

# Post1-Daniel-Kang.Rmd

Daniel Kang

10/28/2017

## The Entry to WindRose

```
library(ggplot2)
library(readr)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
##
## 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
```

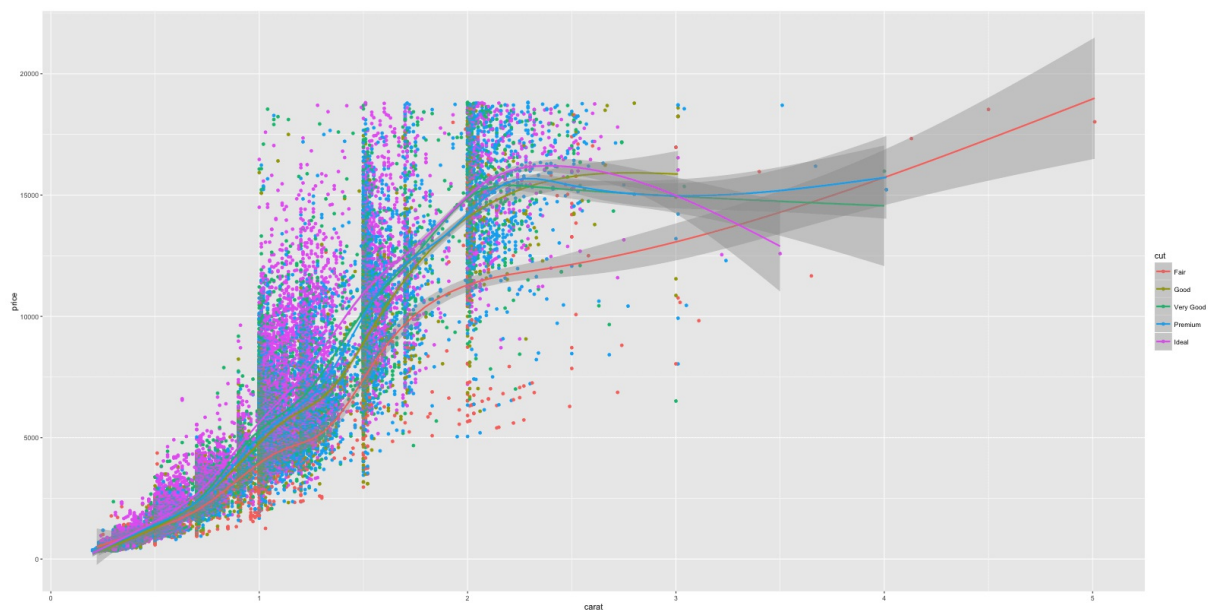
```
post1 <- read.csv('/Users/Danielisgood/stat133-hws-fall17/hw02/data/nba2017-player-statistics.csv')
```

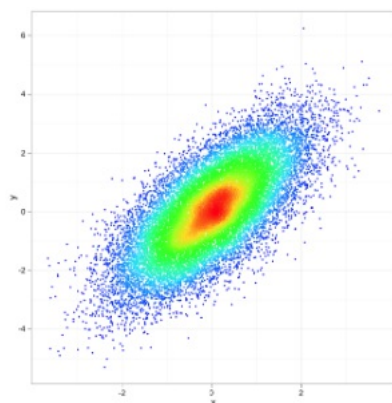
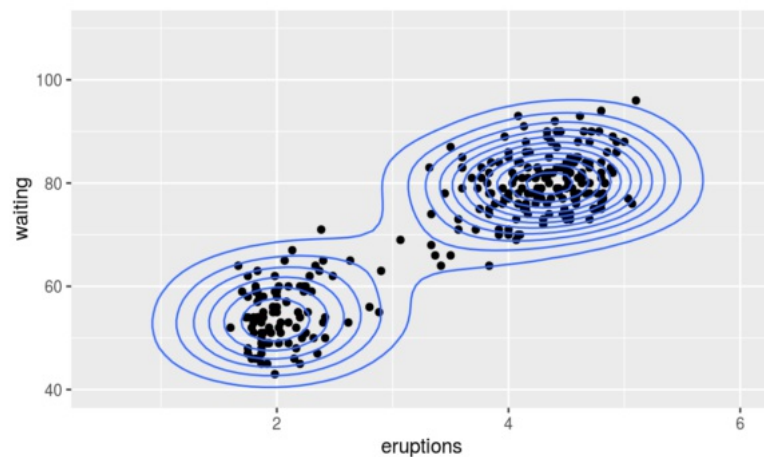
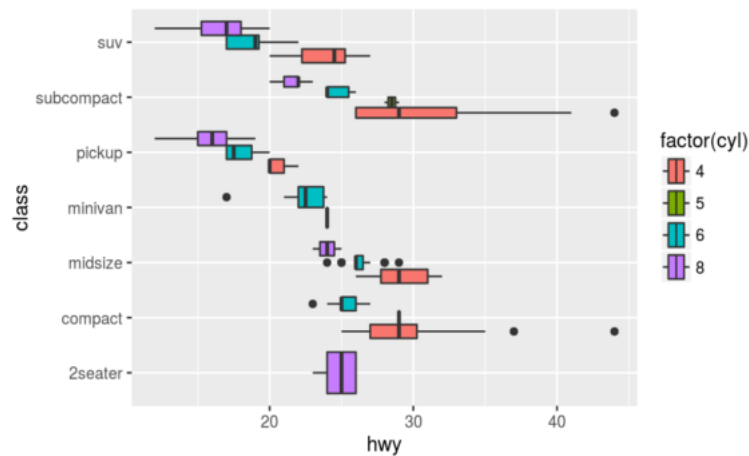
## Introduction:

This is the first time that I learn R language in this class, and we go through a lot of topics such as dplyr, git and some vectors. It was pretty interesting because I am a HUGE NBA fan so I have a lot of fun doing those analyses and all the other things start getting boring until I met `ggplot2`. The main reason that I love `ggplot2` is that `ggplot2` make the data that we made look so much simpler. In another word, `ggplot2` can let any random people understand what are those data about. By using `ggplot2`, you can make a lot of different plots for the same data. Last but not least, you can design color, size and the shape! I can not wait to introduce some fun functions about `ggplot2` to you guys.

The topic I want to introduce today is one of my favorite topic is `geom_bar`, you probably think it is easy but it is gonna be so interesting let the bar plot combine with the polar coordinate, I will introduce how useful and fun in the rest of the post

Here are some basic ggplot functions:





## The history of `ggplot2` and `wind rose`

So let's start from the history of `ggplot2`. `GGplot2` is an R package created by Hadley Wickham in 2005. It can highly improve the quality and aesthetic of your graphs! If you speak French, check this clear quick start page. More over R studio provides a really helpful cheat sheet summarizing the main `ggplot2` functions. I hope you will find the graph you are looking for in the examples below !If it is not the case, don't forget to contribute by requesting or proposing a graph !

## The motivation:

I have to mention again, the reason that `ggplot2` is so successful that is simple and easy to understand. "The structured syntax and high level of abstraction used by `ggplot2` should allow for the user to concentrate on the visualizations instead of creating the underlying code. On top of this central philosophy `ggplot2` has:

1. Increased flexible over many plotting systems.
2. An advanced theme system for professional/publication level graphics.
3. Large developer base – Many libraries extending its flexibility.
4. Large user base – Great documentation and active mailing list.

## Visuals

The logical in `ggplot2`: In the package, there is actually have two different things, `qplot()` and `ggplot2`, they both create to make the graph, but there are also some tiny differences in the code. But in this post, I will focus on the `ggplot2`.

What is `ggplot2`? The package that can make a graph! But how? Image that `ggplot2` is a person how have a pen, paper, and all the stuff to make a graph, to simplify step is, set up the paper, make the graph or plot and finally put some color or but a little but design on it. I will introduce couple kinds of `geom_bar` below.

So first I want to break down a code of `ggplot2` Background: As a student, I always like to learn by myself in some specific topic such as `ggplot2` by myself. `ggplot2` is one of the best package that I have learned, and that's the reason make me want to learn more by my self.

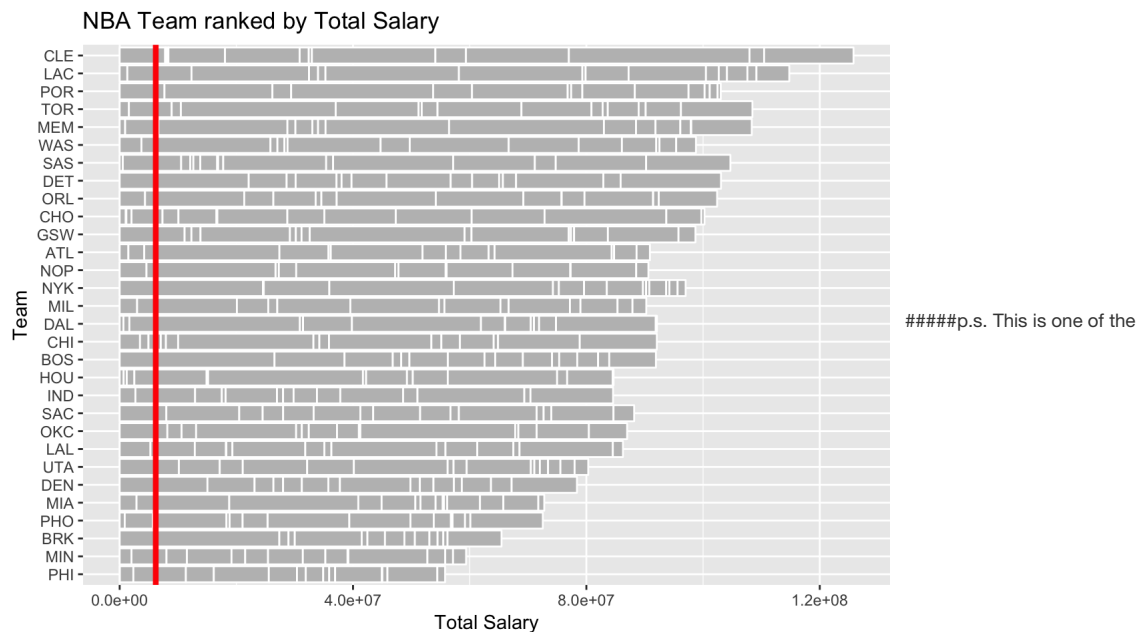
## How does the ggplot works?

Let's See!!!!!!

### Breakdown the Code.

In this paragraph, I want to use an exaple from one of my work to break down the post and you will have some sense about ggplot!

```
ggplot(post1, aes(x = reorder(Team, Salary), y = Salary))+  
  geom_bar(stat = 'identity', color = 'white', fill = 'grey')+  
  coord_flip()+  
  ylab("Total Salary") + xlab("Team")+  
  ggtitle("NBA Team ranked by Total Salary")+  
  geom_hline(yintercept = mean(post1$Salary), color = "red", size = 1.5)
```



questions in homework 3. So you have to understand that `ggplot2()` will not make any plot, `ggplot2()` is just like giving you a blank paper and as you see the code above, you have to assign everything manually. It is not as fast as other graphic application like Excel but this is also the reason why everyone likes to use `ggplot2` because it is simple and it does one thing by one code which means you can fully design your own plot.

As most of you guys known, there is tons of different type of plot you can do for the same data such as `geom_boxplot`, `geom_point`, `geom_bar`... etc. In R, we should use "+" to combine the thing you add to that plot. #####Let's breakdown the code~ As I said, the `ggplot()` is just kind of set up the x-axes and y-axes and which data we are using. So the next step is starting to design your plot! In this case, I use the `geom_bar`, and you can do a lot of thing after you set up these two main things such as `coord_flip` is the function that can make your plot being horizontal or vertical. You can also have a label of what is y and what is x by using `xlab("")` etc. I have to mention again that to accumulate these function we have to put "+" between them, last but not least, the format is various, the way I did is just trying to make it look more organized.

### Things to know before you start making the plot:

Different type of data and the thing you want to analyst will have the different suitable plot. For example, if you want to talk about the average point that each of NBA team score, you probably can make a `geom_boxplot` to know the mean, high and low. But if you want to know the distribution that the points that the player score, `geom_point` is a pretty good choice.

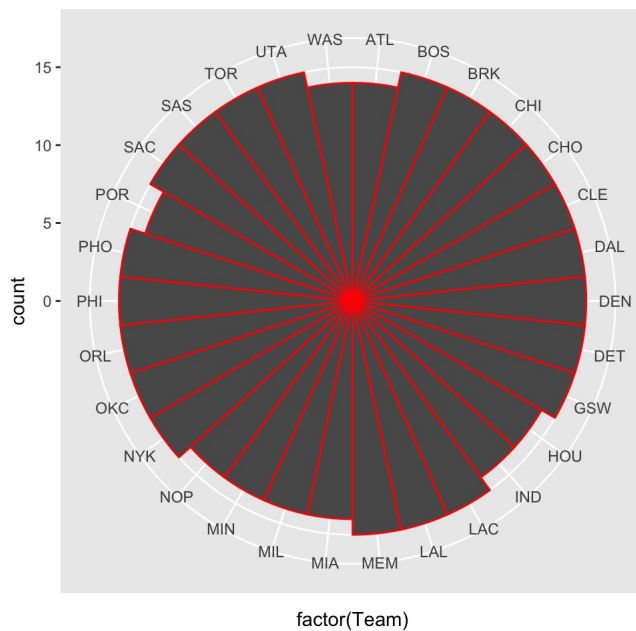
Here is the cool example of `geom_bar`, I know we learn it in the lecture but this is the easiest and one of the most common plot people use for analyst the data.

A bar chart uses height to represent a value, and so the base of the bar must always be shown to produce a valid visual comparison. Naomi Robbins has a nice article on this topic. This is why it doesn't make sense to use a log-scaled y axis with a bar chart.

By default, multiple bar occupying the same x position will be stacked atop one another by `position_stack`. If you want them to be dodged side-to-side, use `position_dodge`. Finally, `position_fill` shows relative proportions at each x by stacking the bars and then standardising each bar to have the same height.

Here is the easy example of the `Wind rose` which is `geom_bar` function, let's take a look and see how simple the bar plot with polar coordinate is!

```
ggplot(post1, aes(x = factor(Team))) +  
  geom_bar(width = 1, colour = "red") +  
  coord_polar()
```

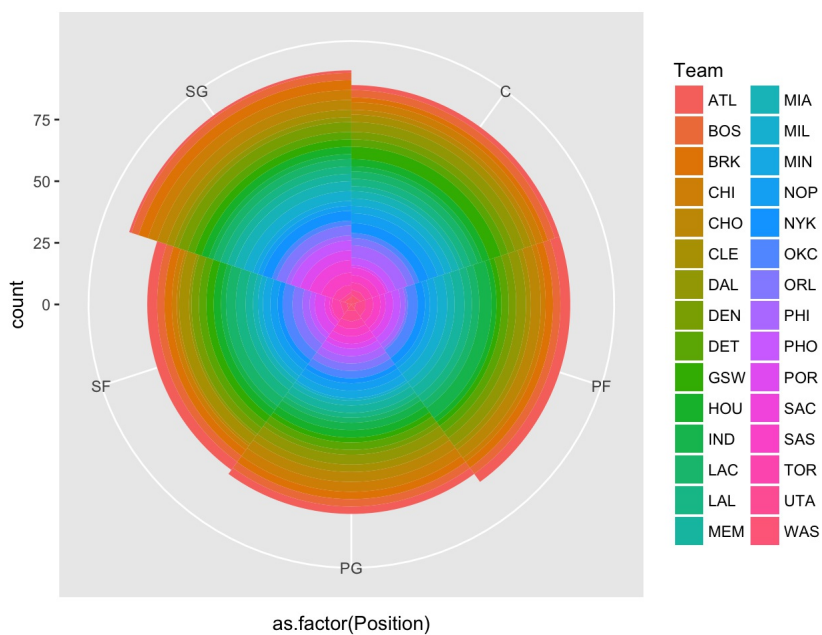


As you can see, we can setup the plot with couple easy lines. The bar plot is useful for the data that you want to compare. Sometimes people use it for such as personality test or NBA players' analysis. In this function you can assign a lot of things such as color, width etc. Also you can read this really quick that the counts of the variable that are in the assigned variable.

If it is boring, that fine!! Let's make it colorful~~ Color is really important that can also represent some other variable. Looks nicer, feel better lol.

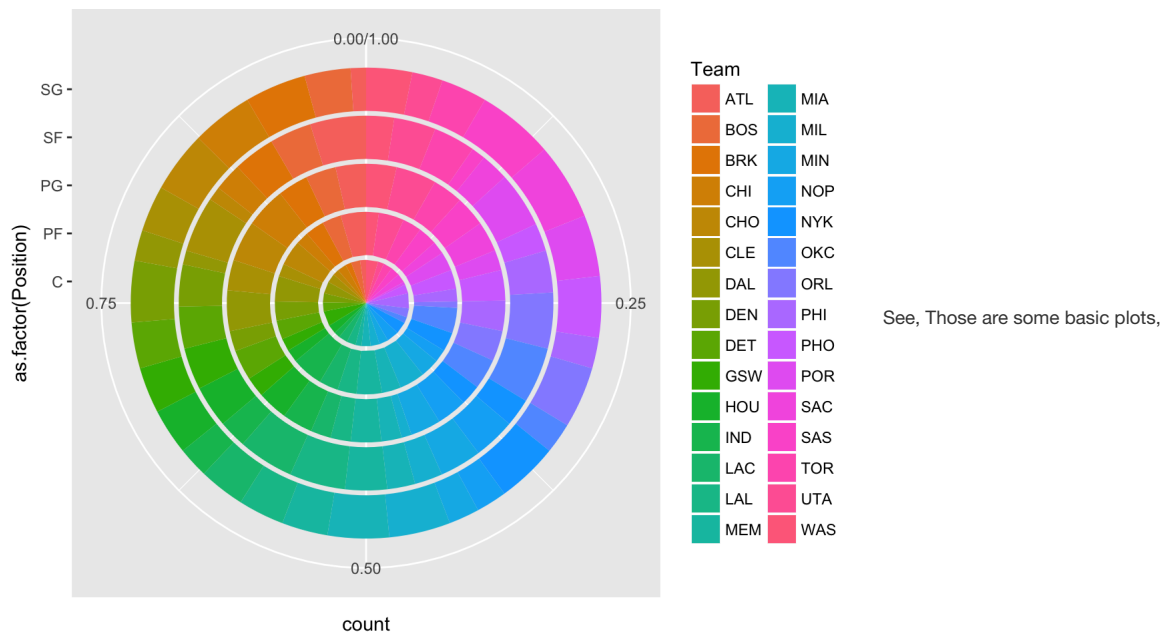
The is the plot include the teams and positions. Sometimes you can also seprate the team by the area of The State. It is gonna be even BETTER

```
yy <- ggplot(post1, aes(x = as.factor(Position), fill = Team))
# Wind rose
yy + geom_bar(width = 1) + coord_polar()
```



The plot below, you can also set the theta as y-axis as well.

```
# Race track plot
yy + geom_bar(width = 0.9, position = "fill") + coord_polar(theta = "y")
```



You can change a lot of things for the same variable. If you are a NBA fan, copy the code and I know you wanna do something!!!!!!

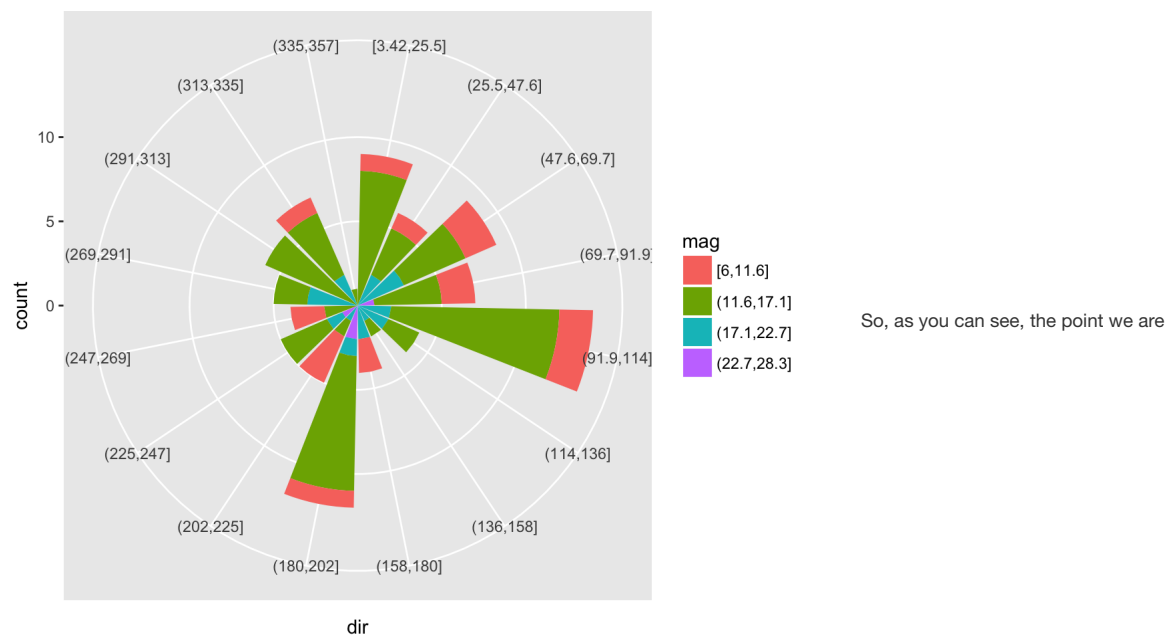
The example of The real Wind Rose plot + `coord_polar()` (advance!!!!)

Wind rose plot. This kind of plot have been use very often by Meteorologist, it basically describes the direction and the speed of the wind and how it distribute. The wind rose plot is an extension of bar plot, it analyt the data by the wind from different direction and the frequency of it. Sometimes people will use the color to show the level of the wind.

```
set.seed(1234)
#creat 100 wind directions ramdonly and make it be inside 16 areas
dir <- cut_interval(runif(100,0,360), n = 16)
c#reat 100 wind speeds and separate it to 4 levels
```

```
## function (...) .Primitive("c")
```

```
mag <- cut_interval(rgamma(100,15), 4)
sample <- data_frame(dir = dir, mag = mag)
#let x axis = wind directions, y = frequency, color represent wind speed.
p <- ggplot(sample, aes(x = dir, y = ..count.., fill=mag))
#after setting it all up we have to make a bar plot with polar coordinate
p + geom_bar() + coord_polar()
```



making those plots is to make people understand our report of some specific easy and clear so you have to make sure the you have to make the difference between the variables you want to compare.

Last but not least, I always see this plot on some research papers or TV, and hope you guys will just like me, enjoy the moment that you finally can say "I can make this thing in R too" next time if you and you friend see this on somewhere lol.

## Take Home Message

In this post, I basically introduce the fundamental of ggplot and also some useful function. The really first thing is you have to know the

advantages of ggplot2. Then you will have to learn the basic code such as `ggplot()`, how to setup the function and remember what is that specific function control. So when you wanna make something on it, you will know what to do. Last but not least, you are able to do a lot of data analyst by the ggplot, it is easy to learn and it is gonna make you excited!

After reading this post, Hope you understand more about ggplot and not scare of it anymore. I always have the error when I am making the plot, but it's because we all will make mistake so you just have to follow the error message and debug it.

Hope ya'll have a good day.

## References

1. [Polar coordinates] ([http://ggplot2.tidyverse.org/reference/coord\\_polar.html](http://ggplot2.tidyverse.org/reference/coord_polar.html))
2. [geom\\_bar1](#)
3. [Reasons use ggplot\(\)](#)
4. [geom\\_bar2](#)
5. [geom\\_bar3](#)
6. [geom\\_bar4](#)
7. [geom\\_bar5](#)
8. [geom\\_bar6](#)
9. [geom\\_bar7](#)
10. [geom\\_bar8](#)
11. [geom\\_bar9](#)