Data Visualization Report of NBA Dataset

Intended audience:

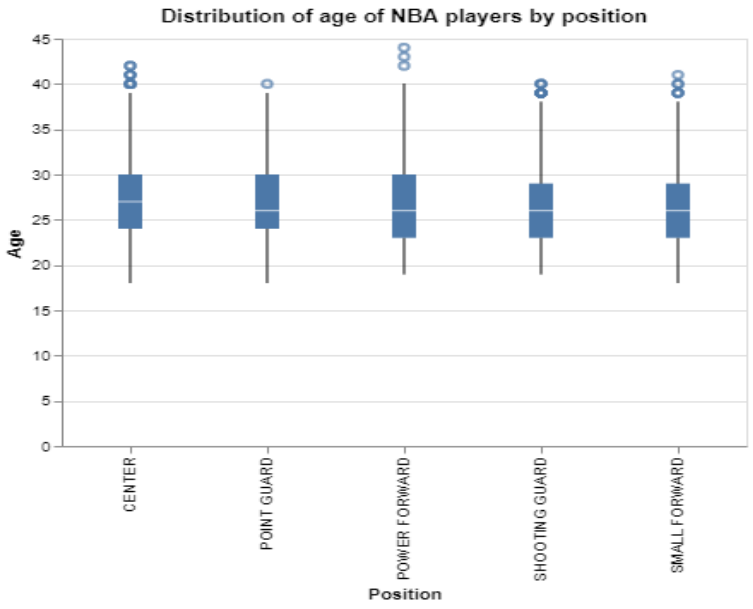
Our visualization based on structured tasks aims to provide the reader with in-depth knowledge of various facets of basketball. The visualization helps evaluate the performances in particular stats for teams and players. Thus, our target audience is coaches/ team managers. Viz 1 helps to identify the prime ages of players. This stat is important for coaches and managers when drafting teams. Next is the multiview of Viz 2 and Viz 3. This would help identify the trends of team performances in the past. Managers might use this to replicate tactics from the most successful years. Viz 4 and Viz 5 addresses the issues of which players to draft based on their past performances. Again this is helpful for drafting purposes for the team manager / management.

List of all Tasks:

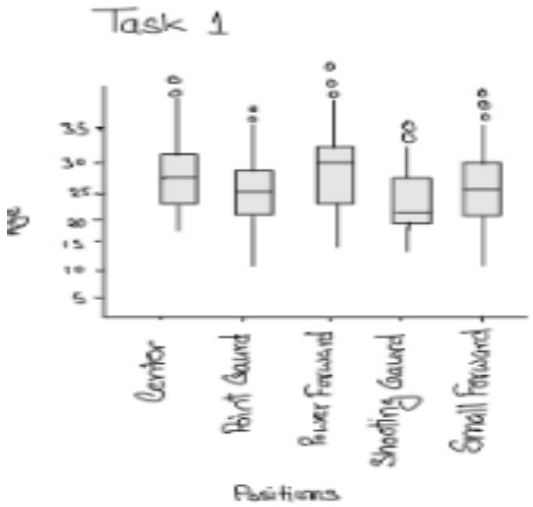
- What is the distribution of age of players (age) across different positions (position)?
  - Characterize Distribution
- What are the all-time top teams with the highest total 3 PM in Eastern and Western Conference?
  - Filter
- How have teamwise 3PA (attempted_three_point_field_goals) changed over the season (year)?
  - Correlate
- What were the highest assists per game across seasons?
  - Compute Derived Value
- How efficient are prolific playmakers in the NBA?
  - Find anomalies
- Lebrons James' scoring performance for his teams over the years.

1. Viz #1

a. Screenshot:



b. Previous Iterations:



c. List of tasks the viz addresses:

i. What is the distribution of age of players (age) across different positions (position)?

d. Explanation of visualization choices from theoretical principles

i. marks: boxplot to compare the age distribution of players in different positions, easy-to-identify range, outliers and median

ii. channels:

- x-axis: encoding position for categorical variable
- y-axis: encoding age as a quantitative variable
- tooltip: name and age given to outliers, detailed overview of boxplot(Q1, Q3, etc…)

iii. characteristics of channel

- **popout** as shape(circle)
- x-axis variables are aggregated and grouped by **similarity**
- **accuracy** may be compromised if we were to compare the spread between distributions
- age and position are **integral**
- the x-axis has 5 unique steps we can perceive

iv. interaction: N/A

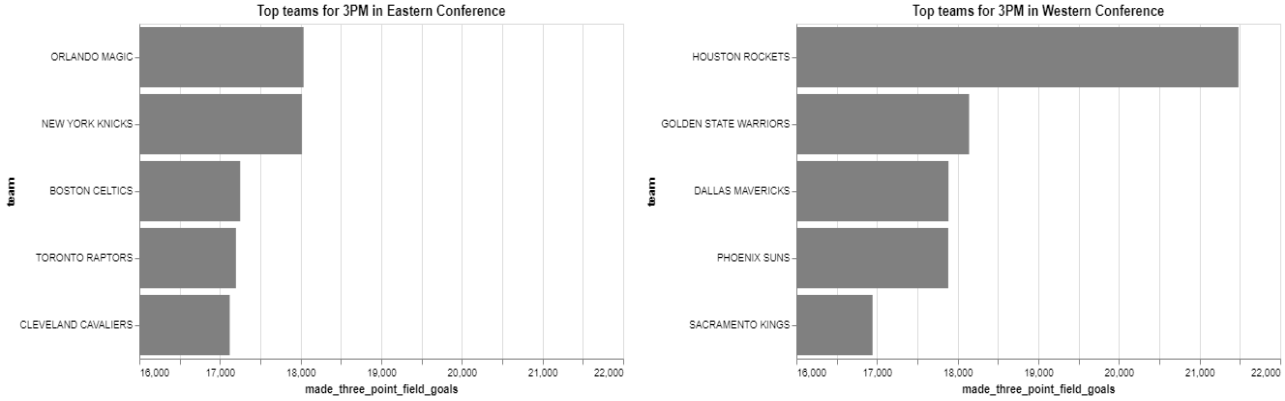
v. characteristics of interaction/interactivity: N/A

vi. critique

- The visualization overall answers the given task effectively but is unable to determine the distribution at a specific age(shape of distribution).

2. Viz #2

a. **Screenshot:**



b. Previous Iterations:

i. N/A (We changed our task)

c. List of tasks the viz addresses:

i. What are the all-time top teams with the highest total 3 PM in Eastern and Western Conference since 2001?

d. Explanation of visualization choices from theoretical principles

i. marks: the bar was used to easily compare the differences in 3PM along a common axis

ii. channels:

- x-axis encodes three-pointers made
- y-axis encodes the team as a categorical variable

iii. characteristics of channel

- no **popout** is used
- conferences the teams are in are aggregated and grouped by **containment(**different graphs)
- **accuracy** is good as we can precisely tell the difference between encoded items within the conference(may be difficult to compare between conferences)
- three-pointers by the team and team name are **integral** three-pointers made by team variable become ambiguous as it is associated with the team name
- the x-axis has 10(10 teams) unique levels we can perceive

iv. interaction: **addressed in VIZ #3**

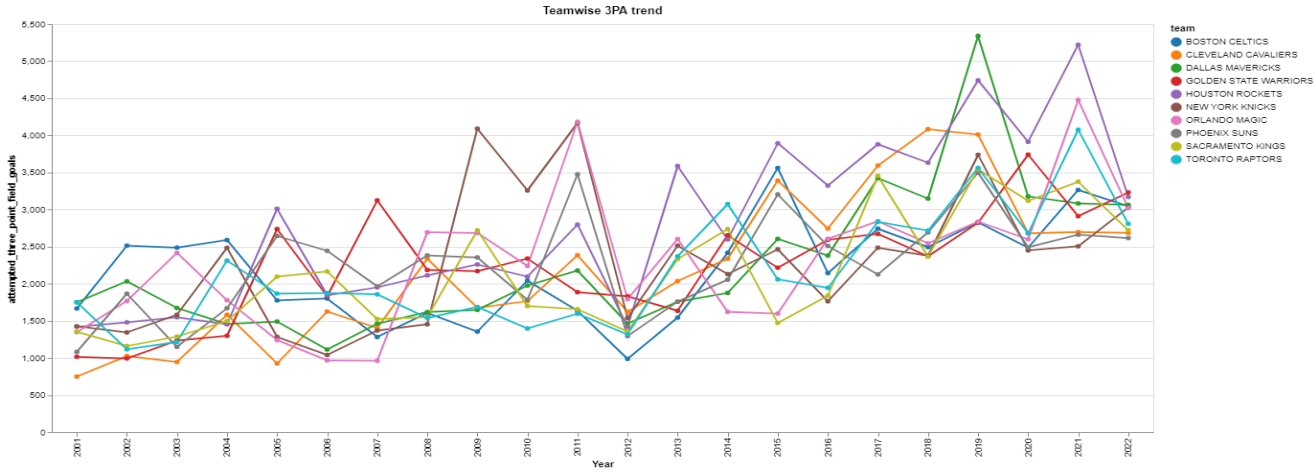
v. Characteristics of interaction/interactivity: **addressed in VIZ #3**

vi. critique

- The visualization overall addresses the task associated with it effectively. You can easily tell the all-time top teams for 3 PM made in the respective conferences.

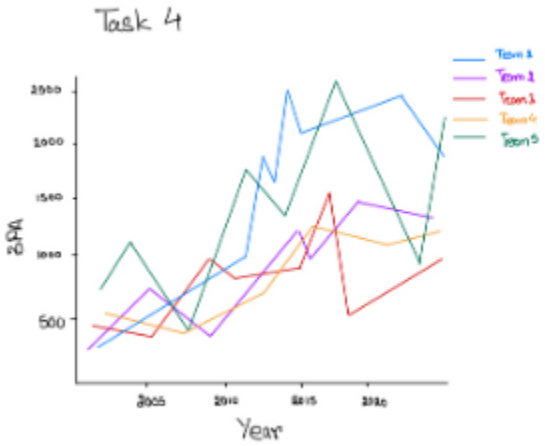
3. Viz #3

a. Screenshot:



b. Previous Iterations:

i.



c. List of tasks the viz addresses:

i. How have teamwise 3PA (attempted_three_point_field_goals) changed over the season (year)?

d. Explanation of visualization choices from theoretical principles

i. marks: line because it is effective in showing trends

ii. channels:

- x-axis: year as ordinal
- y-axis: three pointer attempts
- color: top 10 teams in attempts
- opacity: based on clicking interaction
- tooltip: shows team, 3PA, 3PM, 3% efficiency

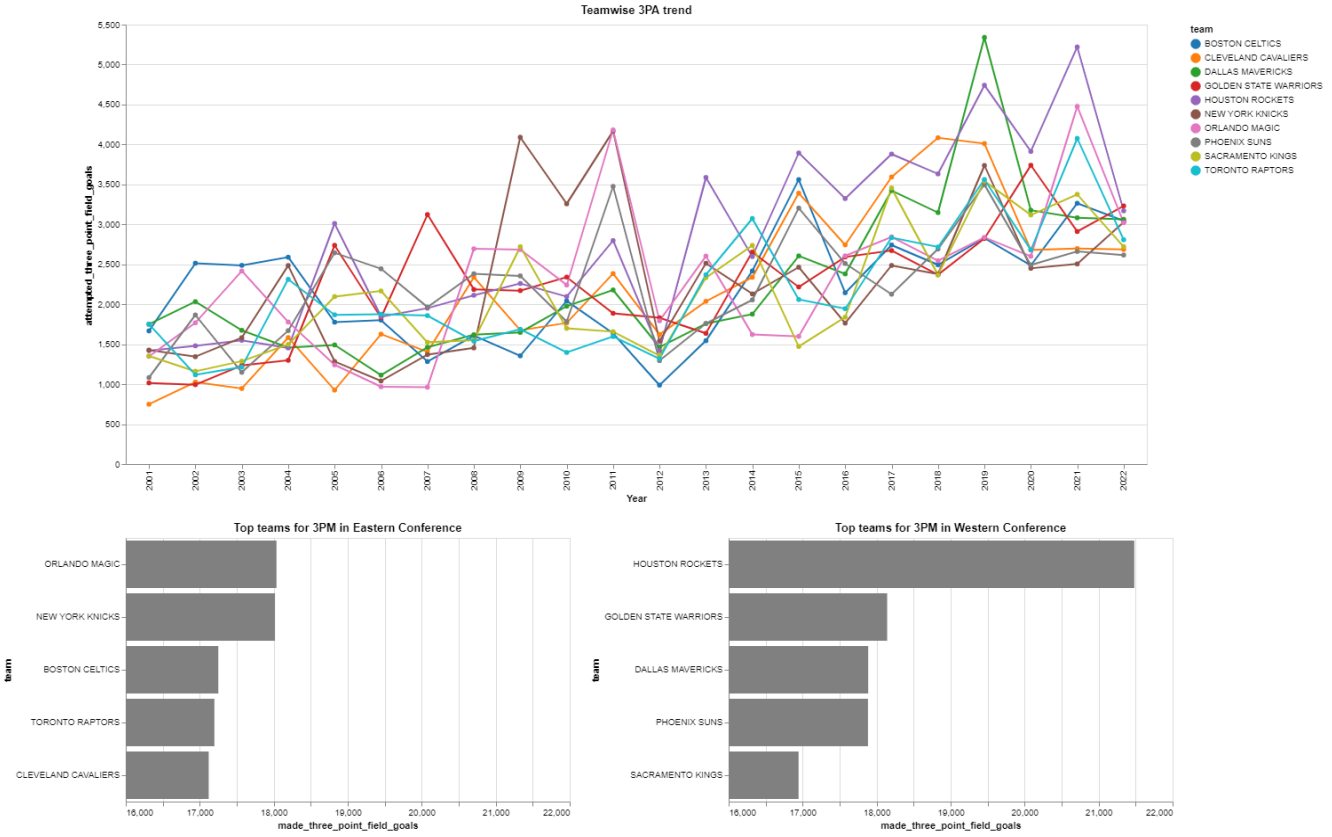
iii. characteristics of channel

- **popout** using circle shape

- grouping by **connection**
- 3PA and year are perceived as **integral** variables
- **accuracy** is not good with lots of trends line, however, it will be fixed with the implementation of interactivity
- the y-axis has 10(10 teams) unique levels we can perceive

iv. interaction

- yes, when **Viz #2** combined with **Viz #3**
- click the line chart points or bars to highlight a specific team



v. characteristics of interaction/interactivity

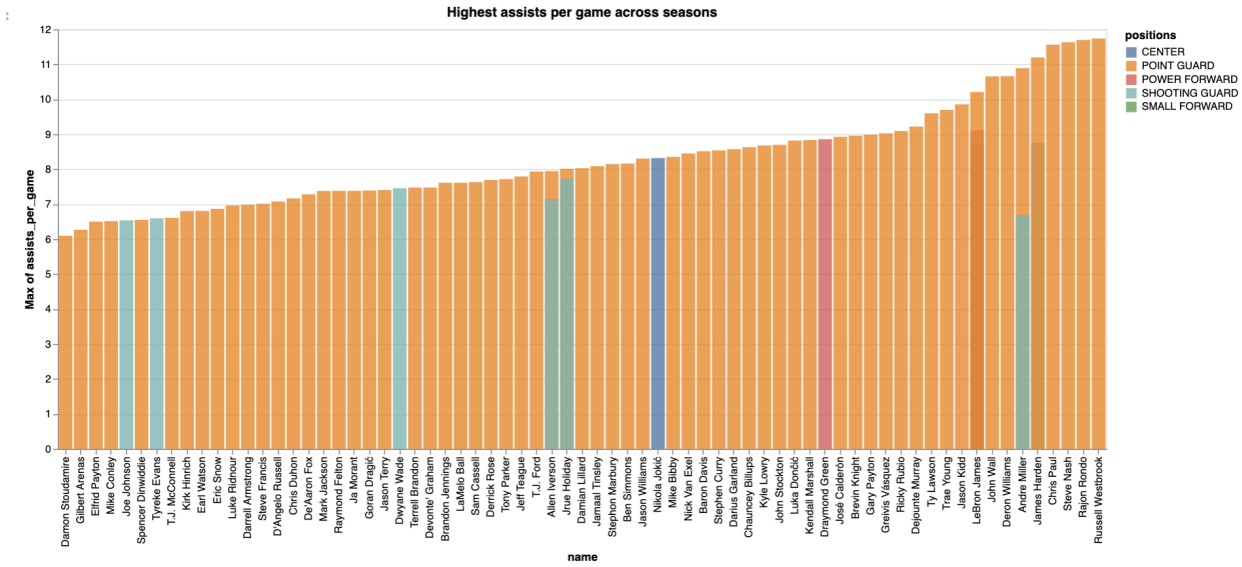
- action: select
- event: click/ shift-click
- reaction type: highlight
- views: juxtapose
- interaction coupling: bidirectional
- views data: share navigation
- interactivity action: direct focus
- interactivity reaction: immediate, propagated spread, discrete flow
- affordance: lack of affordance, invisible functionality and may be difficult to discover

vi. critique

- The visualization effectively answers the task, however, it does violate the rule of using more than 5 colours(interactivity counters this flaw).

4. Viz #4

a. Screenshot:



b. Previous Iterations:

i. N/A (We changed our task)

c. List of tasks the viz addresses:

i. What were the highest assists per game across seasons?

d. Explanation of visualization choices from theoretical principles

i. Marks: bar as it's easy and effective in comparing differences of max assist per player

ii. channels:

- colour used for position
- x-axis used for player names
- y-axis is used for max assists per game

iii. characteristics of channel

- **popout** using colour
- grouping by **proximity**
- player name and max assist attributes are perceived as **integral**
- **accuracy** is good as you can easily determine who has the higher max assist as it is sorted
- colour has 5(5 positions) unique levels we can perceive

iv. interaction

- **addressed in VIZ #5**

v. characteristics of interaction/interactivity

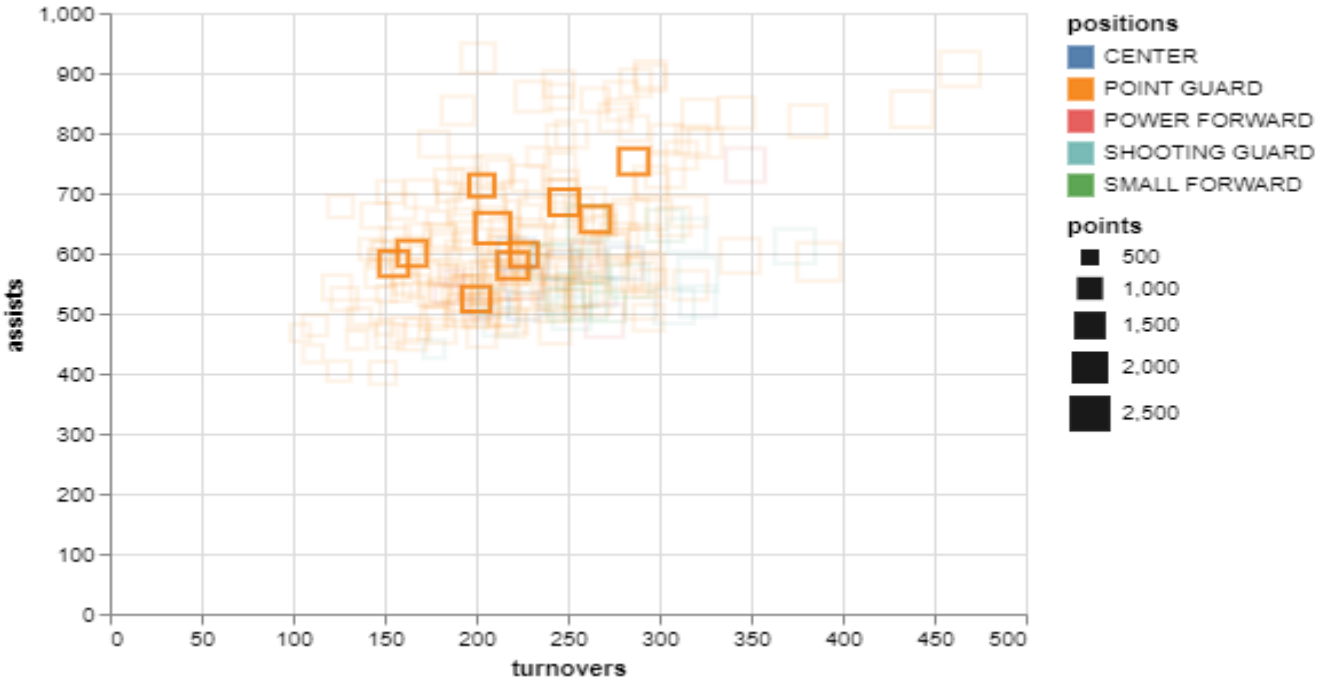
- **addressed in VIZ #5**

vi. critique

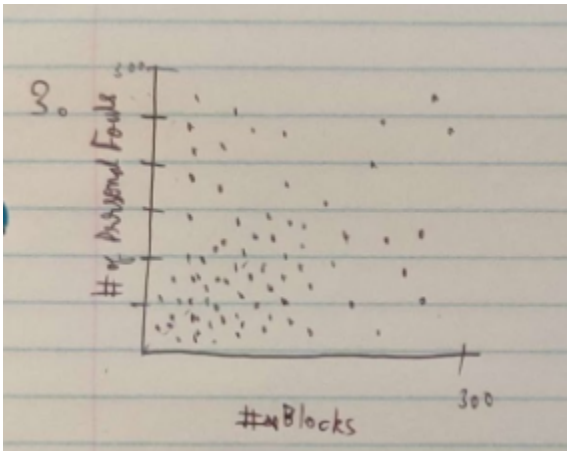
- good that less than 5 colours are used
- visualization effectively addresses the task
- there may be too many players in the bar chart(use a smaller subset maybe top 10?)

5. Viz #5

a. Screenshot:



b. Previous Iterations:

i. 

- changed to assists/turnovers from blocks/fouls

c. List of tasks the viz addresses:

i. How efficient are prolific playmakers in the NBA?

d. Explanation of visualization choices from theoretical principles

i. Marks:

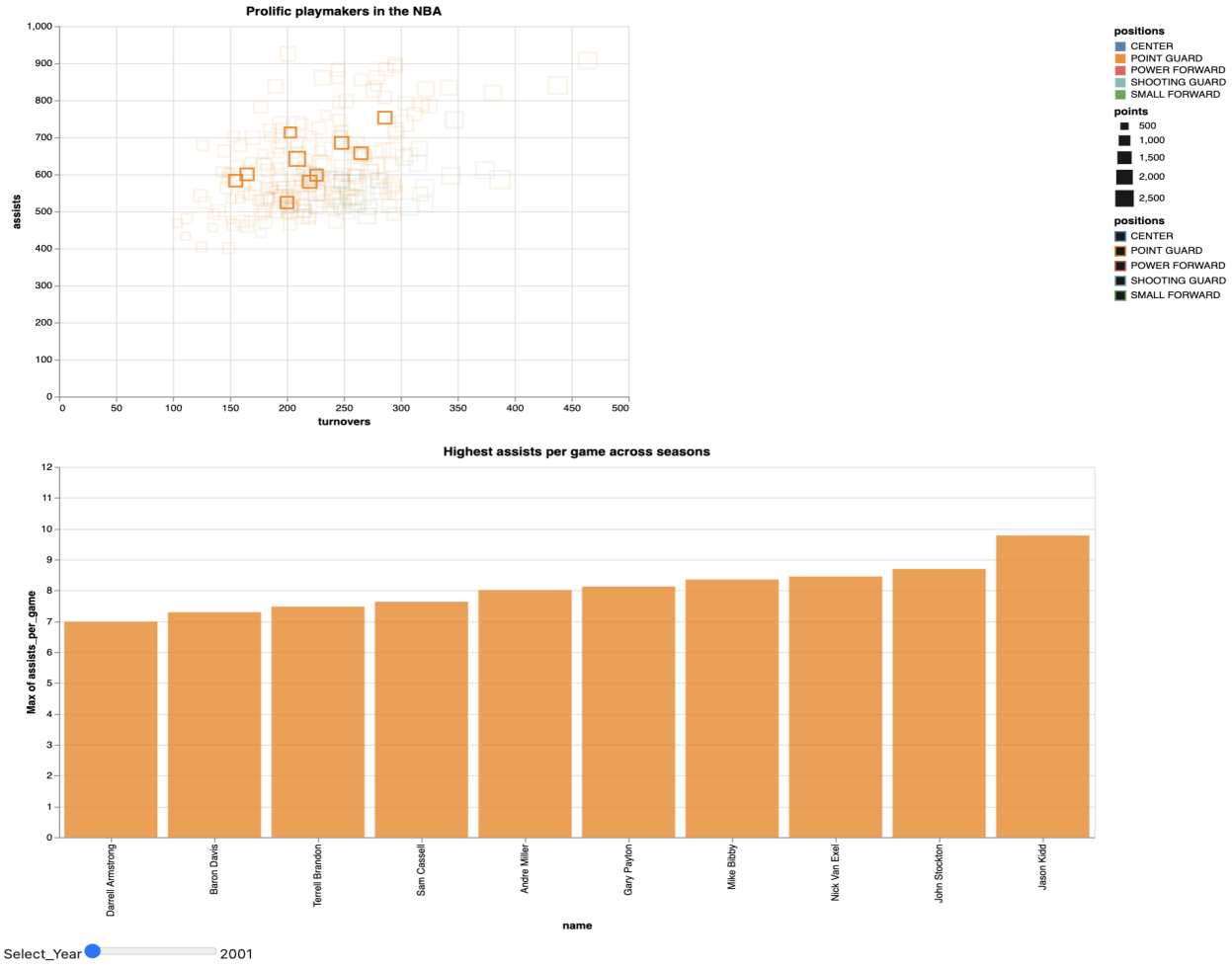
- square as it's easy to differentiate the size of each square compared to circles

- scatterplot of squares is used as assist to turnover ratio is a known correlation in basketball

ii. channels:
- y-axis for assists
- x-axis for turnovers
- colour for position
- opacity based on if a year is being selected
- size based on points
- tooltip for the assist-to-turnover ratio, name and team associated with the player

iii. characteristics of channel
- **popout** using size and opacity
- grouping by **proximity**
- assist and turnover attributes are perceived as **separable**
- **accuracy** is good, low turnover to assist ratio tends to be top left quadrant of the graph and can be realized with tooltip
- colour and size have 5 unique levels we can perceive

iv. interaction
- yes, when **Viz #4** combined with **Viz #5**
- using a slider to see the efficient players combined with the highest assist per year



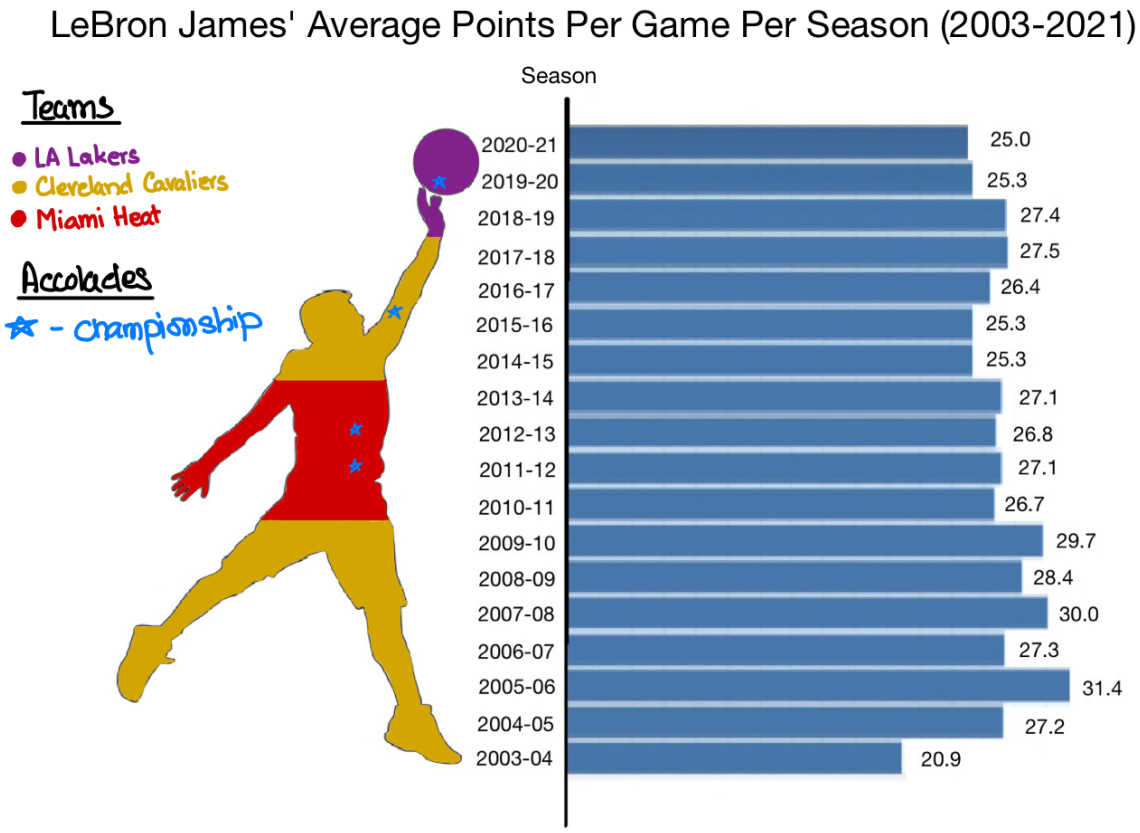
v. characteristics of interaction/interactivity
   - action: filter
   - event: drag
   - views: juxtapose
   - interaction coupling: unidirectional
   - views data: share data
   - interactivity action:
     a. presence, explicit
   - interactivity reaction:
     a. immediate, continuous flow, propagated
   - affordance: good affordance as functionality is easy to discover

vi. critique
   - visualization effectively answers the task as we can effectively determine whether players are prolific playmakers when they have a high assist season along with a low turnover-to-assist ratio
   - Use of squares as mark helps distinguish size difference more easily, additionally the user can zoom in on specific areas and points of interest if desired.

- Although position is not a focus of our task, it is encoded as colour specifically for interactivity with viz #4, allowing us to quickly identify a player on both plots.

6. Novel Visualisation

a. Screenshot:



b. **Previous Iterations:** NA

c. List of tasks the viz addresses:

i. Lebrons James' scoring performance for his teams over the years.

d. Explanation of visualization choices from theoretical principles

i. marks:

- Bar because it helps compare LeBron's PPG per season
- Figure because it allows us to see which team he was playing or and helps us remember this visualization.

ii. channels:

- y-axis for Year
- Color for team
- x-axis for points per game by year
- Shapes for titles and accolades

iii. characteristics of channel

- Expressive: Usage of a figure helps make the plot memorable.

iv. critique

- visualization effectively answers the task as we can effectively determine whether players are prolific playmakers when they have a high assist season along with a low turnover-to-assist ratio