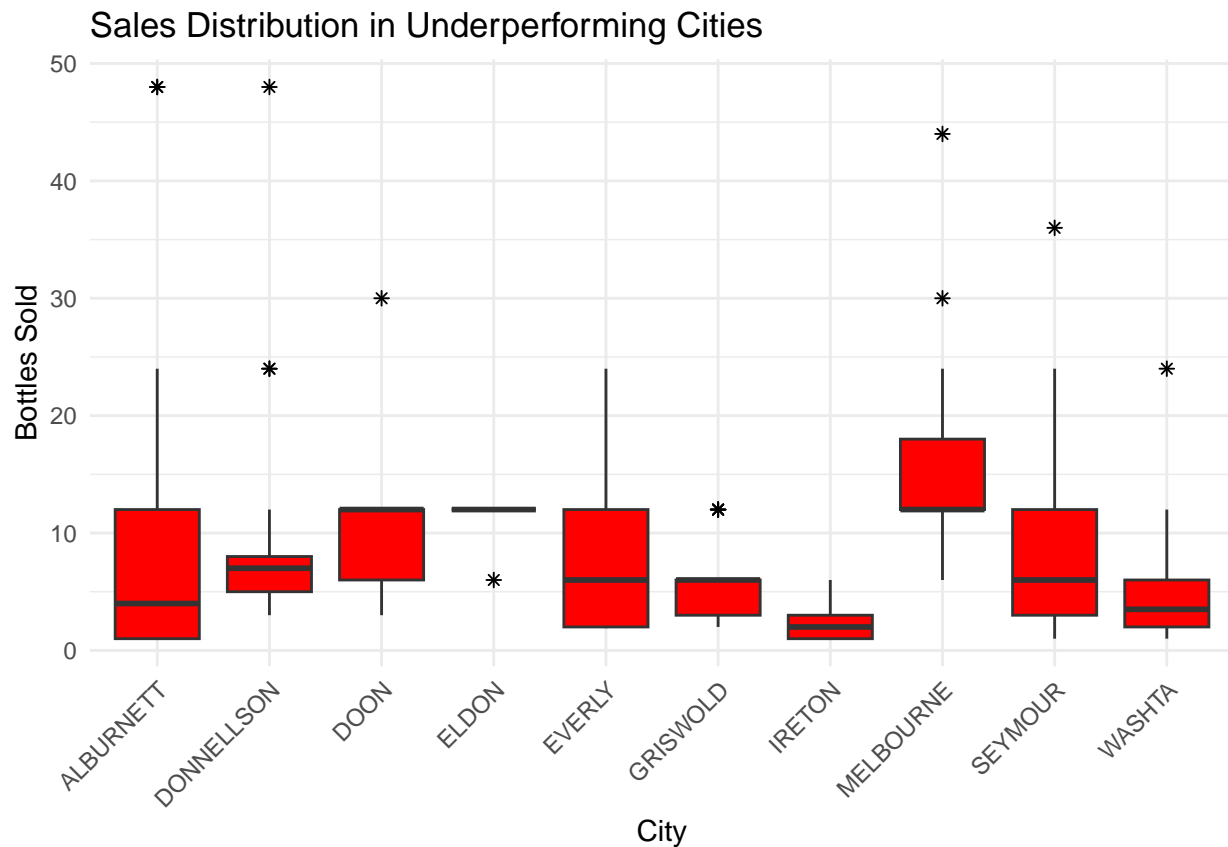
best_cities <- tail(city_sales, 10)

worst_city_sales <- iowa_sales %>%
  filter(City %in% worst_cities$City)

best_city_sales <- iowa_sales %>%
  filter(City %in% best_cities$City)

ggplot(worst_city_sales, aes(x = City, y = Bottles.Sold)) +
  geom_boxplot(fill = "red", outlier.colour = "black", outlier.shape = 8) +
  labs(title = "Sales Distribution in Underperforming Cities",
       x = "City",
       y = "Bottles Sold") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```





```
ggplot(best_cities, aes(x = reorder(City, TotalRevenue), y = TotalRevenue)) +
  geom_point(color = "green", size = 3) +
  labs(title = "Best Performing Cities",
       x = NULL,
       y = "Total Revenue ($)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_y_continuous(labels = scales::label_number(scale = 1e-6, suffix = "M"))
```

