Gas Station A

2023-08-23

Relationship between Temperature Changes and Soda Sales for a Convenience Store

Shamikh Maredia

Introduction

The analysis we were running was based on a convenience store and gas station located in Spring Branch, TX. And its location in the Lone Star State is ideal for looking at temperature data. In a state where blistering summers can last from March to September and winters that can lead to power outages and lows of 10-20 degrees, this varied climate distribution will really be able to extrapolate how temperature can affect sales.

I hypothesize that higher temperature days will lead to greater sales in soda due to the nature of drinking a cold drink to alleviate the unforgiving Texas heat.

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.3     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflic
ts to become errors
```

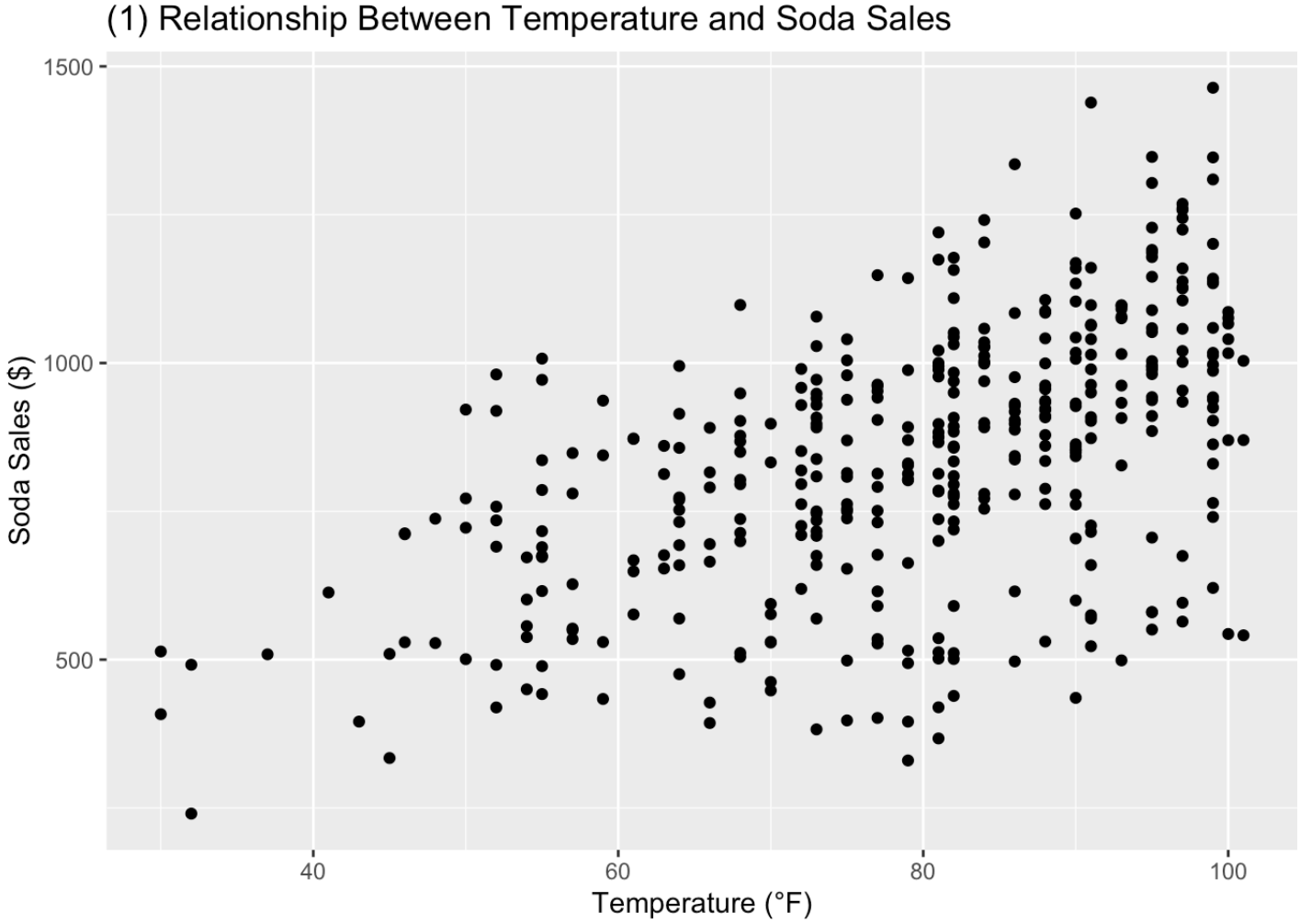
```
library(readxl)
library(forcats)
```

Data was collected from the daily entries of Gas Station A in the Modisoft program containing soda sales, gas sales, and coffee sales. The temperature data for the location was also collected and all entries were categorized by season and day of the week.

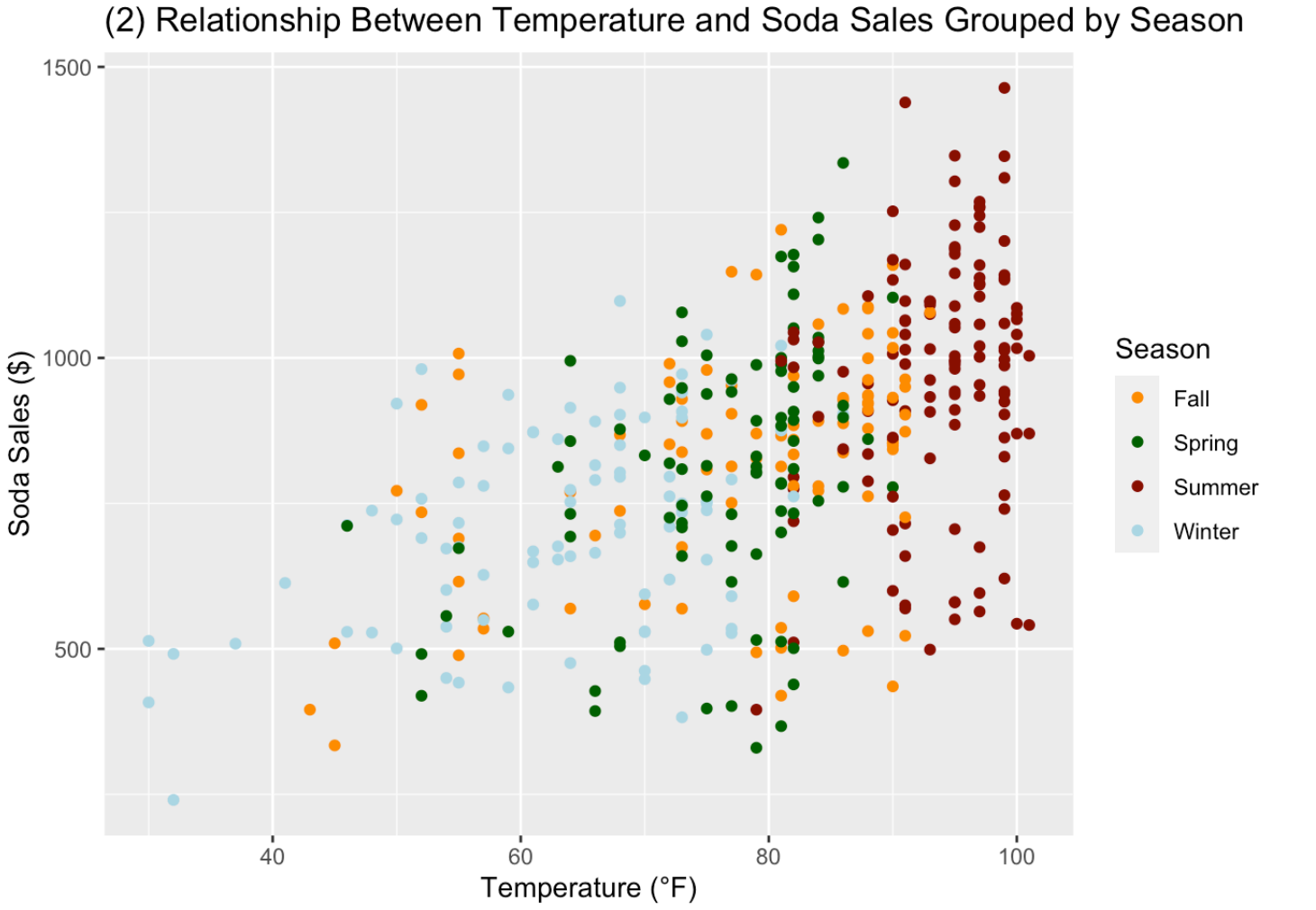
The data was then tidied, cleaned and combined into a single dataset for efficient manipulation. The following are the results with included visualizations.

Visualizations

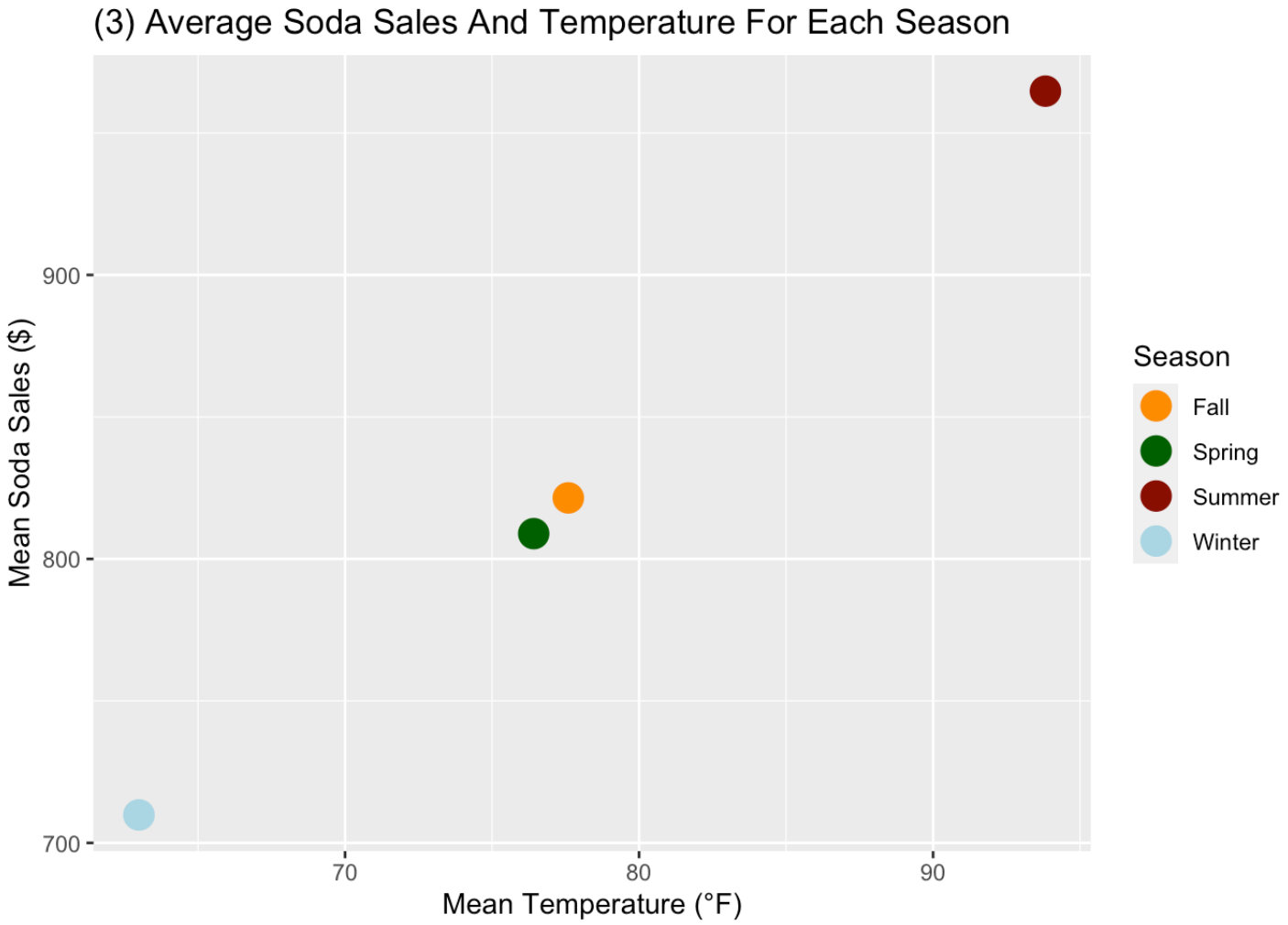
```
data <- read_excel('CDATA.xlsx')

tidydata <- pivot_longer(data,
                  cols = c(`Spring`, `Summer`, `Fall`, `Winter`),
                  names_to = "Season",
                  values_to = "SeasonBinary")

datafinal <- tidydata %>%
  filter(SeasonBinary == 1) %>%
  select(1, 3, 4, 5, 7, 8, 18)

ggplot(datafinal, aes(y = SodaSales, x = Temperature)) +
  geom_point() +
  labs(title = "(1) Relationship Between Temperature and Soda Sales",
       x = "Temperature (°F)",
       y = "Soda Sales ($)")
```



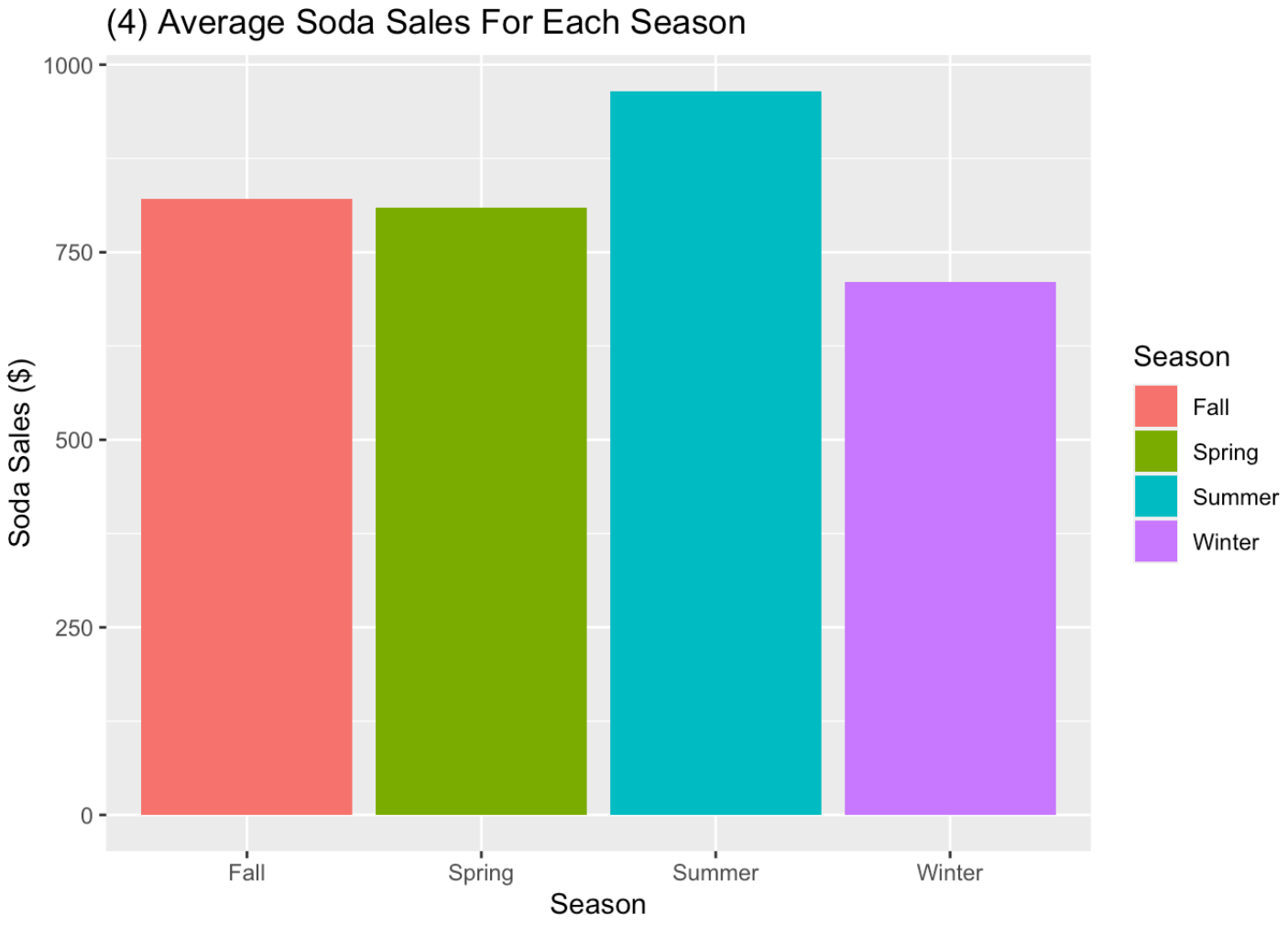

```
ggplot(datafinal, aes(y = SodaSales, x = Temperature, color = Season)) +
  geom_point() +
  labs(title = "(2) Relationship Between Temperature and Soda Sales Grouped by Season
",
       x = "Temperature (°F)",
       y = "Soda Sales ($)") +
  scale_color_manual(values = c("Fall" = "darkorange", "Spring" = "darkgreen", "Summe
r" = "darkred", "Winter" = "lightblue"))
```




```
datafinal %>%
  group_by(Season) %>%
  summarize(meansoda = mean(SodaSales),
            meantemp = mean(Temperature)) %>%
  ggplot(aes(x = meantemp, y = meansoda, color = Season)) +
  geom_point(size = 5) +
  scale_color_manual(values = c("Fall" = "darkorange", "Spring" = "darkgreen", "Summe
r" = "darkred", "Winter" = "lightblue")) +
  labs(title = "(3) Average Soda Sales And Temperature For Each Season",
       x = "Mean Temperature (°F)",
       y = "Mean Soda Sales ($)")
```




```
ggplot(datafinal, aes(x = Season, y = SodaSales, fill = Season)) +
  geom_bar(stat = "summary", fun = "mean") +
  labs(title = "(4) Average Soda Sales For Each Season",
       y = "Soda Sales ($)")
```




```
datafinal %>%
  group_by(Season) %>%
  summarize(meansoda = mean(SodaSales),
            meantemp = mean(Temperature))
```

```
## # A tibble: 4 × 3
##   Season meansoda meantemp
##   <chr>     <dbl>    <dbl>
## 1 Fall       821.     77.6
## 2 Spring     809.     76.4
## 3 Summer     965.     93.8
## 4 Winter     710.     63.0
```

```
data2 <- datafinal %>%
  mutate(TempIntervals = case_when(
    Temperature > 20 & Temperature <= 30 ~ '21-30',
   Temperature > 30 & Temperature <= 40 ~ '31-40',
    Temperature > 40 & Temperature <= 50 ~ '41-50',
   Temperature > 50 & Temperature <= 60 ~ '51-60',
   Temperature > 60 & Temperature <= 70 ~ '61-70',
   Temperature > 70 & Temperature <= 80 ~ '71-80',
   Temperature > 80 & Temperature <= 90 ~ '81-90',
   Temperature > 90 & Temperature <= 100 ~ '91-100',
   Temperature > 100 & Temperature <= 110 ~ '101-110'))
```

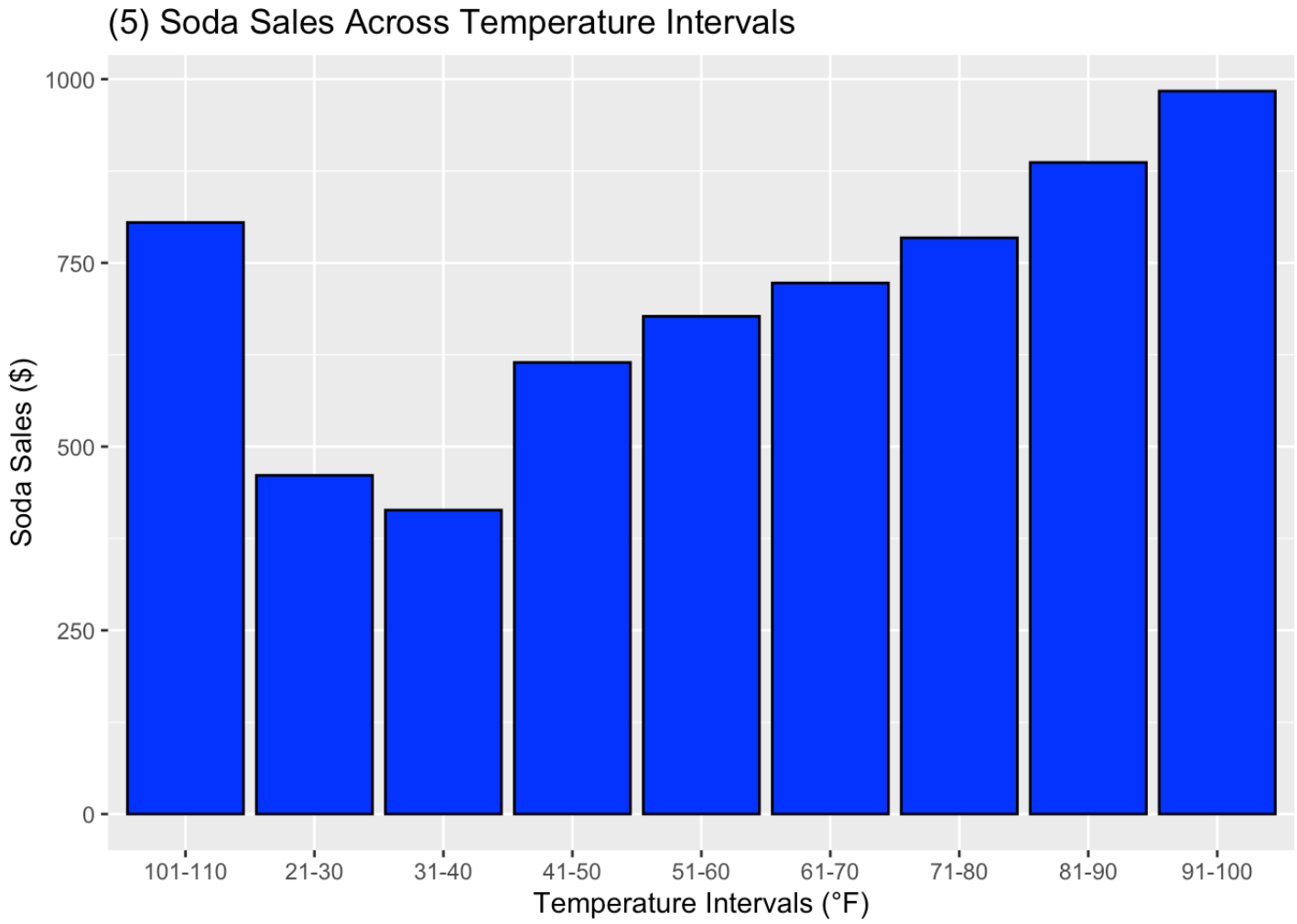
Plot (1) shows a linear, positive relationship between soda sales and temperature as was hypothesized.

Plot (2) is just filtering Plot (1) based on the season each observation was taken in. It shows that summer days (marked in dark red) tend to congregate in the top right corner of high temperatures and greater number of sales in $.

Plot (3) averages season sales and shows a clear distinction between the summer and winter months based on the dollar value of soda sales.

Plot (4) draws upon more of the same conclusions with the hotter summer months having, on average, the highest temperature and soda sales, with ~94°F and ~$965 in sales, respectively.

```
ggplot(data2, aes(x = TempIntervals, y = SodaSales)) +
    labs(title = "(5) Soda Sales Across Temperature Intervals", y = "Soda Sales ($)",
x = "Temperature Intervals (°F)") +
  geom_bar(stat = "summary", fun = "mean", fill = "blue", color = "black")
```



If we were to divide the observations into different temperature intervals, we would see that revenue earned from soda sales almost doubles from the lowest and highest intervals. But not only that, the visualization also shows that higher temperatures, in general, lead to greater soda sales in $.

(I don't know why the highest temperature interval is showing up on the far left of the graph; maybe a glitch in the code.)

Additional Analysis and Business Applications

A regression was also ran on Stata with soda sales as the dependent variable, temperature as the independent variable, and additional variables as controls (coffee sales, gas sales, season and day of the week).

The results are shown below:

```
. regress sodasales temperature coffeesales gassales spring summer fall monday tuesday wednesday thursday friday
> saturday
```

Source	SS	df	MS		
Model	14487994.1	12	1207332.84	Number of obs	= 391
Residual	5037283.86	378	13326.1478	F(12, 378)	= 90.60
				Prob > F	= 0.0000
Total	19525278	390	50064.8153	R-squared	= 0.7420
				Adj R-squared	= 0.7338
				Root MSE	= 115.44

sodasales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temperature	5.786086	.6086097	9.51	0.000	4.589401	6.98277
coffeesales	.8241876	.4004764	2.06	0.040	.036747	1.611628
gassales	.0395294	.0043053	9.18	0.000	.0310641	.0479946
spring	-12.72041	19.15643	-0.66	0.507	-50.38692	24.9461
summer	-.3835215	24.93265	-0.02	0.988	-49.40759	48.64055
fall	-29.18291	19.98322	-1.46	0.145	-68.47512	10.1093
monday	191.8056	24.72679	7.76	0.000	143.1863	240.4249
tuesday	250.6548	25.20108	9.95	0.000	201.1029	300.2066
wednesday	250.2482	26.51855	9.44	0.000	198.1059	302.3906
thursday	275.1577	26.68939	10.31	0.000	222.6794	327.636
friday	284.7257	28.33763	10.05	0.000	229.0066	340.4449
saturday	209.5835	23.58523	8.89	0.000	163.2089	255.9582
_cons	-157.4156	45.67726	-3.45	0.001	-247.229	-67.60224

For every 1°F increase, soda sales increase by about $5.79; therefore if, hypothetically, temperature increased by 20°F from one day to another, soda sales revenue would increase by around 115.80 dollars.

An increase in coffee and gas sales also lead to increases in soda sales, however, they are relatively minimal.

Friday seems to be the best day for soda sales, generating $284.73 in sales more than on Sunday (the worst day), on average.

The season categories were deemed insignificant, therefore no conclusion can be drawn from that.

```
datafinal %>%
  group_by(DayOfWeek) %>%
  summarize(AverageSodaSales = mean(SodaSales)) %>%
  arrange(desc(AverageSodaSales))
```

```
## # A tibble: 7 × 2
##   DayOfWeek AverageSodaSales
##   <chr>                <dbl>
## 1 Friday                973.
## 2 Thursday              944.
## 3 Wednesday             919.
## 4 Tuesday               889.
## 5 Monday                828.
## 6 Saturday              803.
## 7 Sunday                515.
```

Since Texas weather is so volatile, a change of 20°F between consecutive days in not unfeasible. And since average soda sales (depending on the day) varies from $514.89 - $972.51, an increase of 115.80 dollars would be a substantial increase in revenue for the convenience store.

In my opinion, increasing soda prices by $0.05 for every 10 degree increase in temperature would further increase store revenue from a practice that most customers would fail to acknowledge as they would continue to purchase the soda. Even for regular and loyal customers, a very small change in prices would not influence them to go to another store to buy the soda; some may not even notice the price change.

For high temperatures days, it would be even more beneficial to increase prices on days such as Friday as soda sales seem to peak on those calender days.

Acknowledgments: Modisoft for the data entries, weather.com for temperature readings in Spring Branch, TX and the manager of Gas Station A for granting permission of using the data acquired