R Pie Chart for data visualization

Daniel Hua

10/29/2017

```
library(readr)    # importing data
library(dplyr)    # data wrangling
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(ggplot2)  # graphics
library(plotrix)  # 3d graph
```

Introduction

The purpose of this post is to introduce the useage of pie charts in R. Pie charts have long been a major tool of data visualization in many forms of data analysis due to its simplicity. They are very straight forward and easy to understand visually for certain types of data. For example, displaying relative porprotions for a small number of categories would be best accomplished by pie charts. Learning to use pie chart in R would be extremely useful for data analysis purpose.

Pie Chart in base R

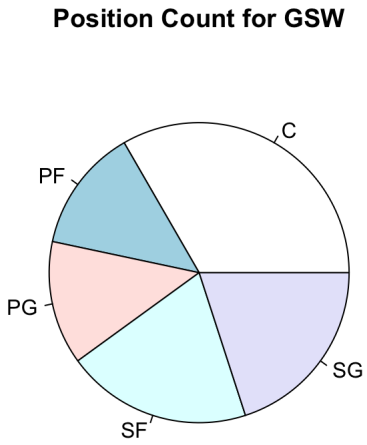
Basic functionality

To create pie chart in the base R, one could simply use the **pie()** function with a vector of positive numbers as the input. To illustrate its functionality, 2017 NBA roster information for the Golden State Worrior would be used here.

```
# getting the inital roster
roster <- read_csv("../data/nba2017-roster.csv")
```

```
## Parsed with column specification:
## cols(
##   player = col_character(),
##   team = col_character(),
##   position = col_character(),
##   height = col_integer(),
##   weight = col_integer(),
##   age = col_integer(),
##   experience = col_integer(),
##   salary = col_double()
## )
```

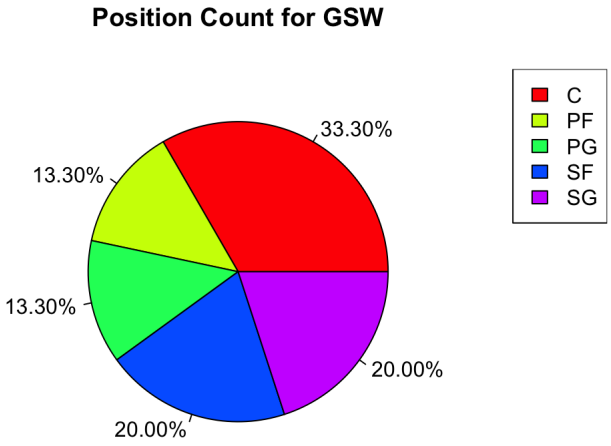
```
# getting the total count for each position for GSW
gsw <- filter(roster, team == "GSW")
gsw_po <- tally(group_by(gsw, position))

# generating pie chart
pie(gsw_po$n, labels = gsw_po$position, main = "Position Count for GSW")
```



This pie chart demonstrate the relative number of people in the 5 positions for the Golden State Warrior team in 2017. ### Showing percentages

The above example might not be the best representation of pie chart. Where pie chart really shine is when showing the precentages of groups and we could accomplish this by executing the code below.

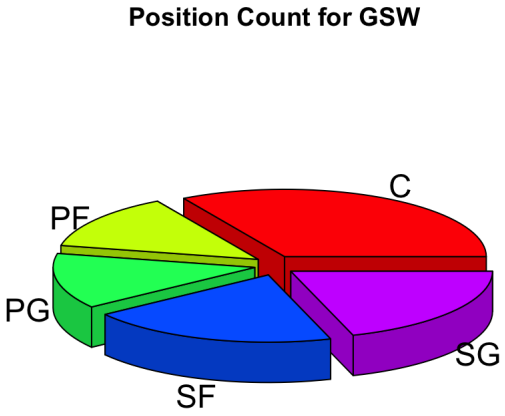
```
# Caculating the percentage of position in team
x <- gsw_po$n
perc <- round(100*x/sum(x), 1)
percent <- function(x, digits = 2, format = "f", ...) {
  paste0(formatC(x, format = format, digits = digits, ...), "%")
}
perc <- percent(perc)
# Generating the graph
pie(x, labels = perc, main = "Position Count for GSW", col = rainbow(length(x)))
legend("topright", gsw_po$position ,fill = rainbow(length(x)))
```



3D pie chart

In addition, 3-dimensional pie chart could also be accomplished in R with the package **plotrix** as shown below

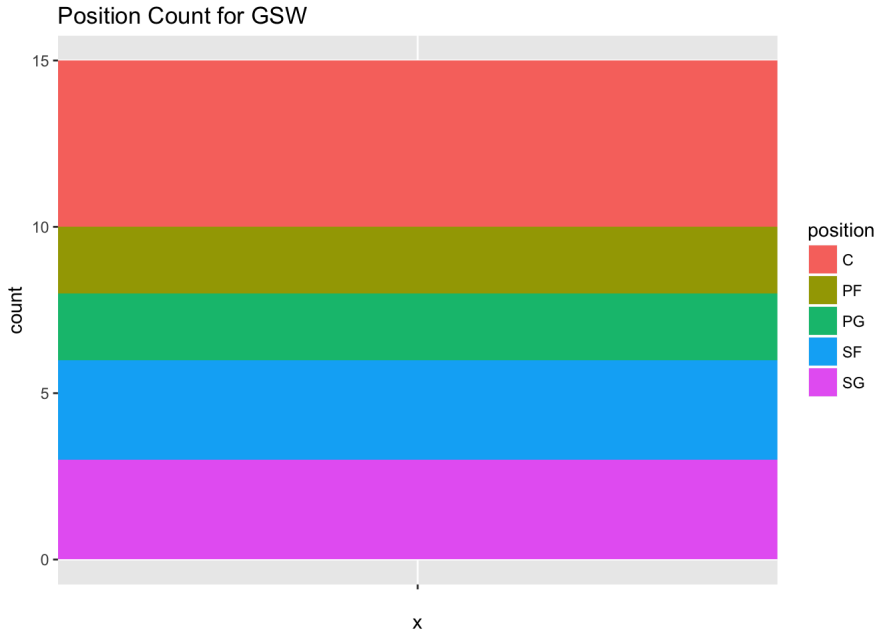
```
pie3D(x, labels = gsw_po$position, explode = 0.1 ,main = "Position Count for GSW")
```



Pie chart using ggplot2

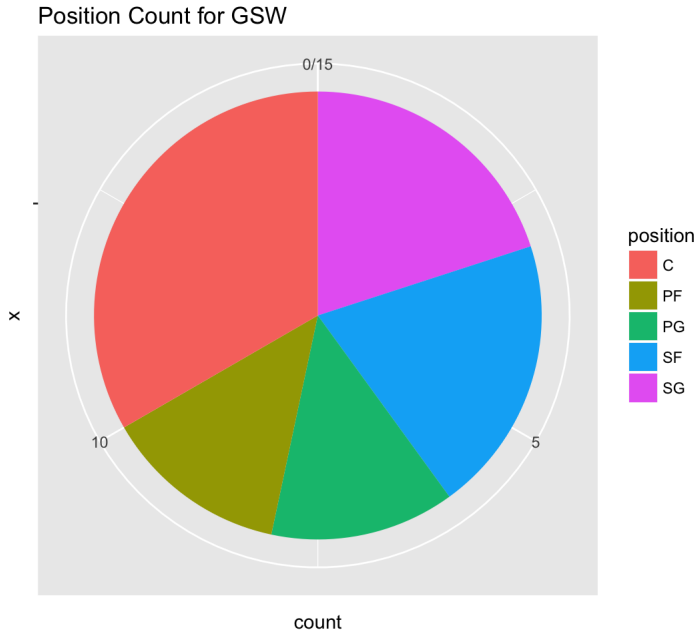
Unlike the base R, the package **ggplot2** offers a higher degree of control over creating plots. The same also goes for pie charts. Although **ggplot2** doesn't actually include a direct function for pie chart, we could easily convert the barplot to pie chart using the function *coord_polar*

```
# first create a barplot
bar <- ggplot(gsw_po, aes(x = "", y = n, fill = position)) +
  geom_bar(width = 1, stat = "identity") +
  labs(y = "count", title = "Position Count for GSW")
bar
```



To convert the barplot to a pie chart, simply do the following:

```
p <- bar + coord_polar("y",start = 0)
p
```



Limitations

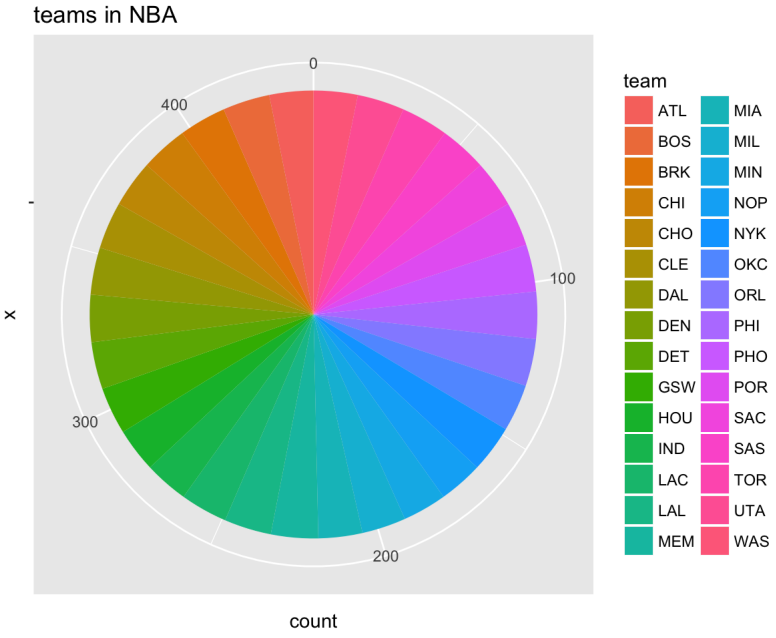
Despite all of its advantages, pie chart also has a lot of limitations. The most obvious one would probably be that pie chart is only valuable for one-dimensional data and could only analysis one single variable. Unlike other method such as scatterplot, pie chart could not demonstrate relationship between multiple variables since it only takes in one single variable as its input.

Another major limitation of pie chart is that pie chart is only straightforward for a small amout of categories. When there are more than 6 categories, pie chart tends to become increasingly confusing and would not be a good way to represent data anymore. To better illustrate this, an example is provided below:

Example of pie chart with a lot of categories

Instead of positions in GSW, we become interested in the porprotion of teams in the whole roster and decided to do a pie chart for visulization

```
ggplot(data = roster, aes(x = "", fill = team)) +
  geom_bar(width = 1) +
  coord_polar("y") + labs(title = "teams in NBA")
```



This is clearly a bad idea because there are way too many teams in NBA for visulization using pie chart so we ended up with this graph that is too complicated to get anything useful.

Summaeize

Pie charts in R could indeed be a powerful and valuable tool for simplistic data visualization. However, the limitations of pie charts are also very significant and even fatal. The best usage of pie charts is probably showing the proportions of a small amount of groups, preferrbly no more than 5 or 6. Nevertheless, when used appropriately and combined with other visualization method, pie charts in R would definitely boost the overall quality of the data analysis.

Reference and further readings:

- http://www.sthda.com/english/wiki/ggplot2-pie-chart-quick-start-guide-r-software-and-data-visualization
- https://www.statmethods.net/graphs/pie.html
- https://stackoverflow.com/questions/20442693/how-to-use-ggplot2-to-generate-a-pie-graph
- https://steemit.com/programming/@dkmathstats/pie-charts-in-r-with-ggplot2
- ggplot2 cheatsheat
- nba 2017 roster data
- https://stackoverflow.com/questions/7145826/how-to-format-a-number-as-percentage-in-r