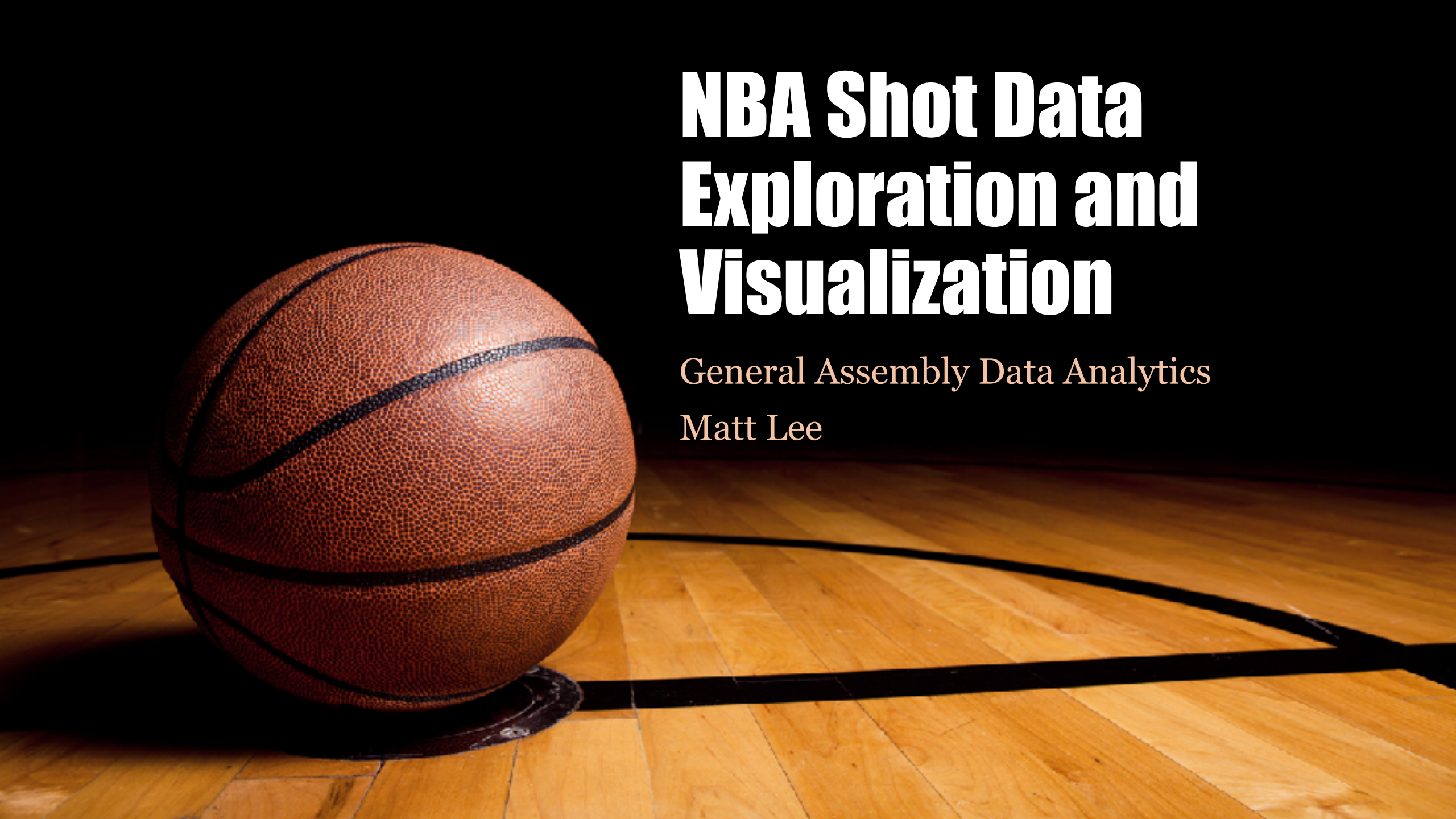


NBA Shot Data Exploration and Visualization

General Assembly Data Analytics

Matt Lee

Shot analysis with selected features

- Attempt to discover player rankings with different features
- Identify player tendencies
- Display statistics not usually found in official statistics sites
- Visualize player shots

Available features from datasets

- Game/Date
- Match Up
- Home/Away
- Final margin
- Shot number
- Period
- Game/shot clock
- Dribbles
- Shot Distance
- Touch Time
- Event Type
- Action Type
- Shot Type
- X/Y Coordinates

Identifying the 'Right' Dataset

- Obtained data from GitHub
  - User collected data from stats.nba.com
  - 30 MB - 203,589 rows, 490 unique players
  - Covers shots from 2014-15 NBA season
- Secondary dataset scraped from stats.nba.com using Python
  - Obtained 11 players' shot information, including X and Y coordinates of all shots in 2015-16 NBA season
- Set up data in Excel and used basic data cleansing techniques
- Imported data to Tableau to generate visualizations and calculated fields

Understanding the Data

Initial Assumptions/Limitations

- **Exclusion of Free Throws**
  - Point totals distorted
  - Unable to identify player strength from drawing fouls
- **Incomplete records (Data set 1)**
  - In addition to exclusion of free throws, some shots were not recorded
    - E.g. Stephen Curry recorded 1,573 points, but official NBA stats show 1,592
      - E.g. December 14, 2014 VS Pelicans, Curry's first shot in OT was not recorded, resulting in a difference of 2 points

Data Cleaning

- Separation of Date, Home Team, and Away Team from Matchup Column
  - E.g. APR 15, 2015 - GSW vs. DEN
- Shot Clock NULL Values
  - When the game clock counts down to 24 seconds, it becomes the shot clock and the shot clock value becomes NULL
  - Filtered 4th quarter and overtimes and replaced NULL shot clock values with ‘CLUTCH’ to measure player effectiveness in high-pressure situations
- Negative Touch Time Values
  - 439 out of 203,589 rows (0.2%) excluded from data analysis

Exploratory Data Analysis

- Shot Number
  - Mean = 6.29
  - Median = 5
  - Max = 43 (Russell Westbrook)
  - Variance = 21.45
  - Standard Dev. = 4.63
- Largest Final Margin
  - 53 (123 – 70 DAL vs. PHI)
- Dribbles
  - Mean = 2.00
  - Median = 1
  - Max = 32 (Mo Williams)
  - Variance = 11.78
  - Standard Dev. = 3.43
- Shot Distance
  - Mean = 13.61
  - Median = 13.6
  - Max = 47.4 (Michael Carter-Williams)
  - Variance = 79.54
  - Standard Dev. = 8.92

Data Visualization with Tableau

Setup

Workbook 1

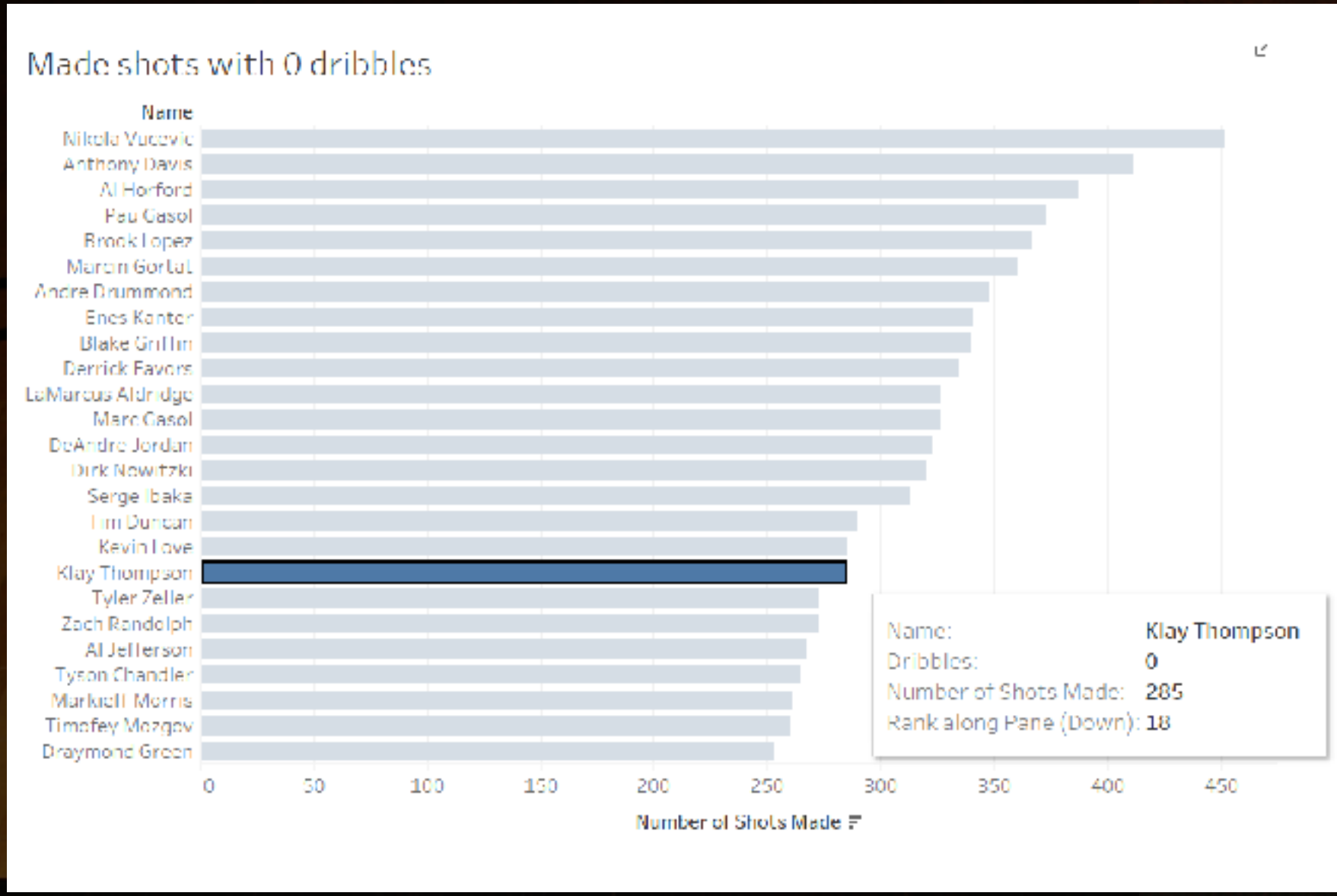
- Left join Shots dataset with CSV file of Player ID, Name, and an image URL of player

Workbook 2

- Inner join CSV file of Player ID, Name, and image URL with CSV file of scraped data of 11 players

Game Data Dashboard

Identifying top catch and shoot player



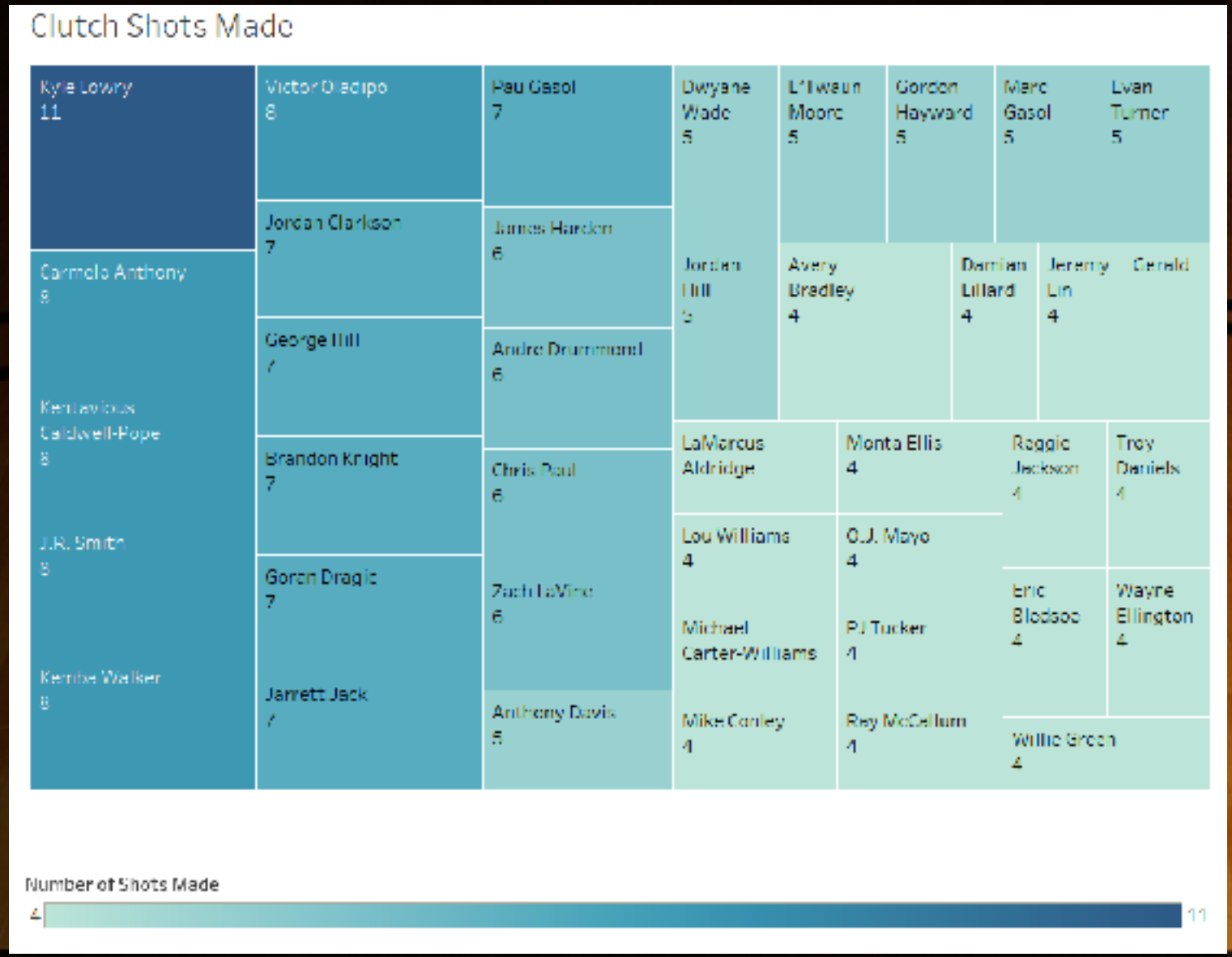
Ranking of Top 25 players composed of either Power Forwards (PF) or Centers (C)

Their short proximity to the basket enables dunks, put-backs, and offensive rebound chances, resulting in 0 dribbles

The exception is **Klay Thompson**, who ranks 18th in this list

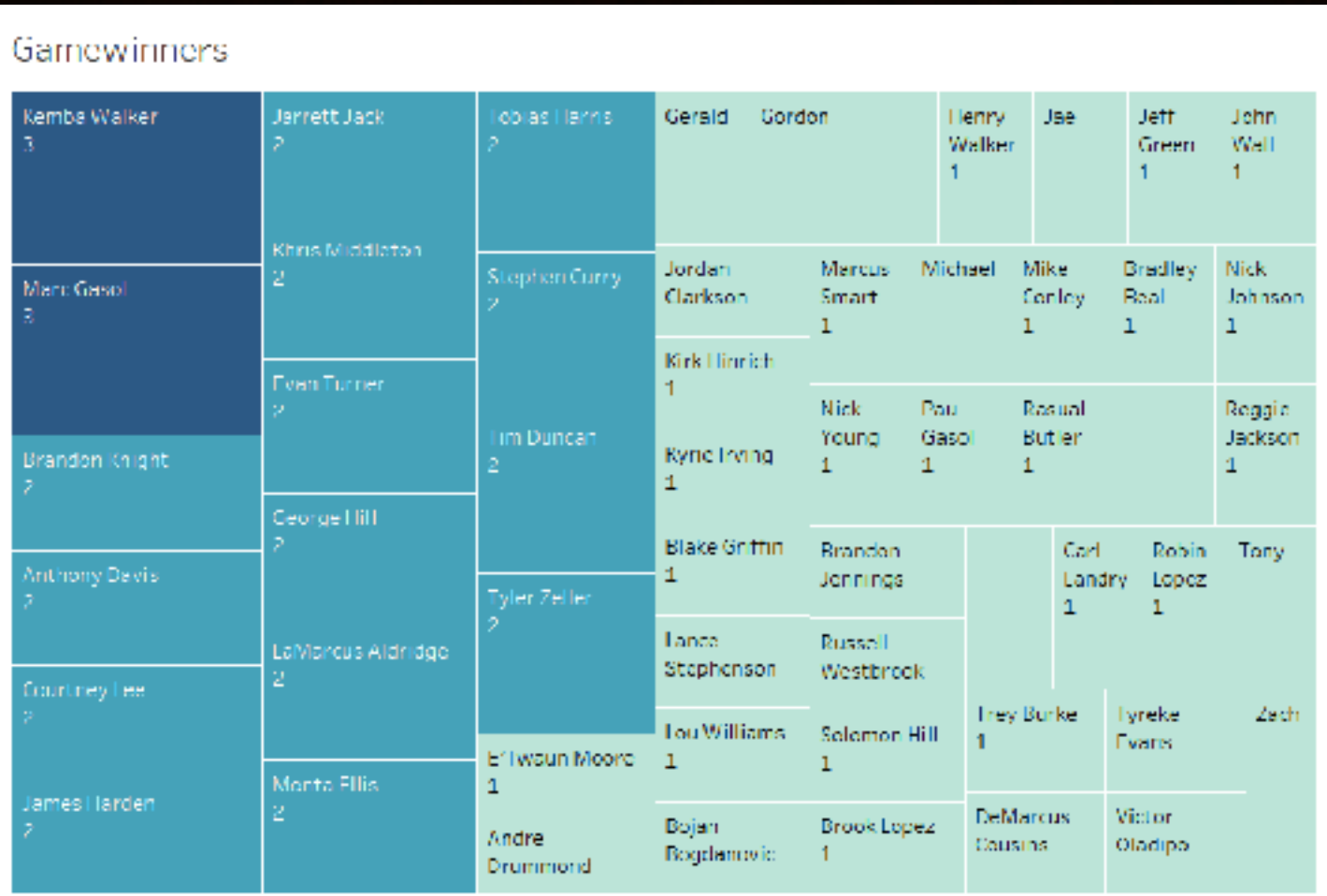
Being able to be placed in this list of PFs and Cs proves that he makes an extremely high volume of catch and shoot shots for a Shooting Guard

Identifying ‘clutch’ shooters



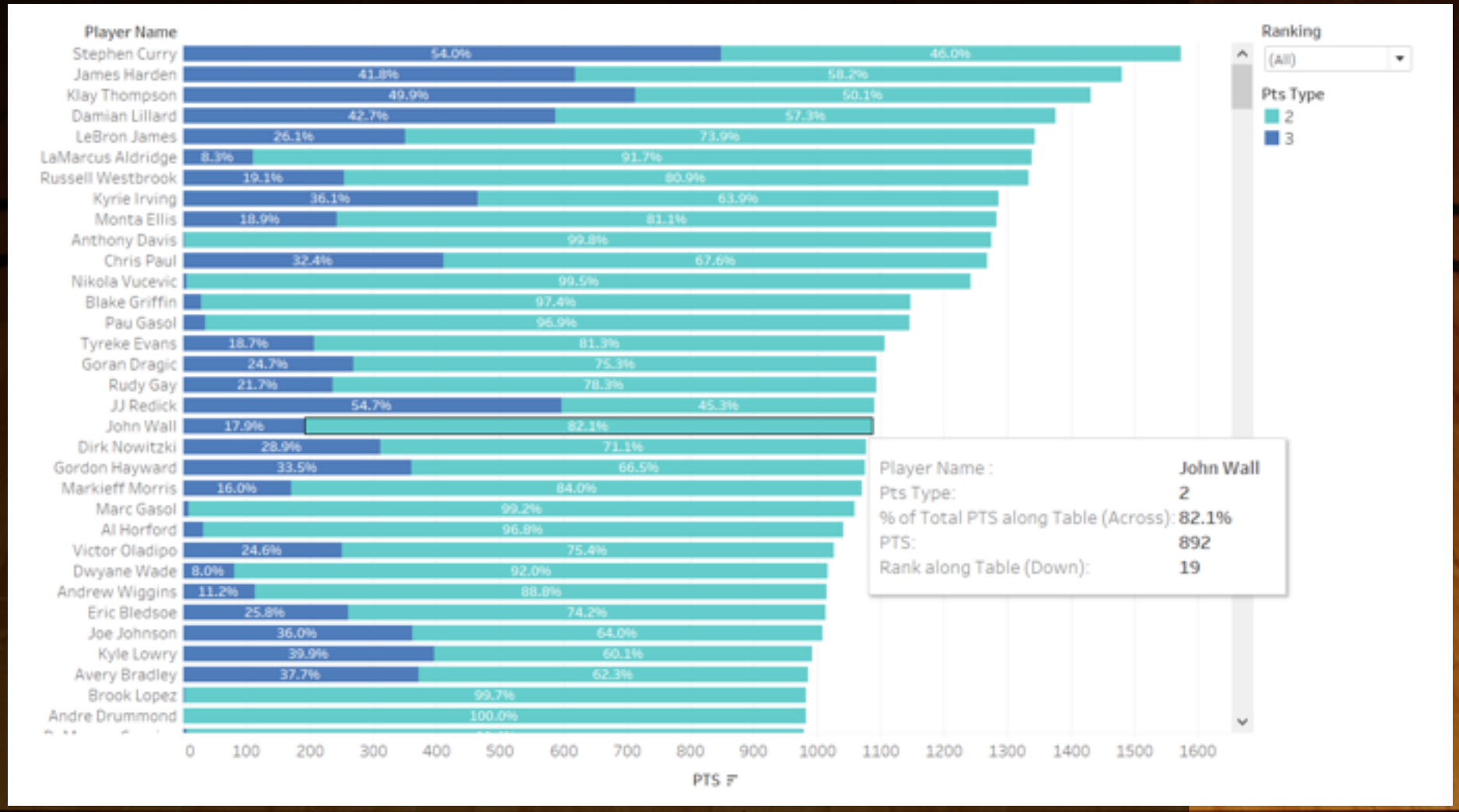
- Labelled shot clock at 24 seconds and under as ‘Clutch’
- Filtered 4th quarter and overtime(s)
- Kyle Lowry leads the league in shots made (11)
- However, does not account for Wins/Losses
  - E.g. Losing team may have free shots as time winds down
  - Does not accurately define ‘clutch’ shooters

Identifying ‘clutch’ shooters



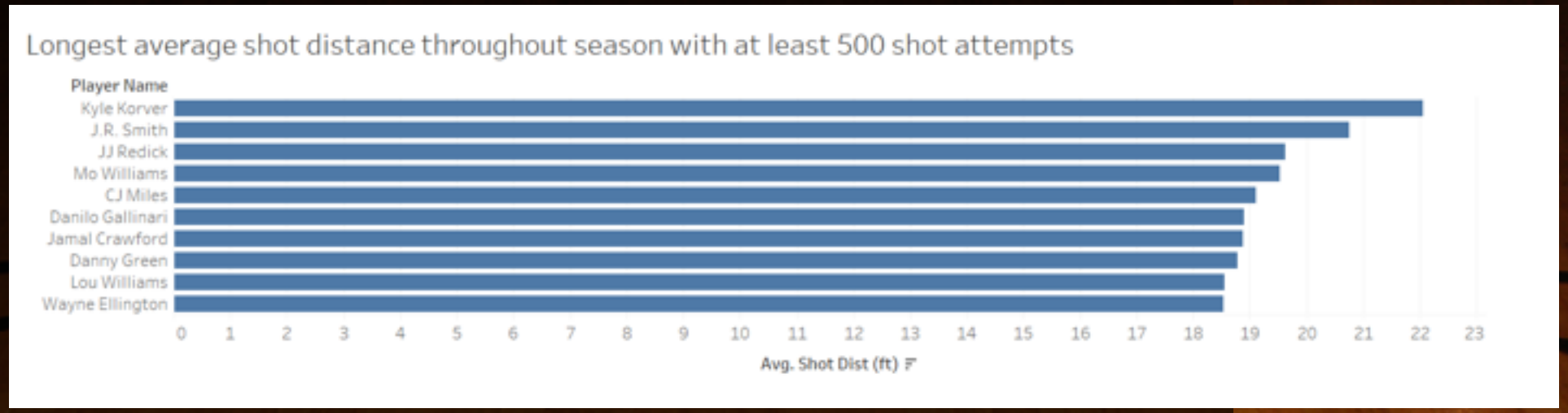
- **Adding on ‘Win’ and Final Margin of 1, 2, and 3 points filter**
- **We identify the leaders in game-winning shots**

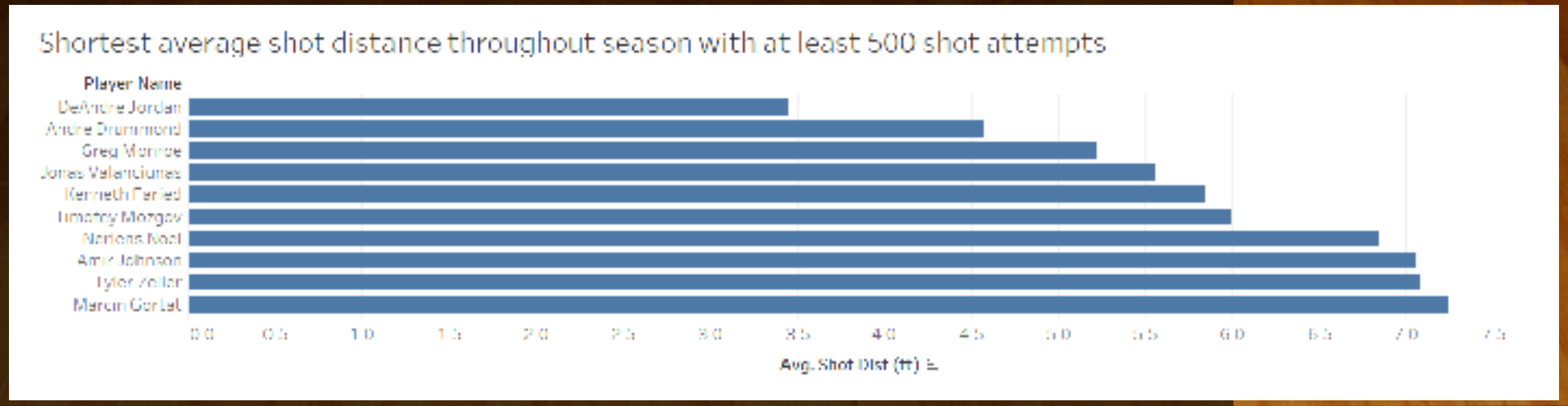
Shot Type Breakdown



- Overall points ranking (kept when filtered)
- Percentage breakdown of 2 and 3 point shots

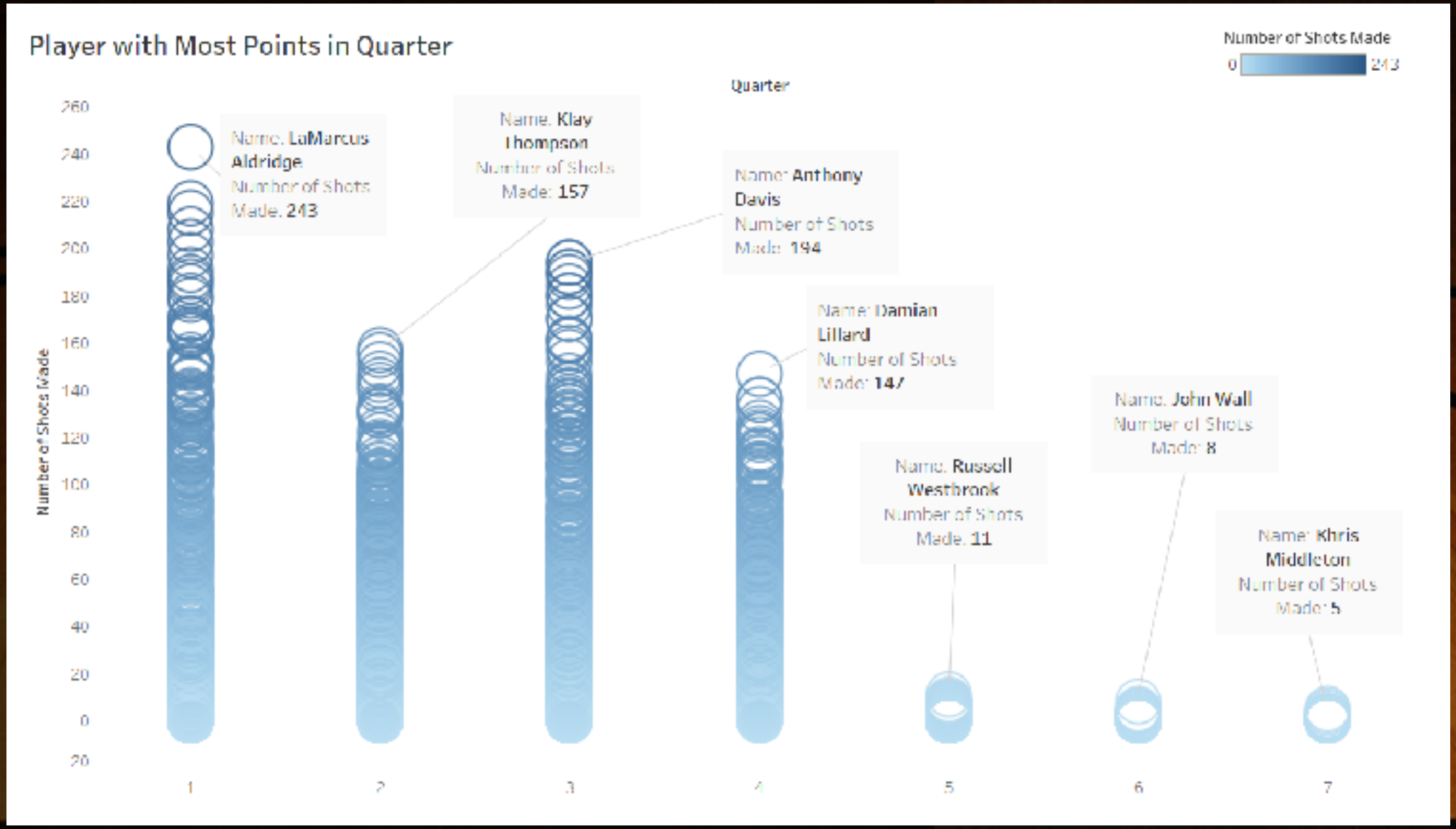
Shot Distance



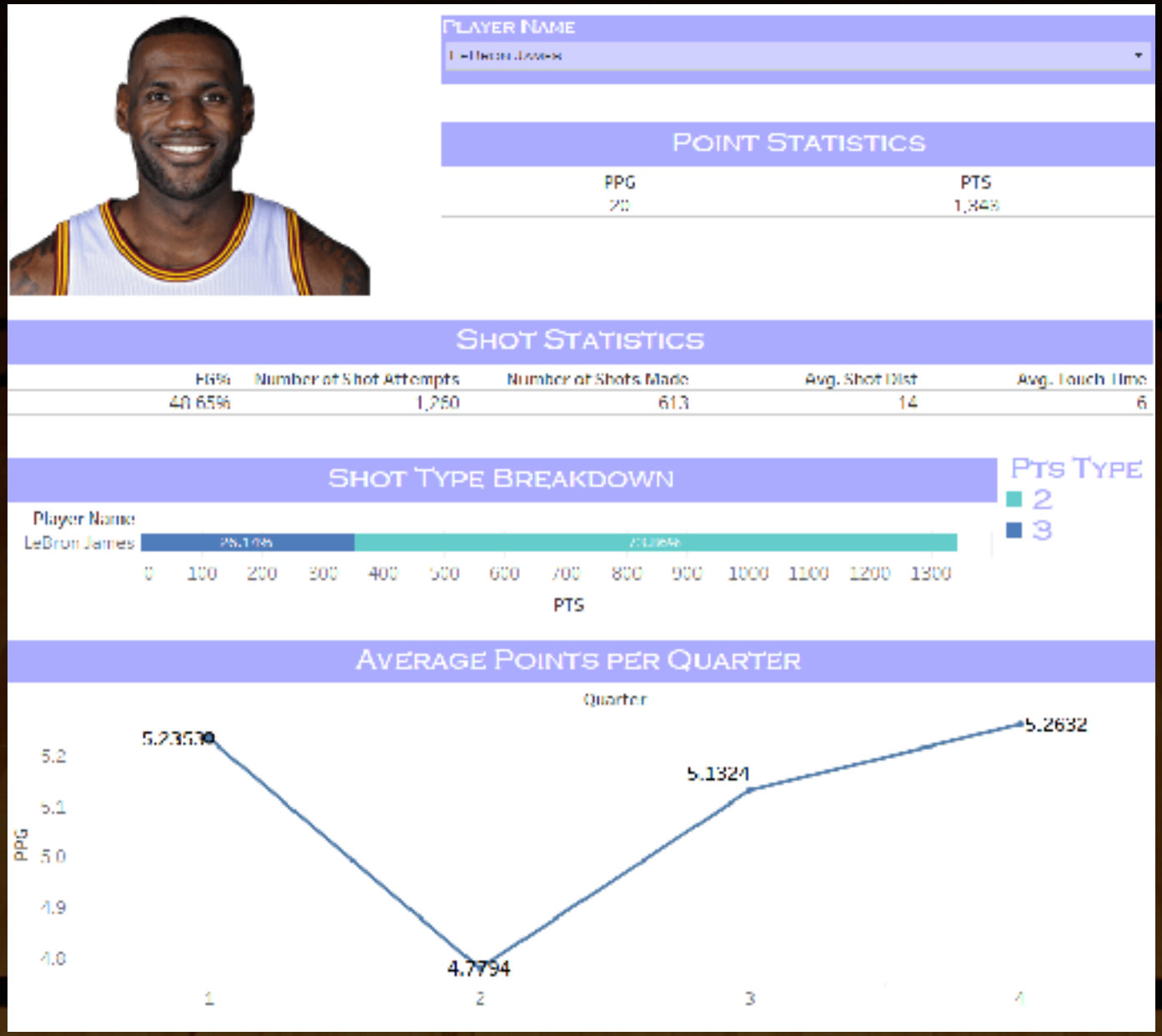


- Tendencies would include shooting jump-shots only
- 3 point line: 22-24 feet
- Free Throw line: 15 feet
- Tendencies include dunks, hook shots, post-moves, short-range jumpers

Leaders by Quarters

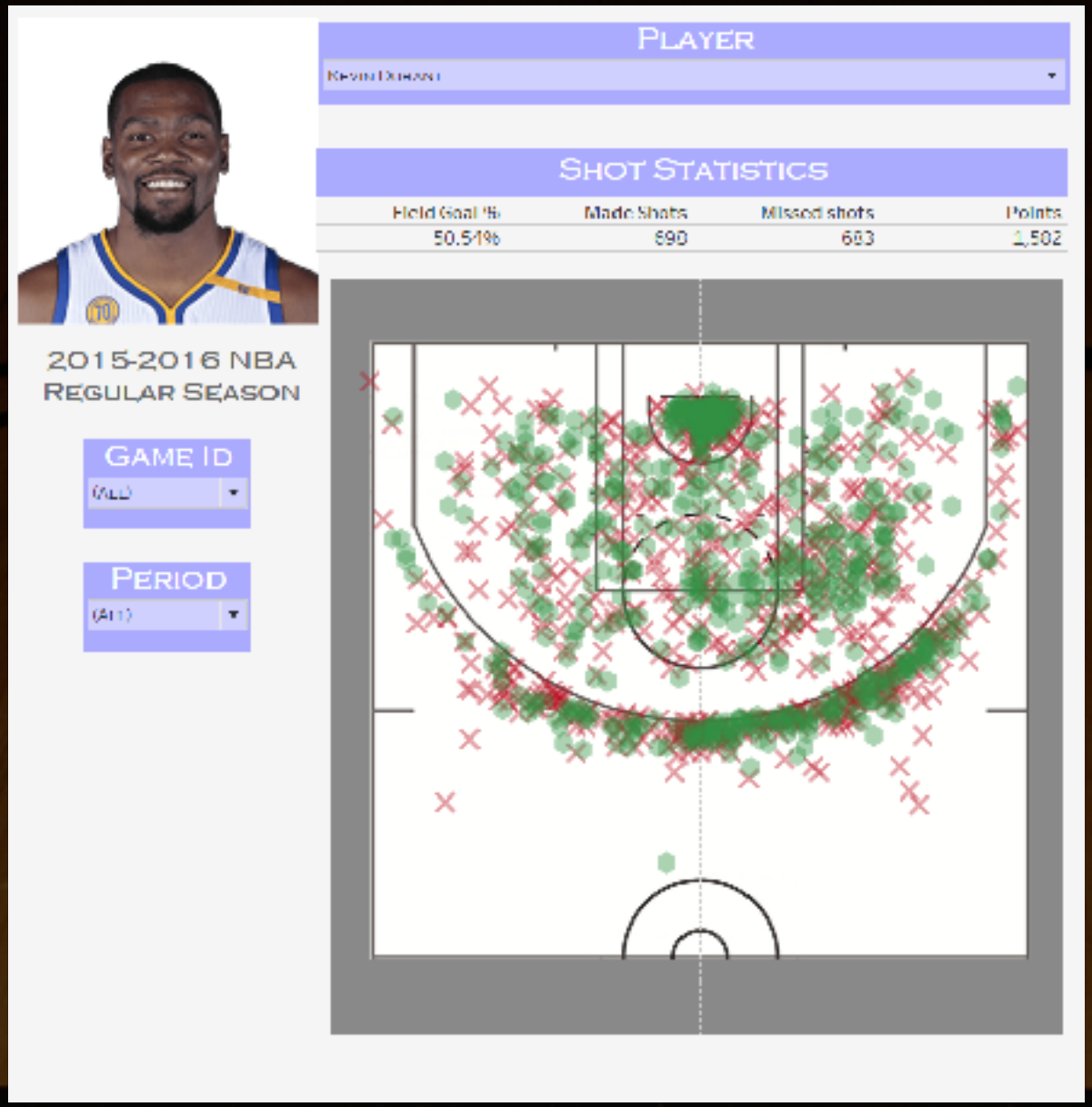


Player Card



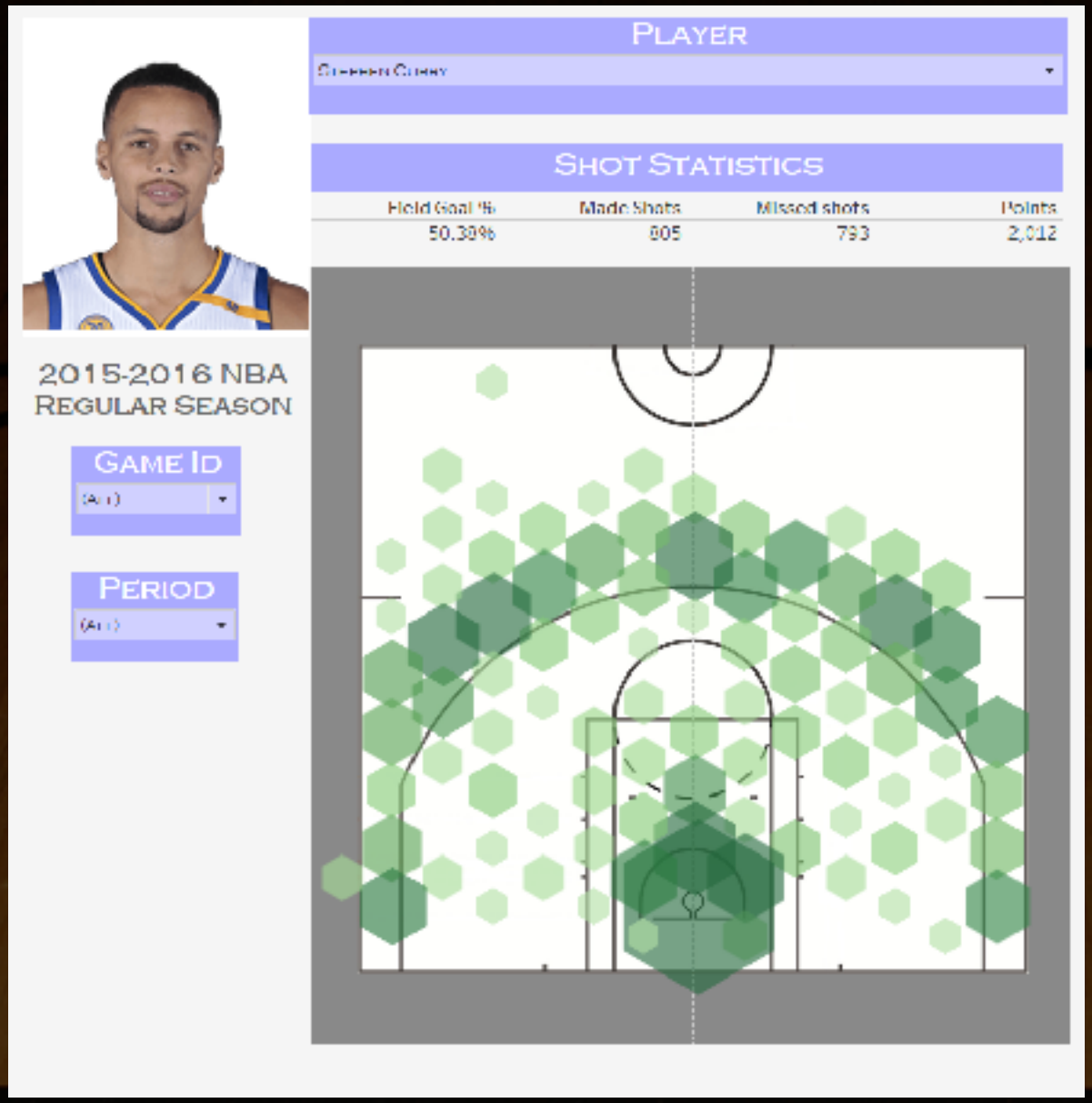
- Created a player profile that show a summary of stats from the data set
- Enabled filter to switch image and stats to any of the 490 players in the data set

Shot Chart v1



- Created player shot chart using second data set
- Enabled filter to switch image, stats, and shot chart to any of the 11 players in the data set
- Included additional filters of Game ID and Period which affect the stats and the shot chart
- Created Calculated Field ‘Points’ using nested IF statements

Shot Chart v2



- Used Hexbins to plot Made Shots
- Different style of visualization with only made shots

Further Analysis

- Collect more data from stats.nba.com
- Analyze more variables, such as ‘Defender Distance’
- Improve Tableau Dashboard design

Thanks!

Q&A

- Sources for help:
  - Dataset - https://github.com/JunWorks/NBAstat
  - stats.nba.com scraping framework - http://savvastjortjoglou.com/nba-shot-sharts.html
  - Dynamic Image Filter - https://community.tableau.com/thread/119079?start=0&tstart=0
  - Keeping Overall Rank - https://community.tableau.com/message/516072?et=watches.email.thread#516072
  - Hexbin Tutorial - http://tableautim.com/tableau-hexbin-tutorial-with-hexbin-shape-pack/