Case Study 2, Part 1

Peter Claussen

10/3/2017

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 3.3.2
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#0072B2", "#D55E00", "#F0E442","#CC79A7","#
```

We have two examples of corn fields planted with multiple varieties. Can we detect differences among varieties?

Processing

The first step is to load each file and process

```
EastQuarter.dat <- read.csv("East Quarter.csv",header=TRUE,comment.char = "#")
head(EastQuarter.dat)
```

```
##   Longitude Latitude        Field                Dataset Product ObjId
## 1 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     1
## 2 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     2
## 3 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     3
## 4 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     4
## 5 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     5
## 6 -96.88461 43.45379 East Quarter East Quarter Corn 2015       A     6
##   Heading Distance Elevation                   Desc Swath Moisture
## 1  269.74    0.070    1479.7 10/10/2015 10:42:59 AM    20        0
## 2  269.74    0.201    1479.7 10/10/2015 10:43:00 AM    20        0
## 3  269.74    0.168    1479.7 10/10/2015 10:43:01 AM    20        0
## 4  269.74    0.135    1479.7 10/10/2015 10:43:02 AM    20        0
## 5  269.74    0.135    1479.8 10/10/2015 10:43:03 AM    20        0
## 6  269.74    0.267    1479.8 10/10/2015 10:43:04 AM    20        0
##   YldMassDry YldVolWet YldVolDry
## 1          0         0         0
## 2          0         0         0
## 3          0         0         0
## 4          0         0         0
## 5          0         0         0
## 6          0         0         0
```

```
North80.dat <- read.csv("North 80.csv",header=TRUE,comment.char = "#")
head(North80.dat)
```

```
##   Longitude Latitude Field         Dataset Product ObjId Heading Distance
## 1 -96.90707 43.44624 North North Corn 2016       A     1  88.220    4.007
## 2 -96.90705 43.44623 North North Corn 2016       A     2  88.660    4.171
## 3 -96.90704 43.44624 North North Corn 2016       A     3  88.600    3.942
```

```
## 4 -96.90702 43.44624 North North Corn 2016       A     4  89.060    4.007
## 5 -96.90700 43.44624 North North Corn 2016       A     5  89.180    4.171
## 6 -96.90702 43.44323 North North Corn 2016       A     6   0.031    3.764
##   Elevation                  Desc Swath Moisture YldMassDry YldVolWet
## 1    1443.4  10/20/2016 6:44:38 PM    20    16.78      21.09     21.54
## 2    1443.4  10/20/2016 6:44:39 PM    20    16.78      20.14     20.57
## 3    1443.5  10/20/2016 6:44:40 PM    20    16.78      19.21     19.62
## 4    1443.5  10/20/2016 6:44:41 PM    20    16.72      17.93     18.30
## 5    1443.5  10/20/2016 6:44:42 PM    20    16.72      20.07     20.48
## 6    1433.6 10/21/2016 10:57:23 AM    20    17.28      13.07     13.43
##   YldVolDry
## 1    204.67
## 2    187.77
## 3    189.55
## 4    174.00
## 5    187.10
## 6    135.00
```

We write a function to standardize each file.

```
metric_from_gps <- function(longitude, latitude, origin=c(0,0)) {
  if(origin[1]==0) {
    origin <- c(min(longitude),min(latitude))
  }
  mid_latitude <- (origin[2] + max(latitude))/2

  easting <- longitude - origin[1]
  northing <- latitude - origin[2]

  m_per_deg_lat = 111132.954 - 559.822 * cos(2.0 * mid_latitude) + 1.175 * cos(4.0 * mid_latitude)
  m_per_deg_lon = (3.14159265359/180 ) * 6367449 * cos(mid_latitude)
  easting <- easting*m_per_deg_lon
  northing <- northing*m_per_deg_lat

  return(list(northing=northing,
              easting=easting,
              origin=origin,
              m_per_deg_lon=m_per_deg_lon,
              m_per_deg_lat=m_per_deg_lat))
}


standardize.field <- function(field.dat,origin=c(0,0)) {

  #we will aggregate those values which can be averaged.
  col.names = names(field.dat)
  drop.columns = c(
    which(col.names=="Field"),
    which(col.names=="Dataset"),
    which(col.names=="Product"),
    which(col.names=="Desc")
  )
  #convert date-time to seconds
  field.dat$DateTime <- as.POSIXct(as.character(field.dat[,1]),
                                   format = "%m/%d/%Y %I:%M:%S %p",tz = "")
```

```
  field.dat$Seconds <- field.dat$DateTime - field.dat$DateTime[1]
  aggregate.dat <- field.dat[,-1]

  #remove trivial values
  field.dat <- subset(field.dat,field.dat$Yield!=0)

  meters <- metric_from_gps(field.dat$Longitude,
                            field.dat$Latitude,
                            origin)
  #convert longitude and latitude to meters
  field.dat$Easting <- meters$easting
  field.dat$Northing <- meters$northing

  #compute a quantile
  field.dat$Rank <- rank(field.dat$Yield)
  field.dat$Quantile <- field.dat$Rank/length(field.dat$Rank)
  field.dat$Percent <- 100*field.dat$Yield/median(field.dat$Yield)
  return(field.dat)
}
```

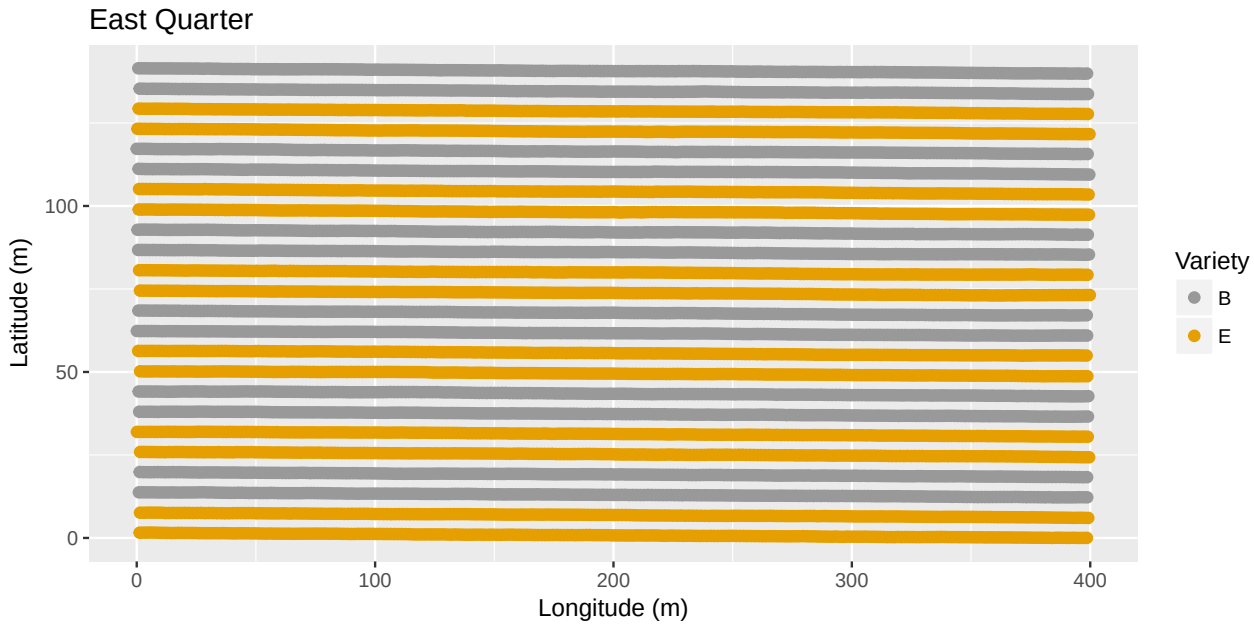
East Quarter

I've already inspected the data to determine which part of the field has strips. We'll truncate to something reasonably square and balanced with regard to variety placement.

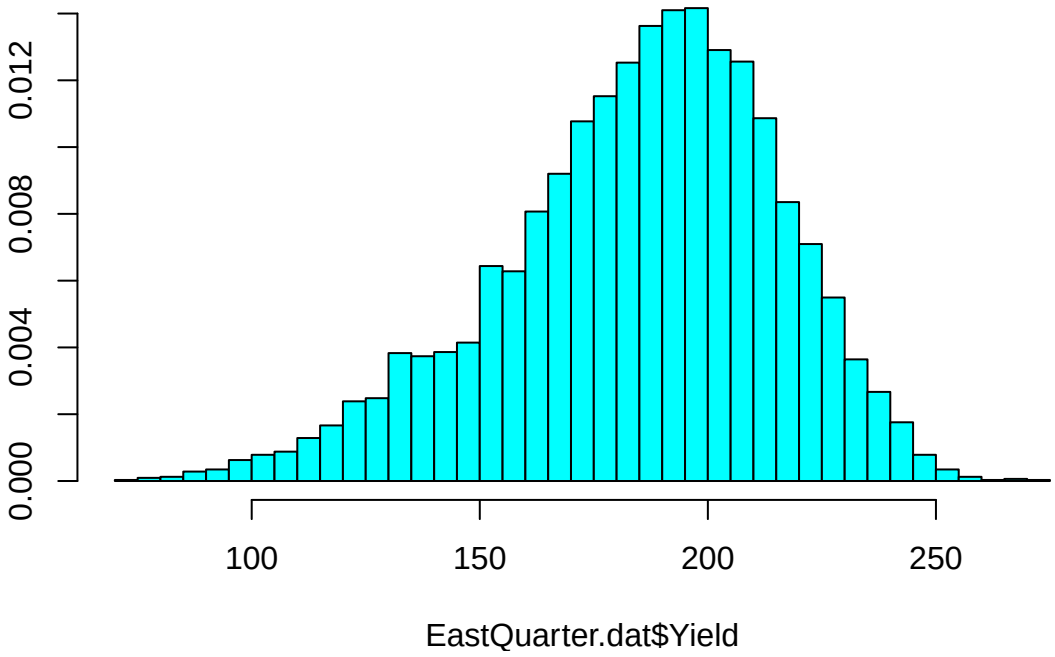
```
EastQuarter.dat$Yield <- EastQuarter.dat$YldVolDry
EastQuarter.dat <- standardize.field(EastQuarter.dat)
EastQuarter.dat$Product <- as.factor(EastQuarter.dat$Product)
```

```
EastQuarter.dat <- subset(EastQuarter.dat,EastQuarter.dat$Easting>=180)
EastQuarter.dat$Easting <- EastQuarter.dat$Easting - min(EastQuarter.dat$Easting)
EastQuarter.dat <- subset(EastQuarter.dat,EastQuarter.dat$Easting<=400)

EastQuarter.dat <- subset(EastQuarter.dat,EastQuarter.dat$Northing>=435)
EastQuarter.dat$Northing <- EastQuarter.dat$Northing - min(EastQuarter.dat$Northing)
EastQuarter.dat <- subset(EastQuarter.dat,EastQuarter.dat$Northing<=145)

EastQuarter.dat$Product <- as.factor(as.character(EastQuarter.dat$Product))
```

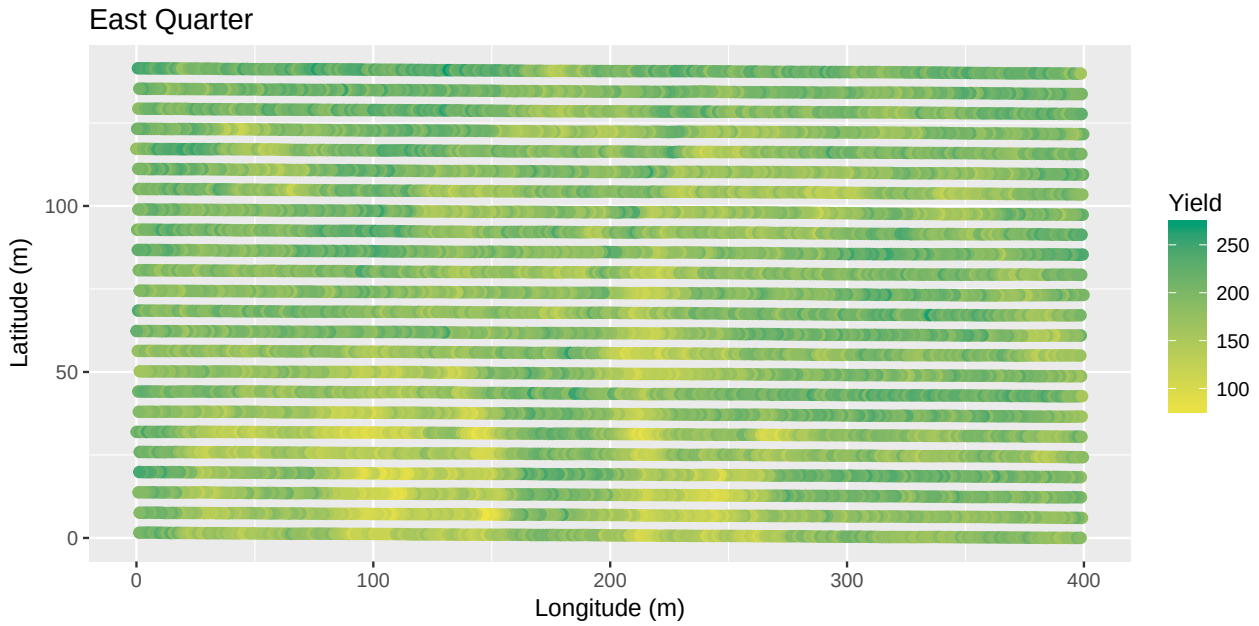
```
ggplot(EastQuarter.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Product),size=2) +
scale_colour_manual(values=cbPalette) +
labs(colour = "Variety", x="Longitude (m)", y="Latitude (m)", title = "East Quarter")
```



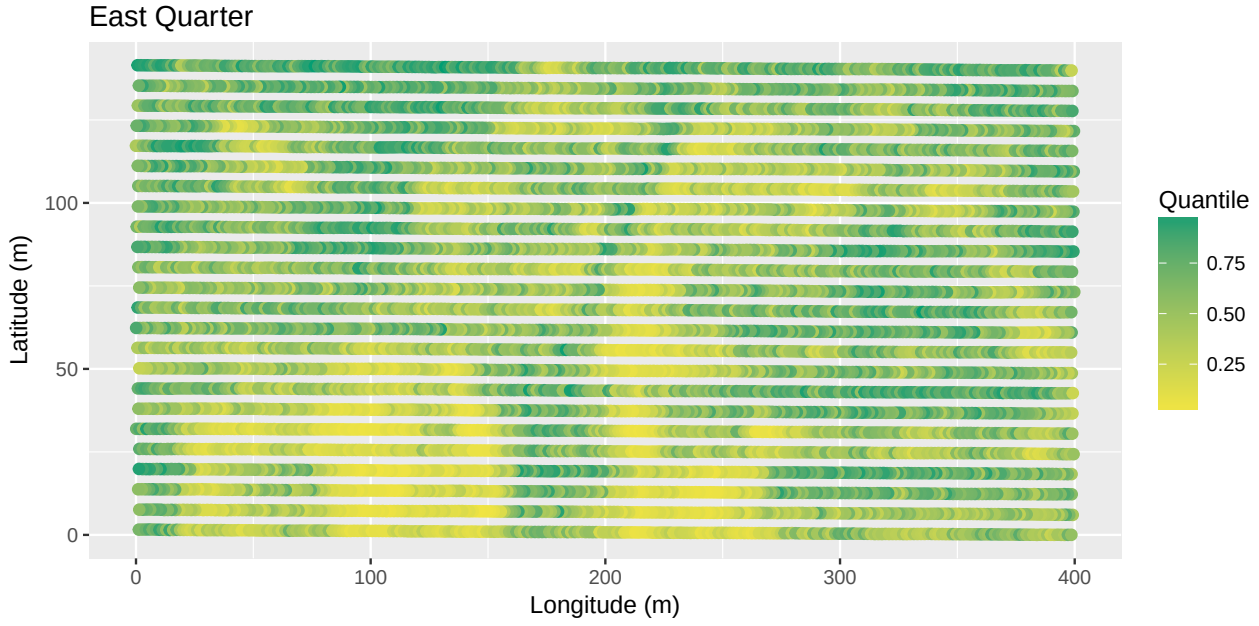

```
truehist(EastQuarter.dat$Yield)
```




```
ggplot(EastQuarter.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Yield),size=2) +
scale_colour_gradient(low=cbPalette[7], high=cbPalette[4]) +
labs(colour = "Yield", x="Longitude (m)", y="Latitude (m)", title = "East Quarter")
```




```
ggplot(EastQuarter.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Quantile),size=2) +
scale_colour_gradient(low=cbPalette[7], high=cbPalette[4]) +
labs(colour = "Quantile", x="Longitude (m)", y="Latitude (m)", title = "East Quarter")
```

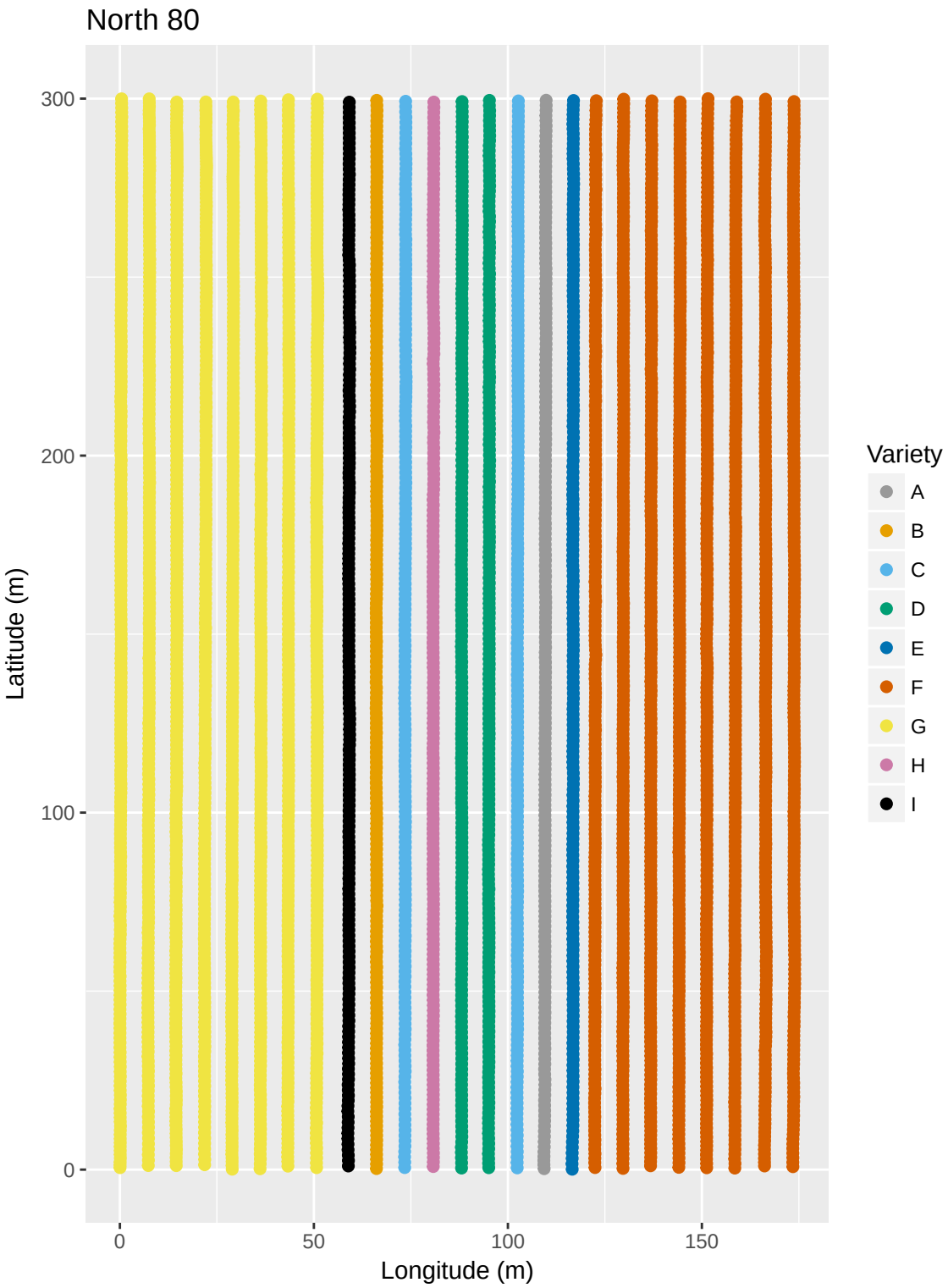


North

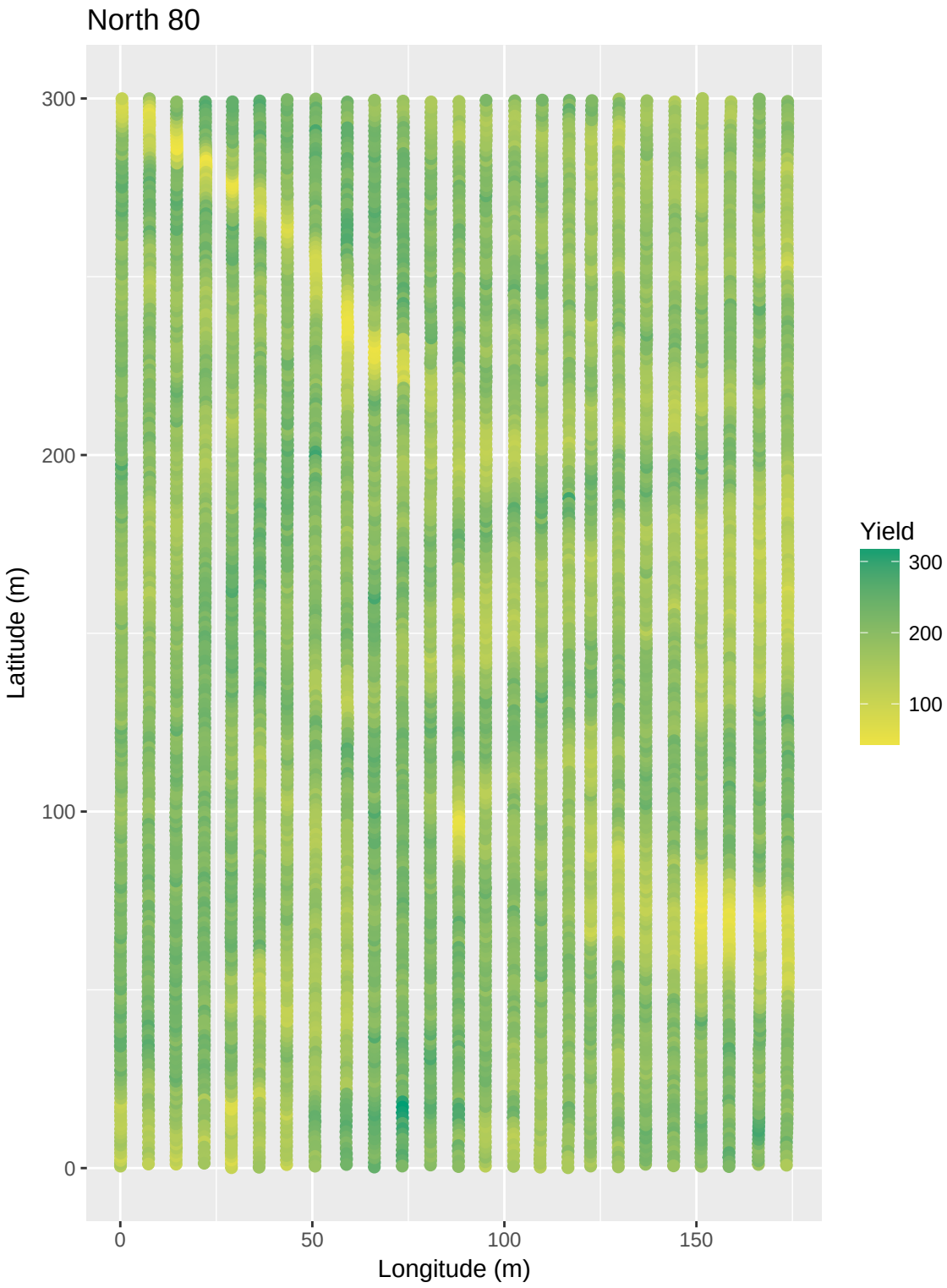
```
North80.dat$Yield <- North80.dat$YldVolDry
North80.dat <- standardize.field(North80.dat)

North80.dat <- subset(North80.dat,North80.dat$Northing>=25)
```

```
North80.dat$Northing <- North80.dat$Northing - min(North80.dat$Northing)
North80.dat <- subset(North80.dat,North80.dat$Northing<=300)

North80.dat <- subset(North80.dat,North80.dat$Easting>=740)
North80.dat <- subset(North80.dat,North80.dat$Easting<=920)

North80.dat$Easting <- North80.dat$Easting - min(North80.dat$Easting)

North80.dat$Product <- as.factor(as.character(North80.dat$Product))
```

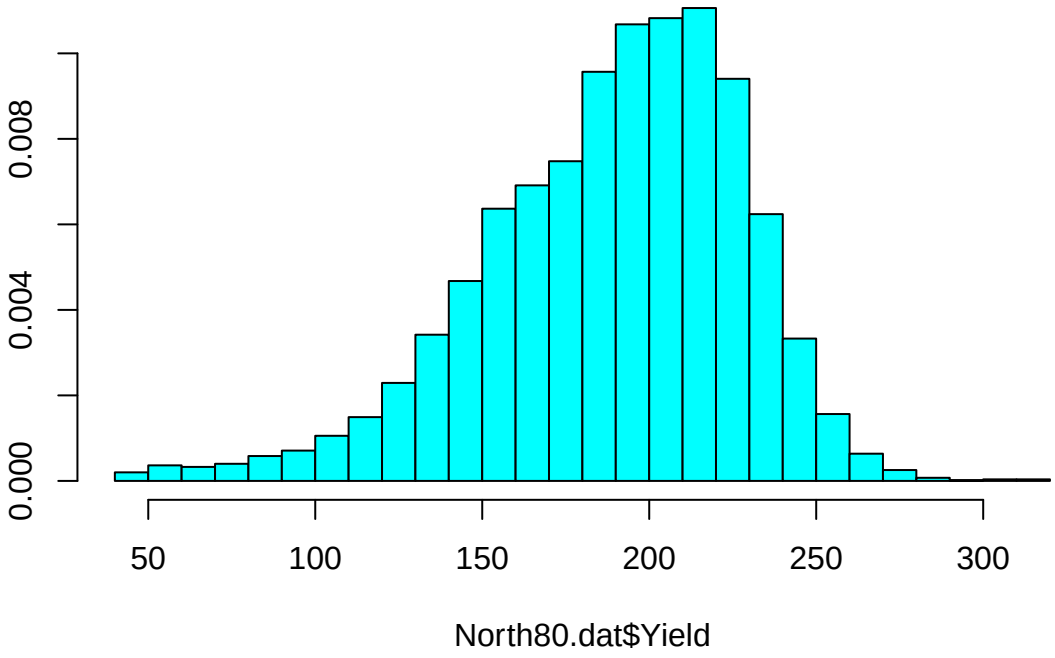
```
ggplot(North80.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Product),size=2) +
scale_colour_manual(values=cbPalette) +
labs(colour = "Variety", x="Longitude (m)", y="Latitude (m)", title = "North 80")
```



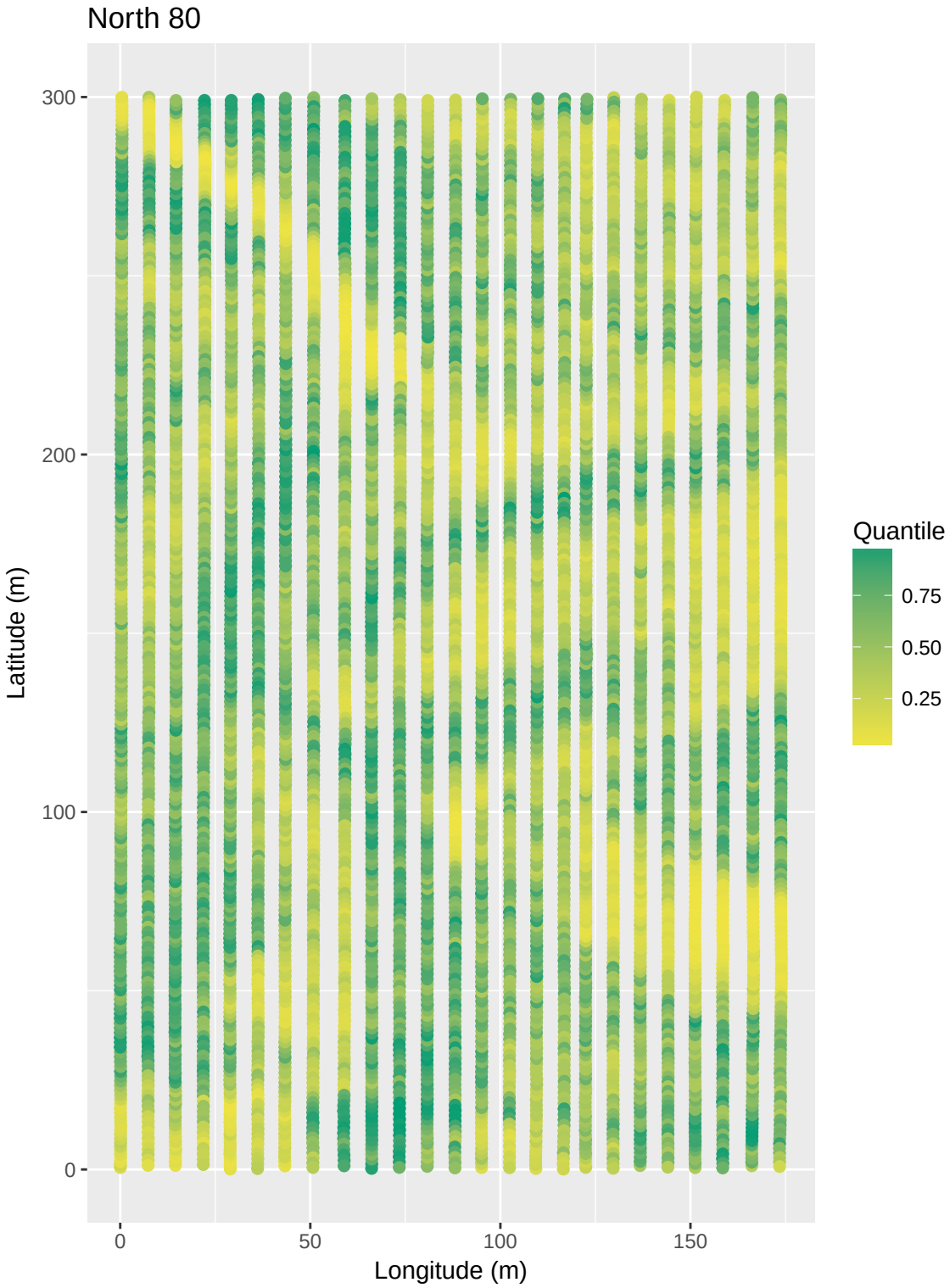

```
ggplot(North80.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Yield),size=2) +
scale_colour_gradient(low=cbPalette[7], high=cbPalette[4]) +
labs(colour = "Yield", x="Longitude (m)", y="Latitude (m)", title = "North 80")
```




```
truehist(North80.dat$Yield)
```




```
ggplot(North80.dat, aes(Easting,Northing)) +
geom_point(aes(colour = Quantile),size=2) +
scale_colour_gradient(low=cbPalette[7], high=cbPalette[4]) +
labs(colour = "Quantile", x="Longitude (m)", y="Latitude (m)", title = "North 80")
```




```
summary(aov(lm(Yield ~ Product,data=EastQuarter.dat)))
```

```
##               Df  Sum Sq Mean Sq F value Pr(>F)
## Product        1  372714  372714   427.1 <2e-16 ***
## Residuals   6367 5556751     873
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(lm(Quantile ~ Product,data=EastQuarter.dat)))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Product        1     26  25.999   478.4 <2e-16 ***
## Residuals   6367    346   0.054
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tapply(EastQuarter.dat$Yield,list(EastQuarter.dat$Product),mean)
```

```
##        B        E
## 192.3915 177.0918
```

```
tapply(EastQuarter.dat$Quantile,list(EastQuarter.dat$Product),mean)
```

```
##         B         E
## 0.5770214 0.4492382
```

```
summary(aov(lm(Yield ~ Product,data=North80.dat)))
```

```
##               Df  Sum Sq Mean Sq F value Pr(>F)
## Product        8  413676   51709   36.15 <2e-16 ***
## Residuals   5488 7850880    1431
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(lm(Quantile ~ Product,data=North80.dat)))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## Product        8   21.1   2.635    42.5 <2e-16 ***
## Residuals   5488  340.2   0.062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tapply(North80.dat$Yield,list(North80.dat$Product),mean)
```

```
##        A        B        C        D        E        F        G        H
## 191.3857 210.2030 199.0592 189.2594 186.8105 178.3901 193.2407 200.6945
##        I
## 182.3015
```

```
tapply(North80.dat$Quantile,list(North80.dat$Product),mean)
```

```
##         A         B         C         D         E         F         G
## 0.5029713 0.6711646 0.5734136 0.5119963 0.4626400 0.4359405 0.5440935
##         H         I
## 0.5854179 0.4772832
```

```
save(EastQuarter.dat,North80.dat,file="Strips.Rda")
```