Agriculture

Brian Colgrove

4/8/2021

```r
#packages
library(gstat)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.4
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```r
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(multcomp)
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: survival
```

```
## Loading required package: TH.data
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser
```

```
library(nlme)
```

```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
library(DataExplorer)
library(sf)
```

```
## Linking to GEOS 3.8.0, GDAL 3.0.4, PROJ 6.3.1
```

```
library(maps)
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
library(mapproj)
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
source("stdres.R")
source("predictgls.R")


#read in the data
agriculture <- read.table("https://mheaton.byu.edu/docs/files/Stat469/Topics/3%20-%20SpatialCorr
elation/1%20-%20PointReference/HWCaseStudy/Data/WaterHoldingCapacity.txt", header = TRUE, sep =
" ")

#summarize the data
head(agriculture)
```

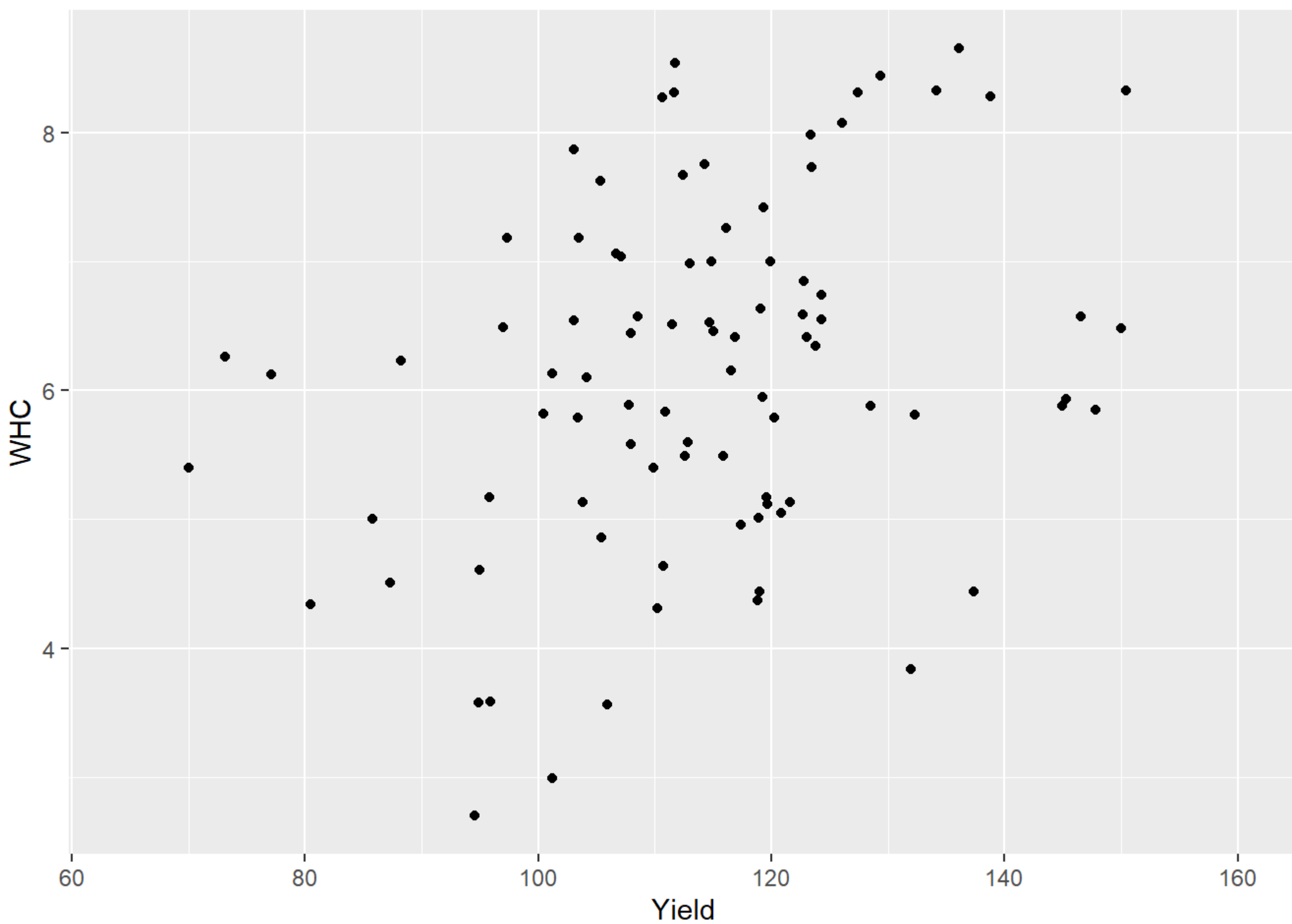
```
##         Lon      Lat  Yield    EC  WHC
## 1 -111.7851 42.60525 121.63 12.40 5.13
## 2 -111.7844 42.60525 109.26 20.90   NA
## 3 -111.7838 42.60525 122.54 23.95   NA
## 4 -111.7831 42.60525 124.81 24.12   NA
## 5 -111.7865 42.60555 142.45 15.10   NA
## 6 -111.7858 42.60555 143.95 17.76   NA
```

```
summary(agriculture)
```

```
##       Lon              Lat             Yield              EC
##  Min.   :-111.8   Min.   :42.61   Min.   : 64.52   Min.   :12.40
##  1st Qu.:-111.8   1st Qu.:42.61   1st Qu.:103.06   1st Qu.:21.25
##  Median :-111.8   Median :42.61   Median :113.01   Median :23.38
##  Mean   :-111.8   Mean   :42.61   Mean   :113.10   Mean   :23.27
##  3rd Qu.:-111.8   3rd Qu.:42.61   3rd Qu.:124.01   3rd Qu.:25.81
##  Max.   :-111.8   Max.   :42.62   Max.   :160.51   Max.   :30.67
##
##       WHC
##  Min.   :2.700
##  1st Qu.:5.170
##  Median :6.150
##  Mean   :6.138
##  3rd Qu.:7.000
##  Max.   :8.650
##  NA's   :439
```

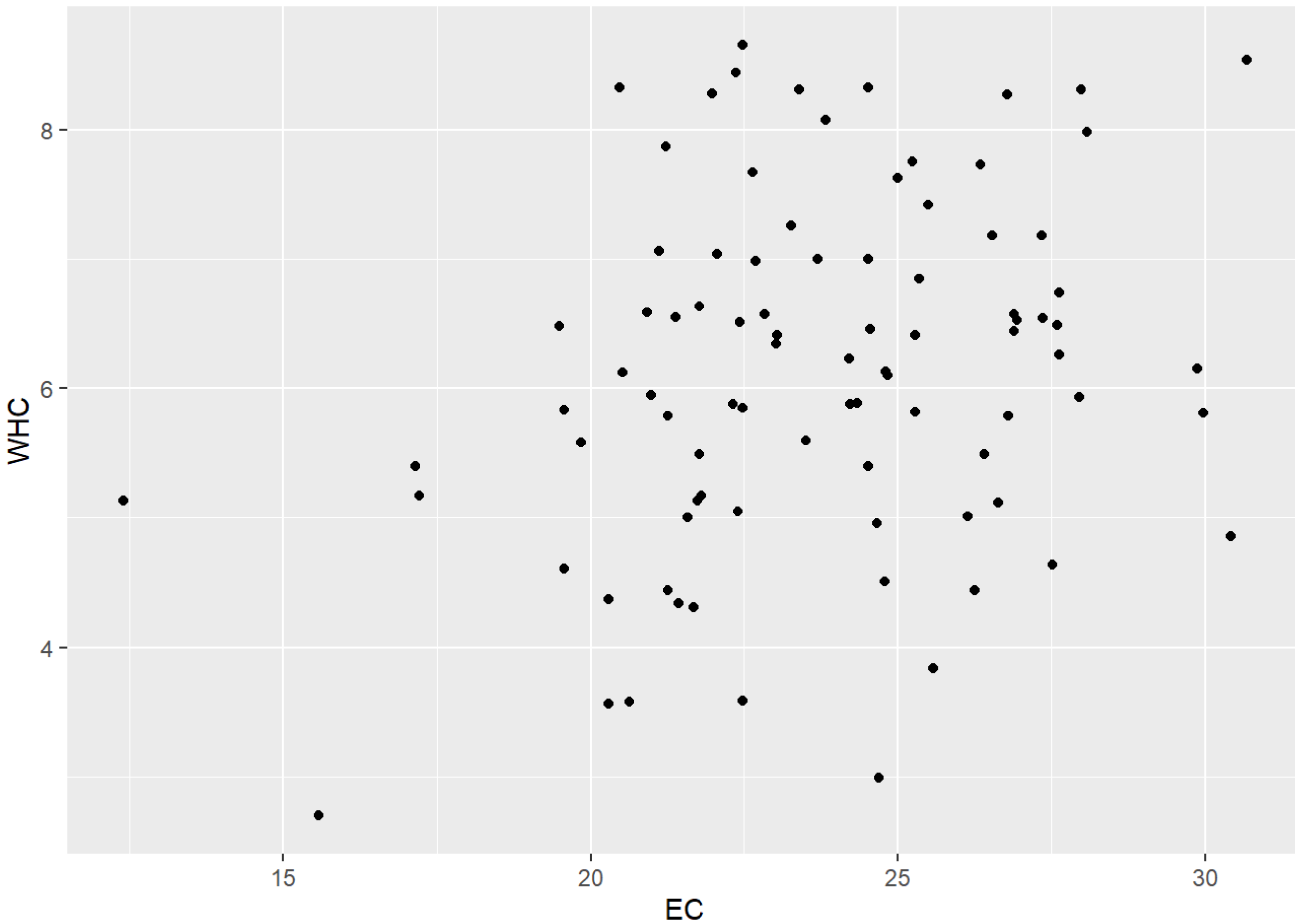
```
#scatter plot Yield vs WHC
ggplot(agriculture, aes(x = Yield, y = WHC)) +
  geom_point()
```

```
## Warning: Removed 439 rows containing missing values (geom_point).
```



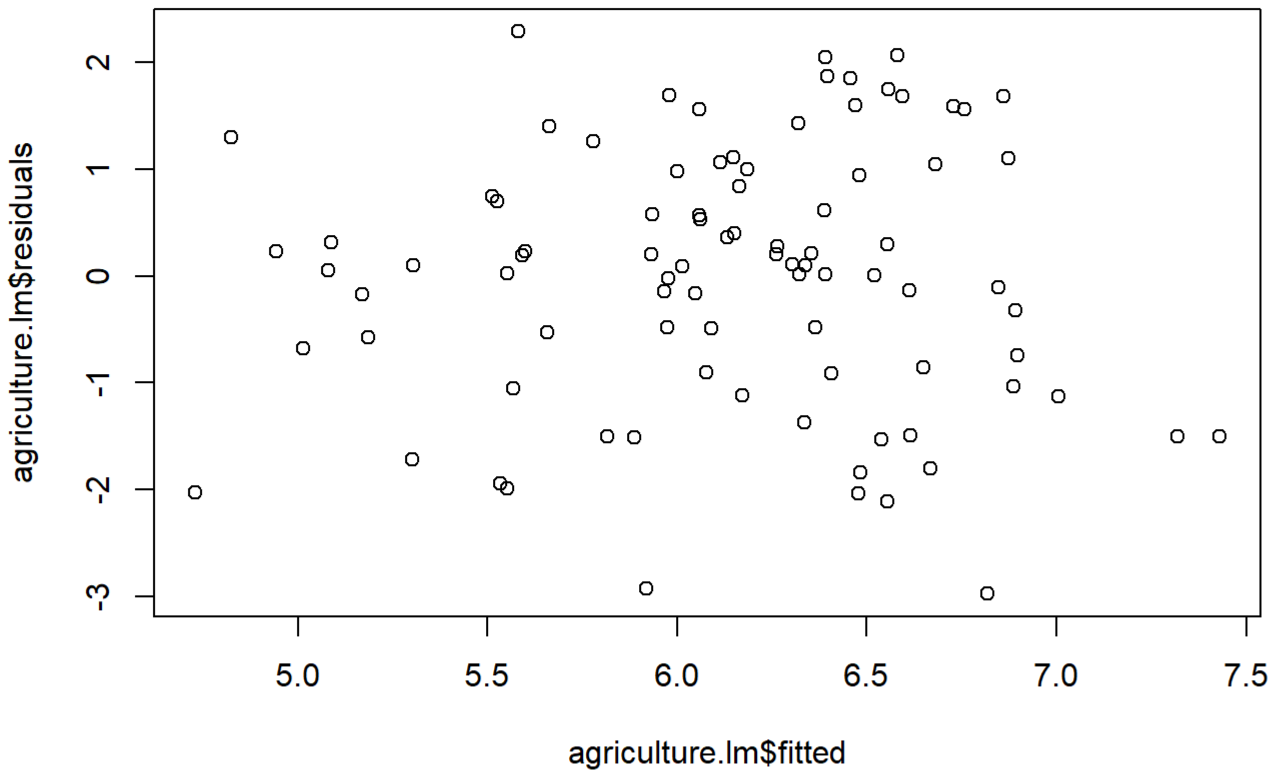
```
#scatter plot EC vs WHC
ggplot(agriculture, aes(x = EC, y = WHC)) +
  geom_point()
```

```
## Warning: Removed 439 rows containing missing values (geom_point).
```

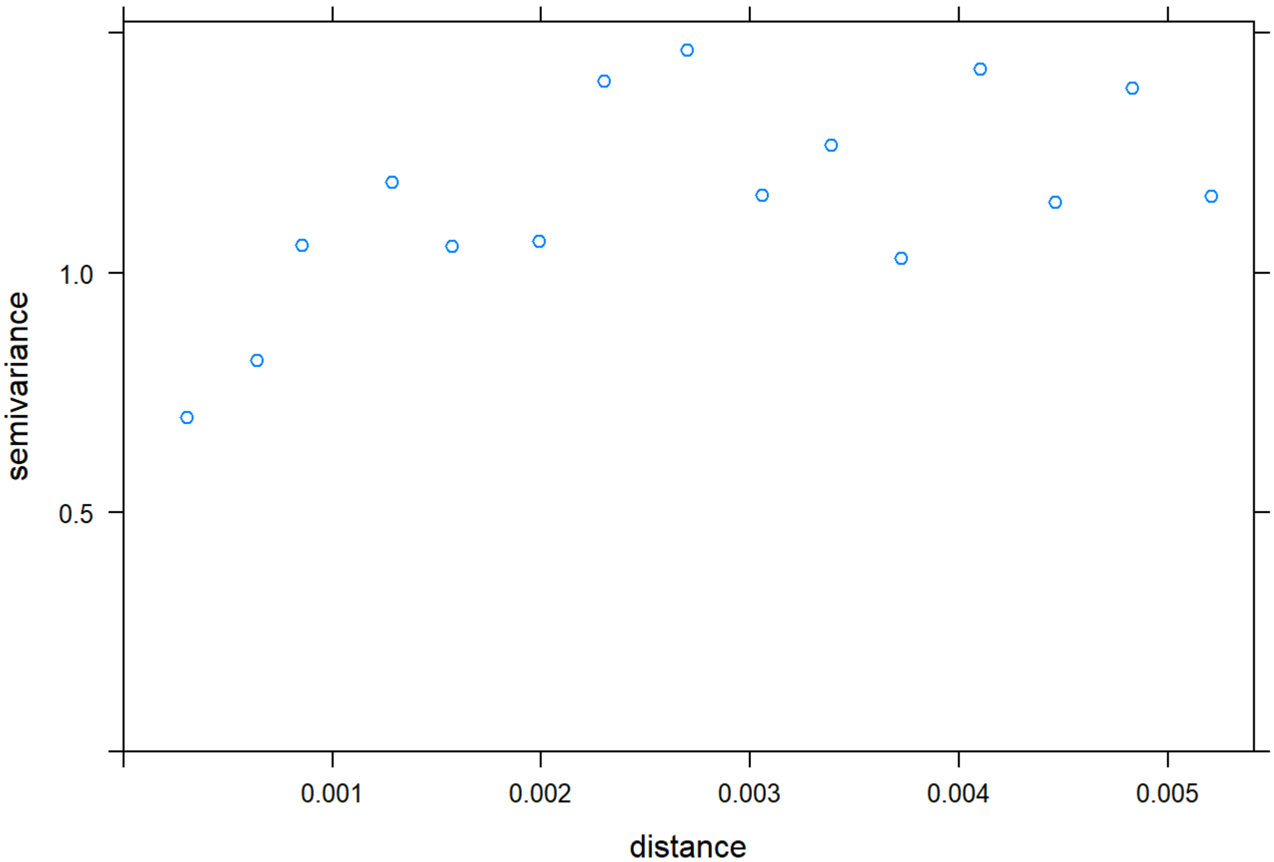


There seems to be a positive linear relationship between Yield and WHC and also EC and WHC.

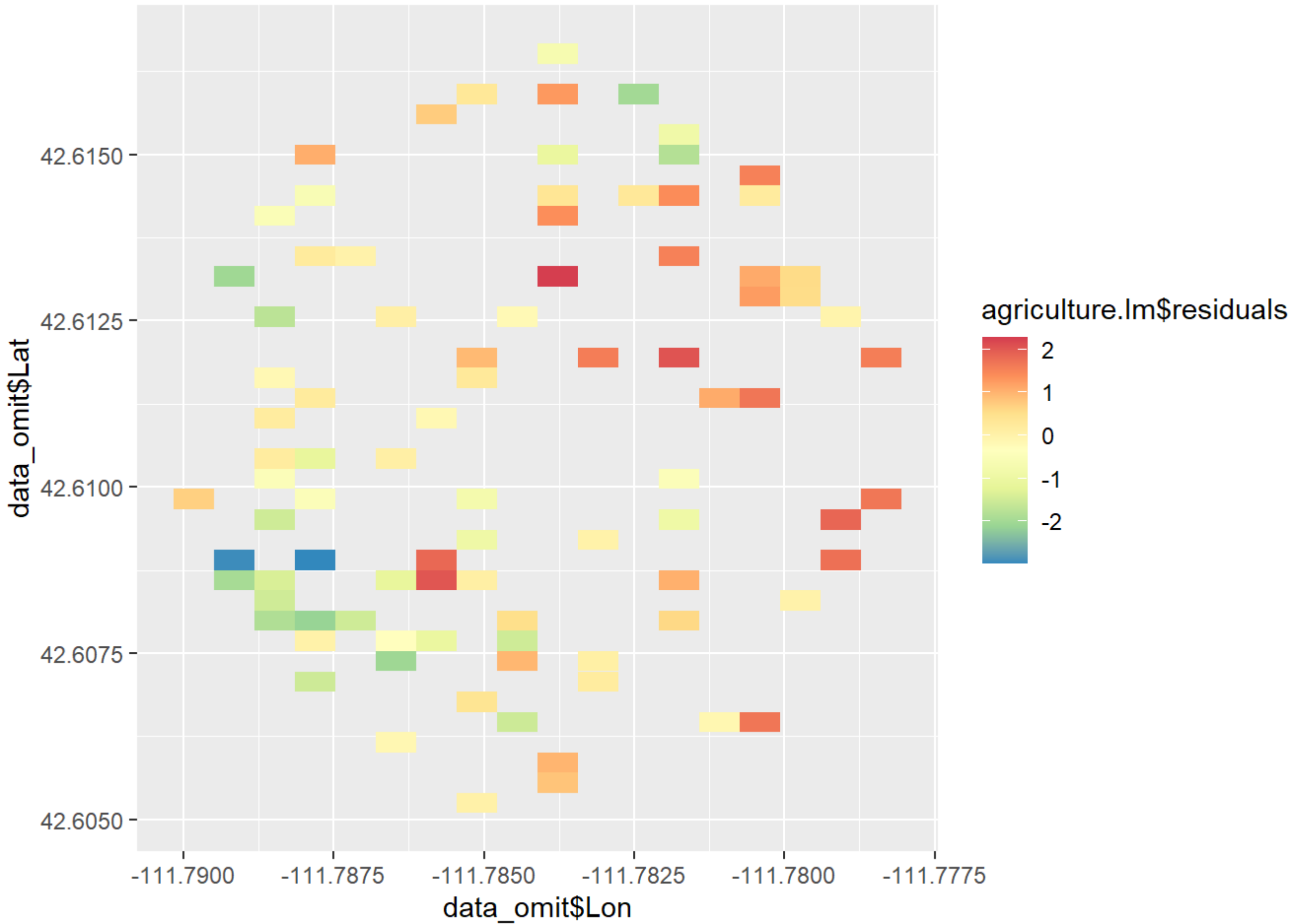
```
#lm model
agriculture.lm <- lm(WHC~EC + Yield, data = agriculture)
plot(agriculture.lm$fitted, agriculture.lm$residuals)
```




```
#variogram
data_omit <- na.omit(agriculture)
myVariogram <- variogram(object=WHC~EC + Yield, locations=~Lon+Lat, data=data_omit)
plot(myVariogram)
```




```
#residual plot
ggplot() + geom_tile(aes(x=data_omit$Lon, y=data_omit$Lat, fill=agriculture.lm$residuals)) + sca
le_fill_distiller(palette="Spectral")
```



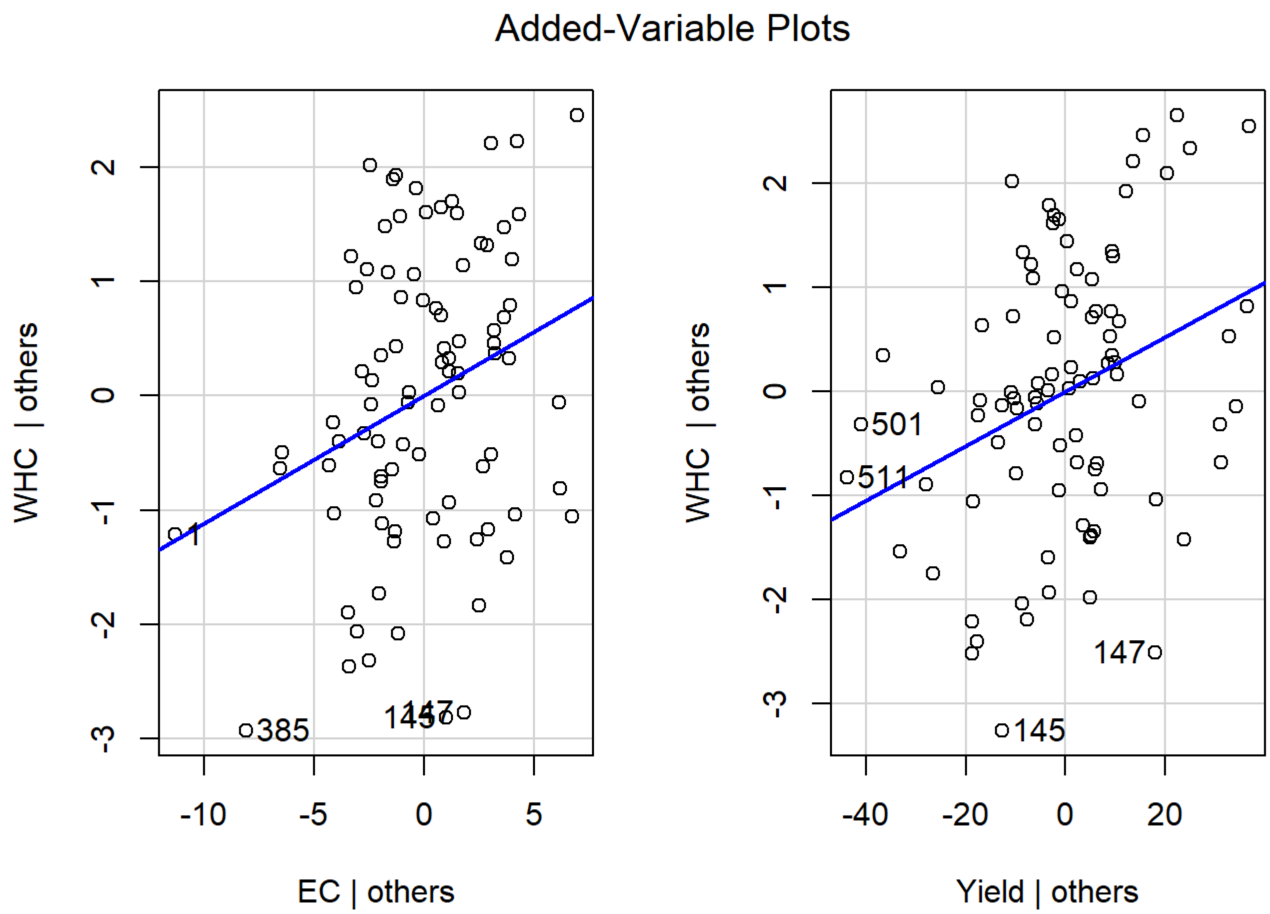
There seems to be spatial correlation for both residual and variogram graphs. The residual one gets higher values the further North East you go. The variogram shows smaller variance at smaller distance.

```
#gls model
agriculture.gls <- gls(model=WHC~EC+Yield, data=data_omit, correlation=corExp(form=~Lon+Lat, nug
get=TRUE), method="ML", na.action = na.omit)

agriculture.gls2 <- gls(WHC~Yield+EC, data=data_omit, correlation=corSpher(form=~Lon+Lat, nugget
=TRUE), method="ML")

agriculture.gls3 <- gls(WHC~Yield+EC, data=data_omit, correlation=corGaus(form=~Lon+Lat, nugget=
TRUE), method="ML")

c(AIC(agriculture.gls), AIC(agriculture.gls2), AIC(agriculture.gls3))
```

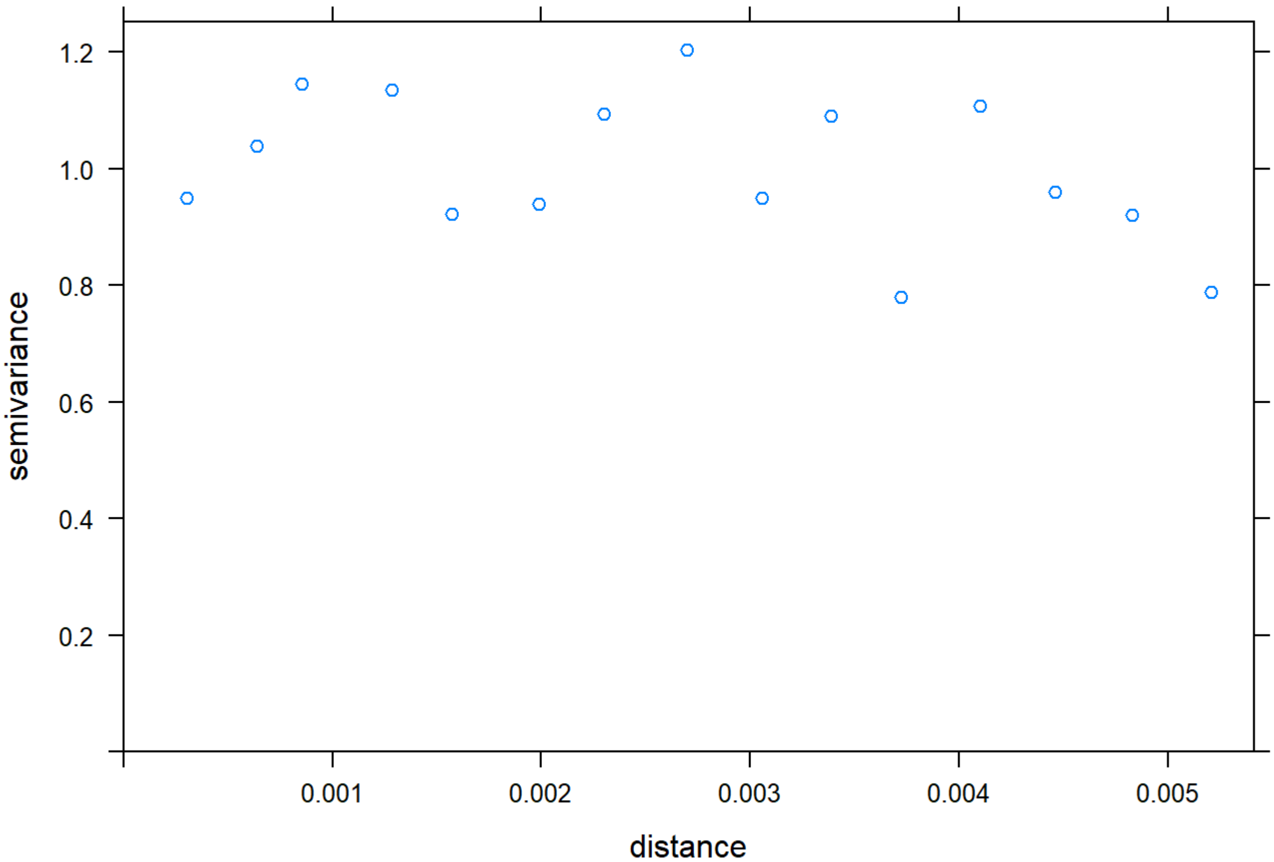
```
## [1] 272.3653 272.9623 273.4355
```

The exponential correlation gls model is just barely better than the other two according to its AIC values.

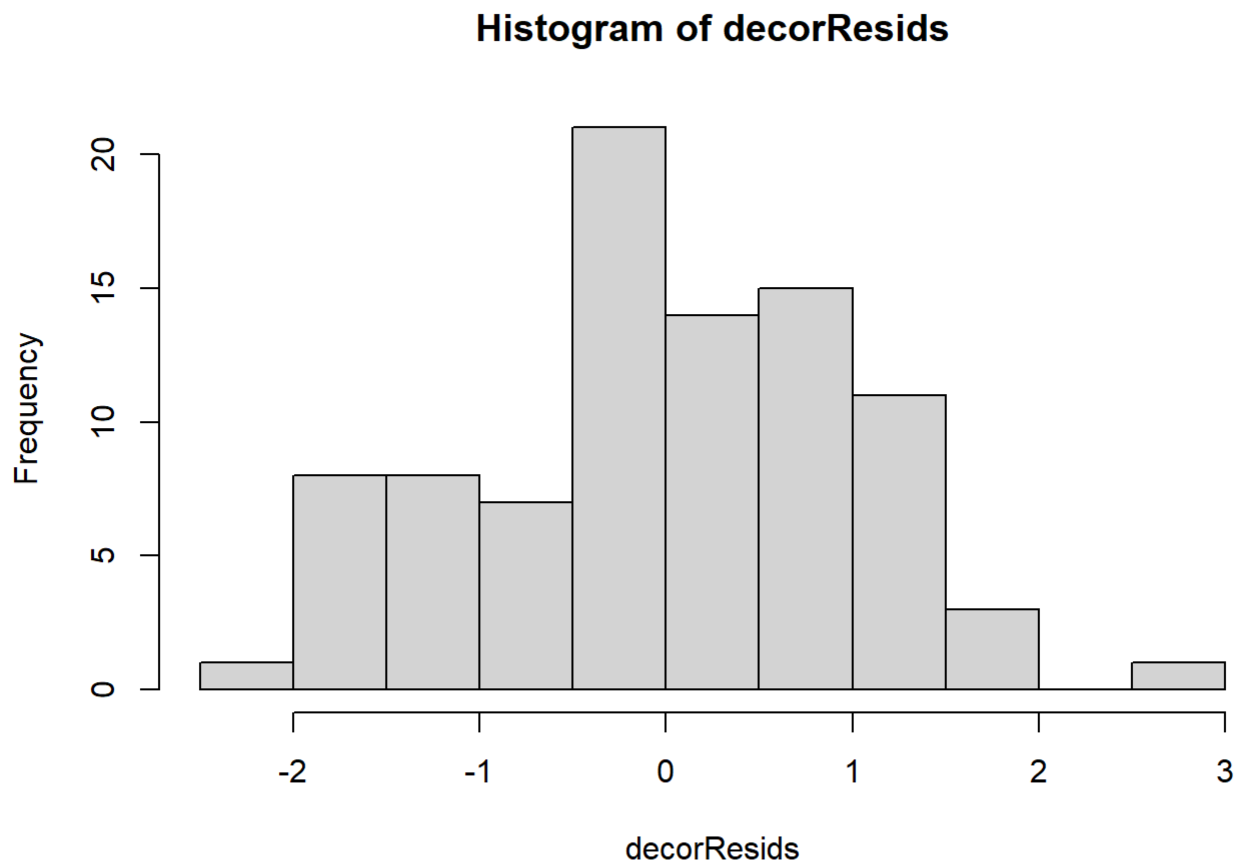
```
#linearity
avPlots(agriculture.lm)
```



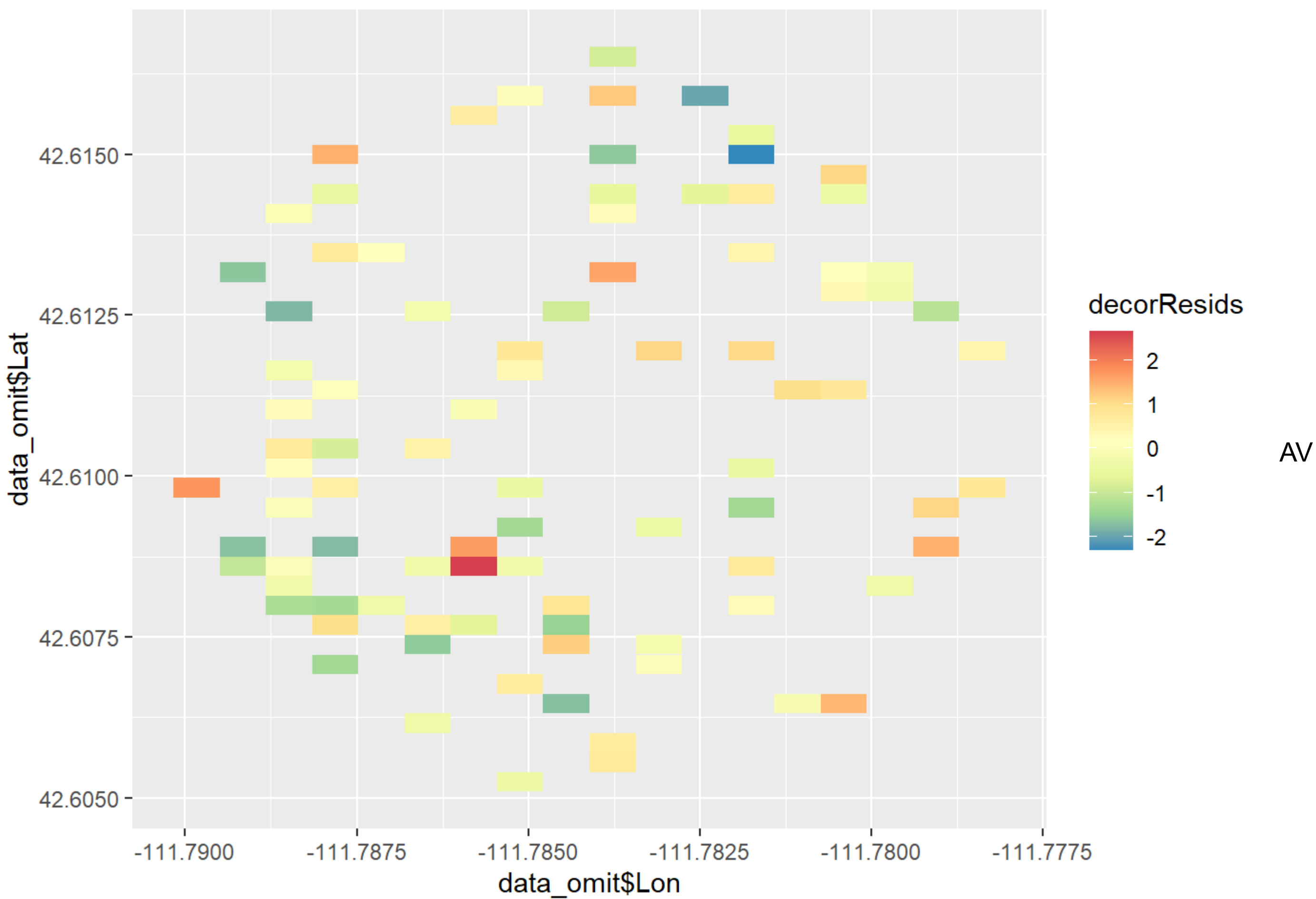

```
#decorrelated residuals
decorResids <- stdres.gls(agriculture.gls)

#variogram of decorrelated residuals
residDF <- data.frame(Lon=data_omit$Lon, Lat=data_omit$Lat, decorrResid=decorResids)
residVariogram <- variogram(object=decorrResid~1, locations=~Lon+Lat, data=residDF)
plot(residVariogram)
```




```
#normality
hist(decorResids)
```




```
#residual plot
ggplot() + geom_tile(aes(x=data_omit$Lon, y=data_omit$Lat, fill=decorResids)) + scale_fill_disti
ller(palette="Spectral")
```



plots are linear, the histogram is normal, the variogram shows a more constant variance along the distances, the residual plot shows pretty good values across the map. No patterns. All assumptions are met.

```
#table
summary(agriculture.gls)$tTable
```

```
##                  Value   Std.Error  t-value     p-value
## (Intercept) 1.56633579 1.395088389 1.122750 0.264668291
## EC          0.07399449 0.036260298 2.040648 0.044352708
## Yield       0.02578200 0.009460142 2.725329 0.007781781
```

```
#confidence interval
intervals(agriculture.gls, which="coef")
```

```
## Approximate 95% confidence intervals
##
##  Coefficients:
##                    lower       est.      upper
## (Intercept) -1.207008138 1.56633579 4.33967972
## EC           0.001911404 0.07399449 0.14607758
## Yield        0.006975860 0.02578200 0.04458814
## attr(,"label")
## [1] "Coefficients:"
```

```
#na predictions
vec <- which(is.na(agriculture$WHC))
na_data <- agriculture[vec,c(1:4) ]

#na.preds <- predict(temp.lm, newdata=na_data)
na.preds <- predictgls(agriculture.gls, newdframe = na_data)
```

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```
#map predictions
colnames(na.preds)[5] <- "WHC"
full <- rbind(na.preds[1:5], data_omit )
mymap <- get_stamenmap(bbox = c(left = min(full$Lon), bottom = min(full$Lat), right =
  max(full$Lon), top =  max(full$Lat)), zoom=7, maptype="watercolor")
```

```
## Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.
```

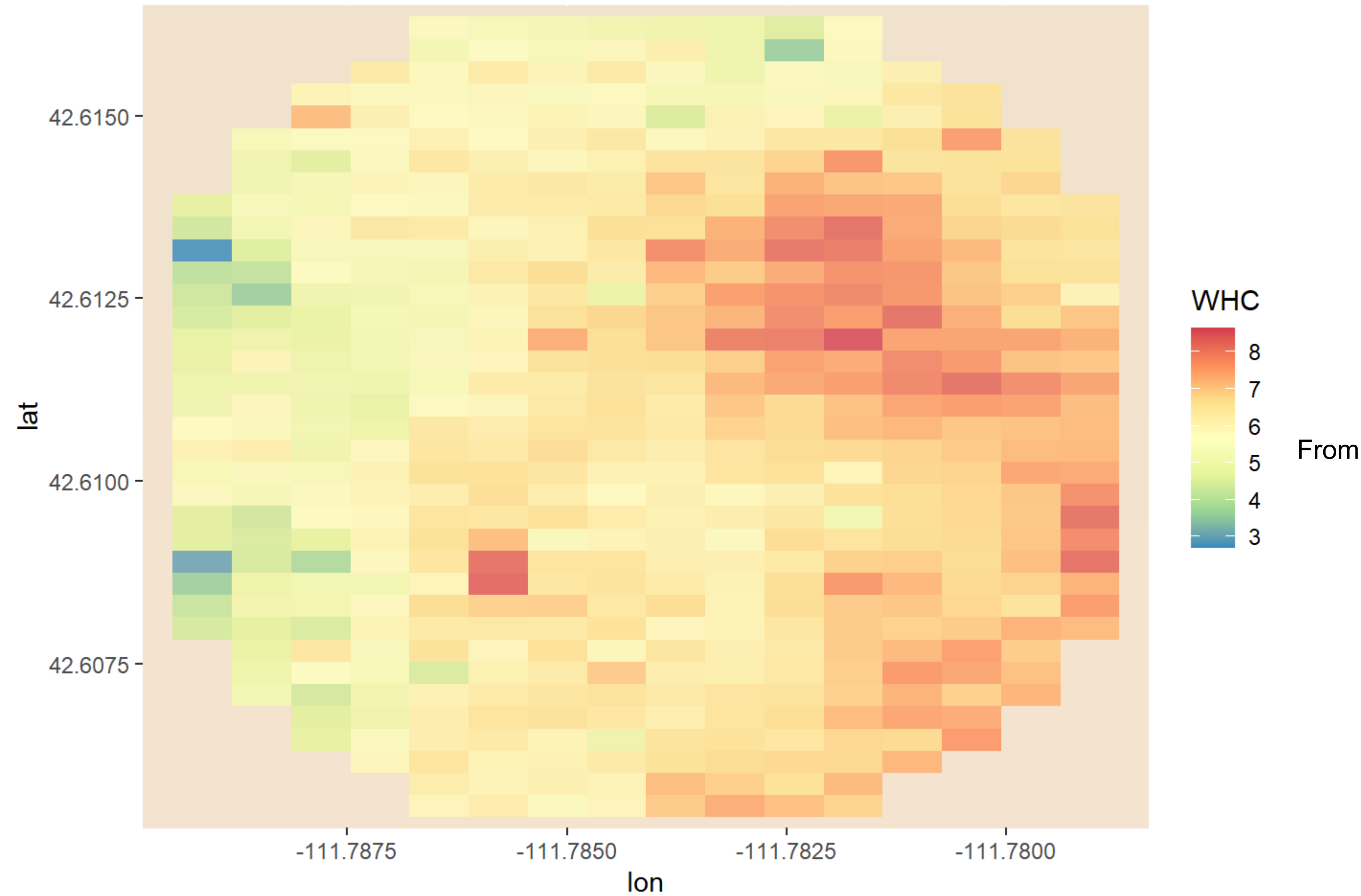
```
ggmap(mymap)+geom_raster(data=full, aes(x=Lon, y=Lat, fill=WHC),  alpha=0.8)+
  geom_tile()+ scale_fill_distiller(palette="Spectral", na.value=NA) + coord_cartesian()
```

```
## Coordinate system already present. Adding new coordinate system, which will replace the exist
ing one.
```

```
## Warning: Raster pixels are placed at uneven horizontal intervals and will be
## shifted. Consider using geom_tile() instead.
```

```
## Warning: Raster pixels are placed at uneven vertical intervals and will be
## shifted. Consider using geom_tile() instead.
```

```
## Warning: Removed 24 rows containing missing values (geom_raster).
```



our hypothesis test it does seem like that higher yield areas increase with higher WHC since our p-value is 0.00778. We are 95% confident that WHC increases per 1 unit of Yield is between 0.006, 0.045.