

# Where's Waldo?

## 16 marks

**Where's Waldo?** This was one of the most popular children's books worldwide in the last few decades. The book would present a picture across two pages crowded with small colourful cartoon drawings of people, animals, and objects. These would be in a variety of poses and would appear anywhere over both pages. Amongst the many people and animals appearing all over both pages would be the character called Waldo (or Wally in some countries). The objective was to find Waldo somewhere in each complex picture.

A Dr. Randy Olson (presently at the University of Pennsylvania) appears to have determined all the locations of Waldo across a series of these children's books (don't ask me why). These are available on the course website in the file `wheres-waldo-locations.csv`.

Read this into R as a data frame as follows

```
waldo <- read.csv("wheres-waldo-locations.csv", header=TRUE)
```

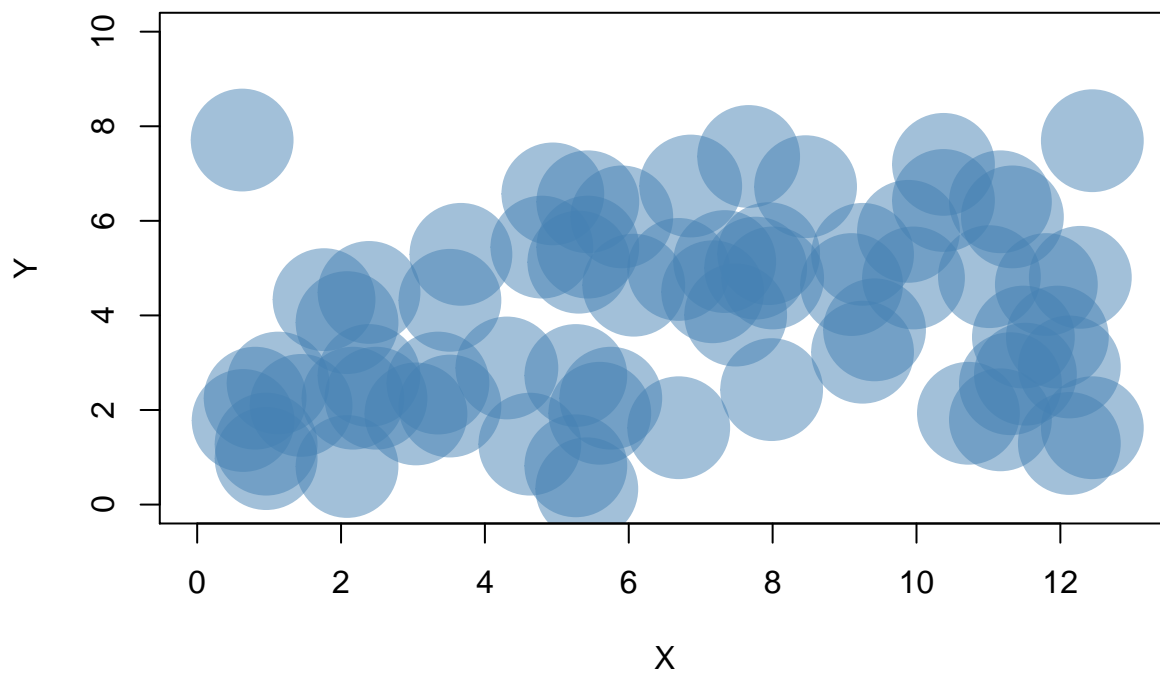
Before the first line above can be executed successfully, the working directory would need to be set (using `setwd(...)`) to whatever directory you stored the csv data file after it was downloaded.

The second line will print the first few rows of the data and reveal that it has four different variates: `Book`, `Page`, `X`, and `Y`. The first two are self explanatory and the last two are Cartesian coordinates for the actual location of Waldo on that page in that book.

In this question, you will be plotting Waldo's locations in several ways and looking for patterns in these locations. You will need to ensure that the limits of the plot remain the same over different plots.

- a. **(3 marks)** Using a common `xlim`, a common `ylim`, `pch = 19`, and alpha blending in all plots, plot the locations of Waldo for different values of the point size. Select one such size so as to best convey the density of the points. Hand in only this one plot. Comment on where Waldo has a tendency to be located on the page.

```
with(waldo, plot(X,Y, xlim=c(0,13), ylim=c(0,10), pch=19, cex=7,  
                col=adjustcolor('steelblue', 0.5)))
```

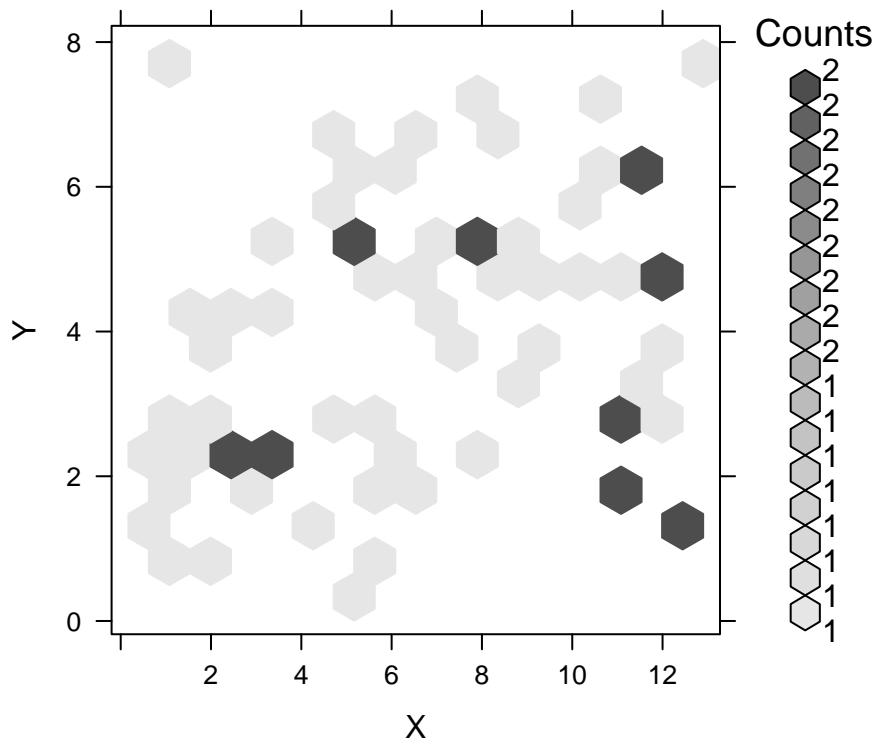


The Waldo tends to be located at anywhere in horizontal but in the middle of vertical axis.

- b. (3 marks) Repeat part (a) but use hexagonal binning in two different ways to convey the density. Comment on where Waldo has a tendency to be located on the page.

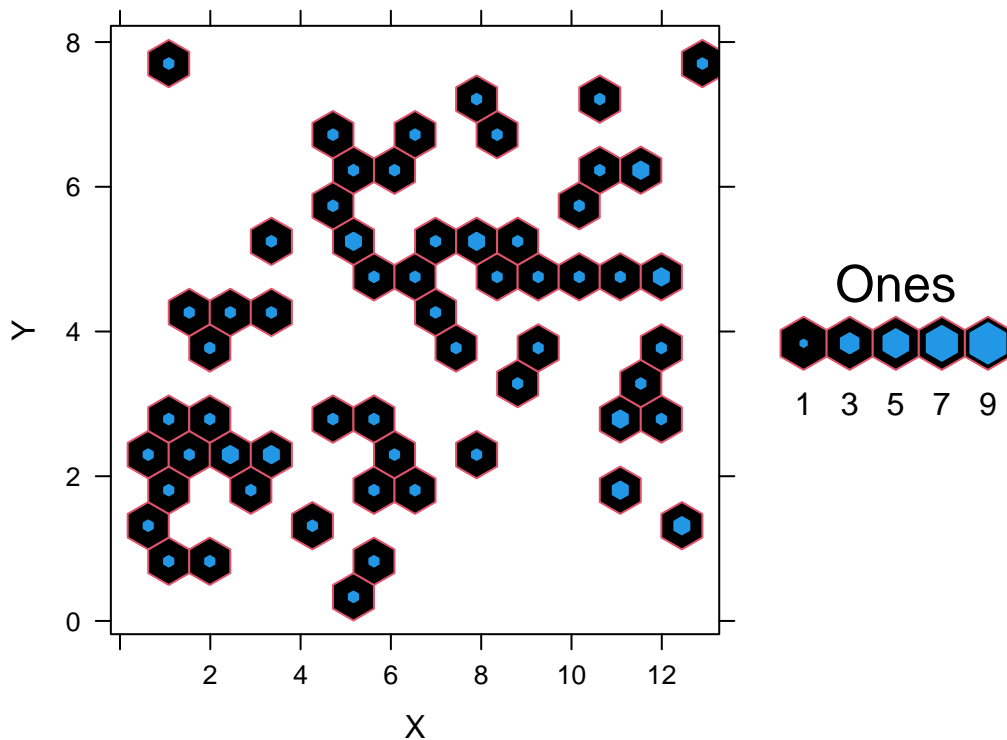
```
library('hexbin')
hexbinplot(Y~X, data=waldo, xbins=13,
           colramp=function(n) rev(gray.colors(n)),
           aspect=1,
           main='Hexagonal bin histogram')
```

## Hexagonal bin histogram



```
hexbinplot(Y~X, data=waldo, xbins=13,  
           colramp=function(n) rev(gray.colors(n)),  
           aspect=1,  
           style='nested.lattice',  
           main='Hexagonal bin histogram')
```

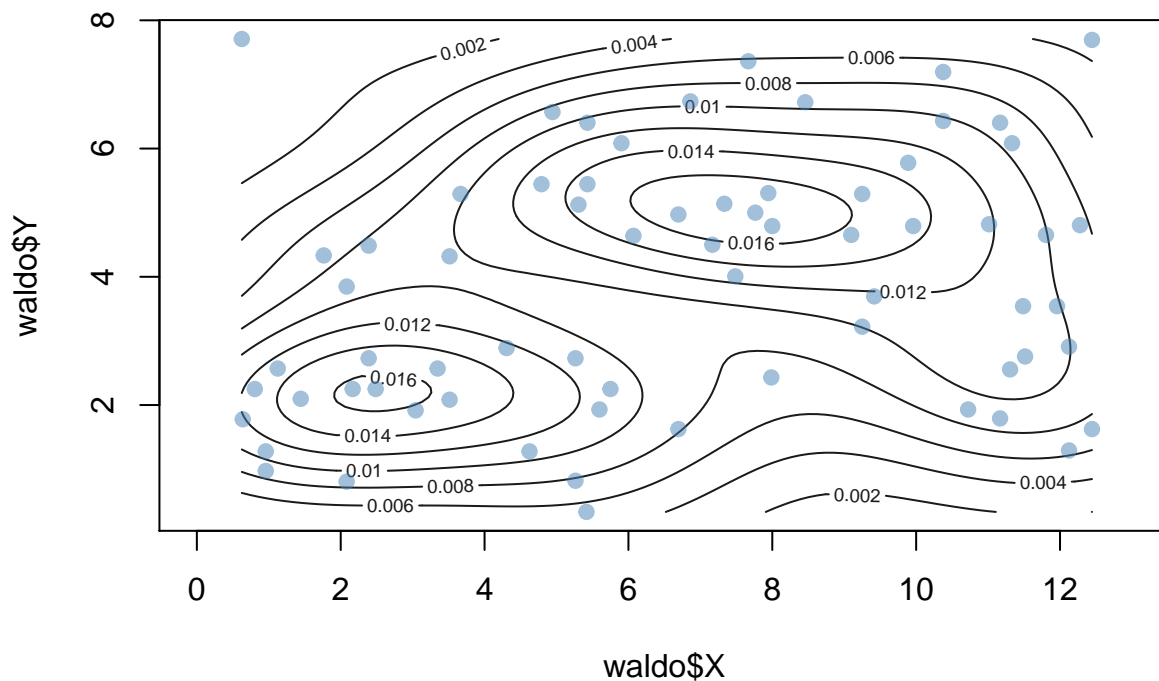
## Hexagonal bin histogram



Waldo has a tendency to be located around (2, 2) and (8, 5).

- c. (5 marks) Use the function `kde2d(...)` from the MASS package and plot the contours of the density estimate for Waldo's positions. Hand in your code and contour plot. Based on this plot alone, describe

```
library(MASS)
den <- kde2d(waldo$X, waldo$Y, n=100)
zlim <- range(den$z)
plot(waldo$X, waldo$Y, pch=19, col=adjustcolor('steelblue', 0.5),
     contour(den$x, den$y, den$z, col='grey10',
             levels=pretty(zlim,10), lwd=1, add=TRUE),
     xlim=c(0,13))
```



- where Waldo is most likely to appear? (i.e. where should you look first for Waldo?)

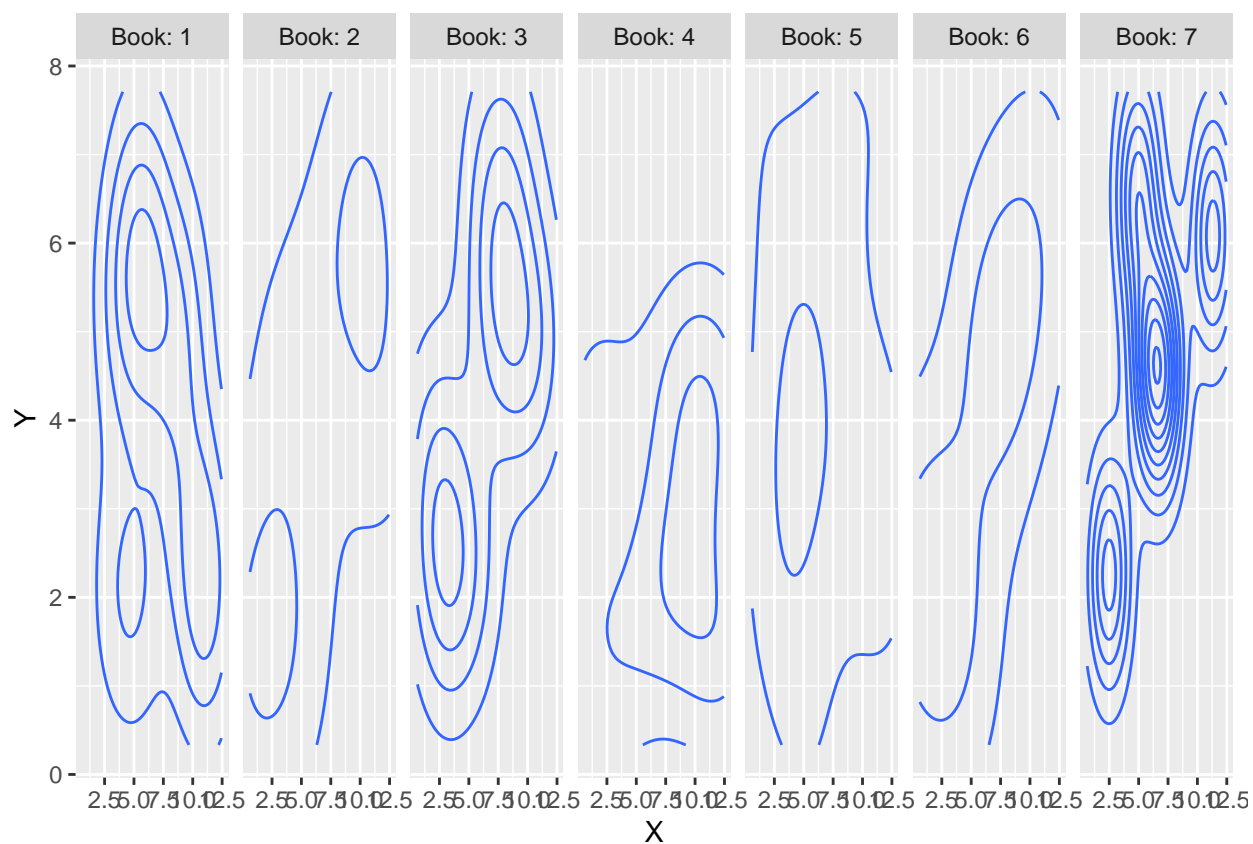
Waldo is most likely to appear around (2.5, 2) and (8, 5).

- where Waldo is least likely? (i.e. where should you look last for Waldo?)

Waldo is least likely to appear at top left and bottom right.

- d. (5 marks) Using `ggplot` construct two dimensional kernel density estimates (in a single conditional plot) for Waldo's positions conditional on each book. Comment on whether the author (Martin Handford) has changed the distribution for Waldo's position with each book.

```
library(ggplot2)
p <- ggplot(data=waldo, mapping=aes(X,Y))
p+geom_density_2d()+facet_grid(.~Book,labeller=label_both)
```



From seven graphs, we are able to see some changes in the locations of hills. We conclude that the author has changed the distribution for Waldo's position. Especially for book 4 and 5, there is only one hill.