El Cerrito Sporting Goods

Ira Sharenow

January 7, 2019

R Markdown

The **goal** of the analysis is to determine if any of the salespersons are performing exceptionally well or exceptionally poorly. In particular, Barb was given more assignments than the others. A major question will be to determine if she did better than the three other salespersons.
For each assignment (row in table), the salesperson was assigned to a region in order to sell a specific type of ball. Prices are fixed.
There is no missing data (so no use of na.rm = TRUE)

Results: The first group of graphs show that Barb did indeed have more assignments than the other salespersons, so she produced more revenue.
Looking further, there was little difference in terms of the median number of balls sold. Barb dominated the positive outliers.

Conclusions: Most of the time there was little difference in the performance of the four salespersons. However, if there was an exceptionally large sale to be made, by an overwhelming margin Barb was the one to make the big sale.

Create the data

```
counts = 1000
set.seed(2019)

orderID = 1:counts
salesDate = sample(seq(as.Date('2016-01-01'), as.Date('2018-12-31'), by="day"), counts)
salesperson = sample(x = c('Al', 'Barb', 'Carol', 'Dave'), size = counts, replace = TRUE)
ball = sample(x = c('baseball', 'basketball', 'football', 'soccer ball'), size = counts, replace = TRUE)
itemsSold = round(rnorm(n = counts, mean = 55, sd = 10),0)


transactionsDF = data.frame(orderID = orderID, salesDate = salesDate,
                            salesperson = as.character(salesperson),
                            ball = ball, stringsAsFactors = FALSE )


transactionsDF$year = as.integer(format(transactionsDF$salesDate, '%Y'))
transactionsDF$region = sample(x = c('North', 'South', 'East', 'West'), size = counts, replace = TRUE)
transactionsDF$itemsSold = itemsSold
transactionsDF$price = ifelse(transactionsDF$ball == 'baseball', 10,
                              ifelse(transactionsDF$ball == 'basketball', 12,
                                     ifelse(transactionsDF$ball == 'football', 15,20)))

transactionsDF$salesTotal = transactionsDF$itemsSold * transactionsDF$price

head(transactionsDF)
```

```
##   orderID  salesDate salesperson        ball year region itemsSold price
## 1       1 2018-04-23        Barb soccer ball 2018   West        75    20
## 2       2 2018-02-19          Al    baseball 2018  North        63    10
```

```
## 3       3 2016-11-27       Carol    football 2016  North        49    15
## 4       4 2017-11-06          Al soccer ball 2017  North        60    20
## 5       5 2016-02-25       Carol    baseball 2016   East        46    10
## 6       6 2016-02-17       Carol soccer ball 2016  North        66    20
##   salesTotal
## 1       1500
## 2        630
## 3        735
## 4       1200
## 5        460
## 6       1320
```

```
# Now give Barbara excellent results and then "forget' what I did as I analyze the data
barbCount = 30
transactionsDF$salesperson[1:barbCount] = rep('Barb', barbCount)
transactionsDF$itemsSold[1:barbCount] = sample(x = 85:100, size = barbCount, replace = TRUE)

# reorganize the columns to that year is next to salesDate
transactionsDF = transactionsDF %>% select(orderID, salesDate, year, salesperson,
                                           ball, region, itemsSold, price, salesTotal)
```

Data Summaries with Charts

Counts Tables and Charts

Simple Counts

```
# Group by salespersons
table(transactionsDF$salesperson)
```

```
##
##    Al  Barb Carol  Dave
##   231   289   247   233
```

```
# Group by region
table(transactionsDF$region)
```

```
##
##  East North South  West
##   250   250   251   249
```

```
# Group by salesperson and region
addmargins(table(transactionsDF$salesperson, transactionsDF$region), c(1,2))
```

```
##
##         East North South West  Sum
##   Al      55    63    66   47  231
##   Barb    68    74    65   82  289
##   Carol   61    58    63   65  247
##   Dave    66    55    57   55  233
##   Sum    250   250   251  249 1000
```

```
# Columns add to 100%
round(prop.table(table(transactionsDF$salesperson, transactionsDF$region), 2),3)
```

```
##
##          East North South  West
##   Al    0.220 0.252 0.263 0.189
##   Barb  0.272 0.296 0.259 0.329
```

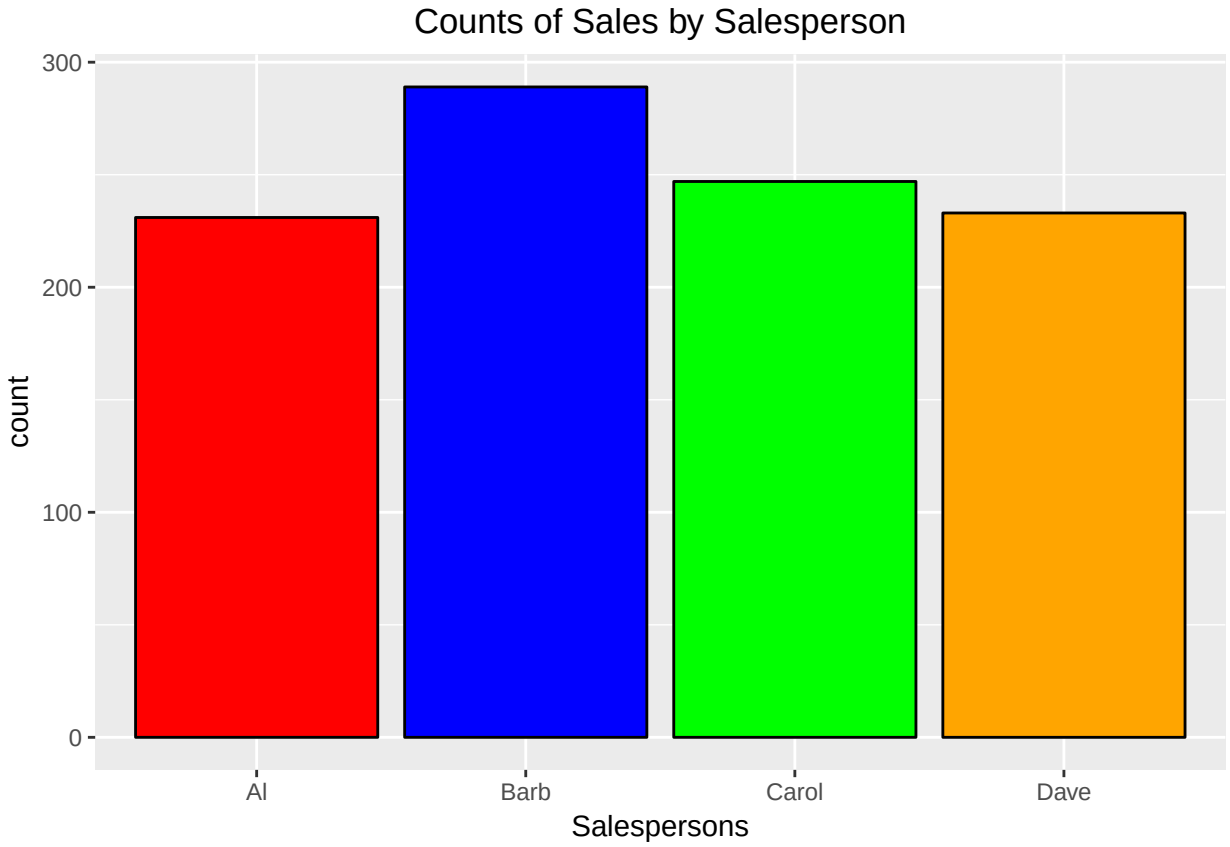
```
##   Carol 0.244 0.232 0.251 0.261
##   Dave  0.264 0.220 0.227 0.221
```

```
# Group by year
table(transactionsDF$year)
```

```
##
## 2016 2017 2018
##  332  333  335
```

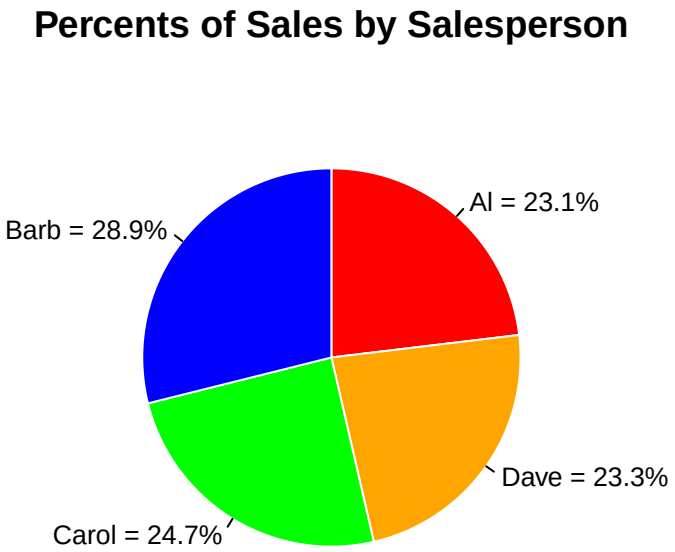
Color palettes defined
Bar charts of counts Some pie charts

```
colorValuesPeople = c("red", "blue", "green", "orange")
colorValuesRegions = c("#E69F00", "#56B4E9", "#009E73", "#F0E442")
colorValuesYears = c('#D55E00', '#009E73', '#56B4E9')


ggplot(transactionsDF, aes(x = factor(salesperson))) + geom_bar(fill = colorValuesPeople, color = 'blac
  ggtitle("Counts of Sales by Salesperson") + xlab("Salespersons")  +
  theme(plot.title = element_text(hjust=0.5))
```



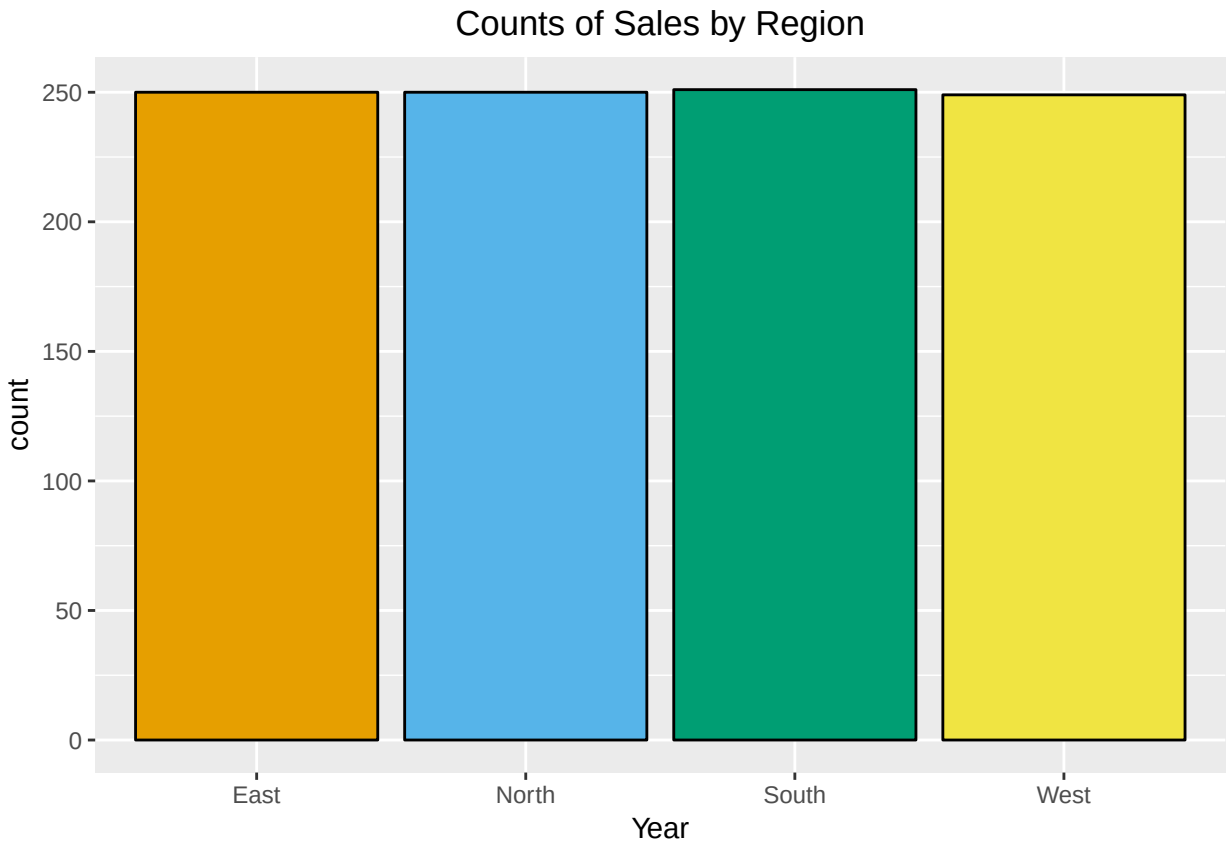

```
# Pie chart
res = transactionsDF %>% group_by(salesperson) %>% summarize(counts = n())
res = data.frame(res)
res = res[order(res[,2]), ]

pielabels <- sprintf("%s = %3.1f%s", res[,1],
```

```
                    100*res[,2]/sum(res[,2]), "%")
pie(res[, 2],
    labels = pielabels,
    clockwise = TRUE,
    col = c('red', 'orange', 'green', 'blue'),
    border = 'white',
    cex = 0.8,
    main = "Percents of Sales by Salesperson")
```

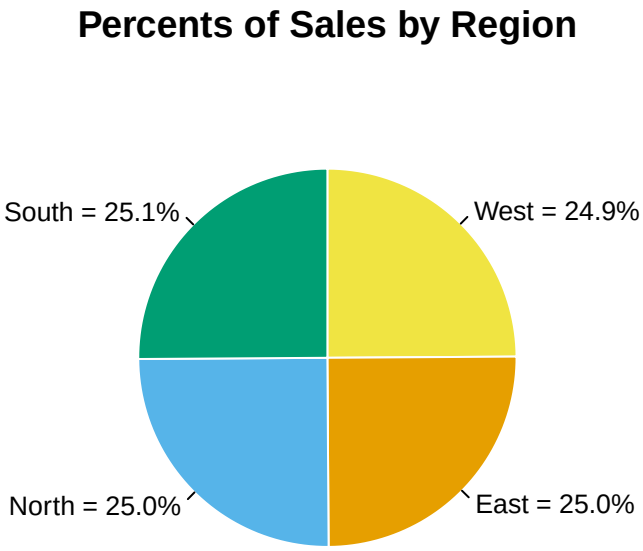


Barchart and pie chart by Region

```
ggplot(transactionsDF, aes(x = factor(region))) + geom_bar(fill = colorValuesRegions, color = 'black') +
  ggtitle("Counts of Sales by Region") + xlab("Year")  +
  theme(plot.title = element_text(hjust=0.5))
```

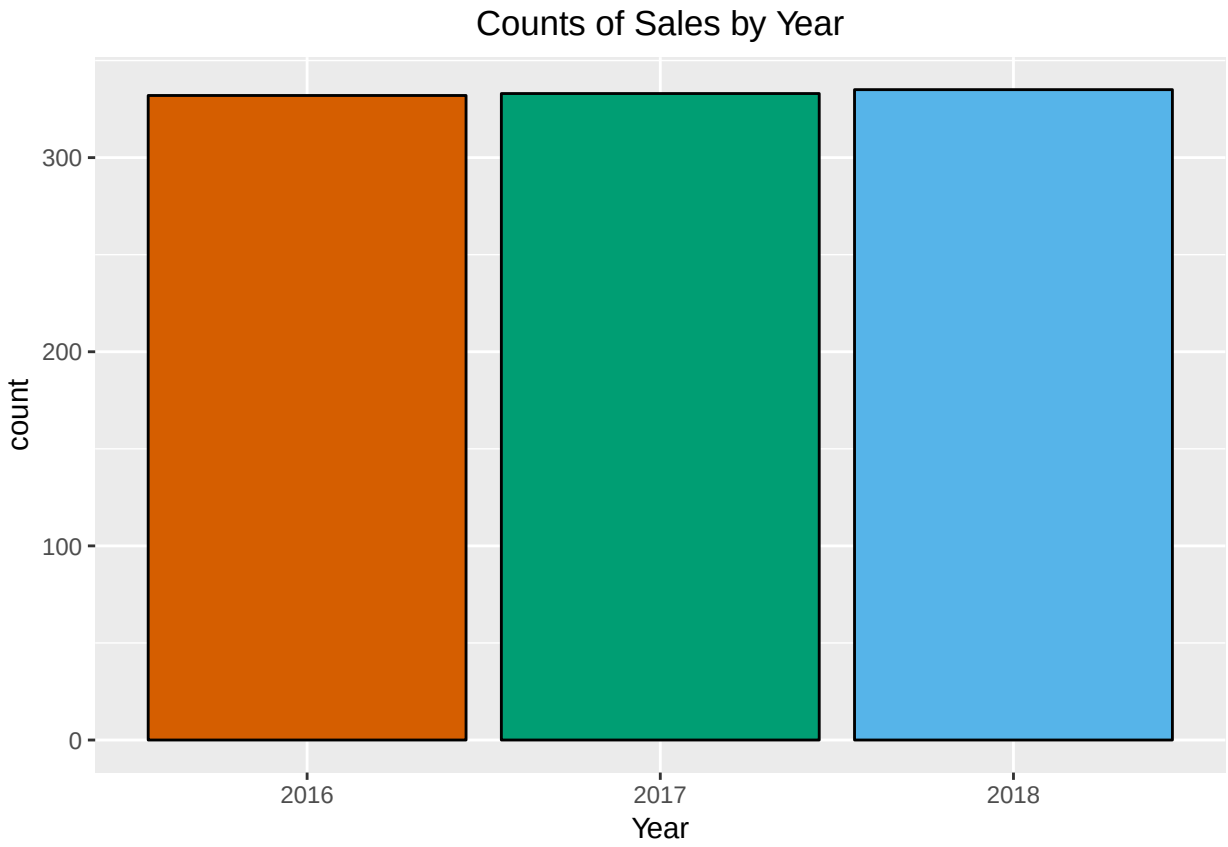



```
# Pie chart
res = transactionsDF %>% group_by(region) %>% summarize(counts = n())
res = data.frame(res)
res = res[order(res[,2]), ]

pielabels <- sprintf("%s = %3.1f%s", res[,1],
                     100*res[,2]/sum(res[,2]), "%")
pie(res[, 2],
    labels = pielabels,
    clockwise = TRUE,
    col = c('#F0E442', '#E69F00', '#56B4E9', '#009E73'),
    border = 'white',
    cex = 0.8,
    main = "Percents of Sales by Region")
```



Barchart and pie chart by year

```
ggplot(transactionsDF, aes(x = factor(year))) + geom_bar(fill = colorValuesYears, color = 'black') +
  ggtitle("Counts of Sales by Year") + xlab("Year")  +
  theme(plot.title = element_text(hjust=0.5))
```



Total sales tables

```
tapply(transactionsDF$salesTotal, transactionsDF$salesperson, sum)
```

```
##     Al   Barb  Carol   Dave
## 178573 229481 192624 182321
```

```
sapply(split(transactionsDF$salesTotal, transactionsDF$salesperson), sum) # alternate base R method
```

```
##     Al   Barb  Carol   Dave
## 178573 229481 192624 182321
```

```
aggregate(transactionsDF$salesTotal, list(transactionsDF$salesperson), sum) # alternate base R method
```

```
##   Group.1      x
## 1      Al 178573
## 2    Barb 229481
## 3   Carol 192624
## 4    Dave 182321
```

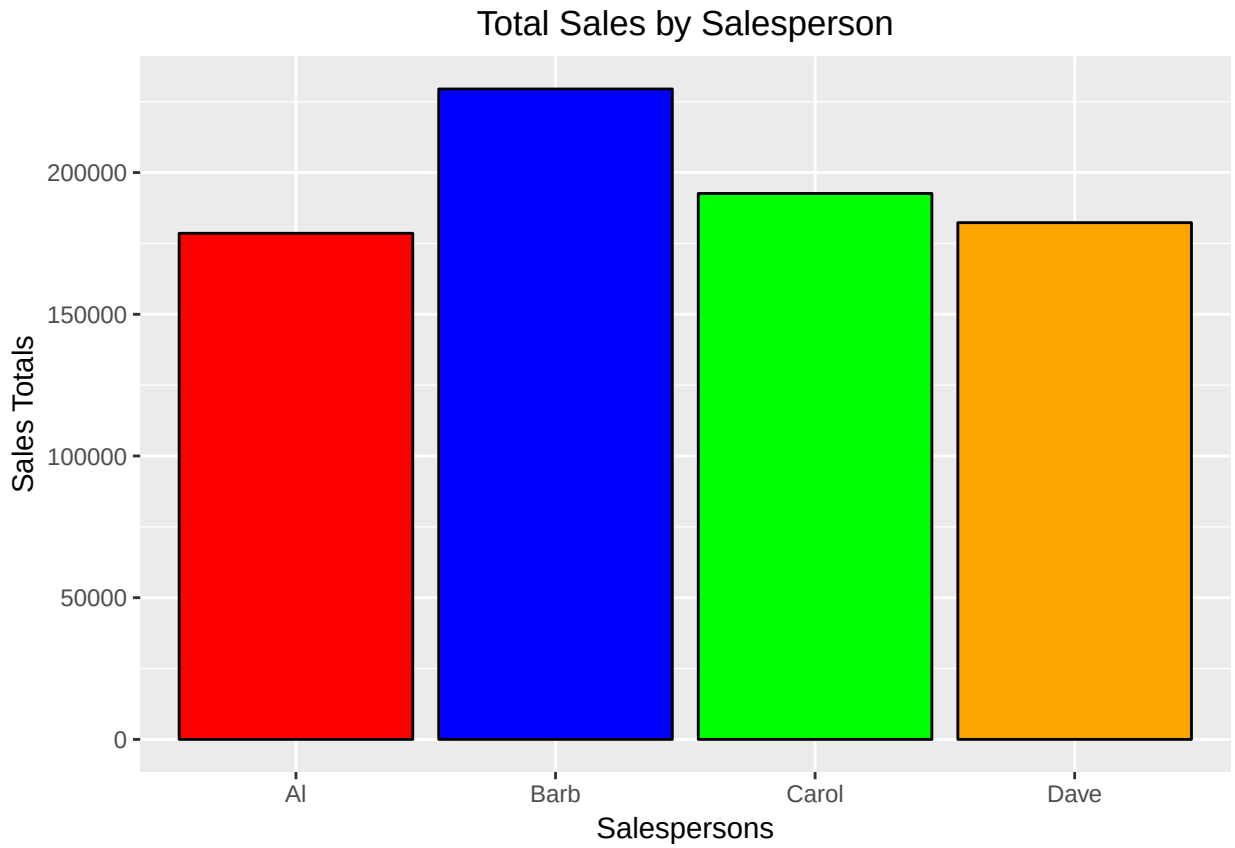
```
transactionsDF %>% group_by(salesperson) %>% summarize(totals = sum(salesTotal)) #dplyr
```

```
## # A tibble: 4 x 2
##   salesperson totals
##   <chr>        <dbl>
## 1 Al          178573
## 2 Barb        229481
## 3 Carol       192624
## 4 Dave        182321
```

```
res = transactionsDF %>% group_by(salesperson) %>% summarize(totals = sum(salesTotal))
res
```

```
## # A tibble: 4 x 2
##   salesperson totals
##   <chr>        <dbl>
## 1 Al          178573
## 2 Barb        229481
## 3 Carol       192624
## 4 Dave        182321
```

```
ggplot(res, aes(x = factor(salesperson), y = totals)) +
  geom_bar(stat = "identity", fill = colorValuesPeople,
           color = 'black') +
  ggtitle("Total Sales by Salesperson") + xlab("Salespersons")  + ylab("Sales Totals") +
  theme(plot.title = element_text(hjust=0.5))
```



Ball is not believed to be important, but just a check.
Ball prices are fixed by the company

```
tapply(transactionsDF$salesTotal, transactionsDF$ball, sum)
```

```
##    baseball  basketball    football soccer ball
##      127260      157584      225315      272840
```

```
transactionsDF %>% group_by(ball) %>% summarize(totals = sum(salesTotal)) #dplyr
```

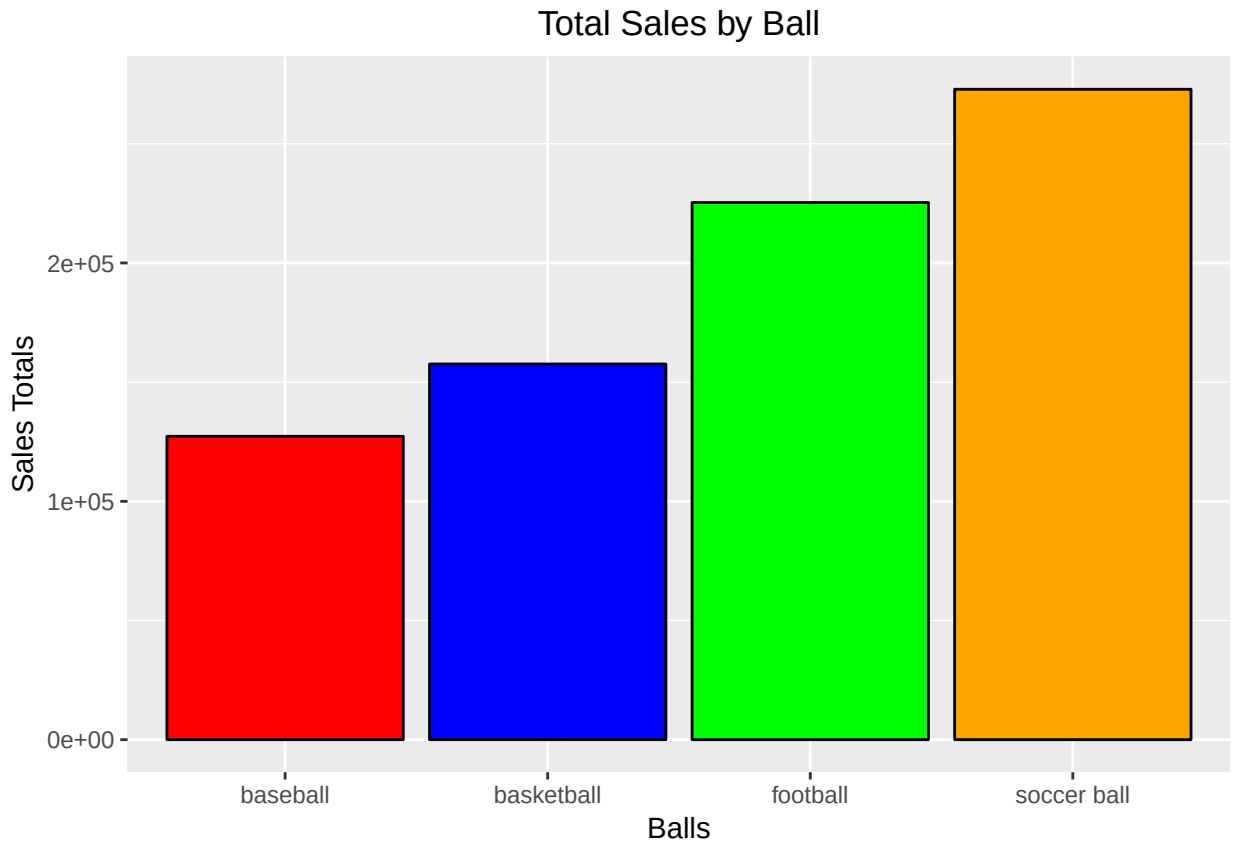
```
## # A tibble: 4 x 2
```

```
##   ball        totals
##   <chr>        <dbl>
## 1 baseball    127260
## 2 basketball  157584
## 3 football    225315
## 4 soccer ball 272840
```

```
res = transactionsDF %>% group_by(ball) %>% summarize(totals = sum(salesTotal)) #dplyr
res
```

```
## # A tibble: 4 x 2
##   ball        totals
##   <chr>        <dbl>
## 1 baseball    127260
## 2 basketball  157584
## 3 football    225315
## 4 soccer ball 272840
```

```
ggplot(res, aes(x = factor(ball), y = totals)) +
  geom_bar(stat = "identity", fill = colorValuesPeople,
           color = 'black') +
  ggtitle("Total Sales by Ball") + xlab("Balls")  + ylab("Sales Totals") +
  theme(plot.title = element_text(hjust=0.5))
```

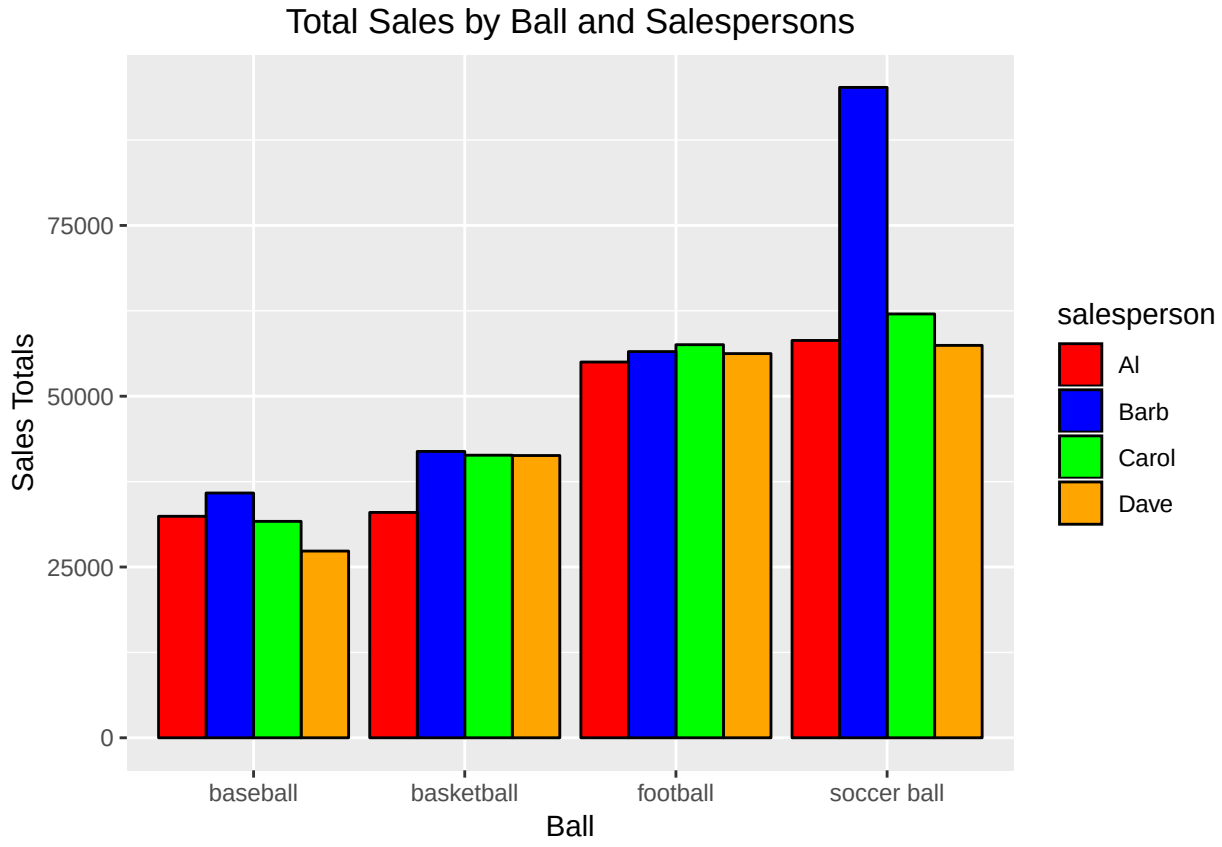



```
res = transactionsDF %>% group_by(ball, salesperson) %>% summarize(totals = sum(salesTotal)) #dplyr
res
```

```
## # A tibble: 16 x 3
```

```
## # Groups:   ball [?]
##    ball        salesperson totals
##    <chr>       <chr>        <dbl>
##  1 baseball    Al           32420
##  2 baseball    Barb         35830
##  3 baseball    Carol        31680
##  4 baseball    Dave         27330
##  5 basketball  Al           32988
##  6 basketball  Barb         41916
##  7 basketball  Carol        41364
##  8 basketball  Dave         41316
##  9 football    Al           55005
## 10 football    Barb         56535
## 11 football    Carol        57540
## 12 football    Dave         56235
## 13 soccer ball Al           58160
## 14 soccer ball Barb         95200
## 15 soccer ball Carol        62040
## 16 soccer ball Dave         57440
```

```
ggplot(res, aes(x = factor(ball), y = totals, fill = salesperson)) +
  geom_bar(stat = "identity", position = "dodge",
           color = 'black') + scale_fill_manual(values = colorValuesPeople) +
  ggtitle("Total Sales by Ball and Salespersons") + xlab("Ball") + ylab("Sales Totals") +
  theme(plot.title = element_text(hjust=0.5))
```



Averages

```
round(tapply(transactionsDF$itemsSold, list(transactionsDF$ball,transactionsDF$salesperson) , mean),0)
```

```
##             Al Barb Carol Dave
## baseball    56   56    54   55
## basketball  53   58    54   55
## football    54   57    57   56
## soccer ball 55   60    54   54
```

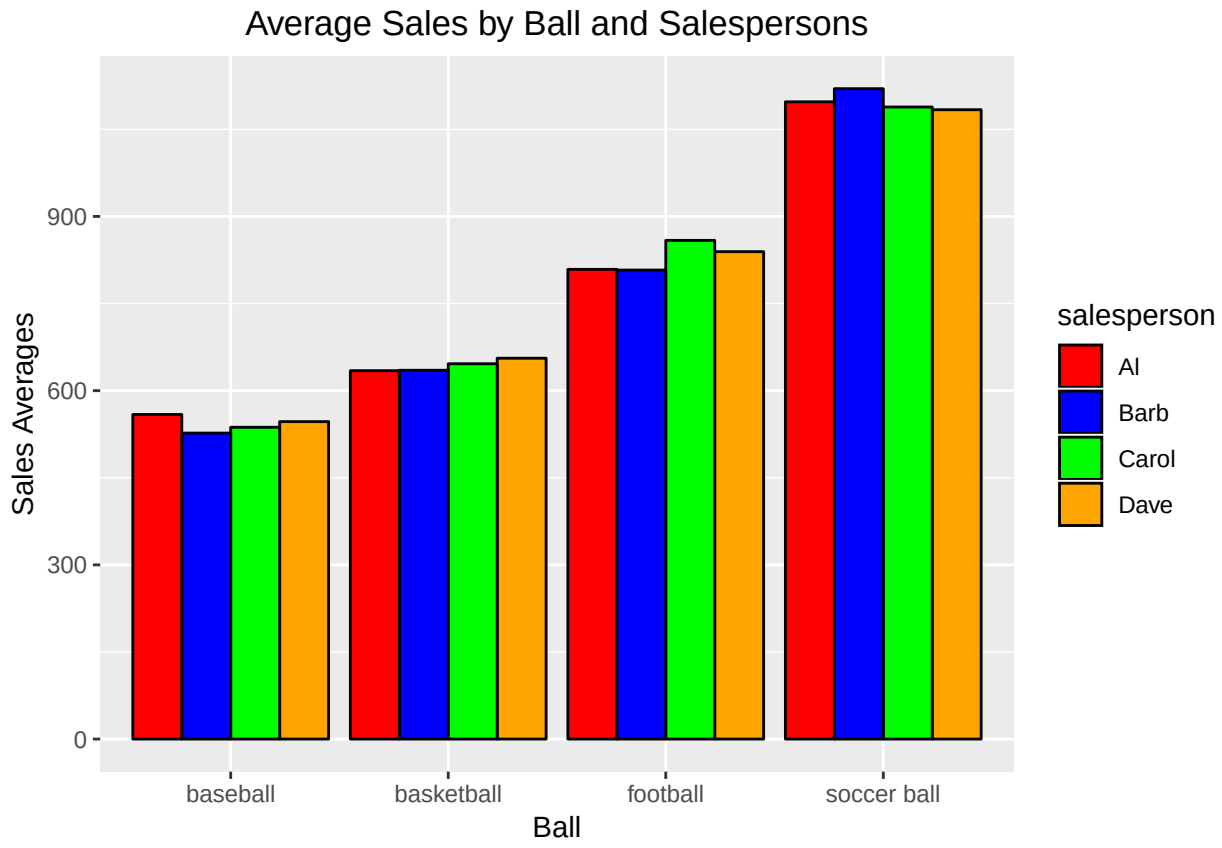
```
transactionsDF %>% group_by(ball, salesperson) %>%   summarize(averages = mean(itemsSold))  %>%
  arrange(ball, desc(averages)) %>% print(n = Inf) # dplyr
```

```
## # A tibble: 16 x 3
## # Groups:   ball [4]
##    ball        salesperson averages
##    <chr>       <chr>          <dbl>
##  1 baseball    Al              55.9
##  2 baseball    Barb            55.8
##  3 baseball    Dave            54.7
##  4 baseball    Carol           53.7
##  5 basketball  Barb            58.3
##  6 basketball  Dave            54.7
##  7 basketball  Carol           53.9
##  8 basketball  Al              52.9
##  9 football    Carol           57.3
## 10 football    Barb            56.6
## 11 football    Dave            56.0
## 12 football    Al              53.9
## 13 soccer ball Barb            59.8
## 14 soccer ball Al              54.9
## 15 soccer ball Carol           54.4
## 16 soccer ball Dave            54.2
```

```
res = transactionsDF %>% group_by(ball, salesperson) %>% summarize(averages = mean(salesTotal)) #dplyr
res
```

```
## # A tibble: 16 x 3
## # Groups:   ball [?]
##    ball        salesperson averages
##    <chr>       <chr>          <dbl>
##  1 baseball    Al              559.
##  2 baseball    Barb            527.
##  3 baseball    Carol           537.
##  4 baseball    Dave            547.
##  5 basketball  Al              634.
##  6 basketball  Barb            635.
##  7 basketball  Carol           646.
##  8 basketball  Dave            656.
##  9 football    Al              809.
## 10 football    Barb            808.
## 11 football    Carol           859.
## 12 football    Dave            839.
## 13 soccer ball Al             1097.
## 14 soccer ball Barb           1120
## 15 soccer ball Carol          1088.
## 16 soccer ball Dave           1084.
```

```
ggplot(res, aes(x = factor(ball), y = averages, fill = salesperson)) +
  geom_bar(stat = "identity", position = "dodge",
           color = 'black') + scale_fill_manual(values = colorValuesPeople) +
  ggtitle("Average Sales by Ball and Salespersons") + xlab("Ball") + ylab("Sales Averages") +
  theme(plot.title = element_text(hjust=0.5))
```




```
# DO THIS GRAPH
transactionsDF %>% group_by(ball, region, salesperson) %>%
  summarize(averages = mean(itemsSold)) %>%
  arrange(ball, region, desc(averages)) %>% print(n = Inf) # dplyr
```

```
## # A tibble: 64 x 4
## # Groups:   ball, region [16]
##    ball        region salesperson averages
##    <chr>       <chr>  <chr>          <dbl>
##  1 baseball    East   Barb            66.8
##  2 baseball    East   Carol           58.5
##  3 baseball    East   Al              57.2
##  4 baseball    East   Dave            56.7
##  5 baseball    North  Al              56.8
##  6 baseball    North  Carol           54.4
##  7 baseball    North  Barb            53.4
##  8 baseball    North  Dave            50.2
##  9 baseball    South  Barb            56.1
## 10 baseball    South  Al              55.2
## 11 baseball    South  Dave            53
```

```
## 12 baseball    South  Carol          50.7
## 13 baseball    West   Dave           56.2
## 14 baseball    West   Carol          52.1
## 15 baseball    West   Al             51.8
## 16 baseball    West   Barb           50.4
## 17 basketball  East   Barb           59.1
## 18 basketball  East   Dave           56.3
## 19 basketball  East   Carol          52.2
## 20 basketball  East   Al             51.2
## 21 basketball  North  Dave           57.2
## 22 basketball  North  Al             55.2
## 23 basketball  North  Barb           53.8
## 24 basketball  North  Carol          52.8
## 25 basketball  South  Barb           56.9
## 26 basketball  South  Carol          55.6
## 27 basketball  South  Dave           54
## 28 basketball  South  Al             52.4
## 29 basketball  West   Barb           62.5
## 30 basketball  West   Carol          54.8
## 31 basketball  West   Al             51.4
## 32 basketball  West   Dave           50.8
## 33 football    East   Barb           60.4
## 34 football    East   Al             59.7
## 35 football    East   Dave           57.5
## 36 football    East   Carol          55.6
## 37 football    North  Barb           60.0
## 38 football    North  Carol          58.7
## 39 football    North  Dave           56.2
## 40 football    North  Al             54.2
## 41 football    South  Carol          57.9
## 42 football    South  Dave           56.3
## 43 football    South  Al             52.2
## 44 football    South  Barb           49.7
## 45 football    West   Carol          57.4
## 46 football    West   Barb           54.8
## 47 football    West   Dave           53.9
## 48 football    West   Al             51.9
## 49 soccer ball East   Barb           61.3
## 50 soccer ball East   Carol          57.1
## 51 soccer ball East   Al             53.9
## 52 soccer ball East   Dave           53.8
## 53 soccer ball North  Barb           59.5
## 54 soccer ball North  Dave           56.2
## 55 soccer ball North  Carol          52.4
## 56 soccer ball North  Al             51.6
## 57 soccer ball South  Barb           60.3
## 58 soccer ball South  Carol          55.4
## 59 soccer ball South  Al             54.7
## 60 soccer ball South  Dave           49.7
## 61 soccer ball West   Al             59.2
## 62 soccer ball West   Barb           58.1
## 63 soccer ball West   Dave           57.7
## 64 soccer ball West   Carol          53.7
```

Medians

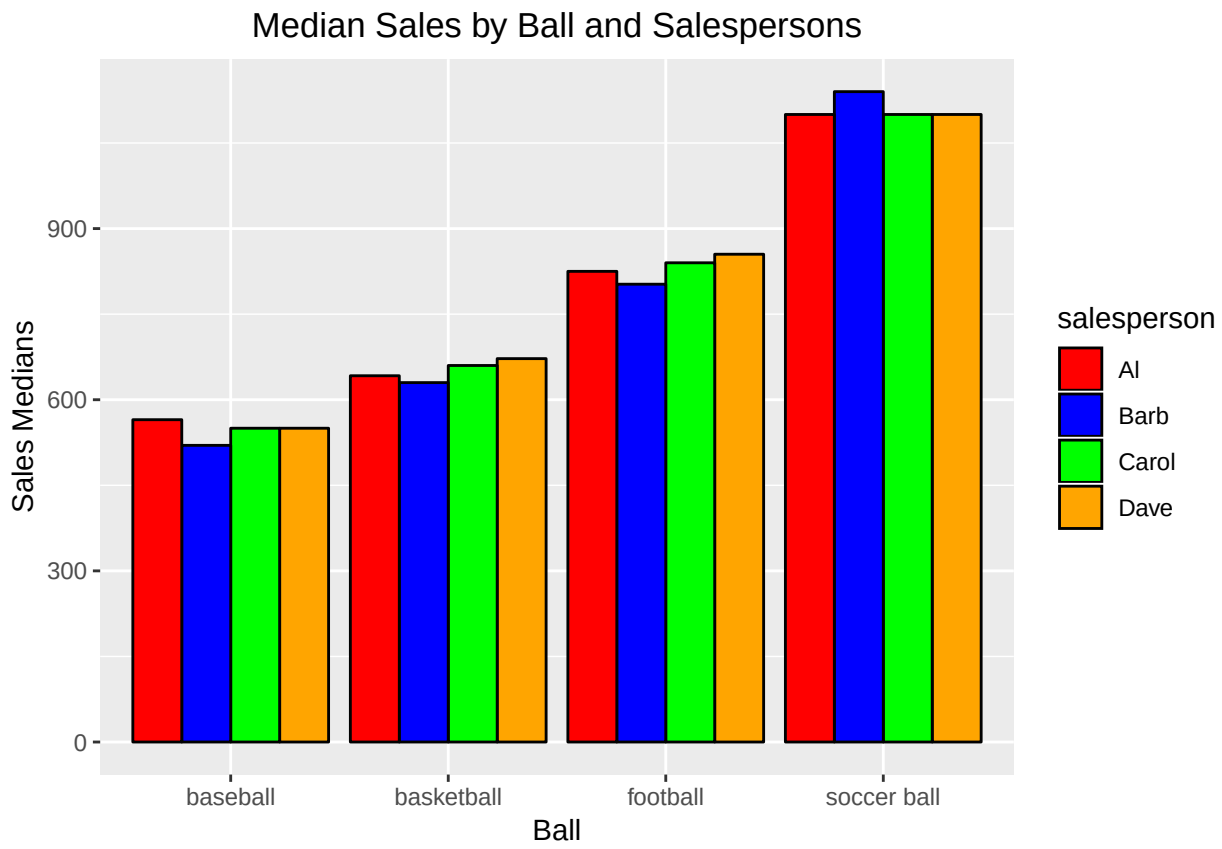
```
transactionsDF %>% group_by(ball, salesperson) %>%   summarize(medians = median(itemsSold))  %>%
  arrange(ball, desc(medians)) %>% print(n = Inf) # dplyr
```

```
## # A tibble: 16 x 3
## # Groups:   ball [4]
##    ball        salesperson medians
##    <chr>       <chr>         <dbl>
##  1 baseball    Al             56.5
##  2 baseball    Carol          55
##  3 baseball    Dave           55
##  4 baseball    Barb           53
##  5 basketball  Dave           56
##  6 basketball  Barb           55
##  7 basketball  Carol          55
##  8 basketball  Al             53.5
##  9 football    Dave           57
## 10 football    Carol          56
## 11 football    Al             55
## 12 football    Barb           55
## 13 soccer ball Barb           57
## 14 soccer ball Al             55
## 15 soccer ball Carol          55
## 16 soccer ball Dave           55
```

```
res = transactionsDF %>% group_by(ball, salesperson) %>% summarize(medians = median(salesTotal)) #dplyr
res
```

```
## # A tibble: 16 x 3
## # Groups:   ball [?]
##    ball        salesperson medians
##    <chr>       <chr>         <dbl>
##  1 baseball    Al             565
##  2 baseball    Barb           520
##  3 baseball    Carol          550
##  4 baseball    Dave           550
##  5 basketball  Al             642
##  6 basketball  Barb           630
##  7 basketball  Carol          660
##  8 basketball  Dave           672
##  9 football    Al             825
## 10 football    Barb           802.
## 11 football    Carol          840
## 12 football    Dave           855
## 13 soccer ball Al            1100
## 14 soccer ball Barb          1140
## 15 soccer ball Carol         1100
## 16 soccer ball Dave          1100
```

```
ggplot(res, aes(x = factor(ball), y = medians, fill = salesperson)) +
  geom_bar(stat = "identity", position = "dodge",
           color = 'black') + scale_fill_manual(values = colorValuesPeople) +
  ggtitle("Median Sales by Ball and Salespersons") + xlab("Ball") + ylab("Sales Medians") +
  theme(plot.title = element_text(hjust=0.5))
```




```
transactionsDF %>% group_by(ball, region, salesperson) %>%
  summarize(medians = median(itemsSold)) %>%
  arrange(ball, region, desc(medians)) %>% print(n = Inf) # dplyr
```

```
## # A tibble: 64 x 4
## # Groups:   ball, region [16]
##    ball       region salesperson medians
##    <chr>      <chr>  <chr>         <dbl>
##  1 baseball   East   Barb           64
##  2 baseball   East   Dave           60
##  3 baseball   East   Al             58
##  4 baseball   East   Carol          57
##  5 baseball   North  Al             58
##  6 baseball   North  Carol          57
##  7 baseball   North  Dave           53.5
##  8 baseball   North  Barb           51.5
##  9 baseball   South  Al             55
## 10 baseball   South  Dave           53.5
## 11 baseball   South  Barb           53
## 12 baseball   South  Carol          49
## 13 baseball   West   Dave           57
## 14 baseball   West   Carol          53
## 15 baseball   West   Al             51
## 16 baseball   West   Barb           48
## 17 basketball East   Dave           57.5
## 18 basketball East   Al             53
```

```
## 19 basketball  East   Barb           53
## 20 basketball  East   Carol          52.5
## 21 basketball  North  Al             56.5
## 22 basketball  North  Dave           56
## 23 basketball  North  Carol          55
## 24 basketball  North  Barb           49
## 25 basketball  South  Barb           56
## 26 basketball  South  Carol          56
## 27 basketball  South  Dave           55
## 28 basketball  South  Al             51.5
## 29 basketball  West   Barb           56
## 30 basketball  West   Carol          55
## 31 basketball  West   Al             54
## 32 basketball  West   Dave           53
## 33 football    East   Al             61
## 34 football    East   Barb           58
## 35 football    East   Dave           57
## 36 football    East   Carol          54.5
## 37 football    North  Dave           58.5
## 38 football    North  Barb           58
## 39 football    North  Al             56.5
## 40 football    North  Carol          55
## 41 football    South  Carol          59
## 42 football    South  Dave           57
## 43 football    South  Al             51.5
## 44 football    South  Barb           50
## 45 football    West   Carol          57
## 46 football    West   Dave           54
## 47 football    West   Al             52
## 48 football    West   Barb           52
## 49 soccer ball East   Carol          57.5
## 50 soccer ball East   Barb           56.5
## 51 soccer ball East   Dave           53.5
## 52 soccer ball East   Al             52
## 53 soccer ball North  Carol          57
## 54 soccer ball North  Barb           56
## 55 soccer ball North  Dave           56
## 56 soccer ball North  Al             53.5
## 57 soccer ball South  Barb           57.5
## 58 soccer ball South  Carol          55
## 59 soccer ball South  Al             54.5
## 60 soccer ball South  Dave           46
## 61 soccer ball West   Dave           58.5
## 62 soccer ball West   Al             57.5
## 63 soccer ball West   Barb           57.5
## 64 soccer ball West   Carol          53
```

```
lm.fit1 = lm(itemsSold ~ salesperson , data = transactionsDF)
summary(lm.fit1)
```

```
##
## Call:
## lm(formula = itemsSold ~ salesperson, data = transactionsDF)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -39.716  -7.870  -0.557   6.284  42.284
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       54.3983     0.7795  69.790  < 2e-16 ***
## salespersonBarb    3.3180     1.0455   3.173  0.00155 **
## salespersonCarol   0.4722     1.0843   0.435  0.66332
## salespersonDave    0.5245     1.0999   0.477  0.63359
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.85 on 996 degrees of freedom
## Multiple R-squared:  0.01318,    Adjusted R-squared:  0.0102
## F-statistic: 4.433 on 3 and 996 DF,  p-value: 0.004192
```

```
lm.fit2 = lm(itemsSold ~ salesperson + ball, data = transactionsDF)
summary(lm.fit2)
```

```
##
## Call:
## lm(formula = itemsSold ~ salesperson + ball, data = transactionsDF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.135  -7.883  -0.329   6.585  42.819
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       53.84709    1.02353  52.609   <2e-16 ***
## salespersonBarb    3.29846    1.04803   3.147   0.0017 **
## salespersonCarol   0.49237    1.08541   0.454   0.6502
## salespersonDave    0.53177    1.10148   0.483   0.6294
## ballbasketball     0.03572    1.08330   0.033   0.9737
## ballfootball       0.98927    1.05624   0.937   0.3492
## ballsoccer ball    1.09800    1.08016   1.017   0.3096
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.85 on 993 degrees of freedom
## Multiple R-squared:  0.01503,    Adjusted R-squared:  0.009078
## F-statistic: 2.525 on 6 and 993 DF,  p-value: 0.01974
```

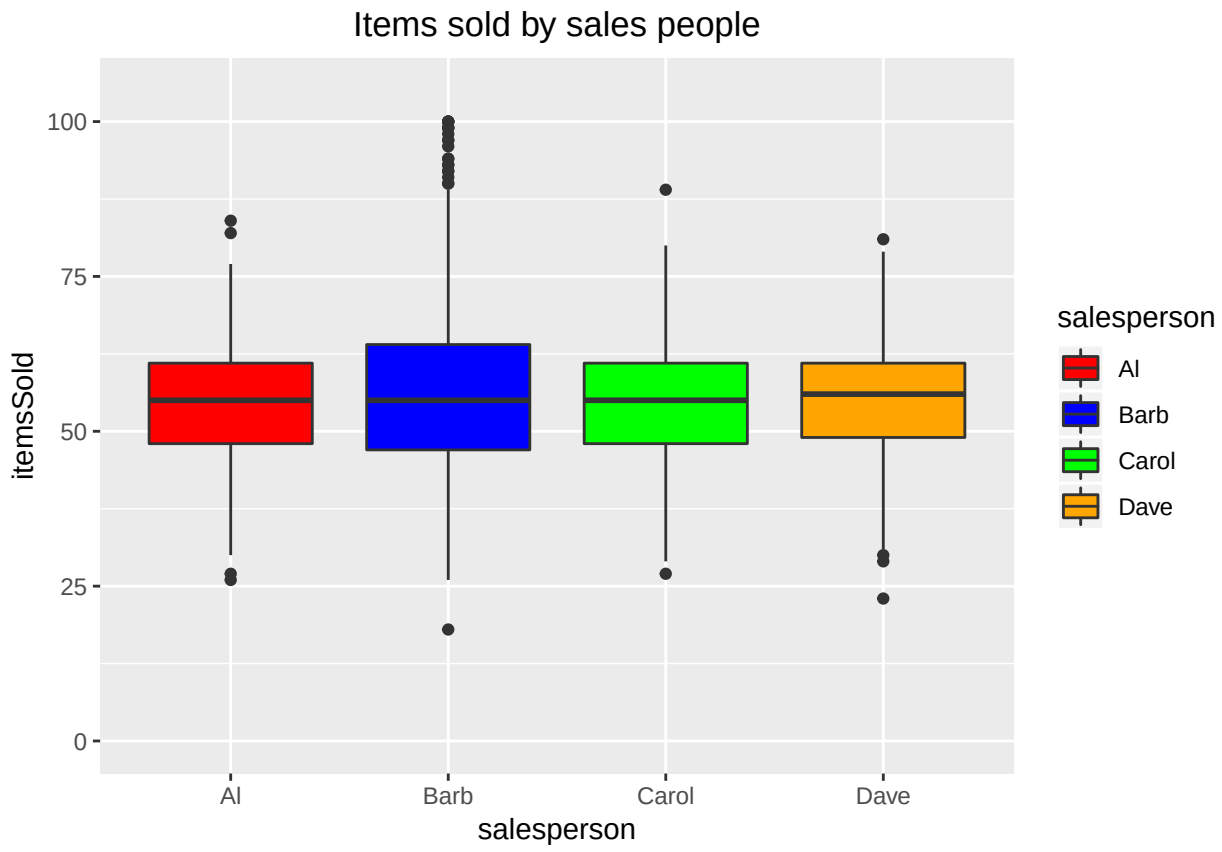
```
lm.fit3 = lm(itemsSold ~ salesperson + ball + region, data = transactionsDF)
summary(lm.fit3)
```

```
##
## Call:
## lm(formula = itemsSold ~ salesperson + ball + region, data = transactionsDF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.152  -7.608  -0.221   6.622  41.799
##
## Coefficients:
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       55.6484     1.2086  46.044  < 2e-16 ***
## salespersonBarb    3.3003     1.0478   3.150  0.00168 **
## salespersonCarol   0.4842     1.0838   0.447  0.65513
## salespersonDave    0.4239     1.0999   0.385  0.70002
## ballbasketball     0.1456     1.0813   0.135  0.89293
## ballfootball       1.0727     1.0536   1.018  0.30886
## ballsoccer ball    1.1353     1.0776   1.054  0.29236
## regionNorth       -1.8930     1.0584  -1.788  0.07400 .
## regionSouth       -2.8697     1.0580  -2.712  0.00679 **
## regionWest        -2.5703     1.0600  -2.425  0.01550 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.82 on 990 degrees of freedom
## Multiple R-squared:  0.0238, Adjusted R-squared:  0.01493
## F-statistic: 2.682 on 9 and 990 DF,  p-value: 0.004411
```

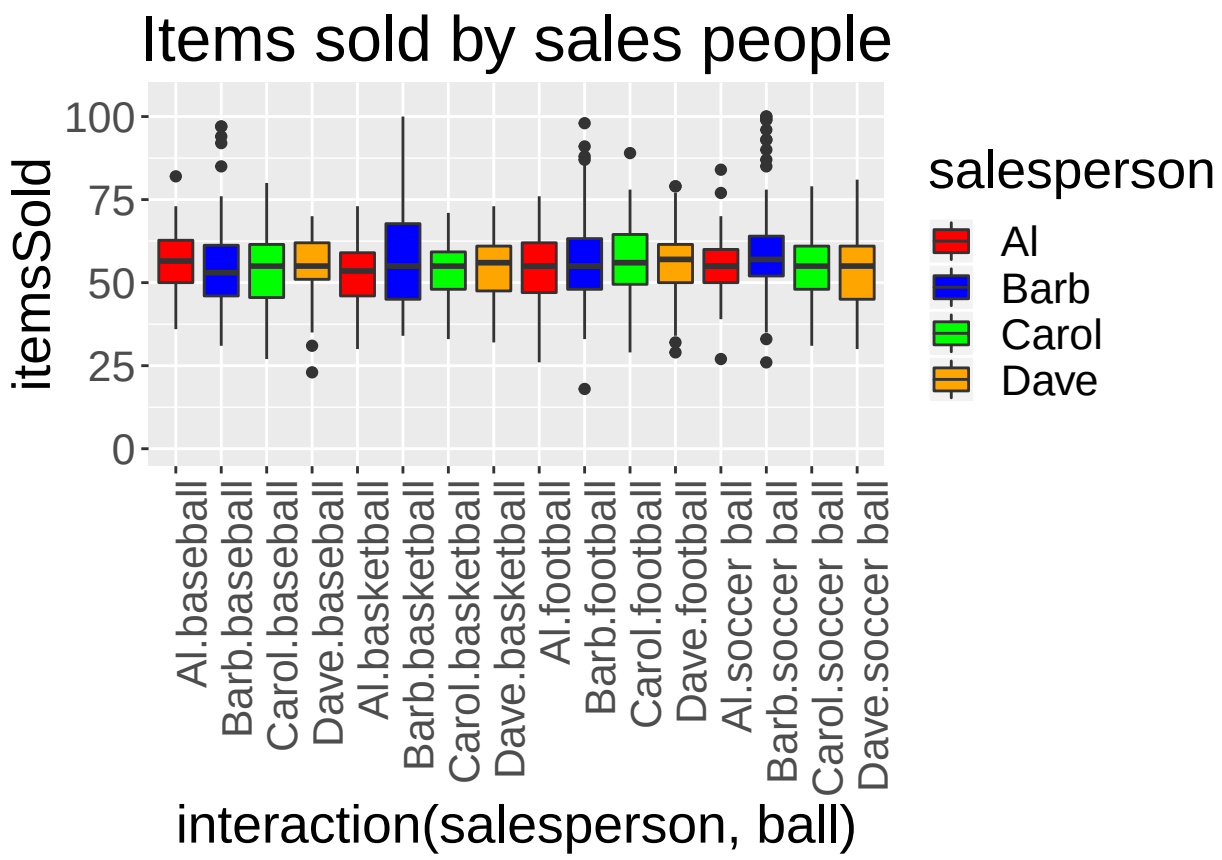
Boxplots of Counts

```
ggplot(transactionsDF, aes(x = salesperson, y = itemsSold)) +
  geom_boxplot(aes(fill=salesperson)) +
  scale_fill_manual(values = colorValuesPeople) + ylim( 0, max(transactionsDF$itemsSold) + 5) +
  ggtitle("Items sold by sales people") + theme(plot.title = element_text(hjust=0.5))
```




```
ggplot(transactionsDF, aes(x = interaction(salesperson, ball), y = itemsSold)) +   geom_boxplot(aes(fil
  scale_fill_manual(values = colorValuesPeople) + ylim( 0, max(transactionsDF$itemsSold) + 5) +
```

```
  ggtitle("Items sold by sales people") + theme(plot.title = element_text(hjust=0.5)) +
  theme(text = element_text(size=20),
        axis.text.x = element_text(angle=90, hjust=1))
```



Scatterplot with jittering

```
p = ggplot(transactionsDF, aes(x = salesperson, y = itemsSold, color = salesperson, shape = ball)) +
  geom_point(alpha = 0.01) + geom_hline(yintercept = 80) +
  geom_jitter() + scale_color_manual(values = colorValuesPeople) +
  ggtitle("Items sold by salesperson") + theme(plot.title = element_text(hjust=0.5))

p + annotate("text", 3.3, y = 95, label = "Barb gets the positive outliers")
```


Items sold by salesperson
Barb gets the positive outliers
itemsSold
100
75
50
25
Al
Barb
Carol
Dave
salesperson
ball
baseball
basketball
football
soccer ball
salesperson
Al
Barb
Carol
Dave