# HW1R

Likhit Raj Yerramsetty

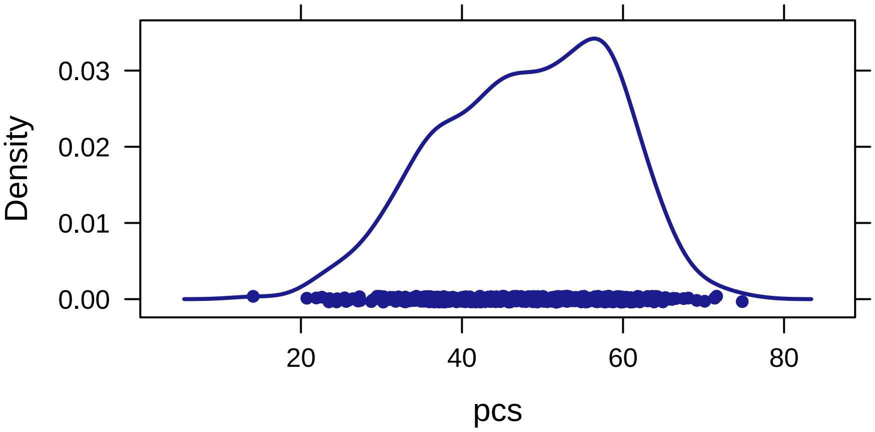
**Problem 1** The **HELPrct** dataset in the mosaicData package includes data from the Health Evaluation and Linkage to Primary Care study, which was conducted in Boston 10 years ago. One of the study variables is a measure of physical function, with higher scores being better (possible scores can range from 0 to 100 points). Describe the sample size plus CENTER, SPREAD and SHAPE of this distribution, providing only a single measure of center and a single measure of spread. Be sure to provide an interpretation in the context of the problem. Could you provide any different graph to describe the distribution of this variable? (Please do it)

```
favstats(~ pcs, data=HELPrct)
```

```
##       min       Q1   median       Q3      max     mean      sd   n missing
##  14.07429 40.38438 48.87681 56.95329 74.80633 48.04854 10.7846 453       0
```

```
densityplot(~ pcs,
  main="Figure 1: Density plot\nof Physical Component Scores from HELP study",
  data=HELPrct)
```

Figure 1: Density plot
of Physical Component Scores from HELP study



SOLUTION:

SAMPLE SIZE: n = 453 participants

CENTER: median = 48.88 pcs

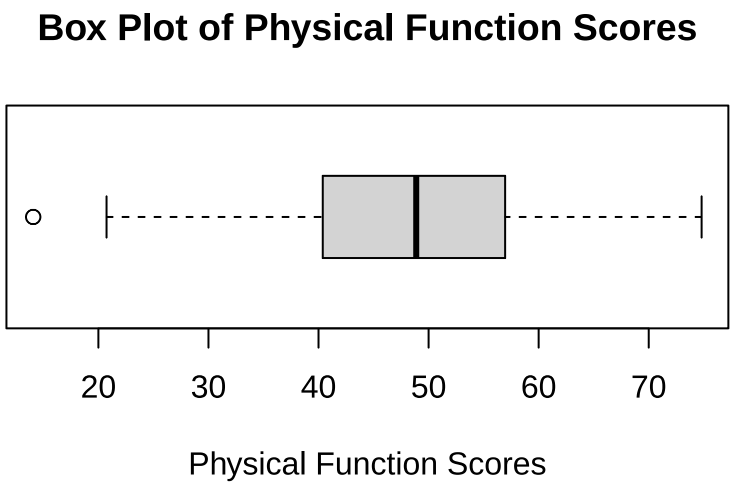
- The median is a bit higher than the mean for this data and as the distribution is left skewed.

SPREAD: sd = 10.78 pcs

-The standard deviation is expressed in the same units as the original data, making it straightforward to interpret. And We can get the sd gives us the values around the mean while also taking the extreme values into account

SHAPE: skewed left

```
boxplot(HELPrct$pcs,
        horizontal = TRUE,
        main = "Box Plot of Physical Function Scores",
        xlab = "Physical Function Scores")
```



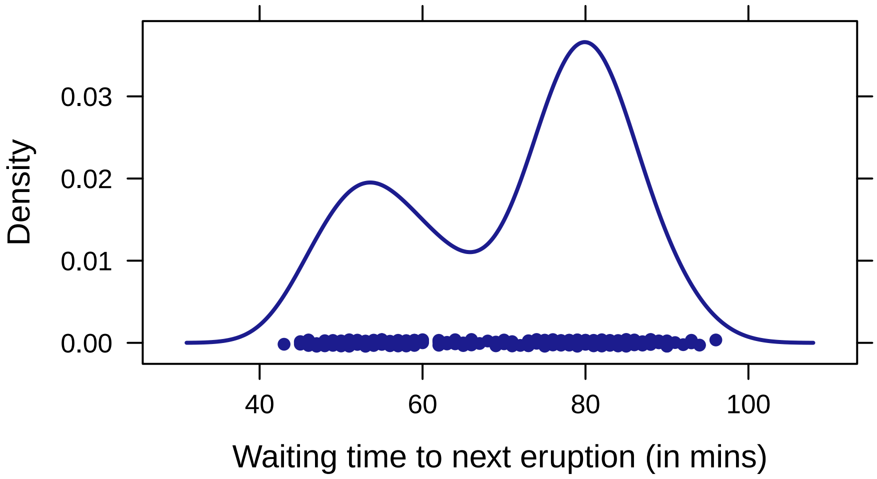
**Problem 2 (Old Faithful)** The **faithful** dataset contains the waiting time (in minutes) to the next eruption of the Old Faithful geyser in Yellowstone National Park in Wyoming. Describe the sample size plus CENTER, SPREAD and SHAPE of this distribution, providing only a single measure of center and a single measure of spread. Be sure to provide an interpretation in the context of the problem (and don't forget to specify units).Could you provide any different graph to describe the distribution of this variable? (Please do it)

```
favstats(~ waiting, data=faithful)
```

```
##  min Q1 median Q3 max     mean       sd   n missing
##   43 58     76 82  96 70.89706 13.59497 272       0
```

```
densityplot(~ waiting,
  xlab="Waiting time to next eruption (in mins)",
  main="Figure 2: Density plot of Old Faithful geyser dataset", data=faithful)
```

**Figure 2: Density plot of Old Faithful geyser dataset**



SOLUTION:

SAMPLE SIZE: n = 272 observations

CENTER: median = 76 minutes

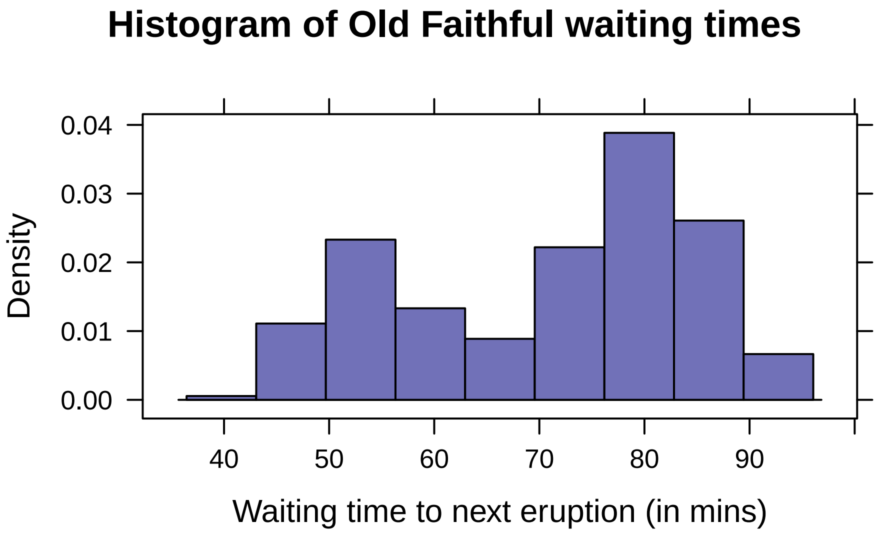
-median is a better measurement of central tendency for this data set since the median is shifted toward the higher density peak of this bimodal distribution.

SPREAD: 13.59 minutes

- The standard deviation provides a measure of the dispersion of the data points around the mean. It takes into account all data points in the dataset and is influenced by extreme values.

SHAPE: bimodal

```r
histogram(~ waiting,
  xlab="Waiting time to next eruption (in mins)",
  main="Histogram of Old Faithful waiting times", data=faithful)
```


Histogram of Old Faithful waiting times
Density
0.04
0.03
0.02
0.01
0.00
40
50
60
70
80
90
Waiting time to next eruption (in mins)