Maize

Adama NDOUR

2023-12-11

## Import libraries

library(tidyverse)  
library(skimr)  
library(gtsummary)  
library(inspectdf)  
library(sf)

## Import data

df<- readxl::read\_excel("input/maize.xls",sheet = "MaizeDec7-2023\_final\_no\_IFDC")

## Overview of the data

str(df)

## tibble [3,065 × 34] (S3: tbl\_df/tbl/data.frame)  
## $ Plot\_ID : chr [1:3065] "MZ\_2317" "MZ\_2318" "MZ\_2319" "MZ\_2320" ...  
## $ ID : chr [1:3065] "OARI\_3" "OARI\_3" "OARI\_3" "OARI\_3" ...  
## $ MD : chr [1:3065] "H" "H" "H" "H" ...  
## $ Region : chr [1:3065] "Oromia" "Oromia" "Oromia" "Oromia" ...  
## $ Zone : chr [1:3065] "West Shewa" "West Shewa" "West Shewa" "West Shewa" ...  
## $ District : chr [1:3065] "Bako Tibe" "Bako Tibe" "Bako Tibe" "Bako Tibe" ...  
## $ Kebele : chr [1:3065] "Sadan Kite" "Sadan Kite" "Sadan Kite" "Sadan Kite" ...  
## $ Farmer\_name : chr [1:3065] "Ashim Feyisa" "Ashim Feyisa" "Ashim Feyisa" "Ashim Feyisa" ...  
## $ Year\_GC : num [1:3065] 2021 2021 2021 2021 2021 ...  
## $ Lat\_DD : num [1:3065] 9.1 9.1 9.1 9.1 9.1 ...  
## $ Long\_DD : num [1:3065] 37.2 37.2 37.2 37.2 37.2 ...  
## $ Alt\_m : num [1:3065] 1699 1699 1699 1699 1699 ...  
## $ Institute : chr [1:3065] "OARI" "OARI" "OARI" "OARI" ...  
## $ Center : chr [1:3065] "Bako" "Bako" "Bako" "Bako" ...  
## $ test\_crop : chr [1:3065] "Maize" "Maize" "Maize" "Maize" ...  
## $ Variety : chr [1:3065] "BH546" "BH546" "BH546" "BH546" ...  
## $ Soil\_table : chr [1:3065] "Nitisols" "Nitisols" "Nitisols" "Nitisols" ...  
## $ AEZ\_code : chr [1:3065] "SH32" "SH4" "SH5" "SH6" ...  
## $ Rep : num [1:3065] 1 1 1 1 1 1 1 1 1 1 ...  
## $ Treat : num [1:3065] 1 2 3 4 5 6 7 8 9 10 ...  
## $ N : num [1:3065] 120 120 120 120 120 120 0 120 0 120 ...  
## $ P : num [1:3065] 34.9 34.9 34.9 34.9 34.9 ...  
## $ K : num [1:3065] 49.8 49.8 49.8 49.8 0 ...  
## $ S : num [1:3065] 10.5 10.5 10.5 0 10.5 10.5 10.5 0 0 30 ...  
## $ Zn : num [1:3065] 5 5 0 5 5 5 5 0 0 0 ...  
## $ B : num [1:3065] 1 0 1 1 1 1 1 0 0 0 ...  
## $ GYld\_Kgpha : chr [1:3065] "6644.5714285714284" "5667.2571428571437" "6977.8285714285712" "7056.7200000000003" ...  
## $ All (Trt1) : chr [1:3065] "6644.5714285714284" "6644.5714285714284" "6644.5714285714284" "6644.5714285714284" ...  
## $ RecNP (Trt8) : chr [1:3065] "7081.7142857142853" "7081.7142857142853" "7081.7142857142853" "7081.7142857142853" ...  
## $ Yld. diff-All (%) : num [1:3065] 0 -14.71 5.02 6.2 7.73 ...  
## $ Yld. diff-RecNP (%): num [1:3065] -6.173 -19.973 -1.467 -0.353 1.076 ...  
## $ BMYld\_ tpha : num [1:3065] 12.9 11.8 14.2 12.7 13.3 ...  
## $ BM gpha : num [1:3065] 12889 11778 14222 12667 13333 ...  
## $ HI calculated : num [1:3065] 51.6 48.1 49.1 55.7 53.7 ...

## Data manipulation

# Rename some variables  
df<- df %>% rename\_with(  
 ~ c("all\_tr1","recNP\_tr8","diff\_all","diff\_recNP"),.cols= c("All (Trt1)","RecNP (Trt8)", "Yld. diff-All (%)", "Yld. diff-RecNP (%)"))  
  
# Convert to numeric  
df <- df %>% mutate(across(c("GYld\_Kgpha","all\_tr1","recNP\_tr8"),  
 as.numeric))

## Warning: There were 3 warnings in `mutate()`.  
## The first warning was:  
## ℹ In argument: `across(c("GYld\_Kgpha", "all\_tr1", "recNP\_tr8"), as.numeric)`.  
## Caused by warning:  
## ! NAs introduced by coercion  
## ℹ Run `dplyr::last\_dplyr\_warnings()` to see the 2 remaining warnings.

## Inspect data

* Categorical variables

df %>% select\_if(is.character) %>% skim()

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 3065 |
| Number of columns | 14 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 14 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| Plot\_ID | 25 | 0.99 | 7 | 7 | 0 | 3040 | 0 |
| ID | 25 | 0.99 | 6 | 8 | 0 | 103 | 0 |
| MD | 25 | 0.99 | 1 | 2 | 0 | 2 | 0 |
| Region | 25 | 0.99 | 4 | 17 | 0 | 4 | 0 |
| Zone | 25 | 0.99 | 3 | 14 | 0 | 13 | 0 |
| District | 25 | 0.99 | 4 | 14 | 0 | 23 | 0 |
| Kebele | 25 | 0.99 | 4 | 21 | 0 | 74 | 0 |
| Farmer\_name | 25 | 0.99 | 4 | 20 | 0 | 103 | 0 |
| Institute | 25 | 0.99 | 4 | 5 | 0 | 3 | 0 |
| Center | 25 | 0.99 | 4 | 11 | 0 | 12 | 0 |
| test\_crop | 25 | 0.99 | 5 | 5 | 0 | 1 | 0 |
| Variety | 25 | 0.99 | 4 | 10 | 0 | 19 | 0 |
| Soil\_table | 25 | 0.99 | 8 | 8 | 0 | 3 | 0 |
| AEZ\_code | 25 | 0.99 | 2 | 4 | 0 | 93 | 0 |

df %>% inspect\_cat()

## # A tibble: 14 × 5  
## col\_name cnt common common\_pcnt levels   
## <chr> <int> <chr> <dbl> <named list>   
## 1 AEZ\_code 94 SH3 29.4 <tibble [94 × 3]>   
## 2 Center 13 Adet 24.4 <tibble [13 × 3]>   
## 3 District 24 Dugda 11.5 <tibble [24 × 3]>   
## 4 Farmer\_name 104 A/Fita A/Fogi 0.979 <tibble [104 × 3]>   
## 5 ID 104 ARARI\_1 0.979 <tibble [104 × 3]>   
## 6 Institute 4 EIAR 42.8 <tibble [4 × 3]>   
## 7 Kebele 75 Tikur Balito 6.82 <tibble [75 × 3]>   
## 8 MD 3 H 79.0 <tibble [3 × 3]>   
## 9 Plot\_ID 3041 <NA> 0.816 <tibble [3,041 × 3]>  
## 10 Region 5 Oromia 51.6 <tibble [5 × 3]>   
## 11 Soil\_table 4 Nitisols 70.1 <tibble [4 × 3]>   
## 12 Variety 20 BH\_661 30.6 <tibble [20 × 3]>   
## 13 Zone 14 West Gojam 20.5 <tibble [14 × 3]>   
## 14 test\_crop 2 Maize 99.2 <tibble [2 × 3]>

# numeric variables  
df %>% select\_if(is.numeric) %>% skim() %>%   
 knitr::kable()

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| skim\_type | skim\_variable | n\_missing | complete\_rate | numeric.mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 | numeric.hist |
| numeric | Year\_GC | 25 | 0.9918434 | 2021.0000000 | 0.0000000 | 2021.0000 | 2021.00000 | 2021.000000 | 2021.000000 | 2021.00000 | ▁▁▇▁▁ |
| numeric | Lat\_DD | 25 | 0.9918434 | 9.5191725 | 1.5462249 | 7.1301 | 8.10810 | 9.031000 | 11.290000 | 11.79400 | ▆▇▃▃▇ |
| numeric | Long\_DD | 25 | 0.9918434 | 37.2310513 | 0.7517056 | 36.2493 | 36.75000 | 37.046100 | 37.547400 | 38.86110 | ▅▇▂▁▂ |
| numeric | Alt\_m | 25 | 0.9918434 | 1755.9513158 | 259.5964474 | 1054.0000 | 1671.00000 | 1795.000000 | 1936.000000 | 2123.00000 | ▂▁▅▇▇ |
| numeric | Rep | 24 | 0.9921697 | 1.9976981 | 0.8177681 | 1.0000 | 1.00000 | 2.000000 | 3.000000 | 3.00000 | ▇▁▇▁▇ |
| numeric | Treat | 1 | 0.9996737 | 5.5045692 | 2.8722919 | 1.0000 | 3.00000 | 6.000000 | 8.000000 | 10.00000 | ▇▇▇▇▇ |
| numeric | N | 25 | 0.9918434 | 92.6049342 | 52.0285710 | 0.0000 | 69.00000 | 115.000000 | 138.000000 | 138.00000 | ▃▁▂▃▇ |
| numeric | P | 25 | 0.9918434 | 36.8866453 | 30.2620610 | 0.0000 | 20.10000 | 30.100000 | 40.174672 | 92.00000 | ▃▇▂▁▃ |
| numeric | K | 25 | 0.9918434 | 31.3321815 | 25.8641715 | 0.0000 | 0.00000 | 49.788399 | 49.800000 | 60.00000 | ▆▁▁▁▇ |
| numeric | S | 25 | 0.9918434 | 9.2703947 | 8.3273306 | 0.0000 | 0.00000 | 10.500000 | 10.500000 | 30.00000 | ▅▇▁▁▁ |
| numeric | Zn | 25 | 0.9918434 | 2.9950658 | 2.4508949 | 0.0000 | 0.00000 | 5.000000 | 5.000000 | 5.00000 | ▆▁▁▁▇ |
| numeric | B | 25 | 0.9918434 | 0.5993421 | 0.4901124 | 0.0000 | 0.00000 | 1.000000 | 1.000000 | 1.00000 | ▆▁▁▁▇ |
| numeric | GYld\_Kgpha | 36 | 0.9882545 | 5263.4554718 | 2336.6345043 | 0.0000 | 3667.10000 | 5176.900000 | 6813.000000 | 12687.90000 | ▃▇▇▃▁ |
| numeric | all\_tr1 | 30 | 0.9902121 | 6252.7160317 | 2169.2446581 | 1298.3000 | 4774.56607 | 6000.000000 | 7780.000000 | 11633.70000 | ▂▆▇▅▂ |
| numeric | recNP\_tr8 | 18 | 0.9941272 | 5948.4751000 | 1959.4374929 | 0.0000 | 4644.70000 | 5878.862339 | 7356.500000 | 12687.90000 | ▁▇▇▃▁ |
| numeric | diff\_all | 51 | 0.9833605 | -13.4783428 | 32.1701197 | -100.0000 | -29.22243 | -8.099023 | 1.123781 | 167.85512 | ▂▇▂▁▁ |
| numeric | diff\_recNP | 54 | 0.9823817 | -10.7693895 | 33.5320567 | -100.0000 | -26.20379 | -4.870344 | 5.991677 | 378.49791 | ▇▇▁▁▁ |
| numeric | BMYld\_ tpha | 159 | 0.9481240 | 17.0537132 | 8.9573722 | 0.4000 | 11.11111 | 15.300000 | 21.700000 | 49.77778 | ▃▇▃▂▁ |
| numeric | BM gpha | 25 | 0.9918434 | 16302.0034424 | 9431.6015306 | 0.0000 | 10443.72149 | 14962.962963 | 21105.000000 | 49777.77778 | ▃▇▃▂▁ |
| numeric | HI calculated | 166 | 0.9458401 | 36.1258278 | 37.8666288 | 0.0000 | 24.63735 | 31.902914 | 41.075068 | 1439.75492 | ▇▁▁▁▁ |

## Descriptive statistics

* Yield per treatment

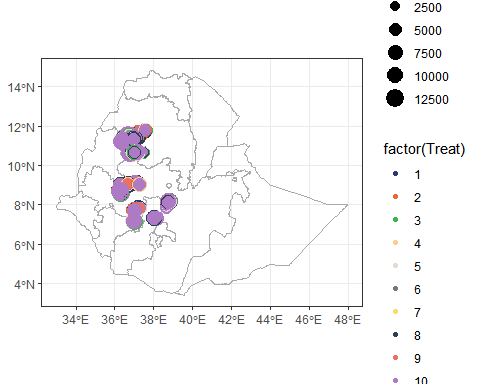
df %>% select(c("Treat","GYld\_Kgpha")) %>%   
 tbl\_summary(  
 by = Treat,  
 statistic = list(  
 all\_continuous() ~ "{mean} ({sd})",  
 all\_categorical() ~ "{n} / {N} ({p}%)"  
 ),  
 digits = all\_continuous() ~ 2,  
 missing\_text = "(Missing)"  
 ) %>%   
 modify\_spanning\_header(all\_stat\_cols() ~ "\*\*Treatments\*\*")

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Treatments | | | | | | | | | | |
| **Characteristic** | **1**,  N = 306 | **2**,  N = 306 | **3**,  N = 306 | **4**,  N = 304 | **5**,  N = 306 | **6**,  N = 308 | **7**,  N = 308 | **8**,  N = 306 | **9**,  N = 308 | **10**,  N = 306 |
| GYld\_Kgpha | 6,253.71 (2,171.25) | 6,069.55 (2,053.17) | 5,979.57 (2,081.14) | 5,912.83 (2,012.28) | 5,888.45 (1,990.17) | 4,465.66 (1,770.20) | 3,244.62 (1,880.14) | 5,970.46 (1,935.45) | 2,759.27 (1,718.02) | 6,066.15 (2,240.88) |
| Missing | 2 | 2 | 1 | 4 | 4 | 3 | 10 | 2 | 4 | 3 |

## Explory data analysis

* Map

maize\_areas<- st\_as\_sf(df %>% select(c("Lat\_DD","Long\_DD",  
 "Treat","GYld\_Kgpha")) %>%   
 drop\_na() %>% filter(GYld\_Kgpha>0),  
 coords=c("Long\_DD","Lat\_DD"), crs=4326)  
  
  
# Loxgb Ethiopia shapefile  
Eth <- sf::st\_read("input/eth\_adm\_2021\_shp", layer = "eth\_admbnda\_adm0\_csa\_bofedb\_itos\_2021", quiet = TRUE) %>%  
 st\_transform(4326)  
Eth\_regions<- sf::st\_read("input/eth\_adm\_2021\_shp", layer = "eth\_admbnda\_adm1\_csa\_bofedb\_2021", quiet = TRUE) %>%  
 st\_transform(4326)  
  
p<-ggplot() +  
 geom\_sf(data = Eth, fill = NA, color = "black") +  
 geom\_sf(data = Eth\_regions, fill = NA, color = "darkgrey")+  
 #geom\_sf\_label(data = Eth\_regions,aes(label=ADM1\_EN))+  
 geom\_sf(data=maize\_areas,aes(color=factor(Treat),size=GYld\_Kgpha), na.rm=T) +  
 scale\_color\_manual(values=c("#283C72","#E76A32","#3BB547","#FBCD93","#E4DAD1","#707B7C","#F7DC6F","#2C3E50","#EC7063","#AF7AC5"))+  
 coord\_sf()+  
 theme\_bw()  
ggsave(p, file="output/Eth\_maize\_yield\_map.png", width = 6, height = 4, dpi = 300)  
p



library(tmap)

## Breaking News: tmap 3.x is retiring. Please test v4, e.g. with  
## remotes::install\_github('r-tmap/tmap')

library(tmaptools)  
# Set tmap to view mode  
tmap\_mode("view")

## tmap mode set to interactive viewing

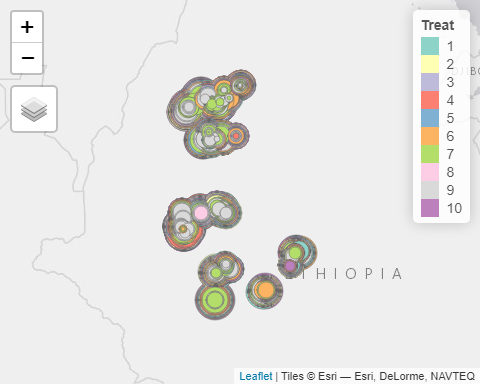
tm\_shape(Eth)+  
 tm\_borders(col = "black") +  
tm\_shape(Eth\_regions) +  
 tm\_borders(col = "gray")

## PhantomJS not found. You can install it with webshot::install\_phantomjs(). If it is installed, please make sure the phantomjs executable can be found via the PATH variable.



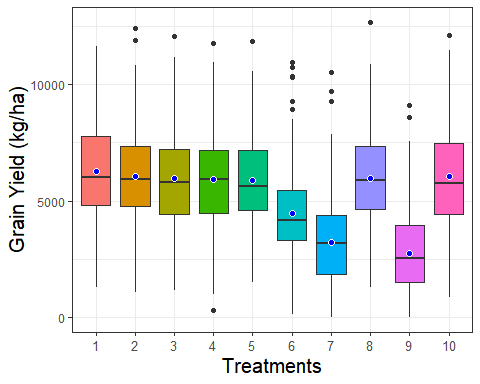
tm\_shape(maize\_areas %>% mutate(Treat=as.factor(Treat)))+  
 tm\_dots(size = "GYld\_Kgpha",col = "Treat")

## Legend for symbol sizes not available in view mode.



* Yield boxplots

df %>% drop\_na(GYld\_Kgpha) %>%   
 ggplot(aes(factor(Treat),GYld\_Kgpha, fill=factor(Treat)))+  
 geom\_boxplot() +  
 stat\_summary(fun = mean, geom = "point", shape = 21, size = 2, color = "white",fill="blue")+  
 theme\_bw() +  
 theme(axis.text.x = element\_text(size = 10),   
 axis.title.y = element\_text(size = 15),  
 axis.title.x = element\_text(size = 15),  
 legend.position = 'none') +  
 labs(y="Grain Yield (kg/ha)",x="Treatments")



## Analysis of Variance (ANOVA)

Let’s check the assumption

# Homogeinty of variance  
bartlett.test(GYld\_Kgpha ~ Treat, data = df)

##   
## Bartlett test of homogeneity of variances  
##   
## data: GYld\_Kgpha by Treat  
## Bartlett's K-squared = 37.916, df = 9, p-value = 1.804e-05

p<0.05 so we can reject the NULL hypothesis which stipulates that there are significant differences of variances accross treatments.

* Robust ANOVA

# Perform robust ANOVA using Welch's test  
welch\_result <- oneway.test(GYld\_Kgpha ~ Treat, data = df, var.equal = FALSE)  
print(welch\_result)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: GYld\_Kgpha and Treat  
## F = 145.58, num df = 9.0, denom df = 1229.5, p-value < 2.2e-16

p of the treatment is statistically significant (p < 0.05), it is likely that treatment type does have a significant effect on average crop yield. In fact there is statistically-significant difference in average crop yield according to treatment.