

For Justin

Abe Kazemzadeh

5/3/2022

```
df <- read_excel("Blacksheepgastronomy Sales data 2019-YTD.xlsm")
```

### average sales by billing\_region

unique billing regions

```
unique(df$billing_region)
```

```
## [1] NA "Michigan" "Indiana" "Wisconsin"
## [5] "Pennsylvania" "Missouri" "Ohio" "Illinois"
## [9] "Arkansas" "New Jersey" "Maryland" "Texas"
## [13] "Oklahoma" "Minnesota" "North Carolina" "Tennessee"
## [17] "Kentucky" "Florida" "Montana" "Kansas"
## [21] "Massachusetts" "Virginia" "Georgia" "New York"
## [25] "California" "Puerto Rico" "Washington" "South Carolina"
## [29] "Nebraska" "Colorado" "Alaska" "South Dakota"
## [33] "New Hampshire" "Iowa" "Alabama" "Arizona"
## [37] "Hawaii" "New Mexico" "Alberta" "Nevada"
## [41] "Mississippi"
```

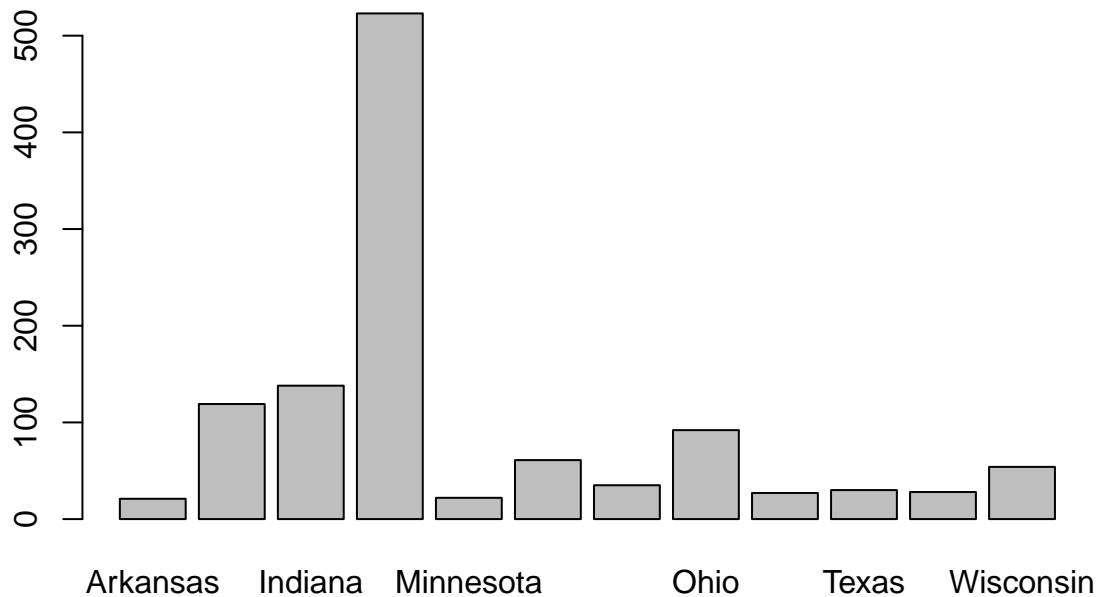
counts of transactions in each region

```
table(df$billing_region)
```

```
##
## Alabama Alaska Alberta Arizona Arkansas
## 7 5 1 3 21
## California Colorado Florida Georgia Hawaii
## 11 6 14 7 1
## Illinois Indiana Iowa Kansas Kentucky
## 119 138 2 7 7
## Maryland Massachusetts Michigan Minnesota Mississippi
## 4 4 523 22 4
## Missouri Montana Nebraska Nevada New Hampshire
## 61 16 3 1 6
## New Jersey New Mexico New York North Carolina Ohio
## 7 3 5 35 92
## Oklahoma Pennsylvania Puerto Rico South Carolina South Dakota
## 6 18 1 7 4
## Tennessee Texas Virginia Washington Wisconsin
## 27 30 28 5 54
```

table of billing regions with greater than 20 transactions

```
tab <- table(df$billing_region)
barplot(subset(tab, tab > 20))
```



some tidy functionality see r4ds

get the mean, sd, and count of sales in each region.

```
df %>%
  filter(units_per_transaction != 0) %>%
  group_by(billing_region) %>%
  filter(n()>10) %>%
  summarize(mean_sales=mean(total_sales), sd_sales=sd(total_sales), n=n())
```

```
## # A tibble: 14 x 4
##   billing_region mean_sales sd_sales    n
##   <chr>          <dbl>    <dbl> <int>
## 1 Arkansas      25.7      24.4    11
## 2 Illinois      35.4      29.3    61
## 3 Indiana       31.4      22.5    76
## 4 Michigan      30.3      37.7   296
## 5 Minnesota      49.9     111.    14
## 6 Missouri      35.8      37.6    34
## 7 Montana       22.2      18.4    11
## 8 North Carolina 36.8      33.1    20
```

```
## 9 Ohio                29.7    26.0    51
## 10 Pennsylvania       39.7    31.0    11
## 11 Tennessee          44.6    48.0    15
## 12 Texas              71.3   151.    15
## 13 Virginia           53.4    53.0    18
## 14 Wisconsin          26.4    24.5    30
```

a knitr table

```
library(knitr)
df %>%
  filter(units_per_transaction != 0) %>%
  group_by(billing_region) %>%
  filter(n()>10) %>%
  summarize(mean_sales=mean(total_sales), sd_sales=sd(total_sales), n=n()) %>%
  kable()
```

billing_region	mean_sales	sd_sales	n
Arkansas	25.68182	24.41519	11
Illinois	35.38852	29.34850	61
Indiana	31.35395	22.54538	76
Michigan	30.25456	37.73287	296
Minnesota	49.90714	111.28217	14
Missouri	35.77794	37.58994	34
Montana	22.23182	18.35323	11
North Carolina	36.79000	33.05203	20
Ohio	29.71765	25.96757	51
Pennsylvania	39.67273	30.97787	11
Tennessee	44.61333	48.02872	15
Texas	71.27333	150.66434	15
Virginia	53.37778	52.99758	18
Wisconsin	26.36333	24.47619	30

## statistic tests

anova

```
anova(aov(total_sales ~ billing_region, data=df))
```

```
## Analysis of Variance Table
##
## Response: total_sales
##          Df Sum Sq Mean Sq F value    Pr(>F)
## billing_region   39  175187   4492.0    3.4222 1.599e-11 ***
## Residuals      1275  1673574   1312.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```