Data Visualization in Basketball Games

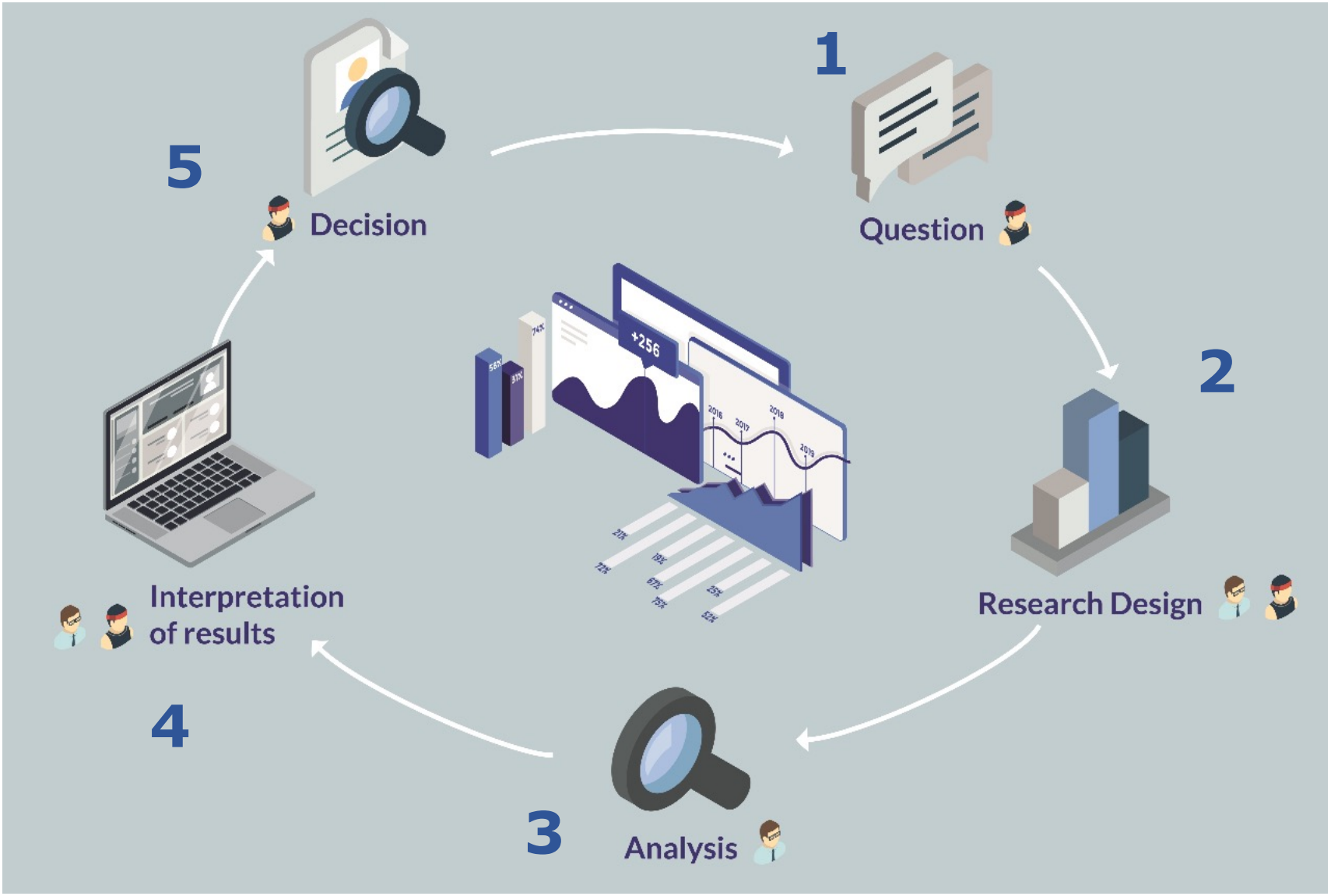
Jingyuan Sophie Li
PSYC 6135
April 3, 2025

Data Science in Basketball

A variety of goals, including but not limited to

- Determining discriminating factors between successful and unsuccessful teams
- Examining the patterns of scoring during the games
- Analyzing players performance and their impact on their teams
- Studying team's tactics and identifying optimal strategies

Zuccolotto & Manisera, 2020



Basketball Experts

Data Scientists

The virtuous cycle of Sports Analytics. Artwork by Gummy Industries Srl (https://gummyindustries.com/)

Agenda

Using data visualization to support communication

Two Parts

1. **Visualizing players' and teams' performance**
   - Radial plots
   - Shot charts
   - Bubble plots
2. **Visualizing teamwork and interactions between players**
   - Assist networks
   - Lorenz curve

Data

```
data(package="BasketballAnalyzeR")
```

Datasets:

- **Team's box scores (Tbox)** – The cases (rows) are the analyzed teams. The variables (columns) are team achievements in the considered games.
- **Player's box scores (Pbox)** – The cases (rows) are the analyzed players. The variables (columns) are individual achievements in the considered games.
- **Play-by-play data (PbP.BDB)** – The cases (rows) are the events occurred during the games. The variables are descriptions of the events.

Zuccolotto & Manisera, 2020

Visualizing performance

Radial Plots

- Visualizing the team's or player's profiles

```
# Radial plots
Pbox.PG <- subset(Pbox, Player=="Russell Westbrook" |
                        Player=="Stephen Curry" |
                        Player=="Chris Paul" |
                        Player=="Kyrie Irving" |
                        Player=="Damian Lillard" |
                        Player=="Kyle Lowry" |
                        Player=="John Wall" |
                        Player=="Rajon Rondo" |
                        Player=="Kemba Walker")
attach(Pbox.PG)
X <- data.frame(P2M, P3M, FTM, REB=OREB+DREB, AST,
                        STL, BLK)/MIN
detach(Pbox.PG)
radialprofile(data=X, title=Pbox.PG$Player, std=TRUE)
```

Key Variables:

P2M: 2-Point Field Goals Made

P3M: 3-Point Field Goals Made

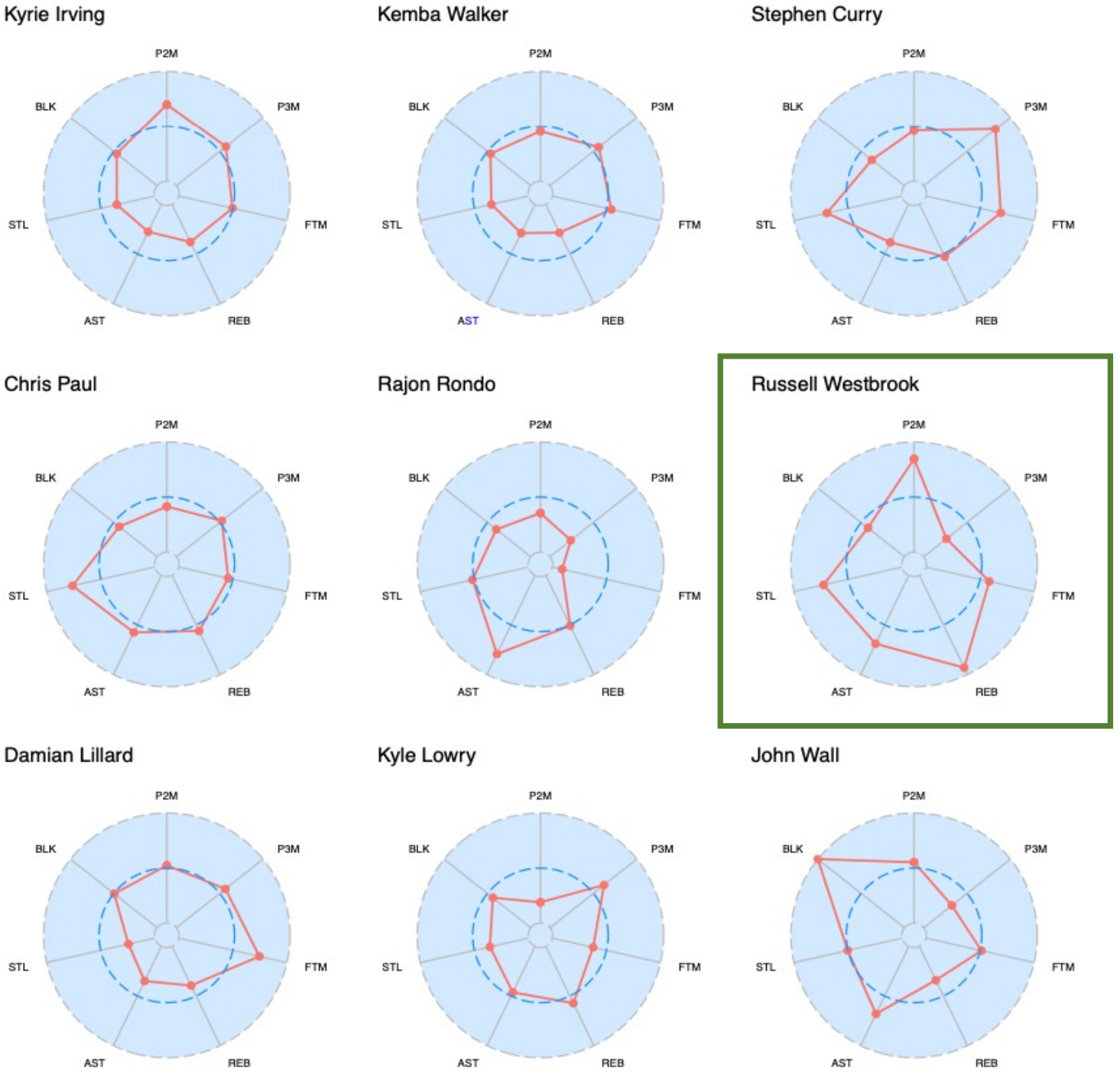
FTM: Free Throws Made

REB: Total number of rebounds (offensive + defensive)

AST: Total assists made

STL: Number of steals (taking the ball from opponents)

BLK: Blocks; number of shots blocked



Radial plots of nine selected points guards, standardized variables.

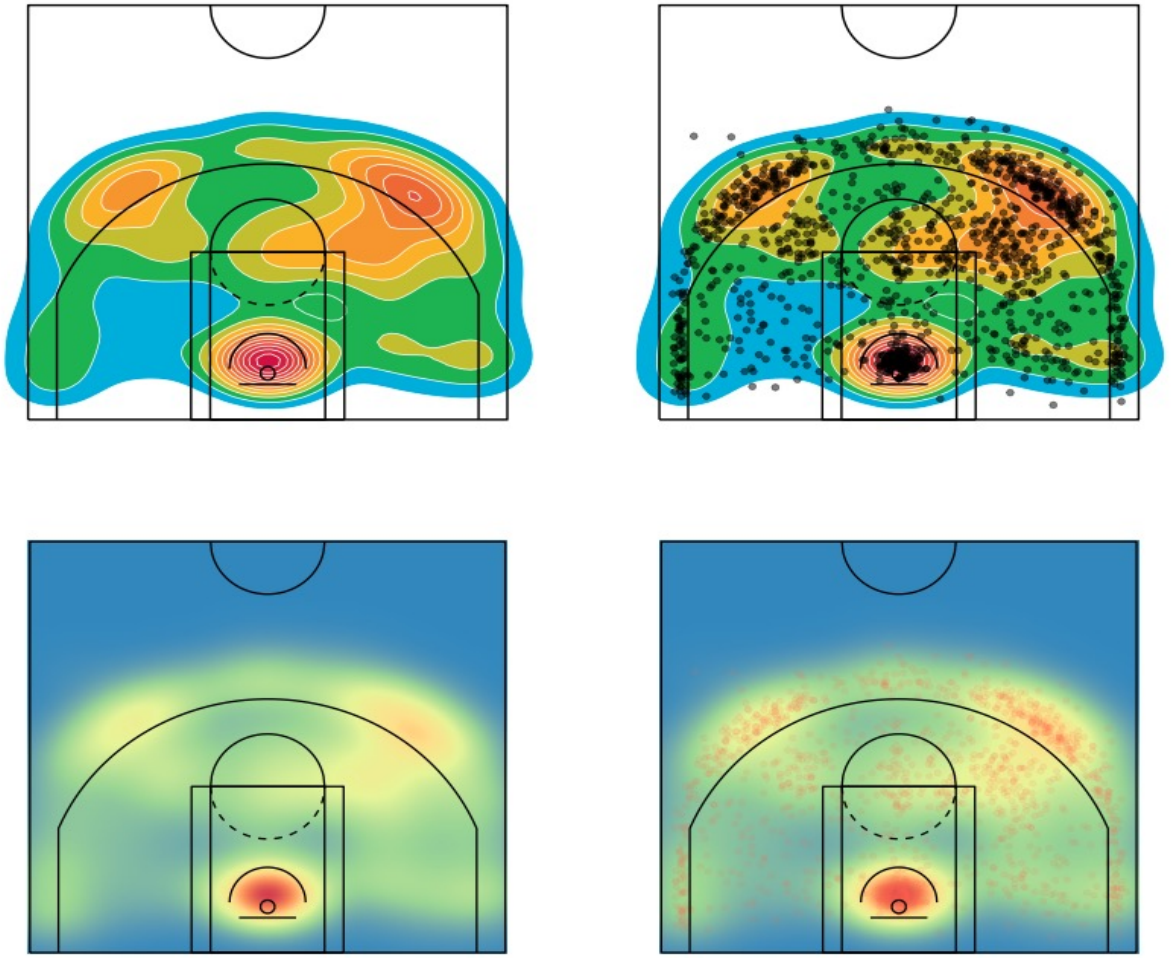
Dashed blue line: zero (average of each variable).

Limitation: It draws the eye to the area and shape, rather than the distance of each point from the center.

Visualizing performance

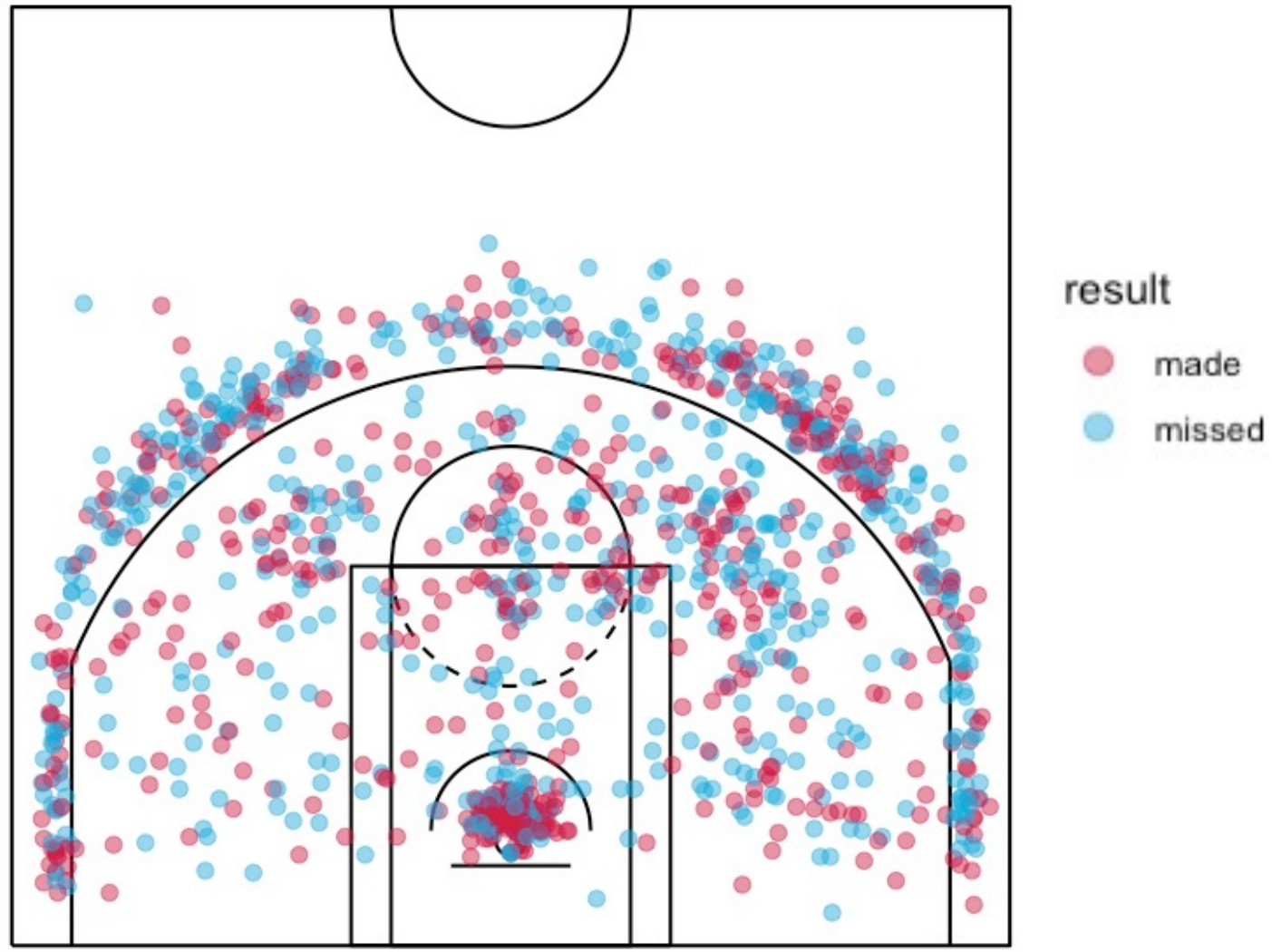
Shot Charts

- A useful tool to show shot patterns across the court.



Spatial density estimation of Klay Thompson’s shots

```r
# Shot Chart
subdata <- subset(PbP, player=="Klay Thompson")
subdata$xx <- subdata$original_x/10
subdata$yy <- subdata$original_y/10-41.75
# Shot Chart with made/missed shots colored
shotchart(
  data = subdata,
  x = "xx",
  y = "yy",
  z = "result",
  type = NULL,
  scatter = TRUE,
  pt.col = c("made" = "red", "missed" = "blue")
)
```



Visualizing performance

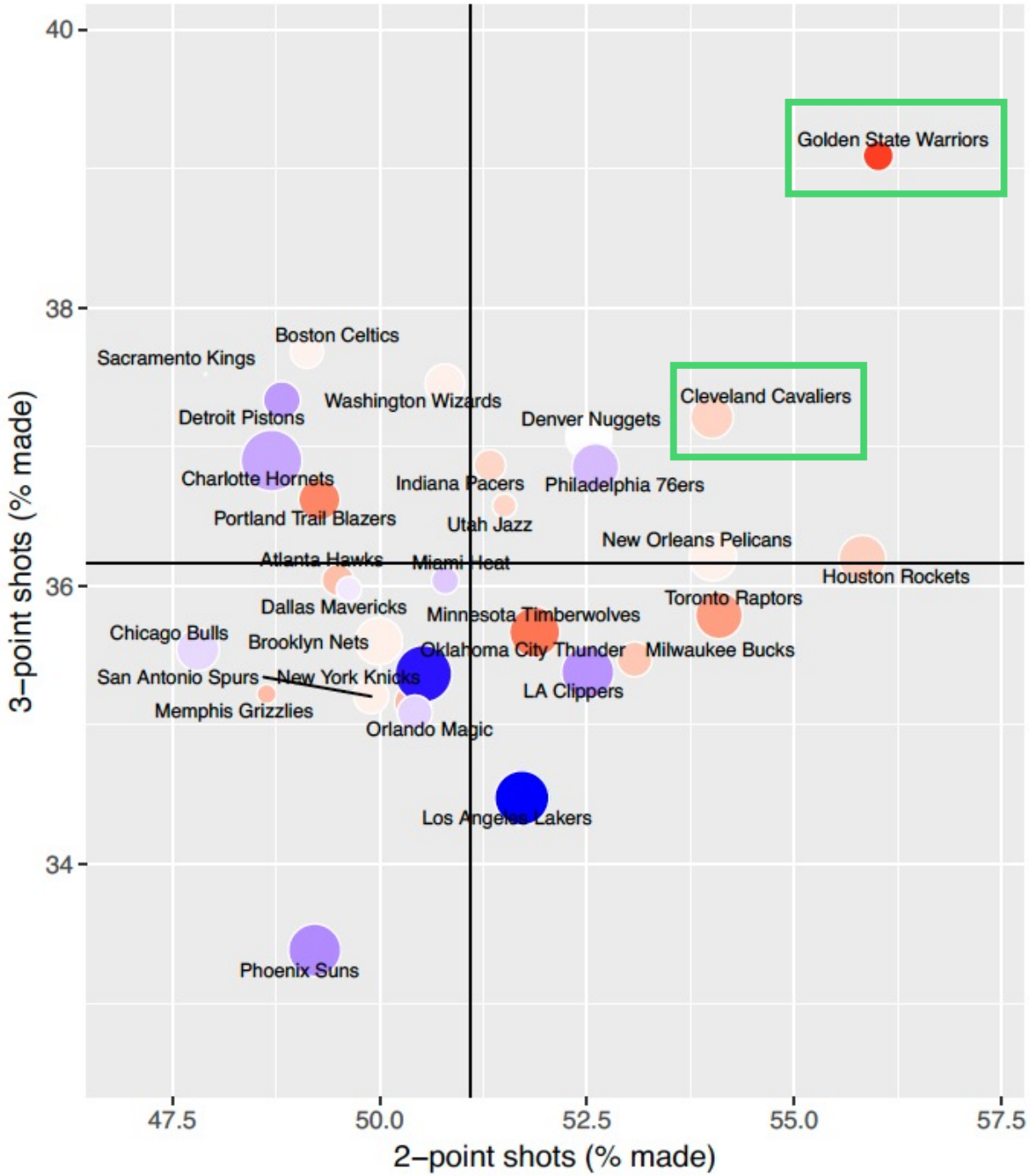
Bubble Plots

- Scatter plots where individual cases (teams or players) are plotted in the plane by means of bubbles instead of points
- The size and color of the bubble are defined according to two additional variables.

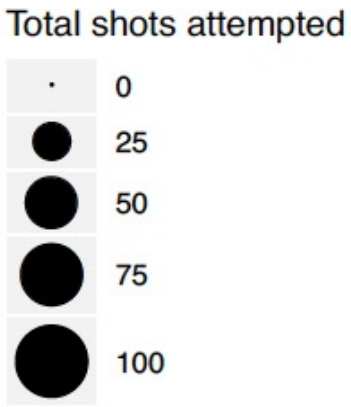
Example

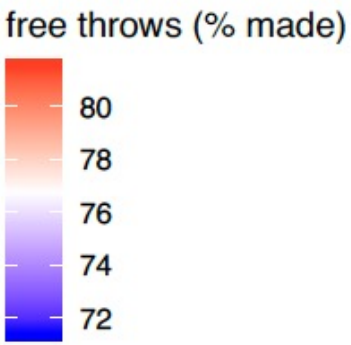
Question: How do NBA teams differ in their shooting efficiency?

Variables: 2-point shots (% made), 3-point shots (% made), total shots attempted, and free throws (% made)



Bubble plot of the teams according to shooting percentages and total shots attempted.






```
# Bubble plots
attach(Tbox)
X <- data.frame(T=Team, P2p, P3p, FTp,
                AS=P2A+P3A+FTA)
detach(Tbox)
labs <- c("2-point shots (% made)",
          "3-point shots (% made)",
          "free throws (% made)",
          "Total shots attempted")
bubbleplot(X, id="T", x="P2p", y="P3p",
           col="FTp",
           size="AS",
           labels=labs)
```

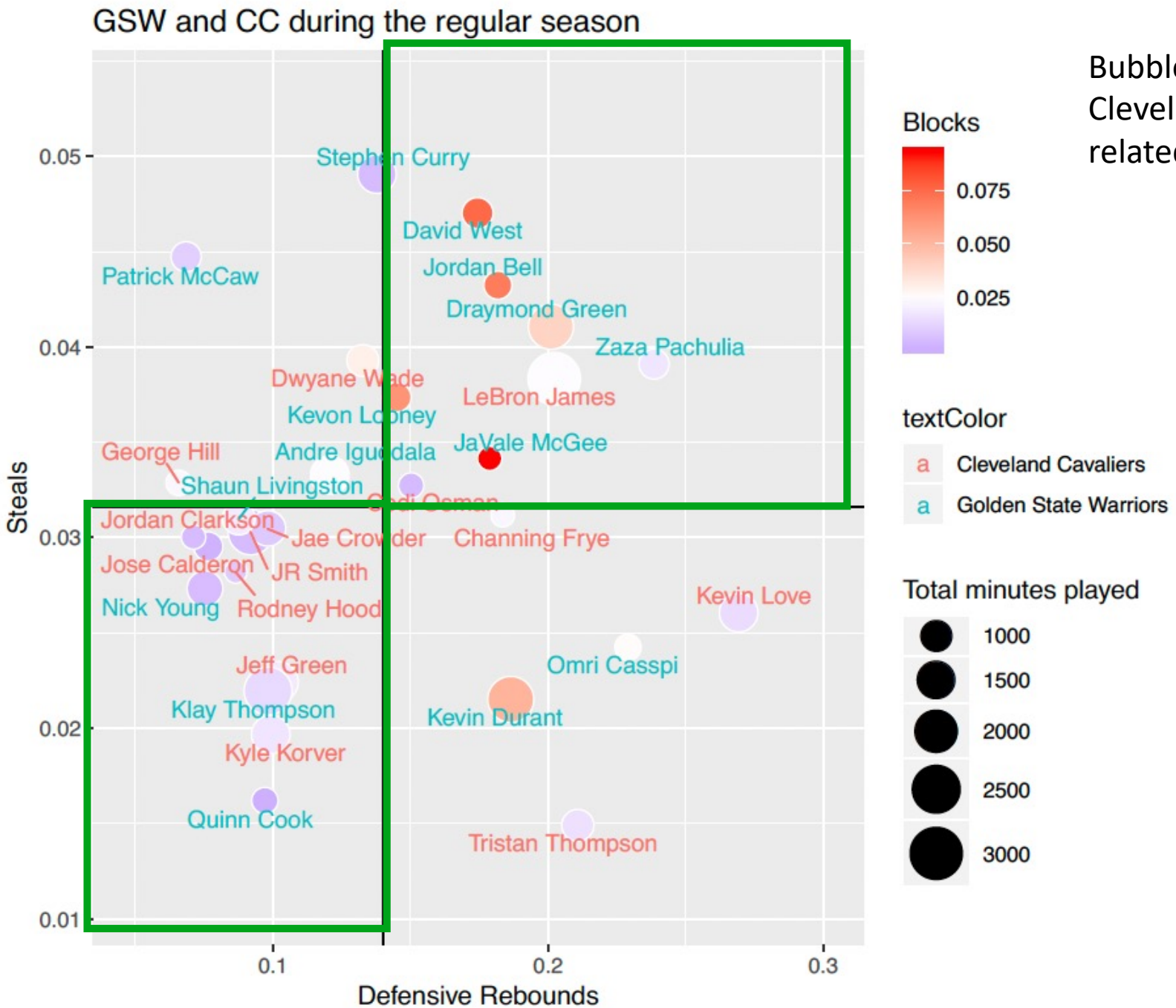
Visualizing performance

Bubble Plots

Example

Question: How do players from the Cleveland Cavaliers and Golden State Warriors differ in terms of **defense performance**?

Variables: team, steals, defensive rebounds, blocks, and total minutes played



Bubble plot of the Golden State Warriors' and Cleveland Cavaliers' players according to statistics related to defense (per minute played).

Visualizing Teamwork

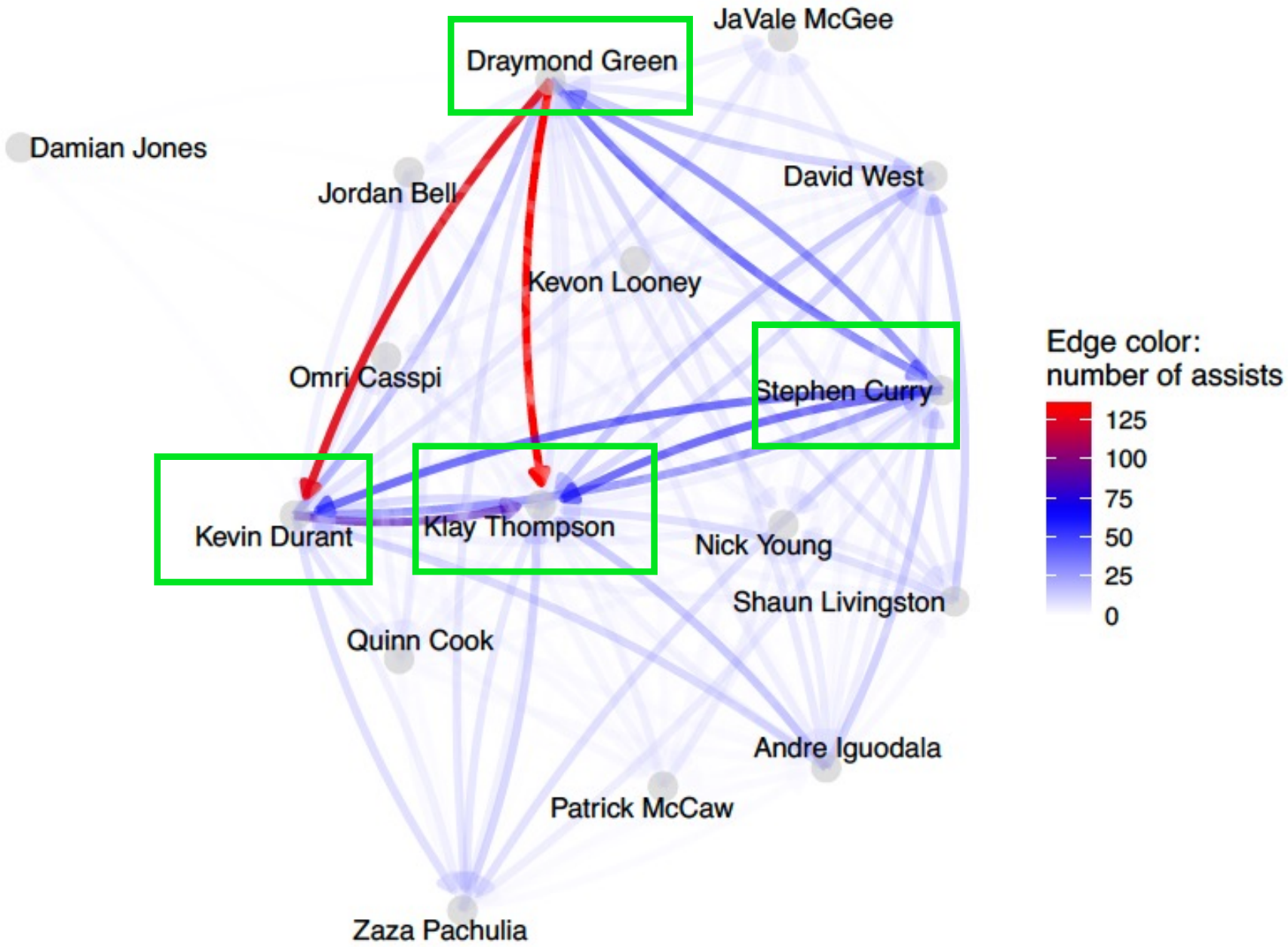
Assist Network

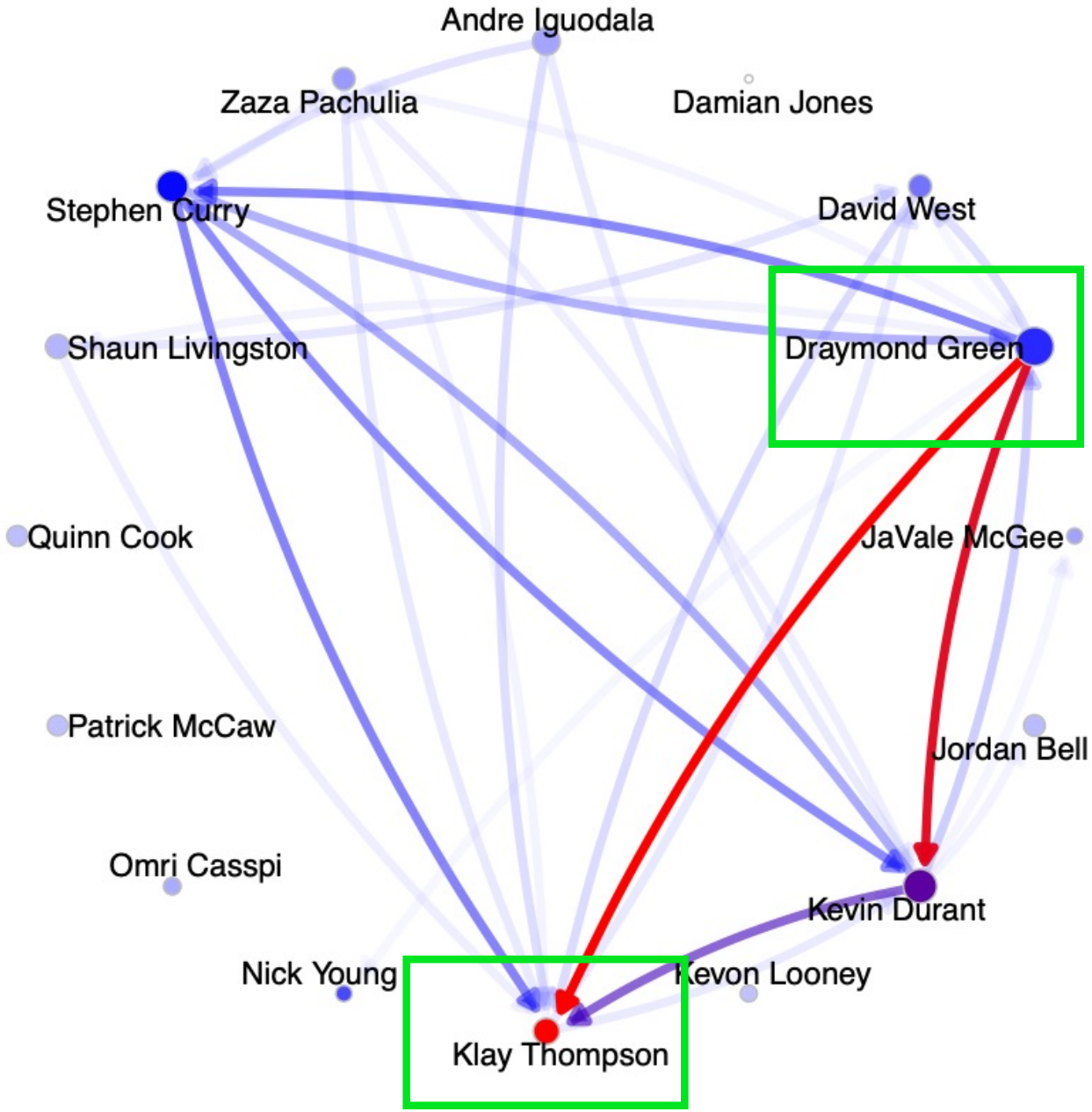
- Using R function *assistnet*, designed to investigate the network of assists in a team using play-by-play data
- **Assist**: A pass that directly leads to a teammate's scoring

Example:

Using Golden State Warriors' play-by-play dataset to build the network and calculate some assist statistics.

```
PbP.GSW <- subset(PbP, team=="GSW")
netdata <- assistnet(PbP.GSW)
set.seed(7)
plot(netdata)
```


JaVale McGee
Draymond Green
Damian Jones
Jordan Bell
David West
Kevon Looney
Omri Casspi
Stephen Curry
Kevin Durant
Klay Thompson
Nick Young
Shaun Livingston
Quinn Cook
Andre Iguodala
Patrick McCaw
Zaza Pachulia
Edge color:
number of assists
125
100
75
50
25
0



Edge color:
number of assists

125
100
75
50
25
0

Node color:
FGPTS_AST

900
600
300

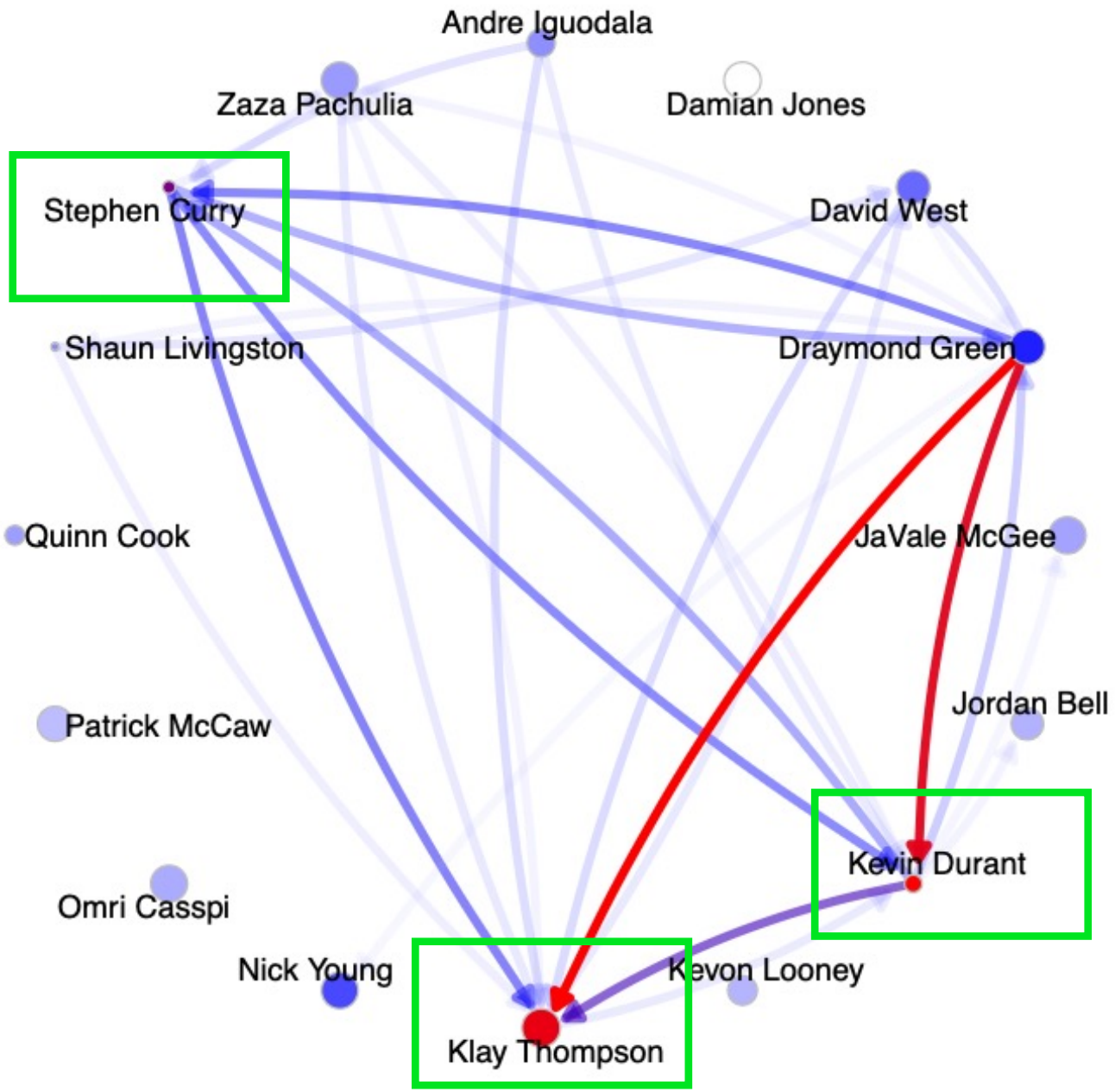
Node Color:
Point scored thanks to assists from teammates

Node size:
ASTPTS

200
400
600
800
1000

Node Size:
Points scored by other players because of this player's assists

How much do players rely on assists to score?



Edge color:
number of assists

125
100
75
50
25

FGPTS →

1000
500

Node Color:
Total points the player scored

Node size:
FGPTS_ASTp →

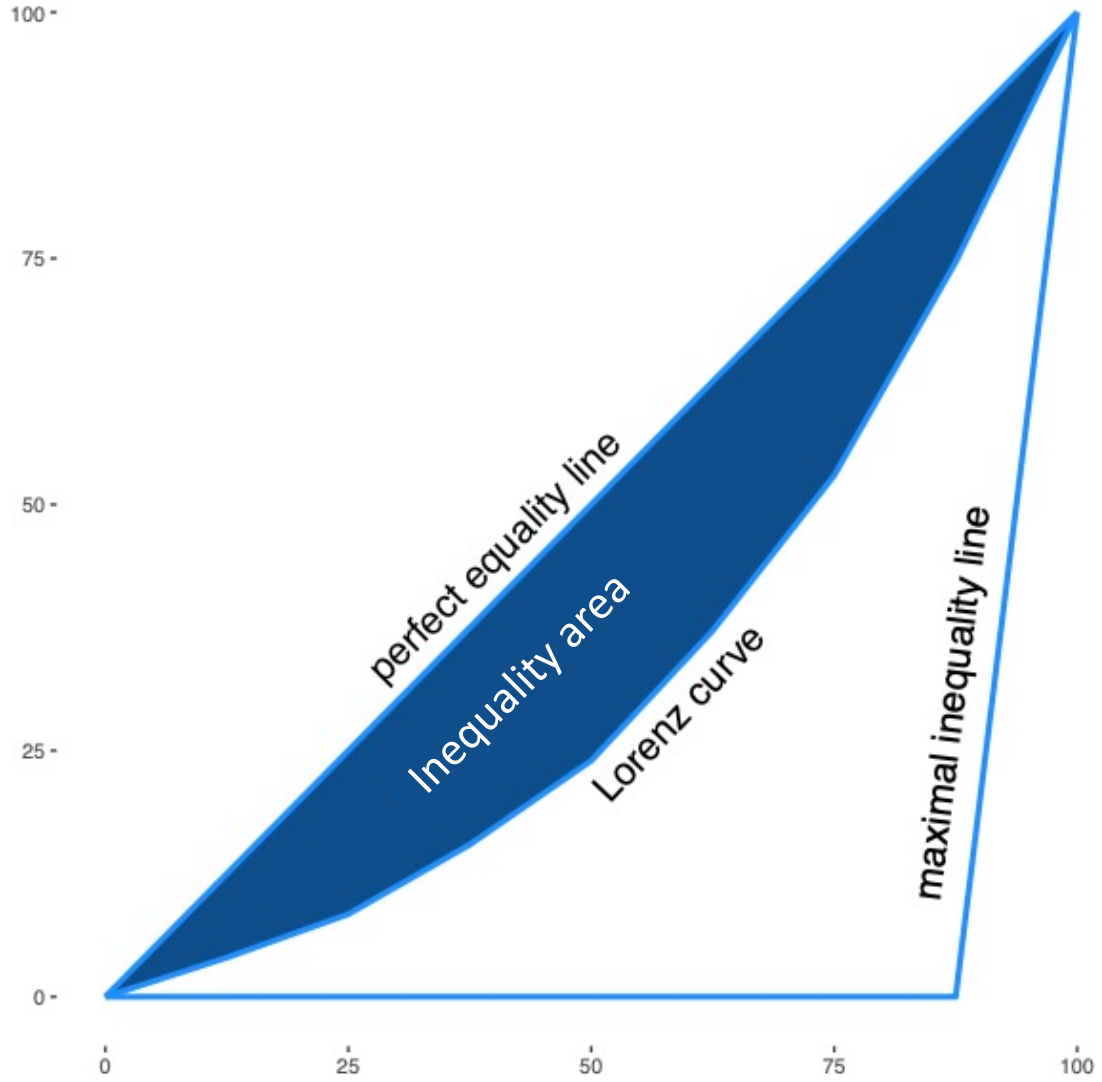
0.55
0.60
0.65
0.70
0.75
0.80

Node Size:
Percentage of points scored thanks to a teammate's assist over total points

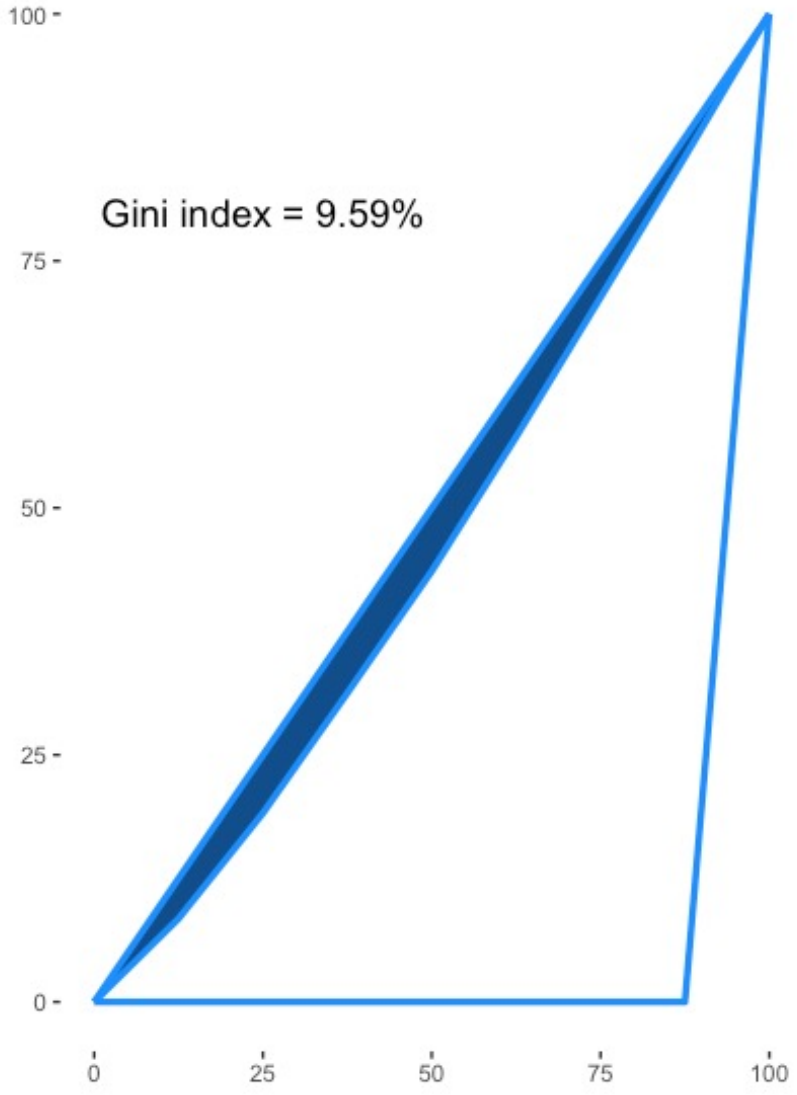
Visualizing Teamwork

Visualizing Inequality: Lorenz curve

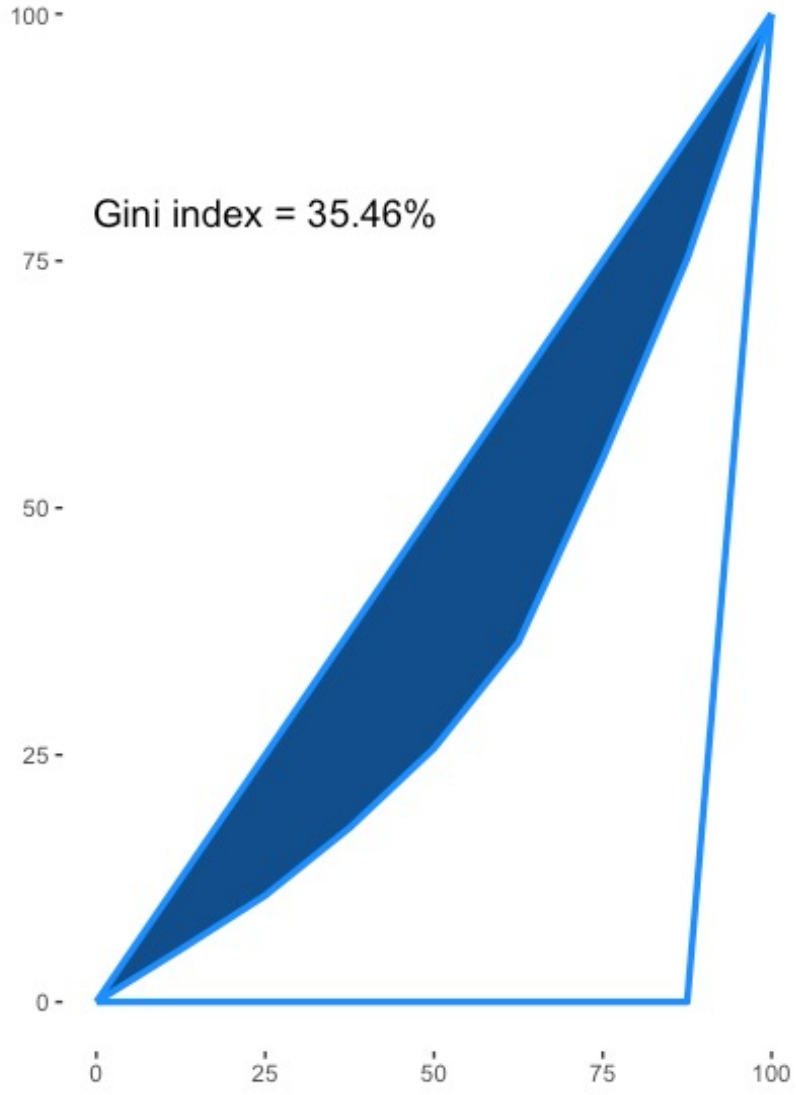
- **Inequality** of points made: There are few players who score a big part of the total points of the team
- **Equality** of points made: All the team member score the same points
- **Lorenz curve** represents the inequality area: the larger its size, the higher the inequality
- **Gini coefficient**: an index ranging from 0 (equal) to 100% (maximal inequality)



Brooklyn Nets



Golden State Warriors



Inequality analysis of scored point of Brooklyn Nets and Golden State Warriors – 8 players

Summary

- Data visualizations help translate complex stats into clear insights
- Different visualizations highlight different aspects:
  - Radial plots, bubble plots, and shot charts provide overviews of player and team performance
  - Network and inequality visualizations reveal interaction dynamics and scoring distribution across players
- **The goals**: enhance communication, support decisions, and bridge data scientists and basketball experts

Other R Resources

- **ballr** (by Ryan Elmore and Peter DeWitt), an R package that provides simple function for assessing data and tables available on http://www.basketball-reference.com https://CRAN.R-project.org/package=ballr
- **bbr** (by Max Joseph), an R package to scrape data from basketball-reference.com https://CRAN.R-project.org/package=ballr
- **nbaTools** (by Chirag Agrawal), an R package for scraping NBA related data from NBA.com https://github.com/ccagrawal/nbaTools
- **ncaahoopR** (by Luke Benz), an R package for working with NCAA Basketball Play-by-Play Data https://github.com/lbenz730/ncaahoopR
- **BallR** (by Todd W. Schneider), Interactive NBA and NCAA Shot Charts with R and Shiny https://github.com/toddwschneider/ballr
- **NBA_SportVu** (by Rajiv Shah), R code for exploring the NBA SportVu motion data https://github.com/rajshah4/NBA_SportVu

Reference

P. Zuccolotto and M. Manisera (2020) Basketball Data Science. Applications with R. CRC Press. ISBN 9781138600799.

Thank You!