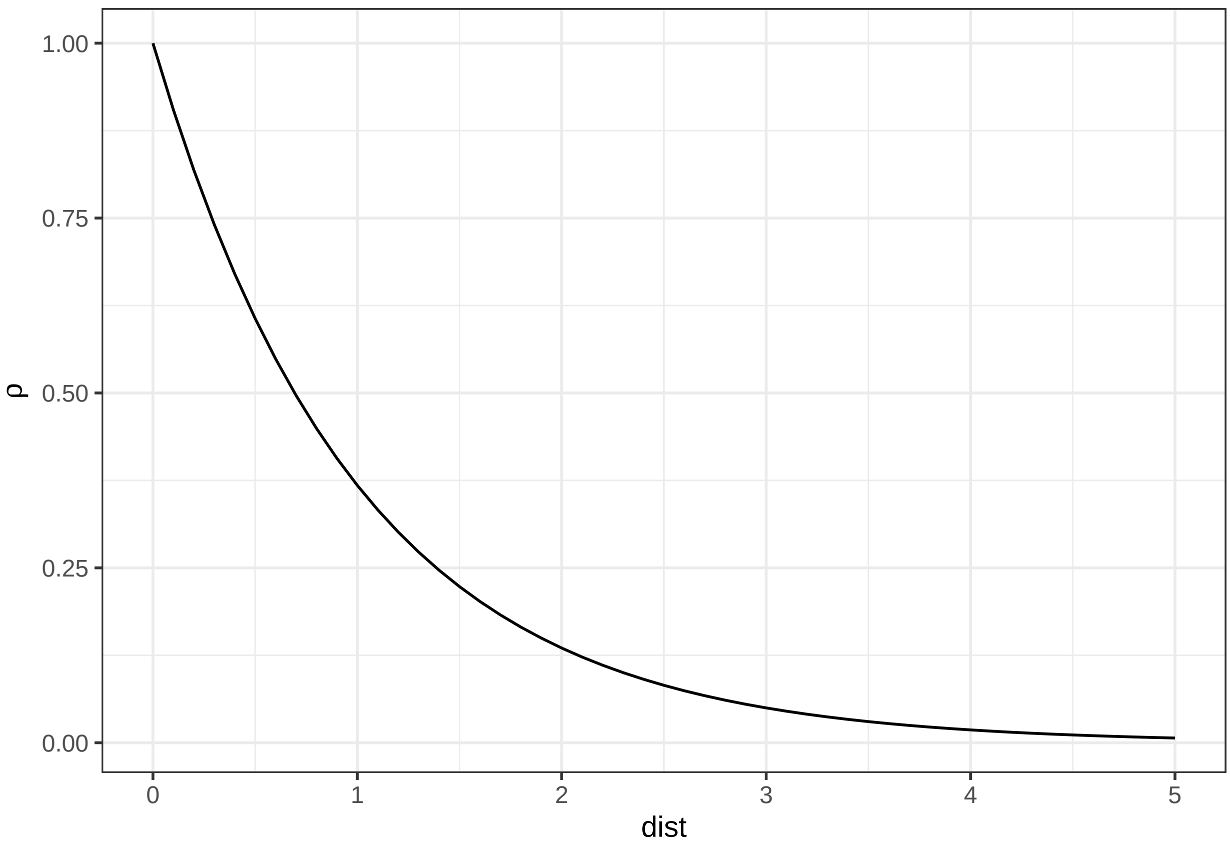
GP Theory Part 2

Correlation function: more details

Recall the variant of the exponential covariance function that we have previously seen. Where d as the Euclidean distance between x_1 and x_2, such that $d = \sqrt{(x_i - x_j)^2}$

$$\rho_{i,j} = \exp\left(-d\right)$$

Lets view the exponential correlation as a function of distance between the two points.



Now let's consider a more general framework for covariance where

$$\sigma_{i,j} = \sigma^2 \exp\left(-d_{ij}/\phi\right)$$

Now we have introduced two new parameters into this function. What do you suppose that they do?

- σ^2: controls the magnitude of the covariance.

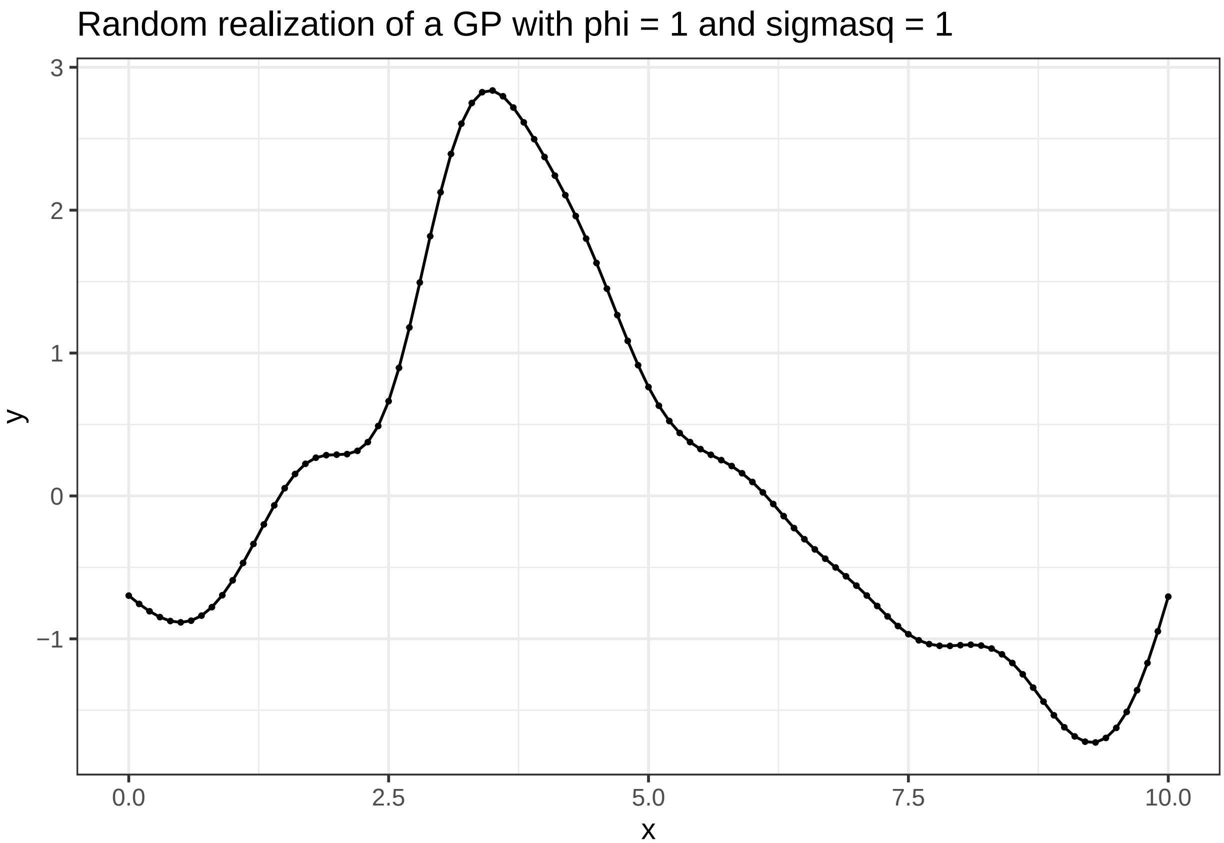
- ϕ: controls the range of the spatial correlation

Modify the following code (from last lecture) to gain an intuition about these parameters.

```
phi <- 1
sigmasq <- 1
x <- seq(0, 10, by = .1)
n <- length(x)
d <- plgp::distance(x)
eps <- sqrt(.Machine$double.eps)
H <- exp(-d/phi) + diag(eps, n)
H[1:3,1:3]
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.9900498 0.9607894
## [2,] 0.9900498 1.0000000 0.9900498
## [3,] 0.9607894 0.9900498 1.0000000
```

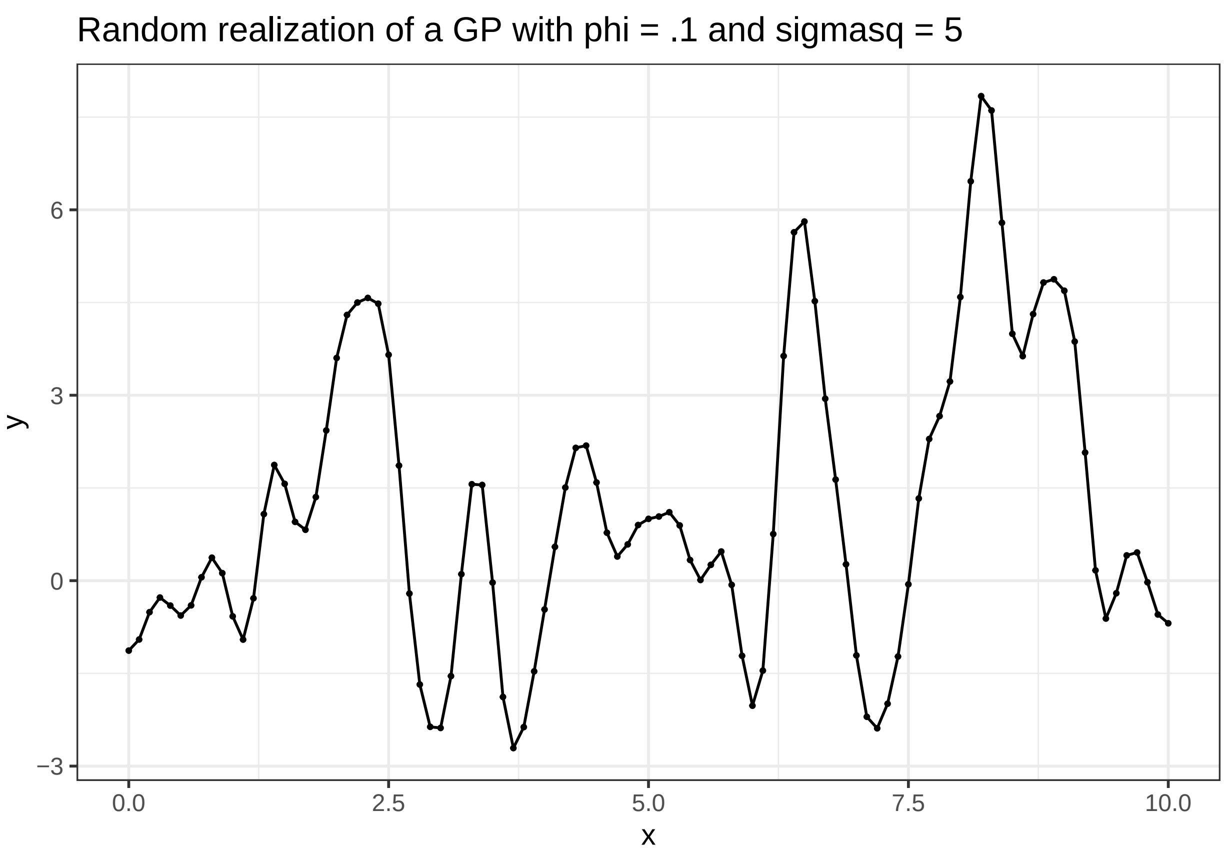
```
y <- rmnorm(1, rep(0,n),sigmasq * H)
tibble(y = y, x = x) %>% ggplot(aes(y=y, x=x)) +
  geom_line() + theme_bw() + ggtitle('Random realization of a GP with phi = 1 and sigmasq = 1') +
  geom_point(size = .5)
```



```
phi <- .1
sigmasq <- 5
H <- exp(-d/phi) + diag(eps, n)
H[1:3,1:3]
```

```
##           [,1]      [,2]      [,3]
## [1,] 1.0000000 0.9048374 0.6703200
## [2,] 0.9048374 1.0000000 0.9048374
## [3,] 0.6703200 0.9048374 1.0000000
```

```
y <- rmnorm(1, rep(0,n),sigmasq * H)
tibble(y = y, x = x) %>% ggplot(aes(y=y, x=x)) +
  geom_line() + theme_bw() + ggtitle('Random realization of a GP with phi = .1 and sigmasq = 5') +
  geom_point(size = .5)
```



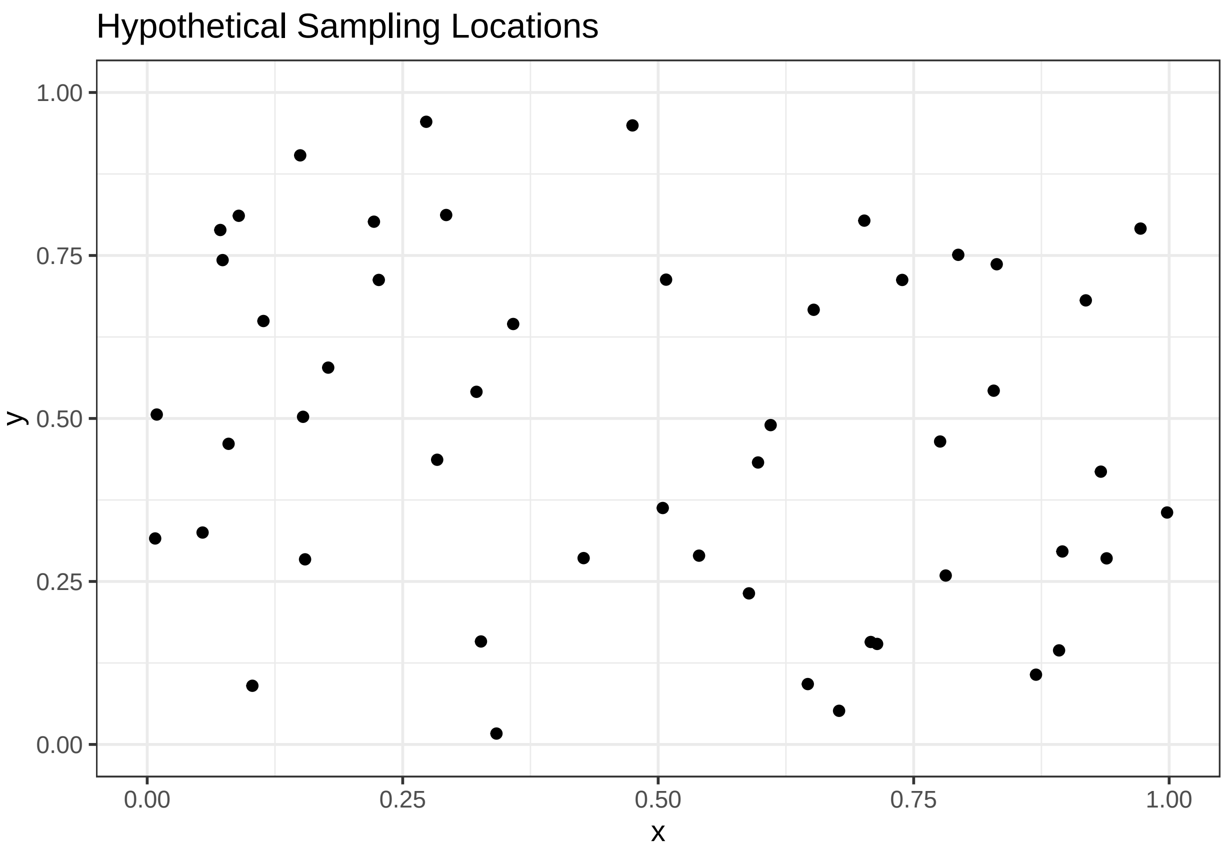
We will soon talk about a more broad set of correlation functions and another parameter that provides flexibility so that predictions do not have to directly through observed points.

Geostatistical Data

At last, we will look at simulated 2-d "spatial" data.

1. Sampling Locations

```
num.locations <- 50
coords <- data.frame(x = runif(num.locations), y = runif(num.locations))
coords %>% ggplot(aes(x=x,y=y)) + geom_point() +
  ggtitle('Hypothetical Sampling Locations') + xlim(0,1) +
  ylim(0,1) + theme_bw()
```



2. Calculate Distances

```
dist.mat <- plgp::distance(coords)
```

3. Define Covariance Function and Set Parameters

Use exponential covariance with no nugget:

```
sigma.sq <- 1
phi <- .1
Sigma <- sigma.sq * exp(- dist.mat/phi) + diag(eps, num.locations)
```

4. Sample realization of the process

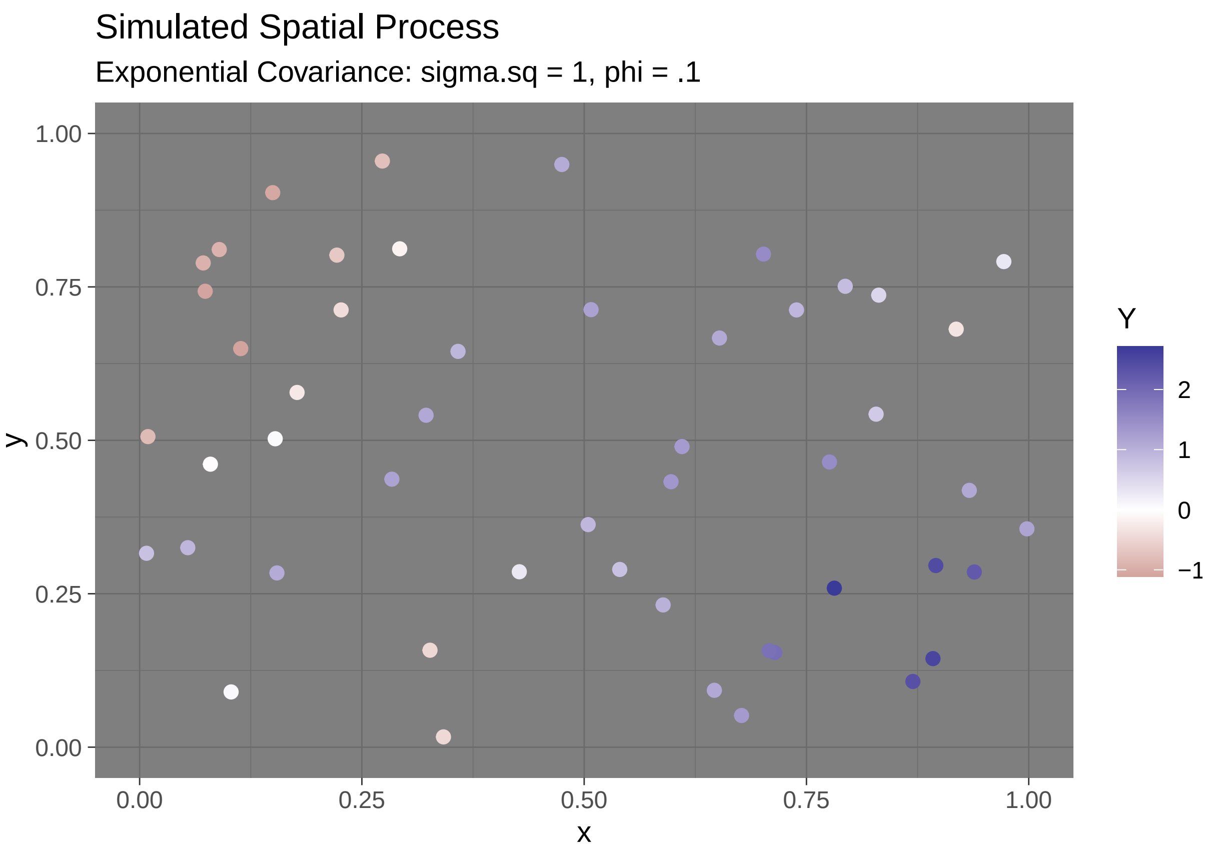
- This requires a distributional assumption, we will use the Gaussian distribution

```
Y <- rmnorm(n=1, mean = 0, varcov = Sigma)
```

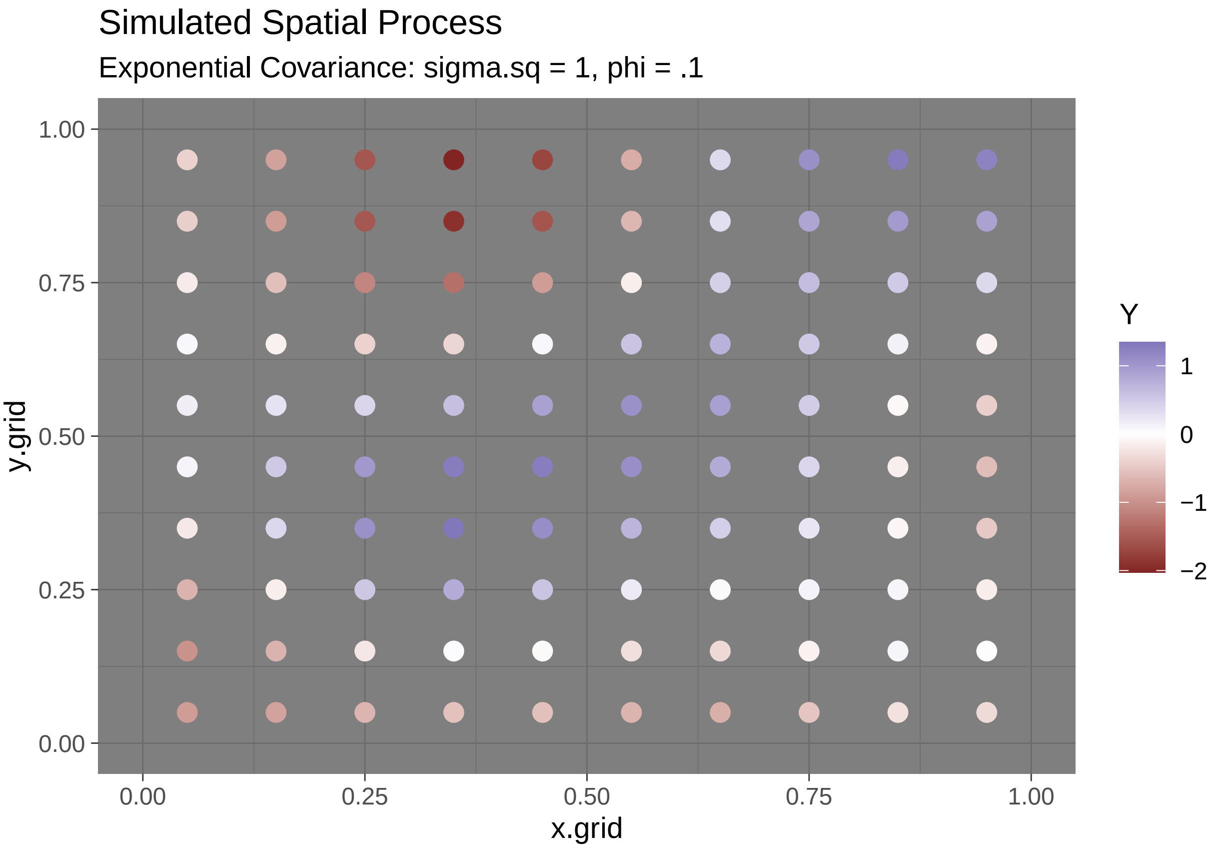
- What about the rest of the locations on the map?

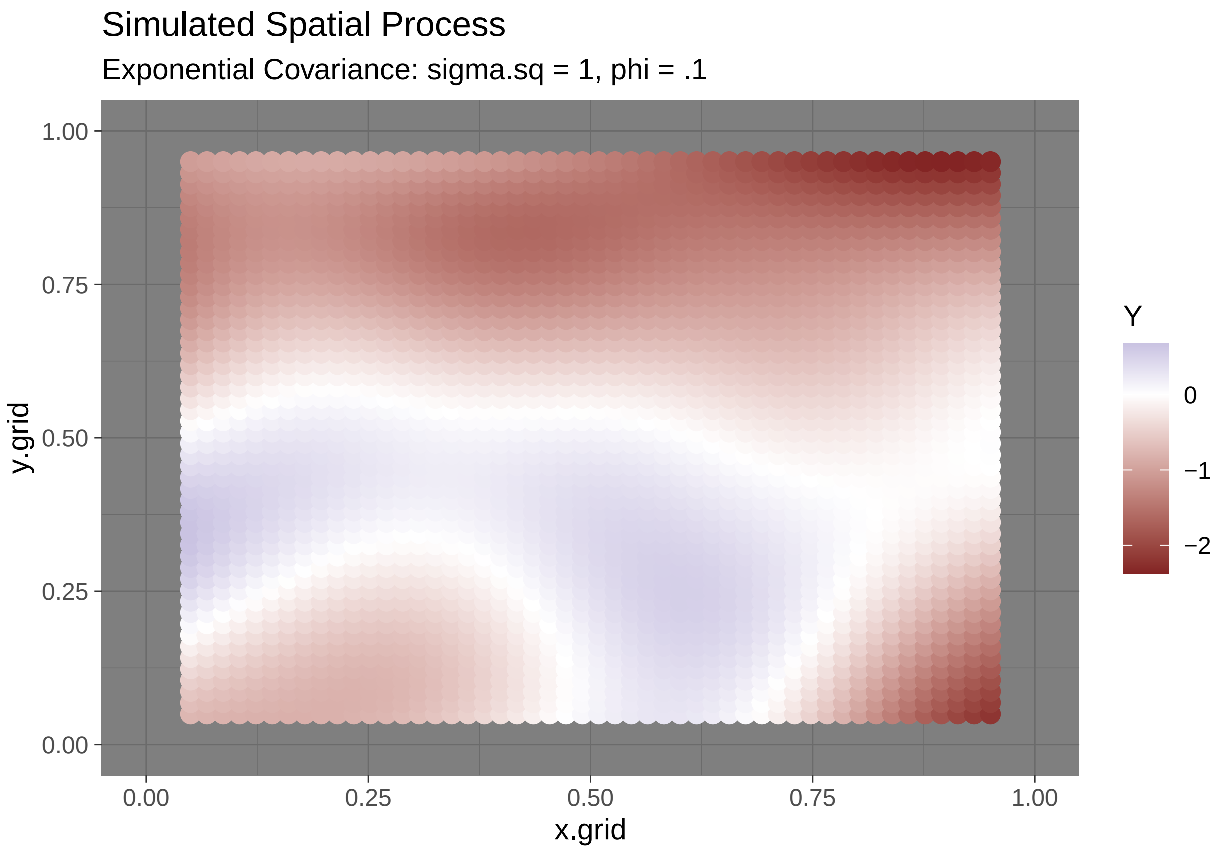
5. Vizualize Spatial Process

```
coords %>% mutate(Y = Y) %>% ggplot(aes(x=x,y=y)) + geom_point(aes(color=Y), size=2) +
  ggtitle(label = 'Simulated Spatial Process',
          subtitle = 'Exponential Covariance: sigma.sq = 1, phi = .1') +
  xlim(0,1) + ylim(0,1) +   scale_colour_gradient2() + theme_dark()
```



Now we can look at more sampling locations





How does the spatial process change with:

- another draw with same parameters?
- a different value of ϕ
- a different value of σ^2