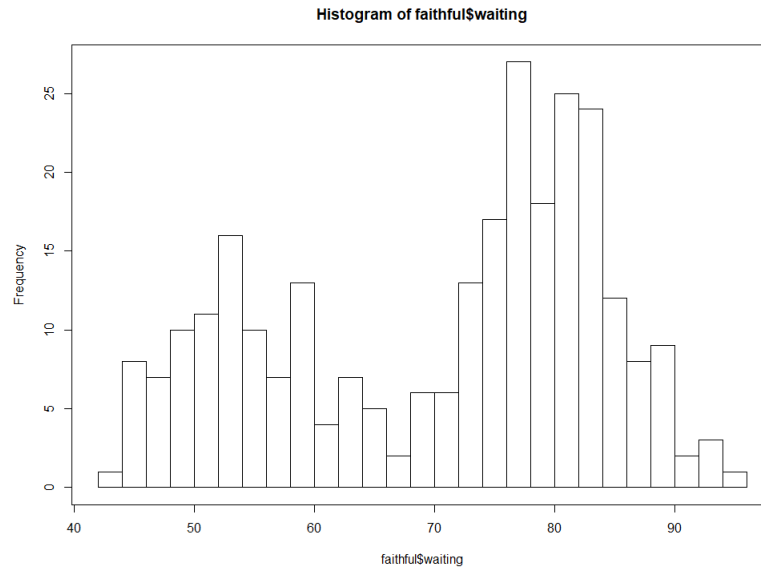


## Statistical Computing: Homework 4

Due on May 12 (Thursday) 8:30am

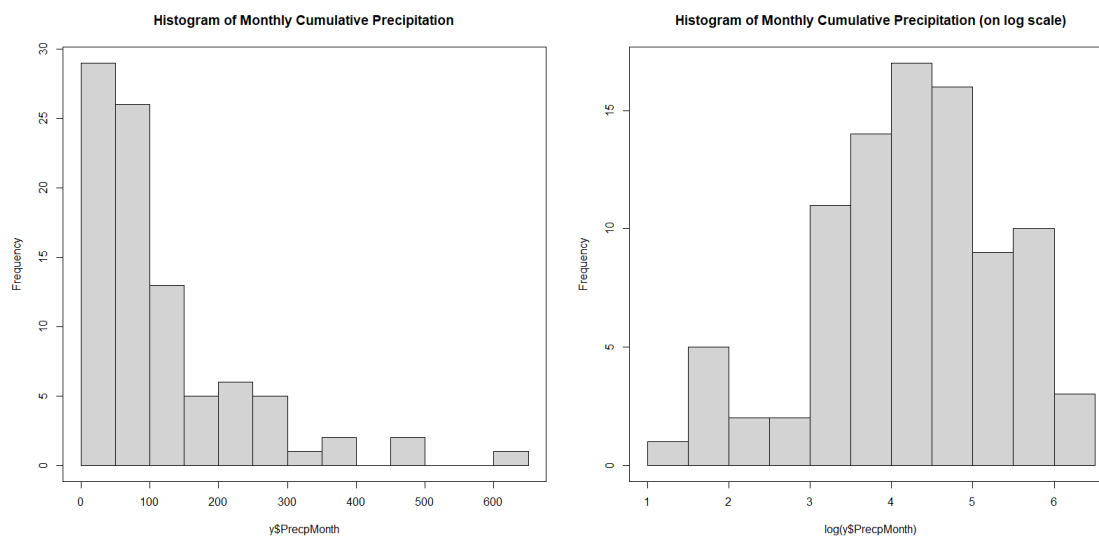
Take one problem from these 2 at your choice. Working on 2 problems gets extra points.

1. The oldfaithful data collect the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, USA. The data can be obtained from R (**faithful**) with 272 observations. In this homework, we ignore the temporal dependence in the data series and treat the 272 observations as independent data. The histogram plot of the **waiting time** between eruptions shows that the distribution has more than one mode.



Fit a mixture model (with 2 components) to the variable **waiting time** by MLE using EM algorithm.

- assume a proper family for the building components in the mixture model
  - show your EM work and report your parameter estimates
  - compare your estimated mixture pdf with the data histogram
2. The precipitation data given in the file “precip.csv” collect the monthly cumulative precipitation at Hsin-Chu from 2014/11/1 to 2022/4/30. The histogram of the 90 observations is shown below. (on the left: original scale with unit mm; on the right: log scale).



Fit a **mixture lognormal distribution** with  $k$  components to the variable **PrecipMonth**. Or, equivalently, fit a **mixture normal distribution** with  $k$  components to the transformed variable  **$\log(\text{PrecipMonth})$** . Use EM for fitting (to estimate the MLE) and suggest a proper  $k$ .

- justify the way you select  $k$  (e.g., AIC or BIC)
- show your EM work and report your parameter estimates
- compare your estimated mixture pdf with the data histogram