We are going to model the wind velocity at Twin Fall, Idaho as a random process. The data we have to build our model is the average wind velocity for each hour of the day for the last three years (2010-2012). The dataframe can be accessed from RStudio on the web using the Import Dataset option on the Workspace tab. The file that you want is /home/stob/data/TwinfallsID.csv. The variable that we are modeling — average hourly wind velocity — is often the one on which wind models built.

Graph the data

```R
> histogram(TwinfallsID$WS, type = "density")
```

![Histogram of Wind Velocity](image)

Decide on the form of the density

The Weibull density function is often used for this sort of data. The Weibull density is given by the function

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(x/\beta\right)^\alpha} \quad 0 < x < \infty$$

This density has two parameters: \( \alpha \) and \( \beta \) which are called the shape and scale parameters. Both \( \alpha \) and \( \beta \) are positive real numbers.

We can get an idea of what the shape of various weibull densities look like with the `plotDist` function of the `mosaic` package.

```R
> plotDist("weibull", params = list(shape = 2, scale = 3))
```

Fit the density to the data

We have to choose the \( \alpha \) and \( \beta \) that fit the data “best.” The function `fitdistr` of the `MASS` package does this.

One problem is that the Weibull density assumes all values of the variable are greater than 0. The wind velocity is 0 for several entries in our dataset. We will change that to a small positive number:

```R
> TwinfallsID$WS[TwinfallsID$WS == 0] = 0.01
> fitdistr(TwinfallsID$WS, "weibull")
```

```
shape      scale
1.6986347  6.6549337
(0.0079714) (0.0255138)
```
Check the fit

\[
\texttt{fit} = \texttt{fitdistr(TwinfallsID$WS, \"weibull\")}
\]
\[
\texttt{a = fit$estimate[1]}
\]
\[
\texttt{b = fit$estimate[2]}
\]
\[
\texttt{x = seq(0, 30, 0.01)}
\]
\[
\texttt{histogram(TwinfallsID$WS, type = \"density\")}
\]
\[
\texttt{ladd(panel.xyplot(x, dweibull(x, a, b), type = \"l\"))}
\]

Using the model

The function `pweibull` answers questions about the area under the curve (i.e., about probability).

\[
\texttt{pweibull(10, a, b)}
\]

[1] 0.86428

A simpler model

The Rayleigh density is just a Weibull density with \( \alpha = 2 \). In this case, the mean of the variable with this density is just \( \beta \sqrt{\pi}/2 \). The mean of our sample is \( \bar{y} \), so it is reasonable to choose \( \beta = 2\bar{y}/\sqrt{\pi} \).