E-Commerce Data Analysis

You have three data sets: central, western and eastern region. You may need to edit the code below for the path to the files.

library(tidyr)  
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.4.3

## Warning: package 'readr' was built under R version 4.4.3

## Warning: package 'dplyr' was built under R version 4.4.2

## Warning: package 'forcats' was built under R version 4.4.2

## Warning: package 'lubridate' was built under R version 4.4.2

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ purrr 1.0.2  
## ✔ forcats 1.0.0 ✔ readr 2.1.5  
## ✔ ggplot2 3.5.1 ✔ stringr 1.5.1  
## ✔ lubridate 1.9.3 ✔ tibble 3.2.1  
## ── 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 conflicts to become errors

#load files  
centralData <- read.csv("C:/Users/Public/pranitha/Orders\_Central.csv", header=TRUE)  
westData <- read.csv("C:/Users/Public/pranitha/Orders\_West.csv", header=TRUE)  
eastData <- read.delim("C:/Users/Public/pranitha/Orders\_East.txt", header=TRUE, sep='\t')

For this assignment, answer the following questions, but please note: the data may or may not be setup such that you can just answer. You may need to perform some “data wrangling” before you can answer. Include the code you use for any wrangling/EDA in your submission. Also include the code you use to answer each question along with text indicating what the answer is. You must provide proof of your answer by showing your R code. You may use online resources and talk to others, but what you submit must be your own work. You should use R to accomplish this work - not other tools.

Understanding the datasets

1. Central Dataset

head(centralData)

## Row.ID Order.ID Ship.Mode Customer.ID Customer.Name Segment  
## 1 15 US-2016-118983 Standard Class HP-14815 Harold Pawlan Home Office  
## 2 16 US-2016-118983 Standard Class HP-14815 Harold Pawlan Home Office  
## 3 17 CA-2015-105893 Standard Class PK-19075 Pete Kriz Consumer  
## 4 22 CA-2017-137330 Standard Class KB-16585 Ken Black Corporate  
## 5 23 CA-2017-137330 Standard Class KB-16585 Ken Black Corporate  
## 6 35 CA-2018-107727 Second Class MA-17560 Matt Abelman Home Office  
## Country City State Postal.Code Product.ID  
## 1 United States Fort Worth Texas 76106 OFF-AP-10002311  
## 2 United States Fort Worth Texas 76106 OFF-BI-10000756  
## 3 United States Madison Wisconsin 53711 OFF-ST-10004186  
## 4 United States Fremont Nebraska 68025 OFF-AR-10000246  
## 5 United States Fremont Nebraska 68025 OFF-AP-10001492  
## 6 United States Houston Texas 77095 OFF-PA-10000249  
## Category Sub.Category  
## 1 Office Supplies Appliances  
## 2 Office Supplies Binders  
## 3 Office Supplies Storage  
## 4 Office Supplies Art  
## 5 Office Supplies Appliances  
## 6 Office Supplies Paper  
## Product  
## 1 Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter  
## 2 Storex DuraTech Recycled Plastic Frosted Binders  
## 3 Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D  
## 4 Newell 318  
## 5 Acco Six-Outlet Power Strip, 4' Cord Length  
## 6 Easy-staple paper  
## Sales Quantity Profit Order.Year Order.Month Order.Day Ship.Year  
## 1 68.810 5 -123.8580 2016 11 22 2016  
## 2 2.544 3 -3.8160 2016 11 22 2016  
## 3 665.880 6 13.3176 2015 11 11 2015  
## 4 19.460 7 5.0596 2017 12 9 2017  
## 5 60.340 7 15.6884 2017 12 9 2017  
## 6 29.472 3 9.9468 2018 10 19 2018  
## Ship.Month Ship.Day Discounts  
## 1 11 26 0.8  
## 2 11 26 0.8  
## 3 11 18 None  
## 4 12 13 None  
## 5 12 13 None  
## 6 10 23 0.2

tail(centralData)

## Row.ID Order.ID Ship.Mode Customer.ID Customer.Name  
## 2318 9952 CA-2018-121559 Second Class HW-14935 Helen Wasserman  
## 2319 9962 CA-2016-168088 First Class CM-12655 Corinna Mitchell  
## 2320 9963 CA-2016-168088 First Class CM-12655 Corinna Mitchell  
## 2321 9973 CA-2017-130225 Standard Class RC-19960 Ryan Crowe  
## 2322 9983 US-2017-157728 Standard Class RC-19960 Ryan Crowe  
## 2323 9984 US-2017-157728 Standard Class RC-19960 Ryan Crowe  
## Segment Country City State Postal.Code  
## 2318 Corporate United States Indianapolis Indiana 46203  
## 2319 Home Office United States Houston Texas 77041  
## 2320 Home Office United States Houston Texas 77041  
## 2321 Consumer United States Houston Texas 77041  
## 2322 Consumer United States Grand Rapids Michigan 49505  
## 2323 Consumer United States Grand Rapids Michigan 49505  
## Product.ID Category Sub.Category  
## 2318 OFF-BI-10002072 Office Supplies Binders  
## 2319 OFF-PA-10000675 Office Supplies Paper  
## 2320 FUR-BO-10004218 Furniture Bookcases  
## 2321 OFF-EN-10000056 Office Supplies Envelopes  
## 2322 OFF-PA-10002195 Office Supplies Paper  
## 2323 TEC-PH-10001305 Technology Phones  
## Product  
## 2318 Cardinal Slant-D Ring Binders  
## 2319 Xerox 1919  
## 2320 Bush Heritage Pine Collection 5-Shelf Bookcase, Albany Pine Finish, \*Special Order  
## 2321 Cameo Buff Policy Envelopes  
## 2322 RSVP Cards & Envelopes, Blank White, 8-1/2" X 11", 24 Cards/25 Envelopes/Set  
## 2323 Panasonic KX TS208W Corded phone  
## Sales Quantity Profit Order.Year Order.Month Order.Day Ship.Year  
## 2318 17.3800 2 8.6900 2018 6 1 2018  
## 2319 65.5840 2 23.7742 2016 3 19 2016  
## 2320 383.4656 4 -67.6704 2016 3 19 2016  
## 2321 99.5680 2 33.6042 2017 9 11 2017  
## 2322 35.5600 7 16.7132 2017 9 22 2017  
## 2323 97.9800 2 27.4344 2017 9 22 2017  
## Ship.Month Ship.Day Discounts  
## 2318 6 3 None  
## 2319 3 22 0.2  
## 2320 3 22 0.32  
## 2321 9 17 0.2  
## 2322 9 28 None  
## 2323 9 28 None

summary(centralData)

## Row.ID Order.ID Ship.Mode Customer.ID   
## Min. : 15 Length:2323 Length:2323 Length:2323   
## 1st Qu.:2564 Class :character Class :character Class :character   
## Median :5029 Mode :character Mode :character Mode :character   
## Mean :5031   
## 3rd Qu.:7612   
## Max. :9984   
## Customer.Name Segment Country City   
## Length:2323 Length:2323 Length:2323 Length:2323   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## State Postal.Code Product.ID Category   
## Length:2323 Min. :46060 Length:2323 Length:2323   
## Class :character 1st Qu.:55125 Class :character Class :character   
## Mode :character Median :65203 Mode :character Mode :character   
## Mean :65341   
## 3rd Qu.:77041   
## Max. :79907   
## Sub.Category Product Sales Quantity   
## Length:2323 Length:2323 Min. : 0.444 Min. : 1.00   
## Class :character Class :character 1st Qu.: 14.620 1st Qu.: 2.00   
## Mode :character Mode :character Median : 45.980 Median : 3.00   
## Mean : 215.773 Mean : 3.78   
## 3rd Qu.: 200.012 3rd Qu.: 5.00   
## Max. :17499.950 Max. :14.00   
## Profit Order.Year Order.Month Order.Day   
## Min. :-3701.893 Min. :2015 Min. : 1.000 Min. : 1.00   
## 1st Qu.: -5.664 1st Qu.:2016 1st Qu.: 5.000 1st Qu.: 8.00   
## Median : 5.184 Median :2017 Median : 8.000 Median :15.00   
## Mean : 17.093 Mean :2017 Mean : 7.693 Mean :15.43   
## 3rd Qu.: 22.456 3rd Qu.:2018 3rd Qu.:11.000 3rd Qu.:23.00   
## Max. : 8399.976 Max. :2018 Max. :12.000 Max. :31.00   
## Ship.Year Ship.Month Ship.Day Discounts   
## Min. :2015 Min. : 1.000 Min. : 1.00 Length:2323   
## 1st Qu.:2016 1st Qu.: 5.000 1st Qu.: 8.00 Class :character   
## Median :2017 Median : 8.000 Median :16.00 Mode :character   
## Mean :2017 Mean : 7.566 Mean :15.77   
## 3rd Qu.:2018 3rd Qu.:11.000 3rd Qu.:24.00   
## Max. :2019 Max. :12.000 Max. :31.00

colSums(is.na(centralData))

## Row.ID Order.ID Ship.Mode Customer.ID Customer.Name   
## 0 0 0 0 0   
## Segment Country City State Postal.Code   
## 0 0 0 0 0   
## Product.ID Category Sub.Category Product Sales   
## 0 0 0 0 0   
## Quantity Profit Order.Year Order.Month Order.Day   
## 0 0 0 0 0   
## Ship.Year Ship.Month Ship.Day Discounts   
## 0 0 0 0

duplicate\_rows <- sum(duplicated(centralData))  
print(paste("Number of duplicate rows:", duplicate\_rows))

## [1] "Number of duplicate rows: 0"

Insights gained for CentralData :

1. From head and tail, i found that there are 24 columns
2. From summary, i found that there are 2323 rows and for numeric columns we can observe minimum value, maximum value, values present in 1st quartile, median, mean and 3rd quartile.
3. Row.ID, Postal.Code, Sales, Quantity, Profit, Order.Year, Order.Month, Order.Day, Ship.Year, Ship.Month, Ship.Day are the numeric columns present in the centralData
4. Order.ID, Ship.Mode, Customer.ID, Customer.Name, Segment, Country, City, State, Product.ID, Category, Sub.Category, Product, Discounts are the non-numeric columns in the centralData
5. There are no null values present in the data
6. There are no duplicate rows present in the dataset
7. Discount column has None value which means no discount is given. so replacing it with 0

For better understanding I will make the following below changes:

1. In Discount Column, None value is replaced with 0
2. Discount column is represented as Character so for analysis I decided to convert it to numeric
3. Adding a new column called Region which has value “Central”
4. Adding two new columns Order.Date and Ship.Date based on Order.Year, Order.Month, Order.Day and Ship.Year, Ship.Month, Ship.Day respectively
5. Deleting the Order.Year, Order.Month, Order.Day, Ship.Year, Ship.Month, Ship.Day
6. Rearrange the columns in specific sequence: Row.ID, Order.ID, Customer.ID, Customer.Name, Product.ID, Product, Category, Sub.Category, Segment, Ship.Mode, Sales, Quantity, Profit, Discounts, Order.Date, Ship.Date, City, State, Country, Postal.Code
7. Converting Order.Date and Ship.Date to Date Format if they are not in Date Format
8. Rename Column Product to Product.Name and Discounts to Discount
9. Formatting Sales and Profit data by rounding off to two decimals and again converting them to numeric

# None value in discount is replaced with 0  
centralData <- centralData %>% mutate(Discounts = gsub("None", "0", Discounts), Discounts = as.numeric(Discounts))   
  
# Discount column is typecast to numeric   
centralData <- centralData %>% mutate(Discounts = as.numeric(Discounts))  
  
# Adding new column called Region which have value 'Central" in it which represents the region of customers  
centralData <- centralData %>% mutate(Region = "Central")  
  
# Adding New columns called Order.Date and Ship.Dates  
centralData <- centralData %>% mutate(Order.Date = sprintf("%02d-%02d-%d", Order.Month, Order.Day, Order.Year))  
centralData <- centralData %>% mutate(Ship.Date = sprintf("%02d-%02d-%d", Ship.Month, Ship.Day, Ship.Year))  
  
# Deleting the Order.Year, Order.Month, Order.Day, Ship.Year, Ship.Month, Ship.Day columns as we   
centralData <- centralData %>% select(-Order.Year, -Order.Month, -Order.Day, -Ship.Year, -Ship.Month, -Ship.Day)  
  
# Rearranging the columns in specific order  
centralData <- centralData %>% select(Row.ID, Order.ID, Customer.ID, Customer.Name, Product.ID, Product, Category, Sub.Category, Segment, Ship.Mode, Sales, Quantity, Profit,   
 Discounts, Order.Date, Ship.Date, City, State, Country, Postal.Code, Region)  
  
# checking the datatypes of all columns  
str(centralData) # Order.Date and Ship.Date are still in character datatype as decided to typecaste to Date datatype

## 'data.frame': 2323 obs. of 21 variables:  
## $ Row.ID : int 15 16 17 22 23 35 36 37 38 39 ...  
## $ Order.ID : chr "US-2016-118983" "US-2016-118983" "CA-2015-105893" "CA-2017-137330" ...  
## $ Customer.ID : chr "HP-14815" "HP-14815" "PK-19075" "KB-16585" ...  
## $ Customer.Name: chr "Harold Pawlan" "Harold Pawlan" "Pete Kriz" "Ken Black" ...  
## $ Product.ID : chr "OFF-AP-10002311" "OFF-BI-10000756" "OFF-ST-10004186" "OFF-AR-10000246" ...  
## $ Product : chr "Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter" "Storex DuraTech Recycled Plastic Frosted Binders" "Stur-D-Stor Shelving, Vertical 5-Shelf: 72\"H x 36\"W x 18 1/2\"D" "Newell 318" ...  
## $ Category : chr "Office Supplies" "Office Supplies" "Office Supplies" "Office Supplies" ...  
## $ Sub.Category : chr "Appliances" "Binders" "Storage" "Art" ...  
## $ Segment : chr "Home Office" "Home Office" "Consumer" "Corporate" ...  
## $ Ship.Mode : chr "Standard Class" "Standard Class" "Standard Class" "Standard Class" ...  
## $ Sales : num 68.81 2.54 665.88 19.46 60.34 ...  
## $ Quantity : int 5 3 6 7 7 3 7 5 9 3 ...  
## $ Profit : num -123.86 -3.82 13.32 5.06 15.69 ...  
## $ Discounts : num 0.8 0.8 0 0 0 0.2 0.2 0.6 0.2 0.32 ...  
## $ Order.Date : chr "11-22-2016" "11-22-2016" "11-11-2015" "12-09-2017" ...  
## $ Ship.Date : chr "11-26-2016" "11-26-2016" "11-18-2015" "12-13-2017" ...  
## $ City : chr "Fort Worth" "Fort Worth" "Madison" "Fremont" ...  
## $ State : chr "Texas" "Texas" "Wisconsin" "Nebraska" ...  
## $ Country : chr "United States" "United States" "United States" "United States" ...  
## $ Postal.Code : int 76106 76106 53711 68025 68025 77095 75080 75080 77041 77041 ...  
## $ Region : chr "Central" "Central" "Central" "Central" ...

# Type-casting Order.Date and Ship.Date to Date format  
centralData <- centralData %>% mutate(Order.Date = as.Date(Order.Date, format = "%m-%d-%Y"), Ship.Date = as.Date(Ship.Date, format = "%m-%d-%Y"))  
  
# Renaming the column PRoduct to Product.Name and Discounts to Discount  
centralData <- centralData %>% rename(Product.Name = Product)  
centralData <- centralData %>% rename(Discount = Discounts)  
  
# Formatting Sales and Profit data and converting back to numeric  
centralData$Sales <- sprintf("%.2f", centralData$Sales)  
centralData$Profit <- sprintf("%.2f", centralData$Profit)  
centralData$Sales <- as.numeric(centralData$Sales)  
centralData$Profit <- as.numeric(centralData$Profit)  
  
head(centralData,15) # Observed that all values have converted as expected

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 15 US-2016-118983 HP-14815 Harold Pawlan OFF-AP-10002311  
## 2 16 US-2016-118983 HP-14815 Harold Pawlan OFF-BI-10000756  
## 3 17 CA-2015-105893 PK-19075 Pete Kriz OFF-ST-10004186  
## 4 22 CA-2017-137330 KB-16585 Ken Black OFF-AR-10000246  
## 5 23 CA-2017-137330 KB-16585 Ken Black OFF-AP-10001492  
## 6 35 CA-2018-107727 MA-17560 Matt Abelman OFF-PA-10000249  
## 7 36 CA-2017-117590 GH-14485 Gene Hale TEC-PH-10004977  
## 8 37 CA-2017-117590 GH-14485 Gene Hale FUR-FU-10003664  
## 9 38 CA-2016-117415 SN-20710 Steve Nguyen OFF-EN-10002986  
## 10 39 CA-2016-117415 SN-20710 Steve Nguyen FUR-BO-10002545  
## 11 40 CA-2016-117415 SN-20710 Steve Nguyen FUR-CH-10004218  
## 12 41 CA-2016-117415 SN-20710 Steve Nguyen TEC-PH-10000486  
## 13 42 CA-2018-120999 LC-16930 Linda Cazamias TEC-PH-10004093  
## 14 45 CA-2017-118255 ON-18715 Odella Nelson TEC-AC-10000171  
## 15 46 CA-2017-118255 ON-18715 Odella Nelson OFF-BI-10003291  
## Product.Name  
## 1 Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter  
## 2 Storex DuraTech Recycled Plastic Frosted Binders  
## 3 Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D  
## 4 Newell 318  
## 5 Acco Six-Outlet Power Strip, 4' Cord Length  
## 6 Easy-staple paper  
## 7 GE 30524EE4  
## 8 Electrix Architect's Clamp-On Swing Arm Lamp, Black  
## 9 #10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes  
## 10 Atlantic Metals Mobile 3-Shelf Bookcases, Custom Colors  
## 11 Global Fabric Manager's Chair, Dark Gray  
## 12 Plantronics HL10 Handset Lifter  
## 13 Panasonic Kx-TS550  
## 14 Verbatim 25 GB 6x Blu-ray Single Layer Recordable Disc, 25/Pack  
## 15 Wilson Jones Leather-Like Binders with DublLock Round Rings  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 1 Office Supplies Appliances Home Office Standard Class 68.81 5  
## 2 Office Supplies Binders Home Office Standard Class 2.54 3  
## 3 Office Supplies Storage Consumer Standard Class 665.88 6  
## 4 Office Supplies Art Corporate Standard Class 19.46 7  
## 5 Office Supplies Appliances Corporate Standard Class 60.34 7  
## 6 Office Supplies Paper Home Office Second Class 29.47 3  
## 7 Technology Phones Corporate First Class 1097.54 7  
## 8 Furniture Furnishings Corporate First Class 190.92 5  
## 9 Office Supplies Envelopes Home Office Standard Class 113.33 9  
## 10 Furniture Bookcases Home Office Standard Class 532.40 3  
## 11 Furniture Chairs Home Office Standard Class 212.06 3  
## 12 Technology Phones Home Office Standard Class 371.17 4  
## 13 Technology Phones Corporate Standard Class 147.17 4  
## 14 Technology Accessories Corporate First Class 45.98 2  
## 15 Office Supplies Binders Corporate First Class 17.46 2  
## Profit Discount Order.Date Ship.Date City State Country  
## 1 -123.86 0.80 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 2 -3.82 0.80 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 3 13.32 0.00 2015-11-11 2015-11-18 Madison Wisconsin United States  
## 4 5.06 0.00 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 5 15.69 0.00 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 6 9.95 0.20 2018-10-19 2018-10-23 Houston Texas United States  
## 7 123.47 0.20 2017-12-08 2017-12-10 Richardson Texas United States  
## 8 -147.96 0.60 2017-12-08 2017-12-10 Richardson Texas United States  
## 9 35.41 0.20 2016-12-27 2016-12-31 Houston Texas United States  
## 10 -46.98 0.32 2016-12-27 2016-12-31 Houston Texas United States  
## 11 -15.15 0.30 2016-12-27 2016-12-31 Houston Texas United States  
## 12 41.76 0.20 2016-12-27 2016-12-31 Houston Texas United States  
## 13 16.56 0.20 2018-09-10 2018-09-15 Naperville Illinois United States  
## 14 19.77 0.00 2017-03-11 2017-03-13 Eagan Minnesota United States  
## 15 8.21 0.00 2017-03-11 2017-03-13 Eagan Minnesota United States  
## Postal.Code Region  
## 1 76106 Central  
## 2 76106 Central  
## 3 53711 Central  
## 4 68025 Central  
## 5 68025 Central  
## 6 77095 Central  
## 7 75080 Central  
## 8 75080 Central  
## 9 77041 Central  
## 10 77041 Central  
## 11 77041 Central  
## 12 77041 Central  
## 13 60540 Central  
## 14 55122 Central  
## 15 55122 Central

colSums(is.na(centralData)) # there are no null values

## Row.ID Order.ID Customer.ID Customer.Name Product.ID   
## 0 0 0 0 0   
## Product.Name Category Sub.Category Segment Ship.Mode   
## 0 0 0 0 0   
## Sales Quantity Profit Discount Order.Date   
## 0 0 0 0 0   
## Ship.Date City State Country Postal.Code   
## 0 0 0 0 0   
## Region   
## 0

Everything is fine with the CentralDataset so proceeding with WestDataset

1. West Dataset

head(westData)

## Row.ID Order.ID Order.Date Ship.Date Ship.Mode  
## 1 3 CA-2017-138688 12-06-2017 0.00 15-06-2017 0.00 Second Class  
## 2 6 CA-2015-115812 09-06-2015 0.00 13-06-2015 0.00 Standard Class  
## 3 7 CA-2015-115812 09-06-2015 0.00 13-06-2015 0.00 Standard Class  
## 4 8 CA-2015-115812 09-06-2015 0.00 13-06-2015 0.00 Standard Class  
## 5 9 CA-2015-115812 09-06-2015 0.00 13-06-2015 0.00 Standard Class  
## 6 10 CA-2015-115812 09-06-2015 0.00 13-06-2015 0.00 Standard Class  
## Customer.ID Customer.Name Segment Country City Postal.Code  
## 1 DV-13045 Darrin Van Huff Corporate United States Los Angeles 90036  
## 2 BH-11710 Brosina Hoffman Consumer United States Los Angeles 90032  
## 3 BH-11710 Brosina Hoffman Consumer United States Los Angeles 90032  
## 4 BH-11710 Brosina Hoffman Consumer United States Los Angeles 90032  
## 5 BH-11710 Brosina Hoffman Consumer United States Los Angeles 90032  
## 6 BH-11710 Brosina Hoffman Consumer United States Los Angeles 90032  
## Region Product.ID Category Sub.Category  
## 1 West OFF-LA-10000240 Office Supplies Labels  
## 2 West FUR-FU-10001487 Furniture Furnishings  
## 3 West OFF-AR-10002833 Office Supplies Art  
## 4 West TEC-PH-10002275 Technology Phones  
## 5 West OFF-BI-10003910 Office Supplies Binders  
## 6 West OFF-AP-10002892 Office Supplies Appliances  
## Product.Name Sales  
## 1 Self-Adhesive Address Labels for Typewriters by Universal 14.62  
## 2 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## 3 Newell 322 7.28  
## 4 Mitel 5320 IP Phone VoIP phone 9.07.152  
## 5 DXL Angle-View Binders with Locking Rings by Samsill 18.504  
## 6 Belkin F5C206VTEL 6 Outlet Surge 114.9  
## Quantity Discount Profit Right\_Row.ID Right\_Order.Date Right\_Ship.Date  
## 1 2 0.0 68.714 3 06-12-2017 0.00 6/16/2017 0:00  
## 2 7 0.0 1.41.694 6 06-09-2015 0.00 6/14/2015 0:00  
## 3 4 0.0 19.656 6 06-09-2015 0.00 6/14/2015 0:00  
## 4 6 0.2 9.07.152 6 06-09-2015 0.00 6/14/2015 0:00  
## 5 3 0.2 57.825 6 06-09-2015 0.00 6/14/2015 0:00  
## 6 5 0.0 34.47 6 06-09-2015 0.00 6/14/2015 0:00  
## Right\_Ship.Mode Right\_Customer.ID Right\_Customer.Name Right\_Segment  
## 1 Second Class DV-13045 Darrin Van Huff Corporate  
## 2 Standard Class BH-11710 Brosina Hoffman Consumer  
## 3 Standard Class BH-11710 Brosina Hoffman Consumer  
## 4 Standard Class BH-11710 Brosina Hoffman Consumer  
## 5 Standard Class BH-11710 Brosina Hoffman Consumer  
## 6 Standard Class BH-11710 Brosina Hoffman Consumer  
## Right\_Country Right\_City Right\_State2 Right\_Postal.Code Right\_Region  
## 1 United States Los Angeles California 90036 West  
## 2 United States Los Angeles California 90032 West  
## 3 United States Los Angeles California 90032 West  
## 4 United States Los Angeles California 90032 West  
## 5 United States Los Angeles California 90032 West  
## 6 United States Los Angeles California 90032 West  
## Right\_Product.ID Right\_Category Right\_Sub.Category  
## 1 OFF-LA-10000240 Office Supplies Labels  
## 2 FUR-FU-10001487 Furniture Furnishings  
## 3 FUR-FU-10001487 Furniture Furnishings  
## 4 FUR-FU-10001487 Furniture Furnishings  
## 5 FUR-FU-10001487 Furniture Furnishings  
## 6 FUR-FU-10001487 Furniture Furnishings  
## Right\_Product.Name Right\_Sales  
## 1 Self-Adhesive Address Labels for Typewriters by Universal 14.62  
## 2 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## 3 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## 4 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## 5 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.86  
## Right\_Quantity Right\_Discount Right\_Profit State  
## 1 2 0 68.714 CA  
## 2 7 0 1.41.694 CA  
## 3 7 0 1.41.694 CA  
## 4 7 0 1.41.694 CA  
## 5 7 0 1.41.694 CA  
## 6 7 0 1.41.694 CA

tail(westData)

## Row.ID Order.ID Order.Date Ship.Date Ship.Mode  
## 9548 9992 CA-2018-121258 26-02-2018 0.00 02-03-2018 0.00 Standard Class  
## 9549 9992 CA-2018-121258 26-02-2018 0.00 02-03-2018 0.00 Standard Class  
## 9550 9993 CA-2018-121258 26-02-2018 0.00 02-03-2018 0.00 Standard Class  
## 9551 9993 CA-2018-121258 26-02-2018 0.00 02-03-2018 0.00 Standard Class  
## 9552 9993 CA-2018-121258 26-02-2018 0.00 02-03-2018 0.00 Standard Class  
## 9553 9994 CA-2018-119914 04-05-2018 0.00 08-05-2018 0.00 Second Class  
## Customer.ID Customer.Name Segment Country City Postal.Code  
## 9548 DB-13060 Dave Brooks Consumer United States Costa Mesa 92627  
## 9549 DB-13060 Dave Brooks Consumer United States Costa Mesa 92627  
## 9550 DB-13060 Dave Brooks Consumer United States Costa Mesa 92627  
## 9551 DB-13060 Dave Brooks Consumer United States Costa Mesa 92627  
## 9552 DB-13060 Dave Brooks Consumer United States Costa Mesa 92627  
## 9553 CC-12220 Chris Cortes Consumer United States Westminster 92683  
## Region Product.ID Category Sub.Category  
## 9548 West TEC-PH-10003645 Technology Phones  
## 9549 West TEC-PH-10003645 Technology Phones  
## 9550 West OFF-PA-10004041 Office Supplies Paper  
## 9551 West OFF-PA-10004041 Office Supplies Paper  
## 9552 West OFF-PA-10004041 Office Supplies Paper  
## 9553 West OFF-AP-10002684 Office Supplies Appliances  
## Product.Name  
## 9548 Aastra 57i VoIP phone  
## 9549 Aastra 57i VoIP phone  
## 9550 It's Hot Message Books with Stickers, 2 3/4" x 5"  
## 9551 It's Hot Message Books with Stickers, 2 3/4" x 5"  
## 9552 It's Hot Message Books with Stickers, 2 3/4" x 5"  
## 9553 Acco 7-Outlet Masterpiece Power Center, Wihtout Fax/Phone Line Protection  
## Sales Quantity Discount Profit Right\_Row.ID Right\_Order.Date  
## 9548 2.58.576 2 0.2 1.93.932 9992 2/26/2018 0:00  
## 9549 2.58.576 2 0.2 1.93.932 9993 2/26/2018 0:00  
## 9550 29.6 4 0.0 13.32 9991 2/26/2018 0:00  
## 9551 29.6 4 0.0 13.32 9992 2/26/2018 0:00  
## 9552 29.6 4 0.0 13.32 9993 2/26/2018 0:00  
## 9553 243.16 2 0.0 72.948 9994 05-04-2018 0.00  
## Right\_Ship.Date Right\_Ship.Mode Right\_Customer.ID Right\_Customer.Name  
## 9548 03-03-2018 0.00 Standard Class DB-13060 Dave Brooks  
## 9549 03-03-2018 0.00 Standard Class DB-13060 Dave Brooks  
## 9550 03-03-2018 0.00 Standard Class DB-13060 Dave Brooks  
## 9551 03-03-2018 0.00 Standard Class DB-13060 Dave Brooks  
## 9552 03-03-2018 0.00 Standard Class DB-13060 Dave Brooks  
## 9553 05-09-2018 0.00 Second Class CC-12220 Chris Cortes  
## Right\_Segment Right\_Country Right\_City Right\_State2 Right\_Postal.Code  
## 9548 Consumer United States Costa Mesa California 92627  
## 9549 Consumer United States Costa Mesa California 92627  
## 9550 Consumer United States Costa Mesa California 92627  
## 9551 Consumer United States Costa Mesa California 92627  
## 9552 Consumer United States Costa Mesa California 92627  
## 9553 Consumer United States Westminster California 92683  
## Right\_Region Right\_Product.ID Right\_Category Right\_Sub.Category  
## 9548 West TEC-PH-10003645 Technology Phones  
## 9549 West OFF-PA-10004041 Office Supplies Paper  
## 9550 West FUR-FU-10000747 Furniture Furnishings  
## 9551 West TEC-PH-10003645 Technology Phones  
## 9552 West OFF-PA-10004041 Office Supplies Paper  
## 9553 West OFF-AP-10002684 Office Supplies Appliances  
## Right\_Product.Name  
## 9548 Aastra 57i VoIP phone  
## 9549 It's Hot Message Books with Stickers, 2 3/4" x 5"  
## 9550 Tenex B1-RE Series Chair Mats for Low Pile Carpets  
## 9551 Aastra 57i VoIP phone  
## 9552 It's Hot Message Books with Stickers, 2 3/4" x 5"  
## 9553 Acco 7-Outlet Masterpiece Power Center, Wihtout Fax/Phone Line Protection  
## Right\_Sales Right\_Quantity Right\_Discount Right\_Profit State  
## 9548 2.58.576 2 0.2 1.93.932 CA  
## 9549 29.6 4 0.0 13.32 CA  
## 9550 91.96 2 0.0 1.56.332 CA  
## 9551 2.58.576 2 0.2 1.93.932 CA  
## 9552 29.6 4 0.0 13.32 CA  
## 9553 243.16 2 0.0 72.948 CA

summary(westData)

## Row.ID Order.ID Order.Date Ship.Date   
## Min. : 3 Length:9553 Length:9553 Length:9553   
## 1st Qu.:2485 Class :character Class :character Class :character   
## Median :4827 Mode :character Mode :character Mode :character   
## Mean :4915   
## 3rd Qu.:7292   
## Max. :9994   
## Ship.Mode Customer.ID Customer.Name Segment   
## Length:9553 Length:9553 Length:9553 Length:9553   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Country City Postal.Code Region   
## Length:9553 Length:9553 Min. :59102 Length:9553   
## Class :character Class :character 1st Qu.:90008 Class :character   
## Mode :character Mode :character Median :92105 Mode :character   
## Mean :91579   
## 3rd Qu.:94513   
## Max. :99301   
## Product.ID Category Sub.Category Product.Name   
## Length:9553 Length:9553 Length:9553 Length:9553   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Sales Quantity Discount Profit   
## Length:9553 Min. : 1.000 Min. :0.0000 Length:9553   
## Class :character 1st Qu.: 2.000 1st Qu.:0.0000 Class :character   
## Mode :character Median : 3.000 Median :0.0000 Mode :character   
## Mean : 3.885 Mean :0.1144   
## 3rd Qu.: 5.000 3rd Qu.:0.2000   
## Max. :14.000 Max. :0.7000   
## Right\_Row.ID Right\_Order.Date Right\_Ship.Date Right\_Ship.Mode   
## Min. : 3 Length:9553 Length:9553 Length:9553   
## 1st Qu.:2485 Class :character Class :character Class :character   
## Median :4827 Mode :character Mode :character Mode :character   
## Mean :4915   
## 3rd Qu.:7292   
## Max. :9994   
## Right\_Customer.ID Right\_Customer.Name Right\_Segment Right\_Country   
## Length:9553 Length:9553 Length:9553 Length:9553   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Right\_City Right\_State2 Right\_Postal.Code Right\_Region   
## Length:9553 Length:9553 Min. :59102 Length:9553   
## Class :character Class :character 1st Qu.:90008 Class :character   
## Mode :character Mode :character Median :92105 Mode :character   
## Mean :91579   
## 3rd Qu.:94513   
## Max. :99301   
## Right\_Product.ID Right\_Category Right\_Sub.Category Right\_Product.Name  
## Length:9553 Length:9553 Length:9553 Length:9553   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Right\_Sales Right\_Quantity Right\_Discount Right\_Profit   
## Length:9553 Min. : 1.000 Min. :0.0000 Length:9553   
## Class :character 1st Qu.: 2.000 1st Qu.:0.0000 Class :character   
## Mode :character Median : 3.000 Median :0.0000 Mode :character   
## Mean : 3.885 Mean :0.1144   
## 3rd Qu.: 5.000 3rd Qu.:0.2000   
## Max. :14.000 Max. :0.7000   
## State   
## Length:9553   
## Class :character   
## Mode :character   
##   
##   
##

colSums(is.na(westData))

## Row.ID Order.ID Order.Date Ship.Date   
## 0 0 0 0   
## Ship.Mode Customer.ID Customer.Name Segment   
## 0 0 0 0   
## Country City Postal.Code Region   
## 0 0 0 0   
## Product.ID Category Sub.Category Product.Name   
## 0 0 0 0   
## Sales Quantity Discount Profit   
## 0 0 0 0   
## Right\_Row.ID Right\_Order.Date Right\_Ship.Date Right\_Ship.Mode   
## 0 0 0 0   
## Right\_Customer.ID Right\_Customer.Name Right\_Segment Right\_Country   
## 0 0 0 0   
## Right\_City Right\_State2 Right\_Postal.Code Right\_Region   
## 0 0 0 0   
## Right\_Product.ID Right\_Category Right\_Sub.Category Right\_Product.Name   
## 0 0 0 0   
## Right\_Sales Right\_Quantity Right\_Discount Right\_Profit   
## 0 0 0 0   
## State   
## 0

duplicate\_rows\_westData <- sum(duplicated(westData))  
print(paste("Number of duplicate rows:", duplicate\_rows\_westData))

## [1] "Number of duplicate rows: 0"

Insights gained from West Dataset:

1. There are 9553 rows and there are duplicates for all columns with prefix “Right\_” ColumnName example Row.ID has duplicate column Right\_Row.ID. And they also have same data but in different rows thats the reason i didnt get result for duplicates. So i want to drop columns with prefix Right\_ and check for duplicates.
2. There are no null values.
3. Row.ID, Postal.Code, quantity, Discount, are numeric columns. Order.ID, Order.Date, Ship.Date, Ship.Mode, Customer.ID, Customer.Name, Segment, Country, City, Region, Product\_ID, Category, Sub.Category, Product.Name, Sales, Profit, State (Considering Right\_State2) are non meric columns.

For better understanding I will make the following below changes:

1. Renaming State as Right\_State and Right\_State2 as State
2. Removing columns with prefix Right\_ (duplicate columns)
3. Converting Order.Date and Ship.Date to Date Format from character type
4. typecasting character columns like Sales, Profit to numeric in a format
5. Rearranging the columns

# Renaming the column name  
westData <- westData %>% rename(Right\_State = State, State = Right\_State2)  
  
# Deleting the columns with Prefix Right\_  
westData <- westData %>% select(-Right\_Row.ID, -Right\_Order.Date, -Right\_Ship.Date, -Right\_Ship.Mode, -Right\_Customer.ID, -Right\_Customer.Name, -Right\_Segment, -Right\_Country,  
 -Right\_Country, -Right\_City, -Right\_Postal.Code, -Right\_Region, -Right\_Product.ID, -Right\_Category, -Right\_Sub.Category, -Right\_Product.Name,  
 -Right\_Sales, -Right\_Quantity, -Right\_Discount, -Right\_Profit, -Right\_State)  
  
# Converting Order.Date and Ship.Date to Date Format  
westData <- westData %>% mutate(Order.Date = as.Date(Order.Date, format = "%d-%m-%Y"), Ship.Date = as.Date(Ship.Date, format = "%d-%m-%Y"))  
  
# Rearranging the columns  
westData <- westData %>% select(Row.ID, Order.ID, Customer.ID, Customer.Name, Product.ID, Product.Name, Category, Sub.Category, Segment, Ship.Mode, Sales, Quantity, Profit,   
 Discount, Order.Date, Ship.Date, City, State, Country, Postal.Code, Region)  
  
head(westData)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 3 CA-2017-138688 DV-13045 Darrin Van Huff OFF-LA-10000240  
## 2 6 CA-2015-115812 BH-11710 Brosina Hoffman FUR-FU-10001487  
## 3 7 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AR-10002833  
## 4 8 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002275  
## 5 9 CA-2015-115812 BH-11710 Brosina Hoffman OFF-BI-10003910  
## 6 10 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AP-10002892  
## Product.Name  
## 1 Self-Adhesive Address Labels for Typewriters by Universal  
## 2 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood  
## 3 Newell 322  
## 4 Mitel 5320 IP Phone VoIP phone  
## 5 DXL Angle-View Binders with Locking Rings by Samsill  
## 6 Belkin F5C206VTEL 6 Outlet Surge  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 1 Office Supplies Labels Corporate Second Class 14.62 2  
## 2 Furniture Furnishings Consumer Standard Class 48.86 7  
## 3 Office Supplies Art Consumer Standard Class 7.28 4  
## 4 Technology Phones Consumer Standard Class 9.07.152 6  
## 5 Office Supplies Binders Consumer Standard Class 18.504 3  
## 6 Office Supplies Appliances Consumer Standard Class 114.9 5  
## Profit Discount Order.Date Ship.Date City State Country  
## 1 68.714 0.0 2017-06-12 2017-06-15 Los Angeles California United States  
## 2 1.41.694 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 3 19.656 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 4 9.07.152 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 5 57.825 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 6 34.47 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## Postal.Code Region  
## 1 90036 West  
## 2 90032 West  
## 3 90032 West  
## 4 90032 West  
## 5 90032 West  
## 6 90032 West

# typecasting character columns Sales, Profit to numeric in a format  
format\_values <- function(value) {  
 value <- as.character(value)   
 last\_dot\_pos <- regexpr("\\.[0-9]+$", value)  
   
 if (last\_dot\_pos > 0) {  
 decimal\_part <- substr(value, last\_dot\_pos, nchar(value))  
 integer\_part <- gsub("\\.", "", substr(value, 1, last\_dot\_pos - 1))  
 formatted\_value <- paste0(integer\_part, decimal\_part)  
 } else {  
 formatted\_value <- gsub("\\.", "", value)  
 }  
 numeric\_value <- suppressWarnings(as.numeric(formatted\_value))  
   
 ifelse(is.na(numeric\_value), NA, sprintf("%.2f", round(numeric\_value, 2)))  
}  
  
# Apply function to both Sales and Profit columns  
westData <- westData %>% mutate(Sales = sapply(Sales, format\_values), Profit = sapply(Profit, format\_values))  
  
# Converting back to numeric   
westData$Sales <- as.numeric(westData$Sales)  
westData$Profit <- as.numeric(westData$Profit)  
head(westData)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 3 CA-2017-138688 DV-13045 Darrin Van Huff OFF-LA-10000240  
## 2 6 CA-2015-115812 BH-11710 Brosina Hoffman FUR-FU-10001487  
## 3 7 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AR-10002833  
## 4 8 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002275  
## 5 9 CA-2015-115812 BH-11710 Brosina Hoffman OFF-BI-10003910  
## 6 10 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AP-10002892  
## Product.Name  
## 1 Self-Adhesive Address Labels for Typewriters by Universal  
## 2 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood  
## 3 Newell 322  
## 4 Mitel 5320 IP Phone VoIP phone  
## 5 DXL Angle-View Binders with Locking Rings by Samsill  
## 6 Belkin F5C206VTEL 6 Outlet Surge  
## Category Sub.Category Segment Ship.Mode Sales Quantity Profit  
## 1 Office Supplies Labels Corporate Second Class 14.62 2 68.71  
## 2 Furniture Furnishings Consumer Standard Class 48.86 7 141.69  
## 3 Office Supplies Art Consumer Standard Class 7.28 4 19.66  
## 4 Technology Phones Consumer Standard Class 907.15 6 907.15  
## 5 Office Supplies Binders Consumer Standard Class 18.50 3 57.83  
## 6 Office Supplies Appliances Consumer Standard Class 114.90 5 34.47  
## Discount Order.Date Ship.Date City State Country  
## 1 0.0 2017-06-12 2017-06-15 Los Angeles California United States  
## 2 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 3 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 4 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 5 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 6 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## Postal.Code Region  
## 1 90036 West  
## 2 90032 West  
## 3 90032 West  
## 4 90032 West  
## 5 90032 West  
## 6 90032 West

#checking null values  
colSums(is.na(westData))

## Row.ID Order.ID Customer.ID Customer.Name Product.ID   
## 0 0 0 0 0   
## Product.Name Category Sub.Category Segment Ship.Mode   
## 0 0 0 0 0   
## Sales Quantity Profit Discount Order.Date   
## 0 0 0 0 0   
## Ship.Date City State Country Postal.Code   
## 0 0 0 0 0   
## Region   
## 0

Everything fine and there are no null values so proceeding to next dataset

1. East Dataset

head(eastData)

## Category City Country Customer.ID Customer.Name  
## 1 Furniture Philadelphia United States SF-20065 Sandra Flanagan  
## 2 Office Supplies Philadelphia United States TB-21520 Tracy Blumstein  
## 3 Office Supplies Philadelphia United States TB-21520 Tracy Blumstein  
## 4 Office Supplies Philadelphia United States FH-14365 Fred Hopkins  
## 5 Technology Medina United States VW-21775 Victoria Wilson  
## 6 Office Supplies Dublin United States JB-15925 Joni Blumstein  
## Discount Order.Date Order.ID Postal.Code Product.ID  
## 1 0.3 7/16/2018 0:00 US-2018-156909 19140 FUR-CH-10002774  
## 2 0.7 9/17/2016 0:00 US-2016-150630 19140 OFF-BI-10000474  
## 3 0.7 9/17/2016 0:00 US-2016-150630 19140 OFF-BI-10001525  
## 4 0.7 7/6/2018 0:00 US-2018-124303 19120 OFF-BI-10000343  
## 5 0.7 1/2/2016 0:00 CA-2016-146262 44256 TEC-MA-10000864  
## 6 0.7 12/24/2016 0:00 CA-2016-169397 43017 OFF-BI-10002852  
## Product.Name  
## 1 Global Deluxe Stacking Chair, Gray  
## 2 Avery Recycled Flexi-View Covers for Binding Systems  
## 3 Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11", Executive Red  
## 4 Pressboard Covers with Storage Hooks, 9 1/2" x 11", Light Blue  
## 5 Cisco 9971 IP Video Phone Charcoal  
## 6 Ibico Standard Transparent Covers  
## Profit Quantity Region Row.ID Segment Ship.Date Ship.Mode  
## 1 -1.0196 2 East 24 Consumer 7/18/2018 0:00 Second Class  
## 2 -7.0532 2 East 29 Consumer 9/21/2016 0:00 Standard Class  
## 3 -5.7150 6 East 33 Consumer 9/21/2016 0:00 Standard Class  
## 4 -2.2586 2 East 199 Corporate 7/13/2018 0:00 Standard Class  
## 5 -950.4000 9 East 216 Corporate 1/9/2016 0:00 Standard Class  
## 6 -13.8432 4 East 222 Consumer 12/27/2016 0:00 First Class  
## State Sub.Category Sales  
## 1 Pennsylvania Chairs USD71.37  
## 2 Pennsylvania Binders USD9.61  
## 3 Pennsylvania Binders USD6.85  
## 4 Pennsylvania Binders USD2.94  
## 5 Ohio Machines USD1188  
## 6 Ohio Binders USD24

tail(eastData)

## Category City Country Customer.ID Customer.Name  
## 2841 Office Supplies Newark United States HE-14800 Harold Engle  
## 2842 Office Supplies Plainfield United States RB-19435 Richard Bierner  
## 2843 Office Supplies Plainfield United States RB-19435 Richard Bierner  
## 2844 Office Supplies Plainfield United States RB-19435 Richard Bierner  
## 2845 Office Supplies Long Beach United States DV-13465 Dianna Vittorini  
## 2846 Office Supplies Long Beach United States DV-13465 Dianna Vittorini  
## Discount Order.Date Order.ID Postal.Code Product.ID  
## 2841 0 12/5/2017 0:00 CA-2017-146374 19711 OFF-EN-10004483  
## 2842 0 12/11/2018 0:00 CA-2018-153871 7060 OFF-BI-10004209  
## 2843 0 12/11/2018 0:00 CA-2018-153871 7060 OFF-BI-10004600  
## 2844 0 12/11/2018 0:00 CA-2018-153871 7060 OFF-AP-10003622  
## 2845 0 5/17/2016 0:00 CA-2016-100251 11561 OFF-LA-10003766  
## 2846 0 5/17/2016 0:00 CA-2016-100251 11561 OFF-SU-10000898  
## Product.Name  
## 2841 #10 White Business Envelopes,4 1/8 x 9 1/2  
## 2842 Fellowes Twister Kit, Gray/Clear, 3/pkg  
## 2843 Ibico Ibimaster 300 Manual Binding System  
## 2844 Bravo II Megaboss 12-Amp Hard Body Upright, Replacement Belts, 2 Belts per Pack  
## 2845 Self-Adhesive Removable Labels  
## 2846 Acme Hot Forged Carbon Steel Scissors with Nickel-Plated Handles, 3 7/8" Cut, 8"L  
## Profit Quantity Region Row.ID Segment Ship.Date Ship.Mode  
## 2841 51.5543 7 East 9967 Corporate 12/10/2017 0:00 Second Class  
## 2842 18.0900 5 East 9968 Consumer 12/17/2018 0:00 Standard Class  
## 2843 331.1910 2 East 9969 Consumer 12/17/2018 0:00 Standard Class  
## 2844 6.5975 7 East 9970 Consumer 12/17/2018 0:00 Standard Class  
## 2845 15.1200 10 East 9985 Consumer 5/23/2016 0:00 Standard Class  
## 2846 16.1240 4 East 9986 Consumer 5/23/2016 0:00 Standard Class  
## State Sub.Category Sales  
## 2841 Delaware Envelopes USD109.69  
## 2842 New Jersey Binders USD40.20  
## 2843 New Jersey Binders USD735.98  
## 2844 New Jersey Appliances USD22.75  
## 2845 New York Labels USD31.5  
## 2846 New York Supplies USD55.60

summary(eastData)

## Category City Country Customer.ID   
## Length:2846 Length:2846 Length:2846 Length:2846   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Customer.Name Discount Order.Date Order.ID   
## Length:2846 Min. :0.000 Length:2846 Length:2846   
## Class :character 1st Qu.:0.000 Class :character Class :character   
## Mode :character Median :0.000 Mode :character Mode :character   
## Mean :0.147   
## 3rd Qu.:0.200   
## Max. :0.700   
## NA's :40   
## Postal.Code Product.ID Product.Name Profit   
## Min. : 1040 Length:2846 Length:2846 Min. :-6599.978   
## 1st Qu.:10011 Class :character Class :character 1st Qu.: 1.497   
## Median :11561 Mode :character Mode :character Median : 8.073   
## Mean :17714 Mean : 32.969   
## 3rd Qu.:19143 3rd Qu.: 28.172   
## Max. :45503 Max. : 5039.986   
## NA's :51 NA's :40   
## Quantity Region Row.ID Segment   
## Min. : 1.000 Length:2846 Min. : 24 Length:2846   
## 1st Qu.: 2.000 Class :character 1st Qu.:2296 Class :character   
## Median : 3.000 Mode :character Median :4942 Mode :character   
## Mean : 3.739 Mean :4938   
## 3rd Qu.: 5.000 3rd Qu.:7510   
## Max. :30.000 Max. :9986   
## NA's :40 NA's :40   
## Ship.Date Ship.Mode State Sub.Category   
## Length:2846 Length:2846 Length:2846 Length:2846   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Sales   
## Length:2846   
## Class :character   
## Mode :character   
##   
##   
##   
##

colSums(is.na(eastData))

## Category City Country Customer.ID Customer.Name   
## 0 0 0 0 0   
## Discount Order.Date Order.ID Postal.Code Product.ID   
## 40 0 0 51 0   
## Product.Name Profit Quantity Region Row.ID   
## 0 40 40 0 40   
## Segment Ship.Date Ship.Mode State Sub.Category   
## 0 0 0 0 0   
## Sales   
## 0

duplicate\_rows <- sum(duplicated(eastData))  
print(paste("Number of duplicate rows:", duplicate\_rows))

## [1] "Number of duplicate rows: 39"

Insights gained for EastData :

1. From head and tail, i found that there are 21 columns
2. From summary, i found that there are 2846 rows and for numeric columns we can observe minimum value, maximum value, values present in 1st quartile, median, mean and 3rd quartile.
3. Category, City, Country, Customer.ID, Customer.Name, Order.Date, Order.ID,, Product.ID, Product.Name, Region, Segment, Ship.Date, Ship.Mode, State, Sub-Category, Sales are the 16 non-numeric columns
4. Discount, Postal.Code, Profit, Quantity, Row.ID are the 5 numeric columns
5. There are null values present in the data
6. There are 39 duplicate rows present in the dataset

For better understanding I will make the following below changes:

1. Remove null values and Duplicate rows
2. Convert Sales column to numeric type from character type
3. Converting Profit column to right format
4. Convert Orde.Date and Ship.Date to Date Format
5. Rearranging the columns

# Removing Null value value rows  
eastData <- na.omit(eastData)  
  
# Formatting the Sales and Profit column  
eastData <- eastData %>% mutate(Sales = as.numeric(gsub("USD", "", Sales)))  
eastData <- eastData %>% mutate(Profit = sapply(Profit, format\_values))  
eastData$Profit <- as.numeric(eastData$Profit)  
  
  
# Converting Order.Date and Ship.Date to Date Format  
eastData <- eastData %>% mutate(Order.Date = as.Date(Order.Date, format = "%m/%d/%Y"), Ship.Date = as.Date(Ship.Date, format = "%m/%d/%Y"))  
  
 # Rearranging columns   
eastData <- eastData %>% select(Row.ID, Order.ID, Customer.ID, Customer.Name, Product.ID, Product.Name, Category, Sub.Category, Segment, Ship.Mode, Sales, Quantity, Profit,   
 Discount, Order.Date, Ship.Date, City, State, Country, Postal.Code, Region)  
  
head(eastData)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 24 US-2018-156909 SF-20065 Sandra Flanagan FUR-CH-10002774  
## 2 29 US-2016-150630 TB-21520 Tracy Blumstein OFF-BI-10000474  
## 3 33 US-2016-150630 TB-21520 Tracy Blumstein OFF-BI-10001525  
## 4 199 US-2018-124303 FH-14365 Fred Hopkins OFF-BI-10000343  
## 5 216 CA-2016-146262 VW-21775 Victoria Wilson TEC-MA-10000864  
## 6 222 CA-2016-169397 JB-15925 Joni Blumstein OFF-BI-10002852  
## Product.Name  
## 1 Global Deluxe Stacking Chair, Gray  
## 2 Avery Recycled Flexi-View Covers for Binding Systems  
## 3 Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11", Executive Red  
## 4 Pressboard Covers with Storage Hooks, 9 1/2" x 11", Light Blue  
## 5 Cisco 9971 IP Video Phone Charcoal  
## 6 Ibico Standard Transparent Covers  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 1 Furniture Chairs Consumer Second Class 71.37 2  
## 2 Office Supplies Binders Consumer Standard Class 9.61 2  
## 3 Office Supplies Binders Consumer Standard Class 6.85 6  
## 4 Office Supplies Binders Corporate Standard Class 2.94 2  
## 5 Technology Machines Corporate Standard Class 1188.00 9  
## 6 Office Supplies Binders Consumer First Class 24.00 4  
## Profit Discount Order.Date Ship.Date City State  
## 1 -1.02 0.3 2018-07-16 2018-07-18 Philadelphia Pennsylvania  
## 2 -7.05 0.7 2016-09-17 2016-09-21 Philadelphia Pennsylvania  
## 3 -5.72 0.7 2016-09-17 2016-09-21 Philadelphia Pennsylvania  
## 4 -2.26 0.7 2018-07-06 2018-07-13 Philadelphia Pennsylvania  
## 5 -950.40 0.7 2016-01-02 2016-01-09 Medina Ohio  
## 6 -13.84 0.7 2016-12-24 2016-12-27 Dublin Ohio  
## Country Postal.Code Region  
## 1 United States 19140 East  
## 2 United States 19140 East  
## 3 United States 19140 East  
## 4 United States 19120 East  
## 5 United States 44256 East  
## 6 United States 43017 East

# Again Checking for Null values and Duplicate Rows  
colSums(is.na(eastData))

## Row.ID Order.ID Customer.ID Customer.Name Product.ID   
## 0 0 0 0 0   
## Product.Name Category Sub.Category Segment Ship.Mode   
## 0 0 0 0 0   
## Sales Quantity Profit Discount Order.Date   
## 0 0 0 0 0   
## Ship.Date City State Country Postal.Code   
## 0 0 0 0 0   
## Region   
## 0

duplicate\_rows <- sum(duplicated(eastData))  
print(paste("Number of duplicate rows:", duplicate\_rows))

## [1] "Number of duplicate rows: 0"

There are no duplicates and Null values. Everything is formatted as i decided

Now Combining All the three datasets as a single dataset before that checking for number of rows for each dataset

dim(centralData)

## [1] 2323 21

dim(westData)

## [1] 9553 21

dim(eastData)

## [1] 2795 21

Before combing the data, decided to create a new column called Shipping\_Days for all three datasets

eastData <- eastData %>% mutate(Shipping\_Days = as.numeric(Ship.Date - Order.Date))  
westData <- westData %>% mutate(Shipping\_Days = as.numeric(Ship.Date - Order.Date))  
centralData <- centralData %>% mutate(Shipping\_Days = as.numeric(Ship.Date - Order.Date))  
  
finalData <- rbind(centralData, westData, eastData)  
  
head(finalData)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 15 US-2016-118983 HP-14815 Harold Pawlan OFF-AP-10002311  
## 2 16 US-2016-118983 HP-14815 Harold Pawlan OFF-BI-10000756  
## 3 17 CA-2015-105893 PK-19075 Pete Kriz OFF-ST-10004186  
## 4 22 CA-2017-137330 KB-16585 Ken Black OFF-AR-10000246  
## 5 23 CA-2017-137330 KB-16585 Ken Black OFF-AP-10001492  
## 6 35 CA-2018-107727 MA-17560 Matt Abelman OFF-PA-10000249  
## Product.Name  
## 1 Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter  
## 2 Storex DuraTech Recycled Plastic Frosted Binders  
## 3 Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D  
## 4 Newell 318  
## 5 Acco Six-Outlet Power Strip, 4' Cord Length  
## 6 Easy-staple paper  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 1 Office Supplies Appliances Home Office Standard Class 68.81 5  
## 2 Office Supplies Binders Home Office Standard Class 2.54 3  
## 3 Office Supplies Storage Consumer Standard Class 665.88 6  
## 4 Office Supplies Art Corporate Standard Class 19.46 7  
## 5 Office Supplies Appliances Corporate Standard Class 60.34 7  
## 6 Office Supplies Paper Home Office Second Class 29.47 3  
## Profit Discount Order.Date Ship.Date City State Country  
## 1 -123.86 0.8 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 2 -3.82 0.8 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 3 13.32 0.0 2015-11-11 2015-11-18 Madison Wisconsin United States  
## 4 5.06 0.0 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 5 15.69 0.0 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 6 9.95 0.2 2018-10-19 2018-10-23 Houston Texas United States  
## Postal.Code Region Shipping\_Days  
## 1 76106 Central 4  
## 2 76106 Central 4  
## 3 53711 Central 7  
## 4 68025 Central 4  
## 5 68025 Central 4  
## 6 77095 Central 4

summary(finalData)

## Row.ID Order.ID Customer.ID Customer.Name   
## Min. : 3 Length:14671 Length:14671 Length:14671   
## 1st Qu.:2456 Class :character Class :character Class :character   
## Median :4856 Mode :character Mode :character Mode :character   
## Mean :4935   
## 3rd Qu.:7394   
## Max. :9994   
## Product.ID Product.Name Category Sub.Category   
## Length:14671 Length:14671 Length:14671 Length:14671   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Segment Ship.Mode Sales Quantity   
## Length:14671 Length:14671 Min. : 0.44 Min. : 1.000   
## Class :character Class :character 1st Qu.: 18.53 1st Qu.: 2.000   
## Mode :character Mode :character Median : 57.90 Median : 3.000   
## Mean : 229.95 Mean : 3.838   
## 3rd Qu.: 215.98 3rd Qu.: 5.000   
## Max. :17499.95 Max. :14.000   
## Profit Discount Order.Date   
## Min. :-8.145e+03 Min. :0.0000 Min. :2015-01-03   
## 1st Qu.: 4.000e+00 1st Qu.:0.0000 1st Qu.:2016-05-28   
## Median : 2.700e+01 Median :0.1500 Median :2017-08-02   
## Mean : 3.220e+11 Mean :0.1406 Mean :2017-05-13   
## 3rd Qu.: 1.110e+02 3rd Qu.:0.2000 3rd Qu.:2018-05-28   
## Max. : 7.920e+14 Max. :0.8000 Max. :2018-12-30   
## Ship.Date City State Country   
## Min. :2015-01-07 Length:14671 Length:14671 Length:14671   
## 1st Qu.:2016-05-31 Class :character Class :character Class :character   
## Median :2017-08-04 Mode :character Mode :character Mode :character   
## Mean :2017-05-17   
## 3rd Qu.:2018-06-01   
## Max. :2019-01-09   
## Postal.Code Region Shipping\_Days   
## Min. : 1040 Length:14671 Min. : 0.000   
## 1st Qu.:60610 Class :character 1st Qu.: 2.000   
## Median :90008 Mode :character Median : 3.000   
## Mean :73352 Mean : 3.354   
## 3rd Qu.:94109 3rd Qu.: 5.000   
## Max. :99301 Max. :12.000

Successfully Combined the three datasets as a single dataset named as finalData

Analysing the data

# Checking for Null values in Shipping\_Days  
colSums(is.na(finalData))

## Row.ID Order.ID Customer.ID Customer.Name Product.ID   
## 0 0 0 0 0   
## Product.Name Category Sub.Category Segment Ship.Mode   
## 0 0 0 0 0   
## Sales Quantity Profit Discount Order.Date   
## 0 0 0 0 0   
## Ship.Date City State Country Postal.Code   
## 0 0 0 0 0   
## Region Shipping\_Days   
## 0 0

# Checking duplicate rows  
dup\_rows\_cnt <- sum(duplicated(finalData))  
print(paste("Number of duplicate rows:", dup\_rows\_cnt))

## [1] "Number of duplicate rows: 6350"

Insight gained

1. Found that there are 6350 duplicate rows and no null values

Decided to Do

1. Remove Duplicate values
2. Create new column called Year based on Order Date

# View Duplicate rows  
dup\_rows <- finalData[duplicated(finalData), ]  
head(dup\_rows,20)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 2332 6 CA-2015-115812 BH-11710 Brosina Hoffman FUR-FU-10001487  
## 2333 7 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AR-10002833  
## 2334 8 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002275  
## 2335 9 CA-2015-115812 BH-11710 Brosina Hoffman OFF-BI-10003910  
## 2336 10 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AP-10002892  
## 2337 11 CA-2015-115812 BH-11710 Brosina Hoffman FUR-TA-10001539  
## 2338 12 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002033  
## 2339 6 CA-2015-115812 BH-11710 Brosina Hoffman FUR-FU-10001487  
## 2340 7 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AR-10002833  
## 2341 8 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002275  
## 2342 9 CA-2015-115812 BH-11710 Brosina Hoffman OFF-BI-10003910  
## 2343 10 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AP-10002892  
## 2344 11 CA-2015-115812 BH-11710 Brosina Hoffman FUR-TA-10001539  
## 2345 12 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002033  
## 2346 6 CA-2015-115812 BH-11710 Brosina Hoffman FUR-FU-10001487  
## 2347 7 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AR-10002833  
## 2348 8 CA-2015-115812 BH-11710 Brosina Hoffman TEC-PH-10002275  
## 2349 9 CA-2015-115812 BH-11710 Brosina Hoffman OFF-BI-10003910  
## 2350 10 CA-2015-115812 BH-11710 Brosina Hoffman OFF-AP-10002892  
## 2351 11 CA-2015-115812 BH-11710 Brosina Hoffman FUR-TA-10001539  
## Product.Name  
## 2332 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood  
## 2333 Newell 322  
## 2334 Mitel 5320 IP Phone VoIP phone  
## 2335 DXL Angle-View Binders with Locking Rings by Samsill  
## 2336 Belkin F5C206VTEL 6 Outlet Surge  
## 2337 Chromcraft Rectangular Conference Tables  
## 2338 Konftel 250 Conference\xa0phone\xa0- Charcoal black  
## 2339 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood  
## 2340 Newell 322  
## 2341 Mitel 5320 IP Phone VoIP phone  
## 2342 DXL Angle-View Binders with Locking Rings by Samsill  
## 2343 Belkin F5C206VTEL 6 Outlet Surge  
## 2344 Chromcraft Rectangular Conference Tables  
## 2345 Konftel 250 Conference\xa0phone\xa0- Charcoal black  
## 2346 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood  
## 2347 Newell 322  
## 2348 Mitel 5320 IP Phone VoIP phone  
## 2349 DXL Angle-View Binders with Locking Rings by Samsill  
## 2350 Belkin F5C206VTEL 6 Outlet Surge  
## 2351 Chromcraft Rectangular Conference Tables  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 2332 Furniture Furnishings Consumer Standard Class 48.86 7  
## 2333 Office Supplies Art Consumer Standard Class 7.28 4  
## 2334 Technology Phones Consumer Standard Class 907.15 6  
## 2335 Office Supplies Binders Consumer Standard Class 18.50 3  
## 2336 Office Supplies Appliances Consumer Standard Class 114.90 5  
## 2337 Furniture Tables Consumer Standard Class 1706.18 9  
## 2338 Technology Phones Consumer Standard Class 911.42 4  
## 2339 Furniture Furnishings Consumer Standard Class 48.86 7  
## 2340 Office Supplies Art Consumer Standard Class 7.28 4  
## 2341 Technology Phones Consumer Standard Class 907.15 6  
## 2342 Office Supplies Binders Consumer Standard Class 18.50 3  
## 2343 Office Supplies Appliances Consumer Standard Class 114.90 5  
## 2344 Furniture Tables Consumer Standard Class 1706.18 9  
## 2345 Technology Phones Consumer Standard Class 911.42 4  
## 2346 Furniture Furnishings Consumer Standard Class 48.86 7  
## 2347 Office Supplies Art Consumer Standard Class 7.28 4  
## 2348 Technology Phones Consumer Standard Class 907.15 6  
## 2349 Office Supplies Binders Consumer Standard Class 18.50 3  
## 2350 Office Supplies Appliances Consumer Standard Class 114.90 5  
## 2351 Furniture Tables Consumer Standard Class 1706.18 9  
## Profit Discount Order.Date Ship.Date City State Country  
## 2332 141.69 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2333 19.66 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2334 907.15 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2335 57.83 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2336 34.47 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2337 853.09 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2338 683.57 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2339 141.69 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2340 19.66 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2341 907.15 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2342 57.83 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2343 34.47 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2344 853.09 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2345 683.57 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2346 141.69 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2347 19.66 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2348 907.15 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2349 57.83 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## 2350 34.47 0.0 2015-06-09 2015-06-13 Los Angeles California United States  
## 2351 853.09 0.2 2015-06-09 2015-06-13 Los Angeles California United States  
## Postal.Code Region Shipping\_Days  
## 2332 90032 West 4  
## 2333 90032 West 4  
## 2334 90032 West 4  
## 2335 90032 West 4  
## 2336 90032 West 4  
## 2337 90032 West 4  
## 2338 90032 West 4  
## 2339 90032 West 4  
## 2340 90032 West 4  
## 2341 90032 West 4  
## 2342 90032 West 4  
## 2343 90032 West 4  
## 2344 90032 West 4  
## 2345 90032 West 4  
## 2346 90032 West 4  
## 2347 90032 West 4  
## 2348 90032 West 4  
## 2349 90032 West 4  
## 2350 90032 West 4  
## 2351 90032 West 4

# Remove duplicate rows  
finalData <- finalData[!duplicated(finalData), ]  
  
# Creating a new column called Year from Order.Date  
finalData <- finalData %>% mutate(Year = format(as.Date(Order.Date, format = "%m/%d/%Y"), "%Y"))  
  
# Checking for null values and duplicates  
colSums(is.na(finalData))

## Row.ID Order.ID Customer.ID Customer.Name Product.ID   
## 0 0 0 0 0   
## Product.Name Category Sub.Category Segment Ship.Mode   
## 0 0 0 0 0   
## Sales Quantity Profit Discount Order.Date   
## 0 0 0 0 0   
## Ship.Date City State Country Postal.Code   
## 0 0 0 0 0   
## Region Shipping\_Days Year   
## 0 0 0

dup\_rows\_cnt <- sum(duplicated(finalData))  
print(dup\_rows\_cnt)

## [1] 0

str(finalData)

## 'data.frame': 8321 obs. of 23 variables:  
## $ Row.ID : int 15 16 17 22 23 35 36 37 38 39 ...  
## $ Order.ID : chr "US-2016-118983" "US-2016-118983" "CA-2015-105893" "CA-2017-137330" ...  
## $ Customer.ID : chr "HP-14815" "HP-14815" "PK-19075" "KB-16585" ...  
## $ Customer.Name: chr "Harold Pawlan" "Harold Pawlan" "Pete Kriz" "Ken Black" ...  
## $ Product.ID : chr "OFF-AP-10002311" "OFF-BI-10000756" "OFF-ST-10004186" "OFF-AR-10000246" ...  
## $ Product.Name : chr "Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter" "Storex DuraTech Recycled Plastic Frosted Binders" "Stur-D-Stor Shelving, Vertical 5-Shelf: 72\"H x 36\"W x 18 1/2\"D" "Newell 318" ...  
## $ Category : chr "Office Supplies" "Office Supplies" "Office Supplies" "Office Supplies" ...  
## $ Sub.Category : chr "Appliances" "Binders" "Storage" "Art" ...  
## $ Segment : chr "Home Office" "Home Office" "Consumer" "Corporate" ...  
## $ Ship.Mode : chr "Standard Class" "Standard Class" "Standard Class" "Standard Class" ...  
## $ Sales : num 68.81 2.54 665.88 19.46 60.34 ...  
## $ Quantity : int 5 3 6 7 7 3 7 5 9 3 ...  
## $ Profit : num -123.86 -3.82 13.32 5.06 15.69 ...  
## $ Discount : num 0.8 0.8 0 0 0 0.2 0.2 0.6 0.2 0.32 ...  
## $ Order.Date : Date, format: "2016-11-22" "2016-11-22" ...  
## $ Ship.Date : Date, format: "2016-11-26" "2016-11-26" ...  
## $ City : chr "Fort Worth" "Fort Worth" "Madison" "Fremont" ...  
## $ State : chr "Texas" "Texas" "Wisconsin" "Nebraska" ...  
## $ Country : chr "United States" "United States" "United States" "United States" ...  
## $ Postal.Code : int 76106 76106 53711 68025 68025 77095 75080 75080 77041 77041 ...  
## $ Region : chr "Central" "Central" "Central" "Central" ...  
## $ Shipping\_Days: num 4 4 7 4 4 4 2 2 4 4 ...  
## $ Year : chr "2016" "2016" "2015" "2017" ...

head(finalData,15)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1 15 US-2016-118983 HP-14815 Harold Pawlan OFF-AP-10002311  
## 2 16 US-2016-118983 HP-14815 Harold Pawlan OFF-BI-10000756  
## 3 17 CA-2015-105893 PK-19075 Pete Kriz OFF-ST-10004186  
## 4 22 CA-2017-137330 KB-16585 Ken Black OFF-AR-10000246  
## 5 23 CA-2017-137330 KB-16585 Ken Black OFF-AP-10001492  
## 6 35 CA-2018-107727 MA-17560 Matt Abelman OFF-PA-10000249  
## 7 36 CA-2017-117590 GH-14485 Gene Hale TEC-PH-10004977  
## 8 37 CA-2017-117590 GH-14485 Gene Hale FUR-FU-10003664  
## 9 38 CA-2016-117415 SN-20710 Steve Nguyen OFF-EN-10002986  
## 10 39 CA-2016-117415 SN-20710 Steve Nguyen FUR-BO-10002545  
## 11 40 CA-2016-117415 SN-20710 Steve Nguyen FUR-CH-10004218  
## 12 41 CA-2016-117415 SN-20710 Steve Nguyen TEC-PH-10000486  
## 13 42 CA-2018-120999 LC-16930 Linda Cazamias TEC-PH-10004093  
## 14 45 CA-2017-118255 ON-18715 Odella Nelson TEC-AC-10000171  
## 15 46 CA-2017-118255 ON-18715 Odella Nelson OFF-BI-10003291  
## Product.Name  
## 1 Holmes Replacement Filter for HEPA Air Cleaner, Very Large Room, HEPA Filter  
## 2 Storex DuraTech Recycled Plastic Frosted Binders  
## 3 Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18 1/2"D  
## 4 Newell 318  
## 5 Acco Six-Outlet Power Strip, 4' Cord Length  
## 6 Easy-staple paper  
## 7 GE 30524EE4  
## 8 Electrix Architect's Clamp-On Swing Arm Lamp, Black  
## 9 #10-4 1/8" x 9 1/2" Premium Diagonal Seam Envelopes  
## 10 Atlantic Metals Mobile 3-Shelf Bookcases, Custom Colors  
## 11 Global Fabric Manager's Chair, Dark Gray  
## 12 Plantronics HL10 Handset Lifter  
## 13 Panasonic Kx-TS550  
## 14 Verbatim 25 GB 6x Blu-ray Single Layer Recordable Disc, 25/Pack  
## 15 Wilson Jones Leather-Like Binders with DublLock Round Rings  
## Category Sub.Category Segment Ship.Mode Sales Quantity  
## 1 Office Supplies Appliances Home Office Standard Class 68.81 5  
## 2 Office Supplies Binders Home Office Standard Class 2.54 3  
## 3 Office Supplies Storage Consumer Standard Class 665.88 6  
## 4 Office Supplies Art Corporate Standard Class 19.46 7  
## 5 Office Supplies Appliances Corporate Standard Class 60.34 7  
## 6 Office Supplies Paper Home Office Second Class 29.47 3  
## 7 Technology Phones Corporate First Class 1097.54 7  
## 8 Furniture Furnishings Corporate First Class 190.92 5  
## 9 Office Supplies Envelopes Home Office Standard Class 113.33 9  
## 10 Furniture Bookcases Home Office Standard Class 532.40 3  
## 11 Furniture Chairs Home Office Standard Class 212.06 3  
## 12 Technology Phones Home Office Standard Class 371.17 4  
## 13 Technology Phones Corporate Standard Class 147.17 4  
## 14 Technology Accessories Corporate First Class 45.98 2  
## 15 Office Supplies Binders Corporate First Class 17.46 2  
## Profit Discount Order.Date Ship.Date City State Country  
## 1 -123.86 0.80 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 2 -3.82 0.80 2016-11-22 2016-11-26 Fort Worth Texas United States  
## 3 13.32 0.00 2015-11-11 2015-11-18 Madison Wisconsin United States  
## 4 5.06 0.00 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 5 15.69 0.00 2017-12-09 2017-12-13 Fremont Nebraska United States  
## 6 9.95 0.20 2018-10-19 2018-10-23 Houston Texas United States  
## 7 123.47 0.20 2017-12-08 2017-12-10 Richardson Texas United States  
## 8 -147.96 0.60 2017-12-08 2017-12-10 Richardson Texas United States  
## 9 35.41 0.20 2016-12-27 2016-12-31 Houston Texas United States  
## 10 -46.98 0.32 2016-12-27 2016-12-31 Houston Texas United States  
## 11 -15.15 0.30 2016-12-27 2016-12-31 Houston Texas United States  
## 12 41.76 0.20 2016-12-27 2016-12-31 Houston Texas United States  
## 13 16.56 0.20 2018-09-10 2018-09-15 Naperville Illinois United States  
## 14 19.77 0.00 2017-03-11 2017-03-13 Eagan Minnesota United States  
## 15 8.21 0.00 2017-03-11 2017-03-13 Eagan Minnesota United States  
## Postal.Code Region Shipping\_Days Year  
## 1 76106 Central 4 2016  
## 2 76106 Central 4 2016  
## 3 53711 Central 7 2015  
## 4 68025 Central 4 2017  
## 5 68025 Central 4 2017  
## 6 77095 Central 4 2018  
## 7 75080 Central 2 2017  
## 8 75080 Central 2 2017  
## 9 77041 Central 4 2016  
## 10 77041 Central 4 2016  
## 11 77041 Central 4 2016  
## 12 77041 Central 4 2016  
## 13 60540 Central 5 2018  
## 14 55122 Central 2 2017  
## 15 55122 Central 2 2017

dim(finalData)

## [1] 8321 23

Insights gained from finalData :

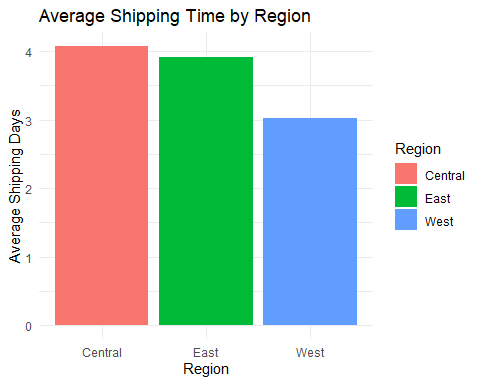
1. There are 23 columns and 8321 rows
2. There are no null values and duplicate values. All columns have proper datatype.

Question 1. Which region, on average, ships products faster?

avg\_shipping\_time <- data.frame(  
 Region = c("Central", "West", "East"),  
 Avg\_Shipping\_Days = c(mean(centralData$Shipping\_Days, na.rm = TRUE), mean(westData$Shipping\_Days, na.rm = TRUE), mean(eastData$Shipping\_Days, na.rm = TRUE))  
)  
print(avg\_shipping\_time)

## Region Avg\_Shipping\_Days  
## 1 Central 4.072751  
## 2 West 3.015388  
## 3 East 3.913417

ggplot(avg\_shipping\_time, aes(x = Region, y = Avg\_Shipping\_Days, fill = Region)) +  
 geom\_col() + scale\_y\_continuous(labels = scales::comma) +   
 labs(title = "Average Shipping Time by Region", x = "Region", y = "Average Shipping Days") +  
 theme\_minimal()



Answer:

On average Region West Ships products faster than the other regions

Insight gain: Central Region takes more time (average shipping time) than East and West regions. Region West takes least time to ship products

Question 2. Which products ship slowest by region?

sl\_sp\_prod <- function(data, region\_name) {  
 data %>% mutate(Shipping\_Days = as.numeric(Ship.Date - Order.Date)) %>%  
 filter(Shipping\_Days == max(Shipping\_Days, na.rm = TRUE)) %>%  
 select(Product.Name, Shipping\_Days) %>%  
 mutate(Region = region\_name)  
}  
  
# Slowest shipped product from each region  
sl\_sp\_central <- sl\_sp\_prod(centralData, "Central") %>% distinct(Product.Name, .keep\_all = TRUE)  
sl\_sp\_west <- sl\_sp\_prod(westData, "West") %>% distinct(Product.Name, .keep\_all = TRUE)  
sl\_sp\_east <- sl\_sp\_prod(eastData, "East") %>% distinct(Product.Name, .keep\_all = TRUE)  
  
print(sl\_sp\_central)

## Product.Name Shipping\_Days  
## 1 Stanley Bostitch Contemporary Electric Pencil Sharpeners 12  
## 2 Bush Westfield Collection Bookcases, Medium Cherry Finish 12  
## Region  
## 1 Central  
## 2 Central

print(sl\_sp\_west)

## Product.Name  
## 1 Avery Durable Slant Ring Binders, No Labels  
## 2 Trav-L-File Heavy-Duty Shuttle II, Black  
## 3 3M Office Air Cleaner  
## 4 Logitech G700s Rechargeable Gaming Mouse  
## 5 Acco D-Ring Binder w/DublLock  
## 6 Eldon Wave Desk Accessories  
## 7 Bush Advantage Collection Racetrack Conference Table  
## 8 Poly Designer Cover & Back  
## 9 Premier Electric Letter Opener  
## 10 Fellowes Premier Superior Surge Suppressor, 10-Outlet, With Phone and Remote  
## 11 Cisco SPA525G2 IP Phone - Wireless  
## 12 Personal Creations Ink Jet Cards and Labels  
## 13 GBC White Gloss Covers, Plain Front  
## 14 Xerox 222  
## 15 Hon Every-Day Series Multi-Task Chairs  
## 16 Fellowes Superior 10 Outlet Split Surge Protector  
## 17 Sharp AL-1530CS Digital Copier  
## 18 AT&T 1070 Corded Phone  
## 19 Dana Halogen Swing-Arm Architect Lamp  
## 20 Staples  
## 21 Belkin F9H710-06 7 Outlet SurgeMaster Surge Protector  
## 22 Staple remover  
## 23 Deluxe Chalkboard Eraser Cleaner  
## 24 Xerox 1987  
## 25 Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11", Dark Blue  
## 26 Bretford CR4500 Series Slim Rectangular Table  
## 27 Avery Hole Reinforcements  
## 28 Wilson Jones Ledger-Size, Piano-Hinge Binder, 2", Blue  
## 29 Avery Durable Slant Ring Binders With Label Holder  
## 30 Round Ring Binders  
## 31 Micropad Numeric Keypads  
## 32 Stanley Contemporary Battery Pencil Sharpeners  
## 33 AT&T 1080 Corded phone  
## 34 Wilson Jones Heavy-Duty Casebound Ring Binders with Metal Hinges  
## 35 iKross Bluetooth Portable Keyboard + Cell Phone Stand Holder + Brush for Apple iPhone 5S 5C 5, 4S 4  
## 36 DAX Value U-Channel Document Frames, Easel Back  
## 37 Boston Heavy-Duty Trimline Electric Pencil Sharpeners  
## 38 4009 Highlighters by Sanford  
## 39 Newell 310  
## 40 GE 48" Fluorescent Tube, Cool White Energy Saver, 34 Watts, 30/Box  
## 41 Dot Matrix Printer Tape Reel Labels, White, 5000/Box  
## 42 Micro Innovations USB RF Wireless Keyboard with Mouse  
## 43 GBC Velobind Prepunched Cover Sets, Regency Series  
## 44 Cardinal Slant-D Ring Binders  
## 45 Avery 494  
## 46 Imation\xa016GB Mini TravelDrive USB 2.0\xa0Flash Drive  
## 47 Alphabetical Labels for Top Tab Filing  
## 48 Eldon Spacemaker Box, Quick-Snap Lid, Clear  
## 49 Plymouth Boxed Rubber Bands by Plymouth  
## 50 SanDisk Ultra 64 GB MicroSDHC Class 10 Memory Card  
## 51 Dixon Ticonderoga Erasable Colored Pencil Set, 12-Color  
## 52 Tennsco 6- and 18-Compartment Lockers  
## 53 BOSTON Ranger #55 Pencil Sharpener, Black  
## 54 GBC DocuBind P50 Personal Binding Machine  
## 55 Holmes HEPA Air Purifier  
## 56 Xerox 1988  
## 57 Harmony HEPA Quiet Air Purifiers  
## 58 Kingston Digital DataTraveler 16GB USB 2.0  
## 59 Panasonic KX MB2061 Multifunction Printer  
## 60 Self-Adhesive Address Labels for Typewriters by Universal  
## 61 Tenex Traditional Chairmats for Medium Pile Carpet, Standard Lip, 36" x 48"  
## 62 Office Impressions Heavy Duty Welded Shelving & Multimedia Storage Drawers  
## 63 Global Stack Chair without Arms, Black  
## 64 Bevis Steel Folding Chairs  
## 65 Acco Pressboard Covers with Storage Hooks, 14 7/8" x 11", Executive Red  
## 66 Satellite Sectional Post Binders  
## 67 Computer Room Manger, 14"  
## 68 Bady BDG101FRU Card Printer  
## 69 Eldon Fold 'N Roll Cart System  
## 70 Okidata C331dn Printer  
## 71 Tyvek Top-Opening Peel & Seel Envelopes, Plain White  
## 72 Xerox 1978  
## 73 Office Star Flex Back Scooter Chair with White Frame  
## 74 Acco Six-Outlet Power Strip, 4' Cord Length  
## 75 Acme Rosewood Handle Letter Opener  
## 76 Global Deluxe Stacking Chair, Gray  
## 77 Cardinal EasyOpen D-Ring Binders  
## 78 50 Colored Long Pencils  
## 79 Acco PRESSTEX Data Binder with Storage Hooks, Light Blue, 9 1/2" X 11"  
## 80 APC 7 Outlet Network SurgeArrest Surge Protector  
## 81 Chromcraft 48" x 96" Racetrack Double Pedestal Table  
## 82 Kensington 4 Outlet MasterPiece Compact Power Control Center  
## 83 Acco Hanging Data Binders  
## 84 Binder Clips by OIC  
## 85 Things To Do Today Spiral Book  
## 86 Apple iPhone 5S  
## 87 Hon 2090 "Pillow Soft" Series Mid Back Swivel/Tilt Chairs  
## 88 Tensor Computer Mounted Lamp  
## 89 Logitech Illuminated Ultrathin Keyboard with Backlighting  
## 90 RCA ViSYS 25423RE1 Corded phone  
## 91 Eldon Expressions Wood Desk Accessories, Oak  
## 92 Crayola Anti Dust Chalk, 12/Pack  
## 93 Newell 338  
## 94 Novimex Swivel Fabric Task Chair  
## 95 Howard Miller 11-1/2" Diameter Brentwood Wall Clock  
## 96 Polycom SoundPoint Pro SE-225 Corded phone  
## 97 Bevis Round Conference Table Top, X-Base  
## 98 Xerox 2000  
## 99 Hewlett-Packard Deskjet D4360 Printer  
## 100 Airmail Envelopes  
## 101 Wilson SignalBoost 841262 DB PRO Amplifier Kit  
## 102 Martin Yale Chadless Opener Electric Letter Opener  
## 103 Chromcraft Round Conference Tables  
## 104 Xerox 1974  
## 105 Lock-Up Easel 'Spel-Binder'  
## 106 24-Hour Round Wall Clock  
## 107 Logitech Media Keyboard K200  
## 108 Green Canvas Binder for 8-1/2" x 14" Sheets  
## 109 Newell 323  
## 110 Newell 317  
## 111 Computer Printout Paper with Letter-Trim Perforations  
## 112 Sanford EarthWrite Recycled Pencils, Medium Soft, #2  
## 113 SAFCO Boltless Steel Shelving  
## 114 Xerox 1911  
## 115 Staple envelope  
## 116 Tennsco Snap-Together Open Shelving Units, Starter Sets and Add-On Units  
## 117 Nu-Dell Executive Frame  
## 118 Acrylic Self-Standing Desk Frames  
## 119 Global Leather Executive Chair  
## 120 Bravo II Megaboss 12-Amp Hard Body Upright, Replacement Belts, 2 Belts per Pack  
## 121 Fiskars 8" Scissors, 2/Pack  
## 122 Newell 343  
## 123 Electrix 20W Halogen Replacement Bulb for Zoom-In Desk Lamp  
## 124 Sensible Storage WireTech Storage Systems  
## 125 Hon Olson Stacker Stools  
## 126 Vtech AT&T CL2940 Corded Speakerphone, Black  
## 127 Rogers Profile Extra Capacity Storage Tub  
## 128 Avery 475  
## 129 Easy-staple paper  
## 130 Panasonic KX-TG6844B Expandable Digital Cordless Telephone  
## 131 Alliance Rubber Bands  
## 132 invisibleSHIELD by ZAGG Smudge-Free Screen Protector  
## 133 Office Star - Contemporary Swivel Chair with Padded Adjustable Arms and Flex Back  
## 134 Fellowes PB300 Plastic Comb Binding Machine  
## 135 Letter Size Cart  
## 136 Contract Clock, 14", Brown  
## 137 Executive Impressions 10" Spectator Wall Clock  
## 138 Panasonic KX TS3282W Corded phone  
## 139 Toshiba IPT2010-SD IP\xa0Telephone  
## 140 Microsoft Wireless Mobile Mouse 4000  
## 141 Fellowes Bankers Box Stor/Drawer Steel Plus  
## 142 Letter or Legal Size Expandable Poly String Tie Envelopes  
## 143 Howard Miller 13" Diameter Pewter Finish Round Wall Clock  
## 144 KI Adjustable-Height Table  
## 145 3M Replacement Filter for Office Air Cleaner for 20' x 33' Room  
## 146 Fellowes Staxonsteel Drawer Files  
## 147 GBC Prepunched Paper, 19-Hole, for Binding Systems, 24-lb  
## 148 KI Conference Tables  
## 149 Panasonic KX MC6040 Color Laser Multifunction Printer  
## 150 Imation\xa0Clip USB\xa0flash drive\xa0- 8 GB  
## 151 Hypercom P1300 Pinpad  
## 152 Global Stack Chair with Arms, Black  
## 153 Avery Durable Slant Ring Binders  
## 154 Motorola HK250 Universal Bluetooth Headset  
## 155 Logitech\xa0Gaming G510s - Keyboard  
## 156 Anker Astro Mini 3000mAh Ultra-Compact Portable Charger  
## 157 Bretford CR8500 Series Meeting Room Furniture  
## 158 Xerox 1997  
## 159 Xerox 1928  
## 160 Telephone Message Books with Fax/Mobile Section, 5 1/2" x 3 3/16"  
## 161 Surelock Post Binders  
## 162 Laser & Ink Jet Business Envelopes  
## 163 Hoover Commercial Soft Guard Upright Vacuum And Disposable Filtration Bags  
## 164 Cardinal Slant-D Ring Binder, Heavy Gauge Vinyl  
## 165 Xerox 1947  
## 166 Hon Metal Bookcases, Putty  
## 167 Eureka The Boss Lite 10-Amp Upright Vacuum, Blue  
## 168 GBC VeloBinder Manual Binding System  
## 169 36X48 HARDFLOOR CHAIRMAT  
## 170 Eberhard Faber 3 1/2" Golf Pencils  
## 171 Xerox 1900  
## 172 Mini 13-1/2 Capacity Data Binder Rack, Pearl  
## 173 Acme Tagit Stainless Steel Antibacterial Scissors  
## 174 Xerox 203  
## 175 Iceberg Nesting Folding Chair, 19w x 6d x 43h  
## 176 OtterBox Defender Series Case - Samsung Galaxy S4  
## 177 Xerox 1962  
## 178 Wilson Jones Clip & Carry Folder Binder Tool for Ring Binders, Clear  
## 179 Sanyo 2.5 Cubic Foot Mid-Size Office Refrigerators  
## 180 Standard Rollaway File with Lock  
## 181 GBC Recycled Grain Textured Covers  
## 182 REDIFORM Incoming/Outgoing Call Register, 11" X 8 1/2", 100 Messages  
## 183 Hot File 7-Pocket, Floor Stand  
## 184 O'Sullivan 2-Door Barrister Bookcase in Odessa Pine  
## 185 Carina Double Wide Media Storage Towers in Natural & Black  
## 186 GBC Wire Binding Combs  
## 187 Hon 5100 Series Wood Tables  
## 188 SanDisk Cruzer 64 GB USB Flash Drive  
## 189 Hewlett Packard 310 Color Digital Copier  
## 190 Xerox 1937  
## 191 Grandstream GXP1160 VoIP phone  
## 192 Personal File Boxes with Fold-Down Carry Handle  
## 193 Perixx PERIBOARD-512B, Ergonomic Split Keyboard  
## 194 Avery Legal 4-Ring Binder  
## 195 Eldon Simplefile Box Office  
## 196 AT&T SB67148 SynJ  
## 197 Memorex Mini Travel Drive 64 GB USB 2.0 Flash Drive  
## 198 Howard Miller Distant Time Traveler Alarm Clock  
## 199 Nortel Business Series Terminal T7208 Digital phone  
## 200 Global Leather & Oak Executive Chair, Burgundy  
## 201 Logitech\xa0P710e Mobile Speakerphone  
## 202 Fellowes 8 Outlet Superior Workstation Surge Protector w/o Phone/Fax/Modem Protection  
## 203 Imation\xa0USB 2.0 Swivel\xa0Flash Drive\xa0USB\xa0flash drive\xa0- 4 GB - Pink  
## 204 Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind  
## 205 Luxo Professional Magnifying Clamp-On Fluorescent Lamps  
## 206 Perma STOR-ALL Hanging File Box, 13 1/8"W x 12 1/4"D x 10 1/2"H  
## 207 AT&T 841000 Phone  
## 208 Rediform S.O.S. 1-Up Phone Message Bk, 4-1/4x3-1/16 Bk, 1 Form/Pg, 40 Messages/Bk, 3/Pk  
## 209 Luxo Economy Swing Arm Lamp  
## 210 DAX Natural Wood-Tone Poster Frame  
## 211 Acco Smartsocket Table Surge Protector, 6 Color-Coded Adapter Outlets  
## 212 GBC Recycled VeloBinder Covers  
## 213 Acme Galleria Hot Forged Steel Scissors with Colored Handles  
## 214 Adams Telephone Message Book W/Dividers/Space For Phone Numbers, 5 1/4"X8 1/2", 300/Messages  
## 215 Newell 346  
## 216 Howard Miller 13-1/2" Diameter Rosebrook Wall Clock  
## 217 Lesro Round Back Collection Coffee Table, End Table  
## 218 Acme Hot Forged Carbon Steel Scissors with Nickel-Plated Handles, 3 7/8" Cut, 8"L  
## Shipping\_Days Region  
## 1 6 West  
## 2 6 West  
## 3 6 West  
## 4 6 West  
## 5 6 West  
## 6 6 West  
## 7 6 West  
## 8 6 West  
## 9 6 West  
## 10 6 West  
## 11 6 West  
## 12 6 West  
## 13 6 West  
## 14 6 West  
## 15 6 West  
## 16 6 West  
## 17 6 West  
## 18 6 West  
## 19 6 West  
## 20 6 West  
## 21 6 West  
## 22 6 West  
## 23 6 West  
## 24 6 West  
## 25 6 West  
## 26 6 West  
## 27 6 West  
## 28 6 West  
## 29 6 West  
## 30 6 West  
## 31 6 West  
## 32 6 West  
## 33 6 West  
## 34 6 West  
## 35 6 West  
## 36 6 West  
## 37 6 West  
## 38 6 West  
## 39 6 West  
## 40 6 West  
## 41 6 West  
## 42 6 West  
## 43 6 West  
## 44 6 West  
## 45 6 West  
## 46 6 West  
## 47 6 West  
## 48 6 West  
## 49 6 West  
## 50 6 West  
## 51 6 West  
## 52 6 West  
## 53 6 West  
## 54 6 West  
## 55 6 West  
## 56 6 West  
## 57 6 West  
## 58 6 West  
## 59 6 West  
## 60 6 West  
## 61 6 West  
## 62 6 West  
## 63 6 West  
## 64 6 West  
## 65 6 West  
## 66 6 West  
## 67 6 West  
## 68 6 West  
## 69 6 West  
## 70 6 West  
## 71 6 West  
## 72 6 West  
## 73 6 West  
## 74 6 West  
## 75 6 West  
## 76 6 West  
## 77 6 West  
## 78 6 West  
## 79 6 West  
## 80 6 West  
## 81 6 West  
## 82 6 West  
## 83 6 West  
## 84 6 West  
## 85 6 West  
## 86 6 West  
## 87 6 West  
## 88 6 West  
## 89 6 West  
## 90 6 West  
## 91 6 West  
## 92 6 West  
## 93 6 West  
## 94 6 West  
## 95 6 West  
## 96 6 West  
## 97 6 West  
## 98 6 West  
## 99 6 West  
## 100 6 West  
## 101 6 West  
## 102 6 West  
## 103 6 West  
## 104 6 West  
## 105 6 West  
## 106 6 West  
## 107 6 West  
## 108 6 West  
## 109 6 West  
## 110 6 West  
## 111 6 West  
## 112 6 West  
## 113 6 West  
## 114 6 West  
## 115 6 West  
## 116 6 West  
## 117 6 West  
## 118 6 West  
## 119 6 West  
## 120 6 West  
## 121 6 West  
## 122 6 West  
## 123 6 West  
## 124 6 West  
## 125 6 West  
## 126 6 West  
## 127 6 West  
## 128 6 West  
## 129 6 West  
## 130 6 West  
## 131 6 West  
## 132 6 West  
## 133 6 West  
## 134 6 West  
## 135 6 West  
## 136 6 West  
## 137 6 West  
## 138 6 West  
## 139 6 West  
## 140 6 West  
## 141 6 West  
## 142 6 West  
## 143 6 West  
## 144 6 West  
## 145 6 West  
## 146 6 West  
## 147 6 West  
## 148 6 West  
## 149 6 West  
## 150 6 West  
## 151 6 West  
## 152 6 West  
## 153 6 West  
## 154 6 West  
## 155 6 West  
## 156 6 West  
## 157 6 West  
## 158 6 West  
## 159 6 West  
## 160 6 West  
## 161 6 West  
## 162 6 West  
## 163 6 West  
## 164 6 West  
## 165 6 West  
## 166 6 West  
## 167 6 West  
## 168 6 West  
## 169 6 West  
## 170 6 West  
## 171 6 West  
## 172 6 West  
## 173 6 West  
## 174 6 West  
## 175 6 West  
## 176 6 West  
## 177 6 West  
## 178 6 West  
## 179 6 West  
## 180 6 West  
## 181 6 West  
## 182 6 West  
## 183 6 West  
## 184 6 West  
## 185 6 West  
## 186 6 West  
## 187 6 West  
## 188 6 West  
## 189 6 West  
## 190 6 West  
## 191 6 West  
## 192 6 West  
## 193 6 West  
## 194 6 West  
## 195 6 West  
## 196 6 West  
## 197 6 West  
## 198 6 West  
## 199 6 West  
## 200 6 West  
## 201 6 West  
## 202 6 West  
## 203 6 West  
## 204 6 West  
## 205 6 West  
## 206 6 West  
## 207 6 West  
## 208 6 West  
## 209 6 West  
## 210 6 West  
## 211 6 West  
## 212 6 West  
## 213 6 West  
## 214 6 West  
## 215 6 West  
## 216 6 West  
## 217 6 West  
## 218 6 West

print(sl\_sp\_east)

## Product.Name  
## 1 Pressboard Covers with Storage Hooks, 9 1/2" x 11", Light Blue  
## 2 Cisco 9971 IP Video Phone Charcoal  
## 3 Newell 3-Hole Punched Plastic Slotted Magazine Holders for Binders  
## 4 GBC DocuBind 300 Electric Binding Machine  
## 5 GBC Wire Binding Strips  
## 6 Angle-D Ring Binders  
## 7 GBC Personal VeloBind Strips  
## 8 Wilson Jones Ledger-Size, Piano-Hinge Binder, 2", Blue  
## 9 GBC Ibimaster 500 Manual ProClick Binding System  
## 10 Cubify CubeX 3D Printer Double Head Print  
## 11 GBC Standard Plastic Binding Systems Combs  
## 12 O'Sullivan 2-Door Barrister Bookcase in Odessa Pine  
## 13 Speck Products Candyshell Flip Case  
## 14 Samsung Galaxy Note 3  
## 15 Bose SoundLink Bluetooth Speaker  
## 16 Plantronics Voyager Pro HD - Bluetooth Headset  
## 17 Hon Rectangular Conference Tables  
## 18 Mitel MiVoice 5330e IP Phone  
## 19 Cisco IP Phone 7961G-GE VoIP phone  
## 20 Global Adaptabilities Conference Tables  
## 21 Office Star - Contemporary Task Swivel Chair  
## 22 Hon Every-Day Series Multi-Task Chairs  
## 23 Global Deluxe Steno Chair  
## 24 Wirebound Message Books, 5-1/2 x 4 Forms, 2 or 4 Forms per Page  
## 25 Avery 505  
## 26 Sony Micro Vault Click 16 GB USB 2.0 Flash Drive  
## 27 Storex Dura Pro Binders  
## 28 Dixon My First Ticonderoga Pencil, #2  
## 29 Stockwell Gold Paper Clips  
## 30 GBC DocuBind P100 Manual Binding Machine  
## 31 Xerox 203  
## 32 Newell 35  
## 33 Avery 488  
## 34 Avery Hi-Liter EverBold Pen Style Fluorescent Highlighters, 4/Pack  
## 35 Xerox 222  
## 36 Acco Data Flex Cable Posts For Top & Bottom Load Binders, 6" Capacity  
## 37 Southworth 100% R\xe9sum\xe9 Paper, 24lb.  
## 38 GBC Poly Designer Binding Covers  
## 39 Avery Durable Plastic 1" Binders  
## 40 Binney & Smith inkTank Erasable Pocket Highlighter, Chisel Tip, Yellow  
## 41 Logitech Trackman Marble Mouse  
## 42 Eldon Cleatmat Plus Chair Mats for High Pile Carpets  
## 43 O'Sullivan Plantations 2-Door Library in Landvery Oak  
## 44 GBC Imprintable Covers  
## 45 Plantronics CS510 - Over-the-Head monaural Wireless Headset System  
## 46 Sharp AL-1530CS Digital Copier  
## 47 Newell 347  
## 48 Avery 486  
## 49 Acco Perma 2700 Stacking Storage Drawers  
## 50 Wilson Jones data.warehouse D-Ring Binders with DublLock  
## 51 Tennsco Lockers, Gray  
## 52 Memorex Mini Travel Drive 8 GB USB 2.0 Flash Drive  
## 53 Wilson Jones Clip & Carry Folder Binder Tool for Ring Binders, Clear  
## 54 Avery 476  
## 55 Avery Binding System Hidden Tab Executive Style Index Sets  
## 56 Black Avery Memo-Size 3-Ring Binder, 5 1/2" x 8 1/2"  
## 57 Wilson Jones Legal Size Ring Binders  
## 58 NETGEAR RangeMax WNR1000 Wireless Router  
## 59 Wilson Jones "Snap" Scratch Pad Binder Tool for Ring Binders  
## 60 Fellowes Neat Ideas Storage Cubes  
## 61 Avery Metallic Poly Binders  
## 62 GBC DocuBind TL300 Electric Binding System  
## 63 Cardinal Holdit Business Card Pockets  
## 64 Wilson Jones Turn Tabs Binder Tool for Ring Binders  
## 65 Iceberg Nesting Folding Chair, 19w x 6d x 43h  
## 66 Situations Contoured Folding Chairs, 4/Set  
## 67 Global Airflow Leather Mesh Back Chair, Black  
## 68 Global Leather Executive Chair  
## 69 Office Star - Mid Back Dual function Ergonomic High Back Chair with 2-Way Adjustable Arms  
## 70 Decoflex Hanging Personal Folder File  
## 71 Staples  
## 72 Avery 473  
## 73 OIC Bulk Pack Metal Binder Clips  
## 74 Cush Cases Heavy Duty Rugged Cover Case for Samsung Galaxy S5 - Purple  
## 75 Razer Kraken PRO Over Ear PC and Music Headset  
## 76 Avery 508  
## 77 3M Hangers With Command Adhesive  
## 78 Advantus Push Pins  
## 79 Woodgrain Magazine Files by Perma  
## 80 SAFCO Commercial Wire Shelving, Black  
## 81 Jabra SPEAK 410  
## 82 Acme Hot Forged Carbon Steel Scissors with Nickel-Plated Handles, 3 7/8" Cut, 8"L  
## 83 Xerox 1913  
## 84 Deluxe Heavy-Duty Vinyl Round Ring Binder  
## 85 Global Value Steno Chair, Gray  
## 86 Deflect-o Glass Clear Studded Chair Mats  
## 87 Adjustable Depth Letter/Legal Cart  
## 88 Flexible Leather- Look Classic Collection Ring Binder  
## 89 Ultra Door Push Plate  
## 90 Compact Automatic Electric Letter Opener  
## 91 Staple holder  
## 92 DAX Two-Tone Silver Metal Document Frame  
## 93 Crayola Anti Dust Chalk, 12/Pack  
## 94 Ames Color-File Green Diamond Border X-ray Mailers  
## 95 Xerox 214  
## 96 Office Star - Contemporary Swivel Chair with Padded Adjustable Arms and Flex Back  
## 97 Hon Metal Bookcases, Putty  
## 98 SANFORD Liquid Accent Tank-Style Highlighters  
## 99 Novimex Turbo Task Chair  
## 100 Fellowes Advanced Computer Series Surge Protectors  
## 101 Vinyl Coated Wire Paper Clips in Organizer Box, 800/Box  
## 102 Ativa MDM8000 8-Sheet Micro-Cut Shredder  
## 103 Acme Elite Stainless Steel Scissors  
## 104 Barrel Sharpener  
## 105 DAX Black Cherry Wood-Tone Poster Frame  
## 106 Xerox 204  
## 107 Xerox 1975  
## 108 Plantronics MX500i Earset  
## 109 Master Big Foot Doorstop, Beige  
## 110 BIC Liqua Brite Liner  
## 111 Fellowes Bases and Tops For Staxonsteel/High-Stak Systems  
## 112 Computer Printout Paper with Letter-Trim Perforations  
## 113 Avery White Multi-Purpose Labels  
## 114 Howard Miller Distant Time Traveler Alarm Clock  
## 115 HP Designjet T520 Inkjet Large Format Printer - 24" Color  
## 116 Permanent Self-Adhesive File Folder Labels for Typewriters by Universal  
## 117 Gear Head AU3700S Headset  
## 118 BIC Brite Liner Grip Highlighters  
## 119 Space Solutions HD Industrial Steel Shelving.  
## 120 Acme Kleen Earth Office Shears  
## 121 Panasonic KP-380BK Classic Electric Pencil Sharpener  
## 122 Fiskars Softgrip Scissors  
## 123 OIC Stacking Trays  
## 124 Xerox 1885  
## 125 Manco Dry-Lighter Erasable Highlighter  
## 126 Xerox 1917  
## 127 Xerox 1970  
## 128 Plantronics Savi W720 Multi-Device Wireless Headset System  
## 129 Sauder Inglewood Library Bookcases  
## 130 Global Fabric Manager's Chair, Dark Gray  
## 131 Samsung Galaxy S4  
## 132 invisibleSHIELD by ZAGG Smudge-Free Screen Protector  
## 133 Avery 485  
## 134 DAX Wood Document Frame  
## 135 Avery 503  
## 136 Eldon Pizzaz Desk Accessories  
## 137 Ibico Laser Imprintable Binding System Covers  
## 138 Fellowes Advanced 8 Outlet Surge Suppressor with Phone/Fax Protection  
## 139 Xerox 208  
## 140 Newell 32  
## 141 Computer Room Manger, 14"  
## 142 Avery Hi-Liter GlideStik Fluorescent Highlighter, Yellow Ink  
## 143 Space Solutions Commercial Steel Shelving  
## 144 Acme Serrated Blade Letter Opener  
## Shipping\_Days Region  
## 1 7 East  
## 2 7 East  
## 3 7 East  
## 4 7 East  
## 5 7 East  
## 6 7 East  
## 7 7 East  
## 8 7 East  
## 9 7 East  
## 10 7 East  
## 11 7 East  
## 12 7 East  
## 13 7 East  
## 14 7 East  
## 15 7 East  
## 16 7 East  
## 17 7 East  
## 18 7 East  
## 19 7 East  
## 20 7 East  
## 21 7 East  
## 22 7 East  
## 23 7 East  
## 24 7 East  
## 25 7 East  
## 26 7 East  
## 27 7 East  
## 28 7 East  
## 29 7 East  
## 30 7 East  
## 31 7 East  
## 32 7 East  
## 33 7 East  
## 34 7 East  
## 35 7 East  
## 36 7 East  
## 37 7 East  
## 38 7 East  
## 39 7 East  
## 40 7 East  
## 41 7 East  
## 42 7 East  
## 43 7 East  
## 44 7 East  
## 45 7 East  
## 46 7 East  
## 47 7 East  
## 48 7 East  
## 49 7 East  
## 50 7 East  
## 51 7 East  
## 52 7 East  
## 53 7 East  
## 54 7 East  
## 55 7 East  
## 56 7 East  
## 57 7 East  
## 58 7 East  
## 59 7 East  
## 60 7 East  
## 61 7 East  
## 62 7 East  
## 63 7 East  
## 64 7 East  
## 65 7 East  
## 66 7 East  
## 67 7 East  
## 68 7 East  
## 69 7 East  
## 70 7 East  
## 71 7 East  
## 72 7 East  
## 73 7 East  
## 74 7 East  
## 75 7 East  
## 76 7 East  
## 77 7 East  
## 78 7 East  
## 79 7 East  
## 80 7 East  
## 81 7 East  
## 82 7 East  
## 83 7 East  
## 84 7 East  
## 85 7 East  
## 86 7 East  
## 87 7 East  
## 88 7 East  
## 89 7 East  
## 90 7 East  
## 91 7 East  
## 92 7 East  
## 93 7 East  
## 94 7 East  
## 95 7 East  
## 96 7 East  
## 97 7 East  
## 98 7 East  
## 99 7 East  
## 100 7 East  
## 101 7 East  
## 102 7 East  
## 103 7 East  
## 104 7 East  
## 105 7 East  
## 106 7 East  
## 107 7 East  
## 108 7 East  
## 109 7 East  
## 110 7 East  
## 111 7 East  
## 112 7 East  
## 113 7 East  
## 114 7 East  
## 115 7 East  
## 116 7 East  
## 117 7 East  
## 118 7 East  
## 119 7 East  
## 120 7 East  
## 121 7 East  
## 122 7 East  
## 123 7 East  
## 124 7 East  
## 125 7 East  
## 126 7 East  
## 127 7 East  
## 128 7 East  
## 129 7 East  
## 130 7 East  
## 131 7 East  
## 132 7 East  
## 133 7 East  
## 134 7 East  
## 135 7 East  
## 136 7 East  
## 137 7 East  
## 138 7 East  
## 139 7 East  
## 140 7 East  
## 141 7 East  
## 142 7 East  
## 143 7 East  
## 144 7 East

Answer:

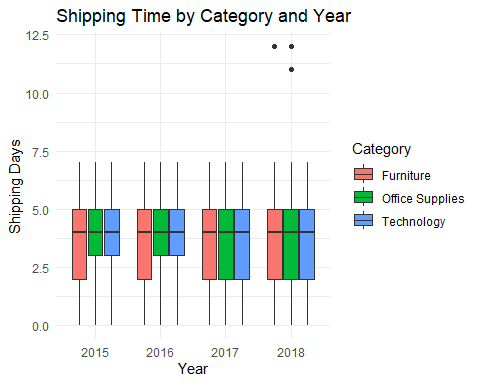
There 2 products from central region that were slowest shipped products which took 12 days, they are Stanley Bostitch Contemporary Electric Pencil Sharpeners and Bush Westfield Collection Bookcases, Medium Cherry Finish

There are 218 products from West region that were slowest shipped products which took 6 days.

There are 144 products from east region that were slowest shipped products which took 7 days.

Question 3. Plot time to ship by category, by year.

ggplot(finalData, aes(x = Year, y = Shipping\_Days, fill = Category)) +  
 geom\_boxplot() + labs(title = "Shipping Time by Category and Year", x = "Year", y = "Shipping Days", fill = "Category") + theme\_minimal()



Answer:

Insights gained:

Across all years (2015-2018), the median shipping time remains between 3 to 4 days for all product categories (Almost same median shipping time for all products).

The Furniture Category products consistently has a wide spread showing that it sometimes takes longer time (days) to ship.This says that Furniture shipments experience more variation in delivery times

Both Office Supplies and Technology have similar distributions, mostly shipping within 3 to 4.5 days.They have shorter whiskers, indicating that their shipping times are more predictable and consistent.

In 2018 there are two extreme outliers for Furniture and Office Supplies Which shows that in some cases, shipments were delayed beyond 10 days, which is unusual compared to previous years.

The overall shipping time distributions remain stable from 2015 to 2017, without drastic changes. The only major change is the outliers in 2018.

Question 4. Which category had the highest profit in each region and in what order year?

pft\_info <- finalData %>%  
 group\_by(Region, Year, Category) %>%  
 summarise(Total\_Profit = sum(Profit, na.rm = TRUE)) %>%  
 arrange(Region, Year, desc(Total\_Profit))

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

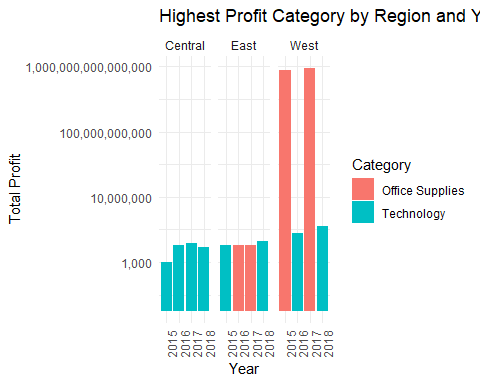
print(pft\_info)

## # A tibble: 36 × 4  
## # Groups: Region, Year [12]  
## Region Year Category Total\_Profit  
## <chr> <chr> <chr> <dbl>  
## 1 Central 2015 Technology 1022.  
## 2 Central 2015 Office Supplies 622.  
## 3 Central 2015 Furniture -1105.  
## 4 Central 2016 Technology 10168.  
## 5 Central 2016 Office Supplies 1780.  
## 6 Central 2016 Furniture -231.  
## 7 Central 2017 Technology 13761.  
## 8 Central 2017 Office Supplies 6392.  
## 9 Central 2017 Furniture -254.  
## 10 Central 2018 Technology 8747.  
## # ℹ 26 more rows

hg\_pft\_cgy <- pft\_info %>%  
 group\_by(Region, Year) %>%  
 slice\_max(Total\_Profit, n = 1)  
print(hg\_pft\_cgy)

## # A tibble: 12 × 4  
## # Groups: Region, Year [12]  
## Region Year Category Total\_Profit  
## <chr> <chr> <chr> <dbl>  
## 1 Central 2015 Technology 1.02e 3  
## 2 Central 2016 Technology 1.02e 4  
## 3 Central 2017 Technology 1.38e 4  
## 4 Central 2018 Technology 8.75e 3  
## 5 East 2015 Technology 1.04e 4  
## 6 East 2016 Office Supplies 1.10e 4  
## 7 East 2017 Office Supplies 1.11e 4  
## 8 East 2018 Technology 1.86e 4  
## 9 West 2015 Office Supplies 6.28e14  
## 10 West 2016 Technology 5.89e 4  
## 11 West 2017 Office Supplies 7.92e14  
## 12 West 2018 Technology 1.50e 5

ggplot(hg\_pft\_cgy, aes(x = as.factor(Year), y = Total\_Profit, fill = Category)) +  
 geom\_col(position = "dodge") + facet\_wrap(~Region) +   
 scale\_y\_log10(labels = scales::comma) +   
 labs(title = "Highest Profit Category by Region and Year", x = "Year", y = "Total Profit") +  
 theme\_minimal()+ theme(axis.text.x = element\_text(angle = 90, hjust = 1))



Answer:

Categories with highest profit from each region and every year:

From Region Central:

In 2015 - Technology

In 2016 - Technology

In 2017 - Technology

In 2018 - Technology

From region East:

In 2015 - Technology

In 2016 - Office Supplies

In 2017 - Office Supplies

In 2018 - Technology

From region West:

In 2015 - Office Supplies

In 2016 - Technology

In 2017 - Office Supplies

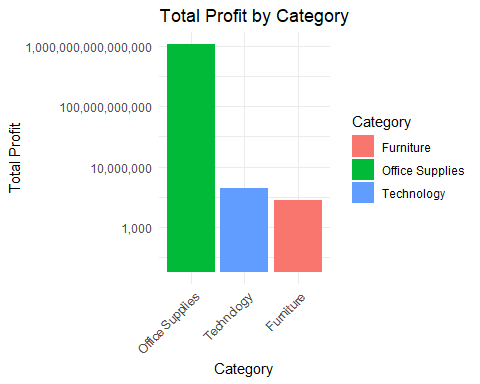
In 2018 - Technology

Question 5. Which category has the highest total profit totaling all years and regions?

hg\_tol\_pft\_cgy <- finalData %>%  
 group\_by(Category) %>%  
 summarise(Total\_Profit = sum(Profit, na.rm = TRUE)) %>%  
 arrange(desc(Total\_Profit))  
print(hg\_tol\_pft\_cgy )

## # A tibble: 3 × 2  
## Category Total\_Profit  
## <chr> <dbl>  
## 1 Office Supplies 1.42e15  
## 2 Technology 4.00e 5  
## 3 Furniture 6.15e 4

ggplot(hg\_tol\_pft\_cgy, aes(x = reorder(Category, -Total\_Profit), y = Total\_Profit, fill = Category)) +  
 geom\_col() + scale\_y\_log10(labels = scales::comma) +  
 labs(title = "Total Profit by Category", x = "Category", y = "Total Profit") +  
 theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))



Answer:

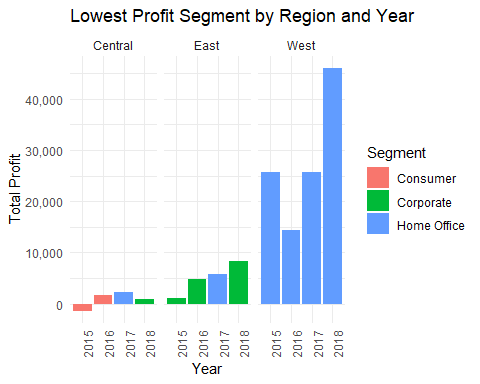
Office Supplies has the highest total profit

Question 6. Which segment had the lowest profit in each region and in what order year?

lw\_pft\_seg <- finalData %>%  
 group\_by(Region, Year, Segment) %>%   
 summarise(Total\_Profit = sum(Profit, na.rm = TRUE), .groups = "drop") %>%   
 group\_by(Region, Year) %>% filter(Total\_Profit == min(Total\_Profit))  
print(lw\_pft\_seg)

## # A tibble: 12 × 4  
## # Groups: Region, Year [12]  
## Region Year Segment Total\_Profit  
## <chr> <chr> <chr> <dbl>  
## 1 Central 2015 Consumer -1419.  
## 2 Central 2016 Consumer 1762.  
## 3 Central 2017 Home Office 2217.  
## 4 Central 2018 Corporate 891.  
## 5 East 2015 Corporate 1040.  
## 6 East 2016 Corporate 4732.  
## 7 East 2017 Home Office 5817.  
## 8 East 2018 Corporate 8319.  
## 9 West 2015 Home Office 25625.  
## 10 West 2016 Home Office 14391.  
## 11 West 2017 Home Office 25625.  
## 12 West 2018 Home Office 46000.

ggplot(lw\_pft\_seg, aes(x = as.factor(Year), y = Total\_Profit, fill = Segment)) +  
 geom\_col(position = "dodge") + facet\_wrap(~Region) +   
 scale\_y\_continuous(labels = scales::comma) +   
 labs(title = "Lowest Profit Segment by Region and Year", x = "Year", y = "Total Profit") +  
 theme\_minimal()+ theme(axis.text.x = element\_text(angle = 90, hjust = 1))



Answer:

Segments with Lowest profit from each region and every year:

From Region Central:

In 2015 - Consumer

In 2016 - Consumer

In 2017 - Home Office

In 2018 - Corporate

From region East:

In 2015 - Corporate

In 2016 - Corporate

In 2017 - home Office

In 2018 - Corporate

From region West:

In 2015 - home Office

In 2016 - home Office

In 2017 - home Office

In 2018 - home Office

Question 7. What are yearly sales by region?

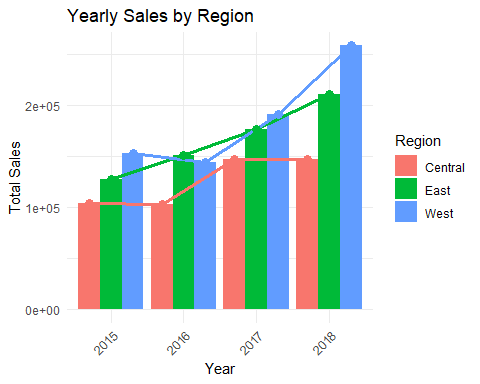
yr\_sales\_rgn <- finalData %>%  
 group\_by(Region, Year) %>%   
 summarise(Total\_Sales = sum(Sales, na.rm = TRUE)) %>%   
 arrange(Region, Year)

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

print(yr\_sales\_rgn)

## # A tibble: 12 × 3  
## # Groups: Region [3]  
## Region Year Total\_Sales  
## <chr> <chr> <dbl>  
## 1 Central 2015 103838.  
## 2 Central 2016 102874.  
## 3 Central 2017 147429.  
## 4 Central 2018 147098   
## 5 East 2015 128093.  
## 6 East 2016 150878.  
## 7 East 2017 176145.  
## 8 East 2018 210937.  
## 9 West 2015 152932.  
## 10 West 2016 144556.  
## 11 West 2017 191305.  
## 12 West 2018 258390.

ggplot(yr\_sales\_rgn, aes(x = Year, y = Total\_Sales, fill = Region)) +  
 geom\_bar(stat = "identity", position = "dodge") + # Clustered bars  
 geom\_line(aes(group = Region, color = Region), position = position\_dodge(width = 0.9), linewidth = 1.2) +   
 geom\_point(aes(color = Region), position = position\_dodge(width = 0.9), size = 3) +   
 labs(title = "Yearly Sales by Region", x = "Year", y = "Total Sales", fill = "Region", color = "Region") +  
 theme\_minimal() + theme(axis.text.x = element\_text(angle = 45, hjust = 1))



Answer:

Yearly sales from each region:

From Region Central

2015 - 103838.1

2016 - 102874.2

2017 - 147429.4 ( Highest Sale )

2018 - 147098.0

From Region East

2015 - 128092.7

2016 - 150877.6

2017 - 176145.3

2018 - 210937.3 ( Highest Sale )

From Region West

2015 - 152931.6

2016 - 144555.8

2017 - 191304.8

2018 - 258389.6 ( Highest Sale )

Insight Gained:

Across all years Region west has more sales followed by East. Region Central has lowest sales across all years compared to all other regions

Question 8.Over all the years which customer purchased the most in a single order (by sales dollar amount) and what did they buy (product and quantity)?

lg\_ord <- finalData %>% arrange(desc(Sales)) %>% slice(1)   
print(lg\_ord)

## Row.ID Order.ID Customer.ID Customer.Name Product.ID  
## 1575 6827 CA-2017-118689 TC-20980 Tamara Chand TEC-CO-10004722  
## Product.Name Category Sub.Category Segment  
## 1575 Canon imageCLASS 2200 Advanced Copier Technology Copiers Corporate  
## Ship.Mode Sales Quantity Profit Discount Order.Date Ship.Date  
## 1575 Standard Class 17499.95 5 8399.98 0 2017-10-02 2017-10-09  
## City State Country Postal.Code Region Shipping\_Days Year  
## 1575 Lafayette Indiana United States 47905 Central 7 2017

ord\_date <- lg\_ord$Order.Date  
cname <- lg\_ord$Customer.Name  
c\_id <- lg\_ord$Customer.ID  
or\_id <- lg\_ord$Order.ID  
pname <- lg\_ord$Product.Name  
qty <- lg\_ord$Quantity  
s\_amt <- lg\_ord$Sales  
  
cat("Order Date:", format(ord\_date, "%Y-%m-%d"), "\n")

## Order Date: 2017-10-02

cat("Customer Name:", cname, "\n")

## Customer Name: Tamara Chand

cat("Customer ID:", c\_id, "\n")

## Customer ID: TC-20980

cat("Order ID:", or\_id, "\n")

## Order ID: CA-2017-118689

cat("Product Name:", pname, "\n")

## Product Name: Canon imageCLASS 2200 Advanced Copier

cat("Quantity:", qty, "\n")

## Quantity: 5

cat("Sales Amount: $", s\_amt, "\n")

## Sales Amount: $ 17499.95

Answer:

On 2017-10-02, Tamara Chand ( customer Id : TC-20980) made the highest purchase where sale amount is $ 17499.95 (single day highest purchase) where he bought 5 \* Canon imageCLASS 2200 Advanced Copier