Univariate Plots

PH345: Winter 2025

Phil Boonstra

Univariate Plots

Univariate plots are used to visualize the distribution of a single variable.

Examples

- Histograms
- Boxplots
- Barplots

Ultra-runner data (Samtleben, 2023)

$n = 288$ ultra-runners (completing 100km ultra-marathons)

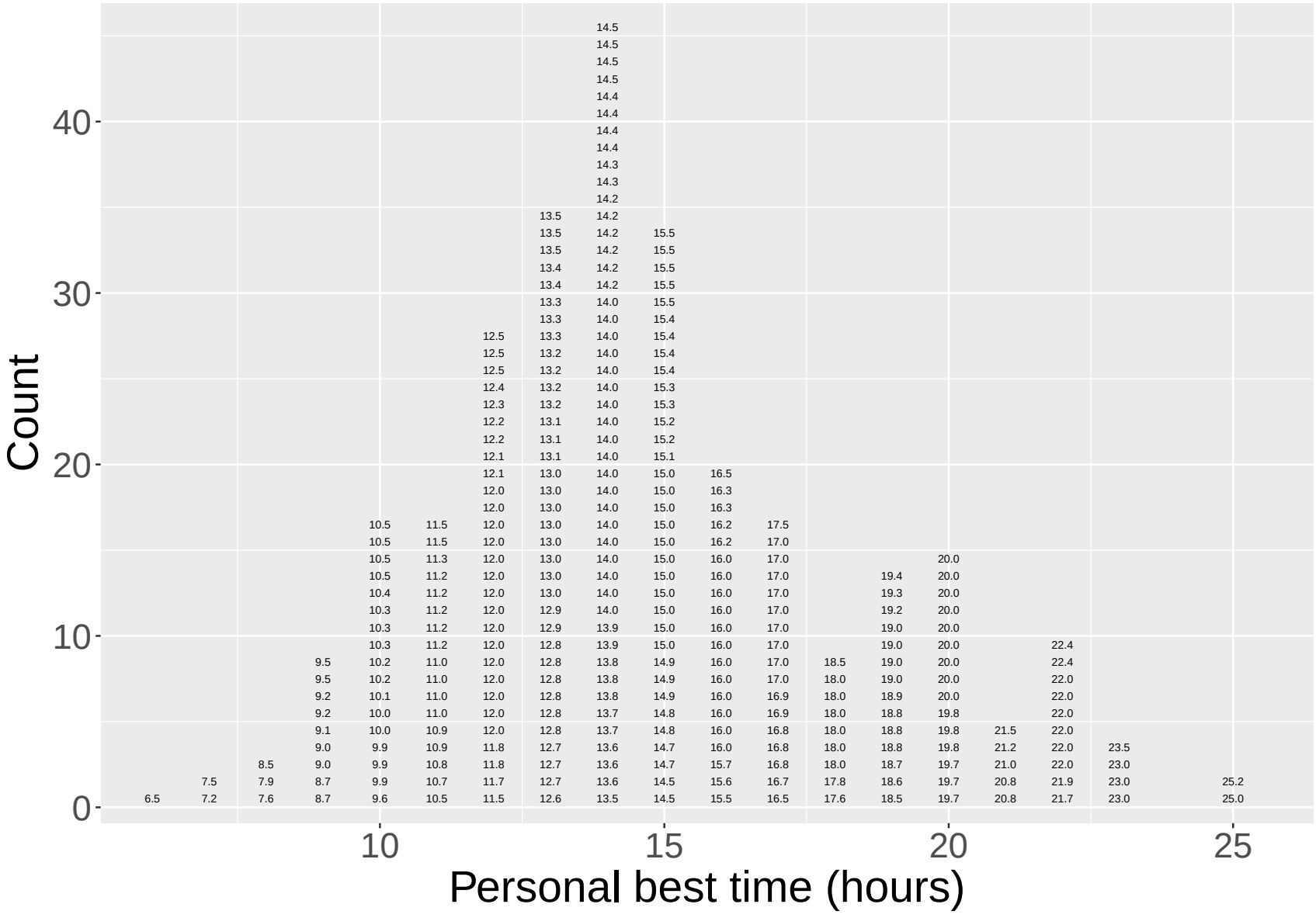
Each runner's personal best (in hours):

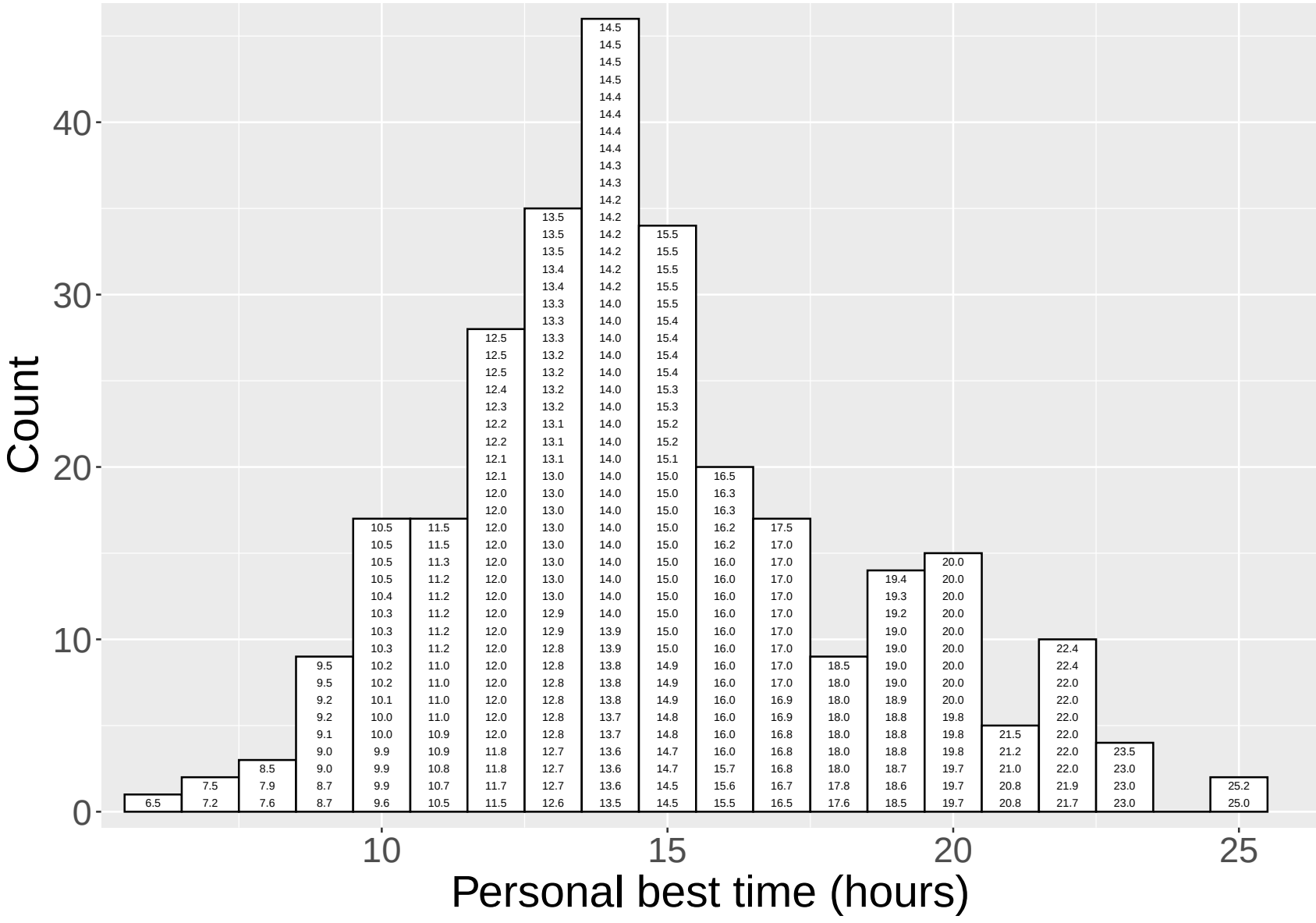
```
14 7.6 14.2 14.33 17 12 16 16.16 9.95 17.55 12.5 23 18.85 8.5 19.33 16 18 12
14.38 15 19.25 14 14.21 14.91 14.5 19 18.5 15 20 12.16 14.82 12.99 13.5 12.98
9.2 10 13.55 14 15 14 22 15.33 15.53 12.26 12 13 10.77 13.98 14 16 12.05 20.83
14 14.3 14.78 13.83 16 12.9 19.67 14 16.67 10.25 15.38 13.35 14 22 7.15 14 12
16 9.5 15.13 12.99 18.77 15 11.25 14 13 14.53 18.75 16 14.5 18.66 15.5 12.77
9.05 16.3 17 22 9.5 15.46 8.7 16.75 12 14.41 10.5 17 11.17 15.5 17 13.86 20
10.45 10.34 13.33 14.5 7.9 11 10.71 12 15.36 19.41 14 9 15.16 12 18.81 10.5 12
14 15 9 20 21.5 11.33 15 21.25 23 22 18.6 21.9 16.16 15.5 13.71 23.5 10.33 8.7
18 12.83 10.49 13.33 14.86 19.99 15.66 22.36 22.4 16 16.52 11.25 13.06 9.6
14.25 20 20 13.75 10.34 12.25 13.25 12 10.95 16.75 13.25 14 13.65 18 18 15 12.7
17.5 19.66 11.51 12.71 12 17 13 6.5 19 19.7 14.25 9.86 23 15.33 14.65 15.6 22
14 14 16.86 14.51 13.51 13.75 18.51 19.75 20.8 15.99 16.34 25 13 16.88 12.95
11.5 12.75 11.16 12.7 10.13 17.01 11.24 12.6 20 14.01 13.05 13.18 12 12 15.38
15 10.52 15.16 9.9 13.5 21.68 20 19 12 14.91 11 14.36 11 17 11.99 12.46 20
15.01 12.41 13.49 14 13.2 13.55 13.96 10.95 16 11.8 17 11.65 13.58 13.09 13.86
16 15 12.08 14.16 11 18 12.85 22 11.5 14.66 10.16 13 7.5 19.84 16.75 12 25.25
15.5 13.36 10 17 12.83 16 12.5 16 9.18 16.5 14.41 14.25 19 15 13.36 17.83 10.5
11.75 12.75 19.75 15.4 21 18 14.46
```

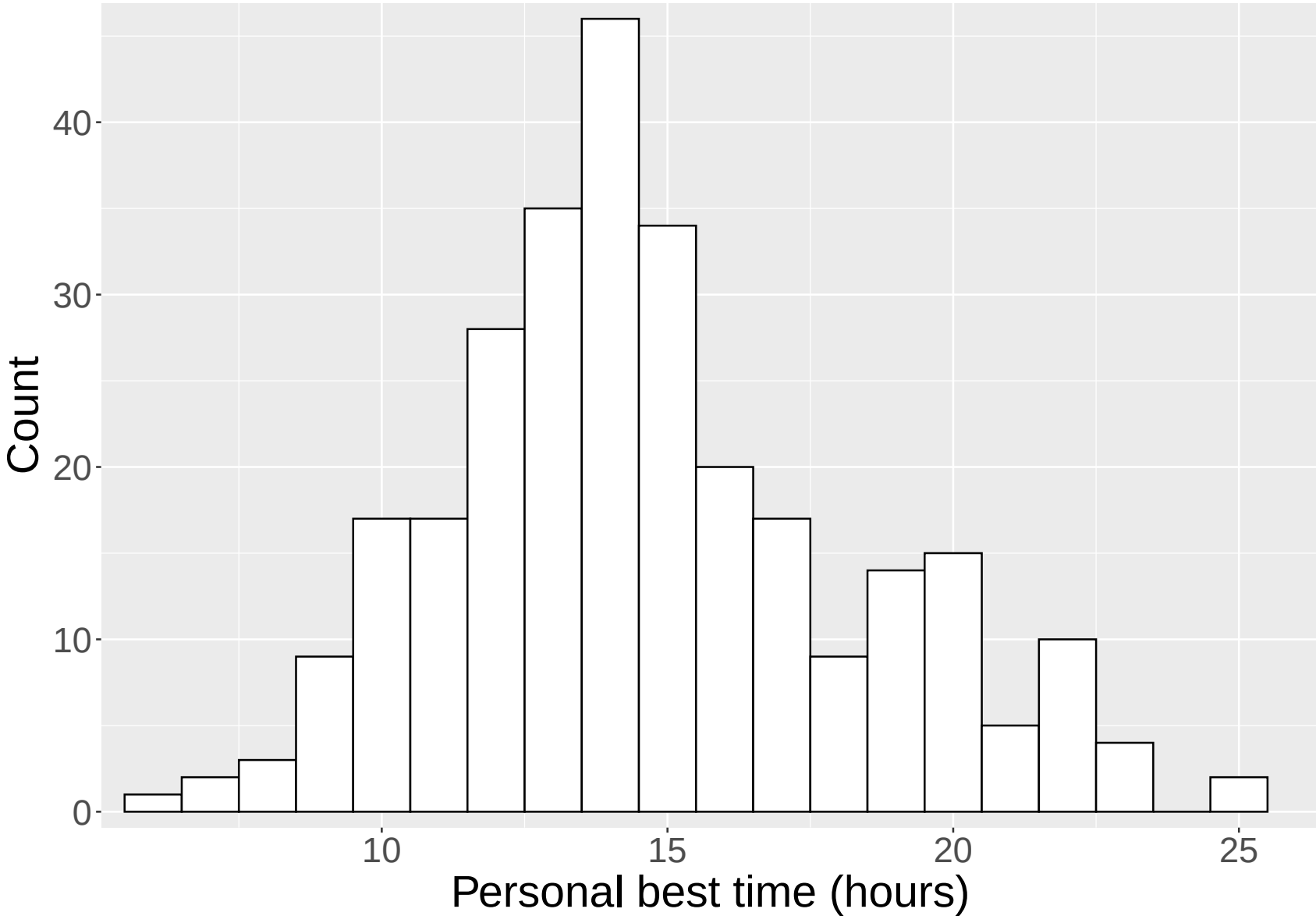
https://causeweb.org/tshs/ultra-running/

Creating a histogram

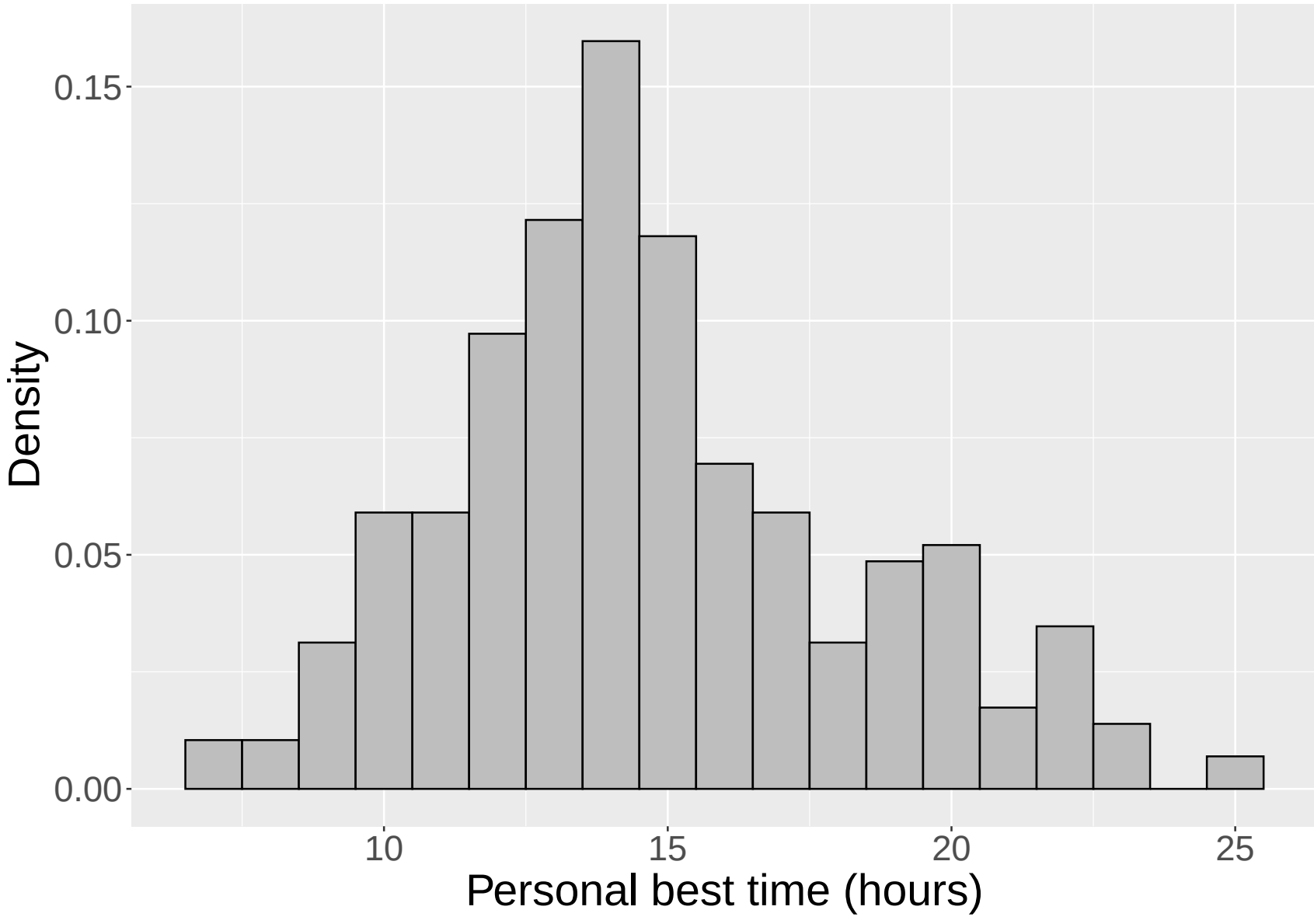
1. Choose a bin size and a center value, e.g. one hour bins centered at the integers would be denoted as $(5.5, 6.5]$, $(6.5, 7.5]$, $(7.5, 8.5]$, etc. Bins must be non-overlapping, and there should be enough bins to completely cover the data.
2. Assign each runner to a bin, e.g. 12.98 goes into the $(12, 13]$ bin and 12.0 goes in to the $(11, 12]$ bin
3. Plot bars for each bin, with the height of the bar corresponding to the number of runners in that bin


Count
0
10
20
30
40
10
15
20
25
Personal best time (hours)

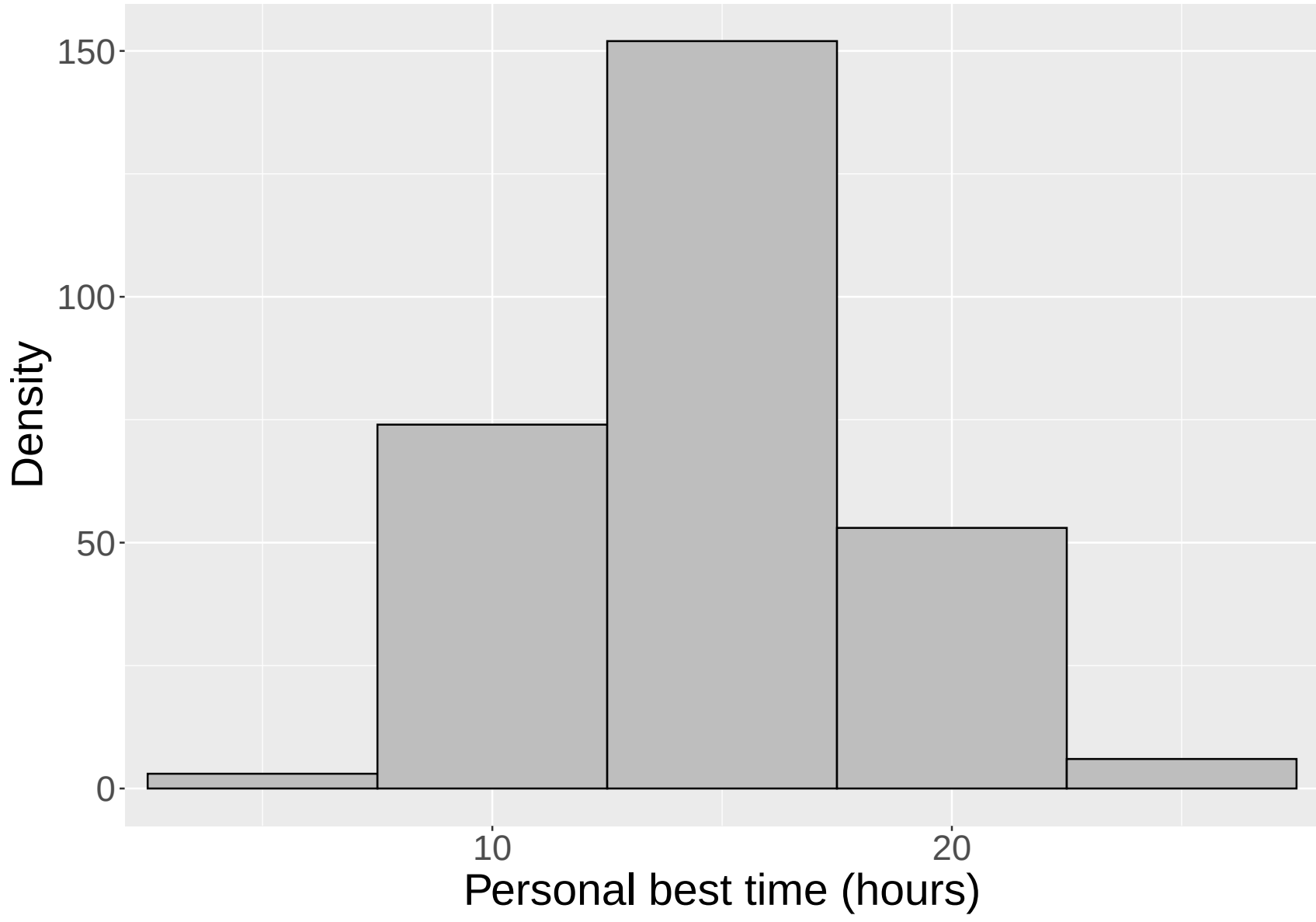

Count
0
10
20
30
40
10
15
20
25
Personal best time (hours)
6.5
7.2
7.5
7.6
7.9
8.5
8.7
9.0
9.1
9.2
9.2
9.5
9.5
9.6
9.9
9.9
10.0
10.1
10.2
10.2
10.3
10.3
10.3
10.4
10.5
10.5
10.5
10.5
10.5
10.5
11.5
11.5
11.3
11.2
11.2
11.2
11.2
11.0
11.0
11.0
10.9
10.9
10.8
10.7
10.5
12.5
12.5
12.5
12.4
12.3
12.2
12.2
12.1
12.1
12.0
12.0
12.0
12.0
12.0
12.0
12.0
12.0
11.8
11.8
11.7
11.5
13.5
13.5
13.5
13.4
13.4
13.3
13.3
13.3
13.2
13.2
13.2
13.1
13.1
13.0
13.0
13.0
13.0
13.0
13.0
12.9
12.9
12.8
12.8
12.8
12.8
12.8
12.7
12.7
12.7
12.6
14.5
14.5
14.5
14.5
14.4
14.4
14.4
14.4
14.4
14.3
14.3
14.2
14.2
14.2
14.2
14.2
14.2
14.0
14.0
14.0
14.0
14.0
14.0
14.0
14.0
14.0
14.0
14.0
13.9
13.9
13.8
13.8
13.8
13.7
13.7
13.6
13.6
13.6
13.5
15.5
15.5
15.5
15.5
15.5
15.4
15.4
15.4
15.4
15.3
15.3
15.2
15.2
15.1
15.0
15.0
15.0
15.0
15.0
15.0
15.0
15.0
15.0
15.0
14.9
14.9
14.9
14.8
14.8
14.7
14.7
14.5
14.5
16.5
16.3
16.3
16.2
16.2
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
16.0
15.7
15.6
15.5
17.5
17.0
17.0
17.0
17.0
17.0
17.0
17.0
17.0
17.0
16.9
16.9
16.8
16.8
16.8
16.7
16.5
18.5
18.0
18.0
18.0
18.0
18.0
18.0
18.0
17.8
17.6
19.4
19.3
19.2
19.0
19.0
19.0
19.0
18.9
18.8
18.8
18.8
18.7
18.6
18.5
20.0
20.0
20.0
20.0
20.0
20.0
20.0
20.0
20.0
19.8
19.8
19.7
19.7
19.7
22.4
22.4
22.0
22.0
22.0
22.0
22.0
22.0
21.9
21.7
21.5
21.2
21.0
20.8
20.8
23.5
23.0
23.0
23.0
25.2
25.0


Count
40
30
20
10
0
10
15
20
25
Personal best time (hours)

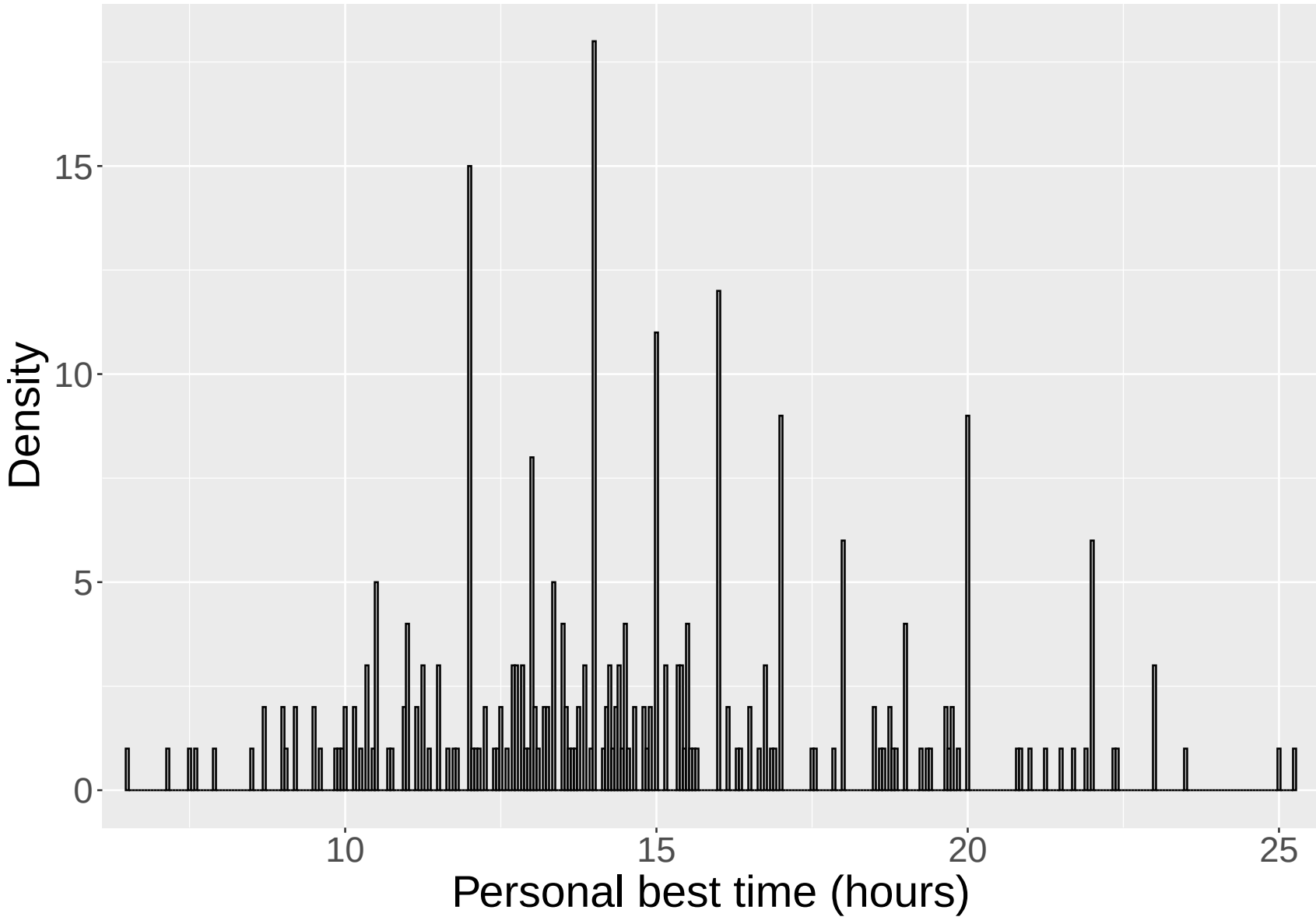
Counts vs density



Bin width of 5 hours (too large!)



Bin width of 3 minutes (too small!)



Boxplots

References

Samtleben, E. (2023) Ultrarunning dataset. Teaching of Statistics in the Health Sciences Resource Portal, Available at https://www.causeweb.org/tshs/ultra-running/.