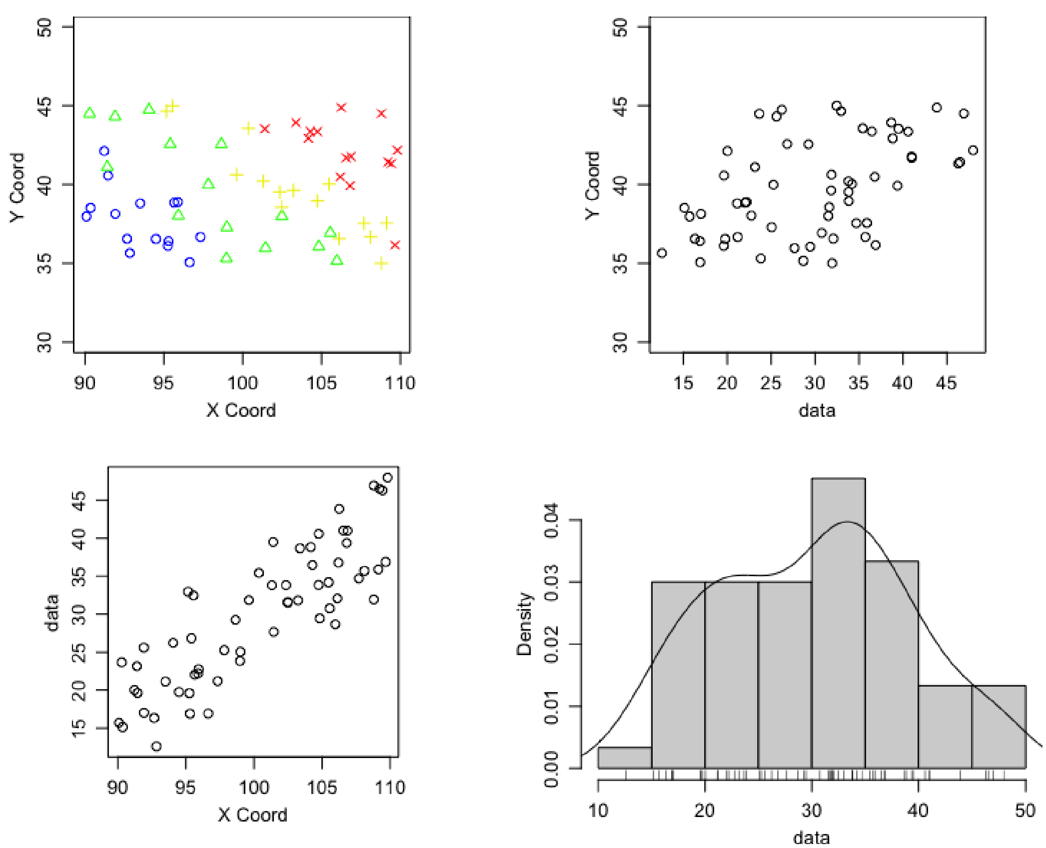
Exercise 1: Use the following r code to simulate and plot a data set consisting of $n = 60$ values from the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, where the ϵ's are independent $N(0, \sigma^2)$ random variables.

Code:

```
library(geoR)
set.seed(32123) # for reproducibility
n <- 60
lons <- runif( n, min=90, max=110 )
lats <- runif( n, min=35, max=45 )
y <- 30.0 + 1.3*(lons-100) + 1.5*(lats-40) + rnorm(n,mean=0,sd=1.6)
mydata <- as.geodata( cbind(lons,lats,y) )
plot(mydata)
```

a. Include the resulting plot in your write-up.

Solution:
Running the given R code produces the following plot,



b. What are the values of $\beta_0, \beta_1, \beta_2$, and σ^2 that are being used to simulate these data?

Solution:
The given R code defines the following function for Y (assuming $lons = X_1$ and $lat = X_2$),

$$
\begin{aligned}
Y &= 30 + 1.3(X_1 - 100) + 1.5(X_2 - 40) + \epsilon, \\
&= 30 + 1.3X_1 - 130 + 1.5X_2 - 60 + \epsilon, \\
&= -160 + 1.3X_1 + 1.5X_2 + \epsilon.
\end{aligned}
$$

Simplifying the function we get that $\beta_0 = -160, \beta_1 = 1.3, \beta_2 = 1.5$. The distribution of ϵ is described by the rnorm() function which gives $\sigma^2 = 1.6^2 = 2.56$.

c. Fit a linear model with independent errors using the R function lm, for example by typing,
Code:

```
myfit <- lm( y ~ lons + lats )
summary(myfit)
```

Solution:
Fitting the model we get the following summary report,
Code:

```
> LinearModel <- lm(y ~ lons + lats)
> summary(LinearModel)

Call:
lm(formula = y ~ lons + lats)

Residuals:
    Min      1Q  Median      3Q     Max
-3.6062 -1.2721 -0.2853  1.2536  3.0766

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -153.67616    4.47966  -34.30   <2e-16 ***
lons           1.25420    0.03443   36.43   <2e-16 ***
lats           1.45169    0.06947   20.90   <2e-16 ***
---

Residual standard error: 1.636 on 57 degrees of freedom
Multiple R-squared:  0.9681,    Adjusted R-squared:  0.967
F-statistic: 864.7 on 2 and 57 DF,  p-value: < 2.2e-16
```

d. State the estimated regression function and the estimate of σ^2 (Be sure to include the R output.)

Solution:
From the summary report, which is included above, we can see that the following estimated regression function is produced,

$$\begin{aligned}\hat{Y} &= \hat{\beta_0} + \hat{\beta_1}X_1 + \hat{\beta_2}X_2, \\ &= (-153.67616) + (1.25420)X_1 + (1.45169)X_2.\end{aligned}$$

e. Construct a 95 percent confidence for all β_i's.

Solution:
We can quickly compute the 95 percent confidence interval for all of our β_i regression paramaters using the R confint() function. Doing so we get the following,
Code:

```
> confint(LinearModel)
                 2.5 %      97.5 %
(Intercept) -162.646523 -144.705802
lons           1.185266    1.323137
lats           1.312573    1.590810
```

f. Make a table with the values of the β's, the estimates of the β's, and 95 percent CIs for the β's. Comment briefly for example, are the $\hat{\beta}$'s 'close to' the true β's? Do the β's lie in the respective 95 percent CIs.

Solution:

i	β_i	$\hat{\beta_i}$	Upper CI	Lower CI
0	-160	-153.67616	-162.646523	-144.705802
1	1.3	1.25420	1.185266	1.323137
2	1.5	1.45169	1.312573	1.590810

g. Perhaps $Y(s)$ is the weight of observed dog fur, in nanograms, at s. Interpret β_0 in terms of $E(Y)$. Explain why ie doesn't really make sense to interpret β_0.(I'm looking for two reasons. Three, if you include "its silly to talk about dog fur.")

Solution:
The term $\beta_0 = -160$ would be interpreted as, the mean weight of the observed dog fur is -160 at point $s = (0, 0)$. In terms of $E(Y)$ we would say $\beta_0 = -160 = E(Y((0, 0)))$. It doesn't really make sense to interpret this coefficient since the data is translation invariant, or in other words $s \in D$ is a generic data location in $D \subset \mathbb{R}^d$ dimensional space.

h. Interpret β_1 and β_2 in terms of $E(Y)$.(β_1 is the change in mean response if ...if what? -If you have question about what I'm looking for please ask.)

Solution:
For every one unit $s \in D$ increases in the longitudinal direction, the $E(Y)$ response changes by $\beta_1 = 1.25420$. Similarly for every one unit $s \in D$ increases in the latitudinal direction, the $E(Y)$ response changes by $\beta_1 = 1.45169$. These coefficient make sense interpret since the describe a relationship between points in D.

Exercise 2: Let,

$$\Sigma = \begin{bmatrix} 5 & -3 \\ -3 & 1 \end{bmatrix}$$

a. Show that Σ cannot be a valid variance-covariance matrix for $(Y_1, Y_2)^T$, by assuming it is and finding the correlation of Y_1 and Y_2.

Solution:
Assuming Σ is a valid variance-covariance matrix we can compute the correlation of Y_1 and Y_2 by the following,

$$\begin{aligned} Cor(Y_1, Y_2) &= \frac{Cov(Y_1, Y_2)}{\sqrt{Var(Y_1)Var(Y_2)}}, \\ &= \frac{-3}{\sqrt{(5)(1)}}, \\ &< -1. \end{aligned}$$

This matrix produces a correlation between Y_1 and Y_2 which is outside of the possible domain $Cor(Y_1, Y_2) \notin [-1, 1]$

b. Show that Σ cannot be a valid variance-covariance matrix by assuming it is and finding $Var(Y_1 + 2Y_2)$.

Solution:
Simplifying the variance expression we get the following,

$$\begin{aligned} Var(Y_1 + 2Y_2) &= Var(Y_1) + 4Var(Y_2) + 2(1)(2)Cov(Y_1, Y_2), \\ &= 5 + 4(1) + 4(-3), \\ &= 5 + 4 + 4(-3), \\ &= -3. \end{aligned}$$

A linear combination of random variables should not be able to produce a negative variance, which suggests that Σ is not a valid variance-covariance matrix.

Exercise 3: Consider the covariogram,

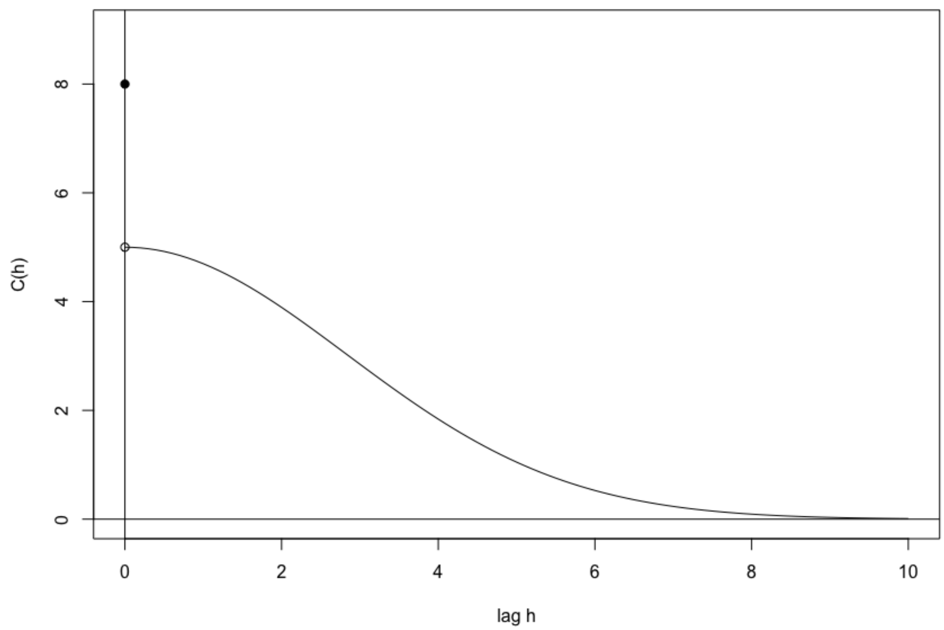
$$C(h) = \begin{cases} 8 & h = 0 \\ 5e^{-h^2/16} & h > 0 \end{cases}$$

a. Sketch $C(h)$, for example, using R code:

```
hh <- seq( 0,10, length=200 )
        curve( 5*exp(-x^2/16), xlab="lag h", ylab="C(h)",
              from=0.0, to=10)
        abline( v = 0 ) # include y-axis on the plot
        abline( h = 0 ) # include x-axis on the plot
```

Solution:
Running the tweaked version of the given R code we produce the following plot,
Code:

```
hh <- seq( 0,10, length=200 )
curve( 5*exp(-x^2/16), xlab="lag h", ylab="C(h)",
       from=0.0, to=10, ylim = c(0, 9))
abline( v = 0 ) # include y-axis on the plot
abline( h = 0 ) # include x-axis on the plot
points(0, 5)
points(0, 8, pch = 16)
```



b. Calculate the corresponding semivariogram. (Be sure to show your work.) You need to be careful about your calculations at $h = 0$.

Solution:
Recall the relationship between a covariogram and the corresponding semivariogram,

$$\gamma(h) = C(0) - C(H).$$

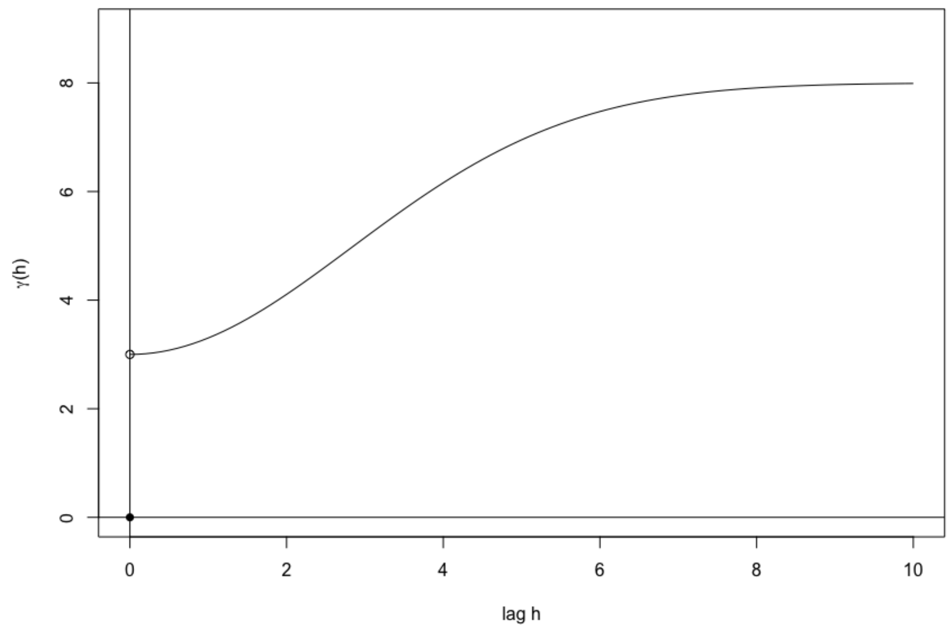
Applying this relationship to both parts of the piecewise function $C(h)$ we get the following,

$$\begin{aligned}\gamma(h) &= \begin{cases} C(0) - C(0) & h = 0 \\ C(0) - C(h) & h > 0 \end{cases}, \\ &= \begin{cases} 0 & h = 0 \\ 8 - 5e^{-h^2/16} & h > 0 \end{cases}\end{aligned}$$

Plotting the resulting semivariogram, we get the following,
Code:

```
hh <- seq( 0,13, length=200 )
curve( 8 - 5*exp(-x^2/16), xlab="lag h",
        ylab=expression(paste(gamma, "(h)")),
        from=0.0, to=10, ylim = c(0, 9))
```

```
abline( v = 0 ) # include y-axis on the plot
abline( h = 0 ) # include x-axis on the plot
points(0, 3)
points(0, 0, pch = 16)
```



c. What kind of (semi)variogram is this, and what are its parameters?

Solution:
This is an example of an exponential semivariogram. It's parameters are $\tau^2 = 3, \sigma^2 = 5$ and $\phi = 16$.

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 8 - 5e^{-h^2/16} & h > 0 \end{cases},$$
$$= \begin{cases} 0 & h = 0 \\ 3 + 5(1 - e^{-h^2/16}) & h > 0 \end{cases},$$
$$= \begin{cases} 0 & h = 0 \\ \tau^2 + \sigma^2(1 - e^{-h^2/\phi}) & h > 0 \end{cases}$$

d. What is the nugget? the partial sill? the sill? the range (or effective range)? $Var(Y(s))$?

Solution:
From the figure we can see that the nugget for the semivariogram is 3. The partial sill is 5. The sill is 8. the range looks to go from $[0, 10]$ maybe even $[0, 8]$.

Exercise 4: Consider the semivariogram,

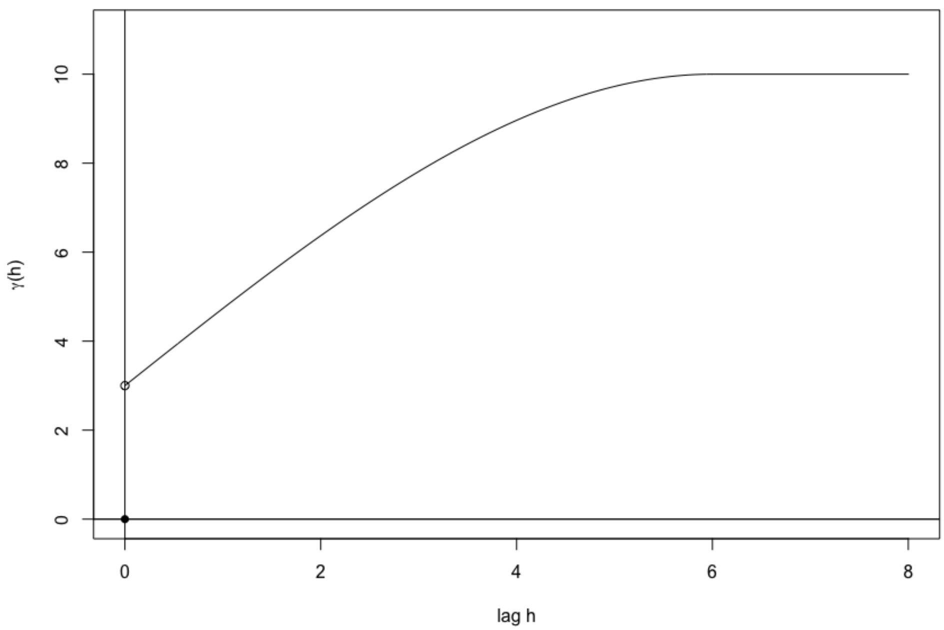
$$\gamma(h) = \begin{cases} 0 & h = 0 \\ 3 + 7(1.5(h/6) - .5(h/6)^3) & 0 < h < 6 \\ 10 & h \geq 6 \end{cases}$$

a. Use R to sketch $\gamma(h)$ for a suitable range of h.

Solution:

Code:

```
hh <- seq( 0,13, length=200 )
curve( 3 + 7*(1.5*(x/6) - .5*(x/6)^3), xlab="lag h",
        ylab=expression(paste(gamma, "(h)")),
        from=0.0, to=6, ylim = c(0, 11), xlim = c(0,8))
segments(6, 10, x1 = 8, y1 = 10)
abline( v = 0 ) # include y-axis on the plot
abline( h = 0 ) # include x-axis on the plot
points(0, 3)
points(0, 0, pch = 16)
```



b. Find the corresponding covariogram, $C(h)$.

Solution:
Recall the relationship between a covariogram and a semivariogram,

$$C(h) = \lim_{R\to\infty} \gamma(R) - \gamma(h).$$

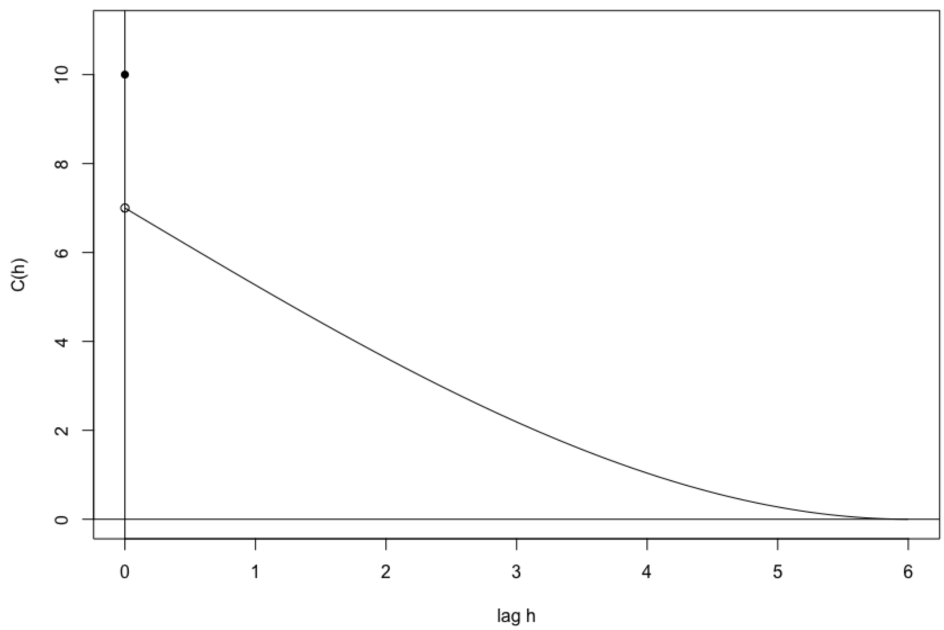
Applying this relationship to each part of the piecewise definition of $\gamma(h)$ we get,

$$C(h) = \begin{cases} 10 - 0 & h = 0, \\ 10 - (3 + 7(1.5(h/6) - .5(h/6)^3)) & 0 < h < 6, , \\ 10 - 10 & h \geq 6. \end{cases}$$
$$= \begin{cases} 10 & h = 0, \\ 10 - 3 - 7(1.5(h/6) - .5(h/6)^3)) & 0 < h < 6, \\ 0 & h \geq 6. \end{cases}$$
$$= \begin{cases} 10 & h = 0, \\ 7 - 7(\frac{3}{2}(h/6) - \frac{1}{2}(h/6)^3)) & 0 < h < 6, \\ 0 & h \geq 6. \end{cases}$$

Plotting in R we get the following,
Code:

```
hh <- seq( 0,13, length=200 )
curve( 7-7*(1.5*(x/6) - .5*(x/6)^3),
          xlab="lag h", ylab="C(h)",
          from=0.0, to=6, ylim = c(0, 11))

abline( v = 0 ) # include y-axis on the plot
abline( h = 0 ) # include x-axis on the plot
points(0, 7)
points(0, 10, pch = 16)
```



c. What kind of (semi)variogram is this, and what are its parameters?

Solution:
This is an example of a spherical semivariogram with parameters $\tau^2 = 3$, $\sigma^2 = 6$, and $\phi = 6$.

d. What is the nugget? the partial sill? the sill? the range? $Var(Y(s))$?

Solution:
From the figure we can see that the nugget for the semivariogram is 3. The partial sill is 7. The sill is 10. The range is from $[0, 6]$.

Exercise 5: Let X be a discrete random variable such that,

x	-2	-1	1	2
$P(X = x)$	.4	.1	.1	.4

Define $Y = X^2$.

a. Show that $cov(X, Y) = 0$ (To do this, you wil need to calculate $E(X)$, $E(Y)$, and $E(XY)$, and use the formula for covariance frm page 25 of the lecture notes.)

Solution:
Recall the formula for the covariance,

$$cov(X, Y) = E(XY) - \mu_x \mu_y = E(X^3) - \mu_x \mu_y.$$

Computing each term we get,

$$E(X^3) = (.4)(-2)^3 + (.1)(-1)^3 + (.1)1^3 + (.4)2^3 = 0.$$

$$\mu_x = \frac{(-2) + (-1) + (1) + (2)}{4} = 0.$$

$$\mu_y = \frac{(-2)^2 + (-1)^2 + (1)^2 + (2)^2}{4} = 2.4.$$

$$cov(X, Y) = E(X^3) - \mu_x \mu_y = 0 - 0(2.4) = 0.$$

b. Show that X and Y are dependent. (E.g., calculate $P(X = 1, Y = 4)$ and $P(X = 1)P(Y = 4)$; if X and Y are independent, these two values- among others- must be equal.)

Solution:
Following the example described in the hint consider the following,

$$P(X = 1, Y = 4) = 0.$$

Since $Y = X^2$ we know that when $X = 1$ it follows that $Y = 1$ and similarly if $y = 4$ it follows that $X = 2, -2$. In anycase the joint probability $P(X = 1, Y = 4)$ must be zero. Furthermore computing the product we get,

$$P(X = 1)P(Y = 4) = (.1)(.8) = .08 \neq 0.$$

Thus X and Y are dependent.