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.

```
library(MASS)
xlim = extendrange(waldo$X)
ylim = extendrange(waldo$Y)
with(waldo, plot(X,Y, col=adjustcolor("steelblue", 0.5),pch=19,cex=6,xlim=xlim, ylim=ylim ,main = "Walfo's
```

Walfo's location
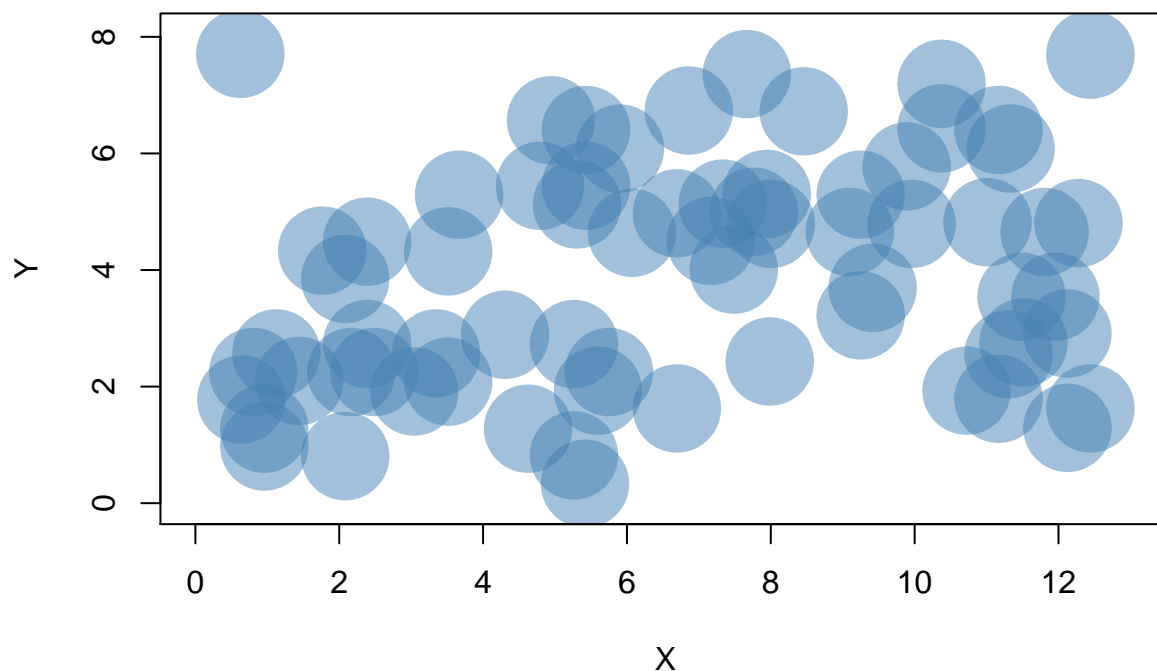
Waldo tends to be around bottomleft(x=2,y=2), bottomright(x=2,y=12) and middle (x=7, y=4) of the page, however the pattern is not very strong except Waldo seldom appear on the top left of the page.
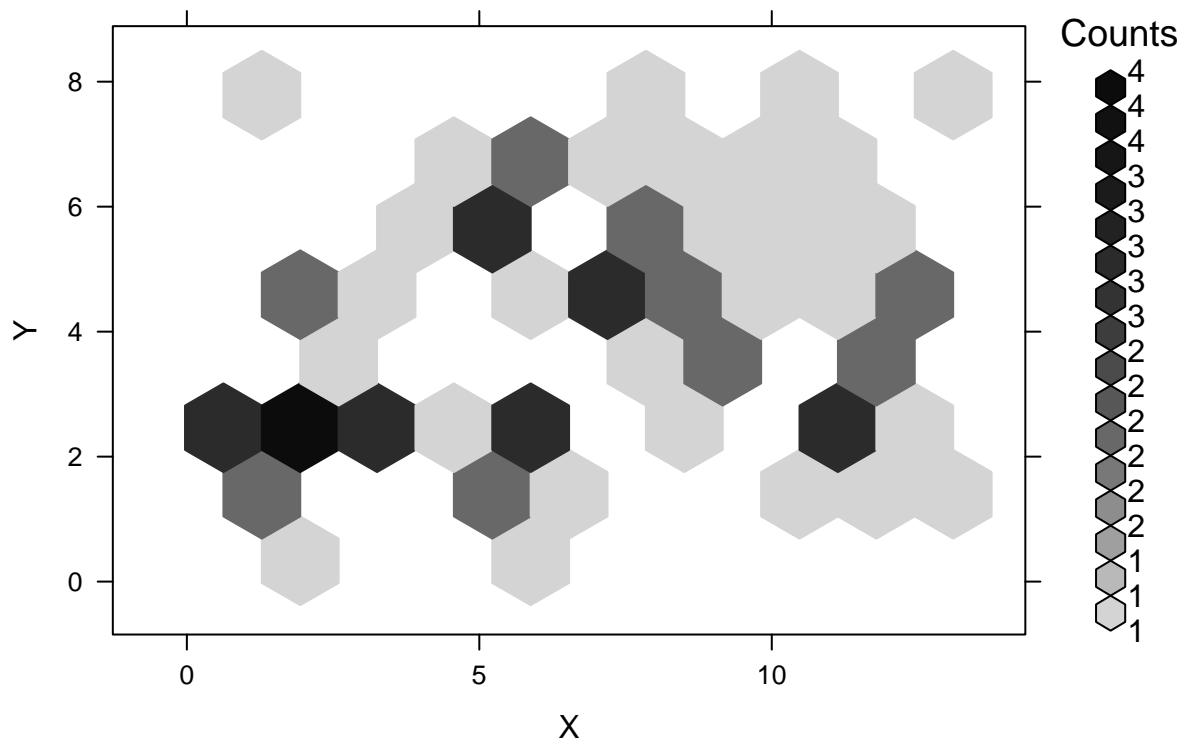
b. **(3 marks)** Repeat part (a) but use hexagonal binning in two different ways to convey the density. Comment on where Waldo has a tendency to be located on the page.

```r
library('hexbin')
```

```
## Warning: package 'hexbin' was built under R version 4.0.4
```
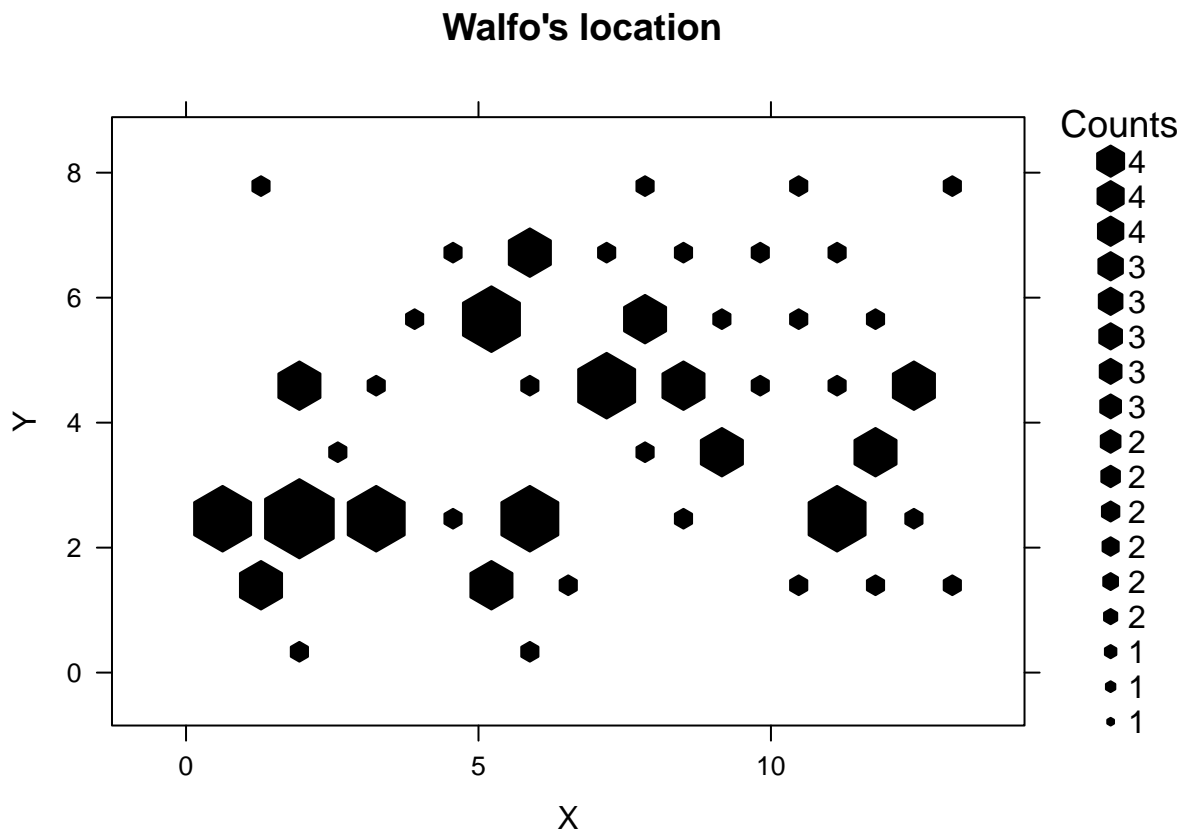
```r
hexbinplot(Y~X, data=waldo, xbins=9, aspect = 8/12, xlim=extendrange( xlim,f=0.1), ylim=extendrange( ylim,f
```

Walfo's location


```
hexbinplot(Y~X, data=waldo, style="lattice", xbins=9, aspect = 8/12, xlim=extendrange( xlim,f=0.1), ylim=ex
```

```
## Warning in fun(maxcnt = base::quote(4L), trans = base::quote(NULL), inv =
## base::quote(NULL), : legend shows relative sizes
```

Walfo's location

Waldo has a high probability to be at lower to middle of y axis (y=2 to y=5) of the page,

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

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

```
x= waldo$X
y= waldo$Y
den = kde2d(x=x, y=y)
zlim = range(den$z)

plot(x,y,pch=19,main = "Walfo's location")
contour(den$x, den$y, den$z, col="grey10", levels = pretty(zlim, 10), lwd=1, add=TRUE)
```

Walfo's location
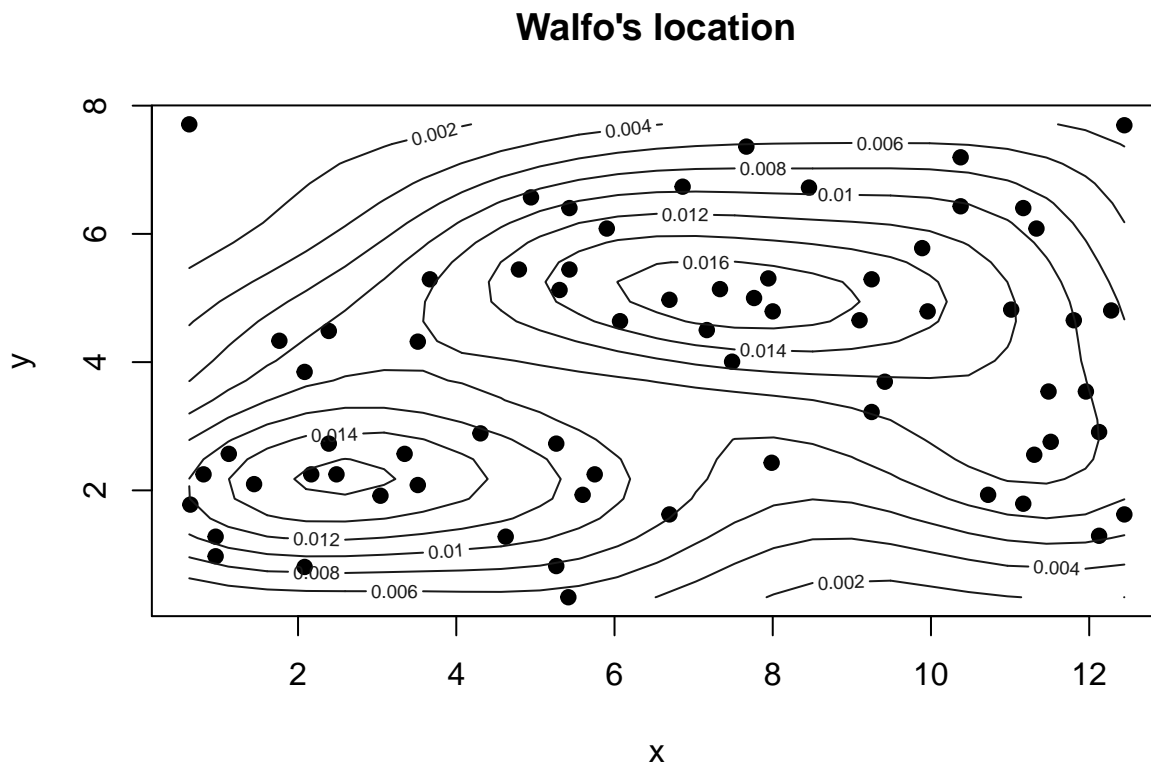
Waldo is most likely to appear at (x=2, y=2) and around (x=7, y=5); he is least likely to appear at top left (x=0,y=8) or bottom right(x=9, y=0)

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.

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```r
par(mfrow=c(3,3))
for (book in 1:7){
    ind = waldo$Book == book
    newdata=waldo[ind,]
    print( ggplot(data = newdata, mapping = aes(X,Y)) + xlim(extendrange(range( newdata$X)))+
        xlim(extendrange(range( newdata$X))) +
        geom_density_2d() + geom_point()+ ggtitle(paste("Walfo's location in book", book)))
}
```
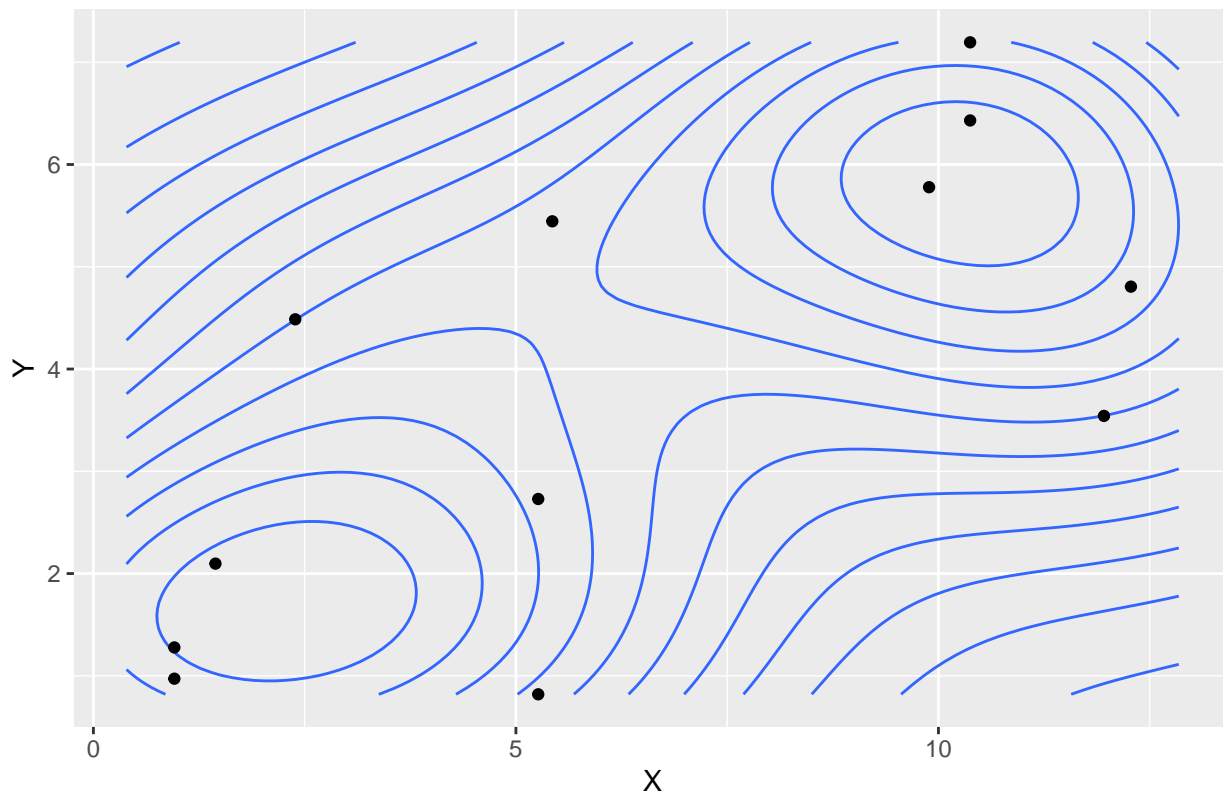
```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```
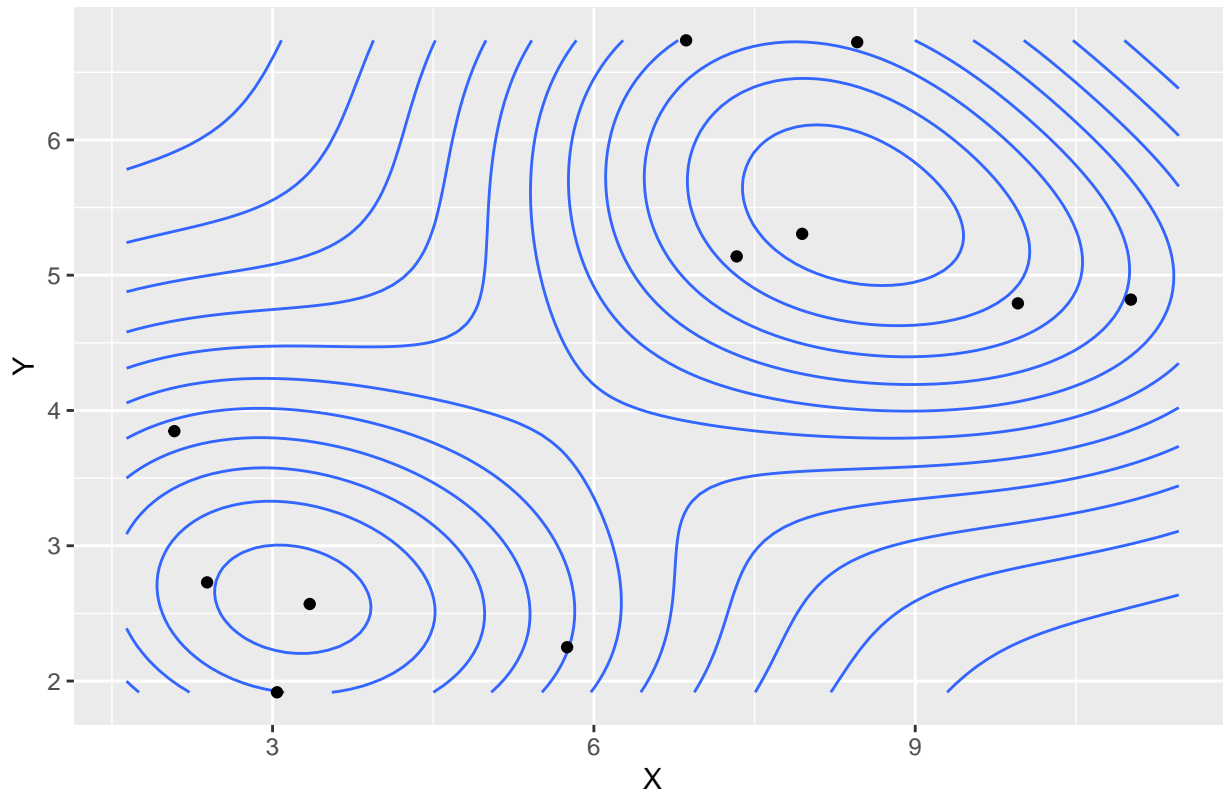
Walfo's location in book 1
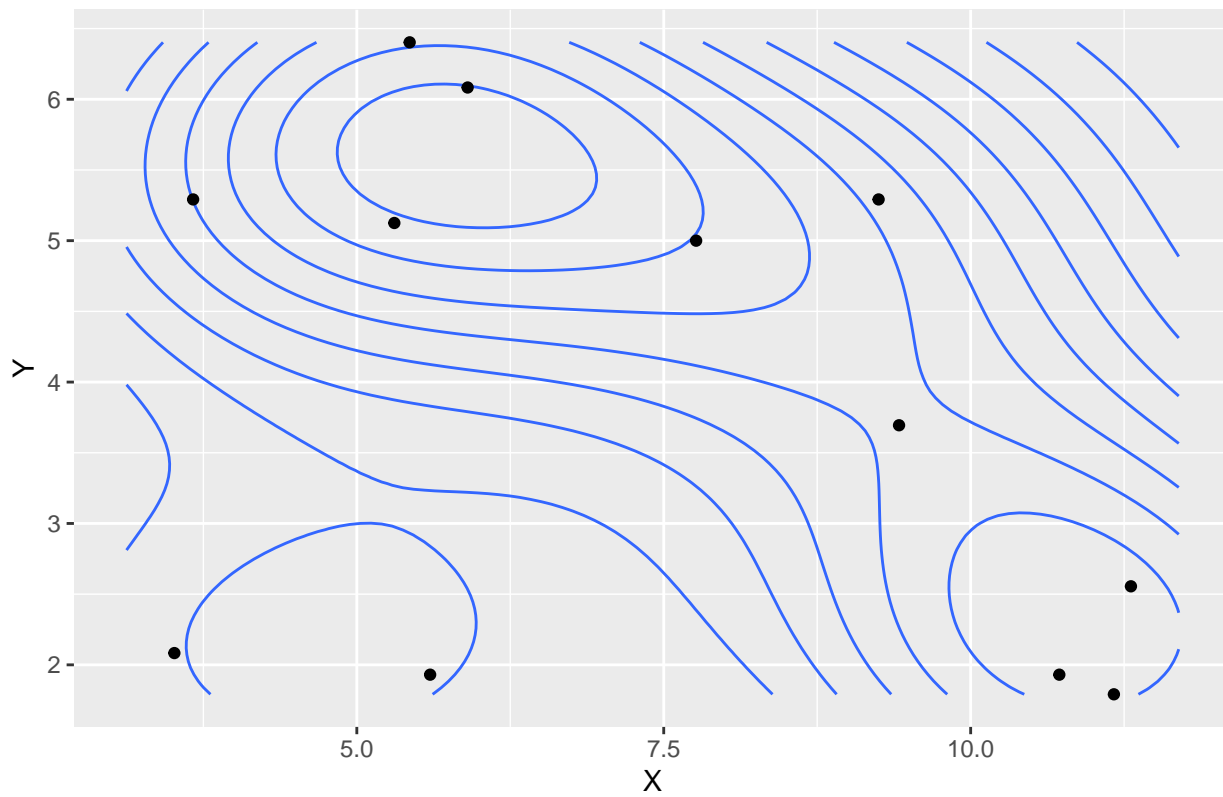

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

Walfo's location in book 2


```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```
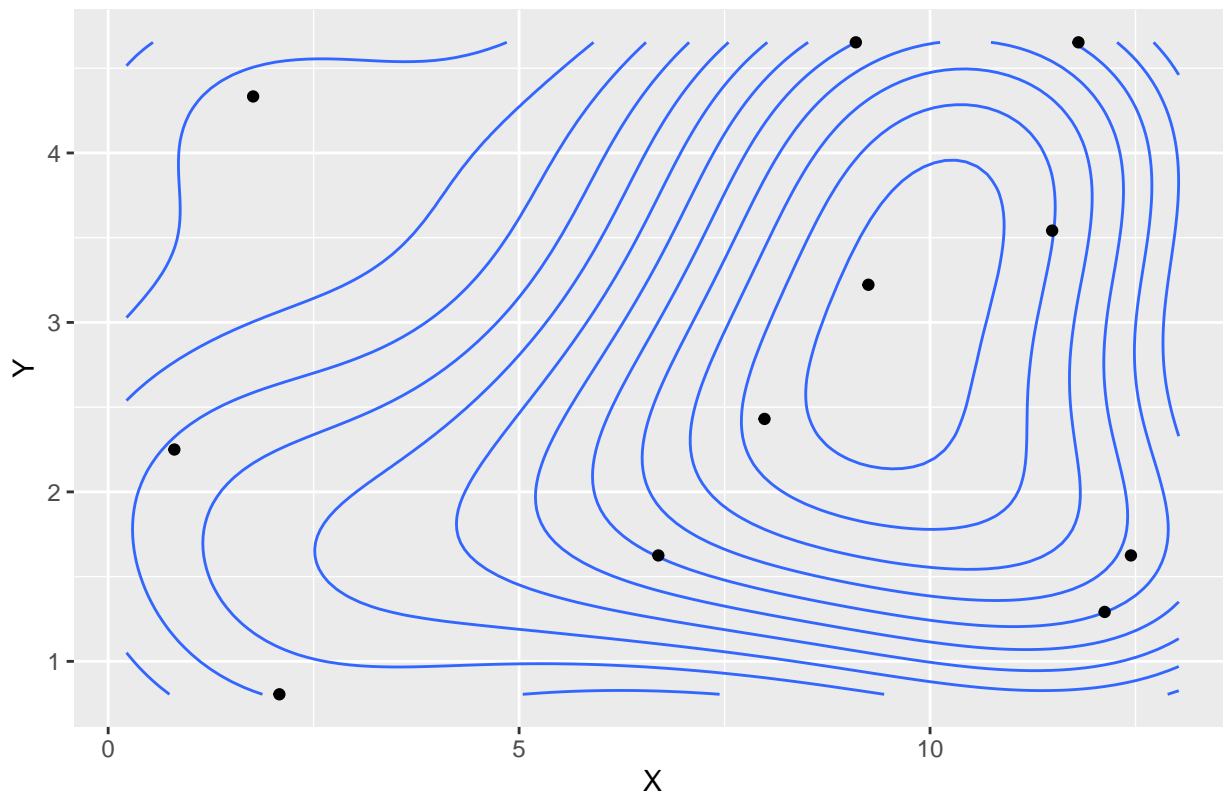
Walfo's location in book 3


```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```
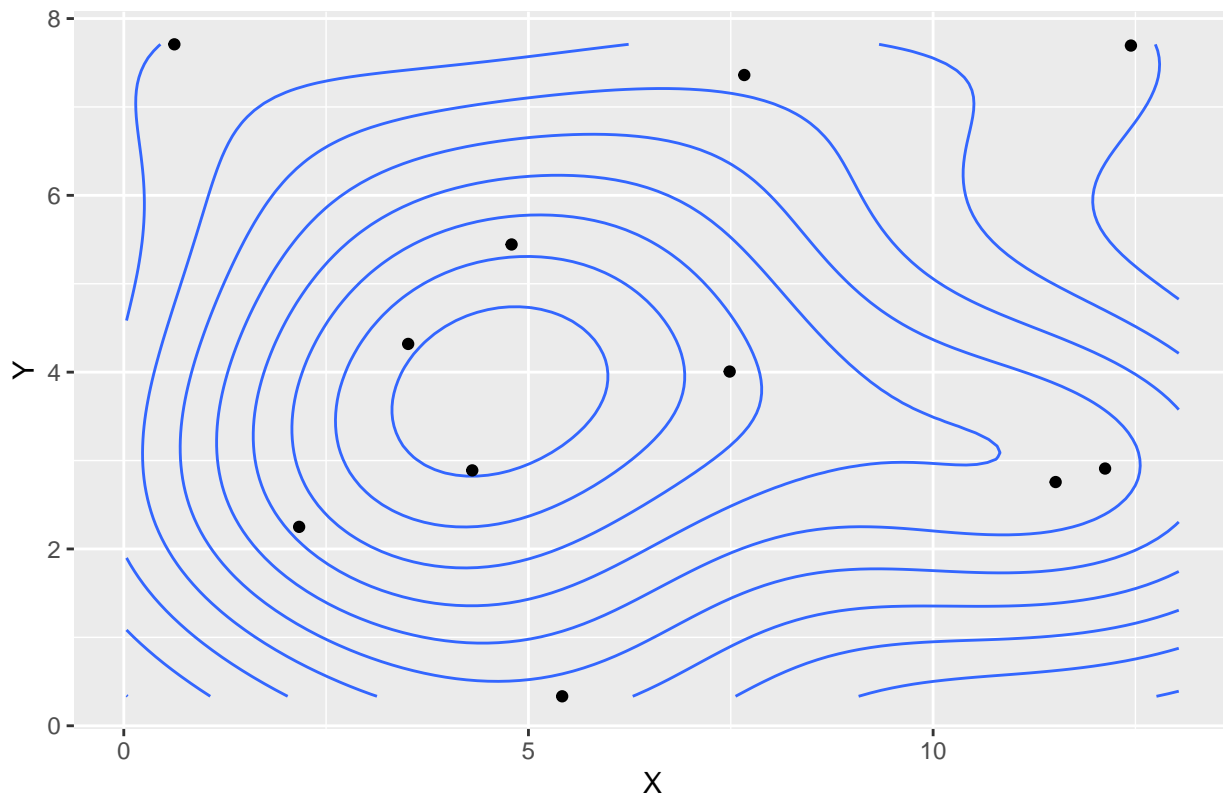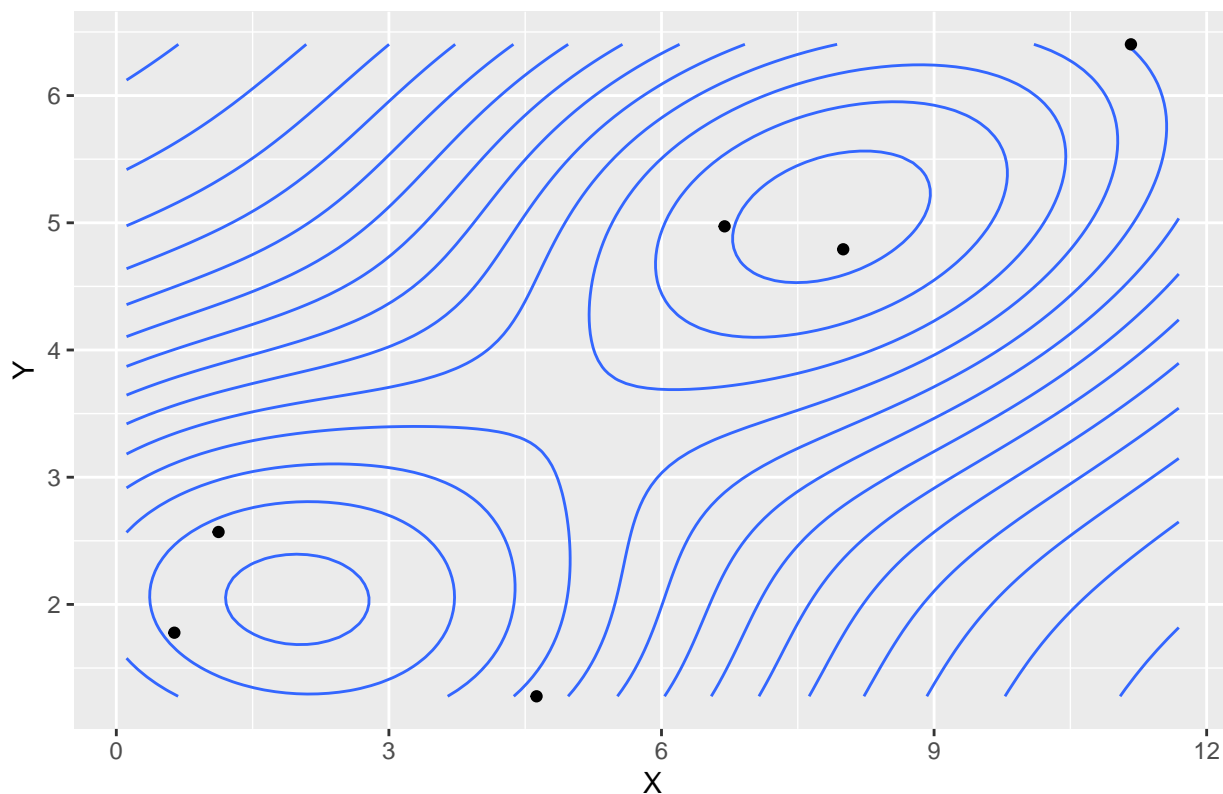
Walfo's location in book 4


```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```
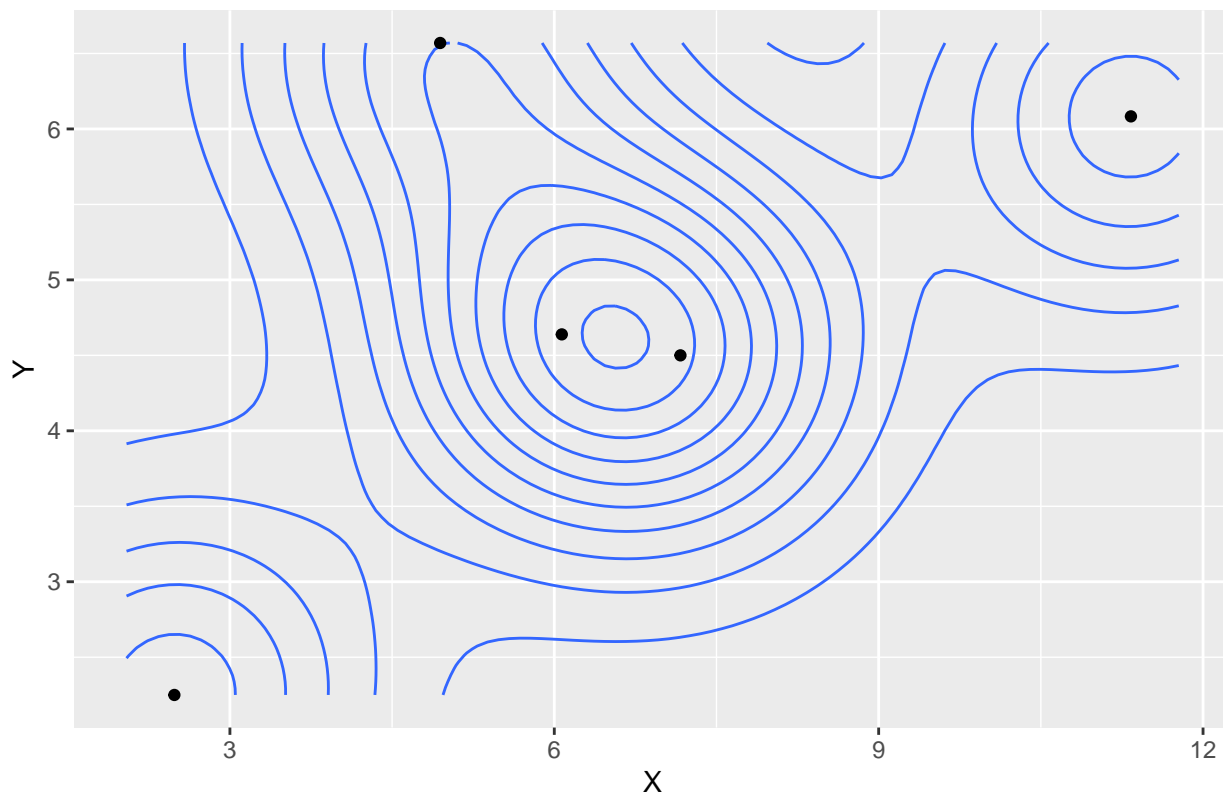
Walfo's location in book 5


```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

Walfo's location in book 6

Walfo's location in book 7

the author seems to be randomly choosing position of Waldo, as there are cases when Waldo's position is similar, and cases when his position vary by a lot in both consequent book(when order of book matter), and through all books

(when order does not matter)