South Dakota State University  
STAT600 Summer 2020  
Midterm Project

Colin Nagengast

July 12, 2020

# Project Purpose

The purpose of this project is to analyze a farm field and determine how many grid cells the field should be split into in order to detect a statistically significant difference between two different groups or treatments applied to the farm field in future years. In doing so I need to keep in mind the number of measurements that can be taken within a grid cell where a grid cell that is too small will result in too few yield measurements available and will also mean more work in applying the treatments to the field. Grid cells that are too large will result in not being able to distinguish if a difference found is due to differences in treatments or just the location within the field. In the end there will be two treatment groups, each with a quantity of (# Grid Cells)/2 sample units.

# Methodology

I was given three datasets for yield results from a farm field for the seasons of 2015, 2016 and 2017. The field measures 400 meters (north to south) by 600 meters (east to west). Each dataset consists of thousands of records that include the yield counts taken by farm equipment and the corresponding coordinates of latitude and longitude of where in the field the yield counts were taken.

My plan is to simulate applying two treatments, A and B, to the farm field. I will do so by splitting the field up into grids ranging in size from four (two across by two top to bottom) to 1,600 (400 across by 400 top to bottom) and doing this 18 times in total. Once the field is split up, I will start from the bottom and take every odd numbered row and ‘assign’ it to Treatment A and every even numbered row will be ‘assigned’ to Treatment B. Once I have all rows assigned, I will then give numbers to all cells based upon their row and column number. Summary statistics will then be computed for each treatment.

As an example, for the case of splitting the field into a 10x10 grid for 100 cells, there will be 50 Treatment A cells and 50 Treatment B cells. I will create and assign the following:

‘M1’ to be the grand mean of the 50 Treatment A Yield amount cell means

‘M2’ to be the grand mean of the 50 Treatment B Yield amount cell means

‘S1’ will be the standard deviation of the 50 cell means for Treatment A

‘S2’ will be the standard deviation of the 50 cell means for Treatment B

‘N1’ will be equal to 50, the number of Treatment A cells

‘N2’ will be equal to 50 for the corresponding Treatment B cells

The result will be a data frame that measures 1x6 for each iteration of splitting up the farm field. And I will have 18 of these iterations in total for a final 18 row data frame of summary data for each season. I will then calculate the required replicates of grid cells necessary to find a significant statistical difference between the two groups. This final piece will be done while assuming the percent difference between the two-treatment group means for Yield counts will be 2.5%, 5% and 10%.

I will begin by researching and building my code using the 2015 yield data. After it is tested, runs correctly, and does everything I want it to do I will then apply it to the 2016 and 2017 datasets. Because we are looking at agriculture data over a three-year time span there will be differences between results from year to year. Looking at three seasons of data will allow me to see best case and worst-case scenarios within the data. This simulation of dividing the field into two Treatment Groups allows me to see how much variation there is within a field assuming that the enitre field was treated the same for each season. Therefore, we would expect the differences of the mean Yield counts between Treatment A and Treatment B to be about zero which would be the same as a typical null-hypothesis.

# R code methodology

Most of the R code in this document and for this analysis was built by examining the 2015 yield dataset. Therefore, I will present all the R code for my analysis of the 2015 dataset along with all output. I then re-ran the R code using the 2016 and 2017 datasets but will only show the code and output for the portions where I introduce the datasets and present the final statistics/plots. The accompanying R Markdown file contains the complete code for the datasets for all three seasons.

Here is a high-level overview of what I did with R (more details to follow).

1. Create a function called ‘**create.grid**’ to divide the field into grid cells.
2. Create a function called ‘**create.cells**’ to aggregate the data to the cell level.
3. Create a function called ‘**create.dataframe**’ that includes means, standard deviations, and sample sizes for all cells.
4. Assign rows to either Treatment A (odd rows) or Treatment B (even rows)
5. Create a function called ‘**create.comparison**’ where I aggregate to the Treatment level and compute an overall mean of Yield counts and standard deviation of cell means for both treatments.
6. Create function called ‘**required.replicates**’ that computes the coefficient of variation and the number of required replicates for percent difference values of 2.5%, 5% and 10% between Treatments A and B. This is run after all grid cell iterations have been run and a summary file has been created.

The above steps 1-6 create the functions used to extract, manipulate, format, and create new attributes and calculate the required replicates for each grid cell size iteration for this project.

The next steps involve the use of the 2015 season dataset and how the above functions are used.

1. Take a high-level look at the dataset, checking for missing values, plotting the data, etc...
2. Run steps 1-5 for each of the 18 iterations of Grid Cell sizes I chose to investigate.
3. Run step 6 (required.replicate function) to create final table of statistics and associated plots of results.
4. Repeat steps 1-9 for the datasets for the years 2016 and 2017.

# Function 1 - ‘create.grid’

This function will create a row number, *row*, column number, *col*, and cell number, *cell*, for each grid cell. The function uses two parameters that the user defines: *numrow* (number of rows desired) and *numcol* (number of columns desired).

For example, if a user calls the function with *numrow*=10 and *numcol*=10 then the function will produce row numbers from 1-10, column numbers from 1-10 and will compute cell numbers.

The formula for *row* is **ceiling(*Latitude* / (400/*numrow*))**, where 400 is the length of the field from north to south in meters.

The formula for *col* is **ceiling(*Longitude* / (600/*numcol*))**, where 600 is the length of the field from east to west in meters.

The formula for *cell* is: **(100)\*(*row*) + *col***.

The function prints out a message reporting what the minimum number of yield samples there are within a cell so I can keep track of which grid sizes produce cells with fewer than 30 yield samples. Lastly, the function returns the original dataset columns along with the new *row*, *col* and *cell* columns.

###########################################################################  
# Function 1 - Create Grid Over Field  
# 1 - Read in data  
# 2 - Assign Row, Column and Cell numbers to grid  
# 3 - Frequency count of yields per cell - print out minimum number  
# 4 - Frequency counts of yields by row and column  
# 5 - Return columns  
###########################################################################  
  
create.grid <- function(numrow, numcol) {  
  
HomePath = "../STAT600/home.2015.csv"  
home.dat <- read.csv(HomePath,header=TRUE)  
  
row <- ceiling(home.dat$Latitude/(400/numrow))  
col <- ceiling(home.dat$Longitude/(600/numcol))  
cell <- 100\*(row) + col  
  
freq\_yield <- aggregate(Yield ~ cell, data=home.dat, length)  
cat("Minimum number of Yield estimates is:",min(freq\_yield[,2]), "\n")  
  
return(list(X=home.dat$X, Yield=home.dat$Yield, Latitude=home.dat$Latitude, Longitude=home.dat$Longitude, row=row, col=col, cell=cell))  
}

# Function 2 - ‘create.cells’

This function, ‘create.cells’, requires no parameters as it takes the updated dataset (created from previous function) and aggregates the data to the cell level. The new aggregated variables include:

*Cell\_Means* - the mean Yield count for each cell

*Cell\_SDs* - the standard deviation of Yield count for each cell

*Cell\_Ns* - the number of Yield counts for each cell

*min\_lat* and *max\_lat* - the minimum and maximum latitude for each cell

*min\_lon* and *max\_lon* - the minimum and maximum longitude for each cell

#######################################################################  
# Function 2 - Summarize Data to Create Data Frame of Cells  
# 1 - Find minimum and maximum of Latitude and Longitude for each cell  
# 2 - Find Mean, SD and N of Yield for each cell  
# 3 - Return list of columns to create summarized data frame at cell level - in later function  
#######################################################################  
   
create.cells <- function() {  
min\_lat <- aggregate(Latitude ~ cell, data=y, min, na.rm=TRUE)  
max\_lat <- aggregate(Latitude ~ cell, data=y, max, na.rm=TRUE)  
  
min\_lon <- aggregate(Longitude ~ cell, data=y, min, na.rm=TRUE)  
max\_lon <- aggregate(Longitude ~ cell, data=y, max, na.rm=TRUE)  
  
Cell\_Means <- aggregate(Yield ~ cell, data=y, mean, na.rm=TRUE)  
Cell\_SDs <- aggregate(Yield ~ cell, data=y, sd, na.rm=TRUE)  
Cell\_Ns <- aggregate(Yield ~ cell, data=y, length)  
  
return(list(min\_lat=min\_lat, max\_lat=max\_lat, min\_lon=min\_lon, max\_lon=max\_lon, Cell\_Means=Cell\_Means,Cell\_SDs=Cell\_SDs, Cell\_Ns=Cell\_Ns))  
  
}

# 

# Function 3 - ‘create.dataframe’

The function, ‘create.dataframe’, requires no parameters and formally creates a proper dataframe from the aggregated data that was created in the previous function. The new datafame is called ‘**Grid.dat**’ and includes the same columns that the previous function created but with new names: *Cell*, *N*, *LatMin*, *LatMax*, *LonMin*, *LonMax*, *M* (the cell means) and *S* (the cell standard deviations). In addition, an indicator (0=no, 1=yes) is calculated to determine if the cell contains 30 or more sample units. A message is displayed to indicate the percent of cells with 30 or more sample units and the first six rows of the dataframe are printed.

###########################################################################  
# Function 3 - Create Data Frame of Summarized Cell Data  
# 1 - Keep necessary columns for data frame  
# 2 - Determine which cells have greater than or equal to 30 observations  
# 3 - Print out few records of data frame and percent of cells with greater than or equal to 30 observations  
###########################################################################  
  
create.dataframe <- function() {  
Grid.dat <- data.frame(z[,1], z[,14], z[,2], z[,4], z[,6], z[,8], z[,10], z[,12])  
colnames(Grid.dat) = c("Cell", "N", "LatMin", "LatMax", "LonMin", "LonMax","M", "S")   
  
Grid.dat$Cell\_Yield30 <- ifelse(Grid.dat$N >= 30, 1, 0)  
  
print(head(Grid.dat))  
  
cat("Percent of Cells with Greater than or Equal to 30 Yields:", 100\*(sum(Grid.dat$Cell\_Yield30) / length(Grid.dat$Cell\_Yield30)))  
  
return(Grid.dat)  
}

# Creating Treatment Groups

In the code for each of the 18 grid cell creation iterations there is R code to separate the ‘Grid.dat’ dataframe into a Treatment A dataframe and a Treatment B dataframe.

Below is R code that is used for the 10x10, 100 cells, iteration of creating grid cells:

**Grida.dat <- Grid.dat[c(c(1:10),c(21:30),c(41:50),c(61:70),c(81:90)),]**

**Gridb.dat <- Grid.dat[c(c(11:20),c(31:40),c(51:60),c(71:80),c(91:100)),]**

Where ‘Grida.dat’ becomes the Treatment A dataframe and ‘Gridb.dat’ becomes the Treatment B dataframe.

Again, code like this is repeated for each of the 18 iterations of grid cell sizes used in this project.

# Function 4 - ‘create.comparison’

This function, ‘create.comparison’ requires no parameters and performs calculations on the two dataframes created in the step above. For each datframe (‘**Grida.dat**’ and ‘**Gridb.dat**’), I calculate the mean of all values in the *M* column for each dataframe. These values are assigned ***M1*** for Treatment A and ***M2*** for Treatment B. I calculate the standard deviation of all *M* values for each dataframe. These values are assigned ***S1*** for Treatment A and ***S2*** for Treatment B. The length of each *M* column is calculated for each dataframe. These values are assigned as ***N1*** for Treatment A and ***N2*** for Treatment B. Finally, these six columns are combined into a one row dataframe.

Also in this function is code to find the average latitude and longitude for each cell within Treatment A and B, by adding the *LatMin* and *LatMax* together and dividing by two and doing the same with adding *LonMin* and *LonMax* together and dividing by two. I then plot these averages of Latitude and Longitude to see how my treatments are distributed across the field. If done correctly, the Treatment A cells will appear as red dots going across the field and the Treatment B cells will appear as blue dots going across the field.

###############################################################  
# Function 4 - Summarize data to Treatment levels (A & B)  
# 1 - For Dataframes A and B, find the average Latitude and Longitude for each cell  
# 2 - Plot the Group A and Group B cells of the grid in red and blue to verify they were created correctly  
# 3 - Calculate Mean, SD and N for Group A (M1, S1, N1) and repeat for Group B (M2, S2, N2)  
# 4 - End Result is a one row data frame simulating summary results for comparing Group A and Group B  
###############################################################  
  
create.comparison <- function() {  
  
# Find average Latitude and Longitude for plotting  
   
Grida.dat$Lat <- round((Grida.dat$LatMin + Grida.dat$LatMax)/2 ,2)  
Grida.dat$Lon <- round((Grida.dat$LonMin + Grida.dat$LonMax)/2 ,2)  
  
Gridb.dat$Lat <- round((Gridb.dat$LatMin + Gridb.dat$LatMax)/2 ,2)  
Gridb.dat$Lon <- round((Gridb.dat$LonMin + Gridb.dat$LonMax)/2 ,2)  
  
#Plot average Latitude and Longitude for both treatment dataframes to check accuracy  
  
plot(Lat ~ Lon, data=Grida.dat, xlim=c(0,600), ylim=c(0,400), type="p", pch=16, col="red",main="Red=Treatment 'A', Blue=Treatment'B'", xlab="Longitude", ylab="Latitude")  
points(Lat ~ Lon, data=Gridb.dat, xlim=c(0,600), ylim=c(0,400), type="p", pch=16, col="blue")  
  
# Calculate final Treatment Means, Standard Deviations and Sample Size for Treatments A & B  
  
M1 <- round(mean(Grida.dat$M),3)  
S1 <- round(sd(Grida.dat$M),3)  
N1 <- length(Grida.dat$M)  
  
M2 <- round(mean(Gridb.dat$M),3)  
S2 <- round(sd(Gridb.dat$M),3)  
N2 <- length(Gridb.dat$M)  
  
return(Grid.compare.dat <- data.frame(M1,S1,N1,M2,S2,N2))  
}

# 

# Function 5 - ‘required.replicates’

This function, ‘required.replicates’ requires nine parameters, six required and three optional. The six required are ***m1*, *m2*, *s1*, *s2*, *n1* and *n2***, are the means, standard deviations and sample sizes for Treatments A and B. The three optional parameters are for **,**  and ***%Diff***, where **alpha=0.05** and **beta=0.2** are default values and ***PctDiff\_fix*** **will equal 2.5, 5 or 10**.

The functions calculate the following:

(the number of grid cells between Treatments A & B)

. (the Coefficient of Variation between Treatments A & B)

(the actual percent difference between Treatments A & B)

(The number of required replicates for the Treatments A & B holding the PctDiff constant at 2.5, 5 or 10)

The function returns the above data elements in addition to the original columns in one table with 18 rows.

# Function for Required Replicates, CV and PctDiff  
  
required.replicates <- function(N1,N2,M1,M2,S1,S2, alpha=0.05, beta=0.20, PctDiff\_fix=5) {  
GridCells <- N1+N2  
CV <- round(100\*(sqrt( (S1^2 + S2^2) / 2 ) / ((M1 + M2)/2)),3)  
PctDiff <- round(100\*((M1 - M2) / ((M1 + M2)/2)),3)  
rr=ceiling(2 \* ((CV / PctDiff\_fix)^2) \* ((qnorm(1-alpha/2,0,1)+qnorm(1-beta,0,1))^2))  
return(list (GridCells=GridCells,N1=N1, M1=M1, S1=S1, N2=N2, M2=M2, S2=S2, CV=CV, PctDiff=PctDiff, rr=rr))  
}

# 

# Introduce the 2015 dataset

Here, I read in the dataset for the 2015 season and print out a summary of the dataset along with the first and last six records to inspect what is in the dataset. I also check for missing values among the variables. Lastly, I plot the points of latitude and longitude to get a sense of what the field looks like.

# Read in the data for 2015  
  
Home2015Path = "../STAT600/home.2015.csv"  
home.2015.dat <- read.csv(Home2015Path,header=TRUE)  
  
# Print a summary, first six and last six observations  
summary(home.2015.dat)

## X Yield Latitude Longitude   
## Min. : 97 Min. : 0.0 Min. : 0.1 Min. : 6.4   
## 1st Qu.: 8790 1st Qu.: 28.8 1st Qu.: 99.7 1st Qu.:135.5   
## Median :15065 Median : 35.1 Median :196.6 Median :292.7   
## Mean :13640 Mean : 35.7 Mean :198.3 Mean :294.6   
## 3rd Qu.:19992 3rd Qu.: 42.8 3rd Qu.:297.0 3rd Qu.:451.4   
## Max. :23419 Max. :103.8 Max. :399.9 Max. :595.8

head(home.2015.dat)

## X Yield Latitude Longitude  
## 1 97 40.86 0.8848 120.9  
## 2 98 43.54 2.1551 120.9  
## 3 99 42.34 3.4425 120.9  
## 4 100 39.22 4.7189 120.9  
## 5 101 38.25 6.0020 120.9  
## 6 102 36.96 7.2819 120.9

tail(home.2015.dat)

## X Yield Latitude Longitude  
## 11586 23414 59.07 8.547 595.8  
## 11587 23415 50.51 7.023 595.7  
## 11588 23416 45.83 5.561 595.7  
## 11589 23417 44.20 4.083 595.7  
## 11590 23418 46.08 2.609 595.7  
## 11591 23419 43.77 1.147 595.7

which(is.na(home.2015.dat$X))

## integer(0)

which(is.na(home.2015.dat$Yield))

## integer(0)

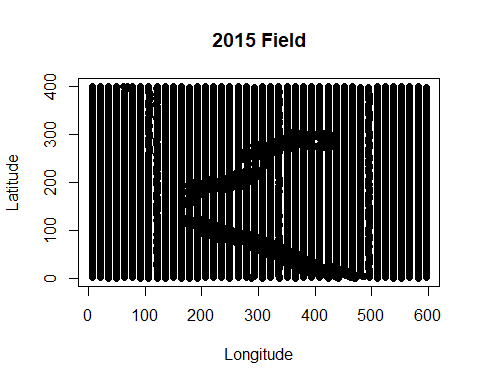
which(is.na(home.2015.dat$Latitude))

## integer(0)

which(is.na(home.2015.dat$Longitude))

## integer(0)

# Plot the data with x=Longitude and y=Latitude  
  
plot(home.2015.dat$Longitude, home.2015.dat$Latitude, type="p", pch=16,main="2015 Field", xlab="Longitude", ylab="Latitude")



# Create 18 sets of grid cells across farm field and corresponding data frames

Following are 18 R code chunks that call the above functions along with code to split the dataframe, summarized at the cell level, into separate Treatment A and Treatment B dataframes and then calling in the final function to create the final one row grid cell summary data frame. There are comments at the top of each R code chunk that tell how many grid cells that the chunk will create.

Within this code chunk the following output will be created:

1. Message saying what the minimum number of yield estimates there are within a cell
2. The first six rows of the original dataset with *row*, *col* and *cell* columns added
3. The first six rows of the dataframe summarized at the cell level
4. Message saying what percent of of cell grids have 30 or more yield counts
5. A scatter plot of latitude and longitude points showing Treatment A cells (red) and Treatment B cells (blue)
6. The final one row summary dataframe for this grid cell iteration

###########################################  
# CREATE 2 ROWS BY 2 COLUMNS FOR 4 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=2, numcol=2))

## Minimum number of Yield estimates is: 2774

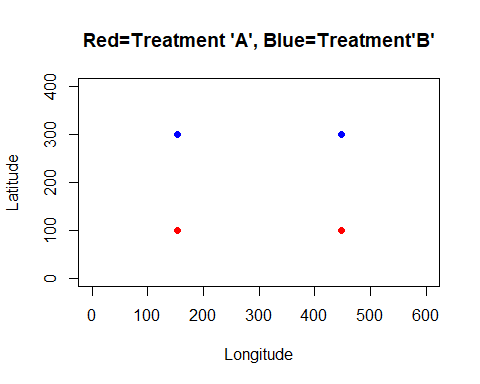
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 1 101  
## 2 98 43.54 2.1551 120.9 1 1 101  
## 3 99 42.34 3.4425 120.9 1 1 101  
## 4 100 39.22 4.7189 120.9 1 1 101  
## 5 101 38.25 6.0020 120.9 1 1 101  
## 6 102 36.96 7.2819 120.9 1 1 101

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 3052 0.05954 200.0 6.404 299.1 34.10 12.202 1  
## 2 102 2842 0.19633 199.9 300.254 595.8 39.33 10.536 1  
## 3 201 2923 200.02463 399.9 6.398 299.3 32.48 8.318 1  
## 4 202 2774 200.11001 399.9 300.341 595.8 36.99 10.526 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(1,2),]  
Gridb.dat <- Grid.dat[c(3,4),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid1.compare.dat <- create.comparison()



Grid1.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.71 3.698 2 34.73 3.19 2

##########################################  
# CREATE 4 ROWS BY 6 COLUMNS FOR 24 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=4, numcol=6))

## Minimum number of Yield estimates is: 444

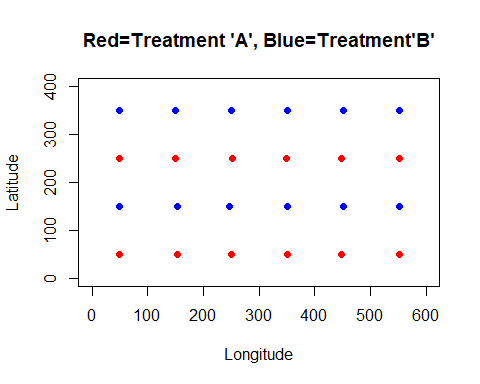
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 2 102  
## 2 98 43.54 2.1551 120.9 1 2 102  
## 3 99 42.34 3.4425 120.9 1 2 102  
## 4 100 39.22 4.7189 120.9 1 2 102  
## 5 101 38.25 6.0020 120.9 1 2 102  
## 6 102 36.96 7.2819 120.9 1 2 102

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 467 0.16869 99.94 6.404 92.33 25.12 5.789 1  
## 2 102 502 0.05954 99.65 106.515 199.81 36.16 8.877 1  
## 3 103 504 0.43137 99.98 200.858 299.09 39.05 14.073 1  
## 4 104 495 0.19633 99.89 300.254 399.56 41.34 11.742 1  
## 5 105 472 0.33789 99.67 401.183 494.83 41.64 9.598 1  
## 6 106 463 0.32352 99.55 509.113 595.79 37.52 8.683 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:6),c(13:18)),]  
Gridb.dat <- Grid.dat[c(c(7:12),c(19:24)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid2.compare.dat <- create.comparison()



Grid2.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.45 4.937 12 35 6.153 12

##########################################  
# CREATE 40 ROWS BY 40 COLUMNS FOR 1600 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=40, numcol=40))

## Minimum number of Yield estimates is: 1

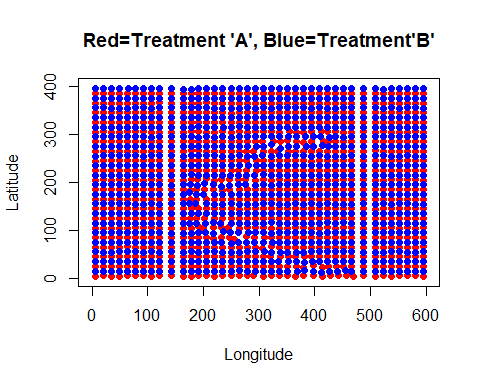
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 9 109  
## 2 98 43.54 2.1551 120.9 1 9 109  
## 3 99 42.34 3.4425 120.9 1 9 109  
## 4 100 39.22 4.7189 120.9 1 9 109  
## 5 101 38.25 6.0020 120.9 1 9 109  
## 6 102 36.96 7.2819 120.9 1 9 109

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 6 1.0210 8.559 6.404 6.487 28.10 3.108 0  
## 2 102 7 1.2993 9.768 20.754 20.798 27.64 3.249 0  
## 3 103 7 0.1687 9.468 35.037 35.074 25.51 3.766 0  
## 4 104 7 0.2977 9.607 49.381 49.396 30.23 3.706 0  
## 5 105 6 1.5105 9.408 63.650 63.730 22.70 1.461 0  
## 6 106 6 1.4256 8.650 77.964 77.999 24.02 1.809 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 0

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:40),c(81:120),c(161:200),c(241:280),c(321:360),c(401:440),c(481:520),c(561:600),c(641:680),c(721:760),c(801:840),c(881:920),c(961:1000),c(1041:1080),c(1121:1160),c(1201:1240),c(1281:1320),c(1361:1400),c(1441:1480),c(1521:1560)),]  
  
Gridb.dat <- Grid.dat[c(c(41:80),c(121:160),c(201:240),c(281:320),c(361:400),c(441:480),c(521:560),c(601:640),c(681:720),c(761:800),c(841:880),c(921:960),c(1001:1040),c(1081:1120),c(1161:1200),c(1241:1280),c(1321:1360),c(1401:1440),c(1481:1520),c(1561:1600)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid3.compare.dat <- create.comparison()



Grid3.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.02 9.118 800 35.88 9.108 800

##########################################  
# CREATE 4 ROWS BY 4 COLUMNS FOR 16 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=4, numcol=4))

## Minimum number of Yield estimates is: 649

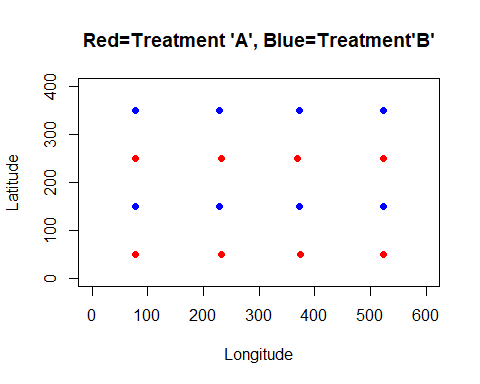
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 1 101  
## 2 98 43.54 2.1551 120.9 1 1 101  
## 3 99 42.34 3.4425 120.9 1 1 101  
## 4 100 39.22 4.7189 120.9 1 1 101  
## 5 101 38.25 6.0020 120.9 1 1 101  
## 6 102 36.96 7.2819 120.9 1 1 101

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 761 0.05954 99.94 6.404 150.0 27.84 6.782 1  
## 2 102 712 0.08697 99.98 164.368 299.1 39.86 12.905 1  
## 3 103 699 0.19633 99.89 300.254 448.8 42.01 11.435 1  
## 4 104 731 0.32352 99.67 450.362 595.8 38.47 8.723 1  
## 5 201 849 100.13851 199.90 6.419 150.0 32.72 12.528 1  
## 6 202 730 100.25308 199.98 164.328 292.8 36.59 12.182 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:4),c(9:12)),]  
Gridb.dat <- Grid.dat[c(c(5:8),c(13:16)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid4.compare.dat <- create.comparison()



Grid4.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.59 4.583 8 35.17 5.426 8

##########################################  
# CREATE 6 ROWS BY 6 COLUMNS FOR 36 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=6, numcol=6))

## Minimum number of Yield estimates is: 295

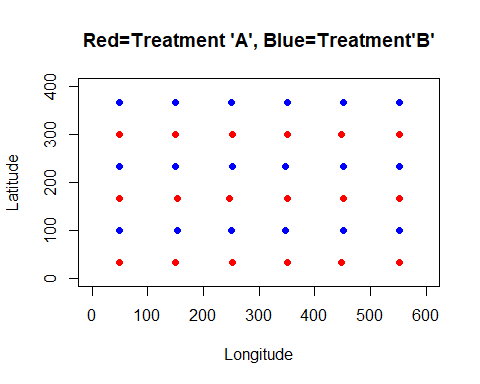
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 2 102  
## 2 98 43.54 2.1551 120.9 1 2 102  
## 3 99 42.34 3.4425 120.9 1 2 102  
## 4 100 39.22 4.7189 120.9 1 2 102  
## 5 101 38.25 6.0020 120.9 1 2 102  
## 6 102 36.96 7.2819 120.9 1 2 102

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 310 0.16869 66.30 6.404 92.32 24.69 5.838 1  
## 2 102 334 0.05954 66.61 106.536 191.76 37.63 9.414 1  
## 3 103 322 0.43137 66.63 206.077 299.09 43.01 13.439 1  
## 4 104 340 0.19633 66.39 301.045 399.56 42.11 13.047 1  
## 5 105 323 0.33789 66.25 401.183 494.83 42.00 10.513 1  
## 6 106 311 0.32352 66.63 509.113 595.79 39.74 9.238 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:6),c(13:18),c(25:30)),]  
Gridb.dat <- Grid.dat[c(c(7:12),c(19:24),c(31:36)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid5.compare.dat <- create.comparison()



Grid5.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.48 6.022 18 34.99 5.467 18

##########################################  
# CREATE 8 ROWS BY 8 COLUMNS FOR 64 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=8, numcol=8))

## Minimum number of Yield estimates is: 153

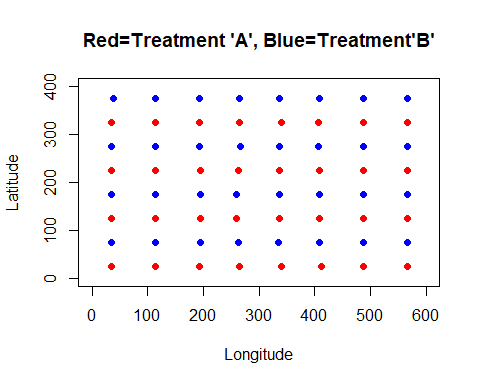
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 2 102  
## 2 98 43.54 2.1551 120.9 1 2 102  
## 3 99 42.34 3.4425 120.9 1 2 102  
## 4 100 39.22 4.7189 120.9 1 2 102  
## 5 101 38.25 6.0020 120.9 1 2 102  
## 6 102 36.96 7.2819 120.9 1 2 102

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 165 0.16869 49.86 6.404 63.73 24.51 4.919 1  
## 2 102 218 0.05954 49.82 77.951 149.96 31.73 7.920 1  
## 3 103 171 0.08697 49.97 164.377 220.60 45.57 10.869 1  
## 4 104 168 0.43137 49.95 234.946 293.04 43.91 13.018 1  
## 5 105 179 0.19633 49.86 307.081 373.76 45.13 12.078 1  
## 6 106 182 0.63460 49.80 375.040 448.85 42.26 11.881 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:8),c(17:24),c(33:40),c(49:56)),]  
Gridb.dat <- Grid.dat[c(c(9:16),c(25:32),c(41:48),c(57:64)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid6.compare.dat <- create.comparison()



Grid6.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36 6.409 32 35.59 6.314 32

##########################################  
# CREATE 6 ROWS BY 8 COLUMNS FOR 48 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=6, numcol=8))

## Minimum number of Yield estimates is: 203

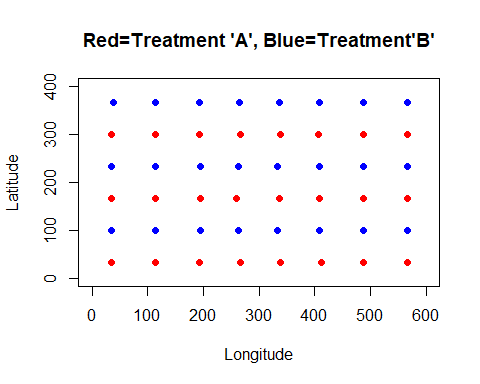
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 2 102  
## 2 98 43.54 2.1551 120.9 1 2 102  
## 3 99 42.34 3.4425 120.9 1 2 102  
## 4 100 39.22 4.7189 120.9 1 2 102  
## 5 101 38.25 6.0020 120.9 1 2 102  
## 6 102 36.96 7.2819 120.9 1 2 102

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 219 0.16869 66.30 6.404 63.73 24.48 4.953 1  
## 2 102 288 0.05954 66.61 77.949 149.96 30.77 7.882 1  
## 3 103 229 0.08697 66.61 164.377 220.60 42.87 10.701 1  
## 4 104 230 0.43137 66.63 234.937 299.09 43.62 13.853 1  
## 5 105 250 0.19633 66.39 301.045 373.76 41.92 13.768 1  
## 6 106 233 0.63460 66.17 375.040 448.85 43.36 10.989 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:8),c(17:24),c(33:40)),]  
Gridb.dat <- Grid.dat[c(c(9:16),c(25:32),c(41:48)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid8.compare.dat <- create.comparison()



Grid8.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.57 6.514 24 35.01 5.757 24

##########################################  
# CREATE 10 ROWS BY 10 COLUMNS FOR 100 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=10, numcol=10))

## Minimum number of Yield estimates is: 97

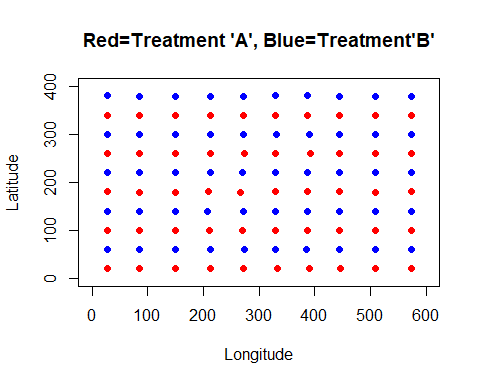
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 3 103  
## 2 98 43.54 2.1551 120.9 1 3 103  
## 3 99 42.34 3.4425 120.9 1 3 103  
## 4 100 39.22 4.7189 120.9 1 3 103  
## 5 101 38.25 6.0020 120.9 1 3 103  
## 6 102 36.96 7.2819 120.9 1 3 103

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 106 0.16869 39.47 6.404 49.42 25.16 4.020 1  
## 2 102 110 0.18063 39.72 63.632 106.61 27.62 6.753 1  
## 3 103 145 0.05954 39.95 120.860 178.15 40.95 10.487 1  
## 4 104 103 0.84160 39.65 191.596 235.04 45.97 9.704 1  
## 5 105 109 0.43137 39.45 249.364 293.04 47.54 14.267 1  
## 6 106 109 0.19633 39.90 307.091 359.70 47.27 8.932 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:10),c(21:30),c(41:50),c(61:70),c(81:90)),]  
Gridb.dat <- Grid.dat[c(c(11:20),c(31:40),c(51:60),c(71:80),c(91:100)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid7.compare.dat <- create.comparison()



Grid7.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.69 7.009 50 34.99 5.961 50

##########################################  
# CREATE 8 ROWS BY 10 COLUMNS FOR 80 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=8, numcol=10))

## Minimum number of Yield estimates is: 121

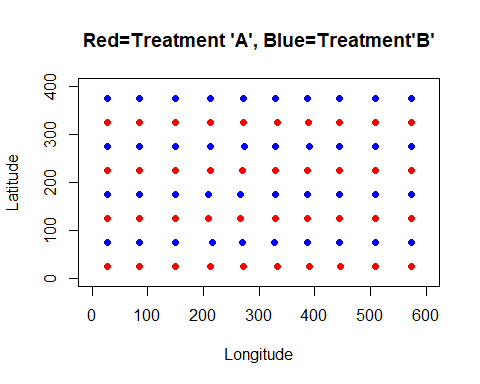
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 3 103  
## 2 98 43.54 2.1551 120.9 1 3 103  
## 3 99 42.34 3.4425 120.9 1 3 103  
## 4 100 39.22 4.7189 120.9 1 3 103  
## 5 101 38.25 6.0020 120.9 1 3 103  
## 6 102 36.96 7.2819 120.9 1 3 103

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 133 0.16869 49.86 6.404 49.43 24.61 3.954 1  
## 2 102 139 0.18063 49.75 63.632 106.61 27.28 6.879 1  
## 3 103 184 0.05954 49.97 120.860 178.15 39.80 9.998 1  
## 4 104 129 0.84160 49.18 191.596 235.04 43.08 11.363 1  
## 5 105 137 0.43137 49.95 249.364 293.04 45.15 13.841 1  
## 6 106 139 0.19633 49.86 307.081 359.70 46.51 9.540 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:10),c(21:30),c(41:50),c(61:70)),]  
Gridb.dat <- Grid.dat[c(c(11:20),c(31:40),c(51:60),c(71:80)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid9.compare.dat <- create.comparison()



Grid9.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36 6.28 40 35.6 6.232 40

##########################################  
# CREATE 12 ROWS BY 12 COLUMNS FOR 144 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=12, numcol=12))

## Minimum number of Yield estimates is: 60

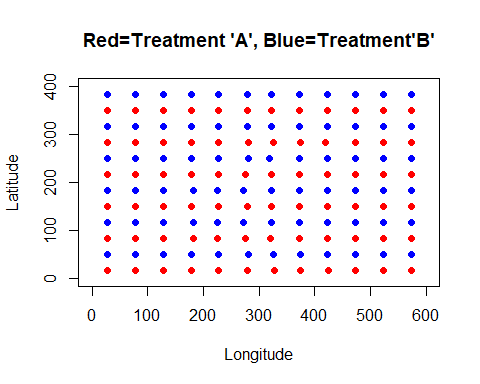
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 3 103  
## 2 98 43.54 2.1551 120.9 1 3 103  
## 3 99 42.34 3.4425 120.9 1 3 103  
## 4 100 39.22 4.7189 120.9 1 3 103  
## 5 101 38.25 6.0020 120.9 1 3 103  
## 6 102 36.96 7.2819 120.9 1 3 103

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 89 0.16869 33.12 6.404 49.42 25.47 4.088 1  
## 2 102 66 0.74595 33.05 63.632 92.32 24.78 6.601 1  
## 3 103 99 0.05954 33.10 106.541 149.96 35.81 7.947 1  
## 4 104 67 0.08697 32.56 164.380 191.76 48.38 8.918 1  
## 5 105 87 0.74240 33.05 206.077 249.45 49.21 10.725 1  
## 6 106 70 0.43137 32.88 263.817 293.04 47.74 14.964 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:12),c(25:36),c(49:60),c(73:84),c(97:108),c(121:132)),]  
Gridb.dat <- Grid.dat[c(c(13:24),c(37:48),c(61:72),c(85:96),c(109:120),c(133:144)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid10.compare.dat <- create.comparison()



Grid10.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 35.75 7.033 72 36.1 6.362 72

##########################################  
# CREATE 14 ROWS BY 14 COLUMNS FOR 196 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=14, numcol=14))

## Minimum number of Yield estimates is: 50

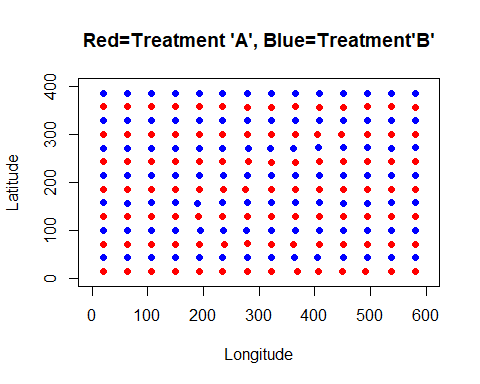
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 3 103  
## 2 98 43.54 2.1551 120.9 1 3 103  
## 3 99 42.34 3.4425 120.9 1 3 103  
## 4 100 39.22 4.7189 120.9 1 3 103  
## 5 101 38.25 6.0020 120.9 1 3 103  
## 6 102 36.96 7.2819 120.9 1 3 103

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 58 0.16869 28.34 6.404 35.12 25.11 3.701 1  
## 2 102 56 0.29774 28.46 49.347 78.00 23.94 6.076 1  
## 3 103 63 0.18063 28.14 92.234 120.91 31.80 4.733 1  
## 4 104 62 0.05954 28.09 135.374 164.49 44.92 8.727 1  
## 5 105 56 0.41640 28.45 178.064 206.18 47.55 9.402 1  
## 6 106 55 0.74240 28.50 220.500 249.44 51.05 11.459 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:14),c(29:42),c(57:70),c(85:98),c(113:126),c(141:154),c(169:182)),]  
Gridb.dat <- Grid.dat[c(c(15:28),c(43:56),c(71:84),c(99:112),c(127:140),c(155:168),c(183:196)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid11.compare.dat <- create.comparison()



Grid11.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.24 7.424 98 35.45 6.36 98

##########################################  
# CREATE 16 ROWS BY 16 COLUMNS FOR 256 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=16, numcol=16))

## Minimum number of Yield estimates is: 30

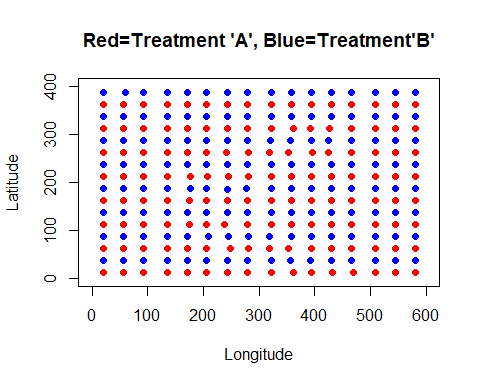
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 4 104  
## 2 98 43.54 2.1551 120.9 1 4 104  
## 3 99 42.34 3.4425 120.9 1 4 104  
## 4 100 39.22 4.7189 120.9 1 4 104  
## 5 101 38.25 6.0020 120.9 1 4 104  
## 6 102 36.96 7.2819 120.9 1 4 104

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 50 0.16869 24.97 6.404 35.12 25.33 3.916 1  
## 2 102 31 0.29774 23.67 49.347 63.73 24.51 6.603 1  
## 3 103 53 0.18063 24.63 77.951 106.61 28.52 5.065 1  
## 4 104 55 0.05954 24.85 120.868 149.96 40.22 7.764 1  
## 5 105 36 0.08697 24.71 164.380 178.15 47.69 8.675 1  
## 6 106 48 0.84160 24.57 191.659 220.58 50.92 9.868 1  
## Percent of Cells with Greater than or Equal to 30 Yields: 100

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:16),c(33:48),c(65:80),c(97:112),c(129:144),c(161:176),c(193:208),c(225:240)),]  
Gridb.dat <- Grid.dat[c(c(17:32),c(49:64),c(81:96),c(113:128),c(145:160),c(177:192),c(209:224),c(241:256)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid12.compare.dat <- create.comparison()



Grid12.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.04 7.281 128 35.53 7.205 128

##########################################  
# CREATE 30 ROWS BY 30 COLUMNS FOR 900 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=30, numcol=30))

## Minimum number of Yield estimates is: 6

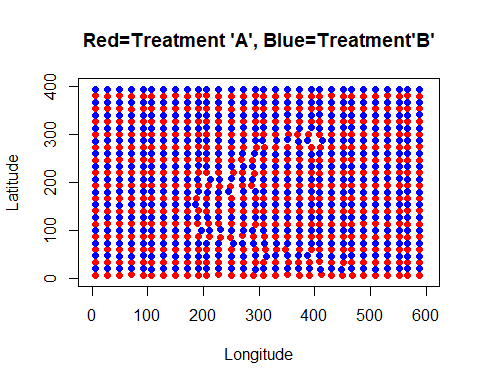
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 7 107  
## 2 98 43.54 2.1551 120.9 1 7 107  
## 3 99 42.34 3.4425 120.9 1 7 107  
## 4 100 39.22 4.7189 120.9 1 7 107  
## 5 101 38.25 6.0020 120.9 1 7 107  
## 6 102 36.96 7.2819 120.9 1 7 107

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 9 1.0210 13.08 6.404 6.544 29.11 3.232 0  
## 2 102 18 0.1687 12.59 20.725 35.074 25.61 3.901 0  
## 3 103 9 0.2977 12.73 49.347 49.396 29.24 3.771 0  
## 4 104 17 1.4256 13.00 63.650 77.999 23.71 2.540 0  
## 5 105 9 0.7459 12.69 92.234 92.316 25.27 1.788 0  
## 6 106 10 0.1806 12.20 106.544 106.598 32.29 2.931 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 0.1111

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:30),c(61:90),c(121:150),c(181:210),c(241:270),c(301:330),c(361:390),c(421:450),c(481:510),c(541:570),c(601:630),c(661:690),c(721:750),c(781:810),c(841:870)),]  
  
Gridb.dat <- Grid.dat[c(c(31:60),c(91:120),c(151:180),c(211:240),c(271:300),c(331:360),c(391:420),c(451:480),c(511:540),c(571:600),c(631:660),c(691:720),c(751:780),c(811:840),c(871:900)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid13.compare.dat <- create.comparison()



Grid13.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.36 8.51 450 35.91 8.165 450

##########################################  
# CREATE 26 ROWS BY 26 COLUMNS FOR 676 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=26, numcol=26))

## Minimum number of Yield estimates is: 8

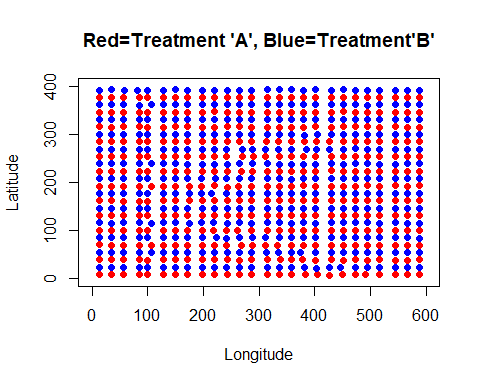
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 6 106  
## 2 98 43.54 2.1551 120.9 1 6 106  
## 3 99 42.34 3.4425 120.9 1 6 106  
## 4 100 39.22 4.7189 120.9 1 6 106  
## 5 101 38.25 6.0020 120.9 1 6 106  
## 6 102 36.96 7.2819 120.9 1 6 106

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 21 1.02098 15.38 6.404 20.80 27.29 3.684 0  
## 2 102 10 0.16869 14.14 35.037 35.07 24.55 4.076 0  
## 3 103 19 0.29774 14.29 49.347 63.73 26.07 4.305 0  
## 4 104 18 0.74595 14.47 77.951 92.31 25.33 2.934 0  
## 5 105 14 0.18063 14.87 92.312 106.60 31.26 3.997 0  
## 6 106 23 0.05954 15.00 120.868 135.43 40.56 4.618 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 1.183

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:26),c(53:78),c(105:130),c(157:182),c(209:234),c(261:286),c(313:338),c(365:390),c(417:442),c(469:494),c(521:546),c(573:598),c(625:650)),]  
  
Gridb.dat <- Grid.dat[c(c(27:52),c(79:104),c(131:156),c(183:208),c(235:260),c(287:312),c(339:364),c(391:416),c(443:468),c(495:520),c(547:572),c(599:624),c(651:676)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid14.compare.dat <- create.comparison()



Grid14.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 35.91 8.246 338 35.88 8.24 338

##########################################  
# CREATE 24 ROWS BY 24 COLUMNS FOR 576 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=24, numcol=24))

## Minimum number of Yield estimates is: 9

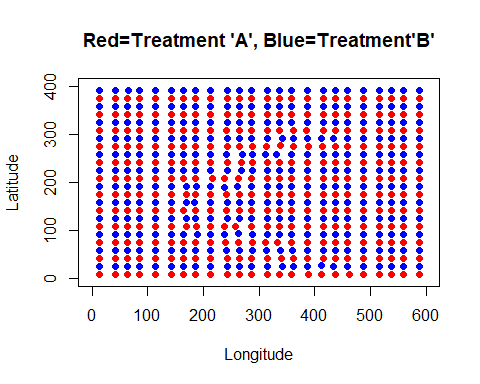
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 5 105  
## 2 98 43.54 2.1551 120.9 1 5 105  
## 3 99 42.34 3.4425 120.9 1 5 105  
## 4 100 39.22 4.7189 120.9 1 5 105  
## 5 101 38.25 6.0020 120.9 1 5 105  
## 6 102 36.96 7.2819 120.9 1 5 105

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 22 1.02098 16.08 6.404 20.80 27.21 3.616 0  
## 2 102 22 0.16869 15.87 35.037 49.40 27.11 4.379 0  
## 3 103 10 1.51054 15.73 63.650 63.73 21.78 3.938 0  
## 4 104 22 0.74595 15.92 77.951 92.32 25.83 3.411 0  
## 5 105 26 0.18063 16.28 106.544 120.91 34.65 3.817 0  
## 6 106 24 0.05954 16.52 135.387 149.96 47.06 4.968 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 1.389

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:24),c(49:72),c(97:120),c(145:168),c(193:216),c(241:264),c(289:312),c(337:360),c(385:408),c(433:456),c(481:504),c(529:552)),]  
  
Gridb.dat <- Grid.dat[c(c(25:48),c(73:96),c(121:144),c(169:192),c(217:240),c(265:288),c(313:336),c(361:384),c(409:432),c(457:480),c(505:528),c(553:576)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid15.compare.dat <- create.comparison()



Grid15.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 35.98 7.966 288 35.67 8.168 288

##########################################  
# CREATE 22 ROWS BY 22 COLUMNS FOR 484 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=22, numcol=22))

## Minimum number of Yield estimates is: 11

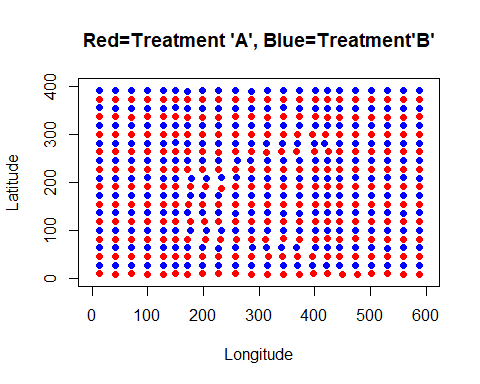
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 5 105  
## 2 98 43.54 2.1551 120.9 1 5 105  
## 3 99 42.34 3.4425 120.9 1 5 105  
## 4 100 39.22 4.7189 120.9 1 5 105  
## 5 101 38.25 6.0020 120.9 1 5 105  
## 6 102 36.96 7.2819 120.9 1 5 105

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 25 1.02098 18.16 6.404 20.8 26.99 3.521 0  
## 2 102 24 0.16869 17.41 35.037 49.4 26.94 4.622 0  
## 3 103 23 1.42561 17.37 63.650 78.0 23.77 4.608 0  
## 4 104 26 0.18063 17.51 92.234 106.6 29.73 4.005 0  
## 5 105 27 0.05954 17.56 120.868 135.4 40.19 4.732 0  
## 6 106 13 1.01523 17.93 149.885 150.0 49.08 6.077 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 4.132

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:22),c(45:66),c(89:110),c(133:154),c(177:198),c(221:242),c(265:286),c(309:330),c(353:374),c(397:418),c(441:462)),]  
  
Gridb.dat <- Grid.dat[c(c(23:44),c(67:88),c(111:132),c(155:176),c(199:220),c(243:264),c(287:308),c(331:352),c(375:396),c(419:440),c(463:484)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid16.compare.dat <- create.comparison()



Grid16.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 35.88 8.225 242 35.94 7.038 242

##########################################  
# CREATE 20 ROWS BY 20 COLUMNS FOR 400 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=20, numcol=20))

## Minimum number of Yield estimates is: 20

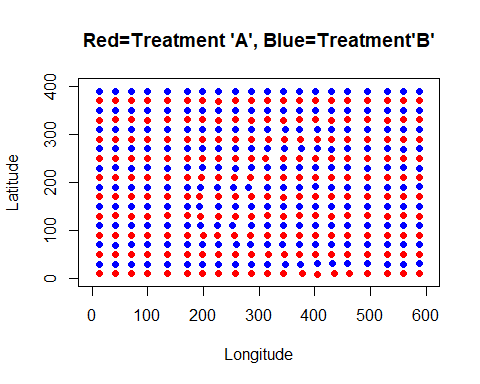
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 5 105  
## 2 98 43.54 2.1551 120.9 1 5 105  
## 3 99 42.34 3.4425 120.9 1 5 105  
## 4 100 39.22 4.7189 120.9 1 5 105  
## 5 101 38.25 6.0020 120.9 1 5 105  
## 6 102 36.96 7.2819 120.9 1 5 105

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 27 1.02098 19.52 6.404 20.8 27.01 3.383 0  
## 2 102 26 0.16869 18.97 35.037 49.4 27.13 4.854 0  
## 3 103 25 1.42561 18.90 63.650 78.0 23.82 5.079 0  
## 4 104 28 0.18063 18.83 92.234 106.6 29.94 3.955 0  
## 5 105 43 0.05954 19.33 120.868 150.0 42.69 6.687 1  
## 6 106 28 0.08697 18.72 164.380 178.2 45.48 8.394 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 25

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:20),c(41:60),c(81:100),c(121:140),c(161:180),c(201:220),c(241:260),c(281:300),c(321:340),c(361:380)),]  
  
Gridb.dat <- Grid.dat[c(c(21:40),c(61:80),c(101:120),c(141:160),c(181:200),c(221:240),c(261:280),c(301:320),c(341:360),c(381:400)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid17.compare.dat <- create.comparison()



Grid17.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.18 8.147 200 35.55 7.175 200

##########################################  
# CREATE 18 ROWS BY 18 COLUMNS FOR 324 CELLS  
##########################################  
  
y <- data.frame(create.grid(numrow=18, numcol=18))

## Minimum number of Yield estimates is: 26

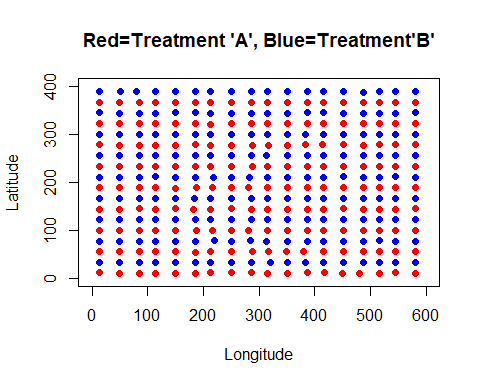
head(y)

## X Yield Latitude Longitude row col cell  
## 1 97 40.86 0.8848 120.9 1 4 104  
## 2 98 43.54 2.1551 120.9 1 4 104  
## 3 99 42.34 3.4425 120.9 1 4 104  
## 4 100 39.22 4.7189 120.9 1 4 104  
## 5 101 38.25 6.0020 120.9 1 4 104  
## 6 102 36.96 7.2819 120.9 1 4 104

z <- data.frame(create.cells())  
  
Grid.dat <- create.dataframe()

## Cell N LatMin LatMax LonMin LonMax M S Cell\_Yield30  
## 1 101 30 1.02098 22.15 6.404 20.80 26.67 3.464 1  
## 2 102 44 0.16869 22.08 35.037 63.73 24.69 5.800 1  
## 3 103 30 0.74595 21.69 77.951 92.32 26.74 4.962 1  
## 4 104 34 0.18063 21.46 106.544 120.91 33.56 3.990 1  
## 5 105 49 0.05954 22.10 135.386 164.45 47.15 6.894 1  
## 6 106 28 0.41640 21.66 178.064 191.76 43.42 10.223 0  
## Percent of Cells with Greater than or Equal to 30 Yields: 71.91

######################################################  
# Divide up Data Frame into two - Group A and Group B  
######################################################  
Grida.dat <- Grid.dat[c(c(1:18),c(37:54),c(73:90),c(109:126),c(145:162),c(181:198),c(217:234),c(253:270),c(289:306)),]  
  
Gridb.dat <- Grid.dat[c(c(19:36),c(55:72),c(91:108),c(127:144),c(163:180),c(199:216),c(235:252),c(271:288),c(307:324)),]  
  
#######################################  
# Create Final Comparison File   
#######################################  
Grid18.compare.dat <- create.comparison()



Grid18.compare.dat

## M1 S1 N1 M2 S2 N2  
## 1 36.3 7.36 162 35.5 7.643 162

# 

# Calculate Required Replicates and Plots for 2015 Dataset

Now that we have created 18 summary dataframes for Treatments A and B the following code will begin the conclusion of the analysis of the 2015 datasett. First, we will stack the 18 dataframes together. Then the ‘required.replicates’ function is run three times, once each for a fixed value of the Percent Difference paramteter, for values of 2.5%, 5% and 10%. The dataframe with all 18 rows is printed three times (once for each value of PctDiff used in the function) along with the additional columns calculated within the function that include: *CV* (Coefficient of Variation), *PctDiff* (actual computed Percent Difference) and *rr* the required replicates required.

Plots of the results are printed and include:

1. Required Replicates vs Coefficient of Variation
2. Number of Grid Cells vs Coefficient of Variation
3. Required Replicates vs Number of Grid Cells (for PctDiff = 5% and 10%)
4. Required Replicates vs Number of Grid Cells (for PctDiff = 5% and 10% and 2.5%)

########################################################################  
# 2015 Analysis  
# 1 - Stack all 17 summary files for varying number of grid cells together  
# 2 - Run Required Replicates 3 times for different values of PctDiff  
# 3 - Print out resulting tables showing actual PctDiff, CV and Required Replicates for each number of Grid Cells  
# 4 - Print Plots  
########################################################################  
  
# Stack all files together  
  
allGrid.dat = rbind(Grid1.compare.dat, Grid2.compare.dat,Grid3.compare.dat,Grid4.compare.dat,Grid5.compare.dat,Grid6.compare.dat, Grid7.compare.dat,Grid8.compare.dat,Grid9.compare.dat,Grid10.compare.dat,Grid11.compare.dat,Grid12.compare.dat,Grid13.compare.dat,Grid14.compare.dat,Grid15.compare.dat,Grid16.compare.dat,Grid17.compare.dat,Grid18.compare.dat)  
  
# Run Required Replicates function 3 times-once each for PctDiff set to 2.5%, 5% and 10%  
  
allGrid.final\_025 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=2.5))  
  
allGrid.final\_05 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat**$**N2,

allGrid.dat**$**M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=5))  
  
allGrid.final\_10 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=10))  
  
# Print out summary tables for PctDiff = 2.5%  
allGrid.final\_025[order(allGrid.final\_025$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 36.71 3.698 2 34.73 3.190 9.667 5.531 235  
## 4 16 8 36.59 4.583 8 35.17 5.426 13.996 3.960 493  
## 2 24 12 36.45 4.937 12 35.00 6.153 15.613 4.047 613  
## 5 36 18 36.48 6.022 18 34.99 5.467 16.092 4.172 651  
## 8 48 24 36.57 6.514 24 35.01 5.757 17.177 4.370 742  
## 6 64 32 36.00 6.409 32 35.59 6.314 17.773 1.129 794  
## 9 80 40 36.00 6.280 40 35.60 6.232 17.474 1.112 767  
## 7 100 50 36.69 7.009 50 34.99 5.961 18.154 4.760 828  
## 10 144 72 35.75 7.033 72 36.10 6.362 18.668 -0.969 876  
## 11 196 98 36.24 7.424 98 35.45 6.360 19.283 2.195 934  
## 12 256 128 36.04 7.281 128 35.53 7.205 20.241 1.422 1030  
## 18 324 162 36.30 7.360 162 35.50 7.643 20.900 2.234 1098  
## 17 400 200 36.18 8.147 200 35.55 7.175 21.404 1.754 1151  
## 16 484 242 35.88 8.225 242 35.94 7.038 21.315 -0.145 1142  
## 15 576 288 35.98 7.966 288 35.67 8.168 22.518 0.854 1274  
## 14 676 338 35.91 8.246 338 35.88 8.240 22.964 0.070 1325  
## 13 900 450 36.36 8.510 450 35.91 8.165 23.077 1.262 1338  
## 3 1600 800 36.02 9.118 800 35.88 9.108 25.348 0.395 1614

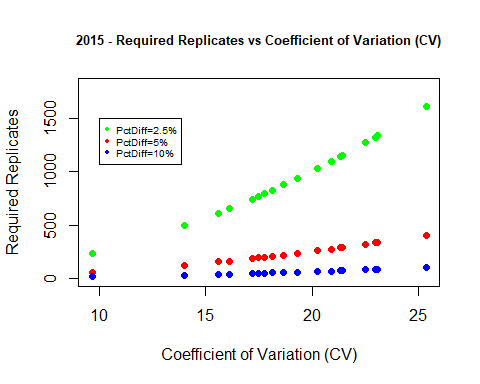
# Print out summary tables for PctDiff = 5%  
allGrid.final\_05[order(allGrid.final\_05$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 36.71 3.698 2 34.73 3.190 9.667 5.531 59  
## 4 16 8 36.59 4.583 8 35.17 5.426 13.996 3.960 124  
## 2 24 12 36.45 4.937 12 35.00 6.153 15.613 4.047 154  
## 5 36 18 36.48 6.022 18 34.99 5.467 16.092 4.172 163  
## 8 48 24 36.57 6.514 24 35.01 5.757 17.177 4.370 186  
## 6 64 32 36.00 6.409 32 35.59 6.314 17.773 1.129 199  
## 9 80 40 36.00 6.280 40 35.60 6.232 17.474 1.112 192  
## 7 100 50 36.69 7.009 50 34.99 5.961 18.154 4.760 207  
## 10 144 72 35.75 7.033 72 36.10 6.362 18.668 -0.969 219  
## 11 196 98 36.24 7.424 98 35.45 6.360 19.283 2.195 234  
## 12 256 128 36.04 7.281 128 35.53 7.205 20.241 1.422 258  
## 18 324 162 36.30 7.360 162 35.50 7.643 20.900 2.234 275  
## 17 400 200 36.18 8.147 200 35.55 7.175 21.404 1.754 288  
## 16 484 242 35.88 8.225 242 35.94 7.038 21.315 -0.145 286  
## 15 576 288 35.98 7.966 288 35.67 8.168 22.518 0.854 319  
## 14 676 338 35.91 8.246 338 35.88 8.240 22.964 0.070 332  
## 13 900 450 36.36 8.510 450 35.91 8.165 23.077 1.262 335  
## 3 1600 800 36.02 9.118 800 35.88 9.108 25.348 0.395 404

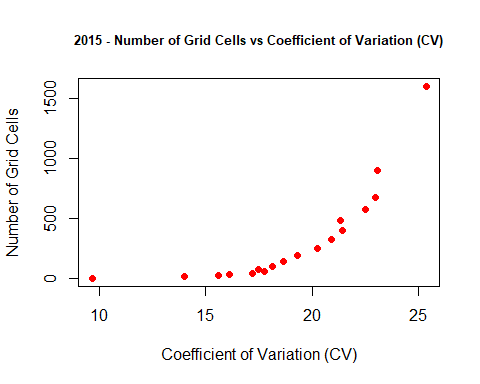
# Print out summary tables for PctDiff = 10%  
allGrid.final\_10[order(allGrid.final\_10$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 36.71 3.698 2 34.73 3.190 9.667 5.531 15  
## 4 16 8 36.59 4.583 8 35.17 5.426 13.996 3.960 31  
## 2 24 12 36.45 4.937 12 35.00 6.153 15.613 4.047 39  
## 5 36 18 36.48 6.022 18 34.99 5.467 16.092 4.172 41  
## 8 48 24 36.57 6.514 24 35.01 5.757 17.177 4.370 47  
## 6 64 32 36.00 6.409 32 35.59 6.314 17.773 1.129 50  
## 9 80 40 36.00 6.280 40 35.60 6.232 17.474 1.112 48  
## 7 100 50 36.69 7.009 50 34.99 5.961 18.154 4.760 52  
## 10 144 72 35.75 7.033 72 36.10 6.362 18.668 -0.969 55  
## 11 196 98 36.24 7.424 98 35.45 6.360 19.283 2.195 59  
## 12 256 128 36.04 7.281 128 35.53 7.205 20.241 1.422 65  
## 18 324 162 36.30 7.360 162 35.50 7.643 20.900 2.234 69  
## 17 400 200 36.18 8.147 200 35.55 7.175 21.404 1.754 72  
## 16 484 242 35.88 8.225 242 35.94 7.038 21.315 -0.145 72  
## 15 576 288 35.98 7.966 288 35.67 8.168 22.518 0.854 80  
## 14 676 338 35.91 8.246 338 35.88 8.240 22.964 0.070 83  
## 13 900 450 36.36 8.510 450 35.91 8.165 23.077 1.262 84  
## 3 1600 800 36.02 9.118 800 35.88 9.108 25.348 0.395 101

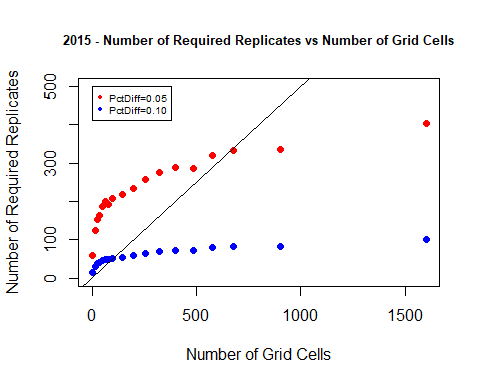
# Plot Required Replicates vs Coefficient of Variation  
  
plot(rr ~ CV, data=allGrid.final\_05, pch=16, col="red", ylim=c(0,1800),   
 main="2015 - Required Replicates vs Coefficient of Variation (CV)", cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Required Replicates")  
points(rr ~ CV, data=allGrid.final\_10, pch=16, col="blue")  
points(rr ~ CV, data=allGrid.final\_025, pch=16, col="green")  
legend(10, 1500, legend=c("PctDiff=2.5%", "PctDiff=5%","PctDiff=10%"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



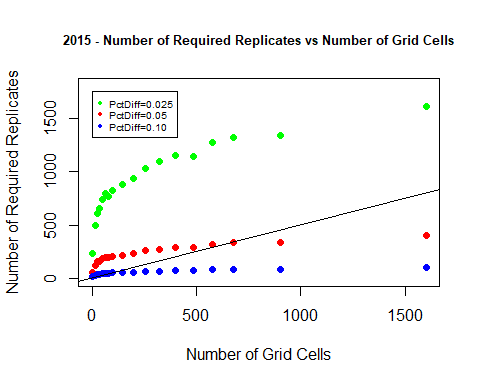
# Plot Number of Grid Cells vs Coefficient of Variation  
  
plot(GridCells ~ CV, data=allGrid.final\_05, pch=16, col="red",  
 main="2015 - Number of Grid Cells vs Coefficient of Variation (CV)", cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Number of Grid Cells")



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,500),  
 main="2015 - Number of Required Replicates vs Number of Grid Cells", cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(0, 500, legend=c("PctDiff=0.05","PctDiff=0.10"),  
 col=c("red", "blue"), pch=c(16,16), cex=0.6)



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10 and .025)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,1800),  
 main="2015 - Number of Required Replicates vs Number of Grid Cells", cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
points(rr ~ GridCells, data=allGrid.final\_025, pch=16, col="green", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(0, 1750, legend=c("PctDiff=0.025", "PctDiff=0.05","PctDiff=0.10"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



# 

# Analysis of 2015 Field Results

The overall mean yield count is near 36 among my treatment divisions. The standard deviations of cell yield means increases as we further divide the field up into smaller and smaller grid cells, from a mean of about 3 for four grid cells up to 9 for 1,600 grid cells. As we divided the field up into grid cells, once we reached 256 (25 meters by 37.5 meters) that was the highest number of grid cells where there were still 30 or more yields per cell. Therefore, we really shouldn’t look at dividing the field into cells smaller than 25x37.5 meters.

The plot of CV versus Required Replicates shows that the number of required replicates increases as the CV increases and it increases more for lower Percent Differences between the two treatments (2.5% vs 5% vs 10%).

The plot of CV versus Number of Grid Cells shows that the CV increases as we increase the number of grid cells, or we divide the field up into smaller and smaller cells.

I have two plots of Number of Grid Cells versus Number of Required Replicates. The first shows the data with separate trends using Percent Difference equal to 5% and 10% while the second plot adds the trend for 2.5%. The plots show the relationship between required replicates and number of grid cells where as the number of grid cells increases the number of required replicates increases but the number of replicates needed levels off and we look for where the number of grid cells is twice the number of required replicates. I added the trend line to show where the number of grid cells is twice that of the number of required replicates.

Now, looking at the results table and the plots, if we seek a Percent Difference between the two treatment means of 5% then it looks like we should use 676 grid cells as that gives us 332 required replicates. But, at 676 grid cells I noticed that only 1.2% of the cells had 30 or more yield samples. So, I will need to make an adjustment. The number of grid cells that gave us a minimum of 30 yield samples was 256. I will use that information in the next R code chunk where I devleop a new required replicates function and find what is the lowest Percent Difference that will work with 256 grid cells.

For a Percent Difference of 10%, the difference requires fewer replicates to detect this difference. It looks like 100 grid cells, dividing the field up by 60x40 meter sections, as that will give us twice the number of replicates needed, which is 52.

For a Percent Difference of 2.5%, the smaller difference requires many more replicates to detect a difference. Based upon the 2015 dataset there are not enough grid cells available to divide the field into in order to detect a 2.5% percent differnece between treatment means.

# Function to Determine Minimum Percent Difference for Required Replicates

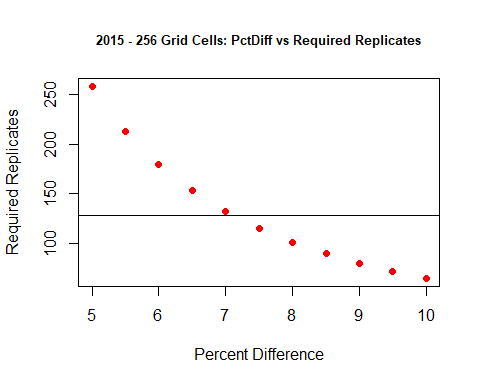
From the above analysis we saw that for a 5% Percent Difference the number of grid cells needed was too high as most cells had fewer than 30 yield samples. Therefore, using the lowest number of grid cells in my analysis that had 30 or more yield samples in all cells, 256, I now can calculate the lowest Percent Difference we can use so that the number of grid cells is twice that of the required replicates.

The below function, ‘**required.replicates2**’, is a slightly modified version of the original ‘**required.replicates**’ function. It only returns the value of required replicates that are calculated. This code chunk will use a loop to calculate the number of required replicates, using the data from the summary table above for Grid Cells = 256 and for PctDiff ranging from 5.0% to 10.0% in increments of 0.5%. A dataframe returning the percent difference and required replicates is created and the relationship between the two is plotted. A horizontal line showing the point at which required replicates are half of the 256 grid cells is plotted. Where the points cross this horizontal line will be a good Percent Difference to use for when dividing the field into 256 grid cells.

# Find Minimum Percent Difference to use with 256 Grid Cells  
  
required.replicates2 <- function(N1,N2,M1,M2,S1,S2, alpha=0.05, beta=0.20, PctDiff\_fix=5) {  
CV <- 100\*(sqrt( (S1^2 + S2^2) / 2 ) / ((M1 + M2)/2))  
PctDiff <- 100\*((M1 - M2) / ((M1 + M2)/2))   
return(rr=ceiling(2 \* ((CV / PctDiff\_fix)^2) \* ((qnorm(1-alpha/2,0,1)+qnorm(1-beta,0,1))^2)))  
}  
  
akeep <- vector()  
a <- 5  
rrkeep <- vector()  
k <- vector()  
  
for (i in 1:11) {   
k[i] <- i  
rr <- required.replicates2(128, 128, 36.04, 35.53, 7.28, 7.21, alpha=0.05, beta=0.20, PctDiff\_fix=a)  
rrkeep[i] <- rr  
akeep[i] <- a  
a <- a + 0.5  
  
}  
  
find.min <- data.frame(akeep,rrkeep)  
colnames(find.min) = c("PctDiff", "ReqRep")   
find.min

## PctDiff ReqRep  
## 1 5.0 258  
## 2 5.5 213  
## 3 6.0 179  
## 4 6.5 153  
## 5 7.0 132  
## 6 7.5 115  
## 7 8.0 101  
## 8 8.5 90  
## 9 9.0 80  
## 10 9.5 72  
## 11 10.0 65

plot(ReqRep ~ PctDiff, data=find.min, pch=16, col="red",  
 main="2015 - 256 Grid Cells: PctDiff vs Required Replicates", cex.main=0.8,  
 xlab="Percent Difference",  
 ylab="Required Replicates")  
abline(h = 256/2)



# Finding PctDiff for Grid Cells = 256

From the plot I see that the Percent Difference is at about 7% for when the number of grid cells (256) is about double the number of required replicates (132). Thus, a 7% Percent Difference between treatment means is what we can use when dividing the field into 256 grid cells, so that every cell has a minimum of 30 yield samples. This is the result for the 2015 dataset. The result could differ when looking at the 2016 and 2017 datasets. I will look at those datasets next.

# Introduce the 2016 dataset

Here, I read in the dataset for the 2016 season and print out a summary of the dataset along with the first and last six records to inspect what is in the dataset. I also check for missing values among the variables. Lastly, I plot the points of latitude and longitude to get a sense of what the field looks like.

# Read in the data for 2016  
  
Home2016Path = "../STAT600/home.2016.csv"  
home.2016.dat <- read.csv(Home2016Path,header=TRUE)  
  
# Print a summary, first six and last six observations  
summary(home.2016.dat)

## X Yield Latitude Longitude   
## Min. : 1697 Min. : 3.76 Min. : 0.0 Min. : 0   
## 1st Qu.: 9242 1st Qu.: 99.77 1st Qu.: 97.1 1st Qu.:152   
## Median :12289 Median :123.27 Median :194.4 Median :316   
## Mean :12103 Mean :117.97 Mean :196.5 Mean :307   
## 3rd Qu.:15089 3rd Qu.:139.71 3rd Qu.:295.5 3rd Qu.:468   
## Max. :17697 Max. :278.75 Max. :400.0 Max. :600

head(home.2016.dat)

## X Yield Latitude Longitude  
## 1 1697 93.43 139.9 600.0  
## 2 1698 85.55 143.7 599.9  
## 3 1699 91.67 146.7 599.9  
## 4 1700 101.74 149.6 599.9  
## 5 1701 111.19 152.6 599.9  
## 6 1702 103.15 155.5 599.9

tail(home.2016.dat)

## X Yield Latitude Longitude  
## 8408 17692 122.7 388.6 589.6  
## 8409 17693 139.3 390.9 589.5  
## 8410 17694 161.7 393.2 589.5  
## 8411 17695 140.9 395.4 589.5  
## 8412 17696 136.5 397.7 589.5  
## 8413 17697 155.1 399.9 589.5

which(is.na(home.2016.dat$X))

## integer(0)

which(is.na(home.2016.dat$Yield))

## integer(0)

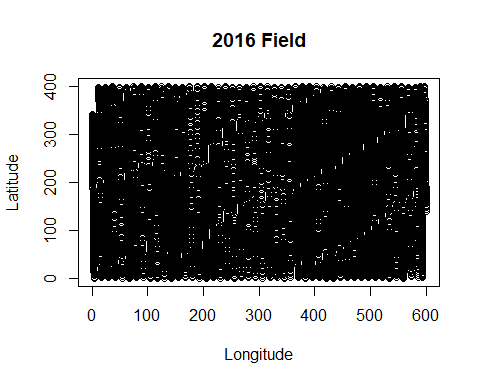
which(is.na(home.2016.dat$Latitude))

## integer(0)

which(is.na(home.2016.dat$Longitude))

## integer(0)

# Plot the data with x=Longitude and y=Latitude  
  
plot(home.2016.dat$Longitude, home.2016.dat$Latitude, type="p", main="2016 Field", xlab="Longitude", ylab="Latitude")



###############################################################  
# 2016 Analysis  
# 1 - Stack all 17 summary files for varying number of grid cells together  
# 2 - Run Required Replicates 3 times for different values of PctDiff  
# 3 - Print out resulting tables showing actual PctDiff, CV and Required Replicates for each number of Grid Cells  
# 4 - Print Plots  
##############################################################  
  
# Stack all files together  
  
allGrid.dat = rbind(Grid1.compare.dat, Grid2.compare.dat,Grid3.compare.dat,Grid4.compare.dat,Grid5.compare.dat,Grid6.compare.dat, Grid7.compare.dat,Grid8.compare.dat,Grid9.compare.dat,Grid10.compare.dat,Grid11.compare.dat,Grid12.compare.dat,Grid13.compare.dat,Grid14.compare.dat,Grid15.compare.dat,Grid16.compare.dat,Grid17.compare.dat,Grid18.compare.dat)  
  
  
# Run Required Replicates function 3 times-once each for PctDiff set to 2.5%, 5% and 10%  
  
allGrid.final\_025 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=2.5))  
  
allGrid.final\_05 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=5))  
  
allGrid.final\_10 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=10))  
  
# Print out summary tables for PctDiff = 2.5%  
allGrid.final\_025[order(allGrid.final\_025$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 122.8 8.18 2 112.3 11.254 8.371 8.933 176  
## 4 16 8 120.1 13.04 8 114.9 8.806 9.464 4.436 225  
## 2 24 12 120.3 13.83 12 114.9 10.984 10.616 4.548 284  
## 5 36 18 119.3 14.96 18 116.1 13.601 12.152 2.708 371  
## 8 48 24 119.0 16.85 24 115.9 14.814 13.507 2.587 459  
## 6 64 32 116.1 19.17 32 118.6 14.619 14.525 -2.093 530  
## 9 80 40 116.4 20.55 40 118.7 15.660 15.545 -2.010 607  
## 7 100 50 119.3 19.35 50 115.7 18.832 16.245 3.071 663  
## 10 144 72 119.4 18.66 72 115.4 21.433 17.119 3.425 737  
## 11 196 98 117.3 22.61 98 118.7 19.193 17.772 -1.137 794  
## 12 256 128 117.2 24.79 128 117.1 21.720 19.898 0.045 995  
## 18 324 162 116.7 24.48 162 118.6 23.154 20.250 -1.645 1030  
## 17 400 200 117.9 25.31 200 117.8 24.815 21.265 0.148 1136  
## 16 484 242 117.8 24.45 242 117.0 27.388 22.108 0.727 1228  
## 15 576 288 116.5 27.20 288 117.6 25.909 22.694 -0.995 1294  
## 14 676 338 117.4 27.61 338 117.2 26.446 23.050 0.185 1335  
## 13 900 450 116.4 28.61 450 116.8 29.363 24.856 -0.346 1552  
## 3 1600 800 117.6 30.98 800 117.0 29.976 25.991 0.540 1697

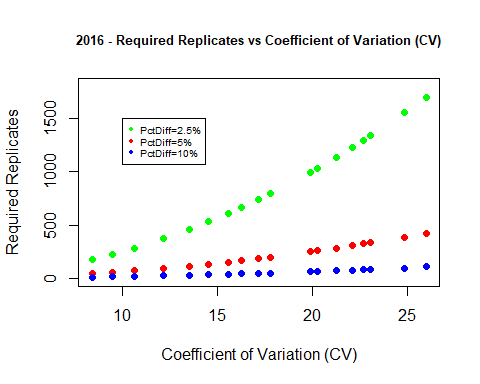
# Print out summary tables for PctDiff = 5%  
allGrid.final\_05[order(allGrid.final\_05$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 122.8 8.18 2 112.3 11.254 8.371 8.933 44  
## 4 16 8 120.1 13.04 8 114.9 8.806 9.464 4.436 57  
## 2 24 12 120.3 13.83 12 114.9 10.984 10.616 4.548 71  
## 5 36 18 119.3 14.96 18 116.1 13.601 12.152 2.708 93  
## 8 48 24 119.0 16.85 24 115.9 14.814 13.507 2.587 115  
## 6 64 32 116.1 19.17 32 118.6 14.619 14.525 -2.093 133  
## 9 80 40 116.4 20.55 40 118.7 15.660 15.545 -2.010 152  
## 7 100 50 119.3 19.35 50 115.7 18.832 16.245 3.071 166  
## 10 144 72 119.4 18.66 72 115.4 21.433 17.119 3.425 185  
## 11 196 98 117.3 22.61 98 118.7 19.193 17.772 -1.137 199  
## 12 256 128 117.2 24.79 128 117.1 21.720 19.898 0.045 249  
## 18 324 162 116.7 24.48 162 118.6 23.154 20.250 -1.645 258  
## 17 400 200 117.9 25.31 200 117.8 24.815 21.265 0.148 284  
## 16 484 242 117.8 24.45 242 117.0 27.388 22.108 0.727 307  
## 15 576 288 116.5 27.20 288 117.6 25.909 22.694 -0.995 324  
## 14 676 338 117.4 27.61 338 117.2 26.446 23.050 0.185 334  
## 13 900 450 116.4 28.61 450 116.8 29.363 24.856 -0.346 388  
## 3 1600 800 117.6 30.98 800 117.0 29.976 25.991 0.540 425

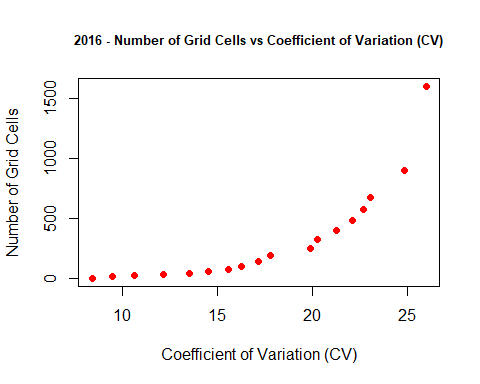
# Print out summary tables for PctDiff = 10%  
allGrid.final\_10[order(allGrid.final\_10$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 122.8 8.18 2 112.3 11.254 8.371 8.933 11  
## 4 16 8 120.1 13.04 8 114.9 8.806 9.464 4.436 15  
## 2 24 12 120.3 13.83 12 114.9 10.984 10.616 4.548 18  
## 5 36 18 119.3 14.96 18 116.1 13.601 12.152 2.708 24  
## 8 48 24 119.0 16.85 24 115.9 14.814 13.507 2.587 29  
## 6 64 32 116.1 19.17 32 118.6 14.619 14.525 -2.093 34  
## 9 80 40 116.4 20.55 40 118.7 15.660 15.545 -2.010 38  
## 7 100 50 119.3 19.35 50 115.7 18.832 16.245 3.071 42  
## 10 144 72 119.4 18.66 72 115.4 21.433 17.119 3.425 47  
## 11 196 98 117.3 22.61 98 118.7 19.193 17.772 -1.137 50  
## 12 256 128 117.2 24.79 128 117.1 21.720 19.898 0.045 63  
## 18 324 162 116.7 24.48 162 118.6 23.154 20.250 -1.645 65  
## 17 400 200 117.9 25.31 200 117.8 24.815 21.265 0.148 71  
## 16 484 242 117.8 24.45 242 117.0 27.388 22.108 0.727 77  
## 15 576 288 116.5 27.20 288 117.6 25.909 22.694 -0.995 81  
## 14 676 338 117.4 27.61 338 117.2 26.446 23.050 0.185 84  
## 13 900 450 116.4 28.61 450 116.8 29.363 24.856 -0.346 97  
## 3 1600 800 117.6 30.98 800 117.0 29.976 25.991 0.540 107

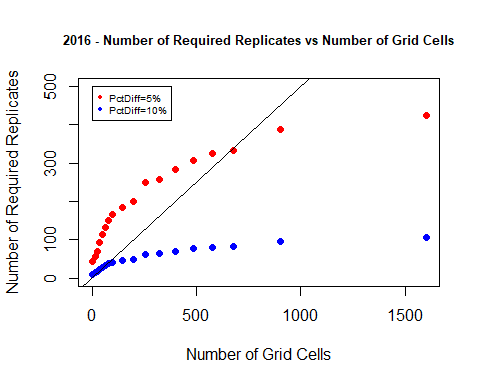
# Plot Required Replicates vs Coefficient of Variation  
  
plot(rr ~ CV, data=allGrid.final\_05, pch=16, col="red", ylim=c(0,1800),   
 main="2016 - Required Replicates vs Coefficient of Variation (CV)",cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Required Replicates")  
points(rr ~ CV, data=allGrid.final\_10, pch=16, col="blue")  
points(rr ~ CV, data=allGrid.final\_025, pch=16, col="green")  
legend(10, 1500, legend=c("PctDiff=2.5%", "PctDiff=5%","PctDiff=10%"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



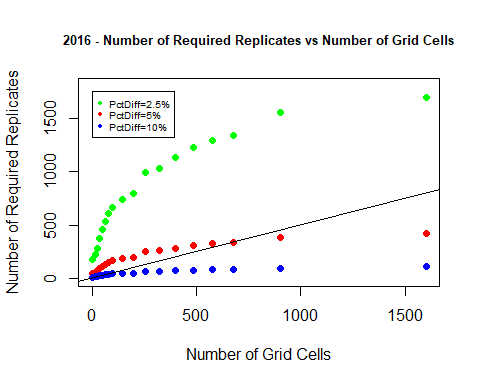
# Plot Number of Grid Cells vs Coefficient of Variation  
  
plot(GridCells ~ CV, data=allGrid.final\_05, pch=16, col="red",  
 main="2016 - Number of Grid Cells vs Coefficient of Variation (CV)",cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Number of Grid Cells")



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,500),  
 main="2016 - Number of Required Replicates vs Number of Grid Cells",cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(0, 500, legend=c("PctDiff=5%","PctDiff=10%"),  
 col=c("red", "blue"), pch=c(16,16), cex=0.6)



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10 and .025)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,1800),  
 main="2016 - Number of Required Replicates vs Number of Grid Cells",cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
points(rr ~ GridCells, data=allGrid.final\_025, pch=16, col="green", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(0, 1750, legend=c("PctDiff=2.5%", "PctDiff=5%","PctDiff=10%"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



# Analysis of 2016 Field Results

The overall mean yield count is around 118 among my treatment divisions, a large increase over the 2015 dataset’s number. The standard deviations of cell yield means increased as we further divide the field up into smaller and smaller grid cells, from about 8 for four cell grids up to 30 for 1,600 grid cells, again larger values when compared to 2015. As we divided the field up into grid cells once we reached 144 (33.3 meters by 50 meters) that was the highest number of grid cells where there were still 30 or more yields per cell. Therefore, we really shouldn’t look at dividing the field into grid cells smaller than 33.3x50 meters.

The four plots for 2016 look very similar to the plots for 2015 in terms of relationships and trends of the variables being plotted.

Now, looking at the final results table and the plots, if we seek a Percent Difference between the two treatment means of 5% then it looks like we should once again use 676 grid cells as that gives us 334 required replicates. But when we divided up the field, I noticed that 0% of the cells had 30 or more yield samples. So, I will need to make an adjustment. The number of grid cells that gave us a minimum of 30 yield samples was 144. I will use that information in the next R code chunk where I find what is the lowest Percent Difference that will work with 144 grid cells similar to what I did with the 2015 dataset for 256 grid cells..

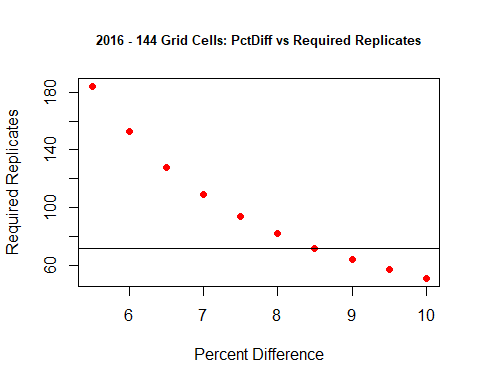
For a Percent Difference of 10%, the difference requires fewer replicates to detect this difference. It looks like 80 grid cells, dividing the field up by 60x50 meter sections will require us to need 38 replicates.

For a Percent Difference of 2.5%, the smaller difference requires many more replicates to detect a difference. Based upon the 2016 dataset there are not enough grid cells available to divide the field into in order to detect a 2.5% percent difference between treatment means.

# Find Minimum Percent Difference to use with 144 Grid Cells  
  
required.replicates2 <- function(N1,N2,M1,M2,S1,S2, alpha=0.05, beta=0.20, PctDiff\_fix=0.05) {  
CV <- 100\*(sqrt( (S1^2 + S2^2) / 2 ) / ((M1 + M2)/2))  
PctDiff <- 100\*((M1 - M2) / ((M1 + M2)/2) )  
return(rr=ceiling(2 \* ((CV / PctDiff\_fix)^2) \* ((qnorm(1-alpha/2,0,1)+qnorm(1-beta,0,1))^2)))  
}  
  
akeep <- vector()  
a <- 5  
rrkeep <- vector()  
k <- vector()  
  
for (i in 1:10) {   
k[i] <- i  
rr <- required.replicates2(72, 72, 119.3964, 115.3747, 18.66, 21.43, alpha=0.05, beta=0.20, PctDiff\_fix=a)  
rrkeep[i] <- rr  
a <- a + 0.5  
akeep[i] <- a  
}  
  
find.min <- data.frame(akeep,rrkeep)  
colnames(find.min) = c("PctDiff", "ReqRep")   
find.min

## PctDiff ReqRep  
## 1 5.5 184  
## 2 6.0 153  
## 3 6.5 128  
## 4 7.0 109  
## 5 7.5 94  
## 6 8.0 82  
## 7 8.5 72  
## 8 9.0 64  
## 9 9.5 57  
## 10 10.0 51

plot(ReqRep ~ PctDiff, data=find.min, pch=16, col="red",  
 main="2016 - 144 Grid Cells: PctDiff vs Required Replicates", cex.main=0.8,  
 xlab="Percent Difference",  
 ylab="Required Replicates")  
abline(h = 144/2)



# Finding PctDiff for Grid Cells = 144

From the plot I see that the Percent Difference is at about 8.5% for when the number of grid cells (144) is about double the number of required replicates (72). Thus, an 8.5% Percent Difference between treatment means is what we can use when dividing the field into 144 grid cells, so that every cell has a minimum of 30 yield samples. This is the result for the 2016 dataset. I will still look at the 2017 dataset next to see what it has to say.

# 

# Introduce the 2017 dataset

Here, I read in the dataset for the 2017 season and print out a summary of the dataset along with the first and last six records to inspect what is in the dataset. I also check for missing values among the variables. Lastly, I plot the points of latitude and longitude to get a sense of what the field looks like.

# Read in the data for 2017  
  
Home2017Path = "../STAT600/home.2017.csv"  
home.2017.dat <- read.csv(Home2017Path,header=TRUE)  
  
# Print a summary, first six and last six observations  
summary(home.2017.dat)

## X Yield Latitude Longitude   
## Min. : 1217 Min. :14.2 Min. : 0.0 Min. : 0   
## 1st Qu.: 5758 1st Qu.:56.0 1st Qu.: 98.7 1st Qu.:152   
## Median :12076 Median :58.8 Median :198.3 Median :299   
## Mean :10787 Mean :58.5 Mean :198.6 Mean :299   
## 3rd Qu.:15492 3rd Qu.:61.5 3rd Qu.:297.4 3rd Qu.:446   
## Max. :18507 Max. :82.0 Max. :400.0 Max. :600

head(home.2017.dat)

## X Yield Latitude Longitude  
## 1 1217 58.69 1.392 256.9  
## 2 1218 58.25 3.152 256.7  
## 3 1219 65.82 4.797 256.5  
## 4 1220 61.13 6.445 256.3  
## 5 1221 58.83 8.111 256.1  
## 6 1222 61.89 9.775 256.0

tail(home.2017.dat)

## X Yield Latitude Longitude  
## 9574 18502 58.14 373.7 599.0  
## 9575 18503 59.19 371.8 599.1  
## 9576 18504 59.04 370.0 599.3  
## 9577 18505 52.75 368.2 599.5  
## 9578 18506 54.88 366.4 599.6  
## 9579 18507 57.09 364.6 599.8

which(is.na(home.2017.dat$X))

## integer(0)

which(is.na(home.2017.dat$Yield))

## integer(0)

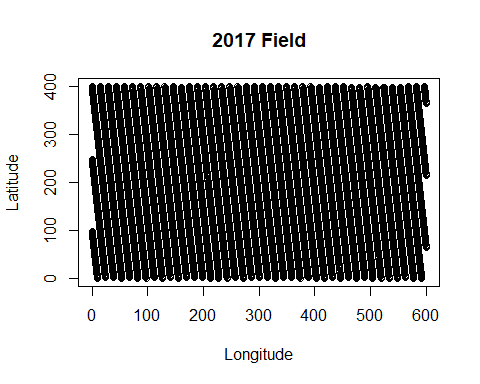
which(is.na(home.2017.dat$Latitude))

## integer(0)

which(is.na(home.2017.dat$Longitude))

## integer(0)

# Plot the data with x=Longitude and y=Latitude  
  
plot(home.2017.dat$Longitude, home.2017.dat$Latitude, type="p",main="2017 Field", xlab="Longitude", ylab="Latitude")



########################################################################  
# 2017 Analysis  
# 1 - Stack all 17 summary files for varying number of grid cells together  
# 2 - Run Required Replicates 3 times for different values of PctDiff  
# 3 - Print out resulting tables showing actual PctDiff, CV and Required Replicates for each number of Grid Cells  
# 4 - Print Plots  
#######################################################################  
  
# Stack all files together  
  
allGrid.dat = rbind(Grid1.compare.dat, Grid2.compare.dat,Grid3.compare.dat,Grid4.compare.dat,Grid5.compare.dat,Grid6.compare.dat, Grid7.compare.dat,Grid8.compare.dat,Grid9.compare.dat,Grid10.compare.dat,Grid11.compare.dat,Grid12.compare.dat,Grid13.compare.dat,Grid14.compare.dat,Grid15.compare.dat,Grid16.compare.dat,Grid17.compare.dat,Grid18.compare.dat)  
  
# Run Required Replicates function 3 times - once each for PctDiff set to 0.025, 0.05 and 0.10 (2.5%, 5% and 10%)  
  
allGrid.final\_025 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=2.5))  
  
allGrid.final\_05 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=5))  
  
allGrid.final\_10 <- data.frame(required.replicates(allGrid.dat$N1,   
 allGrid.dat$N2,  
 allGrid.dat$M1,  
 allGrid.dat$M2,   
 allGrid.dat$S1,   
 allGrid.dat$S2,  
 alpha=0.05, beta=0.20, PctDiff\_fix=10))  
  
# Print out summary tables for PctDiff = 2.5%, 5% and 10%  
  
allGrid.final\_025[order(allGrid.final\_025$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 58.52 0.723 2 58.50 1.128 1.619 0.027 7  
## 4 16 8 58.70 0.967 8 58.29 1.364 2.021 0.713 11  
## 2 24 12 58.70 1.038 12 58.29 1.501 2.206 0.696 13  
## 5 36 18 58.28 1.346 18 58.72 1.447 2.389 -0.749 15  
## 8 48 24 58.25 1.589 24 58.74 1.506 2.646 -0.826 18  
## 6 64 32 58.28 1.843 32 58.72 1.494 2.868 -0.761 21  
## 9 80 40 58.27 1.994 40 58.72 1.529 3.037 -0.759 24  
## 7 100 50 58.65 1.681 50 58.34 1.937 3.100 0.521 25  
## 10 144 72 58.44 1.539 72 58.58 2.209 3.254 -0.231 27  
## 11 196 98 58.48 1.876 98 58.54 2.149 3.448 -0.099 30  
## 12 256 128 58.50 2.242 128 58.52 2.082 3.698 -0.027 35  
## 18 324 162 58.40 2.340 162 58.68 2.251 3.922 -0.482 39  
## 17 400 200 58.53 2.298 200 58.47 2.329 3.955 0.097 40  
## 16 484 242 58.62 2.556 242 58.40 2.720 4.511 0.369 52  
## 15 576 288 58.41 2.757 288 58.50 2.760 4.719 -0.157 56  
## 14 676 338 58.46 3.026 338 58.51 2.760 4.952 -0.079 62  
## 13 900 450 58.54 3.066 450 58.43 3.271 5.421 0.183 74  
## 3 1600 800 58.47 3.642 800 58.48 3.831 6.392 -0.029 103

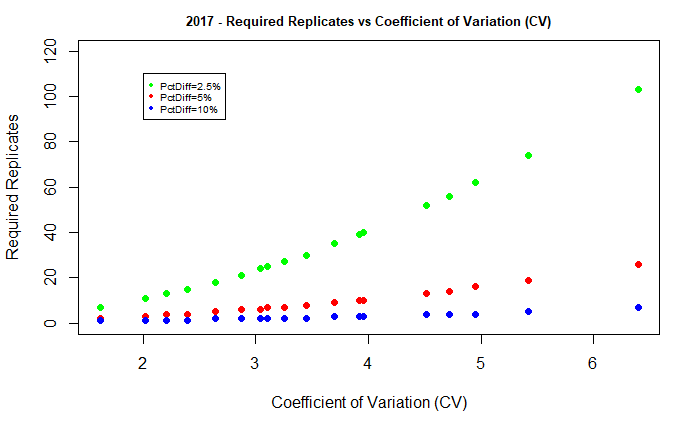
allGrid.final\_05[order(allGrid.final\_05$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 58.52 0.723 2 58.50 1.128 1.619 0.027 2  
## 4 16 8 58.70 0.967 8 58.29 1.364 2.021 0.713 3  
## 2 24 12 58.70 1.038 12 58.29 1.501 2.206 0.696 4  
## 5 36 18 58.28 1.346 18 58.72 1.447 2.389 -0.749 4  
## 8 48 24 58.25 1.589 24 58.74 1.506 2.646 -0.826 5  
## 6 64 32 58.28 1.843 32 58.72 1.494 2.868 -0.761 6  
## 9 80 40 58.27 1.994 40 58.72 1.529 3.037 -0.759 6  
## 7 100 50 58.65 1.681 50 58.34 1.937 3.100 0.521 7  
## 10 144 72 58.44 1.539 72 58.58 2.209 3.254 -0.231 7  
## 11 196 98 58.48 1.876 98 58.54 2.149 3.448 -0.099 8  
## 12 256 128 58.50 2.242 128 58.52 2.082 3.698 -0.027 9  
## 18 324 162 58.40 2.340 162 58.68 2.251 3.922 -0.482 10  
## 17 400 200 58.53 2.298 200 58.47 2.329 3.955 0.097 10  
## 16 484 242 58.62 2.556 242 58.40 2.720 4.511 0.369 13  
## 15 576 288 58.41 2.757 288 58.50 2.760 4.719 -0.157 14  
## 14 676 338 58.46 3.026 338 58.51 2.760 4.952 -0.079 16  
## 13 900 450 58.54 3.066 450 58.43 3.271 5.421 0.183 19  
## 3 1600 800 58.47 3.642 800 58.48 3.831 6.392 -0.029 26

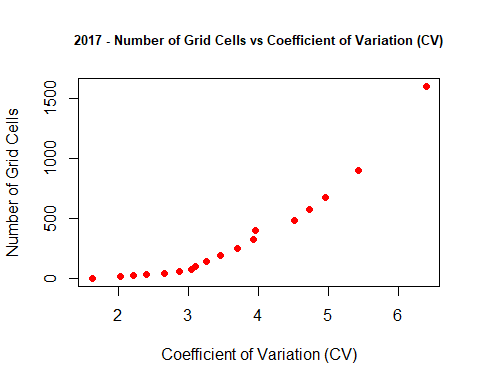
allGrid.final\_10[order(allGrid.final\_10$GridCells),]

## GridCells N1 M1 S1 N2 M2 S2 CV PctDiff rr  
## 1 4 2 58.52 0.723 2 58.50 1.128 1.619 0.027 1  
## 4 16 8 58.70 0.967 8 58.29 1.364 2.021 0.713 1  
## 2 24 12 58.70 1.038 12 58.29 1.501 2.206 0.696 1  
## 5 36 18 58.28 1.346 18 58.72 1.447 2.389 -0.749 1  
## 8 48 24 58.25 1.589 24 58.74 1.506 2.646 -0.826 2  
## 6 64 32 58.28 1.843 32 58.72 1.494 2.868 -0.761 2  
## 9 80 40 58.27 1.994 40 58.72 1.529 3.037 -0.759 2  
## 7 100 50 58.65 1.681 50 58.34 1.937 3.100 0.521 2  
## 10 144 72 58.44 1.539 72 58.58 2.209 3.254 -0.231 2  
## 11 196 98 58.48 1.876 98 58.54 2.149 3.448 -0.099 2  
## 12 256 128 58.50 2.242 128 58.52 2.082 3.698 -0.027 3  
## 18 324 162 58.40 2.340 162 58.68 2.251 3.922 -0.482 3  
## 17 400 200 58.53 2.298 200 58.47 2.329 3.955 0.097 3  
## 16 484 242 58.62 2.556 242 58.40 2.720 4.511 0.369 4  
## 15 576 288 58.41 2.757 288 58.50 2.760 4.719 -0.157 4  
## 14 676 338 58.46 3.026 338 58.51 2.760 4.952 -0.079 4  
## 13 900 450 58.54 3.066 450 58.43 3.271 5.421 0.183 5  
## 3 1600 800 58.47 3.642 800 58.48 3.831 6.392 -0.029 7

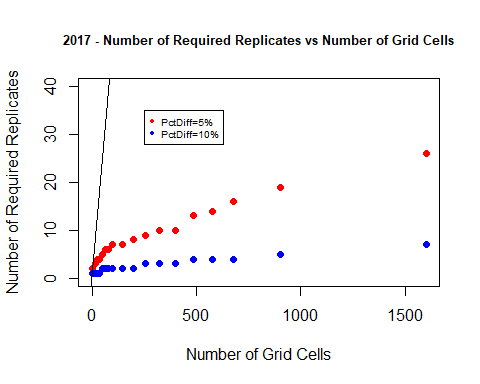
# Plot Required Replicates vs Coefficient of Variation  
  
plot(rr ~ CV, data=allGrid.final\_05, pch=16, col="red", ylim=c(0,120),   
 main="2017 - Required Replicates vs Coefficient of Variation (CV)",cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Required Replicates")  
points(rr ~ CV, data=allGrid.final\_10, pch=16, col="blue")  
points(rr ~ CV, data=allGrid.final\_025, pch=16, col="green")  
legend(0.02, 110, legend=c("PctDiff=2.5%", "PctDiff=5%","PctDiff=10%"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



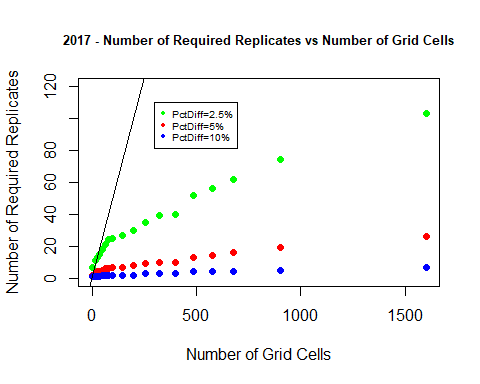
# Plot Number of Grid Cells vs Coefficient of Variation  
  
plot(GridCells ~ CV, data=allGrid.final\_05, pch=16, col="red",  
 main="2017 - Number of Grid Cells vs Coefficient of Variation (CV)",cex.main=0.8,  
 xlab="Coefficient of Variation (CV)",   
 ylab="Number of Grid Cells")



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,40),  
 main="2017 - Number of Required Replicates vs Number of Grid Cells",cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(250, 35, legend=c("PctDiff=5%","PctDiff=10%"),  
 col=c("red", "blue"), pch=c(16,16), cex=0.6)



# Plot Number of Required Replicates vs Number of Grid Cells (for PctDiff = .05 and .10 and .025)  
  
plot(rr ~ GridCells, data=allGrid.final\_05, pch=16, col="red", xlim=(c(0,max(GridCells))), ylim=c(0,120),  
 main="2017 - Number of Required Replicates vs Number of Grid Cells", cex.main=0.8,  
 xlab="Number of Grid Cells",  
 ylab="Number of Required Replicates")  
points(rr ~ GridCells, data=allGrid.final\_10, pch=16, col="blue", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
points(rr ~ GridCells, data=allGrid.final\_025, pch=16, col="green", xlim=(c(0,max(GridCells))), ylim=c(0,max(rr)))  
abline(coef = c(0,0.5))  
legend(300, 110, legend=c("PctDiff=2.5%", "PctDiff=5%","PctDiff=10%"),  
 col=c("green", "red", "blue"), pch=c(16,16,16), cex=0.6)



# Analysis of 2017 Field Results

The overall mean yield count is around 58.5 among my treatment divisions, a decrease from 2016 but slightly higher than 2015. The standard deviations of cell yield means increases as we further divide the field up into smaller and smaller grid cells, from about 1 for four grid cells up to 3.7 for 1,600 grid cells. There is very little variation in this field when compared to the two previous years. As we divided the field up into grid cells, once we reached 196 (42.857 meters by 28.571 meters) that was the highest number of grid cells where there were still 30 or more yields per cell. Therefore, we really shouldn’t look at dividing the field into grid cells smaller than 42.857 meters by 28.571 meters.

The four plots for 2017 look very similar to the plots for the other two years in terms of relationships and trends of the variables being plotted.

Now, looking at the results table and the plots, if we seek a Percent Difference between the two treatment means of 5% then it looks like we could use a field with very few grid cells. But we also must keep in mind that we require a minimum of 12 error degrees of freedom. Therefore, I would use 16 grid cells.

For a Percent Difference of 10%, the answer would be the same, 16 grid cells, so that we would have a minimum of 12 error degrees of freedom available.

And this time, thanks to the low variance among the field, for a 2.5% Percent Difference between treatments, we can find this using 36 grid cells.

# Final Comments

The 2017 field dataset displayed a ‘best’ case scenario of simulating the treatments across the field. There was very little variation across the field so any difference between the treatments could easily be captured with little effort. For planning an experiment, I would rather use the results of 2015 and 2016 in order to be safe.

In looking at the 2015 and 2016 datasets, the 2015 had lower yields and lower variation, while 2016 had higher yields and higher variation. I feel safer in knowing that for 2015 I had to split the field up into 256 grid cells to maintain that all cells had a minimum of 30 yield samples. In my analysis for 2016, once I split the field up into 196 grid cells, I had already encountered some fields with less than 30 yield samples. So, I would rather error on the side of caution and split the field up into more grid cells than less.

Therefore, my goal would be to split the field into 256 grid cells and aim for detecting a Percent Difference of 7% between the two Treatment means.