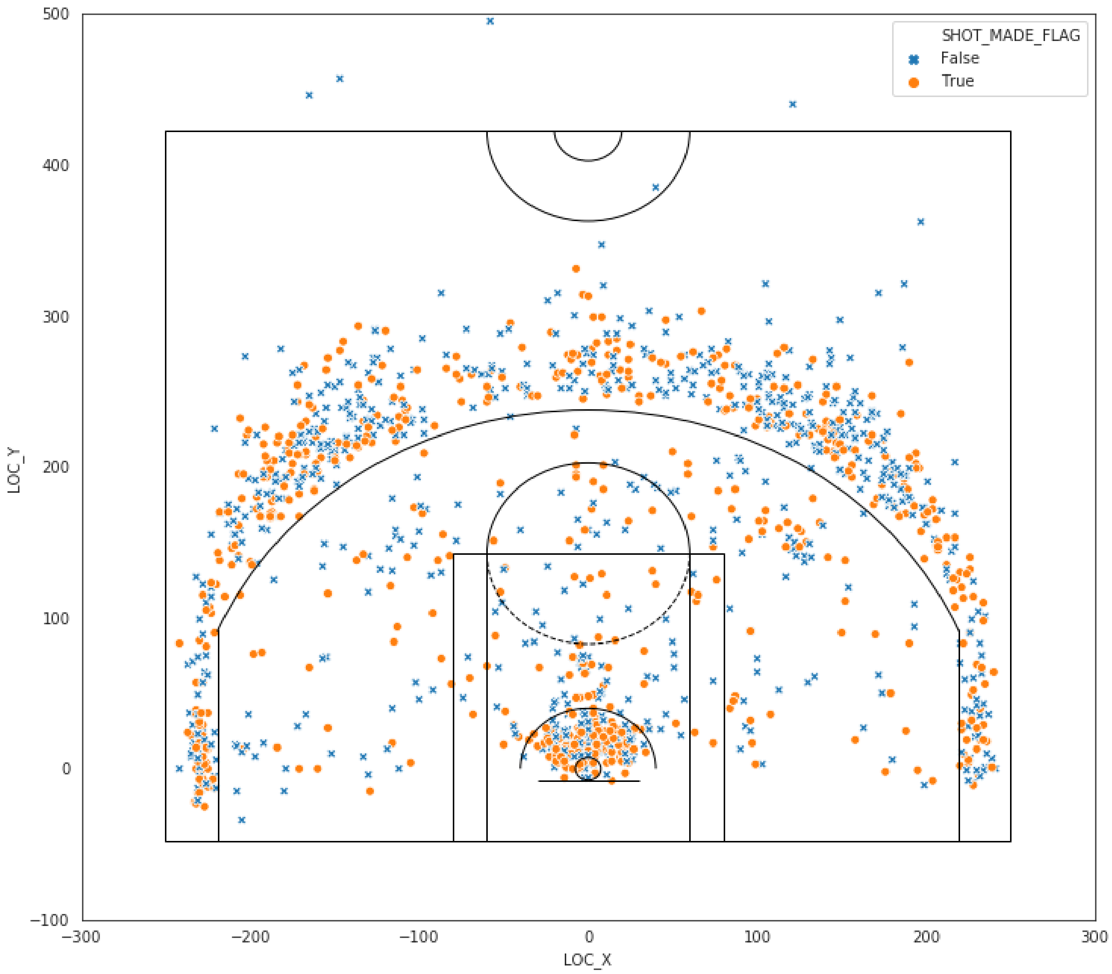
0.0.1 Question 2b: All Shots Scatter Plot + Court Outline

Again use seaborn to make a scatter plot of Stephen Curry's shots. Again, set the x-axis limits to span (-300, 300), the y-axis limits to span (-100, 500) color the points by whether the shot was made or missed. Set the missed shots to have an 'x' symbol and made shots to be a circular symbol. Call the `draw_court` function with `outer_lines` set to to be true. Save the `Axes` returned by the plot call in a variable called `ax`.

In [114]:
```
plt.figure(figsize=(12, 11))
markers = {0 : "X", 1 : "o"}
ax = sns.scatterplot(data=curry_data,x="LOC_X",y="LOC_Y",markers=markers,style="SHOT_MADE_FLAG
draw_court(outer_lines=True)
plt.xlim(-300,300)
plt.ylim(-100,500)

plt.show()
```

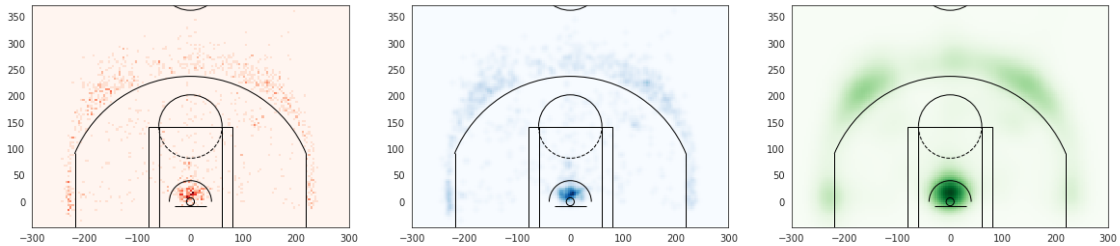


0.0.2 Question 2c: Analyzing the Visualization

In a few sentences, discuss what makes this an effective or ineffective visualization for understanding the types of shots that Stephen Curry likes to take and is good at taking, relative to other players in the league. Are there ways it can be improved?

This is an effective visualization to a certain degree. By looking at this plot we can see that Steph loves to take 3 point shots as well as 2 point shots in the paint. It also shows us the shots he is good at taking (i.e the shots he made). In terms of relative to players in the league, I believe it would be better to plot his shots of made/misses against other players in the league to get a better idea. From this plot we are only able to see the shots Steph has made/missed.

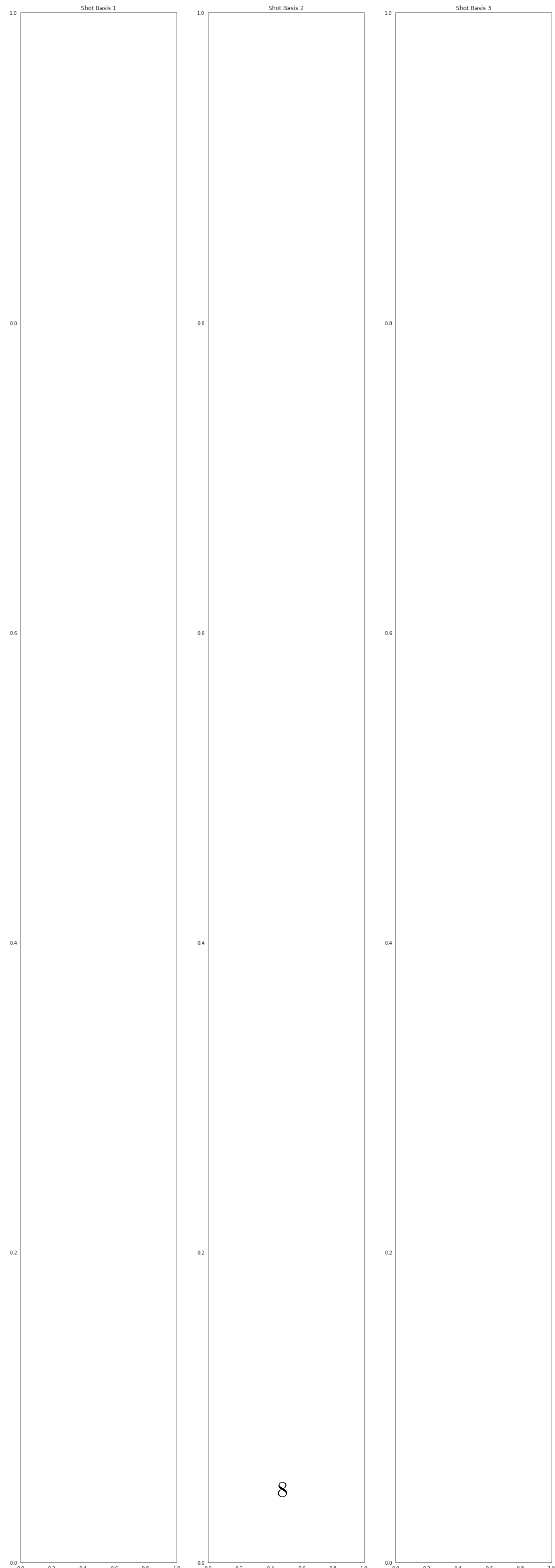
In [121]:
```
fig, ax = plt.subplots(1, 3, figsize=(20,60))
plot_shotchart(curry_binned_unsmoothed[0], xedges, yedges, ax=ax[0],use_log=False, cmap = 'Re
plot_shotchart(curry_binned_smoothed1[0], xedges, yedges, ax=ax[1],use_log=False, cmap = 'Blu
plot_shotchart(curry_binned_smoothed5[0], xedges, yedges, ax=ax[2],use_log=False, cmap = 'Gre
fig.show()
```



0.0.3 Question 4b: Visualizing Shot Types

Plot the first three basis images by calling `plot_vectorized_shot_chart` below on the columns of `W3`.

```
In [126]: def plot_vectorized_shotchart(vec_counts, xedges, yedges, ax=None, use_log=False, cmap = 'Reds
              """Plots 2d heatmap from vectorized heatmap counts

              Args:
                  hist_counts: vectorized output of numpy.histogram2d
                  xedges, yedges: bin edges in arrays
                  ax: figure axes [None]
                  use_log: will convert count x to log(x+1) to increase visibility [False]
                  cmap: Set the color map https://matplotlib.org/examples/color/colormaps_reference.htm
              Returns:
                  ax: axes with plot
              """

              nx = len(xedges)-1
              ny = len(yedges)-1

              # use reshape to convert a vectorized counts back into a 2d histogram
              two_d_counts = np.reshape(vec_counts, (nx, ny))

              return(plot_shotchart(two_d_counts, xedges, yedges, ax=ax, use_log=use_log, cmap=cmap))

          fig, ax = plt.subplots(1, 3, figsize=(20,60))

          ## Write a for loop
          for i in range(3):
              # Call plot_vectorized_shot_chart
              ax[i].set_title('Shot Basis %i' % (i+1))
```


Shot Basis 1
Shot Basis 2
Shot Basis 3
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

0.0.4 Question 4c: Reconstruction Error

Below we re-construct the shooting pattern for a single player. By "reconstructing" we mean use the approximation

$$\hat{X} = WH$$

obtained via NMF. Find $\hat{X}$ by multipling W and H. In python the `@` symbol is used for matrix multiplication.

```
In [127]: X3_hat = W3@H3
```

0.0.5 Question 4d: Choice of Colormap

Why does it make sense to use a *sequential* palette for the original and reconstructed shot charts and a *diverging* palette for the residual? *Hint:* Read the introduction to colormaps here.

We should use sequential palette for the original and reconstructed charts since they are better for values that have ordering. We use diverging palette on the residuals since it is better for values that have a critical value.

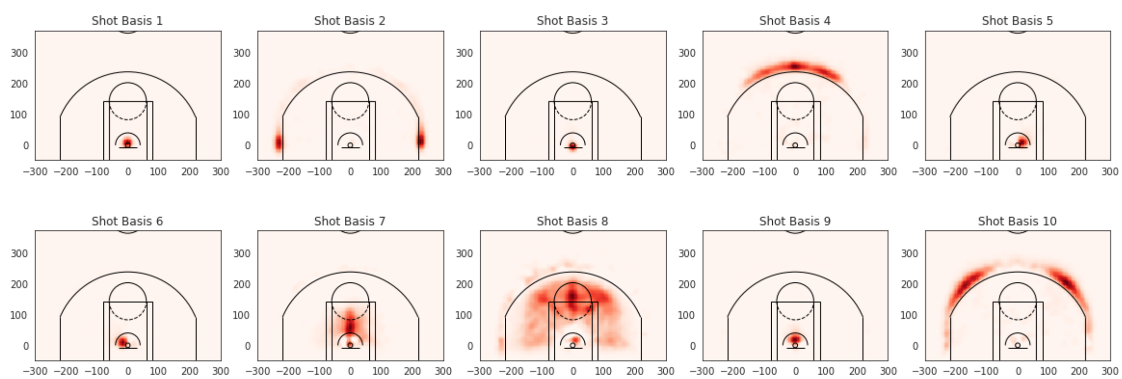
0.0.6 Question 4e: More Detailed Modeling

Re-run the analysis, this time for 10 basis vectors instead of 3. Again plot the bases using `plot_vectorized_shotchart` on the columns of `W10`.

Hint: Study the following code

```
fig, ax = plt.subplots(2, 5, figsize=(20, 7))
ax = ax.flatten() # turn ax into a flat array
ax[0].set_title('hello')
ax[9].set_title('there')
fig.show()
```

```
In [129]: plot_vectorized_shot_chart2 = non_negative_marix_decomp(10, X)
          W10 = plot_vectorized_shot_chart2[0]
          H10 = plot_vectorized_shot_chart2[1]
          fig, ax = plt.subplots(2, 5, figsize=(20, 7))

          ## Write a for loop
          for i in range(10):
              plot_vectorized_shotchart(W10[:,i], xedges, yedges, ax= ax[i//5, i % 5])
              ax[i//5, i % 5].set_title('Shot Basis %i' % (i+1))
```

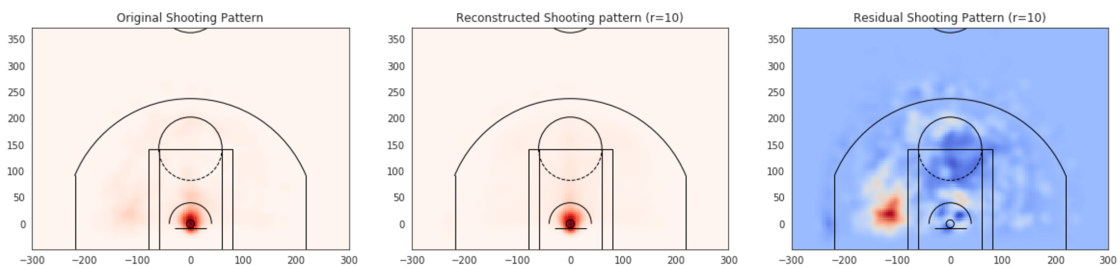


If you did things correctly, you should be really impressed! We've identified potentially interesting patterns of shooting styles without actually specifying anything about the way basketball is played or where the relevant lines are on the court. The resulting images are based only on the actual behavior of the players. Even more impressive is that we're capturing similarity in regions that are far apart on the court. One reason we can do this is that a basketball court is symmetric along the length of the court (i.e. symmetric about x=0). However, people tend to be left or right hand dominant, which might affect their preferences. Look carefuly at the shot basis plots above: is there any evidence of *asymmetry* in player shooting behavior? Refer to specific basis images in your answer.

If we look at the shots we can see some of this behavior. Starting by looking at shot basis 5 the player shoots on the right side of the hoop more. In shot basis 6 the player shoots from the left side of the hoop more. In shot basis 8 it seems the player shoots more on the right side of the court compared to the left.

Repeat part 5b, and again plot original, reconstructed and residual shot chats for LaMarcus Aldridge.

```
In [130]: X10_hat = W10@H10

          fig, ax = plt.subplots(1, 3, figsize=(20,60))

          # I took the first player appearing in first column
          # (you probably want to do more interesting players)
          original_shotchart = plot_vectorized_shotchart(X[:,to_plot_idx], xedges,yedges, ax=ax[0])
          reconstructed_shotchart = plot_vectorized_shotchart(X10_hat[:,to_plot_idx],xedges, yedges, ax=
          residual_chart =  plot_vectorized_shotchart((X-X10_hat)[:, to_plot_idx], xedges,yedges, ax=ax

          ax[0].set_title('Original Shooting Pattern')
          ax[1].set_title('Reconstructed Shooting pattern (r=10)')
          ax[2].set_title('Residual Shooting Pattern (r=10)');
```



0.0.7 Question 4f: Comparing Players

With `H10` matrix, it is possible to compare any pair of players. For all players pairwise, i and j, compare using euclidean distance between their coefficients:

$$\text{player-distance}(i,j) = \|H_i - H_j\|_2 = \left(\sum_{k=1}^{10}(H_{ki} - H_{kj})^2\right)^{1/2}$$

Create a heatmap for comparing pair-wise player distance matrix. Find the two pairs of players with smallest distances. Also, find two pairs of players with largest distances.

In [203]:
```
player_distances = np.zeros(387)
player_distance_sum = 0
for i in range(387):
    for j in range(1,388):
        for k in range(10):
            player_distance_sum += (H10[k,i] - H10[k,j])**2
        player_distances[i] += player_distance_sum**(1/2)
        player_distance_sum = 0
```

Out[203]: 387

Type your answer here, replacing this text.

0.0.8 Question 4g: Residuals

The residual betwene `Xhat` and `X` gives a sense of how well a player is decribed by NMF computed matrices `W` and `H`. Calculate RMSE for each player, and plot the histogram. Comment on this distribution and players with smallest and largest RMSEs (use 10 components).

Type your answer here, replacing this text.

0.0.9 Question 4h: Proposing improvements

One of the main purposes of exploratory data analysis is to generate new ideas, directions, and hypothesis for future analyses and experiments. Take two players of your choice and compare their shooting patterns with various visualizations.

State any insights and defend your conclusions with visual and/or numerical comparisons.

Looking at LaMarcus Aldridge and comparing him to Steph Curry it is easy to see some differences. LaMarcus takes very few to no three point shots, mostly only two pointers either in the paint or outside of it. Steph Curry however, makes multiple three point shots along with some two point shots down in the paint.