Assignment STAT702

Product name: BIC Round Stic Xtra Life Ballpoint Pen, Medium Point (1.0mm), Red, 12-Count Sales sku_id: 219884 Reviews asin: B00006IE7J

#1 Analysis of Sales Data

1(a) For the product (sku_id) which has been assigned to your group (see page 6), compute the total monthly sales from January 2011 − September 2013. Present your results in an appropriate plot and write 2 − 3 sentences describing your results.

Hint: This will require some "wrangling" of the variable week. To do this, format week as a date and then use the appropriate lubridate function to extract the month.

Marking Criteria

- Total monthly sales have been correctly computed and are displayed in an appropriate plot.

- Description of results/plot is correct and provides useful insights.

- Plot is constructed using ggplot2 and has appropriate titles, labels, scales etc.

```
# Load libraries
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

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

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

```
# Read in data and convert to tibbles
reviews_data <- read.csv("reviews_data.csv")
reviews_data <- as_tibble(reviews_data)

sales_data <- read.csv("sales_data.csv")
sales_data <- as_tibble(sales_data)
```

```
# Create a summary table, grouped by month and year (single column with m/y)

sales_data %>%
        filter(sku_id == 219844) %>%
        mutate(week = as_date(week, format = "%d/%m/%y"),
               date = format(week, "%m/%y"),
               date = my(date)) %>%
        group_by(date) %>%
        summarise(total_units_sold = sum(units_sold))-> sales_summary
```

```
# Scatterplot of monthly sales

ggplot(sales_summary) +
        geom_point(aes(x = date, y = total_units_sold)) +
        xlab("Date") +
        ylab("Units Sold") +
        ggtitle("Monthly sales of product (sku_id 219844)")
```


Monthly sales of product (sku_id 219844)
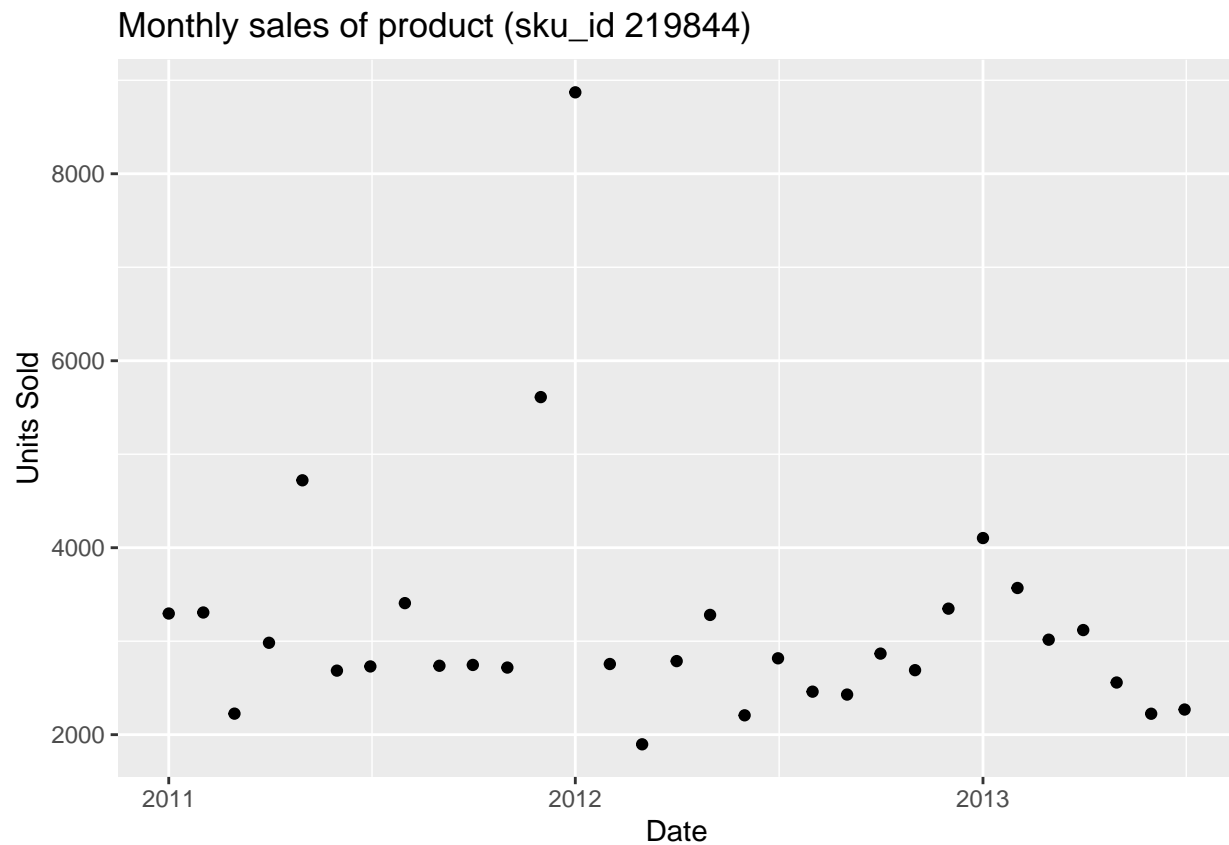
```
# # Print the summary, displaying month and year in separate columns
#
# sales_summary %>%
#         mutate(month = month(date, label = TRUE), year = (year(date))) %>%
#         select(month, year, total_units_sold) %>% view()
```

```
# Compute basic stats
summary(sales_summary$total_units_sold)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1897    2622    2787    3175    3302    8871
```

Mean monthly sales are 3175. Outlier of 8871 sales in Jan 2012, minimum observation is 1897 in March 2012. 50% of observations lie between the values 2622 (first quartile) and 3302 (third quartile). No seasonal variation or trend identified.

1(b) The GM Sales wants to know which stores are performing well and which are not, in terms of product sales. For the product (sku_id) which has been assigned to your group, use appropriate summary statistics and plots to investigate sales performance across the stores and write 2 − 3 paragraphs summarising your findings.

Hint: You will need to decide what it means for a store to be "performing well" and how you will evaluate this using the data.

Marking criteria

- Sales performance is clearly defined.

- Written summary includes relevant and appropriate summary statistics and plots.

- Plot/s are constructed using ggplot2 and have appropriate titles, labels, scales etc.

- Descriptions of results and plots are correct and provides useful insights.