Additional Examples

- **Shoes data:** Experiment reported by Box, Hunter & Hunter (1978). Amounts of “shoe wear”: Two materials (A and B) randomly assigned to left & right shoes (L and R) of 10 boys;

```
> library("BHH2") #reproduces examples in Box, Hunter & Hunter
> data(shoes.data)
> shoes.data
   boy matA sideA matB sideB
1    1 13.2     L 14.0     R
2    2  8.2     L  8.8     R
3    3 10.9     R 11.2     L
4    4 14.3     L 14.2     R
5    5 10.7     R 11.8     L
6    6  6.6     L  6.4     R
7    7  9.5     L  9.8     R
8    8 10.8     L 11.3     R
9    9  8.8     R  9.3     L
10  10 13.3     L 13.6     R
```

Additional Examples

- **One sample t-test:**

```
> attach(shoes.data)
The following objects are masked from shoes.data (pos = 3):

    boy, matA, matB, sideA, sideB

> t.test(matA, mu=10)

   One Sample t-test

data:  matA
t = 0.81272, df = 9, p-value = 0.4373
alternative hypothesis: true mean is not equal to 10
95 percent confidence interval:
  8.876427 12.383573
sample estimates:
mean of x 
    10.63 
```

Additional Examples

Two-sample tests: paired vs. unpaired (equal vs. unequal variances)

- Two-sample unpaired, unequal variances
  - Test for equal variances:

```
> var.test(matA,matB)
   F test to compare two variances

data:  matA and matB
F = 0.94739, num df = 9, denom df = 9, p-value = 0.9372
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2353191 3.8142000
sample estimates:
ratio of variances
         0.9473933
> var(matA)
[1] 6.009
> var(matB)
[1] 6.342667
```

Additional Examples

- t-test with unequal variances (default in **R**)

```
> t.test(matA,matB)

   Welch Two Sample t-test

data:  matA and matB
t = -0.36891, df = 17.987, p-value = 0.7165
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.745046  1.925046
sample estimates:
mean of x mean of y
    10.63     11.04
```

Additional Examples

- t-test with equal variances

```
> t.test(matA,matB,var.equal =T)

   Two Sample t-test

data:  matA and matB
t = -0.36891, df = 18, p-value = 0.7165
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.744924  1.924924
sample estimates:
mean of x mean of y
    10.63     11.04
```

Additional Examples

- Paired t-test

```
> t.test(matA,matB,paired=T)

   Paired t-test

data:  matA and matB
t = -3.3489, df = 9, p-value = 0.008539
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6869539 -0.1330461
sample estimates:
mean of the differences
                  -0.41
```

Additional Examples

Density Estimation

$$\hat{f}(x) = \frac{1}{nb} \sum_{j=1}^{n} K\left(\frac{x - x_j}{b}\right)$$

bandwidth **Kernel**

```
density(x, ...)
## Default S3 method:
density(x, bw = "nrd0", adjust = 1,
        kernel = c("gaussian", "epanechnikov", "rectangular",
                   "triangular", "biweight",
                   "cosine", "optcosine"),
        weights = NULL, window = kernel, width,
        give.Rkern = FALSE,
        n = 512, from, to, cut = 3, na.rm = FALSE, ...)
```

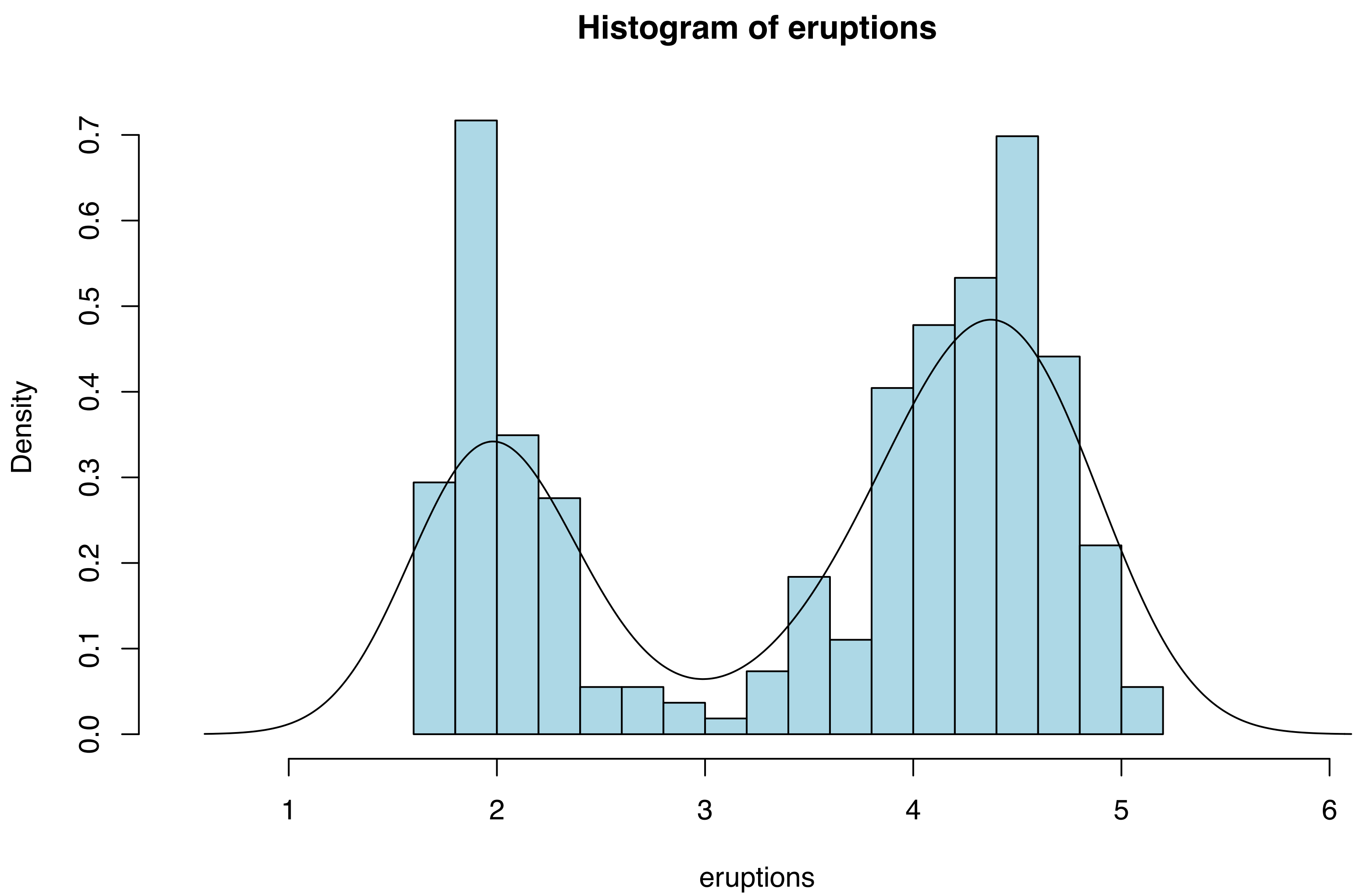
Additional Examples

Default bandwidth choice in **R** is

$$\hat{b} = 0.9\min(\hat{\sigma}, R/1.34)n^{-1/5}$$

with R the IQR.

```
> attach(faithful)
> hist(eruptions,breaks=15,xlim=c(0.5,6),col='lightblue',prob=T)
> lines(density(eruptions))
```

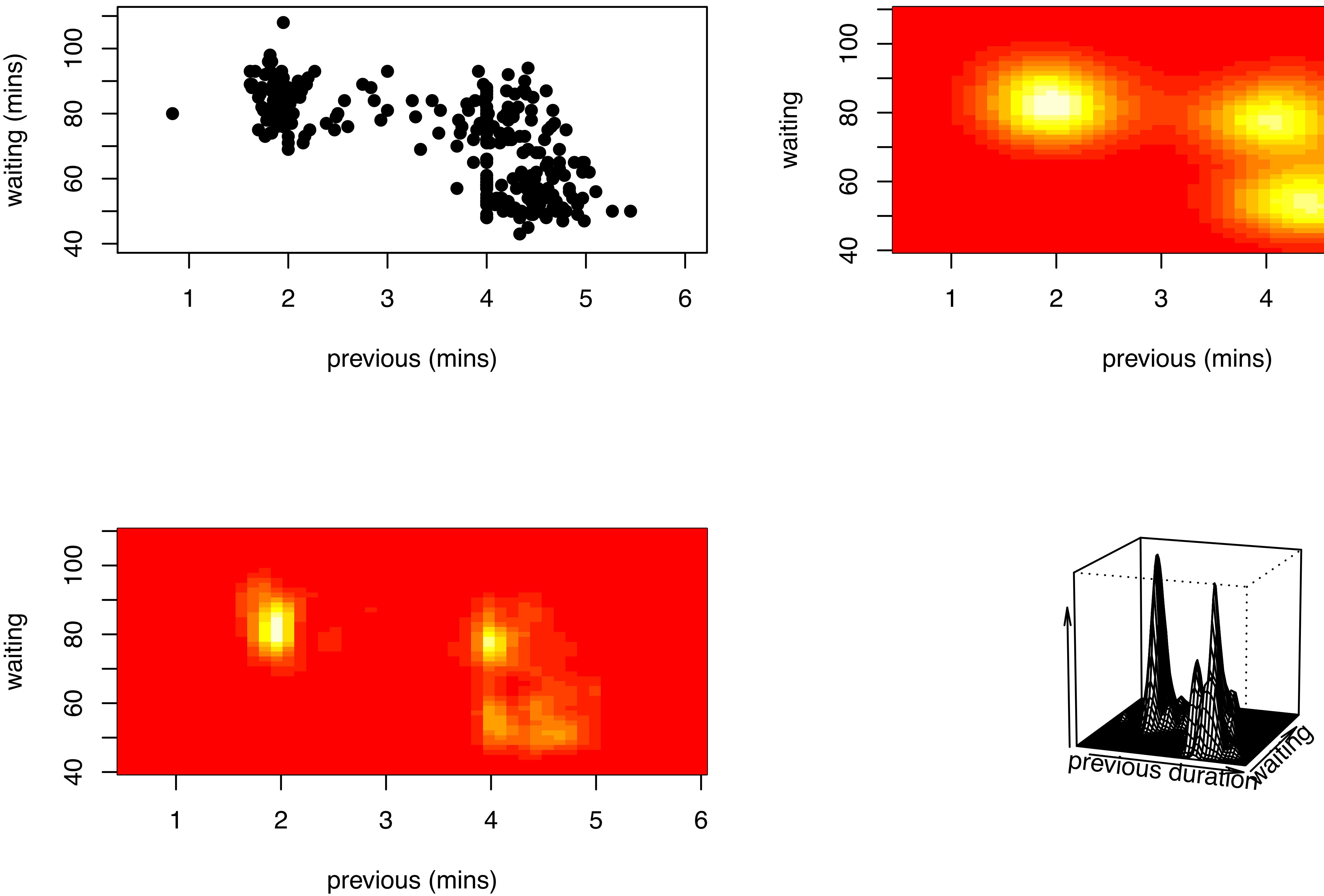


Additional Examples

- 2-dimensional density estimation:

```
> library(MASS)
> attach(geyser)

> par(mfrow=c(2,2))
> plot(duration,
waiting,xlab=“previous(mins)",ylim=c(40,110),xlim=c(0.5,6),
     ylab="waiting (mins)”,pch=19)

> f1=kde2d(duration,waiting,n=50,lims=c(0.5,6,40,110))
> image(f1,xlab="previous (mins)”,ylab="waiting")

> f2=kde2d(duration,waiting,n=50,lims=c(0.5,6,40,110),
    h=c(width.SJ(duration),width.SJ(waiting)))
> image(f2,xlab="previous (mins)”,ylab="waiting")

> persp(f2,phi=15,theta=20,d=10,xlab=“previous duration”,
     ylab="waiting",zlab="")
```

Additional Examples



Additional basic inference tools

- Test for proportions based on large sample approx.:

$$H_0 : \quad p = p_0$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

or equivalently:

$$Z = \frac{y - np_0}{\sqrt{np_0(1 - p_0)}}$$

Additional basic inference tools

- Test for proportions based on large sample approx.:
  - Continuity correction;
  - Exact test;

- Nonparametric methods: read

- Two-sample inference: we have described most of the tests but read about permutation tests.