L09 Coordinates

Data Visualization (STAT 302)

Anna Wagman

Contents


```
library(knitr)
opts_chunk$set(warning = FALSE, message = FALSE, dpi = 300)
```

Overview

The goal of this lab is the use of an alternative coordinate system in `ggplot2` to build a plot.

Datasets

We'll be using the `mod_nba2014_15_advanced.txt` dataset — add to your project's `/data` subdirectory. The `codebook_mod_nba2014_15_advanced.txt` provides a quick description of the variables in the dataset — suggest adding it to the `/data` subdirectory as well.

```
# Load package(s)
library(ggplot2)
library(dplyr)
library(knitr)
library(tidyverse)

# Read in dataset
nba_dat <- read_delim("data/mod_nba2014_15_advanced.txt", delim = "|") %>%
  janitor::clean_names()
```

The Exercise

Using the `mod_nba2014_15.txt` dataset we will begin the process of trying to recreate/approximate the plot type featured in the http://fivethirtyeight.com/ article Kawhi Leonard Is The Most Well-Rounded Elite Shooter Since Larry Bird for any player of your choice for the 2014-2015 season.

Start with data wrangling

When data wrangling we will need a helper function for creating the quartiles for players. Additionally, it will be useful to exclude players that played less than 10 games or played less than 5 minutes a game. That is, we only want to include "qualified" players.

```
# Returns quartile rank
quartile_rank <- function(x = 0:99) {
  # Set quartile
```

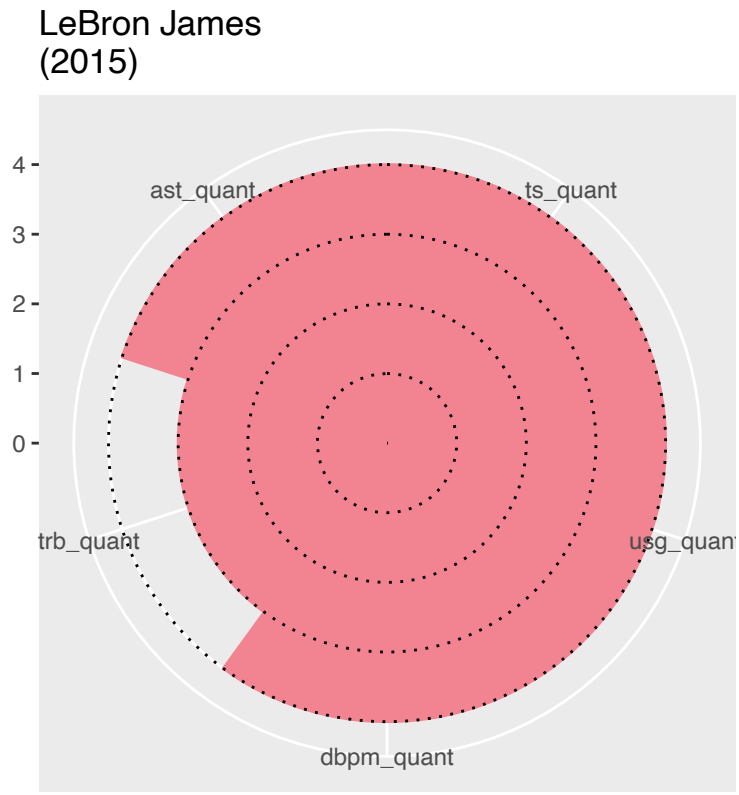
```
  quart_breaks <- c(
    -Inf,
    quantile(x,
      probs = c(.25, .5, .75),
      na.rm = TRUE
    ),
    Inf
  )
  cut(x = x, breaks = quart_breaks, labels = FALSE)
}
```

Add comments to the code below where indicated. The added comments should concisely describe what the following line(s) of code do in the data wrangling process

```
# Graphical data
nba_graph_dat <- nba_dat %>%
  #select players who played 10 or more games for at least 5 min
  filter(g >= 10, mp / g >= 5) %>%
  #use quartile_rank function from above to set range
  mutate(
    ts_quant = quartile_rank(ts_perc),
    trb_quant = quartile_rank(trb_perc),
    dbpm_quant = quartile_rank(dbpm),
    ast_quant = quartile_rank(ast_perc),
    usg_quant = quartile_rank(usg_perc)
  ) %>%
  #select columns we want - that have "_quant"
  select(player, contains("_quant")) %>%
  # setting the pivot
  pivot_longer(
    cols = -player,
    names_to = "variable",
    values_to = "value"
    ) %>%
  # arrange order by player
  arrange(player)
```

```
nba_graph_dat %>%
  filter(player == "LeBron James") %>%
  ggplot(aes(variable, value)) +
  geom_col(fill = "#F28391",
           alpha = 1,
           width = 1,
           color = "#F28391"
           ) +
  geom_hline(linetype = "dotted",
             yintercept = c(0, 1, 2, 3, 4, 5)) +
  scale_x_discrete(limits = c("ts_quant", "usg_quant", "dbpm_quant",
                              "trb_quant", "ast_quant")) +
  coord_polar() +
  labs(x = "",
       y = "",
       title = "LeBron James\n(2015)") +
```

```
  ylim(0, 4)
```



Start building the plot