• A video is at disposal in the folder

• You can play with the visualisation (I provide code), please look at ReadMe.txt

INDIVIDUAL PROJECT – INF552

Milestone 3 (Final)

In my final project I propose a visualization about data on shoots made during the regular NBA season in 2015. To set the context, the NBA (National Basketball Association) is the professional American basketball league, and the largest basketball league in the world. The purpose of this visualization is to observe different characteristics of the shots made in games during the 2015 season. Note that for a better understanding and interactivity you can look at the video, and play with the code itself when reading this report.

I) The dataset

The dataset comes from the NBA and contains information on 128070 shots made in 2015. For each shot made, it provides the following information (I give here only the interesting information):

```
GAME_ID,MATCHUP,LOCATION,W,SHOT_NUMBER,PERIOD,GAME_CLOCK,SHOT_CLOCK,TOUCH_TIME,SHOT_DIST,PTS_TYPE,SHOT_RESULT,CLOSEST_DEFENDER,CLOSEST_DEFENDER_PLAYER_ID,CLOSE_DEF_DIST
```

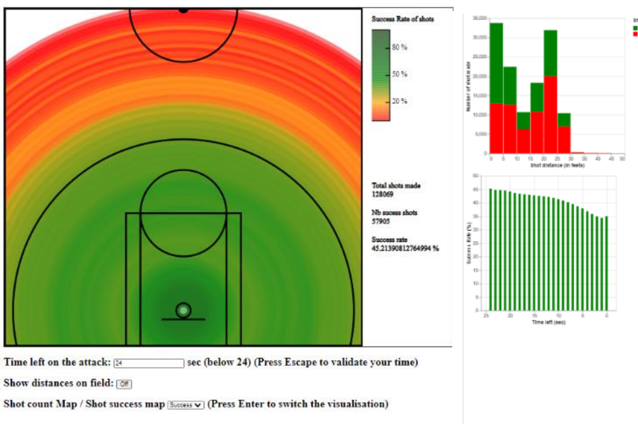
The objective is to represent and aggregate the information given by these shots according to the different variables. More precisely and for the information we are interested in:

• GAME_CLOCK : the time remaining in the game
• SHOT_CLOCK : the time remaining for the attack phase. **It is important to know in this visualisation that each attack phase has a maximum duration of 24sec.** After this time, the attack is stopped and the opponents win the ball.
• SHOT_DIST: The distance at which the shot was taken.
• SHOT_RESULT: The result of the shot (success or failure)
• CLOSEST_DEFENDER: The name of the nearest defender when the shot was taken

We can already feel with these variables that many visualisations will be possible (such as the influence of the remaining time on the attack phase, on the percentage of success or the influence of the distance on the shot). This is what visualization is all about! Whether you are passionate about basketball or you know nothing about it, this visualization will allow you to better understand the NBA basketball games.

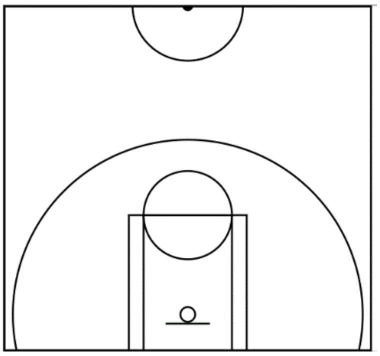
II) The visualisation

This visualisation has been coded from scratch with D3 and Vegalite.The visualisation is an **animated** and **interactive** grid with several charts and representation to allow the user to navigate to navigate freely in the data and its representation. Here is a first preview of the visualization:



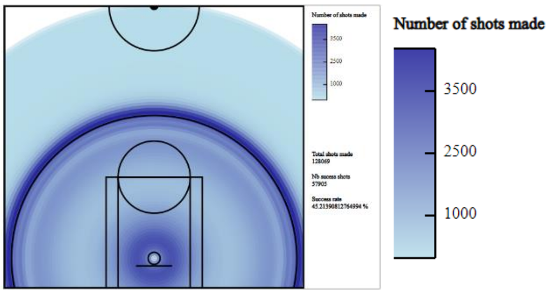
NOTE THAT: A **video** demonstrating the visualisation will be available in the project handover. Also the (functional) **code** will be made available, so that you can **test the visualization**, the dynamic side of the visualization bringing a real plus.

a) The basketball field

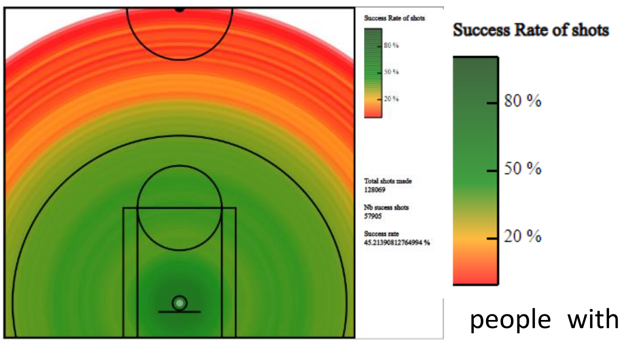


Since the data allows many spatial visualisations thanks to the distance at which the shot was taken, I decided, inspired by PC geo-visualisations (when we were colouring countries or area looking at the data), to project my data and -their representation on a basketball court. With this in mind, a basketball court has been reproduced (**see left**) with the exact and official dimensions (in proportions) that a real basketball court, projecting the data onto this space will give a more faithful and significant representation of reality, and we will see this later.

Now that we have a space to observe our data, we can offer different visualisations. We are going to observe here the shots made from different angles but according to the distance at which they were taken. The first view available with this map will be the number of shots made looking at the distance.



For this purpose, I discretized the space into small distance intervals (the size of the interval is a flexible parameter), counted and represented the number of shots done. As we know, the discretization should me small enough to get a smooth representation of the map, but in this case there was a limit on the minimum size because the data was not really collected on a continuous basis, some distances were much more represented than others (for no real reason). Finally, a smoothing function (by opacity and overlapping) makes it possible to obtain an almost continuous and smooth result on our representation. Looking at the colouring system, I decided to use a classic and efficient method we studied, with the use of a saturation colour scale with obviously and intuitively more saturated colours for distances and areas with more shots and vice versa. I also provided a legend for this colour gradient (**see screen shot above**). Colour scale is like the space and the data represented in linear scale that was the best scale for the visualisation (for example logarithmic scale would have been inappropriate as data is “linearly” distributed and representable. This choice of representation allows us to **know directly where the majority of the shots were made** and thanks to the projection on the basketball court, it is possible to know **immediately which areas correspond to which shots in real life**. For example, with the example given above, we can directly see that a large part of the shots were made close to the basket, which corresponds to the numerous dunks and layups in a game. We can also see that many shots were made at the 3 point line, a very popular area because it maximizes the points of an attack, and that beyond that line only a few shots are taken. Several analysis (that are more precise than the previous one) are also possible with that visualisation.

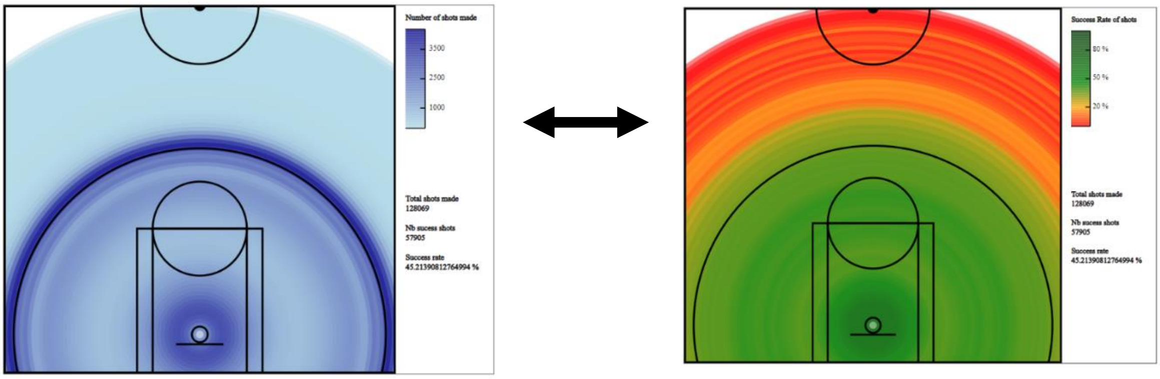


In the same scheme as before, the visualisation allows to observe the success rate of the shots according to the distance, with a smooth representation projected on the basketball court.

This time a colour gradient has been used (not saturation). Same as before I used very few colours to make the visualization readable for people with colour perception deficiency. I also used an intuitive method of colorization. As we know and according to common sense and official statistics this colour gradient and legend have been set to represent accurately the success rate of shots. Indeed, a green colour is associated, to a positive idea so here that the success rate is good. We have obviously the opposite for red colour and a bad success rate. Moreover, according to NBA statistics, a bad shoot rate is between 0 and 15 %, so it is more likely to be red in our case, a mediocre shot rate is around 20% so I chose an intuitive intermediate colour which is orange, and finally a good shooter has a success rate greater or equal to 40% which will be obviously associated to green colours (the further the green is from the red, the better the success rate is). This is exactly the intuitive sense and the information captured in this visualisation and with this colourisation. Again several insights can be gained from this visualization. We can obviously see that the closer the shots are made to the basket, the higher their success rate and vice versa. We can also notice that shots made at a distance corresponding to the distance of free throws are slightly better than the others (excluding of course shots made within 5 feet of the basket).

b) Time, animation and interactivity

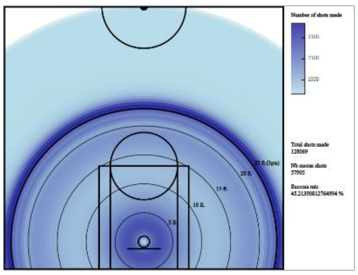
When we look at the previous visualizations made, we get a lot of information about the course of a basketball game. However, it goes without saying that these two statistics are closely related. Indeed, if we take for example the red zones above the 3 points line, corresponding to areas with low success rates, it would be interesting to be able to see how many shots are made at these distances in order to evaluate the impact of this lack of success on the game. To do this, the visualization allows a smooth transition (**to be watched on the video and/or tested on the code provided**), with colours and legend updated. Thus, by pressing the enter key, the user can switch smoothly from one visualization to another and thus observe the different information combined with the 2 visualizations.

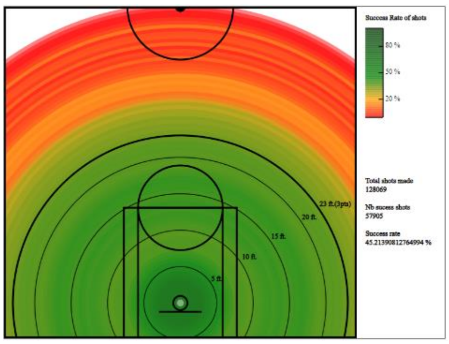


This has a real advantage, if we take again the previous example, we can notice that the zones above the three points have a low success rate but also that very few shots are taken at this distance. The zone actually has a reduced impact on the game. However if we compare the shots taken below 5 feet and at the 3 point line, they are numerous but the success rate is much better for the shots under 5 feet, so they are completely different areas and we can now understand that even if shots made at the 3 point lines maximise the attack phase gain, the success rate is not that good and it might be strategically incorrect to always shot from that distance.
An additional feature of the visualization allows, for a user who is not familiar with basketball or who has difficulties to evaluate the distances on the basketball court, to display or not at any time the distances to the basket on the visualization field in order to get an idea of what is happening (this can be seen as scale like for a geo-visualization).

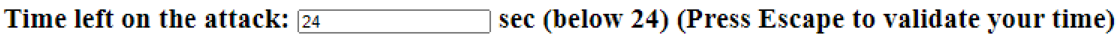
Show distances on field: On

(See better in the ANNEXE section)

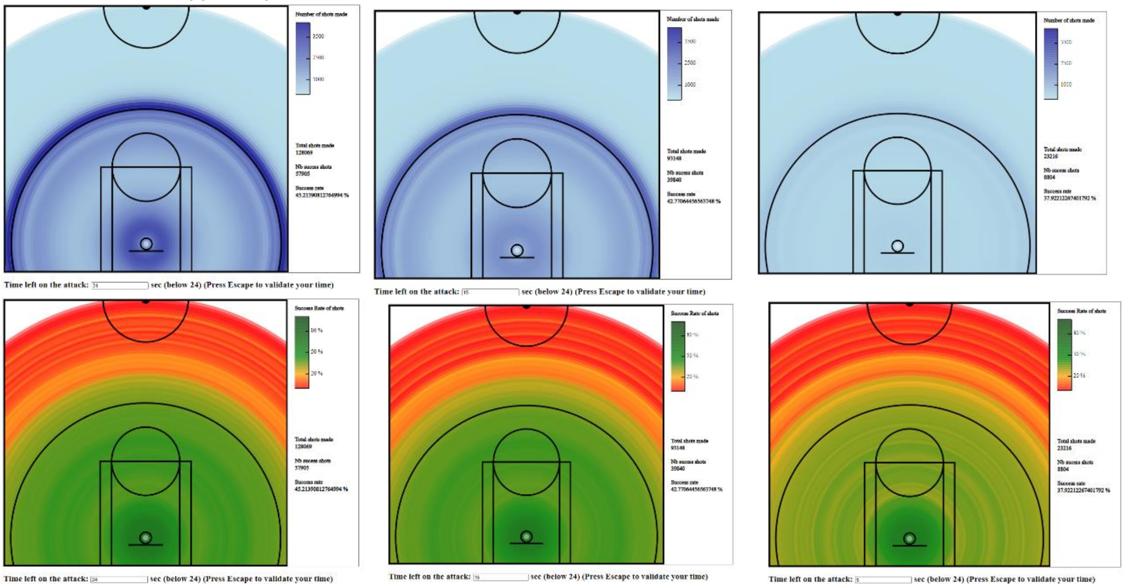




Finally, we mentioned it at the very beginning of this report, but we will now study the impact of time on an attack phase on the different parameters already highlighted. Indeed, over the remaining 24 seconds a shot will be more or less successful depending on whether it was taken at the beginning or at the end of this attack. This is what the visualization allows us to observe. To that purpose an additional interaction field has been set up allowing the user to choose a remaining time (in seconds and less than 24) for the attack phase and to see the impact on the visualizations described above.



After defining this time, a transition to the new visualization is made and he can navigate again between the number of shots made and success rate, but with this additional time condition on the data. (To better visualize this feature I invite you to watch the video of the project and/or to test the visualization). Typically here are the results obtained, for different time selections.

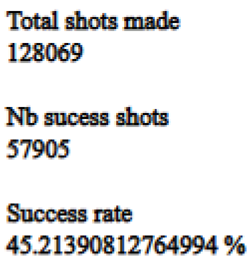


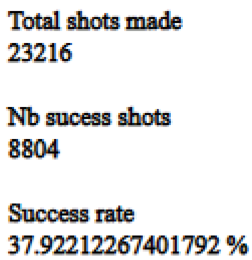
Time left: 24s (full) **Time left: 15s** **Time left: 5s**

First, for the visualization choices, I decided not to change the colour scales and legends and to not readapt them to the new data since this would have distorted and faked the perception of the evolution of the different elements of the visualizations. This allows us to extract a lot of information from this visualisation. Indeed, we can obviously see now that except for shots made under 5 feet, the success rate collapses as time progresses into the attack phase. For the number of shot made, we can see obviously that less shots are made when only 5sec left on the clock, but also that fewer shots are made below 5 feet of distance, and that 3 points shots are important. The issue here is that conserving the same scale to see the number of shot made, allows to see clearly the evolution of the number of shots made but the visualisation (the blue one) becomes less understandable. To tackle this issue, the visualisation proposes complementary charts that gives the complementary information about the number of shot made under a certain time, we will see this in the next part of the report.

We now see how powerful the integration of time is in our main visualisation, but this also allows us to update some information in our grid of visualisation such as the global statistics about the data (put on the right of the field in the visualization), that will be updated in function of the time left chosen by the user:

Global statistics when **24sec** left in the attack phase:





Global statistics when **5sec** left in the attack phase:

These statistics give us important general information about the shots in relation to the time remaining for the given attack phase. They also allow the user to better read the information displayed on the basketball court, for example he will better understand the drop in the success rate of certain areas of shots when there is little time left for the attack phase by comparing it to the overall success rate at that time (which drops from 45% for 24sec left to 38% for 5sec left). Finally, time will allow us to update a complementary charts, which we will explain right now.

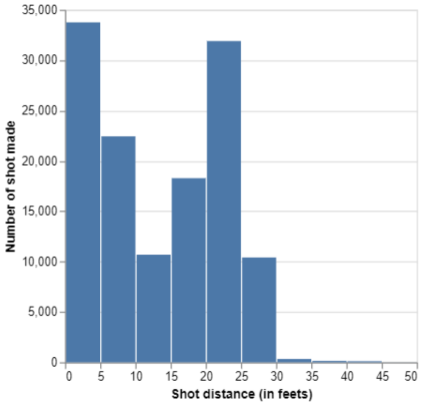
To finish on this part, I would like to explain the design decisions to include animations and transitions in my visualization. As far as I am concerned, and from what I learn from the course animation allows in my visualisation to : maintain elements up to date, hook the user and keep him engage, providing visual comfort and aesthetics (with transistions), make progress visible and last but not least reveal data relationships.

c) Complementary chart

In order to complete my visualization I added 2 bar charts.

The first one is the following:

Figure 1:
24 sec left
Looking at the number of shots made.



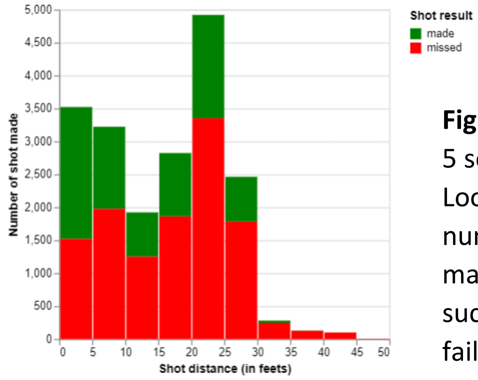
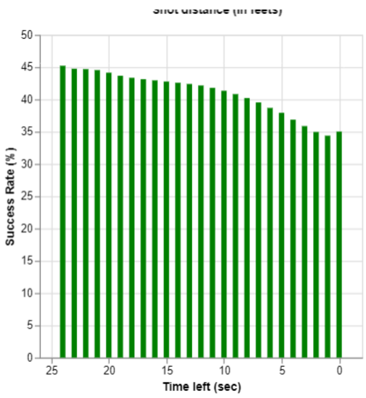


Figure 2:
5 sec left
Looking at the number of shots made and at the success and failure.

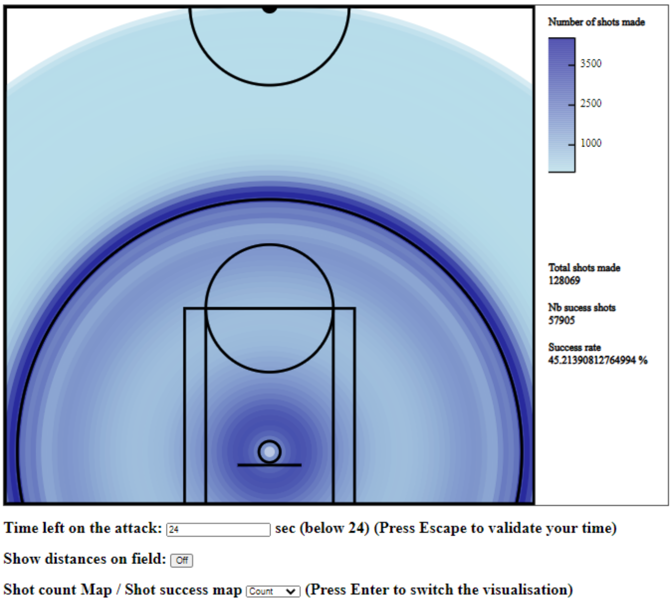
As I explained before, it was difficult to perceive precisely, on the basketball field, the number of shots done in function of the distance when the time set for the attack phase was too short. This bar chart allows you to visualize this more clearly by counting the number of shots per 5 feet interval. Furthermore it is updating according to the time remaining in an attack phase. It is therefore easier to analyse these shots as a function of time. A second feature of this bar chart is that when the user changes the visualization mode to see the success rate (blue to red/orange/green), it allows to encode the colours and to see for each distance interval, the number of shots made, but also the number of shots successful in green and those missed in red (to remind the intuitive colours on the basketball field). Let’s take a look on how powerful this visualisation is. If we compare figure1 and figure2, we can see that the distribution of shots has changed, with much more 3 points shots when only 5 sec left than when 24 sec left. And you can do the same comparison with success rate if you are looking at the success rate visualisation (switching with enter key).

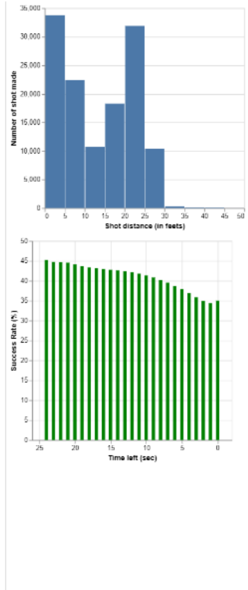


Finally a last bar chart has been added (this one is static), and allows us to perceive the evolution of the success rate in function of time because it really hard to see an overall behaviour for this visualization only by changing and selecting different times. This chart (**on left**) allows us to perceive quickly how the success rate is evolving. As we would expect, the success rate decreases the less time there is left on an attack and this visualization allows us to see that immediately.

ANNEXE (larger size screen shots, examples, see video)

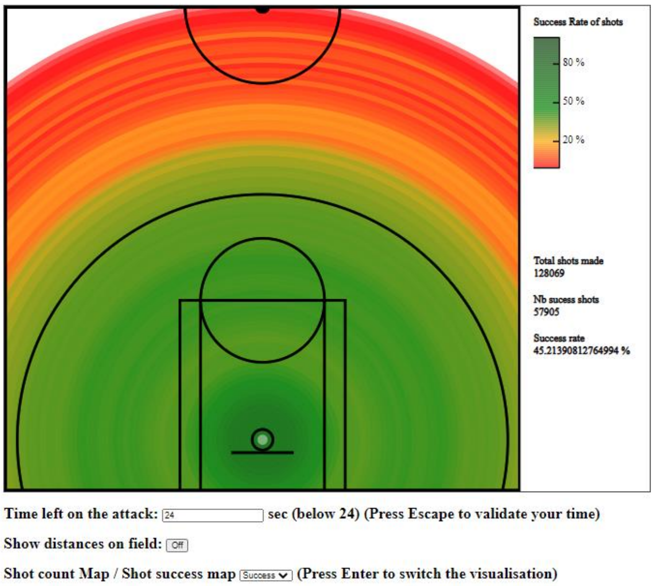
•Number of shots made
•24 sec left

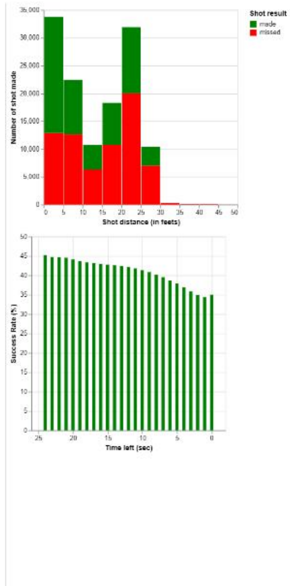




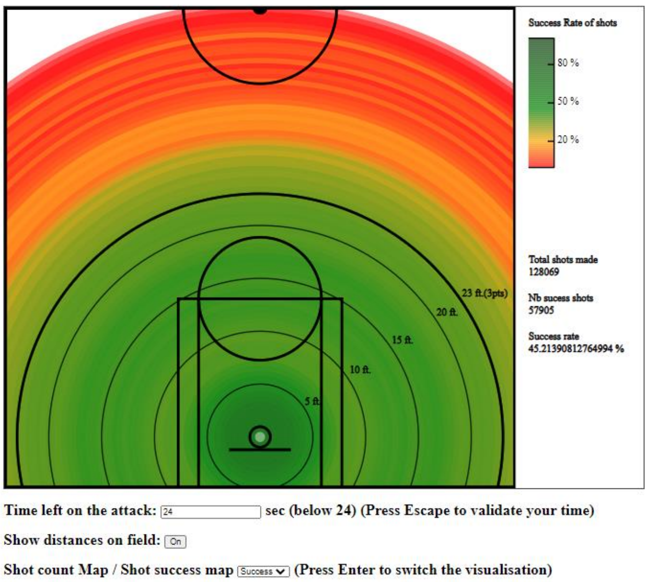
Transition

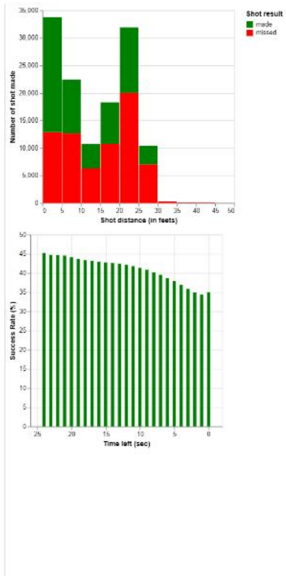
•Shots success rate
•24 sec left





•Shots success rate
•24 sec left
•Show distances ON





- Shots success rate
- 5 sec left