Carrie Huang (ch646), Samin Foysol (saf243), Lev Akabas (la286)

Final Written Submission

Part1:

A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. **Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.** (10 pts)

We used three sets of data for this project, which we found online on the stats.nba.com website. Each set of data provides information on all NBA games from a certain season. The data that we used were from the 1997 - 1998, 2006 - 2007, and 2016 -2017 seasons. These data sets were obtained by iterating through a list of players and getting a list of information about every shot each player took during a given season (including whether the shot was made, the x and y coordinates of the shot, and whether it was a 2 or a 3). Combining the data from all the individual players yielded the aggregate data across the league for an entire season. Using the x and y coordinates for each shot, we grouped all shot attempts into two 50-by-47 two-dimensional arrays, since an NBA half court is 50 feet by 47 feet, with each value in the two-dimensional array representing the number of shots made/attempted from a particular 1-foot by 1-foot section of the court, which would ultimately correspond to one circle on the heat map.

HERE IS AN EXAMPLE OF THE DATA USED:
http://stats.nba.com/stats/shotchartdetail?Period=0&VsConference=&LeagueID=00&LastNGames=0&TeamID=0&Position=&Location=&Outcome=&ContextMeasure=FGA&DateFrom=&StartPeriod=&DateTo=&OpponentTeamID=0&ContextFilter=&RangeType=&Season=2014-15&AheadBehind=&PlayerID=201935&EndRange=&VsDivision=&PointDiff=&RookieYear=&GameSegment=&Month=0&ClutchTime=&StartRange=&EndPeriod=&SeasonType=Regular+Season&SeasonSegment=&GameID=&PlayerPosition=&

Part2:

A description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales. (10 pts)

We integrated the data from each data set into one heat map per NBA season trying to display how efficient each location on the court is for taking a shot. The color scale represents how efficient each shot is in terms of how many points are scored per location on average, with the green meaning more efficient or higher expected points per shot and red meaning less efficient or lower expected points per shot. Then the size of the circles represent the frequency of shots taken at that location. The smaller the circle, the fewer shots that were taken from there, and the larger the circle, the more shots that were taken at that location. The positions of the circles on the SVG represent the coordinates of locations on the basketball court. All the scales used to determine the color or circle size were linear, although we manually determined those properties of some outliers (such as the spot right next to the basket) to make scaling for the rest of the data much easier and make the charts more readable as a whole.

Part3:

The story. What does your visualization tell us? What was surprising about it? (5 pts)

Our visualization shows change in shot selection and the efficiency of those shots in the NBA over the past two decades (by looking at the 1997-98, 2006-07, and 2016-17 seasons). Through time, we see growing frequency of the 3 point shot and diminishing frequency of longer 2 point shots. Although we were not surprised by this shift, as this is a widely known fact among basketball fans that 3 point shooting has been growing in popularity, what is surprising is that 3s have always been the most efficient shot, even when they were less popular, as seen in our chart for the 97-98 season. Statistically, 3 point shooting has always had a better output then longer 2 point shots in terms of average points made per attempt. Frequency of 3 point shots has only grown recently, perhaps because of the information age where players and coaches have better access to analytics.

There is also a consistently high number of shots taken close to the basket, regardless of time period and efficiency — shots right at the basket are highly efficient, but shots several feet from the basket are actually less efficient than 3 point shots. The high frequency of shots close to the basket is likely due to the fact that it's easier to shoot from closer to the basket, as well as the higher probability of getting an offensive rebound or being fouled close to the basket, which would result in the player keeping possession or shooting free throws.

In conclusion, teams have been getting smarter with their shot selection over the past 20 years. Our visualization generally shows the increase over time of the frequency of shots attempted at areas historically efficient and a the decline of the frequency of shots attempted at areas historically inefficient.