Assignment 2 : Exploratory Data Analysis

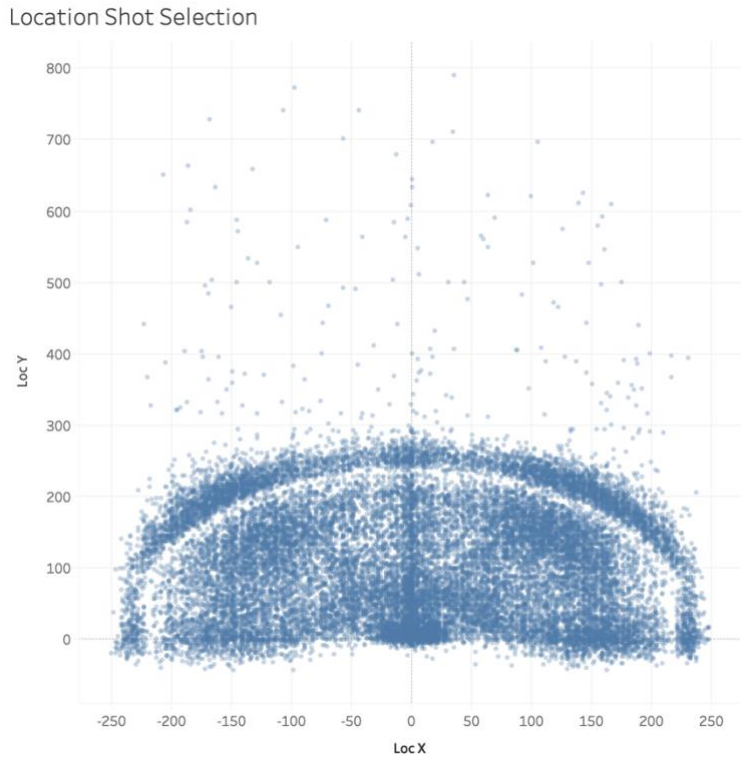
Juan Alvarez

Domain: Coming from Los Angeles and in light of the NBA Playoffs occurring at this moment, I wanted to reflect on how great Kobe Bryant was to watch when he was on the Lakers. One thing that stands out when recalling his matches was how he was able to close out games in the last moments of a game. That being said, I wanted to explore this visually and see what other questions may come up.
Initial Question: How "clutch" was Kobe Bryant when he played for the Los Angeles Lakers? (Clutch: Being able to perform under pressure)
Data Info: This dataset was established for a Kaggle competition three years ago.
https://www.kaggle.com/c/kobe-bryant-shot-selection/
This set contains 20 years' worth of location and circumstances of every field goal attempted by Kobe Bryant. The column names are pretty self-explanatory:

action_type	loc_x	playoffs
combined_shot_type	loc_y	season
game_event_id	minutes_remaining	seconds_remaining
game_id	period	shot_distance
lat	lon	shot_made_flag (this is what you are predicting)
shot_type	shot_zone_range	team_id
shot_zone_area	matchup	team_name
shot_zone_basic	opponent	game_date
shot_id		

Already looking at this dataset, I knew that I had to make some transformations. I wanted to convert the location data of the shot to a distance and angle value that is easier to interpret. Let's look at the location data of each shot.

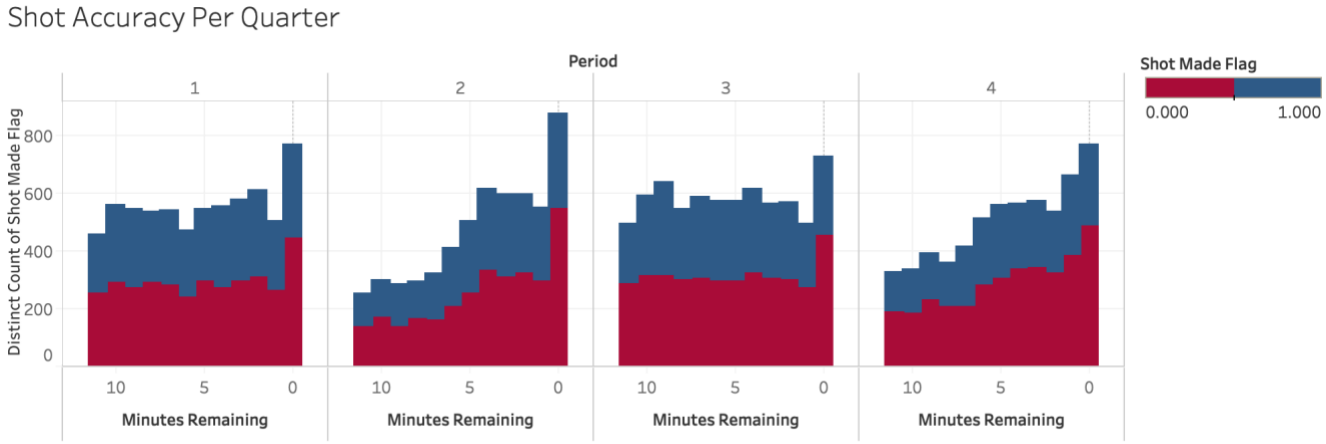
Vis 1) Lets just take a look at the shots he took, just to see what we are working with.



Loc X vs. Loc Y.

On a visualization you can see how difficult a shot is, but if we were looking at the raw latitude and longitude data then it would be impossible to instantly view the difficulty of the shot.
**On second thought, adding an angle value to the shot relative to the basket would not be beneficial for the use of exploratory data analysis and helping us answer our hypothesis.

Vis 2) Lets attempt to answer our high-level question of how clutch Kobe was. One way we can do this is by seeing how accurate his shots were when the game was in the 4th quarter.



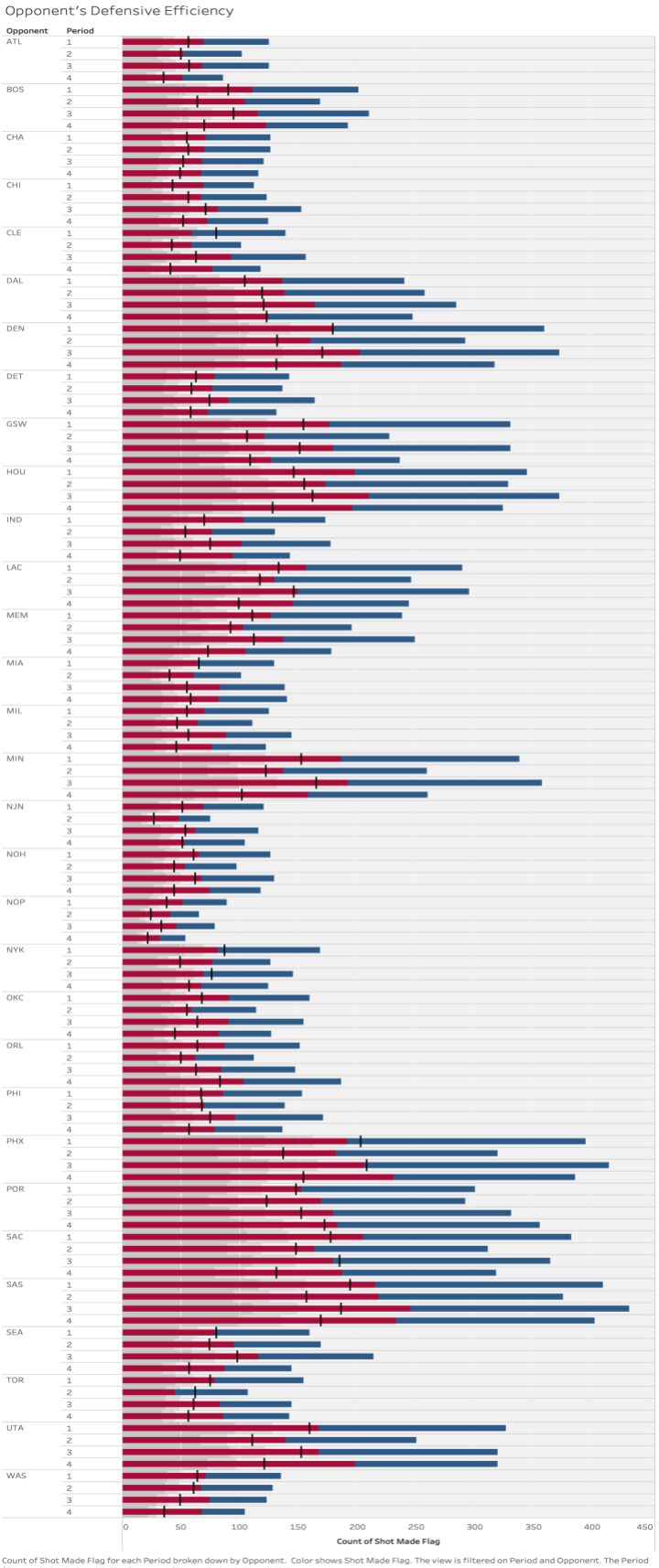
Period and Minutes Remaining vs. Distinct count of Shot Made Flag. Color shows Shot Made Flag. The view is filtered on Period, which excludes 5, 6 and 7.

In this visualization, we can see a couple of things. First, we can see a trend of more shots taken in the end of the 2nd quarter heading towards halftime, and the 4th quarter heading towards the end of the game. Although the accuracy in his shots seemed to slightly decrease as the minutes remaining in the quarter decreased, we can see that he took more shots making the total amount

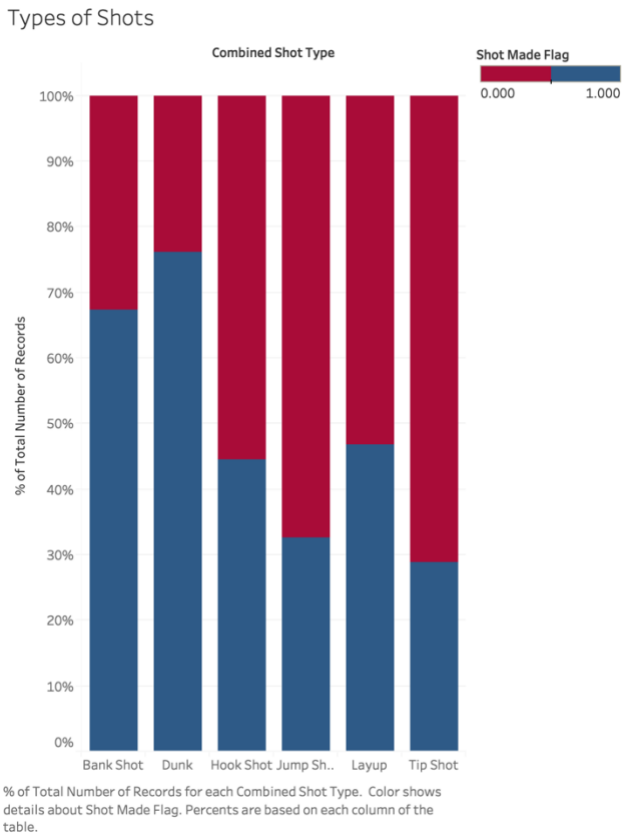
of shots made towards the final stretch of the game increase. Looking at the last bar in the 4th quarter (where there are zero minutes remaining or just seconds left in the game), he had 649 missed shots out of a total of 931. That's a 30% accuracy score, where in the other final stretches of the beginning quarters he had an accuracy score of about 35%, 31%, and 31% retrospectively. A 1% dip in accuracy is remarkable when taking into account that he is forced to make more difficult shots as the game is closing and the opponent's defense is naturally more aggressive.

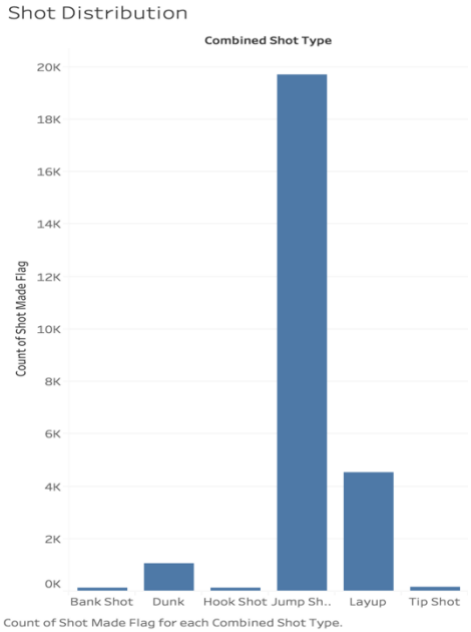
Vis 3) One question that came up when making visualization 2 was that we did not factor in different opponents. Every team has their strength and weaknesses, and defensive ability is one of them. So, let's look at Kobe's accuracy based on who he was playing against. **As new teams came into the NBA, I decided to not show opponents where there is a nominal amount of data**

We can see that the black tick represents the median of the shots made flag. Per opponent we can see that tick change depending on the quarter. The more left it is, then the more efficient the defense was at stopping him. We could not combine visualization 2 and 3 together, because that would unfortunately create a convoluted and illegible graphic.

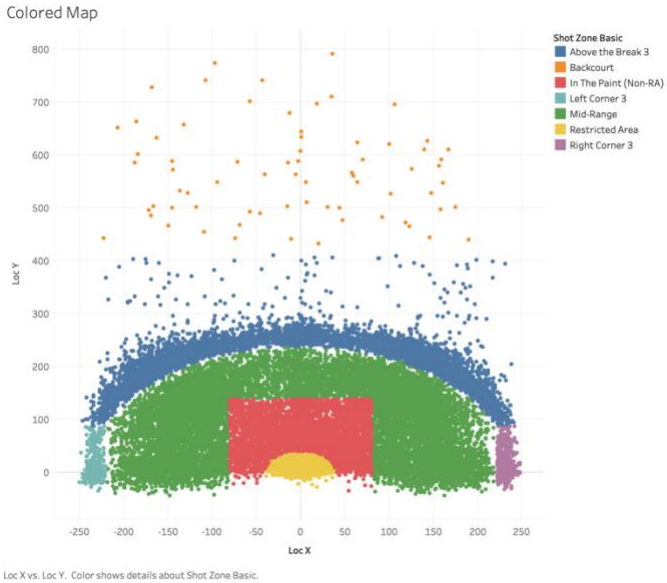


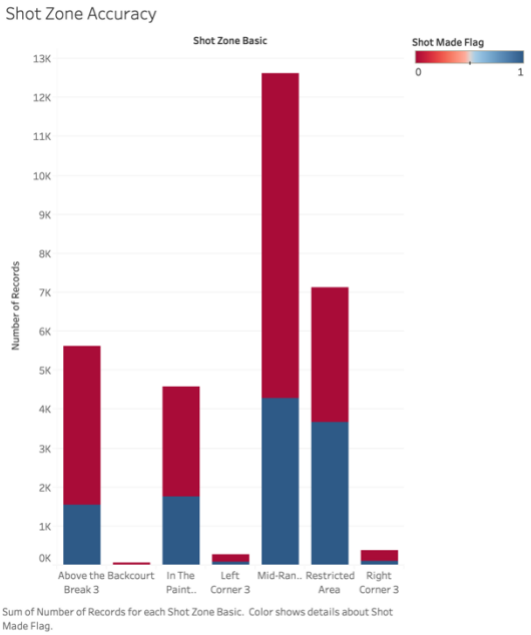
Vis 4) I wanted to see what types of shots he took in his career, and which ones he utilized were more effective. This visual shows how accurate he was in 6 different shot types: bank, dunk, hook, jump, layup, and tip. Having knowledge on the type of player he was, and general basketball knowledge, I can confirm that Bryant was a master of creating space and pulling off a jump shot, which is why this is the most utilized shot (Vis 5).



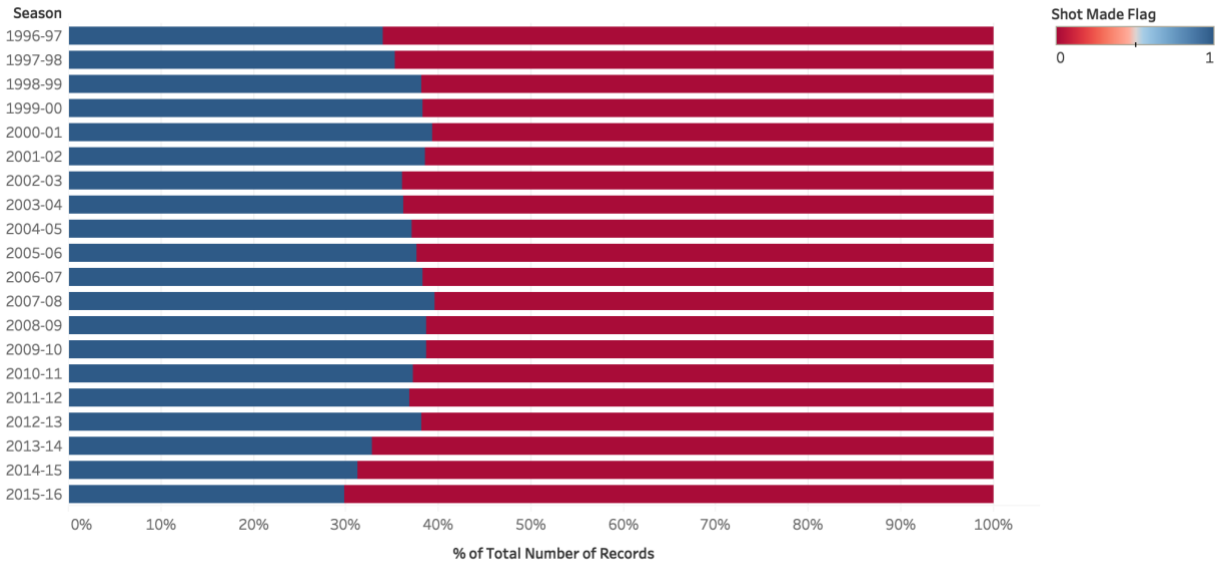


More Visualizations:





Accuracy Per Season

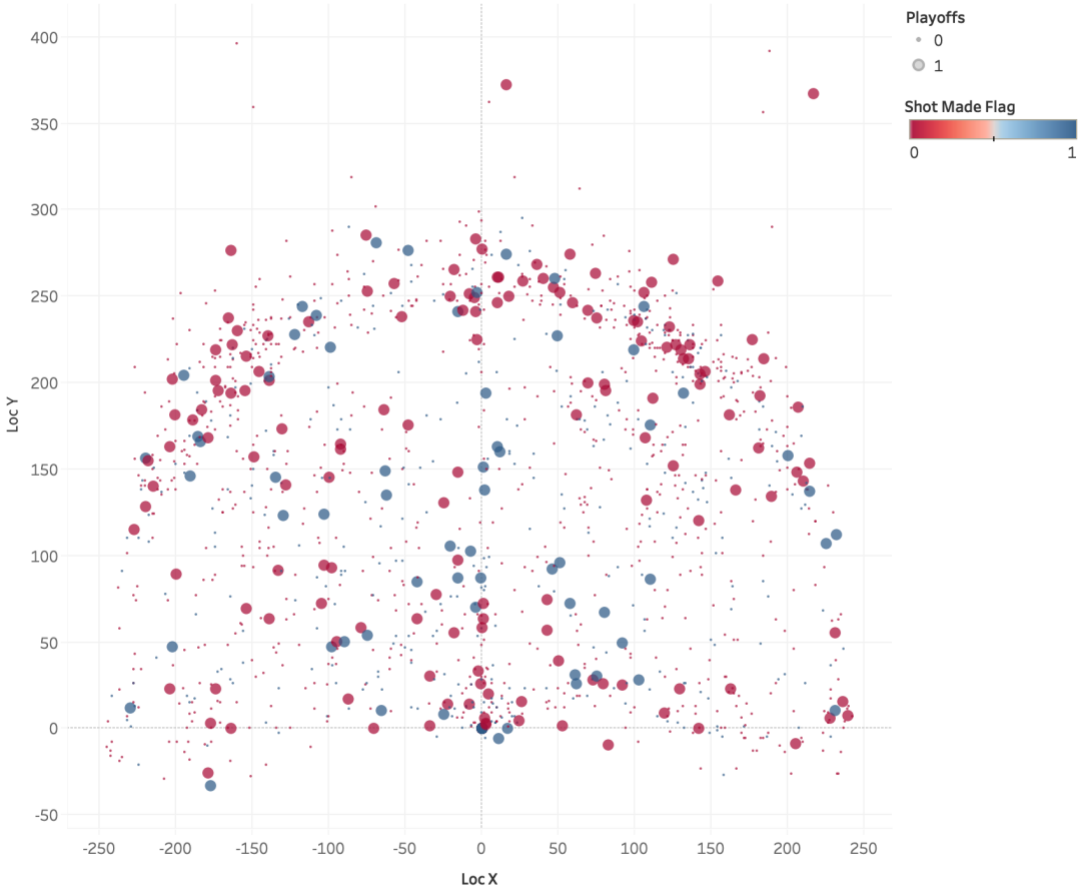


% of Total Number of Records for each Season. Color shows details about Shot Made Flag.

Final Visual:

Looking back at all the visuals I have created, I know have a better understanding of the dataset and how I wanted to approach my final visual in presenting how well Kobe performed in the final moments of his games. I liked the visual of the map showing his shots, and I decided that I can put in a filter that would show which shots he made and missed with 1 minute left in the game in the 4th quarter. The size of the circle indicates whether or not the shot was in the playoffs. Playoff games are more important as they are the ones the matter in order for a team to win a championship. As always, a red circle indicates a missed shot while a blue circle indicates a made shot.

Final Viz



Loc X vs. Loc Y. Color shows Shot Made Flag. Size shows Playoffs. The data is filtered on Minutes Remaining and Period. The Minutes Remaining filter ranges from 0 to 1. The Period filter ranges from 4 to 4.