ggplot2 - Bar Charts

Vincent La

May 25, 2017

Bar Charts

This is a short introduction and reference guide to bar charts in ggplot2 which assumes some experience in R but very little using ggplot2.

Visualizing Sasquatch Sightings

The dataset we will be working with is fictious, just like Sasquatch. Suppose we had a table of months and the number of Sasquatch sightings per month.

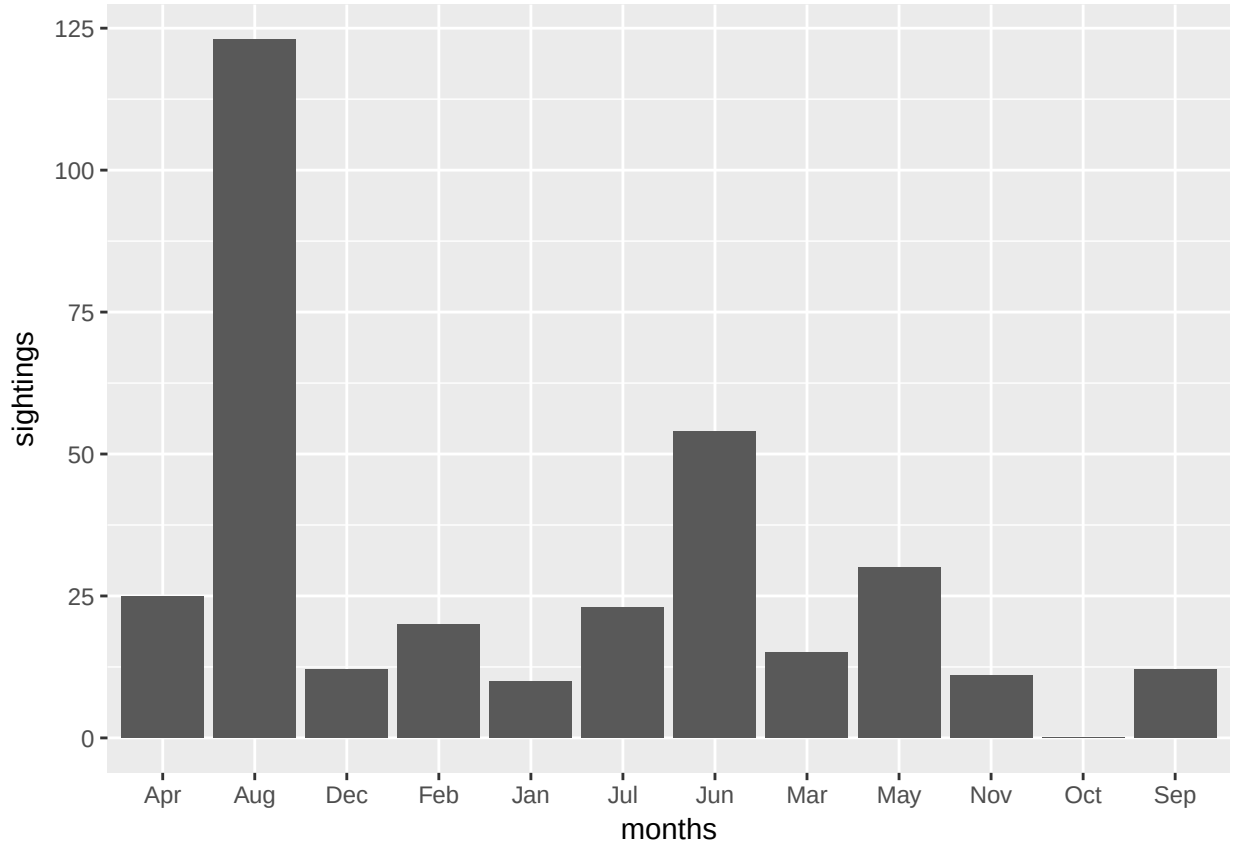
```
sas_dat <- data.frame(
  months = c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep",
             "Oct", "Nov", "Dec"),
  sightings = c(10, 20, 15, 25, 30, 54, 23, 123, 12, 0, 11, 12)
)

sas_dat
```

```
##    months sightings
## 1     Jan        10
## 2     Feb        20
## 3     Mar        15
## 4     Apr        25
## 5     May        30
## 6     Jun        54
## 7     Jul        23
## 8     Aug       123
## 9     Sep        12
## 10    Oct         0
## 11    Nov        11
## 12    Dec        12
```

Using pre-calculated counts with `stat='identity'`

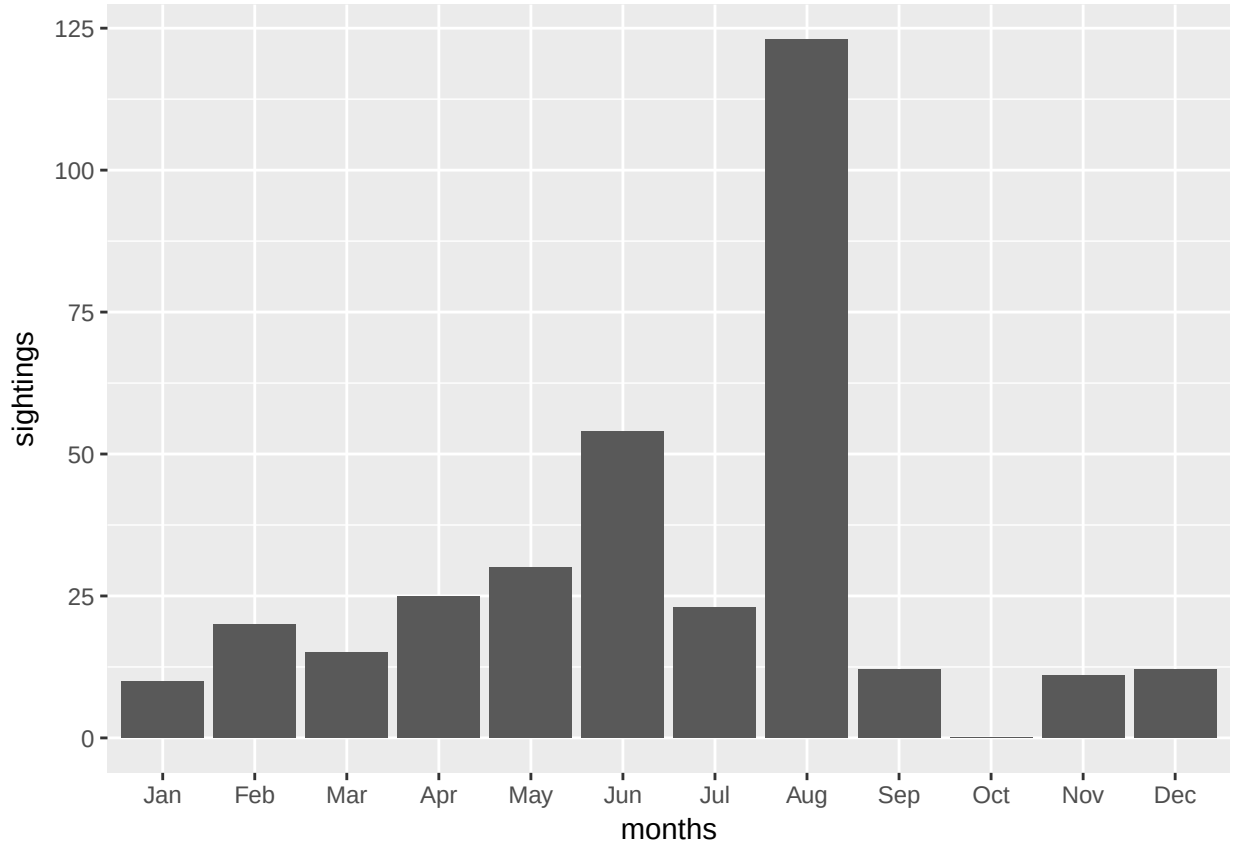
The `stat='identity'` option tells ggplot2 that the y-values should be left as is. On the other hand, ggplot's default behavior is to assume that you are giving it data that needs to be counted.

```
sas_plot <- ggplot(data=sas_dat, aes(x=months, y=sightings)) +
  geom_bar(stat="identity")

sas_plot
```



Re-Ordering

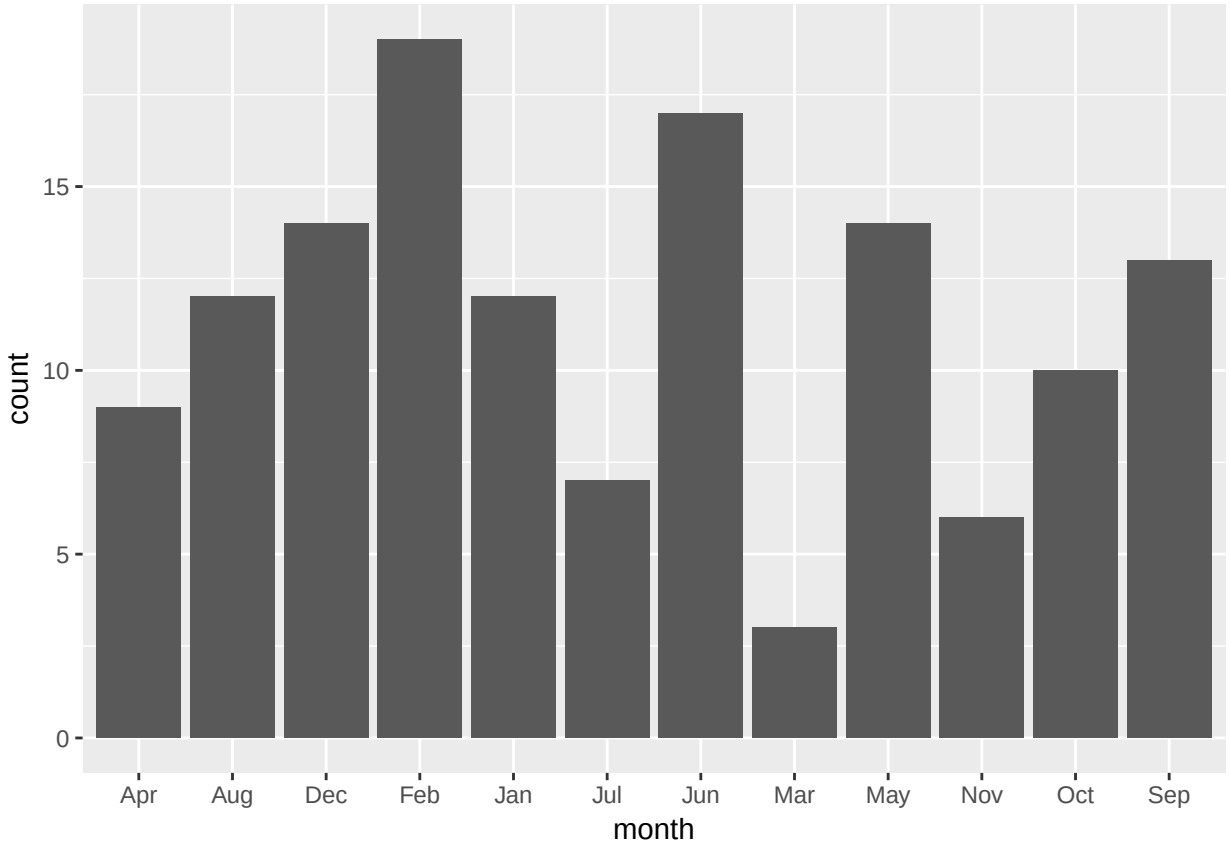
```
month_names <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep",
                 "Oct", "Nov", "Dec")

sas_dat <- data.frame(
  months = factor(month_names, levels=month_names),              # Use levels argument
  sightings = c(10, 20, 15, 25, 30, 54, 23, 123, 12, 0, 11, 12)
)

ggplot(data=sas_dat, aes(x=months, y=sightings)) +
  geom_bar(stat="identity")
```



Letting ggplot2 count for you with `stat='bin'`

```
month <- c()

set.seed(420)

for (m in c("Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep",
            "Oct", "Nov", "Dec")) {
  month <- append(month, rep(m, runif(1, min=0, max=20)))
}
```

```
dat <- data.frame(
 month = month
)

my_plot <- ggplot(data=dat, aes(x=month)) + geom_bar()
my_plot
```



Fancying Up our Plots

Renaming the X and Y Axes

Because ggplot2 names the axes after our data frame's column names, directly changing them will also change the axis labels. Although this is a rather inefficient way of renaming the columns, it is included because it demonstrates use of the sometimes useful built-in `names()` function.

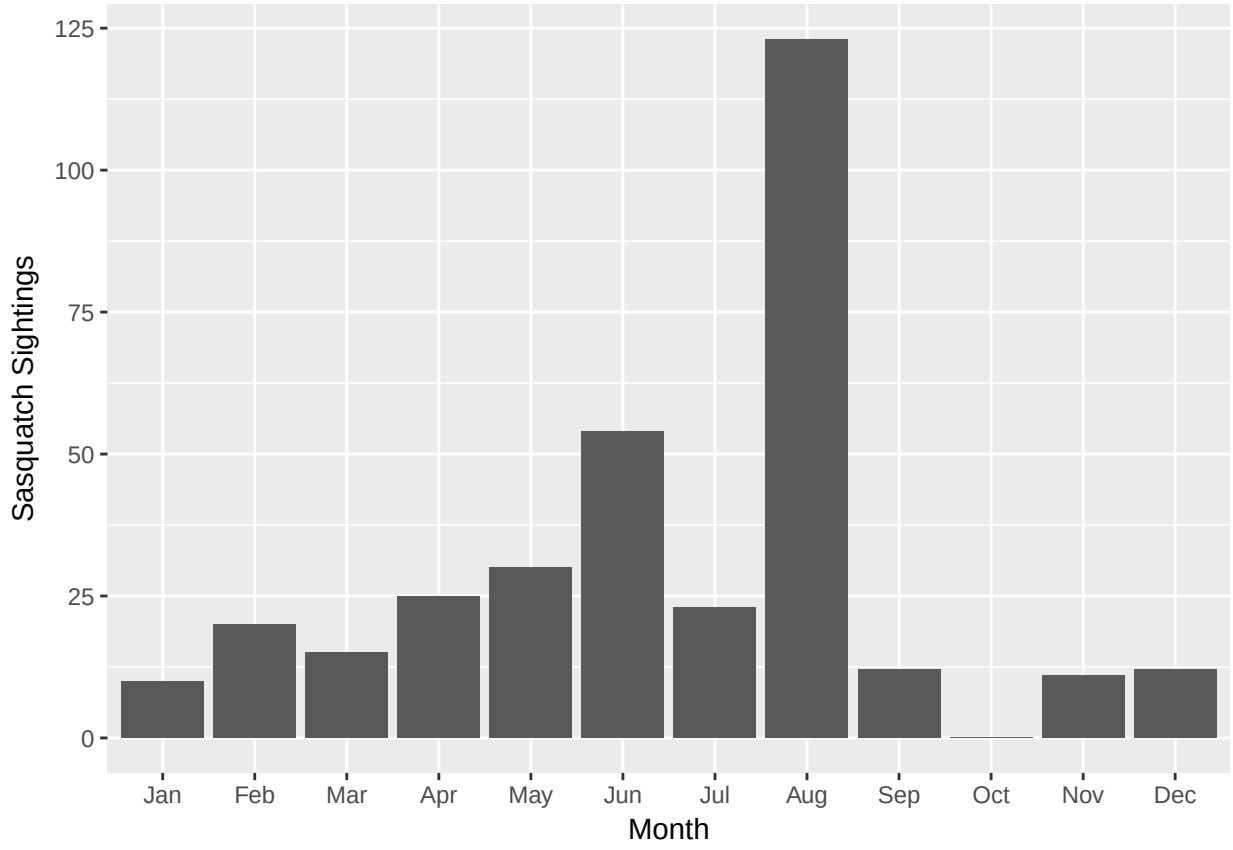
```
sas_dat3 <- sas_dat

# Use names() function to directly access column names
names(sas_dat3)
```

```
## [1] "months"    "sightings"
```

```
# Change column names
names(sas_dat3) = c("Month", "Sasquatch Sightings")
```

However, this method can be inconvenient as you will also have to change how you refer to your columns as well. Notice that the argument y=`Sasquatch Sightings` was used to escape the scape in the column name.

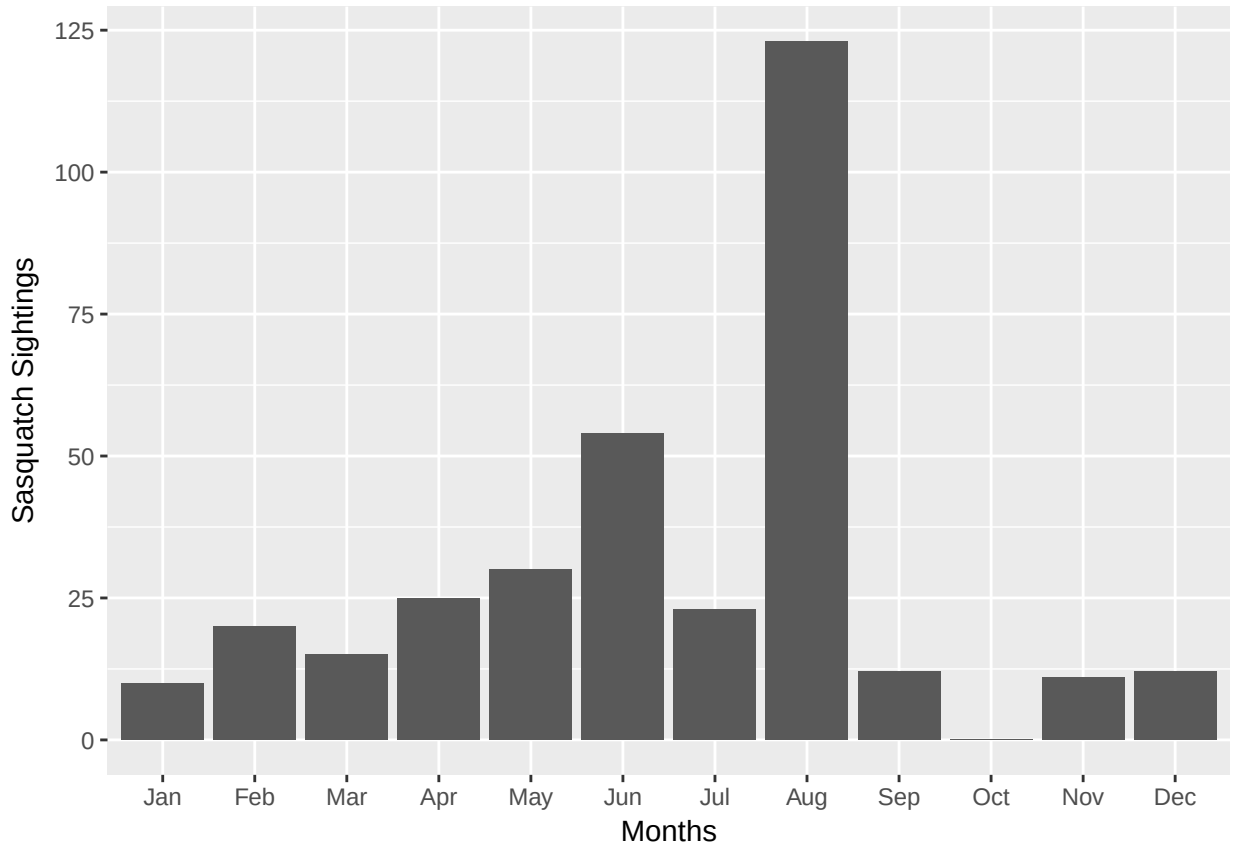
```
my_plot <- ggplot(data=sas_dat3, aes(x=Month, y=`Sasquatch Sightings`)) +
  geom_bar(stat="identity")

my_plot
```



Renaming the X and Y Axes: The Easier Way

This method only adds an extra line of code (if you use the same whitespace rules as I do).

```
ggplot(data=sas_dat, aes(x=months, y=sightings)) +
  geom_bar(stat="identity") +
  xlab("Months") + ylab("Sasquatch Sightings")      # Change x and y labels
```


125
100
75
50
25
0
Sasquatch Sightings
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months