Exercise 1. (6 points)

PART 1: Generate 1000 data points randomly in 100-dimensional space, where each dimension is generated from the uniform distribution in (0,1).

```
> uniform <- replicate(100, runif(1000,0,1))
> uniform <- data.frame(uniform)
> head(uniform)
```

PART 2: Perform PCA with this data set. What is the dimensionality of the subspace required to represent (i) 80% of the variance, (ii) 95% of the variance, and (iii) 99% of the variance.

```
> pr<-prcomp(uniform, scale=TRUE)
> summary(pr)
                          PC70    PC91    PC98
Standard deviation     0.88740  0.77635  0.71854
Proportion of Variance 0.00787  0.00603  0.00516
Cumulative Proportion  0.80730 0.95147 0.99012
```

the dimensionality of the subspace required to represent (i) 80% of the variance is 70 (ii) 95% of the variance is 91, and (iii) 99% of the variance is 98.

Exercise 2. (2 points)

Sketch how you would construct a Variogram Cloud to analyze temperature differences across a random sample of cities in the US. What distance measure would you use for the X-axis? How would you reduce the computational effort?

X-axis can be the spatial distance between two samples(cities). The y-axis semivariance is computed across the network. The closer has the smaller y. To reduce the computational effort, we can the data to calculate variogram. For example each object is classified in one of the different land cover classes etc.

Exercise 3. (2 points)

Describe how you might find spikes in a time-series.

We can draw a histogram over every some interval.

Or If a new point is a few number of standard deviations away from some moving mean.