HW 9: Soccer stats

Stats and sports class

Fall 2019

Overview

In this HW, we'll look at women's world cup data (note – several questions are similar to the most recent lab).

```
library(RCurl)
library(tidyverse)
url <- getURL("https://raw.githubusercontent.com/statsbylopez/StatsSports/master/Data/sb_shot_data.csv")
wwc_shot <- read.csv(text = url)
```

Better shot maps

ggplot() has ample ways to enhance shot maps. Consider the following maps

```
wwc_shot <- wwc_shot %>%
  mutate(is_goal = shot.outcome.name == "Goal")

usa_shot <- wwc_shot %>%
  filter(possession_team.name == "United States Women's")


p1 <- ggplot(usa_shot, aes(location.x, location.y)) +
  geom_point()

p2 <- ggplot(usa_shot, aes(location.x, location.y, colour = is_goal)) +
  geom_point()

p3 <- ggplot(usa_shot, aes(location.x, location.y,
      colour = is_goal, size = shot.statsbomb_xg)) +
  geom_point()

p1
```
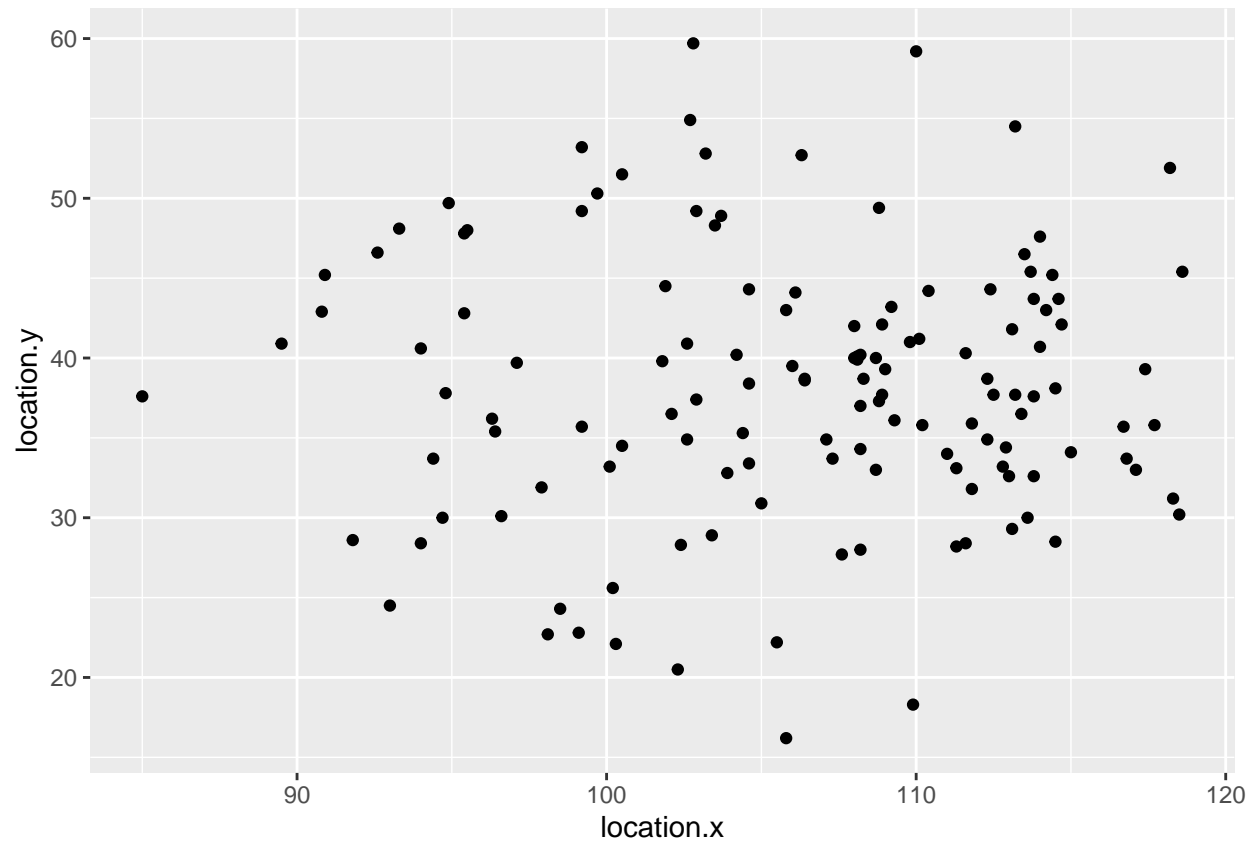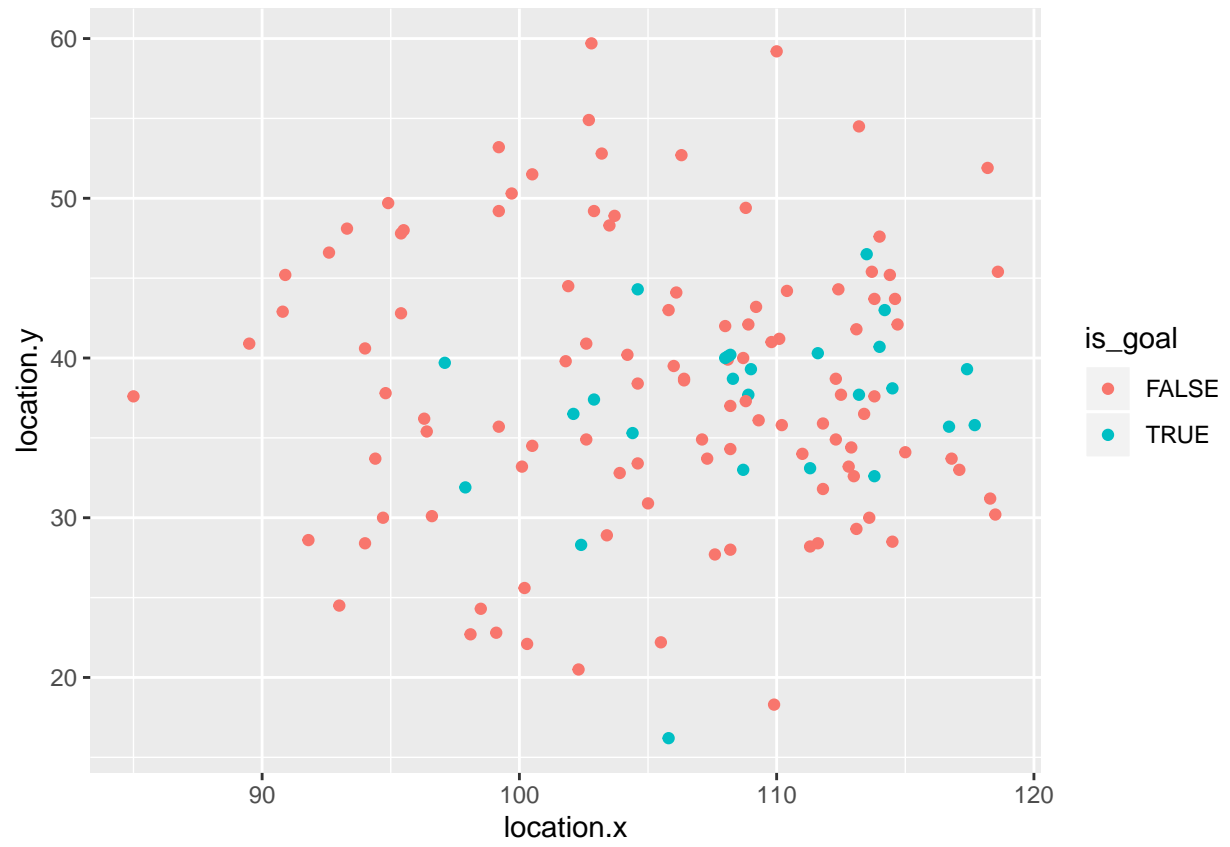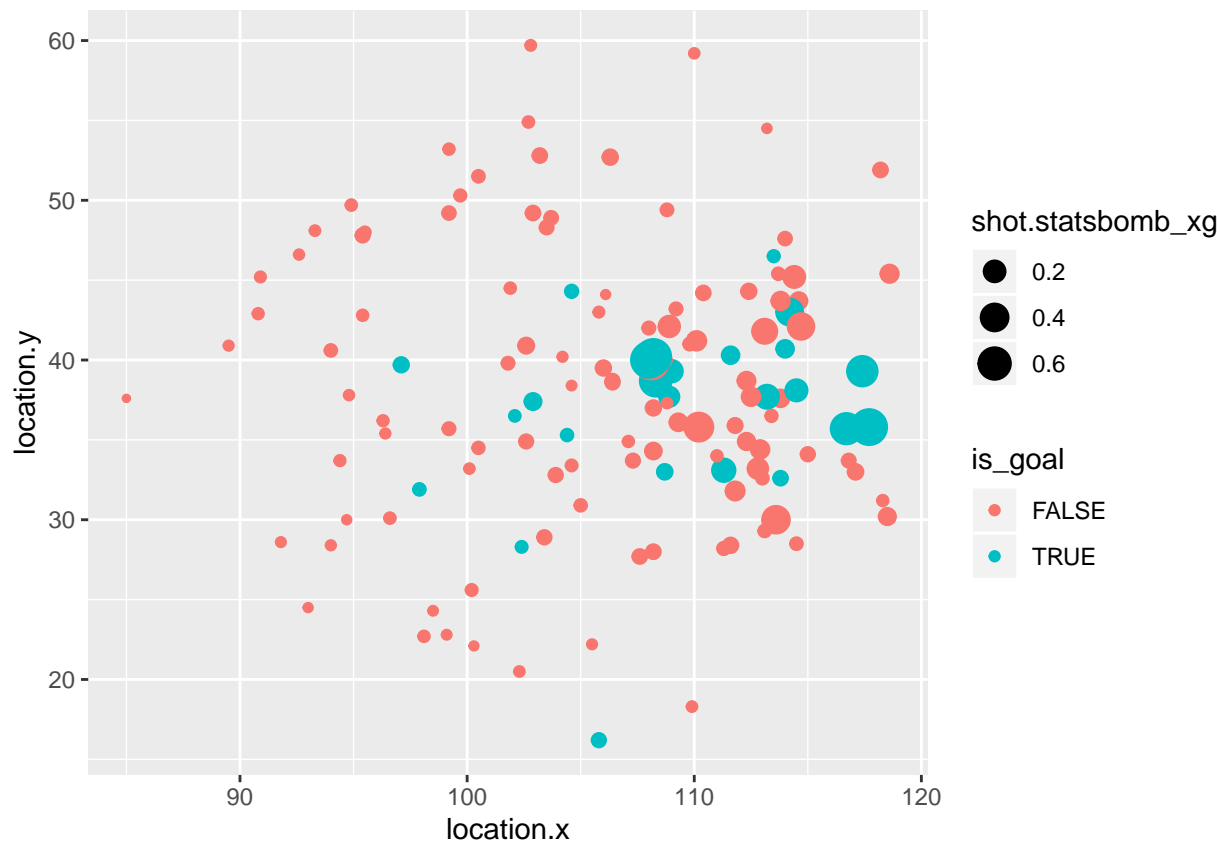
p2

p3

1. What features are apparent in `p2` that aren't apparent in `p1`? What features are apparent in `p3` that aren't apparent in `p2`.

p2 highlights where the actual goals were scored (generally closer to the net) that p1 does not. p3 gets at the size of the expected goal (how likely each show was of scoring) and features both distance and other shot level information

Practice with dplyr

8. For each USA shooter, average the `TimeInPoss` and `DefendersBehindBall` when they took their shot. Filter to make sure you are only looking at players with at least 10 shots. What does this say about how players took shots?

```r
usa_shot %>%
  group_by(player.name) %>%
  summarise(mean_time = mean(TimeInPoss),
            mean_def = mean(DefendersBehindBall),
            n_shots = n()) %>%
  filter(n_shots >= 10)
```

```
## # A tibble: 6 x 4
##   player.name   mean_time mean_def n_shots
##   <fct>             <dbl>    <dbl>   <int>
## 1 Alex Morgan       23.2     3.65      20
## 2 Carli Lloyd       18.6     3.18      17
## 3 Julie Ertz         9.09    5.58      12
## 4 Megan Rapinoe     15.9     3.15      13
## 5 Rose Lavelle      32.1     5.64      14
```

```
## 6 Samantha Mewis        21.0       5           16
```

Julie Ertz tended to take shots quicker, while Alex Morgan and Roze Lavelle's took longer to build up. Ertz has the *most* defenders in front of her, while Carli Lloyd and Megan Rapinoe had the fewest

Exploration

A soccer coach wants to know the best places to shoot from. What would you tell the coach? Create a grid across the field using the `cut()` command (for both x and y), and then, within each location, estimate the goal rate. Next, use `geom_tile()` to make a map of goal rates within each cell of the grid you created. For a reminder on `cut()`, see our notes on Hosmer-Lemeshow, or `?cut()`.

```
wwc_shot %>%
  mutate(cut_x = cut(location.x, 10),
         cut_y = cut(location.y, 10)) %>%
  group_by(cut_x, cut_y) %>%
  summarise(goal_rate = mean(is_goal)) %>%
  ggplot(aes(cut_x, cut_y, fill = goal_rate)) +
  geom_tile()
```
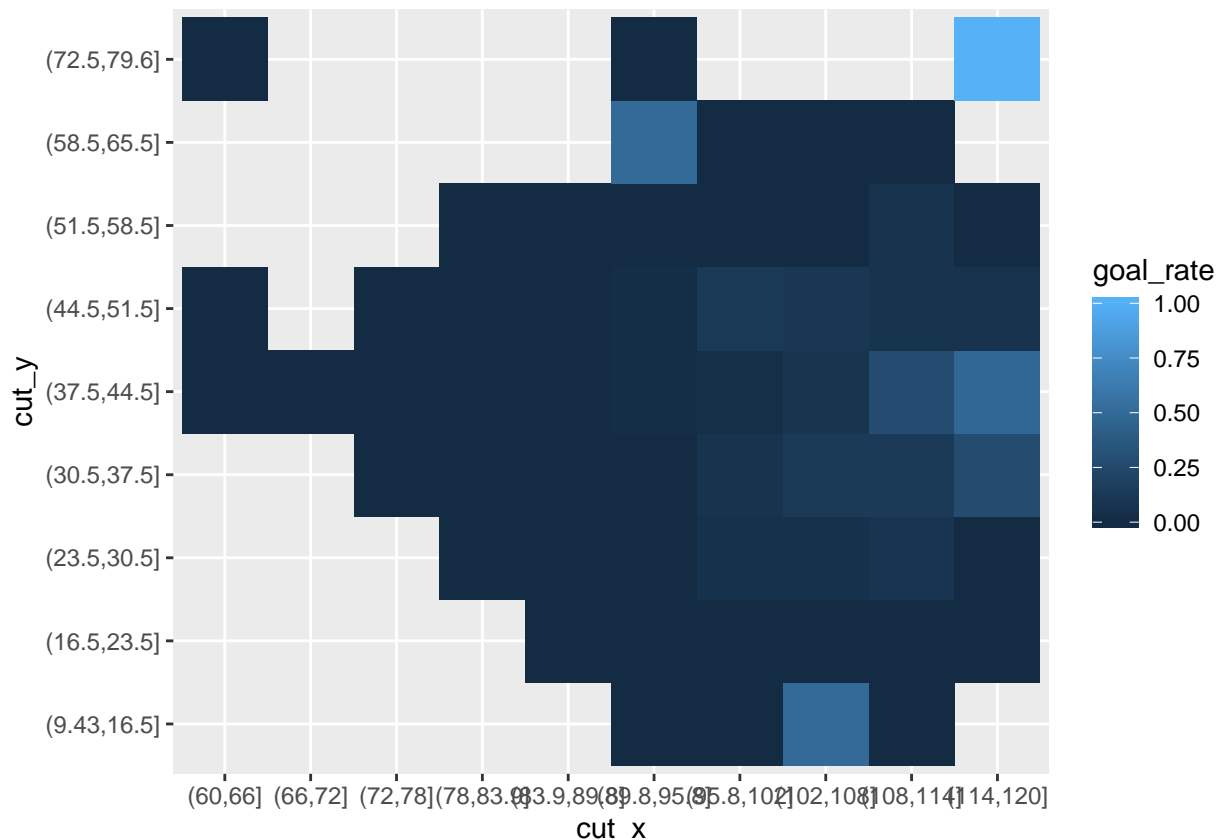

Solutions: The above map creates a grid and plot that is useful for assesing goal likelihood – students that are closest (in idea, not neccessarily in the exact visual) will get full credit